
Get Full Access and More at
ExpertConsult.com
GOLDMAN-CECIL MEDICINE
LEE GOLDMAN
ANDREW I. SCHAFER
VOLUME 1
25TH EDITION
CROW | DOROSHOW | DRAZEN | GRIGGS | LANDRY
LEVINSON | RUSTGI | SCHELD | SPIEGEL
Cecil
ELSEVIER
GOLDMAN-CECIL MEDICINE
GOLDMAN | SCHAFER
VOLUME 1
25TH EDITION
PAGES: 1-1338
Cecil
ELSEVIER
GOLDMAN-CECIL MEDICINE
GOLDMAN | SCHAFER
VOLUME 2
25TH EDITION
PAGES: 1339-2038
Cecil
ELSEVIER

QUICK REFERENCE (QR) VIDEO ACCESS

The images below are QR codes. Each code corresponds to a video from the *Goldman-Cecil Medicine 25* collection. For fast and easy video access, right from your mobile device, follow these instructions. The videos are also available on Expertconsult.com.

What You Need

- **A mobile device, such as a smartphone or tablet, equipped with a camera and Internet access**
- **A QR code reader application (If you do not already have a reader installed on your mobile device, look for free versions in your app store.)**

How It Works

- **Open the QR code reader application on your mobile device.**
- **Point the device's camera at the code and scan.**
- **Each code opens an individual video player for instant viewing—no log-on required.**

Confusion Assessment Method (CAM)
Chapter 28, Video 28-1 – Marcos Mialnez, Jorge G. Ruiz, and Rosanne M. Leipzig

Interlaminar Epidural Steroid Injection
Chapter 30, Video 30-1 – Ali Turabi

Standard Echocardiographic Views: Long Axis Image Plane
Chapter 55, Video 55-1A – Catherine M. Otto

Standard Echocardiographic Views: Short Axis Image Plane
Chapter 55, Video 55-1B – Catherine M. Otto

Standard Echocardiographic Views: Short Axis Image Plane
Chapter 55, Video 55-1C – Catherine M. Otto

Standard Echocardiographic Views: Four-Chamber Image Plane
Chapter 55, Video 55-1D – Catherine M. Otto

Dilated Cardiomyopathy: Long Axis View
Chapter 55, Video 55-2A – Catherine M. Otto

Dilated Cardiomyopathy: Short Axis View
Chapter 55, Video 55-2B – Catherine M. Otto

Dilated Cardiomyopathy: Apical Four-Chamber View
Chapter 55, Video 55-2C – Catherine M. Otto

Three-Dimensional Echocardiography
Chapter 55, Video 55-3 – Catherine M. Otto

Stress Echocardiography: Normal Reaction
Chapter 55, Video 55-4A – Catherine M. Otto

Stress Echocardiography: Normal Reaction
Chapter 55, Video 55-4B – Catherine M. Otto

Stress Echocardiography: Proximal Stenosis of the Left Anterior Descending Coronary Artery
Chapter 55, Video 55-4C – Catherine M. Otto

Stress Echocardiography: Proximal Stenosis of the Left Anterior Descending Coronary Artery
Chapter 55, Video 55-4D – Catherine M. Otto

Pericardial Effusion: Parasternal Long Axis
Chapter 55, Video 55-5A – Catherine M. Otto

Pericardial Effusion: Parasternal Short Axis
Chapter 55, Video 55-5B – Catherine M. Otto

Pericardial Effusion: Apical Four-Chamber Views
Chapter 55, Video 55-5C – Catherine M. Otto

Secundum Atrial Septal Defect
Chapter 69, Video 69-1 – Ariane J. Marelli

Perimembranous Ventricular Septal Defect
Chapter 69, Video 69-2 – Ariane J. Marelli

Coronary Stent Placement
Chapter 74, Video 74-1 – Paul S. Teirstein

Guidewire Passage
Chapter 74, Video 74-2 – Paul S. Teirstein

Delivering the Stent
Chapter 74, Video 74-3 – Paul S. Teirstein

Inflating the Stent
Chapter 74, Video 74-4 – Paul S. Teirstein

Final Result
Chapter 74, Video 74-5 – Paul S. Teirstein

Superficial Femoral Artery (SFA) Stent Procedure
Chapter 79, Video 79-1 – Christopher J. White

Orthotopic Bicaval Cardiac Transplantation
Chapter 82, Video 82-1 – Y. Joseph Woo

Wheezing
Chapter 87, Video 87-1 – Jeffrey M. Drazen

VATS Wedge Resection
Chapter 101, Video 101-1 – Malcolm M. DeCamp

Ventilation of an Ex Vivo Rat Lung
Chapter 105, Video 105-1 – Arthur S. Slutsky, George Volgyesi, and Tom Whitehead

Renal Artery Stent
Chapter 125, Video 125-1 – Renato M. Santos and Thomas D. DuBose, Jr.

Interpretation of a Computed Tomographic Colonography
Chapter 133, Video 133-1 – David H. Kim

Donor Liver Transportation–Donor and Recipient
Chapter 154, Video 154-1 – Igal Kam, Thomas Bak, and Michael Wachs

Snare Polypectomy of a Colon Adenoma
Chapter 193, Video 193-1 – Douglas O. Faigel

Laparascopic-Assisted Double Balloon Enteroscopy with Polypectomy of a Jejunal Adenoma Followed by Surgical Oversew of the Polypectomy Site
Chapter 193, Video 193-2 – Douglas O. Faigel

Endoscopic Mucosal Resection Using Saline Lift Polypectomy of a Colon Adenoma Followed by Closure of the Mucosal Defect with Clips
Chapter 193, Video 193-3 – Douglas O. Faigel

Endoscopic View of Rectal Cancer
Chapter 193, Video 193-4 – Douglas O. Faigel

Endoscopic Ultrasound
Chapter 193, Video 193-5 – Douglas O. Faigel

Laparoscopic Roux-en-Y Gastric Bypass
Chapter 220, Video 220-1 – James M. Swain

Pituitary Surgery
Chapter 224, Video 224-1 – Ivan Ciric

Skin Testing
Chapter 251, Video 251-1 – Larry Borish

Nasal Endoscopy
Chapter 251, Video 251-2 – Larry Borish

Hip Arthroscopy Osteochondroplasty
Chapter 276, Video 276-1 – Bryan T. Kelly

Cervical Provocation
Chapter 400, Video 400-1 – Richard L. Barbano

Spurling Maneuver
Chapter 400, Video 400-2 – Richard L. Barbano

Cervical Distraction Test
Chapter 400, Video 400-3 – Richard L. Barbano

Straight Leg Raise
Chapter 400, Video 400-4 – Richard L. Barbano

Contralateral Straight Leg Raise
Chapter 400, Video 400-5 – Richard L. Barbano

Seated Straight Leg Raise
Chapter 400, Video 400-6 – Richard L. Barbano

Discectomy
Chapter 400, Video 400-7 – Jason H. Huang

Absence Seizure
Chapter 403, Video 403-1 – Samuel Wiebe

Left Rolandic Seizure
Chapter 403, Video 403-2 – Samuel Wiebe

Left Temporal Complex Partial Seizure
Chapter 403, Video 403-3 – Samuel Wiebe

Left Temporal Complex Partial Seizure Postictal Confusion
Chapter 403, Video 403-4 – Samuel Wiebe

Left Temporal Complex Partial Seizure
Chapter 403, Video 403-5 – Samuel Wiebe

Supplementary Sensory-Motor Seizure
Chapter 403, Video 403-6 – Samuel Wiebe

Right Posterior Temporal Seizure - Dramatic Frontal Semiology
Chapter 403, Video 403-7 – Samuel Wiebe

Right Mesial Frontal Seizure
Chapter 403, Video 403-8 – Samuel Wiebe

Nonconvulsive Status Epilepticus
Chapter 403, Video 403-9 – Samuel Wiebe

GTC Seizure Tonic Phase
Chapter 403, Video 403-10 – Samuel Wiebe

GTC Seizure Clonic Phase
Chapter 403, Video 403-11 – Samuel Wiebe

Myoclonic Facial Seizure
Chapter 403, Video 403-12 – Samuel Wiebe

Tonic Seizure Lennox Gastaut
Chapter 403, Video 403-13 – Samuel Wiebe

Atonic Seizure Lennox Gastaut
Chapter 403, Video 403-14 – Samuel Wiebe

Reflex Auditory Seizure
Chapter 403, Video 403-15 – Samuel Wiebe

Four Score
Chapter 404, Video 404-1 – James L. Bernat and Eelco F. M. Wijdicks

Persistent Vegetative State
Chapter 404, Video 404-2 – James L. Bernat and Eelco F. M. Wijdicks

Minimally Conscious State
Chapter 404, Video 404-3 – James L. Bernat and Eelco F. M. Wijdicks

Akinetic Mutism
Chapter 404, Video 404-4 – James L. Bernat and Eelco F. M. Wijdicks

Early Parkinson's Disease
Chapter 409, Video 409-1 – Anthony E. Lang

Freezing of Gait in Parkinson's Disease
Chapter 409, Video 409-2 – Anthony E. Lang

Gunslinger Gait in Progressive Supranuclear Palsy
Chapter 409, Video 409-3 – Anthony E. Lang

Supranuclear Gaze Palsy in Progressive Supranuclear Palsy
Chapter 409, Video 409-4 – Anthony E. Lang

Applause Sign in Progressive Supranuclear Palsy
Chapter 409, Video 409-5 – Anthony E. Lang

Apraxia of Eyelid Opening in Progressive Supranuclear Palsy
Chapter 409, Video 409-6 – Anthony E. Lang

Cranial Dystonia in Multiple System Atrophy
Chapter 409, Video 409-7 – Anthony E. Lang

Anterocollis in Multiple System Atrophy
Chapter 409, Video 409-8 – Anthony E. Lang

Stridor in Multiple System Atrophy
Chapter 409, Video 409-9 – Anthony E. Lang

Alien Limb Phenomenon in Corticobasal Syndrome
Chapter 409, Video 409-10 – Anthony E. Lang

Myoclonus in Corticobasal Syndrome
Chapter 409, Video 409-11 – Anthony E. Lang

Levodopa-Induced Dyskinesia in Parkinson's Disease
Chapter 409, Video 409-12 – Anthony E. Lang

Essential Tremor
Chapter 410, Video 410-1 – Anthony E. Lang

Huntington's Disease
Chapter 410, Video 410-2 – Anthony E. Lang

Hemiballism
Chapter 410, Video 410-3 – Anthony E. Lang

Blepharospasm
Chapter 410, Video 410-4 – Anthony E. Lang

Oromandibular Dystonia
Chapter 410, Video 410-5 – Anthony E. Lang

Cervical Dystonia
Chapter 410, Video 410-6 – Anthony E. Lang

Writer's Cramp
Chapter 410, Video 410-7 – Anthony E. Lang

Embouchure Dystonia
Chapter 410, Video 410-8 – Anthony E. Lang

Sensory Trick in Cervical Dystonia
Chapter 410, Video 410-9 – Anthony E. Lang

Generalized Dystonia
Chapter 410, Video 410-10 – Anthony E. Lang

Tics
Chapter 410, Video 410-11 – Anthony E. Lang

Tardive Dyskinesia
Chapter 410, Video 410-12 – Anthony E. Lang

Hemifacial Spasm
Chapter 410, Video 410-13 – Anthony E. Lang

Wernickes Encephalopathy Eye Movements: Before Thiamine
Chapter 416, Video 416-1 – Barbara S. Koppel

Wernickes Encephalopathy Eye Movements: After Thiamine
Chapter 416, Video 416-2 – Barbara S. Koppel

Limb Symptoms and Signs
Chapter 419, Video 419-1 – Pamela J. Shaw

Bulbar Symptoms and Signs
Chapter 419, Video 419-2 – Pamela J. Shaw

Normal Swallowing
Chapter 419, Video 419-3 – Pamela J. Shaw

Charcot-Marie-Tooth Disease Exam and Walk
Chapter 420, Video 420-1 – Michael E. Shy

GOLDMAN-CECIL MEDICINE

GOLDMAN-CECIL MEDICINE

25TH EDITION

Volume I

EDITED BY

LEE GOLDMAN, MD

Harold and Margaret Hatch Professor
Executive Vice President and Dean of the
Faculties of Health Sciences and Medicine
Chief Executive, Columbia University Medical Center
Columbia University
New York, New York

ANDREW I. SCHAFER, MD

Professor of Medicine
Director, Richard T. Silver Center for Myeloproliferative Neoplasms
Weill Cornell Medical College
New York, New York

ELSEVIER
SAUNDERS

ELSEVIER
SAUNDERS

1600 John F. Kennedy Blvd.
Ste. 1800
Philadelphia, PA 19103-2899

GOLDMAN-CECIL MEDICINE, 25TH EDITION

ISBN: 978-1-4557-5017-7
Volume 1 Part Number: 9996096564
Volume 2 Part Number: 9996096629

International Edition (IE):

ISBN: 978-0-323-28800-2
IE Volume 1 Part Number: 9996118347
IE Volume 2 Part Number: 9996118282

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

With respect to any drug or pharmaceutical products identified, readers are advised to check the most current information provided (i) on procedures featured or (ii) by the manufacturer of each product to be administered, to verify the recommended dose or formula, the method and duration of administration, and contraindications. It is the responsibility of practitioners, relying on their own experience and knowledge of their patients, to make diagnoses, to determine dosages and the best treatment for each individual patient, and to take all appropriate safety precautions.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
Goldman's Cecil medicine.
Goldman-Cecil medicine / [edited by] Lee Goldman, Andrew I. Schafer.—25th edition.
p. ; cm.
Cecil medicine
Preceded by Goldman's Cecil medicine / [edited by] Lee Goldman, Andrew I. Schafer. 24th ed. c2012.
Includes bibliographical references.
ISBN 978-1-4557-5017-7 (hardcover, 2 vol set : alk. paper)—ISBN 978-0-323-28800-2 (international edition : alk. paper)—ISBN 978-9996096563 (volume 1 : alk. paper)—ISBN 9996096564 (volume 1 : alk. paper)—ISBN 978-9996096624 (volume 2 : alk. paper)—ISBN 9996096629 (volume 2 : alk. paper)
I. Goldman, Lee (Physician), editor. II. Schafer, Andrew I., editor. III. Title. IV. Title: Cecil medicine.
[DNLM: 1. Medicine. WB 100]
RC46
616—dc23

2014049904

Executive Content Strategist: Kate Dimock
Senior Content Development Manager: Maureen Iannuzzi
Publishing Services Manager: Anne Altepeter
Senior Project Manager: Cindy Thoms
Design Specialist: Paula Catalano

Printed in the United States of America

Last digit is the print number: 9 8 7 6 5 4 3 2 1

ASSOCIATE EDITORS

Mary K. Crow, MD
Joseph P. Routh Professor of Rheumatic Diseases in Medicine
Weill Cornell Medical College
Physician-in-Chief and Benjamin M. Rosen Chair in Immunology and Inflammation Research
Hospital for Special Surgery
New York, New York

James H. Doroshow, MD
Bethesda, Maryland

Jeffrey M. Drazen, MD
Distinguished Parker B. Francis Professor of Medicine
Harvard Medical School
Senior Physician
Brigham and Women's Hospital
Boston, Massachusetts

Robert C. Griggs, MD
Professor of Neurology, Medicine, Pediatrics, and Pathology and Laboratory Medicine
University of Rochester School of Medicine and Dentistry
Rochester, New York

Donald W. Landry, MD, PhD
Samuel Bard Professor of Medicine
Chair, Department of Medicine
Physician-in-Chief
Columbia University Medical Center
New York, New York

Wendy Levinson, MD
Professor of Medicine
Chair Emeritus
Department of Medicine
University of Toronto
Toronto, Ontario, Canada

Anil K. Rustgi, MD
T. Grier Miller Professor of Medicine and Genetics
Chief of Gastroenterology
American Cancer Society Professor
University of Pennsylvania Perelman School of Medicine
Philadelphia, Pennsylvania

W. Michael Scheld, MD
Bayer-Gerald L. Mandell Professor of Infectious Diseases
Professor of Medicine
Clinical Professor of Neurosurgery
Director, Pfizer Initiative in International Health
University of Virginia Health System
Charlottesville, Virginia

Allen M. Spiegel, MD
Dean
Albert Einstein College of Medicine
Bronx, New York

PREFACE

In the 90 years since the first edition of the *Cecil Textbook of Medicine* was published, almost everything we know about internal medicine has changed. Progress in medical science is now occurring at an ever-accelerating pace, and it is doing so within the framework of transformational changes in clinical practice and the delivery of health care at individual, social, and global levels. This textbook and its associated electronic products incorporate the latest medical knowledge in multiple formats that should appeal to students and seasoned practitioners regardless of how they prefer to access this rapidly changing information.

Even as *Cecil*'s specific information has changed, however, we have remained true to the tradition of a comprehensive textbook of medicine that carefully explains the *why* (the underlying pathophysiology of disease) and the *how* (now expected to be evidence-based from randomized controlled trials and meta-analyses). Descriptions of physiology and pathophysiology include the latest genetic advances in a practical format that strives to be useful to the nonexpert. Medicine has entered an era when the acuity of illness and the limited time available to evaluate a patient have diminished the ability of physicians to satisfy their intellectual curiosity. As a result, the acquisition of information, quite easily achieved in this era, is often confused with knowledge. We have attempted to address this dilemma with a textbook that not only informs but also stimulates new questions and gives a glimpse of the future path to new knowledge. Grade A evidence is specifically highlighted in the text and referenced at the end of each chapter. In addition to the information provided in the textbook, the *Cecil* website supplies expanded content and functionality. In many cases, the full articles referenced in each chapter can be accessed from the *Cecil* website. The website is also continuously updated to incorporate subsequent Grade A information, other evidence, and new discoveries.

The sections for each organ system begin with a chapter that summarizes an approach to patients with key symptoms, signs, or laboratory abnormalities associated with dysfunction of that organ system. As summarized in E-Table 1-1, the text specifically provides clear, concise information regarding how a physician should approach more than 100 common symptoms, signs, and laboratory abnormalities, usually with a flow diagram, a table, or both for easy reference. In this way, *Cecil* remains a comprehensive text to guide diagnosis and therapy, not only for patients with suspected or known diseases but also for patients who may have undiagnosed abnormalities that require an initial evaluation.

Just as each edition brings new authors, it also reminds us of our gratitude to past editors and authors. Previous editors of *Cecil* include a short but remarkably distinguished group of leaders of American medicine: Russell Cecil, Paul Beeson, Walsh McDermott, James Wyngaarden, Lloyd H. Smith, Jr., Fred Plum, J. Claude Bennett, and Dennis Ausiello. As we welcome new associate editors—Mary K. Crow, James H. Doroshow, and Allen M. Spiegel—we also express our appreciation to William P. Arend, James O. Armitage, David R. Clemmons, and other associate editors from the previous editions on whose foundation we have built. Our returning associate editors—Jeffrey M. Drazen, Robert C. Griggs, Donald W. Landry, Wendy Levinson, Anil K. Rustgi, and W. Michael Scheld—continue to make critical contributions to the selection of authors and the review and approval of all manuscripts. The editors, however, are fully responsible for the book as well as the integration among chapters.

The tradition of *Cecil* is that all chapters are written by distinguished experts in each field. We are also most grateful for the editorial assistance in New York of Maribel Lim and Silva Sergenian. These individuals and others in our offices have shown extraordinary dedication and equanimity in working with authors and editors to manage the unending flow of manuscripts, figures, and permissions. We also thank Cassondra Andreychik, Ved Bhushan Arya, Cameron Harrison, Karen Krok, Robert J. Mentz, Gaétane Nocturne, Patrice Savard, Senthil Senniappan, Tejpratap Tiwari, and Sangeetha Venkatarajan, who contributed to various chapters, and we mourn the passing of Morton N. Swartz, MD, co-author of the chapter on "Meningitis: Bacterial, Viral, and Other" and Donald E. Low, MD, author of the chapter "Nonpneumococcal Streptococcal Infections, Rheumatic Fever." At Elsevier, we are most indebted to Kate Dimock and Maureen Iannuzzi, and also thank Maria Holman, Gabriela Benner, Cindy Thoms, Anne Altepeter, Linda McKinley, Paula Catalano, and Kristin Koehler, who have been critical to the planning and production process under the guidance of Mary Gatsch. Many of the clinical photographs were supplied by Charles D. Forbes and William F. Jackson, authors of *Color Atlas and Text of Clinical Medicine*, Third Edition, published in 2003 by Elsevier Science Ltd. We thank them for graciously permitting us to include their pictures in our book. We have been exposed to remarkable physicians in our lifetimes and would like to acknowledge the mentorship and support of several of those who exemplify this paradigm—Eugene Braunwald, Lloyd H. Smith, Jr., Frank Gardner, and William Castle. Finally, we would like to thank the Goldman family—Jill, Jeff, Abigail, Mira, Samuel, Daniel, Robyn, Tobin, and Dashel—and the Schafer family—Pauline, Eric, Melissa, Nathaniel, Pam, John, Evan, Samantha, Kate, and Sean, for their understanding of the time and focus required to edit a book that attempts to sustain the tradition of our predecessors and to meet the needs of today's physician.

LEE GOLDMAN, MD
ANDREW I. SCHAFER, MD

CONTRIBUTORS

Charles S. Abrams, MD
Professor of Medicine, Pathology, and Laboratory Medicine, University of Pennsylvania School of Medicine; Director, PENN-Chop Blood Center for Patient Care & Discovery, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania
Thrombocytopenia

Frank J. Accurso, MD
Professor of Pediatrics, University of Colorado School of Medicine; Attending Physician, Children's Hospital Colorado, Aurora, Colorado
Cystic Fibrosis

Ronald S. Adler, MD, PhD
Professor of Radiology, New York University School of Medicine; Department of Radiology, NYU Langone Medical Center, New York, New York
Imaging Studies in the Rheumatic Diseases

Cem Akin, MD, PhD
Associate Professor, Harvard Medical School; Attending Physician, Director, Mastocytosis Center, Brigham and Women's Hospital, Department of Medicine, Division of Rheumatology, Immunology, and Allergy, Boston, Massachusetts
Mastocytosis

Allen J. Aksamit, Jr., MD
Professor of Neurology, Mayo Clinic College of Medicine, Consultant in Neurology, Mayo Clinic, Rochester, Minnesota
Acute Viral Encephalitis

Qais Al-Awqati, MB ChB
Robert F. Loeb Professor of Medicine, Jay I. Meltzer Professor of Nephrology and Hypertension, Professor of Physiology and Cellular Biophysics, Division of Nephrology, Columbia University, College of Physicians and Surgeons, New York, New York
Structure and Function of the Kidneys

Ban Mishu Allos, MD
Associate Professor of Medicine, Division of Infectious Diseases, Associate Professor, Preventive Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee
Campylobacter Infections

David Altshuler, MD, PhD
Professor of Genetics and of Medicine, Harvard Medical School, Massachusetts General Hospital; Professor of Biology (Adjunct), Massachusetts Institute of Technology, Boston and Cambridge, Massachusetts
The Inherited Basis of Common Diseases

Michael Aminoff, MD, DSc
Professor, Department of Neurology, University of California San Francisco, San Francisco, California
Approach to the Patient with Neurologic Disease

Jeffrey L. Anderson, MD
Professor of Internal Medicine, University of Utah School of Medicine; Vice-Chair for Research, Department of Internal Medicine, Associate Chief of Cardiology and Director of Cardiovascular Research, Intermountain Medical Center, Intermountain Healthcare, Salt Lake City, Utah
ST Segment Elevation Acute Myocardial Infarction and Complications of Myocardial Infarction

Larry J. Anderson, MD
Professor, Division of Infectious Disease, Department of Pediatrics, Emory University School of Medicine and Children's Healthcare of Atlanta, Atlanta, Georgia
Coronaviruses

Aśok C. Antony, MD
Chancellor's Professor of Medicine, Indiana University School of Medicine; Attending Physician, Indiana University Health Affiliated Hospitals and Richard L. Roudebush Veterans Affairs Medical Center, Indianapolis, Indiana
Megaloblastic Anemias

Gerald B. Appel, MD
Professor of Medicine, Division of Nephrology, Department of Medicine, Columbia University College of Physicians and Surgeons, New York, New York
Glomerular Disorders and Nephrotic Syndromes

Frederick R. Appelbaum, MD
Executive Vice President and Deputy Director, Fred Hutchinson Cancer Research Center; President, Seattle Cancer Care Alliance; Professor, Division of Medical Oncology, University of Washington School of Medicine, Seattle Washington
The Acute Leukemias

Suneel S. Apte, MBBS, DPhil
Staff, Cleveland Clinic Lerner College of Medicine at Case Western Reserve University, Cleveland, Ohio
Connective Tissue Structure and Function

James O. Armitage, MD
The Joe Shapiro Professor of Internal Medicine, University of Nebraska Medical Center, Omaha, Nebraska
Approach to the Patient with Lymphadenopathy and Splenomegaly; Non-Hodgkin Lymphomas

M. Amin Arnaout, MD
Professor of Medicine, Departments of Medicine and Developmental and Regenerative Biology, Harvard Medical School; Physician and Chief Emeritus, Division of Nephrology, Massachusetts General Hospital, Boston, Massachusetts
Cystic Kidney Diseases

Robert M. Arnold, MD
Leo H. Criep Professor of Clinical Care, Chief, Section of Palliative Care and Medical Ethics, University of Pittsburgh; Medical Director, UPMC Palliative and Supportive Care Institute, Pittsburgh, Pennsylvania
Care of Dying Patients and Their Families

David Atkins, MD, MPH
Director, Health Services Research and Development, Veterans Health Administration, Washington, D.C.
The Periodic Health Examination

John P. Atkinson, MD
Chief, Division of Rheumatology, Internal Medicine, Washington University School of Medicine in St. Louis, St. Louis, Missouri
Complement System in Disease

Bruce R. Bacon, MD
Endowed Chair in Gastroenterology, Professor of Internal Medicine, Co-Director, Saint Louis University Liver Center; Director, Saint Louis University Abdominal Transplant Center, Saint Louis University School of Medicine, St. Louis, Missouri
Iron Overload (Hemochromatosis)

Larry M. Baddour, MD
Professor of Medicine, Chair, Division of Infectious Diseases, Mayo Clinic, Rochester, Minnesota
Infective Endocarditis

Grover C. Bagby, MD
Professor of Medicine and Molecular and Medical Genetics, Knight Cancer Institute at Oregon Health and Science University and Portland VA Medical Center, Portland, Oregon
Aplastic Anemia and Related Bone Marrow Failure States

Barbara J. Bain, MBBS
Professor in Diagnostic Haematology, Imperial College London; Honorary Consultant Haematologist, St. Mary's Hospital, London, United Kingdom
The Peripheral Blood Smear

Dean F. Bajorin, MD
Attending Physician and Member, Medicine, Memorial Hospital, Memorial Sloan Kettering Cancer Center; Professor of Medicine, Weill Cornell Medical College, New York, New York
Tumors of the Kidney, Bladder, Ureters, and Renal Pelvis

Robert W. Baloh, MD
Professor of Neurology, University of California Los Angeles School of Medicine, Los Angeles, California
Neuro-Ophthalmology; Smell and Taste; Hearing and Equilibrium

Jonathan Barasch, MD, PhD
Professor of Medicine and Pathology and Cell Biology, Department of Medicine, Division of Nephrology, Columbia University College of Physicians & Surgeons, New York, New York
Structure and Function of the Kidneys

Richard L. Barbano, MD, PhD
Professor of Neurology, University of Rochester, Rochester, New York
Mechanical and Other Lesions of the Spine, Nerve Roots, and Spinal Cord

Elizabeth Barrett-Connor, MD
Professor of Community and Family Medicine, University of California San Diego, San Diego, California
Menopause

John R. Bartholomew, MD
Section Head, Vascular Medicine, Cardiovascular Medicine, Cleveland Clinic, Professor of Medicine, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, Ohio
Other Peripheral Arterial Diseases

Mary Barton, MD, MPP
Vice President, Performance Measurement, National Committee for Quality Assurance, Washington, D.C.
The Periodic Health Examination

Robert C. Basner, MD
Professor of Medicine, Columbia University Medical Center; Director, Columbia University Cardiopulmonary Sleep and Ventilatory Disorders Center, Columbia University College of Physicians and Surgeons, New York, New York
Obstructive Sleep Apnea

Stephen G. Baum, MD
Chairman of Medicine, Mount Sinai Beth Israel Hospital; Professor of Medicine and of Microbiology and Immunology, Albert Einstein College of Medicine, New York, New York
Mycoplasma Infections

Daniel G. Bausch, MD, MPH&TM
Associate Professor, Department of Tropical Medicine, Tulane University Health Sciences Center, New Orleans, Louisiana
Viral Hemorrhagic Fevers

Arnold S. Bayer, MD
Professor of Medicine, David Geffen School of Medicine at University of California Los Angeles; LA Biomedical Research Institute; Vice Chair for Academic Affairs, Department of Medicine, Harbor-UCLA Medical Center, Los Angeles, California
Infective Endocarditis

Hasan Bazari, MD
Associate Professor of Medicine, Harvard Medical School, Department of Medicine, Clinical Director, Nephrology, Program Director, Internal Medicine Residency Program, Massachusetts General Hospital, Boston, Massachusetts
Approach to the Patient with Renal Disease

John H. Beigel, MD
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland
Antiviral Therapy (Non-HIV)

George A. Beller, MD
Professor of Medicine, University of Virginia Health System, Charlottesville, Virginia
Noninvasive Cardiac Imaging

Robert M. Bennett, MD
Professor of Medicine, Oregon Health and Science University, Portland, Oregon
Fibromyalgia, Chronic Fatigue Syndrome, and Myofascial Pain

Joseph R. Berger, MD
Professor of Neurology, Chief of the Multiple Sclerosis Division, Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania
Cytomegalovirus, Epstein-Barr Virus, and Slow Virus Infections of the Central Nervous System; Neurologic Complications of Human Immunodeficiency Virus Infection; Brain Abscess and Parameningeal Infections

Paul D. Berk, MD
Professor of Medicine, Department of Medicine, Columbia University College of Physicians and Surgeons, New York, New York
Approach to the Patient with Jaundice or Abnormal Liver Tests

Nancy Berliner, MD
Professor of Medicine, Harvard Medical School; Chief, Division of Hematology, Brigham and Women's Hospital, Boston, Massachusetts
Leukocytosis and Leukopenia

James L. Bernat, MD
Louis and Ruth Frank Professor of Neuroscience, Professor of Neurology and Medicine, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire; Department of Neurology, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire
Coma, Vegetative State, and Brain Death

Philip J. Bierman, MD
Professor, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, Nebraska
Approach to the Patient with Lymphadenopathy and Splenomegaly; Non-Hodgkin Lymphomas

Michael R. Bishop, MD
Professor of Medicine, Director, Hematopoietic Cellular Therapy Program, Section of Hematology and Oncology, Department of Medicine, University of Chicago, Chicago, Illinois
Hematopoietic Stem Cell Transplantation

Bruce R. Bistrian, MD, PhD, MPH
Professor of Medicine, Beth Israel Deaconess Medical Center; Professor of Medicine, Harvard Medical School, Boston, Massachusetts
Nutritional Assessment

Joseph J. Biundo, MD
Clinical Professor of Medicine, Tulane Medical Center, New Orleans, Louisiana
Bursitis, Tendinitis, and Other Periarticular Disorders and Sports Medicine

Adrian R. Black, PhD
Assistant Professor, Director of Tissue Sciences for the Eppley Institute, The Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, Nebraska
Cancer Biology and Genetics

Charles D. Blanke, MD
Professor of Medicine, Oregon Health and Science University, Portland, Oregon
Neoplasms of the Small and Large Intestine

Joel N. Blankson, MD, PhD
Associate Professor, Johns Hopkins University School of Medicine, Baltimore, Maryland
Immunopathogenesis of Human Immunodeficiency Virus Infection

Martin J. Blaser, MD
Muriel and George Singer Professor of Medicine, Professor of Microbiology, Director, Human Microbiome Program, New York University Langone Medical Center, New York, New York
Acid Peptic Disease; Human Microbiome

William A. Blattner, MD
Professor and Associate Director, Institute of Human Virology, School of Medicine, University of Maryland; Professor of Medicine, School of Medicine, University of Maryland; Professor and Head, Division of Cancer Epidemiology, Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, Maryland
Retroviruses Other Than Human Immunodeficiency Virus

Thomas P. Bleck, MD
Professor of Neurological Sciences, Neurosurgery, Internal Medicine, and Anesthesiology, Associate Chief Medical Officer (Critical Care), Rush Medical College, Chicago, Illinois
Arboviruses Affecting the Central Nervous System

Joel A. Block, MD
The Willard L. Wood MD Professor and Director, Division of Rheumatology, Rush University Medical Center, Chicago, Illinois
Osteoarthritis

Henk Blom, MD
Laboratory of Clinical Biochemistry and Metabolism, Department of General Pediatrics, Adolescent Medicine and Neonatology, University Medical Centre Freiburg, Head of Laboratory/Clinical Biochemical Geneticist, Freiburg, Germany
Homocystinuria and Hyperhomocysteinemia

Olaf A. Bodamer, MD
Medical Genetics, University of Miami Hospital, Miami, Florida
Approach to Inborn Errors of Metabolism

William E. Boden, MD
Professor of Medicine, Albany Medical College; Chief of Medicine, Albany Stratton VA Medical Center; Vice-Chairman, Department of Medicine, Albany Medical Center, Albany, New York
Angina Pectoris and Stable Ischemic Heart Disease

Jean Bolognia, MD
Professor of Dermatology, Yale Medical School; Attending Physician, Yale-New Haven Hospital, New Haven, Connecticut
Infections, Hyperpigmentation and Hypopigmentation, Regional Dermatology, and Distinctive Lesions in Black Skin

Robert A. Bonomo, MD
Chief, Medical Service, Louis Stokes Cleveland VA Medical Center; Professor of Medicine, Pharmacology, Biochemistry, Molecular Biology, and Microbiology, Case Western Reserve University School of Medicine, Cleveland, Ohio
Diseases Caused by Acinetobacter and Stenotrophomonas Species

Larry Borish, MD
Professor of Medicine, Allergy, and Clinical Immunology, University of Virginia Health System, Charlottesville, Virgina
Allergic Rhinitis and Chronic Sinusitis

Patrick J. Bosque, MD
Associate Professor of Neurology, University of Colorado Denver School of Medicine; Neurologist, Denver Health Medical Center, Denver, Colorado
Prion Diseases

David J. Brenner, PhD, DSc
Higgins Professor of Radiation Biophysics, Center for Radiological Research, Columbia University Medical Center, New York, New York
Radiation Injury

Itzhak Brook, MD, MSc
Professor of Pediatrics and Medicine, Georgetown University, Georgetown University Medical Center, Washington, D.C.
Diseases Caused by Non–Spore-Forming Anaerobic Bacteria; Actinomycosis

Enrico Brunetti, MD
Assistant Professor of Infectious Diseases, University of Pavia; Attending Physician, Division of Infectious and Tropical Diseases, IRCCS San Matteo Hospital Foundation; Co-Director, WHO Collaborating Centre for Clinical Management of Cystic Echinococcosis, Pavia, Italy
Cestodes

David M. Buchner, MD, MPH
Shahid and Ann Carlson Khan Professor in Applied Health Sciences, Department of Kinesiology and Community Health, University of Illinois at Urbana-Champaign, Champaign, Illinois
Physical Activity

Pierre A. Buffet, MD, PhD
Research Unit Head, Erythrocyte Parasite Pathogenesis Research Team INSERM–University Paris 6, CIMI–Paris Research Center, University Pierre and Marie Curie; Associate Professor of Parasitology, Faculty of Medicine, University Pierre and Marie Curie, Pitié-Salpêtrière Hospital, Paris, France
Leishmaniasis

H. Franklin Bunn, MD
Professor of Medicine, Harvard Medical School; Physician, Brigham and Women's Hospital, Boston, Massachusetts
Approach to the Anemias

David A. Bushinsky, MD
John J. Kuiper Distinguished Professor of Medicine, Chief, Nephrology Division, University of Rochester School of Medicine; Associate Chair for Academic Affairs in Medicine, University of Rochester Medical Center, Rochester, New York
Nephrolithiasis

Vivian P. Bykerk, MD
Associate Professor of Medicine, Weill Cornell Medical College; Associate Attending Physician, Hospital for Special Surgery, New York, New York
Approach to the Patient with Rheumatic Disease

Peter A. Calabresi, MD
Professor of Neurology and Director of the Richard T. Johnson Division of Neuroimmunology and Neuroinfectious Diseases, Johns Hopkins University; Director of the Multiple Sclerosis Center, Johns Hopkins Hospital, Baltimore, Maryland
Multiple Sclerosis and Demyelinating Conditions of the Central Nervous System

David P. Calfee, MD, MS
Associate Professor of Medicine and Healthcare Policy and Research, Weill Cornell Medical College; Chief Hospital Epidemiologist, New York-Presbyterian Hospital/Weill Cornell Medical Center, New York, New York
Prevention and Control of Health Care–Associated Infections

Douglas Cameron, MD, MBA
Professor of Ophthalmology and Visual Neurosciences, University of Minnesota, Minneapolis, Minnesota
Diseases of the Visual System

Michael Camilleri, MD
Atherton and Winifred W. Bean Professor, Professor of Medicine, Pharmacology, and Physiology, College of Medicine, Mayo Clinic, Consultant, Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota
Disorders of Gastrointestinal Motility

Grant W. Cannon, MD
Thomas E. and Rebecca D. Jeremy Presidential Endowed Chair for Arthritis Research, Associate Chief of Staff for Academic Affiliations, George E. Wahlen VA Medical Center, Salt Lake City, Utah
Immunosuppressing Drugs Including Corticosteroids

Maria Domenica Cappellini, MD
Professor of Internal Medicine, University of Milan, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, University of Milan, Milan, Italy
The Thalassemias

Blase A. Carabello, MD
Professor of Medicine, Chairman, Department of Cardiology, Mount Sinai Beth Israel Heart Institute, New York, New York
Valvular Heart Disease

Edgar M. Carvalho, MD
Professor of Medicine and Clinical Immunology, Faculdade de Medicina da Bahia, Universidade Federal da Bahia and Escola Bahiana de Medicina e Saúde Pública, Salvador, Bahia, Brazil
Schistosomiasis (Bilharziasis)

William H. Catherino, MD, PhD
Professor and Research Head, Department of Obstetrics and Gynecology, Uniformed Services University of the Health Sciences Division of Reproductive Endocrinology and Infertility; Program in Reproductive and Adult Endocrinology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland
Ovaries and Development; Reproductive Endocrinology and Infertility

Jane A. Cauley, DrPH
Professor of Epidemiology, University of Pittsburgh Graduate School of Public Health, Vice Chair of the Department of Epidemiology, Pittsburgh, Pennsylvania
Epidemiology of Aging: Implications of the Aging of Society

Naga P. Chalasani, MD
David W. Crabb Professor and Director, Division of Gastroenterology and Hepatology, Indiana University School of Medicine, Indianapolis, Indiana
Alcoholic and Nonalcoholic Steatohepatitis

Henry F. Chambers, MD
Professor of Medicine, University of California San Francisco School of Medicine; Director, Clinical Research Services, Clinical and Translational Sciences Institute, San Francisco, California
Staphylococcal Infections

William P. Cheshire, Jr., MD
Professor of Neurology, Mayo Clinic, Jacksonville, Florida
Autonomic Disorders and Their Management

Ilseung Cho, MD, MS
Assistant Professor of Medicine, Division of Gastroenterology, Department of Medicine, New York University, New York, New York
Human Microbiome

Arun Chockalingam, PhD
Professor of Epidemiology and Global Health, Director, Office of Global Health Education and Training; Dalla Lana Faculty of Public Health, University of Toronto, Toronto, Ontario, Canada
Global Health

David C. Christiani, MD
Professor of Medicine, Harvard Medical School; Physician, Pulmonary and Critical Care, Massachusetts General Hospital; Elkan Blout Professor of Environmental Genetics, Environmental Health, Harvard School of Public Health, Boston, Massachusetts
Physical and Chemical Injuries of the Lung

David H. Chu, MD, PhD
Director, Contact Dermatitis, Division of Dermatology and Cutaneous Surgery, Scripps Clinic Medical Group, La Jolla, California
Structure and Function of the Skin

Theodore J. Cieslak, MD
Pediatric Infectious Diseases, Clinical Professor of Pediatrics, University of Texas Health Science Center at San Antonio; Department of Pediatrics, Fort Sam Houston, Texas
Bioterrorism

Carolyn Clancy, MD
Interim Under Secretary for Health, Veterans Administration, Washington, D.C.
Measuring Health and Health Care

David R. Clemmons, MD
Kenan Professor of Medicine, University of North Carolina School of Medicine; Attending Physician, Medicine, UNC Hospitals, Chapel Hill, North Carolina
Approach to the Patient with Endocrine Disease

David Cohen, MD
Professor of Medicine, Division of Nephrology; Medical Director, Kidney and Pancreas Transplantation, Columbia University Medical Center, New York, New York
Treatment of Irreversible Renal Failure

Jeffrey Cohen, MD
Chief, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland
Varicella-Zoster Virus (Chickenpox, Shingles)

Myron S. Cohen, MD
Associate Vice Chancellor for Global Health, Director, UNC Institute for Global Health and Infectious Diseases, Chief, Division of Infectious Diseases, Yeargan-Bate Eminent Professor of Medicine, Microbiology, and Immunology and Epidemiology, Chapel Hill, North Carolina
Approach to the Patient with a Sexually Transmitted Infection; Prevention of Human Immunodeficiency Virus Infection

Steven P. Cohen, MD
Professor of Anesthesiology and Critical Care Medicine and Physical Medicine and Rehabilitation, Johns Hopkins School of Medicine, Baltimore, Maryland, and Uniformed Services University of the Health Sciences, Bethesda, Maryland; Director, Pain Research, Walter Reed National Military Medical Center, Bethesda, Maryland
Pain

Steven L. Cohn, MD
Professor of Clinical Medicine, University of Miami Miller School of Medicine; Medical Director, UHealth Preoperative Assessment Center; Director, Medical Consultation Service, University of Miami Hospital, Miami, Florida
Preoperative Evaluation

Robert Colebunders, MD
Emeritus Professor, Institute of Tropical Medicine, Antwerp, Belgium
Immune Reconstitution Inflammatory Syndrome in HIV/AIDS

Joseph M. Connors, MD
Clinical Professor, University of British Columbia; Clinical Director, BC Cancer Agency Centre for Lymphoid Cancer, Vancouver, British Columbia, Canada
Hodgkin Lymphoma

Deborah J. Cook, MD, MSc
Professor of Medicine, Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada
Approach to the Patient in a Critical Care Setting

Kenneth H. Cowan, MD, PhD
Director, Fred & Pamela Buffett Cancer Center; Director, The Eppley Institute for Research in Cancer and Allied Diseases; Professor of Medicine, University of Nebraska Medical Center, Omaha, Nebraska
Cancer Biology and Genetics

Joseph Craft, MD
Paul B. Beeson Professor of Medicine and Immunobiology, Section Chief, Rheumatology, Program Director, Investigative Medicine, Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut
The Adaptive Immune Systems

Jill Patricia Crandall, MD
Professor of Clinical Medicine, Division of Endocrinology and Diabetes Research Center, Albert Einstein College of Medicine, Bronx, New York
Diabetes Mellitus

Simon L. Croft, BSc, PhD
Professor of Parasitology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom
Leishmaniasis

Kristina Crothers, MD
Associate Professor, Department of Medicine, Division of Pulmonary and Critical Care, University of Washington School of Medicine, Seattle, Washington
Pulmonary Manifestations of Human Immunodeficiency Virus and Acquired Immunodeficiency Syndrome

Mary K. Crow, MD
Joseph P. Routh Professor of Rheumatic Diseases in Medicine, Weill Cornell Medical College; Physician in Chief and Benjamin M. Rosen Chair in Immunology and Inflammation Research, Hospital for Special Surgery, New York, New York
The Innate Immune Systems; Approach to the Patient with Rheumatic Disease; Systemic Lupus Erythematosus

John A. Crump, MB ChB, MD, DTM&H
McKinlay Professor of Global Health, Centre for International Health, University of Otago, Dunedin, New Zealand
Salmonella Infections (Including Enteric Fever)

Mark R. Cullen, MD
Professor of Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, California
Principles of Occupational and Environmental Medicine

Charlotte Cunningham-Rundles, MD, PhD
Professor of Medicine and Pediatrics, Icahn School of Medicine at Mount Sinai, New York, New York
Primary Immunodeficiency Diseases

Inger K. Damon, MD, PhD
Director, Division of High Consequence Pathogens and Pathology, Centers for Disease Control and Prevention, Atlanta, Georgia
Smallpox, Monkeypox, and Other Poxvirus Infections

Troy E. Daniels, DDS, MS
Professor Emeritus of Oral Pathology and Pathology, University of California San Francisco, San Francisco, California
Diseases of the Mouth and Salivary Glands

Nancy E. Davidson, MD
Hillman Professor of Oncology, University of Pittsburgh; Director, University of Pittsburgh Cancer Institute and UPMC CancerCenter, Pittsburgh, Pennsylvania
Breast Cancer and Benign Breast Disorders

Lisa M. DeAngelis, MD
Chair, Department of Neurology, Memorial Sloan-Kettering Cancer Center; Professor of Neurology, Weill Cornell Medical College, New York, New York
Tumors of the Central Nervous System

Malcolm M. DeCamp, MD
Fowler McCormick Professor of Surgery, Feinberg School of Medicine, Northwestern University; Chief, Division of Thoracic Surgery, Northwestern Memorial Hospital, Chicago, Illinois
Interventional and Surgical Approaches to Lung Disease

Carlos del Rio, MD
Hubert Professor and Chair and Professor of Medicine, Hubert Department of Global Health, Rollins School of Public Health and Department of Medicine, Emory University School of Medicine, Atlanta, Georgia
Prevention of Human Immunodeficiency Virus Infection

Patricia A. Deuster, PhD, MPH
Professor and Director, Consortium for Health and Military Performance, Department of Military and Emergency Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland
Rhabdomyolysis

Robert B. Diasio, MD
William J. and Charles H. Mayo Professor, Molecular Pharmacology and Experimental Therapeutics and Oncology, Mayo Clinic, Rochester, Minnesota
Principles of Drug Therapy

David J. Diemert, MD
Associate Professor, Department of Microbiology, Immunology, and Tropical Medicine, School of Medicine and Health Sciences, The George Washington University, Washington, D.C.
Intestinal Nematode Infections; Tissue Nematode Infections

Kathleen B. Digre, MD
Professor of Neurology, Ophthalmology, Director, Division of Headache and Neuro-Ophthalmology, University of Utah, Salt Lake City, Utah
Headaches and Other Head Pain

James H. Doroshow, MD
Bethesda, Maryland
Approach to the Patient with Cancer; Malignant Tumors of Bone, Sarcomas, and Other Soft Tissue Neoplasms

John M. Douglas, Jr., MD
Executive Director, Tri-County Health Department, Greenwood Village, Colorado
Papillomavirus

Jeffrey M. Drazen, MD
Distinguished Parker B. Francis Professor of Medicine, Harvard Medical School; Senior Physician, Brigham and Women's Hospital, Boston, Massachusetts
Asthma

Stephen C. Dreskin, MD, PhD
Professor of Medicine and Immunology, Division of Allergy and Clinical Immunology, Department of Medicine, University of Colorado Denver, School of Medicine, Aurora, Colorado
Urticaria and Angioedema

W. Lawrence Drew, MD, PhD
Professor Emeritus, Laboratory Medicine and Medicine, University of California San Francisco, San Francisco, California
Cytomegalovirus

George L. Drusano, MD
Professor and Director, Institute for Therapeutic Innovation, College of Medicine, University of Florida, Lake Nona, Florida
Antibacterial Chemotherapy

Thomas D. DuBose, Jr., MD
Emeritus Professor of Internal Medicine and Nephrology, Wake Forest School of Medicine, Winston-Salem, North Carolina
Vascular Disorders of the Kidney

F. Daniel Duffy, MD
Professor of Internal Medicine and Steve Landgarten Chair in Medical Leadership, School of Community Medicine, University of Oklahoma College of Medicine, Tulsa, Oklahoma
Counseling for Behavior Change

Herbert L. DuPont, MD, MACP
Mary W. Kelsey Chair and Director, Center for Infectious Diseases, University of Texas School of Public Health; H. Irving Schweppe Chair of Internal Medicine and Vice Chairman, Department of Medicine, Baylor College of Medicine; Chief of Internal Medicine, St. Luke's Hospital System, Houston, Texas
Approach to the Patient with Suspected Enteric Infection

Madeleine Duvic, MD
Professor and Deputy Chairman, Department of Dermatology, The University of Texas MD Anderson Cancer Center, Houston, Texas
Urticaria, Drug Hypersensitivity Rashes, Nodules and Tumors, and Atrophic Diseases

Kathryn M. Edwards, MD
Sarah H. Sell and Cornelius Vanderbilt Chair in Pediatrics, Vanderbilt University School of Medicine; Director, Vanderbilt Vaccine Research Program, Monroe Carrell Jr. Children's Hospital at Vanderbilt, Nashville, Tennessee
Parainfluenza Viral Disease

N. Lawrence Edwards, MD
Professor of Medicine, Vice Chairman, Department of Medicine, University of Florida; Chief, Section of Rheumatology, Medical Service, Malcom Randall Veterans Affairs Medical Center, Gainesville, Florida
Crystal Deposition Diseases

Lawrence H. Einhorn, MD
Distinguished Professor, Department of Medicine, Division of Hematology/Oncology, Livestrong Foundation Professor of Oncology, Indiana University School of Medicine, Indianapolis, Indiana
Testicular Cancer

Ronald J. Elin, MD, PhD
A.J. Miller Professor and Chairman, Department of Pathology and Laboratory Medicine, University of Louisville School of Medicine, Louisville, Kentucky
Reference Intervals and Laboratory Values

George M. Eliopoulos, MD
Professor of Medicine, Harvard Medical School; Physician, Division of Infectious Diseases, Beth Israel Deaconess Medical Center, Boston, Massachusetts
Principles of Anti-Infective Therapy

Perry Elliott, MD
Professor in Inherited Cardiovascular Disease, Institute of Cardiovascular Science, University College London, London, United Kingdom
Diseases of the Myocardium and Endocardium

Jerrold J. Ellner, MD
Professor of Medicine, Boston University School of Medicine; Chief, Section of Infectious Diseases, Boston Medical Center, Boston, Massachusetts
Tuberculosis

Dirk M. Elston, MD
Director, Ackerman Academy of Dermatopathology, New York, New York
Arthropods and Leeches

Ezekiel J. Emanuel, MD, PhD
Vice Provost for Global Initiatives, Diane V.S. Levy and Robert M. Levy University Professor, Chair, Department of Medical Ethics and Health Policy, University of Pennsylvania, Philadelphia, Pennsylvania
Bioethics in the Practice of Medicine

Joel D. Ernst, MD
Director, Division of Infectious Diseases and Immunology, Jeffrey Bergstein Professor of Medicine, Professor of Medicine, Pathology, and Microbiology, New York University School of Medicine; Attending Physician, New York University Langone Medical Center, New York, New York
Leprosy (Hansen Disease)

Gregory T. Everson, MD
Professor of Medicine, Director of Hepatology, University of Colorado School of Medicine, Aurora, Colorado
Hepatic Failure and Liver Transplantation

Amelia Evoli, MD
Associate Professor of Neurology, Catholic University, Agostino Gemelli University Hospital, Rome, Italy
Disorders of Neuromuscular Transmission

Douglas O. Faigel, MD
Professor of Medicine, Mayo Clinic, Chair, Division of Gastroenterology and Hepatology, Scottsdale, Arizona
Neoplasms of the Small and Large Intestine

Matthew E. Falagas, MD, MSc, DSc
Director, Alfa Institute of Biomedical Sciences, Athens, Greece; Adjunct Associate Professor of Medicine, Tufts University School of Medicine, Boston, Massachusetts; Chief, Department of Medicine and Infectious Diseases, Iaso General Hospital, Iaso Group, Athens, Greece
Pseudomonas and Related Gram-Negative Bacillary Infections

Gary W. Falk, MD, MS
Professor of Medicine, Division of Gastroenterology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania
Diseases of the Esophagus

Gene Feder, MBBS, MD
Professor, Centre for Academic Primary Care, School of Social and Community Medicine, University of Bristol; General Practitioner, Helios Medical Centre, Bristol, United Kingdom
Intimate Partner Violence

David J. Feller-Kopman, MD
Director, Bronchoscopy and Interventional Pulmonology, Associate Professor of Medicine, The Johns Hopkins University, Baltimore, Maryland
Interventional and Surgical Approaches to Lung Disease

Gary S. Firestein, MD
Dean and Associate Vice Chancellor of Translational Medicine, University of California San Diego School of Medicine, La Jolla, California
Mechanisms of Inflammation and Tissue Repair

Glenn I. Fishman, MD
Director, Leon H. Charney Division of Cardiology, Vice-Chair for Research, Department of Medicine, William Goldring Professor of Medicine, New York University School of Medicine, New York, New York
Principles of Electrophysiology

Lee A. Fleisher, MD
Robert D. Dripps Professor and Chair, Anesthesiology and Critical Care, Professor of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania
Overview of Anesthesia

Paul W. Flint, MD
Professor and Chair, Otolaryngology, Head and Neck Surgery, Oregon Health and Science University, Portland, Oregon
Throat Disorders

Evan L. Fogel, MD, MSc
Professor of Clinical Medicine, Indiana University School of Medicine, Indianapolis, Indiana
Diseases of the Gallbladder and Bile Ducts

Marsha D. Ford, MD
Adjunct Professor of Emergency Medicine, School of Medicine, University of North Carolina-Chapel Hill; Director, Carolinas Poison Center, Carolinas HealthCare System, Charlotte, North Carolina
Acute Poisoning

Chris E. Forsmark, MD
Professor of Medicine, Chief, Division of Gastroenterology, Hepatology, and Nutrition, University of Florida, Gainesville, Florida
Pancreatitis

Vance G. Fowler, Jr., MD, MHS
Professor of Medicine, Duke University Medical Center, Durham, North Carolina
Infective Endocarditis

Manuel A. Franco, MD, PhD
Director of Postgraduate Programs, School of Sciences, Pontificia Universidad Javeriana, Bogota, Colombia
Rotaviruses, Noroviruses, and Other Gastrointestinal Viruses

David O. Freedman, MD
Professor of Medicine and Microbiology, University of Alabama at Birmingham; Director, Gorgas Center for Geographic Medicine, Birmingham, Alabama
Approach to the Patient before and after Travel

Martyn A. French, MD
Professor in Clinical Immunology, School of Pathology and Laboratory Medicine, University of Western Australia, Perth, Australia
Immune Reconstitution Inflammatory Syndrome in HIV/AIDS

Karen Freund, MD, MPH
Professor of Medicine, Associate Director, Tufts Clinical and Translational Science Institute, Tufts University School of Medicine, Tufts Medical Center, Boston, Massachusetts
Approach to Women's Health

Cem Gabay, MD
Professor of Medicine, Head, Division of Rheumatology, University Hospitals of Geneva, Geneva, Switzerland
Biologic Agents

Kenneth L. Gage, PhD
Chief, Entomology and Ecology Activity, Centers for Disease Control and Prevention, Division of Vector-Borne Diseases, Bacterial Diseases Branch, Fort Collins, Colorado
Plague and Other Yersinia Infections

John N. Galgiani, MD
Professor of Medicine, Valley Fever Center for Excellence, University of Arizona, Tucson, Arizona
Coccidioidomycosis

Patrick G. Gallagher, MD
Professor of Pediatrics, Pathology, and Genetics, Yale University School of Medicine; Attending Physician, Yale–New Haven Hospital, New Haven, Connecticut
Hemolytic Anemias: Red Blood Cell Membrane and Metabolic Defects

Leonard Ganz, MD
Director of Cardiac Electrophysiology, Heritage Valley Health System, Beaver, Pennsylvania
Electrocardiography

Hasan Garan, MD
Director, Cardiac Electrophysiology, Dickinson W. Richards, Jr. Professor of Medicine, Columbia University Medical Center, New York, New York
Ventricular Arrhythmias

Guadalupe Garcia-Tsao, MD
Professor of Medicine, Yale University School of Medicine; Chief, Digestive Diseases, VA Connecticut Healthcare System, West Haven, Connecticut
Cirrhosis and Its Sequelae

William M. Geisler, MD, MPH
Professor of Medicine, University of Alabama at Birmingham, Birmingham, Alabama
Diseases Caused by Chlamydiae

Tony P. George, MD
Division of Brain and Therapeutics, Department of Psychiatry, University of Toronto; Schizophrenia Division, The Centre for Addiction and Mental Health, Toronto, Ontario, Canada
Nicotine and Tobacco

Lior Gepstein, MD, PhD
Edna and Jonathan Sohnis Professor in Medicine and Physiology, Rappaport Faculty of Medicine and Research Institute, Technion–Israel Institute of Technology, Rambam Health Care Campus, Haifa, Israel
Gene and Cell Therapy

Susan I. Gerber, MD
Team Lead, Respiratory Viruses/Picornaviruses, Division of Viral Diseases/Epidemiology Branch, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia
Coronaviruses

Dale N. Gerding, MD
Professor of Medicine, Loyola University Chicago Stritch School of Medicine, Research Physician, Edward Hines, Jr. VA Hospital, Hines, Illinois
Clostridial Infections

Morie A. Gertz, MD
Consultant, Division of Hematology, Mayo Clinic, Rochester, Minnesota; Roland Seidler, Jr. Professor of the Art of Medicine in Honor of Michael D. Brennan, MD, Professor of Medicine, Mayo Clinic, College of Medicine, Rochester, Minnesota
Amyloidosis

Gordon D. Ginder, MD
Professor, Internal Medicine, Director, Massey Cancer Center, Virginia Commonwealth University, Richmond, Virginia
Microcytic and Hypochromic Anemias

Jeffrey S. Ginsberg, MD
Professor of Medicine, McMaster University, Member of Thrombosis and Atherosclerosis Research Institute, St. Joseph's Healthcare Hamilton, Hamilton, Ontario, Canada
Peripheral Venous Disease

Geoffrey S. Ginsburg, MD, PhD
Director, Duke Center for Applied Genomics and Precision Medicine; Professor of Medicine, Pathology and Biomedical Engineering, Duke University, Durham, North Carolina
Applications of Molecular Technologies to Clinical Medicine

Michael Glogauer, DDS, PhD
Professor, Faculty of Dentistry, University of Toronto, Toronto, Ontario, Canada
Disorders of Phagocyte Function

John W. Gnann, Jr., MD
Professor of Medicine, Department of Medicine, Division of Infectious Diseases, Medical University of South Carolina, Charleston, South Carolina
Mumps

Matthew R. Golden, MD, MPH
Professor of Medicine, University of Washington, Director, HIV/STD Program, Public Health–Seattle & King County, Seattle, Washington
Neisseria Gonorrhoeae *Infections*

Lee Goldman, MD
Harold and Margaret Hatch Professor, Executive Vice President and Dean of the Faculties of Health Sciences and Medicine, Chief Executive, Columbia University Medical Center, Columbia University, New York, New York
Approach to Medicine, the Patient, and the Medical Profession: Medicine as a Learned and Humane Profession; Approach to the Patient with Possible Cardiovascular Disease

Ellie J.C. Goldstein, MD
Clinical Professor of Medicine, David Geffen School of Medicine at University of California Los Angeles, Los Angeles, California; Director, R.M. Alden Research Laboratory, Santa Monica, California
Diseases Caused by Non–Spore-Forming Anaerobic Bacteria

Larry B. Goldstein, MD
Professor of Neurology, Director, Duke Stroke Center, Neurology, Duke University; Staff Neurologist, Durham VA Medical Center, Durham, North Carolina
Approach to Cerebrovascular Diseases; Ischemic Cerebrovascular Disease

Lawrence T. Goodnough, MD
Professor of Pathology and Medicine, Stanford University; Director, Transfusion Service, Stanford University Medical Center, Stanford, California
Transfusion Medicine

Eduardo H. Gotuzzo, MD
Professor of Medicine, Director, Alexander von Humboldt Tropical Medicine Institute, Universidad Peruana Cayetano Heredia; Chief Physician, Department of Infectious, Tropical, and Dermatologic Diseases, National Hospital Cayetano Heredia, Lima, Peru
Cholera and Other Vibrio Infections; Liver, Intestinal, and Lung Fluke Infections

Deborah Grady, MD, MPH
Professor of Medicine, University of California San Francisco, San Francisco, California
Menopause

Leslie C. Grammer, MD
Professor of Medicine, Northwestern University Feinberg School of Medicine; Attending Physician, Northwestern Memorial Hospital, Chicago, Illinois
Drug Allergy

F. Anthony Greco, MD
Medical Director, Sarah Cannon Cancer Center, Nashville, Tennessee
Cancer of Unknown Primary Origin

Harry B. Greenberg, MD
Professor, Departments of Medicine and Microbiology and Immunology, Stanford University School of Medicine, Stanford, California
Rotaviruses, Noroviruses, and Other Gastrointestinal Viruses

Steven A. Greenberg, MD
Associate Professor of Neurology, Harvard Medical School; Associate Neurologist, Brigham and Women's Hospital, Boston, Massachusetts
Inflammatory Myopathies

Robert C. Griggs, MD
Professor of Neurology, Medicine, Pediatrics, and Pathology and Laboratory Medicine, University of Rochester School of Medicine and Dentistry, Rochester, New York
Approach to the Patient with Neurologic Disease

Lev M. Grinberg, MD, PhD
Professor, Chief, Department of Pathology, Ural Medical University; Chief Researcher of the Ural Scientific Research Institute of Phthisiopulmonology, Chief Pathologist of Ekaterinburg, Ekaterinburg, Russia
Anthrax

Daniel Grossman, MD
Vice President for Research, Ibis Reproductive Health, Oakland, California; Assistant Clinical Professor, Bixby Center for Global Reproductive Health, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco, San Francisco, California
Contraception

Lisa M. Guay-Woodford, MD
Hudson Professor of Pediatrics, The George Washington University; Director, Center for Translational Science, Director, Clinical and Translational Institute at Children's National, Children's National Health System, Washington, D.C.
Hereditary Nephropathies and Developmental Abnormalities of the Urinary Tract

Richard L. Guerrant, MD
Thomas H. Hunter Professor of International Medicine, Founding Director, Center for Global Health, Division of Infectious Diseases and International Health, University of Virginia School of Medicine, University of Virginia Health Sciences Center, Charlottesville, Virginia
Cryptosporidiosis

Roy M. Gulick, MD, MPH
Gladys and Roland Harrison Professor of Medicine, Medicine/Infectious Diseases, Weill Cornell Medical College; Attending Physician, New York–Presbyterian Hospital, New York, New York
Antiretrovial Therapy of HIV/AIDS

Klaus D. Hagspiel, MD
Professor of Radiology, Medicine, and Pediatrics, Chief, Noninvasive Cardiovascular Imaging, University of Virginia Health System, Charlottesville, Virginia
Noninvasive Cardiac Imaging

John D. Hainsworth, MD
Chief Scientific Officer, Sarah Cannon Research Institute, Nashville, Tennessee
Cancer of Unknown Primary Origin

Anders Hamsten, MD, PhD
Professor of Cardiovascular Diseases, Center for Molecular Medicine and Department of Cardiology, Karolinska University Hospital, Department of Medicine, Karolinska Institute, Stockholm, Sweden
Atherosclerosis, Thrombosis, and Vascular Biology

Kenneth R. Hande, MD
Professor of Medicine and Pharmacology, Vanderbilt/Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee
Neuroendocrine Tumors and the Carcinoid Syndrome

H. Hunter Handsfield, MD
Professor Emeritus of Medicine, University of Washington Center for AIDS and STD, Seattle, Washington
Neisseria Gonorrhoeae *Infections*

Göran K. Hansson, MD, PhD
Professor of Cardiovascular Research, Center for Molecular Medicine at Karolinska University Hospital, Department of Medicine, Karolinska Institute, Stockholm, Sweden
Atherosclerosis, Thrombosis, and Vascular Biology

Raymond C. Harris, MD
Professor of Medicine, Ann and Roscoe R. Robinson Chair in Nephrology, Chief, Division of Nephrology, Vanderbilt University School of Medicine, Nashville, Tennessee
Diabetes and the Kidney

Stephen Crane Hauser, MD
Associate Professor of Medicine, Internal Medicine, Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, Minnesota
Vascular Diseases of the Gastrointestinal Tract

Frederick G. Hayden, MD
Stuart S. Richardson Professor of Clinical Virology and Professor of Medicine, University of Virginia School of Medicine; Staff Physician, University of Virginia Health System, Charlottesville, Virginia
Influenza

Douglas C. Heimburger, MD, MS
Professor of Medicine, Associate Director for Education and Training, Vanderbilt University School of Medicine, Vanderbilt Institute for Global Health, Nashville, Tennessee
Nutrition's Interface with Health and Disease

Erik L. Hewlett, MD
Professor of Medicine and of Microbiology, Immunology, and Cancer Biology, University of Virginia School of Medicine, University of Virginia Health System, Charlottesville, Virginia
Whooping Cough and Other Bordetella Infections

Richard J. Hift, PhD, MMed
School of Clinical Medicine, University of KwaZulu-Natal, Durban, South Africa
The Porphyrias

David R. Hill, MD, DTM&H
Professor of Medical Sciences, Director of Global Public Health, Frank H. Netter MD School of Medicine at Quinnipiac University, Hamden, Connecticut
Giardiasis

Nicholas S. Hill, MD
Professor of Medicine, Tufts University School of Medicine; Chief, Division of Pulmonary, Critical Care, and Sleep Medicine, Tufts Medical Center, Boston, Massachusetts
Respiratory Monitoring in Critical Care

L. David Hillis, MD
Professor and Chair, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas
Acute Coronary Syndrome: Unstable Angina and Non-ST Elevation Myocardial Infarction

Jack Hirsh, CM, MD, DSc
Professor Emeritus, McMaster University, Hamilton, Ontario, Canada
Antithrombotic Therapy

Steven M. Holland, MD
Chief, Laboratory of Clinical Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland
The Nontuberculous Mycobacteria

Steven M. Hollenberg, MD
Professor of Medicine, Cooper Medical School of Rowan University; Director, Coronary Care Unit, Cooper University Hospital, Camden, New Jersey
Cardiogenic Shock

Edward W. Hook III, MD
Professor and Director, Division of Infectious Diseases, University of Alabama at Birmingham, Birmingham, Alabama
Granuloma Inguinale (Donovanosis); Syphilis; Nonsyphilitic Treponematoses

David J. Hunter, MBBS, MPH, ScD
Vincent L. Gregory Professor of Cancer Prevention, Harvard School of Public Health; Professor of Medicine, Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts
The Epidemiology of Cancer

Khalid Hussain, MBChB, MD, MSc
Developmental Endocrinology Research Group, Clinical and Molecular Genetics Unit, Institute of Child Health, University College London, Department of Paediatric Endocrinology, Great Ormond Street Hospital for Children, London, United Kingdom
Hypoglycemia/Pancreatic Islet Cell Disorders

Steven E. Hyman, MD
Director, Stanley Center for Psychiatric Research, Broad Institute, Distinguished Service Professor of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts
Biology of Addiction

Michael C. Iannuzzi, MD, MBA
Chairman, Department of Internal Medicine, State University of New York Upstate Medical University, Syracuse, New York
Sarcoidosis

Robert D. Inman, MD
Professor of Medicine and Immunology, University of Toronto; Staff Rheumatologist, University Health Network, Toronto, Ontario, Canada
The Spondyloarthropathies

Sharon K. Inouye, MD, MPH
Professor of Medicine, Harvard Medical School; Director, Aging Brain Center, Institute for Aging Research, Hebrew SeniorLife, Boston, Massachusetts
Neuropsychiatric Aspects of Aging; Delirium or Acute Mental Status Change in the Older Patient

Geoffrey K. Isbister, MD, BSc
Associate Professor, Clinical Toxicologist, Calvary Mater Newcastle, Callaghan, Senior Research Academic, School of Medicine and Public Health, University of Newcastle, New South Wales, Australia
Envenomation

Michael G. Ison, MD, MS
Associate Professor in Medicine-Infectious Diseases and Surgery-Organ Transplantation, Northwestern University Feinberg School of Medicine, Chicago, Illinois
Adenovirus Diseases

Elias Jabbour, MD
Associate Professor, Department of Leukemia, Division of Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas
The Chronic Leukemias

Michael R. Jaff, DO
Professor of Medicine, Harvard Medical School, Chair, Institute for Heart, Vascular, and Stroke Care, Massachusetts General Hospital, Boston, Massachusetts
Other Peripheral Arterial Diseases

Joanna C. Jen, MD, PhD
Professor of Neurology, University of California Los Angeles School of Medicine, Los Angeles, California
Neuro-Ophthalmology; Smell and Taste; Hearing and Equilibrium

Dennis M. Jensen, MD
Professor of Medicine, David Geffen School of Medicine at University of California Los Angeles; Staff Physician, Medicine-GI, VA Greater Los Angeles Healthcare System; Key Investigator, Director, Human Studies Core & GI Hemostasis Research Unit, CURE Digestive Diseases Research Center, Los Angeles, California
Gastrointestinal Hemorrhage

Michael D. Jensen, MD
Professor of Medicine, Endocrine Research Unit, Director, Obesity Treatment Research Program, Mayo Clinic, Rochester, Minnesota
Obesity

Robert T. Jensen, MD
Chief, Cell Biology Section, Digestive Disease Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Clinical Center, Bethesda, Maryland
Pancreatic Neuroendocrine Tumors

Stuart Johnson, MD
Professor of Medicine, Loyola University Chicago Stritch School of Medicine; Associate Chief of Staff for Research, Edward Hines, Jr. VA Hospital, Hines, Illinois
Clostridial Infections

Richard C. Jordan, DDS, PhD
Professor of Oral Pathology, Pathology and Radiation Oncology, University of California San Francisco, San Francisco, California
Diseases of the Mouth and Salivary Glands

Ralph F. Józefowicz, MD
Professor, Neurology and Medicine, University of Rochester, Rochester, New York
Approach to the Patient with Neurologic Disease

Stephen G. Kaler, MD
Senior Investigator and Head, Section on Translational Neuroscience, Molecular Medicine Program, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland
Wilson Disease

Moses R. Kamya, MB ChB, MMed, MPH, PhD
Chairman, Department of Medicine, Makerere University College of Health Sciences, Kampala, Uganda
Malaria

Louise W. Kao, MD
Associate Professor of Emergency Medicine, Department of Emergency Medicine, Indiana University School of Medicine, Indianapolis, Indiana
Chronic Poisoning: Trace Metals and Others

Steven A. Kaplan, MD
E. Darracott Vaughan, Jr. Professor of Urology, Chief, Institute for Bladder and Prostate Health, Weill Cornell Medical College; Director, Iris Cantor Men's Health Center, NewYork–Presbyterian Hospital, New York, New York
Benign Prostatic Hyperplasia and Prostatitis

Daniel L. Kastner, MD, PhD
Scientific Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland
The Systemic Autoinflammatory Diseases

Sekar Kathiresan, MD
Associate Professor in Medicine, Harvard Medical School; Director, Preventive Cardiology, Massachusetts General Hospital, Boston, Massachusetts
The Inherited Basis of Common Diseases

David A. Katzka, MD
Professor of and Consultant in Medicine, Gastroenterology, Mayo Clinic, Rochester, Minnesota
Diseases of the Esophagus

Debra K. Katzman, MD
Professor of Pediatrics, Senior Associate Scientist, The Research Institute, The Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada
Adolescent Medicine

Carol A. Kauffman, MD
Professor of Internal Medicine, University of Michigan Medical School; Chief, Infectious Diseases Section, Veterans Affairs Ann Arbor Healthcare System, Ann Arbor, Michigan
Histoplasmosis; Blastomycosis; Paracoccidioidomycosis; Cryptococcosis; Sporotrichosis; Candidiasis

Kenneth Kaushansky, MD
Senior Vice President for Health Sciences, Dean, School of Medicine, Stony Brook University, Stony Brook, New York
Hematopoiesis and Hematopoietic Growth Factors

Keith S. Kaye, MD, MPH
Professor of Medicine, Division of Infectious Diseases, Wayne State University School of Medicine, Detroit, Michigan
Diseases Caused by Acinetobacter and Stenotrophomonas Species

Armand Keating, MD
Professor of Medicine, Director, Division of Hematology, Epstein Chair in Cell Therapy and Transplantation, Professor, Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada
Hematopoietic Stem Cell Transplantation

Robin K. Kelley, MD
Assistant Professor of Medicine, University of California San Francisco, Helen Diller Family Comprehensive Cancer Center, San Francisco, California
Liver and Biliary Tract Cancers

Morton Kern, MD
Chief of Medicine, VA Long Beach Health Care System School of Medicine; Professor of Medicine, Associate Chief, Cardiology, University of California–Irvine, Irvine, California
Catheterization and Angiography

Gerald T. Keusch, MD
Professor of Medicine and International Health and Public Health, Boston University School of Medicine, Boston, Massachusetts
Shigellosis

Fadlo R. Khuri, MD
Professor and Chair, Hematology and Medical Oncology, Deputy Director, Winship Cancer Institute, Emory University, Atlanta, Georgia
Lung Cancer and Other Pulmonary Neoplasms

David H. Kim, MD
Vice Chair of Education, Professor of Radiology, Section of Abdominal Imaging, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin
Diagnostic Imaging Procedures in Gastroenterology

Matthew Kim, MD
Instructor of Medicine, Harvard Medical School; Associate Physician, Brigham and Women's Hospital, Boston, Massachusetts
Thyroid

Louis V. Kirchhoff, MD, MPH
Professor, Departments of Internal Medicine (Infectious Diseases) and Epidemiology, University of Iowa Health Care; Staff Physician, Medical Service, Department of Veterans Affairs Medical Center, Iowa City, Iowa
Chagas Disease

David S. Knopman, MD
Professor of Neurology, Mayo Clinic College of Medicine, Rochester, Minnesota
Regional Cerebral Dysfunction: Higher Mental Function; Alzheimer Disease and Other Dementias

Tamsin A. Knox, MD, MPH
Associate Professor of Medicine, Nutrition/Infection Unit, Tufts University School of Medicine, Boston, Massachusetts
Gastrointestinal Manifestions of HIV and AIDS

D.P. Kontoyiannis, MD, ScD
Professor, Department of Infectious Diseases, Infection Control and Employee Health, The University of Texas MD Anderson Cancer Center, Houston, Texas
Mucormycosis; Mycetoma

Barbara S. Koppel, MD
Professor of Clinical Neurology, New York Medical College, Chief of Neurology, Metropolitan Hospital Center, New York City Health and Hospital Corporation, New York, New York
Nutritional and Alcohol-Related Neurologic Disorders

Kevin M. Korenblat, MD
Associate Professor of Medicine, Department of Medicine, Washington University School of Medicine, St. Louis, Missouri
Approach to the Patient with Jaundice or Abnormal Liver Tests

Bruce R. Korf, MD, PhD
Wayne H. and Sara Crews Finley Chair in Medical Genetics, Professor and Chair, Department of Genetics, University of Alabama at Birmingham, Birmingham, Alabama
Principles of Genetics

Neil J. Korman, MD, PhD
Professor, Dermatology, Case Western Reserve University School of Medicine, University Hospitals Case Medical Center, Cleveland, Ohio
Macular, Papular, Vesiculobullous, and Pustular Diseases

Mark G. Kortepeter, MD, MPH
Associate Dean for Research, Associate Professor of Preventive Medicine and Medicine, Consultant to the Army Surgeon General for Biodefense; Office of the Dean, Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland
Bioterrorism

Joseph A. Kovacs, MD
Senior Investigator and Head, AIDS Section, Critical Care Medicine Department, National Institutes of Health, Bethesda, Maryland
Pneumocystis Pneumonia

Thomas O. Kovacs, MD
Professor of Medicine, Division of Digestive Diseases, David Geffen School of Medicine at University of California Los Angeles, Los Angeles, California
Gastrointestinal Hemorrhage

Monica Kraft, MD
Professor of Medicine, Duke University School of Medicine; Chief, Division of Pulmonary, Allergy, and Critical Care Medicine, Duke University Medical Center, Durham, North Carolina
Approach to the Patient with Respiratory Disease

Christopher M. Kramer, MD
Ruth C. Heede Professor of Cardiology, Professor of Radiology, Director, Cardiovascular Imaging Center, University of Virginia Health System, Charlottesville, Virginia
Noninvasive Cardiac Imaging

Donna M. Krasnewich, MD, PhD
Program Director, National Institute of General Medical Sciences, National Institutes of Health, Bethesda, Maryland
The Lysosomal Storage Diseases

Peter J. Krause, MD
Senior Research Scientist in Epidemiology, Medicine, and Pediatrics, Yale School of Public Health and Yale School of Medicine, New Haven, Connecticut
Babesiosis and Other Protozoan Diseases

John F. Kuemmerle, MD
Chair, Division of Gastroenterology, Hepatology, and Nutrition, Professor of Medicine, and Physiology and Biophysics, Center for Digestive Health, Virginia Commonwealth University, Richmond, Virginia
Inflammatory and Anatomic Diseases of the Intestine, Peritoneum, Mesentery, and Omentum

Ernst J. Kuipers, MD, PhD
Professor of Medicine, Department of Gastroenterology and Hepatology, Chief Executive Officer, Erasmus MC University Medical Center, Rotterdam, The Netherlands
Acid Peptic Disease

Paul W. Ladenson, MD
Professor of Medicine, Pathology, Oncology, and Radiology and Radiological Sciences, John Eager Howard Professor of Endocrinology and Metabolism, University Distinguished Service Professor, The Johns Hopkins University School of Medicine; Physician and Division Director, The Johns Hopkins Hospital, Baltimore, Maryland
Thyroid

Daniel Laheru, MD
Ian T. MacMillan Professorship in Clinical Pancreatic Research, Medical Oncology, Johns Hopkins University School of Medicine, Baltimore, Maryland
Pancreatic Cancer

Donald W. Landry, MD, PhD
Samuel Bard Professor of Medicine, Chair, Department of Medicine, Physician-in-Chief, NewYork-Presbyterian Hospital/Columbia University Medical Center, New York, New York
Approach to the Patient with Renal Disease

Anthony E. Lang, MD
Director, Division of Neurology, Jack Clark Chair for Research in Parkinson's Disease, University of Toronto; Director, Morton and Gloria Shulman Movement Disorders Clinic and the Edmond J. Safra Program in Parkinson's Disease and the Lily Safra Chair in Movement Disorders, Toronto Western Hospital, Toronto, Ontario, Canada
Parkinsonism; Other Movement Disorders

Richard A. Lange, MD, MBA
President and Dean, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center El Paso, El Paso, Texas
Acute Coronary Syndrome: Unstable Angina and Non-ST Elevation Myocardial Infarction

Frank A. Lederle, MD
Core Investigator, Center for Chronic Disease Outcomes Research, Minneapolis VA Medical Center; Professor of Medicine, University of Minnesota School of Medicine, Minneapolis, Minnesota
Diseases of the Aorta

Thomas H. Lee, MD, MSc
Senior Physician, Department of Medicine, Brigham and Women's Hospital; Chief Medical Officer, Press Ganey, Boston, Massachusetts
Using Data for Clinical Decisions

William M. Lee, MD
Professor of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas
Toxin- and Drug-Induced Liver Disease

James E. Leggett, MD
Associate Professor, Department of Medicine, Oregon Health and Science University; Infectious Diseases, Department of Medical Education, Providence Portland Medical Center, Portland, Oregon
Approach to Fever or Suspected Infection in the Normal Host

Stuart Levin, MD
Professor of Medicine, Emeritus Chairman, Department of Medicine, Rush University Medical Center, Chicago, Illinois
Zoonoses

Stephanie M. Levine, MD
Professor of Medicine, Division of Pulmonary Diseases and Critical Care Medicine, The University of Texas Health Science Center San Antonio, South Texas Veterans Health Care System, San Antonio, Texas
Alveolar Filling Disorders

Gary R. Lichtenstein, MD
Professor of Medicine, Perelman School of Medicine at the University of Pennsylvania, Director, Center for Inflammatory Bowel Disease, University of Pennsylvania, Philadelphia, Pennsylvania
Inflammatory Bowel Disease

Henry W. Lim, MD
Chairman and C.S. Livingood Chair, Department of Dermatology, Henry Ford Hospital; Senior Vice President for Academic Affairs, Henry Ford Health System, Detroit, Michigan
Eczemas, Photodermatoses, Papulosquamous (Including Fungal) Diseases, and Figurate Erythemas

Aldo A.M. Lima, MD, PhD
Professor of Medicine and Pharmacology, School of Medicine, Federal University of Ceará, Fortaleza, Ceará, Brazil
Cryptosporidiosis; Amebiasis

Geoffrey S.F. Ling, MD, PhD
Professor of Neurology, Uniformed Services University of the Health Sciences, Bethesda, Maryland
Traumatic Brain Injury and Spinal Cord Injury

William C. Little, MD
Patrick Lehan Professor of Cardiovascular Medicine, Chair, Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi
Pericardial Diseases

Donald M. Lloyd-Jones, MD, ScM
Senior Associate Dean, Chair, Department of Preventive Medicine, Eileen M. Foell Professor of Preventive Medicine and Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois
Epidemiology of Cardiovascular Disease

Bennett Lorber, MD
Thomas M. Durant Professor of Medicine and Professor of Microbiology and Immunology, Temple University School of Medicine, Philadelphia, Pennsylvania
Listeriosis

Donald E. Low, MD[†]
Nonpneumococcal Streptococcal Infections, Rheumatic Fever

Daniel R. Lucey, MD, MPH
Adjunct Professor, Microbiology and Immunology, Georgetown University Medical Center, Washington, D.C.
Anthrax

James R. Lupski, MD, PhD
Cullen Professor of Molecular and Human Genetics, Professor of Pediatrics, Baylor College of Medicine and Texas Children's Hospital, Houston, Texas
Gene, Genomic, and Chromosomal Disorders

Jeffrey M. Lyness, MD
Senior Associate Dean for Academic Affairs, Professor of Psychiatry and Neurology, University of Rochester School of Medicine and Dentistry, Rochester, New York
Psychiatric Disorders in Medical Practice

Bruce W. Lytle, MD
Chair, Heart and Vascular Institute, Professor of Surgery, Thoracic and Cardiovascular Surgery, Cleveland Clinic, Cleveland, Ohio
Interventional and Surgical Treatment of Coronary Artery Disease

[†]Deceased.

C. Ronald MacKenzie, MD
Assistant Attending Physician, Department of Medicine-Rheumatology, C. Ronald MacKenzie Chair in Ethics and Medicine, Hospital for Special Surgery, Associate Professor of Clinical Medicine and Medical Ethics, Weill Cornell Medical College of Cornell University, New York, New York
Surgical Treatment of Joint Disease

Harriet L. MacMillan, MD, MSc
Professor, Departments of Psychiatry and Behavioural Neurosciences, and Pediatrics, Chedoke Health Chair in Child Psychiatry, Offord Centre for Child Studies, McMaster University, Hamilton, Ontario, Canada
Intimate Partner Violence

Robert D. Madoff, MD
Professor of Surgery, Stanley M. Goldberg, MD, Chair, Colon and Rectal Surgery, University of Minnesota, Minneapolis, Minnesota
Diseases of the Rectum and Anus

Frank Maldarelli, MD, PhD
Head, Clinical Retrovirology Section, HIV Drug Resistance Program, National Cancer Institute, National Institutes of Health, Bethesda, Maryland
Biology of Human Immunodeficiency Viruses

Atul Malhotra, MD
Chief of Pulmonary and Critical Care, Kenneth M. Moser Professor of Medicine, Director of Sleep Medicine, University of California San Diego, La Jolla, California
Disorders of Ventilatory Control

Mark J. Manary, MD
Helene B. Roberson Professor of Pediatrics, Washington University School of Medicine; Attending Physician, St. Louis Children's Hospital, St. Louis, Missouri; Adjunct Professor, Children's Nutrition Research Center, Baylor College of Medicine, Houston, Texas; Senior Lecturer in Community Health, University of Malawi College of Medicine, Blantyre, Malawi
Protein-Energy Malnutrition

Donna Mancini, MD
Professor of Medicine, Department of Medicine, Division of Cardiology, Columbia University College of Physicians and Surgeons, Center for Advanced Cardiac Care, Columbia University Medical Center, New York, New York
Cardiac Transplantation

Lionel A. Mandell, MD
Professor of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada
Streptococcus Pneumoniae *Infections*

Peter Manu, MD
Professor of Medicine and Psychiatry, Hofstra North Shore–LIJ School of Medicine at Hofstra University, Hempstead, New York; Adjunct Professor of Clinical Medicine, Psychiatry and Behavioral Sciences, Albert Einstein College of Medicine, Bronx, New York; Director of Medical Services, Zucker Hillside Hospital, Glen Oaks, New York
Medical Consultation in Psychiatry

Ariane Marelli, MD, MPH
Professor of Medicine, McGill University, Director, McGill Adult Unit for Congenital Heart Disease, Associate Director, Academic Affairs and Research, Cardiology, McGill University Health Centre, Montreal, Québec, Canada
Congenital Heart Disease in Adults

Xavier Mariette, MD, PhD
Professor, Rheumatology, Université Paris-Sud, AP-HP, Le Kremlin Bicêtre, France
Sjögren Syndrome

Andrew R. Marks, MD
Wu Professor and Chair, Department of Physiology and Cellular Biophysics, Founding Director, Helen and Clyde Wu Center for Molecular Cardiology, Columbia University College of Physicians and Surgeons, New York, New York
Cardiac Function and Circulatory Control

Kieren A. Marr, MD
Professor of Medicine and Oncology, The Johns Hopkins University, Director, Transplant and Oncology Infectious Diseases, Baltimore, Maryland
Approach to Fever and Suspected Infection in the Compromised Host

Thomas J. Marrie, MD
Dean, Faculty of Medicine, Dalhousie University; Professor of Medicine, Capital District Health Authority, Halifax, Nova Scotia, Canada
Legionella Infections

Paul Martin, MD
Professor of Medicine and Chief, Division of Hepatology, Miller School of Medicine, University of Miami, Miami, Florida
Approach to the Patient with Liver Disease

Joel B. Mason, MD
Professor of Medicine and Nutrition, Tufts University; Staff Physician, Divisions of Gastroenterology and Clinical Nutrition, Tufts Medical Center, Boston, Massachusetts
Vitamins, Trace Minerals, and Other Micronutrients

Henry Masur, MD
Chief, Critical Care Medicine Department, Clinical Center, National Institutes of Health, Bethesda, Maryland
Infectious and Metabolic Complications of HIV and AIDS

Eric L. Matteson, MD, MPH
Professor of Medicine, Mayo Clinic College of Medicine, Consultant, Divisions of Rheumatology and Epidemiology, Mayo Clinic, Rochester, Minnesota
Infections of Bursae, Joints, and Bones

Michael A. Matthay, MD
Professor, Departments of Medicine and Anesthesia, University of California San Francisco, San Francisco, California
Acute Respiratory Failure

Toby A. Maurer, MD
Professor of Dermatology, University of California San Francisco; Chief of Dermatology, San Francisco General Hospital, San Francisco, California
Skin Manifestations in Patients with Human Immunodeficiency Virus Infection

Emeran A. Mayer, MD, PhD
Professor of Medicine, Physiology, and Psychiatry, Division of Digestive Diseases, Department of Medicine, University of California Los Angeles, Los Angeles, California
Functional Gastrointestinal Disorders: Irritable Bowel Syndrome, Dyspepsia, Chest Pain of Presumed Esophageal Origin, and Heartburn

Stephan A. Mayer, MD
Director, Institute for Critical Care Medicine, Icahn School of Medicine at Mount Sinai, New York, New York
Hemorrhagic Cerebrovascular Disease

Stephen A. McClave, MD
Professor of Medicine, Director of Clinical Nutrition, University of Louisville School of Medicine, Louisville, Kentucky
Enteral Nutrition

F. Dennis McCool, MD
Professor of Medicine, The Warren Alpert Medical School of Brown University; Medical Director of Sleep Center, Memorial Hospital of Rhode Island, Pawtucket, Rhode Island
Diseases of the Diaphragm, Chest Wall, Pleura, and Mediastinum

Charles E. McCulloch, PhD
Professor of Biostatistics, Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California
Statistical Interpretation of Data

William J. McKenna, MD
Professor of Cardiology, Institute of Cardiovascular Science, University College London, London, United Kingdom
Diseases of the Myocardium and Endocardium

Vallerie McLaughlin, MD
Kim A. Eagle, MD, Endowed Professor of Cardiovascular Medicine, Director, Pulmonary Hypertension Program, University of Michigan, Ann Arbor, Michigan
Pulmonary Hypertension

John J.V. McMurray, MB, MD
Professor of Cardiology, Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, Scotland, United Kingdom
Heart Failure: Management and Prognosis

Kenneth R. McQuaid, MD
Professor of Clinical Medicine, Marvin H. Sleisenger Endowed Chair, Vice Chairman, University of California San Francisco; Chief, Medical Services and Gastroenterology, San Francisco VA Medical Center, San Francisco, California
Approach to the Patient with Gastrointestinal Disease

Marc Michel, MD
Professor of Internal Medicine, Head of the Unit of Internal Medicine at Henri Mondor University Hospital, National Referral Center for Adult's Immune Cytopenias, Creteil, France
Autoimmune and Intravascular Hemolytic Anemias

Jonathan W. Mink, MD, PhD
Frederick A. Horner, MD Endowed Professor in Pediatric Neurology, Professor of Neurology, Neurobiology & Anatomy, Brain & Cognitive Sciences, and Pediatrics, Chief, Division of Child Neurology, Vice Chair, Department of Neurology, University of Rochester, Rochester, New York
Congenital, Developmental, and Neurocutaneous Disorders

William E. Mitch, MD
Gordon A. Cain Chair in Nephrology, Director of Nephrology, Baylor College of Medicine, Houston, Texas
Chronic Kidney Disease

Mark E. Molitch, MD
Martha Leland Sherwin Professor of Endocrinology, Division of Endocrinology, Metabolism, and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois
Neuroendocrinology and the Neuroendocrine System; Anterior Pituitary

Bruce A. Molitoris, MD
Professor of Medicine, and Cellular and Integrative Physiology Director, Indiana Center for Biological Microscopy, Indiana University, Indianapolis, Indiana
Acute Kidney Injury

Jose G. Montoya, MD
Professor of Medicine, Division of Infectious Disease and Geographic Medicine, Stanford University School of Medicine, Stanford, California; Director, Palo Alto Medical Foundation Toxoplasma Serology Laboratory, National Reference Center for the Study and Diagnosis of Toxoplasmosis, Palo Alto, California
Toxoplasmosis

Alison Morris, MD, MS
Associate Professor of Medicine, Clinical Translational Science, and Immunology, Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania
Pulmonary Manifestations of Human Immunodeficiency Virus and Acquired Immunodeficiency Syndrome

Ernest Moy, MD, MPH
Medical Officer, Center for Quality Improvement and Patient Safety Agency for Healthcare Research and Quality, Rockville, Maryland
Measuring Health and Health Care

Atis Muehlenbachs, MD, PhD
Infectious Diseases Pathology Branch, Centers for Disease Control and Prevention, Atlanta, Georgia
Leptospirosis

Andrew H. Murr, MD
Professor and Chairman, Roger Boles, MD Endowed Chair in Otolaryngology Education, Department of Otolaryngology-Head and Neck Surgery, University of California San Francisco School of Medicine, San Francisco, California
Approach to the Patient with Nose, Sinus, and Ear Disorders

Daniel M. Musher, MD
Professor of Medicine, Molecular Virology, and Microbiology, Distinguished Service Professor, Baylor College of Medicine, Infectious Disease Section, Michael E. DeBakey Veterans Affairs Medical Center, Houston, Texas
Overview of Pneumonia

Robert J. Myerburg, MD
Professor of Medicine and Physiology, Division of Cardiology, Department of Medicine, American Heart Association Chair in Cardiovascular Research, University of Miami Miller School of Medicine, Miami, Florida
Approach to Cardiac Arrest and Life-Threatening Arrhythmias

Sandesh C.S. Nagamani, MD
Assistant Professor, Department of Molecular and Human Genetics, Director, Clinic for Metabolic and Genetic Disorders of Bone, Baylor College of Medicine and Texas Children's Hospital, Houston, Texas
Gene, Genomic, and Chromosomal Disorders

Stanley J. Naides, MD
Medical Director and Interim Scientific Director, Immunology, Quest Diagnostics Nichols Institute, San Juan Capistrano, California
Arboviruses Causing Fever and Rash Syndromes

Yoshifumi Naka, MD, PhD
Professor of Surgery, Department of Surgery, Columbia University College of Physicians and Surgeons, New York, New York
Cardiac Transplantation

Theodore E. Nash, MD
Principal Investigator, Clinical Parasitology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland
Giardiasis

Avindra Nath, MD
Chief, Section of Infections of the Nervous System, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland
Cytomegalovirus, Epstein-Barr Virus, and Slow Virus Infections of the Central Nervous System; Neurologic Complications of Human Immunodeficiency Virus Infection; Meningitis: Bacterial, Viral, and Other; Brain Abscess and Parameningeal Infections

Eric G. Neilson, MD
Vice President for Medical Affairs and Lewis Landsberg Dean, Northwestern University Feinberg School of Medicine, Northwestern Memorial Hospital, Chicago, Illinois
Tubulointerstitial Nephritis

Lawrence S. Neinstein, MD
Professor of Pediatrics and Medicine, Keck School of Medicine of USC; Executive Director, Engemann Student Health Center, Division Head of College Health, Assistant Provost, Student Health and Wellness, University of Southern California, Los Angeles, California
Adolescent Medicine

Lewis S. Nelson, MD
Professor of Emergency Medicine, Director, Fellowship in Medical Toxicology, New York University School of Medicine; Attending Physician, New York University Langone Medical Center and Bellevue Hospital Center, New York, New York
Acute Poisoning

Eric J. Nestler, MD, PhD
Nash Family Professor and Chair, Department of Neuroscience, Director, The Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York
Biology of Addiction

Anne B. Newman, MD, MPH
Professor of Epidemiology, The University of Pittsburgh Graduate School of Public Health; Chair, Department of Epidemiology, Director, University of Pittsburgh Center for Aging and Population Health, Pittsburgh, Pennsylvania
Epidemiology of Aging: Implications of the Aging of Society

Thomas B. Newman, MD, MPH
Professor, Epidemiology & Biostatistics and Pediatrics, University of California San Francisco, San Francisco, California
Statistical Interpretation of Data

William L. Nichols, MD
Associate Professor, Medicine and Laboratory Medicine, Mayo Clinic College of Medicine; Staff Physician, Special Coagulation Laboratory, Comprehensive Hemophilia Center, and Coagulation Clinic, Mayo Clinic, Rochester, Minnesota
Von Willebrand Disease and Hemorrhagic Abnormalities of Platelet and Vascular Function

Lindsay E. Nicolle, MD
Professor of Internal Medicine and Medical Microbiology, University of Manitoba, Health Sciences Centre, Winnipeg, Manitoba, Canada
Approach to the Patient with Urinary Tract Infection

Lynnette K. Nieman, MD
Senior Investigator, Program on Reproductive and Adult Endocrinology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland
Approach to the Patient with Endocrine Disease; Adrenal Cortex; Polyglandular Disorders

Dennis E. Niewoehner, MD
Professor of Medicine, University of Minnesota; Staff Physician, Minneapolis Veterans Affairs Health Care System, Minneapolis, Minnesota
Chronic Obstructive Pulmonary Disease

S. Ragnar Norrby, MD, PhD
Director General, Swedish Institute for Infectious Disease Control, Solna, Sweden
Approach to the Patient with Urinary Tract Infection

Susan O'Brien, MD
Professor, Department of Leukemia, Division of Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas
The Chronic Leukemias

Christopher M. O'Connor, MD
Professor of Medicine and Chief, Division of Cardiology, Director, Duke Heart Center, Durham, North Carolina
Heart Failure: Pathophysiology and Diagnosis

Francis G. O'Connor, MD, MPH
Professor and Chair, Military and Emergency Medicine, Medical Director, Uniformed Services University Consortium for Health and Military Performance, Bethesda, Maryland
Disorders Due to Heat and Cold; Rhabdomyolysis

Patrick G. O'Connor, MD, MPH
Professor and Chief, General Internal Medicine, Yale University School of Medicine, New Haven, Connecticut
Alcohol Abuse and Dependence

James R. O'Dell, MD
Bruce Professor and Vice Chair of Internal Medicine, Chief, Division of Rheumatology, University of Nebraska Medical Center and Omaha VA Nebraska–Western Iowa Health Care System, Omaha, Nebraska
Rheumatoid Arthritis

Anne E. O'Donnell, MD
Professor of Medicine, Chief, Division of Pulmonary, Critical Care, and Sleep Medicine, Georgetown University Medical Center, Washington, D.C.
Bronchiectasis, Atelectasis, Cysts, and Localized Lung Disorders

Jae K. Oh, MD
Professor of Medicine, Director, Echocardiography Core Laboratory and Pericardial Clinic, Division of Cardiovascular Diseases, Co-Director, Integrated Cardiac Imaging, Division of Cardiovascular Diseases, Mayo Clinic, Rochester, Minnesota
Pericardial Diseases

Jeffrey E. Olgin, MD
Gallo-Chatterjee Distinguished Professor of Medicine, Chief, Division of Cardiology, Co-Director, Heart and Vascular Center, University of California San Francisco, San Francisco, California
Approach to the Patient with Suspected Arrhythmia

Walter A. Orenstein, MD
Professor of Medicine, Pediatrics, and Global Health, Emory University School of Medicine, Atlanta, Georgia
Immunization

Douglas R. Osmon, MD, MPH
Professor of Medicine, Mayo Clinic College of Medicine; Consultant, Division of Infectious Diseases, Mayo Clinic, Rochester, Minnesota
Infections of Bursae, Joints, and Bones

Catherine M. Otto, MD
J. Ward Kennedy-Hamilton Endowed Chair in Cardiology, Professor of Medicine, University of Washington School of Medicine; Director, Heart Valve Clinic, University of Washington Medical Center, Seattle, Washington
Echocardiography

Mark Papania, MD, MPH
Medical Epidemiologist, Division of Viral Diseases, Measles, Mumps, Rubella, and Herpes Virus Laboratory Branch, Centers for Disease Control and Prevention, Atlanta, Georgia
Measles

Peter G. Pappas, MD
Professor of Medicine, University of Alabama at Birmingham, Birmingham, Alabama
Dematiaceous Fungal Infections

Pankaj Jay Pasricha, MD
Director, The Johns Hopkins Center for Neurogastroenterology; Professor of Medicine and Neurosciences, The Johns Hopkins School of Medicine; Professor of Innovation Management, Johns Hopkins Carey Business School, Baltimore, Maryland
Gastrointestinal Endoscopy

David L. Paterson, MD
Professor of Medicine, University of Queensland Centre for Clinical Research, Royal Brisbane and Women's Hospital Campus, Brisbane, Queensland, Australia
Infections Due to Other Members of the Enterobacteriaceae, Including Management of Multidrug Resistant Strains

Carlo Patrono, MD
Professor and Chair of Pharmacology, Department of Pharmacology, Catholic University School of Medicine, Rome, Italy
Prostaglandin, Aspirin, and Related Compounds

Jean-Michel Pawlotsky, MD, PhD
Professor of Medicine, The University of Paris-Est; Director, National Reference Center for Viral Hepatitis B, C, and Delta and Department of Virology, Henri Mondor University Hospital; Director, Department of Molecular Virology and Immunology, Institut Mondor de Recherche Biomédicale, Créteil, France
Acute Viral Hepatitis; Chronic Viral and Autoimmune Hepatitis

Richard D. Pearson, MD
Professor of Medicine and Pathology, University of Virginia School of Medicine and University of Virginia Health System, Charlottesville, Virginia
Antiparasitic Therapy

Trish M. Perl, MD, MSc
Professor of Medicine and Pathology, The Johns Hopkins School of Medicine; Professor of Epidemiology, Johns Hopkins Bloomberg School of Public Health; Infectious Diseases Specialist and Senior Epidemiologist, The Johns Hopkins Hospital and Health System, Baltimore, Maryland
Enterococcal Infections

Adam Perlman, MD, MPH
Associate Professor, Department of Medicine, Duke University Medical Center; Executive Director, Duke Integrative Medicine, Duke University Health System, Durham, North Carolina
Complementary and Alternative Medicine

William A. Petri, Jr., MD, PhD
Wade Hampton Frost Professor, Departments of Medicine, Pathology, Microbiology, Immunology, and Cancer Biology, School of Medicine, University of Virginia; Chief, Division of Infectious Diseases and International Health, University of Virginia Hospitals, Charlottesville, Virginia
Relapsing Fever and Other Borrelia Infections; African Sleeping Sickness; Amebiasis

Marc A. Pfeffer, MD, PhD
Dzau Professor of Medicine, Harvard Medical School; Senior Physician, Cardiovascular Division, Brigham and Women's Hospital, Boston, Massachusetts
Heart Failure: Management and Prognosis

Perry J. Pickhardt, MD
Professor of Radiology and Chief, Gastrointestinal Imaging, Section of Abdominal Imaging, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin
Diagnostic Imaging Procedures in Gastroenterology

David S. Pisetsky, MD, PhD
Chief of Rheumatology, Medical Research Service, Durham VA Medical Center; Professor of Medicine and Immunology, Department of Medicine, Duke University Medical Center, Durham, North Carolina
Laboratory Testing in the Rheumatic Diseases

Marshall R. Posner, MD
Professor of Medicine, Director of Head and Neck Medical Oncology, Director of the Office of Cancer Clinical Trials, The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, New York
Head and Neck Cancer

Frank Powell, PhD
Professor of Medicine, Chief of Physiology, University of California San Diego, La Jolla, California
Disorders of Ventilatory Control

Reed E. Pyeritz, MD, PhD
William Smilow Professor of Medicine and Genetics and Vice Chair for Academic Affairs, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania
Inherited Diseases of Connective Tissue

Thomas C. Quinn, MD, MSc
Associate Director for International Research, Head, Section of International HIV/AIDS Research, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health; Professor of Medicine, Pathology, International Health, Molecular Microbiology and Immunology, and Epidemiology, The Johns Hopkins Medical Institutions, Baltimore, Maryland
Epidemiology and Diagnosis of Human Immunodeficiency Virus Infection and Acquired Immunodeficiency Syndrome

Jai Radhakrishnan, MD, MS
Professor of Medicine, Division of Nephrology, Department of Medicine, Columbia University Medical Center; Associate Division Chief for Clinical Affairs, Division of Nephrology, New York-Presbyterian Hospital, New York, New York
Glomerular Disorders and Nephrotic Syndromes

Petros I. Rafailidis, MD, PhD, MSc
Senior Researcher, Alfa Institute of Biomedical Sciences, Attending Physician, Department of Medicine and Hematology, Athens Medical Center, Athens Medical Group, Athens, Greece
Pseudomonas and Related Gram-Negative Bacillary Infections

Ganesh Raghu, MD
Adjunct Professor of Medicine and Laboratory Medicine, University of Washington, Director, CENTER for Interstitial Lung Diseases at the University of Washington; Co-Director, Scleroderma Clinic, University of Washington Medical Center, Seattle, Washington
Interstitial Lung Disease

Margaret Ragni, MD, MPH
Professor of Medicine and Clinical Translational Science, Department of Hematology/Oncology, University of Pittsburgh Medical Center; Director, Hemophilia Center of Western Pennsylvania, Pittsburgh, Pennsylvania
Hemorrhagic Disorders: Coagulation Factor Deficiencies

Srinivasa N. Raja, MD
Professor of Anesthesiology and Neurology, Director, Division of Pain Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland
Pain

S. Vincent Rajkumar, MD
Professor of Medicine, Division of Hematology, Mayo Clinic, Rochester, Minnesota
Plasma Cell Disorders

Stuart H. Ralston, MB ChB, MD
Professor of Rheumatology, Institute of Genetics and Molecular Medicine, Western General Hospital, The University of Edinburgh, Edinburgh, United Kingdom
Paget Disease of Bone

Didier Raoult, MD, PhD
Professor, Aix Marseille Université, Faculté de Médecine; Chief, Hôpital de la Timone, Fédération de Microbiologie Clinique, Marseille, France
Bartonella Infections; Rickettsial Infections

Robert W. Rebar, MD
Professor, Department of Obstetrics and Gynecology, Western Michigan University Homer Stryker MD School of Medicine, Kalamazoo, Michigan
Ovaries and Development; Reproductive Endocrinology and Infertility

Annette C. Reboli, MD
Founding Vice Dean, Professor of Medicine, Cooper Medical School of Rowan University, Cooper University Healthcare, Department of Medicine, Division of Infectious Diseases, Camden, New Jersey
Erysipelothrix Infections

K. Rajender Reddy, MD
Professor of Medicine, Professor of Medicine in Surgery, Perelman School of Medicine at the University of Pennsylvania; Director of Hepatology, Director, Viral Hepatitis Center, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania
Bacterial, Parasitic, Fungal, and Granulomatous Liver Diseases

Donald A. Redelmeier, MD
Professor of Medicine, University of Toronto; Senior Scientist and Staff Physician, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada
Postoperative Care and Complications

Susan E. Reef, MD
Centers for Disease Control and Prevention, Atlanta, Georgia
Rubella (German Measles)

Neil M. Resnick, MD
Thomas P. Detre Endowed Chair in Gerontology and Geriatric Medicine, Professor of Medicine and Division Chief, Geriatrics, Associate Director, University of Pittsburgh Institute on Aging, University of Pittsburgh; Chief, Division of Geriatric Medicine and Gerontology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania
Incontinence

David B. Reuben, MD
Director, Multicampus Program in Geriatric Medicine and Gerontology; Chief, Division of Geriatrics, Archstone Professor of Medicine, David Geffen School of Medicine at University of California Los Angeles, Los Angeles, California
Geriatric Assessment

Emanuel P. Rivers, MD, MPH
Professor and Vice Chairman of Emergency Medicine, Wayne State University; Senior Staff Attending, Critical Care and Emergency Medicine, Henry Ford Hospital, Detroit, Michigan
Approach to the Patient with Shock

Joseph G. Rogers, MD
Professor of Medicine, Senior Vice Chief for Clinical Affairs, Division of Cardiology, Durham, North Carolina
Heart Failure: Pathophysiology and Diagnosis

Jean-Marc Rolain, PharmD, PhD
Professor, Institut Hospitalo-Universitaire Méditerranée-Infection, Aix-Marseille Université, Marseille, France
Bartonella Infections

José R. Romero, MD
Professor of Pediatrics, University of Arkansas for Medical Sciences, Horace C. Cabe Professor of Infectious Diseases; Director, Section of Pediatric Infectious Diseases, Arkansas Children's Hospital, Little Rock, Arkansas
Enteroviruses

Karen Rosene-Montella, MD
Professor and Vice Chair of Medicine, Director of Obstetric Medicine, The Warren Alpert Medical School of Brown University; Senior Vice President, Women's Services and Clinical Integration, Lifespan Health System, Providence, Rhode Island
Common Medical Problems in Pregnancy

Philip J. Rosenthal, MD
Professor, Department of Medicine, University of California San Francisco, San Francisco, California
Malaria

Marc E. Rothenberg, MD, PhD
Director, Division of Allergy and Immunology, Director, Cincinnati Center for Eosinophilic Disorders; Professor of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, Ohio
Eosinophilic Syndromes

James A. Russell, MD
Professor of Medicine, University of British Columbia; Associate Director, Intensive Care Unit, St. Paul's Hospital, Vancouver, British Columbia, Canada
Shock Syndromes Related to Sepsis

Anil K. Rustgi, MD
T. Grier Miller Professor of Medicine and Genetics, Chief of Gastroenterology, American Cancer Society; Professor, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania
Neoplasms of the Esophagus and Stomach

Daniel E. Rusyniak, MD
Professor of Emergency Medicine, Adjunct Professor of Neurology and Pharmacology and Toxicology, Department of Emergency Medicine, Indiana University School of Medicine, Indianapolis, Indiana
Chronic Poisoning: Trace Metals and Others

Robert A. Salata, MD
Professor and Executive Vice Chair, Department of Medicine, Chief, Division of Infectious Diseases and HIV Medicine, Case Western Reserve University, University Hospitals Case Medical Center, Cleveland, Ohio
Brucellosis

Jane E. Salmon, MD
Collette Kean Research Chair, Hospital for Special Surgery, Professor of Medicine, Weill Cornell Medical College, New York, New York
Mechanisms of Immune-Mediated Tissue Injury

Edsel Maurice T. Salvana, MD, DTM&H
Associate Professor of Medicine, Section of Infectious Diseases, Department of Medicine, Philippine General Hospital; Director, Institute of Molecular Biology and Biotechnology, National Institutes of Health, University of the Philippines Manila, Manila, Philippines
Brucellosis

Renato M. Santos, MD
Associate Professor, Cardiology, Wake Forest School of Medicine, Winston-Salem, North Carolina
Vascular Disorders of the Kidney

Michael N. Sawka, PhD
Professor, School of Applied Physiology, Georgia Institute of Technology, Atlanta, Georgia
Disorders Due to Heat and Cold

Paul D. Scanlon, MD
Professor of Medicine, Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, Minnesota
Respiratory Function: Mechanisms and Testing

Carla Scanzello, MD, PhD
Assistant Professor of Medicine, Division of Rheumatology, Perelman School of Medicine at the University of Pennsylvania and Translational Musculoskeletal Research Center, Philadelphia Veterans Affairs Medical Center, Philadelphia, Pennsylvania
Osteoarthritis

Andrew I. Schafer, MD
Professor of Medicine, Director, Richard T. Silver Center for Myeloproliferative Neoplasms, Weill Cornell Medical College, New York, New York
Approach to Medicine, the Patient, and the Medical Profession: Medicine as a Learned and Humane Profession; Approach to the Patient with Bleeding and Thrombosis; Hemorrhagic Disorders: Disseminated Intravascular Coagulation, Liver Failure, and Vitamin K Deficiency; Thrombotic Disorders: Hypercoagulable States

William Schaffner, MD
Professor and Chair, Department of Preventive Medicine, Department of Health Policy; Professor of Medicine (Infectious Diseases), Vanderbilt University School of Medicine, Nashville, Tennessee
Tularemia and Other Francisella Infections

W. Michael Scheld, MD
Bayer-Gerald L. Mandell Professor of Infectious Diseases, Professor of Medicine, Clinical Professor of Neurosurgery, Director, Pfizer Initiative in International Health, University of Virginia Health System, Charlottesville, Virginia
Introduction to Microbial Disease: Host-Pathogen Interactions

Manuel Schiff, MD
Professor, Université Paris 7 Denis Diderot, Sorbonne Paris Cité, Head of Metabolic Unit/Reference Center for Inborn Errors of Metabolism, Robert Debré University Hospital, APHP, Paris, France
Homocystinuria and Hyperhomocysteinemia

Michael L. Schilsky, MD
Associate Professor, Medicine and Surgery, Yale University School of Medicine, New Haven, Connecticut
Wilson Disease

Robert T. Schooley, MD
Professor and Head, Division of Infectious Diseases, Executive Vice Chair for Academic Affairs, Department of Medicine, University of California San Diego, La Jolla, California
Epstein-Barr Virus Infection

David L. Schriger, MD, MPH
Professor, Department of Emergency Medicine, University of California Los Angeles, Los Angeles, California
Approach to the Patient with Abnormal Vital Signs

Steven A. Schroeder, MD
Distinguished Professor of Health and Healthcare and of Medicine, University of California San Francisco, San Francisco, California
Socioeconomic Issues in Medicine

Lynn M. Schuchter, MD
Professor of Medicine, University of Pennsylvania; Chief, Hematology/Oncology Division, Program Leader, Melanoma and Cutaneous Malignancies Program, Abramson Cancer Center, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania
Melanoma and Nonmelanoma Skin Cancers

Sam Schulman, MD, PhD
Professor, Division of Hematology and Thromboembolism, Director of Clinical Thromboembolism Program, Department of Medicine, McMaster University, Hamilton, Ontario, Canada
Antithrombotic Therapy

Lawrence B. Schwartz, MD, PhD
Charles and Evelyn Thomas Professor of Medicine, Internal Medicine, Virginia Commonwealth University, Richmond, Virginia
Systemic Anaphylaxis, Food Allergy, and Insect Sting Allergy

Carlos Seas, MD
Associate Professor of Medicine, Vice Director, Alexander von Humboldt Tropical Medicine Institute, Universidad Peruana Cayetano Heredia; Attending Physician, Department of Infectious, Tropical, and Dermatologic Diseases, National Hospital Cayetano Heredia, Lima, Peru
Cholera and Other Vibrio Infections

Steven A. Seifert, MD
Professor of Emergency Medicine, University of New Mexico School of Medicine, Medical Director, New Mexico Poison and Drug Information Center, University of New Mexico Health Sciences Center, Albuquerque, New Mexico
Envenomation

Julian L. Seifter, MD
Associate Professor of Medicine, Harvard Medical School; Senior Physician, Brigham and Women's Hospital, Boston, Massachusetts
Potassium Disorders; Acid-Base Disorders

Duygu Selcen, MD
Associate Professor of Neurology and Pediatrics, Department of Neurology, Mayo Clinic, Rochester, Minnesota
Muscle Diseases

Clay F. Semenkovich, MD
Herbert S. Gasser Professor and Chief, Division of Endocrinology, Metabolism and Lipid Research, Washington University School of Medicine, St. Louis, Missouri
Disorders of Lipid Metabolism

Carol E. Semrad, MD
Professor of Medicine, The University of Chicago Medicine, GI Section, Chicago, Illinois
Approach to the Patient with Diarrhea and Malabsorption

Harry Shamoon, MD
Professor of Medicine and Associate Dean for Clinical and Translational Research, Albert Einstein College of Medicine; Director, Harold and Muriel Block Institute for Clinical and Translational Research at Einstein and Montefiore, Bronx, New York
Diabetes Mellitus

James C. Shaw, MD
Associate Professor, Department of Medicine, University of Toronto; Head, Division of Dermatology, Department of Medicine, Women's College Hospital, Toronto, Ontario, Canada
Examination of the Skin and an Approach to Diagnosing Skin Diseases

Pamela J. Shaw, DBE, MBBS, MD
Professor of Neurology, University of Sheffield, Consultant Neurologist, Royal Hallamshire Hospital, Sheffield, United Kingdom
Amyotrophic Lateral Sclerosis and Other Motor Neuron Diseases

Robert L. Sheridan, MD
Associate Professor of Surgery, Burn Service Medical Director, Boston Shriners Hospital for Children, Massachusetts General Hospital, Division of Burns, Harvard Medical School, Boston, Massachusetts
Medical Aspects of Injuries and Burns

Stuart Sherman, MD
Professor of Medicine and Radiology, Director of ERCP, Indiana University School of Medicine, Indianapolis, Indiana
Diseases of the Gallbladder and Bile Ducts

Michael E. Shy, MD
Professor of Neurology, Pediatrics, and Physiology, University of Iowa, Iowa City, Iowa
Peripheral Neuropathies

Ellen Sidransky, MD
Chief, Section on Molecular Neurogenetics, Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland
The Lysosomal Storage Diseases

Richard M. Siegel, MD, PhD
Clinical Director, National Institute of Arthritis, Musculoskeletal, and Skin Diseases, National Institutes of Health, Bethesda, Maryland
The Systemic Autoinflammatory Diseases

Robert F. Siliciano, MD, PhD
Professor, The Johns Hopkins University School of Medicine, Howard Hughes Medical Institute, Baltimore, Maryland
Immunopathogenesis of Human Immunodeficiency Virus Infection

Michael S. Simberkoff, MD
Chief of Staff, VA New York Harbor Healthcare System; Professor of Medicine, NYU School of Medicine, New York, New York
Haemophilus and Moraxella Infections

David L. Simel, MD, MHS
Professor of Medicine, Duke University; Chief, Medical Service, Durham Veterans Affairs Medical Center, Durham, North Carolina
Approach to the Patient: History and Physical Examination

Kamaljit Singh, MD
Associate Professor of Medicine, Attending Physician, Infectious Diseases, Rush University Medical Center, Chicago, Illinois
Zoonoses

Karl Skorecki, MD
Annie Chutick Professor in Medicine, Rappaport Faculty of Medicine and Research Institute, Technion–Israel Institute of Technology; Director, Medical and Research Development, Rambam Health Care Campus, Haifa, Israel
Gene and Cell Therapy; Disorders of Sodium and Water Homeostasis

Itzchak Slotki, MD
Associate Professor of Medicine, Hebrew University, Hadassah Medical School; Director, Division of Adult Nephrology, Shaare Zedek Medical Center, Jerusalem, Israel
Disorders of Sodium and Water Homeostasis

Arthur S. Slutsky, MD
Professor of Medicine, Surgery, and Biomedical Engineering, University of Toronto; Vice President (Research), St. Michael's Hospital, Keenan Research Centre, Li Ka Shing Knowledge Institute, Toronto, Ontario, Canada
Acute Respiratory Failure; Mechanical Ventilation

Eric J. Small, MD
Professor of Medicine and Urology, Deputy Director and Director of Clinical Sciences, Helen Diller Family Comprehensive Cancer Center; Chief, Division of Hematology and Oncology, University of California San Francisco School of Medicine, San Francisco, California
Prostate Cancer

Gerald W. Smetana, MD
Professor of Medicine, Harvard Medical School; Division of General Medicine and Primary Care, Beth Israel Deaconess Medical Center, Boston, Massachusetts
Principles of Medical Consultation

Frederick S. Southwick, MD
Professor of Medicine, Division of Infectious Diseases, University of Florida and VF Health, Gainesville, Florida
Nocardiosis

Allen M. Spiegel, MD
Dean, Albert Einstein College of Medicine, Bronx, New York
Principles of Endocrinology; Polyglandular Disorders

Robert F. Spiera, MD
Professor of Clinical Medicine, Weill Cornell Medical College; Director, Scleroderma, Vasculitis, and Myositis Center, The Hospital for Special Surgery, New York, New York
Polymyalgia Rheumatica and Temporal Arteritis

Stanley M. Spinola, MD
Professor and Chair, Department of Microbiology and Immunology, Professor of Medicine, Microbiology and Immunology, and Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, Indiana
Chancroid

David Spriggs, MD
Head, Division of Solid Tumor Oncology, Department of Medicine, Memorial Sloan Kettering Cancer Center; Professor of Medicine, Department of Medicine, Weill Cornell Medical College, New York, New York
Gynecologic Cancers

Paweł Stankiewicz, MD, PhD
Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas
Gene, Genomic, and Chromosomal Disorders

Paul Stark, MD
Professor Emeritus, University of California San Diego; Chief of Cardiothoracic Radiology, VA San Diego Healthcare System, San Diego, California
Imaging in Pulmonary Disease

David P. Steensma, MD
Professor of Medicine, Harvard Medical School, Adult Leukemia Program, Dana-Farber Cancer Institute, Boston, Massachusetts
Myelodysplastic Syndrome

Martin H. Steinberg, MD
Professor of Medicine, Pediatrics, and Pathology and Laboratory Medicine, Boston University School of Medicine; Director, Center of Excellence in Sickle Cell Disease, Boston Medical Center, Boston, Massachusetts
Sickle Cell Disease and Other Hemoglobinopathies

Theodore S. Steiner, MD
Associate Professor, University of British Columbia; Associate Head, Division of Infectious Diseases, Vancouver General Hospital, Vancouver, British Columbia, Canada
Escherichia Coli *Enteric Infections*

David S. Stephens, MD
Stephen W. Schwarzmann Distinguished Professor of Medicine, Emory University School of Medicine and Woodruff Health Sciences Center, Atlanta, Georgia
Neisseria Meningitidis *Infections*

David A. Stevens, MD
Professor of Medicine, Stanford University Medical School; President, Principal Investigator, Infectious Diseases Research Laboratory, California Institute for Medical Research, San Jose and Stanford, California
Systemic Antifungal Agents

James K. Stoller, MD, MS
Chairman, Education Institute, Jean Wall Bennett Professor of Medicine, Cleveland Clinic Lerner College of Medicine; Staff, Respiratory Institute, Cleveland Clinic, Cleveland, Ohio
Respiratory Monitoring in Critical Care

John H. Stone, MD, MPH
Professor of Medicine, Director, Clinical Rheumatology, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts
The Systemic Vasculitides

Richard M. Stone, MD
Professor of Medicine, Harvard Medical School, Clinical Director, Adult Leukemia Program, Dana-Farber Cancer Institute, Boston, Massachusetts
Myelodysplastic Syndrome

Raymond A. Strikas, MD, MPH
Education Team Lead, Immunization Services Division, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia
Immunization

Edwin P. Su, MD
Associate Professor of Clinical Orthopaedics, Orthopaedic Surgery, Weill Cornell University Medical College; Associate Attending Orthopaedic Surgeon, Adult Reconstruction and Joint Replacement, Hospital for Special Surgery, New York, New York
Surgical Treatment of Joint Disease

Roland W. Sutter, MD, MPH&TM
Coordinator, Research, Policy and Product Development, Polio Operations and Research Department, World Health Organization, Geneva, Switzerland
Diphtheria and Other Corynebacteria Infections

Ronald S. Swerdloff, MD
Professor of Medicine, David Geffen School of Medicine at University of California Los Angeles; Chief, Division of Endocrinology, Department of Medicine, Harbor-UCLA Medical Center, Torrance, California
The Testis and Male Hypogonadism, Infertility, and Sexual Dysfunction

Heidi Swygard, MD, MPH
Associate Professor of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina
Approach to the Patient with a Sexually Transmitted Infection

Megan Sykes, MD
Michael J. Friedlander Professor of Medicine, Director, Columbia Center for Translational Immunology, Columbia University Medical Center, New York, New York
Transplantation Immunology

Marian Tanofsky-Kraff, PhD
Associate Professor, Department of Medical and Clinical Psychology, Uniformed Services University of Health Sciences, Bethesda, Maryland
Eating Disorders

Susan M. Tarlo, MBBS
Professor of Medicine, Department of Medicine and Dalla Lana School of Public Health, University of Toronto, Respiratory Physician, University Health Network, Toronto Western Hospital and St. Michael's Hospital, Toronto, Ontario, Canada
Occupational Lung Disease

Victoria M. Taylor, MD, MPH
Professor of Medicine, University of Washington, Fred Hutchinson Cancer Research Center, Seattle, Washington
Cultural Context of Medicine

Ayalew Tefferi, MD
Professor of Medicine, Department of Hematology, Mayo Clinic, Rochester, Minnesota
Polycythemia Vera, Essential Thrombocythemia, and Primary Myelofibrosis

Paul S. Teirstein, MD
Chief of Cardiology, Department of Medicine, Scripps Clinic, La Jolla, California
Interventional and Surgical Treatment of Coronary Artery Disease

Sam R. Telford III, ScD
Professor, Tufts University Cummings School of Veterinary Medicine, North Grafton, Massachusetts
Babesiosis and Other Protozoan Diseases

Rajesh V. Thakker, MD
May Professor of Medicine, University of Oxford; Radcliffe Department of Clinical Medicine, OCDEM, Churchill Hospital, Headington, Oxford, United Kingdom
The Parathyroid Glands, Hypercalcemia, and Hypocalcemia

Antonella Tosti, MD
Professor of Clinical Dermatology, Department of Dermatology and Cutaneous Surgery, University of Miami, Miami, Florida
Diseases of Hair and Nails

Indi Trehan, MD, MPH, DTM&H
Assistant Professor of Pediatrics, Washington University School of Medicine; Attending Physician, St. Louis Children's Hospital, Barnes-Jewish Hospital, St. Louis, Missouri; Visiting Honorary Lecturer in Paediatrics and Child Health, University of Malawi College of Medicine; Consultant Paediatrician, Queen Elizabeth Central Hospital, Blantyre, Malawi
Protein-Energy Malnutrition

Ronald B. Turner, MD
Professor of Pediatrics, University of Virginia School of Medicine, Charlottesville, Virginia
The Common Cold

Thomas S. Uldrick, MD
Staff Clinician, HIV and AIDS Malignancy Branch, National Cancer Institute, Bethesda, Maryland
Hematology and Oncology in Patients with Human Immunodeficiency Virus Infection

Anthony M. Valeri, MD
Professor of Medicine, Columbia University Medical Center; Director, Hemodialysis, Medical Director, Kidney and Pancreas Transplantation, New York-Presbyterian Hospital (CUMC); Director, Hemodialysis, Columbia University Dialysis Center, New York, New York
Treatment of Irreversible Renal Failure

John Varga, MD
John and Nancy Hughes Professor of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois
Systemic Sclerosis (Scleroderma)

Bradley V. Vaughn, MD
Professor of Neurology, Department of Neurology, University of North Carolina, Chapel Hill, North Carolina
Disorders of Sleep

Alan P. Venook, MD
Professor of Medicine, University of California San Francisco, Helen Diller Family Comprehensive Cancer Center, San Francisco, California
Liver and Biliary Tract Cancers

Joseph G. Verbalis, MD
Professor of Medicine, Georgetown University; Chief, Endocrinology and Metabolism, Georgetown University Hospital, Washington, D.C.
Posterior Pituitary

Ronald G. Victor, MD
Professor of Medicine, Burns and Allen Chair in Cardiology Research, Director, Hypertension Center, Associate Director, The Heart Institute, Cedars-Sinai Medical Center, Los Angeles, California
Arterial Hypertension

Angela Vincent, MBBS
Professor of Neuroimmunology, University of Oxford; Honorary Consultant in Immunology, Oxford University Hospital Trust, Oxford, United Kingdom
Disorders of Neuromuscular Transmission

Robert M. Wachter, MD
Professor and Associate Chairman, Department of Medicine, University of California San Francisco, San Francisco, California
Quality of Care and Patient Safety

Edward H. Wagner, MD, MPH
Director Emeritus, MacColl Center for Health Care Innovation, Group Health Research Institute, Seattle, Washington
Comprehensive Chronic Disease Management

Edward E. Walsh, MD
Professor of Medicine, University of Rochester School of Medicine and Dentistry; Head, Infectious Diseases, Rochester General Hospital, Rochester, New York
Respiratory Syncytial Virus

Thomas J. Walsh, MD
Director, Transplantation-Oncology Infectious Diseases Program, Chief, Infectious Diseases Translational Research Laboratory, Professor of Medicine, Pediatrics, and Microbiology and Immunology, Weill Cornell Medical Center; Henry Schueler Foundation Scholar, Sharp Family Foundation Scholar in Pediatric Infectious Diseases, Adjunct Professor of Pathology, The Johns Hopkins University School of Medicine; Adjunct Professor of Medicine, The University of Maryland School of Medicine, Baltimore, Maryland
Aspergillosis

Jeremy D. Walston, MD
Raymond and Anna Lublin Professor of Geriatric Medicine and Gerontology, The Johns Hopkins University School of Medicine, Baltimore, Maryland
Common Clinical Sequelae of Aging

Christina Wang, MD
Professor of Medicine, David Geffen School of Medicine at University of California Los Angeles; Associate Director, UCLA Clinical and Translational Research Institute, Harbor-UCLA Medical Center, Torrance, California
The Testis and Male Hypogonadism, Infertility, and Sexual Dysfunction

Christine Wanke, MD
Professor of Medicine and Public Health, Director, Division of Nutrition and Infection, Associate Chair, Department of Public Health, Tufts University School of Medicine, Boston, Massachusetts
Gastrointestinal Manifestions of HIV and AIDS

Stephen I. Wasserman, MD
Professor of Medicine, University of California San Diego, La Jolla, California
Approach to the Patient with Allergic or Immunologic Disease

Thomas J. Weber, MD
Associate Professor, Medicine/Endocrinology, Duke University, Durham, North Carolina
Approach to the Patient with Metabolic Bone Disease; Osteoporosis

Heiner Wedemeyer, MD
Professor, Department of Gastroenterology, Hepatology, and Endocrinology, Hannover Medical School, Hannover, Germany
Acute Viral Hepatitis

Geoffrey A. Weinberg, MD
Professor of Pediatrics, University of Rochester School of Medicine and Dentistry; Director, Pediatric HIV Program, Golisano Children's Hospital at University of Rochester Medical Center, Rochester, New York
Parainfluenza Viral Disease

David A. Weinstein, MD, MMSc
Professor of Pediatric Endocrinology, Director, Glycogen Storage Disease Program, Division of Pediatric Endocrinology, University of Florida College of Medicine, Gainesville, Florida
Glycogen Storage Diseases

Robert S. Weinstein, MD
Professor of Medicine, Department of Medicine, University of Arkansas for Medical Sciences; Staff Endocrinologist, Department of Medicine, Central Arkansas Veterans Health Care System, Little Rock, Arkansas
Osteomalacia and Rickets

Roger D. Weiss, MD
Professor of Psychiatry, Harvard Medical School, Boston, Massachusetts; Chief, Division of Alcohol and Drug Abuse, McLean Hospital, Belmont, Massachusetts
Drug Abuse and Dependence

Martin Weisse, MD
Chair, Pediatrics, Tripler Army Medical Center, Honolulu, Hawaii; Professor, Pediatrics, Uniformed Services University of the Health Sciences, Bethesda, Maryland
Measles

Jeffrey I. Weitz, MD
Professor of Medicine and Biochemistry, McMaster University; Executive Director, Thrombosis and Atherosclerosis Research Institute, Hamilton, Ontario, Canada
Pulmonary Embolism

Samuel A. Wells, Jr., MD
Medical Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland
Medullary Thyroid Carcinoma

Richard P. Wenzel, MD, MSc
Professor and Former Chairman, Internal Medicine, Virginia Commonwealth University, Richmond, Virginia
Acute Bronchitis and Tracheitis

Victoria P. Werth, MD
Professor of Dermatology and Medicine, Hospital of the University of Pennsylvania and Philadelphia Veterans Administration Medical Center; Chief, Dermatology Division, Philadelphia Veterans Administration Medical Center, Philadelphia, Pennsylvania
Principles of Therapy of Skin Diseases

Sterling G. West, MD, MACP
Professor of Medicine, University of Colorado School of Medicine; Associate Division Head for Clinical and Educational Affairs, University of Colorado Division of Rheumatology, Aurora, Colorado
Systemic Diseases in Which Arthritis Is a Feature

A. Clinton White, Jr., MD
Paul R. Stalnaker Distinguished Professor and Director, Infectious Disease Division, Department of Internal Medicine, University of Texas Medical Branch, Galveston, Texas
Cestodes

Christopher J. White, MD
Professor of Medicine, Ochsner Clinical School, University of Queensland School of Medicine; System Chairman of Cardiovascular Diseases, Ochsner Medical Center, New Orleans, Louisiana
Atherosclerotic Peripheral Arterial Disease; Electrophysiologic Interventional Procedures and Surgery

Perrin C. White, MD
Professor of Pediatrics, The Audry Newman Rapoport Distinguished Chair in Pediatric Endocrinology, University of Texas Southwestern Medical Center, Chief of Endocrinology, Children's Medical Center Dallas, Dallas, Texas
Disorders of Sexual Development

Richard J. Whitley, MD
Distinguished Professor of Pediatrics, Loeb Eminent Scholar Chair in Pediatrics, Professor of Pediatrics, Microbiology, Medicine, and Neurosurgery, The University of Alabama at Birmingham, Birmingham, Alabama
Herpes Simplex Virus Infections

Michael P. Whyte, MD
Professor of Medicine, Pediatrics, and Genetics, Division of Bone and Mineral Diseases, Washington University School of Medicine; Medical-Scientific Director, Center for Metabolic Bone Disease and Molecular Research, Shriners Hospital for Children, St. Louis, Missouri
Osteonecrosis, Osteosclerosis/Hyperostosis, and Other Disorders of Bone

Samuel Wiebe, MD, MSc
Professor of Clinical Neurosciences, University of Calgary; Co-Director, Calgary Epilepsy Program, Alberta Health Services, Foothills Medical Centre, Calgary, Alberta, Canada
The Epilepsies

Jeanine P. Wiener-Kronish, MD
Henry Isaiah Dorr Professor of Research and Teaching in Anaesthesia and Anesthestist-in-Chief, Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts
Overview of Anesthesia

Eelco F.M. Wijdicks, MD, PhD
Professor of Neurology, Division of Critical Care Neurology, Department of Neurology, Mayo Clinic, Rochester, Minnesota
Coma, Vegetative State, and Brain Death

David J. Wilber, MD
George M. Eisenberg Professor of Medicine, Loyola Stritch School of Medicine; Director, Division of Cardiology, Director, Clinical Electrophysiology, Loyola University Medical Center, Maywood, Illinois
Electrophysiologic Interventional Procedures and Surgery

Beverly Winikoff, MD, MPH
President, Gynuity Health Projects; Professor of Clinical Population and Family Health, Mailman School of Public Health, Columbia University, New York, New York
Contraception

Gary P. Wormser, MD
Professor of Medicine and Chief, Division of Infectious Diseases, Department of Medicine, New York Medical College, Valhalla, New York
Lyme Disease

Myron Yanoff, MD
Professor and Chair, Ophthalmology, Drexel University College of Medicine, Philadelphia, Pennsylvania
Diseases of the Visual System

Robert Yarchoan, MD
Branch Chief, HIV and AIDS Malignancy Branch, National Cancer Institute, Bethesda, Maryland
Hematology and Oncology in Patients with Human Immunodeficiency Virus Infection

Neal S. Young, MD
Chief, Hematology Branch, NHLBI and Director, Trans-NIH Center for Human Immunology, Autoimmunity, and Inflammation, National Institutes of Health, Bethesda, Maryland
Parvovirus

William F. Young, Jr., MD, MSc
Professor of Medicine, Mayo Clinic College of Medicine; Chair, Division of Endocrinology, Diabetes, Metabolism, and Nutrition, Mayo Clinic, Rochester, Minnesota
Adrenal Medulla, Catecholamines, and Pheochromocytoma

Alan S.L. Yu, MB, BChir
Harry Statland and Solon Summerfield Professor of Medicine, Director, Division of Nephrology and Hypertension and the Kidney Institute, University of Kansas Medical Center, Kansas City, Kansas
Disorders of Magnesium and Phosphorus

Sherif R. Zaki, MD, PhD
Chief, Infectious Diseases Pathology Branch, Centers for Disease Control and Prevention, Atlanta, Georgia
Leptospirosis

Mark L. Zeidel, MD
Herman L. Blumgart Professor of Medicine, Harvard Medical School; Physician-in-Chief and Chairman, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts
Obstructive Uropathy

Thomas R. Ziegler, MD
Professor, Department of Medicine, Division of Endocrinology, Metabolism, and Lipids, Emory University School of Medicine, Atlanta, Georgia
Malnutrition, Nutritional Assessment, and Nutritional Support in Adult Hospitalized Patients

Peter Zimetbaum, MD
Associate Professor of Medicine, Harvard Medical School; Director of Clinical Cardiology, Beth Israel Deaconess Medical Center, Boston, Massachusetts
Cardiac Arrhythmias with Supraventricular Origin

VIDEO CONTENTS

This icon appears throughout the book to indicate chapters with accompanying video available on ExpertConsult.com. For quick viewing, use your smartphone to scan the QR codes in the front of the book.

1

APPROACH TO MEDICINE, THE PATIENT, AND THE MEDICAL PROFESSION: MEDICINE AS A LEARNED AND HUMANE PROFESSION

LEE GOLDMAN AND ANDREW I. SCHAFER

APPROACH TO MEDICINE

Medicine is a profession that incorporates science and the scientific method with the art of being a physician. The art of tending to the sick is as old as humanity itself. Even in modern times, the art of caring and comforting, guided by millennia of common sense as well as a more recent, systematic approach to medical ethics (Chapter 2), remains the cornerstone of medicine. Without these humanistic qualities, the application of the modern science of medicine is suboptimal, ineffective, or even detrimental.

The caregivers of ancient times and premodern cultures tried a variety of interventions to help the afflicted. Some of their potions contained what are now known to be active ingredients that form the basis for proven medications (Chapter 29). Others (Chapter 39) have persisted into the present era despite a lack of convincing evidence. Modern medicine should not dismiss the possibility that these unproven approaches may be helpful; instead, it should adopt a guiding principle that all interventions, whether traditional or newly developed, can be tested vigorously, with the expectation that any beneficial effects can be explored further to determine their scientific basis.

When compared with its long and generally distinguished history of caring and comforting, the scientific basis of medicine is remarkably recent. Other than an understanding of human anatomy and the later description, albeit widely contested at this time, of the normal physiology of the circulatory system, almost all of modern medicine is based on discoveries made within the past 150 years. Until the late 19th century, the paucity of medical knowledge was perhaps exemplified best by hospitals and hospital care. Although hospitals provided caring that all but well-to-do people might not be able to obtain elsewhere, there is little if any evidence that hospitals improved health outcomes. The term *hospitalism* referred not to expertise in hospital care but rather to the aggregate of iatrogenic afflictions that were induced by the hospital stay itself.

The essential humanistic qualities of caring and comforting can achieve full benefit only if they are coupled with an understanding of how medical science can and should be applied to patients with known or suspected diseases. Without this knowledge, comforting may be inappropriate or misleading, and caring may be ineffective or counterproductive if it inhibits a sick person from obtaining appropriate, scientific medical care. *Goldman-Cecil Medicine* focuses on the discipline of *internal medicine,* from which neurology and dermatology, which are also covered in substantial detail in this text, are relatively recent evolutionary branches. The term *internal medicine,* which is often misunderstood by the lay public, was developed in 19th-century Germany. *Inneren medizin* was to be distinguished from clinical medicine because it emphasized the physiology and chemistry of disease, not just the patterns or progression of clinical manifestations. *Goldman-Cecil Medicine* follows this tradition by showing how pathophysiologic abnormalities cause symptoms and signs and by emphasizing how therapies can modify the underlying pathophysiology and improve the patient's well-being.

Modern medicine has moved rapidly past organ physiology to an increasingly detailed understanding of cellular, subcellular, and genetic mechanisms. For example, the understanding of microbial pathogenesis and many inflammatory diseases (Chapter 256) is now guided by a detailed understanding of the human immune system and its response to foreign antigens (Chapters 45 to 49). Advances in our understanding of the human microbiome raise the possibility that our complex interactions with microbes, which outnumber our cells by a factor of 10, will help explain conditions ranging from inflammatory bowel disease (Chapter 141) to obesity (Chapter 220).[1]

Health, disease, and an individual's interaction with the environment are also substantially determined by genetics. In addition to many conditions that may be determined by a single gene (Chapter 41), medical science increasingly understands the complex interactions that underlie multigenic traits (Chapter 42). The decoding of the human genome holds the promise that personalized health care increasingly can be targeted according to an individual's genetic profile, in terms of screening and presymptomatic disease management, as well as in terms of specific medications and their adjusted dosing schedules.[2]

Although gene therapy has been approved for only one disease, lipoprotein lipase deficiency (Chapter 206), and only in Europe, it has shown promise in other conditions, such as Leber congenital amaurosis (Chapter 423). Cell therapy is now beginning to provide vehicles for the delivery of genes, gene products, and vaccines. It has also opened the way for "regenerative medicine" by facilitating the regeneration of injured or diseased organs and tissues. Such advances and others, such as nanomedicine, have already led to targeted and personalized therapies for a variety of cancers.[3] Knowledge of the structure and physical forms of proteins helps explain abnormalities as diverse as sickle cell anemia (Chapter 163) and prion-related diseases (Chapter 415). Proteomics, which is the normal and abnormal protein expression of genes, also holds extraordinary promise for developing drug targets for more specific and effective therapies.

Concurrent with these advances in fundamental human biology has been a dramatic shift in methods for evaluating the application of scientific advances to the individual patient and to populations. The randomized controlled trial, sometimes with thousands of patients at multiple institutions, has replaced anecdote as the preferred method for measuring the benefits and optimal uses of diagnostic and therapeutic interventions (Chapter 10). As studies progress from those that show biologic effect, to those that elucidate dosing schedules and toxicity, and finally to those that assess true clinical benefit, the metrics of measuring outcome has also improved from subjective impressions of physicians or patients to reliable and valid measures of morbidity, quality of life, functional status, and other patient-oriented outcomes (Chapter 11). These marked improvements in the scientific methodology of clinical investigation have expedited extraordinary changes in clinical practice, such as recanalization therapy for acute myocardial infarction (Chapter 73), and have shown that reliance on intermediate outcomes, such as a reduction in asymptomatic ventricular arrhythmias with certain drugs, may unexpectedly increase rather than decrease mortality. Just as physicians in the 21st century must understand advances in fundamental biology, similar understanding of the fundamentals of clinical study design as it applies to diagnostic and therapeutic interventions is needed. An understanding of human genetics will also help stratify and refine the approach to clinical trials by helping researchers select fewer patients with a more homogeneous disease pattern to study the efficacy of an intervention.

This explosion in medical knowledge has led to increasing specialization and subspecialization, defined initially by organ system and more recently by locus of principal activity (inpatient vs. outpatient), reliance on manual skills (proceduralist vs. nonproceduralist), or participation in research. Nevertheless, it is becoming increasingly clear that the same fundamental molecular and genetic mechanisms are broadly applicable across all organ systems and that the scientific methodologies of randomized trials and careful clinical observation span all aspects of medicine.

The advent of modern approaches to managing data now provides the rationale for the use of health information technology. Computerized health records, oftentimes shared with patients in a portable format, can avoid duplication of tests and assure that care is coordinated among the patient's various health care providers.

Extraordinary advances in the science and practice of medicine, which have continued to accelerate with each recent edition of this textbook, have transformed the global burden of disease.[4] Life expectancies for men and women are increasing, a greater proportion of deaths are occurring among people older than age 70 years, and far fewer children are dying before the age of 5 years. Nevertheless, huge regional disparities remain, and disability from conditions such as substance abuse, mental health disorders, injuries, diabetes, musculoskeletal disease, and chronic respiratory disease have become increasingly important issues for all health systems.

APPROACH TO THE PATIENT

Patients commonly have complaints (symptoms). These symptoms may or may not be accompanied by abnormalities on examination (signs) or on laboratory testing. Conversely, asymptomatic patients may have signs or laboratory abnormalities, and laboratory abnormalities can occur in the absence of symptoms or signs.

Symptoms and signs commonly define *syndromes,* which may be the common final pathway of a wide range of pathophysiologic alterations. The fundamental basis of internal medicine is that diagnosis should elucidate the pathophysiologic explanation for symptoms and signs so that therapy may improve the underlying abnormality, not just attempt to suppress the abnormal symptoms or signs.

When patients seek care from physicians, they may have manifestations or exacerbations of known conditions, or they may have symptoms and signs that suggest malfunction of a particular organ system. Sometimes the pattern of symptoms and signs is highly suggestive or even pathognomonic for a particular disease process. In these situations, in which the physician is focusing on a particular disease, *Goldman-Cecil Medicine* provides scholarly yet practical approaches to the epidemiology, pathobiology, clinical manifestations, diagnosis, treatment, prevention, and prognosis of entities such as acute myocardial infarction (Chapter 73), chronic obstructive lung disease (Chapter 88), obstructive uropathy (Chapter 123), inflammatory bowel disease (Chapter 141), gallstones (Chapter 155), rheumatoid arthritis (Chapter 264), hypothyroidism (Chapter 226), tuberculosis (Chapter 324), and virtually any known medical condition in adults.

Many patients, however, have undiagnosed symptoms, signs, or laboratory abnormalities that cannot be immediately ascribed to a particular disease or cause. Whether the initial manifestation is chest pain (Chapter 51), diarrhea (Chapter 140), neck or back pain (Chapter 400), or a variety of more than 100 common symptoms, signs, or laboratory abnormalities, *Goldman-Cecil Medicine* provides tables, figures, and entire chapters to guide the approach to diagnosis and therapy (see E-Table 1-1 or table on inside back cover). By virtue of this dual approach to known disease as well as to undiagnosed abnormalities, this textbook, similar to the modern practice of medicine, applies directly to patients regardless of their mode of manifestation or degree of previous evaluation.

The patient-physician interaction proceeds through many phases of clinical reasoning and decision making. The interaction begins with an elucidation of complaints or concerns, followed by inquiries or evaluations to address these concerns in increasingly precise ways. The process commonly requires a careful history or physical examination, ordering of diagnostic tests, integration of clinical findings with test results, understanding of the risks and benefits of the possible courses of action, and careful consultation with the patient and family to develop future plans. Physicians can increasingly call on a growing literature of evidence-based medicine to guide the process so that benefit is maximized while respecting individual variations in different patients. Throughout *Goldman-Cecil Medicine,* the best current evidence is highlighted with specific grade A references that can be accessed directly in the electronic version.

The increasing availability of evidence from randomized trials to guide the approach to diagnosis and therapy should not be equated with "cookbook" medicine. Evidence and the guidelines that are derived from it emphasize proven approaches for patients with specific characteristics. Substantial clinical judgment is required to determine whether the evidence and guidelines apply to individual patients and to recognize the occasional exceptions. Even more judgment is required in the many situations in which evidence is absent or inconclusive. Evidence must also be tempered by patients' preferences, although it is a physician's responsibility to emphasize evidence when presenting alternative options to the patient. The adherence of a patient to a specific regimen is likely to be enhanced if the patient also understands the rationale and evidence behind the recommended option.

To care for a patient as an individual, the physician must understand the patient as a person. This fundamental precept of doctoring includes an understanding of the patient's social situation, family issues, financial concerns, and preferences for different types of care and outcomes, ranging from maximum prolongation of life to the relief of pain and suffering (Chapters 2 and 3). If the physician does not appreciate and address these issues, the science of medicine cannot be applied appropriately, and even the most knowledgeable physician will fail to achieve the desired outcomes.

Even as physicians become increasingly aware of new discoveries, patients can obtain their own information from a variety of sources, some of which are of questionable reliability. The increasing use of alternative and complementary therapies (Chapter 39) is an example of patients' frequent dissatisfaction with prescribed medical therapy. Physicians should keep an open mind regarding unproven options but must advise their patients carefully if such options may carry any degree of potential risk, including the risk that they may be relied on to substitute for proven approaches. It is crucial for the physician to have an open dialogue with the patient and family regarding the full range of options that either may consider.

The physician does not exist in a vacuum, but rather as part of a complicated and extensive system of medical care and public health. In premodern times and even today in some developing countries, basic hygiene, clean water, and adequate nutrition have been the most important ways to promote health and reduce disease. In developed countries, adoption of healthy lifestyles, including better diet (Chapter 213) and appropriate exercise (Chapter 16), is the cornerstone to reducing the epidemics of obesity (Chapter 220), coronary disease (Chapter 52), and diabetes (Chapter 229). Public health interventions to provide immunizations (Chapter 18) and to reduce injuries and the use of tobacco (Chapter 32), illicit drugs (Chapter 34), and excess alcohol (Chapter 33) can collectively produce more health benefits than nearly any other imaginable health intervention.

APPROACH TO THE MEDICAL PROFESSION

In a profession, practitioners put the welfare of clients or patients above their own welfare.[5] Professionals have a duty that may be thought of as a contract with society. The American Board of Internal Medicine and the European Federation of Internal Medicine have jointly proposed that medical professionalism should emphasize three fundamental principles: the primacy of patient welfare, patient autonomy, and social justice.[6] As modern medicine brings a plethora of diagnostic and therapeutic options, the interactions of the physician with the patient and society become more complex and potentially fraught with ethical dilemmas (Chapter 2). To help provide a moral compass that is not only grounded in tradition but also adaptable to modern times, the primacy of patient welfare emphasizes the fundamental principle of a profession. The physician's altruism, which begets the patient's trust, must be impervious to the economic, bureaucratic, and political challenges that are faced by the physician and the patient (Chapter 5).

The principle of patient autonomy asserts that physicians make recommendations but patients make the final decisions. The physician is an expert advisor who must inform and empower the patient to base decisions on scientific data and how these data can and should be integrated with a patient's preferences.

The importance of social justice symbolizes that the patient-physician interaction does not exist in a vacuum. The physician has a responsibility to the individual patient and to broader society to promote access and to eliminate disparities in health and health care.

To promote these fundamental principles, a series of professional responsibilities has been suggested (Table 1-1). These specific responsibilities represent practical, daily traits that benefit the physician's own patients and society as a whole. Physicians who use these and other attributes to improve their patients' satisfaction with care are not only promoting professionalism but also reducing their own risk for liability and malpractice.

An interesting new aspect of professionalism is the increasing reliance on team approaches to medical care, as exemplified by physicians whose roles are defined by the location of their practice—historically in the intensive care unit or emergency department and more recently on the inpatient general hospital floor. Quality care requires coordination and effective communication across inpatient and outpatient sites among physicians who themselves now typically work defined hours.[7] This transition from reliance on a single, always available physician to a team, ideally with a designated coordinator, places new challenges on physicians, the medical care system, and the medical profession.

TABLE 1-1 PROFESSIONAL RESPONSIBILITIES

Commitment to:
Professional competence
Honesty with patients
Patient confidentiality
Maintaining appropriate relations with patients
Improving the quality of care
Improving access to care
Just distribution of finite resources
Scientific knowledge
Maintaining trust by managing conflicts of interest
Professional responsibilities

From Brennan T, Blank L, Cohen J, et al. Medical professionalism in the new millennium: a physician charter. *Ann Intern Med.* 2002;1136:243-246.

E-TABLE 1-1 GUIDE TO THE APPROACH TO COMMON SYMPTOMS, SIGNS, AND LABORATORY ABNORMALITIES

	CHAPTER	SPECIFIC TABLES OR FIGURES
SYMPTOMS		
Constitutional		
Fever	280	Tables 280-1 to 280-8
Fatigue	274	E-Table 274-1
Poor appetite	132	Table 132-1
Weight loss	132, 219	Figure 132-4; Tables 132-4, 219-1, 219-2
Obesity	220	Figure 220-1
Snoring, sleep disturbances	100, 405	Table 405-6
Head, Eyes, Ears, Nose, Throat		
Headache	398	Tables 398-1, 398-2
Visual loss, transient	423, 424	Tables 423-2, 424-1
Ear pain	426	Table 426-3
Hearing loss	428	Figure 428-1
Ringing in ears (tinnitus)	428	Figure 428-2
Vertigo	428	Figure 428-3
Nasal congestion, rhinitis, or sneezing	251, 426	Figure 251-1; Table 251-2
Loss of smell or taste	427	Table 427-1
Dry mouth	425	Table 425-7
Sore throat	429	Figure 429-2; Table 429-1
Hoarseness	429	
Cardiopulmonary		
Chest pain	51, 137	Tables 51-2, 137-5, 137-6
Bronchitis	96	
Shortness of breath	51, 83	Figure 83-3
Palpitations	51, 62	Figure 62-1; Tables 51-4, 62-5
Dizziness	51, 62, 428	Figure 62-1; Table 428-1
Syncope	62	Figure 62-1; Tables 62-1, 62-2, 62-4
Cardiac arrest	63	Figures 63-2, 63-3
Cough	83	Figure 83-1; Tables 83-2, 83-3
Hemoptysis	83	Tables 83-6, 83-7
Gastrointestinal		
Nausea and vomiting	132	Figure 132-5; Table 132-5
Dysphagia, odynophagia	132, 138	Table 132-1
Hematemesis	135, 153	Figure 135-3; Table 135-1
Heartburn/dyspepsia	132, 137, 138, 139	Figures 132-6, 138-2; Tables 137-3, 137-4, 139-1
Abdominal pain		
Acute	132, 142	Figures 132-1, 132-2; Tables 132-2, 132-3, 142-1
Chronic	132, 137	Figure 132-3; Tables 132-2, 137-1
Diarrhea	137, 140	Figures 137-1, 140-1 to 140-4
Melena, blood in stool	135	Figures 135-3, 135-4, 135-6; Table 135-4
Constipation	136, 137	Figures 136-3, 136-5, 137-1; Table 136-2
Fecal incontinence	145	Figure 145-5
Anal pain	145	
Genitourinary		
Dysuria	284, 285	Tables 284-3, 284-5, 285-2
Frequency	284	Table 284-3
Incontinence	26	Tables 26-1 to 26-3
Urinary obstruction	123	Tables 123-1 to 123-3
Renal colic	126	Figure 126-1
Vaginal discharge	285	
Menstrual irregularities	236	Figure 236-3; Tables 236-3, 236-4
Female infertility	236	Table 236-5
Hot flushes	240	Table 240-1
Erectile dysfunction	234	Figure 234-10
Male infertility	234	Figures 234-8, 234-9; Table 234-7
Scrotal mass	200	Figure 200-1
Genital ulcers or warts	285	Table 285-1

E-TABLE 1-1 GUIDE TO THE APPROACH TO COMMON SYMPTOMS, SIGNS, AND LABORATORY ABNORMALITIES—cont'd

	CHAPTER	SPECIFIC TABLES OR FIGURES
Musculoskeletal		
Neck or back pain	400	Figures 400-4, 400-5, 400-6; Tables 400-3 to 400-5
Painful joints	256	Figure 256-1; Tables 256-1, 256-3
Extremities		
Swollen feet, ankles, or legs		
Bilateral	51	Figure 51-8
Unilateral	81	Figure 81-2; Table 81-2
Claudication	79	Table 79-3
Acute limb ischemia	79	Figure 79-5; Table 79-1
Neurologic		
Weakness	396, 420, 421, 422	Tables 396-1, 420-2, 421-2, 421-4
Sensory loss	396, 420	Figure 420-1; Tables 420-1, 420-3 to 420-5
Memory loss	402	Figures 402-1, 402-2; Tables 402-1 to 402-6
Abnormal gait	396	Table 396-2
Seizures	403	Tables 403-1 to 403-6
Integumentary		
Abnormal bleeding	171	Table 171-1
Rash	436, 441	Figure 436-1; Tables 436-1 to 436-6, 441-5
Hives	252, 440	Figure 252-2; Tables 252-1, 440-1, 440-2
Abnormal pigmentation	441	Table 441-2
Alopecia and hirsutism	442	Tables 442-1, 442-3
Nail disorders	442	Table 442-4
SIGNS		
Vital Signs		
Fever	280, 281	Figure 281-1; Tables 280-1 to 280-8, 281-2
Hypothermia	8, 109	Table 109-4
Tachycardia/bradycardia	8, 62, 64, 65	Figures 62-2, 62-3; Tables 64-4, 65-2
Hypertension	67	Table 67-5
Hypotension/shock	8, 106	Figures 106-3, 108-1; Tables 106-1, 107-1, 107-2
Altered respiration	8, 86, 104	Tables 86-1, 86-2, 104-2
Head, Eyes, Ears, Nose, Throat		
Eye pain	423	Table 423-3
Red eye	423	Tables 423-4, 423-6
Dilated pupil	424	Figure 424-4
Nystagmus	424	Table 424-5
Papilledema	424	Table 424-2
Strabismus	424	Figure 424-6
Jaundice	147	Figure 147-2; Tables 147-1 to 147-3
Otitis	426	Table 426-3
Sinusitis	251, 426	Tables 251-3, 426-1, 426-2
Oral ulcers and discolorations	425	Tables 425-1 to 425-4
Salivary gland enlargement	425	Table 425-6
Neck		
Neck mass	190	Figure 190-3
Lymphadenopathy	168	Tables 168-1 to 168-6
Thyroid nodule	226	Figure 226-4
Thyromegaly/goiter	226	Figures 226-1, 226-3
Breast		
Breast mass	198	
Lungs		
Wheezes	83	Table 83-4
Cardiac		
Heart murmur or extra sounds	51	Figure 51-6; Tables 51-7, 51-8
Jugular venous distention	51	Table 51-6
Carotid pulse abnormalities	51	Figure 51-5

E-TABLE 1-1 GUIDE TO THE APPROACH TO COMMON SYMPTOMS, SIGNS, AND LABORATORY ABNORMALITIES—cont'd

	CHAPTER	SPECIFIC TABLES OR FIGURES
Abdomen		
Hepatomegaly	146	Figure 146-5
Splenomegaly	168	Tables 168-7, 168-9
Acute abdomen	142, 143	Figure 143-1; Table 142-1
Abdominal swelling/ascites	142, 153	Table 153-3
Rectal bleeding/positive stool	135, 193	Figures 135-3, 135-4, 135-6; Table 135-4
Hemorrhoids	145	Table 145-1
Musculoskeletal/Extremities		
Arthritis	256	Figure 256-1
Edema	51	Figure 51-8
Cyanosis	51	
Clubbing	51	Figure 51-10
Neurologic		
Delirium	28	Figure 28-1; Tables 28-1, 28-2
Psychiatric disturbances	397	Tables 397-1 to 397-4, 397-6 to 397-8, 397-10, 397-11, 397-13, 397-14
Coma	404	Tables 404-1 to 404-4
Stroke	407, 408	Figure 407-1; Tables 407-2, 407-3, 407-5, 407-6, 408-5, 408-6
Movement disorders	409, 410	Tables 409-4, 410-1 to 410-9
Neuropathy	420	Figure 420-1; Tables 420-1 to 420-5, E-Table 420-1
Skin and Nails		
Suspicious mole	203	Table 203-1
Nail diseases	442	Table 442-4
COMMON LABORATORY ABNORMALITIES		
Hematology/Urinalysis		
Anemia	158	Tables 158-2 to 158-6
Polycythemia	166	Figure 166-2; Table 166-4
Leukocytosis	167	Figure 167-4; Table 167-1
Lymphocytosis	167	Table 167-3
Monocytosis	167	Table 167-2
Eosinophilia	170	Figure 170-2; Table 170-1
Neutropenia	167	Figure 167-7; Tables 167-4 and 167-5
With fever	281	Figure 281-1
Thrombocytosis	166	Figure 166-6; Table 166-6
Thrombocytopenia	172	Figure 172-1; Tables 172-1, 172-3
Prolonged PT or PTT	171	Figure 171-4
Urinalysis	114, 120	Tables 114-2, 120-6
Chemistries		
Abnormal liver enzymes	147	Figures 147-2 to 147-4
Elevated BUN/creatinine		
Acute	120	Figure 120-1; Tables 120-1 to 120-5
Chronic	130	Table 130-1
Hyperglycemia	229	Tables 229-1, 229-2
Hypoglycemia	230	Tables 230-1, 230-2
Electrolyte abnormalities	116, 117	Figure 116-4; Tables 116-6, 116-7, 117-2, 117-3
Acid-base disturbances	118	Figures 118-1, 118-2; Tables 118-1 to 118-6
Hypercalcemia	245	Figure 245-3; Tables 245-2 to 245-4
Hypocalcemia	245	Figure 245-4; Table 245-6
Hypo- and hyperphosphatemia	119	Tables 119-2, 119-3
Magnesium deficiency	119	Table 119-1
Elevated P_{CO_2}	86	Figure 86-2
Chest Radiograph/ECG		
Solitary pulmonary nodule	191	Figure 191-2
Pleural effusion	99	Tables 99-4 to 99-6
ECG abnormalities	54	Tables 54-2 to 54-5

BUN = blood urea nitrogen; ECG = electrocardiogram; PT = prothrombin time; PTT = partial thromboplastin time.

The changing medical care environment is placing increasing emphasis on standards, outcomes, and accountability. As purchasers of insurance become more cognizant of value rather than just cost (Chapter 12), outcomes ranging from rates of screening mammography (Chapter 198) to mortality rates with coronary artery bypass graft surgery (Chapter 74) become metrics by which rational choices can be made. Clinical guidelines and critical pathways derived from randomized controlled trials and evidence-based medicine can potentially lead to more cost-effective care and better outcomes.

These major changes in many Western health care systems bring with them many major risks and concerns. If the concept of limited choice among physicians and health care providers is based on objective measures of quality and outcome, channeling of patients to better providers is one reasonable definition of better selection and enlightened competition. If the limiting of options is based overwhelmingly on cost rather than measures of quality, outcomes, and patient satisfaction, it is likely that the historical relationship between the patient and the truly professional physician will be fundamentally compromised.

Another risk is that the same genetic information that could lead to more effective, personalized medicine will be used against the very people whom it is supposed to benefit—by creating a stigma, raising health insurance costs, or even making someone uninsurable. The ethical approach to medicine (Chapter 2), genetics (Chapter 40), and genetic counseling provides means to protect against this adverse effect of scientific progress.

In this new environment, the physician often has a dual responsibility: to the health care system as an expert who helps create standards, measures of outcome, clinical guidelines, and mechanisms to ensure high-quality, cost-effective care; and to individual patients who entrust their well-being to that physician to promote their best interests within the reasonable limits of the system. A health insurance system that emphasizes cost-effective care, that gives physicians and health care providers responsibility for the health of a population and the resources required to achieve these goals, that must exist in a competitive environment in which patients can choose alternatives if they are not satisfied with their care, and that places increasing emphasis on health education and prevention can have many positive effects. In this environment, however, physicians must beware of overt and subtle pressures that could entice them to underserve patients and abrogate their professional responsibilities by putting personal financial reward ahead of their patients' welfare. The physician's responsibility to represent the patient's best interests and avoid financial conflicts by doing too little in the newer systems of capitated care provides different specific challenges but an analogous moral dilemma to the historical American system in which the physician could be rewarded financially for doing too much.

In the current health care environment, all physicians and trainees must redouble their commitment to professionalism. At the same time, the challenge to the individual physician to retain and expand the scientific knowledge base and process the vast array of new information is daunting. In this spirit of a profession based on science and caring, *Goldman-Cecil Medicine* seeks to be a comprehensive approach to modern internal medicine.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Martin R, Miquel S, Langella P, et al. The role of metagenomics in understanding the human microbiome in health and disease. *Virulence*. 2014;5:413-423.
2. Ganesh SK, Arnett DK, Assimes TL, et al. Genetics and genomics for the prevention and treatment of cardiovascular disease: update: a scientific statement from the American Heart Association. *Circulation*. 2013;128:2813-2851.
3. Paoletti C, Hayes DF. Molecular testing in breast cancer. *Annu Rev Med*. 2014;65:95-110.
4. The Global Burden of Disease Study 2010. *Lancet*. 2012-2013;380:2053-2260.
5. Walton M, Kerridge I. Do no harm: is it time to rethink the Hippocratic Oath? *Med Educ*. 2014;48:17-27.
6. Snyder L, for the American College of Physicians Ethics, Professionalism, and Human Rights Committee. American College of Physicians ethics manual: sixth edition. *Ann Intern Med*. 2012; 156:73-104.
7. O'Malley AS, Reschovsky JD. Referral and consultation communication between primary care and specialist physicians: finding common ground. *Arch Intern Med*. 2011;171:56-65.

2

BIOETHICS IN THE PRACTICE OF MEDICINE

EZEKIEL J. EMANUEL

It commonly is argued that modern advances in medical technology, antibiotics, dialysis, transplantation, and intensive care units have created the bioethical dilemmas that confront physicians in the 21st century. In reality, however, concerns about ethical issues are as old as the practice of medicine itself. The Hippocratic Oath, composed sometime around 400 BC, attests to the need of ancient Greek physicians for advice on how to address the many bioethical dilemmas that they confronted. The Oath addresses issues of confidentiality, abortion, euthanasia, sexual relations between physician and patient, divided loyalties, and, at least implicitly, charity care and executions. Other Hippocratic works address issues such as termination of treatments to dying patients and telling the truth. Whether we agree with the advice dispensed or not, the important point is that many bioethical issues are not created by technology but instead are inherent in medical practice. Technology may make these issues more common and may change the context in which they arise, but many, if not most, bioethical issues that regularly confront physicians are timeless and inherent in the practice of medicine.

Many physicians have been educated that four main principles can be invoked to address bioethical dilemmas: autonomy, nonmaleficence, beneficence, and justice. Autonomy is the idea that people should have the right and freedom to choose, pursue, and revise their own life plans. Nonmaleficence is the idea that people should not be harmed or injured knowingly; this principle is encapsulated in the frequently repeated phrase that a physician has an obligation to "first do no harm"—*primum non nocere*. This phrase is not found either in the Hippocratic Oath or in other Hippocratic writing; the only related, but not identical, Hippocratic phrase is "at least, do not harm." Whereas nonmaleficence is about avoiding harm, beneficence is about the positive actions that the physician should undertake to promote the well-being of his or her patients. In clinical practice, this obligation usually arises from the implicit and explicit commitments and promises surrounding the physician-patient relationship. Finally, there is the principle of justice as the fair distribution of benefits and burdens.

Although helpful in providing an initial framework, these principles have limited value because they are broad and open to diverse and conflicting interpretations. In addition, as is clear with the principle of justice, they frequently are underdeveloped. In any difficult case, the principles are likely to conflict. Conflicting ethical principles are precisely why there are bioethical dilemmas. The principles themselves do not offer guidance on how they should be balanced or specified to resolve the dilemma. These principles, which are focused on the individual physician-patient context, are not particularly helpful when the bioethical issues are institutional and systemic, such as allocating scarce vaccines or organs for transplantation or balancing the risks and benefits of mammograms for women younger than 50 years. Finally, these four principles are not comprehensive. Other fundamental ethical principles and values, such as communal solidarity, duties to future generations, trust, and professional integrity, are important in bioethics but not encapsulated except by deformation in these four principles.

There is no formula or small set of ethical principles that mechanically or magically gives answers to bioethical dilemmas. Instead, medical practitioners should follow an orderly analytic process. First, practitioners need to obtain the facts relevant to the situation. Second, they must delineate the basic bioethical issue. Third, it is important to identify all the crucial principles and values that relate to the case and how they might conflict. Fourth, because many ethical dilemmas have been analyzed previously and subjected frequently to empirical study, practitioners should examine the relevant literature, whether it is commentaries or studies in medical journals, legal cases, or books. With these analyses, the particular dilemma should be reexamined; this process might lead to reformulation of the issue and identification of new values or new understandings of existing values. Fifth, with this information, it is important to distinguish clearly unethical practices from a range of ethically permissible actions. Finally, it is important not only to come to some resolution of the case but also to state clearly the reasons behind the decisions, that is, the interpretation of the principles used and how values were balanced. Although unanimity and consensus may be desirable ideals, reasonable people frequently disagree about how to resolve ethical dilemmas without being unethical or malevolent.

A multitude of bioethical dilemmas arise in medical practice, including issues of genetics, reproductive choices, and termination of care. In clinical practice, the most common issues revolve around informed consent, termination of life-sustaining treatments, euthanasia and physician-assisted suicide, and conflicts of interest.

PHYSICIAN-PATIENT RELATIONSHIP: INFORMED CONSENT

History

It commonly is thought that the requirement for informed consent is a relatively recent phenomenon. Suggestions about the need for a patient's informed consent can be found as far back as Plato, however. The first

recorded legal case involving informed consent is the 1767 English case of *Slater v. Baker and Stapleton,* in which two surgeons refractured a patient's leg after it had healed improperly. The patient claimed they had not obtained consent. The court ruled:

> *[I]t appears from the evidence of the surgeon that it was improper to disunite the callous without consent; this is the usage and law of surgeons: then it was ignorance and unskillfulness in that very particular, to do contrary to the rule of the profession, what no surgeon ought to have done.*

Although there may be some skepticism about the extent of the information disclosed or the precise nature of the consent obtained, the notable fact is that an 18th-century court declared that obtaining prior consent of the patient is not only the usual practice but also the ethical and legal obligation of surgeons. Failure to obtain consent is incompetent and inexcusable. In contemporary times, the 1957 case of *Salgo v. Leland Stanford Junior University Board of Trustees* constitutes a landmark by stating that physicians have a positive legal obligation to disclose information about risks, benefits, and alternatives to patients; this decision popularized the term *informed consent.*

Definition and Justification

Informed consent is a person's autonomous authorization of a physician to undertake diagnostic or therapeutic interventions for himself or herself. In this view, the patient understands that he or she is taking responsibility for the decision while empowering someone else, the physician, to implement it. However, agreement to a course of medical treatment does not necessarily qualify as informed consent.

There are four fundamental requirements for valid informed consent: mental capacity, disclosure, understanding, and voluntariness. Informed consent assumes that people have the mental capacity to make decisions; disease, development, or medications can compromise patients' mental capacity to provide informed consent. Adults are presumed to have the legal competence to make medical decisions, and whether an adult is incompetent to make medical decisions is a legal determination. Practically, physicians usually decide whether patients are competent on the basis of whether patients can understand the information disclosed, appreciate its significance for their own situation, and use logical and consistent thought processes in decision making. Incompetence in medical decision making does not mean a person is incompetent in all types of decision making and vice versa. Crucial information relevant to the decision must be disclosed, usually by the physician, to the patient. The patient should understand the information and its implications for his or her interests and life goals. Finally, the patient must make a voluntary decision (i.e., one without coercion or manipulation by the physician). It is a mistake to view informed consent as an event, such as the signing of a form. Informed consent is viewed more accurately as a process that evolves during the course of diagnosis and treatment.

Typically, the patient's autonomy is the value invoked to justify informed consent. Other values, such as bodily integrity and beneficence, have also been cited, especially in early legal rulings.

Empirical Data

Fairly extensive research has been done on informed consent. In general, studies show that in clinical situations, physicians frequently do not communicate all relevant information for informed decision making. In a study of audiotapes from 1057 outpatient encounters, physicians mentioned alternatives in only 11.3% of cases, provided pros and cons of interventions in only 7.8% of situations, and assessed the patient's understanding of the information in only 1.5% of decisions. The more complex the medical decisions, the more likely it was that the elements of informed consent would be fulfilled. Importantly, data suggest that disclosure is better in research settings, both in the informed consent documents and in the discussions. For instance, in recorded interactions between researchers and prospective participants, the major elements of research, such as that the treatment was investigational and the risks and benefits of treatment, were disclosed in more than 80% of interactions. Greater disclosure in the research setting may be the consequence of requiring a written informed consent document that has been reviewed by an independent committee, such as an institutional review board or a research ethics committee. Some have suggested that for common medical interventions, such as elective surgery, standardized informed consent documents should include the risks and benefits as quantified in randomized controlled trials, relevant data on the surgeon, the institution's clinical outcomes for the procedure, and a list of acceptable alternatives.[1]

Patients frequently fail to recall crucial information disclosed, although they usually think they have sufficient information for decision making. Whether patients fail to recall key information because they are overwhelmed by the information or because they do not find much of it salient to their decision is unclear. The issue is what patients understand at the point of decision making, not what they recall later.

Studies aimed at improving informed consent in the clinical setting suggest that interactive media, such as videos and interactive computer software, can improve understanding by patients. Conversely, data on shared decision making show that interactive media do not improve participants' understanding, whereas more personal interaction, whether as an additional telephone call by a research nurse or as an additional face-to-face meeting, does enhance understanding.[2]

One of the most important results of empirical research on informed consent is the gap between information and decision making. Many studies show that most patients want information, but far fewer prefer decision-making authority. One study showed that most patients wanted information, but only about one third desired decision-making authority, and patients' decision-making preferences were not correlated with their information-seeking preferences. Several investigators found that patients' preference for decision-making authority increases with higher educational levels and declines with advancing age. Most important, the more serious the illness, the more likely patients are to prefer that physicians make the decisions. Several studies suggest that patients who have less of a desire to make their own decisions generally are more satisfied with how the decisions were made.

Practical Considerations

Implementing informed consent raises concerns about the extent of information to be disclosed and exceptions to the general requirement. A major area of ethical and legal disagreement has been what information to disclose and how to disclose it. As a practical matter, physicians should disclose at least six fundamental elements of information to patients: (1) diagnosis and prognosis; (2) nature of the proposed intervention; (3) alternative interventions, including no treatment; (4) risks associated with each alternative; (5) benefits of each alternative; and (6) likely outcomes of these alternatives (Table 2-1). Because risk is usually the key worry of physicians, it generally is recommended that physicians disclose (1) the nature of the risks, (2) their magnitude, (3) the probability that each risk will occur, and (4) when the consequence might occur.[3] Increasingly, these disclosures should include data both from clinical trials as well as the actual data from the institution and physician performing the test and treatments. Some argue that minor risks need not be disclosed. In general, all serious risks, such as death, paralysis, stroke, infections, or chronic pain, even if rare, should be disclosed, as should common risks.

The central problem is that the physician should provide this detailed information within reasonable time constraints and yet not overwhelm patients with complex information in technical language. The historical constraint of office time is no longer tenable. Interactive electronic media, which patients can view at home on their own time, can facilitate the transfer of information outside of the physician's office. Different states have adopted two contrasting legal standards defining how much information should be disclosed. The *physician* or *customary* standard, adapted from malpractice law, states that the physician should disclose information "which a reasonable medical practitioner would make under the same or similar circumstances." Conversely, the *reasonable person* or *lay-oriented* standard states that physicians should disclose all information that a "reasonable person in the patient's circumstances would find material to" the medical decision. The physician standard is factual and can be determined empirically, but the patient-oriented standard, which is meant to engage physicians with patients, is hypothetical. Currently, each standard is used by about half the states.

TABLE 2-1 FUNDAMENTAL ELEMENTS FOR DISCLOSURE TO PATIENTS

Diagnosis and prognosis
Nature of proposed intervention
Reasonable alternative interventions
Risks associated with each alternative intervention
Benefits associated with each alternative intervention
Probable outcomes of each alternative intervention

There are exceptions to the requirements of informed consent. In emergency situations, consent can be assumed because patients' interests concentrate on survival and retaining maximal mental and physical functioning; as a result, reasonable persons would want treatment. In some circumstances, physicians may believe the process of informed consent could pose a serious psychological threat. In rare cases, the "therapeutic privilege" promoting a patient's well-being trumps autonomy, but physicians should be wary of invoking this exception too readily.

If patients are deemed incompetent, family members—beginning with spouse, children, parents, siblings, then more distant relatives—usually are selected as surrogates or proxies, although there may be concerns about conflicting interests or knowledge of the patient's wishes. In the relatively rare circumstance in which a patient formally designated a proxy, that person has decision-making authority.

The *substituted judgment* standard states that the proxy should choose what the patient would choose if he or she were competent. The *best interests* standard states that the proxy should choose what is best for the patient. Frequently, it is not clear how the patient would have decided because the situation was not discussed with the patient and he or she left no living will. Similarly, what is best for a patient is controversial because there are usually tradeoffs between quality of life and survival. These problems are exacerbated because a proxy's predictions about a patient's quality of life are poor; proxies tend to underestimate patients' functional status and satisfaction. Similarly, proxy predictions are inaccurate regarding life-sustaining preferences when the patient is mentally incapacitated. Families tend to agree with patients about two thirds of the time in deciding whether to provide life-sustaining treatments if the patient became demented, when chance alone would generate agreement in 50% of the cases. Such confusion about how to decide for incapacitated patients can create conflicts among family members or between the family and medical providers. In such circumstances, an ethics consultation may be helpful.

TERMINATION OF MEDICAL INTERVENTIONS

History

Since the start of medicine, it has been viewed as ethical to withhold medical treatments from the terminally ill and "let nature take its course," while keeping the patient as comfortable as possible.[4] Hippocrates argued that physicians should "refuse to treat those [patients] who are overmastered by their disease." In the 19th century, prominent American physicians advocated withholding of cathartic and emetic "treatments" from the terminally ill and using ether to ease pain at the end of life. John Collins Warren, who wrote *Etherization: with Surgical Remarks* in 1848, included a chapter on using ether to ease the pain of a cancer patient's death. The editors of *The Lancet,* in 1900, argued that physicians should intervene to ease the pain of death and that they did not have an obligation to prolong a clearly terminal life. The contemporary debate on terminating care began in 1976 with the *Quinlan* case, in which the New Jersey Supreme Court ruled that patients had a right to refuse life-sustaining interventions on the basis of a right of privacy and that the family could exercise the right for a patient in a persistent vegetative state.

Definition and Justification

It generally is agreed that all patients have a right to refuse medical interventions. Ethically, this right is based on the patient's autonomy and is implied by the doctrine of informed consent. Legally, state courts have cited the right to privacy, right to bodily integrity, or common law to justify the right to refuse medical treatment. In the 1990 *Cruzan* case and in the subsequent physician-assisted suicide cases, the U.S. Supreme Court affirmed that there is a "constitutionally protected right to refuse lifesaving hydration and nutrition." The Court stated that "[A] liberty interest [based on the 14th Amendment] in refusing unwanted medical treatment may be inferred from our prior decisions." All patients have a constitutional and an ethical right to refuse medical interventions. These rulings were the basis of the consistent state and federal court rulings to permit the husband to terminate artificial nutrition and hydration in the *Schiavo* case.

Empirical Data

Data show that termination of medical treatments is now the norm, and the trend has been to stop medical interventions more frequently based on the preferences of patients and their surrogate decision makers.[5] More than 85% of Americans die without cardiopulmonary resuscitation, and more than 90% of decedents in intensive care units do not receive cardiopulmonary resuscitation. Of decedents in intensive care units, more than 85% die after the withholding or withdrawal of medical treatments, with an average of 2.6 interventions being withheld or withdrawn per decedent.

Despite extensive public support for use of advance care directives and the passage of the Patient Self-Determination Act mandating that health care institutions inform patients of their right to complete such documents, less than 30% of Americans have completed one.[6] Even among severely or terminally ill patients, less than 50% have an advance directive in their medical record. Data suggest that more than 40% of patients required active decision-making about terminating medical treatments in their final days, but more than 70% lacked decision-making capacity, thereby emphasizing the importance of advance directives. Efforts to improve completion of advance care directives have generated mixed results. In La Crosse County, Wisconsin, for example, after health care organizations in the county added an "Advance Directive" section to their electronic medical records, 90% of decedents had some type of advance directive. Unfortunately, even successful pilot efforts like La Crosse County's have not been adopted or easily scaled. A persistent problem has been that even when patients complete advance care directives, the documents frequently are not available, physicians do not know they exist, or they tend to be too general or vague to guide decisions. The increasing use of electronic health records should make it possible for advance directives to be available whenever and wherever the patient presents to a health care provider. Although electronic health records will help in making existing advance directives available, they will not solve the problem of actually having a conversation between the physician and the patient about advance care planning. Starting that conversation still seems to be a persistent barrier.

Just as proxies are poor at predicting patients' wishes, data show that physicians are probably even worse at determining patients' preferences for life-sustaining treatments. In many cases, life-sustaining treatments are continued even when patients or their proxies desire them to be stopped. Conversely, many physicians discontinue or never begin interventions unilaterally without the knowledge or consent of patients or their surrogate decision makers. These discrepancies emphasize the importance of engaging patients early in their care about treatment preferences.

Practical Considerations

There are many practical considerations in enacting this right (Table 2-2). First, patients have a right to refuse any and all medical interventions, from blood transfusions and antibiotics to respirators, artificial hydration, and nutrition. Although initiation of cardiopulmonary resuscitation was the focus of the early court cases, this issue is viewed best as addressing just one of the many medical interventions that can be stopped or withheld.

The question of what medical interventions can be terminated—or not started—is a recurrent topic of debate among physicians and other health care providers. The fact is that any treatment prescribed by a physician and administered by a health care provider can be stopped. The issue is not whether the treatment is ordinary, extraordinary, or heroic, or whether it is high technology or low technology. Treatments that can be stopped include not only ventilators, artificial nutrition, and hydration but also dialysis, pacemakers, ventricular assist devices, antibiotics, and any medications.

Second, there is no ethical or legal difference between withholding an intervention and withdrawing it. If a respirator or other treatment is started because physicians are uncertain whether a patient would have wanted it, they always can stop it later when information clarifies the patient's wishes. Although physicians and nurses might find stopping a treatment to be more difficult psychologically, withdrawal is ethically and legally permitted—and required—when it is consonant with the patient's wishes.

Third, competent patients have the exclusive right to decide about terminating their own care.[7] If there is a conflict between a competent patient and his or her family, the patient's wishes are to be followed. It is the patient's right to refuse treatment, not the family's right. For incompetent patients, the situation is more complex; if the patients left clear indications of their wishes, whether as explicit oral statements or as written advance care directives, these wishes should be followed. Physicians should not be overly concerned about the precise form patients use to express their wishes; because patients have a constitutional right to refuse treatment, the real concern is whether the wishes are clear and relevant to the situation. If an incompetent patient did not leave explicit indications of his or her wishes or designate a proxy decision maker, the physician should identify a surrogate decision maker and rely on the decision maker's wishes while being cognizant of the potential problems noted. There is a potential problem in terminating life-sustaining care to patients who are permanently incompetent but still conscious. Some state courts have restricted what treatments a proxy decision maker can terminate,

TABLE 2-2 PRACTICAL CONSIDERATIONS IN TERMINATION OF MEDICAL TREATMENTS

PRACTICAL QUESTION	ANSWER
Is there a legal right to refuse medical interventions?	Yes. The U.S. Supreme Court declared that competent people have a constitutionally protected right to refuse unwanted medical treatments based on the 14th Amendment.
What interventions can be legally and ethically terminated?	Any and all interventions (including respirators, antibiotics, pacemakers, ventricular assist devices, intravenous or enteral nutrition and hydration) can be legally and ethically terminated.
Is there a difference between withholding life-sustaining interventions and withdrawing them?	No. The consensus is that there is no important legal or ethical difference between withholding and withdrawing medical interventions. Stopping a treatment once begun is just as ethical as never having started it.
Whose view about terminating life-sustaining interventions prevails if there is a conflict between the patient and family?	The views of a competent adult patient prevail. It is the patient's body and life.
Who decides about terminating life-sustaining interventions if the patient is incompetent?	If the patient appointed a proxy or surrogate decision maker when competent, that person is legally empowered to make decisions about terminating care. If no proxy was appointed, there is a legally designated hierarchy, usually (1) spouse, (2) adult children, (3) parents, (4) siblings, and (5) available relatives.
Are advance care directives legally enforceable?	Yes. As a clear expression of the patient's wishes, they are a constitutionally protected method for patients to exercise their right to refuse medical treatments. In almost all states, clear and explicit oral statements are legally and ethically sufficient for decisions about withholding or withdrawing medical interventions.

TABLE 2-3 DEFINITIONS OF ASSISTED SUICIDE AND EUTHANASIA

TERM	DEFINITION
Voluntary active euthanasia	Intentional administration of medications or other interventions to cause the patient's death with the patient's informed consent
Involuntary active euthanasia	Intentional administration of medications or other interventions to cause the patient's death when the patient was competent to consent but did not consent (e.g., the patient may not have been asked)
Nonvoluntary active euthanasia	Intentional administration of medications or other interventions to cause the patient's death when the patient was incompetent and was mentally incapable of consenting (e.g., the patient might have been in a coma)
Passive euthanasia	Withholding or withdrawal of life-sustaining medical treatments from a patient to let him or her die (termination of life-sustaining treatments)—a poor term that should not be used
Indirect euthanasia	Administration of narcotics or other medications to relieve pain with the incidental consequence of causing sufficient respiratory depression to result in the patient's death
Physician-assisted suicide	A physician provides prescription medications or other interventions to a patient with the understanding that the patient can use them to commit suicide

thereby requiring the incompetent patient to have given very specific instructions about the particular treatments he or she does not want to receive and the conditions under which care should be withheld or withdrawn. This requirement severely limits the authority and power of proxy decision makers in these cases.

Fourth, the right to refuse medical treatment does not translate into a right to demand any treatment, especially treatments that have no pathophysiologic rationale, have already failed, or are known to be harmful. Futility has become a justification to permit physicians unilaterally to withhold or withdraw treatments despite the family's requests for treatment. Some states, such as Texas, have enacted futility laws, which prescribe procedures by which physicians can invoke futility either to transfer a patient or to terminate interventions. However, the principle of futility is not easy to implement in medical practice. Initially, some commentators advocated that an intervention was futile when the probability of success was 1% or lower. Although this threshold seems to be based on empirical data, it is a covert value judgment. Because the declaration of futility is meant to justify unilateral determinations by physicians, it generally has been viewed as an inappropriate assertion that undermines physician-patient communication and violates the principle of shared decision making. Similar to the distinction between ordinary and extraordinary, futility is viewed increasingly as more obfuscating than clarifying, and it is being invoked much less often.

ASSISTED SUICIDE AND EUTHANASIA

History

Since Hippocrates, euthanasia and physician-assisted suicide have been controversial issues. In 1905, a bill was introduced into the Ohio legislature to legalize euthanasia; it was defeated. In the mid-1930s, similar bills were introduced and defeated in the British Parliament and the Nebraska legislature. As of January 2014, physician-assisted suicide is legal in Oregon and Washington State, based on statewide public referenda, and in Vermont, based on legislation passed in May 2013. Both euthanasia and physician-assisted suicide are legal in the Netherlands, Belgium, and Luxembourg, and physician-assisted suicide is legal in Switzerland. The Montana Supreme Court did not recognize a constitutional right to physician-assisted suicide, but it ruled that the law permitting the termination of life-sustaining treatment protected physicians from prosecution if they helped hasten the death of a consenting, rational, terminally ill patient.

Definition and Justification

The terms *euthanasia* and *physician-assisted suicide*[8] require careful definition (Table 2-3). So-called passive and indirect euthanasia are misnomers and are not instances of euthanasia, and both are deemed ethical and legal.

There are four arguments against permitting euthanasia and physician-assisted suicide. First, Kant and Mill thought that autonomy did not permit the voluntary ending of the conditions necessary for autonomy, and as a result, both philosophers were against voluntary enslavement and suicide. Consequently, the exercise of autonomy cannot include the ending of life because that would mean ending the possibility of exercising autonomy. Second, many dying patients may have pain and suffering because they are not receiving appropriate care, and it is possible that adequate care would relieve much pain and suffering (Chapter 3). Although a few patients still may experience uncontrolled pain and suffering despite optimal end-of-life care, it is unwise to use the condition of these few patients as a justification to permit euthanasia or physician-assisted suicide for any dying patient. Third, there is a clear ethical distinction between intentional ending of a life and termination of life-sustaining treatments. The actual acts are different—injecting a life-ending medication, such as a muscle relaxant, or providing a prescription for one is not the same as removing or refraining from introducing an invasive medical intervention. Finally, adverse consequences of permitting euthanasia and physician-assisted suicide must be considered. There are disturbing reports of involuntary euthanasia in the Netherlands and Belgium, and many worry about coercion of expensive or burdensome patients to accept euthanasia or physician-assisted suicide. Permitting euthanasia and physician-assisted suicide is likely to lead to further intrusions of lawyers, courts, and legislatures into the physician-patient relationship.

There are four parallel arguments for permitting euthanasia and physician-assisted suicide. First, it is argued that autonomy justifies euthanasia and physician-assisted suicide. To respect autonomy requires permitting individuals to decide when it is better to end their lives by euthanasia or physician-assisted suicide. Second, beneficence—furthering the well-being of individuals—supports permitting euthanasia and physician-assisted suicide. In some cases, living can create more pain and suffering than death; ending a painful life relieves more suffering and produces more good. Just the reassurance of having the option of euthanasia or physician-assisted suicide, even if people do not use it, can provide "psychological insurance" and be

beneficial to people. Third, euthanasia and physician-assisted suicide are no different from termination of life-sustaining treatments that are recognized as ethically justified. In both cases, the patient consents to die; in both cases, the physician intends to end the patient's life and takes some action to end the patient's life; and in both cases, the final result is the same: the patient's death. With no difference in the patient's consent, the physician's intention, or the final result, there can be no difference in the ethical justification. Fourth, the supposed slippery slope that would result from permitting euthanasia and physician-assisted suicide is not likely. The idea that permitting euthanasia and physician-assisted suicide would undermine the physician-patient relationship or lead to forced euthanasia is completely speculative and not borne out by the available data.

In its 1997 decisions, the U.S. Supreme Court stated that there is no constitutional right to euthanasia and physician-assisted suicide but that there also is no constitutional prohibition against states legalizing these interventions. Consequently, the legalization of physician-assisted suicide in Oregon, Vermont, and Washington State was constitutional.

Empirical Data

Attitudes and practices related to euthanasia and physician-assisted suicide have been studied extensively. First, surveys consistently indicate that between 50 and 80% of the American and British public support legalizing euthanasia and physician-assisted suicide for terminally ill patients who are suffering intractable pain.[9] However, public support declines significantly for euthanasia and physician-assisted suicide in other circumstances, such as for psychological reasons.[10] Physicians tend to be much less supportive of euthanasia and physician-assisted suicide, with oncologists, palliative care physicians, and geriatricians among the least supportive. Among American and British physicians, the majority opposes legalizing either practice. Second, approximately 25% of American physicians have received requests for euthanasia or physician-assisted suicide, including about 50% of oncologists. Third, multiple studies indicate that less than 5% of American physicians have performed euthanasia or physician-assisted suicide. Among oncologists, 4% have performed euthanasia and 11% have performed physician-assisted suicide during their careers. Fourth, in many cases, the safeguards are violated. One study found that in 54% of euthanasia cases, it was the family who made the request; in 39% of euthanasia and 19% of physician-assisted suicide cases, the patient was depressed; in only half of the cases was the request repeated.

In the Netherlands and Belgium, where euthanasia and physician-assisted suicide are legal, less than 2% of all deaths are by these measures, with 0.4 to 1.8% of all deaths as the result of euthanasia without the patient's consent.[11] Since the practice of assisted suicide was legalized in Oregon in 1997, a cumulative 0.2% of all deaths are by physician-assisted suicide.

Counterintuitively, data indicate that it is not pain that primarily motivates requests for euthanasia or physician-assisted suicide but rather psychological distress, especially depression and hopelessness. Interviews with physicians and with patients with amyotrophic lateral sclerosis, cancer, or infection with human immunodeficiency virus show that pain is not associated with interest in euthanasia or physician-assisted suicide; instead, depression and hopelessness are the strongest predictors of interest. Studies of patients in Australia and the Netherlands confirm the importance of depression in motivating requests for euthanasia. The desire to avoid dependence and loss of dignity are key motivations.

Finally, data from the Netherlands and the United States suggest that there are significant problems in performing euthanasia and physician-assisted suicide. Dutch researchers reported that physician-assisted suicide causes complications in 7% of cases, and in 15% of cases, the patients did not die, awoke from coma, or vomited up the medication. Ultimately, in nearly 20% of physician-assisted suicide cases, the physician ended up injecting the patient with life-ending medication, converting physician-assisted suicide to euthanasia. These data raise serious questions about how to address complications of physician-assisted suicide when euthanasia is illegal or unacceptable.

Practical Considerations

There is widespread agreement that if euthanasia and physician-assisted suicide are used, they should be considered only after all reasonable attempts at physical and psychological palliation have failed. A series of safeguards have been developed and embodied in the Oregon and the Dutch procedures, as follows: (1) the patient must be competent and must request euthanasia or physician-assisted suicide repeatedly and voluntarily; (2) the patient must have pain or other suffering that cannot be relieved by optimal palliative interventions; (3) there should be a waiting period to ensure that the patient's desire for euthanasia or physician-assisted suicide is stable and sincere; and (4) the physician should obtain a second opinion from an independent physician. Oregon and Washington State require patients to be terminally ill, whereas the Netherlands, Belgium, and Switzerland have no such requirement. Although there have been some prosecutions in the United States, there have been no convictions—except for Dr. Kevorkian—when physicians and others have participated in euthanasia and physician-assisted suicide.

FINANCIAL CONFLICTS OF INTEREST

History

Worrying about how payment and fees affect medical decisions is not new. In 1899, a physician reported that more than 60% of surgeons in Chicago were willing to provide a 50% commission to physicians for referring cases. He subsequently argued that in some cases, this fee splitting led to unnecessary surgical procedures. A 1912 study by the American Medical Association confirmed that fee splitting was a common practice. Selling patent medicines and patenting surgical instruments were other forms of financial conflicts of interest thought to discredit physicians a century ago. In the 1990s, the ethics of capitation for physician services and pharmaceutical prescriptions and payments by pharmaceutical and biotechnology companies to clinical researchers and practitioners raised the issue of financial conflicts of interest.

Definition and Justification

It commonly is argued that physicians have certain primary interests: (1) to promote the well-being of their patients, (2) to advance biomedical research, (3) to educate future physicians, and, more controversially, (4) to promote public health (Table 2-4). Physicians also have other, secondary interests, such as earning income, raising a family, contributing to the profession, and pursuing avocational interests, such as hobbies. These secondary interests are not evil; typically, they are legitimate, even admirable. A conflict of interest occurs when one of these secondary interests compromises pursuit of a primary interest, especially the patient's well-being.

Conflicts of interest are problematic because they can or appear to compromise the integrity of physicians' judgment, compromising the patient's well-being or research integrity. Conflict of interest can induce a physician to do something—perform a procedure, fail to order a test, or distort data—that would not be in a patient's best interest. These conflicts can undermine the trust of patients and the public, not only in an individual physician but also in the entire medical profession. Even the appearance of conflicts of interest can be damaging because it is difficult for patients and the public "to determine what motives have influenced a professional decision." The focus is on financial conflicts of interest, not because they are worse than other types of conflicts, but rather because they are more pervasive and more easily identified and regulated compared with other conflicts. Since ancient times, the ethical norm on conflicts has been clear: the physician's primary obligation is to patients' well-being, and a physician's personal financial well-being should not compromise this duty.

Empirical Data

Financial conflicts are not rare but are frequently under-reported.[12] The increased use of medical services and escalating health care spending, sometimes without clear benefit to patients, have been linked, at least statistically, to ownership of imaging facilities and referral to specialty hospitals owned by physicians. In Florida, it was estimated that nearly 40% of physicians were involved as owners of freestanding facilities to which they referred patients. In one study, 4 to 4.5 times more imaging examinations were ordered by self-referring physicians than by physicians who referred patients to radiologists. Similarly, patients referred to joint-venture physical therapy facilities have an average of 16 visits compared with 11 at non–joint-venture facilities. A recent study of urologists found that those who had integrated radiation

TABLE 2-4 PRIMARY INTERESTS OF PHYSICIANS

Promotion of the health and well-being of their patients
Advancement of biomedical knowledge through research
Education of future physicians and health care providers
Promotion of the public health

facilities into their practices increased their use of the radiation by 2.5 times compared with urologists who did not have financial relationships with radiation facilities.[13] There are no comparable data on the influence of capitation on physicians' judgment.

Similarly, multiple studies have shown that interaction with pharmaceutical representatives can lead to prescribing of new drugs, nonrational prescribing, and decreased use of generic drugs by physicians. Industry funding for continuing medical education payment for travel to educational symposia increases prescribing of the sponsor's drug.

Regarding researcher conflicts of interest, the available data suggest that corporate funding does not compromise the design and methodology of clinical research; in fact, commercially funded research may be methodologically more rigorous than government- or foundation-supported research. Conversely, data suggest that financial interests do distort researchers' interpretation of data. The most important impact of financial interests, however, appears to be on dissemination of research studies. Growing evidence suggests the suppression or selective publication of data unfavorable to corporate sponsors but the repeated publication of favorable results.

Practical Considerations

First, financial conflicts of interest are inherent in any profession when the professional earns income from rendering a service. Second, conflicts come in many different forms, from legitimate payment for services rendered to investments in medical laboratories and facilities, drug company dinners and payment for attendance at meetings, payment for enrolling patients in clinical research trials, and consultation with companies.

Third, in considering how to manage conflicts, it is important to note that people are poor judges of their own potential conflicts. Individuals often cannot distinguish the various influences that guide their judgments, do not think of themselves as bad, and do not imagine that payment shapes their judgments. Physicians tend to be defensive about charges of conflicts of interest. In addition, conflicts tend to act insidiously, subtly changing practice patterns so that they then become what appear to be justifiable norms.

Fourth, rules—whether laws, regulations, or professional standards—to regulate conflicts of interest are based on two considerations: (1) the likelihood that payment or other secondary interests would create a conflict and (2) the magnitude of the potential harm if there is compromised judgment. Rules tend to be of three types: (1) disclosure of conflicts, (2) management of conflicts, and (3) outright prohibition. Federal law bans certain types of self-referral of physicians in the Medicare program. The American Medical Association and the Pharmaceutical Research and Manufacturers of America have established joint rules that permit physicians to accept gifts of minimal value but "refuse substantial gifts from drug companies, such as the costs of travel, lodging, or other personal expenses . . . for attending conferences or meetings." Additionally, the Physician Payment Sunshine Act, which was passed in 2010 as part of the Affordable Care Act and went into effect in August 2013, requires that drug and device manufacturers report all payments and transfers of value given to physicians to the Centers for Medicare and Medicaid Services so that information can be published on a searchable public website.

Fifth, there is much emphasis on disclosure of conflicts, with the implicit idea being that sunshine is the best disinfectant. Disclosure may be useful in publications, but it is unclear whether this is a suitable safeguard in the clinical setting. Disclosure just may make patients worry more. Patients may have no context in which to place the disclosure or to evaluate the physician's clinical recommendation, and patients may have few other options in selecting a physician or getting care, especially in an acute situation. Furthermore, self-disclosure often is incomplete, even when required.

Finally, some conflicts can be avoided by a physician's own action. Physicians can refuse to engage in personal investments in medical facilities or to accept gifts from pharmaceutical companies at relatively little personal cost. In other circumstances, the conflicts may be institutionalized, and minimizing them can occur only by changing the way organizations structure reimbursement incentives. Capitation encourages physicians to limit medical services, and its potentially adverse effects are likely to be managed by institutional rules rather than by personal decisions.

FUTURE DIRECTIONS

In the near future, as genetics moves from the research to the clinical setting, practicing physicians are likely to encounter issues surrounding genetic testing, counseling, and treatment. The use of genetic tests without the extensive counseling so common in research studies would alter the nature of the bioethical issues. Because these tests have serious implications for the patient and others, scrupulous attention to informed consent must occur. The bioethical issues raised by genetic tests for somatic cell changes, such as tests that occur commonly in cancer diagnosis and risk stratification, are no different from the issues raised with the use of any laboratory or radiographic test.

In some cases, ethics consultation services may be of assistance in resolving bioethical dilemmas, although current data suggest that consultation services are used mainly for problems that arise in individual cases and are not used for more institutional or policy problems.

Grade A Reference

A1. Stacey D, Légaré F, Bennett CL, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev*. 2014;1:CD001431.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Krumholz HM. Informed consent to promote patient-centered care. *JAMA*. 2010;303:1190-1191.
2. Flory J, Emanuel E. Interventions to improve research participants' understanding in informed consent for research: a systematic review. *JAMA*. 2004;292:1593-1601.
3. Caring Connections. http://www.caringinfo.org/. Accessed February 9, 2015.
4. Education in Palliative and End-of-life Care. http://www.epec.net. Accessed February 9, 2015.
5. Silveira MJ, Kim SY, Langa KM. Advance directives and outcomes of surrogate decision making before death. *N Engl J Med*. 2010;362:1211-1218.
6. Rao JK, Anderson LA, Lin FC, et al. Completion of advance directives among U.S. consumers. *Am J Prev Med*. 2014;46:65-70.
7. Loggers ET, Starks H, Shannon-Dudley M, et al. Implementing a Death with Dignity program at a comprehensive cancer center. *N Engl J Med*. 2013;368:1417-1424.
8. Boudreau JD, Somerville MA, Biller-Andorno N. Clinical decisions. Physician-assisted suicide. *N Engl J Med*. 2013;368:1450-1452.
9. Seale C. Legalisation of euthanasia or physician-assisted suicide: survey of doctors' attitudes. *Palliat Med*. 2009;23:205-212.
10. Emanuel EJ. Euthanasia and physician-assisted suicide: a review of the empirical data from the United States. *Arch Intern Med*. 2002;162:142-152.
11. van der Heide A, Onwuteaka-Philipsen BD, Rurup ML, et al. End-of-life practices in the Netherlands under the Euthanasia Act. *N Engl J Med*. 2007;356:1957-1965.
12. Okike K, Kocher MS, Wei EX, et al. Accuracy of conflict-of-interest disclosures reported by physicians. *N Engl J Med*. 2009;361:1466-1474.
13. Mitchell JM. Urologists' use of intensity-modulated radiation therapy for prostate cancer. *N Engl J Med*. 2013;369:1629-1637.

3

CARE OF DYING PATIENTS AND THEIR FAMILIES

ROBERT ARNOLD

By 2030, 20% of the U.S. population will be older than 65 years, and people older than 85 years constitute the fastest growing segment of the population. Owing to successes in public health and medicine, many of these people will live the last years of their lives with chronic medical conditions such as cirrhosis, end-stage kidney disease, heart failure, and dementia. Even human immunodeficiency virus (HIV) and many cancers, once considered terminal, have turned into chronic diseases.

The burden associated with these illnesses and their treatments is high. Chronically ill patients report multiple physical and psychological symptoms that lower their quality of life. The economic pressures associated with medical care adversely affect patients' socioeconomic status and cause family stress, especially among caregivers, who spend 20 or more hours a week helping their loved ones.

Palliative care, which was developed to decrease the burden associated with chronic illness, emphasizes patient- and family-centered care that optimizes quality of life by anticipating, preventing, and treating suffering. Palliative care throughout the continuum of illness addresses physical, intellectual, emotional, social, and spiritual needs while facilitating the patient's autonomy, access to information, and choice. Palliative care and services, which are coordinated by an interdisciplinary team, are available concurrently with or independent of curative or life-prolonging care. Palliative and nonpalliative health care providers should collaborate and communicate about care needs while focusing on peace and dignity throughout the course of illness, during the dying process, and after death.

Five points deserve special emphasis. First, palliative care can be delivered at any time during the course of an illness and is often provided concomitantly with disease-focused, life-prolonging therapy. Waiting until a patient is dying to provide palliative care is a serious error. For example, most elderly patients with chronic incurable illnesses, who might benefit from palliative care, are in the last 10 years of their lives but do not consider themselves to be dying. If palliative care is to have an impact on patients' lives, it should be provided earlier in a patient's illness, in tandem with other treatments. Second, prediction is an inexact science. Although many cancers have a predictable trajectory in the last 3 to 6 months of life, for most illnesses, doctors rarely can accurately predict whether a patient is in the last 6 months

of life[1] (E-Fig. 3-1). Third, palliative care primarily focuses on the illness's burden rather than treating the illness itself. Because these burdens can be physical, psychological, spiritual, or social, good palliative care requires a multidisciplinary approach. Fourth, palliative care takes the family unit as the central focus of care. Treatment plans must be developed for both the patient and the family. Fifth, palliative care recognizes that medical treatments are not uniformly successful and that patients die. At some point in a patient's illness, the treatments may cause more burden than benefit. Palliative care recognizes this reality and starts with a discussion of the patient's goals and the development of an individualized treatment plan.

Many people confuse palliative care with hospice—an understandable confusion because hospices epitomize the palliative care philosophy. The two, however, are different. In the United States, hospice provides palliative care, primarily at home, for patients who have a life expectancy of 6 months or less and who are willing to forgo life-prolonging treatments. However, the requirement that patients must have a life expectancy of 6 months or less limits hospice's availability, as does the requirement that patients give up expensive and potentially life-prolonging treatments. Moreover, because doctors and patients often are unwilling to cease these treatments until very late in the disease course, so are most patients.

Palliative care is both a subspeciality and a domain of good internal medicine.[2] Given the need for palliative care, every clinician must be able to provide basic palliative care, and subspecialties such as oncology need special expertise.

PALLIATIVE CARE DOMAINS

Palliative care is a holistic discipline with physical, psychological, spiritual, existential, social, and ethical domains. When caring for patients with chronic life-limiting illness, good palliative care requires that the following questions be addressed:

Is the Patient Physically Comfortable?

Across many chronic conditions, patients have a large number of inadequately treated physical symptoms (Table 3-1). The reasons are multifactorial and range from inadequate physician education, to societal beliefs regarding the inevitability of suffering in chronic illness, to public concerns regarding opioids, to the lack of evidence-based treatments in noncancer patients.

The first step to improve symptom management is a thorough assessment. Standardized instruments such as the Brief Pain Inventory (Fig. 3-1) measure both the patient's symptoms and the effect of those symptoms on the patient's life. Use of standardized instruments assures that physicians will identify overlooked or underreported symptoms and, as a result, will enhance the satisfaction of both the patient and family.

The evidence for the treatment of end-stage symptoms continues to improve. The use of nonsteroidal anti-inflammatory agents and opioids can result in effective pain management in more than 75% of patients with cancer. Advances such as intrathecal pumps and neurolytic blocks are helpful in the remaining 25% (Chapter 30). The use of oxygen is not helpful for refractory dyspnea except when hypoxia has been documented, whereas use of medications for depression often can be helpful (Chapter 397).

Is the Patient Psychologically Suffering?

Patients may be physically comfortable but still suffering. Psychological symptoms and syndromes such as depression, delirium, and anxiety are common in patients with life-limiting or chronic illnesses. It may be difficult to determine whether increased morbidity and mortality are caused by the physical effects of the illness or by the psychological effects of depression and anxiety on energy, appetite, or sleep. Screening questions focusing on mood (e.g., "Have you felt down, depressed, and hopeless most of the time for the past 2 weeks?") and anhedonism (e.g., "Have you found that little brings you pleasure or joy in the past 2 weeks?") have been shown to help in diagnosing depression in this population. Increasing data show that treatment of depression in chronic illness is possible and improves both morbidity and mortality.

For patients and families facing mortality, existential and spiritual concerns are common. Progressive illness often raises questions of love, legacy, loss, and meaning. A physician's role is not to answer these questions or to provide reassurance, but rather to understand concerns of the patient and family, how they are coping, and what resources might help. Spirituality often is a source of comfort, and physicians can ascertain a patient's beliefs using a brief instrument such as the FICA Spiritual Assessment Tool (Table 3-2). A single screening question such as "Are you at peace?" may identify patients who are in spiritual distress and facilitate referrals to chaplains.

Is the Family Suffering?

Families, defined broadly as those individuals who care most for the patient, are an important source of support for most patients. Families provide informal caregiving, often at the expense of their own physical, economic, and psychological health. Good palliative care requires an understanding of how the family is coping and a search for ways to provide family members with the social or clinical resources they need to improve their well-being. Comprehensive and individually targeted interventions can reduce caregivers' burdens, although the absolute benefits are relatively small.

Because patients in palliative care often die, the palliative care team must address bereavement and postdeath family suffering. Good communication and informational brochures in an intensive care unit can decrease family members' adverse psychological outcomes after death. A letter of condolence or a follow-up phone call to the next of kin after a patient's death is respectful and offers the opportunity to clarify questions about the patient's care. Some family members suffer from complicated grief—a recently described syndrome associated with separation and traumatic distress, with symptoms persisting for more than 6 months. Primary care physicians, who have ongoing relationships with the loved one, and hospices, which provide bereavement services for a year after the patient's death, have the opportunity to assess whether the grief symptoms persist or worsen.

Is the Patient's Care Consistent with the Patient's Goals?

The sine qua non for palliative care is ensuring that the treatment plan is consistent with the patient's values. In one European cohort of elderly patients, most preferred longevity over quality of life, and half wanted resuscitation if necessary.[3] However, a large proportion of elderly, seriously ill patients are not focused on living as long as possible. Instead, they want to maintain a sense of control, relieve their symptoms, improve their quality of life, avoid being a burden on their families, and have a closer relationship with their loved ones.

Ensuring that treatment is consistent with a patient's goals requires good communication skills (Table 3-3). The approaches to giving bad news, discussing goals of care, and talking about forgoing life-sustaining treatment have similar structures (Table 3-4). First, the patient needs to understand the basic facts about the diagnosis, possible treatments, and prognosis. The communication skill that helps physicians communicate information is *Ask-Tell-Ask*—exploring what the patient knows or wants to know, then explaining or answering questions, and then providing an opportunity for the patient to ask more. In the hospital, where discontinuity of care is common and misunderstandings frequent, it is important to determine what the patient knows before providing information so as to keep everyone well coordinated. When giving bad news, knowing what the patient knows allows the physician to anticipate the patient's reaction. Finally, information must be titrated based on the patient's preferences. Although most patients want to hear everything about their disease, a minority do not. There is no foolproof way to ascertain what any patient wants to know other than by asking.

When giving patients information, it is important to give small pieces of information, not use jargon, and check the patient's understanding.[4] Giving information is like dosing a medication: one gives information, checks understanding, and then gives more information based on what the patient has heard.

After ensuring that the doctor and the patient have a shared understanding of the medical facts, the physician should engage in an open-ended conversation about the patient's goals as the disease progresses. This strategy requires that the patient be asked about both hopes and fears. One might ask: "What makes life worth living for you?" "If your time is limited, what are the things that are most important to achieve?" "What are your biggest fears or concerns?" "What would you consider to be a fate worse than death?" The clinician can use an understanding of these goals to make recommendations about which treatments to provide and which treatments would not be helpful. As a result, early palliative care can improve quality of life, mood, and even survival.

Physicians find talking about prognosis particularly difficult for two reasons: first, it is hard to foretell the future accurately; and second, they fear this information will "take away patients' hope." Thus, they often avoid talking to patients about these issues unless specifically asked. Although some patients do not want to hear prognostic information, for many patients, this

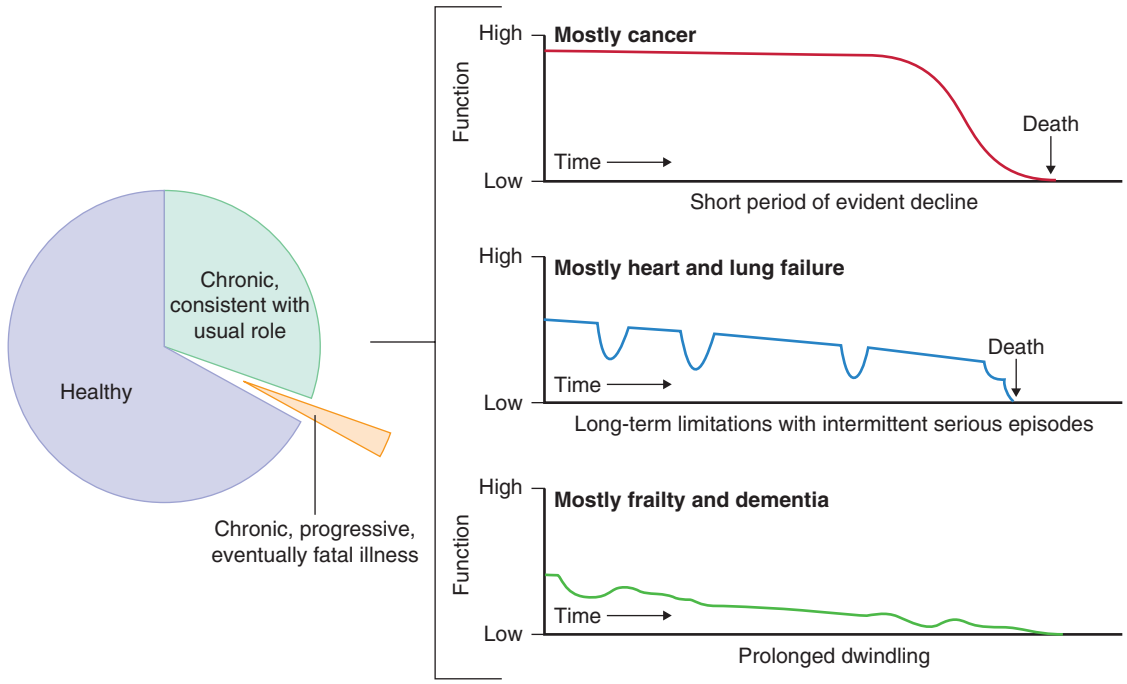

E-FIGURE 3-1. Different disease trajectories for different illnesses. Permission obtained from RAND Corporation © Lynn J. Perspectives on care at the close of life. Serving patients who may die soon and their families: the role of hospice and other services. *JAMA*. 2001;285:925-932.

TABLE 3-1 APPROACHES TO THE MANAGEMENT OF PHYSICAL AND PSYCHOLOGICAL SYMPTOMS

SYMPTOM	ASSESSMENT	TREATMENT
Pain	How severe is the symptom (as assessed with the use of validated instruments) and how does it interfere with the patient's life? What is the etiology of the pain? Is the pain assumed to be neuropathic or somatic? What has the patient used in the past (calculate previous days' equal analgesic dose)?	Prescribe medications to be administered on a standing or regular basis if pain is frequent. For mild pain: use acetaminophen or a nonsteroidal anti-inflammatory agent (see Table 30-3). For moderate pain: titrate short-acting opioids (see Table 30-4). For severe pain: rapidly titrate short-acting opioids until pain is relieved or intolerable side effects develop; start long-acting opiates once pain is controlled. Rescue doses: prescribe immediate-release opioids—10% of the 24-hour total opiate every hour (orally) or every 30 minutes (parenterally) as needed. Concomitant analgesics (e.g., corticosteroids, anticonvulsants, tricyclic antidepressants, and bisphosphonates) should be used when applicable (particularly for neuropathic pain). Consider alternative medicine and interventional treatments for pain.
Constipation	Is the patient taking opioids? Does the patient have a fecal impaction?	Prescribe laxatives for all patients on opiates. If ineffective, add drugs from multiple classes (e.g., stimulant, osmotic laxatives, and enemas). Prescribe methylnaltrexone if still constipated.
Shortness of breath	Ask the patient to assess the severity of the shortness of breath. Does the symptom have reversible causes?	Prescribe oxygen to treat hypoxia-induced dyspnea, but *not* if the patient is not hypoxic. Opioids relieve breathlessness without measurable reductions in respiratory rate or oxygen saturation; effective doses are often lower than those used to treat pain. Aerosolized opiates do not work. Fans or cool air may work through a branch of the trigeminal nerve. Consider anxiolytics (e.g., low-dose benzodiazepines) and use reassurance, relaxation, distraction, and massage therapy.
Fatigue	Is the patient too tired to do activities of daily living? Is the fatigue secondary to depression? Is a disease process causing the symptom or is it secondary to reversible causes?	Provide cognitive education about conserving energy use. Treat underlying conditions appropriately.
Nausea	Which mechanism is causing the symptom (e.g., stimulation of the chemoreceptor trigger zone, gastric stimulation, delayed gastric emptying or "squashed stomach" syndrome, bowel obstruction, intracranial processes, or vestibular vertigo)? Is the patient constipated?	Prescribe an agent directed at the underlying cause (Chapter 132). If persistent, give antiemetic around the clock. Multiple agents directed at various receptors or mechanisms may be required.
Anorexia and cachexia	Is a disease process causing the symptom, or is it secondary to other symptoms (e.g., nausea and constipation) that can be treated? Is the patient troubled by the symptom or is the family worried about what not eating means?	A nutritionist may help find foods that are more appetizing (Chapter 213). Provide counseling about the prognostic implications of anorexia (Chapter 219).
Delirium	Is the confusion acute, over hours to days? Does consciousness wax and wane? Are there behavioral disturbances, marked by a reduced clarity in the patient's awareness of the environment, e.g., a problem of attention? Does the patient have disorganized thinking? Does the patient have an altered level of consciousness—either agitated or drowsy? Is there a reversible reason for the delirium? **D: D**rugs (opioids, anticholinergics, sedatives, benzodiazepines, steroids, chemotherapies and immunotherapies, some antibiotics) **E: E**yes and **E**ars (poor vision and hearing, isolation) **L: L**ow-flow states (hypoxia, myocardial infarction, congestive heart failure, chronic obstructive pulmonary disease, shock) **I: I**nfections **R: R**etention (urine/stool), **R**estraints **I: I**ntracranial (central nervous system metastases, seizures, subdural, cerebrovascular accident, hypertensive encephalopathy) **U: U**nderhydration, **U**ndernutrition, **U**ndersleep **M: M**etabolic disorders (sodium, glucose, thyroid, hepatic, deficiencies of vitamin B_{12}, folate, niacin, and thiamine) and toxic (lead, manganese, mercury, alcohol)	Identify underlying causes and manage symptoms (Chapter 28). Recommend behavioral therapies, including avoidance of excess stimulation, frequent reorientation, and reassurance. Ensure presence of family caregivers and explain delirium to them. Prescribe haloperidol, risperidone, or olanzapine.
Depression	Have you felt down, depressed, or hopeless most of the time during the past 2 weeks? Have you found that little brings you pleasure or joy during the past 2 weeks? (Somatic symptoms are not reliable indicators of depression in this population.)	Recommend supportive psychotherapy, cognitive approaches, behavioral techniques, pharmacologic therapies (see Table 397-5), or a combination of these interventions; prescribe psychostimulants for rapid treatment of symptoms (within days) or selective serotonin reuptake inhibitors, which may require 3 to 4 weeks to take effect; tricyclic antidepressants are relatively contraindicated because of their side effects.
Anxiety (applicable also for family members)	Does the patient exhibit restlessness, agitation, insomnia, hyperventilation, tachycardia, or excessive worry? Is the patient depressed? Is there a spiritual or existential concern underlying the anxiety?	Recommend supportive counseling and consider prescribing benzodiazepines.
Spiritual distress	Are you at peace?	Inquire about spiritual support.

Modified from Morrison RS, Meier DE. Palliative care. *N Engl J Med.* 2004;350:2582-2590.

information helps them plan their lives. Patients who are told that their disease is generally terminal are more likely to spend a longer period of time in hospice and to avoid aggressive technology at the end of life, without adverse psychological consequences. Furthermore, their families usually have fewer postdeath adverse psychological outcomes.

Given that one cannot guess how much information to provide, a physician can start these conversations by asking, "Are you the kind of person who wants to hear about what might happen in the future with your illness or would you rather take it day by day?" If the patient requests the latter, the physician can follow up by asking if there is someone else with whom he or she can talk about the prognosis. Second, before giving prognostic information, it is useful to inquire about the patient's concerns in order to provide information in the most useful manner. Finally, it is appropriate when discussing prognostic information to acknowledge uncertainty: "The course of this cancer can be quite unpredictable, and physicians don't have a crystal ball. I think you should be aware of the possibility that your health may

STUDY ID# _____ HOSPITAL ID# _____

DO NOT WRITE ABOVE THIS LINE

Brief Pain Inventory (Short Form)

Date: ___/___/___ Time: ____

Name: _____ (Last) _____ (First) _____ (Middle Initial)

1. Throughout our lives, most of us have had pain from time to time (such as minor headaches, sprains, and toothaches). Have you had pain other than these everyday kinds of pain today?

1. Yes 2. No

2. On the diagram, shade in the areas where you feel pain. Put an X on the area that hurts the most.

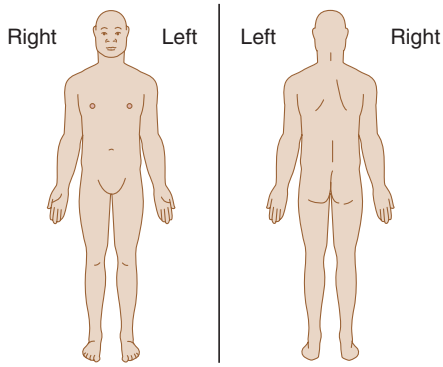

3. Please rate your pain by circling the one number that best describes your pain at its worst in the last 24 hours.

0	1	2	3	4	5	6	7	8	9	10
No pain										Pain as bad as you can imagine

4. Please rate your pain by circling the one number that best describes your pain at its least in the last 24 hours.

0	1	2	3	4	5	6	7	8	9	10
No pain										Pain as bad as you can imagine

5. Please rate your pain by circling the one number that best describes your pain on the average.

0	1	2	3	4	5	6	7	8	9	10
No pain										Pain as bad as you can imagine

6. Please rate your pain by circling the one number that tells how much pain you have right now.

0	1	2	3	4	5	6	7	8	9	10
No pain										Pain as bad as you can imagine

FIGURE 3-1. Brief Pain Inventory (short form).

7. What treatments or medications are you receiving for your pain?

8. In the last 24 hours, how much relief have pain treatments or medications provided? Please circle the one percentage that most shows how much relief you have received.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
No pain										Complete relief

9. Circle one number that describes how, during the past 24 hours, pain has interfered with your:

A. General Activity

0	1	2	3	4	5	6	7	8	9	10
Does not interfere										Completely interferes

B. Mood

0	1	2	3	4	5	6	7	8	9	10
Does not interfere										Completely interferes

C. Walking Ability

0	1	2	3	4	5	6	7	8	9	10
Does not interfere										Completely interferes

D. Normal Work (includes both work outside the home and housework)

0	1	2	3	4	5	6	7	8	9	10
Does not interfere										Completely interferes

E. Relations with Other People

0	1	2	3	4	5	6	7	8	9	10
Does not interfere										Completely interferes

F. Sleep

0	1	2	3	4	5	6	7	8	9	10
Does not interfere										Completely interferes

G. Enjoyment of Life

0	1	2	3	4	5	6	7	8	9	10
Does not interfere										Completely interferes

FIGURE 3-1, cont'd.

deteriorate quickly, and you should plan accordingly. We probably are dealing with weeks to months, although some patients do better, and some do worse. Over time, the course may become clearer, and if you wish, I may be able to be a little more precise about what we are facing."

The physician must discuss these topics in an empathic way. Palliative care conversations are as much about emotions as facts.[5] Talking about disease progression or death may elicit negative emotions such as anxiety, sadness, or frustration. These emotions decrease a patient's quality of life and interfere with the ability to hear factual information. Empathic responses strengthen the patient-physician relationship, increase the patient's satisfaction, and make the patient more likely to disclose other concerns. The first step is recognizing when the patient is expressing emotions. Once the physician recognizes the emotion being expressed, he or she can respond empathically.

It is also important for physicians to recognize their own emotional reactions to these conversations. The physician's emotional reactions color impressions of the patient's prognosis, thereby making it hard to listen to the patient, and may influence the physician to hedge bad news. The physician should become aware of her or his own emotional reactions to ensure that the conversation focuses on the patient rather than the health care provider's needs.

In addition to good communication skills, palliative care requires a basic knowledge of medical ethics and the law. For example, patients have the moral and legal right to refuse any treatment, even if refusal results in their death. There is no legal difference between withholding and withdrawing life-sustaining treatment. When confronted with areas of ambiguity, the physician should know how to obtain either a palliative care or ethics consultation.

TABLE 3-2 FICA SPIRITUAL ASSESSMENT TOOL

F—What is your **faith**/religion? Do you consider yourself a religious or spiritual person? What do you believe in that gives meaning/importance to life?

I—**Importance** and **influence** of faith. Is your faith/religion important to you? How do your beliefs influence how you take care of yourself? What are your most important hopes? What role do your beliefs play in regaining your health? What makes life most worth living for you? How might your disease affect this?

C—Are you part of a religious or spiritual **community**? Is this of support to you, and how? Is there a person you really love or is very important to you? How is your family handling your illness? What are their reactions/expectations?

A—How would you like me to **address** these issues in your health care? What might be left undone if you were to die today? Given the severity or chronicity of your illness, what is most important for you to achieve? Would you like me to talk to someone about religious/spiritual matters?

From Puchalski C, Romer A. Taking a spiritual history. *J Palliat Med.* 2000;3:129-137.

TABLE 3-3 CORE COMMUNICATION SKILLS

RECOMMENDED SKILL	EXAMPLE
A. IDENTIFYING CONCERNS AND RECOGNIZING CUES	
Elicit Concerns	
Open-ended questions	"Is there anything you wanted to talk to me about today?"
Active listening	Allowing patient to speak without interruption; allowing pauses to encourage patient to speak
Recognize Cues	
Informational concerns	Patient: "I'm not sure about the treatment options"
Emotional concerns	Patient: "I'm worried about that"
B. RESPONDING TO INFORMATIONAL CONCERNS	
"Ask-tell-ask"	Topic: communicating information about cancer stage
Ask	"Have any of the other doctors talked about what stage this cancer is?"
Tell	"That's right, this is a stage IV cancer, which is also called metastatic cancer…"
Ask	"Do you have questions about the staging?"
C. RESPONDING TO EMOTIONAL CONCERNS	
Nonverbal Empathy: S-O-L-E-R	
S	Face the patient **S**quarely
O	Adopt an **O**pen body posture
L	**L**ean toward the patient
E	Use **E**ye contact
R	Maintain a **R**elaxed body posture
Verbal Empathy: N-U-R-S-E	
N	**N**ame the emotion: "You seem worried"
U	**U**nderstand the emotion: "I see why you are concerned about this"
R	**R**espect the emotion: "You have shown a lot of strength"
S	**S**upport the patient: "I want you to know that I will still be your doctor whether you have chemotherapy or not"
E	**E**xplore the emotion: "Tell me more about what is worrying you"

From Back AL, Arnold RM, Tulsky JA. *Discussing Prognosis.* Alexandria, VA: American Society of Clinical Oncology; 2008.

During the past 10 years, there has been a societal push to encourage patients to designate health care proxies and to create advance care planning documents, typified by the use of living wills. These documents are meant to protect patients against unwanted treatments and to ensure that as they are dying, their wishes are followed.[6] Unfortunately, there are few empirical data showing that these documents actually change practice. Still, discussions of the documents with health professionals and family members generally provoke important conversations about end-of-life care decisions and may help families confronted with difficult situations know they are respecting their loved one's wishes.

TABLE 3-4 DISCUSSING PALLIATIVE CARE

GENERAL APPROACH

- Plan what to say. Create the right setting, allow adequate time, and determine who else should be present at the meeting.
- Listen carefully. Be prepared for strong emotions, respond empathetically, encourage description of feelings, and allow time for silence and response.

ESTABLISHING GOALS OF MEDICAL CARE

- Determine what the patient knows. Clarify any uncertainties or misconceptions.
- Understand what the patient is hoping to accomplish as well as any fears and worries.
- Repeat the goals back to the patient to make sure they are heard.
- Suggest treatments to meet these goals and clarify what will not be done because it will not help achieve the goals. Focus on the goals that you think you can achieve. Plan follow-up, review and revise plan as needed.

COMMUNICATING BAD NEWS

- Determine what the patient knows, wants to know, and can comprehend.
- Share information, recognizing that people handle information in different ways.
- Avoid jargon, pause frequently, check for understanding, and use silence.
- Recognize and support the patient's emotional reaction.
- Assess the patient's safety.
- Agree to a plan that enlists potential sources of support.

WITHDRAWING TREATMENT

- Discuss the context of the current discussion and what has changed to precipitate it.
- Review prior treatment goals and reassess their virtues.
- Discuss alternative treatments based on the new goals.
- Document a plan for forgoing treatment and share with the patient, the patient's family, and the health care team.

Adapted from Morrison RS, Meier DE. Clinical practice. Palliative care. *N Engl J Med.* 2004;350:2582-2590.

Is the Patient Going to Die in the Location of Choice?

Most patients say that they want to die at home. Unfortunately, most patients die in institutions—either hospitals or nursing homes. Burdensome transitions decrease quality in end-of-life care. Good palliative care requires establishing a regular system of communication to minimize transitional errors. A social worker who knows about community resources is important in the development of a dispositional plan that respects the patient's goals.

Hospice programs are an important way to allow patients to die at home. In the United States, *hospice* refers to a specific, government-regulated form of end-of-life care, available under Medicare since 1982 but subsequently adopted by Medicaid and many other third-party insurers. Hospice care typically is given at home, a nursing home, or specialized acute care unit. Care is provided by an interdisciplinary team, which usually includes a physician, nurse, social worker, chaplain, volunteers, bereavement coordinator, and home health aides, all of whom collaborate with the primary care physician, patient, and family. Bereavement services are offered to the family for a year after the death.

Hospices are paid on a per diem rate and are required to cover all the costs related to the patient's life-limiting illness. Because of this and the fact that their focus is on comfort rather than life prolongation, many hospices will not cover expensive treatments such as inotropic agents in heart failure or chemotherapy in cancer, even if they have a palliative effect. Many hospices are experimenting with different service models in an attempt to enroll patients earlier in the course of their illness and increase access to their services.

Grade A References

A1. Temel JS, Greer JA, Muzikansky A, et al. Early palliative care for patients with metastatic non-small-cell lung cancer. *N Engl J Med.* 2010;363:733-742.

A2. Michna E, Cheng WY, Korves C, et al. Systematic literature review and meta-analysis of the efficacy and safety of prescription opioids, including abuse-deterrent formulations, in non-cancer pain management. *Pain Med.* 2014;15:79-92.

A3. Abernethy AP, McDonald CF, Frith PA, et al. Effect of palliative oxygen versus room air in relief of breathlessness in patients with refractory dyspnoea: a double-blind, randomised controlled trial. *Lancet.* 2010;376:784-793.

A4. Laoutidis ZG, Mathiak K. Antidepressants in the treatment of depression/depressive symptoms in cancer patients: a systematic review and meta-analysis. *BMC Psychiatry.* 2013;13:140.

A5. Gallo JJ, Morales KH, Bogner HR, et al. Long term effect of depression care management on mortality in older adults: follow-up of cluster randomized clinical trial in primary care. *BMJ.* 2013;346:f2570.

A6. Jiang W, Krishnan R, Kuchibhatla M, et al. Characteristics of depression remission and its relation with cardiovascular outcome among patients with chronic heart failure (from the SADHART-CHF Study). *Am J Cardiol*. 2011;107:545-551.
A7. Lautrette A, Darmon M, Megarbane B, et al. A communication strategy and brochure for relatives of patients dying in the ICU. *N Engl J Med*. 2007;356:469-478.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Gill TM, Gahbauer EA, Han L, et al. Trajectories of disability in the last year of life. *N Engl J Med*. 2010;362:1173-1180.
2. Quill TE, Abernethy AP. Generalist plus specialist palliative care: creating a more sustainable model. *N Engl J Med*. 2013;368:1173-1175.
3. Brunner-La Rocca HP, Rickenbacher P, Muzzarelli S, et al. End-of-life preferences of elderly patients with chronic heart failure. *Eur Heart J*. 2012;33:752-759.
4. Center to Advance Palliatrive Care. https://www.capc.org/. Accessed February 9, 2015.
5. National Consensus Project for Quality Palliative Care. Clinical Practice Guidelines for Quality Palliative Care, 2013. http://www.nationalconsensusproject.org. Accessed February 9, 2015.
6. Silveira MJ, Kim SY, Langa KM. Advance directives and outcomes of surrogate decision making before death. *N Engl J Med*. 2010;362:1211-1218.

REVIEW QUESTIONS

1. A 75-year-old man with lung cancer is admitted to the hospital with severe shortness of breath. Work-up reveals no other cause of his shortness of breath other than lymphogenic spread of his cancer. His oxygen saturation is 94%. Which of the following treatments should be instituted for his dyspnea?

A. Morphine
B. Benzodiazepines
C. Oxygen
D. A and C
E. All the above

Answer: A In randomized controlled data, opioids have been shown to decrease dyspnea both in lung cancer patients and in patients with COPD. Oxygen is helpful only if the patient has hypoxia. Benzodiazepines have not been shown to decrease breathlessness.

2. Which of the following is NOT required for a patient to be in hospice?

A. The patient must be DNR.
B. The patient must have a life-limiting illness, which is likely to cause her death in 6 months.
C. The patient wishes to focus on quality of life rather than longevity of life.
D. If the patient lives at home, she must have a primary caregiver.

Answer: A The patient does not have to be DNR to be in hospice. The others are requirements of hospice.

3. Which of the following is true of depression in life-limiting illnesses?

A. It is a normal reaction when people have a life-limiting illness, and it should not be treated.
B. It cannot be improved because the treatments take too long to work in patients with serious illness.
C. Treatment of depression decreases both morbidity and mortality.
D. It requires a psychiatric consult because treatment is very complicated.

Answer: C Data show that the treatment of depression improves both quality of life and mortality.

4. Which of the following is true?

A. Telling patients that they have a terminal illness will result in their losing hope.
B. Telling patients they have a terminal illness has no impact on their desire for future treatment.
C. Telling patients that they have terminal illnesses is associated with their choosing hospice more frequently.
D. Patients have clearly stated that they do not want to be told that they have a terminal illness.

Answer: C Data suggest that telling patients that they have a life-limiting illness is associated with a lower likelihood of choosing aggressive care at the end of life and is not associated with poorer psychiatric outcomes.

4

CULTURAL CONTEXT OF MEDICINE

VICTORIA M. TAYLOR

The 2010 U.S. Census counted about 39 million blacks or African Americans (13% of the population), nearly 15 million Asian Americans (5% of the population), about 3 million American Indians and Alaska Natives, and more than 500,000 Native Hawaiians and other Pacific Islanders. It also counted more than 50 million individuals of Hispanic or Latino origin (16% of the population). Approximately 40 million Americans (13% of the population) were foreign born. One in 2 immigrants to the United States have limited English proficiency (i.e., they do not speak English very well or fluently), and 1 in 10 immigrants do not speak English at all (Fig. 4-1).

During the past two decades, a large body of literature has documented substantial disparities in health status. Although some of these disparities are based on socioeconomic status, many are based on race, ethnicity, or other characteristics. Black men have a substantially higher age-adjusted incidence of prostate cancer than do white men (236 per 100,000 versus 147 per 100,000). American Indians/Alaska Natives are more than twice as likely as non-Latino whites of a similar age to have diabetes. More than half of the Americans who are living with chronic hepatitis B infection are Asians or Pacific Islanders. Lesbian, gay, bisexual, and transgender individuals have higher rates of suicidal behavior compared with heterosexual individuals. A major goal of Healthy People 2020 is to eliminate health disparities for preventable and treatable conditions such as cancer, diabetes, and human immunodeficiency virus infection.

Culture can be defined as a shared system of values, beliefs, and patterns of behavior, and it is not simply defined by race and ethnicity. Culture can also be shaped by factors such as country and region of origin, acculturation, language, religion, and sexual orientation. For instance, the black population of the northeastern United States includes individuals who moved from southern states decades ago as well as recent immigrants from Ethiopia. As the United States population becomes increasingly diverse and as pronounced differences in health status continue to be documented, consideration of the cultural context of medicine is becoming a national priority.

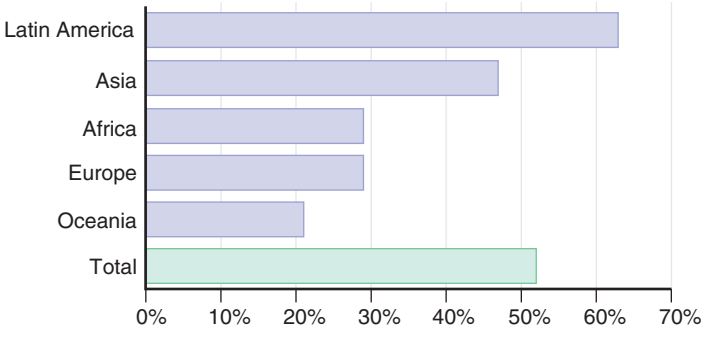

FIGURE 4-1. Proportion of immigrants aged 5 years and older with limited English proficiency by region of origin. (From Grieco EM, Acosta YD, de la Cruz P, et al. The foreign-born population in the United States: 2010. Washington DC: U.S. Department of Commerce; 2012.)

DISPARITIES IN HEALTH CARE ACCESS AND QUALITY

Components of health care access include the ability to get into the health care system as well as to obtain appropriate care once in the system. The availability of health care providers who meet an individual patient's needs is another key component of access to care. Quality care is based on scientific evidence (i.e., is effective), avoids injury to the patient (i.e., is safe), minimizes harmful delays (i.e., is timely), is responsive to the individual patient's needs (i.e., is patient centered), promotes communication among providers (i.e., is coordinated), does not vary because of personal characteristics (i.e., is equitable), and avoids waste (i.e., is efficient).

Access to Health Care

Racial and ethnic minority groups, particularly immigrants, disproportionately have problems accessing health care. Before the implementation of the Affordable Care Act, the proportions of Latinos and Native Americans/Alaska Natives who lacked health insurance was more than twice the proportion among non-Latino whites, and less than two thirds of Americans with limited English proficiency were insured. About 1 in 3 Korean American and Vietnamese American adults had no regular source of medical care compared with about 1 in 10 non-Latino white adults.

Blacks and Latinos are far less likely than are whites and Asians to have access to physicians of their own race and ethnicity. This imbalance is important because racial concordance between physicians and patients can improve the processes of care. For example, patients with race-concordant physicians are more likely to use needed health services, are less likely to postpone or delay seeking care, and are more satisfied with their care than are patients in race-discordant relationships. Whether these differences translate into different health outcomes, however, is less clear.[1]

Quality of Health Care

National surveys confirm population-level disparities in the quality of preventive care. Recent immigrants have far lower levels of interval screening for breast, cervical, and colorectal cancer than do individuals who were born in the United States (Fig. 4-2). The proportion of Native Hawaiians and other Pacific Islanders whose serum cholesterol levels are measured at least once every 5 years is significantly lower than among whites. In 2011, only 40% of Asians aged 65 years and older had ever received the pneumococcal vaccine compared with 67% of non-Latino whites.

Racial and ethnic disparities have been documented for a number of specific clinical situations. For example, Latino women with breast cancer are less likely to receive radiation therapy within a year of breast-conserving surgery than are white women, Native Americans and Alaska Natives are less likely than whites to receive recommended care such as initial antibiotics within 6 hours of hospital arrival, and blacks with end-stage renal disease

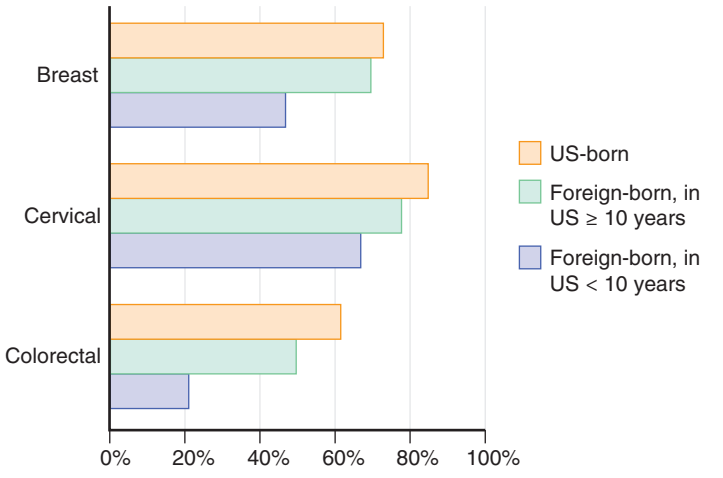

FIGURE 4-2. Adherence to cancer screening guidelines by immigration status. Breast = mammography during last 2 years among women aged 50 to 74 years. Cervical = Papanicolaou test during last 3 years among women aged 21 to 65 years. Colorectal = among individuals aged 50 to 75 years, fecal occult blood test last year; sigmoidoscopy last 5 years and fecal occult blood test last 3 years; or colonoscopy last 10 years. (From Centers for Disease Control and Prevention. Cancer screening—United States, 2010. *MMWR Morb Mortal Wkly Rep.* 2012; 61:41-45.)

are less likely to be entered to a transplant list than are whites. Moreover, disparities in the quality of care are found even when variations in insurance status, income, and comorbid conditions are taken into account.

Disparities in health care quality exist even in systems that are generally believed to provide equal access.[2] For example, in the Veterans Affairs Health System, disparities between blacks and whites have been documented for blood pressure control among patients with hypertension, cholesterol control among patients with coronary heart disease, and glucose control among patients with diabetes. Moreover, these disparities persist even after adjusting for location and socioeconomic status. Similar disparities have been documented in Medicare managed care programs between elderly blacks and whites with diabetes and cardiovascular disorders.

CULTURAL COMPETENCE IN HEALTH CARE

Health disparities can be reduced or perhaps even eliminated by maintaining culturally competent health care systems. Cultural competence may be defined as a set of congruent attitudes, behaviors, and policies that come together both among professionals and within systems to enable effective work in cross-cultural situations (Fig. 4-3). Ongoing efforts to improve cultural competence in the health care system target organizational, structural, and clinical barriers. These initiatives aim to close gaps in health status, to decrease differences in the quality of care, to enhance patients' satisfaction, and to increase patients' trust.

Organizational Barriers and Interventions

Diversity among health care professionals is associated with better access to care for disadvantaged populations. Black and Latino physicians are more likely than their white colleagues to work in medically underserved communities and to have a better understanding of barriers to health care. Because less than 10% of practicing physicians are black or Latino, and only about 15% of medical school students are from one of these groups, many U.S. medical schools have implemented comprehensive programs to infuse diversity among their students, resident physicians, and faculty.

About two thirds of the patients who receive care at federally funded community health centers in medically underserved areas are members of racial and ethnic minority groups. In these health centers, patients are three times more likely to have limited proficiency in English compared with the general population. The community health center model has proved effective not only in increasing access to care but also in improving continuity of care and health outcomes. For example, medically underserved communities with community health centers have fewer preventable hospitalizations and uninsured emergency department visits than do similar communities without health centers. Compared with national rates, community health centers report minimal racial and ethnic disparities in clinical outcomes such as the control of diabetes and hypertension.[3]

Structural Barriers and Interventions

Accumulating evidence suggests that trained professional interpreters can improve the clinical care received by individuals with limited English proficiency. However, interpreter services often remain ad hoc, with family members and untrained nonclinical employees acting as interpreters.[4] Use of ad hoc services has potentially negative clinical consequences, including breach of the patient's confidentiality and inaccurate communication. One major obstacle to the implementation of professional interpreter programs is a lack of reimbursement; Medicare and most private insurers do not pay for interpretation and related services, and most states do not pay for interpretation under Medicaid.

Assistance with navigation represents a promising model to enable racial and ethnic minority patients to move through the health system effectively and to be actively involved in decision making about their medical care.[5] Guides may be nurses, social workers, or volunteers who are familiar with the health care system. They help patients and their families navigate the treatment process, steering them around obstacles that may limit their access to quality care, choice of doctors, and access to treatment options. For example, an American Cancer Society navigation program is effective in reducing the time to diagnostic resolution after abnormal cancer screening tests in medically underserved patients.

Another option for closing the gap in health care among various minority populations is community health workers.[6] In general, community health workers live locally and share the language and culture of the patients being served. Lay community health workers provide cultural mediation between communities and the health care system; culturally appropriate and accessible health education and information; help in obtaining needed medical services, informal counseling, and social support; and advocacy within the health care system. The effectiveness of community health workers is documented by a study in which Mexican American women randomized to

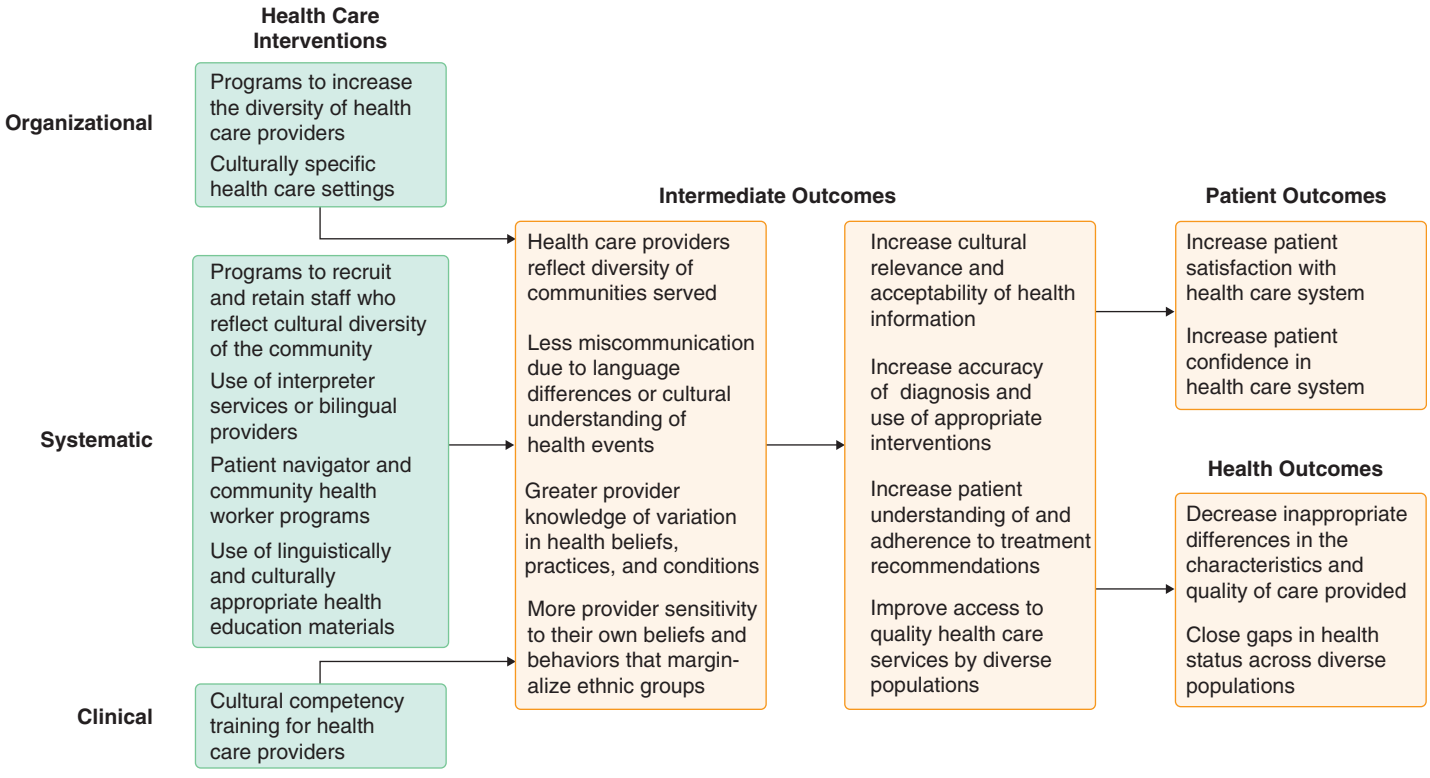

FIGURE 4-3. Analytic framework for evaluating the effectiveness of health care interventions to increase cultural competence. (From Anderson LM, Scrimshaw SC, Fullilove MT, et al., for the Task Force on Community Preventive Services. Culturally competent healthcare systems: a systemic review. *Am J Prev Med.* 2003; 24[suppl]:68-79.)

receive health worker education were significantly more likely to obtain Papanicolaou tests than were women randomized to receive usual care.[A] The largest formal system of community health workers is the Indian Health Service, which currently has about 1400 community health representatives.

Clinical Barriers and Interventions

Patients who are members of racial and ethnic minority groups often understand health and disease (i.e., explanatory model) differently than the general population. For example, many Vietnamese people believe that disease is caused by an imbalance of the humoral forces of yin and yang. When ill, they commonly use Chinese herbal medicine as well as indigenous folk practices known as Southern medicine in an effort to restore the balance of humoral forces. In addition, Vietnamese patients may think that Western medicine is too strong and will upset the internal balance. Consequently, a hypertensive Vietnamese patient may, for example, use Chinese herbal medicines instead of prescribed antihypertensive medication. Alternatively, the patient may take a lower dose of medication than prescribed by his or her physician.

Cultural competency training for health care providers generally includes teaching cross-cultural knowledge and communication skills, while avoiding stereotypes.[7] Examples include the effect of prejudice on gays and lesbians and how this prejudice shapes their interactions with the health care system, and common spiritual practices that might interfere with prescribed therapies (such as Ramadan fasting practices, when observed by diabetic Muslim patients). Communication skills that can be addressed in cultural competence training include approaches to eliciting patients' explanatory models and use of traditional treatments, as well as methods for negotiating different styles of communication and levels of family participation in decision-making. Cultural competency training improves the attitudes and skills of health professionals as well as patient satisfaction, but there is less evidence that it improves clinical outcomes.[A]

SUMMARY

Individual clinical practices should regularly assess their current organizational climate, policies, and training related to diversity. Practices can address health disparities by hiring clinical and office staff who are representative of the communities they serve, by routinely using professional interpreters during clinical encounters with patients who have limited proficiency with English, by offering cultural competency education and training to physicians and staff, and by providing educational and informational materials that are culturally and linguistically appropriate for their patient populations.[8]

National and state efforts to improve cultural competence in health care, whether used alone or in conjunction with socioeconomic initiatives, are likely to play a significant role in reducing health disparities across population subgroups. An important goal of the Affordable Care Act is to reduce health disparities by expanding health insurance coverage, addressing diversity in the health care workforce, increasing the capacity of community health centers, and promoting the use of patient navigators and community health workers.

Grade A References

A1. Bagchi AD, Dale S, Verbitsky-Savitz N, et al. Examining effectiveness of medical interpreters in emergency departments for Spanish-speaking patients with limited English proficiency: results of a randomized controlled trial. *Ann Emerg Med*. 2011;57:248-256.

A2. Paskett ED, Katz ML, Post DM, et al. The Ohio Patient Navigation Research Program: does the American Cancer Society patient navigation model improve time to resolution in patients with abnormal screening tests? *Cancer Epidemiol Biomarkers Prev*. 2012;21:1620-1628.

A3. Byrd TL, Wilson KM, Smith JL, et al. AMIGAS: a multicity, multicomponent cervical cancer prevention trial among Mexican American women. *Cancer*. 2013;119:1365-1372.

A4. Sequist TD, Fitzmaurice GM, Marshall R, et al. Cultural competency training and performance reports to improve diabetes care for black patients: a cluster randomized, controlled trial. *Ann Intern Med*. 2010;152:40-46.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Meghani SH, Brooks JM, Gipson-Jones T, et al. Patient-provider race-concordance: does it matter in improving minority patients' health outcomes. *Eth Health*. 2009;14:107-130.
2. Trivedi AN, Grebla RC, Wright SM, et al. Despite improved quality of care in the Veterans Affairs Health Care System, racial disparity persists for some clinical outcomes. *Health Aff*. 2011;4: 707-715.
3. Lebrun LA, Shi L, Zhu J, et al. Racial/ethnic differences in clinical quality performance among health centers. *J Ambul Care Manage*. 2013;36:24-34.
4. VanderWielen LM, Enurah AS, Rho HY, et al. Medical interpreters: improvements to address access, equity, and quality of care for limited-English-proficient patients. *Acad Med*. 2014;89:1324-1327.
5. Natale-Pereira A, Enard KR, Nevarez L, et al. The role of patient navigators in eliminating health disparities. *Cancer*. 2011;117(suppl):3543-3552.
6. Brownstein JN, Hirsch GR, Rosenthal EL, et al. Community health workers 101 for primary care providers and other stakeholders in health care systems. *J Ambul Care Manage*. 2011;34:210-220.
7. Betancourt JR, Cervantes MC. Cross-cultural medical education in the United States: key principles and experiences. *Kaohsiung J Med Sci*. 2009;25:472-478.
8. Chin MH, Clarke AR, Nocon RS, et al. A roadmap and best practices for organizations to reduce racial and ethnic disparities in health care. *J Gen Intern Med*. 2012;27:992-1000.

REVIEW QUESTIONS

1. What percentage of immigrants to the United States have limited English proficiency?

A. 20%
B. 30%
C. 40%
D. 50%
E. 60%

Answer: D The 2010 Census found that one in two immigrants have limited English proficiency. (Grieco EM, Acosta YD, de la Cruz GP, et al. The foreign born population in the United States: 2010. Washington, DC: U.S. Department of Commerce, 2012.)

2. A 19-year-old Vietnamese man presents for a routine physical examination. He is a recent immigrant, has no symptoms or significant medical history, and has not previously had a physical examination or any blood testing in the United States. Which of the following blood tests should you perform?

A. HIV
B. Hepatitis B
C. Glucose
D. Cholesterol
E. All of the above

Answer: B More than half of the Americans who have chronic hepatitis B infection are Asians or Pacific Islanders. Therefore, all immigrants from Asia should be tested for hepatitis B. (Pollack H, Wang S, Wyatt Le, et al. A comprehensive screening and treatment model for reducing disparities in hepatitis B. *Health Aff.* 2011;30:1974-1983.)

3. Which of the following statements about health insurance is incorrect?

A. Whites are more likely to have insurance than American Indians/Alaska Natives.
B. Latinos are less likely to have insurance than non-Latino whites.
C. Gay/lesbian/homosexual individuals are less likely to have insurance than heterosexual individuals.
D. Citizens are more likely to have insurance than noncitizens.
E. Individuals who are proficient in English are more likely to have insurance than individuals who are not.

Answer: C The California Health Interview Survey provides information about health insurance coverage among population subgroups and shows differences by race/ethnicity, citizenship status, and level of English proficiency but not by sexual orientation. (University of California Los Angeles. Ask CHIS 2009, http://www.chis.ucla.edu. Accessed March 22, 2014.)

4. Which of the following are core community health worker functions?

A. Cultural mediation
B. Health education
C. Informal counseling
D. Social support
E. All of the above

Answer: E Cultural mediation, health education, informal counseling, and social support are all core community health worker functions. (Brownstein JN, Hirsch GR, Rosenthal EL, Rush CH. Community health workers 101 for primary care providers and other stakeholders in health care systems. *J Ambul Care Manage.* 2011;34:210-220.)

5. Your practice has recently started seeing a large number of Hispanic immigrant patients with limited English proficiency. You have decided to make some changes to your practice to accommodate the specific needs of these patients. Which of the following approaches are appropriate?

A. Hire Hispanic staff
B. Ask limited English proficiency patients to bring a family member with them to provide
C. Medical interpretation
D. Provide cultural competency training to providers
E. A and C
F. All of the above

Answer: D Clinical practices can address health disparities in multiple ways, including by hiring staff who are representative of the practice population, providing professional interpreter services, offering cultural competency training to providers, and providing linguistically appropriate patient education materials. Family members should not be asked to provide medical interpretation. (Washington DL, Bowles J, Saha S. Transforming clinical practice to eliminate racial-ethnic disparities in healthcare. *J Gen Intern Med.* 2008;23:685-691.)

5

SOCIOECONOMIC ISSUES IN MEDICINE

STEVEN A. SCHROEDER

All nations—rich and poor—struggle with how to improve the health of the public, obtain the most value from medical services, and restrain rising health care expenditures. Many developed countries also wrestle with the paradox that their citizens have never been so healthy or so unhappy with their medical care. Despite the reality that only about 10% of premature deaths result from inadequate medical care, the bulk of professional and political attention focuses on how to obtain and pay for state-of-the-art medical care. By comparison, 40% of premature deaths stem from unhealthy behaviors—including smoking (about 44%; Chapter 32), excessive or unwise drinking (about 11%; Chapter 33), obesity and insufficient physical activity (about 15% but estimated to rise substantially in the years to come; Chapters 16 and 220), illicit drug use (about 2%; Chapter 34), and imprudent sexual behavior (about 3%; Chapter 285). Genetics (Chapter 40) account for an additional 30%; social factors—discussed next—account for 15%, and environmental factors (Chapter 19) account for 5%. Of the major behavioral causes of premature deaths, tobacco use (Chapter 32) is by far the most important, although recent increases in obesity (Chapter 220) and physical inactivity (Chapter 16) are also alarming. Health is influenced by genetic predisposition, behavioral patterns, environmental exposures, social circumstances, and health care.

SOCIAL STATUS INFLUENCES HEALTH

Socioeconomic status, or class, is a composite of many different factors, including income, net wealth, education, occupation, and neighborhood. In general, people in lower classes are less healthy and die earlier than people at higher socioeconomic levels, a pattern that holds true in a stepwise fashion from the poorest to the richest. In the United States, the association between health and class is usually discussed in terms of racial and ethnic disparities; but in fact, race and class are independently associated with health status, and it can be argued that class is the more important factor. For example, U.S. racial disparities in the prevalence of adult smoking are relatively small among whites, blacks, and Hispanic Americans, whereas there are huge differences among smoking rates by educational level (Fig. 5-1).[1] U.S. physicians have reduced their smoking prevalence to a record low of only 1%. Although both smoking rates and the numbers of cigarettes smoked by those who continue to smoke are gradually declining (Fig. 5-2), more than 43 million Americans and millions more elsewhere continue to smoke.[2] Because people of higher socioeconomic status adopt health-promoting behaviors at a faster rate than people of lower socioeconomic status, overall population health can increase while health disparities also widen (Fig. 5-3).

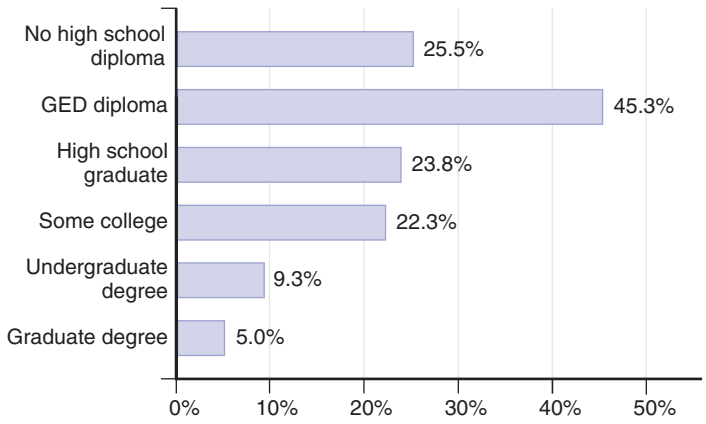

FIGURE 5-1. Prevalence of adult smoking, by education, United States, 2011. GED = General Education Development. (From Centers for Disease Control and Prevention. Current cigarette smoking among adults: United States, 2011. *MMWR Morb Mortal Wkly Rep.* 2012;61:889-894.)

In part, the relationship between class and health is mediated by higher rates of unhealthy behaviors among the poor, such as the inverse relationship between educational attainment and cigarette smoking, but unhealthy behaviors do not fully explain the poor health of those in the lower socioeconomic classes. Even when such behaviors are held constant, people in lower socioeconomic classes are much more likely to die prematurely than are people of higher classes. Of interest is that first-generation immigrants to the United States appear to be more protected from the adverse health consequences of low socioeconomic status than are subsequent generations.

It is unclear which of the components of class—education, wealth (either absolute wealth or the extent of the gap between rich and poor), occupation, or neighborhood—makes the greatest impact on a person's health. Most likely, it is a combination of all of them. For example, the constant stress of a lower class existence—lack of control over one's life circumstances, social isolation, and the anxiety derived from the feeling of having low status—is linked to poor health. This stress may trigger a variety of neuroendocrinologic responses that are useful for short-term adaptation but bring long-term adverse health consequences.

What can clinicians do with this knowledge? Clearly, it is difficult to write prescriptions for more income, a better education, good neighborhoods, or high-paying jobs. Physicians can, however, encourage healthy behavior. At key times of transition, such as during discharge planning for hospitalized patients, clinicians should be attentive to social circumstances. For patients who are likely to be socially isolated, clinicians should encourage or arrange interactions with family, neighbors, religious organizations, or community agencies to improve the likelihood of optimal outcomes. Access points to vital social services, such as child care, disability insurance, and food supplementation, can be provided in clinical settings.[3] In addition, physicians should seek to identify and eliminate any aspects of racism in health care institutions (Chapter 4). Finally, in their role as social advocates, physicians can promote such goals as safe neighborhoods, improved schools, and access to quality health care.

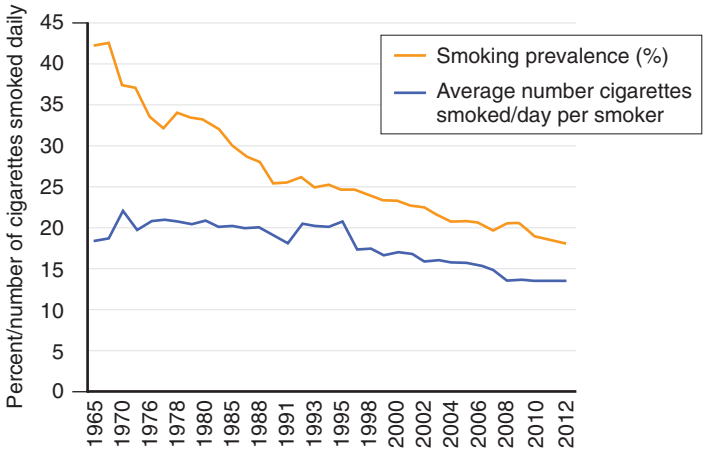

FIGURE 5-2. Smoking prevalence and average number of cigarettes smoked per day per current smoker. (Data based on Centers for Disease Control and Prevention (CDC). Smoking prevalence, 1965-2010. *MMWR Morb Mortal Wkly.* 2011;60:109-113; Current cigarette smoking in the United States: current estimate. CDC; http://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking. Accessed February 10, 2015; National Health Interview Survey. CDC; http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm. Accessed February 10, 2015; Jamal A, Agaku IT, O'Connor E, et al. Current cigarette smoking among adults—United States, 2005-2013. *MMWR Morb Mortal Wkly.* 2014;63:1108-1112.)

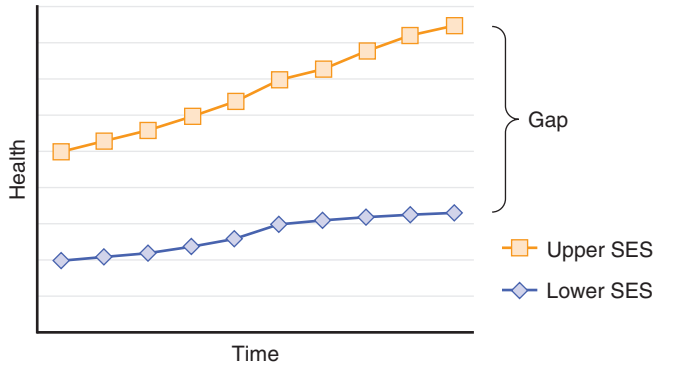

FIGURE 5-3. Health improves while disparities widen. SES = socioeconomic status.

ECONOMIC ISSUES IN MEDICAL CARE

Medical care today is on a collision course. On the one hand, an ever-expanding science base continuously generates new technologies and drugs that promise a longer and healthier life. Add a public eager to obtain the latest breakthroughs touted in the media and over the Internet, plus a well-stocked medical industry eager to meet that demand, and it is easy to understand why expenditures continue to soar. On the other hand, payers for medical care—health insurance companies, government (federal, state, and local), and employers—increasingly bridle at medical care costs.

The United States continues to lead the world in health care expenditures.[4] In 2011, it spent more than $2.7 trillion, amounting to 17.9% of the gross domestic product. Most policy analysts contend that this rate of increase in medical care expenditures is unsustainable, but this claim has been made for many years. A potent combination of supply and demand factors explains why the United States spends so much.[5] On the supply side, the United States far exceeds other countries in the availability and use of expensive diagnostic technologies, such as magnetic resonance imaging and computed tomography. For example, the United States has four times as many magnetic resonance imaging machines per capita as does Canada. Similar patterns exist for therapeutic technologies, whether coronary angioplasty, cancer chemotherapy, or joint prostheses. The differences are especially dramatic in older patients. Other supply factors that drive high medical expenditures in the United States include a fee-for-service payment system that compensates physicians much more when they use expensive technologies than when they do not[6]; a medical professional work force that earns much higher incomes relative to the population than in other nations and that emphasizes specialist rather than generalist practice; accelerated development of new and costly medications that are directly marketed to consumers; much higher administrative costs; higher rates of fraud and abuse; and a high rate of defensive medicine in response to pervasive fears about medical malpractice suits. Supply factors that do not appear to be unique to the United States are the number of physicians or hospitals. Many other developed countries have a much larger physician work force relative to their population, as well as a much higher ratio of primary care physicians to specialists. The number of hospitals and hospital beds, the frequency of hospitalizations, and the length of hospital stay are relatively low in the United States, although it does have a much greater proportion of intensive care beds. Finally, recent analyses suggest that a principal driver of high expenditures on health care in the United States is the much greater price charged per unit of service compared with other developed countries.

Demand factors also drive medical expenditures. The extent to which the media and the medical profession feature medical "breakthroughs" is extensive and one-sided. New promising treatments merit front-page stories and commercial advertisements, whereas subsequent disappointing results are buried or ignored. The cumulative result is to whet patients' appetite for more and to leave the impression that good health depends only on finding the right treatment. This same quest explains the popularity of alternative medicine, for which patients are willing to spend $34 billion annually out of their own pockets (Chapter 39). The cumulative impact of these supply and demand drivers is that there are incentives to do more at every step of the American medical system.[5]

It could be argued that rising expenditures for medical care are not a bad thing. What could be more important than ensuring maximal health? There are several rebuttals to that argument. First, it is not clear that money spent on medical care brings appropriate value in the United States, given that its health statistics are worse than those of virtually every other developed country. Second, there are substantial regional differences in the supply and use of medical care, such as a two-fold difference in the supply of acute hospital beds and a four-fold difference in the risk of being hospitalized in an intensive care unit at the end of life. Similar regional differences exist for procedures such as transurethral prostatectomy, hysterectomy, and coronary artery bypass surgery. Yet there is no evidence that "more is better" on a regional basis.

Consequently, rising health care expenditures are stressing public programs such as Medicare, Medicaid, the Veterans Administration health system, and municipal hospitals, with budget requests outstripping the tax base to pay for them. Medical debt is by far the most important cause of bankruptcy. Finally, as health care becomes less affordable for businesses and

government, the number of people without health insurance will continue to increase.

Cost-Containment Strategies

Since the mid-1970s, a variety of strategies to contain rising medical expenditures have yielded limited success.[5] These attempts have tried to restrict the supply of costly medical technologies as well as the production of physicians, especially specialists; to promote health maintenance organizations that have incentives to spend less on medical care; to ration indirectly by limiting health insurance coverage; to institute prospective payment for hospital care; to use capitation payments or discounted fee schedules for physician reimbursement; to introduce gatekeeper mechanisms to reduce access to costly care; to put patients at more financial risk for their own medical care; to reform malpractice procedures; to reduce administrative costs; and to encourage less aggressive care at the end of life. The most recent suggestions—comparative effectiveness research to curtail the use of unnecessary technology, electronic medical records to avoid duplication of tests, payment for performance, accountable care organizations that change payment incentives—all hold promise to improve quality, but their potential for substantial cost reduction is only theoretical at present.

Recently, however, the rate of increase in health care expenditures has slowed relative to the gross domestic product.[7] Two basic hypotheses have been offered: the recession that began in 2008, and heightened cost consciousness among hospitals, health insurers, and some physician groups.

Payment for medical care varies by country. In the United States, health insurance coverage is an incomplete patchwork, consisting of government-sponsored programs for elderly people (Medicare), poor people (Medicaid), and veterans, plus employer-based coverage for workers and their families. Medicare covers acute care services in the hospital and in physicians' offices but has limited coverage for prescription drugs and long-term care. More than half of all Medicare subscribers also buy supplemental insurance. Medicaid covers more services than Medicare does, but Medicaid payments to physicians and hospitals are so low in many states that patients have restricted access to care. At any given time, more than 44 million Americans have lacked health insurance, and 70 million have been without insurance at some point during the year. In addition, millions of immigrant workers are also uninsured. The lack of health insurance contributes to poor health, such as delayed diagnosis and undertreatment of asthma, diabetes, hypertension, and cancer.

The 2010 Patient Protection and Affordable Care Act (ACA) contains numerous insurance reform features that took effect in 2010 and 2011, as well as coverage expansions that began in 2014.[8] The ACA was originally expected to cover 32 million previously uninsured Americans, with about half enrolling in subsidized private insurance plans and half in expanded state Medicaid programs. However, a 2012 Supreme Court decision gave states the choice of opting out of the Medicaid expansion. As a result, only 27 states plus the District of Columbia accepted that expansion. States that opted out of Medicaid expansion, such as Texas, tend to be those with the highest proportion of uninsured—mainly poor—people. In addition, various coverage components, especially regarding contraception, continue to be litigated. Revenue-generating provisions of the ACA are split about evenly between spending reductions and cost containment. In contrast to what happened after the passage of Medicare and Medicaid, the ACA continues to be highly controversial politically, and thus subject to potential changes, depending on election results.

Because medical care is both so valued and so expensive, physicians everywhere will inevitably become more involved in issues of medical economics. As cost-containment pressures force patients to assume more of their medical expenses, patients will become more aware of costs and more demanding about the price and value of care. In addition, knowledge will continue to accumulate about the real and potential harm from unnecessary or marginally useful medical services. Thus, informed clinical decision making will require that physicians have accurate information about the risks, benefits, and costs of medical care and better ways to communicate what is known and what is not.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Centers for Disease Control and Prevention. Current cigarette smoking among adults: United States, 2011. *MMWR Morb Mortal Wkly Rep*. 2012;61:889-894.
2. Jha P, Ramasundarahettige C, Landsman V, et al. 21st Century hazards of smoking and benefits of cessation in the United States. *N Engl J Med*. 2013;368:341-350.
3. Gottlieb L, Sandel M, Adler NE. Collecting and applying data on social determinants of health in health care settings. *JAMA Intern Med*. 2013;173:1017-1020.
4. Lorenzoni L, Belloni A, Sassi F. Health-care expenditure and health policy in the USA versus other high-spending OECD countries. *Lancet*. 2014;384:83-92.
5. Schroeder SA. Personal reflections on the high cost of American medical care. *Arch Intern Med*. 2011;171:722-727.
6. Schroeder SA, Frist W. Phasing out fee-for-service payment. *N Engl J Med*. 2013;368:2929-2932.
7. Fuchs VR. The gross domestic product and health care spending. *N Engl J Med*. 2013;369:107-109.
8. Shaw FE, Asomugha CN, Conway PH, et al. The Patient Protection and Affordable Care Act: opportunities for prevention and public health. *Lancet*. 2014;384:75-82.

REVIEW QUESTIONS

1. Preventing premature mortality is a prime goal for all clinicians. Of the following statements regarding premature mortality, which one is incorrect?

A. About 10% of premature deaths could be prevented by assuring high-quality medical care to all.
B. Among the various general causes of premature deaths, behavioral factors such as risky sexual behavior, alcohol and drug abuse, smoking, and obesity and physical inactivity are the most important and also offer opportunities for remediation.
C. Among behavioral factors, the most important cause of premature death is smoking cigarettes.
D. As smoking prevalence gradually decreases, the remaining smokers are smoking more cigarettes per day.
E. Based on current trends, obesity and physical inactivity will likely become more important causes of premature deaths.

Answer: D Statement A is correct. As important as good medical care is, the lack of such services only accounts for about 10% of premature deaths. Statements B, C, and E are correct. As regards statement D, at the same time as overall smoking prevalence has declined, the number of daily cigarettes consumed by those who continue to smoke has also declined. This decline is probably of function of several factors, including rising tobacco taxation that makes smoking more expensive, the spread of clean indoor air laws, and the increasing stigma attached to smoking.

2. Increasing evidence points to the important health impacts of social economic status (SES). Which one of the following statements regarding health and social class is correct?

A. Among the various components of SES, racial and ethnic characteristics contribute the most to health disparities.
B. Virtually all of the worst health among low SES populations can be explained by personal behaviors such as cigarette smoking and drug and alcohol abuse.
C. The relationship between SES and health is concentrated among those with low SES. In other words, for people above 400% of the poverty level, SES does not contribute to health status.
D. Because so many of the determinants of low SES (e.g., education, income, housing, net wealth) are outside the purview of clinicians, they should not be concerned about them.
E. Co-location of access to important social services—such as food stamps or disability payments—at medical sites can improve the social status of selected patients.

Answer: E Answer A is incorrect. Class, as measured by income, net wealth, educational status, and neighborhood, is the most important factor leading to health disparities. Regarding answer B, although personal behaviors such as smoking and physical activity are important, many other factors associated with low social class contribute to poor health. Regarding answer C, there is a stepwise association between SES and health, even at the higher levels of SES. For example, those in the top decile of SES enjoy better health than those in the ninth decile, even though both deciles have high SES. For answer D, it is true that clinicians would have a hard time improving these factors. Nevertheless, they should be conscious of them because they often influence treatment strategies (e.g., the ability to obtain nutritious food or to buy medications). Finally, answer E is correct. For example, convenient provision of food stamps for those with food insecurity would help stabilize diabetic patients.

3. Regarding per capita medical expenditures, population health, and access to medical care, which statement best expresses the performance of the United States versus other developed nations?

A. It leads the world in medical expenditures but trails badly in health outcomes and access to health care services.
B. It trails the world in medical expenditures, health outcomes, and access to care.
C. It is about in the middle for expenditures, health, and access.
D. It leads the world in expenditures, health, and access.
E. It leads the world in expenditures and health but lags in access.

Answer: A Answer A is correct as written. Of the three statements in answer B, only the second is correct. Of the three answers in C, none are correct. Of those in D, only the first is correct. And of those in E, only the first is correct.

4. Regarding the causes of rising medical expenditures in the United States, which answer is correct?

A. Fee-for-service payment to physicians is no longer a major driver of cost escalation because it will soon be replaced by bundled or capitated payment.
B. The United States is an outlier in that it has more hospital beds per capita and a higher length of stay.
C. The United States is an outlier in that it features more physicians per capita.
D. There are multiple factors responsible for the patterns of expenditures in the United States, and it is unlikely that any one bears major responsibility for the high expenditure profile.
E. The threat and reality of malpractice suits, in a country with the highest number of lawyers per capita, is the major reason for high U.S. medical expenditures.

Answer: D For answer A, although there is much current talk about the imminent demise of fee-for-service payment, it remains the dominant way of paying physicians. Regarding answer B, the United States actually has fewer hospital beds per capita than other developed countries as well as a shorter length of stay. The cost of a day in the hospital, however, is far higher in the United States. Similarly, for answer C, the United States has fewer physicians per capita than most developed nations but has a much higher proportion of specialists. Answer D, the correct answer, reflects the multiple reasons that medical care is more costly in the United States. Regarding answer E, malpractice suits are indeed more of a factor in the United States than other nations. Nevertheless, if all malpractice costs were to vanish, the United States would still have the most expensive health care system by far.

5. Regarding responses to high medical expenditures in the United States, which of the following statements is most correct?

A. Because expenditures on medical care are so essential and because wealthy countries characteristically spend more on health care as wealth increases, there is little interest in curtailing rising medical expenditures.
B. The cost-containment measures found in the Patient Protection and Affordable Care Act will be sufficient to rein in rising medical costs.
C. The combination of pressures on personal and governmental health care spending and crowding out of other social expenditures makes it likely that more intense cost-containment activity will occur.
D. The combination of widespread electronic medical record use and the spread of Accountable Care Organizations will be sufficient to curb rising medical expenditures.
E. Political obstacles to cost containment, such as the fear of rationing, will not be a problem.

Answer: C Answer A is incorrect. There is emerging consensus that something must be done to curtail rising medical expenditures (as phrased in answer C), but not on how to accomplish that. Although there are some cost-containment features in the Patient Protection and Affordable Care Act (PPACA) (answer B), it is unproved whether they will reduce medical expenditures. Answer C is correct. Regarding answer D, there is no good evidence that either the spread of electronic medical records or Accountable Care Organizations will curb rising medical expenditures, despite the enthusiasm of advocates for those programs. Finally, political obstacles (answer E) remain potent barriers to medical cost containment, such as the assertion during the debate over the PPACA that access to palliative care services would be tantamount to creating death panels.

6

GLOBAL HEALTH

ARUN CHOCKALINGAM

Health is a human right, but more than 2 billion people live with a daily income of less than $2 and have no access to good health care. Health is determined by the context of people's lives. Individuals are unable to control many of the social determinants of health (Chapter 5), such as income and social status, education, physical environment, social support network, genetics, health services, and gender.[1]

In the process of modernization from a less developed to a more developed nation, the epidemiologic transition of modern sanitation, medications, and health care has drastically reduced infant and maternal mortality rates and extended average life expectancy. As a result, the world has progressed from the age of pestilence and famine, with a life expectancy between 20 and 40 years, to the age of receding pandemics, with a life expectancy of 30 to 50 years, and now to the current age of degenerative and man-made diseases, with a life expectancy of 60 years or more.

These trends, coupled with subsequent declines in fertility rates, have driven a demographic transition in which the major causes of death change from infectious diseases to chronic and degenerative diseases.[2] As many countries around the world have undergone globalization, owing to their internal urbanization, modernization, and economic development, an increased proportion of their burden of morbidity and mortality is now due to chronic noncommunicable diseases, including cardiovascular, cerebrovascular, and renovascular diseases as well as cancer, diabetes, chronic respiratory diseases, and mental disorders (Table 6-1).

WHAT IS GLOBAL HEALTH?

The term *global health* is sometimes confused with public health, international health, tropical medicine, and population health. Global health, which is defined as the health of populations in a global context, transcends the perspectives and concerns of individual nations and crosses national borders. Global health depends on the public health efforts and institutions of all countries, including their strategies for improving health, both population-wide and for individuals. Global health depends on multiple factors, including social, political, environmental, and economic determinants of health. Although global health often focuses on improving the health of people who live in low- and middle-income countries, it also includes the health of any marginalized population in any country.

Global health requires use of a wide range of institutions that collaborate in addressing all health issues. Global health also depends on the constructive use of evidence-based information to provide health and health equity, in part by strengthening primary health care and the health care delivery system.

Millennium Development Goals

In an attempt to address global inequity, the United Nations advanced eight millennium development goals with the objective of achieving these goals between 2000 and 2015. These eight goals incorporate 21 targets (Table 6-2), with a series of measurable health and economic indicators for each target.[3] Although many of the targets have not yet been achieved, substantial progress has been made toward all targets.

The millennium development goals emphasize that health and development are interconnected. To address global inequity, fundamental issues include reducing poverty, improving education, and empowering people. In addition to specific goals for reducing infant and child mortality, maternal mortality, and mortality due to infectious diseases such as human immunodeficiency virus infection/acquired immunodeficiency syndrome (HIV/AIDS), malaria, and tuberculosis, the millennium development goals strongly encourage environmental sustainability and global partnership.

GLOBAL BURDEN OF DISEASES

The global burden of disease is measured in terms of total and cause-specific mortality and morbidity as well as the national economic burden for health care. The Global Burden of Diseases, Injuries and Risk Factors Study 2010[4] shows that an estimated 53 million people died from all causes in 2010, with

TABLE 6-1 EPIDEMIOLOGIC TRANSITION IN CARDIOVASCULAR DISEASES

STAGES OF DEVELOPMENT	LIFE EXPECTANCY	BURDEN OF CARDIOVASCULAR DISEASE DEATHS, % OF TOTAL DEATHS	PREDOMINANT CARDIOVASCULAR DISEASES AND RISK FACTORS	MODERN REGIONAL EXAMPLES
1. Age of pestilence and famine	20-40 years	5-10	Infections, rheumatic heart disease, and nutritional cardiomyopathies	Rural India, sub-Saharan Africa, South America
2. Age of receding pandemics	30-50 years	10-35	As above plus hypertensive heart disease and hemorrhagic strokes	China
3. Age of degenerative and man-made diseases	50->60 years	35-65	All forms of strokes; ischemic heart disease at young ages; increasing obesity and diabetes	Aboriginal communities, urban India, former socialist economies
3A. Age of delayed degenerative diseases	>60 years	<50	Stroke and ischemic heart disease at old age	Western Europe, North America, Australia, New Zealand
3B. Age of health regression and social upheaval	50-60 years	35-55	Re-emergence of deaths from rheumatic heart disease, infections, increased alcoholism and violence; increase in ischemic and hypertensive diseases in the young	Russia

During stages 1 to 3A, life expectancy increases, whereas life expectancy decreases in stage 3B compared with stage 3A and even stage 3.
Modified from Omran AR. The epidemiological transition: a theory of the epidemiology of population change. *The Milbank Quarterly*. 2005;83:731-757. Reprinted from *The Milbank Memorial Fund Quarterly*. 1971;49:509-538; and Yusuf S, Reddy S, Ounpuu S, et al. Global burden of cardiovascular diseases: part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization. *Circulation*. 2001;104:2746-2753.

TABLE 6-2 MILLENNIUM DEVELOPMENT GOALS AND TARGETS (2000-2015)

GOAL 1: ERADICATE EXTREME POVERTY AND HUNGER
Target 1A: Halve, between 1990 and 2015, the proportion of people living on less than $1.25 a day. Target 1B: Achieve decent employment for women, men, and young people. Target 1C: Halve, between 1990 and 2015, the proportion of people who suffer from hunger.
GOAL 2: ACHIEVE UNIVERSAL PRIMARY EDUCATION
Target 2A: By 2015, all children (girls and boys) can complete a full course of primary schooling.
GOAL 3: PROMOTE GENDER EQUALITY AND EMPOWER WOMEN
Target 3A: Eliminate gender disparity in primary and secondary education preferably by 2005, and at all levels by 2015.
GOAL 4: REDUCE CHILD MORTALITY RATES
Target 4A: Reduce by two thirds, between 1990 and 2015, the under-five mortality rate.
GOAL 5: IMPROVE MATERNAL HEALTH
Target 5A: Reduce by three quarters, between 1990 and 2015, the maternal mortality ratio. Target 5B: Achieve, by 2015, universal access to reproductive health.
GOAL 6: COMBAT HIV/AIDS, MALARIA, AND OTHER DISEASES
Target 6A: Have halted by 2015 and begun to reverse the spread of HIV/AIDS. Target 6B: Achieve, by 2010, universal access to treatment for HIV/AIDS for all those who need it. Target 6C: Have halted by 2015 and begun to reverse the incidence of malaria and other major diseases.
GOAL 7: ENSURE ENVIRONMENTAL SUSTAINABILITY
Target 7A: Integrate the principles of sustainable development into country policies and programs; reverse loss of environmental resources. Target 7B: Reduce biodiversity loss, achieving, by 2010, a significant reduction in the rate of loss. Target 7C: Halve, by 2015, the proportion of the population without sustainable access to safe drinking water and basic sanitation. Target 7D: By 2020, to have achieved a significant improvement in the lives of at least 100 million slum-dwellers.
GOAL 8: DEVELOP A GLOBAL PARTNERSHIP FOR DEVELOPMENT
Target 8A: Develop further an open, rule-based, predictable, non-discriminatory trading and financial system. Target 8B: Address the special needs of the least developed countries. Target 8C: Address the special needs of landlocked developing countries and Small Island developing States. Target 8D: Deal comprehensively with the debt problems of developing countries through national and international measures in order to make debt sustainable in the long term. Target 8E: In cooperation with pharmaceutical companies, provide access to affordable, essential drugs in developing countries. Target 8F: In cooperation with the private sector, make available the benefits of new technologies, especially information and communications.

From United Nations Millennium Development Goals. http://www.un.org/millenniumgoals/poverty.shtml. 2008. Accessed January 21, 2015.

13.2 million (25%) deaths due to communicable, maternal, neonatal, and nutritional disorders; 34.5 million (65%) due to noncommunicable diseases; and 5.1 million (10%) due to injuries (Table 6-3).[5] Although overall deaths between 1990 and 2010 increased by 13.5%, medical and public health advancements reduced deaths from communicable diseases by 17%, whereas deaths due to noncommunicable disease increased by 30% and deaths due to injury, including war-related deaths, increased by 24%.

CHANGING PATTERNS OF DISEASES

Despite the general trends of declining morbidity and mortality from communicable diseases, parts of Africa, Asia, and Latin America are still facing the challenges of infectious diseases, such as HIV infection, malaria, and tuberculosis, even as their prevalence of chronic noncommunicable diseases has risen—a so-called double burden. Concerted global health efforts and public awareness as well as investments by industrialized countries, multilateral agencies, and nongovernmental organizations have resulted in significant progress against HIV/AIDS (Chapter 384). Despite all of these efforts, however, the worldwide mortality due to HIV/AIDS and tuberculosis rose by 50% in 2010 compared with 1990 (Table 6-3). Although malaria deaths have fallen worldwide during the last decade, malaria is a rising threat in parts of Southeast Asia—especially Cambodia, Myanmar, Thailand, and Vietnam—where drug resistance to antimalaria medications is a problem.

TABLE 6-3 GLOBAL DEATHS IN 1990 AND 2010 FOR ALL AGES AND BOTH SEXES COMBINED

CAUSES OF DEATH	ALL AGES—DEATHS (THOUSANDS) 1990	2010	% CHANGE
All causes	46,511	52,769	+13
Communicable, maternal, neonatal, and nutritional disorders	15,859	13,156	−17
HIV/AIDS and tuberculosis	1770	2661	+50
Diarrhea, lower respiratory infection, and other common IDs	7772	5277	−32
Neglected tropical diseases and malaria	1211	1322	+ 9
Maternal disorders	359	255	−29
Neonatal disorders	3081	2236	−42
Nutritional deficiencies	977	684	−30
Other communicable, maternal, neonatal, and nutritional disorders	690	721	+5
Noncommunicable diseases	26,560	34,540	+30
Neoplasm	5779	7978	+38
Cardiovascular and circulatory diseases	11,903	15,616	+31
Chronic respiratory diseases	3986	3776	−5
Cirrhosis of the liver	778	1031	+33
Digestive diseases (except cirrhosis)	973	1112	+14
Neurologic disorders	595	1274	+14
Mental and behavioral disorders	138	231	+68
Diabetes, urogenital, blood, and endocrine diseases	1544	2726	+77
Musculoskeletal disorders	70	154	+121
Other noncommunicable diseases	794	642	−19
Injuries	4092	5073	+24
Transport injuries	958	1397	+46
Unintentional injuries other than transport injuries	2030	2123	+5
Self-harm and interpersonal violence	1009	1340	+33
Forces of nature, war, and legal intervention	95	214	+125

HIV/AIDS = human immunodeficiency virus infection/acquired immunodeficiency syndrome; ID = infectious diseases.
From Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2095-2128.

Noncommunicable diseases account for nearly two thirds of the global burden of disease. Nearly 80% of all noncommunicable diseases related to death and disability occur in the low- and middle-income countries, where they account for about 14 million deaths in people younger than 60 years. The prevention and control of noncommunicable diseases should involve both upstream and downstream approaches, such as social determinants; national and international policies regarding trade, agriculture, transportation, and environmental and other policies; health care, including accessibility, availability, and affordability; and settings, such as schools and worksites, where health promotion and disease prevention are targeted, as well as media by which health can be influenced.

Among noncommunicable diseases, cardiovascular diseases account for the largest burden of disease. One of the major worldwide risk factors for cardiovascular disease is hypertension (Chapter 67), which has an estimated prevalence of 35% to 45% of the global population—more than 2 billion people older than 25 years. The prevalence of hypertension is highest in Africa, where it is about 45% for both sexes, and lowest in the Americas, where it is about 35% for both sexes (Fig. 6-1). In all regions, men have a slightly higher prevalence of hypertension than do women. Despite significant efforts by global nongovernmental organizations and the World Health Organization, more than 50% of the world's population with hypertension does not even know their condition, and the percentage treated and controlled varies from less than 5% in Zambia to 66% in Canada.

As noncommunicable diseases have reached an epidemic proportion, all 192 United Nations Member States agreed to address their prevention and control worldwide, particularly in developing countries. The emphasis is on four major noncommunicable diseases (cardiovascular diseases, cancer, diabetes, and chronic respiratory diseases) and four key risk factors common to all four of these noncommunicable diseases (tobacco use, unhealthy diets, physical inactivity, and harmful use of alcohol). The World Health Organization developed a global monitoring framework to enable global tracking of progress in preventing and controlling these four major noncommunicable diseases and their key risk factors, aiming for a 25% reduction by 2025—with a slogan of *25 By 25*.[6]

Noncommunicable diseases represent a growing economic threat across the globe and are becoming an acute problem in low- and middle-income countries in which they are estimated to account for nearly $500 billion per year.[7] By contrast, worldwide adoption of best practices could substantially reduce that economic burden. For example, population-based interventions to reduce tobacco and harmful alcohol use as well as to improve unhealthy diets and to increase physical activity are estimated to cost less than $0.40 per person per year. These low-technology, population-wide interventions along with individual-based noncommunicable disease "best buy" interventions, such as individualized counseling and drug therapy, bring the total annual cost to $11.4 billion. Thus, on a per capita basis, the annual investment ranges from less than $1 in low-income countries to $3 in upper middle-income countries.

The growing epidemic of noncommunicable diseases, including mental health conditions, and the unfinished agenda of controlling infectious diseases (HIV/AIDS, malaria, tuberculosis, maternal and child health, and other infectious and parasitic diseases) pose a huge threat to the global population in terms of both human and fiscal losses. Although individual countries theoretically take responsibility for the health of their respective citizens, many low- and middle-income countries are unable to meet their domestic population's basic needs. The sum of public and private health care expenditure by countries based on their gross domestic product varies from 1.6% in South Sudan to 18% in the United States, with many high-income countries spending more than 10%. Thus, the worldwide solution requires a response of all of human society, including strategic domestic and international investments, both within countries and through multilateral agencies. This societal responsibility must be shared by the private sector, nongovernmental organizations, academia, professional societies, and the public themselves.

The modern global health agenda should move beyond false dichotomies—such as prevention versus treatment, infectious versus noncommunicable diseases, primary care versus specialized care, social determinants versus health services, life sciences versus social sciences—toward integration in the sense of common purpose. The approach to noncommunicable diseases should be integrated with the approach to communicable diseases. Health

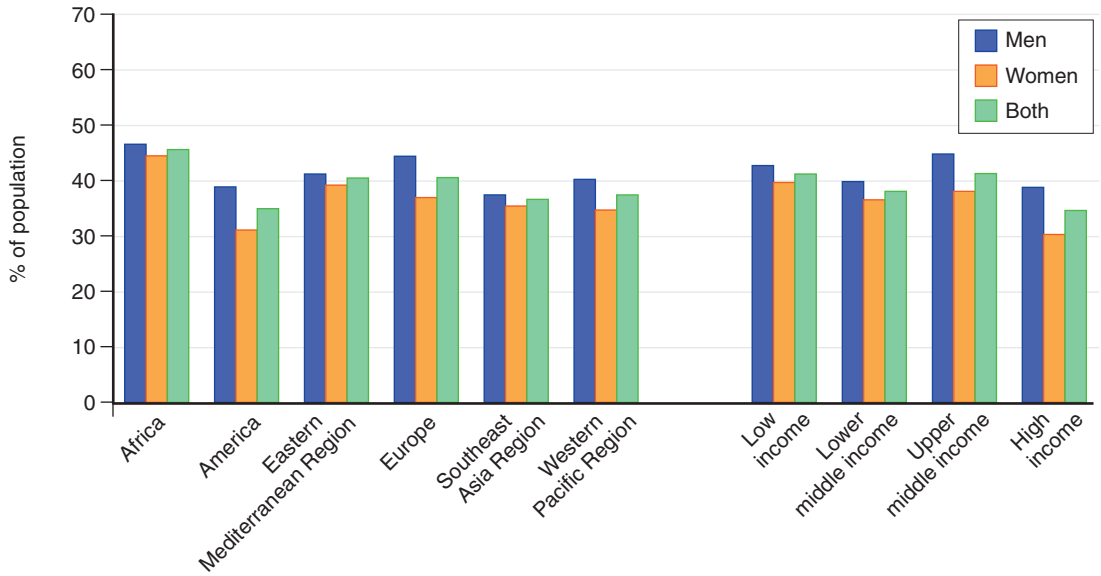

FIGURE 6-1. Global prevalence of hypertension. Age-standardized prevalence of hypertension, defined as systolic blood pressure ≥140 mm Hg or diastolic blood pressure ≥90 mm Hg or taking medications for lowering of blood pressure, in persons aged 25 years and older. (Data source: Noncommunicable Disease Global Monitoring Framework. World Health Organization. 2013. http://www.who.int/nmh/global_monitoring_framework/en/. Accessed January 21, 2015.)

targets should include noncommunicable diseases such as hypertension, diabetes, and cervical cancer, even in low- and middle-income countries. A critical component of health is a healthy lifestyle throughout life, from childhood to adolescence to adulthood, including risk factor reduction as well as treatment when it becomes necessary. Global health must integrate prevention and treatment in a cost-effective manner.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. 2008, World Health Organization. http://whqlibdoc.who.int/publications/2008/9789241563703_eng.pdf. Accessed January 21, 2015.
2. Beaglehole R, Bonita R. What is global health? *Global Health Action*. 2012;3:5142.
3. United Nations Millennium Development Goals. http://www.un.org/millenniumgoals/poverty.shtml. 2008. Accessed January 21, 2015.
4. Lim SS, Vos T, Flaxman AD, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2224-2260.
5. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2095-2128.
6. Noncommunicable Disease Global Monitoring Framework. 2013, World Health Organization. http://www.who.int/nmh/global_monitoring_framework/en/. Accessed January 21, 2015.
7. From burden to "best buys": reducing the economic impact of non-communicable diseases in low- and middle-income countries. 2011, World Health Organization and World Economic Forum. http://www.who.int/nmh/publications/best_buys_summary.pdf. Accessed January 21, 2015.

REVIEW QUESTIONS

1. Based on the Global Burden of Disease Study 2010, which one of the following statements is correct?

A. The global burden of diabetes mellitus remained stable between 1990 and 2010.
B. Cancer is the leading cause of noncommunicable disease-related deaths.
C. Cardiovascular and circulatory deaths accounted for almost 30% of all deaths in 2010.
D. The total number of deaths due to HIV/AIDS and tuberculosis decreased in 2010 compared with 1990.
E. Forces of nature, war, and legal intervention decreased in 2010 compared with 1990.

Answer: C See Table 6-3. In 2010, the number of cardiovascular-related deaths was 15.6 million out of the total deaths of 52.8 million, amounting to 29.6%. The number of diabetes-related deaths in 1990 and 2010 were 1.5 million and 2.7 million, respectively. During those 20 years, diabetes mellitus increased by 76%. Cancer was the second leading cause of deaths in 2010. There were 8.0 million deaths due to cancer in 2010 compared with 15.6 million deaths from cardiovascular and circulatory causes. The number of deaths due to HIV/AIDS and tuberculosis in 2010 (2.7 million) increased by 50% compared with the 1990 number (1.8 million). Compared with 1990, deaths resulting from forces of nature, war, and legal intervention more than doubled (125%) by 2010.

2. Hypertension or high blood pressure is the major risk factor for heart disease, stroke, and kidney diseases worldwide. Which one of the following statements is incorrect?

A. The prevalence of hypertension, globally, after the age of 25 years varies between 35 and 45%.
B. Across all countries, men have a slightly higher prevalence of hypertension than do women.
C. The global hypertension control rate is about 60%.
D. Early diagnosis of hypertension leads to prevention of all forms of vascular complications.
E. According to the World Health Organization, population-attributable deaths due to hypertension are estimated at about 7.5 million per year.

Answer: C The global control rate varies from less than 5% in Zambia to 66% in Canada and is very low overall worldwide. All other statements (answers A, B, D, and E) are correct.

3. Which of the following statement about global health is correct?

A. Global health is not the opposite of domestic health.
B. Global health must integrate both infectious diseases and noncommunicable diseases.
C. Public, private, and societal partnership is necessary to deliver effective global health.
D. Academia has a major role in promoting global health.
E. All of the above.

Answer: E Statement A is true because global health includes domestic health as well, particularly the health of marginalized people in developed countries. Statement B is true because the current health system must address all diseases, particularly when noncommunicable diseases account for 65% of all global causes of death. New evidence shows that one fifth of all cancers worldwide are caused by chronic infections produced by agents such as HIV, human papillomavirus, and hepatitis B virus. Infections and parasitic diseases also cause other noncommunicable diseases, such as rheumatic heart disease, Chagas disease, cardiomyopathy, and peptic ulcer. As HIV/AIDS survivors live longer, they also are exposed to lifestyle-related risk factors and noncommunicable diseases. Statement C is true. The economic burden of diseases is so large that public-private partnership is essential. Statement D is true. Academia should develop and supply needed knowledge and train the next generation of the global health work force.

7

APPROACH TO THE PATIENT: HISTORY AND PHYSICAL EXAMINATION

DAVID L. SIMEL

OVERVIEW

Physicians may have multiple objectives with varying degrees of importance in their encounters with patients. These goals include, but are not limited to, the translation of symptoms and signs into diagnoses, the assessment of stability or change in known conditions, the provision of information and counseling for future prevention, and the reaffirmation or alteration of therapeutic interventions. A general health check will increase the number of diagnoses for a patient, but it may not affect overall morbidity and mortality.[1]

The interaction between the patient and physician represents not only a scientific encounter but also a social ritual centered on locus of control and meeting each other's expectations. Patients expect that their health care needs and concerns will be addressed competently. Physicians also have expectations: a need to feel that they have not missed something important in addressing diagnostic challenges, a need to put limits on the time available for each interaction, and a need to maintain objectivity so that their evaluation and recommendations are not clouded by their emotional feelings about the patient. The expertly performed rational clinical examination enhances the expected social ritual and the likelihood of acquiring relevant data. It also optimizes the physician's ability to understand the patient's symptoms and concerns, as well as to facilitate the healing process.[2]

Physical Examination Begins with the History

It is almost impossible to consider the history as distinct from the physical examination because the clinical examination begins as soon as the physician sees or hears the patient. From the patient's perspective, it matters little whether the physician determines the diagnosis from the history or the physical examination. The concept of deliberate practice, which has been demonstrated to be important in performing arts and in sports, can be applied to the performance of the clinical examination. This deliberate practice, which must be specifically designed to improve the ability to elicit symptoms and signs, requires appropriate training under observation, concentration, and commitment to ensure its consistent and accurate performance. The creation of simulation centers at many medical schools and hospitals provides new opportunities for this training.

Quantitative Principles of the Clinical Examination

The diagnostic accuracy of a symptom or sign is quantified by its sensitivity and specificity, often described as its likelihood ratio (LR) or predictive value (Chapter 10). Data on diagnostic accuracy can be obtained by a literature search for the evaluation of a disease-specific condition (e.g., melanoma) or a clinical finding (e.g., splenomegaly) (E-Tables 7-1 and 7-2). Each component of the history and physical examination has an associated sensitivity (the percentage of patients with a disorder who have an abnormal finding) and specificity (the percentage of patients without a disorder who have a normal finding). Measures of precision, such as the kappa (κ) statistic (0 = random agreement; +1 = perfect agreement), quantify the agreement between observers on a symptom or sign (Chapter 10). Current research on the clinical examination uses LRs that inform clinicians how likely they are to observe a particular finding in a patient with a given condition as opposed to a patient without the condition. A patient with an abnormal glabellar tap has an LR of 4.5 for Parkinson disease (Chapter 409), which means that the risk for Parkinsonism increases 4.5-fold compared with that in a patient who does not have the finding. Similarly, a patient who insists that he or she does not have "shaking in the arms" has an LR of 0.25 for Parkinson disease and is one fourth as likely (a reduced chance) to have the disease compared with the baseline risk.

MEDICAL HISTORY

The history begins by asking patients to describe, in their own words, the reason for seeking medical care (Table 7-1). Although patients may have many reasons for initiating a visit to the physician, they should be encouraged to select the one or two most important concerns they have. The physician should reassure the patient that other concerns will not be ignored, but that it is important to understand what is most concerning to the patient.

History of the Present Illness

Open-ended questions facilitate descriptions of problems in the patient's own words. Subsequently, specific questions fill in gaps and help clarify important points. These questions should be asked in an order dictated by the story the patient tells and targeted to suit the individual problem. When the patient is acutely ill, the physician should limit the amount of time spent in open-ended discussion and move promptly to the most important features that allow quick evaluation and management. In general, the history of the problem under consideration includes the following:

- Description of onset and chronology
- Location of symptoms
- Character (quality) of symptoms
- Intensity
- Precipitating, aggravating, and relieving factors
- Inquiry into whether the problem or similar problems occurred before and, if so, whether a diagnosis was established at that time

It is often helpful to ask patients to express what they believe is the cause of the problem or what concerns them the most. This approach often uncovers other pertinent factors and helps establish that the physician is trying to meet the patient's needs.

Past Medical and Surgical History

An astute clinician recognizes that patients may not report all their prior problems because they may forget, may assume that previous events are unrelated to their current problem, or simply may not want to discuss past events. Open-ended statements such as "Tell me about other medical illnesses that we did not discuss" and "Tell me about any operations you had" prompt the patient to consider other items. The physician should ask the patient about unexplained surgical or traumatic scars.

A list of current medications includes prescriptions, over-the-counter medications, vitamins, and herbal preparations. Patients who do not recall the names of medications should bring all medication bottles to the next visit. Patients may not consider topical medications (e.g., skin preparations or eyedrops) as important, so they may need prompting.

Information about allergies (Chapter 254) is particularly important, but challenging, to collect. Patients may attribute adverse reactions or intolerances to allergies, but many supposed allergic reactions are not truly drug allergies. For example, less than 20% of patients who claim a penicillin allergy are allergic on skin testing. Eliciting the patient's actual response to medications facilitates a determination of whether the response was a true allergic reaction.

Social and Occupational History and Risk Factors

The social history not only reveals important information but also improves understanding of the patient's unique values, support systems, and social situation. It can be helpful to ask the patient to describe what they would do during a typical day. Patients who lack confidence when filling out medical forms by themselves have an LR positive of 5.0 for problems with health care literacy.[3]

Data that may influence risk factors for disease should be gathered, including a nonjudgmental assessment of substance abuse. The tobacco history should include the use of snuff, chewing tobacco, and cigar and cigarette smoking (Chapter 32). Alcohol use should be determined quantitatively and by the effect that it has had on the patient's life (Chapter 33). Past or present use of illicit substances, prescription pain medications or sedatives, and intravenous drugs should be assessed (Chapter 34). The sexual history should address sexual orientation and gender identity, as well as current and past sexual activity. The physician can initiate this discussion with the question, "Do you have any concerns or questions about your sexuality, sexual orientation, or sexual desires?"[4] The employment history should include the current and past employment history, as well as any significant hobbies. All adult patients should be asked if they served in the military. Military veterans should be asked about their combat history, years of service, and areas of deployment (Table 7-2).[5]

The physician should also obtain information on socioeconomic status, insurance, the ability to afford or obtain medications, and past or current barriers to health care because of their impact on care of the patient (Chapter

E-TABLE 7-1 MEDLINE SEARCH STRATEGY FOR IDENTIFYING QUANTITATIVE INFORMATION ON THE CLINICAL EXAMINATION USING THE OVID SEARCH SYSTEM* TO SEARCH FOR STUDIES ON MELANOMA

1. exp physical examination/or physical exam$.mp
2. medical history taking.mp
3. professional competence.mp
4. (sensitivity and specificity).mp or (sensitivity and specificity).tw
5. (reproducibility of results or observer variation).mp
6. diagnostic tests, routine/
7. (decision support techniques or Bayes theorem).mp
8. 1 or 2 or 3 or 4 or 5 or 6 or 7
9. limit 8 to (Ovid full text available and human and English language)
10. exp melanoma
11. 9 and 10

*OVID Technologies, Inc. A condition and a physical finding are given as examples. Abbreviations or search term abbreviations are as follows: "exp" indicates that the topic is "exploded" to include all subheadings for the topic. The "$" is a wildcard designator, so "exam$" would include the words *examination, examining*, and *examiner.* "mp" searches for the word or phrase in the title, abstract, registry number word, or mesh subject heading. Step 9 limits the search to studies that involve humans only and for which the full manuscript is available online and is written in English. If the search yields too few topics, the limitation of full text available can be removed and the search repeated. If too many results are obtained, some of the items from step 8 can be eliminated.

E-TABLE 7-2 PUBMED SEARCH STRATEGY FOR IDENTIFYING QUANTITATIVE INFORMATION ON THE CLINICAL EXAMINATION TO SEARCH FOR STUDIES ON SPLENOMEGALY

(splenomegaly[mh] AND (Humans[Mesh] AND English[lang])) AND ((physical examination[mh] OR physical exam*[title/abstract]) OR (medical history taking[mh]) OR (professional competence[mh]) OR ((sensitivity[title/abstract] AND specificity[title/abstract]) OR (sensitivity and specificity[mh])) OR (reproducibility of results[mh]) OR (observer variation[mh]) OR (diagnostic tests, routine[mh]) OR (decision support techniques[mh]) OR (bayes theorem[mh]))

The PUBMED search strategy may be too sensitive, with many articles found. Strategies to eliminate case reports and to constrain the search to full text availability will reduce high-yield results to a more manageable list.

TABLE 7-1 PATIENT'S MEDICAL HISTORY

Description of the patient
Age, gender, ethnic background, occupation
Chief reason for seeking medical care
State the purpose of the evaluation (usually in the patient's words)
Other physicians involved in the patient's care
Include the clinician that the patient identifies as his or her primary provider or the physician who referred the patient. Record contact information for all physicians who should receive information about the visit
History of the reason for seeking medical care
In chronologic fashion, determine the evolution of the indication for the visit and then each major symptom. It is best to address the patient's reason for seeking care first rather than what the physician ultimately believes is most important
Be careful to avoid "premature closure," in which a diagnosis is assumed before all the information is collected
Past medical and surgical history
List other illnesses and previous surgeries not related to the current problem
List all prescribed and over-the-counter medications with dose
Remember to ask about vitamin and herbal supplements
Allergies and adverse reactions
List allergic reactions to medications and food. Record the specific reaction (e.g., hives). Distinguish allergies from adverse reactions or intolerance to medication (e.g., dyspepsia from nonsteroidal anti-inflammatory agents)
Social, occupational, and military history (see Table 7-2)
Describe the patient's current family and a typical day for patient. The occupational history should focus on current and past employment as it might relate to the current problem
Risk factors
Include history of tobacco use, illegal drug use, and risk factors for sexually transmitted disease (including human immunodeficiency virus and hepatitis)
Family history
History of any diseases in first-degree relatives and a listing of family members with any conditions that could be risk factors for the patient (e.g., cardiovascular disease at a young age, malignancy, known genetic disorders, longevity)
Review of systems (see Table 7-3)

TABLE 7-2 BASIC MILITARY HISTORY

- Tell me about your military experience.
- When and where do you/did you serve?
- What do you/did you do while in the service?
- How has military service affected you?

If the patient answers "yes" to any of the questions below, ask the patient, "Can you tell me more about that?"

- Did you see combat, enemy fire, or casualties?
- Were you or a buddy wounded, injured, or hospitalized?
- Did you ever become ill while you were in the service?
- Were you a prisoner of war?

To screen for post-traumatic stress disorder, ask, "In your life, have you ever had an experience so horrible, frightening, or upsetting that, in the past month you …

- Have had nightmares about it or thought about it when you did not want to?"
- Tried hard not to think about it or went out of your way to avoid situations that reminded you of it?"
- Were constantly on guard, watchful, or easily startled?"
- Felt numb or detached from others, activities, or your surroundings?"

From Department of Veterans Affairs. Military Health History Pocket Card for Clinicians. http://www.va.gov/oaa/pocketcard/military-health-history-card-for-print.pdf. Accessed February 9, 2015.

5). Marital status and the living situation (i.e., whom the patient lives with, significant stressors for that patient) are important as risk factors for disease and to determine how best to care for the patient. A patient's culture (Chapter 4) and values should be known, including any prior advance directives or desire to overrule them (Chapter 3). The physician should explicitly elicit and record information regarding the next of kin; surrogate decision makers; emergency contacts; social support systems; and financial, emotional, and physical support available to the patient.

Family History

The family history is never diagnostic, but it allows risk stratification, which affects the pretest probability for an increasing number of disorders (e.g., heart disease, breast cancer, or Alzheimer disease). For common diseases such as heart disease, additional inquiry into the age of onset in first-degree relatives and death attributed to the disease should be obtained (Chapter 52). When a patient reports that a first-degree relative had a myocardial infarction, the LR is 19 that the patient has a family history of myocardial infarction. Patients may lack appropriate information about the absence of disease, however, so a reported lack of a family history of myocardial infarction reduces the likelihood only by one third. In general, the specificity of the reported family history far exceeds its sensitivity; for example, only two thirds of patients with essential tremor (Chapter 410) report a family history, but 95% of such patients have first-degree relatives with tremor. The expansion of knowledge about genetic diseases (Chapter 40) requires clinicians not only to improve their skills in eliciting the family history but also to develop methods for confirming the information. For example, patients who report that a first-degree relative had carcinoma of the colon (LR 25), breast (LR 14), ovaries (LR 34), or prostate (LR 12) are usually providing accurate information.[6]

TABLE 7-3 REVIEW OF SYSTEMS*

FOCUS all questions on a specific time frame (e.g., within the past "month" or "now") and on items not already addressed during the clinical examination:
Change in weight or appetite
Change in vision
Change in hearing
New or changing skin lesions
Chest discomfort or sensation of skipped beats
Shortness of breath, dyspnea on exertion
Abdominal discomfort, constipation, melena, hematochezia, diarrhea
Difficulty with urination
Change in menses
Joint or muscle discomfort not already mentioned
Problems with sleep
Difficulty with sexual function
Exposure to "street" drugs or medications not already mentioned
Depression (feeling "down, depressed, or hopeless"; loss of interest or pleasure in doing things)
A sensation of unsteadiness when walking, standing, or getting up from a chair

*Clinicians may start with this basic list and adapt the items to their specific patient population by considering factors such as age, gender, medications, and the problems identified during the examination. The process is facilitated by developing a routine personal approach to these questions, typically going through the systems from "head to toe."

Review of Systems

The review of systems, which is the structural assessment of each of the major organ systems, elicits symptoms or signs that are not covered or may be overlooked in the history of the present illness (Table 7-3). Although a review of systems is required in many electronic records systems, it may only yield important diagnoses in less than 10% of patients, and the cost of pursuing false-positive findings is not known. In contrast to the open-ended nature of collecting the medical history, which allows the patient to "claim" or "deny" a variety of symptoms, the direct questioning technique of the review of systems leads the patient to "accept" or "reject" symptoms. The review of systems is more efficient if at least some questions are restricted to a specific time frame (e.g., "Has there been any recent change in your vision?" or "Have you recently had shortness of breath, wheezing, or coughing?") or by having the patient fill out a previsit questionnaire.

PHYSICAL EXAMINATION

Chaperones

Surveys suggest that most patients of either sex and all ages report a lack of preference for a chaperone, but it is not clear whether this response is their true feeling or a desire to give a "correct" response. Nevertheless, many adult women (29%) and adolescent girls (46%) do express a preference for a chaperone during a breast, pelvic, or rectal examination by a male physician (especially during their first examination). Examiners should offer patients the option of a chaperone, and a chaperone should be considered when the clinician and patient are of different genders. Many examiners prefer a chaperone to allay their own anxieties attributable to gender differences or to achieve a perceived need for protection should the patient become concerned during the procedure.

Vital Signs

Vital signs include the pulse rate and rhythm, blood pressure, respiratory rate, body temperature, and the patient's quantitative assessment of pain. Marked abnormalities require a rapid, focused evaluation that may take precedence

over the typical structural approach to the remainder of the evaluation (Chapter 8).

When the blood pressure is abnormal (Chapter 67), the measurement should be repeated, assuring that the cuff size is appropriate. Many adults require a large adult cuff; using a narrow cuff can alter systolic/diastolic blood pressure by −8 to +10/+2 to +8 mm Hg. The appearance of repetitive sounds (Korotkoff sounds, phase 1) constitutes systolic pressure. (Record the value rounded upward to the nearest 2 mm Hg.) After the cuff is inflated about 20 to 30 mm Hg above the palpated pressure, the Korotkoff sounds muffle and disappear as the pressure is released (phase 5). The level at which the sounds disappear is the diastolic pressure.

Respirations should be assessed with the patient unaware that the rate is being observed. The examiner should decide whether patients have tachypnea (a rapid rate of breathing) or hypopnea (a slow or shallow rate of breathing). Tachypnea is not always associated with hyperventilation, which is defined by increased alveolar ventilation resulting in a lower arterial carbon dioxide level (Chapter 103). The subjective sensation of dyspnea (Chapter 83) is caused by an increased work of breathing.[7]

The body temperature of adults is measured with an oral electric thermometer. Rectal thermometers reliably record temperatures 0.4° C higher than oral thermometers. Tympanic thermometers vary too much in comparison with oral thermometers (−1.2° to +1.6° C vs. the oral temperature) to be reliable in hospitalized patients.

As a vital sign measure, patients should self-rate any pain on a scale of 0 to 10 (no pain to worst pain ever) (Chapter 30). However, the validity, usefulness, and value of this approach as a screening tool for clinical diagnosis are uncertain.

Head and Neck

Face

The examiner should note any asymmetrical facial features. Examples of asymmetry include skin lesions (Chapter 436), cranial nerve palsies (Chapter 396), parotid enlargement (Chapter 425), or the ptosis of Horner syndrome (Chapter 424). A variety of disorders may cause symmetrical, abnormal facies; examples include acromegaly (Chapter 224), Cushing syndrome (Chapter 227), and Parkinson disease (Chapter 409).

Ears

Physicians may not recognize their patient's hearing impairment (Chapter 428). The inability to appreciate the whispered voice increases the likelihood of hearing loss (LR 6).[8] Otoscopic evaluation of the tympanic membranes should reveal a translucent membrane and an obvious cone of light reflected where the eardrum meets the malleolus (see Fig. 426-7). Cerumen impaction is an easily treated cause of diminished hearing.

Nose

Patients with nasal symptoms often incorrectly self-diagnose bacterial sinusitis (Chapter 426). The nares should be examined for the presence of polyps, which can be seen as obstructing, glistening mucosal masses. Transillumination performed in a dark room is useful for diagnosing sinusitis, especially when combined with visualization of a purulent discharge, a patient's report of a poor response to decongestants or antihistamines, a maxillary toothache, and the presence of discolored rhinorrhea (Chapter 426). These patients have an LR greater than 6 for rhinosinusitis.

Mouth

The quality of the patient's dentition directly affects nutrition. Premalignant oral lesions (e.g., leukoplakia [see Fig. 190-1], nodules, ulcerations) found by generalist physicians are usually verified by dentists (LR > 6.5) (Chapter 425). Patients who use smokeless tobacco products are at significantly increased risk for premalignant and malignant oral lesions (Chapter 32). Bimanual palpation of the cheeks and floor of the mouth facilitates identification of potentially malignant lesions (Chapter 425).

Eyes

The eye examination begins with simple visual inspection to look for symmetry in the lids, extraocular movements, pupil size and reaction, and the presence of redness (Chapters 423 and 424). Abnormalities in extraocular movements should be grouped into nonparalytic (usually chronic with onset in childhood) or paralytic causes (third, fourth, or sixth cranial nerve palsy). Pupillary abnormalities may be symmetrical or asymmetrical (anisocoria). Red eyes should be categorized by the pattern of ciliary injection, presence of pain, effect on vision, and papillary abnormalities. When the eye examination is approached systematically, the generalist physician can evaluate the likelihood of conjunctivitis, episcleritis or scleritis, iritis, and acute glaucoma.

Routine determination of visual acuity can confirm a patient's report of diminished vision but does not replace the need for formal ophthalmologic evaluation in patients with visual complaints (Chapter 423). Aging patients often experience acute flashes and floaters, especially with posterior vitreal detachments. If acute flashes and floaters are associated with visual loss, the patient should be urgently referred for an ophthalmologic examination for the evaluation of a possible acute retinal detachment.[9] Cataracts can be detected with direct ophthalmoscopy, but the generalist's proficiency in this evaluation is uncertain.

After identifying the optic disc by ophthalmoscopy, the examiner should note the border of the disc for clarity, color, and the size of the central cup in relation to the total diameter (usually less than half the diameter of the disc). A careful observer usually can see spontaneous venous pulsations that indicate normal intracranial pressure, but about 10% of patients with normal intracranial pressure will not have spontaneous pulsations. Abnormalities of the optic disc include optic atrophy (a white disc), papilledema (see Fig. 423-27) (blurry margins with a pink, hyperemic disc), and glaucoma (a large, pale cup with retinal vessels that dive underneath and that may be displaced toward the nasal side). The generalist's examination inadequately detects early glaucomatous changes, so high-risk patients should undergo routine ophthalmologic examination for glaucoma.[10]

After inspecting the disc, the upper and lower nasal quadrants should be examined for the appearance of vessels and the presence of any retinal hemorrhages (see Fig. 423-24) or lesions. Proceeding from the nasal quadrants to the temporal quadrants decreases the risk for papillary constriction from the bright light focused on the fovea. Dilating the pupils leads to an improved examination. Patients with diabetes (Chapter 229) should undergo routine examination by eye care experts because the sensitivity of a generalist's examination is not adequate to exclude diabetic retinopathy or monitor it over time.

Neck

Carotid Pulses

The carotid pulses should be palpated for contour and timing in relation to the cardiac impulse. Abnormalities in the carotid pulse contour reflect underlying cardiac abnormalities (e.g., aortic stenosis) but are generally appreciated only after detecting an abnormal cardiac impulse or murmur (Chapter 51).

Many physicians listen for bruits over the carotid arteries because asymptomatic carotid bruits are associated with an increased incidence of cerebrovascular and cardiac events in older patients (Chapters 406 and 407). In asymptomatic patients, the presence of a carotid bruit increases the likelihood of a 70 to 90% stenotic lesion (LR 4 to 10), but the absence of a bruit is of uncertain value. Unfortunately, clinical data do not provide adequate data for judging the importance of detecting bruits in asymptomatic patients.

Jugular Veins

The examination of the neck veins is an interesting but often unreliable indicator of central venous pressure or fluid responsiveness in hospitalized sick patients (Chapter 51).[11] Inspection of the waveforms may facilitate the interpretation of the cardiac examination for right heart valvular lesions. The waves are seen best by shining a penlight obliquely on the vein while the examiner looks for the dynamic changes of the projected shadow on the bed linen.

Thyroid

The thyroid gland is felt best when standing behind the patient and using both hands to palpate the thyroid gland gently (Chapter 226). Palpation is enhanced when the patient swallows sips of water to allow the thyroid to glide underneath the fingers. When viewed from the side, lateral prominence of the thyroid between the cricoid cartilage and the suprasternal notch indicates thyromegaly. The generalist physician should estimate the size of the thyroid gland as normal or enlarged; the impression of an enlarged thyroid gland by a generalist physician has an LR of almost 4, whereas assessment of normal size makes thyromegaly less likely (LR 0.4). The volume of a normal thyroid gland is no greater than the volume of the patient's distal thumb phalanx.

Lymphatic System

While palpating the thyroid, the examiner may also identify enlarged cervical lymph nodes (Chapter 168). Lymph nodes can also be palpated in the supraclavicular area, axilla, epitrochlear area, and inguinofemoral region. Simple

lymph node enlargement confined to one region is common and does not usually represent an important underlying disorder. Unexpected gross lymph node enlargement in a single area or diffuse lymph node enlargement is more important. Patients with febrile illnesses, underlying malignancy, or inflammatory diseases should routinely undergo an examination of each of the aforementioned areas for lymph node enlargement.

Chest

Inspection of the patient's posture may reveal lateral curves in the back (scoliosis) or kyphosis that may be associated with loss of vertebral height from osteoporosis (Chapter 243). When patients have back pain, the spine and paravertebral muscles should be palpated for spasm and tenderness (Chapter 400). The patient may be placed through maneuvers to assess loss of mobility associated with ankylosing spondylitis (Chapter 265), but a history of loss of lateral mobility may be just as efficient in the early stages of spondylitis.

Lungs

The incremental value of palpation and percussion of the chest to supplement the history, auscultation, and eventual chest radiograph is unknown. Normal vesicular sounds, which approximate a 3:1 inspiratory:expiratory ratio with no pause between phases, are heard throughout most of the normal posterior chest during quiet breathing. Auscultated wheezes are continuous adventitial sounds. Crackles (formerly called rales) are discontinuous sounds heard in conditions that stiffen the lung (heart failure, pulmonary fibrosis, and obstructive lung disease). The best piece of information for increasing the likelihood of chronic obstructive pulmonary disease is a history of more than 40 pack years of smoking (LR 19). The presence of wheezing or downward displacement of the larynx to within 4 cm of the sternum (distance between the top of the thyroid cartilage and the suprasternal notch) increases the likelihood of obstructive pulmonary disease (LR of 4 for either).

Heart

The patient should be examined in the sitting and lying positions (Chapter 51). Palpation of the apical impulse in the left lateral decubitus position helps detect a displaced apical impulse and can reveal a palpable S_3 gallop. When the apical impulse is lateral to the midclavicular line, radiographic cardiomegaly (LR 3.5) and an ejection fraction of less than 50% (LR 6) are more likely. Most examiners auscultate in sequence the second right then the second left intercostal spaces, the left sternal border, and then the apex. The examiner should concentrate on the timing, intensity, and splitting of sounds with respiration. The first and second heart sounds are heard best with the diaphragm, as are pericardial rubs. Gallops (S_3 and S_4) are heard best with the stethoscope bell. High-pitched versus low-pitched murmurs are detected by switching from the diaphragm to the bell. The location, timing, intensity, radiation patterns, and respiratory variation of murmurs should be noted. Special maneuvers during auscultation (e.g., Valsalva, auscultation during sudden squatting or standing) do not usually need to be performed if the results of routine precordial examination are entirely normal.

The presence of an S_3 gallop is useful for detecting left ventricular systolic dysfunction (LR > 4 for identifying patients with an ejection fraction of <30%). The presence of a systolic thrill (palpable murmur, LR 12) or a holosystolic murmur increases the likelihood of moderate to severe aortic stenosis or mitral regurgitation. Quiet systolic murmurs (LR 0.08) are much less likely to herald important cardiac abnormalities. A loud, early diastolic murmur (LR 4) or a diastolic murmur associated with an S_3 suggests severe aortic regurgitation.

Breast

The most important determinants of the accuracy of the breast examination are the duration of the examination; the patient's position; careful evaluation of the breast boundaries; the pattern of the examination; and the position, movement, and pressure of the examiner's fingers (Chapter 198). To obtain the best sensitivity, the duration of the breast examination needs to be 5 to 10 minutes' total time, but few generalist physicians perform such a lengthy examination. Clinicians should recognize that the examination may make them (or their patient) feel uncomfortable—the presence of a chaperone may give the clinician the confidence to perform an intensive examination.

The patient should be examined with the pads of the fingers while she is supine, holding her hand first on her forehead (to flatten the lateral border of the breast) and then on her shoulder (to flatten the medial border). The examiner should make small circular motions with the fingers, moving up and down in parallel rows to span the entire breast-clavicle to the bra line. Cancerous breast lumps are difficult to distinguish from benign breast lumps on examination, but the presence of a fixed mass or a mass 2 cm in diameter has an LR of about 2 to 2.5 for cancer.

Abdomen

When patients have potential abdominal symptoms and the history suggests an acute problem, the examination should focus initially on identifying patients who require surgical evaluation. Palpation and percussion of the abdomen of patients with no symptoms or risk factors for an abdominal disorder seldom reveal important abnormalities (Chapter 132) except for asymptomatic widening of the abdominal aorta in older patients (LR of 16 for detecting aneurysms >4 cm in diameter). However, palpation misses a substantial proportion of small to medium aneurysms (Chapter 78).

The presence of bowel sounds in patients with acute symptoms can be falsely reassuring because the sounds can be present despite an ileus and may be increased early in an obstruction. For patients without gastrointestinal symptoms or abnormalities on palpation, auscultation for bruits is important primarily to detect renal bruits in patients with hypertension (Chapters 67 and 125). The presence of an abdominal bruit in a hypertensive patient, if heard in systole and diastole, strongly suggests renovascular hypertension (LR ≈ 40).

Liver

Detection of liver disease depends mostly on the history and laboratory evaluations (Chapter 146). By the time that signs are present on physical examination, the patient usually has advanced liver disease. The first abnormalities on physical examination associated with liver disease are extrahepatic. The clinician should assess the patient for ascites, peripheral edema, jaundice, or splenomegaly. In patients with an enlarged liver, palpation should begin at the liver edge, but palpation of the edge below the costal margin increases the likelihood of hepatomegaly only slightly (LR 1.7). The upper border of the liver may be detected by percussion, and a span of less than 12 cm reduces the likelihood of hepatomegaly. In the absence of a known diagnosis (e.g., a hepatoma, which may cause a hepatic bruit), auscultation of the liver rarely is helpful.

Spleen

Examination for splenomegaly in patients without findings suggestive of a disorder associated with splenomegaly almost always reveals nothing (Chapter 168). Approximately 3% of healthy teenagers may have a palpable spleen. The examination for an enlarged spleen begins first with percussion in the left upper quadrant to detect dullness. Palpation can be performed by any of the following three approaches ($\kappa \approx 0.2$ to 0.4): palpating with the right hand while providing counterpressure with the left hand behind the spleen, palpating with one hand without counterpressure (with the patient in the right lateral decubitus position for both techniques), or placing the patient supine with the left fist under the left costovertebral angle while the examiner tries to hook the spleen with the hands.

Musculoskeletal System

The musculoskeletal examination in adult patients is almost always driven by symptoms (Chapters 256 and 263). Most patients have back pain at some point during their life (Chapter 400). The patient's history helps assess the likelihood of an underlying systemic disease (age, history of systemic malignancy, unexplained weight loss, duration of pain, responsiveness to previous therapy, intravenous drug use, urinary infection, or fever). The most important physical examination findings for lumbar disc herniation in patients with sciatica all have excellent reliability, including ipsilateral straight leg raising causing pain, contralateral straight leg raising causing pain, and ankle or great toe dorsiflexion weakness.

The generalist physician should evaluate an adult patient with knee discomfort for torn menisci or ligaments. The best maneuver for demonstrating a tear in the anterior cruciate ligament is the anterior drawer or Lachman maneuver, in which the examiner detects the lack of a discrete end point as the tibia is pulled toward the examiner while the femur is stabilized. A variety of maneuvers that assess for pain, popping, or grinding along the joint line between the femur and tibia are used to evaluate for meniscal tears. As with many musculoskeletal disorders, no single finding has the accuracy of the orthopedist's examination, which factors in the history and a variety of clinical findings.

The shoulder examination is directed toward determining range of motion, maneuvers that cause discomfort, and assessment of functional disability.

Hip osteoarthritis is detected by evidence of restriction of internal rotation and abduction of the affected hip. Generalist physicians often rely on radiographs to determine the need for referral to orthopedic physicians, but routine radiographs are not needed early in the course of shoulder or hip disorders. The degree of pain and disability experienced by the patient may prompt confirmation of the diagnosis and referral.

The hands and feet may show evidence of osteoarthritis (local or as part of a systemic process) (Chapter 262), rheumatoid arthritis (Chapter 264), gout (Chapter 273), or other connective tissue diseases. In addition to regional musculoskeletal disorders, such as carpal tunnel syndrome, a variety of medical and neurologic conditions should prompt routine examination of the distal ends of the extremities to prevent complications (e.g., diabetes [neuropathy or ulcers] or hereditary sensorimotor neuropathy [claw toe deformity]).

Skin

The skin should be examined under good lighting (Chapter 436). It is best to ask the patient to point out any spots on the skin of concern. Examiner agreement on some of the most important features of melanoma (asymmetry, haphazard color, border irregularity) is fair to moderate (Chapter 203). A lesion that is symmetrical, has regular borders, is only one color, is 6 mm or smaller, or has not enlarged in size is unlikely to represent a melanoma (LR 0.07). However, an increasing number of findings greatly enhance the likelihood of melanoma (LR 2.6 for two or more findings and LR 98 for the presence of all five findings) (Chapter 203).

Basal cell carcinoma and squamous cell carcinoma occur more frequently than melanoma (Chapter 203). These lesions can be detected during routine examination by paying careful attention to sun-exposed areas of the nose, face, forearms, and hands.

Neurologic Examination

Full details of the neurologic examination are given in Chapter 396.

Psychiatric Evaluation

During the general examination, much of the psychiatric assessment (including cognition) is accomplished while eliciting the routine history and performing the review of systems (Chapter 397). Observation of the patient's mannerisms, affect, facial expression, and behavior may suggest underlying psychiatric disturbances. When a screening survey and review of systems are obtained by a questionnaire completed by the patient, the clinician should review the responses carefully to determine whether the patient exhibits symptoms of depression. Specific questioning for symptoms of depression is appropriate for all adult patients. Military veterans should be screened for post-traumatic stress disorder and possible prior traumatic brain injuries that may affect their behaviors. Delirium (Chapter 28) is common in both medical and surgical inpatients and is recognized by fluctuating mental status. Delirium should be suspected when the patient has trouble carrying on a normal conversation during bedside rounds; but the patient's nurse and visitors may detect delirium before the physician, so their report may contribute to diagnosis.[12]

Genitalia and Rectum

Pelvic Examination

A complete examination includes a description of the external genitalia, appearance of the vagina and cervix as seen through a speculum, and bimanual palpation of the uterus and ovaries (Chapters 199 and 237). About 10 to 15% of asymptomatic women have some abnormality on examination, and 1.5% have abnormal ovaries. However, screening for ovarian cancer is limited by the low sensitivity of the physical examination for detecting early-stage ovarian carcinoma (Chapter 199). In the emergency setting, all women of reproductive age with vaginal bleeding and pelvic pain should have a pregnancy test and an ultrasound to evaluate them for a possible ectopic pregnancy.[13]

Male Genitalia

Examination of the male genitalia should begin with a description of whether the penis is circumcised and whether there are any visible skin lesions (e.g., ulcers or warts). Palpation should confirm the presence of bilateral testes in the scrotum. The epididymis and testes should be palpated for nodules. The low incidence of testicular carcinoma means that most nodules are benign (Chapter 200).

The prostate should be examined in all quadrants, with attention focused on surface irregularities or differences in consistency throughout the prostate (Chapter 201). An estimate of prostate size may be confounded by the size of the examiner's fingers. It may be best to estimate the size of the prostate in centimeters of width and height.

Rectum

Patients can be examined while lying on their side, although this approach may place the examiner in an awkward stance (Chapters 132 and 145). The rectal examination in women can be performed as part of a bimanual examination, with the index finger in the vagina and the third finger in the rectum to permit palpation of the rectovaginal vault. Men may be asked to stand and lean over the examining table; alternatively, they may be examined while on their back with their hips and knees flexed. This latter maneuver is not used often, although it may facilitate examination of the prostate, which falls into the finger in this position.

The rectal examination begins with inspection of the perianal area for skin lesions. A well-lubricated, gloved finger is placed on the anus, and while applying gentle pressure, the examiner asks that the patient bear down as though having a bowel movement. This maneuver facilitates entry of the finger into the rectum. A normal rectal response includes tightening of the anal sphincter around the finger. The examiner should palpate circumferentially around the length of the fully inserted finger for masses. On withdrawing the gloved finger, the finger should be wiped on a stool guaiac card for fecal blood testing to assess for acute blood loss. As a screening test for colorectal carcinoma (Chapter 193), digital examination does not replace the need for testing stool samples collected by the patient (or using alternative screening strategies, such as flexible sigmoidoscopy or colonoscopy).

SUMMARIZING THE FINDINGS FOR THE PATIENT

The physician should summarize the pertinent positive and negative findings for the patient and be willing to express uncertainty to the patient, provided that it is accompanied by a plan of action (e.g., "I will reexamine you on your next visit"). The rationale for subsequent laboratory, imaging, or other tests should be explained. A plan should be established for providing further feedback and results to the patient, especially when there is a possibility that bad news may need to be delivered. Some physicians ask the patient if there is "anything else" to be covered. Patients who express additional new concerns at the end of the visit may have been fearful to address them earlier (e.g., "by the way, doctor, I'm getting a lot of chest pain"); when the problems seem non-urgent, it is acceptable to reassure the patient and offer the promise of evaluating the patient in a follow-up phone call or at the next visit.

FUTURE DIRECTIONS

The common assumption that physicians' diagnostic skills are deteriorating is not supported by evidence. There is considerable evidence that the scientific approach to understanding what is worthwhile and what is not worthwhile during the clinical examination identifies a core set of skills for clinical diagnosticians. Because good patient outcomes at good value are driven primarily by the quality of the information obtained during the clinical examination, continued application of scientific principles to the history and physical examination should improve diagnostic skills.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Krogsboll LT, Jorgensen KJ, Larsen CG, et al. General health checks in adults for reducing morbidity and mortality from disease: Cochrane systematic review and meta-analysis. *BMJ*. 2012;345:e7191.
2. Verghese A, Brady E, Kapur CC, et al. The bedside evaluation: ritual and reason. *Ann Intern Med*. 2011;155:550-553.
3. Powers BJ, Trinh JV, Bosworth HB. Can this patient read and understand written health information? *JAMA*. 2010;304:76-84.
4. Makadon HJ. Ending LGBT invisibility in health care: the first step in ensuring equitable care. *Cleve Clin J Med*. 2011;78:220-224.
5. Department of Veterans Affairs. Military Health History Pocket Card for Clinicians. http://www.va.gov/oaa/pocketcard/military-health-history-card-for-print.pdf. Accessed February 9, 2015.
6. Murff HJ, Spigel DR, Syngal S. Does this patient have a family history of cancer? An evidence-based analysis of the accuracy of family cancer history. *JAMA*. 2004;292:1480-1489.
7. Brenner S, Guder G. The patient with dyspnea. Rational diagnostic evaluation. *Herz*. 2014; 39:8-14.
8. Bagai A, Thavendiranathan P, Detsky AS. Does this patient have hearing impairment? *JAMA*. 2006;295:416-428.
9. Hollands H, Johnson D, Brox AC, et al. Acute-onset floaters and flashes: is this patient at risk for retinal detachment? *JAMA*. 2009;302:2243-2249.
10. Hollands H, Johnson D, Hollands S, et al. Do findings on routine examination identify patients at risk for primary open-angle glaucoma? The rational clinical examination systematic review. *JAMA*. 2013;309:2035-2042.
11. Marik PE, Cavallazzi R. Does the central venous pressure predict fluid responsiveness? An updated meta-analysis and a plea for some common sense. *Crit Care Med*. 2013;41:1774-1781.
12. Wong CL, Holroyd-Leduc J, Simel DL, Straus SE. Does this patient have delirium? Value of bedside instruments. *JAMA*. 2010;304:779-786.
13. Crochet JR, Bastian LA, Chireau MV. Does this woman have an ectopic pregnancy? The rational clinical examination systematic review. *JAMA*. 2013;309:1722-1729.

REVIEW QUESTIONS

1. A 24-year-old-woman with right lower quadrant abdominal pain and vaginal bleeding has a positive home pregnancy. Before a pelvic ultrasound is ordered, the pelvic examination is performed and reveals cervical motion tenderness. Cervical motion tenderness has the following diagnostic characteristics for an ectopic pregnancy: sensitivity 45%, specificity 91%, likelihood ratio positive (LR+) 4.9, LR negative (LR−) 0.62, positive predictive value (PPV) 46%. Which of these values is most helpful for assessing the probability that she has an ectopic pregnancy?

A. Sensitivity
B. Specificity
C. Likelihood ratio positive
D. Likelihood ratio negative
E. Positive predictive value

Answer: C The sensitivity is the percentage of patients who have cervical motion tenderness among women with an ectopic pregnancy. The specificity is the percentage of patients who do not have cervical motion tenderness among women without an ectopic pregnancy. These individual values are not helpful for this particular patient (Crochet JR, Bastian LA, Chireau MV. does this woman have an ectopic pregnancy? The rational clinical examination systematic review. *JAMA*. 2013;309:1722-1729) because without knowing whether or not she has an ectopic pregnancy, we do not know which result applies. The PPV describes the probability of ectopic pregnancy when there is cervical motion tenderness. To use the PPV for this patient requires knowing the prevalence of disease from which the PPV was derived. Without knowing that value, you cannot be certain whether the PPV is appropriate for your patient. The LR+ quantifies the increase in odds of disease among those with cervical motion tenderness, and the LR− quantifies the decrease in odds of disease when cervical motion tenderness is absent. Based on your assessment of the prior probability of ectopic pregnancy, which is 10 to 20% among all pregnant women with abdominal pain and/or vaginal bleeding, the LR+ is the most relevant value because it can be used to determine the increase in likelihood of ectopic pregnancy (Chapter 10).

2. The primary role of completing a review of systems in helping with clinical diagnosis during the clinical evaluation is which of the following?

A. To review symptoms associated with the presenting problem
B. To pick up the presence of concerns that were uncomfortable to address during the history of the present illness
C. To complete a record that enhances patient billing
D. To focus on the presence or absence of findings during a constrained time period
E. To allow open-ended questioning that enhances the patient's relationship with the physician

Answer: D A complete medical history should reveal the most important symptoms experienced by the patient that are pertinent to the presenting problems. Open-ended questioning should occur during elicitation of the medical history. Although completing a review of systems may be required for electronic medical records and billing, simply fulfilling that function does not help with diagnosis. However, about 10% of the time, the review of systems might pick up on an important finding that was not addressed during the rest of the evaluation. The main purpose of the review of systems is to use direct questioning to pick up on a limited set of symptoms, during a specified recent time interval (e.g., over the past week, or the past month). This approach helps the patient focus on a well-defined and recent time period that should enhance the reliability of their answers that pertain to their current condition.

3. While conducting your new patient evaluation on a 35-year-old male veteran who served in combat duty in Afghanistan, you note that he seems anxious and that he is questioning you frequently about a variety of difficulties in making his appointment and the attitudes of your office staff. You recognize the need to screen for post-traumatic stress disorder using a four-item screening instrument. You start by asking, "In your life, have you ever had an experience so horrible, frightening, or upsetting that, in the past month you …" All of the items below are part of the four-item screening instrument for post-traumatic stress disorder that follows this introduction except:

A. Have little interest or pleasure in doing things?
B. Have had nightmares about it or thought about it when you did not want to?
C. Tried hard not to think about it or went out of your way to avoid situations that reminded you of it?
D. Were constantly on guard, watchful, or easily startled?
E. Felt numb or detached from others, activities, or your surroundings?

Answer: A The loss of interest or pleasure in doing things is a symptom that suggests depression. The other symptom that is elicited in a two-item screener for depression asks the patient whether they feel down, depressed, or hopeless (Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care*. 2003;41:1284-1292). Patients who answer "yes" to at least one of the items have an LR+ of 2.7 for depression, whereas answering "no" to both questions makes depression much less likely with an LR of 0.14 (Williams JW Jr, Noël P, Cordes JA, et al. Is this patient clinically depressed? *JAMA*. 2002;287:1160-1170). Although it is important to screen for depression among patients with post-traumatic stress disorder (military related or not) (Campbell DG, Felker BL, Liu CF, et al. Prevalence of depression-PTSD comorbidity: implications for clinical practice guidelines and primary care-based interventions. *J Gen Intern Med*. 2007;22:711-718), screening for depression alone may not detect the associated post-traumatic stress disorder.

8

APPROACH TO THE PATIENT WITH ABNORMAL VITAL SIGNS

DAVID L. SCHRIGER

Care of the patient is guided by integration of the chief complaint, history, vital signs, and physical examination findings (Chapter 7). Physicians should be keenly aware of a patient's vital signs but should seldom make them the centerpiece of the evaluation.

TABLE 8-1 NORMAL AND PANIC RANGES FOR KEY VITAL SIGNS IN ADULTS*

	NORMAL	PANIC
Temperature	36°-38° C (96.8°-100.4° F)	40° C (104° F)
Pulse	60-100 beats/min	<45 beats/min, >130 beats/min
Respirations	12-20 breaths/min	<10 breaths/min, >26 breaths/min
Oxygen saturation	95-100%	<90%
Systolic blood pressure	90-130 mm Hg	<80 mm Hg, >200 mm Hg
Diastolic blood pressure	60-90 mm Hg	<55 mm Hg, >120 mm Hg

*Normal values are for healthy adults. Values outside these ranges are common in patients who are ill or are anxious about their health care encounter. Panic values demand the health care provider's attention in any adult patient. These values are specific (rarely present in healthy patients) but not sensitive (most ill patients' vital signs will not include panic values). All vital signs must be interpreted in the context of the patient's presentation (see text).

THE IMPORTANCE OF VITAL SIGNS

The importance of vital signs in medical care is a conundrum for proponents of an evidence-based approach to the care of patients. No experienced physician would be willing to care for patients without them, yet a formal evaluation of the utility of vital signs for making specific diagnoses would conclude that they are not particularly useful because their likelihood ratios are too close to 1 to differentiate those who have a specific condition from those who do not (Chapter 7). For uncommon conditions, their predictive value is even worse. For example, the probability of tachycardia in a patient in thyroid storm is high, yet the probability of thyroid storm in a patient with isolated tachycardia is low. This application of Bayes theorem (Chapter 10) demonstrates why there is no justification for ordering thyroid tests for every tachycardic patient and why attempts to say "When vital sign *x* is high [low], do *y*" fail. Each vital sign can be normal or abnormal in almost every acute condition (Table 8-1), and vital signs can be transiently abnormal in healthy individuals. An algorithmic approach to testing and treatment in response to abnormal vital signs would be too vague and too complex to be of use.

Predictive Value

How can it be that vital signs are poor predictors of diagnoses but central to the practice of medicine? First, although vital signs are insufficiently predictive to be of use in rigid algorithms, these algorithms are but one of several heuristics used by physicians to diagnose and to treat patients. Pattern recognition and the hypothetical-deductive model are heuristics that are based not on average tendencies of a single factor (e.g., hypotension is present in *x*% of cases of septic shock) or a small number of factors (hypotension and tachycardia are present in *y*% of cases of septic shock) but on the complex interaction of multiple factors (e.g., because this patient is an ill-appearing elderly man with an enlarged prostate and a history of urinary tract infections, is tachycardic and hypotensive, has clear lungs and an enlarged but nontender prostate, and has an oxygen saturation of 97%, he should be treated for urosepsis [Chapter 284] while awaiting results of urinalysis and urine culture). Thus, vital signs can play an important function in medical decision making even though their likelihood ratios for specific conditions are unimpressive.

Despite their poor predictive value for any single diagnosis, abnormal vital signs help identify patients who are sicker.[1] For example, even one abnormal vital sign in the emergency department significantly increases the likelihood of adverse outcomes in elderly patients by about 50%,[2] and abnormal vital signs after admission carry a 20-fold increased risk for subsequent important deterioration.[3] Abnormal vital signs are also a key predictor of which patients are at risk of dying soon after being sent home from an emergency department.[4]

The usefulness of vital signs is also evidenced by the finding that more severe abnormalities are associated with an even worse prognosis. For example, among hospitalized patients, one critically abnormal vital sign carries about a 1% risk for inpatient death, whereas three simultaneously abnormal vital signs carry nearly a 25% risk for inpatient death.[5]

Vital Signs as Symptoms

Abnormal vital signs are seldom the fundamental pathophysiologic problem. In shock (Chapter 106), hypotension and tachycardia are manifestations of pathophysiologic processes occurring at cellular and molecular levels. Given the circuitous links from clinical disease to fundamental pathophysiology to abnormal vital signs, it is not surprising that the relationships between the disease states and vital signs are not strong. Until new technologies enable direct measurement of primary pathologic processes, vital signs remain an important, albeit imperfect, proxy.

The five key vital signs are temperature, pulse, blood pressure, respiratory rate, and oxygen saturation (pulse oximetry). Pulse oximetry is included because it has become widely available in acute care settings, is noninvasive and relatively inexpensive, and provides information unique from the respiratory rate. Advocates have suggested that pain, smoking status, and weight be considered routine vital signs; although a case can be made for each, they are not considered here. Clinicians should never forget that the most important vital sign is what the patient looks like; general appearance is a sign that guides the intensity and urgency of the evaluation.

MEASURING VITAL SIGNS

Although obtaining vital signs is generally straightforward, the validity and reliability of measurement depend on proper technique and, for blood pressure and pulse oximetry, well-maintained equipment. Rectal and oral temperatures are generally accurate (Chapter 280), although oral temperatures can be falsely depressed in patients who breathe through their mouths. Axillary temperature is unreliable and should not be used. There is wide variability in the validity and reliability of measurements performed with tympanic membrane thermometers. A hypothermia thermometer is preferred in patients with suspected hypothermia, and core temperature should be measured with an esophageal, bladder, or rectal temperature sensor in patients with severe hypothermia or hyperthermia (Chapter 109).

Blood pressure must be measured with an appropriately sized cuff (Chapter 67). Automated blood pressure machines occasionally provide spurious results, and questionable values should be confirmed by manual auscultation and by checking other limbs. Pulse is best obtained by palpation because this technique provides the opportunity to assess regularity and contour; the pulse should be counted for sufficient time for an accurate rate to be obtained (at least 15 seconds). High heart rates on the digital readout of a cardiac monitor must be confirmed by palpation because these monitors can spuriously count large P waves, T waves, or pacemaker spikes as R waves, thereby reporting a heart rate double the actual rate. Orthostatic vital signs—the comparison of blood pressure and pulse in the supine, sitting, and standing positions—are advocated by some but have proved to be insensitive and nonspecific for hypovolemia.

Because the typical respiratory rate is between 12 and 20 breaths per minute and because there is considerable breath-to-breath variation, the respiratory rate should be assessed for at least 30 seconds and preferably 1 minute. New technologies purporting to measure the respiratory rate have not proved clinically useful. Oxygen saturation is dependent on technology, so an understanding of the idiosyncrasies of the device being used is critical; valid measurements are unlikely unless there is good correlation of the machine's pulse reading and the patient's pulse. The probe should be placed on a part of the body that is warm and well perfused. Pulse oximeters compare the absorption of light at two wavelengths, so readings may be spuriously high under conditions that change the color of oxygenated or deoxygenated hemoglobin, including carbon monoxide poisoning (Chapter 94), methemoglobinemia (Chapter 161), and some of the less common hemoglobinopathies.

ROLE OF VITAL SIGNS IN MANAGEMENT OF THE PATIENT

Abnormal vital signs should be remeasured. Certain abnormalities require prompt evaluation (Table 8-2). Other vital sign abnormalities should be rechecked in the future unless they have been previously noted, in which case a work-up can be initiated, guided by the patient's past history and physical examination findings. It is critical that the physician always "treat the patient, not the vital signs."

Patients without Systemic Complaints

In patients presenting for a routine evaluation or nonsystemic complaint (e.g., knee injury), an abnormal vital sign will seldom be the harbinger of acute illness. Most commonly, it will be a false reading or a transient finding due to random variation or anxiety that requires no evaluation or treatment and can be rechecked in the future. On occasion, it will be the only or most apparent manifestation of a chronic condition or risk factor. The

TABLE 8-2 ABNORMALITIES REQUIRING RAPID EVALUATION IN THE ASYMPTOMATIC PATIENT

An irregularly irregular rapid pulse (if it is not known to be chronic) should trigger an evaluation of the patient's rhythm so that atrial fibrillation can be identified, evaluated, and treated (Chapter 64), thereby decreasing the patient's risk of stroke.
A heart rate above 130 beats per minute warrants an electrocardiogram to determine the patient's rhythm and a consideration of the differential diagnosis of tachycardia (hypovolemia, anemia, and thyroid disease in particular).
A markedly elevated diastolic blood pressure (e.g., >115 mm Hg) should stimulate an evaluation for hypertensive urgencies (Chapter 67). Note that hypertension in the absence of signs of acute end-organ damage does not require acute treatment, which can reduce intracranial perfusion pressure and cause stroke. Patients with elevated blood pressure should be offered standard evaluation and treatment for chronic hypertension (Chapter 67).
Markedly low pulse or blood pressure in patients receiving cardioactive medications should lead to a confirmation that the patient is truly asymptomatic, an inquiry into the dosing of these medications, and a reconsideration of the regimen.
Markedly low pulse in elderly patients who are not receiving rate-controlling drugs should trigger an evaluation of the patient's cardiac conduction system.
Oxygen saturation below 93% in the absence of known pulmonary problems should prompt an evaluation of the patient's pulmonary status.

measurement of an elevated blood pressure leading to a diagnosis of hypertension is the classic example of the value of vital signs in such patients.

Patients Who Complain of Systemic Illness but Do Not Appear to Be Very Ill

Vital signs serve two additional roles in symptomatic patients who do not appear particularly ill. First, abnormalities in vital signs provide information that may suggest or support a diagnosis. The presence of elevated temperature in a patient with productive cough, shortness of breath, and localized rales and egophony supports a diagnosis of infectious pneumonia. Vital signs may also play a role in defining therapy and triage. For example, guidelines for patients with community-acquired pneumonia (Chapter 97) formally incorporate vital signs.

The second role of vital signs in the stable symptomatic patient is to provide warning that the patient is sicker than he or she appears. For example, the presence of hypotension in a well-appearing patient thought to have pyelonephritis may be an indication of sepsis or hypovolemia. For vital signs to be of use, the physician must be aware of them and must incorporate them explicitly into a thought process that considers the dangerous diagnoses associated with the abnormal vital sign. The physician then must decide whether the likelihood of each potentially dangerous diagnosis is high enough to warrant specific evaluation. Unfortunately, no quick or easy rules differentiate spurious abnormalities that can be ignored from those that should trigger additional testing or treatment. What can be said is that the well-trained physician who is aware of abnormal vital signs and is willing to contemplate a change in treatment or disposition in response to them is less likely to make mistakes.

A few specific points bear mention. First, for most vital signs, "normal" is relative. Blood pressure must be interpreted in the context of the patient. For example, a blood pressure of 88/64 mm Hg may be reasonable for an otherwise healthy, young 50-kg woman but should cause concern in a 90-kg middle-aged man. Similarly, a blood pressure of 128/80 mm Hg would be fine in a 60-year-old man but worrisome in a 34-week pregnant woman. Second, because vital signs are insensitive measures of disease, normal vital signs should not dissuade the physician from pursuing potentially critical diagnoses. For example, young, well-conditioned adults may maintain normal vital signs well into the course of shock.

Use of Vital Signs in Patients Who Appear to Be Ill

For some patients, abnormal vital signs are expected on the basis of their appearance and their symptoms. For patients in extremis, care should proceed according to established guidelines such as Advanced Cardiac Life Support (Chapter 63), Advanced Trauma Life Support, and algorithms for the treatment of shock (Chapters 107 and 108). For other ill-appearing patients, two processes must occur. In one, the physician, armed with knowledge of the differential diagnosis of each abnormal vital sign and the ability to take a thorough history and to perform an appropriate physical examination, narrows the list of potential diagnoses and decides which are of sufficient probability to warrant evaluation. Simultaneously, the physician considers the list of treatment options for all diagnoses associated with the abnormal vital sign and, before establishing a diagnosis, initiates those treatments for which the potential benefit of prompt administration exceeds potential harms. For example, antibiotics for febrile patients at risk for bacterial infection, hydrocortisone for hypotensive patients at risk for hypoadrenalism, and thiamine for hypothermic patients at risk for Wernicke encephalopathy may improve outcome and are unlikely to cause harm even if the patient does not have the suspected condition. Although early presumptive treatment can be life-saving in selected patients, it should not be abused; physicians must avoid knee-jerk responses that can cause harm.

Differential Diagnosis and Treatment Options

Single Abnormal Vital Signs

Because vital signs can be abnormal in virtually any disease process, no differential diagnosis can be encyclopedic. The physician should focus initially on common diseases and diseases that require specific treatment. The thought process should begin with the chief complaint and history and then incorporate information about the vital signs and the remainder of the physical examination.

Multiple Abnormal Vital Signs

Patients who are acutely ill are likely to have several abnormal vital signs. Although certain patterns of abnormal vital signs predominate in specific conditions (e.g., hypotension, tachycardia, and hypothermia in profound sepsis), no pattern can be considered pathognomonic. The physician's goal is to work toward a diagnosis while simultaneously providing treatments whose benefits outweigh potential harms.

Fever is generally accompanied by tachycardia, with the general rule of thumb that the heart rate will increase by 10 beats per minute for every 1° C increase in temperature. The absence of tachycardia with fever is known as pulse-temperature dissociation and has been reported in typhoid fever (Chapter 308), legionnaires disease (Chapter 314), babesiosis (Chapter 353), Q fever (Chapter 327), infection with *Rickettsia* spp (Chapter 327), malaria (Chapter 345), leptospirosis (Chapter 323), pneumonia caused by *Chlamydia* spp (Chapter 318), and viral infections such as dengue fever (Chapter 382), yellow fever (Chapter 381), and other viral hemorrhagic fevers (Chapter 381), although the predictive value of this finding is unknown.

Much can be learned by comparing the respiratory rate with pulse oximetry. Hyperventilation in the presence of high oxygen saturation suggests a central nervous system process or metabolic acidosis rather than a cardiopulmonary process. Low respiratory rates in the presence of low levels of oxygen saturation suggest central hypoventilation, which may respond to narcotic antagonists.

Hypertension and bradycardia in the obtunded or comatose patient are known as the Cushing reflex, a relatively late sign of elevated intracranial pressure. Physicians should strive to diagnose and treat this condition before the Cushing reflex develops.

Approach to Abnormalities of Specific Vital Signs

Elevated Temperature

Normal temperature is often cited as 37° C (98.6° F), but there is considerable diurnal variation and variation among individuals, so 38° C is the most commonly cited threshold for fever. Fever thought to be due to infection should be treated with antipyretics and appropriate antimicrobials (Chapter 280). The importance of early administration of antibiotics to potentially septic patients cannot be overstated (Chapters 280 and 281). Hyperthermia (temperature above 40° C) should be treated with cooling measures such as ice packs, cool misting in front of fans, cold gastric lavage, and, for medication-related syndromes, medications such as dantrolene (Chapter 109). Most hospital anesthesia departments will have a designated kit for the treatment of malignant hyperthermia (Chapters 432 and 434).

Low Temperature

The treatment of hypothermia is guided by its cause (Chapter 109). The body's temperature decreases when heat loss exceeds heat production. Every logically possible mechanism for this phenomenon has been observed. Decreased heat production can result from endocrine hypofunction (e.g., Addison disease [Chapter 227], hypopituitarism [Chapter 224], hypothyroidism [Chapter 226]) and loss of the ability to shiver (e.g., drug-induced or neurologic paralysis or neuromuscular disorders). Malfunction of the hypothalamic regulatory system can be due to hypoglycemia (Chapter 229) and a variety of central nervous system disorders (Wernicke encephalopathy

[Chapter 416], stroke [Chapter 407], tumor [Chapter 189], and trauma [Chapter 399]). Resetting of the temperature set point can occur with sepsis. Increased heat loss can be due to exposure, behavioral and physical disorders that prevent the patient from sensing or responding to cold, skin disorders that decrease its ability to retain heat, and vasodilators (including ethanol). A careful history and physical examination should illuminate which of these possibilities is most likely.

Several considerations are worthy of emphasis. The spine of an obtunded hypothermic patient who is "found down" must be protected and evaluated because paralysis from a fall may have prevented the patient from seeking shelter and may have diminished the ability to produce heat. The physician should not forget to administer antibiotics to patients who may be septic (Chapter 108), thiamine to those who may have Wernicke encephalopathy (Chapter 416), hydrocortisone to those who may be hypoadrenal (Chapter 227), and thyroid hormone to those who may have myxedema coma (Chapter 226). Severely hypothermic patients (Chapter 109) should be treated gently because any stimulation may trigger ventricular dysrhythmias; even in the absence of pulses, cardiopulmonary resuscitation should be used only in patients with ventricular fibrillation or asystole.

Elevated Heart Rate

The rate, rhythm, and electrocardiogram differentiate sinus tachycardia from tachyarrhythmias (Chapters 62 to 65). Tachyarrhythmias can be instigated by conditions that may require specific treatment (e.g., sepsis [Chapter 108], electrolyte disorders [Chapters 116, 117, and 118], endocrine disorders [Chapter 221], and poisonings [Chapters 22 and 110]) before the arrhythmia is likely to resolve. For sinus tachycardia, treatment of the underlying cause is always paramount. Treatments may include antipyretics (for fever); anxiolytics; oral or intravenous fluids (for hypovolemia); nitrates, angiotensin-converting enzyme inhibitors, and diuretics (for heart failure and fluid overload [Chapter 59]); oxygen (for hypoxemia); α-blockers (for stimulant overdose); β-blockers (for acute coronary syndromes [Chapters 72 and 73] or thyroid storm [Chapter 226]); and anticoagulation (for pulmonary embolism [Chapter 98]). Tachycardia is often an appropriate response to a clinical condition and should not be treated routinely unless it is causing or is likely to cause secondary problems.

Low Pulse

Bradycardia can be physiologic (athletes and others with increased vagal tone), due to prescribed cardiac medications (e.g., β-blockers, calcium-channel blockers, digoxin), overdoses (e.g., cholinergics, negative chronotropic agents), disease of the cardiac conducting system, electrolyte abnormalities (severe hyperkalemia), and inferior wall myocardial infarction (Chapters 64 and 73). Asymptomatic patients do not require immediate treatment. The goal of therapy is to produce a heart rate sufficient to perfuse the tissues and alleviate the symptoms (Chapter 63). Overdoses should be treated with specific antidotes (Chapter 110). Endocrine disorders should be treated with replacement therapy. In patients with acute coronary syndrome (Chapter 72), the goal is to restore perfusion and alleviate the ischemia. Patients with profound bradycardia or hypotension may require chronotropic drugs to increase perfusion even if these agents increase myocardial oxygen demand. In normotensive patients with milder bradycardia, chronotropic agents should be used only if symptoms and ischemia cannot be resolved by other means. Atropine is the primary therapy for bradycardia; isoproterenol and cardiac pacing are reserved for those who do not respond (Chapter 63).

Elevated Blood Pressure

Elevated blood pressure does not require acute treatment in the absence of symptoms or signs of end-organ damage (Chapter 67). In patients whose blood pressure is markedly above their baseline, the history and physical examination should assess for the conditions that define "hypertensive emergency": evidence of encephalopathy, intracranial hemorrhage, ischemic stroke, heart failure, pulmonary edema, acute coronary syndrome, aortic dissection, renal failure, and preeclampsia. In the absence of these conditions, treatment should consist of restarting or adjusting the medications of patients with known hypertension and initiating a program of blood pressure checks and appropriate evaluation for those with no prior history of hypertension (Chapter 67).

The patient with a true hypertensive emergency should be treated with agents appropriate for the specific condition. Because rapid decreases in blood pressure can be as deleterious as the hypertensive state itself, intravenous agents with short half-lives, such as nitroprusside, labetalol, nitroglycerin, and esmolol, are preferred (Chapter 67).

Low Blood Pressure

Low blood pressure must be evaluated in the context of the patient's symptoms, general appearance, and physical examination findings. Treatment depends on context. The same blood pressure value may necessitate intravenous inotropic agents in one patient and no treatment in another.

In tachycardic hypotensive patients, the physician must rapidly integrate all available evidence to determine the patient's volume state, cardiac function, vascular capacitance, and primary etiology (Chapter 106). Not all patients with hypotension and tachycardia are in shock, and not all patients in shock will have hypotension and tachycardia. Patients in shock should be treated on the basis of the cause (Chapters 106 to 108).

Symptomatic hypotensive patients thought to be intravascularly volume depleted should receive intravenous fluid resuscitation with crystalloid or blood, depending on their hemoglobin level (Chapter 106). In patients with known heart disease, patients who are frail or elderly, and patients whose volume status is uncertain, small boluses of fluid (e.g., 250 mL of normal saline), each followed by reassessment, are preferred so that iatrogenic heart failure may be avoided. Inotropic support should be reserved for patients who do not respond to fluid resuscitation. High-output heart failure should be kept in mind in patients with possible thyroid storm or stimulant overdose.

Increased Respiratory Rate

Tachypnea is a normal response to hypoxemia (see later). Treatment of tachypnea in the absence of hypoxemia is directed at the underlying cause, which often is pain (Chapter 30). Anxiolytics (e.g., diazepam, 5 to 10 mg PO or IV; lorazepam, 1 to 2 mg PO, IM, or IV) or reassurance can calm patients with behavioral causes of hyperventilation. Breathing into a paper bag has been shown to be an ineffective treatment. Pulmonary embolism (Chapter 98) does not necessarily reduce the oxygen saturation or cause a low P_{O_2} and should always be considered in at-risk patients with unexplained tachypnea.

Decreased Respiratory Rate

Any perturbation of the respiratory center in the central nervous system can slow the respiratory drive (Chapter 86). Narcotics and other sedatives and neurologic conditions are common causes of a decreased respiratory rate. The primary treatment of apnea is mechanical ventilation (Chapter 105), but narcotic antagonists can be tried in patients with a history or physical examination findings (miosis, track marks, opiate patch) suggestive of narcotic use or abuse (Chapter 34). In nonapneic patients, mechanical ventilation is indicated for patients who are breathing too slowly to maintain an acceptable oxygen saturation and for patients who are retaining carbon dioxide in quantities sufficient to depress mental function. Patients who are unable to protect their airway should be intubated. Oxygen should be administered to all hypopneic patients who are hypoxemic (see earlier). Patients with chronic hypoventilation (Chapter 86) may have retained HCO_3^- to compensate for an elevated P_{CO_2} and so may depend on hypoxia to maintain respiratory drive; in these patients, overaggressive administration of oxygen can decrease the respiratory rate, increase the P_{CO_2}, and increase obtundation (Chapter 104).

Decreased Oxygen Saturation

In hypopneic patients, initial efforts should try to increase the respiratory rate (see earlier) and tidal volume. Regardless of etiology, oxygen, in amounts adequate to restore adequate oxygen saturation (P_{O_2} > 60 mm Hg, oxygen saturation >90%), is the mainstay of therapy. When oxygen alone fails, noninvasive methods for improving ventilation or tracheal intubation are required (Chapter 104). Oxygen should increase the P_{O_2} in all patients except those who have severe right-to-left shunting (Chapter 69). Treatment of conditions that cause hypoxemia includes antibiotics (pneumonia), bronchodilators (asthma, chronic obstructive pulmonary disease), diuretics and vasodilators (pulmonary edema), anticoagulants (pulmonary embolism), hyperbaric oxygen (carbon monoxide poisoning), methylene blue (methemoglobinemia, sulfhemoglobinemia), and transfusion (anemia).

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Straede M, Brabrand M. External validation of the simple clinical score and the HOTEL score, two scores for predicting short-term mortality after admission to an acute medical unit. *PLoS ONE*. 2014;9:e105695.
2. Lamantia MA, Stewart PW, Platts-Mills TF, et al. Predictive value of initial triage vital signs for critically ill older adults. *West J Emerg Med*. 2013;14:453-460.
3. Lighthall GK, Markar S, Hsiung R. Abnormal vital signs are associated with an increased risk for critical events in US veteran inpatients. *Resuscitation*. 2009;80:1264-1269.
4. Gabayan GZ, Sun BC, Asch SM, et al. Qualitative factors in patients who die shortly after emergency department discharge. *Acad Emerg Med*. 2013;20:778-785.
5. Bleyer AJ, Vidya S, Russell GB, et al. Longitudinal analysis of one million vital signs in patients in an academic medical center. *Resuscitation*. 2011;82:1387-1392.

REVIEW QUESTIONS

1. A patient presents with malaise, cough, and shortness of breath. Vital signs include temperature 40° C, blood pressure 120/74 mm Hg, respiratory rate 18 breaths per minute, pulse 70 beats per minute, and oxygen saturation 97%. This presentation could be consistent with:

A. Streptococcal pneumonia
B. Pyelonephritis due to *Escherichia coli*
C. Legionella pneumonia
D. Influenza-like illness
E. Mycoplasma pneumonia

Answer: C This patient is exhibiting a pulse-temperature dissociation because the pulse (70) is far lower than one would expect given that the patient is febrile to 40° C. This phenomenon is seen in a number of conditions, including typhoid fever and legionella infection. The other conditions would all be expected to produce tachycardia unless the patient could not become tachycardic because of medications (e.g., β-blockers) or cardiac conduction problems.

2. An 88-year-old man presents from a nursing home with slight agitation and vital signs that include temperature 38.7° C, blood pressure 96/64 mm Hg, respiratory rate 22 breaths per minute, pulse 94 beats per minute, and oxygen saturation 96%. Physical examination reveals dry mucous membranes, clear lungs, a soft abdomen, an indwelling Foley catheter, and slightly cool but noncyanotic extremities. The patient should be given:

A. Antipyretics (e.g., acetaminophen)
B. Intravenous normal saline, 500 mL with additional boluses as tolerated
C. Intravenous antibiotics
D. All of the above
E. Only A and B until urine culture results are available

Answer: D This case is an example of how vital signs can guide treatment in the absence of a firm diagnosis. The patient meets all three of the physical examination criteria for the systemic inflammatory response syndrome (SIRS) and is likely septic. The physician should not wait for his white blood cell count or other laboratory results to initiate antibiotic treatment because evidence suggests that early antibiotics are a crucial step in preventing morbidity and mortality. Although antibiotics should not be overused, the early provision of appropriate broad-spectrum antibiotics before the confirmation of a specific diagnosis is prudent and may be life-saving for this patient.

3. An intern is awakened at 3 AM by the ward nurse regarding a patient who is postoperative day 2 from a hip replacement and is newly tachycardic. Vital signs include temperature 36° C, blood pressure 146/82 mm Hg, respiratory rate 18 breaths per minute, pulse 112 beats per minute, and oxygen saturation 97% on room air. The intern drowsily orders a 1000-mL normal saline fluid challenge for dehydration. Later that morning, the patient is acutely intubated for respiratory distress. What most likely went wrong?

A. The intern failed to consider pulmonary embolism as a possible cause for the tachycardia.
B. The intern failed to consider fat embolism in a patient who had recently undergone hip surgery.
C. The intern failed to consider sepsis in the differential diagnosis.
D. The intern failed to consider failure of the patient controlled anesthesia (PCA) pump in the differential diagnosis.
E. The intern failed to realize that tachycardia can be present in both dehydration and heart failure.

Answer: E First and foremost, the intern's main mistake was not getting out of bed to evaluate the patient in person. Vital signs alone are not sufficient data on which to base an important clinical decision. Although choices A and B are certainly possible in a postoperative orthopedic patient, heart failure is a more likely diagnosis. There is little clinical support for the other choices.

4. A patient arrives in the emergency department comatose with decreased respiratory rate in the winter. Vital signs are temperature 36° C, blood pressure 128/68 mm Hg, respiratory rate 10 breaths per minute, pulse 100 beats per minute, and oxygen saturation 100% on room air. Pupils are 6 mm and reactive, and lungs are clear. What is the single most important initial treatment?

A. High-flow O_2 administered by non-rebreather mask
B. Intravenous normal saline, 1000 mL with additional boluses as tolerated
C. Intravenous antibiotics
D. Naloxone, 0.8 mg IV
E. Immediate endotracheal intubation

Answer: A This patient may have carbon monoxide poisoning. It is winter, a time when people use heating devices that may have incomplete combustion. The pupillary examination is not suggestive of opiate intoxication (D), and no other diagnosis is apparent. Because oxygen is the best treatment for this condition and is generally harmless in adults, it makes sense to initiate this therapy while efforts (e.g., blood gas analysis with co-oximetry) are made to confirm the diagnosis. There is no basis for thinking this patient is dehydrated (B) or infected (C), and intubation would be premature (E). Remember that pulse oximetry is falsely elevated in carbon monoxide poisoning, so the 100% oxygen saturation means nothing.

9 STATISTICAL INTERPRETATION OF DATA

THOMAS B. NEWMAN AND CHARLES E. MCCULLOCH

ROLE AND LIMITATIONS OF STATISTICS

Much of medicine is inherently probabilistic. Not everyone with hypercholesterolemia who is treated with a statin is prevented from having a myocardial infarction, and not everyone not treated does have one, but statins reduce the *probability* of a myocardial infarction in such patients. Because so much of medicine is based on probabilities, studies must be performed on *groups* of people to estimate these probabilities. Three component tasks of statistics are: selecting a sample of subjects for study, describing the data from that sample, and drawing inferences from that sample to a larger population of interest.[1]

SAMPLING: SELECTING SUBJECTS FOR A STUDY

The goal of research is to produce generalizable knowledge, so that measurements made by researchers on samples of individuals will eventually help draw inferences to a larger group of people than was studied. The ability to draw such inferences depends on how the subjects for the study (the sample) were selected. To understand the process of selection, it is helpful to begin by identifying the group to which the results are to be generalized and then work backward to the sample of subjects to be studied.

Target Population

The *target population* is the population to which it is hoped the results of the study will be generalizable. For example, to study the efficacy of a new drug to treat obesity, the target population might be all people with a certain level of obesity (e.g., body mass index [BMI] of ≥30 kg/m^2) who might be candidates for the drug.

Sampling

The *intended sample* is the group of people who are eligible to be in the study based on meeting *inclusion criteria,* which specify the demographic, clinical, and temporal characteristics of the intended subjects, and not meeting *exclusion criteria,* which specify the characteristics of subjects whom the investigator does not wish to study. For example, for the study of a new obesity drug, the intended sample (inclusion criteria) might be men and women 18 years or older who live in one of four metropolitan areas, who have a BMI of 30 kg/m^2 or higher, and who have failed an attempt at weight loss with a standard diet. Exclusion criteria might include an inability to speak English or Spanish, known alcohol abuse, plans to leave the area in the next 6 months, and being pregnant or planning to become pregnant in the next 6 months.

In some cases, particularly large population health surveys such as the National Health and Nutrition Examination Survey (NHANES), the intended sample is a *random* sample of the target population. A *simple random sample* is a sample in which every member of the target population has an equal chance of being selected. Simple random samples are the easiest to handle statistically but are often impractical. For example, if the target population is the entire population of the United States (as is the case for NHANES), a simple random sample would include subjects from all over the country. Getting subjects from thousands of distinct geographic areas to examination sites would be logistically difficult. An alternative, used in NHANES, is *cluster* sampling, in which investigators take a random sample of "clusters" (e.g., specific census tracts or geographic areas) and then try to study all or a sample of the subjects in each cluster. Knowledge of the cluster sampling process must then be used during analysis of the study (see later) to draw inferences correctly back to the target population.

Regardless of the method used to select the intended sample, the *actual sample* will almost always differ in important ways because not all intended subjects will be willing to enroll in the study and not all who begin a study will complete it. In a study on treatment of obesity, for example, those who consent to be in the study probably differ in important, but difficult-to-quantify ways from those who do not (and may be more likely to do well with treatment). Furthermore, subjects who respond poorly to treatment may drop out, thus making the group that completes the study even less representative.

Statistical methods address only some of the issues involved in making inferences from a sample to a target population. Specifically, *most statistical methods address only the effect of random variation on the inference from the intended sample to the target population.* Estimating the effects of differences between the intended sample and the actual sample depends on the quantities being estimated and content knowledge about whether factors associated with being in the actual sample are related to those quantities. One rule of thumb about generalizability is that *associations between variables* are more often generalizable than measurements of single variables. For instance, subjects who consent to be in a study of obesity may be more motivated than average, but this motivation would be expected to have less effect on the *difference* in weight loss between groups than on the average weight loss in either group.

DESCRIBING THE SAMPLE

Types of Variables

A key use of statistics is to describe sample data. Methods of description depend on the *type of variable* (E-Table 9-1). *Numerical* variables include *continuous* variables (those that have a wide range of possible values), *count* variables (e.g., the number of times a woman has been pregnant), and *time-to-event* variables (e.g., the time from initial treatment to recurrence of breast cancer). Whereas *numerical* variables describe the data with numbers, *categorical* variables consist of named characteristics. Categorical variables can be further divided into *dichotomous* variables, which can take on only two possible values (e.g., alive/dead); *nominal* variables, which can take on more than two values but have no intrinsic ordering (e.g., race); and *ordinal* variables, which have more than two values and an intrinsic ordering of the values (e.g., tumor stage). Numerical variables are also ordinal by nature and can be made binary by breaking the values into two disjointed categories (e.g., systolic blood pressure >140 mm Hg or not), and thus sometimes methods designed for ordinal or binary data are used with numerical variable types, often for ease of interpretation.

Univariate Statistics for Numerical Variables: Mean, Standard Deviation, Median, and Percentiles

When describing data in a sample, it is a good idea to begin with *univariate* (one variable at a time) statistics. For numerical variables, univariate statistics typically measure *central tendency* and *variability.* The most common measures of central tendency are the *mean* (or average, i.e., the sum of the observations divided by the number of observations) and the *median* (the 50th percentile, i.e., the value that has equal numbers of observations above and below it).

One of the most commonly used measures of variability is the *standard deviation* (SD). The SD is defined as the square root of the *variance,* which is calculated by subtracting each value in the sample from the mean, squaring that difference, totaling all of the squared differences, and dividing by the number of observations minus 1. Although this definition is far from intuitive, the SD has some useful mathematical properties, namely, that if the distribution of the variable is the familiar bell-shaped, *normal,* or *Gaussian* distribution, about 68% of the observations will be within 1 SD of the mean, about 95% within 2 SD, and about 99.7% within 3 SD. Even when the distribution is not normal, these rules are often approximately true.

For variables that are not normally distributed, including most count and time-to-event variables, the mean and SD are not as useful for summarizing the data. In that case, the median may be a better measure of central tendency because it is not influenced by observations far below or far above the center. Similarly, the range and pairs of percentiles, such as the 25th and 75th percentiles or the 15th and 85th percentiles, will provide a better description of the spread of the data than the SD will. The 15th and 85th percentiles are particularly attractive because they correspond, in the Gaussian distribution, to about −1 and +1 SD from the mean, thus making reporting of the 50th, 15th, and 85th percentiles roughly equivalent to reporting the mean and SD.

Univariate Statistics for Categorical Variables: Proportions, Rates, and Ratios

For categorical variables, the main univariate statistic is the *proportion* of subjects with each value of the variable. For dichotomous variables, only one proportion is needed (e.g., the proportion female); for nominal variables and ordinal variables with few categories, the proportion in each group can be provided. Ordinal variables with many categories can be summarized by

E-TABLE 9-1 TYPES OF VARIABLES AND COMMONLY USED STATISTICAL METHODS

		ASSOCIATED STATISTICAL METHODS	
TYPE OF OUTCOME VARIABLE	**EXAMPLES**	**BIVARIATE**	**MULTIVARIATE**
Categorical (dichotomous)	Alive; readmission to the hospital within 30 days	2×2 table, chi-square analysis	Logistic regression
Categorical (nominal)	Race; cancer, tumor type	Chi-square analysis	Nominal logistic regression
Categorical (ordinal)	Glasgow Coma Scale	Mann-Whitney-Wilcoxon, Kruskal-Wallis	Ordinal logistic regression
Numerical (continuous)	Cholesterol; SF-36 scales*	*t* Test, analysis of variance	Linear regression
Numerical (count)	Number of times pregnant; number of mental health visits in a year	Mann-Whitney-Wilcoxon, Kruskal-Wallis	Poisson regression, linear models
Time to event regression	Time to breast cancer; time to viral rebound in HIV-positive subjects	Log rank	Cox proportional hazards

*Numerical scores with many values are often treated as though they were continuous. HIV = human immunodeficiency virus; SF-36 = short-form 36-item health survey.

using proportions or by using medians and percentiles, as with continuous data that are not normally distributed.

It is worth distinguishing among *proportions, rates,* and *ratios* because these terms are often confused. *Proportions* are unitless, always between 0 and 1 inclusive, and express what fraction of the subjects have or develop a particular characteristic or outcome. Strictly speaking, *rates* have units of inverse time; they express the proportion of subjects in whom a particular characteristic or outcome develops over a specific time period. The term is frequently misused, however. For example, the term *false-positive rate* is widely used for the proportion of subjects without a disease who test positive, even though it is a proportion, not a rate. *Ratios* are the quotients of two numbers; they can range between zero and infinity. For example, the male-to-female ratio of people with a disease might be 3 : 1. As a rule, if a ratio can be expressed as a proportion instead (e.g., 75% male), it is more concise and easier to understand.

Incidence and Prevalence

Two terms commonly used (and misused) in medicine and public health are *incidence* and *prevalence. Incidence* describes the number of subjects who *contract* a disease *over time* divided by the population at risk. Incidence is usually expressed as a rate (e.g., 7 per 1000 per year), but it may sometimes be a proportion if the time variable is otherwise understood or clear, as in the lifetime incidence of breast cancer or the incidence of diabetes during pregnancy. *Prevalence* describes the number of subjects who *have* a disease at *one point in time* divided by the population at risk; it is always a proportion. At any point in time, the prevalence of disease depends on how many people contract it and how long it lasts: prevalence = incidence × duration.

Bivariate Statistics

Bivariate statistics summarize the relationship between two variables. In clinical research, it is often desirable to distinguish between *predictor* and *outcome variables.* Predictor variables include treatments received, demographic variables, and test results that are thought possibly to predict or cause the *outcome variable,* which is the disease or (generally bad) event or outcome that the test should predict or treatment prevent. For example, to see whether a bone mineral density measurement (the predictor) predicts time to vertebral fracture (the outcome), the choice of bivariate statistic to assess the association of outcome with predictor depends on the types of predictor and outcome variables being compared.

Dichotomous Predictor and Outcome Variables

A common and straightforward case is when both predictor and outcome variables are dichotomous, and the results can thus be summarized in a 2 × 2 table. Bivariate statistics are also called *measures of association* (E-Table 9-2).

Relative Risk

The *relative risk* or *risk ratio* (RR) is the ratio of the proportion of subjects in one group in whom the outcome develops divided by the proportion in the other group in whom it develops. It is a general (but not universal) convention to have the outcome be something bad and to have the numerator be the risk for those who have a particular factor or were exposed to an intervention. When this convention is followed, an RR greater than 1 means that exposure to the factor was (on average) bad for the study subjects (with respect to the outcome being studied), whereas an RR less than 1 means that it was good. That is, risk factors that cause diseases will have RR values greater than 1, and effective treatments will have an RR less than 1. For example, in the Women's Health Initiative (WHI) randomized trial, conjugated equine estrogen use was associated with an increased risk for stroke (RR = 1.37) and decreased risk for hip fracture (RR = 0.61).

Relative Risk Reduction

The *relative risk reduction* (RRR) is 1 − RR. The RRR is generally used only for effective interventions, that is, interventions in which the RR is less than 1, so the RRR is generally greater than 0. In the aforementioned WHI example, estrogen had an RR of 0.61 for hip fracture, so the RRR would be 1 − 0.61 = 0.39, or 39%. The RRR is commonly expressed as a percentage and used only when it is positive.

Absolute Risk Reduction

The *risk difference* or *absolute risk reduction* (ARR) is the difference in risk between the groups, defined as earlier. In the WHI, the risk for hip fracture was 0.11% per year with estrogen and 0.17% per year with placebo. Again, conventionally the risk is for something bad, and the risk in the group of interest is subtracted from the risk in a comparison group, so the ARR will be positive for effective interventions. In this case, the ARR = 0.06% per year, or 6 in 10,000 per year.

Number Needed to Treat

The *number needed to treat* (NNT) is 1/ARR. To see why this is the case, consider the WHI placebo group and imagine treating 10,000 patients for a year. All but 17 would not have had a hip fracture anyway because the fracture rate in the placebo group was 0.17% per year, and 11 subjects would sustain a fracture despite treatment because the fracture rate in the estrogen group was 0.11% per year. Thus, with treatment of 10,000 patients for a year, 17 − 11 = 6 fractures prevented, or 1 fracture prevented for each 1667 patients treated for 1 year. This calculation is equivalent to 1/0.06% per year.

Risk Difference

When the treatment *increases* the risk for a bad outcome, the difference in risk between treated and untreated patients should still be calculated, but it is usually just called the risk difference rather than an ARR (because the "reduction" would be negative). In that case, the NNT is sometimes called the number needed to harm. This term is a bit of a misnomer. The reciprocal of the risk difference is still a number needed to treat; it is just a number needed to treat per person harmed rather than a number needed to treat per person who benefits. In the WHI, treatment with estrogens was estimated to cause about 12 additional strokes per 10,000 women per year, so the number needed to be treated for 1 year to cause a stroke was about 10,000/12, or 833.

Odds Ratio

Another commonly used measure of association is the *odds ratio* (OR). The OR is the ratio of the *odds* of the outcome in the two groups, where the definition of the odds of an outcome is $p/(1-p)$, with p being the probability of the outcome. From this definition it is apparent that when p is very small, $1-p$ will be close to 1, so $p/(1-p)$ will be close to p, and the OR will closely approximate the RR. In the WHI, the ORs for stroke (1.37) and fracture (0.61) were virtually identical to the RRs because both stroke and fracture were rare. When p is not small, however, the odds and probability will be quite different, and ORs and RRs will not be interchangeable.

Absolute versus Relative Measures

RRRs are usually more generalizable than ARRs. For example, the use of statin drugs is associated with about a 30% decrease in coronary events in a wide variety of patient populations (Chapter 206). The ARR, however, will usually vary with the baseline risk, that is, the risk for a coronary event in the absence of treatment. For high-risk men who have already had a myocardial infarction, the baseline 5-year risk might be 20%, which could be reduced to 14% with treatment, an ARR of 6%, and an NNT of about 17 for approximately 5 years. Conversely, for a 45-year-old woman with a high low-density lipoprotein cholesterol level but no history of heart disease, in whom the 5-year risk might be closer to 1%, the same RRR would give a 0.7% risk with treatment, a risk difference of 0.3%, and an NNT of 333 for 5 years.

The choice of *absolute* versus *relative measures* of association depends on the intended use of the measure. As noted earlier, RRs are more useful as summary measures of effect because they are more often generalizable across a wide variety of populations. RRs are also more helpful for understanding causality. However, absolute risks are more important for questions about clinical decision making because they relate directly to the tradeoffs between risks and benefits—specifically, the NNT, as well as the costs and side effects that need to be balanced against potential benefits. RRRs are often used in advertising because they are generally more impressive than ARRs. Unfortunately, the distinction between relative and absolute risks may not be appreciated by clinicians, thereby leading to higher estimates of the potential benefits of treatments when RRs or RRRs are used.

Risk Ratios versus Odds Ratios

The choice between RRs and ORs is easier: RRs are preferred because they are easier to understand. Because ORs that are not equal to 1 are always farther from 1 than the corresponding RR, they may falsely inflate the perceived importance of a factor. ORs are, however, typically used in two circumstances. First, in case-control studies (Chapter 11), in which subjects with and without the disease are sampled separately, the RR cannot be

E-TABLE 9-2 COMMONLY USED MEASURES OF ASSOCIATION FOR DICHOTOMOUS PREDICTOR AND OUTCOME VARIABLES*

	OUTCOME		
PREDICTOR	***Yes***	***No***	***Total***
Yes	a	b	**a + b**
No	c	d	**c + d**
Total	**a + c**	**b + d**	***N***
Risk ratio or relative risk (RR)	$\frac{a}{(a+b)} \div \frac{c}{(c+d)}$		
Relative risk reduction (RRR)	1 – RR		
Risk difference or absolute risk reduction (ARR)	$\frac{a}{(a+b)} - \frac{c}{(c+d)}$		
Number needed to treat (NNT)	$\frac{1}{ARR}$		
Odds ratio (OR)	$\frac{ad}{bc}$		

*The numbers of subjects in each of the cells are represented by a, b, c, and d. Case-control studies allow calculation of only the odds ratio.

calculated directly. This situation does not usually cause a problem, however, because case-control studies are generally performed to assess rare outcomes, for which the OR will closely approximate the RR. Second, in observational studies that use a type of multivariate analysis called *logistic regression* (see later), use of the OR is convenient because it is the parameter that is modeled in the analysis.

Dichotomous Predictor Variable, Continuous Outcome Variable

Many outcome variables are naturally continuous rather than dichotomous. For example, in a study of a new treatment of obesity, the outcome might be change in weight or BMI. For a new diuretic, the outcome might be change in blood pressure. For a palliative treatment, the outcome might be a quality-of-life score calculated from a multi-item questionnaire. Because of the many possible values for the score, it may be analyzed as a continuous variable. In these cases, dichotomizing the outcome leads to loss of information. Instead, the *mean difference* between the two groups is an appropriate measure of the effect size. When the outcome is itself a difference (e.g., change in blood pressure over time), the effect is measured by the difference in the *within*-group differences *between* the groups.

Most measurements have units (e.g., kg, mm Hg), so differences between groups will have the same units and be meaningless without them. If the units of measurement are familiar (e.g., kg or mm Hg), the difference between groups will be meaningful without further manipulation. For measurements in unfamiliar units, such as a score on a new quality-of-life instrument, some benchmark is useful to help judge whether the difference between groups is large or small. In that case, authors typically express the difference in relation to the spread of values in the study by calculating the *standardized mean difference* (SMD), which is the difference between the two means divided by the SD of the measurement. It is thus expressed as the number of SDs by which the two groups are apart. To help provide a rough feel for this difference, a 1-SD difference between means (SMD = 1) would be a 15-point difference in IQ scores, a 600-g difference in birthweight, or a 40-mg/dL difference in total cholesterol levels.

Continuous Predictor Variable

When predictor variables are continuous, the investigator can either group the values into two or more categories and calculate mean differences or SMDs between the groups as discussed earlier or use a *model* to summarize the degree to which changes in the predictor variable are associated with changes in the outcome variable. Use of a model may more compactly describe the effects of interest but involves assumptions about the way the predictor and outcome variables are related. Perhaps the simplest model is to assume a linear relationship between the outcome and predictor. For example, one could assume that the relationship between systolic blood pressure (mm Hg) and salt intake (g/day) was linear over the range studied:

$$\mathrm{SBP}_i = a + (b \times \mathrm{SALT}_i) + \varepsilon_i$$

where SBP_i is the systolic blood pressure for study subject *i*, SALT_i is that subject's salt intake, and ε_i is an error term that the model specifies must average out to zero across all of the subjects in the study. In this model, *a* is a constant, the *intercept,* and the strength of the relationship between the outcome and predictor can be summarized by the slope *b,* which has units equal to the units of SBP divided by the units of SALT, or mm Hg per gram of salt per day in this case.

Note that without the units, such a number is meaningless. For example, if salt intake were measured in grams per week instead of grams per day, the slope would only be one seventh as large. Thus, when reading an article in which the association between two variables is summarized, it is critical to note the units of the variables. As discussed earlier, when units are unfamiliar, they are sometimes standardized by dividing by the SDs of one or both variables.

It is important to keep in mind that use of a model to summarize a relationship between two variables may not be appropriate if the model does not fit. In the preceding example, the assumption is that salt intake and blood pressure have a linear relationship, with the slope equal to *b* mm Hg/g salt per day. The value of *b* is about 1 mm Hg/g salt per day for hypertensive patients. If the range of salt intake of interest is from 1 to 10 g/day, the model predicts that blood pressure will increase 1 mm Hg as a result of a 1-g/day increase in salt intake whether that increase is from 1 to 2 g/day or from 9 to 10 g/day. If the effect of a 1-g/day change in salt intake differed in subjects ingesting low- and high-salt diets, the model would not fit, and misleading conclusions could result.

When the outcome variable is dichotomous, the relationship between the probability of the outcome and a continuous predictor variable is often modeled with a *logistic* model:

$$\Pr\{Y_i = 1\} = \frac{1}{1 + e^{-(a + bx_i)}}$$

where the outcome Y_i is coded 0 or 1 for study subject *i,* and x_i is that subject's value of the predictor variable. Once again, *a* is a constant, in this case related to the probability of the disease when the predictor is equal to zero, and *b* summarizes the strength of the association; in this case, it is the natural logarithm of the OR rather than the slope. The OR is the OR *per unit change* in the predictor variable. For example, in a study of lung cancer, an OR of 1.06 for pack years of smoking would indicate that the odds of lung cancer increase by 6% for each pack year increase in smoking.

Because the outcome variable is dichotomous, it has no units, and "standardizing" it by dividing by its SD is unnecessary and counterproductive. On the other hand, continuous predictor variables do have units, and the OR for the logistic model will be per unit change in the predictor variable or, if standardized, per SD change in the predictor variable. Re-expressing predictors in standardized or at least more sensible units is often necessary. For example, suppose 10-year mortality risk decreases by 20% (i.e., RR = 0.8) for each increase in gross income of $10,000. The RR associated with an increase in gross income of $1 (which is what a computer program would report if the predictor were entered in dollars) would be 0.99998, apparently no effect at all because a change of $1 in gross income is negligible and associated with a negligible change in risk. To derive the coefficient associated with a $1 change, the coefficient for a $10,000 change is raised to the $^{1}/_{10,000}$ power: $0.8^{(1/10{,}000)} = 0.99998$.

Multivariable Statistics

In many cases, researchers are interested in the effects of multiple predictor variables on an outcome. Particularly in observational studies, in which investigators cannot assign values of a predictor variable experimentally, it will be of interest to estimate the effects of a predictor variable of interest *independent* of the effects of other variables. For example, in studying whether breastfeeding reduces the mother's risk for subsequent breast cancer, investigators would try to take differences in age, race, family history, and parity into account. Trying to stratify by all these variables would require a massive data set. Instead, models are used because they enable the information about individual predictors to be summarized by using the full data set. In this way, the estimated coefficients from the model are powerful descriptive statistics that allow a sense of the data in situations in which simpler methods fail. These models are similar to those described earlier but include terms for the additional variables.

Multiple Linear Regression

The multiple linear regression model for an outcome variable Y as function or predictor variables x_1, x_2, and so forth is as follows:

$$Y_i = a + (b_1 \times x_{1i}) + (b_2 \times x_{2i}) + \ldots + (b_k \times x_{ki}) + \varepsilon_i,$$

where the subscripts 1, 2, …, k are for the first, second, … k^{th} variables of the model, and the *i* subscripts are for each individual. As before, the relationships between each of these predictor variables and the outcome variable are summarized by coefficients, or slopes, which have units of Y divided by the units of the associated predictor. In addition, the linear combination of predictor variables adds a major simplifying constraint (and assumption) to the model: it specifies that the effects of each variable on the outcome variable are the same regardless of the values of other variables in the model. Thus, for example, if x_1 is the variable for salt intake and x_2 is a variable for sex (e.g., 0 for females and 1 for males), this model assumes that the average effect of a 1-g increase in daily salt intake on blood pressure is the same in men and women. If such is not believed to be the case, either based on previous information or from examining the data, the model should include *interaction* terms, or separate models should be used for men and women.

Multiple Logistic Regression

The logistic model expands to include multiple variables in much the same way as the linear model:

$$\Pr\{Y_i = 1\} = \frac{1}{1 + e^{-(a + b_1 i + b_2 x_{2i} + \ldots + b_k x_{ki})}}$$

Again, the additional assumption when more than one predictor is included in the model is that in the absence of included interaction terms,

the effect of each variable on the odds of the outcome is the same regardless of the values of other variables in the model. Because the logistic model is multiplicative, however, the effects of different predictors on the odds of the outcome are multiplied, not added. Thus, for example, if male sex is associated with a doubling of the odds for heart disease, this doubling will occur in both smokers and nonsmokers; if smoking triples the odds, this tripling will be true in both men and women, so smoking men would be predicted to have $2 \times 3 = 6$ times higher odds of heart disease than nonsmoking women.

Recursive Partitioning

Recursive partitioning, or "*classification and regression trees,*" is a prediction method often used with dichotomous outcomes that avoids the assumptions of linearity. This technique creates prediction rules by repeatedly dividing the sample into subgroups, with each subdivision being formed by further separating the sample on the value of one of the predictor variables. The optimal choice of variables and cut points may depend on the relative costs of false-positive and false-negative predictions, as set by the investigator. The end result is a set of branching questions that forms a treelike structure in which each final branch provides a yes/no prediction of the outcome. The methods of fitting the tree to data (e.g., cross-validation) help reduce overfitting (inclusion of unnecessary predictor variables), especially in cases with many potential predictors.

Proportional Hazards (Cox) Model

A multivariate model often used in studies in which subjects are monitored over time for development of the outcome is the *Cox* or *proportional hazards* model. Like the logistic model, the Cox model is used for continuous or dichotomous predictor variables, but in this case with a time-to-event outcome (e.g., time to a stroke). This approach models the *rate* at which the outcome occurs over time by taking into account the number of people still at risk at any given time. The coefficients in the Cox model are logarithms of *hazard ratios* rather than ORs, interpretable (when exponentiated) as the effect of a unit change in predictors on the *hazard* (risk in the next short time period) of the outcome developing. Like the logistic model, the Cox model is multiplicative; that is, it assumes that changes in risk factors multiply the hazard by a fixed amount regardless of the levels of other risk factors. A key feature of the Cox model and other *survival analysis* techniques is that they accommodate censored data (when the time to event is known only to exceed a certain value). For example, if the outcome is time to stroke, the study will end with many subjects who have not had a stroke, so their time to stroke is known only to exceed the time to their last follow-up visit.

INFERRING POPULATION VALUES FROM A SAMPLE

The next step after describing the data is drawing inferences from a sample to the population from which the sample was drawn. Statistics mainly quantify random error, which arises by chance because even a sample randomly selected from a population may not be exactly like the population from which it was drawn. Samples that were not randomly selected from populations may be unrepresentative because of *bias,* and statistics cannot help with this type of systematic (nonrandom) error.

Inferences from Sample Means: Standard Deviation versus Standard Error

The simplest case of inference from a sample to a population involves estimating a population mean from a sample mean. Intuitively, the larger the sample size, *N*, the more likely it will be that the sample mean will be close to the population mean, that is, close to the mean that would be calculated if every member of the population were studied. The more variability there is in the population (and hence the sample), the less accurate the sample estimate of the population mean is likely to be. Thus, the precision with which a population mean can be estimated is related to both the size of the sample and the SD of the sample. To make inferences about a population mean from a sample mean, the *standard error of the mean* (SEM), which takes both of these factors into account, is as follows:

$$\text{SEM} = \frac{\text{SD}}{\sqrt{N}}$$

To understand the meaning of the SEM, imagine that instead of taking a single sample of *N* subjects from the population, many such samples were taken. The mean of each sample could be calculated, as could the mean of those sample means and the SD of these means. The SEM is the best estimate from a single sample of what that SD of sample means would be.

Confidence Intervals

The SEM expresses variability of sample means in the same way that the SD expresses variability of individual observations. Just as about 95% of *observations* in a population are expected to be within ±1.96 SD of the mean, 95% of *sample means* are expected to be within 1.96 SEM of the population mean, thereby providing the 95% confidence interval (CI), which is the range of values for the population mean consistent with what was observed from the sample.

CIs can similarly be calculated for other quantities estimated from samples, including proportions, ORs, RRs, regression coefficients, and hazard ratios. In each case, they provide a range of values for the parameter in the target population consistent with what was observed in the study sample.

Significance Testing and *P* Values

Many papers in the medical literature include *P* values, but the meaning of *P* values is widely misunderstood and mistaught. *P* values start with calculation of a *test statistic* from the sample that has a known distribution under certain assumptions, most commonly the *null hypothesis,* which states that there is no association between variables. *P* values provide the answer to the question, "If the null hypothesis were true, what would be the probability of obtaining, by chance alone, a value of the test statistic this large or larger (suggesting an association between groups of this strength or stronger)?" When the *P* value is small, there are two possible explanations. First, something with a small possibility of occurring actually happened; or second, the null hypothesis is false, and there is a true association. Values of *P* less than 0.05 are customarily described as "statistically significant."

There are a number of common pitfalls in interpreting *P* values. The first is that because *P* values less than .05 are customarily described as being "statistically significant," the description of results with *P* values less than .05 sometimes gets shortened to "significant" when in fact the results may not be clinically significant (i.e., important) at all. A lack of congruence between clinical and statistical significance most commonly arises when studies have a large sample size and the measurement is of a continuous or frequently occurring outcome.

A second pitfall is concluding that no association exists simply because the *P* value is greater than .05. However, it is possible that a real association exists, but that it simply was not found in the study. This problem is particularly likely if the sample size is small because small studies have low *power,* defined as the probability of obtaining statistically significant results if there really is a given magnitude of difference between groups in the population. One approach to interpreting a study with a nonsignificant *P* value is to examine the power that the study had to find a difference. A better approach is to look at the 95% CI. If the 95% CI excludes all clinically significant levels of the strength of an association, the study probably had an adequate sample size to find an association if there had been one. If not, a clinically significant effect may have been missed. In "negative" studies, the use of CIs is more helpful than power analyses because CIs incorporate information from the study's results.

Finally, a common misconception about *P* values is that they indicate the probability that the null hypothesis is true (e.g., that there is no association between variables). Thus, it is not uncommon to hear or read that a *P* value less than .05 implies at least a 95% probability that the observed association is not due to chance. This statement represents a fundamental misunderstanding of *P* values. Calculation of *P* values is *based on the assumption* that the null hypothesis is true. The probability that an association is real depends not just on the probability of its occurrence under the null hypothesis but also on the probability of another basis for the association (see later)—an assessment that depends on information from outside the study, sometimes called the *prior probability* of an association (of a certain magnitude) estimated before the study results were known and requiring a different approach to statistical inference. Similarly, CIs do not take into account previous information on the probable range of the parameter being estimated. Bayesian methods, which explicitly combine prior knowledge with new information, are beginning to enter the mainstream medical literature.[2]

Appropriate test statistics and methods for calculating *P* values depend on the type of variable, just as with descriptive statistics (see E-Table 9-1). For example, to test the hypothesis that the mean values of a continuous variable are equal in two groups, a *t* test would be used; to compare the mean values across multiple groups, analysis of variance would be used. Because there are many different ways for the null hypothesis to be false (i.e., many different ways that two variables might be associated) and many test statistics that could be calculated, there are many different ways of calculating a *P* value for

the association of the same two variables in a data set, and they may not all give the same answer.

Meta-analysis

Statistical techniques for inferring population values from a sample are not restricted to samples of individuals. *Meta-analysis* is a statistical method for drawing inferences from a sample of *studies* to derive a summary estimate and confidence interval for a parameter measured by the included studies, such as a risk ratio for a treatment effect.[3] Meta-analysis allows the formal combination of results while estimating and accommodating both the within-study and between-study variations. Meta-analysis is most often done when raw data from the studies are not available, as is typically the case when synthesizing information from multiple published results. For example, the previously cited estimate that a 1-g/day change in salt intake is associated with a 1-mm Hg change in blood pressure was obtained from a meta-analysis of randomized trials of low-salt diets in adults.

INFERRING CAUSALITY

In many cases, a goal of clinical research is not just to identify associations but also to determine whether they are *causal,* that is, whether the predictor causes the outcome. Thus, if people who take vitamin E live longer than those who do not, it is important to know whether it is *because* they took the vitamin or for some other reason.

Determination of causality is based on considering alternative explanations for an association between two variables and trying to exclude or confirm these alternative explanations. The alternatives to a causal relationship between predictor and outcome variables are *chance, bias, effect-cause,* and *confounding. P* values and CIs help assess the likelihood of *chance* as the basis for an association. *Bias* occurs when systematic errors in sampling or measurements can lead to distorted estimates of an association. For example, if those making measurements of the outcome variable are not blinded to values of the predictor variable, they may measure the outcome variable differently in subjects with different values of the predictor variable, thereby distorting the association between outcome and predictor.

Effect-cause is a particular problem in cross-sectional studies, in which (in contrast to longitudinal studies) all measurements are made at a single point in time, thereby precluding demonstration that the predictor variable preceded the outcome—an important part of demonstrating causality. Sometimes biology provides clear guidance about the direction of causality. For example, in a cross-sectional study relating levels of urinary cotinine (a measure of exposure to tobacco smoke) to decreases in pulmonary function, it is hard to imagine that poor pulmonary function caused people to be exposed to smoke. Conversely, sometimes inferring causality is more difficult: are people overweight because they exercise less, or do they exercise less because they are overweight (or both)?

Confounding

Confounding can occur when one or more extraneous variables are associated with both the predictor of interest and the outcome. For example, observational studies suggested that high doses of vitamin E might decrease the risk for heart disease. However, this association seems to have been largely due to confounding: people who took vitamin E were different in other ways from those who did not, including differences in factors causally related to coronary heart disease. If such factors are known and can be measured accurately, one way to reduce confounding is to *stratify* or *match* on these variables. The idea is to assemble groups of people who did and did not take vitamin E but who were similar in other ways. Multivariate analysis can accomplish the same goal—other measured variables are held constant statistically, and the effect of the variable of interest (in this case the use of vitamin E) can be examined. Multivariate analysis has the advantage that it can control simultaneously for more potentially confounding variables than can be considered with stratification or matching, but it has the disadvantage that a model must be created (see earlier), and this model may not fit the data well.

A new technique that is less dependent on model fit but still requires accurate measurements of confounding variables is the use of *propensity scores.* Propensity scores are used to assemble comparable groups in the same way as stratification or matching, but in this case the comparability is achieved on the basis of the *propensity* to be exposed to or be treated with the predictor variable of primary interest. Although propensity scores can adjust only for known confounders and are more subject to manipulation by investigators, systematic reviews suggest that they usually give answers that are generally similar to those of randomized trials addressing the same question.[4]

A major limitation of these methods of controlling for confounding is that the confounders must be known to the investigators and accurately measured. In the case of vitamin E, apparent favorable effects persisted after controlling for known confounding variables. It is for this reason that randomized trials provide the strongest evidence for causality. If the predictor variable of interest can be randomly assigned, confounding variables, both known and unknown, should be approximately equally distributed between the subjects who are and are not exposed to the predictor variable, and it is reasonable to infer that any significant differences in outcome that remain in these now comparable groups would be due to differences in the predictor variable of interest. In the case of vitamin E, a recent meta-analysis of randomized trials found no benefit and in fact suggested harm from high doses.

OTHER COMMON STATISTICAL PITFALLS

Missing Data

Research on human subjects is challenging. People drop out of studies, refuse to answer questions, miss study visits, and die of diseases that are not being studied directly in the protocol. Consequently, missing or incomplete data are a fact of medical research. When the particular data that are missing are unrelated to the outcome being studied (which might be true, for example, if the files storing the data got partially corrupted), analyses using only the data present (sometimes called a complete case analysis) are unlikely to be misleading. Unfortunately, such is rarely the case. Subjects refusing to divulge family income probably have atypical values, patients not coming for scheduled visits in a study of depression may be more or less depressed, and patients in an osteoporosis study who die of heart disease probably differ in many ways from those who do not.

Whenever a sizable fraction of the data is missing (certainly if it is above 10 or 15%), there is the danger of substantial bias from an analysis that uses only the complete data. This is the gap noted earlier between the intended and actual samples. Any study with substantial missing data should be clear about how many missing data there were and what was done to assess or alleviate the impact; otherwise, the critical consumer of such information should be suspicious. Multiple imputation is a technique of using observations with nonmissing data to estimate missing values; it can produce less biased estimates than simply excluding the observations with missing data.[5] In a randomized trial, the general rule is that the primary analysis should include all subjects who were randomized, regardless of whether they followed the study protocol, in an *intention-to-treat* analysis.

Clustered or Hierarchical Data

Data are often collected in a clustered (also called hierarchical) manner; for example, NHANES used a cluster sample survey, and a study of patient outcomes might be conducted at five hospitals, each with multiple admission teams. The cluster sample or the clustering of patients within teams within hospitals leads to correlated data. Said another way, and other things being equal, data collected on the same patient, by the same admission team, or in the same cluster are likely to be more similar than data from different patients, teams, or clusters. Failure to use statistical methods that accommodate correlated data can seriously misstate standard errors, widths of CIs, and *P* values, most often leading to overly optimistic estimates, that is, standard errors and *P* values that are incorrectly too small and CIs that are incorrectly too narrow. Statistical methods for dealing with correlated data include *generalized estimating equations* and the use of *robust standard errors* and *frailty models* (for time-to-event data). Studies with obvious hierarchical structure that fail to use such methods may be in serious error.

Multiple Hypothesis Testing

The "multiple hypothesis testing" or "multiple comparisons" issue refers to the idea that if multiple statistical tests are conducted, each at a significance level of .05, the chance that at least one of them will achieve a *P* value of less than .05 is considerably larger than .05, even when all the null hypotheses are true. For example, when comparing the mean value of a continuous variable across many different groups, analysis of variance is a time-tested method of performing an overall test of equality and avoiding making a large number of pairwise comparisons.

Because most medical studies collect data on a large number of predictor variables, performing a test on the association of each one with the outcome may generate false-positive results. The risk for falsely positive results is especially high with genomic studies, in which a researcher may test a million single-nucleotide polymorphisms for association with a disease.

A typical method for dealing with the problem of multiple testing is the *Bonferroni correction,* which specifies that the *P* value at which the null hypothesis will be rejected (e.g., .05) should be *divided by the number of hypothesis tests performed*. Although simple to use, a problem with this approach is that it is overly conservative. Studies with many listed or apparent outcomes or predictors (or both) are subject to inflation of the error rate to well above the nominal .05. Automated stepwise regression methods for choosing predictors in regression models typically *do not* alleviate and may exacerbate this problem. If no adjustment or method for dealing with multiple comparisons is used, the high probability of false-positive results should be kept in mind.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Newman TB, Kohn MA. Evidence-Based Diagnosis. New York: Cambridge University Press; 2009. *A practical text for clinicians*.
2. Goodman SN. Bayesian methods for evidence evaluation: are we there yet? *Circulation*. 2013;127: 2367-2369.
3. Murad MH, Montori VM, Ioannidis JP, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. *JAMA*. 2014;312:171-179.
4. Lonjon G, Boutron I, Trinquart L, et al. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. *Ann Surg*. 2014;259:18-25.
5. Cummings P. Missing data and multiple imputation. *JAMA Pediatr*. 2013;167:656-661.

REVIEW QUESTIONS

1. A study of 105 vegan Buddhist nuns randomly sampled from monasteries around Ho Chi Minh City found that the average femoral neck bone mineral density was 0.62 g/cm^2, with a standard deviation of 0.11 g/cm^2 (Ho-Pham LT, Nguyen PL, Le TT, et al. Veganism, bone mineral density, and body composition: a study in Buddhist nuns. *Osteoporos Int.* 2009;20:2087-2093). Which of the following statements about this result is correct?

A. The 95% confidence interval for the mean bone mineral density in these women is about 0.4 to 0.84 g/cm^2.
B. If bone mineral density is normally distributed, we would expect about 10% of the women in the sample to have bone mineral density outside of the interval: 0.4 to 0.84 g/cm^2.
C. The 95% confidence interval for the mean bone mineral density in these women is about 0.60 to 0.64 g/cm^2.
D. Because the women were sampled randomly, there is a 95% chance that a randomly selected woman from the *population* would have a bone mineral density between 0.60 and 0.64 g/cm^2.
E. Because the women were sampled randomly, there is a 95% chance that a randomly selected woman from the *sample* would have a bone mineral density between 0.60 and 0.64 g/cm^2.

Answer: C We would expect about 95% of *observations* to be within 2 standard deviations of the sample mean, leaving 5% out, so choice B is incorrect. The 95% confidence interval for a sample mean is about mean ± 2 standard errors of the mean (SEM), where the $\text{SEM} = \frac{\text{SD}}{\sqrt{N}}$. In this case, the SD is 0.11 g/cm^2, and the N is about 100, so the SEM will be about 0.11/10 = 0.01 g/cm^2, and the 95% CI will be about 0.60 to 0.64 g/cm^2, as indicated in choice C. Choices D and E are incorrect because the range given is too narrow: it is ± 2 SEM when it should be ± 2 SD.

2. A study of data collected through the Get with the Guidelines-Stroke Program examined the time from onset of stroke symptoms to treatment with tissue-type plasminogen activator (tPA) among 58,353 patients with acute ischemic stroke treated within 4.5 hours of the onset of symptoms (Saver JL, Fonarow GC, Smith EE, et al. Time to treatment with intravenous tissue plasminogen activator and outcome from acute ischemic stroke. *JAMA.* 2013;309:2480-2488). The authors reported that "faster onset-to treatment time, in 15-minute increments, was associated with … increased achievement of independent ambulation at discharge (OR, 1.04; 95% CI 1.03-1.05; P < 0.001) … ." Which of the following is a correct interpretation of these findings?

A. In this study, the effect of a 15-minute reduction in onset-to treatment time was associated with a 4 % (relative) increase in the odds of independent ambulation at discharge.
B. Because the odds ratio is very close to 1.0, the results are not statistically significant.
C. Because the odds ratio is very close to 1.0, the results, although highly statistically significant, are not clinically significant.
D. The 4% increase in odds of independent ambulation translates into a number-needed-to treat (NNT) of 25.
E. None of the above is correct.

Answer: A Choice A exactly expresses the meaning of the odds ratio for this study. Choices B and C are incorrect because the proximity of the odds ratio to 1 is in this case based partly on the choice of the authors to express it per 15 minutes onset-to-treatment time. This illustrates the importance of knowing the units of the predictor variable when it is not dichotomous. If the authors had expressed the difference per hour instead of per 15 minutes, the odds ratios would have been taken to the fourth power, that is, the odds ratio for independent ambulation at discharge would have been about $1.04^4 = 1.17$. Choice D is incorrect because estimation of the NNT requires knowing the absolute risk reduction, which was not provided in this case.

3. Assume a study of both smoking and nonsmoking mothers reports that the effect of smoking on birthweight is about a 32-g decrease in birthweight per cigarette smoked per day during pregnancy (similar to what is reported by Juarez SP, Merlo J. Revisiting the effect of maternal smoking during pregnancy on offspring birthweight: a quasi-experimental sibling analysis in Sweden. *PLoS One.* 2013;8:e61734.). Which of the following statements about this finding is NOT correct?

A. The result is based on a *model*, in which the predicted birthweight is linearly related to the number of cigarettes smoked per day.
B. This model predicts that the difference in birthweight between a baby whose mother did not smoke and one who smoked 5 cigarettes per day is the same as the difference between babies of mothers who smoked 20 and 25 cigarettes per day.
C. The model predicts the same effect of smoking on birthweight, regardless of the mother's age and prepregnancy weight.
D. This model predicts that cutting cigarette smoking in half will lead to a 64-g expected weight increase in the baby.
E. The model could include additional terms that would reflect the effect of mother's age and prepregnancy weight.

Answer: D Choices A, B and C accurately describe characteristics of a linear model for the effect of smoking on birthweight. Choice D is not consistent with a linear model. Choice E reflects that the effects of other variables can be taken into account, while still maintaining a linear model for the effect of cigarettes smoked on birthweight.

4. A case-control study of the relationship of breast cancer and the use of COX-2 inhibitors found that the odds ratio and confidence interval relating cancer to use of baby aspirin were OR = 0.77 and 95% CI (0.42-1.41) (Harris RE, Beebe-Donk J, Alshafie GA. Reduction in the risk of human breast cancer by selective cyclooxygenase-2 [COX-2] inhibitors. *BMC Cancer.* 2006;6:27). The authors then stated, "Neither acetaminophen nor baby aspirin had any effect on the relative risk of breast cancer." This is incorrect because of which of the following?

A. The confidence interval crosses 1.
B. The authors do not give the *P* value.
C. The confidence interval has a lower limit of 0.42.
D. Odds ratios are inappropriate for this study.
E. The authors should have used a higher level of confidence.

Answer: C C is correct because the confidence interval allows for a 58% reduction (from [1 to 0.42]*100%) in the chance of breast cancer associated with the use of baby aspirin, a potentially important effect. Choice A is incorrect because it merely indicates a lack of a statistically significant result and does not bear on the size of the effect. Choice B is incorrect because CIs are more useful for ruling out important effects. Odds ratios are especially useful in case-control studies, so D is incorrect. And E is incorrect because a higher level of confidence would make the CI even wider.

5. In a randomized trial of an intervention to reduce hypertension, researchers selected subjects from a single antihypertensive patient club in Shanghai (Xue F, Yao W, Lewin RJ. A randomised trial of a 5 week, manual based, self-management programme for hypertension delivered in a cardiac patient club in Shanghai. *BMC Cardiovasc Disord.* 2008;8:10). About one third of those approached agreed to be randomized. The results of this study can be safely generalized to which of the following?

A. All hypertensives
B. All hypertensives in China
C. All hypertensives in Shanghai
D. All hypertensives belonging to the single club in Shanghai
E. None of the above.

Answer: E Because two thirds of the participants refused to participate, the extrapolation of the effect of the intervention might not even apply to the particular club in which the study was conducted, much less a broader population.

10 USING DATA FOR CLINICAL DECISIONS

THOMAS H. LEE

TABLE 10-1 KEY DEFINITIONS*

Probability	A number between 0 and 1 that expresses an estimate of the likelihood of an event
Odds	The ratio of [the probability of an event] to [the probability of the event's not occurring]
TEST PERFORMANCE CHARACTERISTICS	
Sensitivity	Percentage of patients with disease who have an abnormal test result
Specificity	Percentage of patients without disease who have a normal test result
Positive predictive value	Percentage of patients with an abnormal test result who have disease
Negative predictive value	Percentage of patients with a normal test result who do not have disease
BAYESIAN ANALYSIS	
Pretest (or prior) probability	The probability of a disease before the information is acquired
Post-test (or posterior) probability	The probability of a disease after new information is acquired
Pretest (or prior) odds	(Pretest probability of disease)/(1 – pretest probability of disease)
Likelihood ratio	(Probability of result in diseased persons)/(Probability of result in nondiseased persons)

**Disease* can mean a condition, such as coronary artery disease, or an outcome, such as postoperative cardiac complications.

Key functions in the professional lives of all physicians are the collection and analysis of clinical data. Decisions must be made on the basis of these data, including which therapeutic strategy is most appropriate for the patient and whether further information should be gathered before the best strategy can be chosen. This decision-making process is a blend of science and art in which the physician must synthesize a variety of concerns, including the patient's most likely outcome with various management strategies, the patient's worst possible outcome, and the patient's preferences among these strategies.

Only rarely does the physician enjoy true certainty regarding any of these issues, so a natural inclination for physicians is to seek as much information as possible before making a decision. This approach ignores the dangers inherent in the collection of information. Some of these dangers are immediate, such as the risk of cerebrovascular accident associated with coronary angiography. Some dangers are delayed, such as the risk of a malignant neoplasm due to radiation exposure from diagnostic tests. And some dangers are subtle, such as the risk of unnecessary anguish for patients due to delays, uncertainty, and confusion.

An additional concern is the cost of information gathering, including the direct costs of the tests themselves and the indirect costs that flow from decisions made on the basis of the test results. Substantial data demonstrate marked variation in use of tests among physicians located in different regions and even within the same group practice. Standards of medical professionalism endorse the need for physicians to exert their influence to minimize inefficiency, but this challenge grows increasingly complex as medical progress leads to proliferation of alternative testing strategies.

For the physician, there are three key questions in this sequence: Should I order a test to improve my assessment of diagnosis or prognosis? Which test is best? Which therapeutic strategy is most appropriate for this patient?

SHOULD I ORDER A TEST?

The decision of whether to order a test depends on the physician's and the patient's willingness to pursue a management strategy with the current degree of uncertainty.[1] This decision is influenced by several factors, including the patient's attitudes toward diagnostic and therapeutic interventions (e.g., a patient with claustrophobia might prefer an ultrasound to magnetic resonance imaging) and the information provided by the test itself. The personal tolerance of the patient and physician for uncertainty also frequently influences test-ordering approaches. A decision to watch and wait rather than to obtain a specific test also should be considered an information-gathering alternative because the information obtained while a patient is being observed often reduces uncertainty about the diagnosis and outcome. In other words, the "test of time" should be recognized as one of the most useful tests available when this tactic does not seem inappropriately risky.

Most tests do not provide a definitive answer about diagnosis or prognosis but instead reduce uncertainty. Accordingly, the impact of information from tests often is expressed as *probabilities* (Table 10-1). A probability of 1.0 implies that an event is certain to occur, whereas a probability of 0 implies that the event is impossible. When all the possible events for a patient are assigned probabilities, these estimates should sum to 1.0.

It is often useful to use *odds* to quantify uncertainty instead of probability. Odds of 1 : 2 suggest that the likelihood of an event is only half the likelihood that the event will not occur, or a probability of 0.33. The relationship between odds and probability is expressed in the following formula:

$$\text{Odds} = P/(1-P)$$

where P is the probability of an event.

Performance Characteristics

Sensitivity and *specificity* are key terms for the description of test performance. These parameters describe the test and are in theory true regardless of the population of patients to which the test is applied. Research studies that describe test performance often are based, however, on highly selected populations of patients; test performance may deteriorate when tests are applied in clinical practice. The result of a test for coronary artery disease, such as an electron beam computed tomography scan, rarely may be abnormal if it is evaluated in a low-risk population, such as high-school students. False-positive abnormal results secondary to coronary calcification in the absence of obstructive coronary disease are common when the test is performed in middle-aged and elderly people. Another increasingly appreciated factor that can distort the performance of screening tests is the phenomenon of "overdiagnosis," in which the "disease" that is detected would not have led to clinical harm if it had not been found.[2]

Although researchers are interested in the performance of tests, the true focus of medical decision making is the patient. Physicians are more interested in the implications of a test result on the probability that a patient has a specific disease or outcome, that is, the predictive values of abnormal or normal test results. These predictive values are extremely sensitive to the population from which they are derived (Table 10-2; see also Table 10-1). An abnormal lung scan result in an asymptomatic patient has a much lower positive predictive value than that same test result in a patient with dyspnea and diminished oxygen saturation. Bayes theorem (see later) provides a

TABLE 10-2 EXAMPLE OF ODDS RATIO FORM OF BAYES THEOREM

Question: What is the probability of coronary disease for a patient with a 50% pretest probability of coronary disease who undergoes an exercise test if that patient develops (a) no ST segment changes, (b) 1 mm of ST segment depression, or (c) 2 mm of ST segment depression?

Step 1. Calculate the pretest odds of disease:

$$P/(1-P) = 0.5/(1-0.5)$$
$$= 0.5/0.5$$
$$= 1$$

Step 2. Calculate the likelihood ratios for the various test results, using the formula LR = sensitivity/(1 − specificity). (Data from pooled literature.)

TEST RESULT	SENSITIVITY	SPECIFICITY	LIKELIHOOD RATIO
No ST segment changes	0.34	0.15	0.4
1-mm ST segment depression	0.66	0.85	4.4
2-mm ST segment depression	0.33	0.97	11

Step 3. Calculate the post-test odds of disease and convert those odds to post-test probabilities.

TEST RESULT	PRETEST ODDS	LIKELIHOOD RATIO	POST-TEST ODDS	POST-TEST PROBABILITY
No ST segment changes	1	0.4	0.4	0.29
1-mm ST segment depression	1	4.4	4.4	0.81
2-mm ST segment depression	1	11	11	0.92

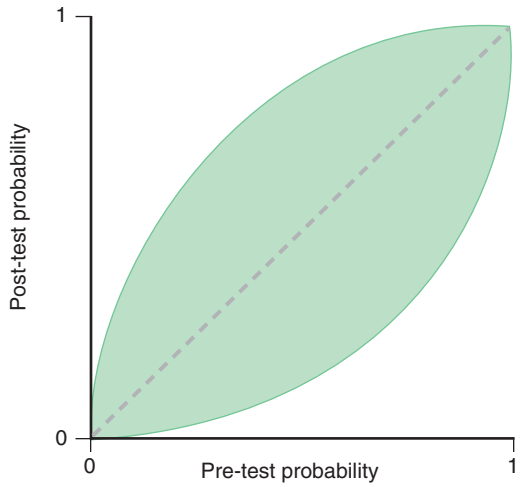

FIGURE 10-1. Impact of various test results on the patient's probability of disease. The *x*-axis depicts a patient's probability of disease before a test. If the test is of no value, the post-test probability (*dotted line*) is no different from the pretest probability. An abnormal test result raises the post-test probability of disease, as depicted by the concave downward arc, whereas a normal test result lowers the probability.

framework for analyzing the interaction between test results and a patient's pretest probability of a disease.

As useful as the performance characteristics may be, they are limited by the fact that few tests truly provide dichotomous (i.e., positive or negative) results. Tests such as exercise tests have several parameters (e.g., ST segment deviation, exercise duration, hemodynamic response) that provide insight into the patient's condition, and the normal range for many blood tests (e.g., a serum troponin level) varies markedly according one's willingness to "miss" patients with disease. Tests that require human interpretation (e.g., radiologic studies) are particularly subject to variability in the reported results.

Bayes Theorem

The impact of a test result on a patient's probability of disease was first quantified by Bayes, an 18th century English clergyman who developed a formula that describes the probability of disease in the presence of an abnormal test result. The classic presentation of Bayes theorem is complex and difficult to use. A more simple form of this theorem is known as the *odds ratio* form, which describes the impact of a test result on the pretest odds (see Table 10-1) of a diagnosis or outcome for a specific patient.

To calculate the post-test odds of disease, the pretest odds are multiplied by the *likelihood ratio* (LR) for a specific test result. The mathematical presentation of this form of Bayes theorem is as follows:

$$\text{Post-test odds} = (\text{Pretest odds}) \times (\text{LR})$$

The LR is the probability of a particular test result in patients with the disease divided by the probability of that same test result in patients without disease. In other words, the LR is the test result's sensitivity divided by the false-positive rate. A test of no value (e.g., flipping a coin and calling "heads" an abnormal result) would have an LR of 1.0 because half of patients with disease would have abnormal test results, as would half of patients without disease. This test would have no impact on a patient's odds of disease. The further an LR is above 1.0, the more that test result raises a patient's probability of disease. For LRs less than 1.0, the closer the LR is to 0, the more it lowers a patient's probability of disease.

When it is displayed graphically (Fig. 10-1), a test of no value (*dotted line*) does not change the pretest probability, whereas an abnormal or normal result from a useful test moves the probability up or down. For a patient with a high pretest probability of disease, an abnormal test result changes the patient's probability only slightly, but a normal test result leads to a marked reduction in the probability of disease. Similarly, for a patient with a low pretest probability of disease, a normal test result has little impact, but an abnormal test result markedly raises the probability of disease.

Consider how various exercise test results influence a patient's probability of coronary disease (see Table 10-2). For a patient whose clinical history, physical examination, and electrocardiographic findings suggest a 50% probability of disease, the pretest odds of disease are 1.0. LRs for various test results are developed by pooling data from published literature. The sensitivity of an exercise test with any amount of ST segment changes is the rate of such test results in patients with coronary disease, and the specificity is the percentage of patients without coronary disease who do *not* have this test result. The LR for no ST change is less than 1, whereas the LRs for patients with ST changes are greater than 1 (see Table 10-2). Therefore, when the LRs for various test results are multiplied by the pretest odds to calculate post-test odds, the odds decrease for patients without ST segment changes but increase for patients with 1 or 2 mm of ST segment change. Post-test odds can be converted to post-test probabilities according to the following formula:

$$\text{Probability} = \text{Odds}/(1+\text{odds})$$

The calculations quantify how the absence of ST segment changes reduces a patient's probability of disease, whereas ST segment depression raises the probability of disease.

This form of Bayes theorem is useful for showing how the post-test probability of disease is influenced by the patient's pretest probability of disease. If a patient's clinical data suggest a *probability* of coronary disease of only 0.1, the *pretest odds* of disease would be only 0.11. For such a low-risk patient, an exercise test with no ST segment changes would lead to post-test probability of coronary disease of 4%, whereas 1-mm or 2-mm ST segment changes would lead to a post-test probability of disease of 33% or 55%, respectively.

Even if clinicians rarely perform the calculations that are described in Bayes theorem, there are important lessons from this theorem that are relevant to principles of test ordering (Table 10-3). The most crucial of these lessons is that the interpretation of test results must incorporate information about the patient. An abnormal test result in a low-risk patient may not be a true indicator of disease. Similarly, a normal test result in a high-risk patient should not be taken as evidence that disease is not present.

Figure 10-2 provides an example of the post-test probabilities for positive and negative results for a test with a sensitivity of 85% and a specificity of 90% (e.g., radionuclide scintigraphy for diagnosis of coronary artery disease). In a high-risk population with a 90% prevalence of disease, the positive predictive value of an abnormal result is 0.99 compared with 0.31 for the same test result obtained in a low-risk population with a 5% prevalence of disease. Similarly, the negative predictive value of a normal test result is greater in the low-risk population than in the high-risk population.

TABLE 10-3 PRINCIPLES OF TEST ORDERING AND INTERPRETATION

The interpretation of test results depends on what is already known about the patient.
No test is perfect; clinicians should be familiar with its diagnostic performance (see Table 10-1) and never believe that a test "forces" them to pursue a specific management strategy.
Tests should be ordered if they may provide *additional* information beyond that already available.
Tests should be ordered if there is a reasonable chance that the data will influence the patient's care.
Two tests that provide similar information should not be ordered.
In choosing between two tests that provide similar data, use the test that has lower costs or causes less discomfort and inconvenience to the patient.
Clinicians should seek all of the information provided by a test, not just an abnormal or normal result.
The cost-effectiveness of strategies using noninvasive tests should be considered in a manner similar to that of therapeutic strategies.

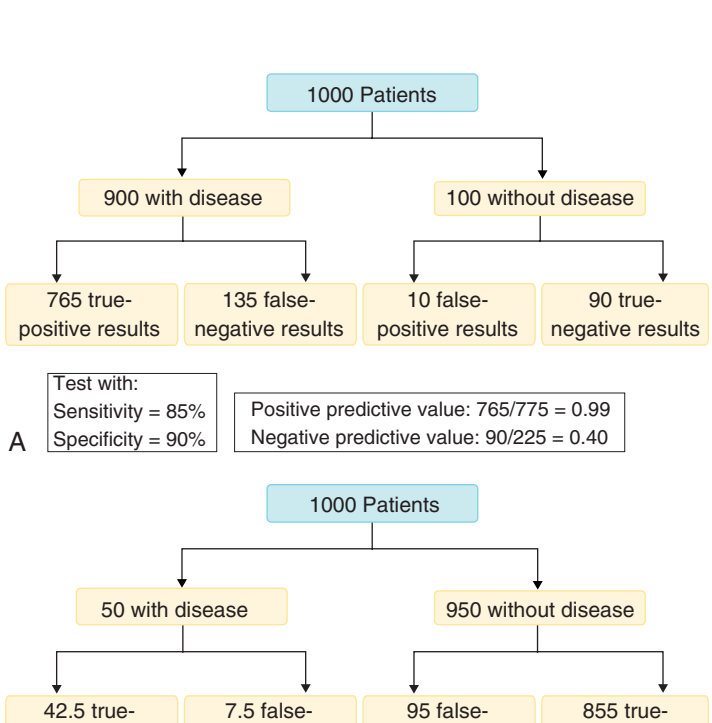

FIGURE 10-2. Interpretation of test results in high-risk and low-risk patients. A, High-risk population (90% prevalence of disease). B, Low-risk population (5% prevalence of disease).

Multiple Testing

Clinicians frequently obtain more than one test aimed at addressing the same issue and at times are confronted with conflicting results. If these tests are truly independent (i.e., the tests do not have the same basis in pathophysiology), it may be appropriate to use the post-test probability obtained through performance of one test as the pretest probability for the analysis of the impact of the second test result.

If the tests are not independent, this strategy for interpretation of serial test results can be misleading. Suppose a patient with chronic obstructive pulmonary disease and a history vaguely suggestive of pulmonary embolism is found to have an abnormal lung ventilation-perfusion scan. Obtaining that same test result over and over would not raise that patient's probability of pulmonary embolism further and further. In this extreme case, the tests are identical; serial testing adds no information. More commonly, clinicians are faced with results from tests with related but not identical bases in pathophysiology, such as ventilation-perfusion scintigraphy and pulmonary angiography.

Regardless of whether tests are independent, the performance of multiple tests increases the likelihood that an abnormal test result will be obtained in a patient without disease. If a chemistry battery includes 20 tests and the normal range for each test has been developed to include 95% of healthy individuals, the chance that a healthy patient will have a normal result for any specific test is 0.95. However, the probability that all 20 tests will be normal is $(0.95)^{20}$, or 0.36. Most healthy people can be expected to have at least one abnormal result. Unless screening test profiles are used thoughtfully, false-positive results can subject patients to unnecessary tests and procedures.

Threshold Approach to Decision Making

Even if a test provides information, that information may not change management for an individual patient. Lumbar spine radiographs of a patient who is not willing to undergo surgery may reveal the severity of disease but expose the patient to needless radiation. Similarly, a test that merely confirms a diagnosis that already is recognized is a waste of resources (see Table 10-3).

Before ordering a test, clinicians should consider whether that test result could change the choice of management strategies. This approach is called the *threshold approach to medical decision making*, and it requires the physician to be able to estimate the threshold probability at which one strategy will be chosen over another. The management of a clinically stable patient with a high probability of coronary disease might not be changed by any of the post-test probabilities shown in Table 10-2. If that patient had no ST segment changes, the post-test probability of 0.29 still would be too high for a clinician to consider that patient free of disease. An abnormal test result that strengthened the diagnosis of coronary disease might not change management unless it suggested a greater severity of disease that might warrant another management strategy.

Testing for Peace of Mind

Physicians frequently order tests even when there is little chance that the outcomes will provide qualitatively new insights into a patient's diagnosis or prognosis or alter a patient's management. In such cases, the cited goal for testing may be to improve a patient's peace of mind. Although a decrease in uncertainty can improve quality of life for many patients, individuals with hypochondriasis and somatization disorders rarely obtain comfort from normal test results; instead, their complaints shift to a new organ system, and their demands focus on other tests. For such patients, management strategies using frequent visits and cognitive tactics are recommended.

WHICH TEST IS BEST?

If the clinician decides that more information is needed to reduce uncertainty, and if it appears possible that tests might lead to a change in management strategies, the question arises as to which test is most appropriate. Note that just because guideline development committees have concluded that a specific test is "appropriate" in a given clinical context, it does not mean that this test is the *most* appropriate option. Several factors influence the choice among diagnostic strategies, including the patient's preferences, the costs and risks associated with the tests, and the diagnostic performance of alternative tests.

Diagnostic performance of a test often is summarized in terms of sensitivity and specificity,[3] but as shown in the example in Table 10-2, these parameters depend on which threshold (e.g., 1 mm vs. 2 mm of ST segment change) is used. A low threshold for calling a test result abnormal might lead to excellent sensitivity for detecting disease but at the expense of a high false-positive rate. Conversely, a threshold that led to few false-positive results might cause a clinician to miss many cases of true disease.

The receiver operating characteristic (ROC) curve is a graphic form of describing this tradeoff and providing a method for comparing test performance (Fig. 10-3). Each point on the ROC curve describes the sensitivity and the false-positive rate for a different threshold for abnormality for a test. A test of no value would lead to an ROC curve with the course of the dotted line, whereas a misleading test would be described by a curve that was concave upward (not shown).

The more accurate the test, the closer its ROC curve comes to the upper left corner of the graph, which would indicate a test threshold that has excellent sensitivity and a low false-positive rate. The closer an ROC curve comes to the upper left corner, the greater the area under the curve. The area under ROC curves can be used to compare the information provided by two tests.

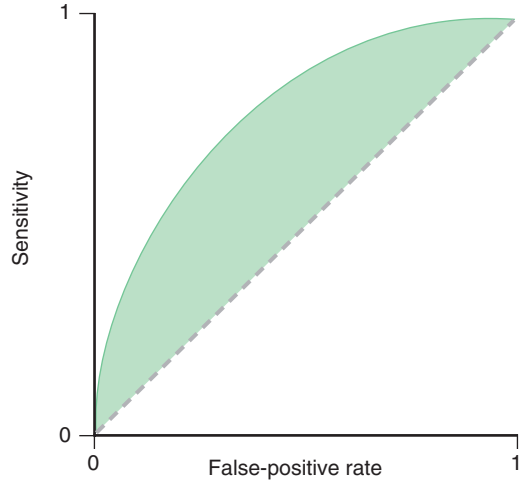

FIGURE 10-3. Receiver operating characteristic curve. The points on the curve reflect the sensitivity and false-positive (1 – specificity) rates of a test at various thresholds. As the threshold is changed to yield greater sensitivity for detecting the outcome of interest, the false-positive rate rises. The better the test, the closer the curve comes to the upper left corner. A test of no value (e.g., flipping a coin) would lead to a curve with the course of the dotted line. The area under the curve is used often to compare alternative testing strategies.

Even if one test is superior to another as shown by a greater area under its ROC curve, the question still remains as to what value of that test should be considered abnormal. The choice of threshold depends on the purpose of testing and on the consequences of a false-positive or false-negative diagnosis. If the goal is to screen the population for a disease that is potentially fatal and potentially curable, a threshold with excellent sensitivity is appropriate even if it leads to frequent false-positive results. In contrast, if a test is used to confirm a diagnosis that is likely to be treated with a high-risk invasive procedure, a threshold with high specificity is preferred. Only 1 mm of ST segment depression might be the appropriate threshold when exercise electrocardiography is used to evaluate the possibility of coronary disease in a patient with chest pain. If the question is whether to perform coronary angiography in a patient with stable angina in search of severe coronary disease that might benefit from revascularization, a threshold of 2 mm or more would be more appropriate.

CHOOSING A STRATEGY

Physicians and patients ultimately must use clinical information to make decisions. These choices usually are made after consideration of a variety of factors, including information from the clinical evaluation, patients' preferences, and expected outcomes with various management strategies. Insight into the impact of these considerations can be improved through the performance of decision analysis (Table 10-4).

The first step in a decision analysis is to define the problem clearly; this step often requires writing out a statement of the issue so that it can be scrutinized for any ambiguity. After the problem is defined, the next step is to define the alternative strategies.

Consider the question of which test is most appropriate to screen patients for breast cancer: mammography with or without breast magnetic resonance imaging—a technology that is highly sensitive for detecting breast cancer but is more costly and less specific. The expected outcomes for these strategies depend on each test's sensitivity and specificity for detecting breast cancer, which is influenced in turn by other factors, such as the frequency with which the test is performed. Patients' outcomes also are influenced by their underlying risk for breast cancer and the likelihood that earlier detection of tumors reduces the risk for death.

Each of these variables must be known or estimated for calculations to be made of each strategy's predicted life expectancy and direct medical costs. These outcomes differ for patients according to age, medical history, family history, and presence or absence of genetic markers such as *BRCA* mutations. Optimal strategies for an elderly patient with a short life expectancy and low clinical risk of cancer are unlikely to be the same as those for a younger patient with inherited mutations of the *BRCA1* or *BRCA2* gene, indicating a cumulative lifetime risk of breast cancer of 50 to 85% (Chapter 198).

TABLE 10-4 STEPS IN PERFORMANCE OF DECISION ANALYSIS

Frame the question.
Create the decision tree.
Identify the alternative strategies.
List the possible outcomes for each of the alternative strategies.
Describe the sequence of events as a series of decision nodes and chance nodes.
Choose a time horizon for the analysis.
Determine the probability for each chance outcome.
Assign a value to each outcome.
Calculate the expected utility for each strategy.
Perform sensitivity analysis.

The credibility of the decision analysis depends on the credibility of these estimates. Published reports often do not provide information on the outcomes of interest for specific subsets of patients, or there may not have been sufficient statistical power within subsets of patients for the findings to be statistically significant. Randomized trial data are relevant to the populations included in the trial; the extension of the findings to other genders, races, and age groups requires assumptions by individuals performing the analysis. For many issues, expert opinion must be used to derive a reasonable estimate of the outcome.

For many diseases, the potential outcomes are more complex than perfect health or death. With chronic diseases, patients may live many years in a condition somewhere between these two, and the goal of medical interventions may be to improve quality of life rather than to extend survival. The value of life in imperfect health must be reflected in decision analyses. These values by convention are expressed on a scale of 0 to 100, where 0 indicates the worst outcome and 100 indicates the best outcome.

Life-expectancy and quality-of-life estimates are combined in many decision analyses to calculate *quality-adjusted life years.* A strategy that leads to a 10-year life expectancy with such severe disability that utility of the state of health is only half that of perfect health would have a quality-adjusted life expectancy of 5 years. With such adjustments to life-expectancy data, the impact of interventions that improve quality of life but do not extend life can be compared with interventions that extend life but do not improve its quality or perhaps even worsen it.[4]

After the value and the probability of the various outcomes have been estimated, the expected utility of each strategy can be calculated. In comparing the different strategies available at a decision node, the analysis generally selects the option with the highest expected utility. At chance nodes, the expected utility is the weighted average of the utility of the various possible branches.

After the analysis has been performed with the baseline assumptions, *sensitivity analyses* should be performed in which these assumptions are varied over a reasonable range. These analyses can reveal which assumptions have the most influence over the conclusions and identify threshold probabilities at which the conclusions would change. For example, the threshold at which breast magnetic resonance imaging should be added to mammography is likely to be influenced by the cost of the magnetic resonance imaging and the accuracy of the radiologists who interpret the images.

Cost-Benefit and Cost-Effectiveness Analyses

For clinicians and health care policymakers, the choices that must be addressed go beyond the choices within any single decision analysis. Because resources available for health care are limited, policymakers may have to choose among many competing options for "investments" in health. Although such decisions frequently are made on the basis of political considerations, cost-benefit and cost-effectiveness analyses can be informative in making the choices.

The methodology of these techniques is similar to that of decision analysis except that costs for the various possible outcomes and strategies also are calculated. *Discounting* is used to adjust the value of future benefits and costs because resources saved or spent currently are worth more than resources saved or expended in the future. In *cost-benefit* analyses, all benefits are expressed in terms of economic impact. Extensions in life expectancy are translated into dollars by estimating societal worth or economic productivity.

Because of the ethical discomfort associated with expressing health benefits in financial terms, *cost-effectiveness* analyses are used more commonly

TABLE 10-5 ESTIMATED COST-EFFECTIVENESS OF SELECTED HEALTH INTERVENTIONS

INTERVENTION	COST PER QUALITY-ADJUSTED LIFE YEAR (QALY) (2010 DOLLARS)
Treating rheumatoid arthritis with drugs that slow disease progression	Saves money and improves health
Using warfarin for 70-year-olds with atrial fibrillation	$3,000 per QALY
Daily dialysis for 60-year old critically ill men with kidney injury	$6,000 per QALY
Using an implantable cardioverter-defibrillator to prevent sudden cardiac death in high-risk patients	$38,000 per QALY
Treating spinal stenosis and leg pain with spine surgery	$90,000 per QALY
Screening 60-year-old heavy smokers with annual CT scans	$140,000 per QALY
Annual HIV screening for people with a low to moderate risk	Increases costs and makes health worse

Modified from the CEA Registry.[6]

than cost-benefit analyses. In these analyses, the ratio of costs to health benefits is calculated; one frequently used method for evaluating a strategy is calculation of cost per quality-adjusted life year.[5] These estimates can be used to identify strategies that are both cost-saving and health-improving, to compare strategies by which the health care system can "purchase" additional quality-adjusted life years, and even to caution about strategies that increase costs while worsening health (Table 10-5).[6]

Cost-effectiveness analyses can provide important insights into the relative attractiveness of different management strategies and can help guide policymakers in decisions about which technologies to make available on a routine basis. No medical intervention can have an attractive cost-effectiveness if its effectiveness has not been proved. The cost-effectiveness of an intervention depends heavily on the population of patients in which it is applied. An inexpensive intervention would have a poor cost-effectiveness ratio if it were used in a low-risk population unlikely to benefit from it. In contrast, an expensive technology can have an attractive cost-effectiveness ratio if it is used in patients with a high probability of benefiting from it. Table 10-5 shows cost-effectiveness estimates from published literature for some selected medical interventions. Such estimates should be used only with understanding of the population for which they are relevant.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Laine C. High-testing begins with a few simple questions. *Ann Intern Med*. 2012;156:162-163.
2. Etzioni R, Gulati R, Mallinger L, et al. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. *Ann Intern Med*. 2013;158:831-838.
3. Otero HJ, Fang CH, Sekar M, et al. Accuracy, risk and the intrinsic value of diagnostic imaging: a review of the cost-utility literature. *Acad Radiol*. 2012;19:599-606.
4. Heijnsdijk EAM, Wever EM, Auvinen A, et al. Quality-of-life effects of prostate-specific antigen screening. *N Engl J Med*. 2012;367:595-605.
5. Ryen L, Svensson M. The willingness to pay for a quality adjusted life year: a review of the empirical literature. *Health Econ*. 2014. [Epub ahead of print].
6. Cost-Effectiveness Analysis Registry. https://research.tufts-nemc.org/cear4/. Accessed February 10, 2015.

REVIEW QUESTIONS

1. A 53-year-old man presents to his primary care physician with a chief complaint of chest pain for the past 2 weeks. The pain is described as intermittent, usually exertional, midline, and aching in nature. His electrocardiogram is completely normal. Which one of the following is the most appropriate next step?

A. Watch and wait
B. Exercise electrocardiography
C. B-type natriuretic peptide
D. Exercise nuclear scintigraphy
E. Coronary angiography

Answer: B The immediate question is whether this patient's symptoms represent new ischemic heart disease. His clinical presentation suggests a moderate probability of coronary artery disease, so watch-and-wait is probably too risky a strategy. At the same time, he does not appear to have unstable ischemic disease, so there is no need to proceed immediately to coronary angiography in preparation for coronary revascularization. Measurement of B-type natriuretic peptide would not alter management. Of the two noninvasive tests for ischemic heart disease, exercise electrocardiography is the least expensive, is the most convenient, and carries no radiation exposure. Guidelines would thus suggest that answer B is the most appropriate next step.

2. A patient undergoes an exercise test and has 2 mm of ST-segment depression on electrocardiogram. In which one of these patients would this finding be most likely to change management?

A. A healthy 19-year-old volunteer in a research study
B. A 62-year-old woman who is completely asymptomatic
C. A 62-year-old woman with frequent nonexertional aching chest pain
D. A 62-year-old woman with recent myocardial infarction and chest pain at rest
E. A 62-year-old woman who is completely symptom free 4 months after coronary artery bypass graft surgery

Answer: C Patients A and B have a sufficiently low probability of coronary disease that the exercise test abnormality is highly likely to be a false-positive result. Patient D has an extremely high probability of coronary disease and likely needs coronary angiography and revascularization as next steps; she does not need exercise electrocardiography because the test is unlikely to change this management plan. Patient E's care is also not likely to be influenced by an exercise test result because she is asymptomatic after major coronary revascularization surgery. Patient C has a low to moderate probability of coronary disease, and this abnormal exercise test result moves her into a mid-range probability. Thus, she is likely to undergo either further testing, initiation of antianginal therapy, or both as a result of this exercise test result.

3. Which one of the following is NOT an important consideration when weighing whether to order a test for a patient?

A. Test may influence decision making for patient's care
B. Test is less expensive than alternative strategies
C. Test is safer than alternative strategies
D. Test reduces patient's or clinician's uncertainty
E. Test is expected to be abnormal in presence of patient's already confirmed diagnosis

Answer: E Tests should be ordered when they are expected to change care and should be chosen on basis of safety, cost, and impact. They should not be ordered simply because they can confirm an already known diagnosis.

11

MEASURING HEALTH AND HEALTH CARE

CAROLYN M. CLANCY AND ERNEST MOY

Physicians routinely quantify a variety of health measures, including symptoms, vital signs, and findings on physical examination, to improve diagnosis, treatment, and prognostication. Similarly, the efficacy and quality of health care also can and should be measured for several reasons.

First, the quality of care delivered is often suboptimal.[1] Persistent variations in practice for patients with the same diagnosis reflect a combination of clinical uncertainty, individualized practice styles, patients' preferences and characteristics (age, race, ethnicity, education, income), and other factors. Both suboptimal care and varied care for the same condition undermine the historical assumption that a combination of highly trained health professionals and accredited facilities is sufficient to ensure consistent high-quality care.

A second major trend is attributable to the successes of biomedical science: the major challenge in health care today is the management of chronic disease for a population with increased life expectancy. For chronic conditions, health benefits are increasingly measured in improvements in functional status or quality of life, rather than simply using mortality rates or life expectancy.

A third trend relates directly to how the increasing costs of health care are now threatening public budgets and investments in other social goals, such as education. Although the United States spends more per capita on health care than any other developed nation (Chapter 5), the outcomes achieved lag far behind.

Finally, advances in communication and information technologies have inspired more people to play an active role in their health and health care. These innovations have accelerated demands for transparency and shared decision making.

As health insurance and health care regulation have expanded, requirements to track and justify health care services have grown. Intensifying urgency to improve the quality of health care, reduce disparities, control costs, and enhance transparency will likely lead patients and insurers to demand more data and to link quality measures to payments for services.

Fortunately, modern technology can help to meet the demand for data. Patients can record and submit their health parameters using hand-held devices connected to personal health records. Automated billing programs can track health care services, while electronic health records can assess the quality of physician care. Ultimately, fully integrated health information systems will allow patient information to be retrieved instantly and seamlessly wherever and whenever it is needed. In addition to assessing care quality today, these tools offer enormous promise for learning as a byproduct of care delivery.

HOW ARE HEALTH AND HEALTH CARE MEASURED?

Three types of measures typically assess health and health care. Measures of *health* quantify the sickness or well-being of a person. Measures of *health care quality* quantify the extent to which a patient receives needed care and does not receive unnecessary care. Health care quality is assessed using measures of *structure* (e.g., education and credentialing of clinicians), *process* (adherence to professional standards and evidence-based recommendations), and *outcomes* (or end results of care, including how patients experience their care and their self-reported health and function). Measures of *health care resources* quantify the resources used (e.g., radiographs, surgery, medication, intensive care) to improve the health of a patient.[2] All measures can be summed up across populations within a practice or community (Table 11-1).

Measures of health and health care often overlap (E-Fig. 11-1). Health measures that can be improved by health care, such as blood pressure or blood glucose levels, are often used as health care quality outcome measures. The delivery of quality health care requires the use of resources and the generation of direct health care costs, which may or may not improve health care at the margins of spending. Impaired health that reduces the ability to do work and earn wages but that could have been prevented by the delivery of health care contributes to the indirect costs of health care. At the intersection of health, health care quality, and health care resources are measures of health care value. These measures compare the health benefits of specific health care services with their costs.[2]

WHERE DO MEASURES OF HEALTH AND HEALTH CARE COME FROM?

Most researchers, provider groups, insurers, regulators, and credentialing organizations that develop measures consult with physicians to ensure that their metrics are consistent with professional standards. Insurers may create measures to allocate health care resources, to plan for future needs, and to identify efficient physicians for inclusion on panels or to be rewarded with performance bonuses. Regulators may develop measures to establish licensure requirements and to identify physicians who might benefit from remedial instruction. Credentialing organizations may construct measures to demonstrate the superior performance of physicians who meet their high standards. For example, the National Committee for Quality Assurance maintains the Healthcare Effectiveness Data and Information Set that is widely used to accredit health plans, and the American Board of Internal Medicine and other specialty boards include measures of practice performance as well as measures of medical knowledge for the maintenance of certification.

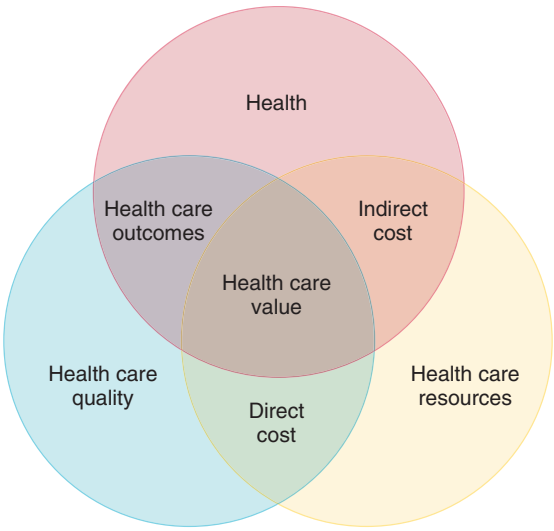

E-FIGURE 11-1. Type of measures of health and health care.

TABLE 11-1 MEASURES OF HEALTH AND HEALTH CARE

MEASURES OF HEALTH

- *Mortality:* rates of death typically adjusted for age and sex
- *Morbidity:* incidence and prevalence rates of diseases and their sequelae
- *Functional status:* assessments of a patient's ability to perform various actions such as activities of daily living or instrumental activities of daily living as observed by a provider or reported by the patient
- *Self-reported health status:* a patients' assessment of their health and well-being

MEASURES OF HEALTH CARE QUALITY

- *Health care outcomes:* the end results or health benefits derived from good health care or the health loss attributable to poor health care
- *Health care processes:* assessments of whether the right care was delivered at the right time and in the right way
- *Health care infrastructure:* the availability of resources needed to deliver good health care
- *Patient perceptions of health care:* a patient's assessment of health care received, usually emphasizing patient-provider communication and shared decision making
- *Access to health care:* the ability of patients to gain entry into health care and navigate to needed resources

MEASURES OF HEALTH CARE RESOURCES

- *Health care utilization:* the quantity of health care services that are used
- *Direct costs:* the costs of providers, supplies, and equipment needed to deliver health care
- *Indirect costs:* the costs of lost wages and decreased productivity due to illness or injury that could have been prevented by appropriate health care
- *Nonmedical costs:* the costs of health care not related to the delivery of services, such as administration, advertising, research, and profits earned by health industries

TYPE OF MEASURE	EXAMPLE
HEALTH	
Mortality	Deaths due to colorectal cancer per 100,000 population
Morbidity	New AIDS cases per 100,000 population
Functional status	% of people unable to perform one or more activities of daily living
Self-reported health status	% of people reporting that their overall health is excellent
HEALTH CARE QUALITY	
Health care outcomes	Death per 1000 hospitalizations with pneumonia
Intermediate outcomes	% of adults with diabetes whose blood pressure is <140/80 mm Hg
Health care processes	% of children who were given all recommended vaccinations
Health care infrastructure	% of office-based physicians with computerized systems for recording clinical notes
Patient perceptions of health care	% of patients who always reported good communication with their regular providers
Access to health care	% of patients who were unable or delayed in receiving medical care
HEALTH CARE RESOURCES	
Health care utilization	% of people with an emergency department visit in the past year
Direct costs	Expenditures for the treatment of depression
Indirect costs	Lost wages and productivity while caring for children with asthma
Nonmedical costs	Profits earned by pharmaceutical companies

WHY MEASURE HEALTH AND HEALTH CARE?

Measures of health and health care can be used for many purposes by different stakeholders (Table 11-2). Patients can use their own health information to track their progress, adjust their lifestyle, and plan for their future health care needs. Patient-level measures are important to physicians, hospitals, health plans, and policymakers, whether they are used to identify sentinel events that represent quality defects (e.g., amputating the wrong leg or giving a patient the wrong medication) or to initiate root cause analyses to improve health care quality.

Hospitals and health plans aggregate measures over their practices to identify opportunities for raising quality, improving efficiency, and reducing care disparities. For measures for which physicians are primarily accountable, these data can be used to acknowledge and reward high-performing physicians, select physicians for inclusion on a panel, and produce report cards to inform the public. Patients can use these report cards to select physicians and health plans that best match their health care needs. Measures that are aggregated to the community level can help policymakers allocate health care resources to locales with the greatest need and assess the success of any interventions.

WHAT ARE THE LIMITATIONS OF HEALTH AND HEALTH CARE MEASURES?

Historically, physicians recorded information on paper and had tremendous autonomy with respect to the level of detail and accuracy. Now, however, diverse stakeholders have a shared interest in standardizing measurements to facilitate fair comparisons. Detailed chart reviews largely have been replaced by data gathered for electronic medical records and by appropriately detailed registries composed of patients with a common condition or intervention. Examples of registries that have been used as platforms for assessing health care quality include the registry developed by the Society for Thoracic Surgeons to improve the quality of cardiothoracic surgery and the National Surgical Improvement Program registry, which was initially developed within the Veterans Health Administration, to assess processes and outcomes of major surgical procedures. Both registries rely on meticulous manual data, collected by trained nurses, that facilitate professional efforts to improve outcomes of care.

SELECTING MEASURES BASED ON EVIDENCE

Data used to measure the structure, process, and outcomes of health care can be gathered as part of rigorous hypothesis-driven research or from secondary sources of data initially collected for routine clinical care billing. The data's rigor and integrity are critical to the reliability of analyses performed on them.

In addition to the quality of the data themselves, however, the method of study design is also critical to the validity of reported findings. The most robust measures are supported by the concurrence of evidence derived from different research methodologies. In randomized controlled trials, random assignment of the intervention ensures that patients who do and do not receive it are as similar as possible. In double-blind studies, neither the patient nor the patient's physician knows the assigned treatment. For more complex interventions, such as the use of a team to manage depression, blinding is impractical. In all settings, controls typically receive the best standard care. In some situations, patients may serve as their own controls in a time-series randomized trial.

Although the randomized trial is a rigorous way to assess treatment effects, enrollment is often limited to selected individuals who meet strict entry criteria. This approach enhances a trial's internal validity but limits its generalizability. As a result, randomized trials are ideal for establishing an intervention's efficacy, which is its potential benefit under ideal conditions, but not necessarily its *effectiveness* in the real world. *Cluster randomized trials*, which randomize the level of the provider or system, represent a practical approach to determining effectiveness.

Although randomized trials are critical for assessing efficacy and effectiveness, *observational* and *case-control* studies are also important. In a *cohort study*, persons or patients are followed to determine their outcomes as a function of whether or not they possess a particular attribute or have been exposed to a particular condition or intervention. Cohort studies can provide a direct estimate of the absolute risk for an outcome in exposed patients but cannot guarantee whether such differences are related to interventions that were not randomly allocated.

In *case-control studies*, cases are defined based on having experienced an outcome not experienced by controls. Information that is gleaned from existing records or interviews can determine the proportion of case and control patients who experienced an exposure of interest. Case-control studies can achieve the same or higher statistical power as cohort studies despite enrolling fewer subjects, so they are especially attractive for investigating uncommon outcomes, such as serious adverse effects of medications when the events are too rare to study with clinical trials or cohort studies. Case-control studies are subject to the misclassification bias and recall bias, and they provide estimates of relative risk but not absolute risk.

TABLE 11-2 USES OF HEALTH AND HEALTH CARE MEASURES BY USER AND UNIT OF ANALYSIS

	USES BY PROVIDERS	USES BY POLICYMAKERS	USES BY PATIENTS
Patient-level measures	Diagnosis and prognosis Monitoring treatment response and compliance	Identifying sentinel events	Tracking health and health needs Retirement and estate planning
Provider practice or health plan–level measures	Improving quality Improving efficiency Reducing disparities	Public reporting Paying for performance Provider credentialing	Selecting providers and health plans
Community-level measures	Selecting practice location	Resource allocation Policy evaluation	Selecting place to live

TABLE 11-3 ATTRIBUTES OF MEASURES OF HEALTH AND HEALTH CARE

MEASURE ATTRIBUTES	CRITERIA FOR PROVIDERS	CRITERIA FOR POLICYMAKERS	CRITERIA FOR PATIENTS
Measure is scientifically sound			
Based on strong evidence	Based on high-quality studies?		
Valid	Measures what is intended and includes key elements?		
Reliable	Reproducible by different measurers and providers?		
Clearly specified	Numerator, denominator, exclusions, and risk adjustment defined?		
Endorsed by independent experts	Well accepted by scientific and medical communities?		
Condition is important			
Affects many people Causes high mortality or morbidity Costs a lot Is very unequal across populations	Important for my practice or the population I serve?	Important for population that will be affected by my policy?	Important to me?
Data collection is feasible			
Already available Could be collected at low cost relative to potential benefit Auditable	Available for my practice or the population I serve?	Available before and after policy implementation for affected population?	Available to me?
Findings are actionable			
Can be understood Can be improved Have been used successfully Have few unintended consequences	Do I know what actions I must change to improve quality? Used in practices like mine to improve quality with few unintended consequences?	Do I know what policies I must change to improve quality? Used for policies like mine to improve quality with few unintended consequences?	Do I know what actions I must take to improve my health? Used by patients like me to improve health with few unintended consequences?
When used as part of measure set, measures are:	Represents multiple conditions, settings of care, populations, types of data (survey, administrative data, medical records), types of measures (structure, process, outcomes), and perspectives (patient, provider, system, society) as appropriate?		
Balanced Able to be disaggregated	Identifies individual measures that can be improved?		

Cross-sectional studies collect data at just one point in time. They can be used to estimate the prevalence of a condition or outcome but not to make valid inferences about whether a given outcome is related to any causal attribute or event.

There are additional methodologic issues beyond the credibility and quality of the source of data for quality measures. When presented with reports suggesting less-than-perfect performance, many physicians believe that their patients are sicker or are less likely to adhere to treatment recommendations. To address these concerns, risk adjustment for severity of illness, accurate identification of patient characteristics, and assessment of outcomes over time are all critical. Case-mix adjustment methods commonly use sophisticated statistical approaches in an attempt to adjust for such potential differences and to estimate whether observed differences result from differing patient populations rather than true differences in quality of care.

The precision of specific measures is also a challenge. Many stakeholders, especially payers and policymakers, are increasingly impatient and are generally more enthusiastic about use of measures that are only "pretty good" than are providers whose care is being judged.

SELECTING MEASURES: WHICH MEASURES ARE RIGHT FOR WHAT PURPOSE?

Creating measures is different from selecting which are most important and meaningful. Numerous publicly available sources provide a wealth of health and health care information (E-Table 11-1). The most prominent consensus-based organization that endorses measures is the National Quality Forum, which convenes panels of experts to assess the quality of a proposed measure based on criteria that include importance, scientific acceptability, feasibility, and usability. Endorsed measures are reassessed periodically to ensure they remain up to date and applicable.

The National Institutes of Health Patient Reported Outcomes Measurement Information System (http://www.nihpromis.org) is a repository of standardized patient-reported health status measures and collection instruments. Domains include physical, mental, and social well-being. The Agency for Healthcare Research and Quality National Quality Measures Clearinghouse (http://www.qualitymeasures.ahrq.gov) is an inventory of evidence-based measures of health care quality. Domains include clinical quality, efficiency, and population health. Also related are the National Guidelines Clearinghouse (http://www.guideline.gov) and the Health Care Innovations Exchange (http://www.innovations.ahrq.gov), collections of clinical guidelines and descriptions of successful uses of measures, respectively.

Potential users should examine the attributes of specific measures to ensure that they are appropriate for their intended application (Table 11-3). A number of organizations have developed scales or rating systems that summarize performance across multiple dimensions. Sometimes referred to as composite measures (e.g., a letter grade from the Leapfrog Group for hospital safety), these scales vary in comprehensiveness with little consensus regarding how components of a summary scale should be weighted. Indeed, it is not uncommon for a hospital to receive an "F" from one group and be rated a "top performer" by another. Thus, although appealing in their simplicity, the optimal use of these approaches remains unresolved.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

E-TABLE 11-1 EXAMPLES OF PUBLICLY AVAILABLE SOURCES OF SECONDARY DATA

FEDERAL AGENCIES

Department of Health and Human Services (HHS) Health System Management Project—government data on critical U.S. health system indicators: https://healthmeasures.aspe.hhs.gov/

Agency for Health Care Research and Quality (AHRQ): http://www.ahrq.gov

- Healthcare Cost and Utilization Project—the largest collection of longitudinal hospital care data in the United States, with all-payer, encounter-level information beginning in 1988: http://www.hcup-us.ahrq.gov/
- Medical Expenditure Panel Survey (MEPS)—a set of large-scale surveys of U.S. families and individuals, medical providers, and employers: http://meps.ahrq.gov/mepsweb
- National CAHPS benchmarking database—the national repository for data from the CAHPS family of surveys: https://cahpsdatabase.ahrq.gov
- National Healthcare Quality and Disparities Reports—the database of the National Reports including the State Snapshots and NHQR/DR query tool: http://nhqrnet.ahrq.gov/inhqrdr/

Centers for Disease Control and Prevention (CDC): http://www.cdc.gov

- CDC Wonder: http://wonder.cdc.gov
- DATA2020—the Healthy People 2020 interactive data tool that allows users to explore the data and technical information related to the Healthy People 2020 objectives: http://www.healthypeople.gov/2020/data/
- Health Data Interactive—tables with national health statistics for infants, children, adolescents, adults, and older adults that can be customized by age, gender, race/ethnicity, and geographic location: http://www.cdc.gov/nchs/hdi.htm
- Health Indicators Warehouse—the data hub for the HHS Community Health Data Initiative, a flagship HHS open government initiative to release data; encourage innovative application development; and catalyze change to improve community health: http://healthindicators.gov

Centers for Medicare and Medicaid Services (CMS): http://www.cms.gov/

- A family of consumer-oriented websites that provides information on quality of care delivered in various settings:
- Dialysis Facility Compare: http://www.medicare.gov/DialysisFacilityCompare/search.html
- Home Health Compare: http://www.medicare.gov/homehealthcompare/search.html
- Hospital Compare: http://www.medicare.gov/hospitalcompare/search.html
- Nursing Home Compare: http://www.medicare.gov/nursinghomecompare/search.html

Indian Health Service: http://www.ihs.gov

U.S. Census Bureau American Fact Finder: http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml

Cancer Data from the National Cancer Institute SEER program: http://seer.cancer.gov

Department of Veterans Affairs: ASPIRE is a web-based dashboard that documents quality and safety goals for all VA hospitals. These data show strengths and opportunities for improvement at the national, regional, and local hospital level. The data are updated regularly.

ASPIRE: http://www.hospitalcompare.va.gov/HOSPITALCOMPARE/aspire/index.asp

PRIVATE, NONPROFIT ORGANIZATIONS

Commonwealth Foundation Why Not the Best?—measures of hospital quality that are publicly reported by Center for Medicare and Medicaid Services: http://www.whynotthebest.org

National Quality Forum (NQF) Quality Positioning System—inventory of quality measures endorsed by NQF: http://www.qualityforum.org/QPS/QPSTool.aspx

The Joint Commission Quality Check—provides hospital-specific quality data from a variety of sources, including Joint Commission core measures (ORYX): http://www.qualitycheck.org/consumer/searchQCR.aspx

GENERAL REFERENCES

1. National Healthcare Quality and Disparities Reports. July 2013. Agency for Healthcare Research and Quality, Rockville, MD. Accessed February 10, 2015 at: http://www.ahrq.gov/research/findings/nhqrdr/index.html.
2. Hussey PS, Wertheimer S, Mehrotra A. The association between health care quality and cost: a systematic review. *Ann Intern Med*. 2013;158:27-34.

12 QUALITY OF CARE AND PATIENT SAFETY

ROBERT M. WACHTER

During the past two decades, scores of studies have demonstrated that the quality and safety of modern health care leave much to be desired, despite the fact that most physicians are well trained and work very hard. Yet the evidence is undeniable, with clear documentation of stunning variations in patterns of care that are neither supported by evidence nor justified by outcomes, major gaps between evidence-based best practices and current practice, and staggering numbers of serious medical errors. The recognition of these quality and safety problems has catalyzed a major transformation in thinking and practice, with new technologies, regulations, training models, incentive systems, and more.

To appreciate the problem and how to address it requires an understanding of quality measurement and improvement, the safety of patients, and value, which is the confluence of safety, quality, and cost.[1]

QUALITY

Definition

Quality of care has been defined by the Institute of Medicine as "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge." It includes six aims for a quality health care system, emphasizing that quality involves more than the delivery of evidence-based care (Table 12-1). Nevertheless, evidence-based medicine (Chapter 10) provides much of the scientific underpinning for quality measurement and improvement.[2] Previously, the lack of clinical evidence and the apprenticeship model of medical training promoted an idiosyncratic practice style by which a senior clinician or a marquee medical center determined the standard of care—a tradition now sometimes termed *eminence-based medicine*. Without discounting the value of experience and mature clinical judgment, the modern paradigm for determining optimal practice has changed, driven by the explosion in clinical research during the past 30 years; for example, the number of randomized clinical trials grew from 350 per year in 1970 to more than 27,000 per year in 2012. This research has helped define "best practices" in many areas of medicine, from preventive strategies for a healthy 62-year-old outpatient (Chapters 14 and 15) to the treatment of a patient with acute myocardial infarction and cardiogenic shock (Chapters 73 and 107).

Donabedian triad, which divides quality measures into *structure* (how care is organized), *process* (what is done), and *outcomes* (what happens to the patient), represents the most popular construct for quality measurement. Each element of the triad has important advantages and disadvantages as a quality measure (Table 12-2). Many of the widely used quality measures are process measures for which clinical research has established a link between such processes and improved outcomes. An example is the rate at which aspirin or a β-blocker is given to survivors of a myocardial infarction before hospital discharge (Chapter 73). However, when processes are less relevant and the science of case-mix adjustment is suitably advanced (e.g., cardiac bypass surgery; Chapter 74), outcome measurement (e.g., risk-adjusted mortality rate or 30-day readmission rate) is increasingly used. In other areas involving complex processes, structural measures are used as proxies for quality; examples here include the presence of intensivists to staff critical care units, a dedicated stroke service, and computerized physician order entry systems.

TABLE 12-1 THE INSTITUTE OF MEDICINE'S SIX QUALITY AIMS

Patient safety
Patient centeredness
Effectiveness
Efficiency
Timeliness
Equity

From Committee on Quality of Health Care in America, Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.

The Epidemiology of Quality-Related Problems

It is now well established that there are large and clinically indefensible variations in care from one city to another. Furthermore, U.S. practice adheres to the best evidence only slightly more than 50% of the time, even when adherence is known to correlate with ultimate clinical outcomes.

Levers for Change

For physicians, policymakers, administrators, and patients, evidence of major problems with quality has led to the recognition of structural problems that prevent the delivery of the highest quality of care. These problems include the lack of information regarding the performance of a provider or institution, the absence of incentives for quality improvement, the challenge for practicing physicians to stay abreast of modern evidence-based medicine, and the absence of an information technology support system for quality.

The first step in quality improvement is the creation of practice standards against which to measure quality. Scores of such measures have been promulgated by a variety of organizations, including payers (such as the Centers for Medicare and Medicaid Services), accreditors (such as the Joint Commission), and medical societies. These measures have identified many opportunities for improvement among individual physicians, practices, and hospitals.

Given the volume of new literature published each year, it is impossible for an individual physician to keep up with all the evidence-based advances in his or her field. *Practice guidelines,* such as those for the treatment of community-acquired pneumonia (Chapter 97) or the prophylaxis of deep venous thrombosis (Chapter 81), aim to synthesize evidence-based best practices into a set of summary recommendations.[3] Although concerns about "cookbook medicine" linger, there is a growing consensus that best practices should be "hard wired" if possible. The major challenges are to update guidelines as new knowledge accumulates and to recognize the complexity of guidelines when patients have multiple, potentially overlapping illnesses. *Clinical pathways* are similar to guidelines but attempt to codify a series of steps, usually temporally (on day 1, do the following; on day 2, do the

TABLE 12-2 COMPARISON OF THREE MEASURES OF CLINICAL QUALITY: THE DONABEDIAN TRIAD

MEASURE	SIMPLE DEFINITION	ADVANTAGES	DISADVANTAGES
Structure	How was care organized?	May be highly relevant in a complex health system	May fail to capture the quality of care by individual physicians Difficult to determine the "gold standard"
Process	What was done?	More easily measured and acted on than outcomes May not require case-mix adjustment No time lag—can be measured when care is provided May directly reflect quality (if carefully chosen)	A proxy for outcomes Not all may agree on "gold standard" processes May promote "cookbook" medicine, especially if physicians and health systems try to "game" their performance
Outcomes	What happened to the patient?	What we really care about	May take years to occur May not reflect quality of care Requires case-mix and other adjustment to prevent "apples-to-oranges" comparisons

Modified from Donabedian A. The quality of care. How can it be assessed? JAMA. 1988;270:1743-1748; and Shojania KG, Showstack J, Wachter R. Assessing hospital quality: A review for clinicians. Eff Clin Pract. 2001;4:82-90.

following; and so forth), making them more useful for stereotypical processes such as the postoperative management of patients after hip replacement. As more health care delivery organizations become computerized, pathways and guidelines are often translated into *order sets* or *clinical decision support systems* to guide clinicians at the point of care.

Although professionalism (Chapter 1) should be a sufficient incentive for physicians to provide high-quality care, reaching this goal typically depends on the existence of a system organized to translate research into practice and to deliver the right care every time. Such a system requires significant investments (in educating physicians, hiring case managers or clinical pharmacists, building information systems, and developing guidelines). The historical payment system, which compensates physicians and hospitals on the basis of volume rather than quality, provides no incentive to make the requisite investments, but this situation is changing rapidly.

The Changing Environment for Quality

The recent recognition of major gaps in quality and of the need for systemic change to improve quality has led to a variety of initiatives to catalyze quality improvement. Virtually all involve several steps: defining reasonable quality measures (evidence-based measures; capturing appropriate structures, process, or outcomes), measuring the performance of providers or systems, and using these results to promote change. This final imperative creates the greatest degree of uncertainty and experimentation.

Although one might hope that simply giving a physician information about prior performance would generate meaningful improvement, this strategy yields only modest change at best. Increasingly, a more aggressive and transparent strategy, such as disseminating the results of quality measurement to key stakeholders, is being adopted. In some cases, simple transparency is the main strategy—the rationale being that providers will find the exposure of their gaps in quality to be sufficiently concerning or embarrassing to motivate improvement. Although there is little evidence that patients use such data to choose among physicians or hospitals, transparency itself has frequently resulted in impressive improvements in some publicly reported quality measures.

The newest strategy in the United States is to tie payments for service to quality performance (pay for performance, or P4P). A number of P4P programs are under way, but early results indicate that differential payment leads to surprisingly modest gains beyond the improvements achieved by simple transparency.[4,5] P4P also raises a host of concerns, including whether presently captured quality data are accurate, whether payments should go to the best performers or those with the greatest improvements, whether existing measures adequately measure quality in patients with complex diseases, and whether P4P will create undue focus on certain measurable practices, leading to relative inattention to other important processes that are not being compensated. Another concern is that an overemphasis on "extrinsic" motivation (i.e., bonus payments) can actually extinguish "intrinsic" motivation (i.e., professionalism). One variation on the P4P theme is Medicare's "no pay for adverse events" program, in which hospital payments are withheld for certain "preventable" adverse events, such as injuries from falls or health care–associated infections. As with P4P more generally, the impact of such programs on quality and safety has been surprisingly modest.

Quality Improvement Strategies

Whether the motivation is professionalism, embarrassment, or economics, the next question is how actually to improve the quality of care. There is no simple answer; successful institutions and physicians have used a variety of strategies. In general, most use a variation of a "plan, do, study, act" (PDSA) cycle, recognizing that quality improvement activities must be carefully planned and implemented, that their impact needs to be measured, and that the results of these activities are often imperfect and require retooling.

In addition to the PDSA cycle, several other types of activities are useful. For quality improvement practices that require predictable repetition, efforts to "hard wire" the practice or to use alternative providers who focus on the activity are often beneficial. For example, the best strategy to increase the rate of pneumococcal vaccination (Chapter 18) among hospitalized patients with pneumonia is to embed it in a standard order set, either paper based or computerized. Another example is that having a nurse remove patients' shoes before the physician's entry can increase rates of diabetic foot examinations in an outpatient practice (Chapter 229).

In some areas, though, quality improvement involves much more complex and interdependent activities. In these circumstances, bringing teams together to examine their practices and to participate in a PDSA cycle is the most likely path to success. For example, a group of cardiac surgeons in the northeastern United States participated in an experiment in which they observed one another's practices, agreed on best practices, and measured one another's outcomes; the result was a 24% reduction in mortality with cardiac surgery. Many health care organizations are adopting one of the more sophisticated methodologies, such as Lean or Six Sigma, which involve mapping out all of the steps of a complex process (e.g., hospital admission) in an attempt to root out waste. For a hospital or clinic, the precise methodology chosen is probably less important than the decision to adopt a single way of approaching complex processes in need of improvement.

PATIENT SAFETY

Epidemiology

The concept of "first, do no harm" began more than 2 millennia ago, and many hospitals host periodic forums (e.g., morbidity and mortality conferences) to discuss errors. Until recently, however, there has been little teaching about the nature of medical mistakes, investment in safety research, regulation of safety standards, or emphasis on safety improvements, despite the fact that an estimated 44,000 to 98,000 Americans die each year of medical mistakes—the equivalent of a jumbo jet crashing each day. Such deaths may be related to medication errors, gaps in the discharge process, communication problems in intensive care units, or retained sponges in surgical patients—in short, virtually every aspect of modern medical care. Moreover, detailed clinical and statistical evidence of suboptimal safety has been reinforced by several high-profile and disquieting errors, sometimes apparently related to inadequate supervision and prolonged duty hours of trainees. These errors include the wrong patient getting a major procedure, the wrong limb being operated on, chemotherapy overdoses, mistaken mastectomies, and more. In the past few years, new classes of errors have emerged because of poorly designed health care information systems.[6] In addition, increasing attention is focusing on areas that were previously underemphasized, such as diagnostic errors.[7]

Because patients may be harmed despite receiving perfect care (i.e., from an accepted complication of surgery or a side effect of medication), it is important to separate *adverse events* from *errors.* The patient safety literature commonly defines an error as "an act or omission that leads to an unanticipated, undesirable outcome or to substantial potential for such an outcome." Adverse events, in contrast, are injuries due to medical management rather than the patient's underlying illness. This distinction is crucial. For example, when a patient who was appropriately prescribed warfarin for chronic atrial fibrillation develops a gastrointestinal bleed despite a therapeutic international normalized ratio, an adverse event, not a medical error, has occurred. Conversely, if the international normalized ratio was supratherapeutic because the physician prescribed a new medication without checking for possible drug interactions, a medical error would have occurred.

The Modern Approach to Patient Safety

The historical approach to medical errors often has been to blame the provider who was most proximate: whoever performed the surgery, hung the intravenous medication, or mixed the chemotherapy. It is now recognized that this approach fails to appreciate that most errors are committed by hard-working, well-trained individuals, and such errors are unlikely to be prevented by admonishing people to be more careful or by shaming and suing them. Instead, the modern approach, known as *systems thinking*, holds that humans will inevitably err and that safety depends on creating systems that anticipate errors and either prevent or catch them before they cause harm. Such an approach has been the cornerstone of safety improvements in other high-risk industries for some time.

The "Swiss cheese" model of accidents, drawn from innumerable investigations of accidents in commercial aviation and the nuclear power industry, for example, emphasizes that single errors by one individual working in an otherwise safety-conscious system rarely cause harm. Instead, such errors must penetrate multiple incomplete layers of protection ("layers of Swiss cheese") to cause terrible harm. The lesson is to focus not on the futile goal of trying to perfect human behavior but rather on creating multiple overlapping layers of protection to decrease the probability that the holes in the Swiss cheese will ever align, allowing an error to slip through.

How to Improve Patient Safety

Drawing on these models, modern thinking emphasizes efforts to design and implement systems to prevent or catch errors. For example, errors in

routine behaviors can best be prevented by building in redundancies and crosschecks in the form of checklists, read-backs, and other standardized safety procedures, such as counting sponges in the operating room, signing a surgical site before an operation, or asking patients their names before administering a medication. In recent years, the use of checklists for the placement of central lines and to prepare patients for surgery has resulted in remarkable reductions in morbidity and mortality.[8] One way to decrease errors at the person-machine interface is by the use of "forcing functions," engineering solutions that decrease the probability of human error. The classic example outside of medicine is the modification of automobile braking systems to make it impossible to place a car in reverse when the driver's foot is off the brake. In health care, forcing functions include changing the gas nozzles and connectors so that anesthesiologists cannot mistakenly hook up the wrong gas, such as nitrogen instead of oxygen, and administer it to a patient. Given the ever-increasing complexity of modern medicine, building in such forcing functions in intravenous pumps, defibrillators, mechanical ventilators, and computerized order entry systems will be crucial to safety.

In addition to better systems, communication and teamwork must be improved. All commercial pilots must take "crew resource management" courses, in which they train for emergencies with other crew members, learn to flatten hierarchies that might stifle open communication, communicate clearly with standard language, and use checklists and other systematic approaches. The evidence that such interventions in medical care will improve the safety of patients is increasingly persuasive.[9] For example, better coordinated pharmacy practices can reduce medication errors after hospital discharge,[A2] and team-based approaches can reduce falls among hospital inpatients.[A3] The goal is a "culture of safety"—an environment in which teamwork, clear communication, and openness about errors, both with other health care professionals and with patients, is the norm.

Another key principle in ensuring the safety of patients is to learn from one's mistakes. Safe systems have a culture in which errors are openly discussed, often in morbidity and mortality conferences. To be most useful, these discussions should be interdisciplinary (involving physicians and other health professionals), identify when the errors occurred, and emphasize systems thinking and solutions; they should not be punitive. In addition to open discussions during conferences, safe organizations build in mechanisms to hear about errors from frontline staff, often through "incident reporting systems"; they also perform detailed "root cause" analyses of major errors or "sentinel events" in an effort to define all the layers of Swiss cheese that need improvement. The importance of open communication extends to patients as well. Disclosure of errors is now required by the Joint Commission. Patients and families value such openness, and reasonably strong evidence indicates that disclosure of errors might decrease the chance of a malpractice suit.[10]

Finally, there is increasing appreciation of the importance of a well-trained, well-staffed, and well-rested work force for the delivery of safe care. Lower nurse-to-patient ratios,[11] long work hours for residents, and lack of board certification are all linked to poor outcomes for patients. Safer systems cannot be created if the providers are overextended or poorly trained or supervised. In the United States, the Accreditation Council for Graduate Medical Education has limited duty hours for residents and has prohibited first-year residents from working 24-hour shifts. Evidence so far confirms that these standards have improved residents' quality of life but not patient safety, probably because of the concomitant increase in risky handoffs.[12]

In the absence of comparative evidence and in light of the high cost of interventions such as improved staffing, computerized order entry, and teamwork training, even institutions committed to safety must often make difficult choices. Given the natural tendency to focus on practices that are measured, publicly reported, and compensated, institutions and physicians tend to focus first on areas that are subject to regulation or on initiatives with multiple potential benefits, such as computerization. For example, computerization has been promoted in the United States by a federal incentive program, which provides billions of dollars to hospitals and physicians that implement computer systems meeting certain "Meaningful Use" standards.

Because improving culture is difficult to measure and to regulate, there is concern that it will not be as high a priority as it should be. Moreover, a safe culture depends on balancing the imperative to improve systems with the need to define and enforce accountability. The safety field is increasingly emphasizing more active enforcement of policies that address problems such as disruptive behavior by clinicians and failure to adhere to evidence-based safety practices.

VALUE: CONNECTING SAFETY AND QUALITY TO COST

Outside of health care, most purchasing decisions are based on perceived value: (quality + safety) ÷ cost. Health care decisions historically have not been made this way, in part because of the limited ability of patients and payers to make rational judgments about the quality and safety of a given provider or system, and in part because health care insurance insulates patients from the full cost of care. In the United States, which spends nearly 20% of its Gross Domestic Product on health care, policy pressures are increasingly being brought to bear on the entire value equation, that is, promoting value, not volume. For example, Medicare's Value-based Purchasing program modifies hospital reimbursement on the basis of quality, safety, and patient satisfaction scores. The organization's readmission initiative threatens substantial cuts in reimbursement to hospitals with higher-than-expected 30-day readmission rates.[13] Other federal programs seek to drive physicians and hospitals into arrangements in which they accept a fixed payment to manage a population of patients (Accountable Care Organizations) or for an episode of illness (bundled payments). Although such programs are controversial and of unproven value, they are part of a growing focus on quality, safety, the patient's experience, and the costs of care.

Grade A References

A1. Jack BW, Chetty VK, Anthony D, et al. A reengineered hospital discharge program to decrease rehospitalizations: A randomized trial. *Ann Intern Med*. 2009;150:178-187.

A2. Kripalani S, Roumie CL, Dalal AK, et al. Effect of a pharmacist intervention on clinically important medication errors after hospital discharge: a randomized trial. *Ann Intern Med*. 2012;157:1-10.

A3. Dykes PC, Carroll DL, Hurley A, et al. Fall prevention in acute care hospitals: a randomized trial. *JAMA*. 2010;304:1912-1918.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bohmer RM. The four habits of high-value health care organizations. *N Engl J Med*. 2011;365:2045-2047.
2. Shekelle PG, Pronovost PJ, Wachter RM, et al. Making Health Care Safer: A critical review of modern evidence supporting strategies to improve patient safety. *Ann Intern Med*. 2013;158(Pt 2):365-440.
3. Wyatt KD, Stuart LM, Brito JP, et al. Out of context: clinical practice guidelines and patients with multiple chronic conditions: a systematic review. *Med Care*. 2014;52(Suppl 3):S92-S100.
4. Jha AK, Joynt KE, Orav EJ, Epstein AM. The long-term effect of Premier pay for performance on patient outcomes. *N Engl J Med*. 2012;366:1606-1615.
5. Kristensen SR, Meacock R, Turner AJ, et al. Long-term effect of hospital pay for performance on mortality in England. *N Engl J Med*. 2014;371:540-548.
6. Sittig DF, Singh H. Defining health information technology-related errors: new developments since To Err Is Human. *Arch Intern Med*. 2011;171:1281-1284.
7. Singh H, Graber ML, Kissam SM, et al. System-related interventions to reduce diagnostic errors: a narrative review. *BMJ Qual Saf*. 2012;21:160-170.
8. Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med*. 2009;360:491-499.
9. Neily J, Mills PD, Young-Xu Y, et al. Association between implementation of a medical team training program and surgical mortality. *JAMA*. 2010;304:1693-1700.
10. Kachalia A, Kaufman SR, Boothman R, et al. Liability claims and costs before and after implementation of a medical error disclosure program. *Ann Intern Med*. 2010;153:213-221.
11. Needleman J, Buerhaus P, Pankratz VS, et al. Nurse staffing and inpatient hospital mortality. *N Engl J Med*. 2011;364:1037-1045.
12. Fletcher KE, Reed DA, Arora VM. Patient safety, resident education and resident well-being following implementation of the 2003 ACGME duty hour rules. *J Gen Intern Med*. 2011;26:907-919.
13. Joynt KE, Jha AK. A path forward on Medicare readmissions. *N Engl J Med*. 2013;368:1175-1177.

REVIEW QUESTIONS

1. Which one of the following is the current number of randomized clinical trials published each year?

A. 1000
B. 5000
C. 13,000
D. 19,500
E. 27,000

Answer: E Without discounting the value of experience and mature clinical judgment, the modern paradigm for determining optimal practice has changed, driven by the explosion in clinical research during the past 30 years; for example, the number of randomized clinical trials grew from 350 per year in 1970 to more than 27,000 per year in 2012.

2. Which one of the following would *not* be characterized as process measures under Donabedian triad?

A. Fraction of patients with myocardial infarction who received β-blockers
B. Fraction of patients with stroke who died within 30 days (risk adjusted)
C. Fraction of hospitalized patients with pneumonia given pneumococcal vaccination
D. Fraction of hospitalized patients with heart failure with documented predischarge instructions
E. Fraction of diabetic outpatients who received eye examinations

Answer: B Many of the widely used quality measures are process measures for which clinical research has established a link between such processes and improved outcomes. An example is the rate at which aspirin or a β-blocker is given to survivors of a myocardial infarction before hospital discharge (Chapter 73). However, when processes are less relevant and the science of case-mix adjustment is suitably advanced (e.g., cardiac bypass surgery; Chapter 74), outcome measurement (e.g., risk-adjusted mortality rate or 30-day readmission rate) is increasingly used. In other areas involving complex processes, structural measures are used as proxies for quality; examples here include the presence of intensivists to staff critical care units, a dedicated stroke service, and computerized physician order entry (CPOE) systems.

3. Which of one the following is a policy lever that has been used to promote improved quality and safety?

A. Public reporting of performance
B. Pay for performance
C. No pay for errors
D. Appeal to professionalism
E. All of the above

Answer: E The recent recognition of major gaps in quality and of the need for systemic change to improve quality has led to a variety of initiatives to catalyze quality improvement. Virtually all involve several steps: defining reasonable quality measures (evidence-based measures; capturing appropriate structures, process, or outcomes), measuring the performance of providers or systems, and using these results to promote change. This final imperative creates the greatest degree of uncertainty and experimentation.

Although one might hope that simply giving a physician information about prior performance would generate meaningful improvement, this strategy yields only modest change at best. Increasingly, a more aggressive and transparent strategy, such as disseminating the results of quality measurement to key stakeholders, is being adopted. In some cases, simple transparency is the main strategy—the rationale being that providers will find the exposure of their gaps in quality to be sufficiently concerning or embarrassing to motivate improvement. Although there is little evidence that patients use such data to choose among physicians or hospitals, transparency itself has frequently resulted in impressive improvements in some publicly reported quality measures.

The newest strategy in the United States is to tie payments for service to quality performance (pay for performance, or P4P). A number of P4P programs are under way, and early results indicate that differential payment leads to surprisingly modest gains beyond the improvements achieved by simple transparency (Jha AK, Joynt KE, Orav EJ, Epstein AM. The long-term effect of Premier pay for performance on patient outcomes. *N Engl J Med*. 2012;366:1606-1615). P4P also raises a host of concerns, including whether presently captured quality data are accurate, whether payments should go to the best performers or those with the greatest improvements, whether existing measures adequately measure quality in patients with complex diseases, and whether P4P will create undue focus on certain measurable practices, leading to relative inattention to other important processes that are not being compensated. The behavioral economics literature also warns that an overemphasis on "extrinsic" motivation (i.e., bonus payments) can actually extinguish "intrinsic" motivation (i.e., professionalism). One variation on the P4P theme is Medicare's "no pay for adverse events" program, in which hospital payments are withheld for certain "preventable" adverse events such as injuries from falls or health care–associated infections (Jack BW, Chetty VK, Anthony D, et al. A reengineered hospital discharge program to decrease rehospitalizations: a randomized trial. *Ann Intern Med*. 2009;150:178-187). As with P4P more generally, the impact of such programs on quality and safety has been surprisingly modest.

4. The "Swiss cheese model" for patient safety refers to which one of the following?

A. The importance of ensuring that the workforce is well fed and rested
B. The ability of Swiss trains and watches to run on time
C. The observation that many errors relate to the failure of multiple incomplete layers of protection
D. The inadequate incentives to promote safety
E. The tendency for computer systems to create new classes of errors

Answer: C The "Swiss cheese" model of accidents, drawn from innumerable investigations of accidents in commercial aviation and the nuclear power industry, for example, emphasizes that single errors by one individual working in an otherwise safety-conscious system rarely cause harm. Instead, such errors must penetrate multiple incomplete layers of protection ("layers of Swiss cheese") to cause terrible harm. The lesson is to focus not on the futile goal of trying to perfect human behavior but rather on creating multiple overlapping layers of protection to decrease the probability that the holes in the Swiss cheese will ever align, allowing an error to slip through.

5. Which one of the following would be an accurate equation to capture health care "value"?

A. Quality × safety ÷ patient satisfaction × cost
B. Quality ÷ patient satisfaction × cost × safety
C. Cost ÷ quality
D. Quality × safety × patient satisfaction ÷ cost
E. Quality × cost ÷ patient satisfaction

Answer: D Outside of health care, most purchasing decisions are based on perceived value: (quality + safety) ÷ cost. Health care decisions historically have not been made this way, in part because of the limited ability of patients and payers to make rational judgments about the quality and safety of a given provider or system, and in part because health care insurance insulates patients from the full cost of care. In the United States, which spends nearly 20% of its Gross Domestic Product on health care, policy pressures are increasingly being brought to bear on the entire value equation ("promoting value, not volume"). For example, Medicare's Value-based Purchasing program modifies hospital reimbursement based on quality, safety, and patient satisfaction scores.

13

COMPREHENSIVE CHRONIC DISEASE MANAGEMENT

EDWARD H. WAGNER

The World Health Organization defines chronic disease as "health problems that require ongoing management over a period of years or decades." This definition encompasses a broad array of physical, mental, and behavioral health problems. Regardless of cause or pathophysiology, chronic conditions require ongoing attention and adjustments by patients and their loved ones as well as care by professionals. Improving care for chronic illnesses has been facilitated by the realization that individuals with a wide variety of chronic health problems have similar needs to minimize morbidity and to optimize quality of life (Table 13-1). Because these needs are shared across conditions, clinical management of these seemingly disparate illnesses requires similar practice capacities and functions. The design and organization of care that leads to better outcomes are remarkably similar for conditions as clinically different as diabetes, depression, and substance abuse disorders.

To live effectively with their health conditions and to manage their treatments, patients must have access to essential information, skills, and encouragement. They must also receive evidence-based therapy and preventive care over time to improve disease control and to reduce the risk of complications and exacerbations. Because chronic illnesses are rarely cured and often change over time, effective care involves continuous monitoring and adjustments by patients and caregivers alike. When chronically ill patients experience periods of increased severity and risk, they often benefit from an intensification of management and support. Care for the chronically ill generally involves multiple health professionals and care settings, and it must be coordinated effectively.

TABLE 13-1 COMMON NEEDS OF PATIENTS WITH CHRONIC ILLNESS

Support and information that enables patients to be competent self-managers of their health and illness
Effective clinical and behavioral treatment that keeps the condition under control and optimizes health status
Effective preventive care to reduce the risk of complications and other morbidity
Ongoing monitoring of the patient's condition to detect and to respond to problems early in their course
Intensification of management and support during high-risk periods
Coordination of care to increase the efficiency and effectiveness of referrals and to prevent the mishaps that commonly occur during care transitions

THE GOALS OF CHRONIC CARE MANAGEMENT

The goals of chronic care management are to meet the aforementioned needs of patients with chronic illness routinely and efficiently. Chronic conditions, whether medical or psychiatric, present challenges for patients and their caregivers very different from those of acute illnesses or injuries. The decisions and behaviors undertaken by patients to deal with their illness, generally called self-management, influence the course and outcomes of most chronic diseases in major ways. Patients make decisions and take action in dealing with symptoms, coping with the social and emotional impacts of illness, monitoring their condition, taking medications, adjusting lifestyle, and interacting with the health care system. Most need training and ongoing support to become competent managers of their health and their illness. The competence and confidence with which patients self-manage have a major impact on outcomes. For example, participants in diabetes or hypertension self-management training programs generally experience clinically significant reductions in hemoglobin A_{1c} or blood pressure levels without major changes in drug therapy.■ Self-management can be enhanced in patients of all socioeconomic groups by empowering, training, and supporting them. Modern self-management support is a collaborative rather than a didactic process that seeks concordance between providers' and patients' perspectives on the goals of treatment and the actions needed to reach those goals.[1]

Whereas the primary goals of acute disease care are cure and recovery, cure is not an option for most chronic diseases, which are often characterized by slowly progressive deterioration, even with excellent care. Nevertheless, control of the metabolic or physiologic abnormalities or symptoms resulting from the underlying pathophysiologic processes (disease control) is now possible for most chronic conditions. Drugs and other therapies attempt to minimize morbidity, to limit further organ damage, to reduce the risk of exacerbations and complications, and to maintain quality of life and function. The metabolic or physiologic abnormalities or symptoms used to assess disease control often serve as clinical targets to guide therapy and as performance indicators to monitor the progress and quality of care for populations of patients (e.g., the percentage of diabetic patients with blood pressure <130/80 mm Hg). Disease control for many conditions requires the rigorous application of evidence-based treatment protocols that carefully step up or intensify treatment until clinical targets are reached.■

Ongoing monitoring and assessment of chronically ill patients are essential to optimize treatment and to prevent losses to follow-up. Regular review of self-management activities reinforces their importance to the patient and to the overall management of the condition. Serious exacerbations and complications of common chronic illnesses are potentially preventable if they are identified early in their course and treated appropriately (e.g., recurrence of major depression, opportunistic infection in patients with HIV, foot ulcers among diabetic patients). The periodicity of assessments must change as the severity of illness waxes and wanes over time.

An increasingly important goal of effective chronic care management is to provide more intensive monitoring and management during periods of high risk, such as transitions from hospital to community or exacerbations. Clinical care or case management by nurses, pharmacists, or other nonphysician health professionals enables closer monitoring of patients, helps with medication adjustment, provides self-management support, and facilitates care coordination. Effective care management programs can reduce the likelihood of rehospitalization for chronically ill patients discharged from the hospital, reduce emergency department visits and hospitalizations among multiproblem ambulatory older adults, and improve disease control.

Chronically ill patients frequently receive medical and supportive services from multiple providers in different settings. Without good coordination, unnecessary hospital readmissions and emergency department visits, gaps and inefficiencies in care, and distress of the patient are all too common. Without assistance from their primary providers, patients or their loved ones must often assume responsibility for the onerous task of coordinating care.

MATCHING PATIENT NEEDS AND CARE DELIVERY

The percentage of chronically ill patients in good control varies widely from practice to practice, even after adjustment for patient differences. When individual practices are audited, failures to follow evidence-based guidelines or losses to follow-up appear almost randomly across different interventions and among patients within a practice, suggesting that they are related to flaws in care systems rather than to cognitive gaps. What, then, distinguishes practices that have high rates of control of major chronic conditions from the majority with much lower control rates?

Ambulatory care systems have for centuries been organized to react to acute problems, not to address the ongoing needs of the chronically ill. The focus on making a diagnosis and initiating treatment for the problem at hand leaves little time for addressing less urgent needs, such as medication adjustment, self-management support, and preventive care. Reimbursement systems that favor multiple short encounters aggravate the problem.

Failure to adhere to evidence-based guidelines and insufficient follow-up and attention to patient self-management are largely responsible for the poor control of chronic illness. The guidelines for most chronic conditions include recommendations to step up or to intensify therapy when clinical targets are not reached. Failure to intensify treatment in patients who have not achieved therapeutic targets, called clinical inertia, has been found in a large percentage of patients with uncontrolled diabetes, hypertension, depression, and other chronic illnesses. Clinicians have an understandable reluctance to increase drug doses or to add new drugs, but concerns about toxicity or nonadherence of the patient can be mitigated if practices have organized approaches to stepping up therapy and closely monitoring its impact.

Deficiencies in practice infrastructure and the professional environment compound the difficulties of meeting the needs of chronically ill patients. Despite electronic medical record systems, many practices still have difficulty in managing patient populations or measuring the quality of their care. Primary care physicians are increasingly uninvolved or even unaware when their patients are hospitalized. Efforts to improve the quality of chronic care must focus on improving the infrastructure and practice systems that busy clinicians require to meet the needs of their patients.

Interventions That Improve Care and Outcomes of Chronic Diseases

A wide array of systemic changes can improve the care and outcomes of major chronic diseases. These changes fall into four general categories. *Patient-directed interventions* try to enhance the knowledge of patients, increase their involvement in care, and alter their behavior. *Provider-directed interventions* give providers feedback about the quality of their care and try to change providers' knowledge and behavior through education and reminders. *Organizational changes* generally focus on the composition and functioning of the care team, the organization of patient encounters (e.g., planned visits), and the management of patients outside the office (e.g., follow-up, care management, and care coordination). *Information technology interventions* include the use of registries, computer reminders, and other decision support programs for management of populations and patients. Growing evidence suggests that the use of patient databases (registries) to measure performance, to identify individuals needing care, and to plan care of individual patients may be the information technology function that contributes the most to improving the care of the chronically ill.

Provider-directed interventions, especially educational programs, generally demonstrate weak effects. Conversely, interventions that promote the patient's self-management and change the composition or functioning of the practice team have the most salutary effects.■■ Across conditions, assigning responsibilities for care of patients to the nonphysician members of practice teams consistently leads to significant improvements in evidence-based care processes, disease control, and other outcomes. Better informed patients have better outcomes when they receive care from more informed providers who are supported by a well-organized clinical team and appropriate information technology.

The Chronic Care Model

Developed in the late 1990s, the chronic care model summarizes the basic elements for improving care of chronically ill individuals (Fig. 13-1). This

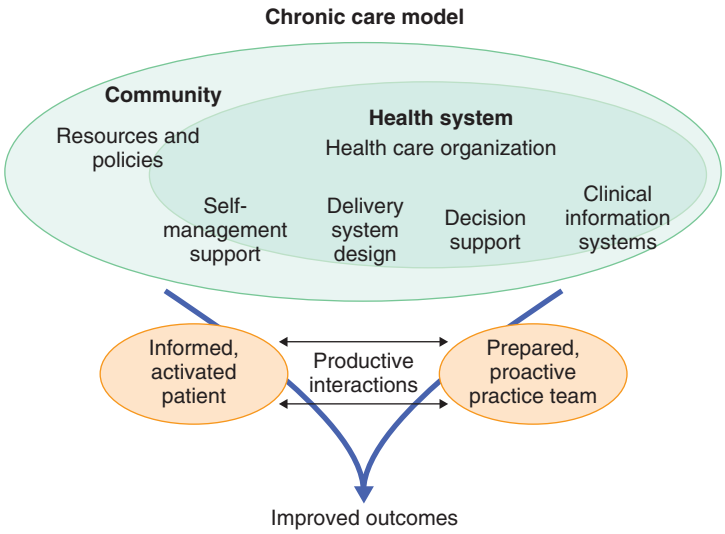

FIGURE 13-1. The chronic care model. (Modified from Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract. 1998;1:2-4. Chronic Care Model © American College of Physicians, Annals of Internal Medicine.)

model identifies six features of a health care system that facilitate high-quality chronic disease care: self-management support, delivery system design, decision support, clinical information systems, health care organization, and community resources. Evidence-based changes in these six areas foster productive interactions between informed patients who take an active part in their care and practice teams organized to meet their needs.

The chronic care model presumes that chronically ill patients have a primary care clinician, either a generalist or a specialist, who assumes responsibility for managing their care and coordinating the activities of other clinicians and institutions that provide advice and services. It posits that important patient outcomes, such as the prevention of morbidity, mortality, and avoidable emergency department visits and hospitalizations, can be affected by productive medical care. Accumulating evidence indicates that implementing the elements of the chronic care model is associated with improvements in the care and outcomes of chronic diseases.[1] Productive interactions are more likely when patients actively participate in their care. To do so, patients benefit from having the relevant information and skills needed to manage their illness as well as ongoing support and encouragement to overcome denial and passivity.

The other partner in productive interactions is the primary care clinician and the practice team. Before visits with chronically ill patients, practice teams should review accessible information to determine what services are needed, have clear assignments or standing orders for the delivery of those services, and include staff trained to perform them. Although all six elements of the chronic care model contribute to a delivery system that can effectively manage chronic illness, changes in two categories—delivery system design and self-management support—are essential to improve disease control and to reduce morbidity and mortality.

Self-Management Support

Didactic education of the patient alone has been demonstrated to have little if any impact on the patient's behavior or disease control. To change their lifestyle, to take medications as directed, to deal with symptoms and stress, and to address the other challenges of living with chronic illness, patients need to participate actively in their care, to understand and agree with the clinician's assessment of the problems and recommendations, and to learn the skills needed to carry out the recommendations. More active patients are more likely to engage in relevant self-management behaviors and to use health care more effectively,[2] so patients should be empowered to take an active role in their health care.

Group and individual interventions can help patients understand their illness and its treatment and give them the skills and confidence needed to be competent self-managers. In general, these programs share common features that include the teaching of critical skills (e.g., self-monitoring, use of medications), the collaborative development of realistic goals, and action plans for meeting these goals. Although self-management courses facilitated by peer or professional leaders improve disease control in diabetes, hypertension, and other chronic diseases, their effects diminish after termination of the program, so sustained follow-up and reinforcement are essential. Long-term self-management support is best accomplished by the primary care team in the context of ongoing chronic illness care. Although a physician's advice and encouragement are important, most primary care clinicians have neither the time nor the training to help patients set behavioral goals, to develop action plans, and then to follow up by telephone or e-mail. Clinical staff such as nurses or medical assistants with good communication skills and additional training in counseling methods can and should perform these functions.

Delivery System Design

Team Care

The goal of delivery system design is to match the organization of care delivery to the needs of the chronically ill. Without the help of a team, it is unlikely that physicians, in the course of a 15- to 20-minute visit, can effectively manage intercurrent problems, review and adjust treatment, provide recommended assessments and preventive care, discuss self-management goals and plans, and plan follow-up. Designing effective team care begins with a consideration of the various tasks required to meet the needs of the chronically ill and to ensure adherence to guidelines. Key tasks are allocated to the most appropriate members of the practice team. The team's routine involvement in patient visits is facilitated by protocols or standing orders that guide independent action by nonclinician staff. Brief meetings, often termed huddles, before clinic sessions allow practice teams to review key data (often from registries) of scheduled patients, to identify needed services, and to plan their delivery. The goal is to maximize the productivity of every patient interaction.

The specific practice change associated with the largest improvements in chronic disease outcomes is the increased involvement of the nonprovider members of practice teams in meeting the patient's clinical care needs.[3,4] For example, a recent trial showed that adding briefly trained lay "care guides" to primary care teams significantly increased the likelihood that patients would achieve disease-specific goals, such as diabetic retinal examinations and effective blood pressure control.[5]

Planned Interactions

Much of chronic care management involves predictable preventive care or disease management activities, which are often postponed when chronically ill patients seek care for an acute problem. The availability of key patient data gives practice teams the opportunity to update chronic illness care even during visits for more urgent problems. Armed with information from an electronic medical record's reminder systems, nonphysician staff can provide flu shots, diabetic foot examinations, and other evidence-based actions. Practices also can initiate planned visits for patients who are noncompliant with guidelines, have failed to reach treatment targets, or have been lost to follow-up. Another format for the delivery of planned care is the group visit, in which patients receive their primary medical care together as a group. Group sessions generally include the same check-in activities that occur during an individual visit, followed by brief individual communications between the primary care provider and each patient, opportunities for private consultations, and an educational session with ample opportunity for peer-to-peer interaction. Approximately 30 to 50% of patients offered group visits attend them, and many prefer to receive much of their care in this setting.

Follow-up and Case Management

Follow-up tailored to a patient's needs and the clinical severity of the disease is a critical component of effective chronic care management. Traditionally, follow-up consisted of periodic in-person physician visits whose frequency is limited by cost and convenience. Ample evidence indicates that follow-up by telephone, e-mail, or telemedicine can be cost-effective and far more flexible than total reliance on repeated face-to-face return visits. Nonphysician team members, guided by protocols with clear referral criteria, can effectively manage much of electronic follow-up. Self-monitoring by patients is an important component of the follow-up plans for many chronic illnesses and is especially useful when patients use the results to adjust their therapy[6] as well as to inform their practice team. Collecting self-monitoring data for the sole purpose of bringing them to the physician's office appears to be far less effective.

Chronically ill individuals at high risk of hospitalization, nursing home placement, major complications, or death often experience periods when more intensive monitoring and management would be helpful. Less severely

ill patients also need closer follow-up during exacerbations or intercurrent illnesses, after hospital discharge, or when medications are being titrated or changed. Severely ill patients commonly have multiple chronic conditions, which often are complicated further by depression and other psychosocial problems. In response to these needs, nurse care management programs that regularly communicate with patients electronically or in person have proliferated. These programs assess a patient's status, review and support self-management goals and action plans, help coordinate care among providers and care settings, and may help manage medications. Case management is more likely to be effective if the case manager collaborates closely with the primary care clinician, has specific management goals (e.g., improve disease control or improve function), influences the medication regimen directly or indirectly, and reviews his or her caseload regularly with clinician experts.[3,4]

Decision Support

Efforts to educate health professionals have a limited impact on clinical performance. Inserting information and alerts directly into the flow of decision making is more helpful, although "alert fatigue" has become a serious problem. Even the most carefully developed evidence-based guidelines will have no impact on practice if they are not integrated into clinical management through computer templates, alerts, protocols, standing orders, and other efforts to standardize practice.

Many chronically ill patients who are cared for by generalists benefit from the advice and involvement of medical specialists. Truly shared care involving interactive communication between primary care clinicians and specialists can improve outcomes.[A10] Interactive communication channels between generalists and specialists (Chapter 430) can be improved by creating systems whereby consultants respond to questions by secure messaging within an electronic medical record system or through a web-based referral system.

Clinical Information Systems

Well-maintained registries that are either independent or incorporated in an electronic medical record enable practices to identify patients who need additional services, to produce rapid summaries of key clinical data and services for future patient encounters, and to measure clinical performance. Registries allow practices both to monitor and to manage their chronically ill populations. For example, individuals overdue for important preventive interventions or who fail to keep appointments can be efficiently identified and contacted.

Electronic two-way communication between patients and providers is playing an increasing role in the management of chronic disease. Telehealth, web portals, and mobile device applications enable providers to obtain and to respond to clinical data from patients, and they give patients efficient mechanisms for addressing their questions and concerns.

Community Resources

Programs and organizations in patients' local communities can most effectively meet many of their needs. Such needs include transportation, homemaker services, smoking cessation (Chapter 32), exercise (Chapter 16), weight control (Chapter 220), peer support, caregiver support and respite, self-management training, and financial counseling and assistance. For commonly needed services, practices should at least be able to provide specific information to advise patients on their best options.

TRANSFORMING PRACTICE

Measures of disease control and other indicators of the quality of chronic care begin to improve only when practices have made system-wide changes in most of the elements of the chronic care model, such that the routine care of all their chronically ill patients has been affected.[5] Practice routines and culture must change in ways that are initially foreign and uncomfortable for many practitioners. Team care means more meetings, new roles, and additional training for staff. To have useful registries and to provide proactive care, busy practices must define their population of patients and learn to manipulate software and data to obtain the information they need. Some of the changes may require additional financial investment. For all these reasons and more, improving chronic care management requires highly motivated physicians and practices that actively engage in continuous quality improvement. Practices must develop approaches that fit their population of patients, resources, and practice style, and they must refine and adapt them using rapid cycle improvement methods.

Health Care Organization

Many practices caring for chronically ill individuals are parts of larger health care organizations that can either encourage and promote the improvement of chronic care or undermine and obstruct it. Helpful organizations promote continuous quality improvement, incorporate a trusted performance measurement system, and provide financial or nonfinancial incentives for high quality.

Grade A References

A1. Chodosh J, Morton SC, Mojica W, et al. Meta-analysis: chronic disease self-management programs for older adults. *Ann Intern Med*. 2005;143:427-438.

A2. McManus RJ, Mant J, Haque MS, et al. Effect of self-monitoring and medication self-titration on systolic blood pressure in hypertensive patients at high risk of cardiovascular disease: the TASMIN-SR randomized clinical trial. *JAMA*. 2014;312:799-808.

A3. Shaw RJ, McDuffie JR, Hendrix CC, et al. Effects of nurse-managed protocols in the outpatient management of adults with chronic conditions: a systematic review and meta-analysis. *Ann Intern Med*. 2014;161:113-121.

A4. Tricco AC, Ivers NM, Grimshaw JM, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet*. 2012;379:2252-2261.

A5. Coleman K, Austin BT, Brach C, et al. Evidence on the Chronic Care Model in the new millennium. *Health Aff (Millwood)*. 2009;28:75-85.

A6. Stellefson M, Dipnarine K, Stopka C. The chronic care model and diabetes management in US primary care settings: a systematic review. *Prev Chronic Dis*. 2013;10:E26.

A7. Miller CJ, Grogan-Kaylor A, Perron BE, et al. Collaborative chronic care models for mental health conditions: cumulative meta-analysis and metaregression to guide future research and implementation. *Med Care*. 2013;51:922-930.

A8. Adair R, Wholey DR, Christianson J, et al. Improving chronic disease care by adding laypersons to the primary care team: a parallel randomized trial. *Ann Intern Med*. 2013;159:176-184.

A9. Green BB, Cook AJ, Ralston JD, et al. Effectiveness of home blood pressure monitoring, Web communication, and pharmacist care on hypertension control: a randomized controlled trial. *JAMA*. 2008;299:2857-2867.

A10. Foy R, Hempel S, Rubenstein L, et al. Meta-analysis: effect of interactive communication between collaborating primary care physicians and specialists. *Ann Intern Med*. 2010;152:247-258.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Battersby M, Von Korff M, Schaefer J, et al. Twelve evidence-based principles for implementing self-management support in primary care. *Jt Comm J Qual Patient Saf*. 2010;36:561-570.
2. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. *Health Aff (Millwood)*. 2013;32:207-214.
3. Hickam DH, Weiss JW, Guise JM, et al. *Outpatient Case Management for Adults With Medical Illness and Complex Care Needs. Comparative Effectiveness Review, No. 99*. Rockville, MD: Agency for Healthcare Research and Quality; 2013.
4. Martinez-Gonzalez NA, Berchtold P, Ullman K, et al. Integrated care programmes for adults with chronic conditions: a meta-review. *Int J Qual Health Care*. 2014;26:561-570.
5. Improving Chronic Illness Care. http://www.improvingchroniccare.org/index.php?p=Toolkit&s=244. Accessed February 11, 2015.

REVIEW QUESTIONS

1. An obese patient with type 2 diabetes consistently has HbA1c levels above 10%. There is strong evidence that he is not following diet or exercise recommendations. He has recently attended diabetes education classes at the local hospital. Which of the following is the most appropriate next step?

A. Collaboratively establish realistic goals and an action plan and monitor closely.
B. Remind him of the dangers of uncontrolled diabetes.
C. Increase his medications.
D. Refer him to a diabetologist.

Answer: A It is highly likely that this gentleman's poor diabetes control is heavily influenced by his poor self-management. Before changing or increasing drug therapy, it would be most appropriate to have him spend time with someone (preferably in the practice) trained to use effective counseling methods like motivational interviewing to try to motivate him, develop achievable behavioral goals and a realistic action plan, and then follow him closely.

2. A physician receives performance metrics about her care of diabetes and hypertension. She is disappointed to find that her measures are below the average for her practice, even for simple things that she thought she was diligent about—for example, flu shots and regular HbA1c testing. She can't work any harder. Which of the following is the most appropriate next step?

A. Increase the time allotted for visits by diabetic or hypertensive patients.
B. Try to increase the involvement of her medical assistant and other available staff in the performance of evidence-based care.
C. Hire a nurse to manage routine diabetes and hypertension visits.
D. Add more frequent alerts to her EMR.

Answer: B There is considerable evidence that the involvement of all members of the practice team in meeting patient needs is the intervention that leads to the largest improvements in the delivery of evidence-based services and disease control for hypertensive and diabetic patients. Medical assistants and LPNs can review patient records before visits to identify needed services and, with standing orders, perform all those functions that state regulations allow. They can also be trained to provide self-management support.

3. Local health plans are providing performance bonuses to practices that can demonstrate reductions in their readmission rates for chronically ill patients. Which of the following is most likely to help a practice reduce their readmission rate?

A. Have practice clinicians make hospital rounds.
B. Tell patients to call for an appointment after discharge.
C. Hire and/or train a nurse care manager in the practice to work with recently hospitalized patients.
D. Ask hospital discharge planners to alert the practice that their patient is ready for discharge.

Answer: C Hospital readmissions can be reduced by ensuring timely, coordinated, and appropriately intensive post-hospital care. Hospitals often don't or can't identify a patient's primary care provider, so community practices must proactively try to identify hospitalized patients and make certain that they receive appropriate follow-up care. A care manager working in or closely with primary care is the best option.

4. Major depression is a common problem seen primarily in primary care. A practice is trying to develop local guidelines for the management of patients with this condition. They are trying to take into account the evidence of effectiveness as well as patient preferences and costs. Which of the following is the best strategy for them to employ?

A. Refer all patients either to a psychiatrist or therapist depending on patient preference.
B. Refer patients that don't respond to initial therapy to a mental health provider.
C. Give all patients a trial of antidepressant medication.
D. Identify a mental health provider consultant and care manager within to practice to help the primary care physicians manage the patients.

Answer: D There is very strong evidence that collaborative care models are superior to other ways of managing depressed patients in primary care. These models include ensuring that primary care physicians understand treatment guidelines, that staff in the practice are trained to provide follow-up assessments for depressed patients, and that a mental health provider is available to provide information and support to the primary care physician or care manager, and to see patients if necessary.

14

COUNSELING FOR BEHAVIOR CHANGE

F. DANIEL DUFFY

Most medical care requires patients to change some behavior. For example, patients may need to keep appointments, stop an addictive behavior, eat different foods, take daily medications, monitor glucose levels, or increase their physical activity. The probability of effectuating change depends on the skills and language of the counselor, recognizing that clinicians typically see patients during brief and relatively infrequent visits for prevention and chronic care.

Behavior change counseling is talk therapy that engages patients in a partnership to execute a plan for change.[1] Behavior change counseling delivered by trained counselors is efficacious in sustaining healthy diets, increased physical activity, reduced alcohol and tobacco use, improved dental outcomes, reduced body weight, and self-care monitoring. Trained physicians who provide brief office-based counseling using patient-centered motivational methods can help patients successfully lose weight and improve diabetes self-care management. Modest success has been achieved with office-based counseling for medication adherence but not for heavy drinking or problem drug use.

MOTIVATIONAL INTERVIEWING

Motivational interviewing is an evidence-based therapy that has been adapted to the clinical setting to counsel patients about behavior change.[2] Using this approach, clinicians avoid giving advice and instead ask open-ended questions and then use "reflective listening" to uncover internal ambivalence that may restrain change. The verbal behaviors of trained clinicians include listening carefully and responding to patients' voiced desires, fears, and ambivalence about changing; using statements to affirm patients' autonomy and ability; summarizing patients' self-arguments for and against change; and helping patients intensify their motivation to make changes. With persistence, these conversations can convert patients' ambivalence into the belief and confidence that they can and will change. The overall advantage of motivational interviewing is about 50% greater than comparative counseling methods.[3]

Change counseling in medical care is based on the transtheoretical model of change (Fig. 14-1), which proposes that people change by moving through a cycle of five cognitive-experiential stages (Table 14-1): *precontemplation*, *contemplation*, *determination*, *action*, and *maintenance*,[4] often with a sixth stage (*relapse*). Many people cycle several times before adopting a new habit. By identifying a patient's stage along this continuum of change, a physician can select the most efficient counseling approach.

FITTING CHANGE COUNSELING INTO MEDICAL PRACTICE

The goal of change counseling is to help patients do what they need to do to achieve their own health goals. Physicians help but cannot make people change. Patients do the work, guided by clinicians who evoke their internal motivation to make the change.

To learn what behaviors patients need to change, the physician starts by taking a history of active problems, by performing an appropriate physical examination (Chapter 7), and by ordering any indicated diagnostic tests (Chapter 10). To optimize motivational interviewing, the history should rely primarily on open-ended questions so patients can tell their own stories. The physician's periodic reflections, expressions of empathy, and summaries of what patients say help patients believe they can and must change.

THE ASK-TELL-ASK APPROACH

The process of change counseling helps patients focus on their role in achieving their own health goals. In contrast to a paternalistic practice of making a diagnosis, ordering treatment, and giving expert advice, it promotes a patient's motivation, autonomy, and responsibility for making changes.

Iterative *ask-tell-ask* cycles remind clinicians first to *ask* patients to engage; second to *tell* information, answer questions, and correct misinformation; and third to *ask* patients to verbalize their understanding and intentions.

To use the ask-tell-ask strategy to initiate a conversation about the diagnosis and treatment, clinicians first ask permission to summarize their findings. Second, if patients agree, clinicians tell their diagnostic impressions and health assessments using clear and simple language. Third, clinicians ask an engaging follow-up question to clarify patients' understanding with a question such as, *What do you know about this situation*? The third ask may take the form of a reflective statement that invites patients to go deeper, affirm their courage, express empathy, or acknowledge nonverbal expressions of emotion. Usually, people know more than expected, thereby reducing the time needed for education of the patient.

To focus patients' attention on their role in making changes, clinicians might ask, *What ideas do you have about what you and I might do to improve your health*? Generating a list of hypothetical options may take several ask-tell-ask iterations. When a satisfactory list has been developed, the physician should summarize the patient's voiced health goals and the shared list of options for achieving them.

Clinicians then initiate change planning with a question such as, *What are you willing to do*? Patients' answers indicate their readiness to take action to change their behavior. Clinicians should guard against the "expert trap," in which patients skirt their own ambivalence for change and shift responsibility for change onto the clinician by stating, *You are the expert, and I'll do whatever you tell me to do.*

To avoid the trap, clinicians can estimate patients' motivational stages using a *conviction-confidence ruler* (Fig. 14-2). Clinicians show patients a ruler with markings from 1 to 10 and ask, *How convinced are you that it is important for you to do what is needed from 1 (not convinced at all) to 10 (totally convinced)*? They then ask, *Using the same scale, how confident are you that you will do it*?

Patients with low (0 to 2) scores are probably in the precontemplation stage. Midrange scores (3 to 7) suggest the contemplation stage. A high conviction score (8 to 10) implies determination to change, and moderate to high confidence scores (5 to 9) imply the preparation stage. High scores for conviction and confidence indicate action or maintenance stages. In the relapse stage, patients may have a low confidence score and, if very discouraged, a low conviction score as well. Scores of less than 7 on either question may indicate insufficient motivation for success.

The conviction-confidence scale can clarify patients' convictions about the importance of changing by asking, *Why four on conviction (or confidence) and not lower*? Or one might ask, *What would it take to raise your conviction (or confidence) score to a nine*? Voiced answers to these question help patients clarify their values and beliefs and see their strengths and resources to make needed changes.

STAGE-SPECIFIC CHANGE COUNSELING

A patient-centered change plan is far more comprehensive than a medical treatment plan, which forms only one part of a person's health plan. For example, filling a prescription is one step in a treatment plan, but remembering to take the medication several times a day, altering eating habits, monitoring medication effects, and quitting a self-destructive behavior are ongoing tasks in a change plan. The most common error in medical care is treating lifestyle changes as a simple prescription and being disappointed with the lack of patient adherence to recommendations. Nevertheless, physicians often must first focus the change plan on ensuring that patients take their medications before addressing more complex and time-consuming lifestyle changes.

Precontemplation stage counseling encourages patients to seek information from reading material and websites and to talk with family, friends, or others who have successfully made similar changes. This process highlights the conflict between the risks of the status quo and the benefits of new lifestyle behaviors. Precontemplation is not influenced by scare tactics, professional argument, or debate.

Contemplation stage counseling, which is more difficult, requires time and training. Trained physicians or counselors help patients explore and resolve the natural ambivalence that keeps them from doing the work needed to change. A *decisional balance table* (Table 14-2) can contrast a patient's reasons for sustaining current behaviors with the reasons for changing to healthier ones. Clinicians or counselors listen to what patients say and verbally reflect a patient's ambivalence to help patients begin to convince themselves to change. This approach helps patients "think out loud" about their conflicting desire to change and their simultaneous wish to sustain the status quo.

When momentum stalls or patients backslide from doing their part, physicians often blame patients for resisting change or label them noncompliant. More likely, the difficulty lies with insufficient resolution of the natural ambivalence about change and evolving discord in the physician-patient relationship. Physicians and patients begin talking at cross purposes, become

TABLE 14-1 STAGES OF CHANGE CHARACTERISTICS

STAGE OF CHANGE	PATIENT RESPONSE	PATIENT CHANGE TASKS	CLINICIAN COUNSELING TASKS	COUNSELING RESOURCES
Precontemplation	Surprise about problem Not thinking about changing Demoralized if in relapse	Learn about condition and change needed Develop self-awareness	Reflective listening Advise Inform	Information media Self-assessment logs or diary
Contemplation	Ambivalence Change talk—concern about health risks Sustain talk—prefers the status quo	Self–re-evaluation Raise conviction of importance of change Raise confidence in ability to change	Empathize with patient's ambivalence Open questions Affirm the positive Reflective listening Summarize patients' talking themselves into change	Conviction-confidence ruler Decisional balance Feedback logs Role models Referral for group training
Determination	I must change I can do it	Raise importance Raise confidence Develop action plan Pick start date Tell others	Support self-efficacy Facilitate action plan Anticipate problems Encourage social support	Menu of options Referral for formal counseling List of role models
Action	What will I do? Who or what might help? What problems might I have?	Take action steps to change Manage withdrawal from addiction	Manage cues to old behavior Be alert to positive effect of new behavior Reward self for change Manage withdrawal	Written plan Menu of cues Menu of consequences Menu of rewards Refer for counseling
Maintenance	Pleased with changed self Spontaneous talk about success and difficulties	Manage cues to go back to old behavior Manage bad effects of the new behavior Seek and use social support for change Become a role model to others	Discuss signs and times of danger for relapse Discuss "relapse thinking" Affirm new lifestyle habits	Role model for others Social support sources Referral for maintenance counseling
Relapse	Tells about thinking old behavior was no longer a concern Loss of control Demoralization	Learn the antecedents of relapse Recognize "relapse thinking" Keep open, do not hide Call for help	Reframe relapse to be a valuable lesson Evoke self-efficiency Move quickly back into action	Refer for formal relapse counseling Encourage finding role models who made change Frequent follow-up

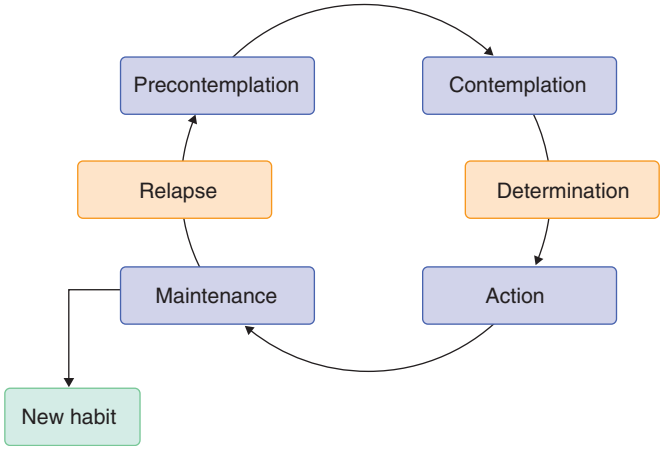

FIGURE 14-1. Cycle of change stages.

Conviction–Confidence Ruler												
On a scale of 1–10, how *convinced* are you that it is important for you to change … *(Name the behavior)* … ?"												
Not at all convinced	0	1	2	3	4	5	6	7	8	9	10	Totally convinced
On a scale of 1–10, how *confident* are you that you have the ability to change … *(Name the behavior)* … ?"												
Not at all confident	0	1	2	3	4	5	6	7	8	9	10	Totally confident

FIGURE 14-2. Motivational counseling aid.

TABLE 14-2 DECISIONAL BALANCE

BEHAVIOR	REASONS NOT TO CHANGE	REASONS TO CHANGE
Unhealthy behavior	What do you like about it?	What are your concerns about it?
Healthy behavior	What are your concerns about it?	What do you like about it?

defensive, blame, interrupt, and disengage. Physicians might resolve the relational discord with an apology, by affirming patients' strengths, or by shifting the focus of counseling to empathy with the patient's ambivalence and encouragement to believe that she or he can change.

Determination and action stage counseling reinforces a patient's confidence. Counseling might begin with the question, *What are you going to do and when will you begin?* Picking a start date and committing to other people generates accountability and social support. Because it is difficult to change alone, asking *Who or what might help you?* encourages patients to solicit support from partners or others who are making similar changes. Patients may use reminders and rewards to reinforce the new behavior while avoiding situations that stir craving for the old behavior. To plan for relapse prevention and recovery and to handle problems and side effects, clinicians might ask, *What problems might arise and how might you handle them?* As with all change counseling, action plans work best when patients identify the people, resources, and coping strategies themselves. When the change involves quitting an addictive behavior, the plan also should include treatments to manage withdrawal and craving (Chapters 32, 33, and 34).

Maintenance stage counseling is frequently ignored on the assumption that once patients take action to change, the work is done. Moving from action to maintenance requires the new behavior to be internalized and to become routine, requiring little thought or effort. Counseling bolsters the new behavior when "change boredom" sets in, life stresses and competing priorities weaken commitment, or old behaviors become attractive and their consequences forgotten. Maintenance counseling keeps vigilance alive with

questions such as, *What is working*? or *Most people have difficulty; how has this change gone for you*? Physicians can talk about relapse with questions like, *Many people begin to think that after a while it's safe to . . . [return to the old habit] just one more time; if you have had these thoughts, how have you handled them*? To engage patients in creative problem solving, physicians might ask, *Who or what has helped you sustain your new behavior*? Planning for recovery after a lapse can be initiated by asking, *Should you slip, what will you do, or who will you call*? A suggestion to call immediately tells patients that relapse is common and can be remediated; it should not be considered a sign of failure. By reframing relapse as a learning opportunity, physicians also help patients reflect on how to strengthen their action plan.

COUNSELING TEAMWORK AND REFERRAL

Behavior change counseling helps patients move from precontemplation to maintaining new habits. This counseling is not a single intervention but rather an ongoing process with long-term follow-up at frequent intervals. Longer counseling visits are appropriate in the first few months for the precontemplation and contemplation stages. Face-to-face or telephone follow-up visits every week or two are useful during the determination and action stages. For the maintenance stage, monthly brief follow-up counseling in the office combined with almost daily contact with supporting persons or role models works well. When intensive help is needed, physicians may refer the patient to education classes, behavioral counseling professionals, or self-help support groups to augment the medical treatment. Medical home primary care practices also can provide counseling for proactive prevention and chronic illness care using multidisciplinary teams of nurse educators, nutritionists, physical therapists, psychologists, and pharmacists. Changing of addictive behaviors usually requires referral to a behavioral change specialist.

Grade A References

A1. Armstrong MJ, Mottershead TA, Ronksley PE, et al. Motivational interviewing to improve weight loss in overweight and/or obese patients: a systematic review and meta-analysis of randomized controlled trials. *Obes Rev*. 2011;12:709-723.

A2. Chen SM, Creedy D, Lin HS, et al. Effects of motivational interviewing intervention on self-management, psychological and glycemic outcomes in type 2 diabetes: a randomized controlled trial. *Int J Nurs Stud*. 2012;49:637-644.

A3. Easthall C, Song F, Bhattacharya D. A meta-analysis of cognitive-based behaviour change techniques as interventions to improve medication adherence. *BMJ Open*. 2013;3:e002749.

A4. Foxcroft DR, Coombes L, Wood S, et al. Motivational interviewing for alcohol misuse in young adults. *Cochrane Database Syst Rev*. 2014;8:CD007025.

A5. Roy-Byrne P, Bumgardner K, Krupski A, et al. Brief intervention for problem drug use in safety-net primary care settings: a randomized clinical trial. *JAMA*. 2014;312:492-501.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Noordman J, van der Weijden T, van Dulmen S. Communication-related behavior change techniques used in face-to-face lifestyle interventions in primary care: a systematic review of the literature. *Patient Educ Couns*. 2012;89:227-244.
2. VanBuskirk KA, Wetherell JL. Motivational interviewing with primary care populations: a systematic review and meta-analysis. *J Behav Med*. 2014;37:768-780.
3. Lundahl B, Moleni T, Burke BL, et al. Motivational interviewing in medical care settings: a systematic review and meta-analysis of randomized controlled trials. *Patient Educ Couns*. 2013;93:157-168.
4. Romano M, Peters L. Understanding the process of motivational interviewing: a review of the relational and technical hypotheses. *Psychother Res*. 2014;10:1-21.

REVIEW QUESTIONS

1. Which of the following statements best describes the purpose of change counseling in the context of medical care?

A. To evoke patient self-motivation and to resolve ambivalence about changing
B. To advise patients of the evidence-based changes that are best for their health
C. To affirm and to support the autonomy of patients regardless of whether they change
D. To make patient-centered arguments to convince patients to change
E. To create a safe supportive environment that encourages patient change

Answer: A The principles of change counseling emphasize the importance of goal-directed counseling that ends in the patient's moving toward more healthy behaviors. The approach uses conversational counseling methods that evoke the patient's motivation and resolve natural ambivalence between the status quo and the effort needed to change. Therefore, A is the best answer. B is wrong because counseling is more than giving advice. C speaks to supporting the patient's autonomy, which is important, but not to making progress toward change. D is wrong because argument is specifically avoided in change counseling and is a sign of dysfunction in the patient-physician relationship. E is important for counseling, but it does not go far enough in defining the goals of change counseling.

2. Which of the following motivational interviewing skills are most useful in helping patients explore their ambivalence and talk themselves into changing?

A. Giving clear, simple explanations about medical conditions and the changes people can make to reduce the risks to their health
B. Using open questions and statements of affirmation and reflection to guess a patient's meaning and summarizing a patient's perspectives
C. Obtaining a complete assessment of risks, prior patient behaviors, and experiences with making other changes in their lives
D. Using the conviction-confidence ruler to assess the patient's reasons for changing and reasons for sustaining the status quo
E. Acknowledging nonverbal emotion caused by clinicians giving advice or telling patients what to do

Answer: B B is the correct answer because it states the four skills that characterize motivational and most patient-centered interviewing: open questions, affirmation, reflection, and summarization. Although a health risk assessment is an important starting point for change counseling, A is incorrect because it is limited to giving advice. C is incorrect because obtaining an assessment, although important, is only part of the task. D is incorrect because the purpose of the conviction-confidence ruler is to assess motivation, and D describes decisional ambivalence. Acknowledging emotions is an important engagement skill, but E is incorrect because the answer is limited to advice given by clinicians.

3. What is the purpose of a decisional balance tool in change counseling?

A. It shows the balance of risks of disease and benefits of treatment.
B. It documents the clinician's decisions for the changes patients need to make.
C. It tracks change in the patient's confidence during the course of counseling.
D. It demonstrates the patient's perceived importance of making a change.
E. It displays a patient's ambivalence between making change and sustaining the status quo.

Answer: E E is correct. The decisional balance is a counseling tool that displays a patient's ambivalence, showing the arguments for sustaining the status quo in one column and the arguments for change in an adjacent column. It is a potential starting point for change counseling in the precontemplation and contemplation stages. A is incorrect because the decisional balance is not about the clinician's decisions. B is not correct; although putting the decisional balance in the medical record documents its use and its results, decisional balance displays the patient's ambivalence, not the physician's decisions. C is wrong because the decisional balance displays ambivalence and does not track changes in commitment. D is partially correct because the decisional balance documents the patient's reasons for change, but it also documents the reasons for sustaining the status quo.

15

THE PERIODIC HEALTH EXAMINATION

DAVID ATKINS AND MARY BARTON

Primary and secondary prevention is an essential part of primary care and the patient-centered medical home (Chapter 13). The appropriate services and their frequency vary with the age, gender, and individual risk factors of each patient. A periodic health examination focusing on prevention increases the delivery of appropriate screening and lifestyle counseling. The most comprehensive prevention recommendations are produced by the U.S. Preventive Services Task Force (USPSTF), an ongoing panel of experts supported by the federal Agency for Healthcare Research and Quality (http://www.uspreventiveservicestaskforce.org).

USPSTF recommendations are used by major primary care subspecialty groups, many health plans, and quality organizations, and their A and B recommendations guide preventive services that are covered under the Affordable Care Act. The USPSTF bases its recommendations on two factors: an estimate of the *net benefits* (benefits minus harms) of a service, and an assessment of the *certainty* of that estimate, based on the rigor of supporting studies. Grade A recommendations require high certainty of a substantial net benefit, most often from large, prospective, controlled studies that measure morbidity or mortality. Grade B recommendations have high certainty of moderate net benefit or have moderate certainty for substantial benefit. USPSTF recommendations are thus more conservative than those of some subspecialty organizations that may give more weight to indirect evidence, such as earlier detection of disease, and less weight to potential harms of interventions. Clinicians can draw several general conclusions from an evidence-based approach to prevention: we should be selective in our use of screening tests, especially in older patients, and involve patients in decisions about specific services for which a small chance of benefit must be balanced against possible harm.

HISTORY AND RISK ASSESSMENT

The history and risk assessment are important tools to identify individuals who may need additional screening tests or immunizations not generally recommended for their age group or who may benefit from specific counseling to address unhealthy behaviors. Formal health risk appraisals should be linked to a system to provide specific feedback and targeted interventions.[1] Risk assessment should address the following:

- Use of tobacco, alcohol, and other drugs (especially injection drugs) (Chapters 32, 33, and 34)
- Diet (Chapters 213 and 220)
- Physical activity (Chapter 16)
- Sexual behavior that may increase the risk for sexually transmitted diseases or unintended pregnancy (Chapters 285, 384, and 387)
- Family history of cancer and heart disease
- Birthplace and current residence (community risk for infectious diseases)
- Presence of chronic diseases, such as diabetes, and of other cardiovascular risk factors

SCREENING FOR EARLY DISEASE OR ASYMPTOMATIC RISK FACTORS

Every year, new screening tests are introduced and marketed on the basis of their ability to detect unrecognized diseases or risk factors for disease. Other conditions need to be fulfilled for a screening test to be worthwhile for routine use, however (Table 15-1). Benefits of screening must be balanced against the potential harms, including false-positive results and the risks and costs of follow-up procedures or treatments. Even when tests have high specificity, the majority of positive results will be false-positive results if the test is used to screen healthy populations for uncommon conditions such as cancer (Chapter 10). There is growing evidence that cancer screening can lead to "overdiagnosis"—detection of slow-growing cancers that would never have caused clinical symptoms in a patient's lifetime.[2] Overdiagnosis subjects patients to the side effects of treatment with no benefits. A relatively small number of screening tests have been proved beneficial for the general population (Table 15-2), but additional tests are indicated for specific populations at risk (Table 15-3). Many commonly used tests are not recommended by the USPSTF (Table 15-4) because they provide little benefit or lack sufficient evidence to prove or disprove their value.

Depression

Depression is common and frequently undetected in primary care (Chapter 397). Simple screening instruments, including the following two questions—During the past 2 weeks, have you felt down, depressed, or hopeless? During the past 2 weeks, have you felt little interest or pleasure in doing things?—increase the detection of major depression. To improve outcomes, however, screening for depression must be linked to an organized system with staff support to ensure follow-up and treatment.[3]

High Blood Pressure, Abnormal Lipids, and Other Coronary Risk Factors

Overall cardiovascular risk, based on age, gender, blood pressure, lipid levels, and diabetes status, should be calculated at regular intervals to guide treatment decisions to reduce cardiovascular risk.[4] Blood pressure should be measured at least every 2 years (Chapter 67). The USPSTF recommends measuring total and high-density lipoprotein cholesterol, which can be done on fasting or nonfasting samples, beginning in middle age or earlier in the presence of other cardiovascular risk factors. Recommendations from the American Heart Association and American College of Cardiology favor fasting lipoprotein analysis, which will identify patients with severe LDL-C

TABLE 15-1 REQUIREMENTS OF AN EFFECTIVE SCREENING TEST

The disease being screened for is an important cause of morbidity and mortality.
Screening can detect disease in an early, presymptomatic phase.
Screening and treatment of patients with early disease or risk factors produce better health outcomes than does treatment of patients when they present with symptoms.
Screening test is acceptable to patients and clinicians—safe, convenient, acceptable false-positive rate, acceptable costs.
Benefits of early detection and treatment are sufficient to justify potential harms and costs of screening.

TABLE 15-2 PREVENTION RECOMMENDATIONS FOR THE GENERAL POPULATION: FROM THE U.S. PREVENTIVE SERVICES TASK FORCE AND OTHER SOURCES

SCREENING

Height, weight, and body mass index (BMI) calculation: periodically
Blood pressure: at least every 2 yr
Screen for alcohol misuse
Brief screen for depression*
Total blood cholesterol and HDL cholesterol: every 5 yr for men and women ≥20 yr with CHD risk factors and men ≥35 yr without CHD risk factors
Colorectal cancer screening: age ≥50 yr (see text for options)
Mammogram: at least every 2 yr for women ≥50 yr; discuss with women 40–49 yr
Papanicolaou (Pap) test: at least every 3 yr for sexually active women aged 21–29 yr; every 5 yr when combined with HPV testing in women 30-65 yr; in absence of HPV testing, continue every 3 yr until age 65
HIV test: at least once in all adults; repeat based on high-risk sexual activity
Chlamydia and gonorrhea: sexually active women ≤24 yr and older women at risk
Bone mineral density test: women ≥65 yr and high-risk women <65 yr

COUNSELING

Substance use
  Tobacco cessation
  Reduction of risky or harmful alcohol use
Diet and exercise
  Limit saturated fat; maintain calorie balance; emphasize grains, fruits, and vegetables‡
  Regular physical activity†
Sexual behavior
  Unintended pregnancy: contraception
  STD prevention: avoid high-risk behavior, use condoms or female barrier with spermicide†
Injury prevention
  Lap and shoulder belts
  Motorcycle, bicycle, and ATV helmets†
  Smoke detector†
Dental health
  Regular visits to dental care provider†
  Floss, brush with fluoride toothpaste daily†

IMMUNIZATIONS

Pneumococcal vaccine (once, age ≥65 yr)
Influenza vaccine (annual)
Tetanus-diphtheria (Td) boosters (every 10 yr)
Measles, mumps, rubella (MMR) vaccine (susceptible adults aged 19–49 yr)§
Varicella (two doses, susceptible adults aged 30–49 yr)§
Human papillomavirus (HPV) vaccine (three doses, women ≤26 yr)

CHEMOPREVENTION

Multivitamin with folic acid (women planning or capable of pregnancy)
Discuss benefits and harms of aspirin to prevent cardiovascular disease in middle-aged adults and others at increased risk for vascular disease

*Depression screening is most effective where systems exist to improve its management.
†The ability of clinician counseling to influence this behavior is uncertain.
‡Diet counseling is most effective when it targets at-risk groups (overweight or with CHD risk factors).
§Immunity can be verified by serologic testing, documented history of illness, or vaccination.
Other sources for Tables 15-2 and 15-3 include U.S. Department of Health and Human Services (diet, physical activity) and Centers for Disease Control and Prevention (injury prevention, immunizations, PPD).
ATV = all-terrain vehicle; CHD = coronary heart disease; HDL = high-density lipoprotein; STD = sexually transmitted disease.

TABLE 15-3 RECOMMENDED SCREENING AND INTERVENTIONS FOR HIGH-RISK POPULATIONS

POTENTIAL INTERVENTION	POPULATION
Ultrasound examination for abdominal aortic aneurysm	Current or former male smokers aged 65-75 yr
Low-dose computed tomography for lung cancer	Smokers with at least 30 pack-year smoking history (current smokers and those who quit smoking within past 15 years)
Hepatitis C test	Born 1945 to 1965; history of injection drug use or other high risk behaviors
Syphilis (RPR/VDRL)	High-risk sexual behavior; consider local epidemiology*
Gonorrhea screen	High-risk sexual behavior; consider local epidemiology*
PPD	Specific immigrant groups, prisoners, HIV patients
Hepatitis B vaccine	Exposure to blood products; IV drug use; high-risk sexual behavior; travelers to high-risk areas
Hepatitis A vaccine	Persons living in or traveling to high-risk areas; institutionalized persons and workers in these institutions; those with certain chronic medical conditions
Meningococcal vaccine	First-year college students in dormitories; military recruits; those with asplenia; travelers to high-risk areas
Varicella vaccine	Adults born after 1980 without evidence of immunity
Breast cancer chemoprevention	Women at increased risk for breast cancer and with low risk for thromboembolic complications
Diabetes screen	Persons at increased risk of diabetes

*Routine screening may be indicated in communities or settings where infection is prevalent.
HIV = human immunodeficiency virus; IV = intravenous; PPD = purified protein derivative; RPR = rapid plasma reagin; VDRL = Venereal Disease Research Laboratory.

TABLE 15-4 INTERVENTIONS NOT RECOMMENDED FOR ROUTINE USE IN ASYMPTOMATIC AVERAGE-RISK ADULTS*

Resting or exercise electrocardiography (D) or helical computed tomography (CT) for asymptomatic coronary disease (I)
Ultrasound examination for asymptomatic carotid artery stenosis (D)
Chest radiograph for lung cancer
Oral examination for oral cancer (I)
Routine blood tests for anemia
Routine urine tests (to screen for infection, cancer, or chronic kidney disease) (I)
Blood tests or ultrasound examination for ovarian cancer (D)
Whole body CT
Brief tests of mental status to detect dementia (I)
Vitamin supplements (I, D)
Blood level of C-reactive protein to predict coronary risk (I)
Prostate-specific antigen for prostate cancer (D)

*These services have either insufficient evidence to support routine use (I) or at least fair evidence that they provide no benefit or that harms outweigh benefits (D).

elevations (>190 mg/dL) (Chapter 206). Either strategy is sufficient to calculate a 10-year cardiovascular risk and to identify patients who may benefit from statin therapy (Chapter 52), which should be considered whenever the 10-year risk exceeds 7.5%. Although factors such as C-reactive protein, homocysteine, and coronary calcification as assessed by computed tomography (CT) are associated with an increased risk for heart disease, the USPSTF found insufficient evidence to recommend their routine use.

Abdominal Aortic Aneurysm

Between 5 and 9% of men older than 65 years have an abdominal aortic aneurysm (Chapter 78). The risk for aneurysm is highest in smokers and is substantially lower in women (1%). The USPSTF recommends one-time screening with ultrasound examination in men aged 65 to 75 years who are

current or former smokers,[5] based on trials demonstrating as much as a 40% lower death rate from abdominal aortic aneurysm rupture in screened men.

Colorectal Cancer

Screening can reduce both the incidence of and mortality from colorectal cancer. Options for screening men and women older than 50 years include an annual highly sensitive fecal occult blood test (FOBT) or fecal immunochemical test, flexible sigmoidoscopy every 5 years plus FOBT every 3 years, or colonoscopy every 10 years (Chapter 193). Colonoscopy combines detection with the opportunity for biopsy and removal of lesions, so it is preferred in some guidelines. It carries higher costs and risks, however, and no single strategy has proved to be more effective or cost-effective than the alternatives. In a randomized trial, a single flexible sigmoidoscopy screening between 55 and 64 years of age reduced the colorectal cancer incidence by 33% and the colorectal cancer mortality by 43% after 11 years of follow-up in people who were screened.[6] In 2008, the USPSTF concluded that evidence was not yet sufficient to support newer technologies such as CT colonography or fecal tests for DNA markers of neoplasia. DNA tests were approved by FDA in 2014 and are more sensitive but less specific than FOBT.[7] The USPSTF recommends stopping routine screening at age 75 years.

Breast Cancer

In large trials, mammography screening (at intervals of 1 to 2 years, with or without clinical breast examination) reduces breast cancer mortality by 15 to 30% (Chapter 198). Most trials suggest that the benefits of screening extend to women in their 40s, but the benefits are smaller and the risks for false-positive results are higher than in women aged 50 to 70 years.[8] If a woman aged 40 to 49 years has a two-fold increased risk for breast cancer, she will have a similar benefit-to-harm ratio for biennial screening mammography as an average-risk woman aged 50 to 74 years.[9] No studies provide data on the benefits of screening women aged 75 and older. In a collaborative study, six independent models predicted that an average of 80% of the benefit of mammography could be achieved with biennial rather than annual mammography, whereas national mammography surveillance data indicate that false-positive results and other harms of screening would be reduced by about half if women were screened every other year instead of yearly. Although many cancers are discovered by patients, teaching women to perform breast self-examination increases the likelihood that a woman will undergo further evaluation for an unimportant finding, but it does not improve outcomes. Widespread screening for *BRCA1* or *BRCA2,* inherited mutations that increase the risk for breast cancer, is not recommended, but the USPSTF recommends that primary care providers screen women who have family members with breast, ovarian, tubal, or peritoneal cancer using one of several recommended family history screening tools designed to identify women at risk for potentially harmful mutations. Women with a positive family history should be referred for genetic counseling.[10]

Cervical Cancer

Papanicolaou (Pap) screening is highly effective in preventing invasive cervical cancer, but many low-risk women in the United States are screened more often than needed. The USPSTF and a multi-specialty-society collaborative[11] endorse delaying screening until age 21 and then screening women aged 21 to 30 years with previous normal test results only every 3 years instead of annually (Chapter 199). In women ages 30 to 65 years, both groups recommend adding testing for the human papillomavirus (Chapters 18 and 373) to Pap testing (co-testing) and extending the interval between screens to 5 years. Women with adequate screening histories should be given the option of discontinuing screening after 65 years of age. Screening is not indicated in women who have undergone a hysterectomy for benign disease. Whether women who have received the human papillomavirus vaccine (Chapter 18) need less frequent screening is not yet known.

Prostate Cancer

Screening with prostate-specific antigen (PSA) can increase the detection of organ-confined prostate cancer, but two large trials provided conflicting results as to whether screening lowers morbidity or mortality from prostate cancer (Chapter 201). An American trial found no benefit of annual PSA testing, but it was hampered by a high rate of screening in the control group. A European study reported that men randomized to screening every 4 years had a 20% lower risk of dying from prostate cancer after 9 years; for each prostate cancer death prevented, 48 additional men were treated for prostate cancer, and 1068 men had to undergo screening. Because of the small benefit and significant morbidity associated with overdiagnosis and overtreatment, including incontinence and impotence, the USPSTF recommends against routine PSA screening,[12] but the American College of Physicians and some specialty groups recommend discussing PSA screening with men who have a life expectancy of at least 15 years.

Lung Cancer

The USPSTF now recommends annual screening with low-dose CT in adults aged 55 to 80 years who have at least a 30 pack-year smoking history and currently smoke or have quit within the last 15 years, based on a large trial in which such screening reduced deaths from lung cancer by 20% and all-cause mortality by 7%.[13] More than 10% of screened persons will have another clinically significant finding, such as a mass in the kidney or adrenal gland or an aortic aneurysm, and at least one false-positive test will occur in more than 40% of screened adults.

Osteoporosis

Tests of bone mineral density can identify men and women with a high risk for fracture due to osteoporosis and who may benefit from medications proved to lower the risk for fracture (Chapter 243). The USPSTF recommends screening for osteoporosis in women older than 65 years as well as in younger women who have risk factors that put them at comparable risk.[14] The most accurate predictor of risk for hip fracture is bone mineral density of the hip assessed with dual-energy x-ray absorptiometry. A tool developed by the World Health Organization (http://www.shef.ac.uk/FRAX/) incorporates bone mineral density and other risk factors, such as age and fracture history, to guide treatment decisions.

Thyroid Disease

Routine thyroid testing occasionally identifies patients with symptomatic but undiagnosed hypothyroidism (Chapter 226), but it more often detects subclinical hypothyroidism, a disorder marked by elevations in thyroid-stimulating hormone with normal levels of free thyroxine. Because the benefits of treating subclinical hypothyroidism remain uncertain, the USPSTF does not recommend routine thyroid testing in the absence of symptoms. Clinicians should be alert to subtle signs of thyroid disease and have a low threshold for testing patients in high-risk groups, including postpartum and postmenopausal women.

Diabetes

Routine screening for diabetes beginning at age 45 years is recommended by some groups, but the USPSTF recommends screening for abnormal blood glucose and type 2 diabetes only in patients at increased risk based on age, overweight, or family history. Although tight glucose control can reduce the incidence of microvascular disease, the benefit of early presymptomatic detection on clinically important retinopathy, neuropathy, and nephropathy is likely to be small. The benefits of early detection of diabetes are based primarily on the benefits of intensive lifestyle modifications.

HIV Infection

The USPSTF and Centers for Disease Control and Prevention (CDC) recommend screening for HIV infection (Chapters 387, 388, and 389) in all adolescents and adults aged 15 to 65 years, including pregnant women. Whether to continue screening after age 65 years should consider risk factors such as new sexual partners. The optimal screening interval is unclear, but annual screening is reasonable in persons at very high risk, such as individuals who are actively engaged in risky sexual behaviors. Rescreening is not necessary in seronegative patients who have not been at increased risk since the last screen.

Hepatitis C Infection

Drugs for chronic hepatitis B and C infection (Chapter 149) have increased the ability to achieve viral suppression and prevent the complications of chronic liver disease in infected individuals. As a result, the USPSTF and CDC now recommend screening for hepatitis B and C in persons at high risk for infection, especially individuals with a history of past or current injection drug use.[15,16] Because persons born between 1945 and 1965 account for three fourths of Americans living with hepatitis C, one-time screening in this cohort is also recommended.

Sexually Transmitted Disease

Screening for chlamydia (Chapter 318) is recommended for all sexually active women aged 24 years and younger and for older women at risk. Nucleic acid amplification tests can be performed on cervical or urine

specimens. Early detection can reduce pelvic inflammatory disease (Chapter 285), a risk factor for infertility and ectopic pregnancy. Similar benefits are likely from screening women for gonorrhea (Chapter 299), but the risk for gonorrhea infection is more concentrated in high-risk urban and southeastern rural populations.

Vision and Hearing

Undetected but correctable vision (Chapter 423) and hearing (Chapter 428) problems are common in older adults and can be discovered by asking about problems and performing simple tests of visual acuity and hearing. Unfortunately, evidence is limited to show that regular screening leads to measureable benefits in function. Regular visual acuity testing is recommended for older adults by many organizations, but a large trial did not find any lasting benefits of screening, despite detecting many correctable causes of vision problems. The USPSTF concluded that evidence was insufficient to recommend routine vision or hearing screening.

BEHAVIORAL INTERVENTIONS

Lifestyle factors contribute to a large proportion of preventable deaths in the United States. Brief interventions are effective for some behaviors such as smoking and problem drinking, but changing other behaviors usually requires more intensive approaches. The *5 As* framework—ask, assess, advise, assist, and arrange—which was developed from smoking cessation research, provides a useful framework for counseling (Chapter 14).

Tobacco Use

Brief interventions can produce small but clinically important increases in quit rates among smokers. Effects increase with more intensive counseling and support, including the use of medication (Chapter 33).

Alcohol Misuse

The USPSTF recommends screening all adults 18 years or older for alcohol misuse with one of three tools: the 10-question AUDIT instrument, its 3-question version AUDIT-C (http://www.integration.samhsa.gov/images/res/tool_auditc.pdf), or a single question: "How many times in the past year have you had five (for men) or four (for women and adults over 65 years old) or more drinks in a day?"[17] Brief multi-contact behavioral interventions can successfully reduce alcohol consumption in at-risk drinkers (Chapter 33).

Diet

Diet counseling can reduce the intake of saturated fat and increase the consumption of fruits and vegetables. Effects are most consistent with more intensive counseling (multiple sessions with trained counselors) and in higher risk patients, such as those with elevated lipid levels (Chapter 213).

Physical Activity

Moderate physical activity reduces the risk for obesity, diabetes, and coronary heart disease, among other benefits. Studies of counseling in the primary care setting, however, have reported inconsistent effects on long-term levels of physical activity (Chapter 16).

Injury Prevention

Motor vehicle injuries are the leading cause of years of potential life lost before age 65 years. In older persons, falls are a leading cause of unintentional injury and can be reduced with targeted interventions (Chapter 25). The USPSTF recommends targeted interventions for at-risk persons, such as exercise, physical therapy, and vitamin D supplementation (Chapter 25).[18] The USPSTF recommends that women of childbearing age be screened for intimate partner violence (Chapter 241)[19] but found insufficient evidence to recommend screening all elderly or other vulnerable adults for abuse and neglect.

IMMUNIZATIONS

Recommendations regarding immunization (Chapter 18) are regularly updated by the Advisory Committee on Immunization Practices of the Centers for Disease Control and Prevention (http://www.cdc.gov/vaccines).[20] Annual influenza immunization is recommended for all adults. Vaccination with pneumococcal polysaccharide (PPSV23) is recommended at least once at age 65 years or after for all adults and for younger adults with asplenia, chronic heart or lung disease, and other immune disorders. Revaccination is not generally recommended unless the initial immunization was before age 65 years or the patient has immune-related risk factors. Adults with immunocompromising conditions, including chronic renal failure, should also receive pneumococcal conjugate 13-valent vaccination (PCV13). Two doses of varicella vaccine are recommended for all adults born in the United States after 1980 without other evidence of immunity, and one dose of zoster vaccine is recommended for all adults at age 60 years. Adults should be revaccinated once with the Tdap vaccine (tetanus, diphtheria, acellular pertussis) and every 10 years with Td; if their vaccination history is uncertain, complete primary immunization with two additional doses of Td is recommended. A series of three doses of a vaccine against human papilloma virus (HPV) is recommended for young women up to age 26 years (with HPV4 or HPV2) to reduce the risk for cervical cancer and for men up to age 21 years (with HPV4) to reduce the risk for genital warts and the transmission of the virus.

CHEMOPREVENTION AND SUPPLEMENTS

Aspirin, postmenopausal hormone replacement therapy, breast cancer chemopreventive drugs, and supplements of vitamins or minerals can carry both benefits and risks. Decisions need to consider the likely benefits (which increase with the underlying risk of the disease being prevented), the probability of harm, and the individual preferences of each patient.

Aspirin

In men and women without known vascular disease, aspirin reduces the combined risk for myocardial infarction, stroke, and other serious cardiovascular events by 12%, but it also increases the risk for serious gastrointestinal bleeding and hemorrhagic stroke and does not significantly reduce cardiovascular mortality. For men ages 45 to 79 years and women ages 55 to 79 years, clinicians should assess whether the benefits, which increase with higher risk for vascular disease, outweigh the bleeding risks, which increase with age (Chapter 38).

Chemoprevention of Breast Cancer

Tamoxifen and raloxifene can reduce the incidence of invasive breast cancer by nearly 50% in women at increased risk, but both agents increase the risk for thromboembolic events and worsen menopausal symptoms; tamoxifen also increases the risk for endometrial cancer (Chapter 198). The USPSTF recommends clinicians discuss the balance of benefits and harms of these medications with women over age 35 years who are at increased risk for breast cancer. Women most likely to benefit are those who are younger than age 60 years and whose 5-year risk of invasive breast cancer is 3% or higher.[21] Online tools can be used to assess this risk (http://www.cancer.gov/bcrisktool/).

Postmenopausal Hormone Therapy

In long-term follow-up studies, estrogen-only therapy reduced the risk for fracture and invasive breast cancer but increased the risk for stroke, thromboembolism, gallbladder disease, and incontinence (Chapter 240). Combination therapy with estrogen and progestin generally has comparable benefits and risks, but it increases the risk for invasive breast cancer and also increases the risk for dementia.[22] Hormone therapy is a reasonable option for younger menopausal women with persistent, troublesome menopausal symptoms, but the USPSTF recommends against its routine use for preventive purposes.

Vitamin and Mineral Supplementation

The USPSTF does not recommend routine use of vitamin or mineral supplements because there is no convincing evidence that multivitamins or individual or paired supplements (e.g., vitamin A, C, D, or E; folic acid; beta-carotene; selenium; or calcium) reduce cardiovascular disease, cancer, or all-cause mortality in community-dwelling average-risk adults.[23] The USPSTF found insufficient evidence to support routine use of calcium and vitamin D supplements for the primary prevention of fractures in ambulatory adults, but primary prevention studies most commonly targeted healthy postmenopausal women. Vitamin D supplementation is recommended for individuals older than 65 years who are at risk for falls (see Injury Prevention), although the mechanism of action is uncertain.

FUTURE ISSUES

There is growing awareness that many older patients receive cancer screening well beyond the ages at which they are likely to benefit. Making clear recommendations for older patients is complicated by lack of good evidence, the difficulty assessing life expectancy, and differences in patients' preferences.[24] As the understanding of genetic factors that modify the risk for disease grows, clinicians may eventually be able to target screening, preventive treatments,

or lifestyle interventions to those at greatest risk. The value of using currently available genomic tests to screen average-risk individuals is limited by an incomplete understanding of the predictive value of specific genotypes in the general population, the uncertain effect of such information on clinical decisions, and concerns about the possible adverse effects of screening (e.g., anxiety or "labeling," false reassurance, discrimination).

Grade A References

A1. James PA, Oparil S, Carter BL, et al. 2014 Evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014;311:507-520.

A2. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129:S1-S45.

A3. Whitlock EP, Lin JS, Liles E, et al. Screening for colorectal cancer: a targeted, updated systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2008;149:638-658.

A4. Moyer VA. Screening for cervical cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2012;156:880-891, W312.

A5. Moyer VA. Screening for HIV: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2013;159:51-60.

A6. LeFevre ML. Screening for chlamydia and gonorrhea: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;161:902-910.

A7. U.S. Preventive Services Task Force. Counseling and interventions to prevent tobacco use and tobacco-caused disease in adults and pregnant women: U.S. Preventive Services Task Force reaffirmation recommendation statement. *Ann Intern Med*. 2009;150:551-555.

A8. U.S. Preventive Services Task Force. Aspirin for the prevention of cardiovascular disease: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2009;150:396-404.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Agency for Healthcare Research and Quality. Health Risk Appraisal: Technology Assessment Report. Rockville, MD: AHRQ; 2011. http://www.cms.gov/Medicare/Coverage/DeterminationProcess/downloads/id79ta.pdf; Accessed February 13, 2015.
2. Esserman LJ, Thompson IM, Reid B. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. *JAMA*. 2013;310:797-798.
3. O'Connor EA, Whitlock EP, Beil TL, et al. Screening for depression in adult patients in primary care settings: a systematic evidence review. *Ann Intern Med*. 2009;151:793-803.
4. American Heart Association/American College of Cardiology. 2013 Prevention Guideline Tools: CV Risk Calculator. http://my.americanheart.org/professional/StatementsGuidelines/PreventionGuidelines/Prevention-Guidelines_UCM_457698_SubHomePage.jsp; Accessed January 29, 2015.
5. Guirguis-Blake JM, Beil TL, Senger CA, et al. Ultrasonography screening for abdominal aortic aneurysms: a systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014;160:321-329.
6. Holme Ø, Løberg M, Kalager M, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. *JAMA*. 2014;312:606-615.
7. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med*. 2014;370:1287-1297.
8. Nelson HD, Tyne K, Naik A, et al. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2009;151:727-737.
9. van Ravesteyn NT, Miglioretti DL, Stout NK, et al. Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: a comparative modeling study of risk. *Ann Intern Med*. 2012;156:609-617.
10. Moyer VA. Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer in women: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2014;160:271-281.
11. Saslow D, Solomon D, Lawson HW, et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *CA Cancer J Clin*. 2012;62:147-172.
12. Moyer VA. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2012;157:120-134.
13. The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Eng J Med*. 2011;365:395-409.
14. Screening for osteoporosis. U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2011;154:356-364.
15. LeFevre ML. Screening for hepatitis B virus infection in nonpregnant adolescents and adults: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;161:58-66.
16. Moyer VA. Screening for hepatitis C virus infection in adults: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2013;159:349-357.
17. Moyer VA. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2013;159:210-218.
18. Moyer VA. Prevention of falls in community-dwelling older adults: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2012;157:197-204.
19. Nelson HD, Bougatsos C, Blazina I. Screening women for intimate partner violence: a systematic review to update the U.S. Preventive Services Task Force recommendation. *Ann Intern Med*. 2012;156:796-808, W-279, W-280, W-281, W-282.
20. Centers for Disease Control and Prevention, Advisory Committee on Immunization Practices. Adult immunization schedules, United States, 2014. http://www.cdc.gov/vaccines/schedules/hcp/adult.html; Accessed January 29, 2015.
21. Moyer VA. Medications for risk reduction of primary breast cancer in women: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2013;159:698-708.
22. Nelson HD, Walker M, Zither B, et al. Menopausal hormone therapy for the primary prevention of chronic conditions: a systematic review to update the U.S. Preventive Services Task Force recommendations. *Ann Intern Med*. 2012;157:104-113.
23. Fortmann SP, Burda BU, Senger CA, et al. Vitamin and mineral supplements in the primary prevention of cardiovascular disease and cancer: An updated systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2013;159:824-834.
24. Leipzig RM, Whitlock EP, Wolff TA, et al. Reconsidering the approach to prevention recommendations for older adults. *Ann Intern Med*. 2010;153:809-814.

REVIEW QUESTIONS

1. A 40-year-old married woman comes for her periodic health examination. She has had negative annual Pap tests for the last 5 years. Which of the following options describes a recommended strategy for cervical cancer screening?

A. Obtain a human papillomavirus (HPV) test along with the Pap test; if both are normal, no screening will be needed for 5 years.
B. Because her last Pap was normal, defer Pap testing until next annual examination.
C. Continue annual screening until the age of 64 years.
D. Continue annual screening until age 50 years.
E. Begin annual HPV testing but lengthen the interval for Pap testing to every 5 years.

Answer: A In the absence of HPV testing, every-3-year Pap testing is recommended by the U.S. Preventive Services Task Force (USPSTF) and the American College of Obstetrics and Gynecology (ACOG) for women with previously normal Pap tests. In women who are HPV negative, the interval can safely be extended to every 5 years. Co-testing (combined Pap and HPV testing) at 5-year intervals is recommended by the USPSTF and ACOG beginning at the age of 30 years and extending to age 65 years.

2. A 52-year-old woman is new to your practice. She is a lifelong nonsmoker, and she had her first mammogram in the past year, with normal results. She has never undergone colorectal cancer screening. Which of the following is true regarding recommendations for screening and options available to her?

A. As an average-risk adult, she could be offered screening with any of the following options, which differ in their risks, intervals, and burdens of testing: high-sensitivity fecal occult blood testing (FOBT), flexible sigmoidoscopy with FOBT, or colonoscopy.
B. Regular colorectal cancer screening in adults should stop at age 65 years.
C. Computed tomography (CT) colonography at 10-year intervals is an acceptable option in place of colonoscopy.
D. Colonoscopy every 5 years is the preferred option because of its higher sensitivity.
E. Flexible sigmoidoscopy at 5-year intervals should be combined with annual FOBT.

Answer: A The U.S. Preventive Services Task Force (USPSTF) recommendations are based on a review of original studies and a commissioned comparative model to determine optimal age for starting and stopping screening for colorectal cancer and the optimal intervals. USPSTF recommendations suggest beginning screening at age 50 years and stopping regular screening at age 75 years. Their analyses indicated that either (1) annual high-sensitivity fecal occult blood testing (FOBT), (2) flexible sigmoidoscopy every 5 years combined with FOBT every 3 years, or (3) colonoscopy every 10 years would yield similar benefits over a population. CT colonography every 5 years and double-contrast barium enema every 5 years are included options in the 2008 guidelines of the Multisociety Task Force because of their ability to detect both cancer and polyps.

3. A 62-year-old African American man comes in for his first visit. Other than a 5-year history of high blood pressure, he is in good health. In response to questions about whether he should get a prostate-specific antigen (PSA) test, which of the following statements reflects information that should be part of a shared decision-making discussion?

A. As an African American, he has a higher risk for prostate cancer but may not benefit more from screening compared with men of other ethnic backgrounds.
B. PSA screening has been shown to reduce mortality from prostate cancer in some studies.
C. False-positive PSA tests are common and may require a biopsy.
D. PSA testing may result in detection and treatment of a cancer that would not have caused clinical problems (overdiagnosis).
E. All of the above

Answer: E Although African American men do have an increased risk for prostate cancer, there is no evidence that benefits of screening are any higher in this group. PSA testing was associated with lower prostate cancer mortality in one European study, in which 48 men needed to be treated for every 1 death prevented after 9 years. By comparison, mortality was not significantly lower in a large American trial. False-positive PSA results and potential overdiagnosis are the major concerns regarding prostate cancer screening. As a result, none of the major organizations recommends routine screening with PSA, although some recommend a shared decision-making approach (http://www.auanet.org/education/guidelines/prostate-cancer-detection.cfm).

4. A 58-year-old male accountant comes in for a periodic examination. He is recently divorced and has a steady female partner with whom he is sexually active. Because his partner is postmenopausal, they do not use condoms. He has never been screened for any sexually transmitted diseases and has no symptoms nor history of any drug use or transfusions. Which of the following tests are indicated as a screening test?

A. HIV
B. Hepatitis C
C. Chlamydia
D. Gonorrhea
E. Hepatitis B

Answer: A and B HIV testing is indicated in all adults who have never been screened. Although the subject has no specific risk factors for hepatitis C, testing is recommended for adults in his age cohort (born between 1945 and 1965). Chlamydia and gonorrhea screening might be indicated for a woman based on having a new sexual partner but is not recommended for men. Hepatitis B screening is recommended only for high-risk groups in nonpregnant adults.

5. A 40-year-old nonsmoking man is seen for his periodic examination. He has a healthy weight (body mass index = 20), blood pressure of 118/78 mm Hg, and no history of diabetes, and he is able to exercise regularly without symptoms. Last year, his nonfasting total cholesterol was 232 with an HDL of 50. Which of the following represents appropriate steps to address cardiovascular prevention in this patient?

A. Perform a baseline electrocardiogram to look for evidence of ischemia or ventricular hypertrophy.
B. Obtain a fasting lipoprotein analysis.
C. Recommend he begin a daily low-dose aspirin regimen.
D. Calculate his 10-year risk for cardiovascular disease to assess need to consider further interventions.
E. Obtain a coronary computed tomography (CT) to look for evidence of calcifications

Answer: D Ten-year cardiovascular risk can be calculated using age, gender, lipids, blood pressure, and history of diabetes and smoking. This patient's 10-year risk is only 1.3%, putting him in a low-risk category; any benefits of aspirin will be offset by risks for gastrointestinal bleeding. Fasting lipoprotein analysis, although it will allow you to assess triglyceride and LDL cholesterol, is not needed because risk calculators require only total cholesterol and HDL. Neither ECG nor coronary CT is recommended by the U.S. Preventive Services Task Force as screening tests for coronary disease. Because findings on coronary CT are associated with risk for coronary events, some organizations suggest this test can be considered in patients who are near a threshold for considering statin therapy (7.5% 10-year risk for cardiovascular disease).

16

PHYSICAL ACTIVITY

DAVID M. BUCHNER

DEFINITIONS

Physical activity can be broadly defined as body movement that is produced by skeletal muscles and expends energy. *Health-enhancing physical activity* is activity that, when added to light-intensity activities of daily life, produces health benefits and involves the large muscle groups of the body and substantial energy expenditure. Herein, *physical activity* refers to health-enhancing physical activity. *Exercise* refers to the subset of physical activity that involves a structured program to improve physical fitness.

Regular physical activity improves *health-related physical fitness*—the physiologic components of fitness that influence risk for disease, functional limitations, disability, and premature mortality. These components include cardiorespiratory endurance (aerobic capacity); skeletal muscle strength, power, and endurance; body composition and bone strength; and balance, flexibility, and reaction time.

The primary attributes of physical activity are *type* (mode), *frequency, duration,* and *intensity.* Types of physical activity (e.g., walking, swimming, lifting, stretching) are grouped according to their main physiologic effects into well-known categories: *aerobic* (or "cardio"), *muscle strengthening, flexibility,* and *balance.* Intensity is the level of effort during activity. For aerobic activity, the *absolute intensity* is measured in metabolic equivalents (METs), with 1 MET being the resting metabolic rate—an oxygen consumption of roughly 3.5 mL/kg/minute. *Absolute* intensity is commonly classified into: sedentary behavior (1.0 to 1.5 MET), light intensity (1.6 to 2.9 MET), moderate intensity (3.0 to 5.9 MET), and vigorous intensity (6.0 MET and higher). The *relative intensity,* which is the percentage of oxygen uptake (aerobic capacity) reserve required to perform an activity, ranges from 0 to 100%. In practice, the heart rate is used to monitor relative intensity because of the generally linear relationship between heart rate and percentage of oxygen uptake.

The *volume* (or amount) of activity is the product of frequency, duration, and intensity. Volume can be measured using self-report (questionnaires) as MET-minutes per week (a sum of the MET intensity of all activities multiplied by the minutes each activity is performed). Volume can also be assessed using objective measures of physical activity in either arbitrary units ("counts per day" from an accelerometer) or as activity-related energy expenditure (kcal/week). A commonly used objective measure is the accelerometer (usually worn at hip or wrist, for 1 week), which detects movement of the body and provides detailed information on the frequency, duration, and intensity of movement.

EPIDEMIOLOGY

Levels of Physical Activity

It is highly likely that Americans' total level of physical activity has declined since the 1950s. Occupational activity has probably declined the most, but activity around the home and as transportation for getting places has also declined. Trends in recreational physical activity are either stable or slightly improving.

Based on self-reported activity, about 60% of adults meet public health guidelines for physical activity,[1] but only about 10% to at most 45% of American adults meet these guidelines based on accelerometer data collected at the same time. Although one accelerometer reading is not a perfect measure of average physical activity levels, the magnitude of this discrepancy suggests that adults commonly over-report their activity levels and that physical inactivity is a larger public health problem than previously realized.

Physical activity levels decline with age, and men report more activity than women. Higher levels of income and education are associated with greater physical activity. Although white Americans self-report higher levels of physical activity than do other racial and ethnic groups, Mexican American adults are the most active group as judged by accelerometer data.[1]

Preventive Health Benefits in Adults

There is strong evidence that regular moderate or vigorous physical activity (Table 16-1) reduces the risk for premature mortality and coronary artery disease (Chapter 52), stroke (Chapter 406), high blood pressure (Chapter 67), adverse lipid profile (Chapter 206), type 2 diabetes mellitus (Chapter 229), metabolic syndrome, osteoporosis (Chapter 243), colon cancer (Chapter 193), breast cancer (Chapter 198), and obesity (Chapter 220).[2] Even 15 minutes per day or 90 minutes per week of moderately intensive exercise is associated with a 14% decrease in mortality.[3] Physical activity also reduces the risk for falls, cognitive impairment in older adults, age-related muscle loss, and depression (Chapters 25 and 397). There is moderate evidence that physical activity reduces the risk for hip fracture, lung cancer, endometrial cancer, and sleep disorders. Physical activity can delay age-related functional limitations and loss of independence. Some evidence also suggests that physical activity reduces the risk for anxiety disorders, osteoarthritis, and back pain.

The benefits of physical activity are independent of other risk factors. For example, a sedentary obese smoker achieves health benefits from exercise, even if smoking and obesity persist.

Epidemiologic studies report that low levels of activity are associated with an increased risk for adverse health outcomes,[4] including up to a 67% increase in overall mortality, a doubling of mortality from cardiovascular disease,[5] and an increased mortality from cancer. When a healthy diet, regular activity, and abstinence from smoking are considered simultaneously, the effect of lifestyle is even more dramatic.

The main determinant of the health benefits of physical activity is volume. Substantial health benefits begin to occur with a volume of 500 to 1000 MET-minutes/week. An adult can accumulate 500 MET-minutes by walking at 3.0 miles per hour (a 3.3-MET activity) on 3 days a week for 50 minutes (3.3 METs $\times$ 3 $\times$ 50 minutes $=\approx$500 MET-minutes). When measured by caloric expenditure, this volume of walking in a 75-kg (165-lb) adult expends an extra 430 kcal above the 190 kcal that would have been expended under resting conditions.

The dose-response relationship between volume and health benefits is curvilinear, such that the marginal benefit of activity at lower levels is large, and the benefit decreases with higher activity levels.[5] Data from the Women's Health Initiative Observational Study (Table 16-2) provide an example of the dose-response effect. Compared with the least active women, the risk for cardiovascular disease was 19% less in women who averaged 600 MET-minutes/week, yet just 28% less in women who averaged 1968 MET-minutes/week. Dose-response effects are also found in intervention studies. For example, a randomized trial of two doses of exercise (doses equivalent to jogging 12 and 20 miles/week) found that the lower volume of exercise significantly improved plasma lipoproteins, but the higher volume had greater beneficial effects.

TABLE 16-1 EXAMPLES OF MODERATE-INTENSITY AND VIGOROUS-INTENSITY ACTIVITIES

MODERATE INTENSITY
Walking briskly (3 miles per hour or faster, but not race-walking)
Water aerobics
Bicycling slower than 10 miles per hour
Tennis (doubles)
Ballroom dancing
General gardening
VIGOROUS INTENSITY
Race-walking, jogging, or running
Swimming laps
Tennis (singles)
Bicycling 10 miles per hour or faster
Jumping rope
Heavy gardening (continuous digging or hoeing, with heart rate increases)
Hiking uphill or with a heavy backpack

From U.S. Department of Health and Human Services. *2008 Physical Activity Guidelines for Americans.* http://www.health.gov/paguidelines.

TABLE 16-2 RELATIVE RISK FOR CARDIOVASCULAR DISEASE IN THE WOMEN'S HEALTH INITIATIVE OBSERVATIONAL STUDY ($N = 73,743$)

MEDIAN MET-MINUTES PER WEEK	MULTIVARIATE ADJUSTED RELATIVE RISK FOR CARDIOVASCULAR DISEASE
0	1.0
252	0.89
600	0.81
1050	0.78
1968	0.72

Data from Manson JE, Greenland P, LaCroix AZ, et al. Walking compared with vigorous exercise for the prevention of cardiovascular events in women. *N Engl J Med.* 2002;347:716-725.

TREATMENT

Therapeutic Health Benefits in Adults

Clinical practice guidelines assign a substantial therapeutic role to physical activity in patients with coronary heart disease (Chapter 52), high blood pressure (Chapter 67), type 2 diabetes (Chapter 229), obesity (Chapter 220), osteoporosis (Chapter 243), osteoarthritis (Chapter 262), claudication (Chapter 79), and chronic obstructive pulmonary disease (Chapter 88).

In individuals with impaired glucose tolerance and high cardiovascular risk, both baseline activity levels and changes in activity levels are associated with a reduction in subsequent cardiovascular events.[6] Physical activity also plays a role in the management of depression and anxiety disorders, elevated cholesterol levels, pain, heart failure, syncope, stroke, back pain, dementia, and constipation and in the prophylaxis of venous thromboembolism.

Although there is more evidence for the therapeutic effects of aerobic activity, muscle-strengthening exercise can also provide benefits. For example, a randomized trial of exercise in adults with diabetes reported that the combination of aerobic and muscle-strengthening exercise was superior to either by itself.

Physical activity effectively opposes age-related loss of fitness and functional limitations. In older adults, low fitness (exercise capacity of <85% of predicted value) is associated with doubled risk for mortality. Randomized controlled trials demonstrate that exercise by sedentary older adults improves health-related physical fitness (e.g., aerobic exercise improves aerobic capacity) and has beneficial effects on functional limitations, such as a slow gait speed. A systematic review of nine observational studies reported that slow gait speed is strongly related to mortality as well, with low-risk adults having gait speeds of 1.0 meter/second or higher.[7] Physical activity, especially balance training, prevents falls in older adults at increased risk for falling, such as those with impaired gait and balance. Because both low fitness and high body weight cause functional limitations, the question arises whether functional ability is best improved by weight loss, exercise, or both. With the doses of weight loss and exercise that are feasible in randomized trials, exercise alone appears more effective than weight loss alone, and combined exercise and weight loss is most effective.

Health Risks of Physical Activity

Physical activity and exercise have risks. Musculoskeletal injuries are by far the most common type of activity-related adverse event. The risk for injury depends on the type and volume of activity. Collision and contact sports have a much higher injury risk than noncontact activities such as walking, which is associated with about one musculoskeletal injury for every 1000 hours of walking for exercise.[8] The weekly volume of activity is directly related to the risk for musculoskeletal injuries, but when following public health guidelines, which involves participating in about 500 to 1000 MET-minutes/week, the risk for injury is low. The risk for injury is directly related to the rate of increase in the dose of activity, with more rapid increases having higher risk.[2] Previous musculoskeletal injuries and low fitness also increase the injury risk.

Overall, regular physical activity decreases the risk for sudden cardiac death and myocardial infarction; active adults have roughly a 70% lower risk for sudden death. However, relatively vigorous physical activity acutely increases the risk for these events in both active and inactive adults. About 5 to 10% of myocardial infarctions are associated with vigorous activity. Yet sudden death is a rare event during exercise, with published rates in the range of one death per year in 18,000 ostensibly healthy men and one death per 2.6 million workouts in fitness facilities.

Physical Activity Guidelines for Adults

In 2008, the U.S. Department of Health and Human Services issued the first national physical activity guidelines: the *2008 Physical Activity Guidelines for Americans*[2] (Table 16-3).

Recommended Amounts of Aerobic Activity for Prevention

To obtain substantial health benefits from physical activity, adults should do at least 150 minutes/week of moderate-intensity aerobic activity or 75 minutes/week of vigorous-intensity activity. Adults can do a combination of both moderate- and vigorous-intensity activity, using the rule of thumb that one vigorous-intensity minute of activity counts the same as two moderate-intensity minutes. Bouts of aerobic activity 10 or more minutes clearly have health benefits, and even short bouts of less than 10 minutes may also provide some health benefits.[9]

Recommended Amounts of Muscle-Strengthening Activity for Prevention

Adults should perform activities that strengthen the major muscle groups of the body at least 2 days each week. The major muscle groups are the legs, hips, back, chest, abdomen, shoulders, and arms.

Flexibility Activity

Flexibility activities are an acceptable part of a physical activity regimen, but there is insufficient evidence that flexibility activities have any health benefits, even for preventing injuries. Flexibility training does increase flexibility, and it may facilitate the types of physical activity that do have health benefits. If so, flexibility training is more important for people with reduced flexibility, such as older adults with age- and disease-related changes in range of motion.

Balance Activity

Balance training is currently recommended only for adults at increased risk for falls (Chapter 24), such as those older than 65 years with impaired gait or balance or frequent falls. Examples of balance exercises include sideways walking, backward walking, heel walking, and standing using a narrow base of support. Tai chi programs include exercises that improve balance, and tai chi exercise is an evidence-based approach to fall prevention. Preferably, adults at risk for falls should do balance training at least 3 days/week and follow an evidence-based program demonstrated to reduce the risk for falls.

Sedentary Behavior

Sedentary behavior can be defined as activities (not including sleeping) with energy expenditure of 1.0 to 1.5 METs; most of these behaviors involve sitting or lying down. Epidemiologic studies report that higher levels of sedentary time (or equivalently lower levels of light-intensity activity) have adverse health effects, independent of a person's volume of moderate- to vigorous-intensity activity. More sedentary time as assessed by accelerometers is associated with greater body weight, worse markers of metabolic health, and higher mortality. Currently, light-intensity activities expend calories and can be recommended as part of a plan to achieve and maintain a healthy weight. However, data are insufficient to issue guidelines on the amounts and patterns of sitting time that cause adverse health effects in adults.

Guidelines for Weight Management

The amount of activity required to maintain a healthy body weight varies widely among adults (Chapter 220). For some adults, the physical activity levels recommended in the guidelines will result in a stable, healthy body weight. Many adults, however, need higher levels of activity to achieve a healthy weight. These individuals should restrict their caloric intake and gradually increase their physical activity each week, to the point that is effective in achieving and maintaining a healthy body weight for them.

Additional Guidelines for Older Adults

In addition to balance training, the 2008 guidelines contain other recommendations specifically for older adults. Older adults who cannot do 150 minutes/week of moderate-intensity activity should be as active as their abilities and conditions allow. Progressive resistance strength training improves physical function in elderly people.[A] Older adults should determine their level of physical activity using relative intensity, not absolute, intensity. This latter guideline seeks to avoid inappropriately high levels of effort in older adults with low fitness.

Recommending Physical Activity in Clinical Settings

Promoting physical activity in clinical settings involves essentially the same steps as the "5 As" of smoking cessation: ask, advise, assess, assist, arrange. (1) Ask about the amount of physical activity a patient typically engages in each week by questionnaire or interview. (2) Advise all patients to participate in at least a moderate amount of physical activity each week. Advise patients who do not meet the recommendations to increase their physical activity gradually to a specified minimal level, and tailor the recommendations according to their medical conditions. (3) Assess the next step or steps a patient needs to take to become more active. (4) Assist the patient in taking these steps. (5) Arrange an appointment to follow up on efforts to increase activity.

Health care providers should ask and advise patients about physical activity regularly. One quality-of-care measure assesses whether asking and advising are done at least once a year in older adults. Some recommend that physical activity should be a "vital sign" that is assessed at every visit. Studies consistently report that adults (especially older adults) identify their health care providers as an important source of advice on physical activity.

Nevertheless, the U.S. Preventive Services Task Force concluded that the benefit of routine counseling in the primary care setting is small.[10] The benefit is higher, however, if counseling is provided only when patients request it.

Exercise Prescription

For patients who want to exercise, a clinician can provide an exercise prescription. The prescription includes (1) the type, frequency, duration, and intensity of aerobic exercise; (2) the exercise movements (e.g., bench press), repetitions, and sets for resistance exercise; (3) other exercises such as stretching, balance exercises, warm-up, and cool-down; and (4) risk management strategies, such as increasing levels of activity gradually over time.

Lifestyle Prescription

A lifestyle prescription refers to approaches that integrate physical activity into daily life. For example, rather than walking specifically for exercise, a person walks to work or walks for pleasure as a recreational activity. Common ways to integrate physical activity into daily life are walking and biking for transportation and performing yard work and gardening.

Tailoring the Recommendation

Recommendations need to be tailored to individual abilities, individual preferences, medical conditions, and behavior techniques that improve adherence. Randomized trials show that home-based programs are superior to center-based programs in terms of long-term adherence.[A] Although center-based programs are preferred by some, in most cases they serve a short-term purpose, such as assisting adults to initiate regular exercise.

A target level of physical activity below that of preventive recommendations is appropriate for adults with very low fitness, a large burden of chronic disease (e.g., severe chronic obstructive lung disease), or major functional limitations. An assessment of the nature of the activity limitation and the individual's capabilities and preferences can determine the target activity level and other details of the activity recommendation. Often, promoting physical activity in such adults relies on health care and community resources designed for people with preexisting limitations, such as cardiac rehabilitation, pulmonary rehabilitation, and exercise classes for adults with arthritis.

Risk Management

Strategies to reduce injury include increasing physical activity gradually over time, selecting activities in which collision or contact with people or objects is unusual, increasing physical fitness, using appropriate gear and sports equipment (e.g., bike helmets), engaging in activities in safe environments, and following basic safety rules and policies. Popular activities such as walking, biking, swimming, and gardening have a low risk for injury. When increasing the level of physical activity, the guidelines recommend using relative intensity to determine the level of effort, starting with relatively moderate-intensity activity, then increasing the duration and frequency first and the intensity last. In general, adding 5 to 15 minutes of moderate-intensity activity per session, two to three times a week, to a person's usual activities carries a low risk for musculoskeletal injury and no known risk for sudden cardiac death.[8]

TABLE 16-3 KEY PHYSICAL ACTIVITY GUIDELINES FOR ADULTS

All adults should avoid inactivity. Some physical activity is better than none, and adults who participate in any amount of physical activity gain some health benefits.
For substantial health benefits, adults should do at least 150 min (2.5 hr)/wk of moderate-intensity aerobic activity or 75 min (1.25 hr)/wk of vigorous-intensity aerobic activity, or an equivalent combination of moderate- and vigorous-intensity aerobic activity.
Aerobic activity should be performed in episodes lasting at least 10 min and should be spread throughout the week.
For additional and more extensive health benefits, adults should increase their aerobic physical activity to 300 min (5 hr)/wk of moderate-intensity or 150 min (2.5 hr)/wk of vigorous-intensity aerobic physical activity, or an equivalent combination of moderate- and vigorous-intensity activity. Additional health benefits are gained by engaging in physical activity beyond this amount.
Adults should also do muscle-strengthening activities that are moderate or high intensity and involve all major muscle groups on 2 or more days/wk because these activities provide additional health benefits.

From U.S. Department of Health and Human Services. *2008 Physical Activity Guidelines for Americans.* http://www.health.gov/paguidelines.

There is no evidence of the protective value of a medical consultation or of supervised exercise in healthy people of any age who seek to increase their level of physical activity. The U.S. Preventive Services Task Force recommended against routine screening for coronary disease in adults at low risk for it and concluded that in adults at high risk for coronary disease, there is insufficient evidence to recommend for or against screening with resting electrocardiography or an exercise treadmill test.[11]

Coordination between Medical Care and Community

Many factors affecting physical activity levels are difficult to influence in medical care settings, such as the characteristics of the communal environment (e.g., parks and recreational facilities) and the social environment (e.g., crime and social support). Community-level interventions that address such characteristics are essential to promoting physical activity. Effective community-based interventions include school physical education, social support interventions, community-wide campaigns, enhancement of access to places where physical activity is possible, and interventions that involve community design, such as improving the connectivity of streets and the walkability of neighborhoods.[12] Medical care and community efforts should be synergistic and mutually supportive. For example, health plans should be advocates for evidence-based community interventions. Community programs should serve as resources for evidence-based therapeutic activity for selected chronically ill adults, such as exercise classes designed to reduce the risk for falls.

Grade A References

A1. Pahor M, Guralnik JM, Ambrosius WT, et al. Effect of structured physical activity on prevention of major mobility disability in older adults: the LIFE study randomized clinical trial. *JAMA*. 2014;311:2387-2396.

A2. Kraus WE, Houmard JA, Duscha BD, et al. Effects of the amount and intensity of exercise on plasma lipoproteins. *N Engl J Med*. 2002;347:1483-1492.

A3. Church TS, Blair SN, Cocreham S, et al. Effects of aerobic and resistance training on hemoglobin A_{1c} levels in patients with type 2 diabetes: a randomized controlled trial. *JAMA*. 2010;304:2253-2262.

A4. Pahor M, Blair SN, Espeland M, et al. Effects of a physical activity intervention on measures of physical performance: results of the Lifestyle Interventions and Independence for Elders Pilot (LIFE-P) study. *J Gerontol A Biol Sci Med Sci*. 2006;61:1157-1165.

A5. Gillespie LD, Robertson MC, Gillespie WJ, et al. Interventions for preventing falls in older people living in the community. *Cochrane Database Syst Rev*. 2012;9:CD007146.

A6. Villareal DT, Chode S, Parimi N, et al. Weight loss, exercise, or both and physical function in obese older adults. *N Engl J Med*. 2011;364:1218-1229.

A7. Liu CJ, Latham NK. Progressive resistance strength training for improving physical function in older adults. *Cochrane Database Syst Rev*. 2009;3:CD002759.

A8. Ashworth NL, Chad KE, Harrison EL, et al. Home versus center based physical activity programs in older adults. *Cochrane Database Syst Rev*. 2005;1:CD004017.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S. adults: compliance with the Physical Activity Guidelines for Americans. *Am J Prev Med*. 2011;40:454-461.
2. U.S. Department of Health and Human Services. *2008 Physical Activity Guidelines for Americans*. http://www.health.gov/paguidelines/guidelines/default.aspx; Accessed January 29, 2015.
3. Wen CP, Wai JP, Tsai MK, et al. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. *Lancet*. 2011;378:1244-1253.
4. Shin SY, Park JI, Park SK, et al. Utility of graded exercise tolerance tests for prediction of cardiovascular mortality in old age: The Rancho Bernardo Study. *Int J Cardiol*. 2014;181C:323-327.
5. Powell KE, Paluch AE, Blair SN. Physical activity for health: what kind? how much? how intense? On top of what? *Annu Rev Public Health*. 2011;32:349-365.
6. Yates T, Haffner SM, Schulte PJ, et al. Association between change in daily ambulatory activity and cardiovascular events in people with impaired glucose tolerance (NAVIGATOR trial): a cohort analysis. *Lancet*. 2014;383:1059-1066.
7. Studenski S, Perera S, Patel K, et al. Gait speed and survival in older adults. *JAMA*. 2011;305:50-58.
8. Physical Activity Guidelines Advisory Committee. Part G. Section 10. Adverse events. In *Physical Activity Guidelines Advisory Committee Report*. http://www.health.gov/paguidelines/Report/G10_adverse.aspx; Accessed January 29, 2015.
9. Loprinzi PD, Cardinal BJ. Association between biologic outcomes and objectively measured physical activity accumulated in ≥10-minute bouts and <10-minute bouts. *Am J Health Promot*. 2013;27:143-151.
10. LeFevre ML; U.S. Preventive Services Task Force. Behavioral counseling to promote a healthful diet and physical activity for cardiovascular disease prevention in adults with cardiovascular risk factors. U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;161:587-593.
11. U.S. Preventive Services Task Force. Screening for coronary heart disease with electrocardiography. http://www.uspreventiveservicestaskforce.org/uspstf/uspsacad.htm; 2012 Accessed January 29, 2015.
12. Community Preventive Services Task Force. Increasing physical activity. http://www.thecommunityguide.org/pa/index.html; Accessed January 29, 2015.

REVIEW QUESTIONS

1. A 50-year-old man seeks advice about increasing his level of physical activity. While reading about exercise on the Internet, he noted that several websites advised: "consult your physician prior to starting an exercise program." Currently, he rarely engages in moderate to vigorous physical activity, except occasionally on weekends when he does several hours of yard work. He has experienced gradual weight gain over the past two decades, and his body mass index is 28. His medical history is unremarkable except for high blood pressure, which is well controlled by hydrochlorothiazide. He is specifically interested in advice about exercise-related injuries. As a young man, he jogged regularly but had problems with pain in his knees. He also hopes he can lose weight by exercising more. Based the history and the *2008 Physical Activity Guidelines for Americans*, which of the following statements is appropriate to include in your advice to this man?

A. He should regularly perform moderate to vigorous intensity aerobic activity. He should choose a type or types he prefers and spread the aerobic activity throughout the week.
B. To determine the intensity (level of effort) of his activity, his only option is to monitor his heart rate.
C. Vigorous-intensity activity is preferable in adults his age. His goal should be to get at least 60 minutes of vigorous activity each week.
D. Aerobic activity is important, but he should also do muscle strengthening exercises (such as weight training) 3 days per week, with two sets of 10 repetitions for each muscle group.
E. If he is considering getting his physical activity by walking, his goal should be to walk at least 30 minutes a day on at least 5 days per week.

Answer: A This advice is relevant given the man's history of concentrated yard work because the guidelines explicitly discourage large amounts of physical activity in a single day (so-called weekend warriors). B is incorrect because adults can use either absolute or relative intensity to guide their level of effort (heart rate monitoring is the most common way to assess relative intensity). C is incorrect in two ways—the guidelines do *not* state that vigorous activity is preferable, and the guidelines for vigorous activity are 75 minutes per week. D is incorrect in two ways—muscle strengthening exercises need to be on only 2 days per week, and the key guidelines do not have a requirement for two sets for each muscle group. E is incorrect because the current guidelines no longer require physical activity on "most days of the week" or "at least 5 days per week."

2. In asymptomatic healthy adults, an evidence-based approach to reducing the risk for activity-related injuries includes which of the following?

A. Gradually increasing the physical activity level over time
B. Prescribing flexibility exercises
C. Screening for coronary artery disease using either an exercise test or electron-beam computed tomography
D. Starting with supervised exercise and then transitioning to unsupervised
E. None of the above

Answer: A Gradually increasing physical activity over time is a key guideline for injury prevention. There is no evidence that flexibility exercises have health benefits, including no evidence that they prevent injuries. The U.S. Preventive Services Task Force does not recommend routine screening for coronary artery disease using an exercise treadmill test or using electron-beam computed tomography. There is insufficient evidence that supervised exercise reduces injury risk in healthy adults.

3. Which one of the following statements about the benefits and risks of physical activity is true?

A. The overall risk for sudden death is higher in people who exercise regularly.
B. The total volume of physical activity is the main determinant of health benefits.
C. Obese adults do not obtain benefits from physical activity unless they lose weight.
D. The benefits of being active at less than recommended levels are so small that it is of little value to recommend and promote physical activity in people who are not capable of meeting guidelines.
E. The benefits of increasing physical activity do not depend on current level of activity. A person who increases physical activity from 0 minutes to 30 minutes a day obtains the same health benefits as a person who increases activity from 120 minutes a day to 150 minutes per day.

Answer: B The overall risk for sudden death is lower in people who have a higher total volume of exercise. The health benefits of physical activity are independent of other risk factors, and obese adults obtain benefits from physical activity even if obesity persists. Consistent with the dose-response effect, there are meaningful benefits from obtaining less than the recommended amounts of activity. There is a curvilinear dose-response relationship, and the marginal health benefits of increasing activity are less at high levels of physical activity.

4. Which one of the following is recommended for the promotion of physical activity in clinical settings?

A. Counseling and assisting virtually all patients to be physically active with techniques such as motivational interviewing
B. Recommending people start with moderate-intensity activities such as walking, but advising virtually all patients gradually to shift to doing mainly vigorous intensity activity
C. Paying little attention to community resources such as access to parks, gyms, and multiuse trials because there is little scientific evidence that they influence activity
D. Advising virtually all patients to engage in regular physical activity, with the specific activity recommendation based on integrating public health guidelines for prevention with any applicable therapeutic guidelines
E. Emphasizing exercise prescription and exercise classes because most people prefer these activities over lifestyle activities such as walking for transportation and gardening

Answer: D The physician should advise patients regarding the types and amounts of activity appropriate for them, specifically taking into account chronic diseases and activity limitations in making these recommendations. The U.S. Preventive Services Task Force does not recommend exercise counseling for all patients. There is strong evidence that access to community resources affects physical activity levels. Most patients prefer home-based exercise rather than exercise classes. Vigorous activity is not recommended for all adults.

5. Which one of the following statements is true about the use of physical activity to reduce the risk for falls and functional limitations?

A. In overweight and obese older adults with physical functional limitations, weight loss is more important than physical activity in improving function.
B. The fact that falls are more likely during activities such as walking explains why exercise programs in older adults do not reduce risk for falls.
C. Because gait speed in older adults is mainly determined by leg length, it is not a useful measure of mobility or mortality risk.
D. Low fitness (low exercise capacity) does not significantly affect the risk for mortality and is mainly important because it causes functional limitations.
E. Performing Tai Chi exercises can reduce the risk for falls in elderly people.

Answer: E Tai Chi is an evidence-based method for reducing the risk for falls in elderly people. Weight loss is not more important than physical activity in promoting function. Exercise programs, including balance exercises, reduce the risk for falls. Exercise increases gait speed, and better gait speed is a strong predictor of lower mortality. Low physical fitness is associated with higher mortality.

17

ADOLESCENT MEDICINE

DEBRA K. KATZMAN AND LAWRENCE S. NEINSTEIN

Adolescence, a period of transition between childhood and adulthood, is marked by critical biologic, psychological, social, and cognitive changes. During this unique developmental period, patterns of behaviors and lifestyle choices are established that can influence current and future health, and unique medical and psychological problems can emerge.[1] Adult-oriented health care providers play a pivotal role in engaging youth in their health and in providing health care to adolescents and young adults.

NORMAL PHYSICAL GROWTH AND DEVELOPMENT

Biologic growth and development in adolescents are signified by the onset of puberty (Chapter 235), which varies temporally among adolescents and explains why adolescents of the same chronologic age can vary greatly in physical appearance. The first visible sign of puberty among girls is usually thelarche, or the development of breast buds, which occurs on average at 10.5 years in white girls and 1 year earlier in African American girls (Chapter 235).

Menarche occurs 2 to 4 years after the initial appearance of breast buds and pubic hair. The average age of menarche is 12.9 years for white girls and 12.2 years for African American girls. Menstrual periods are not always regular during the first 2 years after menarche. At menarche, only 20% of cycles are ovulatory; it may take up to another 4 years for 80% of cycles to be ovulatory. The average length for the completion of puberty in girls is 4 years (range, 1.5 to 8.0 years).

Puberty begins about 2 years later in boys than in girls. The first physical sign of puberty in boys, testicular enlargement and thinning of the scrotum, occurs at 11.5 years. Adrenarche occurs 6 months later at an average age of 12.0 to 12.5 years (Chapter 234). Facial hair starts to grow about 3 years after pubic hair. The completion of puberty in boys can take an average of 3 years (range, 2 to 5 years).

Pubertal weight gain accounts for about half of a person's ideal adult body weight. Peak weight gain follows the linear growth spurt by 3 to 6 months in adolescent girls and by about 3 months in adolescent boys. The average weight gain during puberty among adolescent girls is 15 to 55 lb or 7 to 25 kg (mean gain, 38.5 lb or 17.5 kg). Overall, adolescent boys gain 15 to 65 lb or 7 to 30 kg during puberty (mean gain, 52.2 lb or 23.7 kg).

Boys' body fat levels decrease during adolescence, dropping to 12% body fat by the end of puberty. On average, adolescent girls' lean body mass falls from 80 to 75%, whereas their average body fat levels increase from 16 to 27% by the end of adolescence. By the time adolescent girls are 16 years of age and adolescent boys are 18 years of age, they have accrued more than 90% of their adult skeletal mass.

Sexual maturity ratings (SMRs), also known as Tanner staging, are used to describe the progression of secondary sexual characteristics that occur in adolescents, irrespective of chronologic age. SMR is based on the development of breasts and the appearance of pubic hair among girls (Fig. 17-1) and on testicular and penile development and the appearance of pubic hair among boys (Fig. 17-2). SMR 1 corresponds to the prepubertal stage; puberty has not begun, and no sexual development has occurred. SMR 2 to SMR 5 indicate the progression of puberty to adulthood. Once young people reach SMR 5, they have fully developed secondary sexual characteristics. Sexual maturation correlates with linear growth, changes in weight and body composition, and hormonal changes. Sexual maturation provides important assurance of the normal progression of puberty or identification of abnormal pubertal development.[2]

NORMAL PSYCHOSOCIAL DEVELOPMENT

Adolescence is often divided into three psychosocial developmental phases: early adolescence (11 to 13 years), middle adolescence (14 to 16 years), and late adolescence (17 to 21 years). In early adolescence, adolescents begin to separate from their parents and establish an individual identity. As adolescents pull away from their parents in search of their own identity, their peer group takes on an important and special significance.

At the beginning of adolescence, cognitive abilities are dominated by concrete thinking. Young adolescents lack abstract reasoning capabilities, problem-solving skills needed to overcome barriers to behavioral changes, and the ability to appreciate how their current behaviors can affect their future health status.

Middle adolescence is characterized by growth in emotional autonomy and increasing separation from family. The adolescents' peer groups play a powerful role, and adolescents are increasingly involved in partnering relationships that include dating, sexual experimentation, and adverse health behaviors such as smoking cigarettes, drinking alcohol, using street drugs, and being truant. Abstract reasoning skills emerge but often regress to concrete thinking when adolescents are faced with stressful situations. These adolescents begin to understand the relationship between health behaviors and future health status, but peer pressure may make it challenging for them to make appropriate health-related choices.

During late adolescence, young people become increasingly more economically and emotionally independent. Peer group values become less important, and young people spend more time in a relationship with one person. The late stage of adolescence is characterized by the development of a strong personal identity. Abstract reasoning skills expand, and older adolescents have problem-solving skills that help them overcome challenges to behavioral change.

VITAL STATISTICS

Although adolescents and young adults are generally perceived to be healthy, their morbidity and mortality are often the result of risky behaviors and social forces (Table 17-1).[3] Unintentional injury, homicide, and suicide are the leading causes of death in 15- to 24-year-olds and account for more than 70% of all adolescent and young adult deaths. About 75% of these causes of death and injury are preventable.

Unintentional and Intentional Injuries

Unintentional injuries account for 44% of all injury deaths to children and adolescents. For every childhood death caused by injury, another 34 hospitalizations and 1000 emergency department visits occur. Major causes of unintentional injuries to adolescents include motor vehicle accidents, being struck by or against an object or person, cuts from sharp objects, and falls. Athletic injuries, which also are frequent, can be reduced by targeted prevention programs. Factors that contribute to adolescent injuries include socioeconomic factors (poor adolescents are at greatest risk for injury), environmental factors (hazards such as all-terrain vehicles, backyard swimming pools, firearms, kerosene heaters, and gang activity), school environment, and developmental factors.

About 20% of all adolescents, especially adolescent girls in heterosexual relationships, report having experienced either psychological or physical violence from a dating partner. About 20 to 30% of students in grades 6 to 10 are involved in bullying, as a bully, victim, or bully-victim (those who are both aggressive to peers and victimized by peers). For younger adolescents, bullying can lead to a significantly higher risk of psychosomatic problems.

Suicide (Chapter 397) is the second leading cause of death among adolescents and young adults 15 to 24 years of age. Native Americans and white youth 15 to 24 years of age have the highest suicides rates, whereas African American youth 15 to 19 years of age and Asian youth 20 to 24 years of age have the lowest. The estimated ratio of attempted-to-completed suicides among adolescents ranges between 50:1 and 100:1. The most common methods used in suicide attempts are drugs and alcohol, whereas suffocation, hanging, or use of firearms is associated with completed suicides. Death rates caused by firearms are eight times higher in adolescent boys than in adolescent girls.

Homicide is the third leading cause of death in the 15- to 24-year-old U.S. population and the number one cause of death among African American males 15 to 24 years of age. About 75% of homicides in older adolescents and young adults involve firearms.

Clinicians should discuss the impact of alcohol and drugs on driving and stress the importance of routine use of seat belts when driving in a motor vehicle. Clinicians also should discuss safety during sports activities, such as use of bicycle helmets and sports-related protective gear.

Other Diseases

Excluding intentional and unintentional injuries, cancer is the leading cause of death in adolescents and is the leading cause of death by disease. Cancers with an increased incidence in adolescents include Hodgkin lymphoma

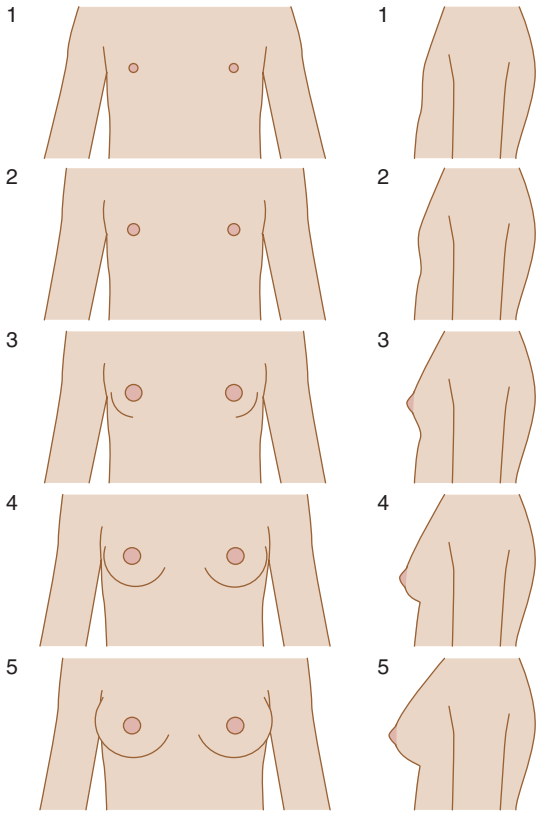

FIGURE 17-1. Female breast development. *Sexual maturity rating 1 (SMR 1)*: Preadolescent; no glandular tissue. Areola and papilla: Areola conforms to general chest line. *SMR 2:* Breast buds appear; areola is slightly widened and projects as small mound. *SMR 3:* Enlargement of the entire breast with protrusion of the papilla or of the nipple. Breast and areola enlarge with no separation of their contours. *SMR 4:* Enlargement of the breast and projection of areola and papilla as a secondary mound. *SMR 5:* Adult configuration of the breast with protrusion of the nipple; areola no longer projects separately from remainder of breast. (Redrawn from Daniel WA, Paulshock BZ. A physician's guide to sexual maturity rating. Patient Care. 1979;30:122. Original illustration by Paul Singh-Roy.)

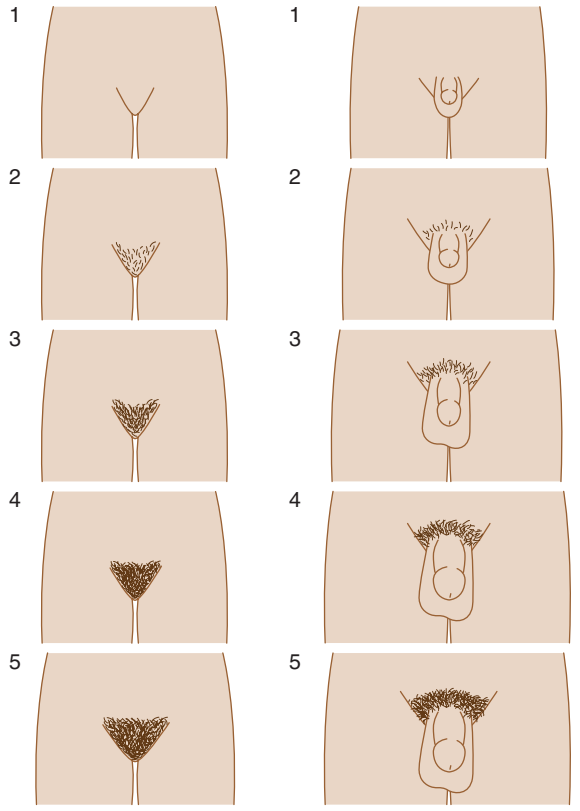

FIGURE 17-2. Female and male pubic hair development. *Sexual maturity rating 1 (SMR 1):* Prepubertal; no pubic hair. *SMR 2:* Small amount of scanty, long, slightly pigmented, along the base of the scrotum and phallus in the male and the medial border of the labia majora in females. *SMR 3:* Darker, coarser, starts to curl, small amount, extending laterally. *SMR 4:* Coarse, curly; resembles adult type but does not extend to the medial surface of the thighs. *SMR 5:* Abundant, adult-type pattern; hair extends onto the medial aspect of the thighs. ***Right,* Male genital development.** *SMR 1:* Penis preadolescent. Testicular volume <4 mL. *SMR 2:* Penis slight or no enlargement. Beginning enlargement of testes; testicular volume 4 to 8 mL; scrotal skin reddened, thinner. *SMR 3:* Penis increased in length. Testicular volume 10 to 15 mL. Further enlargement of scrotum. *SMR 4:* Penis larger in breadth, glans penis develops. Testes and scrotum nearly adult; testicular volume 12-20 mL. *SMR 5:* Penis adult size; testicular volume greater than 25 mL. (Redrawn from Daniel WA, Paulshock BZ. A physician's guide to sexual maturity rating. Patient Care. 1979;30:122. Original illustration by Paul Singh-Roy.)

(Chapter 186), germ cell tumors (Chapter 200), central nervous system tumors (Chapter 189), non-Hodgkin lymphoma (Chapter 185), thyroid cancer (Chapter 226), malignant melanoma (Chapter 203), and acute lymphoblastic leukemia (Chapter 183). HIV infection continues to be one of the 10 leading causes of death among people 15 to 24 years of age.

APPROACH TO THE ADOLESCENT PATIENT

When interviewing adolescent patients, the clinician should consider the adolescents' physical, cognitive, and psychosocial developmental stage; their growing autonomy and increasing role in taking responsibility for their own health and well-being; and their individual progression from childhood to adulthood. Chronologic age is not a guarantee that all adolescents will be at the same stage of physical, cognitive, and psychosocial development. The clinician should offer adolescent health care in a sensitive, flexible, and developmentally and culturally appropriate manner.

The clinician can meet with the adolescent alone at first to give the adolescent the message that she or he is the patient and the clinician is most eager to hear what the adolescent has to say. Conversely, the clinician can meet with the adolescent and her or his parents or guardians together for the initial part of the interview and then meet with the adolescent alone; this approach permits the clinician to develop an understanding of the reason for the visit from the perspective of the adolescent and his or her parents or guardians and to communicate that input from both the adolescent and the parents or guardians is highly valued. This approach also provides an opportunity for the clinician to observe the interaction between the adolescent and parents or guardians. The final approach to the interview is to greet the adolescent and family and request a meeting with the parents or guardians alone. This approach allows the parents or guardians to discuss concerns about the adolescent that they may not feel comfortable raising in the presence of the adolescent. Having this information may improve the focus of the visit. In this approach, the encounter with the parents or guardians is then followed by the clinician's meeting with the adolescent alone. With this interview structure, it is important that the adolescent be present from the time he or she meets with the clinician until the end of the visit, so that the adolescent does not perceive a breach of confidentiality.

TABLE 17-1 LEADING CAUSES OF DEATH IN U.S. ADOLESCENTS

CAUSE	PERCENTAGE OF DEATHS IN YOUTH AGED 15-24 YEARS
Unintentional injury	40.6
Suicide	15.8
Homicide	15.2
Malignant neoplasias	5.4
Diseases of the heart	3.2
Congenital anomalies	1.4
Cerebrovascular diseases	0.6
Pregnancy, childbirth, and the puerperium	0.6
All other causes	15.7

From Hoyert DL, Xu J. Deaths: preliminary data for 2011. National Vital Statistics Reports. Hyattsville, MD: National Center for Health Statistics; 2012.

Confidentiality

Issues of consent and confidentiality are central in the physician-adolescent interaction. Adolescents appreciate clinicians whom they trust and who can

assure them of confidentiality. When guaranteed confidentiality, adolescents are more likely to seek necessary medical care, to disclose sensitive information, and to trust their clinician. Under these circumstances, most adolescents will involve their parents in their care. Parents appreciate education about the concept of confidentiality and recognize the importance of allowing the adolescent the opportunity to speak alone with the clinician. Confidentiality is also important for clinicians. To make an accurate diagnosis and to provide treatment, the clinician must obtain all relevant information from the adolescent. If the adolescent fears that such information will not be kept confidential, she or he may not provide all the necessary factual information.

Confidentiality and defining the limits of confidentiality should be discussed with the adolescent and his or her parents or guardians at the beginning of the interview. The adolescent and family need to know that the clinician will intervene if he or she believes that the adolescent's actions may cause him or her, or another person, significant harm. Examples of situations in which the clinician would not maintain confidentiality when dealing with young people include disclosure of current suicidal or homicidal intent.

Those who work with young people must have a clear understanding of consent and confidentiality and should make sure that adolescents are aware of confidentiality policies and practices. The duty of confidentiality does not preclude encouraging and empowering adolescents to talk to their parents or guardians about important health care issues and to include them in discussions of these issues. The legal definition of confidentiality varies, depending on geographic location, and clinicians should be familiar with the local laws.

Another goal of the interview is to build rapport with the adolescent and his or her parents or guardians. Clinicians can establish rapport with the adolescent at the start of the interview by creating an environment that is nonjudgmental, unintimidating, and supportive. The clinician's genuine interest in the adolescent is paramount. It is helpful to encourage the adolescent to talk about himself or herself—friends, hobbies, school. The clinician should listen carefully to the adolescent's statements and feelings. Sensitivity to and understanding of the adolescent's developmental stage and cultural background are important in interviewing an adolescent and interpreting answers accurately. The clinician should be respectful of the adolescent's growing need to be independent and desire to be treated as an individual person. Taking the time to build rapport is key to engaging the adolescent in a discussion of his or her personal health concerns with the clinician. It is the clinician's responsibility to familiarize himself or herself with the legal issues of confidentiality (as they relate to the adolescent patient and his or her parents or guardians) in his or her locality.

Preventive Health Care

Preventive health care for adolescents (Chapters 15 and 18) should promote physical and mental health and healthy physical, psychological, and social growth and development.[4] Positive behaviors such as exercise (Chapter 16) and nutritious eating (Chapter 213) should be encouraged, and health risk behaviors such as unsafe driving, use of tobacco (Chapter 32), unsafe sexual behaviors, and excess alcohol (Chapter 33) should be discouraged. Because lifelong health habits are established during adolescence, it is an important time to invest in health promotion and preventive services.

The U.S. Preventive Services Task Force recommends screening and counseling of all adolescents about depression and obesity and ensuring that all adolescents are up-to-date on their immunizations for tetanus, diphtheria, pertussis, varicella, measles, mumps, rubella, hepatitis B, meningococcus, polio, and human papillomavirus (HPV; Chapter 18). In addition, at-risk adolescents should be counseled about sexually transmitted diseases (e.g., HIV, chlamydia, gonorrhea, and syphilis) and advised about immunizations for influenza, pneumococcus, and hepatitis A.[5] Influenza immunization is recommended for all adolescents unless there is a vaccine shortage, in which case at-risk adolescents should be vaccinated first.

Components of the Adolescent Care Health Visit

History

Open-ended, nonjudgmental, developmentally appropriate, and gender-neutral questions help put the adolescent at ease and produce informative answers (Chapter 15). In addition to a standard medical history, the assessment should include a psychosocial history from the adolescent, either through a screening questionnaire or during the interview with the adolescent alone. The HEEADSSS assessment, which is a valuable screening tool for obtaining a comprehensive psychosocial history, covers the following topics:

- **H**ome: family members, living arrangements, and relationships
- **E**ducation/**E**mployment: academic or vocational success and future plans
- **E**ating: concerns about weight or body image or disordered eating attitudes and behaviors
- **A**ctivities: recreational activities, dating, and relationships
- **D**rugs: use of tobacco, alcohol, illicit drugs, anabolic steroids, and driving while intoxicated
- **S**exuality: sexual orientation, sexual activity, and sexual abuse
- **S**uicide (mental health): feelings of sadness, loneliness, depression; or suicidal ideation, attempts, and non-suicidal self-injury
- **S**afety: risk of unintentional injury or violence, fighting; or weapon carrying

Physical Examination

In general, the adolescent physical examination should occur without a parent or guardian present. In some situations, the adolescent is asked whether he or she would prefer to have a parent in the room during the physical examination. A male clinician should request that a female health care provider be present during the physical examination of a female patient, especially during the breast and genital examination. In theory, a female clinician should request a male health care provider be present during the genital examination of a male patient.

Care should be taken to ensure the adolescent's privacy. Providing an examination gown that covers the trunk and genital area is important. Talking with the adolescent during the examination also tends to increase comfort; explaining the procedures and commenting on the results are helpful. A comprehensive physical examination (Chapter 15) should include measuring the adolescent's weight and height, calculating his or her body mass index, plotting these measurements on standardized growth charts, and determining the adolescent's SMR.

The most recent American guidelines for cervical cancer screening recommend against screening women younger than 21 years.[6] These recommendations do not apply to young women with previously abnormal test results on cervical screening or to women who are immunosuppressed by HIV infection, organ transplantation, chemotherapy, or chronic use of corticosteroids. Indications for a pelvic examination in an adolescent include symptoms of vaginal or uterine infection, menstrual irregularities (e.g., amenorrhea, dysfunctional uterine bleeding, menorrhagia, and severe dysmenorrhea), undiagnosed abdominal or pelvic pain, tenderness, mass, trauma, sexual abuse, or assault. A careful history and a urine or patient-obtained specimen may be an alternative to the pelvic examination.

Conclusion of the Health Visit

At the conclusion of the adolescent health visit, the clinician should review the findings with the adolescent and talk about what happens next. The adolescent should have the opportunity to ask questions, get clarification, make comments, and respond to suggestions. The clinician should discuss with the adolescent what information will remain confidential. The clinician should then meet with the adolescent and the parents or guardians (if they have accompanied the adolescent to the health visit) to review the outcomes and to discuss the nonconfidential issues.

REPRODUCTIVE HEALTH

Reproductive health care includes issues of adolescent sexual development, adolescent sexual behaviors, adolescent pregnancy, and contraception. In the United States, nearly 50% of high-school boys and girls in grades 9 to 12 have had vaginal intercourse at least once. This number increases from 38% of boys and 28% of girls in grade 9 to about 63% of both boys and girls by their senior year. In addition, 3.4% of girls and 9% of boys in grades 9 to 12 started having sex before they were 13 years old, 15% of adolescents have had four or more sex partners during their lives, and 40% of sexually active high-school students did not use a condom when they last had sexual intercourse. About 60% of adolescent girls who are 13 years of age or younger when they first had sexual intercourse report having involuntary intercourse. Adolescents are more likely to report engaging in oral than in vaginal sex because they perceive oral sex as significantly less risky (fewer health, social, and emotional consequences) than vaginal sex, but adolescents who engage in oral sex also are more likely to engage in vaginal sex.[7]

Every year, about 750,000 pregnancies occur in girls 15 to 19 years of age in the United States. About 50% of these pregnancies result in a live birth, about 35% end with an abortion, and about 15% end with a miscarriage or stillbirth. During the past 20 years, teenage pregnancy rates in the United States have declined by 42%.[8] The birth rate for mothers 15 to 19 years of age decreased by about 50%, from about 62 per 100,000 to about 31 per 100,000.

Abortion rates also have decreased. Most of these decreases are because of increased use of effective contraception and decreased sexual activity. Adolescent pregnancy is more common among black and Hispanic girls as well as among girls from low-income families.

Positive outcomes of adolescent pregnancies are enhanced with good prenatal care, adequate initial and follow-up prenatal visits, nutritional counseling, assessment of psychosocial issues, and substance abuse counseling. Adolescents younger than 15 years are at increased risk for premature and low-birthweight infants; those older than 15 years who have adequate prenatal care do not have increased adverse outcomes.

Contraception

The most commonly used contraceptive methods in adolescents are oral contraceptive pills and condoms, although new hormonal delivery systems, such as transdermal patches, vaginal rings, and long-acting reversible contraceptives, are convenient and effective (Chapter 238). Long-acting reversible contraceptive methods (intrauterine devices and contraceptive implants) are increasingly popular in adolescents, who are at high risk of unintended pregnancy. Routine counseling about emergency contraception is an important part of the public health strategy to reduce teen pregnancy (Chapter 238).

Sexually Transmitted Diseases

Every year, about 4 million adolescents in the United States acquire a sexually transmitted disease (Chapter 285), representing 25% of the total cases of sexually transmitted disease diagnosed annually. *Chlamydia trachomatis,* which is the most commonly reported bacterial sexually transmitted disease, is reported annually in about 3.4% of female adolescents 15 to 19 years of age and 0.6% of male adolescents of the same age. The annual incidence of gonorrhea is about 0.6% for female adolescents 15 to 19 years of age and 0.25 % for adolescent males of the same age. In adolescents, HIV infection (Chapter 384) is contracted primarily as a sexually transmitted disease, and adolescents and young adults 15 to 24 years of age represent 14% of all new diagnoses of HIV infection in the United States. HPV (Chapter 373), which is the most prevalent of all sexually transmitted diseases in 15- to 24-year-olds, has a prevalence of about 20% in adolescent girls 14 to 17 years of age.

Adolescents are at greater risk of acquiring sexually transmitted diseases because they are more likely to engage in unprotected sexual intercourse with concurrent partners who have multiple other partners. Adolescent girls may have persistent vaginal columnar epithelium that is more susceptible than squamous epithelium to *Neisseria gonorrhoeae, C. trachomatis*, and HPV as well as lower levels of immunoglobulin A in their cervical mucus. Treatment may be delayed owing to an inability to access confidential health care services, lack of health care coverage, poverty, and drug trafficking and use.

Chlamydial infections (Chapter 318), often asymptomatic or minimally symptomatic in women, usually are manifested with a vaginal discharge in young adolescent girls. Otherwise, clinical presentations and treatments of sexually transmitted diseases (Chapter 285) in adolescents are similar to those in adults.

About 20% of adolescents given prescriptions for a sexually transmitted disease treatment fail to fill their prescriptions. Therefore, whenever it is available, an observed single-dose therapy is recommended.[9] In addition, expedited partner therapy, which is the treatment of sexual partners without requiring a prior clinical evaluation or prevention counseling, reduces the level of recurrent sexually transmitted diseases better than does standard management. Clinicians should consider the use of expedited partner therapy for partners exposed within 60 days to heterosexual males and females with chlamydia or gonorrhea infections when in-person evaluation and treatment are unlikely (Chapter 285).

ADOLESCENT OBESITY AND EATING DISORDERS

More than one third of children and adolescents are overweight or obese. Obese adolescents are significantly more likely to become severely obese in adulthood. Interventions should focus on changing behavior, increasing physical activity, and modifying the diet. Randomized trials also show that substituting sugar-free beverages for sugar-sweetened beverages can reduce short-term weight gain,■ but family-based behavioral treatment programs offer the best short- and long-term success for weight management.■ In extreme cases, pharmacotherapy and even bariatric surgery may be considered in the therapeutic plan (Chapter 220).

Eating disorders (Chapter 219) commonly begin during adolescence.[10] For adolescent-onset eating disorders, early recognition and aggressive treatment are critical to a successful outcome. The medical complications of eating disorders in adolescents include growth retardation, pubertal delay, low bone mineral density, and changes in brain structure and cognitive function. These complications, which occur early in the disease process, may not be completely reversible, thereby underscoring the need for early and aggressive treatment.

The goals of treatment are to restore physical health, normal eating behavior patterns, and mental health and to reduce the impact of the eating disorder on the quality of life. Adolescents should be treated at a facility where the health professionals understand eating disorders and have experience in treating adolescents. Successful management strategies for adolescents with eating disorders include early restoration of a normal nutritional and physiologic state, involvement of the family in the treatment, and incorporation of an interdisciplinary team in the treatment. Family-based therapy is an effective treatment for adolescents with eating disorders and helps protect against relapse.■

SUBSTANCE ABUSE

Nearly half of adolescents try an illicit drug (Chapter 34) by the time they finish high school. About 75% of adolescents consume alcohol (Chapter 33) by the end of high school, and about half report having been drunk at least once in their life. Adolescents who begin using alcohol or drugs before 15 years of age are more than five times more likely to develop an addictive disorder later in life compared with those who first use alcohol at 21 years of age. About 45% of adolescents use marijuana, and 20% of 12th graders are current smokers.

Adolescents start using drugs or alcohol primarily because of social pressures but report continuing them to feel good or to cope with difficulties.[11] Boys tend to initiate drug and alcohol use at younger ages than girls do, but once girls begin to experiment, they are just as likely as boys to use drugs. Boys are more likely to consume marijuana, steroids, and smokeless tobacco, whereas girls are more likely to abuse amphetamines and methamphetamine.

Adolescents with poor self-esteem, low motivation, and poor academic achievement have a greater propensity for alcohol and drug abuse than those with positive self-esteem. Adolescents with a family member with a history of alcohol or other drug abuse are at greater risk. Adolescents are less likely to succumb to external pressures toward drug use if they have a strong sense of attachment to parents who clearly communicate their disapproval.

Signs and symptoms suggestive of substance abuse include changes in physical appearance, poor hygiene or dress, wearing long-sleeved shirts to hide scarring at injection sites, persistent cough or bronchitis, difficulty sleeping, sudden weight loss or weight gain, sudden changes in personality, aggressive behavior, irritability, nervousness, giddiness, changes in peer group, increased isolation from peers or family, depression, loss of interest in once favorite activities, decline in performance or attendance at school or work, forgetfulness, increased secretiveness, money or objects disappearing from the household, and prescription drugs that seem to be used up too quickly.

A validated screening tool for alcohol and substance abuse (Table 17-2) should be used in adolescents. Adolescents who answer yes to one of the six questions should receive advice on the adverse effects of substance abuse. Those who answer yes to two or more questions are at high risk and require further assessment or referral (Chapters 33 and 34). Urine drug testing has low sensitivity for detection of drug use, and there is no consensus about its role for screening of adolescents.

Alcohol and drug use contributes to more than 40% of adolescent deaths from motor vehicle crashes, to suicide attempts, and to an increased risk for subsequent use and its related problems in adulthood.

Historically, substance abuse treatment has been based on abstinence. More recently, a harm-reduction approach has gained acceptance. Adolescents discontinue drug and alcohol use mainly because of concern about their negative effects, and community-based interventions can be effective.■ To reduce alcohol abuse over time, individual interventions have a larger impact than family-based interventions.■

CHRONIC ILLNESS AND TRANSITION

The prevalence of chronic diseases in adolescents has increased significantly because advances in medical technology and treatments have increased the survival of young people with childhood diseases formerly considered lethal.

TABLE 17-2 CRAFFT SCREENING TOOL FOR DRUG AND ALCOHOL USE IN ADOLESCENTS

The adolescent is instructed, "Please answer these next questions honestly. Your answers will be kept confidential. During the past 12 months, did you …"
Drink any alcohol (more than a few sips)?
Smoke any marijuana or hashish?
Use anything else to get high?

If the adolescent answers no to the three opening questions, the provider only needs to ask the adolescent the first question—the CAR question. If the adolescent answers yes to any one or more of the three opening questions, the provider asks all six CRAFFT questions.

C—Have you ever ridden in a CAR driven by someone (including yourself) who was "high" or had been using alcohol or drugs?
R—Do you ever use alcohol or drugs to RELAX, feel better about yourself, or fit in?
A—Do you ever use alcohol/drugs while you are by yourself, ALONE?
F—Do you ever FORGET things you did while using alcohol or drugs?
F—Do your family or FRIENDS ever tell you that you should cut down on your drinking or drug use?
T—Have you gotten into TROUBLE while you were using alcohol or drugs?

CRAFFT is a mnemonic acronym of the first letters of key words in the six screening questions. The questions should be asked exactly as written.

About 18% of adolescents in the United States live with a chronic illness. Chronic illness may affect the adolescent's development, or the adolescent's development may affect the illness. For example, cystic fibrosis (Chapter 89) may delay puberty and hinder normal peer development. Puberty can exacerbate diabetes mellitus (Chapter 229). Increased risk-taking by adolescents with diabetes, asthma, or chronic renal failure can hinder their compliance with their medication regimen.

Adolescents in general, and those with special health care needs in particular, require a smooth, seamless, coordinated, and developmentally appropriate transition to the adult health care system. Barriers to a successful transition include a lack of adult physicians who can manage chronic childhood conditions (e.g., congenital heart disease; Chapter 69); patients, families, and pediatric subspecialists who are reluctant to terminate long-standing relationships; and patients and families who are poorly equipped to find their way in the adult health care system. As adolescents move closer to the age of transition, professionals should provide developmentally appropriate information and skills to support them and their families as they negotiate the adult health care system.

Adolescents benefit from an introductory visit with their adult-oriented care provider before leaving the pediatric health care system. This visit should give the adolescent a clearer idea about his or her new role as an adult patient, especially about expectations around decision making, giving consent, and the role of the family in his or her health care. The patient, family, and pediatric and adult health care systems should view the transition as a natural part of the developmental process of care for all adolescents.

Grade A References

A1. Rossler R, Donath L, Verhagen E, et al. Exercise-based injury prevention in child and adolescent sport: a systematic review and meta-analysis. *Sports Med*. 2014;44:1733-1748.
A2. Ebbeling CB, Feldman HA, Chomitz VR, et al. A randomized trial of sugar-sweetened beverages and adolescent body weight. *N Engl J Med*. 2012;367:1407-1416.
A3. Sung-Chan P, Sung YW, Zhao X, et al. Family-based models for childhood-obesity intervention: a systematic review of randomized controlled trials. *Obes Rev*. 2013;14:265-278.
A4. Couturier J, Kimber M, Szatmari P. Efficacy of family-based treatment for adolescents with eating disorders: a systematic review and meta-analysis. *Int J Eat Disord*. 2013;46:3-11.
A5. Oesterle S, Hawkins JD, Fagan AA, et al. Testing the universality of the effects of the communities that care prevention system for preventing adolescent drug use and delinquency. *Prev Sci*. 2010;11:411-423.
A6. Tripodi SJ, Bender K, Litschge C, et al. Interventions for reducing adolescent alcohol abuse: a meta-analytic review. *Arch Pediatr Adolesc Med*. 2010;164:85-91.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Sawyer SM, Afifi RA, Bearinger LH, et al. Adolescence: a foundation for future health. *Lancet*. 2012;379:1630-1640.
2. Palmert MR, Dunkel L. Clinical practice. Delayed puberty. *N Engl J Med*. 2012;366:443-453.
3. Hoyert DL, Xu J. *Deaths: preliminary data for 2011. National Vital Statistics Reports*. Hyattsville, MD: National Center for Health Statistics; 2012.
4. Catalano RF, Fagan AA, Gavin LE, et al. Worldwide application of prevention science in adolescent health. *Lancet*. 2012;379:1653-1664.
5. U.S. Preventive Services Task Force. Child and adolescent recommendations. http://www.uspreventiveservicestaskforce.org/tfchildcat.htm; Accessed January 29, 2015.
6. Saslow D, Solomon D, Lawson HW, et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *CA Cancer J Clin*. 2012;62:147-172.
7. Eaton DK, Kann L, Kinchen S, et al. Youth risk behavior surveillance—United States. *MMWR Surveill Summ*. 2011;61:1-162.
8. McCracken KA, Loveless M. Teen pregnancy: an update. *Curr Opin Obstet Gynecol*. 2014;26:355-359.
9. Burstein GR, Eliscu A, Ford K, et al. Expedited partner therapy for adolescents diagnosed with chlamydia or gonorrhea: a position paper of the Society for Adolescent Medicine. *J Adolesc Health*. 2009;45:303-309.
10. Campbell K, Peebles R. Eating disorders in children and adolescents: state of the art review. *Pediatrics*. 2014;134:582-592.
11. Viner RM, Ozer EM, Denny S, et al. Adolescence and the social determinants of health. *Lancet*. 2012;379:1641-1652.

REVIEW QUESTIONS

1. A mother brings her 11-year-old daughter to see her physician because she is worried that her daughter has not started puberty. The physician performs a complete physical examination and assures the mother and her daughter that she has started puberty. What is the first visible sign of puberty among girls?

A. Linear growth spurt
B. Menarche
C. Thelarche, the development of breast buds
D. Pubarche, first appearance of pubic hair
E. Weight gain

Answer: C The first visible sign of puberty among girls is usually thelarche, or the development of breast buds, which occurs on average at 10.5 years in white girls and 1 year earlier in African American girls.

2. A 14-year-old girl comes to see you because she has not started her menstrual period. The following statements about menarche are all true except which one?

A. Menarche occurs 2 to 4 years after the initial appearance of breast buds and pubic hair.
B. The average age of menarche is 12.9 years for white girls.
C. The average age of menarche is 12.2 years for African American girls.
D. Menstrual periods are always regular after the onset of menarche.
E. At menarche, only 20% of cycles are ovulatory.

Answer: D Menarche, or the onset of the first menstrual period, occurs 2 to 4 years after the initial appearance of breast buds and pubic hair. The average age of menarche is 12.9 years for white girls and 12.2 years for African American girls. Menstrual periods are not always regular during the first 2 years after menarche. At menarche, only 20% of cycles are ovulatory; it may take up to 4 years for 80% of cycles to be ovulatory. The average length for the completion of puberty in girls is 4 years (range, 1.5 to 8.0 years).

3. An 18-year-old sexually active woman comes to see her physician for a Papanicolaou smear. Her current partner is her first and only sexual partner, and she reports that she and her partner are monogamous. She is taking the oral contraceptive pill and uses the condom every time she has sexual intercourse. In reviewing the most recent American guidelines for cervical cancer recommendations, you realize that performing a Papanicolaou smear does not apply to which of the following circumstances?

A. Young women with symptoms of cervical cancer
B. Women with previous abnormal test results on cervical screening
C. Women who are immunosuppressed by HIV, organ transplantation, chemotherapy, or chronic use of corticosteroids
D. Sexually active women who are practicing safe sex

Answer: D The most recent American guidelines for cervical cancer screening recommend against screening of women younger than 21 years. These recommendations do not apply to young women with symptoms of cervical cancer or previous abnormal test results on cervical screening or to women who are immunosuppressed by HIV, organ transplantation, chemotherapy, or chronic use of corticosteroids. (Saslow D, Solomon D, Lawson HW, et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *CA Cancer J Clin.* 2012;62:147-172.)

4. You are working at a university health center caring for students. You learn that this semester a 20-year-old student died of testicular cancer (student A), a 23-year-old student committed suicide (student B), and a 19-year-old student was killed in a car accident (student C). Which of these three students fall in the top three causes of death in young adults?

A. Students A and B (cancer and suicide)
B. Students B and C (suicide and car accident)
C. Students A and C (cancer and car accident)
D. All three students (cancer, suicide, and car accident)
E. None of the students

Answer: B The top three causes of death in adolescents and young adults are unintentional car accidents (most from motor vehicle accidents), suicides, and homicides. These three cause more than 71% of deaths in adolescents and young adults. Cancer deaths in this age group are a distant fourth cause of mortality (see Table 17-1). (Hoyert DL, Xu J. Deaths: preliminary data for 2011. *National Vital Statistics Reports*. Hyattsville, MD: National Center for Health Statistics; 2012.)

18

IMMUNIZATION

RAYMOND A. STRIKAS AND WALTER A. ORENSTEIN

Immunization is one of the most cost-effective means of preventing morbidity and mortality from infectious diseases. Routine immunization, particularly of children, has resulted in decreases of 90% or more in reported cases of measles, mumps, rubella, congenital rubella syndrome, poliomyelitis, tetanus, invasive *Haemophilus influenzae* type b, varicella, hepatitis A, and diphtheria. In many circumstances, immunization in children and adults not only prevents morbidity and mortality but also reduces health care costs in the long run.

GENERAL CHARACTERISTICS OF IMMUNIZATIONS

Immunization protects against disease or the sequelae of disease through the administration of an immunobiologic—a vaccine, toxoid, immune globulin preparation, or antitoxin. Protection induced by immunization can be active or passive.[1]

Active Immunization

Administration of a vaccine or toxoid causes the body to produce an immune response against the infectious agent or its toxins. Vaccines consist of suspensions of live (usually attenuated) or inactivated microorganisms or fractions thereof. Toxoids are modified bacterial toxins that retain immunogenic properties but lack toxicity. Active immunization generally results in long-term immunity, although the onset of protection may be delayed because it takes time for the body to respond. With live attenuated vaccines, small quantities of living organisms multiply within the recipient until an immune response cuts off replication. In most recipients, a single dose of a live vaccine generally induces a long-term immune response that closely parallels natural infection. In contrast, inactivated vaccines and toxoids contain large quantities of antigens. Killed (inactivated) vaccines often require multiple doses.

Passive Immunization

Passive immunization with use of immune globulins or antitoxins delivers preformed antibodies to provide temporary immunity. Immune globulins obtained from human blood may contain antibodies to a variety of agents, depending on the pool of human plasma from which they are prepared. Specific immune globulins are made from the plasma of donors with high levels of antibodies to specific antigens (such as tetanus immune globulin). Most immune globulins must be injected intramuscularly, although intravenous and subcutaneous preparations are also available. Antitoxins are solutions of antibodies derived from animals immunized with specific antigens (e.g., diphtheria antitoxin). Passive immunization usually is indicated to protect individuals immediately before anticipated exposure or shortly after known or suspected exposure to an infectious agent (Table 18-1), when active immunization either is not possible or has not been adequate.

Route and Timing of Vaccination

Each immunobiologic has a preferred site and route of administration. In adults, vaccines containing adjuvants should be injected intramuscularly, preferably in the deltoid muscle. For most adults, intramuscular injections should be administered with a 1- to 1½-inch, 22- to 25-gauge needle. Use of the buttocks is discouraged except when large volumes are required because of the potential for damage to the sciatic nerve and because of diminished immune response to some vaccines (such as hepatitis B), probably because vaccine is injected into fat rather than into muscle. Subcutaneous vaccines are usually administered in the triceps area. In general, inactivated vaccines and toxoids can be given at the same visit at different sites. Live and inactivated vaccines usually can be administered at the same time. For example, measles, mumps, and rubella (MMR) vaccine can be administered at the same time as inactivated poliovirus vaccine and live attenuated varicella vaccine. In general, injected and intranasally administered live vaccines not delivered on the same day should be separated by at least 4 weeks to avoid interference by the second vaccine of viral replication and immunity induced by the first vaccine. Orally administered live vaccines (such as oral typhoid vaccine) can be administered

TABLE 18-1 PASSIVE IMMUNIZATIONS FOR ADULTS

DISEASE	NAME OF MATERIAL	COMMENTS AND USE
Tetanus	Tetanus immune globulin, human	Management of tetanus-prone wounds in persons without adequate prior active immunization and treatment of tetanus
Cytomegalovirus	Cytomegalovirus immune globulin, intravenous	Prophylaxis for bone marrow and kidney transplant recipients
Diphtheria	Diphtheria antitoxin, equine	Treatment of established disease; high frequency of reactions to serum of nonhuman origin; in the United States, available only from CDC
Rabies	Rabies immune globulin, human	Postexposure prophylaxis of animal bites
Measles	Immune globulin, human	Prevention or modification of disease in exposed persons, not for control of outbreaks; particularly indicated for unvaccinated infants aged <12 months, pregnant women without evidence of measles immunity, and severely immunocompromised persons
Hepatitis A	Immune globulin, human	Pre-exposure and postexposure prophylaxis for travelers and others who need protection before immunity can be achieved with hepatitis A vaccine
Hepatitis B	Hepatitis B immune globulin, human	Prophylaxis for needle stick or mucous membrane contact with HBsAg-positive persons, for sexual partners with acute hepatitis B or carriers of hepatitis B, for infants born to mothers who are carriers of HBsAg, for infants whose mother or primary caregiver has acute hepatitis B
Varicella	Varicella-zoster immune globulin (VariZIG)	Persons with underlying disease and at risk for complications from chickenpox who have not had varicella or varicella vaccine and who are exposed to varicella; may be given up to 10 days after exposure to known susceptible adults, particularly if antibody negative; VariZIG is available under IND
Vaccinia	Vaccinia immune globulin	Treatment of eczema vaccinatum, vaccinia necrosum, and severe inadvertent inoculations such as ocular vaccinia after vaccinia (smallpox) vaccination; available only from CDC
Erythroblastosis fetalis	Rh immune globulin	Rh-negative women who give birth to Rh-positive infants or who abort
Hypogammaglobulinemia	Immune globulin, intravenous, subcutaneous	Maintenance therapy
Idiopathic thrombocytopenic purpura	Immune globulin, intravenous	Therapy for acute episodes
Botulism	Heptavalent A, B, C, D, E, F, E, G antitoxin, equine	Treatment of botulism; available through CDC
	Botulism immune globulin for infants	Treatment of botulism in infants
Snakebite	Antivenin, equine (North American coral snake antivenin)	Specific for North American coral snake, *Micrurus fulvius*
Crotalidae, polyvalent	Effective for viper and pit viper bites, including rattlesnakes, copperheads, moccasins	
Spider bite	Antivenin, equine	Specific for black widow spider, *Latrodectus mactans,* and other members of the genus
Scorpion	Fab fragments, equine	Specific for *Centruroides* genus scorpions

CDC = Centers for Disease Control and Prevention; HBsAg = hepatitis B surface antigen; IND = Investigational New Drug.

at any interval before or after live injected or intranasal vaccines. Immune globulin may interfere with the replication of injected live vaccine viruses; ideally, most live vaccines should be administered at least 2 weeks before or 3 to 11 months after immune globulin. Immune globulin does not interfere with the response to yellow fever vaccine and is not believed to interfere with orally or intranasally administered live virus vaccines.

Adverse Reactions

No vaccine is completely safe or completely effective. Adverse reactions fall into three general categories: local, systemic, and allergic. Local reactions are generally the most frequent but least severe. Systemic adverse reactions include fever, malaise, myalgias (muscle pain), headache, and loss of appetite. These symptoms, which are common and nonspecific, may occur in vaccinated persons because of the vaccine or because of something unrelated to the vaccine. Allergic reactions, which are the least frequent but most severe, may be caused by the vaccine antigen itself or by another component of the vaccine, such as the cell culture material, stabilizer, preservative, or antibiotic used to inhibit bacterial growth. Life-threatening allergic reactions occur at a rate of about one per million doses. The risk of an allergic reaction can be minimized by good screening before vaccination. All providers who administer vaccines must have an emergency protocol and supplies to treat anaphylaxis.

Antipyretics should not be administered routinely before or at the time of vaccination because they tend to reduce the immune response,■ but they can be used for the treatment of fever and local discomfort that might occur after vaccination. The egg protein contained in vaccines grown in chicken eggs (influenza and yellow fever vaccines) may cause reactions in persons severely allergic to eggs. In general, persons without anaphylactic-type allergies to eggs can be given these vaccines safely, but persons with anaphylactic reactions to eggs generally should not receive these vaccines except when it is absolutely necessary, and then only under established protocols by physicians who have expertise in such situations. Although measles and mumps vaccines are grown in chick embryo tissue culture, the risk of anaphylaxis even in persons with severe hypersensitivity to eggs is low, so they can be vaccinated without prior testing.

Suspected adverse events temporally related to vaccinations should be reported to the Vaccine Adverse Event Reporting System (at *www.vaers.hhs.gov*). The National Vaccine Injury Compensation Program was established in the 1980's to compensate individuals who experience certain health events after vaccination on a "no-fault" basis. No-fault means that persons filing claims are not required to prove negligence on the part of either the health care provider or manufacturer to receive compensation. The program covers all routinely recommended childhood vaccinations, although adults who receive a covered vaccine may also file a claim. Claims may be based on a Vaccine Injury Table (see *http://www.hrsa.gov/vaccinecompensation/vaccinetable.html*), which lists the adverse events associated with vaccines and is updated periodically.

General Considerations

The major group that makes comprehensive, detailed recommendations regarding immunization of adults is the Advisory Committee on Immunization Practices of the Centers for Disease Control and Prevention, which

publishes its information in *Morbidity and Mortality Weekly Report* (also available at http://www.cdc.gov/vaccines/schedules/hcp/adult.html). Immunizations for adults depend on age, lifestyle, occupation, and medical conditions. Two adult immunization schedules are available, one based on age group (Fig. 18-1 and E-Table 18-1) and one based on underlying risk (Fig. 18-2 and Table 18-2).[1,2] All adults who have not received a primary series of diphtheria-tetanus-pertussis–containing vaccine as children should receive one dose of combined tetanus and diphtheria-pertussis vaccine (Tdap), followed by a second dose of tetanus and diphtheria toxoids (Td) 4 weeks later, a third Td dose 6 to 12 months later, and a Td booster every 10 years.[3] All adults (and especially health care personnel) who received a diphtheria-tetanus-pertussis (DTaP) vaccine series as children should have a one-time dose of Tdap, followed by Td every 10 years. Persons born in or after 1957 should have evidence of immunity to measles, mumps, and rubella (e.g., documentation of vaccination or presence of antibodies considered compatible with protection). Adults without evidence of immunity to varicella (documentation of age-appropriate vaccination with two doses of varicella vaccine; laboratory evidence of immunity or confirmation of disease; birth in the United States before 1980, except for health care personnel, pregnant women, or immunocompromised persons; or diagnosis or verification of a history of varicella or herpes zoster disease by a health care provider) should receive varicella vaccine.

Pneumococcal polysaccharide vaccine (PPSV23) is indicated for all adults 65 years and older and for younger adults with certain medical conditions that place them at high risk of complications (see Table 18-2). All adults aged 65 years and older and persons 19 years and older with immunocompromising conditions (including chronic renal failure and nephrotic syndrome), functional or anatomic asplenia, cerebrospinal fluid leaks, or cochlear implants should receive a single dose of pneumococcal conjugate vaccine, ideally before receiving pneumococcal polysaccharide vaccine. Adults 65 years and older should receive pneumococcal polysaccharide vaccine 6 to 12 months later. Adults with immunocompromising conditions should also receive a dose of PPSV23 at least 8 weeks later. Immunocompromised adults who already received PPSV23 should still receive pneumococcal conjugate vaccine 1 year or more after the PPSV23 dose; they should be revaccinated with PPSV23 once after pneumococcal conjugate vaccine when 5 years have passed since their first PPSV23 dose.

Influenza vaccination is recommended annually for all persons 6 months of age or older, including all health care personnel. Health care workers exposed to blood or blood products should receive hepatitis B vaccine. Health care workers should also be immune to measles, mumps, rubella, and varicella.

Text continued on page 73

Recommended Adult Immunization Schedule — United States–2014

Note: These recommendations *must* be read with the footnotes that follow containing number of doses, intervals between doses, and other important information.

Recommended adult immunization schedule, by vaccine and age group†

Vaccine ↓ / Age Group →	19–21 years	22–26 years	27–49 years	50–59 years	60–64 years	≥65 years
Influenza †*	1 dose annually					
Tetanus, diphtheria, pertussis (Td/Tdap) †*	Substitute 1-time dose of Tdap for Td booster; then boost with Td every 10 yrs					
Varicella †*	2 doses					
Human papillomavirus (HPV) Female †*	3 doses					
Human papillomavirus (HPV) Male †*	3 doses	3 doses				
Zoster †					1 dose	
Measles, mumps, rubella (MMR) †*	1 or 2 doses					
Pneumococcal 13-valent conjugate (PCV13) †*	1 dose					
Pneumococcal polysaccharide (PPSV23) †	1 or 2 doses					1 dose
Meningococcal †*	1 or more doses					
Hepatitis A †*	2 doses					
Hepatitis B †*	3 doses					
Haemophilus influenzae type b (Hib) †*	1 or 3 doses					

For all persons in this category who meet the age requirements and who lack documentation of vaccination or have no evidence of previous infection; zoster vaccine recommended regardless of prior episode of zoster

Recommended if some other risk factor is present (e.g., on the basis of medical, occupational, lifestyle, or other indications)

No recommendation

*Covered by the Vaccine Injury Compensation Program

†See E-Table 18-1 for additional information.

Report all clinically significant postvaccination reactions to the Vaccine Adverse Event Reporting System (VAERS). Reporting forms and instructions on filing a VAERS report are available at http://www.vaers.hhs.gov or by telephone, 800-822-7967.

Information on how to file a Vaccine Injury Compensation Program claim is available at www.hrsa.gov/vaccinecompensation or by telephone, 800-338-2382. To file a claim for vaccine injury, contact the U.S. Court of Federal Claims, 717 Madison Place, N.W., Washington, D.C. 20005; telephone 202-357-6400.

Additional information about the vaccines in this schedule, extent of available data, and contraindications for vaccination are also available at http://www.cdc.gov/vaccines, or from the CDC-INFO Contact Center at 800-CDC-INFO (800-232-4636) in English or Spanish, 8:00 a.m. – 8:00 p.m. Eastern Time, Monday – Friday, excluding holidays.

Use of trade-names and commercial sources is for identification only and does not imply endorsement by the U.S. Department of Health and Human Services.

The recommendations in this schedule were approved by the Centers for Disease Control and Preventtion's (CDC) Advisory Committee on Immunization Practices (ACIP), the American Academy of Family Physicians (AAFP), the American College of Physicians (ACP), American College of Obstetricians and Gynecologist (ACOG) and American College of Nurse-Midwives (ACNM).

FIGURE 18-1. Recommended adult immunization schedule by vaccine and age group, United States, 2014. See E-Table 18-1 for footnotes. (Adapted from http://www.cdc.gov/vaccines/schedules/downloads/adult/adult-combined-schedule.pdf.)

E-TABLE 18-1 FOOTNOTES RECOMMENDED IMMUNIZATION SCHEDULE FOR ADULTS AGED 19 YEARS OR OLDER: UNITED STATES, 2014

1. **Additional information**
   - Additional guidance for the use of the vaccines described in this supplement is available at www.cdc.gov/vaccines/hcp/acip-recs/index.html.
   - Information on vaccination recommendations when vaccination status is unknown and other general immunization information can be found in the General Recommendations on Immunization at www.cdc.gov/mmwr/preview/mmwrhtml/rr6002a1.htm.
   - Information on travel vaccine requirements and recommendations (e.g., for hepatitis A and B, meningococcal, and other vaccines) is available at http://wwwnc.cdc.gov/travel/destinations/list.
   - Additional information and resources regarding vaccination of pregnant women can be found at http://wwwnc.cdc.gov/travel/destinations/list.
2. **Influenza vaccination**
   - Annual vaccination against influenza is recommended for all persons aged 6 months or older.
   - Persons aged 6 months or older, including pregnant women and persons with hives-only allergy to eggs, can receive the inactivated influenza vaccine (IIV). An age-appropriate IIV formulation should be used.
   - Adults aged 18 to 49 years can receive the recombinant influenza vaccine (RIV) (FluBlok). RIV does not contain any egg protein.
   - Healthy, nonpregnant persons aged 2 to 49 years without high-risk medical conditions can receive either intranasally administered live, attenuated influenza vaccine (LAIV) (FluMist), or IIV. Health care personnel who care for severely immunocompromised persons (i.e., those who require care in a protected environment) should receive IIV or RIV rather than LAIV.
   - The intramuscularly or intradermally administered IIV are options for adults aged 18 to 64 years.
   - Adults aged 65 years or older can receive the standard-dose IIV or the high-dose IIV (Fluzone High-Dose).
3. **Tetanus, diphtheria, and acellular pertussis (Td/Tdap) vaccination**
   - Administer 1 dose of Tdap vaccine to pregnant women during each pregnancy (preferred during 27 to 36 weeks' gestation) regardless of interval since prior Td or Tdap vaccination.
   - Persons aged 11 years or older who have not received Tdap vaccine or for whom vaccine status is unknown should receive a dose of Tdap followed by tetanus and diphtheria toxoids (Td) booster doses every 10 years thereafter. Tdap can be administered regardless of interval since the most recent tetanus or diphtheria-toxoid containing vaccine.
   - Adults with an unknown or incomplete history of completing a 3-dose primary vaccination series with Td-containing vaccines should begin or complete a primary vaccination series including a Tdap dose.
   - For unvaccinated adults, administer the first 2 doses at least 4 weeks apart and the third dose 6 to 12 months after the second.
   - For incompletely vaccinated (i.e., less than 3 doses) adults, administer remaining doses.
   - Refer to the ACIP statement for recommendations for administering Td/ Tdap as prophylaxis in wound management (see footnote 1).
4. **Varicella vaccination**
   - All adults without evidence of immunity to varicella (as defined below) should receive 2 doses of single-antigen varicella vaccine or a second dose if they have received only 1 dose.
   - Vaccination should be emphasized for those who have close contact with persons at high risk for severe disease (e.g., health care personnel and family contacts of persons with immunocompromising conditions) or are at high risk for exposure or transmission (e.g., teachers; child care employees; residents and staff members of institutional settings, including correctional institutions; college students; military personnel; adolescents and adults living in households with children; nonpregnant women of childbearing age; and international travelers).
   - Pregnant women should be assessed for evidence of varicella immunity. Women who do not have evidence of immunity should receive the first dose of varicella vaccine upon completion or termination of pregnancy and before discharge from the health care facility. The second dose should be administered 4 to 8 weeks after the first dose.
   - Evidence of immunity to varicella in adults includes any of the following:
     - documentation of 2 doses of varicella vaccine at least 4 weeks apart;
     - U.S.-born before 1980, except health care personnel and pregnant women;
     - history of varicella based on diagnosis or verification of varicella disease by a health care provider;
     - history of herpes zoster based on diagnosis or verification of herpes zoster disease by a health care provider; or
     - laboratory evidence of immunity or laboratory confirmation of disease.
5. **Human papillomavirus (HPV) vaccination**
   - Two vaccines are licensed for use in females, bivalent HPV vaccine (HPV2) and quadrivalent HPV vaccine (HPV4), and one HPV vaccine for use in males (HPV4).
   - For females, either HPV4 or HPV2 is recommended in a 3-dose series for routine vaccination at age 11 or 12 years and for those aged 13 through 26 years, if not previously vaccinated.
   - For males, HPV4 is recommended in a 3-dose series for routine vaccination at age 11 or 12 years and for those aged 13 through 21 years, if not previously vaccinated. Males aged 22 through 26 years may be vaccinated.
   - HPV4 is recommended for men who have sex with men through age 26 years for those who did not get any or all doses when they were younger.
   - Vaccination is recommended for immunocompromised persons (including those with HIV infection) through age 26 years for those who did not get any or all doses when they were younger.
   - A complete series for either HPV4 or HPV2 consists of 3 doses. The second dose should be administered 4 to 8 weeks (minimum interval of 4 weeks) after the first dose; thethird dose should be administered 24 weeks after the first dose and 16 weeks after the second dose (minimum interval of at least 12 weeks).
   - HPV vaccines are not recommended for use in pregnant women. However, pregnancy testing is not needed before vaccination. If a woman is found to be pregnant after initiating the vaccination series, no intervention is needed; the remainder of the 3-dose series should be delayed until completion of pregnancy.
6. **Zoster vaccination**
   - A single dose of zoster vaccine is recommended for adults aged 60 years or older regardless of whether they report a prior episode of herpes zoster. Although the vaccine is licensed by the U.S. Food and Drug Administration for use among and can be administered to persons aged 50 years or older, ACIP recommends that vaccination begin at age 60 years.
   - Persons aged 60 years or older with chronic medical conditions may be vaccinated unless their condition constitutes a contraindication, such as pregnancy or severe immunodeficiency.

E-TABLE 18-1 FOOTNOTES RECOMMENDED IMMUNIZATION SCHEDULE FOR ADULTS AGED 19 YEARS OR OLDER: UNITED STATES, 2014—cont'd

7. **Measles, mumps, rubella (MMR) vaccination**
   - Adults born before 1957 are generally considered immune to measles and mumps. All adults born in 1957 or later should have documentation of 1 or more doses of MMR vaccine unless they have a medical contraindication to the vaccine or laboratory evidence of immunity to each of the three diseases. Documentation of provider-diagnosed disease is not considered acceptable evidence of immunity for measles, mumps, or rubella.

   Measles component:
   - A routine second dose of MMR vaccine, administered a minimum of 28 days after the first dose, is recommended for adults who:
     - — are students in postsecondary educational institutions;
     - — work in a health care facility; or
     - — plan to travel internationally.
   - Persons who received inactivated (killed) measles vaccine or measles vaccine of unknown type during 1963-1967 should be revaccinated with 2 doses of MMR vaccine.

   Mumps component:
   - A routine second dose of MMR vaccine, administered a minimum of 28 days after the first dose, is recommended for adults who:
     - — are students in a postsecondary educational institution;
     - — work in a health care facility; or
     - — plan to travel internationally.
   - Persons vaccinated before 1979 with either killed mumps vaccine or mumps vaccine of unknown type who are at high risk for mumps infection (e.g., persons who are working in a health care facility) should be considered for revaccination with 2 doses of MMR vaccine.

   Rubella component:
   - For women of childbearing age, regardless of birth year, rubella immunity should be determined. If there is no evidence of immunity, women who are not pregnant should be vaccinated. Pregnant women who do not have evidence of immunity should receive MMR vaccine upon completion or termination of pregnancy and before discharge from the health care facility.

   Health care personnel born before 1957:
   - For unvaccinated health care personnel born before 1957 who lack laboratory evidence of measles, mumps, and/or rubella immunity or laboratory confirmation of disease, health care facilities should consider vaccinating personnel with 2 doses of MMR vaccine at the appropriate interval for measles and mumps or 1 dose of MMR vaccine for rubella.
8. **Pneumococcal conjugate (PCV13) vaccination**
   - Adults aged 19 years or older with immunocompromising conditions (including chronic renal failure and nephrotic syndrome), functional or anatomic asplenia, cerebrospinal fluid leaks, or cochlear implants who have not previously received PCV13 or PPSV23 should receive a single dose of PCV13 followed by a dose of PPSV23 at least 8 weeks later.
   - Adults aged 19 years or older with the aforementioned conditions who have previously received 1 or more doses of PPSV23 should receive a dose of PCV13 one or more years after the last PPSV23 dose was received. For adults who require additional doses of PPSV23, the first such dose should be given no sooner than 8 weeks after PCV13 and at least 5 years after the most recent dose of PPSV23.
   - When indicated, PCV13 should be administered to patients who are uncertain of their vaccination status history and have no record of previous vaccination.
   - Although PCV13 is licensed by the U.S. Food and Drug Administration for use among and can be administered to persons aged 50 years or older, ACIP recommends PCV13 for adults aged 19 years or older with the specific medical conditions noted above.
9. **Pneumococcal polysaccharide (PPSV23) vaccination**
   - When PCV13 is also indicated, PCV13 should be given first (see footnote 8).
   - Vaccinate all persons with the following indications:
     - — all adults aged 65 years or older;
     - — adults younger than 65 years with chronic lung disease (including chronic obstructive pulmonary disease, emphysema, and asthma), chronic cardiovascular diseases, diabetes mellitus, chronic renal failure, nephrotic syndrome, chronic liver disease (including cirrhosis), alcoholism, cochlear implants, cerebrospinal fluid leaks, immunocompromising conditions, and functional or anatomic asplenia (e.g., sickle cell disease and other hemoglobinopathies, congenital or acquired asplenia, splenic dysfunction, or splenectomy [if elective splenectomy is planned, vaccinate at least 2 weeks before surgery]);
     - — residents of nursing homes or long-term care facilities; and
     - — adults who smoke cigarettes.
   - Persons with immunocompromising conditions and other selected conditions are recommended to receive PCV13 and PPSV23 vaccines. See footnote 8 for information on timing of PCV13 and PPSV23 vaccinations.
   - Persons with asymptomatic or symptomatic HIV infection should be vaccinated as soon as possible after their diagnosis.
   - When cancer chemotherapy or other immunosuppressive therapy is being considered, the interval between vaccination and initiation of immunosuppressive therapy should be at least 2 weeks. Vaccination during chemotherapy or radiation therapy should be avoided.
   - Routine use of PPSV23 vaccine is not recommended for American Indians/Alaska Natives or other persons younger than 65 years unless they have underlying medical conditions that are PPSV23 indications. However, public health authorities may consider recommending PPSV23 for American Indians/Alaska Natives who are living in areas where the risk for invasive pneumococcal disease is increased.
   - When indicated, PPSV23 vaccine should be administered to patients who are uncertain of their vaccination status and have no record of vaccination.
10. **Revaccination with PPSV23**
    - One-time revaccination 5 years after the first dose of PPSV23 is recommended for persons aged 19 through 64 years with chronic renal failure or nephrotic syndrome, functional or anatomic asplenia (e.g., sickle cell disease or splenectomy), or immunocompromising conditions.
    - Persons who received 1 or 2 doses of PPSV23 before age 65 years for any indication should receive another dose of the vaccine at age 65 years or later if at least 5 years have passed since their previous dose.
    - No further doses of PPSV23 are needed for persons vaccinated with PPSV23 at or after age 65 years.
11. **Meningococcal vaccination**
    - Administer 2 doses of quadrivalent meningococcal conjugate vaccine (MenACWY-D [Menactra]) at least 2 months apart to adults of all ages with functional asplenia or persistent complement component deficiencies. HIV infection is not an indication for routine vaccination with MenACWY-D. If an HIV-infected person of any age is vaccinated, 2 doses of MenACWY-D should be administered at least 2 months apart.
    - Administer a single dose of meningococcal vaccine to microbiologists routinely exposed to isolates of Neisseria meningitidis, military recruits, persons at risk during an outbreak attributable to a vaccine serogroup, and persons who travel to or live in countries in which meningococcal disease is hyperendemic or epidemic.
    - First-year college students up through age 21 years who are living in residence halls should be vaccinated if they have not received a dose on or after their 16th birthday.
    - MenACWY-D is preferred for adults with any of the preceding indications who are aged 55 years or younger as well as for adults aged 56 years or older who a) were vaccinated previously with MenACWY-D and are recommended for revaccination, or b) for whom multiple doses are anticipated. Meningococcal polysaccharide vaccine (MenACWY-CRM [Menveo]) is preferred for adults aged 56 years or older who have not received MenACWY-D previously and who require a single dose only (e.g., travelers).
    - Revaccination with MenACWY-D every 5 years is recommended for adults previously vaccinated with MenACWY-D or MenACWY-CRM who remain at increased risk for infection (e.g., adults with anatomic or functional asplenia, persistent complement component deficiencies, or microbiologists).

E-TABLE 18-1 FOOTNOTES RECOMMENDED IMMUNIZATION SCHEDULE FOR ADULTS AGED 19 YEARS OR OLDER: UNITED STATES, 2014—cont'd

12. **Hepatitis A vaccination**
    - Vaccinate any person seeking protection from hepatitis A virus (HAV) infection and persons with any of the following indications:
        - men who have sex with men and persons who use injection or noninjection illicit drugs;
        - persons working with HAV-infected primates or with HAV in a research laboratory setting;
        - persons with chronic liver disease and persons who receive clotting factor concentrates;
        - persons traveling to or working in countries that have high or intermediate endemicity of hepatitis A; and
        - unvaccinated persons who anticipate close personal contact (e.g., household or regular babysitting) with an international adoptee during the first 60 days after arrival in the United States from a country with high or intermediate endemicity. (See footnote 1 for more information on travel recommendations.) The first dose of the 2-dose hepatitis A vaccine series should be administered as soon as adoption is planned, ideally 2 or more weeks before the arrival of the adoptee.
    - Single-antigen vaccine formulations should be administered in a 2-dose schedule at either 0 and 6 to 12 months (Havrix), or 0 and 6 to 18 months (Vaqta). If the combined hepatitis A and hepatitis B vaccine (Twinrix) is used, administer 3 doses at 0, 1, and 6 months; alternatively, a 4-dose schedule may be used, administered on days 0, 7, and 21 to 30 followed by a booster dose at month 12.
13. **Hepatitis B vaccination**
    - Vaccinate persons with any of the following indications and any person seeking protection from hepatitis B virus (HBV) infection:
        - sexually active persons who are not in a long-term, mutually monogamous relationship (e.g., persons with more than 1 sex partner during the previous 6 months); persons seeking evaluation or treatment for a sexually transmitted disease (STD); current or recent injection drug users; and men who have sex with men;
        - health care personnel and public safety workers who are potentially exposed to blood or other infectious body fluids;
        - persons with diabetes who are younger than age 60 years as soon as feasible after diagnosis; persons with diabetes who are age 60 years or older at the discretion of the treating clinician based on the likelihood of acquiring HBV infection, including the risk posed by an increased need for assisted blood glucose monitoring in long-term care facilities, the likelihood of experiencing chronic sequelae if infected with HBV, and the likelihood of immune response to vaccination;
        - persons with end-stage renal disease, including patients receiving hemodialysis, persons with HIV infection, and persons with chronic liver disease;
        - household contacts and sex partners of hepatitis B surface antigen-positive persons, clients and staff members of institutions for persons with developmental disabilities, and international travelers to countries with high or intermediate prevalence of chronic HBV infection; and
        - all adults in the following settings: STD treatment facilities, HIV testing and treatment facilities, facilities providing drug abuse treatment and prevention services, health care settings targeting services to injection drug users or men who have sex with men, correctional facilities, end-stage renal disease programs and facilities for chronic hemodialysis patients, and institutions and nonresidential day care facilities for persons with developmental disabilities.
    - Administer missing doses to complete a 3-dose series of hepatitis B vaccine to those persons not vaccinated or not completely vaccinated. The second dose should be administered 1 month after the first dose; the third dose should be given at least 2 months after the second dose (and at least 4 months after the first dose). If the combined hepatitis A and hepatitis B vaccine (Twinrix) is used, give 3 doses at 0, 1, and 6 months; alternatively, a 4-dose Twinrix schedule, administered on days 0, 7, and 21 to 30 followed by a booster dose at month 12 may be used.
    - Adult patients receiving hemodialysis or with other immunocompromising conditions should receive 1 dose of 40 mcg/mL (Recombivax HB) administered on a 3-dose schedule at 0, 1, and 6 months or 2 doses of 20 mcg/mL (Engerix-B) administered simultaneously on a 4-dose schedule at 0, 1, 2, and 6 months.
14. ***Haemophilus influenzae type b* (Hib) vaccination**
    - One dose of Hib vaccine should be administered to persons who have functional or anatomic asplenia or sickle cell disease or are undergoing elective splenectomy if they have not previously received Hib vaccine. Hib vaccination 14 or more days before splenectomy is suggested.
    - Recipients of a hematopoietic stem cell transplant should be vaccinated with a 3-dose regimen 6 to 12 months after a successful transplant, regardless of vaccination history; at least 4 weeks should separate doses.
    - Hib vaccine is not recommended for adults with HIV infection since their risk for Hib infection is low.
15. **Immunocompromising conditions**
    - Inactivated vaccines generally are acceptable (e.g., pneumococcal, meningococcal, and inactivated influenza vaccine) and live vaccines generally are avoided in persons with immune deficiencies or immunocompromising conditions. Information on specific conditions is available at http://www.cdc.gov/vaccines/hcp/acip-recs/index.html.

Vaccines That Might Be Indicated for Adults Based on Medical and Other Indications†

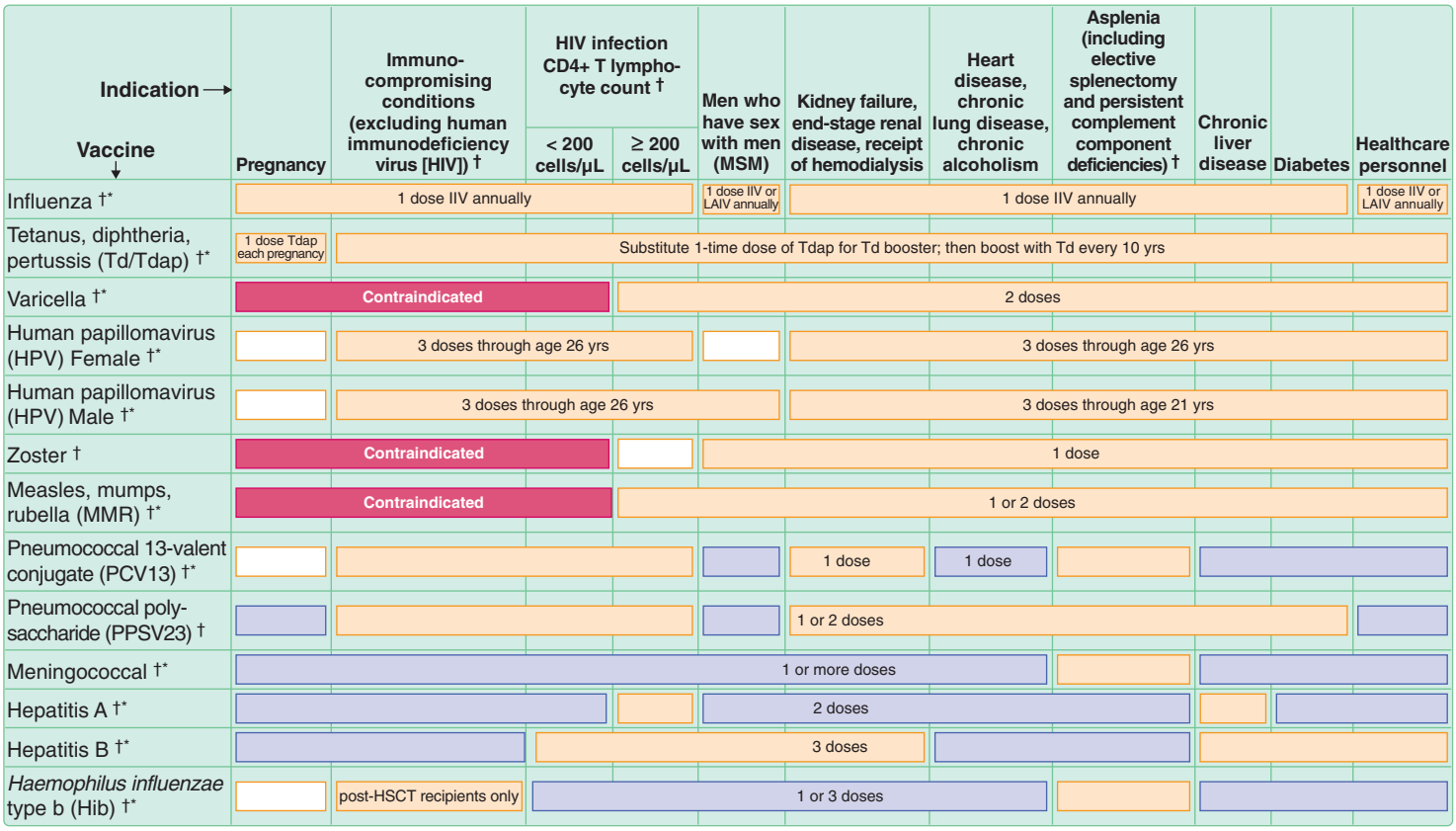

*Covered by the Vaccine Injury Compensation Program †For additional information please see http://www.cdc.gov/vaccines/schedules/downloads/adult/adult-schedule.pdf

These schedules indicate the recommended age groups and medical indications for which administration of currently licensed vaccines is commonly indicated for adults ages 19 years and older, as of Febuary 1, 2014. For all vaccines being recommended on the Adult Immunization Schedule: a vaccine series does not need to be restarted, regardless of the time that has elapsed between doses. Licensed combination vaccines may be used whenever any components of the combination are indicated and when the vaccine's other components are not contraindicated. For detailed recommendations on all vaccines, including those used primarily for travelers or that are issued during the year, consult the manufacturers' package inserts and the complete statements from the Advisory Committee on Immunization Practices (www.cdc.gov/vaccines/hcp/acip-recs/index.html). Use of trade names and commerical sources is for identification only and does not imply endorsement by the U.S. Department of Health and Human Services.

FIGURE 18-2. Recommended adult immunization schedule by vaccine and medical and other indications, United States, 2014. See E-Table 18-1 for footnotes. (Adapted from http://www.cdc.gov/vaccines/schedules/downloads/adult/adult-combined-schedule.pdf.)

TABLE 18-2 SELECTED IMMUNIZING AGENTS INDICATED FOR ADULTS*

DISEASE	IMMUNIZING AGENT	INDICATIONS	SCHEDULE	MAJOR CONTRAINDICATIONS AND PRECAUTIONS	COMMENTS
Anthrax	Anthrax vaccine, adsorbed, an inactivated vaccine	Pre-exposure prophylaxis of persons at high risk of exposure (e.g., military, certain laboratory workers) Consider with antibiotics for postexposure prophylaxis	0.5-mL dose IM at 0, 4, and 6 wk and boosters at 12 and 18 mo, then booster annually thereafter If used after exposure, three doses at 0, 2, and 4 wk with 60 days of antimicrobial therapy; antibiotics should be continued for 14 days after third dose	Severe allergic reaction to a vaccine component or after a prior dose Moderate or severe acute illness is a precaution to vaccination	Effectiveness against aerosol exposure inferred primarily from animal data Limited data on the benefits of postexposure use
Diphtheria	Tetanus and diphtheria toxoids combined	All adults	For incompletely immunized adults, three doses IM needed for primary series: two doses IM 4 wk apart, third dose 6-12 mo after second dose; one of these doses should be Tdap Booster every 10 yr No need to repeat if schedule is interrupted	History of neurologic reaction after a previous dose Severe allergic reaction to a vaccine component or after a prior dose	Tetanus and diphtheria toxoids combined with acellular pertussis vaccine (Tdap) preferred as one-time booster for all persons Moderate or severe acute illness is a precaution

TABLE 18-2 SELECTED IMMUNIZING AGENTS INDICATED FOR ADULTS—cont'd

DISEASE	IMMUNIZING AGENT	INDICATIONS	SCHEDULE	MAJOR CONTRAINDICATIONS AND PRECAUTIONS	COMMENTS
Hepatitis A	Inactivated hepatitis A vaccine	Travelers to highly or intermediately endemic countries Men who have sex with men Illegal drug users (injection and noninjection) Persons who work with hepatitis A virus–infected primates or who do research with the virus Persons with chronic liver disease Recipients of clotting factors	Two doses at least 6 mo apart for persons aged ≥1 yr	Severe allergic reaction to a vaccine component or after a prior dose Moderate or severe acute illness is a precaution to vaccination	Recommended for all children Should be considered for outbreak control
Hepatitis B	Inactivated hepatitis B virus subunit vaccine containing HBsAg	Adolescents Health care and public safety workers potentially exposed to blood Clients and staff of institutions for the developmentally disabled Hemodialysis patients Men who have sex with men Users of illicit injectable drugs Recipients of clotting factors Household and sexual contacts of HBV carriers Inmates of long-term correctional facilities Heterosexuals treated for sexually transmitted diseases or with multiple sexual partners Travelers with close contact for ≥6 mo with populations with high prevalence of HBV carriage Adults 19-59 yr with diabetes mellitus	Three doses IM at 0, 1, and 6 mo	Severe allergic reaction to a vaccine component or after a prior dose Moderate or severe acute illness is a precaution to vaccination	Pregnancy is not a contraindication. Health care workers who have contact with patients or blood, sexual contacts of persons with chronic HBV infection, hemodialysis patients, other immunosuppressed persons, and recipients of clotting factor concentrates should be tested 1-2 mo after vaccination to determine serologic response.
Human papillomavirus	Inactivated L1 capsid proteins of types 6, 11, 16, and 18 (quadrivalent) and types 16 and 18 (bivalent)	Females at 11-12 yr; catch-up vaccination of females through 26 yr Males 11-12 yr (quadrivalent vaccine only) and catch-up vaccination through 21 yr; consider catch-up through 26 yr to prevent anogenital and oropharyngeal cancers and genital warts	Three 0.5-mL doses IM at 0, 1 to 2, and 6 mo	Severe allergic reaction to a vaccine component or to a prior dose Vaccine is not recommended for pregnant women	The vaccine will not protect against existing infections. Because the types in the vaccine are not responsible for about 30% of infections associated with cervical cancer, screening for cancer should occur as for unvaccinated women.
Influenza	Inactivated virus vaccine	All persons ≥6 months of age, with greatest priority for those at higher risk for influenza complications (e.g., ≥65 yr old, persons with underlying medical conditions, pregnant women) or in contact with those at higher risk (e.g., health care workers, and persons with close contact with children <5 yr)	Annual vaccination; see annual ACIP recommendation	Severe allergic reaction to an influenza vaccine component (including eggs) or after a prior dose Recombinant vaccine (RIV) containing no egg protein may be given to persons 18-49 yr with severe egg allergies Moderate or severe acute illness is a precaution to vaccination GBS within 6 wk of prior dose of influenza vaccine	Optimum timing for vaccination is October. However, vaccination can occur throughout the influenza season, particularly for persons at high risk for complications and their contacts who were not vaccinated earlier. Only one dose of influenza vaccine per season is recommended for adults.

TABLE 18-2 SELECTED IMMUNIZING AGENTS INDICATED FOR ADULTS—cont'd

DISEASE	IMMUNIZING AGENT	INDICATIONS	SCHEDULE	MAJOR CONTRAINDICATIONS AND PRECAUTIONS	COMMENTS
Influenza (cont'd)	Live attenuated influenza virus	Persons 2 through 49 yr without underlying conditions that place them at high risk of complications from influenza	Annual vaccination; see annual ACIP statement Administered intranasally	Persons <2 yr or ≥50 yr Persons with underlying disorders that place them at high risk of influenza complications History of GBS within 6 wk of prior dose of influenza vaccine Pregnant women Hypersensitivity to eggs or components of vaccine	Can be used for household contacts and health care workers caring for patients without severe immunocompromise
Japanese encephalitis	Inactivated Japanese encephalitis virus vaccine	Travelers to Asia spending at least 1 mo in endemic areas during transmission season	Two 0.5-mL doses IM on days 0 and 28 for persons 18 yr and older	Pregnancy	
Measles	Live virus vaccine	All adults born after 1956 without history of live vaccine on or after first birthday or detectable measles antibody Persons born before 1957 generally can be considered immune	One dose sufficient for most adults; two doses at least 1 mo apart indicated for persons entering college or medical facility employment, traveling abroad, or at risk of measles during outbreaks	Altered immunity (e.g., leukemia, lymphoma, generalized malignant disease, congenital immunodeficiency, immunosuppressive therapy) Immune globulin or other blood products within prior 3-11 mo, depending on dose of immune globulin or blood product received Untreated tuberculosis Anaphylactic hypersensitivity to neomycin or gelatin Pregnancy Thrombocytopenia	Persons with anaphylactic allergies to eggs may be vaccinated (see text). Vaccine should be administered to persons with asymptomatic HIV infection and should be considered for patients except those with severe immunocompromise.
Meningococcal disease (two vaccines)	1. Meningococcal conjugate vaccines containing polysaccharide of serogroups A, C, W, and Y (age 2 mo–55 yr) 2. Polysaccharide vaccine containing tetravalent A, C, W, and Y (ages 56 yr and older, if never vaccinated and only one vaccination expected to be necessary)	All 11- to 18-yr-old persons and all persons with persistent complement component deficiencies or anatomic or functional asplenia Persons who will travel to areas with hyperendemic or epidemic diseases Certain laboratory workers May be useful during localized outbreaks	One dose with revaccination at age 16 yr of children who receive conjugate vaccine at 11-12 yr of age and every 5 yr for persons at high risk	Allergic reactions to a component of the vaccine, including diphtheria toxoid and latex	Conjugate vaccine is preferred to polysaccharide alone for persons aged 2 mo through 55 yr and for persons 56 yr and older previously vaccinated who continue to be at risk.
Mumps	Live virus vaccine	All adults born after 1956 without history of live vaccine on or after first birthday or detectable mumps antibody Persons born before 1957 generally can be considered immune	One dose sufficient for most adults Two doses at least 1 mo apart indicated for persons entering college or medical facility employment or traveling abroad	Altered immunity (e.g., leukemia, lymphoma, generalized malignant disease, congenital immunodeficiency, immunosuppressive therapy) Immune globulin or other blood products within prior 3-11 mo Anaphylactic hypersensitivity to neomycin or gelatin Pregnancy Thrombocytopenia if administered with measles vaccine	Although persons born after 1956 are generally immune, vaccine can be given to adults of all ages and may be particularly indicated for postpubertal males who are thought to be susceptible. Persons with anaphylactic allergies to eggs may be vaccinated.

TABLE 18-2 SELECTED IMMUNIZING AGENTS INDICATED FOR ADULTS—cont'd

DISEASE	IMMUNIZING AGENT	INDICATIONS	SCHEDULE	MAJOR CONTRAINDICATIONS AND PRECAUTIONS	COMMENTS
Pertussis	Adult preparation of pertussis antigens combined with tetanus and diphtheria toxoids (Tdap)	All 11- to 12-yr-olds Catch-up vaccination for all persons ≥13 yr	One dose Pregnant women: one dose during each pregnancy between 27 and 36 weeks of gestation	Severe allergic reaction to a vaccine component or after a prior dose Moderate or severe acute illness is a precaution to vaccination	Two preparations are available, one licensed for all persons ≥10 yr, one for 10- to 64-yr-olds.
Pneumococcal disease	23-Valent polysaccharide vaccine (PPSV23)	All adults with cardiovascular disease, pulmonary disease (including asthma), diabetes mellitus, alcoholism, cirrhosis, cerebrospinal fluid leaks, splenic dysfunction or anatomic asplenia, Hodgkin disease, lymphoma, multiple myeloma, chronic renal failure, nephrotic syndrome, immunosuppression, HIV infection, cigarette smokers 19 yr and older High-risk populations, such as certain Native Americans, and All adults ≥65 yr	One dose IM or SC; a second dose should be considered ≥5 yr later for adults at high risk of disease (e.g., asplenic patients) and those who lose antibody rapidly (e.g., nephrotic syndrome, renal failure, transplant recipients) Revaccinate adults who received a first dose when <65 yr who are now ≥65 yr and who received their vaccine at least 5 yr earlier	Severe allergic reaction to a vaccine component or after a prior dose Moderate or severe acute illness is a precaution to vaccination	
	Pneumococcal conjugate vaccine (PCV13)	One dose recommended for all adults age 65 years and older, for immunocompromised adults ≥19 yr, and those with functional/anatomic asplenia, cochlear implants, cerebrospinal fluid leaks	One dose if not previously vaccinated with PCV13) as an adult Give ≥1 year after PPSV23; or if PPSV23-naïve, if 65 years or older and not immunocompromised, give PPSV23 6-12 months after PCV13; if >19 yrs old and immunocompromised, give PPSV23 ≥8 wk after PCV13	Severe allergic reaction (e.g., anaphylaxis) to any component of PCV13 or any diphtheria toxoid-containing vaccine Moderate or severe acute illness is a precaution to vaccination	Although licensed for use in adults ≥50 yr, it is recommended only for adults ≥19 yr with the noted medical conditions.
Poliomyelitis	Inactivated poliovirus vaccine (IPV)	Certain adults who are at greater risk of exposure to wild poliovirus than the general population, including travelers to countries where poliomyelitis is epidemic or endemic or specific populations with disease caused by wild poliovirus	For unvaccinated adults, two doses IM or SC 4 wk apart and a third dose 6-12 mo after the second; if <4 wk available before protection is needed, a single dose of IPV For incompletely immunized adults, complete primary series of three doses of IPV or prior oral poliovirus vaccine (OPV); no need to restart interrupted series A single dose of IPV can be given to adults who previously received a primary series but now are at high risk, such as those traveling to an endemic area	On theoretical grounds, pregnant women should not receive IPV, but if immediate protection is needed, IPV can be used Severe allergic reaction to a vaccine component or after a prior dose Moderate or severe acute illness is a precaution to vaccination	
Rabies	Inactivated vaccine, HDCV or PCEC	High-risk persons, including animal handlers, selected laboratory and field workers, and persons traveling for ≥1 mo to areas with high risk of rabies	Pre-exposure prophylaxis: three doses of 1 mL IM on days 0, 7, and 21 or 28	History of severe hypersensitivity reaction	Further doses needed after exposure

TABLE 18-2 SELECTED IMMUNIZING AGENTS INDICATED FOR ADULTS—cont'd

DISEASE	IMMUNIZING AGENT	INDICATIONS	SCHEDULE	MAJOR CONTRAINDICATIONS AND PRECAUTIONS	COMMENTS
Rubella	Live virus vaccine	Adults, particularly women of childbearing age, who lack history of rubella vaccine and detectable rubella-specific antibodies in serum Males and females in institutions where rubella outbreaks may occur, such as hospitals, the military, and colleges Persons born before 1957, except women who can become pregnant, generally can be considered immune	One dose SC	Pregnancy, altered immunity (e.g., leukemia, lymphoma, generalized malignant disease, congenital immunodeficiency, immunosuppressive therapy) Immune globulin or other blood products within 3-11 mo before vaccination Anaphylactic hypersensitivity to neomycin Administration of blood products should not contraindicate postpartum vaccination Thrombocytopenia if administered with measles vaccine	Women should be counseled to avoid pregnancy for 1 mo after vaccination.
Smallpox	Live vaccinia virus	Persons working with orthopox viruses Members of public health and health care response teams	One dose intracutaneously with a bifurcated needle Boosters every 10 yr and perhaps every 3 yr for persons working with virulent orthopox viruses	History or presence of eczema or other acute, chronic, or exfoliative skin condition in patient or a close household or personal contact Immunosuppression or pregnancy in patient or a close household or personal contact History of heart disease Breast-feeding Age <1 yr Allergy to a vaccine component No contraindications if exposed to smallpox	Some complications of vaccination are treatable with vaccinia immune globulin. Vaccine is effective 3-4 days after exposure to variola and perhaps longer to prevent or to modify the illness. Serious adverse events are rare but significant, including eczema vaccinatum, progressive vaccinia, myopericarditis, autoinoculation, and encephalitis. Vaccinia is transmissible.
Tetanus	Tetanus and diphtheria toxoids combined	All adults	Three doses IM needed for primary series: two doses 4 wk apart, third dose 6-12 mo after second dose Booster every 10 yr; no need to repeat if schedule is interrupted	History of neurologic or severe allergic reaction after a prior dose	Special recommendations for wound treatment (see text) Persons with GBS within the first 6 wk after immunization, particularly adults who received a prior primary series, probably should not be revaccinated in most circumstances. Tetanus and diphtheria toxoids combined with acellular pertussis (Tdap) vaccine preferred for booster at age 11-12 yr One-time booster of Tdap for all adults
Typhoid fever	Vi capsular polysaccharide vaccine Live attenuated Ty21a oral vaccine	Travelers to areas where the risk of prolonged exposure to contaminated food and water is high May be considered for family and intimate contacts of carriers and laboratory workers who work with *Salmonella typhi*	Vi polysaccharide vaccine: one dose IM 0.5 mL; boosters every 2 yr Oral vaccine: four doses on alternate days; repeat series every 5 yr if risk continues	Severe local or systemic reaction to a prior dose Ty21a vaccine should not be administered to persons with altered immunity or those receiving antimicrobial agents	Efficacy only 50-77% Food and water precautions essential

TABLE 18-2 SELECTED IMMUNIZING AGENTS INDICATED FOR ADULTS—cont'd

DISEASE	IMMUNIZING AGENT	INDICATIONS	SCHEDULE	MAJOR CONTRAINDICATIONS AND PRECAUTIONS	COMMENTS
Varicella: chickenpox strain	Attenuated Oka strain of varicella virus	All persons without evidence of varicella immunity, especially health care personnel, childbearing-age women, and persons with household or other contact with persons at high risk of complications of varicella (e.g., susceptible immunosuppressed persons)	Two 0.5-mL doses SC 4-8 wk apart for persons ≥13 yr A second dose is recommended for all persons who previously received one dose	Immunocompromise Pregnancy Allergy to vaccine components	Adults with a history of prior clinician-diagnosed or verified varicella can be considered immune. Vaccine virus has rarely been transmitted to contacts from healthy vaccinees in whom rash developed. Women who receive vaccine should not become pregnant for 1 mo.
Varicella: zoster	Attenuated Oka strain of varicella virus, approximately 14 times more potent than varicella vaccine	Persons ≥60 years of age	One 0.65-mL dose SC	Immunocompromise Pregnancy Allergy to vaccine components	May be administered regardless of a prior history of shingles
Yellow fever	Live attenuated virus (17D strain)	Persons living or traveling in areas where yellow fever exists	One dose; booster every 10 yr	Immunocompromised persons History of anaphylactic allergies to eggs Pregnancy on theoretical grounds, although may be given if risk is high	Fever, jaundice, and multiple-organ system failure (viscerotropic disease) have been rarely reported in first-time recipients of 17D-derived yellow fever vaccinations. Vaccinate only persons traveling to areas endemic to yellow fever.

*See the text and package inserts for further details, particularly regarding indications, dosage, administration, side effects, and adverse reactions and contraindications.
ACIP = Advisory Committee on Immunization Practices; GBS = Guillain-Barré syndrome; HBsAg = hepatitis B surface antigen; HBV = hepatitis B virus; HDCV = human diploid cell vaccine for rabies; HIV = human immunodeficiency virus; IM = intramuscularly; IPV = inactivated poliovirus vaccine; MMR = measles, mumps, and rubella vaccine; OPV = live trivalent oral poliovirus vaccine; PCEC = purified chick embryo cell culture rabies vaccine; SC = subcutaneously.

Immunocompromised Persons

Patients with conditions that compromise their immune systems—immunodeficiency diseases, leukemia, lymphoma, and generalized malignant disease, and those who are immunosuppressed from therapy with corticosteroids, alkylating agents, antimetabolites, and radiation—generally should not receive live attenuated vaccines. An exception is infection with human immunodeficiency virus (HIV). Two doses of MMR vaccine are recommended for all persons aged 12 months and older who have HIV infection and do not have evidence of immunity to measles, rubella, and mumps and who are not severely immunosuppressed (i.e., a CD4 percentage ≥15% and $CD4^+$ lymphocyte counts ≥200 cells/μL for ≥6 months for persons aged >5 years). Varicella vaccination (two doses, 3 months apart) may be considered in HIV-infected persons with $CD4^+$ T-lymphocyte counts above 200 cells/μL. Patients with leukemia in remission who have not been receiving any chemotherapy for at least 3 months may receive live virus vaccines. Short-course therapy (<2 weeks) with corticosteroids, alternate-day regimens with low to moderate doses of short-acting corticosteroids, and topical applications or tendon injections are not ordinarily contraindications to the administration of live vaccines.

Immunocompromised patients can receive inactivated vaccines and toxoids, although the efficacy of such preparations may be diminished. Patients with known HIV infection should receive pneumococcal vaccine and annual influenza vaccination.

Pregnancy

In general, live vaccines should not be given to pregnant women because of the theoretical concern that the vaccines could adversely affect the fetus. No significant adverse events have been documented as attributable to vaccination of pregnant women with rubella-containing or varicella vaccines; nevertheless, pregnant women should not receive MMR or varicella vaccine, and women who do receive these vaccines should wait 1 month before becoming pregnant. Poliomyelitis and yellow fever vaccines usually should not be given to pregnant women unless the risk of disease is substantial. Tdap vaccination is indicated for pregnant women during each pregnancy, preferably between 27 and 36 weeks of gestation, to prevent pertussis in their infants and themselves. The safety of hepatitis A vaccination during pregnancy has not been determined, so the risk should be weighed against the benefit. All pregnant women should be screened for hepatitis B surface antigen (HBsAg). If HBsAg is positive, their children should receive hepatitis B vaccine and hepatitis B immune globulin within 12 hours of birth. All women who are or will be pregnant during the influenza season should receive inactivated influenza vaccine to protect themselves and their babies.

INDIVIDUAL IMMUNOBIOLOGICS

Hepatitis A

Two inactivated hepatitis A (Chapter 148) vaccines are available in the United States. Seroconversion rates after a single dose of either vaccine in persons older than 1 year exceed 95%, and protection is expected to persist for at least 25 years.

Indications

For adults, the vaccine is indicated primarily for persons traveling to countries (generally those in the developing world) with high or intermediate

endemicity for hepatitis A, but it is also recommended for other groups at high risk for infection or for development of severe hepatitis. In addition, vaccine is routinely recommended for children 12 to 23 months of age. Vaccination also is recommended for persons aged 1 to 40 years for postexposure prophylaxis after close contact with an infected person or exposure to a contaminated food or water source. Health care workers have not been shown to be at higher risk than the general population for hepatitis A and do not need routine immunization. Although food handlers are not at increased risk for hepatitis A compared with the general population, the consequences of infection or suspected infection in this group, which can lead to extensive public health investigations, may make vaccination cost-effective in some settings. Doses vary by age and product. All schedules call for a second dose at least 6 months after the first dose.

Adverse Events

The most common adverse reaction after hepatitis A vaccination is tenderness and soreness at the injection site. Although rare and more serious adverse events have been reported in temporal association with vaccination, no causal relationships have been established.

Hepatitis B

Hepatitis B (Chapter 148) vaccine was the first vaccine known to prevent cancer. It also can prevent acute and chronic complications of hepatitis B, including an estimated 1800 deaths annually in the United States from liver cancer, cirrhosis, and fulminant hepatic disease. Currently produced vaccines are derived from insertion of the gene for HBsAg into *Saccharomyces cerevisiae.* When it is administered in a three-dose series, hepatitis B vaccine produces adequate antibody responses (anti-HBs ≥ 10 IU/L) in more than 90% of healthy adults younger than 40 years and in more than 95% of normal infants, children, and adolescents. By age 60 years, protective levels of antibody develop in only 75% of vaccinated persons. The duration of vaccine-conferred immunity is not known, although follow-up of vaccinees for more than 20 years indicates persistence of protection against clinically significant infections (i.e., detectable viremia and clinical disease). Booster doses are not currently recommended. Vaccine must be injected intramuscularly in the deltoid.

Indications

Hepatitis B vaccine is indicated for adults at increased risk of infection (see Fig. 18-1 and E-Table 18-1), including all persons 19 through 59 years of age with diabetes. Vaccination may be offered at the physician's discretion to persons 60 years of age or older with diabetes. Universal screening for HBsAg is recommended for all pregnant women; three doses of vaccine and one dose of hepatitis B immune globulin are recommended for infants of acutely or chronically infected mothers.

Adverse Events

The major adverse reaction is soreness at the injection site. Rare instances of Guillain-Barré syndrome, leukoencephalitis, optic neuritis, transverse myelitis, rheumatoid arthritis, type 1 diabetes, and autoimmune disease have been reported, but causal associations have not been confirmed with any systemic immune complications.

Human Papillomavirus

Two licensed human papillomavirus (HPV) vaccines contain the L1 capsid protein of types 16 and 18, which account for about 70% of cases of cervical cancer. The quadrivalent vaccine (HPV4), which also contains the L1 capsid protein of types 6 and 11 that are the most common causes of anogenital warts, is the only HPV vaccine licensed for boys and men.[4] Routine vaccination of 11- to 12-year-old girls and boys is recommended in a three-dose schedule at 0, 1 to 2, and 6 months. Catch-up vaccination should be undertaken for females through 26 years of age, all males through 21 years of age, and immunocompromised males and men who have sex with men through age 26 years. Catch-up vaccination may also be considered for other men 22 to 26 years of age. Women with a prior abnormal Papanicolaou smear and persons with genital warts should be vaccinated to prevent persistent infection with types of HPV they may not yet have acquired.

Adverse Events

The most commonly reported local symptoms are injection site pain, redness, and swelling. The most common generalized symptoms are dizziness, syncope, nausea, vomiting, fatigue, headache, fever, and urticaria. Anaphylaxis is very rare.

Influenza

Influenza vaccines for seasonal influenza include intramuscular and intradermal inactivated influenza vaccine, which may contain three or four influenza split or subvirion virus types—A(H3N2), A(H1N1), and one or two B strains—and intranasal live attenuated influenza vaccine. Live attenuated vaccine, which consists of four cold-adapted, temperature-sensitive attenuated viruses, one for each of the expected circulating strains, uses viruses that have been reassorted with circulating strains to contain six internal genes from the parent virus and genes for the surface hemagglutinin and neuraminidase of an A(H3N2), A(H1N1), and two B strains.

A recombinant inactivated influenza vaccine contains no egg protein and may be given to persons 18 to 49 years of age who experience hives after eating eggs; the alternative is to give the standard, egg-derived, inactivated vaccine followed by 30 minutes of observation. Persons with more severe symptoms after eating eggs may either receive the recombinant vaccine or be referred to a physician with expertise in management of allergic conditions.

Indications

Annual influenza (Chapter 364) vaccination is indicated for everyone 6 months of age and older, but especially persons at high risk of complications from the disease: all children aged 6 through 59 months; all persons aged 50 years and older; adults and children who have chronic pulmonary, cardiovascular, renal, hepatic, neurologic, hematologic, or metabolic disorders; persons who are immunosuppressed; women who are or will be pregnant during the influenza season; children and adolescents who are receiving long-term aspirin therapy and who might be at risk for experiencing Reye syndrome after influenza virus infection; residents of long-term care facilities; American Indians/Alaska Natives; and persons who are morbidly obese. To reduce transmission of influenza to high-risk patients, health care personnel[5] and household contacts of high-risk patients, including contacts of children younger than 5 years, also should be vaccinated annually.

The efficacy of inactivated influenza vaccine varies with the host's condition and the degree to which antigens in the vaccine match viruses that circulate during the following season. Provided the match is good, the vaccine's efficacy is usually 50% to 70% for healthy adults younger than 65 years. Effectiveness varies but averages about 60% in preventing laboratory-confirmed outpatient illness in persons 50 years and older but is only about 35% in the institutionalized elderly.[6] Live attenuated vaccine is licensed only for nonpregnant persons aged 2 through 49 years without underlying conditions that place them at high risk of complications from influenza. It also can be used for appropriately aged contacts of high-risk patients but is not recommended for contacts of severely immunosuppressed patients, such as patients with bone marrow transplants. Live attenuated vaccine is more than 85% effective in young children. In healthy adults, live attenuated vaccine and inactivated influenza vaccine are similarly effective.

Influenza vaccination also reduces cardiovascular events. Vaccination efforts should begin as soon as the vaccine becomes available, usually by October, and should continue throughout the season. Peak influenza activity usually recurs in January or February, and influenza season continues through March.

Adverse Events

The most common side effect of inactivated vaccine is soreness at the injection site. Fever, malaise, and myalgia may begin 6 to 12 hours after vaccination and persist for 1 or 2 days, although such reactions are most common in children exposed to vaccine for the first time. The most common adverse events after live attenuated influenza vaccine in adults are runny nose, headache, and sore throat. Severe allergic reactions are rare, including Guillain-Barré syndrome in about one case per 1 million doses.

Measles

Indications

Measles (Chapter 367) immunization is recommended for all persons born in or after 1957 without laboratory evidence of immunity, laboratory confirmation of prior disease, or prior appropriate vaccination. Children should routinely receive two doses of MMR vaccine—one at 12 to 15 months of age and one at 4 to 6 years of age. Most adults are considered to have been appropriately vaccinated if they received one dose of vaccine on or after their first birthday. However, adults who are at increased risk of exposure to measles or transmission of it (e.g., health care workers, college students, international travelers) should receive a second dose if they did not get a second dose as children, unless they have serologically documented immunity, have laboratory confirmation of disease, or were born before 1957. Persons embarking

on foreign travel should have received two doses of MMR vaccine or have other evidence of measles immunity as defined before. Persons born before 1957 are usually immune as a result of natural infection and do not require vaccination, although vaccination is not contraindicated if they are believed to be susceptible.

During institutional outbreaks of measles, all persons who have not received two doses or who lack other evidence of measles immunity should be vaccinated. Although measles vaccine can be administered only with mumps and rubella as MMR, individuals already immune to one or more of the components may receive MMR without harm.

Measles vaccine is contraindicated for pregnant women on theoretical grounds, for persons with moderate to severe acute febrile illnesses, and for persons with altered immunocompetence, except those with HIV infection who are not severely immunocompromised. Patients with anaphylactic reactions to eggs can be vaccinated without prior skin testing.

Adverse Events

MMR vaccine can cause fever (<15%), rash (5%), transient lymphadenopathy (20% of adults), or parotitis (<1%). Febrile reactions, which usually are otherwise asymptomatic, generally occur 7 to 12 days after vaccination and persist for 1 or 2 days. MMR vaccination can rarely cause anaphylaxis, febrile seizures in children, thrombocytopenic purpura, transient arthralgia, and measles inclusion body encephalitis in persons with demonstrated immunodeficiencies.

Meningococcal Vaccines

Three quadrivalent meningococcal vaccines are available against disease caused by serogroups A, C, Y, and W135: meningococcal polysaccharide vaccine (MPSV4), which consists of 50 μg of polysaccharide of each of the four serogroups and is licensed for persons 2 years of age and older, and two meningococcal conjugate vaccines (MenACWY). A fourth vaccine, Hib-MenCY-TT, including *H. influenzae* type b plus meningococcal serotypes C and Y, was licensed in 2012 only for children ages 6 weeks through 18 months.

One conjugate vaccine consists of 4 μg of each polysaccharide covalently linked to 48 μg of diphtheria toxoid and licensed for persons 9 months to 55 years of age. The other conjugate vaccine consists of polysaccharide linked to CRM_{197} and is licensed for persons 2 months to 55 years of age. The four serogroups in each vaccine account for approximately two thirds of meningococcal disease in the United States and about 75% of the disease in persons 11 years of age or older (Chapter 298). Serogroup A and C polysaccharide vaccines are 85 to 100% efficacious in epidemic settings; vaccination with Y and W polysaccharides induces bactericidal antibodies and is presumed to be efficacious. In contrast to polysaccharide vaccines, conjugate vaccines induce immunologic memory, result in higher and more durable levels of high-avidity antibodies, and induce herd immunity. The duration of immunity for both MPSV4 and MenACWY is estimated at between 3 and 5 years.

Indications

Routine vaccination with MenACWY is recommended for all adolescents at 11 through 18 years of age, with a first dose administered at age 11 or 12 years and revaccination at age 16 years, or a first dose between 13 and 15 years and revaccination at age 16 to 18 years. Meningococcal vaccination once is also recommended for college freshmen who have not previously been vaccinated and will live in dormitories, military recruits, persons at risk during a community outbreak attributable to a vaccine serogroup, and persons who travel to or live in areas with hyperendemic or epidemic disease (e.g., the "meningitis belt" of sub-Saharan Africa, stretching from Mauritania to Ethiopia). For some persons at very high risk of meningococcal disease (e.g., microbiologists with frequent exposure to *Neisseria meningitidis* in culture and persons with persistent complement component deficiencies, splenic dysfunction, or asplenia), revaccination is recommended every 5 years.

MenACWY can be used for persons 2 months to 10 years of age with high-risk conditions but is not recommended routinely for this age group. For persons older than 55 years with an indication for vaccine, a single dose of MPSV4 is preferred. For persons now aged 56 years and older who were vaccinated previously with MenACWY and are recommended for revaccination or for whom multiple doses are anticipated (e.g., persons with asplenia and microbiologists), MenACWY is preferred.

Adverse Events

The major adverse reactions to MPSV4 are local reactions and systemic symptoms, such as headache and malaise, which generally persist for 1 or 2 days. The incidence of local reactions and low-grade fever is slightly higher after MenACWY than after MPSV4. An excess risk of Guillain-Barré syndrome was initially reported after MenACWY vaccines but has not been confirmed in subsequent studies.

Mumps

Indications

Mumps (Chapter 369) vaccine is indicated for all persons without evidence of immunity. For most adults, such evidence consists of a prior history of vaccination on or after the first birthday, laboratory evidence of immunity, or laboratory confirmation of disease. For adults at high risk, including health care workers, international travelers, and students at post-high school educational institutions, two doses of mumps vaccine constitute acceptable evidence of immunity. Most persons born before 1957 can be considered immune as a result of natural infection, although vaccination is not contraindicated if such persons are thought to be susceptible.

Adverse Events

Adverse events after the vaccine strain used in the United States are uncommon but include fever, parotitis, and allergic manifestations. Thrombocytopenic purpura has been reported rarely after MMR. Mumps vaccine is contraindicated for pregnant women on theoretical grounds, for persons with moderate to severe acute febrile illnesses, and for persons with altered immunocompetence. When it is combined with measles vaccine, it may be given to persons with asymptomatic HIV infection and considered for persons with symptomatic infection if they are not severely immunocompromised. Patients with anaphylactic reactions to eggs can be vaccinated without skin testing.

Pertussis Vaccine

Each of the two vaccines licensed for boosting immunity to pertussis in adults is combined with tetanus and diphtheria toxoids, and they have a lower content of pertussis antigens compared with the childhood pertussis-containing vaccines (Tdap). Boostrix (GlaxoSmithKline), which is licensed for adolescents and adults 10 years of age and older, contains three pertussis antigens—toxoid (PT), filamentous hemagglutinin (FHA), and pertactin (PRN). Adacel (Sanofi Pasteur), which is licensed for 10- through 64-year-olds, contains five pertussis antigens: PT, FHA, PRN, and two fimbriae. Both vaccines, when they are administered to previously vaccinated adolescents and adults, induce serologic responses that are comparable to those induced with effective childhood vaccination.

Indications

A single dose of Tdap is indicated for all adolescents at 11 to 12 years of age. Older adolescents and adults who have not received Tdap should get it instead of their next scheduled Td booster. Tdap can be given at any interval after a prior Td. All pregnant women should receive one dose of Tdap vaccine during each pregnancy, optimally between 27 and 36 weeks of gestation to maximize maternal antibody response and passive antibody transfer to the infant. If they have not been vaccinated during pregnancy, women should receive Tdap post partum if they have never previously been vaccinated with it. With the exception of pregnant women, booster doses of Tdap are not recommended.

Adverse Events

Adverse events, usually local reactions, are similar with the adult preparation of tetanus and diphtheria toxoids (Td) alone (see tetanus and diphtheria section later).

Pneumococcal Vaccines

Two pneumococcal vaccines are available for adults. Pneumococcal polysaccharide vaccine (PPSV23) consists of purified polysaccharide capsular antigens from the 23 types of *Streptococcus pneumoniae* that are responsible for 85 to 90% of the bacteremic disease in the United States (Chapter 289). Most adults, including elderly patients and those with alcoholic cirrhosis and diabetes mellitus, have a two-fold or greater rise in type-specific antibodies within 2 to 3 weeks of vaccination. Vaccination is approximately 60% effective against invasive pneumococcal disease, but its efficacy against pneumonia in high-risk populations, such as patients with alcoholic cirrhosis or Hodgkin disease, is not clear.

Pneumococcal conjugate vaccine (PCV13), licensed in 2011 for adults aged 50 years and older, should be administered once to eligible adults

because it is expected to add protection to that offered by PPSV23. Ideally, PCV13 should precede PPSV23 by at least 8 weeks in immunocompromised persons over age 19 years and by at least 6 months in persons who are age 65 years or older and are not immunocompromised; if PPSV23 is administered first, an interval of 1 year should elapse before PCV13 is administered.

Indications

The preponderance of information supports the use of pneumococcal vaccines in high-risk populations, including all persons 65 years and older, persons 19 years or older who smoke cigarettes, and patients with asthma. Special efforts should target hospitalized patients. Approximately two thirds of patients who are admitted later with pneumococcal disease had been hospitalized for other reasons within the preceding 5 years. The newer PCV13 vaccine is recommended for all adults over age 65 years and for adults 19 years of age and older with immunocompromising conditions, functional or anatomic asplenia, cerebrospinal fluid leaks, or cochlear implants.

Because immunity may decrease 5 years or more after initial vaccination with PPSV23, a single booster dose of PPSV23 should be considered at that time for adults at highest risk of disease (such as asplenic patients) and for adults who lose antibody rapidly (such as patients with nephrotic syndrome or renal failure). Persons 65 years or older who received a dose more than 5 years earlier when they were younger than 65 years also should be revaccinated.

Adverse Events

Local reactions to PPSV23 are frequent, but less than 1% of vaccinees experience severe local reactions or systemic symptoms such as fever and malaise. Severe events, such as anaphylaxis or Arthus-like reactions at the site of injection, are rare. Because of the rarity of severe reactions in revaccinated patients, persons with indications for vaccination but with unknown histories of prior vaccination should be vaccinated. PCV13 reactions include pain, redness, and swelling at the injection site; limitation of movement of the injected arm; fatigue; and headache.

Poliomyelitis

Since 2000, inactivated poliovirus vaccine (IPV) has replaced the live attenuated oral poliovirus vaccine (OPV) in the United States because OPV vaccine caused about eight polio cases per year in the United States among recipients or their contacts (Chapter 379). OPV is still the vaccine used in most countries around the world.

Indications

Routine vaccination of persons 18 years of age or older is not recommended. If adults who are unvaccinated, are incompletely vaccinated, or have unknown vaccination status travel to areas where wild poliovirus is endemic or epidemic, they should receive a series of three doses: two doses of IPV administered at an interval of 4 to 8 weeks, with a third dose adminitered 6 to 12 months after the second. If three doses of IPV cannot be administered, alternatives include the following: three doses of IPV administered 4 weeks or more apart; if less than 8 weeks remain before protection is needed, two doses of IPV administered 4 weeks or more apart; and if less than 4 weeks remain before protection is needed, a single dose of IPV. If fewer than three doses are administered, the remaining doses needed to complete a three-dose series should be administered when feasible, at the intervals recommended, especially if the person remains at increased risk for poliovirus exposure.

As a precaution, adults (≥18 years of age) who are traveling to areas where poliomyelitis cases are occurring and who have received a routine series with either IPV or OPV in childhood should receive another dose of IPV before departure.[7] For adults, available data do not indicate the need for more than a single lifetime booster dose with IPV.

Adverse Events

Minor local reactions (pain, redness) commonly occur after IPV. Because IPV contains trace amounts of streptomycin, polymyxin B, and neomycin, allergic reactions may occur in persons allergic to these antibiotics.

Rabies

Two inactivated rabies vaccines are licensed in the United States.[8] Human diploid cell vaccine (HDCV), which is prepared from the Pitman-Moore strain of rabies virus, also contains small amounts of neomycin sulfate, albumin, and phenol red indicator. The purified chick embryo cell vaccine (PCECV), which is prepared from the Flury LEP rabies virus strain, also contains small amounts of polygeline, human serum albumin, potassium glutamate, and sodium EDTA. Both vaccines are given intramuscularly.

Indications

Rabies (Chapter 414) vaccine is indicated for pre-exposure prophylaxis of high-risk persons, including animal handlers, selected laboratory and field workers, and persons traveling for more than 1 month to areas where rabies is a constant threat. For both rabies vaccines, the pre-exposure regimen consists of three 1-mL intramuscular injections on days 0, 7, and 21 or 28. Post-exposure treatment depends on prior exposure to vaccine (Chapter 414). Persons being treated for the first time should be given human rabies immune globulin as well as four doses of vaccine at days 0, 3, 7, and 14.

Adverse Events

Local reactions (e.g., pain at the injection site, redness, swelling, and induration) are common after both rabies vaccine preparations. Hypersensitivity reactions can occur after booster doses.

Rubella

Indications

One dose of rubella vaccine (Chapter 368) is indicated for adults born in 1957 or later without evidence of immunity and for women of any age who lack evidence of immunity and who are considering becoming pregnant. Persons without a prior history of vaccination on or after the first birthday, laboratory evidence of immunity, or laboratory confirmation of disease should be considered as lacking evidence of immunity. A single dose of vaccine is 95% or more effective. Many persons receive two doses of rubella vaccine by the two-dose schedule of MMR.

Adverse Events

Follow-up of pregnant women who had no evidence of immunity and who received rubella vaccines within 3 months of the estimated date of conception show no evidence of defects compatible with congenital rubella syndrome in their offspring. Nevertheless, vaccine is contraindicated in pregnant women on theoretical grounds, and conception should be delayed for 1 month after rubella vaccination.

Arthralgia develops among about 25% of nonimmune postpubertal females after vaccination with rubella vaccine. Symptoms generally begin 1 to 3 weeks after vaccination, usually are mild, persist for about 2 days, and rarely recur. These symptoms are less common in postpubertal males compared with females. Other infrequent adverse events include transient peripheral neuritis and pain in the arms and legs. Thrombocytopenic purpura is rare when rubella vaccine is administered as MMR. Rubella vaccine is contraindicated for persons with moderate to severe acute illnesses and for persons with reduced immunocompetence. When it is given with measles vaccine, it may be administered to persons with asymptomatic HIV infection and considered for symptomatic persons who are not severely immunocompromised. Rubella vaccine is grown in human diploid cells and can be administered without problems to persons with allergy to eggs.

Tetanus and Diphtheria

Tetanus (Chapter 296) toxoid is one of the most effective immunizations, with more than 95% protection after a primary series of three doses. In persons aged 7 years or older, it should always be used in combination with diphtheria (Chapter 292) toxoid (Td), which is more than 85% effective in preventing disease. Combinations that also include pertussis antigens (Tdap) are preferred to Td for routine immunization of adolescents and adults who have yet to receive Tdap as well as for pregnant women (for whom Tdap is recommended during every pregnancy).[9] Doses need not be repeated if the schedule is interrupted. A booster dose of Td is recommended every 10 years.

Indications

After a wound, persons of unknown immunization status or persons who have received fewer than three doses of tetanus toxoid should receive a dose of Tdap or Td regardless of the severity of the wound. Td also is indicated for persons who have previously received three or more doses if more than 10 years has elapsed since the last dose, in the case of clean and minor wounds, and if more than 5 years has elapsed for all other wounds. Persons who have never received a dose of Tdap should receive it instead of Td for wound management. Tetanus immune globulin should be administered simultaneously at a separate site to persons who have wounds that are not clean and minor if they have not previously received at least three doses of toxoid.

Adverse Events

Most reactions to Td consist of local inflammation and low-grade fever. Exaggerated local (Arthus-like) reactions, with extensive painful swelling, often from shoulder to elbow, are occasionally reported 2 to 8 hours after receipt of a diphtheria- or tetanus-containing vaccine, particularly in persons who have received frequent doses of diphtheria or tetanus toxoid. Severe systemic reactions, such as generalized urticaria, anaphylaxis, Guillain-Barré syndrome, and other neurologic complications, rarely have been reported after receipt of tetanus toxoid.

Varicella: Chickenpox

A single dose of live attenuated varicella vaccine (Oka strain), which can be combined with MMR vaccine, protects 70 to 90% of recipients against any disease and more than 95% of recipients against severe disease, but a two-dose schedule is recommended. Whether immunity wanes with increasing time after vaccination is unclear. Use of vaccine has been associated with dramatic decreases in the incidence of varicella.

Indications

Varicella vaccine is indicated routinely for all children without a contraindication. The two-dose schedule includes vaccination at 12 to 15 months of age and again at 4 to 6 years of age. For persons who previously received just a single dose, a catch-up second vaccination is recommended, preferably at least 3 months after the first dose. Persons 13 years or older without evidence of immunity to varicella should receive two doses at least 4 weeks apart. Evidence of immunity to varicella includes documentation of age-appropriate vaccination with two doses of varicella vaccine at least 28 days apart, laboratory evidence of immunity or laboratory confirmation of disease, birth in the United States before 1980, or diagnosis or verification of a history of varicella or herpes zoster disease by a health care provider. Serologic screening of adults in some situations may be cost-effective, provided that identified susceptible adults are vaccinated. Serologic testing is not indicated after vaccination. The vaccine is contraindicated in persons who are immunocompromised, persons with anaphylactic allergies to vaccine components, and pregnant women. Post-exposure vaccination within 3 days of exposure can reduce the likelihood of symptomatic infection by about two-thirds. Varicella vaccine is temperature sensitive, so it must be stored at −15° C or colder to retain potency and should be discarded if it is not used within 30 minutes of reconstitution.

Adverse Events

The most common side effect is soreness at the injection site, which is reported in 25 to 35% of recipients 13 years or older. Varicella-like rashes at the injection site (median of two lesions) have been reported in 3% of recipients in this age group after the first dose and in 1% after the second dose. Nonlocalized rashes with a median of five lesions have been reported in 5.5% of recipients after the first dose and in 0.9% after the second dose. Although the vaccine virus can cause herpes zoster (shingles), especially in children, the incidence is substantially lower than would be expected after natural varicella (Chapter 375). More severe events in temporal relation to the vaccine have been reported rarely, but a causal relationship has not been established. Transmission of vaccine virus to a contact is extremely rare and probably occurs only from vaccinees in whom a varicella-like rash developed.

Varicella: Zoster

The varicella-zoster virus vaccine, which is approximately 14 times more potent than the varicella vaccine used routinely in children, reduces zoster by about 50% and post-herpetic neuralgia by about two thirds in persons 60 years of age or older. The efficacy against zoster declines after age 70 years, but protection against post-herpetic neuralgia continues. As for the varicella-chickenpox vaccine, special freezer storage is required.

Indications

A single dose of zoster vaccine is recommended for persons 60 years of age or older, even if they have a prior history of zoster. It is not necessary to elicit a history of varicella or to test for varicella immunity before administering the vaccine. The vaccine is not recommended for immunocompromised persons or pregnant women.

Adverse Events

Local reactions (erythema, pain or tenderness, and swelling at the injection site) are common. Severe reactions were of similar incidence in vaccine and placebo recipients in clinical trials.

VACCINES INTENDED PRIMARILY FOR INTERNATIONAL TRAVELERS

Evaluation of people before travel should include a review and provision of routine vaccines recommended on the basis of age and other individual characteristics. Recommendations for specific vaccines related to travel will depend on itinerary, duration of travel, and host factors. Detailed recommendations are available at *www.cdc.gov/travel*.[10]

Japanese Encephalitis Vaccine

Indications

Japanese encephalitis (Chapter 383) vaccine is indicated primarily for travelers to Asia who will spend a month or longer in endemic areas during the transmission season, especially if travel will include rural areas. In all instances, travelers should be advised to take personal precautions to reduce exposure to mosquito bites. An older vaccine was reported to be 80 to 91% effective in preventing clinical disease, and the current whole virus inactivated vaccine (Ixiaro, Intercell Biomedical) was licensed on the basis of comparable immunogenicity. The primary series consists of two 0.5-mL doses given intramuscularly on days 0 and 28, with the second dose administered at least 1 week before travel (see Table 18-2). If the primary series was administered more than 1 year previously, a booster dose may be given before the next potential exposure to the virus.

Adverse Events

Headache and myalgia and local reactions (pain and tenderness) occur in more than 10% of vaccinees. However, the incidence rates of these events were similar to those in a comparison group that received a placebo with aluminum hydroxide.

Typhoid Vaccine

Indications

Two types of vaccines, a live attenuated Ty21a oral vaccine and a capsular polysaccharide vaccine (ViCPS), appear to be of comparable efficacy (50 to 77%). Typhoid (Chapter 308) vaccine is indicated primarily for travelers to areas where the risk of prolonged exposure to contaminated food and water is high. Because the vaccine is not always effective, food and water precautions are still essential. The vaccine also may be considered for family or other intimate contacts of typhoid carriers and laboratory workers who work with *Salmonella typhi.* For adults and children 6 years and older, either of the vaccines may be used. For Ty21a, one enteric-coated capsule is taken every other day for four doses. Alternatively, a single dose of the ViCPS vaccine may be given. The duration of protection with Ty21a is not known; repetition of the primary series is recommended every 5 years for persons at risk. Boosters are recommended every 2 years for the ViCPS vaccine if persons continue to be at risk. The ViCPS vaccine can be given to children as young as 2 years.

Adverse Events

Fever and headache may occur after receipt of typhoid vaccines. Stomach pain, nausea, and rash are rarely observed.

Yellow Fever Vaccine

Indications

Yellow fever (Chapter 381) occurs only in areas of South America and Africa. Vaccination with a single dose of the live attenuated 17D strain of virus confers protection to almost all recipients for at least 10 years. A booster dose is recommended every 10 years for persons at continued risk of exposure to yellow fever.

Adverse Events

Adverse reactions (fever, aches, and soreness, redness, or swelling where the injection was given) occur in up to 25% of vaccinees. Anaphylaxis has been reported in 0.8 to 1.8 persons per 100,000 doses of vaccine distributed. A rare syndrome (0.25 case per 100,000 doses distributed) of multiple organ system failure or viscerotropic disease with high rates of mortality has been reported, primarily among older adults and persons who have undergone thymectomy or have severe thymic dysfunction. Meningoencephalitis, Guillain-Barré syndrome, acute disseminated encephalomyelitis, and bulbar palsy have been reported in 1 to 2 persons per 100,000 doses and are more common in older vaccines. In patients 60 years and older who are going to spend time in yellow fever–endemic zones, yellow fever vaccine should be administered with caution and only after careful counseling. Yellow fever vaccine should not be

given to immunocompromised persons or persons with anaphylactic allergies to eggs. The vaccine is contraindicated in pregnant women on theoretical grounds, although pregnant women who must travel to a high-risk area may be vaccinated.

VACCINES FOR POSSIBLE BIOTERRORISM AGENTS

Anthrax Vaccine

Anthrax (Chapter 294) vaccine adsorbed (AVA) is prepared from a cell-free filtrate of a nonencapsulated strain of anthrax and contains many cell products, including protective antigen. Protective antigen is responsible for binding to cells, allowing transport of lethal factor and edema factor into host cells. A recombinant protective antigen (rPA) vaccine is in clinical trials.

Indications

Pre-exposure prophylaxis consists of a three-dose primary intramuscular schedule at 0, 4 weeks, and 6 months, with booster doses at 12 and 18 months, followed by annual boosters. Protective efficacy of an earlier form of the vaccine against cutaneous anthrax was 92.5%. Animal models suggest efficacy against inhalation anthrax. Pre-exposure vaccination is recommended for persons engaged in work involving exposure to high concentrations of *Bacillus anthracis* or in activities with high potential for aerosol production. Vaccine is recommended in conjunction with antibiotics for postexposure prophylaxis after exposure to aerosolized *B. anthracis* spores. The recommended regimen is three doses of AVA administered at 0, 2, and 4 weeks, combined with at least 60 days of antibiotics, which should be continued for at least 14 days after the third dose of vaccine (Chapter 294).

Adverse Events

The most common adverse events are local reactions, including subcutaneous nodules, which are thought to be due to the deposition of the aluminum-containing adjuvant in subcutaneous tissue. These adverse events are less common with intramuscular injections than with subcutaneous injections.

Smallpox Vaccine

Smallpox vaccine uses vaccinia virus, an orthopox virus that is distinct from variola and cowpox viruses and that provides cross-protection from smallpox. Smallpox vaccine is close to 100% effective when it is administered properly with a bifurcated needle. Vaccination also prevents or modifies disease when it is administered within 3 to 4 days of exposure and perhaps even after greater delays. The skin usually does not need any special preparation. If alcohol is used for cleaning, the skin should be allowed to dry before vaccination to avoid inactivation of the vaccine. The needle is held perpendicular to the skin with 15 punctures for all vaccinees, made rapidly with enough vigor to ensure that a trace of blood appears within 15 to 20 seconds. With a primary take, the vaccination site should become reddened and pruritic within 3 or 4 days after vaccination; a large vesicle with a red areola forms and becomes pustular by 7 to 11 days. The lesion scabs by the third week.

Indications

The vaccine is indicated for persons who work with orthopox viruses. To increase preparedness for a smallpox attack, vaccination is often recommended for persons who will serve on public health or health care response teams. The duration of immunity is unclear. Revaccination is recommended at least every 10 years for persons who continue to be at risk. Contraindications include history or presence of eczema, other chronic or exfoliative skin conditions, and immunosuppression or pregnancy in the patient or a close household or other contact. Persons who are younger than 1 year, are breastfeeding, or have allergies to vaccine components should not be vaccinated. Because of reports of post-vaccination cardiac events, vaccination should be deferred in persons with ischemia or other severe heart diseases or persons at high risk for ischemic heart disease events (*http://emergency.cdc.gov/agent/smallpox/vaccination/index.asp*). In the event of exposure to variola, there are no contraindications. Should variola be introduced into a community, vaccination would be indicated for all exposed persons and their close contacts to prevent further spread, and recommendations for more widespread vaccination would have to be evaluated on a case-by-case basis.

Adverse Events

Fever is the most common adverse event. Other more serious complications include eczema vaccinatum, which is a local or disseminated vaccinia infection in persons with a history of eczema or other exfoliative dermatitis; vaccinia necrosum, which occurs in immunocompromised persons; autoinoculation, especially of the eye, which can cause keratitis and scarring; generalized vaccinia; myopericarditis; and encephalitis. The risk for death from vaccinia is about one case per 1 million primary vaccinations.

Other Agents

Other organisms or products that have been considered potential bioterrorism threats include plague (Chapter 312) and botulinum toxin (Chapter 296). Poisoning with botulinum toxin can be treated with a trivalent antitoxin available from the Centers for Disease Control and Prevention (see *http://www.cdc.gov/laboratory/drugservice/formulary.html*). An experimental heptavalent botulinum toxoid can be obtained from the Centers for Disease Control and Prevention for laboratory workers at high risk for exposure to toxin. Pre-exposure vaccination is not warranted or feasible for the general population.

OTHER VACCINES

A protein-conjugated vaccine for *H. influenzae* type b (Hib) should be considered for some adults at high risk for invasive Hib disease (e.g., asplenia, sickle cell disease, or recipient of a hematopoietic stem cell transplant) if they have not previously received Hib vaccine.

Grade A References

A1. Prymula R, Siegrist CA, Chlibek R, et al. Effect of prophylactic paracetamol administration at time of vaccination on febrile reactions and antibody responses in children: two open-label, randomised controlled trials. *Lancet*. 2009;374:1339-1350.

A2. Udell JA, Zawi R, Bhatt DL, et al. Association between influenza vaccination and cardiovascular outcomes in high-risk patients: a meta-analysis. *JAMA*. 2013;310:1711-1720.

A3. Thierry-Carstensen B, Jordan K, Uhlving HH, et al. A randomised, double-blind, non-inferiority clinical trial on the safety and immunogenicity of a tetanus, diphtheria and monocomponent acellular pertussis (TdaP) vaccine in comparison to a tetanus and diphtheria (Td) vaccine when given as booster vaccinations to healthy adults. *Vaccine*. 2012;30:5464-5471.

A4. Moberley S, Holden J, Tatham DP, et al. Vaccines for preventing pneumococcal infection in adults. *Cochrane Database Syst Rev*. 2013;1:CD000422.

A5. Macartney K, Heywood A, McIntyre P. Vaccines for post-exposure prophylaxis against varicella (chickenpox) in children and adults. *Cochrane Database Syst Rev*. 2014;6:CD001833.

A6. Gagliardi AM, Gomes Silva BN, Torloni MR, et al. Vaccines for preventing herpes zoster in older adults. *Cochrane Database Syst Rev*. 2012;10:CD008858.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bridges CB, Coyne-Beasley T. Advisory Committee on Immunization Practices recommended immunization schedule for adults aged 19 years or older: United States, 2014. *Ann Intern Med*. 2014;160:190.
2. Centers for Disease Control and Prevention. Advisory Committee on Immunization Practices (ACIP) recommended immunization schedule for adults aged 19 years and older—United States. 2014. http://www.cdc.gov/vaccines/schedules/downloads/adult/adult-combined-schedule.pdf; Accessed January 29, 2015.
3. Tomovici A, Barreto L, Zickler P, et al. Humoral immunity 10 years after booster immunization with an adolescent and adult formulation combined tetanus, diphtheria, and 5-component acellular pertussis vaccine. *Vaccine*. 2012;30:2647-2653.
4. Giuliano AR, Palefsky JM, Goldstone S, et al. Efficacy of quadrivalent HPV vaccine against HPV infection and disease in males. *N Engl J Med*. 2011;364:401-411.
5. Black CL, Yue X, Ball SW, et al. Influenza vaccination coverage among health care personnel—United States, 2013-14 influenza season. *MMWR Morb Mortal Wkly Rep*. 2014;63:805-811.
6. Osterholm MT, Kelley NS, Sommer A, et al. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. *Lancet Infect Dis*. 2012;12:36-44.
7. Wallace GS, Seward JF, Pallansch MA. Interim CDC guidance for polio vaccination for travel to and from countries affected by wild poliovirus. *MMWR Morb Mortal Wkly Rep*. 2014;63:591-594.
8. Rupprecht CE, Briggs D, Brown CM, et al. Use of a reduced (4-dose) vaccine schedule for postexposure prophylaxis to prevent human rabies: recommendations of the advisory committee on immunization practices. *MMWR Recomm Rep*. 2010;59:1-9.
9. Terranella A, Asay GR, Messonnier ML, et al. Pregnancy dose Tdap and postpartum cocooning to prevent infant pertussis: a decision analysis. *Pediatrics*. 2013;131:e1748-e1756.
10. Centers for Disease Control and Prevention. Traveler's health. www.cdc.gov/travel; Accessed January 29, 2015.

REVIEW QUESTIONS

1. Tdap vaccine is recommended to be given only once for all groups except which of these?

A. Pregnant women
B. Parents of children younger than 2 years
C. Adults 65 years of age and older
D. Children 11 years of age or older who did not complete their DTaP vaccine series
E. Adolescents and adults entering college

Answer: A Pregnant women should receive Tdap vaccine during each pregnancy to offer protection to each newborn. Available data indicate fairly rapid waning of pertussis immunity after Tdap vaccination and no significant adverse events after repeated Td or Tdap vaccinations with short intervals between them. For more information, see *http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6207a4.htm*.

2. Persons who are at increased risk of meningococcal disease and for whom meningococcal vaccine is recommended include all except which of the following?

A. Persons without spleens or whose spleens are not functional (e.g., have sickle cell disease)
B. Persons with persistent complement deficiencies
C. Persons 65 years and older
D. Microbiologists regularly working with *Neisseria meningitidis* isolates
E. Persons traveling to sub-Saharan Africa

Answer: C Available data indicate increased risk of meningococcal disease for all the listed groups except persons 65 years and older. *http://www.cdc.gov/mmwr/pdf/rr/rr6202.pdf*.

3. Pneumococcal conjugate vaccine is recommended for which of these adult groups?

A. Persons with nephrotic syndrome
B. Persons with liver cirrhosis
C. Persons with chronic obstructive pulmonary disease
D. Persons with coronary artery disease
E. Persons with diabetes

Answer: A Most of the additional benefit of pneumococcal conjugate vaccination, when it is used to supplement pneumococcal polysaccharide vaccination, accrues to persons with iatrogenic or disease-related immunocompromising conditions, including nephrotic syndrome, as well as persons with functional or anatomic asplenia, cerebrospinal fluid leaks, and cochlear implants. For more information, see *http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6140a4.htm*.

4. Adults at increased risk of hepatitis B infection who should receive hepatitis B vaccine include which of the following?

A. Persons with chronic obstructive pulmonary disease
B. Persons with coronary artery disease
C. Persons with nephrotic syndrome not on dialysis
D. Persons 19 to 59 years old with diabetes mellitus
E. Persons 65 years and older

Answer: D Persons who have frequent contact with contaminated blood or body fluids are at increased risk of hepatitis B infection. Adults 19 to 59 years of age with diabetes mellitus also are at increased risk of hepatitis B infection, at least in part because of poor infection control practices with assisted blood glucose monitoring (see *http://www.cdc.gov/mmwr/pdf/wk/mm6050.pdf*). Therefore, the Centers for Disease Control and Prevention recommend that adults with diabetes mellitus younger than 60 years receive hepatitis B vaccine.

5. Human papillomavirus (HPV) vaccine is recommended for

A. Routine and catch-up vaccination for boys beginning at 11 years and men through 21 years
B. Persons 65 years and older
C. Only men who have sex with men
D. Girls and women 11 to 50 years
E. Girls and boys beginning at 9 years of age

Answer: A HPV vaccine is recommended for persons primarily before they begin sexual activity and have been infected with HPV. The upper age range for both licensed vaccines is 26 years. Although the vaccine is licensed for both boys and girls beginning at the age of 9 years, it is recommended to begin at age 11 to 12 years, at the same time as other adolescent vaccines (Tdap and meningococcal vaccines). For more information, see *http://www.cdc.gov/mmwr/pdf/wk/mm6050.pdf* and *http://www.cdc.gov/mmwr/PDF/wk/mm5920.pdf*.

19

PRINCIPLES OF OCCUPATIONAL AND ENVIRONMENTAL MEDICINE

MARK R. CULLEN

In the first several decades after World War II, when many American workers came to enjoy coverage by health insurance—for everything *but* workplace injuries and illnesses—the myth grew that modern work is largely free of the risks of the industrial horrors of past eras. Starting in the 1970s, however, resurgence of societal and medical interest in these consequences of work found that diseases related to work are not truly extinct, just not well observed or studied. Occupational physicians, often cut off from mainstream medical practice, had difficulty in changing the perception, and most practicing internists were largely oblivious. It is now recognized that a substantial burden of ill health and disability is due to work-associated physical, chemical, and biologic hazards. Psychosocial aspects of work also may be injurious to health.

Although tens of thousands of toxic chemicals and other hazards can potentially cause or exacerbate a wide range of acute and chronic conditions, certain basic principles and clinical approaches apply broadly to general and specialty medical practice. This chapter outlines these basics, then briefly summarizes the most common occupational disorders seen by internists in

developed countries, and finally reviews the effects of the environmental exposures most likely to be encountered.

PRINCIPLES OF OCCUPATIONAL AND ENVIRONMENTAL DISEASE

It is widely imagined that the major health effects of environmental and occupational exposures are unique disorders best recognized by their failure to fit easily into other diagnostic categories (e.g., arsenic poisoning). In reality, *the major consequences of chemical and physical exposures are, without further exploration of an environmental connection, indistinguishable in clinical presentations from disorders that make up the bulk of outpatient and inpatient medical practice:* common rashes (Chapter 438), nonspecific liver function abnormalities (Chapter 147), wheezing and irritative symptoms of the upper and lower respiratory tract (Chapter 87), various cancers (Chapter 180), peripheral neuropathies (Chapter 420), dysphoria (Chapter 397), and nonspecific cognitive dysfunction (Chapter 402). Although a handful of pathologically distinct disorders still occur, such as silicosis (Chapter 93) and lead poisoning (Chapter 22), when an environmental or workplace agent causes overt disease, physiologic and radiographic studies typically reveal manifestations completely consistent with common diagnoses such as asthma (Chapter 87), contact dermatitis (Chapter 438), fatty liver (Chapter 152), and lung cancer (Chapter 191).

The underlying cause of such conditions will inevitably remain obscure unless the clinician adheres to a disciplined approach designed to investigate and to exclude occupational or environmental causes whenever it is appropriate. The best approach is consistent use of the occupational and environmental history, a short series of questions that can be expanded on the basis of the responses (see later). The point is that the internist cannot "wait" to consider occupational or environmental issues until other diseases have been ruled out without running the risk of missing almost every occupational and environmental effect that he or she will encounter.

Whatever the pathway or time course, exposure dose is the major determinant of the risk for development of disease. As in pharmacology (Chapter 29), it is impossible to make any meaningful statement about cause and effect without appreciation of dose. Consider, for example, the difference in health effects of aspirin at 65 mg, 650 mg, and 6500 mg (Chapter 37). Over this two-order magnitude of change, the chemical goes from having one therapeutic target organ to having many to being lethal. It is no different with lead or organophosphate pesticides or solvents, except that there is rarely as simple a way to determine dose as in the drug situation, where pill bottles are labeled, drug prescriptions are recorded, and blood or urine levels are readily available in most laboratories. This limitation is exacerbated because, unlike with drugs, the range of toxic exposures may vary far more widely. For example, water in a contaminated drinking well or poor indoor air in an office could have toxins at a level that is two, three, or even four orders of magnitude (i.e., 10,000 times) lower than the level that may have been evaluated in epidemiologic studies of workers or tested in animals. Fortunately, it is much easier to "range find" than one might presuppose (see later discussion of history), and eagerness for precision—often unattainable—should not interfere with obtaining the great amount of information that *can* be readily gleaned from the patient and is often sufficient to act on. The key point is that no attempt to apply clinical information in relation to work or environment can be useful without some effort to characterize exposure dose.

Environmental hazards may affect preferentially vulnerable populations—those with underlying disease, those at the extremes of life, those with atopy, and those with other serious health risks such as smoking or diabetes. Genetic variability may underlie some of these differences, but few relevant genes have been sufficiently characterized for use in practice. Clinical studies of a host of common occupational diseases have identified behavioral and constitutional cofactors; for example, smoking dramatically increases the risk of lung cancer in asbestos-exposed workers (Chapter 191). This interaction creates a double demand on the clinician—the presence of smoking or atopy in a young woman with cough not only does not *preclude* the possibility of an occupational cause of her asthma but rather actually *increases* the likelihood that such an exposure may be important.

The Occupational and Environmental History and Exposure Assessment

Key to determining whether work and other environmental exposures may be causing or contributing to adverse health is the exposure history. The approach to obtaining this information and to the use of available resources to corroborate and complement it depends on the clinical context. In primary and much specialty medical care, where it is anticipated that a patient will be observed during a long period into the future, the most important step is to establish the hazards to which the patient may be exposed at work presently, the activities that may have resulted in past harmful exposures potentially relevant to future health (because of a latency with tobacco), and whether the present residential environment (including air and water and food sources) is thought to be contaminated by harmful materials. The recommended approach is to use a simple questionnaire, which can be self-administered or supervised by a medical extender (E-Fig. 19-1). These instruments can then be reviewed together by the patient and physician as time permits and updated over time. When jobs or materials are noted but the actual generic exposures are unknown, the patient and available reference sources can be enlisted to "translate" the history into specifics, such as which metals are being welded or what is actually contained in a cleaning agent or plastic. This information is obligatorily maintained and supplied on request by employers in most developed countries in the form of fact sheets termed Material Safety Data Sheets, many of which can be easily found online as well. In this way, the ongoing and former exposures, which may have an impact on health, can be noted and, where important, incorporated into routine preventive care or clinical surveillance for sequelae.

For patients with new clinical complaints or recently diagnosed conditions, the question of an environmental cause looms more urgently, so the approach must be more focused. If symptoms or signs of acute or subacute illness are suggested, the *timing* of recent or unusual environmental exposures in relation to the symptoms is key—more important than specific chemical detail. For example, if the patient develops shortness of breath shortly after the introduction of a new chemical or process at work or after a leak or spill, that fact should drive further questions, such as Did others get sick as well? For recurrent symptoms, such as cough or rash, cyclic changes are most often the strongest clue: Do symptoms get worse on workdays and improve on days off or holidays? For more insidious symptoms, such as weakness or numbness of the extremities or new-onset hepatic dysfunction, the appropriate question would be whether the onset of the abnormality has followed by weeks or months some demonstrable change in the work or home environment. Again, the coincidence of others similarly affected may be more valuable than detailed knowledge of the constituents of that environment. When such a temporal pattern is suggested, further efforts are warranted to establish what exposure may have occurred and what its dose may have been, often in conjunction with a specialty consultation.

In the elucidation of evidently more chronic conditions, such as pulmonary fibrosis, chronic renal insufficiency, or a malignant neoplasm, an alternative approach is suggested because the exposure, if relevant, is usually remote. In this situation, a detailed query about current work or ambient environments is *not* likely to be helpful in differential diagnosis, although knowledge of a past exposure to an important hazard (such as silica, asbestos, or cadmium) might, on the basis of the knowledge of its effects, influence the sequence of the evaluation. However, it is generally more efficient to explore past exposures *after* the pathophysiologic disturbance has been characterized, focusing inquiry on factors known to cause or suspected of causing that disorder—as easily found in suggested texts or literature searches.

In acute or chronic cases, information about *what* the exposure has been (generically) must be augmented by an estimation of exposure dose. A brief exposure to a fume containing a small percentage of lead will not, in general, cause acute lead poisoning (although hosts may differ in their responses), nor will trace contamination of a drinking well with benzene typically cause blood dyscrasias. The patient will rarely be able to supply detailed information about past or even current "dose" but often can provide valuable clues: Did the exposure continue during many years? Were fumes or fibers grossly visible in the air? Were respirators or other protective gear necessary or offered? Have episodes of unprotected exposure ever resulted in irritation or acute discomfort? A positive reply to any of these questions would suggest "high" exposure, where the reference point is the level at which the risk for development of a health effect becomes substantial. Conversely, if exposure has occurred in an otherwise typical office or around a home renovation, the levels of exposure are more likely "low." Nevertheless, such low-level exposure does not exclude a health effect, especially one caused by idiosyncratic mechanisms or occurring in hosts who are more "sensitive" to chemical exposures, a health characteristic found in 2 to 10% of the population. Although not to be condoned because of potential broader public health consequences, exposures to trace contaminants in food and drinking water are uncommon causes of *perceptible* clinical problems. When concern about the exposure is high, information from patients can be readily supplemented by information from employers (with the patient's consent!) and regulatory or health

Occupational and Environmental Exposure History

Current employment (if not currently employed, fill out for last job held)

1. Job title _____
2. Name and address of employer _____
3. Nature of work _____
4. How many hours a week do you usually work? _____
5. Start date? _____ If ended, last day worked _____
   If ended, why did you leave? _____
6. Do you have to wear personal protective equipment (like a respirator or gloves or hearing protection) at your job? If yes, what kind(s) _____
7. Exposures at this job. For each, record whether "none," "a little," or "a lot"

	None	A little	A lot
Dust or fumes			
Organic solvents or thinners			
Other chemicals			
Noise			
Excessive heat			
Physical strain			
Mental stress			

Conflict between work and family

Any hazard at this job that you think might be making you sick? _____

Past employment

Please list all the jobs you had before this last one, working backward toward school. For each, list the title, years you held job, and whether there was any exposure you are worried about from that job.

Job title	Years held	Exposure of concern?

Were you ever in the military service?

If yes, what years? _____ Were you in an active war zone? _____

Where? _____

Did hazardous exposures in any of these jobs make you sick? _____

If so, explain _____

Environmental exposures

Are there any specific exposures in your home or neighborhood you are worried about?

If so, what? _____

Please tell us about your home environment:

1. Is your drinking water from a city or town source? Private well? Any health problems with the water? _____
2. Is your heating system oil? Gas? Electric? Any health problems with the heat? _____
3. Are there sources of pollution near your home? _____ What? _____
4. Does anyone smoke inside your home? _____ Who? _____
5. Does anyone have hobbies at home involving chemicals or industrial materials? If so, describe. _____
6. Does anyone bring dust or other materials home on clothes from work? If yes, what? _____

E-FIGURE 19-1. A questionnaire for use during the "intake" of new patients for ongoing primary or specialty care. The instrument may be self-administered or supervised by a medical extender; it should then be reviewed by the patient and physician and periodically updated.

authorities or by consultation with specialists who should know the levels of most workplace hazards in the community. Finally, with an appropriate understanding of the limits of testing and awareness of "timing" issues in relation to exposure (as with measuring drug levels), an increasing number of hazardous chemicals can be biologically measured in blood or urine. Reliable testing is currently available for most metals and some pesticides, and testing may become available for a broad array of organic chemicals in the foreseeable future. Random sampling for "unknowns" is rarely helpful and most often leads to erroneous inferences because trace chemicals are ubiquitous, that is, almost everyone will have a higher than average level of "something."

OCCUPATIONAL AND ENVIRONMENTAL HEALTH DISORDERS COMMON IN PRACTICE

Although almost any medical complaint or condition could in theory have an occupational or environmental cause or contribution, certain conditions encountered in medical practice *commonly* do (Table 19-1). For these conditions, attention to the history is most important and most often rewarding.

Asthma

Atopic men and women with preexisting airways disease tolerate irritants in the workplace poorly and may experience exacerbations in temporal relation to one or more exposures. More important, numerous antigens are extant in the workplace, from large proteins, such as latex and animal danders, to small molecules, such as isocyanates needed to set polyurethane. More than 250 agents have been well characterized, and many others are suspect. Virtually no profession or work is immune, and up to 20% of all adult-onset asthma may have a work component. Presentation is often nonspecific; timing of symptoms during or slightly staggered from exposure is the clue to diagnosis, keeping in mind that there may be a lag of several hours between exposure and cough or other symptoms. The reward for early recognition of such causes is the likelihood that airway inflammation will abate when the noxious exposure is eliminated[1,2]; otherwise, lifelong, often generalized asthma is the rule (Chapter 87).

Chronic Interstitial, Parenchymal, and Inflammatory Lung Disorders

The rounded opacities of silicosis (Chapter 93) and coal workers' pneumoconiosis (Chapter 93) radiographically resemble sarcoid (Chapter 95); chronic beryllium disease (Chapter 93), a granulomatous disorder caused by sensitization to this widely used light metal, is clinically identical to sarcoid in almost all respects, but a reasonably specific test for blood and bronchoalveolar lavage fluid is now available to distinguish them. Asbestosis (Chapter 93) is identical to idiopathic pulmonary fibrosis (Chapter 92) in every clinical way except that benign pleural changes often accompany asbestosis, and asbestosis tends to be more indolent and usually stops progressing when exposure ceases or within a few years thereafter. Hypersensitivity pneumonitis (Chapter 93) is rarely suspected outside of agricultural settings but is occurring far more often; the causes are likely to be microbial contaminants of work materials, but some chemicals, such as the isocyanates, may also be causal. Occupational constrictive bronchiolitis, which can cause indolent or rapidly progressive dyspnea, is seen after exposure to a variety of noxious chemicals.[3] Recently, manufacturing and inorganic chemicals have been associated with outbreaks of allergic alveolitis, and synthetic fibers and food flavorings have precipitated severe and sometimes fatal airways responses. These observations support careful investigation of the environment in all cases of adult-onset lung disease.

TABLE 19-1 COMMON OCCUPATIONAL AND ENVIRONMENTAL HEALTH CONDITIONS IN GENERAL PRACTICE

CONDITION	EXPOSURE SETTINGS	COMMENT
Asthma	Virtually any indoor or outdoor workplace	New-onset, recrudescent, or exacerbated asthma
Interstitial, parenchymal, and inflammatory lung disorders	Dusts, metals, and organic materials	All parenchymal disorders have one or more environmental causes
Cancers of the respiratory tract	Asbestos, radon, silica, combustion fumes, tars, and some metals	Smokers are more likely to be affected
Sensorineural hearing loss	Noise, metals, and solvents	High-frequency loss, especially in younger workers
Musculoskeletal disorders of trunk and limbs	Heavy or repetitive activities or postures	Cold, vibration, and work stress contribute
Upper airway irritation	Dust and fumes	More common in smokers and atopic persons
Nonspecific building-related illness	Office work	Must exclude *specific* causes
Dermatitis, allergic or irritant	Repeated exposure to unprotected skin	Work and environmental exposures should be considered in every case
Multiple chemical sensitivities	Any	Complication of adverse environmental exposure

Cancers of the Respiratory Tract

Although most carcinomas of the lung and upper airway occur in smokers, occupational exposures to asbestos, silica, and the polyaromatic hydrocarbons in particulate air pollution, diesel exhaust, pitch, and asphalt contribute to the burden, as do radon and carcinogenic metals such as chromium and nickel found in most alloys (Chapter 191). Some organic materials, such as formaldehyde, are also likely culprits. Until there is an established strategy for secondary prevention, patients with these exposures should be observed expectantly; at a minimum, extraordinary efforts should be made to control smoking in these exposed individuals. Asbestos-exposed workers—smokers or otherwise—are additionally at risk for malignant mesothelioma (Chapters 99 and 191), but other than primary prevention, the only clinical implication is awareness for early diagnosis and compassionate care for this still largely incurable industrial disease.

Fatty Liver

With the widespread use of abdominal imaging, fatty liver has been recognized as more common than previously thought (Chapter 152). This disorder is common among individuals exposed regularly to organic solvents, a possibility that should be considered at the same time that infectious, metabolic, and pharmaceutical causes are considered. Once it is suspected, whether or not other factors are also present, chemical exposure should be reduced. Improvement tends to be slow, often during a period of many months, but the risk of progression is likely to have been averted or at least diminished.

Sensorineural Hearing Loss

Aside from aging, noise is the most important cause of high-frequency sensorineural hearing loss, recognizable as early as in adolescence (Chapter 426). Hobbies such as shooting and loud music may combine with industrial and agricultural noise to accelerate hearing loss. Although it is the responsibility of employers to conduct routine audiograms and to control exposure, clinicians should test noise-exposed patients periodically and reinforce whatever control strategies may be in place at work. Exposure to metals such as lead and organic solvents may compound the risk further.

Musculoskeletal Disorders of the Upper Extremity and Trunk

The most common cause of work disability, including permanent disability, is an injury to the back (Chapter 400) or upper extremity; the annual loss to the U.S. economy from disability and health care expenditures is estimated at a staggering 1 to 2% of the gross domestic product. Repetitive, heavy, awkward, and time-pressured activities are notorious contributors, as are cold and vibration.[4] A majority of cases, however, occur in workers without extremely physical jobs, such as health care or other service workers. Although an anatomically localized lesion may be identified and specifically treated in a small fraction of cases, as in carpal tunnel syndrome (Chapter 420) or thoracic outlet obstruction, the most important modalities of care in most cases are *early recognition* and *reduction of further insult*.[5] Physical therapy and medications may hasten recovery but cannot prevent recurrences and even progression unless the causal work and avocational activities are modified. Employers, who are increasingly familiar with these ergonomic issues, share an incentive to modify tasks or work stations.

Upper Airway Irritation

Virtually any smoke, fume, dust, or chemical has potential to irritate the upper respiratory tract (Chapter 93), causing acute or chronic symptoms indistinguishable from common allergic manifestations (Chapter 249) or upper respiratory infections (Chapter 96). Although the mucosae of the eyes, nose, sinuses, and throat tend to be forgiving, recurrent episodes are extremely nettlesome and cause substantial work disability. Atopic patients and patients with frequent infections are often the most sensitive to these ubiquitous environmental insults, which must ultimately be addressed along with the symptoms themselves and secondary infections.

Dermatitis

Erythematous rashes are a common consequence of topical exposures to workplace, avocational, and household materials, including latex, plastics, and many foods (Chapter 440). Although the keys to recognition are timing and the anatomic relation to clothing, allergenic and irritating chemicals can find their way into unlikely places, such as the groin and belt lines. Specialty consultation and patch testing are warranted in intractable cases but should not supplant careful observation and history taking in most situations. Acneiform lesions and folliculitis (Chapter 440) are also often caused by chemical irritation or physical factors at work, such as heat, pressure, or friction.

Sick Building Syndrome and Nonspecific Building-Related Illness

The effort to reduce the influx of "fresh" air into buildings to save heating and air-conditioning costs has resulted in upper airway and dermal irritation as well as vague central nervous system symptoms such as headache and fatigue, occurring shortly after beginning work and clearing minutes to hours after leaving the affected building. Many occupants are typically affected, especially those who spend the most time in one place. The cause is unknown, but recent evidence suggests that microbial materials may be the most common culprits. In every instance, a search for a specific allergen or irritant is worth undertaking (Chapter 249), but the most remedial sources are poor overall ventilation and dampness in which molds fester. When the cause is remedied, most building occupants typically experience symptomatic improvement. From a clinical perspective, the major consideration is whether any more serious problem, such as asthma, may have also developed.

Multiple Chemical Sensitivities

An environmental illness as transient as a single noxious inhalation or as persistent as a protracted course of nonspecific building-related illness can initiate a cycle of similar symptoms after exposures to odors or irritants at very low levels, thereby rendering everyday tasks such as shopping or driving problematic. A patient typically complains of feeling "allergic" to everything, although there is no evidence for allergic mechanisms (Chapter 249); the cause of this vexing complication, most prevalent in women and also seen in veterans of conflicts in the Middle East, is unknown and may involve psychological as well as physiologic factors. Despite the severity of complaints, which often include fatigue, muscle pain, stridor, chest tightness, and palpitations, laboratory test results are normal; many patients will meet clinical criteria for fibromyalgia (Chapter 274). Coexistent anxiety and depression often prompt psychiatric referral (Chapter 397), but the disorder has proved relatively refractory to all treatment modalities. Sympathetic support, environmental modification as needed to provide some symptomatic relief, and candor regarding the unknown nature of the disorder are appropriate; extensive clinical investigations often serve only to reinforce the patient's "sick" role and are best avoided. Despite all efforts, the most severely affected individuals will often seek the care of alternative practitioners (Chapter 39) with compelling if unproven theories and expensive, potentially harmful remedies.

COMMON HAZARDOUS EXPOSURES IN THE WORKPLACE AND AMBIENT ENVIRONMENT

Tens of thousands of chemicals in the workplace as well as important physical and biologic hazards may be encountered in the general environment (Table 19-2). Several of these hazards are of major current concern in industrialized countries.

TABLE 19-2 COMMON HAZARDS IN THE WORKPLACE AND AMBIENT ENVIRONMENT

HAZARD	HEALTH EFFECTS OF GREATEST CONCERN	COMMENTS
Metals	Neurotoxicity, cancer	Most can be measured in blood or urine to assess dose
Organic solvents	Respiratory and dermal irritation, neurotoxicity, hepatotoxicity	Benzene and a few others have unique effects
Organohalides (e.g., DDT, PCBs)	Cancer	Ubiquitous suspect carcinogens of high population concern
Herbicides and pesticides	Rare acute neurotoxicity, unknown long-term effects	Widespread hazards of high population concern
Electromagnetic radiation	Leukemia, glioblastoma	Ubiquitous exposures with unproven effects
Particulate matter	Acute and chronic atherosclerotic cardiovascular disease	Air pollution, workplace
Mold	Allergy	High population concern regarding putative chronic effects
Mineral dusts	Cancer	Old hazards still of high concern (e.g., asbestos, silica)

DDT = dichlorodiphenyltrichloroethane; PCBs = polychlorinated biphenyls.

Metals

Exposures to lead and arsenic (Chapter 22), once commonplace in industry, are now generally controlled; concern remains highest for environmental settings, especially for children. There is now greater concern for mercury—entrained in large ocean fish worldwide—and manganese, a potent neurotoxin that is found in welding fumes and various alloys and that affects extrapyramidal and autonomic function. For most metals—manganese being a notorious exception—blood or urine tests are available to quantify a patient's burden, but these tests must be mindful of timing, the form of metal, and possible "confounders," such as the largely benign form of arsenic excreted in urine for several days after even a single shellfish meal.

Organic Solvents

These petroleum derivatives remain ubiquitous in workplace and household products. All are irritating, potentially neurotoxic, and, to varying degrees, hepatotoxic (Chapter 110). Several more serious toxins, such as trichloroethylene and *n*-hexane, are no longer widely used. Benzene and the ethers of ethylene glycol are bone marrow toxins (Chapter 165).

Organohalides

Although these complex organic pesticides and industrial materials are no longer made and sold in developed countries, their remarkable biopersistence has resulted in entrainment into everyone's fat. Worse, the dread byproduct dioxin, once associated with herbicide manufacture, has now been recognized as a predictable consequence of combustion of any chlorine-containing materials. All are suspect carcinogens, although debate remains whether this effect is limited to soft tissue sarcomas (Chapter 202)—a relationship established for dioxin—or promotes cancers more globally. Some toxicologic and epidemiologic evidence links this class of agents to type 2 diabetes mellitus and dyslipidemias.

Herbicides and Pesticides

The acute neurotoxicity and irritant properties of most herbicides and pesticides have been well studied (Chapter 110). These agents are generally well controlled, although both occupational and residential overexposures occasionally occur. In developing countries, these substances remain a widespread vehicle for both suicide and homicide.

Nonionizing Electromagnetic Radiation

Electric wires, appliances, and, notoriously, cell phones emit low-frequency electromagnetic radiation at levels far below those that cause local thermal injuries (Chapter 20). These radiations are nonionizing, but there is some

epidemiologic evidence of an increased risk of childhood leukemia with high-level exposure from household wiring and of excess brain tumors in adult workers with regular exposures. These data are difficult to interpret because study results differ according to how exposure is assessed; the only conclusion is that there is basis for concern and need for further study but not cause for widespread alarm or action other than precaution in the placement of new heavy power lines near schools and residences.

Particulate Matter

Evidence accumulated in the past decade points to the likelihood that ambient air pollution contributes measurably to the population risk of cardiovascular disease. Focus has turned from the well-established respiratory irritants—the gases sulfur dioxide and ozone—to the smallest particles, so-called $PM_{2.5}$. These particles may be laden with polyaromatic hydrocarbons from diesel exhaust, coal burning, and industrial sources, which are proinflammatory. This risk also accrues to more heavily exposed industrial workers, although it remains unclear if the risk is associated with particles of any origin or just those that evolve from combustion.

Mold

Molds are ubiquitous and long known for their unpleasant odors and potential for inducing allergic responses (Chapter 249), including asthma. Recently, concern has arisen over the potential for serious effects from various mycotoxins, long problems in veterinary medicine when domestic animals consume contaminated feed; however, a consensus panel concluded that there is no evidence of human risks beyond those well established from living or working in a moldy environment. Mold formation should be prevented wherever possible, especially in schools and offices, where molds contribute to problems with indoor air quality. Identification, with eradication of leaks and other sources of water accumulation, is key.

Mineral Dusts

Although asbestos has been largely abated, silica and human-made mineral fibers remain widely distributed in the environment. Silica (Chapter 93), present in virtually every form of "rock," is a potent cause of lung injury and cancer, so respiratory exposure should be carefully controlled in every setting. The evidence of serious risk from fibrous glass, mineral wool, and other human-made mineral fibers is less clear; probably only the finest fibers, such as slag wool, have cancer-causing potential, but many are potent dermal and upper respiratory irritants and should be well controlled for that reason alone.

SUMMARY

Occupational and environmental health problems remain prevalent, although their spectrum and nature have changed as rapidly as any in medicine and are likely to change even faster as technology, work, and knowledge continue to evolve. Physicians need not necessarily develop a large base of *facts*—themselves subject to revision frequently—but rather an *approach* that incorporates key elements and provides a foundation for efficient recognition and management of current and future clinical syndromes.

Grade A Reference

A1. Schaafsma FG, Whelan K, van der Beek AJ, et al. Physical conditioning as part of a return to work strategy to reduce sickness absence for workers with back pain. *Cochrane Database Syst Rev*. 2013;8:CD001822.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Fishwick D. Work aggravated asthma; a review of the recent evidence. *Br Med Bull*. 2014; 110:77-88.
2. de Groene GJ, Pal TM, Beach J, et al. Workplace interventions for treatment of occupational asthma: a Cochrane systematic review. *Occup Environ Med*. 2012;69:373-374.
3. Kreiss K. Occupational causes of constrictive bronchiolitis. *Curr Opin Allergy Clin Immunol*. 2013;13:167-172.
4. Kapellusch JM, Garg A, Boda S, et al. Association between lifting and use of medication for low back pain: results from the backworks prospective cohort study. *J Occup Environ Med*. 2014;56:867-877.
5. Dick FD, Graveling RA, Munro W, et al. Workplace management of upper limb disorders: a systematic review. *Occup Med (Lond)*. 2011;61:19-25.

REVIEW QUESTIONS

1. A 35-year-old carpenter complains of new-onset headaches shortly after beginning a new job in a small boat-refurbishing shop. His job involves cleaning debris from the hulls of sailboats and resurfacing them. What is the most useful approach to establish whether some exposure or condition in the shop is causing his headaches?

A. Proceed as with any new headache evaluation. Then, once a diagnosis is made, explore online toxicologic databases to see if any chemicals are known to cause such a headache pattern.
B. Ask the patient to obtain a complete list of all chemicals and processes in the shop and research each to see if it can cause headaches.
C. Question the patient closely about the temporal relationship between work activities and symptoms.
D. Test the patient's blood and urine for common industrial solvents likely to be used in his work.
E. Call the employer to inquire if other workers have made similar complaints in the past.

Answer: C For the assessment of acute symptoms in relation to an environmental exposure, the history of temporal relationship between exposure and onset/cessation of symptoms is the most compelling basis for pursuit of a work connection for a new complaint. If headaches typically occur after arriving at work or starting a specific activity and are absent at other times, the physician should obtain a detailed list of chemicals or other potential culprits and explore their toxicology.

2. Which of the following is true regarding the relevance of "dose of exposure" to a workplace or environmental chemical?

A. Unlike pharmaceuticals, environmental toxins typically cause injury in susceptible people irrespective of the actual dose.
B. Unlike with pharmaceuticals, the possible doses to which a worker might be exposed at work range over many orders of magnitude.
C. Unlike with pharmaceuticals, there is no simple way of estimating exposure dose from talking to a patient; the workplace or exposure environment itself must be evaluated.
D. Unlike pharmaceuticals, environmental toxins typically affect everyone exposed in a similar way.
E. Unlike with pharmaceuticals, blood and urine levels are the mainstay for assessing exposure dose to chemicals.

Answer: B For drugs and toxins, dose matters. What differs is the sheer number, probably 100,000 or more, of different chemicals and the enormous range of possible exposure doses, from trivial parts per billion or trillion in a well-controlled environment to gram quantities in an unventilated workplace or in an accidental release. However, as noted in the text, strong inference about the range of exposure dose can be made from talking to the patient; directly measuring the environment and biomarkers of exposure is rarely the first-line approach. As with drugs, idiosyncratic responses are common, and no two people respond alike.

3. Which of the following is true regarding occupation-related asthma?

A. Nonatopic individuals with previously normal respiratory health may develop de novo sensitivity to one of hundreds of workplace proteins, metals, or other chemicals.
B. Smokers and atopic individuals are the ones most likely to develop airway reactions at work. Cough or wheeze in the presence of one or more such risk factors demands *increased* vigilance about workplace exposures.
C. Workplace factors cause or exacerbate asthma in up to 20% of adult-onset asthma cases.
D. A and C
E. A, B, and C

Answer: E Work exposures commonly cause airway lability in susceptible individuals, sometimes by primary sensitization in atopic or nonatopic men and women and sometimes by irritating already abnormal airways in smokers and patients with preexisting airway disease. Most studies estimate that almost one in five newly symptomatic adult asthmatics has an occupational or environmental cause, so this possibility must always be explored because an overlooked and preventable exposure may lead to lifelong, irreversible injury.

4. A 60-year-old woman is admitted to the hospital with new onset of a large pleural effusion and incapacitating chest pain. What is the best approach to determine if a work exposure may have been involved?

A. Perform a detailed occupation and environmental history of *current* exposures on admission to identify a possible cause for malignant or nonmalignant disease consistent with the presentation, such as asbestos.
B. Perform a detailed occupation and environmental history of *past* exposures on admission to identify a possible cause for malignant or nonmalignant disease consistent with the presentation, such as asbestos.
C. Proceed with the evaluation as you would for any patient with the observed constellation of signs and symptoms. Take a detailed environmental history of current and past exposures *after* a pathologic diagnosis has been made, focusing on established causes of that diagnosis, for example, lung cancer.
D. After the evaluation, ask the laboratory to analyze surgical and other specimens for evidence of fibers or metals that are known to cause the patient's diagnosed condition.
E. After the diagnosis is made, seek information from the local health department or similar agencies about the occurrence of similar cases in the area because patients are unlikely to recall what they were exposed to years before.

Answer: C In the assessment of a new-onset chronic disease, it is not efficient to take an exhaustive history of either current or former exposures as the evaluation will quickly reveal a clear target (diagnosis) on which to focus. Although recognition of a prominent work exposure, such as asbestos, might make the possibility of malignant mesothelioma more likely in this case, it is still necessary to make a pathologic diagnosis before worrying about its cause. This approach is in contrast with the approach to new onset of acute or recurring symptoms (question 1). D and E are important in some cases, but the first and most direct approach is to inquire whether the patient may have been exposed to the (invariably) small number of substances known to cause the just-diagnosed condition. These lists can easily be found in reference texts or online search tools (e.g., "occupational causes of lung cancer").

5. Which of the following is true of musculoskeletal disorders among working-age patients?

A. Because these disorders are most often associated with nonspecific findings on examination or imaging studies, they rarely lead to significant dysfunction or disability.
B. Now that most work in developed economies is less physical, musculoskeletal disorders have become relatively rare.
C. The best approach to prevent long-term disability is early recognition, intensive evaluation, and definitive treatment where possible.
D. Assessment and modification of offending work activities are more important in the long run than is a definitive anatomic diagnosis.
E. When a musculoskeletal disorder is recognized, the best way to establish whether it is related to work activity is to visit the workplace or to call the employer.

Answer: D Musculoskeletal disorders remain extremely common and extremely disabling even in highly developed economies and in jobs—like health care, food service, and education—that are not intensely physical. The key to diagnosis is the relationship between physical activities at work, which the patient can easily demonstrate to the physician, and the pattern of symptoms and signs. Intervening in the work by job modification or change is far more valuable in the long run than finding a drug or surgery to control the symptoms, which often recur when the patient returns to the old environment and tasks.

20

RADIATION INJURY

DAVID J. BRENNER

IONIZING RADIATION

DEFINITION

Ionizing radiations include x-rays, gamma rays, beta particles, alpha particles, and neutrons. Unlike non-ionizing radiations, including radio waves, microwaves, infrared, visible light, and ultraviolet (UV) radiations, these radiations have enough energy to knock electrons out of atomic or molecular orbits, thereby breaking chemical bonds in biomolecules such as DNA.

A key difference among the various ionizing radiations is their ability to penetrate matter (Fig. 20-1). For example, alpha particles have very limited range, measured in micrometers. Because alpha particles cannot penetrate skin, their health significance is entirely in the context of internal exposure by inhalation or ingestion of alpha-emitting radioactive materials. Beta particles, which have intermediate ranges, can typically be stopped by a piece of paper. By contrast, x-rays, gamma rays, and neutrons are highly penetrating and are hard to shield.

Ionizing radiations are emitted by radioactive materials, which can either be naturally occurring or can be man-made, using machines such as nuclear reactors. The amount of ionizing radiation emitted by a radioactive material is determined by its activity, measured in becquerels (1 Bq = 1 radioactive disintegration/second) or Curies (Ci).

Radioactivity should be distinguished from absorbed dose, which is a measure of how much ionizing energy is actually absorbed by a structure of interest, such as a person or an organ. The basic measurement unit of absorbed dose, which is the energy deposited per unit mass, is the Gray (1 Gy = 1 J/kg). *Equivalent dose* takes into account the fact that not all types of radiation are equally effective, *effective dose* takes into account the fact that not all organs in the body are equally sensitive to ionizing radiation, and *collective dose* is a measure of the effective dose delivered summed over a whole exposed population (Table 20-1).

EPIDEMIOLOGY

Radiation exposure can come from natural sources, therapeutic or diagnostic medical exposures, or accidental exposures to individuals or to large populations. Natural sources and radiologic examinations are the major contributors to the overall collective radiation dose to the U.S. population (Fig. 20-2).[1]

Natural Sources of Ionizing Radiation

The largest naturally occurring source of radiation is from radon gas, which is an alpha-particle emitter. Radon is a constituent component of all rocks, particularly granite-type rocks. Radon is a member of a chain of radioactive elements that starts with naturally occurring uranium and radioactively decays from one element to the next; uniquely in this chain of elements, radon is a gas, which can emerge from the ground and result in high levels in the basement and higher floors of houses. Long-term exposure to radon gas can cause lung cancer because emitted alpha particles in the lung can damage cells in the bronchial epithelium. The three other main sources of naturally occurring radiation exposure are from terrestrial exposure (gamma rays emitted from naturally occurring radioactive materials in the ground), from space radiations (mainly intergalactic cosmic rays that penetrate through the atmosphere), and internal radiation (from naturally occurring radioactive elements in the body, such as potassium-40).

Medical Exposures

The single largest source of medical exposure is computed tomography (CT), which on average contributes about 25% of the collective dose to the U.S. population. This source of exposure is recent because CT has been commonly used only since the 1980s. Now, however, about 75 million CT scans are performed annually in the United States. In addition to CT, nuclear medicine, interventional fluoroscopy, and conventional radiography together constitute about 25% of the overall collective dose to the U.S. population.

Radiologic Accidents

Since the widespread introduction of nuclear power plants, major radiologic accidents have become an intermittent, although rare, source of radiation

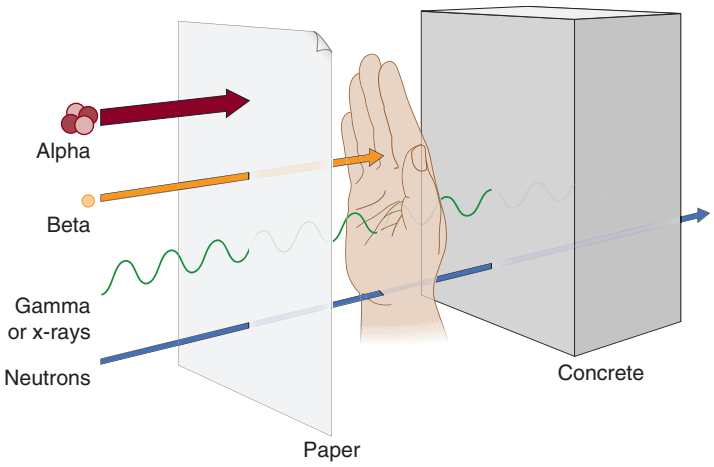

FIGURE 20-1. Schematic illustrating the relative ranges of alpha particles, beta rays, x-rays or gamma rays, and neutrons as they penetrate through matter.

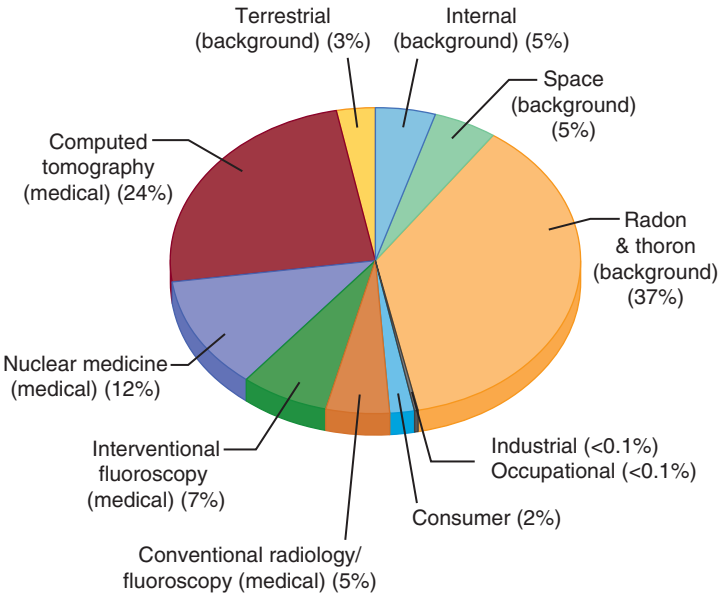

FIGURE 20-2. The major sources of radiation exposure to an average individual in the United States. (From Hall EJ. *Radiobiology for the Radiologist*. 7th ed. Philadelphia: Lippincott Williams & Wilkins; 2012.)

TABLE 20-1 STANDARD QUANTITIES AND UNITS ASSOCIATED WITH EXPOSURE TO IONIZING RADIATION

QUANTITY	DEFINITION	UNITS
Absorbed dose	Energy per unit mass	Gray (Gy)
Equivalent dose	Average dose × radiation-type-specific radiation weighting factor	Sievert (Sv)
Effective dose	Sum of equivalent doses to all exposed organs, each multiplied by an appropriate organ relative sensitivity factor	Sievert (Sv)
Collective dose	Sum of effective doses to a population of exposed individuals	Person-Sv

exposure. The accidents at Chernobyl in 1986 and at Fukushima in 2011 released far more radioactivity than any other accidents (Table 20-2).

Radiologic Terrorism

In the broadest of terms, two different types of such threats exist, one from an improvised nuclear device and the other from a radiologic dispersal device, sometimes called a *dirty bomb*.[2] The two are very different in terms of their potential consequences and also their likelihood of occurrence.

TABLE 20-2 ESTIMATED RELEASES OF RADIOACTIVE IODINE AND RADIOACTIVE CESIUM FROM THE FOUR HISTORICALLY BIGGEST NUCLEAR POWER PLANT ACCIDENTS

	^{131}I (TBq)	$^{137/134}$Cs (TBq)
Windscale, 1957	750	20
Three Mile Island, 1979	1	0.000001
Chernobyl, 1986	1,800,000	110,000
Fukushima, 2011	~500,000	~20,000

The consequences of exploding an improvised nuclear device, perhaps similar to those detonated at Hiroshima or Nagasaki, although more likely exploded at ground level, would be devastating in terms of loss of life and major injuries. For example, a 20-kT daytime surface improvised nuclear device explosion in midtown Manhattan might involve 2.2 million people exposed to high radiation doses, of whom perhaps only 1.6 million might survive.

By contrast, the consequences of exploding a radiologic dispersal device or dirty bomb, dispersing what could be very small or very large amounts of radioactive material, would likely involve a relatively small spatial area in which individuals are significantly exposed, as well as a much larger area in which individuals are exposed to very low levels of radioactivity. The significance of a dirty bomb depends on the amount of radioactivity in the device and the efficiency with which the explosion aerosolizes and disperses the radioactive material. Of course, the consequences of a radiologic dispersal device extend beyond health effects and could potentially involve large-scale social and economic disruption.[3]

PATHOBIOLOGY

Ionizing radiation can both mutate cells and kill them. The key initial damage site is a DNA double-strand break, although other types of damage are possible. This strand break is generally repaired by a variety of mechanisms, the most important of which is nonhomologous end rejoining. Imprecise nonhomologous end rejoining can lead to small mutations, whereas double-strand break misrepair (i.e., the wrong ends of breaks being rejoined to each other) can lead to chromosomal translocations, inversions, and telomere fusions.

Although chromosomal translocations, inversions, and point mutations are typically nonlethal lesions, such radiation-induced chromosomal aberrations can be the initial lesions associated with carcinogenesis. For example, the initial event in many hematopoietic cancers (Chapters 183 and 184) is often a recurrent translocation, which in turn results in the fusion of genes located at the translocation breakpoints. Radiation-induced point mutations and deletions also are often associated with loss of heterozygosity in tumor suppressor genes.

Misrepair of double-strand breaks can also induce dicentric chromosome aberrations and centric rings (Chapter 181), which are generally lethal lesions that cause reproductive cell death in which a cell loses its ability to divide. A second, quite different mechanism of radiation-induced cell death is apoptosis, or programmed cell death that, for example, is important after radiation exposure of hematopoietic cells and jejunal crypt cells.

CLINICAL MANIFESTATIONS

The consequences of radiation exposure are dose dependent and are also dependent on whether the exposure is to the whole body or only part of the body. After high doses of whole-body exposure, the major consequences are acute radiation syndromes, which may be fatal. After high doses of partial-body exposure, such as after radiotherapy, the consequences are highly organ specific. These high-dose tissue effects are deterministic, meaning that above a certain dose threshold, the probability of the effect rapidly increases to 100%, with the severity of the effect increasing with increasing dose. By contrast, the major concern at lower doses is with stochastic effects, in which there is no dose threshold, and it is the probability of the effect that increases with increasing dose. The major example here is radiation-induced cancer.

Acute Radiation Syndromes

The radiation dose needed to kill mitotically active dividing cells is far less than the dose needed to remove the functioning ability of differentiated cells. As a result, acute radiation syndromes generally relate to the failure of

precursor cells to provide replacements for functional cells within an organ, when the latter need replacement. Acute radiation syndromes typically progress through four stages: prodromal, clinical latency, manifest illness, and recovery or death.

Prodromal Stage

Prodromal (early) symptoms occur shortly after irradiation, with the radiation dose determining the severity, duration, and onset. Typical prodromal symptoms include nausea, vomiting, anorexia, fatigue, diarrhea, abdominal cramping, and dehydration. At very high doses, these symptoms can appear within a few minutes of exposure; but at somewhat lower doses, they may not appear for many hours, if at all.

Latency Period

After the prodromal stage, a latency period typically precedes manifest illness. This delay occurs because the subsequent syndromes (hematopoietic and gastrointestinal) involve the failure of relevant tissues to self renew, and manifest illness does not occur until after the typical cellular turnover times in these tissues.

Manifest Illness

The *hematopoietic syndrome* results from whole-body or significant partial-body exposure to 3 to 9 Gy. At these doses, radiation kills some or all of the mitotically active hematopoietic precursor cells. Symptoms then result from lack of circulating blood elements some 4 to 8 weeks later as circulating cells die off and are not replaced. At this point, the hematopoietic syndrome may manifest as infections and possible hemorrhage, impairment of immune mechanisms, and potentially multiple-organ failure.

The *gastrointestinal syndrome* usually occurs after whole-body or significant partial-body exposure of about 8 Gy and above. These doses lead to death of intestinal stem cells in the regenerating crypts. Symptoms then result when differentiated cells of the villi are naturally sloughed off, but stem cells are unable to produce new cells, thereby leading to depopulation of the epithelial lining of the gastrointestinal tract. After a latency period of about 7 days, loss of, or significant shortening of, the villi on the intestinal epithelium leads to bacterial growth and increased risk for sepsis. Common symptoms include anorexia, nausea, vomiting, prolonged bloody diarrhea, abdominal cramps, dehydration, and weight loss. Should death occur, it is typically 7 to 10 days later.

At extremely high radiation doses, death is typically from the *cerebrovascular syndrome* within a few days after exposure. The cause of death is believed to be because changes in permeability of small blood vessels in the brain lead to cerebral edema.

Organ-Specific Late Effects after Radiotherapy

Iatrogenic organ-specific late effects follow quite predictably after targeted radiation therapy in which the organ in question is irradiated, although the severity of the response is less predictable. These tissue effects can be divided into global organ responses and focal responses, in which a particular part of the organ is differentially irradiated (Table 20-3).

As one example, significant radiation exposure to the heart increases the risk for subsequent major heart disease. These risks have been most notable in women who have had radiotherapy for left-sided breast cancer (Chapter 198),[4] patients treated with mantle radiation for Hodgkin lymphoma (Chapter 186), and some patients treated for lung tumors adjacent to the heart.[5] The effects to the heart can be wide ranging, including to the coronary arteries, the myocardium, the heart valves, and the pericardial sac. The increased risk, which is proportional to the cardiac dose, begins within a few years after exposure and continues throughout life. These radiation risks appear to be multiplicative of the natural background cardiac risks, so patients with preexisting cardiac risk factors are likely to have greater absolute increases in their risk for radiation-induced cardiac disease.

At lower radiation doses, good evidence indicates that atomic bomb survivors have an increased lifetime risk for heart disease as well as an increased risk for stroke.[6] Among atomic bomb survivors, radiation-induced heart disease and stroke are estimated to account for about one third of the radiation-induced excess deaths compared with cancers.

TABLE 20-3 TYPICAL ORGAN-SPECIFIC DETERMINISTIC LATE EFFECTS AFTER HIGH-DOSE RADIOTHERAPY

ORGAN	LOCALIZED END POINT	GLOBAL END POINT
Brain, cranial nerves	Focal weakness, vision loss	Neurocognitive deficit
Lung	Bronchial stricture	Shortness of breath
Heart	Coronary stenosis	Pericarditis, cardiomyopathy
Bladder	Bleeding	Urinary frequency, diarrhea
Bowel	Ischemia, bleeding	Enteritis

Adapted with changes from Rubin P, Constine LS, Marks LB, eds. *ALERT—Adverse Late Effects of Cancer Treatment. Vol. 1: General Concepts and Specific Precepts.* Heidelberg: Springer; 2014.

Stochastic Effects of Radiation

In contrast to the deterministic organ-based effects of radiation, ionizing radiation also produces stochastic effects, whereby the severity of the disease is not dose dependent, but the probability of an effect increases with increasing dose. These effects are largely due to radiation-induced mutations of target cells, as opposed to cell killing, with by far the most important effect being radiation carcinogenesis. Other potential radiation-induced stochastic health effects include deleterious radiation-induced genetic effects to subsequent generations.

Radiation Carcinogenesis

In general, radiation acts as a multiplier of natural background cancer rates, although there are exceptions, such as radiation-induced breast cancer, in which the radiation risks are largely independent of the natural background rates, and prostate cancer, which generally is not a radiogenic tumor. Most radiation-induced cancers occur many years after radiation exposure, typically in the same "cancer-prone" years (about 55 to 75 years of age), as do naturally occurring cancers. Exceptions include radiation-induced hematopoietic cancers and radiation-induced thyroid cancers, both of which typically occur within a few years of radiation exposure. Sensitivity to radiation-induced cancer is age dependent, with people exposed at younger ages having higher lifetime cancer risks.

Most of what is known quantitatively about the radiation-induced risks for cancer comes from studies of atomic bomb survivors, who had measurable increased subsequent risks for cancer.[6] At low radiation doses, for example, below 0.1 Gy, epidemiologic studies are difficult to perform and interpret because it is difficult to quantify a small extra risk for cancer above the natural human lifetime cancer risk of about 40%. More recently, large-scale epidemiologic studies in children who received CT scans suggest small but statistically significant increased risks for radiation-induced cancer, at least down to the radiation doses relevant to CT scans.[7,8]

Radiation Cataractogenesis

Ionizing radiation induces cataracts (Chapter 423) with a probability, severity, and latency related to the radiation dose.[9] Studies of atomic bomb survivors, Chernobyl cleanup workers, astronauts, and radiologic technicians have all shown increased incidence of cataracts, even at radiation doses below 1 Gy. Radiation cataractogenesis is a result of radiation damage or killing of dividing cells in the lens. Because there are essentially no mechanisms in the lens for removal of damaged cells, the resulting nontranslucent lens fibers migrate toward the posterior pole of the lens, where they represent the first stage of a cataract. Posterior subcapsular cataracts are less common than nuclear or cortical cataracts in the general population but are the type most commonly associated with ionizing radiation.

DIAGNOSIS

The diagnosis of radiation injury can be considered in two different contexts. The first context is when the radiation dose is not known, and the goal is to estimate the dose that was delivered in order to optimize treatment. Examples include accidents and radiologic terrorism scenarios. The second context is when the radiation dose is known, in particular after radiotherapy. Here the sequelae (see Table 20-3) are predictable, but the timing and the severity of the injury are less predictable.

Radiation Biodosimetry

For an individual exposed to a large whole-body radiation dose, the first critical task is to estimate the radiation dose that the individual received, because different syndromes (hematopoietic, gastrointestinal, cerebrovascular) are associated with different radiation dose ranges. The technique for estimating exposures to individuals is known as *radiation biodosimetry*. In this diagnostic process, a small amount of a body tissue, such as blood or urine, is analyzed,

and dose-dependent end points in the sample are measured to provide an estimated dose. The most common end point, which is chromosomal aberrations in blood samples, can now be assessed by high-throughput technologies.[10] Other practical radiation-dose dependent assays include gene expression in blood or induction of metabolites in urine.

Post-Radiotherapy Sequelae

Deterministic radiation-induced organ syndromes are seen at reasonably predictable intervals after radiation therapy. For example, radiation pericarditis (Chapter 77) can develop during treatment or months to years later, initially with an effusion and progressing to pericardial constriction. Radiation enterocolitis (Chapter 140) can develop 6 to 12 months after radiation doses of 40 to 60 Gy for prostate cancer or gynecologic malignancies. Radiation-induced pulmonary injury can be detected in up to 50% of patients who receive thoracic radiation therapy, but only a minority of them will develop clinical symptoms (Chapter 94). In addition, the majority of all radiotherapy patients experience acute skin or mucosal toxicity, in particular radiation dermatitis, oral mucositis, and xerostomia.

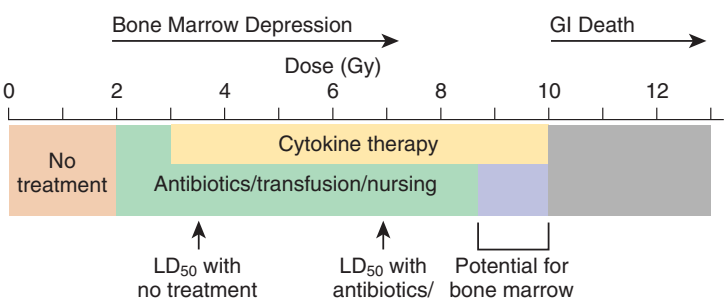

FIGURE 20-3. Schematic recommended treatment for acute radiation syndromes, as a function of the estimated dose that the individual received. Note the very narrow dose window for which bone marrow transplantations are potentially useful. (Adapted with changes from Hall EJ. *Radiobiology for the Radiologist*. 7th ed. Philadelphia: Lippincott Williams & Wilkins; 2012.)

TREATMENT Rx

Whole-Body Exposure Syndromes

For exposures below 2 Gy, no immediate treatment is needed (Fig. 20-3). For the hematopoietic syndrome, two primary treatment approaches are designed to reduce red blood cell depression, by using transfusions and cytokine therapy, and to minimize infection, by using antibiotics and possibly barrier nursing. Antibiotics and appropriate nursing care can raise the human LD_{50} (the dose at which 50% of exposed individuals die) from about 4 to 6 Gy, and the addition of human granulocyte colony-stimulating factor (e.g., filgrastim, 5 µg/kg daily for up to 2 weeks, subcutaneously or by IV infusion) may increase the LD_{50} to about 9 Gy.

The role of bone marrow transplantation in treating individuals with the hematopoietic syndrome is probably quite limited at doses less than about 9 Gy. In the aftermath of the Chernobyl accident, for example, 13 individuals underwent bone marrow transplantation, and three deaths were directly attributed to the sequelae of the procedure in patients in whom the transplantation was not indicated based on their exposure. Conservative treatment regimens, including antibiotics, cytokine therapy, transfusions, and nursing are clearly effective, but current treatment options are very limited at doses above about 10 Gy, when the gastrointestinal syndrome becomes the dominant cause of death.

Radiotherapy Sequelae

For radiotherapy-induced dermatitis, mucositis, and xerostomia, treatments include topical agents for dermatitis (Chapter 438), analgesics for mucositis, and saliva substitutes for xerostomia. Beyond dermatitis, mucositis, and xerostomia, radiation-induced deterministic tissue effects are typically treated in the same way as their non-radiation-induced counterparts. For example, the treatment of radiation-induced coronary artery disease is no different than the approach to coronary diseases in general (Chapters 71 to 74). However, the management of acute and chronic gastrointestinal symptoms after radiotherapy requires a more specialized approach because of the risk for local stricture (Chapter 142).[11]

PREVENTION

Countermeasures after Radiologic Accidents or Terrorism

Three FDA-approved pharmaceuticals can limit the radiation dose caused by exposures to specific radioisotopes: potassium iodide for radioactive iodine, Prussian blue for radioactive cesium or thallium, and diethylenetriamine pentaacetate (DTPA) for plutonium, americium, or curium. Although these pharmaceutical countermeasures can limit the organ doses and therefore the health consequences produced by these specific radioisotopes, they cannot be used to treat any of the adverse clinical effects caused by them.

The administration of potassium iodide (KI), which contains nonradioactive iodine, saturates the thyroid with nonradioactive iodine and thereby prevents radioactive iodine being absorbed into the thyroid, where it can cause thyroid cancer (Chapter 226). The protective effects of KI last about 24 hours, so it should be taken daily (130 mg for adults, 65 mg for children >3 years of age, 32 mg for children 1 to 3 years of age, 16 mg for infants) while there is a likelihood of exposure.

Prussian blue (ferric hexacyanoferrate [II]) binds cesium in the gastrointestinal tract and limits its reabsorption into the blood stream, thereby reducing the biologic half-life of cesium in the body from about 110 days to about 30 days and limiting the organ toxicities that it produces. The recommended dose of Prussian blue is 3 g orally three times daily for adults and 1 g orally three times daily for children 2 to 12 years of age, for a minimum of 30 days.

DTPA is a chelating agent that binds to radioactive plutonium, americium, and curium, thereby decreasing their persistence in the body. DTPA comes in two forms, Ca-DTPA and Zn-DTPA, with the calcium form considered more effective over the first 24 hours of exposure, and the zinc form preferable at later times, although both are effective over several weeks. They are administered intravenously (or with a nebulizer), with a daily intravenous dose of 1 g for adults and 14 mg/kg for children younger than 12 years.

Medical Imaging

Because CT is such a superb diagnostic tool and because individual CT risks are small, the nearly 90 million CT scans performed annually in the United States generally provide far more benefits than risks. However, organ doses from CT are typically far larger than those from conventional radiographic examinations, so CT should operate under the principle of using as low a dose as reasonably achievable. The first opportunity is to reduce the radiation dose per scan, so-called optimization, with improved technology. The second opportunity is justification—making sure that all CT scans are clinically justified. A substantial fraction of CT scans could potentially be replaced by other modalities, such as magnetic resonance imaging for some head examinations[12] and selective use of ultrasound for diagnosing appendicitis.[13] In addition, some CT scans could be avoided entirely by avoiding CT scans when imaging is not clinically necessary.

Radiation Therapy

Amifostine (200 mg/m^2 given intravenously daily, before each fraction of radiation therapy) has been approved by the U.S. Food and Drug Administration (FDA) for minimizing radiotherapy-associated xerostomia, but its use has been limited by significant toxicity. Several drugs have been designed to mitigate or treat the damage caused by radiation therapy, particularly to the mucosal barrier. Palifermin, which is a recombinant human keratinocyte growth factor, is FDA approved (60 µg/kg/day administered as an IV bolus injection for 3 consecutive days before therapy and 3 consecutive days after therapy for leukemia or lymphoma) to decrease the incidence or duration of severe oral mucositis.

NON-IONIZING RADIATION INJURY

Ionizing radiations uniquely have sufficient energy to cause DNA strand breaks, which in turn can lead to cell killing and mutations. By contrast, non-ionizing radiations, such as radio waves, microwaves, infrared radiation, and visible light, do not have sufficient energy to break DNA strands and have not been associated with health hazards. One exception is UV radiation, which is clearly a mutagen and a carcinogen. A second possible exception is radiofrequency radiation, although in this case the epidemiologic evidence and mechanistic considerations do not provide strong evidence of health risks.

Ultraviolet Radiation

UV radiations, which are part of the non-ionizing electromagnetic spectrum, cover a wavelength from about 150 to 400 nm (intermediate between visible light and x-rays). In turn, these UV wavelengths are subcategorized as UVA (400 to 320 nm), UVB (315 to 280 nm), and UVC (280 to 150 nm).

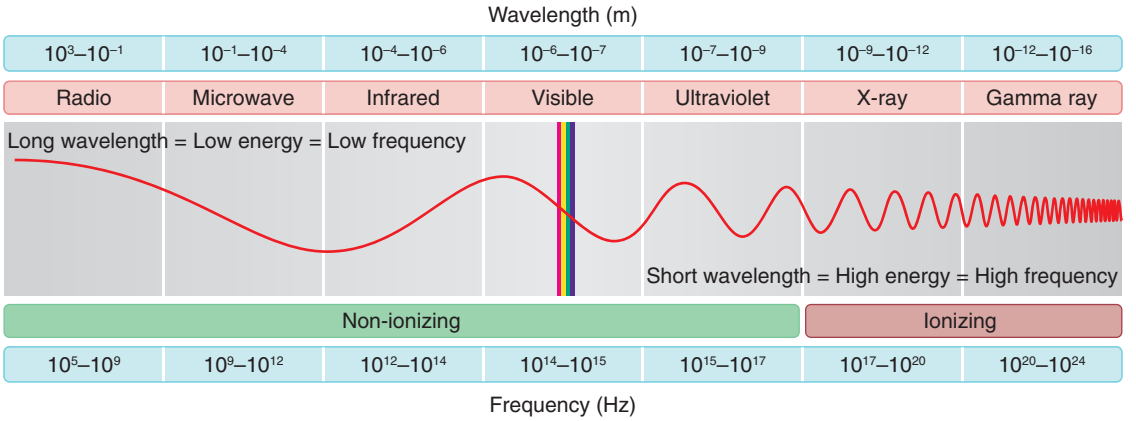

FIGURE 20-4. The electromagnetic spectrum.

Sunlight is by far the largest source of UV radiation to humans. Because of filtering in the atmosphere, about 95% of the UV spectrum that reaches the earth's surface is UVA, the remainder is nearly all UVB, and almost no UVC penetrates the ozone layer to reach the earth. Other sources of UV exposures include tanning beds and booths. Industrial exposures include welding arcs, plasma torches, germicidal and blacklight lamps, electric arc furnaces, hot-metal operations, mercury-vapor lamps, and some lasers.

PATHOBIOLOGY

UV radiations do not penetrate deeply into human tissues, so the injuries they cause are principally to the skin and eyes. The biologic effects of UV radiation are primarily attributable to its absorption in DNA, where pyrimidine dimers and their products are produced. These products are generally very efficiently removed, primarily through nucleotide excision repair but also through base excision repair, but errors in these repair processes ultimately can result in mutations. Although most UV radiation in sunlight is UVA, UVB is much more efficiently absorbed by nucleic acids, and most of the biologic effects of sunlight are associated with UVB.

The damaging effects of UV radiation are exacerbated in patients who have underlying defects in nucleotide excision repair. For example, studies of patients with xeroderma pigmentosum (Chapter 436) have provided key evidence of the link between UV light and skin cancer. In addition, the effects of UV are often enhanced in patients who are taking medications that contain photodynamic agents, such as tetracyclines, fluoroquinolones, nonsteroidal-anti-inflammatory drugs, diuretics, statins, phenothiazines, thioxanthenes, and antifungals. Such drug-related phototoxic reactions typically resolve after the drug is discontinued.

CLINICAL MANIFESTATIONS

The major acute effects of UVB exposure are sunburn and erythema, whereas UVA exposure is rarely the direct cause of sunburn. All ultraviolet radiations are well-established skin carcinogens (Chapter 203). Squamous cell carcinomas and basal carcinomas are typically associated with multiple or chronic exposures to sunlight, whereas melanoma is linked primarily to episodes of acute sunburn. Lighter skinned individuals, whose skin and eyes contain less melanin, are more prone to all UV effects because melanin acts as a photoprotector. Acute exposures of the eye to UVB (Chapter 423) can cause "welder flash" (photokeratitis), and chronic exposures are associated with cataracts as well as pterygium.

PREVENTION

Excessive exposure to sunlight or other sources of UV radiation should be avoided, especially in fair-skinned individuals. UV radiation–screening lotions or creams and UV radiation–blocking sunglasses should be used when appropriate.

During significant exposure to sunlight, it is important to use a broad-spectrum sunscreen that provides protection from UVB as well as UVA. Standard recommendations are use of sunscreen with a sun protection factor (SPF) of 30, with reapplication at least every 2 hours. Likewise, sunglasses should be rated to block 99% or 100% of UVB and UVA radiation.

Radiofrequency Radiation

Radiofrequency radiations are also part of the non-ionizing electromagnetic spectrum (Fig. 20-4). Examples are radiation emitted from mobile phones (800 MHz to 2 GHz) and radiation from electric power lines (50/60 Hz). Biologic studies of low-level radiofrequency exposure have shown slightly increased heating but no increase in DNA damage.[14]

Numerous epidemiologic studies have focused on the effects of electric power lines. Although some early studies showed an association between the proximity of residence to high-voltage power lines and the risk for childhood leukemia, these early studies have generally not proved repeatable.[15] Likewise, numerous studies of a possible relationship between cell phone use and head or neck tumors have not shown a convincing link. A recently launched prospective European study will follow 290,000 adult cell phone users over a 20-year period.

Grade A References

A1. Gu J, Zhu S, Li X, et al. Effect of amifostine in head and neck cancer patients treated with radiotherapy: a systematic review and meta-analysis based on randomized controlled trials. *PLoS ONE*. 2014;9:e95968.

A2. Le QT, Kim HE, Schneider CJ, et al. Palifermin reduces severe mucositis in definitive chemoradiotherapy of locally advanced head and neck cancer: a randomized, placebo-controlled study. *J Clin Oncol*. 2011;29:2808-2814.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Mettler FA Jr, Bhargavan M, Faulkner K, et al. Radiologic and nuclear medicine studies in the United States and worldwide: frequency, radiation dose, and comparison with other radiation sources—1950-2007. *Radiology*. 2009;253:520-531.
2. Anderson PD, Bokor G. Nuclear and radiological terrorism: continuing education article. *J Pharm Pract*. 2013;26:171-182.
3. Rogers MB, Amlot R, Rubin GJ. The impact of communication materials on public responses to a radiological dispersal device (RDD) attack. *Biosecur Bioterror*. 2013;11:49-58.
4. Brenner DJ, Shuryak I, Jozsef G, et al. Risk and risk reduction of major coronary events associated with contemporary breast radiotherapy. *JAMA Intern Med*. 2014;174:158-160.
5. Jaworski C, Mariani JA, Wheeler G, et al. Cardiac complications of thoracic irradiation. *J Am Coll Cardiol*. 2013;61:2319-2328.
6. Ozasa K, Shimizu Y, Suyama A, et al. Studies of the mortality of atomic bomb survivors, Report 14, 1950-2003: an overview of cancer and noncancer diseases. *Radiat Res*. 2012;177:229-243.
7. Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *Lancet*. 2012; 380:499-505.
8. Mathews JD, Forsythe AV, Brady Z, et al. Cancer risk in 680,000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. *BMJ*. 2013;346:f2360.
9. Kleiman NJ. Radiation cataract. *Ann ICRP*. 2012;41:80-97.
10. Repin M, Turner HC, Garty G, et al. Next generation platforms for high-throughput biodosimetry. *Radiat Prot Dosimetry*. 2014;159:105-110.
11. Andreyev HJ, Davidson SE, Gillespie C, et al. Practice guidance on the management of acute and chronic gastrointestinal problems arising as a result of treatment for cancer. *Gut*. 2012;61: 179-192.
12. Tahvonen P, Oikarinen H, Paakko E, et al. Justification of CT examinations in young adults and children can be improved by education, guideline implementation and increased MRI capacity. *Br J Radiol*. 2013;86:20130337.
13. Aspelund G, Fingeret A, Gross E, et al. Ultrasonography/MRI versus CT for diagnosing appendicitis. *Pediatrics*. 2014;133:586-593.
14. Waldmann P, Bohnenberger S, Greinert R, et al. Influence of GSM signals on human peripheral lymphocytes: study of genotoxicity. *Radiat Res*. 2013;179:243-253.
15. Pedersen C, Raaschou-Nielsen O, Rod NH, et al. Distance from residence to power line and risk of childhood leukemia: a population-based case-control study in Denmark. *Cancer Causes Control*. 2014;25:171-177.

21

BIOTERRORISM

MARK G. KORTEPETER AND THEODORE J. CIESLAK

The likelihood that an individual physician would be called on to respond to a bioterror or biocrime incident is remote, but primary care physicians play critical roles as the point of entry into the medical system and as consultants for emergency services for any potential victims. Therefore, it is important for internal medicine physicians to be familiar with potential epidemiologic clues that an attack has occurred, general clinical aspects of agents considered to be the greatest threats, and how to alert the appropriate public health resources if they suspect an event.

HISTORY

Over centuries, biologic pathogens have repeatedly been used as weapons of warfare, ranging from tossing a dead animal carcass or feces into an adversary's water supply to releasing infectious agents by aerosol spraying.[1] Although biologic weapons can be used by governments as agents of warfare, a bigger concern now is their use by terrorists on civilian populations.

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as necessarily reflecting the views of the Department of Defense, the United States Army, or the U.S. Government.

EPIDEMIOLOGIC CLUES

As with any other outbreak, the response to a potential act of bioterrorism relies on basic public health fundamentals. Although no single feature can be considered definitive, several features should raise the suspicion of an unnatural event (Table 21-1).

Victims of a chemical or conventional weapon release would likely become ill shortly after a release or explosion, whereas release of a biologic weapon would initially be silent because of its incubation period, with victims likely presenting for care in a delayed fashion to different care providers, dispersed in space and time. One way in which a bioterror attack might differ from a natural outbreak is in the unusual presentation of illness. For example, a disease that typically occurs on the skin or causes gastrointestinal illness might instead be manifested with respiratory features (e.g., anthrax, plague). Another clue could be illnesses that are unresponsive to standard therapies because of an unusual antibiotic sensitivity profile. Suspicion should also be raised by finding a disease outside the location or season in which it is typically found.

TABLE 21-1 EPIDEMIOLOGIC CLUES OF A BIOWEAPON ATTACK

A large outbreak with a similar disease or syndrome, especially in a discrete population
Many cases of unexplained diseases or deaths
More severe disease than expected for a specific pathogen or failure to respond to standard therapy
Unusual route of exposure for a pathogen, such as the inhalational route for diseases that normally occur through other exposures
A disease that is unusual for a given geographic area or transmission season, especially in the absence of a competent vector
Multiple simultaneous or serial epidemics
A single case of an uncommon agent (smallpox, some viral hemorrhagic fevers, inhalational anthrax, pneumonic plague)
Unusual strains or variants of organisms, or antimicrobial resistance patterns different from those known to be circulating
A similar or exact genetic type among agents isolated from distinct sources at different times or locations
Higher attack rates among those exposed in certain areas, such as inside a building after an indoor release, or lower rates in those inside a sealed building with an external release
Outbreaks of the same disease occurring simultaneously in noncontiguous areas
A zoonotic disease occurring in humans but not in animals
Direct evidence or intelligence of a release (equipment, munitions, tampering) or other potential vehicle of spread (spray device, contaminated letter)
A downwind pattern of casualty location

Modified from Dembek ZF, Alves DA, Cieslak TJ, et al. USAMRIID's Medical Management of Biological Casualties Handbook, 7th ed. September 2011. Found at www.usamriid.army.mil under reference materials tab. Accessed January 29, 2015.

AGENTS OF CONCERN

Among the myriad human infectious pathogens, a relatively small number possess the requisite properties to be considered potential weapons that could cause widespread disease, so-called weapons of mass destruction (Table 21-2). These agents typically remain stable in aerosol, which allows their relatively efficient spread, potentially over a large population.

ANTHRAX

EPIDEMIOLOGY

Anthrax, caused by infection with the gram-positive bacillus *Bacillus anthracis* (Chapter 294), is a worldwide scourge of herbivores. Most human cases have had direct contact with infected animals or their hides, hair, bone, or skins, although a recent outbreak occurred among injection drug users in the United Kingdom. Endemic areas tend to include underdeveloped parts of sub-Saharan Africa and Southeast Asia, where humans interact closely with animals that have not been vaccinated. Sporadic cases occur in the United States, typically along the trails of the historic cattle drives in the central plains. Animals are typically infected while grazing in areas contaminated with anthrax spores, which can survive in the soil for decades and are stable in a desiccated powder form. Although most cases worldwide are cutaneous, the spore form of the organism can be milled into the ideal particle size (2 to 6 μm) for infecting the lung and causing inhalational disease.

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

Exposure to spores can occur through a break in the skin, through the gastrointestinal tract, or by inhalation. Spores are taken up by macrophages and replicate in the local skin, or they can be transported to the regional lymph nodes in the gastrointestinal tract or lungs. While replicating, the bacilli secrete lethal toxin, which causes local necrosis, and edema toxin, which causes significant local edema. In untreated or unrecognized disease, the organism can cause bacteremia, systemic toxemia, and death.

Cutaneous anthrax can be readily recognized by the astute clinician on the basis of two clinical manifestations caused by its toxins: a black eschar, from which anthrax gets its name (from the Greek *anthrakis*, for coal), and surrounding edema that is out of proportion to the size of the lesion. The lesion starts as a papule that becomes a vesicle and eventually develops a central, black, necrotic area. Although this form of disease is readily treatable, it can lead to a 20% case-fatality rate if it is not recognized and treated appropriately. Gastrointestinal anthrax, which occurs after ingestion of infected, undercooked meat, is much more difficult to recognize and can have a case-fatality rate of 50% or more. Afflicted individuals develop fever, abdominal

TABLE 21-2 CENTERS FOR DISEASE CONTROL AND PREVENTION CATEGORY A BIOTERROR AGENTS

DISEASE (AND AGENT)	DIAGNOSTIC ASSOCIATIONS	WEAPONIZATION RATIONALE	TREATMENT (SEE TEXT FOR DETAIL)
Anthrax (*Bacillus anthracis*; Chapter 294)	Hemorrhagic mediastinitis	Highly lethal inhalational disease; stable spores survive desiccation; can be formulated as aerosol	Ciprofloxacin (400 mg IV q12h) or doxycycline (100 mg IV q12h) + clindamycin (600 mg IV q8h) + penicillin G (4 million units IV q4h)
Smallpox (variola virus; Chapter 372)	Synchronous exanthem	Virions stable in environment; population immunologically naïve	Supportive care; cidofovir and ST-246 are promising investigational drugs
Plague (*Yersinia pestis*; Chapter 312)	Hemoptysis	Contagious through respiratory droplets; "Black Death" conjures fear	Gentamicin (5 mg/kg IV qd) or ciprofloxacin (400 mg IV q12h) or doxycycline (100 mg IV q12h)
Tularemia (*Francisella tularensis*; Chapter 311)	Plague-like illness	Bacteria stable in environment; very low infectious dose	Same as for plague
Botulism (botulinum toxins; Chapter 296)	Descending, flaccid paralysis	Highly potent toxin; lends itself to food and water use; cases consume vast resources	Supportive care; ventilator support; botulinum antitoxin may halt (but will not reverse) symptom progression
Viral hemorrhagic fevers (Chapter 381; see Table 21-3)	Hemorrhagic diatheses	Fear factor is paramount; few countermeasures exist	Supportive care; ribavirin may be beneficial in select cases (e.g., Lassa fever, New World arenaviruses, Crimean-Congo hemorrhagic fever virus, hemorrhagic fever renal syndrome) when it is given under an experimental protocol: 30 mg/kg IV load, then 16 mg/kg q6h for 4 days, then 8 mg/kg q8h for 6 days

pain, diarrhea, hematochezia, hematemesis, and ascites. Paracentesis may yield hemorrhagic ascites.

Inhalational anthrax is of greatest concern after an aerosolized bioweapons attack. After an incubation period that averages 1 to 6 days but can be as long as 43 days, patients present with fever, profound drenching sweats, nausea, vomiting, diarrhea, shortness of breath, cough, and chest pain. Without prompt recognition and appropriate therapy, individuals can deteriorate rapidly with increasing dyspnea, stridor, cyanosis, and respiratory failure. Engorgement and hemorrhage of mediastinal lymph nodes with accompanying mediastinitis lead to the clinical hallmark of a widened mediastinum on the chest radiograph in about 60% of cases. Patients may also develop large hemorrhagic pleural effusions and infiltrates. Bacteremic dissemination can lead to widely metastatic infection in the gastrointestinal tract and meninges.

DIAGNOSIS

B. anthracis can be identified by Gram stain or culture of tissue or body fluids, including skin biopsy for cutaneous disease, peripheral blood, pleural effusions, and cerebrospinal fluid but not sputum. A chest radiograph may provide initial clues to inhalational disease with manifestations such as hilar or paratracheal fullness and pleural effusions. A chest computed tomography scan can confirm mediastinal adenopathy. Other techniques, such as fluorescent antibody staining or polymerase chain reaction (PCR), may provide more rapid diagnosis.

TREATMENT Rx

Antibiotics licensed for the treatment of inhalational anthrax include penicillin, ciprofloxacin, doxycycline, and levofloxacin, each for a total duration of 60 days (Table 21-2). After initial empirical therapy, antibiotic selection should be guided by sensitivity testing. In the 2001 outbreaks, patients who received more than one drug appeared to have better outcomes, so one of these drugs should be administered in combination with at least one other drug to which the organism is sensitive. Potential additional drugs include rifampin (300 to 600 mg orally once or twice daily), vancomycin (1 g or 15 mg/kg IV every 12 hours), ampicillin (2 g IV every 4 hours), chloramphenicol (500 mg IV or orally, every 6 hours), imipenem (1 g IV every 6 hours), clindamycin (600 mg IV every 8 hours), and clarithromycin (500 mg orally every 12 hours). Actual treatment doses and durations of combination therapy are based on clinical judgment or expert consultations.[2] A new monoclonal antibody, raxibacumab (single dose of 40 mg/kg IV during 2 hours and 15 minutes after premedicating with diphenhydramine 25 to 50 mg IV), is approved for additional therapy of inhalational anthrax along with antibiotics on the basis of animal efficacy data. Anthrax immune globulin remains an investigational product at this time.

PREVENTION

The anthrax vaccine, which is effective at preventing all forms of disease, has been reserved primarily for high-risk ranchers or veterinarians who have regular contact with herbivores, scientific personnel in research laboratories, and military personnel. The vaccine is now given as five doses intramuscularly during 18 months (day 0, followed by 1, 6, 12, and 18 months), with annual boosters if continued risk exists.

After a known exposure, chemoprophylaxis is recommended with ciprofloxacin (500 mg orally twice daily) or doxycycline (100 mg orally twice daily) for 60 days, although shorter durations are likely to be effective, especially if combined with vaccination. If the vaccine is available, three doses may be given concomitantly with antibiotics, but it is not licensed for this purpose and must be administered under an investigational protocol. Raxibacumab (dosed as before) is also approved for prophylaxis if antibiotics are not available.

PROGNOSIS

Factors associated with survival from anthrax include antibiotics or anthrax antiserum given during the prodromal phase, pleural fluid drainage, and a multiple drug regimen. Case-fatality rates in untreated patients approach 100%, although 55% of the victims of inhalation anthrax survived in a 2001 outbreak because of antibiotics and modern intensive care.

SMALLPOX

EPIDEMIOLOGY

Although smallpox was officially eradicated in 1980, concerns exist about undeclared stores outside approved repositories in the United States and Russia. The recent discoveries of unaccounted for vials of variola in the United States provide some credence to this concern. In addition to near-universal susceptibility, the environmental stability of smallpox viral particles makes it a formidable potential weapon.

Smallpox is caused by variola virus, an orthopoxvirus (Chapter 372) that is closely related to cowpox, vaccinia, and monkeypox. Variola generally spreads to household and other close contacts, but it can spread farther distances through aerosolized droplet nuclei as well as by direct contact with secretions of patients or fomites. Health care providers, who can be infected in the nosocomial setting, or unsuspecting individuals who have contact with contaminated objects are also at risk.

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

Infection initially occurs in the respiratory mucosa, followed by replication in the regional lymph nodes and then by an asymptomatic primary viremia that seeds the reticuloendothelial system. Approximately 1 week after infection (range, ~1 to 2 weeks), a secondary viremia seeds the skin. At this time, the sudden onset of illness includes fever, headache, backache, and vomiting. Within 2 or 3 days of this prodrome, and often as the temperature falls, the characteristic rash begins with small papules in a centrifugal pattern (more on the face and extremities than on the trunk) that progress in synchronous fashion during the next week to vesicles, umbilicated pustules, and finally scabs. Lesions are deep seated and can be intensely painful. Individuals are contagious at the onset of the rash and are considered to be free of contagion once the scabs have separated.

DIAGNOSIS

Chickenpox (varicella; Chapter 375), which can be confused with smallpox, occurs predominantly on the trunk rather than on the extremities (a centripetal pattern), and the lesions occur in successive crops, so macules, papules, pustules, and scabs can be seen simultaneously. In addition, varicella is transmissible before onset of the rash, but contagiousness abates when all the lesions are scabbed. Other diseases in the differential diagnosis of smallpox include monkeypox (Chapter 372), other poxviruses, disseminated vaccinia in a vaccine recipient or contact, disseminated herpes zoster (Chapter 375) or herpes simplex (Chapter 374), impetigo (Chapter 441), drug eruptions (Chapter 440), contact dermatitis (Chapter 438), erythema multiforme (Chapter 439), and rickettsialpox (Chapter 327).

Clinical recognition of the characteristic rash (Fig. 21-1) should raise suspicion of smallpox. Diagnosis would likely come from PCR or viral culture of skin or blood samples or from acute and convalescent serologies.

TREATMENT Rx

The primary therapy for smallpox remains supportive care. Potential therapies include cidofovir (5 mg/kg IV; duration is determined on the basis of clinical response and potential side effects), a product licensed for treatment of cytomegalovirus retinitis in HIV-infected patients, and the investigational drug ST-246, which has been added to the strategic national stockpile for use in case of a smallpox outbreak.

PREVENTION

Vaccination uses vaccinia, an orthopoxvirus related to variola. In the event of an outbreak, postexposure vaccination within 4 days can prevent or ameliorate disease.[3] The U.S. government has stockpiled enough doses for the entire U.S. population of a newer, cell-cultured product for use in a national emergency. The surveillance and containment or ring vaccination method combines active case finding with vaccination of potential contacts within a certain radius around the cases. Health care providers would need to be vaccinated, and airborne and contact precautions (HEPA filter masks, negative-pressure rooms, gowns, gloves, eye protection) would be recommended in the hospital environment. Although there was an attempt to encourage health care providers to be vaccinated in the aftermath of the 2001 anthrax attacks, this effort was discontinued because of concerns about vaccine side effects, such as myocarditis, and skepticism about the true risk of an outbreak. Because the vaccine is a live agent, it is contraindicated in individuals with eczema or significant exfoliative skin conditions or immune compromise in the pre-exposure setting. Even in normal hosts, it can be associated with inoculation of other locations on the body, spread to close contacts, disseminated disease, postvaccine encephalitis, progressive vaccinia in individuals with defective cell-mediated immunity, eczema vaccinatum in people with prior eczema, and pericarditis or myocarditis. In the postexposure setting,

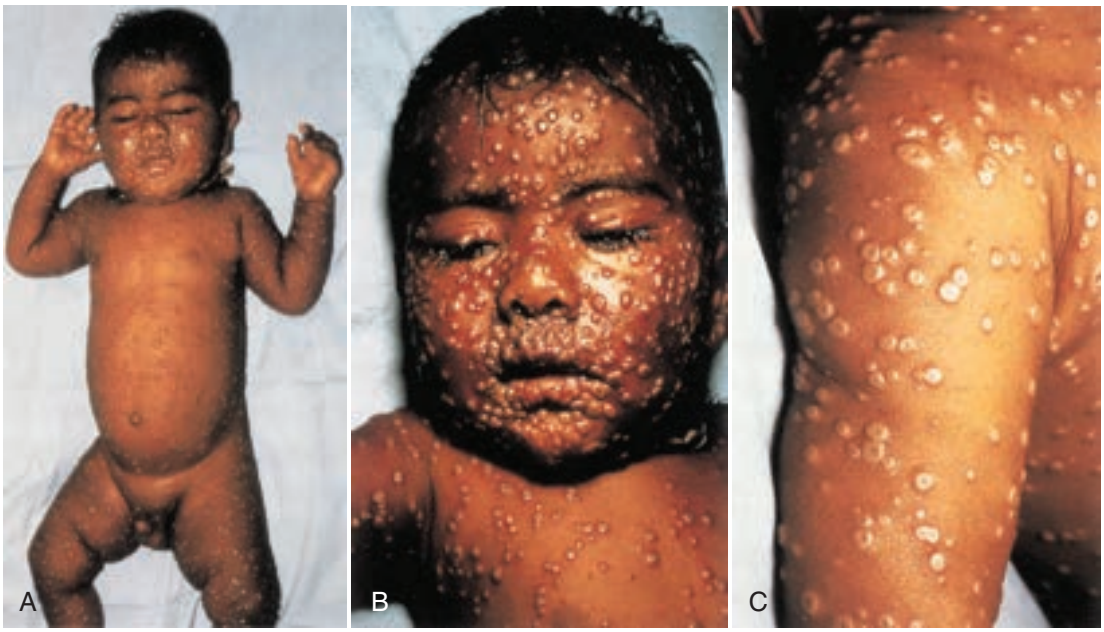

FIGURE 21-1. Smallpox. ***A,*** Demonstrates centrifugal nature of the rash. ***B,*** Demonstrates coalescences of some pustules. ***C,*** Demonstrates umbilicated pustules. (From Fenner F, Henderson DA, Arita A, et al. Smallpox and its eradication. Geneva: World Health Organization; 1988.)

risk benefits would have to be weighed regarding which individuals at risk for complications might receive vaccine and whether vaccinia immune globulin might be administered concomitantly. First-line therapy for significant vaccine reactions is vaccinia immune globulin.

PROGNOSIS

The case-fatality rate for the typical form of smallpox, known as variola major, is approximately 30%, but wide variations exist among races. In the minority of individuals who develop hemorrhagic smallpox or flat-type smallpox, the case-fatality rate approaches 100%. A different viral strain, variola minor, has case-fatality rates of only about 1%. Survivors of smallpox are often left with lifelong scarring from the lesions. Blindness and bone deformities, especially in children, are also known complications.

PLAGUE

EPIDEMIOLOGY

Plague was responsible for millions of deaths during three pandemics. Plague was used as a weapon by the Japanese before World War II when they released infected fleas in Chinese cities, and it was included in the former Soviet Union's biologic weapon arsenal. Plague is transmissible through the respiratory route, so an aerosol release could have devastating effects.

Plague is caused by the Gram-negative coccobacillus *Yersinia pestis* (Chapter 312). On light microscopy, the organism can have a bipolar appearance, making it look like a safety pin. In general, humans are infected during close proximity with rodents in areas where plague is enzootic or when their pets serve as vehicles for bringing infected fleas or the disease into their household. The "Black Death" of the 14th and 15th centuries owed much of its persistence to the ubiquitous presence of rats in homes and a lack of appreciation for their role in disease transmission.

Each year in the United States, several human cases are reported, usually in the Southwest. Most of these cases are bubonic, although occasional pneumonic plague is linked to infections in domestic cats.

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

Plague is manifested in three ways. Bubonic plague, which is the most common form, occurs after the bite of an infected flea. The organism spreads through the local lymphatics to the regional lymph nodes. As replication occurs, the lymph node or group of lymph nodes becomes swollen and extremely tender. As most flea bites occur on the lower extremities, the inguinal and femoral lymph nodes are most often affected. Untreated bubonic plague can progress to septicemic plague after organisms gain entry to the blood stream. In addition, patients can develop necrosis of cooler areas of the body, such as the tip of the nose, ears, or digits, as a result of a temperature-dependent coagulase produced by the organism. Septicemic plague may also occur in the absence of an antecedent bubo. Meningitis may occur from seeding through the blood stream. In pneumonic plague, the lungs are seeded secondarily by bacteremia from septicemic plague or primarily when a person inhales infected droplets. Although naturally occurring pneumonic plague is rare, this form of disease is the major concern after intentional aerosol release. In patients with pneumonic plague, viable organisms are found in the sputum, and the disease can then be spread to others. One of the hallmarks of the disease is the potential for purulent sputum to become hemorrhagic.

DIAGNOSIS

Plague bacilli are readily identified with Gram, Wright-Giemsa, or Wayson stains along with culture of the sputum or other infected body fluid, such as blood, fluid from a bubo, or cerebrospinal fluid. Other diagnostic methods include PCR for the F1 antigen, direct fluorescent antibody staining of body fluids, and serology by enzyme-linked immunosorbent assay or passive hemagglutination. However, serology is primarily useful in retrospect, because patients need to be treated empirically before a serologic response occurs.

TREATMENT

Rx

The historic drug of choice has been streptomycin (1 g IM twice daily for 7 to 10 days or for 3 days after the fever remits), but suitable alternatives include gentamicin (5 mg/kg IM or IV daily or a 2 mg/kg load followed by 1.7 mg IM or IV three times daily), doxycycline (200 mg IV load, then 100 mg IV every 12 hours), ciprofloxacin (400 mg IV every 12 hours), and levofloxacin (500 to 750 mg IV once daily) (Table 21-2). The drug of choice for plague meningitis is chloramphenicol (25 to 30 mg/kg IV load, followed by 50 to 60 mg/kg/day every 6 hours; on favorable clinical response, the dose can be reduced to 25 to 30 mg/kg/day every 6 hours).[4]

PREVENTION

Use of standard precautions is appropriate for bubonic plague. Droplet precautions should be applied for patients with pneumonic plague; this generally includes a private room where caregivers wear masks, gowns, gloves, and eye protection within 3 to 6 feet of the patients to minimize spread. These precautions should be continued until the patient demonstrates improvement and has been receiving effective antibiotics for 72 hours. Oral ciprofloxacin, levofloxacin, or doxycycline is recommended for postexposure prophylaxis of household contacts or individuals suspected of exposure in either the endemic or bioterrorism setting. Prophylaxis should continue for 7 days beyond the period of exposure.

No licensed vaccine for plague currently exists in the United States and a previously-licensed whole cell killed vaccine is no longer available. A new

vaccine that uses the F1 and V antigens, which has demonstrated protection in animals against aerosol challenge, is being developed by the Department of Defense and is currently undergoing phase II testing.

PROGNOSIS

Case-fatality rates are 50% or more for untreated bubonic plague and nearly 100% for untreated septicemic and pneumonic plague. These fatality rates can be significantly reduced with prompt recognition and appropriate therapy.

TULAREMIA

Tularemia (Chapter 311) is caused by infection with the Gram-negative, aerobic nonmotile coccobacillary organism *Francisella tularensis*.[5] A zoonotic disease of rabbits, ground squirrels, and other small mammals, tularemia can be acquired by humans by skin or mucous membrane contact with body fluids or tissues of infected animals or from being bitten by infected deerflies, mosquitoes, or ticks. Disease can rarely occur by inhalation of contaminated dusts or ingestion of contaminated food or water. Relatively uncommon in the United States, fewer than 150 cases of naturally occurring human tularemia are typically reported each year. However, the environmental stability of *F. tularensis* as well as its very low infectious dose (as few as 10 organisms) makes it a potential weaponization threat.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Clinical manifestations of tularemia depend on the route of exposure. Although six different forms (glandular, oculoglandular, ulceroglandular, pharyngeal, pneumonic, and typhoidal) have been described, tularemia is perhaps more simplistically compared with plague, with glandular forms analogous to bubonic plague as well as pneumonic and typhoidal forms that present a clinical picture similar to what is seen in pneumonic and septicemic plague. A preponderance of these latter forms would be expected after an intentional release by aerosol. After an incubation period of 2 to 10 days, symptoms begin with fever and proceed to include severe exhaustion, substernal chest pain, nonproductive cough, and weight loss.

Diagnosis, which requires isolation of the organism in blood, sputum, skin, or mucus membrane lesions, may be difficult because of unusual growth requirements and overgrowth of commensal organisms. Therefore, serology is the mainstay of diagnosis, often in retrospect. Because of the risk of spread to microbiology laboratory personnel, it is important to notify the laboratory if infection with this organism is suspected.

TREATMENT AND PROGNOSIS

Rx

Treatment of tularemia is similar to that of plague, with aminoglycosides for 10 to 14 days; streptomycin and gentamicin are considered the drugs of choice, although ciprofloxacin is a newer potential alternative (Table 21-2). Doxycycline/tetracycline and chloramphenicol are second-line choices because their use has been associated with relapses, and the recommended duration of therapy is 14 to 21 days.

Person-to-person spread is unusual, and standard precautions are adequate in caring for ill individuals. For known exposures, prophylaxis with tetracycline (500 mg orally four times a day for 2 weeks) is effective if it is begun within 24 hours of exposure; consensus recommendations in a bioterrorism setting are doxycycline 100 mg or ciprofloxacin 500 mg orally twice a day as postexposure prophylaxis.[6] A live attenuated vaccine has proven efficacy in preventing disease due to laboratory exposures and aerosol challenge in human volunteers, but it is not readily available to the public.

The case-fatality rate for untreated pneumonic and typhoidal tularemia is 35%, but this rate can be reduced to less than 5% with appropriate treatment.

BOTULISM

Botulism is caused by exposure to one of eight related neurotoxins (A to H) produced by certain strains of *Clostridium botulinum* (Chapter 296), a ubiquitous anaerobic spore-forming Gram-positive bacillus.[7] Botulinum toxins, which are among the most toxic substances known, can be lethal in doses as low as 0.001 µg/kg. These toxins, which act at presynaptic nerve terminals, block the release of acetylcholine, thereby resulting in a generalized flaccid paralysis with autonomic dysfunction. Although most toxins are unstable in the environment and thus constitute dubious weaponization threats, the potent nature of botulinum toxin, coupled with the ease with which it might be used to contaminate food and water, makes it an agent of concern. Naturally occurring human botulism is limited to types A, B, and E, although other toxin serotypes can produce an identical clinical syndrome. For this reason, a heptavalent antitoxin with activity against all seven serotypes has been licensed recently.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Botulism typically develops after a latent period ranging from hours to several days. Initial manifestations involve bulbar palsies, ptosis, photophobia, blurred vision, and other signs of cranial nerve dysfunction. Symptoms progress craniocaudally, leading to dysphonia, dysphagia, and, ultimately, a descending symmetrical paralysis. In fatal cases, death typically results from respiratory muscle failure.

Although botulism may mimic other neurologic disorders, such as myasthenia gravis and the Guillain-Barré syndrome, the occurrence of an outbreak involving multiple cases of descending, symmetrical, flaccid paralysis should make diagnosis of botulism relatively straightforward.

TREATMENT AND PROGNOSIS

Rx

Supportive care measures plus antitoxin, including meticulous attention to ventilatory support, are the mainstay of botulism management. Patients may require such support for several months, so the management of a large-scale botulism outbreak would be especially problematic in terms of medical resources. Botulism is not contagious, so standard precautions are adequate in managing botulism victims. The first victims of an outbreak may have higher case-fatality rates because of delays in recognition. Those who recover can have long-term sequelae, such as dyspnea on exertion, fatigue, and weakness.

VIRAL HEMORRHAGIC FEVERS

EPIDEMIOLOGY

Viral hemorrhagic fever (Chapter 381) is a term used to describe the clinical syndrome caused by four families of RNA viruses: filoviruses (Ebola and Marburg), arenaviruses ("Old World" Lassa fever and "New World" South American viruses, Machupo, Junin, and others), bunyaviruses (hantaviruses, Crimean Congo hemorrhagic fever, and Rift Valley fever), and flaviviruses (yellow fever and dengue) (Table 21-3).[8] Agents of greatest concern from a bioterrorism perspective include the filoviruses and arenaviruses, which are infectious by the aerosol route and can replicate well in cell culture for large-scale production and whose clinical syndrome of hemorrhage and fever engenders fear. The massive outbreak of Ebola virus disease in West Africa in 2014 demonstrates the devastation that a hemorrhagic fever virus can have on a population, even in the absence of bioterrorism, especially when the medical infrastructure is sub-optimal.[9]

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

Once virus gains entry through the skin, mucous membranes, or respiratory tract, replication occurs in macrophages and dendritic cells. Organisms are then transported to regional lymph nodes and eventually through the lymphatics and blood stream to target organs, such as the liver and spleen, where local necrosis occurs. The body responds with activation of cytokines and chemokines, and massive depletion of lymphocytes disables the host's adaptive immune response. A sepsis-like picture ensues, with declining mean arterial pressure, increased vascular permeability, and a bleeding diathesis.

Clinical features of the hemorrhagic fevers include the acute onset of malaise, sore throat, fever, skin flushing, conjunctival injection, prostration, myalgias, and nonbloody diarrhea. These viruses have significant differences in the prominence of individual clinical findings. For example, dengue, the filoviruses, and Lassa more commonly cause a maculopapular rash in the first week. The filoviruses and South American viruses more typically cause obtundation and encephalitis. Although all of these produce modest elevations of aminotransferase levels, yellow fever and Rift Valley fever are known for causing significant hepatic dysfunction, thereby resulting in icterus and jaundice. Depending on a number of factors, including virulence, route of exposure, inoculum, and other host factors, illness can progress in the second week to overt signs of bleeding, including petechiae, purpura, ecchymoses, and oozing from venipuncture sites. If massive bleeding occurs, the source is usually the gastrointestinal tract. Severe cases will demonstrate a combination of neurologic and hematologic abnormalities.

TABLE 21-3 COMPARISON OF VIRAL HEMORRHAGIC FEVER AGENTS[4]

FAMILY	VIRUS	ENDEMIC AREA	FATALITY RATE	NOSOCOMIAL TRANSMISSION	COUNTERMEASURES
Filoviruses	Ebola	Africa, Philippines (Reston)	50-90% (Sudan/Zaire)	Yes	Anecdotal success with immune plasma; monoclonal antibodies demonstrate benefit after disease onset
	Marburg	Africa	23-70%	Yes	
Arenaviruses	Lassa	West Africa	1-2%	Yes	Ribavirin effective in a clinical trial with nonrandomized controls
	Junin	Argentine Pampas	30%	Rare	Immune plasma; reports of benefit with ribavirin; Candid 1 vaccine protective; not available in the United States
	Machupo	Bolivia	25-35%	Rare	Immune plasma; reports of benefit with ribavirin
Bunyaviruses	Crimean-Congo hemorrhagic fever	Africa, southeast Europe, Central Asia, India	30%	Yes	Anecdotal success with ribavirin treatment
	Rift Valley fever	Africa	<0.5%	No	Livestock vaccines in Africa; U.S. Department of Defense has two experimental vaccines
	Hantaviruses	Europe, Asia, South America (rare)	5% (Asian hemorrhagic fever with renal syndrome)	No*	Vaccines available in Asia; ribavirin effective in randomized trial but not licensed for that purpose
Flaviviruses	Yellow fever	Africa, South America	3-12%,(20-50% if second phase develops)	No	17D live attenuated vaccine
	Kyasanur Forest	Southern India	3-5%	No	Formalin-inactivated vaccine in India
	Omsk	Siberia	0.2-3%	No	Tick-borne encephalitis vaccines (not in United States) may offer some cross protection

*Exception is Andes, which causes hantavirus pulmonary syndrome (not addressed in this chapter).

DIAGNOSIS

Numerous more common diseases should be considered in the differential diagnosis of a viral hemorrhagic fever, including rickettsial infections (Rocky Mountain spotted fever, ehrlichia, anaplasma, and African tick typhus; Chapter 327), leptospirosis (Chapter 323), meningococcemia (Chapter 298), typhoid (Chapter 308), and falciparum malaria (Chapter 345). However, a viral hemorrhagic fever should be considered in patients who present with a clinically compatible syndrome, especially if a cluster of cases is seen. Laboratory features may include thrombocytopenia, elevated aminotransferase levels, leukopenia, anemia (although an elevated hematocrit may be seen in patients who have significant vascular leakage, as is sometimes seen with dengue and hantaviruses), hematuria, and proteinuria. PCR, viral culture, immunoglobulin M–specific enzyme-linked immunosorbent assay, acute and convalescent serology, or immunohistochemistry on autopsy specimens can establish the diagnosis. Routine clinical specimens and any attempt at viral isolation require a biosafety level 3 or 4 laboratory with expert consultation from the Centers for Disease Control and Prevention or the U.S. Army Medical Research Institute of Infectious Diseases.

TREATMENT

Rx

No licensed therapies exist for the viral hemorrhagic fevers, so supportive care is the primary means of management with close attention to fluid status, avoidance of procedures that cause bleeding and medications that impair platelet function, use of blood products as necessary, use of vasopressors for hypotension, and dialysis for renal failure. Close attention to fluid status has significantly reduced mortality from dengue, and aggressive fluid repletion has been needed to address significant loss of fluid volume due to diarrhea, electrolyte abnormalities, and protein/nutritional depletion in Ebola infections.[10] Oral ribavirin or intravenous ribavirin (see Table 21-2) under an approved experimental protocol can reduce mortality in Lassa hemorrhagic fever, and it also appears to be beneficial against Argentine hemorrhagic fever (Junin virus) and hantaviruses that cause hemorrhagic fever with renal syndrome. It has no apparent efficacy against filoviruses (Ebola or Marburg) or flaviviruses (dengue, yellow fever). Monoclonal antibodies are a potentially promising option for Ebola virus.[11]

PREVENTION

Yellow fever vaccine is the only licensed vaccine against a viral hemorrhagic fever in the United States. Experimental vaccines exist against Junin virus and Rift Valley fever virus, but these vaccines have been used primarily to protect laboratory workers.

Certain viral hemorrhagic fevers (primarily Ebola, Marburg, Lassa, and Crimean-Congo hemorrhagic fever) produce high-level viremia during the period of greatest bleeding risk, thereby making them notorious for causing nosocomial outbreaks. In general, these outbreaks occur in environments, such as sub-Saharan Africa, where basic infection control practices are inadequate because of limited resources (e.g., lack of gowns, gloves, and eye protection as well as reuse of unsterilized needles and syringes). Spread of these viruses in the nosocomial environment can be significantly reduced with standard precautions that limit contact with blood and body fluids. In caring for patients at higher risk of spread (those who are coughing, vomiting, or hemorrhaging), enhanced precautions are recommended, using negative-pressure isolation and N-100 masks along with gowns, gloves, and eye protection. Even with these precautions, the infection of two nurses in Dallas in 2014 from a severely ill patient demonstrates the challenge of doing this right, as there is little room for error.

PROGNOSIS

Hemorrhagic viruses exhibit a wide range of mortality rates, ranging from less than 1% with Rift Valley fever to 80 to 90% with Ebola Zaire (Table 21-3). Despite the notoriety of these viruses for causing hemorrhage, the majority of individuals do not die of blood loss. Rather, death results from a sepsis-like picture, including a loss of vascular hemostasis, disseminated intravascular coagulation, hypotension, renal failure, shock, and death.

RESPONSE TO A BIOTERRORISM ATTACK

Even a relatively small bioterror event can have profound consequences. After the 2001 anthrax mailings, thousands of people were prescribed antibiotic prophylaxis, and government buildings were closed for decontamination. Since that event, significant improvements in future responses to a bioterrorism event include the licensure, manufacture, and storage of enough smallpox vaccine for everyone in the United States, production of anthrax and botulinum antitoxins for the strategic national stockpile, and development of protocols and authority for response. However, no country has yet achieved the ability to detect a pathogen release reliably in real time.[12] The 2014 Ebola outbreak also has identified shortcomings in preparedness in the U.S.

In any future bioterror attack, early warning signs might include clusters of patients with mediastinal adenopathy or widened mediastinum on the chest radiograph, a centrifugal pustular rash (smallpox), pneumonia and hemoptysis (plague), descending flaccid paralysis (botulism), or hemorrhagic manifestations and impaired clotting (viral hemorrhagic fever). Should such a cluster be identified or suspected, physicians should contact the local county or city health department, the state health department, and potentially the Centers for Disease Control and Prevention. Unlike a natural

outbreak, a potential bioterror outbreak also requires a rapid dialogue between public health and law enforcement authorities, in part because patient and environmental samples constitute evidence against a perpetrator and must be handled under chain of custody. Physicians also must be prepared not only to protect themselves and their patients but also to serve as community resources, regardless of whether an outbreak is natural or terrorist in origin.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Christian M. Biowarfare and bioterrorism. *Crit Care Clin*. 2013;29:717-756.
2. Stern EJ, Uhde KB, Shadomy SV, et al. Conference report on public health and clinical guidelines for anthrax [conference summary]. *Emerg Infect Dis*. 2008;14:e1. http://wwwnc.cdc.gov/eid/article/14/4/07-0969.htm; Accessed January 29, 2015.
3. Sato H. Countermeasures and vaccination against terrorism using smallpox: pre-event and post-event smallpox vaccination and its contraindications. *Environ Health Prev Med*. 2011;16:281-289.
4. Raoult D, Mouffok N, Bitam I, et al. Plague: history and contemporary analysis. *J Infect*. 2013;66:18-26.
5. Egan JR, Hall IM, Leach S. Modeling inhalational tularemia: deliberate release and public health response. *Biosecur Bioterror*. 2011;9:331-343.
6. Kman NE, Nelson RN. Infectious agents of bioterrorism: a review for emergency physicians. *Emerg Med Clin North Am*. 2008;26:517-547, x-xi.
7. Leclair D, Fung J, Isaac-Renton JL, et al. Foodborne botulism in Canada, 1985-2005. *Emerg Infect Dis*. 2013;19:961-968.
8. Ftika L, Maltezou HC. Viral haemorrhagic fevers in healthcare settings. *J Hosp Infect*. 2013;83:185-192.
9. Schieffelin JS, Shaffer JG, Goba A, et al. Clinical illness and outcomes in patients with Ebola in Sierra Leone. *N Engl J Med*. 2014;371:2092-2100.
10. Lyon GM, Mehta AK, Varkey JB, et al. Clinical care of two patients with Ebola virus disease in the United States. *N Engl J Med*. 2014;371:2402-2409.
11. Qiu X, Wong G, Audet J, et al. Reversion of advanced Ebola virus disease in nonhuman primates with ZMapp. *Nature*. 2014;514:47-53.
12. Russell PK, Gronvall GK. U.S. medical countermeasure development since 2001: a long way yet to go. *Biosecur Bioterror*. 2012;10:66-76.

REVIEW QUESTIONS

1. An air detector in your city has indicated that there has been an aerosol attack with smallpox. This has been verified by the state health department. You are seeing a number of individuals suspected of having been exposed during the attack that occurred 3 days ago. What would be the appropriate management for these individuals?

A. Give vaccinia immune globulin at the earliest opportunity.
B. Vaccinate exposed individuals with the smallpox vaccine (ACAM2000).
C. Vaccinate exposed individuals with the smallpox vaccine, except those with eczema.
D. There is nothing you can offer at this point except reassurance.
E. You can treat them with cidofovir.

Answer: B Studies of historical outbreaks have demonstrated that after individuals are exposed to smallpox, it has been possible to prevent or to ameliorate smallpox disease by vaccination. When it is administered by scarification, replication of the smallpox vaccine (vaccinia virus) can provide immunity faster than the replication of the variola (smallpox) virus. However, the vaccination must occur relatively quickly after exposure, usually within 4 or 5 days, to have maximum likelihood of benefit. Although the vaccine can cause adverse effects in immunocompromised people and individuals with skin disorders, such as eczema, there are no contraindications to vaccination in the event of a true exposure. Cidofovir is potentially nephrotoxic and has been used for treatment but not for prevention. Vaccinia immune globulin is primarily used for treatment of vaccine adverse events, not for smallpox disease. Finally, a newer vaccine, modified vaccinia Ankara, which is believed to have less risk for immunocompromised individuals, is under study as a potential alternative.

2. You become aware of a large outbreak of nontyphoidal *Salmonella* in your community. Which of the following is not true?

A. More cases than would be expected in the community might be a clue of intentional spread.
B. A genetic match between patient isolates, a laboratory at a suspicious local cult, and local salad bars may indicate intentional spread.
C. This outbreak is unlikely to be spread intentionally because *Salmonella* is not considered a high-threat agent.
D. Disease in a specific ethnic group may indicate intentional targeting of the infectious agent.
E. An unusual antibiotic susceptibility pattern of the organism might indicate intentional tampering.

Answer: C Public health authorities have designated certain organisms as most concerning (category A agents) on the basis of characteristics such as the ability to grow in large quantities, stability in aerosol, and ability to infect by the aerosol route. However, we are unable to understand a specific perpetrator's criminal intentions; the perpetrators may not have access to such classic bioweapon agents, and they may instead use an organism that is easier to obtain. This is what occurred in the Dalles, Oregon, outbreak in 1984, when the Rajneesh cult contaminated local salad bars with *Salmonella typhimurium*. These types of outbreaks are difficult to pinpoint as artificially caused, so it is useful to look at historical outbreaks and potential clues as to why this might deviate from what would be expected.

22

CHRONIC POISONING: TRACE METALS AND OTHERS

LOUISE W. KAO AND DANIEL E. RUSYNIAK

About 80% of the elements in the periodic table are metals or metalloids, and various metals are in contact with humans in the home, workplace, and environment. Metals are part of a number of normal physiologic processes, such as iron in hemoglobin, but also can cause a number of toxic adverse effects.[1]

LEAD TOXICITY

EPIDEMIOLOGY

Lead (Pb) is a silver-gray, malleable metal that is resistant to corrosion. It has no known physiologic use, so any lead in the human body should be considered contamination.

The most common sources of lead poisoning are lead-based paints, lead-contaminated dust in older buildings, and lead-contaminated soil. Other sources include plumbing, solder, batteries, bullets, toys, curtain weights, necklace charms, food containers, and cosmetics, including Nigerian eyeliners. Adults in occupations such as battery manufacturing, welding, construction, mining, glass blowing, and shipbuilding have the highest lead exposures. Other less common but important routes of exposure include lead-contaminated moonshine and lead-containing ethnic folk remedies (e.g., greta and azarcon in Mexico).

Children in lower socioeconomic areas have the highest exposures to lead, and the majority of the scientific literature is biased toward pediatric studies. Much of what we have learned, however, is applicable to adults.

PATHOBIOLOGY

Lead's most devastating effects are on the central nervous system (CNS) of children. By disrupting the intracellular junction of capillary endothelium, lead impairs the blood-brain barrier, which further increases CNS absorption. Lead also increases capillary leak, which in severe cases can result in cerebral edema.

Lead inhibits several enzymes involved in heme synthesis: aminolevulinic acid synthetase, δ-aminolevulinic acid dehydratase, coproporphyrinogen decarboxylase, and ferrochelatase. Lead also inhibits the enzyme erythrocyte pyrimidine-5-nucleotidase, which impairs RNA degradation and contributes to red cell hemolysis. Collectively, the inhibition of these enzymes results in decreased concentrations of hemoglobin and a shorter life span of red blood cells.

By depositing lead-protein complexes in renal proximal tubular cells, lead interferes with normal mitochondrial function, thereby resulting in decreased reabsorption of glucose, amino acid, and phosphate. Chronically exposed persons can develop renal failure from tubular atrophy, interstitial fibrosis, and glomerular sclerosis.

Although lead has not been associated with specific bone-related disorders, the skeletal system serves as the main reservoir for lead; with chronic exposure, lead stores in bone can have a half-life of 5 to 19 years. As a result, soft tissues may be subjected to increased lead exposure during times of accelerated bone turnover, such as during childhood growth, after a long bone fracture, or during pregnancy.

CLINICAL MANIFESTATIONS

Lead poisoning can be subtle and difficult to diagnose because many of its historical clinical associations (e.g., neuropathy, gout, abdominal colic) are rarely seen today. In children, the most concerning long-term complications of environmental lead exposure are developmental cognitive deficits, especially when lead concentrations in young children exceed 100 µg/dL. Subacute neurologic problems include ataxia, lethargy, seizures, and coma. Adults can also experience severe neurologic symptoms (e.g., seizures and cerebral edema) with plumbism, but typically only when whole blood lead levels are higher than 150 µg/dL. More commonly, the neurologic problems caused by lead in adults are manifested as memory problems, insomnia, depression, and personality changes.

Other clinical effects of lead poisoning, in both children and adults, include a normocytic or microcytic anemia, abdominal pain, hepatotoxicity, and pancreatitis. The classic finding of peripheral neuropathy with footdrop and wristdrop is well described in adults but only occasionally seen in children. Nephrotoxicity most commonly is manifested as Fanconi syndrome (Chapter 128) with aminoaciduria, glycosuria, and phosphaturia. A consequence of this toxicity is impaired uric acid clearance and gout.[2] High lead levels also probably increase the risk of hypertension.

DIAGNOSIS

Diagnosis of lead intoxication involves a high level of suspicion. The best initial test is a whole blood lead level collected in a certified lead-free tube (Table 22-1). The blood lead level can be falsely elevated in individuals who have received chelation therapy, which mobilizes tissue lead stores, in the prior 7 days.

Radiographic imaging can support the diagnosis of lead poisoning and sometimes detect a lead-containing object or a retained bullet. Bone x-ray fluorescence technology can estimate bone lead levels, which can reflect chronic lead exposure, but this test is mostly a research tool and is not commonly available in clinical practice.

TREATMENT

Rx

The first step in management of patients with elevated lead levels is prompt removal of the source, although the effectiveness of environmental cleanup and educational initiatives is disappointing.[3] If lead comes from a work-related source, the Occupational Safety and Health Administration should be contacted. Chelation increases excretion of lead in a 24-hour urine sample and thereby decreases blood concentrations, but studies to date with the chelating agent succimer (2,3-dimercaptosuccinic acid, DMSA) in children have not shown improved neuropsychological function[3,4] despite success in reducing blood lead levels. The current recommendation is that asymptomatic adults with blood lead levels of less than 70 µg/dL do not require chelation, patients with mild symptoms or blood lead levels between 70 and 100 µg/dL should receive only oral chelation, and patients with lead-induced encephalopathy or blood lead levels above 100 µg/dL should be considered for parenteral chelation. Although DMSA is approved by the Food and Drug Administration only for children, it also is commonly used in adults (Table 22-2).

Whole bowel irrigation with a polyethylene glycol–electrolyte solution should be considered if a patient has retained a gastrointestinal lead object. Patients who have lead-containing bullet fragments lodged in soft tissues or joint spaces may need to have them surgically removed.

PROGNOSIS

Despite controversy about long-term recovery from lead-associated cognitive deficits in children, adults with neurologic symptoms improve with removal of the exposure. The hematologic and nephrotoxic effects of lead also are reversible.

MERCURY TOXICITY

EPIDEMIOLOGY

Mercury (Hg) is a naturally occurring metal with three distinct forms. Although all forms are neurotoxic, elemental mercury (Hg^0) also causes pulmonary toxicity, inorganic mercury (Hg^+ or Hg^{++}) causes gastrointestinal and

renal toxicity, and organic mercury (ethyl, methyl, alkyl, or phenyl) is a teratogen.

Elemental mercury is found in dental amalgams, thermometers, fluorescent lamps, batteries, and some paints. Occupational exposures may occur in dentists and dental hygienists, painters, gold extractors, bronzers, electroplaters, metallurgists, paper pulp manufacturers, miners, ceramic workers, and other workers who use mercury in processing. Mercury thermometers are becoming increasingly rare, and many countries have banned them because of potential mercury toxicity. Toxicity from elemental mercury is caused primarily from breathing its volatilized vapor.

Inorganic mercury has been used historically as a medicinal, cosmetic, and topical antiseptic in the forms of HgCl (calomel) and $HgCl_2$ (mercuric chloride). It is also used in the tanning industry, in taxidermy, in manufacturing of fireworks, and in dye manufacturing. Inorganic mercury may also be found in patent medicines, folk remedies (such as empacho among Mexican Americans), skin lightening creams, and Asian herbal remedies. Toxicity may occur from ingestion, inhalation, or dermal absorption.

Organic mercury compounds can be classified as short-chain alkyl (such as methylmercury), long-chain alkyl, and aryl compounds. The aryl compounds, such as thimerosal, behave like inorganic mercurial compounds. Methyl mercury has caused several large-scale human poisoning epidemics.

Soil and marine organisms methylate inorganic mercury from the air and industrial waste. Through a process known as bioamplification, methyl mercury concentrates in the tissues of marine life, reaching highest concentrations in large predatory fish such as tuna and swordfish. Fish consumption is now the largest route of organic mercury exposure in the general population. Organic mercury is also used as a fungicide, pesticide, wood preservative, and medicinal antiseptic or preservative (Mercurochrome and thimerosal). Toxicity is typically by ingestion.

TABLE 22-1 DIAGNOSTIC TESTING FOR METALS: REFERENCE RANGES

METAL	SERUM LEVEL	WHOLE BLOOD LEVEL	SPOT URINE	24-HOUR URINE
Lead		<10 µg/dL		
Mercury		<10 µg/L	<20 µg/L	<5 µg/g creatinine
Arsenic		<5 µg/L		<50 µg/L or <100 µg/g creatinine
Cadmium		<5 µg/L		<3 µg/g creatinine
Aluminum	<2 µg/L	<12 µg/L		4-12 µg/g creatinine
Bismuth	<1 µg/dL	<5 µg/dL	<20 µg/L	
Cobalt	0.1-1.2 µg/L		0.1-2.2 µg/L	
Manganese	0.9-2.9 µg/L	4-15 µg/L		<10 µg/L
Silver	<1 µg/L			<2 µg/L
Thallium		<2 µg/L		<5 µg/L
Zinc	109-130 µg/dL	600-1000 µg/dL		<500 µg/day

Modified from Nelson L, Goldfrank LR. Goldfrank's Toxicologic Emergencies. 9th ed. New York: McGraw-Hill Medical; 2011.

PATHOBIOLOGY

The pathophysiology of mercury poisoning is related to its disruption of cell physiology by binding to sulfhydryl, phosphoryl, carboxyl, and amide groups, thereby causing widespread dysfunction of normal cellular mechanisms. Each type of mercury causes differential effects depending on its route of exposure, solubility, and lipophilicity.

Elemental mercury is poorly absorbed by ingestion. If gastrointestinal motility or mucosal integrity is compromised, however, elemental mercury

TABLE 22-2 CHELATORS FOR ADULT HEAVY METAL POISONING

SYMPTOMS	CHELATOR	RECOMMENDED DOSE†	ADVERSE EFFECTS
LEAD			
Asymptomatic, BLL* < 70	None		
Mild-moderate, BLL 70-00	DMSA	10 mg/kg tid PO × 5 days, then bid for 14 days	Nausea, vomiting, diarrhea; mild elevations in aminotransferase levels
Encephalopathy, BLL > 100	BAL +	4 mg/kg deep IM q 4 h × 5days	Local injection site (pain, redness, sterile abscess); nausea, vomiting, diarrhea; anxiety, hypertension, tachycardia, fever; contraindicated in peanut allergy
	$CaNa_2EDTA$‡	1500 mg/m^2/day (approximately 50-75 mg/kg/day) either continuous infusion or in 2-4 divided IV doses for 5 days	Renal toxicity (from metal chelate), constitutional symptoms, transient hypotension
MERCURY			
Acute (elemental or inorganic) with moderate/severe symptoms	BAL	5 mg/kg deep IM then 2.5 mg/kg every 12-24 hours for 10 days or until symptoms improve and patient is able to take PO	Local injection site (pain, redness, sterile abscess); nausea, vomiting, diarrhea; anxiety, hypertension, tachycardia, fever; contraindicated in peanut allergy
Chronic (elemental or inorganic) or organic with symptoms	DMSA	10 mg/kg tid PO × 5 days, then bid for 14 days	Nausea, vomiting, diarrhea, mild elevations in aminotransferase levels
ARSENIC			
Acute exposure with moderate to severe symptoms	BAL	3 mg/kg deep IM q 4 h × 2 days, then bid for 7 to 10 days	Local injection site (pain, redness, sterile abscess); nausea, vomiting, diarrhea; anxiety, hypertension, tachycardia, fever; contraindicated in peanut allergy
Chronic exposure with moderate symptoms§	DMSA	10 mg/kg tid PO × 5 days, then bid for 14 days	Nausea, vomiting, diarrhea, mild elevations in aminotransferase levels
ALUMINUM			
Acute or chronic exposure in dialysis patient with encephalopathy	Deferoxamine	See *http://www2.kidney.org/professionals/kdoqi/guidelines_bone/Guide12.htm*	Hypotension, increased risk of sepsis, acute lung injury
THALLIUM			
Acute thallium poisoning with moderate to severe symptoms	Prussian Blue	3 g tid PO until the urinary thallium excretion is less than 0.5 mg/day	Constipation, blue stool

*Adult recommendations.
†Chelation doses in general are not well studied or validated, with the exception of lead in children. Optimal dosing is not well established, particularly for BAL. Doses listed are author's suggestions.
‡Start 4 hours after first dose of BAL is administered.
§Benefit in this setting is not established.
BAL = British anti-Lewisite, dimercaprol; BLL = blood lead level; DMSA = 2,3-dimercaptosuccinic acid; succimer; EDTA = ethylenediaminetetraacetic acid.

may ionize into more readily absorbed forms. Elemental mercury toxicity occurs primarily by inhalation of the vapor into the alveoli, where it is well absorbed in the pulmonary circulation and readily crosses the blood-brain barrier. In the brain, it interferes with multiple cellular processes, including protein and nucleic acid synthesis. Mercury also inhibits catechol-*O*-methyltransferase, thereby elevating circulating catecholamine levels.

Inorganic mercury is absorbed after ingestion, initially binds the gastrointestinal mucosa, and accumulates in the kidneys, where it exerts direct oxidative damage. Although it does not readily cross the blood-brain barrier, CNS toxicity can be seen with chronic exposures because of the prolonged elimination rate. Mercuric ions do not appear to cross the placental barrier.

Organic mercury compounds are readily absorbed after ingestion and inhalation and are moderately absorbed after dermal exposure, particularly if skin is not intact. Owing to its high lipid solubility, organic mercury crosses the blood-brain barrier and concentrates in the CNS. It can also cross the placental barrier and concentrate in the fetus. Organic mercury concentrates in red blood cells, distributes throughout the body, and is primarily eliminated in the feces. Some forms of organic mercury are metabolized in the body to inorganic mercury compounds. Toxicity results from its inhibition of enzyme systems and interference with cellular maturity, microtubule function, and neurotransmitter synthesis and uptake.

CLINICAL MANIFESTATIONS

The manifestations of elemental mercury toxicity vary depending on the dose and chronicity of the exposure.[5] Inhalation of high concentrations of mercury vapor, as may occur in an industrial setting, may result in cough, chills, fever, and shortness of breath. Nausea, vomiting, and weakness may result. This syndrome may progress to a severe acute lung injury and respiratory as well as renal failure. Chronic exposure to lower concentrations of mercury vapor produces a classic triad of tremor, gingivostomatitis, and neuropsychiatric disturbances. The mercurial tremor may be both static and intentional. Sudden episodic bursts of tremor, also called tetanus mercurialis, have been described. Neuropsychiatric manifestations of mercury poisoning, also known as erethism, include fatigue, insomnia, memory dysfunction, social withdrawal, shyness, and depression.

Inorganic mercury poisoning can occur after dermal or mucosal absorption of mercury-containing cosmetics and teething powders as well as after accidental or intentional ingestion of mercuric chloride antiseptics. After acute ingestion, a hemorrhagic gastroenteritis is typically followed by renal failure from acute tubular necrosis. Chronic exposure is associated with erethism, renal dysfunction, and neurologic manifestations such as sensorimotor neuropathy, constriction of visual fields, tremor, and delirium. Acrodynia, or pink disease, has been described after elemental or inorganic mercury exposure, most notably in children after exposure to mercurial teething powder and diaper ointment. This syndrome is manifested as an erythematous, hyperkeratotic, often desquamating rash on the palms, soles, and face in conjunction with a papular rash. Acrodynia is also associated with an idiosyncratic hypersensitivity to mercury ions and mercury poisoning, which itself can increase circulating catecholamines and mimic pheochromocytoma (Chapter 228). Common findings include tremor, diaphoresis, tachycardia, and hypertension.

Human toxicity from organic mercury was first recognized when the dumping of mercury-containing waste into Minamata Bay in Japan led to a large-scale poisoning of the population, whose primary dietary staple was fish from the bay. In what is now called Minamata disease, patients presented with paresthesia, ataxia, dysarthria, tremor, and constriction of visual fields or "tunnel vision." These symptoms can be progressive and sometimes fatal. Children born to exposed mothers (congenital Minamata disease) suffer mental retardation, limb deformities, chorea, seizures, and microcephaly.

With organic mercury toxicity, symptoms are typically delayed for weeks to months. Other symptoms that follow organic mercury poisoning include mucous membrane irritation after ingestion and dermatitis after cutaneous exposure. However, no cardiovascular toxicities have been reported with up to moderate elevation in blood mercury levels.[6]

DIAGNOSIS

The diagnosis of mercury poisoning should be considered on recognition of the characteristic signs and symptoms coupled with a known or suspected exposure. Whole blood and urine mercury levels can confirm the exposure, but the correlation between levels and symptoms is inconsistent. In the setting of elemental mercury ingestion or injection, radiographs may also be useful. In a nonoccupationally exposed patient, whole blood mercury concentration should not exceed 10 µg/L, and a 24-hour urine mercury concentration should not exceed 20 µg/L (Table 22-1); urine must be collected in an acid-washed container. In general, urine concentrations are more reliable for monitoring of exposure and the response to therapy, but whole blood concentration is the preferred monitoring test for methyl mercury poisoning. The ratio of red blood cell to plasma mercury can differentiate organic mercury toxicity from inorganic mercury toxicity because organic mercury concentrates significantly in red blood cells. Hair analysis is unreliable because of the potential for external contamination. The practice of obtaining urine mercury levels after a chelation challenge is not recommended because levels obtained in this way are difficult to interpret.

TREATMENT

Rx

Treatment is primarily symptomatic and supportive, with removal from the source of exposure. In cases of acute exposure, decontamination may be required. In cases of occupational or environmental exposure, environmental decontamination and surveillance by local or federal agencies may be necessary. After inhalation of elemental mercury vapor, support of respiratory function is crucial. After ingestion of inorganic mercury salts, fluid resuscitation, usually with normal saline, is needed to correct intravascular depletion, and renal replacement therapy may be required in cases of oliguric renal failure. Mercury compounds are poorly cleared by extracorporeal elimination measures, such as hemodialysis or peritoneal dialysis.

If dietary fish is the likely cause of an elevated mercury level, affected individuals should avoid eating any fish or shellfish for 1 month, after which blood or urine mercury levels should be reanalyzed. If mercury levels have declined into the normal range, as is usually the case, low-mercury fish (shrimp, canned light tuna, salmon, pollock, catfish) can be reintroduced into the diet at a frequency of no more than two meals per week.

Chelation therapy will increase urinary elimination of mercury (Table 22-2), and limited clinical data support its use early after acute poisonings. Oral DMSA, which is the chelator of choice, is generally well tolerated. If oral administration is not possible, dimercaprol (British anti-Lewisite, BAL) can be used except after methyl mercury exposure, in which it is contraindicated because it may shift mercury into the brain.

PROGNOSIS

The neurotoxicity that follows significant poisoning from all types of mercury may be irreversible, particularly in the setting of organic mercury poisoning, in which the diagnosis is often delayed. Nevertheless, 33 of 40 symptomatic children affected in the contaminated grain outbreak of 1971 in Iraq improved during a 2-year observation period. Acute renal failure after elemental or inorganic mercury poisoning can sometimes resolve. Patients with acrodynia have been reported to recover completely after chelation therapy and removal of the exposure.

ARSENIC TOXICITY

EPIDEMIOLOGY

Arsenic (As), which is found in soil, minerals, rocks, and metal ores, is present in all living organisms. It exists in several forms: elemental, inorganic (As^{3+} trivalent arsenite and As^{5+} pentavalent arsenate), gaseous (arsine, AsH^3), and organic. Elemental and organic arsenic have low toxicity, whereas gaseous arsine and inorganic arsenic are highly toxic. Arsenic's medicinal properties were recognized as early as 400 BC, and it has been used throughout history as a medicinal as well as a component of pigments, cosmetics, and famously as a poison.

Human exposure to arsenic may occur through contaminated air, groundwater, soil, and food, particularly seafood, rice, and produce. Marine organisms, especially shellfish, contain organic arsenicals arsenobetaine and arsenocholine, which are commonly reported in laboratory assays as elevated arsenic level but exert no known toxic effects.

In Bangladesh, ongoing epidemic arsenic poisoning from contaminated groundwater has affected millions of persons. Taiwan, Chile, the Cordoba province in Argentina, West Bengal, and other regions in the Ganga plain also have elevated levels of naturally occurring arsenic and cases of arsenic poisoning.

Occupational exposure to arsenic occurs in the microelectronics industry, where arsenide crystals are used to etch circuits on microchips. Arsine gas is liberated when inorganic arsenic contacts acid, as may occur in occupations such as metal smelting, galvanizing, and semiconductor manufacturing. Arsenic compounds are also used in the production of paint, fungicide,

insecticide, pesticide, herbicide, wood preservatives, ceramics, and glass, and employees in these industries may also be potentially exposed.

Arsenic compounds may still be found in folk remedies and patent medicines. Modern medicinal uses of arsenic include arsenic trioxide (Trisenox) for the treatment of acute promyelocytic leukemia (Chapter 183) and melarsoprol, an organic arsenical, for the treatment of African trypanosomiasis (Chapter 346).

PATHOBIOLOGY

The toxicologically significant arsenic compounds are inorganic (trivalent and pentavalent). Arsine gas, which is also toxic, causes acute hemolysis. After absorption, inorganic arsenic binds to hemoglobin and is distributed to liver, kidney, heart, and lungs. In the liver, arsenic is methylated to form monomethylarsenoic and dimethylarsinic acid, both of which are less toxic. Arsenic concentrates in keratin-rich tissues, such as hair, skin, and nails. Much of an ingested dose of arsenic is eliminated in the urine. The mechanism of toxicity is by binding sulfhydryl groups of critical enzymes, including those of the Krebs cycle, thereby resulting in impaired gluconeogenesis, impaired oxidative phosphorylation, and ultimately depletion of cellular energy stores. Pentavalent arsenate may substitute for phosphate in biochemical reactions and disrupt normal oxidative phosphorylation. Arsenic also affects cardiac conduction by blocking cardiac potassium channels. Arsenic can alter gene expression through induction, downregulation, and upregulation of various genes involved in apoptosis, cell signaling, and growth factor response.

Arsine is a colorless, nonirritating gas. After inhalation, it is absorbed rapidly and binds to erythrocytes, where it exerts oxidative stress and causes severe Coombs-negative intravascular hemolysis. Renal failure is due to hemoglobin pigment deposition as well as to the direct toxic effects of arsine on renal tubular cells.

CLINICAL MANIFESTATIONS

The initial clinical features after ingestion of inorganic arsenic are nausea, vomiting, bloody diarrhea, and abdominal pain. Within several days, hematologic findings such as pancytopenia can be seen. QT prolongation, which can develop acutely or chronically, can lead to dysrhythmias such as torsades de pointes. After gastrointestinal symptoms improve, distal symmetrical peripheral neuropathy develops, potentially accompanied by weakness or encephalopathy.

Chronic exposures affect the bone marrow, skin, and peripheral nervous system.[7] Dermatologic effects include patchy or diffuse alopecia, hyperpigmentation, and melanosis as well as hyperkeratosis on the palms and soles. The pigmentation of chronic poisoning commonly appears in a finely freckled, "raindrop" pattern of symmetrical pigmentation or depigmentation that is particularly pronounced on the trunk and extremities. Nails may exhibit transverse white bands, which are known as Mees lines (Chapter 442) and reflect growth interruption during poisoning. Anemia, pancytopenia, neutropenia, thrombocytopenia, and eosinophilia can be seen. Neuropathy, which is a hallmark of arsenic poisoning, is described as a diffuse, symmetrical, ascending, painful sensorimotor neuropathy, most prominent in a stocking-glove distribution. In severe poisoning, ascending weakness and paralysis may result in respiratory failure that mimics the Guillain-Barré syndrome (Chapter 420). Arsenic exposure also causes a dose-dependent decline in lung function. Peripheral vascular disease, including peripheral vascular gangrene (black foot disease), can develop in chronically exposed patients. Even low to moderate chronic arsenic exposure increases the long-term risk of cardiovascular disease by about 30%.[8] Arsenic is a human carcinogen, and exposed populations have an increased risk of developing malignant neoplasms in the lung, skin, and bladder.

Arsine gas produces a clinical triad of abdominal pain, hemolysis, and hematuria, typically occurring hours after exposure. Patients may initially have headache, weakness, nausea, and vomiting. Several weeks after an acute exposure, peripheral neuropathy may develop.

DIAGNOSIS

Because arsenic clears from the blood quickly, an arsenic level above 100 µg/24 hours in urine collected in an acid-washed container requires further scrutiny (Table 22-1). Arsenic levels above 50 µg/L in a spot urine test warrant a 24-hour test. Patients who have recently ingested seafood may have urine arsenic levels exceeding 1500 µg/L, exclusively from organic arsenic, so differentiating inorganic from nontoxic organic arsenic is often critical. Total arsenic levels in hair or nails, where arsenic accumulates, are useful indicators of past exposures. Blood arsenic, urine arsenic, and urine arsenic metabolites can be used to confirm recent or ongoing exposure. In the unexposed individual, blood arsenic should be below 1 µg/L; hair and nail levels should be less than 1 ppm.

In patients with chronic arsenic exposure, a complete blood count may show anemia (normocytic, normochromic, or megaloblastic), leukopenia, and thrombocytopenia. A peripheral smear may show basophilic stippling (Fig. 157-14) or karyorrhexis. Renal dysfunction, hepatic enzyme elevation, and hyperbilirubinemia may be seen as well. The electrocardiogram may show QT prolongation and nonspecific ST-T wave changes. Nerve conduction studies typically show evidence of a distal symmetrical sensorimotor axonopathy. Conduction slowing may be seen in severe poisoning.

TREATMENT

Rx

Initial treatment of arsenic toxicity includes supportive care, fluid repletion, decontamination if needed, and removal of the source of exposure. Hemodialysis may be required in patients who have significant renal dysfunction.

In cases of severe acute poisoning from inorganic arsenic, chelation is beneficial if it is instituted early. Dimercaprol (BAL; Table 22-2), which is the traditional chelating agent for arsenic, is effective in decreasing morbidity and mortality if it is administered within minutes to hours of acute exposure. In a small randomized trial, 2,3-dimercapto-1-propanesulfonate, which is not commercially available in the United States, significantly improved clinical symptoms, especially weakness, skin pigmentation, and lung disease, when given as 100 mg orally four times daily every other week for four cycles.[■] The oral analogue of BAL, dimercaptosuccinic acid (DMSA, succimer) is also useful for subacute or chronic arsenic poisoning. In chronic inorganic arsenic intoxication, however, a clear benefit of chelation therapy has not been demonstrated.

Hemolysis caused by arsine gas poisoning should be treated with prompt exchange transfusion. Exchange transfusion can restore functional erythrocytes, remove hemoglobin pigments, remove arsenic itself, and remove toxic products formed in the arsine-hemoglobin reaction.

PROGNOSIS

Outcomes after arsenic poisoning are influenced by the dose, type of arsenic compound, and route and chronicity of exposure. Acute high-dose arsine gas inhalation with severe and rapid systemic toxicity can be fatal. With appropriate treatment, however, recovery has been reported. After acute inorganic arsenic ingestion, rapid diagnosis and treatment, including chelation, can reduce mortality from about 75% to about 45%. After acute or chronic arsenic poisoning, electrocardiographic abnormalities and bone marrow suppression are generally reversible after exposure ceases, but encephalopathy and neuropathy may be permanent. Skin changes, such as hyperpigmentation and hyperkeratoses, can progress to cancer but also can improve if exposure is reduced.

CADMIUM TOXICITY

EPIDEMIOLOGY

Cadmium (Cd) is primarily found in zinc ores as cadmium sulfide. Serious human exposures usually come from industrial use, including the production of nickel-cadmium batteries, electroplating, soldering, and welding. Workplace exposures occur through inhalation of dust containing oxides of cadmium or fumes from welding or smelting of metals containing cadmium. Depending on the particle size, more than 50% of cadmium can be absorbed through the lungs, but little cadmium is absorbed through the gastrointestinal tract. Many plants take up cadmium from the environment, but poor oral bioavailability and relatively low food concentrations make ingestion of food an unlikely source of significant cadmium, except when cadmium-containing industrial waste is dumped in agricultural regions, such as happened when cadmium-containing mining waste in waterways that irrigated rice fields caused a large outbreak of renal disease. Patients with renal disease often developed osteomalacia (Chapter 244), which can lead to painful fractures. Today, the most common source of environmental cadmium exposure is through smoking, because tobacco concentrates cadmium from the soil, and smoking exposes the lungs to these high levels of cadmium. Compared with nonsmokers, an average smoker has about double the concentration of total body cadmium.

PATHOBIOLOGY

Cadmium commonly exists as a divalent ion (Cd^{2+}) in salts such as cadmium sulfide and cadmium oxide. Like other divalent metals, it binds to

and inhibits sulfhydryl-containing proteins and enzymes, thereby resulting in oxidative stress, cellular apoptosis, or necrosis. To survive in an environment where metals are ubiquitous, mammals have developed numerous antioxidant systems for protection. The one best studied in cadmium is the protein metallothionein, which is an intracellular, cysteine-rich protein especially found in the liver and kidney, where it is a potent protective antioxidant that binds to ionized cadmium.

CLINICAL MANIFESTATIONS

The most life-threatening manifestation of cadmium toxicity is acute chemical pneumonitis (Chapters 93 and 94). In persons who are exposed to high concentrations of cadmium fumes (e.g., smelter or welders), cadmium pneumonitis is manifested like metal fume fever with fever, malaise, myalgias, and elevated white blood cell counts that develop within 12 hours of exposure. Unlike metal fume fever, which typically resolves within a few days, cadmium pneumonitis can progress, with symptoms of dyspnea, hemoptysis, pulmonary edema, and respiratory distress as well as diffuse bilateral alveolar infiltrates on the chest radiograph.

The most problematic manifestation of chronic cadmium exposure is renal failure. Like lead, cadmium can damage the proximal tubules of the kidney and result in Fanconi syndrome (Chapter 128). Cadmium, bound to metallothionein, accumulates in the kidney, where its half-life is decades. Once metallothionein binding is saturated, toxicity can develop. Unlike the acute pulmonary manifestations, renal symptoms of cadmium toxicity may have a latency of 10 years of longer. Clinically, the signs of renal disease from cadmium are not different from those of other causes of proximal tubular disease. Patients will often manifest secondary symptoms of renal dysfunction, such as osteoporosis or ureteral stones, from impaired calcium metabolism. Hypertension and anemia from cadmium exposure are also likely secondary to its renal toxicity.

DIAGNOSIS

No laboratory tests can aid in the diagnosis of cadmium pneumonitis, so physicians must rely on their clinical suspicion or on alveolar infiltrates on a chest radiograph in a worker who presents with influenza-like symptoms.

Diagnosis of renal toxicity from chronic cadmium exposure relies predominantly on a 24-hour urine cadmium level standardized to grams of creatinine in a certified laboratory experienced in heavy metal testing (Table 22-1). In nonexposed nonsmokers, cadmium values should average about 0.08 µg per gram of creatinine. Levels of 7 µg cadmium per gram of creatinine and higher require removal from the workplace. For patients with levels above 3 µg cadmium per gram of creatinine, a medical evaluation and renal testing are indicated. Renal dysfunction has been reported in patients with lower levels, and the best test for evaluating renal function is urinary β_2-microglobulin levels, which serve as a useful early marker of toxicity; levels above 300 µg cadmium per gram of creatinine should raise concern for early kidney disease.

TREATMENT

Rx

Because of the similarities between cadmium pneumonitis and metal fume fever, any patient presenting with influenza-like symptoms after working with heated cadmium should be admitted for observation. If the metal a patient was welding or cutting is not known, any clinical or radiographic sings of noncardiogenic pulmonary edema would also warrant hospital admission for supplemental oxygen and pulmonary support.

There are no effective treatments for renal toxicity caused by cadmium. Chelation therapy is not recommended.

PROGNOSIS

Patients with cadmium pneumonitis can survive with aggressive treatment, but severe cases can result in death, and survivors can have persistent restrictive lung disease. Because of its long half-life in the kidney, cadmium-induced renal damage is largely irreversible unless it is detected early when body burden is low (urinary cadmium < 10 µg per gram of creatinine).

SYNDROMES SPECIFIC TO OTHER TOXIC METALS

Aluminum (Dialysis Dementia)

In the 1970s, renal failure patients exposed to aluminum-containing phosphate binders or aluminum-contaminated dialysis fluid developed progressive neurologic impairment and multifocal seizures. Also known as dialysis dementia, this syndrome developed during weeks to years and was often fatal if undetected. Aluminum-containing phosphate binders and dialysis fluids are no longer used in patients with renal failure, and dialysis patients are routinely screened for aluminum toxicosis. However, occasional outbreaks continue to be reported, typically due to contamination of dialysis fluid with aluminum from electric pumps or drums.

Aluminum is found in air, soil, and water and is a component of metal alloys used in the home, such as cookware. Aluminum is also found in some antacids (sucralfate) and as aluminum potassium sulfate (Alum), which is used to treat hemorrhagic cystitis. Additional iatrogenic sources of aluminum include total parenteral nutrition solutions and vaccines. Ingested aluminum is not metabolized by the body and is excreted unchanged by the kidney. In patients with impaired renal function, aluminum is bound to transferrin and concentrates primarily in bone. Smaller amounts, however, are distributed to heart, liver, kidney, and brain. Ingested aluminum is thought to affect several biochemical functions, including neurotransmitter manufacture, uptake, and release. Aluminum also affects erythropoiesis and the normal function of bone, likely partially by interference with parathyroid function.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Clinical effects in patients with chronic exposures include anemia, osteomalacia, and neurologic effects including memory loss, tremor, dyspraxia, encephalopathy, and seizures. The relationship between aluminum and Alzheimer dementia has been debated, but the complex characteristics of aluminum bioavailability make it difficult to produce conclusive evidence.

Serum and urine aluminum levels can be obtained to estimate exposure. A serum aluminum concentration should not exceed 2 µg/L, and a 24-hour urine concentration is expected to be 4 to 12 µg per gram of creatinine in a patient with typical background aluminum exposure (Table 22-1).

TREATMENT AND PROGNOSIS

Rx

Management of aluminum toxicity centers on removal from the source of exposure. The only chelator with proven benefit is deferoxamine (Table 22-2), which has a high affinity for aluminum and forms a dialyzable aluminum-deferoxamine complex.[9,10] If it is detected early, successful treatment with full recovery of neurologic function has been reported with deferoxamine.

Bismuth (Bismuth Encephalopathy)

Bismuth salts used to treat gastrointestinal disorders include bismuth subgallate, bismuth citrate, bismuth subnitrate, and bismuth subsalicylate (the active ingredient in Pepto-Bismol). Bismuth (Bi) is poorly absorbed from the gastrointestinal tract and is still used today to treat peptic ulcer disease and diarrhea. It also is used as an oral deodorant for patients with colostomies. Idiopathic bismuth toxicity is almost always associated with the chronic ingestion of over-the-counter preparations.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Bismuth toxicity is manifested primarily as a subacute encephalopathy with ataxia and incoordination followed by progressive memory loss, behavioral changes, insomnia, and muscle cramps. As symptoms progress, a prevalent feature is limb myoclonus when patients are startled or as they fall asleep. As symptoms worsen, myoclonus can progress to involve the whole body, including the tongue, and can occur without stimulus. Seizures can also develop.

The diagnosis of bismuth encephalopathy can be difficult. Blood and urine levels can indicate exposure to bismuth (Table 22-1) but do not correlate with symptoms. Furthermore, electroencephalographic findings, laboratory tests, and imaging studies are not specific. Because of its similarity in onset and symptoms, bismuth encephalopathy is often misdiagnosed as Creutzfeldt-Jakob disease (Chapter 415) or other progressive encephalopathies. Therefore, any patient in whom these diagnoses are being entertained should be screened for the use of bismuth products. As bismuth causes black stools, the finding of heme-negative black stool in a patient with a rapidly progressing encephalopathy should also warrant an investigation for bismuth toxicity.

Cobalt (Thyroid and Cardiac Toxicity)

Cobalt (Co) is an essential trace element that serves as the catalytic center of vitamin B_{12}. Toxicity is most commonly associated with chronic ingestion of

TREATMENT AND PROGNOSIS

Rx

The treatment of bismuth encephalopathy is to stop the use of the offending agent. Although there have been case reports purporting the benefits of metal chelators (e.g., DMSA, BAL, and 2,3-dimercapto-1-propansulfonic acid; Table 22-2) the rarity of the disorder makes a clinical trial impossible. Ethylenediaminetetraacetic acid (EDTA) may increase brain bismuth concentrations, and it should not be used. Although deaths have occurred from bismuth encephalopathy, full recovery is possible if the offending drug is stopped and the patient receives modern supportive care.

cobalt salts and more recently from cobalt-containing metal prostheses. One of the side effects of chronic ingestion is polycythemia. By stabilizing hypoxia-inducible transcription factors that normally respond to low concentrations of oxygen, cobalt mimics hypoxia and stimulates erythropoietin production. In the 1950s, cobalt chloride was used to treat iron deficiency anemia. Some of these patients developed hypothyroidism and goiter because of cobalt's ability to inhibit tyrosine iodinase. In addition, cobalt has caused dilated cardiomyopathy with pericardial effusions in adults who consumed large amounts of beer in which cobalt salts had been added as a foam stabilizer. Cobalt can also cause neurotoxicity with symptoms of hearing loss, visual impairment, and polyneuropathy.

Because cobalt is no longer used as a therapeutic or as an additive to beer, the concern for cobalt toxicity is now isolated to patients with cobalt-chromium prosthetic joints.[11] Cobalt liberated from a prosthesis has been linked to hypothyroidism, dilated cardiomyopathy, and neurotoxicity with metal-on-ceramics arthroplasty and metal-on-metal prosthetics. When it is viewed in the context of the millions of prosthetic joints placed each year and the few reported cases, this complication seems rare. The diagnosis in a patient with a cobalt-containing joint replacement requires symptoms of polycythemia, hypothyroidism, cardiomyopathy, and neurotoxicity. In these patients, a blood chromium level above 7 μg/L should stimulate a referral for a possible joint revision. There is little evidence that chelation improves outcomes in cobalt-poisoned patients.

Silver (Argyria)

Silver (Ag) is a precious metal long used in coinage and for its antibacterial properties. Silver is a broad-spectrum antimicrobial and commonly used as a topical antimicrobial in bandages, catheters, and medical devices. The ingestion of colloidal silver preparations as a "natural" supplement accounts for the majority of recent cases of significant toxicity.

When it is ingested continually, silver will be deposited in the skin and the liver.[12] The primary chronic toxicity of silver is argyria, which is a permanent blue-gray discoloration of the skin due to silver deposition over time (Fig. 22-1). Argyria can result from inhalation, ingestion, mucosal absorption, or dermal application or exposure. Argyria may be localized to the site of exposure, such as with corneal argyria, which was seen in the past from the use of colloidal silver-containing eye drops, or at the site of silver earrings and rings. Rarely, argyria has been reported after the implantation of silver-containing medical devices.

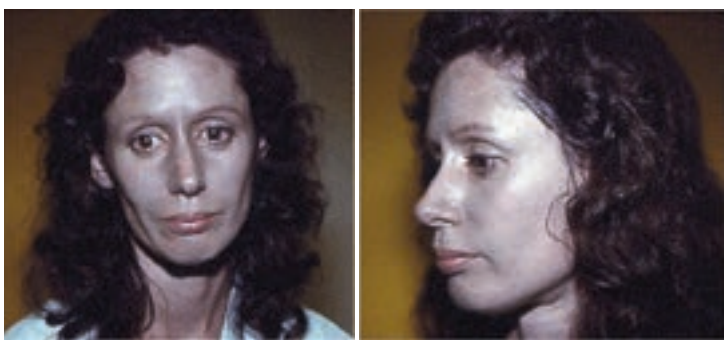

FIGURE 22-1. Frontal and side views of a 36-year-old woman with argyria show the gray discoloration of her face and neck. (From Jacobs R. "Argyria: my life story," *Clin Dermatol.* Elsevier, 2006;24:66-69. Figure 3).

The diagnosis of argyria is primarily based on history and physical examination. Localized skin biopsy, typically to differentiate argyria from malignant lesions, will show characteristic black-brown globules that are adherent to the dermal elastic fibers, blood vessels, basement membranes, hair follicles, and sweat glands on light microscopy; refractile particles on darkfield microscopy; and the presence of silver on scanning electron microscopy with energy dispersive radiography.

Chelation therapy is not effective for argyria, but successful laser treatment has been reported.[13]

Thallium (Neuropathy and Alopecia)

Thallium (Tl) has no beneficial role in the human body. Its salts are odorless, tasteless, well absorbed in the gastrointestinal tract, and very toxic. These characteristics make it a potential homicidal agent, and patients with thallium poisoning should always be considered victims of a crime. By interfering with potassium and sulfhydryl-containing enzymes, thallium interrupts normal energy production. Although it is toxic to all organs, the peripheral nervous system and integumentary system are the most sensitive.

The earliest clinical sign of thallium poisoning is a rapidly progressive, painful sensory polyneuropathy. Within 2 or 3 days of exposure, patients will describe painful burning paresthesias in their feet. These paresthesias can progress up the legs and, over time, can involve the hands. Motor nerves can also be affected, and profound weakness, including in respiratory muscles, can be misdiagnosed as Guillain-Barré syndrome (Chapter 420). Cranial neuropathies have also been reported. The best-known complication of thallium is painless hair loss, typically beginning 5 to 14 days after exposure and sometimes progressing to total body alopecia. Skin and nails can also be affected, with scaling of palms, acne-like lesions of the face, and Mees lines in the nails. Other neurologic symptoms of thallium poisoning include hallucinations, altered mental status, insomnia, psychosis, ataxia, and coma. Constipation, myalgias, pleuritic chest pain, arrhythmias, and hypotension can also occur.

Because systemic toxicities often develop before hair loss and because not all patients lose their hair, clinicians should consider thallium as a possible cause of any rapidly progressive painful neuropathy. In patients with thallium poisoning, a pulled hair will often have darkening of the hair root when it is visualized under a low-power light microscope. The definitive diagnosis of thallium poisoning requires the identification of elevated concentrations of thallium in a 24-hour urine sample (normal, <20 μg/specimen). As thallium undergoes enterohepatic circulation and is eliminated in the feces, treatment requires binding of thallium in the gut. The most effective antidote is Prussian Blue (a complex of potassium hexacyanoferrate), with recommended doses ranging from 3 g orally three times a day up to 250 mg/kg three times a day.[14]

Zinc (Myelopathy)

Zinc (Zn) is an essential mineral that is necessary for normal cellular functioning. It is involved in the catalytic activity of more than 100 enzymes and plays an important role in olfaction, taste, immune function, protein synthesis, and DNA synthesis. It is found naturally in a wide variety of foods and is absorbed in the jejunum, where it binds to metallothioneins. Zinc deficiency is manifested as a triad of dermatitis, diarrhea, and alopecia, which is reversible after zinc repletion.

In the setting of zinc overload, metallothioneins are upregulated; because copper has a higher affinity for metallothioneins, the result is increased copper binding and subsequent elimination of copper. Zinc-induced copper deficiency has been reported after overuse of zinc-containing supplements, with the improper use of zinc-containing denture adhesives, and from the presence of retained zinc-containing coins such as pennies.

The hematologic manifestations of zinc-induced copper deficiency include sideroblastic anemia, leukopenia, neutropenia, and myelodysplastic syndrome,[15] all of which are reversible with zinc cessation alone. In addition, a syndrome of myeloneuropathy, characterized by a spastic gait and sensory ataxia, is also associated with zinc overload/copper deficiency and may improve with copper supplementation and zinc cessation.

Diagnostic testing in patients suspected of zinc overload should include serum or urine zinc concentrations (Table 22-1), serum copper levels, and ceruloplasmin levels. In patients with myeloneuropathy, nerve conduction studies may show a sensory neuropathy, and magnetic resonance imaging (MRI) may show increased T2 signal in the cervical cord. Treatment of zinc overload–induced copper deficiency is primarily by cessation of zinc exposure and repletion of copper, such as with 6 mg/day of elemental copper

orally daily for the first week, then 4 mg/day for the next week, and finally 2 mg/day after that until serum levels return to normal. Various chelation regimens have been reported, but data are insufficient to support their routine use in the setting of chronic toxicity.

Manganese (Parkinsonism)

Manganese (Mn) is an essential element necessary for the function of several enzymes, including superoxide dismutase and glutamine synthetase. Although there is no clear syndrome associated with manganese deficiency, a well-described constellation of symptoms and findings is associated with manganese toxicity. Associated primarily with chronic occupational exposure, manganese toxicity results from inhalation of high concentrations of either manganese dust (e.g., in miners) or manganese fumes (e.g., smelting, grinding, and rarely welding).[16] More recently, manganese toxicity has been reported in abusers of methcathinone, when this drug was illegally synthesized with potassium permanganate. Although the mechanism is not clear, manganese causes selective toxicity in the globus pallidus, striatum, and substantia nigra pars reticulata. In contrast to idiopathic Parkinson disease (Chapter 409), manganese does not affect the substantia nigra pars compacta.

The symptoms of manganism develop insidiously after years of exposure. Patients first develop changes in appetite, muscle weakness, and apathy. On occasion, patients will develop signs of central excitation and can have what is termed manganese psychosis. Symptoms often progress to difficulties with gait, speech, and facial expression. The gait abnormalities involve a slow and clumsy gait with the inability to walk backward. Patients often freeze while turning and frequently fall. Symptoms can progress to muscle hypertonia in extension and a peculiar gait in which patients walk on the balls of the feet with their ankles extended. Patients also develop stuttering speech, masked facies, sleep disturbances, and myalgias. Unlike with idiopathic Parkinson disease, manganese-exposed patients do not typically develop a resting tremor, and these symptoms are almost always symmetrical at onset.

Although there is no definitive test of manganese toxicity, a history of exposure, confirmatory laboratory testing (Table 22-1), radiographic studies, and clinical symptoms can support the diagnosis. The best laboratory test is a whole blood manganese level (normal, <15 μg/L), with elevated levels indicative of exposure but not necessarily correlating directly with symptoms. Because manganese can concentrate in the globus pallidus, MRI often will show an intense T1-weighted signal in the globus pallidus bilaterally. As with blood, however, MRI is indicative of manganese exposure but does not correlate with clinical symptoms. Because of the differences in the affected brain regions between manganese toxicity and Parkinson disease, fluorodopa positron emission tomography (PET) scanning can help differentiate between these disorders: patients with manganese toxicity will typically have normal PET scans, whereas patients with Parkinson disease have abnormal scans.

Manganism is not typically responsive to treatment. Although chelation with $CaNa_2EDTA$ has been associated with increased urinary elimination of manganese, there are no convincing data as to whether patient outcomes are improved.

Grade A References

A1. Yeoh B, Woolfenden S, Lanphear B, et al. Household interventions for preventing domestic lead exposure in children. *Cochrane Database Syst Rev*. 2012;4:CD006047.

A2. Rogan WJ, Dietrich KN, Ware JH, et al. The effect of chelation therapy with succimer on neuropsychological development in children exposed to lead. *N Engl J Med*. 2001;344:1421-1426.

A3. Guha Mazumder DN, De BK, Santra A, et al. Randomized placebo-controlled trial of 2,3-dimercapto-1-propanesulfonate (DMPS) in therapy of chronic arsenicosis due to drinking arsenic-contaminated water. *J Toxicol Clin Toxicol*. 2001;39:665-674.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Department of Health and Human Services, Public Health Service, Agency for Toxic Substances and Disease Registry. Toxicological Profiles. http://www.atsdr.cdc.gov/toxprofiles/index.asp; Toxic Substances Portal. http://www.atsdr.cdc.gov/substances/index.asp; Accessed January 29, 2015.
2. Krishnan E, Lingala B, Bhalla V. Low-level lead exposure and the prevalence of gout: an observational study. *Ann Intern Med*. 2012;157:233-241.
3. Kosnett MJ. The role of chelation in the treatment of arsenic and mercury poisoning. *J Med Toxicol*. 2013;9:347-354.
4. Jang DH, Hoffman RS. Heavy metal chelation in neurotoxic exposures. *Neurol Clin*. 2011;29:607-622.
5. Rice KM, Walker EM Jr, Wu M, et al. Environmental mercury and its toxic effects. *J Prev Med Public Health*. 2014;47:74-83.
6. Mozaffarian D, Shi P, Morris JS, et al. Mercury exposure and risk of cardiovascular disease in two U.S. cohorts. *N Engl J Med*. 2011;364:1116-1125.
7. Tyler CR, Allan AM. The effects of arsenic exposure on neurological and cognitive dysfunction in human and rodent studies: a review. *Curr Environ Health Rep*. 2014;1:132-147.
8. Moon KA, Guallar E, Umans JG, et al. Association between exposure to low to moderate arsenic levels and incident cardiovascular disease. A prospective cohort study. *Ann Intern Med*. 2013;159:649-659.
9. Smith SW. The role of chelation in the treatment of other metal poisonings. *J Med Toxicol*. 2013;9:355-369.
10. National Kidney Foundation. KDOQI Clinical Practice Guidelines for Bone Metabolism and Disease in Chronic Kidney Disease. http://www.kidney.org/sites/default/files/docs/boneguidelines.pdf. Accessed January 29, 2015.
11. Brent J, Devlin JJ. Dilemmas about the toxicological consequences of metal-on-metal hip prostheses—what we do and do not know, and what we should do? *Clin Toxicol (Phila)*. 2013;51:195-198.
12. Hadrup N, Lam HR. Oral toxicity of silver ions, silver nanoparticles and colloidal silver—a review. *Regul Toxicol Pharmacol*. 2014;68:1-7.
13. Saager RB, Hassan KM, Kondru C, et al. Quantitative near infrared spectroscopic analysis of Q-Switched Nd:YAG treatment of generalized argyria. *Lasers Surg Med*. 2013;45:15-21.
14. Zhang HT, Qiao BP, Liu BP, et al. Study on the treatment of acute thallium poisoning. *Am J Med Sci*. 2014;347:377-381.
15. Gabreyes AA, Abbasi HN, Forbes KP, et al. Hypocupremia associated cytopenia and myelopathy: a national retrospective review. *Euro J Haematol*. 2013;90:1-9.
16. Furbee B. Welding and parkinsonism. *Neurol Clin*. 2011;29:623-640.

REVIEW QUESTIONS

1. A 58-year-old patient seeks your care for management of his hypertension. In his evaluation, you discover that he operates a gun range and often cleans up casings at the end of the day. You request a blood lead level that comes back elevated at 60 µg/dL. He has no other medical complaints. Along with pharmacologic management for his hypertension, which of the following lead treatment strategies would be the most efficacious?

A. Chelation with BAL and CaNa2EDTA
B. Chelation with DMSA
C. Chelation challenge, and if the 24-hour urine level is above 100 µg per gram of creatinine, chelate with DMSA
D. No intervention is warranted.
E. Proper respiratory protection at work and repeat lead levels in 1 month

Answer: E Adults who are asymptomatic and have blood lead levels below 70 µg/dL require no treatment besides removal of the source of exposure and decrease in workplace exposure through proper respiratory protection.

2. A 38-year-old health-conscious man is referred to you for arsenic poisoning. He has been seeing his primary care physician for symptoms of mild depression. He asked his physician to check him for heavy metal poisoning because he has well water and read that lead could cause symptoms similar to his. A 24-hour urine test result showed lead and mercury levels to be within normal limits but revealed an extremely elevated urine arsenic level (530 µg per gram of creatinine). What is the most prudent next step?

A. Admit him to the hospital and immediately begin chelation with DMSA.
B. Ask about any history of seafood ingestion and send his urine for arsenic speciation.
C. Begin outpatient treatment with DMSA and repeat testing in 3 weeks.
D. Obtain blood arsenic levels to determine if he has had a recent exposure.
E. Obtain hair arsenic levels to determine if he has had a remote exposure.

Answer: B Organic arsenic is considered largely nontoxic to humans. Found in shellfish and seafood, it is a common cause of elevated urinary arsenic levels. In patients without symptoms of acute arsenic poisoning, a detailed history of seafood ingestion and sending a urine sample for speciation of arsenic would be the next step.

3. A 32-year-old chemical worker ingests a vial of methylmercury in a suicide attempt. She is admitted to the hospital, lavaged, and begun on DMSA chelation. The development of which of the following symptoms would concern you the most in respect to her long-term outcome?

A. Hypertension and tachycardia
B. Myalgias
C. Prolonged QTc on her electrocardiogram
D. Proteinuria
E. Visual field constriction and perioral numbness

Answer: E The most devastating effect of organic mercury intoxication is neural degeneration. Early symptoms of this include constriction of visual fields, tremor, and paresthesias. Symptoms can progress to ataxia and ultimately to encephalopathy and death. Elemental mercury exposure can mimic pheochromocytoma and can be manifested with muscle weakness and myalgias; these symptoms can improve when the source of mercury exposure is removed. Proteinuria would be a potential manifestation of inorganic mercury toxicity, but renal damage tends to resolve over time. Arsenic, not mercury, is associated with a prolonged QTc.

4. A 68-year-old Hispanic woman with end-stage renal disease requiring hemodialysis three times a week develops a progressive encephalopathy. The water used for dialysis was tested and found to be normal. Which one of the following would you think is most responsible for her symptoms?

A. Occasional use of Pepto-Bismol
B. A prosthetic hip joint
C. Her sucralfate prescription
D. The homeopathic use of colloidal silver
E. Traditional Hispanic folk remedy for gastrointestinal distress

Answer: C Sucralfate is most likely to cause encephalopathy in a dialysis patient from aluminum toxicity. A prosthetic hip with cobalt has been linked to cardiomyopathy and hypothyroidism. Pepto-Bismol can cause a progressive encephalopathy but typically only after prolonged and regular use. Hispanic folk and Chinese herbal medicines can occasionally contain arsenic and elemental mercury, but either would be less likely to cause the patient's symptoms. Colloidal silver ingestion typically affects only the skin.

23

EPIDEMIOLOGY OF AGING: IMPLICATIONS OF AN AGING SOCIETY

ANNE B. NEWMAN AND JANE A. CAULEY

DEMOGRAPHY: AGING OF SOCIETIES WORLDWIDE

More than 40 million adults older than 65 years are now living in the United States, and they account for 13% of the population. Even more notably, 5.5 million people (1.8% of the population) are older than 85 years, and more than 50,000 people (0.02% of the population) are older than 100 years. Although only 13% of the U.S. population is older than 65 years, they account for 36% of all medical expenditures.[1]

The trend toward an aging society has expanded to become a global health issue. The global population aged 65 years and older is more than 500 million people, about 7% of the world's population. By 2040, the world is projected to have 1.3 billion older people, accounting for 14% of the total population. In 2040, 28% of the population in western Europe will be 65 years and older, including about 9.3% older than 80 years. Japan is currently the oldest country in the world, with 22% of the population aged 65 years and older, compared with 18% in western Europe and 21% in North America (Table 23-1). Between 2005 and 2040, the overall world population is projected to increase by 35%, whereas the percentage aged 85 years or older will increase by 300% and the percentage aged 100 years or older will increase by 750%. The estimated number of centenarians was about 270,000 in 2005, but this number is projected to reach 2.3 million by 2040.[2]

Population aging, which has been accelerating in the United States and western Europe for more than 100 years, is now affecting countries such as India and China, where the population older than 65 years is projected to increase more than three-fold by 2040. The population aged 65 years and older was already 166 million in China and India in 2008, nearly one third of the world's total, and the absolute number will increase to 550 million by 2040. In the United States, increases in the elderly population are most prominent in minorities, especially Hispanics.

Life expectancy is also changing dramatically. Life expectancy is highest in Japan at 82 years and is in the 78- to 80-year range for other developed countries. Life expectancy has continued to increase at all ages and especially for people older than 85 years, except in some parts of the developing world, such as sub-Saharan Africa, where HIV/AIDS has had a major negative impact. In the United States, an average 65-year-old woman can expect to live for an additional 19.6 years, and the average 65-year-old man can expect to live for an additional 16.8 years. These changes have many important impacts on society because rates of illness and disability increase with age, and the relative proportion of younger adults who can help care for this older population is not increasing.

THE HEALTH OF OLDER ADULTS

Mortality

The current generation of older adults is healthier than previous generations because of the successful elimination of many infectious diseases and the prevention of injuries, but the population at risk for chronic diseases is increasing. In the United States in 1900, infectious diseases accounted for the largest share of deaths after 65 years of age, whereas heart disease, cancer, stroke, and lung disease account for more than half of all deaths today.

As people age, their mortality rate increases. Death rates are 1.8% per year in men and 1.2% per year in women between the ages of 65 and 70 years, increasing to 13% per year in women and 15% per year in men older than 85 years.[3] With increasing age, causes of death in older adults shift away from cancer and cardiovascular disease and toward stroke and dementia. Nevertheless, cardiovascular disease and cancer remain the top two causes of death in older adults into advanced old age (Table 23-2).

However, many factors contribute to mortality in old age, and death is usually multifactorial, with contributions from other chronic underlying conditions. Mortality in older adults is related to sociodemographic status and health habits; cardiovascular risk factors; and pulmonary, vascular, kidney, physical, and cognitive function. The inflammatory marker interleukin-6 is consistently associated with death from many different primary causes, thereby suggesting that inflammation is a common underlying pathway to death.[4]

Morbidity

The most common chronic conditions in older adults are hypertension, high cholesterol levels, and ischemic heart disease (Fig. 23-1). Prevalence varies by sex, with men having more heart disease than women and women having more arthritis than men. Because women live longer than men do, the prevalence of dementia is higher in women than in men, but incidence rates are more similar. Older African Americans and Hispanics have higher rates of hypertension, diabetes, and metabolic syndrome compared with older whites.[5]

The majority of older adults have one or more chronic health conditions, and 50% have two or more. Because of this high prevalence of multiple conditions, the care of older adults requires a balanced approach that considers the impact of treatment of one condition on the other conditions and the potential that some conditions may be masked by multiple overlapping symptoms. The presence of multiple conditions is commonly referred to as comorbidity, although multimorbidity may be a more accurate term.

Multimorbidity is strongly associated with disability, with a stepwise increase in the proportion of individuals who have self-reported difficulty in activities of daily living according to their number of medical conditions. Prospectively, the number of comorbid conditions predicts future disability. Multimorbidity also is associated with having a slower gait speed and poorer lower extremity strength and balance. Furthermore, disability and comorbidity provide distinct information and should not be considered equivalent constructs. The combined influence of multiple chronic diseases on physical functioning can be greater than the simple sum of their effects.

Subclinical Disease

Subclinical diseases are also common and increasingly recognized as important contributors to the risk of disability and mortality. With the advent of noninvasive functional testing and imaging, it is clear that many conditions can be advanced in older people, even if they are not symptomatic. Symptoms may also be nonspecific and attributed to aging rather than to underlying disease. For example, population studies of cardiovascular disease reveal that strokes are present on magnetic resonance imaging in about one third of older adults with no history of stroke, and advanced atherosclerosis is present also in one third of older adults with no history of a prior myocardial infarction. Ankle-brachial index screening documents that peripheral artery disease is 5 to 10 times more common than classic intermittent claudication. Other conditions, such as chronic obstructive pulmonary disease and osteoarthritis, are also more common when noninvasive testing is performed compared with clinically diagnosed disease. Importantly, subclinical disease burden is strongly related to mortality.

Obesity

Obesity (Chapter 220) is increasing in older adults at a rate faster than in any other age group. Obesity and disability disproportionately affect women and especially women of color. About one third of whites older than 60 years are obese, compared with about 50% of blacks and 40% of Hispanics. Obesity has a major impact on disability, including difficulty in walking or climbing steps. The epidemic of obesity is also contributing to the large increases in the incidence of diabetes in older adults. Increasing rates of obesity may halt or reverse increases in life expectancy, reduce quality of life, and increase disability.

Geriatric Syndromes

Many multifactorial conditions in older adults do not fall into a single disease category and are called geriatric syndromes. For example, urinary incontinence (Chapter 26) is a functional loss of bladder control that can be influenced by poor mobility, poor vision, loss of urinary concentrating ability, and certain medications as well as impaired bladder structure or function. Other important geriatric syndromes include impaired sleep (Chapter 405), delirium (Chapter 28), falling (Chapter 25), and weight loss. Rather than focusing on identification of a primary cause, all contributing factors need to be addressed in the evaluation and treatment of these syndromes.

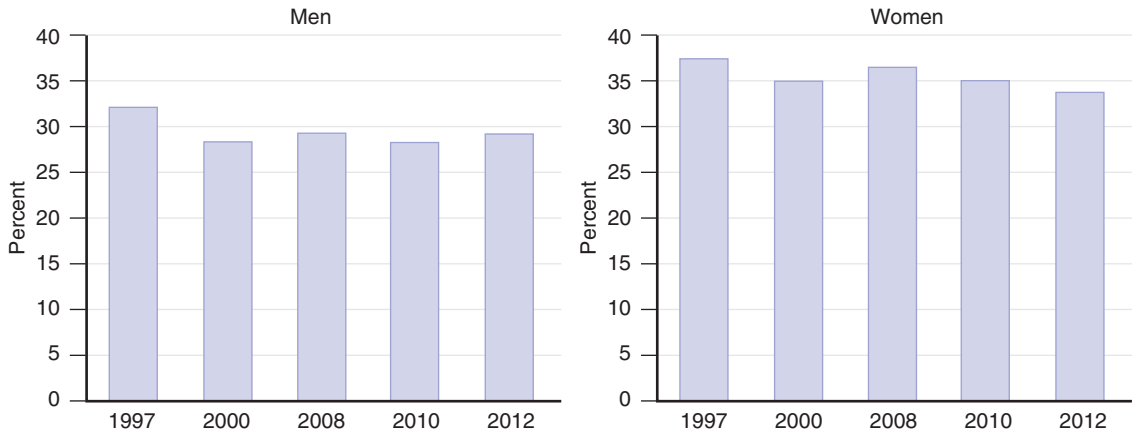

FIGURE 23-1. Prevalence of complex activity limitation among those 65 years and older in the United States. A complex activity limitation is defined as having one or more of the following limitations: self-care (activities of daily living or instrumental activities of daily living), social, or work. (Source: Centers for Disease Control and Prevention, National Center for Health Statistics: Health, United States, 2012, and Health, United States, 2010. Data from the National Health Interview Survey. http://www.cdc.gov/nchs/data/hus/hus12.pdf. Accessed January 29, 2015.)

TABLE 23-1 PERCENTAGE OF OLDER POPULATION BY REGION, 2014-2050

REGION	≥65 YEARS	≥75 YEARS	≥80 YEARS
NORTHERN AFRICA			
2014	4.9	1.7	0.8
2050	13.6	5.5	3.1
EASTERN AFRICA			
2014	2.8	0.9	0.4
2050	5.3	1.8	0.9
MIDDLE AFRICA			
2014	2.8	0.8	0.3
2050	5.4	1.7	0.8
SOUTHERN AFRICA			
2014	6	2.1	1
2050	11	4.9	3.1
WESTERN AFRICA			
2014	3.1	0.9	0.4
2050	5.3	1.8	0.8
ASIA (EXCLUDES NEAR EAST)			
2014	7.6	2.8	1.4
2050	18.8	9	5.2
WESTERN ASIA (NEAR EAST)			
2014	5.2	2	1
2050	14.2	6	3.4
EASTERN EUROPE			
2014	14.5	6.7	3.5
2050	27.9	13.7	8.5
WESTERN EUROPE			
2014	19.5	9.5	5.5
2050	27.5	16.6	11.6

Source: U.S. Census Bureau, International Data Base. http://www.census.gov/population/international/data/idb/informationGateway.php. Accessed January 29, 2015.

TABLE 23-2 LEADING CAUSES OF DEATH: U.S. POPULATION, 2011

CAUSE	RATE PER 100,000
Diseases of heart	173.7
Malignant neoplasms	168.6
Chronic lower respiratory diseases	42.7
Cerebrovascular diseases	37.9
Accidents (unintentional injuries)	38.0
Alzheimer disease	24.6
Diabetes mellitus	21.5
Influenza and pneumonia	15.7
Nephritis, nephrotic syndrome, and nephrosis	13.4
Intentional self-harm (suicide)	12.0
Septicemia	10.5
Chronic liver disease and cirrhosis	9.7
Essential hypertension and hypertensive renal disease	8.0
Parkinson disease	7.0
Pneumonitis due to solids and liquids	5.3

From Hoyert DL, Xu JQ. Deaths: preliminary data for 2011. *Natl Vital Stat Rep.* 2012;61:6.

Disability

In older adults, disability is usually related to multiple chronic conditions. Disability can be slowly progressive and chronic, or it can be sudden and catastrophic, as in the case of a stroke or hip fracture. Levels of disability include loss of basic self-care skills, such as bathing and toileting; loss of ability to live in the community for skills such as shopping and paying bills; loss of mobility; and loss of ability to function in high-level tasks. Classification schemes for disability in geriatrics are reviewed in Chapter 24.

The majority of older adults are not disabled. Data from the National Health Interview Survey show that disability levels declined in the 1980s and 1990s but have largely plateaued in recent years (Fig. 23-2). In 2012, about 28% of men and 34% of women older than 65 years reported difficulty with complex activities. Disability for activities of daily living appears to be declining worldwide, although evidence suggests that less severe levels of disability are becoming more common. It is possible, however, that the obesity epidemic will result in more years with disability.

Frailty versus Vigor

Frailty has been described as a state of decreased physiologic reserve and increased vulnerability to stress. Frailty is a wasting syndrome characterized by weakness, fatigue, low activity, slow movement, and weight loss. It has also been characterized by poor physical function alone or by a high burden of chronic disease. By the definition of a wasting syndrome, frailty is present in 7% of community-dwelling adults aged 65 years and older and 25% of adults aged 85 years and older. These estimates undoubtedly underestimate the prevalence of frailty because frail individuals are more likely to be temporarily in a medical care facility and unable to participate at any given time.

Loss of muscle mass, or sarcopenia, is an underlying component of frailty,[6] but excess fat, especially visceral fat and muscle fat, has a greater influence on physical function, inflammation, and metabolism than does low muscle mass. In the elderly, muscle strength, including contractile force, mitochondrial function, and speed of contraction or power, is more important to physical

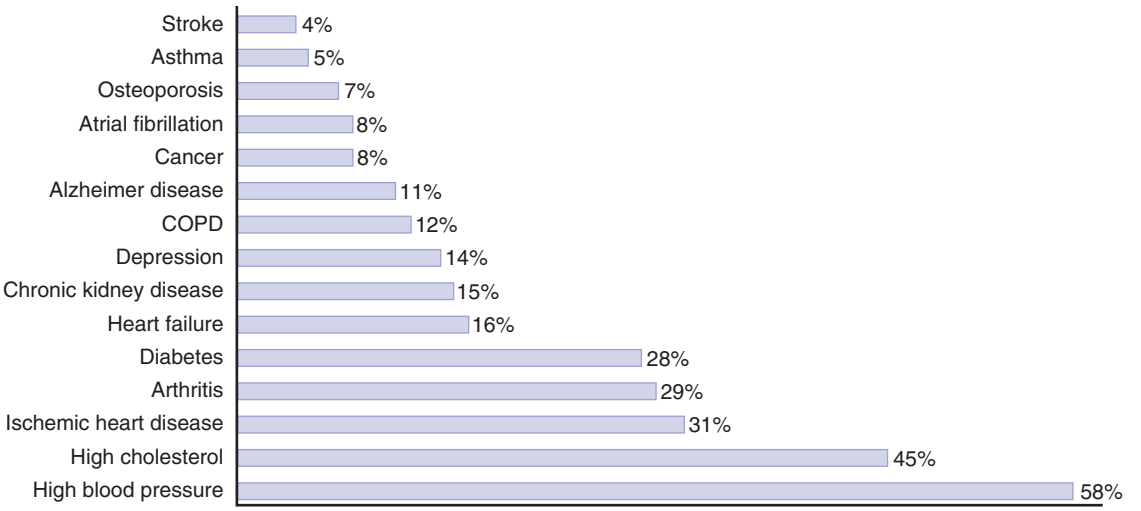

FIGURE 23-2. Percentage of Medicare fee-for-service beneficiaries with 15 selected chronic conditions: 2010. (Source: Centers for Medicare and Medicaid Services: Chronic Conditions among Medicare Beneficiaries, Chartbook: 2012 Edition. Based on 2010 Centers for Medicare and Medicaid Services administrative claims data for 100% of Medicare beneficiaries enrolled in the fee-for-service program.) COPD = chronic obstructive pulmonary disease.

function than muscle mass. Important weakness can be assessed in an office setting by measuring a grip strength (<30 kg in men and <20 kg in women).

Longevity and Healthy Aging

About 20 to 30% of longevity is heritable, similar to many complex chronic diseases.[7] Animal models of aging show that lifespan can be extended several-fold by altering genes, especially in the metabolic pathway. In lower organisms and mice, calorie restriction, achieved through long-term daily restriction of available calories, extends life, but this approach has mixed results in primates. Many hormonal factors are reduced with aging, but hormone replacement with estrogen, growth hormone, testosterone, and adrenal androgens is not beneficial for longevity and may be harmful. Among genetic factors, absence of the *ApoE4* allele remains the most robust predictor of longevity, and recent genome-wide association studies of centenarians also point to genes such as *FOX03A* in the pathway for metabolic control.

Many health conditions in late life may stem from early life exposures, including the health of the mother during pregnancy. For example, low birthweight or famine during pregnancy predicts early cardiovascular disease and diabetes when babies are later exposed to higher calorie intake. In Sweden, a lower lifetime exposure to infectious disease was associated with increased life expectancy. More research is needed to determine the critical periods for development and the optimal environmental exposures that can enhance healthy aging later in life.

PREVENTION

Osteoporosis treatment is effective in reducing fracture risk in advanced old age,[8] and treatment of hypertension is effective in decreasing mortality after 80 years of age.[A1] Statins are effective in reducing coronary heart disease in older adults.[A2] Physical activity significantly decreases the risk of disability[A3] and may reduce the risk of dementia, although the latter finding is primarily from observational studies that may be confounded by the fact that healthier people are more able to exercise. Ongoing research is evaluating the potential for aspirin to prolong active life in older adults and whether testosterone replacement in hypogonadal men will improve physical, cognitive, or sexual function in older men.

Grade A References

A1. Briasoulis A, Agarwal V, Tousoulis D, et al. Effects of antihypertensive treatment in patients over 65 years of age: a meta-analysis of randomised controlled studies. *Heart*. 2014;100:317-323.
A2. Savarese G, Gotto AM Jr, Paolillo S, et al. Benefits of statins in elderly subjects without established cardiovascular disease: a meta-analysis. *J Am Coll Cardiol*. 2013;62:2090-2099.
A3. Pahor M, Guralnik JM, Ambrosius WT, et al. Effect of structured physical activity on prevention of major mobility disability in older adults: the LIFE study randomized clinical trial. *JAMA*. 2014;311:2387-2396.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Agency for Healthcare Research and Quality. The High Concentration of U.S. Health Care Expenditures. Research in Action, Issue 19. June 2006. http://www.ahrq.gov/research/findings/factsheets/costs/expriach/index.html; Accessed January 29, 2015.
2. Kinsella K, Wan H. International population reports, P95/09-1, An aging world: 2008. Washington, DC: U.S. Government Printing Office; 2009. http://www.aicpa.org/research/cpahorizons2025/globalforces/socialandhumanresource/downloadabledocuments/agingpopulation.pdf; Accessed January 29, 2015.
3. Murphy SL, Xu J, Kochanek KD. Deaths: final data for 2010. *Natl Vital Stat Rep*. 2013;61:1-117.
4. Beavers KM, Hsu FC, Houston DK, et al. The role of metabolic syndrome, adiposity, and inflammation in physical performance in the Health ABC Study. *J Gerontol A Biol Sci Med Sci*. 2013; 68:617-623.
5. Klijs B, Nusselder WJ, Looman CW, et al. Contribution of chronic disease to the burden of disability. *PLoS ONE*. 2011;6:e25325.
6. Kim TN, Choi KM. Sarcopenia: definition, epidemiology, and pathophysiology. *J Bone Metab*. 2013;20:1-10.
7. Newman AB, Murabito JM. The epidemiology of longevity and exceptional survival. *Epidemiol Rev*. 2013;35:181-197.
8. Cauley JA. Public health impact of osteoporosis. *J Gerontol A Biol Sci Med Sci*. 2013;68:1243-1251.

24

GERIATRIC ASSESSMENT

DAVID B. REUBEN

Geriatric assessment is a broad term used to describe the evaluation of older patients, a process that recognizes the diverse medical and psychosocial conditions that influence the health status of older persons. In addition to the diseases that are common in the elderly population, these influences include social, psychological, and environmental factors. Geriatric assessment can range from brief screens by individual clinicians to an intensive interdisciplinary process that includes both evaluation and management.

Three fundamental concepts guide geriatric assessment and the resulting medical management. At the core of geriatric assessment is functional status, both as a dimension to be evaluated and as an outcome to be improved or maintained. A second overarching concept guiding geriatric assessment is prognosis, particularly life expectancy. Finally, geriatric assessment must be guided by the patient's goals.

FUNCTIONAL STATUS

Functional status can be viewed as a summary measure of the overall impact of health conditions in the context of an elderly person's environment and social support network. The underlying framework of functional status is a hierarchy of increasing complexity, beginning with specific physical movements (e.g., lifting, walking) that are integrated into higher level activities (e.g., fulfilling occupational and social roles). Impairment of functional status can be triggered by the onset of disease, deconditioning, changes in social support or environment, and advanced age.

Most commonly, older adults' functional status is assessed at two levels: activities of daily living (ADLs) and instrumental activities of daily living (IADLs). ADLs refer to self-care tasks such as bathing, dressing, toileting, maintaining continence, grooming, feeding, and transferring. Dependency in these tasks, which is present in up to 10% of older persons, usually requires full-time help at home or placement in a nursing home.

IADLs refer to tasks that are integral to maintaining an independent household, such as using the telephone, doing laundry, shopping for groceries, driving or using public transportation, preparing meals, taking medications, performing housework, and handling finances. Dependency in IADLs is more common, and almost 20% of persons aged 75 years or older are impaired in at least one task. With the progressive loss of multiple IADL functions, older persons find it more difficult to remain in their homes. Accordingly, many social services (e.g., Meals On Wheels, homemaker

TABLE 24-1 LIFE EXPECTANCY IN VARIOUS STATES OF FUNCTIONAL HEALTH ACCORDING TO AGE, GENDER, AND INITIAL FUNCTIONAL STATUS*

AGE	INITIAL FUNCTIONAL CATEGORY	LIFE EXPECTANCY IN YEARS IN EACH FUNCTIONAL STATUS							
		WOMEN				MEN			
		Independent Years	*Mobility-Disabled Years*	*ADL-Disabled Years*	*Total Years*	*Independent Years*	*Mobility-Disabled Years*	*ADL-Disabled Years*	*Total Years*
70 years	Independent	10.0	4.0	2.7	16.7	8.5	2.6	1.0	12.1
	Mobility disabled	7.3	5.6	2.8	15.7	5.6	4.1	1.1	10.7
	ADL disabled	3.0	2.9	5.6	11.5	1.6	1.5	3.4	6.5
75 years	Independent	7.0	3.6	2.6	13.2	6.0	2.4	1.0	9.4
	Mobility disabled	4.0	5.2	2.8	12.0	2.9	3.8	1.1	7.9
	ADL disabled	1.1	1.8	5.3	8.2	0.5	0.8	3.1	4.4
80 years	Independent	4.7	3.2	2.4	10.3	4.1	2.2	0.9	7.2
	Mobility disabled	2.0	4.4	2.7	9.0	1.4	3.3	1.0	5.7
	ADL disabled	0.4	1.0	4.7	6.0	0.2	0.4	2.6	3.1
85 years	Independent	3.3	2.9	1.8	8.0	2.9	2.1	0.7	5.8
	Mobility disabled	1.0	3.6	2.3	6.9	0.7	2.8	0.9	4.4
	ADL disabled	0.1	0.5	4.0	4.6	0.0	0.2	2.1	2.3

*Using 1988-1990 Established Populations for Epidemiologic Studies of the Elderly Data
ADL (activities of daily living) disabled = self-report of being unable to do or requiring human help to perform any of the following: bathing, transferring from bed to chair, dressing, eating, and using the toilet. Mobility disabled = self report of being unable to walk a half-mile or to walk up and down stairs to the second floor without help.

services, transportation services) are available to compensate for these deficiencies. A move to an assisted living facility can provide most IADL functions, but many facilities do not routinely provide assistance with ADLs except at an additional cost. In the United States, the costs of assisted living facilities are not covered by Medicare.

At a higher level of function, advanced activities of daily living (AADLs) refer to the ability to fulfill societal, community, and family roles as well as to participate in recreational or occupational tasks. These advanced activities vary considerably from individual to individual but may be valuable in monitoring functional status before the development of disability.

The choice of functional assessment tool depends on the characteristics of the population being assessed. For example, nursing home residents are almost always completely dependent in IADLs, so the focus should be on assessing ADLs and other basic dimensions of health. Hospitalized older persons should be assessed with regard to their prehospitalization functional status to provide insight into what may be achievable as well as their functional status at the time of discharge to identify any existing gap and to facilitate plans to close it.

Functional status is usually measured by self-report or proxy report. However, physical and occupational therapists often add objective information using structured clinical examinations or assessments. In addition, dimensions such as mobility and balance that contribute to function can be assessed by objective measures (described later).

Functional status should be assessed periodically: at the time of an initial visit; after a major illness; and at the time of social milestones, such as the illness of a spouse or a change in living or working situation. Changes in functional status should always prompt further diagnostic evaluation and intervention unless the change is expected and reflects a trajectory that is consistent with the patient's wishes. Measurement of functional status can be valuable in monitoring response to treatment (especially of chronic diseases) and can provide prognostic information that is useful in planning short- and long-term care.

PROGNOSIS

Life expectancy affects both the assessment process and the management decisions based on that assessment. For some older patients, comorbidities can worsen prognosis, such that screening tests (e.g., mammography) and treatments (e.g., for hypertension) with demonstrated effectiveness would not be beneficial within the expected survival period.

The probability that a patient will survive for a specified time (e.g., 5 years from the time of the assessment) or to a specified age (e.g., to age 100 years) can be estimated on the basis of age, gender, and race with government-generated life tables. Life expectancy also can be estimated with online calculators (see, for example, *eprognosis.org*) or by incorporating clinical characteristics. For example, better cognitive status and less comorbidity are the best predictors of 5-year survival in nonagenarians.[1] One simple approach is to incorporate functional status in addition to age and gender (Table 24-1).[2,3]

PATIENT GOALS

As people age, their current and future health may become a prominent factor in determining and achieving their life goals. Among very old patients, goals may be limited to achievement of a functional or health state (e.g., being able to walk independently), control of symptoms (e.g., pain, dyspnea), maintenance of their living situation (e.g., remaining at home), or short-term survival (e.g., living long enough to reach a personal milestone, such as an upcoming holiday). Sometimes the goals of patients and physicians differ. For example, a patient may want a cure, whereas the physician believes that only symptom management is possible. Conversely, the physician may believe that a better outcome is possible, but the patient declines to pursue the recommended path (e.g., hip replacement to restore mobility).[4] Physicians and patients alike must recognize that the elderly may receive less benefit from some interventions than younger do patients because of other adverse prognostic factors. For frail older persons with multiple chronic conditions or a short life expectancy, clinicians and patients should work together to identify the patient's personal goals within and across a variety of dimensions (e.g., symptoms, physical functional status, social and role functioning).[5] The care plan can then be developed and framed in the context of meeting these goals.

COMPONENTS OF GERIATRIC ASSESSMENT

Geriatric assessment begins with a medical evaluation. Some aspects (described later) that are rarely abnormal in younger adults (e.g., mobility, cognition) may cause substantial morbidity in older persons. In addition, some clusters of abnormal findings, such as muscle wasting, poor hygiene, bruises, pressure sores, and contractures, should raise the suspicion of elder mistreatment, neglect, or abuse. In such cases, patients should be questioned and examined without family members or caregivers present. Patients should then be queried as to whether anyone has threatened or hurt them, whether they have been receiving enough care, and whether anyone has taken their things. Answers to these questions may provide confirmatory information and prompt a report to adult protective services.

Nonmedical assessments are also important because they can identify problems that should be addressed and may be key to achieving the patient's goals. In addition to an assessment of functional status, older persons' environmental, financial, and nonfinancial support should be assessed. In some situations, particularly when older persons become acutely ill or experience

caregiver stress or loss of a loved one, assessment of spiritual needs, with appropriate referral, may be valuable.

Advance Directives

Clinicians should discuss older patients' preferences for specific treatments while they still have the cognitive capacity to make these decisions. Patients should be asked to identify a spokesperson to make medical decisions for them if they cannot speak for themselves. This information should be conveyed through a durable power of attorney for health care, which also allows patients to specify treatments they do not want. Many states have allowed the use of Physician Orders for Life-Sustaining Treatment (POLST), a specific advance directive that documents a patient's end-of-life treatment preferences and serves as an order sheet. The standardized form, which is signed by both the physician and the patient, must be honored in all settings of care, including by emergency medical technicians responding to 911 calls.

Medical Assessment

The medical assessment includes vision; hearing; cognition; mood and affect; falls, mobility, and balance; medication review; nutrition; and urinary incontinence. Numerous screening instruments have been developed that assess many of these dimensions. Another important medical component of geriatric assessment includes[5] a determination of whether preventive services are up-to-date.

Vision

Each of the four major eye diseases—cataract, age-related macular degeneration, diabetic retinopathy, and glaucoma (Chapter 423)—increases in prevalence with age. Moreover, presbyopia is virtually universal, and most older persons require eyeglasses. Visual impairment has been associated with increased risk of falls, functional and cognitive decline, immobility, and depression.[6] Corrective lenses or other treatments may restore vision or prevent further decline of visual function.

A single question can be used to screen for visual impairment: Do you have difficulty driving, watching television or reading, or doing any of your daily activities because of your eyesight, even while wearing glasses?

The Snellen eye chart is the standard method of screening for visual acuity. The patient is asked to stand 20 feet from the chart and to read letters. Inability to read letters on or below the 20/40 line with the best-corrected vision (using glasses) indicates the need for further evaluation.

Hearing

Hearing loss (Chapter 428), which affects more than 60% of community-dwelling adults older than 70 years and 75% of residents living in nursing home settings, is associated with reduced cognitive, social, emotional, and physical functioning. When hearing loss is detected, amplification by hearing aids or assistive listening devices can improve quality of life and functional status.[7]

To identify hearing loss, simple questions (e.g., Would you say you have any difficulty hearing? Do you feel you have hearing loss?) can be asked, or the 10-item self-reported Hearing Handicap Inventory for the Elderly instrument can be administered. If either of these screens is positive or if the clinician suspects hearing loss, an objective test is the whisper voice test, which involves whispering three different random words in each ear at distances of 6, 12, and 24 inches from the patient's ear and then asking the patient to repeat the words. Patients who fail the test—that is, are unable to repeat half the whispered words correctly—should be referred to an audiologist for further evaluation.

A more accurate screening tool for hearing impairment is the Welch Allyn AudioScope. This handheld otoscope has a built-in audiometer that can be set at different levels of intensity. A pretone at 60 dB is delivered to the patient, and then four tones of 500, 1000, 2000, and 4000 Hz at 40 dB are presented. The inability to hear either the 1000- or 2000-Hz frequency in both ears or both the 1000- and 2000-Hz frequencies in one ear indicates a positive screen and identifies the need for formal audiometric testing.

Cognitive Assessment

The incidence of dementia (Chapters 27 and 402) increases with age, especially among those older than 85 years. Early detection of memory problems can lead to the identification of treatable conditions that contribute to cognitive impairment and the development of a proactive management plan with the patient's full participation.[8] Because the diagnosis of dementia has not been documented for as many as 80% of patients who meet diagnostic criteria for it, clinicians should routinely evaluate elderly patients using any of a variety of brief screening instruments (Chapter 27).[9]

Another component of cognitive assessment is decision-making capacity. In cognitively intact older persons, capacity is assumed. However, among those with cognitive impairment, decision making must be determined before many treatments are initiated. As a rule, capacity is specific to the decision the patient is being asked to make. Ask the following questions to determine a patient's decision-making capacity:

- Can the patient make and express personal preferences at all?
- Can the patient comprehend the risks and benefits?
- Does the patient comprehend the implications?
- Can the patient give reasons for the alternative selected?
- Are supporting reasons rational?

If the answer to all these questions is yes, the patient is competent to make the decision at hand. If not, a surrogate (the durable power of attorney for health care, if the patient has identified one) should make the decision. If a durable power of attorney for health care has not been identified, the decision can be made by a family member, friend, or caregiver who knows the patient well. The order of surrogates is determined by state law. If no one is available, the health care provider should make the decision on the basis of the patient's known value system or what the health care provider believes to be in the best interest of the patient.

Mood and Affect

Although major depression (Chapter 397) is no more common among the elderly than among the younger population, minor depression and other affective disorders are common and cause considerable morbidity. Moreover, the clinical manifestations of depression may be atypical, and depression may be masked in patients with cognitive impairment or other neurologic disorders such as Parkinson disease (Chapter 409).

A two-item version of the Patient Health Questionnaire can effectively screen for depression symptoms.[10] The screener asks the patient: Over the past 2 weeks, how often have you been bothered by any of the following problems?

- Little interest or pleasure in doing things
- Feeling down, depressed, or hopeless

Responses are scored as follows: 0, not at all; 1, several days; 2, more than half the days; 3, nearly every day. Persons who score a total of 3 points or higher on the two-item screen have a 75% probability of having a depressive disorder and should be evaluated in more detail (see Table 27-3 and Chapter 397).

Falls, Mobility, and Balance

Approximately one third of community-dwelling persons older than 65 years and half of those older than 80 years fall each year. Ten percent of these falls result in a serious injury. Patients who have fallen or have gait or balance problems are at higher risk of another fall. Performing a falls assessment (measuring orthostatic blood pressure; assessing vision; reviewing medications; and testing balance, gait, and lower extremity strength) and treating risk factors for falling can reduce falls by 30 to 40%.

All older patients should be asked at least annually if they have fallen,[11] and frail older persons should be asked about falls at every visit. In addition, asking about fear of falling can identify patients at risk of future falls.

Patients who have fallen or have a fear of falling should have their balance and gait assessed by direct observation of their ability to perform specific tasks. Tests of balance include the ability to maintain a side-by-side, semitandem, and full-tandem stance for 10 seconds; resistance to a nudge; and stability during a 360-degree turn. Quadriceps strength can be assessed by observing an older person rising from a hard armless chair without using his or her hands.

In addition, direct qualitative and quantitative observation of gait to determine stability is a quick and important component of assessment.[12] Qualitative aspects include evaluation of hesitancy; sway; step length, height, symmetry, and continuity; and path deviation. Gait speed is also a helpful marker for recurrent falls. Patients who take more than 13 seconds to walk 10 meters (0.8 meter per second) are more likely to have recurrent falls and to have a shorter life expectancy.[13]

The timed up-and-go test combines some features of strength and gait. It is a timed test of the patient's ability to rise from a standard armchair, walk 3

meters (10 feet), turn, walk back, and sit down again. Patients who take longer than 20 seconds to complete the test should receive further evaluation.

Medication Review

Older persons often see several different health care providers who prescribe multiple medications that increase the risk for drug-drug interactions and adverse drug events. At a minimum, the clinician should review the patient's updated and accurate medication list at each visit. A good method of detecting potential problems is to have patients bring in all their medications (prescription and nonprescription) in their bottles. Entering a patient's medication list into commercially available computer drug interaction programs can help prevent adverse events. Many electronic health records also alert clinicians to potential drug interactions.

Nutrition

Malnutrition in older adults includes obesity (Chapter 220), undernutrition (Chapter 214), and specific vitamin deficiencies (Chapter 218). Obesity in older adults is defined as a body mass index (BMI) of 30 kg/m^2 or greater. High BMI is associated with poorer function and more comorbidities, such as type 2 diabetes mellitus (Chapter 229), osteoarthritis (Chapter 262), hyperlipidemia (Chapter 206), coronary artery disease (Chapter 52), and sleep apnea (Chapter 100).

At the initial visit, patients should be weighed and asked about weight loss in the previous 12 months. They should be weighed at all follow-up visits. BMI should be calculated on the initial visit and periodically thereafter (e.g., yearly or when a change in weight suggests the need to recalculate). Serum markers, including serum albumin, prealbumin, and cholesterol levels, are nonspecific indicators of nutritional status and can be affected by inflammatory states, physiologic stress, and trauma.

Weight loss of 4% or more during 12 months predicts increased mortality and should prompt an evaluation of medical (e.g., malignant disease, gastrointestinal disorders, hyperthyroidism, diabetes), psychiatric (e.g., depression, dementia), dental, and social or functional (e.g., poverty, inability to shop or to prepare meals) causes. In community-based older adults, a serum 25-hydroxyvitamin D level below 50 nmol/L (20 ng/mL) is associated with an increased risk for relevant clinical disease events.[14]

Urinary Incontinence

Approximately one third of community-dwelling older adult women have some degree of urinary incontinence (Chapter 26), and 75% of older men have abnormal urinary tract symptoms. Complications of incontinence include skin irritation, pressure ulcers, urinary tract infections, sleep disruption, and falls. Identification of urinary incontinence is important because effective behavioral and pharmacologic treatments are available.

A simple screen for urinary incontinence asks the following question: Have you had urinary incontinence (do you "lose" your urine) to the extent that it is bothersome and you would like to know how it can be treated? This may be valuable in determining which patients want further evaluation and therapy. Patients who answer yes should be asked how much urine is leaked, how much it interferes with daily life, and when the leakage occurs. Further evaluation and treatment depend on whether the incontinence is overflow incontinence, urge incontinence, or stress incontinence (see Table 26-3).

PREVENTIVE SERVICES

Preventive services include lifestyle advice, screening tests to detect asymptomatic disease, and vaccinations. Recommended adult immunization schedules for older adults include annual influenza vaccination, one-time pneumonia vaccination, one-time pneumococcal vaccination, one-time herpes zoster vaccination (even if the patient reports a past episode of herpes zoster), and tetanus toxoid vaccination every 10 years after receiving one dose of tetanus, diphtheria, and pertussis vaccination (Tdap). (Chapter 18).

To help the clinician decide what is appropriate for a specific patient, the U.S. Preventive Services Task Force has created an interactive website (*http://epss.ahrq.gov/ePSS/search.jsp*) with recommendations based on the patient's age, gender, tobacco use, and current sexual activity (Chapter 14). For younger elderly persons, recommendations include screening for blood pressure, diabetes (if blood pressure >135/80 mm Hg), hyperlipidemia, obesity, alcohol misuse, colorectal cancer (to age 75 years), vitamin D supplementation, and osteoporosis as well as breast cancer in women (to age 75 years), with appropriate counseling and treatment if these disorders are detected. Exercise or physical therapy is recommended to prevent falls in patients who are at increased risk. Aspirin may be recommended to prevent cardiovascular disease in this population. Other recommendations depend on patient-specific risk factors.

With increasing age, fewer preventive services are recommended because of limited life expectancy and greater comorbidities. Accordingly, some preventive measures (e.g., aspirin to prevent cardiovascular disease) and cancer screenings are not recommended in the very elderly. When the evidence base for preventive services is sparse, decisions should be individualized, on the basis of the patient's personal values, goals, and preferences.

NONMEDICAL ASSESSMENTS

Environmental Assessment

Assessing a patient's environment includes evaluating three components: the person's ability to access community services (e.g., getting to the bank and stores if the patient cannot drive), the safety of the physical environment, and the appropriateness of the living situation for the person's functional ability and cognitive status. A brief screen for safety of the physical environment can be accomplished by using a checklist that is in the public domain (*http://www.cdc.gov/ncipc/pub-res/toolkit/Falls_ToolKit/DesktopPDF/English/booklet_Eng_desktop.pdf*). For home-bound older patients, a home safety evaluation by a home health agency is more appropriate.

Social Support Assessment

When older persons become frail, the adequacy of their social support network may be the determining factor in whether they can remain at home or need institutionalization. A brief screen of social support includes taking a social history, including asking who would be available to help if the patient becomes ill. For patients with functional impairment, the clinician should ascertain who can help the patient perform ADLs or IADLs. Early identification of problems with social support can help avoid crisis situations if the patient has a sudden medical or functional decline. Caregivers should be screened periodically for symptoms of depression or caregiver burnout and referred for counseling or support groups if necessary.

Financial Assessment

Although clinicians do not have the training or expertise to explore financial resources in detail, knowing the patient's insurance status may be helpful. For example, patients with Medicaid coverage may qualify for additional medical or social support benefits. Some may be eligible for other state or local benefits, depending on their income. Others may have long-term care insurance or veterans' benefits that can help pay for caregivers, obviating the need for institutionalization.

A STRATEGIC APPROACH TO GERIATRIC ASSESSMENT FOR THE PRACTICING CLINICIAN

Although assembling an interdisciplinary assessment team is beyond the capability of most practitioners, even small group practices can use teamwork and simple practice design to perform geriatric assessments efficiently and comprehensively. These assessments can lead to local implementation of or referral to comprehensive care models that can improve the outcomes of elderly individuals with a variety of chronic conditions.[■■] Because of time constraints, screening is increasingly delegated to staff and to patients and their families by standing orders, forms, and questionnaires (Table 24-2). For example, previsit questionnaires can be used to gather information about past medical and surgical history; medications and allergies; social history, including available social support resources, preventive services, ability to perform functional tasks, and need for assistance; home safety; and advance directives. In addition, the previsit questionnaire can include specific questions that assess vision, hearing, falls, urinary incontinence, and depressive symptoms. A reasonable approach is to assess these issues annually beginning at age 75 years. Persons who are younger than 75 years but have multiple comorbidities should also be screened and reassessed annually. In addition, some elements of geriatrics assessment (assessing ADLs and IADLs; gait, balance, and falls; mood and affect; and cognition) should be performed after major illnesses, especially for those requiring hospitalization.

TABLE 24-2 APPROACHES TO ASSESSMENT OF FUNCTION IN ELDERLY INDIVIDUALS

ASPECT BEING ASSESSED	*PREVISIT QUESTIONNAIRE*		*OFFICE STAFF ADMINISTERED*	
	LENGTH OF SCREEN*	INSTRUMENTS	LENGTH OF SCREEN*	INSTRUMENTS
Functional status	D	Activities of daily living Instrumental activities of daily living		
Advance directives	B	Specific question about advance directives		
MEDICAL ASSESSMENT				
Visual impairment	B	Single-item question	B	Snellen eye chart (see Table 423-1)
Hearing impairment	B	Hearing Handicap Inventory for the Elderly	B (if needed)	Whisper test Audioscope
Cognitive problems	B		D	Mini-cog (see Chapter 27)
Mind, affective problems	D	PHQ	B	
Falls, mobility, balance	B	Simple questions	B	Timed up-and-go test
Medication review	D			Inspection of medication bottles
Malnutrition	D	Single question	B	Weight
Urinary incontinence	B	Single question	B	International Consultation on Incontinence Modular Questionnaire short form if single question is positive (see Chapter 26)
Preventive services	D	Specific questions		
OTHER DIMENSIONS				
Environment	D	Home safety checklist		
Social support	B	Single question		
Financial status			B	Insurance status

*B = brief screen (e.g., <2 minutes); D = detailed evaluation (usually ≥5 minutes); PHQ = Patient Health Questionnaire (see text).

Grade A References

A1. Boult C, Green AF, Boult LB, et al. Successful models of comprehensive care for older adults with chronic conditions: evidence for the Institute of Medicine's "Retooling for an Aging America" report. *J Am Geriatr Soc*. 2009;57:2328-2337.

A2. Deschodt M, Flamaing J, Haentjens P, et al. Impact of geriatric consultation teams on clinical outcome in acute hospitals: a systematic review and meta-analysis. *BMC Med*. 2013;11:48.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Formiga F, Ferrer A, Chivite D, et al. Predictors of long-term survival in nonagenarians: the Nona-Santfeliu study. *Age Ageing*. 2011;40:111-116.
2. Yourman LC, Lee SJ, Schonberg MA, et al. Prognostic indices for older adults: a systematic review. *JAMA*. 2012;307:182-192.
3. Keeler E, Guralnik JM, Tian H, et al. The impact of functional status on life expectancy in older persons. *J Gerontol A Biol Sci Med Sci*. 2010;65:727-733.
4. Reuben DB, Tinetti ME. Goal-oriented patient care—an alternative health outcomes paradigm. *N Engl J Med*. 2012;366:777-779.
5. Smith AK, Williams BA, Lo B. Discussing overall prognosis with the very elderly. *N Engl J Med*. 2011;365:2149-2151.
6. Christ SL, Zheng DD, Swenor BK, et al. Longitudinal relationships among visual acuity, daily functional status, and mortality: the Salisbury Eye Evaluation study. *JAMA Ophthalmol*. 2014; 132:1400-1406.
7. Pacala JT, Yueh B. Hearing deficits in the older patient: "I didn't notice anything." *JAMA*. 2012;307:1185-1194.
8. Moyer VA. Screening for cognitive impairment in older adults: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160:791-797.
9. Cordell CB, Borson S, Boustani M, et al. Medicare Detection of Cognitive Impairment Workgroup. Alzheimer's Association recommendations for operationalizing the detection of cognitive impairment during the Medicare Annual Wellness Visit in a primary care setting. *Alzheimers Dement*. 2013;9:141-150.
10. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire–2: validity of a two-item depression screener. *Med Care*. 2003;41:1284-1292.
11. Panel on Prevention of Falls in Older Persons, American Geriatrics Society and British Geriatrics Society. Summary of the Updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. *J Am Geriatr Soc*. 2011;59:148-157.
12. Donoghue OA, Savva GM, Cronin H, et al. Using timed up and go and usual gait speed to predict incident disability in daily activities among community-dwelling adults aged 65 and older. *Arch Phys Med Rehabil*. 2014;95:1954-1961.
13. Studenski S, Perera S, Patel K, et al. Gait speed and survival in older adults. *JAMA*. 2011;305:50-58.
14. de Boer IH, Levin G, Robinson-Cohen C, et al. Serum 25-hydroxyvitamin D concentration and risk for major clinical disease events in a community-based population of older adults: a cohort study. *Ann Intern Med*. 2012;156:627-634.

REVIEW QUESTIONS

1. A Physician Orders for Life-Sustaining Treatment (POLST) form orders must be respected in which of the following settings?

A. Hospital
B. Emergency department
C. Skilled nursing facility
D. Home (by emergency medical technicians)
E. All of the above settings

Answer: E Physician Orders for Life-Sustaining Treatment forms must be respected in all settings of care.

2. For frail older persons, the most important determining factor of whether a specific preventive service (e.g., mammography, colonoscopy, bone mineral density assessment, cholesterol lowering for primary prevention of heart disease) should be provided is which of the following?

A. Life expectancy
B. U.S. Preventive Services Task Force recommendations
C. Professional society (e.g., American College of Physicians) guidelines
D. Pay-for-performance incentives

Answer: A If a patient is unlikely to live long enough to accrue the benefits of treatment, the preventive service is not indicated. Some guidelines issued by professional societies do not consider comorbidity or life expectancy. Pay-for-performance initiatives do not consider age. The U.S. Preventive Services considers life expectancy only based on current age and only for some recommendations.

3. Which of the following statements about functional status is incorrect?

A. It is always impaired in dementia.
B. Impairment of instrumental activities (e.g., shopping, managing finances, performing housework) is more common than impairment of basic activities (e.g., bathing, dressing, transferring).
C. It can be assessed by self-report or direct observation.
D. It is has prognostic value in outpatient but not inpatient settings.

Answer: D Functional status has prognostic value in both outpatient and hospital settings. Functional impairment is required for the diagnosis of dementia and is used to differentiate dementia from mild cognitive impairment. Functional status can be assessed by self-report or objectively (e.g., by a physical or occupational therapist or with structured, performance-based instruments).

25

COMMON CLINICAL SEQUELAE OF AGING

JEREMY D. WALSTON

EPIDEMIOLOGY

Older adults make up the majority of patients actively treated in the health care system, in large part owing to their increased burden of chronic diseases but also because of their marked vulnerability to adverse health outcomes, such as functional and cognitive decline, falls, delirium (Chapter 28), and frailty. This age-related vulnerability is thought to have its basis in altered biology that results in tissue and physiologic system changes, which in turn may contribute to many of these conditions and to the chronic disease states commonly seen in older adults. Importantly, these biologic changes are likely to be heterogeneous and to occur at different chronologic ages and in different organs at different rates (Fig. 25-1). These complex age-related biologic changes represent a source of vulnerability that sets the stage for the marked increase in clinical sequelae observed in older adults.

PATHOBIOLOGY

Age-Related Cellular and Molecular Changes

The multiple biologic changes of aging affect important homeostatic functions and can lead to altered cellular function, declines in tissue resiliency, and cellular dropout. Several pathways become dysregulated with increasing age. First, autophagy, an intracellular process responsible for the recycling of damaged or redundant organelles or proteins, becomes less effective with age.[1] The result is an intracellular accumulation of dysfunctional mitochondria and proteins, which in turn can trigger cellular dysregulation through increased levels of free radicals, lower mitochondrial energy production, and programmed cell death or apoptosis. Second, senescent cell populations arise with increasing age and may lead to alterations in tissue and immune system function. For example, fibroblasts and fat cells evolve toward a phenotype whereby reproduction and cell death are less likely to take place, and survival persists in an altered, less functional state.[2] These senescent cells no longer function normally and often chronically secrete inflammatory cytokines and other bioactive molecules that alter surrounding tissues. Senescent T-cell populations also evolve with increasing age, perhaps related to early life viral infections. Although they are often normal in appearance and number, these T cells are less able to respond appropriately to immunogenic signals, thereby increasing vulnerability to infections. Third, some tissues become more sensitive to apoptosis, or programmed cell death. Although apoptosis is a normal cellular program that kills and disassembles damaged or redundant cells in all tissues, it accelerates with age and likely contributes to the vulnerability to chronic disease states, such as Parkinson disease, heart failure, and the generalized loss of cell number in many tissues. Fourth, evidence suggests that increased activity in transforming growth factor-β signaling may play an important role in the fibrotic changes that are observed in the heart, skeletal muscle, and lung tissue and that result in functional decrements with increasing age. Importantly, these aging-related molecular and cellular changes are heterogeneous and may affect individuals at different ages and in different tissues. However, the end result is increased susceptibility to chronic disease states, frailty, declines in function and cognition, and ultimately mortality.

Physiologic System Dysregulation and Its Consequences

Dysfunction in multiple physiologic stress response systems also plays a role in late-life vulnerability. Chronic diseases such as diabetes, vascular disease, chronic obstructive pulmonary disease, depression, and heart failure activate the innate immune system, the sympathetic nervous system, and the

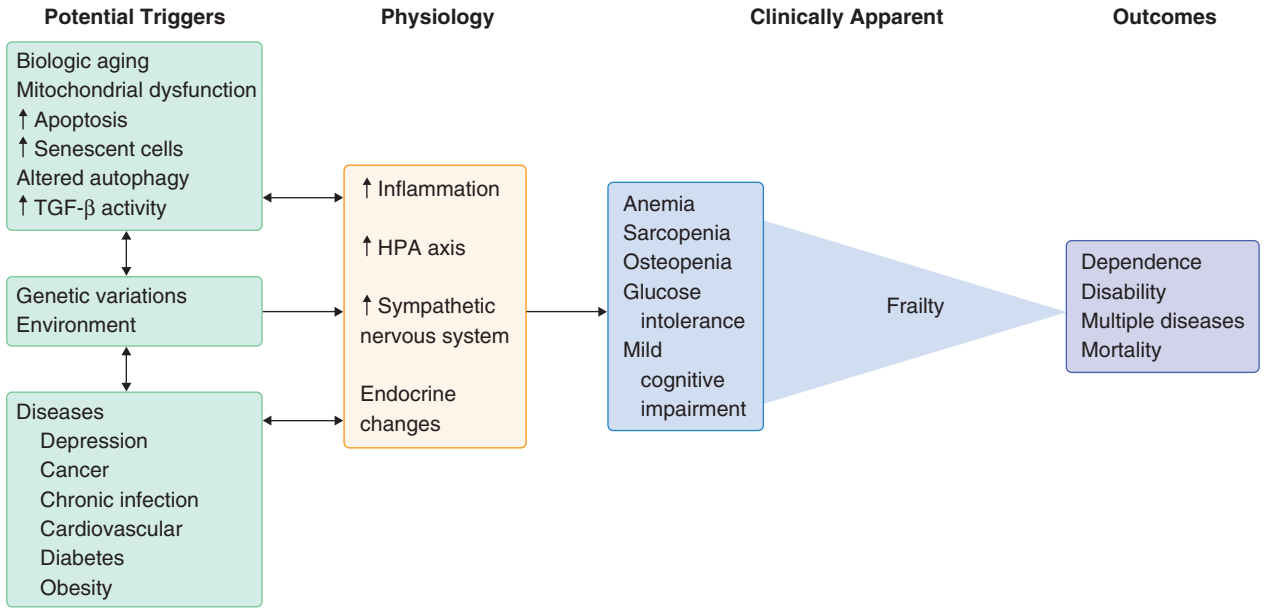

FIGURE 25-1. Comprehensive model pathway for biologic vulnerability to adverse outcomes in aging. HPA = hypothalamic-pituitary-adrenal; TGF-β = transforming growth factor-β.

hypothalamic-pituitary-adrenal axis, which in turn increase cortisol and the inflammatory cytokine interleukin-6 (IL-6).[3] These responses further exacerbate aging-related clinical conditions, such as osteoporosis and hypertension, and increase vulnerability to frailty, functional decline, accidental injury, and worsening chronic disease states[4] (E-Fig. 25-1).

CLINICAL MANIFESTATIONS

Some clinical manifestations are related not to physiologic aging but rather to cumulative exposures, such as with sun-related skin cancer (Chapter 203), or the delayed expression of genetic abnormalities, such as Huntington disease (Chapter 410) or polycystic kidney disease (Chapter 127). However, many organs and systems become less functional with age. For example, a healthy 70-year-old will have only about 50% of the lung function (Chapter 83) and renal function (Chapter 115) of a young adult. The resultant lack of physiologic reserve capacity does not affect day-to-day function but can greatly affect the ability to recover from a severe illness that exhausts the body's reserve capacity.

Despite normal baseline temperatures, older adults are more susceptible to hypothermia or hyperthermia (Chapter 109) after environmental exposure, even though they are less likely to develop fever with infections. For example, patients with pneumonia (Chapter 97) may present with confusion and dehydration rather than with fever and cough.

EFFECTS OF AGING ON SPECIFIC ORGANS AND SYSTEMS

Cardiovascular System

Between the ages of 20 and 80 years, left ventricular systolic function does not change, but the left ventricle gradually thickens. The result is that left ventricular filling in early diastole declines by 50%, and ventricular filling becomes more dependent on atrial contraction (Chapter 53). Although atherosclerosis is the most important cause of symptomatic cardiac disease in elderly people, the age-associated vascular stiffness results in an age-related increase in heart failure despite normal systolic function (Chapter 58). With the gradual loss of up to 90% of sinus node pacemaker cells by the age of 80 years, both the resting heart rate and the maximal heart rate with exercise decline. Conduction system dysfunction contributes to an increase in the prevalence of atrial fibrillation, which is seen in about 4% of community-dwelling older individuals (Chapter 64) and can develop in up to one third of the elderly after surgery (Chapter 433). Heart valves thicken and stiffen, and the prevalence of aortic stenosis and mitral annular calcifications rises (Chapter 75), often causing heart murmurs. Stiffening of the aorta causes an increase in systolic blood pressure, whereas diastolic blood pressure often stays stable or even declines (Chapter 67).

The combination of impaired ventricular filling and the inability to increase the heart rate with stress contributes to the postural hypotension that is seen in 20% of older individuals (Chapter 62) as well as their predisposition to syncope with stresses that younger individuals would tolerate. The reduced ability of the elderly to tolerate cardiovascular stress must be recognized and anticipated whenever they experience a major illness. In addition, coronary artery disease can limit cardiac reserve and increase the risk that hypotension will cause a secondary myocardial infarction.

Respiratory System

The chest wall stiffens with advancing age, and the lungs lose elastic recoil (Chapter 85). Maximal vital capacity declines by about 40%, but oxygen exchange declines by about 50% because of the additive effect of progressive ventilation-perfusion mismatching (Chapter 85). As a result, the arterial Po_2 of many 80-year-olds is about 70 to 75 mm Hg. The clinical manifestations are often progressive shortness of breath with exercise (Chapter 83) and an increased susceptibility to community-acquired pneumonia (Chapter 97) and even to aspiration pneumonia.

Gastrointestinal System

Taste and smell (Chapter 427) decline with advancing age. Food tends to taste less sweet and more bitter.

The esophageal sphincter can become lax (Chapter 138), thereby increasing reflux and even aspiration. Atrophic gastritis reduces the risk of duodenal ulcer but also the absorption of iron (Chapter 159) and vitamin B_{12} (Chapter 164). Delayed gastric emptying can lead to a sense of early satiety and decreased appetite.

A gradual decline in the number of hepatocytes decreases the weight of the liver by about one third by the age of 90 years and decreases the liver's ability to metabolize drugs (Chapter 29). Distal colonic motility from the rectosigmoid to the anal canal declines, and more than 60% of elderly individuals develop constipation (Chapter 136). Diverticula (Chapter 142) become more common with age and are seen in up to 50% of people older than age 80 years.

Urinary System

Glomerular filtration declines by about 1% per year, and kidney size declines by about one third in older adults (Chapter 115). Maximal concentrating capacity declines, and it becomes more difficult to excrete a salt load or to conserve water in the face of dehydration.

The bladder becomes more irritable with advancing age and may generate less power, which is especially a problem in men with prostatic hypertrophy (Chapter 129). By comparison, urinary incontinence (Chapter 26) is more prevalent in women. Residual bladder urine volume increases and nocturia

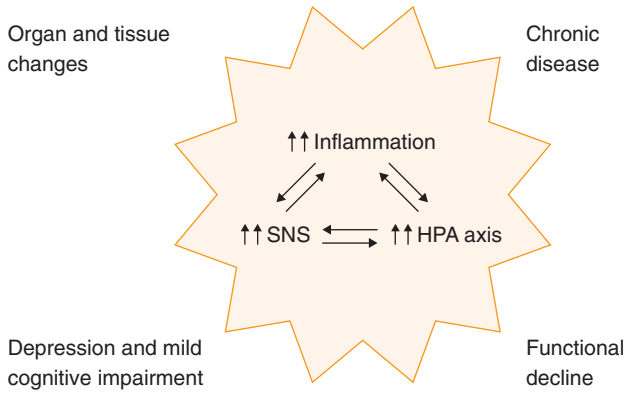

E-FIGURE 25-1. **Dysregulation of stress response systems and their influence on vulnerability to adverse outcomes. HPA = hypothalamic-pituitary-adrenal; SNS = sympathetic nervous system.**

is common. Vaginal and urethral atrophy predispose women to urinary tract infections (Chapter 284).[5]

The kidney is more susceptible to the effects of medications, particularly nonsteroidal anti-inflammatory drugs, which can result in sodium and fluid retention and subsequent hypertension. In elderly individuals, a slight acidemia results from impaired acid excretion and may contribute to the development of osteoporosis.

Endocrine System

Growth hormone levels fall with advancing age (Chapter 224), thereby resulting in decreased muscle strength, thinning of bones and skin, and increased central fat. However, growth hormone replacement does not appear to result in improved muscle strength. Levels of thyroid hormones do not decline with age (Chapter 226). Parathyroid hormone levels, however, commonly increase, especially in women, probably in response to the kidney's declining ability to maintain normal serum levels of phosphorus and calcium (Chapters 199 and 245).

The ability of the pancreas to release insulin is blunted with advancing age (Chapter 229), but the kidney's clearance of insulin also decreases. The net result is maintained plasma insulin levels in the fasting state but an increased likelihood of postprandial or stress-induced hyperglycemia. Dramatic declines in estrogen and progesterone production precipitate menopause at an average age of 51 years (Chapter 240). Testosterone levels begin to decrease in men by about 50 years of age, often with resulting declines in sexual function but not in the potency of semen (Chapter 234). Testosterone supplements reverse the muscle loss and sexual implications of declining testosterone levels but increase the risk of cardiovascular complications five-fold.

Immune System

Declines in the responsiveness of the immune system explain why the incidence of autoimmune conditions, such as systemic lupus erythematosus (Chapter 266) and multiple sclerosis (Chapter 411), declines in the elderly. However, this same decline explains increased morbidity and mortality with infectious diseases and the increased risk of reactivating infections such as tuberculosis (Chapter 324) and herpes zoster (Chapter 375). These risks emphasize the importance of vaccination against herpes zoster, influenza, pneumococcal pneumonia, and tetanus in the elderly (Chapter 18).

Hematopoietic System

Hematopoiesis (Chapter 156) is generally sustained with aging, except in response to marked stress. The one exception is that the hematocrit declines somewhat in elderly men, presumably owing to their lower testosterone levels.

Integumentary System

With aging, the epidermis and dermis adhere less tightly and the subcutaneous tissue thins, thereby making the skin feel looser and more likely to wrinkle and ulcerate. Clinical sequelae include senile purpura (Fig. 25-2) due to tears in small venules after bumps or abrasions (Chapter 440). Ultraviolet light exposure also predisposes to skin cancer (Chapter 203), rosacea (Chapter 439), xerosis, and hair loss (Chapter 442).

Wound healing is also compromised, and complete skin healing can take 5.5 weeks instead of 3.5 weeks in individuals older than 65 years. As a result, older adults are more prone to pressure sores when they are bedridden. Pressure sores, which are necrotic areas of muscle, subcutaneous fat, and skin, usually occur between underlying bone and a hard surface (or a soft surface during a prolonged time) as a result of compression and subsequent ischemia (Fig. 25-3). A continuous-pressure threshold of only 30 to 35 mm Hg is needed to cause pressure sores, and a standard mattress can generate pressures five times as high. In addition to pressure injury, other contributing factors include shear injury from rubbing constantly against underlying surfaces; burning injury from friction of the superficial skin layers; and moisture that softens the skin, makes it stick to underlying surfaces, and provides easy access for infection.

Safe positioning, regular turning, avoidance of direct pressure, pressure-reducing beds, deep foam mattresses, and air suspension beds can reduce the incidence of pressure sores. Pressure sores should be photographed to establish a baseline. The wound should be freed of any pressure to prevent additional pressure ulcers. Wet-to-dry dressings are a mainstay, and semiocclusive and occlusive dressings also can be helpful. Surgical or chemical débridement is often required. Topical or systemic antibiotics (Chapter 282) may be needed. Pressure ulcers usually heal within 6 months, but surgical repair is sometimes required.

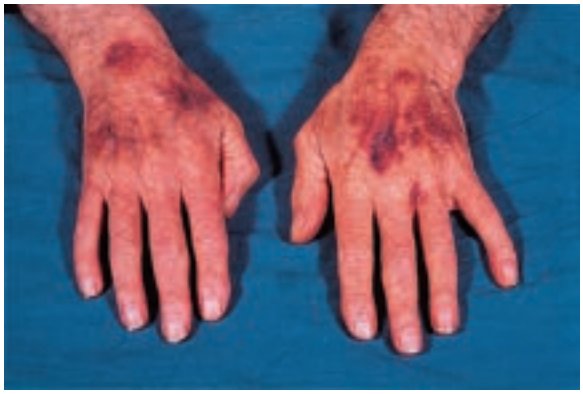

FIGURE 25-2. Senile purpura is a common and benign condition that results from impaired collagen production and capillary fragility in some older adults. In the absence of other signs of disease, no investigation is necessary. (From Forbes CD, Jackson WF. Color Atlas and Text of Clinical Medicine. 3rd ed. London: Mosby; 2003.)

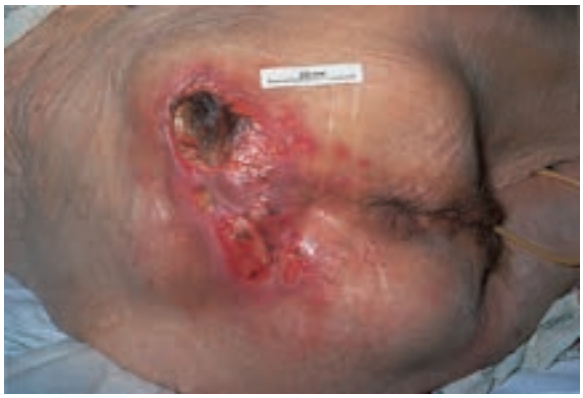

FIGURE 25-3. Severe sacral pressure sore, one of the serious but preventable complications of immobility. (From Forbes CD, Jackson WF. Color Atlas and Text of Clinical Medicine. 3rd ed. London: Mosby; 2003.)

Musculoskeletal System

Bone mass and density decrease by about 1% per year but by up to 2 to 3% per year in the first 5 to 10 years after menopause in women (Chapters 243 and 240). Tendons and ligaments become less elastic, thereby contributing to a higher incidence of rupture, especially of the Achilles tendon. Muscle mass declines by about 25% by the age of 70 years and by 30 to 40% by the age of 80 years unless it is offset by exercise.

Clinical Pharmacology

Older adults take a disproportionate share of all prescription (Chapter 29) and nonprescription medications (Chapter 39), and they are at increased risk of drug-drug interactions. Because the elderly have less muscle mass and more fat as a proportion of total body weight, they are more sensitive to the effects of water-soluble drugs and have prolonged effects from lipophilic drugs. Declines in renal and hepatic function reduce the clearance of most drugs, although drugs that are conjugated and glucuronidated are cleared relatively normally. Elderly people also are more likely to be nonadherent to prescribed regimens owing to the number of medications and their cost, mental impairment, and medication side effects.

Cancer and Cancer Screening

The incidence of most cancers rises with age, although many cancers may be less aggressive in the elderly than in younger persons. Some screening options, such as colonoscopy, are associated with their own intrinsic risks (Chapter 134), and all screening tests carry the potential side effects of false-positive results. All of these risks are greater in elderly persons, who also have fewer years of life to gain from early detection and treatment of cancer. As a result, screening recommendations should be adjusted in elderly persons. Examples include discontinuing colon cancer screening after the age of 75 years (Chapter 15), limiting screening mammography to women with at least a 10-year life expectancy,[6] advising any prostate cancer screening only to men

with a 15-year life expectancy (Chapter 15), and often discontinuing Pap smear testing in women older than 65 years if prior active screening results have been negative.

Sensory and Sleep

In addition to age-related cognitive decline (Chapters 27 and 28), hearing loss develops in about 25% of individuals older than 65 years (Chapter 428), with decreased neural transmission leading to difficulty in discriminating important sounds from background noise. Presbycusis diminishes the ability to hear high-frequency sounds.

The thickening and stiffening of the lens diminish the ability to focus on nearby objects and increase glare (Chapter 423). Transmission of light through the lens may decline by 50% or more, so elderly individuals require more ambient light. Vitreal detachment causes floaters that also can interfere with vision. Decreased tear production causes dryness of the eyes. All of these conditions contribute to a decline in visual acuity to the extent that 40% of men and 60% of women older than 65 years have a visual acuity of 20/70 or worse.

Older adults tend to have difficulty in sleeping (Chapter 405) yet spend much more time in bed. Sleep apnea also becomes more common with advancing age (Chapter 100).

Frailty as a Clinical Marker of Vulnerability in Older Adults

Frailty, which is a late-life syndrome of weakness, slowness, and weight loss, is associated with high risk of adverse health outcomes and mortality. For example, diminished heart rate variability, which is a marker of dysregulated sympathetic nervous system activity, is associated with aging, frailty, and cardiac arrhythmias. Frail older adults have significantly higher levels of salivary cortisol during the afternoon nadir period, thereby suggesting chronically increased activity of the hypothalamic-pituitary-adrenal axis. Elevated levels of inflammatory cytokines, especially IL-6, tumor necrosis factor-α receptor 1 (TNFR1), and C-reactive protein, are strongly related to functional decline, frailty, chronic disease, and mortality in older adults,[7] probably owing to increased fat, more senescent cells, and free radical production from altered mitochondria. IL-6 is likely to have a negative impact on stem cells and satellite cells, which in turn may contribute to the chronic anemia and age-related declines in skeletal muscle (sarcopenia) and bone mass (osteopenia) commonly observed in frail, older adults. TNFR1 stimulates apoptosis and necroptosis, which are cell programs that lead to cell death and possibly tissue depletion and vulnerability later in life.

In addition to stress response systems, endocrine factors that normally maintain muscle mass also play a role in frailty. For example, the adrenal androgen dehydroepiandrosterone sulfate and insulin-like growth factor 1 are significantly lower in frail adults.

Frailty serves as a clinical indicator of which older adults are at high risk for adverse outcomes, including delirium (Chapter 28), falls, and mortality. Frailty is characterized by signs and symptoms of increasing weakness, fatigue, and declines in activity. Validated frailty screening tools (Table 25-1) enable physicians to identify patients at highest risk of adverse outcomes and mortality and to develop preventive interventions that will decrease risk and improve quality of life.

One approach is to measure physiologic parameters such as grip strength, walking speed, and weight loss as well as to gather information about activity and fatigue levels. With use of this approach, the prevalence of frailty rises with increasing age; approximately 10% of community-dwelling adults older than 65 years meet these frailty criteria and subsequently are at increased risk of functional decline, falling, hospitalization, and death, even after adjustment for age, socioeconomic and smoking status, and multiple common disease states.[8] The Fried or the Cardiovascular Health Study approach has been widely used to identify elders for whom interventions may be helpful.

Frailty increases the likelihood for development of influenza or influenza-like illness in the 6 months after vaccination; the likelihood of requiring care in a skilled nursing or long-term care facility after hospitalization for general surgery; poor renal transplant graft function and early hospital readmission after transplantation; falls, hospitalization, and mortality in patients on hemodialysis for chronic renal failure; and the risk of death in aging intravenous drug users. Biologic differences between frail and non-frail older adults (see Fig. 25-1) drive the marked vulnerability to adverse outcomes observed in the frail subjects.[9] Interventions should be targeted to the specific characteristics of a patient's frailty. Increasing physical exercise is the mainstay of most programs.

TABLE 25-1 FOUR COMMONLY USED INSTRUMENTS TO MEASURE FRAILTY IN OLDER ADULTS, WITH RELEVANT MEASUREMENT DOMAINS AND SCORING CRITERIA

INSTRUMENT	DOMAINS	SCORING
Fried frailty phenotype (biologic syndrome model)*	Physical function (slowness, low activity, weakness), nutritive (weight loss), and exhaustion	Score range: 0 to 5 Frail = ≥3 criteria present Intermediate/pre-frail = 1-2 criteria present Robust = 0 criteria present
Frailty index/ accumulation of deficits (burden model)†	Diseases, ability in ADL, health attitudes/values, and symptoms/signs from the clinical and neurologic examinations	Number of deficits present and divided by the number of deficits taken into consideration Higher proportion equates to a higher level of frailty
Vulnerable Elders Survey (VES-13)‡	Physical function and ADL/IADL disability	Score range: 0 to 10 Frail = score ≥3
1994 Frailty measure (functional domains model)§	Physical function, nutritive function, cognitive function, sensory problems	Subject scoring a 3 or higher on at least one item in any domain is considered to have a problem or difficulty in that domain Frail = problems/difficulties in ≥2 domains

Details on the use and implementation of these tools can be found in the referenced articles.
ADL = activities of daily living; IADL = instrumental activities of daily living.
*Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci.* 2001;56:M146-M156.
†Rockwood K, Andrew M, Mitnitski A. A comparison of two approaches to measuring frailty in elderly people. *J Gerontol A Biol Sci Med Sci.* 2007;62:738-743.
‡Saliba D, Elliott M, Rubenstein LZ, et al. The Vulnerable Elders Survey: a tool for identifying vulnerable older people in the community. *J Am Geriatr Soc.* 2001;49:1691-1699.
§Strawbridge WJ, Shema SJ, Balfour JL, et al. Antecedents of frailty over three decades in an older cohort. *J Gerontol B Psychol Sci Soc Sci.* 1998;53:S9-S16.

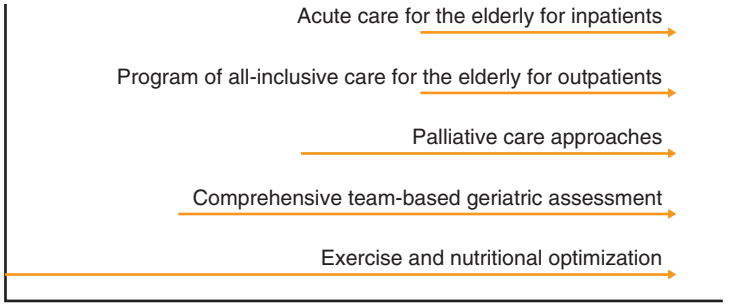

FIGURE 25-4. Assessment and treatments in frail and vulnerable older adults.

The Care of Frail Older Patients in Clinical Settings

Focused prevention of iatrogenic injuries or recurrent hospitalizations and remediation of symptoms are often warranted for frail vulnerable older adults[10] (Fig. 25-4). Physical activity, exercise interventions, and nutritional supplementation can reduce disability and symptoms and improve quality of life across the spectrum of robust to frail older adults. In addition, exercise and weight loss together improve function in obese and frail older adults. Other options include a team-based approach that engages the patient, family members, caregivers, health care providers, and social workers. Classic palliative care, with appropriate pain management, less invasive treatment plans, limited hospital visits, and organized home care plans, can greatly improve quality of life. If the patient has frequently been admitted to an inpatient setting, a program for all-inclusive care or a medical daycare setting may help prevent recurrent admissions and improve quality of life. Once admitted to the hospital, frail older adults may benefit from being congregated in a unit that specializes in their care and can provide attention to their functionality, continence, sleep disturbances, delirium, and palliative care issues.

TABLE 25-2 EVALUATION OF A PATIENT WHO HAS FALLEN

Blood tests to exclude anemia, infection, and metabolic problems such as diabetes and thyroid disease
Electrocardiography to evaluate heart disease
24-Hour electrocardiographic recording or loop monitoring to evaluate arrhythmias (Chapter 62)
Echocardiography for patients with significant heart murmurs (Chapter 55)
Drug levels to determine whether a patient is being undertreated or overtreated with a particular drug
If focal neurologic signs or symptoms are present, a computed tomography scan of the brain
If suggestive symptoms are present, radiography of the neck or spine to look for spinal stenosis

Falls

Falls are a common manifestation of frailty. Contributing factors include cardiopulmonary disease, poor eyesight, hearing loss, balance disturbances, weakness, movement disorders, neuropathies, poor judgment, depression, osteoporosis, arthritis, and foot disorders. The risk of falling also increases in patients who take more prescription medications, especially hypnotics, muscle relaxants, antihypertensive agents, diuretics, and antidepressant medications. Environmental risks include stairs, loose objects, rugs, poor lighting, poorly fitting shoes, uneven pavements, and slippery surfaces.

The approach to a patient who has fallen must include a careful history of the circumstances surrounding the fall and all conditions that could contribute to it. The physical examination should test vision, gait, balance, muscle strength, neurologic function, and pulmonary function. Further testing is recommended whenever the history and physical examination do not reveal the cause of falling or if they reveal conditions that require further evaluation (Table 25-2). In general, this evaluation is similar to the evaluation for syncope (Chapter 62).

Given the strong relationship between vitamin D deficiency and frailty,[11] clinical trials are under way with some evidence that vitamin D_3 supplementation may reduce the mortality in elderly individuals.[A6] Exercise programs, including Tai Chi, improve strength and balance as well as help prevent falls that are often associated with frailty.[A4,A8] Hip protectors to reduce fractures in nursing home patients and in ambulatory older individuals have shown mixed results, and compliance with these cumbersome devices is only 25 to 70%. Further studies of types and doses of exercise may help elucidate optimal regimens for preventing frailty and for optimizing health and well-being of frail, older adults.

Grade A References

A1. Giannoulis MG, Martin FC, Nair KS, et al. Hormone replacement therapy and physical function in healthy older men. Time to talk hormones? *Endocr Rev*. 2012;33:314-377.

A2. Basaria S, Coviello AD, Travison TG, et al. Adverse events associated with testosterone administration. *N Engl J Med*. 2010;363:109-122.

A3. Cameron ID, Fairhall N, Langron C, et al. A multifactorial interdisciplinary intervention reduces frailty in older people: randomized trial. *BMC Med*. 2013;11:65.

A4. Gine-Garriga M, Roque-Figuls M, Coll-Planas L, et al. Physical exercise interventions for improving performance-based measures of physical function in community-dwelling, frail older adults: a systematic review and meta-analysis. *Arch Phys Med Rehabil*. 2014;95:753-769.

A5. Pahor M, Guralnik JM, Ambrosius WT, et al. Effect of structured physical activity on prevention of major mobility disability in older adults: the LIFE study randomized clinical trial. *JAMA*. 2014;311:2387-2396.

A6. Bjelakovic G, Gluud LL, Nikolova D, et al. Vitamin D supplementation for prevention of mortality in adults. *Cochrane Database Syst Rev*. 2014;1:CD007470.

A7. Moyer VA. Prevention of falls in community-dwelling older adults: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2012;157:197-204.

A8. El-Khoury F, Cassou B, Charles MA, et al. The effect of fall prevention exercise programmes on fall induced injuries in community dwelling older adults: systematic review and meta-analysis of randomised controlled trials. *BMJ*. 2013;347:f6234.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Tan CC, Yu JT, Tan MS, et al. Autophagy in aging and neurodegenerative diseases: implications for pathogenesis and therapy. *Neurobiol Aging*. 2014;35:941-957.
2. Tchkonia T, Zhu Y, Van DJ, et al. Cellular senescence and the senescent secretory phenotype: therapeutic opportunities. *J Clin Invest*. 2013;123:966-972.
3. Silverman MN, Sternberg EM. Glucocorticoid regulation of inflammation and its functional correlates: from HPA axis to glucocorticoid receptor dysfunction. *Ann N Y Acad Sci*. 2012;1261:55-63.
4. Barnes JN, Hart EC, Curry TB, et al. Aging enhances autonomic support of blood pressure in women. *Hypertension*. 2014;63:303-308.
5. Mody L, Juthani-Mehta M. Urinary tract infections in older women: a clinical review. *JAMA*. 2014;311:844-854.
6. Walter LC, Schonberg MA. Screening mammography in older women: a review. *JAMA*. 2014;311:1336-1347.
7. Varadhan R, Yao W, Matteini A, et al. Simple biologically informed inflammatory index of two serum cytokines predicts 10 year all-cause mortality in older adults. *J Gerontol A Biol Sci Med Sci*. 2014;69:165-173.
8. Bandeen-Roche K, Xue QL, Ferrucci L, et al. Phenotype of frailty: characterization in the women's health and aging studies. *J Gerontol A Biol Sci Med Sci*. 2006;61:262-266.
9. Afilalo J, Alexander KP, Mack MJ, et al. Frailty assessment in the cardiovascular care of older adults. *J Am Coll Cardiol*. 2014;63:747-762.
10. Ko FC. The clinical care of frail, older adults. *Clin Geriatr Med*. 2011;27:89-100.
11. Wong YY, McCaul KA, Yeap BB, et al. Low vitamin D status is an independent predictor of increased frailty and all-cause mortality in older men: the Health in Men Study. *J Clin Endocrinol Metab*. 2013;98:3821-3828.

REVIEW QUESTIONS

1. A 74-year-old man who is in robust health and who exercises frequently asks what supplements will keep him strong and fit. Which of the options has been proved effective in maintaining quality of life and fitness in older adults?

A. Growth hormone
B. Testosterone
C. Vitamin D and calcium
D. Regular exercise
E. Coenzyme Q

Answer: D Many studies have evaluated the safety and efficacy of hormones and other treatments for maintenance of strength and function in older age groups. Although growth hormone, testosterone, and vitamin D may be helpful in those with diagnosed deficiencies, there is no evidence that giving these agents to a relatively healthy person without signs or symptoms of such deficits would be beneficial. Regular exercise is the only listed intervention that has been shown to be safe and efficacious for building strength and enhancing quality of life in almost any setting.

2. An 82-year-old patient presents with a chief complaint of "I just don't have any energy." He notes muscle wasting, fatigue, and functional decline during the past year. Extensive medical evaluation in the next month reveals no obvious medical conditions that explain this decline. What should you recommend?

A. Careful medical monitoring
B. Assessment of his risk for falling
C. Exercise and activity regimen
D. Social engagement
E. All of the above

Answer: E This obviously frail patient is at high risk for adverse outcomes, social isolation, and mortality in the coming years. He warrants careful medical monitoring, assessment of the risk for falling, and increasing social engagement. In addition, exercise has been demonstrated to be helpful in even the frailest older adults. Hence, all of the answers are appropriate in this patient.

26

INCONTINENCE

NEIL M. RESNICK

DEFINITION

Urinary incontinence is the involuntary leakage of urine sufficient to be a health or social problem.

EPIDEMIOLOGY

More than twice as common in women as in men, the prevalence of incontinence increases with age. Incontinence afflicts 15 to 30% of older adults living at home, one third of those in acute care settings, and half of those in nursing homes. It predisposes to perineal rashes, pressure ulcers, urinary tract infections, urosepsis, falls, and fractures, and it is associated with embarrassment, stigmatization, isolation, depression, anxiety, sexual dysfunction, and risk for institutionalization. Its cost in the United States exceeds $26 billion annually.

Despite these considerations, geriatric incontinence remains largely neglected, by patients and physicians alike. This is unfortunate because its increased prevalence with age relates more to age-associated diseases and functional impairments than to age itself. Most important, incontinence is usually treatable and often curable at all ages, even in frail elderly people, although the approach in older patients must be broader than that employed in younger patients.

PATHOBIOLOGY

At any age, continence depends not only on the integrity of lower urinary tract function but also on the presence of adequate mentation, mobility, motivation, and manual dexterity. Although incontinence in younger patients is rarely associated with deficits in these domains, such deficits occur commonly in older patients, in whom they can cause or exacerbate incontinence or influence therapeutic approaches.

With age, bladder capacity does not change, but bladder sensation and contractility decrease. At the cellular level, detrusor smooth muscle develops a "dense band pattern" characterized by dense sarcolemmal bands with depleted caveolae. This depletion may mediate the age-related decline in bladder contractility. In addition, an incomplete disjunction pattern characterized by scattered protrusion junctions develops and may underlie the high prevalence of involuntary bladder contractions (detrusor overactivity) in older adults of both sexes. Bladder ischemia and/or inflammation may also contribute.[1,2] Urethral length and sphincter strength decrease in women, whereas the prostate enlarges in most men and causes measurable obstruction in about half. The postvoid residual volume in the bladder also increases in both sexes but normally to less than 100 mL. In addition, elderly people often excrete most of their fluid intake at night, even in the absence of venous insufficiency, renal disease, heart failure, or prostatism. Because this shift in nocturnal fluid excretion is coupled with an age-associated increase in sleep disorders, most older adults have one or two episodes of nocturia per night.

None of these changes causes incontinence, but all predispose to it. This predisposition, combined with the increased likelihood that an older person will encounter an additional pathologic, physiologic, or pharmacologic insult, explains the increased prevalence of incontinence with age. Thus, the onset or exacerbation of incontinence in an older person is likely to be due to precipitants that are outside the lower urinary tract and that are amenable to medical intervention. Furthermore, treatment of the precipitants alone may be sufficient to restore continence, even if there is coexisting urinary tract dysfunction. For example, a flare of hip arthritis in a woman with age-related detrusor overactivity may decrease mobility sufficiently to convert her urinary urgency into incontinence. Treatment of the arthritis, rather than the involuntary detrusor contractions, will not only restore continence but also lessen pain and improve mobility. Because of their frequency, reversibility, and association with morbidity beyond incontinence, the transient precipitant causes should be addressed first.

Causes of Transient Incontinence

Incontinence is transient in up to one third of community-dwelling elderly people and in up to half of acutely hospitalized patients. Although most

TABLE 26-1 CAUSES OF TRANSIENT INCONTINENCE: DIAPERS MNEMONIC

Delirium	Result of underlying illness or medication; incontinence is secondary and abates once the cause of delirium is corrected	
Infection—*symptomatic* UTI	Acute, symptomatic UTI causes incontinence, but the far more common asymptomatic bacteriuria does not	
Atrophic urethritis/vaginitis	Characterized by vaginal erosions, telangiectasia, petechiae, and friability; may cause or contribute to incontinence. Although oral estrogen may worsen incontinence, a 3- to 12-month course of topical estrogen can be useful	
Pharmaceuticals	*Drug type:*	*Potential effects on continence:*
	Sedative-hypnotics (e.g., long-acting benzodiazepines; alcohol)	Sedation, delirium, decreased mobility
	Anticholinergics (dicyclomine, disopyramide, sedating antihistamines, antipsychotics, tricyclic antidepressants, anti-Parkinson, antidepressants) (*not* SSRIs)	Urinary retention, overflow incontinence, delirium, impaction; the antipsychotics also decrease mobility
	Opiates	Urinary retention, stool impaction, sedation, delirium
	α-Adrenergic antagonists	Relax sphincter; may induce stress incontinence in women
	α-Adrenergic agonists	Urinary retention in men (tighten sphincter, prostate)
	Calcium-channel blockers, especially the dihydropyridines	Urinary retention; nocturnal diuresis due to fluid retention
	"Loop" diuretics (thiazide-like agents only rarely cause it)	Polyuria, frequency, urgency
	NSAIDs	Nocturnal diuresis due to fluid retention
	Thiazolidinediones	Nocturnal diuresis due to fluid retention
	Some nociceptives (gabapentin, pregabalin)	Nocturnal diuresis due to fluid retention; sedation; delirium
	Dopamine receptor agonists (e.g., ropinirole, pramipexole)	Nocturnal diuresis due to fluid retention
	Angiotensin-converting enzyme inhibitors	Drug-induced cough leads to stress incontinence in women
	Vincristine	Urinary retention due to neuropathy
Excess urine output	From large intake, diuretic agents (theophylline, caffeinated beverages, alcohol), and metabolic disorders (hyperglycemia, hypercalcemia); nocturnal incontinence may result from mobilization of peripheral edema (heart failure, venous insufficiency, side effects of medications)	
Restricted mobility	Often results from overlooked, correctable conditions such as arthritis, pain, foot problems, postprandial hypotension, or fear of falling	
Stool impaction	May cause both fecal and urinary incontinence that remit with disimpaction	

NSAIDs = nonsteroidal anti-inflammatory drugs; SSRI = selective serotonin reuptake inhibitor; UTI = urinary tract infection.
Adapted from Resnick NM, Tadic SD, Yalla SV. Geriatric incontinence and voiding dysfunction. In: Wein AJ, Novick AC, Partin AW, et al, eds. *Campbell-Walsh Urology*, 10th ed. St. Louis: Elsevier; 2010.

transient causes are outside the lower urinary tract (Table 26-1), three points warrant emphasis. First, the risk for transient incontinence is increased if, in addition to physiologic changes of the lower urinary tract, there also are pathologic changes. Anticholinergic agents are more likely to cause overflow incontinence in individuals with a weak or obstructed bladder, whereas excess urine output is more likely to cause urge incontinence in people with detrusor overactivity or impaired mobility. Second, these transient causes may persist if left untreated and should not be dismissed merely because incontinence is long-standing. Third, identification of the most common cause is of little value because causes vary among individuals, and geriatric incontinence is rarely due to just one cause.

Causes of Established Incontinence Related to the Lower Urinary Tract

Detrusor overactivity, also called involuntary bladder contraction or *overactive bladder*, generally causes *urge incontinence* and is the most common type of lower urinary tract dysfunction in incontinent elderly people, in whom it accounts for about two thirds of cases. Histologically, detrusor overactivity is associated with the complete disjunction pattern, with widening of the intercellular space, reduction of normal (intermediate) muscle cell junctions, and emergence of novel protrusion junctions and ultraclose abutments that connect cells together in chains. These connections may mediate a change in cell coupling from a mechanical to an electrical mechanism that results in involuntary bladder contraction. Other potential causes include ischemia, abnormalities in suburothelial myofibroblasts, and changes in central nervous system structural and functional control mechanisms.

At any age, detrusor overactivity is usually idiopathic, but it can be associated with a variety of other causes that may affect prognosis and management. Such conditions include an upper motor neuron lesion (Chapters 400 and 419), urethral obstruction, stress incontinence, bladder calculus, and bladder carcinoma (Chapter 197).

Detrusor overactivity exists as two subsets in elderly people: one in which contractile function is preserved and one in which it is impaired. The latter condition, termed *detrusor hyperactivity with impaired contractility*, has several implications. First, because the bladder is weak, these patients commonly develop urinary retention, which is also seen in patients with outlet obstruction and detrusor underactivity. Second, even in the absence of retention, detrusor hyperactivity with impaired contractility mimics other lower urinary tract causes of incontinence. For instance, if the involuntary detrusor contraction occurs coincident with a stress maneuver and if the weak contraction is not detected, detrusor hyperactivity with impaired contractility will be misdiagnosed as stress incontinence. Alternatively, because detrusor hyperactivity with impaired contractility may be associated with urinary urgency, frequency, weak flow rate, elevated residual urine, and bladder trabeculation, in men it may mimic prostatic obstruction. Third, anticholinergic therapy of detrusor hyperactivity with impaired contractility may result in urinary retention owing to bladder weakness, thereby requiring alternative therapeutic approaches.

Stress incontinence, which is the second most common cause of incontinence in older women and the dominant cause in middle-aged women, usually reflects urethral hypermobility plus some degree of sphincter weakness. Stress incontinence is rare in men but can result from sphincter damage following radical but not transurethral prostatectomy.

Urethral obstruction is the second most common cause of established incontinence in older men, although most obstructed men are not incontinent. When obstruction is associated with incontinence, it usually presents as urge incontinence owing to the associated detrusor overactivity; overflow incontinence is uncommon. Outlet obstruction is rare in women but may result from a bladder neck suspension or from urethral kinking associated with a large cystocele.

Detrusor underactivity is usually idiopathic. When it causes incontinence, it is associated with overflow incontinence (<10% of incontinence).

Damage to lower urinary tract innervation can cause several types of dysfunction. A brain lesion may cause detrusor overactivity. A spinal cord lesion (Chapters 189 and 400) above the sacral level can cause both detrusor overactivity and detrusor-sphincter dyssynergia, a condition in which the sphincter contracts rather than relaxes during detrusor contraction; the result can be severe outlet obstruction and hydronephrosis. A spinal cord lesion below the sacral level can cause detrusor underactivity, sphincter weakness, or both. Peripheral and autonomic nerve damage can cause still additional problems. Because *neurogenic bladder* is such a nonspecific term, it is preferable to refer to the specific dysfunction that it causes.

Causes of Incontinence Unrelated to the Lower Urinary Tract (Functional Incontinence)

"Functional" incontinence, which is often cited as a distinct type of geriatric incontinence and attributed to deficits of cognition and mobility, implies that urinary tract function is normal. However, normal urinary tract function is the exception, even in continent elderly people, and is rarely observed in incontinent elderly people. Moreover, incontinence is not inevitable, even with dementia or immobility. Among the most severely demented

institutionalized residents, nearly 20% are continent; among those who can transfer from a bed to a chair, nearly half are continent. Functionally impaired individuals also are the most likely to suffer from factors that cause transient incontinence, and a diagnosis of functional incontinence may result in failure to detect these reversible causes. Finally, if functionally impaired individuals also have urethral obstruction or stress incontinence, they may benefit from targeted therapy. Nonetheless, functional impairment often contributes to incontinence, and addressing its causes and those of transient incontinence may ameliorate incontinence sufficiently to obviate the need for further investigation.

CLINICAL MANIFESTATIONS

The manifestations of transient incontinence depend on the underlying condition. For established incontinence, detrusor overactivity usually manifests as *urge incontinence,* characterized by leakage that follows the *abrupt* onset or intensification of a desire to void, leakage of a moderate to large amount, urinary frequency (>8 voids/day), nocturia, and nocturnal incontinence. However, some patients with detrusor overactivity may present without the urge component. *Stress incontinence* causes leakage that coincides *instantaneously* with both the onset and cessation of a cough or other cause of increased abdominal pressure; nocturnal leakage is rare. Some patients report both types of incontinence, or *mixed incontinence*, but it is useful to determine which component is the most bothersome. In men with sphincter damage following radical prostatectomy, leakage resembles the intermittent drip of a leaky faucet. Occasionally, patients present with incontinence that is more difficult to characterize clinically without further testing.

DIAGNOSIS

In addition to a targeted clinical evaluation (Table 26-2), a bladder diary can provide diagnostic clues and guide therapy (Fig. 26-1). For example, incontinence occurring only between 8 AM and noon may be caused by a morning loop diuretic. Incontinence that occurs at night in a demented man with heart failure, but not during a 4-hour nap in his wheelchair, is likely due to nocturnal diuresis associated with his heart failure and not to dementia, impaired mobility, or prostatic obstruction. A woman with volume-dependent stress incontinence may leak only on the way to void after a full night's sleep, when

TABLE 26-2 CLINICAL EVALUATION OF THE INCONTINENT PATIENT

HISTORY
Type (urge, stress, overflow, or mixed)
Incontinence frequency, severity, duration
Pattern (diurnal, nocturnal, or both; also, e.g., after taking medications)
Associated symptoms (straining to void, incomplete emptying, dysuria, hematuria, suprapubic/perineal discomfort)
Alteration in bowel habit/sexual function (because of proximity to the bladder and shared innervation)
Other relevant factors (cancer, acute illness, neurologic disease, pelvic or lower urinary tract surgery/radiation therapy)
Medications, including nonprescription agents (see Table 26-1)
Brief assessment of cognitive and physical function
PHYSICAL EXAMINATION
Identify other relevant medical conditions (e.g., congestive heart failure, peripheral edema)
If stress incontinence suspected, determine whether leakage *coincides* with the onset *and* cessation of a single, forceful cough
Palpate for bladder distention after voiding
Pelvic examination to detect atrophic vaginitis, pelvic muscle laxity, pelvic mass
Rectal examination (skin irritation, resting tone and voluntary control of anal sphincter, prostate nodule; fecal impaction (*note:* prostate size correlates poorly with presence of urethral obstruction)
Neurologic examination (mental status and elemental examination, including sacral reflexes and perineal sensation)
INITIAL INVESTIGATION
Bladder diary (see Fig. 26-1)
Metabolic survey (electrolytes, calcium, glucose, and urea nitrogen as appropriate)
Measure postvoid residual volume, by portable ultrasound if available
Urinalysis to detect sterile hematuria or infection; culture if new-onset or worsening incontinence
Renal ultrasound to detect hydronephrosis in men whose postvoid residual volume urine exceeds about 200 mL
Urine cytology for patients with hematuria, pain, or unexplained new-onset or worsening incontinence
Uroflowmetry for men in whom urethral obstruction is suspected
Cystoscopy for patients with hematuria, suspicion of lower urinary tract pathology (e.g., bladder fistula, stone, or tumor; urethral diverticulum), or need for lower urinary tract surgery

Adapted from Resnick NM, Yalla SV. Management of urinary incontinence in the elderly. *N Engl J Med.* 1985;313:800-805.

Date	Time	Volume Voided (mL)	Are You Wet or Dry?	Approximate Volume of Incontinence	Comments
4/5	3:40 pm	240	Wet	Slight	
	6:05 pm	210	Dry		
	8:15 pm	150	Dry		Running water
	10:20 pm	150	Wet	15 mL	Bowel movement
	10:30 pm	30	Dry		
4/6	3:15 am	270	Wet	Slight	
	6:05 am	300	Wet	Slight	
	7:40 am	200	Dry		
	9:50 am	?	Dry		
	11:20 am	200	Dry		
	12:50 pm	180	Dry		
	1:40 pm	240	Dry		
	3:35 pm	160	Dry		
	6:00 pm	170	Dry		
	8:20 pm	215	Wet	Slight	Running water
	10:25 pm	130	Dry		

FIGURE 26-1. Sample bladder diary. Bladder diary of an incontinent 75-year-old man. Urodynamic evaluation excluded urethral obstruction and confirmed a diagnosis of detrusor hyperactivity with impaired contractility (detrusor hyperactivity with impaired contractility). Note the 24-hour urine output of nearly 3 liters due to the belief that drinking 10 glasses of fluid per day was "good for my health." (He did not mention this until queried about the voiding record.) Given the typical voided volume of 150 to 250 mL and a measured postvoid residual of 150 mL, excess fluid intake was overwhelming his usual bladder capacity of 400 mL (150+250 mL). Although involuntary bladder contractions were present, the easily reversible volume component of the problem, combined with the risk for precipitating urinary retention with an anticholinergic agent, prompted treatment with volume restriction alone. After daily urinary output dropped to 1500 mL, frequency abated, and incontinence resolved. (Adapted from DuBeau CE, Resnick NM. Evaluation of the causes and severity of geriatric incontinence: a critical appraisal. *Urol Clin North Am*. 1991;18:243-256.)

her bladder contains more than 400 mL—more than it ever does during her continent waking hours.

Because urinary retention is difficult to detect by examination and can affect diagnosis and therapy, the postvoid residual volume should be determined routinely, except possibly in middle-aged women with a classic presentation of stress incontinence. For example, in a randomized trial of women with uncomplicated, demonstrable stress urinary incontinence and a postvoid residual volume of less than 150 mL, preoperative office evaluation alone provided similar 1-year outcomes as did evaluation with urodynamic testing.■ Urodynamic testing is generally recommended only when diagnostic certainty is required, such as before most surgical repairs in older patients, or if there is evidence of a serious underlying cause of the incontinence, such as a brain or spinal cord lesion, carcinoma of the bladder or prostate, hydronephrosis, or bladder calculus.[3] Urodynamic evaluation comprises a battery of tests designed to assess the lower urinary tract during the filling and voiding phases of micturition. The selection among tests depends on the clinical setting and question to be answered; for instance, measuring detrusor pressure and urine flow during voiding can determine whether urethral obstruction is present, whereas monitoring bladder and urethral pressures during the filling phase and with coughing may be helpful for patients with an atypical presentation of mixed incontinence.

TREATMENT

Optimal therapy requires a multifactorial approach (Table 26-3), including treatment of transient causes, underlying medical conditions, functional impairments, and the urinary tract abnormality itself. Although pads and diapers have a role, they remain an adjunct to more specific therapy.

Behavioral Therapy

Behavioral therapy includes education, self-monitoring with a bladder diary, adjustment of the intake of fluid and caffeine,[3,4] weight loss for overweight women with stress incontinence,■ use of aids (e.g., a bedside urinal), and various types of bladder retraining and urethral sphincter exercises (e.g., progressively increasing voiding intervals, strategies to cope with urgency, and pelvic muscle exercises).[4,5]■ The efficacy of behavioral therapy is equivalent to pharmacotherapy for urge incontinence. For stress incontinence, the efficacy of behavioral therapy is superior to drugs but inferior to surgery.[5]■■ Moreover, combining behavioral and pharmacologic therapy may prove more beneficial than either treatment alone, especially for urge incontinence,■■ because neither therapy generally abolishes involuntary bladder contractions.[6] For institutionalized patients who are cognitively impaired but can state their name and are partly mobile, regular daytime reminders to void ("prompted voiding")

TABLE 26-3 STEPWISE APPROACH TO TREATMENT OF URINARY INCONTINENCE*

CONDITION	CLINICAL TYPE OF INCONTINENCE†	TREATMENT
Detrusor overactivity with normal contractility	Urge	1. Bladder retraining or prompted voiding regimens 2. ± Bladder relaxant medication if needed and not contraindicated (see drug list below). If treatment fails, consider posterior tibial neurostimulation, sacral neuromodulation, or intradetrusor injection of onabotulinumtoxinA 3. Indwelling catheterization alone is often unhelpful because detrusor spasms often increase, leading to leakage around the catheter 4. In selected cases, induce urinary retention pharmacologically and add intermittent or indwelling catheterization‡
Detrusor hyperactivity with impaired contractility	Urge§	1. If bladder empties adequately, behavioral methods (as above) ± bladder relaxant medication (low doses; especially feasible if sphincter incompetence coexists) 2. If residual urine >150 mL, augmented voiding techniques¶ or intermittent catheterization (± bladder relaxant medication). If neither feasible, undergarment or indwelling catheter‡ 3. In selected cases, induce urinary retention pharmacologically and add intermittent or indwelling catheterization‡
Stress incontinence	Stress	1. Conservative methods (weight loss if obese; treatment of cough or atrophic vaginitis; physical maneuvers to prevent leakage [e.g., tighten pelvic muscles before cough, cross legs]; occasionally, use of tampon or pessary is helpful) 2. If leakage threshold ≥150 mL identified, adjust fluid excretion and voiding intervals appropriately 3. Pelvic muscle exercises ± biofeedback/weighted intravaginal cones; must continue indefinitely 4. Surgery (sling, artificial sphincter, periurethral bulking injections)
Urethral obstruction	Urge/overflow‖	1. Conservative methods (including adjustment of fluid excretion, bladder retraining/prompted voiding) if hydronephrosis, recurrent symptomatic UTI, and hematuria have been excluded 2. α-Adrenergic antagonist 3. Also consider adding a bladder relaxant if detrusor overactivity coexists, postvoid residual volume is small, and surgery is not desired/feasible; *monitor postvoid residual volume!* 4. Finasteride, if not contraindicated and the patient either prefers it or is not a surgical candidate 5. Surgery (incision, prostatectomy) is an effective alternative before or after these steps
Underactive detrusor	Overflow	1. Decompress for at least several days (the larger the postvoid residual volume, the longer should be the decompression [up to a month]) and then perform a voiding trial 2. Exclude urethral obstruction if this has not already been done 3. If cannot void or if postvoid residual volume remains large, try augmented voiding techniques¶ ± α-adrenergic antagonist, but only if some voiding possible; bethanechol *rarely* useful 4. If fails, or voiding is not possible, intermittent or indwelling catheterization‡

BLADDER RELAXANT AGENTS FOR URGE INCONTINENCE

- Anticholinergic
  - Oxybutynin IR, 7.5-20 mg daily (2.5-5 mg tid-qid); oxybutynin XL, 5-30 mg once daily; oxybutynin patch (3.9 mg/day) twice weekly; oxybutynin 10% gel (1 g topically once per day)
  - Tolterodine, 1-2 mg twice daily; tolterodine LA, 4 mg once daily
  - Darifenacin, 7.5-15 mg once daily
  - Solifenacin, 5-10 mg once daily
  - Trospium, 20 mg daily to twice daily; 60 mg (extended release) once daily
  - Fesoterodine, 4-8 mg once daily
- β_3-Adrenergic agonist
  - Mirabegron, 25-50 mg once daily

*These treatments should be initiated only after adequate toilet access has been ensured, contributing conditions have been treated (e.g., atrophic vaginitis, UTI, fecal impaction, heart failure), fluid management has been optimized, and unnecessary or exacerbating medications have been addressed. For additional details, see text.
†*Urge:* leakage in the absence of stress maneuvers and urinary retention, usually preceded by *abrupt* onset or intensification of the need to void; *stress:* leakage that coincides *instantaneously* with stress maneuvers, in the absence of urinary retention or detrusor contraction; *overflow:* frequent leakage of small amounts associated with urinary retention.
‡UTI prophylaxis can be used for recurrent symptomatic UTIs, but only if catheter is not indwelling.
§May also mimic stress or overflow incontinence.
‖Also can cause postvoid "dribbling" alone, which is treated conservatively (e.g., by sitting to void and allowing more time, "double voiding," and in men by gently "milking" the urethra after voiding).
¶Augmented voiding techniques include Credé (application of suprapubic pressure) and Valsalva (straining) maneuvers, and double voiding. They should be performed only *after* voiding has begun.
UTI = urinary tract infection.
Adapted and updated in 2014 from Resnick NM. Voiding dysfunction and urinary incontinence. In: Beck JC, ed. *Geriatric Review Syllabus.* New York: American Geriatrics Society; 1991:141-154.

have proved effective for daytime incontinence; pads and diapers are appropriate for the others.[7]

Pharmacotherapy

Currently approved drugs have not proved effective for stress incontinence and overflow incontinence. For urge incontinence, however, several bladder relaxants have proved modestly and equally effective (see Table 26-3), even in trials that targeted older patients.[5,A1,A2,A3] All of the anticholinergics have antimuscarinic properties, such as dry mouth, constipation, visual blurring, and occasional confusion. Yet each is well tolerated even in cognitively impaired elderly patients when prescribed properly, although cognitive status should be monitored.[A4] These drugs also can be well tolerated in patients taking cholinesterase inhibitors. The choice among these drugs often hinges on other considerations. For instance, immediate-release oxybutynin has the quickest onset of action, making it an inexpensive and effective choice for patients who need excellent control at predictable times. The other drugs, although more expensive, can be used less often and can be better tolerated for daily use. Mirabegron, a β_3-adrenergic agonist, is a newer drug with an efficacy similar to the anticholinergic agents.[8]

Regardless of the drug selected, the key is to begin with a low dose and increase it slowly, realizing that the full benefit is generally not apparent for about 2 months and that side effects may offset the benefit. With such titration, urge incontinence can be controlled in about one third of patients and substantially improved in another one third.

Surgical Procedures

Surgery for stress incontinence has proved effective for women of all ages, including elderly women, and is relatively durable.[A1] Periurethral bulking injections can help frail women or those with mild stress incontinence, but it does not generally restore continence. However, urethral sling and mid-urethral tape suspension procedures can cure most women for at least 5 years. For women with more complex stress incontinence and for men with stress incontinence more than 1 year after radical prostatectomy, an artificial sphincter has proved effective and relatively durable. Experience with the "male sling" is still limited.

Surgical interventions for urge incontinence, including neuromodulation,[A9] tibial nerve stimulation,[A10] and injections of onabotulinum toxin,[A11,A12] are third-line agents. However, these procedures have not yet been studied adequately in elderly people. In addition, the limited available data suggest that their efficacy may be only modestly better than that of pharmacotherapy, and that older patients may not fare as well as younger ones.

PREVENTION

There are scant data regarding the prevention of incontinence, but one randomized trial of an educational and behavioral modification program for women older than 55 years found that it reduced the risk for incontinence for 1 year.[A13] A secondary analysis of the Diabetes Prevention Program found that, at the end of 3 years, an intensive lifestyle intervention was associated with reduced risk for self-reported incontinence, with most of the benefit explained by weight loss and a reduced risk for stress incontinence.[A14]

PROGNOSIS

Limited data suggest that incontinence progresses in about one third of patients and remits in about 10 to 15%, although it is unclear how much of the remission reflects intervention or improvement in functional or medical status.

Grade A References

A1. Nager CW, Brubaker L, Litman HJ, et al. A randomized trial of urodynamic testing before stress-incontinence surgery. *N Engl J Med*. 2012;366:1987-1997.

A2. Subak LL, Wing R, West DS, et al. Weight loss to treat urinary incontinence in overweight and obese women. *N Engl J Med*. 2009;360:481-490.

A3. Shamliyan T, Wyman J, Kane RL. AHRQ Comparative Effectiveness Reviews. Nonsurgical treatments for urinary incontinence in adult women: diagnosis and comparative effectiveness. Rockville, MD: Agency for Healthcare Research and Quality; 2012;11(12):EHC074-EF.

A4. Labrie J, Berghmans BL, Fischer K, et al. Surgery versus physiotherapy for stress urinary incontinence. *N Engl J Med*. 2013;369:1124-1133.

A5. Rai BP, Cody JD, Alhasso A, et al. Anticholinergic drugs versus non-drug active therapies for non-neurogenic overactive bladder syndrome in adults. Cochrane Database System Review publications. 2012;12:CD003193.

A6. Myers DL. Female mixed urinary incontinence: a clinical review. *JAMA*. 2014;311:2007-2014.

A7. DuBeau CE, Kraus SR, Griebling TL, et al. Effect of fesoterodine in vulnerable elderly subjects with urgency incontinence: a double-blind, placebo controlled trial. *J Urol*. 2014;191:395-404.

A8. Wagg A, Dale M, Tretter R, et al. Randomised, multicentre, placebo-controlled, double-blind crossover study investigating the effect of solifenacin and oxybutynin in elderly people with mild cognitive impairment: the SENIOR study. *Eur Urol*. 2013;64:74-81.

A9. Herbison GP, Arnold EP. Sacral neuromodulation with implanted devices for urinary storage and voiding dysfunction in adults. Cochrane Database System Review publications. 2009;2:CD004202.

A10. Peters KM, Carrico DJ, Wooldridge LS, et al. Percutaneous tibial nerve stimulation for the long-term treatment of overactive bladder: 3-year results of the STEP study. *J Urol*. 2013;189:2194-2201.

A11. Visco AG, Brubaker L, Richter HE, et al. Anticholinergic therapy vs. onabotulinumtoxinA for urgency urinary incontinence. *N Engl J Med*. 2012;367:1803-1813.

A12. Nitti VW, Dmochowski R, Herschorn S, et al. OnabotulinumtoxinA for the treatment of patients with overactive bladder and urinary incontinence: results of a phase 3, randomized, placebo controlled trial. *J Urol*. 2013;189:2186-2193.

A13. Diokno AC, Sampselle CM, Herzog AR, et al. Prevention of urinary incontinence by behavioral modification program: a randomized, controlled trial among older women in the community. *J Urol*. 2004;171:1165-1171.

A14. Brown JS, Wing R, Barrett-Connor E, et al. Lifestyle intervention is associated with lower prevalence of urinary incontinence: the Diabetes Prevention Program. *Diabetes Care*. 2006;29:385-390.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Kupelian V, Rosen RC, Roehrborn CG, et al. Association of overactive bladder and C-reactive protein levels. Results from the Boston Area Community Health (BACH) Survey. *BJU Int*. 2012;110: 401-407.
2. Yamaguchi O, Nomiya M, Andersson KE. Functional consequences of chronic bladder ischemia. *Neurourol Urodyn*. 2014;33:54-58.
3. Huang AJ. Nonsurgical treatments for urinary incontinence in women: summary of primary findings and conclusions. *JAMA Intern Med*. 2013;173:1463-1464.
4. Smith A, Bevan D, Douglas HR, et al. Management of urinary incontinence in women: summary of updated NICE guidance. *BMJ*. 2013;347:f5170.
5. Gormley EA, Lightner DJ, Burgio KL, et al. Diagnosis and treatment of overactive bladder (non-neurogenic) in adults: AUA/SUFU guideline. *J Urol*. 2012;188:2455-2463.
6. Resnick NM, Perera S, Tadic S, et al. What predicts and what mediates the response of urge urinary incontinence to biofeedback? *Neurourol Urodyn*. 2013;32:408-415.
7. Wagg A, Gibson W, Johnson T 3rd, et al. Urinary incontinence in frail elderly persons: report from the 5th International Consultation on Incontinence. *Neurourol Urodyn*. 2014;[Epub ahead of print].
8. Wagg A, Cardozo L, Nitti VW, et al. The efficacy and tolerability of the beta3-adrenoceptor agonist mirabegron for the treatment of symptoms of overactive bladder in older patients. *Age Ageing*. 2014;43:666-675.

REVIEW QUESTIONS

1. Which of the following is true about acute urinary retention in an older man?

A. Indicates the need for surgical decompression
B. Is treated effectively with α-adrenergic blockers
C. Is treated effectively with bethanechol
D. Requires treatment of the underlying urinary tract abnormality
E. None of the above

Answer: E The differential diagnosis for urinary retention extends beyond urethral obstruction, particularly in elderly people. Many older patients may develop retention owing to a weak detrusor associated with age, autonomic neuropathy, or previous urethral obstruction. In addition, fecal impaction, pain (e.g., following hip replacement), and medications with urinary tract side effects (e.g., anticholinergics, sedating antihistamines, decongestants, opiates) may induce acute urinary retention, particularly in patients with underlying bladder weakness or obstruction. Thus, the bladder should be decompressed for at least a week while reversible causes are addressed; the larger the postvoid residual volume, the longer the decompression should be. Decompression allows some restoration of detrusor strength, which also facilitates urodynamic testing should it be necessary. α-Adrenergic blockers are effective for men with symptoms of prostatism, but clinical trials of its efficacy have excluded patients with significant urinary retention. Bethanechol has not proved effective in controlled clinical trials. Decompression in some elderly patients can reduce but not eliminate residual urine. Provided it does not cause symptoms or renal compromise, subclinical retention need not necessarily be treated in all elderly patients, even if obstruction is present.

2. A 58-year-old obese woman with significant daily stress incontinence comes for a follow-up appointment. Her bladder diary shows maximal voided volume of 125 mL during the daytime. Each of these measures is appropriate except which of the following?

A. Adjustment of fluid excretion and voiding intervals
B. Advise weight reduction
C. Teach her postural maneuvers
D. Consideration of surgical correction
E. Pelvic floor muscle exercises

Answer: A Recent evidence suggests that weight loss will improve stress incontinence in obese women, and data support the use of postural maneuvers, pelvic floor muscle exercises, and surgical correction as well. Adjusting fluid excretion and voiding intervals can also be useful, especially for women with volume-dependent stress leakage. It can work particularly well for women with a threshold of at least 150 mL and best in those with a threshold of more than 250 mL. When the threshold for stress incontinence is this low, however, the extent of fluid restriction required is usually not feasible and might even lead to dangerous dehydration.

3. A 78-year-old woman with dementia has responded modestly to donepezil (Aricept, a cholinesterase inhibitor) for the past year. The recent onset of urge incontinence led her primary physician to prescribe tolterodine last month while awaiting your assessment. Her incontinence has responded well. The next appropriate step is which of the following?

A. Discontinue tolterodine because of its interaction with donepezil.
B. Discontinue donepezil because her cognitive function is stable.
C. Discontinue both drugs because she is stable and the urge incontinence may reflect an adverse effect of the donepezil.
D. Continue both drugs and monitor her for deterioration in cognitive function.
E. Taper the tolterodine.

Answer: D Because cholinesterase inhibitors block the metabolism of acetylcholine, they can provoke urge incontinence, especially in older adults who already may have underlying age-related involuntary detrusor contractions that have not yet caused incontinence. However, despite prescription of these agents to millions of demented patients, there is little evidence that they cause incontinence. Moreover, because the benefits of these drugs for dementia are modest at best and not seen in most patients who use them, patients and families may decide that the benefit of the bladder relaxant outweighs the risk. Particularly in this patient, who has already benefited from tolterodine without notable cognitive deterioration, it is worth continuing therapy and monitoring her cognitive status.

4. A 53-year-old woman says that she has frequent leakage. It most commonly occurs about 10 seconds after she coughs or laughs. Having suffered for years, she is interested in the approach with the best chance of success.

A. Pelvic muscle strengthening exercises ("Kegels")
B. An α-adrenergic agonist (e.g., pseudoephedrine)
C. A pessary
D. Urethral suspension surgery
E. A bladder relaxant drug

Answer: E Most stress incontinence results from weakness of intrinsic sphincter muscle or pelvic muscle support. In all such cases, however, leakage coincides instantaneously with the increase in abdominal pressure and ceases as soon as the cough ends. In this case, however, there is a delay between the cough and leakage. This finding suggests that the cough has triggered an involuntary detrusor contraction, which then leads to leakage. Therapy should focus on reducing or eliminating the precipitant of coughing (e.g., bronchospasm, use of an angiotensin-converting enzyme inhibitor) or the underlying detrusor overactivity.

5. Which of the following is true of incontinence in demented elderly patients?

A. Is inevitable
B. Is virtually always due to detrusor overactivity/overactive bladder
C. Is unlikely to respond to therapy
D. Is multifactorial and often reversible
E. Treatment should focus primarily on preventing skin breakdown

Answer: D Incontinence is never normal, even in patients with dementia. Detrusor overactivity is the most common type of lower urinary tract dysfunction among demented and incontinent nursing home residents, but it is also the most common dysfunction among their dry peers. Moreover, in 40% of these individuals, incontinence is not associated with detrusor overactivity but with obstruction (in men), stress incontinence (in women), or a combination of an outlet and a detrusor problem, and the cause does not correlate with either the presence or severity of dementia. Thus, it is not tenable to attribute incontinence *a priori* to detrusor overactivity. Because incontinence in elderly people is usually multifactorial, involving urinary tract as well as non–urinary tract contributions, it is often treatable. Even among nursing home patients, randomized studies of just a toileting regimen have shown more than a 50% reduction in incontinent episodes overall and full daytime continence in nearly 40% of residents. Because causes of transient incontinence are particularly common in demented individuals, additional benefit can be achieved by addressing other common contributors to their incontinence, including medication use, depression, fecal impaction, urinary tract infection, atrophic vaginitis, and disorders of fluid excretion. For those with detrusor overactivity, a bladder relaxant may provide additional help; for those with urethral obstruction or incompetence, minimally invasive procedures can be helpful. It is also important to prevent skin breakdown, but such treatment should not be the primary approach to the incontinent nursing home resident.

27 NEUROPSYCHIATRIC ASPECTS OF AGING

SHARON K. INOUYE

DEFINITION

The process of aging produces important physiologic changes in the central nervous system (Table 27-1), including neuroanatomic, neurotransmitter, and neurophysiologic changes. These processes result in age-related symptoms and manifestations (Table 27-2) for many older persons. These physiologic changes develop at dramatically variable rates among older persons, however, and the decline may be modified by factors such as diet, exercise, environment, lifestyle, genetic predisposition, disability, disease, and side effects of drugs. These changes can result in the common age-related symptoms of benign senescence, slowed reaction time, postural hypotension, vertigo or giddiness, presbyopia, presbycusis, stiffened gait, and sleep difficulties. In the absence of disease, these physiologic changes usually result in relatively modest symptoms and little restriction in activities of daily living. These changes decrease physiologic reserve, however, and increase the susceptibility to challenges posed by disease-related, pharmacologic, and environmental stressors. As a result, mild (relative risk, 1.2) and moderate (relative risk, 1.4) cognitive impairment are associated with increased mortality.[1]

EPIDEMIOLOGY

Neuropsychiatric disorders, the leading cause of disability in older persons, account for nearly 50% of functional incapacity. Severe neuropsychiatric conditions have been estimated to occur in 15 to 25% of older adults worldwide. These conditions are due to diseases that increase with age but are not part of the normal aging process. Alzheimer disease and related dementias occur in approximately 10% of adults aged 65 years and older and in 40% of

TABLE 27-1 AGE-RELATED PHYSIOLOGIC CHANGES IN THE CENTRAL NERVOUS SYSTEM

Neuroanatomic changes
Brain atrophy
Decreased neuron counts
Increased neuritic plaques
Increased lipofuscin and melanin
Neurotransmitter changes
Decline in cholinergic transmission
Decreased dopaminergic synthesis
Decreased catecholamine synthesis
Neurophysiologic changes
Decreased cerebral blood flow
Electrophysiologic changes (slowing of alpha rhythm, increased latencies in evoked responses)

TABLE 27-2 NEUROPSYCHIATRIC MANIFESTATIONS OF AGE-RELATED PHYSIOLOGIC CHANGES

SYSTEM	MANIFESTATION
Cognition	Forgetfulness Processing speed declines throughout adult life Neuropsychological declines: selective attention, verbal fluency, retrieval, complex visual perception, logical analysis
Reflexes	Stretch reflexes lose sensitivity Decreased or absent ankle reflexes Decreased autonomic and righting reflexes, postural instability
Sensory	Presbycusis (high-frequency hearing loss), tinnitus Deterioration of vestibular system, vertigo Presbyopia (decreased lens elasticity) Slowed pupil reactivity, decreased upgaze Olfactory system deterioration Decreased vibratory sensation
Gait and balance	Gait stiffer, slowed, forward flexed Increased body sway and mild unsteadiness
Sleep	Decreased sleep efficiency, fatigue Increased awakenings, insomnia Decrease in sleep stages 3 and 4 Sleep duration more variable, more naps

TABLE 27-3 GERIATRIC DEPRESSION SCALE—SHORT FORM

1. Are you basically satisfied with your life?	yes/**NO**
2. Have you dropped many of your activities and interests?	**YES**/no
3. Do you feel that your life is empty?	**YES**/no
4. Do you often get bored?	**YES**/no
5. Are you in good spirits most of the time?	yes/**NO**
6. Are you afraid that something bad is going to happen to you?	**YES**/no
7. Do you feel happy most of the time?	yes/**NO**
8. Do you feel helpless?	**YES**/no
9. Do you prefer to stay home rather than going out and doing new things?	**YES**/no
10. Do you feel you have more problems with memory than most?	**YES**/no
11. Do you think it is wonderful to be alive now?	yes/**NO**
12. Do you feel pretty worthless the way you are now?	**YES**/no
13. Do you feel full of energy?	yes/**NO**
14. Do you feel that your situation is hopeless?	**YES**/no
15. Do you think that most people are better off than you are?	**YES**/no

Scoring: Answers indicating depression are capitalized; six or more capitalized answers indicate depressive symptoms.
Modified with permission from Yesavage J, Brink T, Rowe T, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res.* 1983;17:37-49.

those older than 85 years (Chapter 402). Delirium occurs in 5 to 10% of all persons older than 65 years and in up to 80% of older persons during hospitalizations for acute illnesses (Chapter 28). Severe depression (Chapter 397) occurs in approximately 5% of older adults, with 15% having significant depressive symptoms. Anxiety disorders occur in 10% of older adults. Older individuals are also subject to substantial morbidity and functional disability from cerebrovascular disease (Chapters 406 through 408), Parkinson disease (Chapter 409), peripheral neuropathies (Chapter 420), degenerative myelopathies (Chapters 400 and 422), spinal stenosis and disc disease (Chapter 400), seizure disorders (Chapter 403), sleep apnea (Chapter 100), visual disturbances (Chapter 423), falls (Chapter 25), incontinence (Chapter 26), and impotence (Chapter 234).

DIAGNOSIS

To diagnose these neuropsychiatric conditions, physicians must understand and perform a mental status examination and an assessment of functional capacity and know the uses and side effects of psychoactive drugs in geriatric patients.

Mental Status Examination

In addition to a detailed neurologic examination, evaluation of neuropsychiatric disturbances in older persons requires a careful mental status examination, including an assessment of mood, affect, and cognition. Brief screening tests are available to evaluate these domains and to assist in the detection of potential problems requiring further evaluation and treatment (see Chapter 24). Scores of 6 or more on the 15-item short-form Geriatric Depression Scale (Table 27-3) indicate substantial depressive symptoms requiring further evaluation. For cognitively impaired patients, observer (proxy)–rated depression scales, such as the Hamilton Depression Scale or Cornell Scale, are recommended.

Early cognitive deficits can easily be missed during conversation because intellectual impairment can be masked with intact social skills. Given the high frequency of cognitive impairment, formal cognitive screening is reasonable but not required for all older persons.[2] Ideally, cognitive testing should evaluate at least the general domains of attention, orientation, language, memory, visuospatial ability, and conceptualization. To exclude delirium, attention should be assessed first by asking the patient to perform a task, such as repeating digits (normal span: more than five forward or more than three backward) or reciting the months backward (allow one error maximum); the remainder of cognitive testing would not be useful in an inattentive or delirious patient. For further cognitive testing, many brief, practical screening instruments are available. Historically, the most widely used instrument has been the Mini-Mental State Examination, a 19-item, 30-point scale that can be completed in 10 minutes. This copyrighted instrument now requires a per-use fee if the official version is used. Useful, brief alternative instruments include the Short Portable Mental Status Questionnaire (Table 27-4) and the Mini-Cog

TABLE 27-4 SHORT PORTABLE MENTAL STATUS QUESTIONNAIRE

QUESTION	RESPONSE			ERROR?
What are the date, month, and year?*	Date	Month	Year	
What is the day of the week?				
What is the name of this place?				
What is your phone number?				
How old are you?				
When were you born?				
Who is the current president?				
Who was the president before him?				
What was your mother's maiden name?				
Can you count backward from 20 by 3s?				

*A mistake on *any* part of this question should be scored as an error.
Total possible errors: 10; more than three errors indicates cognitive impairment.
From Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *J Am Geriatr Soc.* 1975;23:433-441. Copyright © E. Pfeiffer 1994. Reproduced with permission of the author.

TABLE 27-5 MINI-COG TEST

1. Instruct the patient to listen carefully to and remember three unrelated words and then to repeat the words: banana, sunrise, chair.
2. Instruct the patient to draw the face of a clock, either on a blank sheet of paper or on a sheet with a large circle already drawn on the page. After the patient puts the numbers on the clock face, ask him or her to draw the hands of the clock to read a specific time, such as 11:10. These instructions can be repeated, but no additional instructions should be given. Allow up to 3 minutes to complete the clock drawing.
3. Ask the patient to repeat the three previously presented words.
4. Give 1 point for each correct word and 2 points for a correctly drawn clock. Scores <3 suggest cognitive impairment.

Modified from Borson S, Scanlan J, Watanabe J, et al. Improving identification of cognitive impairment in primary care. *Int J Geriatr Psychiatry.* 2006;21:349-355. Reprinted by permission of the copyright holder (S. Borson).

test (Table 27-5), both of which can be completed in less than 5 minutes. More detailed testing can be conducted with the Montreal Cognitive Assessment (Fig. 27-1), which requires 15 to 20 minutes; scores less than 26 indicate cognitive impairment. Questions to evaluate judgment and problem-solving ability in hypothetical situations, such as in a fire or when driving,

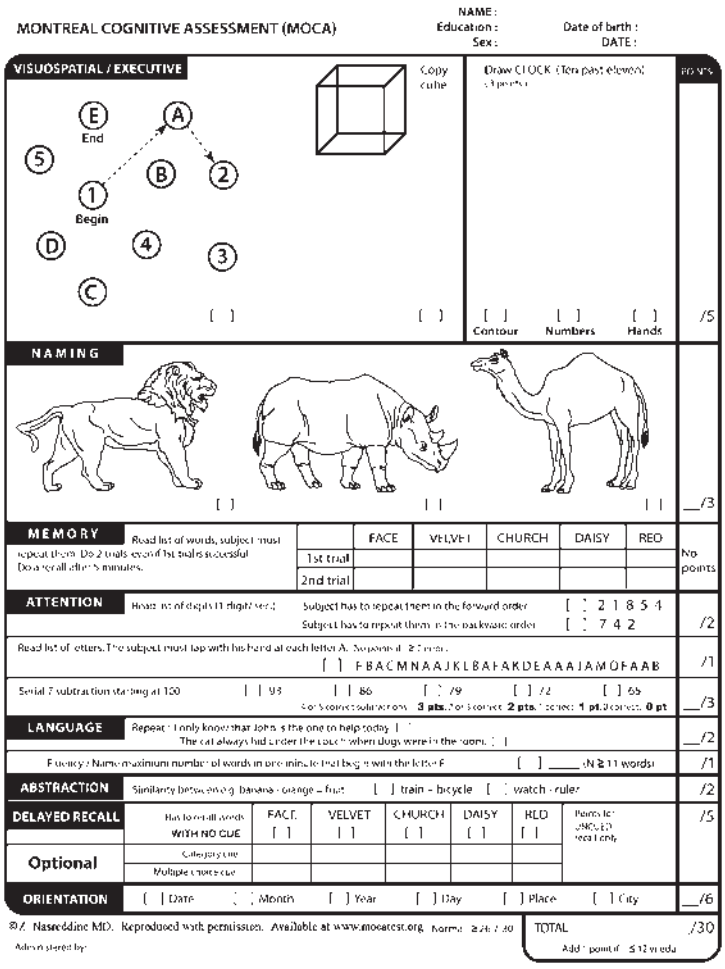

FIGURE 27-1. Montreal Cognitive Assessment. (Reproduced with permission from Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* 2005;53:695-699.)

can provide crucial insight into the patient's ability to function safely and independently.

Functional Assessment

Functional impairment, defined as difficulty in performing daily activities, is common among older persons. Although it is not routinely evaluated in the standard medical assessment, determination of the patient's degree of functional incapacity based on medical and neuropsychiatric conditions is crucial to diagnosis of mild cognitive impairment and dementia as well as to understanding of the burden of the disease and its impact on the individual's daily life. The important relationship between functional status and health in older persons is reflected by the finding that functional measures are stronger predictors of mortality after hospitalization than are admitting diagnoses. Functional measures strongly predict other important long-term outcomes in the elderly, such as future care needs, caregiver burden, risk for institutionalization, and long-term prognosis.[3] Functional independence is critical if patients are to remain living independently in the community, and functional decline represents the leading risk factor for nursing home placement.

The functional assessment should evaluate the patient's ability to perform basic self-care activities of daily living and instrumental activities of daily living, the higher level activities needed for independent living. Activities of daily living include basic self-care skills, such as feeding, grooming, bathing, dressing, toileting, transferring, and walking. Instrumental activities of daily living are more complex tasks, including shopping, preparing meals, managing finances, housekeeping, using the telephone, taking medications, and driving or using public transportation. The functional assessment is conducted with the patient, with corroboration from a family member or caregiver. Other related domains that should be assessed include vision, hearing, continence, nutritional status, safety, falls, living situation, availability of social support, and socioeconomic status.

The onset of acute cognitive or functional decline is often the first and sometimes the only sign of serious acute illness in older persons and warrants immediate medical attention. Similarly, the onset or worsening of related conditions, such as delirium, falls, incontinence, depression, frailty, or failure to thrive, heralds the need for prompt medical evaluation.

TREATMENT

Psychoactive Effects of Drugs in Older Patients

Adverse Drug Events in the Elderly

Iatrogenic complications occur in 29 to 38% of older hospitalized patients, with a 3- to 5-fold increased risk in older compared with younger patients. Adverse drug events, the most common type of iatrogenic complication, account for 20 to 40% of all complications, and at least 27% of these are preventable.[4] The elderly are particularly vulnerable to adverse drug reactions because of multiple-drug regimens, multiple chronic diseases, relative renal and hepatic insufficiency, decreased physiologic reserve, and altered drug metabolism and receptor sensitivity with aging.[5] Inappropriate drug use has been reported in up to 40% of older patients who are hospitalized, with more than one quarter of these patients having absolute contraindications to the drug and the others being given a drug that was unnecessary. Among an estimated 100,000 annual drug-related emergency hospitalizations in patients older than 65 years, the most commonly implicated drugs are warfarin (33%), insulin (14%), oral antiplatelet agents (13%), and oral hypoglycemic agents (11%).[6] Because up to 50% of adverse drug events occur in patients receiving inappropriate drugs, the potential for reducing these adverse events is substantial.

Drugs with Psychoactive Effects

Nearly every class of drug has the potential to cause mental status changes in a vulnerable patient, but specific drugs are commonly implicated (Table 27-6) and should be used with caution in older patients.[6] Many cases of delirium or cognitive decline in older patients may be preventable through avoidance, substitution, or dose reduction of these psychoactive drugs. Long-acting benzodiazepines (e.g., diazepam, clonazepam, chlordiazepoxide) are particularly problematic medications for the elderly and should not be used to treat insomnia. If nonpharmacologic approaches to the management of insomnia are unsuccessful, short-term use of an intermediate-acting benzodiazepine without active metabolites (e.g., lorazepam 0.5 mg, half-life of 10 to 15 hours) is recommended. Drugs with anticholinergic effects (e.g., antihistamines, antidepressants, neuroleptics, antispasmodics) produce a panoply of poorly tolerated side effects in older patients, including delirium, postural hypotension, urinary retention, constipation, and dry mouth. Of the narcotics, meperidine causes delirium more frequently than other agents because of an active metabolite, normeperidine. Cardiac drugs, such as digitalis and antiarrhythmic agents, have prolonged half-lives, narrowed therapeutic windows, and decreased protein binding in older patients. The clinician should be aware that toxicity with these agents (e.g., digoxin) can occur even at therapeutic drug levels. The H_2-receptor antagonists (e.g., cimetidine, ranitidine, famotidine, nizatidine) are among the most common causes of drug-induced delirium in the elderly because of their frequent use; clinicians should strongly consider the use of less toxic alternatives (e.g., sucralfate, antacids) or dosage reductions for older patients, especially when the medication is being used for prophylaxis rather than treatment of active disease. Proton pump inhibitors have been associated with delirium in case reports but have less neuropsychiatric toxicity than H_2-receptor antagonists.

Psychoactive drugs account for nearly 50% of preventable adverse drug events, often in patients who have been prescribed three or more psychoactive drugs, frequently at inappropriately high doses in the elderly. Delirium and cognitive impairment are the most frequent adverse outcomes of psychoactive drugs. The use of any psychoactive drug is associated with a 4-fold increased risk of delirium or cognitive decline, but the outcomes of these conditions depend on the type or class of drug administered and the total number of drugs received. Sedative-hypnotic drugs are associated with a 3- to 12-fold increased risk for delirium or cognitive decline, narcotics are associated with a 2- to 3-fold increased risk, and anticholinergic drugs are associated with a 5- to 12-fold increased risk. Each drug carries its own individual risk for adverse outcomes, and when multiple drugs are used, the overall risk is compounded by the heightened potential for drug-drug and drug-disease interactions. If more than three drugs are added in a 24-hour period, the risk of delirium increases 4-fold. Similarly, the risk of cognitive decline increases directly with the number of drugs prescribed, from a 3-fold increased risk with two or three drugs to a 14-fold increased risk with six or more drugs.

Principles of Drug Therapy in the Elderly

Physicians always should consider whether nonpharmacologic approaches (Chapter 39) may be used as alternatives to medications in older persons. Relaxation techniques, massage, and music are highly effective for the treatment of insomnia and anxiety; localized pain can often be managed effectively

TABLE 27-6 DRUGS WITH PSYCHOACTIVE EFFECTS

Sedative-hypnotics
Benzodiazepines (especially flurazepam, diazepam)
Barbiturates
Sleeping medications (chloral hydrate)
Narcotics (especially meperidine)
Anticholinergics
Antihistamines (diphenhydramine, hydroxyzine)
Antispasmodics (belladonna, Lomotil)
Heterocyclic antidepressants (amitriptyline, imipramine, doxepin)
Neuroleptics (chlorpromazine, haloperidol, thioridazine)
Antiparkinson drugs (benztropine, trihexyphenidyl)
Atropine, scopolamine
Cardiac drugs
Digitalis glycosides
Antiarrhythmics (quinidine, procainamide, lidocaine)
Antihypertensives (β-blockers, methyldopa)
Gastrointestinal drugs
H_2-receptor antagonists (cimetidine, ranitidine, famotidine, nizatidine)
Proton pump inhibitors (esomeprazole, lansoprazole, omeprazole, pantoprazole)
Metoclopramide (Reglan)
Miscellaneous drugs
Nonsteroidal anti-inflammatory drugs
Corticosteroids
Anticonvulsants
Levodopa
Lithium
Over-the-counter drugs
Cold and sinus preparations (antihistamines, pseudoephedrine)
Sleep aids (diphenhydramine, alcohol-containing elixirs)
Stay Awake (caffeine)
Nausea, gastrointestinal relief (Donnagel, meclizine, H_2-receptor antagonists, loperamide)

TABLE 27-7 GUIDELINES FOR DRUG THERAPY IN THE ELDERLY

GENERAL PRINCIPLES
Remember that the elderly are highly sensitive to the psychoactive effects of all drugs.
Know the pharmacology of the drugs you prescribe. Know a few drugs well.
RECOMMENDED APPROACH
Use nonpharmacologic approaches whenever possible.
Avoid *routine* use of "as needed" drugs for sleep, anxiety, pain.
Choose the drug with the least toxic potential.
Substitute less toxic alternatives whenever possible (antacid or sucralfate for an H_2-blocker or proton pump inhibitors, Metamucil or Kaopectate for Imodium or Lomotil, scheduled acetaminophen regimen for pain management).
Reduce the dosage.
"Start low and go slow."
Start with 25 to 50% of the standard dose of psychoactive drugs in the elderly.
Titrate the drug slowly.
Set realistic end points: titrate to improvement, not elimination of symptoms.
Keep the regimen simple.
Regularly reassess the medication list. Have the patient bring in all bottles and review what is being taken.
Reevaluate long-time drug use because the patient is changing.
Review over-the-counter medications, including herbal remedies.

with local measures such as injection, heat, ultrasound, and transcutaneous electrical stimulation.

When drug therapy is required in the elderly, physicians should choose the drug with the least toxic potential and emphasize drugs that have been well tested in older populations (Table 27-7). It is often wise to start with 25 to 50% of the standard adult dosage for a psychoactive drug and to increase the dose slowly. Drug regimens should be kept simple, with the fewest drugs and the fewest number of pills possible. Most important, the medication list should be reassessed frequently. Involving a clinical pharmacist in the care team can significantly improve healthcare outcomes.[A1] Systematic interventions involving geriatricians, clinical pharmacists, and computer-based monitoring systems also can significantly reduce the frequency of adverse drug reactions in older persons.[A2]

For persons with dementia or cognitive impairment, the following classes of drugs should be avoided or discontinued if possible: benzodiazepines, H_2-receptor antagonists, meperidine, sedative-hypnotics, and thioridazine.[4] Even long-standing medications should be reevaluated as the patient changes with age and illness. Long-term use does not justify continued use. The physician should review with the patient all prescribed and over-the-counter medications on a regular basis, preferably by having the patient bring in all medication bottles and indicate how each is being taken. Patients frequently underestimate the toxic potential of over-the-counter medications and herbal remedies, and they may be using a variety of such agents that could potentiate the side effects or directly counteract the desired effects of prescription medications (Chapter 29). For example, high-risk over-the-counter medications for older persons include nonsteroidal anti-inflammatory agents, H_2-blockers, and antihistamines. In addition, herbal remedies may interact with warfarin either to increase or to decrease its effect (St. John's wort decreases the effect and makes anticoagulation more difficult; gingko biloba increases the effect and may cause bleeding)[7]; others (such as kava kava, echinacea, and Chinese herbal preparations) have been associated with the risk of hepatotoxicity.

Grade A References

A1. Lee JK, Slack MK, Martin J, et al. Geriatric patient care by U.S. pharmacists in healthcare teams: systematic review and meta-analysis. *J Am Geriatr Soc*. 2013;61:1119-1127.

A2. Dalleur O, Boland B, Losseau C, et al. Reduction of potentially inappropriate medications using the STOPP criteria in frail older inpatients: a randomised controlled study. *Drugs Aging*. 2014; 31:291-298.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Sachs GA, Carter R, Holtz LR, et al. Cognitive impairment: an independent predictor of excess mortality: a cohort study. *Ann Intern Med*. 2011;155:300-308.
2. Moyer VA. Screening for cognitive impairment in older adults: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160:791-797.
3. Kåreholt I, Lennartsson C, Gatz M, et al. Baseline leisure time activity and cognition more than two decades later. *Int J Geriatr Psychiatry*. 2011;26:65-74.
4. American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society Updated Beers Criteria for potentially inappropriate medication use in older adults. *J Am Geriatr Soc*. 2012;60:616-631.
5. Permpongkosol S. Iatrogenic disease in the elderly: risk factors, consequences, and prevention. *Clin Interv Aging*. 2011;6:77-82.
6. Budnitz DS, Lovegrove MC, Shehab N, et al. Emergency hospitalizations for adverse drug events in older Americans. *N Engl J Med*. 2011;365:2002-2012.
7. Shi S, Klotz U. Drug interactions with herbal medicines. *Clin Pharmacokinet*. 2012;51:77-104.

REVIEW QUESTIONS

1. An 82-year-old woman with atrial fibrillation, congestive heart failure, hyperlipidemia, diabetes mellitus, and hypertension is brought to you by her daughter for evaluation of the acute onset of disorientation and agitation. Her medications include furosemide, metoprolol, atorvastatin, sertraline, warfarin, ranitidine, amlodipine, and an over-the-counter sleep remedy containing diphenhydramine. Which of her medications are *most likely* to be contributing to her altered mental status?

A. Furosemide, metoprolol, and amlodipine
B. Warfarin, atorvastatin, and sertraline
C. Sertraline, ranitidine, and diphenhydramine
D. Diphenhydramine, warfarin, and metoprolol
E. Atorvastatin, furosemide, and sertraline

Answer: C Drugs with psychoactive side effects are most likely contributing to this patient's altered mental status. Anticholinergic drugs (such as diphenhydramine) are common contributors to mental status changes. Over-the-counter medications containing these agents should be used with caution in older persons. H_2-receptor blockers are among the most common drugs contributing to delirium. Antidepressants may also contribute to mental status changes, even at therapeutic levels. Choice A is not correct because these drugs are not highly correlated with mental status changes, although metoprolol can cross the blood-brain barrier. A major adverse effect of this combination in the elderly is orthostatic hypotension. All of the other combinations have only one drug listed that has marked psychoactive effects.

2. A 78-year-old man with hypertension, peripheral vascular disease, osteoarthritis of the knees, and macular degeneration is brought in by his daughter for evaluation of confusion and "failure to thrive" after a flu-like illness 2 months ago. The daughter reports that during the past few weeks, her father has been drowsy and sleeping most of the day. His memory is poor at times, but he remains independent in all self-care activities. Normally active, he has stopped his usual gardening and walking and shows little interest in interacting with his grandchildren, which he normally enjoys. His appetite is poor, and he has lost weight. Your examination reveals a flattened affect with poor eye contact. Neurologic examination is nonfocal without cogwheeling or tremor. On the Short Portable Mental Status Questionnaire, he has two errors after initially poor effort, which improved after extensive prompting and encouragement. He scores 7 of 15 on the Geriatric Depression Scale. What is the most likely diagnosis?

A. Early-onset Parkinson disease
B. Mild cognitive impairment
C. Early dementia (combined vascular and Alzheimer type)
D. Post–viral encephalitis
E. Major depressive disorder

Answer: E This patient demonstrates symptoms of a major depressive disorder, including anhedonia, loss of interest in his usual activities, excessive sleep, loss of appetite, and weight loss. His cognitive screening test reveals cognitive "dilapidation" with poor effort, sometimes called a pseudodementia. His depression screening reveals marked depressive symptoms. In older persons, somatic symptoms are often more prominent than mood complaints. This patient should be referred for immediate geropsychiatric evaluation and management. Precipitants are not always present, but bereavement and acute or chronic medical conditions can contribute. There is no evidence of Parkinson disease, and the patient does not meet criteria for mild cognitive impairment or dementia.

3. Older persons are at least three times more likely than younger persons to experience an adverse drug reaction. All of the following contribute to this propensity *except* which one?

A. Greater number of drugs
B. Multiple chronic conditions
C. Decreased glomerular filtration rate
D. Decreased hepatic function
E. Increased physiologic reserve

Answer: E Because of a greater number of drugs and chronic conditions, the elderly are at increased risk of drug-drug and drug-disease interactions. In addition, renal and hepatic functioning are relatively decreased, as is physiologic reserve capacity. Drug receptor sensitivity can be either increased or decreased, depending on the drug and receptor regulation.

4. An 86-year-old widowed woman living alone with peripheral vascular disease, lower extremity amputation, hypertension, and old stroke without residua is brought in by her neighbor for increasing forgetfulness during the past year. Your examination reveals a cheerful, interactive, and well-groomed woman with normal findings on neurologic examination. She has five errors on the Short Portable Mental Status Questionnaire. What is the most appropriate next step in her evaluation?

A. Obtain brain imaging and lumbar puncture
B. Assess her activities of daily living and level of independence
C. Initiate treatment with an acetylcholinesterase inhibitor
D. Initiate treatment with an antidepressant
E. Refer her for nursing home placement

Answer: B Assessment of activities of daily living (both basic and instrumental) is key to the diagnosis of dementia, and information must often be obtained from a reliable informant. This is the important next step in the evaluation. Dementia can be diagnosed only in the presence of a functional impairment. After this assessment is complete, further evaluation can include laboratory work, neuroimaging, and other testing as required. Her ability to live independently must also be assessed, and social work involvement is recommended after a diagnosis is established.

28

DELIRIUM OR ACUTE MENTAL STATUS CHANGE IN THE OLDER PATIENT

SHARON K. INOUYE

Mental status change, one of the most common presenting symptoms in acutely ill elders, is estimated to account for 30% of emergency evaluations among older patients. Mental status often serves as a barometer of the underlying health of an elderly patient and is commonly the only symptom of serious underlying disease. A broad range of medical, neurologic, and psychiatric conditions can lead to mental status changes (Chapters 397 and 402). A systematic approach aids in the evaluation of suspected mental status change in an older patient (Fig. 28-1).

The first step in evaluating suspected altered mental status in an older patient is to obtain a detailed history from a reliable informant to establish the patient's baseline level of cognitive function and the clinical course of any cognitive changes. Chronic changes (those occurring during months to years) most likely represent an underlying dementing illness and should be evaluated accordingly (Chapter 402). Acute changes (those occurring during days to weeks)—even if superimposed on an underlying dementia—should be evaluated by a formal cognitive assessment to determine whether delirium is present. If features of delirium (e.g., inattention, disorganized thinking, altered level of consciousness, fluctuating symptoms) are not present, further evaluation for depression, acute nonorganic psychotic disorders, or other psychiatric conditions is indicated.

DELIRIUM

Delirium, a clinical syndrome characterized as an acute disorder of attention and cognitive function, is the most frequent complication of hospitalization for elders and is a potentially devastating problem. Delirium is often unrecognized despite sensitive methods for its detection, and its complications may be preventable.

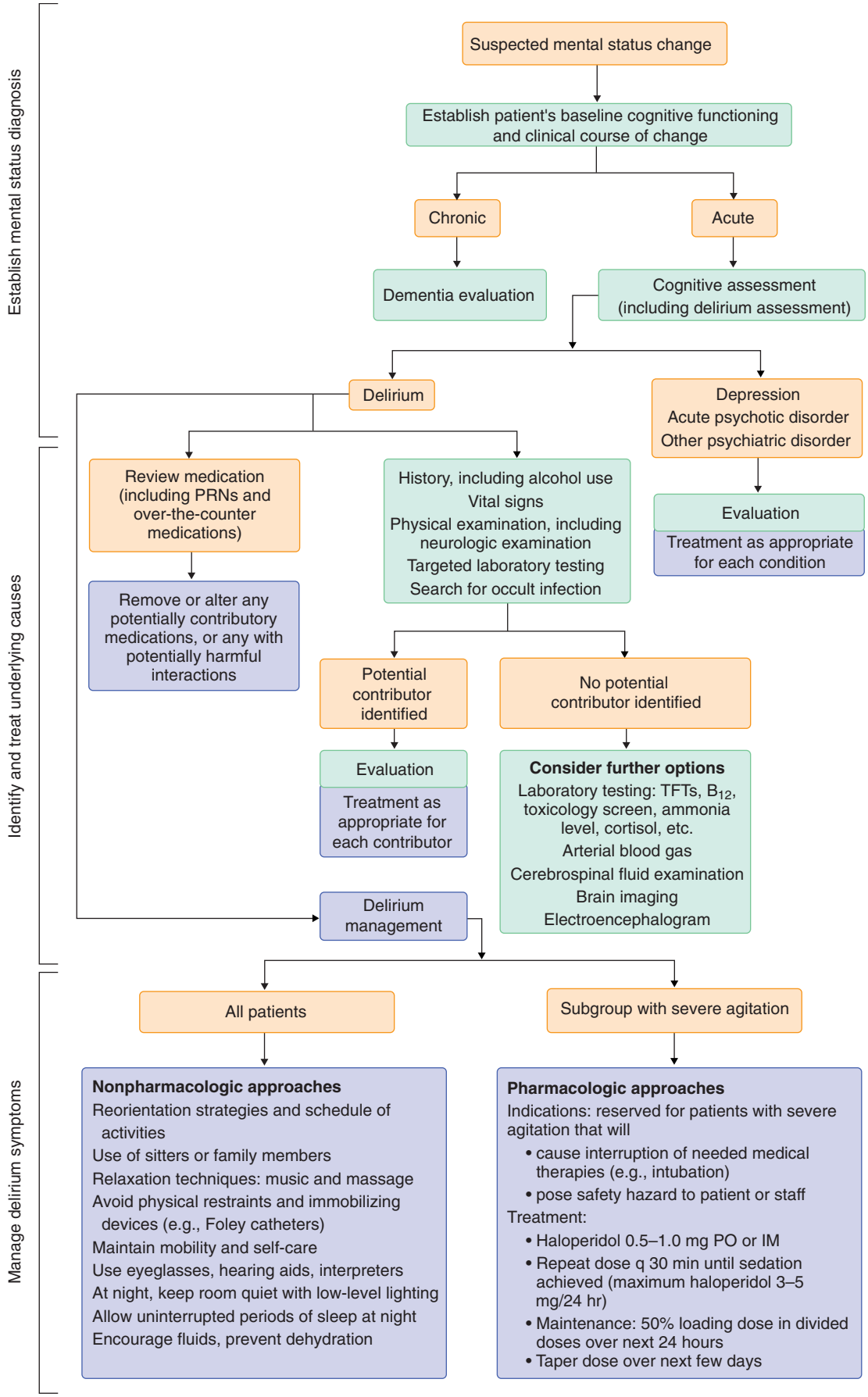

FIGURE 28-1. Algorithm for the evaluation of suspected mental status change in older patients. PRN = as needed; TFTs = thyroid function tests.

TABLE 28-1 DIAGNOSTIC CRITERIA FOR DELIRIUM

DSM-5 CRITERIA*

- Disturbed attention (i.e., reduced ability to direct, focus, sustain, and shift attention) and awareness (reduced orientation to the environment) that has developed in a short time (usually hours to a few days), represents a change from baseline attention and awareness, and tends to fluctuate in severity during the course of a day
- An additional disturbance in cognition (e.g., memory deficit, disorientation, language, visuospatial ability, or perception)
- Evidence from the history, physical examination, or laboratory findings that the disturbance is a direct physiologic consequence of another medical condition, substance intoxication or withdrawal, exposure to a toxin, or multiple causes
- No evidence for another preexisting, established, or evolving neurocognitive disorder to explain the inattention and cognitive disturbance
- No evidence for a reduced level of arousal, such as coma

CAM DIAGNOSTIC ALGORITHM†

Feature 1. Acute onset and fluctuating course. This information is usually obtained from a family member or nurse and is shown by positive responses to the following questions: Is there evidence of an acute change in mental status from the patient's baseline? Did the (abnormal) behavior fluctuate during the day—that is, tend to come and go or increase and decrease in severity?

Feature 2. Inattention. This feature is shown by a positive response to the following question: Did the patient have difficulty focusing attention—for example, was he or she easily distracted or did he or she have difficulty keeping track of what was being said?

Feature 3. Disorganized thinking. This feature is shown by a positive response to the following question: Was the patient's thinking disorganized or incoherent, such as rambling or irrelevant conversation, unclear or illogical flow of ideas, or unpredictable switching from subject to subject?

Feature 4. Altered level of consciousness. This feature is shown by any answer other than "alert" to the following question: Overall, how would you rate this patient's level of consciousness: alert (normal), vigilant (hyperalert), lethargic (drowsy, easily aroused), stupor (difficult to arouse), or coma (unable to arouse)?

*Modified with permission from American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Washington, DC: American Psychiatric Association; 2013.
†The diagnosis of delirium requires the presence of features 1 and 2 and either 3 or 4. CAM = Confusion Assessment Method. From Inouye SK, van Dyck CH, Alessi CA, et al. Clarifying confusion: The Confusion Assessment Method. A new method for detection of delirium. *Ann Intern Med.* 1990;113:941-948. Copyright 2003, Hospital Elder Life Program, LLC. Not to be reproduced without permission.

DEFINITION

The diagnostic criteria for delirium are evolving (Table 28-1). The *Diagnostic and Statistical Manual of Mental Disorders* is based on expert consensus, but the diagnostic sensitivity and specificity of its criteria have not been tested. The Confusion Assessment Method provides a simple, operationalized diagnostic algorithm.[1] In studies of more than 1000 subjects, it had a sensitivity of 94%, a specificity of 89%, and a high interrater reliability.

EPIDEMIOLOGY

In persons older than 65 years admitted to general medical services, the prevalence of delirium at hospital admission is 18 to 35%. Delirium develops anew in an additional 11 to 14% of these patients during hospitalization. Higher rates are found when frequent surveillance is performed in older, surgical, and intensive care populations. Delirium occurs in 10 to 70% of postoperative patients, up to 80% of patients in medical intensive care units, up to 35% of nursing home patients, and at least 45% of patients at the end of life.[1]

The hospital mortality rates for delirium are 25 to 33%, as high as those associated with sepsis. The problem of delirium in hospitalized elderly patients has assumed particular prominence because patients aged 65 years and older currently account for more than 50% of all inpatient days of hospital care. Based on U.S. vital health statistics, delirium complicates hospital stays for at least 20% of the 12.5 million older persons hospitalized each year and increases hospital costs by more than $3000 per patient, amounting to more than $9 billion of Medicare expenditures yearly. Substantial additional costs are incurred after hospital discharge because of the increased need for rehabilitation services, nursing home placement, home care, and rehospitalization. Health care costs associated with delirium range from $50 billion to $200 billion per year. These extrapolations highlight the extensive economic and health policy implications of delirium.

PATHOBIOLOGY

Similar to other common geriatric syndromes (Chapter 25), delirium usually has multiple causes. A search for the innumerable potential underlying contributors requires clinical astuteness and a thorough medical evaluation, especially because many of these factors are treatable but may result in substantial morbidity and mortality if left untreated. The process is made more challenging by the frequently nonspecific, atypical, or muted features of the underlying illness in older persons. Delirium is commonly the only initial sign of an underlying life-threatening illness, such as pneumonia (Chapter 97), urosepsis (Chapter 284), or myocardial infarction (Chapter 73) in the older population.

The basic pathogenesis of delirium remains unclear. Recent evidence suggests that interacting biologic factors result in disruption of large-scale neuronal networks in the brain, thereby leading to acute cognitive disruption and delirium. Some of the leading proposed mechanisms include disruption in neurotransmitter systems, inflammation, physiologic stressors, metabolic derangements, electrolyte and acid-base disorders, and genetic factors. Many neurotransmitter systems are potentially involved, but relative cholinergic deficiency and dopamine excess are the most frequently linked. Inflammation may operate through both peripheral and central (brain) inflammatory cascades. Neuroimaging studies coupled with cognitive testing demonstrate a generalized disruption in higher cortical function, with dysfunction in the prefrontal cortex, frontal and temporoparietal cortex, fusiform cortex, lingual gyri, subcortical structures, thalamus, and basal ganglia.

The development of delirium usually involves a complex interrelationship between a vulnerable patient with pertinent predisposing factors and exposure to noxious insults or precipitating factors. Delirium may develop in vulnerable patients, such as cognitively impaired or severely ill patients, after a relatively benign insult, such as a single dose of sleeping medication. Conversely, in patients who are not vulnerable, delirium may develop only after exposure to multiple noxious insults. Previous studies have shown that the effects of these risk factors may be cumulative. Recognition of this multifactorial causation is important to the clinician because the removal or treatment of one factor in isolation usually is not sufficient to resolve the delirium. The full spectrum of vulnerability and precipitating factors should be addressed.

Factors that predispose patients to delirium include preexisting cognitive impairment or dementia, history of delirium, functional impairment, visual or hearing impairment, multiple comorbid conditions, severe underlying illness, depression, history of a stroke or transient ischemic attack, alcohol abuse, and advanced age. Dementia is an important and consistent risk factor for delirium; persons with dementia have a two-fold to five-fold increased risk for delirium, and 30 to 50% of delirious patients have underlying dementia.

Medications, the most common remediable causes of delirium, contribute to delirium in 40% of cases (Chapter 27). Insufficiency or failure of any major organ system, particularly renal or hepatic failure, can precipitate delirium. Surgical procedures are leading risk factors for delirium. Hypoxemia and hypercarbia have been associated with delirium. Clinicians must be attuned to occult respiratory failure, which in the elderly often lacks the usual signs and symptoms of dyspnea and tachypnea and can be missed by the measurement of oxygen saturation alone. Acute myocardial infarction or heart failure can be manifested as delirium in an elderly patient without the usual symptoms of chest pain or dyspnea. Occult infection is a particularly notable cause of delirium. Older patients frequently fail to mount the febrile or leukocytotic response to infection, and clinicians must assess them carefully for signs of pneumonia, urinary tract infection, endocarditis, abdominal abscess, or infected joints. A variety of metabolic disorders may contribute to delirium, including hypernatremia and hyponatremia, hypercalcemia, acid-base disorders, hypoglycemia and hyperglycemia, and thyroid or adrenal disorders. Immobilization and immobilizing devices (e.g., indwelling bladder catheters, physical restraints, bed alarms) are important factors in precipitating delirium. Dehydration and volume depletion and nutritional decline during hospitalization (e.g., weight loss, fall in serum albumin concentration) are well-documented factors contributing to delirium. Drug and alcohol withdrawal are important and often unsuspected causes of delirium in the elderly. Environmental factors, such as unfamiliar surroundings, sleep deprivation, deranged schedule, frequent room changes, sensory overload, and sensory deprivation, may aggravate delirium in the hospital. Psychosocial factors, such as depression, psychological stress, pain, and lack of social supports, also may precipitate delirium.

CLINICAL MANIFESTATIONS

The cardinal features of delirium include acute onset and inattention. Establishing the acuteness of onset requires accurate knowledge of the patient's baseline cognitive function. Patients with delirium are inattentive; that is,

they have difficulty focusing, maintaining, and shifting attention. They appear easily distracted and have difficulty in maintaining conversation and following commands. Objectively, patients may have difficulty with simple repetitive tasks, digit spans, and recitation of months backward. Other key features include disorganized thought processes, which are usually a manifestation of underlying cognitive or perceptual disturbances; altered level of consciousness, which typically consists of lethargy with reduced awareness of the environment; and fluctuation of cognitive symptoms. Although not cardinal elements, other features that frequently occur during delirium include disorientation, cognitive deficits, psychomotor agitation or retardation, perceptual disturbances such as hallucinations and illusions, paranoid delusions, and sleep-wake cycle reversal.

DIAGNOSIS

The cornerstone of the evaluation of delirium is a comprehensive history and physical examination (Table 28-2). The first step is to establish the diagnosis of delirium through cognitive assessment and to determine whether the present condition represents an acute change from the patient's baseline cognitive function, such as from a family member. Because cognitive impairment may not be apparent during conversation, brief cognitive screening tests, such as the Short Portable Mental Status Questionnaire (Table 27-4) and the Confusion Assessment Method (CAM), should be used (Video 28-1). Attention should be assessed further with other simple tests, such as a forward digit span (inattention is indicated by an inability to repeat five digits forward) or recitation of the months backward (allow maximum of one error). A delirium assessment specifically for nonverbal (e.g., intubated) patients, called the CAM-ICU, has been developed. The history, which should be obtained from a reliable informant, is targeted to establish the patient's baseline cognitive function and the time course of any mental status change and to obtain clues about potential precipitating factors, such as recent medication changes, intercurrent infection, or medical illness. Physical examination should include a detailed neurologic examination for focal deficits and a careful search for signs of occult infection or an acute abdominal process.

Review of the patient's medication list, including over-the-counter medications and herbal remedies, is crucial, and the use of medications with psychoactive effects should be discontinued or minimized whenever possible.[2] In the elderly, these medications may cause psychoactive effects even at doses and measured drug levels within the "therapeutic range." Consideration should also be given to the possibility that withdrawal from alcohol or other medications is a contributor to delirium.

TABLE 28-2 EVALUATION OF DELIRIUM IN ELDERLY PATIENTS

Perform cognitive testing and determine baseline cognitive functioning: establish the diagnosis of delirium.
Obtain a comprehensive history and perform a physical examination, including a careful neurologic examination for focal deficits and a search for occult infection.
Review the patient's medication list: discontinue or minimize all psychoactive medications; check the side effects of all medications.
Perform a laboratory evaluation (tailored to the individual): complete blood count, electrolytes, blood urea nitrogen, creatinine, glucose, calcium, phosphate, liver enzymes, oxygen saturation.
Search for occult infection: physical examination, urinalysis, chest radiography, selected cultures (as indicated).
When no obvious cause is revealed after these steps, further targeted evaluation is considered in selected patients, as follows: Laboratory tests: magnesium, thyroid function, vitamin B_{12} level, drug levels, toxicology screen, ammonia level Arterial blood gas analysis: indicated in patients with dyspnea, tachypnea, any acute pulmonary process, or history of significant respiratory disease Electrocardiography: indicated in patients with chest or abdominal discomfort, shortness of breath, or cardiac history Cerebrospinal fluid examination: indicated when meningitis or encephalitis is suspected Brain imaging: indicated in patients with new focal neurologic signs or with a history or signs of head trauma Electroencephalography: useful in diagnosis of occult seizure disorder and in differentiating delirium from nonorganic psychiatric disorders

Laboratory Findings

Laboratory evaluation must be tailored to the individual situation (Table 28-2). In patients with preexisting cardiac or respiratory diseases or related symptoms, electrocardiography or arterial blood gas determination may be indicated. The need for cerebrospinal fluid examination is controversial except for the clear indication of the febrile delirious patient, in whom meningitis or encephalitis is suspected. Brain imaging should be reserved for patients with new focal neurologic signs, those with a history or signs of head trauma (e.g., upper body bruising), and those without another identifiable cause of the delirium. Electroencephalography, with a false-negative rate of 17% and a false-positive rate of 22% for distinguishing delirious from nondelirious patients, has a limited role and is most useful for detecting an occult seizure disorder and differentiating delirium from psychiatric disorders.

Differential Diagnosis

A crucial difficulty is distinguishing a long-standing confusional state (dementia) from delirium alone or delirium superimposed on dementia (see Fig. 28-1). These two conditions are differentiated by the acute onset of symptoms in delirium (dementia is much more insidious) and the impaired attention and altered level of consciousness associated with delirium. The differential diagnosis also includes depression and nonorganic psychotic disorders. Although paranoia, hallucinations, and affective changes can occur with delirium, the key features of acute onset, inattention, altered level of consciousness, and global cognitive impairment assist in the recognition of delirium. At times, the differential diagnosis can be difficult, particularly with an uncooperative patient or when an accurate history is unavailable. Because of the potentially life-threatening nature of delirium, it is prudent to manage the patient as if he or she has delirium and to search for and treat underlying precipitants (e.g., intercurrent illness, metabolic derangement, drug toxicity) until further information can be obtained.

TREATMENT Rx

Prevention

The most effective strategy to reduce delirium and its associated complications is primary prevention before delirium occurs. Preventive strategies should address important risk factors and target moderate- to high-risk patients at baseline (Table 28-3). Clinical trials document that multicomponent nonpharmacologic interventions targeted toward delirium risk factors can reduce the incidence of delirium by 30 to 40%.[1] These components typically include strategies designed to improve orientation, to provide therapeutic activities, to increase mobilization and exercise, to enhance sleep with nonpharmacologic interventions, to optimize vision and hearing, and to manage dehydration. Preoperative geriatric consultation and lighter anesthesia also can reduce postoperative delirium. By comparison, no drug treatments, including antipsychotics or cholinesterase inhibitors, have been consistently effective in preventing delirium. Preventive efforts require system-wide changes to educate physicians and nurses, to improve their recognition of delirium and heighten their awareness of its clinical implications, to provide incentives to change practice patterns that lead to delirium (e.g., immobilization, sleep medications, bladder catheters, physical restraints), and to create systems that enhance high-quality geriatric care (e.g., geriatric expertise, case management, clinical pathways, quality monitoring).

Medical Therapy

In general, nonpharmacologic approaches should be used in all delirious patients, and these are usually successful in managing symptoms. Pharmacologic approaches should be reserved for patients whose symptoms may result in the interruption of needed medical therapies (e.g., intubation, intravenous lines) or may endanger the safety of the patient or other persons. No drug is ideal for the treatment of delirium, however; any drug can cloud the patient's mental status further and obscure efforts to monitor the course of the mental status change. The drug should be given at the lowest dose and for the shortest time possible. Neuroleptics are the preferred agents. Haloperidol, the most widely used agent, causes less orthostatic hypotension and fewer anticholinergic side effects than thioridazine and is available in parenteral form; however, it has a higher rate of extrapyramidal side effects and acute dystonias. Second-generation antipsychotics have not proved superior to haloperidol. If parenteral administration is required, intravenous haloperidol should be administered in a monitored setting because its use results in a rapid onset of action, short duration of effect, and risk of hypotension and torsades de pointes; oral and intramuscular administration has a more optimal duration of action, and these are the preferred routes. The recommended starting dose is 0.25 to 0.5 mg of haloperidol orally or intramuscularly, repeated every 30

VIDEO 28-1. Confusion Assessment Method

TABLE 28-3 *NICE* GUIDELINES FOR PREVENTION OF DELIRIUM

CLINICAL FACTOR	RECOMMENDED PREVENTIVE INTERVENTIONS*
Cognitive impairment or disorientation	Provide orienting cues: calendars, clocks, photos. Reorient the patient to time, place, person, schedule. Provide cognitively stimulating activities, like reminiscence. Encourage regular visits from family and friends.
Dehydration or constipation	Encourage patients to drink fluids. Advise team about use of parenteral fluids if necessary. Carefully monitor fluid balance in patients with heart failure or renal disease.
Hypoxia	Assess for symptoms of hypoxia and monitor oxygen saturation levels.
Immobility or limited mobility	Encourage early mobilization and regular ambulation. Keep walking aids (canes, walkers) readily available at all times. Encourage all patients to conduct active range-of-motion exercises.
Infection	Look for and treat infection. Avoid unnecessary catheterization. Implement infection control procedures.
Multiple medications	Review medication for both the type and the number of medications. Monitor for potential interactions.
Pain	Assesses for pain, especially in those who have communication difficulties. Begin and monitor pain management in those with known or suspected pain.
Poor nutrition	Follow general nutrition guidelines and seek input from dietitian early on if needed. Ensure availability and proper fit of dentures.
Sensory impairment	Resolve reversible causes or contributors to impairment. Ensure that working hearing and visual aids are available and used by those who need them. Educate staff.
Sleep disturbance	Avoid medical/nursing procedures during sleep if possible. Schedule medications and procedures to avoid disturbing sleep. Reduce noise level at night unit-wide. Educate staff.

*Operationalized protocols for all recommended preventive interventions are available at www.hospitalelderlifeprogram.org.
Modified with permission from National Institute for Health and Care Excellence (NICE) Clinical Guideline 103. www.nice.org.uk/cg103.

minutes after the vital signs have been rechecked, until sedation has been achieved. The end point should be an awake but manageable patient. The average elderly patient who has not been treated previously with neuroleptics should receive no more than 3 to 5 mg of haloperidol in a 24-hour period. Subsequently, a maintenance dose consisting of half the loading dose should be administered in divided doses during the next 24 hours, with doses tapered during the next few days as the agitation resolves.

Benzodiazepines are not recommended as the first-line treatment of delirium because of their tendency to cause oversedation and to exacerbate the confusional state. They remain the drugs of choice, however, for the treatment of withdrawal syndromes from alcohol and sedative drugs (Chapters 33 and 34).

Nonpharmacologic Management

Multicomponent geriatric interventions may be effective in improving quality of life with the potential for significant cost savings.[1] Nonpharmacologic management techniques recommended for every delirious patient include encouraging the presence of family members, using "sitters" as orienting influences, and transferring a disruptive patient to a private room or closer to the nurses' station for increased supervision. Interpersonal contact and communication, including verbal reorientation strategies, simple instructions and explanations, and frequent eye contact, are vital. Patients should be involved in their own care and allowed to participate in decision making as much as possible. Eyeglasses and hearing aids may reduce sensory deficits. Mobility, self-care, and independence should be encouraged, and physical restraints and bed alarms should be avoided, if possible, because of their tendency to increase agitation, their lack of efficacy, and their potential to cause injury. Attention must be focused on minimizing the disruptive influences of the hospital environment. Clocks and calendars should be provided to assist with orientation. Room and staff changes should be kept to a minimum. A quiet environment with low-level lighting is optimal for delirious patients, and use of ear plugs may be helpful for management of delirium.[A4] Perhaps the most important intervention is to schedule the checking of vital signs, the administration of medications, and the performance of procedures to allow the patient's uninterrupted sleep at night.[3] Nonpharmacologic approaches to relaxation, including music, relaxation tapes, and massage, can be highly effective in managing agitation.

End-of-Life Care

Delirium occurs in at least 80% of patients at the end of life and is considered part of the dying process by many hospice care providers (Chapter 3). Establishing the goals for care in advance with the patient and family is critical to guide appropriate management. For example, some patients may prioritize the preservation of alertness and the ability to communicate with loved ones as long as possible; others may prioritize comfort above all else. Physicians must be aware that even in terminal patients, many causes of delirium are potentially reversible with simple interventions such as adjusting medications, providing oxygen, or treating dehydration; however, aggressive diagnostic evaluation is usually inappropriate in this population. Nonpharmacologic measures to treat agitation and delirium should be instituted in all patients (including massage, music, and relaxation therapies). Haloperidol remains the first-line therapy for delirium in terminally ill patients. If more sedation is indicated, a short-acting benzodiazepine such as lorazepam (starting dose, 0.5 to 1.0 mg PO, IM, or SL), which is easily titrated, is recommended in this setting. Because sedation may result in decreased interaction and communication, increased confusion, and respiratory depression, this choice should be made in conjunction with the family while honoring the patient's preferences.

PROGNOSIS

Delirium increases the risk of death about 2-fold, of institutionalization about 2.5-fold, and of dementia more than 12-fold[4] even after controlling for age, sex, severity of illness, comorbid conditions, and baseline dementia. Delirium also prolongs hospital stays and increases health care costs.

Delirium was previously considered a reversible, transient condition, but symptoms typically last for 30 days or more, only 20% of patients have complete resolution by 6 months, and its detrimental effects may persist at 1 year.[5] Delirium has even more pronounced effects in patients with underlying dementia. The rate of cognitive decline in patients with dementia more than doubles after an episode of delirium,[6] and about one in eight hospitalized patients with dementia who develop delirium will have at least one severe adverse outcome, including a five-fold increased risk of death and a nine-fold increased risk of institutionalization.[7] These studies suggest that strategies to prevent delirium in persons with dementia should be a priority to prevent future cognitive deterioration and adverse outcomes. The long-term detrimental effects are most likely related to the duration, severity, and underlying cause of the delirium and the vulnerability of the patients.

FUTURE DIRECTIONS

Because delirium is common, frequently iatrogenic, and linked to processes of care, it is a marker of the quality of care and safety of the patient in the hospital setting. Research is needed to elucidate the pathophysiologic mechanisms of delirium by the use of neuroimaging modalities, neuropsychological testing, and genetic and laboratory markers; to clarify the contribution of delirium to irreversible cognitive impairment and dementia; and to improve the evidence-based management of delirium.

Grade A References

A1. O'Mahony R, Murthy L, Akunne A, et al. Synopsis of the National Institute for Health and Clinical Excellence guideline for prevention of delirium. *Ann Intern Med*. 2011;154:746-751.
A2. Deschodt M, Braes T, Flamaing J, et al. Preventing delirium in older adults with recent hip fracture through multidisciplinary geriatric consultation. *J Am Geriatr Soc*. 2012;60:733-739.
A3. Moyce Z, Rodseth RN, Biccard BM. The efficacy of peri-operative interventions to decrease post-operative delirium in non-cardiac surgery: a systematic review and meta-analysis. *Anaesthesia*. 2014;69:259-269.
A4. Van Rompaey B, Elseviers MM, Van Drom W, et al. The effect of earplugs during the night on the onset of delirium and sleep perception: a randomized controlled trial in intensive care patients. *Crit Care*. 2012;16:R73.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. *Lancet*. 2014;383:911-922.
2. American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society Updated Beers Criteria for potentially inappropriate medication use in older adults. *J Am Geriatr Soc*. 2012;60:616-631.
3. Yoder JC, Yuen TC, Churpek MM, et al. A prospective study of nighttime vital sign monitoring frequency and risk of clinical deterioration. *JAMA Intern Med*. 2013;173:1554-1555.
4. Witlox J, Eurelings LS, de Jonghe JF, et al. Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. *JAMA*. 2010;304:443-451.
5. Saczynski JS, Marcantonio ER, Quach L, et al. Cognitive trajectories after post-operative delirium. *N Engl J Med*. 2012;367:30-39.
6. Gross AL, Jones RN, Habtemariam D, et al. Delirium and long-term cognitive trajectory among persons with dementia. *Arch Intern Med*. 2012;172:1324-1331.
7. Fong TG, Jones RN, Marcantonio ER, et al. Adverse outcomes after hospitalization and delirium in persons with Alzheimer Disease. *Ann Intern Med*. 2012;156:848-856.

REVIEW QUESTIONS

1. An 85-year-old woman with a history of hypertension, diabetes, transient ischemic attacks, prior cerebrovascular event without residua, hearing loss, and remote breast cancer is brought in to your office by her daughter for increasing memory lapses. The patient lives alone, but her daughter is nearby and very attentive. What would be the most appropriate next steps in the evaluation?

A. Recent history and urgent brain MRI scan
B. Recent history, cognitive screening, and neurologic examination
C. Recent history, neurologic examination, and brain MRI scan
D. Chest and head CT scans for metastatic disease
E. Recent history, carotid ultrasound, and brain MRI scan

Answer: B The first step is to establish a diagnosis of the memory problem, which requires a history of the mental status change from the daughter, a careful neurologic examination for focal changes, and cognitive screening tests. If the changes are acute, with inattention and cognitive dysfunction on testing, delirium must be excluded. If the changes are chronic with memory impairment on cognitive testing, dementia is more likely. Neuroimaging tests would not be appropriate first steps in the evaluation.

2. You obtain the recent history and conduct cognitive screening of the patient in question 1. The daughter indicates to you that the patient has long-standing short-term memory problems but has been having more difficulty with her regular activities (like forgetting items at the grocery store and getting lost while driving) and has had episodes of disorientation during the past 1 to 2 weeks; at times, she has not made sense on the telephone. On your examination, you find that her neurologic examination is nonfocal, and she makes six errors on the Short Portable Mental Status Questionnaire. She has difficulty in following instructions and gives nonsense answers to some questions. She is disoriented to date, month, and year; she makes five errors on months of the year backwards. What are the key aspects of the history and evaluation that support a diagnosis of delirium?

A. Disorientation, memory loss, and functional impairments (driving, shopping)
B. Disorientation, memory loss, and inattention
C. Acute onset, disorientation, memory loss
D. Acute onset, inattention, and evidence of disorganized thinking
E. Acute onset, inattention, and functional impairments (driving, shopping)

Answer: D The key features of a delirium include the acute onset, inattention (in this case manifested by her difficulty in following instructions and errors on the sustained attention task of months of the year backwards), and evidence of disorganized thinking (in this case manifested by the errors on the mental status questionnaire and her nonsense answers to questions). Acute onset of disorientation and the functional impairments are supportive features but not key diagnostic elements. The memory impairment has been chronic, so it is not helpful diagnostically in this case. Formal cognitive screening is key in establishing the diagnosis of delirium.

3. You are called as the medical consultant to evaluate JS, an 89-year-old retired accountant admitted to the urology service for removal of a benign ureteral tumor, because of a recent history of falling and weight loss. He is ready for discharge, but the family is concerned about his "failure to thrive." He has a history of chronic renal insufficiency, hypertension, intermittent atrial fibrillation, diastolic dysfunction, previous transient ischemic attacks, positive response to PPD, legal blindness due to macular degeneration, osteoarthritis, chronic reflux esophagitis, depression, and mild cognitive impairment. He smokes about three cigars per day and drinks two or three glasses of wine with dinner each evening. His current daily medications include aspirin, famotidine, paroxetine, metoprolol, amiodarone, amlodipine, naproxen, clopidogrel, and a multivitamin with minerals. He takes furosemide about three times a week for ankle edema. His daughter tells you that during the past month, his appetite has been poor in the context of severe heartburn, and he has become increasingly withdrawn and refuses to eat. He has had two falls in the past week and one fall in the hospital last night. In addition, she has found him wandering and disoriented several times during the past week. On your examination, he has temporal wasting and bruised areas on his shoulders, but his neurologic examination is nonfocal. He is disoriented to place, date, day, and year; he makes three errors on serial sevens. He reports seeing butterflies on the bed. He is unable to cooperate with the full mental status testing, stating that he has to get to the bank immediately. What is the *most likely* factor contributing to this patient's delirium?

A. Malnutrition and dehydration
B. Postoperative status
C. Multifactorial etiology
D. Polypharmacy
E. Multimorbidity and depression

Answer: C This is a classic example of delirium resulting from multifactorial etiology in an older person, that is, the complex interplay between baseline, predisposing (vulnerability) factors and precipitating (noxious) factors. Although each of the other answers is likely to be contributing to some degree, the fundamental principle is that delirium is typically of multifactorial etiology in the elderly.

4. You have completed your evaluation of JS (question 3) and made your recommendations to the urologist, including recommendations for a metabolic profile, which reveals a sodium level of 126 mg/dL. The urologist insists that the patient is ready to be discharged and recommends that any further evaluation be completed on an outpatient basis. What are the most appropriate management strategies for the patient at this point?

A. Discharge from the hospital with completion of the evaluation on an outpatient basis by the primary care physician
B. Discharge from the hospital with completion of evaluation on an outpatient basis by the geriatric assessment clinic
C. Keep patient on the urology service and attempt to complete the evaluation and to institute a management plan
D. Obtain a social work consultation for post-acute placement
E. Transfer to the medical service to complete the evaluation and to institute a management plan

Answer: E Because delirium can be a medical emergency, it is important to complete the evaluation in the inpatient setting on a timely basis. Given the expertise needed, this goal may be best achieved on the medical service. The most important next steps are to ensure the patient's safety and to prevent further falls, to evaluate the patient for possible significant head trauma with neuroimaging, to correct hyponatremia and other metabolic derangements, to eliminate or to reduce nonessential psychoactive medications, and to provide appropriate hydration and nutrition. It would not be appropriate at this point to discharge the patient or to transfer him to a post-acute setting.

5. Before JS (case in question 3) was admitted for surgery, what were the *most appropriate* strategies that would have been helpful to prevent delirium?

A. Treatment with oral low-dose haloperidol before, during, and for 3 days after surgery
B. Treatment with oral olanzapine before surgery and sublingual olanzapine after surgery for 48 hours
C. Treatment with donepezil beginning 2 to 4 weeks before surgery and continuing for 1 month postoperatively
D. Multicomponent nonpharmacologic prevention strategies to address orientation, mobility, medications, vision and hearing impairments, nutrition, and hydration
E. Interdisciplinary geriatric consultation beginning on the first postoperative day to address functional and cognitive status

Answer: D To date, there is no grade A evidence that any pharmacologic strategies are effective for preventing delirium. The current recommendation is to use multicomponent nonpharmacologic interventions to address delirium risk factors, such as the Hospital Elder Life Program (HELP) model (which allows fulfillment of all of the NICE guidelines for delirium prevention). Proactive delirium consultation is effective if it is started preoperatively.

29

PRINCIPLES OF DRUG THERAPY

ROBERT B. DIASIO

Under different conditions, a drug may produce diverse effects ranging from no effect to a desirable effect to an undesirable toxic effect. Physicians must learn how to choose the correct drug dosage for different conditions to ensure effective and safe therapy. This necessitates understanding the pharmacokinetics—the movement of a drug over time through the body—and the pharmacodynamics—the relationship between drug concentration and drug effect (Fig. 29-1). This chapter reviews the basic concepts of pharmacokinetics and pharmacodynamics, followed by guidelines on how to use this information to optimize therapeutic applications. Drug interactions and adverse drug responses are briefly discussed, with advice on how both can be recognized and minimized in clinical practice. Lastly, the increasing role of pharmacogenomics in individualizing therapy beyond preventing or predicting adverse drug responses is discussed.

PHARMACOKINETIC PRINCIPLES

Administration

The most efficient and straightforward means of administering a drug into the systemic circulation is by intravenous injection of the drug as a bolus. With this route, the full amount of a drug is delivered to the systemic circulation almost immediately. The same dose also may be administered as an intravenous infusion over a longer period, resulting in a decrease in the peak plasma concentration and an accompanying increase in the time the drug is present in the circulation. Many other routes of administration can be used, including sublingual, oral, transdermal, rectal, inhalational, subcutaneous, and intramuscular; each of these routes carries not only a potential delay in the time it takes the drug to enter the circulation but also the possibility that a large fraction of it will never reach the circulation.

Absorption

Absorption refers to the transfer of a drug from the site of administration to the systemic circulation. Many drugs cross a membrane barrier by passive diffusion and enter the systemic circulation. Because passive diffusion in this setting depends on the concentration of the solute at the membrane surface, the rate of drug absorption is affected by the concentration of free drug at the absorbing surface. Factors that influence the availability of free drug thus affect drug absorption from the administration site; this effect can be exploited to design medications that release a drug slowly into the circulation by prolonging drug absorption. With certain sustained-released oral preparations, the rate of dissolution of the drug in the gastrointestinal tract determines the rate at which the drug is absorbed (e.g., timed-release antihistamines). Similarly, a prolonged drug effect can be obtained by the use of transdermal medications (e.g., nitroglycerin) or intramuscular depot preparations (e.g., benzathine penicillin G).

First-Pass Effect

Some drugs that are administered orally are absorbed relatively efficiently into the portal circulation but are metabolized by the liver before they reach the systemic circulation. Because of this "first-pass" or "presystemic" effect, the oral route may be less suitable than other routes of administration for such drugs. A good example is nitroglycerin, which is well absorbed but efficiently metabolized during the first pass through the liver. However, the same drug can achieve adequate systemic levels when it is given sublingually or transdermally.

Bioavailability

The extent of absorption of a drug into the systemic circulation may be incomplete. The bioavailability of a particular drug is the fraction (F) of the total drug dose that ultimately reaches the systemic circulation from the site of administration. This fraction is calculated by dividing the amount of the drug dose that reaches the circulation from the administration site by the amount of the drug dose that would enter the systemic circulation after direct intravenous injection into the circulation (essentially the total dose). Bioavailability, or F, can range from 0, in which no drug reaches the systemic circulation, to 1.0, in which essentially all of the drug is absorbed. The bioavailability of a drug may vary in different formulations because the overall absorption differs. This variability has become a concern with the increasing use of generic preparations.

Distribution

After delivery of a drug into the systemic circulation either directly by intravenous injection or after absorption, the drug is transported throughout the body, initially to the well-perfused tissues and later to areas that are less perfused. The distribution phase can be assessed best by plotting the drug's plasma concentration on a log scale versus time on a linear scale (Fig. 29-2). For an intravenously administered drug, when absorption is not a factor, the initial phase—from immediately after administration through the rapid fall in concentration—represents the distribution phase, during which a drug rapidly disappears from the circulation and enters the tissues. This is followed by the elimination phase (see later), when drug in the plasma equilibrates with drug in the tissues. During this latter phase, the drug's plasma concentration is thought to be related to drug effect.

Volume of Distribution

The volume of distribution (VD) relates the amount of drug in the body to the concentration of drug in the plasma. It is calculated by dividing the dose that ultimately gets into the systemic circulation by the plasma concentration at time zero (C_{p0}):

$$VD = dose / C_{p0} \quad (1)$$

The C_{p0} can be calculated by extrapolating the elimination phase back to time zero (see Fig. 29-2). The VD is best considered the "apparent VD" because it represents the apparent volume needed to contain the entire amount of the drug, assuming it is distributed throughout the body at the same concentration as in the plasma. Table 29-1 lists pharmacokinetic data for commonly used drugs from several drug classes, showing the wide variation in VD. Digoxin, for example, has a large VD (>5 L), whereas glimepiride has a relatively small VD (0.18 L). As discussed later, VD is a useful pharmacokinetic tool for calculating the loading dose and appreciating how various changes can affect a drug's half-life.

Elimination

Drugs are removed from the body by two major mechanisms: hepatic elimination, in which drugs are metabolized in the liver and excreted through the biliary tract; and renal elimination, in which drugs are removed from the circulation by either glomerular filtration or tubular secretion. For most drugs, the rates of hepatic and renal elimination are proportional to the plasma concentration of the drug. This relationship is often described as a "first-order" process. Two measurements, clearance and half-life, are used to evaluate elimination.

Clearance

The efficiency of elimination can be assessed by quantifying how fast the drug is cleared from the circulation. Drug clearance is a measure of the volume of plasma cleared of drug per unit of time. It is similar to the clinical measurement used to assess renal function—creatinine clearance, which is the volume of plasma from which creatinine is removed per minute. Total drug clearance (Cl_{tot}) is the rate of elimination by all processes (El_{tot}) divided by the plasma concentration of the drug (C_p):

$$Cl_{tot} = El_{tot} / C_p \quad (2)$$

Drugs may be cleared by several organs, but as noted earlier, renal clearance and hepatic clearance are the two major mechanisms. Total drug clearance (Cl_{tot}) can best be described as the sum of clearances by each organ. For most drugs, this is essentially the sum of renal clearance and hepatic clearance:

$$Cl_{tot} = El_{Ren} + Cl_{Hep} \quad (3)$$

Table 29-1 shows the wide variation in clearance values among commonly used medications; some drugs (e.g., phenobarbital) have relatively low clearances (<5 mL/minute), and other drugs (e.g., aspirin) have relatively high clearances (>500 mL/minute). Tobramycin is cleared almost entirely by the kidneys, whereas aspirin, carbamazepine, and phenytoin are cleared less than 5% by the kidneys.

Drug clearance is affected by several factors, including blood flow through the organ of clearance, protein binding to the drug, and activity of the clearance processes in the organs of elimination (e.g., glomerular filtration rate and

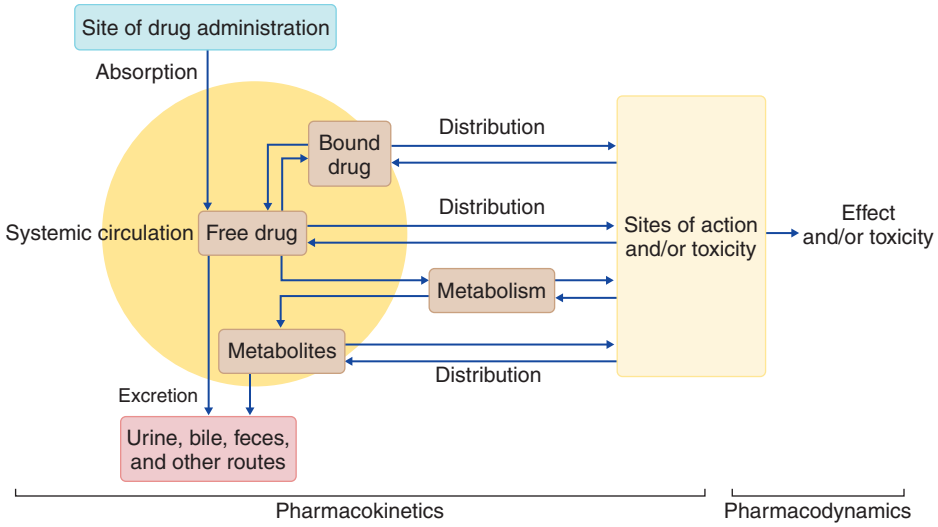

FIGURE 29-1. Schematic of a drug's movement through the body, from the site of administration to production of a drug effect. The relationship between pharmacokinetics and pharmacodynamics is shown.

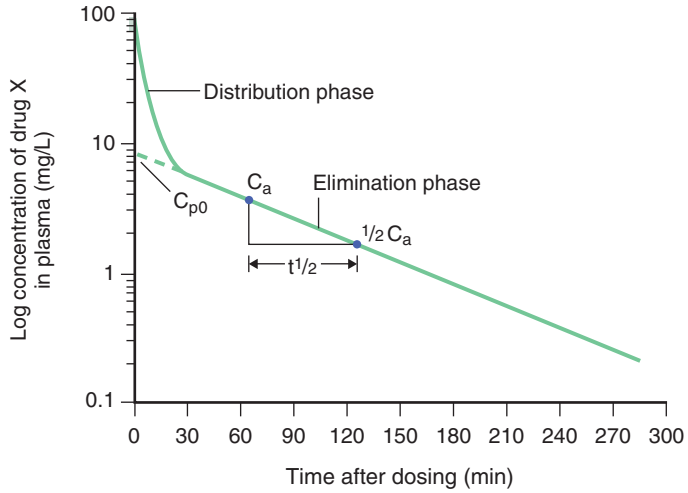

FIGURE 29-2. Representative drug concentration versus time plot used in pharmacokinetic studies. Concentration of drug is plotted with a logarithmic scale on the ordinate, and time is plotted with a linear scale on the abscissa. The resultant curve has two phases: the distribution phase, which is the initial portion of the plotted line when the concentration of drug decreases rapidly; and the later elimination phase, during which there is an exponential disappearance of drug from the plasma over time. The *dotted line* extrapolated from the elimination phase back to time zero is used to calculate plasma concentration at time zero (C_{p0}). During the elimination phase, the half-life (t½) can be calculated as the time it takes to decrease the concentration by half (shown here as the time needed to decrease from concentration C_a to ½ C_a).

tubular secretion in the kidney, enzyme activity in the liver). Drug clearance is not affected by the distribution of drug throughout the body (VD) because clearance mechanisms act only on drug in the circulation.

Half-Life

The amount of time needed to eliminate a drug from the body depends on the clearance and the VD. The first-order elimination constant (K_e) represents the proportion of the apparent VD that is cleared of drug per unit of time during the drug's exponential disappearance from the plasma over time (elimination phase):

$$K_e = \mathrm{Cl}/\mathrm{VD} \qquad (4)$$

The value of this constant for a particular drug can be determined by plotting drug concentration versus time on a log-linear plot (see Fig. 29-2) and measuring the slope of the straight line obtained during the exponential (elimination) phase.

The time needed to eliminate the drug is best described by its half-life ($t_{1/2}$), which is the time required during the elimination phase (see Fig. 29-2) for the plasma concentration of the drug to be decreased by half. Mathematically, the half-life is equal to the natural logarithm of 2 (representing a reduction of drug concentration to half) divided by K_e. Substituting for K_e from Equation 4 and calculating the natural logarithm of 2, the half-life can be represented by the following equation:

$$t_{1/2} = 0.693\ \mathrm{VD}/\mathrm{Cl} \qquad (5)$$

From this equation, one can predict that at a given clearance, as the VD increases, the half-life increases. Similarly, at a given VD, as the clearance increases, the half-life decreases. Clinically, many disease states (see later) can affect VD and clearance. Because disease affects VD and clearance differently, the half-life may increase, decrease, or not change much at all. Therefore, the half-life by itself is not a good indicator of the extent of abnormality in elimination.

The half-life is useful to predict how long it takes for a drug to be eliminated from the body. For any drug that has a first-order elimination, one would expect that by the end of the first half-life, the drug would be reduced to 50%; by the end of the second half-life, to 25%; by the end of the third half-life, to 12.5%; by the end of the fourth half-life, to 6.25%; and by the end of the fifth half-life, to 3.125%. In general, a drug can be considered essentially eliminated after three to five half-lives, when less than 10% of the effective concentration remains. Table 29-1 shows the wide variation in half-life for several commonly used drugs.

CLINICAL APPLICATION OF PHARMACOKINETIC PRINCIPLES

Using a Loading Dose

To attain a desired therapeutic concentration rapidly, a loading dose is often used. In determining the amount of drug to be given, the clinician must consider the "volume" within the body into which the drug will be distributed. This volume is best described by the apparent VD. The loading dose can be calculated by multiplying the desired concentration by the VD:

$$\text{Loading dose} = \text{desired concentration} \times \mathrm{VD} \qquad (6)$$

Rapid administration of the entire loading dose may produce an initially high peak concentration that results in toxicity. This problem can be avoided either by administering the loading dose as a divided dose or by varying the rate of access to the circulation, such as by administering the drug as an infusion (with an intravenous drug) or by taking advantage of the slower access to the circulation from various other routes (e.g., oral dosing). This approach is illustrated by phenytoin (see Table 29-1), which may need to be administered with a loading dose to achieve a therapeutic level (10 to 20 mg/L) rapidly. Because the VD for phenytoin is approximately 0.6 L/kg, the loading dose calculated from Equation 6 is 420 mg/L to attain a minimally

TABLE 29-1 PHARMACOKINETIC PARAMETERS FOR SOME COMMONLY USED DRUGS

DRUG	VD (L/kg)	PROTEIN BINDING (%)	TOTAL CLEARANCE (mL/min)	% OF TOTAL CLEARANCE AS RENAL CLEARANCE	HALF-LIFE (hr)	THERAPEUTIC RANGE (mg/L)
Amoxicillin	0.47	17-18		86	1.2	2-8
Aspirin (acetylsalicylic acid)	0.14-0.18	80-90	575-725	<2	0.2-0.3	20-250
Carbamazepine	1.2	75-90	50-125	1-3	12-17	4-12
Digoxin	5-7.3	20-30	75	50-70	34-44	0.5-2
Glimepiride	0.18	>99.5	0.62 ± 0.26	<0.5	3.4 ± 2.0	
Lidocaine	3	60-80	700	<10	1.5-2	1-5
Lithium carbonate	0.7-1	0	20-40	95-99	20-270	0.4-1.4*
Penicillin G	0.5-0.7	45-68	—	20	0.4-0.9	Variable
Phenobarbital	0.6-0.7	20-45	4	25	2-6 days	<10-40*
Phenytoin	0.4-0.8	88-93	—	<5	7-26	10-20
Procainamide	2.2	14-23	470-600	40-70	2.5-4.7	4-8
Theophylline	0.3-0.7	60	36-50	<10	4-16	5-20
Tobramycin	0.25-0.30	<10	70	>95	2-4	0.5-2 (TR) 4-8 (PK)
Vancomycin	0.4-1	52-60	65	85	4-6	5-10 (TR) 25-35 (PK)

*Therapeutic range varies according to the indication for the drug. For example, lithium carbonate in the range of 0.4-1.3 mg/L is appropriate for affective schizophrenia disorder; a range of 1.0-1.4 mg/L is appropriate for mania. Phenobarbital concentration below 10 mg/mL is appropriate for anticonvulsant therapy; 40 mg/L is appropriate as a hypnotic.
PK = peak value; TR = trough value; VD = volume of distribution.

therapeutic level of 10 mg/L in a 70-kg adult. However, administration of 420 mg of phenytoin by intravenous bolus carries the risk for cardiac arrest and death. By taking advantage of the reduced bioavailability (F = 0.8) and slow absorption of oral phenytoin, the loading dose can be administered safely as an oral dose of 500 mg.

The equation for the loading dose can also be used to calculate the dose needed to "boost" an inadequate blood level of drug to a desired therapeutic range. If therapeutic monitoring (see later under Drug Monitoring as a Guide to Therapy) shows that the phenytoin level is 5 mg/L and the desired level is 15 mg/L, it is necessary to multiply the difference needed to achieve the desired concentration (10 mg/L) by the VD (0.6 L/kg) to determine the dose (in mg/kg) necessary to achieve this drug level after distribution. In a 70-kg individual, 0.6 mg/kg is multiplied by 70 kg to obtain the calculated loading dose (420 mg) that can be administered safely. A 500-mg oral dose with a bioavailability of less than 1 (e.g., F = 0.8) would deliver to the systemic circulation the approximate amount needed and avoid the risks associated with rapid intravenous administration.

Determining Drug Accumulation

Continuing to administer a drug, either as a prolonged infusion or as repeated doses, results in accumulation until a steady state occurs. Steady state is the point at which the amount of drug being administered equals the amount being eliminated so that the plasma and tissue levels remain constant. The elimination half-life determines not only the time course of drug elimination but also the time course of drug accumulation. This mirror-image pattern of drug accumulation and elimination is illustrated in Figure 29-3. As with drug elimination, three to five half-lives determine the time it takes to reach steady state during drug accumulation. Whereas drugs with short half-lives accumulate rapidly, drugs with long half-lives require a longer time to accumulate, with a potential delay in achieving therapeutic levels. For drugs with long half-lives, a loading dose may be needed to obtain rapid drug accumulation and a more rapid therapeutic effect.

With each change in drug dose or rate of infusion, a change in steady state occurs. Although it is not obvious for drugs with short half-lives, the effects of dose adjustments for drugs with longer half-lives are delayed, and the time varies directly with the drug's half-life.

Using a Maintenance Dose

After steady state is reached in three to five half-lives with either a continuous infusion or intermittent doses, the rate of drug administered equals the rate of drug eliminated. For an intravenous drug, the administration rate is the infusion rate (I); for a drug administered by another route (e.g., orally), the administration rate is the dose per unit of time (D/t). Equation 7 shows

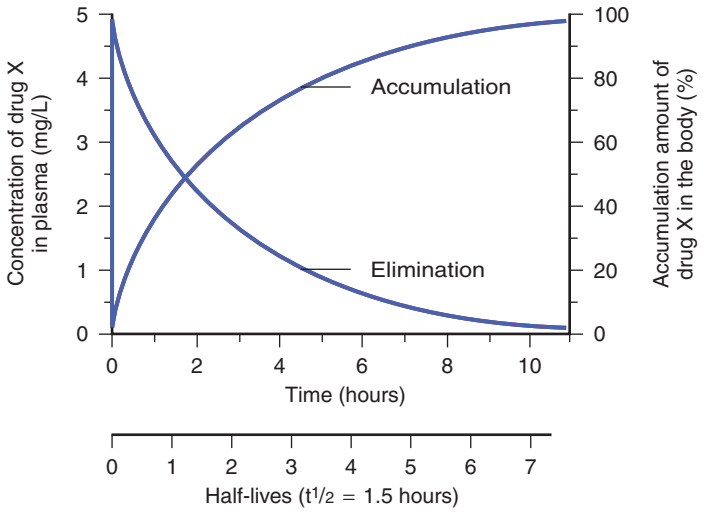

FIGURE 29-3. Representative plot of the mirror-image relationship between the elimination of drug (after drug is discontinued) and the accumulation of drug (during infusion). The plot shows the concentration on the left *y*-axis and time on the upper *x*-axis. The lower *x*-axis shows the time in half-lives, and the *y*-axis on the right shows the percentage of drug in the body. After three to five half-lives, elimination is essentially complete, and accumulation is essentially at a steady state.

that the rate of elimination (total) equals $Cl_{tot} \times C_p$. With an intravenously administered drug, because the infusion rate equals the elimination rate at steady state, it follows that

$$I = Cl_{tot} \times C_p \quad (7)$$

Similarly, with an orally administered drug, the dose administered per unit of time equals the elimination rate at steady state, with the result that

$$D/t = Cl_{tot} \times C_p \quad (8)$$

These equations show the direct relationship between the dose and the resultant plasma concentration at steady state. This relationship is independent of the distribution of the drug. By use of these equations, it is possible to determine the infusion rate or the interval and dose needed to achieve and maintain a specified drug concentration in the plasma.

When a drug is administered intermittently, it approaches steady-state concentration over time, with a pattern similar to that observed with continuous infusion (Fig. 29-4). With intermittent drug administration, such as with an

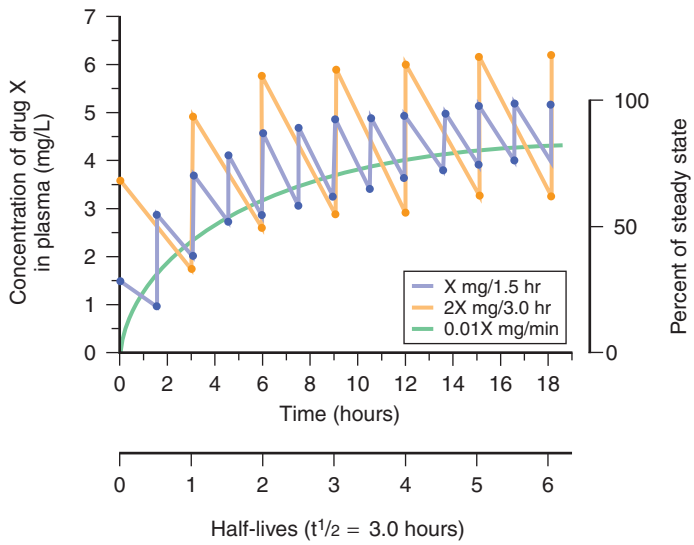

FIGURE 29-4. Accumulation of drug over time, approaching a steady state. Time is depicted in hours (upper *x*-axis) and half-lives (lower *x*-axis, showing that steady state is reached in three to five half-lives). The *green line* depicts the pattern produced by an infusion of a hypothetical drug at a dose of 0.01X. The *orange line* shows the pattern resulting from oral administration of a 2X dose every 3 hours, and the *blue line* represents the pattern produced by oral administration of dose X every 1.5 hours.

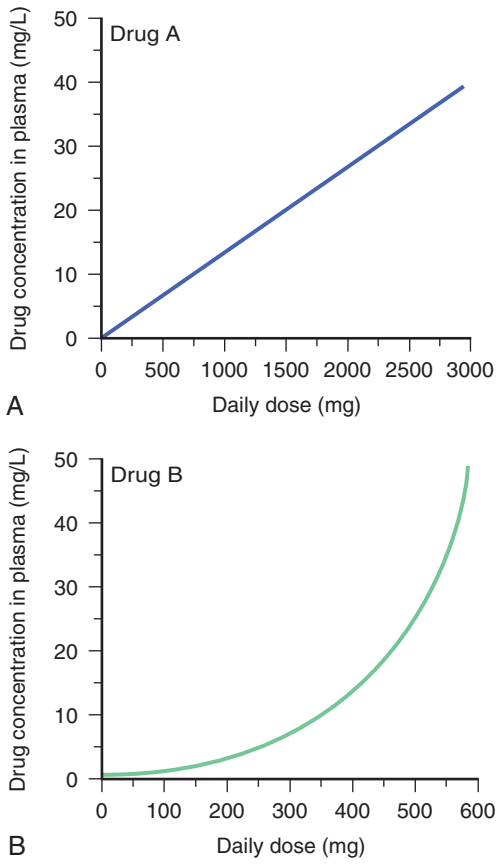

FIGURE 29-5. Effect of increasing the dose of a drug on its serum concentration. A, Drug A follows first-order or linear kinetics. B, Drug B follows zero-order or nonlinear (or saturable) kinetics.

oral dose, the drug concentration fluctuates; the magnitude of fluctuation between the peak and trough concentrations depends on the interval of administration, drug half-life, absorption characteristics, and site of administration. The effect of a change in the interval of administration for an oral drug is shown in Figure 29-4. As the intervals decrease below the half-life, the fluctuation decreases and approaches the curve produced by an intravenous infusion. Orally administered drugs may reach the blood stream more rapidly, attaining a higher peak concentration with one formulation, whereas the same drug administered as a timed-release formulation is absorbed more slowly, with a lower peak concentration but lasting longer in the plasma. Finally, the same drug administered by different routes may have different plasma profiles not only because of differing absorption characteristics but also because of other effects, such as first-pass metabolism.

Decreasing the Drug Level

At times, it may be necessary to decrease the plasma drug level while maintaining therapy (e.g., when signs of toxicity become apparent or a potentially dangerously high concentration of drug is noted; see later). The most effective and rapid response is to discontinue the drug; the length of time for which the drug is discontinued is determined by the estimated half-life of the drug in the specific patient. After discontinuation of the drug for a time based on its half-life, the total clearance (Cl_{tot}) of the drug can be used to determine what infusion rate (I, Equation 7) or dose and interval (D/t, Equation 8) must be used to achieve the new desired concentration (C_p).

Effect of Dose Increases on Elimination Kinetics

Although the previously discussed pharmacokinetic principles can be a guide to the dose of most drugs, not all drugs behave the same when the dose is increased. The elimination of most drugs follows first-order or linear kinetics; the amount of drug eliminated is directly proportional to the concentration of drug in the plasma (Fig. 29-5A). A few drugs have a different pattern of elimination. Three of the most commonly used drugs that exhibit this different pharmacokinetic pattern are ethanol, phenytoin, and salicylate. These drugs have dose-dependent, nonlinear saturation kinetics. As the dose of drug increases and the concentration of drug in the plasma rises, the relative amount of drug being eliminated falls (i.e., the clearance decreases) until the rate of drug metabolism is at its maximum. At this point, drug elimination is said to be zero order, and the drug concentration in plasma starts to increase much more (no longer linearly) with each subsequent increase in dose (Fig. 29-5B).

DRUG MONITORING AS A GUIDE TO THERAPY

Although published pharmacokinetic data (usually population averages) such as those in Table 29-1 are useful to determine initial drug dosing, modification of the dose may be needed in an individual patient. For some drugs (e.g., certain antihypertensives or anticoagulants), the therapeutic effects (e.g., blood pressure or coagulation) can be quantified easily over a range of concentrations, permitting adequate drug adjustment. For many other drugs (e.g., some antiarrhythmics or antiseizure medications), therapeutic effects over a range of concentrations are not readily detectable. With these drugs, the plasma concentration may provide further guidance in optimizing therapy if the plasma concentration of the drug is a reflection of its concentration at the site of action and the drug effects are reversible. A third, much smaller group of drugs produces irreversible effects (e.g., aspirin inhibition of platelet aggregation). With these drugs, plasma drug concentration does not correlate with drug effect, and drug monitoring is not useful.

To use drug concentration as a guide to therapy, it is necessary to establish a range of concentrations from minimally to maximally efficacious with tolerable toxicity. This range of concentrations, or the *therapeutic window*, is usually determined from a dose-response curve generated from a population of patients who have been examined closely for therapeutic and toxic effects (Fig. 29-6). This graph also may be used to determine the *therapeutic index*, a useful measure of drug toxicity calculated by dividing the 50% value from the toxicity curve by the 50% value from the efficacy curve. Because these curves are generated from population data, the values may not be applicable to all individuals.

Table 29-1, in addition to providing useful pharmacokinetic data, lists therapeutic ranges of several common drugs for which measuring the concentration and knowing the therapeutic range may be useful in clinical management. Many of these drugs are used to treat serious or life-threatening diseases. In these cases, it is essential to avoid inadequate doses because a therapeutic effect is often critical. Excessive doses must also be avoided because of the risk for toxicity with drugs that have a small therapeutic index. In contrast, it is not necessary to assay levels of drugs used to treat noncritical diseases (when inadequate treatment is not a serious problem) or for which the therapeutic index is large (when relative overtreatment is not likely to produce toxicity).

Problems with Interpreting Drug Concentration

The time of blood collection, perhaps more than any other factor, contributes to the misinterpretation of drug levels. As can be seen in Figure 29-2, if

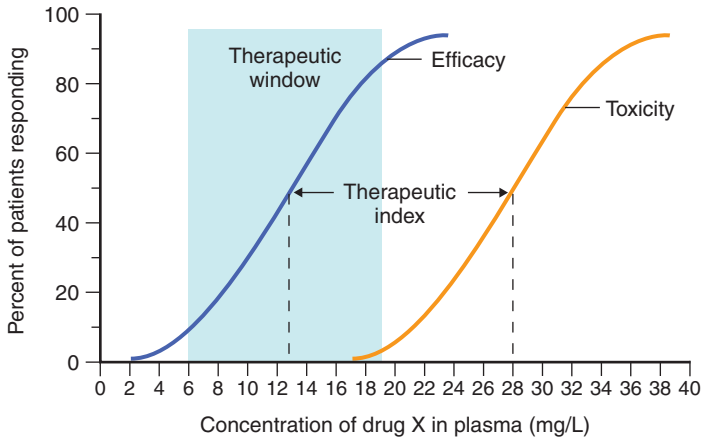

FIGURE 29-6. Pattern produced in a dose-response population study in which both effect and toxicity are measured. The therapeutic window is shown as the range of therapeutically effective concentrations, which includes most of the efficacy curve and less than 10% of the toxicity curve. The therapeutic index is calculated by dividing the 50% value on the toxicity curve by the 50% value on the efficacy curve.

sampling is performed too early, while the drug is still in the distribution phase, the drug level may be high and not reflect drug concentration at the site of action. It is therefore important to sample after the distribution phase.

For many drugs administered intermittently, a trough level, obtained immediately before the next dose is administered, is most useful for making decisions about dose adjustments (see Table 29-1). For drugs administered by infusion or intermittently at short intervals (see Fig. 29-4), the best time to draw blood is during steady state.

Protein binding is another major factor that contributes to the misinterpretation of drug levels. Free drug (not bound to protein and able to equilibrate with tissues and to interact with the site of action) is the critical drug concentration when therapeutic decisions are being made. Many drugs are tightly bound to plasma protein, however. Table 29-1 shows that many commonly used drugs, such as aspirin, carbamazepine, phenytoin, and glimepiride, have protein binding of more than 75%. Because many of the commonly used drug assays determine total drug concentration (which includes protein-bound drug and free drug), assessment of the "true" free drug concentration may be inaccurate, particularly if the fraction of drug bound to protein varies. In addition, the drug's binding may be decreased by disease or by other drugs, leading to increased unbound drug levels that alter the interpretation of the measured drug concentrations. Kidney and liver disease can change the binding of certain drugs (e.g., phenytoin) to protein because of a decrease in protein (e.g., decreased albumin, as in nephrotic syndrome or liver disease) or as a result of competition for protein binding by endogenously produced substances (e.g., uremia in kidney disease, hyperbilirubinemia in liver disease). Similarly, other drugs being administered may compete for binding to protein. A major problem secondary to these changes in protein binding is that free drug is not typically measured in many of the common drug assays used by clinical laboratories. Lastly, changes in drug binding to protein can affect the pharmacokinetics of the drug, the main effect being on the VD, which increases as protein binding decreases.

The usefulness of a drug assay is also limited by physiologic changes that may alter the response at a particular drug concentration. An example of this pharmacodynamic change is the response produced by a certain level of digoxin in the presence of altered electrolyte concentrations (e.g., potassium, calcium, and/or magnesium). Tolerance, a reduced response to a given concentration of drug with continued use, is another pharmacodynamic change that may alter how a drug concentration is interpreted. Tolerance is commonly observed with the continued use of narcotics (e.g., in terminal cancer patients); initially, adequate pain control is noted at a given drug concentration, but after long-term administration, the same drug concentration is no longer associated with pain relief.

ADJUSTMENTS OF DRUG DOSE WITH DISEASE

Kidney Disease

The major questions to be answered when determining whether a drug dosage needs to be adjusted in the setting of kidney disease are the following: Is the drug primarily excreted through the kidneys? Are increased drug levels likely to be associated with toxicity? If the answer to both is yes, it is likely that with decreased renal clearance, a drug will accumulate and become toxic. With renal failure, it is necessary to adjust the dosing regimen of such drugs, particularly for a drug with a long half-life and a small therapeutic index (e.g., digoxin).

To obtain the desired concentration over time in the presence of decreased clearance, adjustments can be made by decreasing the dose while maintaining the dose interval (DD), maintaining the dose but increasing the interval between doses (II), or a combination of both (DD and II). Table 29-2 shows how these three different methods are used to adjust the dosages of several common drugs to account for renal dysfunction (see Table 29-1 for their pharmacokinetic properties with normal renal function). With these adjustments, it may be possible to achieve an average concentration similar to that obtained with normal renal function; however, there may be concomitant marked changes in the magnitude of peak and trough values. In choosing the type of drug adjustment, the clinician should consider not only the therapeutic index of the drug but also (1) whether an effective concentration must be achieved quickly and maintained within a narrow range (i.e., maintaining an average drug concentration and avoiding trough levels at which the drug is ineffective) and (2) whether toxicity is associated with elevated (i.e., peak) drug concentrations.

Renal drug clearance correlates with creatinine clearance (whether the drug uses glomerular filtration or tubular secretion); therefore, any adjustment of drug dose in kidney disease can use the creatinine clearance to calculate the dose needed because renal drug clearance is proportional to creatinine clearance. The creatinine clearance (Cl_{cr}), which is used as an estimate of glomerular filtration rate, may be calculated directly from the serum creatinine concentration by the following equation:

$$Cl_{cr} = [(140 - age) \times weight\,(kg)]/[72 \times serum\ creatinine\,(mg/dL)] \quad (9)$$

The calculated creatinine clearance should be multiplied by 0.85 for females. (*Note:* This calculation applies only when the serum creatinine concentration is less than 5 mg/dL and renal function is not rapidly changing.)

Using Clearance for Dose Adjustment

The dose of a drug used in renal insufficiency ($dose_{D\text{-}RI}$) is proportional to the dose used with normal renal function ($dose_D$) in the same ratio as clearance of the drug in renal insufficiency ($Cl_{D\text{-}RI}$) to clearance with normal renal function (Cl_D). By rearranging, $dose_{D\text{-}RI}$ is defined as:

$$Dose_{D\text{-}RI} = dose_D \times [Cl_{D\text{-}RI}/Cl_D] \quad (10)$$

One can estimate the $Cl_{D\text{-}RI}$ by multiplying Cl_D by the ratio of the creatinine clearance in renal insufficiency ($Cl_{cr\text{-}RI}$) over Cl_{cr} with normal renal function:

$$Cl_{D\text{-}RI} = Cl_D \times [Cl_{cr\text{-}RI}/Cl_{cr}] \quad (11)$$

As shown in Equation 3, total clearance is the sum of clearance by renal and nonrenal (typically hepatic) mechanisms. Any nonrenal clearance is assumed to remain normal, and only the renal clearance is adjusted, with total clearance being reduced only to the extent that renal clearance is reduced. The dose may be calculated from the total (adjusted) clearance and the desired plasma concentration by either Equation 7 or Equation 8. The calculated dose is only an initial guide to the dose needed, however. By monitoring the drug response or the plasma drug concentration at various times after initial dosing, further dose adjustments can be made as necessary. From a practical perspective, most clinical dose adjustments in the presence of renal dysfunction can be guided by published tables based on changes in glomerular filtration rate (see Table 29-2) and the effectiveness of dialysis in removing the drug. Computerized decision support systems are particularly effective in guiding medication dosing for inpatients with renal insufficiency.

Loading Dose in Renal Insufficiency

For drugs typically administered with a loading dose in patients with normal renal function, the same approach can be used in those with renal insufficiency to ensure that the desired concentration is achieved rapidly. For drugs typically administered without a loading dose, the presence of a prolonged half-life resulting from renal insufficiency may delay drug accumulation to steady state. In this setting, a loading dose (equal to the amount needed to reach steady state with normal renal function) is required.

Additional Considerations in Renal Insufficiency

Because of individual differences among patients, the approaches outlined earlier should be considered only initial approximations to prevent ineffective

TABLE 29-2 ADJUSTMENT OF DRUG DOSAGE IN RENAL FAILURE

DRUG	Type of Elimination	HALF-LIFE (hr)		METHOD*	ADJUSTMENT FOR RENAL FAILURE			
		Normal	End-Stage Renal Disease		GFR > 50 mL/min	GFR 10-50 mL/min	GFR < 10 mL/min	Removed by Dialysis†
Amikacin	Renal	2-3	30	DD II	60-90% 12 hr	30-70% 12-18 hr	20-30% 24 hr	Yes
Aspirin	Hepatic (renal)	2-19	Unchanged	II	4 hr	4-6 hr	Avoid	Yes
Carbamazepine	Hepatic (renal)	35	?	DD	Unchanged	Unchanged	75%	No
Digoxin	Renal (nonrenal 15-40%)	36-44	80-120	DD II	Unchanged 24 hr	25-75% 36 hr	10-25% 48 hr	No
Lidocaine	Hepatic (renal < 20%)	1.2-2.2	1.3-3	DD	Unchanged	Unchanged	Unchanged	No
Lithium carbonate	Renal	14-28	Prolonged	DD	Unchanged	50-75%	25-50%	Yes
Penicillin G	Renal (hepatic)	0.5	6-20	DD	Unchanged 6-8 hr	75% 8-12 hr	25-50% 12-16 hr	Yes
Phenobarbital	Hepatic (renal 30%)	60-150	117-160	II	Unchanged	Unchanged	12-16 hr	Yes
Phenytoin	Hepatic (renal)	24	8	DD	Unchanged	Unchanged	Unchanged	No
Procainamide	Renal (hepatic 7-24%)	2.5-4.9	5.3-5.9	II	4 hr	6-12 hr	8-24 hr	Yes
Theophylline	Hepatic	3-12	?	DD	Unchanged	Unchanged	Unchanged	Yes
Tobramycin	Renal	2.5	56	DD II	60-90% 8-12 hr	30-70% 12 hr	20-30% 24 hr	Yes
Vancomycin	Renal	6-8	200-250	II	24-72 hr	72-240 hr	240 hr	No

*DD (alone) = decrease dose (maintain same interval); II (alone) = increase interval between doses (maintain dose); DD and II (together) = combination of both approaches.
†Dialysis refers to hemodialysis.
GFR = glomerular filtration rate.

(too low) or toxic (too high) doses. For maintenance therapy, it is desirable to monitor blood levels to guide dosing.

If a metabolite of the drug is responsible for its effect or toxicity and the metabolite accumulates in the setting of renal failure, the drug level alone may not provide sufficient guidance for planning therapy. For example, the major metabolite of procainamide is *N*-acetylprocainamide, which has a toxicity similar to that of the parent drug but only modest antiarrhythmic activity. In the setting of renal failure, *N*-acetylprocainamide may accumulate dramatically because it is more dependent on renal elimination. Measurement of procainamide levels alone does not accurately assess either the levels needed for antiarrhythmic effect or the risk for toxicity.

Liver Disease

Although many drugs are biotransformed in the liver, it is not possible to make any general recommendations for drug dose adjustments in liver disease. In contrast to renal disease, no useful laboratory test is available on which to base dose adjustments. It has been suggested that if the liver's capacity to produce protein (reflected by albumin concentration and prothrombin time) is reduced significantly, the clearance of drugs metabolized by the cytochrome P-450 enzymes is probably reduced as well.

One special situation that can develop with chronic liver disease and may require dose adjustment is the portacaval shunt. This condition produces not only a potential hemodynamic alteration, leading to decreased hepatic blood flow and accompanying decreased clearance, but also a possible bypassing of the first-pass effect, resulting in higher concentrations of drug reaching the systemic circulation. Drugs with a large hepatic extraction that are typically administered orally (e.g., propranolol) may appear in the systemic circulation at higher, potentially toxic concentrations.

Hemodynamic Diseases

Decreased cardiac output and hypotensive conditions lead to decreased perfusion of the organs, including those responsible for eliminating drugs. As noted earlier with regard to primary kidney disease, the dose can be adjusted for decreased renal perfusion by the use of creatinine clearance. The effect of decreased hepatic blood flow on pharmacokinetics is more difficult to assess. For drugs that have a high hepatic extraction (e.g., lidocaine), decreased hepatic blood flow suggests a need for dose reduction.

Altered hemodynamics also may affect the distribution of selected drugs. Drugs that have a relatively large VD (e.g., lidocaine, procainamide, quinidine) may be affected by conditions leading to hypotension, such as shock, resulting in a decrease in the apparent VD. With a reduced VD, the loading dose of a drug should be reduced to avoid potentially toxic drug levels.

In general, in the setting of severely compromised hemodynamics, it is advisable to be conservative, avoiding potentially toxic loading and maintenance doses of drugs. Drug levels and the clinical status should be monitored closely, and drug doses should be adjusted as necessary.

APPROACH TO DRUG OVERDOSE

The pharmacokinetic principles discussed earlier can be used to determine the best approach to drug removal in the setting of a drug overdose, particularly if hemodialysis or hemoperfusion is contemplated. The major goal is to increase the overall clearance of the drug, removing a substantial fraction of the total body load of drug. Examination of the VD and clearance values can provide some guidance. For drugs with a large VD (e.g., digoxin; see Table 29-1), only a small amount can be removed because clearance affects only the amount of drug present in the plasma, and a large portion of the drug is outside the plasma compartment. Similarly, for drugs with high clearance values, hemoperfusion may increase the overall clearance only minimally and is not indicated. Table 29-2 provides data for determining whether hemodialysis is likely to be useful to remove several commonly prescribed drugs.

DRUG USE IN ELDERLY PATIENTS

Administering drugs to elderly patients is perhaps the most challenging area of adult therapeutics because of several factors: the increasing likelihood of multiple illnesses, often with multisystemic involvement; the need for these patients to take multiple drugs (often prescribed by different physicians); and the increasing probability of altered pharmacokinetics and pharmacodynamics. These factors together contribute to a significantly increased frequency of drug interactions and adverse drug responses in this group of patients.

Pharmacokinetic Changes with Age

These changes can be secondary to the general physiologic effects of aging, such as alteration in body composition, or to specific changes in pharmacokinetically important organs (e.g., kidneys, liver). The distribution of drugs

tends to change dramatically with age, mainly because of changes in body composition. Most typical is the increase in total body fat, with the accompanying decrease in lean body mass and total body water. The concentration of plasma proteins may also change; in particular, albumin decreases as the liver ages. Changes in drug distribution are manifested as a change in the apparent VD. For water-soluble drugs that are not bound to plasma proteins, the apparent VD is reduced; in contrast, for lipid-soluble drugs, the VD is increased. Minimal changes in metabolism accompany aging, but these alone cannot account for altered pharmacokinetics.

Excretion can be altered in the elderly, and the clearance of many drugs is decreased. Cardiac output and blood flow to the kidneys and liver also may be decreased. Glomerular filtration rate may be reduced by 50%. Hepatic elimination of drugs is less affected, except for drugs with a high hepatic clearance (e.g., lidocaine). The elimination half-life of many drugs is increased with aging as a consequence of a larger apparent VD and a decreased hepatic or renal clearance (see Equation 5).

Pharmacodynamic Changes with Age

These changes are a result of changes in the responsiveness of the target organ. They require the use of smaller drug doses in elderly patients, even if the pharmacokinetics are unchanged. This affects many drugs commonly used in elderly people; for example, antianxiety drugs and drugs from the sedative-hypnotic class may produce increased central nervous system depression in elderly patients at concentrations that are well tolerated in younger adults. Similarly, anticoagulants (e.g., warfarin) may produce hemorrhage in elderly people at concentrations that are well tolerated in younger adults.

General Recommendations for Drug Use in Elderly Patients

- Clearance of drugs eliminated by the kidneys may be reduced by 50%.
- Drugs eliminated primarily by the liver typically do not require dose adjustments for age, except for drugs with high hepatic clearances, which may be affected by the age-related decrease in hepatic blood flow.
- Because of the potential for increased target organ sensitivity in elderly people, only the lowest effective dose should be used.
- Frequent reviews of the patient's drug history should be conducted, including both prescription and over-the-counter medications, keeping in mind the increased potential risk for drug interactions and adverse drug responses.

INTERACTIONS BETWEEN DRUGS

Because patients are typically treated with multiple agents, even for a single disease, the possibilities for drug interactions are great. Many clinically important drug interactions typically involve a drug with a low therapeutic index (e.g., warfarin) and an easily detectable pharmacologic effect (e.g., bleeding), such that a small increase in the amount of drug produces a significant effect (toxicity).

It is difficult to accurately assess the prevalence of drug interactions in either the inpatient or the ambulatory setting, particularly because no formal and comprehensive surveillance mechanism is available. The risk for drug interactions seems to be increasing, particularly for critically ill, hospitalized patients, who are frequently taking more than 10 medications.[1]

There are basically two types of drug interaction: (1) pharmacokinetic drug interactions, caused by a change in the amount of drug or active metabolite at the site of action; and (2) pharmacodynamic drug interactions (without a change in pharmacokinetics), caused by a change in drug effect.

Pharmacokinetic Drug Interactions

Less Drug at the Site of Action

Decreased Absorption

The gastrointestinal lumen is perhaps the best example of an area where drug interactions can result in decreased drug absorption. Some commonly used drugs can illustrate this type of interaction. For many drugs, a physicochemical interaction prevents the drug from being absorbed. Drugs such as colestipol and cholestyramine (resins used to lower cholesterol and bind bile acids) can also bind other drugs present in the gastrointestinal lumen, including digoxin and warfarin. Because of the potential for many other drugs to be bound, it is generally recommended that other drugs not be administered within 2 hours of colestipol or cholestyramine. Another type of interaction occurs when metal ions (e.g., aluminum, calcium, and magnesium in antacids and iron in supplements to treat iron deficiency) form insoluble complexes with tetracyclines, which can act as chelating agents. Other commonly used medications that decrease absorption include kaolin-pectin suspensions to treat diarrhea. These medications can significantly inhibit the absorption of coadministered drugs (e.g., digoxin).

Drugs that are particularly susceptible to pH changes may have decreased absorption when they are administered with other drugs that either affect gastric acidity or alter the extent of exposure to low pH. Protein pump inhibitors or histamine-2 receptor antagonists may elevate gastric pH, which can inhibit the dissolution and subsequent absorption of drugs that are weak bases (e.g., ketoconazole). Medications that delay gastric emptying (e.g., belladonna alkaloids) can increase the degradation of a coadministered acid-labile drug (e.g., levodopa), resulting in decreased absorption.

Altered Distribution

Drugs that use the same active transport process to reach their site of action can compete at the level of transport, resulting in lower levels of drug reaching that site. The classic example of this type of interaction is the coadministration of guanidinium-type antihypertensives with tricyclic antidepressants, phenothiazines, and certain sympathomimetic amines (e.g., ephedrine), which block the effects of the antihypertensive drug.

Increased Metabolism

Many drugs (e.g., phenobarbital, phenytoin, ethanol, glutethimide, griseofulvin, rifampin) and toxic compounds (e.g., cigarette smoke, certain chlorinated hydrocarbons) can increase the hepatic metabolism of other drugs (e.g., corticosteroids, cyclophosphamide, cyclosporine, certain β-adrenergic blockers, theophylline, warfarin) by inducing the activity of the cytochrome P-450 super family of monooxygenase enzyme.

More Drug at the Site of Action

Increased Absorption

Any drug that increases the rate of gastric emptying (e.g., metoclopramide) can potentially increase the absorption of acid-unstable drugs. Also, drugs that decrease intestinal motility (e.g., anticholinergics) may increase the absorption of drugs that are relatively poorly absorbed (e.g., digoxin tablets) by increasing the drug's contact time with the absorbing surface.

Altered Distribution

Drugs bound to protein are limited in their distribution (particularly to the site of action) and are not available for metabolism or excretion. Drugs can compete with each other for binding to plasma proteins, resulting in drug interactions. Sulfonamides can displace barbiturates bound to serum albumin, leading to increased levels of free barbiturates and possible toxicity.

Decreased Metabolism

One of the most impressive drug interactions is produced when one drug inhibits the metabolism of another, leading to the second drug's accumulation and a significant risk for toxicity. This type of interaction results when 6-mercaptopurine, an antileukemic drug with a low therapeutic index, is used with allopurinol, often administered to control hyperuricemia. The interaction may result in potentially life-threatening toxicity.

Some drugs can inhibit the metabolism of many other drugs. For example, cimetidine can inhibit the metabolism of diazepam, imipramine, lidocaine, propranolol, quinidine, theophylline, and warfarin. Amiodarone inhibits the metabolism of calcium-channel blockers, phenytoin, quinidine, and warfarin. Of particular importance with amiodarone is its half-life of 1 to 2 months; it continues to inhibit drug metabolism for several months after it has been discontinued.

Other drugs are notable because their metabolism is inhibited by a variety of different drugs. The metabolism of the commonly used anticoagulant warfarin is inhibited not only by cimetidine and amiodarone but also by many other drugs, including alcohol, allopurinol, disulfiram, metronidazole, phenylbutazone, sulfinpyrazone, and trimethoprim-sulfamethoxazole. Similarly, the metabolism of phenytoin is inhibited by chloramphenicol, clofibrate, dicumarol, disulfiram, isoniazid (slow acetylators), phenylbutazone, and valproic acid.

Although most of these examples involve enzymes that metabolize the drug in the liver, drug-metabolizing enzymes outside the liver also may be affected by certain drugs. The best-known example is monoamine oxidase, which can be affected by nonspecific monoamine oxidase inhibitors, resulting in the accumulation of catecholamines at multiple sites after their release in response to the eating of tyramine-containing foods such as aged cheese, aged or cured meats, and any spoiled meat, poultry, or fish.

Decreased Excretion

Drugs can compete for the active transporters present in the kidney. Most of these interactions involve the acid transporters. The best-known interaction is probenecid's inhibition of penicillin transport, leading to decreased penicillin clearance and thus higher plasma levels, an interaction that was used in the past to maximize penicillin therapy. A similar inhibitory effect on the renal excretion of methotrexate can be produced by salicylates, phenylbutazone, and probenecid. The active transport of basic drugs (e.g., procainamide) can also be inhibited by other drugs (e.g., cimetidine, amiodarone).

Pharmacodynamic Drug Interactions

With pharmacodynamic interactions, drugs interact at the level of the receptor (target) or may produce additive effects by acting at separate sites on cells. An example of the first is the interaction of propranolol and epinephrine, which blocks β-adrenergic receptors; as a result, the α-adrenergic effects of epinephrine are unopposed. This undesirable interaction can result in severe hypertension.

Many examples exist of the additive effects of drugs. Aspirin, which can produce increased bleeding time by acting on platelets, can interact with warfarin, which affects clotting. The result is an increased risk for hemorrhage. Similarly, cardiac drugs, such as β-adrenergic blockers and calcium-channel blockers, have additive negative inotropic effects when they are coadministered, resulting in an increased risk for cardiac failure.

DIAGNOSIS AND PREVENTION OF DRUG INTERACTIONS

For a drug interaction to be recognized, the index of suspicion must be high whenever multiple drugs are used together. Because of the ever-increasing list of known and suspected drug interactions, it is impossible for a clinician to remember all or even many of the possible interactions.

Several clinical settings should raise concern about the possibility of drug interactions:

- The use of any drug with a low therapeutic index (Table 29-3) should be suspect.
- As the number of drugs being used concurrently increases, there is a disproportionately greater risk for drug interactions, particularly with more than 10 drugs.
- Critically ill patients who have multisystemic disease with compromised renal, hepatic, cardiac, or pulmonary function have an increased risk for drug interactions. This risk may be higher for patients with acquired immunodeficiency syndrome, who have an immunocompromised state and take a large number of drugs.
- Patients with various behavioral and psychiatric disorders (e.g., drug abusers taking a large number of prescription drugs as well as illicit drugs and alcohol) are at risk for drug interactions.

Another type of drug interaction that is becoming increasingly important is that between components of food (e.g., grapefruit juice) or natural products (e.g., herbs) and drugs. By inhibiting the intestinal cytochrome P-450 3A4 enzyme system, grapefruit juice can raise levels of drugs metabolized by this pathway (e.g., saquinavir, cyclosporine, verapamil) and result in toxicity or adverse drug effects.

Several steps can be taken to prevent drug interactions:

- When taking the medical history, it is important to document all drugs the patient is taking (and has recently taken), including prescription, over-the-counter, and other addictive drugs.
- It is desirable to minimize the number of drugs being taken by frequently reviewing the patient's drug list to ensure that each drug continues to be needed.
- There should be a high degree of suspicion when medications with a low therapeutic index known to have a high risk for drug interactions (see Table 29-3) are used.
- High-risk clinical settings, such as occur with critically ill patients, should raise the suspicion of adverse drug interactions.
- Adverse drug interactions should be considered in the differential diagnosis whenever any change occurs in a patient's course.

TABLE 29-3 DRUGS WITH LOW THERAPEUTIC INDICES AT HIGH RISK FOR ADVERSE DRUG RESPONSES AND DRUG INTERACTIONS

Anticoagulants
Antiarrhythmics
Anticonvulsants
Digoxin
Lithium carbonate
Oral hypoglycemics
Theophylline

ADVERSE REACTIONS TO DRUGS

An adverse drug response (ADR) is an undesired effect produced by a drug at standard doses, which typically necessitates reducing or stopping the suspected agent and may require treatment of the noxious effect produced. Further harm may occur with continued or future therapy with the drug.

EPIDEMIOLOGY

The financial impact of ADRs is estimated to be more than $100 billion per year. The actual incidence of ADRs is difficult to quantify because many cases are either not recognized or not reported. Several large studies have shown that the incidence may approach 20% for outpatients (even higher for patients taking more than 15 drugs) and 2 to 7% for inpatients. The incidence of ADRs increases exponentially with more than four drugs. Meta-analyses of several prospective studies suggest that ADRs are now the third leading cause of death in hospitalized patients. It is clear from more recent surveys that a relatively small group of drugs (see Table 29-3) continues to be implicated in most of the reported ADRs. Current trends suggest that the incidence of ADRs is likely to increase as a result of more prescribed and over-the-counter medications being used. Although still in an evolutionary phase, systems approaches have the potential to greatly advance the understanding of adverse drug events and the ability to predict them.[2]

PATHOBIOLOGY

Most ADRs are caused by an exaggerated (but predictable) pharmacologic effect of the drug or by a toxic or immunologic effect of the drug or a metabolite (not typically expected). Among an estimated 100,000 annual drug-related emergency hospitalizations in patients older than 65 years, the most commonly implicated drugs are warfarin (33%), insulin (14%), oral antiplatelet agents (13%), and oral hypoglycemic agents (11%).

Predictable Toxic Responses to Drugs

Exaggerated drug responses that cause adverse drug effects may be due to any condition that causes altered pharmacokinetics or pharmacodynamics (discussed earlier). There has been increasing interest in the discipline of pharmacogenomics, which explores the role of genetic factors in altered pharmacokinetics or pharmacodynamics and the resultant increased susceptibility to ADRs. Furthermore, a developing discipline of pharmacometabolomics should complement pharmacogenomics in personalized drug therapy by capturing environmental and microbiome-level influences on responses to drugs.[3] There is now ample evidence that molecular changes in the genes coding for drug-metabolizing enzymes can account for the variability in pharmacokinetics and drug effects observed in population studies. There are now many examples of the role of pharmacogenomics in altered drug response and effect. Three of the best-studied examples are genetic polymorphisms associated with debrisoquine-sparteine, *N*-acetylation, and mephenytoin. Each is associated with autosomal recessive inheritance, and together they are responsible for the metabolism of approximately 40 drugs (Table 29-4). Individuals with autosomal recessive genes are typically "poor metabolizers," with potentially altered pharmacokinetics that result in elevated plasma drug concentrations and can lead to toxicity. Recent studies have focused on genomic alterations associated with the variability in response to some commonly used drugs. Thus, the variability of response to warfarin (Chapter 38) is now understood to be due to single-nucleotide polymorphisms in the cytochrome P-450 2C9 (*CYP2C9*) and vitamin K epoxide reductase (*VKOR*) genes. These single-nucleotide polymorphisms have a significant effect on warfarin dose requirements. Similarly, polymorphisms in transporter genes can have profound effects on the pharmacokinetics of statins (Chapter 206). A common genetic variant of the organic anion-transporting polypeptide 1B1 can reduce the hepatic uptake of many statins, increasing the risk for statin-induced myopathy. Also, it is now appreciated that genetically impaired adenosine triphosphate (ATP)-binding cassette G2 transporter efflux activity can result in an increase in systemic exposure to various statins. Of particular importance to therapeutics is that the effects of these genetic polymorphisms differ, depending on the specific statin used.

TABLE 29-4 GENETIC POLYMORPHISMS OF DRUG-METABOLIZING ENZYMES

TYPE	PRIMARY DRUG EXAMPLES	OTHER DRUGS THAT ARE SUBSTRATES	INCIDENCE OF "POOR METABOLIZERS" IN WHITES (%)	ENZYME INVOLVED
Debrisoquine-sparteine polymorphism	Amitriptyline, codeine, tamoxifen	Antidepressants, antiarrhythmics, β-adrenergic receptor–blocking drugs, codeine, dextromethorphan, neuroleptics	5-10	Cytochrome P-450 IID6 (CYP2D6)
Mephenytoin polymorphism	Mephenytoin	Mephobarbital, hexobarbital, diazepam, omeprazole	4 (Japanese, Chinese, 15-20)	Cytochrome P-450 IIC (CYP2C)
N-acetylation polymorphism	Isoniazid, sulfadiazine	Hydralazine, phenelzine, procainamide, dapsone, sulfamethazine, sulfapyridine, aminoglutethimide, aminosalicylic acid, sulfasalazine	40-70 (Japanese, 10-20)	*N*-acetyltransferase (NAT2)
Methyl conjugation polymorphism	Catecholamines	L-Dopa, methyldopa	25-30	Catechol-*O* methyltransferase (COMT)

This provides a rational basis for an individualized approach to the use of lipid-lowering therapeutic agents.

A particularly impressive example occurs with certain cancer chemotherapy agents that have a relatively narrow therapeutic window and the potential to produce severe cytotoxicity (e.g., deficiency in dihydropyrimidine dehydrogenase activity can result in life-threatening toxicity after the administration of 5-fluorouracil). These defects typically are not recognized until the patient is given the drug. They are often described as "pharmacogenetic" syndromes.

Other genetic alterations do not affect metabolism specifically and do not produce a range of quantitative changes. These defects can produce "qualitative" defects and are often associated with structural defects. The classic example is glucose-6-phosphate dehydrogenase. Individuals who are deficient in this enzyme cannot tolerate the oxidative stress produced by some drugs, leading to hemolysis (Chapter 161). Drugs that can produce this clinical picture include aspirin, nitrofurantoin, primaquine, probenecid, quinidine, quinine, sulfonamides, sulfones, and vitamin K. Another similar defect is deficiency of methemoglobin reductase, which results in an inability to maintain iron in hemoglobin in the ferrous state, causing methemoglobinemia (Chapter 158) after exposure to oxidizing drugs such as nitrites, sulfonamides, and sulfones.

Unpredictable Toxic Responses to Drugs

Other toxic or immunologic ADRs are not predictable and are not obviously due to an increase in drug concentration (pharmacokinetic) or drug effect (pharmacodynamic). Unpredictable toxic responses include direct reactions between a drug and a specific organ (e.g., platinum-containing drugs, such as cisplatin, can produce direct toxicity in the kidney and the eighth cranial nerve). With other drugs, metabolism to an active intermediate must occur first. With a standard dose of acetaminophen, no untoward effects occur because the relatively small amount of reactive metabolite formed by oxidative metabolism is detoxified rapidly by reduced glutathione. In the presence of an overdose, the glutathione is depleted, and the remaining reactive metabolite can damage the liver. Understanding the mechanism of this toxicity has provided a rationale for treating acetaminophen overdose. Sulfhydryl-containing compounds (e.g., *N*-acetylcysteine), which can complex with the reactive metabolite, can be administered to reduce the amount of free toxic metabolite present, protecting the liver.

Immunologic reactions to drugs are generally not produced by the drug alone. Similar to other low-molecular-weight compounds (<1000 D), they are typically not antigenic themselves. When a drug or reactive metabolite combines with a protein to form a drug-protein complex, it can become antigenic, capable of eliciting an immune response.

Perhaps the most impressive form of drug allergy is anaphylaxis, which is due to an immunoglobulin E–mediated hypersensitivity. Many drugs from different classes have been shown to produce this type of drug allergy. The best-known example is the anaphylactic response produced by penicillin, which can occur after its administration by any route. Skin testing with penicillin G, penicilloic acid, or penicilloyl polylysine can identify patients at risk and should be performed in those with a suspected penicillin allergy who need treatment with penicillin. If the skin test result is positive, the patient must undergo desensitization before receiving penicillin. If the skin test result is negative, penicillin can be administered with caution.

DIAGNOSIS OF ADVERSE DRUG RESPONSES

Although many of the well-known adverse drug effects are due to a relatively small group of drugs, every drug has the potential to cause an ADR. A physician should always consider the possibility of an ADR in the differential diagnosis even if none has been reported previously for the particular drug. In earlier editions of this book, we provided a table listing many diverse clinical presentations associated with ADRs. The reader is now referred to the numerous websites providing more complete and up-to-date information on ADRs than can be detailed here. For example, *www.fdable.com* allows one to monitor drug safety with unlimited access to up-to-date information from the U.S. Food and Drug Administration (FDA) Adverse Event Reporting System (AERS).

In many instances, it is readily apparent that a specific drug has produced an ADR, such as the appearance of a rash in an otherwise healthy patient who was recently prescribed a single drug (e.g., penicillin). In other cases, the effect produced by the drug may be difficult to discern from other disease states. In still other cases, the adverse effect may mimic the illness being treated (e.g., development of an arrhythmia in a patient being treated with an antiarrhythmic drug).

From a public health perspective, it is highly desirable to have a mechanism available to detect, catalog, and track the incidence and severity of ADRs not only for drugs at various stages of development but also for drugs that were approved earlier. The FDA tracks adverse drug events through a voluntary reporting program, MedWatch. Health care professionals are encouraged to report any adverse events or product problems on a one-page form that can be sent by mail, fax, or electronically to the FDA. Nontraditional resources that are generated by patients via the Internet could supplement existing pharmacovigilance approaches.[4] Although various methods for surveying ADRs have been proposed, ultimately, the cooperation of alert clinicians, health care professionals, and patients must be encouraged.[4]

GENOMIC DATA PROVIDING GUIDANCE FOR CANCER THERAPEUTIC DECISION

In addition to genetic data being useful in understanding and increasingly predicting adverse drug effects (see earlier), it is becoming increasingly clear that genomic data (e.g., the presence of a specific genetic variant) can also help inform the clinician of the appropriate drug or dose of drug that should be used for the specific patient.[5,6] There is no therapeutic area in which this is more apparent than with oncologic drugs and particularly many of the new targeted therapy agents, for which the genomic profile of the tumor (vs. the host tissue) is critical.[7] A representative example is vemurafenib, a targeted agent that has recently been demonstrated to have therapeutic value in melanoma and potentially other malignancies as well. Vemurafenib is effective in tumors that have a specific mutation in their *BRAF* gene, a valine-to-glutamic acid mutation at residue 600 (V600E), that results in the oncogene protein product, BRAF(V600E) kinase, exhibiting a markedly elevated activity that overactivates the MAPK signaling pathway. Vemurafenib is an orally bioavailable, ATP-competitive, small-molecule inhibitor of BRAF(V600E) kinase with antineoplastic activity. It selectively binds to the ATP-binding site of BRAF(V600E) kinase and inhibits its activity, resulting in inhibition of an overactivated MAPK signaling pathway downstream in BRAF(V600E) kinase–expressing tumor cells, thereby reducing tumor cell proliferation. There are multiple other examples in oncology demonstrating that genomic

data from the tumor of an individual patient can be useful in making a therapeutic decision.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hennessy S, Flockhart DA. The need for translational research on drug-drug interactions. *Clin Pharm Ther*. 2012;91:771-773.
2. Lesko LJ, Zheng S, Schmidt S. Systems approaches in risk assessment. *Clin Pharm Ther*. 2013;93: 413-424.
3. Kaddurah-Daouk R, Weinshilboum RM. Pharmacometabolomics: implications for clinical pharmacology and systems pharmacology. *Clin Pharmacol Ther*. 2014;95:154-167.
4. White RW, Harpaz R, Shah NH, et al. Toward enhanced pharmacovigilance using patient-generated data on the internet. *Clin Pharmacol Ther*. 2014;96:239-246.
5. Johnson JA, Cavallari LH. Pharmacogenetics and cardiovascular disease: implications for personalized medicine. *Pharmacol Rev*. 2013;65:987-1009.
6. Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. *N Engl J Med*. 2011; 364:1144-1153.
7. Wheeler HE, Maitland ML, Dolan ME, et al. Cancer pharmacogenomics: strategies and challenges. *Nat Rev Genet*. 2013;14:23-34.

REVIEW QUESTIONS

1. A drug is considered to be effectively cleared from the circulation after how many half-lives?

A. 1.5
B. 2
C. 5
D. 7
E. 10

Answer: C A drug is considered to be effectively cleared when less than 10% of drug remains in the circulation. This is typically 3 to 5 half-lives.

2. Which one of the following drugs may be difficult to remove in an overdose setting because of a large volume of distribution (VD)?

A. Procainamide
B. Aspirin
C. Digoxin
D. Phenobarbital
E. Phenytoin

Answer: C A large VD indicates that a large fraction of the specific drug is present outside the circulation. Examination of drugs in Table 29-1 demonstrates that digoxin's V_D is much higher than the other four drugs such that, in an overdose situation, easy removal of digoxin from the affected individual would be difficult because a large fraction of the drug is outside the circulation and therefore relatively inaccessible.

3. The antihistamine terfenadine (Seldane) should not be administered with which one of the following fruit juices?

A. Lemon juice
B. Orange juice
C. Prune juice
D. Grapefruit juice
E. Apple juice

Answer: D Many drugs are metabolized by CYP3A4 in the small intestine. Certain chemicals in grapefruit juice, but not in the other juices listed, block the action of CYP3A4, so instead of being metabolized, more of the drug enters the blood stream and stays in the body longer. The result is that potentially dangerous levels of the drug accumulate. This finding has led to more attention to foods, herbs, and other natural substances that are coadministered with drugs.

4. Which one of the following drugs has pharmacokinetics characterized as dose-dependent, nonlinear saturation kinetics?

A. Digoxin
B. Tobramycin
C. Lidocaine
D. Glimepiride
E. Phenytoin

Answer: E Not all drugs behave the same when drug dose is increased. The elimination of most drugs follows first-order or linear kinetics, such that the amount of drug eliminated is directly proportional to the plasma concentration of the drug. A few drugs (e.g., phenytoin) have a different pattern of elimination—that is, dose-dependent, nonlinear kinetics—such that as the drug dose increases and the plasma concentration of the drug rises, the relative amount of the drug being eliminated falls (i.e., clearance decreases) until the rate of drug metabolism is at a maximum. At this point, drug elimination is said to be zero order, and the drug concentration in plasma starts to increase much more (no longer linear). This is shown in Figure 29-5B.

5. Before administration of vemurafenib, which one of the following genes should be monitored for the presence of a V600E single-nucleotide polymorphism?

A. EGFR
B. BRAF
C. NMYC
D. VEGF
E. CMYC

Answer: B The effectiveness of vemurafenib in tumors has been demonstrated to be dependent on the presence of a specific mutation in the *BRAF* gene of the tumor, a valine-to–glutamic acid mutation at residue 600 (V600E). This mutation results in the oncogene protein product, BRAF(V600E) kinase, exhibiting a markedly elevated activity that overactivates the MAPK signaling pathway. Vemurafenib selectively binds to the adenosine triphosphate–binding site of BRAF(V600E) kinase and inhibits its activity, resulting in inhibition of an overactivated MAPK signaling pathway downstream in BRAF(V600E) kinase–expressing tumor cells, thereby reducing tumor cell proliferation. Before considering this drug, specific testing for the presence of this mutation is recommended.

30 PAIN

STEVEN P. COHEN AND SRINIVASA N. RAJA

Pain is ubiquitous in life, usually serving as a warning sign of impending or actual injury to the organism. As such, pain is older than humans, dating back to our most primitive ancestors. Pain is also a vital diagnostic clue for physicians. Physicians should be intimately familiar with pain, for it is the most common symptom for which patients seek medical attention.

It is difficult to overestimate the impact that pain has on society. According to an Institute of Medicine report released in 2011, one in three Americans suffer from chronic pain, which is more than the total affected by heart disease, cancer, and diabetes combined. Among the various types of pain, back pain is the most common, followed by severe headaches, arthralgias, and neck pain. The estimated economic cost of chronic pain, to include medical costs and lost productivity, ranges between $560 and $635 billion (in 2010 dollars).[1] Spinal pain is the leading cause of disability in individuals younger than 45 years in industrialized nations, with the economic costs exceeding $100 billion annually in the United States by some estimates. Special populations at risk for chronic pain include the elderly, and individuals with physical and psychological morbidities. Several conditions characterized by chronic pain, such as headaches, irritable bowel syndrome, fibromyalgia, and complex regional pain syndrome, are more prevalent in women than in men.

DEFINITION

The International Association for the Study of Pain defines pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage." This definition recognizes that pain may be experienced in some circumstances in the absence of ongoing tissue damage (e.g., phantom pain after a healed amputation). One implication of this construct is the assumption that pain is always subjective; hence, a patient's report of pain should always be accepted at face value in the absence of evidence to the contrary.

PATHOBIOLOGY

Classification of Pain States

Multiple classifications have been used to describe pain states on the basis of duration, anatomic source, and etiology. *Acute pain* usually results from injury or inflammation, has survival value, and may play a role in healing by promoting behaviors that minimize reinjury. In contrast, *chronic pain* is perhaps best construed as a "disease" that serves no useful purpose. Although there is no clear threshold at which acute pain transitions to a chronic state, it is generally accepted that pain persisting beyond the expected healing period is pathologic. In most cases, this period is between 3 and 6 months. The intensity of pain can be classified as mild (1 to 3), moderate (4 or 5), or severe (≥6 on a 0 to10 numerical rating scale).

Somatic and Visceral Pains

Pain can originate from somatic or visceral structures. *Somatic pain* is typically well localized and generally results from injury or disease of the skin, musculoskeletal structures, and joints. Different types of stimulation can evoke pain by binding to distinct receptors (also known as nociceptors), which can be broadly categorized as chemosensitive, thermosensitive, mechanosensitive, or polymodal. *Visceral pain* arises from internal organ dysfunction and can result from inflammation, ischemia, occlusion of flow resulting in capsular or organ distention (e.g., renal stones, bowel obstruction, cholecystitis), or functional disease (e.g., irritable bowel syndrome). In contrast to somatic pain, visceral pain is usually diffuse and poorly localized, is often referred to somatic regions (e.g., myocardial ischemia radiating into the arm), and tends to be associated with exaggerated autonomic reflexes and greater emotional features. Some of the reasons for these differences between somatic and visceral pain include a lower density and different types (e.g., distention sensitive) of nociceptors in visceral structures and convergence with afferent pathways predominantly populated by somatic pain in the spinal cord.

Neuropathic, Nociceptive, and Mixed Pain

Pain can be etiologically classified as neuropathic, nociceptive, or mixed (Table 30-1). *Neuropathic pain* has been defined as pain arising as a direct consequence of a lesion or disease affecting the somatosensory system. Common peripheral neuropathic pain states include post-herpetic neuralgia, diabetic neuropathy, and radicular pain. One subtype of neuropathic pain is central pain, which is manifested as a constellation of symptoms requiring a primary lesion to the central nervous system as a necessary (but not sufficient) inciting event. Owing to its high prevalence, the most common overall cause of central pain is central post-stroke pain (occurring after approximately 8% of cerebrovascular accidents), although spinal cord lesions (e.g., syringomyelia or spinal cord injury) are associated with a higher incidence of central pain (>50%). Epidemiologic studies indicate that neuropathic pain is prevalent in 7% to 8% of the population, and that 20% to 25% of all chronic pain is neuropathic in nature. *Nociceptive pain* usually results from an injury or disease affecting somatic structures such as skin, muscle, tendons and ligaments, bone, and joints. Pain associated with cancer can result either from the tumor itself or as a consequence of therapy (e.g., surgery, chemotherapy, and radiation therapy). In light of the often multiple different causes, advanced cancer pain is a typical example of a mixed pain state.

Dysfunctional Pain

There is a group of pain syndromes that have been characterized by amplification of pain signaling in the absence of either inflammation or injury (as in nociceptive pain) or damage to the nervous system (as in neuropathic pain). These conditions include pain states such as fibromyalgia, irritable bowel syndrome and interstitial cystitis. The precise pathophysiologic mechanisms of pain in these disorders are still being elucidated, although they share some features of neuropathic pain such as augmented sensory perception and altered central neurotransmission.

Pain Mechanisms

Pain results from activation of specialized peripheral receptors (*nociceptors*) by a noxious event (stimulus). These receptors respond to various external stimuli: *mechanical* (e.g., pressure, tumor growth, incision), *thermal* (e.g., hot or cold), or *chemical* (e.g., ischemia or infection). In addition to distinct nociceptors that respond to each type of stimulus, there are also *polymodal* nociceptors that respond to multiple stimulus modalities. Once a stimulus is detected, it is converted into an electrical nerve signal (*transduction*), which is conveyed along the axons of thinly myelinated (A delta) or unmyelinated (C) nerve fibers through specific pathways (*transmission*). *Modulation* generally refers to the attenuation of pain signals through intrinsic inhibitory activity within the peripheral and the central nervous systems before being perceived as an unpleasant sensation (*perception*), though in some cases pain signals may be amplified during this process. *Pathologic pain* is the result of injury- or disease-induced changes in the peripheral or central nervous systems leading to alterations in pain signaling. One important example of pathologic pain from injury to the nervous system is *peripheral sensitization*. This form of pain is characterized by the development of spontaneous ectopic activity in injured nerves and dorsal root ganglion cells as well as enhanced sensitivity to mechanical, thermal, or chemical stimuli.

Prolonged and repeated activation of nociceptive afferent fibers produces *central sensitization*, a state of increased sensitivity of central pain signaling neurons. Activation of *N*-methyl-D-aspartate (NMDA) receptors by glutamate is thought to be an important mechanism for central sensitization. Studies indicate that in addition to functional changes in neurons, microglia and astrocytes may also play an important role in the central sensitization process. Other central neuroplastic changes that may contribute to neuropathic pain states include deafferentation hyperactivity that may occur after spinal cord or avulsion injuries, loss of large-fiber inhibition, reorganization of central connections of primary afferent fibers, and excitatory descending modulatory mechanisms. Central and peripheral sensitization are considered to be the prime culprits responsible for pain induced by innocuous stimuli (*allodynia*) and increased pain to normally noxious stimuli (*hyperalgesia*)

TABLE 30-1 CLASSIFICATION AND PREVALENCE OF COMMON PAIN CONDITIONS

NEUROPATHIC		*NOCICEPTIVE*		*MIXED*
Peripheral*	***Central***	***Somatic***	***Visceral***	
Peripheral neuropathy (1-3%)	Central post-stroke pain (8%)	Arthritis (25-40% in people > 40 years)	Endometriosis (10% in women of reproductive age)	Headache (15% for migraine, 20-30% for tension type)
Post-herpetic neuralgia (annual incidence 0.1-0.2%)	Spinal cord injury (30-50%)	Myofascial pain (5-10%)	Irritable bowel syndrome (5-15%)	Cancer[§] (lifetime prevalence 30-40%)
Chronic postsurgical pain (5-20% after surgery)	Multiple sclerosis (25%)	Fibromyalgia[†] (2-4%)	Interstitial cystitis (0.2-1% of women)	Low back pain[‖] (point prevalence 10-30%)
Phantom limb pain (30-60%)	Parkinson disease (10%)	Connective tissue disorders (0.2-0.5%)	Ulcers, gastritis, esophagitis (3-9%)	Neck pain[‖] (annual incidence 20-30%)
Trigeminal neuralgia (0.01%)	Seizure disorder (1-3%)	Burn pain[‡] (annual incidence of burns requiring hospitalization 0.01%)	Cholecystitis, appendicitis	Ischemic pain[¶]
Radiculopathy, spinal stenosis (3%-10%)				
Complex regional pain syndrome (0.03%, 3-20% after orthopedic surgery)				
Nerve entrapment syndromes (e.g., carpal tunnel, thoracic outlet, meralgia paresthetica; 2-4%)				

*Prevalence rates represent proportion of patients with condition who develop pain.
†Some cases may represent a variant of central pain.
‡Third-degree burns are often associated with neuropathic pain.
§Neuropathic pain occurs in 20 to 50% of cases and may be secondary to tumor invasion, surgery, chemotherapy, and radiation treatment.
‖Neuropathic pain may accompany nociceptive pain in 10 to 35% of cases.
¶Typically nociceptive, but long-standing pain may result in ischemic neuropathy.

that are commonly observed in neuropathic pain states. Studies have shown an important role for neurotrophins, prostaglandins, cytokines such as tumor necrosis factor-α and interleukins, and glial cells in the sensitization processes.[2]

Genetics

Chronic pain is a prototypical example of the interplay between genes and environment. Although tissue or nerve injury is necessary for the development of most pain syndromes, by itself it is not sufficient, as only a small percentage of injuries result in chronic pain. There are many ways in which one's genetic makeup can influence pain to include differences in pain sensitivity, susceptibility to disease, immunomodulatory response to injury, psychological predisposition to pain persistence, interactions between genotype differences and environment, and response to analgesic therapy. Not only genetics but also epigenetics can affect how painful stimuli are perceived.

Heritability is estimated to account for between 30% and 60% of the variance in pain response and has been demonstrated to play a role in acute pain perception and the transition from acute injury to chronic pain.[3] Even when the variability associated with gender and racial differences is taken into consideration (i.e., women and African Americans are more likely to report pain than are men and whites, respectively), genetics continues to play a major role in explaining pain differences. Heritable pain conditions may occur through either dominant or recessive gene transmission, and involve an assortment of different phenotypes.

Although rare, examples of monogenic (single-gene mutation) pain conditions include hereditary sensory and autonomic neuropathies, familial hemiplegic migraine, and neurologic channelopathies such as primary erythromelalgia, which involves a mutation at the gene encoding the voltage-gated sodium channel Nav 1.7. This latter mutation causes increased excitability and ongoing activity in afferent sensory neurons.

More common are those conditions associated with multiple gene abnormalities and incomplete penetrance, with about a dozen genes accounting for more than half of the identified gene candidates. In most cases, genes may promote a predisposition to pain, which requires a subsequent environmental inciting event (e.g., an injury in the context of depression) for manifestation.

Compared with other specialties, the study of the genetic basis for pain variability is still in its infancy, with most research revolving around conditions containing pain as a major symptom (e.g., endometriosis, osteoarthritis) rather than pain itself as an independent measure. In the future, the translational application of genetic studies may involve the suppression of pain-facilitating alleles, increasing the expression of pain-protective alleles, and targeted analgesic therapies.

DIAGNOSIS

History

Similar to the work-up of any symptom, the evaluation of pain begins with a thorough history. One of the primary tenets of pain assessment is that subjective complaints should always be taken seriously. There is no diagnostic test that can measure pain or even ascertain its existence. The most promising techniques involve functional brain imaging that reflects cerebral metabolism. These research tools have helped us understand that the brains of chronic pain patients are different from those of normal individuals and undergo morphologic alterations such as a diminution in gray matter in the areas involved in pain perception. Studies have confirmed that these deleterious changes are a consequence, not a cause, of chronic pain and may be reversed by effective treatment.[4]

A comprehensive history should include the location of pain and its quality, exacerbating and relieving factors, temporal aspects, associated symptoms and signs (e.g., numbness or weakness), interference with activities of daily living, and response to prior treatments. The temporal aspects of pain can provide valuable clues to etiology. Most cases of acute pain develop subsequent to a specific inciting event (e.g., surgery, trauma), whereas chronic pain conditions are usually more insidious in onset. Because acute pain tends to be self-limited and the relationship to a precipitant event is more tangible, it is generally better tolerated and associated with fewer psychological sequelae.

The severity of pain can be measured by a variety of different rating scales. Some of the more common instruments include categorical scales, verbal and numerical rating scales (0 to 10), and the visual analog scale, in which a 10-cm line is anchored on each side by two points designated "no pain" and "worst possible pain." Because there are subtle differences between different scales, repeated assessments of response to therapy are ideally gauged by the same instrument. For young children and mentally incapacitated patients, the use of age-appropriate substitute scales or facial expressions has been validated.

Recent guidelines from experts across multiple specialties have concluded that pain scores represent only one component of pain management. Other important aspects of treatment include assessments of functional capacity (e.g., Oswestry disability index for back pain), psychological and emotional functioning, satisfaction ratings, adverse treatment effects and disposition (i.e., work status). It is therefore imperative that realistic goals be established and individually tailored treatment regimens be developed to achieve these ends.

Distinguishing between neuropathic and nociceptive pain can have important treatment implications (Table 30-2). Neuropathic pain is characterized by positive and negative symptoms. Negative symptoms, such as a loss of sensation, are usually the result of axon or neuron loss, whereas positive

TABLE 30-2 CATEGORIZATION OF NEUROPATHIC AND NOCICEPTIVE PAIN

CLINICAL CHARACTERISTIC	NEUROPATHIC PAIN	NOCICEPTIVE PAIN
Etiology	Nerve injury or peripheral or central sensitization	Tissue or potential tissue damage
Descriptors	Lancinating, shooting, electrical-like, stabbing	Throbbing, aching, pressure-like
Sensory deficits	Frequent (e.g., numbness, tingling, pricking)	Infrequent and, if present, in nondermatomal or non-nerve distribution
Motor deficits	Neurologic weakness may be present if motor nerve affected	May have pain-induced weakness
Hypersensitivity	Pain frequently evoked with nonpainful (allodynia) or painful (exaggerated response) stimuli	Uncommon except for hypersensitivity in the immediate area of an acute injury
Character	Distal radiation common	Distal radiation less common; proximal radiation frequent
Paroxysms	Exacerbations common and unpredictable	Exacerbations less common and associated with activity
Autonomic signs	Color changes, temperature changes, swelling, or sudomotor (sweating) activity occurs in one third to one half of patients	Autonomic signs uncommon in chronic nociceptive pain

symptoms reflect abnormal excitability of the nervous system. Numbness, tingling, and other symptoms suggestive of sensory dysfunction are strongly indicative of neuropathic pain, especially when they occur in a dermatomal or nerve distribution. Descriptors such as "burning," "shooting" and "electrical" are more likely to be associated with neuropathic pain, whereas adjectives such as "throbbing" and "aching" tend to be identified with nociceptive pain states such as arthralgias. Other positive symptoms observed in neuropathic pain states include pain evoked by normally innocuous stimuli (allodynia), and an exaggerated or prolonged pain response to noxious stimuli (hyperalgesia, hyperpathia). Although neuropathic pain tends to be more intermittent than nociceptive pain, mechanical spinal pain is classically exacerbated by movement. As alluded to earlier, some conditions, such as cancer, may be characterized by aspects of both nociceptive and neuropathic pain. There are several patient-report instruments available that may help distinguish neuropathic from nociceptive pain (e.g., painDETECT, s-LANSS, and DN4),[5] although determination by the physician through history, examination, and diagnostic testing remains the reference standard.[6]

A thorough history should evaluate sleep patterns. In epidemiologic studies, more than 50% of chronic pain patients exhibit some form of sleep disturbance, and sleep abnormalities are nearly ubiquitous in some conditions such as fibromyalgia. It is well known that pain can interfere with sleep, but what is not as commonly appreciated is that sleep deprivation may enhance pain sensitivity by lowering nociceptive thresholds and reduce response to analgesic therapy. Evidence also exists for a relationship between poor sleep and the development of chronic pain after an acute event. Although there is some evidence that improved sleep may lessen chronic pain, these studies are mostly correlational.

The proper evaluation of the patient in pain must include a psychosocial history. Between one half and two thirds of chronic pain patients exhibit varying degrees of psychopathology, with depression being the most common comorbidity, followed by anxiety disorders, somatoform disorders, and substance abuse. Many of these coexisting psychological conditions have been associated with poor treatment prognosis. People seeking chronic pain care may also be more likely to carry a concomitant axis II diagnosis (i.e., personality disorder), which can act as an additional barrier to effective treatment. Potential social factors that can have a negative impact on treatment should be identified, including low job satisfaction and secondary gain. A focused psychosocial history that includes prior psychiatric diagnoses, suicidal ideation, work history, legal history, and substance abuse is therefore essential in the formulation of a treatment plan.

Physical Examination

Examination of the patient in pain should encompass all body systems because pain is a frequent manifestation of systemic disease. A physical examination finding by itself is almost never pathognomonic but usually functions to confirm suspicions garnered from the history and to select patients for imaging studies or invasive diagnostic testing. Unlike acute pain, chronic pain is usually not associated with autonomic activation resulting in altered vital signs or facial grimacing.[7] Sensory symptoms can precede other neurologic findings by months or weeks. The most common forms of neuropathy are associated with sensory deficits in a glove-stocking distribution, but other patterns occur as well (Chapter 420). Numbness in the distribution of a nerve root or single nerve strongly suggests neuropathic pain, but nondermatomal sensory changes can accompany nociceptive pain (e.g., fibromyalgia) as well. Allodynia and hyperalgesia are hallmarks of neuropathic pain. Postural and gait abnormalities may be either causative factors for rheumatologic conditions (e.g., bursitis) or consequences of the underlying condition.

A careful evaluation of passive and active range of motion is useful because generalized and regional pain complaints are often accompanied by deconditioning. Distinctions should be drawn between pain-induced and neurologic weakness, with the latter often occurring in conjunction with muscle atrophy or asymmetry in reflexes. Sometimes reflex assessment is the only way to distinguish between a true neurologic condition and nonorganic causes.

Diagnostic Tests

Imaging has largely supplanted history and physical examination as the "gold standard" for diagnosis of disease but is not without drawbacks. There is a poor correlation between findings on magnetic resonance imaging (MRI) and the intensity of spinal pain, with more than 50% of asymptomatic individuals having abnormalities on lumbar, thoracic, and cervical films. Systematic reviews have found that early imaging for back pain does not improve outcomes or affect decision making and should be reserved for those patients with indications of a serious underlying condition.■ Absolute indications for MRI in patients with back pain are serious or progressive neurologic deficits, new-onset bowel and bladder dysfunction, suspected metastatic disease or infection, and referral of patients for procedural interventions. Yet, some have even questioned the utility of imaging as a prognostic tool before procedures. In a randomized trial evaluating whether MRI improved treatment outcomes in patients with sciatica who were referred for epidural steroid injections, preprocedure MRI was not shown to improve outcomes and only infrequently altered decision making.■ For back pain, the presence of "red flags" suggestive of more serious disease (e.g., trauma history, infection, history of intravenous drug abuse) should also alert the practitioner to seek further evaluation. For nonspinal pain conditions, MRI is ideal for discerning inflammation and soft tissue disease, whereas computed tomography scanning may be better for detecting bleeding and bone disease. The main advantage of ultrasound is that it is inexpensive and not associated with radiation. For chronic pain disorders, such as primary headaches and chronic abdominal or pelvic pain, imaging should be reserved for the evaluation of acute processes or a significant change in symptoms.

Electromyography and nerve conduction studies (Chapter 420) can be used to diagnose injury to large nerve fibers. However, because these studies are associated with significant false-negative and false-positive rates and are not sensitive in detecting impairment of small-fiber function, normal findings on a neurophysiologic study do not rule out neuropathic pain. For small-fiber neuropathies, a skin biopsy demonstrating decreased density of epidermal nerve fibers is a sensitive test.

TREATMENT

Rx

The goals of treatment should include elucidating the cause of pain and alleviating suffering. The biopsychosocial model posits that biologic, psychological, and social factors all play a role in chronic pain and should be addressed. Consequently, the goal of pain treatment should not be limited to pain reduction but also encompass improving function, psychological comorbidities, sleep, and social interaction. It has been argued that treatment should be mechanism based rather than etiology based, but at present, simple clinical tools to correlate symptoms and signs with mechanisms are lacking. The future development of diagnostic methods to identify mechanisms (e.g., intravenous

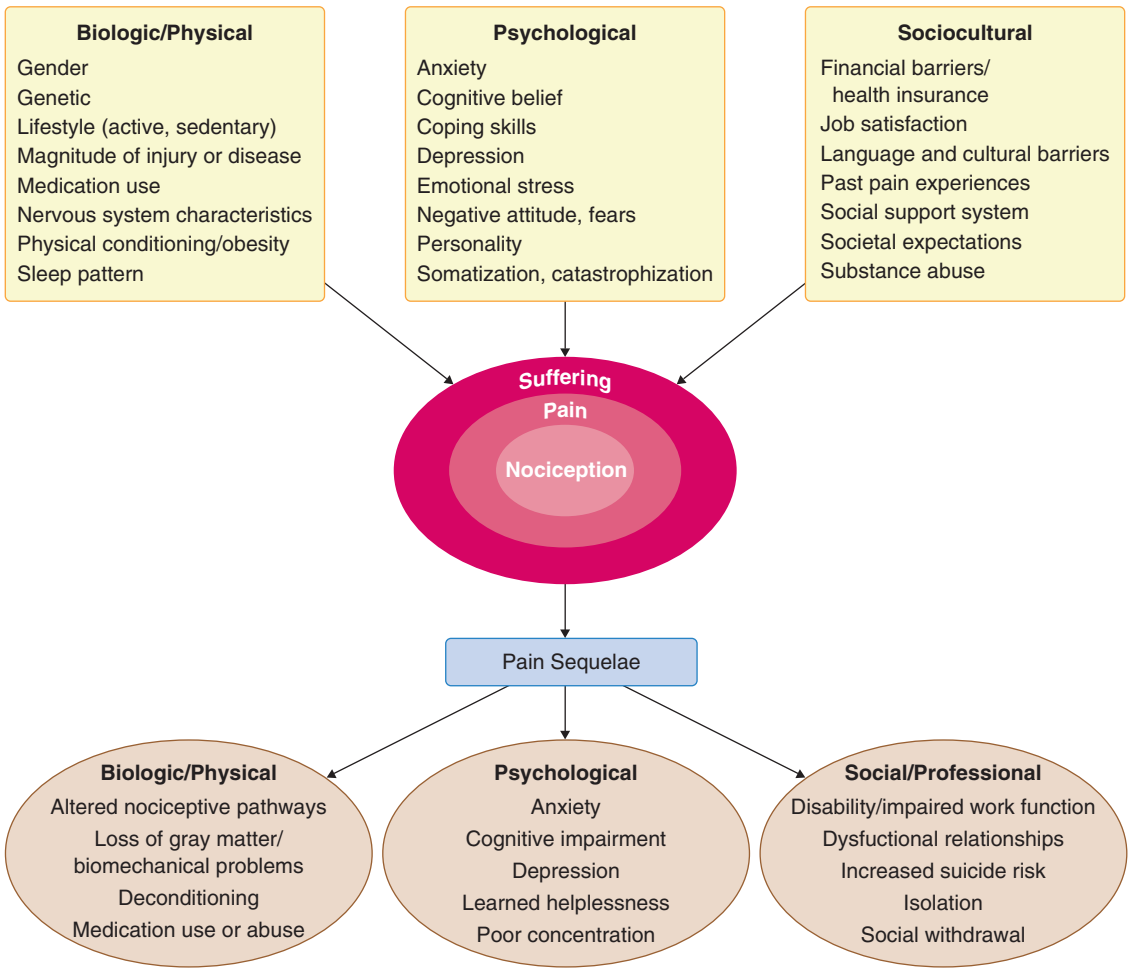

FIGURE 30-1. Schematic drawing of the biopsychosocial model of pain.

infusion tests) may help develop novel target-specific pharmacologic agents (Fig. 30-1).

Pain is a complex perceptual experience affected by a multitude of factors that include not only activation of nociceptors but also emotions, memory and cognition, social and cultural context, and expectations. Therefore, despite a paucity of studies evaluating a multidisciplinary approach to pain management, a strong consensus exists that this approach is beneficial.

Pharmacologic Therapies

Antipyretic Analgesics

Aspirin (Chapter 37) is the most widely used analgesic in the world. Along with its pharmacologic cousins nonsteroidal anti-inflammatory drugs (NSAIDs) and the antipyretic drugs acetaminophen and phenacetin, this group forms the backbone of pharmacologic pain treatment. Antipyretic analgesics exert their antinociceptive effects by the inhibition of cyclooxygenase, the rate-limiting enzyme in the production of prostaglandins, which sensitize nociceptors and regulate inflammation. There are several pharmacologic distinctions between NSAIDs and their counterparts, phenazone and acetaminophen. Whereas NSAIDs act both centrally and peripherally, making them effective topical agents for nociceptive inflammatory conditions, the primary site of enzyme inhibition for acetaminophen is in the central nervous system. Acetaminophen is also a weaker analgesic than NSAIDs and is largely devoid of anti-inflammatory effects.

The main drawback of nonopioid antipyretic analgesics is their ceiling effect, which can render them ineffective as stand-alone agents for severe pain. For cancer pain, the World Health Organization treatment paradigm advocates adding opioids to an analgesic regimen uncontrolled by NSAIDs, not replacing them. NSAIDs may act synergistically with opioids and have proven opioid-sparing effects. It is well acknowledged that aspirin, NSAIDs, and acetaminophen are more effective in treating nociceptive than neuropathic pain, although many neuropathic pain sufferers regularly take NSAIDs.

A major concern about nonopioid antipyretic agents is side effects. For NSAIDs, these include bleeding, gastrointestinal ulceration, renal toxicity, and increased risk of cardiovascular events (Chapter 37). Although the use of cyclooxygenase 2–selective inhibitors, such as celecoxib, or the addition of protein pump inhibitors to conventional NSAIDs may attenuate the risk of bleeding and ulcers, they do not affect the incidence of adverse renal or cardiovascular events. These risks are significantly increased in the elderly, in patients with multiple comorbidities, and with polypharmacy. In view of their considerable risks, NSAIDs should be prescribed in the lowest dose possible, for the shortest duration of time, and with periodic surveillance. Because it possesses a more favorable safety profile, acetaminophen is often considered a first-line therapy before NSAIDs, even for pain conditions associated with inflammation. Topical NSAIDs (e.g., diclofenac) have also been shown in randomized controlled studies to be beneficial for the treatment of rheumatologic disorders and have been touted as having comparable efficacy for regional pain conditions but with fewer adverse effects than systemic NSAIDs.

Adjuvant Analgesics

Multiple evidence-based guidelines for the treatment of chronic pain states, particularly neuropathic pain, have been published (Table 30-3). In general, these suggest that antidepressants and anticonvulsants should be the two first-line classes of medications for chronic neuropathic pain.[8,9] Depending on the particular drug and condition, these medications have been demonstrated to provide significant pain relief in carefully selected patients beyond that observed with placebo in 10 to 40% of ideal pharmacologic candidates. Whereas opioids have shown similar efficacy for neuropathic pain, anticonvulsants and antidepressants carry a lower risk of serious adverse events and long-term tolerance, rendering them preferable for long-standing noncancer pain (Table 30-4). In terms of efficacy, tricyclic antidepressants (TCAs) are superior to serotonin-norepinephrine reuptake inhibitors, which in turn are more efficacious than serotonin-specific reuptake inhibitors. However, because of their more favorable side effect profile, serotonin-norepinephrine reuptake inhibitors may be more effective than TCAs in some patients. Among the various TCAs, amitriptyline is the most studied but is probably comparable in efficacy to its metabolite nortriptyline and cousin imipramine. However, the last two drugs' more favorable side effect profiles (e.g., less sedation and anticholinergic activity) make them the preferred choices for neuropathic pain. In patients who cannot tolerate TCAs, serotonin-norepinephrine reuptake inhibitors such as duloxetine can be beneficial. In addition to neuropathic pain, antidepressants have also been shown to be effective in headache prophylaxis, abdominal and pelvic pain, fibromyalgia, and musculoskeletal pain.

Anticonvulsants are effective for neuropathic pain by virtue of their membrane-stabilizing properties. Although anticonvulsant drugs may be

TABLE 30-3 GUIDELINES FOR PHARMACOLOGIC TREATMENT OF CHRONIC NEUROPATHIC PAIN

NeuPSIG (IASP, 2010)*,[10]	CANADIAN PAIN SOCIETY (2007)*,[11]	EFNS (2010)[8]
First Line†	**First Line**	**DIABETIC NEUROPATHIC PAIN**
TCA (nortriptyline, desipramine)	TCA: amitriptyline, nortriptyline, imipramine, desipramine	**First Line**
Gabapentin	Anticonvulsants: gabapentin, pregabalin	Duloxetine
Pregabalin	Carbamazepine for tic douloureux (idiopathic trigeminal neuralgia)	Gabapentin
Duloxetine		Pregabalin
Venlafaxine		TCA
Topical 5% lidocaine		Venlafaxine ER
		Second or Third Line
		Opioids
		Tramadol
Second Line	**Second Line**	**POST-HERPETIC NEURALGIA**
Opioids	SNRI: Duloxetine, venlafaxine	**First Line**
Tramadol	Topical lidocaine 5%	Gabapentin
		Pregabalin
		TCA
		Lidocaine plasters
		Second or Third Line
		Capsaicin
		Opioids
Third Line	**Third Line**	**TRIGEMINAL NEURALGIA**
Antiepileptics	Tramadol	**First Line**
Mexiletine	Opioid analgesics	Carbamazepine
Memantine		Oxcarbazepine
Antidepressants		**Second or Third Line**
Dextromethorphan		Surgery
Topical capsaicin		
	Fourth Line	**CENTRAL PAIN**
	Cannabinoids	**First Line**
	Methadone	Gabapentin
	SSRI: citalopram, paroxetine	Pregabalin
	Other anticonvulsants: lamotrigine, topiramate, valproic acid	TCA
		Second or Third Line
	Miscellaneous agents: mexiletine, clonidine	Cannabinoids (multiple sclerosis)
		Lamotrigine
		Opioids
		Tramadol (spinal cord injury)

*Guidelines not disease specific, but for neuropathic pain in general.
†Consider lower starting dosages and slower titration in geriatric patients.
EFNS = European Federation of Neurological Societies; ER = extended release; NeuPSIG (IASP) = Neuropathic Pain Special Interest Group of the International Association for the Study of Pain; SNRI = serotonin norepinephrine reuptake inhibitors; SSRI = selective serotonin reuptake inhibitors; TCA = tricyclic antidepressant.

better than antidepressants for prototypical "lancinating-type" neuropathic pain, antidepressants are more versatile in that they have proven benefit in myriad other pain conditions. Owing to their high efficacy and favorable side effect profiles, gabapentin and pregabalin are first-line agents for most forms of neuropathic pain. In addition to independent pain-relieving properties, randomized controlled trials suggest that these drugs may act synergistically with opioids and antidepressants, provide anxiolysis, and exhibit preemptive analgesic effects when administered before surgery.

When gabapentinoid drugs are ineffective or intolerable, alternative anticonvulsants that act by different cellular mechanisms, such as lamotrigine and oxcarbazepine, may be employed. For trigeminal neuralgia, carbamazepine remains the treatment of choice, although adverse effects, such as the risk of agranulocytosis, limit its utility for other conditions. Other classes of adjuvants that may be effective in certain contexts include topical creams (e.g., capsaicin), *N*-methyl-D-aspartate antagonists (e.g., dextromethorphan), skeletal muscle relaxants (e.g., baclofen), cannabinoids, and antiarrhythmics (mexiletine). Topical lidocaine patches have been shown to reduce pain and allodynia in patients with post-herpetic neuralgia, and anecdotal evidence has suggested that they may be useful in the treatment of certain types of back pain. A high-concentration (8%) topical patch of capsaicin (the pungent chemical in chili pepper) has been demonstrated in clinical trials to provide significant pain relief compared with placebo for post-herpetic neuralgia and HIV neuropathy. A single 1-hour application can result in attenuation of pain for up to 12 weeks.

Opioid Analgesics

Opioid analgesics are the cornerstone of treatment for cancer pain (Table 30-5). Although randomized studies have found that opioids are effective in noncancer pain conditions, several reviews[12,13] and guidelines have concluded that they provide long-term improvement in only a minority of individuals, and that their superiority to other analgesics and ability to improve function are limited or inconclusive. Maximizing the therapeutic effects of opioid analgesics requires careful attention to balancing the beneficial effects with the undesirable adverse effects, including addiction. Understanding the clinical pharmacology of opioids, including their relative potency, duration of action, bioavailability, and pharmacokinetics, is essential for rational use. Their use for chronic pain management is limited primarily by myriad side effects that include nausea, constipation, sedation, itch, respiratory depression, and endocrine deficiency leading to sexual dysfunction and accelerated osteoporosis. Aggressive management with stool softeners and agents that enhance bowel motility, such as docusate and senna, can minimize constipation. Whereas most opioids are devoid of end-organ toxicity, an exception is meperidine. A metabolite, normeperidine, can accumulate after several days of treatment, causing myoclonus and anxiety; at higher concentrations, confusion, delirium, and seizures can ensue. Opioids that are predominantly renally eliminated, such as morphine, should be used with caution in patients with renal dysfunction. Two metabolites of morphine, morphine-6-glucuronide, which contains analgesic properties, and morphine-3-glucuronide, which may amplify pain in certain contexts, can accumulate in patients with renal dysfunction and contribute to the adverse effects of morphine. Alternative drugs include fentanyl and methadone.

For treatment of acute pain or an exacerbation of chronic pain in hospitalized patients (e.g., sickle cell crisis), patient-controlled analgesia (PCA) provides a convenient means of administering opioids. Intravenous morphine, fentanyl, or hydromorphone is commonly administered with an infusion device wherein a basal infusion, bolus dose, lockout interval, and maximum dose per hour can be programmed. PCA devices are safe and allow patients to control their pain management with less dependence on health care providers. Studies comparing PCA with conventional administration of opioids have generally found PCAs to be associated with better pain relief and higher satisfaction rates, albeit with larger amounts of medication consumed.

Critical steps in managing chronic pain patients with opioids include appropriate evaluation, clear and detailed documentation, a function-based treatment plan with well-defined patient goals that include an exit strategy, a written patient-physician agreement that includes informed consent and patient education, periodic review that focuses on progress toward functional goals, and specialist referrals in managing difficult patients. When considering a trial with opioids, health care providers need to consider the disease process, to assess traits such as compliance and responsibility, to perform risk stratification, and to monitor for predefined treatment goals and aberrant drug-related behaviors (Fig. 30-2).

The long-term use of opioids can be associated with tolerance and physical dependence. Cross-tolerance among opioids is not complete, and a strategy often used when tolerance is suspected is rotation to an alternative opioid, which may result in a 30 to 50% reduction in equianalgesic dose. Rates of addiction (Chapters 31 and 34) in reported studies vary widely, from less than 5% to 50%, which reflects differences in populations studied, definitions, and surveillance methods. A synthesis of evidence suggests that among chronic pain patients receiving opioid therapy, up to 40% will exhibit aberrant drug-related behaviors, 20% will abuse their drugs, and around 10% may become addicted.

Tramadol and tapentadol represent a newer class of analgesic drugs that have a dual mechanism of action. Tramadol is a weak agonist that inhibits the reuptake of norepinephrine and serotonin. Along with the usual side effects associated with opioids, seizures have been reported with tramadol, and adverse drug interactions can occur in patients taking coumadin and selective serotonin reuptake inhibitors. Tapentadol also has a dual mode of action as a μ-opioid agonist and a norepinephrine reuptake inhibitor and is slightly stronger than tramadol. Tramadol is presently approved for moderate to moderately severe severe pain. Tapentadol is approved for moderate to severe acute pain, while its extended-release formulation is approved for moderate to severe chronic pain and diabetic neuropathy.

Butorphanol, nalbuphine, and pentazocine are opioid agonist-antagonist drugs that can antagonize the actions of μ-opioid agonists and cause psychotomimetic effects due to their actions on the κ-opioid receptor. These drugs should be used with caution, particularly in patients receiving other μ-opioid agonists as they may precipitate withdrawal or reduce the effectiveness of pure opioid agonists. Buprenorphine, a partial agonist at the μ-opioid receptor and an antagonist at other opioid receptors, is available in various formulations, including a once per week transdermal patch for pain, and in combination with naloxone for the treatment of opioid addiction.

Combination Therapies

Most clinical trials have studied individual drugs in specific chronic pain states. Because no drug is universally effective and most provide only partial

TABLE 30-4 ADJUVANT ANALGESIC DRUGS FOR CHRONIC PAIN

DRUG	DOSAGE	INDICATIONS	ADVERSE EFFECTS	COMMENTS
TRICYCLIC ANTIDEPRESSANTS				
Amitriptyline, imipramine, desipramine, nortriptyline	10-150 mg/day	Peripheral neuropathy, post-herpetic neuralgia, other types of peripheral neuropathic pain, central pain, facial pain, fibromyalgia, headache prophylaxis, irritable bowel syndrome, and chronic low back pain with or without radiculopathy	Sedation, dry mouth, confusion, weight gain, constipation, urinary retention, ataxia, cardiac conduction delay (QTc prolongation)	First-line agents for neuropathic pain and headache prophylaxis Secondary amine drugs (e.g., nortriptyline) have fewer side effects than tertiary amines (e.g., amitriptyline) Contraindicated in glaucoma
SEROTONIN-NOREPINEPHRINE REUPTAKE INHIBITORS				
Venlafaxine	75-225 mg/day	Peripheral neuropathy, headache prophylaxis	Sedation, dry mouth, constipation, ataxia, hypertension, hyperhidrosis	Dose adjustment in patients with renal dysfunction
Duloxetine	60-120 mg/day	Peripheral neuropathy, fibromyalgia	Sedation, dry mouth, constipation, hyperhidrosis	U.S. Food and Drug Administration (FDA) approved for fibromyalgia and diabetic neuropathy Contraindicated in glaucoma
ANTICONVULSANTS				
Gabapentin	600-3600 mg/day	Peripheral neuropathy, post-herpetic neuralgia, other types of peripheral neuropathic pain, central pain, pelvic pain, headache prophylaxis, radiculopathy, chronic postsurgical pain	Sedation, weight gain, dry mouth, ataxia, edema	First-line agent for neuropathic pain FDA approved for post-herpetic neuralgia Effective preemptively for postoperative pain
Pregabalin	150-600 mg/day	Peripheral neuropathy, post-herpetic neuralgia, central pain, fibromyalgia	Sedation, weight gain, dry mouth, ataxia, edema	First-line agent for neuropathic pain FDA approved for diabetic neuropathy, post-herpetic neuralgia. and fibromyalgia Effective preemptively for postoperative pain. Same mechanism of action as gabapentin
Carbamazepine	200-1600 mg/day	Facial neuralgias, diabetic neuropathy	Sedation, ataxia, diplopia, hyponatremia, agranulocytosis, diarrhea, aplastic anemia, hepatotoxicity, Stevens-Johnson syndrome	First-line agent and FDA approved for trigeminal & glossopharyngeal neuralgia Contraindicated in patients with porphyria and atrioventricular conduction block
Topiramate	50-400 mg/day	Headache prophylaxis, chronic low back pain with or without radiculopathy	Sedation, ataxia, diplopia, weight loss, diarrhea, metabolic acidosis, kidney stones	First-line agent and FDA approved for migraine prophylaxis Often used as appetite suppressant
CORTICOSTEROIDS (SYSTEMIC)				
	5-60 mg/day (prednisone)	Inflammatory arthritis, other inflammatory pain conditions (e.g., inflammatory bowel disease), traumatic nerve injury, complex regional pain syndrome	Myriad psychiatric, gastrointestinal, neurologic, and cardiac side effects; immunosuppression, weakness, edema, weight gain, elevated glucose, poor wound healing, others	Stronger evidence supports local (i.e., injection) administration More effective for acute pain Strong anti-inflammatory effects
MISCELLANEOUS				
Muscle relaxants	Variable, depending on drug	Skeletal muscle spasm, acute spinal pain, temporomandibular disorder Baclofen effective for spasticity, dystonia, and trigeminal neuralgia	Sedation, ataxia, blurred vision, confusion, asthenia, xerostomia and other gastrointestinal effects, palpitations	First-line agents for acute back pain and skeletal muscle spasm
Lidocaine patch	1-3 patches every 12 hours	Post-herpetic neuropathy, peripheral neuropathy, other types of neuropathic and possibly myofascial pain associated with allodynia	Minimal systemic side effects when applied appropriately	Second-line agent and FDA approved for post-herpetic neuralgia
Capsaicin cream	0.025% applied 3 or 4 times per day	Post-herpetic neuralgia, peripheral neuropathy and other types of neuropathic pain, chronic postsurgical pain, arthritis and other musculoskeletal conditions	Burning on application Minimal systemic side effects when applied appropriately	FDA approved for arthritis Second-line agent for post-herpetic neuralgia and third-line agent for peripheral neuropathy
Cannabinoids	Variable, depending on drug and delivery route	Strongest evidence is for multiple sclerosis May be effective for peripheral neuropathy and other types of neuropathic pain and spasticity	Myriad psychiatric, neurologic and cardiac effects; xerostomia, abdominal pain, and other gastrointestinal effects	Fourth-line agent with narrow therapeutic index Modest analgesic effect comparable to codeine

pain relief, in clinical practice, two or more drugs are often used in combination. The rational use of polypharmacy should include drugs that act at different sites in the pain signaling process or modulate different neurotransmitter systems and ideally have antagonistic adverse effects (e.g., sedation and stimulation) (Fig. 30-3). For gabapentinoids, controlled trials suggest that combination therapy with either opioids or TCAs may be more effective than either drug alone.[■■] For antidepressants, controlled studies have failed to establish a benefit for combination therapy with opioids. In a placebo-controlled crossover study performed in a population with chronic low back pain, treatment with pregabalin and celecoxib was found to be superior to monotherapy.[■]

Psychological Treatment

The relationship between pain and psychopathology is complex. The lifetime prevalence of coexisting psychiatric illness in chronic pain patients

TABLE 30-5 FORMULATIONS, DOSAGES, AND PHARMACOLOGIC INFORMATION ON COMMONLY PRESCRIBED OPIOIDS

DRUG	EQUIANALGESIC DOSAGE (ORAL UNLESS SPECIFIED)	READILY AVAILABLE ROUTES OF ADMINISTRATION	DURATION OF ACTION	COMMENTS
PURE OPIOID AGONISTS				
Morphine	30 mg	IV, IM, PO, PR SR formulation	3-6 hr for short-acting 8-12 hr for SR	Reference standard for all opioids Renally excreted active metabolite
Oxycodone	20 mg	PO, PR SR formulation	3-6 hr for short-acting 8-12 hr for SR	Widely available in combination form with nonopioid analgesics Greater euphoric effects than morphine Both IR and SR forms popular among recreational users
Hydromorphone	3-6 mg	PO, PR, IV, IM SR formulation	3-6 hr 18-24 hr for SR	Higher PO : IV conversion ratio than other opioids SR form reserved for opioid-tolerant patients, and contraindicated in patients with recent monoamine oxidase inhibitor use
Hydrocodone	30-60 mg	PO SR formulation under development	3-6 hr 8-12 hr for SR	Wide variation in morphine-equivalent dose Most commonly prescribed opioid in the United States; typically used in combination form with nonopioid analgesic acetaminophen Formulations containing <15 mg hydrocodone are schedule III in the United States SR form being tested with and without acetaminophen
Oxymorphone	10mg	PO, IV, IM, Pr; SR formulation	4-8 hr, 12 hr for SR	Co-ingestion of alcohol with SR formulation can lead to "dose-dumping", or very high plasma levels and overdose. Very low (10%) oral bioavailability. Longer duration of action than morphine or oxycodone.
Methadone	2-20 mg	PO, PR, IV	6-12 hr for pain	Morphine:methadone conversion varies according to dose and length of opioid use, ranging from 2 : 1 to >20 : 1 in patients receiving very high doses Any physician with a schedule II DEA license may prescribe for pain May take 5-7 days to reach steady state because of extended half-life (i.e., accumulation) Electrocardiographic monitoring recommended with higher doses Other properties, such as NMDA receptor antagonism and reuptake inhibition of serotonin and norepinephrine, may slow the development of tolerance and increase efficacy for neuropathic pain
Fentanyl	12.5 µg/hr (TD) 800-1000 µg (TM) 200-400 µg (B)	TD, TM, B	72 hr for TD 1.5-3 hr for TM and B	TD, TM, and B formulations may be useful in patients with poor bowel function TD: Wide variation in conversion ratios; delivery system may be associated with fewer gastrointestinal side effects TM and B: delivery systems associated with more rapid (10 min) onset than IR oral opioids U.S. Food and Drug Administration approved for breakthrough cancer pain in opioid-tolerant patients
Codeine	200 mg	PO, PR, IM SR codeine combination products available as cough suppressant	3-6 hr 12 hr for SR	Often used in combination with nonopioid analgesics Efficacy and side effects may be affected by rate of metabolism to active metabolite morphine, which varies significantly Popular as cough suppressant
Propoxyphene	200 mg	PO, PR	3-6 hr	Wide variation in morphine-equivalent dose Often used in combination form with nonopioid analgesic Toxic metabolite may accumulate with excessive use, especially in elderly Weak antagonist at NMDA receptor
Meperidine	300 mg	PO, PR, IM, IV	2-4 hr	Toxic metabolite may accumulate with excessive use, especially in patients with renal insufficiency Associated with tachycardia and hypertension Concurrent use with monoamine oxidase inhibitors may result in fatal reactions May cause more "euphoria" than other opioids IM absorption erratic and injections painful
AGONIST-ANTAGONISTS, PARTIAL AGONISTS				
Buprenorphine Buprenorphine/ Naloxone (Suboxone) (4 : 1 ratio of buprenorphine to naloxone)	0.3-24 mg SL 5-70 µg/hr (TD) 2-24 mg/day	SL, PR, IV, TD	6-8 hr 7 days for SR	Partial µ-agonist and κ-antagonist that may precipitate withdrawal in opioid-dependent patients receiving high doses Lower abuse potential and fewer psychomimetic effects than pure agonists Not readily reversed by naloxone Schedule III drug in the United States Primary use of SL preparation is to treat addiction Used in combination with naloxone (Suboxone, Subutex) for opioid dependence May prolong QT interval
Butorphanol	1 mg/ spray, repeat after 60-90 minutes (NS) 1-2 mg (IV or IM)	NS, IM, IV, PO	3-4 hr	Partial agonist and antagonist at µ-receptor and antagonist at κ-receptor Commonly used as nasal spray to treat migraine headache and less commonly for labor pain Significant abuse potential
Nalbuphine	1 : 1 parenteral ratio	SC, IM, IV	3-6 hr	Mixed agonist-antagonist, often used for labor and delivery Sometimes used to treat refractory opioid-induced pruritus
Levorphanol	4 mg	PO, IM, IV	4-8 hr	2 : 1 oral to IV conversion ratio Multiple mechanisms of action including inhibition of serotonin and norepinephrine reuptake, NDMA receptor antagonism, and σ-receptor agonism

TABLE 30-5 FORMULATIONS, DOSAGES, AND PHARMACOLOGIC INFORMATION ON COMMONLY PRESCRIBED OPIOIDS—cont'd

DRUG	EQUIANALGESIC DOSAGE (ORAL UNLESS SPECIFIED)	READILY AVAILABLE ROUTES OF ADMINISTRATION	DURATION OF ACTION	COMMENTS
Pentazocine Pentazocine/ naloxone (Talwin) (100 : 1 ratio of pentazocine to naloxone)	90 mg	PO, SC, IM, IV	3-4 hr parenteral 8 hr PO	Mixed agonist-antagonist Naloxone added in 1970s to prevent abuse Also prescribed in preparation with acetaminophen
WEAK, DUAL-ACTION OPIOID AGONISTS				
Tramadol	150-300 mg	PO	4-6 hr 24 hr for SR	Dual action involves inhibition of serotonin and norepinephrine reuptake Affinity for μ-opioid receptor 6000× less than morphine Reduced side effects compared with morphine Analogue of codeine with active metabolite in which differences in metabolism may affect analgesia and side effects Available in combination form with acetaminophen Avoid concomitant use of serotonergic drugs Maximum recommended dose of 400 mg/day Not a federally controlled substance in the United States, although certain states have classified it as schedule IV
Tapentadol	75-110 mg	PO	4-6 hr 12 hr for SR	Dual action involves inhibition of norepinephrine reuptake May have fewer of certain side effects than morphine, such as gastrointestinal and respiratory depression Maximum dose 600 mg/day Avoid concurrent use of monoamine oxidase inhibitors

B = buccal; IM = intramuscular; IR = immediate release; IV = intravenous; NMDA = *N*-methyl-D-aspartate; NS = nasal spray; PO = oral; PR = rectal; SL = sublingual; SR = sustained release; TD = transdermal; TM = transmucosal.

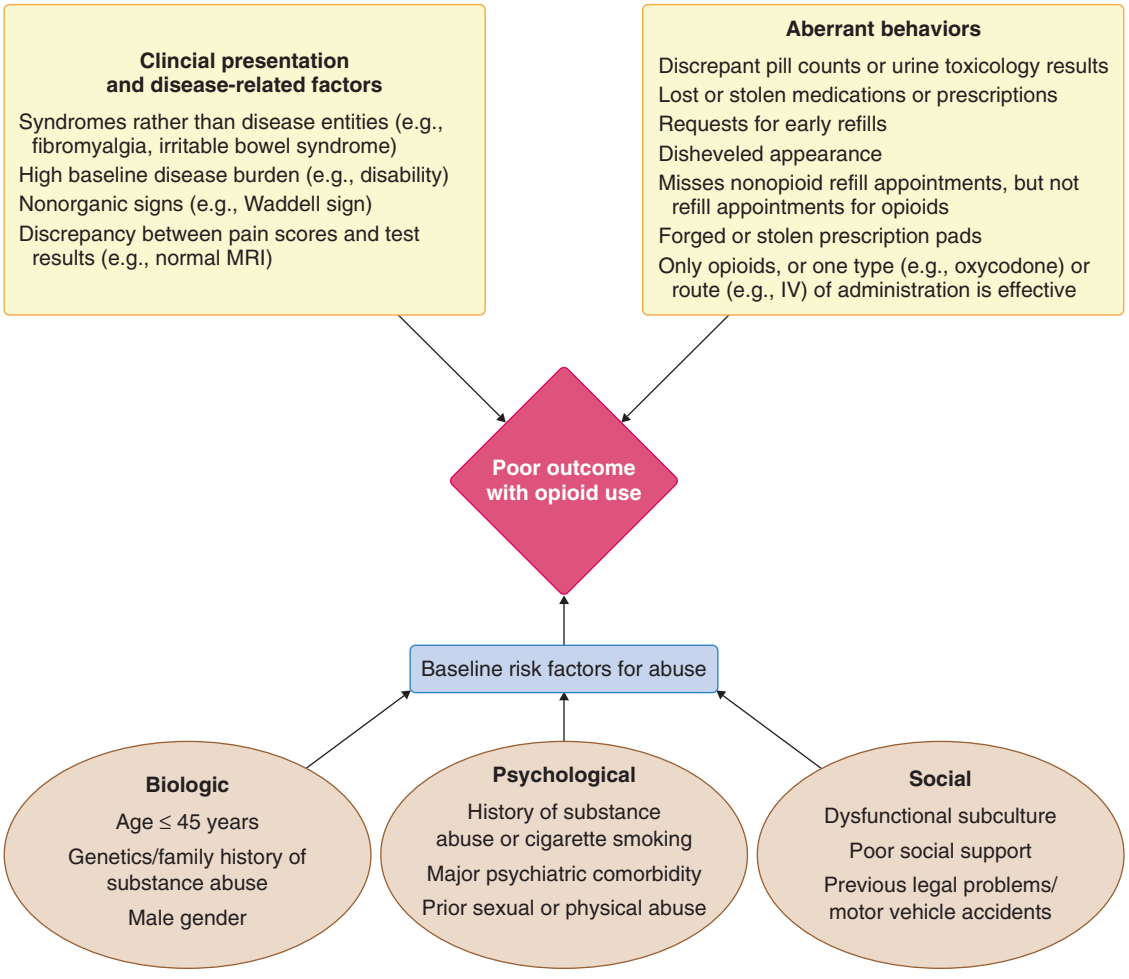

FIGURE 30-2. Factors associated with opioid abuse and treatment failure. IV = intravenous; MRI = magnetic resonance imaging.

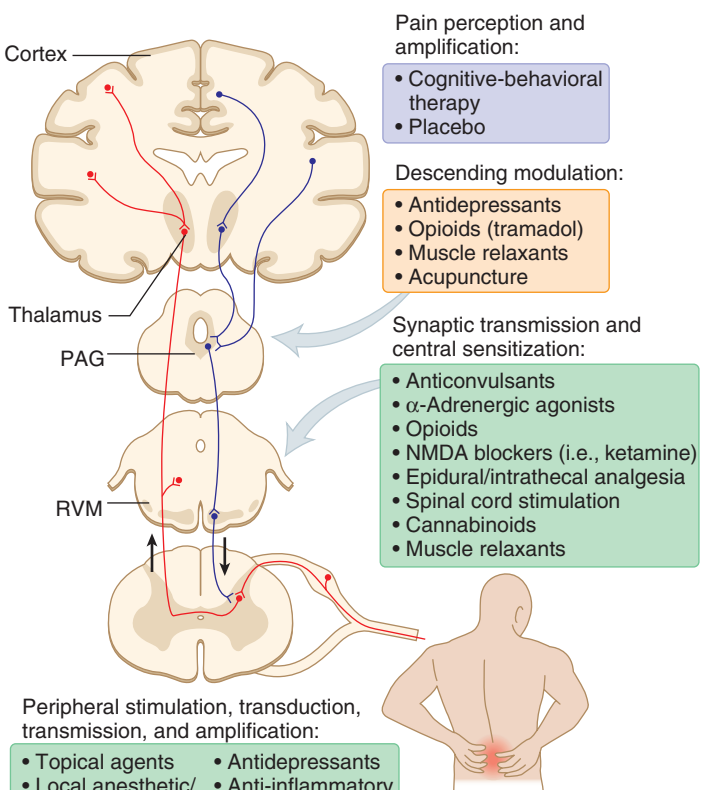

FIGURE 30-3. **Rational choice of combination therapies for pain should be based on the mechanisms of drug actions. Combining drugs with disparate actions can have additive or synergistic analgesic effects and minimize adverse effects. NMDA = *N*-Methyl-D-aspartate; PAG = periaqueductal gray; RVM = rostral ventromedial medulla.**

ranges from 50% to more than 80%. Between 30% and 60% of chronic pain sufferers experience symptoms of depression, making it the most common comorbidity. For anxiety disorders and substance abuse, co-prevalence rates are around 30% and 10 to 15%, respectively.

Viewed from a different perspective, the relationship is even more striking. More than 60% of patients with major depression and more than half of all patients with anxiety and substance abuse disorders experience chronic pain. Although it is widely acknowledged that chronic pain can predispose patients to depression, anxiety, and self-destructive behaviors such as substance abuse and suicide, what is less commonly appreciated is the effect that preexisting psychiatric conditions have on pain perception. There is a plethora of literature demonstrating that coexisting psychopathology is a strong predictor for the development of chronic pain after an acute, traumatic event (e.g., back pain episode, surgery) and is associated with poorer treatment outcomes.

It is incumbent on practitioners to screen all pain patients for psychological conditions that can adversely affect treatment (Table 30-6). Not only major psychiatric conditions, such as depression and generalized anxiety, but also maladaptive behaviors and secondary diagnoses, such as somatization disorder and poor coping skills, can negatively influence treatment.

Relaxation techniques such as biofeedback, self-hypnosis, and guided imagery have proved effective in a wide array of acute and chronic pain conditions but may be especially useful in those with high levels of anxiety. Cognitive-behavioral therapy is a highly structured form of psychotherapy predicated on the replacement of negative thought patterns and behaviors with more constructive ones. These therapies enhance modulation of pain signals and can also be effective for commonly associated symptoms such as fatigue and sleep abnormalities. Ideal candidates include educated, motivated patients in whom distorted thinking (e.g., catastrophization) and counterproductive behaviors serve to amplify pain behavior. In patients with personality disorders and ingrained maladaptive behaviors, long-term psychotherapy may be necessary.

Interventional Therapies: Nerve Blocks, Neuromodulation, and Surgery

Nerve Blocks

Injections may be performed for therapeutic, diagnostic, and sometimes prognostic purposes. Mechanistically, injections performed with local anesthetic may work by releasing entrapped nerves, enhancing blood flow, and interrupting processes involved in central sensitization (i.e., "breaking the cycle of pain"). Additional benefits of adding corticosteroid to local anesthetic include blocking the inflammatory cascade, suppressing ectopic discharges from injured nerves, and inhibiting the synthesis of prostaglandins, some of which serve to sensitize nociceptors.

Nerve blocks are almost never a panacea for noncancer pain but in appropriate candidates may provide intermediate-term pain relief, facilitate rehabilitative therapy, and improve quality of life for several weeks to months. Translating this relief into long-term improvement necessitates addressing the underlying causes and predisposing factors, which often entails psychotherapy and rehabilitation. Injections that can afford benefit in well-selected individuals include trigger point injections with local anesthetic for myofascial pain, intra-articular injection of steroids or viscosupplementation for chronic osteoarthritis, and nerve blocks with corticosteroid for entrapment syndromes (e.g., carpal tunnel syndrome). Among spinal injections, the strongest evidence is for epidural steroid injections in patients with radicular pain less than 6 months in duration, and radiofrequency denervation for facet joint pain. Neuroablative procedures such as celiac plexus neurolysis have been shown to provide significant pain relief lasting several months in patients with pain associated with upper abdominal cancers.

Electrical Stimulation

Electrical stimulation in various forms has been used for millennia to treat pain. The most commonly used type of electrical stimulation is transcutaneous electrical nerve stimulations (TENS), in which an electrical current is used to stimulate nerves for therapeutic purposes. The evidence supporting TENS to treat chronic pain is mixed, with one of the main criticisms being that any benefit is short-lasting. Spinal cord stimulation is a minimally invasive neuromodulatory technique for managing neuropathic pain states refractory to conservative measures in which an electrode(s) is inserted into the epidural space to stimulate the dorsal column. It was developed based on the gate-control theory, which postulates that activation of peripheral sensory A-fibers can attenuate pain signaling by slower-conducting pain C-fibers. Common indications include failed back surgery syndrome, complex regional pain syndrome, and outside of the United States, ischemic pain associated with peripheral vascular disease or angina. In patients with intractable chronic pain unresponsive to conservative measures, brain stimulation techniques have also shown promise. Deep brain stimulation has been demonstrated in uncontrolled studies to provide benefit for chronic neuropathic and nociceptive pain conditions, including phantom limb pain and cluster headache, whereas motor cortex stimulation has been successfully employed for trigeminal neuralgia and central pain.

Surgery

Surgical interventions are often advocated in chronic pain patients who have failed to respond to more conservative measures. A traumatic neuroma is an inexorable consequence of cutting or burning of a nerve, formed as a result of unregulated and disorganized nerve regeneration. Neuromas have been shown to fire ectopic pain signals and are often quite painful. Hence, neurolytic procedures are rarely successful in the long-term treatment of neuropathic

TABLE 30-6 PSYCHOSOCIAL FACTORS ASSOCIATED WITH CHRONIC PAIN

Multiple pain complaints
Poor job satisfaction, low pay
Inadequate coping skills
Fear-avoidance behavior
Manual labor, physically stressful job
Obesity
Somatization
Smoking
Low baseline activity levels
Ongoing litigation
Older age
Low education level
Higher presenting pain intensity, disability
Neurological symptoms
Anxiety
Depressed mood
Emotional distress

TABLE 30-7 EVIDENCE FOR TREATMENTS OF DIFFERENT CAUSES OF CHRONIC LOW BACK PAIN

CONDITION	PREVALENCE	TREATMENT
Lumbosacral radiculopathy from herniated disc	Annual prevalence 5-15%	Moderate evidence that epidural steroids may provide short-term relief, weak evidence for long-term benefit Weak evidence for percutaneous intradiscal procedures except chymopapain (strong evidence for a modest effect) Strong evidence that surgery may provide benefit for up to 2 years, but conflicting evidence for long-term benefit Negative or weak evidence for pharmacologic treatment
Spinal stenosis	5-10% of adults ≥ 65 years Rare in patients < 50 years	Moderate evidence that epidural steroids may provide short-term relief, weak evidence for long-term benefit Weak evidence for percutaneous therapy Strong evidence that surgery may provide benefit for at least 2 years Negative or weak evidence for pharmacologic treatment
Discogenic pain from degenerative disc disease	20-40% of patients with axial low back pain	Weak, conflicting evidence for modest short-term benefit with intradiscal treatments Weak, conflicting evidence that surgery may provide modest benefit for up to 2 years compared with no or unstructured treatment and that surgery is not more effective than structured care to include exercise and cognitive-behavioral treatment Negative evidence for epidural steroids
Facet arthropathy	10-15% of patients with axial low back pain, increasing with age	Moderate evidence for benefit with radiofrequency denervation Negative evidence for steroid injections and surgery
Sacroiliac joint pain	15-30% of patients with axial low back pain below L5 vertebra, increasing with age and inflammatory arthritis	Moderate evidence for short-term relief with steroid injections Moderate evidence for benefit with radiofrequency denervation Negative evidence for surgery
Myofascial pain	20%, but may be superimposed on a primary cause in more than 75% of patients	Strong evidence for exercise, muscle relaxants, and NSAIDs Weak, conflicting evidence for antidepressants Pharmacologic treatment more effective for acute than chronic low back pain
Vertebral fracture	6-18%, increasing with age and in certain ethnic groups (e.g., Asian women)	Conflicting evidence for vertebral augmentation Moderate extrapolated evidence for NSAIDs, although concerns exist for effect on bone healing Strong evidence for arthroses and bisphosphonates Anecdotal evidence for surgery, which may be necessary for stabilization Facet pain may play a role in pain after vertebral fracture

NSAIDs = nonsteroidal anti-inflammatory drugs.

pain. Part of the challenge in deciding when operative therapy is indicated revolves around the difficulty involved in establishing a causative relationship between the targeted pathologic process and pain. Roughly 10% of women of reproductive age have endometriosis (Chapter 236), but many patients with endometriosis have minimal symptoms, and pelvic pain is a frequent occurrence in young women with no detectable disease. With respect to inguinal hernia repair and spinal decompression, the incidence of chronic postsurgical pain is inversely correlated with the size of the bowel and disc herniation, respectively. This illustrates that in many individuals, the targeted disease may not be the primary cause of symptoms. On a similar note, scar tissue is a predictable sequela of surgical treatment, but lysis of adhesions is only infrequently associated with long-term symptom palliation because of the high recurrence rate and absence of any means to correlate the presence of adhesions with pain. Not surprisingly, surgery performed solely to remove a painful body part (e.g., hysterectomy, orchiectomy) rarely results in long-term benefit.

Back Pain

Back pain (Chapter 400) is the leading cause of disability in people younger than 45 years in the world. In patients who present with serious spinal disease (tumor, trauma), decompression, stabilization and fusion can be beneficial, but outcomes are strongly dependent on patient selection. Randomized studies have determined that decompression procedures done for radiculopathy and spinal stenosis are effective for short-term pain relief, but most demonstrate no long-term (>2 years) improvement compared with conservative treatment. With respect to fusion or disc replacement performed for mechanical pain associated with common degenerative changes, randomized trials suggest that less than one third of patients can expect significant pain relief or a highly functional outcome, with the results diminishing over time (Table 30-7).

Physical Treatments

The use of "physical" therapies to provide pain relief and to enhance function is a cornerstone in the multimodal approach to the patient with pain. Physical therapists evaluate, educate, and provide minimally invasive procedural interventions to help prevent and alleviate pain and dysfunction. These include addressing causative mechanisms of pain (e.g., correcting gait abnormalities) and providing treatments (e.g., hot and cold packs, joint manipulation).

Exercise has been used for decades as a treatment for chronic pain and a means to prevent injury. Exercise works through a variety of mechanisms, including enhancing blood flow, releasing endorphins, exerting anti-inflammatory effects, activating inhibitory pathways, and improving sleep and mood. Whereas the largest body of research has been conducted for spinal pain, benefits have also been demonstrated in fibromyalgia, headaches, arthritis, neuropathic pain, and cancer.

Complementary and Alternative Therapies

Patients are seeking complementary and alternative medical (CAM) treatments (Chapter 39) with increasing frequency, with utilization rates around 40%. Pain is the most common indication for CAM therapies. Some of the most popular CAM modalities are acupuncture, chiropractic, and yoga, all of which have been shown to be beneficial in certain contexts. However, the effect size tends to be modest for these treatments, and there is little evidence to support one modality over another or against conventional medical treatments.

Pain Management in Older Persons

Older adults are more likely to report pain than are younger cohorts and often present with multiple comorbidities. The increased risk for drug-related adverse effects in this population is the basis for the popular recommendation "start low and go slow" in the titration of analgesic drugs. Age-related physiologic factors that can decrease the therapeutic dose of opioid and nonopioid analgesics include changes in volume of distribution and protein binding, decreased metabolism, decreased excretion, and increased pharmacologic sensitivity. For these reasons, nonpharmacologic interventions, such as ergonomic modifications, tailored physical therapy and exercise programs, nutritional consultation, psychobehavioral approaches, and injections, should be considered in appropriate patients.[14]

FUTURE DRUGS AND PREVENTION OF PAIN

Much research is being devoted to development of new routes of drug administration (e.g., transdermal, transmucosal), abuse-deterrent opioids, and novel nonopioid drug treatments, which can optimize outcomes and reduce risks and side effects. Two other areas receiving significant research attention involve the genotyping and phenotyping of chronic pain patients. The conceptual appeal of these endeavors is that pain treatment tailored to individuals may result in greater benefit and less harm than the shotgun approach.

Regenerative therapies, which seek to facilitate the body's ability to repair, replace, restore, or regenerate diseased or damaged tissue, are another frontier in pain medicine. Whereas many of these treatments are currently in preliminary stages of development, they may someday be used to treat central, joint,

and spinal pain. Examples that are as yet unproven include stem cell therapy and the infusion of platelet-rich plasma, which involves stimulation of the body's natural healing processes.

Another area that has been underinvestigated is identifying patients at high risk for development of pain and employing strategies either to prevent pain or to minimize disease burden. In patients at high risk for chronic postsurgical pain (e.g., young patients with preexisting pain and psychological comorbidities), these measures might include the use of preemptive analgesics such as NSAIDs, anticonvulsants, and cytokine inhibitors and employing surgical techniques associated with less trauma. For patients with musculoskeletal complaints, this might entail educational initiatives and extensive rehabilitation.

Finally, the specialty of pain medicine must come to terms with the reality of spiraling health care costs by developing cost-effectiveness models by which to gauge benefit. Clinical trials evaluating new treatments should measure not only subjective outcomes such as pain scores but also objective ones that can result in societal cost savings, such as return to work or prevention of surgery. Proving that new treatments are more effective than existing or placebo treatments and result in cost-containment should be a cornerstone of future clinical research.

Grade A References

A1. Chou R, Fu R, Carrino JA, Deyo RA. Imaging strategies for low-back pain: systematic review and meta-analysis. *Lancet*. 2009;373:463-472.

A2. Cohen SP, Gupta A, Strassels SA, et al. Does MRI affect outcomes in patients with lumbosacral radiculopathy referred for epidural steroid injections? A randomized, double-blind, controlled study. *Arch Intern Med*. 2012;172:134-142.

A3. Derry S, Moore RA, Rabbie R. Topical NSAIDs for chronic musculoskeletal pain in adults. *Cochrane Database Syst Rev*. 2012;9:CD007400.

A4. Lunn MP, Hughes RA, Wiffen PJ. Duloxetine for treating painful neuropathy, chronic pain or fibromyalgia. *Cochrane Database Syst Rev*. 2014;1:CD007115.

A5. Moore RA, Derry S, Aldington D, et al. Amitriptyline for neuropathic pain and fibromyalgia in adults. *Cochrane Database Syst Rev*. 2012;12:CD008242.

A6. Moore A, Wiffen P, Kalso E. Antiepileptic drugs for neuropathic pain and fibromyalgia. *JAMA*. 2014;312:182-183.

A7. Chaparro LE, Wiffen PJ, Moore RA, Gilron I. Combination pharmacotherapy for the treatment of neuropathic pain in adults. *Cochrane Database Syst Rev*. 2012;7:CD008943.

A8. Clarke H, Bonin RP, Orser BA, et al. The prevention of chronic postsurgical pain using gabapentin and pregabalin: a combined systematic review and meta-analysis. *Anesth Analg*. 2012;115:428-442.

A9. Derry S, Sven-Rice A, Cole P, et al. Topical capsaicin (high concentration) for chronic neuropathic pain in adults. *Cochrane Database Syst Rev*. 2013;2:CD007393.

A10. Noble M, Treadwell JR, Tregear SJ, et al. Long-term opioid management for chronic noncancer pain. *Cochrane Database Syst Rev*. 2010;1:CD006605.

A11. Gilron I, Bailey JM, Tu D, et al. Morphine, gabapentin, or their combination for neuropathic pain. *N Engl J Med*. 2005;352:1324-1334.

A12. Gilron I, Bailey JM, Tu D, et al. Nortriptyline and gabapentin, alone and in combination for neuropathic pain: a double-blind, randomised controlled crossover trial. *Lancet*. 2009;374:1252-1261.

A13. Romano CL, Romano D, Bonora C, et al. Pregabalin, celecoxib, and their combination for treatment of chronic low-back pain. *J Orthop Traumatol*. 2009;10:185-191.

A14. Jacobs WC, van Tulder M, Arts M, et al. Surgery versus conservative management of sciatica due to a lumbar herniated disc: a systematic review. *Eur Spine J*. 2011;20:513-522.

A15. Kovacs FM, Urrútia G, Alarcón JD. Surgery versus conservative treatment for symptomatic lumbar spinal stenosis: a systematic review of randomized controlled trials. *Spine (Phila Pa 1976)*. 2011;36:E1335-E1351.

A16. Jacobs W, Van der Gaag NA, Tuschel A, et al. Total disc replacement for chronic back pain in the presence of disc degeneration. *Cochrane Database Syst Rev*. 2012;9:CD008326.

A17. Vickers AJ, Linde K. Acupuncture for chronic pain. *JAMA*. 2014;311:955-956.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Institute of Medicine Report from the Committee on Advancing Pain Research, Care, and Education. Relieving Pain in America, A Blueprint for Transforming Prevention, Care, Education and Research. The National Academies Press; 2011. Available at: http://books.nap.edu/openbook.php?record_id=13172. Accessed January 29, 2015.
2. Tiwari V, Guan Y, Raja SN. Modulating the delicate glial-neuronal interactions in neuropathic pain: Promises and potential caveats. *Neurosci Biobehav Rev*. 2014;45C:19-27.
3. Denk F, McMahon SB, Tracey I. Pain vulnerability: a neurobiological perspective. *Nat Neurosci*. 2014;17:192-200.
4. Seminowicz DA, Wideman TH, Naso L, et al. Effective treatment of chronic low back pain in humans reverses abnormal brain anatomy and function. *J Neurosci*. 2011;31:7540-7550.
5. Tampin B, Briffa NK, Goucke R, Slater H. Identification of neuropathic pain in patients with neck/upper limb pain: Application of a grading system and screening tools. *Pain*. 2013;154:2813-2822.
6. Haanpää M, Attal N, Backonja M, et al. NeuPSIG guidelines on neuropathic pain assessment. *Pain*. 2011;152:14-27.
7. Kang JH, Chen HS, Chen SC, et al. Disability in patients with chronic neck pain: heart rate variability analysis and cluster analysis. *Clin J Pain*. 2012;28:797-803.
8. Attal N, Cruccu G, Baron R, et al. EFNS guidelines on the pharmacological treatment of neuropathic pain: 2010 revision. *Eur J Neurol*. 2010;17:1113-e88.
9. Kalso E, Aldington DJ, Moore RA. Drugs for neuropathic pain. *BMJ*. 2013;347:f7339.
10. Dworkin RH, O'Connor AB, Audette J, et al. Recommendations for the pharmacological management of neuropathic pain: an overview and literature update. *Mayo Clin Proc*. 2010;85(suppl):S3-S14.
11. Moulin DE, Clark AJ, Gilron I, et al. Pharmacological management of chronic neuropathic pain—consensus statement and guidelines from the Canadian Pain Society. *Pain Res Manag*. 2007;12:13-21.
12. Stein C, Reinecke H, Sorgatz H. Opioid use in chronic non-cancer pain: guidelines revisited. *Curr Opin Anaesthesiol*. 2010;23:598-601.
13. Nuckols TK, Anderson L, Popescu I, et al. Opioid prescribing: a systematic review and critical appraisal of guidelines for chronic pain. *Ann Intern Med*. 2014;160:38-47.
14. Rastogi R, Meek BD. Management of chronic pain in elderly, frail patients: finding a suitable, personalized method of control. *Clin Interv Aging*. 2013;8:37-46.

REVIEW QUESTIONS

1. Which one of the following describes central neuropathic pain?

A. A type of nociceptive pain
B. Most often seen after a stroke
C. A major cause of the pain of irritable bowel syndrome
D. Almost never genetic in etiology
E. Usually not responsive to anticonvulsants like gabapentin and pregabalin

Answer: B The most common overall cause of central pain is central post-stroke pain, occurring after about 8% of cerebrovascular accidents. Central pain and peripheral pain are the two major types of "neuropathic pain," whereas "nociceptive pain" usually results from an injury or disease affecting somatic structures like skin, muscle, tendons, bone, joints, and ligaments. The pain of irritable bowel syndrome is one of the visceral forms of nociceptive pain (see Table 30-1). Gabapentin and pregabalin are considered first-line therapies for central neuropathic pain (see Table 30-3). Heritability is a component of practically all types of pain, including central neuropathic pain, estimated to account for between 30 and 60% of the variance in pain response (see Genetics). See Neuropathic, Nociceptive, and Mixed Pain.

2. Which of the following is true regarding diagnostic testing for pain?

A. Imaging has little if any role in diagnosing the cause of pain.
B. MRI findings correlate well with the intensity of spinal pain.
C. Early imaging for back pain guides decision making and improves outcomes.
D. Normal neurophysiologic (EMG, nerve conduction) test results are useful to rule out a neuropathic etiology of pain.
E. MRI is ideal for discerning inflammation and soft tissue pathology as causes of pain.

Answer: E For nonspinal pain conditions, MRI is ideal for identifying inflammation and soft tissue pathology. (In contrast, CT scanning may be better for detecting bleeding and bone pathology as causes of pain.) There is poor correlation, however, between MRI findings and the intensity of spinal pain, with more than 50% of asymptomatic individuals having abnormalities on lumbar, thoracic, and cervical views. Systematic reviews have found that early imaging for back pain does not improve outcomes or affect decision making and should be reserved for those individuals with indications of a serious underlying condition (see reference A1). EMG and nerve conduction studies are associated with significant false-negative rates and are not sensitive in detecting impairment of small fiber conduction; therefore, normal results do not rule out neuropathic pain. See Diagnostic Tests.

3. Which of the following statements is true regarding opioid analgesics?

A. Most opioids are devoid of end-organ toxicity.
B. Opioids are comparably effective for long-term pain relief in cancer and noncancer patients.
C. Opioids delivered by patient-controlled analgesia (PCA) require intensive supervision.
D. Most patients on opioid therapy for chronic pain become addicted and/or abuse their drug.
E. Opioids are cleared primarily by the liver, not the kidneys.

Answer: A Most opioids are not associated with end-organ toxicity, with the notable exception being meperidine. Abuse of drugs and addiction are potentially serious complications of long-term opioid use but are estimated to occur in only 20% and 10% of such individuals, respectively. Morphine is one of the opioids eliminated predominantly by the kidneys and should be used with caution in patients with renal dysfunction. Opioid analgesics are the cornerstone of treatment for cancer pain. In contrast, several recent reviews and guidelines have concluded that they provide long-term improvement in only a minority of individuals with noncancer pain conditions (see reference A10). PCA devices are safe; are programmed with basal infusion, bolus dose, lockout interval, and maximum dose per hour controls; and allow patients to control their pain management with less dependence on health care providers. See Opioid Analgesics.

31

BIOLOGY OF ADDICTION

ERIC J. NESTLER AND STEVEN E. HYMAN

DEFINITION

Drug addiction is compulsive substance use despite serious negative consequences. Harmful drug use and addiction are significant contributors to medical morbidity and mortality both directly as a result of the toxic effects of abused drugs and indirectly through accidents, violence, nonsterile needle use, smoking, and other health hazards. By far the greatest contributors to illness and death are the widely used legal drugs, tobacco (Chapter 32), and alcohol (Chapter 33), although illegal addictive drugs and abused prescription drugs also exact a significant toll. In addition, addiction creates enormous burdens on society by impairing the function of the addicted person in multiple life roles, disrupting families and neighborhoods, and motivating crime. Globally, drug use is not distributed evenly and is not simply related to stringency of drug policy.[1]

Compulsive use, the cardinal feature of addiction, means that the affected person cannot control substance use for a significant time despite powerful reasons to do so, such as drug-related health problems, drug-associated arrests, or the threat of losing one's job or spouse. In the clinic, addiction can be remarkably frustrating to treat: drug seeking and administration are apparently voluntary behaviors that an otherwise sentient person seems unwilling to control. Even after significant efforts have been exerted to get an addicted patient into drug treatment, relapse is common, even long after the last withdrawal symptom has cleared. Relapses are often precipitated by stress or by reminders of drug use (cues) that may range from familiar drug use contexts (such as smoking after a meal) to interactions with drug-using friends, the smell of marijuana or tobacco smoke, and body feelings previously associated with drug seeking (so-called interoceptive cues). Molecular, cellular, and behavioral studies of drug action in animal models and noninvasive human neuroimaging studies are providing significant insights into the neurobiology underlying compulsive drug taking and its persistence. Other important frontiers of research include the human genetics of addiction risk and more recently, the neurobiology of what have been called behavioral addictions, such as compulsive gambling.

Drug users may repeatedly take drugs to gain pleasure or to escape from negative feelings, including the aversive feelings that may occur when drugs wear off. As a result of repeated drug administration over time, initial protein targets for the drugs and their downstream signaling pathways (Table 31-1) are excessively stimulated and may thus undergo homeostatic adaptations. These adaptations can produce tolerance (the need for increasing drug doses to achieve desired affects) or dependence (revealed by withdrawal symptoms between drug doses or with drug cessation). Both tolerance and dependence can contribute to ongoing drug use and to dosage increases; however, neither tolerance nor dependence alone explains compulsive use. First, tolerance and dependence occur not only with repeated use of many addictive drugs (e.g., heroin) but also with many nonaddictive drugs (e.g., β-adrenergic antagonists [propranolol], α_2-adrenergic agonist antihypertensive agents [clonidine], nitrates, selective serotonin reuptake inhibitor antidepressants). Second, some highly addictive drugs, such as cocaine, may produce little physical dependence and withdrawal in some individuals who nonetheless exhibit compulsive use. Finally, if dependence and withdrawal were necessary factors in addiction, the phenomenon of late, post-detoxification relapse would not be the major clinical problem that it is.[2] Although these forms of homeostatic adaptation play a role in addiction, as will be described, other types of plasticity within the nervous system are more significant.

RISK FACTORS FOR ADDICTION

Only a minority of individuals who use drugs go on to become addicted. The best-established risk factors for addiction are male sex and family history. Across countries and cultures, males have a greater risk for both heavy drug use and addiction, with risk ratios in the range of 1.4 : 1 to 2 : 1. In recent years, however, the sex ratios have narrowed in many countries, especially for tobacco use and alcohol. Moreover, females who succumb to addiction tend to do so more quickly after initial drug exposures, and drug use during pregnancy can have enormous deleterious effects on the fetus.

Genes play the preponderant role in familial risk as evidenced by twin and adoption studies.[3,4] Twin studies consistently show higher rates of concordance for heavy drug use and addiction within monozygotic twin pairs than within dizygotic twin pairs. Adoption studies that have been performed in several Scandinavian countries and in the United States have focused mostly on alcoholism. These studies demonstrate that individuals adopted early in life tend to resemble their biologic rather than their adoptive parents with respect to patterns of alcohol use. Large population genetic studies suggest that the heritable risk for addiction to any of several substances, including opiates, stimulants, nicotine, alcohol, and marijuana, is roughly the same and varies between 20 and 60%, depending on the study.

Although genes clearly play a significant role in vulnerability to addiction, few of the specific genetic variants that confer risk have been identified with certainty. Like all common neuropsychiatric disorders, addiction risk is highly

TABLE 31-1 PROPERTIES OF ADDICTIVE DRUGS

DRUG	NEUROTRANSMITTER	DRUG TARGET	EFFECT AFTER BINDING
Opiates (morphine, heroin, oxycodone)	Endorphins; enkephalins	μ and δ opioid receptor (agonist)	Activate G_i/G_o; activate K^+ channels
Psychostimulants (cocaine, amphetamines)	Dopamine (DA)	Dopamine transporter (DAT)* (antagonist)	Increase synaptic DA; stimulate presynaptic and postsynaptic DA receptors
Nicotine	Acetylcholine	Nicotinic acetylcholine receptor (nAChR) (agonist)	Stimulate cation channel (may desensitize)
Alcohol	γ-Aminobutyric acid (GABA) Glutamate	$GABA_A$ receptor (agonist) *N*-methyl-D-aspartate (NMDA) receptor (antagonist)	Activate Cl^- channel Inhibit Ca^{2+} entry
Marijuana (Δ^9-tetrahydrocannabinol)	Anandamide; 2-arachidonoylglycerol	Cannabinoid CB_1 (agonist)	Activate G_i/G_o; activate K^+ channels
Phencyclidine, ketamine		NMDA receptor (antagonist)	Inhibit Ca^{2+} entry

*The psychostimulants also interact with the norepinephrine and serotonin transporters, but under normal conditions, it is the DAT that is critical for rewarding and addictive properties. Unlike cocaine, amphetamines enter dopamine nerve terminals through the DAT and interact with a second target, the vesicular monoamine transporter (VMAT), to release DA into the cytoplasm and thence, through the DAT, to release it into the synapse.

Modified from Hyman SE, Malenka RC, Nestler EJ. Neural mechanisms of addiction: the role of reward-related learning and memory. *Annu Rev Neurosci.* 2006;29:565-598.

genetically complex; there is evidence from linkage and association studies for contributions by a large number of genetic variants of relatively small effect. Large genome-wide association studies and other applications of modern genomic methods are continuing. However, the task of gene identification is also complicated by the challenges of phenotype definition. There are no objective medical tests with which to make the diagnosis, and there may be independent genetic and nongenetic risk factors for different stages of substance use disorders, such as drug experimentation, addiction, and treatment responsiveness. Moreover, twin and family studies suggest that there may be both shared and unshared genetic risk factors underlying addictions to different drugs. As in other genetically complex disorders, it is hoped that with the identification of multiple risk-conferring variants, it will be possible to identify biochemical pathways involved in addiction pathogenesis, which will then suggest potential targets for new and more effective treatments.

REWARD CIRCUITRY: THE NEURAL SUBSTRATE OF ADDICTION

The survival and perpetuation of species require that animals, including humans, learn to predict threats and also learn the circumstances under which they can obtain "rewards," such as food, water, shelter, and opportunities for mating. A simple operational definition of reward is a stimulus that elicits approach and consummatory behaviors. Several interconnected neural circuits, highly conserved in evolution, control an individual's responses to rewarding and aversive stimuli (Fig. 31-1). Dopamine-releasing neurons in the ventral tegmental area (VTA) of the midbrain and their major target neurons in the nucleus accumbens (NAc) in the ventral striatum serve as a rheostat that detects and drives responses to rewards and threats, the amygdala and hippocampus are crucial for forming reward- and fear-related memories, the dorsal striatum (caudate and putamen) mediates well-learned behaviors and habits, and several regions of prefrontal cortex exert executive control over these subcortical systems. Addiction involves abnormal functioning of this entire circuit.[5]

The neurotransmitter dopamine, released from VTA nerve terminals in the NAc, plays the key (albeit not the only) role in binding rewards and reward-associated cues to adaptive reward-seeking responses. In animals, implanted electrodes can record firing of dopamine neurons; microdialysis catheters and electrochemical methods can be used to detect dopamine that has been released from presynaptic terminals. In humans, positron emission tomography permits indirect measures of dopamine release by observing the displacement of a positron-emitting D_2 dopamine receptor ligand previously bound to receptors after a stimulus or pharmacologic challenge. By use of such methods in multiple paradigms, it has been well established that natural rewards cause firing of VTA neurons and dopamine release in the NAc and other forebrain regions. When dopamine action is blocked, whether by lesioning of dopamine neurons, blocking of postsynaptic dopamine receptors, or inhibition of dopamine synthesis, rewards no longer motivate the behaviors necessary to obtain them.

New insights into the role of dopamine have emerged from studies of patients with Parkinson disease (Chapter 409). Parkinson disease results from the death of midbrain dopamine neurons; however, neurons within the substantia nigra, which project to the dorsal striatum, are more severely affected than neurons within the VTA. Patients are generally treated with

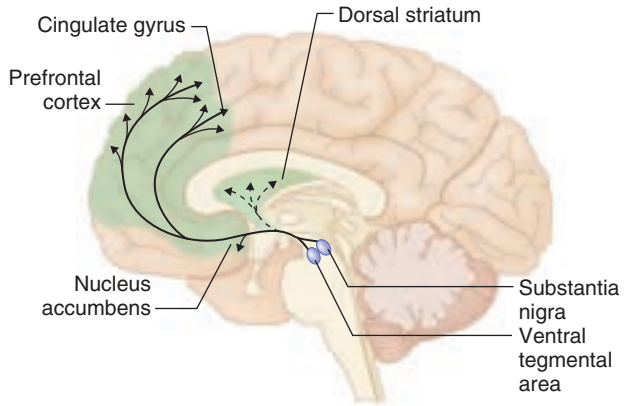

FIGURE 31-1. Brain reward circuits. The major dopaminergic projections to the forebrain that underlie brain reward are shown superimposed on a diagram of the human brain: projections from the ventral tegmental area to the nucleus accumbens, amygdala (not shown), hippocampus (not shown), and prefrontal cerebral cortex. Also shown are projections from the substantia nigra to the dorsal striatum (caudate and putamen and related structures) that play a role in habit formation and other deeply ingrained motor behaviors, including those related to drug consumption. (From Hyman SE, Malenka RC, Nestler EJ. Neural mechanisms of addiction: the role of reward-related learning and memory. *Annu Rev Neurosci.* 2006;29:565-598.)

L-dopa, a dopamine precursor, but as the disease progresses, other drugs may be needed, including selective D_2 dopamine receptor agonists. Relevant to this discussion is that a minority of patients treated with D_2 dopamine receptor agonists develop new risky, goal-directed behaviors, such as compulsive gambling or shopping. These behaviors generally cease when the drug is withdrawn. It is thought that whereas dopamine receptor agonists produce therapeutic effects on motor behavior in the more fully denervated dorsal striatum, they can combine with endogenous dopamine from preserved VTA neurons to overstimulate the NAc and other components of reward circuitry. These observations not only underscore the role of dopamine in motivation and reward seeking but also suggest that what have been called behavioral addictions share neural substrates with drug addiction.[6]

Much evidence suggests that the precise pattern of dopamine neuron firing and the resulting synaptic release of dopamine in forebrain circuits act to shape behavior so as to maximize future reward. In a basal state, dopamine neurons have a slow tonic pattern of firing. When a reward is encountered that is new, unexpected, or greater than expected, there is a phasic burst of firing of dopamine neurons causing a transient increase in synaptic dopamine. When a reward is predicted from known cues and is exactly as expected, there is little change from the tonic pattern of firing, that is, only small additional dopamine release. When a predicted reward is omitted or less than expected, dopamine neurons pause their firing to levels below their tonic rate. Phasic increases in synaptic dopamine signify that the world is better than expected, facilitate learning of new predictive information, and bind the newly learned predictive cues to action.

Dopamine is not the only neurotransmitter that signals reward. Others, including acetylcholine, endogenous opioid peptides (e.g., enkephalin and

endorphin), and endogenous lipid substances called endocannabinoids (because cannabinoid drugs like marijuana are agonists at their receptors), are also released in the reward circuitry in response to natural rewards.

PROPERTIES OF ADDICTIVE DRUGS

Addictive drugs are chemically diverse and interact with different molecular targets in the nervous system (Table 31-1). They also exhibit significant differences from each other in many of their physiologic and behavioral effects. For example, cocaine and amphetamines are stimulants; they increase arousal, may cause anxiety, and at lower doses enhance cognitive performance. Alcohol is a depressant, is anxiolytic at low doses, and degrades cognitive performance. Heroin and other opiates are analgesic and cause drowsiness, constipation, and pupillary constriction. The shared behavioral effect of all addictive drugs is the liability, in vulnerable individuals, of causing compulsive use. The shared pharmacologic property that is required to cause addiction is the ability to increase levels of synaptic dopamine in the NAc and other forebrain regions. For example, cocaine blocks the dopamine uptake transporter that normally clears dopamine from synapses. Amphetamines cause reverse transport of dopamine into synapses through the dopamine uptake transporter. Opiates, nicotine, alcohol, and cannabinoids cause dopamine release, acting by different initial mechanisms to stimulate VTA dopamine neurons or to release them from resting inhibitory control. Each of these other drugs also induces reward through nondopamine mechanisms, that is, through activating cholinergic, opioid, or cannabinoid receptors within the reward circuitry (Table 31-1).

Natural rewards, such as food or sexual opportunities, regulate the firing of dopamine neurons through highly processed sensory information, both external and interoceptive. Addictive drugs short-circuit this kind of information processing by acting directly on proteins that control dopamine and other signals in the reward circuitry. Acting by such direct pharmacologic mechanisms, addictive drugs typically produce greater quantities of synaptic dopamine and other reward-related neurotransmitters over longer times than natural rewards do. In addition, addictive drugs provide a grossly pathologic learning signal by occluding pauses in dopamine and other neuron firing even when drug use proves less pleasurable than expected or even aversive. For example, when the inhalation of a smoker causes painful coughing, it might seem that the brain would signal an experience that is worse than expected with a resulting decrement in VTA neuron firing rate. However, because nicotine causes dopamine release pharmacologically, independent of the smoker's actual experience, reward circuits, unavailable to conscious introspection, still receive a positive message that reinforces nicotine seeking and nicotine use. In short, addictive drugs, by virtue of their effects on dopamine and related neurotransmitters, always signal "better than expected."

DRUG-INDUCED NEURAL PLASTICITY RELEVANT TO ADDICTION

Neurobiologic research on addiction has focused intensely on compulsive aspects of drug use, the ability of specific cues to activate drug seeking and craving, and the long persistence of stress and cue-dependent relapse risk. As described previously, compulsive drug use and the power of drug-associated cues reflect the usurpation of the brain's reward circuitry by drugs of abuse.[7] The persistence of addiction reflects long-term changes in neurons and synapses and their interacting circuits. Research during more than a decade has identified long-term changes in gene expression resulting from use of addictive drugs; recently, some long-lived alterations in gene expression have been attributed to drug-induced epigenetic mechanisms, such as modifications of chromatin.[8]

Long-term changes in gene expression may render the addicted person susceptible to stress and may also create persistent changes in hedonic state and mood regulation that may motivate drug taking. By themselves, however, changes in gene expression do not explain the ability of exquisitely specific cues to activate drug seeking or, if seeking is impeded, intense subjective drug craving. The ability of specific cues to activate drug seeking and wanting is based on long-term associative memories consolidated under the influence of dopamine and other reward signals. Long-term memory formation represents perhaps the most persistent changes in brain function that may occur in adult life. The neural substrates of memory are likely to include alterations in synaptic weights, such as long-term potentiation or long-term depression, and physical remodeling of dendritic spines. Processes of drug sensitization (or reverse tolerance), demonstrated in animal models and in humans, may contribute to these memory-related phenomena.[9]

The centrality of associative learning mechanisms for addiction was first recognized from clinical observation: much drug taking and, most notably, late relapses follow exposure to cues previously associated with drug use. Cues that can reinitiate drug use include environmental stimuli (e.g., persons with whom drugs have been used, drug paraphernalia) and body feelings. Because addictive drugs reliably increase synaptic dopamine and other reward-related neurotransmitters as a result of their direct pharmacologic actions—indeed, they produce excessive and grossly distorted reward signals—the brain receives a powerful impetus to connect the circumstances in which the drugs have been used with the motivation to take drugs again. Even if the drug is no longer pleasurable, the signals continue to reinforce drug wanting and seeking. Moreover, the certainty and magnitude of these signals give drugs a marked advantage over natural rewards and other learned goals, including prosocial activities.

In the laboratory, it has been possible to study the effects of drugs and drug cues on neural circuits, physiology, and subjective responding in addicted human subjects. For example, drug-associated cues have been shown to elicit drug urges and physiologic responses (such as sympathetic activation) as well as activation of reward circuits in addicted human subjects. By positron emission tomography, cocaine-related cues have been shown to elicit dopamine release in the dorsal striatum in addicted subjects.

Investigations at the cellular and molecular levels have begun to identify the physiologic and molecular changes that underlie the effects of addictive drugs on reward-related memory processes. Among psychotropic drugs that have been examined, only those drugs that can cause addiction produce long-term potentiation in brain reward circuits including the VTA. Addictive drugs also activate transcription factors, such as the cyclic adenosine monophosphate response element binding protein (CREB), and alter the composition of activator protein 1 (AP-1) complexes (composed of Fos and Jun families of transcription factors) in brain reward circuits. Cocaine, opiates, and other addictive drugs have been shown to regulate many genes downstream of CREB, AP-1, and other transcription factors. As well, it has been increasingly possible, through the use of genetically modified mice and virus-mediated gene transfer, to directly demonstrate the involvement of these transcriptional mechanisms in the range of behavioral abnormalities induced by repeated drug exposure in animal models. However, it is far more challenging to determine which drug-regulated proteins are causally involved in human addiction. Part of the challenge is that, whereas rodent models have provided many insights into drug action and behavior, it is difficult to model human compulsion in animals, that is, to model a free-living and independent person who loses control over drug use while experiencing the negative consequences of that use. That said, a growing body of knowledge is emerging about the neural and molecular processes that produce addiction, with the hope of translating these advances into medical diagnostic tests and more effective treatments of addictive disorders.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Degenhardt L, Whiteford HA, Ferrari AJ, et al. Global burden of disease attributable to illicit drug use and dependence: findings from the Global Burden of Disease Study 2010. *Lancet*. 2013;382:1564-1574.
2. Steketee JD, Kalivas PW. Drug wanting: behavioral sensitization and relapse to drug-seeking behavior. *Pharmacol Rev*. 2011;63:348-365.
3. Haberstick BC, Zeiger JS, Corley RP, et al. Common and drug-specific genetic influences on subjective effects to alcohol, tobacco and marijuana use. *Addiction*. 2011;106:215-224.
4. Agrawal A, Verweij KJ, Gillespie NA, et al. The genetics of addiction—a translational perspective. *Transl Psychiatry*. 2012;2:e140.
5. Marinelli M, McCutcheon JE. Heterogeneity of dopamine neuron activity across traits and states. *Neuroscience*. 2014;282C:176-197.
6. Leeman RF, Potenza MN. A targeted review of the neurobiology and genetics of behavioural addictions: an emerging area of research. *Can J Psychiatry*. 2013;58:260-273.
7. Hyman SE, Malenka RC, Nestler EJ. Neural mechanisms of addiction: the role of reward-related learning and memory. *Annu Rev Neurosci*. 2006;29:565-598.
8. Robison AJ, Nestler EJ. Transcriptional and epigenetic mechanisms of addiction. *Nat Rev Neurosci*. 2011;12:623-637.
9. Kandel ER, Kandel DB. Shattuck Lecture. A molecular basis for nicotine as a gateway drug. *N Engl J Med*. 2014;371:932-943.

REVIEW QUESTIONS

1. In addition to dopamine, all of following are important mediators of reward signals in the brain EXCEPT which?

A. Acetylcholine
B. Endocannabinoids
C. Dynorphin
D. Endorphins and enkephalins

Answer: C Dynorphin, although an endogenous opioid peptide, does not activate μ (mu) and δ (delta) opioid receptors. Rather, it activates κ (kappa) opioid receptors, which mediate an aversive effect. See reference 7.

2. Which of following brain regions is NOT important in controlling responses to rewarding stimuli in the environment?

A. Amygdala
B. Occipital cortex
C. Hippocampus
D. Prefrontal cortex
E. Dorsal striatum

Answer: B All of the other brain regions listed are key components of the brain's reward circuitry. In contrast, the occipital cortex is most important for detecting and interpreting visual stimuli. See reference 7.

3. Which of the following statements best describes the activity of dopamine neurons in the ventral tegmental area in response to rewards?

A. The neurons show large increases in phasic firing in response to an unexpected reward.
B. The neurons show large increases in phasic firing in response to the withholding of an expected reward.
C. The neurons show large increases in tonic firing in response to the withholding of an expected reward.
D. The neurons show large decreases in tonic firing in response to an unexpected reward.

Answer: A Unexpected rewards induce an increase in phasic firing of the neurons. All other answers are incorrect responses. See reference 7.

4. Which of the following properties is unique to a drug of abuse?

A. It increases dopamine signaling from the ventral tegmental area to the nucleus accumbens.
B. It increases dopamine signaling from the substantia nigra to the dorsal striatum.
C. It decreases dopamine signaling from the substantia nigra to the dorsal striatum.
D. Its repeated use is associated with tolerance.
E. Its repeated use is associated with dependence and withdrawal.

Answer: A All drugs of abuse share only one property: increasing dopamine transmission from the ventral tegmental area to the nucleus accumbens. Choices B and C are incorrect. Although drugs of abuse do cause tolerance and dependence/withdrawal (choices D and E), many nonabused drugs do so as well. See reference 7.

5. Which of the following is NOT a known mechanism of a drug of abuse?

A. Cannabinoid receptor agonist
B. Nicotinic cholinergic receptor agonist
C. NMDA glutamate receptor antagonist
D. $GABA_A$ receptor agonist
E. Dopamine receptor antagonist

Answer: E All drugs of abuse increase dopamine signaling; a dopamine receptor antagonist would block some of the actions of a drug of abuse. All other answers are known mechanisms of a drug of abuse. See reference 7.

32

NICOTINE AND TOBACCO

TONY P. GEORGE

DEFINITIONS

Cigarette smoking is the most common (>90%) method of tobacco use, although other forms of tobacco use, including pipe tobacco, cigars, and smokeless tobacco, are common. Nicotine is the active ingredient in tobacco that acts as a reinforcer for repeated use for all forms of tobacco.

EPIDEMIOLOGY

Cigarette smoking is the most preventable cause of morbidity and mortality in the Western world. In the United States, approximately 20% of the general population currently uses tobacco compared with 47% in 1965. Since the

release of the U.S. Surgeon General's report in 1965, smoking prevalence has been substantially reduced, but this reduction appears to have slowed in recent years, likely because the remaining 20% of smokers are refractory to tobacco treatment. Approximately 450,000 people in the United States die each year as a result of smoking-attributable medical illnesses, including lung cancer, chronic obstructive pulmonary disease, cardiovascular disease, and stroke; and economic and health care costs of tobacco use exceed $400 billion annually. Smokeless tobacco (e.g., chewing tobacco) use has also increased, which has contributed to higher rates of oral pathologies, including precancerous oral lesions and cancers of the mouth and nasopharynx. Moreover, the health risks of environmental tobacco smoke (ETS) have become increasing clear, prompting the development of widespread tobacco bans in public settings.

Worldwide, it is estimated that approximately 1.1 billion people use tobacco on a regular basis, including approximately 65 million in the United States. Tobacco smoking is increasing rapidly throughout the developing world, and it is estimated that cigarette smoking will cause about 450 million deaths worldwide in the next 50 years.[1] In particular, the onset of smoking occurs at a younger age, the rates of smoking in women are increasing, and more smokers are of a lower socioeconomic status. Reducing smoking prevalence by 50% would prevent 20 to 30 million premature deaths in the first quarter of this century and 150 million in the second quarter.

For most smokers, quitting is the single most important thing they can do to improve their health.[2] A prospective cohort study in Norway suggests that even with sustained reductions (>25 to 75%) in daily smoking consumption, there is little if any decrease in cardiovascular disease and lung or other smoking-related cancer risk, further substantiating the merits of quitting versus reducing smoking.

PATHOBIOLOGY

Nicotine is the primary reinforcer in tobacco smoke, with contributions from more than 4000 components to the sensory (non-nicotine) aspects of cigarette smoking. The primary site of action of nicotine is the $\alpha_4\beta_2$ nicotinic acetylcholine receptor (nAChR), and the endogenous neurotransmitter acting on nAChRs is acetylcholine. nAChRs in the central nervous system (CNS) are pentameric ion channel complexes comprising two α- and three β-subunits; the seven α-subunits are designated α_2 to α_9 and the three β-subunits are designated β_2 to β_4. This produces considerable diversity in subunit combinations, which may explain the region-specific and functional selectivity of nicotinic effects in the CNS.[3] Activation of nAChRs leads to Na^+/Ca^{2+} ion channel fluxes and neuronal membrane depolarization. nAChRs are located presynaptically on several neurotransmitter-secreting neuron types in the CNS, including mesolimbic dopaminergic (DA) neurons that project from the ventral tegmental area (VTA) to the nucleus accumbens (NAc). Activation of nAChRs on mesolimbic DA neurons leads to DA secretion in the nucleus accumbens.

At low concentrations of nicotine, $\alpha_4\beta_2$ nAChR stimulation of afferent GABAergic projections onto mesoaccumbal DA neurons predominates, leading to reduced mesolimbic DA neuron firing and DA release. At higher nicotine concentrations, $\alpha_4\beta_2$ nAChRs desensitize, and predominant activation of α_7 nAChRs on glutamatergic projections occurs, leading to increased mesolimbic DA neuron firing and release. Within milliseconds of activation by nicotine, nAChRs desensitize. After overnight abstinence, nAChRs resensitize; this may explain why most smokers report that the first cigarette in the morning is the most satisfying. Interestingly, positron emission tomography (PET) neuroimaging studies have shown that smoking 2 or 3 puffs from a cigarette produces saturation of nAChRs in the brain reward system, suggesting that although binding to central nAChRs is an important first step in the effects of nicotine, it is not a complete explanation for continued smoking behaviors.

CLINICAL MANIFESTATIONS

Although there is a subset of cigarette smokers who do not smoke every day (e.g., "chippers"), most cigarette smokers are daily users and have some degree of physiologic dependence on nicotine. Smokers typically describe a "rush" and feelings of alertness, relaxation, and "satisfaction" when smoking, and it is well known that nicotine has both stimulating and anxiolytic effects depending on basal level of arousal. Airway stimulation is an important aspect of smoking behavior, and additives such as menthol enhance the experience by increasing the taste and reducing the harshness of smoked tobacco.

Interestingly, the positive effects of cigarette smoking (e.g., taste, satisfaction) appear to be mediated by non-nicotine components of tobacco such as tar. Besides positive reinforcement, withdrawal, and craving, there are several secondary effects of nicotine and tobacco use that may contribute to both maintenance of smoking and smoking relapse, including mood modulation (e.g., reduction of negative affect), stress reduction, and weight control. In addition, conditioned cues can elicit the urge to smoke even after prolonged periods of abstinence. Specific effects might be most relevant to smokers wishing to lose weight and to those with psychiatric presentations (mood modulation, cognitive enhancement, stress reduction). These secondary effects may present additional targets for pharmacologic intervention in certain subgroups of smokers (e.g., those with schizophrenia or depression, or those concerned about their weight).

DIAGNOSIS

The *Diagnostic and Statistical Manual*, 5th edition (DSM-5),[4] which was released in 2013 by the American Psychiatric Association, has changed the diagnostic terminology for nicotine and tobacco, eliminating the term *dependence* and instead using the term *tobacco use disorder*. Tobacco use disorder is established clinically by historical documentation of 2 of the following 11 criteria:

1. Tobacco often taken in larger amounts or over a longer period than was intended
2. Persistent desire or unsuccessful efforts to cut down or control tobacco use
3. A great deal of time spent in activities necessary to obtain or use tobacco
4. Presence of craving, or a strong desire or urge to use tobacco
5. Recurrent tobacco use resulting in failure to fulfill major obligations at work, school, or home
6. Continued tobacco use despite persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of tobacco
7. Important social, occupational, or recreational activities given up or reduced because of tobacco use
8. Recurrent tobacco use in situations in which it is physically hazardous (e.g., smoking in bed)
9. Continued tobacco use despite persistent or recurrent physical or psychological problems that are caused or exacerbated by tobacco use
10. Tolerance, as defined by either a need for markedly increased amounts of tobacco to achieve desired effects, or markedly diminished effects with continued use of the same amount of tobacco
11. Withdrawal, manifested by the presence of the characteristic tobacco abstinence syndrome (e.g., four of the following: irritability, anxiety, difficulty concentrating, increased appetite, restlessness, dysphoric mood, insomnia), or tobacco (or nicotine) taken to relieve or avoid tobacco withdrawal symptoms.

For abstinent smokers, remission is classified as early (between 3 and 12 months of abstinence) or sustained (>12 months of abstinence). Moreover, current severity of tobacco use disorder is coded as mild (2 or 3 symptoms), moderate (4 or 5 symptoms) or severe (6 or more symptoms).

In addition, most physiologically dependent tobacco smokers state that they smoke their first cigarette of the day within the first 5 minutes of awakening (e.g., time to first cigarette <5 minutes after awakening). Timeline follow-back procedures and smoking diaries have been used successfully to monitor smoking consumption over time. Scales such as the Fagerstrom Test for Nicotine Dependence allow assessment of the level of nicotine dependence, with scores higher than 4 on a scale of 0 to 10 being consistent with physiologic dependence to nicotine. Nicotine craving and withdrawal can be reliably monitored using validated scales such as the Tiffany Questionnaire for Smoking Urges and the Minnesota Nicotine Withdrawal Scale. These scales have excellent test-retest reliability and internal consistency in smokers with schizophrenia compared with nonpsychiatric control smokers, suggesting that they can be used in psychiatric populations.

TREATMENT

Rx

Psychosocial Treatments

Behavioral therapies are based on the theory that learning processes operate in the development, maintenance, and cessation of smoking (Table 32-1). Behavioral treatments for smoking can facilitate motivation to quit, provide an emphasis on the social and contextual aspects of smoking, and enhance overall success at smoking cessation.[5] In most reviews, 6-month quit rates with behavior therapies are 20% to 25%, and behavior therapy typically increases quit rates up to two-fold over standard medical advice. The primary

goals of behavioral therapies in treatment of tobacco dependence include providing necessary skills to smokers to aid them in quitting smoking and teaching skills to avoid smoking in high-risk situations.

Brief Interventions

Brief advice has been found to increase smoking cessation rates and has been strongly endorsed in the latest U.S. Department of Health and Human Services Guidelines on Tobacco Dependence Treatment. Therefore, it is recommended that physicians use the "5 As" with all patients (*Ask* patients if they smoke, *Advise* patients to quit, *Assess* patients' motivation level for quitting, *Assist* with quit attempts, and *Arrange* follow-up contacts). Providing self-help material is a form of brief intervention used to increase motivation to quit and impart smoking cessation skills. Several recent studies have documented that minimal behavioral interventions such as community support groups, telephone counseling, and computer-generated tailored self-help materials can augment smoking cessation rates in controlled settings.

Motivational Interventions

The goal of motivational interviewing (MI) interventions is to elicit change through addressing ambivalence, increasing intrinsic motivation for change, and creating an atmosphere of acceptance in which patients take responsibility for making changes happen. Brief MI interventions have been developed for smoking cessation, and there is some evidence for increased smoking cessation using MI techniques.

TABLE 32-1 BEHAVIORAL TREATMENTS FOR TOBACCO DEPENDENCE

BEHAVIORAL TREATMENTS	MECHANISM OF ACTION	
Brief interventions	Increase motivation to quit and impart cessation skills (e.g., community support, telephone counseling)	2
Cognitive-behavioral and relapse-prevention therapies	Behavioral strategies are developed to manage triggers; cognitive coping strategies target maladaptive thoughts to prevent relapse	1
Motivational interviewing	Therapist promotes patient's self-motivational statements, and in turn, patient gains greater awareness of the problems with smoking; increases intention for smoking cessation	2

Effectiveness rating: 1 = strong evidence to support efficacy; 2 = moderate evidence to support efficacy; 3 = little evidence to support efficacy

Cognitive-Behavioral Therapies

In cognitive behavioral therapy (CBT), patients learn to anticipate situations in which they are likely to smoke and then plan to cope with these situations using behavioral (e.g., substitution of behavior) and cognitive (e.g., challenging thoughts) techniques. Some degree of efficacy of CBT in smokers has been observed for both individual and group counseling formats.

Relapse-Prevention (Coping Skills) Therapies

A large number of smokers relapse within 6 months of quitting. Focusing on relapse prevention skills, including recognizing high-risk situations and coping with lapses, can be included in initial smoking cessation treatment or following a quit attempt.

Pharmacologic Treatments

There are three U.S. Food and Drug Administration (FDA)-approved classes of smoking cessation pharmacotherapies: nicotine replacement therapy (NRT), sustained-release bupropion, and varenicline[6] (Table 32-2).

Nicotine Replacement Therapies

The goal of NRT is to alleviate tobacco withdrawal, which allows smokers to focus on habit and conditioning factors when attempting cessation. NRTs rely on systemic venous absorption and so do not produce the rapid high levels of arterial nicotine achieved when cigarette smoke is inhaled. Thus, individuals are unlikely to become addicted to NRT. Safety concerns regarding smoking while using an NRT patch appear to be less serious than previously thought. The most recent evidence suggests that treatment with a nicotine patch while concomitantly smoking is not hazardous, and that use of NRT *before* the quit date may actually facilitate smoking cessation compared with when given at the time of quitting smoking. All commercially available forms of NRT are effective and increase quit rates approximately 1.5- to 2.5-fold compared with placebo. The transdermal patch, gum, and lozenge are available over the counter (OTC), whereas the nasal spray is available by prescription.

Nicotine Gum

Nicotine ingested orally is extensively metabolized on first pass through the liver. Nicotine polacrilex gum avoids this problem through buccal absorption. Nicotine gum was approved as an OTC medication in the United States in 1996 and contains 2 or 4 mg of nicotine that can be released from a resin by chewing. Nicotine gum should be administered by scheduled dosing (e.g., 1 piece of 2-mg gum/hour). The original recommended duration of treatment was 3 months, although many experts believe longer treatment is more effective. Nicotine absorption from the gum peaks 30 minutes after beginning to use the gum. Venous nicotine levels from 2- and 4-mg gum are about one third and two thirds, respectively, of the steady-state (e.g., between cigarettes) levels of nicotine achieved with cigarette smoking. Nicotine delivered by cigarettes is absorbed directly into the pulmonary arterial circulation. Thus, arterial

TABLE 32-2 PHARMACOLOGIC TREATMENTS FOR TOBACCO DEPENDENCE

NICOTINE REPLACEMENT THERAPIES*		
Gum (OTC)	Slow nicotine absorption gradually reduces nicotine craving and withdrawal	1
Transdermal nicotine patch (OTC)	Slow nicotine absorption gradually reduces nicotine craving and withdrawal	1
Lozenge (OTC)	Slow nicotine absorption gradually reduces nicotine craving and withdrawal	1
Vapor inhaler (prescription) and e-cigarettes (nonprescription)	Fast nicotine absorption leads to stimulation of nAChR, which rapidly reduces nicotine craving and withdrawal. Electronic cigarettes (e-cigarettes) are not approved by the FDA and are under study to compare with other nicotine replacement therapies.	1
Nasal spray (prescription)	Fast nicotine absorption leads to stimulation of nAChR, which reduces craving and withdrawal	1
NON-NICOTINE PHARMACOTHERAPIES		
Bupropion SR*	Blocks reuptake of DA and NA; high-affinity, noncompetitive nAChR antagonism reduces nicotine reinforcement, withdrawal, and craving	1
Varenicline*	Acts as a partial agonist of $\alpha_4\beta_2$ nAChRs	1
Nortriptyline	Blocks reuptake of NA and 5-HT; probably reduces withdrawal symptoms and comorbid depressive symptoms; side effects limit utility.	1-2
Clonidine	α_2-Adrenoreceptor agonist reduces nicotine withdrawal symptoms	2
Mecamylamine	Noncompetitive, high-affinity nAChR antagonist combined with TNP reduces nicotine reinforcement, craving, and withdrawal and may increase smoking cessation rates versus TNP alone.	2-3
Naltrexone	Minimal evidence that this μ-opioid peptide receptor antagonist improves smoking cessation outcomes alone, or in combination with TNP. It may reduce alcohol use, and obviate cessation-induced weight gain.	3
Cytisine	Nicotinic partial agonist appears to be safe and efficacious for smoking cessation.	2
Nicotine vaccine	Limited evidence of efficacy for smoking cessation in early human trials, and recent phase III trials have been negative.	3

*Approved by the U.S. Food and Drug Administration (FDA).
Effectiveness rating: 1 = strong evidence to support efficacy; 2 = moderate evidence to support efficacy; 3 = little evidence to support efficacy.
5-HT = serotonin; DA = dopamine; NA = norepinephrine; nAChR= nicotine acetylcholine receptor; OTC = over the counter; TNP = transdermal nicotine patch.

levels from smoking are 5 to 10 times higher than those from the 2- and 4-mg gums. Absorption of nicotine in the buccal mucosa is decreased by an acidic environment, and patients should not drink beverages (e.g., coffee, soda, juice) immediately before, during, or after nicotine gum use.

Several placebo-controlled trials established the safety and efficacy of nicotine gum for smoking cessation.[6] There appears to be some evidence to support using higher doses of nicotine gum (4 mg pieces) in more highly dependent cigarette smokers (≥25 cigarettes per day [cpd]), which supports the idea of matching nicotine gum dose to dependence level of the smoker. Side effects from nicotine gum are rare and include those of mechanical origin (e.g., difficulty chewing, sore jaw) or of local pharmacologic origin (e.g., burning in mouth, throat irritation). Tolerance develops to most side effects over the first week, and education about proper use of the gum (e.g., do not chew too vigorously) decreases side effects.

Nicotine Lozenges

Nicotine lozenges that deliver nicotine (2 and 4 mg preparations) by buccal absorption were approved for OTC use in the United States in 2002. Lozenges offer further flexibility for nicotine replacement options for smokers and are known to allow greater absorption of nicotine compared with nicotine gum. Mild throat and mouth irritation have been reported in preliminary trials. Nicotine lozenges have shown their superiority to placebo lozenges, with significant reduction in nicotine craving and withdrawal. Furthermore, high lozenge doses may be more efficacious in more highly dependent smokers, suggesting that, similar to nicotine gum, lozenge dose can be matched with dependence level. Interestingly, the combination of nicotine lozenge and nicotine patch may lead to the higher long-term quit rates compared with both NRT monotherapies and bupropion.

Transdermal Nicotine Patch

Transdermal nicotine patch (TNP) formulations take advantage of ready absorption of nicotine across the skin. Three of the formulations are for 24-hour use, and one is for 16-hour use. Starting doses are 21 to 22 mg/24-hour patch and 15 mg/16-hour patch. Patches are applied daily each morning. Nicotine administered by patches is slowly absorbed so that, on the first day, venous nicotine levels peak 6 to 10 hours after administration. Thereafter, nicotine levels remain fairly steady, with a decline from peak to trough of 25% to 40% with 24-hour patches. Nicotine levels obtained with the use of patches are typically half those obtained by smoking. After 4 to 6 weeks on high-dose patch (21 or 22 mg/24 hours, and 15 mg/16 hours), smokers are tapered to a middle dose (e.g., 14 mg/24 hours or 10 mg/16 hours) and then to the lowest dose after an additional 2 to 4 weeks (7 mg/24 hours or 5 mg/16 hours). Most studies suggest that abrupt cessation of the use of patches often causes no significant withdrawal; thus, tapering does not appear to be necessary. The recommended total duration of treatment is usually 6 to 12 weeks.

The overall efficacy of the TNP for smoking cessation has been well documented.[6] The effects of active TNP are independent of patch type, treatment duration, tapering procedures, and behavioral therapy format or intensity, although it should be noted that behavioral treatment enhances outcomes with TNP compared with TNP alone. Severe adverse events with nicotine patches have not been found, with the most common minor side effects being skin reactions (50%), insomnia and increased or vivid dreams (15% with 24-hour patches), and nausea (5-10%). Tolerance to these side effects usually develops within a week. Rotation of patch sites decreases skin irritation. Insomnia reported in the first week after cessation appears to be mostly a result of nicotine withdrawal rather than the nicotine patch itself. A 24-hour patch can be removed before bedtime to determine whether the insomnia is caused by the nicotine patch. Without treatment, insomnia usually abates after 4 to 7 days. There is minimal dependence liability associated with patch use: only 2% of patch users continue to use it for an extended period after a cessation trial.

Nicotine Nasal Spray

Nicotine nasal spray is a nicotine solution in a nasal spray bottle similar to the type used with saline sprays. This NRT was approved for treatment of nicotine dependence in the United States in 1996. Nasal spray delivers about 1 mg nicotine per administration, and the patient administers the spray (10 mg/mL) to each nostril every 4 to 6 hours. This formulation produces a more rapid rise in nicotine levels than does nicotine gum, and the rise in nicotine levels produced by nicotine spray falls between the levels produced by nicotine gum and cigarettes. Peak nicotine levels occur within 10 minutes, and venous nicotine levels are about two thirds those of between-cigarette levels. Smokers may use the nasal spray as needed up to 30 times/day for 12 weeks.

Randomized, double-blind, placebo-controlled trials of nasal spray versus placebo spray[6] have established the safety and efficacy of the nasal spray for smoking cessation. Both trials employed treatment for 3 to 6 months, and active nasal spray led to a doubling of quit rates during active use. Differences were reduced or absent with extended follow-up, suggesting the need for maintenance use of this agent. However, such long-term studies to date have not been published. Major side effects from nicotine nasal spray are nasal and throat irritation, rhinitis, sneezing, coughing, and watering eyes. Nicotine nasal spray may have modest dependence liability; prolonged use occurs in about 10% of smokers using the nasal spray, so follow-up of smokers using nasal spray is recommended.

Nicotine Inhaler

Nicotine inhalers are cartridges (plugs) of nicotine (containing about 1 mg of nicotine each) placed inside hollow cigarette-like plastic rods. The cartridges produce a nicotine vapor when warm air is passed through them. Absorption from a nicotine inhaler is primarily buccal rather than respiratory. More recent versions of inhalers produce a rise in venous nicotine levels more rapidly than with nicotine gum but less rapidly than with nicotine nasal spray, with nicotine blood levels of about one third that of between-cigarette levels. Smokers are instructed to puff continuously on the inhaler (0.013 mg/puff) during the day, and recommended dosing is 6 to 16 cartridges daily. The inhaler is to be used as needed for about 12 weeks. No serious medical side effects have been reported with nicotine inhalers. Fifty percent of subjects report throat irritation or coughing. Double-blind, randomized controlled trials have demonstrated the superiority of inhaler to placebo inhalers for smoking cessation. Results revealed a two- to three-fold increase in quit rates (17 to 26%) at trial end point compared with placebo inhalers, and smaller differences at follow-up periods of 1 year or longer. These data support the short-term efficacy of the inhaler in cigarette smokers, but longer-term trials with the inhaler are needed, and there is some modest concern about abuse liability based on long-term use of the product in more than 10% of smokers.

Since their introduction in 2004, electronic cigarettes (e-cigarettes) to deliver nicotine have been purchased by millions (and rapidly increasing numbers) of people. Their role in tobacco control has been controversial, and there is a dearth of strong data to support or refute their efficacy in mitigating tobacco withdrawal and assisting quit attempts by smokers. One recent randomized controlled trial found that 13 weeks of nicotine e-cigarette use resulted in increased smoking abstinence at 6 months compared with use of patches or placebo e-cigarettes, although these differences were not statistically significant. Additional research will be required to determine their benefits and harms at both individual and population levels.[7]

Sustained-Release Bupropion

The phenylaminoketone, atypical antidepressant agent bupropion, in the sustained-release (SR) formulation (Zyban), is a non-nicotine first-line pharmacologic treatment approved by the FDA for nicotine dependent smokers who want to quit smoking. The mechanism of action of this antidepressant agent in the treatment of nicotine dependence likely involves dopamine and norepinephrine reuptake blockade, as well as antagonism of high-affinity nAChRs. The exact mechanism by which bupropion exerts antismoking effects is unclear. The goals of bupropion therapy are smoking cessation, reduction of nicotine craving and withdrawal symptoms, and prevention of cessation-induced weight gain.

The target dose of bupropion in nicotine dependence is 300 mg daily (150 mg bid), and it is typically started 7 days before the target quit date (TQD) at 150 mg daily, then increased to 150 mg twice daily after 3 or 4 days. Unlike with the NRTs, there is no absolute requirement that smokers completely cease smoking by the TQD, although many smokers report a significant reduction in urges to smoke and craving that facilitates cessation at the time of the TQD when drug levels reach steady-state plasma levels. Some smokers gradually reduce their cigarette smoking over several weeks before quitting.

A pivotal multicenter study established the efficacy and safety of bupropion SR for treatment of nicotine dependence, which led to its FDA approval in the United States in 1998. In a 7-week double-blind, placebo-controlled multicenter trial, four doses of bupropion SR (0, 100, 150, and 300 mg/day in twice-daily dosing), in combination with weekly individual cessation counseling, were prescribed to 615 cigarette smokers using at least 15 cpd. At 1-year follow-up, cessation rates were 12.4%, 19.6%, 22.9%, and 23.1%, respectively. Bupropion SR treatment dose-dependently reduced weight gain associated with smoking cessation and significantly reduced nicotine withdrawal symptoms at the 150- and 300-mg/day doses.

Subsequently, the efficacy of the combination of bupropion SR and TNP was studied in a double-blind, placebo-controlled, randomized multicenter trial. A total of 893 cigarette smokers, using at least 15 cpd, were randomized to one of four experimental groups: (1) placebo bupropion + placebo patch; (2) placebo bupropion + TNP; (3) bupropion (300 mg/day) + placebo patch; or (4) bupropion + TNP. Cessation rates at the 1-year follow-up assessment were 15.6%, 16.4%, 30.3%, and 35.5%, respectively. The bupropion groups were significantly better than the placebo-alone and TNP-alone conditions, but the combination of bupropion and TNP was not significantly better than bupropion alone. Weight suppression after cessation was most robust in the combination therapy group. Finally, a randomized controlled trial has demonstrated the efficacy of sustained-release bupropion in relapse prevention after smoking cessation. In individuals who quit smoking with 7 weeks of bupropion (300 mg/day) treatment, bupropion SR versus placebo for 12 months delayed smoking relapse and resulted in weight gain.

Common side effects reported with bupropion administration in cigarette smokers are headache, nausea and vomiting, dry mouth, insomnia, and activation, most of which occur during the first week of treatment. The main contraindication for the use of bupropion is a past history of seizures of any etiology. The rates of de novo seizures are low with this agent (<0.5%) at doses of 300 mg daily or less, and seizures have been observed when daily dosing exceeds 450 mg/day.

Varenicline

Varenicline tartrate (Chantix in the United States, Champix in Europe and Canada), an $\alpha_4\beta_2$ nAChR partial agonist, was approved as a first-line smoking cessation agent by the FDA in 2006 and in Canada and Europe in 2007. Varenicline (2 mg/day) is at least as good if not better than bupropion SR (300 mg/day) for providing long-term abstinence from cigarette smoking. Continuous abstinence over the follow-up period (weeks 9 to 52) were lower, and participants taking varenicline continued to show a higher rate of abstinence than participants taking bupropion and placebo. Varenicline has been also found to be effective in preventing smoking relapse compared with placebo.

Varenicline reduces tobacco cravings and smoking satisfaction and is generally well tolerated. The most common adverse events reported in the initial studies were nausea and insomnia. However, since approval of the drug, concerns have arisen over treatment-emergent neuropsychiatric events, including agitation, suicidal and homicidal ideation, mania, and psychosis. Thus, close monitoring of smokers, especially those with a history of psychiatric illness, has been strongly advised when prescribing this agent. In fact, there is recent evidence that vareniciline appears to be safe and efficacious in smokers with comorbid psychiatric disorders[8] and specifically disorders such as schizophrenia and bipolar disorder.

Combination Pharmacotherapies

There is substantial evidence to suggest that combining various formulations of NRTs, combining NRTs with bupropion SR or varenicline, or combining bupropion with varenicline may lead to enhanced smoking cessation outcomes compared with monotherapies. Although, less is known about how best to switch medications in the case of initial nonresponse, a recent study that used an adaptive trial design demonstrated that initial response to transdermal nicotine (e.g., a 50% reduction in smoking in the first 2 weeks after patch application) predicted success in subsequent smoking cessation. Moreover, those who failed to respond initially to transdermal nicotine responded to augmentation to bupropion SR or a switch to varenciline. Further studies of combination pharmacotherapies and strategies for switching among approved medications are warranted.

Off-Label Medications

Other medications (see Table 32-2) have demonstrated some evidence of efficacy for tobacco treatment, but are otherwise not approved for use for this indication and should be considered second-line treatments.

PREVENTION

Tobacco dependence remains one of the leading preventable causes of morbidity and mortality in the Western world. Nonetheless, smoking cessation therapies are amongst the most cost-effective and proven therapies in medicine. Yet, most health care providers do not identify tobacco use in their patients. Recent recommendations suggest that all smokers should be approached about quitting smoking. Moreover, effective tobacco policies such as state taxes and prevention efforts targeted to reducing the initiation of tobacco use by youth and adults are critical elements in the overall strategy to reduce the burden of tobacco-related disease and related social and health care costs.

PROGNOSIS

The prognosis for tobacco smokers who quit is excellent in terms of years of life and quality of life gained, and earlier quitting appears to be associated with increases in lifespan and quality of life.[9] Furthermore, although medication and behavioral treatments have documented efficacy in treating tobacco dependence, it is important that these therapies be used in combination to achieve the best overall results and ensure adequate skill acquisition and treatment adherence. The promise of novel treatment and prevention interventions and the employment of personalized medicine matching the right medications to smokers with preferential responses using pharmacogenetic (e.g. polymorphisms in genes for catecholamine-*O*-methyltransferase and the nicotinic receptor subunits CHRNA3 and CHRNA4) and neuroimaging biomarkers (e.g., positron emission tomography and functional magnetic resonance imaging) are of considerable excitement for the tobacco treatment field[10] and may be realized within the next decade. Furthermore, given the high rates of tobacco use and dependence in psychiatric populations and the lower rates of quitting in this subset of smokers compared with the general population, specific adaptation of tobacco treatments to mentally ill smokers will be of paramount importance to improving prognosis and outcomes in these hard-to-treat groups of smokers.[11] Additional future challenges include developing safer and more effective smoking cessation therapies and making these therapies available to all smokers who wish to quit.

Grade A References

A1. Nicotine Replacement Therapy for Smoking Cessation or Reduction. A Review of the Clinical Evidence. Ottawa ON: Canadian Agency for Drugs and Technologies in Health; 2014.

A2. Bullen C, Howe C, Laugesen M, et al. Electronic cigarettes for smoking cessation: a randomized controlled trial. *Lancet*. 2013;382:1629-1637.

A3. Hughes JR, Stead LF, Hartmann-Boyce J, et al. Antidepressants for smoking cessation. *Cochrane Database Syst Rev*. 2014;1:CD000031.

A4. Hartmann-Boyce J, Stead LF, Cahill K, et al. Efficacy of interventions to combat tobacco addiction: Cochrane update of 2013 reviews. *Addiction*. 2014;109:1414-1425.

A5. Koegelenberg CF, Noor F, Bateman ED, et al. Efficacy of varenicline combined with nicotine replacement therapy vs varenicline alone for smoking cessation: a randomized clinical trial. *JAMA*. 2014;312:155-161.

A6. Rose JE, Behm FM. Combination treatment with varenicline and bupropion in an adaptive smoking cessation paradigm. *Am J Psychiatry*. 2014;171:1199-1205.

A7. Rose JE, Behm FM. Adaptive smoking cessation according to initial response to precessation nicotine patch. *Am J Psychiatry*. 2013;170:860-867.

A8. Cahill K, Stead LF, Lancaster T. Nicotine receptor partial agonists for smoking cessation. *Cochrane Database Syst Rev*. 2012;4:CD006103.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Thun MJ, Carter BD, Feskanich D, et al. 50-Year trends in smoking-related mortality in the United States. *N Engl J Med*. 2013;368:351-364.
2. Jha P, Peto R. Global effects of smoking, of quitting, and of taxing tobacco. *N Engl J Med*. 2014; 370:60-68.
3. Leslie FM, Mojica CY, Reynaga DD. Nicotinic receptors in addiction pathways. *Mol Pharmacol*. 2013;83:753-758.
4. Kupfer DJ, Regier DA. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Arlington, VA: American Psychiatric Association; 2013.
5. Zwar NA, Mendelsohn CP, Richmond RL. Supporting smoking cessation. *BMJ*. 2014;348: f7535.
6. Carson KV, Brinn MP, Robertson TA, et al. Current and emerging pharmacotherapies for smoking cessation. *Subst Abus*. 2013;7:85-105.
7. Hajek P, Etter JF, Benowitz N, et al. Electronic cigarettes: review of use, content, safety, effects on smokers and potential for harm and benefit. *Addiction*. 2014;109:1801-1810.
8. David SP, Lancaster T, Stead LF, et al. Opioid antagonists for smoking cessation. *Cochrane Database Syst Rev*. 2013;6:CD003068.
9. Jha P, Ramasundarahettige C, Landsman V, et al. 21st-Century hazards of smoking and benefits of cessation in the United States. *N Engl J Med*. 2013;368:341-350.
10. Bough KJ, Lerman C, Rose JE, et al. Biomarkers for smoking cessation. *Clin Pharmacol Ther*. 2012;93:526-538.
11. Mackowick KM, Lynch MJ, Weinberger AH, George TP. Treatment of tobacco dependence in people with mental health and addictive disorders. *Curr Psychiatry Rep*. 2012;14:478-485.

REVIEW QUESTIONS

1. Which of the following is the level of reduction in smoking that is associated with significant lowering of cardiovascular and respiratory disease risk?

A. 25%
B. 50%
C. 75%
D. 100%

Answer: D See Epidemiology section.

2. Which of the following signs and symptoms is NOT associated with nicotine withdrawal?

A. Irritability
B. Insomnia
C. Weight gain
D. Gastrointestinal discomfort

Answer: D See Diagnosis section.

3. Polymorphisms in the following genes have been associated with successful smoking cessation outcomes, EXCEPT which gene?

A. CHRNA4
B. Neuregulin
C. COMT
D. CHRNA3

Answer: B See Prognosis section.

4. Which of the following is NOT an evidence-based behavioral interventions for smoking cessation?

A. Cognitive behavioral therapy (CBT)
B. Motivational interviewing
C. Dialectical behavioral therapy (DBT)
D. Brief interventions

Answer: C See Table 32-1.

5. Which of the following medications may be effective as adjunctive treatment to behavioral therapies for smoking cessation?

A. Naltrexone
B. Nicotine vaccine
C. Mecamylamine
D. Nortriptyline

Answer: D See Table 32-2.

33 ALCOHOL USE DISORDERS

PATRICK G. O'CONNOR

DEFINITION

A variety of terms have been used to describe the spectrum of medical, psychological, behavioral, and social problems associated with excessive consumption of alcohol (*alcohol problems*). *Alcoholism* is perhaps the most widely used term to describe patients with alcohol problems. In an attempt to define *alcoholism* more precisely, an expert panel of the National Council on Alcoholism and Drug Dependence and the American Society of Addiction Medicine developed a definition of alcoholism that included "a primary chronic disease with genetic, psychosocial and environmental factors … often progressive and fatal … characterized by impaired control over drinking, preoccupation with the drug alcohol, use of alcohol despite future consequences, and distortions of thinking, most notably denial." Because the term *alcoholism* is so broad, it also can be imprecise in defining the entire spectrum of alcohol problems.

Abstainers are individuals who consume no alcohol. *Moderate drinking* is defined by the National Institute on Alcohol Abuse and Alcoholism as the average number of drinks consumed daily that places an adult at low risk for alcohol problems. There is some epidemiologic evidence to suggest that moderate drinking may have some health benefits by reducing the risk for cardiovascular disease (Chapter 52). The scope of alcohol consumption that imparts this benefit may be low, however (e.g., less than one drink per day).

At-risk drinking is a level of alcohol consumption that imparts health risks (Table 33-1). This category of drinking behavior has been identified on the basis of epidemiologic evidence that certain threshold levels of alcohol consumption are associated with increased risk for specific health problems.

TABLE 33-1 TERMS AND CRITERIA FOR PATTERNS OF ALCOHOL USE

AT-RISK DRINKING
Men: >14 drinks/week or >4 drinks/day Women: >7 drinks/week or >3 drinks/day
ALCOHOL USE DISORDER CRITERIA*
Tolerance Withdrawal More use than intended Craving Unsuccessful attempts to cut down Excessive time acquiring alcohol Activities given up because of use Use despite negative effects Failure to fulfill major role obligations Recurrent use in hazardous situations Continued use despite social or intrapersonal problems

*Mild = 2-3 criteria, moderate = 4-5 criteria, severe = 6 or more criteria.
(From American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.)

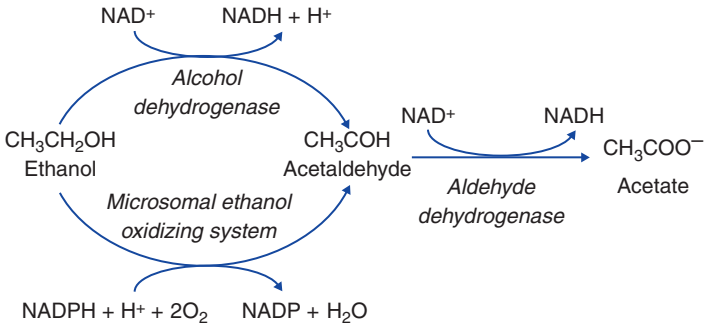

FIGURE 33-1. Ethanol metabolism. Alcohol dehydrogenase predominates at low to moderate ethanol doses. The microsomal ethanol-oxidizing system is induced at high ethanol levels of chronic exposure and by certain drugs. Aldehyde dehydrogenase inhibition (genetic or drug induced) leads to acetaldehyde accumulation, particularly in the latter group.

At-risk drinking is defined differently for men younger than 65 years than for women of all ages because of generally lower body weights and lower rates of metabolism of alcohol in women; the definition in men older than 65 years is the same as in women because of the age-related increased risk for alcohol problems, in part owing to changes in alcohol metabolism in older individuals. *Binge drinking* or *heavy drinking* is the episodic consumption of large amounts of alcohol, usually five or more drinks per occasion for men and four or more drinks per occasion for women. One standard drink contains 12 g of pure alcohol, an amount equivalent to that contained in 5 ounces of wine, 12 ounces of beer, or 1.5 ounces of 80-proof spirits.

The recently published *Diagnostic and Statistical Manual of Mental Disorders,* 5th edition (DSM-5) replaced the previous terminology of *alcohol abuse* and *alcohol dependence* with the term *alcohol use disorders* (see Table 33-1) in order to more clearly describe the spectrum of symptoms experienced by patients. Patients who meet 2 or 3 criteria are considered to have *mild,* 4 or 5 criteria *moderate,* and 6 to 11 criteria *severe* alcohol use disorder.[1]

EPIDEMIOLOGY

In national surveys, 52% of American adults reported that they use alcoholic beverages (liquor, wine, or beer), whereas 23% reported binge drinking, and 6.5% reported heavy drinking in the past 30 days.[2] Among individuals who use alcohol, many experience problems because of their drinking. It has been estimated that more than $100 billion is spent by American society each year to treat alcohol use disorders and to recover the costs of alcohol-related economic losses. Excessive alcohol consumption ranks as the third leading preventable cause of death in the United States after cigarette smoking and obesity. More than 100,000 deaths per year in the United States are attributed to alcohol use disorders.

Population-based epidemiologic studies have shown that alcohol use disorders are among the most prevalent medical, behavioral, or psychiatric disorders in the general population. An epidemiologic survey of the general population in the United States estimated a prevalence of alcohol abuse and dependence (using the older DSM-IV criteria) to be between 7.4 and 9.7%. The lifetime prevalence of abuse and dependence is estimated to be even higher. Despite higher thresholds and tolerance, men are at least twice as likely as women to meet criteria for alcohol abuse and dependence by standard diagnostic survey techniques. Although sociodemographic features, such as young age, low income, and low education level, have been associated with an increased risk for problem drinking, alcohol use disorders are prevalent throughout all sociodemographic groups, and all individuals should be screened carefully. The "skid row" stereotype of the alcohol-dependent patient is much more the exception than the rule.

The prevalence of alcohol use disorders is higher in most health care settings than it is in the general population because alcohol problems often result in treatment-seeking behaviors. The prevalence of problem drinking in general outpatient and inpatient medical settings has been estimated between 15 and 40%. These data strongly support the need for physicians to screen all patients for alcohol use disorders.

PATHOBIOLOGY

Beverage alcohol contains ethanol, which acts as a sedative-hypnotic drug. Alcohol is absorbed rapidly into the blood stream from the stomach and intestinal tract. Because women have lower levels of gastric alcohol dehydrogenase, the enzyme primarily responsible for metabolizing alcohol, they experience higher blood alcohol concentrations than do men who consume similar amounts of ethanol per kilogram of body weight. The absorption of alcohol can be affected by other factors, including the presence of food in the stomach and the rate of alcohol consumption. By means of metabolism in the liver, alcohol is converted to acetaldehyde and acetate (Fig. 33-1). Metabolism is proportional to an individual's body weight, but a variety of other factors can affect how alcohol is metabolized. A genetic variation in a significant proportion of the Asian population alters the structure of an aldehyde hydrogenase isoenzyme, resulting in the development of an alcohol flush reaction, which includes facial flushing, hot sensations, tachycardia, and hypotension.

In the brain, alcohol seems to affect a variety of receptors, including γ-aminobutyric acid (GABA), *N*-methyl-D-aspartate, and opioid receptors. Glycinuric and serotoninergic receptors also are thought to be involved in the interaction between alcohol and the brain. The phenomena of reinforcement and cellular adaptation are thought, at least in part, to influence alcohol-dependent behaviors. Alcohol is known to be reinforcing because withdrawal from ethanol and ingestion of ethanol itself are known to promote further alcohol consumption. After chronic exposure to alcohol, some brain neurons seem to adapt to this exposure by adjusting their response to normal stimuli. This adaptation is thought to be responsible for the phenomenon of tolerance, whereby increasing amounts of alcohol are needed over time to achieve desired effects. Although much has been learned about the variety of effects alcohol can have on various brain receptors, no single receptor site has been identified. A variety of neuropsychological disorders are seen in association with chronic ethanol use, including impaired short-term memory, cognitive dysfunction, and perceptual difficulties.

Although the brain is the primary target of alcohol's actions, a variety of other tissues have a major role in how alcohol affects the human body. Direct liver toxicity may be among the most important consequences of acute and chronic alcohol use (Chapter 152). A variety of histologic abnormalities ranging from inflammation to scarring and cirrhosis have been described. The pathophysiologic mechanism of these effects is thought to include the direct release of toxins and the formation of free radicals, which can interact negatively with liver proteins, lipids, and DNA. Alcohol also has substantial negative effects on the heart and cardiovascular system. Direct toxicity to myocardial cells frequently results in heart failure (Chapter 58), and chronic heavy alcohol consumption is considered to be a major contributor to hypertension (Chapter 67). Other organ systems that experience significant direct toxicity from alcohol include the gastrointestinal tract (esophagus, stomach), immune system (bone marrow, immune cell function), and endocrine system (pancreas, gonads).

CLINICAL MANIFESTATIONS

Alcohol has a variety of specific acute and chronic effects. The acute effects seen most commonly are alcohol intoxication and alcohol withdrawal. Chronic clinical effects of alcohol include almost every organ system.

Acute Effects

Alcohol Intoxication

After entering the blood stream, alcohol rapidly passes through the blood-brain barrier. The clinical manifestations of alcohol intoxication are related directly to the blood level of alcohol. Because of tolerance, individuals chronically exposed to alcohol generally experience less severe effects at a given blood alcohol level than do individuals who are not chronically exposed to alcohol.

The symptoms of mild alcohol intoxication in nontolerant individuals typically occur at blood alcohol levels of 20 to 100 mg/dL and include euphoria, mild muscle incoordination, and mild cognitive impairment. At higher blood alcohol levels (100 to 200 mg/dL), more substantial neurologic dysfunction occurs, including more severe mental impairment, ataxia, and prolonged reaction time. Individuals with blood alcohol levels in these ranges can be obviously intoxicated, with slurred speech and lack of coordination. These effects progress as the blood alcohol level rises to higher levels, to the point at which stupor, coma, and death can occur at levels equal to or greater than 300 to 400 mg/dL, especially in individuals who are not tolerant to the effects of alcohol. The usual cause of death in individuals with very high blood levels of alcohol is respiratory depression and hypotension.

Alcohol Withdrawal Syndrome

Alcohol withdrawal can occur when individuals decrease their alcohol use or stop using alcohol altogether. The severity of symptoms can vary greatly. Many individuals experience alcohol withdrawal without seeking medical attention, whereas others require hospitalization for severe illness. Because ethanol is a central nervous system depressant, the body's natural response to withdrawal of the substance is a hyperexcitable neurologic state. This state is thought to be the result of adaptive neurologic mechanisms being unrestrained by alcohol, with an ensuing release of a variety of neurohumoral substances, including norepinephrine. In addition, chronic exposure to alcohol results in a decrease in the number of GABA receptors and impairs their function.

The clinical manifestations of alcohol withdrawal include hyperactivity resulting in tachycardia and diaphoresis. Patients also experience tremulousness, anxiety, and insomnia. More severe alcohol withdrawal can result in nausea and vomiting, which can exacerbate metabolic disturbances. Perceptual abnormalities, including visual and auditory hallucinations and psychomotor agitation, are common manifestations of more moderate to severe alcohol withdrawal. Grand mal seizures commonly occur during alcohol withdrawal, although they do not generally require treatment beyond the acute withdrawal phase.

The time course of the alcohol withdrawal syndrome can vary within an individual and by symptom complex, and the overall duration of symptoms can be a few to several days (Fig. 33-2). Tremor is typically among the earliest symptoms and can occur within 8 hours of the last drink. Symptoms of tremulousness and motor hyperactivity typically peak within 24 to 48 hours. Although mild tremor typically involves the hands, more severe tremors can involve the entire body and greatly impair a variety of basic motor functions. Perceptual abnormalities typically begin within 24 to 36 hours after the last drink and resolve within a few days. When withdrawal seizures occur, they are typically generalized tonic-clonic seizures and most often occur within 12 to 24 hours after reduction of alcohol intake. Seizures can occur, however, at later time periods as well.

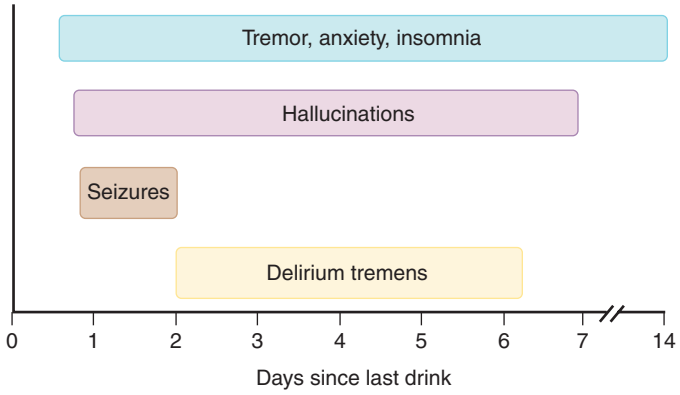

FIGURE 33-2. Time course of alcohol withdrawal.

The most severe manifestation of the alcohol withdrawal syndrome is delirium tremens. This symptom complex includes disorientation, confusion, hallucination, diaphoresis, fever, and tachycardia. Delirium tremens typically begins after 2 to 4 days of abstinence, and the most severe form can result in death.[3]

Chronic Effects

Acute manifestations, including intoxication and withdrawal, are generally stereotypical in their appearance and time course, but chronic manifestations tend to be more varied. Many patients with alcohol dependence may be without evidence of any chronic medical manifestations for many years. As time goes on, however, the likelihood that one or more of these manifestations will occur increases considerably. All major organ systems can be affected, but the primary organ systems involved are the nervous system, cardiovascular system, liver, gastrointestinal system, pancreas, hematopoietic system, and endocrine system (Table 33-2). Patients who drink are at risk for a variety of malignant neoplasms, such as head and neck, esophageal, colorectal, breast, and liver cancers (see individual chapters on those cancers). Excessive alcohol use often causes significant psychiatric and social morbidity that can be more common and more severe than the direct medical effects, especially earlier in the course of problem drinking.

Nervous System

In addition to the acute neurologic manifestations of intoxication and withdrawal, alcohol has major chronic neurologic effects. About 10 million Americans have identifiable nervous system impairment from chronic alcohol use. Individual predisposition to these disorders is highly variable and is related to genetics, environment, sociodemographic features, and gender; the relative contribution of these factors is unclear.

In the central nervous system, the major effect is cognitive impairment. Patients may present with mild to moderate short-term or long-term memory problems or may have severe dementia resembling Alzheimer disease (Chapter 402). The degree to which the direct toxic effect of alcohol is responsible for these problems or the impact of alcohol-related nutritional

TABLE 33-2 ALCOHOL-RELATED COMPLICATIONS

SYSTEM/REALM OF PROBLEM	COMPLICATIONS
Nervous system	Intoxication Withdrawal Cognitive impairment Cerebellar degeneration Peripheral neuropathy
Cardiovascular system	Cardiac arrhythmias Chronic cardiomyopathy Hypertension
Liver	Fatty liver Alcoholic hepatitis Cirrhosis
Gastrointestinal tract, esophagus	Chronic inflammation Malignant neoplasms Mallory-Weiss tears Esophageal varices
Stomach	Gastritis Peptic ulcer disease
Pancreas	Acute pancreatitis Chronic pancreatitis
Other medical problems	Cancers: mouth, oropharynx, esophagus, colorectal, breast, hepatocellular carcinoma Pneumonia Tuberculosis
Psychiatric	Depression Anxiety Suicide
Behavioral and psychosocial	Injuries Violence Crime Child or partner abuse Tobacco, other drug abuse Unemployment Legal problems

deficiencies is uncertain (Chapter 416). The deficiency of vitamins such as thiamine may plays a major role in promoting alcoholic dementia and severe cognitive dysfunction, as is seen in Wernicke encephalopathy and Korsakoff syndrome (Chapter 416). Alcohol also causes a polyneuropathy that can present with paresthesias, numbness, weakness, and chronic pain (Chapters 416 and 420). As with the central nervous system, peripheral nervous system effects are thought to be caused by a combination of the direct toxicity of alcohol and nutritional deficiencies. A small proportion (<1%) of patients with alcohol dependence may develop midline cerebellar degeneration, which presents as an unsteady gait.

Cardiovascular System

The most common cardiovascular complications of chronic alcohol consumption are cardiomyopathy, hypertension, and supraventricular arrhythmias. Alcoholic cardiomyopathy can present clinically in a manner similar to other causes of heart failure (Chapter 58). It is the most common cause of nonischemic cardiomyopathy in Western countries, accounting for about 45% of cases. Like these other causes, alcoholic cardiomyopathy also responds to conventional treatments of heart failure (Chapter 59). Abstinence from alcohol can result in significant improvement in cardiomyopathy in some patients. Increasing levels of alcohol consumption also are associated with increasing levels of systolic and diastolic hypertension (Chapter 67).

The most common arrhythmias associated with chronic alcohol use include atrial fibrillation and supraventricular tachycardia; these are seen commonly in the setting of acute intoxication and withdrawal (Chapter 64). The prevalence of alcohol-induced arrhythmias is unclear. Alcoholic cardiomyopathy also is associated with arrhythmias, in particular, ventricular arrhythmias (Chapter 65). Interestingly, the association between high levels of alcohol intake and increased cardiovascular mortality may be lessened in the presence of high physical activity.[4]

Liver

Alcohol abuse is the major cause of morbidity and mortality from liver disease in the United States. It has been estimated that there are more than 2 million people with known alcoholic liver disease in the United States. Factors that predispose to early liver disease include the quantity and duration of alcohol exposure, female gender, and malnutrition. The range of clinical manifestations includes acute fatty liver, alcoholic hepatitis, and cirrhosis (Chapter 153). Fatty liver associated with alcohol ingestion can be asymptomatic or associated with nonspecific abdominal discomfort; it generally improves with abstinence from alcohol. Alcoholic hepatitis can present as an asymptomatic condition identified through abnormalities in liver enzymes or as an acute episode with abdominal pain, nausea, vomiting, and fever. Patients with alcoholic hepatitis have particularly high levels of aspartate aminotransferase in the blood and elevated levels of γ-glutamyltransferase. Alcoholic hepatitis typically improves with abstinence from alcohol along with supportive care.

Alcohol-related cirrhosis is a major cause of death in the United States (Chapter 154). Although patients are often asymptomatic, patients with more advanced cirrhosis may present with a variety of symptoms and signs, including jaundice, ascites, and coagulopathy. Cirrhosis also is associated with gastrointestinal bleeding from esophageal varices (Chapter 138). Although there is some controversy about the use of liver transplantation to treat patients with alcoholic cirrhosis, many believe that patients in established recovery are good candidates for liver transplantation (Chapter 154). One randomized trial demonstrated that motivational enhancement therapy limited the quantity and frequency of pretransplantation alcohol consumption in patients with alcohol dependence.

Gastrointestinal Disease

Chronic alcohol use is associated with a variety of esophageal problems, including esophageal varices, Mallory-Weiss tears, and squamous cell carcinoma of the esophagus. The risk for squamous cell carcinoma is increased further in patients who smoke tobacco and drink alcohol (Chapter 192). Patients with these problems can present with difficulty swallowing, chest pain, gastrointestinal blood loss, and weight loss. Acute alcoholic gastritis typically presents with abdominal discomfort, nausea, and vomiting (Chapter 135).

Pancreas

The risk for pancreatitis in individuals with alcohol dependence is approximately four times that in the general population. Quantity and duration of alcohol exposure and a history of pancreatitis are predictive of future episodes. Acute alcoholic pancreatitis, which may present with severe abdominal pain, nausea, vomiting, fever, and hypotension, can be life-threatening (Chapter 144). Individuals who have recurrent acute pancreatitis may develop chronic pancreatitis, which typically presents with chronic abdominal pain, malabsorption, weight loss, and malnutrition.

Hematopoietic System

The anemia that commonly is seen in patients with chronic alcohol problems can be multifactorial (e.g., blood loss, nutrient deficiency, secondary to liver disease and hypersplenism). Studies of selected inpatients with alcohol dependence showed the prevalence of anemia to range from about 10 to 60%. Gastrointestinal blood loss due to Mallory-Weiss tears (Chapter 135), alcoholic gastritis (Chapter 132), or esophageal varices (Chapters 135 and 153) may be a key factor, and many patients develop subsequent iron deficiency. Dietary folate deficiency can be associated with megaloblastic anemias (Chapter 164). Alcohol also has a direct toxic effect on the bone marrow, which can lead to sideroblastic anemia that resolves after abstinence. Alcohol can suppress megakaryocyte production and cause thrombocytopenia, which may manifest as petechiae or bleeding (Chapter 172); the thrombocytopenia is particularly sensitive to abstinence, with platelet counts usually rebounding or returning to normal within 5 to 7 days after cessation of alcohol intake. Alcohol also appears to interfere directly with platelet function. Alcohol-related immune dysfunction, as evidenced by decreased production and function of white blood cells and derangement in humoral and cell-mediated immunity, partly explains why alcohol-dependent individuals are at higher risk for infectious diseases, such as pneumonia and tuberculosis.

Malignant Neoplasms

Alcohol intake has been associated with numerous cancers, including cancer of the upper digestive, respiratory, and liver malignant neoplasms. The amount of alcohol exposure that increases cancer risk may vary widely and may not correlate with what might be considered to be "safe" levels of alcohol consumption.[5] Concerning specific cancers, alcohol use is associated with squamous cell carcinomas of the esophagus (Chapter 192) and of the head and neck (Chapter 190). The co-occurrence of alcohol and tobacco abuse seems to be synergistic. Either heavy alcohol use or smoking individually increases the rate of oropharyngeal cancer by about six or seven times that of the general population, whereas the rate for people with both risk factors is about 40 times that of the general population. Patients with alcohol-induced liver disease who also have a history of hepatitis B or C are at particularly increased risk for hepatocellular carcinoma (Chapter 196).

Chronic alcohol use also has been associated with malignant neoplasms of the breast (Chapter 198), prostate (Chapter 201), pancreas (Chapter 194), cervix (Chapter 199), lung (Chapter 191), and colon (Chapter 193). Women who have more than one or two alcoholic drinks per day may increase their breast cancer risk 1.5-fold or more. Hormonal mechanisms and direct carcinogenic effects of alcohol have been postulated as causes of this association. The association of cervical cancer with alcohol dependence may be due to alcohol-associated high-risk sexual behaviors that are thought to increase the risk for cervical cancer.

Other Medical Issues

Gout has been associated with alcohol abuse, and flares can occur at lower serum urate levels than in nonalcoholic patients (Chapter 273). Alcoholic ketoacidosis (Chapter 118), which usually follows an alcoholic binge, presents as nausea, vomiting, abdominal pain, and volume depletion. Typically, ketoacidosis is seen with low or normal glucose readings. Mild or nonspecific abnormalities in thyroid function, especially in patients with underlying liver disease, may reflect abnormalities in the clearance of thyroid-stimulating hormone or the impact of elevated circulating estrogens. Infertility and menstrual irregularities have been associated with chronic alcohol consumption, presumably due to alcohol-induced hypothalamic-pituitary dysfunction, gonadal toxicity, and impaired hepatic metabolism of circulating hormones. Hypogonadism is highly prevalent in male alcoholics with cirrhosis. Alcohol dependence also is associated with higher rates of dental and periodontal disease (Chapter 425) and with a variety of dermatologic problems, including spider angiomas and, in patients with poor hygiene, skin infestations. Both at-risk drinking and alcohol dependence have been shown to be associated with an increased risk for hospital-acquired infections, sepsis, and mortality, especially in intensive care patients.

Psychiatric Issues

Psychiatric symptoms and illnesses are exceedingly common among individuals with alcohol problems. The prevalence of anxiety disorders is about 40%, and the prevalence of affective disorders is about 30%. Antisocial personality disorder is also more common in individuals with alcohol problems than in the general population. These psychiatric problems are more prevalent during periods of heavy drinking and withdrawal. All patients with alcohol use disorders require careful screening for psychiatric illnesses. Although alcohol dependence may be associated with worse outcomes in patients with common psychiatric problems such as depression and anxiety,[6] effective treatment of underlying psychiatric disorders may result in improved drinking behaviors.

Other Behavioral and Psychosocial Issues

Alcohol commonly is the underlying cause of domestic abuse, injuries, trauma, motor vehicle crashes, and burns. For example, of all substances of abuse, alcohol and cannabis appear to have the strongest association with intimate partner violence.[7] Patients presenting with injuries should be questioned carefully about their alcohol use. Tobacco use (Chapter 32) and other drug abuse (Chapter 34) are more prevalent in people with alcohol problems than in the general population.

DIAGNOSIS

Data from the history, physical examination, and laboratory generally are needed to provide a complete picture of the extent of alcohol problems in affected patients (Table 33-3).

Discussing the Diagnosis with Patients

In discussing alcohol problems, it is crucial that physicians be sensitive to the stigma and shame that may be felt by patients with alcohol problems and by their families. Alcohol-related diagnoses or problems should be discussed in a nonjudgmental manner, which forges a partnership and indicates commitment to helping with whatever problems the patients might have. Setting the stage for the discussion should include educating patients about the various levels of alcohol problems (e.g., at-risk drinking, alcohol use disorder) so that patients have an understanding of the spectrum of alcohol problems. Many patients may have a skewed view of what qualifies as problem drinking and may believe that only individuals with severe alcohol problems are truly problem drinkers. The history, physical examination, and laboratory studies should be provided as "proof" that a problem may or does exist.

TABLE 33-3 DIAGNOSIS OF ALCOHOL PROBLEMS

HISTORY
Step 1: Ask all patients about current and past use. Do you drink alcohol (ever or currently)? Do you have a family history of alcohol problems? Step 2: Obtain detailed history regarding quantity and frequency of alcohol use. What types of alcohol do you consume? How often do you drink? How much do you usually drink? Do you ever drink more, and if so, how much? Step 3: Standardized questionnaire CAGE questions: • Have you ever felt that you should ***c***ut down on your drinking? • Have people ***a***nnoyed you by criticizing your drinking? • Have you ever felt bad or ***g***uilty about drinking? • Have you ever taken a drink first thing in the morning (***e***ye opener) to steady your nerves or get rid of a hangover? Step 4: Assess specific areas in suspected or known problem drinkers. Criteria for alcohol abuse and dependence Evidence of medical and psychiatric problems Evidence of behavioral or social problems Use of other substances • Tobacco • Mood-altering prescription drugs • Illicit drugs (e.g., heroin, cocaine) • Prior alcohol or substance abuse treatment
PHYSICAL EXAMINATION
Thorough and complete examination important in all patients Focus attention to system with identified problems In all patients, carefully examine: Central and peripheral nervous systems Cardiovascular system Liver Gastrointestinal tract
LABORATORY STUDIES (IN SELECTED PATIENTS)
Liver enzymes Coagulation studies Complete blood count Carbohydrate-deficient transferrin

History

A four-step approach to the alcohol history includes comprehensive questions about alcohol use and a thorough evaluation for alcohol-related problems.

Step 1: Ask All Patients about Current and Past Alcohol Use

A single question—Do you currently or have you ever used alcohol?—can identify quickly patients who are not lifetime abstainers and require further screening. Patients who answer yes to this question should proceed through the subsequent three steps. Patients who answer no can be classified as lifetime abstainers from alcohol and require no further questioning unless their answer changes over time. It is crucial to ask about current and past alcohol use because many patients who meet lifetime criteria for alcohol use disorder but who are currently in recovery answer no to the question about current use; unless it is specifically asked about, important past use information may be missed.

Step 2: Obtain Detailed History Regarding Quantity and Frequency of Alcohol Use

A question to be asked routinely is, What type or types of alcoholic drinks (beer, wine, spirits) do you consume? Many patients do not consider the use of beer or wine "drinking." Quantity should be determined for typical use—How much do you usually drink on a typical drinking day?—and for range of use—Do you ever drink more than your usual amount, and if so, how much? This second question can be particularly important for identifying binge drinking. Quantity questions offer easy identification of at-risk drinking. Asking about the frequency of alcohol consumption—How often do you drink?—helps distinguish daily from nondaily alcohol users. Binge drinkers who drink only on weekends tend to have significant alcohol problems yet not be daily drinkers. A major goal of step 2 is to acquire a complete characterization of current alcohol use behaviors and the pattern of quantity and frequency of alcohol use during the patient's lifetime.

Step 3: Use Standardized Screening Instruments

Many standardized questionnaires have been developed to detect alcohol abuse and dependence. The two questionnaires that have been evaluated most extensively in medical settings are the CAGE (Cut down, Annoyed, Guilty, and Eye opener) questionnaire (see Table 33-3) and the Alcohol Use Disorder Identification Test (AUDIT). The CAGE questionnaire includes four questions and is scored by giving 1 point for each positive response. Given that the word *ever* is used in the CAGE questions, by definition this instrument is designed to detect lifetime alcohol problems and does not distinguish between lifetime problems and current problems. To screen for DSM-IV criteria alcohol abuse and dependence, the CAGE has a sensitivity of 43 to 94% and a specificity of 70 to 97% when a cutoff score of 2 is used to indicate a "positive" result. How this and other instruments relate to the new DSM-5 criteria for alcohol use disorder is the subject of ongoing inquiry.

The AUDIT's 10 questions cover the quantity and frequency of alcohol use, drinking behaviors, adverse psychological symptoms, and alcohol-related problems. The AUDIT was developed by the World Health Organization to identify hazardous (e.g., at-risk) drinking and harmful (e.g., alcohol use that results in physical or psychological harm) drinking. In contrast to the CAGE questionnaire, the AUDIT focuses on recent (current to past year) drinking behaviors. Each question is scored 0 to 4 (range for total score is 0 to 40), and a total score of 8 is considered to be a positive result.

Step 4: Assess Specific Areas in Suspected or Known Problem Drinkers

Questions asked in step 4 are based on the results of the questions asked in steps 2 and 3 to obtain more detailed information in patients with potential alcohol problems. Even patients who do not screen positive on the CAGE questionnaire may warrant detailed questioning about alcohol abuse and

dependence (see Table 33-1), especially if they are drinking at or above at-risk levels or there is other evidence of possible alcohol problems. A detailed review for evidence of alcohol-related medical and psychiatric problems should occur, and the need for further medical and psychiatric evaluation should be determined. The physician should look for evidence of behavioral and social problems commonly associated with alcohol use and screen for family and occupational dysfunction and other problems, such as domestic violence. Patients should be asked about their use of tobacco, mood-altering prescription medications, and illicit drugs such as heroine and cocaine.

Finally, many patients with alcohol problems have prior treatment episodes that should be detailed. The inquiry should include questions not only about formal alcohol treatment (including number of episodes, duration of treatment, and inpatient versus outpatient treatment) but also about more informal treatments, such as attendance at self-help groups like Alcoholics Anonymous (AA). For patients who require a referral for treatment, knowledge of prior treatment experience is a crucial determinant of future referral recommendations.

The National Institute on Alcohol Abuse and Alcoholism has published *Helping Patients Who Drink Too Much: A Clinician's Guide,* which provides a similar approach to screening and evaluating patients for alcohol-related problems and includes an appendix of useful supporting materials.

Physical Examination

Patients with potential alcohol use disorders require a detailed physical examination to complement the history. In addition, attention should be focused on detecting common alcohol-related problems, including the nervous system, cardiovascular system, liver, and gastrointestinal system (see Table 33-2).

Laboratory Findings

A variety of laboratory tests have been proposed to aid screening for alcoholic abuse and dependence. Aminotransferase levels, red blood cell, mean corpuscular volume, and carbohydrate-deficient transferrin, alone or in combination, are not as effective as screening questionnaires, such as the CAGE and the AUDIT.

Laboratory tests do have a role in diagnosis and assessment of patients with potential alcohol problems. Routine laboratory testing, including liver enzymes (Chapter 147), bilirubin, complete blood count, and prothrombin time should be obtained in all patients with alcohol problems on a regular basis so that an appropriate and complete picture of the effects of alcohol on the individual can be obtained.

PREVENTION AND TREATMENT

Rx

The relationship of change in alcohol use with prevention of subsequent problems has been well established. Treatment of alcohol use disorders should be based on the severity of potential or actual alcohol problems and tailored to meet the needs of individual patients. Separate advice and management approaches are suggested for nondependent at-risk or problem drinkers compared with individuals who are alcohol dependent (DSM-IV criteria) (Table 33-4). As with the screening instruments described previously, the manner in which the new DSM-5 criteria relate to specific treatment recommendations will become clearer in the near future.

Treatment of At-Risk Drinkers

Evidence confirms that generalist physicians, in a cost-effective manner, can help patients reduce their alcohol intake and prevent subsequent alcohol-related problems by using brief (5 to 20 minutes), focused counseling techniques (brief interventions) that are well suited for primary care and other medical settings.[8] More recent research has examined the use of brief interventions in more specialized settings, supporting their use in general medical inpatient settings and emergency departments.[■] Newer approaches to decreasing alcohol use in heavy drinkers include providing personalized advice by computer[■] and using text messaging and social media.

The brief counseling strategy includes four main components: motivational techniques, feedback about the problems with alcohol use, discussion of the adverse effects of alcohol, and setting recommended drinking limits. Motivational techniques are designed to motivate patients to change their alcohol use behavior by identifying potential or actual problems with which their alcohol use is associated. Feedback about these problems can make it clear to the patient that the problems exist. For at-risk and problem drinkers who do not meet criteria for alcohol dependence, setting recommended drinking limits below at-risk levels (e.g., less than one drink per day for women and less than two drinks per day for men) is a realistic and suitable goal. Epidemiologic evidence suggests that drinking below these levels is less likely to be associated with problems. Several randomized clinical trials confirm that patients who receive brief interventions significantly decrease their alcohol intake, often to "safe" levels, and can decrease health care use as well.

TABLE 33-4 ADVICE FOR PATIENTS WITH ALCOHOL PROBLEMS

State your medical concern:
Be specific about your patient's drinking patterns and related health risks.
Ask: How do you feel about your drinking?

Agree on a plan of action:
Ask: Are you ready to try to cut down or abstain?
Talk with patients who are ready to make a change in their drinking about a specific plan of action.

For patients who are not alcohol dependent:
Advise the patient to cut down if drinking is at or above at-risk drinking amounts (see Table 33-1) and there is no evidence of alcohol dependence.
Ask the patient to set a specific drinking goal: Are you ready to set a drinking goal? Some patients choose to abstain for a period of time or for good; others prefer to limit the amount they drink. What do you think will work best for you?
Provide patient education materials and tell the patient: It helps to think about your reasons for wanting to cut down and examine what situations trigger unhealthy drinking patterns. These materials will give you some useful tips on how to maintain your drinking goal.

For patients with evidence of alcohol dependence:
Advise to abstain if:
- Evidence of alcohol dependence
- History of repeated failed attempts to cut down
- Pregnant or trying to conceive
- Contraindicated medical condition or medication

Refer for additional diagnostic evaluation or treatment.
Procedures for patient in making referral decisions:
Involve your patient in making referral decisions.
Discuss available alcohol treatment services.
Schedule a referral appointment while the patient is in the office.

Treatment of Alcohol Use Disorders

Patients who meet criteria for an alcohol use disorder, in particular those who are dependent, usually require more intensive services than do patients who meet criteria for at-risk drinking. Most patients can be managed in outpatient treatment settings, whereas patients with a more severe alcohol use disorder or comorbid problems initially will likely require inpatient management, specific counseling programs, and pharmacologic therapy. Before entering a formal program to maintain remission, many patients first require medical management of alcohol withdrawal. Professional organizations have published practice guidelines that provide useful recommendations for how to select among treatment options for patients with alcohol dependence.

Management of Alcohol Withdrawal

Many patients may not present for medical management of alcohol withdrawal and deal with it on their own. However, a substantial subset do present for alcohol withdrawal treatment. Patients with mild to moderate withdrawal generally can be managed safely as outpatients with close follow-up. Patients with moderate to severe withdrawal, as manifested by hypertension, tremor, and any mental status changes, especially patients with significant comorbid medical or psychiatric illnesses, generally are treated best as inpatients. Patients who have a history of severe withdrawal in the past (e.g., delirium tremens) or who have a history of alcohol withdrawal seizures also generally should be managed as inpatients. The three major goals of medical management of alcohol withdrawal are to minimize the severity of withdrawal-related symptoms; to prevent specific withdrawal-related complications, such as seizures and delirium tremens; and to provide referral to relapse prevention treatment.

A wide variety of medications have been evaluated for their effectiveness in managing the alcohol withdrawal syndrome (Table 33-5). Longer-acting benzodiazepines are preferred because they provide a smoother withdrawal. Shorter-acting benzodiazepines, such as oxazepam, may be indicated in individuals with severe liver disease. The most common approach is to administer a standing dose of a benzodiazepine, with additional medication being given "as needed" on the basis of withdrawal symptoms. The specific benzodiazepine and dose often depend on the experience of the prescribing physician and the characteristics of the patient, including the severity of withdrawal (higher doses are used if withdrawal is more severe), the presence of liver disease (patients with severe liver disease should receive lower doses or shorter-acting medications), and the response to prior doses of medication (higher doses are given if symptom control is inadequate; lower doses are

TABLE 33-5 MEDICATIONS FOR THE TREATMENT OF ALCOHOL DEPENDENCE*

MEDICATION	DOSE AND ROUTE	FREQUENCY	EFFECTS	MAJOR COMMON ADVERSE EFFECTS
ALCOHOL WITHDRAWAL				
Benzodiazepines‡				
Chlordiazepoxide*	25-100 mg, PO/IV/IM‡	Every 4-6 hr	Decreased severity of withdrawal; stabilization of vital signs; prevention of seizures and delirium tremens	Confusion, oversedation, respiratory depression
Diazepam†	5-10 mg, PO/IV/IM‡	Every 6-8 hr		
Oxazepam†	15-30 mg, PO‡	Every 6-8 hr		
Lorazepam†	1-4 mg, PO/IV/IM‡	Every 4-8 hr		
β-Blockers				
Atenolol	25-50 mg, PO	Once a day	Improvement in vital signs	Bradycardia, hypotension
Propranolol	10-40 mg, PO	Every 6-8 hr	Reduction in craving	
α-Agonists				
Clonidine	0.1-0.2 mg, PO	Every 6 hr	Decreased withdrawal symptoms	Hypotension, fatigue
Antiepileptics				
Carbamazepine	200 mg, PO	Every 6-8 hr	Decreased severity of withdrawal; prevention of seizures	Dizziness, fatigue, red blood cell abnormalities
PREVENTION OF RELAPSE				
Disulfiram†	125-500 mg, PO	Daily	Decreased alcohol use among those who relapse	Disulfiram-alcohol reaction, rash, drowsiness, peripheral neuropathy
Naltrexone†	50 mg, PO 380 mg, IM	Daily Every 4 wk	Increased abstinence, decreased drinking days	Nausea, abdominal pain, myalgias-arthralgias
Acamprosate†	666 mg, PO	Three times a day	Increased abstinence	Diarrhea

*Most commonly used medications listed.
†Currently approved by U.S. Food and Drug Administration for the indication noted.
‡Dose and routes given for standard fixed-dose regimens, which include dose tapers over time.

given if adverse effects, such as oversedation, have occurred). In general, the amount of medication per dosing period is decreased gradually as the withdrawal syndrome abates. An individualized "symptom-triggered" dosing approach, in which benzodiazepines are administered on a dose-by-dose basis as guided by withdrawal symptoms, is safe and effective in certain patients and can reduce the total doses of benzodiazepines needed to treat withdrawal. β-Blockers (atenolol and propranolol), α-agonists (clonidine), and antiepileptics (carbamazepine) improve signs and symptoms of alcohol withdrawal but are viewed best as adjunctive medications to be used in addition to benzodiazepines.

Prevention of Relapse

Counseling Strategies Used by Alcohol Treatment Programs

Three commonly used psychotherapeutic techniques are motivational enhancement therapy, 12-step facilitation, and cognitive-behavioral coping skills. Two of these techniques are designed to give patients specific tools to help them avoid relapse to alcohol use. In motivational enhancement therapy, patients identify reasons for staying away from alcohol. The 12-step facilitation therapy uses the principles of AA to help patients focus their attention on abstinence. In cognitive-behavioral coping skills therapy, the patient identifies triggers to alcohol use and develops strategies to help deal with the triggers when they are present.

Project MATCH (Matching Alcohol Treatments to Client Heterogenicity) showed equivalence among three counseling approaches (cognitive-behavioral coping skills therapy, motivational enhancement therapy, or 12-step facilitation therapy) to treat alcohol dependence. At 1-year follow-up, most enrolled patients either remained abstinent or significantly decreased their alcohol use.

Self-Help Groups

Self-help groups such as AA and Rational Recovery are an important source of support and treatment for many patients with alcohol dependence. AA has the advantage of being widely available throughout the United States and is free of charge. The overall approach to treatment is based on the 12 steps for maintaining abstinence and dealing with the various effects of alcohol. AA meetings can be either "open" to anybody in the community or "closed" for active members only. The meetings vary in format, size, location, and demographic makeup. In counseling patients about attending AA, it is important for physicians to make them aware that variations in the nature of specific meetings, especially location and demographics of participants, require patients to be willing to attend more than one meeting site on a trial basis so that they find a comfortable setting.

Research on the effectiveness of AA has been limited, and there are no large controlled studies. Indirect evidence suggests, however, a significant improvement in alcohol use behaviors.

Pharmacotherapy to Prevent Relapse to Alcohol Use

The addition of medication to enhance the effectiveness of counseling therapies has been the subject of research for the past 40 years. As the neurobiology of alcohol use disorders has become more clearly understood, the potential to develop medications that may promote abstinence or decreased alcohol use has grown.[9,10] Three medications—disulfiram, naltrexone, and acamprosate—are approved for the treatment of alcohol dependence in the United States (see Table 33-5).

Disulfiram

Disulfiram is designed to prevent alcohol use by causing a severe adverse reaction when patients use alcohol. The disulfiram reaction, which includes flushing, nausea, vomiting, and diarrhea, is mediated by the inhibition of alcohol dehydrogenase and the resulting increase in serum levels of acetaldehyde and acetate after ingestion of alcohol (see Fig. 33-1). Disulfiram also affects monoamine metabolism, and the alcohol-disulfiram reaction may be related to changes in central monoamine functioning. Although disulfiram offers little benefit to most patients, it is effective in reducing alcohol intake in highly motivated patients who are supervised in an alcohol treatment program.

Naltrexone

Naltrexone is thought to decrease alcohol use by diminishing the euphorigenic effects of alcohol and by decreasing craving in alcohol-dependent patients. Randomized, placebo-controlled trials generally have shown that alcohol-dependent patients who receive naltrexone (50 mg/day) are more likely to decrease their alcohol use or remain abstinent compared with patients who receive placebo, and the effects persist after discontinuation of treatment, although one randomized trial did not show benefit in male veterans with severe alcohol dependence. Although most studies of naltrexone were performed in a specialty alcohol treatment setting and observed subjects for only 10 to 12 weeks, naltrexone also is effective in outpatient and primary care settings in patients who are followed for up to 34 weeks. Side effects of naltrexone are infrequent, most notably self-limited nausea in about 10% of patients. Dose-related hepatotoxicity has been reported in patients treated for obesity with high-dose naltrexone (300 mg/day). Mild liver enzyme abnormalities are not a contraindication to naltrexone, but patients should be followed with repeated liver enzyme studies. Patients with acute hepatitis or liver failure should not use naltrexone. In addition to the oral naltrexone, a newer long-acting injectable form of naltrexone was approved by the U.S. Food and Drug Administration (FDA) in 2006. Injectable naltrexone is typically administered at a dose of 380 mg intramuscularly every 4 weeks. Before beginning naltrexone, it is important to be sure that the patient is not opioid dependent in order to avoid a potentially severe opioid withdrawal reaction. Complete opioid abstinence for at least 7 to 10 days is recommended. Adverse reactions seen most commonly in patients who receive injectable naltrexone include injection

site reactions (e.g., induration, itching) and symptoms such as nausea and headache, which are generally self-limited. This formulation of naltrexone may be particularly effective in those with more severe alcohol use disorders.[11]

Acamprosate

Approved by the FDA in 2004, acamprosate (calcium acetylhomotaurinate) has been identified as an effective agent for treatment of alcohol dependence. The precise mechanism of action of acamprosate is uncertain but may be related to its effects on neuroexcitatory amino acids and the inhibitory GABA system. In randomized, placebo-controlled clinical trials, subjects who received acamprosate were more likely to remain abstinent compared with subjects who received placebo. Side effects are minimal and typically include diarrhea. Like naltrexone, acamprosate is given as an adjunctive therapy to psychological treatments for alcohol dependence. Acamprosate appears to be effective in both men and women.

Other Pharmacologic Approaches to Prevent Relapse

There has been much interest in evaluation of the effectiveness of combinations of drug therapies to treat alcohol dependence. One study of 160 patients suggested that the combination of naltrexone and acamprosate was more effective than either medication alone. A larger federally funded study that enrolled 1383 subjects, Project COMBINE, examined naltrexone and acamprosate alone and in combination with two different psychological therapies to see which combination of pharmacologic and behavioral therapies is most effective. The behavioral therapies were medical management, which was designed to approximate counseling that can be provided in primary care and other medical settings, and combined behavioral intervention, which incorporated counseling techniques that are provided in alcohol treatment specialty settings. Results from this study demonstrated that patients receiving medical management with naltrexone, combined behavioral intervention, or both fared best, lending further support to the idea that alcohol-dependent patients can be effectively treated in primary care and other medical settings. Interestingly, acamprosate was not shown to be effective in this study. Gabapentin is effective as a single drug therapy for treating alcohol dependence and relapse-related symptoms, such as insomnia, and the addition of gabapentin to naltrexone improves drinking outcomes in alcohol-dependent patients.

Topiramate, a fructopyranose derivative, is an effective treatment of alcohol dependence at a dose of up to 300 mg/day and is especially effective in patients with a *GRIK1* polymorphism.[12] Other medications that have shown promise include ondansetron, bromocriptine, and sodium valproate. Other drugs have shown possible benefits in patients with alcohol use disorders, perhaps more specifically those with concurrent depression (e.g., fluoxetine) or anxiety (e.g., buspirone).

PROGNOSIS

Alcohol abuse and dependence are chronic disorders that are characterized by exacerbations and remissions. The prognosis is better for patients who seek treatment and receive it in a systematic way (Table 33-6), but it can be poor for patients with advanced liver disease and continued alcohol use. In addition, the use of combinations of medications (e.g., naltrexone plus acamprosate) is under investigation.

FUTURE DIRECTIONS

To date, most studies have focused on shorter-term outcomes, from a few months to a year. It is important to understand more clearly what happens to these patients over time, especially the need for "booster sessions" to sustain improvements provided by brief interventions. Newer pharmacologic therapies may help many patients. More recently, there has been increased interest in expanding chronic care management approaches like those used in managing patients with diabetes or heart failure to patients with substance use disorders.[13] Although one large early trial failed to find a benefit, this approach may be beneficial in selected populations of patients with substance use disorders. Newer technology-based therapies, such as a smartphone application to support recovery in individuals with alcohol use disorders, hold promise for improving treatment outcomes in the future.[14]

TABLE 33-6 OVERVIEW OF TREATMENT APPROACH FOR PATIENTS WITH ALCOHOL PROBLEMS

Evaluate all patients
For patterns of problem alcohol use (Table 33-1)
For alcohol-related complications, if indicated (Table 33-2)
With use of data collected from history, physical examination, and laboratory testing (Table 33-3)
For at-risk and nondependent problem drinkers
Advise to decrease alcohol use to below at-risk levels (Table 33-4)
Advise patients who cannot decrease use to below at-risk levels to abstain
For patients who are alcohol dependent
Assess for need for withdrawal management medications (Table 33-5)
Refer to an alcohol treatment program
Consider medication to prevent relapse (Table 33-5)

Grade A References

A1. D'Onofrio G, Fiellin DA, Pantalon MV, et al. A brief intervention reduces hazardous and harmful drinking in emergency department patients. *Ann Emerg Med*. 2012;60:181-192.

A2. Boon B, Risselada A, Huiberts A, et al. Curbing alcohol use in male adults through computer generated personalized advice: randomized controlled trial. *J Med Internet Res*. 2011;13:e43.

A3. Jonas DE, Amick HR, Feltner C, et al. Pharmacotherapy for adults with alcohol use disorders in outpatient settings: a systematic review and meta-analysis. *JAMA*. 2014;311:1889-1900.

A4. Mason BJ, Quello S, Goodell V, et al. Gabapentin treatment for alcohol dependence: a randomized clinical trial. *JAMA Intern Med*. 2014;174:70-77.

A5. Anton RF, Myrick H, Wright TM, et al. Gabapentin combined with naltrexone for the treatment of alcohol dependence. *Am J Psychiatry*. 2011;168:709-717.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorder. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.
2. Substance Abuse and Mental Health Services Administration. Results from the 2012 National Survey on Drug Use and Health: Summary of National Findings. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2013.
3. Carlson RW, Kumar NN, Wong-McKinstry E, et al. Alcohol withdrawal syndrome. *Crit Care Clin*. 2012;28:549-585.
4. Soedamah-Muthu SS, De Neve M, Shelton NJ, et al. Joint associations of alcohol consumption and physical activity with all-cause and cardiovascular mortality. *Am J Cardiol*. 2013;112: 380-386.
5. Latino-Martel P, Arwidson P, Ancellin R, et al. Alcohol consumption and cancer risk: revisiting guidelines for sensible drinking. *Can Med Assoc J*. 2011;183:1861-1865.
6. Boschloo L, Vogelzangs N, van den Brink W, et al. Alcohol use disorders and the course of depressive and anxiety disorders. *Br J Psychiatry*. 2012;200:476-484.
7. Afifi TO, Henriksen CA, Asmundson GJG, et al. Victimization and perpetration of intimate partner violence and substance use disorders in a nationally representative sample. *J Nerv Mental Dis*. 2012;200:684-691.
8. Johnson NA, Kypri K, Latter J, et al. Prevalence of unhealthy alcohol use in hospital outpatients. *Drug Alcohol Depend*. 2014;144:270-273.
9. Spanagel R, Vengelien V. New pharmacological treatment strategies for relapse prevention. *Curr Top Behav Neurosci*. 2013;13:583-609.
10. Miller PM, Book SW, Stewart SH. Medical treatment of alcohol dependence: a systematic review. *Int J Psychiatry Med*. 2011;42(3):227-266.
11. Fiedmann PD. Clinical practice: alcohol use in adults. *N Engl J Med*. 2013;368:365-373.
12. Kranzler HR, Covault J, Feinn R, et al. Topiramate treatment for heavy drinkers: moderation by a GRIK1 polymorphism. *Am J Psychiatry*. 2014;171:445-452.
13. O'Connor PG. Managing substance dependence as a chronic disease: is the glass half full or half empty? *JAMA*. 2013;310:1132-1134.
14. Gustafson DH, McTavish FM, Chih MY, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. *JAMA Psychiatry*. 2014;71:566-572.

REVIEW QUESTIONS

1. A 58-year-old man presents to his primary care physician's office for follow-up of hypertension. During the visit, he reports that he consumes 3 to 4 glasses (approximately 5 ounces each) of red wine every day but denies any medical, behavioral, or social problems related to his drinking. He is a nonsmoker and does not have a history of alcohol or other substance use disorders. On examination, his blood pressure is 144/96 mm Hg, his pulse is 80 beats per minute, and the remainder of his examination is unremarkable. What would be the most appropriate next step concerning his alcohol consumption?

A. Refer him to Alcoholics Anonymous.
B. Suggest that his current red wine consumption might have health benefits, but he should not drink more than this amount.
C. Prescribe naltrexone to help him stop drinking.
D. Refer him to an alcohol treatment program for counseling.
E. Advise him to decrease his consumption to no more than 2 drinks a day and reassess his alcohol use at the next visit.

Answer: E This patient exhibits "at-risk" drinking—a level of alcohol consumption that increases his risk for alcohol-related health or social problems. Randomized clinical trials have demonstrated that brief physician advice in a primary care setting can result in reduced alcohol consumption, often to levels that are below those considered to be of risk in a patient such as this (e.g., ≤2 drinks/day). In this patient's case, decreasing his alcohol use may also improve his blood pressure control. The health-related risks associated with this level of alcohol consumption are likely to far outweigh any potential (and likely minimal if any) benefit from consumption of red wine. Alcoholics Anonymous, naltrexone, and referral to an alcohol treatment program for counseling should be considered in patients who meet criteria for an alcohol use disorder, which is not the case with this patient.

2. A 30-year-old woman presents for evaluation of "shakiness." She reports several months' history of heavy alcohol consumption, typically 8 to 10 drinks per day including wine and vodka. Her presenting complaint began when she decided to quit drinking 1 day ago because she was recently placed on probation at work because of frequent tardiness related to her alcohol use. She notes that she has been drinking increased amounts of alcohol to "keep going" and that when she does try to decrease her alcohol use, especially with her recent trouble at work and some "close calls" while driving, she is unsuccessful. Which of the following best describes her alcohol consumption pattern and its related effects?

A. She is suffering from alcohol-related delirium tremens.
B. She has a severe alcohol use disorder.
C. She meets criteria for alcohol abuse.
D. She has a mild alcohol use disorder.
E. She is not a binge drinker.

Answer: B This patient meets at least six of the DSM-5 criteria for an alcohol use disorder and thus would be considered in the severe category. Her presentation as described includes several of these criteria: evidence of tolerance, withdrawal, more use than intended, unsuccessful attempts to cut down, tolerance, use despite negative effects, failure to fulfill a major role obligation, and use in hazardous situations. Although she does present with symptoms likely related to alcohol withdrawal, there is nothing in the presentation to suggest delirium tremens. *Alcohol abuse* is a term from earlier versions of DSM and is no longer used as diagnosis. Patients with a "mild" alcohol use disorder have only two or three of the DSM-5 criteria. She does not meet criteria for binge drinking.

3. The same 30-year-old woman described in question 2 eventually develops severe enough withdrawal symptoms to be hospitalized for the medical management of her alcohol withdrawal. Which of the following is true concerning the course of her alcohol withdrawal?

A. If present, withdrawal seizures are likely to occur between 48 and 72 hours after the last drink.
B. Chlordiazepoxide and diazepam are the only benzodiazepines that have been approved by the FDA for the treatment of alcohol withdrawal.
C. Benzodiazepines have been demonstrated to effectively reduce the occurrence of alcohol withdrawal seizures.
D. Longer acting benzodiazepines are always preferred to those that are shorter acting.
E. β-Blockers are an effective alternative to benzodiazepines for the treatment of alcohol withdrawal.

Answer: C Benzodiazepines are the treatment of choice for the alcohol withdrawal syndrome. There is strong evidence that they decrease the frequency of alcohol withdrawal seizures, and numerous randomized trials have demonstrated their effectiveness in controlling withdrawal-related symptoms and vital sign abnormalities. Alcohol withdrawal seizures tend to occur within 12 to 24 hours after alcohol consumption is reduced or stopped. There are four benzodiazepines that have been approved by the FDA—in addition to the two mentioned in choice B, oxazepam and lorazepam are also FDA approved. Short-acting benzodiazepines may be preferred in patients with severe liver disease. β-Blockers are effective adjuncts to benzodiazepines in the treatment of alcohol withdrawal but should not be used alone as an alternative to benzodiazepines.

4. The same 30-year-old patient presented in questions 2 and 3 is discharged from the hospital after successful treatment of the alcohol withdrawal syndrome. Which of the following is the appropriate "next step" in the management of her alcohol-use disorder?

A. Assess her motivation to remain alcohol free and determine her interest in receiving treatment for relapse prevention.
B. Refer her to the Alcoholics Anonymous program that your father went to.
C. Continue her benzodiazepines as an outpatient to help her stay away from alcohol.
D. Discharge her on disulfiram and see her back in your office in a month.
E. Advise her to abstain from alcohol and see her back in your office in a month.

Answer: A The first step in managing patients with alcohol use disorders who have been successfully detoxified is to assess their motivation to remain alcohol free and discuss various follow-up treatment options in order to assess their interest in treatment and help identify the best approach to getting relapse prevention therapy. Patients who are suitably motivated and interested in engaging in ongoing treatment should then be referred as appropriate. Those who are not ready to accept treatment may benefit from motivational interviewing approaches and close follow-up in order to engage them in treatment in the future. Alcoholics Anonymous may be helpful to many patients; however, without assessing a patient's motivation, it is difficult to tell if such a referral will be followed. In addition, AA meetings differ in terms of their content and membership, and patients may need to try several meeting sites to find a group that they are comfortable with. Benzodiazepines are not indicated for the prevention of relapse to alcohol use. Disulfiram has been shown to be effective in highly motivated patients who receive treatment in a highly supervised manner. There is no evidence that brief advice alone is effective in the treatment of alcohol use disorders.

5. A 47-year-old man presents to his primary care physician for a routine preventive health visit. He has a history of an alcohol use disorder and has been hospitalized three times for alcohol withdrawal, is participating in Alcoholics Anonymous, and is currently getting cognitive behavioral therapy from a therapist. Despite this, he continues to struggle with craving for alcohol and has had numerous relapses interrupting episodes of sobriety. He has heard about medications that might help him prevent these relapses and would like to try something. Which of the following is the most appropriate recommendation?

A. Don't use a relapse prevention medication because this will be frowned upon by his Alcoholics Anonymous group.
B. Give him a prescription for disulfiram and ask him to follow up in 3 months.
C. Start with a combination of naltrexone and acamprosate because this will yield the best results.
D. Consider the use of intramuscular naltrexone, 380 mg per month.
E. Prescribe topiramate, 300 mg a day.

Answer: D In several clinical trials, naltrexone has been demonstrated to decrease alcohol consumption in individuals who are dependent on alcohol. Although generally well tolerated, contraindications include current opioid dependence and severe liver disease. Although it is true that some Alcoholics Anonymous participants or groups may frown upon the use of medication to treat alcohol use disorders, this is not a reason to not use these medications. However, patients should be made aware of this possibility and counseled on how to deal with this if it occurs. Disulfiram has been shown to be effective in highly structured treatment programs. Although it may be helpful in some primary care patients, there is no evidence that it is effective when used by primary care physicians in an unstructured manner. There is no significant evidence that combinations of relapse prevention medications improve outcomes in patients who are dependent on alcohol. Topiramate has been demonstrated to be effective in preventing relapse to alcohol use in some studies, but it is not yet approved for this indication.

34

DRUGS OF ABUSE

ROGER D. WEISS

DEFINITION

The term *substance use disorder* has replaced *substance abuse* and *dependence* in the diagnostic lexicon; recent research has shown that the previous hierarchical distinction between abuse and dependence, with dependence representing a more severe form of the disorder, was problematic and not warranted. A substance use disorder is a clinical syndrome characterized by the following statement from the fifth edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*: "The essential feature of a substance use disorder is a cluster of cognitive, behavioral, and physiological symptoms indicating that the individual continues using the substance despite significant substance-related problems."[1] There is no single pathognomonic symptom that is diagnostic of a substance use disorder. Rather, the syndrome is a series of 11 symptoms, of which the individual needs to meet two or more in the same 12-month period to warrant a diagnosis of a substance use disorder (Table 34-1). These 11 symptoms can be grouped into four general categories:

Impaired control: taking a substance in larger amounts or for a longer time than intended; persistent desire or unsuccessful attempts to stop or to reduce use; a great deal of time spent using a substance or recovering from the effects of its use; craving

Social impairment: failure to fulfill role obligations at home, work, or school as a result of repeated substance use; continued substance use despite experiencing interpersonal problems; or reducing or giving up important social, recreational, or occupational activities

Risky use: recurrent use in hazardous situations (e.g., driving) or despite knowledge that substance use is causing or exacerbating a physical or psychological problem

Pharmacological criteria: tolerance and physical dependence, when relevant; not all drugs cause these symptoms

For people who are prescribed medications that can cause tolerance and physical dependence, the diagnostic term *substance use disorder* should be used only with people whose medication use is problematic; tolerance and physical dependence are not considered criteria for a diagnosis of a substance use disorder in patients who are taking medications such as opioids or

TABLE 34-1 DSM-5 CRITERIA FOR SUBSTANCE USE DISORDERS

IMPAIRED CONTROL
1. Took larger amounts/for longer period of time than intended
2. Persistent desire or unsuccessful attempts to stop or reduce use
3. Much time spent using a substance or recovering from its effects
4. Craving
SOCIAL IMPAIRMENT
5. Failure to fulfill home, work, or school obligations because of repeated substance use
6. Continued use despite experiencing interpersonal problems
7. Reducing/giving up important social, recreational, or occupational activities
RISKY USE
8. Recurrent use in hazardous situations
9. Physical/psychological problems related to use
PHARMACOLOGICAL CRITERIA
10. Tolerance, when relevant*
11. Physical dependence, when relevant*

Substance use disorder is diagnosed if two or more of these 11 criteria are met within a 12-month period.
*"When relevant" indicates that these criteria are not counted, even if the symptoms are present, when an individual is using legitimately prescribed medications as intended and those medications are helping the person to function better.
From Hasin DS, O'Brien CP, Auriacombe M, et al. DSM-5 criteria for substance use disorders: recommendations and rationale. *Am J Psychiatry.* 2013;170:834-851.

sedative-hypnotics exclusively as part of appropriate medical care. If someone is using legitimately prescribed medications as intended (e.g., opioids for chronic pain or benzodiazepines for panic disorder) and those medications are helping the person to function better, that person will not meet criteria for a substance use disorder, even if the person is tolerant to the medication and is physically dependent.

EPIDEMIOLOGY

Use of illicit drugs and non-medical use of prescribed drugs are common. In 2011, approximately 22.5 million Americans reported using an illicit drug in the previous month, representing approximately 9% of the population. When asked about their substance use in the past month, 18 million people reported using marijuana, 6 million reported using potentially psychoactive prescription drugs nonmedically, 1.4 million people used cocaine, and 1 million people used hallucinogens; in fact, 10% of youths aged 12 to 17 years reported using an illicit drug in the past month. Illicit drug use is an important contributor to the global burden of disease, accounting for 20 million disability-adjusted life years in 2010. Worldwide, more people were dependent on opioids and amphetamines than other drugs.[2] Drug abuse produces substantial medical morbidity and mortality as well as tremendous social and economic costs.

PATHOBIOLOGY

Drug use disorders involve complex interactions between the pharmacology of a specific drug, an individual's genetic makeup, psychological strengths and weaknesses, environmental circumstances, and societal influences (such as physical and perceived drug availability, legal status and cost of the drug, religious and cultural mores, and presence of alternative rewarding activities). Thus, one can conceptualize the etiology of drug abuse by employing the public health model frequently cited in the study of infectious disease, that is, as an interaction among the host (i.e., the potential drug user), the agent (a specific drug in this case, as opposed to an infectious microorganism), and the environment (the person's family life and peer group and the social, cultural, and religious attitudes toward use of that substance).

The Host

A host factor that is well known to heighten vulnerability to drug abuse problems is a positive family history of a substance use disorder, which has been shown to increase the likelihood of development of both alcohol and drug dependence. Twin studies and adoption studies have shown that both genetic and environmental factors contribute to this vulnerability, although the precise nature by which this occurs is still unknown and a subject of active research. One area of great research interest relates to whether people can be vulnerable to drug dependence in general (e.g., as a result of a risk-taking temperament or poor decision making) or whether they are at high risk to abuse particular substances (perhaps because of a highly reinforcing response to a specific drug).

Psychiatric illness has been shown to influence the likelihood for development of drug abuse problems. For example, conduct disorder in childhood and adolescence and antisocial personality disorder in adulthood have both been found to predispose to subsequent drug abuse problems. Psychiatric disorders such as mood disorders are frequently noted in people with drug abuse problems. However, the presence of these two disorders in the same person does not necessarily imply causation, even if one of the disorders is manifested first.

In addition to the risk factors mentioned previously, certain individual protective factors may reduce the likelihood of a substance use disorder. Individuals who have positive familial relationships, success in academic activities, and meaningful religious affiliations have a lower likelihood for development of drug problems. The fact that many people have a mixture of risk and protective factors speaks to the complex etiologic nature of drug use disorders.

The Agent

Most drugs of abuse are inherently reinforcing; animals typically will self-administer most of the commonly abused drugs. Not all drugs are equally reinforcing in general, however, and there is a great deal of individual variation in drug preference. Some people like the stimulating effects of drugs such as cocaine and amphetamine, whereas others experience that level of stimulation as extremely uncomfortable. Some people like the relaxation induced by drugs such as marijuana and sedative-hypnotics, whereas others feel deadened and overly slowed down by these drugs. Although some people gravitate toward particular drugs of abuse because of their specific pharmacologic properties, other will use a variety of drugs indiscriminately, based on level of availability; some of these people are primarily seeking to alter their current emotional state, regardless of the direction in which it is changed. The reinforcing properties of many drugs of abuse appear to be mediated through dopaminergic pathways, although other neurotransmitters, including γ-aminobutyric acid, serotonin, and norepinephrine, are also involved in mediating drug-induced reinforcement.

The Environment

The third critical factor in the development, maintenance, and perhaps cessation of drug abuse is the environment in which the use occurs. Drug use does not occur in a vacuum. Rather, many societal factors, including legal status, availability, price, perception of dangerousness, social desirability, peer group, and religious beliefs, influence behavior relating to substance use. Drug availability is known to be a substantial influence on likelihood of substance use. For example, alcohol consumption has been shown to increase when the hours during which alcohol can be sold are extended. The restriction of alcohol availability by restricting hours of sale or by increasing its cost through taxation in turn reduces consumption. Illicit drugs are, of course, by definition less available than alcohol or tobacco. A major factor that influences use of these agents is the potential user's perception of the drug's safety, social cachet (or lack thereof), likelihood of incurring legal consequences, and peer group behavior. Treatment research has shown that environmental influences can have a powerful effect on drug use. Studies have shown, for example, that offering an alternative positive reward (e.g., a voucher that can be exchanged for desired goods and services such as movie tickets or clothes) in response to abstaining from drugs may help drug-dependent individuals overcome their severe craving and reduce their substance use. In fact, this type of treatment approach, based on the use of motivational incentives for abstinence, has been shown to be one of the most powerful treatment interventions available for the treatment of drug use disorders. The impact of environmental contingencies demonstrates the importance of appreciating the complexity of the interaction among the individual, the drug, and the environment in the determination of drug use.

CLINICAL MANIFESTATIONS

Medical Complications Related to Drug Use Disorders

Drug use disorders are associated with significant medical morbidity and sometimes with mortality. Medical complications are often directly related to the pharmacology of the abused agent, for example, the vasoconstrictor properties of cocaine; drug-specific complications are described later in the sections focusing on particular drugs of abuse.

In addition to these drug-specific sequelae, however, many medical complications incurred by patients with drug use disorders occur not as a result of the particular drug being abused. Rather, serious complications may occur as a result of three factors that cut across many of the drugs of abuse: paraphernalia, particularly unsterile needles; adulterants; and lifestyle issues.

Paraphernalia

Some of the most serious medical problems that occur in individuals with drug use disorders are a result of the route of administration rather than of the actual drug being used. The use of unsterile needles, particularly if they are shared with other drug users, can lead to a variety of localized and systemic infections, some of which can be life-threatening. Skin infections and cellulitis are relatively common among injection drug users. Systemic infections related to needle use are often serious; individuals who inject drugs may develop infective endocarditis (Chapter 76). Other relatively common infections among injection drug users include hepatitis B, hepatitis C, and HIV infection.

Adulterants

Drugs that are purchased and sold illicitly are often adulterated or "cut" with other similar-looking products, with the intention of increasing the dealer's profit margin. For example, other white powdery substances are typically added to cocaine and heroin during the dealing process to dilute their purity. Some of these adulterants can in turn cause medical problems.[3] At times, these complications occur because of the combined toxicity of the adulterant and the route of administration. Thus, for example, a patient may have granulomas in the lung or liver as a result of talc use; talc is commonly added to street heroin and can also cause difficulties in users who crush talc-containing pharmaceutical tablets (e.g., opioids) and then inject them. Other common adulterants in street drugs include quinine (frequently used with heroin) and lidocaine or levamisole (often added to cocaine), but such toxic materials as strychnine and ground glass have been found in samples of street drugs, leading to serious medical sequelae.

Lifestyle Issues

Many patients with drug use disorders expose themselves to multiple risks due to intoxication, participating in dangerous illegal activities, and associating with potentially violent people. As a result, these individuals experience a high rate of traumatic injuries and are at greater risk of being victims of assault, homicide, or suicide. Suicide is far more common among people with substance use disorders than among the general population; this may be related to a combination of the effects of acute intoxication, the high prevalence of depression among these individuals, and the higher rate of antisocial personality disorder in this population, which is associated with a propensity toward impulsiveness, risk taking, and violence. Although it is well known that intoxication can lead to motor vehicle crashes, intoxication can also serve as a risk factor for becoming a victim of someone else's vehicle; one study reported that one third of pedestrians who are killed by motor vehicles have alcohol in their blood, perhaps a reflection of the combination of risk taking, poor judgment, and impaired motor coordination that can occur during periods of intoxication.

TREATMENT

General Treatment Principles

Drug use disorders represent a heterogeneous group of disorders. These are based on type of drug or drugs used; frequency and amount of use; severity of medical, behavioral, and social consequences; presence and severity of comorbid medical and psychiatric illness; and motivation to change. Treatment thus requires a careful medical and psychiatric assessment, including a detailed substance use history and laboratory testing. It is often helpful to enlist the help of a family member or significant other (with the patient's permission) in obtaining historical information. Intoxication and withdrawal syndromes need to be treated acutely; longer-term treatment involves helping the patient reduce or ideally abstain from substances of abuse and thus improve overall functioning.

Among the common drugs of abuse, medications with approval by the U.S. Food and Drug Administration (FDA) are available only for opioids and nicotine (nicotine is discussed in Chapter 32). However, researchers are actively studying a number of compounds for the treatment of other drugs of abuse, particularly stimulants and marijuana. Behavioral treatments are critically important in the treatment of substance use disorders. A number of behavioral treatments have a substantial evidence base supporting their efficacy; these include cognitive-behavioral therapy, motivational enhancement therapy, contingency management (also referred to as motivational incentive) therapy, 12-step facilitation therapy, and behavioral couples therapy. In addition to professional treatment, peer support groups such as the 12-step–oriented Alcoholics or Narcotics Anonymous and non–12-step groups such as SMART Recovery can be extremely helpful in facilitating recovery from drug abuse problems.

MAJOR DRUGS OF ABUSE

Opioids

For centuries, opioids have been a core part of the medical pharmacopoeia, primarily because of their capacity to treat pain but also because of their antitussive and antidiarrheal properties. Unfortunately, opioids are also powerful euphoriants and thus have substantial abuse liability. Although opium itself has been used for centuries, the isolation of morphine and codeine from opium in the 19th century along with the introduction of the hypodermic needle led to the increased prevalence of intravenous opioid use. Ironically, heroin was introduced near the end of the 19th century as a treatment for morphine addiction.

Opioids can be divided into four categories: natural opium alkaloids, including opium, morphine, and codeine; semisynthetic derivatives of morphine, including heroin and oxycodone; synthetic opioids that are not derived from morphine, including methadone and meperidine; and opioid-containing preparations, such as elixir of terpin hydrate.

EPIDEMIOLOGY

Opioid dependence represents a significant public health problem and accounts for more admissions for substance use disorder treatment than any substance other than alcohol. In the past two decades, there has been a shift in the epidemiology of opioid use disorders, however, with a reduction in heroin use and an increase in the abuse of opioid analgesic drugs; the latter has occurred as a result of either misuse of prescription opioids or illicit use of these agents. Approximately 620,000 people in the United States reported using heroin in 2011, with 178,000 trying the drug for the first time. During the same year, more than 11 million people either misused prescription opioids or used them illicitly, with 1.9 million doing so for the first time; opioids are currently the most commonly misused prescription drugs. Most people who use opioid analgesics in this way report that they initially obtained them from a friend or relative. It is thus likely that a portion of these people might, for example, have used a relative's opioid for the treatment of a temporary painful condition such as a migraine headache. However, the number of people seeking treatment as a result of an opioid analgesic use disorder has increased dramatically in the past decade.

PATHOBIOLOGY

Opioids are readily absorbed when they are taken orally, intranasally, or by smoking or injection. Heroin, which is almost immediately converted to morphine in the liver, is most commonly injected but may be smoked or used intranasally.

Opioids work by binding to specific opioid receptors and then exerting their activity. The major subtypes of opioid receptors have been identified and well described. Most of the commonly abused opioids bind as agonists to the μ-receptor and typically produce the effects most commonly associated with opioids: miosis, respiratory depression, analgesia, euphoria, and drowsiness. Opioids that bind to the κ-receptor, unlike μ-receptor agonists, often produce dysphoria rather than euphoria. The other two receptors, δ- and N/OFQ-receptors, do not appear to play a known significant role in opioid use disorders.

Opioid analgesics are ordinarily taken by the oral route, but they may be altered to be used through a different route of administration. This is particularly common with the extended- release preparations, which may be altered by chewing the pill (facilitating a rapid release of the opioid medication) or by crushing the pill, dissolving it in water, and then injecting it or using it intranasally.

CLINICAL MANIFESTATIONS

The initial response to the administration of heroin, particularly when it is used intravenously, is a "rush," often described as orgasmic, lasting 30 to

60 seconds. This sensation is generally followed by a profound sense of relaxation that is sometimes referred to as being "wrapped in warm cotton." During this period, the user generally feels drowsy and may be seen to be "nodding," with mental clouding and a sense of tranquility. A reduction in respiratory rate occurs, along with miosis, reduced contractility of smooth muscle, and reduced secretions in the stomach, pancreas, and biliary tract. Thus, constipation and urinary hesitancy may occur. Itching is commonly seen during opioid intoxication. Many people experience nausea and vomiting in their initial use of opioids, although tolerance tends to develop to this effect over time. Tolerance also occurs to some other effects rather quickly, particularly the analgesic, respiratory depressant, and euphoriant properties of opioids. In contrast, relatively little tolerance occurs to constipation or to pupillary constriction. It is important to be aware, then, that miosis is a manifestation of opioid use, but it is not diagnostic of opiate overuse or intoxication.[4]

Physical Dependence

Physical dependence on opioids leads to a characteristic withdrawal syndrome, the key signs of which include elevated heart rate and blood pressure, mydriasis, abdominal cramps, sweating, gooseflesh, rhinorrhea, lacrimation, and gastrointestinal distress, particularly diarrhea, nausea, and vomiting. Insomnia is common, particularly difficulty in falling asleep; this is often the most long-lasting complaint among people who experience opioid withdrawal. Yawning, muscle twitches, and difficulty with body temperature regulation are also commonly seen. The severity of withdrawal can be highly variable, depending on the dose of opioids taken, the length of time that they have been taken, and individual factors. For short-acting opioids such as heroin and hydrocodone, the earliest stages of withdrawal typically occur approximately 6 to 12 hours after the last use. Peak symptoms tend to occur 48 to 72 hours after the last dose, and most clinical symptoms usually resolve within 7 to 10 days. For longer-acting opioids such as methadone, each of these time periods associated with withdrawal from short-acting opioids should be approximately doubled or tripled.

Other Medical Complications

The most common serious medical complications that occur from opioid use are typically related to factors other than the opioids themselves, particularly needle use and adulterants; these were discussed previously. Common medical problems among heroin users include hepatitis B, hepatitis C, infective endocarditis, talc granulomatosis, HIV infection, cellulitis, and abscesses, all typically related to needle use.

An important noninfectious complication that has been reported with opioid use disorders is alteration of the cardiac conduction system, with a prolongation of the QT interval; this can lead to potentially serious arrhythmias, including torsades de pointes. This complication has been particularly noted with methadone.

Chronic pain is commonly seen among individuals with opioid use disorders, not just those who have received opioids for the treatment of pain. Pain can occur in these individuals for numerous reasons. In addition to the possibility that a chronic painful condition led to the use of opioids in the first place, those dependent on opioids are more likely to experience accidents, violence, and other forms of physical trauma that could produce chronic pain. There is also some evidence that chronic opioid use may lead to hyperalgesia, although there is some controversy regarding this issue. As with all substance use disorders, psychiatric illnesses (particularly mood disorders) are more common in those with opioid use disorders than in the general population. Moreover, the use of multiple drugs is common in patients with opioid use disorders, particularly among those using heroin. Indeed, the use of more than one drug is typically the rule rather than the exception in most substance use disorders.

TREATMENT

Opioid Withdrawal

Opioid detoxification can be accomplished by switching patients from their current drug of abuse (e.g., heroin, hydrocodone) to methadone or buprenorphine and then tapering that medication. Although the details of accomplishing this vary, one method commonly used in hospital settings is to administer methadone 10 mg orally whenever a patient experiences objective signs of opioid withdrawal (e.g., mydriasis, tachycardia, hypertension, and sweating). This process can be repeated every 2 to 4 hours for 24 hours after the initial dose; the total amount of methadone given in that 24-hour period is the "stabilization dose," which should not ordinarily exceed 40 mg. The stabilization dose is then reduced by 5 mg a day until the detoxification is completed.

Buprenorphine can also be used successfully for opioid detoxification[■]; patients who demonstrate objective signs of opioid withdrawal (often measured with a standardized withdrawal severity scale) can be stabilized with buprenorphine during a 1- to 2-day period; the subsequent taper from buprenorphine may occur either right away or after a period of stabilization with buprenorphine. The dose of buprenorphine will depend on whether the medication will be used for a brief, several-day detoxification or for longer-term stabilization or maintenance treatment.

Longer-Term Opioid Dependence Treatment

Three effective medications are approved by the FDA for the treatment of opioid dependence: methadone (a full opioid agonist), buprenorphine (a partial agonist), and naltrexone (an opioid antagonist). Methadone has been used successfully for both detoxification from opioids and maintenance treatment for many years.[■] Unlike buprenorphine and naltrexone, which can be prescribed by physicians in their offices (although physicians wishing to prescribe buprenorphine for the treatment of opioid dependence must receive specialized training and certification to do so), methadone is available for the treatment of opioid dependence only in specially licensed treatment programs. Methadone is a long-acting μ-receptor agonist with a slow onset of peak effects (typically approximately 2 to 6 hours) and a slow offset of action, allowing for once-a-day administration. Methadone reduces opioid craving and induces cross-tolerance, thus blocking or attenuating the effects of other opioid use. Although the therapeutic dose of methadone for a particular individual may vary, doses of 60 mg or higher have typically been shown to be more effective than lower doses; there is some evidence that even higher doses (e.g., 80 mg a day or more) may be more effective than 60 mg. Methadone treatment has been shown to reduce opioid use, to increase employment, to decrease criminal behavior, and to reduce the rate of development of HIV infection.

When a patient enrolled in a methadone treatment program experiences pain (e.g., postoperatively) requiring opioid analgesia, the patient should continue to receive the baseline methadone maintenance treatment dose for the addiction and should receive a different opioid for treatment of the pain (Chapter 30); before administering methadone, it is a good idea to confirm the methadone dose with the patient's treatment program. The fact that the patient is receiving methadone every day does not obviate the need for opioid analgesia, however. In fact, many patients receiving methadone treatment for opioid dependence will require a dose of opioids that is relatively high as a result of cross-tolerance to other opioid drugs.

The Drug Addiction Act of 2000 revolutionized the treatment of opioid dependence by enabling the approval of the partial opioid agonist buprenorphine for the treatment of opioid dependence and allowing treatment with buprenorphine to be administered in physicians' offices rather than exclusively in specialized opioid treatment programs. To prescribe buprenorphine, physicians must apply to the Substance Abuse and Mental Health Services Administration for a waiver that allows them to prescribe buprenorphine, after taking an 8-hour training course on buprenorphine. At the time of this writing, physicians may treat up to 100 patients with buprenorphine in their office practice. Buprenorphine, a partial μ-agonist and κ-antagonist, has a more favorable safety profile than methadone because of its partial agonist properties. Respiratory depression, which can be induced by full agonists and is responsible for some overdose deaths, is far less likely to occur with buprenorphine because its partial agonist properties cause a plateau of opioid effects as the dose increases. Buprenorphine is administered sublingually, in either tablet or film form, for the treatment of opioid dependence either as buprenorphine alone (sometimes referred to as the "mono" product) or (more commonly, in the United States) as a combination product of buprenorphine and naloxone; the naloxone is added to discourage users from dissolving and injecting the medication because the naloxone in the combination product will precipitate withdrawal when it is injected. Buprenorphine has been shown to be effective for both opioid detoxification and maintenance treatment. Randomized trials have shown that extended treatment with buprenorphine-naloxone has produced far better outcomes compared with short-term detoxification in opioid-addicted youths aged 15 to 21 years.[■] Typical doses of 12 to 16 mg of sublingual buprenorphine per day appear to be as effective as methadone in doses up to approximately 60 mg a day. Buprenorphine is also effective for maintenance therapy of heroin addiction[■], but individuals who require much higher doses of methadone may respond better to that agent than to buprenorphine.[■] To address problems with adherence, diversion, and nonmedical use, an implantable formulation of buprenorphine has been developed that provides a low, steady level of the drug during 6 months.[■]

Naltrexone, a pure opioid antagonist, blocks the effects (including euphoria) of opioids. As a result, individuals taking naltrexone should have a reduced desire to use opioids because they will have no desired effect. When it is used orally (50 mg/day) or in its long-acting form (380 mg intramuscularly every 4 weeks), naltrexone is highly effective at suppressing illicit opioid use. A

randomized clinical trial in prescription opioid-dependent outpatients has also demonstrated the effectiveness of naltrexone maintenance after buprenorphine taper. However, naltrexone has traditionally suffered from low acceptability; few patients have been interested in being treated with oral naltrexone. Moreover, among those who initially accept this treatment, the dropout rate is extremely high. An advantage of extended-release injectable naltrexone is that it can mitigate the traditionally poor adherence associated with the oral formulation; naltrexone can be a useful medication for patients who are willing (either as the result of external pressure or for internal motivation) to use it.

Methadone, buprenorphine, and naltrexone are not designed to be delivered alone but should be given in conjunction with counseling to be effective. It has been well demonstrated that the administration of methadone in the absence of counseling is an inadequate treatment approach; less is known about the optimal combination of buprenorphine and counseling. Two studies have demonstrated that counseling delivered within a medical office setting can be effective in conjunction with buprenorphine treatment, suggesting that general physicians who are trained to use buprenorphine can effectively treat at least a portion of patients with opioid dependence in their offices with a combination of buprenorphine and counseling.

Central Nervous System Stimulants: Cocaine and Amphetamines

The two most important central nervous system stimulants, cocaine and amphetamine (including methamphetamine[5]), are derived from different sources; cocaine is extracted from coca leaves, whereas amphetamine is a synthetic compound. However, both induce similar psychoactive activity when they are taken illicitly and can produce similar adverse consequences. Amphetamine has been used over the years to treat obesity and to combat fatigue and depression. Cocaine is still used as a topical anesthetic for otolaryngologic surgery. Ironically, its vasoconstrictor action, which is responsible for many of the cocaine-related medical complications described later, can be valuable for surgeons because of the resultant reduction of blood flow in the operating field. Although cocaine was not extracted from the coca leaf until the 19th century, coca leaves have been chewed for more than 1500 years for medicinal and religious purposes as well as to combat work-related fatigue. Sigmund Freud was one of the foremost advocates of cocaine, both extolling its psychoactive properties and discovering its ability to relieve pain, thus eventually leading to its discovery as the first local anesthetic. Cocaine was seen in the late 19th century as a "cure-all" and was included in numerous patent medicines as well as in Coca-Cola. The Harrison Narcotic Act of 1914 restricted the use of cocaine, and the drug was not widely used until the late 1970s, when there was a resurgence in cocaine use in the United States.

Like cocaine,[6] amphetamine was synthesized for the first time in the late 19th century. It was used for clinical purposes for the first time in the 1920s. Reports of amphetamine abuse first occurred in the 1930s, with intermittent epidemics since that time. In recent years, methamphetamine[5] abuse has been particularly prevalent and worrisome in the United States, with particularly high concentration of its use in the Midwestern and Western states, including Hawaii.

Cocaine can be used intranasally, by intravenous injection, or by smoking. Cocaine hydrochloride, which is the form of the drug used in medical therapeutics, is a water-soluble compound that can be used intranasally ("snorted") or injected. Adding an alkaline compound such as baking soda to an aqueous solution of cocaine hydrochloride produces a rocklike compound known as crack, which can be smoked. Smoking cocaine produces the most rapid onset of intoxication (6 to 10 seconds) and the shortest period of drug effect (10 to 15 minutes). Methamphetamine can also be used in multiple ways—orally, by smoking, or intravenously. Methamphetamine effects last much longer than those produced by cocaine; psychiatric symptoms such as paranoia that typically last only a matter of hours in cocaine users may persist for days to weeks after methamphetamine use and occasionally may result in a chronic psychotic state.

EPIDEMIOLOGY

Approximately 37 million Americans have used cocaine during their lifetime. In 2011, just less than 4 million people used cocaine; 1.4 million people, or 0.5% of the population, reported using cocaine in the past month. Of these past-month cocaine users, 17% (228,000) used crack, a decrease from 2010. Cocaine had the third highest rate of drug abuse or dependence in 2011 and was the third most common drug (behind opioids and cannabis) associated with a recent treatment episode; cocaine is also the drug of abuse most commonly involved in emergency department visits.

Other stimulant use is less common; approximately 20 million Americans have used stimulants nonmedically during their lifetime, with 12 million methamphetamine users. Methamphetamine use in the past month is reported by more than 400,000 people in the United States.

PATHOBIOLOGY

Both cocaine and amphetamine increase the accumulation and activity of specific neurotransmitters in the synaptic cleft, including dopamine, norepinephrine, and serotonin. Cocaine is believed to exert this effect by binding to the dopamine transporter. Increased dopaminergic activity, particularly in the nucleus accumbens, is thought to be responsible for the reinforcing effects of cocaine. Amphetamines appear to increase the level of dopamine in the synaptic cleft primarily by stimulating presynaptic dopamine release as opposed to reuptake blockade.

CLINICAL MANIFESTATIONS

Both cocaine and amphetamines reliably produce euphoria, wakefulness, a sense of initiative, increased self-confidence (sometimes to the point of grandiosity), and, in some instances, sexual stimulation. With higher doses, users may feel "wired," a syndrome characterized by anxiety, irritability, and perhaps paranoia. Withdrawal from either of these agents leads to opposite effects from those of intoxication: increased appetite, hypersomnia, and depression, which can occasionally be serious. Medical complications related to cocaine use[6] are related to a combination of cocaine's stimulant activity (increased heart rate and blood pressure) and its vasoconstrictor properties. Local complications that result from the drug's vasoconstrictor activity include ulcerations of the nasal mucosa, perforation of the nasal septum, and decreased pulmonary diffusion capacity. Systemic complications include myocardial infarction, intracranial hemorrhage, grand mal seizures (as a result of intoxication, not withdrawal), and ventricular tachyarrhythmias, which may be responsible for sudden death. Physicians seeing a patient in an emergency department for an unexplained seizure should consider drug abuse as a potential cause. (Not only cocaine but also phencyclidine and meperidine intoxication may lead to seizures, as can sedative-hypnotic or alcohol withdrawal). A serum or urine toxicology screen may thus be an important diagnostic tool in such a situation.

TREATMENT

Rx

The treatment of stimulant use disorders primarily consists of behavioral therapies, including individual and group therapy, and self-help groups. Specific forms of treatment, such as cognitive-behavioral therapy, individual drug counseling by a 12-step–oriented disease model, and a behavioral treatment in which patients are reinforced for positive outcomes (e.g., drug-free urine screens), have been found to be successful. A great deal of research has been conducted in search of an effective pharmacotherapeutic treatment for stimulant dependence, but there is as yet no medication that has consistently been found to be effective enough to warrant approval by the FDA for this purpose.

Sedative-Hypnotic and Anxiolytic Drugs

Benzodiazepines and other sedative-hypnotic and anxiolytic medications such as barbiturates and zolpidem are frequently prescribed for the treatment of anxiety and sleep difficulties. Although different classifications of these drugs have very different chemical structures, they are grouped together according to their therapeutic applications. Most of these drugs act at the γ-aminobutyric acid type A receptor and can cause physical dependence and both dispositional and pharmacodynamic tolerance.

Because the benzodiazepines are far and away the most commonly prescribed sedative-hypnotics, they are also the most widely abused. There are two major patterns of benzodiazepine abuse. Many people who ultimately abuse these medications have initially received a legitimate benzodiazepine prescription for the treatment of anxiety or insomnia. However, a combination of tolerance and decreased effectiveness of the agent over time may lead some people to increase the dose on their own. In such circumstances, attempts by the physician to taper the person off of the medication can be very difficult.

A second pattern of benzodiazepine abuse occurs among individuals who are using other drugs of abuse, most commonly opioids or stimulants. For example, many individuals who are dependent on heroin or other opioids

may use benzodiazepines as a means of either enhancing the opioid effect or buffering symptoms of opioid withdrawal. Such individuals typically use relatively large doses of benzodiazepines intermittently, and therefore many of these patients do not develop physical dependence on benzodiazepines, unlike the first category of patients, for whom physical dependence is common.

TREATMENT

The treatment of people who are abusing benzodiazepines depends to some extent on the pattern of abuse. For individuals who have an anxiety disorder and have been misusing a legitimately prescribed medication, a common approach would be to taper the benzodiazepine and to institute a different type of treatment, such as an antidepressant along with cognitive-behavioral therapy. Tapering a benzodiazepine that a person has been taking for an extended time (sometimes many years) is often a slow process, with careful monitoring of withdrawal symptoms (anxiety, agitation, insomnia, tachycardia, palpitations). Because benzodiazepine withdrawal, like alcohol withdrawal, can precipitate a seizure, gradual withdrawal is preferred. Most patients tolerate a benzodiazepine dose reduction initially with relatively little difficulty. However, as with most drug withdrawal regimens, people experience their greatest discomfort toward the end of the taper. One reason for this is that the percentage dose reduction at the low end of a taper regimen continues to increase over time; a reduction from 2 mg to 1.5 mg of clonazepam, for instance, is a 25% reduction, whereas the same half-milligram dose reduction from 1 mg to 0.5 mg represents a 50% drop.

For patients who are abusing benzodiazepines as part of a pattern of multiple substance use, medical detoxification from the benzodiazepine itself will often be unnecessary; for this population, psychosocial approaches that advocate abstinence from all substances of abuse, in conjunction with appropriate pharmacotherapy as needed (e.g., in the case of opioid dependence), are preferred. However, some deaths have been reported in France and elsewhere as a result of combinations of buprenorphine and benzodiazepines, usually used parenterally. Thus, physicians who are treating patients who are abusing both opioids and benzodiazepines need to be mindful of this issue when considering the use of buprenorphine.

Marijuana

Marijuana, which refers to the dried leaves and flowers of the plant *Cannabis sativa*, has been used for its psychoactive and medicinal properties for centuries. The major psychoactive substance in marijuana is Δ^9-tetrahydrocannabinol (THC); the concentration of THC has increased from 1 to 3% in 1970 to nearly 10% in recent years.

EPIDEMIOLOGY

Marijuana is the most commonly used illicit drug in the United States; more than 120 million Americans have used marijuana, and approximately 22 million report that they have used marijuana in the past month. With the recent trend toward legalization of medical marijuana and outright legalization of marijuana in some states, those numbers will likely continue to increase.

PATHOBIOLOGY

Marijuana and other cannabinoids such as hashish (dried cannabis resin) exert their effects by binding to the cannabinoid receptors, of which two are currently known. Binding to the CB1 receptor, which is located primarily in the brain, appears to be responsible for the psychoactive effects of THC, whereas the CB_2 receptor may be associated with immune system responses.

CLINICAL MANIFESTATIONS

When marijuana is smoked, its psychoactive effects occur almost immediately, with peak intensity approximately 30 minutes later; effects tend to disappear within 3 hours. Oral administration of marijuana leads to a delayed onset of action, but the effects of the drug persist for a longer time. Because THC is highly soluble in lipids, it can be stored in fat depots of regular users for several weeks, sometimes longer, with resultant positive urine test results for THC. Physiologic effects of marijuana intoxication include increased heart rate and conjunctival injection. Psychological effects include a sense of euphoria and well-being, friendliness, increased appetite, a distorted sense of time, impaired short-term memory, and sometimes a feeling of having achieved special insights. Cannabis has the capacity to cause tolerance in regular users, and some regular heavy users experience withdrawal symptoms on cessation of use, including irritability, difficulty in sleeping, and anxiety.[7]

TREATMENT

The most common acute adverse event that occurs in marijuana smokers is a sense of acute panic, most common in inexperienced smokers, when the user's level of intoxication is greater than expected and the individual feels out of control. This can be best managed with reassurance that the effects will go away as the drug wears off. Recent evidence has shown that cannabis use, particularly during adolescence, may increase the likelihood for development of a psychotic disorder such as schizophrenia later in life.

Compared with alcohol, opioids, and stimulants, it has been relatively uncommon for people to seek treatment of cannabis use disorder itself. However, that situation has gradually changed in recent years, and an increasing number of people have sought treatment because of difficulty in stopping marijuana use. There are no medications approved by the FDA for the treatment of cannabis use disorder. Psychosocial approaches similar to the treatment of other substance use disorders are currently the treatment of choice.

Hallucinogens

Hallucinogens are a group of plant-based and synthetic drugs that lead to primarily visual perceptual alterations, such as illusions and hallucinations, along with an alteration in the experience of external stimuli; ordinary events can appear profound to people while they are under the influence of these agents. The most common hallucinogens are lysergic acid diethylamide (LSD), mescaline, and psilocybin. Methylene dioxymethamphetamine (MDMA), also known as ecstasy, has both mild stimulant and potentially hallucinogenic properties and is thus sometimes categorized with the hallucinogens and sometimes as a stimulant.

EPIDEMIOLOGY

Approximately 1 million people in the United States report having used a hallucinogen in the previous month; approximately 36 million people have used these drugs during their lifetime. LSD is the most commonly used hallucinogen, with 23 million lifetime users in the United States; approximately 60% of that number have used MDMA.

PATHOBIOLOGY

LSD is thought to exert its action through serotonin agonist activity, particularly at the 5-HT_{2A} receptor. Other neurotransmitters may be involved in hallucinogenic activity as well. Hallucinogens can produce tolerance in a matter of days, but they do not produce physical dependence.

CLINICAL MANIFESTATIONS

In addition to their effects on perception and behavior, hallucinogens can produce sympathomimetic effects such as tachycardia, increased blood pressure and body temperature, and pupillary dilation. Hyperreflexia and muscle weakness can also be seen. The most commonly seen medical consequence of hallucinogen use is hyperthermia, which can occur most commonly in users of MDMA.

The most common acute psychological adverse event, similar to marijuana, is a feeling of panic over the sense of loss of control that a person may feel as a result of intoxication; as is the case with marijuana, this is most likely to occur in inexperienced users. Some hallucinogen users will develop psychotic symptoms that fail to remit after the drug has worn off. Some hallucinogen use can also lead to longer-term perceptual difficulties. When these occur, a spontaneous return of very brief hallucinogen-induced symptoms long after the drug has worn off is known as a flashback. People who have perceptual difficulties that are much more pervasive may be said to have *hallucinogen persisting perception disorder*, which can at times be quite disabling.

TREATMENT

Symptomatic treatment is focused on the specific adverse medical and psychiatric sequelae described previously. If a psychotic episode that occurs after use of a hallucinogen persisted over time, it would be treated like any other psychotic disorder. There is no specific treatment for hallucinogen use disorder, and it is uncommon for people to seek treatment specifically because they want to stop using hallucinogens.

Phencyclidine

Phencyclidine (PCP) was originally developed as a human general anesthetic, but its use for that purpose was stopped in the 1960s because it frequently led to psychosis and hallucinations in the postoperative period. Approximately 120,000 individuals report that they used PCP during 2011. Low doses of PCP can lead to symptoms that resemble alcohol intoxication, with slurred speech, ataxia, and a subjective feeling sometimes described as "feeling dead." PCP intoxication typically is accompanied by increased muscle tone, hyperreflexia, nystagmus, and ataxia.

When it is taken in high doses, PCP can have serious medical and psychiatric consequences. High-dose users may experience psychosis, catatonia, and extremely violent behavior. Medical sequelae of PCP intoxication can include muscle rigidity, seizures, hyperthermia, coma, and occasionally death.

Anabolic-Androgenic Steroids

Anabolic-androgenic steroids differ from other drugs described in this chapter because the motivation for use is typically related to the drug's physical rather than behavioral effects. Anabolic-androgenic steroids such as testosterone and its synthetic analogues have traditionally been used primarily to enhance strength and thus athletic performance, although in recent years, an increasing number of people have used these drugs primarily in an attempt to improve their physical appearance. Anabolic-androgenic steroids can have a legitimate medical purpose; they have most commonly been used to treat testosterone deficiency in men and more recently have been used to treat wasting syndromes in patients with AIDS.

Abuse of anabolic-androgenic steroids can cause a number of medical and psychiatric problems, including hypertension, elevated low-density lipoprotein cholesterol, cardiomyopathy, hepatotoxicity, acne, feminization (gynecomastia and reduced testicular size) in men, and masculinization (hirsutism, reduction in breast tissue, deeper voice) in women. Behavioral effects include aggressiveness (sometimes leading to violence) and an increased prevalence of mood disorders. There is no specific treatment to help people abusing anabolic steroids to stop. Rather, behavioral treatment approaches that are commonly used to treat other substance use disorders should be employed with this population.

CONCLUSIONS AND FUTURE DIRECTIONS

The ever-changing epidemiology of drug abuse means that the next decade will likely present new challenges as new drugs of abuse become increasingly popular. Recent research has focused on development and testing of effective pharmacologic and behavioral treatments for substance use disorders. Screening for these disorders in general medical practice and combining office-based interventions with referrals to specialty substance use disorder treatment as indicated can lead to successful outcomes for many of these patients.

Grade A References

A1. Gowing L, Ali R, White JM. Buprenorphine for the management of opioid withdrawal. *Cochrane Database Syst Rev*. 2009;3:CD002025.

A2. Mattick RP, Breen C, Kimber J, et al. Methadone maintenance therapy versus no opioid replacement therapy for opioid dependence. *Cochrane Database Syst Rev*. 2009;3:CD002209.

A3. Minozzi S, Amato L, Bellisario C, et al. Maintenance treatments for opiate-dependent adolescents. *Cochrane Database Syst Rev*. 2014;6:CD007210.

A4. Mattick RP, Breen C, Kimber J, et al. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. *Cochrane Database Syst Rev*. 2014;2:CD002207.

A5. Weiss RD, Potter JS, Fiellin DA, et al. Adjunctive counseling during brief and extended buprenorphine-naloxone treatment for prescription opioid dependence a 2-phase randomized controlled trial. *Arch Gen Psychiatry*. 2011;68:1238-1246.

A6. Ling W, Casadonte P, Bigelow G, et al. Buprenorphine implants for treatment of opioid dependence: a randomized controlled trial. *JAMA*. 2010;304:1576-1583.

A7. Sigmon SC, Dunn KE, Saulsgiver K, et al. A randomized, double-blind evaluation of buprenorphine taper duration in primary prescription opioid abusers. *JAMA Psychiatry*. 2013;70:1347-1354.

A8. Krupitsky E, Nunes EV, Ling W, et al. Injectable extended-release naltrexone for opioid dependence: a double-blind, placebo-controlled, multicentre randomised trial. *Lancet*. 2011;377:1506-1513.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hasin DS, O'Brien CP, Auriacombe M, et al. DSM-5 criteria for substance use disorders: recommendations and rationale. *Am J Psychiatry*. 2013;170:834-851.
2. Degenhardt L, Whiteford HA, Ferrari AJ, et al. Global burden of disease attributable to illicit drug use and dependence: findings from the Global Burden of Disease Study 2010. *Lancet*. 2013;382:1564-1574.
3. Cole C, Jones L, McVeigh J, et al. Adulterants in illicit drugs: a review of empirical evidence. *Drug Test Anal*. 2011;3:89-96.
4. Boyer EW. Management of opioid analgesic overdose. *N Engl J Med*. 2012;367:146-155.
5. Panenka WJ, Procyshyn RM, Lecomte T, et al. Methamphetamine use: a comprehensive review of molecular, preclinical and clinical findings. *Drug Alcohol Depend*. 2013;129:167-179.
6. Karila L, Petit A, Lowenstein W, et al. Diagnosis and consequences of cocaine addiction. *Curr Med Chem*. 2012;19:5612-5618.
7. Volkow ND, Baler RD, Compton WM, et al. Adverse health effects of marijuana use. *N Engl J Med*. 2014;370:2219-2227.

REVIEW QUESTIONS

1. Which of the following substances is NOT associated with withdrawal seizures?

A. Cocaine
B. Alprazolam
C. Butalbital
D. Alcohol
E. Chlordiazepoxide

Answer: A Cocaine causes a seizure upon intoxication, not withdrawal. Sedative-hypnotic agents cause withdrawal seizures.

2. Which of the following symptoms is seen as a result of opioid withdrawal?

A. Grand mal seizure
B. Miosis
C. Constipation
D. Tachycardia
E. Synesthesia

Answer: D Miosis and constipation occur with intoxication, seizures are not seen with opioid withdrawal, and synesthesia (seeing sounds, hearing colors) is seen in hallucinogen intoxication.

3. Which of the following symptoms does NOT count as a symptom of a substance use disorder?

A. Physical dependence in a patient taking opioids as prescribed for pain
B. Craving
C. Taking more than intended
D. Use despite family problems
E. Persistent desire to cut down on drug use

Answer: A Physical dependence and tolerance are not considered criteria for a substance use disorder if they occur as a result of appropriate medical care.

4. Which of the following is NOT an FDA-approved medication for treatment of opioid dependence?

A. Acamprosate
B. Oral naltrexone
C. Extended-release injectable naltrexone
D. Methadone
E. Sublingual buprenorphine

Answer: A Acamprosate is approved for the treatment of alcohol use disorders; the other medications are approved for the treatment of opioid dependence.

5. Other than marijuana, what type of drug is most widely abused?

A. Prescription drugs
B. Hallucinogens
C. Cocaine
D. Methamphetamine
E. Heroin

Answer: A Prescription drugs, especially opioids, are much more widely used nonmedically than the other drugs of abuse.

35

IMMUNOSUPPRESSING DRUGS INCLUDING CORTICOSTEROIDS

GRANT W. CANNON

IMMUNOSUPPRESSIVE DRUGS

The immune response is an essential host defense mechanism to control and fight infection. The ability to suppress immune reactions is a critical component in autoimmune disease treatment and transplantation management. During autoimmune diseases, the basic immune physiology is altered, and one or more components of this process do not function properly. Current challenges with the selection, dosing, monitoring, and development of immunosuppressing drugs involve identifying the component of the immune system to be altered by the immunosuppressive therapy while at the same time maintaining a competent immune response to fight infection and perform other important immunoregulatory functions. After organ transplantation, most patients have an immune response to reject the organ that has been implanted. Immunosuppression during transplantation management, in contrast to autoimmune disease, involves the suppression of normal immune reactions rather than an effort to alter a pathologic process. In this latter case, immunosuppressive therapy is directed to act on the altered immune system. Suppression of natural host immune responses affects the ability of these protective mechanisms to fight infection. The general principles for selecting immunosuppressive therapy in transplantation patients involve specific monitoring for organ rejection with subsequent selection of immunosuppressive therapy proportionate to the degree of rejection or for maintaining tolerance to the implanted organ. The selection of the most effective therapy requires an individualized treatment program. These decisions require an understanding of the underlying pathophysiologic process, prognosis, and potential adverse events for the agents selected.

This chapter describes the mechanism of action—including the components of the immune response affected by the therapy, indications, and adverse events associated with commonly used immunosuppressive agents in autoimmune diseases and transplantation. Whereas each of these agents has an individual discussion, these drugs are frequently used in combination as their complementary effects are employed. The understanding of the principles and adverse events associated with immunosuppressive therapy is important for all physicians as the use of these drugs becomes more widespread; however, their specific management and initiation, particularly in patients with organ transplantation and severe autoimmune diseases, should generally be limited to specialists with specific training in immunosuppressive therapies.

Corticosteroids

Corticosteroids are highly effective immunosuppressive agents and have had a major impact on the treatment of autoimmune diseases and the effort to avoid rejection of transplants. Historically,[1] the initial enthusiasm for the marked clinical benefit of corticosteroids in the treatment of rheumatoid arthritis was dampened by the significant adverse events that developed after prolonged use. As opposed to many adverse events that may develop as allergic and idiosyncratic reactions to medications, most adverse events with corticosteroid therapy are a direct consequence of the physiologic effects of the drug. This observation sparked a determined effort to understand the mechanisms of actions of corticosteroids at physiologic levels and therapeutic doses. The objectives of these investigations have been to develop modifications of the naturally occurring hormones to exploit the clinical benefits provided by corticosteroids while avoiding the associated adverse effects.

Mechanism of Action

Corticosteroids affect multiple physiologic functions at the molecular, cellular, and organ level. The final result of corticosteroid treatment represents the composite effects of the drug on these multiple functions, which vary with the particular agent, dose, route, and duration of treatment[2] (Table 35-1).

TABLE 35-1 CLASSIFICATION OF IMMUNOSUPPRESSIVE AGENTS

CORTICOSTEROIDS
Binding to cytosol glucocorticoid receptor: results in suppression of pro-inflammatory cytokines (genomic effects)
Inhibition of arachidonic acid release and binding to surface receptor (nongenomic effects)
Results of corticosteroid actions
Leukocyte numbers
Increase in circulating neutrophils
Decrease in circulating lymphocytes, monocytes, eosinophils, and basophils
Leukocyte function
Neutrophils: decrease in trafficking
Lymphocytes: decrease in cellular immune functions and immunoglobulin production
Cytokines
Decrease in pro-inflammatory cytokines: IL-1, IL-2, IL-6, and tumor necrosis factor-α
Increase in anti-inflammatory cytokines: IL-4, IL-10, and IL-13
Prostaglandins and leukotrienes: decreased production
PURINE PATHWAY INHIBITORS
Azathioprine: inhibition of DNA synthesis and purine synthesis
Mycophenolate mofetil: inhibition of purine synthesis
PYRIMIDINE PATHWAY INHIBITORS
Leflunomide: inhibits pyrimidine synthesis by inhibiting dihydroorotate dehydrogenase
IMMUNOPHILIN-BINDING AGENTS
Calcineurin inhibition
Cyclosporine: binds with cyclophilin to inhibit calcineurin, resulting in decreased T-cell activation
Tacrolimus: binds with FKBP12 to inhibit calcineurin, resulting in decreased T-cell activation
Mammalian target of rapamycin (mTOR) inhibition
Sirolimus: binds to FKBP12 to inhibit mTOR, resulting in decreased T-cell activation
ALKYLATING AGENTS
Cyclophosphamide: alkylation of nucleic acids with cytotoxic action

FKBP12 = 12-kD FK-binding protein; IL = interleukin.

Molecular Action

Genomic Effects

Corticosteroids are lipophilic and rapidly cross cell membranes into the cytosol, where they bind to the glucocorticoid receptor. The complex of glucocorticoid and its receptor then enters the nucleus and affects gene transcription by binding to glucocorticoid response elements. The complex may either stimulate or suppress gene transcription and subsequent protein production. This mechanism may have an impact on the function of 1% of all genes and suppresses the production of cytokines and other important inflammatory proteins. In addition to binding to glucocorticoid response elements, the glucocorticoid-glucocorticoid receptor complex also suppresses signal transduction pathways such as transcription factor activator protein-1 (AP-1), nuclear factor-κB (NF-κB), and nuclear factor of activator of T cells (NF-AT). Corticosteroids may also act to affect post-transcription and post-translation steps of protein synthesis. The anti-inflammatory actions of corticosteroids may be related to their action on the NF-κB and AP-1 pathways; the adverse events produced are related more to the activation or suppression of gene transcription. More recently, it has been recognized that corticosteroids can also regulate the mitochondrial genome. Glucocorticoid receptors are likewise present in mitochondria, and the mitochondrial genome contains glucocorticoid response elements.[3]

Nongenomic Effects

The genomic effects of corticosteroids require the diffusion of the drug into the cell, binding to the receptor, entry into the nucleus, and alteration of transcription. The ultimate effect on protein synthesis is not immediate, and it generally takes at least 30 minutes before any response is seen. The observation that some actions of corticosteroids are seen immediately has directed a search for nongenomic effects of corticosteroids. The glucocorticoid-glucocorticoid receptor complex can inhibit arachidonic acid release. In addition to the cytosol glucocorticoid receptor, membrane-bound receptors may be present that interact with cytoplasmic kinase–signaling molecules and G proteins and may mediate the rapid, nonclassic steroid effects and nongenomic functions.[2]

Systemic Effects

Impact on Leukocytes

Corticosteroids affect the activation, production, circulation, function, and survival of leukocytes. Whereas these impacts appear to be principally modulated by the genomic effects of corticosteroids on cytokines, corticosteroids also act on adhesion molecules and other mechanisms. The effects are on neutrophils, monocytes, macrophages, lymphocytes, eosinophils, and basophils. With corticosteroid therapy, neutrophils increase in the peripheral circulation (Chapter 167), primarily because of demargination, in contrast to a decrease in monocytes, lymphocytes, eosinophils, and basophils. Although the number of circulating neutrophils may increase, trafficking appears to be impaired. The impact on T cells is more pronounced than the effects on B cells, with the induction of apoptosis particularly in immature and activated T cells. Although function of B cells and neutrophils is not affected as strongly as that of T cells, high-dose prolonged use of corticosteroids can lead to suppression of antibody production.

Changes in Inflammatory Mediators

Corticosteroids result in a decrease in multiple pro-inflammatory cytokines and interleukins (ILs).[4] The cytokines affected include IL-1, IL-2, IL-6, and tumor necrosis factor-α, at the same time that there is an increase in anti-inflammatory cytokines—IL-4, IL-10, and IL-13. Corticosteroids have been associated with a reduction in the production of prostaglandins, leukotrienes, and other arachidonic acid metabolites, probably related to the reduced production of cyclooxygenase-2 and phospholipase A_2–related pro-inflammatory compounds (Chapter 37).

TREATMENT

Specific Issues with Corticosteroid Therapy

Multiple corticosteroid compounds and preparations are available. Many of the commonly employed compounds are listed in Table 35-2. These compounds have differences in potency, half-life, and sodium-retaining properties. More recently, delayed-release prednisone has been licensed for clinical use and may have a special role in the treatment of inflammatory conditions with circadian features.[5] In many conditions, the local administration of corticosteroids will provide clinical benefit without the systemic toxicity associated with oral therapy. Local therapies include topical, ophthalmic, inhaled, and local injection, such as soft tissue and intra-articular injections. Although the potential for adverse events is generally reduced with local therapy, local toxicities can develop, as can systemic effects if large doses of topical corticosteroids are used.

Whereas most conditions can be treated with local or oral corticosteroids, intravenous administration can provide pulse doses if desired. Intravenous therapy may also be used in situations in which the patient cannot take oral medication or the absorption of oral agents is impaired. Pulse therapy is generally administered in high intravenous doses that are often given as divided treatments during 3 to 5 days. High-dose pulse therapy has particularly been advocated in acute organ transplantation rejection, severe systemic lupus erythematosus (SLE), aggressive vasculitis, and other acute and severe autoimmune disorders. The use of high-dose pulse corticosteroid therapy has been associated with the development of sudden cardiac arrhythmias and sudden death. Many of the patients described have had serious concurrent diseases that could in part be responsible for electrolyte abnormalities and other associated morbidities that could contribute to these observations. Despite the confusion about this association, close monitoring of patients receiving high-dose pulse corticosteroid therapy is warranted.

Indications

Corticosteroids are employed in a large range of autoimmune disorders and transplantation procedures. A listing of each indication for which corticosteroids have been employed and proven effective is beyond the scope of this chapter, but indications range from multiple rheumatologic disorders to transplantation and many other inflammatory conditions. For example, the demonstration that inflammation plays a significant role in reactive airway disease has dramatically increased the use of systemic and inhaled corticosteroids in asthma and chronic obstructive pulmonary disorders. The challenge is to determine the appropriate dose, route, and duration of therapy with these disorders. For severe autoimmune disorders, high doses of corticosteroids are employed. In many cases, oral administration of prednisone, 60 to 80 mg/day as single or divided doses, can be employed. If patients cannot take oral

TABLE 35-2 GLUCOCORTICOID PREPARATIONS

	ANTI-INFLAMMATORY POTENCY	EQUIVALENT DOSE (mg)	SODIUM-RETAINING POTENCY	PLASMA HALF-LIFE (min)	BIOLOGIC HALF-LIFE (hr)
Hydrocortisone	1	20	2+	90	8-12
Cortisone	0.8	25	2+	30	8-12
Prednisone	4	5	1+	60	12-36
Prednisolone	4	5	1+	200	12-36
Methylprednisolone	5	4	0	180	12-36
Triamcinolone	5	4	0	300	12-36
Betamethasone	20-30	0.6	0	100-300	36-54
Dexamethasone	20-30	0.75	0	100-300	36-54

From Garber EK, Targoff C, Paulus HE. Glucocorticoid preparations. In: Paulus HE, Furst DE, Droomgoole SH, eds. *Drugs for Rheumatic Diseases.* New York: Churchill Livingstone; 1987:446.

medication or high doses of corticosteroids are indicated, higher doses can be administered intravenously. Comparative data on the most appropriate doses and routes of administration are generally not available. Clinical judgment and empirical literature have formed the basis of these treatment regimens.

Adverse Effects

As noted previously, most adverse events associated with corticosteroid use are caused by the physiologic action of these drugs. Investigations are ongoing to determine whether separate mechanisms might be more associated with the therapeutic benefits whereas other pathways are more involved in the adverse events of corticosteroids. Currently available forms of corticosteroids do not selectively allow separation of the adverse events from the therapeutic effects. For example, the increase in infection associated with corticosteroids is the result of the impact of these drugs on leukocyte function and antibody production and is not an allergic reaction. The prevalence and severity of these adverse effects increase in proportion to the dose and duration of the therapy. The key to reducing adverse events with corticosteroids is to use the lowest needed dose and shortest possible duration for the required indication. Data also suggest that intermittent and every-other-day dosing may be associated with less toxicity than daily or divided daily doses.

Despite these limitations, corticosteroids are the only viable treatment option in many conditions, and efforts must be made to prevent, to detect development of, and to monitor for these adverse events. In most situations, education of the patient coupled with vigilant surveillance can detect these adverse events and often reduce their serious impact.

The following description of adverse events is not intended to be a comprehensive list of all reported adverse events associated with corticosteroids. The problems with infection, osteoporosis, metabolic abnormalities, and cardiovascular effects are highlighted because specific interventions can have an impact on these complications through proper education of the patient, monitoring, or prophylactic therapy.

Infections

Infections are increased in patients taking corticosteroids, particularly bacterial, fungal, and mycobacterial infections. The increased incidence and general severity of these infections are complicated by the anti-inflammatory actions of corticosteroids that can mask many of the cardinal signs of infection, such as fever, inflammation, and local discomfort. Patients taking corticosteroids should be alerted to the possibility of these "subclinical infections," and the provider must be vigilant in investigating signs and symptoms that may be less concerning in patients not taking corticosteroids. Management of infections in patients receiving corticosteroids requires close monitoring. Appropriate diagnostic procedures, antimicrobial therapy, and supportive measures are keys to successful management of infections in immunocompromised hosts. Patients with adrenal suppression may require "stress doses" of corticosteroids during the initial treatment of an infection. However, when possible, a reduction in corticosteroid dose may help restore a host immune response to the underlying infection, particularly with chronic infections.

Preventive measures to reduce infections include proper immunizations. If possible, for example, in a patient being evaluated for future transplantation, immunizations should be given before the immunosuppressive therapy. In many cases, a delay of immunosuppressive therapy is not possible to allow immunizations to be updated. However, in patients receiving chronic immunosuppressive therapy, routine immunizations should be offered when the disease is stable. Prophylactic antibiotics are generally not recommended to prevent infections in patients receiving corticosteroid therapy. However, two notable exceptions are the use of antituberculosis therapy in purified protein derivative–positive patients and trimethoprim-sulfamethoxazole therapy in patients receiving high-dose corticosteroid therapy for *Pneumocystis jirovecii* (previously named *Pneumocystis carinii*) prophylaxis.

Osteoporosis

Bone loss (Chapter 243) with corticosteroids affects multiple sites and has a greater impact on trabecular bone than on cortical bone. Common sites for involvement are the spine and femur, with associated fracture rates that may be as high as 20%, depending on the duration and dose of therapy. On initiation of long-term corticosteroid therapy, all patients should be evaluated for osteoporosis prophylactic therapy (Chapter 243). Unless it is contraindicated, patients should have adequate calcium and vitamin D through diet or supplements. In many patients, bisphosphonates will provide significant protection. Postmenopausal women should be evaluated for potential estrogen replacement therapy that may be helpful if benefits of therapy with corticosteroids are considered to outweigh the risks for cardiovascular disease and malignant disease. If osteoporosis develops in a patient receiving corticosteroids, efforts should be made to discontinue therapy or to reduce dose. Agents employed for treatment of osteoporosis should also be employed. Effective therapies include bisphosphonates, hormone replacement therapy, and anabolic agents such as parathyroid hormone preparations. These agents are generally used in association with calcium and vitamin D supplementation. An algorithm for prevention and treatment of glucocorticoid-induced osteoporosis is provided (Fig. 35-1).[6]

Metabolic Effects

The metabolic effects of corticosteroids may affect glucose metabolism, with results ranging from mild glucose intolerance to frank diabetes. Patients beginning corticosteroid therapy should be monitored for glucose intolerance and treated if significant hyperglycemia develops. The management of patients with existing diabetes is particularly challenging during corticosteroid therapy and requires close monitoring and adjustments of the diabetes management program. Other metabolic complications of corticosteroid treatment include weight gain with truncal obesity, electrolyte abnormalities including hypokalemia, and fluid retention.

Tapering Steroids

Exogenous corticosteroids will suppress the hypothalamic-pituitary-adrenal (HPA) axis. The likelihood of adrenal suppression increases with dose and duration of therapy. HPA axis suppression should be considered in patients receiving doses of 20 mg/day or more for 3 weeks or longer, although suppression can occur with lower doses. Formal evaluations can be performed to test the integrity of the HPA axis, but in most cases a scheduled taper of corticosteroid dose during several weeks will allow the return of HPA axis function without signs of adrenal insufficiency. The ideal rate for tapering corticosteroids has not been evaluated by clinical trials. The rate of reduction

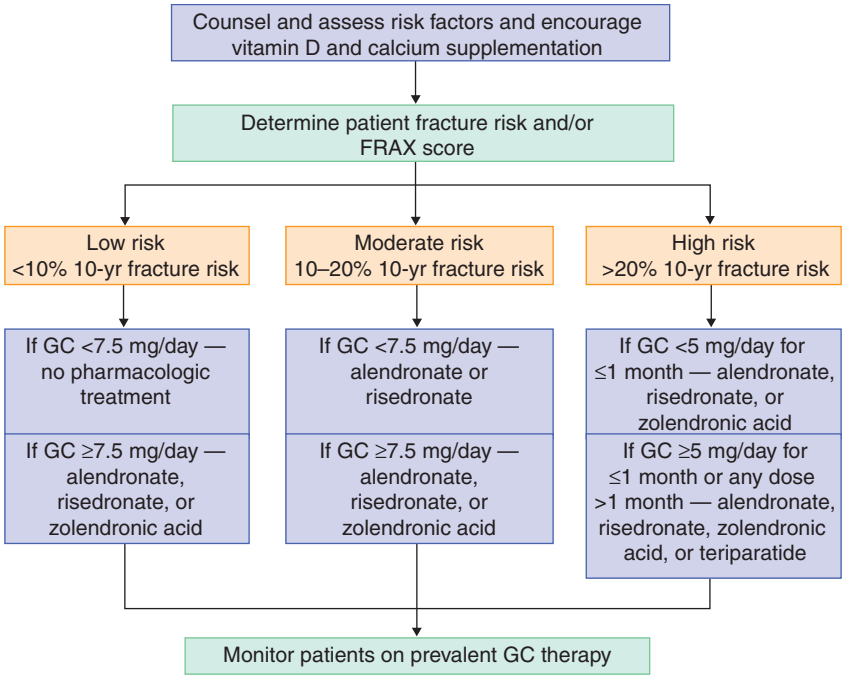

FIGURE 35-1. Algorithm for prevention of steroid-induced osteoporosis on initiation of glucocorticoid (GC) therapy. FRAX = World Health Organization Fracture Risk Assessment Tool. (Modified from American College of Rheumatology 2010 recommendations for the prevention and treatment of glucocorticoid-induced osteoporosis. *Arthritis Rheum.* 2010;62:1515-1526.)

in corticosteroid dose is often limited more by concerns over the potential relapses of disease activity than by the development of adrenal insufficiency. In general, if there is not a significant concern for a flare of the disease, a rapid reduction in dose to about 10 mg prednisone-equivalent per day is well tolerated. This dose replaces the normal physiologic production of cortisol. After this point, reductions in dose of 1 to 2.5 mg/day every 1 to 2 weeks will generally be well tolerated and may be accomplished by decreasing the dose of alternate-day treatment over 6 to 8 weeks. However, with acute medical illness, patients who have received corticosteroid doses in the past sufficient to cause HPA suppression should receive stress doses of steroids for a period of up to 1 year after the steroid treatment.

Cardiovascular Effects

Corticosteroids may induce or exacerbate cardiovascular risk factors, including hypertension, hyperlipidemia, and diabetes. Multiple mechanisms have been proposed for these effects, but the end result is that patients taking corticosteroids have an increased prevalence of atherosclerotic diseases and their associated complications. This problem is further complicated by the increased risk for cardiovascular disease in patients with inflammatory conditions, including rheumatoid arthritis and SLE, above that risk predicted by traditional risk factor assessment. These observations have emphasized the need to monitor corticosteroid-treated patients closely for cardiovascular risk factors with aggressive treatment of detected abnormalities.

Other Adverse Effects

Many other adverse events have been noted with corticosteroid therapy and are listed in Table 35-3. Osteonecrosis (avascular necrosis) is common during corticosteroid therapy and particularly involves the femoral head. Peptic ulcer disease is increased independently of concurrent nonsteroidal anti-inflammatory therapy. Cataracts and a variety of dermatologic abnormalities are more common. Muscle weakness or steroid myopathy, alteration of mood and behavior, and psychosis may develop. This broad spectrum of clinical complications requires the prescriber of corticosteroids to be aware of and alert to the development of these adverse events.

Purine Inhibition

Purines are critical components of nucleic acids and are particularly important in proliferating cells as part of cell growth and division. The inhibition of purines by competitive inhibitors (azathioprine and 6-mercaptopurine) or blocking critical enzymes (mycophenolate mofetil) in the purine pathway is an effective method of immunosuppression.

Azathioprine and 6-Mercaptopurine

Mechanism of Action

Azathioprine is an inactive compound that is metabolized to the active compound 6-mercaptopurine (6-MP). The exact mechanisms of action of 6-MP and its metabolites have not been fully established. At high doses, 6-MP may be incorporated into RNA and DNA, resulting in a cytotoxic effect; however, this effect is probably not the major action of the drug at the doses generally employed for immunosuppression. Most likely through feedback inhibition of de novo purine synthesis, 6-MP and its metabolites may reduce cell proliferation and thus produce immunosuppression. Genetically controlled differences in the activity of enzymes involved in the metabolism of 6-MP have been identified. The enzyme thiopurine *S*-methyltransferase is responsible for the metabolism of 6-MP to the metabolite methyl-6-MP. A rare homozygous (0.3%) and heterozygous (10%) defect in thiopurine *S*-methyltransferase is associated with increased toxicity, with severe hematologic toxicity in the homozygous patients.

The drug is metabolized eventually by xanthine oxidase. Because xanthine oxidase is inhibited by allopurinol, the concurrent use of allopurinol and azathioprine or 6-MP can result in a significant reduction in the metabolism of the active compounds and a significant increase in drug toxicity. For this reason, the combination of allopurinol and azathioprine should be avoided.

Indications

Azathioprine is approved by the U.S. Food and Drug Administration (FDA) for prevention of rejection of renal transplantation and treatment of rheumatoid arthritis. Clinical trials and reports have suggested that azathioprine also has efficacy during other types of organ transplantation and for autoimmune diseases. Azathioprine has been particularly effective at providing an adjunct to corticosteroid therapy, allowing a reduction in corticosteroid dose and avoiding the associated adverse events.

Adverse Effects

Serious infections are reported during treatment with azathioprine, similar to those seen with other immunosuppressive drugs. Opportunistic infections are a particular concern. Hematologic abnormalities include leukopenia, thrombocytopenia, and anemia. A complete blood count (CBC) is recommended on a regular basis, with increased frequency at the initiation of therapy. Current guidelines recommend a CBC weekly during the first month of azathioprine therapy, twice weekly during the second and third months, and monthly thereafter. Genotyping of the enzyme thiopurine *S*-methyltransferase may identify subjects with the highest risk for

TABLE 35-3 MAJOR ADVERSE EVENTS ASSOCIATED WITH IMMUNOSUPPRESSIVE THERAPIES*

CORTICOSTEROIDS
Serious and opportunistic infections
Osteoporosis
Metabolic disorders: hyperglycemia, adrenal suppression, hyperlipidemia, electrolyte abnormalities, fluid retention, hypertension, truncal obesity
Cardiovascular
Miscellaneous: osteonecrosis, peptic ulcer disease, cataracts, dermatologic abnormalities, steroid myopathy, psychosis, growth retardation, altered mood and behavior
AZATHIOPRINE
Serious and opportunistic infections
Hematologic abnormalities: leukopenia, thrombocytopenia, anemia
Gastrointestinal: nausea, vomiting, rare hepatitis
Reproductive: pregnancy class D
Miscellaneous: pancreatitis, interstitial pneumonitis, rashes
MYCOPHENOLATE MOFETIL
Serious and opportunistic infections
Leukopenia
Gastrointestinal: diarrhea, nausea, dyspepsia, elevated transaminases
Reproductive: pregnancy class C
CYCLOSPORINE
Serious and opportunistic infections
Renal disease and hypertension
Potential for increased malignant neoplasms
Reproductive: pregnancy class C
Miscellaneous: hirsutism, gingival hyperplasia, hyperuricemia, electrolyte abnormalities
TACROLIMUS
Serious and opportunistic infections
Renal disease and hypertension, perhaps lower than with cyclosporine
Potential for increased malignant neoplasms
Post-transplantation diabetes mellitus
Neurotoxicity: tremor, headaches, motor function abnormalities, mental status alteration, sensory changes
Reproductive: pregnancy class C
Miscellaneous: hirsutism, gingival hyperplasia, myocardial hypertrophy
SIROLIMUS
Serious and opportunistic infections
Renal disease and hypertension
Potential for increased malignant neoplasms
Reproductive: pregnancy class C
Miscellaneous: hyperlipidemia, pneumonitis, interstitial lung disease
CYCLOPHOSPHAMIDE
Serious and opportunistic infections
Increased incidence of malignant neoplasms
Hematologic toxicity: leukopenia, thrombocytopenia, anemia
Reproductive: pregnancy class D, premature ovarian failure, oligospermia, fetal abnormalities
Urologic: hemorrhagic cystitis, bladder cancer
Miscellaneous: nausea, vomiting, diarrhea, pulmonary fibrosis

*This list highlights the most serious and common adverse events but does not include all reported adverse events with these agents.

hematologic toxicity but does not substitute for monitoring of the CBC. Gastrointestinal toxicity is usually minor, but patients may have significant symptomatic complaints of nausea, vomiting, diarrhea, and epigastric pain that are often self-limited and reversible. A severe hepatic toxicity has been rarely reported, leading to a recommendation for regular monitoring of serum transaminases, alkaline phosphatase, and bilirubin, particularly during the first 6 months of therapy. Rare complications of azathioprine treatment include fever, arthralgia, rash, pancreatitis, and interstitial pneumonitis.

Azathioprine use in pregnancy is classified category D and has been associated with fetal abnormalities in animals. The use of azathioprine should be avoided, if possible, in pregnant women and nursing mothers, although patients with organ transplantation and autoimmune disease have had successful pregnancies while receiving azathioprine.

The association of azathioprine treatment with the development of malignant neoplasms has been controversial. In most clinical situations, azathioprine is used in conditions and in combination or temporal sequence with other drugs that are associated with the development of malignant diseases. For example, patients with systemic vasculitis or SLE may initially be treated with cyclophosphamide and then are often subsequently treated with azathioprine. An increase in malignant neoplasms in this population could be related to the concurrent azathioprine treatment, but the prior cyclophosphamide therapy may be a higher risk factor. In addition, patients with solid organ transplantation, another circumstance in which azathioprine is frequently employed, appear to have a higher rate of malignant neoplasms separate from the use of immunosuppressive drugs. Extensive efforts to identify an independent increased risk for malignant neoplasms with azathioprine have not produced consistent results. These results suggest that the risk for malignant transformation with azathioprine is very low if this risk exists at all.

Azathioprine is an important and effective therapy in organ transplantation and autoimmune disease. Often this drug is used in combination with other agents and as a corticosteroid-sparing drug. Surveillance for infection and monitoring for hematologic toxicity are important. Concurrent treatment with allopurinol should be avoided to prevent serious toxicity from an interaction of the two drugs.

Mycophenolate Mofetil

Mechanism of Action

Mycophenolate mofetil is a prodrug that is converted in vivo to the active compound mycophenolic acid. Mycophenolic acid acts through inhibition of the enzyme inosine monophosphate dehydrogenase, resulting in an increase in 6-thionosinic acid that is normally metabolized by this enzyme. The accumulation of 6-thionosinic acid acts through a negative feedback loop to suppress the de novo synthesis of purines and associated DNA production. Although it is not a cytotoxic agent, the actions of mycophenolic acid are most pronounced on proliferating cells, such as lymphocytes, to reduce cell division and associated functions of these critical cells in the immune response.

Indications

Mycophenolate mofetil is approved by the FDA for the prevention of allograft rejection in renal, hepatic, and cardiac transplantation. In addition to these approved indications, mycophenolate mofetil has been evaluated in SLE, for which its principal use has been in patients with lupus nephritis, although the drug has been used for other manifestations. In the treatment of SLE, it can be used successfully in patients who no longer respond to azathioprine[7] and can be used as initial and maintenance therapy for lupus nephritis. Limited use has been reported in other autoimmune diseases such as rheumatoid arthritis, vasculitis, and polymyositis.

Adverse Effects and Monitoring

Common adverse events with mycophenolate mofetil include hematologic and gastrointestinal complications. Leukopenia is reported in 20 to 35% of patients receiving this drug for organ transplantation; however, severe neutropenia is seen in only 2 to 3% of subjects. Patients receiving mycophenolate mofetil have a higher susceptibility to infections, similar to that seen with other immunosuppressive agents, including opportunistic infections. Diarrhea, nausea, and dyspepsia are frequent. Abnormalities of liver enzymes are commonly noted and appear to be dose dependent. Rare complications include pulmonary fibrosis and malignant neoplasms. Mycophenolate mofetil is classified as category C by the FDA for use in pregnant women. Although the agent should be avoided in patients who are pregnant or not practicing adequate contraception, this drug appears to have less impact on the reproductive system than does cyclophosphamide. Patients with SLE who are concerned about a potential severe impact of cyclophosphamide on reproductive organs might elect to use mycophenolate mofetil instead.

Monitoring of patients receiving mycophenolate mofetil should include a monthly CBC and hepatic enzyme activities. Clinical monitoring for gastrointestinal and infectious complications should be conducted during regular clinical follow-up.

Mycophenolate mofetil is an important and effective agent in the management of transplantation patients. Ongoing work is studying the role of this drug in other autoimmune diseases. It is hoped that mycophenolate mofetil can provide effective therapy for rheumatic and autoimmune disorders with less toxicity than is seen with currently available agents, with a particular potential for use in lupus nephritis.

Immunophilin-Binding Agents

The development of immunophilin inhibitors has significantly advanced organ transplantation.[8] Each of these drugs—cyclosporine, tacrolimus (also known as FK506), and sirolimus (also known as rapamycin)—has significant

immunosuppressive activity on T-cell-mediated functions. Whereas the mechanisms of action for each drug differ, they all have the common action of binding to a cytosolic protein. This binding results in a decrease in T-cell cytokine production and T-cell proliferation. These separate sites of action allow the use of these agents in combination in transplantation management. In addition, the differences in action and specific binding for each drug have resulted in unique adverse events profiles.

Cyclosporine

Mechanism of Action

Cyclosporine acts by binding to the cytosolic protein cyclophilin to form a cyclosporine-cyclophilin complex. The cyclosporine-cyclophilin complex inhibits the enzyme calcineurin. Calcineurin is an enzyme involved in multiple T-cell functions and is of particular importance in enhancing the transcription of genes for pro-inflammatory cytokines. The use of cyclosporine inhibits the production of IL-2 with a resulting decrease in T-cell activation. Cyclosporine also inhibits the production of other cytokines, including IL-3, IL-4, granulocyte-macrophage colony-simulating factor, tumor necrosis factor-α, and interferon-α. The overall impact of these actions is to reduce immune function and inflammation.

Indications

The use of cyclosporine and other calcineurin inhibitors has revolutionized solid organ transplantation. The specific FDA-approved indications include renal, liver, and heart transplantation. Most often the drug is used in conjunction with other immunosuppressive agents including corticosteroids and azathioprine. The critical clinical challenge is to balance the potent immunosuppressive effects of the drug against the adverse events produced by this agent, with particular attention to avoidance of infectious complications and monitoring for hypertension and renal toxicity. Because of these critical issues, prescribing information specifically limits this agent to physicians experienced with the use of immunosuppressive agents.

Oral cyclosporine is also approved for the treatment of rheumatoid arthritis as a disease-modifying antirheumatic drug (DMARD), either alone or in combination with methotrexate. Although it is effective in the treatment of rheumatoid arthritis and other rheumatic diseases, cyclosporine does not provide a substantially greater efficacy than the other DMARDs. Cyclosporine is also approved for the treatment of psoriasis. Because of the significant adverse event profile, cyclosporine is generally reserved for patients with autoimmune diseases whose therapy with more traditional and less toxic agents has failed.

Topical cyclosporine is effective and approved for treatment to increase tear production presumed to be suppressed secondary to inflammation in patients with keratoconjunctivitis sicca syndrome (Chapter 423). The topical treatment is associated with a much lower frequency of adverse drug events than is systemic cyclosporine therapy.

Adverse Effects

Close monitoring is required during cyclosporine therapy. Blood levels can be measured, which is useful in ensuring that the drug remains within a therapeutic range and below levels associated with increased toxicity. Because of multiple potential drug interactions that can both raise and lower cyclosporine levels, as well as food interactions, particularly increased levels with grapefruit and grapefruit juices, patients should be constantly monitored and blood levels obtained when indicated. These evaluations should ensure that when medical therapy is added, changed, or deleted, these adjustments will not have an impact on the effects of the cyclosporine, with its associated therapeutic and toxicity issues.

Infection

All infections have an increased potential for developing in patients receiving cyclosporine and other calcineurin inhibitors as with other immunosuppressive agents. Opportunistic infections associated with impaired cell-mediated immunity are particularly increased.

Renal Disease and Hypertension

Renal disease and hypertension are common adverse events during cyclosporine therapy and are increased in prevalence with increases in dose and duration of therapy. Close monitoring of blood pressure and serum creatinine concentration is critical during treatment with cyclosporine. In many cases, these conditions are reversible if they are detected early and appropriate dose adjustments are implemented. In many patients, a mild increase in serum creatinine concentration may be tolerated if the level remains stable. Whereas renal abnormalities and hypertension are commonly identified during treatment of patients with cyclosporine, most patients will not require discontinuation of the drug if adjustments in dose or other interventions are undertaken to avoid these complications.

Malignant Neoplasia

Malignant neoplasms, particularly lymphomas, have been noted to be more common in patients receiving solid organ transplantation with associated immunosuppressive therapy. However, the exact cause of these malignant neoplasms has not been determined. In vitro mutagenesis assays with cyclosporine have been negative. In vivo animal studies have yielded equivocal results, with some data suggesting a possible increased rate of malignant neoplasms in rats and mice.

Issues with Reproduction

Cyclosporine is a pregnancy class C drug. Data in animals have demonstrated toxicity to both the embryo and fetus. Patients should practice effective contraception while receiving this drug. Data from pregnant transplant patients are difficult to interpret because cyclosporine is generally not the only medical therapy received by these women and the impact of the disease associated with the organ transplantation is at times difficult to separate from the impact of therapy. Despite these limitations, normal pregnancies and early childhood development have been reported in many women receiving cyclosporine. Premature birth and low birth weight are more common in women receiving cyclosporine. Therefore, although the use of cyclosporine during pregnancy and breast-feeding should be avoided if at all possible, in patients who become pregnant, an assessment should determine whether the immunosuppressive therapy should be continued. In some cases, the risk for organ rejection with discontinuation of cyclosporine therapy may exceed the risk for exposure to the fetus during pregnancy.

Other Adverse Effects

Cyclosporine has also been associated with the development of hirsutism, gingival hyperplasia, hyperuricemia, and electrolyte abnormalities.

Tacrolimus

Mechanism of Action

Whereas cyclosporine binds to cyclophilin, tacrolimus binds to a different protein, the 12-kD FK-binding protein (FKBP12). The binding of FKBP12 and tacrolimus forms a complex that inhibits calcineurin in a fashion similar to cyclosporine. Through this mechanism, tacrolimus has similar inhibitory effects on T-cell function and cytokine production.

Indications

Tacrolimus is approved for the prophylactic treatment of organ rejection after kidney and liver transplantation. The efficacy and safety in rheumatoid arthritis is limited but encouraging. Experience with this agent is less extensive than the experience with cyclosporine.

Adverse Effects

The major adverse events with tacrolimus are similar to those of cyclosporine and include increased susceptibility to infection, renal disease, and hypertension. An increase in malignant neoplasia is also reported with a pattern similar to that reported in patients receiving cyclosporine.

Whereas many adverse events with tacrolimus are similar to those of other calcineurin inhibitors, some specific complications have been reported, including the development of post-transplantation diabetes mellitus. This adverse event was seen in 20% of subjects in phase III clinical trials of tacrolimus with an onset generally within the first 3 months of therapy. In many patients, post-transplantation diabetes mellitus will resolve after the drug is discontinued. Neurotoxicity is also reported, including tremor, headache, motor function abnormalities, mental status alterations, and sensory changes. Myocardial hypertrophy has been reported. Tacrolimus, like cyclosporine, is pregnancy class C. Patients receiving this drug should practice effective contraception.

Sirolimus

Mechanism of Action

Sirolimus (or rapamycin) in not a calcineurin inhibitor but has many actions and mechanisms similar to calcineurin inhibitors. The mechanism of action of sirolimus involves binding to FKBP12, the binding protein for tacrolimus, but the effect of this binding is different. Instead of acting on calcineurin, the sirolimus-FKBP12 complex binds to another protein, the mammalian target

of rapamycin (mTOR), which is a key regulatory kinase. The inhibition of mTOR results in significant immunosuppression by decreasing T-cell proliferation and the progression from G_1 phase to S phase in the cell cycle. Because sirolimus works through a mechanism different from the mechanism of action for cyclosporine, the two drugs have been studied in combination. In renal transplantation patients, the use of these two agents in combination has a greater immunosuppressive effect than that of cyclosporine alone.

Indications

Sirolimus is approved for the prophylaxis of organ rejection in renal transplantation in combination with cyclosporine and corticosteroids for patients older than 12 years. Data with the use of sirolimus in other populations, including subjects with autoimmune diseases, are limited and generally are derived only from animal models.

Adverse Effects

The adverse event profile for infection and malignant neoplasms with sirolimus is similar to that with calcineurin inhibitors. Renal disease with cyclosporine and sirolimus in combination has been reported more frequently than in patients receiving cyclosporine alone. Adverse reactions specific for sirolimus include hyperlipidemia, interstitial lung disease, and the syndrome of calcineurin-induced hemolytic-uremic syndrome, thrombotic thrombocytopenic purpura, and thrombotic microangiopathy.

Alkylating Agents

Alkylating agents are an important component of immunosuppressive therapy in autoimmune diseases. The use of these agents is limited by their associated toxicity, particularly the potential for development of malignant neoplasia, reproductive toxicity, and increased incidence of infection. Patients considered candidates for alkylating therapy should be fully informed of the potential risks and benefits of these drugs and concur with the decision for their use. Regular monitoring of blood counts and urinalysis are important in observing patients receiving these drugs.

Cyclophosphamide

Cyclophosphamide has been used primarily as a cytotoxic drug for the treatment of malignant neoplasms; it is currently approved for this indication as well as for biopsy-proven minimal change nephrotic disease in children. Several severe autoimmune diseases have also been shown to be responsive to cyclophosphamide, often given in conjunction with initial high-dose corticosteroid therapy. Because of the significant toxicities associated with these drugs, a benefit-to-risk assessment and discussion should be undertaken with each patient as cyclophosphamide use is being considered. In most cases, the diseases warranting cyclophosphamide therapy will be life-threatening conditions with poor prognosis to justify the potential severe toxicity of this agent.

Mechanism of Action

Cyclophosphamide is an inactive compound that can be administered either orally or intravenously and is metabolized to the active drug by the cytochrome P-450 mixed function oxidase system. This process produces the active compounds phosphoramide mustard and the toxic metabolite acrolein. The cytotoxic effect of this drug results from the alkylation of various cellular constituents, especially nucleic acids. Changes in immune function with cyclophosphamide include depletion of lymphoid tissues, with decreases in both B and T cells, suppression of cellular immune function, and decreased antibody production.

Indications

Treatment of lupus nephritis and severe systemic vasculitis is the most well studied and common use of cyclophosphamide in the rheumatic diseases. Controlled clinical trials have demonstrated improved outcomes in these conditions, particularly with reduced progression to end-stage renal disease in patients with SLE and improved survival with systemic vasculitis such as granulomatosis with polyangiitis (GPA, formerly known as Wegener granulomatosis) and other forms of antineutrophil cytoplasmic autoantibody-associated vasculitis (AAV). However, there has been an increase use of rituximab (Chapter 36) in the treatment of AAV in place of cyclophosphamide with similar clinical benefit and less adverse events.[A2] The evaluation of cyclophosphamide in patients with SLE has demonstrated the benefit of intermittent, usually monthly, intravenous pulse therapy as a method to avoid some of the most severe toxic side effects while maintaining therapeutic efficacy. However, pulse intravenous cyclophosphamide has not been found to be as effective as continuous oral cyclophosphamide in all conditions, particularly GPA. Other conditions, such as polyarteritis nodosa, Takayasu arteritis, and Churg-Strauss syndrome, have been reported to benefit from cyclophosphamide; however, the low prevalence of these diseases has prohibited the conduct of controlled clinical trials to prove efficacy. Cyclophosphamide has also been demonstrated to be effective in the treatment of patients with rheumatoid arthritis, but it is generally not used in this condition because of the associated toxicity and availability of other effective agents. Although the use of cyclophosphamide has been a tremendous advance in the treatment of these life-threatening diseases, the use of less toxic medications, such as rituximab and mycophenolate mofetil, is under investigation to determine their relative clinical efficacy.

Adverse Effects

Before the initiation of cyclophosphamide therapy, a frank discussion with the patient about potential serious and even life-threatening adverse events should be undertaken and documented.

Malignant Neoplasia

Malignant neoplasms may develop in patients receiving cyclophosphamide for the treatment of both malignant and nonmalignant diseases. The risk for malignant change appears to increase with the duration and dose of cyclophosphamide. The most common malignant neoplasms are bladder, myeloproliferative, and lymphoproliferative disorders. The use of intravenous pulse cyclophosphamide may reduce but not eliminate the risk for bladder cancer. These malignant neoplasms may develop years after the discontinuation of the drug.

Reproductive Issues

Cyclophosphamide can be teratogenic, affect female reproduction, and reduce male fertility. Cyclophosphamide is pregnancy category D and should not be used in pregnant women unless life-threatening disease is present that warrants this treatment. Whereas successful pregnancies have been reported in patients receiving cyclophosphamide during pregnancy, fetal abnormalities are well documented secondary to chromosome damage in patients receiving cyclophosphamide. The use of cyclophosphamide in premenopausal women can induce premature ovarian failure. The use of gonadotropin-releasing hormone analogues during intravenous pulse cyclophosphamide therapy may reduce premature ovarian failure in patients with SLE; however, the risk of this complication is still present. In males, temporary and permanent decrease in sperm count may occur with cyclophosphamide. Because the recovery of fertility after cyclophosphamide is variable, sperm banking should be considered before therapy is begun.

Other Adverse Effects

As with other immunosuppressive agents, infections are more frequent and potentially more serious in patients receiving cyclophosphamide. Opportunistic infections are more likely to be seen in these subjects. Hematologic abnormalities can involve all cell lines. The CBC should be monitored regularly and cyclophosphamide discontinued or the dose reduced when cytopenias develop. Cyclophosphamide treatment may be complicated by hemorrhagic cystitis and bladder cancer, which are probably related to the toxic metabolite acrolein. Efforts to reduce the potential for hemorrhagic cystitis and bladder malignant neoplasms include the use of intravenous pulse therapy, hydration, frequent voiding, and treatment with agents containing sulfhydryl groups to scavenge acrolein. Regular urinalysis for blood is indicated to monitor for bladder toxicity. Patients may develop severe nausea, vomiting, and diarrhea with cyclophosphamide therapy.

Chlorambucil

Chlorambucil has not been evaluated as extensively as cyclophosphamide, but this agent appears to have properties and an adverse event profile similar to those of cyclophosphamide. Like cyclophosphamide, chlorambucil is associated with the development of malignant neoplasms and should be avoided during pregnancy.

Miscellaneous Agents

Although principally prescribed for other indications, methotrexate and leflunomide have been evaluated and used for their potential immunosuppressive effects. Leflunomide blocks the enzyme dihydroorotate dehydrogenase, resulting in an inhibition of pyrimidine synthesis and a reduction in

T-cell activation. Methotrexate has multiple mechanisms of action, but its major effect in autoimmune diseases is mediated by inhibition of the enzyme aminoimidazole-4-carboxamide ribonucleotide transformylase. This inhibition affects purine synthesis, increasing the intracellular concentration of aminoimidazole-4-carboxamide ribonucleotide, which stimulates the release of adenosine, a potent anti-inflammatory compound. Methotrexate has principally been advocated for its corticosteroid-sparing effects. Both methotrexate and leflunomide are effective DMARDs in the treatment of rheumatoid arthritis and are discussed in other chapters.

Grade A References

A1. Maneiro JR, Lopez-Canoa N, Salgado E, et al. Maintenance therapy of lupus nephritis with mycophenolate or azathioprine: systematic review and meta-analysis. *Rheumatology (Oxford)*. 2014; 53:834-838.

A2. Specks U, Merkel PA, Seo P, et al. Efficacy of remission-induction regimens for ANCA-associated vasculitis. *N Engl J Med*. 2013;369:417-427.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Rousseau GG. Fifty years ago: the quest for steroid hormone receptors. *Mol Cell Endocrinol*. 2013;375:10-13.
2. Mani SK, Mermelstein PG, Tetel MJ, Anesetti G. Convergence of multiple mechanisms of steroid hormone action. *Horm Metab Res*. 2012;44:569-576.
3. Lee SR, Kim HK, Song IS, et al. Glucocorticoids and their receptors: insights into specific roles in mitochondria. *Prog Biophys Mol Biol*. 2013;112:44-54.
4. Coutinho AE, Chapman KE. The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights. *Mol Cell Endocrinol*. 2011;15:2-13.
5. Buttgereit F, Gibofsky A. Delayed-release prednisone: a new approach to an old therapy. *Expert Opin Pharmacother*. 2013;14:1097-1106.
6. Grossman JM, Gordon R, Ranganath VK, et al. American College of Rheumatology 2010 recommendations for the prevention and treatment of glucocorticoid-induced osteoporosis. *Arthritis Care Res*. 2010;62:1515-1526.
7. Al Maimouni H, Gladman DD, Ibanez D, et al. Switching treatment between mycophenolate mofetil and azathioprine in lupus patients: indications and outcomes. *Arthritis Care Res (Hoboken)*. 2014;66:1905-1909.
8. Halleck F, Friedersdorff F, Fuller TF, et al. New perspectives of immunosuppression. *Transplant Proc*. 2013;45:1224-1231.

36

BIOLOGIC AGENTS AND SIGNALING INHIBITORS

CEM GABAY

Biologic agents are a new class of therapeutic agents that target different mediators involved in the pathogenesis of human diseases. The development of these therapies has markedly improved the management of many diseases. In addition, their use has greatly increased our understanding regarding the pathophysiology of these diseases. One of the most compelling examples is the efficacy of tumor necrosis factor (TNF)-α inhibitors in rheumatoid arthritis and in inflammatory bowel diseases. In addition, new drugs targeting specifically intracellular enzymatic signaling pathways (kinases) have been developed. Most of these signaling inhibitors have been developed for the treatment of cancer. Recently, a kinase inhibitor was approved for the treatment of rheumatoid arthritis.

The United States Adopted Names Council has established a common classification of these different therapies (Table 36-1). The pharmaceutical company usually gives the prefix of the name, and the suffix defines whether this is a monoclonal antibody (mab), a soluble receptor (cept), or a kinase inhibitor (inib). Monoclonal antibodies, by far the largest group of biologic agents today, include also in their name the type of target (immune system, cancer, cardiovascular, system, and bone) as well as their origin (chimeric, humanized, and human).

TABLE 36-1 NOMENCLATURE OF MONOCLONAL ANTIBODIES

TARGET	ABBREVIATION	EXAMPLE
Immunology	li/l	ada**li**mumab
Cardiovascular	ci/c	beva**ci**zumab
Bone	os/s	den**os**umab
Interleukin	ki/k	cana**ki**numab
Miscellaneous tumors	tu/t	ofa**tu**mumab
SOURCE		
Chimeric	xi	ritu**xi**mab
Humanized	zu	trastu**zu**mab
Human	u	golim**u**mab

From the United States Adopted Names Council website (www.ama-assn.org).

To make mouse- or other animal-derived monoclonal antibodies less immunogenic, chemical and genetic engineering methods can replace two thirds of the mouse antibody with human antibody, creating "chimeric" monoclonal antibodies, or can replace 90 to 95% with human antibody to create "humanized" monoclonal antibodies. Chimeric and humanized monoclonal antibodies, being less immunogenic, elicit the production of less neutralizing human antimouse antibodies in the recipient present.

All biologic agents are large molecules that are administered by intravenous or subcutaneous routes, whereas kinase inhibitors are small chemicals that are prescribed as oral drugs. For safety reasons, combination of biologic agents or kinase inhibitors is not recommended. Rather than providing an exhaustive review of all tested approaches, this chapter aims to review the currently approved treatments.

TUMOR NECROSIS FACTOR-α INHIBITORS

The cytokine TNF-α binds to two different receptors, TNF-R55 and TNF-R75, and exerts important functions in the control of host responses against infections. However, uncontrolled TNF-α production may lead to chronic inflammation and subsequent tissue damage.

Types of Tumor Necrosis Factor-α Inhibitors

Different agents have been developed to inhibit the biologic activity of TNF-α, including monoclonal antibodies and soluble receptors. Monoclonal antibodies include infliximab (Remicade), adalimumab (Humira), and golimumab (Simponi). Infliximab is a chimeric antibody, whereas the two others are fully human antibodies. Certolizumab-pegol (Cimzia) is a pegylated humanized anti-TNF-α antibody Fab′ fragment. Etanercept (Enbrel) is a fusion protein that contains the extracellular portion of TNF-R75 coupled to the Fc domain of human immunoglobulin G1 (IgG1). All these agents bind to TNF-α and block its biologic activity. Etanercept binds also to lymphotoxin-α (previously termed TNF-β). Infliximab has been shown to exert cytotoxic effects on macrophages and T lymphocytes expressing TNF-α on their surface.

Indications

TNF-α antagonists are approved for the treatment of rheumatoid arthritis (Chapter 264). They exert marked anti-inflammatory effects and prevent the progression of structural joint damage. TNF-α inhibitors are efficacious in early rheumatoid arthritis and in long-standing disease refractory to conventional disease-modifying antirheumatic drugs (DMARDs) such as methotrexate. TNF-α inhibitors are more efficacious when administered in combination with methotrexate than alone. TNF-α inhibitors are also approved for the treatment of ankylosing spondylitis refractory to nonsteroidal anti-inflammatory drugs and psoriatic arthritis refractory to DMARDs (Chapter 265). Anti-TNF-α antibodies, but not etanercept, have proven efficacy in severe Crohn disease (Chapter 141) by reducing the disease activity score, inducing closure of draining fistulas, and allowing a decrease in the dose of chronic glucocorticoid medication. Infliximab, adalimumab, and golimumab are approved for the treatment of ulcerative colitis refractory to conventional therapy. Infliximab, adalimumab, and etanercept are also approved for the treatment of chronic plaque psoriasis (Chapter 438) refractory to phototherapy or systemic therapy.

Adverse Effects

TNF-α inhibitors can be associated with allergic reactions. Postmarketing data have shown that TNF- α inhibitors are associated with an increased risk for infections, including all types of bacterial and opportunistic infections. In particular, the use of TNF-α inhibitors has been associated with an increased risk for reactivation of latent tuberculosis. The results of cohort studies do not support an overall increased risk for cancer. Rare cases of demyelinating disorders, lupus-like manifestations, and cytopenia have been reported.

INTERLEUKIN-1 INHIBITORS

IL-1 (both IL-1α and IL-1β) binds to IL-1 receptors (type I IL-1R and IL-1R accessory protein) to induce a vast array of inflammatory signals. IL-1 receptor antagonist (IL-1Ra) competitively inhibits the interaction of IL-1 with its receptors.[1]

Types of Interleukin-1 Inhibitors

Anakinra (Kineret), recombinant human IL-1Ra, was the first IL-1 inhibitor used in clinical trials. Rilonacept (Arcalyst) is a fusion protein including the

IL-1 binding motifs of IL-1 receptors coupled to the Fc domain of human IgG1. Canakinumab (Ilaris) is a fully human monoclonal antibody against IL-1β.

Indications

Anakinra is approved for the treatment of rheumatoid arthritis refractory to conventional DMARDs but exhibits relatively modest efficacy. Anakinra, canakinumab, and rilonacept are approved for the treatment of cryopyrin-associated periodic syndromes, a set of hereditary systemic autoinflammatory diseases associated with *NLRP3* gene (encoding for cryopyrin) mutations (Chapter 261) and characterized by enhanced IL-1β release. Canakinumab is approved for the treatment of systemic-onset juvenile idiopathic arthritis in children aged 2 years and older. Anakinra is also effective in adult-onset Still disease and other autoinflammatory conditions. Clinical trials have reported encouraging results with IL-1 antagonists in crystal-induced arthritis such as gout and chondrocalcinosis. Promising results have also been reported from the use of gevokizumab, another fully human anti-IL-1β monoclonal antibody in Behçet uveitis.

Adverse Effects

Anakinra, canakinumab, and rilonacept have frequently been associated with injection site reactions and upper airway infections. A modest increase in serious infections has been reported with the three IL-1 antagonists. The combination of anakinra and etanercept (and probably also other TNF-α antagonists) increases the risk for infections and is not recommended.

INTERLEUKIN-6 INHIBITORS

IL-6 is a pro-inflammatory cytokine that binds to a heterodimeric receptor, including IL-6Rα and gp130. Tocilizumab (Actemra) is a humanized monoclonal antibody against IL-6Rα that inhibits its interaction with IL-6.

Indications

Tocilizumab is approved for the treatment of rheumatoid arthritis refractory to conventional DMARDs and TNF-α antagonists.[2] In clinical trials, tocilizumab prevented the progression of radiographic damage. Tocilizumab in monotherapy seems as effective as in combination with methotrexate and is superior to adalimumab (TNF-α inhibitor) in monotherapy. Tocilizumab is also approved for the treatment of systemic-onset juvenile idiopathic arthritis and polyarticular juvenile idiopathic arthritis in patients aged 2 years and older. In Japan, tocilizumab is approved for the management of Castelman disease. Tocilizumab is also effective for adult-onset Still disease. Clinical trials are in progress in other inflammatory rheumatic diseases. In addition, clinical trials examining the efficacy and safety of other monoclonal antibodies against IL-6R (sarilumab) or IL-6 (sirukumab, clazakizumab) in rheumatoid arthritis are in progress.[3]

Adverse Effects

Tocilizumab has been associated with an increased risk for serious infections. A transient increase in transaminase levels has been reported. Hypercholesterolemia and cytopenia may appear in some patients.

ANTIBODY AGAINST INTERLEUKIN-12 AND INTERLEUKIN-23

IL-12 and IL-23 are heterodimeric cytokines, including a common p40 subunit and a specific subunit—p35 for IL-12 and p19 for IL-23. IL-23 participates in the differentiation of T_H17 cells that produce IL-17. Experimental findings indicated that IL-23 and IL-17 are critical for the development of autoimmune pathologies. Ustekinumab (Stelara) is a human monoclonal antibody against p40 that targets both IL-12 and IL-23 and blocks their biologic activities.

Indications

Ustekinumab is approved for the treatment of psoriasis patients who have had prior exposure to phototherapy or use of other systemic therapies. In clinical trials, ustekinumab was associated with successful treatment of Crohn disease. In addition, ustekinumab alone or in combination with methotrexate is also approved for the treatment of active psoriatic arthritis. Anti-IL-17 and anti-IL-17R antibodies are in clinical trials for the treatment of psoriasis, psoriatic arthritis, ankylosing spondylitis, and rheumatoid arthritis.[4]

Side Effects

Ustekinumab has been associated with injection site reaction and allergic adverse events. Upper respiratory airway infections have been reported.

JANUS KINASE INHIBITOR

The janus kinase (JAK)-signal transducer and activator of transcription (STAT) signaling pathway transmits information from various receptors at the cell surface, including those of cytokines, interferons, and growth factors, to the DNA promoters in the nuclei, thus stimulating cell activation and gene expression. The JAK family comprises several members, such as JAK1, JAK2, JAK3, and Tyk2, that are associated as heterodimers or homodimers to the different receptors. Tofacitinib (Xeljanz) is a competitive inhibitor of JAK autophosphorylation, leading to the absence of STAT phosphorylation by JAK and the nuclear translocation of STAT. Tofacitinib preferentially inhibits the signaling associated with JAK3 and JAK1 activation, including the receptors involved in the immune and inflammatory responses.

Indications

Tofacitinib is approved for the oral treatment of moderate and severe rheumatoid arthritis refractory to conventional DMARDs and TNF antagonists.[5] Tofacitinib can be used alone or in combination with conventional DMARDs. Clinical trials in psoriasis, psoriatic arthritis, Crohn disease, and ulcerative colitis are ongoing. The JAK½ inhibitor ruxolitinib was recently approved for the treatment of myelofibrosis.[6]

Adverse Effects

Tofacitinib is associated with an increased rate of infectious events, including different types of bacterial infections, herpes zoster, and opportunistic infections. Increase in transaminase levels, hypercholesterolemia, and cytopenia has also been observed. Some of these adverse effects are potentially related to the inhibition of signals induced by IL-6 and hematopoietic growth factors.

INHIBITORS OF ANGIOGENESIS

Vascular endothelial growth factor (VEGF) is a family of growth factors involved in vasculogenesis and angiogenesis. VEGF-A binds to the tyrosine kinase receptors, VEGFR1 and VEGFR2. VEGF has been implicated in tumor growth and metastasis, in diabetic retinopathy, and in age-related macular retinopathy.

Types of Angiogenesis Inhibitors

Bevacizumab (Avastin) is a humanized monoclonal antibody against VEGF that acts as an angiogenesis inhibitor. Aflibercept (Zaltrap) is a fusion protein containing extracellular immunoglobulin motifs of VEGFR1 and VEGFR2 coupled to human IgG1. Several VEGF receptor–associated tyrosine kinase inhibitors have been approved for the treatment of cancer, including Sorafenib (Nexavar), Sunitinib (Sutent), Axitinib (Inlyta), Pazopanib (Votrient), Vandetanib (Caprelsa), and Sorafenib (Nexavar). Some of these tyrosine kinase inhibitors block not only VEGF receptors but also platelet-derived growth factor receptor-β, c-Kit, and Flt3 receptor.

Indications

Bevacizumab is approved for the treatment of glioblastoma (Chapter 189), and also for some patients with non–small cell lung cancer (Chapter 191), and metastatic colorectal cancer (Chapter 193). In Europe, bevacizumab is also approved for metastatic breast and ovarian cancer therapy. Intraocular injections of ranibizumab (Lucentis), an antibody fragment derived from bevacizumab, are effective in maintaining vision in most patients with wet macular degeneration (Chapter 423) and are approved for this indication. Aflibercept is approved for the treatment of metastatic colorectal cancer. Sunitinib is approved for the treatment of metastatic renal cell carcinoma (Chapter 197), gastrointestinal stromal tumor refractory to imatinib (Chapter 202), and unresectable pancreatic neuroendocrine tumors. Axitinib and pazopanib are approved for the treatment of renal cell carcinoma. Pazopanib is also approved for advanced soft tissue sarcoma. Vandetanib is approved for late-stage medullary thyroid cancer. Sorafenib is approved for advanced renal cell carcinoma, some cases of hepatocarcinoma (Chapter 196), and late-stage thyroid cancer.

Adverse Effects

Bevacizumab has been associated with a heightened risk for bleeding and gastrointestinal tract perforation. Delayed wound healing and hypertension have also been reported. Reported adverse events for kinase inhibitors include hypertension, fatigue, asthenia, diarrhea, and some abnormal laboratory tests, including transaminases, lipase, amylase, and leukocytes and platelets. Vandetanib is associated with prolonged QT, and more rarely with cases of torsades de pointes and sudden death.

INHIBITORS OF TUMOR GROWTH FACTORS

The epidermal growth factor receptor (EGFR), a member of the ErbB-1 (human epithelial growth factor receptor-1 [HER-1]) family of receptors, is a cell surface receptor for epidermal growth factor (EGF) and transforming growth factor-α. Binding of EGFR leads to tyrosine kinase activation. In tumor cells, EGFR overexpression or dysregulation due to *EGFR* gene mutations (exon 19 deletion or exon 21 [L858R] substitution) can lead to cell proliferation, apoptosis blockade, neovascularization, and tumor metastasis. HER-2 (also known as ErbB2) is a receptor with tyrosine kinase activity that is expressed at high levels in 20 to 30% of breast cancers and other types of cancers and may cause uncontrolled tumor cell proliferation.

Types of Inhibitors

Two types of strategies target the EGFR pathway: monoclonal antibodies to the extracellular domain of EGFR and orally available tyrosine kinase inhibitors. Cetuximab (Erbitux) is a chimeric monoclonal IgG1 anti-EGFR antibody that inhibits EGFR signaling and may also induce tumor cell death by antibody-dependent cellular cytotoxicity. Panitumumab (Vectibix) is a fully human monoclonal IgG2 antibody directed against EGFR. Erlotinib (Tarceva), gefitinib (Iressa), and afatinib (Gilotrif) are the three most commonly studied EGFR tyrosine kinase inhibitors. Trastuzumab (Herceptin) is a humanized monoclonal IgG1 antibody that binds to the extracellular domain of HER-2. Lapatinib (Tykerb) is a small molecule with inhibitory activity on EGFR and HER-2 tyrosine kinases.

Indications

Cetuximab is approved as second- and third-line treatment in colorectal and squamous cell carcinoma of the head and neck cancers. Panitumumab is approved for the treatment of metastatic colorectal cancer. Erlotinib and afatinib are approved for the treatment of metastatic non–small cell lung carcinoma with EGFR mutations, whereas gefitinib is approved for the same indication but with some additional limitations. Erlotinib is also approved for the treatment of pancreatic cancer. Trastuzumab and lapatinib are approved for the treatment of metastatic breast cancer expressing HER-2, and trastuzumab for metastatic stomach and gastroesophageal junction cancer expressing HER-2.

Adverse Effects

Cetuximab is sometimes associated with severe allergic reactions during infusion. Acne-like skin rash has been commonly reported. Erlotinib, gefitinib, and afatinib are associated with acne-like rash and mild gastrointestinal symptoms. Cardiotoxicity is a major problem in patients treated with trastuzumab. Approximately 10% of patients are unable to tolerate this drug because of pre-existing heart problems. The risk for cardiomyopathy is increased when combined with anthracycline (which itself is associated with cardiac toxicity). Lapatinib is in general well tolerated, but some cases of hepatotoxicity have been reported. Active research is in progress to develop more selective angiogenesis inhibitors in order to overcome acquired resistance to specific inhibitors of VEGF and VEGFR and to reduce their negative off-target effects.[7]

OTHER TYROSINE KINASE INHIBITORS

Imatinib mesylate (Gleevec or Glivec) works by binding to the adenosine triphosphate (ATP) binding site of BCR-ABL, resulting in competitive inhibition of its enzymatic activity. Imatinib inhibits also the tyrosine kinase activity of c-Kit and platelet-derived growth factor receptor.

Indications

Imatinib represents a major advance over conventional treatments for chronic myelogenous leukemia (Chapter 184), with more than 90% of patients obtaining complete hematologic response and 70 to 80% of patients achieving a complete cytogenetic response. Resistance to imatinib[8] represents a clinical challenge and is often a result of point mutations causing a conformation change in BCR-ABL, which impairs imatinib binding. Dasatinib (Sprycel) and Nilotinib (Tasigna), two other tyrosine kinase inhibitors, have been approved for the treatment of chronic myelogenous leukemia as first-line therapy or in patients who are not responsive or intolerant to imatinib. Bosutinib (Bosulif) and ponatinib (Iclusig) are second-generation tyrosine kinase inhibitors that are approved as second-line treatment for chronic myelogenous leukemia.

Imatinib is also approved for the treatment of patients with c-Kit-positive advanced gastrointestinal stromal tumor (Chapter 202). Early clinical trials also showed potential beneficial effect of imatinib in systemic mastocytosis, hypereosinophilic syndrome, and dermatofibrosarcoma protuberans.

Adverse Effects

Imatinib has been associated with the occurrence of cytopenia, edema, nausea, and rash, as well as rare cases of congestive heart failure. Cytopenia and congestive heart failure are also reported with all tyrosine kinase inhibitors. Nilotinib also carries a potential risk for severe cardiac arrhythmias. Ponatinib is associated with severe adverse events, including arterial and venous blood clots, leading to its approval only in patients refractory to all other tyrosine kinase inhibitors or in patients who carry a specific gene mutation (*T315I*).

BIOLOGIC AGENTS TARGETING T LYMPHOCYTES

T cells play a central role in immune responses and are thus important targets for therapies against graft rejection and autoimmune diseases.

Types of Agents Modulating T-Cell Activity

The agents targeting T cells include antibodies against T cells, soluble receptors inhibiting costimulation signals, and antibodies blocking T-cell migration. Therapies targeting CD3 were developed more than 30 years ago with a mouse monoclonal IgG2 antibody called OKT3. These antibodies were used in the treatment of kidney transplant rejection but were associated with limiting side effects due to the occurrence of a "cytokine release syndrome." Then, a series of non-Fc-binding humanized anti-CD3 monoclonal antibodies were developed to overcome this problem as well as the development of antimurine immunoglobulin antibodies.

Abatacept (Orencia) and belatacept (Nulojix) are fusion proteins containing the extracellular portion of cytotoxic T lymphocyte antigen-4 coupled to the Fc portion of human IgG1, and they inhibit costimulatory signals between CD28 and CD80/86. Alefacept (Amevive) is a recombinant fully human fusion protein, in which the extracellular domain of CD2 has been linked to the Fc portion of IgG1, and it inhibits costimulatory signals between CD2 and LFA-3.

Natalizumab (Tysabri), a humanized monoclonal antibody against α_4-integrin, inhibits the migration of T cells by blocking the interaction between α_4-integrin and adhesion molecules expressed on endothelial cells.

Indications

Different monoclonal anti-CD3 antibodies are used in clinical trials for the prevention of transplant rejection (Chapter 49) as well as for the treatment of acute graft-versus-host disease and type 1 diabetes mellitus.

Abatacept is approved for the treatment of patients with rheumatoid arthritis and polyarticular juvenile idiopathic arthritis aged 6 years and older refractory to conventional DMARDs and TNF-α inhibitors. Belatacept is approved for the prevention of acute kidney transplant rejection. Alefacept is approved for the treatment of chronic plaque psoriasis. Natalizumab is approved for the treatment of patients with highly active remitting-relapsing multiple sclerosis in monotherapy (Chapter 411) and for severe Crohn disease.

Adverse Effects

Abatacept is generally well tolerated, but serious infectious adverse events may occur. Belatacept has been associated with anemia, urinary tract infections, and post-transplant lymphoproliferative disorders in Epstein-Barr virus–unexposed patients. Lymphopenia may occur with alefacept, and infections, malignancies, and hepatotoxicity have been reported in clinical trials. Natalizumab can be associated with allergic reactions and hepatotoxicity. Some cases of multifocal progressive leucoencephalopathy (Chapter 370), a fatal neurologic complication of JC virus reactivation in immunosuppressed individuals, have been reported, thus limiting the use of natalizumab.

AGENTS TARGETING B LYMPHOCYTES

The B-cell lineage plays a critical role in immune responses through the production of immunoglobulins, the presentation of antigen to T cells, and production of cytokines. Therapies aiming to target B cells are used in B-cell lymphoma and in autoimmune diseases.

Mode of Action

B cells can be targeted either by the use of monoclonal antibodies depleting B cells or by inhibiting cytokines essential for their maturation and survival, such as B-cell-activating factor of the TNF family (BAFF). Rituximab

(Rituxan or MabThera), ofatumumab (Arzerra), and ocrelizumab are B-cell-depleting monoclonal antibodies by binding to CD20. Belimumab (Benlysta) is a fully human monoclonal antibody against BAFF.

Indications

Rituximab is approved for the treatment of non-Hodgkin B-cell lymphoma (Chapter 185), for rheumatoid arthritis patients with inadequate response to TNF-α inhibitors, and in antineutrophil cytoplasmic autoantibody–associated vasculitis, including granulomatosis with polyangiitis (formerly Wegener granulomatosis) and microscopic polyangiitis. Disappointing results were obtained in clinical trials with rituximab in systemic lupus erythematosus. Ofatumumab is approved for the treatment of chronic lymphocytic leukemia (Chapter 184). Ocrelizumab is currently in clinical trials in multiple sclerosis. Epratuzumab, a humanized monoclonal anti-CD22 B-cell-depleting antibody, is in clinical trials for non-Hodgkin B-cell lymphoma. Positive results were reported in systemic lupus erythematosus, suggesting that epratuzumab might be eventually approved for this indication. Belimumab is the first biologic agent approved for the treatment of systemic lupus erythematosus refractory to conventional therapy. However, phase III studies leading to the registration of belimumab in this indication did not include more severe cases with renal and central nervous system involvement. Belimumab did not show efficacy in rheumatoid arthritis, but encouraging results were reported in Sjögren syndrome. The efficacy of tabalumab, a monoclonal antibody targeting membrane-bound and soluble BAFF, is being tested in a clinical trial in systemic lupus erythematosus.

Adverse Effects

Rituximab is associated with allergic reaction during infusions, systemic reaction due to tumoral B-cell lysis, and an increased risk for infectious events. Rare cases of multifocal progressive leucoencephalopathy have been reported. Hypogammaglobulinemia may occur, particularly after several series of infusions. Belimumab is associated with allergic reaction during infusion and an increased risk for infections.

CONCLUSION

The use of biologic agents has led to major advances in the management of many severe diseases refractory to conventional therapies. These agents have also provided a unique way to confirm the role of basic mechanisms in human diseases. New developments in this field will follow different directions, including the refinement of existing strategies (using more selective inhibitors or human rather than chimeric antibodies), the extension to other indications than those for which the agents were primarily designed, the selection of novel targets, and the identification of biologic markers of response.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Gabay C, Lamacchia C, Palmer G. IL-1 pathways in inflammation and human diseases. *Nat Rev Rheumatol*. 2010;4:232-241.
2. Smolen JS, Schoels MM, Nishimoto N, et al. Consensus statement on blocking the effects of interleukin-6 and in particular by interleukin-6 receptor inhibition in rheumatoid arthritis and other inflammatory conditions. *Ann Rheum Dis*. 2013;72:482-492.
3. Scheller J, Garbers C, Rose-John S. Interleukin-6: from basic biology to selective blockade of pro-inflammatory activities. *Semin Immunol*. 2014;26:2-12.
4. Miossec P, Kolls JK. Targeting IL-17 and Th17 cells in chronic inflammation. *Nat Rev Drug Discov*. 2012;11:762-776.
5. O'Shea JJ, Laurence A, McInnes IB. Back to the future: oral targeted therapy for RA and other autoimmune diseases. *Nat Rev Rheumatol*. 2013;9:173-182.
6. Meyer SC, Levine RL. Molecular pathways: molecular basis for sensitivity and resistance to JAK kinase inhibitors. *Clin Cancer Res*. 2014;20:2051-2059.
7. Limaverde-Sousa G, Stemberg C, Ferreria CG. Antiangiogenesis beyond VEGF inhibition: a journey from antiangiogenic single-target to broad-spectrum agents. *Cancer Treat Rev*. 2014;40:548-557.
8. Marfe G, Di Stefano C. Bypass mechanisms of resistance to tyrosine kinase inhibition in chronic myelogenous leukaemia. *Curr Drug Discov Technol*. 2014;11:145-153.

REVIEW QUESTIONS

1. Which of the following TNF inhibitors has not shown efficacy in Crohn disease?

A. Infliximab
B. Adalimumab
C. Golimumab
D. Etanercept

Answer: D The monoclonal anti-TNF antibodies (A, B, and C) reduce disease activity and requirement for glucocorticoid therapy in Crohn disease. Etanercept is a fusion protein containing the extracellular portion of TNF-R75 coupled to the Fc domain of IgG1. In addition to binding TNF, it also binds lymphotoxin-α. For reasons that are not clarified, etanercept has not been efficacious in inflammatory bowel disease.

2. Previous or current infection with which one of the following microbial organisms should be assessed before initiating anti-TNF therapy?:

A. Meningococcus
B. Mycobacterium tuberculosis
C. Mycoplasma
D. Histoplasmosis

Answer: B Anti-TNF therapy has been associated with increased risk for reactivation of latent tuberculosis. Patients should be evaluated and appropriately treated for presence of tuberculosis infection before initiation of anti-TNF therapy. Although histoplasmosis has occasionally occurred in patients treated with anti-TNF agents, testing for infection with this organism in not routinely performed before initiating therapy.

3. A monoclonal antibody that inhibits the biologic activities of IL-12 and IL-23 has shown efficacy in which of the following diseases?

A. Psoriasis
B. Crohn disease
C. Systemic lupus erythematosus (SLE)
D. A and B
E. A, B, and C

Answer: D Ustekinumab has shown positive efficacy in trials in patients with psoriasis or Crohn disease. Although this agent might have potential use in SLE, data from clinical trials are not available.

4. Imatinib mesylate, a kinase inhibitor that has changed the prognosis of patients with chronic myelogenous leukemia, is also approved for which of the following diseases?

A. Scleroderma
B. Idiopathic pulmonary fibrosis
C. C-Kit-positive advanced gastrointestinal tumor
D. Rheumatoid arthritis
E. Inflammatory bowel disease

Answer: C Imatinib mesylate (Gleevec) has antifibrotic activity in some murine models, providing rationale for testing this agent in diseases such as scleroderma or idiopathic pulmonary fibrosis. However, randomized controlled clinical trials of this agent have not been performed in those diseases at this time.

5. Blockade of costimulatory signals between CD28 and CD80/86 has led to approval of abatacept in which of the following diseases?

A. Rheumatoid arthritis
B. Chronic myelogenous leukemia
C. Lupus nephritis
D. Crohn disease

Answer: A Abatacept (Orencia) has shown efficacy in rheumatoid arthritis, as well as a related disease in children, polyarticular juvenile idiopathic arthritis. Despite a strong rationale for costimulatory blockade in SLE, trials of this agent were not effective in lupus nephritis nor in those with non-nephritic manifestations. Crohn disease can be treated with an agent that blocks migration of T cells, natalizumab. Chronic myelogenous leukemia has a different molecular pathophysiology and is effectively targeted by specific tyrosine kinase inhibitors.

37

PROSTANOIDS, ASPIRIN, AND RELATED COMPOUNDS

CARLO PATRONO

Arachidonic acid, or 5,8,11,14-eicosatetraenoic acid, is a 20-carbon polyunsaturated fatty acid esterified in the phospholipid domain of cell membranes. In response to chemical, physical, and hormonal stimuli, arachidonic acid is released from the glycerol backbone sn2 position by the action of various phospholipases A_2 and can be subjected to rapid enzymatic conversion to a series of oxygenated derivatives collectively called eicosanoids. The enzymes catalyzing various structural modifications of free arachidonic acid include prostaglandin H synthases, commonly known as cyclooxygenases (COX), lipoxygenases, and cytochrome P-450 isozymes. The resulting eicosanoids include prostanoids (prostaglandins and thromboxane A_2), leukotrienes, lipoxins, and epoxylins. Esterified arachidonic acid can also be subjected to in situ peroxidation, catalyzed by oxygen radicals, to form a series of corresponding isomers called isoeicosanoids.

Eicosanoids are not stored but are produced in response to diverse stimuli, with a pattern that reflects the cell-specific distribution of arachidonic acid–metabolizing enzymes and downstream isomerases and synthases. Eicosanoids are not circulating hormones but rather ubiquitous autacoids that modulate the intensity and duration of many important cellular responses in an autocrine (acting on the same cells that produced them) or paracrine (acting on nearby cells) fashion.

BIOSYNTHESIS AND ACTION OF PROSTANOIDS

Prostanoids include prostaglandins D_2, E_2, $F_{2\alpha}$, and I_2 (prostacyclin) and thromboxane A_2. They are formed through the sequential actions of phospholipase A_2 to release arachidonic acid from membrane phospholipids, prostaglandin H synthase to catalyze the cyclooxygenation of arachidonic acid to form the unstable intermediate prostaglandin G_2 and its reduction to prostaglandin H_2, and specific isomerases and synthases to catalyze the conversion of prostaglandin H_2 to different prostanoids (Fig. 37-1). Once formed, prostanoids interact with specific G protein–coupled receptors to evoke a variety of cellular responses, depending on the site of their biosynthesis (Fig. 37-1).

Two prostaglandin H synthases have been identified: prostaglandin H synthase 1 is expressed constitutively in all cells; prostaglandin H synthase 2 is constitutively expressed in some cells (e.g., neurons and renal cells) and is induced in other cell types in response to cytokines (e.g., in monocytes), tumor promoters (e.g., in intestinal epithelial cells), growth factors (e.g., in bone marrow–derived stem cells), and laminar shear stress (e.g., in endothelial cells). Because nonsteroidal anti-inflammatory drugs (NSAIDs) target the COX activity of these enzymes, they have become known colloquially as COX-1 and COX-2.[1] Both enzymes are homodimers that convert arachidonic acid, two oxygen molecules, and two electrons from one or more unknown reductants to prostaglandin H_2. COX-1 and COX-2 are found predominantly in the same cellular organelles, at the luminal surface of the endoplasmic reticulum and nuclear envelope of cells. Only one monomer of a dimer catalyzes arachidonic acid oxygenation at any given time. The cross-talk between monomers serves as a way for COX-2 to exhibit selectivity toward arachidonic acid, even when it is a minor component of the available fatty acid pool, and to sustain a "late phase" of prostanoid production. In contrast, COX-1 may efficiently oxygenate arachidonic acid in the early phase of prostanoid production, when this substrate represents a large fraction of free fatty acids, but may be inhibited by other competing fatty acids in the late phase.

Phenotypical analyses of COX-1–deficient and COX-2–deficient mice as well as studies with isoform-selective inhibitors suggest that there are processes in which each isozyme is uniquely involved (e.g., platelet aggregation for COX-1, ovulation and neonatal development for COX-2) and others in which both isozymes function coordinately (e.g., inflammation and its resolution, gastrointestinal ulceration and healing, and carcinogenesis). A way in which the two biosynthetic pathways may be dissociated metabolically is by a preferential coupling of the COX isozymes to various upstream phospholipases and downstream synthases, conditioning preferential formation of a particular prostanoid by a given cell type (e.g., COX-2–dependent prostacyclin by vascular endothelial cells).

Moreover, COX-2 oxygenation may play a unique role in a novel signaling pathway dependent on agonist-induced release of endocannabinoids.[2] Among the products of COX-2 oxygenation of endocannabinoids are glyceryl prostaglandins, some of which (e.g., glyceryl prostaglandin E_2 and glyceryl prostaglandin I_2) exhibit interesting biologic activities in inflammatory, neurologic, and vascular systems.

MEASUREMENTS OF THE PROSTAGLANDIN H SYNTHASE PATHWAY

Prostanoids are formed in vivo at a relatively low rate (e.g., 0.1 ng/kg per minute for both prostacyclin and thromboxane A_2) and are metabolized extensively by lung and liver enzymes to form chemically stable but biologically inactive derivatives that are excreted primarily through the kidney. Given the chemical instability and extremely low concentrations (1 to 2 pg/mL) of prostanoids in the systemic circulation, assessment of their production in humans is largely based on measurements of stable urinary metabolites (e.g., 11-dehydro-thromboxane B_2, a major enzymatic derivative of thromboxane A_2, and 2,3-dinor-6-keto-$PGF_{1\alpha}$, a major enzymatic derivative of prostacyclin). These analytical measurements have established that thromboxane metabolite excretion is primarily derived from platelet

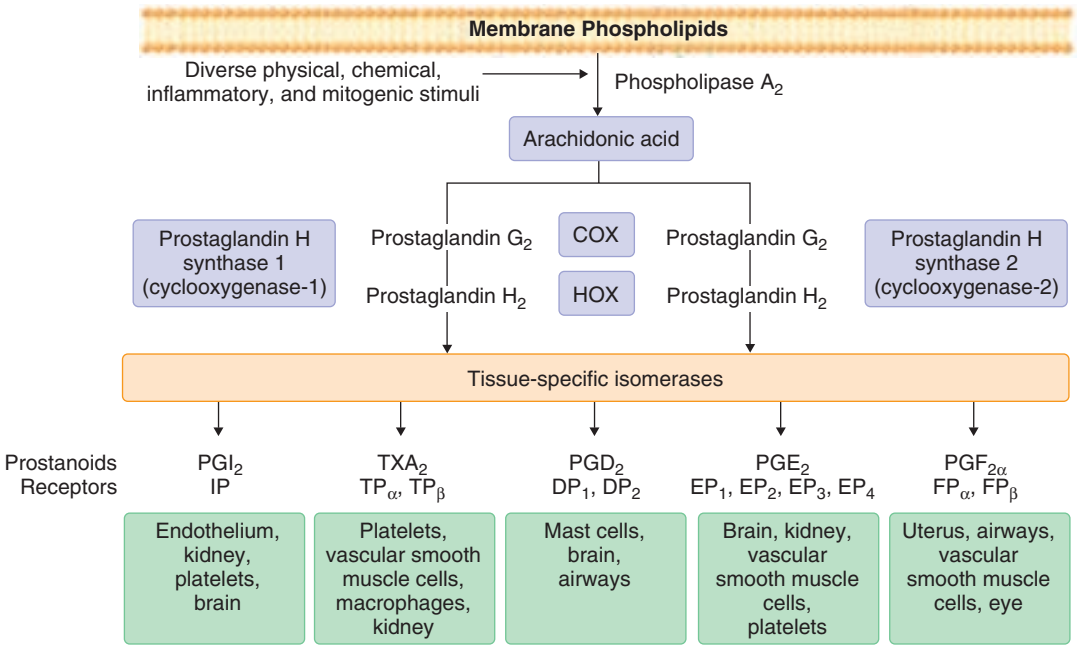

FIGURE 37-1. Production and actions of prostaglandins and thromboxane. Arachidonic acid, a 20-carbon fatty acid containing four double bonds, is liberated from the sn2 position in membrane phospholipids by phospholipase A_2, which is activated by diverse stimuli. Arachidonic acid is converted by prostaglandin H synthases, which have both cyclooxygenase (COX) and hydroperoxidase (HOX) activity, to the unstable intermediate prostaglandin H_2. The synthases are colloquially termed cyclooxygenases and exist in two forms, cyclooxygenase 1 and cyclooxygenase 2. Prostaglandin H_2 is converted by tissue-specific isomerases to multiple prostanoids. These bioactive lipids activate specific cell membrane receptors of the superfamily of G protein–coupled receptors. Some of the tissues in which individual prostanoids exert prominent effects are indicated. DP = prostaglandin D_2 receptor; EP = prostaglandin E_2 receptor; FP = prostaglandin $F_{2\alpha}$ receptor; IP = prostacyclin receptor; TP = thromboxane receptor.

thromboxane biosynthesis and largely reflects the rate of platelet activation in vivo. Similarly, prostacyclin metabolite excretion largely reflects the rate of vascular prostacyclin biosynthesis in vivo. Thromboxane biosynthesis is persistently enhanced in association with the major cardiovascular risk factors (e.g., diabetes mellitus). The biosynthesis of both thromboxane and prostacyclin is episodically increased in patients with acute coronary syndromes, perhaps reflecting a homeostatic response to accelerated platelet-vascular interactions.

Studies of the human pharmacology of COX inhibitors have largely relied on the development of whole blood assays of platelet COX-1 (based on serum thromboxane B_2 measurements) and monocyte COX-2 (based on lipopolysaccharide-induced prostaglandin E_2 production) activities. These assays have been useful in characterizing the variable potency of NSAIDs in inhibiting COX-1 and COX-2 in vitro (a measure of isozyme selectivity) and in determining ex vivo the dose and time dependence of their inhibitory effects in health and disease.

CLINICAL PHARMACOLOGY OF PROSTAGLANDIN H SYNTHASE INHIBITION

Most traditional NSAIDs inhibit COX-1 and COX-2 with similar potency (Fig. 37-2). Some traditional NSAIDs (e.g., nimesulide and diclofenac) and a class of COX-2 inhibitors called coxibs (e.g., celecoxib and etoricoxib) are more potent in inhibiting COX-2 than COX-1 (Fig. 37-2). These drugs fall into three general categories on the basis of their mechanism of action. One category of inhibitors includes freely reversible competitive inhibitors, such as ibuprofen and mefenamic acid. Binding of these inhibitors to the COX sites of both monomers composing a dimer is required for inhibition of COX-2 oxygenation of the substrate. A second group of inhibitors, including flurbiprofen, meclofenamate, diclofenac, and indomethacin, comprises time-dependent, noncovalent inhibitors. These NSAIDs are allosteric inhibitors that bind to one monomer of COX to inhibit its activity. Aspirin is unique to a third group of inhibitors that cause a time-dependent, covalent inhibition. Binding of aspirin to COX-1 or COX-2 leads to irreversible acetylation of a highly conserved serine residue (Ser-529 and Ser-516 in human COX-1 and COX-2, respectively). Aspirin acetylates only one monomer of a COX-1 dimer to cause complete loss of COX activity. Aspirin also maximally acetylates one monomer of human COX-2. The acetylated monomer of aspirin-treated COX-2 forms 15-hydroperoxyeicosatetraenoic acid from arachidonic acid, whereas the nonacetylated partner monomer forms mainly prostaglandin H_2 but only at 15 to 20% of the rate of native COX-2. Thus, the effect of aspirin on COX-2 is an incomplete allosteric inhibition compared with that seen with COX-1.

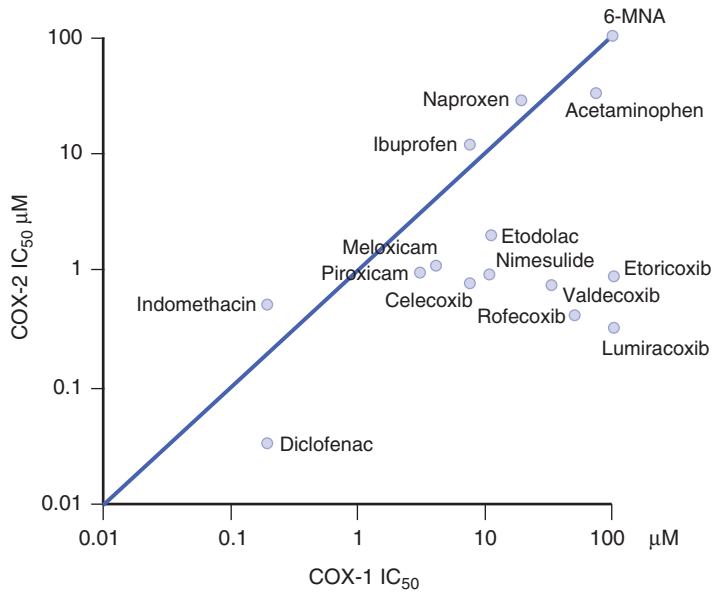

FIGURE 37-2. COX-2 selectivity as a continuous variable. Concentrations of various COX-2 inhibitors to inhibit the activity of platelet COX-1 and monocyte COX-2 by 50% (IC_{50}) are plotted on the abscissa and ordinate scales, respectively. The *solid line* describes equipotent inhibition of both COX-1 and COX-2. Symbols to the left of this line denote greater inhibition of COX-1 than of COX-2. Symbols to the right of this line indicate progressively greater inhibition of COX-2 than of COX-1, that is, increasing degrees of COX-2 selectivity. Aspirin is not shown on the figure because the long-term incubation required for monocyte COX-2 expression in human whole blood affects the chemical stability of the drug and underestimates its inhibitory potency. 6-MNA = 6-methoxy-2-naphthylacetic acid, the active metabolite of nabumetone.

Traditional NSAIDs typically inhibit platelet COX-1 and monocyte COX-2 by 50 to 90%, depending on dose; this effect is usually transient, depending on dose and half-life (Fig. 37-3). Coxibs inhibit monocyte COX-2 to the same extent as other NSAIDs while substantially sparing platelet (and presumably other cells) COX-1 in most patients exposed to therapeutic

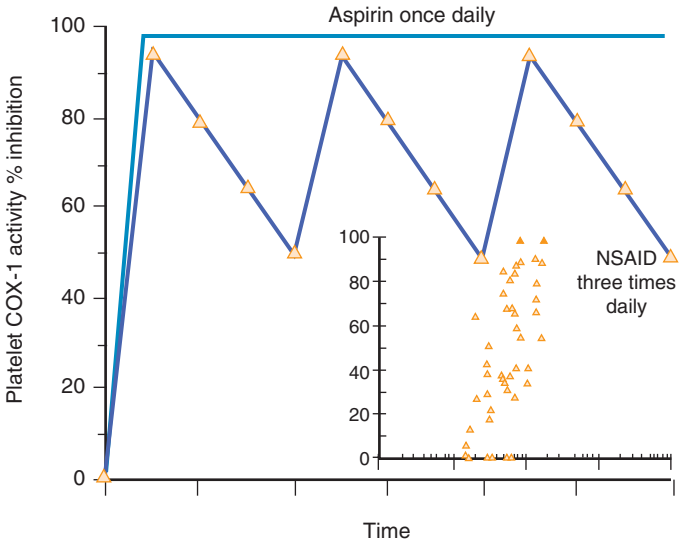

FIGURE 37-3. Time-dependent inhibition of platelet COX-1 activity by aspirin and a traditional nonsteroidal anti-inflammatory drug (NSAID). The average time course of inhibition of serum thromboxane B_2, an ex vivo index of platelet COX-1 activity, is depicted during 24 hours after the administration of low-dose aspirin once daily and a traditional NSAID with short half-life given every 8 hours. The *inset* depicts the interindividual variability in the relationship between NSAID plasma levels plotted on the abscissa log scale and the corresponding level of inhibition of platelet COX-1 plotted on the ordinate.

doses. In contrast, aspirin achieves virtually complete (i.e., >97%) and persistent (i.e., ≥24 hour) inactivation of platelet COX-1 by virtue of its irreversible mechanism of action and inability of anucleate platelets to resynthesize the enzyme. Aspirin is equally potent in acetylating COX-1 and COX-2 in vitro. However, its unique mechanism of action and unusual pharmacokinetic features (20-minute half-life; presystemic encounter with the platelet target in the portal blood before first-pass liver metabolism) allow selective, cumulative inhibition of platelet COX-1 at low doses while substantially sparing vascular COX-2. The effect of aspirin on COX-1–dependent thromboxane A_2 production is saturable at daily doses as low as 30 to 50 mg; in contrast, its inhibitory effect on COX-2–dependent prostaglandin I_2 biosynthesis is dose dependent up to daily doses of 650 to 1300 mg.

The relationships among inhibition of COX isozyme activity, reduced prostanoid formation, and changes in prostanoid-dependent cell function in vivo are not necessarily linear. The strikingly nonlinear relationship between inactivation of platelet COX-1 and inhibition of thromboxane-dependent platelet activation in vivo has important clinical implications for the cardiovascular effects of low-dose aspirin versus traditional NSAIDs (see later). In addition, interindividual variability in drug plasma levels as well as in the corresponding level of COX isozyme inhibition contributes to substantial unpredictability of the individual clinical response to COX inhibitors (E-Fig. 37-1).

Low-Dose Aspirin as an Antithrombotic and Anti-Cancer Agent

Antithrombotic Effects

The efficacy and safety of aspirin as an antithrombotic agent have been evaluated in several populations, ranging from apparently healthy persons at low risk of vascular complications (so-called primary prevention) to high-risk patients presenting with or surviving an acute myocardial infarction or an acute ischemic stroke (so-called secondary prevention). The clinical efficacy of aspirin was demonstrated at doses ranging from 50 to 162 mg given once daily (Table 37-1), consistent with the irreversible nature of its mechanism of action. Furthermore, higher doses (e.g., 300 to 325 mg) were not found to confer additional benefits, consistent with saturability of platelet COX-1 acetylation at low doses.[3,4]

In the six primary prevention trials among 95,000 low-risk individuals, aspirin allocation yielded a 12% relative risk reduction in serious vascular events (myocardial infarction, stroke, or vascular death). This protective effect was mainly due to a reduction in nonfatal myocardial infarction. The net effect on stroke was not significant, reflecting a small reduction in ischemic stroke and counterbalancing effects on hemorrhagic stroke. There was no significant reduction in vascular mortality. Aspirin increased gastrointestinal (or other extracranial) bleeds by approximately 50%. The balance of cardiovascular benefits and bleeding risk associated with low-dose aspirin in specifically primary prevention is uncertain.[5]

TABLE 37-1 VASCULAR DISORDERS FOR WHICH ASPIRIN HAS BEEN SHOWN TO BE EFFECTIVE AND THE LOWEST EFFECTIVE DOSE

DISORDER	LOWEST EFFECTIVE DAILY DOSE (mg)
TIA and ischemic stroke*	50
Men at high cardiovascular risk	75
Essential hypertension	75
Chronic stable angina	75
Unstable angina or NSTEMI*	75
Severe carotid artery stenosis*	75
Polycythemia vera	100
Acute ischemic stroke*	160
Acute STEMI	162

*Higher doses were tested and not found to confer any greater risk reduction.
NSTEMI = non–ST elevation myocardial infarction; STEMI = ST elevation myocardial infarction; TIA = transient ischemic attack.

In 16 secondary prevention trials in 17,000 high-risk patients with prior myocardial infarction, or prior stroke or transient cerebral ischemia, aspirin allocation yielded 19% fewer serious vascular events, with similar proportional reductions in coronary events (20% relative risk reduction) and ischemic stroke (22% relative risk reduction) but a nonsignificant increase in hemorrhagic stroke. The absolute benefit of aspirin was about 25 times larger in secondary than in primary prevention (15 vs. 0.6 fewer vascular events per 1000 per year). In both primary and secondary prevention trials, the proportional reductions in serious vascular events appeared similar for men and women and for older and younger people. The risks of serious vascular events and of major extracranial bleeds were predicted by the same independent risk factors (age, male gender, diabetes mellitus, current smoking, blood pressure, and body mass index), so those with high risk of vascular complications also had a high risk of bleeding. For secondary prevention of cardiovascular disease, the net benefits of adding aspirin to other preventive measures (e.g., statins) substantially exceed the bleeding hazards, irrespective of age and gender.

Anti-Cancer Effects

Aspirin, 75 mg daily or more for at least several years, reduces the incidence and mortality of colorectal cancer. Long-term follow-up of randomized vascular prevention trials of daily aspirin versus control showed that aspirin reduced not only the incidence of and mortality due to colorectal cancer but also death due to several other common cancers. Furthermore, meta-analysis has shown short-term reduction by daily aspirin in cancer incidence and mortality in women as well as in men and in non-smokers as well as in smokers. In a randomized trial of carriers of hereditary colorectal cancer genes, aspirin (600 mg daily) for a mean of 25 months substantially reduced the incidence of cancer after about 5 years. Low-dose aspirin also reduces the risk of metastases, especially for adenocarcinomas. The mechanism underlying the chemopreventive effect of low-dose aspirin might involve inhibition of platelet activation.[6] More recent evidence shows that the risk of major bleeding with aspirin diminishes with prolonged use, suggesting that the balance of risk and benefit favors the use of daily aspirin in primary prevention of colorectal and other cancers.[7] A regulatory review of the existing evidence as well as treatment guidelines are needed for this potential chemopreventive strategy.

Traditional Nonsteroidal Anti-inflammatory Drugs and Coxibs

NSAIDs constitute a chemically heterogeneous group of compounds that provide symptomatic relief of pain and inflammation associated with a variety of human disorders, including the rheumatic diseases. Their shared therapeutic actions (i.e., analgesic, anti-inflammatory, and antipyretic) are usually accompanied by mechanism-based adverse effects on gastrointestinal, cardiovascular, and renal functions. Prostanoids reproduce the main signs and symptoms of the inflammatory response and cause hyperalgesia and fever. Because of the redundancy of mediators of these responses, it is not surprising that NSAIDs exert only a moderate anti-inflammatory effect, are effective only against pain of low to moderate intensity, and reduce fever but do not interfere with the physiologic control of body temperature. The analgesic, anti-inflammatory, and antipyretic actions of traditional NSAIDs are largely reproduced by coxibs, a class of selective inhibitors of COX-2.

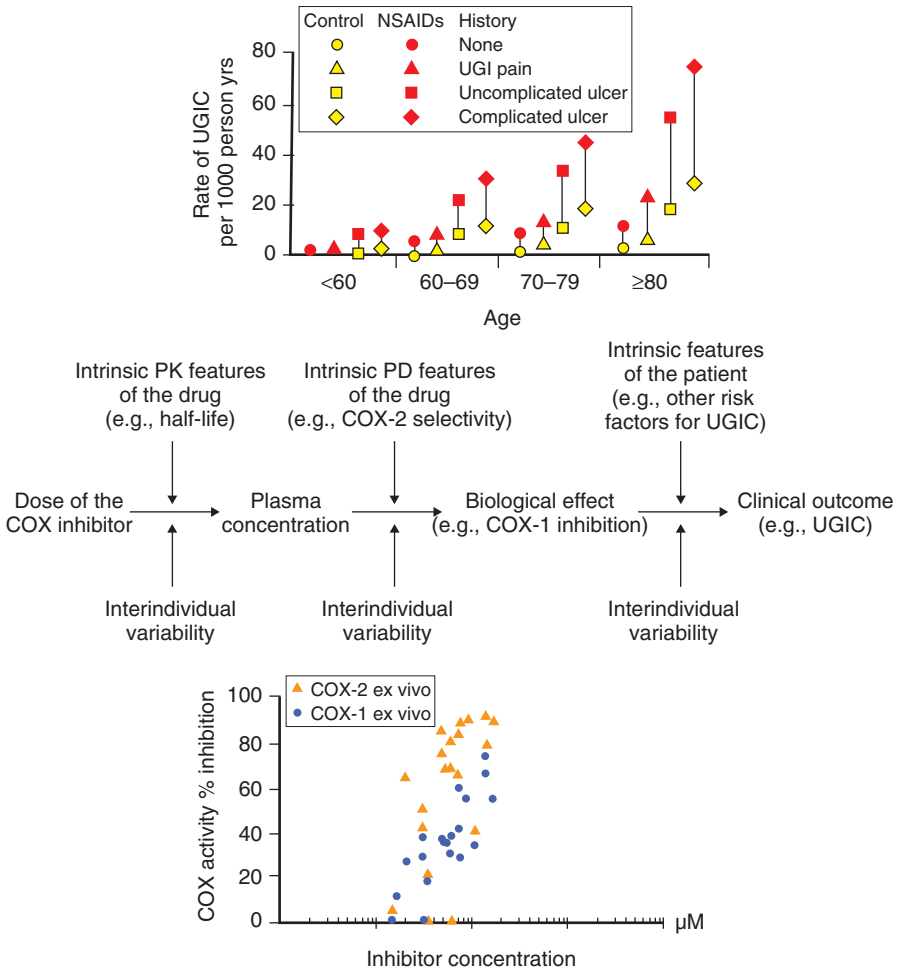

E-FIGURE 37-1. Determinants and sources of variability in the complications of COX-inhibitory drugs. Major determinants of the likelihood of an upper gastrointestinal complication (UGIC) resulting from administration of a cyclooxygenase (COX) inhibitor include pharmacokinetic (PK) and pharmacodynamic (PD) variables as well as the interaction of the drug with preexisting risk factors for UGIC. PK features, such as half-life of the drug, and PD features, such as its selectivity for the COX-2 isoform, are intrinsic to the COX inhibitor. The presence or absence of risk factors for UGIC will clearly vary between patients. Significant interindividual variability arises from several sources and is superimposed on each of these effects. In addition to the variable plasma levels achieved after oral dosing of a COX inhibitor, the inhibition of the platelet COX-1 isozyme (*circles*) and the monocyte COX-2 isozyme (*triangles*) in response to any given plasma level of the inhibitor is highly variable in individual subjects, as shown in the lower panel. The estimated rates of UGIC in women, according to age and the presence or absence of a history of such complications and regular use of NSAIDs, are represented in the upper panel. The solid lines connecting each pair of yellow and red symbols depict the absolute excess of complications related to NSAID therapy.

COX-2 selectivity is a continuous variable (see Fig. 37-2). Thus, one can pragmatically characterize three levels of COX-2 selectivity in terms of the probability of sparing COX-1 at therapeutic plasma levels: low (e.g., acetaminophen), intermediate (e.g., celecoxib, nimesulide, and diclofenac), and high (e.g., rofecoxib, etoricoxib, and lumiracoxib).

Drug Interactions

NSAIDs can modify the pharmacokinetics or pharmacodynamics of other drugs given concurrently, resulting in clinically important drug interactions. A pharmacodynamic interaction may occur between most NSAIDs and several classes of antihypertensive drugs. Reduced production of vasodilator prostacyclin and natriuretic prostaglandin E_2, as a consequence of renal COX-2 inhibition, results in vasoconstriction and sodium and water retention that in turn tend to elevate blood pressure, regardless of the mechanism of action of antihypertensive drugs. This pharmacodynamic interaction has been described with most traditional NSAIDs (including acetaminophen) and coxibs, but not with low-dose aspirin.

Some NSAIDs favoring COX-1 over COX-2 inhibition, such as ibuprofen and naproxen, may interfere with the antiplatelet effect of low-dose aspirin by competing with acetylsalicylic acid for a common docking site (arginine-120) within the COX-1 channel. Drugs favoring COX-2 versus COX-1 inhibition, such as acetaminophen and diclofenac, do not interfere with the pharmacodynamic effect of low-dose aspirin, similar to celecoxib and rofecoxib.

Gastrointestinal and Bleeding Complications

Upper gastrointestinal complications (bleeds, perforations, and obstructions)[8,9] occur in 1 to 2% of NSAID-treated patients (Chapter 139). The mortality rate associated with hospitalization due to major gastrointestinal events is 5 to 6% in recent studies. Mortality rates associated with upper or lower gastrointestinal complications due to NSAIDs are similar. The major risk factors for upper gastrointestinal bleeding are represented by age and a prior history of gastrointestinal disorders (E-Fig. 37-1). Male gender, cigarette smoking, and heavy alcohol intake increase this risk by less than two-fold, as do oral glucocorticoids. Oral anticoagulants, thienopyridines, and low-dose aspirin increase the risk of NSAID-induced bleeding complications by two- to three-fold. The excess of these complications due to traditional NSAIDs has been estimated to range between 3 and 30 events per 1000 patients treated per year, depending on the absence or presence of risk factors (E-Fig. 37-1).

Highly selective COX-2 inhibitors are associated with a statistically significant 50 to 66% relative risk reduction in ulcer complications compared with naproxen or ibuprofen. However, no such agent is currently available on the U.S. market.

Cardiovascular Complications

A meta-analysis of individual participant data from randomized trials of five different coxibs has revealed that in placebo comparisons, allocation to a coxib was associated with a 37% increased risk of major vascular events with no statistically significant heterogeneity among the different coxibs.■ This excess risk of vascular events was derived primarily from a two-fold increased risk of myocardial infarction. Overall, there was no significant difference in the incidence of vascular events between a coxib and any traditional NSAID, but there was evidence of significant heterogeneity between naproxen and the other traditional NSAIDs (largely represented by ibuprofen and diclofenac). The proportional effects on major vascular events were independent of baseline characteristics, including vascular risk. Therefore, the absolute excess of major vascular events caused by coxibs and some traditional NSAIDs varied between 2 and 9 for every 1000 patients allocated to a year of treatment, depending on the baseline level of cardiovascular risk.

Current evidence suggests that the risk of myocardial infarction depends on the extent of COX-2 inhibition and not on the variable COX-2 selectivity of the inhibitor.[10] This risk appears to be modulated by concomitant high-grade and persistent inhibition of platelet COX-1 activity, as suggested by the neutral cardiovascular phenotype associated with a high-dose regimen of naproxen.■ However, in patients at high cardiovascular risk, whose platelet COX-1 is completely and persistently inactivated by low-dose aspirin, the administration of any COX-2 inhibitor (including naproxen) is likely to produce detrimental cardiovascular consequences.

Given the nonlinear relationship between inhibition of platelet COX-1 activity and inhibition of platelet activation in vivo, it is perhaps not surprising that the cardiovascular safety profiles of coxibs and some traditional NSAIDs (e.g., diclofenac) appear similar because they both fail to inhibit platelet activation adequately irrespective of their COX-2 selectivity. Early appearance, dose dependence, and slow dissipation of risk are important features of COX-2-related cardiotoxicity.

Grade A References

A1. Baigent C, Blackwell L, Collins R, et al. for the Antithrombotic Trialists' (ATT) Collaboration. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta-analysis of individual participant data from randomised trials. *Lancet*. 2009;373:1849-1860.

A2. Rothwell PM, Wilson M, Elwin CE, et al. Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. *Lancet*. 2010;376:1741-1750.

A3. Rothwell PM, Price JF, Fowkes FG, et al. Short-term effects of daily aspirin on cancer incidence, mortality, and non-vascular death: analysis of the time course of risks and benefits in 51 randomised controlled trials. *Lancet*. 2012;379:1602-1612.

A4. Burn J, Gerdes AM, Macrae F, et al. Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial. *Lancet*. 2011; 378:2081-2087.

A5. Rothwell PM, Wilson M, Price JF, et al. Effect of daily aspirin on risk of cancer metastasis: a study of incident cancers during randomised controlled trials. *Lancet*. 2012;379:1591-1601.

A6. CNT Collaboration. Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomised trials. *Lancet*. 2013;382: 769-779.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Rouzer CA, Marnett LJ. Cyclooxygenases: structural and functional insights. *J Lipid Res*. 2009; 50:S29-S34.
2. Blankman JL, Cravatt BF. Chemical probes of endocannabinoid metabolism. *Pharmacol Rev*. 2013; 65:849-871.
3. CURRENT-OASIS 7 Investigators, Mehta SR, Bassand JP, et al. Dose comparisons of clopidogrel and aspirin in acute coronary syndromes. *N Engl J Med*. 2010;363:930-942.
4. Patrignani P, Tacconelli S, Piazuelo E, et al. Reappraisal of the clinical pharmacology of low-dose aspirin by comparing novel direct and traditional indirect biomarkers of drug action. *J Thromb Haemost*. 2014;12:1320-1330.
5. Patrono C. Low-dose aspirin in primary prevention: cardioprotection, chemoprevention, both or neither? *Eur Heart J*. 2013;34:3403-3411.
6. Thun MJ, Jacobs EJ, Patrono C. The role of aspirin in cancer prevention. *Nat Rev Clin Oncol*. 2012;9:259-267.
7. Rothwell PM. Aspirin in prevention of sporadic colorectal cancer: current clinical evidence and overall balance of risks and benefits. *Recent Results Cancer Res*. 2013;191:121-142.
8. Thiagarajan P, Jankowski JA. Aspirin and NSAIDs; benefits and harms for the gut. *Best Pract Clin Gastroenterol*. 2012;26:197-206.
9. García Rodríguez LA, Lin KJ, Hernández-Díaz S, Johannsson S. Risk of upper gastrointestinal bleeding with low-dose acetylsalicylic acid alone and in combination with clopidogrel and other medications. *Circulation*. 2011;123:1108-1115.
10. Patrono C, Baigent C. Nonsteroidal anti-inflammatory drugs and the heart. *Circulation*. 2014;129: 907-916.

REVIEW QUESTIONS

1. The concomitant administration of which of the following drugs can interfere with the antiplatelet effect of low-dose aspirin?

A. All traditional nonsteroidal anti-inflammatory drugs (NSAIDs)
B. Celecoxib
C. Ibuprofen and naproxen
D. Acetaminophen
E. Diclofenac

Answer: C Ibuprofen and naproxen are somewhat more potent in inhibiting COX-1 than COX-2 (as shown in Fig. 37-2) and can compete with aspirin for binding to a common docking site within the COX-1 channel, thereby interfering with permanent acetylation of the platelet enzyme. Other COX inhibitors that have some degree of COX-2 selectivity, such as celecoxib, acetaminophen, and diclofenac, do not display this pharmacodynamic interaction.

2. Which of the following is true regarding the increased risk for major vascular events associated with the use of COX-2 inhibitors?

A. The increased risk is restricted to coxibs.
B. The increased risk is shared by all NSAIDs.
C. The increased risk is related to COX-2 selectivity.
D. The increased risk is shared by NSAIDs that do not adequately inhibit platelet COX-1 activity.
E. The increased risk requires a treatment duration of at least 18 months.

Answer: D Coxibs and traditional NSAIDs that do not adequately inhibit platelet COX-1 activity (e.g., diclofenac and ibuprofen) similarly increase the risk for major vascular events, regardless of their variable COX-2 selectivity. The increased risk appears early and does not require prolonged exposure. See reference A6.

3. The main features of the antithrombotic effect of low-dose aspirin include which of the following?

A. Saturability of the effect at low doses
B. A lowest effective daily dose in the range of 160 to 325 mg
C. Differential effects in men versus women
D. Differential effects in older versus younger subjects
E. None of the above

Answer: A As indicated in Table 37-1, the lowest effective daily dose is in the range of 50 to 162 mg. The antithrombotic effect of aspirin is saturable at such low doses, inasmuch as higher doses were tested and not found to confer any greater risk reduction. The effect is similar in men and women, as well as in subjects younger and older than 65 years.

4. Prostaglandin H synthases do which of the following?

A. Catalyze the conversion of PGH_2 to PGE_2
B. Catalyze the first committed step in prostanoid biosynthesis from arachidonic acid
C. Catalyze the release of arachidonic acid from membrane phospholipids
D. Are inhibited irreversibly by all NSAIDs
E. None of the above

Answer: B As shown in Figure 37-1, prostaglandin H synthases catalyze the conversion of arachidonic acid to PGG_2 and its reduction to PGH_2, that is, the first committed step in prostanoid biosynthesis. Other enzymes catalyze the release of arachidonic acid from membrane phospholipids and the conversion PGH_2 to PGE_2. Aspirin, but not other NSAIDs, inhibits prostaglandin H synthases irreversibly.

5. Which of the following is true of low-dose aspirin and traditional NSAIDs?

A. They share the same molecular mechanism of permanent inactivation of COX-isozymes.
B. They share the same pharmacologic effects.
C. They differentially inhibit platelet COX-1 activity as a function of their mechanism of action and half-life.
D. They have the same cardiovascular effects.
E. None of the above

Answer: C As show in Figure 37-3, low-dose aspirin and a traditional NSAID with a short half-life (e.g., ibuprofen) differentially inhibit platelet COX-1 activity during the dosing interval because of the irreversible versus reversible mechanism of action. Aspirin is primarily an antiplatelet agent at low doses administered once daily and requires higher doses and more frequent dosing to produce the same pharmacologic effects of traditional NSAIDs. The cardioprotective effects of low-dose aspirin are not shared by traditional NSAIDs.

38

ANTITHROMBOTIC THERAPY

SAM SCHULMAN AND JACK HIRSH

Antithrombotic therapy suppresses the natural hemostatic mechanisms (Chapter 171) and is effective for preventing and treating venous, cardiac, and arterial thromboembolism. A variety of medications are now available that interfere with different steps in coagulation and platelet activation, sometimes with synergistic effects. Randomized clinical trials have produced a substantial body of evidence-based literature to guide the use of antithrombotic therapy for a wide range of clinical conditions. Updated recommendations have been published in the new 2012 American College of Chest Physicians (ACCP) evidence-based clinical practice guidelines, 9th edition (also accessible at http://journal.publications.chestnet.org/ss/guidelines.aspx). (See Grade A Recommendations for Antithrombotic Therapy at the end of the text in this chapter.)

PHARMACOLOGIC AGENTS

Vitamin K Antagonists

For more than 60 years, vitamin K antagonists have been the only oral anticoagulants available for clinical use. Now, with the development of new oral agents that target single coagulation enzymes (see later), the situation is changing. Coumarins are vitamin K antagonists, of which warfarin is the most widely used. Coumarins inhibit a vitamin K reductase that catalyzes the reduction of 2,3-epoxide (vitamin K epoxide), thereby leading to the depletion of vitamin KH_2, which is required for the production of functionally active (γ-carboxylated) coagulation proteins (factors II [prothrombin], VII, IX, and X) and anticoagulant proteins (protein C and protein S) (Chapter 175). Vitamin K_1 in food sources can reverse these effects of coumarins because it is reduced to vitamin KH_2 by a warfarin-insensitive vitamin K reductase (Fig. 38-1).

Warfarin is rapidly and almost completely absorbed from the gastrointestinal tract. It has a half-life of about 40 hours, a delayed onset of action (2 to 7 days, depending on dose), and a residual anticoagulant effect for up to 5 days after treatment is discontinued. The dose-response relationship of warfarin varies widely among individuals and is influenced by many factors, including age, body weight, liver disease, dietary vitamin K_1, genetic factors, concomitant drug use, compliance of the patient, and inappropriate dosage adjustments. Of these factors, inappropriate dosage adjustment and improved compliance through patient education are the most readily correctable.

The effect of warfarin must be monitored closely to prevent overdosing or underdosing. Laboratory monitoring is performed by measuring the

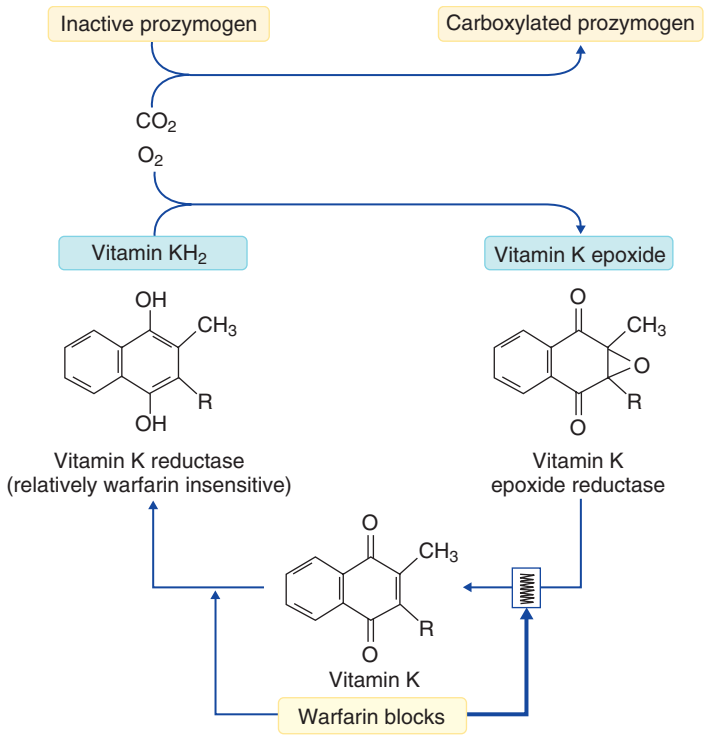

FIGURE 38-1. Warfarin inhibits vitamin K epoxide reductase and leads to the intracellular depletion (in the hepatocyte) of vitamin KH_2. Vitamin KH_2 is required for the conversion (by γ-carboxylation) of functionally inactive to active coagulation proteins. The anticoagulant effect of warfarin can be reversed by vitamin K_1 in food because it is reduced to vitamin KH_2 by a warfarin-insensitive vitamin K reductase.

prothrombin time and is reported as an international normalized ratio (INR). During initiation of warfarin therapy, the INR reflects primarily the depression of factor VII, which has a half-life of only 6 hours. The reliability of warfarin monitoring appears to be improved by using paper nomograms or computer-assisted algorithms. The convenience of monitoring is increased with a portable point-of-care instrument. Pharmacogenetic-guided dosing of warfarin has been evaluated in four randomized trials without producing evidence that the rate of major hemorrhage or thromboembolic complications is reduced, and the current cost is $100 to $200 per patient. The 2012 ACCP practice guidelines (http://journal.publications.chestnet.org/ss/guidelines.aspx) summarize published literature concerning the laboratory and clinical characteristics of vitamin K antagonists.

Indications for Warfarin

Warfarin is effective in the primary and secondary prevention of systemic embolism in patients with atrial fibrillation (Chapter 64),[■] with rheumatic mitral valve complicated stenosis or by left atrial thrombus[■]; with bioprosthetic or mechanical heart valves (Chapter 75); in the primary and secondary prevention of venous thromboembolism (VTE) (Chapters 81 and 98); in the prevention of acute myocardial infarction in high-risk patients (Chapters 72 and 73); and in the prevention of stroke (Chapter 407), recurrent infarction, and death in patients with acute myocardial infarction (Chapter 73). A target INR of 2.5 (range, 2.0 to 3.0) is recommended for almost all indications. Exceptions are mechanical prosthetic heart valve in the mitral position or caged-ball or caged-disk valve in the aortic position or any mechanical aortic valve in combination with atrial fibrillation, anterior myocardial infarction, left atrial enlargement, or low ejection fraction, when an INR of 3.0 (range, 2.5 to 3.5) is recommended.

Dosing and Monitoring

If a rapid anticoagulant effect is required, heparin and warfarin should be started at the same time and overlapped for at least 5 days. Warfarin is started with the estimated maintenance dose of about 5 mg/day, with the first INR measurement after 2 to 3 days, and patients usually reach an INR of 2.0 in 4 or 5 days. If there is no increase of the INR after two or three doses, the daily dose should be progressively increased until an INR response is observed. In patients with a low risk for bleeding, warfarin may be started at a dose of 10 mg and then adjusted according to daily INR results. Heparin treatment is discontinued when the INR has been in the therapeutic range for 2 days.

TABLE 38-1 RECOMMENDED MANAGEMENT OF ELEVATED INR WITH OR WITHOUT BLEEDING IN PATIENTS TREATED WITH WARFARIN

INR	BLEEDING	WARFARIN	VITAMIN K_1	FFP/PCC/rFVIIa
<5.0	Negligible	Hold 1 dose or reduce dose	No	No
5.0-9.9	Negligible	Hold 1-2 doses	Generally no*	No
≥10	Negligible	Hold	2.5-5 mg PO	No
Any	Serious or life threatening	Hold	10 mg IV† and repeat PRN	Yes

*1-2.5 mg PO for patients at increased risk for bleeding.
†Intravenous (IV) infusion should be given slowly.
FFP = fresh-frozen plasma; INR = international normalized ratio; PCC = prothrombin complex concentrate; PO = orally; PRN = as needed; rFVIIa, recombinant factor VIIa.

The INR is then performed two or three times weekly for 1 to 2 weeks and then weekly up to a maximal interval of 4 weeks, depending on the stability of INR results, and more frequently when a new drug is added to the treatment. Once the dose is stable, warfarin assessment up to every 12 weeks appears as safe and effective as every 4 weeks.[■■]

Adjustments to the dose when the INR drifts out of the therapeutic range should be gradual and based on the weekly dose (e.g., 10 to 20% changes in weekly dose). Patients should be encouraged to keep a log of their dose and their INR response.

Adverse Effects

Warfarin-related bleeding is increased by the level of the INR. The risk for bleeding is also increased with concomitant aspirin use, in persons older than 65 years, in those with a history of stroke or gastrointestinal bleeding, and in those with serious comorbid conditions. Elderly patients are more sensitive to warfarin, requiring lower doses to reach the therapeutic range, and have an increased tendency to bleed, including intracranially, even when their INR is in the therapeutic range (Chapter 24).

Warfarin-induced skin necrosis (Fig. 176-1) occurs in 1 in 5000 patients, more frequently in women, and affects mainly breasts, buttocks, and thighs. An imbalance between moderately reduced procoagulant factors and severely depressed natural inhibitors may cause this hypercoagulable state in patients with congenital deficiency of protein C or protein S (Chapter 176), dietary deficiency of vitamin K, cancer, or heparin-induced thrombocytopenia with premature start of vitamin K antagonists (Chapter 172).

Reversing the Effect of Warfarin

The anticoagulant effect of warfarin can be reversed in one of three ways: by discontinuation of therapy, with the expectation that the INR will return to baseline in about 5 days; by administration of vitamin K_1, with the expectation that the anticoagulant effect will be reduced in 6 hours and reversed in 24 hours; and by infusion of fresh-frozen plasma (FFP), prothrombin complex concentrate (PCC), or recombinant coagulation factor VIIa (rFVIIa), which produce immediate reversal (Table 38-1).

Heparin and Low-Molecular-Weight Heparins

Heparin

Heparin binds to antithrombin (AT), thereby increasing the rate at which AT inactivates thrombin, activated factor X (factor Xa), and other coagulation enzymes. Heparin accelerates the inactivation of thrombin by AT by providing a template to which both the enzyme and the inhibitor bind to form a ternary complex (Fig. 38-2). In contrast, the inactivation of factor Xa by the AT-heparin complex does not require ternary complex formation and is achieved by binding of the heparin-bound AT to factor Xa. Heparin binds to a number of plasma, platelet, and endothelial cell–derived proteins that compete with AT for heparin binding. Binding of heparin to plasma proteins contributes to the variability of its anticoagulant response, whereas binding to hepatic macrophages is responsible for its dose-dependent clearance. Both properties contribute to the unpredictable anticoagulant effect of heparin and the need for laboratory monitoring. The plasma half-life is about 60 minutes at therapeutic concentrations.

Heparin is effective for the prevention and treatment of VTE,[■] for the early treatment of patients with unstable angina and acute myocardial infarction,[■] for patients who have cardiac surgery under cardiopulmonary bypass,

for patients undergoing vascular surgery, and during and after coronary angioplasty and coronary stent placement.

The anticoagulant effects of heparin are usually monitored by the activated partial thromboplastin time (aPTT). A therapeutic effect is achieved when the aPTT ratio is equivalent to a heparin level of 0.3 to 0.7 anti–factor Xa units, which for many reagents is an aPTT ratio of 1.5 to 2.5. The risk for bleeding complications is increased with increasing heparin dosage, which in turn is related to the anticoagulant response. However, other clinical factors, such as recent surgery, trauma, and invasive procedures, are also important as predictors of bleeding during heparin treatment.

For the treatment of VTE, heparin is given in doses of 80 U/kg followed by 18 U/kg per hour by continuous infusion; the dose is adjusted according to the aPTT result at 6 hours by use of a validated nomogram. Lower doses (70 U/kg or 5000 U followed by 15 U/kg/h or 1000 U/hour) of heparin are used in patients with acute myocardial ischemia, who also receive aspirin and platelet glycoprotein IIb/IIIa complex (GPIIb-IIIa) antagonists or thrombolytic therapy.

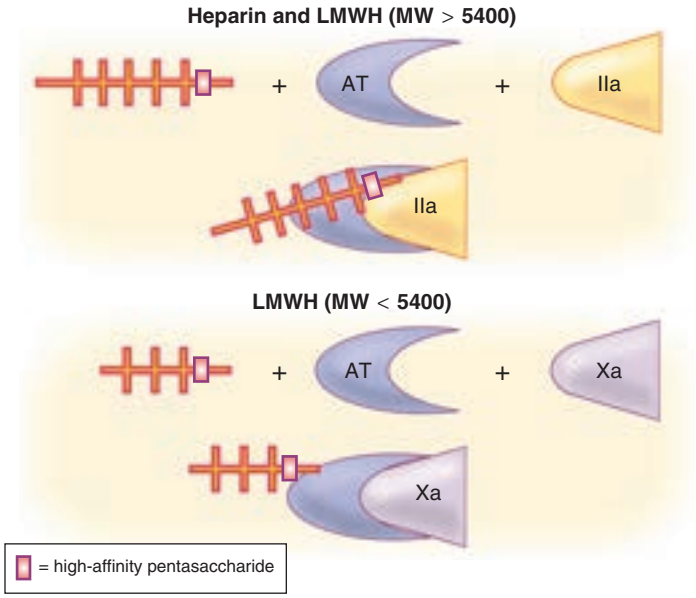

FIGURE 38-2. Only one third of high-affinity pentasaccharide-containing heparin molecules and one fifth of pentasaccharide-containing low-molecular-weight heparin (LMWH) molecules activate antithrombin (AT). Virtually all of the high-affinity heparin molecules are large enough to bridge between AT and factor IIa (thrombin). In contrast, only 25 to 50% of LMWH molecules have a molecular weight (MW) of 5400 or more, and although these smaller molecules inactivate factor Xa, they do not inactivate factor IIa. Although heparin has equal anti–factor IIa and anti–factor Xa activities, LMWH has reduced anti–factor IIa activity.

The main complications of heparin are bleeding and heparin-induced thrombocytopenia. Less common complications are heparin-induced osteoporosis and hyperkalemia. Heparin-related bleeding is dose related, and the risk is increased in patients who undergo an invasive procedure and if heparin is used in combination with a platelet GPIIb-IIIa antagonist or a thrombolytic agent.

If heparin-induced thrombocytopenia (Chapter 172) is suspected on clinical grounds and anticoagulant treatment is indicated, heparin should be stopped and replaced with a thrombin inhibitor: hirudin (lepirudin), argatroban, or danaparoid. Warfarin should not be used alone to treat acute heparin-induced thrombocytopenia because it can aggravate the thrombotic process, but it is safe in combination with a thrombin inhibitor after the platelet count has risen above 100×10^9/L.

Low-Molecular-Weight Heparins

Low-molecular-weight heparins (LMWHs) are fragments produced by either chemical or enzymatic depolymerization of heparin. LMWHs are approximately one third the size of heparin (Table 38-2). Depolymerization of heparin changes the anticoagulant profile. As a result, LMWHs have less protein and cellular binding and, as a consequence, have a more predictable dose response, better bioavailability, and a longer plasma half-life than regular heparin. LMWHs can therefore be administered subcutaneously once daily without laboratory monitoring.

Compared with heparin, which has a ratio of anti–factor Xa to anti–factor IIa activity of approximately 1 : 1, the various commercial LMWHs have ratios of anti–factor Xa to anti–factor IIa varying between 4 : 1 and 2 : 1, depending on their molecular size distribution. LMWHs are cleared principally by the renal route. They are associated with a lower incidence of heparin-induced thrombocytopenia and heparin-induced osteoporosis than is heparin.

LMWHs are effective in the prevention and treatment of VTE (Chapter 81),■ in the treatment of patients with unstable angina and non–ST elevation myocardial infarction (Chapters 72 and 73),■ and as an adjunct to fibrinolytic therapy in patients with acute ST elevation myocardial infarction (Chapter 73).

Pentasaccharides

On the basis of knowledge of the AT-binding sequence on heparin, a pentasaccharide, fondaparinux, with high affinity for AT has been synthesized. The structure of fondaparinux has been modified to increase its affinity to AT. Fondaparinux inactivates factor Xa through an AT-mediated mechanism. Because it is too short to bridge AT to thrombin, fondaparinux has no activity against thrombin.

After subcutaneous injection, fondaparinux is rapidly and completely absorbed and exhibits a bioavailability of 100%. The volume of distribution is similar to the blood volume. The drug is mainly excreted unchanged in the urine, with a terminal half-life of 17 hours in young volunteers and 21 hours in elderly volunteers.

TABLE 38-2 ANTICOAGULANT PROFILES, MOLECULAR WEIGHTS, PLASMA HALF-LIVES, AND RECOMMENDED DOSES OF COMMERCIAL LOW-MOLECULAR-WEIGHT HEPARINS AND HEPARINOID

AGENT	ANTI-X_a/ANTI-II_a RATIO	MOLECULAR WEIGHT	PLASMA HALF-LIFE (min)	*RECOMMENDED DOSE (INTERNATIONAL ANTI-XA UNITS)*		
				General Surgery Prophylaxis	*Orthopedic Surgery Prophylaxis*	ACUTE TREATMENT
Enoxaparin	2.7 : 1	4500	129-180	4000 U SC daily	4000 U SC daily or 3000 U SC bid	7000 U SC bid*† or 10,500 U SC daily*
Dalteparin	2 : 1	5000	119-139	2500 U SC daily	2500 U SC bid or 5000 U SC daily	8400 U SC bid*† or 14,000 U SC daily*
Nadroparin	3.2 : 1	4500	132-162	2850 U SC daily	2700 U SC daily,* 4000 U SC daily* from day 4	13,300 U SC daily*
Tinzaparin	1.9 : 1	4500	111	3500 U SC daily	3500 U SC daily* or 4500 U SC daily*	12,250 U daily*
Ardeparin	2 : 1	6000	200		50 U/kg SC bid	
Danaparoid‡	20 : 1	6500	1100	750 U SC daily	750 U SC bid	2500 U IV, then 4 hr each of 400 U/hr and 300 U/hr, then 200 U/hr; or 2000 U SC bid

*Weight-adjusted dose; stated dose for 70-kg patient.
†The higher daily dose is for acute coronary syndromes; the lower dose is for deep vein thrombosis.
‡Danaparoid sodium is a heparinoid.
IV = intravenously; SC = subcutaneously.

Fondaparinux circulates extensively bound to AT with minimal binding to other plasma proteins. Limited experimental and clinical studies suggest that fondaparinux has a lower risk for heparin-induced thrombocytopenia than that of heparin or LMWH as well as a lower risk for bone loss and of local skin reactions.

Fondaparinux is effective in the prevention and treatment of VTE and is also effective and safe in the treatment of acute coronary syndromes.

New Anticoagulants

The limitations of established anticoagulants have prompted the development of a variety of new anticoagulant agents that target various specific steps in the coagulation mechanism.[1,2]

Direct Thrombin Inhibitors

Direct thrombin inhibitors act independently of AT to inactivate both free thrombin and thrombin bound to fibrin. The direct thrombin inhibitors include hirudin, synthetic hirudin fragments (hirugen, lepirudin, and bivalirudin [Hirulog]), and low-molecular-weight inhibitors that react with the active site of thrombin (dabigatran and argatroban).[3]

Bivalirudin is approved for use in coronary angioplasty (Chapter 73) and reduces the risk for major bleeding in comparison with heparin. Argatroban and lepirudin are approved in patients with heparin-induced thrombocytopenia (Chapter 172). Data comparing argatroban with hirudin in heparin-induced thrombocytopenia are too limited for conclusions to be drawn about their relative efficacy and safety. Argatroban is metabolized in the liver and can be used in patients with renal failure, whereas the other direct thrombin inhibitors depend on elimination through renal excretion, and they may be used in patients with liver disease.

Dabigatran etexilate is administered orally and metabolized by ubiquitous esterases to dabigatran, a reversible, active-site thrombin inhibitor. The pharmacokinetic characteristics of dabigatran and the oral factor Xa inhibitors are summarized in Table 38-3. Dabigatran etexilate has been approved in Europe and several other countries for prophylaxis against VTE after hip or knee arthroplasty (first dose, 110 mg; then 220 mg once daily; for patients older than 70 years or with creatinine clearance of 30 to 50 mL/minute, first dose 75 mg and then 150 mg daily). The effect is similar to that of enoxaparin (40 mg subcutaneously daily). Dabigatran (150 mg twice daily) is more effective than warfarin for stroke prophylaxis in atrial fibrillation, whereas a lower dose (110 mg twice daily) is equally effective with lower risk for major bleeding. Both doses resulted in fewer intracranial hemorrhages than warfarin. The 110-mg regimen has not been approved in the United States, but for patients with a calculated creatinine clearance of 15 to 30 mL/minute, the dose of 75 mg twice daily is available. In the treatment of VTE, dabigatran (150 mg twice daily) has comparable effect to warfarin and is at least as safe regarding bleeding. In terms of periprocedural bleeding, dabigatran and warfarin are equivalently safe, and dabigatran facilitates a shorter interruption of oral anticoagulation.

Direct Factor Xa Inhibitors

A number of orally available low-molecular-weight active site-directed factor Xa inhibitors have been designed and are in various stages of clinical development (rivaroxaban, apixaban, betrixaban, edoxaban). Unlike heparins and pentasaccharide, direct inhibitors inactivate factor Xa without the need for AT as a cofactor.

Rivaroxaban, an oxazolidinone derivative (10 mg daily for 30 to 39 days), is more effective than enoxaparin in patients undergoing hip or knee arthroplasty but with a trend to more bleeding. For stroke prophylaxis in atrial fibrillation, rivaroxaban when compared with warfarin showed similar efficacy and safety. In patients with acute coronary syndromes, rivaroxaban, when added to standard antiplatelet therapy, reduced the risk for thromboembolic events at the cost of increased bleeding. In the treatment of VTE, rivaroxaban had similar efficacy to vitamin K antagonists, and in the trial with patients with pulmonary embolism, rivaroxaban reduced the risk for major bleeding.

Apixaban (2.5 mg twice daily) for prophylaxis against VTE after hip or knee arthroplasty has similar or better efficacy than LMWH and similar or lower risk for bleeding. In atrial fibrillation, apixaban (5 mg twice daily) reduces the risk for stroke and for major bleeding compared with warfarin, and it reduces the risk for stroke without any increase of major bleeding in comparison with aspirin. For patients with VTE, apixaban (5 mg twice daily) has similar effect as warfarin but with a reduced risk for major bleeding.

Edoxaban (60 mg daily after initial LMWH) for patients with VTE has similar effect and risk for major bleeding as warfarin. For patients with extensive pulmonary embolism, edoxaban appears to be more effective than warfarin. For patients with atrial fibrillation, edoxaban (60 mg daily) has similar risk for stroke and lower risk for major bleeding compared with warfarin.

All four new anticoagulants have in common that they were associated with a relative risk reduction for death by 10% and for intracranial hemorrhage by approximately 50% compared with warfarin (Table 38-4). In the 2012 ACCP practice guidelines, only dabigatran received a grade A recommendation because there were not sufficient data available for the other anticoagulants. It is, however, difficult to claim that one drug is better than the other based solely on indirect comparisons. A recent review of the combined results of phase 3 trials comparing direct oral anticoagulants (dabigatran, rivaroxaban, apixaban, or edoxaban) with vitamin K antagonists in the treatment of patients with acute symptomatic VTE concluded that direct oral anticoagulants have similar efficacy but significantly reduce the risks of major bleeding compared with vitamin K antagonists. None of the new oral anticoagulants have, however, demonstrated equivalent or superior efficacy and safety compared with traditional anticoagulants in thromboprophylaxis for patients with mechanical heart valves, and therefore their use in this setting is not recommended at this time.

Platelet-Active Drugs

The platelet-active drugs inhibit different steps in either platelet activation (aspirin, ticlopidine, clopidogrel, prasugrel, ticagrelor, cilostazol, and dipyridamole) or platelet recruitment (GPIIb-IIIa antagonists abciximab, tirofiban, and eptifibatide) (Fig. 38-3).

Aspirin and Other Cyclooxygenase Inhibitors

Mechanism of Action and Pharmacology

Aspirin permanently inactivates cyclooxygenase isoenzymes (COX-1 and COX-2) that catalyze the conversion of arachidonic acid to prostaglandin H_2, a precursor of a variety of eicosanoids, including thromboxane A_2 in platelets and prostacyclin (prostaglandin I_2) in vascular endothelial cells (Chapter 37).

Aspirin is rapidly absorbed in the stomach and upper intestine, attaining peak plasma levels about 30 minutes after ingestion; it has a half-life of about 15 minutes. Inhibition of platelet function is evident by 1 hour with uncoated aspirin but can be delayed after administration of enteric-coated aspirin. Therefore, if only enteric-coated tablets are available when a rapid effect is required, the tablets should be chewed.

Aspirin potentiates the antithrombotic effects of warfarin (in high-risk subjects), dipyridamole (in those with ischemic stroke), clopidogrel (in those

TABLE 38-3 PHARMACOKINETIC CHARACTERISTICS AND DRUG INTERACTIONS OF THE ORAL THROMBIN AND FACTOR Xa INHIBITORS

AGENT	BIOAVAILABILITY (%)	TIME TO PEAK PLASMA CONCENTRATION (hr)	PATHWAYS OF ELIMINATION	PLASMA HALF-LIFE (hr)	PLASMA PROTEIN BINDING (%)	MECHANISMS FOR DRUG INTERACTIONS
Dabigatran	6	2	80% renal excretion active	14-17	35	Induction or inhibition of P-gp
Rivaroxaban	80	2-3	⅓ renal excretion active ⅓ metabolized, renal excretion inactive ⅓ liver metabolism, fecal excretion	7-11	92-95	Induction or inhibition of P-gp Induction or inhibition of CYP3A4
Apixaban	50	3	Renal and fecal excretion, oxidative metabolism	8-14	87	Induction or inhibition of P-gp Induction or inhibition of CYP3A4
Edoxaban	50	1-2	⅓ renal excretion active, ⅔ fecal excretion	8-10	40-59	Induction or inhibition of P-gp

P-gp = P-glycoprotein.

TABLE 38-4 ABSOLUTE ANNUAL RISK REDUCTION* OF MAIN CLINICAL OUTCOMES WITH THE ORAL THROMBIN AND FACTOR Xa INHIBITORS COMPARED WITH WARFARIN IN PATIENTS WITH ATRIAL FIBRILLATION

AGENT	ISCHEMIC STROKE (%)	ALL-CAUSE MORTALITY (%)	MAJOR BLEEDING (%)	INTRACRANIAL BLEEDING (%)	GASTROINTESTINAL BLEEDING (%)
Warfarin	Reference	Reference	Reference	Reference	Reference
Dabigatran 110 mg	+0.14	−0.38	**−0.65**	**−0.26**	+0.10
Dabigatran 150 mg	**−0.28**	−0.47	−0.25	**−0.28**	**+0.49**
Rivaroxaban	−0.08	−0.4	+0.2	**−0.18**	**+0.61**
Apixaban	−0.08	**−0.42**	**−0.96**	**−0.47**	−0.10
Edoxaban 30 mg	**+0.52**	**−0.55**	**−1.82**	**−0.31**	**−0.41**
Edoxaban 60 mg	0.0	−0.36	**−0.68**	**−0.21**	**+0.28**

*Positive numbers correspond to absolute increase in risk. Statistically significant differences are in bold.

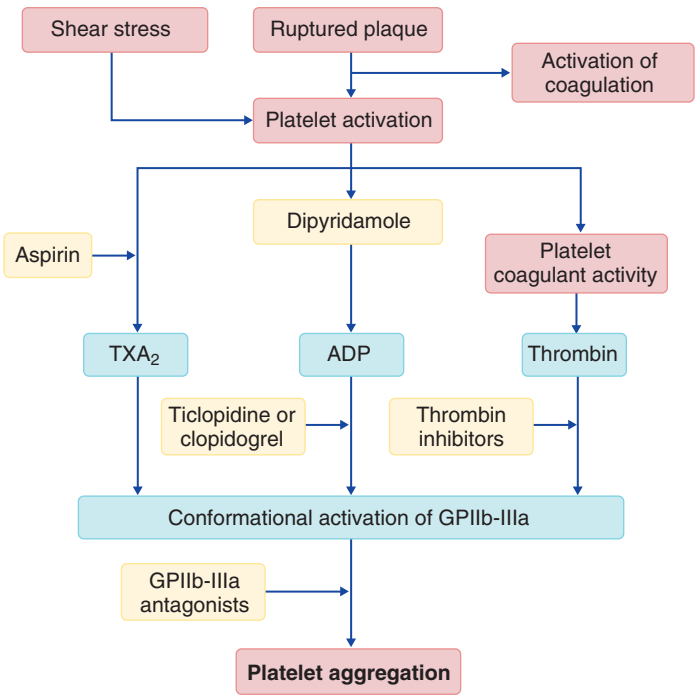

FIGURE 38-3. Sites of action of platelet inhibitors. ADP = adenosine diphosphate; GPIIb-IIIa = glycoprotein IIb/IIIa complex; TXA_2 = thromboxane A_2.

with coronary stents or acute myocardial ischemia), and heparin (in the prevention of recurrent miscarriages in pregnant women with antiphospholipid antibody syndrome and in patients with acute coronary ischemia). Aspirin produces a small increase in major bleeding and a very small increase in the risk for cerebral hemorrhage. It also potentiates bleeding when it is added to another antithrombotic agent.

Aspirin causes gastrointestinal side effects that are dose dependent. Aspirin is contraindicated in individuals with active peptic ulcer disease or aspirin-induced asthma or if gastrointestinal side effects are severe.

Clinical Uses

Based on the results of a meta-analysis, there is evidence that aspirin reduces vascular death by approximately 15% and nonfatal vascular events by about 30% in patients with cardiovascular disease.[■] These effects are achieved in patients with silent myocardial ischemia or stable angina,[■] unstable angina,[■] non–ST elevation myocardial infarction,[■] ST elevation myocardial infarction,[■] noncardioembolic ischemic cerebrovascular disease,[■] and peripheral arterial disease.[■] Aspirin is also effective in patients after coronary angioplasty,[■] coronary[■] or peripheral artery stenting,[■] angioplasty,[■] or bypass graft,[■] symptomatic carotid artery stenosis,[■] or for 1 year after coronary artery bypass surgery[■] and in preventing symptomatic coronary events in asymptomatic men and women older than 50 years.[■] Aspirin has a favorable risk-to-benefit ratio for secondary prevention in patients with overt vascular disease, but the risk-to-benefit ratio is marginal when aspirin is used as primary prevention in asymptomatic individuals, even individuals with type 2 diabetes. Aspirin is less effective than oral anticoagulants in the prevention of recurrent stroke in atrial fibrillation.

Phosphodiesterase Inhibitors

Dipyridamole and cilostazol are phosphodiesterase inhibitors, which elevate platelet cyclic adenosine and guanine monophosphate (cAMP and cGMP) levels. They block platelet reactivity and also inhibit vasoconstriction. The most common side effect is headache.

Dipyridamole is a pyrimidopyrimidine derivative with a terminal half-life of 10 hours and elimination primarily by biliary excretion. Favorable results were obtained with a modified-release preparation in combination with aspirin in patients with carotid stenosis[■] and in patients with prior noncardioembolic stroke or transient ischemic attack,[■] in whom the risk for stroke was reduced by 16% with dipyridamole alone and by 37% with aspirin and dipyridamole in combination, compared with placebo.

In a meta-analysis, cilostazol (100 mg twice daily) was shown to reduce vascular events, driven by fewer cerebrovascular events,[■] and in another meta-analysis to improve maximal and pain-free walking distance in patients with intermittent claudication.[■] Cilostazol is contraindicated in patients with congestive heart failure, owing to reports of fatal events with similar drugs.

Thienopyridines

Ticlopidine, clopidogrel, and prasugrel are thienopyridines that inhibit adenosine diphosphate–induced platelet aggregation through the action of their active metabolites at the $P2Y_{12}$ receptor level. The drugs are administered orally, but their onset of action is delayed until their active metabolites are formed. Similarly, recovery of platelet function is delayed until the circulating affected platelets are replaced by newly formed, unaffected platelets.

Based on a better safety profile and equal efficacy, clopidogrel has replaced ticlopidine. Clopidogrel is rapidly absorbed and metabolized, producing inhibition of platelet aggregation as soon as 90 minutes after an oral loading dosing of 300 mg. With repeated daily administration of low doses (75 mg), there is cumulative inhibition of platelet function with a return to normal 7 days after the last dose of clopidogrel.

Clopidogrel is marginally more effective than aspirin in patients who have experienced a recent stroke[■] or recent myocardial infarction[■] and in patients presenting with symptomatic peripheral arterial disease.[■] The additional benefit over aspirin is modest and similar to that observed with ticlopidine (about 10% relative risk reduction). Clopidogrel is also recommended for patients with symptomatic carotid artery stenosis[■] or peripheral artery angioplasty.[■] The combination of clopidogrel and aspirin is also more effective than aspirin alone in patients with unstable angina and non–ST elevation myocardial infarction (20% risk reduction) and in patients with atrial fibrillation, but at a cost of a modest increase in bleeding. The combination of clopidogrel and aspirin is also more effective than aspirin alone in patients who have percutaneous coronary intervention procedures and is recommended for 1 month after insertion of a bare metal stent, and for 3 to 6 months after a drug eluting stent.[■] Clopidogrel appears to be as well tolerated as aspirin.

Prasugrel is more potent and has a more rapid onset of action than clopidogrel. Given as a bolus dose of 60 mg, followed by 10 mg daily, when compared with clopidogrel, prasugrel reduces the absolute risk for cardiovascular death or nonfatal myocardial infarction or stroke by 2.2%, which is partly offset by an absolute increase of serious bleeding of 0.5%.

Ticagrelor, an oral direct and reversible $P2Y_{12}$ receptor inhibitor, provides a faster response than the thienopyridines. In a randomized trial in patients with acute coronary syndromes, ticagrelor was more effective than clopidogrel, providing an absolute risk reduction for cardiovascular death, myocardial infarction, or stroke of 1.9% without any significant increase of major bleeding.

Integrin $\alpha_{IIb}\beta_3$ (GPIIb-IIIa) Receptor Antagonists

The final common pathway of platelet aggregation is mediated by the binding of fibrinogen to the functionally active integrin $\alpha_{IIb}\beta_3$ (GPIIb-IIIa) on the platelet surface. Inhibitors of this process include monoclonal antibodies, synthetic peptides containing Arg-Gly-Asp (RGD) or Lys-Gly-Asp (KGD), and peptidomimetic and nonpeptide RGD mimetics. These compounds are administered intravenously, and they inhibit platelet function by competing with fibrinogen (and von Willebrand factor) for occupancy on the platelet integrin receptor. Abciximab (ReoPro), a mouse-human chimeric 7E3 Fab antibody, inhibits platelet aggregation in a concentration-dependent manner. Platelet function is impaired rapidly after an intravenous bolus of abciximab and gradually recovers over 24 to 48 hours. Tirofiban (MK-383, Aggrastat) is a nonpeptide derivative of tyrosine. It has a plasma half-life of 1.6 hours, and its effect on hemostasis is reversed within 4 hours of stopping treatment. Eptifibatide (Integrilin) is a synthetic disulfide-linked cyclic heptapeptide. It has a rapid onset and offset of action, and its effect on platelet function is reduced by more than 50% after 4 hours.

All three GPIIb-IIIa receptor antagonists are effective intravenous agents in patients undergoing percutaneous coronary interventions (Chapter 74), and tirofiban and eptifibatide are effective in patients with unstable angina or non–ST elevation myocardial infarction. The GPIIb-IIIa receptor antagonists are administered in combination with heparin and aspirin. Orally active nonpeptide GPIIb-IIIa inhibitors have been developed for long-term use, but the results of clinical trials have been disappointing.

Fibrinolytic Agents

Fibrinolytic agents convert plasminogen to the enzyme plasmin, which then degrades fibrin to soluble fragments, thereby lysing the thrombus. Of the available fibrinolytic agents, streptokinase and urokinase are not fibrin specific; in contrast, recombinant tissue-type plasminogen activator (rt-PA, alteplase) and the rt-PA variant tenecteplase are relatively fibrin specific (Chapter 73).

Streptokinase is an indirect fibrinolytic agent. It binds to plasminogen, converting it into a plasmin-like molecule that in turn converts plasminogen to plasmin. Streptokinase has a number of disadvantages. It is antigenic, rendering its repeated use problematic, and allergenic, producing chills, fever, and rigors in some patients and, in rare instances, anaphylaxis. Anistreplase (APSAC) is an acylated complex of streptokinase and Lys-plasminogen. Compared with streptokinase, it is more fibrin specific, has a longer plasma half-life, and is inactive until it is selectively activated by deacylation on the fibrin surface. Its side-effect profile, antigenicity, and efficacy are similar to those of streptokinase.

Urokinase is a naturally occurring plasminogen activator that differs from streptokinase in that it directly activates plasminogen and is not antigenic. Urokinase was used extensively to treat peripheral vascular occlusions, but production problems have curtailed its availability.

In its natural state, tissue plasminogen activator is produced by vascular endothelium; rt-PA (alteplase) is produced by recombinant DNA technology. Alteplase is not antigenic or allergenic, and it has greater fibrin specificity than streptokinase. It has a short half-life of about 3.5 minutes and therefore is given as a continuous intravenous infusion.

Truncated forms of rt-PA have been developed; the first was reteplase (r-PA), a single-chain deletion mutant that lacks certain domains. As a result, its half-life is about twice that of rt-PA, permitting double-bolus therapy 30 minutes apart. r-PA has lower affinity for fibrin than does rt-PA, but fibrinogen depletion with r-PA is less than that with streptokinase. No antigenicity has been reported with this compound.

Tenecteplase (TNK-tPA) is a mutant tissue plasminogen activator with amino acid substitution at three sites. Compared with rt-PA, it has a longer half-life, allowing single-bolus administration, increased fibrin specificity, and increased resistance to inhibition by plasminogen activator inhibitor 1.

Clinical Uses

Thrombolytic therapy reduces mortality in patients with ST-elevation myocardial infarction and for whom primary percutaneous coronary intervention is unavailable, as well as in patients with extensive pulmonary embolism and hypotension. Treatment with rt-PA intravenously within 3 hours from onset of symptoms of ischemic stroke increases the likelihood of good functional outcome despite a small increase in the risk for intracerebral hemorrhage. Thrombolysis can be justified for severe pulmonary embolism without contraindications, but thrombolysis in patients with moderate-risk pulmonary embolism does not improve outcomes despite better hemodynamics, because of increases in major bleeding and stroke.

SUMMARY OF GRADE A RECOMMENDATIONS FOR ANTITHROMBOTIC THERAPY

Grade 1A indicates that experts are certain that benefits do or do not outweigh risks, burdens, and costs. The reduction in number of grade A recommendations compared with previous editions is due to several changes in the development of the American College of Chest Physicians (ACCP) guidelines. In particular, (a) the GRADE methodology was strictly adhered to, (b) much less importance was paid to surrogate outcomes, and (c) intellectual conflicts were taken into account. Recommendations for the Prevention and Treatment of Venous Thromboembolic Disease that were not considered grade 1A are cited in the General References.[4-9]

ATRIAL FIBRILLATION, INCLUDING PAROXYSMAL (CHAPTER 64)

Patients with Previous Ischemic Stroke, Transient Ischemic Attack (TIA) or At Least Two Other Risk Factors For Stroke

- Standard approach: oral anticoagulation with dabigatran, 150 mg twice daily, or warfarin (INR range, 2.0-3.0) (grade 1A)

VALVULAR AND STRUCTURAL HEART DISEASE (CHAPTER 75)

Mitral Valve Disease

- Rheumatic mitral valve disease with atrial fibrillation, previous systemic embolism, or left atrial thrombus: warfarin anticoagulation (INR range, 2.0-3.0) (grade 1A)
- Mitral stenosis with left atrial thrombus before percutaneous mitral balloon valvotomy: warfarin anticoagulation (INR range, 2.0-3.0) until thrombus resolution is documented (grade 1A)

ANTITHROMBOTIC AND THROMBOLYTIC THERAPY FOR ISCHEMIC STROKE (CHAPTER 407)

- Acute ischemic stroke treatment within 3 hours of onset of symptoms, in eligible patients: thrombolytic therapy with IV rt-PA in a dose of 0.9 mg/kg (maximum of 90 mg), with 10% of the total dose given as an initial bolus and the remainder infused during 60 minutes (grade 1A)
- Acute ischemic stroke or TIA within 48 hours: aspirin in a dose of 160 to 325 mg once daily (grade 1A)
- Noncardioembolic ischemic stroke or TIA: long-term treatment with aspirin, 75 to 100 mg once daily, or clopidogrel, 75 mg once daily; or a combination of aspirin and extended-release dipyridamole, 25 mg/200 mg twice daily, or cilostazol, 100 mg twice daily (grade 1A)
- Cryptogenic stroke and patent foramen ovale or atrial septal aneurysm: long-term treatment with aspirin 50 to 100 mg once daily (grade 1A)
- Ischemic stroke or TIA and atrial fibrillation, including paroxysmal: long-term oral anticoagulation with dabigatran, 150 mg twice daily, or warfarin (INR range, 2.0-3.0) (grade 1A)

CORONARY ARTERY DISEASE (CHAPTER 72, 73, 74)

- Established coronary artery disease, including coronary artery stenoses >50% by coronary angiogram, cardiac ischemia on diagnostic testing, 1 year after acute coronary syndrome or after coronary artery bypass graft: long-term aspirin, 75 to 100 mg daily, or clopidogrel, 75 mg daily (grade 1A)
- Elective percutaneous coronary intervention with placement of bare metal stent: dual antiplatelet therapy for 1 month with aspirin, 75 to 325 mg daily, and clopidogrel, 75 mg daily (grade 1A)
- Elective percutaneous coronary intervention with placement of drug eluting stent: dual antiplatelet therapy for 3 to 6 months with aspirin, 75 to 325 mg daily, and clopidogrel, 75 mg daily (grade 1A)

PERIPHERAL ARTERIAL DISEASE (CHAPTER 79)

- Symptomatic peripheral arterial disease: long-term aspirin, 75 to 100 mg daily, or clopidogrel, 75 mg daily (grade 1A)
- Peripheral artery percutaneous transluminal angioplasty with or without stenting: long-term aspirin, 75 to 100 mg daily, or clopidogrel 75 mg daily (grade 1A)
- Peripheral artery bypass graft surgery: long-term aspirin, 75 to 100 mg daily, or clopidogrel, 75 mg daily (grade 1A)
- Symptomatic carotid artery stenosis, including recent carotid endarterectomy: long-term aspirin, 75 to 100 mg daily, or clopidogrel, 75 mg daily, or aspirin with extended-release dipyridamole, 25 mg/200 mg twice daily (grade 1A)

From: Antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2012.

Grade A References

A1. You JJ, Singer DE, Howard PA, et al. Antithrombotic therapy for atrial fibrillation: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e531S-e575S.

A2. Whitlock EP, Sun JC, Fremes SE, et al. Antithrombotic and thrombolytic therapy for valvular disease: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e576S-e600S.

A3. Schulman S, Parpia S, Stewart C, et al. Warfarin dose assessment every 4 weeks versus every 12 weeks in patients with stable international normalized ratios: a randomized trial. *Ann Intern Med*. 2011;155:653-659.

A4. Holbrook A, Schulman S, Witt DM, et al. Evidence-based management of anticoagulant therapy: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e89S-119S.

A5. Kearon C, Akl EA, Comerota AJ, et al. Antithrombotic therapy for VTE disease: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e419S-94S.

A6. Vandvik PO, Lincoff AM, Gore JM, et al. Primary and secondary prevention of cardiovascular disease: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e637S-e668S.

A7. Steg PG, van't Hof A, Hamm CW, et al. Bivalirudin started during emergency transport for primary PCI. *N Engl J Med*. 2013;369:2207-2217.

A8. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2009;361:1139-1151.

A9. Patel MR, Mahaffey KW, Garg J, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med*. 2011;365:883-891.

A10. Buller HR, Prins MH, Lensing AW, et al. Oral rivaroxaban for the treatment of symptomatic pulmonary embolism. *N Engl J Med*. 2012;366:1287-1297.

A11. Granger CB, Alexander JH, McMurray JJ, et al. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2011;365:981-992.

A12. Connolly SJ, Eikelboom J, Joyner C, et al. Apixaban in patients with atrial fibrillation. *N Engl J Med*. 2011;364:806-817.

A13. Agnelli G, Buller HR, Cohen A, et al. Oral apixaban for the treatment of acute venous thromboembolism. *N Engl J Med*. 2013;369:799-808.

A14. The Hokusai-VTE Investigators. Edoxaban versus warfarin for the treatment of symptomatic venous thromboembolism. *N Engl J Med*. 2013;369:1406-1415.

A15. Giugliano RP, Ruff CT, Braunwald E, et al. Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2013;369:2093-2104.

A16. van Es N, Coppens M, Schulman S, et al. Direct oral anticoagulants compared with vitamin K antagonists for acute venous thromboembolism: evidence from phase 3 trials. *Blood*. 2014;124:1968-1975.

A17. Lansberg MG, O'Donnell MJ, Khatri P, et al. Antithrombotic and thrombolytic therapy for ischemic stroke: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e601S-636S.

A18. Alonso-Coello P, Bellmunt S, McGorrian C, et al. Antithrombotic therapy in peripheral artery disease: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e669S-e690S.

A19. Robless P, Mikhailidis DP, Stansby GP. Cilostazol for peripheral arterial disease. *Cochrane Database Syst Rev*. 2008;1:CD003748.

A20. Eikelboom JW, Hirsh J, Spencer FA, et al. Antiplatelet drugs: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e89S-119S.

A21. Meyer G, Vicaut E, Danays T, et al. Fibrinolysis for patients with intermediate-risk pulmonary embolism. *N Engl J Med*. 2014;370:1402-1411.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Schulman S. Advantages and limitations of the new anticoagulants. *J Intern Med*. 2014;275:1-11.
2. Weitz JI, Eikelboom JW, Samama MM. New antithrombotic drugs: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e120S-51S.
3. Cowell RP. Direct oral anticoagulants: integration into clinical practice. *Postgrad Med J*. 2014;90: 529-539.
4. Guyatt GH, Eikelboom JW, Gould ML, et al. Approach to outcomes measurement in the prevention of thrombosis in surgical and medical patients: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e185S-e194S.
5. Kahn SR, Lim W, Dunn AS, et al. Prevention of VTE in nonsurgical patients: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e195S-e226S.
6. Gould MK, Garcia DA, Wren SM, et al. Prevention of VTE in nonorthopedic surgical patients: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e227S-e277S.
7. Falck-Ytter Y, Francis CW, Johanson NA, et al. Prevention of VTE in orthopedic surgery patients: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e278S-e325S.
8. Bates SM, Jaeschke R, Stevens SM, et al. Diagnosis of DVT: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e351S-e418S.
9. Linkins L-A, Dans AL, Moores LK, et al. Treatment and prevention of heparin-induced thrombocytopenia: antithrombotic therapy and prevention of thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e495S-e530S.

REVIEW QUESTIONS

1. Based on the half-life and onset of action of warfarin, which of the following is the optimal management of this drug in the case of uncomplicated major surgery?

A. Warfarin should be stopped 5 days before surgery and restarted as soon as the patient can take oral medications after surgery.
B. Warfarin should be stopped 2 days before surgery and restarted as soon as the patient can take oral medications after surgery.
C. Warfarin should be stopped 5 days before surgery and restarted at least 5 days after surgery.
D. Warfarin should be stopped 2 days before surgery and restarted at least 5 days after surgery.

Answer: A With a half-life of 40 hours, warfarin has to be stopped 5 days before surgery to eliminate the anticoagulant effect. It takes 5 to 7 days for warfarin to achieve therapeutic anticoagulant effect, and thus it can be started very shortly after surgery provided that there is no bowel paralysis or active bleeding. (Douketis JD, Spyropoulos AC, Spencer FA, et al. *Chest.* 2012;141:e326S-e350S.)

2. For a patient requiring treatment for pulmonary embolism but with a high risk for bleeding, for whom quick elimination of the anticoagulant effect is desirable, which one of the heparins is preferable?

A. Unfractionated heparin (as intravenous infusion)
B. Fondaparinux
C. Danaparoid
D. Low-molecular-weight heparin

Answer: A With a half-life of 1 hour at therapeutic concentration, unfractionated heparin is the heparin that will be eliminated fastest. Low-molecular-weight heparins have half-lives of 2 to 3 hours, and fondaparinux and danaparoid about 20 hours.

3. For which one of the new oral anticoagulants would screening with a thrombin time analysis be sensitive to identify clinically important plasma concentrations of the drug?

A. Dabigatran
B. Rivaroxaban
C. Apixaban
D. Edoxaban
E. Argatroban

Answer: A Dabigatran and argatroban are direct thrombin inhibitors, for which a thrombin time is a sensitive test, but argatroban is not an orally available drug. The others are factor Xa inhibitors, for which thrombin time is less sensitive.

4. Which combinations of antiplatelet therapy are recommended because they provide better effect than single-antiplatelet therapy?

A. Aspirin + dipyridamole in noncardioembolic stroke and aspirin + clopidogrel in unstable angina/non–ST elevation myocardial infarction.
B. Aspirin + clopidogrel in noncardioembolic stroke and ticagrelor + clopidogrel in unstable angina/non–ST elevation myocardial infarction.
C. Aspirin + dipyridamole in noncardioembolic stroke and aspirin + clopidogrel in peripheral arterial disease.
D. Aspirin + dipyridamole in peripheral arterial disease and aspirin + ticagrelor in unstable angina/non–ST elevation myocardial infarction.
E. Aspirin + clopidogrel in peripheral arterial disease and aspirin + dipyridamole in cardioembolic stroke.

Answer: A In noncardioembolic stroke, aspirin plus dipyridamole provides a 37% reduction of recurrent stroke compared with 16% with dipyridamole alone. In unstable angina and non–ST elevation myocardial infarction, the combination of aspirin plus clopidogrel provides 20% greater risk reduction than aspirin alone.

5. What is the objective of thrombolytic therapy in patients with acute ischemic stroke (treatment within 3 hours from onset of symptoms)?

A. To improve the functional outcome
B. To reduce mortality
C. To reduce cardiovascular death
D. To eliminate the need for anticoagulation
E. To shorten the duration of hospitalization

Answer: A Treatment with tissue plasminogen activator within 3 hours of ischemic stroke is associated with a significant increase in the proportion of patients with good functional outcome. Even if duration of hospitalization may thereby be shortened, this is not the objective of the treatment. Thrombolysis does not reduce mortality and does not eliminate the need for ensuing anticoagulation in case of cardioembolic stroke.

39

COMPLEMENTARY AND ALTERNATIVE MEDICINE

ADAM PERLMAN

The National Institutes of Health's National Center for Complementary and Alternative Medicine (NCCAM) has defined complementary and alternative medicine (CAM) as " a group of diverse medical and health care systems, practices, and products that are not generally considered to be part of conventional medicine." The use of CAM by the general public has continued to grow. In the United States, approximately 12% of children and 38% of adults are using some form of CAM, and when the use of megavitamins as well as the use of prayer specifically for health reasons is added, the number increases to 62%. Use of CAM is higher in women and those with more education, but CAM use cuts across all socioeconomic levels, races, and ethnicities. In certain populations, such as patients with cancer or rheumatologic conditions, use of CAM can be significantly higher. In one study, 75% of cancer patients surveyed had used at least one CAM modality, and 58% of those using CAM initiated use after they were diagnosed.[1]

Perhaps motivated by growing patient interest in an era of increased consumerism in health care or frustration with the current evolution of health care, many conventional providers have developed interest and expertise in the integration of CAM into patient care. However, it is important to differentiate between integrating CAM and another growing field within health care, integrative medicine. Integrative medicine has been defined as "an approach to care that puts the patient at the center and addresses the full range of physical, emotional, mental, social, spiritual, and environmental influences that affect a person's health. Employing a personalized strategy that considers the patient's unique conditions, needs, and circumstances, it uses the most appropriate interventions from an array of scientific disciplines to heal illness and disease and help people regain and maintain optimum health."[2] Integrative medicine and integrative medicine programs are increasingly prevalent within the academic medical community. Formed in 1999, the Consortium of Academic Health Centers for Integrative Medicine now includes more than one third of all Academic Health Centers in North America.

Many of the principles as defined in Table 39-1 are not unique to integrative medicine, and interest in them has been increasing as a part of the evolving transformation of the U.S. health care system. This has led to a growing interest in integrative medicine as well as CAM and the need for physicians and all health care providers to have, at a minimum, a basic understanding of CAM. Health care providers must be comfortable engaging in a dialogue with patients about their potential use of CAM and the evidence base or lack thereof for the more popular CAM modalities as well as any potential safety concerns.

Most recently, NCCAM has used the term *complementary health approaches* to describe the practices and products that are studied as a part of NCCAM's research portfolio. In general, that portfolio can be separated into two main subgroups: natural products and mind and body practices.

NATURAL PRODUCTS

Natural products, often referred to as dietary supplements, include vitamins and minerals, herbs or botanicals, and a category referred to as nonvitamin, nonmineral natural products. After prayer, use of natural products was the most common complementary health approach among adults surveyed in 2007, with 17.7% of adults having reported using natural products during the previous 12 months. The 2007 National Health Interview Survey (NHIS) also revealed that 83 million U.S. adults spent almost $44 billion dollars out of pocket on visits to CAM practitioners and purchases of CAM products, classes, or materials. Of that out-of-pocked spending, $14.8 billion, or 43.7%, was for nonvitamin, nonmineral natural products (Table 39-2).

Commonly used natural products in adults include such substances as fish oil, glucosamine, and probiotics, and in children, fish oil, probiotics, and Echinacea. There is a growing body of research literature on numerous natural products with mixed conclusions regarding efficacy. As with any substance that has a physiologic effect on the body, many natural products, although typically safe, do have the potential for side effects as well as the potential to interact with medication. Many commonly used dietary supplements, such vitamin E, Ginkgo, and fish oil, can affect platelet function and therefore lead to an increased risk for bleeding. Patients are often unaware of these potential side effects or interactions with medications.

Currently, natural products are regulated under the Dietary Supplement Health and Education Act (DSHEA). Enacted by Congress in 1994, this act gives the U.S. Food and Drug Administration (FDA) the power to regulate both finished dietary supplement products and dietary ingredients. Dietary supplements are defined as products (other than tobacco) that are intended to supplement the diet and include one or more of the following ingredients: a vitamin, a mineral, an herb or other botanical, an amino acid, a substance for use by humans to supplement the diet by increasing the total dietary intake; or a concentrate, metabolite, constituent, extract, or combination of any of the aforementioned ingredients.

Manufacturers are responsible for ensuring that products are safe before bringing them to market, and the FDA is responsible for taking action against any unsafe product after it has reached the market. Although the FDA has a system in effect for the collection and review of adverse effects linked to dietary supplements, that system is voluntary, and concerns have been raised that the agency does not have adequate resources to ensure safety of products currently on the market in a timely and effective way.

TABLE 39-1 DEFINING PRINCIPLES OF INTEGRATIVE MEDICINE

The defining principles of integrative medicine are as follow:

- The patient and practitioner are partners in the healing process.
- All factors that influence health, wellness, and disease are taken into consideration.
- The care addresses the whole person, including body, mind, and spirit in the context of community.
- Practitioners use all appropriate healing sciences to facilitate the body's innate healing response.
- Effective interventions that are natural and less invasive are used whenever possible.
- Because good medicine is based in good science, integrative medicine is inquiry driven and open to new models of care.
- Alongside the concept of treatment, the broader concepts of health promotion and the prevention of illness are paramount.
- Care is individualized to best address the person's unique conditions, needs, and circumstances.
- Practitioners of integrative medicine exemplify its principles and commit themselves to self-exploration and self-development.

Data from Horrigan, B, Lewis, S, Abrams D, et al. *Integrative Medicine in America: How Integrative Medicine Is Being Practiced in Clinical Centers across the United States*. Encinitas, CA: The Bravewell Collaborative; 2012.

TABLE 39-2 USE OF COMPLEMENTARY OR ALTERNATIVE MEDICINE BY U.S. ADULTS IN 2007

MODALITY	PERCENTAGE OF ADULTS WHO USED IT
BIOLOGICALLY BASED THERAPIES	
Herbal or natural products	17.7
Dietary supplements	N/A
Diet-based therapy	3.5
BODY-BASED PRACTICES	
Chiropractic or osteopathic manipulation	8.6
Massage	8.3
Movement therapies*	1.5
MIND-BODY THERAPIES	
Biofeedback	0.2
Hypnosis	0.2
Meditation	9.4
Guided imagery	2.2
Progressive relaxation	2.9
Deep breathing	12.7
Yoga	6.1
Tai chi	1.0
Qi gong	0.3
ENERGY MEDICINE†	
Reiki, biofield, and other therapies	0.5
WHOLE MEDICAL SYSTEMS	
Naturopathy	0.3
Homeopathy	1.8
Ayurveda	0.1
Traditional Chinese medicine (acupuncture)	1.4
Traditional healers	0.4

*Pilates, Trager, Feldenkrais, and Alexander.
†Energy medicine is based on the theory that there are energy fields surrounding and penetrating the human body. Energy therapies are intended to manipulate these energy fields.

DSHEA also allowed for the enactment of regulations to ensure that manufacturers follow good manufacturing practices. In addition, the act clarified which claims are permissible for dietary supplement labels. It does not allow claims that a dietary supplement will "diagnose, prevent, mitigate, treat, or cure a specific disease" but does allow assertions that a dietary ingredient will affect the structure or function of the body. The Federal Trade Commission has responsibility and authority to regulate advertising for dietary supplements.

DSHEA also established the Office of Dietary Supplements (ODS) at the National Institutes of Health. This Mission of ODS is to "strengthen knowledge and understanding of dietary supplements by evaluating scientific information, stimulating and supporting research, disseminating research results, and educating the public to foster an enhanced quality of life and health for the U.S. population."

MIND AND BODY PRACTICES

As defined by NCCAM, mind and body practices "include a diverse group of procedures or techniques administered or taught by a trained practitioner or teacher." Mind and body practices include such therapies as meditation, acupuncture, massage therapy, movement therapy, relaxation techniques, spinal manipulation, tai chi, yoga, and various energy therapies, such as healing touch, Reiki, or qi gong. Mind and body practices as defined by NCCAM should not be confused with the commonly used term *mind-body medicine*. Mind-body medicine is typically used to describe techniques that are specifically designed to enhance the mind's ability to cause physiologic effects that will lead a positive therapeutic outcome, such as decreased pain or anxiety.

Meditation

Meditation, which involves various techniques to self-regulate attention, has been used for thousands of years by various religions and cultures, primarily in Asia, to increase awareness and ultimately improve self-understanding, inner peace, and enlightenment. In Western culture, meditation has gained in popularity since the 1960s and is often used without a religious context to help manage stress and improve overall health.

The physiologic effect of meditation has been extensively studied. Meditation has been shown to increase activity of the autonomic nervous system and bring about what Benson has termed "the relaxation response."[3] This response can lead not only to the subjective sense of decreased stress but also to measurable effects such as a lowering of blood pressure and heart rate. Other investigators have found evidence of increased blood flow in the brain and altered brain chemistry (see Relaxation Techniques, later). Regular meditation is associated with increased α-wave activity as well as decreased levels of hormones associated with stress, such as cortisol and epinephrine, and increased levels of melatonin.

Many meditation techniques exist, and meditation can be taught in individual or group sessions. Meditation has been shown to have potential benefits for managing conditions such as stress, anxiety, cognitive function in elderly people, gastrointestinal disorders, chronic pain, addictions, and even psoriasis. Although some techniques, such as transcendental meditation, have been more extensively studied, evidence comparing the potential effectiveness of various techniques is largely lacking.

Although safe for most patients, limited evidence suggests that meditation should be approached cautiously for anyone at risk for seizures, symptomatic low blood pressure, or psychotic illness. In one small study of meditators involved in an intensive meditation retreat, more than half of the participants experienced at least one adverse effect.

Acupuncture

Practiced in China for more than 5000 years, acupuncture involves the insertion of very fine needles at specific points in the body. These approximately 360 acupoints reside along 14 channels in the body called *meridians*. In Chinese medical theory, the insertion of the needles is intended to stimulate or improve the balance of the flow of "life energy" or *qi* (pronounced chi). Symptoms or disease are thought to be related to a blockage of flow or imbalance of *qi*. Although very different from a Western medical view of pathophysiology, acupuncture has been shown to have various physiologic effects on the body, including stimulation of endorphins and various brain centers.

There is a growing body of research investigating the potential benefits of acupuncture for a number of conditions. To date, there is evidence suggesting that acupuncture may be beneficial for pain from conditions such as dental pain, fibromyalgia, and headache,[4] as well as beneficial in stroke, analgesia during childbirth, and infertility treatment. A meta-analysis suggests that stimulation of the P6 acupuncture point at the wrist is a potentially effective intervention for reducing postoperative nausea and vomiting.[5] Randomized trials evaluating acupuncture for osteoarthritis of the knee show conflicting results, in part depending on study design, but a recent carefully controlled, blinded trial showed no benefit.

It is important to explain to patients interested in a trying acupuncture, that the needles are typically ultra-fine and often not painful. An

acupuncturist's assessment of the patient will determine the exact location of the needles to be placed. Repeat treatment most commonly occurs once a week, and it often requires 8 to 10 treatments to assess whether acupuncture will have a therapeutic effect.

Modern acupuncture using primarily sterile, disposable needles is generally safe. Risk for infection, although rare, does exist, and electro-acupuncture, which involves stimulation of the acupuncture point by passing a very weak electrical current along the needle, should be avoided in patients with electronic implantable devices such as pacemakers.

Massage Therapy

Massage therapy is perhaps one of the oldest healing modalities. Hippocrates is quoted as saying "the physician must be experienced in many things, but most assuredly in rubbing." Massage is most commonly used to relieve pain from musculoskeletal and other conditions as well as to improve function or relieve stress and aid in relaxation. However, there are more than 80 different types of therapeutic massage, and certain techniques may be more beneficial than others for specific conditions or complaints. Massage has a high use and acceptability in the United States, with approximately 18 million U.S. adults receiving massage in 2007.

The exact mechanism by which massage may exert a therapeutic effect is not clear. Massage is reported to improve local circulation, tone of supportive musculature, and joint flexibility. One commonly held belief was that lactic acid build-up led to delayed-onset muscle soreness and that massage removed lactic acid from muscle. Lactic acid is only present substantially during and immediately after high-intensity anaerobic exercise. It is metabolized within 60 minutes after such exercise ceases and converted back to pyruvate for processing in the Krebs cycle to produce further energy. Some research has suggested that massage may impair lactic acid removal from muscle after strenuous exercise by mechanically impeding blood flow. One study found that massage appeared to exert a clinical benefit by reducing inflammation and promoting mitochondrial biogenesis.[6] In a study of 11 young male athletes, massage was found to activate the mechanotransduction signaling pathways, focal adhesion kinase (FAK) and extracellular signal–regulated kinase 1/2 (ERK1/2), potentiate mitochondrial biogenesis signaling (nuclear peroxisome proliferator–activated receptor γ coactivator 1α [PGC-1α]), and mitigate an increase in nuclear factor κB (NFκB) (p65) nuclear accumulation caused by exercise-induced muscle trauma. Massage was also found to decrease the production of the inflammatory cytokines, tumor necrosis factor-α (TNF-α) and interleukin-6 (IL-6), and to reduce heat shock protein 27 (HSP27) phosphorylation, ultimately decreasing the cellular stress resulting from muscle fiber injury. Massage is also believed to decrease emotional stress through activation of the autonomic nervous system, leading to a variety of neuroendocrine effects.

Clinically, massage has been shown to be of potential benefit for a number of conditions, including neck pain, low back pain, constipation, high blood pressure, lymphedema, stress, and depression. A randomized controlled trial of massage for osteoarthritis of the knee found that an 8-week course of massage decreased pain and improved function, with many of the effects persisting for weeks after cessation of treatment.■ Massage appeared to be a viable option as an adjunct to more conventional treatment modalities.

Massage is safe in most settings. Although massage is not entirely risk free, serious adverse events are probably true rarities. Massage should be avoided over rashes, open wounds, fractures, blood clots, or a tumor and is controversial in patients with lymphatic malignancies. Massage can result in increased soreness or bruising and should done with caution in anyone with a bleeding disorder such as thrombocytopenia.

Movement Therapy

Movement therapy is a term used to describe a broad category of approaches that address health and disease by focusing on restoring balance to the body using physical movement. It includes such therapies as yoga, tai chi, Alexander Technique, Feldenkrais Method, and others. Although some therapies involve complex movements and require a trained instructor, others can be self-directed using instructional materials such as videos or books. Yoga and tai chi in particular have gained in popularity in the West and have a growing body of research suggesting positive health benefits.

Originating in India, yoga has been practiced for more than 5000 years. There are hundreds of different types of yoga, which typically involve principles of proper exercise, relaxation, breathing, diet, and meditation and were traditionally practiced to develop one physically, emotionally, and spiritually. In the West, yoga practices have focused on exercise or physical postures (asana), breathing exercises (pranayama), and meditation or relaxation (dharana). Popular forms of yoga in the West include Ashtanga, a vigorous style often taught in the United States as power yoga; Bikram, practiced in rooms heated to about 100° F; Iyengar, a slow form of yoga with strict attention to posture and alignment; Kundalini, a more spiritual form using postures combined with hand positions, breathing, and meditation; and Hatha, which has a focus on postures and breathing exercises to promote a balance of physical health and mental calmness.

Numerous studies have explored the potential health benefits of yoga. A study in the United Kingdom involving 313 adults with chronic low back pain found that yoga led to more improvement in function than usual care when offered once a week for 3 months.■ In a meta-analysis, yoga reduced low back pain and back-specific disability but did not improve overall health-related quality of life compared with usual care, educational programs, and exercise programs.■

When guided by a well-trained instructor, yoga is generally safe for most healthy individuals. People with certain chronic conditions, such as glaucoma, hypertension, neck pain, or sciatica, should modify or avoid certain poses, as should women who are pregnant. Certain forms of yoga may be safer or more appropriate for people with particular conditions. For example, Bikram, or hot yoga, is best avoided in individuals with known heart disease, lung disease, or history of heat stroke. A review of comparison studies of yoga and exercise concluded that yoga may be as effective or superior to exercise in improving a number of health-related outcomes in both healthy and patient populations.[7]

Tai chi, also know as tai chi chuan, is an ancient Chinese practice involving a series of movements coordinated with breathing and practiced to strengthen the physical body, improve mental sharpness, and enhance the flow of energy or qi. The healthy flow of qi or this vital energy is thought to be a critical aspect of maintaining health in traditional Chinese medicine. A gentle form of movement that emphasizes continuous slow, often symmetrical flexion and extension of the upper and lower body, tai chi can often be practiced even by individuals with conditions such as heart disease or arthritis as well as by seniors at risk for falls.

Tai chi has been studied as an adjunct to conventional treatments for a number of conditions. A systematic review of the efficacy of tai chi for mostly healthy seniors found limited evidence that tai chi is effective in decreasing falls or blood pressure. A systemic review of tai chi for osteoarthritis came to a similar conclusion. There was some encouraging evidence to support efficacy, but future trials with larger patient samples and longer treatment periods were needed. However, a randomized trial in patients with fibromyalgia found that a 12-week course of tai chi compared with wellness education and stretching exercises led to significant improvement in pain as well as quality of life.■ Another randomized controlled trial in patients with mild to moderate Parkinson disease found that tai chi, compared with resistance training or stretching, reduced balance impairments, with additional benefits of improved functional capacity and reduced falls.■

As with other movement therapies, tai chi is best practiced under the guidance of a trained instructor, although it can be learned from videos or books. Safe for most populations, guidelines for appropriate practice of tai chi are the same as those for other land-based exercise programs.

Relaxation Techniques

Relaxation techniques involve a broad range of therapies and techniques, including meditation, yoga, and tai chi, which have been practiced for thousands of years for their purported mental, physical, and spiritual benefits. In more modern times, a number of techniques have been developed with the intent of eliciting the "relaxation response." The relaxation response leads to decreased sympathetic nervous system activation and has been shown to increase α waves on electroencephalogram. Through an effect on the limbic system and its influence on the hypothalamic-pituitary-adrenal axis, there is a subsequent slowing of heart rate and respiratory rate, as well as numerous other neuroendocrine effects, including decreased plasma cortisol. Common techniques include progressive relaxation, breathing exercises, guided visualization, biofeedback, and autogenic training.

Although commonly used to control or manage stress, relaxation techniques have been studied for a number of medical conditions, including anxiety, pain, irritable bowel syndrome, diabetes, premenstrual syndrome, tension headaches, and smoking cessation. Relaxation techniques and meditation programs can provide small to moderate reductions in stress from a wide range of conditions. Relaxation techniques can lead to improvement in both acute and chronic pain, but little evidence exists that the improvement is sustained over time.[8]

Relaxation techniques are a safe, typically low-cost option for patients in need of managing stress more effectively or as a part of an overall plan to manage any of a myriad of stress-related conditions. Given the broad diversity of options, it is important for patients to find a technique that feels most comfortable to them based on goals, personality, beliefs, and lifestyle.

Spinal Manipulation

Spinal manipulation is a method based on the belief that misalignment of the spine can have deleterious effects on health. The technique typically involves correction of a subluxation of the spine by applying a sudden force to the vertebrae or other joint while the patient is lying in various positions on an examining table. Chiropractic manipulation is similar to osteopathic manual therapy practiced by osteopathic physicians. However, osteopaths are medical physicians who may or may not use manipulation as a part of their treatment options. Whereas chiropractors may use a range of modalities, their main focus is on manipulation and restoring of proper alignment of the spine. According to the 2007 NHIS study, more than 18 million adults and more than 2 million children had undergone chiropractic or osteopathic manipulation during the previous 12 months.

The efficacy of manipulation has been studied for a number of diverse conditions, with most studies focused on musculoskeletal disorders. A review of the evidence concluded that manipulation was effective for acute, subacute, and chronic low back pain; migraine and cervicogenic headache; cervicogenic dizziness; several extremity joint conditions; and acute or subacute neck pain.[9] The existing evidence was found to be inconclusive for cervical neck pain of any duration, and for mid-back pain, sciatica, tension-type headache, cocydynia, temporomandibular joint disorders, fibromyalgia, premenstrual syndrome, and pneumonia in older adults.

Although minor side effects such as soreness or light-headedness are not uncommon, overall the risk for a serious adverse event is very low. Concern has been raised that manipulation of the cervical spine may put patients at increased risk for vertebrobasilar artery stroke. However, any small increased risk for such a stroke associated with both chiropractic care and visits to primary care physicians may be because such patients were already having headaches or neck pain because of impending or ongoing vertebral artery dissection.

Energy Therapies

Energy therapies include a number of approaches in which the practitioner intends to channel healing energy (typically through the hands) into the person seeking help in order to restore balance of energy in the body and health. The core concept is that all humans have a subtle vital energy or biofield that flows through them and can be manipulated or used to influence health. Examples of therapies that use this concept are therapeutic touch, healing touch, Reiki, qi gong, and intercessory prayer (prayer for an individual that is specifically directed at that person's health).

NCCAM distinguishes two categories of energy therapies or energy medicine: the veritable and the putative. Veritable energy therapies involve energy that can be measured, such as light therapy or electromagnetic radiation (radiation therapy), and are not considered CAM. Putative energy therapies are not measurable in a reliable way and involve theoretical manipulation or modulation of the vital force or biofield as described previously.

Putative energy therapies have been challenged as being nonplausible biologically and as such are perhaps among the more controversial therapies categorized as CAM. Despite that, prayer for health was the most commonly used intervention in the 2007 NHIS report, with approximately 30% of respondents having had others pray for their health. Approximately 1% had used Reiki and 0.5% qi gong.

A review of the literature on energy healing and pain, focused on Reiki, therapeutic touch, and healing touch, concluded that despite interest in these modalities, particularly in the nursing practice literature, few well-conducted studies existed. A 2008 study funded by NCCAM assessed the efficacy of qi gong in the treatment of osteoarthritis of the knee. This randomized controlled trial comparing two qi gong masters and a "sham" master found that although qi gong may have benefit for patients with osteoarthritis of the knee, the two healers were not equivalent.[10] This study points out the challenge of determining the qualification and competency of energy therapy providers for those interested in pursuing energy therapies as a possible therapeutic modality. Despite the lack of definitive research, energy therapies remain popular with patients, and putative energy therapies are generally regarded as offering no measurable risk.

WHOLE MEDICAL SYSTEMS

It is important to recognize that many of the therapies and approaches described previously come from complete systems of healing, or *whole medical systems*. These systems, with their own particular paradigm or way of viewing disease and health, include systems such as traditional Chinese medicine, Ayurvedic medicine, homeopathy, Native American healing, and naturopathy.

CONCLUSION

Although still controversial, use of CAM by both patients and conventionally trained providers has continued to increase. As evidenced by the growth of the Consortium of Academic Health Centers for Integrative Medicine, as well as such events as the Institute of Medicine's 2009 Summit on Integrative Medicine and the Health of the Public, integration and acceptance of the concepts and principles of integrative medicine into the mainstream and academic medical settings has also grown. This is perhaps a result of a realization that many of the principles of integrative medicine, such as a partnering of patients and providers, care that addresses the whole person, and an emphasis on not only treatment but also the broader concepts of health promotion and illness prevention, offer at least part of the solution to the challenges of our evolving health care system. All health care providers, present and future, should be familiar with these concepts, including the safe and appropriate use of a broad range of healing sciences and providers to facilitate the body's innate healing response, relieve suffering, and optimize vitality.

Grade A References

A1. Goyal M, Singh S, Sibinga EM, et al. Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. *JAMA Intern Med*. 2014;174:357-368.

A2. Corbett MS, Rice SJ, Madurasinghe V, et al. Acupuncture and other physical treatments for the relief of pain due to osteoarthritis of the knee: network meta-analysis. *Osteoarthritis Cartilage*. 2013;21:1290-1298.

A3. Hinman RS, McCrory P, Pirotta M, et al. Acupuncture for chronic knee pain: a randomized clinical trial. *JAMA*. 2014;312:1313-1322.

A4. Perlman AI, Sabina A, Williams A, et al. Massage therapy for osteoarthritis of the knee: a randomized controlled trial. *Arch Intern Med*. 2006;166:2533-2538.

A5. Tilbrook HE, Cox H, Hewitt CE, et al. Yoga for chronic low back pain: a randomized trial. *Ann Intern Med*. 2011;155:569-578.

A6. Cramer H, Lauche R, Haller H, et al. A systematic review and meta-analysis of yoga for low back pain. *Clin J Pain*. 2013;29:450-460.

A7. Wang C, Schmid CH, Rones R, et al. A randomized trial of tai chi for fibromyalgia. *N Engl J Med*. 2010;363:743-754.

A8. Li F, Hammer P, Fitzgerald K, et al. Tai chi and postural stability in patients with Parkinson's disease. *N Engl J Med*. 2012;366:511-519.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Perlman A, Lontok O, Huhmann M, et al. Prevalence and correlates of postdiagnosis initiation of complementary and alternative medicine among patients at a comprehensive cancer center. *J Oncol*. 2012;9:34-41.
2. Horrigan B, Lewis S, Abrams D, et al. *Integrative Medicine in America: How Integrative Medicine Is Being Practiced in Clinical Centers across the United States*. Encinitas, CA: The Bravewell Collaborative; 2012.
3. Bhasin MK, Dusek JA, Chang BH, et al. Relaxation response induces temporal transcriptome changes in energy metabolism, insulin secretion and inflammatory pathways. *PLoS ONE*. 2013;8:e62817.
4. Vickers AJ, Linde K. Acupuncture for chronic pain. *JAMA*. 2014;311:955-956.
5. Cheong KB, Zhang JP, Huang Y, et al. The effectiveness of acupuncture in prevention and treatment of postoperative nausea and vomiting—a systematic review and meta-analysis. *PLoS ONE*. 2013;8:e82474.
6. Crane JD, Ogborn DI, Cupido C, et al. Massage therapy attenuates inflammatory signaling after exercise-induced muscle damage. *Sci Transl Med*. 2012;4:119.
7. Ross A, Thomas S. The health benefits of yoga and exercise: a review of comparison studies. *J Altern Complement Med*. 2010;16:3-12.
8. Dunford E, Thompson M. Relaxation and mindfulness in pain: a review. *Br J Pain*. 2010;4:18-22.
9. Bronfort G, Haas M, Evans R, et al. Effectiveness of manual therapies: the UK evidence report. *Chiropr Osteopat*. 2010;18:1-33.
10. Fazzino DL, Griffin MT, McNulty RS, et al. Energy healing and pain: a review of the literature. *Holist Nurs Pract*. 2010;24:79-88.

40

PRINCIPLES OF GENETICS

BRUCE R. KORF

The elucidation of the structure and function of the genome is one of the great scientific triumphs of the 20th century. The relevance of inheritance to health and disease probably has been recognized throughout history, but it is only during the last century that the rules governing inheritance and the mechanisms whereby genetic information is stored and used have come to light. The application of this knowledge to medical practice had long been focused on relatively rare monogenic and chromosomal disorders. Major contributions have been made in these areas in the form of approaches to genetic counseling, genetic testing, prenatal diagnosis, newborn screening, carrier screening, and, to a limited extent, treatment. As important as these contributions are, however, their impact has been limited by the rarity of these disorders. Powerful tools resulting from the sequencing of the human genome are changing this situation (Chapter 43).[1] Genetic factors that contribute to common and rare disorders are being identified, leading to new approaches to diagnosis, prevention, and treatment. Genetics and genomics are increasingly occupying center stage in medical practice, guiding treatment decisions and preventive strategies. This chapter reviews the paradigm whereby genetics is being integrated into the routine practice of medicine.

GENETIC CONTRIBUTION TO DISEASE

It may be argued that no disorder is either completely determined genetically or completely determined by non-genetic factors. Even monogenic conditions, such as phenylketonuria, are modified by the environment, in this case by dietary intake of phenylalanine. Genetically determined host factors are known to modify susceptibility to infection or other environmental agents. Even individuals who are victims of trauma may find themselves at risk in part because of genetic traits that affect behavior or ability to perceive or escape from danger.

Multifactorial Inheritance

Complex traits that are important for both health and disease are the result of an interaction of multiple genes with one another and with the environment (Fig. 40-1). In some cases, individual genes or environmental factors contribute overwhelmingly to the cause of a disorder, as with a genetic condition, such as neurofibromatosis or Marfan syndrome, or an acquired disorder, such as bacterial infection or trauma. Other times, there may be interplay among many factors, making it difficult to dissect out the specific genes or environmental exposures.

From a medical perspective, it is helpful to divide the genetic contribution to disease into three categories: (1) high-penetrance monogenic or chromosomal disorders; (2) monogenic versions of common disorders; and (3) complex, multifactorial disorders. Each of these has an impact on medical practice in distinctive ways.

High-Penetrance Monogenic or Chromosomal Disorders

High-penetrance monogenic or chromosomal disorders are the disorders that most clinicians think of as "genetic conditions" (Chapter 41). They include rare but familiar single-gene disorders, such as neurofibromatosis, Marfan syndrome, and cystic fibrosis, and chromosomal abnormalities, such as trisomy 21 (Down syndrome). Several thousand distinct human genetic disorders have been described and cataloged in *Mendelian Inheritance in Man* (available at www.omim.org). These include mendelian dominant or recessive disorders, sex-linked disorders, and conditions that are due to mutations within the 16.6-kilobase mitochondrial genome. They also include major chromosomal aneuploidy syndromes and syndromes associated with duplication or deletion of small regions of the genome that result in either reproducible syndromes, such as Williams syndrome (deletion of contiguous loci from a region of chromosome 7), or nonspecific intellectual disability or autism spectrum disorder.

ROLE OF THE NONSPECIALIST

Because of the rarity of many of these conditions, most practitioners have limited experience with a given disorder and are likely to need to refer the patient to an appropriate specialist for assistance with diagnosis and management. Nevertheless, the nonspecialist has many distinct roles in the care of these patients. These roles begin with recognition of the fact that the patient may have such a disorder and arrangement for appropriate diagnostic evaluation.[2] Many genetic disorders produce obvious signs or symptoms that at least prompt referral even if they are not immediately suggestive of a diagnosis. Others can be more subtle, with nevertheless significant consequences if the diagnosis is missed. An example is Marfan syndrome (Chapter 260). The physician needs to be alert to the physical characteristics of patients with Marfan syndrome because life-threatening aortic dissection can be avoided with appropriate monitoring and treatment. Table 40-1 lists examples of some adult-onset monogenic conditions with which the internist should be familiar.

Treatment of Patients with Genetic Disorders

The treatment of patients with genetic disorders may require the assistance of a specialist, but the nonspecialist is likely to be the first contact when an affected individual is ill. The primary care physician needs to be familiar with the disorder and major potential complications. For example, the patient with neurofibromatosis who experiences chronic back pain may be presenting with a malignant peripheral nerve sheath tumor, requiring more aggressive evaluation than would be typical for an unaffected individual with back pain. Formation of a good working relationship between the specialist and nonspecialist is crucial to ensure effective care.

The nonspecialist also has an important role in supporting the patient and helping to explain the difficult choices that may be offered for management. This includes providing support for patients who have disorders that cannot be treated and for the emotional impact that accompanies knowledge that a disorder may be transmitted to one's offspring or shared with other relatives. Most patients have little understanding of the mechanisms of genetics and genetic disease. Although the responsibility to explain these issues may reside with specialists and counselors, the primary care physician has an important supportive role.

ADVANCES IN GENETICS

Many of the disorders in this group have been known for a long time, but more recent advances in genetics have had a substantial impact on approaches to diagnosis and management. Genetic testing has been refined with the advent of molecular diagnostic tests that detect mutations within individual genes. Even rare disorders may be amenable to diagnostic testing; a database of testing laboratories can be found on the Internet (available at www.genetests.org or at www.ncbi.nlm.nih.gov/gtr/). Whole-genome scanning using cytogenomic microarrays is revealing small deletions or duplications in patients with disorders such as autism spectrum disorder, for whom standard chromosomal analysis had previously been unrevealing. Sequencing of the entire coding region of the genome ("*whole exome sequencing*") or the entire genome itself ("*whole genome sequencing*") is now being applied clinically.[3] (Actually, these tests do not detect every possible gene or DNA base, hence the use of the word "whole" is disputed, but the techniques do look across the entire genome, so the word "whole" distinguishes these from approaches that target specific genes.) Population screening for carrier status for autosomal recessive disorders has been offered for many years, with specific ethnic groups being offered testing for conditions of high prevalence in the group. Genomic approaches are now making it possible to vastly expand the scope of testing, increasing the numbers of conditions tested and making it possible to offer a similar comprehensive screen of dozens or even hundreds of genes regardless of ethnicity. Prenatal screening for trisomy can be offered noninvasively by sequencing of fetal DNA isolated from maternal blood. Newborn screening is being expanded beyond inborn errors of metabolism such as phenylketonuria and galactosemia, with the advent of tandem mass spectrometry and the availability of a standardized panel of tests.

Finally, treatment of some monogenic disorders is becoming feasible. Life expectancy for patients with cystic fibrosis has been increasing gradually with better treatments for chronic lung disease; dietary therapy is available for many inborn errors of metabolism; novel therapies that use either pharmaceuticals or gene or enzyme replacement strategies are in use or being tested for many conditions. The principles of management of genetic disorders are evolving rapidly, and care of patients increasingly requires active partnership of specialists and primary care physicians. Moreover, individuals with congenital disorders such as Down syndrome are routinely surviving to adulthood and require primary care physicians who are familiar with their special needs.

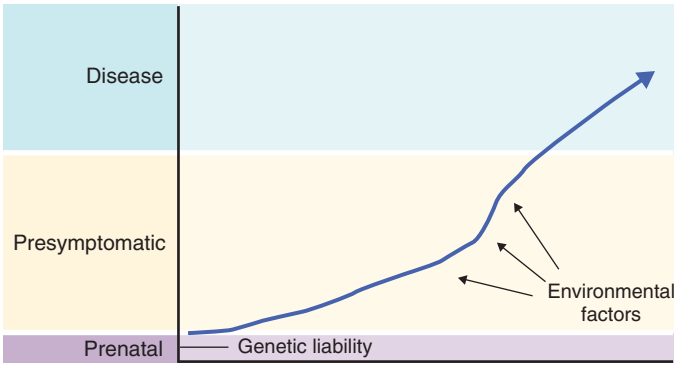

FIGURE 40-1. Multifactorial etiology of disease. An individual is born with a genetic liability but remains in a presymptomatic state for some time until additional events occur, including exposure to environmental factors, that result in crossing a threshold that is identified as *disease*. In instances of high-penetrance monogenic disorders, the genetic liability may be overwhelming. In other instances, genetic factors may contribute only slightly to disease risk.

TABLE 40-1 HIGH-PENETRANCE SINGLE-GENE DISORDERS THAT MAY PRESENT IN ADULTHOOD, WITH SOME MAJOR MEDICAL IMPLICATIONS*

DISORDER	INHERITANCE	MAJOR MEDICAL IMPLICATIONS
CARDIOVASCULAR		
Marfan syndrome	AD	Risk for aortic dissection; lens dislocation
Long QT syndrome	AD, AR	Arrhythmia, sudden death
RENAL		
Adult polycystic kidney disease	AD	Renal failure
PULMONARY		
α_1-Antitrypsin deficiency	AR	Emphysema, cirrhosis
NEUROLOGIC		
NF1	AD	Benign and malignant nerve sheath tumors, gliomas
NF2	AD	Schwannomas (especially vestibular), meningiomas
Von Hippel-Lindau	AD	Hemangioblastoma of cerebellum, brain stem, eye; pheochromocytoma; renal cell carcinoma
Huntington disease	AD	Movement disorder, psychiatric disorder, dementia
HEMATOLOGIC		
Globin disorders	AR	Stroke, iron overload
ENDOCRINE		
MEN syndromes	AD	Tumors of thyroid and parathyroid, pheochromocytoma

*See Table 40-2 for examples of lower penetrance disorders.
AD = autosomal dominant; AR = autosomal recessive; MEN = multiple endocrine neoplasia; NF = neurofibromatosis.

Monogenic Versions of Common Disorders

Not all monogenic disorders produce obscure phenotypes, and not all common disorders have complex multifactorial causes. Some common disorders occur in some families as single-gene traits (Table 40-2). This is usually true for only a proportion of affected individuals, but in some cases, it is a significant proportion and represents an important group of patients to be recognized.

Breast Cancer

An example is breast cancer (Chapter 198). Familial predisposition to breast and ovarian cancer in many cases is attributable to mutation of *BRCA1* or *BRCA2*. Women who inherit a mutation in one of these genes face a high risk for eventually developing breast or ovarian cancer—more than 80% by age 70 years for breast cancer. Women at risk because of mutation do not look different from women with sporadic breast cancer but can be distinguished by many features, including family history of breast or ovarian cancer in multiple relatives, early age at onset of cancer, and multifocality of the cancer (e.g., bilateral breast cancer or breast and ovarian cancer).

TABLE 40-2 SINGLE-GENE DISORDERS WITH INCOMPLETE PENETRANCE THAT MAY ACCOUNT FOR INHERITED FORMS OF SELECTED COMMON DISORDERS

DISORDER	INHERITANCE: GENES	MAJOR MEDICAL IMPLICATIONS
Hemochromatosis	AR: *HFE*	Cirrhosis, cardiomyopathy, diabetes mellitus
Thrombophilia	AD, AR: multiple genes	Deep vein thrombosis
Breast and ovarian cancers	AD: *BRCA1, BRCA2*	Breast and ovarian cancers
Familial adenomatous polyposis	AD: *APC*	Multiple colonic polyps, colon cancer
Lynch syndrome	AD: DNA mismatch repair genes	Colorectal cancer, endometrial cancer
Maturity-onset diabetes of the young	AD: multiple genes	Diabetes mellitus
Cardiomyopathy	AD: genes involved in cardiac contractile apparatus	Arrhythmia, heart failure

AD = autosomal dominant; AR = autosomal recessive.

Colon Cancer and Other Common Disorders

Another example from cancer genetics is colon cancer (Chapter 193). Two syndromes, familial adenomatous polyposis and Lynch syndrome, are autosomal dominantly inherited and convey a high risk for colon cancer. Other noncancer examples are hemochromatosis (Chapter 212), in which cirrhosis, cardiomyopathy, diabetes, joint disease, and other problems ensue from excessive iron absorption; 10% of whites carry an allele that predisposes to this recessive disorder. Mutations in the factor V gene or the prothrombin gene occur commonly and predispose to deep vein thrombosis (Chapter 176). Rarer examples include inherited forms of cardiomyopathy, hypertension, and familial hypercholesterolemia.

Management

The physician may be called on to address these disorders in many ways. There is a compelling reason to make an early diagnosis of hemochromatosis because the complications can be prevented, but not reversed, by phlebotomy and subsequent monitoring of iron stores. Individuals at risk for colon cancer can be offered surveillance with colonoscopy or surgical resection of the colon to reduce the risk for cancer. Individuals at risk for breast and ovarian cancer likewise can be offered surveillance, chemoprevention, or surgery. The benefits of knowledge of genetic risks are less clear in some instances. Carriers of the factor V Leiden mutation would not be treated with anticoagulation until after an event of thrombosis, and the treatment may not be different for a carrier versus a noncarrier. In some cases, however, knowledge of carrier status may help ensure prompt diagnosis or avoid situations of high risk.

Genetic Testing

As with other medical tests, the physician should carefully consider risks, benefits, and clinical utility in deciding to use a genetic test. Some distinct ethical and legal risks may apply to some genetic tests. These may include anxiety, stigmatization, guilt, and possibly discrimination for insurance or employment. Some of these risks may be addressed by legislation to maintain privacy of genetic information, such as the Genetic Information Nondiscrimination Act of 2008, but the risks for anxiety, guilt, and stigmatization cannot be legislated away. To some extent, further research may improve the basis for surveillance or lead to effective treatments. For now, many of these disorders present a double-edged sword of potentially useful knowledge and potentially harmful information.

ROLE OF THE PHYSICIAN

The role of the physician in dealing with monogenic disorders includes recognition of individuals at risk and participation in formulation of a care plan.[4] Individuals at risk may not be identifiable by physical appearance and usually are not evident from medical history or physical examination findings. The most valuable screening tool is the family history. Directed questioning about a family history of major monogenic disorders, especially breast, ovarian, and colon cancer, as well as hypercholesterolemia, hypertension, deep vein thrombosis, cirrhosis, and diabetes, can identify the occasional patient with mendelian segregation of these common disorders. Even if the information is of uncertain reliability, eliciting a family history can prompt referral for further evaluation, documentation of the family history, and consideration for genetic testing. The physician's job is not simply to identify individuals at risk; some people believe they are at high risk even in the absence of well-documented risk factors. Addressing these misconceptions can bring peace of mind and usually does not require genetic testing.

Complex, Multifactorial Disorders

Understanding the genetics of common disorders is one of the great challenges of modern medicine, with the promise of major returns in terms of prevention, diagnosis, and treatment. The etiology of these disorders is complex in that they result from an interaction of multiple genes with one another and with environmental factors. The specific genes that are relevant may be different from one person to the next. Identification of these genes is difficult given this heterogeneity and the relatively small impact that any particular gene may have in a particular person.

POPULATION STUDIES

Dissection of the genetic contribution to common disease cannot be accomplished by the standard genetic approaches involving study of rare variants or family-based linkage studies. Most recent efforts have focused on study of large groups of patients, comparing the prevalence of particular genetic markers in case patients and control subjects. The availability of markers has been boosted by the identification of *single-nucleotide polymorphisms* (SNPs) (Chapter 43). These are differences in single DNA bases between individuals that occur every several hundred bases. Some of these account for common genetic differences between people, including differences that may contribute to disease. The catalog of SNPs currently includes several million variants; it has been found that the genome has evolved as blocks of clusters of genes, making it possible to use only a limited number of SNPs within a given region to determine whether there is a gene in that region that is associated with a disease. Since completion of the HapMap Project, there has been a dramatic increase in the number of SNPs found to be associated with common disorders. For most disorders, however, the total contribution to heritability of the condition has not been accounted for by SNP association studies.

GENETIC RISK ASSESSMENT

The goal of genetic risk assessment is the identification of individuals at risk for disease before the onset of signs or symptoms. In principle, the genetic factors could be identified at birth, or any time in life, by testing a DNA sample. Any individuals found to be at risk might be offered treatment in advance of onset of the disease to avoid complications or might be advised to modify their lifestyle to avoid exposure to environmental factors that might increase risk for disease. Genetic testing has been offered on a direct-to-consumer basis by some companies, although the clinical validity and clinical utility of such testing is a matter of debate.

Although the concept of genomic risk assessment would appear to be an attractive paradigm, many questions may be raised about its practicality and implementation. First, predictive testing is useful only insofar as it guides further management. This is likely to be a moving target because ability to test for risk can be developed more quickly than ability to modify that risk. The utility of interventions may be valued differently by different people. This already has been the case for testing of disorders such as breast cancer. Some women at risk choose not to know their *BRCA* status because the options, including surveillance or prophylactic surgery, are unacceptable to them. If there were a low-cost, safe, and effective treatment that would neutralize any risk the decision to test would be simple, but short of that, there are reasonable arguments on both sides of the issue of whether to test. For many disorders, it will take a long time to show the efficacy of any intervention because there may be a period of many years between the test and the onset of a disorder. Unless surrogate markers can be identified and followed, the task of proving a benefit to predictive testing may require years to decades in some instances.

Predictive Value of Genetic Testing

A second issue surrounds the degree to which genetic testing would be predictive. In patients with suspected genetic conditions, whole exome sequencing currently can make a diagnosis in about 25% of cases.[5,6] Most genetic tests, however, are likely to involve detection of relatively common polymorphic alleles that account for small increments of odds of getting a disease. The predictive value of these tests would be modest, perhaps too low to induce an individual to modify behavior or to take medication. Here, again, much depends on the efficacy of any intervention that can be offered. There may be some disorders for which testing would have substantial predictive value and clinical utility and others for which testing would not be justified.

Social and Ethical Issues

A third concern relates to social and ethical issues. Will people use test results as an excuse to pursue self-destructive behaviors, having received what may be false reassurance of "immunity"? Will people misinterpret results of testing in terms of a simplistic notion of genetic determinism, erroneously believing that their futures have been written, leaving them no recourse but to meet their fate? The rapid pace of technologic change is going to challenge the ability of the social and legal systems to keep pace.

Service Models

Finally, there are questions of the ideal context in which to offer such testing. The personal genomics companies provide their services directly to the consumer in most cases. This creates the obvious risk for incorrect interpretation of results by the patient, although it is not clear that the health care work force is otherwise prepared to deal with the challenges of interpretation of genome-wide studies. The challenges are increasing as the cost of genome sequencing continues to plummet. Genome sequencing also raises the complex question of how to handle incidental findings (i.e., discovery of an unexpected disease risk that may or may not be amenable to medical intervention).

DISEASE STRATIFICATION

A second application of genomics in medical practice entails stratification of disease. Even if genetic testing is not used to predict individuals at risk, it may well be used to determine the most appropriate treatment for a clinically diagnosed disorder. Most common disorders, such as hypertension and diabetes, are symptom complexes that probably result from a variety of causes. The particular combination of causes may differ in different individuals and may respond to different types of treatments. Choice of antihypertensive drug may come to depend on genetic testing to determine the specific cause of hypertension in a patient. The concept of disease stratification is particularly well developed in treatment of cancer, where targeted multigene tests and even genome sequencing is increasingly being used to guide therapy. It is likely that genetic tests eventually will accompany many if not most treatment decisions.

EFFECTS AND IDENTIFICATION OF DRUGS

Aside from helping to choose the most efficacious drug, genetic testing may play a role in avoidance of side effects and in appropriate dosing. Many drugs are known to be associated with rare side effects, some of which are sufficiently severe as to lead the drug to be withdrawn from use. Some of these side effects may occur only in individuals who are susceptible on the basis of having a particular allele at a polymorphic locus. An example is the association of polymorphisms in certain sodium or potassium channel genes with risk for arrhythmia on exposure to specific drugs.

Absorption and metabolism of drugs are largely under genetic control. Several polymorphisms are known to lead to particularly rapid or slow metabolism, accounting for individuals who experience dose-related side effects or lack of efficacy at standard dosages (Table 40-3). Detection of these polymorphisms would allow customization of drug dosage to an individual's pattern of metabolism, increasing the likelihood of efficacy without a prolonged period of trial-and-error dosing.[7]

The greatest gift of genetics and genomics to medicine may be in the ability to identify new drug targets and develop new approaches to treatment. Identification of genes that contribute to common disorders is revealing the cellular mechanisms that lead to disease. This knowledge offers the opportunity to develop new pharmaceutical agents that would target the physiologic mechanisms more precisely, leading to drugs that work better and cause fewer

TABLE 40-3 GENES IN WHICH COMMON POLYMORPHISMS AFFECT RATES OF DRUG METABOLISM OR ACTION

GENE	MEDICATIONS (EXAMPLES)
CYP2C9	Phenytoin, warfarin
CYP2D6	Debrisoquin, β-blockers, antidepressants
VKORC1	Warfarin
UGT1A1	Irinotecan
Thiopurine methyltransferase	Mercaptopurine, azathioprine
N-acetyltransferase	Isoniazid, hydralazine
CYP2C19	Clopidogrel

side effects. New approaches to gene replacement or insertion of genes into cells as localized drug delivery systems also may be developed. The treatment of common disorders likely would entail the use of approaches developed as a result of genomics even in cases in which genetic testing is not used to predict individuals who are at risk.

CONCLUSION

Most physicians in practice today trained before the elucidation of the sequence of the human genome. Nevertheless, physicians will be using the products of the genome project increasingly in their day-to-day practice during the coming years. Whether they are providing care for a patient with a rare genetic disorder or for a patient with a common condition not usually regarded as genetic, management choices increasingly will be informed by tests and treatments that in some way are based on information from the genome sequence.

The essence of the encounter between a physician and a patient can be distilled to two questions: Why this person? Why this time? A person who seeks medical care is doing so as the product of human evolution, having an ancestry associated with certain genetic vulnerabilities, because of inheritance of certain familial risk factors, because of exposure to some environmental factors, because of a particular physiologic process gone awry, because of behavioral traits that lead the person to seek medical care, because of prompting by family or friends to go to the doctor, because society makes medical services available, and because the person can afford to seek care. Genetics cannot answer all of these questions, but it is providing the key to addressing many of the biologic questions that underlie the medical mysteries that have puzzled humankind for generations.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Feero WG. Genomics in medicine: maturation, but not maturity. *JAMA*. 2013;309:1522-1524.
2. Stoppa-Lyonnet D. A guide to cancer genetics in clinical practice. *Eur J Hum Genet*. 2013;21:120.
3. Korf BR, Rehm HL. New approaches to molecular diagnosis. *JAMA*. 2013;309:1511-1521.
4. Manolio TA, Chisholm RL, Ozenberger B, et al. Implementing genomic medicine in the clinic: the future is here. *Genet Med*. 2013;15:258-267.
5. Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA*. 2014;312:1880-1887.
6. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014;312:1870-1879.
7. Dunnenberger HM, Crews KR, Hoffman JM, et al. Preemptive clinical pharmacogenetics implementation: current programs in five United States medical centers. *Annu Rev Pharmacol Toxicol*. 2014;55:89-106.

REVIEW QUESTIONS

1. A couple have a child with autism spectrum disorder in whom chromosomal analysis has been done and was normal. Which of the following is recommended to search for a possible genetic cause?

A. Cytogenomic microarray testing
B. Fluorescence in situ hybridization
C. High-resolution chromosome banding
D. Whole exome sequencing
E. Whole genome sequencing

Answer: A Cytogenomic microarray testing now can detect small copy-number changes (deletions or duplications) that would not be detected by standard cytogenetic analysis, including high-resolution banding.

2. A man has a direct-to-consumer genomic test and is found to have a 1.2 increased odds of type 2 diabetes. The basis for this testing is which of the following?

A. Sequencing a gene known to be involved in type 2 diabetes
B. Single-nucleotide polymorphism genotyping
C. Analysis of RNA sequences in a saliva sample
D. Whole exome sequencing
E. Whole genome sequencing

Answer: B Direct-to-consumer genomic testing is currently based on genotyping of a large number of single nucleotide polymorphisms. (See Population Studies.)

3. A woman requests genetic testing for risk for breast cancer. Which of the following would constitute a risk factor for familial breast and ovarian cancer?

A. History of tobacco use
B. Sister with breast cancer diagnosed at 40 years of age
C. Mother with ovarian cancer diagnosed at 80 years of age
D. Father with history of lung cancer
E. Maternal uncle with colon cancer

Answer: B A first-degree relative with early-onset breast cancer would constitute a risk factor for hereditary breast and ovarian cancer. (See Breast Cancer under Role of the Physician.)

4. A 40-year-old patient with Marfan syndrome seeks advice regarding medical surveillance. Which of the following represents the greatest health risk that may be subject to medical intervention?

A. Dislocation of joints
B. Dislocation of lens
C. Aortic root dilation
D. Cardiomyopathy
E. Skin laceration

Answer: C Aortic root dilation and risk for aortic dissection are the most important life-threatening risks for individuals with Marfan syndrome and are subject to monitoring and intervention. (See Role of the Nonspecialist.)

41

GENE, GENOMIC, AND CHROMOSOMAL DISORDERS

SANDESH C. S. NAGAMANI, PAWEŁ STANKIEWICZ, AND JAMES R. LUPSKI

THE HUMAN GENOME

Unprecedented technologic advances in molecular biology during the past two decades have enabled the determination of the entire DNA (deoxyribonucleic acid) sequence content of the human genome (Human Genome Project, HGP; http://web.ornl.gov/sci/techresources/Human_Genome/index.shtml) and establishment of a reference haploid genome.[1,2] Sequencing of other personal diploid human genomes and international collaborative efforts (http://www.1000genomes.org/) have generated DNA sequence data that have revolutionized our views on human history, evolution, and the genetic and genomic bases of disease.[3]

Human genomic DNA is packaged within the nucleus in 23 chromosome pairs, 22 autosomes, and 2 sex chromosomes, XX in females and XY in males. The diploid genome (2n) in each cell consists of two identical haploid copies of about 3×10^9 base pairs (bp), thus equaling in total 6 billion nucleotides. Most of the human genome consists of repetitive elements: tandem repeats (e.g., satellite sequences in centromeres), telomeric repeats, microsatellites, minisatellites, and short and long interspersed retrotransposable elements (e.g., *Alu* elements and LINE elements, respectively) (Table 41-1). These elements form constitutive heterochromatin, and their functional roles are not yet well elucidated. The unique "coding" DNA sequences comprise the minority of our genome and include about 23,000 protein-coding genes, conserved sequences that encode noncoding RNAs (i.e., not translated to protein), including microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), and long noncoding RNAs (lncRNAs), as well as conserved regulatory elements. Although protein-coding sequences occupy only about 1 to 2% of the human genome, it has been demonstrated recently that most of our DNA may be transcribed into RNA.

Approximately 4 to 5% of the human genome, including both repetitive and unique sequences, is present in two or more copies in the haploid genome. DNA fragments larger than 1 kb and of DNA sequence identity greater than 90% have been termed *low-copy repeats* (LCRs) or *segmental duplications* (SDs). Most LCRs have arisen during primate speciation. A subset of LCRs with DNA sequence identity greater than 95% and longer than 10 kb can lead to local genome instability during both meiotic (constitutional) and mitotic (somatic) cell divisions, resulting in genomic rearrangements and conveying genomic disorders.

GENE

The concept of a gene can be traced back to 1865 when Gregor Mendel observed the inheritance of phenotypic traits in the garden pea, *Pisum sativum*. Mendel noted that two *factors* that we now know to be corresponding DNA loci (alleles) located on homologous chromosomes separate from each other during meiosis and segregate to two different gametes. This phenomenon of independent segregation is now known as *Mendel's first law*. *Mendel's second law* described the independent segregation of two different (nonallelic) loci during gamete formation. The *inheritance factors* or *units of heredity* encoding the genetic information were later defined as *genes*. We now define a *gene* as a fragment of DNA that carries the information used to transcribe it into a functional RNA (ribonucleic acid).

The DNA double helix is built of four nucleotides: two purine bases, adenine (A) and guanine (G), and two pyrimidine bases, thymine (T) and cytosine (C), all connected to deoxyribose sugars and linked by phosphodiester bonds at the 5′ and 3′ carbon of the sugar. (In RNA, thymine is replaced by uracil, U.) Three consecutive nucleotides (triplet codon) of the coding DNA encode an amino acid. There are 64 possible different codons (4^3 combinations) but only 20 amino acids; therefore, the genetic code has been termed *degenerate*. Most of the protein-coding genes in our genome comprise several coding regions or *exons* that are separated by noncoding *introns*. The entire gene (exons and introns) is transcribed into messenger RNA (mRNA) by RNA polymerase II starting from its 5′ end and continuing beyond the poly A recognition signal at the 3′ end. Typically, mRNA begins with a cap and terminates with a polyadenylated (polyA) tail at the 3′ end. In the subsequent process of splicing, the intervening noncoding introns are deleted, and the spliced, mature mRNA is translated into a polypeptide. The polypeptides start at the 5′ end (NH_2 terminus) with a methionine encoded by the AUG triplet. At the 3′ end (COOH terminus), the polypeptides are terminated by one of three terminating codons, UAA, UAG, or UGA (Fig. 41-1).

Micro-RNA (miRNA) (about 22 bp single-stranded RNA), *small nucleolar RNA* (snoRNA) (60 to 300 bp single-stranded RNA), and *long noncoding RNA* (lncRNA) (>200 bp RNA) are transcribed but are not translated. These noncoding RNAs are involved in many important biologic processes. There is evidence that dysregulation of noncoding RNA may have a role in cancer, cardiovascular, neurologic, and developmental disorders.

GENETIC AND GENOMIC VARIATION IN HUMANS

In addition to the Human Genome Project, the International HapMap (http://hapmap.ncbi.nlm.nih.gov), Human Genome Diversity (http://www.stanford.edu/group/morrinst/hgdp.html), ENCODE (http://www

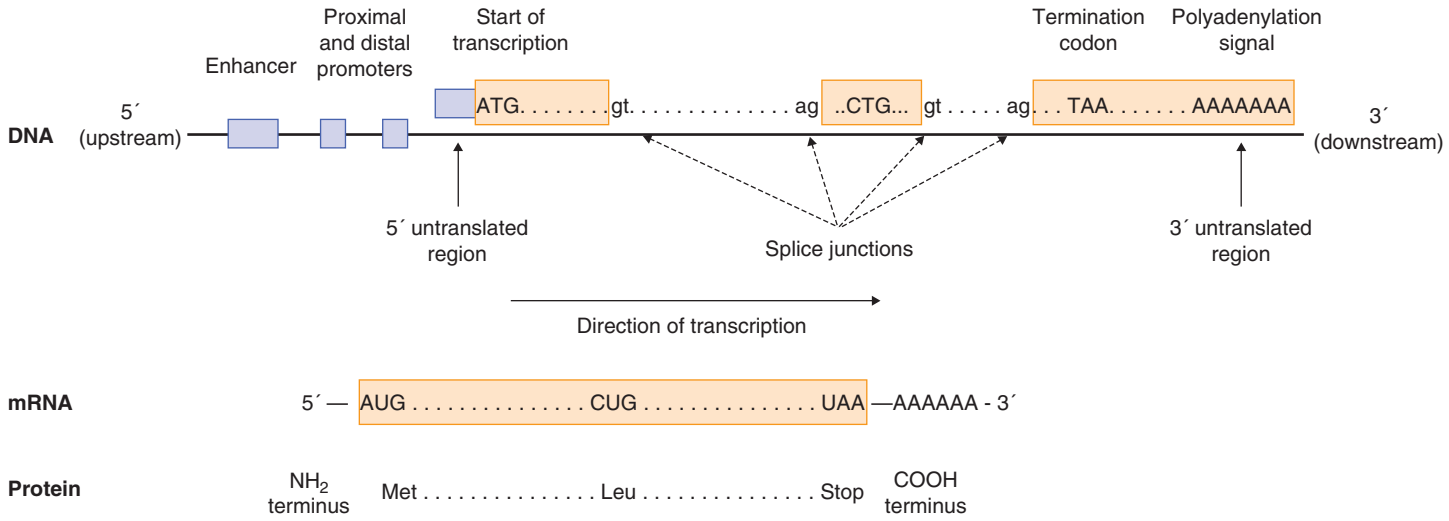

FIGURE 41-1. Gene structure. Schematic representation of the general structure of a typical human gene. Three exons are depicted as open rectangles. Note that the translation usually starts with an ATG triplet encoding methionine. The 5′ (upstream) portion of a gene corresponds to the NH_2 terminus, and the 3′ (downstream) segment encodes the COOH terminus of the polypeptide. Enhancers and promoters are shown as blue rectangles.

TABLE 41-1 STRUCTURE OF THE HUMAN GENOME

CHROMATIN FEATURE	SEQUENCE TYPE	HUMAN GENOME (HAPLOID)	% of Haploid Genome*
Euchromatin	Protein coding	20,000-25,000 genes	~2
	Noncoding	RNA genes	
		Regulatory elements	
		Pseudogenes	~38
		Gene fragments	
		Conserved sequences	
Heterochromatin	Repetitive		~60
		Tandem: satellite DNA, minisatellites, microsatellites	~14
		Interspersed (transposons):	~45
		Retrotransposons	~8
		LTR	
		Non-LTR	
		SINE (*Alu*)	~13
		LINE	~21
		DNA transposons	~3

*Estimated.
LINE = long interspersed nuclear elements; LTR = long-terminal repeat; SINE = short interspersed nuclear elements.

.genome.gov/10005107), 1000 Genomes project (http://www.1000genomes.org/), and other collaborative efforts, including personal genome sequencing projects, have revealed the tremendous and underappreciated extent of variation in the human genome.[3] Genetic variation consists of two major types: (1) nucleotide sequence changes, or single nucleotide variants (SNVs); and (2) genome structural changes, or copy-number variants (CNVs) (Fig. 41-2).

Single Nucleotide Variants

A genetic *polymorphism* is defined as a heterozygous DNA variation present in greater than 1% of the population (http://www.ncbi.nlm.nih.gov/SNP/, http://www.1000genomes.org/, http://evs.gs.washington.edu/EVS/). Genome-wide nucleotide variation uncovered in the early phase of DNA sequencing analyses showed that human genomes differ from the haploid reference genome mainly by single nucleotide changes. These differences have been termed *single nucleotide polymorphisms* (SNPs) and defined as a nucleotide change at a given position generated by substitution. Any two human beings differ on average by about 3.5 million SNPs (0.1% of the 3.0×10^9 reference haploid genome). Whereas most of these SNPs map outside of the exons, on average about 20,000 SNPs occur in coding regions, and among these, about 7000 to 10,000 are nonsynonymous (i.e., change the encoded amino acid).[3] It is important to note that SNPs located outside of the protein-coding regions can still exert phenotypic effects, such as by modifying gene regulatory elements, transcription factor–binding sites, generating splicing mutations, or affecting noncoding RNAs.

A set of consecutive SNPs (or other markers) is defined as a *haplotype*. A nonrandom association of markers in a population not interrupted by meiotic recombination (*crossing over*) is described as *linkage disequilibrium*. (Note that linkage disequilibrium exemplifies the exception to Mendel's second law).

Copy-Number Variants

A more recently characterized group of major polymorphic genetic variation in the human genome is represented by *structural changes*. High-resolution genome-wide analysis of human genome sequences has revealed higher-order architectural features, with a potential to cause genomic instability and extensive submicroscopic structural variations.[4] These structural variations consist of unbalanced CNVs, including deletions, duplications, triplications, insertions, and translocations, that differ from the normal diploid state, as well as balanced rearrangements such as genomic inversions. Recent analyses have revealed that about 11,700 CNVs (size > 443 bp) cover more than 112 million base pairs (Mb) (3.7%) of the reference human genome. A validated subset of these CNVs overlap 13% of the Reference Sequence (RefSeq) (http://www.ncbi.nlm.nih.gov/projects/RefSeq/RSG) genes and 12% of the Online Mendelian Inheritance in Man (OMIM) (http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim) genes and was predicted to alter the structure of 12.5% gene transcripts and 5.5% mRNAs. On average, each individual harbors about 1000 CNVs that range in size between 500 bp and 1.2 Mb; the frequency of smaller CNVs (<1 kb) and indels (insertion or deletion of bases < 100 bp) is much higher than the larger rearrangements. It is noteworthy that any two human genomes contain more base-pair differences due to CNVs than to SNVs.

Despite all these recent achievements, the total number, position, size, gene content, and population distribution of CNVs remain obscure because we still do not have accurate and reliable molecular methods to study smaller

CNVs on a genome-wide scale in different populations, particularly when copy-number changes are greater than $n = 4$ or 5.

CNVs have been shown to be responsible for Mendelian diseases, non-Mendelian traits such as complex diseases, and common traits (including neurobehavioral traits), or to represent benign polymorphic variants (Chapter 40).[5] CNVs can lead to abnormal phenotypes by disrupting the gene structure or changing the copy-number of dosage-sensitive genes. However, long-range effects of CNVs involving nongenic sequences, leaving a gene intact, have been also demonstrated. Furthermore, evidence suggests that a combination of two or more CNVs at the same or different loci may be responsible for phenotypic variation. The genome-wide scale of phenotypic effects exerted by CNVs (genomic load) is unknown and awaits further studies.

A summary of CNVs can be found in the Toronto Database of Genomic Variants (http://projects.tcag.ca/variation). Many clinically relevant CNVs can be found in the Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (DECIPHER) (https://decipher.sanger.ac.uk/information).

Tandem Repeats

Variable number of tandem repeats (VNTR), or *minisatellites*, and short tandem repeats (STRs), such as unstable dinucleotides, trinucleotides, and tetranucleotides—$(GT)_n$, $(CAA)_n$, or $(GATA)_n$—referred to as *microsatellites,* are highly variable. Both minisatellites and microsatellites have been successfully used in linkage and association studies that enable the mapping of traits and the identification of genes and loci responsible for both Mendelian disorders and complex traits. These highly polymorphic sequence repeats are extremely variable in the copy number of their repeating subunits; this property enables the use of a number of such markers to derive a unique pattern of marker genotypes for each human individual. Thus, such markers have been useful in DNA fingerprinting for identity testing and DNA forensics.

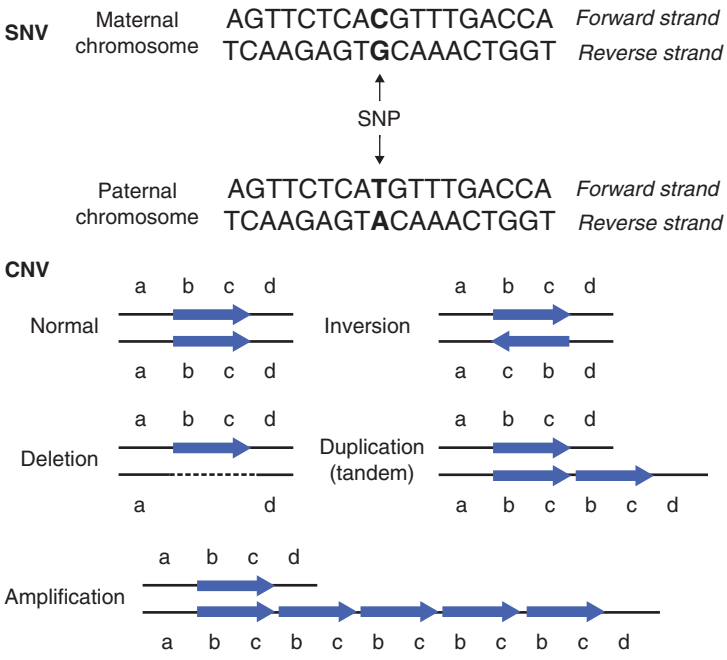

FIGURE 41-2. Genetic variation. (*Upper*) Heterozygous single nucleotide polymorphism (SNP, or single nucleotide variant, SNV) representing the most common transition C→T is shown. (*Bottom*) Structural genomic changes: a balanced inversion and the unbalanced copy-number variants (CNVs), deletion, duplication, and amplification are shown with *blue arrows* on two homologous chromosomes (*black lines*). The *dashed line* represents a deleted fragment of one chromosome.

Repetitive Elements

The other group of polymorphic elements in the human genome is represented by retrotransposons, long and short interspersed nuclear elements (LINEs and SINEs) (see Table 41-1). The most common *Alu* and L1 elements introduce recombinogenic genomic instability and insertional mutagenic activity; their positions within an individual human personal diploid genome can vary tremendously.[6] It has been recently estimated that repetitive elements may represent more than two thirds of the human genome.[7]

CHROMOSOMES

The recombined haploid (1n) human genome formed during meiosis is stored as chromosomes in female and male gametes. They merge at conception, and this diploid genome instructs the development of a zygote; the diploid human genome is subsequently transmitted to the mitotically dividing daughter cells. Human chromosomes can be distinguished from each other in a light microscope by differences in size and characteristic banding patterns after specific chemical staining (e.g., G-banding with Giemsa) when the chromosomes are arrested in a condensed phase (metaphase) of mitotic divisions.

Each human metaphase chromosome is composed of two chromatids that form short (p) and long (q) arms connected by a centromere built with α-satellite DNA. Based on the relative position of the centromere along the chromosome, chromosomes have been described as metacentric (similar-sized p and q arms), submetacentric (q arm significantly longer than p), and acrocentric (chromosomes 13, 14, 15, 21, and 22, with centromeres located close to the end of a chromosome) (Fig. 41-3).

Telomeres consist of repetitive DNA sequences (thousands of copies of TTAGGG repeats) located at the ends of both chromosome arms that are stabilized by a reverse transcriptase enzyme, *telomerase*, which adds TTAGGG sequence to the 3′ end of DNA strands. In contrast to germline and cancer cells, human somatic cells lacking telomerase gradually lose the telomeric sequences. As a result, cells reach the limit of their replicative capacity and fall into senescence.[8]

Chromosomal Aberrations

Microscopically visible chromosomal aberrations have been divided into numerical and structural aberrations and are found in about 1 in 160 live births.

Numerical Aberrations

The numerical aberrations typically result in lethality. Numerical aberrations are classified as either polyploidy (number of chromosomes are in multiples of haploid set of 23 chromosomes) or aneuploidy (with extra or missing chromosomes). Polyploidies such as triploidies (3n), 69,XXX, 69,XXY, and

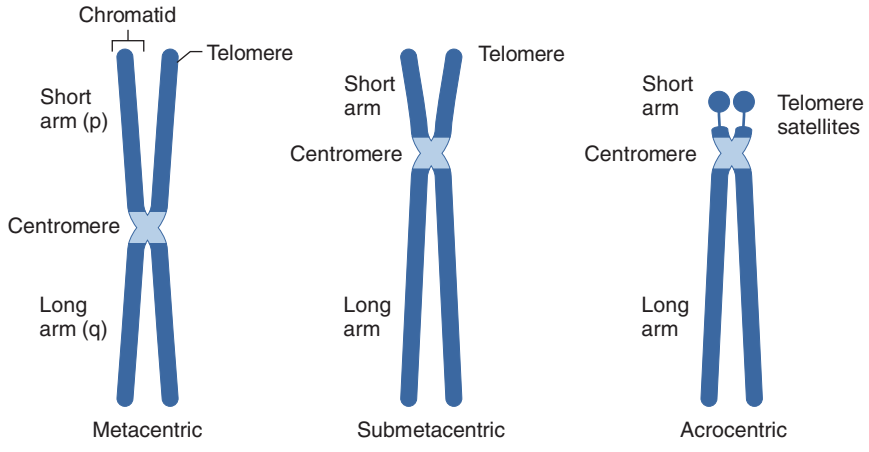

FIGURE 41-3. Types of metaphase chromosomes. Metacentric, submetacentric, and acrocentric chromosomes are composed of two arms connected by a centromere. Each chromosome arm consists of two chromatids.

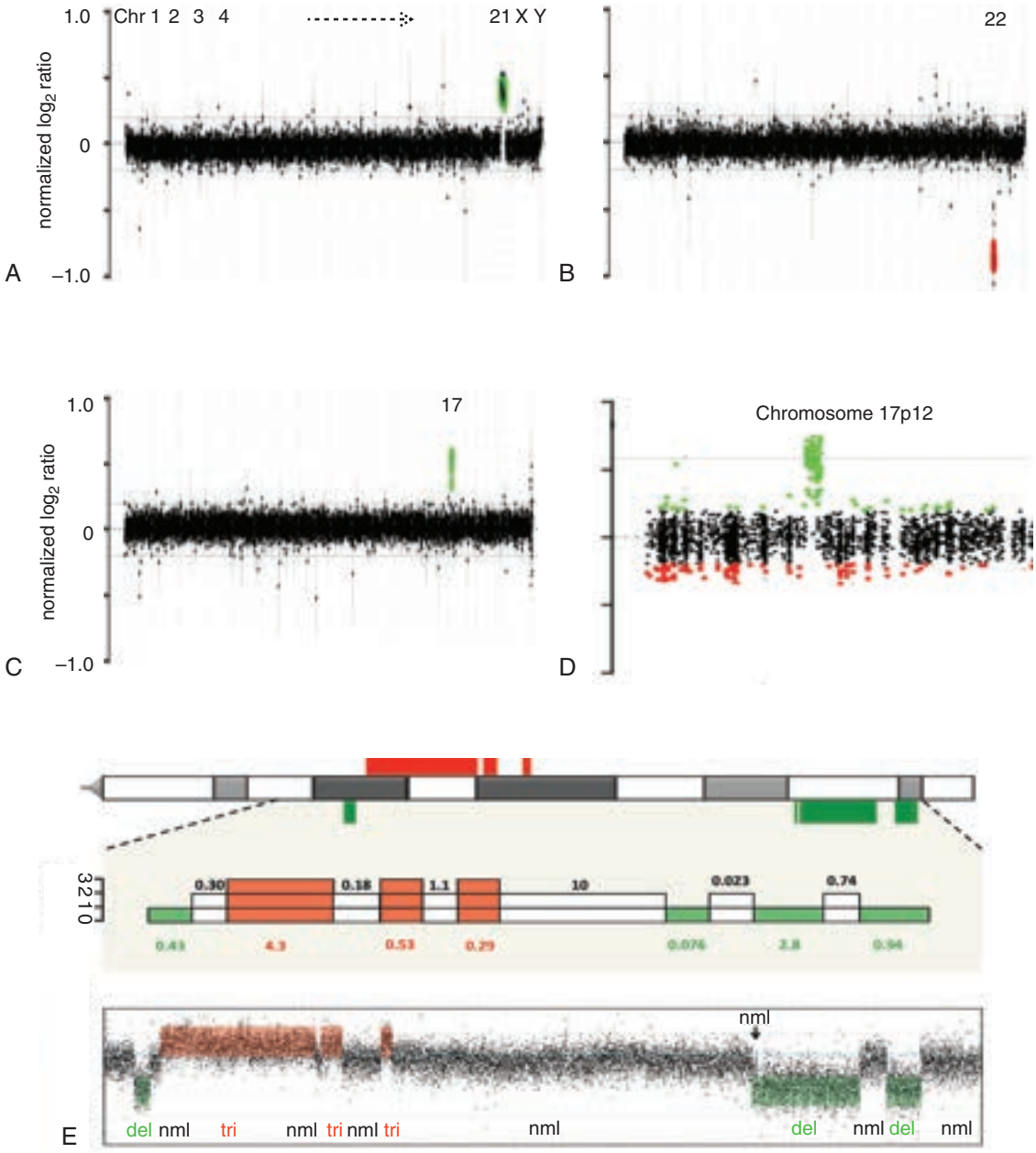

FIGURE 41-4. Assaying copy-number variants (CNVs). Plots of array-based comparative genomic hybridization from patients with genomic rearrangements. Each "dot" on the plot represents an oligonucleotide that interrogates specific regions of the human genome from chromosome 1 (depicted on left) to the sex chromosomes depicted on the right). Gain of copy number is depicted in green, and loss of copy number is depicted in red. A, Patient with Down syndrome showing gain (three copies) of chromosome 21. B, Patient with congenital heart disease and velocardiofacial syndrome with a 3 Mb deletion of chromosome 22q11.2. C, Patient with autosomal dominant sensorimotor neuropathy (CMT1A) with a gain of copy number (duplication) on chromosome 17p12. D, Expanded view of the 17p12 duplication shown in C. E, Complex genomic rearrangement in a pediatric patient with epilepsy and hypotonia showing multiple breakpoints in a single chromosome. The rectangular boxes below the ideogram of the chromosome show copy-number of genomic segments, and the numbers depict the size of the CNVs in Mb. del = Deletion; nml = normal; tri = triplication.

69,XYY, and tetraploidies (4n), 92,XXYY or 92,XXXX, are caused by abnormal fertilization of the egg by two sperms, or by a failure in zygote division, respectively. The most commonly detected viable chromosomal aneuploidies, trisomies and monosomies, involve chromosomes X, Y, 21, 18, and 13 and arise as a result of meiotic nondisjunctions.

Sex chromosome aneuploidies are more common and are found in 1 in 440 newborns. Monosomy X (45,X cell line) in female patients with Turner syndrome is identified in one in every 4000 female newborns. However, this birth rate represents only 1% of all fetuses with 45,X because more than 99% result in miscarriage. (This is similar to the most frequent fetal aneuploidy, trisomy 16, that results in 100% miscarriages.) In most cases, the 45,X cell line is found as a mosaic along with another cell line that has either a normal karyotype or a structural rearrangement of the X chromosome (e.g., deletion of the short arm, ring chromosome, or isochromosome of the long or short arms). One in every 1000 males has a 47,XXY or 47,XYY chromosome complement; the former results in Klinefelter syndrome, whereas the latter typically has mild if any manifestations.

In contrast to gonosomes, monosomies of all autosomes are lethal. The only autosomal trisomies compatible with life are found in patients with Down syndrome (trisomy 21, 1 in every 670 newborns) (Fig. 41-4A), Edwards syndrome (trisomy 18, 1 in 7500 newborns), and Patau syndrome (trisomy 13, 1 in every 22,700 newborns).

Incomplete supernumerary chromosomes are termed *marker chromosomes*. They usually originate from acrocentric autosomes (~50% from chromosome 15) and are found in 1 in 4000 newborns. The severity of the abnormal phenotypes in carriers of marker chromosomes varies among different chromosomes.

Structural Aberrations

Chromosomal deletions and duplications have been categorized as microscopically visible or submicroscopic, terminal or interstitial, recurrent or nonrecurrent. The most frequent are recurrent common-sized rearrangements flanked by directly oriented LCRs or SDs that mediate nonallelic homologous recombination (NAHR). For example, a 3 Mb microdeletion

in chromosome 22q11.2 that results in DiGeorge (velocardiofacial) syndrome is found in 1 in 4000 newborns (Fig. 41-4B); a 1.4 Mb duplication of chromosome 17p12 accounts for greater than half of all adult-onset forms of inherited Charcot-Marie-Tooth (CMT) neuropathies (Fig. 41-4C and D). Evidence suggests that ectopic crossovers that lead to NAHR are preceded by an ectopic synapsis. Nonrecurrent rearrangements are of variable sizes with different breakpoint junctions in each patient reflecting distinct mechanisms of formation that include nonhomologous end joining (NHEJ) and replicative mechanisms such as FoSTeS (fork stalling and template switching) and MMBIR (microhomology-mediated break-induced replication). Whereas most recurrent rearrangements are typically "simple," a significant fraction of nonrecurrent genomic rearrangements with two or more breakpoint junctions referred to as *complex genomic rearrangements* (CGRs) have been implicated in causation of human disease phenotypes (Fig. 41-4E). These CGRs often occur as de novo events, vary in size from those involving a single exon to megabases of genomic sequence, and typically involve loss (i.e., deletion) and/or gain (i.e., duplication, triplication, etc.) in a single or multiple gene loci.[9] An extreme example of CGRs is a chromosome catastrophe wherein several copy-number changes and multiple breakpoints are concentrated on a single chromosome. This phenomenon, termed *chromothripsis*, has been noted in about 3% of all cancers and up to 25% of bone cancers and may portend a more severe disease.[10]

Balanced reciprocal translocations result from an exchange of the DNA material between two chromosomes and are found in about 1 in 600 individuals. During meiosis, the translocation chromosomes form a pachytene tetrad structure, and depending on the segregation type (alternate or adjacent, symmetrical or asymmetrical), either balanced or unbalanced products are transmitted to progeny. The unbalanced products often lead to either spontaneous miscarriage or birth of a child with significant clinical consequences. Recently, it has been shown by high-resolution genome analyses that up to 40% of apparently balanced translocations found in subjects with abnormal phenotypes are associated with additional imbalances at or near the translocation breakpoint or elsewhere in the genome.

Translocations involving short arms (or centromeres) of acrocentric chromosomes are described as *Robertsonian translocations*. Balanced Robertsonian translocations (45 chromosome complement) are present in 1 in 900 newborns and are thus the most common chromosome rearrangements in humans. The most frequent Robertsonian translocation, t(13;14), is found in 1 in 1300 individuals. The carriers of balanced Robertsonian translocations have a significantly increased risk for an unbalanced karyotype in progeny (e.g. trisomy 21 or trisomy 13) or uniparental disomy for chromosomes 14 and 15 that are known to contain imprinted genes.

Constitutional non-Robertsonian chromosomal translocations are nonrecurrent with the exception of three recurrent translocations: t(11;22)(q11.2;q23.3), which utilize AT-rich cruciforms, and t(4;8)(p16;p23) and t(4;11)(p16;p15.2), which are mediated by low-copy repeat gene clusters.

When a fragment of one chromosome is translocated into another chromosome's arm, the aberration is termed an *insertion* or *insertional translocation*. Insertional translocations have been recently shown by high-resolution human genome analyses to occur more than 100 times more frequently than recognized previously. The carrier of a balanced insertion has up to a 50% chance of an abnormal offspring.

An *inversion* is defined as a chromosome fragment that is reversed end to end. Inversions harboring the centromere are termed *pericentric*, and those with breakpoints mapping in the same chromosome arm are termed *paracentric*. Usually, only the products of pericentric inversions (unbalanced terminal deletion of one chromosome arm accompanied by a terminal duplication of the second arm) are found in a progeny. The acentric or dicentric products of paracentric inversions are unstable and thus not transmitted.

Other less common structural chromosomal abnormalities include ring chromosomes, isochromosomes, complex chromosome rearrangements, and heterochromatin variants. Rings arise when two broken ends of the same chromosome fuse. Usually, chromosome material telomeric to the breakpoints is lost and leads to an abnormal phenotype. Rings are commonly unstable mitotically and often form double ring structures. Isochromosomes arise when one part of the chromosome is duplicated and separated from the other. Isochromosomes can be monocentric (breakpoint in the centromere) or dicentric and thus unstable unless one of the centromeres becomes inactivated (pseudoisodicentric).

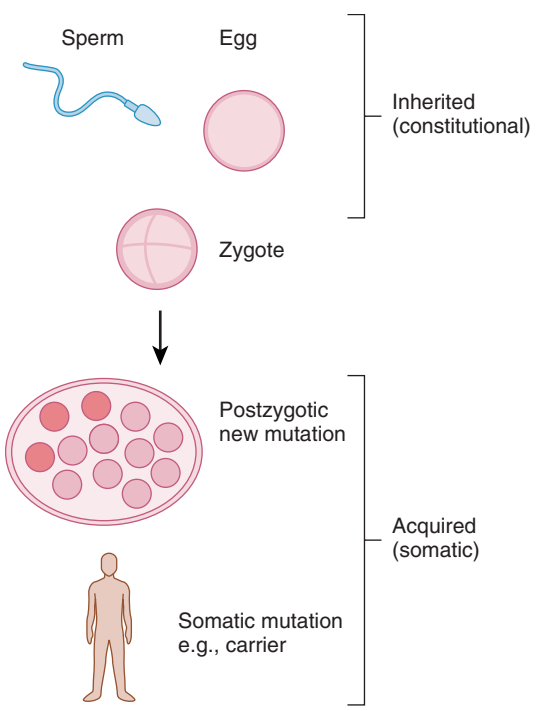

FIGURE 41-5. Mutation. Constitutional mutations are inherited from one of the parents. They can be present in the somatic cells of a parent (carrier) or can arise during gametogenesis (de novo). Mutations that occur postzygotically (acquired, somatic) are usually found in a mosaic state.

Mosaicism and Chimeras

The presence of two or more cell lines with different chromosome complements in one individual is termed *mosaicism* when they originate from the same zygote or *chimeras* when the cells originate from different zygotes. Chromosomal mosaicism is a common phenomenon and is observed in about 50% of embryos at the eight-cell stage and up to 75% of blastocysts. Somatic chromosomal mosaicism is found, for example, in patients with hypomelanosis of Ito and Pallister-Killian syndrome (tetrasomy 12p). Genome mosaicism in an organism may be more prevalent than appreciated and can be responsible for disease.[11]

MUTATIONS

Mutation is defined as a change in nucleotide sequence due to errors of DNA replication, recombination, repair, or radiation, chemical mutagens, viruses, or transposons. Gene mutations can be inherited from a parent (hereditary, constitutional, or *germline mutations*) and thus present in every cell or can be acquired in some tissues during development or any time throughout a person's life (*somatic mutations*) (Fig. 41-5). Whole genome sequencing of lung cancer tissue suggests one new point mutation for every pack of cigarettes smoked. Point mutations, usually involving only one or a few nucleotides, have been divided into substitutions, insertions, and deletions. Mutations mapping in protein-coding sequences and changing the protein structure have been termed *nonsynonymous,* whereas those that do not lead to protein change are known as *synonymous* or silent mutations. The latter mutations can still have functional consequences, for example, by generating a cryptic splice site, an exon splice enhancer, or affecting the regulatory elements.

Based on the functional consequences, mutations have been divided into loss-of-function and gain-of-function mutations. The former, also known as *hypomorphic* (partial loss) or *amorphic* or *null* (complete loss), affect the dosage-sensitive or haploinsufficient genes, in which a decreased amount of protein is not sufficient for normal function. Gain-of-function mutations increase or add a new function for the protein (*neomorphic*), whereas dominant negative mutations encode a protein that interacts antagonistically with the normal product from the other allele (*antimorphic*).

The situation in which one allele is mutated and the second is normal (wild-type) is referred to as *heterozygous*. A combination of the same two mutations in each of the alleles of the same locus (e.g., in consanguineous families) is defined as *homozygous*, or *compound heterozygous* when the two mutant alleles are distinct. Two mutant alleles at different loci are described

as *double heterozygous*. When one of the alleles is absent (e.g., because of a deletion CNV or for most of the X chromosome genes in males), the locus is referred to as *hemizygous*.

Different mutations in one gene can manifest with the same or distinct phenotypes, phenomena called *allelic heterogeneity* or *allelic affinity*, respectively. By contrast, the same abnormal clinical phenotype can be caused by mutations in different genes (*genetic* or *locus heterogeneity*).

Single Nucleotide Variants

Nonsynonymous mutations can either lead to a single amino acid change (*missense*), alter the reading frame leading to alteration in the downstream protein structure (*frameshift*), introduce a stop codon (*premature termination codon* [PTC]) that truncates the protein prematurely (*nonsense*), or abolish the specific site at which splicing of an intron takes place during the processing of precursor mRNA into mature mRNA (splice site). The PTC mutated mRNAs are inactivated and removed from cells by a surveillance mechanism called *nonsense mediated decay* (NMD) that is initiated by a premature termination codon in any exon except the last and a 50- to 55-bp portion of the second to last (i.e., penultimate) exon that usually escape NMD.

Transition mutations refer to changes from pyrimidine to pyrimidine (e.g., C to T) or purine to purine (e.g., A to G) bases and are more frequent than transversions, which are exchanges from purine to pyrimidine (e.g., A to C) or pyrimidine to purine (e.g., T to G) bases. The most common transition, C to T, is about 10-fold more common than other base changes and occurs in the methylated CpG dinucleotide because methylated C can be deaminated and converted to T (see Fig. 41-2).

Unstable Repeat Expansions

Mutations that are unstable have been termed *dynamic*. Pathogenic dynamic expansion of trinucleotide, tetranucleotide, and pentanucleotide repeat sequences can be located in coding (e.g., CAG triplet in Huntington disease) or noncoding regions such as introns (e.g., GAA in Friedreich ataxia), or untranslated regions, either 5′ (e.g., CGG in fragile X syndrome) or 3′ (e.g., CTG in myotonic dystrophy). The mutations convey phenotypes that can be inherited as autosomal dominant (e.g., myotonic dystrophy), autosomal recessive (e.g., Friedreich ataxia), or X-linked (e.g., fragile X syndrome) traits due to gain or loss of function of the encoded protein. For each of the dynamic mutation diseases, there is a specific repeat number limit, above which the disease is manifested. The number of repeats below that threshold but greater than normal is referred to as a *premutation*. However, in some "disease genes," premutations are also associated with a milder, later onset, and sometimes distinct phenotype (e.g., ovarian failure in females and late-onset neurologic disorders in males with premutations in the fragile X syndrome *FMR1* gene). The number of repeats tends to expand in the next generations, a phenomenon called *anticipation*; this typically occurs in a sex-specific manner.

Copy-Number Variants

The Watson-Crick DNA base-pair changes are not the only mutational mechanism responsible for Mendelian monogenic diseases and complex traits. Higher order genomic architectural features can lead to a regional intrinsic instability of the human genome and susceptibility to DNA rearrangements, that is, CNVs that can be a frequent cause of diseases in humans. Such conditions that result from structural genome changes or CNVs have been referred to as *genomic disorders*.

A major mechanism by which rearrangements convey phenotypes is alteration of gene dosage because of a variation in gene copy-number. CNVs can lead to deletion, duplication, or disruption of the dosage-sensitive gene, generate gene fusions, exert position effects, or unmask mutations in the coding region or other functional SNPs in the second allele, as when a deletion CNV results in a hemizygous state.

Different calculations have shown that the de novo locus-specific mutation rates for genomic rearrangements are between 10^{-4} and 10^{-5}, at least 1000 to 10,000 fold more frequent than de novo point mutations. Thus, new-mutation CNV can contribute significantly to sporadic disease[12], including various common human neurodevelopmental conditions such as schizophrenia, autism, and intellectual disability, as well as sporadic cases of rare Mendelian disorders.[13]

Many genomic disorders occur sporadically and are often caused by de novo rearrangements. Recurrent rearrangements (deletions, duplications, or inversions) are caused by NAHR between low-copy repeats that are located less than 5 to 10 Mb from each other and have greater than 97% DNA sequence identity. The fixed position of these low-copy repeats or segmental duplications in the human genome result in recurrent rearrangements having a common size for a given region. NAHR between directly oriented low-copy repeats leads to deletions or reciprocal duplications of the genomic region located between them, whereas NAHR between the oppositely oriented low-copy repeats results in an inversion of the intervening genomic segment. Interestingly, the strand exchanges for NAHR sites are not scattered throughout the entire length of homology within low-copy repeats but instead cluster in recombination hotspots.

Most nonrecurrent CNVs appear to occur by nonhomologous recombination mechanisms, and one often observes microhomology at the breakpoints. The remainder of nonrecurrent different-sized rearrangements likely result from a NHEJ recombination mechanism. One prominent mechanism, particularly for complex (e.g., deletion/normal/duplication) rearrangements is the microhomology-mediated break-induced replication (MMBIR) mechanism.

Microduplication and Microdeletion Syndromes

Some of the microduplication and microdeletion syndromes are caused by a copy-number change of the dosage-sensitive or haploinsufficient gene. Among the best characterized genomic disorders are common autosomal dominant peripheral neuropathies, CMT1A and hereditary neuropathy with liability to pressure palsies (HNPP), that are caused by duplication and deletion CNV, respectively, of an about 1.4 Mb genomic interval within 17p12 harboring a dosage sensitive myelin gene *PMP22*. This genomic segment is flanked by two approximately 24 kb and 98.7% identical LCRs, termed the *proximal* CMT1A-REP and the *distal* CMT1A-REP, which serve as substrates for NAHR. Another example of common predominantly monogenic reciprocal microdeletion/microduplication syndromes is Potocki-Lupski syndrome, which can present clinically as autism and occurs due to dup(17)(p11.2p11.2), the recombination reciprocal to del(17)(p11.2p11.2) found in patients with Smith-Magenis syndrome. When two or more dosage-sensitive genes that are usually functionally unrelated are involved, these are referred to as contiguous gene deletion or duplication syndromes, for example, Potocki-Shaffer syndrome resulting from deletion del(11)(p11.2p11.2). LCR-mediated recurrent microdeletion and microduplication syndromes usually have similar prevalence in different populations; however, for a few genomic disorders, significant differences in incidences in different world populations have been observed, likely demonstrating that variation of genomic architecture is a significant factor for disease susceptibility (e.g., 17q21.31 microdeletion syndrome, Sotos syndrome, and 5q35). Examples of well-known and characterized microdeletion syndromes include Williams-Beuren syndrome (7q11.23), Prader-Willi and Angelman syndromes (15q11.2q12), DiGeorge syndrome (22q11.2), microdeletion 17q21.31 syndrome, microdeletion 1q21.1 syndrome, and Sotos syndrome. For all these microdeletions, the reciprocal microduplications predicted by the NAHR model have been reported, with phenotypes typically being milder. Whereas the role of CNVs in the causation of disorders like the aforementioned syndromes has been known for a while, recent studies have shown that a proportion of patients with neuropsychiatric manifestations including autism and schizophrenia harbor CNVs involving specific loci (e.g., 1q21.1, 15q13.3, and 16p11.2).[13] It is now becoming apparent that CNVs may be important for the some of the complex human traits.

PATTERNS OF INHERITANCE

Mendelian Inheritance

Most characterized disease-associated mutations in humans can be assigned to a single gene (monogenic) or locus and segregate as a Mendelian trait in an autosomal dominant, autosomal recessive, or X-linked fashion.

Autosomal dominant mutation is present in only one allele and thus is transmitted in meiosis to 50% of the gametes and is expected to manifest in half the offspring unless the trait is incompletely penetrant (e.g., in BRCA-related breast and ovarian cancer), represents variable expressivity (e.g., in Marfan syndrome), is age dependent (e.g., in Huntington disease), or is lethal (e.g., alveolar capillary dysplasia). In pedigree analysis, autosomal dominant inheritance is revealed as a vertical transmission of the trait.

In an autosomal recessive trait, the affected individuals carry two mutant alleles at a specific locus that are either the same (homozygous) or different (compound heterozygous). In general, both mutations are inherited from unaffected carrier parents (but note that occasionally heterozygous carriers of the mutated allele may manifest a mild phenotype or have an increased susceptibility to complex or multifactorial traits). Theoretically, affected

probands represent 25% of the progeny; one half of the unaffected siblings carry one mutated allele, and the remaining one fourth of all progeny (one third of unaffected) have two wild-type (normal) alleles. In pedigree analysis, autosomal recessive inheritance is observed as horizontal transmission of the trait.

In X-linked (both dominant and recessive) diseases, no male-to-male transmission is observed, and all daughters of affected fathers are obligate carriers of the mutated allele. X-linked dominant diseases are more rare than X-linked recessive disorders and present both in males and in females. Usually, there are twice as many affected females as males; however, if the disease is lethal in males, only females are affected (e.g., Rett syndrome). Because of X inactivation, the phenotype in females is milder than in males in X-linked dominant diseases. In an X-linked recessive trait, only males are affected; in female carriers, the X chromosome harboring a mutated recessive allele is preferentially inactivated by nonrandom X inactivation. However, females with an incomplete or skewed X inactivation, females with only one X chromosome (Turner syndrome), or females carrying a balanced translocation between the X chromosome and an autosome (X material on the derivative chromosomes is not inactivated) can manifest the X-linked recessive disease.

Non-Mendelian Inheritance

The occurrence of sporadic cases of the disease can be explained by a classic Mendelian inheritance, such as de novo autosomal dominant, autosomal recessive, or X-linked mutation. However, one has to consider other possibilities, including non-Mendelian inheritance—genomic imprinting, uniparental disomy, mosaicism, mitochondrial DNA mutations, and digenic or triallelic inheritance.

Some genes acquire different activity status (usually methylation) after passage through spermatogenesis compared with oogenesis. As a result, a gene can be silenced (*imprinted*) depending on the parent of origin. This parent-of-origin effect is observed for the *UBE3A* gene on chromosome 15q12 that is imprinted during spermatogenesis, and only the maternal copy is active. When the active maternal copy of *UBE3A* is mutated, deleted, or inactivated in a different way, the offspring is affected with Angelman syndrome.

Sporadically, a chromosome pair may not be inherited from both parents. This distortion from biparental inheritance, termed *uniparental disomy* (UPD), may have clinical consequences when the uniparental chromosomes contain an autosomal recessive mutation or an imprinted gene. When both homologues are inherited from one parent, it is referred as *heterodisomy*. In *isodisomy*, both homologues in an offspring originate from only one of the parental homologues. The most frequent mechanism for UPD is trisomy rescue, in which an early postzygotic embryo is trisomic as a result of chromosome nondisjunction in meiosis I, and the extra chromosome is then lost during further development to restore disomy. Because this is a random event, in one third of cases, the disomic chromosomes remaining after trisomy to disomy rescue will represent UPD. Consequently, UPD is associated with advanced maternal age.

In some diseases, pathogenic mutations have been found in single alleles of two different genes with the other alleles at each given locus being normal. This double heterozygous phenomenon of two interacting genes has been reported, for example, for *ROM1* and *RDS* in retinitis pigmentosa and *GJB6* and *GJB2* in deafness.

In some patients, three abnormal alleles in two different genes have been identified. The phenomenon of triallelic (or oligogenic) inheritance has been observed, for example, in Bardet-Biedl syndrome, familial hypercholesterolemia, and cortisone reductase deficiency. Monogenic chromosomal microduplication syndromes (e.g., Charcot-Marie-Tooth type 1A) can also be categorized as triallelic given the presence of three alleles at a given locus because of duplication CNV.

Another distortion from Mendelian inheritance can be caused by mosaicism. Two or more cell lines can be present either in the gonads only (germline mosaicism) or in somatic cells. Mosaicism should be suspected when healthy parents have two or more children with a dominant disease. Mosaicism can be particularly relevant when mutational processes involve DNA replication errors and occur mitotically (e.g., point mutation and MMBIR).

Very rarely, a disease trait is transmitted to daughters and sons only from mothers. In such cases, one should consider a mitochondrial disease due to mutations in mitochondrial DNA (mtDNA). Multiple copies of mtDNA are present in the cell cytoplasm and are transmitted to progeny only through the oocytes. Initial clinical signs and symptoms typically originate from the most energy-dependent tissues (e.g., eyes, brain, skeletal muscle, and heart), and the phenotypic expression among family members varies and depends mainly on the proportion of mtDNA in the cytoplasm that carries the mutation, that is, *heteroplasmy*.

ASSAYING GENETIC VARIATION

Chromosome aberrations larger than about 5 Mb can be detected by light microscopy after specific staining that reveals characteristic banding patterns (e.g., G-banded karyotype analysis). Submicroscopic rearrangements, such as microdeletions or microduplications (30 kb to 5 Mb), have been analyzed previously using molecular cytogenetic techniques such as fluorescence in situ hybridization. In these routine clinical cytogenetic techniques, usually subpopulations of peripheral blood T lymphocytes stimulated by phytohemagglutinin are analyzed. Rearrangements of similar size (30 kb to 5 Mb) can be analyzed also using pulsed-field gel electrophoresis. However, both of these technologies are limited to the analysis of specific genomic regions, that is, locus-specific testing.

The development of array-based comparative genomic hybridization (array CGH) has enabled screening of the entire human genome for imbalances, with the level of genome resolution depending only on the number, size, and distance between the arrayed interrogating probes. These genome-wide imaging techniques are analogous to digital photography wherein the resolution observed is dependent on the pixels used. Initial clinical array CGH used large genomic clones, BACs and PACs (bacterial or P1 artificial chromosomes), as interrogating probes. These were rapidly replaced by oligonucleotides, of which millions can be synthesized on one glass slide. Oligonucleotide probes are also used on SNP arrays, which, in contrast to microarray-based CGH, enable association studies or detection of uniparental disomies. The widespread use of array CGH for diagnostic purposes not only has increased the sensitivity in detecting CNV associated with disease but also has led to the discovery of many new genomic disorders.

For detection of genomic imbalances, an alternative quantitative polymerase chain reaction–based technique, multiplex ligation-dependent probe amplification (MLPA), has been developed. MLPA is an inexpensive, simple, rapid, and sensitive tool to detect dosage alterations in selected genomic regions.

Most recently, several next-generation sequencing (NGS) technologies have been developed that enable simultaneous and massively parallel DNA sequencing. Such technologies have ushered in the era of "panel testing" (simultaneous assay of multiple genes implicated in a particular phenotype), whole exome sequencing (sequencing the entire coding regions of the genome), and whole genome sequencing (sequencing of the entire genome of an individual). In NGS, DNA sequencing uses chemistries other than the traditional Sanger dideoxy chain termination method. NGS methods generate far larger quantities of data at less expense; however, the individual raw sequence reads that are generated from individual amplified DNA template sequences have shorter read lengths and lower quality. Nevertheless, massive redundant sequencing of a personal diploid human genome (e.g., 30-fold coverage with respect to the haploid human reference genome sequence) provides robust and accurate personal genome sequencing. The NGS technologies have led to an explosion in gene discovery and understanding of the mechanistic bases of genetic disorders.[14] The use of whole exome sequencing alone for clinical diagnosis at the present time has resulted in a diagnostic yield of about 25%, a significant improvement over the currently available testing modalities.[15] NGS technologies have also made it possible to amplify cell-free DNA from maternal serum to facilitate prenatal screening for trisomies involving chromosomes 13, 18, or 21.

Recent advances and the relatively widespread use of array CGH and NGS technologies have transformed our understanding and diagnosis of human disease. With decreasing costs of sequencing and more extensive use of these methodologies, our understanding of the human history, evolution, disease, and treatment will inevitably advance to new levels of sophistication.

CONCLUSION

Mutations in humans are caused by SNVs and CNVs. New mutations can contribute to sporadic disease. The total genomic load can be important to clinical phenotype. Individual genetic variation is extensive. It is a sobering thought that for about 80 to 90% of the annotated genes in the reference human genome, a function remains to be elucidated for the potential clinical consequences of mutations. Furthermore, 98% of the human genome is noncoding, and the functional consequences of variation within it cannot be deciphered using the genetic code.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Lander ES. Initial impact of the sequencing of the human genome. *Nature*. 2011;470:187-197.
2. Venter JC. Genome sequencing anniversary. The human genome at 10: successes and challenges. *Science*. 2011;331:546-547.
3. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. *Ann Rev Med*. 2012;63:35-61.
4. Mikhail FM. Copy number variations and human genetic disease. *Curr Opin Pediatr*. 2014;26:646-652.
5. Riggs ER, Ledbetter DH, Martin CL. Genomic variation: lessons learned from whole-genome CNV analysis. *Curr Genet Med Rep*. 2014;2:146-150.
6. Lupski JR. Retrotransposition and structural variation in the human genome. *Cell*. 2010;141:1110.
7. de Koning AP, Gu W, Castoe TA, et al. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet*. 2011;7:e1002384.
8. Kong CM, Lee XW, Wang X. Telomere shortening in human diseases. *FEBS J*. 2013;280:3180-3193.
9. Liu P, Erez A, Nagamani SC, et al. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell*. 2011;146:889-903.
10. Kloosterman W, Koster J, Molenaar JJ. Prevalence and clinical implications of chromothripsis in cancer genomes. *Curr Opin Oncol*. 2014;26:64-72.
11. Lupski JR. Genome mosaicism: one human, multiple genomes. *Science*. 2013;341:358-359.
12. Watson CT, Marques-Bonet T, Sharp AJ, et al. The genetics of microdeletion and microduplication syndromes: an update. *Annu Rev Genomics Hum Genet*. 2014;15:215-244.
13. Ku CS, Tan EK, Cooper DN. From the periphery to centre stage: de novo single nucleotide variants play a key role in human genetic disease. *J Med Genet*. 2013;50:203-211.
14. Lupski JR, Reid JG, Gonzaga-Jauregui C, et al. Complete genome sequencing reveals *SH3TC2* mutations causing CMT1 neuropathy. *N Engl J Med*. 2010;362:1181-1191.
15. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013;369:1502-1511.

REVIEW QUESTIONS

1. Which of the following genic regions is protein coding, i.e., translated into a polypeptide?

A. Introns
B. Micro-RNAs
C. Small nucleolar RNAs
D. Long noncoding RNAs
E. Exons

Answer: E Only 1 to 2% of the genome consists of sequences that encode proteins. These protein-coding genic regions have exons and introns that are transcribed into messenger RNA (mRNA). However, only the exonic regions code for the amino acid sequences in the protein, and thus, the intervening intronic sequences are spliced out from the mature mRNA before translation. Micro-RNAs, small nucleolar RNAs, and long noncoding RNAs are transcribed but not translated; nevertheless, they are involved in many important biologic processes. (See Gene.)

2. Which of the following genetic variations in humans involves a change in a single base pair?

A. Copy-number variants
B. Single nucleotide polymorphism (SNP)
C. Inversion
D. Microsatellite
E. Repetitive element

Answer: B Genetic diversity in humans can be due to variations in single nucleotides (e.g., SNP), copy-number (e.g., duplications, triplications), copy-neutral changes (e.g., inversions, absence of heterozygosity), microsatellites (e.g., short tandem repeats), and repetitive elements. Single nucleotide changes that have allelic frequency of more than 1% are termed *single nucleotide polymorphisms*. These SNPs have been used in association studies that have tried to assess whether particular alleles confer a higher risk for common disorders like diabetes. (See Genetic and Genomic Variation in Humans.)

3. Which of these diagnostic techniques cannot be used to detect deletions or duplications involving one or more genes?

A. Sanger sequencing of genomic region of interest
B. Array comparative genomic hybridization (aCGH)
C. Multiplex ligation-dependent probe amplification (MLPA)
D. Fluorescence in situ hybridization (FISH)
E. SNP array

Answer: A Copy-number variations that include deletions or duplications can cause many human phenotypes. Sanger sequencing that involves amplification of specific regions of the genome followed by reading of the sequence typically cannot detect such copy-number changes. Molecular cytogenetic techniques that assess copy-number such as aCGH and SNP arrays (at the whole genome level) or FISH and MLPA (at specific regions of interest) should be employed in diagnosis of such deletions or duplications. (See Assaying Genetic Variation).

4. Two sisters and their brother who are affected with a severe form of neurologic disease presenting with structural brain malformation, intellectual disability, and behavioral abnormalities are evaluated for a genetic diagnosis. The parents and another sibling are asymptomatic and have no medical problems. A detailed family history shows no other members affected with a similar condition. Physical examination, laboratory tests, and imaging techniques do not point to any specific diagnosis. Which of these testing modalities is likely to be most useful in evaluation of this family?

A. Array comparative genomic hybridization (aCGH)
B. SNP array
C. Whole exome sequencing (WES)
D. Assaying for unstable repeat expansion
E. High-resolution G-banded karyotyping

Answer: C Multiple affected individuals in the same generation without other family members being affected is most suggestive of an autosomal recessive disease. When the clinical phenotype is not distinct to narrow down the putative genes, WES is the diagnostic test most likely to provide a molecular confirmation. (See Assaying Genetic Variation).

5. A 50-year-old man presents with early-onset diabetes, hearing loss, muscle weakness, and stroke-like episodes. His brother, mother, two maternal uncles, maternal grandmother, and others have a history of a similar condition. Review of the extended pedigree shows no male-to-progeny transmission. With which molecular mechanism would this be most consistent?

A. Autosomal dominant mutation that results in haploinsufficiency
B. Autosomal dominant mutation that confers neomorphic property
C. Autosomal dominant mutation that confers antimorphic property
D. Mitochondrial DNA mutation
E. Copy-number variants

Answer: D Whereas all the mutations mentioned can present with multiple members across generations being affected, the transmission of disease to progeny only through females is classic for mitochondrial DNA mutations. The phenotypic manifestations in individuals can vary based on the load of mutant mitochondrial DNA (heteroplasmy). (See Patterns of Inheritance.)

42

THE INHERITED BASIS OF COMMON DISEASES

SEKAR KATHIRESAN AND DAVID ALTSHULER

A central question in medicine is to understand why some people get sick and others do not. We seek these answers for multiple reasons: to provide explanations to our patients, to predict disease risk early enough to prevent it, and most important, to understand pathophysiology so as to design rational approaches to prevention and therapy. In some cases, a single environmental exposure is found to play a major role in disease (e.g., smoking and lung cancer, or HIV infection and AIDS). In others, such as Huntington disease or cystic fibrosis, mutation of a single gene is both necessary and sufficient to cause illness. Of course, singular answers are the exception rather than the rule; in most cases, disease arises from the combined action of inborn and somatically acquired alterations in genome sequence, environmental and behavioral exposures, and bad luck. Such disorders, which explain most morbidity and mortality in human populations, are termed *complex traits*.[1]

As a tool for generating new hypotheses about the root causes of disease, human genetics has a number of unique features. First, it is now possible to systematically query the entire genome sequence of an individual in a manner unlimited by any prior assumption about underlying genes and pathophysiologic processes responsible. Second, because the constitutional genome sequence is established at conception and unaltered throughout life, associations between genome sequence and human phenotype can be interpreted as causal rather than reactive in their relationship to disease. However, although we have entered an era in which the specific genes and variants that contribute to risk for common human diseases can be identified, much work is needed to understand their functions and to learn whether and how this knowledge can improve the practice of medicine.[2]

HERITABILITY: INHERITED VARIATION IN DISEASE RISK

Susceptibility to disease varies within and across human populations. Studies of *familial aggregation* can determine the extent to which inherited difference in the genome sequence contributes to variation in disease risk. Such studies are simple in concept and ask whether members of the same family are more similar in disease risk compared with individuals chosen at random from the population. Of course, familial clustering can reflect not only shared genotype but also shared environment. The contribution of shared genotype can be dissected further by examining concordance of disease in proportion to the extent of genetic relatedness. The simplest such design involves comparing rates of disease concordance among dizygotic and monozygotic twin pairs. More sophisticated methods have now been developed in which the relatedness of individuals is estimated directly from genotype data (rather than based on pedigrees) and concordance compared with these empirically derived estimates of relatedness. With each of these approaches, common diseases such as types 1 and 2 diabetes mellitus, obesity, hypertension, coronary artery disease, autoimmune diseases, common cancers, schizophrenia, and bipolar disease show rates of disease concordance that rise with genetic similarity. However, many other traits of clinical interest (e.g., most drug responses) have not been studied with these methods, and thus the role of inheritance in these characteristics cannot be assumed. That is, variability in a clinical phenotype (such as drug response) cannot be assumed to be inherited in nature—family studies or molecular genetic studies are needed to draw any such conclusion.

Data about familial aggregation allow the calculation of *heritability*, defined as the fraction of interindividual variability in disease risk attributable to additive genetic influences. In this framework, the remaining variability among individuals is due to the sum of all other contributions to disease risk: environmental influences on disease, nonadditive (*epistatic*) genetic effects (e.g., gene-gene interactions or gene-environment interactions), error in the measurement of relatedness or disease, and random chance. For most clinically important traits (diseases and risk factors), empirical estimates of heritability range from 20 to 80% (see Online Mendelian Inheritance in Man, available at *www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=OMIM*, for comprehensive information).

In interpreting estimates of heritability,[3] it is important to consider two crucial factors: the effect of measurement errors and the environmental context. *Measurement errors* decrease the estimate of the observed heritability of a trait. For example, a single measurement of blood pressure is much less heritable than a composite score based on serial measures of blood pressure over time. That is, estimates of heritability are lower bounds because day-to-day variability and imprecision in clinical measures can obscure the underlying biologic susceptibility entrained by inheritance. For the patient and physician, this means that although the blood pressure on a given day may not be particularly heritable, the blood pressure over time (which is the relevant risk factor for vascular disease) is heritable to a greater extent.

Second, estimates of heritability apply only to the context of the environment in which the study was performed. In the case in which environmental triggers of disease are relatively constant across a study population, inherited factors may explain much of the variation in rates of disease. In contrast, in the case in which exposure to environmental causes of disease is highly varied across the study population, nongenetic factors may outweigh the contribution of that same extent of variability in inborn susceptibility. For example, the rate and diversity of smoking behavior have a major impact on how much of the variability in rates of lung cancer (in any given study or patient cohort) is explained by inheritance. If smoking was absent from a given population (or ubiquitous), little of the variation in lung cancer risk would be due to smoking behavior; if, in contrast, half the population smoked multiple packs a day and the other half not at all, smoking behavior would dominate over inborn susceptibility.

For these reasons, heritability is not a fixed characteristic of a given disease but an assessment of a given population, a set of measurements, and the extent to which variability in genetic and environmental exposure explains disease risk. Thus, there is no contradiction between a disease's being highly heritable (in a given population) and yet having rates that vary dramatically across populations separated by time, geography, or socioeconomic status. In broad comparisons across groups, environmental exposure and methods of clinical ascertainment often vary substantially and contribute to secular changes in patterns of disease. Conversely, within a group exposed to a relatively uniform environment and studied in a standardized manner, genetic susceptibility may play a major role in determining individual risk.

HETEROZYGOSITY: INHERITED VARIATION IN GENOME SEQUENCE

Heritability expresses the inherited variation in rates of disease; *heterozygosity* expresses the rate of inherited variation in genome sequences (Table 42-1). Heterozygosity is defined as the proportion of sites on the chromosome at which two randomly chosen copies differ in DNA sequence. Because cells are *diploid* (carry two copies of the genome sequence) and because these two copies were selected in a semirandom manner from the population, heterozygosity is equivalent to the fraction of base pairs that vary between the two copies each of us inherited from our mother and our father. That is, heterozygosity is the rate of genetic variation in the individual.

Single-nucleotide polymorphisms (SNPs) are sites at which a single letter in the DNA code has been swapped for a single alternative letter. Such variants are observed at approximately 1 in 1000 positions in the human genome sequence. In the protein-coding regions of genes, rates of genetic variation are lower—less than 1 in every 2000 bases; the rate of variation that substantially alters the sequence of the encoded protein is lower still (see Table 42-1). The lower rate of variation in coding regions is due to natural selection against alteration in the amino acid sequence of encoded proteins.

Our genomes also contain other types of sequence variation: insertions and deletions of nucleotides; alteration in the number of copies of particular genes and sequences; and larger-scale alterations, such as inversions and translocations. All types of DNA sequence change can influence gene function and contribute to disease.

The genetic variation in each individual is largely attributable to variants that are common. Empirically, more than 98% of the heterozygous sites in each individual display frequency of greater than 1% in the worldwide human population. During the last 15 years, a public database has been built that

TABLE 42-1 CHARACTERISTICS OF HUMAN GENOME SEQUENCE VARIATION

Length of the human genome sequence (base pairs)	3,000,000,000
Number of human genes (estimated)	20,000
Fraction of base pairs that differ between the genome sequences of a human and a chimpanzee	1.3% (1 in 80)
Fraction of base pairs that vary between the genome sequences of any two humans	0.1% (1 in 1000)
Fraction of coding region base pairs that vary in a manner that substantially alters the sequence of the encoded protein	0.2% (1 in 5000)
Number of sequence variants present in each individual as heterozygous sites	3,000,000
Number of amino acid–altering variants present in each individual as heterozygous sites	12,000
Number of sequence variants in any given human population with frequency of >1%	10,000,000
Number of amino acid polymorphisms present in the human genome with a population frequency of >1%	75,000
Fraction of all human heterozygosity attributable to variants with a frequency of >1%	98%

contains essentially all common sequence variants in the human population (with frequency >1%). At the time of this writing, this public database contains more than 44 million human genetic variants (*www.ncbi.nlm.nih.gov:80/SNP/index.html*). Not all these entries represent common variants (some are rare), and a small fraction may represent technical false-positive findings.

The major contribution of common variation in human sequence diversity is explained by the unique demographic history of the human population. Despite the global distribution of the current human population, it is now clear that all humans are the descendants of a single population that lived in Africa only 10,000 to 40,000 years ago. The ancestral population was small (with an effective size of perhaps 10,000 individuals), lived a hunter-gatherer existence at low population densities (relative to other humans and later domesticated animals), and had evolved in Africa during millions of years. Most human genetic variation arose in this phase of human history, before the more recent migrations, expansions, and invention of technologies (e.g., farming) that resulted in widespread population of the globe. Most common human genetic variation predates the Diaspora and is shared by all populations on earth.

A second factor is the slow rate of change in human DNA. Mutation and recombination occur at very low rates, on the order of 10^{-8} per base pair per generation; and yet, any pair of human genes traces a lineage back to a shared ancestor who lived on the order of 10^3 to 10^4 generations ago (if a generation is 20 years, then 10^4 generations is 200,000 years). In other words, considering the typical nucleotide in two unrelated humans, it is more likely that they trace back to a shared ancestor without any mutation having occurred than it is that a mutation has arisen in the intervening time. This explains why 99.9% of base pairs are identical when any two copies of the human genome are compared.

Another aspect of human variation is explained by these simple mathematical and population genetic relationships: the extent of human DNA sequence diversity attributable to rare and common variants. Each of us inherits from our parents some 3 million common polymorphisms (classically defined as those with frequency of >1%). We inherit genetic variants that are shared by apparently unrelated individuals but are at frequencies less than 1%, and we inherit thousands of variants that are limited only to the individual and the individual's closest relatives.

The shared ancestry of human populations explains another aspect of human genetic variation: the correlations among nearby variants known as *linkage disequilibrium.* Empirically, individuals who carry a particular common variant at one site in the genome are observed to be more likely than chance to carry a particular set of variants at nearby positions along the chromosome. That is, not all combinations of nearby variants are observed in the population but rather only a small subset of the possible combinations. These correlations reflect the fact that most variants in our genomes arose once in human history (typically long ago) and did so on an arbitrary but unique copy carried by the individual in whom the mutation first arose. This unique ancestral copy can be recognized in the current population by the stretch of particular alleles (known as a *haplotype*). These ancestral haplotypes, passed down from shared prehistoric ancestors in Africa, offer a practical tool in association studies of human disease because it is not necessary to measure directly each nucleotide to capture much of the information.

THE SEARCH FOR GENES UNDERLYING MONOGENIC DISEASES

The *genetic architecture* of a disease refers to the number and magnitude of genetic risk factors that exist in each patient and their frequencies and interactions in the population. Diseases can be due to a single gene in each family (*monogenic*) or to multiple genes (*polygenic*). It is easiest to identify genetic risk factors when only a single gene is involved and this gene has a large impact on disease in that family. In cases in which a single gene is necessary and sufficient to cause disease, the condition is termed a *mendelian* disorder because the disease tracks perfectly with a mutation (in the family) that obeys Mendel's simple laws of inheritance.

Some single-gene disorders are caused by the same gene in all affected families; for example, cystic fibrosis is always caused by mutations in *CFTR*. Although many individuals with cystic fibrosis carry the same founder mutation (δ-508), others carry any pair of a wide variety of different mutations in *CFTR*. The existence of many different mutations at a given disease gene is known as *allelic heterogeneity.*

A mendelian disorder can be due to a single genetic lesion in any given family but in different families can be due to mutations in a variety of genes. This phenomenon, termed *locus heterogeneity*, is illustrated by retinitis pigmentosa. Although mutation in a single gene is typically necessary and sufficient to cause retinitis pigmentosa, there are dozens of different genes in which retinitis pigmentosa mutations have been found (Online Mendelian Inheritance in Man #268000). In each family, however, only one such gene is mutated to cause disease.

Most single-gene disorders are rare (present in <1% of the population) and are manifested early in life. Many are severe and cause death before reproduction in the absence of modern medical care. The fact that most monogenic disorders are severe in childhood and rare in the population is not a coincidence but reflects the impact of *natural selection.* The deleterious effect of these mutations results in a decrease in reproductive fitness (in individuals unlucky enough to inherit them), and the mutations and the disease are therefore unlikely to drift to high frequency in the population.

There are exceptions to this general idea: cases in which the mutation causing a severe monogenic disease (such as hemoglobin S, the cause of sickle cell anemia) is common in populations. Such cases appear to be the result of a different kind of selection, known as *balancing selection*—situations in which a gene mutation is beneficial in one circumstance (a genotype or environment) but deleterious in another. Heterozygous carriers of hemoglobin S are relatively protected against malaria, and this benefit balances the deleterious effect of sickle cell disease in homozygotes.

Starting in the 1980s, the advent of genome-wide linkage analysis led to rapid success at identifying the specific genetic mutations that cause mendelian disorders, and now thousands of genes have been identified for clinically important conditions (for comprehensive information, see *www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=OMIM*). Progress was sparked by the development of a suite of powerful research techniques—*family-based linkage analysis* followed by *positional cloning*—in which a genome-wide search is undertaken for the causal gene, which is first localized to a chromosomal region. (The initial idea of genetic linkage mapping traces to Sturtevant in fruit flies in 1913 but did not become practical in humans until the 1980s.)

Once the search discovered linkage between a chromosomal region and a disease, that chromosomal neighborhood was scoured for the genetic culprit, which was recognized by the observation of mutations that altered the protein-coding sequence, enriched in cases of disease compared with unaffected relatives and population-based controls. The power of these approaches prompted and was fueled by the Human Genome Project, which provided the foundation of information on DNA structure, sequence, and genetic variation required to undertake such searches.

More recently, it has become possible to search for the mutations underlying mendelian diseases by skipping the step of family-based linkage, instead sequencing the genome of the individual and searching for mutations that might explain the disease. If the gene is already known and the mutation easily interpreted (e.g., truncating the protein), this approach is highly efficient and successful. If the gene is rarely mutated and not yet known to cause the disease, or if the mutations are in noncoding regions, direct sequencing still runs up against the analytical and clinical challenge of genome interpretation.

GENETIC INVESTIGATION OF COMMON DISEASES

Similar to mendelian disorders, most common diseases are influenced by inheritance. In contrast to mendelian disorders, the genetic contribution to common diseases is typically due to the action of many genes rather than a single gene in each family. Empirical evidence in favor of this model comes from classical family studies, which failed to observe classical mendelian ratios for common diseases. In the 1990s, the tools of family-based linkage analysis were applied to nearly all common disorders. Much of this work was done in isolated founder populations (such as Finland and Iceland) with the goal of simplifying the genetic architecture and accessing extended pedigrees. Excepting a few notable successes, these studies revealed few strong signals that localized the genes responsible for disease, indicating that few cases of common diseases are due to individual genes of large effect. If a single gene contained rare mutations of large effect that explained 20% or more of the inherited risk for type 2 diabetes, hypertension, or schizophrenia, it would long since have been found with linkage analysis.

The next shortcut to understanding the genetic determinants of common diseases was to identify and study rare families with early-onset forms of common diseases that clearly demonstrate mendelian patterns of inheritance. Important examples include the role of *BRCA1* and *BRCA2* in early-onset breast cancer, maturity-onset diabetes of the young as a form of type 2 diabetes, many monogenic disorders of blood pressure and electrolyte regulation, early-onset Alzheimer disease, and many others.

These successes provide diagnostic information for families burdened with severe, early-onset forms of disease and insight into the underlying pathways responsible for disease. For example, more than 20 genes have been identified that, when mutated, cause rare mendelian disorders of blood pressure and electrolyte regulation. So far, every one of these genes is active in the kidney, and most are involved in the renin-angiotensin-aldosterone pathway. This result is a compelling demonstration of the central importance of the kidney in human blood pressure regulation and has suggested new therapeutic targets of substantial promise.

It was hoped that the genes found to be responsible for early-onset, monogenic forms of common diseases would contribute to the more common forms of disease in the population. In this scenario, severe mutations might cause early-onset forms, and more prevalent but subtle alterations in the same genes might contribute to common forms of disease. A comprehensive test of this hypothesis awaited tools from the Human Genome Project and improved methods of genetic epidemiologic analysis.

ASSOCIATION STUDIES: FROM CANDIDATE GENES TO GENOME-WIDE ASSOCIATION STUDIES

Genome-wide association studies (GWAS) became possible in the mid-2000s on the basis of the sequencing of the human genome, cataloguing of common genetic variants, and high-throughput tools for measuring genetic variation. However, genetic association studies long predated genomic technologies and are simple in concept: the frequency of a common variant is measured in individuals with the disease of interest and compared with well-matched controls (drawn from the population at large or unaffected family members). Now, this process is routinely performed with hundreds of thousands or millions of genetic variants from the genome-wide collection.

Genetic association studies were pioneered in the context of the human leukocyte antigen (HLA) locus on chromosome 6. The HLA complex was discovered on the basis of its role in transplantation tolerance and is characterized by diverse allelic variation that can be measured by interactions of antibodies and antigens. By measurement of these protein-based (immunologic) readouts of the underlying genetic variation, HLA alleles were found to be a major determinant of susceptibility to infectious and autoimmune diseases. Starting in the 1960s, empirical data on human population genetics and genetic association studies were developed in the context of the HLA complex.

By the 1980s, tools of molecular biology made it possible to directly measure DNA variation (rather than using protein or phenotype measurements as surrogates for the underlying genetic variation), ushering in the modern era of human genetic research. In this pregenomic era, it was only practical to measure one or a small number of genetic variations in each study, limiting association studies to incomplete assessments of individual "candidate" genes selected on the basis of biologic criteria.

The study of candidate genes led to a modest number of robust and reproducible associations, such as the contribution of apolipoprotein E4 to Alzheimer disease; factor V Leiden to deep venous thrombosis; a 32-base deletion in the chemokine receptor CCR5 to HIV infection; common variants in the insulin gene to type 1 diabetes; and SNPs in the peroxisome proliferator–activated receptor γ and the β-cell potassium channel Kir6.2 to the risk for type 2 diabetes.

Early in the 2000s, comprehensive surveys of published genetic association studies showed that valid associations were few and far between, with many initial claims of association proving irreproducible, likely representing false-positive claims. One such analysis estimated that in the pre-GWAS era, only 10 to 20 bona fide associations had been documented of common genetic variants with common diseases.

A major reason for the state of this literature was the intrinsically low likelihood of finding a gene and variant contributing to any given disease. Each genome contains millions of genetic variants, and presumably only a small fraction of these influence disease. This is often described as a problem of "multiple hypothesis testing," with the investigative community searching for associations between multiple genes, multiple variants in each gene, and multiple diseases. An alternative (bayesian) statistical framework frames this issue on the basis of low prior probabilities of association. Regardless, it is conceptually clear that much more stringent statistical thresholds (than the traditional $P < .05$) are required for declaring association of genetic variants and disease.

As in linkage analysis for mendelian traits, a key to success in association studies was the advent of genome-wide search, unbiased by prior hypotheses about biologic mechanisms. With the sequencing of the human genome, development of large-scale SNP databases, and tools for genotyping up to one million SNPs per individual, by 2005 it became practical to perform GWAS to identify genomic loci harboring allelic variation. With a recognition that any given variant had a very low likelihood of truly being associated with disease, much more stringent statistical thresholds were deployed (typically requiring a P value of 10^{-7} or lower to declare "genome-wide significance").

Age-related macular degeneration (AMD) provided an early success of GWAS.[4] AMD is a typical common, polygenic disease (Chapter 423); siblings of affected patients are perhaps three to six times as likely as unrelated individuals to become afflicted, and yet family-based linkage analysis revealed only modestly significant (and modestly reproducible) linkage results. The pathophysiologic defects that underlie AMD were largely unknown until it was found that a common coding polymorphism in the gene for complement factor H is a major risk factor for AMD. The variant (*Y402H*) has a high population frequency (approximately 35% in European populations) and increases risk by 2.5- to 3-fold in heterozygotes and by 5- to 7-fold in homozygotes. Multiple other complement factors have since been found to harbor common genetic variation that influences the risk for AMD in a highly reproducible manner, providing unambiguous information about the primary role of complement in this common disease.

Since 2005, GWAS has been used to identify literally hundreds of novel genetic variants that show reproducible associations to a large variety of common human diseases. The field evolved a set of criteria and standards that largely eliminated the previous difficulties with irreproducible claims of association, making association studies a reliable method to identify genomic loci related to human diseases. The National Human Genome Research Institute of the National Institutes of Health maintains a catalogue of GWAS findings (*www.genome.gov/26525384*) that, at the time of this writing, included 12,987 such associations across 1871 publications. This represents dramatic progress compared with the two dozen or so such findings known at the start of the decade.

The results of GWAS support a number of conclusions about the role of common genetic variants in common disease. First, nearly all diseases investigated by GWAS have yielded novel findings, in many cases yielding dozens to more than 100 independent common variants associated with risk of disease. Second, only a small fraction of these findings were previously known, confirming that an unbiased genetic mapping approach can provide new clues about the etiology of common diseases. Third, most of the associations demonstrate modest odds ratios (on the order of 1.1-fold to 1.5-fold), indicating that the genetic nature of common disease is highly polygenic and that natural selection has likely purged alleles of large effect from the pool of common variants. Fourth, in only a few cases (perhaps 10%) does the associated haplotype carry a variant that alters protein structure; this suggests that much of the risk of common disease acts through effects on gene regulation rather

than protein sequences. Fifth, in sum, the variants thus far identified explain only a modest fraction (ranging between 1% and 20%) of the estimated heritability of each disease, indicating that the rest of the inherited risk is due to some combination of common variants of more modest effect, rare variants not yet discovered, nonadditive interactions between genotypes and between genotype and the environment, or other (as yet unanticipated) factors.

Genome-wide approaches (not limited to candidate genes) can be thought of as testing the completeness of the sets of genes previously discovered for each disease by other approaches. For example, in the case of autoimmune diseases, many (perhaps half) of the findings from GWAS lie near a gene previously known to play a role in the immune system. Similarly, a substantial fraction of the genetic variants found to influence lipid levels lie near genes that were previously known to play a role in lipid biology (because they either carry rare mutations that contribute to mendelian forms of hyperlipidemia or were discovered on the basis of laboratory studies).[5] Examples such as autoimmune disease and lipids provide a reassuring alignment of mendelian genetics, biologic investigation, and the genes mapped by GWAS.

However, for the majority of diseases and of disease-associated genetic variants, the genomic regions showing association to disease are novel and do not contain any genes previously studied. One such case is type 2 diabetes, for which more than 80 independent genomic loci have been found to influence risk for disease, and yet only a handful were previously implicated by other methods. A second is myocardial infarction, for which perhaps one third of the SNPs lie near genes involved in low-density lipoprotein cholesterol, but the other two thirds do not contain any previously studied gene. These examples indicate that there are important gaps in our previous knowledge of pathophysiology and biologic mechanisms and that genome-wide approaches can point to high-priority candidates for study.

Although tantalizing, the results of GWAS have raised many more questions than they have answered. Each discovery implicates a particular genomic region, but it has proved challenging to establish which gene is responsible for the association. This is challenging in large part because so many of these common variants are noncoding, and methods to connect noncoding variation to the genes they regulate remain in their infancy. Even where novel genes are identified, much work is needed to discover their biologic and physiologic functions. Finally, GWAS findings explain only a fraction of the estimated heritability of most diseases, leaving open the question of which genes, and which types of variants and genetic effects, explain the remainder.

FROM COMMON VARIANTS TO INDIVIDUAL GENOMES

Although much of human genetic variation is due to common DNA variants (such as those tested through GWAS), each of us also inherits many thousands of variants that arose more recently and that tend to be lower in frequency and more population specific. To the extent that such variants have large effects on phenotype, they might have been previously identified on the basis of family-based linkage studies of mendelian disorders. However, there certainly exists a large universe of lower-frequency variations that have effects too modest to have been recognized and identified in family-based linkage analyses and are too rare to have been captured by the first generation of GWAS.

The study of lower-frequency and rare variants is now practical owing to advances in technology for DNA sequencing. With dramatic drops in price and increases in throughput, it is increasingly routine to sequence individual genomes in the context of medical research (and, in the future, clinical practice).[6] Such an approach will provide a much more complete assessment of genetic variation than was previously obtainable and will incorporate common as well as rare variants—and points to the major challenge of genome interpretation.[7]

For mendelian diseases, the sequencing of individual genomes has made it possible to bypass family-based linkage analysis and positional cloning and instead directly to sequence all protein-coding genes in the genome (so-called exome sequencing) among affected and unaffected individuals. Since 2009, the use of exome sequencing has led to the identification of numerous genes for mendelian disorders that had proved intractable with previous methods. For example, we studied a family in whom four siblings displayed extremely low blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and triglyceride levels—an apparently recessive disorder termed familial combined hypolipidemia. Previous linkage studies had identified a chromosomal region in which the causal gene lay, but because of the prohibitively large number of genes in the region, causal mutations had not been found. Exome sequencing of DNA samples from two of the siblings revealed only one gene, angiopoietin-like 3 (*ANGPTL3*), that harbored rare DNA variants in both alleles in both siblings.[8] Subsequent studies confirmed the presence of additional *ANGPTL3* mutations in unrelated individuals with the same disease.

For common diseases, elucidation of the role of low-frequency and rare variants is just beginning. At the time of this writing, initial genome sequencing studies of hundreds or a thousand cases of common diseases (compared with appropriate controls) have yielded few findings. This is likely due to some combination of (1) the causal rare variants being lower in frequency and more modest in effect size (that is, not deterministic) and thus requiring large samples to achieve statistical significance; (2) the current limitations in our ability to recognize functional mutations from the sea of benign DNA variants, which is needed to increase signal compared with noise; (3) the need for improved statistical methods for relating rare variants to disease; and (4) the natural selection during human evolution, which shaped the overall balance of rare and common variants that contribute to each disease.

CLINICAL IMPACT: PREDICTION, PREVENTION, AND DRUG TARGETS

Much has been written about the future use of genetic prediction in clinical medicine, but a sober appraisal requires consideration of the natural history of each disease, the available approaches for presymptomatic prevention, and the predictive nature of each test. Where genetic prediction is strong, disease outcomes are serious, and prevention exists, the combination can be of great clinical value. For example, in hemochromatosis, knowledge of genetic risk and measurement of iron stores allow presymptomatic phlebotomy, a safe and effective approach that reduces the development of end-organ damage and that would otherwise not be used. Similarly, testing for *BRCA* mutations in at-risk individuals provides valuable information about cancer risk, allowing women to choose between intensive monitoring and preventive surgery (mastectomy and oophorectomy) to reduce risk of cancer. What these examples share is that the disease is relatively rare, a robustly measured genetic risk factor dramatically increases risk, and an established prevention exists that otherwise (because of cost, convenience, or risk) would not be used.

For most common diseases, the role of genetic prediction remains unclear. This is because the disease is common, and genetic risk (as we understand it today) is probabilistic rather than deterministic in nature. Thus, the discrimination in risk due to genetics is much more limited. Moreover, in many cases, it is the characteristic of available interventions (rather than the genetic test per se) that limits utility. For example, some prevention strategies, such as diet and lifestyle modification for type 2 diabetes, are useful for everyone. In such settings, identification of a high-risk population is either of limited use or could even be counterproductive (if a focus on high-risk individuals ended up denying the rest of the population a worthwhile and safe prevention strategy). In other cases, we simply lack a proven preventive intervention, and thus risk estimation alone is not what limits progress. For example, the genetics of AMD has identified common variants with substantial effects on risk and a cumulative score of such variants that can stratify risk in the population by dozens-fold. However, at present, prevention for AMD involves smoking cessation, diet, and exercise, all of which are best deployed widely in the population rather than in a targeted manner.

To realize the value of genetic insights into disease, it will be necessary to develop new and more effective approaches to prevention that target causal mechanisms. One encouraging example involves the gene encoding proprotein convertase subtilisin/kexin type 9 (*PCSK9*) and risk of myocardial infarction. Mutations in *PCSK9* were first identified through genetic mapping studies of rare families with very high levels of low-density lipoprotein cholesterol. Soon, candidate gene association studies of *PCSK9* revealed the existence of common variants that reduced or eliminated the function of the *PCSK9* protein; in one study, 2.6% of African Americans carried nonsense mutations in *PCSK9*. These "loss of function" variants in *PCSK9*, being common, could be studied in large populations for impact on clinical phenotypes and were soon shown to reduce plasma low-density lipoprotein cholesterol and to protect against coronary heart disease. This indicated that reduction in *PSCK9* function would be expected to reduce risk of myocardial infarction through its effects on low-density lipoprotein cholesterol. Moreover, a small number of people were found to be homozygous for these loss-of-function *PCSK9* mutations and, despite lacking immunoreactive *PCSK9* protein, to be healthy and well. This indicated that even complete reduction in risk of *PSCK9* would likely be safe.

On the basis of these results, several companies have developed monoclonal antibody–based drugs targeting the *PCSK9* protein.[9] Preliminary data

from clinical trials of these agents have demonstrated large reductions in blood low-density lipoprotein cholesterol levels, in some cases surpassing even the most potent statin drugs. A reduction in risk of myocardial infarction is predicted on the basis of the genetic data for loss-of-function *PCSK9* mutations as well as the experience with other drugs that lower low-density lipoprotein cholesterol. However, definitive outcomes trials remain important and, at the time of this writing, have not yet been completed.

IMPLICATIONS AND FUTURE DIRECTIONS

Inherited factors contribute substantially to common as well as to rare diseases. Mendelian disorders are typically caused by rare mutations in the protein-coding regions of genes. On the basis of the results of GWAS, it is clear that common variants play a role in common disease, with typically modest effects that often act through effects on gene regulation rather than on protein structure. Each person carries a deep reservoir of less common and rare genetic variations that will be tested in coming years for a role in disease. It seems reasonable to expect that in the coming decade, systematic and integrative analyses of millions of genome sequences will define lists of genes and variants (both common and rare) that contribute to each human disease. If this international effort incorporates large and epidemiologically valid samples and takes into account factors that could bias results, such as case ascertainment, it should provide reference information needed to annotate each individual genome sequence for disease risk.

However, success in identifying genes and mutations will prove of value only if it leads to improved prediction, diagnosis, understanding, and treatment. Biologic understanding requires bedside-to-bench research, in which genes found mutated in patients are studied in the laboratory. It will be necessary to place new genes into known (and as yet unrecognized) biologic pathways and to understand how dysfunction and dysregulation lead to disease. In some cases, such as the role of complement in AMD (see earlier), initial answers may come quickly; in others, in which the relevant pathobiology is as yet unknown, the information to be gleaned from following these clues is unpredictable. In the fullness of time, genetic insights gleaned from patients should lead to a new generation of therapies that more directly target the underlying root causes of risk in the population.

New approaches to "precision" medicine will require not only development of predictive models and new therapies but also a foundation of clinical trial evidence that demonstrates benefit. That is, it is not sufficient simply to hypothesize that a genetic test or targeted therapy benefits patients, but it will be necessary to test this hypothesis in controlled trials. Such clinical trials will involve measuring DNA variation in study participants and testing approaches to intervention (prevention or treatment) based on such information. Genetic tests may prove predictive without being useful, and only careful research can demonstrate value and justify society's investment in their use.

Whereas much remains uncertain, it is clear that genetic and genomic information is accumulating at a staggering rate and holds much potential as well as challenges for the future of medicine. Rather than leaping to deploy genomics in medicine before value has been shown, it is incumbent on us to carefully develop and critically evaluate the use of this new technology to inform and improve the understanding, prevention, and treatment of disease.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet*. 2014;15:34-48.
2. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science*. 2008;322:881-888.
3. Tenesa A, Haley CS. The heritability of human disease: estimation, uses and abuses. *Nat Rev Genet*. 2013;14:139-149.
4. Gorin MB. Genetic insights into age-related macular degeneration: controversies addressing risk, causality, and therapeutics. *Mol Aspects Med*. 2012;33:467-486.
5. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical, and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707-713.
6. Lifton RP. Individual genomes on the horizon. *N Engl J Med*. 2010;362:1235-1236.
7. Dewey FE, Grove ME, Pan C, et al. Clinical interpretation and implications of whole-genome sequencing. *JAMA*. 2014;311:1035-1045.
8. Musunuru K, Pirruccello JP, Do R, et al. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med*. 2010;363:2220-2227.
9. Seidah NG, Awan Z, Chrétien M, Mbikay M. PCSK9: a key modulator of cardiovascular health. *Circ Res*. 2014;114:1022-1036.

REVIEW QUESTIONS

1. For most common diseases, concordance in disease status is higher in monozygotic twins than in dizygotic twins. This is an example of which phenomenon?

A. Heritability
B. Heterozygosity
C. Epistasis
D. Phenocopy

Answer: A Heritability is the fraction of interindividual variability in disease risk attributable to additive genetic influences. The contribution of shared genotype to a disease can be dissected further by comparing rates of disease within families as a function of the extent of genetic relatedness. The cleanest such design involves the comparison of disease concordance among dizygotic and monozygotic twin pairs. For common diseases such as types 1 and 2 diabetes mellitus, obesity, hypertension, coronary artery disease, autoimmune diseases, common cancers, schizophrenia, and bipolar disease, twin studies have documented that rates of concordance are significantly higher in monozygotic than in dizygotic twin pairs.

2. Heterozygosity is defined as the proportion of sites on the chromosome at which two randomly chosen copies differ in DNA sequence. Most heterozygosity in humans is due to

A. Mutations common in frequency (>1%)
B. Rare mutations (<1%)
C. Mutations private to specific individuals
D. Mutations that lead to loss of protein function

Answer: A The genetic variation in each of us is due largely to common variants. Empirically, more than 98% of the heterozygous sites in each individual display frequency of greater than 1% in the worldwide human population.

3. In cases in which a single gene is necessary and sufficient to cause disease, the condition is termed

A. Mendelian disorder
B. Polygenic condition
C. Phenocopy
D. Allelic heterogeneity
E. Locus heterogeneity

Answer: A In cases in which a single gene is necessary and sufficient to cause disease, the condition is termed a mendelian disorder because the disease tracks perfectly with a mutation (in the family) that obeys Mendel's simple laws of inheritance.

4. Whereas a P value threshold of .05 is considered statistically significant in most of medicine, genome-wide association studies use a P value threshold of $<10^{-7}$ as "genome-wide significance." The principal reason for the more stringent statistical threshold in genome-wide association studies is to

A. Account for multiple testing
B. Account for population stratification
C. Account for potential mismatching of cases and controls
D. Account for technical artifacts

Answer: A Each genome contains millions of genetic variants, and presumably only a small fraction of these influence disease. This is often described as a problem of "multiple hypothesis testing," with the investigative community searching for associations between multiple genes, multiple variants in each gene, and multiple diseases. As a result, much more stringent statistical thresholds (than the traditional $P < .05$) are required for declaring association of genetic variants and disease, and typically $P < 10^{-7}$ is used.

5. Most variants associated with common diseases from genome-wide association studies are of which type?

A. Noncoding variants
B. Missense mutations
C. Loss-of-function mutations
D. Insertion-deletion polymorphisms in coding sequence

Answer: A Most of the associated single-nucleotide polymorphisms from genome-wide association studies lie in noncoding regions, suggesting that they act through effects on gene regulation rather than directly altering protein-coding sequences.

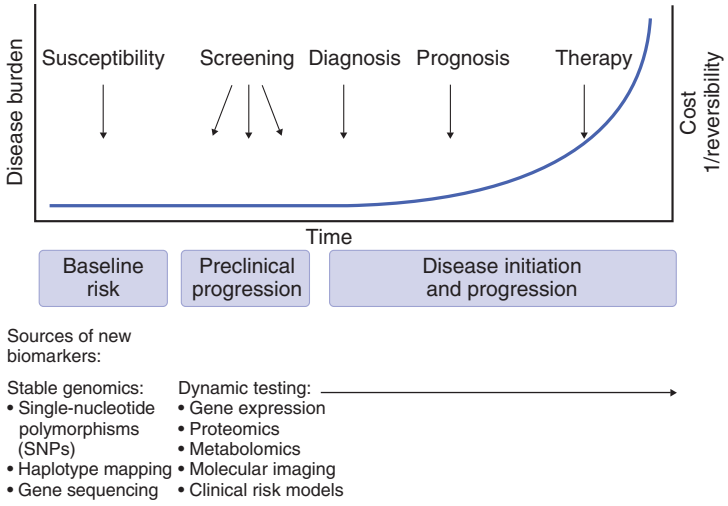

FIGURE 43-1. Use of molecular technologies across the continuum from health to disease. Various molecular technologies may be used to complement the traditional approach to evaluating at the time points indicated. (Adapted from Ginsburg GS, Willard HF. *Genomic and Personalized Medicine.* 2nd ed. Philadelphia, PA: Elsevier; 2013.)

43

APPLICATION OF MOLECULAR TECHNOLOGIES TO CLINICAL MEDICINE

GEOFFREY S. GINSBURG

The completion of the Human Genome Project more than a decade ago has become an enabler of the systematic exploration of the molecular underpinnings of disease and the expectation that these insights would lead to a transformation of medical practice. Until we were able to probe the molecular architecture of disease using genome technologies, most diseases were defined by anatomic location and clinical symptoms and treated with one-size-fits-all therapies that failed to account for the unique biologic background of the individual. The Human Genome Project laid the foundation for molecular medicine along with advances in genotyping and sequencing technologies, bioinformatics, systems biology, and computational biology. Today, molecular medicine aims to build on this foundation, translating these discoveries into clinical practice, with the ultimate goal of personalized and precision medicine.

MOLECULAR TECHNOLOGIES ALONG THE CONTINUUM FROM HEALTH TO DISEASE

Along the continuum from health to disease (as shown in Fig. 43-1), there are several important points where clinical decision making is now directly influenced and advantaged by molecular technologies (Table 43-1).[1,2] Risk estimates for developing some diseases can be defined during health and possibly even at birth using a variety of DNA analyses. Molecular signatures from technology platforms that measure the expressed genome (RNA, proteins, metabolites) can be used to define physiologic states in response to our environment and predict future clinical outcomes. These approaches also form the basis for a new molecular classification and taxonomy of disease and diagnosis. They can also provide more precise ways to screen for and detect disease at its earliest molecular manifestations, often preclinically. In addition, the selection of certain drugs may now be guided by a patient's underlying genetic makeup as well as the molecular makeup of the disease. Given that a disease's evolution from baseline risk often occurs over many years (see Fig. 43-1), periodic molecular profiling defines a novel form of health care monitoring that focuses on disease prevention and proactive management rather than the current paradigm of acute intervention and crisis response.

GENOMES, DISEASE, AND TREATMENT

A key question in medicine is to what extent genetic variation influences the likelihood of disease onset, affects the natural history of disease in combination with the environment, or provides clues relevant to the management of disease. In addition, it is not just the *human* genome that is relevant to an individual's state of health. The genomes of thousands of microorganisms that constitute *our microbiota* are also relevant to human phenotypes, and insights from *their* genomes are providing new approaches for the diagnosis, study, and treatment of disease (see later).

Sequencing: A Driver of Molecular Medicine

Genome-wide association study (GWAS) debuted in 2005 with the identification of variants in the complement factor H gene as a cause of age-related macular degeneration, (http://www.genome.gov/gwastudies). GWAS has been a transformative approach to identifying common genetic variation across the entire human genome in an unbiased fashion, offering an unprecedented opportunity to uncover new biologic pathways of disease. GWAS has been carried out on large cohorts of patients and controls across numer-

TABLE 43-1 APPLICATION OF MOLECULAR DIAGNOSTICS ALONG THE CONTINUUM FROM HEALTH TO DISEASE: EXAMPLES

TIME POINT IN CLINICAL DECISION MAKING	CANCER		CARDIOVASCULAR DISEASE	
	Test	*Indication*	*Test*	*Indication*
Risk/susceptibility	*BRCA1, BRCA2* *HNPCC* *TP53, PTEN*	Breast Colon Sarcomas	*KIF6, 9p21* Familion five-gene profile	CAD LQTS
Screening	HPV genotypes	Cervical	Corus CAD	CAD
Diagnosis	Cancer Type ID OVA1	Cancer of unknown primary Ovarian mass malignancy	Corus CAD	CAD
Prognosis	Oncotype DX (21-gene assay) MammaPrint (70-gene assay) HER2/neu, ER, PR	Breast	TnI, BNP, CRP	ACS
Pharmacogenomics	HER2/neu *UGT1A1* *KRAS* *EGFR* *ALK* *BRAF* NGS of somatic variation for targeted therapy AmpliChip; DMET *CYP2D6/CYP2C19*	Herceptin Irinotecan Cetuximab Erlotinib, gefitinib Crozotinib Vemurafenib Various (see E-Table 43-1)	*KIF6, SLCO1B1* AmpliChip; DMET *CYP2D6/CYP2C19* *VKORC1*	Statins Various (see E-Table 43-1) Warfarin
Monitoring	CTCs	Tumor recurrence or progression	AlloMap gene profile	Transplant rejection

ACS = acute coronary syndromes; BNP = brain natriuretic peptide; CAD = coronary artery disease; CRP = C-reactive protein; CTCs = circulating tumor cells; ER = estrogen receptor; HPV = human papillomavirus; LQTS = long QT syndrome; NGS = next-generation sequencing; PR = progesterone receptor; TnI = troponin I.
Data from Ginsburg GS, Willard H, eds. *Genomic and Personalized Medicine*. 2nd ed. New York, NY: Academic Press; 2012.

ous traits and diseases, revealing hundreds of common genetic variants associated with those traits.

Since 2001, the cost of sequencing a human genome dropped from $3 billion to less than $10,000. Next-generation sequencing (NGS) technology can now read approximately 250 billion bases in a week and allows direct measurement not just of common variants but also theoretically of *all* variations in a genome. It is estimated that the population frequency of germline variants is approximately 1 in every 1000 of the 3.2 billion nucleotide positions, giving rise to approximately 3 million variants in a given human genome. The challenge lies in figuring out the meaning of variants, many of which occur in noncoding regions (introns) of the genome whose function is largely unknown. Until the significance of the noncoding variants is understood, the focus clinically has been on exome sequencing (which examines variation in the coding sequence of exons that are translated into proteins), where mutations have predictable effects on downstream protein structure. Exome sequencing, which represents only about 1.5% of the 3 billion nucleotides that constitute the human genome, is still less expensive to perform than *whole-genome sequencing*. Individuals typically carry several hundred rare and potentially deleterious coding region variants. The first successful clinical applications of exome sequencing in 2009 revealed the diagnosis of patients with Freeman-Sheldon syndrome, and it is increasingly being explored for clinical applications, including the clinical diagnosis of rare genetic diseases,[3] the selection of cancer treatments based on molecular characterization of the tumor,[4] and the tracking of infectious disease outbreaks in real time[5] (see later).

Genetics of Common Complex Diseases

Despite strong statistical associations linking genetic variants with complex diseases, the low relative risk of the disease alleles (generally <2) limits their use for disease predisposition testing and risk assessment. There are notable exceptions, including genetic variation underlying breast cancer, Lynch syndrome, and celiac disease, where some variants have enabled preventive treatment or screening of family members. Despite these technologic advances in genomics, a simple family history continues to be among the best tools to identify risks for common diseases. In fact, for conditions with high heritability, such as cardiovascular disease, family history is a much stronger predictor of disease than any single or combination of genetic/genomic markers.[6] One model suggests that neither family history nor genetic testing should be used as a standalone but that the real power for disease prediction, risk assessment, and differential diagnosis comes from their combined use.

Clinical Sequencing for Diagnostic Dilemmas

More than 3500 mendelian disorders have a known molecular basis (http://omim.org/). However, there are nearly as many suspected mendelian traits for which the molecular basis remains to be identified. The potential for clinical sequencing to find the underlying cause and identify treatment options for these rare, sometimes debilitating diseases has led to the formation of various large national and international rare disease consortia. In some specialized clinical centers and through programs such as the National Human Genome Research Institute Undiagnosed Diseases Program (http://www.genome.gov/27544402), clinical sequencing is being offered to patients with suspected genetic diseases, the so-called diagnostic dilemmas. Early results from these clinical sequencing programs suggest that the success rate of disease gene identification is about 50%, offering hope for a diagnosis to thousands of individuals with previously undiagnosed or untreated rare disorders.[7]

Newborn Screening, Prenatal Diagnosis, and Preconception Carrier Testing

A natural outcome of identifying genes for rare mendelian disorders is the application of these findings to earlier detection, at birth (newborn screening), in utero (prenatal diagnosis) or before conception (carrier testing). Newborn screening—mandatory, state-supported public health programs meant to protect newborn children by screening them for rare, treatable (and thus preventable) disorders at birth—has been steadily increasing from an average of five conditions in 1995, to a panel of 31 core disorders and 26 secondary disorders currently recommended by the U.S. Department of Health and Human Services.[8] Rapid whole-genome sequencing (about 50 hours from test to result) was recently reported, and because the number of conditions considered for newborn screening will undoubtedly grow, rapid whole-genome sequencing can potentially broaden and foreshorten differential diagnoses, resulting in fewer empirical treatments and faster progression to genetic and prognostic counseling.

Cell-free fetal DNA circulating in maternal blood was isolated, amplified, and sequenced noninvasively through a sample of maternal plasma in 1997, In 2008, NGS technologies were used successfully to identify fetal aneuploidy from cell-free fetal DNA in maternal plasma. Clinical trials of the new method rapidly followed, and by late 2011, noninvasive prenatal testing of trisomy 21 by sequencing of maternal plasma DNA was being offered on a clinical and commercial basis in the United States and China. Noninvasive prenatal testing eliminates the need for invasive procedures, while also greatly expanding the number of genetic variants that have traditionally been detected in utero.[9]

Before conception, carrier screening enables couples to assess their risk for having a child with a recessive mendelian disorder and to use this information to guide their reproductive decisions. There are more than 1000 rare, *recessive* mendelian disorders for which the underlying genetic mutation is known. Although individually rare, these can have a sizable public health impact

considering that each person is estimated to carry on average 2.8 mutations for known severe recessive disorders,[10] and the impact of screening could be substantial in terms of reduced disease morbidity and mortality in the population.

Pharmacogenomics: Germline Genetics Variants and Drug Response

Several genomic markers of efficacy, adverse events, and dosing of therapeutics have been discovered (E-Table 43-1), but their uptake into clinical practice has been variable, despite their clear actionability. In some cases, such as with the *HLA-B*5701* genotype for the HIV drug abacavir and *HLA-B*1502* for the antiseizure drug carbamazepine, carriers of these genotypes should avoid the drug entirely to eliminate a specific serious adverse event. In other cases, such as thiopurine *S*-methyltransferase (*TPMT*) for mercaptopurine or *CYP2C9/VKORC1* for warfarin, adjusting the dose of drug based on genotype can help to avoid toxicity and improve efficacy. *Actionability* is not enough to ensure diffusion of pharmacogenomics testing into clinical practice, as exemplified by the antiplatelet drug clopidogrel, for which despite having a U.S. Food and Drug Administration black box warning for efficacy in individuals carrying the *CYP2C19* genetic variant, there is no clear consensus among physicians on its use. In hepatitis C treatment, on the other hand, the *IL28B* genotype test not only has proved highly predictive of response to pegylated interferon/ribavirin used to treat chronic hepatitis C virus infection but also has seen rapid and widespread adoption in the clinic.[11] Genetic markers that predict reduced therapeutic efficacy may face a high hurdle for established drugs, unless evidence supporting clinical validity and utility of the test is indisputable.

Cancer Pharmacogenomics: Somatic Sequencing of Tumor DNA for Targeting Drug Therapies

Cancer arises as a result of somatic DNA mutations that confer a growth advantage on the cells in which they have occurred, giving rise to tumors. Comparison of the genetic profiles of tumors and the surrounding normal tissue (*gene expression profiling*) can reveal the acquired DNA variation that drives growth and that may reveal targets for treatment.

Targeted Therapies for Cancer

The idea of pairing medicines with specific tumor markers in a targeted fashion became a reality in the mid-1980s when detailed molecular studies of breast tumors led to the discovery of human epidermal growth factor receptor-2 (HER-2), a biomarker overexpressed in approximately 30% of breast tumors and associated with adverse outcomes. Subsequently, trastuzumab (Herceptin), a humanized monoclonal antibody targeting HER-2, was developed in 1998 and was shown to have increased efficacy in patients whose tumors tested positive. HER-2 testing of tumor is now part of the standard work-up and management of breast cancer. In the past decade, other examples of cancer therapies with companion diagnostics have emerged (Table 43-2). For example, *EGFR* mutation testing has markedly improved the efficacy of gefitinib and erlotinib, small molecule drugs for the treatment of non–small cell lung cancer that target *EGFR*. In metastatic colorectal cancer, tumors with mutated *KRAS* are usually resistant to treatment with cetuximab and panitumumab, leading the American Society of Clinical Oncology and the U.S. Food and Drug Administration (FDA) to recommend withholding the drugs in these patients. NGS now allows a comprehensive assessment of actionable tumor markers that indicate the potential for a specific therapeutic to have efficacy in a given tumor (see Table 43-1). In 2011, two cancer drugs received accelerated approval by the FDA for use with a companion diagnostic test: (1) crizotinib for the treatment of patients with locally advanced or metastatic non–small cell lung cancer with its companion diagnostic designed to detect the *EML4-ALK* fusion gene, and (2) vemurafenib for the treatment of patients with metastatic or unresectable melanoma positive for *BRAF* V600E mutations. The International Cancer Genome Consortium (https://www.icgc.org/icgc) and the Cancer Genome Atlas (http://cancergenome.nih.gov/) represent international collaborative efforts to define the spectrum of mutations found in tumors, mapping the genomic landscape of cancer. These efforts will provide a foundation from which to develop additional therapeutic strategies against new targets. However, even when successful, the results may be short-lived as therapeutic resistance evolves. Thus, although NGS is a promising new tool for surveying cancer genomes, it may not be a panacea for cancer genomic medicine.

Microbial Genomes: Friends or Foes?

We can now rapidly sequence the genomes of microorganisms—both the commensal bacteria that regularly inhabit our bodies (the *human microbiome*)[12] as well as the pathogenic infectious agents that cause acute and sometimes fatal diseases.[13] The Human Microbiome Project recently published a study of the microbial populations inhabiting various human body sites[14] and provided reference sequences for many taxa in health individuals as well as their correlation with host characteristics, including ethnicity, age, and body mass index. There are now emerging associations of human microbiota and diseases such as diabetes, asthma, psoriasis, atherosclerosis, and obesity.[15,16] Moreover, strategies to modify the gut microbiome are being explored as treatments for inflammatory bowel disease, including the use of fecal transplantation or engraftment of microbiota from a healthy donor into a recipient.[17] The human microbiome will play an important role in molecular medicine because microbial composition can be altered noninvasively through diet or the use of probiotics or antibiotics.

In infectious disease, diagnosis by NGS may supplant the need to first grow microorganisms in culture, previously a major impediment to pathogen identification. For example, in 2003, sequencing of samples from infected patients with the severe acute respiratory syndrome identified the causative agent as a coronavirus. Comparison of sequences of multiple isolates of an organism from a single epidemic gives a picture of the organism's evolution, allowing one to infer where the outbreak began and how the infection spread. Sequencing has been used to determine the origins of historical outbreaks of cholera, tuberculosis, and the 2009 H1N1 influenza. The clinical application of NGS to infectious disease was highlighted recently when the source of carbapenem-resistant *Klebsiella pneumoniae* in a hospital outbreak was identified by sequencing isolates of the bacteria—in real time—from infected individuals and examining the genetic differences.[18]

TABLE 43-2 MOLECULAR MARKER INFORMED CANCER THERAPIES (TARGETED THERAPEUTICS)

BIOMARKER	DRUG	CANCER TYPE	FDA DRUG LABELING RECOMMENDED OR REQUIRED
Estrogen receptor	Tamoxifen	Breast	Yes
Her2/neu	Trastuzumab	Breast	Yes
EGFR	Cetuximab	Colorectal	Yes
Kras	Cetuximab	Colorectal	Yes
EGFR	Panitumumab	Colorectal	Yes
Kras	Panitumumab	Colorectal	Yes
DPYD	5-FU	Breast/colorectal	No
EGFR	Erlotinib	Lung	No
EGFR	Gefitinib	Lung	No
BCR-ABL	Imatinib	CML	Yes
C-KIT	Imatinib	CML/ALL	Yes
ALK	Crozotinib	Lung	Yes
BRAF	Vemurafenib	Melanoma	Yes

5-FU = 5-flurouracil; ALL = acute lymphoblastic leukemia; CML = chronic myelogenous leukemia; FDA, U.S. Food and Drug Administration.

E-TABLE 43-1 SELECTED MOLECULARLY GUIDED THERAPEUTICS AND INDICATIONS

THERAPEUTIC	MOLECULAR TEST	INDICATION
Cancer treatment regimens	Oncotype DX 21-gene assay	**Breast cancer**: 21-gene expression score linked to the likelihood of breast cancer recurrence in women and the magnitude of benefit from certain types of chemotherapy and hormonal therapy
Irinotecan	*UGT1A1*	**Colon cancer**: variations in the *UGT1A1* gene can influence a patient's ability to break down irinotecan, which can lead to increased blood levels of the drug and a higher risk for side effects.
Carbamazepine	*HLA-B*1502*	**Epilepsy and bipolar disorder**: serious dermatologic reactions and *HLA-B*1502* allele. Before initiating therapy, testing for *HLA-B*1502* should be performed in patients with ancestry in populations in which *HLA-B*1502* may be present.
Abacavir	*HLA-B*5701*	**HIV**: test determines patients most likely to experience an adverse hypersensitivity reaction.
Omeprazole Pantoprazole Esomeprazole Diazepam Nelfinavir Rabeprazole Voriconazole Clopidogrel *CYP2D6* Atomoxetine Fluoxetine HCl Olanzapine Cevimeline hydrochloride Tolterodine Terbinafine Tramadol Acetaminophen Clozapine Aripiprazole Metoprolol Propranolol Carvedilol Propafenone Thioridazine Protriptyline HCl Venlafaxine Risperidone Tiotropium bromide Tamoxifen Timolol maleate	AmpliChip DMET *CYP2D6/CYP2C19*	**Multiple diseases**: these tests are used as an aid in determining treatment choice and individualizing treatment dose for therapeutics that are metabolized primarily by the specific enzyme about which the system provides genotypic information.
Warfarin	*CYP2C9*	**Venous thrombosis/stroke**: increased bleeding risk for patients carrying either the *CYP2C9*2* or *CYP2C9*3* alleles
Warfarin	*VKORC1* (vitamin K epoxide reductase)	**Venous thrombosis/stroke**: single-nucleotide polymorphisms in the *VKORC1* gene (*-1639G/A* allele) associated with lower dose requirements
Clopidogrel	*CYP2C19*	**Coronary artery disease:** increased risk for stent thrombosis and secondary events following percutaneous interventions in patients with the *CYP2C19*2* variant
Immunosuppressive drugs	AlloMap gene profile	**Heart transplantation**: blood gene expression score to monitor a patient's immune response after cardiac transplantation and to guide immunosuppressive therapy
Pharmaceutical and surgical prevention options and surveillance	*BRCA1, BRAC2*	**Breast cancer**: guides surveillance and preventive treatment based on susceptibility risk for breast and ovarian cancer
Pharmaceutical and lifestyle options for disease prevention	Familion 5-gene profile	**Heart disease**: guides prevention and drug selection for patients with inherited cardiac "channelopathies" such as long QT syndrome that may lead to cardiac rhythm abnormalities
Mercaptopurine Thioguanine Mercaptopurine Azathioprine	*TPMT* (thiopurine *S*-methyltransferase)	**Acute lymphoblastic leukemia**: patients with inherited little or no TPMT activity are at increased risk for severe toxicity from conventional doses.
Maraviroc	CCR5 promoter and coding polymorphisms	**HIV:** determines whether the patient is likely to respond or not

THE EXPRESSED GENOME

Complex Multimarker Genomic Tests for Disease Diagnosis and Prognosis

Beyond DNA sequence, measures of gene expression, proteins, metabolites, and epigenetic changes are being used to generate comprehensive profiles of biologic systems in health and disease. Many of the computational challenges of analyzing these large, complex data sets are being addressed to yield next-generation biomarkers that are multianalyte, diagnostic, prognostic, and predictive. A growing number of marketed tests now typically measure protein or RNA levels, often with complex algorithms, enabling diagnosis and prognosis (see Table 43-1). One example is Oncotype DX (Genomic Health Inc., Redwood City, CA), a test that examines expression of 21 genes in tumor tissue to determine the likelihood of disease recurrence in women with early-stage hormone estrogen receptor–positive breast cancer. The test, which is currently covered by many major insurance companies, analyzes expression levels and converts them to a recurrence risk score that has been shown to help guide treatment in patients, reduce overall health care costs, and improve outcomes.[19] Other examples include MammaPrint (Agendia Inc., Irvine, CA), which analyzes the expression of 70 genes to determine whether patients are at high or low risk for breast cancer recurrence; OVA1 (Vermillion, Inc., Austin, TX), a five-protein test that gauges whether a woman's ovarian mass is malignant and requires surgery; AlloMap (XDx Expression Diagnostics, Inc., Brisbane, CA), an 11 blood gene RNA signature for monitoring rejection after cardiac transplantation; and Corus CAD (CardioDx, Inc., Palo Alto, CA), a 23-gene blood RNA signature to screen for obstructive coronary artery disease.

Despite their complexity, in vitro diagnostic multianalyte index assays (IVDMIAs) like these are finding their way to the clinic. The 2007 draft guidance from the FDA suggested that IVDMIAs are used to make critical health care decisions and thus should be regulated by the FDA. Some of the marketed IVDMIAs have demonstrated analytical and clinical validity, but evidence of clinical utility is usually lagging. Moreover, the very nature of IVDMIAs presents challenges to insurers, who grapple not only with limited data on clinical utility but also with how to reimburse such tests that comprise both a laboratory component and an associated algorithm used to score risk, the latter part being integral to realizing the test's value.

The success of some IVDMIAs is evidence of the power of computational biology but also of the importance of advocacy and financial resources that the commercial developers of these tests must bring. Companies developing IVDMIAs are able to finance key studies aimed at demonstrating clinical validity, navigate regulatory hurdles, advocate for coverage by insurance companies, and disseminate their tests through marketing to health care providers. Their efforts offer valuable lessons on the effective translation of complex molecular tests to medicine.

Proteomics

The large-scale study of proteins, *proteomics*, allow for both protein identification and differential expression between two physiologic states (such as health and a specific disease). Quantitative proteomics, in which global differences in protein abundances are measured, continues to be a priority area for biomarker discovery and molecular medicine. This area has been dominated by stable isotope approaches, but recent label-free quantitative methods have been developed that rely on the measured intensity of a peptide ion and compares this to its intensity in other samples. Label-free methods have the advantage of higher throughput and fewer sample manipulation steps. Multiple—or selected-reaction monitoring of specific peptides within biofluids allows quantitation of absolute abundance of proteins in clinical samples. Although this technology is relatively immature in its applications to human health and disease, compared with RNA and metabolic profiling, it is anticipated that these methods, combined with the development of mass spectroscopy technology, will advance proteomics to more routine use in disease classification and diagnosis, prognosis, and pharmacogenomics within the next several years.

Metabolic Profiling

A metabolic profile is very similar to some of the traditional targeted profiles, such as a lipid profile, although it is more comprehensive. *Metabolomics* measures changes in the metabolic or chemical milieu that are downstream of genomic and proteomic alterations. It is estimated that humans contain approximately 5000 discrete small molecule metabolites, and the identification of metabolic fingerprints for specific diseases may have particular practical utility for the development of therapies because metabolic changes immediately suggest enzymatic drug targets. Similar to genomics and proteomics, metabolomics may be useful in disease diagnosis, prognosis, and drug development. In particular, metabolomics will likely be a valuable tool in assessing drug toxicity. Targeted mass-spectroscopy-based metabolic profiling has also been increasingly applied to studies of human diseases and conditions. These tools are being applied to diverse areas, such as diabetes, obesity, cardiovascular disease, cancer, and mental disorders.

CLINICAL IMPLEMENTATION OF MOLECULAR PROFILES

In order for molecular medicine to be practiced, it must be woven into current systems of health care delivery, with due consideration given not only to the providers of health care but also to the organizations in which they practice as well. Implementation scientists have outlined various aspects that need to be considered in order for molecular medicine to take hold in the clinical setting. Beyond the scientific soundness of the molecular or genomic test, measured by a strong evidentiary base and regard for potential benefits and harms, there is consideration of how the new test will integrate into the clinical workflow. Consideration should be given to aspects such as access to a laboratory certified by the Clinical Laboratory Improvement Amendments of 1988, methods for sample preparation and transport, test ordering, and receipt and delivery of results. Genomic test implementation is complicated by issues of privacy, complex interpretation of results, and the need to involve third parties for counseling in some cases; they may require the development of new systems to accommodate them.

A robust means of integrating genomic and molecular data into electronic health records will be required, with consideration of not only data storage formats and privacy issues but also appropriate decision support tools for prompting their use at the point of care and delivering results in an easily interpretable format. Currently, there are several examples of decision support tools, such as Warfarin Dosing (www.WarfarinDosing.org), but they are typically standalone tools and not part of routine clinical workflow. To maximize their effectiveness, such tools should be integrated into electronic health records. Tapping into the collective knowledge and experience of various institutions working in this space would greatly facilitate this effort. Ultimately, a national, standardized technical architecture for integrating clinical decision support into electronic health records will be required. Notable efforts in this space include those of Health Level 7 (http://www.hl7.org), an organization that provides interoperability standards for the exchange, integration, sharing, and retrieval of electronic health information. Through their Clinical Genomics Workgroup, this organization has developed a standards guide for genetic testing that includes document templates to support integration of genetic testing into electronic health records.[1] Appropriate clinical decision support, provided in the context of the electronic health record, will greatly facilitate the diffusion and uptake of genomic medicine.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ginsburg GS, Willard H, eds. *Genomic and Personalized Medicine*. 2nd ed. Philadelphia, PA: Elsevier; 2013.
2. Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. *N Engl J Med*. 2014;370:2418-2425.
3. Rabbani B, Mahdieh N, Hosomichi K, et al. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet*. 2012;57:621-632.
4. Ong FS, Das K, Wang J, et al. Personalized medicine and pharmacogenetic biomarkers: progress in molecular oncology testing. *Expert Rev Mol Diagn*. 2012;12:593-602.
5. Harris SR, Cartwright EJ, Torok ME, et al. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. *Lancet Infect Dis*. 2013;13:130-136.
6. Do CB, Hinds DA, Francke U, Eriksson N. Comparison of family history and SNPs for predicting risk of complex disease. *PLoS Genet*. 2012;8:e1002973.
7. Duncan E, Brown M, Shore EM. The revolution in human monogenic disease mapping. *Genes (Basel)*. 2014;5:792-803.
8. Health Resources and Services Administration, Government Advisory Committees. Newborn screening: towards a uniform screening panel and system. 2013. Available at: http://www.hrsa.gov/advisorycommittees/mchbadvisory/heritabledisorders/uniformscreening.pdf; Accessed September 30, 2014.
9. Porreco RP, Garite TJ, Maurel K, et al. Noninvasive prenatal screening for fetal trisomies 21, 18, 13 and the common sex chromosome aneuploidies from maternal blood using massively parallel genomic sequencing of DNA. *Am J Obstet Gynecol*. 2014;211:365.e1-365.e12.
10. Bell CJ, Dinwiddie DL, Miller NA, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med*. 2011;3:65ra4.
11. Booth DR, Ahlenstiel G, George J. Pharmacogenomics of hepatitis C infections: personalizing therapy. *Genome Med*. 2012;4:99.
12. Guarner F. Decade in review-gut microbiota: The gut microbiota era marches on. *Nat Rev Gastroenterol Hepatol*. 2014;11:647-649.
13. Relman DA. Microbial genomics and infectious diseases. *N Engl J Med*. 2011;365:347-357.
14. Consortium HMP. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486:207-214.
15. Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. 2012;490:55-60.
16. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. *Nat Rev Genet*. 2012;13:260-270.
17. Merenstein D, El-Nachef N, Lynch SV. Fecal microbial therapy: promises and pitfalls. *J Pediatr Gastroenterol Nutr*. 2014;59:157-161.
18. Snitkin ES, Zelazny AM, Thomas PJ, et al. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Sci Transl Med*. 2012;4:148ra116.
19. Albain KS, Barlow WE, Shak S, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol*. 2010;11:55-65.

ADDITIONAL RESOURCES

WEBSITE	URL	DESCRIPTION
CPIC	http://www.pharmgkb.org/page/cpic	Provides freely available, peer-reviewed, updatable, and detailed gene/drug clinical practice guidelines; sixe currently published, eight underway
EGAPP	http://www.egappreviews.org/	Synthesizes scientific evidence and makes recommendations on appropriate use of genetic tests in clinical practice; eight evidence reports and six recommendations currently available
FDA Biomarkers	http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm	List of pharmacogenomic biomarkers on drug labels (link to drug labels provided); currently includes >100 biomarker-drug pairs
Genetic Testing Registry	http://www.ncbi.nlm.nih.gov/gtr/	Central location for voluntary submission of genetic test information by providers; includes information on test methodology, validity, evidence of the test's usefulness, and laboratory contacts and credentials; currently includes >1200 fully registered tests for >500 conditions
PharmGKB	http://www.pharmgkb.org	Information on potentially clinically actionable gene-drug associations and genotype-phenotype relationships; currently lists 186 well-known pharmacogenomic associations and provides 46 summaries for very important genes

REVIEW QUESTIONS

1. Genome sequencing of DNA can be used for all of the following *except* which one?

A. Diagnosis of rare diseases
B. Detection of pharmacogenetic variants
C. Newborn screening
D. Prognosis of cancer
E. Diagnosis of infection

Answer: D Various types of genome sequencing have served special clinical functions. Exome sequencing is being increasingly used to diagnose rare genetic diseases and patients with diagnostic dilemmas. Germline genotyping has allowed prediction of adverse events and dosing of many commonly used drugs; although these discoveries have been "actionable," most have failed to diffuse widely into clinical practice at this time. Screening newborns, prenatal diagnosis, and preconception carrier testing are feasible for identified mendelian disorders. In infectious diseases, diagnosis of causative pathogens can be performed by next-generation sequencing, which may in the future supplant the need to first grow microorganisms in culture. Genome sequencing has not been demonstrated to be useful for prognosis of cancer at this time.

2. Transcriptional (RNA) profiling may be useful for all of the following *except* which one?

A. Susceptibility
B. Diagnosis
C. Prognosis
D. Pharmacogenetics
E. Monitoring

Answer: A Next-generation sequencing has allowed sequencing of RNA transcripts in cells. The so-called transcriptome, unlike the static genome, is continually changing. It allows examination, at any given time, of alternatively gene-spliced transcripts, post-transcriptional changes, gene fusion, and changes in gene expression. Transcriptional (RNA) profiling has been applied to diagnosis, prognosis, pharmacogenomics, and monitoring, but not susceptibility analysis.

3. Among the great challenges to implementing genomic diagnostics in the clinic is which one of the following?

A. The precision of the results
B. The evidence to support use
C. The ability of the laboratories to perform the test
D. Their integration into electronic medical records
E. Lack of patient understanding

Answer: B Surmountable challenges to implementing genomic diagnostics at the bedside include increasing the precision of results, expanding the availability of expert laboratories to perform the tests, the ability to integrate the information into electronic medical records, and lack of understanding by patients who are, however, now becoming increasingly savvy about this aspect of their own health care. Evidence to support the utility and cost-effectiveness of wide application of genomic diagnostics in clinical practice is only now being taken up by investigators in comparative effectiveness and health services research.

4. Which of the following describes the microbiome?

A. A small genome
B. A community of microbes colonizing humans
C. A tool used to make thinly sliced materials (e.g., paraffin blocks)
D. The sequence of a virus or bacterium
E. An organelle of the cell

Answer: B The microbiome is the community of microbes that colonize a human host. It is the ecological community of commensal, symbiotic, and pathogenic microorganisms that actually inhabit our body space and outnumber our own native human cells by a ratio of 10 : 1.

5. Which of the following methods may be used for detection of microbial pathogens?

A. Transcriptional profiling
B. Metabolomics
C. Sequencing
D. DNA methylation
E. Proteomics

Answer: C At this time, sequencing is used to detect and identify a microbial pathogen.

44

REGENERATIVE MEDICINE, CELL, AND GENE THERAPIES

LIOR GEPSTEIN AND KARL SKORECKI

CELL THERAPY

Introduction and Definitions

A remarkable clinical need exists for the development and clinical assessment of various methods to facilitate the regeneration of injured or diseased tissues and organs. This need derives from the unrelenting prevalence of trauma, congenital disorders, ischemia, and degenerative processes, which becomes increasingly urgent as the global population expands and ages. Cancer is tied to this field both directly (e.g., replacement of lost vital organ function as a result of cancer invasion or treatment modalities, cell-based delivery of cancer immune and gene therapies) and indirectly (e.g., role of stem cells in cancer

pathogenesis, risk for tumorigenesis in stem cell–based therapies). The recent developments in stem cell biology, molecular interventions, biopolymers, and other related biologic and engineering disciplines have paved the way to the emerging research and clinical discipline of regenerative medicine.[1]

Regenerative Medicine

Regenerative medicine seeks to harness methods for the replacement or repair of dysfunctional cells, tissue, or organs in an attempt to restore normal function. It therefore draws on therapies from the three conventional pillars of medical therapeutics (pharmaceuticals, biologics, and medical devices) as well as from the newest platform technology, namely, cell therapy. The long-term goal of regenerative medicine is to cure disease by replacing the lost functions of tissues and organs, and thus it truly represents a paradigm shift from conventional therapies aiming to alter the natural course of disease or to provide symptomatic control. Consequently, regenerative medicine aims to develop curative strategies for unmet clinical needs such as diabetes, heart failure, and neurodegenerative disorders, among others.

Cell Therapy

Cell therapy involves the application of cells to achieve a therapeutic benefit, regardless of the cell type or clinical indication. Although achieving tissue and organ regeneration through cell replacement represents an important goal of cell therapy technology, its applications may reach far beyond the field of regenerative medicine. Hence, the spectrum of cell therapy approaches may range from permanent cell replacement strategies (attempting to replace lost or dysfunctional cells) to more transient cell therapies aiming to modulate disease progression or to protect tissues at risk, to achieve immunomodulatory effects (e.g., for prevention of graft-versus-host disease), to act as vehicles for the delivery of genes or gene products (cell-based gene therapy strategies), and even to act as cell-based cancer vaccines. This chapter focuses on the use of cell therapy for regenerative medicine and specifically concentrates on the potential role of different stem cell types to meet this challenge.

Stem Cells

Stem cells possess two defining properties: (1) the capacity for self-renewal and (2) the ability to differentiate into cell types with specialized cellular functions (Fig. 44-1). This may occur at the individual stem cell level through the process of asymmetrical cell division or at the cell population level wherein a subset of cells differentiate and the remaining stem cells remain dormant or replicate themselves as stem cells. After asymmetrical cell division, non–stem cell derivatives may either generate a pool of organ system–restricted, transit-amplifying cells with enhanced proliferative capacity or continue to differentiate by epigenetic and gene expression profile changes until reaching the terminally differentiated state. This conceptual framework was developed after the discovery of bone marrow cells that were capable of reconstituting the adult hematopoietic system. These hematopoietic stem cells constitute the basis for hematopoietic stem cell transplantation, the only form of stem cell therapy currently routinely well established in clinical practice (Chapter 178).

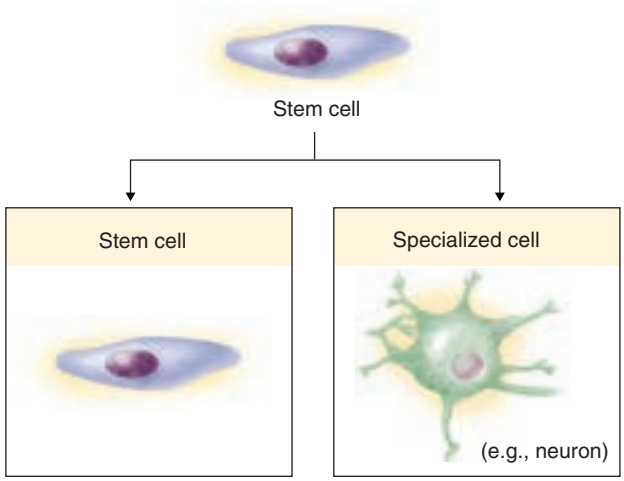

FIGURE 44-1. Asymmetrical cell division. Although this first characteristic was considered a required characteristic for stem cells based on their original description in the adult hematopoietic system, not all cell types currently named as stem cells necessarily display this property. For instance, human embryonic stem cells divide by symmetrical cell division.

The different stem cells types are routinely classified based on the protein or transcription factors they express, but also according to three basic additional attributes. These include replicative capacity (limited vs. unlimited), the scope or potency of differentiation (e.g., pluripotent, multipotent, oligopotent, unipotent), and their place in the life history of the organism (developmental or postdevelopmental). Thus, more recent terminology has broadened use of the term *stem cells* to cover a wider array of cell types that contribute to organ development or have the capacity to repopulate tissues and organ systems. The term *stem cells*, together with the formulations noted previously, has also recently been extrapolated to describe certain cellular subpopulations that may be principally responsible for the growth of malignant tumors. However, because cancer stem cells have no role in tissue regeneration, they are considered in Chapter 181.

Adult (Postnatal) Stem Cells

After birth, many tissues are thought to contain a subpopulation of cells with the capacity for extended self-renewal, combined with the ability to differentiate into more mature cell types with specialized functions (Fig. 44-2). Adult stem cells, thought to represent less than 0.01% of the total number of cells, are located in specialized supportive niche compartments at various sites within the hematopoietic system and elsewhere, and respond to cues in their local microenvironment. As a result of the success of hematopoietic stem cell transplantation in the treatment of bone marrow failure or in conjunction with myeloablative therapy in malignancy, scientists have been motivated to find adult stem cells in other organs. Adult tissues and organ systems reported to contain putative stem cells include bone marrow (hematopoietic and mesenchymal compartments) and peripheral blood, blood vessel endothelium, dental pulp, epithelia of the skin, adipose tissue, digestive system, cornea, retina, testis, and liver. Similar stem/progenitor cells were also reported in organs historically not thought to contain such cells, such as the central nervous system, the heart, and the kidney. Whether adult stem cells represent remnants of developmental stem cells that persist into adulthood for purposes of organ maintenance and repair or represent a distinct cell type dedicated for this latter purpose is not clear. Importantly, in many organs, despite the presence of such tissue-specific stem cells, their regenerative capacity is still inadequate to deal with massive cell loss such as occurs, for example, after a large myocardial infarction or after ischemic brain injury.

Embryonic and Induced Pluripotent Stem Cells

In contrast to adult stem cells that have relatively limited differentiation potency, cells in the developing preimplantation embryo still retain the capacity to differentiate into derivatives of all three germ layers (ectoderm, mesoderm, and endoderm), eventually contributing to all tissues in the body (Fig. 44-3). In normal development, however, such cells do not persist beyond the blastocyst stage. When isolated from unused preimplantation blastocysts generated for in vitro fertilization, the inner cell mass cells isolated can be used to generate human embryonic stem cell (hESC) lines (see Fig. 44-3). The generated hESCs exhibit unlimited self-renewal in cell culture in the undifferentiated state, while retaining the capacity to differentiate into cell derivatives of all three germ layers, essentially giving rise to any cell type in the body. Taking advantage of lessons learned from embryology, scientists were able to utilize the sequential application of different combinations of growth factors to achieve efficient differentiation systems from hESCs, yielding purified populations of different types of neurons, glial cells, cardiomyocytes, vascular endothelial and smooth muscle cells, pancreatic β cells, hepatocytes, different blood cells (platelets, red blood cells), and several other cell lineages.

One of the limitations of the hESC technology is the inability to derive such cells from an adult individual, preventing their utilization in a patient-specific manner. These limitations can be overcome with the introduction of induced pluripotent stem cell (iPSC) technology.[2] This approach allows adult somatic cells (e.g., fibroblasts) to be reprogrammed into pluripotent stem cells by introduction of a set of transcription factors linked to pluripotency (the originally reported combination of factors included OCT3/4, SOX2, c-MYC, and KLF4). The human iPSCs (hiPSCs) generated in this manner can then be coaxed to differentiate into a variety of cell types, using differentiation protocols similar to those already in place for hESC (Fig. 44-4). Importantly, because the hiPSCs can be generated in a patient-specific manner, this technology can potentially be used to develop autologous cell-replacement strategies that can evade the immune system, to generate patient- and disease-specific models of different genetic disorders, and to establish screens for drug testing and drug discovery.

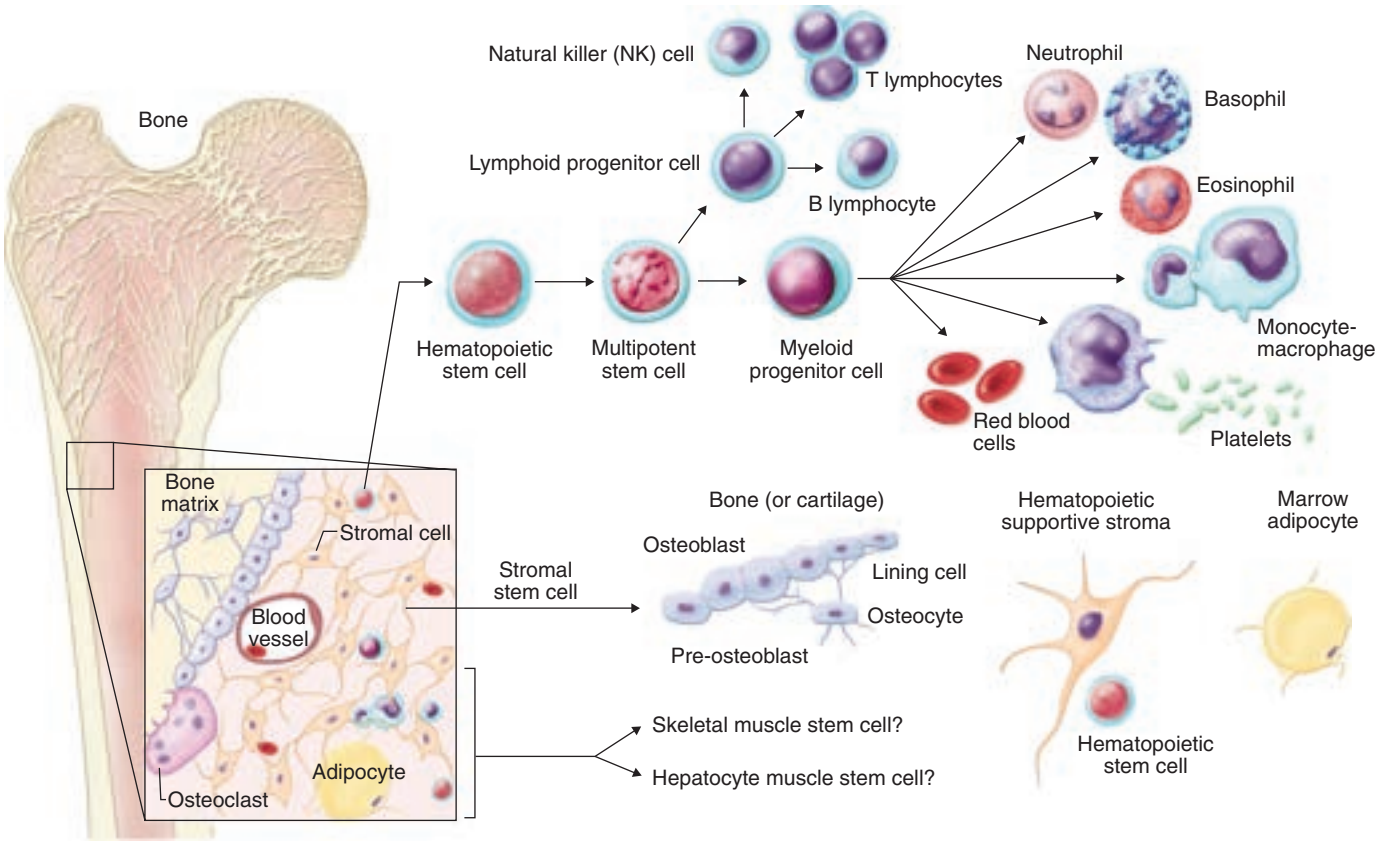

FIGURE 44-2. Adult stem cells. Adult stem cells can be multipotent and have the capacity to differentiate into a limited number of different cell types, often restricted to a given tissue or organ system, as in the case of adult hematopoietic or epidermal stem cells. Two stem cell types have been isolated from adult bone marrow—the hematopoietic stem cell and the mesenchymal stem cell. Adult mesenchymal stem cells of bone marrow origin, although their range of differentiation has been shown to be broader than that of any other adult stem cell type, do not reach pluripotency. It is thought that in some organ systems, such as the gastrointestinal epithelium, a unipotent pool of progenitors exists for repopulating a rapid population turnover of only one type of cell—although it is difficult to be certain whether such progenitors can be distinguished from the overall population of fully differentiated cells in tissues with high cellular turnover.

Cell Therapy Approaches to Regenerative Medicine

Historically, the field of cell therapy can be traced to the transfusion of blood and blood products (Chapter 177), solid organ transplantation (Chapter 49), in vitro fertilization, and bone marrow transplantation (Chapter 178). Nevertheless, beyond the aforementioned therapies, which have become the mainstay treatments in several medical fields, additional cell therapy approaches are considered highly experimental and are still at different stages of preclinical and clinical development. These ongoing efforts can be conceptually grouped into six different approaches (Fig. 44-5).

Delivery of Bone Marrow– and Blood-Derived Stem/Progenitor Cells

A flurry of studies during the past decade evaluated the ability of bone marrow–derived hematopoietic or mesenchymal stem cells to achieve tissue repair after delivery to a variety of organs. These studies were based initially on the assumption that these types of adult stem cells may display some degree of plasticity, allowing them to transdifferentiate into the relevant cell types (e.g., heart cells, nerve cells, and liver cells) after transplantation into the appropriate tissue environment. Although mounting evidence suggests that such transdifferentiation probably does not occur to a significant extent, many of these studies appeared to result in some degree of functional improvement after stem cell delivery to different organs. This clinical benefit may stem from the secretion of different growth factors by the engrafted cells ("paracrine hypothesis"); these factors in turn are thought to augment endogenous tissue repair mechanisms, improve tissue vascularization, modulate inflammation, and protect tissues at risk.

Delivery or Activation of Tissue-Specific Stem/Progenitor Cells or Induction of Cell Proliferation

In contrast to the conventional dogma, recent evidence suggests that a number of organs previously believed to lack any regenerative capacity (e.g., the brain, pancreas, kidney, and heart) in fact do possess such ability, albeit at a limited capacity. Whether this capability is due to the presence of tissue-specific stem/progenitor cells or due to some replication capability of terminally differentiated cells is still a matter of debate for each organ.

Significant efforts have been made in recent years to isolate such putative tissue-specific stem/progenitor cells based on the expression of general or specific stem cell markers or based on their unique culturing properties. These studies also highlighted the potential of such cells to be cultured in a clonal manner and to give rise to one or more cell types relevant to the organs from which they were isolated. Current efforts to utilize the aforementioned findings for regenerative medicine are focused either on the isolation, ex vivo expansion, and transplantation of such putative stem/progenitor cells back to their respective native organs or on the augmentation of their endogenous reparative potential in vivo. The former strategy can be exemplified in the central nervous system where progenitor cells are harvested, cultivated in culture (as neurospheres), and give rise to different types of neurons and supporting glial cells. Similar efforts have followed in other organs. In the heart, for example, such efforts have already reached early clinical trials, in which autologous cardiac stem cells were harvested from the heart, expanded ex vivo, and then engrafted back to the heart. The latter approach, in contrast, aims to influence putative stem cell niches within damaged organs to enhance the endogenous reparative properties of those stem/progenitor cells. Such an effect may underlie the potential therapeutic benefit of bone marrow–derived stem cells after their delivery to different organs.

The final strategy aims to boost endogenous organ repair through the replication of terminally differentiated tissue-specific cells. Such strategies can either augment the inherent physiologic capability of a given organ (e.g., insulin secretagogues for pancreatic β cells) or attempt to induce replication in cells that have already withdrawn from the cell cycle. Caution is warranted with respect to the latter approach because induction of uncontrolled proliferation (e.g., by genetic manipulation) may increase the risk for tumorigenesis.

Engraftment of Fetal Tissue

The most straightforward approach to organ repair would be to replace the missing cells with identical counterparts. Harvesting and expanding adult

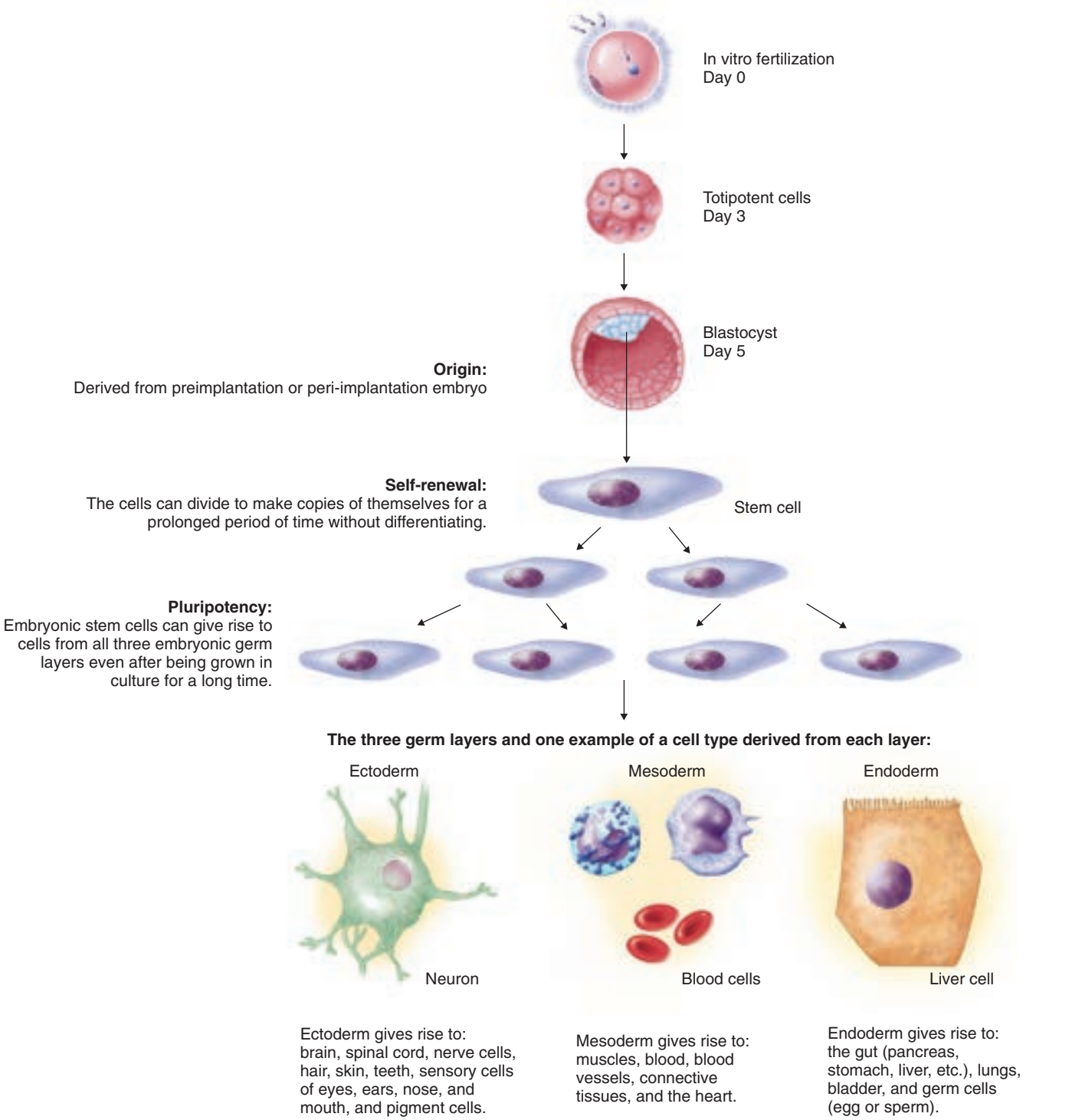

FIGURE 44-3. Embryonic stem cells. ***Totipotency*** **refers to the capacity to differentiate into all cell types in an organism, including extraembryonic tissues, placenta, and umbilical cord, a property confined to the fertilized egg itself, including the cells derived from the first few cell divisions after fertilization.** ***Pluripotency*** **refers to the capacity to differentiate into all the specialized cell types derived from the three germ layers (ectoderm, mesoderm, endoderm) of the developing embryo and is a hallmark feature of embryonic stem and germ cells.**

human cells for transplantation, however, may not be possible in the case of several organs with limited regenerative capacity. During prenatal human development, cells of fetal origin often show enhanced proliferative capacity as well as the ability to differentiate into more than one type of mature or specialized cell. Moreover, animal studies have demonstrated that transplantation of tissues harvested from developing organs (harvested within a specific time window during embryonic development) may give rise to entire functioning organs such as kidneys, lungs, and pancreas. Nevertheless, to date, the only fetal-derived cells that have been used in human clinical applications are the dopaminergic cells derived from the developing fetal nervous system for the treatment of Parkinson disease (Chapter 409). The broader use of fetal tissues for regenerative medicine may be hampered by the limited access to such cells for both technical and ethical reasons, the allogeneic nature of such procedures (requiring immune suppression), and the potential for tumor formation as already described in some case reports.

Transplantation of Ex Vivo Differentiation of Pluripotent Stem Cells

Unlike fetal tissues, hESCs are truly pluripotent (can give rise to advanced cell derivatives of all three germ layers). Importantly, hESCs can be propagated in the undifferentiated state and then coaxed to differentiate into a variety of cell types, giving rise to a potentially unlimited number of specialized cell types for transplantation. Consequentially, numerous preclinical studies have demonstrated the ability of hESC derivatives to engraft, survive,

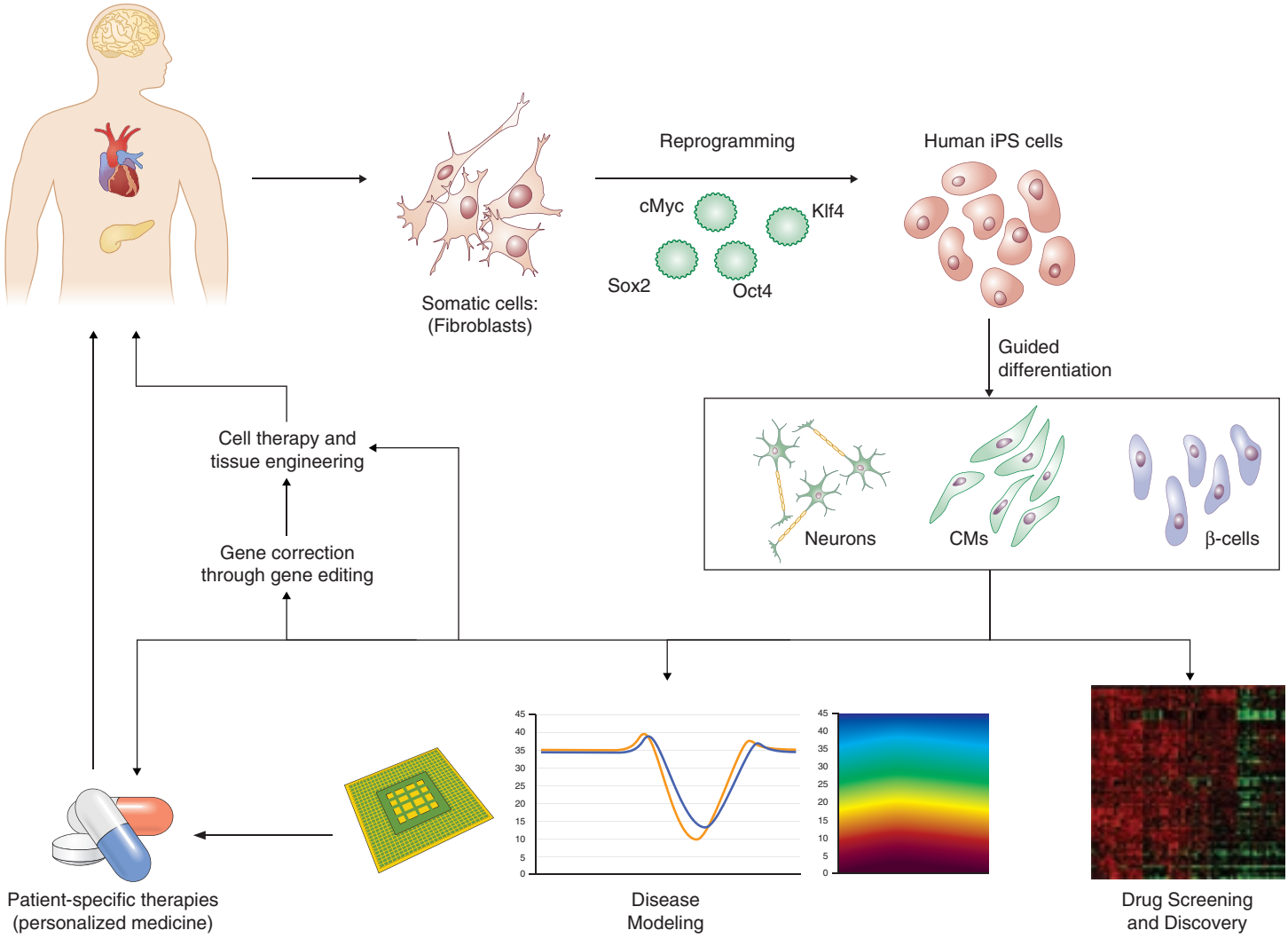

FIGURE 44-4. Application of the induced pluripotent stem cells (iPSC) technology. Patient-specific human iPSC can be generated by reprogramming of adult somatic cells (fibroblasts) with a set of transcription factors and then coaxed to differentiate into a variety of cell lineages. The patient-specific human iPSCs can then be transplanted back to the patient in an autologous manner for regenerative medicine applications. In a similar manner disease- and patient-specific human iPSC models of inherited disorders could be generated ("disease-in-a-dish models") and used for better understanding of genetic disorders, for drug development, and for optimizing patient-specific therapies. Gene editing techniques can be used for mutation correction and for transplantation of healthy cells. CM = cardiomyocytes; iPS = induced pluripotent stem cells.

and improve organ performance in a wide spectrum of relevant animal disease models (e.g., heart failure, Parkinson disease and other neurodegenerative disorders, diabetes). Early clinical studies using hESC derivatives are just emerging and have been focused so far on the retina (transplantation of retinal pigment epithelium [RPE] cells) and spinal cord injury (using oligodendrocyte progenitors).

Despite the significant achievements made with hESCs, the inability to create patient-specific hESCs from adult individuals, the ethical issues arising from destructive use of human embryos, and the anticipated immune rejection associated with such allogeneic cell transplantation impose important hurdles for their clinical utilization. The hiPSC technology provides a potential solution to these challenges. As noted, the patient's own somatic cells (fibroblasts, hair follicles, urine epithelial cells, or blood cells) could be reprogrammed by a set of transcription and chemical factors to yield pluripotent stem cells. The patient-specific hiPSCs could then be coaxed to differentiate to a variety of cell lineages, using protocols similar to those already in place for hESCs. In turn, these differentiated derivatives could then be transplanted either in an autologous or allogeneic manner. Clinical trials using hiPSC-derived cell lineages are expected to be initiated in the coming few years, with the initial targets being macular degeneration (RPE cells), Parkinson disease (dopaminergic neurons), blood product transfusion (hiPSC-derived platelets and red blood cells), and heart failure (cardiomyocytes).

One of the concerns in translating hESCs and hiPSCs into a therapeutic platform is the oncogenic risk. This concern stems from the potential for remaining undifferentiated cells within the cell grafts to form teratomas, from the use of oncogenic reprogramming factors, from the random integration of the viral vectors used in cellular reprogramming ("insertional oncogenesis"), and from genetic instability, potentially leading to both chromosomal aberrations and mutations. Progress to clinical trials requires definitive clarification of this key concern.

Direct Reprogramming

In contrast to the iPSC approach, which seeks to initially reprogram somatic cells to a pluripotent state followed by differentiation of the generated iPSCs to specific cell lineages, recently described direct reprogramming strategies aim to convert the phenotype of one mature cell type (fibroblasts) directly to another. The prototype for such a strategy was the demonstration that *MyoD*, a master regulator of skeletal muscle formation, can convert fibroblasts directly to skeletal muscle. Progress to derive other cell types after this report was delayed for many years because, unlike skeletal muscle, a single master developmental regulatory gene does not exist for most cell lineages.

Based on the experimental approach used to identify the combination of transcription factors that can reprogram somatic cells into iPSCs, researchers evaluated the ability to achieve analogous transcription factor reprogramming strategies to convert the cell fate of somatic cells directly. Consequentially, using a combination of lineage-specific developmental transcription factors, scientists were able to convert terminally differentiated fibroblasts or other somatic cells directly to neurons, β cells, different hematopoietic cell lineages, and cardiomyocyte-like cells. Recent studies have taken this concept a further step forward by demonstrating that transcription factor–based

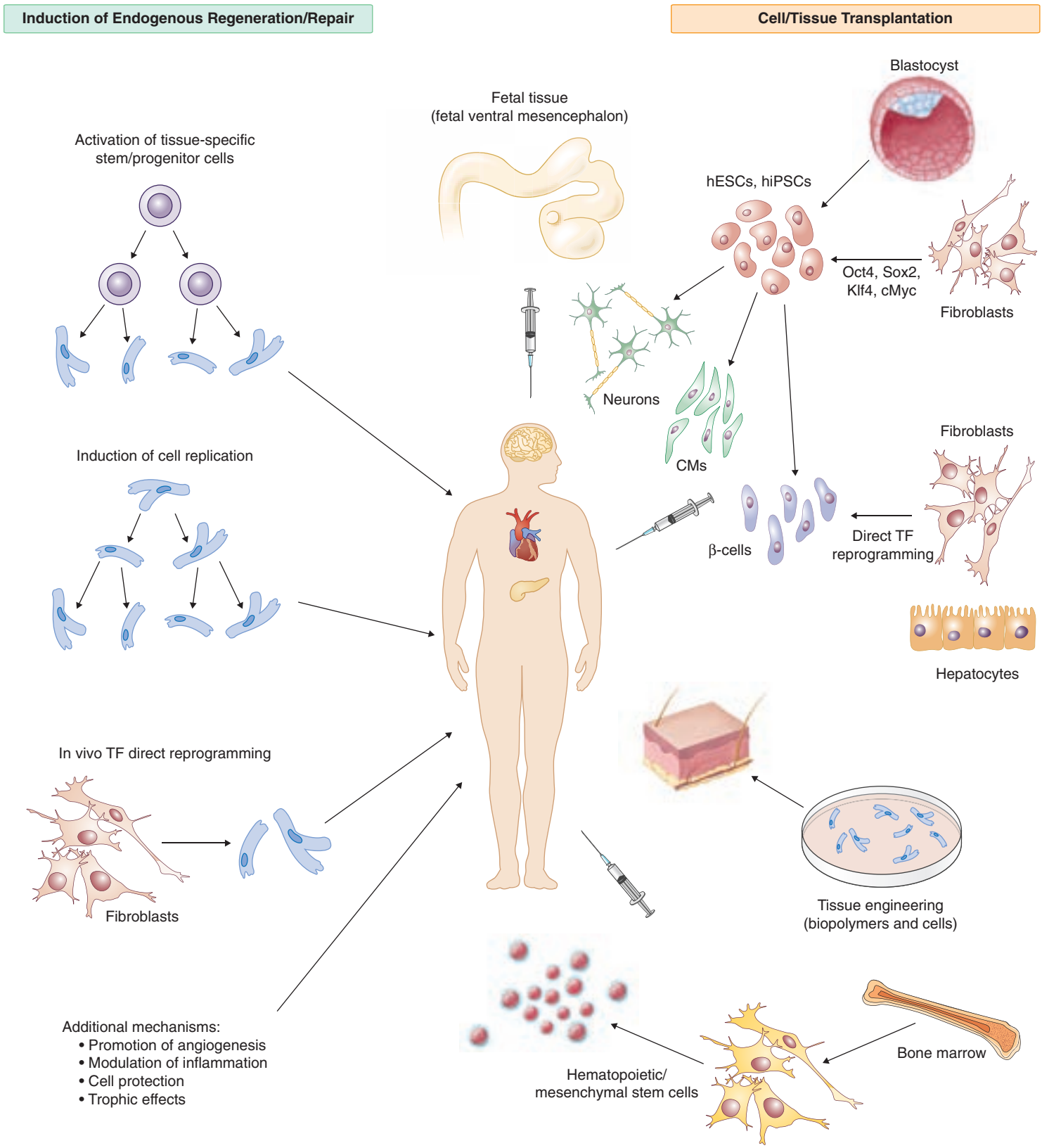

FIGURE 44-5. Conceptual framework for regenerative medicine approaches. These strategies can be divided into those attempting to augment endogenous regeneration *(left side)* and those focusing on transplantation of cells *(right side)*. The former could be achieved through the activation of putative tissue-specific stem/progenitor cells, through induction of cell replication, by in vivo transcription factor (TF)-based direct reprogramming (directly converting one somatic cell [fibroblast] into another), and by several other indirect means (e.g., modulation of inflammation, induction of angiogenesis, trophic effect, protection of tissue at risk). Cell sources that can be used for cell transplantation include fetal tissues (e.g., dopaminergic-rich fetal ventral mesencephalon for Parkinson disease), pluripotent stem cells (human embryonic stem cells [hESCs] and human induced pluripotent stem cells [hIPSCs]), derived cell-lineages, and somatic cells that can be generated ex vivo by direct transcription factor–based reprogramming of fibroblasts. CMs = cardiomyocytes.

transdifferentiation can also be achieved in vivo, suggesting a method whereby resident cells (fibroblasts, hepatic cells, or other cells) could be converted to the appropriate cell types for organ repair. Development of the latter approach for clinical application may be considered more analogous to gene therapy, with the associated advantages, shortcomings, and challenges of this discipline (see later).

Tissue Engineering

Tissue engineering is an interdisciplinary technology combining principles from life sciences and engineering with the goal of developing functional substitutes for damaged tissues and organs.[3,4] Rather than simply introducing cells into a diseased area, in tissue engineering, cells are embedded or seeded onto three-dimensional scaffolds (derived from different biomaterials) before

transplantation. Regardless of the specific clinical application, tissue-engineering strategies usually involve the utilization of combinations of biomaterials, cells, and biologically active factors. The scaffold serves many purposes, including the control of the shape and size of the engrafted tissue, the delivery of biologic signals and adequate biomechanical support to the cells, the induction of vascularization of the graft, and the protection of the cells from physical damage. Scaffolds used in tissue engineering approaches are commonly divided into two general categories: (1) cellular scaffolds that are seeded ex vivo with cells before their in vivo transplantation; and (2) acellular scaffolds that depend on cells in the recipient to repopulate the scaffold with subsequent reconstitution after transplantation. Such tissue-engineered efforts have already reached proof-of-concept clinical trials. These efforts have mainly concentrated on the musculoskeletal system (bone and cartilage repair) but have also targeted other organs such as the heart and even complex organ structures such as the esophagus, trachea, and urinary bladder.

Specific Disease Applications in Cell Therapy

Although a growing number of experimental cell therapies have reached various stages of clinical trials, as yet none has become an established or approved treatment, with the aforementioned exception of hematopoietic stem cells and solid organ transplantation. Nonetheless, with the expectation of significant advances on the horizon, examples of some of the current cell therapy efforts being made in the fields of neurodegenerative disorders, heart failure, and diabetes are provided.

Neurodegenerative Disorders

The central nervous system has limited capacity for regenerating lost tissue both in slowly progressive degenerative neurologic conditions such as Parkinson disease, Alzheimer disease, and amyotrophic lateral sclerosis (ALS) and in acute injuries leading to rapid cell loss (ischemic stroke or traumatic spinal cord injury). Stem cell–based therapies are being explored as potential novel therapeutic paradigms for both acute and chronic neurodegenerative disorders.[5] Consistent with the spectrum of cell therapy–related mechanistic actions described above, these procedures could potentially act through the following mechanisms: (1) cell replacement, whereby cells (precommitted to specific neuronal or glial lineages) are transplanted to replace the specific subtypes of cells that were lost (i.e., dopaminergic neurons in Parkinson disease, motor neurons in ALS, or a mixture of different neuronal and glial subtypes in other disorders); (2) trophic support, whereby the engrafted cells are used to promote the survival of affected neurons or glia or stimulate endogenous repair of the diseased central nervous system through the secretion of neurotrophic factors; and (3) modulation of the inflammatory process thought to contribute to the pathogenesis of many neurodegenerative processes. Achieving the first mechanistic goal, despite being the most attractive, is probably also the most challenging because one would need not only to derive a clinically relevant number of the specific glial or neuronal subtypes or a combination of these cells but also to deliver them to the appropriate site (either focally or diffusely throughout the brain), as well as to assure cell-graft survival, its continuous and appropriate function, and importantly, its integration with the host neuronal network.

Parkinson disease (Chapter 409) involves loss of melanin-containing dopaminergic neurons within the substantia nigra pars compacta of the midbrain, coupled with accompanying depletion of striatal dopamine. This cellular loss is responsible for the major motor features of the disease. In the search for a more definitive therapy than pharmacology, early reports of cell replacement therapy suggested significant improvement in motor function after intrastriatal implantation of mesencephalic dopamine-rich tissue, obtained from aborted human fetuses aged 6 to 9 weeks. Long-term immunosuppressive treatment is essential to allow transplanted dopaminergic neurons to develop into their full functional potential despite the notion of an immunologic sanctuary within the brain. Clinical assessment standards have provided evidence of long-lived graft survival, morphologic and functional integration, and clinical benefit after therapy with cells of fetal origin that have now lasted up to 10 years or longer in some patients. Further progress, however, has been limited by lack of sufficient source tissue to treat a large number of affected patients, prohibitive variability in functional outcome, reports of serious dyskinesias in a subset of treated patients, and ethical considerations.

Given the aforementioned limitations of fetal tissue engraftment, stem cell derivatives could offer a viable alternative for the treatment of Parkinson disease by either replacing the dopaminergic neurons or slowing the degeneration process and restoring the integrity of the nigrostriatal pathway through the release of trophic factors. Importantly, dopaminergic neuroblast-like cells have been generated ex vivo from different stem cell sources, including pluripotent stem cells (hESCs and hiPSCs) by direct fibroblast reprogramming, neural stem cells and progenitors from the embryonic ventral mesencephalon, and adult neural stem cells from the subventricular zone. Preclinical engraftment studies demonstrated that such cells could survive in animal models of Parkinson disease and exert beneficial functional effects after cell maturation. Nevertheless, some properties that are fundamental for successful clinical translation have not been fully met in animal transplantation trials employing human stem cell–derived dopaminergic neurons. Additional challenges that should be addressed include development of methods to prevent the disease process from also destroying the grafted neurons (e.g., engineering the cells to secrete neurotrophic factors) and limiting graft-induced dyskinesia (e.g., by minimizing the number of serotonergic neuroblasts in the grafted tissue).

Investigations of stem cell–based approaches for the treatment of other neurodegenerative diseases, including ALS, Alzheimer disease, Batten disease, stroke, and brain and spinal cord injury, are now moving from experimental animal model studies to planning of clinical trials. Recent reports have shown a major clinical benefit in animal models of directed differentiation and transplantation of hESCs and hiPSCs toward retinal pigment epithelum. Human studies with these cells were recently initiated for patients suffering from macular degeneration.

Heart Disease

Although recent studies have challenged the dogma of the heart being a completely terminally differentiating organ, the endogenous repair mechanisms of the adult heart are usually inadequate in dealing with an extensive myocardial infarction. The resulting decrease in the contractile mass, which is associated with the loss of approximately 1 billion cardiomyocytes, may lead to the development of clinical heart failure (Chapters 58 and 59). With heart failure being the leading cause of hospitalization and with the paucity of donor organs limiting the number of heart transplantations worldwide, it is not surprising that the heart has become the focus of various regenerative medicine efforts.[6]

The first cells that reached clinical trials for heart failure were skeletal myoblasts. Such cells could be harvested (satellite cells) in an autologous manner, expanded ex vivo, and transplanted to the heart. However, skeletal myoblasts display different physiologic properties than cardiomyocytes and cannot form electromechanical connections with host cardiac tissue. Consequentially, these clinical efforts have largely been abandoned because of lack of efficacy as well as evidence suggesting increasing arrhythmogenicity in some patients.

The largest clinical experience in myocardial cell therapy comes from the use of bone marrow–derived stem cells (primarily hematopoietic stem cells and more recently also mesenchymal stem cells). The effects of delivery of such cells (mainly through the coronary circulation) were studied in thousands of patients, primarily in the setting of acute or recent myocardial infarction. These studies revealed either a neutral effect on myocardial performance or mild functional improvement. Although a recent meta-analysis of 33 randomized controlled trials studying transplantation of adult bone marrow–derived cells revealed a statistically significant improvement in left ventricular ejection fraction, this improvement was not associated with a change in mortality.

Bone marrow–derived stem cells are thought to exert their beneficial effects through the secretion of different growth factors rather than transforming to become new heart cells. Consequentially, a cell source that could truly re-muscularize the heart is direly needed. A potential candidate for such a task could be the recently described cardiac progenitor cells. Several reports have described cardiac progenitor cells as multipotent clonogenic cells that could be isolated based on different markers or culturing properties and potentially differentiate into cardiomyocytes and vascular cells. Such cells can be harvested from the heart (during surgery or a percutaneous cardiac catheterization biopsy approach), expanded ex vivo, and then transplanted back to the left ventricle in an autologous manner.

In contrast to the aforementioned cell types, human pluripotent stem cell lines (hESCs and hiPSCs) can undoubtedly become cardiomyocytes during ex vivo differentiation. Research efforts in recent years established efficient directed differentiation systems that could give rise to clinically relevant numbers of cardiomyocytes and demonstrated the ability of the generated cells to engraft, functionally integrate with host cardiac tissue, and improve myocardial performance in animal models of myocardial infarction. Nevertheless, issues related to ethics and the allogeneic nature of the graft (hESCs),

to the inefficient and incomplete reprogramming process (iPSCs), to the heterogeneous and relatively immature properties of the generated cardiomyocytes, to the tumorigenic risk, and to the complex regulatory and financial issues have hindered clinical development of these cells to date.

Most recent efforts in the field have focused on attempting to induce mature cardiomyocytes to reenter the cell cycle (directly or after an initial dedifferentiation phase) or to convert the phenotype of nonmyocytes (fibroblasts) into cardiomyocytes. In the latter approach, recent studies have demonstrated the ability to convert the phenotype of murine fibroblasts both in vitro and in vivo into cardiomyocyte-like cells by the expression of a combination of cardiomyocyte-specific transcription factors (GATA-4, MEF-2C, and TBX-5 in one study). Although these and other efforts have the potential to augment the number of cardiomyocytes and consequentially improve contraction of the failing heart, they are still in the early phase of discovery.

Diabetes Mellitus

Successful pancreatic transplantation and improved glucocorticoid-free protocols for transplantation of islets of Langerhans have been shown not only to restore glucose control in patients with diabetes mellitus but also to prevent or even reverse some of the disease's complications (Chapters 229). However, whole organ or islet-based transplantation approaches are limited both by immunologic rejection and by limitation of an available source of transplantable tissues. This has motivated the search for cell types that can replace (type 1 diabetes mellitus) or augment (type 2 diabetes mellitus) deficient β-cell function.[7]

The development of hESCs and hiPSCs, coupled with improved understanding of β-cell development, has provided a potentially unique cell source to derive β cells for transplantation therapy. β cells make an especially attractive case for cell replacement strategies because only a single cell type is missing, cell replacement does not necessarily need to be performed in the native environment (pancreas), and, theoretically, such cells could even be engrafted subcutaneously. Harnessing lessons from embryology, efficient protocols were developed to promote differentiation of pluripotent cells in vitro into precursor or early-stage β-cell phenotype. More recent efforts have moved the field even closer to clinical application by tackling the challenge of creating more mature and functional β cells.

One of the problems with using β cells for cell replacement therapy is that the autoimmune destruction of endogenous β cells, which underlies the pathogenesis of type 1 diabetes, will probably also result in the destruction of the pluripotent cell–derived β cells, even when derived from an autologous (hiPSCs) source. Consequentially, significant efforts are being made to develop the biotechnologic means (encapsulation technologies) to deliver the cells in an immunoprotective environment that will prevent cell rejection but will retain the capacity of the engrafted β cells to sense glucose and to secrete insulin.

Beyond the derivation of new β cells from pluripotent stem cells, progress has also been made in reprogramming closely related cell types to β cells by the overexpressing of master regulatory transcription factors. Early studies focused on the conversion of hepatocytes to β-like cells through the overexpression of PDX1, the transcription factor MAFA, and NeuroD. In vivo transdifferentiation of mouse acinar cells to β cells has been achieved by transient viral overexpression of three transcription factors (PDX1, NGN3, and MAFA), whereas overexpression of a single transcription factor, PAX4, has successfully converted murine α cells to β cells.

Regenerative strategy focuses on increasing pancreatic β-cell mass by inducing the replication of existing β cells. This therapeutic approach would probably mainly target type 2 diabetic patients by decreasing the burden on existing overworked β cells but may also be beneficial for some patients with type 1 diabetes who still retain some β-cell mass. Whereas several tissues are regenerated by differentiation of tissue-specific stem cells, new pancreatic β cells are derived from the replication of existing β cells. Promising candidates for augmenting β-cell replication were recently identified and include the use of glucokinase activators or betatrophin, a protein secreted by the liver.

Stem Cell–Derived Platforms for Disease Modeling, Personalized Medicine, and Drug Discovery

In addition to the generation of cells for regenerative applications, the ability to grow a wide variety of different specialized cell types of human origin in culture provides unparalleled opportunities for gene and drug discovery and testing. For example, the ability to grow human cardiomyocytes in culture provides a preclinical human cellular-based experimental platform for screening newly developed drugs in terms of their potential to cause QT-interval prolongation and hence the risk for arrhythmia in the clinical setting. Other examples include the creation of an experimental tissue microenvironment of human origin for studying the stromal response to tumor growth and testing anticancer drugs that target tumorigenic responses such as angiogenesis.[8]

The hiPSC technology has further revolutionized this field because it allows for the first time the generation of disease/genotype- and patient-specific hiPSC models of a wide array of inherited disorders. Initial studies focused on diseases with monogenic inheritance, but more recent studies have included diseases with more complex inheritance patterns.[9] Consequently, different types of patient-specific hiPSC-derived neurons, cardiomyocytes, skeletal muscle, blood cells, hepatocytes, and other cell types were demonstrated to recapitulate in a culture dish the abnormal phenotype of a wide array of genetic disorders, including neurodegenerative disorders (e.g., spinal muscular atrophy, familial dysautonomia, ALS, schizophrenia, and even late-onset diseases such as Parkinson and Alzheimer disease), different cardiomyopathies and arrhythmogenic syndromes, a wide array of blood disorders, and several other genetic disorders. These models have already yielded important insights into the mechanisms underlying these disease states and have established unique experimental platforms that will enable the testing of existing therapies in a patient-specific manner (personalized medicine) to evaluate evolving therapies ("clinical studies in the culture dish") and to develop new therapeutic strategies.

GENE THERAPY

Gene therapy can be broadly defined as the transfer of genetic material into cells to restore or correct a cellular dysfunction or to provide a new cellular function in an attempt to cure a disease or at least to improve the clinical status of a patient. The use of genes as therapeutic platforms emerged during the mid-20th century, and in the 1990s, the first regulated registered studies were performed in the United States. In the first clinical study, a 4-year-old girl with adenosine deaminase (ADA) deficiency was treated by transfecting the ADA gene into her white blood cells, resulting in improvements in her immune system. Since then, more than 10,000 patients have been involved in more than 1700 gene therapy clinical studies performed throughout the world. The most common patient populations targeted in these studies have been cancer patients (more than 1000 studies), with another important category being monogenic inherited disorders (more than 100 studies). Although gene therapy initially was conceived as a way to treat life-threatening disorders (inborn errors, cancers) refractory to conventional treatment, it is now being explored for non-life-threatening conditions that adversely affect a patient's quality of life.

Although early clinical failures and a number of reported deaths (only two of which were actually attributed directly to gene therapy) and cases of gene therapy–related leukemic transformation led many to dismiss gene therapy as hazardous and premature, recent clinical successes have bolstered new optimism in the promise of this discipline. These include entirely novel initiatives in treating primary immunodeficiency syndromes, the improvement of vision in patients with the retinal disease (e.g., Leber congenital amaurosis), the successful treatment of X-linked adrenoleukodystrophy,[10] and the encouragement of experimental results in treating different forms of cancer.

Despite these success stories, only a few gene therapy agents are currently approved and available. Fomivirsen (Vitravene) is used for the treatment of cytomegalovirus retinitis in patients with AIDS. In 2012, Glybera became the first gene therapy treatment to be approved for clinical use in either Europe or the United States. Glybera uses a virus injected into a patient to deliver a working copy of a gene for producing lipoprotein lipase (LPL) to treat the rare inherited disorder of LPL deficiency. Finally, the *p53* tumor suppressor coding sequence in an adenovirus vector is used for the treatment of head and neck cancer but is registered only in China.

Classifications and Mechanisms of Action

In general, somatic gene therapy applications can be divided into those aiming to treat or correct various genetic disorders and those targeting non-genetic diseases by attempting to alter cell, tissue, and organ function in a favorable manner. According to the World Health Organization, there are more than 10,000 disorders with monogenic inheritance described in humans (http://www.who.int/genomics/public/geneticdiseases/en/index2.html), but only a small fraction of these may be amenable to gene therapy. Traditional gene therapy efforts for inherited disorders have mainly focused on the

exogenous expression of genes encoding the missing or abnormal proteins and to a lesser extent also on altering the abnormal gene expression patterns. Future efforts are expected to shift the focus from uncontrolled overexpression of the missing protein to directly correcting the mutation at the DNA level (gene-editing strategies) in affected cells, using the newly emerging technologies known as the TALEN and CRISPR[11] approaches described in greater detail later under Gene Editing. With successful widespread use in research studies and proven applications in editing of gene sequence in stem cells, the transition to clinical application is sure to follow.

For nongenetic disorders, gene therapy efforts are aimed at overexpressing a specific protein in an attempt to alter cellular function favorably (e.g., to increase contractility in heart failure by overexpression of the sarcoplasmic reticulum calcium ATPase SERCA2a), protect tissue at risk (e.g., in acute kidney injury), exert paracrine effects through local secretion of specific proteins by the engineered cells (e.g., promote angiogenesis in ischemic tissues or induce neurotrophic effects in neurodegenerative disorders), and even secrete proteins systemically (e.g., in gene therapy trials attempting to correct bleeding disorders by secretion of coagulation factors or for systemic delivery of hormones such as erythropoietin for the treatment of anemia). Major efforts in the gene therapy arena to date have been in designing various methods to treat cancer (see later).

Progress in the field of gene therapy has developed into two different strategies: ex vivo and in vivo gene therapy. The ex vivo gene therapy approach (combined cell and gene therapy strategy) involves the initial harvesting of cells from a given patient followed by genetic modification of these cells in the laboratory. The genetically modified cells can then be selected, amplified in numbers, and returned to the same patient in an attempt to achieve the desired therapeutic effect. This strategy is particularly attractive for the genetic modification of stem cells that could reconstitute the relevant tissues, organs, and organ systems after transplantation. The most prominent example is using hematopoietic stem cell grafts in gene therapy trials for hematopoietic disorders.

The in vivo gene therapy approach, in contrast, involves the delivery of the relevant transgene (through the use of various vectors) directly to the targeted tissue, followed by the stable or transient expression of the transgene in the relevant cells. The expression of the transgene only in the relevant cells/tissues can be achieved by a combination of localized delivery (injection), a particular tropism of the vector used for the tissue of interest, and the expression of the transgene under the control of a cell/tissue-specific promoter.

Gene Therapy Delivery Methods

Gene therapy agents are often composed of two elements: the genetic material itself (i.e., the DNA expression cassette [the most common therapeutic payload used], short interfering RNA, or an antisense molecule) and the vector delivery system. The latter is usually the more complex and limiting component, and it is important to select the most efficient delivery method for any genetic therapy as well as to be aware of the potential adverse effects of each vector type, thus tailoring the therapy to specific clinical considerations. There are formidable barriers to successful gene transfer, such as crossing the cellular membrane, escaping from the endosome, moving through the nuclear membrane, and integrating into the host genome. Vectors that have been developed to try to overcome these obstacles fall into two broad categories: nonviral and viral vectors.

Gene therapy mediated by nonviral vectors is referred to as *transfection* and consists of the direct delivery of naked DNA by injection, the use of liposomes (cationic lipids mixed with nucleic acids), nanoparticles, and other means. Although nonviral vectors can be produced in relatively large amounts and are likely to present minimal toxic or immunologic problems, their major shortcoming is inefficient gene transfer. In addition, expression of the foreign gene tends to be transient, precluding the application of nonviral vectors to many disease states in which sustained and high-level expression of the transgene is required. The efficiency of nonviral vector delivery could be enhanced by the use of different physical methods that have evolved, such as electroporation (for well-circumscribed body compartments or masses such as muscle, skin, and tumors), gene gun (for DNA vaccination), and ultrasound delivery (for cardiovascular and tumor-related applications).

Gene therapy mediated by viral vectors is referred to as *transduction*, and this approach has been the main conduit for transferring genes to human cells in most gene therapy trials. The basic concept of viral vectors is to harness the innate ability of viruses to deliver genetic material into the infected cell. Viruses used in gene therapy have been modified to enhance safety, increase specific uptake, and improve efficiency. However, for each specific virus-based gene therapy vector, there have been major disadvantages that should be balanced against potential therapeutic benefits. For example, in cancer gene therapy, the immune response to the delivery vehicle carrying the anticancer genetic material can be used to advantage by serving as an adjuvant. In contrast, the system for delivery of a gene to be expressed for a prolonged period to replace or supplement a missing gene product in monogenic disease states should preferably be ignored by the immune system.

Viral vectors are derived from viruses with either RNA (retroviruses and lentiviruses) or DNA (adenovirus, adeno-associated virus [AAV], herpes simplex virus [HSV], and poxvirus [vaccine virus]) genomes. Viral vectors also fall into one of two main categories: integrating vectors, which insert themselves into the recipient's genome, and nonintegrating vectors, which often (although not always) form an extrachromosomal genetic element. Integrating vectors, such as γ-retroviral vectors and lentiviral vectors, are generally used to transfect actively dividing cells because they are stably inherited. Integrating vectors, however, may carry the risk for insertional mutagenesis (with clinical oncogenic transformations reported with the use of retroviruses). Nonintegrating vectors, such as adenoviral vectors and AAV vectors, can be used to transfect quiescent or slowly dividing cells, but they are quickly (in the case of adenoviral vectors) lost from cells that divide rapidly. Finally, efficient gene transduction can also be achieved using vectors that are maintained as episomes, especially in nondividing cells.

Adenoviral vectors and retroviral vectors based on Moloney murine leukemia virus featured prominently in early gene therapy trials. There has been a movement away from both, however, after the case of a fatality, which was linked to the toxicity of the adenoviral vector (used to introduce the ornithine transcarbamylase gene in that specific study) and the leukemia cases in SCID-X1 patients (which were linked with activation of *LMO2*, an oncogene on chromosome 11, due to insertional mutagenesis associated with the murine leukemia viral vector). Consequentially, these vectors have been largely been replaced with AAV and lentiviral vectors, respectively, which have become the most common vectors used in clinical trials today. Other viral vectors may have applications in specific settings. For example, in gene therapy applications being developed for pain management, a replication-defective HSV vector is being used because of its tropism for nerve tissues. Also, different oncolytic viruses with a preferential tropism to cancer cells are being used for gene therapy applications in cancer.

Diseases Treated by Gene Therapy

Inherited Immunodeficiency

More than 30 patients reported to date worldwide have undergone treatment with different retroviral vectors for inherited immunodeficiencies (Chapter 250). Patients with one of the following three diseases are included in this group: two types of severe combined immunodeficiency (SCID), both of which are characterized by dysregulation of lymphocyte development, and X-linked chronic granulomatous disease (X-CGD), an inherited immune deficiency with absent phagocyte reduced nicotinamide adenine diphosphate oxidase activity caused by mutations in the *gp91* (*phox*) gene. Individuals with adenosine deaminase (ADA) SCID suffer from premature death of T, B, and natural killer (NK) cells as a result of the accumulation of purine metabolites; patients with this condition have been treated with vectors expressing the *ADA* gene. In the first patients with ADA SCID, transduced T cells expressing transgenic *ADA* have been shown to persist for longer than 10 years; however, the therapeutic effect of gene therapy resulted in incomplete correction of the metabolic defect. More recently, an improved gene transfer protocol of bone marrow CD34-positive cells, combined with low-dose busulfan, resulted in multilineage, stable engraftment of transduced progenitors at substantial levels, restoration of immune function, correction of the *ADA* metabolic defect, and proven clinical benefit.[12] Overall, no adverse effect or toxicity has been observed in patients treated with *ADA* gene transfer in mature lymphocytes or hematopoietic progenitors.

The X-linked type (X-SCID group), in which there is defective cytokine-dependent survival signaling in T and NK cells, was shown to be corrected by introduction of the wild-type sequence of the common γ-C chain, which is an essential component of five cytokine receptors. In one clinical study, hematologic malignancies developed in four patients. One of the four died of this complication. Ten patients were successfully treated with a different viral transduction protocol, with one reported malignancy in up to 8 years of follow-up. Two adult X-CGD patients who suffered recurrent bacterial infections have been treated with CD34-positive cells transduced with a γ-retroviral vector expressing *gp91 phox*, with significant clinical improvement in the

short term. However, in both these patients, there was an expansion of gene-transduced cells caused by the transcriptional activation of growth-promoting genes leading to myelodysplasia and gradual loss of efficacy. In summary, of the nearly 30 patients worldwide treated with gene therapy for immunodeficiency disorders, significant clinical improvement has been observed in many. However, severe and even life-endangering adverse consequences have been encountered with certain viral vectors and protocols. Additional clinical information from long-term observation and new clinical studies will be important for a clearer assessment of clinical benefit.[13]

Visual Loss

Both cell- and gene-based therapy approaches are leading areas for promising inroads in retinal disease. Although early trials of stem cell–based retinal cell therapy have not yet achieved proof of efficacy, at the level of gene therapy, clinical scientists have used gene augmentation therapy with direct subretinal injection of a recombinant AAV expressing *RPE65* complementary DNA in adults and children with Leber congenital amaurosis. This rare inherited eye disease destroys photoreceptors (Chapter 424), and the gene therapy results have shown medical evidence of visual preservation despite continued retinal degeneration.[14]

Cardiovascular and Pulmonary Conditions

Gene therapy efforts in the cardiovascular field have focused on achieving therapeutic angiogenesis in patients suffering from chronic ischemic heart disease or from critical limb ischemia (CLI) and for improving cardiac function in heart failure patients. The use of genes to revascularize the ischemic myocardium due to coronary artery disease and CLI due to peripheral artery disease has been the focus of two decades of preclinical research with a variety of angiogenic mediators, including vascular endothelial growth factor, fibroblast growth factor, hepatocyte growth factor, and others, encoded by DNA plasmids or adenovirus vectors. Overall, these gene therapy studies in animal experimental models of ischemia were very encouraging, leading eventually to several clinical trials. Despite the established proof of concept and reasonable safety, however, results of the latest clinical trials on therapeutic angiogenesis for myocardial ischemia and CLI have provided inconsistent results, and the definite means of inducing clinically useful therapeutic angiogenesis remain elusive. These less than optimal results may stem from a number of reasons, including the application of a single growth factor that may not be sufficient to meet the multifaceted challenge for developing efficient induction of collateral vessels, the need for more sustained growth factor delivery in order to establish more stable vessels, and the need to target arteriogenesis rather than angiogenesis to achieve a more significant increase in perfusion. Therefore, efforts in the field are moving toward the use of different cell therapies for these ischemic conditions, as well as using combined cell and gene delivery strategies to achieve better outcomes. For example, a recent trial has used combined delivery of endothelial and smooth muscle cells (each cell type modified to secrete a different angiogenic growth factor) in CLI patients.

For heart failure, gene therapy trials have focused on restoring the abnormal calcium handling characteristic of failing human cardiomyocytes.[15] Because a reduction in levels of the sarcoplasmic reticulum calcium ATPase (SERCA2a), the sarcoplasmic reticulum calcium pump, was found to be a key factor in the alteration of calcium cycling in heart failure, this protein became an attractive clinical target for gene delivery purposes. Overexpression of SERCA2a levels by cardiomyocyte gene delivery has led to the restoration of previously abnormal calcium transients and to improved cardiac contractility, reduction of the frequency of arrhythmias, and improved oxygen utilization in animal models of heart failure. More recently, the clinical benefits of overexpressing SERCA2a have been demonstrated in phase I and II of the Calcium Upregulation by Percutaneous Administration of Gene Therapy in Cardiac Disease (CUPID) trials. These studies demonstrated that AAV delivery of the SERCA2a transgene by intracoronary delivery is feasible and safe, results in persistent expression of the transgene, and is associated with a significant improvement in associated biochemical alterations and clinical symptoms of heart failure in the treated patients.

Cystic Fibrosis

Experimental protocols for gene therapy for cystic fibrosis (CF) (Chapter 89) have been implemented since 1990. The cystic fibrosis transmembrane conductance regulator (CFTR) protein is mutated in patients with CF. Transducing the epithelium of the nasal and bronchial tree is potentially feasible through nonsystemic approaches. Nonviral gene therapy methods that deliver a copy of the *CFTR* gene to the airway of CF patients have been developed. Several placebo-controlled clinical trials of liposome-mediated *CFTR* gene transfer to the nasal epithelium have confirmed its safety and demonstrated variable degrees of functional correction. In addition, several clinical studies have assessed the potential of retrovectors, adenovectors, and AAV vectors for gene therapy for CF. With both nonviral and viral delivery systems, there were only mild side effects. However, the long-term clinical benefit has been marginal. Improved vectors are being assessed in preclinical studies.

Cancer

One of the most exciting opportunities for gene therapy lies in the cancer arena. Gene therapy strategies targeting cancer can be grouped according to their proposed mechanisms of action and include gene therapies aiming to directly induce cytotoxic effects in cancer cells (through the use of oncolytic viruses or by the delivery of apoptotic inducers and suicide genes), gene therapies aiming to boost the immune response to tumor antigens, and gene therapies targeting the tumor microenvironment.

Direct Cytotoxic Effects

An interesting approach for cancer gene therapy is to harness the action of oncolytic viruses.[16] Oncolytic viruses are therapeutically useful anticancer viruses that will selectively infect, amplify, and then damage cancerous tissues without causing harm to normal tissues. Cancer selectivity of the different oncolytic viruses takes advantage of defects commonly found across many tumor types, such as lack of antiviral responses, activation of Ras pathways, loss of tumor suppressors, and defective apoptosis. Oncolytic viruses can kill infected cancer cells in many different ways, ranging from direct virus-mediated cytotoxicity through a variety of cytotoxic immune effector mechanisms.

Several viruses such as the Newcastle disease virus (which activates the innate or adaptive immune response), reovirus (which activates host protein kinases to shut down protein production), and mumps virus have an inherent ability to specifically target cancer cells and, upon virus replication, cause significant cell death and tumor regression. Other viruses (HSV, adenovirus, vaccinia virus, vesicular stomatitis virus, and poliovirus) need to be genetically engineered to engender oncolytic activity. Genetically engineered viruses and inherently antitumor-selective viruses are being tested in early and late clinical conditions to determine their effectiveness in specific types of cancer (e.g., metastatic melanoma and different brain tumors).

Beyond the direct viral cytopathic effect, viral vectors can be used to deliver genes to cancer cells that will result in tumor cell death. The relevant transgenes encode for cellular proteins that are involved in apoptosis or prevent proliferation. The selectivity for the activation of such genes only in tumor cells is achieved either through the use of the aforementioned oncolytic viruses or by the expression of the transgenes under the control of promoters that are activated only in cancer cells, either as a general property of cancer (e.g., human telomerase or survivin) or in specific types of tumors (probasin in prostate cancer, ceruloplasmin in ovarian cancer, HER2 in breast cancer, and carcinoembryonic antigen in colon cancer). The most clinically advanced gene therapy drug against cancer is the replication-deficient adenovector expressing the human *p53* gene. This therapy (Gendicine) is approved in China for the treatment of patients with head and neck squamous cell carcinoma by direct administration into the tumor bed.

Another attractive approach is the use of suicide genes. Suicide gene therapy involves delivery of a pro-drug activating enzyme (suicide gene) that converts nontoxic pro-drugs to cytotoxic metabolites. The prototype for such a suicide gene/pro-drug combination is HSV thymidine kinase (TK)/ganciclovir (GCV). The TK gene is selectively expressed only in cancer cells (by one of the methods described previously), and after application of GCV, it converts it to the cytotoxic agent phosphorylated GCV. Interestingly, phosphorylated GCV is only toxic to dividing cells, further increasing the selectivity to the cancer cells. Other cytotoxic strategies are to express secreted pro-apoptotic proteins, such as tumor necrosis factor–related apoptosis-inducing ligand (TRAIL) or cytotoxins such as *Pseudomonas* exotoxin.

Immunomodulatory Cell and Gene Therapy for Cancer and Autoimmune Disease

In recent years, the focus of gene- and cell-based therapy for cancer has shifted away from directly manipulating or targeting the cancer cells toward modulation of the immune system itself. Cytotoxic T-lymphocyte antigen 4 (CTLA-4) and programmed death 1 (PD-1) are two T-lymphocyte

proteins that have long been known to attenuate immune destruction of cancer cells. Blocking monoclonal antibodies to circumvent this attenuation have been shown to induce limited remissions in several forms of previously intractable metastatic tumors, including malignant melanoma. These partial successes have now revived still more sophisticated therapeutic approaches, based on the ex vivo personalized genetic engineering of cytotoxic T lymphocytes of cancer patients to enable the immune system to target tumor cells. Chimeric antigen receptor (CAR) therapy releases the encumbrance of major histocompatibility complex restriction in cancer antigen recognition by combining the antigen-binding site of a monoclonal antibody with the signal-activating machinery of the cytotoxic T lymphocytes. This enables combining a high level of target specificity typical of monoclonal antibodies with in vivo expansion and the potential for a durable response, as has been demonstrated in clinical treatment protocols in leukemia and other malignancies.[17] It can be anticipated that CAR-modified T lymphocytes might also prove useful as a combined genetic engineering/cell therapy approach to the management of autoimmune disease.

Disrupting Tumor Microenvironment

Targeting the tumor microenvironment is another attractive approach for cancer gene therapy because it consists of normal cells that should not develop resistance to the therapy. The most obvious target is the tumor neovascularization process. The use of antiangiogenic drugs such as bevacizumab (Avastin), an anti–vascular endothelial growth factor monoclonal antibody, has shown success in clinical trials for some cancer cell types, but the effect may be transient or negligible in others. This may be because the angiogenesis process is complex, and inhibiting just one aspect may not be sufficient. Developing alternative strategies such as combination therapies, including targeting multiple angiogenic pathways, might be a better strategy, especially because inhibiting angiogenesis is cytostatic and not cytotoxic. A number of antiangiogenic factors (e.g., angiostatin) have been expressed in viral vectors and have been used in preclinical studies but have not reached the clinic yet.

Other Forms of Molecular Therapies: RNA Interference and Gene Editing

RNA Interference

RNA interference (RNAi) regulates gene expression by a highly precise mechanism of sequence-directed gene silencing at the stage of translation by degrading specific messenger RNAs or by blocking their translation into protein. Research on the use of RNAi for therapeutic applications has gained considerable momentum. It has been suggested that many of the novel disease-associated targets that have been identified are amenable to conventional small molecule drug blockade and can potentially be targeted with RNAi. In the coming years, the concept of RNAi will be actively translated into a therapeutic option, with numerous early-phase trials already underway.

Gene Editing

The center of gravity for gene therapy may be shifting from gene restoration (where a whole new gene is pasted into the genome) to genome editing, whereby the pathogenic mutation is corrected in its natural gene location with zinc finger nucleases, transcription activator–like effector nucleases (TALENs), or clustered regularly interspaced short palindromic repeats (CRISPRs).[18,19] These hybrid molecules act as highly specific "molecular scissors," which are engineered to target a specific location in the genome and introduce a double-strand break in the DNA proximal to the targeted mutation. The cleavage in the DNA is then resolved by homologous recombination between the endogenous genes and an exogenously introduced donor fragment containing the normal sequence. In this fashion, the pathogenic mutation is permanently changed back to the normal sequence. This also preserves the architecture of the genome and maintains gene control under the normal cellular regulatory elements.

Consequentially, gene editing represents a paradigm shift in the way gene therapy could be performed. To date, gene editing techniques have been used to correct the disease-causing mutations associated with X-linked SCID, hemophilia B, sickle cell disease,[20] and α_1-antitrypsin deficiency and to repair Parkinson disease–associated mutations (*SNCA* gene) in patient-derived hiPSCs or in preclinical mouse models. Targeted gene knockout through similar technologies promises to be a potentially powerful strategy for combating HIV/AIDs. Zinc finger nucleases have been used to confer HIV-1 resistance by disabling the HIV coreceptor C-C chemokine receptor type 5 (CCR5) in primary T cells and hematopoietic stem/progenitor cells. This approach is currently used in clinical trials. Additionally, zinc finger nucleases have been used to improve the performance of T-cell-based immunotherapies by inactivating the expression of endogenous T-cell-receptor genes, thereby enabling the generation of tumor-specific T cells with improved efficacy profiles.

Finally, site-specific nucleases may also bring a unique value to the conventional gene-adding approach by enabling insertion of therapeutic transgenes into specific "safe harbor" locations in the human genome, ensuring long-term expression of the transgene as well as reducing the potential for random insertional mutagenesis.

It is important to mention that the use of site-specific nuclease technology at its current state requires the presence of proliferating cells, and its utility is therefore still relatively limited for nonproliferating somatic cells and for direct in vivo applications. Continued progress in stem cell research, including the production and manipulation of hiPSCs cells, will ultimately open countless new directions for gene therapy, including treatments based on autologous stem cell transplantation.

Grade A References

A1. Clifford DM, Fisher SA, Brunskill SJ, et al. Stem cell treatment for acute myocardial infarction. *Cochrane Database Syst Rev*. 2012;2:CD006536.

A2. Fisher SA, Brunskill SJ, Doree C, et al. Stem cell therapy for chronic ischaemic heart disease and congestive heart failure. *Cochrane Database Syst Rev*. 2014;4:CD007888.

A3. Wang ZX, Li D, Cao JX, et al. Efficacy of autologous bone marrow mononuclear cell therapy in patients with peripheral arterial disease. *J Atheroscler Thromb*. 2014;21:1183-1196.

A4. Zsebo K, Yaroshinsky A, Rudy JJ, et al. Long-term effects of AAV1/SERCA2a gene transfer in patients with severe heart failure: analysis of recurrent cardiovascular events and mortality. *Circ Res*. 2014;114:101-108.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Daley GQ. The promise and perils of stem cell therapeutics. *Cell Stem Cell*. 2012;10:740-749.
2. Takahashi K, Yamanaka S. Induced pluripotent stem cells in medicine and biology. *Development*. 2013;140:2457-2461.
3. Atala A, Kasper FK, Mikos AG. Engineering complex tissues. *Sci Transl Med*. 2012;4:160rv112.
4. Doulatov S, Daley GQ. Development. A stem cell perspective on cellular engineering. *Science*. 2013;342:700-702.
5. Lindvall O, Barker RA, Brustle O, et al. Clinical translation of stem cells in neurodegenerative disorders. *Cell Stem Cell*. 2012;10:151-155.
6. Xin M, Olson EN, Bassel-Duby R. Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair. *Nat Rev Mol Cell Biol*. 2013;14:529-541.
7. Pagliuca FW, Melton DA. How to make a functional beta-cell. *Development*. 2013;140:2472-2483.
8. Abelson S, Shamai Y, Berger L, et al. Intratumoral heterogeneity in the self-renewal and tumorigenic differentiation of ovarian cancer. *Stem Cells*. 2012;30:415-424.
9. Imaizumi Y, Okano H. Modeling human neurological disorders with induced pluripotent stem cells. *J Neurochem*. 2014;129:388-399.
10. Cartier N, Hacein-Bey-Abina S, Bartholomae CC, et al. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science*. 2009;326:818-823.
11. Sheridan C. Gene therapy finds its niche. *Nat Biotechnol*. 2011;29:121-128.
12. Candotti F, Shaw KL, Muul L, et al. Gene therapy for adenosine deaminase-deficient severe combined immune deficiency: clinical comparison of retroviral vectors and treatment plans. *Blood*. 2012;120:3635-3646.
13. Zhang L, Thrasher AJ, Gaspar HB. Current progress on gene therapy for primary immunodeficiencies. *Gene Ther*. 2013;20:963-969.
14. Cideciyan AV, Jacobson SG, Beltran WA, et al. Human retinal gene therapy for Leber congenital amaurosis shows advancing retinal degeneration despite enduring visual improvement. *Proc Natl Acad Sci U S A*. 2013;110:E517-E525.
15. Tilemann L, Ishikawa K, Weber T, et al. Gene therapy for heart failure. *Circ Res*. 2012;110:777-793.
16. Russell SJ, Peng KW, Bell JC. Oncolytic virotherapy. *Nat Biotechnol*. 2012;30:658-670.
17. Porter DL, Levine BL, Kalos M, et al. Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia. *N Engl J Med*. 2011;365:725-733.
18. Gaj T, Gersbach CA, Barbas CF 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*. 2013;31:397-405.
19. Tebas P, Stein D, Tang WW, et al. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med*. 2014;370:901-910.
20. Romero Z, Urbinati F, Geiger S, et al. beta-globin gene transfer to human bone marrow for sickle cell disease. *J Clin Invest*. 2013;123:3317-3330.

REVIEW QUESTIONS

1. Meniscus repair by constructing an implantable scaffold seeded with mesenchymal stem cells that differentiate into chondrocytes would be an example of which of the following?

A. Ex vivo gene therapy
B. Autologous tissue meniscus implantation
C. Tissue engineering
D. Direct stem cell reprogramming
E. Somatic gene therapy

Answer: C Rather than simply introducing cells into a diseased area, in tissue engineering, cells are embedded or seeded onto three-dimensional scaffolds (derived from different biomaterials) before transplantation. This can involve either (i) cellular scaffolds that are seeded ex vivo with cells before their in vivo transplantation, or (ii) acellular scaffolds that require the recipient's cells to repopulate the scaffold to reconstitute it after transplantation. To date, these efforts have mainly concentrated on the musculoskeletal system, as in the example presented. Ex vivo and somatic gene therapies by definition involve the transfer of specific genetic material into cells to correct or restore a cellular defect. Direct stem cell programming is the direct conversion of the phenotype of one cell type (e.g., fibroblasts) to another (e.g., chondrocytes). Autologous tissue meniscus implantation does not involve any type of cell and gene therapy. (See Tissue Engineering.)

2. To date, the only established clinical application of fetal-derived cell therapy has been in patients with which of the following?

A. Parkinson disease
B. Myocardial infarction
C. Heart failure
D. Blood product transfusion
E. Cartilage repair

Answer: A Cells of fetal origin show enhanced proliferative capacity and enhanced ability to differentiate into mature or specialized cells. Although this form of cell therapy represents a form of regenerative medicine with great potential, the only fetal-derived cells that have been used in clinical applications to date are the dopaminergic cells derived from the developing fetal nervous system for the treatment of Parkinson disease. (See Engraftment of Fetal Tissue.)

3. Which one of the following in *not* a form of gene therapy?

A. Transfection
B. Transduction
C. RNA interference
D. Gene editing
E. DNA electroporation

Answer: E Electroporation is a strictly in vitro research method in cell biology that electropermeabilizes cell membranes by an externally applied electrical field to introduce a piece of DNA (or other agents) into a cell. All the other choices are methods of gene therapy. Transfection is the direct delivery of naked DNA by injection, the use of liposomes, nanoparticles, and other means. Transduction is gene therapy that is mediated by viral vectors. RNA interference is a highly precise mechanism of regulating gene expression by sequence-directed silencing (by degrading specific messenger RNAs or by blocking their translation into protein). Gene editing is a way of correcting a pathogenic mutation in its natural gene location (rather than the more conventional gene therapy method of gene restoration, in which a whole gene is inserted into the genome). (See Gene Therapy Delivery Methods and Other Forms of Molecular Therapies: RNA Interference and Gene Editing under the main heading of Gene Therapy.)

45

THE INNATE IMMUNE SYSTEM

MARY K. CROW

THE INNATE IMMUNE SYSTEM IN HOST DEFENSE AND DISEASE PATHOGENESIS

The immune system, comprising cells, the molecules they produce, and the organs that organize those components, evolved over millions of years in response to infections with pathogenic microorganisms.[1] Its essential role in maintaining health is based on its recognition and elimination or control of those foreign microbes. Central to the success of the protective role of the immune system is its capacity to distinguish foreign and dangerous invaders from self-components.[2,3] In addition to its contributions to host defense, the immune system is involved in the prevention of malignancy by surveying and recognizing self-cells that express novel antigens,[4] and it also plays a role in resolution and repair of tissue damage.

The immune system is generally described as including an *innate immune system* and an *adaptive immune system*. The former provides the first and rapid line of defense and cellular response to a foreign stimulus. The latter, dependent on activation by the innate immune response, develops a more specific response targeted to the offending organism and generates memory for that stimulus that can be elicited rapidly should that organism be encountered again on a later occasion.

Immune system cells derive from precursor cells of the hematopoietic lineage and populate discrete lymphoid organs, including lymph nodes, spleen, and thymus, as well as skin and intestine. Cells of the innate immune system serve as sentinels at locations that are likely to encounter foreign organisms, and after activation they will often travel to a local lymphoid organ. The induction of the adaptive immune response occurs in the context of structured aggregates of innate and adaptive immune cells in the lymphoid organs. Once activated and differentiated to produce effector molecules, immune system cells can be sampled in blood as they travel to sites of infection or tissue damage. There they can interact directly with target cells to mediate cell death or, alternatively, provide activating signals to expand or regulate a response, or secrete high local levels of immunomodulatory substances called cytokines. Cytokines are small soluble proteins that communicate among cells within the immune system or between immune system cells and cells in other tissues.[5] The cells and products of the immune system function as an exquisitely regulated complex system.[6] Inherited variations in hundreds of genes have evolved, under pressure of microbial challenge, to ensure adequate defense against pathogenic organisms across the human population.[1] However, in any one individual, the composite genetic profile can generate predisposition to infection or, alternatively, autoimmune or inflammatory disease.

The innate immune response was traditionally viewed as mediating nonspecific protection through the production of preformed effector molecules. However, important advances in characterization of the cell surface and intracellular pattern recognition receptors (PRR), particularly the toll-like receptor (TLR) family, and signaling pathways used by innate immune cells to implement a defensive response are now understood to have relative specificity for pathogen-associated molecular patterns (PAMPs) that are characteristic of categories of microbes.[7,8] In contrast to those receptor systems that initiate an innate immune response, the protein products that implement the response, whether to expand the reaction to additional cells, promote trafficking to the most relevant location, or shape the differentiation programs of adaptive immune system cells, do not show specificity based on the initial triggering stimulus. The products of the innate immune response can be highly effective at ablating or limiting the extent of infection and can generate a tissue repair program that establishes a satisfactory resolution of the episode of infection. However, when sustained or poorly regulated, they can represent an important pathophysiologic mechanism for many autoimmune and inflammatory diseases.

Cells of the Innate Immune System

Monocytes and Macrophages

Monocytes circulate in the peripheral blood with a half-life of 1 to 3 days. Macrophages arise from monocytes that have migrated out of the circulation and have proliferated and differentiated in tissue. Tissue macrophages include alveolar macrophages in the lung, Kupffer cells in the liver, osteoclasts in bone, microglia in the central nervous system, and type A synoviocytes in the synovial membrane. Macrophages secrete myriad products, including hydrolytic enzymes, reactive oxygen species, cytokines, and chemokines. Macrophages engulf microorganisms and foreign particles directly or are activated by protein complexes containing antibodies that bind to cell surface receptors for the Fc portion of immunoglobulin molecules (Fc receptors, or FcRs). These encounters activate intracellular signaling pathways that induce transcription of target genes, primarily those encoding mediators that promote inflammation or enzyme-mediated death of the microbe. Cytokines from other immune system cells, including interferon (IFN)-γ or interleukin (IL)-4, can drive macrophage differentiation toward the production of mediators that are primarily pro-inflammatory or to a wound healing functional profile. Researchers have characterized those functional phenotypes as M1 or M2, although it is recognized that the context of an innate immune response will determine the functional response, with composite profiles common.[9]

In addition to responding to foreign microbes, macrophages contribute to the elimination of senescent or apoptotic cells in a manner that avoids induction of an inflammatory response. Macrophages also interact with other cell types through complementary cell surface adhesion or costimulatory receptors. After capturing antigen, they can function as antigen-presenting cells for T lymphocytes, and they can interact with non–immune system cells such as endothelial cells or fibroblasts.

Dendritic Cells

Dendritic cells (DCs) comprise a complex family of cells that perform essential functions in the innate immune response and serve as a bridge to activation of an adaptive immune response. Myeloid dendritic cells can incorporate antigens derived from invading microbes, travel to nearby lymph nodes, and present processed antigenic peptides to T lymphocytes (T cells) in the form of peptide–major histocompatibility complex (MHC) molecule complexes. They are the most effective antigen-presenting cells based on expression of cell surface costimulatory molecules, and they produce cytokines, including IL-12 and IL-23, after interaction with PAMPs. They thereby contribute to the shaping of the T-cell differentiation program to generate effector cell functions. Plasmacytoid dendritic cells (pDCs) have been identified as highly effective producers of type I interferon, a key mediator of host defense against viral infections.

Natural Killer Cells

Natural killer (NK) and NK T cells provide early defense against viral infections and other intracellular pathogens while adaptive responses are developing.[10] NK cells are sensitized by cytokines, including type I interferons, released from pDCs and macrophages, and secrete abundant IFN-γ, which activates macrophages and other cells. They also are poised to kill virus-infected cells by injecting pore-forming enzymes and granzymes. Activation of NK cells is inhibited by interaction with self-MHC class I molecules on target cells. When those self-histocompatibility antigens are not present, NK cell–mediated killing is implemented. NK cells are important in tumor surveillance because they are able to kill MHC class I–deficient tumor cells that are no longer susceptible to adaptive immune responses. In addition to NK cells, a type of lymphocytes, so-called innate lymphoid cells, which participate early in innate immune responses but do not express rearranged receptors, is a focus of current study.[11]

Neutrophils

Neutrophils are the most abundant circulating white blood cells. They are recruited rapidly to inflammatory sites and can phagocytose and digest microbes (Chapters 167 and 169). Activation of neutrophils and phagocytosis is facilitated through the triggering of FcRs or complement receptors. Microbe-containing phagosomes fuse with lysosomes, which contain enzymes, proteins, and peptides that inactivate and digest microbes. Beyond their phagocytic capability, neutrophils produce a variety of toxic products. The release of toxic products is known as the respiratory burst because it is accompanied by an increase in oxygen consumption. During the respiratory burst, oxygen radicals are generated by nicotinamide adenine dinucleotide phosphate (NADPH) oxidases. Neutrophils also contribute to host defense through extrusion of DNA and associated proteins in the form of neutrophil extracellular traps, or NETs, to which bacteria can stick, facilitating their clearance. Despite their effective contributions to the innate immune

response and microbial host defense, neutrophils can generate considerable collateral damage. NETs have the capacity to induce production of cytokines by pDCs and may damage vascular endothelial cells. Secretion of neutrophil granule contents, particularly their enzymes (myeloperoxidase, elastase, collagenase, and lysozyme), causes direct cellular injury and damages macromolecules at inflamed sites.

Eosinophils

In contrast to macrophages and neutrophils, eosinophils are only weakly phagocytic but are potent cytotoxic effector cells against parasites. Their major effector mechanism is the secretion of cationic proteins (major basic protein, eosinophil cationic protein, and eosinophil-derived neurotoxin). These proteins are released into the extracellular space, where they directly destroy the invading microorganism but can also damage host tissue (Chapter 170).

Basophils and Mast Cells

Basophils and tissue mast cells secrete inflammatory mediators such as histamine, prostaglandins, leukotrienes, and some cytokines.[12] Release of these substances is triggered when cell surface immunoglobulin E (IgE) receptors encounter monomeric IgE. They play a role in atopic allergies, in which allergens bind immunoglobulin (IgE) and cross-link FcεRs. Mast cells have been observed in rheumatoid arthritis synovial tissue and have been implicated in local inflammatory responses (Chapter 255). Like pDCs and macrophages, mast cells express TLRs and FcRs and produce cytokines after encountering immune complexes composed of TLR ligands.

Recognition Receptors and Triggers of an Innate Immune Response

Toll-like Receptors

The innate immune system utilizes both cell surface and intracellular PRRs to recognize conserved structures on microbes (PAMPs). Examples of PAMPs are bacterial lipopolysaccharides, peptidoglycans, mannans, bacterial DNA, double-stranded RNA, and glucans. The discovery and characterization of the TLR family of receptors and their relevant ligands has focused attention on the mechanisms that allow an innate immune response to shape the nature of the resulting inflammatory or repair programs, as well as the T-cell effector cell functions that follow recognition of antigens from the relevant pathogen. The TLRs have in common leucine-rich domains and bind PAMPs common to classes of pathogenic organisms.[7,8] For example, TLR-4, a cell surface–expressed PRR, binds lipopolysaccharide of gram-negative bacteria, and TLR-2 recognizes bacterial peptidoglycans and lipoproteins, often based on dimerization with other TLR family members. Important advances in understanding systemic autoimmune diseases have followed the characterization of endosomal TLRs with relative specificity for single-stranded RNA (TLR-7 and TLR-8), demethylated CpG-enriched DNA (TLR-9), and double-stranded RNA (TLR-3, which has both cell surface and endosomal forms). The distribution of particular TLRs among cells of the innate immune system varies, and additional members of the TLR family may still be discovered and characterized. The TLRs play central roles in alerting the immune system that a microbe, typically a bacteria in the case of TLR-2 and TLR-4 or a virus in the case of TLR-3, TLR-7, TLR-,8 and TLR-9, is threatening the host. But in some cases, when an immune complex with self-nucleic acid gains access to an endosomal TLR, a self-directed innate immune response can be initiated or amplified.

Cytoplasmic Nucleic Acid Sensors

Following the description of the TLR family and the capacity of the endosomal TLRs to recognize microbial and self-nucleic acids, a second category of intracellular innate immune system receptors was defined that recognize RNA or DNA from microbes, primarily viruses, that gain access to the cell cytoplasm. The DExD/H-box family of helicases include retinoic acid–inducible gene I (RIG-I) and melanoma differentiation–associated protein 5 (MDA5), described as members of the RIG-I-like receptor (RLR) family that recognizes viral RNAs with particular structural characteristics that distinguish the viral RNA from most host RNAs (Fig. 45-1).[13,14] Cytoplasmic DNA receptors have also been defined, with cyclic guanosine monophosphate–adenosine monophosphate synthase (cGAS) recently identified as an important sensor of cytoplasmic DNA that triggers an innate immune response after interacting with the stimulator of interferon genes (STING).[15] Whether RNA or DNA triggers these cytoplasmic sensors, the result is transcription and production of interferon-β and other pro-inflammatory cytokines that orchestrate the early phase of an antiviral immune response.

NOD Receptors

Another category of intracellular receptors is proving important in antimicrobial defense as well as contributing to activation of inflammatory states. The nucleotide-binding oligomerization domain (NOD)-like receptor (NLR) family comprises components of an intracellular structure called the inflammasome, a signaling platform that organizes innate immune system activation in response to some stimuli.[16] The inflammasome can activate caspase 1, an enzyme important for maturation of the pro-inflammatory cytokines IL-1β and IL-18. The NLRP3-containing inflammasome has been best studied and implicated in the inflammatory response to monosodium urate crystals, the triggers of gout attacks (Chapter 273). Mutations in the *NLRP3* gene are the basis of chronic autoinflammatory syndromes that are associated with exaggerated production of IL-1 (reviewed in Chapter 261).

C-Type Lectin Receptors

Members of the C-type lectin receptor family have a carbohydrate recognition domain and a calcium-binding domain that promotes signaling after interaction with carbohydrate-expressing microbes as well as self-molecules. DC-SIGN (DC-specific intracellular adhesion molecule-3 grabbing nonintegrin) is an example of a family member that recognizes high-mannose-containing structures on foreign antigens and supports DC activation. Mannose receptors on macrophages, dendritic cells, and other cell types, such as renal mesangial cells, participate in clearance of microbes as well as antigen trapping for presentation to adaptive immune system cells. The selectin family of proteins have a lectin domain, bind to carbohydrate ligands, and mediate the first steps of leukocyte migration. L-selectin is present on virtually all leukocytes; P-selectin and E-selectin are expressed on activated endothelial cells, and P-selectin is also stored in platelets. Selectins capture floating leukocytes and initiate their attachment and rolling on activated endothelial cells.

Scavenger Receptors

Scavenger receptors comprise a diverse family of receptors with the common functional role of binding various ligands and transporting or removing nonself or altered-self targets.[17] They can participate in clearance of microorganisms and cholesterol transport but can also contribute to disease pathology. For example, among the scavenger receptors is the receptor for oxidized low-density lipoproteins, which can promote generation of lipid-laden macrophages and atherosclerosis when accumulated in excess, and receptors for relatively inert substances such as silicon, which can drive an inflammatory response once taken into phagocytic cells. Scavenger receptors can also participate in activation of the inflammasome, as can occur after binding serum amyloid A protein.

Inhibitory Natural Killer Cell Receptors

The immunoglobulin-like killer inhibitory receptor (KIR) family of receptors participates in distinguishing self-cells from cells of foreign origin or tumor cells expressing modified-self-molecules. NK cells are ready to produce their toxic mediators, but they are held in check by inhibitory receptors that recognize MHC class I or MHC class I–like molecules.[10] Recognition of MHC class I molecules provides a negative signal that suppresses cell activity. The observation that NK cells kill target cells lacking MHC class I molecules recognized as self led to the missing-self hypothesis. By screening cell surfaces for the expression of MHC class I molecules, the innate immune system collects information about the intactness of tissues, emphasizing the crucial role of MHC class I molecules as markers of tissue integrity.

Fc and Complement Receptors

Most cells of the innate immune system possess receptors (FcRs) that specifically interact with the constant region (Fc portion) of immunoglobulins and can bind antibodies attached to antigens. The isotype of the antibody determines which cell type is activated in a given response. Triggering of most FcRs transmits activating signals; however, inhibitory FcRs on B lymphocytes (B cells) and macrophages can limit responses. Ligation of an FcγR on macrophages or neutrophils triggers phagocytosis of the antigen, activation of respiratory burst, and induction of cytotoxicity. On NK cells, FcγRs initiate antibody-dependent cell-mediated cytotoxicity. FcRs on pDCs are important for bringing immune complexes into intracellular compartments containing endosomal TLRs. FcRs on mast cells, basophils, and activated

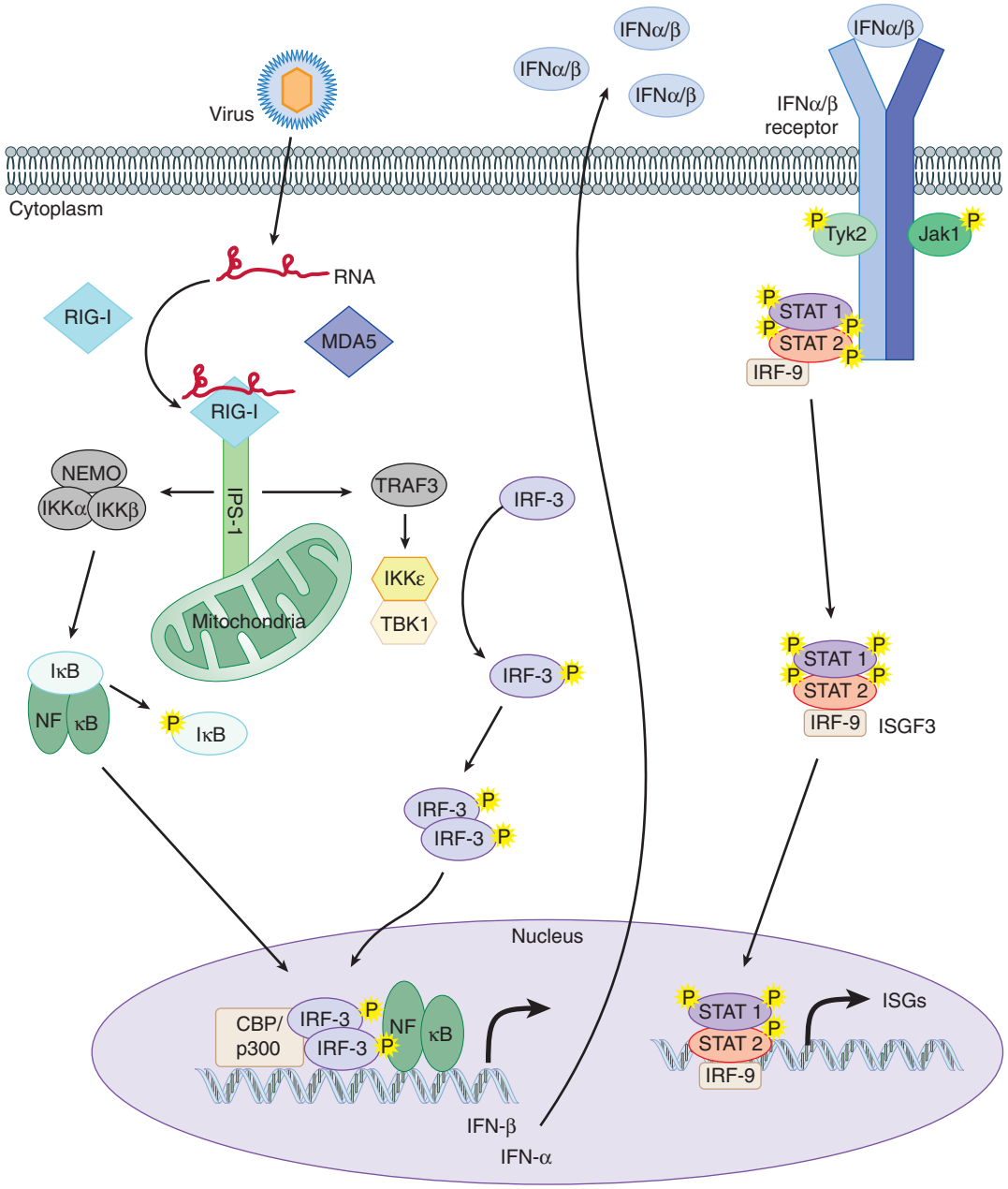

FIGURE 45-1. Induction of antiviral type I interferon response. Cytoplasmic sensors of RNA, including RIG-I and MDA5, trigger a signaling cascade that results in translocation of IRF-3 to the nucleus and transcription of interferons. Those cytokines promote an antiviral immune response after binding to their receptor and activating the JAK-STAT pathway. CBP/p300 = CREB binding protein; NEMO = NF-κB essential modulator; IFN = interferon; IKK = inhibitor of nuclear factor κB kinase subunit; IPS-1 = interferon-β promoter stimulator-1; IRF = interferon response factor; ISG = interferon stimulated gene; ISGF3 = interferon-stimulated gene factor 3; JAK = Janus kinase; MDA5 = melanoma differentiation-associated protein 5; RIG-1 = retinoic acid–inducible gene 1; STAT = signal transducer and activator of transcription; TRAF3 = TNF (tumor necrosis factor) receptor–associated factor; Tyk = tyrosine kinase. (From Wilkins C, Gale M Jr. Recognition of viruses by cytoplasmic sensors. *Curr Opin Immunol.* 2010;22:41-47.)

eosinophils bind monomeric IgE with extremely high affinity. Cross-linking of the constitutively cell surface–bound IgE induces cell activation and the release of cytoplasmic granules. Some immunoglobulin isotypes fix complement, and complement receptors on monocytes amplify cell activation induced by antigen-antibody-complement immune complexes[18] (Chapter 50). Complement receptor 1 (CR1) binds C3b and C4b, initial degradation products of complement activation, and when activated promotes phagocytosis of a complement-bearing immune complex. CR3 and CR4 are β_2-integrins and bind the degradation product iC3b.

Cytokine and Chemokine Receptors

Cells of the innate immune system express receptors for many cytokines, soluble, low-molecular-weight glycoproteins that derive from many cellular sources.[5] Binding of IFN-γ, produced by NK or type 1 helper T cells (T_H1 cells), by its receptor on monocytes activates a differentiation program that expands an inflammatory response. Receptors for IL-4 on monocytes induce a gene transcription program that is more supportive of a wound healing and repair program. Tumor necrosis factor-α (TNF-α) is a product of activated macrophages but also binds to those cells through its specific receptor, expanding an inflammatory response. Innate immune cells also express receptors for IL-6, which induces acute phase reactants and type I interferon, which orchestrates a broad host defense program in response to virus infection (see Fig. 45-1). Chemokine receptors include many family members that are differentially distributed among immune system cells and sense the gradient generated by soluble chemokines, resulting in attraction of cells to sites where they are needed to implement inflammatory or immune functions.

Signaling Pathways and Effector Mediators of the Innate Immune System

Each family of innate immune system receptors utilizes a complex network of molecules to transmit information from the cell surface or its cytoplasm to the nucleus, resulting in induction of a broad gene transcription and protein synthesis program that implements the next phase of the response. The contributions of each of the signal transduction pathways to the overall innate immune response will depend on the proteins produced and will determine whether the resulting cell products focus the overall immune

function on ablating the damaging effects of virus infection on the host, limiting the inflammation and tissue damage that follow a bacterial or fungal infection, or healing a tissue wound through the production of scar tissue.

Receptor-Mediated Signaling Pathways

Certain common cell signaling systems are utilized by many cells and receptor systems.[6,7] Arguably the most important is the nuclear factor κ light-chain enhancer of activated B cells (NF-κB) pathway. NF-κB is a rapid-acting transcription factor because it is preformed in cells of the innate immune system and does not require new protein synthesis to take action. Its activity is induced by ligation of TLRs and many cytokine receptors. Its component transcription factors translocate to the cell nucleus after degradation of an inhibitory component, inhibitor of κB (IκB), and bind to promoter regions of genes encoding mediators of inflammation and cell proliferation. Another important pathway is mediated by the interferon regulatory factor (IRF) family, including transcription factors that are activated by endosomal TLRs in response to ligation by DNA or RNA, or by cytoplasmic nucleic acid sensors, usually from viral sources. IRF-3 is particularly important for promoting transcription of interferon-β, typically produced early in an antivirus innate immune response. IRF-7 is particularly supportive of interferon-α production induced by endosomal TLRs and is constitutively present in pDCs, the most active producers of IFN-α.

The Janus kinase (JAK)-signal transducer and activator of transcription (STAT) pathway is utilized by many cytokine receptors and involves sequential enzymatic reactions by kinases that eventuate in translocation of STAT proteins to the nucleus, where they bind to gene promoters and induce transcription and production of products important in implementing immunoregulation and inflammation.

TNF receptor family members activate a complex signaling pathway that involves proteins called TNF receptor–associated death domain (TRADD) proteins and TNF receptor–associated factors (TRAFs), ultimately activating the NF-κB and the mitogen-activated protein (MAP) kinase pathways.

The TGF-β receptor is a serine/threonine receptor kinase that phosphorylates cytoplasmic proteins of the SMAD family, which act as transcription factors after receptor engagement by TGF-β. TGF-β signaling can play an important role in terminating an innate immune response and initiating a wound healing or tissue repair program.

It is apparent that common intracellular signaling strategies are used by many of the receptor systems that activate and regulate the innate immune system, with ligand-receptor engagement triggering the activation of kinases that phosphorylate downstream pathway proteins, and result in translocation of important transcription factors from cytoplasm to nucleus where new gene transcription takes place.

Soluble Products of the Innate Immune Response

Cells of the innate immune system are the principal producers of many proinflammatory and regulatory cytokines already mentioned, and are also their targets. In addition to the cytokines described, cells of the innate immune system produce chemokines that attract immune system cells to sites of tissue damage or infection, and they produce cell survival and differentiation factors that help to develop an adaptive immune response. Macrophages and dendritic cells produce IL-12 and IL-23 to support development of effector T-cell programs, and they produce B-cell-activating factor (BAFF), a soluble mediator of the TNF family. BAFF supports B-cell survival and can provide costimulatory signals to B cells that have received antigen-specific activation signals through their surface B-cell antigen receptors, promoting differentiation to antibody-producing plasma cells.

A particularly important set of products includes components of the complement system, a group of plasma enzymes and regulatory proteins that are converted from inactive pro-enzymes to active enzymes in a controlled and systematic cascade, which is crucial in linking microbial recognition to cellular effector function (Chapter 50). Mannose-binding lectin circulates in the plasma, functioning as an opsonin, and is involved in activation of the complement pathway. C-reactive protein, an acute phase protein, participates in opsonization by binding to bacterial phospholipids. Macrophages and neutrophils are important in the initiation phase of an innate immune response through their production of antimicrobial defensins, cysteine-rich cationic proteins, and cathelicidin peptides, such as LL37.[19] Both categories of mediators can assist in killing of microbes in phagosomes. Neutrophils extrude stimulatory DNA in the form of NETs or release mitochondrial DNA, along with DNA-associated proteins like high mobility group box 1 (HMGB1) that amplifies TLR responses in pDCs or macrophages.

Role of the Innate Immune System in Localization, Extension, and Resolution of a Host Defense Reaction

Localization of Innate Immune System Cells

Most cells of the innate immune response are free agents, moving through blood or lymph in transit from one site to another. Mobility of the cellular constituents of the innate immune system is required for effective initiation of a response to invading microbes. Cells use a multistep process of adherence and activation. Initially, leukocytes roll on activated endothelial cells, activate chemokine receptors, increase adhesiveness, and eventually migrate through the endothelial layer across a chemokine gradient. The selectin family of proteins mediates the first steps of leukocyte migration. P-selectin and E-selectin are expressed on activated endothelial cells, and P-selectin is also stored in platelets. Selectins capture floating leukocytes and initiate their attachment and rolling on activated endothelial cells. To transform attachment and rolling into firm adhesion, the concerted action of chemokines, chemokine receptors, and integrins is necessary. Integrins are heterodimers formed of many different α chains and β chains; different α/β combinations are expressed on different cell subsets. Only after activation can integrins interact with ligands on endothelial cells. Activation involves modification of the cytoplasmic domain of the β chain, which leads to a structural change of the extracellular domains. This process is termed *inside-out signaling*. The last step of homing is transendothelial migration. Here, the firmly attached leukocytes migrate through the endothelial cell monolayer and the basement membrane of the vessel wall.

Transition to an Adaptive Immune Response

Movement of innate immune system cells is also required to transition a host response from primarily one depending on cells of the innate immune system to one that engages T and B lymphocytes. Dendritic cells resident in the skin and gut serve as sentinels and a first line of defense against invading organisms. When those cells are activated following sensing of PAMPs by PRRs and following uptake of microbial components by those cells, the DCs migrate to local lymph nodes where their contents, by now expressed on their surface in association with MHC class I or II molecules, can be sampled by T cells. As noted, activated macrophages, DCs, and pDCs produce cytokines that shape the differentiation program of T cells. In addition, cell surface costimulatory molecules induced after TLR-mediated activation, such as CD80 and CD86, provide essential accessory activation signals to T cells to ensure their effective activation. Macrophages and DCs also support the development of an adaptive immune response through their production of survival and differentiation factors. Chapter 46 provides a full description of the adaptive immune system and its implementation.

Role of Innate Immune System Cells in Resolution of an Immune Response and Wound Repair

Macrophages are particularly important in resolving an immune response and organizing the repair of damaged tissue. A classic paradigm describing pro-inflammatory/classically activated (M1) and anti-inflammatory/alternatively activated (M2) macrophages (see earlier under Monocytes and Macrophages) is likely to be overly simplistic. Yet it is clear that in the course of a chronic infection, macrophages can shift their functional profile from M1 to M2, in some cases promoted by the T-cell cytokines IL-4 and IL-13, to develop a gene expression program that includes production of TGF-β, supportive of a fibrotic response, and IL-10, a cytokine that inhibits antigen-presenting cell function.[9] Although an M1-like profile driven by IFN-γ is highly productive in achieving initial control over a pathogenic invading microbe, and M2-derived mediators promote wound healing, it should be recognized that either macrophage phenotype, and complex in-between profiles, can also be associated with pathologic states (Fig. 45-2). Current research is unraveling the innate immune mechanisms that account for such diverse diseases as atherosclerosis (Chapter 70), viewed as associated with M1 macrophages, and idiopathic pulmonary fibrosis (Chapter 92), possibly involving M2-like macrophages.

Contribution of the Innate Immune Response to Pathogenesis of Autoimmune Disease

Among the most significant insights of the past decade is the essential contribution of the innate immune system to the pathogenesis of autoimmune and inflammatory diseases. As described, the cells of the innate immune system are integral players in the early recognition of invading pathogenic microbes, and when the functions of this complex system are carefully

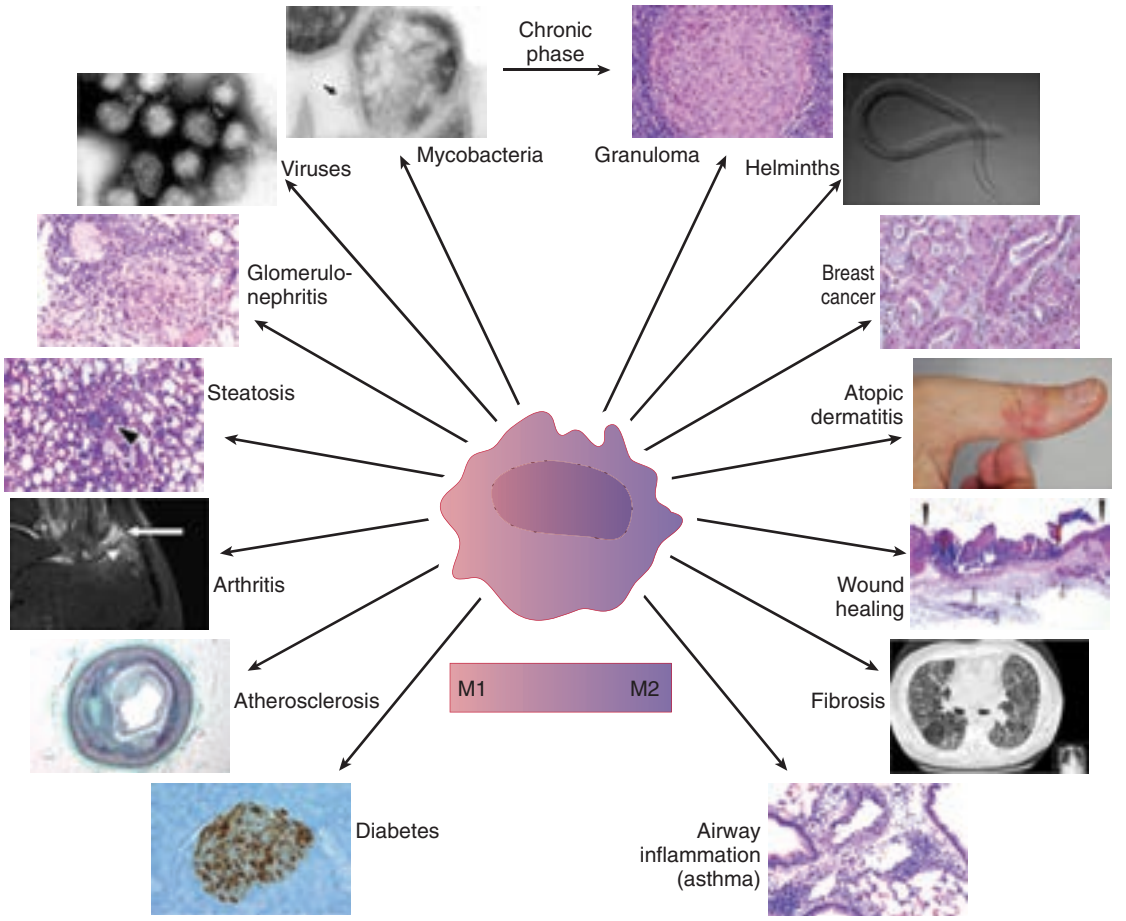

FIGURE 45-2, Schematic representation of macrophage plasticity and polarization in pathology. Dynamic changes occur over time with evolution of pathology: for instance, a switch from M1 to M2 macrophage polarization characterizes the transition from early to chronic phases of infection. Moreover, mixed phenotypes or populations with different phenotypes can coexist. (From Sica A, Mantovani A. Macrophage plasticity and polarization: in vivo veritas. *J Clin Invest.* 2012;122:787-795.)

orchestrated and balanced, the result is efficient ablation, or at least isolation, of the microbe. However, if the microbe is not effectively cleared from the system and persists, a chronic state of infection associated with immune activation and tissue damage is the result. Interestingly, many parallels can be seen between the immune alterations observed in the setting of chronic viral infection and the impaired immunoregulation characteristic of the prototypic autoimmune disease systemic lupus erythematosus. Excessive production of interferon-α is a feature of most patients with that disease, and it is now understood that activation of the endosomal TLRs by nucleic acid–containing immune complexes amplifies the activity of the innate immune response and drives production of interferon-α and other pro-inflammatory cytokines. Neutrophils are now recognized to contribute to the induction of that response through their production of HMGB1, cathelicidins, and extrusion of stimulatory DNA aggregates. TLR activation is proposed to contribute to many additional autoimmune and inflammatory diseases; as endogenous TLR ligands can act as effective TLR stimuli in the setting of a pro-inflammatory environment associated with oxidative cell damage. The inflammasome and its component proteins, including the NOD-like receptors, are recognized as mediators of inflammatory responses induced by urate crystals that result in gout attacks (Chapter 273), and they are targets of mutations that define dramatic autoinflammatory syndromes (Chapter 261), particularly seen in children.

Conclusion

The cells and products of the innate immune response, for many years viewed as less sophisticated and important than the highly specific T and B lymphocytes of the adaptive immune response, have taken their place as essential defenders against pathogenic microbes. Through the recognition of common molecular patterns characteristic of microbes by members of receptor families, some still being discovered, the cells of the innate immune response orchestrate the effector programs that are fine-tuned to target the vulnerabilities of each pathogen and kill, or at least limit the expansion of, that microbe. Advances in understanding the mechanisms utilized by the innate immune response and the clinical syndromes that result when components of that system are genetically altered, have elucidated the central role that receptors and products of the innate immune system play in the pathogenesis of autoimmune and inflammatory diseases. These insights are guiding efforts to develop targeted therapies that will leverage the new knowledge to control or even prevent human diseases in which the innate immune system plays an important pathogenic role.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Quintana-Murci L, Clark AG. Population genetic tools for dissecting innate immunity in humans. *Nat Rev Immunol*. 2013;13:280-293.
2. Iwasaki A, Pillai PS. Innate immunity to influenza virus infection. *Nat Rev Immunol*. 2014; 14:315-328.
3. Busca A, Kumar A. Innate immune responses in hepatitis B virus (HBV) infection. *Virol J*. 2014;11:22.
4. Marcus A, Gowen BG, Thompson TW, et al. Recognition of tumors by the innate immune system and natural killer cells. *Adv Immunol*. 2014;122:91-128.
5. Torrado E, Cooper AM. Cytokines in the balance of protection and pathology during mycobacterial infections. *Adv Exp Med Biol*. 2013;783:121-140.
6. Zak DE, Tam VC, Aderem A. Systems-level analysis of innate immunity. *Annu Rev Immunol*. 2014;32:547-577.
7. Broz P, Monack DM. Newly described pattern recognition receptors team up against intracellular pathogens. *Nat Rev Immunol*. 2013;13:551-565.
8. O'Neill LA, Golenbock D, Bowie AG. The history of toll-like receptors: redefining innate immunity. *Nat Rev Immunol*. 2013;13:453-460.
9. Sica A, Mantovani A. Macrophage plasticity and polarization: in vivo veritas. *J Clin Invest*. 2012;122:787-795.
10. Terabe M, Berzofsky JA. The immunoregulatory role of type I and type II NKT cells in cancer and other diseases. *Cancer Immunol Immunother*. 2014;63:199-213.
11. Hazenberg MD, Spits H. Human innate lymphoid cells. *Blood*. 2014;124:700-709.
12. Cromheecke JL, Nguyen KT, Huston DP. Emerging role of human basophil biology in health and disease. *Curr Allergy Asthma Rep*. 2014;14:408.
13. Schlee M. Master sensors of pathogenic RNA - RIG-I like receptors. *Immunobiology*. 2013;218: 1322-1335.
14. Wu J, Chen ZJ. Innate immune sensing and signaling of cytosolic nucleic acids. *Annu Rev Immunol*. 2014;32:461-488.
15. Cai X, Chiu YH, Chen ZJ. The cGAS-cGAMP-STING pathway of cytosolic DNA sensing and signaling. *Mol Cell*. 2014;54:289-296.
16. Latz E, Xiao TS, Stutz A. Activation and regulation of the inflammasomes. *Nat Rev Immunol*. 2013;13:397-411.
17. Canton J, Neculai D, Grinstein S. Scavenger receptors in homeostasis and immunity. *Nat Rev Immunol*. 2013;13:621-634.
18. Holers VM. Complement and its receptors: new insights into human disease. *Annu Rev Immunol*. 2014;32:433-459.
19. Silva PM, Gonçalves S, Santos NC. Defensins: antifungal lessons from eukaryotes. *Front Microbiol*. 2014;5:97.

REVIEW QUESTIONS

1. Which of the following cell types is *not* considered to be a component of the innate immune system?

A. Macrophages
B. Neutrophils
C. T lymphocytes
D. Plasmacytoid dendritic cells
E. Eosinophils

Answer: C T lymphocytes are important components of the adaptive immune system. T and B lymphocytes use mechanisms that rearrange DNA to form novel specific antigen-binding receptors. Cells of the innate immune system express pattern recognition receptors but do not express specific antigen-binding receptors.

2. Endosomal toll-like receptors (TLRs) recognize which of the following pathogen-associated molecular patterns (PAMPs)?

A. Flagellin
B. Lipopolysaccharide
C. Antigenic peptides
D. Nucleic acids

Answer: D TLRs are important innate immune system receptors that recognize patterns expressed by pathogenic microbes and some endogenous molecules. Cell surface–expressed TLRs recognize molecules that are typically expressed on the surface of microbes. Intracellular endosomal TLRs, such as TLR-3, -7, -8, and -9, recognize RNA or DNA. The sequestering of those nucleic acid–responsive TLRs protects the immune system from inadvertent activation by self-nucleic acids. However, in diseases such as systemic lupus erythematosus, nucleic acid–containing immune complexes can gain access to the endosomal TLRs and induce an innate immune response.

3. Which of the following innate immune system stimuli utilizes the inflammasome to trigger an inflammatory disease?

A. Peptide–major histocompatibility class (MHC) class II complex
B. Monosodium urate crystals
C. Interleukin-6 (IL-6)
D. Immunoglobulin E
E. Complement

Answer: B Urate crystals access the components of the NOD-like receptors of the inflammasome, activate caspase I, and induce the formation of IL-1, a pro-inflammatory mediator that can amplify an innate immune response. In some patients, this response leads to the acute inflammatory arthritis known as gout. Peptide–MHC class II complexes are stimuli for activation of an adaptive immune response. IL-6 is a broadly active cytokine, and immunoglobulin E is a component of an allergic response.

4. M2 macrophages participate in which of the following?

A. Wound healing responses
B. Antigen presentation
C. Production of IL-12
D. Recognition of oxidized low-density lipoprotein (LDL)
E. Complement activation

Answer: A Although the designation of M1 and M2 macrophages is overly simplistic, macrophages do shift their functional program as the course of an immune response progresses toward a more chronic state, with M2 macrophages expanding and producing mediators, such as transforming growth factor-β and IL-10, that contribute to resolution of responses and repair of damaged tissue. Antigen presentation, production of IL-12, and recognition of oxidized LDL are more likely to be features of M1 macrophages.

5. Cells of the innate immune system produce soluble mediators that contribute to activation and expansion of an adaptive immune response. Among those mediators are which of the following?

A. BAFF
B. IL-23
C. LL37
D. All of the above
E. A and B

Answer: E Macrophages and dendritic cells produce mediators that influence both the T and B cell arms of the adaptive immune response. IL-23 can promote generation of T-cell effector programs. BAFF is a B-cell survival and differentiation factor. LL37 is a cathelicidin that is produced by neutrophils and participates in the killing of phagocytosed microbes.

46

THE ADAPTIVE IMMUNE SYSTEM

JOSEPH CRAFT

PRINCIPLES OF ADAPTIVE IMMUNE SYSTEM ACTIVATION: RECOGNITION OF ANTIGEN

Structure of Antigen-Specific Receptors

The innate immune system recognizes structural patterns that are common in the microbial world, whereas the adaptive immune system is designed to respond to the entire continuum of antigens. This goal is achieved through two principal types of antigen recognition receptors: antibodies and T-cell receptors (TCRs). Antibodies, or immunoglobulins, are expressed as cell surface receptors on B cells or are secreted, both of which have the same

specificity for antigen. They recognize conformational structures formed by the tertiary configuration of proteins. In contrast, α/β TCRs, the most abundant class of TCRs, fit specifically to epitopes formed by a small linear peptide embedded into major histocompatibility complex (MHC) molecules on the surface of antigen-presenting cells.

Antibodies

Antibodies consist of two identical heavy chains and two identical light chains, which are covalently linked by disulfide bonds. The amino (N)-terminal domain of each chain is variable and represents the recognition structure that interacts with the antigen. Each antibody has two binding arms of identical specificity. The carboxy (C)-terminal ends of the heavy and light chains form the constant region, which defines the subclass of the antibody (κ or λ for light chains; immunoglobulin M (IgM), IgA, IgD, IgE, or IgG for heavy chains). Additional subclasses can be distinguished for IgG and IgA. The constant region of antibodies includes the Fc region. Fc regions can polymerize (IgA) or pentamerize in the presence of a J (joining) chain (IgM). Fc regions are also the ligand for Fc receptors (FcRs) on cells of the innate immune system.

T-Cell Receptors

TCRs are dimers of α chains and β chains or of γ chains and δ chains, each of which contains three complementary-determining binding sites in the N-terminal domain. These complementary-determining sites define the specificity. α/β TCRs recognize peptide fragments in the context of MHC molecules, although certain ones bind glycolipid antigens, for example from mycobacteria, displayed by molecules with structural similarity to MHC. γ/δ TCRs are more variable and can recognize peptides or certain glycolipid antigens in the context of MHC-like molecules, or even unprocessed antigens, functioning similar to antibodies; the latter is a reflection of their structural similarity.

Specificities of Antibodies and T-Cell Receptors

The repertoires, or total number of specificities, of antibodies and TCRs are extremely diverse and have been estimated in the human to account for up to 10^{11} or higher, and 10^{18}, respectively, combinations. This enormous diversity reflects the anticipatory nature of adaptive immune receptors and must be acquired; it cannot be genetically encoded in contrast to that of innate receptors. Its foundation consists of fewer than 400 genes that are recombined and modified. Immunoglobulin heavy chains are formed from four gene segments—the variable, diversity, joining, and constant region gene segments. Also, TCR β chains and δ chains are assembled by the recombination of variable, diversity, joining, and constant region segments of TCR genes. Immunoglobulin light chains and TCR α chains and γ chains lack the diversity segment and are composed of three gene segments. During antibody or TCR rearrangement, gene segments are cut out by nucleases and recombined at the DNA level to form linear coding units for each receptor gene. Through the combination of several different mechanisms, an enormous diversity of receptors is generated. First, the genome contains multiple forms of gene segments; each receptor or antibody uses a different combination of these gene segments. Second, the splicing process is imprecise, introducing nucleotide variations at the variable-diversity, diversity-joining, and variable-joining junctions. These inaccuracies lead to frame shifts and result in completely different amino acid sequences. Finally, random nucleotides can be inserted at the junctional region by an enzyme, deoxyribonucleotidyl transferase.

Once generated, TCR sequences remain unchanged. This rule does not apply to immunoglobulins, which undergo modification. Immunoglobulin modification includes (1) replacement of an entire variable region, or receptor editing, typically occurring in the bone marrow during B-cell development to modify those immunoglobulin receptors that inadvertently bind self-antigens on initial recombination of gene segments; (2) class switching, in which the variable-diversity-joining unit combines with different constant region genes (isotype switching); or (3) somatic hypermutation, in which the antigen-contact areas of the antibody undergo mutations during an immune response to improve the affinity (affinity maturation). The latter two events occur in secondary lymphoid tissues, such as the spleen, lymph nodes, and mucosal lymphoid tissue, where immune responses to antigens are initiated.

Antigen Processing

T cells bearing α/β TCRs recognize peptide fragments that are displayed in the context of MHC class I and class II molecules through a process named *antigen presentation*. The two classes of MHC molecules are used as restriction elements by two different subsets of T cells. $CD4^+$ T cells recognize antigenic peptides embedded into MHC class II molecules, whereas $CD8^+$ T cells bind peptides complexed with MHC class I molecules. Generally, MHC class II molecules are expressed only on specialized, so-called professional, antigen-presenting cells, such as dendritic cells, monocytes, macrophages, and B cells, whereas class I proteins are displayed by virtually all nucleated cells, facilitating recognition by $CD8^+$ T cells of peptides from viruses that often have a broad range of target tissues. Peptides bound to MHC class II molecules typically derive from extracellular antigens that are captured and internalized into endosomes to be digested by proteinases, notably cathepsin. Occasionally, however, intracellular proteins or membrane proteins are also funneled into this pathway. MHC class II molecules are assembled in the endoplasmic reticulum in association with a protein called the *invariant chain* (Fig. 46-1). The molecules are transported to the endosome, where the invariant chain is removed from the peptide-binding cleft, making the cleft accessible to peptides derived from extracellular proteins. MHC class II molecules, stabilized by peptides of 10 to 30 amino acids in length, are displayed on the cell surface, where they are recognized by $CD4^+$ T cells.

MHC class I–associated peptides are produced in the cytosol by the proteasome, a large cytoplasmic multiprotein enzyme complex (see Fig. 46-1). Specialized transporter proteins, called *transporter in antigen processing* (TAP), facilitate translocation of peptides from the cytosolic proteasome to the endoplasmic reticulum. There, the peptides bind to newly formed MHC class I molecules and are transported to the cell surface, where they are recognized by antigen-specific $CD8^+$ T cells. MHC class I–associated peptides may also originate in the extracellular environment and be presented to T cells through the appropriately named *cross-presentation pathway*. This enables $CD8^+$ T cells to recognize foreign peptides, for example, from viruses, that

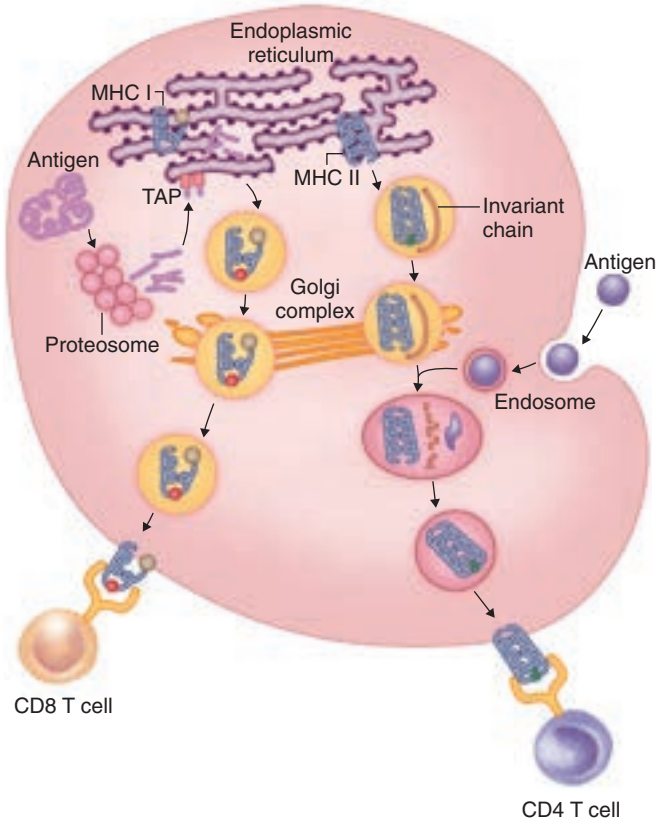

FIGURE 46-1. Pathways of antigen processing and delivery to major histocompatibility complex (MHC) molecules. Cytosolic proteins are broken down by the proteosome to generate peptide fragments, which are transported into the endoplasmic reticulum by specialized peptide transporters (TAP). After peptides are bound to MHC class I molecules, MHC-peptide complexes are released from the endoplasmic reticulum and travel to the cell surface, where they are ligands for $CD8^+$ T-cell receptors (TCRs). Extracellular foreign antigens are taken into intracellular vesicles, called *endosomes*. As the pH in the endosomes gradually decreases, proteases are activated that digest antigens into peptide fragments. After fusing with vesicles that contain MHC class II molecules, antigenic peptides are placed in the antigen-binding groove. Loaded MHC class II–peptide complexes are transported to the cell surface, where they are recognized by the TCRs of $CD4^+$ T cells.

are derived from infected and dying cells that are ingested by myeloid cells and then presented by MHC class I molecules.

The nature of the antigen-processing pathway determines the sequence of events in immune responses. Extracellular antigens, in general, enter the endosomal pool and associate with MHC class II molecules to stimulate $CD4^+$ T cells. Cytosolic antigens, including antigens from intracellular infectious agents, are degraded and displayed in the context of MHC class I molecules to initiate $CD8^+$ T-cell responses.

CELLULAR ELEMENTS OF THE ADAPTIVE IMMUNE SYSTEM

T Cells

T-Cell Development

T precursor cells are derived from hematopoietic stem cells that migrate to the thymus, a primary lymphoid tissue, where all the subsequent stages of T-cell maturation occur (Fig. 46-2). Pre-T cells express two enzymes, recombinase and terminal deoxynucleotidyl transferase, enabling them to recombine TCR genes. The β chain of the TCR is rearranged first and is expressed together with a pre-TCR α chain. Signals from the immature TCR complex inhibit rearrangement of the second β-chain allele and induce thymocyte proliferation and expression of both CD4 and CD8 molecules, so-called double positive thymocytes. Subsequently, the TCR α chain is recombined, with formation of a mature TCR. From here, the thymocyte undergoes many differentiation and selection steps modulated by the thymic microenvironment, with the end result being formation of a T cell that is ready to migrate to secondary lymphoid tissues and to be poised to recognize antigenic peptides. Early-stage thymocytes reside in the thymic cortex, where they mostly interact with epithelial cells. They then migrate toward the medulla, encountering dendritic cells and macrophages at the corticomedullary junction. Thymic stromal cells regulate T-cell proliferation by secreting lymphopoietic growth factors, such as interleukin-7 (IL-7). Interactions of the TCR with MHC molecules expressed on epithelial cells and on dendritic cells or macrophages determine the fate of the thymocyte.[1] Low-avidity recognition of peptide-MHC complexes on thymic epithelial cells by the TCR results in positive selection.[2] This recognition event rescues cells from apoptotic cell death and ensures that only T cells with functional receptors that can recognize MHC molecules, critical for T-cell activation on subsequent residence in the spleen and lymph nodes, survive. Thymocytes that express a receptor not fitting any MHC antigen complex die by neglect. High-affinity interaction between the TCR and peptide-MHC complex induces apoptotic death of the recognizing T cell. This process of negative selection eliminates T cells with specificity for self-antigens and is responsible for central tolerance to many autoantigens. It has been estimated that approximately 1% of thymocytes survive the stringent selection process. While undergoing selection, T cells continue to differentiate, with orderly expression of cell surface molecules. Double-positive thymocytes expressing both CD4 and CD8 molecules downregulate one or the other, developing into single-positive $CD4^+$ helper T cells that have been selected on MHC class II complexes or $CD8^+$ cytotoxic T cells that are restricted to MHC class I complexes. These single-positive cells are now mature T cells that are ready for exit and migration through the circulation to secondary lymphoid organs, including the spleen, lymph nodes, and mucosal lymphoid tissues, following chemokine cues and using adhesion molecules to enter. They exist in these tissues as inactivated, or naïve, cells until receiving the appropriate antigenic signal for activation and subsequent effector function.

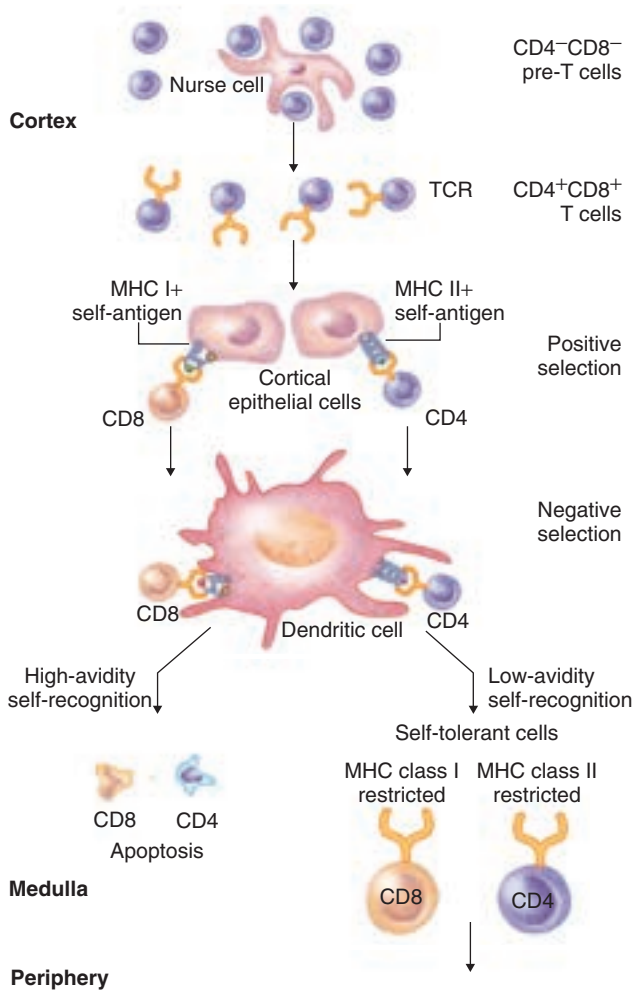

FIGURE 46-2. Maturation of T cells in the thymus. Precursors committed to the T-cell lineage arrive in the thymus and begin to rearrange their T-cell receptor (TCR) genes. Immature T cells with receptors binding to self–major histocompatibility complex (MHC) on cortical epithelial cells receive signals for survival (positive selection). At the corticomedullary junction, surviving T cells probe self-antigens presented by dendritic cells and macrophages. T cells reacting strongly to self-antigens are deleted by apoptosis (negative selection). T cells released into the periphery are tolerant toward self and recognize foreign antigens in the context of self-MHC.

T-Cell Stimulation and Accessory Molecules

T-cell activation is initiated when TCR complexes recognize antigenic peptides in the context of the appropriate MHC molecule on the surface of an antigen-presenting cell in secondary lymphoid organs. The principal antigen-presenting cells for activation of naïve T cells are dendritic cells. MHC-peptide recognition by the TCR, the first signal for T-cell activation, leads to receptor clustering and phosphorylation of the intracellular portion of the CD3 protein complex, the signaling component of the TCR, by receptor-associated tyrosine kinases. These events transmit signals to the nucleus of the T cell and initiate its activation. The coreceptors CD4 and CD8 are also critical for the initial events in T-cell activation, through their interaction with MHC class II and class I molecules, respectively, supporting CD3-mediated signals. Yet, this first activation signal delivered by the TCR and coreceptors is not alone sufficient for robust T-cell survival and differentiation. It needs to be complemented by the interaction of accessory molecules on the T cell and their ligands on the antigen-presenting cell. A spectrum of accessory molecules is known, of which the best known is CD28, which are engaged by CD80 and CD86 (also known as B7.1 and B7.2, respectively) on antigen-presenting cells (E-Table 46-1). Engagement of CD28 provides to the T cells a second, or costimulatory, signal to the T cell.[3] This second signal, delivered by the antigen-presenting dendritic cell, ensures T-cell survival and expansion. CD28-mediated signals are mandatory for the expression of many activation markers on the responding T cells and, in particular, for the secretion of IL-2. In the absence of such a second signal, T cells are rendered nonresponsive and anergic or undergo apoptosis. Finally, adhesion molecules (integrins) stabilize the interactions between T cells and antigen-presenting cells.

Signals from the TCR result in the activation of many genes and entry of the T cell into the cell cycle. The signals are transmitted by a cascade of cytoplasmic events. Cross-linking of the TCR and associated CD3 molecules results in the recruitment and activation of phosphotyrosine kinases and the phosphorylation of molecular constituents of the TCR and various adapter molecules. Signals mediated through the TCR then activate several biochemical pathways, which collectively lead to the activation of transcription factors that regulate gene expression

Three major variables determine the outcome of TCR stimulation: the duration and affinity of the TCR-antigen interaction, the maturation stage of the responding T cell, and the nature of the antigen-presenting cell. Antigen-presenting cells are gatekeepers in the initiation of T-cell responses. They can

E-TABLE 46-1 CYTOKINES AND CYTOKINE FUNCTION

CYTOKINES	MAJOR PRODUCER CELLS	PRINCIPAL ACTION
HEMATOPOIETIN FAMILY		
IL-2	T cells	Proliferation of T cells, B cells, and NK cells
IL-3	T cells	Early hematopoiesis
IL-4	T cells, mast cells	B-cell activation, IgE switch, inhibition of T_H1 cells
IL-5	T cells, mast cells	Eosinophil growth and differentiation
IL-6	Macrophages, endothelial cells	T-cell and B-cell growth and differentiation, induction of acute phase proteins
IL-7	Bone marrow, thymic epithelium	Growth of pre-B cells and pre-T cells
IL-9	T cells	Stimulation of mast cells and T_H2 cells
IL-11	Stromal fibroblasts	Hematopoiesis
IL-13	T cells	B-cell growth and differentiation, inhibition of T_H1 cells and macrophages
G-CSF	Fibroblasts and monocytes	Neutrophil development and differentiation
IL-15	Non-T cells	Growth of T cells and NK cells
GM-CSF	Macrophages, T cells	Growth and differentiation of myelomonocytic lineage cells
INTERFERON FAMILY		
IFN-α	Leukocytes	Antiviral, increases MHC class I expression
IFN-β	Fibroblasts	Antiviral, increases MHC class I expression
IFN-γ	T cells, NK cells	Macrophage activation, increases expression of MHC molecules, Ig class switching, inhibition of T_H2 cells
TUMOR NECROSIS FACTOR FAMILY		
TNF-α	Macrophages, NK cells, T cells	Induction of pro-inflammatory cytokines, endothelial cell activation, apoptosis
TNF-β (LT-α)	T cells, B cells	Cell death, endothelial activation, lymphoid organ development
LT-β	T cells, B cells	Cell death, lymphoid organ development
OTHERS		
TGF-β	Monocytes, T cells	Anti-inflammatory, inhibits cell growth, induces IgA secretion
IL-1α, IL-1β	Macrophages, endothelial cells	Acute phase response, fever, macrophage activation, costimulation
IL-10	T cells, macrophages	Suppression of macrophage functions
IL-12	Macrophages, dendritic cells	NK cell activation, T_H1 cell differentiation
IL-16	T cells, mast cells, eosinophils	Chemoattractant for CD4 T cells, monocytes, and eosinophils
IL-17	CD4 memory cells	Cytokine production by epithelia, endothelial cells, and fibroblasts
IL-18	Macrophages	IFN-γ production by T cells and NK cells
IL-23	Macrophages, dendritic cells	T_H17 cell differentiation

CD = cluster of differentiation; G-CSF = granulocyte colony-stimulating factor; GM-CSF = granulocyte-macrophage colony-stimulating factor; IFN = interferon; Ig = immunoglobulin; IL = interleukin; LT = lymphotoxin; MHC = major histocompatibility complex; NK = natural killer; TGF = transforming growth factor; T_H = helper T lymphocyte; TNF = tumor necrosis factor.

upregulate the expression of accessory molecules that provide costimulatory signals. MHC-peptide complexes are particularly dense on dendritic cells, enabling them to activate naïve T cells. In contrast, memory and effector cells have a lower threshold for activation and can react to antigens presented on peripheral tissue cells.

T-Cell Differentiation and Effector Functions

T-cell activation induces T-cell proliferation, with the goal of clonally selecting and expanding antigen-specific T cells. The extent of clonal proliferation is impressive. Antigen-specific $CD8^+$ T cells expand several thousand–fold; $CD4^+$ T cells expand somewhat less. During the phase of rapid growth, T cells differentiate from naïve T cells that are essentially devoid of effector functions into effector T cells that are needed for clearance of infectious organisms, or pathogens. The transition into effector cells is associated with a fundamental shift in functional profiles. First, effector T cells have a lower activation threshold; they do not require costimulation and can scan tissues that lack professional antigen-presenting cells. Second, they switch the expression of chemokine receptors and adhesion molecules to gain access to peripheral tissues. Finally, they gain effector functions.

The principal effector function of $CD8^+$ T cells is to lyse infected, antigen-bearing target cells. This commitment to eventual cytotoxic function is made during development in the thymus. Upon emigration from the thymus in the naïve, or inactivated, state, $CD8^+$ T cells circulate through secondary lymphoid tissues, surveying antigen-presenting dendritic cells for the appropriate MHC class I–peptide complex that can engage the TCR and that can supply costimulatory signals. On activation, $CD8^+$ T cells acquire cytotoxic functions and, using a variety of receptors and adhesion molecules, can emigrate from secondary lymphoid organs to peripheral tissue sites seeking cells infected by viruses or intracellular bacteria displaying pathogen-derived peptides on MHC class I molecules. On recognizing the appropriate MHC class I–peptide complex, $CD8^+$ T cells induce apoptosis of target cells. The T cell polarizes toward the area of antigen contact; specialized lytic granules are clustered in the contact area. A pore-forming protein, perforin, is released from the lytic granules and inserted into the target cell membrane. Proteases (granzymes) are injected into the target cells to initiate the apoptotic process by activating enzyme cascades. Mechanisms deployed by $CD8^+$ T cells are essentially identical to those of natural killer (NK) cells. $CD4^+$ T cells can also induce apoptosis but by a different mechanism than $CD8^+$ T cells. On activation, they express cell surface molecules such as Fas ligand (CD178) and TRAIL, which initiate the apoptotic cascade selectively in cells expressing the respective ligands Fas (CD95) or the death receptors DR4 and DR5.

Compared with $CD8^+$ T cells, the spectrum of options for $CD4^+$ T cells is larger. They are generally characterized as helper T, or T_H cells, because they produce cytokines and express cell surface molecules that promote the effector function of other lymphocytes and phagocytes. Like $CD8^+$ T cells, they are initially activated in secondary lymphoid tissues on contact with dendritic cells displaying the MHC–peptide complex (MHC class II, compared with MHC class I for $CD8^+$ T-cell activation) bound by a specific TCR along with the proper costimulatory signals. On activation, different subsets of $CD4^+$ effector T cells can be distinguished based on the preferential production of certain cytokines (see E-Table 46-1). T_H1 T cells predominantly produce interferon-γ (IFN-γ) and tumor necrosis factor-α (TNF-α) and are involved in cell-mediated immunity, such as delayed-type hypersensitivity reactions. These cytokines, among other actions, promote macrophage activation that is critical for protective responses against intracellular pathogens such as mycobacteria and listeria. T_H2 T cells preferentially produce IL-4, IL-5, and IL-13, cytokines that promote eosinophil maintenance, expansion, and tissue accumulation, as well as macrophage function; these are all important for host protection following infection with helminths, such as schistosomes and other worms. T_H17 T cells produce IL-17, critical for neutrophil expansion and function, with killing of extracellular bacteria, such as streptococci, and pathogenic fungi.[4] These cells may also produce IL-22 that promotes host-protective function at barrier surfaces, such as the skin and gut. Follicular helper T (T_{FH}) cells home to lymphoid follicles, where B cells congregate, where they express CD40 ligand (CD154) and other surface proteins along with cytokines, including IL-21, IL-4, and IFN-γ, that are critical for B-cell maturation to plasma cells and memory B cells. The decision as to which differentiation pathway to take is made during the early stages of naïve T-cell activation by antigen-presenting cells in secondary lymphoid organs. Pathway differentiation depends on several factors, including (1) the cytokines produced by the activating antigen-presenting cell and other innate cells in the microenvironment, (2) the nature of costimulatory signals, and (3) the avidity of the TCR–MHC antigen interaction. $CD4^+$ T-cell subset, or lineage, development is generally correlated with the expression of specific transcription factors (T-bet for T_H1, GATA3 for T_H2, RORγt for T_H17, and Bcl6 for T_{FH} cells). However, lineage commitment among differentiated $CD4^+$ T cells is not absolute and is not terminal, and transition between different effector types is possible.

Regulatory T Cells

Depending on their cytokine profile, $CD4^+$ T cells have the ability to cross-regulate each other, influence T-cell differentiation, and suppress T-cell effector activity. Classic examples of T cells with regulatory activity generated during the normal immune response are IL-10– and transforming growth factor-β (TGF-β)–producing cells. In addition, specialized subsets of regulatory T (Treg) cells are characterized by expression of the transcription factor forkhead box P3 (Foxp3). Naturally occurring $Foxp3^+$ Treg cells are generated during T-cell development in the thymus and recognize self-antigens. $Foxp3^+$ Treg cells can also arise from conventional $CD4^+$ T cells in the periphery. Natural and inducible Treg cells are in many ways indistinguishable, particularly because their development and function depend on Foxp3, and they are able to suppress T-cell expansion and constitutively express several cell surface markers, albeit markers that are not necessarily specific for Treg because activated T cells can also express them. Treg cells are important in peripheral tolerance, controlling the expansion of autoreactive T cells. They also play a role in immune responses to pathogens by virtue of their ability to suppress T-cell effector function and consequently downmodulate the inflammatory response incited by the former, a natural consequence of pathogen elimination. A principal difference between natural and induced Treg cells is that the latter largely survey mucosal and other environmentally exposed surfaces. Despite extensive studies in various models, the mechanism by which Treg cells function in vivo remains incompletely understood, although it is certainly a consequence of secretion of regulatory cytokines, like IL-10 and TGF-β, that can dampen inflammatory responses. Tregs may also express the T-cell molecule cytotoxic T-lymphocyte antigen (CTLA)-4 (CD152) that, like CD28, engages CD80 and CD86 on antigen-presenting cells. In contrast to CD28, which receives a positive signal from CD80 and CD86, leading to robust T-cell activation, engagement of these molecules on antigen-presenting cells by CTLA-4 on Tregs suppresses the ability of antigen-presenting cells to activate naïve T cells.

T-Cell Homeostasis

Effective immunity depends on the ability of the immune system to generate large numbers of antigen-specific T cells rapidly, yet the space in the T-cell compartment is limited. To avoid competition for space and resources and to prevent perturbation of T-cell diversity by lifelong exposure to antigens, the adaptive immune system employs several counterbalancing mechanisms. In the later stages of the activation process, a strong negative signal derives from interaction of CTLA-4 with CD80/CD86 on antigen-presenting cells. In addition, T cells undergo activation-induced cell death. Activated $CD4^+$ T cells begin to secrete Fas ligand and acquire sensitivity to Fas-mediated death, inducing apoptotic suicide and fratricide in neighboring T cells. These mechanisms impose constraints in the early stages of the T-cell antigen response. Other mechanisms control the rapid decline of expanded antigen-specific T cells when elimination of the antigen has been achieved. Removal of the driving antigen causes a deprivation of cytokines and costimulatory molecules, and growth factor–deprived T cells die from apoptosis. It has been estimated that only 5% of the antigen-expanded population survives after antigen clearance, becoming memory cells that are poised to respond if the host is again challenged by the same offending pathogen.

B Lymphocytes

B-Cell Development

B cells are generated in the bone marrow, like the thymus, a primary lymphoid organ. Lymphoid stem cells differentiate into distinctive B-lineage cells in the marrow, supported by a specialized microenvironment of nonlymphoid stromal cells supplying necessary chemokines, including stromal cell–derived factor 1 and cytokines (IL-7). Precursor B cells enter a process of tightly controlled sequential rearrangements of heavy chain and light chain immunoglobulin genes. On pre-B cells, the membrane μ chain is associated with a surrogate light chain to form a pre-B-cell receptor (BCR). Signals provided through this receptor induce proliferation of progeny that subsequently rearrange different light chain gene segments.

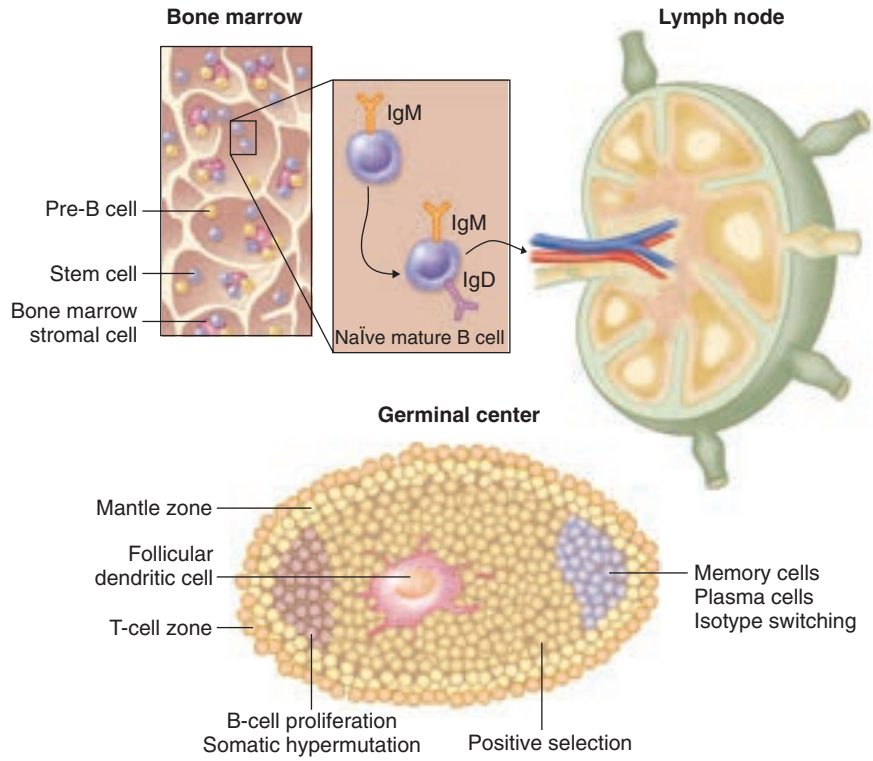

FIGURE 46-3. B-cell development and differentiation. The early stages of B-cell development occur in the bone marrow, with cells progressing through a developmental program determined by the rearrangement and expression of immunoglobulin (Ig) genes. Immature B cells with receptors for multivalent self-antigens die in the bone marrow. Surviving B cells coexpress IgD and IgM surface receptors. They are seeded into peripheral lymphoid organs, where they home to selected locations and receive signals to survive and become longer-lived naïve B cells. Antigen-binding B cells and antigen-presenting B cells that receive help from antigen-specific T cells are activated through membrane-bound and secreted molecules. Activated B cells migrate into the follicles, leading to the formation of germinal centers. B cells in germinal centers undergo somatic hypermutation of immunoglobulin genes; cells with high affinity for antigens presented on the surface of follicular dendritic cells are selected to differentiate into either memory B cells or plasma cells.

It is estimated that only 10% of B cells generated in the bone marrow reach the circulating pool. Losses are mostly due to negative selection and clonal deletion of immature B cells that express receptors directed against self-antigens. Cross-linking of surface IgM by multivalent self-antigens causes immature B cells to die. Such self-reactive B cells can be rescued from death by replacing the light chain with a newly rearranged light chain that is no longer self-reactive, a process named *receptor editing*.[5] On maturation, B cells begin to express surface IgD. B cells positive for IgD and IgM are exported from the bone marrow and migrate to peripheral lymphoid tissues following a chemokine gradient, in a process analogous to the migration of naïve T cells from the thymus to the same tissues (Fig. 46-3). There, colocalization of both types of lymphocytes facilitates their interaction following pathogen challenge. This enables B cells to receive T-cell help for the former's activation and subsequent function, including memory development and antibody secretion, required for responses to protein antigens.

B-Cell Stimulation

Mature, but naïve, B cells in secondary lymphoid organs are activated by soluble and cell-bound antigens to develop into antibody-secreting effector cells. B cells respond to a large variety of antigens, including proteins, polysaccharides, and lipids. Binding of antigen to cell surface IgM molecules induces BCR clustering, the initial step in B-cell activation. In addition to the antigen-binding immunoglobulin, the BCR comprises two proteins, Ig-α and Ig-β. The Ig-α/Ig-β heterodimer functions to transduce a signal and initiates the intracellular signaling cascade, analogous to the CD3 molecule of the TCR. Thus, the composition of the BCR, with ligand-binding and signal-transducing units, and the signaling events that lead to gene induction, are similar to those of the TCR. BCR triggering is enhanced by coreceptors, as for the TCR. The BCR-coreceptor complex is composed of CD81, CD19, and CD21, analogous to the TCR coreceptors CD4 and CD8. CD21 binds to complement fragments on opsonized antigens that are bound by the BCR, resulting in phosphorylation of the intracellular tail of CD19 by tyrosine kinases and augmentation of the BCR-mediated signal.

Like naïve T cells, naïve B cells require accessory signals in addition to triggering of their antigen-binding receptor. They receive second signals either from follicular helper T cells or from microbial components. Microbial constituents, such as bacterial polysaccharides, can induce antibody production in the absence of helper T cells, comprising thymus-independent, or T-independent, antigens.[6] In contrast, in the case of protein antigens, which are thymus- or T-dependent, the initial BCR stimulation prepares the cell for subsequent interaction with follicular helper T cells. These activated B cells start to enter the cell cycle; upregulate cell surface molecules, such as CD80 and CD86, that provide costimulatory signals to T cells; and upregulate certain cytokine receptors. As such, these B cells are prepared to activate helper T cells and to respond to cytokines secreted by those T cells, but they cannot differentiate into antibody-producing cells in the absence of T-cell help. Survival and differentiation factors produced by myeloid cells, such as B-cell-activating factor (BAFF), also stimulate B cells and help to maintain the B-cell pool.[7]

B-Cell Differentiation

Differentiation of B cells activated by protein antigens depends on interaction with helper T cells. B cells use their antigen receptor not only to recognize antigens but also to internalize them. After processing endocytosed antigens, MHC class II–peptide complexes appear on the cell surface, where antigen-specific $CD4^+$ T cells detect them. Also, B cells express costimulatory molecules and provide optimal conditions for T-cell activation. On activation, $CD4^+$ T cells express CD154, also known as CD40 ligand, on their surface and are able to stimulate the CD40 molecule on their B-cell partner. CD40-CD154 interaction is essential for subsequent B-cell proliferation and differentiation. Cytokines secreted by the helper T cells act in concert with CD154 to amplify B-cell differentiation and to determine the antibody type by controlling isotype switching. Isotypes greatly influence the versatility of antibodies as effector molecules, and cytokines drive isotype switching by stimulating the transcriptional activation of heavy chain constant region genes and enabling switching from transcription of the IgM heavy chain gene to that of IgG, IgA, or IgE.

T-cell-dependent B-cell differentiation and maturation take place in germinal centers, specialized areas in secondary lymphoid tissues where B cells rapidly proliferate, with mutation of the variable, or antigen-binding portion, of their immunoglobulin surface receptors (BCRs) (see Fig. 46-3). Those B cells bearing receptors with the highest affinity for antigen are selected for

survival with the help of specific signals delivered by follicular helper T cells, whereas those with lesser affinity die by apoptosis. This process enables affinity maturation of B cells that most efficiently bind antigen and thereby facilitate its removal. As somatic hypermutation and affinity maturation proceed in the germinal center, isotype class switching of the immunoglobulin receptors is also occurring.[8]

Lymphocytes and Lymphoid Tissue

The initiation of adaptive immune responses depends on rare antigen-specific T cells and B cells meeting antigen-presenting cells and their relevant antigen. The recognition of a specific antigen in the tissue by uncommon T cells has a low probability, and it is unlikely that sufficient numbers of antigen-presenting cells and lymphocytes can be brought together to provide crucial momentum. The immune system uses specialized lymphoid microstructures to bring antigens to the site of lymphocyte traffic and accumulation. Secondary lymphoid organs include the spleen for blood-borne antigens, the lymph nodes for antigens encountered in peripheral tissues, and the mucosa-associated, bronchial-associated, and gut-associated lymphoid tissues, where antigens from epithelial surfaces are collected. Lymphocytes circulate through secondary lymphoid organs, constantly searching for their antigen. Their homing to lymph nodes is facilitated by specialized microvessels, called *high endothelial venules*, which provide the proper structure for them to leave the circulation and enter the tissue. Secondary lymphoid tissues have developed several strategies to sequester the relevant antigen. Antigens in peripheral tissue are encountered first by dendritic cells that, after activation, are mobilized to transport antigens into the local lymph nodes by the draining lymph. These antigen-bearing dendritic cells enter the lymph nodes through the afferent lymphatic vessel and settle in the T-cell-rich zones to present processed antigens to T cells. The net result of this process is an accumulation and concentration of the antigen in an environment that can be readily screened by infrequent antigen-specific T cells.

B cells are segregated from T cells in the lymph nodes and are localized in follicles. If, on antigen engagement, B cells find their cooperating (cognate) T cells at the borders of the T-cell-rich areas and the follicle, they receive cues to enter germinal centers along with their cognate follicular helper T cells. Germinal centers contain a network of follicular dendritic cells that capture particulate antigen or immune complexes on the cell surface. This unprocessed antigen is taken up by antigen-specific B cells, processed and presented, and recognized by antigen-specific T_{FH} cells. These T cells provide cytokines and cell-cell contact signals to support the germinal center reaction, a process that includes somatic hypermutation, affinity selection, and isotype switching (see Fig. 46-3). Germinal centers are essential for generating long-lived antibody-secreting plasma cells and memory B cells.

Lymphoid organ development is highly dependent on environmental cues. The symbiotic relationship between the host immune system and microorganisms is best exemplified in the gastrointestinal tract. Development of gut-associated lymphoid tissue is absolutely dependent on bacterial colonization. Increasing evidence suggests that host-symbiont interactions regulate adaptive immune functions throughout life. Disturbances in the bacterial microbiota and failure to maintain intestinal homeostasis are important in diverse diseases, including inflammatory bowel disease (Chapter 143) and HIV-associated immune defects.

Memory

An important consequence of adaptive immunity is the generation of immunologic memory, the basis for long-lived protection after a primary infection. Memory induction by vaccination is one of the landmark successes in medicine. Immunologic memory is defined as the ability to respond more rapidly and effectively to pathogens that have been encountered previously. The bases of immunologic memory are qualitative and quantitative changes in antigen-specific T cells and B cells. As a direct result of clonal expansion and selection in antigen-driven responses, the frequencies of antigen-specific memory B cells and memory T cells are increased 10-fold to 1000-fold compared with the naïve repertoires. The mechanisms through which memory T cells and B cells escape clonal downsizing in the terminal stages of the primary immune response are consequences of upregulation of a selected group of transcription factors that ensure survival. The enrichment of antigen-specific B cells and T cells enhances the sensitivity of the system to renewed challenges and provides a head start of 4 to 10 cell divisions. In addition to increased frequencies, memory T cells and B cells are functionally different from their naïve counterparts. Memory cells are long-lived and survive in the presence of certain cytokines without the need for continuous antigenic stimulation, guaranteeing immunologic memory for the life expectancy of the individual cell. Memory B cells produce predominantly IgG and IgA antibodies with evidence of somatic hypermutation and high affinity for the antigen. Cell surface expression of high-affinity antibodies allows more efficient antigen uptake, which enhances the crucial interaction with T cells. On antigen encounter, memory B cells change to antibody-secreting plasma cells, or re-enter the germinal center, where the high affinity of their immunoglobulin receptor gives them a competitive advantage over naïve B cells in antigen binding, leading to progressive affinity maturation of somatically mutated antibody molecules.

Because the TCR does not undergo isotype switching or affinity maturation, memory T cells are more difficult to distinguish from naïve or effector T cells. In contrast to effector cells, memory T cells lack activation markers and need antigen stimulation to resume effector functions. In contrast to naïve T cells, memory T cells have a lower activation threshold and are less dependent on costimulatory signals. In essence, their requirements for antigen stimulation are fewer, and their clonal size is larger, permitting fast, efficient responses to secondary antigen encounters. Also, memory T cells resume effector functions without having to undergo cell divisions.

Immunologic Tolerance and Autoimmunity

Unresponsiveness to self is a fundamental property of the immune system and is a condition, *sine qua non*, to maintain tissue integrity of the host. Self/nonself distinction is relatively straightforward for the innate immune system, in which receptors to nonself molecules are genetically encoded and evolutionarily selected. Self/nonself discrimination is much more complex for the adaptive immune system, in which antigen-specific receptors are generated randomly and the entire spectrum of antigens can be recognized. Thus, the adaptive immune system must acquire the ability to distinguish between self and nonself. Several different mechanisms are used, collectively called *tolerance*. Tolerance is antigen specific; its induction requires the recognition of antigen by lymphocytes in a defined setting. Failure of self-tolerance results in immune responses against self-antigens. Such reactions are called *autoimmunity* and may give rise to chronic inflammatory autoimmune disease.

Central and peripheral tolerance mechanisms can be distinguished. In central tolerance, self-reactive lymphocytes are deleted during development. This process of negative selection is particularly important for T cells. During thymic development, T cells that recognize antigen with high affinity, in particular antigens that are constitutively expressed on antigen-presenting cells, are deleted. Central tolerance for B cells follows the same principles. Recognition of antigen by developing B cells in the bone marrow induces apoptosis, or receptor editing that replaces the self-reactive receptor with one containing the product of a newly rearranged light chain gene. Negative selection is particularly important for B cells that recognize multivalent antigens because they do not depend on T-cell help and cannot be controlled peripherally.

Not all self-reactive T cells are centrally purged from the repertoire; certain antigens are not encountered at sufficient densities in the thymus. Also, all T cells have some degree of self-reactivity, which is necessary for positive selection in the thymus and for peripheral survival. Mechanisms of peripheral T-cell tolerance include (1) anergy, (2) peripheral deletion, (3) clonal ignorance, and (4) suppression of immune responses by regulatory T cells. T-cell anergy is transient and is actively maintained. It is induced if $CD4^+$ T cells recognize antigens presented by MHC class II molecules without receiving costimulatory signals. In general, costimulatory molecules such as CD80 and CD86 are restricted to antigen-presenting cells, and their expression is dependent on microbial recognition, leading to activation of the antigen-presenting cells. MHC-peptide presentation to T cells by immature or inactivated, resting antigen-presenting cells or on any cell other than peripheral antigen-presenting cells results in anergy because these cells typically lack expression of costimulatory molecules. Tissue-residing immature dendritic cells need to be activated by cytokines or recognition of pathogen-associated molecular patterns (PAMPs) to stimulate and not to anergize T cells. A second tolerance mechanism, peripheral deletion, is induced as a consequence of hyperstimulation. Hyperstimulation of T cells (e.g., by high doses of antigen and high concentrations of IL-2) preferentially activates pro-apoptotic pathways and causes elimination of the responding T-cell specificity. This mechanism may be responsible for the elimination of T cells specific for plentiful peripheral self-antigens and for foreign antigens abundantly present during infection. Whereas induction of anergy and activation-induced cell death are active consequences of antigen recognition, the third tolerance mechanism, clonal ignorance, is less well understood. Clonal

ignorance is defined as the presence of self-reactive lymphocytes that fail to recognize or to respond to peripheral antigens. These cells remain responsive to antigenic challenge if given in the right setting. An example of clonal ignorance is nonresponsiveness to sequestered antigens that are not accessible to the immune system. Other mechanisms must exist, however, because clonal ignorance has also been shown for accessible antigens. Fourth, Treg cells play a pivotal role in maintaining peripheral tolerance. During an immune response, T cells can acquire the ability to produce regulatory cytokines, such as TGF-β, IL-10, or IL-4, that dampen or suppress immune responses. A dedicated subset of Treg cells, Foxp3 $CD4^+$ T cells, has been identified and characterized. Harnessing the frequencies and function of these cells may offer a promising approach to restoring peripheral tolerance in treating autoimmune diseases or facilitating transplantation tolerance; their elimination or functional suppression may potentiate cancer immunotherapy.

A critically important mechanism of peripheral tolerance of B cells is maintained through the absence of T-cell help. B cells require signals from T cells to differentiate into effector cells. B lymphocytes that recognize self-antigens in the periphery in the absence of T-cell help are rendered anergic or are unable to enter lymphoid follicles, where they could receive T-cell help, effectively excluding them from immune responses.

Generation and maintenance of self-tolerance can fail, in which case autoimmune responses are generated. Overall, chronic inflammatory diseases induced by tolerance failure occur in about 5% of the general population. Given the complexity of regulation, it is surprising that autoimmune diseases are not more frequent. It is thought that most autoimmune diseases result from dysfunction of the adaptive immune system, although activation of the innate immune system can set the stage for a self-reactive adaptive immune response. Many models of autoimmunity rely on the hypothesis that peripheral anergy is broken. Aberrant expression of costimulatory molecules on nonprofessional antigen-presenting cells or inappropriate activation of tissue-residing dendritic cells sets the stage for the induction of "forbidden" T-cell responses. Also, autoreactive B cells that recognize self-antigen complexed with foreign antigen may engulf this complex and receive help from T cells specific for the foreign antigen. Autoimmunity also may emerge if antigen ignorance is broken. This could happen if tissue barriers break down and antigens that are usually sequestered from the immune system, such as antigens from the central nervous system or the eye, become accessible. Tolerance mechanisms of anergy or clonal ignorance can also fail if a foreign antigen is sufficiently different from a self-antigen to initiate an immune response but sufficiently similar for activated T cells to elicit T-cell and B-cell effector functions (molecular mimicry).

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Klein L, Kyewski B, Allen PM, et al. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol*. 2014;14:377-391.
2. Fu G, Rybakin V, Brzostek J, et al. Fine-tuning T cell receptor signaling to control T cell development. *Trends Immunol*. 2014;35:311-318.
3. Hubo M, Trinschek B, Kryczanowsky F, et al. Costimulatory molecules on immunogenic versus tolerogenic human dendritic cells. *Front Immunol*. 2013;4:82.
4. Peters A, Yosef N. Understanding Th17 cells through systematic genomic analyses. *Curr Opin Immunol*. 2014;28:42-48.
5. Luning Prak ET, Monestier M, Eisenberg RA. B cell receptor editing in tolerance and autoimmunity. *Ann N Y Acad Sci*. 2011;1217:96-121.
6. Bortnick A, Allman D. What is and what should always have been: long-lived plasma cells induced by T cell-independent antigens. *J Immunol*. 2013;190:5913-5918.
7. Rickert RC, Jellusova J, Miletic AV. Signaling by the tumor necrosis factor receptor superfamily in B-cell biology and disease. *Immunol Rev*. 2011;244:115-133.
8. Matthews AJ, Zheng S, DiMenna LJ, et al. Regulation of immunoglobulin class-switch recombination: choreography of noncoding transcription, targeted DNA deamination, and long-range DNA repair. *Adv Immunol*. 2014;122:1-57.

REVIEW QUESTIONS

1. IgG antibodies are characterized by all of the following *except* which one?

A. Two heavy chains and two light chains
B. A common γ chain
C. Disulfide bonds
D. Subclasses

Answer: B IgG molecules consist of two identical heavy and two identical light chains, linked by disulfide bonds. IgG and IgA molecules include distinct subclasses encoded by distinct genomic DNA sequences. A common γ chain is a signaling component of certain cytokine and Fc receptors but is not a component of IgG antibodies.

2. $CD4^+$ T cells recognize antigenic peptides presented on which of the following?

A. Epithelial cells
B. Major histocompatibility complex (MHC) class I molecules
C. MHC class II molecules
D. Costimulatory molecules
E. CD80

Answer: C Antigenic peptides processed by specialized antigen-presenting cells associate with the peptide-binding cleft of MHC class II molecules and are recognized by the T-cell receptor expressed on $CD4^+$ T cells. $CD8^+$ T cells recognized antigenic peptides associated with MHC class I molecules. CD80 is an example of a costimulatory molecule expressed on antigen-presenting cells that supports the activation of T cells that have interacted with peptide-MHC complexes. Epithelial cells do not typically activate $CD4^+$ T cells.

3. Which of the following functions is *not* provided by differentiated T cells?

A. Production of interleukin-17 (IL-17)
B. Lysis of virus-infected target cells
C. Help for B-cell activation and differentiation
D. Production of BAFF
E. Production of interferon-γ

Answer: D The cytokine milieu of T cells contributes to differentiation of those cells to provide various effector functions. These include support for B-cell activation and differentiation (through CD154-CD40 interactions and production of IL-21), secretion of cytokines (interferon-γ) that support macrophage activation and delayed-type hypersensitivity reactions, induction of inflammatory responses (IL-17), and mediating the killing of virus-infected cells. Myeloid cells, rather than T cells, are the major producers of BAFF.

4. B cells minimize self-reactivity through a process called which of the following?

A. Isotype switching
B. Costimulation
C. Somatic hypermutation
D. Recombination
E. Receptor editing

Answer: E Isotype switching, somatic hypermutation, and recombination are processes that promote the diversity of the B-cell repertoire. Costimulation of T-cell activation supports amplification of an adaptive immune response. Receptor editing modifies the sequence of the B-cell receptor to avoid production of autoreactive B cells.

5. Immunologic tolerance is mediated by all of the following *except* which of the following?

A. Thymic deletion of T cells that recognize antigen with high affinity
B. Regulatory T cells
C. Toll-like receptor activation
D. T-cell activation in the absence of costimulation
E. It is mediated by all of the above.

Answer: C Central and peripheral tolerance mechanisms include thymic deletion of self-reactive T cells, T-cell activation without costimulation, and control of self-reactive cells by regulatory T cells. Toll-like receptor activation is an important feature of innate immune system activation.

47

MECHANISMS OF IMMUNE-MEDIATED TISSUE INJURY

JANE E. SALMON

THE ADAPTIVE IMMUNE RESPONSE

Definition

The adaptive immune response is a crucial component of host defense against infection. Its distinguishing and unique feature is the ability to recognize pathogens specifically, based on clonal selection of lymphocytes bearing antigen-specific receptors. Antigens unassociated with infectious agents also may elicit adaptive immune responses. Many clinically important diseases are characterized by normal immune responses directed against an inappropriate antigen, typically in the absence of infection. Immune responses directed at noninfectious antigens occur in allergy, in which the antigen is an innocuous foreign substance, and in autoimmunity, in which the response is to a self-antigen.

Effector mechanisms that eliminate pathogens in adaptive immune responses are essentially identical to those of innate immunity. The specific antigen recognition feature of the adaptive immune response seems to have been appended to the preexisting innate defense system. As a result, the inflammatory cells and molecules of the innate immune system are essential for the effector functions of B and T lymphocytes. In addition to initiating protective responses, they mediate tissue injury in allergy, hypersensitivity, and autoimmunity.

Effector Mechanisms

Effector actions of antibodies depend on recruiting cells and molecules of the innate immune system. Antibodies are adapters that bind antigens to nonspecific inflammatory cells and direct their destructive effector responses. Antibodies also activate the complement system, which enhances opsonization of antigens, recruits phagocytic cells, and amplifies (or "complements") antibody-triggered damage. The isotype or class of antibodies produced determines which effector mechanisms are engaged.

Cell-bound receptors for immunoglobulin (Ig) constitute the link between humoral and cellular aspects of the immune cascade and play an integral part in the process by which foreign and endogenous opsonized material is identified and destroyed. These cell-based binding sites for antibodies, termed *Fc receptors,* interact with the constant region (Fc portion) of the immunoglobulin heavy chain of a particular antibody class regardless of its antigen specificity. Accessory cells that lack intrinsic specificity, such as neutrophils, macrophages, and mast cells, are recruited to participate in inflammatory responses through the interaction of their Fc receptors with antigen-specific antibodies. Distinct receptors for different immunoglobulin isotypes are expressed on different effector cells.

Receptors for IgG (FcγRs) are a diverse group of receptors expressed as hematopoietic cell surface molecules on phagocytes (macrophages, monocytes, neutrophils), platelets, mast cells, eosinophils, and natural killer (NK) cells. FcγRs often are expressed as stimulatory and inhibitory pairs.[1] Triggering of stimulatory FcγRs initiates a series of events, including phagocytosis; antibody-dependent, cell-mediated cytotoxicity; secretion of granules; and release of inflammatory mediators, such as cytokines, reactive oxidants, and proteases. Extensive structural diversity among FcγR family members leads to differences in binding capacity, signal transduction pathways, and cell type–specific expression patterns. This diversity allows IgG complexes to activate a broad program of cell functions relevant to inflammation, host defense, and autoimmunity. Phagocyte activation is triggered by stimulatory FcγRs, facilitating the recognition, uptake, and destruction of antibody-coated targets, whereas multivalent IgG binding to FcγRs on platelets leads to platelet aggregation and thrombosis, and binding to FcγRs on NK cells mediates cytotoxicity of antibody-coated targets.

IgE binds to high-affinity FcεRs on mast cells, basophils, and activated eosinophils.[2] In contrast to FcγRs, which are low affinity and bind to multivalent IgG rather than circulating individual IgG molecules, FcεRs can bind monomeric IgE. A single mast cell may be armed with IgE molecules specific for different antigens, all bound to surface FcεRs. Mast cells, localized beneath the mucosa of the gastrointestinal and respiratory tracts and the dermis of the skin, await exposure to multivalent antigens, which cross-link surface IgE bound to FcεRs and cause release of histamine-containing granules and generation of cytokines and other inflammatory mediators. IgE-mediated activation of eosinophils, cells normally present in the connective tissue of underlying respiratory, urogenital, and gut epithelium, leads to the release of highly toxic granule proteins, free radicals, and chemical mediators such as prostaglandins, cytokines, and chemokines. These amplify local inflammatory responses by activating endothelial cells and recruiting and activating more eosinophils and leukocytes. Prepackaged granules and high-affinity FcεRs that bind to free monomeric IgE enable an immediate response to pathogens or allergens at the first site of entry, a location where FcεR-bearing cells reside.

Inhibitory FcγRs, which modulate activation thresholds and terminate stimulating signals, are key elements in the regulation of effector function. Given that inhibitory and stimulatory Fc receptors are often coexpressed on the same cells, the effector response to a specific stimulus in a particular cell represents the balance between stimulatory and inhibitory signals. Inhibitory FcγRs can dampen responses triggered by FcεRs on mast cells and FcγR-mediated inflammation at sites of immune complex deposition.

Effector activities targeted by IgG and IgM also may be mediated by components of the complement system (Chapter 50). Antigen-bound multimeric immunoglobulin can initiate activation of the classic pathway of

complement, causing enhanced phagocytosis of antigen-antibody complexes, increased local vascular permeability, and recruitment and activation of inflammatory cells. The target of injury is specified by the antibody, and the extent of damage is determined by the synergistic activities of immunoglobulin and complement.

Antigen-specific effector T cells also may initiate tissue injury. On exposure to an appropriate antigen, memory T cells are stimulated to release cytokines and chemokines that activate local endothelial cells and recruit and activate macrophages and other inflammatory cells. The effector cells directed by T-cell-derived cytokines, or cytolytic T cells themselves, mediate tissue damage. T helper 1 (T_H1) cells produce interferon-γ (IFN-γ) and activate macrophages to cause injury, whereas T_H2 cells produce interleukin-4 (IL-4), IL-5, and eotaxin (an eosinophil-specific chemokine) and trigger inflammatory responses in which eosinophils predominate. T_H17 cells secrete several effector molecules, including IL-17, which act on both immune and nonimmune cells to trigger differentiation; release of antimicrobial molecules, cytokines, and chemokines; and recruitment to sites of inflammation.[3] New T_H effector subsets have recently been identified, including follicular T helper cells (T_{FH}), which provide help to B cells in germinal centers and thus are key regulators of humoral responses and antibody production.

HYPERSENSITIVITY REACTIONS

In predisposed individuals, innocuous environmental antigens may stimulate an adaptive immune response, immunologic memory, and, on subsequent exposure to the antigen, inflammation. These "overreactions" of the immune system to harmless environmental antigens (allergens), called *hypersensitivity* or *allergic reactions,* produce tissue injury and can cause serious disease. Hypersensitivity reactions are grouped into four types according to the effector mechanisms by which they are produced (Table 47-1). The effectors for types I, II, and III hypersensitivity reactions are antibody molecules, whereas type IV reactions are mediated by antigen-specific effector T cells.[4]

Autoimmune disease is characterized by the presence of antibodies and T cells specific for self-antigens expressed on target tissues. The mechanisms of antigen recognition and effector function that lead to tissue damage in autoimmune disease are similar to the mechanisms elicited in response to pathogens and environmental antigens. These mechanisms resemble certain hypersensitivity reactions and may be classified accordingly (Table 47-2). Autoimmune disease caused by antibodies directed against cell surface or extracellular matrix antigens corresponds to type II hypersensitivity reactions; disease caused by formation of soluble immune complexes that subsequently are deposited in tissue corresponds to type III hypersensitivity; and disease caused by effector T cells corresponds to type IV hypersensitivity. Typically, several of these pathogenic mechanisms are operative in autoimmune disease. However, IgE responses are not associated with damage in autoimmunity.

Type I Hypersensitivity Reactions

Type I hypersensitivity reactions (Fig. 47-1) are triggered by the interaction of antigen with antigen-specific IgE bound to FcεRs on mast cells, which

TABLE 47-1 FOUR MAJOR TYPES OF IMMUNOLOGICALLY MEDIATED HYPERSENSITIVITY REACTIONS*

IMMUNOLOGIC SPECIFICITY	TYPE I (IgE ANTIBODY)	TYPE II (IgG ANTIBODY)	TYPE III (IgG ANTIBODY)	TYPE IV (T CELLS)			
				T_H1 Cells	*T_H2 Cells*	*T_H17 Cells*	*T Cells*
Antigen	Soluble antigen allergen	Cell- or matrix-associated antigen	Soluble antigen	Soluble antigen	Soluble antigen	Soluble antigen	Cell-associated antigen
Effector mechanism	FcεRI- or FcγRIII-dependent mast cell activation, with release of mediators/ cytokines	$Fc\gamma R^+$ cells (phagocytes, NK cells), complement	$Fc\gamma R^+$ cells, complement	Macrophage activation	Eosinophil activation	Macrophage activation Neutrophil activation	Direct cytotoxicity
Examples	Systemic anaphylaxis, asthma, allergic rhinitis, urticaria, angioedema	Certain drug reactions and reactions to incompatible blood transfusions	Arthus reaction and other immune complex–mediated reactions (e.g., serum sickness, subacute bacterial endocarditis)	Contact dermatitis, tuberculin reaction	Chronic allergic inflammation (e.g., chronic asthma, chronic allergic rhinitis)	Contact dermatitis, atopic dermatitis, asthma, rheumatoid arthritis	Contact dermatitis (e.g., poison ivy), reactions to certain virus-infected cells, some instances of graft rejection

*Hypersensitivity reactions were classified into four types by Coombs and Gell (1963) and modified by Janeway and colleagues (2001).
FcγR = Fc receptor for immunoglobulin G; FcεR = Fc receptor for immunoglobulin E; NK = natural killer.
From Coombs RRA, Gell PGH: Classification of allergic reactions responsible for clinical hypersensitivity and disease. In: Gell PGH, Coombs RA, eds. *Clinical Aspects of Immunology.* Oxford, UK: Blackwell; 1963; and Janeway C, Travers P, Walport M, Shlomchick M: *Immunobiology: The Immune System in Health and Disease.* 5th ed. New York: Garland Publishing; 2001.

TABLE 47-2 CLASSIFICATION OF AUTOIMMUNE DISEASES ACCORDING TO MECHANISM OF TISSUE INJURY

HYPERSENSITIVITY REACTION	AUTOIMMUNE DISEASE	AUTOANTIGEN
TYPE II		
Antibody against cell surface antigens	Autoimmune hemolytic anemia	Rh blood group antigens, I antigen
	Autoimmune thrombocytopenic purpura	Platelet integrin glycoprotein IIb/IIIa
Antibody against receptors	Graves disease	Thyroid-stimulating hormone receptor (agonistic antibodies)
	Myasthenia gravis	Acetylcholine receptor (antagonistic antibodies)
Antibody against matrix antigens	Goodpasture syndrome	Basement membrane collagen (α_3-chain of type IV collagen)
	Pemphigus vulgaris	Epidermal cadherin (desmoglein)
TYPE III		
Immune complex diseases	Mixed essential cryoglobulinemia	Rheumatoid factor IgG complexes (with or without hepatitis C antigens)
	Systemic lupus erythematosus	DNA, histones, ribosomes, binuclear proteins
TYPE IV		
T-cell-mediated diseases	Insulin-dependent diabetes mellitus	Pancreatic B-cell antigen
	Rheumatoid arthritis	Unknown synovial joint antigen
	Multiple sclerosis	Myelin basic protein, proteolipid protein

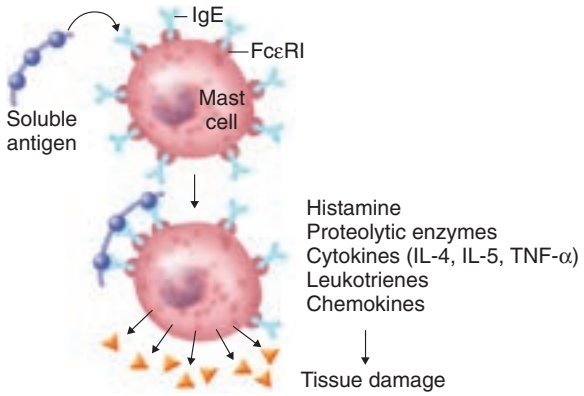

FIGURE 47-1. Type I hypersensitivity. Type I responses are mediated by immunoglobulin E (IgE), which induces mast cell activation. Cross-linking of the Fc receptor for IgE (FcεR) on mast cells, triggered by the interaction of multivalent antigen with antigen-specific IgE bound to FcεR, causes the release of preformed granules containing histamine and proteases. Cytokines, chemokines, and lipid mediators are synthesized after cell activation. IL = interleukin; TNF = tumor necrosis factor.

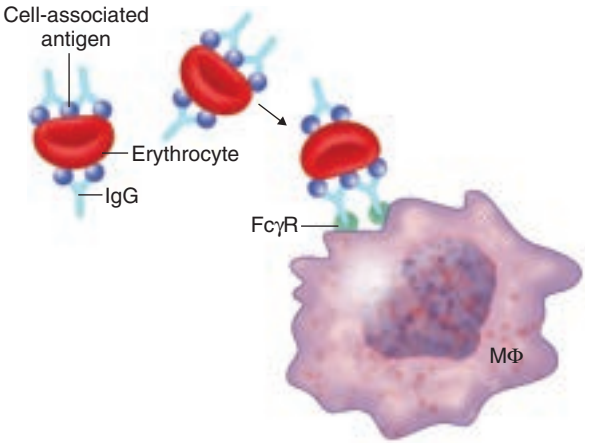

FIGURE 47-2. Type II hypersensitivity. Type II responses are mediated by immunoglobulin G (IgG) directed against cell surface or matrix antigens, which initiates effector responses through the Fc receptor for IgG (FcγR) and complement. The relative contributions of these pathways vary with the IgG subclass and the nature of the antigen. Only FcγR-mediated phagocytosis by macrophages (MΦ) is depicted in this figure. Activation of complement components would result in binding of C3b to the red blood cell membrane, rendering red blood cells susceptible to phagocytosis and leading to formation of the membrane attack complex and cell lysis.

causes mast cell activation. Proteolytic enzymes and toxic mediators, such as histamine, are released immediately from preformed granules, and chemokines, cytokines, and leukotrienes are synthesized after activation. Together, these mediators increase vascular permeability, break down tissue matrix proteins, promote eosinophil production and activation (IL-3, IL-5, and granulocyte-macrophage colony-stimulating factor [GM-CSF]), and cause influx of effector leukocytes (tumor necrosis factor-α [TNF-α], platelet-activating factor, and macrophage inflammatory protein [MIP-1]), constriction of smooth muscle, stimulation of mucus secretion, and amplification of T_H2 cell responses (IL-4 and IL-13). Eosinophils and basophils, activated through cell surface FcεRs, rapidly release highly toxic granular proteins (major basic protein, eosinophil peroxidase, and collagenase) and, over a longer period, produce cytokines (IL-3, IL-5, and GM-CSF), chemokines (IL-8), prostaglandins, and leukotrienes that activate epithelial cells, leukocytes, and eosinophils to augment local inflammation and tissue damage.

FcεR-bearing effectors act in a coordinated fashion. The immediate allergic inflammatory reaction initiated by mast cell products is followed by a late-phase response that involves recruitment and activation of eosinophils, basophils, and T_H2 lymphocytes.[5] The manifestations of IgE-mediated reactions depend on the site of mast cell activation. Mast cells reside in vascular and epithelial tissue throughout the body. In a sensitized host (an individual with IgE responses to antigens), re-exposure to antigen leads to type I hypersensitivity responses only in the mast cells exposed to the antigen. Inhalation of antigens produces bronchoconstriction and increased mucus secretion (asthma and allergic rhinitis); ingestion of antigens causes increased peristalsis and secretion (diarrhea and vomiting); and the presence of subcutaneous antigens initiates increased vascular permeability and swelling (urticaria and angioedema). Blood-borne antigens cause systemic mast cell activation, increased capillary permeability, hypotension, tissue swelling, and smooth muscle contraction—the characteristics of systemic anaphylaxis.

Type II Hypersensitivity Reactions

Type II hypersensitivity reactions (Fig. 47-2) are caused by chemical modification of cell surface or matrix-associated antigens that generates "foreign" epitopes to which the immune system is not tolerant. B cells respond to this antigenic challenge by producing IgG, which binds to these modified cells and renders them susceptible to destruction through complement activation, phagocytosis, and antibody-dependent cytotoxicity.

This phenomenon is seen clinically when drugs interact with blood constituents and alter their cellular antigens. Hemolytic anemia caused by immune-mediated destruction of erythrocytes (Chapter 160) and thrombocytopenia caused by destruction of platelets (Chapter 172), both type II hypersensitivity reactions, are adverse effects of certain drugs. Chemically reactive drug molecules bind covalently to the surface of red cells or platelets creating new epitopes that in a small subset of individuals are recognized as foreign antigens by the immune system and stimulate production of IgM and IgG antibodies reactive with the conjugate of drug and cell surface protein. Penicillin-specific IgG binds to penicillin-modified proteins on red blood cells and triggers activation of the complement cascade. Activation of complement components C1 through C3 results in covalent binding of C3b to the red cell membrane and renders circulating red cells susceptible to phagocytosis by FcγR and complement receptor–bearing macrophages in the spleen or liver. Activation of complement components C1 through C9 and formation of the membrane attack complex cause intravascular lysis of red cells. The factors that predispose only some people to drug-induced type II hypersensitivity reactions are unknown. Penicillin, quinidine, and methyldopa have been associated with hemolytic anemia and thrombocytopenia through this mechanism. Another example is heparin-induced thrombocytopenia or thrombosis, a severe, life-threatening complication that occurs in 1 to 3% of patients exposed to heparin (Chapter 172). Interactions among heparin, human platelet factor 4, antibodies to the human platelet factor 4–heparin complex, platelet FcγRIIA, and splenic FcγRs (which remove opsonized platelets) are involved in the pathogenesis of this disease.

Autoantibodies directed at antigens on the cell surface or extracellular matrix cause tissue damage by mechanisms similar to type II hypersensitivity reactions. IgG or IgM antibodies against erythrocytes lead to cell destruction in autoimmune hemolytic anemia because opsonized cells (coated with IgG or IgM and complement) are removed from the circulation by phagocytes in the liver and spleen or are lysed by formation of the membrane attack complex. Platelet destruction in autoimmune thrombocytopenic purpura occurs through a similar process. Because nucleated cells express membrane-bound complement regulatory proteins, they are less sensitive to lysis through the membrane attack complex, but when coated with antibody, they become targets for phagocytosis or antibody-dependent cytotoxicity. This mechanism is responsible for autoimmune and alloimmune neutropenia (Chapter 167).

IgM and IgG antibodies recognizing antigens within tissue or binding to extracellular antigens cause local inflammatory damage through FcγR and complement mechanisms. Pemphigus vulgaris (Chapter 439) is a serious blistering disease that results from a loss of adhesion between keratinocytes caused by autoantibodies against the extracellular portions of desmoglein 3, an intercellular adhesion structure of epidermal keratinocytes. Another example of a type II hypersensitivity reaction is Goodpasture disease (Chapter 121), in which antibodies against the α_3-chain of type IV collagen (the collagen in basement membranes) are deposited in glomerular and lung basement membrane. Tissue-bound autoantibodies activate monocytes, neutrophils, and basophils through FcγRs, initiating release of proteases, reactive oxidants, cytokines, and prostaglandins. Local activation of complement, particularly C5a, recruits and activates inflammatory cells and amplifies tissue injury. Neighboring cells are lysed by assembly of the membrane attack complex or by FcγR-initiated, antibody-dependent cytotoxicity.

Autoantibodies against cell surface receptors produce disease by stimulating or blocking receptor function. In myasthenia gravis (Chapter 422), autoantibodies against the acetylcholine receptors on skeletal muscle cells bind the receptor and induce its internalization and degradation in lysosomes, reducing the efficiency of neuromuscular transmission and causing progressive muscle weakness. In contrast, Graves disease (Chapter 226) is

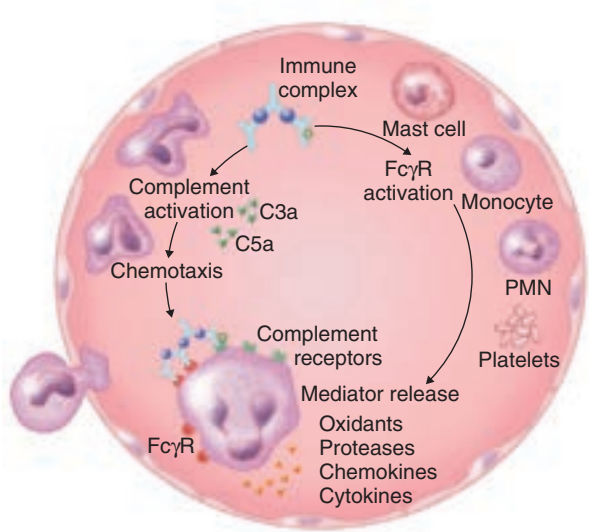

FIGURE 47-3. Type III hypersensitivity. Type III responses are mediated by immunoglobulin G (IgG) directed against soluble antigens. Localized deposition of immune complexes activates mast cells, monocytes, neutrophils, and platelets bearing the Fc receptor for IgG (FcγR), and initiates the complement cascade, all effectors of tissue damage. Generation of complement components C3a and C5a recruits and stimulates inflammatory cells and amplifies effector functions. PMN = polymorphonuclear leukocyte (also called *neutrophil*).

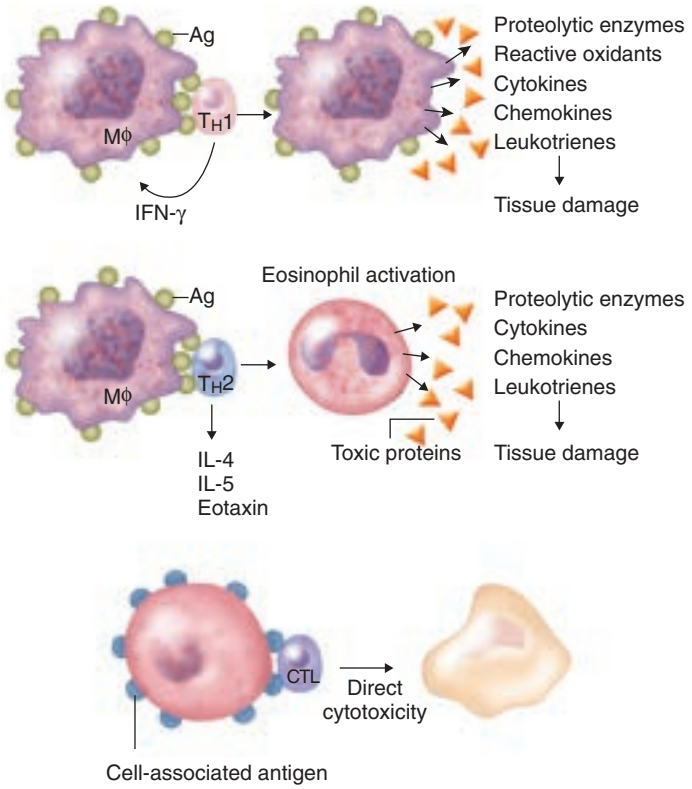

FIGURE 47-4. Type IV hypersensitivity. Type IV responses are mediated by T cells through three different pathways. In the first, type 1 helper T (T_H1) cells recognize soluble antigens (Ag) and release interferon-γ (IFN-γ) to activate effector cells, in this case macrophages (MΦ), and cause tissue injury. In T_H2-mediated responses, eosinophils predominate. T_H2 cells produce cytokines to recruit and activate eosinophils, leading to their degranulation and tissue injury. In the third pathway, damage is caused directly by cytolytic T lymphocytes (CTL). IL = interleukin.

characterized by autoantibodies that act as agonists. Autoantibodies to thyroid-stimulating hormone receptors bind the receptor, mimicking the natural ligand, inducing thyroid hormone overproduction, disrupting feedback regulation, and causing hyperthyroidism.

Type III Hypersensitivity Reactions

Type III hypersensitivity reactions (Fig. 47-3) are caused by tissue deposition of small soluble immune complexes that contain antigens and high-affinity IgG antibodies directed at these antigens. Localized deposition of immune complexes activates FcγR-bearing mast cells and phagocytes and initiates the complement cascade, all effectors of tissue damage.[6]

Immune complexes are generated in all antibody responses. The formation and the fate of immune complexes depend on the biophysical and immunologic properties of the antigen and the antibody. These properties include the size, net charge, and valence of the antigen; the class and subclass of the antibody; the affinity of the antibody-antigen interaction; the net charge and concentration of antibody; the molar ratio of available antigen and antibody; and the ability of the immune complex to interact with the proteins of the complement system. The lattice size of the immune complex is influenced strongly by the physical size and valence of the antigen, the association constant of antibody for that antigen, the molar ratio of antigen and antibody, and the absolute concentrations of the reactants. Larger aggregates fix complement more efficiently, present a broader multivalent array of ligands for complement and FcγRs to bind, and are taken up more readily by mononuclear phagocytes in the liver and spleen and thereby removed from the circulation. Smaller immune complexes, which form in antigen excess—as occurs early in an immune response—circulate in the blood and are deposited in blood vessels, where they initiate inflammatory reactions and tissue damage through interactions with FcγRs and complement receptors.

Serum sickness is a systemic type III hypersensitivity reaction, historically described in patients injected with therapeutic horse antiserum for the treatment of bacterial infections. In general, serum sickness occurs after the injection of large quantities of a soluble antigen. Clinical features include chills, fever, rash, urticaria, arthritis, and glomerulonephritis. Disease manifestations become evident 7 to 10 days after exposure to the antigen, when antibodies are generated against the foreign protein and form immune complexes with these circulating antigens. Immune complexes are deposited in blood vessels, where they activate phagocytes and complement, producing widespread tissue injury and clinical symptoms. The effects are transient, however, and resolve after the antigen is cleared.

A syndrome similar to serum sickness occurs in chronic infections in which pathogens persist in the face of continued immune response. In subacute bacterial endocarditis (Chapter 76), antibody production continues but fails to eliminate the infecting microbes. As the pathogens multiply, generating new antigens, immune complexes form in the circulation and are deposited in small blood vessels, where they lead to inflammatory damage of skin, kidney, and nerve. Hepatitis B virus infection (Chapters 148 and 149) may be associated with immune complex deposition early in its course, during a period of antigen excess, because antibody production in response to hepatitis B surface antigen is as yet relatively insufficient; some anicteric patients may present with acute arthritis. Mixed essential cryoglobulinemia, which may be associated with hepatitis C viral infection, is an immune complex–mediated vasculitis in which deposition of complexes containing IgG, IgM, and hepatitis C antigens causes inflammation in peripheral nerves, kidneys, and skin. Serum sickness also can develop in transplant recipients who are treated with mouse monoclonal antibodies specific for human T cells to prevent rejection, and in patients with myocardial infarction who are treated with the bacterial enzyme streptokinase to effect thrombolysis.

Systemic lupus erythematosus (Chapter 266), the prototypical immune complex–mediated autoimmune disease, is characterized by circulating IgG directed against common cellular constituents, typically DNA and DNA-binding proteins. Small immune complexes are deposited in skin, joints, and glomeruli and initiate local tissue damage.

Type IV Hypersensitivity Reactions

Type IV hypersensitivity reactions (Fig. 47-4), also known as *delayed-type hypersensitivity reactions,* are mediated by antigen-specific effector T cells. They are distinguished from other hypersensitivity reactions by the lag time from exposure to the antigen until the response is evident (1 to 3 days). Antigen is taken up, processed, and presented by macrophages or dendritic cells. T_H1 effector cells that recognize the specific antigen (these are scarce and take time to arrive) are stimulated to release chemokines, which recruit macrophages to the site, release cytokines that mediate tissue injury and growth factors that stimulate monocyte production. IFN-γ activates macrophages and enhances their release of inflammatory mediators, whereas TNF-α and TNF-β activate endothelial cells, enhance vascular permeability, and damage local tissue. The prototypical type IV hypersensitivity reaction

is the tuberculin test, but similar reactions can occur after contact with sensitizing antigens (e.g., poison ivy, certain metals) and lead to epidermal reactions characterized by erythema, cellular infiltration, and vesicles. $CD8^+$ T cells also may mediate damage by direct toxicity.

In contrast to T_H1-mediated hypersensitivity reactions, in which the effectors are macrophages, eosinophils predominate in T_H2-mediated responses. T_H2 effector T cells are associated with tissue damage in chronic asthma (Chapter 87). T_H2 cells produce cytokines to recruit and activate eosinophils (IL-5 and eotaxin), leading to degranulation, further tissue injury, and chronic, irreversible airway damage.

Additional T_H effector cells, such as T_H17 cells, mediate tissue damage. T_H17 cells produce IL-17 family cytokines, as well as IL-21, IL-22, and GM-CSF, that regulate innate effectors and orchestrate local inflammation by inducing release of proinflammatory cytokines and chemokines, proliferation and activation of effector cells and other target cells, recruitment of neutrophils, and enhanced T_H2-mediated inflammation, all of which amplify allergic and autoimmune responses.[3,7] T_H17 cells have been implicated in allergic disorders (atopic dermatitis, asthma) and autoimmune and inflammatory diseases (psoriasis, inflammatory bowel disease, rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis).

In some autoimmune diseases, effector T cells specifically recognize self-antigens to cause tissue damage, either by direct cytotoxicity or by inflammatory responses mediated by activated macrophages. In type 1 insulin-dependent diabetes mellitus, T cells mediate destruction of β cells of the pancreatic islets. IFN-γ-producing T cells specific for myelin basic proteins have been implicated in multiple sclerosis. Rheumatoid arthritis is another autoimmune disease caused, at least in part, by activated T_H1 cells.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hogarth PM, Anania JC, Wines BD. The FcγR of humans and non-human primates and their interaction with IgG: implications for induction of inflammation, resistance to infection and the use of therapeutic monoclonal antibodies. *Curr Top Microbiol Immunol*. 2014;382:321-352.
2. Salazar F, Ghaemmaghami AM. Allergen recognition by innate immune cells: critical role of dendritic and epithelial cells. *Front Immunol*. 2013;4:356.
3. Singh RP, Hasan S, Sharma S, et al. Th17 cells in inflammation and autoimmunity. *Autoimmun Rev*. 2014;13:1174-1181.
4. Shah A. The pathologic and clinical intersection of atopic and autoimmune disease. *Curr Allergy Asthma Rep*. 2012;12:520-529.
5. Voehringer D. Protective and pathological roles of mast cells and basophils. *Nat Rev Immunol*. 2013;13:362-375.
6. Karsten CM, Kohl J. The immunoglobulin, IgG Fc receptor and complement triangle in autoimmune diseases. *Immunobiology*. 2012;217:1067-1079.
7. Maddur MS, Miossec P, Kaveri SV, et al. Th17 cells: biology, pathogenesis of autoimmune and inflammatory diseases, and therapeutic strategies. *Am J Pathol*. 2012;181:8-18.

REVIEW QUESTIONS

1. Type III hypersensitivity reactions are caused by tissue deposition of immune complexes which activate local effectors and lead to injury. Which of the following mediators are NOT characteristic of this pathway of inflammation?

A. Fc receptors for IgE
B. Complement
C. Neutrophils
D. Mononuclear phagocytes

Answer: A Type III responses are mediated by IgG directed against soluble antigens. Localized deposition of immune complexes activates mast cells, monocytes, neutrophils, and platelets bearing the Fc receptor for IgG (FcγR) and initiates the complement cascade, all effectors of tissue damage. Generation of complement components C3a and C5a recruits and stimulates inflammatory cells and amplifies effector functions. IgE does not participate in the inflammatory response. Of note, mast cells may be activated in type III responses through receptors for IgG or complement.

2. Type IV hypersensitivity reactions, as exemplified by the tuberculin test, does NOT require which of the following elements?

A. IgG
B. Macrophages
C. T_H1 effector cells
D. IFN-γ

Answer: A The prototypic type IV hypersensitivity reaction is the tuberculin test, in which antigen is taken up, processed, and presented by macrophages. Type 1 helper T (T_H1) effector cells that recognize the specific antigen are stimulated to release chemokines, which recruit macrophages to the site, and release cytokines including IFN-γ, which activate macrophages and enhance their release of inflammatory mediators. IgG does not participate in type IV hypersensitivity reactions.

3. In which of the following clinical situations are autoantibodies NOT critical triggers of tissue damage?

A. Asthma
B. Pemphigus vulgaris
C. Graves disease
D. Systemic lupus erythematosus

Answer: A IgE antibodies against innocuous environmental antigens initiate asthma, not autoantibodies. Autoantibodies against TSH receptors act as agonists in Graves disease. Autoantibodies against desmoglein-3 effect adhesion between epidermal keratinocytes and cause blister formation. In systemic lupus erythematosus, inflammation is initiated by deposition of immune complexes containing autoantibodies directed against common cellular constituents (e.g., DNA and ribonucleolar proteins).

4. A syndrome similar to serum sickness occurs in which of the following infections?

A. Hepatitis B
B. Tuberculosis
C. Pneumococcal pneumonia
D. Influenza

Answer: A A syndrome similar to serum sickness occurs in chronic infections in which pathogens persist in the face of continued immune response. Hepatitis B virus infection may be associated with immune complex deposition early in its course, during a period of antigen excess, because antibody production in response to hepatitis B surface antigen is as yet relatively insufficient. Efficiency of clearance of immune complexes depends on their size and valence. Smaller immune complexes, which form in antigen excess—as occurs early in an immune response—circulate in the blood and are deposited in blood vessels, where they initiate inflammatory reactions and tissue damage through interactions with FcγRs and complement receptors. In influenza and pneumococcal pneumonia, there is no clinical evidence of systemic immune complex deposition. Tuberculosis generates T-cell responses.

5. The effector mechanisms that lead to tissue damage in autoimmune diseases are similar to those elicited in response to environmental antigens that result in allergy. Which hypersensitivity reaction is correctly matched to an autoimmune condition that occurs through a similar mechanism?

A. Systemic lupus erythematosus and serum sickness
B. Idiopathic thrombocytopenia purpura and asthma
C. Multiple sclerosis and heparin-induced thrombocytopenia
D. Myasthenia gravis and atopic dermatitis

Answer: A Systemic lupus erythematosus, the prototypical immune complex–mediated autoimmune disease, is characterized by circulating IgG directed against common cellular constituents, typically DNA and DNA-binding proteins. Like in the case of serum sickness, small immune complexes are deposited in skin, joints, and glomeruli and initiate local tissue damage. Idiopathic thrombocytopenia purpura, like type II hypersensitivity reactions, is mediated by autoantibodies that are directed to platelet surface antigens, whereas asthma is a triggered by IgE against innocuous environmental antigens (type I hypersensitivity reaction). T cells are key effectors in multiple sclerosis, whereas heparin-induced thrombocytopenia is caused by IgG autoantibodies. Myasthenia gravis is caused by IgG autoantibodies that recognize acetylcholine receptors and impair neuromuscular signaling, whereas atopic dermatitis is mediated by T cells.

48

MECHANISMS OF INFLAMMATION AND TISSUE REPAIR

GARY S. FIRESTEIN

Host defense mechanisms have evolved to recognize pathogens rapidly, render them harmless, and repair the damaged tissue. This complex and highly regulated sequence of events can also be triggered by environmental stimuli such as noxious mechanical and chemical agents. Under normal circumstances, tightly controlled responses protect against further injury and clear damaged tissue. In disease states, however, pathologic inflammation can lead to marked destruction of the extracellular matrix (ECM) and organ dysfunction.

INITIATION OF THE INFLAMMATORY RESPONSE

When normal tissue encounters a pathogen, resident cells are stimulated by engagement of pattern recognition receptors that activate an ancient arm of host defense known as *innate immunity*. In contrast to *adaptive immunity*, which provides exquisite antigen specificity, innate immune responses recognize common motifs on pathogens (Chapter 45). Additional cytoplasmic receptors can sense "danger" signals from a toxic environment or cellular stress, such as urate or adenosine triphosphate (ATP). Innate mechanisms are designed for rapid responses (minutes to hours) compared with the more leisurely adaptive system that can take days to weeks to develop. In addition to orchestrating early events that are critical to host defense, cells of the innate system like dendritic cells orchestrate the subsequent adaptive cascade through the generation of chemokines that organize lymphoid tissue and presentation of antigens to lymphocytes. Innate immunity provides intergenerational continuity in that the receptors are encoded in the germline and are passed unchanged to progeny to protect the species. In contrast, each individual must generate his or her own adaptive immune system through complex somatic mutations and gene rearrangements. This provides defense tailored for each member of the species; its complexity and beauty permit specificity but also provide opportunities for error such as responses against self-antigens in autoimmunity.

Pathogen-Associated Molecular Pattern Recognition

The toll-like receptor (TLR) family of proteins binds common patterns of molecular structures on microbial pathogens that normally are not found in mammalian cells. The TLRs are critical members of the innate immune system and serve as sentinels that initiate a rapid response.[1] Some are expressed on the cell surface, such as TLR2, which is activated primarily by bacterial peptidoglycan and lipoproteins, and TLR4, which is activated by lipopolysaccharide (LPS, or endotoxin). Others are expressed mainly on the inner leaflet of cytoplasmic vesicles, like TLR9, which is activated by unmethylated bacterial sequences that are enriched for CpG motifs (regions of DNA where cytosine and guanine nucleotides in the linear sequence of bases along its length are separated by one phosphate), or TLR3 and TLR7, which are important for antiviral defense because they bind double-stranded and single-stranded viral RNA, respectively. In addition to exogenous molecules, some endogenous structures can bind to TLRs, including heat shock proteins and oxidized low-density lipoproteins (oxLDLs). The latter might be especially important in the pathogenesis of atherosclerosis, in which LDL activates TLR4 within vascular plaques. Local endothelial cell– and macrophage-derived chemotactic factors can then recruit activated T cells into the atheroma.

Signaling by TLR2 and TLR4 progresses through adaptor proteins and often converges on a kinase known as MyD88, which orchestrates several downstream cascades. By directing the phosphorylation of IκB kinase-β (IKKβ), MyD88 activates nuclear factor-κB (NF-κB), a master switch for inflammatory genes.[2] Translocation of NF-κB to the cell nucleus stimulates the production of cytokines (e.g., interleukin-6 [IL-6], IL-8, and tumor necrosis factor [TNF]), the machinery for prostaglandin release (e.g., cyclooxygenase 2 [COX2]), and genes that regulate the ECM (e.g., metalloproteinases). This rapid response is normally transient, although it can persist in pathogenic states. MyD88-independent pathways that stimulate innate immunity also exist. For instance, TLR3 stimulation by RNA viruses uses a separate pathway involving IKKε and interferon regulating factor-3 (IRF-3). IRF-3, in combination with several other transcription factors, induces the expression of genes such as interferon-β (IFN-β) to establish an antiviral state.

These genes primarily offer protection against pathogens by initiating key defense mechanisms. However, these same pathways can create a hazardous milieu that is toxic to normal cells through the production of oxygen radicals, nitric oxide, and other reactive intermediaries. These molecules can damage DNA and harm bystander cells, or even lead to neoplasia (E-Table 48-1). For instance, long-standing inflammation in the colon, as in ulcerative colitis, is associated with adenocarcinoma. Increased COX2 expression as a result of NF-κB translocation is another mechanism that contributes to the development of tumors at inflammatory sites. An unanticipated finding is that NF-κB itself can also directly augment carcinogenesis by serving as a survival signal for damaged cells that would normally be deleted by apoptosis.

The TLR signal transduction mechanisms integrate the environmental stimuli and generate a broadly antipathogen response. Fine-tuning of host defenses against unique pathogen structures to provide long-lived immunity requires the slower, more precise adaptive immune system. Although it is more cumbersome and primitive, innate immunity provides signals that activate adaptive responses. For instance, TLRs can direct dendritic cells (Chapter 45), which have internalized and processed antigen, to migrate from peripheral tissues to central lymphoid organs. The dendritic cells can also produce cytokines and, after maturation, present antigens to T cells in the context of class II major histocompatibility molecules and surface costimulatory proteins. The activated T cells can then migrate to the tissue to enhance and amplify the host response. T cells also provide help to B cells, thereby stimulating antibody production and activating other components of innate immunity (e.g., the complement system, Chapter 50).

Other non-TLR cytoplasmic sensors also serve a similar purpose in the environment. For instance, retinoic acid–inducible gene 1 (RIG-1) and melanoma differentiation–associated gene 5 (MDA5) can detect RNA viruses and initiate an inflammatory response. These are, in some cases, partially redundant with TLR3 and TLR7 and can activate similar signaling mechanisms, such as NF-κB through the IKKß and IRFs through a distinct pathway involving IKKε and TBK1.

E-TABLE 48-1 EXAMPLES OF INFLAMMATION PATHWAYS IN DISEASE

DISEASE	ACTIVATED PATHWAYS
Atherosclerosis	Toll-like receptor activation (e.g., oxLDL) Chemokine-mediated leukocyte recruitment (e.g., MCP-1)
Cancer	Reactive oxygen and nitrogen intermediate-induced mutations Cyclooxygenase 2–mediated neoplasia (e.g., colon, breast) NF-κB prolonging survival of damaged cells
Asthma	IgE-mediated mast cell activation T_H2 cytokine-mediated leukocyte activation Leukotriene-induced bronchospasm Protease-induced airway remodeling
Rheumatoid arthritis	Toll-like receptor activation (e.g., peptidoglycan) Macrophage/fibroblast cytokine production, including IL-1, TNF, and IL-6 Cyclooxygenase 2 induction Protease-mediated cartilage destruction Synovial complement activation
Systemic lupus erythematosus	Complement activation in multiple organs α-Interferon production and interferon signature
Autoinflammatory diseases, including psoriasis	Inflammasome activation, including production of IL-1, IL-18, and IL-33 T_H17 cell activation IL-17A-, IL-12-, and IL-23-mediated inflammation

IgE = immunoglobulin E; MCP-1 = monocyte chemoattractant protein 1; NF-κB = nuclear factor-κB; oxLDL = oxidized low-density lipoprotein; T_H17 = helper T lymphocyte type 17; T_H2 = helper T lymphocyte type 2; TNF = tumor necrosis factor.

Environmental Stress and Danger-Associated Molecular Patterns

Danger-associated molecular pattern molecules serve as a mechanism to detect and respond to damage to the microenvironment. Tissue injury due to direct trauma or a noxious stimulus initiates an inflammatory response and is associated with microvascular damage, extravasation of leukocytes through vascular walls, and leakage of plasma and proteins into the tissue. Endogenous proteins, including ATP receptors, S100, heat shock proteins, and high mobility-group box 1 (HMGB1), mediate release of molecules that reflect cellular toxicity and induce a cellular response. Acid-sensitive ion channels (ASICs) on the cell surface can also detect the environmental stress caused by a decrease in tissue pH. ASICs can mediate a variety of cellular functions, including cell death through apoptosis or pain responses that can lead to adaptive pain behaviors that limit further exposure to noxious stimuli.

Proteases, Coagulation, and Inflammation

Although the coagulation system's primary function is to maintain vascular integrity (Chapter 171), the proteases that regulate its functions also play an important role in the early responses to tissue damage and inflammation. For example, plasminogen is a circulating proenzyme that can be cleaved to plasmin by enzymes in the coagulation pathway, including factors XIa and XIIa. Tissue plasminogen activating factor and kallikrein also have this capacity. When activated, the serine protease plasmin can digest fibrin, fibronectin, thrombospondin, and laminin as well as activating pro-matrix metalloproteinases like collagenase (MMP1). By remodeling the extracellular matrix, this system can ultimately regulate cell recruitment and tissue damage.

Thrombus formation at the site of vascular damage can begin the inflammatory cascade through the release of vasoactive amines (e.g., serotonin), release of lysosomal proteases, and formation of eicosanoid products. The platelets can also later regulate healing with release of growth factors such as platelet-derived growth factor (PDGF) and transforming growth factor-β (TGFβ).

Inflammasome

The inflammasome[3] is among the best characterized mechanisms for sensing danger and includes the 22-member human Nod-like receptor (NLR) family of cytoplasmic proteins. The activated NLR proteins recruit additional proteins to form a complex with caspase-1 and adaptor molecule apoptosis-associated specklike protein (ASC). Activation of caspase-1 is a key function of inflammasomes, with resultant cleavage and activation of IL-1, IL-18, and IL-33. The latter molecule is also known as an "alarmin" because of its rapid release in the presence of tissue damage or a pathogen. Alarmins are often preformed in cells, such as mast cells, and can be either released directly into the microenvironment or quickly processed and secreted. Other alarmins include products of cell destruction, such as ATP or uric acid.

Disorders of the inflammasome are associated with a group of conditions known as autoinflammatory diseases (Chapter 261). The prototypic syndromes known as familial cold autoinflammatory disease, Muckle-Wells disease, and neonatal-onset multisystem inflammatory disease (NOMID) are due to nonconserved mutations in the *NLR* gene that encodes cryopyrin (also known as *NALP3*). These rare diseases are characterized by abnormal inflammasome activation with aberrant release of processed IL-1β. The clinical manifestations, including fever, rash, hearing impairment, and arthritis, depend on the specific amino acid substitution as well as other less well-defined genetic influences. The critical role of IL-1 has been proved by studies using treatment with IL-1 inhibitors, which prevent flares and can reverse end-organ damage. The inflammasome also participates in some common diseases, such as gout (Chapter 273), in which urate crystals can activate the inflammasome.

Immune Complexes and Complement

The complement system (Chapter 50) is another ancient defense mechanism that links innate immunity and the humoral arm of adaptive immunity. Both the classical complement pathway, activated by immunoglobulin G (IgG)- and IgM-containing immune complexes, and the alternative pathway, activated by bacterial products, converge at the third component of complement, C3, with proteolytic release of fragments that amplify the inflammatory response and mediate tissue injury. C3a and C5a directly increase vascular permeability and contraction of smooth muscle. C5a induces mast cell release of histamine, thereby indirectly mediating increased vascular permeability. C5a also activates leukocytes and enhances their chemotaxis, adhesion, and degranulation, with release of proteases and toxic metabolites. C5b attaches to the surface of cells and microorganisms and is the first component in the assembly of the C5b-9 membrane attack complex.

Individuals with abnormalities of the early complement components, especially C1q, C2, and C4, usually have a minimally increased incidence of infection but demonstrate an enhanced risk for developing autoimmune diseases such as systemic lupus erythematosus (SLE) (Chapter 266). The mechanism of increased disease susceptibility is probably related to inefficient clearance of immune complexes. Enhanced activation and consumption of complement proteins can also occur in SLE accompanied by low plasma C3 and C4 levels, especially in association with disease exacerbations. C3 or C5 deficiency increases susceptibility to bacterial infections, whereas defects in the late components that form the membrane attack complex result in an increased incidence of *Neisseria* sp bacteremia (Chapter 298).

SECOND WAVE OF THE INFLAMMATORY RESPONSE

Activation of innate immunity quickly leads to the robust influx of inflammatory cells. Resident cells, such as vascular endothelial cells, mast cells, dendritic cells, and interstitial fibroblasts, respond by releasing soluble mediators, including eicosanoids and pro-inflammatory cytokines (E-Table 48-2). These mediators amplify the inflammatory response and recruit additional leukocytes. Locally stimulated cells, along with the newly arrived inflammatory cells, release toxic reactive intermediates of nitrogen and oxygen as well as a myriad of proteases, principally matrix metalloproteinases (MMPs), serine proteases, and cysteine proteases. These molecules help destroy infectious agents and remove damaged cells, thus clearing the injured site for tissue repair. In most situations, the normal physiologic response is an exquisitely coordinated program that uses proteolytic enzymes to remodel the ECM and promote a supportive environment for wound healing rather than tissue damage.

Cellular Response

Inflammatory cell infiltration at the site of initial tissue damage typically begins with release of chemokines and soluble mediators from resident cells, including interstitial fibroblasts, mast cells, and vascular endothelial cells. Signaling from these events alters the local adhesion molecule profile and creates a chemotactic gradient that recruits cells from the blood stream. Mast cells, in particular, act as sentinels that degranulate within seconds after ligation of immunoreceptors and activation of the signaling molecule spleen tyrosine kinase (Syk) to release vasoactive amines. In most acute responses, polymorphonuclear leukocytes (PMNs) are the first inflammatory cells to arrive at the site of injury, followed later by mononuclear cells.

Most tissue fibroblasts and vascular endothelial cells are generally quiescent before migration of PMNs into the tissue. However, these resident cells can be triggered to proliferate and migrate toward the site of injury as well as to synthesize cytokines, proteases, and ECM components. Growth factors are released, such as basic fibroblast growth factor (bFGF) and vascular endothelial growth factor (VEGF), stimulating new blood vessel formation. Together with granulocyte-macrophage colony-stimulating factor (GM-CSF), these locally released growth factors contribute to cellular proliferation and amplification of the inflammatory response and also induce maturation of dendritic cells that process antigens. In addition, fibroblasts and endothelial cells secrete new ECM proteins, MMPs, and other ECM-digesting enzymes. Initially, the response favors proteolytic activity to clear damaged infrastructure. This is followed by a shift to increased production of new ECM to allow tissue repair and wound healing.

Increased vascular permeability, caused by disruption of endothelial cell tight junctions, allows blood-borne proteins such as fibrinogen, fibronectin, and vitronectin to extravasate into the perivascular ECM. Interaction with preexisting ECM allows the assembly of new ligands for a subset of adhesion molecules (e.g., integrins $\alpha_5\beta_1$ and $\alpha_v\beta_3$). This increased vascular permeability and change in the profiles of adhesion molecules and ligands, in conjunction with release of chemoattractant molecules, leads to the recruitment of leukocytes to sites of inflammation. Some of the chemokines involved are IL-8 (for neutrophils), macrophage chemoattractant protein-1 (MCP-1) for monocytes, RANTES (regulated on activation, T-cell expressed and secreted) for monocytes and eosinophils, and IL-16 (for $CD4^+$ T cells).

Chemokines have the capacity to recruit specific subsets of cells by binding to G protein–coupled chemokine receptors. Directly targeting chemokines, either with biologics or with small molecules, has met with limited success in clinical trials, perhaps because the system is quite complex and highly

E-TABLE 48-2 SIGNALS FOR INDUCTION AND REPAIR OF INFLAMMATION

INFLAMMATION	RESOLUTION AND TISSUE REPAIR
CYTOKINES AND GROWTH FACTORS	
TNF	TGF-β
IL-1 family (IL-1, IL-18, IL-33)	IL-10
IL-6 family (IL-6, IL-11, LIF, osteopontin)	FGF
IL-4, IL-13	Osteoprotegerin
IL-15	IL-1RII
IL-17 family (IL-17A-F)	IL-1Ra
IL-12 family (IL-12, IL-23, IL-27)	Soluble TNF-R
VEGF	IL-18 binding protein
Chemokines HMBG1	
PROTEASES	
Matrix metalloproteinases Collagenases Gelatinases Stromelysins Matrilysins	TIMPs
Serine proteases Trypsin Chymotrypsin Kallikrein Plasmin	SERPINs, α_2-macroglobulin
Cysteine proteases	
ADAMTS family Aggrecanases	
SMALL MOLECULE MEDIATORS	
Prostaglandins (especially PGE_2)	Lipoxins Cyclopentenone Antioxidants
Leukotrienes (especially LTB4)	
C3a and C5a	
Histamine	
Bradykinin	
Reactive oxygen	
Reactive nitrogen	
APOPTOSIS REGULATORS	
Soluble Fas ligand	Fas TRAIL Reactive oxygen Reactive nitrogen

ADAMTS = a disintegrin and metalloproteinase family; FGF = fibroblast growth factor; IL = interleukin; LIF = leukemia inhibitory factor; R = receptor; Ra = receptor antagonist; SERPINs = serine protease inhibitors; TGF = transforming growth factor; TIMPs = tissue inhibitors of metalloproteinase; TNF = tumor necrosis factor; TRAIL, TNF-related apoptosis-inducing ligand; VEGF = vascular endothelial growth factor; HMBG1= high mobility-group box 1.

redundant. An alternative approach might be to target intracellular mechanisms distal to receptor ligation. Chemokine receptors generally signal through the phosphoinositide-3 kinase (PI3K) system, especially the gamma isoform. PI3Kγ is mainly expressed in bone marrow–derived cells and is the convergence point for multiple chemotactic factors. Preclinical studies suggest that blocking this pathway decreases inflammatory cell recruitment in models of lupus and rheumatoid arthritis.

The precise combination of chemokines and vascular adhesion molecules present in an inflammatory lesion determines the timing for recruitment of individual inflammatory cell types. Ligation of integrins on leukocytes also prolongs cell survival after they have moved into the tissue, by preventing apoptosis. The central role of certain specific adhesion molecule–ligand pairs has been confirmed in human diseases. For instance, $\alpha_4\beta_1$ plays a key role in the recruitment of lymphocytes to the central nervous system in multiple sclerosis, and blocking this interaction suppresses disease activity (Chapter 411). Eosinophils use the same adhesion receptors to migrate into the lung in allergen-induced asthma (Chapter 87).

Increased expression of intracellular adhesion molecule 1 (ICAM-1) and vascular cell adhesion molecule 1 (VCAM-1), as well as increased chemokine expression, is evident in other cell types, such as the airway epithelium after allergen challenge in asthma. Rapid transient influx of neutrophils occurs in allergic airway disease, along with activation of the local T cells and mast cells. These neutrophils produce lipid mediators, reactive oxygen intermediates, and proteases such as elastase, which may contribute to airflow obstruction, epithelial damage, and remodeling. Neutrophil elastase, together with chemokines released by both recruited and allergen-activated T cells and mast cells, serves to recruit eosinophils.

Soluble Mediators

PRO-INFLAMMATORY CYTOKINES

Pro-inflammatory cytokines, often derived from macrophages and fibroblasts, are mediators that activate the immune system. The pro-inflammatory members of the IL-1 family (e.g., IL-α, IL-1β, IL-18, and IL-33) and TNF have pleiotropic activities and can enhance adhesion molecule expression on endothelial cells, induce proliferation of endogenous cells, and stimulate antigen presentation. IL-1 and TNF also increase expression of matrix-degrading enzymes, such as collagenase and stromelysin. In addition, they stimulate synthesis of other inflammatory mediators such as prostaglandins from fibroblasts. TNF inhibitors (Chapter 36) are effective in inflammatory diseases such as psoriasis, rheumatoid arthritis, and inflammatory bowel disease, and IL-1 inhibitors (Chapter 36) are beneficial in genetic diseases such as Muckle-Wells syndrome and familial cold autoinflammatory syndrome.

IL-1 and TNF comprise only a small fraction of the acute cytokine response. Many other factors also participate, including IL-6 and its related cytokines (IL-11, osteopontin, and leukemia inhibitory factor), which can both induce acute phase reactants and bias an immune response toward a helper T type 1 (T_H1) or T_H2 phenotype (Chapter 47). GM-CSF can regulate dendritic cell maturation, increase expression of human leukocyte antigen (HLA-DR) on these cells, and enhance antigen presentation. The T_H1 lymphokine IFN-γ, although often considered part of the secondary wave that ensues after T-cell activation, can also induce expression of HLA-DR, increase expression of endothelial cell adhesion molecules, and inhibit collagen production. IL-1, IL-6, and IL-23 can coordinate differentiation toward T_H17 cells, a phenotype that is thought to play a major role in inflammation and autoimmunity owing to the production of IL-17 family members (IL-17A through F). Of these, IL-17A and perhaps IL-17F are especially important because they can synergize with IL-1 and TNF. The growth factor TGF-β biases cells toward the regulatory T cell (Treg) phenotype, which can suppress antigen-specific responses of other T cells (see later). The benefit of individual cytokine inhibitors varies depending on the disease. For instance, IL-6 blockade is effective in rheumatoid arthritis, whereas IL-12/23 and IL-17A inhibition suppresses skin inflammation in psoriasis. Clinical trials now clearly show that IL-17A antibodies are effective in psoriasis.

Many cytokines activate cells by ligating their receptors and engaging the Janus kinase (JAK) family of signaling molecules, including JAK1, JAK2, JAK3, and Tyk2. These kinases, in turn, phosphorylate the signal transducer and activator of transcription (STAT) proteins. The STATs serve as transcription factors that initiate expression of many other cytokines and mediators of the inflammation and amplify the response. JAK inhibition represents an alternative approach to abrogating the inflammatory response.

Cytokines play a key role in the establishment and perpetuation immune-mediated diseases. As noted earlier, autocrine and paracrine cytokine networks play a critical role in the perpetuation of inflammation in rheumatoid arthritis[4] (Chapter 264). MCP-1 recruits and activates macrophages into atheromas containing oxLDLs and foam cells. In allergic asthma (Chapter 87), IL-13 is emerging as a central inflammatory cytokine. IL-13 functions through binding to cell surface IL-4 receptors, and IL-4R–deficient mice are relatively resistant to the development of asthma.

EICOSANOIDS

In addition to cytokines and immune complexes, local inflammatory responses lead to the release of eicosanoids, which are lipid-derived molecules. Because lipids are present in the cell membrane, they are readily available substrates for the synthesis of mediators. These molecules are produced adjacent to sites of injury, and their half-lives range from seconds to minutes. Eicosanoids are not stored but are produced de novo from membrane lipids when cell activation by mechanical trauma, cytokines, growth factors, or other stimuli leads to release of arachidonic acid. Cytosolic phospholipase A_2 ($cPLA_2$) is the key enzyme in eicosanoid production. Cell-specific and agonist-dependent events coordinate the translocation of $cPLA_2$ to the nuclear envelope, endoplasmic reticulum, and Golgi apparatus, where interaction with COX (in the case of prostaglandin synthesis) or 5-lipoxygenase (in the case of leukotriene synthesis) can occur.

PROSTAGLANDINS

Prostanoids[5] are produced when arachidonic acid is released from the plasma membrane of injured cells by phospholipases and metabolized by cyclooxygenases and specific isomerases (Chapter 37). These molecules act both at peripheral sensory neurons and at central sites within the spinal cord and brain to evoke pain and hyperalgesia. Their production is increased in most acute inflammatory conditions, including arthritis and inflammatory bowel disease. In response to exogenous and endogenous pyrogens, prostaglandin E_2 (PGE_2) derived from COX2 mediates a central febrile response. In addition, prostaglandins synergize with bradykinin and histamine to enhance vascular permeability and edema. The levels of prostaglandins are usually very low in normal tissues and increase rapidly with acute inflammation, well before leukocyte recruitment. COX2 induction with inflammatory stimuli most likely accounts for the high levels of prostanoids in chronic inflammation.

COX2 also plays a key role in platelet–endothelial cell interactions by increasing the production of prostacyclin (PGI_2) in endothelial cells (Chapter 37). Increased risk for myocardial infarction associated with the use of selective COX2 inhibitors may be related to unopposed production of thromboxane A_2 by COX1 in platelets. Prostacyclin also protects against atherosclerosis in mice, and COX2 blockade abrogates this beneficial effect. Thus, COX inhibitors can potentially increase thrombotic events.

LEUKOTRIENES

A distinct set of enzymes direct arachidonic acid metabolites toward the synthesis of leukotrienes (Chapter 87). Their relative importance depends on the specific target organ of an inflammatory response. For instance, leukotriene receptor antagonists are effective in asthma, whereas similar approaches have been less impressive in rheumatoid arthritis. Unlike prostaglandins, leukotrienes are primarily produced by inflammatory cells such as neutrophils, macrophages, and mast cells. 5-Lipoxygenase is the key enzyme in this cascade, transforming released arachidonic acid to the epoxide leukotriene A_4 (LTA_4) in concert with 5-lipoxygenase-activating protein (FLAP). LTA_4 can be hydrolyzed by cytosolic LTA_4 hydrolase to LTB_4, a potent neutrophil chemoattractant and stimulator of leukocyte adhesion to endothelial cells. LTA_4 can also conjugate with glutathione to form LTC_4 by LTC_4 synthase at the nuclear envelope. LTC_4 can be metabolized extracellularly to LTD_4 and LTE_4. These three cysteinyl leukotrienes promote plasma leakage from postcapillary venules, upregulation of expression of cell surface adhesion molecules, and bronchoconstriction.

HISTAMINE

Histamine is a vasoactive amine produced by basophils and mast cells that markedly increases capillary leakage. In basophils, histamine is released in response to bacterial formylmethionyl-leucyl-phenylalanine (f-MLP) sequences, complement fragments C3a and C5a, and IgE. The resultant edema can be readily observed clinically in urticaria (Chapters 252 and 440) and allergic rhinitis (Chapter 251). The stimulus for release of histamine from

mast cell granules is the same as in basophils, except for the absence of f-MLP receptors in this cell type. Histamine can also synergize with locally produced LTB_4 and LTC_4. In addition, histamine enhances leukocyte rolling and firm adhesion, and induces gaps in the endothelial cell lining, enhancing leukocyte extravasation.

Despite the production of histamine in asthma and in acute synovitis, currently available histamine blockers have minimal therapeutic effect in these conditions. Targeting the more recently described histamine type 4 receptor (HR4), which has a variety of immunomodulatory effects on bone marrow–derived cells, suggests that more precise inhibition of this novel histamine pathway might have greater success.[6]

KININS

Kinins induce vasodilation, edema, and smooth muscle contraction, as well as pain and hyperalgesia, through stimulation of C fibers. They are formed from high- and low-molecular-weight kininogens by the action of serine protease kallikreins in plasma and peripheral tissues. The primary products of kininogen digestion are bradykinin and lysyl-bradykinin. These products have high affinity for the B2 receptor, which is widely expressed and is responsible for the most common effects of kinins. The peptides desArg-BK and Lys-desArg-BK are generated by carboxypeptidases and bind the kinin B1 receptor subtype, which is not expressed in normal tissues but is rapidly upregulated by TLR ligands and cytokines. The kinin B2 receptor is internalized rapidly and desensitized, whereas the B1 receptor remains highly responsive. Kinin actions are associated with the secondary production of other mediators of inflammation, including nitric oxide, mast cell–derived products, and the pro-inflammatory cytokines IL-6 and IL-8. In addition, kinins can increase IL-1 production through initial stimulation of TNF and can increase prostanoid production through activation of phospholipase A_2 and release of arachidonic acid.

NEURAL NETWORKS

Neural outflow also can rapidly activate inflammatory mechanisms and alter vascular permeability at sites of tissue damage. Pain receptors can activate type δ fibers and carry information to the spinal cord about noxious stimuli where cytokines like IL-1 or TNF are produced. Spinal cytokines lead to phosphorylation of signal molecules in the central nervous system like mitogen activated protein kinases (MAPKs). Reflex neural loops, including sympathetic and parasympathetic nerves, release mediators like substance P, acetylcholine, epinephrine or norepinephrine into the immediate location as well as surrounding tissue. Vascular permeability and activation of resident cells like macrophages can help recruit additional cells to the affected region.

MECHANISMS OF TISSUE DAMAGE IN INFLAMMATION

Reactive Oxygen and Nitrogen

Macrophages, neutrophils, and other phagocytic cells can generate large amounts of toxic reactive oxygen intermediates (ROIs) and reactive nitrogen intermediates (RNIs) that can directly kill pathogens. ROIs and RNIs also serve as critical signal transduction molecules that regulate expression of inflammatory genes.

These molecules can also have deleterious effects on normal tissue by damaging DNA, oxidizing membrane lipids, and nitrosylating proteins. Release of reactive intermediates can be initiated by microbial products such as LPS and lipoproteins, by cytokines such as IFN-γ and IL-8, and by engagement of Fc receptors by IgG. These events cause translocation of several cytosolic proteins, including Rac2 and Rho-family guanosine triphosphatase (GTPase) to the membrane-bound complex carrying cytochrome *c*, with subsequent activation of reduced nicotinamide adenine dinucleotide phosphate (NADPH) oxidase. The reaction catalyzed by NADPH oxidase leads to superoxide production, which, in turn, increases hydrogen peroxide, hydroxyl radicals and anions, hypochlorous acid, and chloramines.

In some cases, ROIs can contribute directly to the initiation of chronic disease. Lipid oxidation produces aldehydes that substitute lysine residues in apolipoprotein B-100. This altered moiety either binds to TLR2 to induce cytokine production or is internalized by macrophages, leading to the production of foam cells and fatty streaks, the primary lesions of atherosclerosis (Chapter 70). Subsequently, altered epitopes in damaged host proteins can be presented to T cells to initiate an adaptive immune response that amplifies the inflammatory vascular lesion.

Nitric oxide synthases (NOS) convert L-arginine and molecular oxygen to L-citrulline and nitric oxide (NO). There are three known isoforms of NOS: neuronal NOS (ncNOS or NOS1) and endothelial cell NOS (ecNOS or NOS3) are both constitutively expressed, whereas macrophage NOS (macNOS, iNOS, or NOS2) is induced by inflammatory cytokines such as TNF and IFN-γ, as well as by products of viruses, bacteria, protozoa, and fungi and by low oxygen tension and low environmental pH.

Together with prostaglandins, the production of NO by NOS2 and ROIs by NADPH oxidase is a key mechanism by which macrophages paradoxically impair T-cell proliferation. This might control inflammatory processes or delete autoreactive T cells and partially accounts for the immunosuppression observed in certain infections and malignancies.

Proteases and Matrix Damage

Production of enzymes that degrade the ECM regulates tissue turnover in inflammation. Reconfiguring of the matrix remodels damaged tissue, releases matrix-bound growth factors and cytokines, prepares the tissue for the ingrowth of new blood vessels, and alters the local milieu to permit adherence and retention of newly recruited cells.

The MMPs are a family of more than 20 extracellular endopeptidases that participate in degradation and remodeling of the ECM matrix (Table 48-1). They are produced as pro-enzymes and require limited proteolysis or partial denaturation to expose the catalytic site. Their name is derived from their dependence on metal ions (zinc/metzincin superfamily) for activity and from their potent ability to degrade structural ECM proteins. MMPs can also cleave cell surface molecules and other pericellular nonmatrix proteins, thereby regulating cell behavior. For instance, MMPs can alter cell growth by digesting matrix proteins associated with growth factors. FGF and TGF-β have high affinities for matrix molecules that serve as depots for storage of these cytokines. Matrix proteolysis releases some growth factors and can make them available to cell surface receptors. In addition, MMPs can directly cleave and activate growth factors. MMPs affect cell migration by altering cell-matrix or cell-cell receptor sites. The adhesion molecule β_4 integrin is

TABLE 48-1 COMMON MATRIX METALLOPROTEINASES AND THEIR SUBSTRATES

MMP FAMILY	MATRIX SUBSTRATES	OTHER SUBSTRATES
Collagenases	Collagen I, II, III, VII, and X Aggrecan	Pro-MMP-1, -2, -8, -9, and -13 Pro-TNF
Entactin	α_1-Proteinase inhibitors Gelatin Tenascin	
Gelatinases	Aggrecan Denatured collagen Elastin Fibronectin Laminin Vitronectin	Pro-MMP-1, -2, and -13 Pro-TNF Pro-IL-1β Latent TGF-β
Matrilysins	Proteoglycans Denatured collagens Entactin Fibrin, fibrinogen Fibronectin Gelatin Laminin Tenascin Vitronectin	Pro-MMP-2 and -7 Pro-TNF Membrane-bound Fas ligand (FasL) Plasminogen β_4 Integrins
Stromelysins	Proteoglycans Aggrecan Collagen III, IV, V, IX, X, and XI Pro-IL-1β Entactin Fibrin, fibrinogen Fibronectin Gelatin Laminin Tenascin Vitronectin	Pro-MMP-1, -3, -7, -8, -9, -10, and -13 Pro-TNF Plasminogen α_1-Proteinase inhibitors

IL = interleukin; MMP = matrix metalloproteinase; TGF = transforming growth factor; TNF = tumor necrosis factor.

cleaved by MMP-7. MMP-3 and MMP-7 digest E-cadherin and not only disrupt endothelial cell junctions but also stimulate cell migration.

Degradation of the ECM is usually initiated by collagenases, which cleave native collagen. Denatured collagen is then recognized and further degraded by gelatinases and stromelysins. Unlike the collagenases, stromelysins demonstrate broad substrate specificity and act on many ECM proteins, such as proteoglycan, fibronectin, laminin, and many cartilage proteins. Stromelysins can also amplify the remodeling process by activating collagenase through limited proteolysis.

MMP gene expression can be induced by many pro-inflammatory cytokines, including TNF, IL-1, IL-17A, and IL-18. One common element in MMP promoters that regulates transcription is activator protein-1 (AP-1). AP-1 is a dimer that includes members of the Jun and Fos families. Cytokines can regulate the *MMP* gene by activating MAPKs, especially c-Jun amino terminal kinase (JNK), which, in turn, phosphorylates c-Jun and markedly enhances MMP production. NF-κB and NF-κB-like binding sites also can contribute to protease transcription.

Several other classes of proteases remodel the matrix, including serine proteases and cysteine proteases. High levels of active serine proteases, such as trypsin, chymotrypsin, and elastase, are released by infiltrating PMNs at sites of inflammation and can directly digest the ECM or activate the proenzyme forms of secreted MMPs. The ADAM (a disintegrin and metalloproteinase) family can cleave the extracellular domain of cytokine receptors. These ECM proteases include two members of the aggrecanase family. One of the aggrecanases (aggrecanase 2, or ADAMTS5) has been implicated in osteoarthritis because mice deficient in this enzyme have decreased cartilage destruction in models of osteoarthritis (Chapter 262).

TISSUE REPAIR AND RESOLUTION OF INFLAMMATION

Inflammation is a normal physiologic response but can cause serious host injury if allowed to persist. Additional mechanisms are required to reestablish homeostasis after this response is initiated. Suppression of acute inflammation by removal or deactivation of mediators and effector cells permits the host to repair damaged tissues through elaboration of appropriate growth factors and cytokines (Fig. 48-1). As in the initial generation of an inflammatory response, components of resolution include a cellular response (apoptosis and necrosis), formation of soluble mediators (such as anti-inflammatory cytokines and antioxidants), and production of direct effectors (such as protease inhibitors).

Deletion of Inflammatory Cells

Cells can be removed from an inflammatory site by several mechanisms. First, the influx of cells can be decreased by suppressing chemotactic factor production and vascular adhesion molecule expression. Second, cells, especially lymphocytes, can be released from the tissue and return to the circulation through lymphatics. Third, stressed cells can undergo necrosis with the release of their contents into the local environment. A fourth mechanism, known as autophagy,[7] can lead to digestion of internal organelles and ultimately to cell death. Perhaps the most critical and effective method for clearing cells from an inflammatory site is programmed cell death, or apoptosis.

Apoptosis is a highly regulated process in eukaryotic cells that leads to cell death and marks the surface membrane for rapid removal by phagocytes. This clearance process does not elicit an inflammatory response, in contrast to cell death by necrosis. PMN phagocytes have a very short half-life in the tissue, and the persistence or release of their contents into the microenvironment after death can be deleterious. In some pathologic conditions, such as leukocytoclastic vasculitis (Chapter 270), abundant neutrophil apoptosis is readily apparent on histopathologic examination. Other cells, including T lymphocytes, undergo postactivation apoptosis to prevent an overwhelming persistent host response. Defective apoptosis or even persistence of apoptotic cells that escape clearance may contribute to chronic inflammatory and autoimmune diseases. For instance, loss of tolerance to self-antigens might participate in autoimmune responses in SLE.

Commitment of a cell to apoptosis can be initiated by a number of factors, including the ROIs in the cellular microenvironment as well as signaling through several death receptor pathways (e.g., FasL/Fas and TNF-related apoptosis-inducing ligand [TRAIL]). The former can damage DNA, which is a common byproduct of the genotoxic environment created by inflammation. If DNA damage is excessive, repair by tightly regulated mismatch repair mechanisms is terminated, and programmed cell death can be initiated by genes such as the p53 tumor suppressor. The burden of mutations induced by ROIs or RNIs in chronic inflammation can potentially accumulate over time and eventually lead to amino acid substitutions in key regulatory proteins. Ultimately, as has been observed in ulcerative colitis, neoplastic disease can ensue.

Removal of apoptotic bodies, or the remnants of packaged apoptotic cells, is rapid and can be accomplished by macrophages, fibroblasts, epithelial and endothelial cells, muscle cells, and dendritic cells. The surface receptors used in recognition and engulfment of apoptotic cells include integrins (e.g., $\alpha_v\beta_3$), lectins, scavenger receptors, ATP-binding cassette transporter 1, LPS receptor, CD14, and complement receptors CR3 and CR4. However, some of these membrane molecules can be used in both pro-inflammatory and apoptotic pathways, the divergence of which may be based on differing ligands and accessory molecules. Apoptotic cells display a series of membrane-associated molecular patterns that interact with receptors on phagocytes. A general feature of apoptotic cells is loss of phospholipid asymmetry, with external presentation of phosphatidylserine. Externalized phosphatidylserine

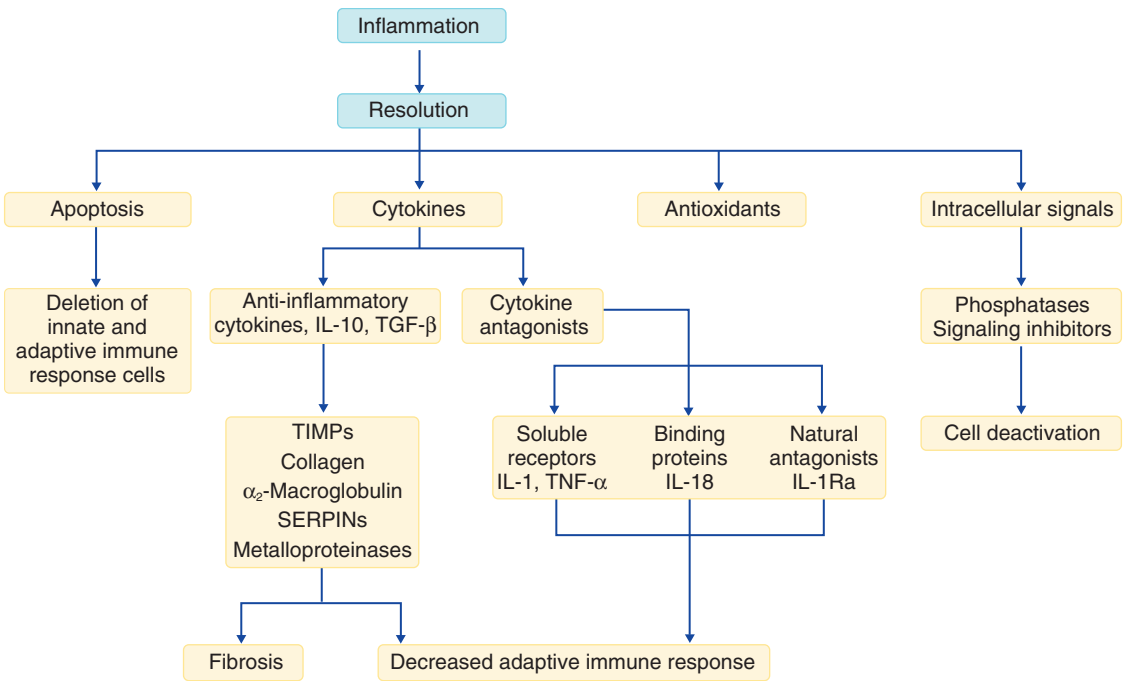

FIGURE 48-1. Anti-inflammatory mechanisms that resolve inflammation and lead to repair of the extracellular matrix. IL = interleukin; SERPINs = serine protease inhibitors; TGF = transforming growth factor; TIMPs = tissue inhibitor of metalloproteinases; TNF = tumor necrosis factor.

may be sufficient to trigger phagocytosis, but other apoptotic cell surface structures exist.

Although some inflammatory and immune cells are being deleted, other cell lineages expand during the resolution phase. Mesenchymal cells, especially fibroblasts, proliferate and produce new matrix that can contract to form a fibrotic scar. Locally produced growth factors such as PDGF induce DNA synthesis of these stromal cells through activation of PI3Ks. TGF-β[8] also stimulates fibroblast proliferation and converts cell phenotype to matrix formation rather than matrix destruction by increasing collagen production and suppressing MMP expression. In addition, mesenchymal stem cells that either reside in the tissue or migrate from the peripheral blood can differentiate into the appropriate organ-specific lineage. The pluripotential cells, in the presence of the appropriate milieu, can become adipocytes, chondrocytes, bone cells, or other terminally differentiated stromal cells.

Soluble Mediators

ANTI-INFLAMMATORY CYTOKINES

A variety of anti-inflammatory cytokines are released by resident and infiltrating cells. TGF-β and IL-10 are examples that are produced by macrophages, interstitial fibroblasts, or T cells. Some T-cell cytokines, including IL-4, IL-10, and IL-13, suppress the expression of MMP by cells stimulated by IL-1 or TNF. In addition to increasing fibroblast proliferation, TGF-β suppresses collagenase production, increases collagen deposition, and decreases MMP activity by inducing production of the tissue inhibitors of metalloproteinases (TIMPs). The repair phase is abnormal in diseases in which tissue fibrosis represents a major pathologic manifestation. For example, scleroderma (Chapter 267) is marked by diffuse fibrosis and is accompanied by high levels of TGF-β and increased production of ECM.

Cytokine decoy receptors can also downregulate the inflammatory response. Receptors can also be shed from the cell surface after proteolytic cleavage and can absorb cytokines, thereby preventing them from ligating functional receptors on cell membranes. These cytokine inhibitors can be released as a coordinated attempt to prevent unregulated inflammation, as in septic shock (Chapter 108), in which endotoxin induces production of soluble receptors after initial massive production of TNF and IL-1. Other types of cytokine-binding proteins are also produced as counter-regulatory mechanisms, including IL-18-binding protein (IL-18BP), which is an Ig superfamily–related receptor that captures IL-18. In bone remodeling (Chapter 243), interactions of receptor activator of NF-κB (RANK) with RANK ligand are required for osteoclast-mediated resorption. The competitive antagonist osteoprotegerin is a member of the TNF receptor family that binds to RANK ligand and inhibits osteoclast activation.

At least two distinct mechanisms contribute to natural IL-1 inhibition. An IL-1 decoy receptor (type II IL-1R) has both cell membrane and soluble forms that neutralize IL-1 activity. In addition, a natural IL-1 antagonist, IL-1Ra, can bind to functional IL-1 receptors and compete with IL-1α or IL-1β. However, IL-1Ra does not transduce a signal to the cell and blocks the biologic functions of ambient IL-1. The balance of IL-1 and IL-1Ra production depends on many influences. For instance, monocytes produce more IL-1, whereas mature macrophages produce IL-1Ra.

DEACTIVATION OF SIGNALING PATHWAYS

The signaling pathways described previously that initiate an inflammatory response have intracellular mechanisms to ensure that the process is self-limited. Many kinases, such as the MAPKs, require post-translational modification through phosphorylation to increase enzyme activity. A system of phosphatases that remove these phosphates can return the kinase to its resting form. For example, dual specificity phosphatase 1 (DUSP1) is an enzyme that dephosphorylates p38 MAPK as well as other MAPKs. DUSP1 expression is increased by p38 MAPK; thus, the very process of activating the cell through p38 is responsible for its own counter-regulatory mechanism. NF-κB activation is typically initiated by phosphorylation of the inhibitor of kB (IkB), which targets it for proteolysis. IkB expression later increases dramatically and stops the signaling through this pathway. JAK-STAT signaling is inhibited by the suppressor of cytokine stimulation (SOCS) proteins. Thus, cellular defense mechanisms have evolved to prevent persistent cell activation.

ANTI-INFLAMMATORY PROSTANOIDS AND CYCLOOXYGENASE

COX2 induced by pro-inflammatory mediators appears early and can contribute to inflammatory responses. However, COX2 expression late in the process has led to speculation that it also functions in the resolution of inflammation. This regulation might occur through formation of the cyclopentenone prostaglandins (CyPG). The prostanoids can serve as ligands for peroxisome proliferator-activated receptors (PPARs) (Chapter 206). There are three main classes of PPAR receptors—PPARα, PPARβ/δ, and PPARγ—all of which bind to DNA as heterodimers in association with the retinoid X receptor. Activation of PPARγ by CyPG is associated with the suppression of AP-1 and STAT transcriptional pathways in macrophages. A variety of natural and synthetic PPAR agonists have demonstrated efficacy in models of ischemia-reperfusion injury, arthritis, and inflammatory airway disease.

Inhibitors of Direct Effectors

ANTIOXIDANTS

Antioxidant enzymes that can inactivate the toxic intermediates and protect normal tissues include catalase and superoxide dismutase. Catalase is a peroxisomal enzyme that catalyzes the conversion of hydrogen peroxide to water and oxygen. Superoxide dismutases (SODs) catalyze the dismutation of superoxide to hydrogen peroxide, which is then removed by catalase or glutathione peroxidase. Glutathione peroxidases and glutathione reductase are additional mechanisms for maintaining redox balance and removal of toxic metabolites. Insufficient production of intracellular antioxidants such as glutathione can suppress lymphocyte responses and could account for defective T-cell receptor signaling and blunted immunity in T cells derived from rheumatoid arthritis synovium (Chapter 264).

Interactions of free radicals with surrounding molecules can generate secondary radical species in a self-propagating chain reaction. Chain-breaking antioxidants are small molecules that can receive or donate an electron and thereby form a stable byproduct with a radical. These antioxidant molecules are categorized as either aqueous phase (vitamin C, albumin, reduced glutathione) or lipid phase (vitamin E, ubiquinol-10, carotenoids, and flavonoids). In addition, transition metal-binding proteins (ceruloplasmin, ferritin, transferrin, and lactoferrin) can serve as antioxidants by sequestering cationic iron and copper and thereby inhibiting the propagation of hydroxyl radicals.

PROTEASE INHIBITORS

Protease inhibitors regulate the function of endogenous proteases and reduce the likelihood of collateral damage to tissues. These proteins form two functional classes, active site inhibitors and α_2-macroglobulin (α_2M). The latter class of protease inhibitors acts by covalently linking the protease to the α_2M chain and thereby blocking access to substrates. α_2M binds to all classes of proteases and, after forming a covalent bond, conveys them to cells through receptor-mediated endocytosis with subsequent enzymatic inactivation. The family of inhibitors of serine proteases (SERPINs) are the most abundant members of the former class of protease inhibitors and play a major role in regulation of blood clot resolution and inflammation, as indicated by many of their names: antithrombin III, plasminogen activator inhibitors 1 and 2, α_2-antiplasmin, α_1-antitrypsin, and kallistatin.

The TIMP family blocks the function of most MMPs. The TIMPs bind to activated MMPs and irreversibly block their catalytic sites. Examples of disease states with an unfavorable balance between TIMPs and MMPs include loss of cartilage in arthritis and regulation of tumor metastasis. TIMP-MMP imbalance in destructive forms of arthritis appears be caused by the limited production capacity for protease inhibitors, which is overwhelmed by the prodigious expression of MMPs. Whereas IL-1 and TNF induce MMPs, IL-6 and TGF-β suppress production of MMPs and increase levels of TIMPs. Therefore, the cytokine profile has a profound influence on the status of remodeling. When pro-inflammatory cytokines predominate, the balance favors matrix destruction; in the presence of pro-inflammatory cytokine inhibitors and growth factors, matrix protein production increases, and MMPs are inhibited by TIMPs.

Grade A Reference

A1. Langley RG, Elewski BE, Lebwohl M, et al. Secukinumab in plaque psoriasis—results of two phase 3 trials. *N Engl J Med*. 2014;371:326-338.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bryant CE, Symmons M, Gay NJ. Toll-like receptor signalling through macromolecular protein complexes. *Mol Immunol*. 2015;63:162-165.
2. Pal S, Bhattacharjee A, Ali A, et al. Chronic inflammation and cancer: potential chemoprevention through nuclear factor kappa B and p53 mutual antagonism. *J Inflamm (Lond)*. 2014;11:23.
3. Strowig T, Henao-Mejia J, Elinav E, Flavell R. Inflammasomes in health and disease. *Nature*. 2012;481:278-286.
4. Arend W, Firestein GS. Pre-rheumatoid arthritis: predisposition and transition to chronic synovitis. *Nature Rev Rheumatol*. 2012;8:573-586.
5. Aoki T, Narumiya S. Prostaglandins and chronic inflammation. *Trends Pharmacol Sci*. 2012;33:304-311.
6. Yamaura K, Shigemori A, Suwa E, et al. Expression of the histamine H4 receptor in dermal and articular tissues. *Life Sci*. 2013;92:108-113.
7. Choi AM, Ryter SW, Levine B. Autophagy in human health and disease. *N Engl J Med*. 2013;368:651-662.
8. Samarakoon R, Overstreet JM, Higgins PJ. TGF-β signaling in tissue fibrosis: redox controls, target genes and therapeutic opportunities. *Cell Signal*. 2013;25:264-268.

49 TRANSPLANTATION IMMUNOLOGY

MEGAN SYKES

DEFINITION

Clinical transplantation encompasses transplantation of organs and islets of Langerhans containing insulin-producing β cells, in which it is necessary to overcome the host-versus-graft (HVG) immune response to avoid rejection, as well as hematopoietic cell transplantation (HCT) (Chapter 178), in which it is necessary to contend with not only the HVG but also the graft-versus-host (GVH) immune response. Because preparations of bone marrow or mobilized peripheral blood stem cells (mPBSCs) contain mature T cells, their administration to conditioned and consequently immunoincompetent recipients is associated with the risk for GVH disease. Organs transplanted include corneas, kidneys, livers, hearts, lungs, small intestines, pancreases, and composite tissue allografts such as hands and faces. The list of transplanted allogeneic cells is likely to expand in the future to include other cell types, such as hepatocytes, myoblasts, and stem cell–derived replacement cells. Transplants originating from a member of the same species are referred to as *allotransplants*. However, transplants from other species, termed *xenografts,* are believed by many to be a promising solution to the severely inadequate supply of allogeneic organs and tissues, and such grafts may be used in the future. Transplants of tissues or cells originating from the recipient, either by processing of cells from the recipient's own organ (e.g., islets of Langerhans following pancreatectomy for chronic pancreatitis) or cell populations (e.g., $CD34^+$ hematopoietic progenitor and stem cells collected from leukapheresis products following mobilization from the bone marrow before high-dose radiation or chemotherapy to treat cancer) are referred to as *autologous*. In the future, these transplants may include stem cell–derived autologous cells used for therapeutic purposes.

ANTIGENS IN TRANSPLANTATION

The major antigens recognized during graft rejection and the cell types targeting them are summarized in Table 49-1.

Major Histocompatibility Antigens

The major histocompatibility complex (MHC; human leukocyte antigens [HLAs] in the human) controls adaptive and some innate immune responses and is of central importance in many immune-mediated diseases. The MHC also presents the strongest immunologic obstacle to all types of allografts. The HLA molecules include two major isoforms, termed class I and class II, and are all encoded in the MHC complex of chromosome 6. Although all HLA molecules have a similar general structure, class I and class II molecules show different expression patterns, with class I MHC expressed on most cells of the body, whereas class II antigens are expressed mainly on antigen-presenting cell (APC) populations, such as dendritic cells, macrophages, and B cells, as well as thymic epithelial cells involved in T-cell selection. Class II MHC can also be expressed on vascular endothelial cells and activated T cells of some species, including humans.

The specialized function of both classes of MHC molecules is the presentation of peptide antigens to T-cell receptors (TCRs), allowing adaptive immune responses to occur. In general terms, the peptides presented by class I molecules are 8- to 9-amino acid peptides derived from cytosolic proteins (e.g., viral proteins) that are transported into the endoplasmic reticulum, where they are processed and loaded onto class I molecules during their synthesis. CD8 molecules interact with the α_3 domain of the class I heavy chain, thereby strengthening the interaction of $CD8^+$ T cells that recognize class I–peptide complexes. Peptides presented by class II MHC molecules, on the other hand, are mostly 10 to 20mers derived from exogenous proteins (e.g., phagocytosed bacteria) that are processed through the endosomal processing pathway, and these are recognized by TCRs of T cells whose CD4 molecules strengthen the overall T cell–APC interaction. The class I presentation pathway is of particular importance in allowing destruction of virally infected cells, consistent with the expression of class I MHC on almost all cell types in the body. However, there are exceptions to this paradigm that account for the phenomenon of cross-priming and cross-presentation, wherein peptides from exogenous proteins are presented by class I molecules, a phenomenon that may have significance for transplantation. Class II MHC presentation of exogenous antigens takes place primarily on professional APCs and B cells, consistent with the role of $CD4^+$ T cells in initiating immune responses by activating APCs, providing direct and indirect (through activated APCs that also present peptides on class I molecules) "help" for $CD8^+$ cytotoxic T lymphocytes (CTLs), and providing help for antibody-producing B cells. B cells are able to focus the antigens recognized by their specific surface immunoglobulin receptors by binding and internalizing these antigens, which thereby predominate in the endosomal antigen-processing pathway and become presented by a high proportion of class II molecules on that B cell. This ability of B cells to preferentially present peptides derived from their cognate antigens to CD4 T cells recognizing those alloantigens is very important in driving alloantibody production.

A number of MHC molecules have been crystallized, both alone and with TCRs that recognize them. The TCR binding structure of class I and II MHC molecules is similar overall and includes both the peptide binding cleft formed by a β-pleated sheet and two α-helices forming the sides of the cleft (E-Fig. 49-1). However, class I and II MHC molecules also have significant structural differences, as summarized in E-Table 49-1. Although class I molecules are formed by the combination of a highly variable heavy (45-kD) chain (α chain) noncovalently linked to a nonpolymorphic, smaller (12-kD) light chain (β_2-microglobulin), class II molecules are heterodimers of two polymorphic chains, a 32-kD α chain and a noncovalently bound 28-kD β chain. TCRs interact physically with both the α-helices of the MHC molecules and side chains of the peptide that is bound in the groove, representing a trimolecular MHC-peptide-TCR interaction (see E-Fig. 49-1). It is the most variable ("hypervariable") portion of the TCR, produced by V-D-J somatic rearrangements and N insertions in the TCR α and β chains, known

TABLE 49-1 LYMPHOCYTES INVOLVED IN GRAFT REJECTION

CELL TYPE	ANTIGENS RECOGNIZED	FUNCTION	RELEVANCE
$CD4^+$ T cells	Allogeneic class II MHC (+ peptide) Self class II MHC + donor peptide	Antigen-presenting cell activation Help (cytokines and costimulation) Proinflammatory cytokine production Cytotoxicity Regulatory function	Organ allografts Cellular allografts Xenografts GVHD
$CD8^+$ T cells	Allogeneic class I MCH (+ peptide) Self class I MHC + donor peptide	Cytotoxicity Cytokine production Regulatory function	Organ allografts Cellular allografts Xenografts GVHD
NK cells	Class I MHC (activates or inhibits NK cell function) Other activating ligands	Cytotoxicity Cytokine production	? Organ allografts Cellular allografts Xenografts
B cells	Class I and class II MHC blood group antigens Xenogeneic carbohydrates	Antibody-mediated rejection (hyperacute, acute humoral, and chronic rejection)	Organ allografts Cellular allografts Xenografts

CTL = cytotoxic T lymphocyte; GVHD = graft-versus-host disease; MHC = major histocompatibility complex; NK = natural killer.

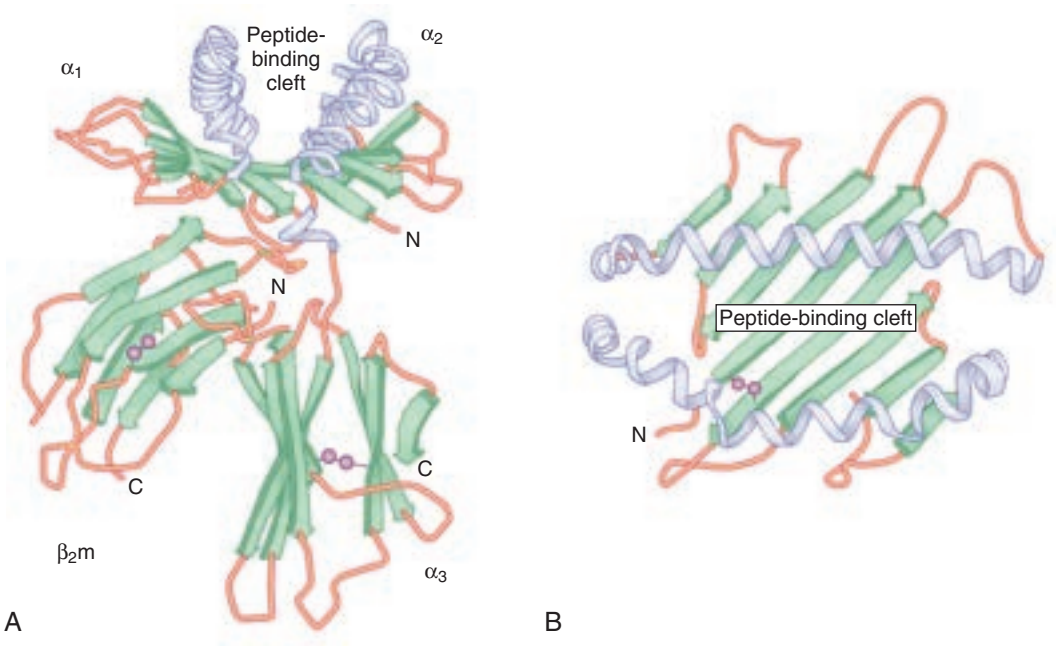

E-FIGURE 49-1. Two views of an HLA class I molecule. A, Ribbon diagram showing the x-ray crystallographic structure of an HLA class I molecule (side view). The β-strand structures are indicated by *thick green arrows* (oriented in an amino to carboxy direction), whereas connecting loops are indicated as *thin lines*. The α-helices are shown flanking a peptide-binding cleft at the top (membrane distal portion) of the molecule. The base (membrane proximal portion) of the molecule is formed by the noncovalent association between the α_3 domain of the class I α chain and β_2-microglobulin (β_2m). **B,** View from the top of the molecule emphasizing that the base of the peptide-binding cleft consists of β-pleated sheets flanked by α-helical structures. C = C terminal; N = N terminal. (Adapted from Bjorkman PJ, Saper MA, Samraoi B, et al. Structure of the class I histocompatibility antigen HLA-A2. *Nature*. 1987;329:506-512.)

E-TABLE 49-1 COMPARISON OF STRUCTURAL AND FUNCTIONAL FEATURES OF HLA CLASS I AND CLASS II ISOTYPES

FEATURE	HLA CLASS I	HLA CLASS II
Chain structure of heterodimer	45-kD α chain 12-kD β_2-microglobulin	34-kD α chain 28-kD β chain
Tissue distribution	All nucleated cells	Antigen-presenting cells (monocytes, B cells, dendritic cells, Langerhans cells), thymic epithelium, and some T cells; inducible on other cell types by interferon-γ
Size of bound peptides	8-9 amino acids	10-20 amino acids
Source peptides	Cytosolic	Endosomal
Functions	Presentation of antigenic peptides to $CD8^+$ T cells; ligands for natural killer cell receptors	Presentation of antigenic peptides to $CD4^+$ T cells

as complementarity-determining region 3 (CDR3), that recognizes specific MHC-peptide complexes.

The HLA molecules are all encoded within a 3.6-million base-pair region that encodes more than 200 genes, including complement and tumor necrosis factor (TNF) genes and many others in addition to MHC that have immunologic functions. The organization of the HLA region is illustrated in E-Figure 49-2, which shows that the heavy chains for "classic" class I HLA-A, B, and C and "nonclassic" class I molecules are encoded in a region that is telomeric to the "central MHC" region that includes complement and TNF genes among others, and lies between the class II and class I HLA regions. The class II region contains two α- and β-chain genes, only one of which is functional, for each of HLA-DQ and DP. However, the DR locus contains different numbers of β chains for different HLA alleles. Some of these DR β chains are pseudogenes, but various HLA-DR alleles contain either one or two functional β-chain genes.

One of the striking features of HLA molecules (and the MHC of most mammalian species) is their extensive polymorphism. There are thousands of defined HLA alleles in the class I and class II regions. Because the primary function of antigen presentation to T cells is to permit responsiveness to and clearance of pathogenic microorganisms, this polymorphism may have evolved to maintain the diversity of immune responsiveness to various pathogens within a population, avoiding annihilation of that population by a single microorganism that might not be presented well by a particular MHC. HLA alleles were originally distinguished by panels of highly sensitized human sera containing multiple alloantibodies. Although it effectively identified structurally related HLA alleles, this method failed to distinguish many allelic differences that are of functional importance for antigen binding and T-cell recognition. It was only with the development of molecular methods to distinguish alleles at the genomic level, eventually through specific genomic sequences, that the full extent of the polymorphism in this region was revealed. In association with this knowledge, it has been necessary to continually revise and refine the system of nomenclature defining these alleles. According to the most recently accepted nomenclature,[1] HLA alleles are identified by the locus (e.g., HLA-A), followed by an asterisk, and then a unique number with up to four sets of digits separated by colons. The first set describes the allele group (e.g., HLA-A*02), which usually corresponds to a serologically defined antigen, and the second set indicates the specific allele (e.g., HLA-A*02:101). The third and fourth set of digits are of less practical importance because they identify silent nucleotide substitutions in different alleles and variations in the nontranslated regions of the gene, respectively.

Within certain populations, however, the level of diversity within allele groups may be quite limited because of the common genetic origin of the allele. For example, for the originally serologically defined HLA-DR3 allele group, there is little diversity among Northern Europeans, such that most carry the DRB1*0301 allele. Thus, for this population, it is reasonable to refer to the serologic HLA-DR3 type as defining this allele. There are certain alleles that predominate within racial groups. For example, as few as five DRB1 alleles predominate among Northern Europeans, with each allele represented in 10 to 30% of this population. E-Table 49-2 summarizes the major DRB1 allelic groups defined initially at the serologic level and later at the level of genomic sequencing.

Most organ transplantations are performed across HLA disparities, and the strong immunosuppressive regimens used in transplant recipients are designed to prevent rejection by this exceptionally strong immune response. In contrast to T-cell responses to peptide antigens derived from foreign proteins, which are recognized by a very small fraction of naïve T cells (in the range of 1 in 10^5), a very high proportion, estimated at 1 to 10% of the T-cell repertoire, recognizes MHC alloantigens. The strong immunogenicity of allogeneic MHC molecules relates to the manner in which T cells are selected in the thymus; developing thymocytes do not survive unless they can weakly recognize a self MHC/peptide complex on a thymic stromal cell. This process is termed *positive selection*. Thymocytes whose receptors have high affinity for self/MHC complexes are deleted, however, so strongly autoreactive T cells rarely make it into the peripheral T-cell pool. Allogeneic antigens are not part of this *negative selection* process. The net result of these two selection steps is that the human T-cell "repertoire" is strongly biased to have cross-reactivity to allogeneic MHC molecules, providing a barrier to organ and hematopoietic cell transplantation. In the case of organ transplantation, in which long-term pharmacotherapy with powerful immunosuppressive drugs is used in an effort to prevent graft rejection, this can translate into improved results with matched organs in some situations. However, for unrelated, deceased donor transplantation, the benefits of HLA matching may be counterbalanced by the disadvantages associated with prolonged graft ischemia when attempts are made to transport organs to the most closely matched recipient.[2]

For hematopoietic cell transplantation (Chapter 178), the risks for GVH disease and marrow graft failure are so greatly amplified in the presence of extensive HLA mismatches that such transplantations have been avoided whenever possible; if a sufficiently matched, related donor cannot be found, a search is conducted through large registries containing millions of volunteer unrelated donors. Because of its extensive polymorphism, truly MHC-identical, unrelated donors can be difficult to find in the human population at large. For individuals with common HLA genotypes, the likelihood of finding a matched unrelated donor is markedly greater than that for individuals with rare genotypes. This situation relates in part to the phenomenon of *linkage disequilibrium*, wherein alleles at nearby loci are found together on the same chromosomal segment, or *haplotype*, more frequently than would be predicted by chance. The pattern of linkage disequilibrium is different in different racial groups, so the chance of finding a truly genotypically identical haplotype is greatest within the same population. For example, among whites, the DRB1*0301 allele is in linkage disequilibrium with DQB1*0201, which is located several hundred thousand base pairs away on chromosome 6; this complex, in addition to the DR4 alleles that are in linkage disequilibrium with DQB1*0302, confers the greatest genetic component of risk for the development of type 1 diabetes. Many autoimmune diseases demonstrate similarly strong HLA associations. Although there are data to indicate that HLA-specific autoantigen presentation plays a major role in determining disease susceptibility, non-HLA genes in linkage disequilibrium likely account for a significant component of these genetic risk factors.

The use of alternative donors has also increased the availability of hematopoietic cell transplantation (HCT) in individuals for whom an HLA-identical related or unrelated donor cannot be identified (Chapter 178). The use of cord blood transplantation, which has reduced GVH disease–inducing activity compared with adult stem cell products, as well as advances in avoiding GVH disease in haploidentical related donor HCT, has recently increased the safety and use of HLA-mismatched HCT.[3]

Minor Histocompatibility Antigens

"Minor" histocompatibility antigens are peptides derived from polymorphic peptides presented by an MHC molecule. Even genotypically HLA-identical siblings have different minor histocompatibility antigens. These are sufficient to induce graft rejection if immunosuppressive pharmacotherapy is not used. In the case of HCT, significant GVH disease frequently (about 30 to 50% of the time) complicates transplantation between HLA-identical siblings, even with the use of pharmacologic immunoprophylaxis.

Other Antigens

The major blood group (ABO) antigens can be the targets of a dramatic "hyperacute" rejection process that occurs when mismatched vascularized grafts are transplanted. Recognition of blood group antigens on the endothelial surface of the graft vessels by recipient "natural" antibodies (antibodies that are present without known sensitization to the antigens) activates the complement and coagulation cascades, resulting in rapid graft thrombosis and ischemia. A similar outcome can occur after transplantation to an individual with preformed anti–donor HLA antibodies resulting from presensitization by prior transplantations, transfusions, or pregnancies. Antibodies against other polymorphic antigens, such as MHC class I–related chain A (MICA), have been associated with graft rejection. In the past, transplantation could not be successfully performed in the presence of a positive anti-donor crossmatch. However, considerable success has been achieved in the transplantation of ABO-mismatched kidneys, livers, and hearts (the latter in the neonatal period only), and in transplantation of kidneys to highly presensitized patients.[4,5] In the case of kidney and liver transplantation, initial removal of the antibody and sometimes depletion of B cells, as well as the infusion of intravenous immunoglobulin (IVIG), has led to these successes. ABO-mismatched neonatal heart transplantation has succeeded because the transplantations are performed before the recipient has developed high levels of anti–blood group antigen antibodies, and the B cells seem to be rendered tolerant to the donor blood group antigen by the grafting process. Recognition of blood group antigens can also be of significance in HCT, in which ABO barriers are routinely crossed in both directions. This can cause hemolysis of recipient erythrocytes if the mismatch is in the GVH direction, but this complication can be avoided by washing the cellular product before infusion. Mismatches in the HVG direction can cause more persistent problems due to ongoing destruction of donor erythropoietic cells, resulting in

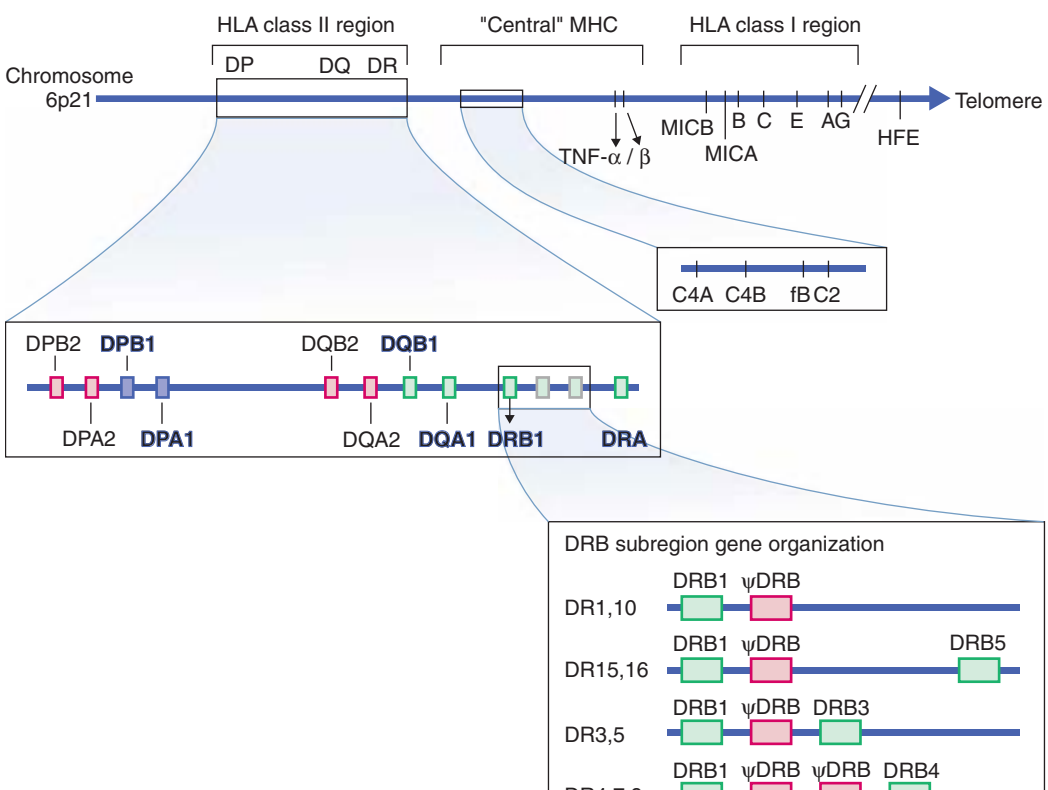

E-FIGURE 49-2. Map of the human major histocompatibility complex (MHC) spanning approximately 3.5 million base pairs on the short arm of chromosome 6. The HLA class I and class II molecules are encoded in distinct regions of the MHC. The HLA class II region contains three subregions: DR, DQ, and DP. Each of these subregions contains a variable number of α- and β-chain genes. HLA class II loci with known functional protein products are labeled in bold. In the case of DR, different numbers of DRB genes are present in different haplotypes, some of which are nonfunctional pseudogenes (ψ). A summary of the most common of these is shown in the *box*. The DQ and DP subregions each contain one pair of functional α- and β-chain genes. The HLA class I region contains the three "classic" class I genes—HLA-A, HLA-B, and HLA-C—as well as other related "nonclassic" class I molecules such as MICA, MICB, HLA-E, and HLA-G. The gene for familial hemochromatosis (HFE) is found just telomeric to the HLA class I region, about 3 million base pairs distant from HLA-A. The "central" MHC also contains a number of genes related to immune function, including the complement components (C4A, C4B, C2, and factor B), as well as tumor necrosis factor (TNF)-α and -β. Not shown in the figure are more than 100 additional genes, many of which are located in the central MHC. A complete listing of MHC-encoded genes can be found in Horton R, Wilming L, Rand V, et al. Gene map of the extended human MHC. *Nat Rev Genet*. 2004;5:889-899.

E-TABLE 49-2 SUMMARY OF MAJOR ALLELIC GROUPS AT THE DRB1 LOCUS AND THEIR RELATIONSHIP TO COMMON DRB1 ALLELES DEFINED AT THE SEQUENCE LEVEL

ALLELIC GROUPS (SEROLOGIC TYPING)		**EXAMPLES OF COMMON ALLELES (NORTHERN EUROPEAN WHITE INDIVIDUALS) DEFINED BY SEQUENCE***
Major Groups	***Serologic "Splits"***	
DR1		DRB1***0101**, 0102, 0103
DR2	DR15 DR16	DRB1***1501**, 1502 DRB1*1601
DR3		DRB1***0301**
DR4		DRB1***0401**, 0402, 0403, 0404, 0405, 0406, 0407, 0408
DR5	DR11 DR12	DRB1*1101, 1102, 1103, 1104 DRB1*1201
DR6	DR13 DR14	DRB1*1301, 1302, 1303 DRB1*1401
DR7		DRB1***0701**
DR8		DRB1*0801, 0802, 0803, 0804, 0806
DR9		DRB1*0901
DR10		DRB1*1001

*Alleles in bold are found in at least 10% of individuals in the population.
From Williams F, Meenagh A, Single R, et al. High resolution HLA-DRB1 identification of a Caucasian population. *Hum Immunol.* 2004;65:66-77.

pure red cell aplasia. More often, however, donor erythropoiesis is successfully established, and antidonor isohemagglutinins disappear from the circulation.

A and B blood group antigens are the consequence of the presence or absence of specific glycosylation enzymes in different individuals. Likewise, an antigenic specificity of the utmost importance in xenotransplantation is a carbohydrate epitope, Galα1–3Galβ1–4GlcNAc (αGal), which is produced by a specific galactosyl transferase. Humans and Old World monkeys lack a functional αGal transferase and produce high levels of natural antibodies against the ubiquitous αGal epitope. Because animals of interest as xenograft sources (e.g., pigs) express αGal at high levels on their vascular endothelium, transplantation of vascularized organs from pigs results in hyperacute rejection unless something is done to absorb the antibodies or inactivate complement. The development of αGal-knockout pigs, therefore, was an important milestone, and encouraging results have been obtained in pig-to-primate transplantation in initial studies.

In another type of transplant reaction, recognition as foreign results not from the presence of an antigen, but paradoxically from the absence of a self MHC molecule. Natural killer (NK) cells express a series of surface inhibitory and activating receptors that, collectively, determine whether the NK cell does or does not kill a potential target cell. The ligands for the inhibitory receptors are MHC class I molecules, and the receptors recognize specific groups of alleles. An NK cell may kill an allogeneic target that lacks a self MHC inhibitory ligand. This phenomenon has been shown in animal models to result in rapid bone marrow rejection when the donor marrow cells are not given in excess numbers or when a fraction of them are destroyed by an incompletely suppressed T-cell response. A similar phenomenon has not been clearly demonstrated in clinical HCT. The possibility that NK cells play a role in organ allograft rejection has long been an area of controversy. NK cells may be of particular importance in xenotransplantation, where they appear early in infiltrates of organ xenografts undergoing acute vascular rejection. NK cells clearly play a strong role in rejection of xenogeneic bone marrow, an observation that is relevant in one approach to inducing tolerance (see later discussion).

MECHANISMS OF REJECTION AND GRAFT-VERSUS-HOST DISEASE

Cellular Mediators

Many different cell types participate in rejection responses, and there is considerable redundancy. T cells are key players in most forms of rejection, with the exception of rejection that can be induced by antibodies in the absence of T-cell help. These include hyperacute and acute vascular rejection processes that may be induced by natural antibodies, as described earlier, or by antibodies that are present due to presensitization. The possible role of NK cells has already been discussed.

Direct and Indirect Allorecognition

T-cell responses are induced by APCs that present alloantigens. There are two forms of alloantigen recognition, termed *direct* and *indirect* (Fig. 49-1). Direct allorecognition denotes recognition of donor antigens on donor APCs provided by the graft. The extraordinarily high frequency of T cells with alloreactivity is caused by direct recognition of allogeneic MHC. Indirect recognition is the recognition of donor antigens that are picked up and presented on recipient MHC molecules on recipient APCs. The indirect response is more similar to "normal" T-cell responses, in which professional APCs present peptide antigens to T cells that are present at relatively low frequency in the naïve repertoire.

In organ transplantation, direct alloreactivity is particularly important in the early post-transplantation period, when APCs within the transplanted organ are still present; many of these cells migrate to the lymphoid tissues, where they initiate the alloresponse. However, the APC supply that comes with the donor graft is not renewable, so if the direct response is not maintained by recognition of donor antigens on endothelial cells or other cells in the graft, it may recede in importance. The indirect response, on the other hand, can be maintained by the constantly renewed pool of recipient APCs. The indirect response is of particular importance in inducing antibody responses.

Effector Mechanisms of Rejection

T cells can promote graft rejection through several effector mechanisms. One is the antibody-dependent processes that have already been discussed, which can be induced by $CD4^+$ helper T cells that promote differentiation and

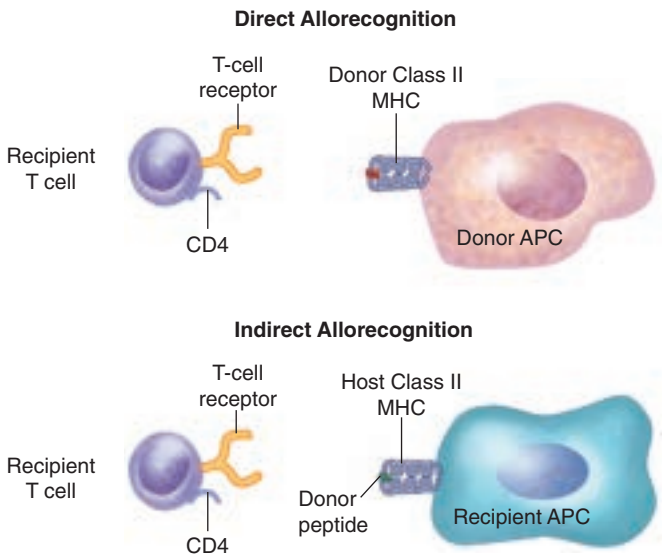

FIGURE 49-1. Direct and indirect allorecognition. Direct allorecognition involves the recognition by a T-cell receptor of major histocompatibility complex (MHC) molecules (with or without a peptide) on a donor antigen-presenting cell (APC). Indirect allorecognition involves recognition by the T-cell receptor of a donor peptide presented on a recipient APC that has picked up and processed donor antigens.

immunoglobulin (Ig) class switching of B cells that recognize other specificities on the same alloantigens. T cells provide cognate help to B cells when the TCRs recognize complexes of self MHC with donor MHC–derived peptide antigens (produced by B cells whose surface Ig receptors recognize and pick up the donor MHC antigen). If antidonor antibody is not present before transplantation but is induced afterward, the response can lead to the pathologic picture of acute humoral rejection. Antibodies may also participate in a slower, poorly understood process of chronic rejection, which, in the case of kidney and heart, is characterized by unique vascular lesions with intimal thickening and loss of the vessel space, and in the case of lung transplantation, by obliterative bronchiolitis. The mechanisms underlying these chronic rejection lesions are not well understood, and several different immune processes may in fact lead to similar lesions.

Another major effector pathway leading to graft rejection involves CTLs, which are predominantly members of the $CD8^+$ T-cell subset but also include $CD4^+$ T cells. Several effector mechanisms lead to killing of target cells by CTLs, and these include the granzyme/perforin-mediated pathway and the pathways involving Fas/Fas ligand (FasL) and other members of the TNF receptor family and their ligands (Chapter 47). Because $CD8^+$ cells recognize class I MHC molecules, which are widely expressed, it is not difficult to envision graft destruction by $CD8^+$ CTLs. $CD8^+$ CTLs may be activated through an APC that is stimulated initially through contact with an alloreactive $CD4^+$ cell. This is one form of CD4 "help" for $CD8^+$ cells. In addition, $CD8^+$ cells may be dependent on cytokines such as interleukin-2 (IL-2) from $CD4^+$ cells for their expansion and cytotoxic differentiation. However, there are also many examples of $CD8^+$ cell–mediated rejection that is independent of "help" from $CD4^+$ cells. Class II MHC, which is recognized by $CD4^+$ T cells, is less widely expressed on graft tissues than is class I MHC, although it may be induced on endothelial cells and graft parenchymal cells in the presence of inflammatory cytokines such as interferon-γ (IFN-γ).

In addition to cytotoxic mechanisms resulting from direct allorecognition, $CD4^+$ and $CD8^+$ T cells with indirect specificity seem also to be capable of causing graft destruction under some circumstances. Cytokines such as IFN-γ have been implicated in some instances, but in general, the pathways of indirect graft destruction are not well understood. A $CD8^+$ cell–mediated form of skin graft rejection that is dependent on donor antigens cross-presented on recipient MHC molecules (a form of indirect allorecognition for $CD8^+$ cells) has been described in an animal model. This form of graft rejection may be directed at antigen presented on endothelial cells of recipient vessels that revascularize the graft. This mechanism would not apply to primarily vascularized organ allografts.

The Role of T-Cell Trafficking

All the rejection processes described require trafficking of T cells into the graft. This process is made possible after the initial activation of naïve T cells

in the lymphoid tissues. Naïve T cells can migrate into lymph nodes because of their expression of the CCR7 chemokine receptor and the adhesion molecule L-selectin. These T cells are activated by migratory graft APCs that also enter the lymph nodes. T-cell activation is associated with loss of CCR7 and L-selectin expression and acquisition of a new set of chemokine receptors and adhesion molecules that allow rolling and adhesion on the graft endothelium and entry into the graft parenchyma (Chapter 47). Inflammation in the graft, such as that induced by ischemia-reperfusion injury and the transplantation procedure, as well as that induced by initially responding T cells, is associated with upregulation of chemokines and adhesion ligands that promote entry of lymphocytes into the graft. Nevertheless, well healed-in grafts can be slowly rejected by adoptively transferred memory T cells, demonstrating that acute graft injury and inflammation are not essential for rejection in the presence of an established memory T-cell response. Rejection of hematopoietic cell grafts may involve many of the same mechanisms as those discussed for solid organs, although less detailed work has been done in this area.

Mechanisms of Graft-versus-Host Disease

Initiation of GVH disease (Chapter 178) requires that donor T cells recognize host alloantigens. The disease involves attacks on a variety of recipient epithelial tissues, namely skin, the intestine, and liver. Animal models have demonstrated clear roles for both $CD4^+$ and $CD8^+$ cells in initiating GVH disease, and each subset is able to do so independently of the other. The mechanisms of GVH disease include activation of alloreactive donor T cells by recipient APCs, leading to the differentiation of effector cells with direct cytotoxic activity and cytokine production in response to host antigens. A prominent role is played by TNF-α, whose production is induced in part by the translocation of bacteria across the intestinal wall, promoting innate immune system activation through toll-like receptors (Chapter 45). An intensely pro-inflammatory environment is produced by the combination of conditioning-induced tissue injury and disruption of mucosal barriers, bacterial activation of the innate immune system, and the GVH alloresponse. An important role is now appreciated for the inflamed microenvironment in target tissues in promoting the trafficking of GVH-reactive T cells into these tissues.[6]

STRATEGIES TO PREVENT GRAFT-VERSUS-HOST DISEASE

In view of the critical role of donor T cells in inducing GVH disease, an obvious strategy for preventing this complication is to remove mature T cells from the marrow graft. This approach has indeed been shown in both animal models and clinical studies to prevent GVH disease effectively. However, there are several disadvantages to this approach. One is that adult humans, particularly those who have undergone prior chemotherapy and radiotherapy, have little remaining thymic tissue and therefore demonstrate sluggish T-cell recovery, leading to serious opportunistic infections.

The second disadvantage applies to the most common indication for allogeneic HCT, namely the treatment of hematologic malignancies (Chapter 178). In this setting, T-cell depletion is often associated with an increased relapse rate due to loss of a graft-versus-tumor (GVT) effect, which is in large part mediated by GVH alloreactivity. Separation of GVH disease from GVT effects is a major goal of research in HCT, and some promising strategies are being explored (E-Table 49-3). These include control of T-cell trafficking so that the GVH alloresponse is confined to the lymphohematopoietic tissues where the tumor resides and a number of other approaches.[6,7]

The third disadvantage of donor T-cell depletion in HCT is that it increases the rate of engraftment failure. GVH alloreactivity and a "veto" effect of donor T cells help to overcome host resistance to donor engraftment. A veto cell, which may be a T cell or an NK cell, kills a CTL that attacks it. Although the phenomenon has been well established in animal models, its mechanisms are not clearly established, and its potential role in humans is uncertain. NK-cell recognition in the GVH direction resulting from the absence in the recipient of a class I MHC ligand (E-Fig. 49-3) that can trigger a donor NK-cell inhibitory receptor (KIR) may promote donor marrow engraftment and antitumor effects against acute myeloid leukemias in the setting of T-cell-depleted, HLA-mismatched HCT.

Clinically, pharmacologic immunosuppressive prophylaxis is usually used in at least the first 6 months after HCT to minimize the complication of GVH disease. Additionally, HLA-matched or closely matched donors are chosen whenever possible because GVH disease increases in frequency and severity as increased HLA barriers are transgressed. These measures, nevertheless, are insufficient, and GVH disease remains a major complication of HCT. Therefore, many of the new strategies being explored in organ transplantation and other fields are also being examined for the prevention of GVH disease in experimental models. It should be borne in mind, however, that tolerance of donor T cells to recipient alloantigens (see later discussion) might not be entirely beneficial in the HCT setting for the treatment of malignant disease because loss of GVH alloreactivity is likely to come with loss of antitumor effects.

STRATEGIES TO PREVENT ALLOGRAFT REJECTION

Nonspecific Immunosuppression

Immunosuppressive drugs are the mainstay of clinical organ transplantation, and improvements in these drugs following the discovery of cyclosporine have extended organ transplantation to include hearts, lungs, pancreases, livers, and other organs and tissues in the past 30 years. The mechanisms of action of these agents are discussed in Chapter 35. However, it is noteworthy that, despite these improvements and their enormous impact on early graft survival, these agents have been less effective in attenuating late graft loss. Because chronic immunologic rejection processes and side effects of the immunosuppressive drugs themselves are responsible for much of this late graft loss, improved immunosuppressive agents and induction of immune tolerance (see later discussion) are major research goals in transplantation.

Costimulatory Blockade

As understanding of immune responses has increased, recent years have seen the exploration of numerous biologic agents, including antibodies and small molecules targeting receptors of the immune system as well as cell-based therapies, in efforts to improve allograft survival. Because of the central role played by T cells in the immune response, considerable attention has been focused on blockers of T-cell costimulation. When a naïve T cell recognizes antigen through its unique TCR, additional "costimulatory" signals are required to allow full activation, expansion, and differentiation to occur. These signals are often provided by APCs in the form of ligands (e.g., B7-1, B7-2) for costimulatory receptors (e.g., CD28) on the T cell. Cross-talk between the T cell and the APC (e.g., due to CD40 activation by CD154 upregulation on the activated T cell) further amplifies the costimulatory activity of the APC, allowing it to effectively activate other T cells as well. The CD154 (T cell)–CD40 (B cell) interaction also promotes Ig class switching and functioning of B cells as APCs. Blockade of these processes (e.g., by CTLA4Ig and anti-CD154 monoclonal antibodies [mAbs]) has led to marked prolongation of allograft survival in stringent rodent and large-animal models. Robust, systemic tolerance to donor antigens has been achieved in rodents receiving bone marrow transplantation with costimulatory blockade and little or no additional conditioning. Some of these agents have joined the armamentarium of immunosuppressive agents in clinical trials in transplantation and autoimmune diseases.[8] Although anti-CD154 antibodies have been associated with thromboembolic complications, precluding further evaluation in transplantation trials, recently developed anti-CD40 antibodies have shown promise in animal studies. Numerous additional costimulatory and inhibitory pathways that affect T-cell responses have been described, and these all are potential targets for further manipulation of the alloresponse.

Immune Tolerance

Immune tolerance denotes a state in which the immune system is specifically unreactive to the donor graft (or recipient in the case of GVH reactivity) while remaining normally responsive to other antigens.[9-11] Tolerance is distinct from the state produced by nonspecific immunosuppressive agents, which increase risks for infection and malignancy. Numerous approaches to tolerance induction have been described in rodent models, largely owing to the strong tolerogenicity of primarily vascularized heart, liver, and kidney grafts in these animals. Because such grafts are less tolerogenic in humans, none of these strategies has been effectively applied clinically to date. Therefore, tolerance strategies that are appropriate for clinical evaluation must first be tested in "stringent" models, including relatively nontolerogenic grafts such as MHC-mismatched skin in rodents and vascularized organ graft models in large animals. In most of the models, only a superficial understanding of the mechanisms leading to tolerance is currently available.

The three major mechanisms of T-cell tolerance are deletion, anergy, and suppression (often referred to as "regulation"). *Deletion* denotes the

E-TABLE 49-3 EXPERIMENTAL STRATEGIES TO PREVENT GRAFT-VERSUS-HOST DISEASE

STRATEGY	ADVANTAGES	LIMITATIONS
Donor T-cell T_H2 polarization (e.g., conditioning with ATG and TLI; in vitro stimulation with cytokine exposure)	May preserve GVL	May limit GVL; T_H2 can contribute to acute and chronic GVHD
Tolerance induction of donor T cells (e.g., costimulatory blockade; regulatory cells)	Some strategies may selectively tolerize GVH-reactive T cells (e.g., in vitro antigen exposure with costimulatory blockade)	Global immunosuppression may limit GVL and anti-infectious immunity; tolerance (i.e., GVH protection) may be incomplete
Donor T-cell depletion plus NK-cell infusion with class I mismatched transplantation	NK cells do not cause GVHD but may mediate antitumor effects; donor NK cells may eliminate host APCs that trigger GVHD	Antitumor effect against only certain types of malignancies; requires appropriate MHC disparity and expression of polymorphic NK-cell receptors; insufficient T-cell immunity to infection
Donor T-cell depletion followed by delayed donor lymphocyte infusion (DLI)	Preserves high level of GVL due to GVH reactivity. GVHD does not occur if host inflammation from conditioning has subsided and initial HCT was devoid of donor T cells	Antitumor effect delayed until time of DLI; most applicable for indolent lymphohematopoietic tumors. GVHD more difficult to control in humans than animal models, probably owing to occult or overt infection resulting from T-cell deficiency before DLI
Depletion of donor T cells recognizing host alloantigens by in vitro or in vivo activation/depletion (i.e., "allodepletion")	Preserves anti-infectious immunity and tumor antigen–specific responses while limiting GVHD	Loss of GVH reactivity will limit GVL and engraftment; highly efficient allodepletion methods not yet available. Residual T cells may cause GVHD
Donor T-cell depletion with infusion of expanded infection-specific T cells (e.g., CMV or EBV specific)	Reduces GVHD potential while protecting against significant infectious organisms	Lack of GVL effect; lack of broad anti-infectious immunity; expense and inefficiency of in vitro T-cell expansion; loss of survival/homing potential of cultured T cells
Donor T-cell depletion with infusion of expanded tumor antigen-specific T cells (expanded from natural repertoire or transduced with a T-cell receptor or chimeric antigen receptor)	GVL without GVHD	Lack of anti-infectious immunity; expense and inefficiency of in vitro expansion of tumor-specific T cells; loss of survival/homing potential of cultured T cells
Insertion of suicide gene (e.g., thymidine kinase) into donor T cells	Drug targeting inserted gene (e.g., ganciclovir) kills donor T cells to treat GVHD after GVL initiated.	Expense and inefficiency of in vitro transduction of T cells; loss of function/survival/homing potential of cultured T cells; risk for GVHD if transduction incomplete; curtailment of GVL when donor T cells killed in vivo
Block T-cell trafficking to epithelial GVHD target tissues (e.g., blockade of adhesion molecules or chemokines, sphingosine 1 phosphate agonists)	Permits lymphohematopoietic GVH reactions to occur, with associated GVL effects	Redundancy of trafficking pathways in inflammatory environment may limit efficacy; tumors outside of lymphohematopoietic system not targeted
Block injury/promote repair in epithelial target tissues (e.g., keratinocyte growth factor)	Permits lymphohematopoietic GVH reactions to occur, with associated GVL effects	Efficacy may be limited

APC = antigen-presenting cell; ATG = antithymocyte globulin; CMV = cytomegalovirus; DLI = donor lymphocyte infusion; EBV = Epstein-Barr virus; GVHD = graft-versus-host disease; GVL = graft-versus-leukemia effects; HCT = hematopoietic cell transplantation; MHC = major histocompatibility complex; NK = natural killer; T_H2 = helper T lymphocytes type 2; TLI = total lymphoid irradiation.

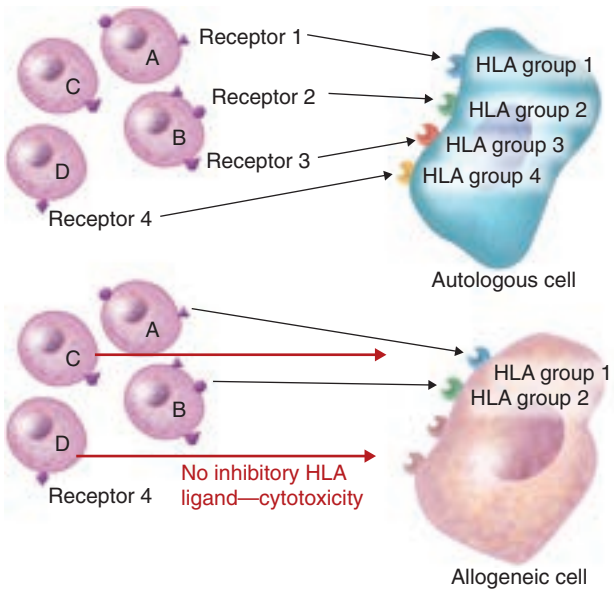

E-FIGURE 49-3. **Killing of allogeneic targets by natural killer (NK) cells due to "missing self." NK cells express clonally distributed inhibitory receptors (KIRs) with specificity for different groups of major histocompatibility complex (MHC) class I alleles, referred to in the figure as human leukocyte antigen (HLA) groups 1, 2, 3, and 4. Four different NK cells (A, B, C, and D) are shown, each with a different set of KIRs (referred to as receptors 1, 2, 3, and 4). Examples of HLA allele groups in the human are the HLA-Cw4, HLA-Cw3, and HLA-Bw4 groups; examples of KIRs are the ligands for these allele groups—namely, KIR2DL1, KIR2DL2/3, and KIR3DL1, respectively. Each functional NK cell has one or more inhibitory receptors that recognize a "self" (autologous) HLA molecule. Although some of the NK cells (e.g., cells A and B in the figure) will also find an HLA ligand to which their receptors bind on allogeneic cells, others (e.g., cells C and D) will not. The latter cells therefore will not receive inhibitory signals from the allogeneic cells and will kill them due to recognition by other (activating) receptors.**

destruction of T cells with receptors that recognize donor antigens; it can be achieved during T-cell development in the thymus, for example, by induction of mixed chimerism in T-cell-depleted hosts. Deletion can also be applied to mature T cells in the periphery, for example, by transplantation of a tolerogenic organ or marrow graft in combination with blockade of costimulatory molecules. *Anergy* denotes the inability of T cells to respond fully to antigens they recognize, and it can be induced by antigen presentation without costimulation. *Suppression* has attracted considerable interest since the discovery that constitutively $CD25^+$ T cells of the $CD4^+$ subset have suppressive activity that is dependent on expression of the transcription factor Forkhead Box Protein 3 (FoxP3). These and other types of suppressive T cells (e.g., NKT cells, regulatory $CD8^+$ cells and B cells, myeloid-derived suppressor cells) have been implicated in rodent transplantation tolerance models and in prevention of autoimmunity. The use of expanded regulatory cells has recently entered clinical trials, and both the ultimate practicality of the approach and the relative advantages of antigen-specific versus nonspecific regulatory cell therapy remain to be determined. There is also interest in strategies for activating or expanding regulatory T cells in vivo, thereby favoring the suppressive immune response over destructive alloimmunity.[12]

The developments in animal models and understanding of immune mechanisms described here have provided impetus for efforts to achieve immune tolerance in clinical transplantation. Every transplantation center has anecdotal cases of patients who have removed themselves from chronic immunosuppression without experiencing graft rejection. However, for every such patient, there are dozens more who have experienced rejection on dose reduction or removal of immunosuppressive drugs. Although trials of minimization and slow withdrawal of nonspecific immunosuppressive therapy are underway in organ transplant recipients, a major current limitation is the absence of good predictors of success. It remains to be seen whether recently identified molecular "tolerance signatures" will provide markers with sufficient predictive value to allow such withdrawal to be safely undertaken.

One approach developed in animal models has been successfully applied to the induction of immune tolerance in a small group of patients receiving renal allografts. This approach, involving bone marrow transplantation after nonmyeloablative conditioning, which is much less toxic than standard HCT conditioning, was shown to be effective in the most stringent rodent and large-animal models before being evaluated clinically. Initial success using combined kidney and bone marrow transplantation in patients with renal failure due to multiple myeloma led to pilot studies in patients with renal failure without malignant disease, with encouraging preliminary results. This approach and others that have emerged from ongoing investigations provide hope that, in the future, transplantation might be routinely performed without the need for chronic immunosuppressive therapy, with its attendant complications and limited ability to control chronic rejection.[11] Because autoimmune diseases are major contributors to end-stage renal disease, diabetes, and other types of organ failure, the potential for tolerance strategies to reverse autoimmunity while inducing allograft tolerance is also a source of hope. All these approaches must, however, be undertaken with the caution that successful regimens could also lead to immune tolerance to active infectious organisms.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Marsh SG, Albert ED, Bodmer WF, et al. Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*. 2010;75:291-455.
2. Susal C, Opelz G. Current role of human leukocyte antigen matching in kidney transplantation. *Curr Opin Organ Transplant*. 2013;18:438-444.
3. Brunstein CG, Fuchs EJ, Carter SL, et al. Alternative donor transplantation after reduced intensity conditioning: results of parallel phase 2 trials using partially HLA-mismatched related bone marrow or unrelated double umbilical cord blood grafts. *Blood*. 2011;118:282-288.
4. Montgomery JR, Berger JC, Warren DS, et al. Outcomes of ABO-incompatible kidney transplantation in the United States. *Transplantation*. 2012;93:603-609.
5. Montgomery RA, Lonze BE, King KE, et al. Desensitization in HLA-incompatible kidney recipients and survival. *N Engl J Med*. 2011;365:318-326.
6. Li HW, Sykes M. Emerging concepts in haematopoietic cell transplantation. *Nat Rev Immunol*. 2012;12:403-416.
7. Blazar BR, Murphy WJ, Abedi M. Advances in graft-versus-host disease biology and therapy. *Nat Rev Immunol*. 2012;12:443-458.
8. Pilat N, Schwarz C, Wekerle T. Modulating T-cell costimulation as new immunosuppressive concept in organ transplantation. *Curr Opin Organ Transplant*. 2012;17:368-375.
9. Ferrer IR, Hester J, Bushell Wood KJ. Induction of transplantation tolerance through regulatory cells: from mice to men. *Immunol Rev*. 2014;258:102-116.
10. Fuchs EJ. Transplantation tolerance: from theory to clinic. *Immunol Rev*. 2014;258:64-79.
11. Zachary AA, Leffell MS. Desensitization for solid organ and hematopoietic stem cell transplantation. *Immunol Rev*. 2014;258:183-207.
12. Issa F, Robb RJ, Wood KJ. The where and when of T cell regulation in transplantation. *Trends Immunol*. 2013;34:107-113.

50 COMPLEMENT SYSTEM IN DISEASE

JOHN P. ATKINSON

The complement system consists of plasma and membrane proteins that participate in host defense against infections and in clearance of cellular and extracellular debris, as well as in a wide variety of autoimmune and inflammatory states (Fig. 50-1).[1,2] Complement is essential in innate immunity and a potent effector arm of adaptive (humoral) immunity. It is a *first* responder, especially in blood, to bacterial and viral invasion (Table 50-1). It helps to maintain sterility ("guardian of the intravascular space") by depositing within seconds its opsonic and membrane-perturbing fragments on a pathogen's surface. A second major activity of complement is to promote the inflammatory response via the release of soluble fragments *(anaphylatoxins)*. They bind to their receptors, leading to cellular activation, including chemokinesis and chemotaxis by phagocytic cells, and thereby enhance protection against infections. Furthermore, the deposition of complement fragments on immune complexes keeps them from precipitating and promotes their adherence to red blood cells (RBCs) for a hand-off to monocytes and dendritic cells in the liver and spleen.

Through these interactions, complement also instructs the adaptive immune response. Antigens decorated by complement proteins are taken up

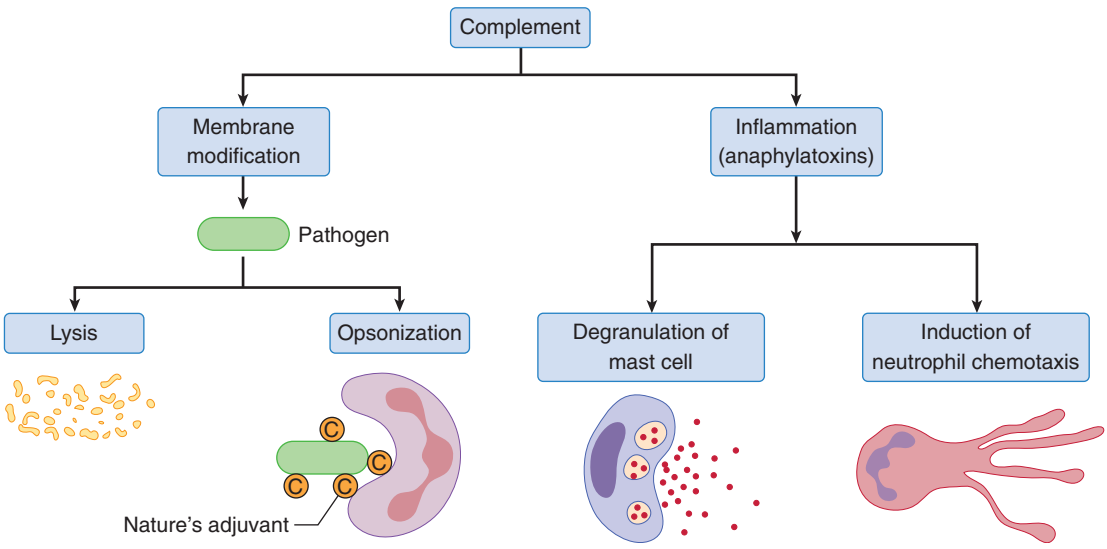

FIGURE 50-1. Function of the complement system. The most important function of the complement system is to alter the membrane of the pathogen by coating its surface with clusters of activation fragments. In one case, they facilitate the key process of opsonization in which C4b and C3b interact with complement receptors. In the other case, as with certain gram-negative bacteria and viruses, the membrane attack complex lyses the organism. The second critical function of complement is to activate cells and thus promote inflammatory and immune responses. The complement fragments C3a and C5a (known as anaphylatoxins) stimulate many cell types such as mast cells to release their contents and stimulate phagocytic cells to migrate to sites of inflammation (chemotaxis). Through these phenomena of opsonization and cell activation, complement serves as nature's adjuvant to prepare, facilitate, and instruct the host's adaptive immune response. Because complement activation occurs in a few seconds, this innate immune system initially engages most pathogens, especially those that try to enter the vascular space. As will be illustrated, these basic functions are also required to handle immune complexes and prevent autoimmunity. (Modified from Arthritis Foundation. *Primer on the Rheumatic Diseases*. 12th ed. Arthritis Foundation; Atlanta, Ga 2001.)

TABLE 50-1 COMPLEMENT SYSTEM IN HOST DEFENSE AGAINST BACTERIA AND VIRUSES

THE ACTIVITY	THE PLAYERS
Opsonization	(C3b > C4b, C1q, MBL)*
Membrane perturbation including lysis (the membrane attack complex)	(C5b-C9)
Proinflammatory via cellular activation (the anaphylatoxins and their receptors)	(C3a, C5a)

*C3b is the major opsonin of the complement system. C1q and MBL (mannose- or mannan-binding lectin) both participate in classical and lectin pathway activation, respectively, but also bind to their specific receptors upon attachment to a target.

TABLE 50-2 SALIENT FEATURES OF THE COMPLEMENT SYSTEM

Ancient innate system of immunity predominantly found in blood (the "guardian of the intravascular space")
Capable of rapidly opsonizing and lysing bacteria and viruses (millions of active fragments can be deposited on a target)
Works in seconds!
Most proteins are synthesized by the liver
Constantly turning over (AP protein C3 "ticks over" at a rate of 1% to 2% per hr)
The AP also features a feedback or amplification loop, which requires tight control
Effector arm of the humoral immune system (IgM and IgG)
Critical for clearance of self-debris (garbage removal)
After immunoglobulins and albumin, complement proteins are among the most abundant in blood
Nature's adjuvant (almost all foreign antigens are coated with complement fragments); instructs the adaptive immune response
A deficiency of an activator leads to bacterial infections or autoimmunity (SLE)
A deficiency of a regulator leads to undesirable cellular and tissue damage at sites of injury or degeneration (excessive activation)

AP = alternative pathway; SLE = systemic lupus erythematosus.

TABLE 50-3 PARTICIPATION OF THE COMPLEMENT SYSTEM IN HUMAN DISEASE

Activation by autoantibody (formation of immune complexes)

Engagement with modified self (clearance of debris or garbage)
- Degenerative processes (diseases of aging such as age-related macular degeneration)
- Cell and tissue damage (ischemia-reperfusion injury; atypical hemolytic uremic syndrome)

TABLE 50-4 PATHOLOGIC CONDITIONS ASSOCIATED WITH COMPLEMENT ACTIVATION

Examples of diseases in which complement activation contributes to the immunopathology:
- Atypical hemolytic uremic syndrome*†
- Paroxysmal nocturnal hemoglobinuria†
- Age-related macular degeneration*†
- Membranoproliferative glomerulonephritis (types 1, 2 and 3)†‡
- Myasthenia gravis‡
- Bullous pemphigoid‡
- Systemic lupus erythematosus/antiphospholipid syndrome‡
- Rheumatoid arthritis‡
- Immune hemolytic anemias‡
- Immune vasculitis (the ANCA-positive syndromes)‡
- Ischemia reperfusion injury*†
- Allotransplantation‡
- Serum sickness‡
- Exposures to foreign materials (e.g., membranes, nanoparticles)*

*Injury, ischemia, trauma, degeneration, or foreign body is the trigger (innate immune activation).
†Lack of adequate regulation contributes to disease pathogenesis.
‡Antibody dependent activation of the complement system (adaptive humoral immune activation).

by monocytes, follicular-dendritic cells, B lymphocytes, and other antigen-presenting cells, resulting in an adaptive immune response. (The complement system is often called "nature's adjuvant.") Thus, complement activation is required for an optimal antibody response to most foreign antigens. Individuals lacking a functional complement system are predisposed to bacterial infections, predominantly by encapsulated organisms, including streptococcus, staphylococcus, *Haemophilus* spp., and *Neisseria* spp.[3] Surprisingly, a complete deficiency in an early component of the classical complement pathway predisposes to autoimmune diseases, particularly systemic lupus erythematosus (SLE).[4] This association suggests that complement is required not only for host defense against foreign agents but also to identify and safely clear self-materials (debris removal), particularly RNA and DNA species.

A remarkable feature of the complement system is that it reacts within seconds (Table 50-2). In less than 2 minutes, it can coat an encapsulated gram-positive bacterium with several million C3b opsonic fragments and lyse gram-negative bacteria by insertion of its terminal components (the membrane attack complex [MAC]). It works even more efficiently if driven by IgM or IgG binding to an antigen on a microbial membrane to activate the cascade. Antibodies and lectins direct the activation process to the pathogen's surface. Overall, the complement cascade is designed to become engaged on the surface of a pathogen, particularly bacteria. Plasma and membrane *regulators of complement activation* inhibit formation on normal "self" cells.

Much of the complement-mediated pathology revolves around the alternative pathway's (AP's) amplification loop. This feedback amplification loop is key in triggering activation early in an immune response; however, it must be rigorously regulated to prevent activation on normal self and excessive activation on injured self.[5] Approximately half of the proteins associated with the complement system are dedicated to the control of its activation and effector functions, especially to maintain homeostasis of the AP's amplification loop.

In clinical medicine (Table 50-3), the complement system participates in three pathologic processes (Table 50-4): (1) an inherited decrease in functional activity leading to increased susceptibility to bacterial infections and to autoimmunity, (2) mediating undesirable tissue damage upon activation by autoantibodies and immune complexes, and (3) excessive activation at sites of tissue injury in individuals carrying genetic variants in regulators.

Knowledge of how complement is activated and how it can be controlled points to opportunities for the development of therapeutic agents such as anti-C5 monoclonal antibody (mAb) therapy, which has been recently approved to treat several complement-dependent hemolytic disorders.

ACTIVATION OF COMPLEMENT

Classical Pathway

The binding of IgM or IgG to a target antigen activates this exceptionally powerful and quick acting pathway to destroy microbes (Figs. 50-2 and 50-3). The classical pathway (CP) reaction cascade is designed to opsonize and perturb the surface membrane of microorganisms. Of course, autoantibodies also trigger this highly efficient CP. Complement action mediated by immune complexes may then lead to cellular and tissue damage. Instructive examples of autoantibodies and complement-mediated diseases are immune hemolytic anemias, myasthenia gravis, and bullous pemphigoid. The basic problem or pathologic defect in this type of human disease is, of course, the formation of the autoantibody. A misidentification of self that has occurred because of a breaking of tolerance. In this pathologic situation, the complement system is working at the behest of the autoantibody.

The CP is also activated by means other than the formation of IgM- and IgG-bearing immune complexes. β-Amyloid in the neuritic plaques of patients with Alzheimer disease directly engages the CP via an interaction with C1q. Likewise, C-reactive protein (CRP) and serum amyloid protein (SAP) bind to chromatin and other ribonucleoprotein complexes released from apoptotic cells, and these types of complexes activate the CP. As noted, the CP plays a key role in the opsonization and removal of nuclear debris. Approximately 80% of patients with hereditary absence of C1q or C4 develop SLE. Deposits of CRP and activated C1 have been demonstrated in ischemic tissue such as infarcted human myocardium. These observations indicate that CP activation via these antibody-independent means is critical in protecting against autoimmune responses by facilitating debris clearance.

Regulation of the CP activation occurs at two levels. First, the *ser*ine *p*rotease *in*hibitor (serpin) known as the C1-inhibitor (C1-INH) blocks the activity of many proteases, including factor XIIa, kallikrein, and factor XIa of

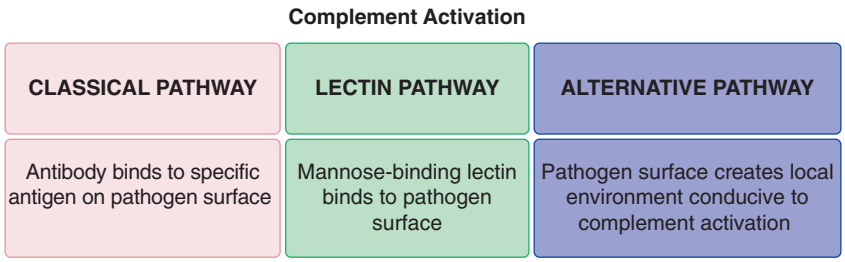

FIGURE 50-2. The three pathways of complement activation.

TABLE 50-5 TISSUE INJURY OR DEGENERATION AND COMPLEMENT ACTIVATION*

Age-related macular degeneration
Osteoarthritis (degenerative joint disease)
Ischemic stroke
Myocardial infarction
Traumatic brain injury (e.g., liver, kidney, gut)
Ischemia-reperfusion injury
Burns
Acute respiratory distress syndrome
Septic shock
Multiorgan failure syndromes
Alzheimer disease

*In these conditions, complement activation leads to deposition of fragments at the site of injury; however, how much of the tissue injury is attributable to complement system is unknown. In many cases, animal models support a pathologic role for the complement system. Only in age-related macular degeneration do we also have powerful genetic evidence in humans to indicate a key role for the complement system.

the clotting system as well as C1r, C1s, and MASP2 of the complement system. The importance of C1-INH is exemplified by its role in hereditary angioedema (Table 50-6). In this dominantly inherited disease, a deficiency of C1-INH allows uncontrolled proteolysis of C4 and C2 and generation of bradykinin, leading to recurrent swelling episodes. This serpin prevents chronic activation of the CP cascade and, after a few minutes, helps to shut down the system. CP activation is also regulated by multiple inhibitors at the key step of C3 activation. These plasma and membrane proteins inhibit C3 convertase formation on healthy self. Membrane regulators are highly expressed on most cell types, where they prevent activation on normal self and overexuberant activation on altered and nonself.

Lectin Pathway

The protein mannose-binding lectin (MBL) is a member of the collectin family that also includes pulmonary surfactants A and D.[6] MBL has a structure similar to C1q in that it consists of several subunits; namely, a globular recognition head domain for carbohydrates and a collagen-like tail that interacts with serine proteases. In the case of MBL, the globular domain is a lectin (protein) that binds to repeating mannose and *N*-acetylglucosamine residues on the surface of pathogens (see Figs. 50-2 and 50-3). Many microorganisms are recognized by MBL, including gram-positive and gram-negative bacteria, mycobacteria, fungi, parasites, and viruses (including human immunodeficiency virus 1 [HIV-1]). In general, as would be expected, mammalian glycoproteins and glycolipids are not readily recognized by MBL and the related lectins (ficolins and collectins) that activate the lectin pathway.

Three serine proteases, MASP-1, MASP-2, and MASP-3, associate with MBL (and the ficolins and collectins) through their collagen-like domain. This is analogous to the association of C1r and C1s with C1q. Activation of MASP-2, with some help from MASP-1, results in cleavage of C2 and C4, leading to formation of the classical/lectin pathway C3 convertase (C4b2a).

Genetic variations in the structural and regulatory portions of the MBL gene lead to wide differences in serum levels. A low level of MBL is associated with recurrent infections in children and adults and is a risk factor for the development of SLE. More striking is the association of low levels of MBL with infections in the setting of the treatment of SLE. For example, heterozygous MBL deficiency has been associated with a fourfold increase in the risk of bacterial pneumonia and homozygous deficiency with a more than 100-fold increase.

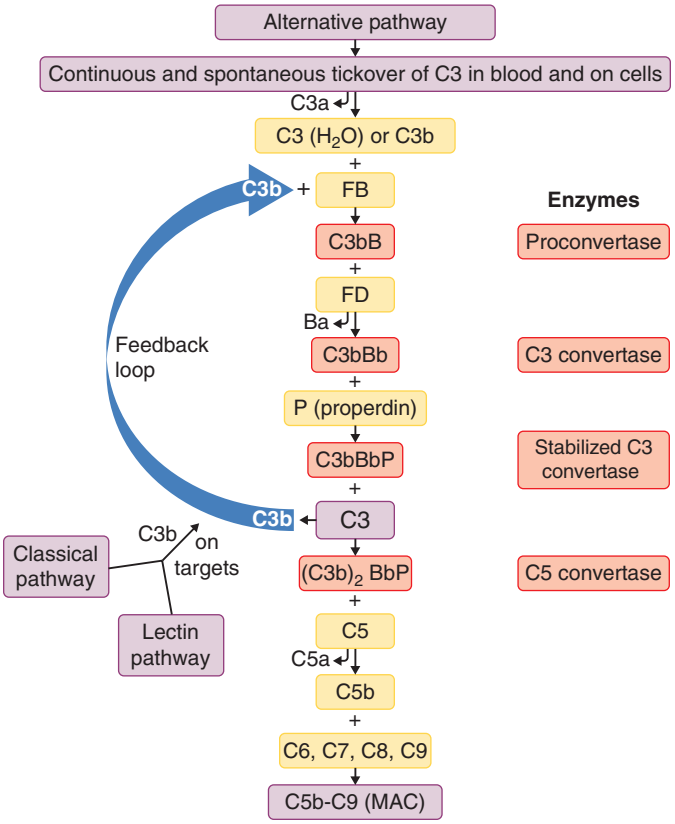

FIGURE 50-3. Complement activation pathways. In the reaction cascade shown, C3b or C3 (H_2O) binds the proenzyme factor B (FB), and the C36B complex then cleaved by the protease factor D (FD). The addition of properdin (P) to the enzyme complex increases the half-life of the enzyme complex approximately 10-fold. Although the source of the C3b can be from spontaneous turnover or via lectin pathway (LP) and classical pathway (CP) activation, the alternate pathway (AP) feedback loop commonly takes over to generate most of the C3b that binds to a target. The alternate pathway is continuously turning over. If activated C3b or C3 (H_2O) remains in the fluid phase, it is rapidly inhibited by the plasma regulator factor H. if activated C3 binds to normal or healthy self, it is prevented from forming a convertase by the ubiquitously expresses membrane cofactor protein (MCP [CD46]) and decay-accelerating factor (DAF [CD55]). DAF "kicks out" the catalytic Bb domain (a temporary stop), but MCP is a permanent stop because, upon its binding, the C3b is proteolytically cleaved to inactive C3b (iC3b) by a serine protease known as factor I. The feedback loop is a powerful amplification system. A single *Escherichia coli* organism in blood can be coated with several million C3bs in a couple of minutes!

Alternative Pathway

The AP takes advantage of the fact that C3 undergoes spontaneous, chronic, low-grade activation (Figs. 50-2 to 50-4). This C3b may covalently attach to any cell; however, on normal self, amplification of the cascade is blocked by inhibitors. In contrast, deposition on polysaccharides of bacterial membranes and to other targets, such as endotoxin and virally infected cells, leads to a rapid engagement of this pathway. These sites, similar to immune complexes and almost any type of biomaterial (cardiopulmonary bypass and hemodialysis membranes, nanoparticles, and so on), lack regulators, so rapid, massive activation may occur.

During spontaneous activation, called *tickover*, small amounts of activated C3 are continuously generated (C3 turns over in blood at 1% to 2%/hr). It can initiate a feedback loop and cleave more C3 to C3b. Also, the initial C3b

TABLE 50-6 SOLUBLE AND MEMBRANE FACTORS REGULATING COMPLEMENT

SOLUBLE FACTORS REGULATING COMPLEMENT			
NAME	**LIGAND OR BINDING FACTOR**	**FUNCTIONAL ACTIVITY**	**PATHOLOGY, IF DEFICIENT**
C1-INH	C1r, C1s, MASP-2	Binds to and displaces C1r and C1s from C1q and MASP-2 from MBL	HAE
C4bp* (C4 binding protein)	C4b, GAGs	Displaces C2a (DAA); cofactor for C4b cleavage by factor I (CA)	No clinical syndrome clearly defined
CPN-1 (carboxypeptidase-N)	C3a, C5a	Inactivates C3a and C5a	Urticaria and angioedema
Factor H*†	C3b, C3d, GAGs	Displaces Bb from AP C3 and C5 convertases (DAA) and is a cofactor for factor I to cleave C3b (CA)	AMD, aHUS, C3 glomerulopathies; bacterial infections secondary to low C3
Factor I†	C3b, C4b	Serine protease; cleaves C3b and C4b, requires a cofactor protein (CA)	AMD, aHUS; bacterial infections secondary to low C3
Protein S (vitronectin)	C5b67	Inhibits membrane attachment by C5b67	None defined
MEMBRANE-BOUND FACTORS REGULATING COMPLEMENT			
NAME	**LIGAND OR BINDING FACTOR**	**FUNCTIONAL ACTIVITY**	**DISEASE, IF DEFICIENT**
DAF (CD55)	C3 and C5 convertases	Displaces Bb from AP convertase and C2a from CP or LP convertases, respectively	PNH
Membrane cofactor protein (MCP, CD46)	C3b, C4b	Cofactor for factor I (CA)	aHUS
Protectin (CD59)	C8, C9	Inhibits MAC formation or insertion	PNH
CR1 (CD35) (immune adherence or C4b/C3b receptor)	C3b, C4b, C3, and C5 convertases	Cofactor for factor I to cleave C4b and C3b (CA); displaces Bb from C3b and C2a from C4b to inhibit convertases (DAA)	No complete deficiency described; decreased levels in immune complex–mediated diseases such as lupus
CRIg	C3b, iC3b, C3c	Inhibits activation of AP	None defined

*Factor H and C4bp also bind to surfaces, particularly at sites of tissue and cellular injury, where they also carry out regulatory activity.
†If heterozygous deficient, individual is predisposed to AMD and aHUS. If homozygous deficient, the AP turns over excessively, resulting in kidney disease (C3 glomerulopathies) and bacterial infections (secondary to the very low C3).
aHUS, atypical hemolytic uremic syndrome; AMD, age-related macular degeneration; AP, alternative pathway; $(C3b)_2$ Bb, alternative pathway C5 convertase; C3bC4bC2a, classical and lectin pathway C5 convertase; C4bC2a, classical and lectin pathway C3 convertase; CA, cofactor activity; CR1, complement receptor type 1; CRIg, complement receptor of the Ig superfamily; DAA, decay-accelerating activity; GAG, glycosaminoglycan; HAE, hereditary angioedema; MAC, membrane attack complex; MASP, mannan-binding lectin-associated serine protease; MBL, mannan or mannose binding lectin; PNH, paroxysmal nocturnal hemoglobinuria.

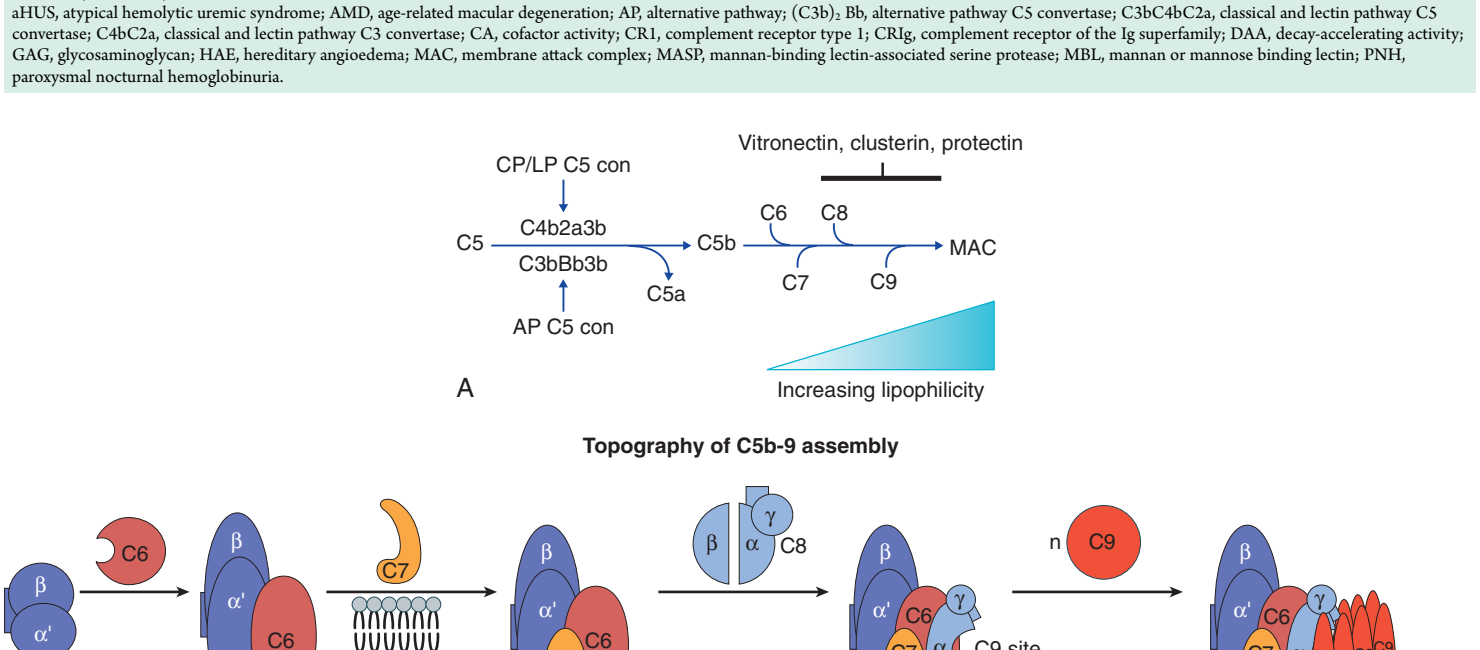

FIGURE 50-4. Activation of C5 and the membrane attack complex (MAC). A, The C5 convertases ("con") are the same as C3 convertases except a C3b has been attached to C4bC2a or a second C3b in the case of AP C5 convertase. **B,** Schematic representation of the assembly of the assembly of the MAC on a cell membrane. C5b (composed of two chains) binds C6 and then C7. The C5b-7 complex can insert into a membrane and then bind C8 (composed of three chains) and multiple C9s to form a pore or channel in the membrane. (Modified from Liszewski MK, et al. *The Human Complement System in Health and Disease*. Marcel Dekker; New York, NY 1998.)

may be derived from either the classical or lectin pathway. Thus, activation of complement by any one of the three pathways has the potential to be rapidly magnified.[7] The AP C3 convertase is negatively controlled (to maintain homeostasis) both in the fluid phase and on host cells by two abundant plasma proteins and two widely expressed membrane proteins.[8]

The central role of the AP as an amplifier of complement activation is borne out by its association with a number of clinicopathologic states in the setting of deficient regulation (Tables 50-6 and 50-7). For example, multiple forms of membranoproliferative glomerulonephritis are associated with excessive C3 fragment deposition in the kidney because of either the

TABLE 50-7 COMPLEMENT SYSTEM MEDIATES THE DISEASE PROCESS AND ITS INHIBITION TREATS THE CONDITION

DISEASE	PATHOPHYSIOLOGY	ETIOLOGY	TREATMENT	FDA APPROVED
PNH	Lyse RBCs	Acquired hemopoietic somatic stem cell mutation in gene required for synthesis of GPI anchor	mAb to C5	Yes
aHUS	Damage to endothelial cells	Inherited loss of function variants in AP regulators or gain of function variants in AP activators	mAb to C5	Yes
HAE	Bradykinin generation	Autosomal dominant variants in the C1-inhibitor gene	C1 inhibitor replacement Bradykinin receptor blockage Kallikrein inhibitor	Yes Yes Yes
AMD	Degeneration of the retina	Inherited variants in a regulator (FH or FI) or gain of function in an alternative pathway component (C3 or FB)	Clinical trials in progress	No

aHUS= atypical hemolytic uremic syndrome; AMD = age-related macular degeneration; FDA = Food and Drug Administration; HAE= hereditary angioedema; GPI = glycosyl phosphatidylinositol; mAb = monoclonal antibody; PNH= paroxysmal nocturnal hemoglobinuria; RBC = red blood cell.

presence of autoantibodies (C3 or C4 nephritic factors) that stabilize C3 convertases or a genetic deficiency in complement regulatory protein (factor H or factor I).[9] Likewise, atypical hemolytic uremic syndrome (i.e., not associated with a preceding enteropathic infection featuring a *Shiga*-like toxin) occurs in individuals who harbor heterozygous missense mutations in factor H or I or have gain-of-function mutations in factor B or C3.[10] Genome-wide association and targeted deep sequencing studies have also linked age-related macular degeneration (AMD) to functional coding mutations in factor H and factor I and more uncommonly in factor B and C3.[11] Finally, rodent models of rheumatoid arthritis, SLE, and ANCA-positive vasculitic syndromes are ameliorated if the AP is disrupted.

C3 and C5 Convertases

The three activation pathways converge at C3. A remarkable feature of C3 is the presence of a thioester bond. Buried within the three-dimensional structure of the C3 protein lies a γ-carboxy group of a reactive glutamic acid residue linked to a cysteine in an "internal thioester." Upon its cleavage, for a few microseconds, a covalent attachment can occur via an ester or amide linkage to any nearby hydroxyl or amino group. Most of the cleaved thioester bonds are hydrolyzed by water to produce a form of C3 (known as C3 [H_2O]); however, a substantial percentage forms an amide or ester bond to amino groups or carbohydrates thereby covalently attaching C3b to a target's surface. Also, the addition of a C3b to C4b2a (classical/lectin pathway C3 convertase) or to C3bBb (AP C3 convertase) then forms a convertase for C5 (C3bBbC3b for the AP and C4bC2aC3b for the CP/LP).

Regulators of Complement Activation at the C3 and C5 steps

The regulators of complement activation (RCA) (see Table 50-2) limit the production of C3b, primarily by the AP C3 convertases. Because the addition of C3b to a C3 convertase makes it a C5 convertase, regulation of the two enzyme complexes is linked. Modulation of their activity on host cells limits tissue destruction and the production of inflammatory mediators.

The RCA proteins control complement activation by two processes. *Decay-accelerating activity* refers to when the inhibitor transiently binds to C3b or C4b in the convertase and thereby dissociates the other members of the complex, rendering it enzymatically inactive (as the component released is the catalytic domain of the protease). The second is cofactor activity, which requires recognition of C3b or C4b by a plasma cofactor protein. Upon this interaction, the protease, factor I, cleaves C3b or C4b. Cleavage of C3b by factor I renders the convertase irreversibly inactive (generates iC3b which cannot participate in convertase formation).

Membrane Attack Complex

The cleavage of C5 generates C5a, the most potent of the complement anaphylatoxins, and C5b. C5b associates with C6 and C7 to create a lipophilic trimer as the initial part of the MAC (Fig. 50-4). The C5b67 trimer inserts into the lipid bilayer and serves as a binding site for C8 and C9. C9 self-polymerizes, leading to 12 to 18 C9 molecules that form a ring structure (completing the MAC). The MAC resembles a doughnut with a 10-nm pore running through the center. This pore allows water and ions to enter the cells, ultimately leading to osmotic lysis. Many pathogens such as gram-positive bacteria possess a capsule that makes them resistant to lysis.[12] Opsonization leading to phagocytosis is thus the major means of eliminating such organisms.

TABLE 50-8 DISTRIBUTION OF ANAPHYLATOXIN RECEPTORS AND THEIR CELLULAR RESPONSES

CELL TYPE	RESPONSES
C5aR (CD88)	
Neutrophils	Chemotaxis
Eosinophils	Enzyme release
Basophils	Generation of reactive oxygen species
Mast cells	Upregulation of adhesion molecules
Monocytes	Increased synthesis of IL-1, IL-6, and IL-8 Prostaglandin and leukotriene synthesis
Hepatocytes	Increased synthesis of acute phase reactants
Pulmonary epithelium	Increased IL-8
Neuronal cells	Cellular activation
Endothelial cells	Increased expression of P-selectin
Renal epithelial/mesangial cells	Proliferation Synthesis of growth factors
C3aR	
Eosinophils	Chemotaxis
Mast cells	Enzyme release
Platelets	Generation of reactive oxygen species Upregulation of adhesion molecules
Epithelial, endothelial, etc.	Cellular activation

CNS = central nervous system; IL = interleukin.

The MAC appears to be essential only for elimination of *Neisseria* spp. Individuals completely deficient in C5, C6, C7, C8, or C9 are at an increased risk only for meningococcal and gonococcal infections. C9 deficiency is a common immunodeficiency in Japan, with a heterozygote frequency of 3% to 5%. Thus, heterozygous deficiency seems to not be deleterious to the population in general but may have a selective advantage.

Extensive complement activation during an inflammatory response can result in sufficient MAC deposition to produce host cell lysis. Host cells, however, have mechanisms in place to resist the osmotic changes caused by the MAC and to block assembly of the MAC as it is formed (the protein is known as protectin or CD59). Rather, the nonlethal effects of sublytic MAC deposition are more likely to contribute to pathology. In most cells, this occurs through a general activation of multiple cell signaling pathways.

The response to MAC deposition at sites of complement activation depends on the cell type (Table 50-8). In phagocytic cells, such as neutrophils or macrophages, sublytic MAC insertion leads to the production of reactive oxygen species (e.g., superoxide, hydrogen peroxide) as well as release of prostaglandins and leukotrienes. Platelets undergoing a "MAC attack" incorporate phosphatidylserine on their outer membrane, facilitating formation of blood coagulation enzyme complexes with a potentially procoagulant effect. On endothelial cells, MAC deposition induces the synthesis of interleukin-1α (IL-1α), which leads to further autocrine and paracrine endothelial cell activation. It stimulates a procoagulant state by (1) altering the phospholipid composition of the endothelial membrane; (2) inducing the synthesis of tissue factor and upregulating the synthesis of plasminogen activator inhibitor; (3) upregulating the expression of adhesion molecules, including intercellular adhesion molecule 1 (ICAM-1) and E-selectin; and (4) stimulating endothelial cells to proliferate through growth factor production. In summary,

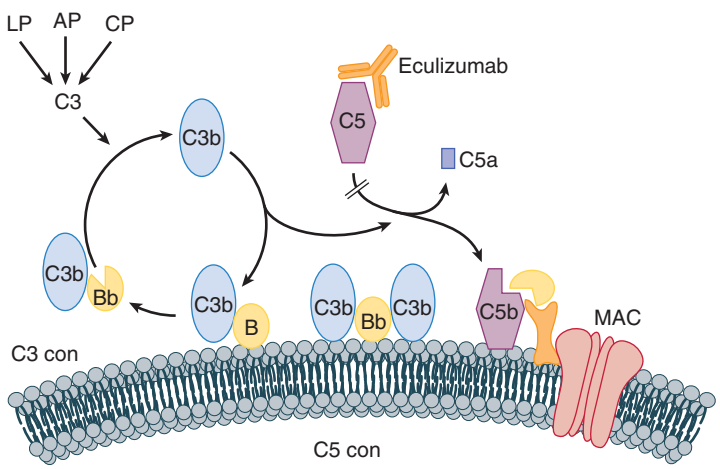

FIGURE 50-5. Complement activation and the mechanism of action of eculizumab (monoclonal antibody [mAb] to C5). The alternate pathway (AP) constantly undergoes "tickover" but can also be primed by the classical pathway (CP) and lectin pathway (LP). The C3b that is formed interacts with factor B (B), which is then cleaved by factor D to form a C3 convertase (C3Bb). As more C3b is generated, some binds to the C3 convertase to form a C5 convertase. mAb to C5 (eculizumab) prevents the cleavage of C5 by the C5 convertase. Not shown is properdin that binds to both the C3 and C5 convertases to increase their half-lives approximately five- to 10-fold (from ≈30 seconds to several minutes).[7] (Modified from Wong EK, Goodship TH, Kavanagh D. Complement therapy in atypical haemolytic uraemic syndrome (aHUS). *Mol Immunol*. 2013;56(3):199-212.)

although cell death does not usually occur, deposition of the sublytic levels of MAC leads to a potentially dangerous situation, with increased inflammation, a procoagulant state, and cellular proliferation. Of course, part of this response is necessary at sites of injury to eliminate pathogens and debris and to facilitate wound repair. The short duration of complement activation and the presence of inhibitors help to maintain homeostasis.

Regulation of MAC formation is important clinically (Fig. 50-5). Two plasma proteins, clusterin and S-protein (vitronectin), bind the C5b-7 complex and prevent its association with the lipid membrane. C8 and multiple C9 molecules adhere to this soluble complex, termed *soluble C5b-9*, which is lytically inactive. CD59 (protectin) is a membrane-bound inhibitor of MAC formation. This small glycoprotein is attached to the cell membrane through a glycosyl phosphatidylinositol tail (GPI anchor). It binds to C5b-8, inserted in the cell membrane, to prevent binding to and polymerization of C9. The expression of CD59 is defective in patients with paroxysmal nocturnal hemoglobinuria (PNH), owing to the failure to synthesize the GPI anchor used by this and many other membrane proteins (including decay-accelerating factor [DAF]) to insert on the cell. The clinical features of PNH are primarily chronic hemolysis and intermittent thrombosis. Hemolysis is caused by complement activation on RBCs because of a lack of DAF and particularly CD59. Thrombosis is likely secondary to intravascular complement activation, leading to endothelial cell activation. The primary defect is an acquired hematopoietic stem cell mutation of a gene on the X chromosome responsible for encoding the first enzyme in the pathway to synthesize a GPI-anchor.

Anaphylatoxins

The anaphylatoxins serve a key early role in initiating a local inflammatory response as they trigger pathways to prepare a cell to face a pathogen or injury (Table 50-9). Similar to the MAC, anaphylatoxins are another major source of potential pathologic damage to self that results from complement activation. These peptides, C3a and C5a, are cleaved from their respective proteins during complement activation. They were named in 1910 to describe their toxic effects, including shock after the transfer of complement-activated serum into laboratory animals. They are 77 (C3a) or 74 (C5a) amino acids long and contain a key carboxy (C)-terminal arginine. They interact with the anaphylatoxin receptors. In plasma, the C-terminal arginine is removed by carboxypeptidase-N from anaphylatoxins not bound to their receptors. Depending on the response studied, this removal totally inactivates the anaphylatoxin or reduces its potency by about 1000-fold.

The C5a receptor (C5aR [CD88]) is a seven-transmembrane-spanning protein that couples ligand binding to G-protein signaling. Expressed on myeloid cells, particularly neutrophils and eosinophils, it mediates the potent chemoattractant property of C5a for both of these cell types. Signaling through CD88 leads to rapid secretion of all granule contents. These include lipases and proteases as well as lactoferrin from neutrophils, and peroxidase, major basic protein, and cationic protein from eosinophils. C5a also induces the release of cytokines, such as tumor necrosis factor (TNF), IL-1, IL-6, IL-8, and adhesion molecules, promoting the inflammatory response. The C5aR is expressed by numerous other tissues, including hepatocytes, bronchial and alveolar epithelium, vascular endothelium, renal mesangial and tubular epithelial cells, and brain neuronal cells. These cells are activated by receptor engagement, leading to production and release of cytokines, chemokines, and prostaglandins and to cellular proliferation.

TABLE 50-9 RESPONSES TO SUBLYTIC MEMBRANE ATTACK COMPLEX ACTIVATION

CELL TYPE	EFFECTS
Most cells	Increased intracellular calcium flux Activation of G proteins Activation of protein kinases Activation of transcription factors Proliferation
Neutrophils and macrophages	Release of reactive oxygen species Activation of phospholipase A_2 Release of prostaglandins, thromboxane, and leukotrienes
Platelets	Release of ATP Increased P-selectin expression Procoagulant membrane changes
Endothelial cells	Increased synthesis of IL-1α Increased release of tissue factor Increased release of von Willebrand factor Increased synthesis of basic fibroblast and platelet-derived growth factors
Synoviocytes	Increased synthesis of prostaglandin Increased synthesis of IL-6 Increased production of matrix metalloproteinase
Glomerular epithelium	Activation of phospholipase A_2 Synthesis of prostaglandin Increased synthesis of collagen and fibronectin
Oligodendrocytes	Increased synthesis of myelin basic protein and proteolipids Increased proliferation

ATP = adenosine triphosphate; IL = interleukin.

The C3a receptor is also a seven-transmembrane-domain protein. It is expressed on almost all myeloid cells, including mast cells, where it mediates the release of allergic mediators. The C3aR also has been detected on many tissues, including in the brain and lung.

The anaphylatoxins have multiple biologic effects. In general, they cause smooth muscle contraction and recruitment of granulocytes, monocytes, and mast cells. In theory, they can contribute to the pathophysiology of any inflammatory condition. In disease models, C3a and C5a have been shown to play a role in diseases such as acute respiratory distress syndrome (ARDS), multisystem organ failure, septic shock, myocardial ischemia-reperfusion injury, asthma, rheumatoid arthritis, SLE, and inflammatory bowel disease. The anaphylatoxin peptides also are responsible for the "postpump" syndrome seen in patients undergoing cardiopulmonary bypass or hemodialysis. Exposure of blood to dialysis or perfusion membranes leads to complement activation. Within minutes of starting bypass, there is a sharp increase in the levels of C3a and C5a in the extracorporeal circuit being returned to the patient. This increase can be associated with respiratory distress, pulmonary hypertension, and pulmonary edema. It has been shown that the length of time that patients stay on the ventilator after bypass surgery correlates with the level of C3a generated during reperfusion.

C3a and C5a have been implicated in the initiation and prolongation of ARDS and multisystem organ failure. After severe trauma, levels of C3a have been measured that suggest activation of the entire circulating C3 pool. This activation leads to bronchoconstriction, increased vascular permeability, hypotension, and vascular plugging with leukocytes. The activation of white blood cells continues the cycle of tissue damage with further complement activation. Continued elevation of C3a in shock or ARDS is a poor prognostic sign. C3a and C5a also appear to play a major role in the pathogenesis of asthma.

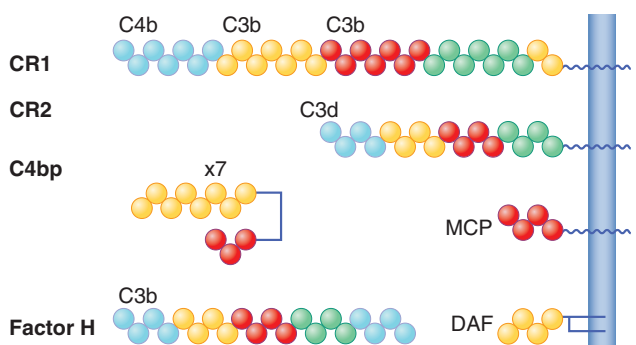

FIGURE 50-6. Model of the complement regulators required to inhibit complement activation on self at the steps of C3 and C5 cleavage. *Circles* represent individual complement control repeats (~60 amino acids each), and *shading* indicates higher organizational units composed of several repeats. The approximate locations for binding of C3b and C4b fragments are indicated. Factor H and C4bp are plasma proteins. C4bp = C4-binding protein; CR = complement receptor; DAF = decay-accelerating factor; MCP = membrane cofactor protein.

Complement Receptors

Opsonization of target by C4b and C3b is effective in preventing infections because these two complement fragments (and the further cleavage products in the case of C3b) are ligands for complement receptors (Fig. 50-6). After covalent attachment of C4b and C3b, immune adherence occurs between the opsonized microbe and immune cells, predominantly neutrophils, monocytes, and macrophages. Complement opsonins are highly effective mediators of immune adherence. On phagocytic cells, this is the prelude to the ingestion and destruction of the target antigen. On RBCs, immune adherence is followed by transfer of the C4b/C3b-coated cargo to monocytes and macrophages in the liver and spleen. CR1 is particularly efficient at immune adherence. Proteolytic modification of C3b leads to iC3b, which is a ligand for the highly phagocytic CR3 and CR4. A further degradation of iC3b to C3d leads to an interaction with CR2 to lower the threshold of B cell activation. Overall, the process is designed with two goals in mind: first is destruction by phagocytosis of the microbe and second is to coat microbial antigens for an adaptive immune response. For example, follicular dendritic cells and B lymphocytes express CR1, CR2, CR3, and CR4 that facilitate complement-coated antigens to be bound, internalized, and presented to other immune cells. CR3 and CR4 facilitate phagocytosis, and CR2 on follicular dendritic cells facilitates immunologic memory generation.

COMPLEMENT INHIBITORS

Given the many disease states in which complement is one of the central mediators of pathology, it is no surprise that complement inhibitors are in preclinical or clinical development for treatment of human diseases (see Table 50-2 and Fig. 50-6). These inhibitors take several different forms. Whereas some are variations of physiologic inhibitors, others are the products of molecular biologic searches for novel compounds.

It is important to consider where in the complement pathway to design an inhibitor to act. Inhibition of the activation pathways limits the production of biologically active peptides. Inhibiting the activation of C3 not only prevents the generation of the C3a anaphylatoxin but also may leave the patient susceptible to infection by limiting the deposition of C3b on targets as an opsonin. Inhibition of C3b deposition would decrease the patient's ability to clear immune complexes, potentially resulting in renal, pulmonary, and vascular damage. It also might promote the development of antibodies to self-antigens.

Inhibition of the C5 convertases is an attractive goal because it would prevent the generation of the C5a anaphylatoxin and the MAC (see Fig. 50-5). This strategy would inhibit complement activation without limiting C3b deposition. Inhibitors based on this concept have been successful; the mAb to C5 is approved by the Food and Drug Administration (FDA) to treat PNH and atypical hemolytic-uremic syndrome (aHUS).

Other concerns about complement inhibition include whether it is short- or long term and whether it is systemic or local. Long-term inhibition of complement, particularly at one of the early steps, is likely to predispose to infection and possibly autoimmunity. Short-term (hours to days) inhibition at any step is unlikely to cause problems. Given that inflammation is usually a local phenomenon, several mechanisms are being tested to target complement inhibitors to these sites. In this way, higher levels of inhibition can be achieved when needed and with lower doses of inhibitor.

Natural Complement Inhibitors

Naturally occurring compounds that control complement activation include products or extracts of plants, fungi, insects, bacteria, viruses, and venoms.[12] The mechanism of complement inhibition by some of these natural products is known and is of clinical and experimental importance. In particular, to protect themselves from the host's complement system, poxviruses, herpesviruses, and flaviviruses produce either mimics of the human regulators that they at one time hijacked from their hosts or proteins that bind the hosts' regulators such as the plasma protein factor H. Bacteria also express a wide variety of inhibitors of the human complement system. *Staphylococcus aureus,* for example, synthesizes up to 10 distinct proteins that inhibit at almost every key step of the complement cascade. Cobra venom factor (CVF) is a modified form of cobra C3b secreted by venom glands in the oral cavity. It is an 144,000-dalton glycoprotein that forms an AP convertase in association with host factor B. Upon injection, CVF leads to massive activation of the AP, leading to shock and pulmonary microvascular injury in experimental animals. It is resistant to the host's inhibitors because the site for that interaction is altered on CVF. Perhaps the most widely used natural inhibitor of complement activation is heparin. It decreases activation of the CP and AP. In clinical practice, the anticomplementary effect of heparin has been used to prevent complement activation during cardiopulmonary bypass. Measurement of complement activation products such as C3a or soluble C5b-9 after bypass showed decreases of 35% to 70% for adult and pediatric patients when heparin-coated extracorporeal circuits were used. Although numerous studies have looked at the decrease in complement activation by heparin-coated bypass circuits, there have been few attempts to correlate this with clinical outcome.

Anti-C5

The complement inhibitor that has achieved the widest attention as a therapeutic agent to stop complement activation is a mAb to C5 that prevents its cleavage to C5a (potent anaphylatoxin) and C5b (initiator of the MAC) (see Fig. 50-5). The generation of the C3b and C4b still occurs, allowing opsonization of pathogens and formation of immune complexes. Because activation of early complement components is also important for the maintenance of tolerance to self-antigens, inhibition of C5 activation is less worrisome than inhibition of C3 activation. The one consequence of C5 deficiency in humans is an increased risk of *Neisseria* infections that can largely be mitigated through vaccination. The anti-C5 mAb eculizumab has been approved for use in patients with PNH and aHUS.

The complement system is undergoing a renaissance. There are several reasons but probably the foremost is the discovery of mutations in complement regulators leading to aHUS and AMD. Second is the therapeutic success of a mAb to C5 in the treatment of aHUS and PNH. Third is the introduction of purified C1-Inhibitor, kallikrein inhibitors, and a bradykinin receptor antagonist to prevent and treat swelling attacks in hereditary angioedema. Last, intriguing recent data implicate the complement system in the pathophysiology of multiple disorders, including AMD, ischemia/reperfusion injury, organ regeneration, brain development (pruning of undesirable synapses), obesity, asthma, T-cell activation phenomena associated with allergic and rheumatic diseases,[13] and more.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ricklin D, Hajishengallis G, Yank K, et al. Complement: a key system for immune surveillance and homeostasis. *Nat Immunol*. 2010;11:785-797.
2. Frank MM. Complement disorders and hereditary angioedema. *J Allergy Clin Immunol*. 2010; 125:5262-5271.
3. Clarke EV, Tenner AJ. Complement modulation of T cell immune responses during homeostasis and disease. *J Leukoc Biol*. 2014;96:745-756.
4. Cozzani E, Drosera M, Gasparini G, et al. Serology of lupus erythematosus: Correlation between immunopathological features and clinical aspects. *Autoimmune Dis*. 2014;2014:321359.
5. Sethi S, Fervenza FC. Pathology of renal diseases associated with dysfunction of the alternative pathway of complement: C3 glomerulopathy and atypical hemolytic uremic syndrome (aHUS). *Semin Thromb Hemost*. 2014;40:416-421.
6. Genster N, Takahashi M, Sekine H, et al. Lessons learned from mice deficient in lectin complement pathway molecules. *Mol Immunol*. 2014;61:59-68.
7. Kemper C, Atkinson JP, Hourcade DE. Properdin: emerging roles of a pattern-recognition molecule. *Annu Rev Immunol*. 2010;28:131-155.
8. Nesargikar PN, Spiller B, Chavez R. The complement system: history, pathways, cascade and inhibitors. *Eur J Microbiol Immunol (Bp)*. 2012;2:103-111.
9. Sethi S, Fervenza FC. Membranoproliferative glomerulonephritis—a new look at an old entity. *N Engl J Med*. 2012;366:1119-1131.
10. Joseph C, Gattineni J. Complement disorders and hemolytic uremic syndrome. *Curr Opin Pediatr*. 2013;25:209-215.
11. Seddon JM, Yu Y, Miller EC, et al. Rare variants in CFI, C3 and C9 are associated with high risk of advanced age-related macular degeneration. *Nat Genet*. 2013;45:1366-1370.
12. Okumura CY, Nizet V. Subterfuge and sabotage: evasion of host innate defenses by invasive gram-positive bacterial pathogens. *Annu Rev Microbiol*. 2014;68:439-458.
13. Liszewski MK, Kolev M, Le Friec G, et al. Intracellular complement activation sustains T cell homeostasis and mediates effector differentiation. *Immunity*. 2013;39:1143-1157.

REVIEW QUESTIONS

1. What statement is false about the alternate pathway of complement activation?

A. It serves as a feedback or amplification loop for each pathway.
B. It mediates lysis and cell damage in several types of hemolytic disorders.
C. It is involved in debris removal and wound repair.
D. It requires lectins and antibodies to be initiated.

Answer: D The alternate pathway is the original complement system. For example, it is present in insects and echinoderms and functions in hemolymph similar to its role in blood vertebrates. It is often called the "guardian of the intravascular space," being particular designed to prevent invasion by bacteria. It is continuously "turning over" at a low rate. Antibody and lectins are more evolutionary recent means to specifically target complement activation.

2. Which statement is *not* true of the classical pathway?

A. In conjunction with IgG and IgM, the classical pathway mediates tissue damage in autoimmune diseases.
B. A complete deficiency of an early component predisposes to SLE.
C. A total complement titer (CH50 or THC) measures the quantity of hemolytic activity in this pathway.
D. Proteins such as CRP and serum amyloid protein (SAP) also can activate the classical pathway.
E. All are true.

Answer: E It requires only a single IgM or two IgGs in close proximity to activate the classical complement pathway. IgA and IgE do not activate the classical complement pathway.

3. A deficiency of a complement inhibitor at the steps of C3 and C5 activation leads to all of the following except:

A. excessive production of the C3a and C5a anaphylatoxins.
B. excessive production of opsonins and the membrane attack complex.
C. systemic lupus erythematosus and closely related rheumatic diseases.
D. renal disease (atypical hemolytic uremic syndrome, membranoproliferative glomerulonephritis, C3 glomerulonephritis, and C3 glomerulopathies.

Answer: C Excessive activation at sties of injury or degeneration is a hallmark of inadequately regulated C3 and C5 convertases. The diseases for which this type of pathogenesis has been most studies are atypical hemolytic uremic syndrome and age-related macular degeneration. The other commonly observed phenotype is renal disease as described in Answer D.

4. Which statement is *not* true?

A. CP is activated by the Fc portion of IgG or IgM engaging C1q, which is part of C1 complex.
B. Complement receptor one (CR1, CD35) on RBCs serves as a taxi or ferry to take C4b and C3b bearing immune complexes to liver and spleen for disposal.
C. Only a limited number of specific foreign antigens can be coated with complement fragments.
D. Complement proteins in blood are synthesized by the liver, but most cell types also express complement proteins locally.

Answer: C Almost all foreign materials become coated with complement C3 fragments. Of note, a role in host defense and response to injury for local synthesis of complement components remains to be definitely established.

5. The C3a and C5a anaphylatoxins accomplish all but which one of the following?

A. Engage their specific receptors to activate many cell types.
B. If liberated in substantial amounts, they can lead to shock and death.
C. They are responsible for opsonic activity of the complement system.
D. Both are produced by all complement pathways.

Answer: C C3a and C5a are liberated in substantial amounts and this occurs in seconds. It is an early warning system to the host cell to develop a proinflammatory environment. The major opsonin of the complement system is C3b and its subsequent limited proteolytic degradation fragments. C4b is also an opsonin if the classical pathway or lectin pathway is activated.

51 APPROACH TO THE PATIENT WITH POSSIBLE CARDIOVASCULAR DISEASE

LEE GOLDMAN

Patients with cardiovascular disease may present with a wide range of symptoms and signs, each of which may be caused by noncardiovascular conditions. Conversely, patients with substantial cardiovascular disease may be asymptomatic. Because cardiovascular disease is a leading cause of death in the United States and other developed countries, it is crucial that patients be evaluated carefully to detect early cardiovascular disease, that symptoms or signs of cardiovascular disease be evaluated in detail, and that appropriate therapy be instituted. Improvements in diagnosis, therapy, and prevention have contributed to a 70% or so decline in age-adjusted cardiovascular death rates in the United States since the 1960s. Furthermore, among people age 65 years and older, regular visits to a primary care physician are associated with a 25 to 30% reduction in overall mortality. However, the absolute number of deaths from cardiovascular disease in the United States has not declined proportionately because of the increase in the population older than 40 years as well as the aging of the population in general.

In evaluating a patient with known or suspected heart disease, the physician must determine quickly whether a potentially life-threatening condition exists. In these situations, the evaluation must focus on the specific issue at hand and be accompanied by the rapid performance of appropriately directed additional tests. Examples of potentially life-threatening conditions include acute myocardial infarction (MI) (Chapter 73), unstable angina (Chapter 72), suspected aortic dissection (Chapter 78), pulmonary edema (Chapter 59), and pulmonary embolism (Chapter 98).

USING THE HISTORY TO DETECT CARDIOVASCULAR SYMPTOMS

Patients may complain spontaneously of a variety of cardiovascular symptoms (Table 51-1), but sometimes these symptoms are elicited only by obtaining a careful, complete medical history. In patients with known or suspected cardiovascular disease, questions about cardiovascular symptoms are key components of the history of present illness; in other patients, these issues are a fundamental part of the review of systems.

Chest Pain

Chest discomfort or pain is the cardinal manifestation of myocardial ischemia resulting from coronary artery disease or any condition that causes myocardial ischemia by an imbalance of myocardial oxygen demand compared with myocardial oxygen supply (Chapter 71). New, acute, often ongoing pain may indicate an acute MI, unstable angina, or aortic dissection; a pulmonary cause, such as acute pulmonary embolism or pleural irritation; a musculoskeletal condition of the chest wall, thorax, or shoulder; or a gastrointestinal abnormality, such as esophageal reflux or spasm, peptic ulcer disease, or cholecystitis (Table 51-2). The chest discomfort of MI commonly occurs without an immediate or obvious precipitating clinical cause and builds in intensity for at least several minutes; the sensation can range from annoying discomfort to severe pain (Chapter 73). Although a variety of adjectives may be used by patients to describe the sensation, physicians must be suspicious of any discomfort, especially if it radiates to the neck, shoulder, or arms. The probability of an acute MI can be estimated by integrating information from the history, physical examination, and electrocardiogram (Fig. 51-1).

The chest discomfort of unstable angina is clinically indistinguishable from that of MI except that the former may be precipitated more clearly by activity and may be more rapidly responsive to antianginal therapy (Chapter 72). Aortic dissection (Chapter 78) classically presents with the sudden onset of severe pain in the chest and radiating to the back; the location of the pain often provides clues to the location of the dissection. Ascending aortic dissections commonly present with chest discomfort radiating to the back, whereas dissections of the descending aorta commonly present with back pain radiating to the abdomen. The presence of back pain or a history of hypertension or other predisposing factors, such as Marfan syndrome, should prompt a careful assessment of peripheral pulses to determine whether the great vessels are affected by the dissection and of the chest radiograph to evaluate the size of the aorta. If this initial evaluation is suggestive, further testing with transesophageal echocardiography, computed tomography (CT), or magnetic resonance imaging (MRI) is indicated. The pain of pericarditis (Chapter 77) may simulate that of an acute MI, may be primarily pleuritic, or may be continuous; a key physical finding is a pericardial rub. The pain of pulmonary embolism (Chapter 98) is commonly pleuritic in nature and is associated with dyspnea; hemoptysis also may be present. Pulmonary hypertension (Chapter 68) of any cause may be associated with chest discomfort with exertion; it commonly is associated with severe dyspnea and often is associated with cyanosis.

Recurrent, episodic chest discomfort may be noted with angina pectoris and with many cardiac and noncardiac causes (Chapter 71). A variety of stress tests (Table 51-3) can be used to provoke reversible myocardial ischemia in susceptible individuals and to help determine whether ischemia is the pathophysiologic explanation for the chest discomfort (Chapter 71).

Dyspnea

Dyspnea, which is an uncomfortable awareness of breathing, is commonly caused by cardiovascular or pulmonary disease. A systematic approach (see Fig. 83-3) with selected tests nearly always reveals the cause. Acute dyspnea can be caused by myocardial ischemia, heart failure, severe hypertension, pericardial tamponade, pulmonary embolism, pneumothorax, upper airway obstruction, acute bronchitis or pneumonia, or some drug overdoses (e.g., salicylates). Subacute or chronic dyspnea is also a common presenting or accompanying symptom in patients with pulmonary disease (Chapter 83). Dyspnea also can be caused by severe anemia (Chapter 158) and can be confused with the fatigue that often is noted in patients with systemic and neurologic diseases (Chapters 256 and 396).

In heart failure, dyspnea typically is noted as a hunger for air and a need or an urge to breathe. The feeling that breathing requires increased work or effort is more typical of airway obstruction or neuromuscular disease. A feeling of chest tightness or constriction during breathing is typical of bronchoconstriction, which is commonly caused by obstructive airway disease (Chapters 87 and 88) but also may be seen in pulmonary edema. A feeling of heavy breathing, a feeling of rapid breathing, or a need to breathe more is classically associated with deconditioning.

In cardiovascular conditions, chronic dyspnea usually is caused by increases in pulmonary venous pressure as a result of left ventricular failure (Chapters 58 and 59) or valvular heart disease (Chapter 75). Orthopnea, which is an exacerbation of dyspnea when the patient is recumbent, is caused by increased work of breathing because of either increased venous return to the pulmonary vasculature or loss of gravitational assistance in diaphragmatic effort. Paroxysmal nocturnal dyspnea is severe dyspnea that awakens a patient at night and forces the assumption of a sitting or standing position to achieve gravitational redistribution of fluid.

Palpitations

Palpitations (Chapter 62) describe a subjective sensation of an irregular or abnormal heartbeat. Palpitations may be caused by any arrhythmia (Chapters 64 and 65) with or without important underlying structural heart disease. Palpitations should be defined in terms of the duration and frequency of the episodes; the precipitating and related factors; and any associated symptoms of chest pain, dyspnea, lightheadedness, or syncope. It is crucial to use the history to determine whether the palpitations are caused by an irregular or a regular heartbeat. The feeling associated with a premature atrial or ventricular contraction, often described as a "skipped beat" or a "flip-flopping of the heart," must be distinguished from the irregularly irregular rhythm of atrial fibrillation and the rapid but regular rhythm of supraventricular tachycardia. Associated symptoms of chest pain, dyspnea, lightheadedness, dizziness, or diaphoresis suggest an important effect on cardiac output and mandate further evaluation. In general, evaluation begins with ambulatory electrocardiography (ECG) (Table 51-4), which is indicated in patients who have palpitations in the presence of structural heart disease or substantial accompanying symptoms. Depending on the series, 9 to 43% of patients have important underlying heart disease. In such patients, more detailed evaluation is warranted (see Fig. 62-1).

Lightheadedness or *syncope* (Chapter 62) can be caused by any condition that decreases cardiac output (e.g., bradyarrhythmia, tachyarrhythmia, obstruction of the left ventricular or right ventricular inflow or outflow, cardiac tamponade, aortic dissection, or severe pump failure), by reflex-mediated vasomotor instability (e.g., vasovagal, situational, or carotid sinus syncope), or by orthostatic hypotension (see Table 62-1). Neurologic

TABLE 51-1 CARDINAL SYMPTOMS OF CARDIOVASCULAR DISEASE

Chest pain or discomfort
Dyspnea, orthopnea, paroxysmal nocturnal dyspnea, wheezing
Palpitations, dizziness, syncope
Cough, hemoptysis
Fatigue, weakness
Pain in extremities with exertion (claudication)

diseases (e.g., migraine headaches, transient ischemic attacks, or seizures) also can cause transient loss of consciousness. The history, physical examination, and ECG are often diagnostic of the cause of syncope (see Table 62-2). Syncope caused by a cardiac arrhythmia usually occurs with little warning. Syncope with exertion or just after conclusion of exertion is typical of aortic stenosis and hypertrophic obstructive cardiomyopathy. In many patients, additional testing is required to document central nervous system disease, the cause of reduced cardiac output, or carotid sinus syncope. When the history, physical examination, and ECG do not provide helpful diagnostic information that points toward a specific cause of syncope, it is imperative that patients with heart disease or an abnormal ECG be tested with continuous ambulatory ECG monitoring to diagnose a possible arrhythmia (see Fig. 62-1); in selected patients, formal electrophysiologic testing may be indicated

TABLE 51-2 CAUSES OF CHEST PAIN

CONDITION	LOCATION	QUALITY	DURATION	AGGRAVATING OR RELIEVING FACTORS	ASSOCIATED SYMPTOMS OR SIGNS
CARDIOVASCULAR CAUSES					
Angina	Retrosternal region; radiates to or occasionally isolated to neck, jaw, epigastrium, shoulder, or arms (left common)	Pressure, burning, squeezing, heaviness, indigestion	<2-10 min	Precipitated by exercise, cold weather, or emotional stress; relieved by rest or nitroglycerin; atypical (Prinzmetal) angina may be unrelated to activity, often early morning	S_3 or murmur of papillary muscle dysfunction during pain
Rest or unstable angina	Same as angina	Same as angina but may be more severe	Usually <20 min	Same as angina, with decreasing tolerance for exertion or at rest	Similar to stable angina but may be pronounced; transient heart failure can occur
Myocardial infarction	Substernal and may radiate like angina	Heaviness, pressure, burning, constriction	≥30 min but variable	Unrelieved by rest or nitroglycerin	Shortness of breath, sweating, weakness, nausea, vomiting
Pericarditis	Usually begins over sternum or toward cardiac apex and may radiate to neck or left shoulder; often more localized than the pain of myocardial ischemia	Sharp, stabbing, knifelike	Lasts many hours to days; may wax and wane	Aggravated by deep breathing, rotating chest, or supine position; relieved by sitting up and leaning forward	Pericardial friction rub
Aortic dissection	Anterior chest; may radiate to back	Excruciating, tearing, knifelike	Sudden onset, unrelenting	Usually occurs in setting of hypertension or predisposition, such as Marfan syndrome	Murmur of aortic insufficiency, pulse or blood pressure asymmetry; neurologic deficit
Pulmonary embolism (chest pain often not present)	Substernal or over region of pulmonary infarction	Pleuritic (with pulmonary infarction) or angina-like	Sudden onset; minutes to <1 hr	May be aggravated by breathing	Dyspnea, tachypnea, tachycardia; hypotension, signs of acute right ventricular failure, and pulmonary hypertension with large emboli; rales, pleural friction rub, hemoptysis with pulmonary infarction
Pulmonary hypertension	Substernal	Pressure; oppressive	Similar to angina	Aggravated by effort	Pain usually associated with dyspnea; signs of pulmonary hypertension
NONCARDIAC CAUSES					
Pneumonia with pleurisy	Localized over involved area	Pleuritic, localized	Brief or prolonged	Painful breathing	Dyspnea, cough, fever, dull to percussion, bronchial breath sounds, rales, occasional pleural friction rub
Spontaneous pneumothorax	Unilateral	Sharp, well localized	Sudden onset, lasts many hours	Painful breathing	Dyspnea; hyperresonance and decreased breath and voice sounds over involved lung
Musculoskeletal disorders	Variable	Aching	Short or long duration	Aggravated by movement; history of muscle exertion or injury	Tender to pressure or movement
Herpes zoster	Dermatomal in distribution	Burning, itching	Prolonged	None	Vesicular rash appears in area of discomfort
Esophageal reflux	Substernal, epigastric	Burning, visceral discomfort	10-60 min	Aggravated by large meal, postprandial recumbency; relief with antacid	Water brash
Peptic ulcer	Epigastric, substernal	Visceral burning, aching	Prolonged	Relief with food, antacid	
Gallbladder disease	Epigastric, right upper quadrant	Visceral	Prolonged	May be unprovoked or follow meals	Right upper quadrant tenderness may be present
Anxiety states	Often localized over precordium	Variable; location often moves from place to place	Varies; often fleeting	Situational	Sighing respirations, often chest wall tenderness

Modified from Andreoli TE, Carpenter CCJ, Griggs RC, et al. Evaluation of the patient with cardiovascular disease. In: *Cecil Essentials of Medicine*, 6th ed. Philadelphia: WB Saunders; 2004:34-35.

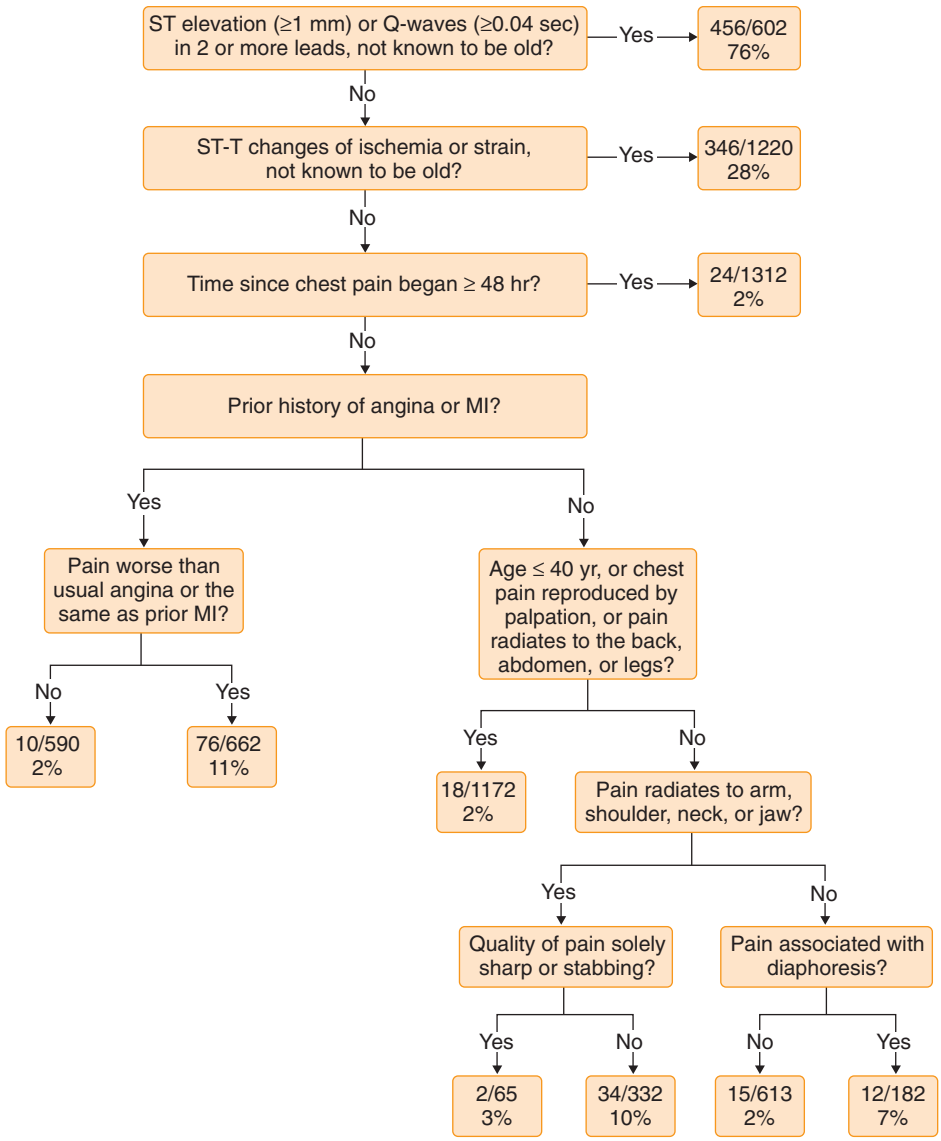

FIGURE 51-1. Flow diagram to estimate the risk for acute myocardial infarction (MI) in emergency departments in patients with acute chest pain. For each clinical subset, the numerator is the number of patients with the set of presenting characteristics who had an MI; the denominator is the total number of patients presenting with that characteristic or set of characteristics. CHF = congestive heart failure; DVT = deep vein thrombosis. (Modified from Pearson SD, Goldman L, Garcia TB, et al. Physician response to a prediction rule for the triage of emergency department patients with chest pain. *J Gen Intern Med*. 1994;9:241-247.)

TABLE 51-3 COMMON EXERCISE TEST PROTOCOLS*

PROTOCOL	STAGE	DURATION (min)	GRADE (%)	RATE (mph)	METABOLIC EQUIVALENTS AT COMPLETION	FUNCTIONAL CLASS
Modified Bruce protocol†	1	3	0	1.7	2.5	III
	2	3	10	1.7	5	II
	3	3	12	2.5	7	I
	4	3	14	3.4	10	I
	5	3	16	4.2	13	I
Naughton protocol‡	0	2	0	2	2	III
	1	2	3.5	2	3	III
	2	2	7	2	4	III
	3	2	10.5	2	5	II
	4	2	14	2	6	II
	5	2	17.5	2	7	I

*Ramp protocols in which the workload is gradually increased on the basis of the patient's estimated functional capacity to achieve maximal effort in approximately 10 minutes are also useful.
†Commonly used in ambulatory patients.
‡Commonly used in patients with recent myocardial infarction, unstable angina, or other conditions that are expected to limit exercise.
Modified from Braunwald E, Goldman L, eds. *Primary Cardiology*. 2nd ed. Philadelphia: WB Saunders; 2003.

TABLE 51-4 AMERICAN HEART ASSOCIATION/AMERICAN COLLEGE OF CARDIOLOGY GUIDELINES FOR USE OF DIAGNOSTIC TESTS IN PATIENTS WITH PALPITATIONS*

AMBULATORY ELECTROCARDIOGRAPHY	
Class I	Palpitations, syncope, dizziness
Class II	Shortness of breath, chest pain, or fatigue (not otherwise explained, episodic, and strongly suggestive of an arrhythmia as the cause because of a relation of the symptom with palpitation)
Class III	Symptoms not reasonably expected to be caused by arrhythmia
ELECTROPHYSIOLOGIC STUDY	
Class I	Patients with palpitations who have a pulse rate documented by medical personnel as inappropriately rapid and in whom electrocardiographic recordings fail to document the cause of the palpitations Patients with palpitations preceding a syncopal episode
Class II	Patients with clinically significant palpitations, suspected to be of cardiac origin, in whom symptoms are sporadic and cannot be documented; studies are performed to determine the mechanisms of arrhythmias, to direct or provide therapy or to assess prognosis
Class III	Patients with palpitations documented to have extracardiac causes (e.g., hyperthyroidism)
ECHOCARDIOGRAPHY	
Class I	Arrhythmias with evidence of heart disease Family history of genetic disorder associated with arrhythmias
Class II	Arrhythmias commonly associated with, but without evidence of, heart disease Atrial fibrillation or flutter
Class III	Palpitations without evidence of arrhythmias Minor arrhythmias without evidence of heart disease

*Class I, general agreement the test is useful and indicated; class II, frequently used, but there is a divergence of opinion with respect to its utility; class III, general agreement the test is not useful.
From Braunwald E, Goldman L, eds. *Primary Cardiology.* 2nd ed. Philadelphia: WB Saunders; 2003:132.

TABLE 51-5 A COMPARISON OF THREE METHODS OF ASSESSING CARDIOVASCULAR DISABILITY

CLASS	NEW YORK HEART ASSOCIATION FUNCTIONAL CLASSIFICATION	CANADIAN CARDIOVASCULAR SOCIETY FUNCTIONAL CLASSIFICATION	SPECIFIC ACTIVITY SCALE
I	Patients with cardiac disease but without resulting limitations of physical activity Ordinary physical activity does not cause undue fatigue, palpitation, dyspnea, or anginal pain.	Ordinary physical activity, such as walking and climbing stairs, does not cause angina. Angina with strenuous or rapid or prolonged exertion at work or recreation	Patients can perform to completion any activity requiring ≥7 metabolic equivalents, e.g., can carry 24 lb up 8 steps; carry objects that weigh 80 lb; do outdoor work (shovel snow, spade soil); do recreational activities (skiing, basketball, squash, handball, jog or walk 5 mph)
II	Patients with cardiac disease resulting in slight limitation of physical activity They are comfortable at rest. Ordinary physical activity results in fatigue, palpitations, dyspnea, or anginal pain.	Slight limitation of ordinary activity Walking or climbing stairs rapidly, walking uphill, walking or stair climbing after meals, in cold, in wind, or when under emotional stress, or only during the few hours after awakening Walking >2 blocks on the level and climbing >1 flight of ordinary stairs at a normal pace and in normal conditions	Patient can perform to completion any activity requiring ≥5 metabolic equivalents but cannot and does not perform to completion activities requiring ≥7 metabolic equivalents, e.g., have sexual intercourse without stopping, garden, rake, weed, roller skate, dance foxtrot, walk at 4 mph on level ground
III	Patients with cardiac disease resulting in marked limitation of physical activity They are comfortable at rest. Less than ordinary physical activity causes fatigue, palpitations, dyspnea, or anginal pain.	Marked limitation of ordinary physical activity Walking 1 or 2 blocks on the level and climbing >1 flight in normal conditions	Patient can perform to completion any activity requiring ≥2 metabolic equivalents but cannot and does not perform to completion any activities requiring ≥5 metabolic equivalents, e.g., shower without stopping, strip and make bed, clean windows, walk 2.5 mph, bowl, play golf, dress without stopping
IV	Patients with cardiac disease resulting in inability to carry on any physical activity without discomfort Symptoms of cardiac insufficiency or of the anginal syndrome may be present even at rest. If any physical activity is undertaken, discomfort is increased.	Inability to carry on any physical activity without discomfort—anginal syndrome may be present at rest	Patient cannot or does not perform to completion activities requiring ≥2 metabolic equivalents; cannot carry out activities listed above (Specific Activity Scale, class III)

From Goldman L, Hashimoto B, Cook EF, et al. Comparative reproducibility and validity of systems for assessing cardiovascular functional class: advantages of a new specific activity scale. *Circulation.* 1981;64:1227-1234. Reproduced by permission of the American Heart Association.

(Chapter 62). In patients with no evident heart disease, tilt testing (Chapter 62) can help detect reflex-mediated vasomotor instability.

Other Symptoms

Nonproductive *cough* (Chapter 83), especially a persistent cough (see Fig. 83-1), can be an early manifestation of elevated pulmonary venous pressure and otherwise unsuspected heart failure. *Fatigue* and *weakness* are common accompaniments of advanced cardiac disease and reflect an inability to perform normal activities. A variety of approaches have been used to classify the severity of cardiac limitations, ranging from class I (little or no limitation) to class IV (severe limitation) (Table 51-5). *Hemoptysis* (Chapter 83) is a classic presenting finding in patients with pulmonary embolism, but it is also common in patients with mitral stenosis, pulmonary edema, pulmonary infections, and malignant neoplasms (see Table 83-6). *Claudication,* which is pain in the extremities with exertion, should alert the physician to possible peripheral arterial disease (Chapters 79 and 80).

Complete Medical History

The complete medical history should include a thorough review of systems, family history, social history, and past medical history (Chapter 15). The

review of systems may reveal other symptoms that suggest a systemic disease as the cause of any cardiovascular problems. The family history should focus on premature atherosclerosis or evidence of familial abnormalities, such as may be found with various causes of the long QT syndrome (Chapter 65) or hypertrophic cardiomyopathy (Chapter 60).

The social history should include specific questioning about cigarette smoking, alcohol intake, and use of illicit drugs. The past medical history may reveal prior conditions or medications that suggest systemic diseases, ranging from chronic obstructive pulmonary disease, which may explain a complaint of dyspnea, to hemochromatosis, which may be a cause of restrictive cardiomyopathy. A careful history to inquire about recent dental work or other procedures is crucial if bacterial endocarditis is part of the differential diagnosis.

PHYSICAL EXAMINATION FOR DETECTION OF SIGNS OF CARDIOVASCULAR DISEASE

The cardiovascular physical examination, which is a subset of the complete physical examination, provides important clues to the diagnosis of asymptomatic and symptomatic cardiac disease and may reveal cardiovascular manifestations of noncardiovascular diseases. The cardiovascular physical examination begins with careful measurement of the pulse and blood pressure (Chapter 8). If aortic dissection (Chapter 78) is a consideration, blood pressure should be measured in both arms and, preferably, in at least one leg. When coarctation of the aorta is suspected (Chapter 69), blood pressure must be measured in at least one leg and in the arms. Discrepancies in blood pressure between the two arms also can be caused by atherosclerotic disease of the great vessels. Pulsus paradoxus, which is more than the usual 10-mm Hg drop in systolic blood pressure during inspiration, is typical of pericardial tamponade (Chapter 77).

General Appearance

The respiratory rate may be increased in patients with heart failure. Patients with pulmonary edema are usually markedly tachypneic and may have labored breathing. Patients with advanced heart failure may have Cheyne-Stokes respirations.

Systemic diseases, such as hyperthyroidism (Chapter 226), hypothyroidism (Chapter 226), rheumatoid arthritis (Chapter 264), scleroderma (Chapter 267), and hemochromatosis (Chapter 212), may be suspected from the patient's general appearance. Marfan syndrome (Chapter 260), Turner syndrome (Chapter 235), Down syndrome (Chapter 41), and a variety of congenital anomalies also may be readily apparent.

Ophthalmologic Examination

Examination of the fundi may show diabetic (see Fig. 423-24) or hypertensive retinopathy (see Fig. 67-8) or Roth spots (see Fig. 423-28) typical of infectious endocarditis. Beading of the retinal arteries is typical of severe hypercholesterolemia. Osteogenesis imperfecta, which is associated with blue sclerae, also is associated with aortic dilation and mitral valve prolapse. Retinal artery occlusion (see Fig. 423-29) may be caused by an embolus from clot in the left atrium or left ventricle, a left atrial myxoma, or atherosclerotic debris from the great vessels. Hyperthyroidism may present with exophthalmos and typical stare (see Fig. 423-6), whereas myotonic dystrophy, which is associated with atrioventricular block and arrhythmia, often is associated with ptosis and an expressionless face (see Fig. 421-2).

Jugular Veins

The external jugular veins help in assessment of mean right atrial pressure, which normally varies between 5 and 10 cm H_2O; the height (in centimeters) of the central venous pressure is measured by adding 5 cm to the height of the observed jugular venous distention above the sternal angle of Louis (Fig. 51-2). The normal jugular venous pulse, best seen in the internal jugular vein (and not seen in the external jugular vein unless insufficiency of the jugular venous valves is present), includes an *a* wave, caused by right atrial contraction; a *c* wave, reflecting carotid artery pulsation; an *x* descent; a *v* wave, which corresponds to isovolumetric right ventricular contraction and is more marked in the presence of tricuspid insufficiency; and a *y* descent, which occurs as the tricuspid valve opens and ventricular filling begins (Fig. 51-3). Abnormalities of the jugular venous pressure (Fig. 51-4) are useful in detecting heart failure, and they correlate well with brain natriuretic peptide levels (Chapter 58) and echocardiographic evidence of an elevated pulmonary artery pressure (Chapter 55).[1] The jugular venous pressure also helps in the diagnosis of pericardial disease, tricuspid valve disease, and pulmonary hypertension (Table 51-6).

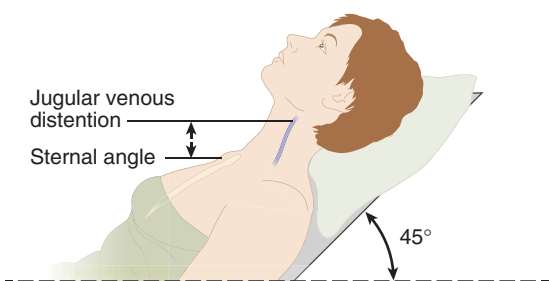

FIGURE 51-2. Jugular venous distention is defined by engorgement of the internal jugular vein more than 5 cm above the sternal angle at 45 degrees. The central venous pressure is the observed venous distention above the sternal angle plus 5 cm.

FIGURE 51-3. Typical distention of the internal jugular vein. (From http://courses.cvcc.vccs.edu/WisemanD/jugular_vein_distention.htm.)

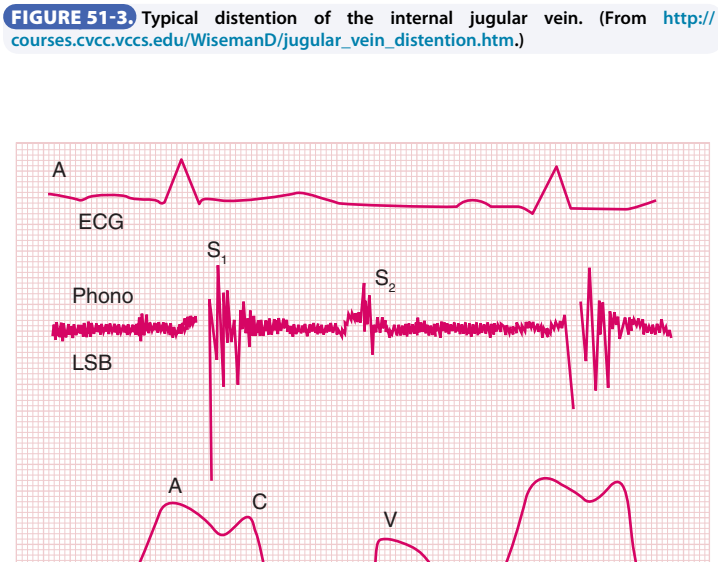

FIGURE 51-4. Normal jugular venous pulse. ECG = electrocardiogram; JUG = jugular vein; LSB = left sternal border; phono = phonocardiogram; S_1 = first heart sound; S_2 = second heart sound.

Carotid Pulse

The carotid pulse should be examined in terms of its volume and contour. The carotid pulse (Fig. 51-5) may be increased in frequency and may be more intense than normal in patients with a higher stroke volume secondary to aortic regurgitation, arteriovenous fistula, hyperthyroidism, fever, or anemia. In aortic regurgitation or arteriovenous fistula, the pulse may have a bisferious quality. The carotid upstroke is delayed in patients with valvular aortic

stenosis (Chapter 75) and has a normal contour but diminished amplitude in any cause of reduced stroke volume.

Cardiac Inspection and Palpation

Inspection of the precordium may reveal the hyperinflation of obstructive lung disease or unilateral asymmetry of the left side of the chest because of right ventricular hypertrophy before puberty. Palpation may be performed with the patient either supine or in the left lateral decubitus position; the latter position moves the left ventricular apex closer to the chest wall and increases the ability to palpate the point of maximal impulse and other phenomena. Low-frequency phenomena, such as systolic heaves or lifts from the left ventricle (at the cardiac apex) or right ventricle (parasternal in the third or fourth intercostal space), are felt best with the heel of the palm. With the patient in the left lateral decubitus position, this technique also may allow palpation of an S_3 gallop in cases of advanced heart failure or an S_4 gallop in cases of poor left ventricular distensibility during diastole. The left ventricular apex is more diffuse and sometimes may be frankly dyskinetic in patients with advanced heart disease. The distal palm is best for feeling thrills, which are the tactile equivalent of cardiac murmurs. By definition, a thrill denotes a murmur of grade 4/6 or louder. Higher-frequency events may be felt best with the fingertips; examples include the opening snap of mitral stenosis or the loud pulmonic second sound of pulmonary hypertension.

TABLE 51-6 ABNORMALITIES OF VENOUS PRESSURE AND PULSE AND THEIR CLINICAL SIGNIFICANCE

Positive hepatojugular reflux	Suspect heart failure, particularly left ventricular systolic dysfunction (echocardiography recommended)
Elevated systemic venous pressure without obvious *x* or *y* descent, quiet precordium, and pulsus paradoxus	Suspect cardiac tamponade (echocardiography recommended)
Elevated systemic venous pressure with sharp *y* descent, Kussmaul sign, and quiet precordium	Suspect constrictive pericarditis (cardiac catheterization and MRI or CT recommended)
Elevated systemic venous pressure with a sharp brief *y* descent, Kussmaul sign, and evidence of pulmonary hypertension and tricuspid regurgitation	Suspect restrictive cardiomyopathy (cardiac catheterization and MRI or CT recommended)
A prominent *a* wave with or without elevation of mean systemic venous pressure	Exclude tricuspid stenosis, right ventricular hypertrophy caused by pulmonary stenosis, and pulmonary hypertension (echo-Doppler study recommended)
A prominent *v* wave with a sharp *y* descent	Suspect tricuspid regurgitation (echo-Doppler or cardiac catheterization to determine etiology)

CT = computed tomography; MRI = magnetic resonance imaging.
From Braunwald E, ed. *Heart Disease: A Textbook of Cardiovascular Medicine*. 5th ed. Philadelphia: WB Saunders; 1997.

Auscultation

The first heart sound (Fig. 51-6), which is largely produced by closure of the mitral and—to a lesser extent—the tricuspid valves, may be louder in patients with mitral valve stenosis and intact valve leaflet movement and less audible in patients with poor closure caused by mitral regurgitation (Chapter 75). The second heart sound is caused primarily by closure of the aortic valve, but closure of the pulmonic valve is also commonly audible. In normal individuals, the louder aortic closure sound occurs first, followed by pulmonic closure. With expiration, the two sounds are virtually superimposed. With inspiration, by comparison, the increased stroke volume of the right ventricle commonly leads to a discernible splitting of the second sound. This splitting may be fixed in patients with an atrial septal defect (Chapter 69) or a right bundle branch block. The split may be paradoxical in patients with left bundle branch block or other causes of delayed left ventricular emptying. The aortic component of the second sound is increased in intensity in the presence of systemic hypertension and decreased in intensity in patients with aortic stenosis. The pulmonic second sound is increased in the presence of pulmonary hypertension.

Early systolic ejection sounds are related to forceful opening of the aortic or pulmonic valve. These sounds are common in congenital aortic stenosis, with a mobile valve; in hypertension, with forceful opening of the aortic valve; and in healthy young individuals, especially when cardiac output is increased. Midsystolic or late systolic clicks are caused most commonly by mitral valve prolapse (Chapter 75). Clicks are relatively high-frequency sounds that are heard best with the diaphragm of the stethoscope.

An S_3 corresponds to rapid ventricular filling during early diastole. It may occur in normal children and young adults, especially if stroke volume is

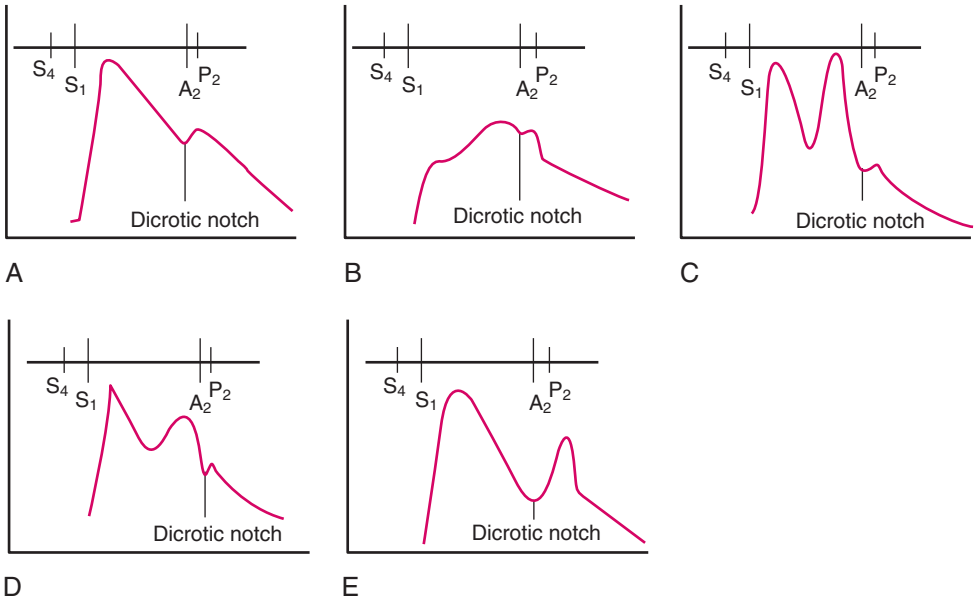

FIGURE 51-5. Schematic diagrams of the configurational changes in the carotid pulse and their differential diagnosis. Heart sounds also are illustrated. **A,** Normal. **B,** Anacrotic pulse with slow initial upstroke. The peak is close to the second heart sound. These features suggest fixed left ventricular outflow obstruction, such as valvular aortic stenosis. **C,** Pulsus bisferiens, with percussion and tidal waves occurring during systole. This type of carotid pulse contour is observed most frequently in patients with hemodynamically significant aortic regurgitation or combined aortic stenosis and regurgitation with dominant regurgitation. It rarely is observed in patients with mitral valve prolapse or in normal individuals. **D,** Pulsus bisferiens in hypertrophic obstructive cardiomyopathy. This finding rarely is appreciated at the bedside by palpation. **E,** Dicrotic pulse results from an accentuated dicrotic wave and tends to occur in sepsis, severe heart failure, hypovolemic shock, and cardiac tamponade and after aortic valve replacement. A_2 = aortic component of the second heart sound; P_2 = pulmonary component of the second heart sound); S_1 = first heart sound; S_4 = atrial sounds. (From Chatterjee K. Bedside evaluation of the heart: the physical examination. In: Chatterjee K, Chetlin MD, Karliner J, et al, eds. *Cardiology: An Illustrated Text/Reference*. Philadelphia: JB Lippincott; 1991:3.11-3.51.)

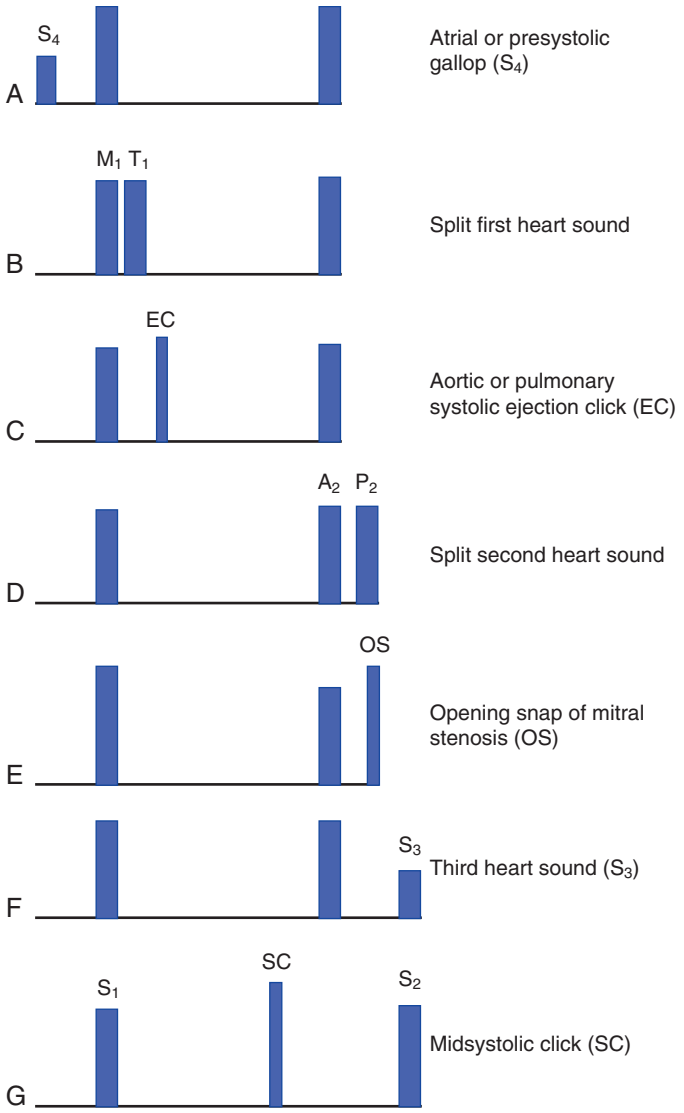

FIGURE 51-6. Timing of the different heart sounds and added sounds. (Modified from Wood P. *Diseases of the Heart and Circulation*. 3rd ed. Philadelphia: JB Lippincott; 1968.)

increased. After about 40 years of age, however, an S_3 should be considered abnormal; it is caused by conditions that increase the volume of ventricular filling during early diastole (e.g., mitral regurgitation) or that increase pressure in early diastole (e.g., advanced heart failure). A left ventricular S_3 gallop is heard best at the apex, whereas the right ventricular S_3 gallop is heard best at the fourth intercostal space at the left parasternal border; both are heard best with the bell of the stethoscope. An S_4 is heard rarely in young individuals but is common in adults older than 40 or 50 years because of reduced ventricular compliance during atrial contraction; it is a nearly ubiquitous finding in patients with hypertension, heart failure, or ischemic heart disease.

The opening snap of mitral and, less commonly, tricuspid stenosis (Chapter 75) occurs at the beginning of mechanical diastole, before the onset of the rapid phase of ventricular filling. An opening snap is high pitched and is heard best with the diaphragm; this differential frequency should help distinguish an opening snap from an S_3 on physical examination. An opening snap commonly can be distinguished from a loud pulmonic component of the second heart sound by the differential location (mitral opening snap at the apex, tricuspid opening snap at the left third or fourth intercostal space, pulmonic second sound at the left second intercostal space) and by the longer interval between S_2 and the opening snap.

Heart murmurs may be classified as systolic, diastolic, or continuous (Table 51-7). Murmurs are graded by intensity on a scale of 1 to 6. Grade 1 is faint and appreciated only by careful auscultation; grade 2, readily audible; grade 3, moderately loud; grade 4, loud and associated with a palpable thrill; grade 5, loud and audible with the stethoscope only partially placed on the chest; and grade 6, loud enough to be heard without the stethoscope on the chest. Systolic ejection murmurs usually peak in early to mid systole when left ventricular ejection is maximal; examples include fixed valvular, supravalvular, or infravalvular aortic stenosis and pulmonic stenosis. The murmur of hypertrophic obstructive cardiomyopathy has a similar ejection quality, although its peak may be later in systole when dynamic obstruction is maximal (Chapter 60). Pansystolic murmurs are characteristic of mitral or tricuspid regurgitation or with a left-to-right shunt from conditions such as a ventricular septal defect (left ventricle to right ventricle). A late systolic murmur is characteristic of mitral valve prolapse (Chapter 75) or ischemic papillary muscle dysfunction. Ejection quality murmurs also may be heard in patients with normal valves but increased flow, such as occurs with marked anemia, fever, or bradycardia secondary to congenital complete heart block; they also may be heard across a valve that is downstream from increased flow because of an intracardiac shunt. Maneuvers such as inspiration, expiration, standing, squatting, and hand gripping can be especially useful in the differential diagnosis of a murmur; however, echocardiography commonly is required to make a definitive diagnosis of cause and severity (Table 51-8).

High-frequency, early diastolic murmurs are typical of aortic regurgitation and pulmonic regurgitation from a variety of causes. The murmurs of mitral and tricuspid stenosis begin in early to mid diastole and tend to diminish in intensity later in diastole in the absence of effective atrial contraction, but they tend to increase in intensity in later diastole if effective atrial contraction is present.

Continuous murmurs may be caused by any abnormality that is associated with a pressure gradient in systole and diastole. Examples include a patent ductus arteriosus, ruptured sinus of Valsalva aneurysm, arteriovenous fistula (of the coronary artery, pulmonary artery, or thoracic artery), and a mammary soufflé. In some situations, murmurs of two coexistent conditions (e.g., aortic stenosis and regurgitation, atrial septal defect with a large shunt and resulting flow murmurs of relative mitral and pulmonic stenosis) may mimic a continuous murmur.

Unfortunately, the physical examination is limited for detecting meaningful valvular heart disease.[2] As a result, echocardiography (Chapter 55) is critical to the evaluation of patients with suspected structural heart disease.

Abdomen

The most common cause of hepatomegaly in patients with heart disease is hepatic engorgement from elevated right-sided pressures associated with right ventricular failure of any cause. Hepatojugular reflux is elicited by pressing on the liver and showing an increase in the jugular venous pressure; it indicates advanced right ventricular failure or obstruction to right ventricular filling. Evaluation of the abdomen also may reveal an enlarged liver caused by a systemic disease, such as hemochromatosis (Chapter 212) or sarcoidosis (Chapter 95), which also may affect the heart. In more severe cases, splenomegaly and ascites also may be noted. Large, palpable, polycystic kidneys (Chapter 127) commonly are associated with hypertension. A systolic bruit suggestive of renal artery stenosis (Chapter 125) or an enlarged abdominal aorta (Chapter 78) is a clue of atherosclerosis.

Extremities

Extremities should be evaluated for peripheral pulses, edema, cyanosis, and clubbing. Diminished peripheral pulses suggest peripheral arterial disease (Chapters 79 and 80). Delayed pulses in the legs are consistent with coarctation of the aorta and are seen after aortic dissection.

Edema (Fig. 51-7) is a cardinal manifestation of right-sided heart failure.[3] When it is caused by heart failure, pericardial disease, or pulmonary hypertension, the edema is usually symmetrical and progresses upward from the ankles; each of these causes of cardiac edema commonly is associated with jugular venous distention and often with hepatic congestion. Unilateral edema suggests thrombophlebitis or proximal venous or lymphatic obstruction (Fig. 51-8). Edema in the absence of evidence of right-sided or left-sided heart failure suggests renal disease, hypoalbuminemia, myxedema, or other noncardiac causes. Among unselected patients with bilateral edema, about 40% have an underlying cardiac disease, about 40% have an elevated pulmonary blood pressure, about 20% have bilateral venous disease, about 20% have renal disease, and about 25% have idiopathic edema.

Cyanosis (Fig. 51-9) is a bluish discoloration caused by reduced hemoglobin exceeding about 5 g/dL in the capillary bed. Central cyanosis is seen in patients with poor oxygen saturation resulting from a reduced inspired oxygen concentration or inability to oxygenate the blood in the lungs (e.g., as a result of advanced pulmonary disease, pulmonary edema, pulmonary arteriovenous fistula, or right-to-left shunting); it also may be seen in patients with marked erythrocytosis. Methemoglobinemia (Chapter 158) also can present with cyanosis. Peripheral cyanosis may be caused by reduced blood flow to the extremities secondary to vasoconstriction, heart failure, or shock.

TABLE 51-7 SOME COMMON CAUSES OF HEART MURMURS*

	USUAL LOCATION	COMMON ASSOCIATED FINDINGS
SYSTOLIC		
Holosystolic		
Mitral regurgitation	Apex → axilla	↑ with handgrip; S_3 if marked mitral regurgitation; left ventricular dilation common
Tricuspid regurgitation	LLSB	↑ with inspiration; right ventricular dilation common
Ventricular septal defect	LLSB → RLSB	Often with thrill
Early to mid systolic		
Aortic valvular stenosis	RUSB	
Fixed supravalvular or subvalvular	RUSB	Ejection click if mobile valve; soft or absent A_2 if valve immobile; later peak associated with more severe stenosis
Dynamic infravalvular	LLSB → apex + axilla	Hypertrophic obstructive cardiomyopathy; murmur louder if left ventricular volume lower or contractility increased, softer if left ventricular volume increased†; can be later in systole if obstruction delayed
Pulmonic valvular stenosis	LUSB	↑ with inspiration
Infravalvular (infundibular)	LUSB	↑ with inspiration
Supravalvular	LUSB	↑ with inspiration
"Flow murmurs"	LUSB	Anemia, fever, increased flow of any cause‡
Mid to late systolic		
Mitral valve prolapse	LLSB or apex → axilla	Preceded by click; murmur lengthens with maneuvers that decrease left ventricular volume†
Papillary muscle dysfunction	Apex → axilla	Ischemic heart disease
DIASTOLIC		
Early diastolic		
Aortic regurgitation	RUSB, LUSB	High-pitched, blowing quality; endocarditis, diseases of the aorta, associated aortic valvular stenosis; signs of low peripheral vascular resistance
Pulmonic valve regurgitation	LUSB	Pulmonary hypertension as a causative factor
Mid to late diastolic		
Mitral stenosis, tricuspid stenosis	Apex, LLSB	Low pitched; in rheumatic heart disease, opening snap commonly precedes murmur; can be caused by increased flow across normal valve‡
Atrial myxomas	Apex (L), LLSB (R)	"Tumor plop"
Continuous		
Venous hum	Over jugular or hepatic vein or breast	
Patent ductus arteriosus	LUSB	
Arteriovenous fistula		
Coronary	LUSB	
Pulmonary, bronchial, chest wall	Over fistula	
Ruptured sinus of Valsalva aneurysm	RUSB	Sudden onset

*See also Chapters 69 and 75.
†Left ventricular volume is decreased by standing or during prolonged, forced expiration against a closed glottis (Valsalva maneuver); it is increased by squatting or by elevation of the legs; contractility is increased by adrenergic stimulation or in the beat after an extrasystolic beat.
‡Including a left-to-right shunt through an atrial septal defect for tricuspid or pulmonic flow murmurs, and a ventricular septal defect for pulmonic or mitral flow murmurs.
LLSB = left lower sternal border (fourth intercostal space); LUSB = left upper sternal border (second and third intercostal spaces); RLSB = right lower sternal border (fourth intercostal space); RUSB = right upper sternal border (second and third intercostal spaces).

TABLE 51-8 SENSITIVITY AND SPECIFICITY OF BEDSIDE MANEUVERS IN THE IDENTIFICATION OF SYSTOLIC MURMURS

MANEUVER	RESPONSE	MURMUR	SENSITIVITY (%)	SPECIFICITY (%)
Inspiration	↑	RS	100	88
Expiration	↓	RS	100	88
Valsalva maneuver	↑	HC	65	96
Squat to stand	↑	HC	95	84
Stand to squat	↓	HC	95	85
Leg elevation	↓	HC	85	91
Handgrip	↓	HC	85	75
Handgrip	↑	MR and VSD	68	92
Transient arterial occlusion	↑	MR and VSD	78	100

HC = hypertrophic cardiomyopathy; MR = mitral regurgitation; RS = right sided; VSD = ventricular septal defect.
Modified with permission from Lembo NJ, Dell'Italia IJ, Crawford MH, et al. Bedside diagnosis of systolic murmurs. *N Engl J Med.* 1988;318:1572-1578. Copyright 1988 Massachusetts Medical Society. All rights reserved.

Clubbing (Fig. 51-10), which is loss of the normal concave configuration of the nail as it emerges from the distal phalanx, is seen in patients with pulmonary abnormalities such as lung cancer (Chapter 191) and in patients with cyanotic congenital heart disease (Chapter 69).[4]

Examination of the Skin

Examination of the skin may reveal bronze pigmentation typical of hemochromatosis (Chapter 212); jaundice (see Fig. 146-1) characteristic of severe right-sided heart failure or hemochromatosis; or capillary hemangiomas typical of Osler-Weber-Rendu disease (see Fig. 173-1), which also is associated with pulmonary arteriovenous fistulas and cyanosis. Infectious endocarditis may be associated with Osler nodes (see Fig. 76-2), Janeway lesions, or splinter hemorrhages (Fig. 51-11) (Chapter 76). Xanthomas (Fig. 51-12) are subcutaneous deposits of cholesterol seen on the extensor surfaces of the extremities or on the palms and digital creases; they are found in patients with severe hypercholesterolemia.

Laboratory Studies

All patients with known or suspected cardiac disease should have an ECG and chest radiograph. The ECG (Chapter 54) helps identify rate, rhythm, conduction abnormalities, and possible myocardial ischemia. The chest radiograph (Chapter 56) yields important information on chamber enlargement, pulmonary vasculature, and the great vessels.

Blood testing in patients with known or suspected cardiac disease should be targeted to the conditions in question. In general, a complete blood cell count, thyroid indices, and lipid levels are part of the standard evaluation. Point-of-care biomarker measurements in the emergency department can

decrease unnecessary admissions and reduce median length-of-stay. For example, among patients who are being evaluated for an acute MI, an undetectable high-sensitivity troponin level at presentation reduces the probability of acute MI to less than 1%.[5] A protocol in which the ECG and troponin level are repeated in 2 hours is as good as longer observation periods for evaluating patients with acute chest pain and suspected MI. However, the advent of high-sensitivity troponin assays has also greatly increased the risk for a false-positive diagnosis of MI,[5] especially because of chronic troponin elevations in many cardiac conditions and in elderly patients (Chapter 72).[6]

Echocardiography (Chapter 55) is the most useful test to analyze valvular and ventricular function. By use of Doppler flow methods, stenotic and regurgitant lesions can be quantified. Hand-held ultrasonography performed by generalists can improve the assessment of left ventricular function, cardiomegaly, and pericardial effusion. Transesophageal echocardiography is the preferred method to evaluate possible aortic dissection and to identify clot in the cardiac chambers. Radionuclide studies (Chapter 56) can measure left ventricular function, assess myocardial ischemia, and determine whether ischemic myocardium is viable. CT can detect coronary calcium, which is a risk factor for symptomatic coronary disease (Chapter 56). In the setting

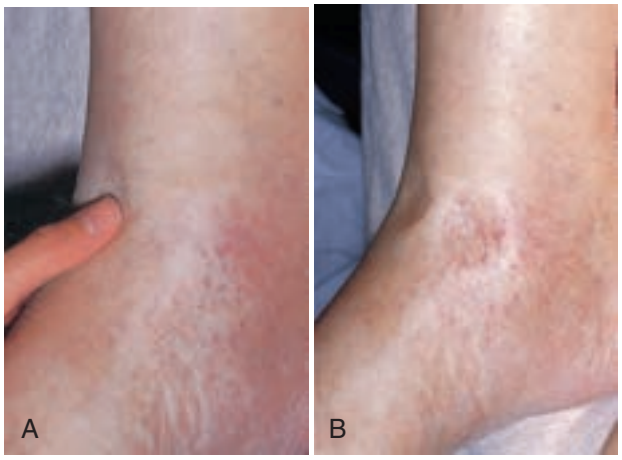

FIGURE 51-7. Pitting edema in a patient with cardiac failure. A depression ("pit") remains in the edema for some minutes after firm fingertip pressure is applied. (From Forbes CD, Jackson WD. *Color Atlas and Text of Clinical Medicine*. 3rd ed. London: Mosby; 2003.)

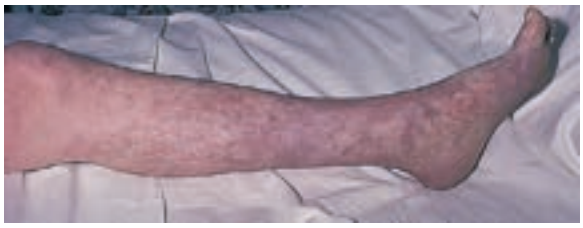

FIGURE 51-9. Arterial embolism causing acute ischemia and cyanosis of the leg. Initial pallor of the leg and foot was followed by cyanosis. (From Forbes CD, Jackson WD. *Color Atlas and Text of Clinical Medicine*. 3rd ed. London: Mosby; 2003.)

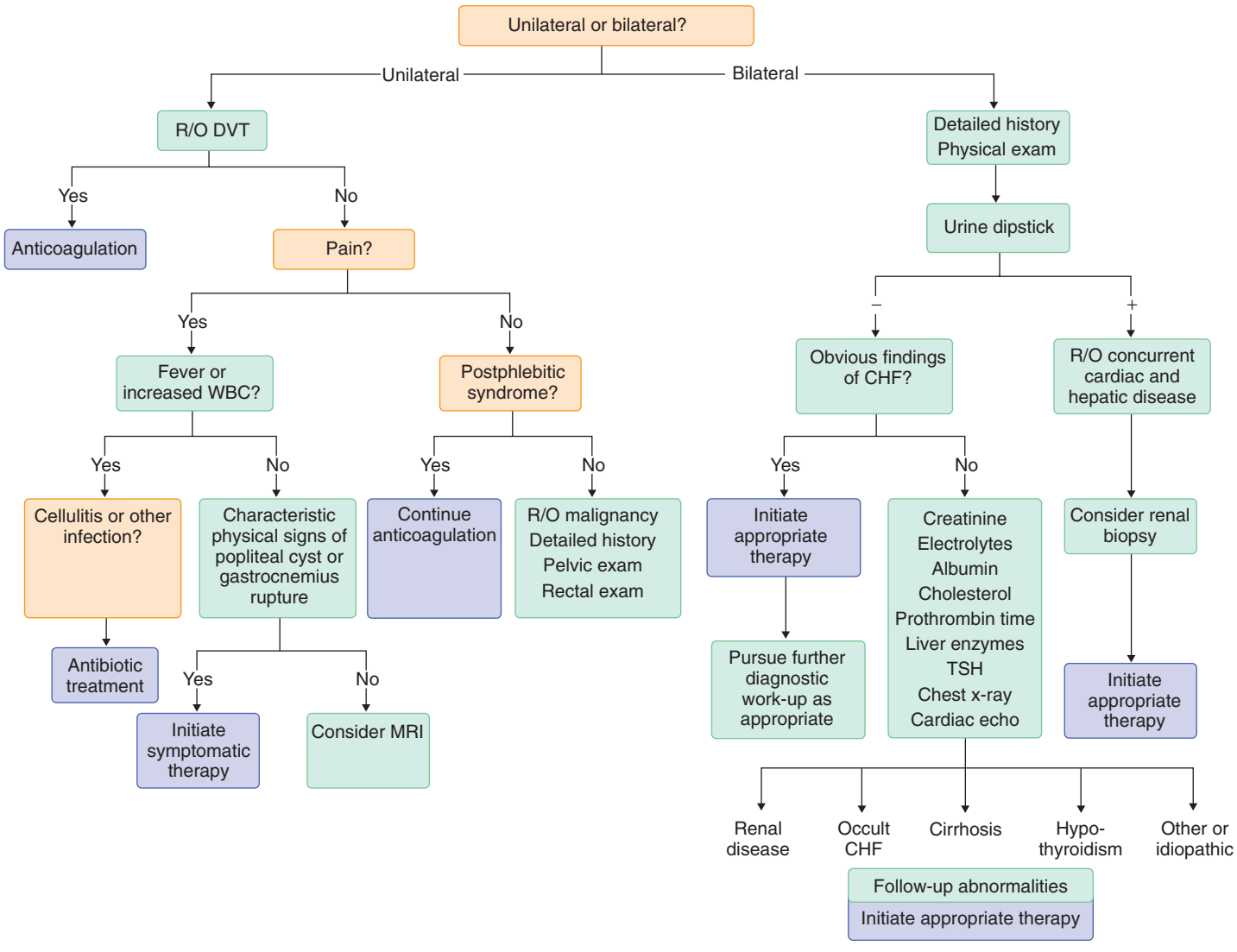

FIGURE 51-8. Diagnostic approach to patients with edema. CHF = congestive heart failure; DVT = deep vein thrombosis; MRI = magnetic resonance imaging; R/O = rule out; TSH = thyroid-stimulating hormone; WBC = white blood cell count. (From Chertow G. Approach to the patient with edema. In: Braunwald E, Goldman L, eds. *Primary Cardiology*. 2nd ed. Philadelphia: WB Saunders; 2003.)

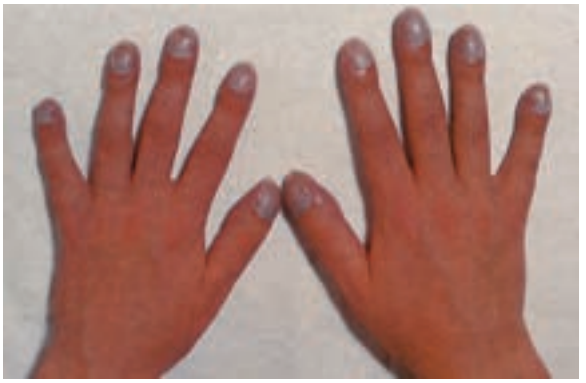

FIGURE 51-10. Severe finger clubbing in a patient with cyanotic congenital heart disease. (From Forbes CD, Jackson WD. *Color Atlas and Text of Clinical Medicine*. 3rd ed. London: Mosby; 2003.)

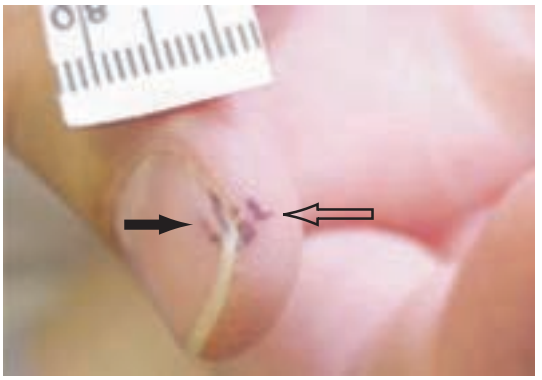

FIGURE 51-11. Splinter hemorrhage (*solid arrow*) and Janeway lesions (*open arrow*). These findings should stimulate a work-up for endocarditis. (Courtesy of Daniel L. Stulberg, MD.)

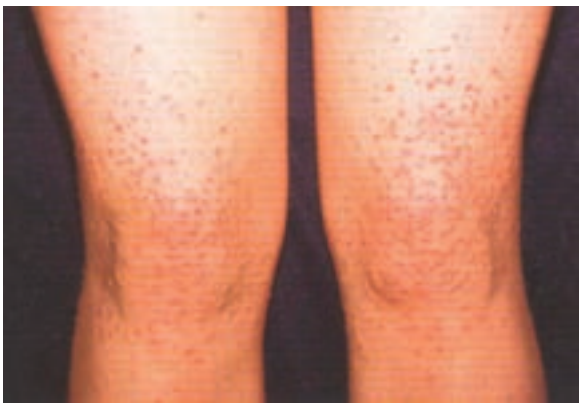

FIGURE 51-12. Eruptive xanthomas of the extensor surfaces of the lower extremities. This patient had marked hypertriglyceridemia. (From Massengale WT, Nesbitt LT Jr. Xanthomas. In: Bolognia JL, Jorizzo JL, Rapini RP, eds. *Dermatology*. Philadelphia: Mosby; 2003:1449.)

of acute chest pain, multislice CT is effective in diagnosing coronary disease.[A] In a randomized trial of emergency department patients at low to intermediate risk for a possible acute coronary syndrome, coronary CT angiography resulted in a higher rate of discharge from the emergency department (50% vs. 23), a shorter length of stay (median, 18 vs. 24.8 hours), and a higher rate of detection of coronary disease (9% vs. 3.5%) without any change in the rate of serious adverse events.[A] However, in a subsequent randomized trial of emergency department patients with symptoms suggestive of acute coronary syndromes but without ischemic ECG changes or an initially positive troponin test, incorporating coronary CT angiography into the triage strategy did not decrease overall costs of care.[A]

Stress testing by exercise or pharmacologic stress is useful to precipitate myocardial ischemia that may be detected by ECG abnormalities, perfusion abnormalities on radionuclide studies, or transient wall motion abnormalities on echocardiography.[7,8] These tests are often crucial in diagnosis of possible myocardial ischemia (Chapter 71) and in establishment of prognosis in patients with known ischemic heart disease. However, they are not recommended for the screening of asymptomatic individuals[9] or prior to participation in sports.[10]

Cardiac catheterization (Chapter 57) can measure precise gradients across stenotic cardiac valves, judge the severity of intracardiac shunts, and determine intracardiac pressures. Coronary angiography provides a definitive diagnosis of coronary disease and is a necessary prelude to coronary revascularization with a percutaneous coronary intervention or coronary artery bypass graft surgery (Chapter 74).

Continuous ambulatory ECG monitoring can help diagnose arrhythmias. A variety of newer technologies allow longer-term monitoring in patients with important but infrequently occurring symptoms (Chapter 62). Formal invasive electrophysiologic testing can be useful in the diagnosis of ventricular or supraventricular wide-complex tachycardia, and it is crucial for guiding a wide array of new invasive electrophysiologic therapies (Chapter 66).

SUMMARY

The history, physical examination, and laboratory evaluation should help the physician establish the cause of any cardiovascular problem; identify and quantify any anatomic abnormalities; determine the physiologic status of the valves, myocardium, and conduction system; determine functional capacity; estimate prognosis; and provide primary or secondary prevention. Key preventive strategies, including diet modification, recognition and treatment of hyperlipidemia, cessation of cigarette smoking, and adequate physical exercise, should be part of the approach to every patient, with or without heart disease.

Grade A References

A1. Than M, Aldous S, Lord SJ, et al. A 2-hour diagnostic protocol for possible cardiac chest pain in the emergency department: a randomized clinical trial. *JAMA Intern Med*. 2014;174:51-58.

A2. Goodacre SW, Bradburn M, Cross E, et al. The randomised Assessment of Treatment using Panel Assay of Cardiac Markers (RATPAC) trial: a randomised controlled trial of point-of-care cardiac markers in the emergency department. *Heart*. 2011;97:190-196.

A3. Litt HI, Gatsonis C, Snyder B, et al. CT angiography for safe discharge of patients with possible acute coronary syndromes. *N Engl J Med*. 2012;366:1393-1403.

A4. Hoffmann U, Truong QA, Schoenfeld DA, et al. Coronary CT angiography versus standard evaluation in acute chest pain. *N Engl J Med*. 2012;367:299-308.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Pellicori P, Kallvikbacka-Bennett A, Zhang J, et al. Revisiting a classical clinical sign: jugular venous ultrasound. *Int J Cardiol*. 2014;170:364-370.
2. Roberts KV, Brown AD, Maguire GP, et al. Utility of auscultatory screening for detecting rheumatic heart disease in high-risk children in Australia's Northern Territory. *Med J Aust*. 2013;199:196-199.
3. Clark AL, Cleland JG. Causes and treatment of oedema in patients with heart failure. *Nat Rev Cardiol*. 2013;10:156-170.
4. Rutherford JD. Digital clubbing. *Circulation*. 2013;127:1997-1999.
5. Storrow AB, Christenson RH, Nowak RM, et al. Diagnostic performance of cardiac troponin I for early rule-in and rule-out of acute myocardial infarction: Results of a prospective multicenter trial. *Clin Biochem*. 2014; [Epub ahead of print].
6. Korley FK, Jaffe AS. Preparing the United States for high-sensitivity cardiac troponin assays. *J Am Coll Cardiol*. 2013;61:1753-1758.
7. Mancini GB, Gosselin G, Chow B, et al. Canadian Cardiovascular Society guidelines for the diagnosis and management of stable ischemic heart disease. *Can J Cardiol*. 2014;30:837-849.
8. Mieres JH, Gulati M, Bairey Merz N, et al. Role of noninvasive testing in the clinical evaluation of women with suspected ischemic heart disease: a consensus statement from the American Heart Association. *Circulation*. 2014;130:350-379.
9. Chou R, Arora B, Dana T, et al. Screening asymptomatic adults with resting or exercise electrocardiography: a review of the evidence for the U.S. Preventive services task force. *Ann Intern Med*. 2011;155:375-385.
10. Sharma S, Estes NA 3rd, Vetter VL, et al. Clinical decisions: cardiac screening before participation in sports. *N Engl J Med*. 2013;369:2049-2053.

52

EPIDEMIOLOGY OF CARDIOVASCULAR DISEASE

DONALD M. LLOYD-JONES

Cardiovascular diseases are the leading cause of death, disability, and medical costs in the world, and they are expected to remain so for the foreseeable future. Cardiovascular disease manifests in a number of different ways, including congenital heart and vascular malformations (Chapter 69); coronary heart disease (Chapters 70, 71, 72, 73, and 74); heart failure (Chapter 59); cardiomyopathies (Chapter 60); valvular heart disease (Chapter 75); dysrhythmias (Chapters 62, 63, 64, and 65); pericardial diseases (Chapter 77); aortic (Chapter 78), peripheral (Chapter 79), and cerebrovascular diseases (Chapter 406); systemic hypertension (Chapter 67); vasculitides (Chapter 270); venous thromboembolic disease (Chapter 81); and pulmonary vascular hypertension (Chapter 68). Of these, coronary heart disease, stroke, and heart failure, which share many common underlying risk factors, have by far the largest impact on the population in terms of incidence, prevalence, quality of life, and medical costs.

BURDEN IN THE UNITED STATES

Cardiovascular diseases have been the leading cause of death in the United States in every year of the 20th and 21st century except for 1918, when the influenza epidemic surpassed them. Cardiovascular diseases account for 1 in 3 deaths in America annually, or about 790,000 deaths, including about 400,000 in women and about 390,000 in men.[1] The overall rate of death due to cardiovascular disease in the United States is about 230 per 100,000 persons, with higher rates in men than in women and in blacks than in whites. Because of secular trends over the past 40 to 50 years, coronary heart disease alone may soon fall below all cancers combined, but all cardiovascular diseases combined are expected to remain the leading causes of death in the United States and globally for the foreseeable future.

Cardiovascular diseases also are the leading cause of hospitalizations and medical costs in the United States. Each year, about 5.8 million Americans are hospitalized for cardiovascular disease, more than 1.3 million cases of which are due to coronary heart disease and another 1 million or more due to heart failure. The United States currently spends more than $300 billion annually on direct and indirect costs for cardiovascular diseases, and these total costs are projected roughly to triple to more than $1 trillion annually by 2030.

In the United States, about 15.4 million adults have coronary heart disease, roughly half of whom have had a myocardial infarction. Each year, Americans suffer more than 900,000 new and recurrent myocardial infarctions, with about 380,000 deaths due to coronary heart disease, a large percentage of which are sudden cardiac deaths. There are about 6.8 million stroke survivors in the United States, with 800,000 new or recurrent strokes occurring every year. Strokes are especially prominent in the so-called "stroke belt" in the southeastern United States, where many African Americans live. With aging, the risks for stroke and heart failure tend to increase earlier in women and African Americans than in white men, whose coronary risk increases earlier. At present, more than 5 million Americans suffer from chronic heart failure, with approximately equal numbers of men and women affected. However, the prevalence of heart failure is about twice as high in blacks as in whites.

GLOBAL BURDEN

Cardiovascular diseases, including coronary heart disease and stroke, became the leading cause of death and disability globally in the early 21st century.[2] About 80% of cardiovascular deaths and events now occur in low- and middle-income countries, and the onset of cardiovascular disease tends to be at an earlier age in these countries. For example, about 50% of coronary deaths occur before age 70 years in India, whereas only 25% occur by that age in high-income countries. Unfavorable global trends in eating patterns, high rates of smoking, and increasing burdens of obesity, diabetes, and hypertension are driving the burden of cardiovascular disease.[3] Whereas stroke was the dominant cause of death and disability in East Asian countries for decades owing to high sodium intake and resulting hypertension, recent changes in diet, activity levels, and smoking have made coronary heart disease an equivalent or greater health burden in this area of the world.

RISK FACTORS FOR CARDIOVASCULAR DISEASE

Established Risk Factors

A number of factors have been established for cardiovascular disease based on their strength and consistency of associations, specificity, temporality, and biologic plausibility.[4,5] Furthermore, these established risk factors explain the vast majority of risk for incident myocardial infarction. Longitudinal cohort studies demonstrate that 90% of individuals who suffer a myocardial infarction have at least one established clinical risk factor before their first event, and adverse levels of nine risk factors and behaviors collectively account for 90% or more of the risk for myocardial infarction in men and women, in older and younger individuals, and in all regions of the world. These nine risk factors and behaviors include smoking (Chapter 32), elevated apolipoprotein B–to–apolipoprotein A1 ratio (Chapter 206), hypertension (Chapter 67), diabetes (Chapter 229), abdominal obesity (Chapter 220), psychosocial factors, lower consumption of fruits and vegetables (Chapter 213), alcohol intake (Chapter 33), and physical inactivity (Chapter 16). Many of the established risk factors tend to cluster in a *metabolic syndrome*, which is characterized by abdominal obesity, insulin resistance, hyperglycemia, elevated blood pressure, elevated triglyceride levels, and lower high-density lipoprotein (HDL) cholesterol levels.

Age is the most powerful risk factor for the development of most cardiovascular diseases, especially stroke (Chapter 407), heart failure (Chapters 58 and 59), and atrial fibrillation (Chapter 64). Chronologic age represents a person's aggregate exposure to multiple physiologic and environmental effects on the cardiovascular system. The incidence of cardiovascular disease at least doubles with each additional decade of age in adulthood until the oldest ages, when the heavy burden of competing causes of mortality (Chapter 23) limits further progression.

The impact of a person's *sex* on cardiovascular disease is important. More women than men die of cardiovascular diseases annually. However, women tend to develop risk factors later in life than do men, and women's incidence rates lag men's by approximately 10 years. The precise contributions of sex hormones to these age trends are uncertain, but many women develop worsening risk factor levels, particularly with regard to lipids, blood pressure, weight, and insulin resistance, during and after the menopausal transition (Chapter 240).

Race per se is not thought to be an independent risk factor for cardiovascular disease, and the established causal risk factors have broadly similar effects in all race and ethnic groups. Nevertheless, hypertension tends to be more prevalent in individuals of African ancestry, especially in environments with higher sodium intake, and to have a somewhat stronger association with cardiovascular events, especially heart failure and stroke. Compared with whites, individuals of East Asian and South Asian descent have a greater risk for developing the metabolic syndrome, insulin resistance, and diabetes at a lower overall body mass index. However, some of the cardiovascular risk differences observed across race and ethnic groups can be attributed to differences in socioeconomic status, rather than race or ethnicity.

Blood lipid levels (Chapter 206), including the total serum cholesterol level and its subfractions, particularly low-density lipoprotein (LDL) cholesterol, have significant, continuous, and graded associations with the risk for coronary heart disease and peripheral arterial atherothrombotic disease. By comparison, independent associations of blood lipids with stroke and heart failure events are much weaker, indicating a potentially lesser role in the pathogenesis of these diseases when they occur independently of their relationship to coexisting coronary heart disease. Apolipoprotein B–containing particles make up the subpopulation of circulating cholesterol-containing particles that represent the atherogenic lipoprotein fractions. These particles are considered to be the central actors in the initiation and promotion of atherogenesis on the basis of a substantial body of epidemiologic, clinical, and basic science evidence. Among U.S. adults aged 20 years and older, 43% (or nearly 100 million) have total cholesterol levels above the desirable range of less than 200 mg/dL, and 14% (31 million) have elevated levels of 240 mg/dL or higher. Mean total cholesterol levels have been falling sharply in recent decades, mostly because of changes in dietary composition but also because of more widespread use of lipid-lowering medications. In the 1970s, mean total cholesterol concentrations were approximately 220 mg/dL, whereas currently they are just under 200 mg/dL. These improvements have been a major contributor to the decline in coronary death rates over the same time period. Randomized clinical trials have unequivocally established LDL cholesterol as a causal agent for coronary heart disease, and statins are effective at reducing rates of both coronary heart disease and stroke, significantly and substantially.[A1] By comparison, niacin is of no apparent added value[A2] and other medications are being actively investigated (Chapter 206).

Blood pressure (Chapter 67) has a continuous, graded association with incident coronary heart disease, stroke, and heart failure events. In worldwide studies of nearly 1 million individuals, the risk at every age for all types of cardiovascular disease death doubled with each 20-mm Hg higher systolic blood pressure and each 10-mm Hg higher diastolic blood pressure, beginning at a blood pressure of 115/75 mm Hg.[6] Although the relationship with outcomes is linear, hypertension is typically defined by blood pressures of 140 mm Hg or higher systolic or 90 mm Hg diastolic (Chapter 67). Using this definition, hypertension is the most prevalent modifiable cardiovascular risk factor worldwide. Among people who are normotensive at age 55 years, the remaining lifetime risk for development of hypertension is 90%. Approximately one third of all American adults currently have hypertension, and its prevalence has been increasing owing to the obesity epidemic. Hypertension has stronger relative associations with stroke and heart failure than with coronary heart disease, in part because of its effects on myocardial and cerebrovascular remodeling. In the United States, rates of treatment and control for hypertension have been gradually increasing. The effective treatment of hypertension reduces the risk for stroke, heart failure, and coronary heart disease events.[A3]

Cigarette smoking (Chapter 32) is one of the strongest risk factors for cardiovascular disease events. After adjustment for other risk factors, smoking confers two- to three-fold higher risk for all manifestations of cardiovascular

disease, especially coronary heart disease and peripheral arterial disease. Fortunately, consistent public health efforts have reduced the prevalence of smoking in the United States from about 45% in the 1960s to just under 20% currently. The prevalence of smoking remains higher in many European and Asian countries, and its continued increase in some parts of the world drives unfavorable trends in cardiovascular morbidity and mortality. A large body of evidence indicates that environmental exposure to tobacco smoke in nonsmokers ("second-hand" or "passive" smoking) also increases risk for cardiovascular events substantially (Chapter 32) and contributes to the population burden of disease. Substantial data also support the benefits of smoking cessation for reducing the risks for a subsequent coronary event and death.[7]

Overweight and obesity have been increasing in the United States and worldwide. Before 1985, fewer than 10% of Americans were obese, defined as having a body mass index of 30 kg/m^2 or higher. Now, however, about 35% of Americans are obese, and another 35% are overweight (Chapter 220). Major societal changes in the availability of food and in dietary content, coupled with reductions in physical activity, have produced this unprecedented epidemic. Although overweight and obesity themselves tend to be weak independent predictors of cardiovascular events in the short term, they are major drivers of elevated blood pressure, elevated blood glucose levels, and adverse lipid profiles that are themselves major contributors to the incidence of cardiovascular disease.[8]

Blood glucose and its surrogate marker, hemoglobin A1c, have a continuous and graded association with cardiovascular events. People with diabetes (Chapter 229), whether diagnosed or undiagnosed, have two- to three-fold higher adjusted risk for cardiovascular events compared with persons without diabetes, and they also have substantially higher risks for developing chronic renal disease (Chapter 130). Whereas diabetes was relatively uncommon before the 1980s, the obesity epidemic has led to a dramatic increase in the prevalence of type 2 diabetes and of impaired fasting glucose levels, termed *pre-diabetes*. At present in the United States, nearly 20 million people, representing more than 8% of all adults, have diagnosed diabetes, and another 8 million (about 3.5% of adults) have undiagnosed diabetes. Fully 87 million more adults, or about 38% of the adult U.S. population, currently have prediabetes. If current trends continue, an estimated 77% of men and 53% of women in the U.S. could have pre-diabetes by 2020. Diabetes affects nonwhite racial and ethnic groups, such as American Indians, African Americans, South Asians, East Asians, and Latinos, who appear to have greater sensitivity to insulin resistance at lower body mass index, in much greater proportions than whites. Unfortunately, tight control of glucose levels in persons with diabetes has not been associated with significant reductions in risk for macrovascular cardiovascular disease.[A4,A5]

Adverse diet (Chapter 213) is a major contributor to obesity, diabetes, hypertension, and hyperlipidemia. Healthy eating patterns emphasize a lower caloric intake and focus on fruits and vegetables, healthy fats from nuts and olive oil, lean sources of protein such as fish, whole grains, a reduced sodium intake, and limiting the intake of processed foods, unhealthy fats, and simple sugars. This eating pattern is typical of the "Mediterranean diet," which has been shown to be associated with a lower incidence of cardiovascular disease.[A6] By comparison, no vitamin or mineral supplement has been shown conclusively to reduce cardiovascular risk.[9]

Alcohol (Chapter 33) has a complex association with cardiovascular events. Moderate intake of one serving of alcohol per day is associated with a modestly lower risk for cardiovascular disease. At higher levels of intake, however, risks for total mortality, hypertension, stroke, and heart failure tend to increase.

Physical inactivity (Chapter 16) *and a sedentary lifestyle* are also significant risk factors for cardiovascular disease. Individuals who participate in no physical activity are at highest risk for events. The risk is significantly lower for people who participate in even minimal physical activity, and risks decrease further with greater activity levels, particularly to the extent that they contribute to improvement in objective physical fitness. The biology and risks of sedentary time may be more than just the absence of physical activity because sedentary lifestyle, measured best by the hours of time spent in front of a television or computer screen, seems to have an adverse effect independent of time spent doing physical activity.

Family history is clearly an important cardiovascular risk factor, independent of other measurable risk factors. However, ideal levels of cardiovascular health do not appear to be genetically programmed nor inexorably compromised as a consequence of aging. Data indicate that the heritability of ideal cardiovascular health is less than 20%, thereby suggesting strong environmental and behavioral influences on this trait.

Novel Risk Markers

Blood markers of inflammation, thrombosis, and target organ damage also appear to characterize the atherosclerotic process (Chapter 70). Serum biomarkers such as C-reactive protein, fibrinogen, plasminogen activator inhibitor-1, interleukin-6, and lipoprotein-associated phospholipase A_2 have significant associations with incident cardiovascular events that are independent of established risk factors.[10] However, because of their lack of specificity and their relatively weak independent associations with incident disease, none of these markers has yet proved useful for routine screening or for incorporation into risk assessment algorithms in primary or secondary prevention. To date, none has provided meaningful reclassification of risk in individuals after quantitative assessment using traditional established risk factors. Newer biomarkers that indicate the presence of existing target organ damage, such as high-sensitivity troponin or natriuretic peptide levels, hold promise for screening and targeting of prevention efforts in older, asymptomatic individuals (Chapter 23).

Noninvasive cardiac testing and imaging holds the potential for detecting preclinical disease and potentially guiding early intervention. For example, electrocardiographic evidence of left ventricular hypertrophy confers significant excess risk for coronary heart disease over and above the presence of hypertension and other risk factors. High levels of coronary calcification on computed tomography (CT) imaging of the heart (Chapter 56) or greater carotid intima-media thickness measured by B-mode ultrasound of the carotid arteries portends a higher risk for future cardiovascular events. Because these imaging markers detect evidence of the actual underlying diseases of interest (i.e., left ventricular hypertrophy or atherosclerosis), rather than nonspecific risk factors, they are more effective at identifying individuals at high risk for incident clinical events, such as heart failure, stroke, and myocardial infarction. Of the available modalities, CT screening for coronary artery calcification appears to be the best widely available means for detecting individuals at near-term risk. For example, in the Multi-Ethnic Study of Atherosclerosis, asymptomatic individuals with coronary artery calcium scores of more than 100 Agatston units had relative hazards for a coronary event that were 7- to 10-fold higher than in individuals without any coronary calcification, even after adjustment for major established risk factors.[11] Coronary calcium scoring also has been shown to be the most effective and reliable means for reclassifying risk after a quantitative risk assessment using established risk factors, with the ability to identify otherwise low-risk individuals who nonetheless will have a cardiovascular event. Although noninvasive screening for cardiovascular disease holds much promise for the future, its precise role remains uncertain at the present time (Chapter 56).

Assessment of Risk for Cardiovascular Disease

Estimation of Short-Term Risk

Adverse levels of any single risk factor or risk marker are associated with elevated risk for incident cardiovascular events. However, combinations of adverse risk factors are additive and sometimes synergistic for increasing risk. To improve the prediction of cardiovascular events and provide quantitative risk assessment, a number of multivariable risk equations or scores, such as the Framingham equations (E-Tables 52-1 and 52-2), have been developed. The vast majority of risk scores available have focused on predicting 10-year absolute risk, and essentially all include age, sex, smoking status, cholesterol, and blood pressure, with some also including diabetes, family history, body mass index, socioeconomic status, or novel biomarkers. The end points of interest for diverse risk equations have varied widely, from the prediction of cardiovascular death alone to the prediction of fatal and nonfatal major coronary events, major atherosclerotic events (coronary disease and stroke), and a broader range of cardiovascular events (including heart failure, coronary revascularization, angina, or claudication). For example, the 10-year risk for incident atherosclerotic cardiovascular disease can be predicted in 50-year-old men and women according to sex, race, and different levels of risk factors (Fig. 52-1), and the risks are dramatically higher with a greater risk factor burden.

Lifetime Risk Estimation

Despite the widespread use of 10-year risk estimates to guide prevention strategies, this approach has important limitations. For example, one consequence of the substantial weighting of age in 10-year risk equations is that younger men and women, even those with substantial risk factor burden, do not tend to have a high short-term risk. When treatment thresholds are applied to quantitative risk estimates for clinical guidelines, men younger

E-TABLE 52-1 FRAMINGHAM RISK SCORE FOR CARDIOVASCULAR DISEASE PREDICTION ACCORDING TO TRADITIONAL RISK FACTORS

POINTS	AGE (yr)	HDL-C (mg/dL)	TOTAL CHOLESTEROL (mg/dL)	SBP NOT TREATED	SBP TREATED	SMOKER	DIABETES
WOMEN							
–3				<120			
–2		60+					
–1		50-59			<120		
0	30-34	45-49	<160	120-129		No	No
1		35-44	160-199	130-139			
2	35-39	<35		140-149	120-129		
3			200-239		130-139	Yes	
4	40-44		240-279	150-159			Yes
5	45-49		280+	160+	140-149		
6					150-159		
7	50-54				160+		
8	55-59						
9	60-64						
10	65-69						
11	70-74						
12	75+						
MEN							
–2		60+		<120			
–1		50-59					
0	30-34	45-49	<160	120-129	<120	No	No
1		35-44	160-199	130-139			
2	35-39	<35	200-239	140-159	120-129		
3			240-279	160+	130-139		Yes
4			280+		140-159	Yes	
5	40-44				160+		
6	45-49						
7							
8	50-54						
9							
10	55-59						
11	60-64						
12	65-69						
13							
14	70-74						
15	75+						

HDL-C = high-density lipoprotein cholesterol; SBP = systolic blood pressure.
From Klag MJ. Epidemiology of cardiovascular disease. In: Goldman L, Schafer AI, eds. *Goldman's Cecil Medicine*. 24th ed. Philadelphia: Elsevier Saunders; 2012:256-260.

E-TABLE 52-2 CARDIOVASCULAR RISK ACCORDING TO TOTAL POINTS OF FRAMINGHAM RISK SCORE

WOMEN		MEN	
Points	***10-Year Risk (%)***	***Points***	***10-Year Risk (%)***
		≤–3	<1
≤–2	<1	–2	1.1
–1	1.0	–1	1.4
0	1.2	0	1.6
1	1.5	1	1.9
2	1.7	2	2.3
3	2	3	2.8
4	2.4	4	3.3
5	2.8	5	3.9
6	3.3	6	4.7
7	3.9	7	5.6
8	4.5	8	6.7
9	5.3	9	7.9
10	6.3	10	9.4
11	7.3	11	11.2
12	8.6	12	13.2
13	10.0	13	15.6
14	11.7	14	18.4
15	13.7	15	21.6
16	15.9	16	25.3
17	18.5	17	29.4
18	21.5	18+	>30
19	24.8	—	—
20	28.5	—	—
21+	>30	—	—

From Klag MJ. Epidemiology of cardiovascular disease. In: Goldman L, Schafer AI, eds. *Goldman's Cecil Medicine*. 24th ed. Philadelphia: Elsevier Saunders; 2012:256-260.

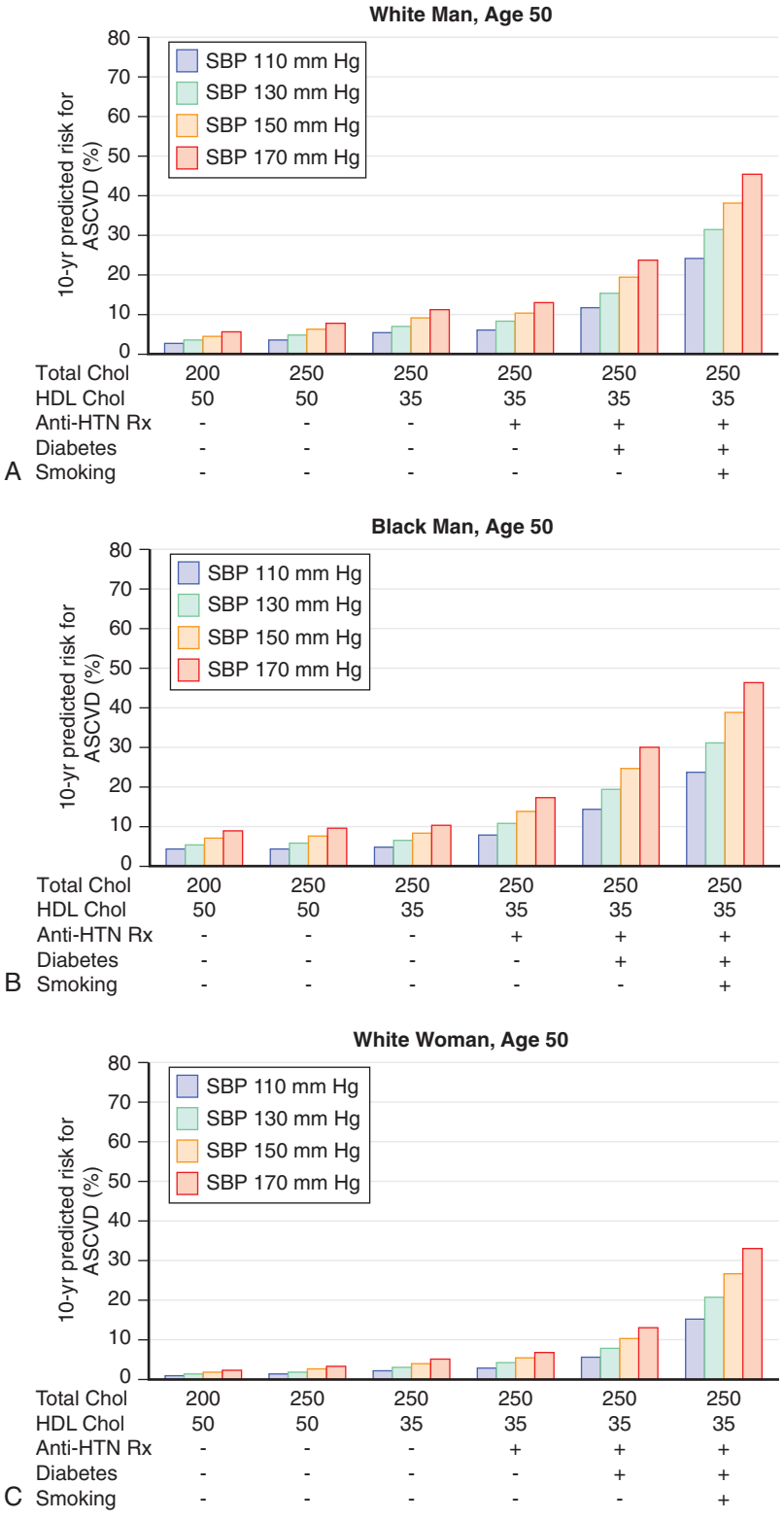

FIGURE 52-1. Predicted 10-year risks for atherosclerotic cardiovascular disease (ASCVD), including fatal coronary heart disease, nonfatal myocardial infarction, and fatal or nonfatal stroke, as a function of selected risk factor levels in a 50-year-old black man (**A**), white man (**B**), white woman (**C**), or black woman (**D**). Chol = cholesterol; HDL = high-density lipoprotein; HTN = hypertension; Rx = medication. (Predicted risks are derived from the Pooled Cohort Equations from the 2013 American College of Cardiology/American Heart Association Guideline on the Assessment of Cardiovascular Risk. Goff DC Jr, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2014;63(25 Pt B):2935-2959.)

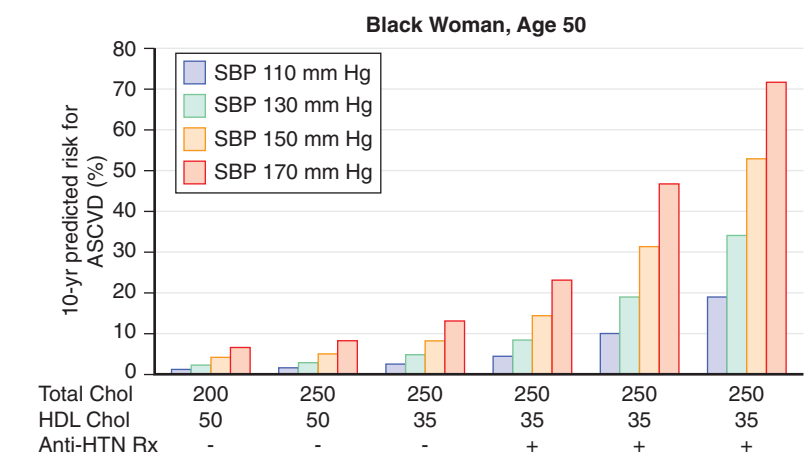

D

FIGURE 52-1, cont'd.

than 50 years and women younger than 60 years will infrequently exceed those thresholds. Therefore, recent guidelines have considered longer risk horizons, such as 30 years or the remaining lifespan.[12] The established risk factors are all associated with lifetime risks for cardiovascular disease, but the nature of the relationships sometimes differs from short-term associations because of competing risks. For example, smoking is a strong risk factor for near-term cardiovascular events but is a weaker predictor of lifetime risk for cardiovascular events because of the simultaneous and substantial risk for cancer death, which limits the lifetime risk for cardiovascular disease among smokers. As a result, lifetime risk estimates may enhance communication for individual patients, but how they should be used in decision making regarding the institution of preventive drug therapy is less certain.

Prevention of Cardiovascular Disease

Because patients with prevalent symptomatic cardiovascular disease are at the highest risk, preventive interventions such as lifestyle modifications and drug therapy to reduce risk are most effective and cost-effective when used as therapy for secondary prevention of recurrent events. Examples of intensive lifestyle modification and of proven therapies include aspirin,[A7] statins,[13,■■] and antihypertensive medications,[14] as well as other medications and implantable devices that may prevent complications such as heart failure or fatal ventricular dysrhythmias.

For primary prevention, asymptomatic individuals may be at risk because of a heritable family history of premature cardiovascular disease, one or more markedly elevated risk factors, or multiple modestly elevated risk factors. The current primary prevention paradigm is to match the intensity of prevention efforts to the absolute risk of the patient. Appropriate lifestyle interventions (e.g., smoking cessation, weight loss, dietary modification) are recommended for all individuals, whereas drug therapy is recommended only for individuals in whom the absolute benefits can be expected to outweigh any potential adverse drug effects and to be cost-effective in doing so. Treatment that restores optimal risk factor levels does not always imply that the treated individual will now have the very low incidence rates observed in people who have maintained optimal risk factor levels throughout young adulthood and into middle age. As a result, another concept is *primordial prevention*, which is the prevention of the development of risk factors in the first place. Primordial prevention requires a focus on health behaviors that may prevent the development of dyslipidemia, diabetes, and hypertension, as well as population-level strategies that attempt to create an environment conducive to favorable health behaviors.

Cardiovascular Health: A New Paradigm

After decades of declining mortality rates from cardiovascular diseases and stroke in the United States, the new goal is to promote cardiovascular health in individuals and the population, monitor it over time, and improve it by concerted action. Central to the concept of cardiovascular health is the observation that optimal levels of seven health behaviors and health factors (Table 52-1) are associated with ideal cardiovascular health. Although about 40% of American adults believe they are in ideal cardiovascular health, fewer than 1% have all seven metrics at ideal levels, principally because of a poor-quality diet.[15] Persons who maintain high levels of cardiovascular health from young age to middle age have extremely favorable outcomes from middle to older ages, including a markedly increased longevity; a better quality of life; a substantially lower incidence of fatal and nonfatal cardiovascular disease events; a lower incidence of other chronic diseases of aging, including cancer and venous thromboembolism; a lower burden of subclinical atherosclerosis (e.g., carotid intima-media thickness, coronary artery calcification); higher levels of cognitive function in middle and older ages; and reduced medical care costs. These outcomes are observed in all segments of the population, as well as across all ages and both sexes.

People who pursue healthy lifestyles from young adulthood to middle age are far more likely to preserve ideal cardiovascular health than those who pursue none: 60% of the former group compared with only 3% of the latter group maintained ideal cardiovascular health factors into middle age. Cardiovascular health promotion thus represents a major paradigm shift and opportunity in public health efforts.

Future of Cardiovascular Epidemiology

Decades of success in observational epidemiology research continue to provide novel insights into trends and risk markers for cardiovascular disease, as well as the influence of in utero and early life exposures on the life course of cardiovascular diseases. New techniques to characterize environmental and behavioral exposures, physiology, health status, and precursors of disease include functional genomics, proteomics, metabolomics, and high-resolution imaging. With these tools, epidemiologic research has advanced to improve the characterization of the life course of cardiovascular diseases in living individuals and populations. For example, studies of the genotypes of individuals at the extremes of the distribution of LDL cholesterol levels have led to the discovery of polymorphisms in a novel gene termed proprotein convertase subtilisin/kexin type 9 (*PCSK9*). Although such polymorphisms are uncommon, specific missense and nonsense mutations in white and African American men and women are associated with substantially lower lifelong levels of LDL cholesterol. In turn, individuals with these polymorphisms have a 47 to 88% lower incidence of coronary heart disease over 15 years of follow-up through middle age compared with individuals without them. *PCSK9* has since become a novel potential therapeutic target.

A second emerging focus in cardiovascular epidemiologic research has been the study of effects of interventions in populations through public health and social policies. For example, studies have demonstrated marked reductions in hospitalizations for acute myocardial infarction occurring rapidly after the initiation of indoor smoking bans in diverse settings. Modeling studies have synthesized data from numerous epidemiologic sources to show that approximately 50 to 75% of the reductions in coronary death rates in Western countries may be attributable to population changes in risk factor levels, despite being offset by a recent worsening in obesity and diabetes prevalence, with the remainder likely attributable to advances in medical and surgical therapies.

TABLE 52-1 DEFINITIONS OF POOR, INTERMEDIATE, AND IDEAL CARDIOVASCULAR HEALTH FOR EACH OF SEVEN METRICS, AND UNADJUSTED PREVALENCE IN THE UNITED STATES

	POOR HEALTH		INTERMEDIATE HEALTH		IDEAL HEALTH	
Goal/Metric	***Definition***	***Prevalence %***	***Definition***	***Prevalence %***	***Definition***	***Prevalence %***
CURRENT SMOKING						
Adults >20 yr of age	Yes	24	Former ≤12 mo	3	Never or quit >12 mo	73 (51 never; 22 former >12 mo)
Children 12-19 yr of age	Tried prior 30 days	17			Never tried; never smoked whole cigarette	83
BODY MASS INDEX						
Adults >20 yr of age	≥30 kg/m^2	34	25-29.9 km/m^2	33	<25 kg/m^2	33
Children 2-19 yr of age	>95th percentile	17	85th-95th percentile	15	<85th percentile	69
PHYSICAL ACTIVITY						
Adults >20 yr of age	None	32	1-149 min/wk moderate intensity or 1-74 min/wk vigorous intensity or 1-149 min/wk moderate + vigorous	24	≥150 min/wk moderate intensity or ≥ min/wk vigorous intensity or ≥150 min/wk moderate + vigorous	44
Children 2-19 yr of age	None	10	>0 and <60 min of moderate or vigorous activity every day	46	≥60 min of moderate of vigorous activity every day	44
HEALTHY DIET SCORE						
Adults >20 yr of age	0-1 components	76	2-3 components	24	4-5 components	<0.5
Children 5-19 yr of age	0-1 components	91	2-3 components	9	4-5 components	<0.5
TOTAL CHOLESTEROL						
Adults >20 yr of age	≥240 mg/dL	16	200-239 mg/dL or treated to goal	38 (27; 12 treated to goal)	<200 mg/dL	45
Children 6-19 yr of age	≥200 mg/dL	9	170-199 mg/dL	25	<170 mg/dL	67
BLOOD PRESSURE						
Adults >20 yr of age	SBP ≥140 or DBP ≥90 mm Hg	17	SBP 120-139 or DBP 80-89 mm Hg or treated to goal	41 (28; 13 treated to goal)	<120/<80 mm Hg	42
Children 8-19 yr of age	>95th percentile	5	90th-95th percentile or SBP ≥120 or DBP ≥80 mm Hg	13	<90th percentile	82
FASTING PLASMA GLUCOSE						
Adults >20 yr of age	≥126 mg/dL	8	100-125 mg/dL or treated to goal	34 (32; 3 treated to goal)	<100 mg/dL	58
Children 12-19 yr of age	≥126 mg/dL	0.5	100-125 mg/dL	18	<100 mg/dL	81

DBP = diastolic blood pressure; SBP = systolic blood pressure.
From National Health and Nutrition Examination Survey (NHANES) data and the American Heart Association. Reproduced from Lloyd-Jones DM, Hong Y, Labarthe D, et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation*. 2010;121:586-613.

Grade A References

A1. Baigent C, Blackwell L, Emberson J, et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. *Lancet*. 2010;376:1670-1681.

A2. Landray MJ, Haynes R, Hopewell JC, et al. Effects of extended-release niacin with laropiprant in high-risk patients. *N Engl J Med*. 2014;371:203-212.

A3. Thomopoulos C, Parati G, Zanchetti A. Effects of blood pressure lowering on outcome incidence in hypertension. 1. Overview, meta-analyses, and meta-regression analyses of randomized trials. *J Hypertens*. 2014;32:2285-2295.

A4. Fullerton B, Jeitler K, Seitz M, et al. Intensive glucose control versus conventional glucose control for type 1 diabetes mellitus. *Cochrane Database Syst Rev*. 2014;2:CD009122.

A5. Hemmingsen B, Lund SS, Gluud C, et al. Targeting intensive glycaemic control versus targeting conventional glycaemic control for type 2 diabetes mellitus. *Cochrane Database Syst Rev*. 2013;11:CD008143.

A6. Estruch R, Ros E, Salas-Salvado J, et al. Primary prevention of cardiovascular disease with a Mediterranean diet. *N Engl J Med*. 2013;368:1279-1290.

A7. Sutcliffe P, Connock M, Gurung T, et al. Aspirin for prophylactic use in the primary prevention of cardiovascular disease and cancer: a systematic review and overview of reviews. *Health Technol Assess*. 2013;17:1-253.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics—2014 update: a report from the American Heart Association. *Circulation*. 2014;129:e28-e292.
2. Murray CJ, Lopez AD. Measuring the global burden of disease. *N Engl J Med*. 2013;369:448-457.
3. Yusuf S, Rangarajan S, Teo K, et al. Cardiovascular risk and events in 17 low-, middle-, and high-income countries. *N Engl J Med*. 2014;371:818-827.
4. Lewington S, Whitlock G, Clarke R, et al. Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. *Lancet*. 2007;370:1829-1839.
5. Eckel RH, Jakicic JM, Ard JD, et al. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129:S76-S99.
6. Lewington S, Clarke R, Qizilbash N, et al. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*. 2002;360:1903-1913.
7. Shields M, Wilkins K. Smoking, smoking cessation and heart disease risk: a 16-year follow-up study. *Health Rep*. 2013;24:12-22.
8. Jahangir E, De Schutter A, Lavie CJ. The relationship between obesity and coronary artery disease. *Transl Res*. 2014;164:336-344.
9. Fortmann SP, Burda BU, Senger CA, et al. Vitamin and mineral supplements in the primary prevention of cardiovascular disease and cancer: an updated systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2013;159:824-834.
10. Kaptoge S, Di Angelantonio E, Pennells L, et al. C-reactive protein, fibrinogen, and cardiovascular disease prediction. *N Engl J Med*. 2012;367:1310-1320.
11. Criqui MH, Denenberg JO, Ix JH, et al. Calcium density of coronary artery plaque and risk of incident cardiovascular events. *JAMA*. 2014;311:271-278.
12. Goff DC Jr, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014;63(25 Pt B):2935-2959.
13. Stone NJ, Robinson J, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014;63(25 Pt B):2889-2934.
14. James PA, Oparil S, Carter BL, et al. 2014 Evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014;311:507-520.
15. Lloyd-Jones DM, Hong Y, Labarthe D, et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation*. 2010;121:586-613.

53

CARDIAC FUNCTION AND CIRCULATORY CONTROL

ANDREW R. MARKS

The heart has the daunting task of pumping sufficient amounts of blood to meet both its own metabolic demands and those of the other organs. Uniquely among all the organs, the heart's failure to perform its task for even a few minutes causes death. The heart continuously fulfills this physiologic role with a variety of electrical, contractile, and structural functions that control the flow of blood to the organs.

STRUCTURE OF THE HEART

Cardiac Development

In humans, the formation of a linear heart tube from the primary cardiac crescent occurs between days 21 and 23 of gestation. Looping of the heart

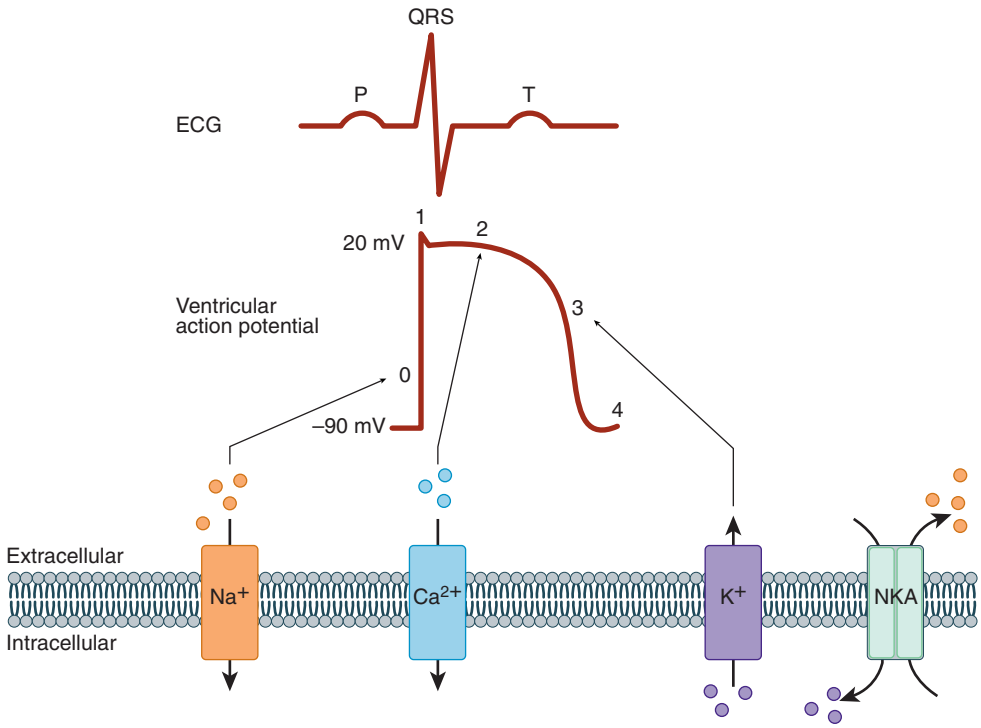

FIGURE 53-1. Cardiac action potential and ion channels. Myocardial contraction begins when sodium channels open and positively charged sodium ions flow into the cell and cause membrane depolarization (phase 0). During phases 1, 2, and 3, calcium ions flow into the cell through L-type calcium channels, while potassium flows out of the cell through voltage-gated potassium channels. These three phases correspond to the myocardial contraction, which corresponds to the QRS complex on the surface electrocardiogram (ECG). The sodium-potassium adenosine triphosphatase (NKA) helps return the system to its resting state.

tube and trabecular formation of the ventricle occur at 26 days of gestation (E-Fig. 53-1). At 6 weeks, the embryonic interventricular communication closes, followed by thickening and remodeling of the ventricular walls in the first trimester. By the end of week 7, heart development is essentially finished, although the heart continues to enlarge throughout gestation.[1,2]

Electrical Cells

The heart is a muscular pump controlled by regular electrical discharges from specialized muscle cells in the conduction system (Chapter 61). The molecular basis for the electrical activity of the heart is the activation of specific ion-conducting channels (Fig. 53-1). Coordinated activation and inactivation of cardiac ion channels regulate the membrane potential of the cardiac cells, thereby resulting in a rapid sequence of depolarization followed by repolarization. This electrical activity, which is manifested on the body surface as the electrocardiogram (ECG), is known as the action potential, and it is responsible for activating the contraction of the cardiac muscle. At a typical heart rate of 70 beats per minute, the heart beats about 100,000 times per day, or 37 million beats a year, corresponding to 3 billion beats during a lifespan of 80 years. Failure to propagate the signal throughout the heart (e.g., heart block) or abnormal rhythms (arrhythmias) that are either too slow (bradycardia) or too fast (tachycardia) can result in death (Chapter 62). Studies indicate that cardiac arrhythmias may be triggered by leak of calcium inside the cardiomyocytes, thereby suggesting a possible novel therapeutic target for a new generation of antiarrhythmic agents.

Ion Channels

Sodium, potassium, and calcium channels determine the electrical activity of the heart by opening and closing in a highly choreographed pattern that determines the action potential of the heart. The electrical regulation of the heart, which is reflected in the relative concentrations of ions inside and outside the heart muscle cells, determines the five phases of the action potential. The action potential is initiated when the opening of sodium channels results in a rapid influx of sodium (phase 0) down its concentration gradient (~145 mmol outside the heart muscle cell, ~10 mmol inside). After a brief early repolarization due to activation of potassium channels (phase 1), the rapid sodium influx depolarizes the cell, thereby activating calcium channels that allow calcium influx (phase 2) down its concentration gradient (~3 mmol outside, ~100 nmol inside). This calcium influx triggers excitation-contraction coupling that results in pumping by the heart. Potassium channels then open and cause repolarization (phase 3) as potassium fluxes out of the cell down its concentration gradient (~4 mmol outside, ~135 mmol inside). The membrane potential returns to the resting level of about −90 mV (phase 4).

Conduction System

Specialized pacemaker cells in the sinoatrial node (Fig. 53-2) have slightly higher (less negative) resting potentials and gradually depolarize during phase 5 owing to the activity of the potassium and calcium channels and the hyperpolarization-activated cyclic nucleotide–gated channels that are responsible for a small inward (depolarizing) current. In the normal heart, pacemaker cells are the first cells to depolarize, and they trigger the subsequent depolarization of the cells in specialized conducting fibers that propagate the electrical signal throughout the heart muscle in a highly regular and integrated fashion. Electrical activation (depolarization) spreading through the atria to the atrioventricular (AV) node is reflected as the P wave on the ECG (Chapter 54). The slowing of conduction in the AV node accounts for the PR interval on the ECG. After passing through the AV node, the depolarizing signal enters the bundle of His, where conduction is rapid. The bundle of His divides into the right and left bundle branches, which conduct the depolarizing signals into the ventricles and account for the QRS complex on the ECG. Repolarization is represented by the ST segment and the T and U waves of the ECG.

Contractile Cells

Heart muscle is composed of millions of individual cells known as cardiomyocytes, which contain an elaborate machinery required for coordinated contraction that pumps blood. Each cardiomyocyte is connected to its neighbors through specialized junctions that enable them to work as a single contractile unit.

The cardiomyocytes are filled with specialized contractile proteins arranged in highly regulated units, called sarcomeres, that give the muscles characteristic patterns known as striations (E-Fig. 53-2). Hence, like skeletal muscle, cardiac muscle is termed striated, as opposed to smooth muscles that form the vasculature and other organs such as the bladder, uterus, and stomach. Cardiomyocytes are also loaded with mitochondria that provide the energy (adenosine triphosphate [ATP]) required to fuel the heart's lifelong contractions (systole) and relaxations (diastole).

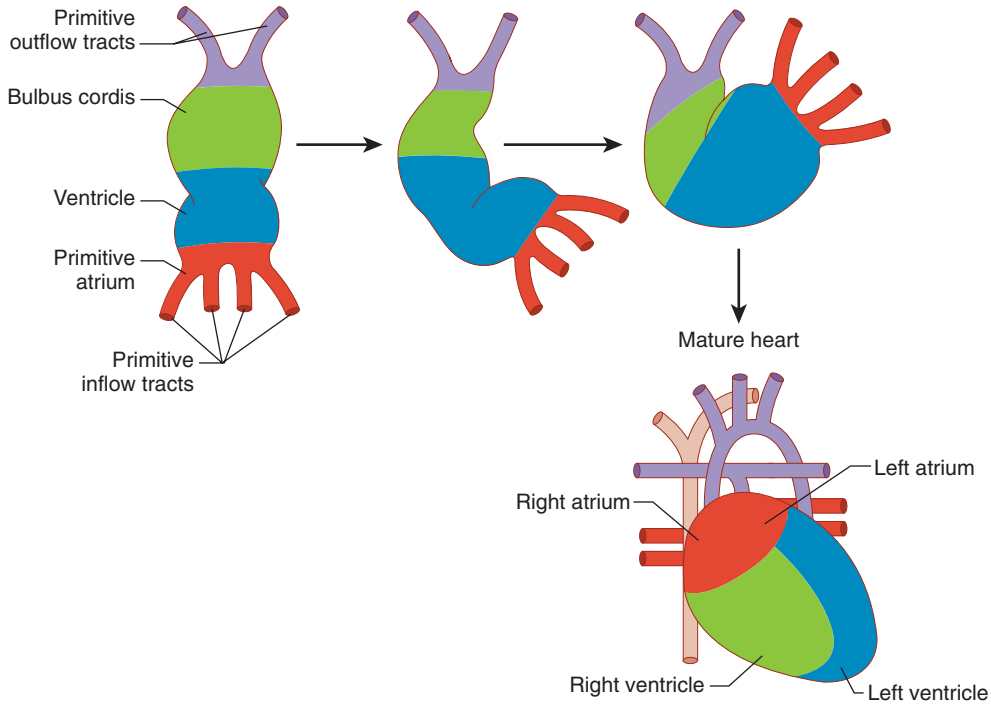

E-FIGURE 53-1. Folding of the heart tube during development. The primitive outflow tracts develop into the pulmonary artery and aorta. The left and right atria, along with the pulmonary veins and the superior and inferior vena cava, arise from the primitive inflow tracts. The area covered by the bulbus cordis and ventricle contributes to the development of the left and right ventricles.

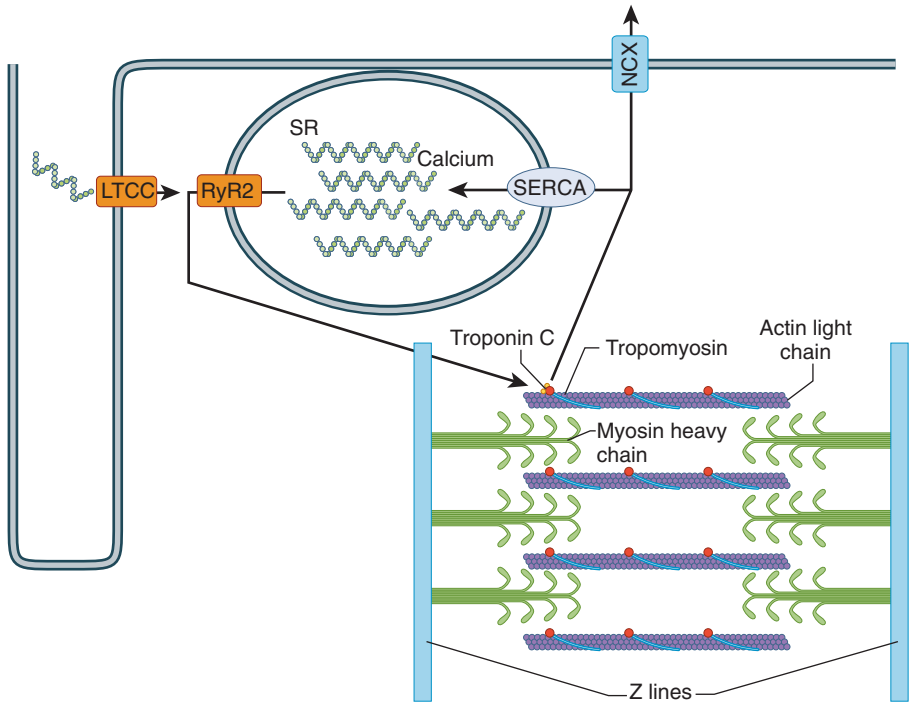

E-FIGURE 53-2. Sarcomere and excitation-contraction coupling. Depolarization of the cell membrane leads to calcium influx through L-type calcium channels (LTCC), which activate ryanodine receptors (RyR2) to release the large calcium stores from the sarcoplasmic reticulum (SR). Calcium binds to troponin C and induces a conformational change in tropomyosin, which enables myosin heads to interact with actin filaments and to cause contraction. During relaxation, calcium is pumped back into the sarcoplasmic reticulum through the sarcoendoplasmic reticulum calcium ATPase (SERCA) and out of the cell through the sodium-calcium exchanger (NCX).

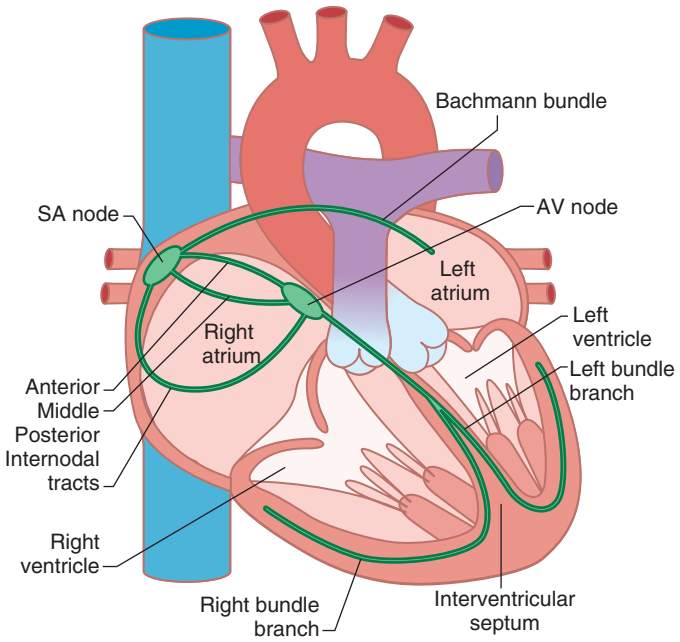

FIGURE 53-2. Cardiac anatomy. Cardiac anatomy comprises electrical and structural components. The electrical impulse that directs cardiac contraction originates in the sinoatrial (SA) node and is rapidly conducted through the atria by specialized conduction tracts. The impulses merge at the atrioventricular (AV) node, where, after a brief pause, they are rapidly conducted into the ventricles through the bundle of His, which is composed of specialized Purkinje cells. Blood moves from the atria into the ventricles through the tricuspid and mitral valves respectively, during diastole. During systole, blood from the ventricles is pumped into the pulmonary artery and aorta through the pulmonic and aortic valves, respectively.

Ultrastructure

The basic unit of the contractile system is the sarcomere, which is defined anatomically as the distance between two Z lines that anchor thin filaments composed of actin, tropomyosin, and troponin. Thin filaments slide past thick filaments (composed of myosin and titin) in a calcium-dependent manner to shorten the sarcomere length. The contractile proteins are surrounded by a calcium-filled membrane called the sarcoplasmic reticulum. The sarcoplasmic reticulum forms specialized associations with the transverse tubules, which are invaginations of the plasma membrane and contain voltage-gated calcium channels. When the muscle is activated by depolarization of its membrane, this electrical signal travels deep into the muscle through the transverse tubules. Inside the muscle, the electrical depolarizing signal activates the voltage-gated channels, which open to allow a small amount of calcium to enter the muscle cells. This influx of calcium in turn activates the type 2 ryanodine receptor (RyR2), calcium-release channels on the sarcoplasmic reticulum. The RyR channels open and release enough calcium from the sarcoplasmic reticulum to raise the calcium concentration in the myoplasm about 10-fold. As a result, calcium binds to troponin C in the thin filaments and causes a conformational change that enables cross-bridging between actin and myosin, thereby leading to sliding of the filaments, shortening of the sarcomere, and muscle contraction. Hydrolysis of ATP provides the energy required for the generation of force by the actin-myosin interaction. The conversion of electrical energy (depolarization of the cell membrane) to mechanical energy is known as excitation-contraction coupling. Relaxation of the heart muscle occurs when calcium is pumped back into the sarcoplasmic reticulum through the sarcoendoplasmic reticulum ATPase.

Signals That Regulate Contraction

Contractile force can be enhanced during stress by activation of the β-adrenergic pathway, which increases both the amount of calcium released and the rate of calcium uptake in the sarcoplasmic reticulum (E-Fig. 53-3). β-Agonists (e.g., epinephrine or norepinephrine) bind to β-adrenergic receptors to activate adenylyl cyclase, which generates cyclic adenosine monophosphate and activates protein kinase A. Protein kinase A phosphorylates phospholamban, the voltage-gated calcium channel, the ryanodine receptor, and sarcomeric regulatory proteins, thereby resulting in increased release of calcium from the sarcoplasmic reticulum and enhanced contractility of the heart.

Nonmuscle Cells

Although the heart is a muscular pump, 60 to 70% of its cells are cardiac fibroblasts, not muscle cells. These fibroblasts provide critical components of the extracellular matrix that determine the structure of the heart. Collagen, which is produced by the cardiac fibroblasts, is a major component of the extracellular matrix, where it forms a network that surrounds the cardiomyocytes and creates tissue that is able to withstand the stress of constant pumping. In certain pathologic conditions, including hypertension, myocardial infarction, and heart failure, the cardiac fibroblasts respond to stress by generating excess extracellular matrix, resulting in fibrosis, which can impair cardiac function.[3] Indeed, a number of commonly used therapies for heart disease, including lipid lowering with statins or fibrates and antihypertensive treatments with angiotensin-converting enzyme inhibitors, β-blockers, and angiotensin receptor blockers, exert part of their beneficial effects on cardiac fibroblasts by reducing fibrosis, thereby resulting in favorable "reverse" remodeling of the heart. Resident fibroblast lineages mediate pressure overload-induced cardiac fibrosis.[4]

ANATOMY OF THE HEART

The primary pumping chamber of the heart is the thick-walled left ventricle, which is composed of billions of cardiomyocytes connected end to end through gap junctions. The right ventricle is a thinner-walled chamber, divided from the left ventricle by the *interventricular septum*. Above the ventricles are the right and left atria, which are thin-walled chambers that receive low-pressure venous blood; they are separated from the ventricles by the *tricuspid valve* on the right side and the *mitral valve* on the left side. These valves are attached to *papillary muscles* that emerge from the ventricular walls through *chordae tendineae*. The pressure gradient between the ventricles and the atria opens the AV valves. The papillary muscles help establish the positions of the valve leaflets and prevent regurgitant flow during contraction. The *aortic and pulmonary valves* separate the left and right ventricles from their arterial connections and enable blood flow out of the ventricles.[5]

Coronary Blood Flow

The coronary arteries receive blood from the aorta, directly above the aortic valve, and travel through the epicardium that surrounds the heart to supply blood to the heart muscle (see Fig. 57-4). The diastolic blood pressure in the ascending aorta just above the aortic valve determines most of the flow of blood into the normal (nonstenosed) coronary arteries while the heart is relaxed. During systole, coronary flow is determined by the left ventricular intracavitary pressure, which equals the pressure within the inner myocardial wall, where coronary arteries are compressed during systole. Coronary blood flows to the epicardium during both systole and diastole but flows to the endocardium predominantly during diastole.

Metabolic Regulation of the Cardiovascular System

Cardiac muscle requires constant coronary perfusion to supply oxygen and other metabolites. Increased energy consumption due to enhanced contractility necessitated by increased pressures or higher heart rates (e.g., during exercise) can be met only by increased coronary blood flow. Signals that augment coronary blood flow (by up to six-fold) include nitric oxide, adenosine, bradykinins, prostaglandins, and carbon dioxide. The breakdown of ATP is the source of adenosine, whereas nitric oxide is produced by the action of nitric oxide synthases that metabolize the amino acid L-arginine. Autoregulatory mechanisms, including constriction in response to increased luminal pressures and dilation in response to reduced pressure, also play a role in determining coronary artery blood flow. Other metabolic factors that cause vasoconstriction include endothelin peptides, serotonin, 5-hydroxytryptamine, thromboxane, angiotensin II, and β_1-adrenergic stimulation.

Sympathetic and parasympathetic pathways of the autonomic nervous system and the renin-angiotensin system exert potent regulatory effects on cardiovascular function. The sympathetic nervous system plays the key role in the response to stress (e.g., the fight-or-flight response) by increasing heart rate and myocardial contractility and decreasing vascular tone. Regulation of cardiovascular function by the sympathetic nervous system is mediated by norepinephrine that is released at the nerve endings and by epinephrine from the adrenal gland. β-Adrenergic signaling is mediated by epinephrine, which increases the heart rate and vasodilates the central arterial bed, thereby resulting in reduced afterload, which in turn helps augment cardiac output.

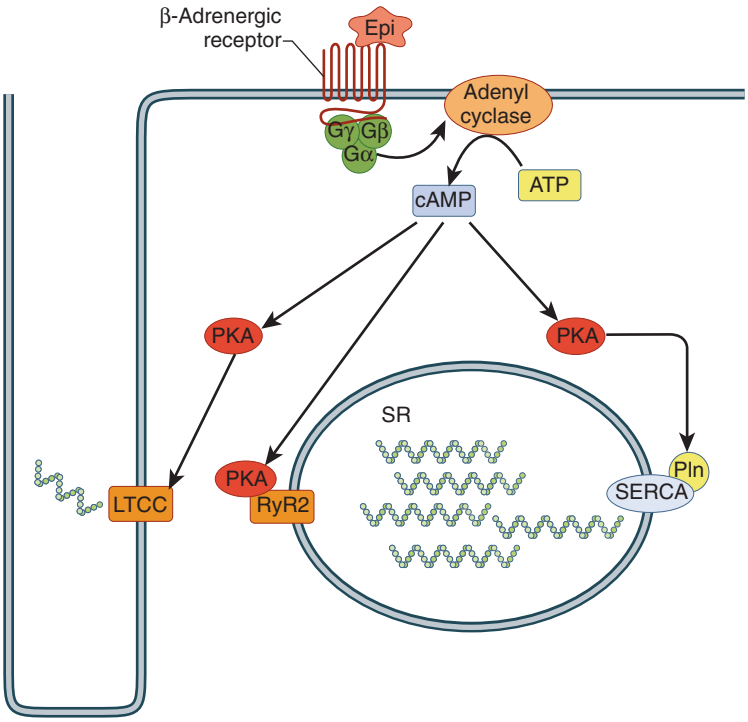

E-FIGURE 53-3. Cardiomyocyte signaling. Binding of catecholamines (epinephrine, norepinephrine) to β-adrenergic receptor, a G-coupled protein receptor, leads to activation of adenyl cyclase, which converts adenosine triphosphate (ATP) to cyclic adenosine monophosphate (cAMP). cAMP activates protein kinase A (PKA), which phosphorylates multiple calcium-handling proteins, including L-type calcium channels (LTCC), ryanodine receptors (RyR2), and phospholamban (Pln), leading to a general increase in their activity and greater contractility. Epi = epinephrine; SERCA = sarcoendoplasmic reticulum calcium ATPase; SR = sarcoplasmic reticulum.

The sinoatrial and AV nodes are regulated by parasympathetic innervation that slows the pacemaker's rate of firing and conduction through the AV node by the release of acetylcholine. Vasoconstriction of the venous system is mediated by the sympathetic nervous system, which limits fluid and blood loss after trauma.

The renin-angiotensin system also regulates blood pressure, peripheral vasoconstriction, and contractility in coordination with the sympathetic nervous system. Both the sympathetic nervous system and the renin-angiotensin system are chronically activated in heart failure (Chapter 58), in which the resulting maladaptive remodeling of the cardiovascular system promotes the progression of heart failure. Decreased perfusion to the kidney, decreased delivery of sodium to the macula densa, or increased sympathetic activity results in the release of the hormone renin from the macula densa cells within the juxtaglomerular apparatus of the kidney. Renin results in the production of angiotensin II, a potent constrictor of peripheral and coronary arteries. In turn, angiotensin II causes the release of the sodium-retaining hormone aldosterone from the adrenal gland (Chapter 227). Together, these signals result in sodium retention and increased arterial blood pressure.

PHYSIOLOGY OF THE HEART AND CIRCULATORY CONTROL

Cardiac Energetics

The major immediate source of energy in the heart is the oxidation of fatty acids and glucose. When oxygen supply is limited, glucose metabolism is favored because it generates more ATP per oxygen consumed. The heart has virtually no ability to conduct anaerobic metabolism (i.e., glycolysis) and therefore is dependent on oxygen for its function. For example, heart function deteriorates immediately under conditions of hypoxia, ischemia, and carbon monoxide poisoning.

Basal metabolism, total mechanical work performed by the heart, contractility, and heart rate determine the oxygen and energy consumption of the heart. During excitation-contraction coupling, two key steps require energy consumption (ATP hydrolysis): release of the myosin head–actin interaction and reuptake of calcium into the sarcoplasmic reticulum.

The mechanical work of the heart is determined by the total *pressure-volume area*, which is related to the number of actin-myosin cross-bridges formed during the contraction. It is the sum of the external work performed by the heart in pumping blood from the ventricle to the aorta (represented by the area inside the pressure-volume loop) plus energy stored in the myocardium at the end of contraction. Enhanced contractility requires increased oxygen consumption because an increased amount of calcium released from the sarcoplasmic reticulum requires increased ATP and oxygen consumption to pump the released calcium back into the sarcoplasmic reticulum through the sarcoplasmic reticular ATPase. On the basis of these principles, increasing the heart rate requires increased oxygen consumption. If the heart rate increases from 70 to 140 beats per minute during exercise or stress, oxygen consumption increases almost two-fold above the basal value.

Contractility and Relaxation

The Cardiac Cycle

In resting humans, the heart beats approximately once per second. With each beat, the heart cycles through a series of four hemodynamic events represented by changes in pressures and volumes (Fig. 53-3) as well as electrical activity as represented by the ECG. When the heart muscle is relaxed at end diastole, the ventricular pressure is at its resting level *(end-diastolic pressure)* and the ventricular volumes are at their maximal value *(end-diastolic volume)*. Aortic pressure declines as the blood ejected into the aorta during the previous ventricular contraction flows to the peripheral circulation. Atrial contraction provides a final boost to ventricular volume immediately before ventricular systole. Ventricular contraction increases the pressure in the ventricle; when this pressure exceeds the pressure in the atrium, the mitral valve closes. However, because ventricular pressure remains less than aortic pressure, the aortic valve remains closed, and no blood enters or leaves the ventricle during this first phase of the cardiac cycle, the *isovolumic contraction* phase. During systole, ventricular pressure eventually exceeds aortic pressure, at which time the aortic valve opens, blood is ejected into the aorta, and ventricular volume decreases during the *ejection* phase of the cycle. At the end of systole when contraction is maximum, ejection ends, and the ventricular volumes are at their lowest *(end-systolic volume)*. The volume of the ejected blood, which is termed the stroke volume (SV), is defined as the difference between the end-diastolic and end-systolic volumes. The ejection fraction (EF), defined as the percentage of end-diastolic volume (EDV) ejected during a contraction (EF = 100 × SV/EDV), is an index of heart function. The next phase in the cycle occurs when the heart muscle relaxes, ventricular pressures are less than the aorta pressure, and the aortic valve closes. During this *isovolumic relaxation* phase, ventricular volumes remain constant because, once again, both the mitral and aortic valves are closed. When ventricular pressures fall below atrial pressures, the mitral and tricuspid valves open, and blood flows from the atria into the ventricles during the *filling* phase.

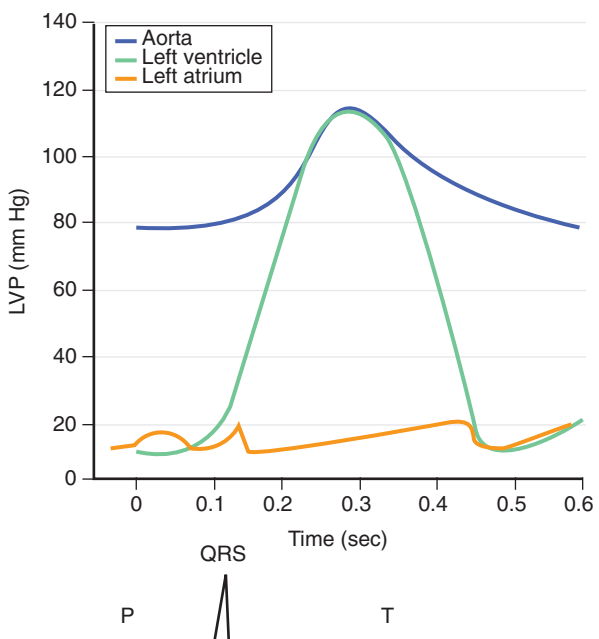

FIGURE 53-3. Wiggers diagram. Changes in aortic, left ventricular, and left atrial pressures represented graphically as a function of time, with the corresponding electrocardiogram signal for each. LVP = left ventricular pressure.

These four phases of the cardiac cycle can be represented by a *pressure-volume diagram* (Fig. 53-4), which plots the instantaneous ventricular pressure versus volume to calculate the *pressure-volume loop*. Similar effects occur on the left and right sides of the heart, but with higher pressures on the left side (Table 53-1).

Pressure-Volume Relationships

The volume of a ventricular chamber correlates with length of its muscles and sarcomeres. In the left ventricle, with its circular cross section, Laplace's law defines the relationship among pressure in the chamber (P), muscle tension (T, force/unit cross-sectional area of the muscle), chamber wall thickness (h), and internal radius of the chamber (R): $P \approx 2 \cdot T \cdot h/R$. Both calcium and the length of the heart muscle determine force (Fig. 53-5). Each muscle is composed of a linear array of sarcomere bundles. Maximal force is achieved at a sarcomere length of about 2.2 to 2.3 mm, which results in the optimal overlap of thick and thin filaments. When the sarcomere length is less than 2.0 mm, the ends of the thin filaments contact each other, thereby resulting in a reduction in force. Conversely, when sarcomeres are stretched beyond 2.3 mm, force decreases owing to reduced overlap between myosin heads and actin.

Force-length relationships, which are determined by measuring the force developed at different muscle lengths while preventing the muscle from shortening (isometric contractions), characterize the systolic and diastolic contractile properties of cardiac muscle. With increasing muscle length, end-systolic force increases to a greater degree than does end-diastolic force. The difference in force at end diastole versus end systole increases as muscle length increases as a result of the greater developed force of the stretched muscle. This relationship of force to length is referred to as the Frank-Starling law of the heart.

Work of the Heart

Cardiovascular performance is reflected in the arterial blood pressure and cardiac output (mean arterial blood flow), which in turn are dependent on four factors: preload, afterload, ventricular contractility, and heart rate.

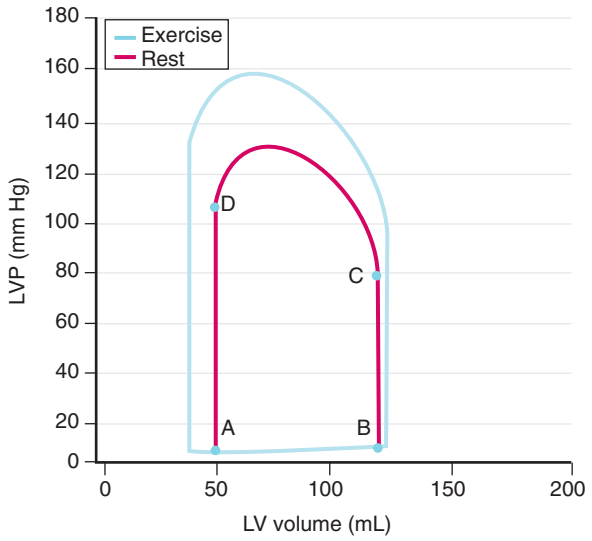

FIGURE 53-4. Pressure-volume loop. The left ventricle (LV) begins to fill when pressure in the chamber falls below that of the left atrium, and the mitral valve opens (point A). Pressure in the ventricle slowly rises as the muscle fibers are stretched by the increasing volume. When the myocardium contracts (point B), pressure in the left ventricle rises, causing the mitral valve to close and trapping the blood inside the chamber (isovolumic contraction). When the pressure in the left ventricle is higher than in the aorta, the aortic valve opens (point C), and blood is ejected out of the left ventricle. As the left ventricle stops contracting, pressure in the aorta becomes higher than that in the left ventricle, and the aortic valve closes (point D). During this period of isovolumic relaxation, the ventricle rapidly relaxes until it starts filling again. During exercise, the release of norepinephrine from sympathetic nerve terminals leads to enhanced myocardial contractility. As a result, the left ventricle generates higher pressures and ejects a greater volume of blood during each beat. LVP = left ventricular pressure.

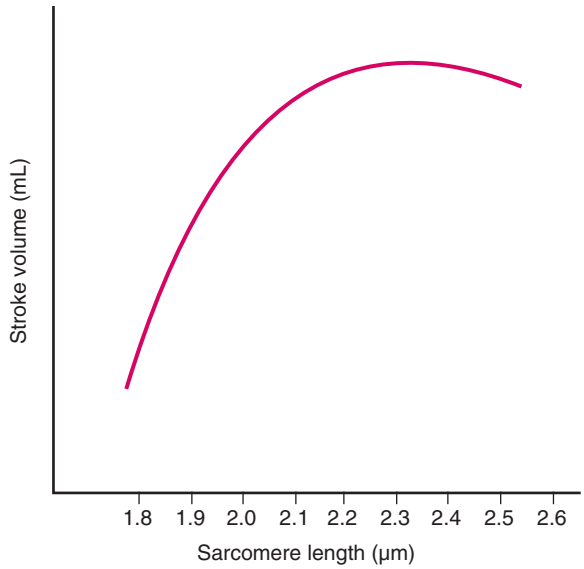

FIGURE 53-5. Starling law. Cardiac output, represented as stroke volume (end-systolic volume minus end-diastolic volume), as a function of initial sarcomere stretch. The greater the initial stretch on the fibers during diastole, referred to as preload, the more force is generated during systole.

TABLE 53-1 RANGE OF NORMAL RESTING HEMODYNAMIC VALUES

PRESSURE
Central venous (mean): 0-5 mm Hg
Right atrial (mean): 0-5 mm Hg
Right ventricular (systolic/diastolic): 20-30/0-5 mm Hg
Pulmonary artery (systolic/diastolic): 20-30/8-12 mm Hg
Left atrial (mean): 8-12 mm Hg
Left ventricular (systolic/diastolic): 100-150/8-12 mm Hg
Aortic (systolic/diastolic): 100-150/70-90 mm Hg
VOLUME-RELATED MEASURES
Right ventricular end-diastolic volume: 70-100 mL
Left ventricular end-diastolic volume: 70-100 mL
Stroke volume: 40-70 mL
Cardiac index: 2.5-4 L/min/m^2
Ejection fraction: 55-70%
ARTERIAL RESISTANCE
Systemic vascular resistance: 10-20 mm Hg · min/L
Pulmonary vascular resistance: 0.5-1.5 mm Hg · min/L

Preload, which refers to the degree to which sarcomeres are stretched just before systole, is defined as the end-diastolic pressure or volume. The Frank-Starling law of the heart dictates that ventricular pressure and output vary with preload, so a decrease in preload decreases end-diastolic volume and pressure, peak pressure, and stroke volume. Conversely, increased preload increases ventricular pressure and output, subject to the limits to which preload pressures can be increased. Left ventricular end-diastolic pressures of 20 to 25 mm Hg and greater cause exudation of fluid into the alveoli and pulmonary edema (Chapter 58).

Afterload refers to the stress that the ventricle must overcome to eject blood. Peak arterial pressure reflects the peak stress imposed on cardiomyocytes according to Laplace's law (described previously as $P \approx 2 \cdot T \cdot h/R$). As long as there is no left ventricular outflow obstruction, arterial pressure reflects myocyte afterload, as does *total peripheral resistance* (TPR), which corresponds to the tone of the resistance vessels. TPR is the ratio between the mean pressure decrease across the arterial system (mean arterial pressure [MAP] minus mean central venous pressure [CVP]) and cardiac output (CO): TPR = (MAP − CVP)/CO. When TPR is increased, the pressure-volume relationship shifts such that peak pressure is increased, whereas stroke volume and ejection fraction are decreased.

Contractility of cardiac muscle *(myocardial contractility)* or a ventricle *(ventricular contractility)* is the intrinsic ability to generate force independent of preload or afterload. When contractility is increased, the pressure-volume relationship shifts so that pressure, stroke volume, and ejection fraction are increased at constant preload volume and arterial resistance.

Cardiac output is measured in liters per minute and is equal to the amount of blood ejected at each heartbeat (stroke volume in liters per beat) multiplied by the number of beats per minute. As a result, *heart rate* is a powerful determinant of cardiac performance. Cardiac output and mean arterial pressure can be related to preload, afterload, contractility, and heart rate through the Frank-Starling curves, which plot end-diastolic pressure versus cardiac output or mean arterial pressure, to yield an overall picture of left ventricular function.

CARDIOVASCULAR RESPONSES TO STRESSORS

Exercise

Exercise requires dramatic increases in cardiac function combined with remodeling of the peripheral circulation to meet the enhanced metabolic demands of critical organs and to redirect blood flow to those organs. Indeed, the oxygen consumption during exercise can increase as much as 18-fold. About one third of the requirement for increased oxygen consumption is met by improved extraction of oxygen from the blood in the muscles (reducing venous saturation from about 75% to about 25%) and the remainder by increasing cardiac output as much as six-fold. Increased cardiac function is achieved largely through sympathetic stimulation and reduction in vagal tone, which combine to increase the heart rate, contractility, ejection fraction, filling rates, and systolic blood pressure and to decrease aortic impedance. In young healthy individuals, heart rate can increase from a baseline of 60 to 70 beats per minute at rest to as much as 170 to 200 beats per minute with exercise. To increase rather than to decrease cardiac output at these high heart rates, which can limit ventricular filling and stroke volume, contractility must also increase, through a phenomenon known as the positive force frequency relationship or Bowditch phenomenon. Along with increased cardiac contractility, arterial vasodilation in the aorta and other major arteries reduces the resistance to cardiac outflow. Both enhanced cardiac contractility and arterial vasodilation are triggered by the same sympathetic nervous system signals. With increased cardiac outflow, venous return also must increase so that preload can be maintained as well as possible to enhance cardiac function by the Frank-Starling mechanism. In response to the stress of repeated exercise (e.g., in trained athletes), the heart may undergo physiologic hypertrophy, which should be distinguished from the pathologic hypertrophy that is

triggered by hypertension, myocardial infarction, and chronic activation of neurohormonal pathways (e.g., the renin-angiotensin system) (E-Fig 53-4).

Heart Failure

Heart failure can be defined as the inability of the heart to provide sufficient blood flow to meet the metabolic demands of the organs (Chapter 58). Heart failure can be due to systolic dysfunction with volume overload, most often as the consequence of ischemic heart disease (myocardial infarction) or as the end-stage consequence of hypertension.[6] Systolic heart failure is characterized by increases in the size of the various cardiac chambers (rightward shift of the end-diastolic pressure-volume relationship). In another form of heart failure, known as *diastolic heart failure*, the heart is not necessarily increased in size, and systolic function is preserved. Emerging data implicate altered regulation of calcium inside the cardiomyocyte in the pathogenesis of both the cardiac and skeletal muscle weakness that is seen in heart failure.

Aging

Prolongation of contraction and relaxation times, which are common abnormalities in older individuals, may be related to cardiac hypertrophy as a consequence of the high prevalence of hypertension with advancing age (Chapter 67). A progressive "stiffening" of the large arteries with advancing age increases resistance, although the mechanism underlying this change is not understood. The heart rate and contractile responses to sympathetic signals are reduced and lead to a diminished ability to respond to conditions of acute overload, such as increased blood pressure or an acute myocardial infarction.

Cardiac Regeneration

Some animals, such as zebrafish, can regenerate substantial portions of their hearts after injury, sometimes by recruiting atrial myocytes to replace damaged ventricular myocytes.[7] In mammals, however, cardiomyocytes stop proliferating right after birth. Any subsequent enlargement of the heart in response to stress (e.g., hypertension) or loss of myocardium (e.g., myocardial infarction) is limited to hypertrophy of existing cardiomyocytes. Studies, however, have shown that micro-RNAs (miRNAs) can activate cardiomyocyte proliferation and repair by enabling the terminally differentiated cardiomyocytes to re-enter the cell cycle and proliferate.[8] These miRNAs are short noncoding RNAs that downregulate target mRNAs by binding to partially complementary sequences and reducing the expression of the encoded proteins.[9] In mice, miRNAs can induce cardiac regeneration and prevent loss of heart function after myocardial infarction, thereby raising the possibility for a better understanding and potential treatment of cardiac diseases.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

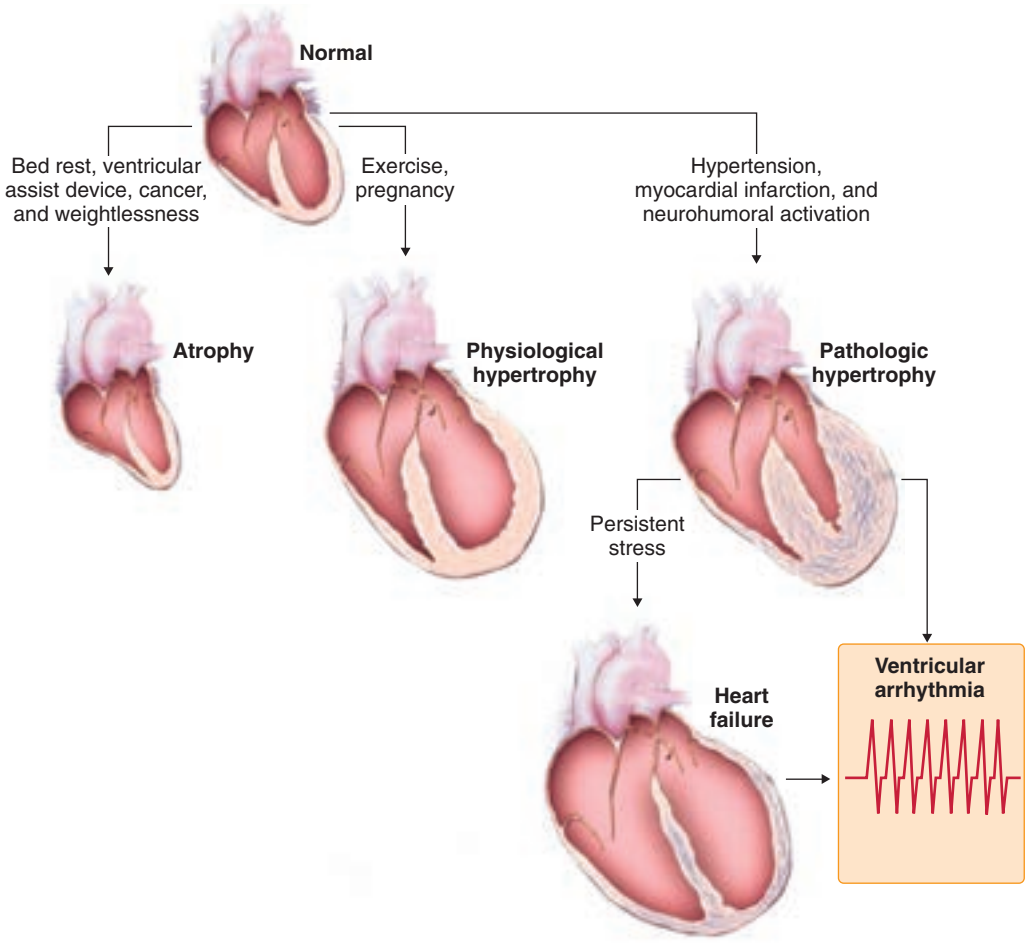

E-FIGURE 53-4. Remodeling of the heart in response to stress. The heart can undergo physiologic remodeling in response to normal stress (repeated exercise), or it can undergo pathophysiologic remodeling after a myocardial infarction or in response to chronic hypertension. Atrophy, or shrinking of the heart, can occur in microgravity (space travel), weight loss, cancer cachexia, and other conditions in which the pumping work of the heart is diminished. (From Hill JA, Olson EN. Cardiac plasticity. *N Engl J Med.* 2008;358:1370-1380.)

GENERAL REFERENCES

1. Burkhoff D, Dickstein ML, Ferber P. *The Heart Simulator*. http://www.columbia.edu/itc/hs/medical/heartsim/; Accessed March 14, 2015.
2. David R, Franz WM. From pluripotency to distinct cardiomyocyte subtypes. *Physiology (Bethesda)*. 2012;27:119-129.
3. von Gise A, Pu WT. Endocardial and epicardial epithelial to mesenchymal transitions in heart development and disease. *Circ Res*. 2012;110:1628-1645.
4. Moore-Morris T, Guimaraes-Camboa N, Banerjee I, et al. Resident fibroblast lineages mediate pressure overload-induced cardiac fibrosis. *J Clin Invest*. 2014;124:2921-2934.
5. Hinton RB, Yutzey KE. Heart valve structure and function in development and disease. *Annu Rev Physiol*. 2011;73:29-46.
6. Heusch G, Libby P, Gersh B, et al. Cardiovascular remodelling in coronary artery disease and heart failure. *Lancet*. 2014;383:1933-1943.
7. Zhang R, Han P, Yang H, et al. In vivo cardiac reprogramming contributes to zebrafish heart regeneration. *Nature*. 2013;498:497-501.
8. Eulalio A, Mano M, Dal Ferro M, et al. Functional screening identifies miRNAs inducing cardiac regeneration. *Nature*. 2012;492:376-381.
9. Pasquinelli AE. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat Rev Genet*. 2012;13:271-282.

REVIEW QUESTIONS

1. What types of muscle are found in the heart?

A. Skeletal and cardiac
B. Cardiac and smooth
C. Only smooth
D. Only cardiac
E. Smooth, cardiac, and skeletal

Answer: B Cardiac muscle is termed striated muscle, and it is anatomically and mechanistically similar to but distinct from skeletal muscle. Cardiac muscle accounts for the contractile function of the heart. Smooth muscle composes the media of the arteries in the heart and is responsible for arterial contraction and tone. There is no skeletal muscle in the heart.

2. Which of the following functions is *not* a property of adult mammalian cardiac muscle?

A. Rhythmic contractions to determine cardiac output
B. Relaxation to allow a cardiac chamber to refill
C. Hypertrophy in response to increased afterload (e.g., hypertension)
D. Rate-dependent increase in contractility
E. Regeneration after injury

Answer: E Adult cardiomyocytes are terminally differentiated, which means that they no longer can divide to form additional cells. They respond to stress by undergoing hypertrophy, a process in which individual cells enlarge, but the number of cells remains constant.

3. Which of the following is *not* true?

A. Troponin C is a critical protein in cardiac muscle.
B. Potassium is the ion that activates skeletal and cardiac muscle contraction.
C. The mechanism of cardiac muscle contraction involves shortening of the sarcomeres.
D. Thick and thin filaments are present in cardiac, skeletal, and smooth muscle sarcomeres.
E. Both pharmacomechanical and electromechanical processes can activate cardiac muscle.

Answer: B Potassium channels are involved in repolarization of the cardiac action potential, but calcium is the ion that activates muscle contraction by binding to troponin C and allowing actin-myosin cross-bridging to occur, thereby shortening the sarcomere. The source of calcium is intracellular calcium release from the sarcoplasmic reticulum through the ryanodine receptor/calcium release channel.

4. Phase 0 of the action potential, when the cardiac muscle cell is depolarized, represents influx of which of the following ions?

A. Calcium
B. Copper
C. Potassium
D. Magnesium
E. Sodium

Answer: E Opening of the sodium channel, which is the initial event in the action potential, allows sodium to rush into the cardiomyocyte and to depolarize the membrane potential. Depolarization of the cell membrane opens voltage-gated calcium channels; repolarization occurs when potassium channels are activated.

5. Which of the following steps in cardiac muscle excitation-contraction coupling does *not* require ATP consumption?

A. Pumping calcium back into the sarcoplasmic reticulum
B. Calcium release from the sarcoplasmic reticulum
C. Actin-myosin cross-bridge formation and contraction
D. Sodium-potassium exchange across the plasma membrane
E. Extrusion of calcium from the cell across the plasma membrane

Answer: B The release of calcium from the sarcoplasmic reticulum through the ryanodine receptor/calcium release channel occurs when the channel opens, allowing calcium to flow out of the sarcoplasmic reticulum and down its concentration gradient (millimolar inside the sarcoplasmic reticulum and nanomolar in the cytoplasm). Thus, no energy is required.

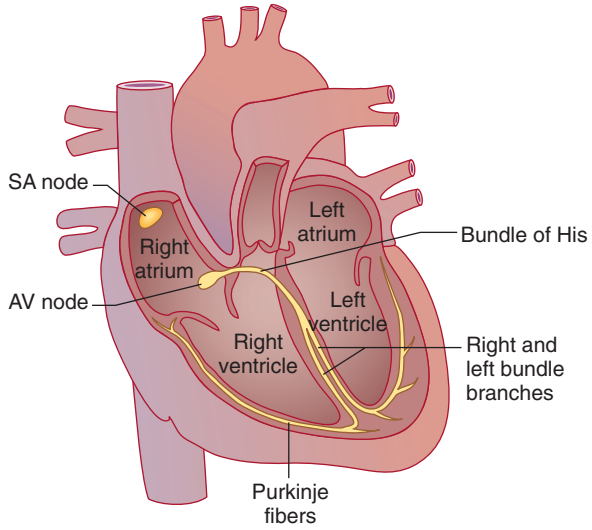

FIGURE 54-1. Cardiac conduction system. The normal conducting system consists of pacemaker cells in the sinoatrial (SA) nodal complex, specialized intra-atrial conducting tracts (including Bachmann bundle), the atrioventricular (AV) node, the His-Purkinje system, and working atrial and ventricular myocardium.

54

ELECTROCARDIOGRAPHY

LEONARD GANZ

Electrocardiography, which has changed surprisingly little since initially introduced by Einthoven in the early 1900s, allows simultaneous recording of myocardial activation from several vantage points on the body's surface, thereby permitting analysis of electrical activation in different myocardial regions. Surface electrocardiography may be supplemented with intracardiac recordings, which are particularly helpful in the diagnosis and management of cardiac arrhythmias (Chapter 62).

NORMAL FUNCTION AND ELECTROCARDIOGRAM

Normal Cardiac Activation

Electrical activation of the heart depends on the spread of a depolarizing wave front from pacemaker cells through cardiac muscle as well as through specialized conducting tissues (Fig. 54-1). Under normal circumstances, cells in the sinoatrial nodal complex in the high lateral epicardial right atrium spontaneously depolarize at the highest rate and therefore constitute the dominant cardiac pacemaker (Chapter 61). This electrical wave front spreads throughout the right and left atria; specialized conducting tracts called Bachmann bundle speed the depolarizing wave front to the left atrium. Electrical atrial activation triggers atrial muscle contraction, which propels blood through the tricuspid and mitral valves into the right and left ventricles. Normally, the atrioventricular (AV) node, where conduction delay is physiologic, serves as the only electrical connection linking the atria and ventricles; the AV valve rings are insulated. The depolarizing wave front exits the AV node into the bundle of His, a specialized conducting tissue capable of rapid conduction. The bundle of His bifurcates into right and left bundle branches; the left bundle branch divides into the left anterior and left posterior fascicles. The bundle branches and their more distal ramifications of specialized conducting tissue are called the Purkinje system. From these specialized conducting tissues, the depolarizing wave front enters into and then moves through ventricular muscle. As in the atria, ventricular electrical activation begets muscle contraction, which pumps blood through the semilunar valves into the pulmonary and systemic circulations. After electrical activation, or depolarization, a period of electrical recovery, or repolarization, is necessary before repeated activation.

At the cellular level, a complex orchestration of ion channels opening and closing determines the membrane potential throughout this process. The flow of ions into and out of the myocardial cells inscribes an action potential that reflects depolarization and repolarization as well as the spontaneous depolarization of pacemaker cells (Chapter 61).

Electrocardiographic Waves

Labeled alphabetically, beginning with the P wave, the basic waves of the electrocardiogram (ECG) correspond to these electrical events (Fig. 54-2). The P wave represents atrial muscle depolarization; in severe hyperkalemia, atrial electrical activation may be unaccompanied by atrial muscle activation, and no P wave is inscribed. The QRS complex represents ventricular muscle depolarization; the disparity between ventricular and atrial muscle mass typically yields a QRS complex much larger in voltage amplitude than the P wave. Recorded from multiple vantage points, the QRS complex harbors tremendous information about the structure and function of ventricular tissue. Under normal circumstances, the PR interval, which is the segment from the onset of the P wave to the onset of the QRS complex, represents the delay between atrial and ventricular depolarization. The ST segment and T wave (and occasionally the U wave) reflect ventricular repolarization, a process of electrical recovery that must take place before the ventricle can be depolarized again. The J (junction) point denotes the end of the QRS complex and beginning of the ST segment. Atrial muscle also requires repolarization before the next depolarizing wave front. Because ventricular mass far exceeds

atrial muscle mass, the low-amplitude atrial repolarization wave is buried underneath the QRS complex and is rarely manifested on the ECG.

One rarely seen ECG finding, the J wave (of Osborn), breaks with the alphabetic convention of the other electrocardiographic waves. Defined as a positive deflection on the QRS downstroke or at the J point, the J wave is seen most commonly in hypothermia (Fig. 54-3). It has also been described in hypercalcemia and brain injury and may increase the risk of idiopathic ventricular fibrillation (see later).

Electrocardiography Standards

A standard ECG is recorded on paper with 1-mm ("small" boxes) as well as 5-mm ("big" boxes) gridlines (see Fig. 54-2). Voltage amplitude is measured on the vertical axis (typically 10 mm equaling 1 mV) and time on the horizontal axis. Because the usual ECG recording speed is 25 mm/sec, each 1-mm gridline (small box) represents 0.04 second (40 msec), and each 5-mm gridline (big box) equals 0.2 second (200 msec). These standard calibrations can be modified in unusual circumstances, but such modifications are typically printed on the ECG.

A standard ECG is recorded during a 10-second period, although a rhythm or monitor strip can be recorded for substantially longer if necessary. Multiple leads are typically recorded simultaneously from the top to the bottom of the page. The usual groupings of leads include I, II, and III; aVR, aVL, and aVF; V_1, V_2, and V_3; and V_4, V_5, and V_6 (see later). Each group of leads is recorded for 2.5 seconds. A single lead (or multilead) rhythm strip is recorded below for the entire 10 seconds. Thus, as the ECG is scanned from left to right, one sees 10 seconds of cardiac activity, with each complex recorded simultaneously in multiple leads.

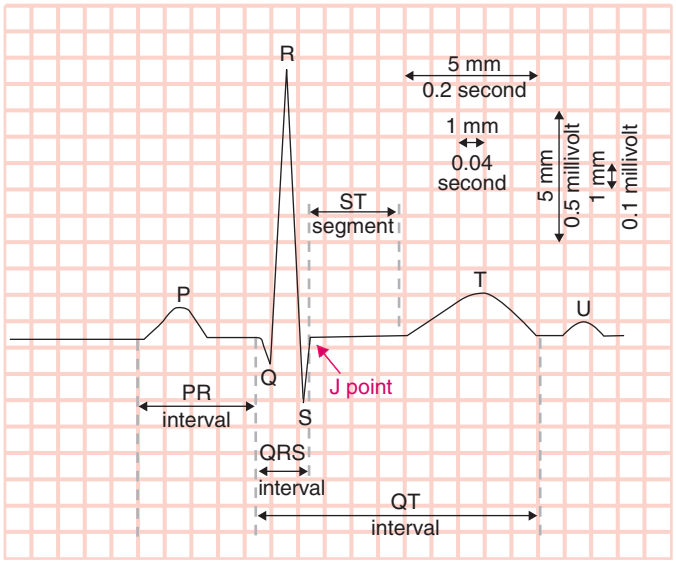

FIGURE 54-2. Inscription of a normal electrocardiogram (ECG). Sinoatrial nodal depolarization is not visible on the surface ECG; the P wave corresponds to atrial muscle depolarization. The PR interval denotes conduction through the atrial muscle, atrioventricular node, and His-Purkinje system. The QRS complex reflects ventricular muscle depolarization. The ST segment, T wave, and U wave (if present) represent ventricular repolarization. The J point lies at the junction of the end of the QRS complex and beginning of the ST segment. The QT interval is measured from the onset of the QRS to the end of the T wave. Note the gridlines. On the horizontal axis, each 1-mm line ("small" box) denotes 0.04 second (40 msec); a "big" box denotes 0.2 second (200 msec). On the vertical axis, 1 mm (small box) corresponds to 0.1 mV; 10 mm (two big boxes) therefore denotes 1 mV.

Normal Intervals

Each of the various ECG waves and intervals has normal ranges, defined from large numbers of electrocardiographic recordings in (presumably) healthy subjects (Table 54-1; see Fig. 54-2).

The RR interval (or PP interval), which is the measurement from R wave to R wave (or P wave to P wave), allows calculation of the heart rate. Because there are 60,000 msec in a minute, the heart rate (HR) in beats per minute can be easily calculated from the RR or PP interval in milliseconds:

$$HR = \frac{60{,}000}{RR}$$

Although the normal resting heart rate has traditionally been defined as being 60 to 100 beats per minute, a range of 50 to 90 at rest may actually be more reflective of normal physiology. When the heart rate is grossly irregular, as in atrial fibrillation (Chapter 64), the RR interval can be averaged over a number of cardiac cycles to estimate the heart rate. Because a standard ECG records 10 seconds in time, the heart rate (beats per minute) will equal the number of QRS complexes recorded on a standard ECG multiplied by 6. Alternatively, in a regular rhythm, the heart rate can be quickly estimated by

TABLE 54-1 NORMAL ELECTROCARDIOGRAPHIC INTERVALS

Heart rate	50-100 beats per minute
P wave duration	< 0.12 sec (120 msec)
PR interval	0.09-0.20 sec (90-200 msec)
QRS duration	0.075-0.11 sec (75-110 msec)
QTc	males: 0.39-0.45 sec (390-450 msec); females: 0.39-0.46 sec (390-460 msec)
QRS axis	−30 to +90 degrees

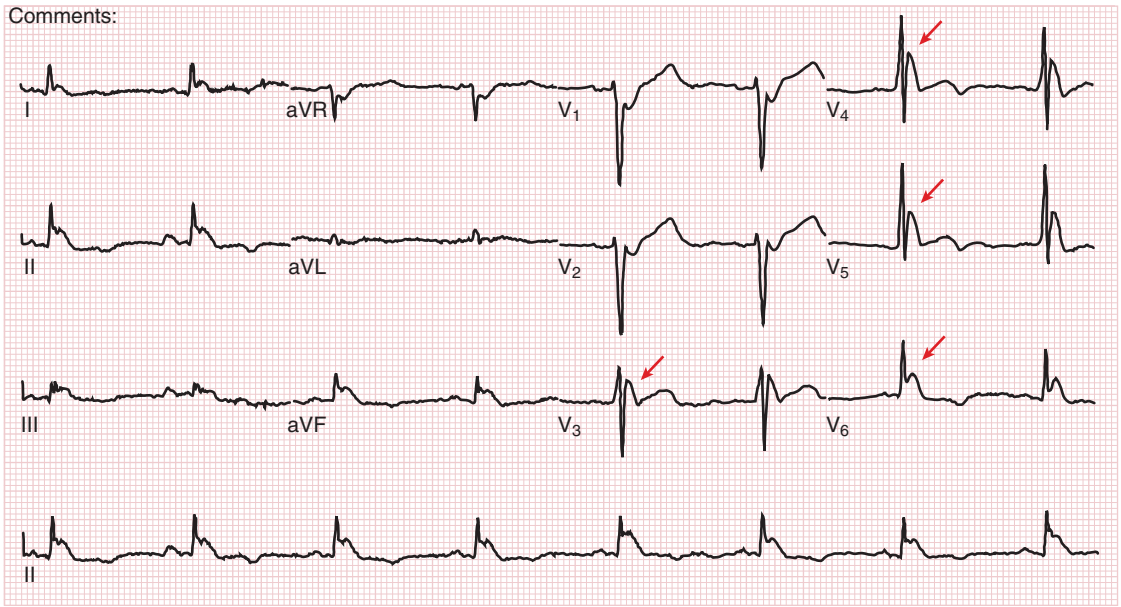

FIGURE 54-3. J wave (of Osborn). This ECG was recorded in a 40-year-old diabetic woman with profound hypothermia (26.6°C), diabetic ketoacidosis, and hypokalemia. Note the massive J waves in leads V_3 to V_6 (*arrows*) and smaller J waves in leads I, II, III, and aVF. Other notable findings include sinus bradycardia and QT prolongation.

counting the number of big boxes between consecutive QRS complexes or P waves (i.e., 2 large boxes = 150 beats per minute, 3 large boxes = 100 beats per minute, 4 large boxes = 75 beats per minute, 5 large boxes = 60 beats per minute, and so on).

P Wave Duration

The P wave duration, from the beginning to the end of a P wave, is typically less than 0.12 second (120 msec, three small boxes) in length. A broader P wave reflects an intra-atrial or interatrial conduction delay, or both. Abnormalities in P wave amplitude, morphology, and axis may reflect atrial enlargement.

PR Interval

The PR interval, measured from the onset of the P wave to the onset of the QRS complex, normally lasts between 0.09 and 0.2 second (90 to 200 msec). One-to-one AV conduction with a PR interval longer than 0.2 second has traditionally been called *first-degree AV block*, but *delayed AV conduction* may be a more appropriate term. Conduction through the atrial tissue, the AV node, and the His-Purkinje system contributes to the PR interval. When the PR interval is prolonged, delay is usually present in the AV node, although other sites of delay are possible. In the Framingham Heart Study, PR interval prolongation was associated with an increased risk of atrial fibrillation, a higher likelihood of later needing a pacemaker, and a higher overall mortality. A short PR interval may reflect ventricular preexcitation (Wolff-Parkinson-White syndrome), a junctional rhythm, or enhanced AV nodal conduction.

QRS Complex

The QRS complex, which reflects ventricular muscle electrical activation, carries important information in patients with coronary artery disease, cardiomyopathy, metabolic abnormalities, and other conditions. Capital letters (Q, R, S) denote large-amplitude deflections (≥5 mm or 0.5 mV), whereas lowercase letters (q, r, s) signify low-amplitude deflections (<5 mm or 0.5 mV). Q, q, S, and s waves are negative excursions from the isoelectric baseline, whereas R and r waves are positive deflections. Q and q waves are initial negative deflections, and S and s waves are negative deflections that follow a positive deflection (R or r wave); a QS complex is an entirely negative deflection. Q waves may reflect prior myocardial infarction (Chapter 73). An R′ or r′ wave refers to a second positive deflection after an S (or s) wave. The duration of the QRS complex reflects the time required for ventricular depolarization. Ventricular activation usually requires at least 0.075 second (75 msec, nearly two small boxes). There is some debate about the upper limit of the normal range for QRS duration; a consensus document specified 0.11 second (110 msec, nearly three small boxes). If the QRS duration is prolonged, an intraventricular or interventricular conduction delay (or both) is present. Particular patterns of interventricular conduction delay are termed bundle branch block (see later).

QT Interval

The QT interval, which reflects ventricular repolarization, is measured from the onset of the QRS complex to the end of the T wave. The QT interval is generally measured in leads II, V_5, and V_6 (see later) and reported as the longest interval among the three, averaged over three to five cycles. If the QT interval cannot be accurately measured in these leads, other leads may be used. The QT interval must be corrected to allow comparison of this interval at differing heart rates. Bazett's formula defines a corrected QT interval (QTc):

$$QTc = \frac{QT}{\sqrt{RR}}$$

Bazett's formula works reasonably well at heart rates in the normal range but overcorrects at high rates and undercorrects at low rates. Although more complex regression formulas have been developed to correct the QT interval at different heart rates, none has achieved widespread clinical use. Irregular rhythms (notably atrial fibrillation) complicate calculation of the QTc. Some investigators recommend measuring at least three QT intervals to get an average and then using an RR interval averaged over 10 cycles in Bazett's formula. The Fridericia formula,

$$QTc = \frac{QT}{\sqrt[3]{RR}}$$

may actually be more accurate than Bazett's formula in atrial fibrillation.[1]

The presence of a U wave complicates measurement of the QT (and therefore QTc) interval because it is not always clear where the T wave ends and whether the U wave should be included in a QTU interval. If the isoelectric baseline is reached between the T and U waves, the U wave is not generally included in the QT interval. If the T wave "merges" into the U wave without reaching the isoelectric baseline, the U wave is included in the QT (or QTU) interval. The QTc in a given patient may vary somewhat during the course of the day and tends to be slightly longer in young and middle-aged women than in men. The upper limit of a normal QTc is somewhat debatable, but a cutoff of 0.45 second (450 msec) in men and 0.46 second (460 msec) in women is generally used. The QT interval is sensitive to drug effects as well as to electrolyte and metabolic derangements. Patients with widened QRS complexes frequently have prolonged QT and QTc intervals. In these patients, the JT interval (from J point to the end of T wave) may be a more accurate index of repolarization, but normal standards have not been established.

Patients with a prolonged QTc, whether congenital or acquired, may be at risk for torsades de pointes ventricular tachycardia (Chapter 65). A short QTc interval (<390 msec) is unusual, and the rare patient with the short QT syndrome is at risk for malignant ventricular arrhythmias. Both short and longer QTc intervals are associated with a higher risk for development of atrial fibrillation, even in the absence of underlying structural heart disease.[2]

Electrocardiographic Leads

Recording a single ECG lead allows calculation of the heart rate and, frequently, accurate diagnosis of the heart rhythm. When the ECG is recorded from multiple skin leads simultaneously, the direction (or vector) of activation as the electrical wave front moves through the heart can be inferred. Although a number of different lead systems are possible (and some are actually used in research settings), standard electrocardiography uses 12 leads from 12 vantage points, recorded with 10 electrodes, six on the chest wall and four on the limbs. In reality, only three limb leads are actually used to generate recordings; the right leg lead serves as an electrical ground. The limb leads, called the frontal plane leads, generate bipolar and augmented unipolar lead recordings. The chest or precordial electrodes record unipolar recordings. Bipolar leads record the potential difference between two skin electrodes. In unipolar recordings, the lead of interest, the exploring electrode, is compared with a reference electrode. By convention, a positive deflection is recorded if the electrical wave front is moving toward the positive electrode in a bipolar pair or toward the exploring electrode in a unipolar lead.

The bipolar limb leads measure potential differences between electrodes on pairs of limb electrodes and closely resemble Einthoven original string galvanometer recordings. Lead I compares the right arm (negative) and left arm (positive); lead II, the right arm (negative) and left leg (positive); and lead III, the left arm (negative) and left leg (positive) (E-Fig. 54-1). Because the direction of both atrial and ventricular depolarization is away from the right arm and toward the left arm, a positive P wave and QRS complex are generally recorded in lead I. Similarly, the P wave and QRS complexes are positive in leads II and III in normal sinus rhythm because atrial and ventricular activation proceeds in a craniocaudal direction.

Leads aVR, aVL, and aVF are augmented unipolar leads in which the potential in each limb is compared with a reference electrode. For lead aVR, the potential of the right arm is compared with a reference composed of the left arm and left leg electrodes. Lead aVL compares the left arm potential with a reference combining the right arm and left leg; aVF compares the left leg with a right and left arm reference. Because atrial and ventricular activation normally moves from right to left and in a craniocaudal direction, the P wave and QRS complex are negative in lead aVR but positive in lead aVF. In lead aVL, P waves and QRS complexes are generally upright, although an rS complex may be recorded, particularly in young patients.

The precordial electrodes are positioned at specific points on the chest wall (E-Fig. 54-2A). These unipolar leads compare electrical potential between the chest electrode and a reference electrode called the Wilson central terminal. The Wilson central terminal combines the right arm, left arm, and left leg potentials through 5000-Ω resistors. The six precordial leads define atrial and ventricular activation with respect to a somewhat transverse plane through the chest wall (E-Fig. 54-2B). In this plane, atrial activation moves from right to left. Initial ventricular activation involving the septum is directed from left to right; left ventricular depolarization, which dominates right ventricular depolarization because of the differential in myocardial mass, then moves apically and laterally. In lead V_1, to the right of the sternum, the P wave is biphasic (reflecting right and then left atrial activation). Initial ventricular activation of the septum inscribes an r wave, whereas subsequent activation away from lead V_1 records a dominant S wave. In lead V_6, the P wave is positive, and initial ventricular septal depolarization inscribes a tiny "septal" q

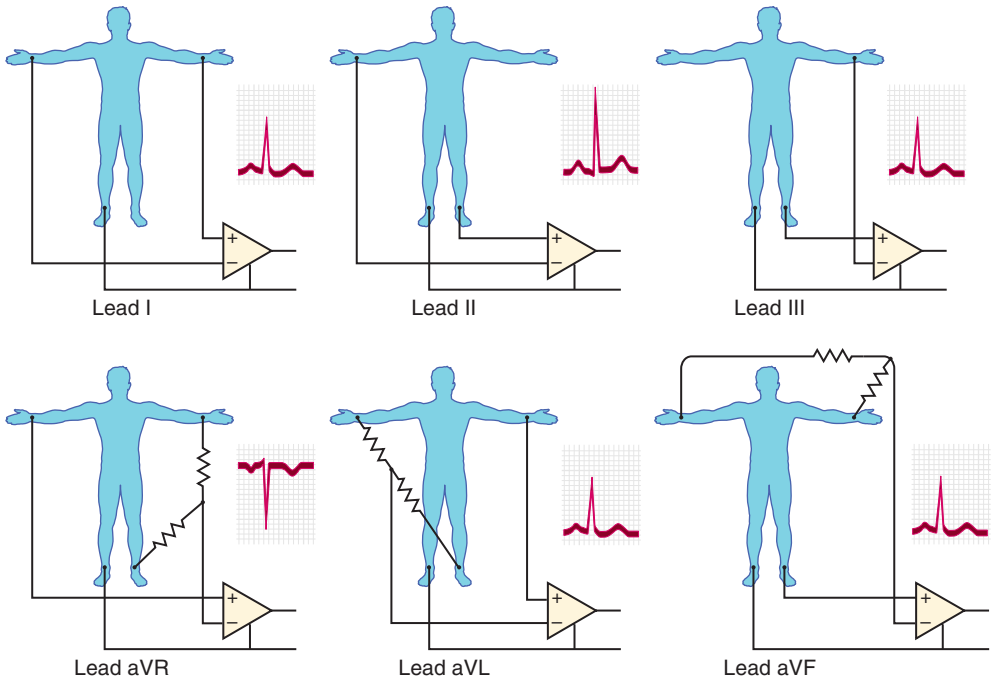

E-FIGURE 54-1. Normal cardiac activation as manifested in the limb leads. Under normal circumstances, P waves and QRS complex are typically upright in leads I, II, III, and aVF and inverted in aVR. In lead aVL, P waves are usually upright, although QRS complexes may be either upright or inverted. The right leg electrode serves to ground the system.

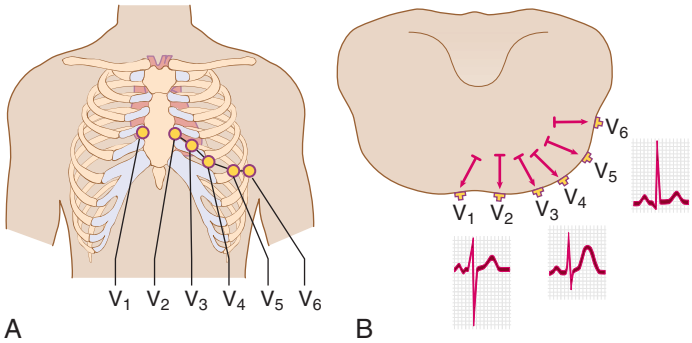

E-FIGURE 54-2. Precordial leads. A, Positioning of the precordial leads on the chest wall. **B,** Normal cardiac activation as manifested in the precordial leads. Note the small r wave and deep S wave in lead V_1, the transition at around V_3 or V_4, and the "septal" q wave and large R wave in lead V_6.

wave (usually ≤0.02 second). Subsequent ventricular depolarization records a dominant R wave.

Right-sided chest leads should be recorded when right ventricular abnormalities are suspected. RV_3, the mirror image of lead V_3, is routinely recorded in pediatric patients because of the possibility of congenital heart disease. In adults, ST elevation in lead RV_3 is specific for acute right ventricular infarction in those being evaluated for an acute inferior wall myocardial infarction.

Axis

An axis of electrical activation can be defined in the frontal plane axis by combining the bipolar and augmented unipolar limb leads (E-Fig. 54-3A). By convention, the axis parallel to lead I, toward the left, is called 0 degrees. A frontal plane axis between −30 and +90 degrees is normal, whereas other axes are abnormal (Fig. 54-4) in adults. Right axis deviation beyond +90 degrees is often a normal variant in children and adolescents. The frontal plane axis can be estimated by identifying the limb lead in which the QRS complex is most nearly isoelectric (similar positive and negative deflections); the axis is perpendicular to this lead (E-Fig. 54-3B). Because two lines pointing 180 degrees apart can be drawn perpendicular to any given line, examination of the other limb leads defines the direction in which the axis points. If the QRS complex is positive in any given limb lead, the axis will be oriented toward that limb lead, not away from it. Alternatively, the axis is in the normal range if the QRS complexes are primarily positive in both leads I and II.

An axis is not defined in the precordial leads. Rather, because the typical progression from leads V_1 to V_6 is from a predominantly negative to a positive QRS complex, the transition point is usually defined as the point at which the amplitude of the R wave first exceeds the amplitude of the S wave. Clockwise rotation (transition zone at V_4 or later) may portend a higher risk of future coronary events, and counterclockwise rotation (transition zone at V_3 or earlier) a lower risk of events.[3]

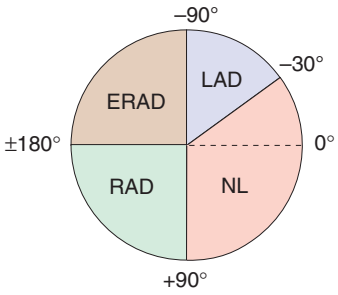

FIGURE 54-4. Chart of frontal plane axes. Normal (NL) = −30 to +90 degrees; left axis deviation (LAD) = −30 to −90 degrees (moderate, −30 to −45 degrees; marked, −45 to −90 degrees); right axis deviation (RAD) = +90 to +180 degrees (moderate, +90 to +120 degrees; marked, +120 to +180 degrees); extreme right axis deviation (ERAD) = −90 to ±180 degrees. Mild RAD is considered normal in children, adolescents, and young adults.

APPROACH TO INTERPRETING THE ELECTROCARDIOGRAM

A stepwise approach to interpreting the ECG ensures that no features of the tracing will be overlooked (Table 54-2).

Normal Electrocardiogram

Figure 54-5 is an example of a normal ECG. Sinus rhythm occurs at about 78 beats per minute, with minor variations in the RR intervals (sinus arrhythmia). The PR interval, QRS duration, and QTc are all normal. The QRS complex is most nearly isoelectric in lead aVL, so the QRS axis will be perpendicular to lead aVL. Because aVL points to −30 degrees, the QRS axis must be approximately −120 or +60 degrees. Because the QRS complex is positive in leads I and II (large R waves), the QRS axis is approximately +60 degrees. The transition in the precordial leads is typically at lead V_3 or V_4. The P wave is biphasic in lead V_1 and then positive in the other precordial leads. Septal q waves, reflecting not lateral infarction but rather normal early septal depolarization, are present in leads V_5 and V_6. Tiny q waves, a normal variant, are seen in the inferior leads.

Abnormal Electrocardiogram

Electrocardiography in patients with coronary artery disease is reviewed in Chapters 71 to 73.[4] Arrhythmias are reviewed in Chapters 61 to 66.

Conduction Abnormalities and Axis Deviation

Abnormalities of the specialized conduction system (i.e., His-Purkinje system) reflect slow or absent conduction in a particular structure (Table

TABLE 54-2 STEPWISE APPROACH TO INTERPRETING THE ELECTROCARDIOGRAM

Estimate the heart rate
Define the heart rhythm (regular vs. irregular; relationship of P waves to QRS complexes)
Measure intervals (PR, QRS duration, QT)
Calculate/estimate QTc
Estimate QRS axis
Examine P wave morphology, duration, and axis
Examine QRS progression and transition in precordial leads
Examine QRS complexes in regional groupings (septal leads [V_1,V_2], anterior leads [V_2, V_3, V_4], lateral leads [I, aVL, V_5, V_6], inferior and posterior leads [II, III, aVF, V_1,V_2])
Examine ST segments in regional groupings
Examine T waves in regional groupings

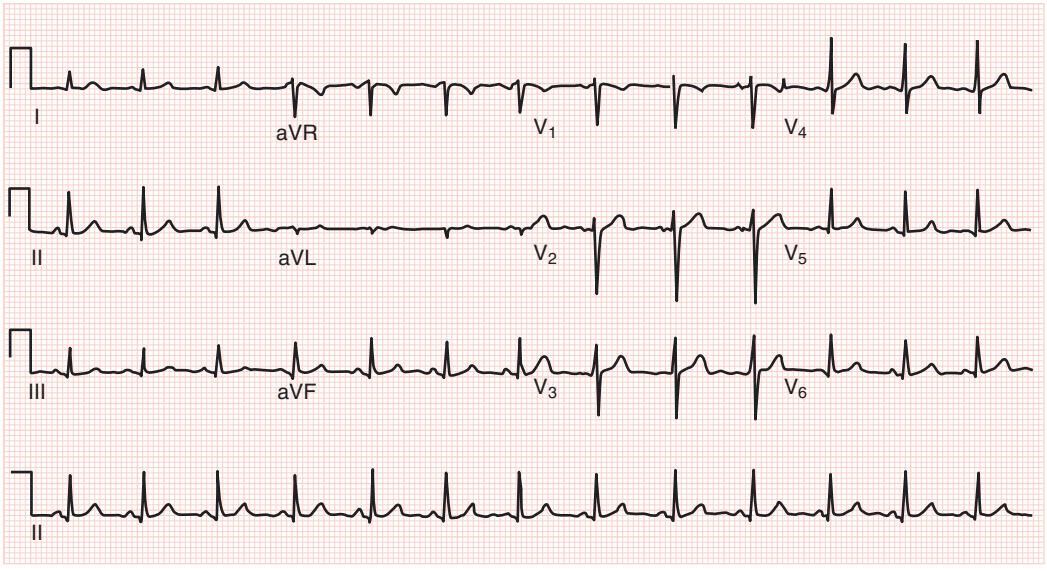

FIGURE 54-5. Normal electrocardiogram. The heart rate is approximately 78 beats per minute, with minor irregularity. Sinus arrhythmia is present. The axis is approximately +60 degrees. The PR, QRS, and QT intervals are approximately 140, 90, and 360 msec, respectively. P wave morphology, duration, and axis are normal. The transition is at lead V_4. No abnormal Q waves are present. ST segments are isoelectric, and T waves are concordant with QRS complexes.

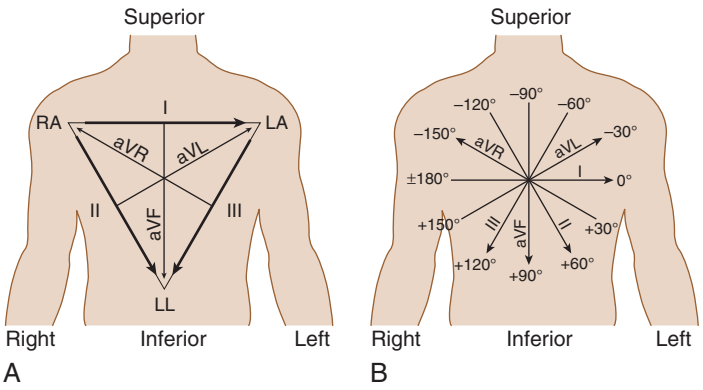

E-FIGURE 54-3. Axis of electrical activation. A, Vectors for the limb leads in the frontal plane. **B,** Hexaxial reference for determining the frontal plane axis. Note that the vectors for leads I, II, and III are in the same direction as in **A**, but now, like the augmented limb leads, these standard limb lead vectors have been moved so that they emanate from the center of the figure. LA = left atrium, LL = limb leads, RA = right atrium.

54-3 and Fig. 54-6). Left anterior or posterior fascicular block does not prolong the QRS duration beyond 120 msec. Incomplete bundle branch refers to QRS patterns that are morphologically similar to left or right bundle branch block, but with a duration of less than 0.12 sec (120 msec). An interventricular conduction delay is generally defined as a QRS duration of more than 0.11 second (110 msec). When the QRS has a duration of at least 0.12 second (120 msec), it often has the configuration of a specific bundle branch block. An isolated left bundle branch block in an otherwise healthy person is associated with a two-fold higher risk for development of a cardiovascular event or dying of a cardiovascular cause. As a result, this finding should trigger an evaluation for possible cardiac disease. By comparison, a complete right bundle branch block generally has not been associated with an increased risk, although one study suggested up to a 30% increased risk in cardiovascular mortality.[5]

TABLE 54-3 FASCICULAR AND BUNDLE BRANCH BLOCKS

	QRS DURATION	AXIS	QRS MORPHOLOGY	ST SEGMENTS AND T WAVES
LAFB	<0.12 sec (120 msec)	−45 to −90 degrees	Delayed transition across the precordium qR aVL	Normal
LPFB	<0.12 sec (120 msec)	+90 to +180 degrees	Delayed transition across the precordium rS I, aVL qR in III, aVF	Normal
RBBB	≥0.12 sec (120 msec)	Normal	rsr′, rsR′, rSR′ in V_1 (and usually V_2); wide S in V_6 and I	Discordant in V_1 and V_2
RBBB with LAFB	≥0.12 sec (120 msec)	−45 to 90 degrees	rsr′, rsR′, rSR′ in V_1 (and usually V_2); wide S in V_6 and I	Discordant in V_1 and V_2
RBBB with LPFB	≥0.12 sec (120 msec)	+ 90 to +180 degrees	rsr′, rsR′, rSR′ in V_1 (and usually V_2); wide S in V_6 and I	Discordant in V_1 and V_2
LBBB	≥0.12 sec (120 msec)	Variable	rS or QS in V_1 (S wide and notched); wide notched R without q in V_5, V_6, and I Wide notched R with or without small q in aVL	Discordant in V_1 to V_6

LAD = left axis deviation; LAFB = left anterior fascicular block; LBBB = left bundle branch block; LPFB = left posterior fascicular block; RBBB = right bundle branch block.

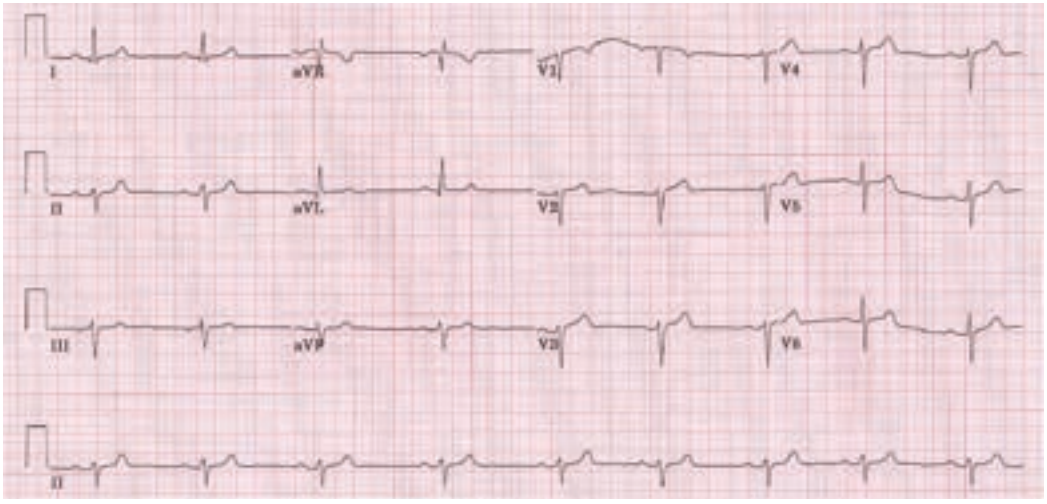

A

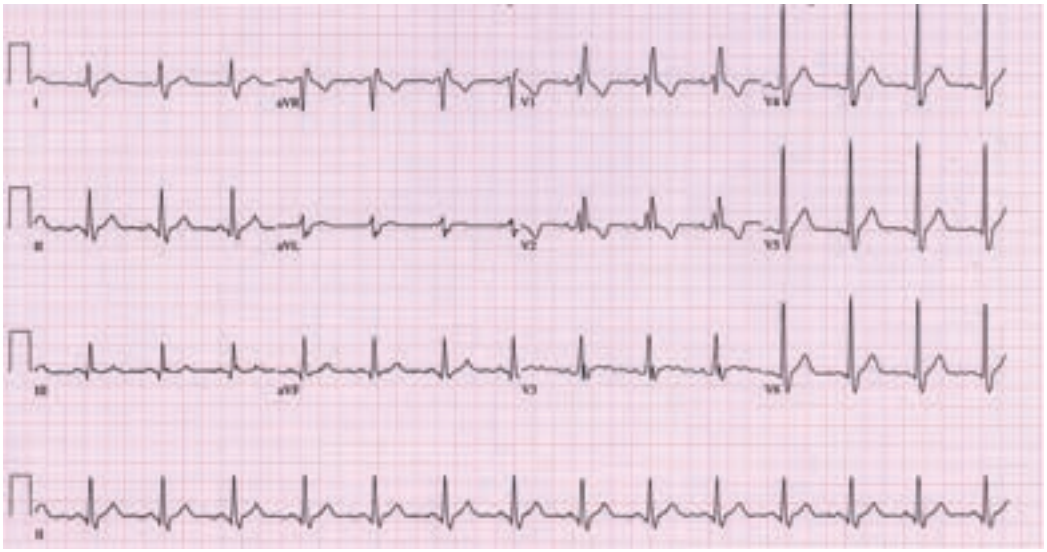

B

FIGURE 54-6. Fascicular and bundle branch blocks. A, Left anterior fascicular block (LAFB). Left axis deviation is present; the axis is approximately −60 degrees. The QRS duration is normal, and there is a delay in R wave progression across the precordial leads (late transition). Small q waves are present in leads I and aVL and small r waves in leads II, III, and aVF. **B,** Right bundle branch block (RBBB). The QRS is widened, with an rsR′ pattern in lead V_1 and a wide terminal S wave in lead V_6. ST segments are downsloping, and T waves are discordant with the QRS complex in the right precordial leads. The axis is normal, and signs of normal septal activation (q waves in lead V_6) are present.

Continued

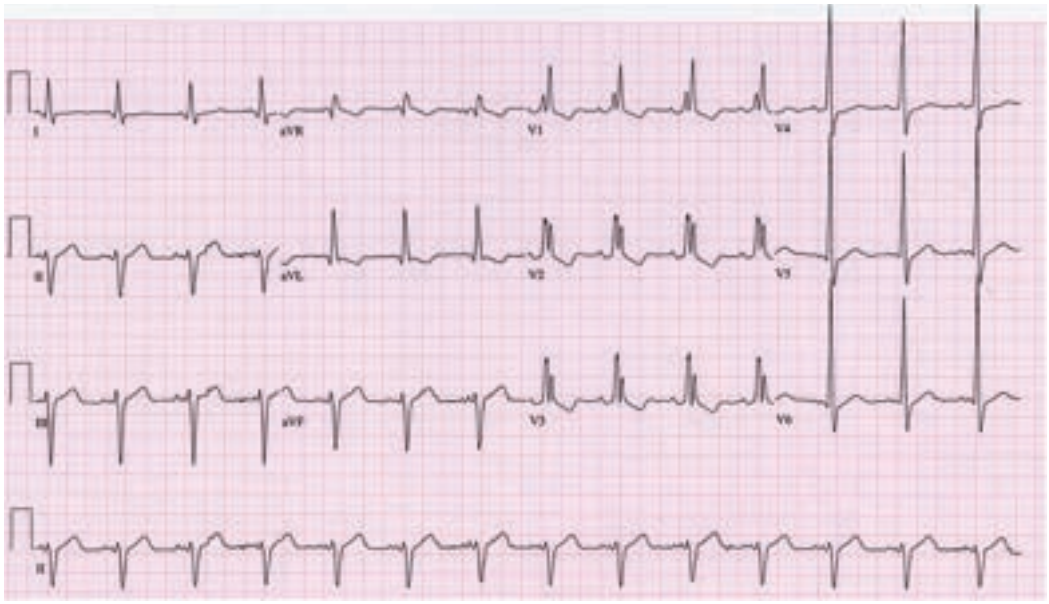

C

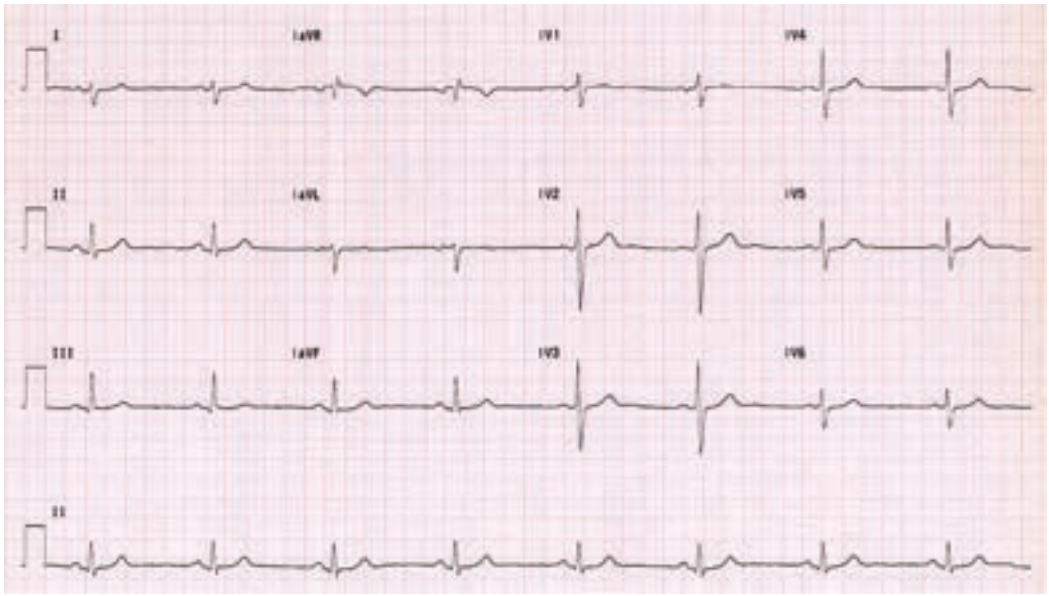

D

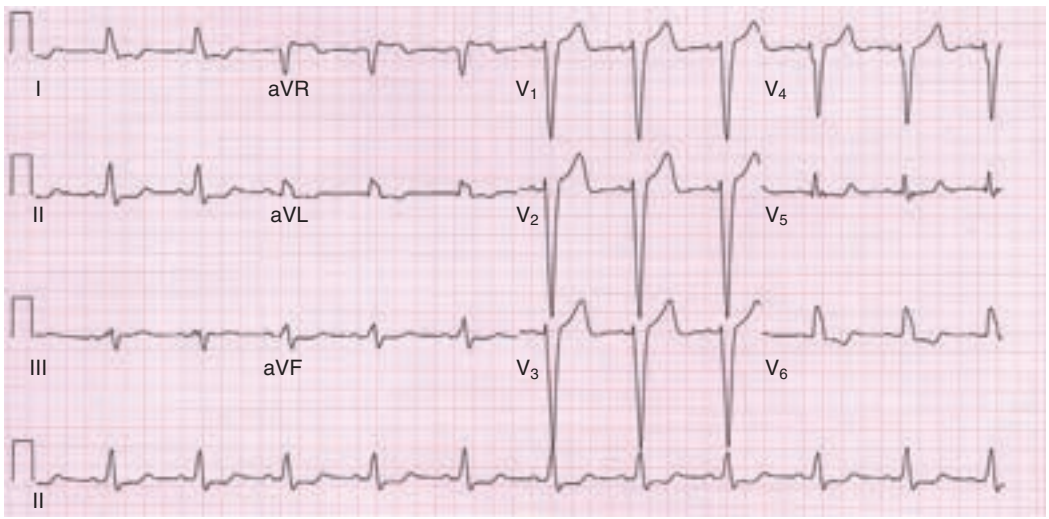

E

FIGURE 54-6, cont'd. **C,** RBBB and LAFB. In addition to features diagnostic of RBBB, an axis of –60 degrees is present. **D,** Left posterior fascicular block (LPFB). Right axis deviation (+120 degrees) is present. QRS duration is normal, and R wave progression across the precordial leads is delayed. Leads I and AVL have rS complexes, and the inferior leads have insignificant q waves. **E,** Left bundle branch block (LBBB). The QRS is widened, with a broad, notched complex in leads I and aVL and the left precordial leads. Small r waves and broad, deep S waves are present in the right precordial leads. With LBBB, the axis is usually normal or deviated to the left. ST segments and T waves are discordant with the QRS complex throughout the precordium.

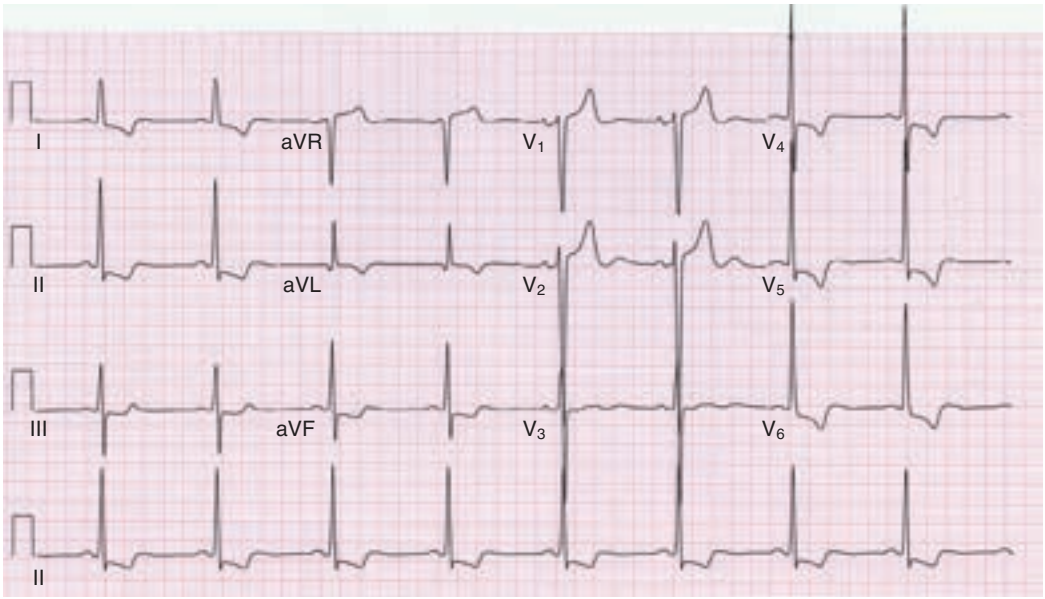

FIGURE 54-7. Left ventricular hypertrophy. Note the striking S wave amplitude in the right precordial leads and R wave amplitude in the left precordial leads. Repolarization abnormalities are present in the left precordial leads as well as in the limb leads. The S wave amplitude in V_3 (2.4 mV) plus the R wave amplitude in aVL (1.0 mV) totals 3.4 mV, easily satisfying the Cornell voltage criteria in this 76-year-old hypertensive man. Sinus bradycardia (50 beats per minute) is present as well.

Chamber Hypertrophy

A number of criteria for defining left ventricular hypertrophy (LVH; Fig. 54-7) and right ventricular hypertrophy (RVH) have been proposed. All of the LVH criteria suffer from poor sensitivity (ranging from 30 to 50%), although the specificity is good (85 to 95%). The Cornell voltage criterion, developed with an echocardiographic standard for LVH, simply adds the S wave amplitude in V_3 and the R wave amplitude in aVL; a total of more than 2.0 mV in women and 2.8 mV in men implies LVH. In many clinical settings, the Cornell criterion has replaced the more complicated Romhilt-Estes criteria, which assign points for QRS amplitude, repolarization abnormalities ("strain" pattern), left axis deviation, and other electrocardiographic features. RVH is much less common than LVH. Electrocardiographic criteria for diagnosis of RVH have even lower sensitivity (10 to 20%) than for LVH, although the specificity is similar. The Sokolow-Lyon criterion for RVH adds the R wave amplitude in lead V_1 to the S wave amplitude in lead V_5 or V_6; a sum of 1.05 mV or more implies RVH.

Low QRS Voltage

Low QRS voltage is defined as limb lead voltage of less than 5 mm (0.5 mV) in all leads or precordial voltage of less than 10 mm (1 mV) in all leads. The differential diagnosis is broad (Table 54-4), and many patients will not have a clinically apparent underlying explanation.

TABLE 54-4 CAUSES OF LOW QRS VOLTAGE

Normal variant
Pericardial effusion
Myocardial infarction
Cardiomyopathy
Hypothyroidism
Obesity
Sarcoidosis
Amyloidosis
Chronic obstructive pulmonary disease
Anasarca

Repolarization Abnormalities

Abnormalities of the ST segment or T waves, or both, are extremely common (Table 54-5). T waves may be in the same direction (concordant) with the QRS complex or discordant. Electrolyte and other metabolic abnormalities, drug effects (particularly digoxin and antiarrhythmic drugs), and secondary effects caused by LVH, bundle branch block, or pacing are all commonly responsible. Furthermore, abnormal depolarization patterns frequently beget abnormal repolarization.

Early repolarization, a relatively common pattern of ST segment elevation, occurs more commonly in patients with idiopathic ventricular fibrillation compared with controls and has also been associated with an increased risk of cardiac mortality.[6] The risk is about 30% higher with 0.1 mV of ST elevation but three-fold higher with more than 0.2 mV of ST elevation. J waves in the absence of hypothermia also increase the risk of idiopathic ventricular fibrillation about four-fold.[7]

TABLE 54-5 CAUSES OF REPOLARIZATION ABNORMALITIES

Athlete's heart
Early repolarization (normal variant)
Myocardial ischemia/injury
Pericarditis
Electrolyte abnormalities
Left ventricular hypertrophy
Intraventricular conduction delay/bundle branch block
Drug effects (digitalis, antiarrhythmic drugs)
Long QT syndrome
Stroke/neurologic catastrophe

Pitfalls of Automated Computerized Electrocardiographic Readings

Automatic computerized ECG interpretations are generally accurate for calculating heart rates, axes, and intervals but have a sensitivity of only about 70% and a positive predictive value of only about 75% for diagnosis of acute myocardial infarction on the first electrocardiogram.[8] Computerized readings are not reliable for diagnosis of rhythm disturbances, a striking weakness of these programs. Over-reading by a physician, including comparison with previous tracings when available, remains mandatory. Formal over-reading by a cardiologist is also recommended, even though it may not alter clinical care very often compared with over-reading by an emergency physician or internist.[9]

Electrocardiograms in Athletes

Extensive physical training leads to structural, electrophysiologic, and autonomic adaptations that can appear abnormal on an uninformed ECG reading.[10] Among the most important ECG findings are rhythms suggesting hypervagotonia, early repolarization, and increased chamber size (Table 54-6 and Fig. 54-8).[11] Differentiation between physiologic adaptations to exercise

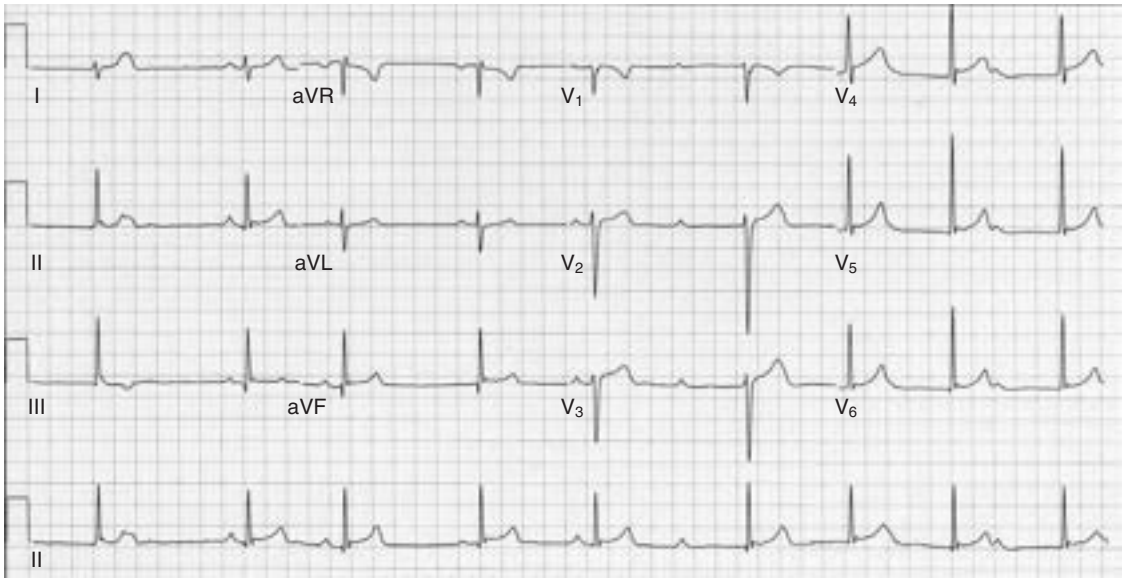

FIGURE 54-8. Athlete's heart. Bradycardia, variable P wave morphology, Mobitz I (Wenckebach) second-degree atrioventricular block, and junctional escape beats all reflect hypervagotonia in this thin, 18-year-old athlete. Mild right axis deviation is present, not uncommon in adolescents and young adults. Prominent S waves are present in leads V_2 and V_3, although formal voltage criteria for left ventricular hypertrophy are absent. Note blocked sinus P waves following junctional beats; this is not abnormal physiology.

TABLE 54-6 GENERALLY BENIGN FINDINGS IN AN ATHLETE'S ELECTROCARDIOGRAM

Sinus arrhythmia, sinus bradycardia, wandering atrial rhythm, junctional rhythm
First-degree atrioventricular block
Mobitz I (Wenckebach) second-degree atrioventricular block
Incomplete right bundle branch block
Isolated voltage criteria for left ventricular hypertrophy (e.g., without repolarization abnormalities, left axis deviation, left atrial abnormality, pathologic Q waves)
Early repolarization pattern

and potentially life-threatening abnormalities can be difficult and often requires expert consultation.

Screening Electrocardiograms

Although frequently recommended by cardiologists and primary care physicians, screening ECGs and exercise ECGs (i.e., exercise stress test) have not been shown to improve outcomes in asymptomatic adults. The U.S. Preventive Services Task Force recommends against screening resting or exercise ECG in asymptomatic adults who are at low risk of coronary heart disease events (Chapter 52).[12,13] For asymptomatic adults at intermediate or high risk, evidence is insufficient to make a recommendation.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Musat DL, Adhaduk M, Preminger MW, et al. Correlation of QT interval correction methods during atrial fibrillation and sinus rhythm. *Am J Cardiol*. 2013;112:1379-1383.
2. Nielsen JB, Graff C, Pietersen A, et al. J-shaped association between QTc interval duration and the risk of atrial fibrillation: results from the Copenhagen ECG study. *J Am Coll Cardiol*. 2013;61: 2557-2564.
3. Nakamura Y, Okamura T, Higashiyama A, et al. Prognostic values of clockwise and counterclockwise rotation for cardiovascular mortality in Japanese subjects: a 24-year follow-up of the National Integrated Project for Prospective Observation of Noncommunicable Disease and Its Trends in the Aged, 1980-2004 (NIPPON DATA80). *Circulation*. 2012;125:1226-1233.
4. Wagner GS, Macfarlane P, Wellens H, et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part VI: acute ischemia/infarction: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology. *Circulation*. 2009;119:e262-e270.
5. Bussink BE, Holst AG, Jespersen L, et al. Right bundle branch block: prevalence, risk factors, and outcome in the general population: results from the Copenhagen City Heart Study. *Eur Heart J*. 2013;34:138-146.
6. Obeyesekere MN, Klein GJ, Nattel S, et al. A clinical approach to early repolarization. *Circulation*. 2013;127:1620-1629.
7. Rosso R, Glikson E, Belhassen B, et al. Distinguishing "benign" from "malignant early repolarization": the value of the ST-segment morphology. *Heart Rhythm*. 2012;9:225-229.
8. de Champlain F, Boothroyd LJ, Vadeboncoeur A, et al. Computerized interpretation of the prehospital electrocardiogram: predictive value for ST segment elevation myocardial infarction and impact on on-scene time. *CJEM*. 2014;16:94-105.
9. Proano L, Sucov A, Woolard R. Cardiology electrocardiogram overreads rarely influence patient care outcome. *Am J Emerg Med*. 2014;32:1311-1314.
10. Drezner JA, Fischbach P, Froehlicher V, et al. Normal electrocardiographic findings: recognising physiologic adaptations in athletes. *Br J Sports Med*. 2013;47:124-136.
11. Drezner JA, Ackerman MJ, Anderson J, et al. Electrocardiographic interpretation in athletes: the 'Seattle criteria'. *Br J Sports Med*. 2013;47:122-124.
12. Moyer VA. Screening for coronary heart disease with electrocardiography: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2012;157:512-518.
13. Maron BJ, Friedman RA, Kligfield P, et al. Assessment of the 12-lead ECG as a screening test for detection of cardiovascular disease in healthy general populations of young people (12-25 years of age): a scientific statement from the American Heart Association and the American College of Cardiology. *Circulation*. 2014;130:1303-1334.

55

ECHOCARDIOGRAPHY

CATHERINE M. OTTO

Echocardiography is the clinical standard for evaluation of cardiac function in patients with known or suspected heart disease. This chapter reviews the basic principles of echocardiography, echocardiographic approaches, quantitative measurements, and clinical indications. The specific indications for echocardiography and additional echocardiographic images are presented in other chapters on individual types of cardiovascular diseases.

ECHOCARDIOGRAPHIC IMAGING

Principles

Echocardiography is based on the use of a piezoelectric crystal that converts electrical to mechanical energy, and vice versa, allowing both transmission and reception of an ultrasound signal. The frequency of ultrasound waves used for diagnostic imaging ranges from 2 to 10 MHz, with lower frequencies having greater tissue penetration and higher frequencies providing better image resolution. Each transducer consists of a complex array of piezoelectric crystals arranged to provide images in a fanlike two- or three-dimensional image, with the narrow top of this sector scan indicating the origin of the ultrasound signal. Transducers also include an acoustic lens that determines the focal depth, height, and width of the ultrasound beam.

Images are generated on the basis of the reflection of ultrasound from acoustic interfaces, for example, the boundary between the blood in the left ventricle and the myocardium. The time delay between transmission and reception is used to determine the depth of origin of the ultrasound reflection. The depths of the reflected signals from multiple ultrasound beams are combined to generate an image. The speed of signal analysis allows acquisition of two-dimensional ultrasound images at frame rates of 30 to 60 per second and of three-dimensional images at slower frame rates. Ultrasound is strongly attenuated by bone and air, so echocardiography relies on acoustic "windows," where, for example, ultrasound can penetrate to the heart while avoiding the ribs and lungs. With transthoracic imaging, the patient is positioned to bring the cardiac structures close to the chest wall, usually in a left lateral decubitus position, and the transducer is placed on the chest, with use of gel to provide acoustic coupling between the transducer and skin. Standard acoustic windows are parasternal, apical, subcostal, and suprasternal notch.[1]

Standard Image Planes

From the parasternal window, the image plane is adjusted manually by an experienced physician or sonographer to provide long and short axis views. Standard cardiac imaging planes are aligned relative to the axis of the heart, with the long axis defined as the plane that intersects the cardiac apex and the middle of the aortic valve. Short axis views are perpendicular to this long axis, with standard image planes at the cardiac base (aortic valve level), mitral valve, and midventricular levels. From the apical window, the transducer is rotated to provide three views oriented 60 degrees from each other, producing a four-chamber, a two-chamber, and a long axis view (Fig. 55-1; Video 55-1). These image planes also can be acquired with three-dimensional ultrasound transducers with standard two-dimensional plane reconstructed from the three-dimensional data set.

Measurements

Echocardiography provides accurate cardiac dimensions from three-dimensional, two-dimensional, or two-dimensional–guided linear depth

VIDEO 55-1. Standard echocardiographic views. The basic image planes used for echocardiography are the long axis (**A**), short axis (**B** and **C**), and four-chamber (**D**) image planes. In the long axis view (**A**), the normal diastolic opening of the mitral valve is well seen with rapid early motion and a smaller opening after atrial contraction. The left atrium (LA) and basal segments of the left ventricle (LV) are well seen. The aortic sinuses and aortic valve are seen with thin aortic leaflets that open widely in systole. Only a small portion of the right ventricle (RV) outflow tract is seen. In a short axis view at the level of the LV (**B**), LV wall thickness, chamber size, and regional and global systolic function can be evaluated. In a short axis view at the aortic base (**C**), the three aortic valve leaflets opening in systole are seen. This view also demonstrates the close relationship between the aortic root and both the left atrium (LA) and right atrium (RA) as well as the tricuspid valve (between the RA and RV) and the pulmonic valve (between the RV and pulmonary artery [PA]). The four-chamber view is recorded with the apical transducer position shown at the top of the image. This view allows assessment of LV and RV size and systolic function as well as evaluation of mitral and tricuspid valve anatomy and LA and RA size.

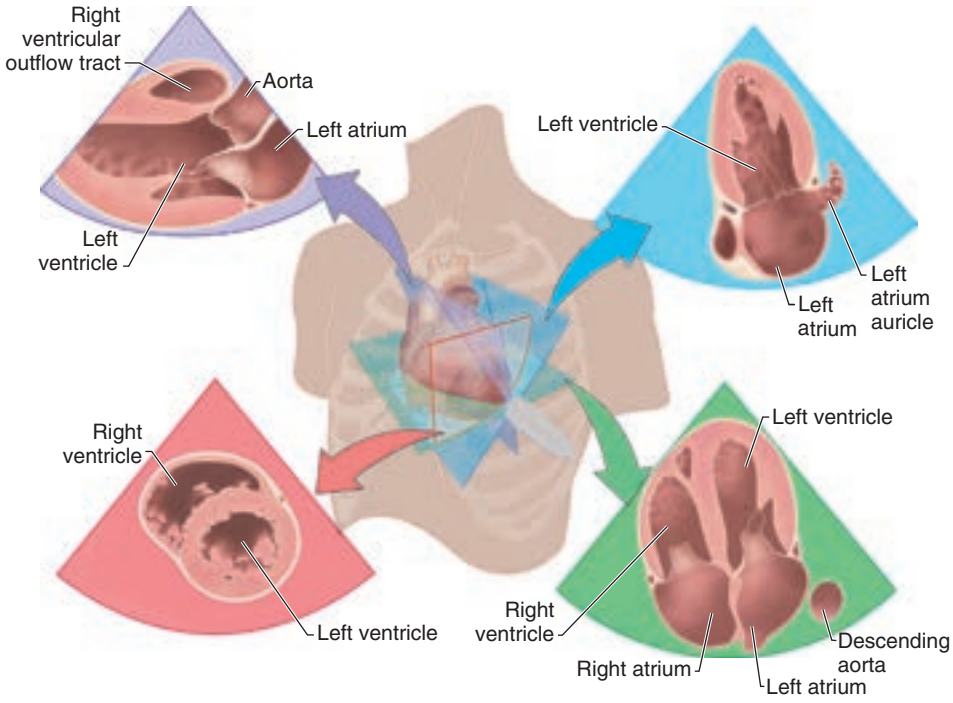

FIGURE 55-1. The four basic image planes used in transthoracic echocardiography. A parasternal transducer position or "window" is used to obtain long and short axis views. The long axis view (*purple outline*) extends from the left ventricular apex through the aortic valve plane. The short axis view is perpendicular to the long axis view, resulting in a circular view of the left ventricle (*red outline*). The transducer is placed at the ventricular apex to obtain the two-chamber (*blue outline*) and four-chamber (*green outline*) views, each of which is about a 60-degree rotation from the long axis view and perpendicular to the short axis view. The four-chamber view includes both ventricles and both atria. The two-chamber view includes the left ventricle and left atrium; sometimes the atrial appendage is visualized. See Video 55-1. (From Otto CM. *Textbook of Clinical Echocardiography.* 5th ed. Philadelphia: Elsevier Saunders; 2013:32.)

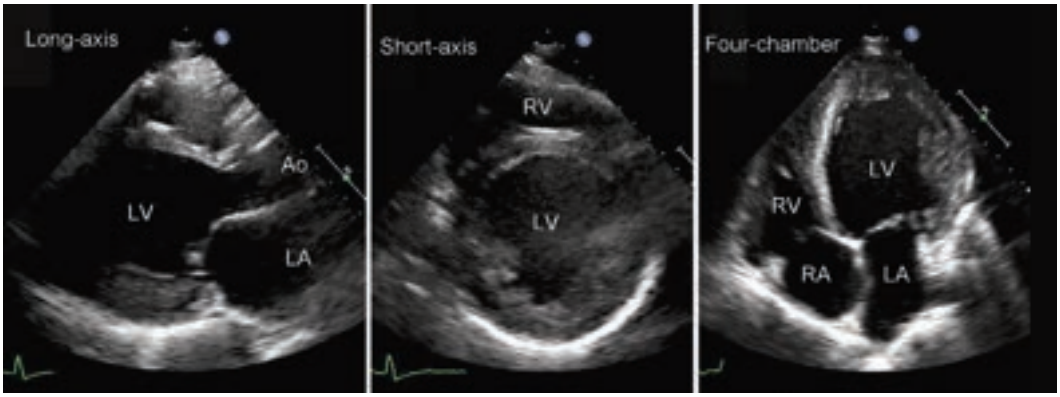

FIGURE 55-2. Dilated cardiomyopathy on echocardiography. This example shows severe left ventricular dilation and systolic dysfunction in standard image planes of parasternal long axis (*left*), parasternal short axis (*center*), and apical four chamber (*right*). Ao = aorta; LA = left atrium; LV = left ventricle; RA = right atrium; RV = right ventricle. See Video 55-2.

(M-mode) recordings. The measurements typically provided include left ventricular end-diastolic and end-systolic internal dimensions, left ventricular wall thickness, left atrial anterior-posterior diameter, and aortic sinus dimension. Left ventricular ejection fraction (EF) is determined by automated border detection from three-dimensional images or by tracing two-dimensional echocardiographic endocardial borders at end diastole and end systole in two orthogonal views (Figs. 55-2 and 55-3; Video 55-2).[2] End-diastolic and end-systolic ventricular volumes (EDV and ESV, respectively) are calculated by validated formulas, and the EF is determined as follows:

$$EF = (EDV - ESV)/EDV$$

Limitations

Echocardiography is an accurate, widely available, and widely used imaging approach. However, the quality of images can be suboptimal because of poor tissue penetration (e.g., excessive adipose tissue, position of the lungs relative to the heart), although images are nondiagnostic in less than 5% of patients with current instrumentation. Reflections are stronger when the interface is perpendicular to the ultrasound beam, so structures that are parallel to the beam may not be visible, an artifact called *echo dropout*. This potential limitation may be avoided by the use of appropriate imaging planes and the integration of data from multiple transducer positions. Ultrasound artifacts, such as beam width, shadowing, and reverberations, may be misinterpreted by inexperienced observers.

DOPPLER ECHOCARDIOGRAPHY

Principles

Ultrasound energy that is backscattered from moving red blood cells is shifted to a higher frequency when the blood is moving toward the transducer and a lower frequency when it is moving away. The magnitude of this Doppler shift corresponds to the velocity of blood flow.

Modalities

Pulsed Doppler allows measurement of flow velocity at a specific intracardiac site with the advantages of high spatial and temporal resolution. However, spatial localization is based on intermittent sampling at a time interval corresponding to the depth of interest. The sampling frequency, which is depth dependent, limits the maximum detectable velocity because

VIDEO 55-2. Dilated cardiomyopathy. In a patient with dilated cardiomyopathy, severe left ventricular (LV) dilation and systolic dysfunction are seen in the parasternal long axis view (**A**), short axis view (**B**), and apical four-chamber view (**C**). There also is marked left atrial (LA) enlargement. The right ventricle (RV) and right atrium (RA) are dilated with severely reduced RV systolic function.

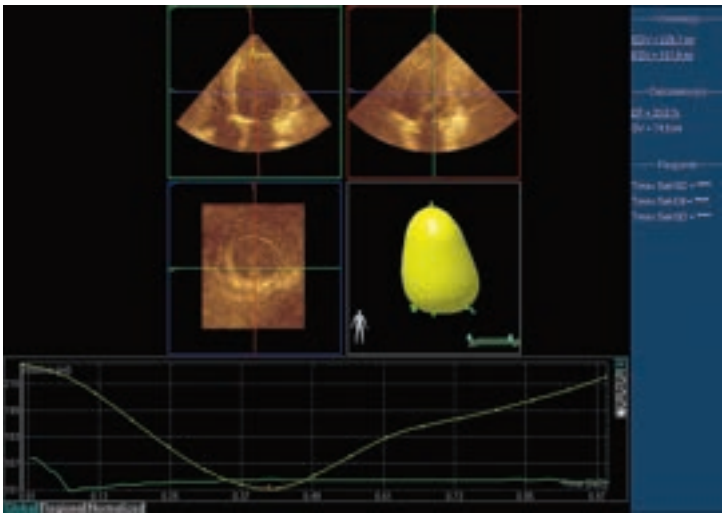

FIGURE 55-3. Three-dimensional echocardiographic measurement of left ventricular volumes and ejection fraction. In the same patient as in Figure 55-2, three-dimensional imaging shows the ventricle in four-chamber (*upper left*), two-chamber (*upper right*), and short axis (*lower left*) views derived from the three-dimensional image acquisition. The left ventricular chamber is reconstructed in the lower right. The ejection fraction is 33%.

of a phenomenon called *signal aliasing*. Normal intracardiac flow velocities are about 1 m/second, which can usually be recorded with pulsed Doppler.

Continuous-wave Doppler allows measurement of high velocities along the entire length of the ultrasound beam, but the origin of the high-velocity signal must be inferred from the two-dimensional images. With stenotic and regurgitant valves, blood flow velocities may be as high as 5 to 6 m/second, requiring the use of the continuous-wave Doppler mode. Both pulsed and continuous-wave Doppler velocities are displayed as a graph of velocity versus time, with the density of the spectral display corresponding to signal strength.

Color flow Doppler imaging is a modification of pulsed Doppler in which the flow velocity is displayed across a two-dimensional or three-dimensional image with a color scale to indicate direction and velocity. The advantage is a visually appealing display of intracardiac flow patterns. Disadvantages are low temporal resolution (frame rates of 10 to 30 per second) and poor velocity resolution due to signal aliasing.

Tissue Doppler uses the Doppler principle to record the velocity of motion of the myocardial wall. Tissue Doppler recordings of the myocardium adjacent to the mitral annulus are used to evaluate diastolic ventricular function. Speckle tracking strain imaging allows direct evaluation of myocardial mechanics (E-Fig. 55-1).[3]

Measurements

A standard echocardiographic study includes pulsed Doppler measurement of antegrade flow velocities (transmitral and transaortic) and evaluation for valve regurgitation by continuous-wave and color Doppler modalities. Other Doppler measurements depend on the specific clinical indication.

Quantitative measurements using Doppler data are derived from two basic concepts: volume flow rate and the pressure-velocity relationship. Stroke volume (SV, in cubic centimeters) can be calculated as the volume of a cylinder, where the base is the spatial cross-sectional area (CSA, in square centimeters) of flow, determined as the area of a circle from a two-dimensional diameter measurement. The height of the cylinder is the distance the average blood cell travels in one cardiac cycle, which is the velocity-time integral (VTI, in centimeters) of flow. Therefore,

$$SV\,(cm^3) = CSA\,(cm^2) \times VTI\,(cm)$$

This approach has been validated for measurement of transaortic, transmitral, and transpulmonic flow. Measurement of volume flow rate at two different intracardiac sites allows quantitation of intracardiac shunts and valvular regurgitation.

The relationship between the pressure gradient (ΔP) across a narrowing and the velocity (v) of blood flow is described by the simplified Bernoulli equation:

$$\Delta P = 4v^2$$

TABLE 55-1 INDICATIONS FOR TRANSTHORACIC ECHOCARDIOGRAPHY IN THE ACUTE SETTING AND IN PATIENTS WITH CARDIAC SIGNS OR SYMPTOMS

CARDIAC SIGNS AND SYMPTOMS
• Cardiac symptoms including chest pain, shortness of breath, palpitations, syncope/presyncope, TIA, stroke, or peripheral embolic event
• Abnormal cardiac murmur (any diastolic murmur or systolic murmur grade 3 or louder)
• Prior test results suggesting structural heart disease
• Atrial fibrillation, SVT, VT, frequent or exercise-induced VPCs
• Evaluation of pulmonary hypertension
• Suspected infective endocarditis (native or prosthetic valve) with positive blood cultures or a new murmur
ACUTE SETTING
• Hypotension or hemodynamic instability of suspected cardiac etiology
• Acute chest pain with suspected MI but nondiagnostic ECG
• Elevated cardiac biomarkers without other features of ACS
• Suspected complications of acute MI
• Evaluation of ventricular function after ACS
• Respiratory failure of uncertain etiology
• Guidance of therapy with acute pulmonary embolism
• Chest trauma or severe deceleration injury with possible cardiac consequences

ACS = acute coronary syndrome; ECG = electrocardiogram; MI = myocardial infarction; SVT = supraventricular tachycardia; TIA = transient ischemic attack; VPCs = ventricular premature contractions; VT = ventricular tachycardia.
Summarized from Douglas PS, Garcia MJ, Haines DE, et al. ACCF/ASE/AHA/ASNC/HFSA/HRS/SCAI/SCCM/SCCT/SCMR 2011 Appropriate Use Criteria for Echocardiography. A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, American Society of Echocardiography, American Heart Association, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, Society of Critical Care Medicine, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance Endorsed by the American College of Chest Physicians. *J Am Coll Cardiol.* 2011;57:1126-1166. Reproduced from Otto CM. *Textbook of Clinical Echocardiography.* 5th ed. Philadelphia: Elsevier Saunders; 2013:119.

This equation allows calculation of maximum and mean gradients across stenotic valves, estimation of pulmonary systolic pressure, and detailed evaluation of intracardiac hemodynamics with regurgitant valves.

ECHOCARDIOGRAPHIC APPROACHES

Several echocardiographic modalities are in clinical use. If it is unclear which modality is optimum in a specific clinical setting, consultation with the echocardiographer is appropriate.

Transthoracic echocardiography is the standard clinical approach in most patients with suspected or known cardiac disease (Table 55-1). Advantages are that it is noninvasive, has no known adverse effects, and provides detailed data on cardiac anatomy and physiology. Limitations include poor image quality in some patients, limited visualization of structures distant from the transducer (e.g., atrial septum, left atrial appendage), and inability to visualize structures immediately distal to prosthetic heart valves (acoustic shadowing).[4]

Transesophageal echocardiography offers superior image quality because of a shorter distance between the transducer and the heart, the absence of interposed bone or lung, and the use of a higher-frequency transducer (Table 55-2). Transesophageal echocardiography usually is well tolerated, but intubation of the esophagus entails some risk, and most clinicians do this procedure with the patient under moderate sedation. Transesophageal echocardiography is much more sensitive than transthoracic echocardiography for detection of left atrial thrombus (95% vs. 50%), valvular vegetations (99% vs. 60%), and prosthetic mitral valve regurgitation (Fig. 55-4).

Point-of-care echocardiography refers to the use of smaller, less expensive ultrasound systems that can be carried by the physician, who can perform quick, limited examinations in the emergency department, at the inpatient bedside, or in the outpatient setting (Table 55-3). Point-of-care echocardiography units range from pocket sized to laptop sized. Some are very simple, with only two-dimensional imaging and limited controls; other systems provide high-quality imaging and all Doppler modalities. Point-of-care echocardiography does not replace a complete imaging study but can serve as an adjunct to the physical examination, particularly in the acute care setting, such as to distinguish ventricular dilation from a pericardial effusion, to estimate ventricular systolic performance (Table 55-4), or to screen for critical aortic stenosis.[5,6]

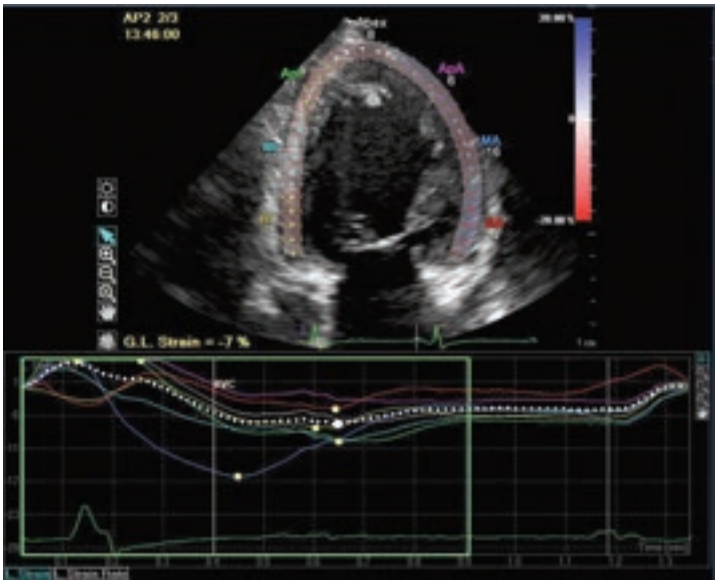

E-FIGURE 55-1. Speckle tracking strain imaging. In an apical four-chamber view, myocardial strain is calculated by the change in distance (divided by the original distance) between echogenic myocardial speckles for each segment as shown by the colored dots. The relationship of time (*x*-axis) with strain (*y*-axis) is shown in colors corresponding to the colors on the two-dimensional image, with global longitudinal strain shown in white. This image is from the same patient as in Figures 55-2 and 55-3 and shows reduced (only 7%) global longitudinal strain (normal is –16% to –22%).

TABLE 55-2 INDICATIONS FOR USE OF TRANSESOPHAGEAL ECHOCARDIOGRAPHY AS THE INITIAL OR SUPPLEMENTAL TEST

- Patients with a high likelihood of nondiagnostic TTE due to patient characteristics or ability to visualize the structures of interest
- Suspected acute aortic disease including dissection and transection
- Suspected endocarditis with a moderate or high pretest probability (e.g., staphylococcal bacteremia, fungemia, prosthetic heart valve or intracardiac device)
- Evaluation of valve structure and function to evaluate suitability for surgical or transcatheter valve interventions
- Guidance of percutaneous noncoronary cardiac interventions, including but not limited to septal ablation, mitral valvuloplasty, PFO/ASD closure, radio frequency ablation
- Evaluation of patients with atrial fibrillation or flutter to facilitate clinical decision making with regard to anticoagulation, cardioversion, or radio frequency ablation
- Evaluation for cardiac source of embolus with no identified source on TTE
- Reevaluation for interval changes compared with prior TEE when a change in therapy is anticipated.
- Suspected complications of endocarditis (e.g., abscess, fistula)*
- Suspected prosthetic mitral valve dysfunction*
- Evaluation of posterior structure (e.g., atrial baffles) in congenital heart disease patients*

*Not considered in the appropriateness guidelines document but generally accepted as appropriate indications for transesophageal echocardiography as the initial approach.
PFO/ASD = patent foramen ovale/atrial septal defect; TEE = transesophageal echocardiography; TTE = transthoracic echocardiography.
Abstracted from Douglas PS, Garcia MJ, Haines DE, et al. ACCF/ASE/AHA/ASNC/HFSA/HRS/SCAI/SCCM/SCCT/SCMR 2011 Appropriate Use Criteria for Echocardiography. A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, American Society of Echocardiography, American Heart Association, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, Society of Critical Care Medicine, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance Endorsed by the American College of Chest Physicians. *J Am Coll Cardiol.* 2011;57:1126-1166 with modification. Reproduced from Otto CM. *Textbook of Clinical Echocardiography.* 5th ed. Philadelphia: Elsevier Saunders; 2013:121.

Contrast echocardiography may be performed with intravenous injection of agitated saline to opacify the right-sided heart chambers. These microbubbles are relatively large and do not pass through pulmonary capillaries. Therefore, appearance of contrast material in the left side of the heart within one or two beats after right-sided heart opacification is consistent with an intracardiac shunt. Although most atrial-level shunts are predominantly left-to-right shunts, a small amount of right-to-left shunting occurs, which is the basis of this approach.

Contrast echocardiography also may be performed with commercially available microbubbles in the range of 1 to 5 μm. Because these microbubbles are smaller than the pulmonary capillaries, right-sided heart opacification is followed by left-sided heart opacification, which can enhance the evaluation of systolic function when image quality is suboptimal, especially during stress echocardiography (Fig. 55-5).

Three-dimensional echocardiography is increasingly available and is recommended for quantitation of left ventricular function, evaluation of complex structural heart disease, and guidance of transcatheter interventions (Fig. 55-6; Video 55-3).

Stress echocardiography is a standard approach for evaluating patients with known or suspected coronary artery disease; it has a sensitivity (85 to 95%) and a specificity (80 to 90%) similar to those of radionuclide stress imaging (Chapters 56 and 71). Myocardial infarction results in thinning and akinesis of the affected wall. However, in the absence of infarction, resting myocardial function is normal, even when severe epicardial coronary disease is present. The increased myocardial demand associated with exercise or pharmacologic stress leads to myocardial ischemia, which results in a regional wall motion abnormality, often before the onset of chest pain or electrocardiographic changes (Fig. 55-7; Video 55-4).

In patients who can exercise, standard views of the left ventricle are recorded at baseline and immediately after maximal treadmill or bicycle exercise. If endocardial definition is suboptimal, left-sided contrast is used. The rest and exercise images are compared in a side-by-side cine loop format. Myocardial ischemia is present if resting wall motion is normal but hypokinesis or akinesis is seen after exercise. The pattern of regional wall

TABLE 55-3 INDICATIONS FOR POINT-OF-CARE ECHOCARDIOGRAPHY

To complement a physical examination, especially in an intensive care unit
Rapid initial screening in an emergency setting or ambulance
Screening programs in schools, industry, and community activities
Triaging candidates for a complete echocardiographic examination
Teaching tool, especially to correlate with a cardiac examination

Modified from Sicari R, Galderisi M, Voigt JU, et al. The use of pocket-size imaging devices: a position statement of the European Association of Echocardiography. *Eur J Echocardiogr.* 2011;12:85-87.

TABLE 55-4 GOALS OF POINT-OF-CARE ECHOCARDIOGRAPHY IN THE SYMPTOMATIC EMERGENCY DEPARTMENT PATIENT

Assess possible pericardial effusion and guide pericardiocentesis
Assess global cardiac systolic function
Identify ventricular enlargement
Assess intravascular volume
Confirm potential positioning of a transvenous pacing wire

Modified from Labovitz AJ, Noble VE, Bierig M, et al. Focused cardiac ultrasound in the emergent setting: a consensus statement of the American Society of Echocardiography and American College of Emergency Physicians. *J Am Soc Echocardiogr.* 2010;23:1225-1230.

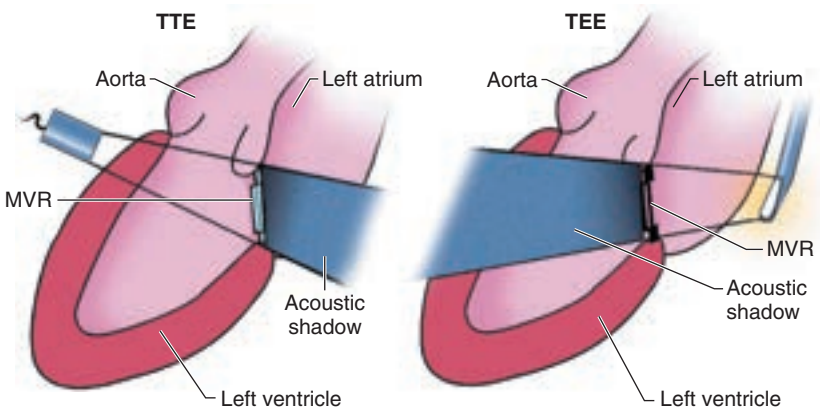

FIGURE 55-4. The problem of acoustic shadowing from a prosthetic mitral valve replacement (MVR). ***Left,*** **With transthoracic echocardiography (TTE), the acoustic shadow distal to the prosthetic valve obscures the left atrium, limiting assessment of valve regurgitation by Doppler techniques.** ***Right,*** **With transesophageal echocardiography (TEE), the left atrium now can be evaluated for valvular regurgitation. However, the acoustic shadow now obscures the left ventricle. (From Otto CM.** ***Textbook of Clinical Echocardiography.*** **5th ed. Philadelphia: Elsevier Saunders; 2013:121.)**

VIDEO 55-3. Three-dimensional echocardiography. The surgeon's view of the mitral valve from the left atrial (LA) side of the valve with the aortic valve at the top of the image. In diastole, the anterior leaflet and posterior leaflet (with P1, P2, and P3 scallops) are seen in the open position with the normal mitral valve orifice (MVO). In systole, severe prolapse of the anterior leaflet is seen, particularly one bulging section (*asterisk*); a flail segment with two small ruptured chords (*arrow*), resulting in severe posteriorly directed mitral regurgitation, is well visualized.

VIDEO 55-4. Stress echocardiography. With exercise or pharmacologic stress, the left ventricle (LV) is imaged in standard views to evaluate global and regional function with the cine images compared side by side at the same cine speed. The normal response to stress is an increase in endocardial motion and wall thickening, with a decrease in chamber size (**A** and **B**). When coronary disease is present, wall motion becomes abnormal with stress in the region of myocardium supplied by the affected vessel. In this patient with a proximal stenosis of the left anterior descending coronary artery, a contrast image of the LV at rest shows normal motion (**C**), but the apical two thirds of the septum is akinetic with stress (*arrows*, **D**).

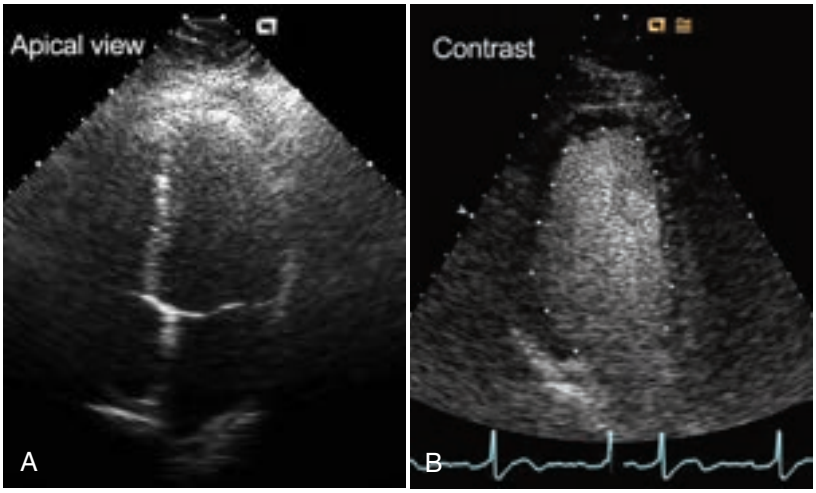

FIGURE 55-5. Poor-quality apical view (A) with marked improvement in definition of the left ventricular cavity after opacification by contrast echocardiography (B). The *dots* indicate the left ventricular endocardial tracing for calculation of ejection fraction.

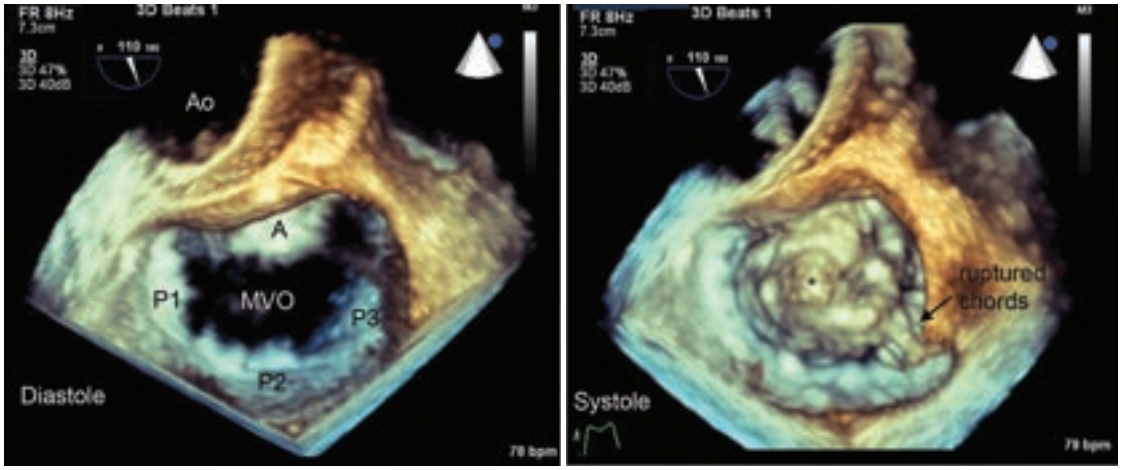

FIGURE 55-6. The surgeon's three-dimensional view of the mitral valve from the left atrial side of the valve with the aortic valve at the top of the image. In diastole, the anterior (A) leaflet and posterior leaflet (with P1, P2, and P3 scallops) are seen in the open position, with the normal mitral valve orifice (MVO). In systole, severe prolapse of the anterior leaflet is seen, particularly one bulging section (*asterisk*), and a flail segment with two small ruptured chords (*arrow*) is well visualized. These abnormalities cause severe, posteriorly directed mitral regurgitation. Ao = aorta. See Video 55-3. (From Otto CM. *Textbook of Clinical Echocardiography.* 5th ed. Philadelphia: Elsevier Saunders; 2013:327.)

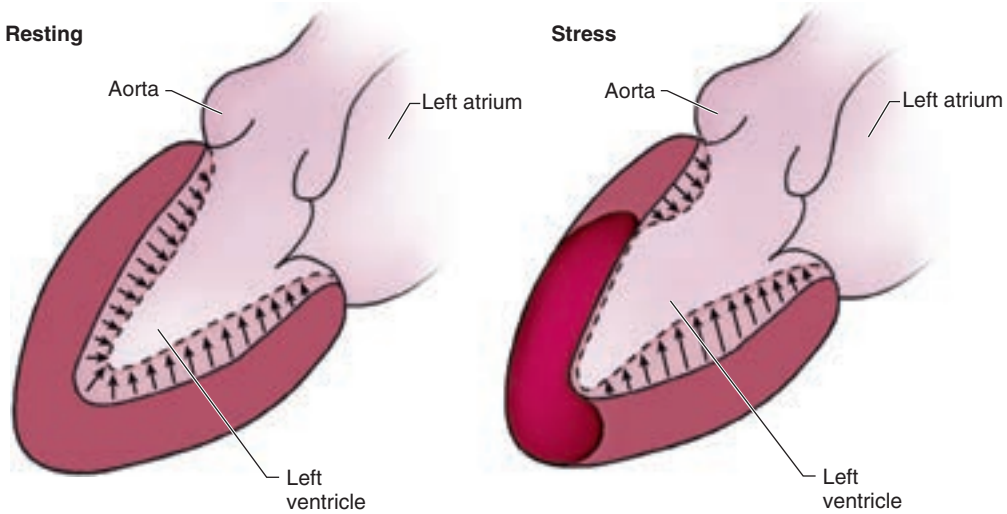

FIGURE 55-7. The concept of stress echocardiography in a patient with 70% stenosis in the proximal third of the left anterior descending coronary artery. At rest (*left*), endocardial motion and wall thickening are normal. After stress (*right*), either exercise or pharmacologic, the middle and apical segments of the anterior wall become ischemic, showing reduced endocardial wall motion and wall thickening. If the left anterior descending coronary artery extends around the apex, the apical segment of the posterior wall also will be affected, as shown here. The normal segment of the posterior wall shows compensatory hyperkinesis. See Video 55-4. (From Otto CM. *Textbook of Clinical Echocardiography.* 5th ed. Philadelphia: Elsevier Saunders; 2013:199.)

motion accurately identifies the area of myocardium at risk and is reasonably reliable for identification of the affected coronary artery. With three-vessel coronary disease, rather than a regional wall motion abnormality, the only clue on imaging may be an absence of the expected decrease in chamber size at peak exercise, caused by diffuse ischemia. Interpretation of an exercise echocardiogram includes exercise duration, hemodynamic response, symptoms, and electrocardiographic changes in addition to the echocardiographic images.

In patients who are unable to exercise, stress testing is performed with a graded intravenous infusion of dobutamine, beginning at 5 to 10 μg/kg/minute and increasing every 3 minutes to a maximum dose of 40 μg/kg/minute. If needed, atropine is used to achieve 85% of the maximum predicted heart rate. In addition to evaluation for myocardial ischemia, dobutamine stress echocardiography can assess myocardial viability in areas of stunning or hibernation, based on an improvement in endocardial motion from baseline to low-dose dobutamine, with subsequent worsening of function at higher doses—the "biphasic" response.

Intracardiac echocardiography is performed with an ultrasound probe on a catheter that is inserted into the right side of the heart through the femoral vein. Intracardiac echocardiography is used in the cardiac catheterization laboratory to guide percutaneous closure of a patent foramen ovale (PFO) and other procedures. In the electrophysiology laboratory, intracardiac echocardiography helps guide catheter positioning and identify complications.

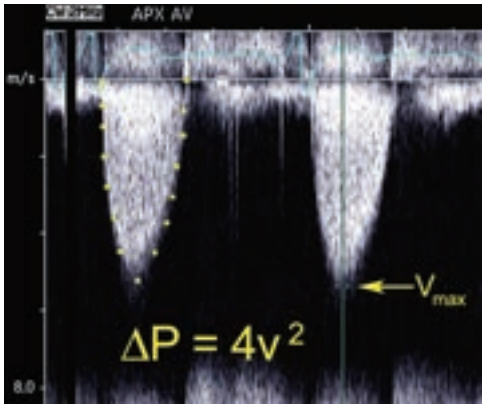

FIGURE 55-8. **In a patient with aortic stenosis, the aortic jet velocity is recorded with continuous-wave Doppler from the window that yields the highest velocity signal. Maximum velocity (V_{max}) is used to calculate the maximum systolic gradient. The Doppler curve is traced, as shown, to calculate the mean systolic gradient with the Bernoulli equation, by which the pressure gradient (ΔP) equals four times the square of the velocity.**

CARDIAC FUNCTION MEASUREMENTS

In addition to qualitative descriptions of cardiac anatomy and physiology, echocardiography provides precise and accurate quantitation of cardiac function (including ventricular systolic and diastolic function), an estimate of the severity of valve stenosis and regurgitation, and a noninvasive estimate of pulmonary pressures.

Systolic Ventricular Function

Overall left ventricular systolic function is graded by visual estimation, with an approximate correspondence to EF as follows: normal (EF > 55%), mildly reduced (EF, 40 to 55%), moderately reduced (EF, 20 to 40%), and severely reduced (EF < 20%). More precise quantitation is performed when it is clinically indicated by calculation of a three-dimensional or biplane ejection fraction. Cardiac output calculations are not routine but may be helpful for noninvasive monitoring of therapy in patients with heart failure. Because EF measurements are affected by preload and afterload, measures that are less dependent on loading conditions, including the end-systolic dimension or volume, are generally preferred for clinical decision making in situations such as the timing of surgery for chronic valvular regurgitation.

Evaluation of right ventricular function by echocardiography is more challenging because of the complex shape of the chamber. Useful measurements include the basal diameter, the tricuspid annular plane systolic excursion, and the systolic Doppler tissue velocity at the annulus. Cardiac magnetic resonance imaging provides more accurate quantitation of right ventricular volumes and function when it is clinically needed.

Diastolic Ventricular Function

Evaluation of diastolic ventricular function is challenging because the patterns of ventricular filling are affected by preload, heart rate, and coexisting valvular regurgitation in addition to the diastolic properties of the ventricle. However, echocardiography can classify diastolic function on the basis of the combination of left ventricular inflow, pulmonary vein flow, tissue Doppler velocities, and isovolumic relaxation time. An estimate of left ventricular filling pressure (e.g., left ventricular end-diastolic pressure) also can be inferred by these approaches.

Valvular Stenosis

Echocardiography is the clinical standard for evaluation of aortic valvular heart disease (see Fig. 75-2). Cardiac catheterization is reserved for cases in which echocardiography is nondiagnostic, clinical data are discrepant with echocardiographic findings, or the coronary anatomy needs to be assessed (Chapter 75).

In patients with aortic stenosis, the most direct measure of stenosis severity is the antegrade velocity across the valve, indicating mild (<3 m/second), moderate (3 to 4 m/second), or severe (>4 m/second) valve obstruction. The maximum and mean transaortic pressure gradients also can be calculated by the Bernoulli equation. Accurate evaluation depends on a careful examination by an experienced echocardiographer.

Aortic valve area (AVA) is calculated by the continuity equation, based on the concept that volume flow rates proximal to and within the narrowed orifice are equal:

$$AVA \times VTI_{AS} = CSA_{LVOT} \times VTI_{LVOT}$$

or

$$AVA = (CSA_{LVOT} \times VTI_{LVOT})/VTI_{AS}$$

where LVOT is left ventricular outflow tract, VTI is velocity-time integral, CSA is cross-sectional area, and AS is aortic stenosis (Fig. 55-8). It is especially important to calculate the AVA when left ventricular systolic dysfunction accompanies aortic valve disease. In some patients, low-dose dobutamine stress echocardiography is helpful in distinguishing ventricular dysfunction caused by severe aortic stenosis from primary myocardial disease with concurrent moderate stenosis.

The evaluation of mitral stenosis (see Fig. 75-4) includes measurement of the mean transmitral gradient from the velocity curve and calculation of the valve area, both from two-dimensional planimetry of a short axis image of the orifice and from the deceleration slope of the Doppler curve (pressure half-time method).

Valvular Regurgitation

The current approach to evaluating valvular regurgitation (see Fig. 75-8) is based on the proximal geometry of the regurgitant jet, with measurement of the narrowest jet width (vena contracta). When further quantitation is needed, regurgitant volume, regurgitant fraction, and regurgitant orifice area are calculated. Although color flow visualization of the flow disturbance may be helpful for detection of regurgitation and for understanding the mechanism of valve dysfunction, this approach should no longer be used to evaluate severity.

For aortic regurgitation, a narrow vena contracta (<3 mm) indicates mild regurgitation, whereas a wide vena contracta (>6 mm) indicates severe regurgitation. Additional evaluation of the severity of aortic regurgitation is based on the presence of holodiastolic flow reversal in the abdominal aorta and the density and slope of the continuous-wave Doppler velocity curve. The approach to evaluating mitral regurgitation (see Fig. 75-6) is similar, beginning with measurement of the vena contracta. In addition to calculation based on transmitral versus transaortic volume flow rates, the proximal acceleration of flow into the regurgitant orifice allows evaluation with central regurgitant jets. Color flow shows a proximal isovelocity surface area.

Pulmonary Pressures

Estimation of pulmonary artery systolic pressure (PAP) is a standard component of a complete examination.[7] The systolic pressure difference between the right ventricle and right atrium is calculated from the peak velocity in the tricuspid regurgitant (V_{TR}) jet, with use of the Bernoulli equation. Then, the right atrial pressure (RAP) is estimated from the size and appearance of

the inferior vena cava. Because right ventricular and pulmonary artery systolic pressures are equal (in the absence of pulmonic stenosis),

$$PAP = 4(V_{TR})^2 + RAP$$

A small amount of tricuspid regurgitation is present in most patients, so pulmonary pressures can be estimated with this approach in more than 90% of patients. Because this approach measures only pulmonary systolic pressure, not pulmonary vascular resistance, invasive evaluation may still be needed in some clinical situations (Chapter 68).

THE ECHOCARDIOGRAPHIC EXAMINATION

Clinical Indications

Echocardiography is not useful for screening of the general population, but it is an effective approach to the initial evaluation of many cardiac signs and symptoms (Table 55-5).[8] Even when transesophageal imaging might be helpful, most clinicians begin with a transthoracic examination; exceptions are for the patient with a possible acute aortic dissection (Chapter 78), in whom transesophageal echocardiography should be performed as quickly as possible, and in the evaluation of possible left atrial thrombosis before cardioversion without anticoagulation (Chapter 64). Resting echocardiography is not helpful for diagnosis of coronary artery disease; stress imaging is needed if this diagnosis is suspected (Chapter 71). In patients with known cardiac disease, echocardiography is used to evaluate severity, to assess the results of medical and surgical interventions, and to guide procedures. Point-of-care ultrasound is useful for rapid screening to evaluate overall left ventricular function and to detect pericardial effusion (Fig. 55-9; Video 55-5).

Normal Findings

Trace to mild valve regurgitation is considered "physiologic" and is seen with 70 to 80% of mitral valves, 80 to 90% of tricuspid valves, and 70 to 80% of pulmonic valves in normal individuals. The prevalence of aortic regurgitation increases with age, but it is found in only 5% of young normal adults; the presence of aortic regurgitation raises the possibility of subtle aortic valve or root abnormalities.

A PFO (Chapter 69) is present in 25 to 35% of normal individuals and may be identified by color Doppler or by contrast echocardiography. Use of

TABLE 55-5 INDICATIONS FOR TRANSTHORACIC ECHOCARDIOGRAPHY (TTE) BY KNOWN DIAGNOSIS

CLINICAL DIAGNOSIS	KEY ECHOCARDIOGRAPHIC FINDINGS	LIMITATIONS OF ECHOCARDIOGRAPHY	ALTERNATIVE APPROACHES
VALVULAR HEART DISEASE (CHAPTER 75)			
Valve stenosis	Cause of stenosis, valve anatomy Transvalvular ΔP, valve area Chamber enlargement and hypertrophy LV and RV systolic function Associated valvular regurgitation	Possible underestimation of the stenosis severity Possible coexisting coronary artery disease	Cardiac catheterization; CMR
Valve regurgitation	Mechanism and cause of regurgitation Severity of regurgitation Chamber enlargement LV and RV systolic function PA pressure estimate	TEE may be needed to evaluate mitral regurgitant severity and valve anatomy (especially before MV repair)	Cardiac catheterization; CMR
Prosthetic valve function	Evidence for stenosis Detection of regurgitation Chamber enlargement Ventricular function PA pressure estimate	TTE is limited by shadowing and reverberations TEE is needed for suspected prosthetic MR due to "masking" of the LA on TTE	Cardiac catheterization; fluoroscopy
Endocarditis (Chapter 76)	Detection of vegetations (TTE sensitivity, 70-85%) Presence and degree of valve dysfunction Chamber enlargement and function Detection of abscess Possible prognostic implications	TEE more sensitive for detection of vegetations (>90%) A definite diagnosis of endocarditis also depends on bacteriologic criteria TEE more sensitive for detection abscess	Blood cultures and clinical findings also are diagnostic criteria for endocarditis
CORONARY ARTERY DISEASE			
Acute myocardial infarction (Chapters 72 and 73)	Segmental wall motion abnormality reflects "myocardium at risk" Global LV function (EF) Complications Acute MR vs. VSD Pericarditis LV thrombus, aneurysm, or rupture RV infarct	Coronary artery anatomy itself is not directly visualized	Coronary angiography (catheterization or CT) Radionuclide or PET imaging for myocardial perfusion
Angina (Chapter 71)	Global and segmental LV systolic function Exclude other causes of angina (e.g., AS, HCM)	Resting wall motion may be normal despite significant CAD Stress echocardiography is needed to induce ischemia and wall motion abnormality	Coronary angiography Radionuclide or PET imaging ETT
Pre-revascularization/ post-revascularization	Assess wall thickening and endocardial motion at baseline Improvement in segmental function after procedure	Dobutamine stress or contrast echocardiography is needed to detect viable but nonfunctioning myocardium	CMR Coronary angiography Radionuclide or PET imaging Contrast echocardiography
End-stage ischemic disease	Overall LV systolic function (EF) PA pressures Associated MR LV thrombus RV systolic function	—	Coronary angiography (Cath or CT) Radionuclide or PET imaging CMR for myocardial viability
CARDIOMYOPATHY (CHAPTERS 58-60)			
Dilated	Chamber dilation (all four) LV and RV systolic function (qualitative and EF) Coexisting atrioventricular valve regurgitation PA systolic pressure LV thrombus	Indirect measures of LVEDP Accurate EF may be difficult if image quality is poor	Radionuclide EF LV and RV angiography

VIDEO 55-5. Pericardial effusion. A moderate pericardial effusion (PE) is seen in parasternal long axis (**A**), short axis (**B**), and apical four-chamber (**C**) views.

TABLE 55-5 INDICATIONS FOR TRANSTHORACIC ECHOCARDIOGRAPHY (TTE) BY KNOWN DIAGNOSIS—cont'd

CLINICAL DIAGNOSIS	KEY ECHOCARDIOGRAPHIC FINDINGS	LIMITATIONS OF ECHOCARDIOGRAPHY	ALTERNATIVE APPROACHES
Restrictive	LV wall thickness LV systolic function LV diastolic function PA systolic pressure	Must be distinguished from constrictive pericarditis	Cardiac catheterization with direct, simultaneous RV and LV pressure measurement after volume loading CMR
Hypertrophic	Pattern and extent of LV hypertrophy Dynamic LVOT obstruction (imaging and Doppler) Coexisting MR Diastolic LV dysfunction	Exercise echo to detect inducible LV outflow tract obstruction	CMR Strain and strain rate imaging
HYPERTENSION (CHAPTER 67)			
	LV wall thickness and chamber dimensions LV mass LV systolic function Aortic root dilation	Diastolic dysfunction precedes systolic dysfunction, but detection is challenging because of age and other factors	Speckle tracking; strain and strain rate imaging LV twist and torsion
PERICARDIAL DISEASE (CHAPTER 77)			
	Pericardial thickening Detection, size, and location of PE Two-dimensional signs of tamponade physiology Doppler signs of tamponade physiology	Diagnosis of tamponade is a hemodynamic and clinical diagnosis Constrictive pericarditis is a difficult diagnosis Not all patients with pericarditis have an effusion	Intracardiac pressure measurements for tamponade or constriction CMR or CT to detect pericardial thickening
DISEASES OF THE AORTA (CHAPTER 78)			
Aortic root dilation	Cause of aortic dilation Accurate aortic root diameter measurements Anatomy of sinuses of Valsalva (especially Marfan syndrome) Associated aortic regurgitation	The ascending aorta is only partially visualized on TTE in most patients	CT CMR TEE
Aortic dissection	Two-dimensional images of ascending aorta, aortic arch, descending thoracic and proximal abdominal aorta Imaging of dissection "flap" Associated aortic regurgitation Ventricular function	TEE more sensitive (97%) and more specific (100%) Cannot assess distal vascular beds	Aortography CT CMR TEE
CARDIAC MASSES (CHAPTER 60)			
LV thrombus	High sensitivity and specificity for diagnosis of LV thrombus Suspect with apical wall motion abnormality or diffuse LV systolic dysfunction	Technical artifacts can be misleading 5-MHz or higher frequency transducer and angulated apical views needed	LV thrombus may not be recognized on radionuclide or contrast angiography
LA thrombus	Low sensitivity for detection of LA thrombus, although specificity is high Suspect with LA enlargement, MV disease	TEE needed to detect LA thrombus reliably	TEE
Cardiac tumors	Size, location, and physiologic consequences of tumor mass	Extracardiac involvement is not well seen Cannot distinguish benign from malignant tumor or tumor from thrombus	TEE CT CMR (with cardiac gating) Intracardiac echocardiography
PULMONARY HYPERTENSION (CHAPTER 68)			
	PA pressure estimate Evidence of left-sided heart disease to account for increased PA pressures RV size and systolic function (cor pulmonale) Associated TR	Indirect PA pressure measurement Cannot determine pulmonary vascular resistance accurately	Cardiac catheterization
CONGENITAL HEART DISEASE (CHAPTER 69)			
	Detection and assessment of anatomic abnormalities Quantitation of physiologic abnormalities Chamber enlargement Ventricular function	No direct intracardiac pressure measurements Complicated anatomy may be difficult to evaluate if image quality is poor (TEE is helpful)	CMR with three-dimensional reconstruction Cardiac catheterization TEE Three-dimensional echocardiography

AS = aortic stenosis; CAD = coronary artery disease; CMR = cardiac magnetic resonance; CT = computed tomography; EF = ejection fraction; ETT = exercise treadmill test; HCM = hypertrophic cardiomyopathy; LA = left atrium; LV = left ventricle; LVEDP = left ventricular end-diastolic pressure; LVOT = left ventricular outflow tract; MR = mitral regurgitation; MV = mitral valve; ΔP = pressure gradient; PA = pulmonary artery; PE = pericardial effusion; PET = positron emission tomography; RV = right ventricle; TEE = transesophageal echocardiography; TR = tricuspid regurgitation; TTE = transthoracic echocardiography; VSD = ventricular septal defect.
From Otto CM. *Textbook of Clinical Echocardiography.* 5th ed. Philadelphia: Elsevier Saunders; 2013:507-509.

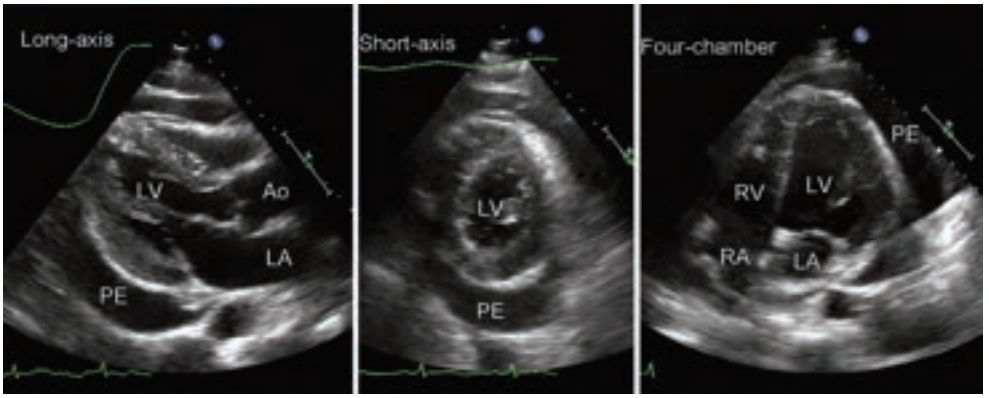

FIGURE 55-9. Pericardial effusion. A large echo-free space is seen anterior and posterior to the cardiac structure in the parasternal long axis view (*left*), short axis view (*center*), and apical four-chamber view (*right*) consistent with a pericardial effusion (PE). Ao = aorta; LA = left atrium; LV = left ventricle; RA = right atrium; RV = right ventricle. See Video 55-5.

the Valsalva maneuver enhances identification of a PFO because the slight elevation in right atrial pressure may lead to a brief right-to-left shunt. The significance of a PFO in patients without clinical events is unclear. Other common anatomic variants seen on echocardiography include aberrant chords (or "webs") in the left ventricle; small, linear, mobile echoes associated with the valves (Lambl's excrescences); and normal ridges in the left and right atria.

Unexpected abnormal findings also may be found on studies requested for other indications. A bicuspid aortic valve is present in 1 to 2% of the population; most of these patients are asymptomatic until late in life, so many cases are diagnosed "incidentally" by echocardiography. Aortic valve sclerosis, which is a frequent unexpected echocardiographic diagnosis, is a marker of cardiovascular disease and an increased risk of myocardial infarction even if valve function is normal.

INTEGRATING THE ECHOCARDIOGRAPHIC AND CLINICAL FINDINGS

The echocardiographic request should indicate the specific reason for the study and any relevant symptoms or signs. The echocardiographic examination then can be tailored to answer the clinical question. The echocardiographic results should be interpreted in conjunction with other clinical data.[8] If the echocardiographic data seem discrepant with the clinical data, the requesting physician should review the images with the echocardiographer to identify areas of uncertainty and to determine the next best diagnostic step.

Grade A Reference

A1. Lindekleiv H, Lochen ML, Mathiesen EB, et al. Echocardiographic screening of the general population and long-term survival: a randomized clinical study. *JAMA Intern Med*. 2013;173:1592-1598.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Otto CM. *Textbook of Clinical Echocardiography*. 5th ed. Philadelphia: Elsevier Saunders; 2013.
2. Dorosz JL, Lezotte DC, Weitzenkamp DA, et al. Performance of 3-dimensional echocardiography in measuring left ventricular volumes and ejection fraction: a systematic review and meta-analysis. *J Am Coll Cardiol*. 2012;59:1799-1808.
3. Mor-Avi V, Lang RM, Badano LP, et al. Current and evolving echocardiographic techniques for the quantitative evaluation of cardiac mechanics: ASE/EAE consensus statement on methodology and indications endorsed by the Japanese Society of Echocardiography. *Eur J Echocardiogr*. 2011;12:167-205.
4. Rosca M, Lancellotti P, Popescu BA, et al. Left atrial function: pathophysiology, echocardiographic assessment, and clinical applications. *Heart*. 2011;97:1982-1989.
5. Via G, Hussain A, Wells M, et al. International evidence-based recommendations for focused cardiac ultrasound. *J Am Soc Echocardiogr*. 2014;27:683.e1-683.e33.
6. Abe Y, Ito M, Tanaka C, et al. A novel and simple method using pocket-sized echocardiography to screen for aortic stenosis. *J Am Soc Echocardiogr*. 2013;26:589-596.
7. Milan A, Magnino C, Veglio F. Echocardiographic indexes for the non-invasive evaluation of pulmonary hemodynamics. *J Am Soc Echocardiogr*. 2010;23:225-239.
8. Douglas PS, Garcia MJ, Haines DE, et al. ACCF/ASE/AHA/ASNC/HFSA/HRS/SCAI/SCCM/SCCT/SCMR 2011 Appropriate Use Criteria for Echocardiography. A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, American Society of Echocardiography, American Heart Association, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, Society of Critical Care Medicine, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance Endorsed by the American College of Chest Physicians. *J Am Coll Cardiol*. 2011;57:1126-1166.

REVIEW QUESTIONS

1. Echocardiographic three-dimensional left ventricular volumes measured in a 55-year-old man receiving cardiotoxic chemotherapy are 120 mL at end diastole and 60 mL at end systole. Left ventricular ejection fraction is

A. 30%
B. 40%
C. 50%
D. 60%
E. 70%

Answer: C Left ventricular ejection fraction (EF) is calculated as stroke volume (end-diastolic volume [EDV] minus end-systolic volume [ESV]) divided by end-diastolic volume multiplied by 100%. In this case:

$$EF = (EDV - ESV)/EDV \times 100\% = (120 - 60)/60 \times 100\% = 50\%$$

2. Echocardiographic evaluation in an 82-year-old woman with a systolic murmur shows a heavily calcified aortic valve. The severity of valve obstruction is best evaluated by

A. Three-dimensional imaging
B. Pulsed Doppler
C. Tissue Doppler
D. Color flow Doppler imaging
E. Continuous-wave Doppler

Answer: E Valve obstruction with reduced systolic opening of the aortic valve results in an increased velocity across the valve. Velocity is a direct measure of the severity of the stenosis and predicts clinical outcome. The velocity (v) across a stenotic valve is related to the pressure gradient (ΔP) as stated in the Bernoulli equation, $\Delta P = 4v^2$. Measurement of high velocities requires continuous-wave Doppler ultrasound, which typically is not available on point-of-care ultrasound systems. Three-dimensional imaging allows more accurate identification of the number of valve leaflets and may show the degree of valve opening, but it is not accurate for measuring the severity of stenosis. Pulsed Doppler velocity measurements are accurate only at low velocities. Tissue Doppler measures the velocity of myocardial motion, which reflects left ventricular systolic and diastolic function, not the severity of stenosis. Color Doppler flow imaging is useful for visualizing the location of intracardiac flow disturbances but cannot measure high-flow velocities.

3. A 68-year-old man presents with a 2-week history of increasing dyspnea and is found to be in atrial fibrillation with a ventricular rate of 120 beats per minute. The most appropriate test to evaluate for left atrial thrombus before cardioversion is

A. Point-of-care echocardiography
B. Transthoracic echocardiography
C. Intracardiac echocardiography
D. Transesophageal echocardiography
E. No testing is needed.

Answer: D In patients being considered for cardioversion for atrial fibrillation, clinical approaches to avoiding systemic embolization with restoration of normal sinus rhythm include effective anticoagulation for several weeks before cardioversion or exclusion of a left atrial thrombus by visualizing the left atrial appendage. Left atrial thrombi cannot be accurately diagnosed by transthoracic or point-of-care echocardiography because atrial thrombi occur most often in the atrial appendage, which is distant from the transducer and difficult to visualize. Transesophageal echocardiography provides superior images of the left atrial appendage and is reliable for exclusion of atrial thrombus when images are obtained in at least two orthogonal views with use of zoom mode and a high-frequency transducer. Intracardiac echocardiography also may provide images of the left atrial appendage and may be used during an electrophysiologic or interventional cardiology procedure. However, intracardiac echocardiography is not used as a stand-alone diagnostic test.

4. A 63-year-old woman presents with a 6-month history of midsternal chest pressure that occurs both at rest and with exertion. Her only cardiac risk factor is chronic hypertension with medical therapy. She also has osteoarthritis of her knees and is able to walk only short distances. The baseline electrocardiogram shows left ventricular hypertrophy with a strain pattern. An appropriate diagnostic test to evaluate this patient for coronary disease is

A. Treadmill electrocardiographic stress test
B. Treadmill echocardiographic stress test
C. Supine bicycle nuclear perfusion stress test
D. Dobutamine stress echocardiography
E. Coronary angiography

Answer: D Coronary artery disease may explain symptoms in this patient, so stress testing is appropriate for diagnosis and risk stratification. Electrocardiographic stress testing will not be diagnostic because her abnormal baseline electrocardiogram limits the accuracy of changes in ST segments with exercise for diagnosis of coronary disease. This patient also has significant lower extremity arthritis, which limits her ability to exercise. Diagnostic accuracy with an exercise test (either treadmill or bicycle) requires that the patient reach 85% of the maximum predicted heart rate for age. Exercise testing is likely to be nondiagnostic in this patient with lower extremity arthritis. Pharmacologic stress testing, either dobutamine stress echocardiography or pharmacologic nuclear perfusion imaging, is the most appropriate test choice. Coronary angiography would be appropriate only if stress testing suggests high-risk coronary disease that might benefit from revascularization.

5. A 68-year-old man presents to the emergency department with a 1-hour history of severe chest pain and diaphoresis. The electrocardiogram shows ST depression in the lateral leads. Point-of-care echocardiography is appropriate to evaluate for

A. Aortic dissection
B. Aortic valve stenosis
C. Acute anterior infarction
D. Mitral valve vegetation
E. Pericardial effusion

Answer: E All of these diseases can be accurately diagnosed by a complete transthoracic or transesophageal echocardiographic study under the supervision of a cardiologist. Point-of-care ultrasound systems typically have fewer features and limited image quality compared with complete diagnostic ultrasound systems. In addition, point-of-care ultrasound is performed and interpreted by physicians with limited training in this modality. The recommended scope of practice for point-of-care ultrasound studies includes diagnosis of pericardial effusion. In contrast, diagnosis of aortic dissection requires transesophageal echocardiographic or computed tomographic or magnetic resonance imaging. Diagnosis of aortic valve stenosis requires continuous-wave Doppler recording of the transvalvular velocity by an experienced sonographer. Acute anterior myocardial infarction results in hypokinesis or akinesis of the anterior wall, which can be visualized by echocardiography. Point-of-care ultrasound allows evaluation of global left ventricular systolic function, but evaluation of regional myocardial dysfunction is more challenging and requires additional training and experience. A mitral valve vegetation might be diagnosed on transthoracic imaging, but transesophageal imaging often is required. Mitral valve vegetations are likely to missed with current point-of-care ultrasound systems or by less experienced clinicians.

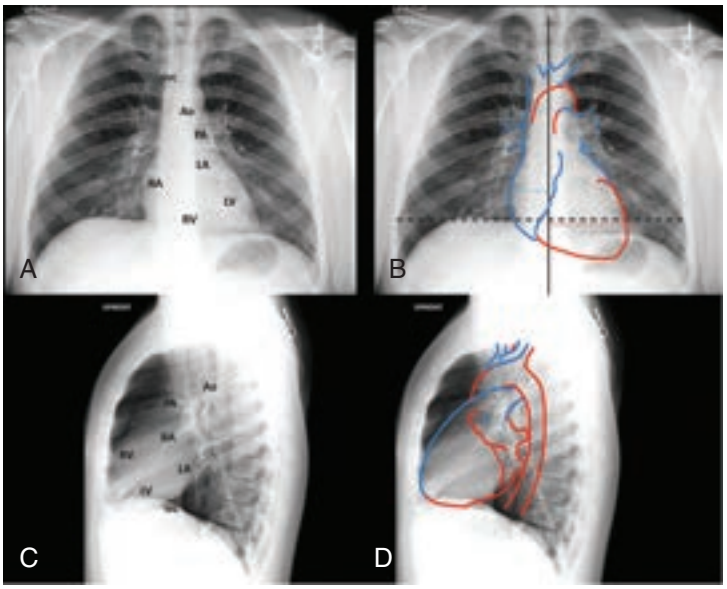

FIGURE 56-1. Normal anatomy. Posteroanterior (**A**) and lateral (**B**) chest radiograph projections in a healthy 28-year-old man. The cardiac chambers and the great vessels are marked on the corresponding drawings (**C, D**). The cardiothoracic ratio (**C**) is calculated by dividing the maximum transverse diameter of the cardiac silhouette (*blue line:* widest distance of the right heart border from the midpoint of the spine; *orange line:* widest distance from the left heart border to the midpoint of the spine) through the distance between the internal margin of the ribs at the top of the right diaphragm (*black line*). Ao = aorta; IVC = inferior vena cava; LA = left atrium; LV = left ventricle; RA = right atrium; RV= right ventricle; PA = pulmonary artery; SVC = superior vena cava.

56

NONINVASIVE CARDIAC IMAGING

CHRISTOPHER M. KRAMER, GEORGE A. BELLER, AND KLAUS D. HAGSPIEL

RADIOGRAPHY OF THE HEART

Chest radiography is a widely available, relatively inexpensive, and rapid imaging modality, with an average effective radiation dose of 0.03 to 0.1 mSv. The heart is best evaluated on posteroanterior (PA) and lateral radiographs, with the heart closest to the image detector.

On the chest radiograph, the heart appears as a homogeneous shadow surrounded by lung, so diagnostic assessment of the heart and great vessels is based on the size and shape of the cardiac silhouette, rather than on direct visualization of the heart's internal anatomy. Nevertheless, the size and shape of the heart and their changes over time, together with the appearance of the pulmonary vasculature, aid in the diagnosis of cardiac diseases. The radiographic appearance of the heart is also influenced by the radiographic technique, projection, body habitus, degree of inspiration, and whether the patient is supine or erect during the examination.

On the PA projection, the normal heart is located in the middle mediastinum, with approximately two thirds projecting to the left of the mid sternum (Fig. 56-1). The superior segment of the right heart border is a more or less straight line formed by the superior vena cava and right innominate vein. The inferior segment is convex and formed by the right atrium. The left heart border consists of three segments: the aortic arch superiorly, the main pulmonary artery in the middle, and the left ventricle (LV) in the longest segment inferiorly. The left atrial appendage is situated in the junction between the lower and middle segments; if enlarged, it can appear as a separate, prominent segment. The inferior border of the heart sometimes is not well differentiated from the diaphragm (see Fig. 56-1).

On lateral chest images, the right ventricle (RV) and the RV outflow tract form the anterior heart border, with the RV in contact with the lower third of the sternum. Lung between the sternum and the posteriorly curving RV and RV outflow tract forms the retrosternal clear space. The posterior heart border, which is seen between the carina and the diaphragm on lateral images, consists of the left atrium superiorly and the LV inferiorly. The inferior vena cava courses obliquely upward before it joins the right atrium. The ascending aorta, aortic arch, and proximal descending thoracic aorta are usually well seen on lateral images (see Fig. 56-1).

Alterations of the contour of the heart are generally caused by dilation of the atria, ventricles, or blood vessels. Chest images are not sensitive for detecting cardiac hypertrophy unless it is severe.

Comprehensive cardiovascular analysis of chest radiographs requires evaluation of the size and morphology of the heart and the great vessels, the pulmonary vasculature, and the presence and positioning of any calcifications or implanted devices such as valves, pacemakers, and defibrillators.

Radiographic Assessment of Heart Size

The cardiothoracic ratio (see Fig. 56-1) estimates the size of the heart. A value of less than 0.5 is considered normal on a radiograph taken during deep inspiration. Pectus excavatum deformities (Chapter 99) and epicardial fat pads can result in an abnormally large cardiothoracic ratio despite a normal-sized heart. Dilation of the LV, which increases the cardiothoracic ratio, appears as a concave mid left heart border and lengthening of the entire left heart border, with a downward pointing apex projecting below the diaphragm on PA views (Fig. 56-2). Extension of the posterior margin of the left ventricle more than 2 cm posterior to the inferior vena cava on the lateral film is considered a sign of left ventricular enlargement. Localized rather than global LV enlargement usually indicates the presence of ventricular aneurysms (Chapter 73). LV hypertrophy without dilation usually is not detectable on chest radiographs.

The most common cause of left atrial enlargement is secondary to LV dysfunction, especially LV dilation. Isolated enlargement of the left atrium is usually a sequela of mitral valve abnormalities or atrial fibrillation. Enlargement of the left atrium can also occur with LV hypertrophy without dilation in patients with aortic stenosis (Chapter 75) or hypertrophic cardiomyopathy (Chapter 60). Straightening of the left heart border between the main pulmonary artery and the LV just below the left main bronchus owing to enlargement of the left atrial appendage is one of the earliest signs of left atrial enlargement; with increasing size, this segment becomes convex. A double contour within the right cardiac border (double density sign), splaying of the carina, and elevation of the right main bronchus are less frequent signs (Fig. 56-3).

The RV normally does not form part of the heart's border on PA radiographs, but significant RV enlargement can lead to an abnormal convexity of the left heart border, with elevation and leftward displacement of the cardiac apex. The best radiographic indication of RV enlargement is obliteration of the retrosternal clear space in the lateral view owing to dilation of the RV outflow tract (Fig. 56-4).

Dilation of the right atrium causes the lower segment of the right heart border to become more prominent and increasingly round. In more severe cases, the entire right heart border is enlarged, and in extreme cases, the right atrium can become border-forming on the lateral view (see Fig. 56-3).

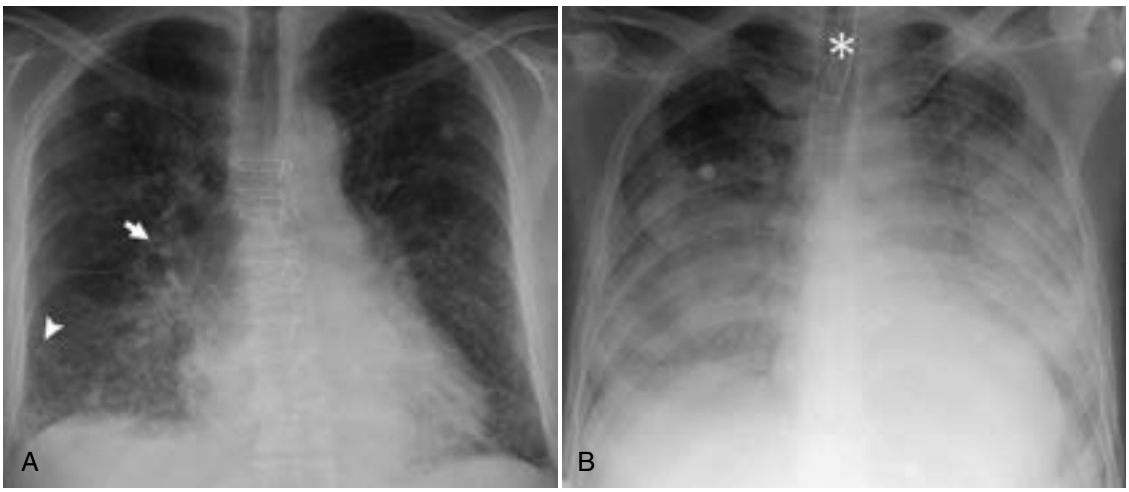

FIGURE 56-2. Left ventricular enlargement and pulmonary edema in two different patients. The image on the left (**A**) was obtained in a patient who had acute myocardial infarction and who had previously undergone aortic bypass grafting (sternotomy wires and bypass clips are seen). Interstitial edema is evidenced by the presence of vascular redistribution, Kerley B lines (*arrowhead*), and peribronchial cuffing (*arrow*). The image on the right (**B**) was obtained in a patient with cardiogenic alveolar edema. The left ventricle is significantly enlarged, and there is extensive bilateral air space consolidation with air bronchogram. Note normal position of an endotracheal tube (*asterisk*) and a nasogastric tube.

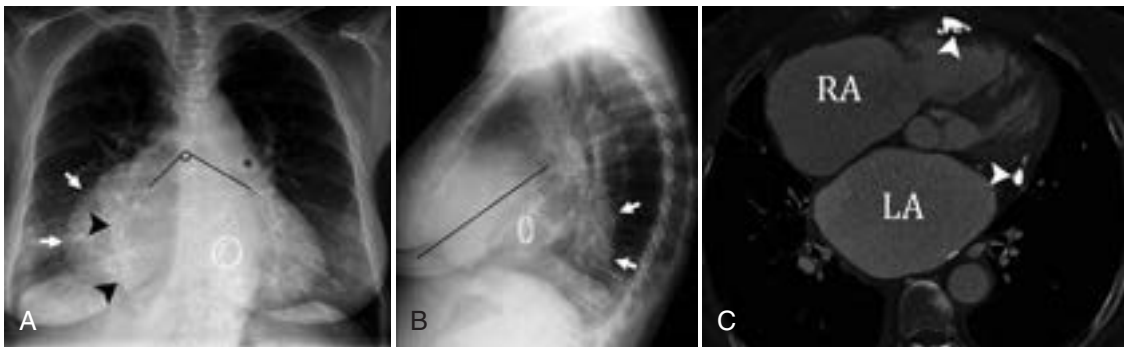

FIGURE 56-3. Biatrial enlargement in a patient who has undergone mitral valve repair. The posteroanterior (**A**) chest radiograph shows a prominent left atrial appendage (*asterisk*), the double density sign (*arrowheads*), and splaying of the carina. There is also enlargement of the right atrium, as evidenced by prominence and round shape of the right side border on the posteroanterior view (*arrows*). Posterior bulging of the heart (*arrows*) owing to biatrial enlargement is seen on the lateral view (**B**). Interstitial pulmonary edema is present. The location of the cardiac valves is best assessed on a lateral view (**B**) by drawing a line from the carina to the anterior costophrenic recess (*black line*). The pulmonic and aortic valves usually are positioned superior to this line, and the tricuspid and mitral valves are inferior to it. A computed tomographic image (**C**) also demonstrates biatrial enlargement as well as pericardial calcifications. LA = left atrium; RA = right atrium.

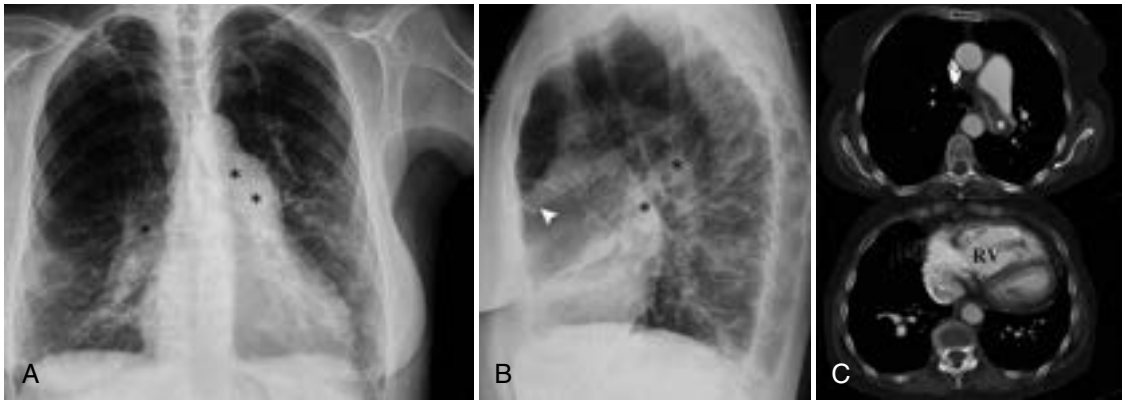

FIGURE 56-4. Pulmonary arterial hypertension in a patient with severe pulmonary emphysema and chronic pulmonary embolism. Enlarged central pulmonary arteries (*asterisks*) and the pruned-tree sign can be seen on the posteroanterior (**A**) and lateral (**B**) projections. Hyperlucency in both upper lobes is present owing to extensive pulmonary emphysema. Right ventricular enlargement has obliterated the retrosternal clear space on the lateral view (*arrowhead*) (**B**). The computed tomographic scan demonstrates dilated central pulmonary arteries with wall adherent chronic pulmonary embolus (*asterisk*) (**C**) as well as right ventricle (RV) dilation and hypertrophy.

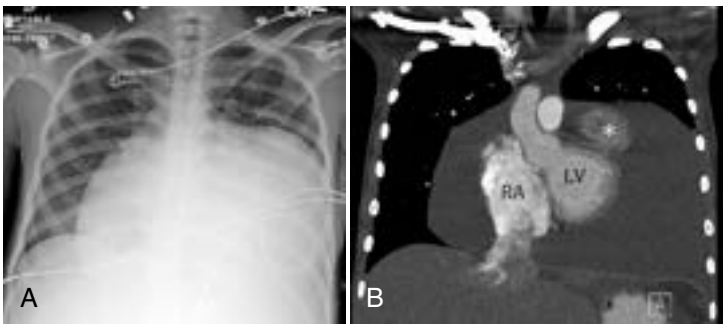

FIGURE 56-5. A 17-year-old girl presenting with massive pericardial effusion. The posteroanterior radiograph (**A**) shows the water-bottle sign; the hilar vessels are obscured. A coronal reformatted computed tomographic scan (**B**) shows the fluid in the pericardial sac as well as dilation of the right atrium and the left atrial appendage (*asterisk*). LV, left ventricle; RA, right atrium.

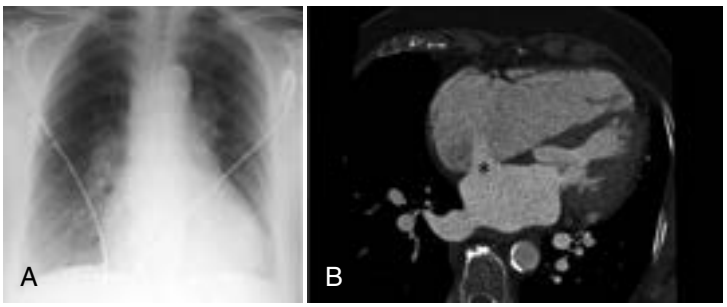

FIGURE 56-6. Female patient with known secundum type atrial septal defect. The posteroanterior radiograph (**A**) shows enlargement of the pulmonary arteries, shunt vascularity, and enlargement of the right heart border. A computed tomographic scan (**B**) shows the septal defect (*asterisk*) with left-to-right shunt. Enlargement of the right heart is also seen.

Pericardial Effusion

Large pericardial effusions cause significant enlargement of the cardiac silhouette despite a normal superior mediastinum—the so-called water-bottle sign. The fluid in the pericardial sac will obscure the hilar vessels on a PA chest film (Fig. 56-5), unlike cardiomegaly without effusion, in which the hilar structures often are relatively conspicuous. Posterior displacement of the pericardial fat line on lateral images is also a valuable finding for the detection of pericardial effusions.

Pulmonary Vasculature

The large and medium-sized pulmonary arteries and veins can be seen on the radiograph as linear shadows, and their size and appearance correlate with pulmonary blood flow and pulmonary venous pressure. The vessels in the lower lung zones are normally larger than in the upper zones as a result of the normal distribution of pulmonary blood flow (see Fig. 56-1). In patients with right-to-left shunts (Chapter 69), the pulmonary vasculature is decreased in caliber. In left-to-right shunts, the vascularity is increased, and the vessels are sharply outlined if the patient does not have heart failure (Fig. 56-6). With increasing pulmonary venous pressure, as is seen in heart failure (Chapters 58 and 59), the vessels in the upper zone enlarge on the chest radiograph. With further increases in pulmonary venous pressures, fluid extravasates into the pulmonary interstitium, and the pulmonary vessels lose their sharp demarcation. Horizontal lines in the periphery of the lungs (Kerley B) and vertical lines in the upper lobes (Kerley A) represent thickened interlobular septae. Fluid in the interstitium of the bronchial walls causes peribronchial thickening or cuffing (see Fig. 56-2). In more advanced cases, the lungs show a diffuse ground-glass appearance that masks the vascular structures, and pulmonary edema ultimately develops (see Fig. 56-2). Long-standing pulmonary arterial hypertension leads to dilation of the central pulmonary arteries with abrupt change in caliber instead of the normal tapering. The size and number of the peripheral arterial branches diminish, resulting in a pruned-tree appearance (see Fig. 56-4).

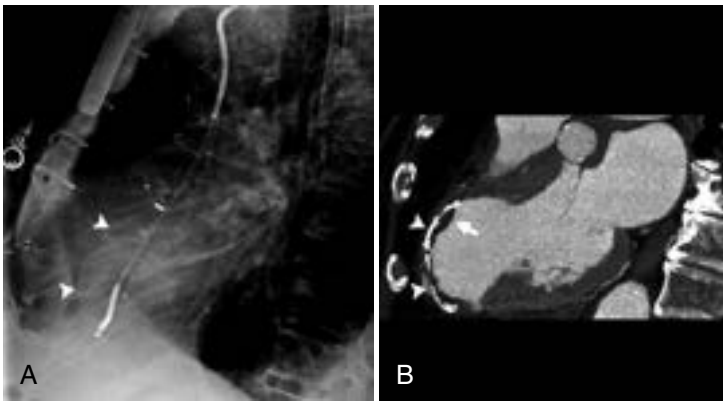

FIGURE 56-7. Calcified chronic myocardial infarction. Lateral radiograph (**A**) demonstrates thin curvilinear calcifications (*arrowheads*) in the region of the left ventricular apex. A reformatted cardiac computed tomographic scan (**B**) shows aneurysmal dilation and thinning of the apex, curvilinear calcifications, and thrombus (*arrow*). Dual-chamber pacer leads, aortocoronary bypass clips, and fracture of the most inferior sternotomy wire (*asterisk*) are also seen on the lateral radiograph (**A**).

Calcifications

Calcifications can often be seen on chest radiographs. In the heart, calcifications most frequently involve the valves and the mitral annulus (Chapter 75). Most aortic valvular calcifications are degenerative in nature and occur in otherwise normal valves, but their incidence is increased in bicuspid valves or in patients who have had rheumatic fever. Calcification of the mitral annulus is common and usually an asymptomatic finding. Coronary artery calcifications are frequent but are rarely seen on chest radiographs. Calcifications of the ventricles are most often seen in patients with prior myocardial infarction or ventricular aneurysms (Fig. 56-7). Pericardial calcifications tend

to be thicker than calcifications in the myocardium and, in severe cases, can entirely surround the heart (see Fig. 56-3).

Implanted Devices

A great number of devices can be seen on a chest radiograph, and it is important to be familiar with their appearance to assess their correct position and integrity. However, accurate assessment of device position can occasionally require cross-sectional imaging because the radiographic projection, body habitus, degree of inspiration, and patient positioning (supine versus erect) greatly impact appearance of a device on the film (see Figs. 56-2, 56-3, and 56-7).

Characteristic Appearance of Cardiac Silhouette

Certain constellations of findings on chest radiographs can be characteristic of specific disorders. For example, in mitral stenosis, left atrial enlargement, pulmonary venous hypertension, a small aortic knob, and enlargement of the main pulmonary artery are typical findings. In aortic stenosis, LV enlargement, calcifications of the aortic valve, and dilation of the ascending aorta are often present. In pulmonic valve stenosis, enlargement of the pulmonary trunk is the most common radiographic sign, with or without signs of RV enlargement. In atrial septal defects, the heart is usually normal in size, with prominent pulmonary vasculature (shunt vascularity). In ventricular septal defects, the left atrium and LV are prominent, and shunt vascularity is present. A number of classic chest radiographic signs have been described for patients with more complex congenital heart disease (Chapter 69). These findings include the egg-on-a-string sign in transposition of the great arteries, the gooseneck sign in endocardial cushion defects, the boot-shaped heart in tetralogy of Fallot (Fig. 56-8), the figure-of-3 and reversed figure-of-3 signs in coarctation of the aorta, the box-shaped heart in Ebstein anomaly, the snowman sign in total anomalous pulmonary venous return, and the scimitar sign in partial anomalous pulmonary venous return (Fig. 56-8). Although these classic signs are useful when present, they are less frequently encountered than previously because congenital anomalies are diagnosed and treated earlier in life. As a result, chest radiography currently has a limited role in the diagnosis of congenital and acquired heart disease, but it remains a useful technique for monitoring the progression of disease and its response to treatment.

NUCLEAR CARDIOLOGY

The techniques of nuclear cardiology permit the noninvasive imaging of myocardial perfusion under stress and resting conditions and of resting regional and global function by use of radionuclide imaging agents and gamma or positron cameras with associated computer processing. All these techniques are based on acquiring images of radioactivity emanating from tracers localized in heart muscle or in the blood pools of the LV and RV. Myocardial perfusion imaging is the most commonly performed nuclear cardiology technique, most often in conjunction with either exercise or pharmacologic stress intended to produce flow heterogeneity between relatively hypoperfused and normally perfused myocardial regions. Single-photon emission computed tomography (SPECT) is one technique used to image uptake of tracers in the myocardium. Radionuclide angiography, in which technetium-99m (^{99m}Tc)-labeled red blood cells or other ^{99m}Tc-labeled agents are injected intravenously, is used for measurement of LV ejection fraction and assessment of regional wall motion, especially to monitor changes in global LV function in patients undergoing chemotherapy with cardiac toxic drugs. Positron emission tomography (PET) can assess regional myocardial metabolism to estimate myocardial viability, most often with fluorine-18-labeled 2-deoxyglucose (FDG), as well as myocardial perfusion by use of either nitrogen-13 (^{13}N) ammonia or rubidium-82 (^{82}Rb).

Myocardial Perfusion Imaging

IMAGING AGENTS

For the assessment of myocardial perfusion using SPECT technology, ^{99m}Tc-labeled perfusion agents, which provide higher-quality images more quickly, are used more commonly than thallium-201 (^{201}Tl) for exercise or pharmacologic stress perfusion imaging to evaluate patients with suspected or known coronary heart disease (CHD). Of the various ^{99m}Tc-labeled agents, ^{99m}Tc-sestamibi and ^{99m}Tc-tetrofosmin are the most common. These ^{99m}Tc agents permit simultaneous assessment of regional and global LV function and volumes with gated SPECT technology. Advantages of PET myocardial perfusion imaging compared with SPECT include higher sensitivity and specificity for the detection of coronary artery disease with a lower dose of radiation. By comparing stress blood flow to resting flow, PET also permits the quantification of absolute regional myocardial blood flow in mL/min/g and coronary flow reserve.

DETECTION OF CORONARY HEART DISEASE

The major indications for stress and rest myocardial perfusion imaging are to diagnose CHD, to assess prognosis, and to detect myocardial viability.[1] Exercise or pharmacologic stress myocardial perfusion imaging in patients with chest pain yields a sensitivity for detecting CHD of 88% for SPECT and 93% for PET.[2] The specificity for excluding CHD is 76% for SPECT and 81% for PET. Exercise or pharmacologic stress SPECT perfusion imaging has sensitivities and specificities that are superior to those of exercise electrocardiogram (ECG) testing alone. Specificity for SPECT myocardial perfusion imaging is enhanced by inspection of regional function on ECG-gated images and with computer algorithms, which correct for attenuation. Both the sensitivity and specificity for detecting CHD are enhanced by image quantitation.

Myocardial perfusion imaging is of particular value compared with exercise ECG testing alone in (1) patients with resting ECG abnormalities, such as those seen with LV hypertrophy, digitalis effect, Wolff-Parkinson-White syndrome, and intraventricular conduction abnormalities; and (2) patients who fail to achieve more than 85% of maximal predicted heart rate. The addition of stress perfusion imaging can assist in differentiating true-positive from false-positive ST depression. Detection of proximal left anterior descending stenoses and proximal multivessel CHD is enhanced by identifying regional systolic thickening or wall motion abnormalities on the gated SPECT images compared with assessment based on perfusion alone. If possible, drugs such as long-acting nitrates, β-blockers, and rate-lowering calcium blockers should be discontinued for 24 hours before exercise stress testing that is performed to diagnose or to exclude CHD as the cause of chest pain. Advances in gamma camera technology have yielded a new generation of high-speed gamma SPECT cameras that use cadmium zinc telluride semiconductor detectors.

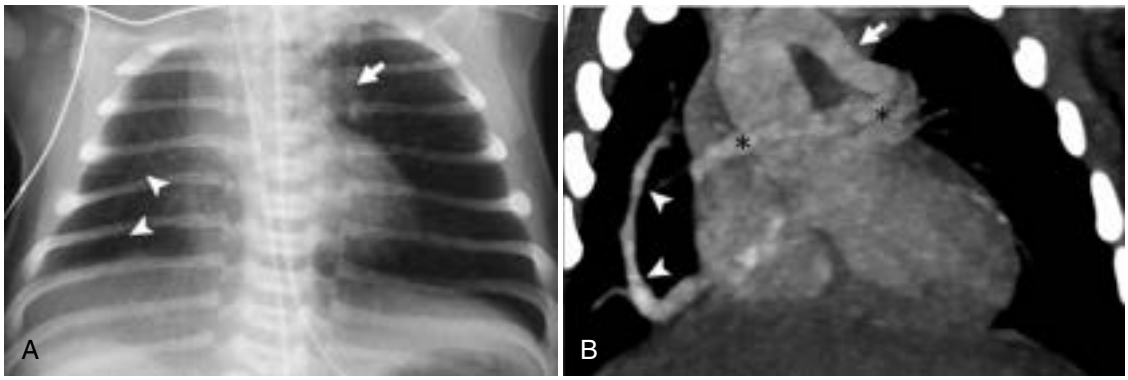

FIGURE 56-8. Infant with tetralogy of Fallot, pulmonary atresia, and partial anomalous pulmonary venous return. Anteroposterior radiograph (**A**) shows the classic boot-shaped heart seen with tetralogy of Fallot; the scimitar sign, owing to abnormal right pulmonary vein draining into the inferior vena cava (*arrowheads*); and abnormal superior parasternal left-sided mediastinal density caused by a left upper pulmonary vein draining into the left innominate vein (*arrow*). Coronal reformatted computed tomography (**B**) confirms these findings. The pulmonary arteries are diminutive (*asterisks*).

This technology has better spatial resolution, permits use of a lower effective dose of tracer for imaging, and requires less time compared with a conventional gamma camera (E-Fig. 56-1).

Employing a "stress-only" protocol for SPECT myocardial perfusion imaging in patients with a low or low-to-intermediate pretest probability of CHD based on clinical variables has reduced radiation exposure and imaging time. The conventional 1-day SPECT protocol entails performing a resting SPECT study first, followed by a stress study a few hours later using higher doses of the tracer. With the stress-only approach, the stress study is done first. If the stress study is normal, the resting study is not performed; but if the stress study is abnormal, then the patient is brought back the following day for the resting study to see if the perfusion defect in question is reversible and indicative of inducible ischemia. The two approaches are equally good for predicting future CHD event rates (E-Fig. 56-2).

Some gamma cameras are combined with a computed tomography (CT) scanner, which allows for multimodality hybrid imaging of anatomy and physiology. Myocardial perfusion imaging can be added to the coronary CT angiographic study if the latter shows an intermediate coronary stenosis of 50 to 70% in diameter. Conversely, the CT angiogram can be performed after an equivocal SPECT study to distinguish between true-positive and false-positive perfusion defects.

PHARMACOLOGIC STRESS IMAGING

In patients who are unable to exercise to 85% of their age-prediction maximum heart rate on exercise stress testing protocols, pharmacologic stress testing with use of vasodilators or dobutamine is an alternative to exercise for detecting physiologically significant coronary artery stenoses. Vasodilator stress SPECT myocardial perfusion imaging can use dipyridamole, adenosine, or regadenoson. The most commonly used vasodilator stress agent is now regadenoson, an A2A adenosine receptor agonist, administered as an intravenous bolus. The addition of limited exercise to adenosine or regadenoson imaging can attenuate the vasodilator-induced decrease in blood pressure and enhance image quality by increasing the heart-to-liver ratio of tracer uptake. Dobutamine stress is preferred in patients who have bronchospasm or a history of asthma or who have consumed caffeine, which is an adenosine receptor antagonist, within 12 hours before testing. Patients who experience side effects such as hypotension and chest pain during vasodilator infusion should be treated with intravenous aminophylline, an adenosine antagonist that immediately reverses these side effects. Regadenoson administration normally increases the heart rate, and its failure to do so is a bad prognostic sign.

ASSESSMENT OF PROGNOSIS

The extent of hypoperfusion on post-stress SPECT perfusion images provides important incremental prognostic information when added to clinical characteristics, the resting LV ejection fraction, exercise ECG stress test variables, and even coronary artery anatomy. Nondiabetic patients with chest pain and a normal myocardial perfusion scan at peak exercise or under vasodilator stress have a subsequent cardiac death or infarction rate of less than 1% per year and are generally appropriate candidates for medical therapy (Chapter 71) or require further diagnostic evaluation for a noncardiac cause of chest pain (Chapters 51 and 137). Conversely, patients with high-risk imaging results may benefit from early referral for invasive strategies, including coronary revascularization (Chapter 74), even if symptoms are mild. Patients who show inducible ischemia (Fig. 56-9) involving more than 10% of the LV myocardium may have a better outcome with coronary revascularization compared with medical therapy.

Transient ischemic LV cavity dilation, by which the LV cavity appears more dilated on stress images compared with rest images, occurs when subendocardial ischemia after stress causes a decrease in tracer uptake in the subendocardium. This finding is particularly predictive of poorer outcomes in diabetic patients who undergo SPECT myocardial perfusion imaging but is of limited value in low-risk patients whose scans are otherwise normal with no defects noted.

Assessment of regional LV function on post-stress gated SPECT images enhances the detection of multivessel CHD. LV ejection fraction and end-systolic and end-diastolic volumes can also be measured by gated SPECT imaging. Stress myocardial perfusion imaging does not appear to add any useful prognostic information over exercise ECG testing alone in patients who achieve 10 metabolic equivalents or more of workload without ischemic ST segment depression on the exercise ECG. Similarly, in low-risk women with good physical capacity, a diagnostic strategy using the exercise ECG alone is as good as exercise SPECT myocardial perfusion imaging for predicting 2-year outcomes at lower cost.[3] One significant limitation of SPECT myocardial perfusion imaging is failure to identify multivessel and left main coronary artery disease in some patients who have balanced ischemia, in which uptake of the imaging agent at peak stress is homogenously diminished throughout the LV myocardium because of diffusely reduced flow in regions supplied by all three major coronary arteries. Quantitative PET myocardial perfusion imaging will identify these patients by demonstrating global reduction in absolute coronary flow.

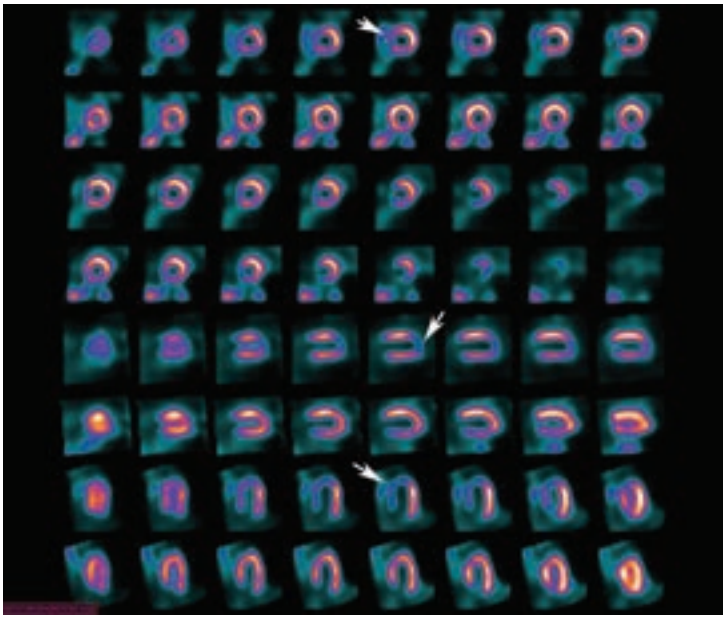

FIGURE 56-9. Exercise stress (rows 1 and 3 from top to bottom) and rest (rows 2 and 4) short axis, stress (row 5) and rest (row 6) vertical long axis, and stress (row 7) and rest (row 8) horizontal long axis single-photon emission computed tomography images showing reversible perfusion defects in the interventricular septum (*white arrows*) and the anteroapical region (*green arrows*). (Reproduced from Beller GA, Bateman TM. Provisional use of myocardial perfusion imaging in patients undergoing exercise stress testing. *J Nucl Cardiol.* 2013;20:711-714.)

DETERMINATION OF MYOCARDIAL VIABILITY WITH SINGLE-PHOTON EMISSION COMPUTED TOMOGRAPHY OR POSITRON EMISSION TOMOGRAPHY

SPECT perfusion imaging is performed in the resting state to identify residual myocardial viability in zones corresponding to severe regional wall motion abnormalities in patients with CHD and depressed LV function. When severe LV dysfunction is caused by "hibernation" (a state of chronic reduced contractility because of substantial ischemia), and not by irreversible myocardial necrosis, areas of resting hypoperfusion that are viable and contributing to hibernation show initial defects on early images but no or less severe defects on 3-hour delayed images. If uptake ultimately exceeds 50 or 60% of peak uptake in these regions, there is a high probability (65 to 75%) that regional myocardial function will improve after successful revascularization, compared with only a 10 to 20% probability for myocardial zones showing less than 50% of peak uptake on resting images.

Regional myocardial metabolism can be assessed noninvasively by PET with FDG and a flow tracer such as [^{13}N]ammonia or ^{82}Rb. FDG is a glucose analogue that is taken up initially in myocardial cells and is trapped by conversion to FDG-6-phosphate. FDG is cell membrane impermeable and remains within viable cells at high concentrations for more than 40 to 60 minutes. Increased FDG activity on clinical PET images in areas of diminished regional blood flow, as determined by [^{13}N]ammonia imaging, is characteristic of myocardial viability. These areas of blood flow–FDG mismatch usually show improved regional function after coronary revascularization. When the extent of viability (hibernation) by PET-FDG exceeds 10% of the LV, revascularization is associated with improved long-term survival compared with medical therapy.[4] Regions of the heart that show both diminished perfusion and FDG uptake (a "match" pattern) represent predominantly nonviable myocardium, with only a 10 to 15% probability of showing improved systolic function after revascularization. Patients who have an ischemic cardiomyopathy with poor viability on either resting SPECT or PET have a worse outcome after coronary revascularization compared with patients with predominantly viable myocardium.

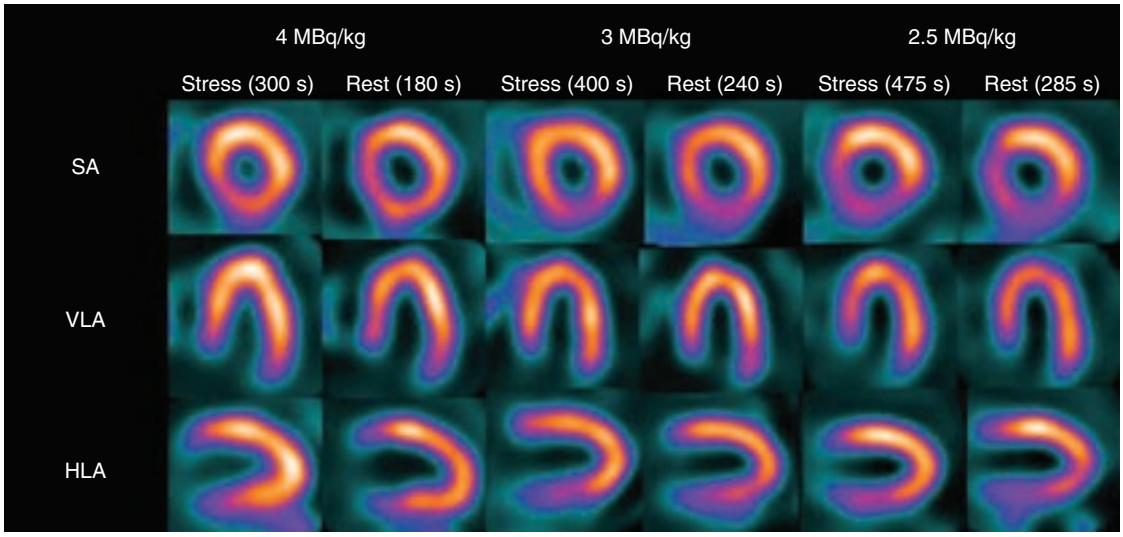

E-FIGURE 56-1. Stress and rest ^{99m}Tc single-photon emission computed tomography short axis (SA), vertical long axis (VLA), and horizontal long axis (HLA) perfusion images. These images depict normal perfused myocardium in patients receiving either low (2.5 MBq/kg) or higher (3 MBq/kg and 4 MBq/kg) stress doses of ^{99m}Tc tetrofosmin with maintenance of good image quality. The total effective radiation dose (stress + rest) decreased by 9.3 mSv, comparing the lowest with the highest administered dose of tracer. Image acquisition time in seconds (s) was increased in proportion to the decrease in administered activity. (Reproduced from Oddstig J, Hedeer F, Jögi J, et al. Reduced administered activity, reduced acquisition time, and preserved image quality for the new CZT camera. *J Nucl Cardiol.* 2013;20:38-44.)

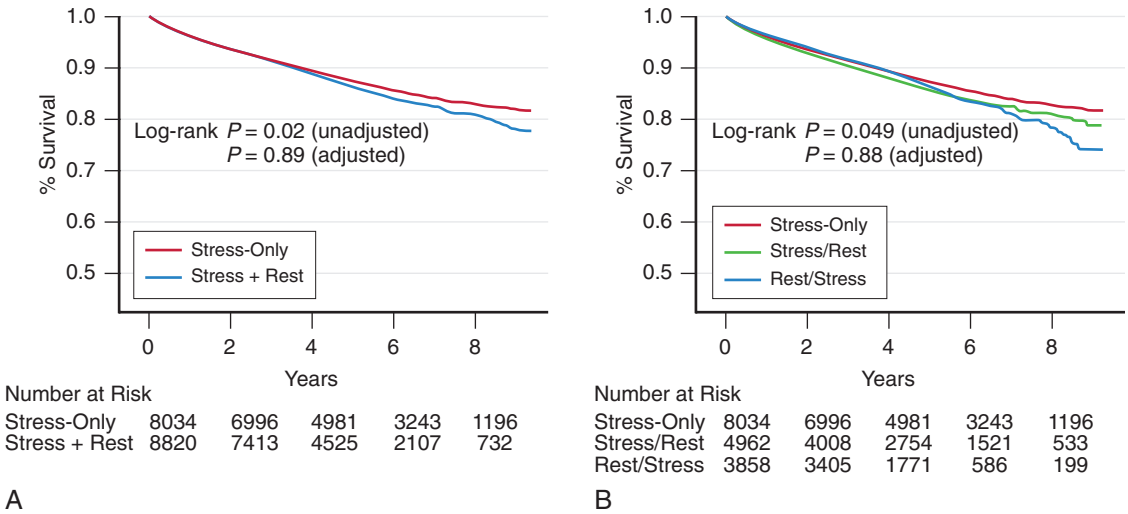

E-FIGURE 56-2. Survival curves in patients with normal gated single-photon emission computed tomography stress-only images compared with patients who underwent both stress and rest imaging. Survival curves are unadjusted or adjusted for baseline clinical characteristics (age, sex, body mass index, history of coronary artery disease, smoking, hyperlipidemia, hypertension, diabetes, chest pain symptoms, and stress electrocardiogram results). After adjustment for baseline clinical characteristics, the two imaging protocols were equally predictive. (Reproduced from Oddstig J, Hedeer F, Jögi J, et al. Reduced administered activity, reduced acquisition time, and preserved image quality for the new CZT camera. *J Nucl Cardiol.* 2013;20:38-44.)

Imaging of Ventricular Function

Global and segmental left and right ventricular function can be evaluated accurately by gated cardiac blood pool imaging to provide a radionuclide angiogram or ventriculogram. The equilibrium radionuclide angiographic approach is performed after thorough mixing of ^{99m}Tc-labeled blood cells within the intravascular compartment. Because ^{99m}Tc remains within the blood pool, serial imaging studies can be acquired during several hours. Acquisition of the images is synchronized with the QRS complex on the ECG through a multigated approach by which each cardiac cycle is divided into multiple frames. A uniform diminution of LV systolic function without segmental wall motion abnormalities suggests nonischemic dilated cardiomyopathy (Chapter 60), whereas depressed global LV function associated with segmental wall motion abnormalities suggests ischemic heart disease. LV function can also be assessed by gated SPECT myocardial perfusion imaging, and the LV ejection fraction and extent of wall motion abnormalities measured by this approach add prognostic information for risk stratification compared with perfusion alone.

CARDIAC COMPUTED TOMOGRAPHY

For CT, imaging with high spatial and temporal resolution and ECG gating during a breath-hold yields snapshots of the heart reconstructed from the same phase of the cardiac cycle. Coronary CT angiography can capture a three-dimensional image of the heart in one to two heartbeats on the latest generation scanners as well as provide coronary artery calcium scoring without the use of contrast and with little radiation.[5] Coronary CT angiography requires 60 to 100 mL of iodinated contrast and average radiation doses of less than 5 mSv. With latest generation scanners, radiation doses of less than 1 mSv are achievable. However, without careful planning of the imaging approach, radiation doses on the order of 5 to 20 mSv are typical on older 64-detector scanners. β-Blockade is often used to achieve a heart rate of 60 beats per minute or less to optimize imaging, and irregular rhythms such as atrial fibrillation may diminish image quality.

Coronary Artery Calcium Scoring

Coronary calcium is an indicator of the burden of atherosclerotic plaque, although there is no correlation of the amount of local coronary calcium with the physiologic or anatomic significance of an underlying coronary stenosis. Calcium scores are generally calculated as an Agatston score, which corresponds to each coronary lesion's calcium area multiplied by the maximal CT attenuation value of that lesion, and then summed for the entire coronary tree. Very high scores confer an increased risk for future cardiac events[6] (Fig. 56-10). Calcium scores are age, gender, and race dependent, and they must be normalized by these factors. Coronary calcium scores predict CHD events independently of standard risk factors, C-reactive protein levels, or the Framingham risk score. Calcium scores above 300 are especially associated with an increased risk for myocardial infarction and cardiac death. The utility of calcium scoring is highest in patients who are at intermediate risk for CHD, based on Framingham risk data (Chapter 52), and increasing scores over time portend a higher risk for a CHD event.[7] By comparison, a calcium score in otherwise low- or high-risk patients will rarely change management, although very high calcium scores may sometime encourage cardiac stress testing.

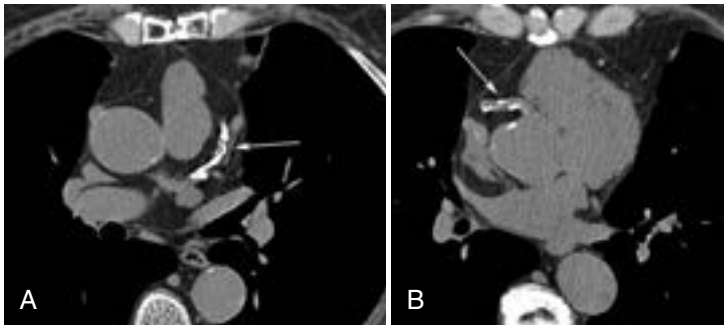

FIGURE 56-10. Non-contrast-enhanced computed tomography axial slices through the heart at two locations for coronary calcium scoring as risk assessment in an asymptomatic patient. **A,** The slice includes the left anterior descending artery with extensive calcification in its proximal portion *(arrow)*. **B,** The slice includes the right coronary artery with proximal spotty calcification *(arrow)*. This patient's calcium score was 457, putting him in a higher risk group regardless of his Framingham risk score.

Coronary Computed Tomographic Angiography

Coronary CT angiography is an excellent technique to diagnose anomalous coronary arterial anatomy in adults (Fig. 56-11). For detection of coronary artery disease, positive-predictive values are in the range of 64 to 91%, and negative-predictive values approach 99%. Thus, the technique is an excellent way to exclude (Fig. 56-12) significant coronary artery disease in the three major coronary vessels (Fig. 56-13). A limitation of the technique, however, is its lower specificity in heavily calcified vessels, which are more common in elderly patients. CT angiography tends to overestimate the percentage of stenosis compared with intravascular ultrasound. The accuracy for detecting stenoses in bypass grafts is quite high, although evaluation of native vessel coronary artery disease is limited in these patients owing to extensive calcification and smaller size vessels. Imaging within most coronary stents has proved difficult.

Coronary CT angiography is not recommended as a routine screening test, but it is useful in selected situations, such as in low- or low-intermediate-risk patients who present to the emergency department with chest pain but without ECG changes or elevations of cardiac biomarkers (Chapters 51 and 72). The high negative-predictive value of CT angiography often can exclude important CHD and avoid the need for other testing in this patient group, and randomized trials show that CT angiography in these patients allows earlier discharge from the emergency department without any increased risk.

CT angiography also may be useful in patients with equivocal or nondiagnostic stress testing or new-onset heart failure. The extent and severity of coronary disease found at CT angiography correlate with subsequent all-cause mortality in a similar fashion to catheter-based coronary angiography. In asymptomatic patients, however, CT angiography does not improve risk stratification over calcium scoring.

Other Cardiac Applications

The same data acquired by coronary CT angiography can be reformatted and used for functional cardiac imaging, including measurement of LV volumes, ejection fraction, wall thickness, and global and segmental wall motion. In acute and chronic myocardial infarction, contrast-enhanced CT can demonstrate late enhancement in a manner similar to cardiovascular magnetic resonance imaging (CMR), albeit with a significantly lower signal and contrast-to-noise ratio. CT angiography can complement echocardiography to evaluate cardiac anatomy in patients with congenital heart disease, especially in patients with contraindications to CMR. CT can evaluate pericardial thickness and calcification in patients with suspected constrictive pericarditis (see Fig. 77-10) and can evaluate native and prosthetic valvular

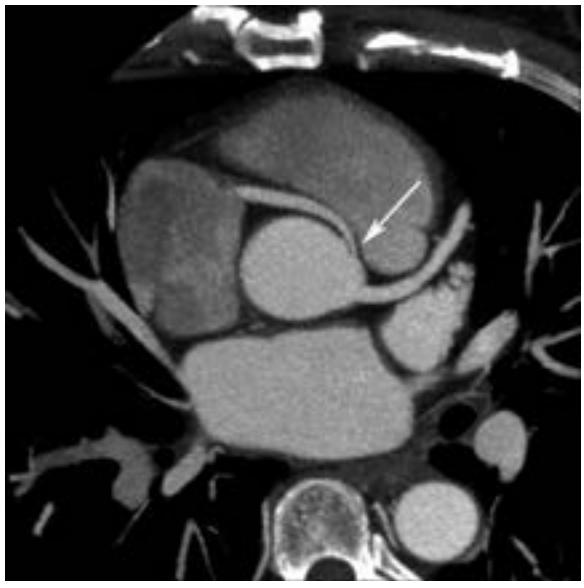

FIGURE 56-11. Contrast-enhanced computed tomographic angiogram in a young patient with chest pain and an anomalous right coronary artery (RCA). The RCA originates with a slitlike origin from the left coronary cusp *(arrow)* and passes anteriorly between the aorta and right ventricular outflow tract. The left main coronary artery originates normally from the left cusp.

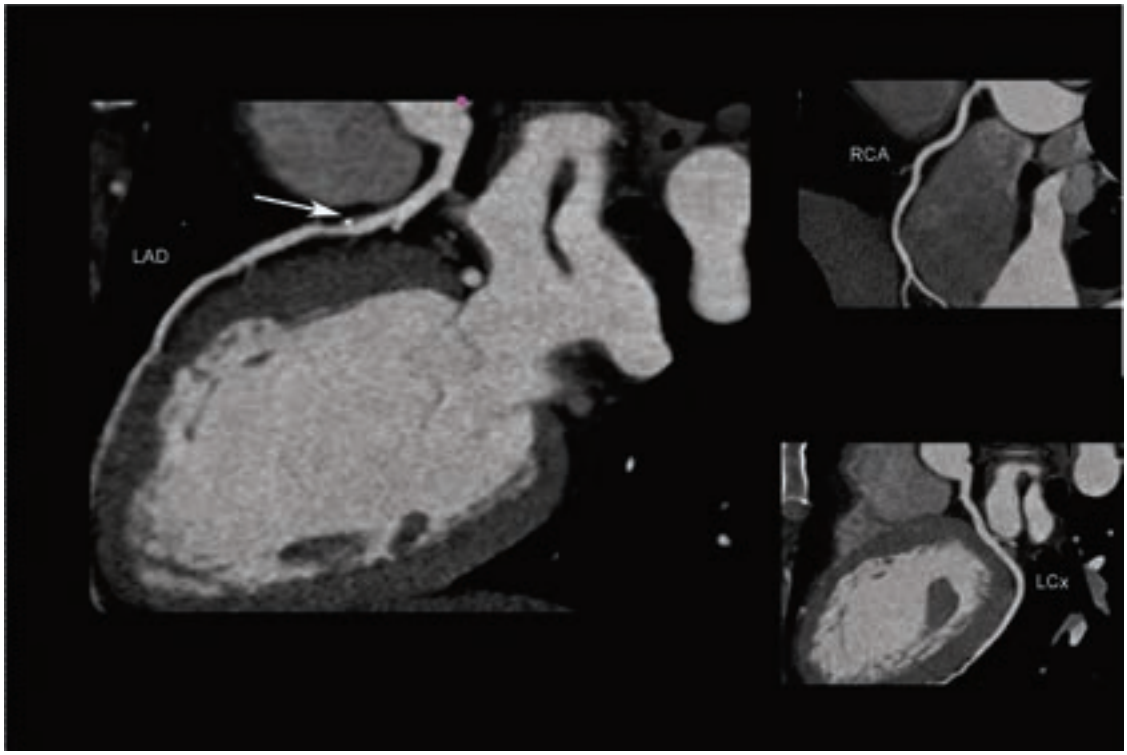

FIGURE 56-12. Contrast-enhanced computed tomographic coronary angiogram obtained on a dual-source 64-detector scanner in a patient with atypical chest pain. The left anterior descending (LAD) artery has a nonobstructive lesion *(arrow)* containing both noncalcified (soft) plaque, which appears dark, and a focal area of calcification. The right coronary artery (RCA) and left circumflex coronary artery (LCx) are normal.

structures (Fig. 56-14) and cardiac masses when imaging with other modalities is inadequate.

CT angiography is often used to image the left atrium and pulmonary venous anatomy for preprocedural planning for pulmonary vein ablation for atrial fibrillation (Chapter 66) or for post-procedural assessment of the possible complication of pulmonary vein stenosis (Fig. 56-15). Cardiac venous anatomy may be imaged to aid in the implantation of LV pacemakers in the cardiac venous system for biventricular pacing for heart failure (Chapters 59 and 66).

CARDIOVASCULAR MAGNETIC RESONANCE IMAGING

Indications, Contraindications, and Pulse Sequences

CMR is a versatile and flexible imaging modality that can be applied in diverse cardiovascular conditions,[8] especially using newer 1.5- or 3-Tesla (T) scanners. Advantages of CMR include the lack of ionizing radiation, the variety of tissues that can be characterized, and the ability to image the heart in any arbitrary plane. Images typically are obtained using ECG gating and breath-hold techniques.

In addition to general restrictions regarding magnetic resonance (e.g., certain intracranial aneurysm clips, transcutaneous electrical nerve stimulation units, intra-auricular implants), patients with cardiac pacemakers and implantable cardioverter-defibrillators generally should *not* undergo CMR because of safety concerns. Newer pacemaker systems under development may be compatible with magnetic resonance imaging. In addition, some non-pacemaker-dependent patients with newer pacemaker and defibrillator models have been scanned safely under controlled conditions, with close monitoring and then testing and reprogramming of the device after the procedure. CMR is safe for all prosthetic heart valves, although image distortions immediately around the prosthesis may obscure nearby pathology. CMR is safe for patients with intracoronary stents. Gadolinium-based contrast agents are contraindicated in patients with a glomerular filtration rate of less than 30 mL/minute/1.83 m^2, owing to their association with nephrogenic systemic fibrosis (Chapter 267).

A comprehensive CMR study includes evaluation of cardiac structure, function, tissue characteristics, perfusion, and scarring or fibrosis. CMR is highly accurate for the noninvasive quantitative assessment of LV and RV volumes and ejection fraction. The components of the examination are tailored to the particular diagnostic question at hand.

Specific Clinical Applications

CORONARY ARTERY DISEASE

For detection of myocardial ischemia, first-pass gadolinium perfusion imaging during vasodilator stress with adenosine or regadenoson shows defects, generally in the subendocardium, that persist for at least five heartbeats during the first pass of contrast (Fig. 56-16). Imaging is often repeated at rest after approximately 10 minutes to be sure any defect seen with stress was not either artifactual or caused by an infarct (the latter is excluded in combination with late gadolinium enhancement). Head-to-head comparisons of vasodilator stress perfusion CMR and dobutamine stress CMR suggest a higher sensitivity for contrast-enhanced perfusion imaging and higher specificity for dobutamine wall motion imaging. CMR stress testing is more accurate than SPECT[9] and is a strong predictor of cardiac events and survival. Coronary CT angiography has superior spatial resolution and accuracy compared with CMR for imaging of the coronary arteries, but CMR coronary imaging is useful for the diagnosis of anomalous coronary arteries.

CMR with late gadolinium enhancement is the gold standard technique for assessment of myocardial scar caused by myocardial infarction, with a better accuracy than nuclear imaging approaches, especially for smaller non–Q wave infarctions (Fig. 56-17). In acute myocardial infarction, CMR T2-weighted techniques can assess the myocardium at risk and estimate the amount of salvaged myocardium owing to reperfusion. Areas of low signal in the subendocardial core of the infarction represent regions of microvascular obstruction with severe capillary destruction and are a marker of subsequent adverse LV remodeling and poorer outcome. To assess myocardial viability in patients with chronic CHD, CMR with late gadolinium enhancement has the best sensitivity for recovery of function with revascularization, but low-dose dobutamine contractile reserve has better specificity.

CARDIOMYOPATHIES

CMR is often used to identify the underlying etiology of cardiomyopathies (Chapter 60). In patients who present in acute heart failure or with chest pain, elevated troponin levels, but a negative coronary arteriogram, CMR is ideally suited to identify myocarditis (Chapter 60) (Fig. 56-18). In patients

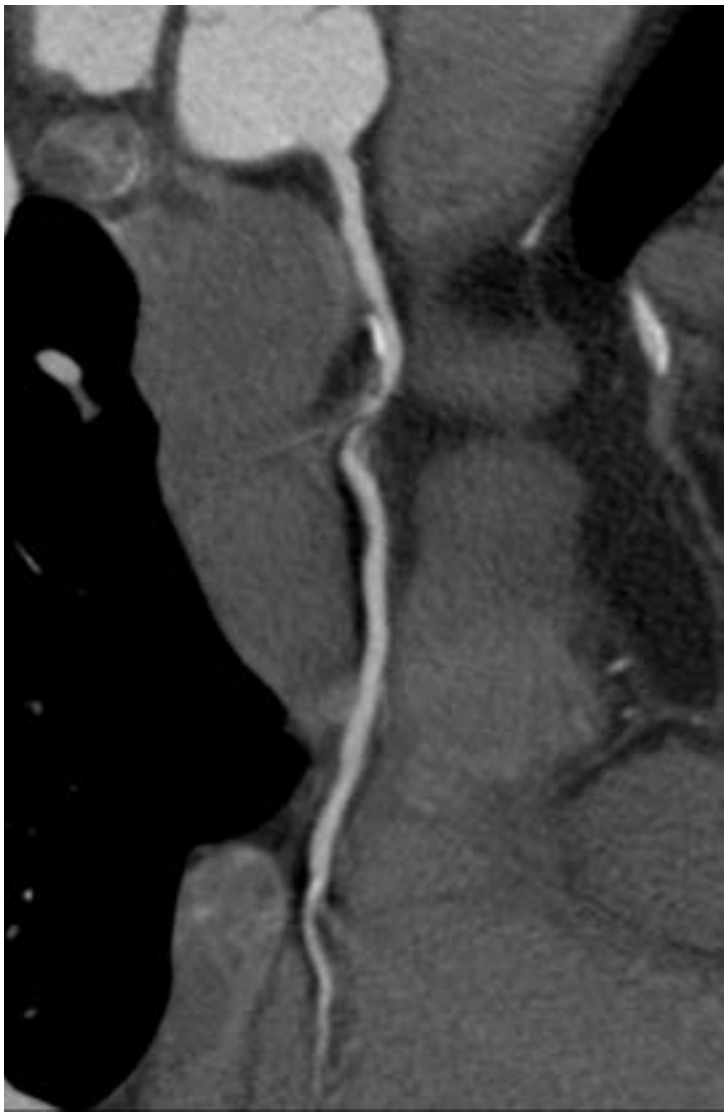

FIGURE 56-13. Contrast-enhanced computed tomographic coronary angiogram of the left anterior descending artery (LAD) in a 54-year-old man who presented to an emergency department with risk factors but atypical chest pain. The mid LAD demonstrates mixed obstructive plaque with calcified and noncalcified components.

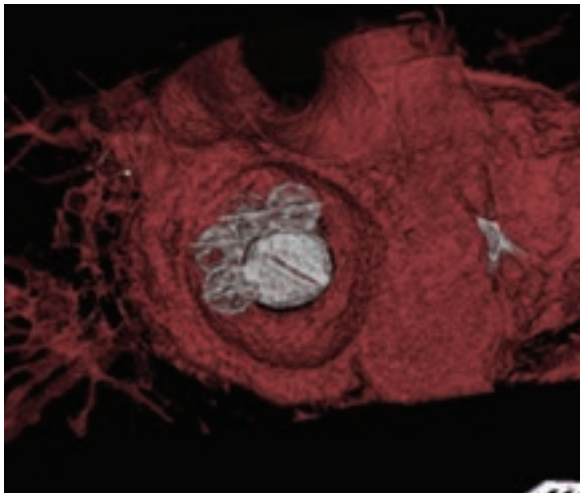

FIGURE 56-14. Three-dimensional reconstruction of a computed tomographic angiogram in a patient with a St. Jude's mechanical mitral valve with a paravalvular leak that has been closed with four Amplatzer devices.

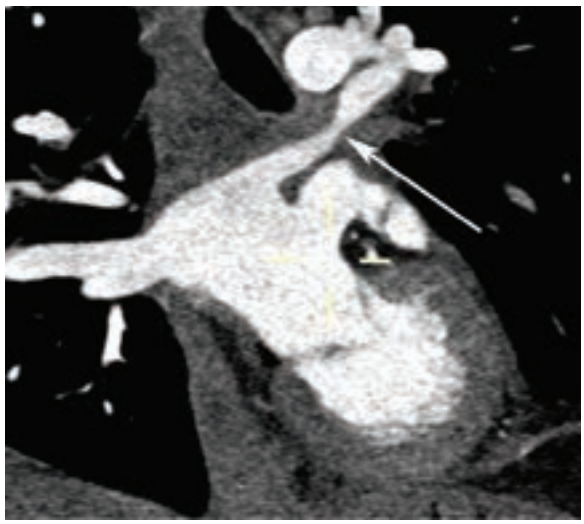

FIGURE 56-15. Contrast-enhanced computed tomographic coronary angiogram in a patient after pulmonary vein ablation for atrial fibrillation demonstrating stenosis of the left upper pulmonary vein toward its origin *(arrow)* relative to the more distal vessel.

with hypertrophic cardiomyopathy (Chapter 60), CMR is more sensitive than echocardiography for identifying increased regional wall thickness and can demonstrate late gadolinium enhancement (Fig. 56-19). Cardiac amyloidosis (Chapters 60 and 188) can be seen as diffuse subendocardial or patchy enhancement, or simply difficulty in nulling normal myocardium. A new noncontrast method entitled native T1 mapping can identify amyloidosis and other cardiomyopathies by measuring elevated T1 values in the myocardium. Patchy fibrosis is readily identified in cardiac sarcoidosis (Chapters 60 and 95) (Fig. 56-20) and is more sensitive than endomyocardial biopsy. The finding of fibrosis by late gadolinium enhancement in almost any form of heart disease is associated with adverse prognosis compared with those without fibrosis.[10]

CMR is often used in the diagnosis of arrhythmogenic right ventricular cardiomyopathy (Fig. 56-21; Chapters 60 and 65), which is characterized by global RV dilation and regional RV akinesis or dyskinesis. Late gadolinium enhancement sometimes may be seen but can be difficult to identify in the thin-walled right ventricle. Fat is a nonspecific finding. In iron overload conditions such as thalassemia (Chapter 162), multi-echo T1-weighted imaging of T2* can identify the extent of iron overload and can be used to follow effects of chelation therapy. Rarer causes of cardiomyopathy, such as ventricular noncompaction, Chagas disease (Chapters 60 and 347), and Takotsubo cardiomyopathy (Chapter 60), also have characteristic CMR findings.

AORTIC DISEASE, PERICARDIAL DISEASE, AND MASSES

CMR is an excellent test to image aneurysms but is a second-line test to detect acute aortic dissection (see Fig. 78-6) or intraluminal aortic hematoma (see Fig. 78-7) in stable patients. CMR is also an excellent test for the evaluation of chronic pericardial disease (Fig. 56-22) because it accurately identifies pericardial thickness as well as adherence of the pericardium to the epicardium in constrictive pericarditis. Real-time imaging can demonstrate ventricular interdependence, a hallmark of this disease. CMR is also an ideal tool to diagnose intracardiac (Fig. 56-23) and extracardiac masses such as myxomas, thrombus, and tumors (Chapter 60), owing to its high spatial resolution and ability to perform tissue characterization.

CONGENITAL HEART DISEASE

CMR is useful for the assessment of both simple and complex congenital heart disease and is often used as an adjunct to echocardiography (Chapter 69). For example, phase velocity CMR readily quantifies blood flow through the major blood vessels, thereby facilitating accurate assessment of the ratio of pulmonary to systemic blood flow in atrial or ventricular septal defects. CMR is particularly valuable for assessing abnormalities of the great vessels, such as aortic coarctation (Fig. 56-24), extracardiac anatomy, or anomalous pulmonary venous drainage, and in patients with complex congenital heart disease who have undergone prior corrective or palliative shunt surgery, such as in tetralogy of Fallot or hypoplastic left heart syndrome. CMR is uniquely

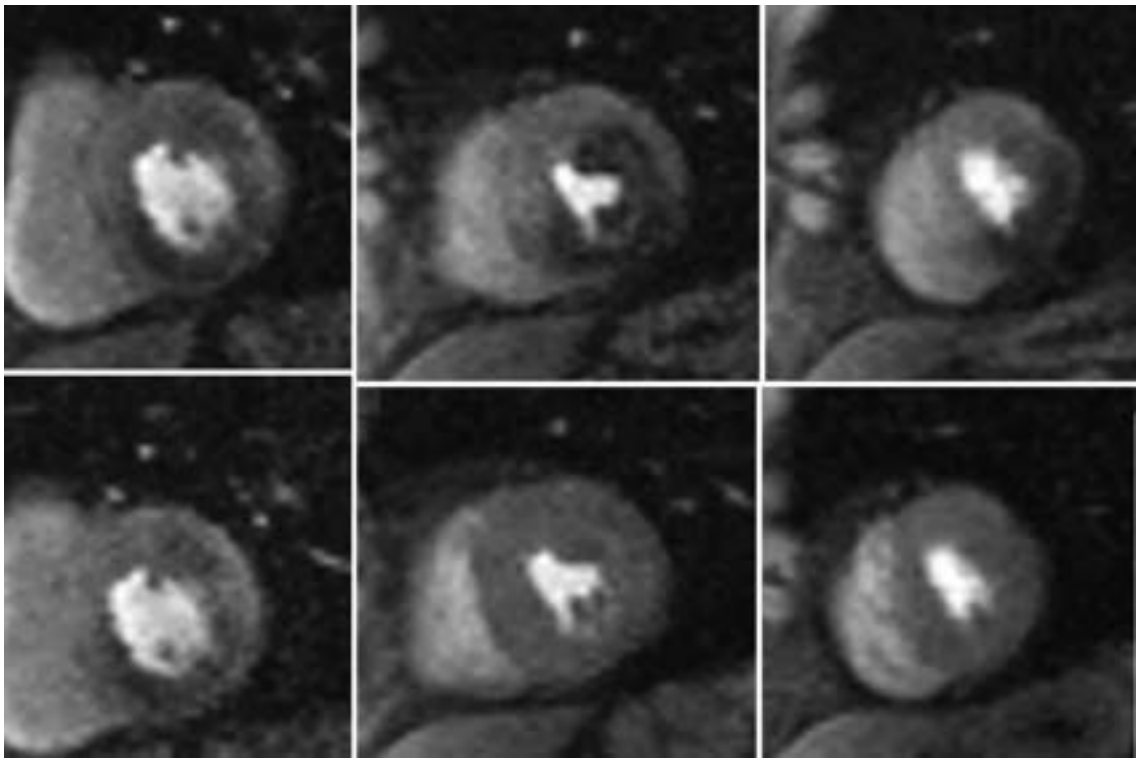

FIGURE 56-16. Set of first-pass gadolinium-enhanced magnetic resonance perfusion images during adenosine stress *(top row)* and at rest *(bottom row)* at the base *(left),* mid ventricle *(center),* and apex *(right).* The stress images demonstrate a large perfusion deficit in the anterolateral and inferolateral walls (from 1 o'clock to 7 o'clock), especially in the mid ventricle and apex, whereas the same regions appear normal at rest.

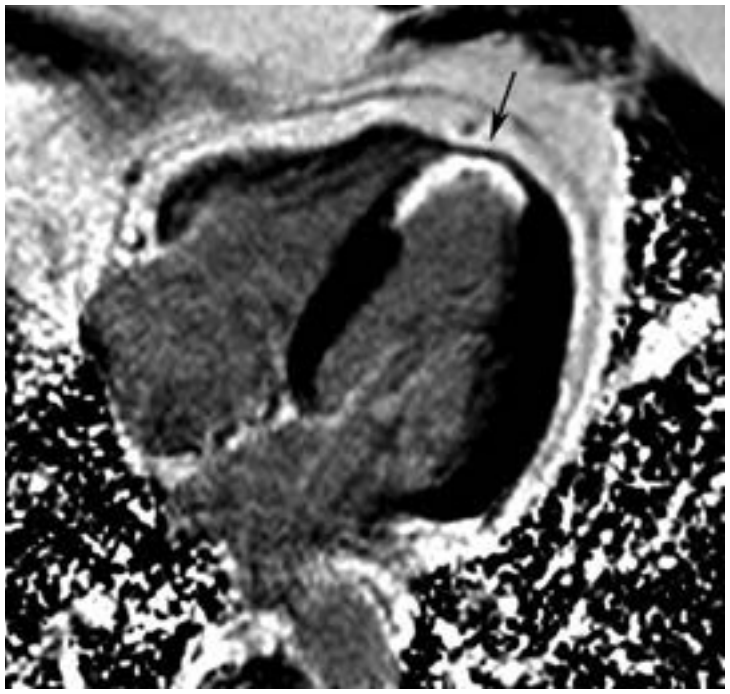

FIGURE 56-17. Late gadolinium-enhanced four-chamber long axis image in a patient with a scar *(arrow)* from a prior anterior myocardial infarction.

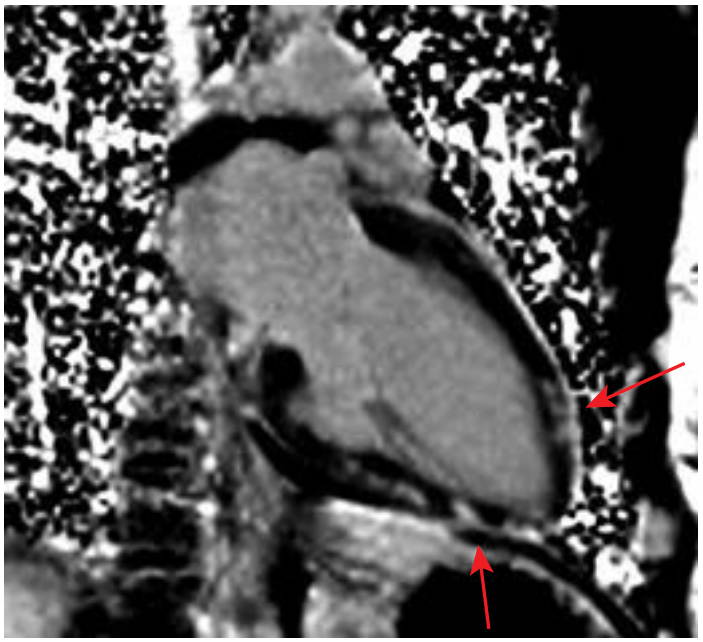

FIGURE 56-18. Gadolinium-enhanced image in a 22-year-old man shows patchy subepicardial enhancement *(arrows),* characteristic of acute myocarditis.

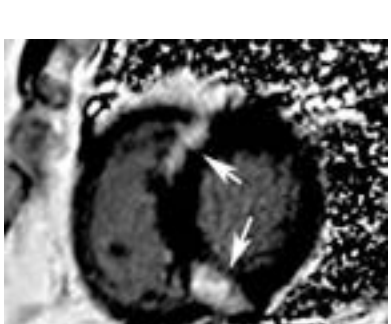

FIGURE 56-19. Late gadolinium-enhanced magnetic resonance short axis image of a 35-year-old man with hypertrophic cardiomyopathy demonstrates a classic pattern of enhancement, which identifies fibrosis *(arrows),* in the right ventricular insertion sites.

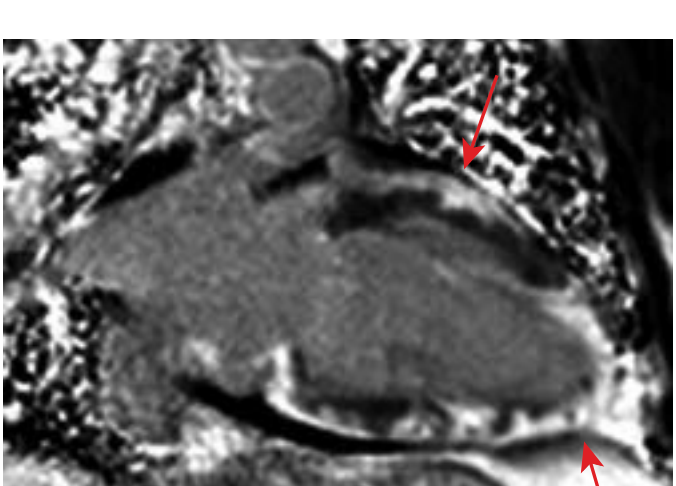

FIGURE 56-20. Late gadolinium-enhanced magnetic resonance image in a 47-year-old man with heart failure, heart block, and hilar lymphadenopathy shows patchy late gadolinium enhancement in a noncoronary distribution, including subepicardial anterior wall enhancement *(upper arrow)* and near transmural apical enhancement *(lower arrow)* consistent with myocardial sarcoidosis.

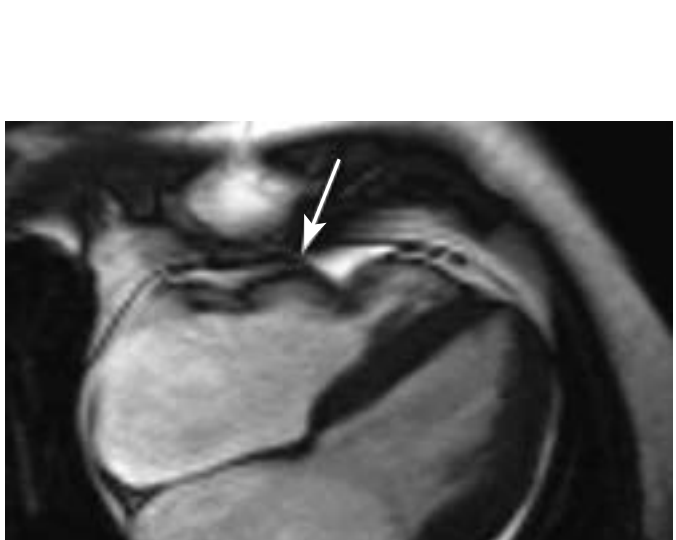

FIGURE 56-21. Magnetic resonance image in a 27-year-old woman with arrhythmogenic right ventricular cardiomyopathy demonstrates regional right ventricular (RV) systolic dysfunction, which is the hallmark of the disease. The *arrow* points to a region of RV dyskinesis at end systole, and the RV appears like an accordion at end systole.

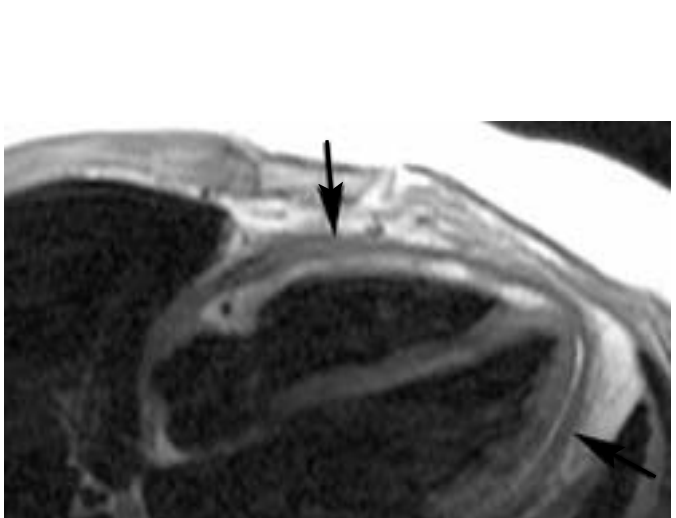

FIGURE 56-22. Magnetic resonance image of a 35-year-old man with dyspnea many years after mantle radiation for Hodgkin lymphoma shows a thickened pericardium *(arrows)* circumferentially around the left and right ventricles.

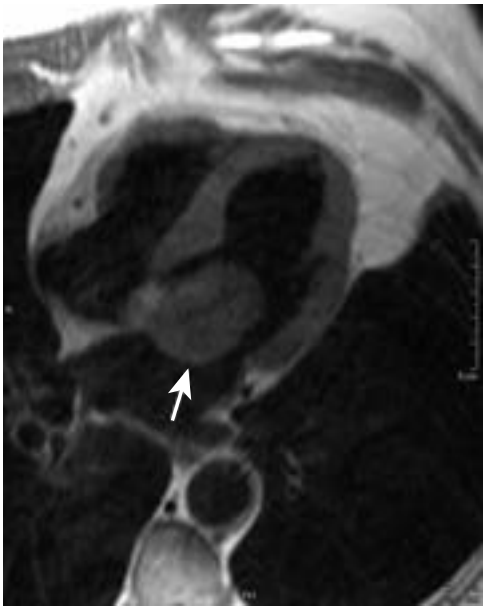

FIGURE 56-23. Diastolic magnetic resonance image of a large left atrial myxoma *(arrow)* showing that it is attached to the atrial septum and is prolapsing across the mitral valve.

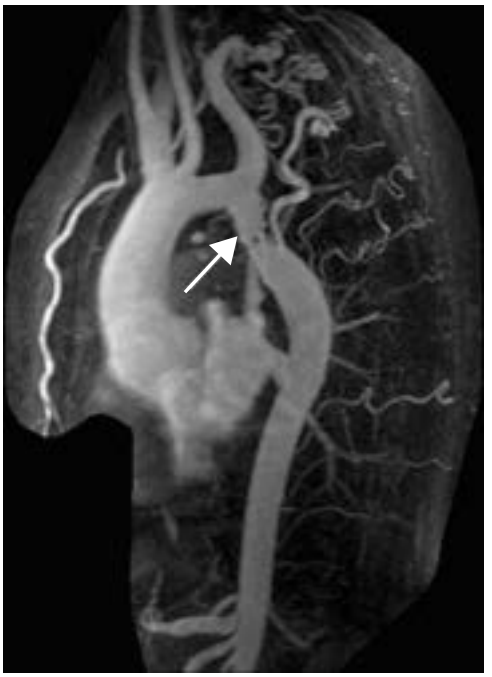

FIGURE 56-24. Three-dimensional contrast-enhanced magnetic resonance angiogram in a patient with an aortic coarctation *(arrow)*.

able to measure right ventricular volumes accurately, an often important determination in this setting.

Grade A References

A1. Litt HI, Gatsonis C, Snyder B, et al. CT angiography for safe discharge of patients with possible acute coronary syndromes. *N Engl J Med*. 2012;366:1393-1403.

A2. Hoffmann U, Truong QA, Schoenfeld DA, et al. Coronary CT angiography versus standard evaluation in acute chest pain. *N Engl J Med*. 2012;367:299-308.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hendel RC, Berman DS, Di Carli MF, et al. ACCF/ASNC/ACR/AHA/ASE/SCCT/SCMR/SNM 2009 appropriate use criteria for cardiac radionuclide imaging: a report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, the American Society of Nuclear Cardiology, the American College of Radiology, the American Heart Association, the American Society of Echocardiography, the Society of Cardiovascular Computed Tomography, the Society for Cardiovascular Magnetic Resonance, and the Society of Nuclear Medicine. *Circulation*. 2009;119:e561-e587.
2. Parker MW, Iskandar A, Limone B, et al. Diagnostic accuracy of cardiac positron emission tomography versus single photon emission computed tomography for coronary artery disease: a bivariate meta-analysis. *Circ Cardiovasc Imaging*. 2012;5:700-707.
3. Shaw LJ, Mieres JH, Hendel RH, et al. Comparative effectiveness of exercise electrocardiography with or without myocardial perfusion single photon emission computed tomography in women with suspected coronary artery disease: results from the What Is the Optimal Method for Ischemia Evaluation in Women (WOMEN) trial. *Circulation*. 2011;124:1239-1249.
4. Ling LF, Marwick TH, Flores DR, et al. Identification of therapeutic benefit from revascularization in patients with left ventricular systolic dysfunction: inducible ischemia versus hibernating myocardium. *Circ Cardiovasc Imaging*. 2013;6:363-372.
5. De Cecco CN, Meinel FG, Chiaramida SA, et al. Coronary artery computed tomography scanning. *Circulation*. 2014;129:1341-1345.
6. Madhavan MV, Tarigopula M, Mintz GS, et al. Coronary artery calcification: pathogenesis and prognostic implications. *J Am Coll Cardiol*. 2014;63:1703-1714.
7. Budoff MJ, Young R, Lopez VA, et al. Progression of coronary calcium and incident coronary heart disease events: MESA (Multi-Ethnic Study of Atherosclerosis). *J Am Coll Cardiol*. 2013;61:1231-1239.
8. Hundley WG, Bluemke DA, Finn JP, et al. ACCF/ACR/AHA/NASCI/SCMR 2010 expert consensus document on cardiovascular magnetic resonance: a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents. *J Am Coll Cardiol*. 2010;55:2614-2662.
9. Greenwood JP, Maredia N, Younger JF, et al. Cardiovascular magnetic resonance and single-photon emission computed tomography for diagnosis of coronary heart disease (CE-MARC): a prospective trial. *Lancet*. 2012;379:453-460.
10. Kuruvilla S, Adenaw N, Katwal AB, et al. Late gadolinium enhancement on cardiac magnetic resonance predicts adverse cardiovascular outcomes in nonischemic cardiomyopathy: a systematic review and meta-analysis. *Circ Cardiovasc Imaging*. 2014;7:250-258.

REVIEW QUESTIONS

1. Which statement regarding the evaluation of the heart on chest radiography is correct?

A. The main value of a chest radiograph is the ability to diagnose congenital cardiac disorders owing to a large number of disease-specific imaging signs.
B. Since the advent of echocardiography, the chest radiograph no longer plays a role in the evaluation of the patient with cardiac disease.
C. The chest radiograph is useful to monitor the progression of cardiac disease and its response to treatment.
D. The chest radiograph is an accurate imaging modality to assess the exact position of an implanted device.
E. All of the above

Answer: C Although echocardiography and the other cross-sectional imaging modalities have replaced chest radiography for the diagnosis of most cardiac diseases, the value of chest radiography is its ability to monitor the progression of cardiac disease and its response to treatment rapidly and inexpensively. The diagnosis of congenital cardiac disorders is made with echocardiography, cardiac catheterization, and cardiac magnetic resonance imaging and computed tomography. Despite the ready availability of echocardiography, the chest radiograph is valuable for rapid evaluation of changes in heart size and pulmonary vascularity, as well as signs of heart failure. The radiograph shows the position of a device on a two-dimensional image, but assessment of the exact position may require cross-sectional imaging.

2. Which of the following groups of patients are good candidates for coronary artery calcium scoring?

A. after coronary artery bypass surgery
B. after coronary stenting
C. High risk Framingham score
D. Intermediate risk Framingham score
E. Low risk Framingham score

Answer: D In patients at intermediate risk, knowing the coronary artery calcium score might change management in terms of aggressive lipid lowering. High-risk patients should already be receiving aggressive therapy, and no additional therapy is needed in low-risk patients.

3. Coronary computed tomographic angiography (CTA) is useful for the following indication:

A. Evaluating native vessels in post-CABG patients
B. Evaluating in-stent restenosis
C. Screening asymptomatic patients with strong family histories of CHD
D. Evaluating low-to-intermediate risk patients in the emergency department
E. All of the above

Answer: D Recent randomized trials have demonstrated the utility of CTA compared with standard evaluation in the emergency department. Native vessels are difficult to image post-CABG because they tend to shrink and calcify. Imaging within stents remains a challenge. Coronary CTA is not appropriate as a screening test in asymptomatic individuals.

4. A 62-year-old white male presents with substernal chest discomfort that is "aching" in nature and occurs with mild exertion but also occasionally at rest. It seems to radiate to the throat and is relieved in a few minutes with rest. He has also noted some recent shortness of breath when walking up hills. He has a history of gastroesophageal reflux disease and a hiatal hernia. He has no history of hypertension, diabetes, or smoking. His BMI is 35, and his LDL cholesterol is 162 mg/dL. His father died suddenly of an unknown cause when he was 60 years old. He has been on a Mediterranean diet for the past 2 months. On physical examination, his blood pressure is 130/80 and the remainder of his exam is unremarkable. The patient is referred by his physician for an exercise SPECT myocardial perfusion scan. He exercises to a workload of 6 METS and a peak heart rate representing 77% of his age-predicted maximum. He stopped because of dyspnea and some slight chest discomfort. His exercise ECG showed some ST segment flattening, but not depressed more than 0.5 mm. These changes resolved in recovery. His myocardial perfusion scan showed no perfusion defects. The left ventricular cavity size appeared larger on the stress than on the rest images. There was diffuse hypokinesis, and the left ventricular ejection fraction was 42%. Which one of the following statements is INCORRECT regarding this patient?

A. His blood pressure does not need to be reduced further
B. He most likely has a nonischemic cardiomyopathy
C. He most likely has multivessel coronary artery disease
D. His diet has been shown to reduce the risk for developing coronary artery disease
E. According to guidelines, he should be started on statin therapy

Answer: B He does not have a nonischemic cardiomyopathy. His SPECT scan shows transient ischemic cavity dilation of the left ventricle in association with diffuse left ventricular hypokinesis. The reason he has no perfusion defects is because of balanced ischemia in the supply regions of all 3 major coronary vessels. No focal defects will be observed if the reduction in perfusion is rather homogeneous in the myocardium. He has an intermediate-high pretest likelihood of coronary disease by clinical criteria. His blood pressure is adequately controlled. Diets such as the Mediterranean diet have been shown to reduce the risk of coronary disease. Due to the high likelihood of coronary disease, he should be started on statin therapy and is a good candidate for coronary arteriography because of likely multivessel disease with a reduced ejection fraction.

57

CATHETERIZATION AND ANGIOGRAPHY

MORTON KERN

Cardiac catheterization is insertion and passage of small plastic tubes (catheters) into arteries and veins to the heart to obtain radiographic images of coronary arteries and cardiac chambers (angiography and ventriculography) and to measure pressures in the heart (hemodynamics). Coronary angiography defines the site, severity, and morphology of atherosclerotic lesions, and it identifies collateral blood supply beyond occluded vessel segments. Cardiac catheterization is used not only to diagnose coronary artery, valvular (Chapter 75), and myocardial diseases (Chapter 60) but also to perform therapeutic (interventional) procedures to relieve obstructing arterial stenoses (Chapter 74), to open or replace narrowed valves, or to close intracardiac defects (Chapter 69) through catheter-based, minimally invasive percutaneous techniques. These same diagnostic and therapeutic techniques are also used in the peripheral arterial circulation in a modified fashion to address carotid, renal, and peripheral vascular disease (Chapters 79 and 80), aortic aneurysms (Chapter 78), and vascular shunts (Table 57-1).

INDICATIONS FOR AND CONTRAINDICATIONS TO CARDIAC CATHETERIZATION

The indications to perform cardiac catheterization include the need to diagnose atherosclerotic coronary artery disease, abnormalities of cardiac muscle function, valvular abnormalities, and congenital heart disease (Table 57-2).[1] Contraindications to cardiac catheterization are few. Absolute contraindications involve only inadequate facilities or equipment for catheterization. Relative contraindications depend on the urgency of the procedure and conditions.

TECHNIQUE OF CATHETERIZATION

After the procedure and its indications, risks, and benefits are explained to the patient, the patient is placed on the cardiac catheterization table and centered under the C-arm of the radiographic gantry (E-Fig. 57-1). After sterile preparation and draping, local anesthetic is administered over the vascular access site—commonly the femoral artery but increasingly the radial artery, which is associated with better outcomes in patients with ST segment elevation myocardial infarction. The artery is punctured, and a vascular sheath is inserted, through which the angiographic catheter is advanced over a soft spring-tipped 0.035-inch guidewire that permits safe, atraumatic passage of the catheter to the heart. The specially shaped catheters are seated and connected to a manifold to measure pressure and to inject radiographic contrast media.

Coronary arteriography records the images from multiple angles by rotation of the C-arm. The images are displayed and preserved on digital imaging systems.

After coronary angiography, the catheter is exchanged for a ventriculography catheter that is inserted into the left ventricle. After left ventricular (LV) pressure is measured, radiographic contrast medium (approximately 25 to 45 mL) is injected under high pressure (1000 psi) to assess LV wall motion, chamber size, presence of mitral valve regurgitation, and shape of the aortic root. The LV ejection fraction (normal is 50 to 70%), a measure of the heart function, is computed as a percentage of the diastolic volume ejected.

After diagnostic angiography is completed, the need for coronary revascularization (Chapter 74) is assessed. If suitable symptomatic coronary artery obstructions are present, percutaneous coronary intervention (PCI) may be performed at the same time if it was discussed with the patient and consented to in advance. Alternatively, the patient may be referred for later PCI or for coronary artery bypass graft surgery.

At the conclusion of the catheterization procedure, the catheters are removed. For the femoral artery, hemostasis is achieved either by manual compression, which requires the patient to remain stationary in bed for 4 hours, or by use of a vascular closure device while the patient remains in bed for 1 to 2 hours. For the radial approach, the arterial sheath is removed with the simple application of a specialized compression wrist band; the patient can ambulate immediately thereafter.

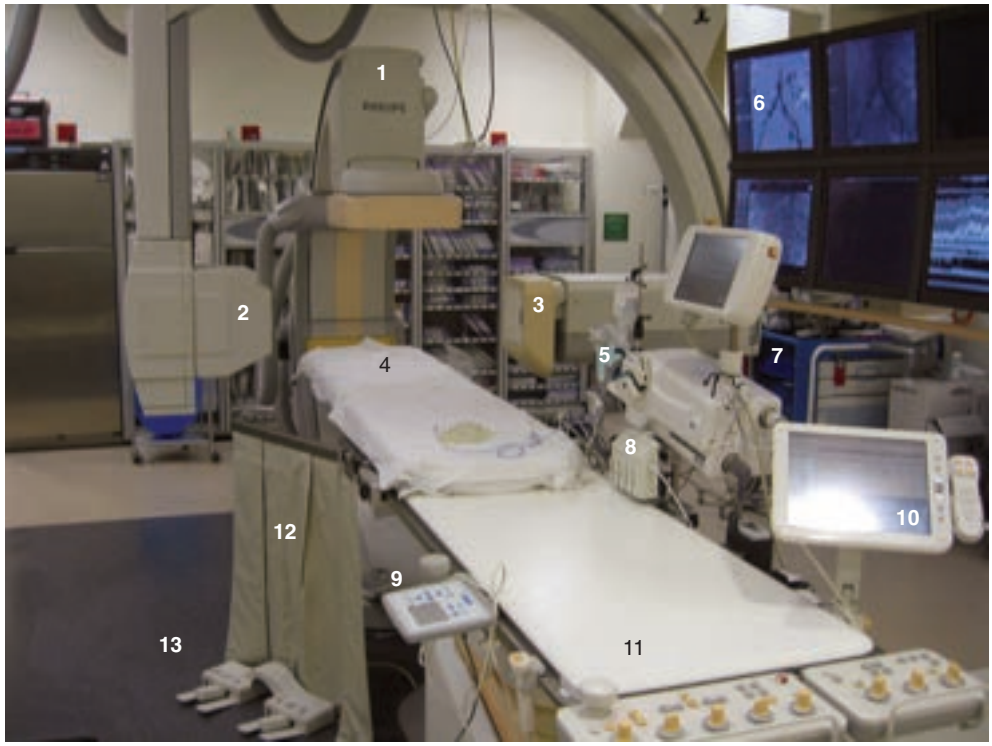

E-FIGURE 57-1. Modern cardiac catheterization laboratory. (1) Anteroposterior imaging gantry. The image intensifier is above the patient and the x-ray tube below. (2) Lateral imaging gantry with the x-ray system to the patient's right and (3) the lateral plane imaging intensifier on the patient's left side. (4) Catheterization laboratory table pad position for patient's head. (5) Power injector for radiographic contrast material. (6) Angiographic and hemodynamic monitor bank. (7) Crash cart. (8) Pressure transducer holders. (9) Monitor displaying intravascular ultrasound image during percutaneous coronary intervention. (10) Control panel for x-ray images. (11) Table controls for movement of x-ray system. (12) Under-table lead shield. (13) Foot pedal controls for fluoroscopy and cineangiography imaging.

TABLE 57-1 PROCEDURES THAT MAY ACCOMPANY CORONARY ANGIOGRAPHY

PROCEDURE	COMMENTS
Central venous access: femoral, internal jugular, subclavian	Uses IV access for emergency medications or fluids, temporary pacemaker; pacemaker not mandatory for most coronary angiography
Hemodynamic assessment, left-sided heart pressures, aorta, and left ventricle	Routine for all studies
Right- and left-sided heart combined pressures	Not routine for coronary artery disease but mandatory for valvular heart disease and routine for heart failure, right ventricular dysfunction, pericardial disease, cardiomyopathy, intracardiac shunts, and congenital abnormalities
Left ventriculography	Routine for all studies; may be excluded with high-risk patients and those with left main coronary or aortic stenosis, severe congestive heart failure, or renal failure
Internal mammary and saphenous vein graft angiography	Routine for coronary bypass conduit
Pharmacologic studies: Assessment of coronary spasm (use of ergonovine or acetylcholine conducted in specialized centers for research only) Use of vasodilators	Conducted routinely for coronary angiography with nitroglycerin (for coronary spasm) Nitric oxide used for pulmonary hypertension
Aortography	Routine for aortic insufficiency, aortic dissection, and aneurysm and may be performed in patients with aortic stenosis; routine to locate bypass graft conduits not visualized by selective angiography
Cardiac pacing electrophysiologic studies	Arrhythmia evaluation
Interventional and special techniques	Percutaneous coronary intervention: includes balloon angioplasty, bare metal or drug-eluting stent, and rotational coronary atherectomy Assessment of fractional flow reserve to assess the functional severity of a coronary stenosis Balloon catheter valvuloplasty; transcatheter aortic valve replacement; myocardial biopsy; atrial septal defect or patent foramen ovale defect closure; transseptal puncture to assess valvular heart disease; electrophysiologic catheter ablation
Vascular closure devices	Routinely available for patients prone to femoral artery access bleeding

Modified from Kern MJ, ed. The Interventional Cardiac Catheterization Handbook. 3rd ed. Philadelphia: Elsevier; 2012.

TABLE 57-2 INDICATIONS FOR AND CONTRAINDICATIONS TO CARDIAC CATHETERIZATION

INDICATIONS
Identification of the extent and severity of coronary artery disease and evaluation of left ventricular function
Assessment of the severity of valvular or myocardial disorders, such as aortic stenosis or insufficiency, mitral stenosis or insufficiency, and various cardiomyopathies, to determine the need for surgical correction
Collection of data to confirm and to complement noninvasive studies
Determination of the presence of coronary artery disease in patients with confusing clinical presentations or chest pain of uncertain origin
ABSOLUTE CONTRAINDICATIONS
Inadequate facilities
Patient refusal
RELATIVE CONTRAINDICATIONS
Severe uncontrolled hypertension
Ventricular arrhythmias
Recent acute stroke
Severe anemia
Active gastrointestinal bleeding
Allergy to radiographic contrast agents
Acute renal failure
Uncompensated congestive failure (patient cannot lie flat)
Unexplained febrile illness or untreated active infection
Electrolyte abnormalities (e.g., hypokalemia)
Severe coagulopathy
Pregnancy
Uncontrolled arrhythmias, hypertension
Uncooperative patient or patient refusal

TABLE 57-3 COMPLICATIONS OF CARDIAC CATHETERIZATION

MAJOR COMPLICATIONS
Death
Cerebrovascular accident
Myocardial infarction, shock
Ventricular tachycardia or fibrillation
RARE BUT SERIOUS COMPLICATIONS
Aortic dissection
Cardiac perforation
Tamponade
Heart failure
Reaction to contrast agents, anaphylaxis
Nephrotoxicity
Arrhythmias, including heart block, asystole, supraventricular tachyarrhythmias
Hemorrhage, including local or retroperitoneal
Infection
Protamine reaction
Vascular complications, including thrombosis, embolus, vascular injury, pseudoaneurysm

COMPLICATIONS OF CARDIAC CATHETERIZATION

For diagnostic cardiac catheterization, risks are less than 0.2% for death, less than 0.5% for myocardial infarction, less than 0.07% for stroke, less than 0.5% for serious arrhythmia, and less than 1% for major vascular complications, including thrombosis and bleeding requiring transfusion or pseudoaneurysm (Table 57-3). Vascular complications occur more frequently with the femoral artery approach than with the radial artery approach; the brachial artery approach, which is used only when neither femoral nor radial access is possible, has the highest rate of vascular complications, and the radial artery has the lowest.

Patients are not routinely anticoagulated for a diagnostic catheterization, and special preparations must be made for patients who are receiving anticoagulants, patients who have diabetes or renal insufficiency, and patients who may have a potential allergy to radiographic contrast media. For the anticoagulated patient, provisions to withhold warfarin or heparin must be made to reduce the potential for femoral puncture site bleeding complications (e.g., retroperitoneal hematoma or pseudoaneurysm). For example, a patient taking warfarin after an aortic valve replacement would have the warfarin withheld for about 3 days before the catheterization, the international normalized ratio would be monitored, and the patient might be administered bridging heparin to the time of the procedure; warfarin would be restarted after the procedure.

For elective catheterizations in patients with insulin-dependent diabetes, half of the usual morning dose of insulin generally is given the morning of the procedure to provide reasonable diabetic coverage and to avoid hypoglycemia. Metformin should be withheld before the study. Medications for hypertension and other medical conditions are continued up to and including the morning of the procedure.

Contrast-induced nephropathy, which generally becomes clinically apparent 2 to 3 days after the catheterization, is uncommon. Patients with diabetes or renal insufficiency and patients who are dehydrated from any cause are at a three- to five-fold increased risk for contrast-induced renal failure. For

patients with renal insufficiency and those at high risk for contrast-induced renal failure because of diabetes or dehydration, hydration with sodium chloride is recommended to increase urine output.[1] Treatment with *N*-acetylcysteine (Mucomyst) is no longer recommended to prevent contrast nephropathy.[1] Recent data suggest a benefit from oral rosuvastatin, either 40 mg on admission followed by 20 mg per day for patients with an acute coronary syndrome[1] or 10 mg for two days before and three days afterwards for patients with diabetes and chronic kidney disease.[1]

Contrast media reactions are rare, with an overall incidence of 5% or less, but potentially serious. Adverse reactions occur in 10 to 12% of patients with a history of allergy and in 15% of patients with reported reaction on a previous contrast radiographic examination. There are three types of allergies to contrast media: cutaneous and mucosal manifestations, smooth muscle and minor anaphylactoid responses, and cardiovascular and major anaphylactoid responses involving laryngeal or pulmonary edema. Pretreatment with corticosteroids is helpful in reducing all types of reactions except urticaria. Patients reporting previous allergic reactions to contrast media should be premedicated with prednisone (60 mg orally the evening before and morning of the procedure) and diphenhydramine (25 to 50 mg orally the morning of the procedure). Patients with known prior anaphylactoid reactions should be pretreated with steroids in the same dose. Routine treatment with a histamine$_2$-receptor blocker (e.g., cimetidine) does not appear to have any benefit.

Hypotension during and after cardiac catheterization may occur from a vasovagal response, occult retroperitoneal bleeding, myocardial ischemia or infarction, or cardiac tamponade. Vasovagal hypotension is treated with volume and atropine (0.5 to 1.0 mg intravenously). Hypotension with back pain suggests retroperitoneal hematoma. Hypotension due to cardiac tamponade, which may occur during or after PCI, requires rapid diagnosis, reversal of anticoagulation, and urgent pericardiocentesis (Chapter 77).

Pulmonary congestion may develop in patients with marginal LV function or critical valvular heart disease. Congestion compromising respiratory and hemodynamic function is an emergency treated with oxygen, diuretics, nitroglycerin, inotropic agents, intubation, and intra-aortic balloon pumping as indicated (Chapter 107).

Chest pain during coronary angiography is unusual, but myocardial ischemia with pain and ST segment changes may occur during PCI (Chapter 74). Treatment with nitroglycerin, heparin, and antiplatelet medications usually controls myocardial ischemia before revascularization (Chapter 72). Minor arrhythmias (e.g., atrial or ventricular premature beats, brief episodes of supraventricular tachycardia) are common and usually resolve without treatment. Ventricular tachycardia or fibrillation is a rare occurrence but requires prompt defibrillation (Chapter 63).

HEMODYNAMIC DATA OBTAINED DURING CARDIAC CATHETERIZATION

Hemodynamic data are the recordings of the pressure and flow signals generated by the heart during the catheterization procedure.[2] A pressure wave is created by cardiac muscle contraction and is transmitted through the arterial circuit (Fig. 57-1). The pressure waves are measured by the fluid-filled catheters with a pressure transducer, which converts the mechanical pressure to an electrical signal that is displayed on a video monitor.

Simultaneous pressure measurements across the heart valves are used to diagnose valve function. In addition to pressure measurements, hemodynamic data also include analysis of multiple blood oxygen saturations sampled throughout the right and left sides of the heart to identify possible intracardiac shunting. Cardiac output is commonly measured by a thermodilution technique but can also be computed by knowing oxygen consumption and comparing systemic arterial oxygenation content to pulmonary arterial oxygen content by the Fick equation. Hemodynamic data permit calculation of vascular and pulmonary resistances and cardiac valve areas (see Table 53-1). Complete hemodynamic data requiring catheterization of the right and left sides of the heart are indicated to evaluate dyspnea of any cause; to confirm echocardiographic findings when data are not concordant with clinical or other testing results; and to determine the status of valvular heart disease, cardiomyopathy, and constrictive or restrictive cardiac physiology.

Complications of right-sided heart catheterization are rare. The most common problem is transient arrhythmia resulting from mechanical stimulation by the catheter as it passes through the right ventricular outflow tract. In patients with left bundle branch block, a temporary pacemaker may be needed if right bundle branch block occurs during right-sided heart catheterization.

Examples of Hemodynamics for Valvular Heart Disease

The severity of a valvular stenosis is based on the pressure gradient and flow across the valve. In patients with suspected aortic stenosis, a transvalvular pressure gradient should be obtained whenever there is conflicting clinical or echocardiographic data. Although the pressure recordings needed to assess the aortic valve gradient can be obtained with a catheter in the left ventricle and the side arm of the sheath in the femoral artery, more accurate recordings can be obtained with a dual-lumen pigtail catheter (Fig. 57-2). The normal aortic valve area is 2.5 to 3.5 cm^2 in adults. Severe aortic valve stenosis is associated with valve areas smaller than 1.0 cm^2.

In patients with mitral stenosis, the valve gradient often is measured with the LV and pulmonary capillary wedge pressures. The most accurate method to compute the mitral stenosis gradient uses the left atrial (obtained by transseptal puncture) and LV pressures (Fig. 57-3). The normal mitral valve area is 4 to 6 cm^2. Valve areas smaller than 1.0 to 1.2 cm^2 are considered severe mitral stenosis.

In patients with mitral regurgitation, the hemodynamics often show a characteristic large *v* wave on the pulmonary capillary wedge tracing

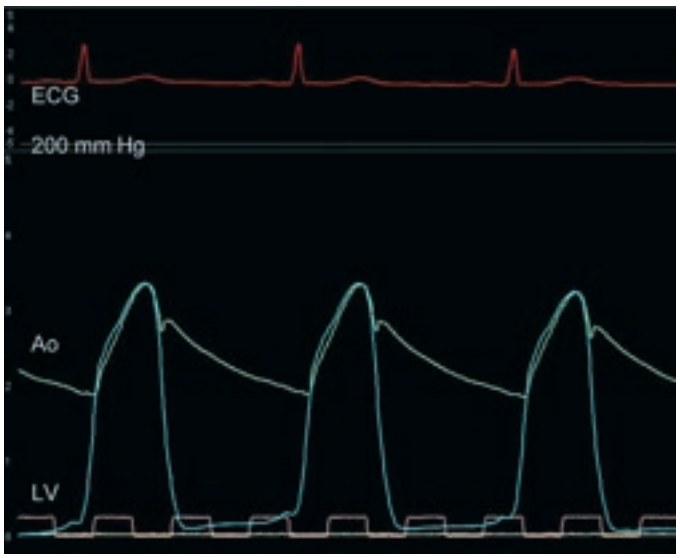

FIGURE 57-1. Normal left ventricular (LV) and aortic (Ao) pressures measured with a high-fidelity dual-transducer catheter.

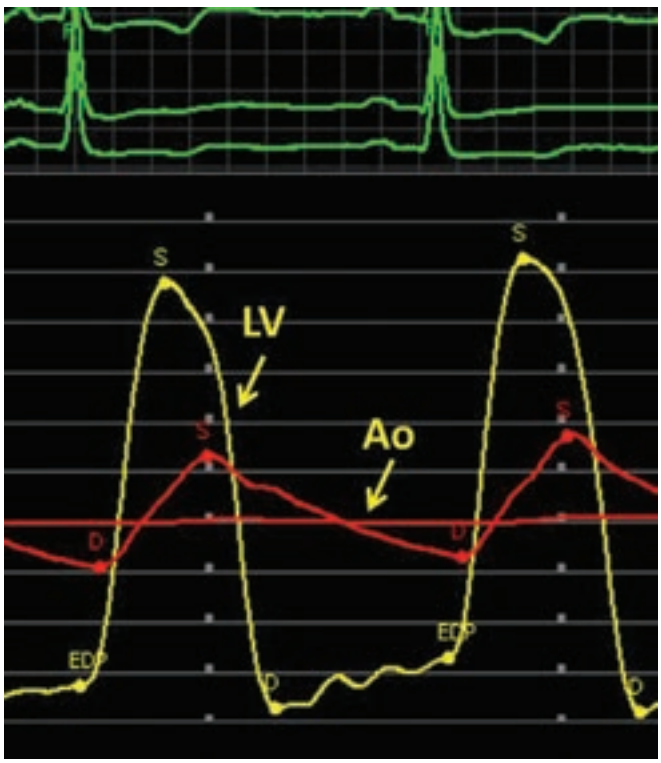

FIGURE 57-2. **Hemodynamics of aortic stenosis.** Aortic (Ao) and left ventricular (LV) pressure in patient with aortic stenosis measured with single dual-lumen catheter. Note delay in aortic pressure relative to left ventricular pressure. D = earliest point of diastolic LV pressure; EDP = end-diastolic pressure; S = peak LV pressure (systole).

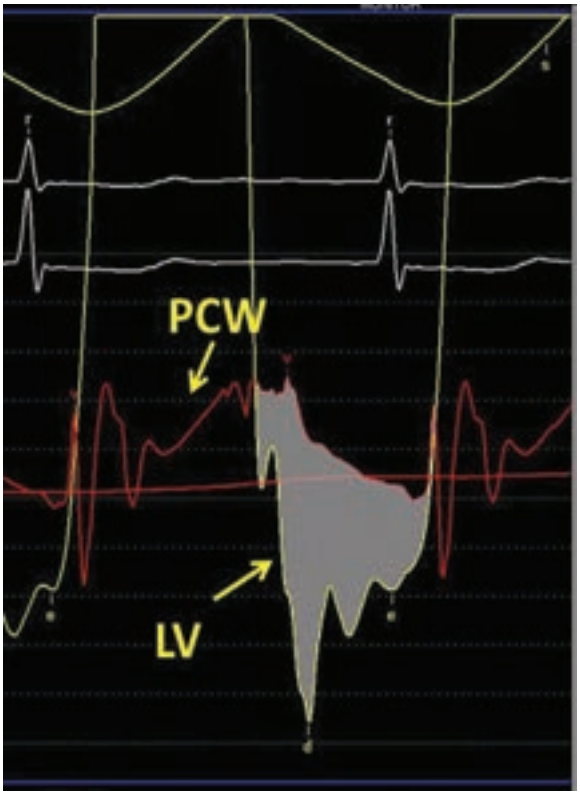

FIGURE 57-3. Mitral stenosis. The difference between the pulmonary capillary wedge (PCW) pressure and the left ventricular (LV) pressure during diastole defines the gradient in mitral valve stenosis.

(E-Fig. 57-2). During left ventriculography, the angiographic grading of mitral regurgitation is on a semiquantitative severity scale of 1 to 4 based on the amount of contrast material seen passing backward from the left ventricle through the incompetent mitral valve into the left atrium. Grade 1 angiographic mitral regurgitation demonstrates a brief puff of contrast material filling the left atrium and emptying immediately; grade 2 shows contrast material filling the left atrium on three beats with moderate density; grade 3 shows contrast material filling immediately and moderate density persisting for two or three beats; and grade 4 shows contrast material filling the entire left atrium, often including the appendage, with a density equal to that of the left ventricle for several beats after injection.

CORONARY ANGIOGRAPHY

Coronary angiography visualizes the epicardial arteries, branches, collaterals, and anomalies to diagnose and to treat patients with coronary artery disease (Fig. 57-4). In anticipation of PCI, the angiogram documents not only the presence and location of stenoses but also proximity to major and minor side branches, luminal abnormalities (e.g., thrombi), areas of calcification, and collateral supply, which will influence the decision and techniques used for revascularization. For the two-dimensional radiographic images to depict the three-dimensional coronary tree, multiple angulations of the radiographic imaging system are required.

Assessment of Coronary Stenoses

The degree of a stenosis is most often reported as the estimated percentage diameter luminal reduction of the most severely narrowed segment compared with the adjacent angiographically "normal" or unobstructed vessel segment, seen in the worst radiographic projection (Fig. 57-5). Because the operator uses visual estimations, an exact evaluation is impossible with less than 20% variation between readings of two or more experienced angiographers. The severity of a stenosis alone should not always be assumed to be associated with abnormal physiology (flow) and ischemia. Moreover, because coronary artery disease is a diffuse process, minimal luminal irregularities on angiography may represent significant, albeit nonobstructive, coronary artery disease at the time of angiography. The precise physiologic impact or morphologic detail of stenosis can be made with specialized catheters and sensor-tipped guidewires. For example, intermediately severe lesions (40 to 70% narrowed) without prior evidence of ischemia can be assessed with a pressure sensor guidewire to measure the translesional pressure during maximal blood flow (e.g., hyperemia induced by adenosine). The ratio of coronary to aortic pressure measured at maximal flow, called the fractional flow reserve (FFR), represents the percentage of normal flow across the stenosis. Lesions with FFR values higher than 0.80 are considered nonischemic and do not require revascularization. Conversely, PCI may be justified for a lesion with an FFR lower than 0.80. After PCI is undertaken, the lesion's length, true diameter, eccentricity, and degree of calcium involvement can be determined by intravascular ultrasound catheter-based imaging or with reflected light by optical coherence tomography to depict intravascular anatomy and to guide subsequent PCI (Fig. 57-6). However, neither intravascular ultrasound nor optical coherence tomography can substitute for a functional determination of ischemia by FFR calculation. The degree of collateralization detected on angiography is also important. For example, stable coronary artery disease patients with a high degree of collateral vessels have a 36% lower mortality compared with patients with low collateralization.[3]

Coronary Artery Anomalies

Coronary artery anomalies may be present in patients with chest pain syndromes or in young individuals who have survived sudden death. The angiographic appearance of coronary anomalies is not always straightforward, and computed tomographic angiography (Chapter 56) has become the diagnostic modality best suited to delineate the origin and course of the anomaly (Fig. 57-7).

At the time of coronary angiography, the misdiagnosis of an unsuspected anomalous origin of a coronary artery is a potential problem. Because the natural history of a patient with an anomalous origin of a coronary artery may depend on the anatomic pathway of the anomalous vessel, it is important to define accurately the origin and course of the vessel. Even experienced angiographers may have difficulty in delineating the true course of some anomalous vessels.

The most critical coronary anomaly is when the left main coronary artery arises from the right cusp and traverses a course between the aorta and pulmonary artery. The artery's initial course through the aortic wall creates a narrow oval opening; with aortic stretch during exercise, coronary blood flow is limited.

Ventriculography

The left ventriculogram, an integral part of nearly every coronary angiographic study, provides information on the motion of the walls of the heart (Fig. 57-8), LV volumes during systole and diastole, LV ejection fraction, rate of ejection, quality of contractility, presence of hypertrophic myopathy, and mitral valvular regurgitation. The normal pattern of LV contraction is a coordinated, uniform, almost concentric inward motion of all points along the ventricular inner surface during systole. Uncoordinated contractions are named according to their severity (e.g., moderate or severe hypokinesis, akinesis, and aneurysm-dyskinesis). Focal abnormal wall motion indicates the presence of ischemia, infarction, or aneurysm.

ADDITIONAL PROCEDURES PERFORMED IN THE CATHETERIZATION LABORATORY

Noncardiac Angiography

Other cardiovascular angiographic studies that may accompany coronary angiography and left ventriculography include aortography (Chapter 78), pulmonary angiography (Chapter 98), and peripheral vascular angiography of iliac and lower extremity arteries (Chapters 79 and 80) or renal arteries (Chapter 125).

Electrophysiologic Studies and Ablation Techniques

An electrophysiologic study (EPS) is an invasive procedure that involves the placement of multipolar catheter electrodes at various intracardiac sites (Chapter 62). The general purposes of an EPS are to characterize the electrophysiologic properties of the conduction system, to induce and analyze the mechanism of arrhythmias, and to evaluate the effects of therapeutic interventions. Electrode catheters are routinely placed in the right atrium, across the tricuspid valve annulus in the area of the atrioventricular node and His bundle, in the right ventricle, in the coronary sinus, and sometimes in the left ventricle. EPS is routinely used in the clinical management of patients who have supraventricular and ventricular arrhythmias (Chapters 64 and 65).

Transseptal Heart Catheterization

Transseptal access by use of a long catheter with a needle to puncture the thin atrial septal membrane at the fossa ovalis permits placement of a catheter into

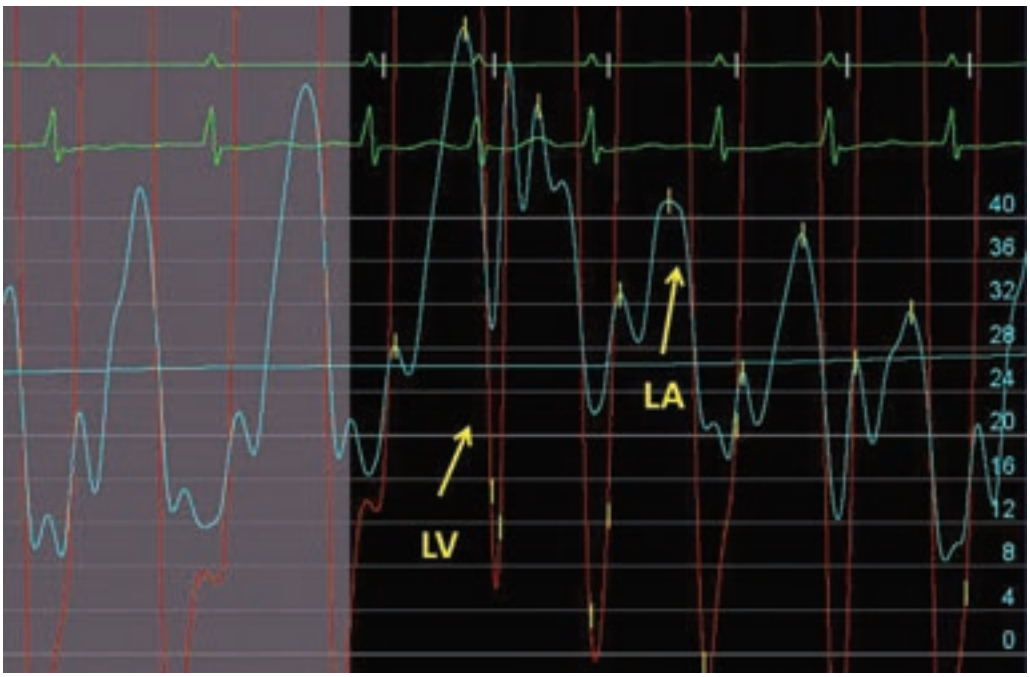

E-FIGURE 57-2. Hemodynamic tracings in a patient with mitral regurgitation characterized by giant *v* wave in the left atrial pressure. Giant *v* wave of severe mitral regurgitation seen on left atrial (LA) pressure tracing (*blue*) compared with left ventricular (LV) pressure tracing (*red*) on a scale of 0 to 40 mm Hg.

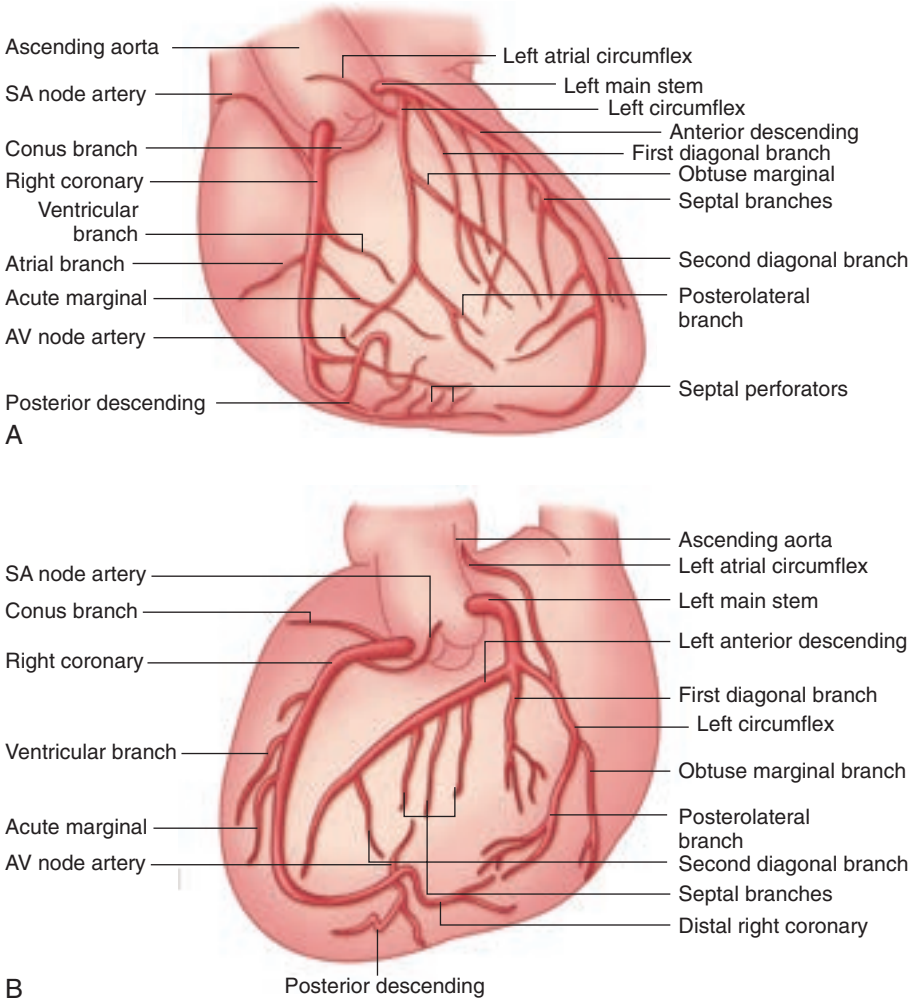

FIGURE 57-4. **Coronary vessels.** The right anterior oblique (**A**) and left anterior oblique (**B**) views are shown. The major arteries are the left main, left anterior descending, circumflex, and right coronary arteries. AV = atrioventricular; SA = sinoatrial. (Modified from Yang SS, Bentivoglio LG, Maranhao V, et al, eds. From Cardiac Catheterization Data to Hemodynamic Parameters. Philadelphia: Oxford University Press; 1988.)

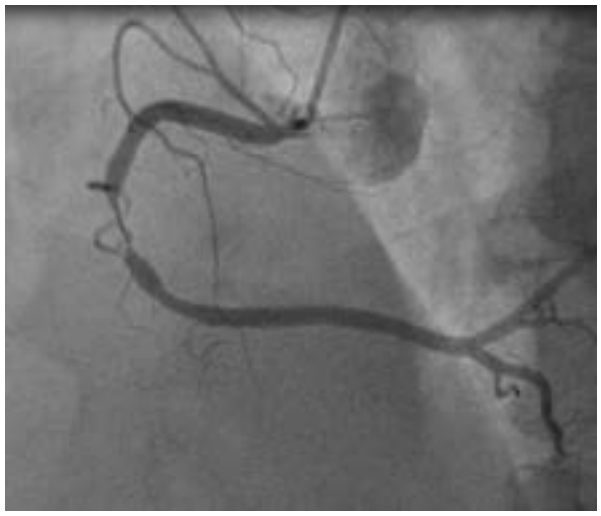

FIGURE 57-5. Example of a significant stenosis in the right coronary artery.

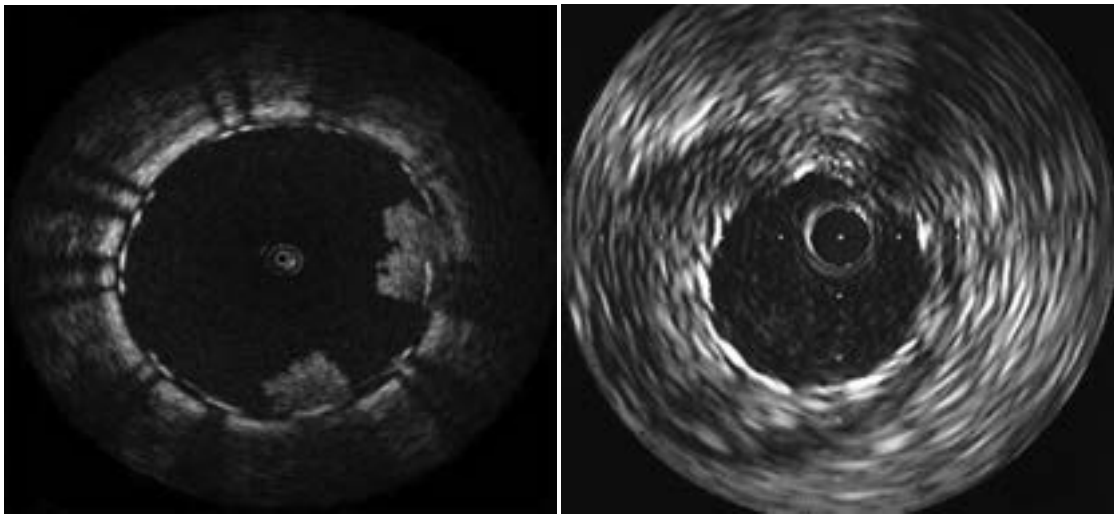

FIGURE 57-6. Cross-sectional image within a coronary artery by intravascular ultrasound (*right*) and optical coherence tomography (*left*).

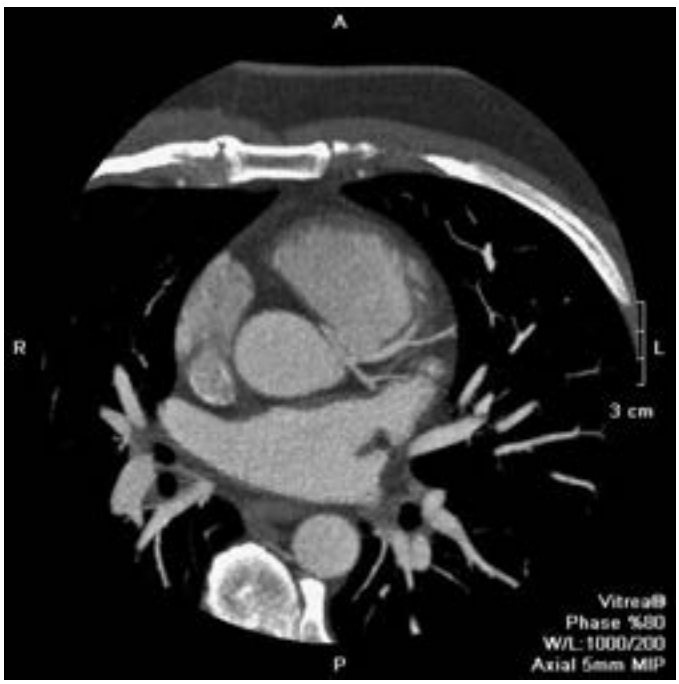

FIGURE 57-7. Frame from a computed tomographic angiographic study showing the origin of the left main coronary artery arising from the right sinus of Valsalva and coursing anteriorly between the aorta and pulmonary artery.

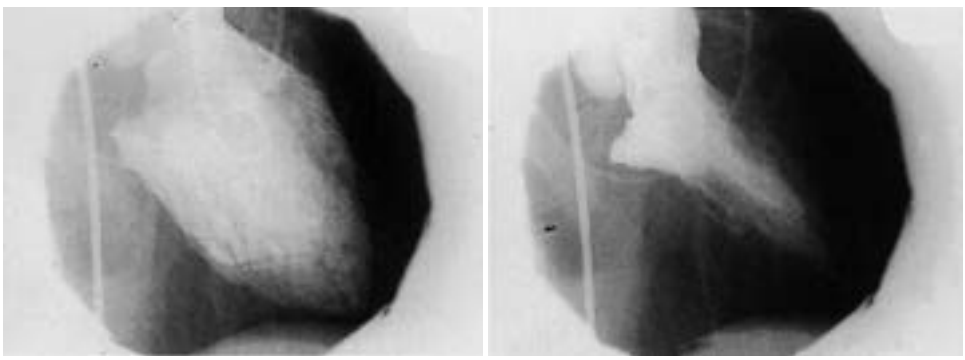

FIGURE 57-8. Example of left ventriculography. The ventricular contour is seen in diastole (*left*) and in systole (*right*).

the left atrium and then the left ventricle. It is an established technique used to acquire precise, high-quality hemodynamic data for patients with aortic stenosis, mitral valve disease (both stenosis and regurgitation), and hypertrophic cardiomyopathy (outflow tract obstructive gradient) and to provide access for valvuloplasty techniques (Chapter 75). The risks of transseptal catheterization, which include punctures of the aortic root, the coronary sinus, or the posterior free wall of the atrium, are potentially lethal problems.

Endomyocardial Biopsy

Endomyocardial biopsy procedures use venous access from the internal jugular vein or femoral vein to insert a flexible metal bioptome to obtain four to six 1-mm^3 pieces of right ventricular myocardium. There are two definitive indications for endomyocardial biopsy: monitoring for cardiac transplant rejection (Chapter 82) and detection of anthracycline cardiotoxicity (Chapter 60). Other indications in selected patients include diagnosis of cardiomyopathy and myocarditis and differentiation between restrictive and constrictive cardiomyopathies.

Other Procedures

Pericardiocentesis (Chapter 77) and PCI (Chapter 74) are performed in the catheterization laboratory for specific indications. Other procedures include balloon valvuloplasty and transcatheter aortic valve replacement for stenotic heart valves (Chapter 75),[4] and closure of an atrial septal defect or patent foramen ovale (Chapter 69), which is confirmed by typical oxygen saturations detected at catheterization.

Grade A References

A1. Romagnoli E, Biondi-Zoccai G, Sciahbasi A, et al. Radial versus femoral randomized investigation in ST-segment elevation acute coronary syndrome: the RIFLE-STEACS (Radial Versus Femoral Randomized Investigation in ST-Elevation Acute Coronary Syndrome) study. *J Am Coll Cardiol*. 2012;60:2481-2489.

A2. Mehta SR, Jolly SS, Cairns J, et al. Effects of radial versus femoral artery access in patients with acute coronary syndromes with or without ST-segment elevation. *J Am Coll Cardiol*. 2012;60:2490-2499.

A3. Koc F, Ozdemir K, Altunkas F, et al. Sodium bicarbonate versus isotonic saline for the prevention of contrast-induced nephropathy in patients with diabetes mellitus undergoing coronary angiography and/or intervention: a multicenter prospective randomized study. *J Investig Med*. 2013;61:872-877.

A4. Acetylcysteine for prevention of renal outcomes in patients undergoing coronary and peripheral vascular angiography: main results from the randomized Acetylcysteine for Contrast-induced nephropathy Trial (ACT). *Circulation*. 2011;124:1250-1259.

A5. Leoncini M, Toso A, Maioli M, et al. Early high-dose rosuvastatin for contrast-induced nephropathy prevention in acute coronary syndrome: Results from the PRATO-ACS Study (Protective Effect of Rosuvastatin and Antiplatelet Therapy On contrast-induced acute kidney injury and myocardial damage in patients with Acute Coronary Syndrome). *J Am Coll Cardiol*. 2014;63:71-79.

A6. Han Y, Zhu G, Han L, et al. Short-term rosuvastatin therapy for prevention of contrast-induced acute kidney injury in patients with diabetes and chronic kidney disease. *J Am Coll Cardiol*. 2014;63:62-70.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Patel MR, Bailey SR, Bonow RO, et al. ACCF/SCAI/AATS/AHA/ASE/ASNC/HFSA/HRS/SCCM/SCCT/SCMR/STS 2012 appropriate use criteria for diagnostic catheterization: a report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, Society for Cardiovascular Angiography and Interventions, American Association for Thoracic Surgery, American Heart Association, American Society of Echocardiography, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society of Critical Care Medicine, Society of Cardiovascular Computed Tomography, Society for Cardiovascular Magnetic Resonance, and Society of Thoracic Surgeons. *J Am Coll Cardiol*. 2012;59:1995-2027.
2. Nishimura RA, Carabello BA. Hemodynamics in the cardiac catheterization laboratory of the 21st century. *Circulation*. 2012;125:2138-2150.
3. Meier P, Hemingway H, Lansky AJ, et al. The impact of the coronary collateral circulation on mortality: a meta-analysis. *Eur Heart J*. 2012;33:614-621.
4. Nishimura RA, Otto CM, Bonow RO, et al. 2014 AHA/ACC guideline for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Thorac Cardiovasc Surg*. 2014;148:e1-e132.

REVIEW QUESTIONS

1. Which of the following patients is suitable for cardiac catheterization?

A. A 60-year-man with atypical chest discomfort, no risk factors for coronary artery disease, a negative stress test result, and relief of symptoms with famotidine
B. A 67-year-man with prior coronary artery bypass surgery who is asymptomatic and fully active and now wants to buy new life insurance
C. A 39-year-old woman with shortness of breath and a midsystolic click on physical examination
D. A 41-year-old man with palpitations
E. A 53-year-old man with palpitations, hypertension with β-blockers and aspirin, and reversible defect on stress perfusion imaging

Answer: E Indications for cardiac catheterization and coronary angiography according to guidelines recommend invasive angiography for patients with objective evidence of myocardial ischemia after treatment with medical therapy. Patient A has no ischemia and likely has gastrointestinal symptoms. Patient B is asymptomatic and is not in need of angiography. Patient C needs an echocardiogram before consideration of invasive assessment. Patient D has palpitations of unknown cause and could be evaluated with an echocardiogram, ambulatory arrhythmia monitoring, or a stress test, but cardiac catheterization is not indicated unless an ischemic origin is demonstrated for his palpitations.

2. Which patient has the highest risk of complications of cardiac catheterization?

A. An 82-year-old woman with hypertension
B. A 76-year-old man with creatinine concentration of 1.7 mg/dL
C. A 61-year-old woman with body mass index of 28
D. A 58-year-old man with a heart murmur
E. A 40-year-old woman with a history of syncope

Answer: B Cardiac catheterization requires administration of radiocontrast media, which can impair renal function. In patients with preexisting renal failure, worsening renal failure is likely. Measures to prevent contrast-induced renal failure include preprocedural hydration to increase urine output.

3. Which of the following conditions does not require a right-sided heart hemodynamic study?

A. Left ventricular systolic failure
B. Left ventricular diastolic failure
C. Pericardial disease
D. Aortic stenosis
E. Severe three-vessel coronary artery disease

Answer: E Right-sided heart catheterization is not routine for coronary artery disease but is mandatory for valvular heart disease and routine for heart failure, right ventricular dysfunction, pericardial disease, cardiomyopathy, intracardiac shunts, and congenital abnormalities.

4. Which of the following is an absolute contraindication to cardiac catheterization?

A. Ventricular tachycardia
B. Sepsis
C. Inadequate imaging equipment
D. Uncontrolled hypertension
E. Creatinine concentration of 2.4 mg/dL

Answer: C Absolute contraindications to catheterization include inadequate facilities and patient refusal. Relative contraindications include severe uncontrolled hypertension, ventricular arrhythmias, recent acute stroke, severe anemia, active gastrointestinal bleeding, allergy to radiographic contrast material, acute renal failure, uncompensated heart failure such that the patient cannot lie flat, unexplained febrile illness or untreated active infection, electrolyte abnormalities (e.g., hypokalemia), severe coagulopathy, pregnancy, uncontrolled arrhythmias or hypertension, and an uncooperative patient.

5. Which of the following coronary anomalies is associated with sudden death?

A. A right coronary artery originating from the left coronary sinus
B. A right coronary artery originating from the anterior coronary cusp
C. A left coronary artery that originates from the right coronary cusp and traverses between the aorta and pulmonary artery
D. A circumflex coronary artery that arises from the right coronary cusp and traverses between the aorta and pulmonary artery
E. A circumflex coronary artery that arises from the right coronary cusp and traverses behind the aorta

Answer: C The most critical coronary anomaly is when the left main coronary artery arises from the right cusp and traverses a course between the aorta and pulmonary artery. The artery's initial course through the aortic wall creates a narrow oval opening; when the aorta stretches during exercise, coronary blood flow is limited.

58

HEART FAILURE: PATHOPHYSIOLOGY AND DIAGNOSIS

CHRISTOPHER M. O'CONNOR AND JOSEPH G. ROGERS

DEFINITION

Heart failure is a clinical syndrome that results when abnormalities in the structure and function of the myocardium impair cardiac output or decrease filling of the ventricles. Characteristic features of the heart failure syndrome include dyspnea (shortness of breath), fatigue, fluid retention, impaired exercise performance, and edema. Pulmonary congestion is a common but not universal feature, so the term *congestive heart failure* is no longer used.

EPIDEMIOLOGY

The lifetime risk of heart failure is at least 20% for Americans and 25% for Europeans.[1] Although these cumulative incidences have remained stable during the past 30 years, the prevalence of heart failure has increased because of the improved long-term survival of patients with ischemic and other forms of heart disease as well as the strong association between heart failure and advancing age.[2] For example, more than 10% of the population older than 80 years has heart failure, and patients are living progressively longer with clinical heart failure. In the United States today, more than 5 million patients have clinical heart failure, and they generate more than one million hospitalizations for heart failure annually. Heart failure is the leading cause of hospitalization for patients older than 65 years. Once a patient is hospitalized for heart failure, the 30-day risk of rehospitalization is 25%, with a 10% risk of 30-day postdischarge mortality. Although survival has improved, the absolute mortality rates for heart failure remain approximately 50% within 5 years of diagnosis.

Approximately 40 to 50% of patients with heart failure have a preserved ejection fraction. Compared with patients who have heart failure with reduced ejection fraction, heart failure patients with a preserved ejection fraction tend to be older and are more likely to be women with a prior history of hypertension and diabetes. The mortality rate of patients with a preserved ejection fraction is somewhat lower than the rate for patients with reduced ejection fraction but is still higher than for an age-matched population.[3] Hospitalization and rehospitalization rates are comparable for heart failure patients regardless of their ejection fractions.

Disparities

African Americans are at increased risk for development of heart failure compared with white Americans and also have more frequent hospitalizations.[4] African Americans with heart failure have a similarly high mortality rate compared with white populations. The reasons for these disparities are multifactorial and include differences in heart failure etiology, prevalence of comorbidities, socioeconomic status, and response to therapies.

Hispanics represent a growing number of heart failure patients in the United States. Although large-scale studies are currently limited, preliminary data suggest this population is particularly vulnerable to heart failure.

Classification

The classification of heart failure considers the disease state and progression, degree of exercise intolerance (Chapter 51), measurement of heart function by ejection fraction, and its cause.[5] In the first approach, heart failure is classified by its progression through four stages from pre-disease to advanced symptoms (Fig. 58-1). The second approach evaluates functional states by classification systems such as the New York Heart Association classification, the Canadian Cardiovascular Society system, or other validated measures of exercise tolerance (see Table 51-5). Functional status is the most appropriate way to understand a patient's limitations and the impact of heart failure on quality of life. It also correlates with prognosis and is often used as a secondary end point in clinical trials. The third approach is to measure ejection fraction. This method, usually obtained by echocardiography (Chapter 55), dichotomizes the heart failure patient as either having a reduced (<50%) or preserved ejection fraction. This simple dichotomous classification helps in understanding the patient's underlying pathophysiologic process and in identifying appropriate treatment strategies. Classification based on cause, usually ischemic versus nonischemic (Chapter 60), is useful in guiding the diagnostic evaluation in tailoring treatment strategies. Deeper phenotypic characterization is a focus of ongoing research.[6]

Causes of Heart Failure

Heart failure has numerous causes (Table 58-1). In the United States and other developed countries, coronary artery disease causes about 70% of heart failure cases usually related to a myocardial infarction (Chapter 73). Although hypertension is the second leading cause of heart failure in Western developed countries, it represents the leading cause of heart failure in many developing countries. Diseases of the myocardium (Chapter 60), valves (Chapter 75), and pericardium (Chapter 77) as well as endocrinopathies, metabolic abnormalities, and genetic conditions may cause heart failure. Chemotherapy (e.g., anthracyclines and trastuzumab; Chapter 179) and chest radiation therapy (Chapter 20) may damage heart muscle. Reversible myocardial dysfunction from toxin exposure (e.g., cocaine [Chapter 34] and excessive alcohol intake [Chapter 33]), persistent tachycardia (Chapters 64 and 65), and severe mental/emotional stress (e.g., Takotsubo cardiomy-

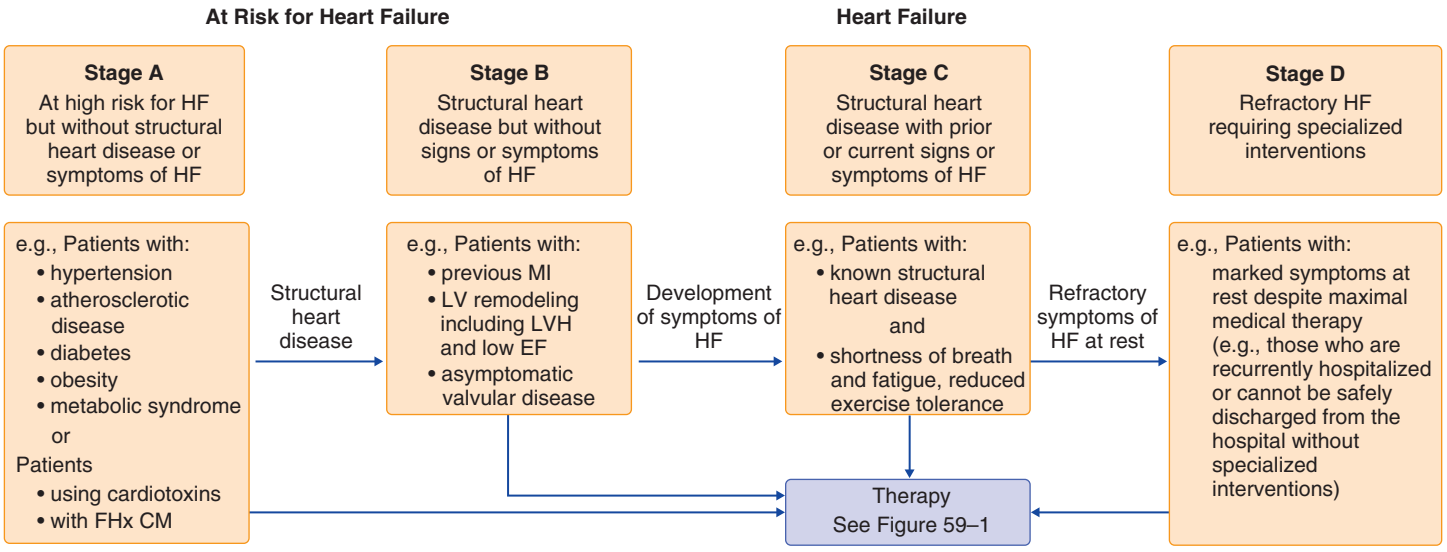

FIGURE 58-1. Stages of heart failure. EF = ejection fraction; FHx CM = family history of cardiomyopathy; HF = heart failure; LV = left ventricle; LVH = left ventricular hypertrophy; MI = myocardial infarction. (Modified from Hunt SA. ACC/AHA 2005 guideline update for the diagnosis and management of chronic heart failure in the adult: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines [Writing Committee to Update the 2001 Guidelines for the Evaluation and Management of Heart Failure]. *J Am Coll Cardiol.* 2005;46:e1-e82.)

TABLE 58-1 CAUSES OF HEART FAILURE

Coronary artery disease or prior myocardial infarction or ischemic injury
Hypertension
Familial and genetic disorders, including dilated cardiomyopathies, hypertrophic cardiomyopathies, storage diseases, and muscular dystrophies
Valvular disease: valvular stenosis or regurgitation
Toxic/drug-induced damage, including prior chemotherapy
Infiltrative processes, such as sarcoid, amyloid, and hemochromatosis (i.e., restrictive cardiomyopathy)
Arrhythmia-related dysfunction, including premature ventricular contraction–induced cardiomyopathy and atrial tachyarrhythmia–related dysfunction
Arrhythmogenic right ventricular cardiomyopathy
Pulmonary heart disease, including cor pulmonale
Infectious agents, including viral infections and Chagas disease
Immunologically mediated myocardial processes
Shunting: intracardiac or extracardiac, including arteriovenous fistulas
Constrictive pericarditis (i.e., nonmyocardial processes)
Age-related changes
Nutritional disorders, such as beriberi
High-output states, such as chronic anemia, thyrotoxicosis

TABLE 58-2 FACTORS THAT MAY PRECIPITATE ACUTE DECOMPENSATION OF CHRONIC HEART FAILURE

Myocardial ischemia or infarction
Arrhythmias
Worsening hypertension
Worsening mitral or tricuspid regurgitation
Initiation of medications that worsen heart failure (calcium antagonists, β-blockers, nonsteroidal anti-inflammatory drugs, antiarrhythmic agents)
Discontinuation of therapy (patient noncompliance or physician initiated)
Dietary indiscretion
Iatrogenic volume overload (transfusion, fluid administration)
Alcohol consumption
Increased activity
Fever or infection
Anemia
Thyroid abnormalities
Exposure to high altitude
Pregnancy

opathy [Chapter 60]) may cause ventricular dysfunction that resolves following withdrawal of the inciting etiology.

The relative pathophysiologic mechanisms that contribute to the burden of heart failure vary by world region. For instance, infectious causes such as Chagas cardiomyopathy (Chapter 60) are prevalent in Central and South America. Rheumatic valvular disease is a common cause of heart failure in developing countries but is infrequent in developed countries.

Because of the large number of possible causes of heart failure, identification of the primary contributing factors may pose diagnostic challenges for the clinician. Furthermore, an acute decompensation of heart failure may be related to worsening of the primary cause, to a new cardiac or pulmonary problem, to the influence of cardiac conditions that demand increased cardiac output, or to poor adherence to otherwise successful therapies (Table 58-2). As a result, a systematic approach to the clinical evaluation is critical. Some patients will have decompensated heart failure that requires hospitalization, whereas others will have heart failure as an important comorbidity during hospitalization for another problem.

PATHOBIOLOGY

A number of factors contribute to the pathobiologic development and progression of heart failure (Table 58-3). The relative contribution of these different mechanisms to the overall pathophysiologic manifestations of heart failure (Fig. 58-2) varies in individual patients.

TABLE 58-3 PATHOBIOLOGIC MECHANISMS OF HEART FAILURE

Hemodynamics
Neurohormones
Cardiorenal interactions
Abnormal calcium cycling
Cell death
Myocardial genetics

Heart Failure with a Reduced Ejection Fraction

A central tenet of the hemodynamic contribution to heart failure is that an abnormality of myocardial function is responsible for the failure of the heart to pump blood at a rate commensurate with the requirements of the metabolizing tissues or that it can do so only in the setting of elevated cardiac filling pressures.[7] In failing hearts, an increase in hemodynamic load causes maladaptive regulation of neurohormonal pathways that reduce intrinsic muscle

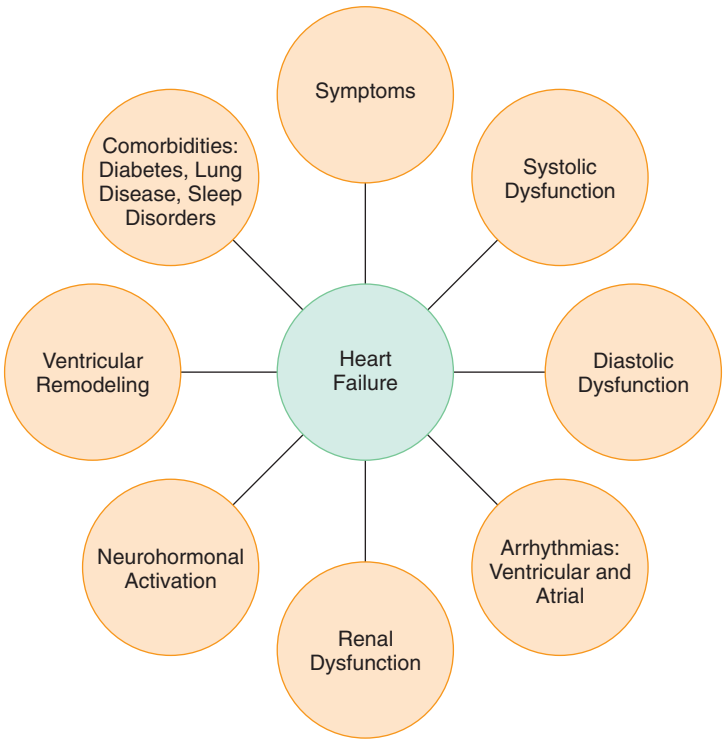

FIGURE 58-2. Pathophysiology of heart failure. The contributing and exacerbating factors to the pathophysiologic process of heart failure.

contractility (Chapter 53), in part because of ventricular remodeling. The structure of the extracellular matrix of the heart plays a pivotal role in ventricular scaffolding and overall pumping function. Replacement fibrosis after myocardial injury (e.g., myocardial infarction, toxin exposure, or chronic renin-angiotensin-aldosterone system activation) increases the connective tissue content and further impairs pump function (Chapter 53). Several neurohormonal pathways have been identified as regulators of myocardial fibrosis including aldosterone, matrix metalloproteinases, the tissue inhibitors of metaloproteinases (TIMPS), TNF-alpha and ST2.

Neurohumoral mechanisms emphasize the importance of adrenergic nervous system activation in the development and progression of left ventricular dysfunction. The initial activation of the sympathetic nervous system probably results from reduced pulse pressure, which activates arterial baroreceptors. Evidence for its activation comes from elevated levels of circulating norepinephrine, direct sympathetic nerve recordings showing increased activity, and increased release of norepinephrine by several organs, including the heart. Elements of the renin-angiotensin-aldosterone system are activated relatively early in heart failure. The presumptive mechanisms of induction include renal hypoperfusion, β-adrenergic system stimulation, and hyponatremia. Adrenergic and renin-angiotensin-aldosterone system activation stimulates the failing heart while also causing peripheral vasoconstriction and the retention of sodium and fluid. The initial result is improved circulation and perfusion of the vital organs. Over time, however, the prolonged activation of these systems causes maladaptive remodeling of the left ventricle and further dysfunction. Further, the sympathetic nervous system and renin-angiotensin-aldosterone system are co-regulated, such that increased activity of each pathway stimulates a simultaneous increase in the other. As cardiac function deteriorates, responsiveness to norepinephrine diminishes, as evidenced by baroreceptor desensitization and downregulation of cardiac adrenergic receptors. This desensitization may further stimulate sympathetic responses. When excessive vasoconstriction depresses left ventricular function, sodium retention increases already elevated ventricular filling pressures. As a result, heart failure is characterized by hypoperfusion associated with hypervolemia. The deleterious effects of vasoconstriction and volume retention are hallmarks of clinical heart failure syndromes.

Natriuretic peptides may counterbalance the vasoconstricting and sodium-retaining actions of the renin-angiotensin-aldosterone system and sympathetic nervous systems by causing both arterial and venous vasodilation as well as natriuresis and diuresis. Similar to other activated neurohormonal pathways in heart failure, the degree of natriuretic pathway activation is associated with adverse patient outcomes, including the risk for hospitalization and death. The favorable counterbalancing effects of the natriuretic pathways suggest that further stimulation or administration of exogenous natriuretic hormones might improve heart failure signs, symptoms, or outcomes. However, the intravenous infusion of the recombinant natriuretic peptide nesiritide does not provide significant clinical benefit.

TABLE 58-4 CONTRIBUTIONS TO HEART FAILURE WITH PRESERVED EJECTION FRACTION

Left ventricular hypertrophy
Hypertension
Myocardial fibrosis
Subendocardial fibrosis (from intermittent ischemia, especially with diabetes)
Arterial stiffness
Endothelial dysfunction

Many patients with heart failure have elevated levels of endothelin and arginine vasopressin. Arginine vasopressin induces vasoconstriction through a vascular (V_1) receptor and reduces free water clearance through a renal tubular (V_2) receptor. Endothelin causes prolonged vasoconstriction, reductions in glomerular filtration, and pulmonary arteriolar constriction. Although both endothelin and vasopressin are attractive targets for therapy, clinical trials with antagonists of these systems have been negative, thereby indicating that interdiction of neurohormonal activation is not uniformly beneficial.

Cardiorenal mechanisms emphasize the integral role of kidney dysfunction in the worsening of heart failure, with both organs contributing to the retention of sodium and water. In most patients with chronic heart failure, the kidneys are anatomically and structurally normal, but passive venous congestion and reduced renal perfusion lead to worsening renal function and a vicious circle of progressive dysfunction of both organ systems. Renal failure also compromises the ability to use inhibitors of the renin-angiotensin-aldosterone system to treat heart failure.

Abnormal calcium cycling reduces the calcium content in the cardiac sarcoplasmic reticulum owing to diastolic leak through altered ryanodine receptors. The reduced calcium content of the sarcoplasmic reticulum alters the contraction interactions between cardiac myosin and actin myofilaments. In addition, a loss of function of cardiac SERCA2a (sarcoendoplasmic reticulum calcium transport ATPase 2a) pumps, which are responsible for removal of cytoplasmic calcium, affects ventricular relaxation and causes diastolic dysfunction.

In all forms of heart failure, a variety of stressors (e.g., elevated neurohormone levels, adrenergic activation, inflammatory mechanisms, and toxin exposure) enhance cell death. Elevated serum levels of many proinflammatory cytokines (including tumor necrosis factor-α, interleukin-1β, and interleukin-6) may induce contractile dysfunction, myocardial fibrosis, and myocyte necrosis, perhaps by mediating some of the deleterious responses to catecholamines and angiotensin II. Cell death is also a cause of heart failure after myocardial infarction.

Persistent tachycardia (e.g., atrial fibrillation with rapid ventricular response) also can result in heart failure, presumably related to chronic hyperadrenergic stimulation. In some patients, the tachycardic response to heart failure may itself contribute to worsening heart failure. These forms of heart failure may be reversible, depending on the relative contribution of the tachycardia to the underlying myocardial dysfunction.

Genetic mutations are increasingly identified as important mediators of the cardiac structural and functional abnormalities linked to symptomatic heart failure (see Table 60-2). The majority of familial cardiomyopathies, most of which are inherited in an autosomal dominant fashion, are related to defects in the cytoskeleton or nuclear proteins. Inherited cardiomyopathy is also associated with muscular dystrophies, infiltrative diseases such as hemochromatosis, and mitochondrial disorders.

Heart Failure with a Preserved Ejection Fraction

The pathophysiologic mechanism of heart failure with a preserved ejection fraction is complex and includes alterations in cardiac structure and function, vascular abnormalities, end-organ dysfunction, and interrelated comorbidities (Table 58-4). In principle, diastolic dysfunction in heart failure with a preserved ejection fraction can result from increased left ventricular stiffness due to hypertrophy and interstitial fibrosis as well as from abnormal left ventricular relaxation due to dysfunctional calcium cycling. Although there are many potential causes of heart failure with a preserved ejection fraction,

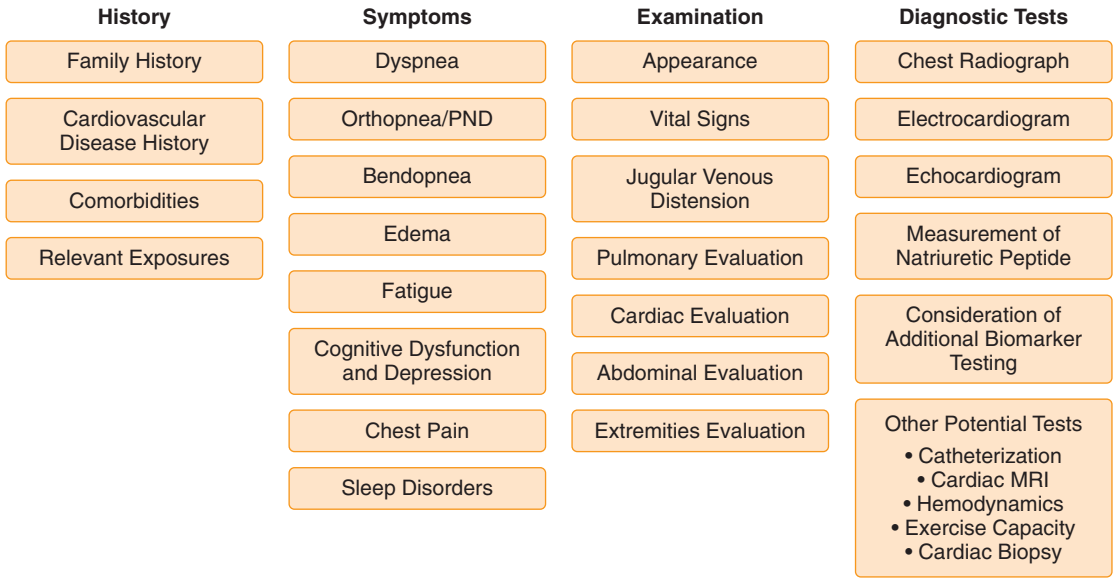

FIGURE 58-3. Diagnostic evaluation of heart failure. The various components of the diagnostic evaluation of heart failure patients are presented from the history, symptoms, and examination to diagnostic testing. MRI = magnetic resonance imaging; PND = paroxysmal nocturnal dyspnea.

most patients have current or prior hypertension; the resulting left ventricular hypertrophy and fibrosis are responsible for increased chamber stiffness. Ischemic heart disease also may contribute to heart failure with a preserved ejection fraction, by virtue of subendocardial fibrosis or as a result of intermittent ischemic dysfunction. Diabetes mellitus is often present, especially in women. Age itself is a crucial predisposing factor because it causes loss of myocytes (apoptosis), increased fibrosis with shifts to more rigid forms of collagen, and loss of vascular compliance.

Myocardial relaxation is an adenosine triphosphate–dependent process. Processes that interfere with myocardial energy metabolism (e.g., ischemia) compromise myocardial relaxation. These changes result in reduced ventricular compliance and elevated filling pressures. Elevated filling pressures increase the pulmonary capillary wedge pressure and contribute to the sensation of dyspnea. Diastolic dysfunction may remain asymptomatic for years, but increasing age, renal dysfunction, hypertension, and progressive left ventricular dysfunction are associated with the development of symptoms of heart failure. The resting hemodynamic profile in heart failure with preserved ejection fraction is frequently normal, but physiological perturbations such as tachycardia or exercise result in exaggerated increases in filling pressures. This phenomenon accounts for the marked exertional impairment seen in this syndrome. Furthermore, because atrial contraction is responsible for a disproportionately large percentage of the diastolic filling of a noncompliant ventricle, atrial fibrillation (i.e., loss of atrial contraction) can severely worsen patients' symptoms. Abnormalities in ventricular relaxation and myocardial stiffness limit ventricular filling, so patients may have a narrow window for optimal fluid volume. Modest volume overload can substantially exacerbate symptoms of dyspnea, and therapeutic diuresis may precipitate symptomatic hypotension owing to ventricular underfilling. Although diastolic dysfunction can occur alone as heart failure with a preserved ejection fraction, the majority of patients with significant diastolic dysfunction also have systolic dysfunction.[8] As a result, it is preferable to characterize patients as having heart failure with a preserved ejection fraction or heart failure with a reduced ejection fraction rather than as having systolic or diastolic heart failure.

CLINICAL MANIFESTATIONS

The diagnosis of heart failure may be apparent when patients present with classic symptoms of shortness of breath in combination with a clinical examination consistent with volume overload. A previous history of a myocardial infarction or poorly controlled hypertension should increase the clinician's suspicion for the diagnosis. In contrast, the diagnosis may be missed or delayed in patients who experience a more insidious course with vague symptoms, such as fatigue and exercise intolerance. The use of biomarkers such as natriuretic peptides has significantly improved the diagnostic yield in recent years.

The evaluation of a patient with suspected heart failure should proceed in an organized and focused manner (Fig. 58-3 and Table 58-5).

TABLE 58-5 APPROACH TO THE DIAGNOSIS OF HEART FAILURE

Obtain a history and physical examination to identify disorders or behaviors that might cause or exacerbate heart failure.
Obtain a family history to aid in diagnosis of familial causes of heart failure.
Assess volume status and vital signs at each encounter.
Patients with suspected new-onset or acute heart failure should undergo chest radiography to assess heart size and pulmonary congestion and to detect other diseases that may cause or contribute to the patient's symptoms.
A 12-lead electrocardiogram should be obtained for all patients presenting with heart failure.
Echocardiography should be performed during the initial evaluation of patients with heart failure to assess ventricular function and valve function.
Measurement of natriuretic peptides is recommended for the following indications: To support the diagnosis of heart failure in ambulatory patients with dyspnea as well as in those with possible acute heart failure, especially in the setting of an uncertain diagnosis. To establish prognosis or disease severity in heart failure.
The initial evaluation of patients presenting with heart failure should include a complete blood count, urinalysis and renal function, serum electrolytes, glucose and lipid profile, liver function tests, and thyroid-stimulating hormone.
Hemodynamic monitoring is recommended to guide therapy in patients who have respiratory distress or clinical evidence of impaired perfusion in whom the adequacy or excess of intracardiac filling pressures cannot be determined from clinical assessment.

Modified from Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2013;62:e147-e239.

Symptoms

Shortness of Breath

Dyspnea (Chapter 83) is the most common but nonspecific symptom of heart failure because patients with predominant lung disease or anemia may have similar symptoms. In most heart failure patients, dyspnea is present only with activity. It is the most common reason that patients seek care for heart failure, both during the chronic state and with acutely decompensated heart failure. The most important cause of dyspnea is pulmonary congestion that increases the accumulation of interstitial or intra-alveolar fluid, reduces lung compliance, and increases the work of breathing (Fig. 58-4). Dyspnea relief is a primary therapeutic target of heart failure treatment. Dyspnea can be quantified and monitored by a validated Likert dyspnea scale, which typically consists of 5- or 7-point demarcations that ask patients to rate their degree of improvement from baseline ranging from markedly better to markedly worse,

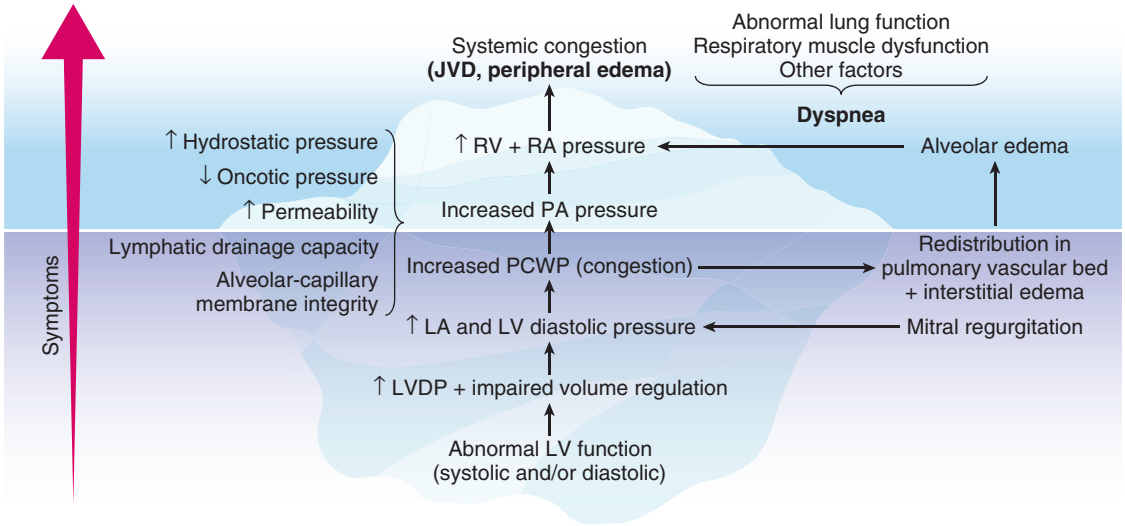

FIGURE 58-4. **Role of congestion in heart failure.** JVD = jugular venous distention; LA = left atrial; LV = left ventricular; LVDP = left ventricular diastolic pressure; PA = pulmonary artery; PCWP = pulmonary capillary wedge pressure; RA = right atrial; RV = right ventricular. (From Gheorghiade M, Follath F, Ponikowski P, et al. Assessing and grading congestion in acute heart failure: a scientific statement from the Acute Heart Failure Committee of the Heart Failure Association of the European Society of Cardiology and endorsed by the European Society of Intensive Care Medicine. *Eur J Heart Fail.* 2010;12:423-433.)

or a visual analog scale, which asks patients to rate their level of breathing difficulty on a vertical spectrum from 0 at the bottom to 100 at the top, with 100 being the best ability to breathe and 0 being the worst dyspnea. Improvement in dyspnea represents a major patient-reported outcome, and more severe dyspnea is associated with worse in-hospital and postdischarge outcomes.

When dyspnea occurs in the recumbent position, it is called *orthopnea*. This symptom is most commonly elicited by asking patients about their breathing while trying to lie flat during the night. Orthopnea results from the increase in venous return from the extremities and splanchnic circulation to the central circulation with changes in posture. The increase in ventricular preload raises pulmonary venous and pulmonary capillary hydrostatic pressures. Orthopnea is typically classified by an ordinal scale based on the number of pillows a patient requires to sleep comfortably without shortness of breath. Patients with prominent orthopnea may report an inability to sleep in a bed and instead may sleep in a recliner. Orthopnea is a specific symptom of heart failure, and it correlates well with the severity of pulmonary congestion.

Bendopnea is defined as severe dyspnea that occurs while bending over. The mechanism of bendopnea is poorly defined but appears to be related to increases in left ventricular filling pressures during bending in patients with a baseline elevation in pulmonary capillary wedge pressure.[9] There appears to be an association between bendopnea and baseline mismatch of left- and right-sided filling pressures (i.e., elevated wedge pressure out of proportion to right atrial pressure). The clinical spectrum of this symptom and its relationship to outcomes are not well understood, but recent data suggest that this symptom may be a more common than has previously been recognized.

Paroxysmal nocturnal dyspnea is acute, severe shortness of breath that wakes the patient from sleep. These symptoms should be distinguished from periods of apnea related to sleep-disordered breathing (Chapter 100). Paroxysmal nocturnal dyspnea usually is manifested about 1 hour after the patient goes to sleep and begins to subside shortly after awakening. Paroxysmal nocturnal dyspnea results from increased venous return and the mobilization of interstitial fluid from the splanchnic circulation and lower extremities, with accumulation of alveolar edema. Paroxysmal nocturnal dyspnea is relatively uncommon but almost always represents severe heart failure, and it appears to be associated with increased mortality.

Fatigue

Fatigue, which is one of the most common symptoms in heart failure, occurs in more than 90% of patients. Although fatigue is difficult to quantify and is not a specific symptom for heart failure, the severity of fatigue is associated with prognosis. As a result, clinicians should pay careful attention to this symptom as an occult manifestation of heart failure.

Chest Pain

Chest pain (Chapter 51) may be mediated by myocardial ischemia from underlying coronary artery disease, but also can occur in patients without obstructive coronary artery disease because of increased wall stress that is proportional to the degree of left ventricular dilation. Patients who have heart failure with a preserved ejection fraction and left ventricular hypertrophy may develop chest pain from a mismatch of oxygen supply and demand. Chest pain in amyloid heart disease results when deposition of amyloid protein in the medial layer of myocardial arterioles causes transient ischemia.

Cardiac Cachexia

Patients with heart failure can develop constitutional symptoms, including nausea, vomiting, anorexia, and diffuse abdominal pain. Muscle wasting is a frequent comorbidity among patients with advanced chronic heart failure.[10] In some patients, these symptoms can cause significant muscle mass and weight loss, termed cardiac cachexia, which is associated with a very poor prognosis. In many patients, these symptoms arise from prominent right-sided heart failure and resulting passive venous congestion in the abdominal vasculature or liver. In some patients, severe tricuspid regurgitation is a contributing factor. In patients with significant hepatic congestion, abdominal pain may be localized to the right upper quadrant, and jaundice may be observed. The group of patients most likely to experience these symptoms includes those with prominent right-sided heart failure. When patients present with this symptom complex including right upper quadrant tenderness due to hepatic congestion, these heart failure symptoms initially may be falsely attributed to gallbladder disease (Chapter 155) or other abdominal disease. In patients with advanced cardiogenic shock (Chapter 107), severe abdominal pain can be a particularly ominous sign of abdominal ischemia (Chapter 143).

Cognitive Dysfunction and Mood Disorders

Cognitive dysfunction (Chapter 402) is common, particularly in elderly heart failure patients. Confusion can be a manifestation of worsening heart failure related to relative hypotension precipitated by medications used to treat heart failure or of a specific complication of an individual drug, such as a β-blocker. Although intrinsic brain function itself is not affected in most patients with heart failure, cerebral hypoperfusion in advanced heart failure can cause memory impairment, limited attention span, and altered mentation.

Depressive symptoms (Chapter 397) occur in up to 25% of patients with heart failure. Depressive symptoms can be detected by simple questions such as the Patient Health Questionnaire depression module (see Chapter 24), which has an 80% predictive value for depression and is associated with worse outcomes in heart failure patients.

Sleep Disorders

Sleep-disordered breathing is observed in upward of 70% of heart failure patients and is associated with increased morbidity and mortality. Patterns of sleep-disordered breathing include obstructive sleep apnea (Chapter 100) and central sleep apnea/Cheyne-Stokes respiration (Chapter 86). Patients may have both types, and the relative proportion of each type varies with the severity of heart failure and its treatment. Nocturnal rostral fluid movement

from the lower extremities of heart failure patients may worsen obstructive sleep apnea. Central sleep apnea is due in part to the instability of the ventilatory control systems in heart failure.

Chronic sleep-disordered breathing also causes a series of derangements that may precipitate or exacerbate heart failure. Sleep-disordered breathing increases blood pressure and the risk of arrhythmias.

Symptoms of sleep-disordered breathing include hypersomnolence, choking or gasping during sleep, recurrent awakenings, unrefreshing sleep, daytime fatigue, and impaired concentration or memory. The symptoms may be difficult to distinguish from other symptoms of heart failure, including unrefreshing sleep due to orthopnea and paroxysmal nocturnal dyspnea. Preliminary data suggest that attention to the diagnosis and management of sleep-disordered breathing with positive-pressure ventilation improves quality of life in some patients with heart failure.

Physical Examination

A carefully performed physical examination is critical to make an accurate diagnosis, to assess possible additional comorbid conditions, and to begin to estimate prognosis.

Global Observation and Vital Signs

The patient's general appearance may provide details related to the acuity and severity of the heart failure. Patients with severe symptoms may be pale or diaphoretic and unable to speak in complete sentences. In extreme circumstances, they may be unable to lie recumbent in bed because of severe dyspnea or pulmonary edema. The heart rate may be elevated (>100 beats per minute), and premature ventricular beats or atrial arrhythmias are common. Approximately 30% of heart failure patients have atrial fibrillation (Chapter 64). Pulsus alternans (alternating amplitude of successive beats) is an uncommon sign but is virtually diagnostic for advanced heart failure. Blood pressure is most commonly normal or high, but it may be low (systolic blood pressure <90 mm Hg) in advanced low-output heart failure. Blood pressure has historically been identified as an important prognostic marker (i.e., higher blood pressure is associated with improved long-term outcomes), but these associated data may not be applicable to all subgroups of patients, especially the elderly. A narrow pulse pressure (e.g., <30 to 35 mm Hg) also indicates more severe heart failure. Weight should be assessed and compared with the patient's known dry weight or recent weight trajectory. The assessment of weight not only helps the clinician to appreciate the severity of volume overload, but it may also assist with quantifying the degree of cardiac cachexia. The respiratory rate should be measured. Both low rates and high rates may be seen in heart failure. Likewise, both hypothermia and hyperthermia can be informative of impending shock or secondary causes of heart failure.

Jugular Veins

Examination of the jugular veins is a critical part of the heart failure physical examination (Chapter 51) both initially and serially.[11] The patient should be positioned in the partially recumbent position with his or her neck turned to the left. The patient's head should be rested on a pillow to limit tension in the neck muscles, which could obscure visualization of the venous pulsations (see Fig. 51-2). The ideal method for measurement is quantifying in centimeters of water (normal = 8 cm H_2O) and estimating the level of pulsations above the sternal angle (and adding 5 cm H_2O; see Fig. 51-3). To distinguish the venous pulsations from the carotid pulsations, clinicians should look for a double pulsation in the venous waveform and can compare the timing with the arterial pulsation in the wrist. If the top of the jugular venous pulsation cannot be appreciated in the initial position, it may be necessary to reposition the patient to see the peak of the pulsation. For instance, in some circumstances, it may be necessary to place the patient in the upright position to visualize the peak of the pulsation near the tragus. The presence of abdominal-jugular reflux should be assessed by putting sustained pressure on the abdomen for 30 seconds; a positive finding is at least a 1-cm rise in the jugular pressure, which then slowly declines when pressure is removed. These findings are a sign of abnormally elevated right ventricular filling pressures. Either an elevated jugular venous pressure or an abnormal abdominal jugular reflux has been reported in 80% of patients with advanced heart failure. No other simple sign is nearly as sensitive.

An additional important finding in the neck is evidence of tricuspid regurgitation, which is visualized as a large *cv* wave (see Fig. 51-4). This finding is confirmed by hepatic pulsations, which can be detected during the abdominal-jugular reflux determination. The carotid pulses should be evaluated for evidence of aortic stenosis (see Fig. 51-5), and thyroid abnormalities should be sought.

Pulmonary Examination

Despite having an elevated left ventricular filling pressure that is transmitted back into the left atrium and pulmonary vasculature, most patients with compensated heart failure do not have evidence of pulmonary congestion on their physical examination. The lungs of chronic heart failure patients undergo adaptive changes and have robust lymphatic drainage to compensate for elevated filling pressures. However, a subset of patients may develop alveolar fluid accumulation, which is appreciated as rales or "crackles" on the clinical examination (Chapter 83). These findings are more common in patients with acute pulmonary edema due to sudden decompensation from an inciting event, such as ischemia or worsening hypertension.

Fluid may also accumulate in the pleural space related to the increased transudation of fluid and impaired lymphatic drainage in the setting of elevated systemic venous pressures (Chapter 99). Pericardial effusions (Chapter 77) may also occur in the setting of heart failure, particularly in inflammatory cardiomyopathy, but this pattern of fluid accumulation is relatively uncommon overall. The clinician should listen and percuss for the possible presence of pleural effusions (Chapter 99), which tend to lateralize to the right side owing to the greater surface area of the lungs and the position of the diaphragm. The diagnosis of a moderate or large pleural effusion is critical because draining of the effusion may represent an important intervention to relieve dyspnea.

Cardiac Examination

The cardiac examination is the cornerstone of the evaluation of the patient with heart failure. Visual inspection may reveal a right ventricular heave, which provides information on right ventricular dysfunction and underlying pulmonary hypertension. Palpation of the location, size, and duration of the point of maximal impulse against the chest wall may provide details related to the degree of left ventricular dilation; the impulse is typically laterally displaced in the setting of ventricular enlargement and may be sustained in the setting of left ventricular hypertrophy.

Auscultation of the heart sounds provides important information related to the underlying rhythm and frequency of ectopic beats. Disorders such as pulmonary hypertension can be appreciated on the basis of an increase in the intensity of the second heart sound. The presence (or worsening) of valvular disorders (Chapter 75) can be characterized for their potential contribution to cardiac dysfunction. An apical S_3 gallop (see Fig. 51-6 in Chapter 51) is common in severe LV dysfunction, and its presence is correlated with an elevated left ventricular end-diastolic pressure and a poor prognosis. As patients are treated for volume overload, the intensity of and ability to detect an S_3 gallop may diminish. The S_4 gallop is common in patients who have ischemic heart disease and hypertension, and it is more likely indicative of diastolic dysfunction.

Abdomen

The physical examination should estimate the size of the liver and spleen and elicit the presence of ascites (Chapter 146). Intra-abdominal hypertension and abdominal venous congestion may cause renal venous hypertension and subsequent renal dysfunction. An enlarged and pulsatile liver is seen in individuals with markedly elevated right heart pressures and in patients with significant tricuspid insufficiency. Hepatic enlargement and dysfunction represent an important step in determining the most appropriate timing of intervention on a regurgitant tricuspid valve. In the setting of irreversible liver disease (i.e., cardiac cirrhosis), heart failure patients are at a substantially increased risk during surgical interventions. Thus, a thorough abdominal examination represents a critical component of the evaluation of heart failure patients.

Extremities

Edema (Chapter 51) results from the retention of sodium due to low cardiac output and reduced renal perfusion pressures, which ultimately result in elevated right-sided filling pressures, increased hydrostatic pressures in the venous circulation, and transudation of fluid into dependent interstitial spaces, especially in the ankles or lower extremities. Edema is commonly measured on a scale of 0 to 3+, but this system has marked interobserver variation. Peripheral edema is a nonspecific finding, and edema due to heart failure must be distinguished from edema related to medication use (e.g., calcium-channel blockers, thiazolidinediones, or nonsteroidal anti-inflammatory drugs), venous insufficiency, or hypoproteinemia.

The temperature of the extremities should also be assessed. Cool extremities suggest low cardiac output or concomitant peripheral arterial disease (Chapter 79).

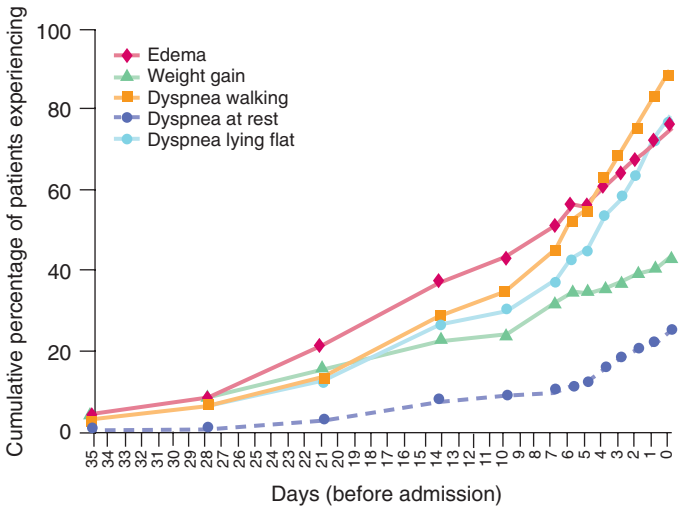

FIGURE 58-5. **Number of days from onset of worsening of selected symptoms of heart failure to admission to the hospital: cumulative percentage of patients. (From Schiff GD, Fung S, Speroff T, et al. Decompensated heart failure: symptoms, patterns of onset, and contributing factors. *Am J Med.* 2003;114:625-630.)**

DIAGNOSIS

Patterns of Presentation

The initial presentation ranges from subtle outpatient findings to acute decompensation that requires hospitalization. An initial presentation may represent the gradual progression of known but previously asymptomatic (stage A or B) heart failure or be the first indication of altered cardiac function. In patients who do not have an antecedent history of heart failure, precipitating factors, such as acute myocardial infarction (Chapter 73), tachyarrhythmias (Chapters 64 and 65), previously unrecognized or new valvular abnormalities (Chapter 75), toxic damage (including alcohol excess), or acute myocarditis (Chapter 60), should be considered.

When patients with stage C or D heart failure present with worsening symptoms, precipitating factors may also include myocardial ischemia, arrhythmias, or worsening of valvular function (see Table 58-2). Other conditions can include anemia, infection, hyperthyroidism, and any conditions that stimulate an increase in cardiac output. In many patients, however, the worsening may be gradual, augured by a sometimes subtle increase in outpatient signs and symptoms (Fig. 58-5).

Electrocardiography

A 12-lead electrocardiogram (Chapter 54) should be obtained for all patients who present with possible heart failure. The major importance of the electrocardiogram is to evaluate the cardiac rhythm, to identify current ischemia or prior myocardial infarction, and to detect evidence of left ventricular hypertrophy. Rhythm abnormalities may be responsible for the development or exacerbation of underlying cardiac dysfunction. For example, an underlying tachyarrhythmia can lead to the development of left ventricular systolic dysfunction that is reversible with appropriate intervention. Q waves suggest coronary artery disease as a likely contributor to ventricular dysfunction. The presence of voltage criteria for left ventricular hypertrophy supports a diagnosis of hypertensive heart disease, including heart failure with a preserved ejection fraction. Underlying conduction abnormalities, such as delayed ventricular conduction (i.e., bundle branch morphology), determine eligibility criteria for cardiac resynchronization therapy (Chapters 59 and 66) and have important prognostic implications. Holter monitoring sometimes may be helpful to determine the burden of ventricular arrhythmias or ectopic beats because tachycardia-mediated cardiomyopathies may be reversible with medical therapy or ablation therapy.

Chest Radiography

Patients with suspected new-onset or worsening heart failure should undergo chest radiography to assess heart size and pulmonary congestion as well as to detect other diseases that may cause or contribute to the patient's symptoms. Many patients with acute heart failure but only a minority of those with chronic heart failure have clear evidence of pulmonary venous hypertension (upper lobe redistribution, enlarged pulmonary veins) or pulmonary edema (perihilar or patchy peripheral infiltrates; see Fig. 56-2). Pleural effusions, usually on the right if unilateral but often bilateral, are also identified on the chest radiograph (see Fig. 99-3). Chest radiography may also play a role in identifying the location of the lead placement of intracardiac devices, such as biventricular pacemakers. Inappropriate lead placement may be observed in patients with worsening heart failure symptoms.

Laboratory Testing

The initial evaluation of patients presenting with heart failure should include a complete blood count to detect anemia and systemic diseases with hematologic manifestations; urinalysis and tests of renal function to assess renal states; serum electrolyte values to identify abnormalities needing treatment and to provide a baseline for subsequent therapy; glucose level and lipid profile to diagnose diabetes and dyslipidemia, which should be carefully managed in patients with heart failure; and thyroid-stimulating hormone level. Markers of hepatic congestion, such as elevated serum aminotransferase and bilirubin levels (Chapter 147), also should be measured because they are important prognostic signs in patients with heart failure. Screening for hemochromatosis (Chapter 212) or human immunodeficiency virus (HIV) infection is reasonable in selected patients with heart failure. Diagnostic tests for rheumatologic diseases (Chapter 256), amyloidosis (Chapter 188), or pheochromocytoma (Chapter 228) are not routinely indicated but rather should be targeted to patients with other ancillary findings suggestive of these conditions. Viral antibody titers yield relatively little incremental information and are rarely indicated in the evaluation of heart failure.

Natriuretic Peptides

Brain natriuretic peptide (BNP) and its amino-terminal fragment (NT-proBNP) provide incremental diagnostic and prognostic information above and beyond the history and physical examination in patients with heart failure. A BNP level should be measured to support the diagnosis of heart failure in ambulatory patients with dyspnea as well as in patients with possible acute heart failure, especially in the setting of an uncertain diagnosis. It also is useful to estimate the severity of heart failure and its prognosis.

Although BNP levels are relatively sensitive and specific markers for clinically confirmed heart failure, circulating levels are influenced by co-morbid processes. For example, obesity reduces BNP levels, whereas advancing age and renal dysfunction are associated with higher levels. Most heart failure therapies reduce BNP levels, but the usefulness of BNP-guided heart failure therapy is not well established.

Troponin

Owing to the increased sensitivity of currently available troponin assays, the majority of patients admitted with acute heart failure have elevations in circulating troponin even without any obvious myocardial ischemia. These elevations, which suggest ongoing myocyte injury or necrosis, are associated with worse clinical outcomes and mortality.[12]

Other Biomarkers: Galectin-3 and ST2

A number of additional biomarkers characterize inflammation, myocyte injury, neurohormonal upregulation, and extracellular matrix turnover in patients with heart failure (E-Table 58-1 and E-Fig. 58-1). For example, biomarkers of myocardial fibrosis, including soluble ST2 and galectin-3, are associated with hospitalization and death in patients with heart failure. In the future, strategies that combine multiple biomarkers into a risk stratification model may prove additive to clinical judgment.[13]

Echocardiography

An echocardiogram should be obtained during the initial evaluation of patients with heart failure to assess ventricular and valve function. Repeated echocardiograms are also indicated when patients have a significant change in their clinical status or receive treatment that may have had a significant effect on cardiac function. In contrast, routine repeated measurements of left ventricular function in the absence of a change in clinical status or treatment should not be performed.

Echocardiography (Chapter 55) allows the assessment of left ventricular systolic and diastolic function (see Figs. 55-2 and 55-3). Wall thickness, ventricular dilation, and regional wall motion abnormalities provide evidence of the underlying etiology and chronicity of heart failure. Right ventricular failure, which is associated with worse prognosis, can also be evaluated to assess the relative contribution of right-sided dysfunction. Echocardiography also evaluates valvular dysfunction (Chapter 75), which may be the result of or cause of worsening ventricular function. Quantitative measurements of

E-TABLE 58-1 EMERGING BIOMARKERS IN CHRONIC HEART FAILURE

BIOMARKER	PROPOSED PATHOPHYSIOLOGY	POTENTIAL ROLE IN PATIENTS WITH CHRONIC HF
Copeptin	Stable plasma surrogate for vasopressin, which is released from the hypothalamus in response to changes in plasma osmolarity and reduced cardiac output Elevated levels of vasopressin contribute to development of hyponatremia, vasoconstriction, and adverse cardiac remodeling	Independent role as a prognostic marker in patients with chronic HF Potential role of molecular marker for tailored therapies with vasopressin antagonism
Cystatin C	Cysteine protease inhibitor that is more sensitive and specific to changes in GFR than creatinine Produced at a constant rate by all nucleated cells of the body	Independent predictor of mortality in patients with chronic HF Identifies poor prognosis even in patients with a normal creatinine level, pointing to a role as a marker of the CRS
Endothelin 1	Produced by the endothelium in response to ATII, inflammation, and vascular shear stress Causes vasoconstriction and adverse cardiac remodeling	Independent predictor of mortality in patients with chronic HF Clinical application requires measurement of the more stable precursor (CT-proET-1) Whether patients with elevated levels of CR-proET-1 would benefit from endothelin 1 antagonism is unclear
Galectin-3	β-Galactoside-binding lectin produced by several tissues Promotes cardiac fibroblast proliferation and collagen synthesis (maladaptive remodeling)	Predicts mortality independent of natriuretic peptides in patients with chronic HF Does not appear to be modified by treatment Potential role as a target for therapy
GDF-15	Member of the TGF-β cytokine family Expressed in most tissues Expression in myocytes is triggered by ischemia, stretch, inflammation, and neurohormonal activation May protect myocytes against apoptosis and hypertrophy	Predictive of mortality in patients with chronic HF independent of established biomarker Serial changes are associated with prognosis Could lead to the development of unique treatments for patients with chronic HF
hs-cTn	Proteins involved in the regulation of cardiac and skeletal muscle contraction Released after loss of myocytes from necrosis, apoptosis, or reversible injury with increased membrane permeability	Predictive of chronic HF mortality in patients with no known CHD Independently predictive of mortality in patients with chronic HF The effect of serial changes in hs-cTn levels on prognosis, and how therapy should be modified on the basis of increased levels of hs-cTn, is unknown.
MR-proADM	Produced by various tissues within the cardiovascular system in response to hemodynamic stress Favorable effects on the vasculature—positive inotropy and vasodilation	Could add prognostic value to predictive models for patients with chronic HF beyond that provided by natriuretic peptides and traditional clinical risk factors
Neuregulin 1	Belongs to a family of growth factors with an important role in cardiac development and the pathogenesis of chronic HF Has an important role in promoting cardiomyocyte growth in function and regulation of the stress response	Preliminary studies show independent associations with disease severity and risk of adverse outcomes in patients with chronic HF A promising target for drug therapy Trials with recombinant neuregulin are already under way, with a phase II trial showing improvement in cardiac structure and function.
NGAL	Small glycoprotein released by multiple cell types during inflammation and injury Involved in cell survival, inflammation, and matrix degradation Early marker of renal tubular damage Increased levels found in patients with chronic HF	Early detection of renal injury in patients with chronic HF A more effective marker of CRS than creatinine Independent predictor of mortality in patients with chronic HF
sST2	Member of the IL-1 receptor family that is secreted from myocytes during biomechanical strain Possible mediator of myocardial hypertrophy and fibrosis by compromising the favorable effects of IL-33/ST2 signaling	Independent predictor of morbidity and mortality in patients with chronic HF Possibly role as a novel therapeutic target

From Ahmad T, Fiuzat M, Felker GM, et al. Novel biomarkers in chronic heart failure. *Nat Rev Cardiol.* 2012;9:347-359.

ATII = angiotensin II; CHD = coronary heart disease; CRS = cardiac-renal syndrome; CT-proET-1 = C-terminal pro-endothelin 1; GDF = growth differentiation factor; GFR = glomerular filtration rate; HF = heart failure; hs-cTn = cardiac troponins measured by a highly sensitive assay; IL = interleukin; MR-proADM = mid-regional pro-adrenomedullin; NGAL = neutrophil gelatinase-associated lipocalin; TGF = transforming growth factor.

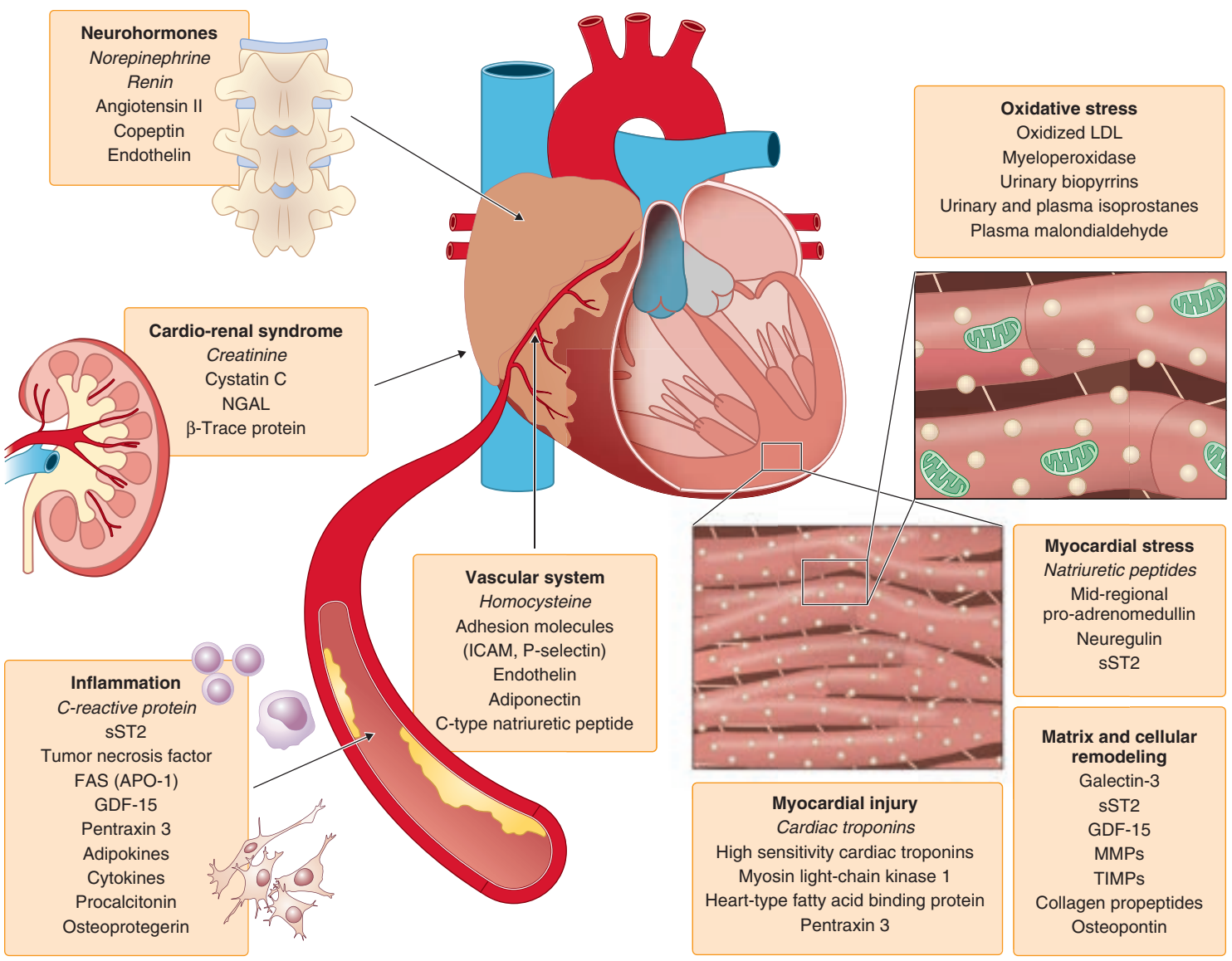

E-FIGURE 58-1. Biomarkers in heart failure. Heart failure involves interplay between myocardial factors, systemic inflammation, renal dysfunction, and neurohormonal activation. Biomarkers can be classified according to broad categories of processes that are involved in the development and progression of chronic heart failure. Several biomarkers have been characterized for each of these processes; these vary greatly in their ease of measurement, cost, turnaround time, and evaluation in the clinical setting. Traditional biomarkers that have been studied fairly rigorously appear in *italics*. APO = apoptosis antigen; GDF = growth differentiation factor; ICAM = intercellular adhesion molecule; LDL = low-density lipoprotein; MMPs = matrix metalloproteinases; NGAL = neutrophil gelatinase-associated lipocalin; TIMPs = matrix metalloproteinase tissue inhibitors. (From Ahmad T, Fiuzat M, Felker GM, et al. Novel biomarkers in chronic heart failure. *Nat Rev Cardiol.* 2012;9:347-359.)

pulmonary artery pressure and central venous pressure help characterize the degree of pulmonary hypertension and may guide diuretic therapies in circumstances in which the jugular veins are difficult to visualize. The presence of an atrial or ventricular thrombus requires anticoagulation. Novel methods using ventricular strain analysis and three-dimensional echocardiography can provide more detailed information about ventricular dyssynchrony and compliance and may, in the future, prove useful in patients with heart failure with a preserved ejection fraction.

Nuclear Cardiology and Coronary Angiography

When myocardial ischemia may be contributing to heart failure, coronary arteriography (Chapter 57) is reasonable to assess eligibility for revascularization. The most powerful predictors of prognosis for patients with ischemic cardiomyopathy are a history of prior myocardial infarction or revascularization, stenosis of 75% or greater of the left main or proximal left anterior descending artery, stenosis of 75% or greater of two or more epicardial vessels, and the severity of ventricular dysfunction. Coronary computed tomographic angiography (Chapter 56) may represent a non-invasive modality for assessing coronary disease in appropriately selected patients.

Noninvasive imaging (Chapter 56) to detect myocardial ischemia and viability is reasonable in patients who present with de novo heart failure and in patients who have known coronary artery disease and no angina, unless they are not eligible for revascularization. Viability assessment is reasonable in select situations in planning revascularization for patients who have heart failure and coronary artery disease. However, stress testing to assess myocardial viability so far has not been able to identify patients who will benefit from revascularization compared with medical therapy alone.

The single-photon emission computed tomography tracer *m*-iodobenzylguanidine (mIBG) has been widely used for studying causes and effects of cardiac sympathetic hyperactivity. Cardiac sympathetic imaging with mIBG is a noninvasive tool that may assist with the risk stratification of patients with heart failure. With mIBG imaging, the myocardial uptake and distribution can be visually assessed and quantified by calculating a heart-to-mediastinum ratio. This approach provides a highly reproducible index of cardiac sympathetic activity. Further study is required to determine the role of this imaging modality in heart failure risk stratification and clinical care.

Cardiac Magnetic Resonance Imaging

Cardiac magnetic resonance imaging (Chapter 56), which provides accurate data on left ventricular volume and ejection fraction, can be useful when echocardiography is inadequate. It is also helpful to assess for potential infiltrative cardiomyopathies when the cause of heart failure, especially heart failure with a preserved ejection fraction, is unclear. Late gadolinium enhancement adds important prognostic information related to ventricular arrhythmia and mortality risk (see Fig. 56-21).

Myocardial Biopsy

Guidelines indicate that endomyocardial biopsy should not be performed in the routine evaluation of patients with heart failure. However, endomyocardial biopsy may be useful in patients who present with heart failure when a specific suspected diagnosis would influence therapy. For example, in patients with acute myocarditis[14] or giant cell myocarditis and in patients with sarcoid or amyloid cardiomyopathy (Chapter 60), the appropriate pathologic diagnosis may inform treatment recommendations and prognosis. In certain circumstances, biopsy may be performed along with genetic testing (e.g., transthyretin gene mutation) to inform decisions on management and counseling of family members.

Assessment of Exercise Capacity

A heart failure patient's exercise capacity can be quantified by several testing modalities, including 6-minute walk distance and cardiopulmonary exercise testing (Chapter 85). Although these tests are not routinely recommended, they are helpful for determining the relative contribution of cardiac compared with pulmonary causes of functional limitation. The results of cardiopulmonary exercise testing are critical to determine the severity of disease in patients who are being considered for therapies such as heart transplantation or ventricular assist device placement (Chapter 82). A maximal oxygen consumption of less than 14 mL/kg/minute is associated with a poor enough prognosis that survival is probably better with transplantation or implantation of a left ventricular assist device compared with medical therapy. Serial cardiopulmonary exercise testing or 6-minute walk testing also can be useful to follow the disease course of specific patients objectively.

Invasive Diagnostics and Hemodynamic Monitoring

Invasive monitoring may be useful in selected patients who have acute heart failure, who have persistent symptoms despite adjustment of standard therapies, and whose fluid status or perfusion is uncertain. However, the routine use of pulmonary artery catheters should be discouraged.

Up to 30% of patients with heart failure with a reduced ejection fraction have an implantable device that detects arrhythmias and can provide hemodynamic assessment. Indirect measures, such as changes in impedance or heart rate variability, are precursors to worsening heart failure symptoms. Chronic hemodynamic monitors (i.e., direct measures of pulmonary pressures and right ventricular pressures) have been approved by the U.S. Food and Drug Administration and are now available for the serial measurement of left ventricular filling pressures. The interrogation of these devices may become part of the routine diagnostic follow-up to guide therapy in patients who have respiratory distress or clinical evidence of impaired perfusion in whom the adequacy or excess of intracardiac filling pressures cannot be determined from clinical assessment.

CONSEQUENCES OF MISDIAGNOSIS

The diagnosis of heart failure may be straightforward in a patient with dyspnea, signs of congestion, and elevated BNP level. Conversely, many patients have multiple comorbid conditions that make the assessment of shortness of breath more of a diagnostic dilemma. Pulmonary conditions such as chronic obstructive pulmonary disease (Chapter 88) represent the most common reason for misdiagnosis. The BNP level and echocardiogram results are useful in this situation. Other potential causes of edema or volume overload include renal failure (Chapter 131), venous thrombosis (Chapters 81 and 98), and venous insufficiency. If left ventricular systolic function is normal, it may be difficult to make a conclusive determination of the relative role of heart failure with preserved ejection fraction compared with other concomitant conditions, such as severe obesity, deconditioning, chronic anemia, or other systemic illnesses. BNP levels may be helpful in some circumstances. In other situations, exercise testing or invasive hemodynamic testing may be necessary to establish the appropriate diagnosis. Misdiagnosis can result in excessive and unnecessary diagnostic testing, higher costs, and increased morbidity and mortality because of the inappropriate use or nonuse of heart failure therapies.

Grade A References

A1. O'Connor CM, Starling RC, Hernandez AF, et al. Effect of nesiritide in patients with acute decompensated heart failure. *N Engl J Med*. 2011;365:32-43.

A2. Bonow RO, Maurer G, Lee KL, et al. Myocardial viability and survival in ischemic left ventricular dysfunction. *N Engl J Med*. 2011;364:1617-1625.

A3. Binanay C, Califf RM, Hasselblad V, et al. Evaluation study of congestive heart failure and pulmonary artery catheterization effectiveness: the ESCAPE trial. *JAMA*. 2005;294:1625-1633.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Huffman MD, Berry JD, Ning H, et al. Lifetime risk for heart failure among white and black Americans: cardiovascular lifetime risk pooling project. *J Am Coll Cardiol*. 2013;61:1510-1517.
2. Go AS, Mozaffarian D, Roger VL, et al. Executive summary: heart disease and stroke statistics—2014 update: a report from the American Heart Association. *Circulation*. 2014;129:399-410.
3. Basuray A, French B, Ky B, et al. Heart failure with recovered ejection fraction: clinical description, biomarkers, and outcomes. *Circulation*. 2014;129:2380-2387.
4. Mentz RJ, Bittner V, Schulte PJ, et al. Race, exercise training, and outcomes in chronic heart failure: findings from Heart Failure—A Controlled Trial Investigating Outcomes in Exercise TraiNing (HF-ACTION). *Am Heart J*. 2013;166:488-495.
5. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2013;62:e147-e239.
6. Ahmad T, Pencina MJ, Schulte PJ, et al. Clinical implications of chronic heart failure phenotypes defined by cluster analysis. *J Am Coll Cardiol*. 2014;64:1765-1774.
7. Braunwald E. Heart failure. *JACC Heart Fail*. 2013;1:1-20.
8. Gupta DK, Shah AM, Castagno D, et al. Heart failure with preserved ejection fraction in African-Americans—the Atherosclerosis Risk in Communities (ARIC) study. *JACC Heart Fail*. 2013;1:156-163.
9. Thibodeau JT, Turer AT, Gualano SK, et al. Characterization of a novel symptom of advanced heart failure: bendopnea. *JACC Heart Fail*. 2014;2:24-31.
10. Fülster S, Tacke M, Sandek A, et al. Muscle wasting in patients with chronic heart failure: results from the studies investigating co-morbidities aggravating heart failure (SICA-HF). *Eur Heart J*. 2013;34:512-519.
11. Drazner MH, Hellkamp AS, Leier CV, et al. Value of clinician assessment of hemodynamics in advanced heart failure: the ESCAPE trial. *Circ Heart Fail*. 2008;1:170-177.
12. Gravning J, Askevold ET, Nymo SH, et al. Prognostic effect of high-sensitive troponin T assessment in elderly patients with chronic heart failure: results from the CORONA trial. *Circ Heart Fail*. 2014;7:96-103.
13. Ahmad T, Fiuzat M, Neely B, et al. Biomarkers of myocardial stress and fibrosis as predictors of mode of death in patients with chronic heart failure. *JACC Heart Fail*. 2014;2:260-268.
14. Caforio AL, Pankuweit S, Arbustini E, et al. Current state of knowledge on aetiology, diagnosis, management, and therapy of myocarditis: a position statement of the European Society of Cardiology Working Group on Myocardial and Pericardial Diseases. *Eur Heart J*. 2013;34:2636-2648.

REVIEW QUESTIONS

1. Which of the following statements is true regarding heart failure epidemiology?

A. White patients have an increased incidence of heart failure compared with black patients.
B. Black patients are at an increased risk for rehospitalization for heart failure.
C. Women have better long-term survival after the diagnosis of heart failure compared with men.
D. The heart failure with preserved ejection fraction population represents approximately a third of overall heart failure cases.
E. The absolute mortality rate after the diagnosis of heart failure is approximately 25% at 5 years.

Answer: B Black patients are at an increased risk for development of heart failure and have an increased risk for heart failure rehospitalization compared with white patients. Median survival is lower in women than in men. Heart failure with preserved ejection fraction represents approximately half of heart failure cases. The absolute mortality rate after heart failure diagnosis is approximately 50% at 5 years.

2. Which is the correct statement regarding guideline recommendations for the diagnosis of heart failure?

A. Routine repeated measurements of left ventricular function in the absence of a clinical status change are reasonable.
B. Measurement of natriuretic peptides is recommended to support the diagnosis of heart failure in ambulatory patients with dyspnea as well as in those with possible acute heart failure.
C. There is an established role for routine or periodic invasive hemodynamic measurements in the management of heart failure.
D. Endomyocardial biopsy should be performed in the routine evaluation of patients with heart failure, given the need to diagnose the underlying etiology.
E. On cardiopulmonary exercise testing, a peak oxygen uptake of more than 20 mL/kg/minute is associated with a relatively poor prognosis.

Answer: B Routine repeated measurements of left ventricular function in the absence of a clinical status change or treatment intervention should not be performed. An echocardiogram should be obtained during the initial evaluation of patients with heart failure to assess ventricular function and valve function. Measurement of natriuretic peptides is recommended to support the diagnosis of heart failure in ambulatory patients with dyspnea as well as in those with possible acute heart failure. Routine or periodic invasive hemodynamic measurements in the management of heart failure are not recommended. Guidelines indicate that endomyocardial biopsy should not be performed in the routine evaluation of patients with heart failure. However, endomyocardial biopsy may be useful in patients presenting with heart failure when a specific diagnosis is suspected that would influence therapy. A peak oxygen uptake of less than 14 mL/kg/minute is associated with a relatively poor prognosis.

3. Which of the following statements is correct regarding the history and physical examination in heart failure patients?

A. Sleep-disordered breathing has a similar prevalence in heart failure patients and in the general population.
B. Lower extremity edema is a relatively specific sign of underlying heart failure.
C. Rales on the pulmonary examination are common in chronic heart failure patients because of elevated ventricular filling pressures.
D. An apical S_3 gallop is a strong indicator of left ventricular systolic dysfunction and elevated left ventricular filling pressures.
E. The blood pressure is most commonly low (systolic blood pressure <90 mm Hg) in chronic heart failure patients.

Answer: D An apical S_3 gallop is a strong indicator of left ventricular systolic dysfunction and elevated left ventricular filling pressures. Sleep-disordered breathing is prevalent in heart failure patients and is associated with increased morbidity and mortality. Upward of 70% of heart failure patients have sleep-disordered breathing. The majority of patients with heart failure have normal findings on pulmonary examination. Despite having an elevation in left ventricular filling pressures that is transmitted back into the left atrium and pulmonary vasculature, most patients with compensated heart failure do not have evidence of pulmonary congestion. In heart failure patients, the blood pressure is most commonly normal or high, but it may be low (systolic blood pressure <90 mm Hg) in advanced low-output heart failure. Orthopnea (not lower extremity edema) is a relatively specific sign of heart failure.

59

HEART FAILURE: MANAGEMENT AND PROGNOSIS

JOHN J. V. MCMURRAY AND MARC A. PFEFFER

EVALUATION AND MANAGEMENT OF HEART FAILURE

Heart failure is an overarching term for a syndrome (i.e., a constellation of signs and symptoms) that encompasses a vast spectrum of cardiovascular disorders and is associated with a greatly heightened risk for death and nonfatal adverse cardiovascular events (Chapter 58). Treatment is initially directed toward prevention of cardiac injury (e.g., due to hypertension or

myocardial infarction) or toward limiting structural progression if cardiac damage has already occurred (e.g., left ventricular remodeling with declining left ventricular ejection fraction) and delaying the development of symptomatic heart failure. When symptoms develop, treatments are also directed at improving functional status as well as prognosis.

Approximately one in five adults will develop heart failure. In the United States, 5.8 million people have heart failure, and U.S. hospitals annually admit 1.0 million patients with a primary diagnosis of heart failure. The estimated cost of heart failure in the United States is about $24 billion per year. Randomized controlled clinical trials (RCTs) supply the framework for quantifying what different therapeutic approaches can offer. Even when they are definitive, RCTs only generate data about average risks and benefits of the tested therapeutic option in a selected cohort. Because an individual patient's responses can only be implied from the overall estimated group responses, RCTs cannot definitively direct the approach of every patient or answer the myriad questions that confront the practitioner regarding the specific circumstances of the patient. Another major limitation of RCTs is the relatively narrow time frame of observation, generally only months to several years, compared with epidemiologic experiences during decades. Despite these limitations, RCTs are the premier tool of evidence-based medicine, and the field of heart failure has fortunately been the focus of relatively high-quality RCTs that have provided robust evidence to improve clinical care and prognosis (Table 59-1 and E-Table 59-1). Indeed, the implementation of evidence from RCTs into clinical practice has resulted in impressive temporal improvements in survival after discharge from a first hospital admission for heart failure. Moreover, the age at which symptomatic heart failure first becomes evident has increased. Despite these tangible advances, heart failure continues to be a leading cause of morbidity and mortality in elderly people.

STAGES OF HEART FAILURE

The American Heart Association/American College of Cardiology Guidelines for the Evaluation and Management of Chronic Heart Failure in the Adult use a staging classification to underscore the evolution and progression of heart failure severity (Fig. 59-1).[1] This classification emphasizes the use of

TABLE 59-1 THERAPIES OF PROVEN BENEFIT IN HEART FAILURE*

Angiotensin-converting enzyme inhibitors
Angiotensin receptor blockers
β-Blockers
Mineralocorticoid receptor antagonists
Sacubitril-valsartan
Hydralazine-isosorbide dinitrate
Ivabradine
Digitalis
Cardiac resynchronization therapy
Cardioverter-defibrillator
Ventricular assist device
Exercise training

*See E-Table 59-1 for more details.

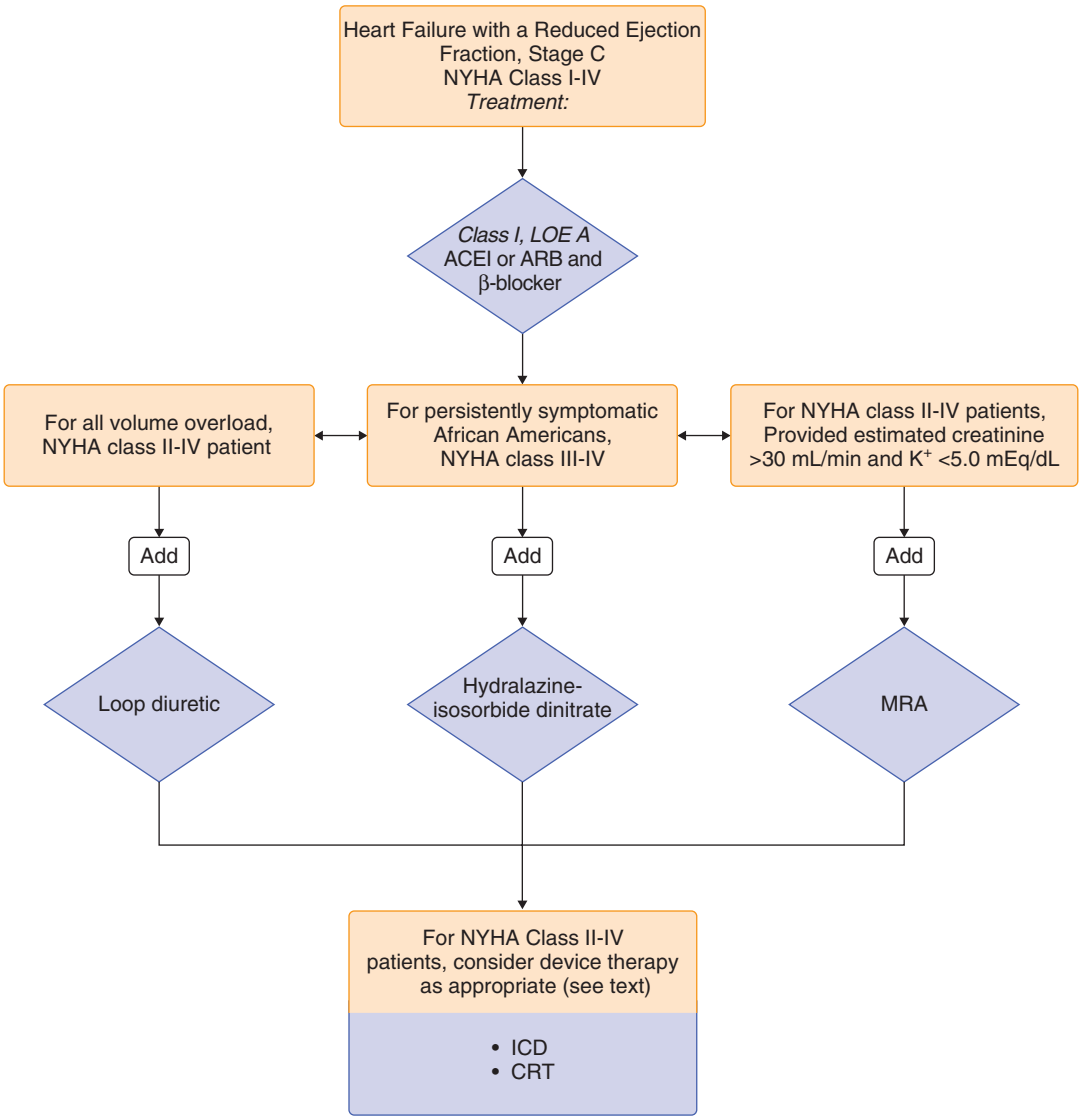

FIGURE 59-1 General approach to heart failure. Stage C heart failure with a reduced ejection fraction. Evidence-based, guideline-directed medical therapy. ACEI = angiotensin-converting enzyme inhibitor; ARB = angiotensin-receptor blocker; CRT = cardiac resynchronization therapy; ICD = implantable cardiovascular-defibrillator; MRA = mineralocorticoid receptor antagonist; NYHA = New York Heart Association. (Adapted from Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. *Circulation.* 2013;128:e240-327.)

E-TABLE 59-1 CONTROLLED TRIALS* IN SYMPTOMATIC HEART FAILURE WITH REDUCED SYSTOLIC FUNCTION

TRIAL, TREATMENT, AND YEAR PUBLISHED	N	SEVERITY OF HEART FAILURE SYMPTOMS	ESTIMATED FIRST-YEAR PLACEBO/CONTROL GROUP MORTALITY	BACKGROUND TREATMENT†	TREATMENT ADDED	TRIAL DURATION (yr)	PRIMARY END POINT	RELATIVE RISK REDUCTION (%)¶	EVENTS PREVENTED PER 1000 PATIENTS TREATED§		
									Death	*HF Hospitalization*	*Death or HF Hospitalization*
ACE INHIBITORS											
CONSENSUS, 1987	253	End stage	52	Spironolactone	Enalapril, 20 mg bid	0.54‡	Death	40	146	—	—
SOLVD-T, 1991	2569	Mild-severe	15.7	—	Enalapril, 20 mg bid	3.5	Death	16	45	96	108
β-BLOCKERS											
CIBIS-2, 1999	2647	Moderate-severe	13.2	ACE-I	Bisoprolol, 10 mg qd	1.3‡	Death	34	55	56	—
MERIT-HF, 1999	3991	Mild-severe	11.0	ACE-I	Metoprolol CR/XL, 200 mg qd	1.0‡	Death	34	36	46	63
COPERNICUS, 2001	2289	Severe	19.7	ACE-I	Carvedilol, 25 mg bid	0.87‡	Death	35	55	65	81
SENIORS, 2005	2128	Mild-severe	10.4	ACE-I + spiro	Nebivolol 10 mg qd	1.75	Death or CV hospitalization	14	23	0	—
ANGIOTENSIN RECEPTOR BLOCKERS (ARBs) OR INHIBITORS											
Val-HeFT, 2001	5010	Mild-severe	~8.0	ACE-I	Valsartan, 160 mg bid	1.9	CV death or morbidity‖	13	0	35	33
CHARM-Alternative, 2003	2028	Mild-severe	12.6	BB	Candesartan, 32 mg qd	2.8	CV death or HF hospitalization	23	30	31	60
CHARM-Added, 2003	2548	Moderate-severe	10.6	ACE-I + BB	Candesartan, 32 mg qd	3.4	CV death or HF hospitalization	15	28	47	39
Neprilysin vs enalapril, 2014	8442	Mild-severe		ACE-I variable	LCZ696, 10 mg bid	2.3	CV death or HF hospitalization	16	32	28	47
MINERALOCORTICOID RECEPTOR ANTAGONISTS (MRAs)											
RALES, 1999	1663	Severe	~25	ACE-I	Spironolactone, 25-50 mg qd	2.0[‡]	Death	30	113	95	—
EMPHASIS-HF, 2011	2737	Mild	~7.0	ACE-I + BB	Eplerenone 25-50 mg qd	1.8	CV death or HF hospitalization	27	30	64	76
HYDRALAZINE-ISDN											
V-HeFT-1, 1986	459	Mild-severe	26.4	—	Hydralazine, 75 mg tid-qid	2.3	Death	34	52	0	—
					ISDN, 40 mg qid						
A-HeFT, 2004	1050	Moderate-severe	~9.0	ACE-I + BB + spironolactone	Hydralazine, 75 mg tid	0.83‡	Composite	—	40	80	—
					ISDN, 40 mg tid						
DIGITALIS GLYCOSIDES											
DIG, 1997	6800	Mild-severe	~11.0	ACE-I	Digoxin	3.1	Death	0	0	79	73
I_F CHANNEL BLOCKER											
SHIFT, 2010	6558	Mild-severe	~7.0%	ACE-I + BB + spironolactone	Ivabradine 5.0-7.5 mg bid	1.91	CV death or HF hospitalization	18	14	47	—

E-TABLE 59-1 CONTROLLED TRIALS IN SYMPTOMATIC HEART FAILURE WITH REDUCED SYSTOLIC FUNCTION—cont'd

TRIAL, TREATMENT, AND YEAR PUBLISHED	*N*	SEVERITY OF HEART FAILURE SYMPTOMS	ESTIMATED FIRST-YEAR PLACEBO/CONTROL GROUP MORTALITY	BACKGROUND TREATMENT†	TREATMENT ADDED	TRIAL DURATION (yr)	PRIMARY END POINT	RELATIVE RISK REDUCTION (%)¶	*EVENTS PREVENTED PER 1000 PATIENTS TREATED*§		
									Death	*HF Hospitalization*	*Death or HF Hospitalization*
N-3 PUFA											
GISSI-HF, 2008	6975	Mild-severe	~7.0%	ACE-I + BB + spironolactone	n-3 PUFA 1g qd	3.9	Death Death or CV hospitalization	9 8	18 —	0 —	— —
CRT											
COMPANION, 2004	925	Moderate-severe	19.0	ACE-I + BB + spironolactone	CRT	1.35‡	Death or any hospital admission	19	38	—	87
CARE-HF, 2005	813	Moderate-severe	12.6	ACE-I + BB + spironolactone	CRT	2.45	Death or CV hospital admission	37	97	151	184
CRT-D											
COMPANION, 2004	903	Moderate-severe	19.0	ACE-I + BB + spironolactone	CRT-ICD	1.35[‡]	Death or any hospital admission	20	74	—	114
MADIT-CRT, 2009	1820	Mild	~3.0%	ACE-I + BB + spironolactone + ICD	CRT-ICD	2.4§	Death or HF event**	34	5	—	—
RAFT, 2011	1798	Mild-severe	~4.0%	ACE-I + BB + spironolactone + ICD	CRT-ICD	3.3	Death or HF hospitalization	25	53	66	70
IMPLANTABLE CARDIOVERTER DEFIBRILLATOR											
SCD-HeFT, 2005	1676	Mild-severe	~7.0	ACE-I + BB	ICD	3.8	Death	23	—	—	—
VENTRICULAR ASSIST DEVICE											
REMATCH, 2001	129	End stage	75	ACE-I + spironolactone	LVAD	1.8	Death	48	282	—	—
EXERCISE TRAINING											
HF-ACTION, 2009	2231	Mild-severe	~6.0	ACE-I + BB + spironolactone	Exercise training	2.5	Death or any hospitalization	7	6	—	—

ACE-I = ACE inhibitor; BB = β-blocker; CRT = cardiac resynchronization therapy (biventricular pacing); CRT-D = CRT device that also defibrillates; CV = cardiovascular; HF hospitalization = patients with at least one hospital admission for worsening heart failure—some patients had multiple admissions; ICD = implantable cardioverter-defibrillator; I_f = hyperpolarization activated pacemaker ("funny") current; ISDN = isosorbide dinitrate; LVAD = left ventricular assist device.

*Excluding active-controlled trials.

†In more than one third of patients, ACE-I + BB means that ACE inhibitors were used in almost all patients and BB in the majority; most patients were also taking diuretics, and many digoxin (except in DIG). Spironolactone was used at baseline in 5% Val-HeFT, 8% MERIT-HF, 17% CHARM-Added, 19% SCD-HeFT, 20% COPERNICUS, and 24% CHARM-Alternative.

‡Relative risk reduction in primary end point.

§Stopped early for benefit.

¶Individual trials may not have been designed or powered to evaluate effect of treatment on these outcomes.

‖Primary end point that also included treatment of heart failure with intravenous drugs for 4 hours or more without admission and resuscitated cardiac arrest (both added small numbers).

**Heart failure hospitalization or treated as an outpatient with intravenous therapy.

Modified from McMurray JJ. Clinical practice. Systolic heart failure. *N Engl J Med.* 2010;362:228-238. *All trials are cited in full in* Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. *Circulation.* 2013;128:e240-327.; and McMurray JJ, Adamopoulos S, Anker SD, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur Heart J.* 2012;33:1787-1847.

different strategies and therapeutic options across the full spectrum of the syndrome, from prevention of heart failure to palliation of patients with end-stage disease.

Stage A: Individuals at Risk for Development of Heart Failure

Stage A designates patients at risk for development of heart failure based on concomitant cardiovascular diseases such as hypertension, coronary artery disease, and diabetes mellitus. Also included in stage A are individuals with prior exposure to cardiotoxic agents such as doxorubicin (Chapter 179) and those with a family history of a cardiomyopathy (Chapter 60). Although these predisposing factors do not by themselves technically constitute the syndrome of heart failure, the guidelines stress the importance of identifying individuals with modifiable factors because this represents an important opportunity to reduce the reservoir of patients at risk.

Population-based preventive approaches can reduce the incidence of heart failure. For example, public health programs targeting the eradication of the insect vector for *Trypanosoma cruzi* (Chapter 347) have reduced the incidence of Chagas cardiomyopathy (Chapter 60) in endemic regions of South and Central America.

Other population-based approaches to reduce the incidence of heart failure require specific screening efforts to identify individuals with modifiable risk factors. The most important although unfortunately nonmodifiable risk factor for the development of heart failure is advanced age; the incidence of heart failure rises sharply per decade after the age of 45 years (Chapter 58). For each decade of age after 45 years, the incidence of heart failure doubles, and heart failure is the leading hospital diagnosis for patients older than 65 years in the United States.

HYPERTENSION

Of the modifiable factors, hypertension (Chapter 67) undoubtedly contributes the greatest population attributable risk for heart failure. In other words, even though the increased risk for heart failure in an individual with hypertension is modest, the high prevalence of hypertension in the general population means that at a population level, hypertension is the major cause of heart failure.

The contribution of hypertension to the risk for heart failure was a consistent finding from all major cardiovascular epidemiologic studies, and the earliest RCTs of antihypertensive therapy showed unambiguous reductions in the risk for heart failure. Of the components of blood pressure, elevated systolic pressure has a greater influence on the incidence of heart failure than does diastolic pressure. In fact, aging is associated with a progressive rise in systolic blood pressure and fall in diastolic pressure as the compliance of the arterial tree diminishes (Chapter 67). In community-based studies, isolated systolic hypertension and elevated pulse pressure have been the most predictive blood pressure measurements for development of heart failure. In the Systolic Hypertension in the Elderly Program, antihypertensive treatment with chlorthalidone followed by atenolol reduced the incidence of new heart failure by about 50%, a treatment effect size recently exceeded (relative risk reduction 64%) with indapamide followed by perindopril in the Hypertension in Very Elderly Trial, probably the last placebo-controlled antihypertensive trial. In general, the actual extent of blood pressure lowering achieved, not the agent used, is the most important factor for preventing heart failure and reducing overall rates of major cardiovascular events. However, the greatest reduction in risk for heart failure seems to be seen when initial therapy is based on a diuretic and angiotensin-converting enzyme (ACE) inhibitor. By comparison, treatment with α-blockers increases the risk for heart failure compared with other antihypertensive drugs. Most important, it is estimated that effective treatment of hypertension (Chapter 67) will substantially reduce the age-adjusted incidence of heart failure by approximately 60% in women and 50% in men.

OTHER RISK FACTORS

Treatment of atherosclerotic risk factors, such as hypercholesterolemia (Chapter 206), and promotion of measures that encourage healthier lifestyles, such as smoking cessation (Chapter 32), weight control (Chapter 220), adoption of a Mediterranean diet (Chapter 213), and aerobic exercise (Chapter 16), should also reduce the number of individuals who progress from stage A to stage B (structural heart disease but without symptoms of heart failure). ACE inhibitors protect against the development of heart failure in patients with diabetes mellitus or with evidence of atherosclerosis. Although obesity is correlated with hypertension, lipid abnormalities, and glucose intolerance, an elevated body mass index is also an independent risk factor for the development of heart failure.

Stage B: Asymptomatic Structural or Functional Heart Disease

Stage B identifies asymptomatic (New York Heart Association or Canadian Cardiovascular Society class I; Chapter 58) patients who have a structural or functional cardiac disorder (e.g., left ventricular hypertrophy, enlargement, or dysfunction and valvar abnormalities) but do not have the signs and symptoms, such as dyspnea and fatigue, of the heart failure syndrome. In addition to history, physical examination, and electrocardiography (Chapter 54), more extensive screening with echocardiography (Chapter 55) or other imaging modalities (Chapter 56) is often required to detect patients with asymptomatic cardiac structural abnormalities.

A patient who has an acute myocardial infarction not complicated by early heart failure is an obvious example of someone who transitions from stage A to stage B. Rapid pharmacologic or mechanical coronary reperfusion is one of the immediate goals of therapy, with the aim of limiting the extent of myocardial injury and reducing the risk for death and future development of heart failure (Chapters 72 and 73). Survivors of the acute phase of myocardial infarction, a well-studied stage B cohort, are at particularly high risk for the future development of heart failure, with an overall annual incidence of 2% per year—but higher in patients who are older, have a lower left ventricular ejection fraction, do not routinely perform at least moderate exercise, or have concomitant hypertension or diabetes mellitus. For example, a clinically stable asymptomatic patient who has recovered from a myocardial infarction but who is older than 60 years with a left ventricular ejection fraction of less than 50% and a history of diabetes and hypertension has an estimated 30% 5-year likelihood of experiencing death or heart failure; without diabetes or hypertension, the 5-year estimated rate becomes 12%. By comparison, a younger myocardial infarction survivor who has a left ventricular ejection fraction over 50% and does not have hypertension or diabetes would be anticipated to have a 5-year rate for heart failure or death of only 3%. Data also suggest that an assessment of right ventricular function provides further independent incremental prediction for the risk for developing heart failure. With the continued improvements in care of patients with acute myocardial infarction (Chapters 72 and 73) and the use of implantable cardioverter-defibrillators (ICDs) after myocardial infarction in patients with reduced left ventricular ejection fraction, this pool of stage B patients, who represent a reservoir for new-onset heart failure, has been expanding. An ICD is recommended in patients who have a left ventricular ejection fraction of 35% or less and who survive at least 40 days after an acute myocardial infarction to reduce the risk for death (Chapter 73).

The impaired left ventricle, often due to a prior myocardial infarction, can undergo progressive chamber enlargement. This process, also termed left ventricular remodeling, describes the time-dependent and often insidious structural alterations of the impaired left ventricle, whereby the relationship of the left ventricular cavity volume increases out of proportion to mass, so the overall ventricular geometry becomes more distorted, usually more spherical. This distortion of left ventricular geometry often leads to mitral regurgitation. These structural changes produce regional and global increases in myocardial wall stress, which can promote further remodeling and contribute to the progressive deterioration of cardiac function and structure often associated with the later stages of symptomatic heart failure.

TREATMENT

Rx

The treatment of heart failure is guided by the stage of symptoms and signs (see Fig. 59-1) as well as a robust literature of therapies proved to be beneficial by randomized trials (Fig. 59-2 and see E-Table 59-1).

Angiotensin-Converting Enzyme Inhibitors and Angiotensin Receptor Blockers

Mechanistic studies confirm that ACE inhibitors inhibit progressive left ventricular enlargement by reducing wall stress during the entire cardiac cycle as well as by more direct inhibition of the intracellular signaling pathways involved in myocardial hypertrophy and interstitial fibrosis. This attenuation of ventricular remodeling by ACE inhibitors reduces the development of symptomatic heart failure and death in stage B asymptomatic patients with left ventricular dysfunction by about 20%. In addition, deaths, often sudden and unexpected, attributed to cardiovascular causes, are reduced in stage B patients by ACE inhibitor therapy.

Trials

Therapy	Stage A	Stage B	Stage C	Stage D
Antihypertensive agents	✓	✓	✓	
Statins	✓	✓	(✓)	(✓)
β-Blockers		✓	✓	✓
ACE inhibitors	✓	✓	✓	✓
Angiotensin II receptor blockers (ARBs)		✓	✓	✓
Hydralazine/ nitrates			✓	✓
Digoxin			✓	✓
Ivabradine			✓	✓
Mineralocorticoid antagonists			✓	✓
Implantable cardioverter-defibrillator (ICD)		✓	✓	(✓)
Cardiac resynchronization therapy (CRT)			✓	✓
Left ventricle assist device (LVAD)				✓
	Stage A High risk for HF without structural heart disease or symptoms of HF	**Stage B** Structural heart disease but without signs or symptoms of HF	**Stage C** Structural heart disease with prior or current symptoms of HF	**Stage D** Refractory HF requiring specialized interventions
	Patients with • Hypertension • Artherosclerotic disease • Diabetes • Obesity • Metabolic syndrome or Patients using • Cardiotoxins with family history of cardiomyopathy (FHx CM)	Patients with • Previous MI • LV remodeling including LVH and low EF • Asymptomatic valvular disease	Patients with • Known structural heart disease • Shortness of breath and fatigue, reduced exercise tolerance	Patients with marked symptoms at rest despite maximal medical therapy (e.g., those who are recurrently hospitalized or cannot be safely discharged from the hospital without specialized interventions)
	THERAPY **Goals** • Treat hypertension • Encourage smoking cessation • Treat lipid disorders • Encourage regular exercise • Discourage alcohol intake, illicit drug use • Control metabolic syndrome **Drugs** • ACEI or ARB as appropriate for patients with vascular disease or diabetes	**THERAPY** **Goals** • Treat hypertension • Encourage smoking cessation • Treat lipid disorders • Encourage regular exercise • Discourage alcohol intake, illicit drug use • Control metabolic syndrome **Drugs** • ACEI or ARB as appropriate for patients with vascular disease or diabetes • β-Blockers in appropriate patients **Devices in selected patients** • Implantable defibrillators	**THERAPY** **Goals** • Treat hypertension • Encourage smoking cessation • Treat lipid disorders • Encourage regular exercise • Discourage alcohol intake, illicit drug use • Control metabolic syndrome • Dietary salt restriction **Drugs for routine use** • Diuretics for fluid retention • ACEI • β-Blockers in appropriate patients • Mineralocorticoid antagonists **Drugs in selected patients** • ARBs • Digitalis • Hydralazine/nitrates • Ivabradine **Devices in selected patients** • Biventricular pacing • Implantable defibrillators	**THERAPY** **Goals** • Appropriate measures under stages A, B, C • Decision re: appropriate level of care **Options** • Compassionate end-of-life care/hospice • Extraordinary measures • Heart transplantation • Chronic inotropes • Permanent mechanical support • Experimental surgery or drugs

FIGURE 59-2. Stages of heart failure and therapies at various stages of heart failure (HF). Check marks indicate therapies proved by randomized trials to be beneficial; check marks in brackets indicate benefits uncertain. ACEI, angiotensin converting enzyme inhibitor; ARB, angiotensin receptor blocker; EF, ejection fraction; LV, left ventricle; LVH, left ventricular hypertrophy; MI, myocardial infarction.

Several ACE inhibitors are effective as prophylactic therapy for high-risk stage B patients, and the target dose of each agent is established (Table 59-2). Therefore, patients with left ventricular systolic dysfunction, heart failure, or both complicating acute myocardial infarction should receive an ACE inhibitor to reduce the risk for chronic heart failure, reinfarction, stroke, and death.[1,2] The angiotensin receptor blocker (ARB) valsartan (Table 59-3) is as effective as captopril in reducing risk for cardiovascular death and other nonfatal cardiovascular outcomes, thereby providing an alternative pharmacologic class of agents for patients who cannot tolerate an ACE inhibitor because of cough or angioedema. Importantly, in patients with left ventricular dysfunction or acute heart failure in the context of a myocardial infarction, the combination of an ACE inhibitor and ARB is not better than either alone, so combination therapy is not recommended in this setting.

β-Blockers

β-Adrenergic receptor blockers (β-blockers) have long been known to reduce death and recurrent myocardial infarction when they are administered during the acute phase of myocardial infarction in patients without pulmonary congestion (Chapter 73). However, carvedilol (Table 59-4) also improves survival, reduces subsequent nonfatal myocardial infarctions, and has a favorable trend for reduced hospitalizations for heart failure in patients with a recent myocardial infarction and reduced left ventricular ejection fraction (≤40%) when it is added to an ACE inhibitor and should be considered in such patients. For stage B patients whose left ventricular dysfunction does not have an ischemic etiology, the evidence for β-blockers is less firm.

Treatment of Arrhythmias

Functional as well as structural problems may lead to the development of heart failure. For example, a persistently rapid ventricular rate in patients with atrial fibrillation can cause a rate-related (tachycardia-induced) cardiomyopathy (Chapter 64). Adequate pharmacologic control of the ventricular rate or interventions to restore sinus rhythm or to ablate re-entry pathways (Chapter 66) may reduce the risk for heart failure.

Other Therapies

Any treatments that control hypertension or reduce the risk for myocardial infarction will benefit stage B patients. Examples include statins, antiplatelet agents, and smoking cessation.

Stages C and D: Symptomatic Heart Failure

The development of symptoms and signs of the heart failure syndrome defines the transition from patients in the asymptomatic "at-risk" stages (A and B) to those who fulfill the clinical diagnosis of symptomatic heart failure (Chapter 58). This transition to the symptomatic phase underscores the progressive nature of heart failure and heralds a marked decline in prognosis. In one study, for example, the 2-year mortality rate was 27% in symptomatic patients compared with 10% in asymptomatic patients despite similarly reduced left ventricular ejection fractions and comorbidities.

TREATMENT

Rx

The goals of treatment for patients with stage C and stage D heart failure are relief of symptoms, avoidance of hospital admission, and prevention of premature death. In general, the preventive measures that are of value during stages A and B should be sustained in patients with stages C and D heart failure.

Heart Failure with Reduced Left Ventricular Ejection Fraction

Pharmacologic Treatment

Drugs are the mainstay of the treatment of patients with symptomatic heart failure on the basis of the cumulative experiences from RCTs (see E-Table 59-1), particularly for patients with reduced left ventricular ejection fraction. However, devices and surgery have an important and increasing role in patients with advanced symptomatic heart failure (stages C and D; see Fig. 59-1). Exercise clearly improves well-being and clinical outcomes (see E-Table 59-1), but the evidence base for other lifestyle interventions is less robust. The organization and delivery of care can also have a substantial impact on outcomes.

Diuretics

Mechanism of Action

Diuretics act by blocking sodium reabsorption at specific sites in the renal tubule, thereby enhancing urinary excretion of sodium and water.

Clinical Benefits

Although not proven to improve mortality and morbidity in large trials, diuretics are required in nearly all patients with symptomatic heart failure (stages C and D) to relieve dyspnea and the signs of sodium and water retention ("congestion"), that is, peripheral and pulmonary edema. No other treatment relieves symptoms and the signs of sodium and water overload as rapidly and effectively. Once a patient needs a diuretic, treatment is usually necessary for the rest of the patient's life, although the dose and type of diuretic may vary.

Practical Use

The key principle is to prescribe the minimum dose of diuretic needed to maintain an edema-free state ("dry weight"). Excessive use can lead to electrolyte imbalances, such as hyponatremia, hypokalemia (and risk for digitalis toxicity), hyperuricemia (and risk for gout), and uremia. The risk for renal dysfunction is increased by concomitant use of nonsteroidal anti-inflammatory drugs (NSAIDs). Diuretic-induced hypovolemia may also cause symptomatic hypotension and prerenal azotemia. Restriction of dietary sodium intake may help reduce but does not eliminate the requirement for diuretics. Diuretic dosing should be flexible, with temporary increases for evidence of fluid retention (e.g., increasing symptoms, weight gain, edema) and decreases for evidence of hypovolemia (e.g., as a consequence of increased electrolyte loss due to gastroenteritis, decreased fluid intake, or both).

In some patients with milder symptoms of heart failure and preserved renal function (stage C), a thiazide diuretic such as chlorthalidone may suffice. In more advanced heart failure (stage D) or in patients with concomitant renal dysfunction, a loop diuretic such as furosemide is often needed. Loop diuretics cause a rapid onset of an intense but relatively short-lived diuresis compared with the longer lasting but gentler effect of a thiazide diuretic. The timing of administration of a loop diuretic, which need not be taken first thing every morning, can be adjusted according to the patient's social activities. The dose may be postponed or even temporarily omitted if the patient has to travel or has another activity that might be compromised by the prompt action of the diuretic. In severe heart failure (stage D), the effects of long-term administration of a loop diuretic may be diminished by increased sodium reabsorption at the distal tubule. This problem can be offset by use of the combination of a loop diuretic and a thiazide or thiazide-like diuretic (e.g., hydrochlorothiazide or metolazone), which act in synergy with a loop diuretic by blocking sodium reabsorption in different segments of the nephron. This combination requires more frequent monitoring of electrolytes and renal function for diuretic-induced hyponatremia, abnormalities of the serum potassium level, and prerenal azotemia.

A period of intravenous loop diuretic, given either as bolus injections or by continuous infusion, may be required in patients who become resistant to the action of oral diuretics. Why this resistance develops is uncertain, but factors thought to be important include impaired absorption of oral diuretics due to gut edema, hypotension, reduced renal blood flow, renal venous congestion, and adaptive changes in the nephron.

Patients with symptomatic heart failure (stages C and D) should be also considered for treatment with a mineralocorticoid receptor (aldosterone) antagonist, such as spironolactone, which increases excretion of sodium but not of potassium (see later). Patients receiving a combination of diuretics require careful monitoring of blood chemistry and clinical status. The use of a mineralocorticoid receptor (or, rarely, a potassium-sparing diuretic) along with an ACE inhibitor or ARB (treatment with all three is not recommended) requires particular care and surveillance for hyperkalemia.

Although they are highly effective in relieving symptoms and signs, diuretics alone are not sufficient for treatment of heart failure. In cases of severe resistant volume overload, mechanical removal of fluid by ultrafiltration may be considered. The addition of other disease-modifying treatments will better maintain clinical stability, slow structural progression, and reduce the risk for hospital admission and premature death.

ACE Inhibitors

Mechanism of Action

These drugs act by inhibiting the enzyme that converts the inactive decapeptide angiotensin I to the active octapeptide angiotensin II (and that also breaks down bradykinin). In patients with heart failure, excessive angiotensin II is thought to exert myriad harmful actions mediated through stimulation of the angiotensin II type 1 receptor subtype (AT1R), including vasoconstriction (which increases ventricular afterload), excessive growth of myocytes and the extracellular matrix (contributing to maladaptive left ventricular remodeling), activation of the sympathetic nervous system, prothrombotic actions, and augmentation of the release of arginine vasopressin and the retention of sodium (both directly and through stimulation of secretion of aldosterone, which activates the mineralocorticoid receptor).

ACE inhibitors also reduce the breakdown of bradykinin (because ACE is identical to kininase II), and the resultant accumulation of bradykinin is directly or indirectly responsible for two of the specific adverse effects of ACE inhibitors, cough, and angioedema. Bradykinin may, however, also have beneficial effects (vasodilation, inhibition of adverse cardiovascular remodeling, and

TABLE 59-2 PRACTICAL GUIDANCE ON THE USE OF ANGIOTENSIN-CONVERTING ENZYME INHIBITORS AND ANGIOTENSIN RECEPTOR BLOCKERS IN PATIENTS WITH HEART FAILURE DUE TO LEFT VENTRICULAR SYSTOLIC DYSFUNCTION

WHY?

Two major randomized trials (CONSENSUS I and SOLVD-T) and a meta-analysis of smaller trials have conclusively shown that angiotensin-converting enzyme (ACE) inhibitors increase survival, reduce hospital admissions, and improve New York Heart Association (NYHA) class and quality of life in patients with *all* grades of symptomatic heart failure. Other major randomized trials in patients with systolic dysfunction after acute myocardial infarction (SAVE, AIRE, TRACE) have shown that angiotensin-converting enzyme (ACE) inhibitors increase survival. In patients with heart failure (ATLAS), the composite end point of death or hospital admission was reduced by higher doses of ACE inhibitor compared with lower doses. ACE inhibitors have also been shown to delay or to prevent the development of symptomatic heart failure in patients with *asymptomatic* left ventricular systolic dysfunction.

In patients previously intolerant of an ACE inhibitor, candesartan has been shown to reduce the risk for the composite outcome of cardiovascular death or heart failure hospitalization, to reduce the risk for heart failure hospital admission, and to improve NYHA class. These findings in heart failure are supported by another randomized trial in patients with left ventricular systolic dysfunction, heart failure, or both complicating acute myocardial infarction (VALIANT) in which valsartan was as effective as the ACE inhibitor captopril in reducing mortality and cardiovascular morbidity.

Added to standard therapy, including an ACE inhibitor, in patients with all grades of symptomatic heart failure, the angiotensin receptor blockers (ARBs) valsartan and candesartan have been shown, in two major randomized trials (Val-HeFT and CHARM), to reduce heart failure hospital admissions, to improve NYHA class, and to maintain quality of life. The two CHARM low–left ventricular ejection fraction trials (CHARM-Alternative and CHARM-Added) also showed that candesartan reduced all-cause mortality.

IN WHOM AND WHEN?

ACE Inhibitors

Indications
- Potentially all patients with heart failure and a low ejection fraction
- First-line treatment (along with β-blockers) in patients with NYHA class II to IV heart failure; start as early as possible in course of disease. ACE inhibitors are also of benefit in patients with asymptomatic left ventricular systolic dysfunction (NYHA class I).

Contraindications
- History of angioedema
- Known bilateral renal artery stenosis

Cautions/seek specialist advice
- Significant hyperkalemia ($K^+ > 5.0$ mmol/L)
- Significant renal dysfunction (creatinine 221 μmol/L or >2.5 mg/dL)
- Symptomatic or severe asymptomatic hypotension (systolic blood pressure <90 mm Hg)

Drug interactions to look out for
- K^+ supplements/K^+-sparing diuretics, e.g., amiloride and triamterene (beware combination preparations with furosemide)
- Mineralocorticoid receptor antagonists (spironolactone, eplerenone), angiotensin receptor blockers, NSAIDs*
- "Low-salt" substitutes with a high K^+ content

Angiotensin Receptor Blockers

Indications
- First-line treatment (along with β-blockers) in patients with NYHA class II to IV heart failure intolerant of an ACE inhibitor because of cough or angioedema
- Second-line treatment (after optimization of ACE inhibitor and β-blocker*) in patients with NYHA class II–IV heart failure intolerant of a mineralocorticoid receptor antagonist

Contraindications
- Known bilateral renal artery stenosis

Cautions/seek specialist advice
- As for ACE inhibitors

Drug interactions to look out for
- As for ACE inhibitors

WHERE?

In the community for most patients
Exceptions—see Cautions/seek specialist advice

WHICH ACE INHIBITOR AND WHAT DOSE?

	STARTING DOSE	TARGET DOSE
Captopril	6.25 mg thrice daily	50 mg thrice daily
Enalapril	2.5 mg twice daily	10-20 mg twice daily
Lisinopril	2.5-5.0 mg once daily	20-35 mg once daily
Ramipril	2.5 mg once daily	5 mg twice daily or 10 mg once daily
Trandolapril	0.5 mg once daily	4 mg once daily

WHICH ARB AND WHAT DOSE?

	STARTING DOSE	TARGET DOSE
Candesartan	4 or 8 mg once daily	32 mg once daily
Valsartan	40 mg twice daily	160 mg twice daily
Losartan	50 mg once daily	150 mg daily

HOW TO USE?

Start with a low dose (see above).
Double dose at not less than 2-week intervals.
Aim for target dose (see above) or, failing that, the highest tolerated dose.
Remember: *some* ACE inhibitor/ARB is better than *no* ACE inhibitor/ARB.
Monitor blood pressure and blood chemistry (urea/blood urea nitrogen, creatinine, K^+).
Check blood chemistry 1-2 weeks after initiation and 1-2 weeks after final dose titration.
When to stop up-titration, reduce dose, stop treatment—see Problem Solving.
A specialist heart failure nurse may assist with education of the patient, follow-up (in person or by telephone), biochemical monitoring, and dose up-titration.

ADVICE TO PATIENT

Explain expected benefits (see Why?).
Treatment is given to improve symptoms, to prevent worsening of heart failure leading to hospital admission, and to increase survival.
Symptoms improve within a few weeks to a few months of starting treatment.
Advise patients to report principal adverse effects, (i.e., dizziness/symptomatic hypotension, cough)—see Problem Solving.
Advise patients to avoid NSAIDs* not prescribed by a physician (self-purchased over-the-counter) and salt substitutes high in K^+—see Problem Solving.

PROBLEM SOLVING

Asymptomatic low blood pressure
- Does not usually require any change in therapy

Symptomatic hypotension
- If dizziness, lightheadedness, or confusion and a low blood pressure, reconsider need for nitrates, calcium-channel blockers,† and other vasodilators.
- If no signs or symptoms of congestion, consider reducing diuretic dose.
- If these measures do not solve problem, seek specialist advice.

Cough
- Cough is common in patients with heart failure, many of whom have smoking-related lung disease.
- Cough is also a symptom of pulmonary edema, which should be excluded when a new or worsening cough develops.
- ACE inhibitor–induced cough rarely requires treatment discontinuation.
- When a troublesome cough does develop (e.g., one stopping the patient from sleeping) and can be proved to be due to ACE inhibition (i.e., recurs after ACE inhibitor withdrawal and rechallenge), substitution of an ARB can be considered

Worsening renal function
- Some rise in urea (blood urea nitrogen), creatinine, and potassium is to be expected after initiation of an ACE inhibitor/ARB; if an increase is small and asymptomatic, no action is necessary.
- An increase in creatinine of up to 50% above baseline, or 266 μmol/L (3 mg/dL), whichever is the smaller, is acceptable.
- An increase in potassium to ≤5.5 mmol/L is acceptable.
- If urea, creatinine, or potassium does rise excessively, consider stopping concomitant nephrotoxic drugs (e.g., NSAIDs*) and other potassium supplements or retaining agents (triamterene, amiloride, spironolactone-eplerenone‡) and, if no signs of congestion, reducing the dose of diuretic.
- If greater rises in creatinine or potassium than those outlined above persist despite adjustment of concomitant medications, the dose of the ACE inhibitor should be halved and blood chemistry rechecked within 1 to 2 weeks; if there is still an unsatisfactory response, specialist advice should be sought.
- If potassium rises to >5.5 mmol/L or creatinine increases by >100% or to above 310 μmol/L (3.5 mg/dL), the ACE inhibitor/ARB should be stopped and specialist advice sought.
- Blood chemistry should be monitored frequently and serially until potassium and creatinine have plateaued.

Note: It is rarely necessary to stop an ACE inhibitor/ARB, and clinical deterioration is likely if treatment is withdrawn. Ideally, specialist advice should be sought before treatment discontinuation.
*Avoid nonsteroidal anti-inflammatory drugs (NSAIDs) unless essential.
†Calcium-channel blockers should be discontinued unless absolutely essential (e.g., for angina or hypertension).
‡The safety and efficacy of an ACE inhibitor used with an angiotensin receptor blocker and spironolactone (as well as β-blocker) are uncertain, and the use of all three inhibitors of the renin-angiotensin-aldosterone system together is not recommended.
Modified from McMurray J, Cohen-Solal A, Dietz R, et al. Practical recommendations for the use of ACE inhibitors, β-blockers, mineralocorticoid receptor antagonists and angiotensin receptor blockers in heart failure: putting guidelines into practice. *Eur J Heart Fail.* 2005;7:710-721.

TABLE 59-3 PRACTICAL GUIDANCE ON THE USE OF β-BLOCKERS IN PATIENTS WITH HEART FAILURE DUE TO LEFT VENTRICULAR SYSTOLIC DYSFUNCTION

WHY?

Several major randomized controlled trials (i.e., USCP, CIBIS II, MERIT-HF, COPERNICUS) have shown, conclusively, that certain β-blockers increase survival, reduce hospital admissions, and improve New York Heart Association (NYHA) class and quality of life when added to standard therapy (diuretics, digoxin, and angiotensin-converting enzyme [ACE] inhibitors) in patients with *stable* mild and moderate heart failure and in some patients with severe heart failure. In the SENIORS trial, which differed substantially in design from the aforementioned studies (older patients, some patients with preserved left ventricular systolic function, longer follow-up), nebivolol appeared to have a smaller treatment effect, although direct comparison is difficult. One other trial (BEST) did not show a reduction in all-cause mortality but did report a reduction in cardiovascular mortality and is otherwise broadly consistent with the aforementioned studies. The COMET trial showed that carvedilol was substantially more effective than a low dose of short-acting metoprolol tartrate* (long-acting metoprolol succinate was used in MERIT-HF).

IN WHOM AND WHEN?

Indications
- Potentially *all* patients with *stable* mild and moderate heart failure; patients with severe heart failure should be referred for specialist advice.
- First-line treatment (along with ACE inhibitors) in patients with *stable* NYHA class II to III heart failure; start as early as possible in course of disease.

Contraindications
- Asthma
- Second- or third-degree atrioventricular block

Cautions/seek specialist advice
- Severe (NYHA class IV) heart failure
- Current or recent (<4 weeks) exacerbation of heart failure (e.g., hospital admission with worsening heart failure, heart block, or heart rate <60 beats/minute).
- Persisting signs of congestion, hypotension/low blood pressure (systolic < 90 mm Hg), raised jugular venous pressure, ascites, marked peripheral edema

Drug interactions to look out for
- Verapamil, diltiazem (should be discontinued)†
- Digoxin, amiodarone

WHERE?

In the community in stable patients (NYHA class IV/severe heart failure patients should be referred for specialist advice)
Not in unstable patients hospitalized with worsening heart failure
Other exceptions—see Cautions/seek specialist advice

WHICH β-BLOCKER AND WHAT DOSE?

	STARTING DOSE	TARGET DOSE
Bisoprolol	1.25 mg once daily	10 mg once daily
Carvedilol	3.125 mg twice daily	25-50 mg twice daily
Metoprolol CR/XL	12.5-25 mg once daily	200 mg once daily*
Nebivolol	1.25 mg once daily	10 mg once daily

HOW TO USE?

Start with a low dose (see above).
Double dose at *not less than* 2-week intervals.
Aim for target dose (see above) or, failing that, the highest tolerated dose.
Remember: *some* β-blocker is better than *no* β-blocker.
Monitor heart rate, blood pressure, and clinical status (symptoms, signs—especially signs of congestion, body weight).
Check blood chemistry 1 to 2 weeks after initiation and 1 to 2 weeks after final dose titration.
When to stop up-titration, reduce dose, stop treatment—see Problem Solving.
A specialist heart failure nurse may assist with education of the patient, follow-up (in person or by telephone), and dose up-titration.

ADVICE TO PATIENT

Explain expected benefits (see Why?).
Treatment is given to improve symptoms, to prevent worsening of heart failure leading to hospital admission, and to increase survival.
Symptomatic improvement may develop slowly after starting treatment, taking 3 to 6 months or longer.
Temporary symptomatic deterioration *may* occur during initiation or up-titration phase; in the long term, β-blockers improve well-being.
Advise patient to report deterioration (see Problem Solving) and that deterioration (tiredness, fatigue, breathlessness) can usually be easily managed by adjustment of other medication; patients should be advised not to stop β-blocker therapy without consulting the physician.
To detect and to treat deterioration early, patients should be encouraged to weigh themselves daily (after waking, before dressing, after voiding, before eating) and to increase their diuretic dose should their weight increase, persistently (>2 days), by >1.5–2.0 kg.‡

PROBLEM SOLVING

Worsening symptoms or signs (e.g., increasing dyspnea, fatigue, edema, weight gain)
- If increasing congestion, increase dose of diuretic or halve dose of β-blocker (if increasing diuretic does not work).
- If marked fatigue (or bradycardia—see below), halve dose of β-blocker (rarely necessary); review patient in 1 to 2 weeks; if not improved, seek specialist advice.
- If serious deterioration, halve dose of β-blocker or stop this treatment (rarely necessary); seek specialist advice.

Low heart rate
- If <50 beats/minute and worsening symptoms, halve dose of β-blocker or, if severe deterioration, stop β-blocker (rarely necessary).
- Review need for other heart rate–slowing drugs (e.g., digoxin, amiodarone, diltiazem, or verapamil†).
- Arrange electrocardiogram to exclude heart block.
- Seek specialist advice.

Asymptomatic low blood pressure
- Does not usually require any change in therapy.

Symptomatic hypotension
- If dizziness, lightheadedness, or confusion and a low blood pressure, reconsider need for nitrates, calcium-channel blockers,† and other vasodilators.
- If no signs or symptoms of congestion, consider reducing diuretic dose or ACE inhibitor.
- If these measures do not solve problem, seek specialist advice.

Note: β-Blockers should not be stopped suddenly unless absolutely necessary (there is a risk for a "rebound" increase in myocardial ischemia or infarction and arrhythmias). Ideally, specialist advice should be sought before treatment discontinuation.
*Metoprolol tartrate should not be used in preference to an evidence-based β-blocker in heart failure.
†Calcium-channel blockers should be discontinued unless absolutely necessary, and diltiazem and verapamil are generally contraindicated in heart failure.
‡This is generally good advice for all patients with heart failure.
Modified from McMurray J, Cohen-Solal A, Dietz R, et al. Practical recommendations for the use of ACE inhibitors, β-blockers, mineralocorticoid receptor antagonists and angiotensin receptor blockers in heart failure: putting guidelines into practice. *Eur J Heart Fail.* 2005;7:710-721.

TABLE 59-4 PRACTICAL GUIDANCE ON THE USE OF MINERALOCORTICORTICOID RECEPTOR ANTAGONISTS IN PATIENTS WITH HEART FAILURE DUE TO LEFT VENTRICULAR SYSTOLIC DYSFUNCTION

WHY?

The RALES study showed that low-dose spironolactone increased survival, reduced hospital admissions, and improved New York Heart Association (NYHA) class when added to standard therapy (diuretic, digoxin, angiotensin-converting enzyme [ACE] inhibitor, and, in a minority of cases, β-blocker) in patients with severe (NYHA class III or IV) heart failure symptoms. The findings of RALES are supported by another randomized trial in patients with heart failure, reduced ejection fraction, and mild symptoms (NYHA class II) in which another mineralocorticoid receptor (MRA), antagonist, eplerenone, increased survival and reduced hospital admissions for heart failure when added to an ACE inhibitor (or angiotensin receptor blocker [ARB]) and β-blocker (EMPHASIS-HF). These findings in heart failure are supported by another randomized trial in patients with left ventricular systolic dysfunction and heart failure (or diabetes) complicating *acute* myocardial infarction (EPHESUS), in which eplerenone increased survival and reduced hospital admission for cardiac causes.

IN WHOM AND WHEN?

Indications
- Potentially all patients with symptomatic heart failure (class II to IV NYHA)
- Second-line therapy (after ACE inhibitors and β-blockers*) in patients with symptomatic heart failure (NYHA class II to IV); second-line therapy (after ACE inhibitors and β-blockers) in patients with a LVEF ≤ 40%

Cautions/seek specialist advice
- Significant hyperkalemia (K^+ > 5.0 mmol/L)†
- Significant renal dysfunction (creatinine > 221 µmol/L or 2.5 mg/dL)†

Drug interactions to look out for
- K^+ supplements/K^+-sparing diuretics (e.g., amiloride and triamterene; beware combination preparations with furosemide)
- ACE inhibitors, angiotensin receptor blockers, NSAIDs‡
- "Low-salt" substitutes with a high K^+ content

WHERE?

In the community or in the hospital
Exceptions—see Cautions/seek specialist advice

WHICH DOSE?†

	STARTING DOSE	TARGET DOSE
Spironolactone	25 mg once daily or on alternate days	25-50 mg once daily
Eplerenone	25 mg once daily	50 mg once daily

HOW TO USE?

Start with a low dose (see above).
Check blood chemistry at 1, 4, 8, and 12 weeks; 6, 9, and 12 months; 6-monthly thereafter.
If K^+ rises above 5.5 mmol/L or creatinine rises to 221 µmol/L (2.5 mg/dL), reduce dose to 25 mg on alternate days and monitor blood chemistry closely.
If K^+ rises to >6.0 mmol/L or creatinine to >310 µmol/L (3.5 mg/dL), stop spironolactone immediately and seek specialist advice.
A specialist heart failure nurse may assist with education of the patient, follow-up (in person or by telephone), biochemical monitoring, and dose up-titration.

ADVICE TO PATIENT

Explain expected benefits (see Why?).
Treatment is given to improve symptoms, to prevent worsening of heart failure leading to hospital admission, and to increase survival.
Symptom improvement occurs within a few weeks to a few months of starting treatment.
Avoid NSAIDs‡ not prescribed by a physician (self-purchased over-the-counter agent) and salt substitutes high in K^+.
If diarrhea or vomiting occurs, patients should stop the mineralocorticoid receptor and contact the physician.

PROBLEM SOLVING

Worsening renal function/hyperkalemia
- See How to Use? section.

Major concern is hyperkalemia (>6.0 mmol/L); although this was uncommon in RALES and EMPHASIS-HF, it has been seen more commonly in clinical practice.
- Conversely, a high-normal potassium level may be desirable in patients with heart failure, especially if they are taking digoxin.

It is important to avoid other K^+-retaining drugs (e.g., K^+-sparing diuretics such as amiloride and triamterene) and nephrotoxic agents (e.g., NSAIDs‡)
The risk for hyperkalemia and renal dysfunction when a mineralocorticoid receptor antagonist is given to patients already taking an ACE inhibitor and ARB is higher than when a mineralocorticoid receptor antagonist is added to just an ACE inhibitor or ARB given singly; close and careful monitoring is mandatory.*
Some "low-salt" substitutes have a high K^+ content.
Male patients treated with spironolactone may develop breast discomfort or gynecomastia (these problems are significantly less common with eplerenone).

Modified from McMurray J, Cohen-Solal A, Dietz R, et al. Practical recommendations for the use of ACE inhibitors, β-blockers, aldosterone antagonists and angiotensin receptor blockers in heart failure: putting guidelines into practice. *Eur J Heart Fail.* 2005;7:710-721.
*The safety and efficacy of spironolactone used with an ACE inhibitor and an ARB (as well as a β-blocker) are uncertain, and the use of all three inhibitors of the renin-angiotensin-aldosterone system together is not recommended.
†It is extremely important to adhere to these cautions and doses in light of recent evidence of serious hyperkalemia with spironolactone in usual clinical practice in Ontario.
‡Avoid nonsteroidal anti-inflammatory drugs (NSAIDs) unless essential.

antithrombotic actions), although the importance of these bradykinin-mediated actions to the clinical benefits of ACE inhibition is uncertain.

Clinical Benefits

Clinical trials have shown that treatment with an ACE inhibitor, when it is used alone or added to diuretics and digoxin, decreases left ventricular size, improves ejection fraction, reduces symptoms and hospital admissions, and prolongs survival (see E-Table 59-1). These agents also reduce the risk for development of myocardial infarction and possibly diabetes, and atrial fibrillation. Consequently, treatment with an ACE inhibitor is recommended for all patients with left ventricular systolic dysfunction, irrespective of symptoms or etiology. ACE inhibitors are not a substitute for a diuretic but mitigate diuretic-induced hypokalemia.

Practical Use

ACE inhibitors should be introduced as early as possible in a patient's treatment. The only contraindications are current symptomatic hypotension and bilateral renal artery stenosis (Chapter 125); the latter is often associated with a prompt and marked increase in serum levels of blood urea nitrogen and creatinine when renal perfusion is reduced precipitously by inhibiting the production and actions of angiotensin II. Treatment should be started in a low dose (see Table 59-2), with the dose gradually increased toward a target dose of proven benefit in a clinical trial. The patient should be evaluated for symptomatic hypotension, uremia, and hyperkalemia after each dose increment; these adverse effects are uncommon and can usually be resolved by reduction in the dose of diuretic (if the patient is edema free) or concomitant hypotensive or nephrotoxic medications (e.g., nitrates, calcium-channel blockers, or NSAIDs). A dry, nonproductive cough occurs in approximately 15% of patients treated with an ACE inhibitor, and if it is troublesome, substitution of an ARB is recommended. In the rare cases of angioedema (Chapter 252), the ACE inhibitor should be stopped and not used again; an ARB can be cautiously substituted (see later).

β-Blockers

Mechanism of Action

Heart failure is characterized by excessive activation of the sympathetic nervous system, which causes vasoconstriction and sodium retention, thereby increasing cardiac preload and afterload and often inducing myocardial ischemia or arrhythmias. In addition, norepinephrine can cause hypertrophy of myocytes and augment their apoptosis. β-Blockers counteract many of these harmful effects of the hyperactivity of the sympathetic nervous system. A rapid heart rate is an important prognostic factor in heart failure among patients in sinus rhythm, and β-blockers reduce heart rate.

Clinical Benefits

The long-term addition of a β-blocker to an ACE inhibitor (and diuretic, digoxin, and mineralocorticoid receptor antagonist) further improves left ventricular function and symptoms, reduces hospital admissions, and strikingly improves survival. Consequently, a β-blocker is recommended for all patients with symptomatic systolic dysfunction, irrespective of etiology and severity, and the combination of a β-blocker with an ACE inhibitor is now the cornerstone of the treatment of symptomatic heart failure (see Fig. 59-1). Treatment with a β-blocker, added to an ACE inhibitor, is recommended for all patients with symptoms (NYHA classes II to IV) and left ventricular systolic dysfunction, irrespective of etiology.

Practical Use

The major contraindications to use of a β-blocker in heart failure are asthma (although it is important to note that the dyspnea caused by pulmonary congestion can be confused with reactive airway disease) and second- or third-degree atrioventricular block. Initiation of treatment during an episode of acute decompensated heart failure should also be avoided until the patient is stabilized. In addition, caution is advised in patients with a heart rate below 60 beats per minute or a systolic blood pressure below 90 mm Hg. It is recommended that a β-blocker shown to produce benefits in a randomized trial be used (see E-Table 59-1).

Like ACE inhibitors, β-blockers should be introduced as early as possible in a patient's treatment, started in a low dose (see Table 59-3), and increased gradually toward a target dose used in a clinical trial (the "start low–go slow" approach). The patient should be checked for symptomatic hypotension and excessive bradycardia after each dose increment, but both of these side effects are uncommon, and hypotension can often be resolved by reduction in the dose of other nonessential blood pressure–lowering medications (e.g., nitrates and calcium-channel blockers). Bradycardia is more likely in patients who are also taking digoxin or amiodarone, and the simultaneous use of these agents should be reviewed if excessive bradycardia occurs. On occasion, symptomatic worsening and fluid retention (e.g., weight gain or edema) may occur after initiation of a β-blocker or during dose up-titration; these side effects usually can be resolved by a temporary increase in the diuretic dose without necessitating discontinuation of the β-blocker.

Treatment with a β-blocker should be given for life, although the dose may need to be decreased (or, rarely, treatment discontinued) temporarily during episodes of acute decompensation if the patient shows signs of circulatory underperfusion or refractory congestion.

Mineralocorticoid Receptor (Aldosterone) Antagonists

Mechanism of Action

Aldosterone, which is the second effector hormone in the renin-angiotensin-aldosterone cascade, has detrimental vascular, renal, autonomic, and cardiac actions when it is produced in excess in patients with heart failure. Excessive aldosterone promotes sodium retention and hypokalemia, and it is believed to contribute to myocardial fibrosis, all of which predispose to arrhythmias. Aldosterone mediates its effects by activating the mineralocorticoid receptor, which is also stimulated by other endogenous corticosteroids. Mineralocorticoid receptor antagonists block these undesirable actions and, at high doses, also act as potassium-sparing diuretics.

Clinical Benefits

The mineralocorticoid receptor antagonist spironolactone (see E-Table 59-1) improves symptoms, reduces hospital admissions, and increases survival when it is added to an ACE inhibitor (and diuretics and digoxin) in patients with a reduced left ventricular ejection fraction and severely symptomatic heart failure. Eplerenone, another mineralocorticoid receptor antagonist, reduces mortality and morbidity when it is added to both an ACE inhibitor and β-blocker in patients with a reduced left ventricular ejection fraction and heart failure with mild symptoms (NYHA class II) (see E-Table 59-1). Consequently, a mineralocorticoid receptor antagonist should be considered in all patients who remain symptomatic (class II to IV) despite treatment with a diuretic, ACE inhibitor (or ARB), and β-blocker.[2] Addition of a mineralocorticoid receptor antagonist to an ACE inhibitor (and diuretic, digoxin, and β-blocker) is preferred to the addition of an ARB because of the greater benefit of a mineralocorticoid receptor antagonist, particularly in reducing all-cause mortality. When begun, a mineralocorticoid receptor antagonist should be given indefinitely. The combination of an ACE inhibitor, an ARB, and a mineralocorticoid receptor antagonist has not been adequately evaluated and is not recommended.

Practical Use

Treatment with a mineralocorticoid receptor antagonist should be initiated with a low dose (see Table 59-4) with careful monitoring of serum electrolytes and renal function. Hyperkalemia and uremia are the adverse effects of greatest concern (as with ACE inhibitors and ARBs), and a mineralocorticoid receptor antagonist should not be given to patients with a serum potassium concentration of more than 5.0 mmol/L, serum creatinine concentration above 2.5 mg/dL (>221 μmol/L), or other evidence of markedly impaired renal function. The importance of selection of patients and dose is underscored by reports of a worrisome incidence of serious hyperkalemia in community practice settings. Spironolactone can have antiandrogenic effects, especially painful gynecomastia, in men; because eplerenone has less of an action on the androgen receptor, it is a reasonable substitute in patients who experience this adverse effect.

Angiotensin Receptor Blockers

Mechanism of Action

Instead of inhibiting the production of angiotensin II through ACE, ARBs block the binding of angiotensin II to the AT1R. This pharmacologically distinct mechanism of action may be important because angiotensin II is also believed to be produced by other enzymes, such as chymase. ARBs do not inhibit kinase II or the breakdown of bradykinin, so they do not cause cough and cause less angioedema than do ACE inhibitors.

Clinical Benefits

When they are used as the sole agent in heart failure, ARBs produce benefits similar to those of ACE inhibitors. An ARB may be used as a substitute in patients who have cough or angioedema with an ACE inhibitor. When they are used in clinically effective doses, other adverse effects such as hypotension, renal dysfunction, and hyperkalemia are encountered as frequently as with an ACE inhibitor. As with an ACE inhibitor, the specific agents, dosing regimens, and target doses that were of demonstrable benefit in clinical trials are recommended (see E-Table 59-1).

In the broader population of patients with persistent heart failure symptoms (stage C or stage D, functional class II to IV) that can be treated with an ACE inhibitor, an ARB in combination with an ACE inhibitor (and β-blocker) further improves the left ventricular ejection fraction, relieves symptoms, reduces the risk for hospital admission for worsening heart failure, and can also reduce the risk for cardiovascular death (see Table 59-2), but the incremental benefits are not as great as adding a mineralocorticoid receptor antagonist (see later). Although a mineralocorticoid receptor antagonist is the preferred additional therapy because of its greater benefits (see above), an ARB is an alternative as the third disease-modifying drug (in addition to an ACE inhibitor and β-blocker) in patients who have persistent symptoms (stages C and D) and who do not tolerate a mineralocorticoid receptor antagonist. The efficacy and safety of the four-drug combination of an ACE inhibitor, β-blocker, ARB, and mineralocorticoid receptor antagonist are uncertain. Consequently, either a mineralocorticoid receptor antagonist or an ARB, but not both, should be added to an ACE inhibitor and a β-blocker in such patients.

The approach to initiation, titration, and monitoring of an ARB is similar to that of an ACE inhibitor (see Table 59-2). The adverse effects, with the exception of cough and angioedema, are similar. Use of multiple inhibitors of the renin-angiotensin-aldosterone system requires even more diligent monitoring, especially in patients at higher risk for uremia, hypotension, or hyperkalemia (i.e., patients 75 years of age and older or with a systolic blood pressure below 100 mm Hg, diabetes, or renal impairment) because combined treatment with an ACE inhibitor and an ARB significantly increases the risks for worsening renal function, hyperkalemia, and symptomatic hypotension. As with ACE inhibitors, β-blockers, and mineralocorticoid receptor antagonists, treatment with ARBs should be indefinite unless there is intolerance.

Neprilysin is an enzyme that breaks down natriuretic peptides and other vasoactive substances, including adrenomedullin and bradykinin. Inhibiting neprilysin augments the concentrations of these substances, which have vasodilator and natriuretic actions. Because neprilysin also degrades angiotensin II, a neprilysin inhibitor must be combined with an agent that blocks the renin-angiotensin system. Since ACE and neprilysin each breakdown bradykinin, inhibiting both enzymes leads to a significant increase in the risk of angioedema. For that reason, the angiotensin receptor neprilysin inhibitor (ARNI) sacubitril-valsartan (LCZ696) was developed. When 200 mg twice daily of sacubitril-valsartan was compared with enalapril 10 mg twice daily, the ARNI reduced cardiovascular mortality and heart failure hospitalization, as well as reducing other measures of progressive worsening of heart failure, including symptom deterioration.[3] Sacubitril-valsartan causes more hypotension and slightly more angioedema than enalapril. Sacubitril-valsartan is currently undergoing regulatory review in the US and Europe.

Ivabradine

Mechanism of Action

Ivabradine is the first of a new class of drugs developed to inhibit the mixed sodium-potassium channel or current (also known as the funny channel, abbreviated as I_f or I_{kf}) in the sinoatrial node and, in so doing, reduce heart rate. Reduction in heart rate is the only known cardiac action of ivabradine, which has this effect only in patients in sinus rhythm.

Clinical Benefits

Only one large RCT has examined the effect of ivabradine on mortality and morbidity in patients with symptomatic heart failure (NYHA class II to IV), a reduced ejection fraction (≤35%), and sinus rhythm with a rate of 70 beats per minute or greater. That trial showed that ivabradine improved symptoms and ejection fraction and reduced the risk for hospitalization for heart failure (but

not mortality) when added to an ACE inhibitor (or ARB), a β-blocker, and a mineralocorticoid receptor antagonist.

Practical Use

Although not approved for use in the United States, ivabradine should be considered where it is available in patients who have persistent symptoms (NYHA class II to IV) despite treatment with other disease-modifying therapies, that is, an ACE inhibitor (or ARB), β-blocker, and mineralocorticoid receptor antagonist, and who are in sinus rhythm with a heart rate of 70 beats per minute or greater. Treatment should be started at 5 mg twice daily, increased to 7.5 mg twice daily after 14 days unless the heart rate is 60 beats per minute or less, and reduced usually to 2.5 mg twice daily if the rate is less than 50 beats per minute. Symptomatic bradycardia and visual disturbance (phosphenes) are uncommon but require the dose be reduced or ivabradine be discontinued. Ivabradine also may increase the risk for atrial fibrillation, which should prompt discontinuation of the drug. Ivabradine should not be used in combination with agents that prolong the QT interval (e.g., amiodarone) and must be used cautiously with inhibitors (including grapefruit juice) or inducers of CYP3A4.

Digoxin

Mechanism of Action

Digitalis glycosides inhibit the cell membrane Na^+, K^+-ATPase pump, thereby increasing intracellular calcium and myocardial contractility. In addition, digoxin is thought to enhance parasympathetic and reduce sympathetic nervous activity as well as to inhibit renin release.

Clinical Benefits

Only one large RCT has examined the effects of starting (as opposed to withdrawing) digoxin on mortality and morbidity in patients with heart failure in sinus rhythm. In that trial, digoxin did not reduce mortality but did decrease the risk for admission to hospital for worsening heart failure when it was added to a diuretic and an ACE inhibitor. In patients in sinus rhythm, addition of digoxin is recommended only for those whose heart failure remains symptomatic despite standard treatment with a diuretic and three disease-modifying drugs, that is, an ACE inhibitor (or ARB), a β-blocker, and a mineralocorticoid receptor antagonist (or ARB). In patients with atrial fibrillation, digoxin may be used at an earlier stage if a β-blocker fails to control the ventricular rate during exercise; Chapter 64). Digoxin can also be used to control the ventricular rate when β-blocker treatment is being initiated or up-titrated.

If the effect of digoxin is needed urgently, loading with 10 to 15 μg/kg lean body weight, given in three divided doses 6 hours apart, may be used. The maintenance dose should be one third of the loading dose. Smaller maintenance doses (e.g., one fourth of the loading dose and not more than 62.5 μg/day) should be used in elderly patients and in patients with reduced renal function as well as in patients with a low body mass. Monitoring of the serum digoxin concentration is recommended because of the narrow therapeutic window. A steady state is reached 7 to 10 days after treatment is started; blood should be collected at least 6 hours (and ideally 8 to 24 hours) after the last dose. The currently recommended therapeutic range is 0.5 to 1.0 ng/mL).

Digoxin can cause anorexia, nausea, arrhythmias, confusion, and visual disturbances, especially if the serum concentration is above 2.0 ng/mL. Hypokalemia increases susceptibility to the adverse effects. The dose of digoxin should be reduced in elderly patients and in patients with renal dysfunction. Certain drugs increase serum digoxin concentration, including amiodarone.

Hydralazine and Isosorbide Dinitrate

Mechanism of Action

Hydralazine is a powerful direct-acting arterial vasodilator. Its mechanism of action is not understood, although it may inhibit enzymatic production of superoxide, which neutralizes nitric oxide and may induce nitrate tolerance. Nitrates dilate both veins and arteries, thereby reducing preload and afterload by stimulating the nitric oxide pathway and increasing cyclic guanosine monophosphate in vascular smooth muscle. Neither drug on its own nor any other direct-acting vasodilator has been demonstrated to be beneficial in heart failure.

Clinical Benefits

Although this combination has been known for some time to improve systolic function and probably to reduce death in class II to IV heart failure compared with placebo, head-to-head comparison showed that an ACE inhibitor is superior for improving survival. Nevertheless, on the basis of subgroup analyses suggesting that African Americans responded better to hydralazine and isosorbide dinitrate, a subsequent RCT showed that the addition of hydralazine and isosorbide dinitrate in African Americans, most of whom were receiving an ACE inhibitor and β-blocker and many of whom were taking spironolactone, further reduced mortality and hospital admissions for heart failure and improved quality of life. The combination of hydralazine and isosorbide dinitrate is recommended in African American patients with NYHA class III or IV symptoms and an ejection fraction of 45% or less despite treatment with standard disease-modifying drugs, that is, an ACE inhibitor, β-blocker, and mineralocorticoid receptor antagonist.

A fixed combination of 37.5 mg of hydralazine and 20 mg of isosorbide dinitrate was used in the trial; one tablet was given, and if tolerated, a second was given 12 hours later. One tablet was then prescribed three times daily for 3 to 5 days, at which point the dose was increased to the target maintenance of two tablets three times daily, that is, a daily dose of 225 mg hydralazine and 120 mg isosorbide dinitrate. Because of the limited inclusion criteria of this RCT, however, it is uncertain whether this combination of vasodilators is an effective addition in other populations of patients.

Practical Use

Other than for African Americans, the main indication for hydralazine and isosorbide dinitrate is as a substitute in patients with intolerance to an ACE inhibitor and an ARB. Hydralazine and isosorbide dinitrate should be used as additional treatment in African Americans and considered for other patients who remain symptomatic with other proven therapies. The main dose-limiting adverse effects of hydralazine and isosorbide dinitrate are headache and dizziness. A rare adverse effect of higher doses of hydralazine, especially in slow acetylators, is a systemic lupus erythematosus–like syndrome (Chapter 266).

Omega-3 Polyunsaturated Fatty Acids

In one trial, 1 gram of n-3 PUFA (850 to 852 mg eicosapentaenoic acid and docosahexaenoic acid as ethyl esters in the average ratio of 1 : 1.2) per day led to a small reduction in cardiovascular morbidity and mortality in patients with heart failure. Although this agent may have beneficial anti-inflammatory and antiarrhythmic effects, its current role in the treatment of heart failure is uncertain, especially because trials in survivors of acute myocardial infarction have not shown benefit (Chapter 73).

The aforementioned treatments are the only pharmacologic therapies shown to be of benefit in patients with heart failure and a reduced left ventricular ejection fraction. Other treatments have been tested in randomized trials and shown to have a neutral (e.g., amlodipine) or uncertain (e.g., bosentan and etanercept) effect on mortality and morbidity or to increase mortality (e.g., dronedarone, milrinone, flosequinan, vesnarinone, and moxonidine).

Other Pharmacologic Issues

Some therapies that are of proven value for cardiovascular conditions that underlie or are associated with heart failure are of uncertain benefit (antiplatelet treatment, Chapter 38) or do not improve outcomes (statins, Chapter 206) in patients with persistent, symptomatic heart failure. Vitamin K antagonists such as warfarin and other non–vitamin K oral anticoagulants that inhibit factor Xa or thrombin are indicated in patients with atrial fibrillation to reduce the risk for thromboembolism, provided patients have no contraindications to their use (Chapter 64). Anticoagulants may also be used in patients with evidence of intracardiac thrombus (e.g., detected during echocardiographic examination) or systemic thromboembolism. The many interactions of warfarin with other drugs, including some statins and amiodarone (Chapter 38), must always be considered. The non–vitamin K oral anticoagulants are contraindicated in patients with severe renal impairment and should be given at a reduced dose in patients with less severe impairment (Chapter 38), keeping in mind that no therapy is available to reverse the actions of these agents. Heparin prophylaxis (Chapter 38) against deep vein thrombosis is indicated when patients with heart failure are bed bound, such as during hospital admission. Vaccination against influenza and pneumococcal infection is advised (Chapter 18) in all patients with heart failure because these infections can lead to severe clinical deterioration.

Drugs to Use with Caution in Heart Failure

Patients with heart failure, especially if it is severe, often have renal and hepatic dysfunction, so any drug excreted predominantly by the kidneys or metabolized by the liver may accumulate (Chapter 29). Similarly, because of their extensive comorbidity, patients with heart failure are inevitably treated with multiple drugs, thereby increasing the risk for drug interactions.

Drugs that should be avoided, if possible, in heart failure include thiazolidinediones (because of the risk for fluid retention), most antiarrhythmic drugs (including dronedarone, although amiodarone and dofetilide may be used), most calcium-channel blockers (with the possible exception of amlodipine, although this drug may increase the risk for pulmonary edema), corticosteroids, NSAIDs, cyclooxygenase-2 inhibitors, many antipsychotics (e.g., clozapine), and antihistamines. The U.S. Food and Drug Administration recently raised the concern that the dopamine agonist pramipexole used to treat Parkinson disease (Chapter 409) also might increase the risk for developing heart failure. Metformin (because of the risk for lactic acidosis) should be used with caution. In a recent trial, saxagliptin was shown to increase the risk for developing heart failure in patients with diabetes. Some salt substitutes contain substantial amounts of potassium and must be used cautiously. Other dietary constituents (e.g., grapefruit and cranberry juice) and supplements such as St. John's wort can interact with drugs taken by patients with heart failure, especially warfarin and digoxin.

Organization of Care

Several studies have shown that organized, nurse-led, multidisciplinary care can improve outcomes in patients with heart failure, particularly by reducing recurrent hospital admissions. Thus it is recommended that all patients with

heart failure be enrolled in a disease management program.[1] The most successful disease management approach seems to involve education of the patients, their families, and caregivers about heart failure and its treatment (including flexible diuretic dosing and reinforcing the importance of adherence), recognizing (and acting on) early deterioration (dyspnea, sudden weight gain, edema), and optimizing proven pharmacologic treatments. A home-based rather than clinic-based approach may be best, although trials are needed to compare these types of interventions directly. Even telephone follow-up is of value.[2] New technology enabling noninvasive home telemonitoring of physiologic measures (e.g., heart rate and rhythm, blood pressure, temperature, respiratory rate, weight, and estimated body water content) and implanted devices, which collect similar data and may be interrogated remotely, are also being tested as aids to monitoring and management, but studies to date have not given consistent results. However, in one moderate-sized trial, use of an implanted sensor for the noninvasive measurement of pulmonary artery pressure resulted in a 30% reduction in the number of heart failure hospitalizations.[1] Despite the usefulness of brain-type natriuretic peptide (BNP) in the diagnosis of heart failure and as a prognostic measure, treatment guided by BNP levels has not been shown in randomized trials to be consistently better than standard, evidence-based care.

Education

Education of the patient, family, and caregivers is invaluable (Table 59-5). Self-detection of early signs and symptoms of deterioration provides for earlier intervention. Counseling on the proper use of therapies, with an emphasis on adherence, is critical.

Useful patient-oriented material is available from several reliable sources the Heart Failure Society of America (http://www.hfsa.org/hfsa-wp/wp/patient/education-modules/), American Heart Association (http://www.americanheart.org/presenter.jhtml?identifier=1486), National Heart, Lung and Blood Institute (http://www.nhlbi.nih.gov/health/dci/Diseases/Hf/HF_WhatIs.html), and the Heart Failure Association of the European Society of Cardiology (http://www.heartfailurematters.org/EN/Pages/index.aspx), and other organizations.

TABLE 59-5 TOPICS THAT SHOULD BE DISCUSSED WITH A PATIENT WITH HEART FAILURE AND WITH HIS OR HER FAMILY AND CAREGIVERS

GENERAL ADVICE
Explain what heart failure is and why symptoms occur
Causes of heart failure
How to recognize symptoms
What to do if symptoms occur
Self-weighing (to identify fluid retention)
RATIONALE FOR TREATMENTS
Importance of adhering to pharmacologic and nonpharmacologic (e.g., dietary) treatments
Smoking advice
Prognosis
DRUG COUNSELING
Rationale (i.e., benefits of individual drugs)
Dose and time of administration
Potential adverse affects (and what, if any, action to take)
What to do in case of missed or skipped doses
Self-management (e.g., flexible diuretic dosing)
REST AND EXERCISE
Rest
Exercise and activities related to work
Daily physical activity
Sexual activity
Rehabilitation
PSYCHOSOCIAL ASPECTS
Depression
Cognitive function
Social support
VACCINATIONS AND IMMUNIZATIONS
Travel
Driving
Dietary and social habits
Control sodium intake when necessary (e.g., some patients with severe heart failure)
Avoid excessive fluids in severe heart failure
Avoid excessive alcohol intake and illicit drugs

Modified from McMurray JJ, Adamopoulos S, Anker SD, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur Heart J.* 2012;33:1787-1847.

Medication Use Counseling

When appropriate, a patient should be taught how to adjust the dose of diuretic within individualized limits. The dose should be increased (or a supplementary diuretic added) if there is evidence of fluid retention (symptoms of congestion) and decreased if there is evidence of hypovolemia (e.g., increased thirst associated with weight loss or postural dizziness, especially during hot weather or an illness causing decreased fluid intake or sodium and water loss). If hypovolemia is more marked, the doses of other medications also will have to be reduced.

The expected effects, beneficial and adverse, of other drugs should also be explained in detail (e.g., possible association of cough with ACE inhibitor). It is useful to inform patients that improvement with many drugs is gradual and may become fully apparent only after several weeks or even months of treatment. It is also important to explain the need for gradual titration with ACE inhibitors, ARBs, and β-blocking drugs to a desired dose level, which again may take weeks or even months to achieve. Patients should be advised not to use NSAIDs without consultation and to be cautious about using herbal or other nonproprietary preparations (Chapter 39).

Adherence

Education and counseling of the patient, caregiver, and family promotes adherence, which is associated with better outcomes. Drug adherence can also be helped by home visits, specialized follow-up programs[3], and certain pharmacy aids, such as dose allocation (pill-organizing) boxes.

Lifestyle Modification

Exercise

Tailored, structured, supervised aerobic exercise is safe and improves functional capacity and quality of life in patients with heart failure (see E-Table 59-1). An appropriate exercise prescription may also reduce hospitalizations and mortality in patients with heart failure. Regular physical activity or exercise training if available is recommended for all patients with heart failure who are able to participate.[4]

Diet, Nutrition, and Alcohol

Most guidelines advocate avoidance of foods containing relatively high salt content in the belief that doing so may reduce the need for diuretic therapy. This recommendation is based on clinical experience, which suggests that excess sodium intake can be a precipitant of clinical decompensation. Some salt substitutes have a high potassium content, which can lead to hyperkalemia.

Restriction of fluid intake is indicated only during episodes of decompensation associated with peripheral edema or hyponatremia. In these situations, daily intake should be restricted to 1.5 to 2.0 L to help facilitate reduction in extracellular fluid volume and to avoid hyponatremia.

Reducing excessive weight will reduce the work of the heart and may lower blood pressure (Chapter 67). Conversely, malnutrition is common in severe heart failure, and the development of cardiac cachexia is an ominous sign. Reduced food intake is sometimes caused by nausea (e.g., related to digoxin use or hepatosplenic congestion) or abdominal bloating (e.g., due to ascites). In these cases, small frequent meals and high-protein and high-calorie liquids may be helpful. In severe decompensated heart failure, eating and bending may be difficult because of dyspnea.

Moderate alcohol intake is not thought to be harmful in heart failure, although excessive intake can cause cardiomyopathy and atrial arrhythmias in susceptible individuals. In patients with suspected alcoholic cardiomyopathy, abstinence from alcohol may improve cardiac function.

Smoking

Smoking causes peripheral vasoconstriction, which is detrimental in heart failure. Nicotine replacement therapy (Chapter 32) is believed to be safe in heart failure. The safety of bupropion in heart failure is uncertain, especially because it is known to increase blood pressure, and varenicline may increase cardiovascular risk.

Sexual Activity

Sexual activity need not be restricted in patients with compensated heart failure, although dyspnea may be limiting. In men with erectile dysfunction (Chapter 234), treatment with a cyclic guanine monophosphate phosphodiesterase type 5 inhibitor can be useful, but these drugs must not be taken within 24 hours of prior nitrate use, and nitrates must not be restarted for at least 24 hours afterward.

Driving

Patients with heart failure can continue to drive, provided their condition does not induce undue dyspnea, fatigue, or other incapacitating symptoms. Patients with recent syncope, cardiac surgery, percutaneous coronary intervention, or device placement may be restricted from driving, at least

temporarily, according to local regulations. Patients holding an occupational or commercial license may also be subject to additional restrictions.

Traveling

Short flights are unlikely to cause problems for a patient with compensated heart failure. Cabin pressure is generally maintained to provide an oxygen level no lower than equivalent to 6000 feet above sea level, which should be well tolerated in patients without severe pulmonary disease or pulmonary hypertension. Longer journeys may cause limb edema and dehydration, thereby predisposing to venous thrombosis. Adjustment of the dose of diuretics and other treatments should be discussed with the patient wishing to travel to a warm climate or a country where the risk for gastroenteritis is high. It is also advisable for heart failure patients to carry a list of medications and contact information for their health care provider.

Comorbidity

Comorbid conditions, which are common and important in patients with heart failure, may be due to the underlying cardiovascular disease that caused or contributed to heart failure (e.g., hypertension, coronary artery disease, diabetes mellitus), may arise as a complication of heart failure (e.g., arrhythmias), or can result from an adverse effect of treatment given for heart failure (e.g., gout). The exact causes of other comorbidities in heart failure, such as diabetes (Chapter 229), depression (Chapter 397), sleep apnea (Chapter 100), renal dysfunction (Chapter 130), and anemia (Chapter 158), are complex and uncertain. These and other comorbid conditions, such as chronic obstructive pulmonary disease and asthma, are important because they are a major determinant of prognosis and may limit the use of certain treatments for heart failure (e.g., renal dysfunction limiting use of ACE inhibitors or asthma limiting β-blockers) and because treatment of comorbidities may affect the stability of heart failure (e.g., NSAIDs needed for rheumatic conditions can cause salt and water retention and renal dysfunction). Both prevention (e.g., diabetes mellitus) and treatment (e.g., anemia) of comorbidities are being evaluated as a potential new therapeutic goal in heart failure.

Angina

β-Blockers are of benefit in both angina (Chapter 71) and heart failure. Similarly, ivabradine, which reduces heart rate by inhibiting the I_f current in the sinus node, is also beneficial in both angina and heart failure. Nitrates relieve angina but on their own are not of proven value in chronic heart failure. Calcium-channel blockers should generally be avoided in heart failure because they have a negative inotropic action and cause peripheral edema; only amlodipine has been shown to have no adverse effect on survival, but it may increase the risk for pulmonary edema. Trimetazidine, ranolazine, and nicorandil are antianginal drugs that are available in certain countries; their safety in patients with heart failure is uncertain. Percutaneous and surgical (Chapter 74) revascularization is also of value in relieving angina in selected patients with heart failure (see later). Coronary artery bypass grafting may reduce the risk for death from cardiovascular causes and cardiovascular hospitalization (including heart failure hospitalization) in selected heart failure patients with angina and a reduced ejection fraction (see later).

Atrial Fibrillation

Atrial fibrillation (Chapter 64) may be the cause of or a consequence of heart failure in a patient presenting with atrial fibrillation and a rapid ventricular rate, and the distinction can be difficult, especially because prolonged atrial fibrillation may lead to a rate-related cardiomyopathy. Thyrotoxicosis (Chapter 226) and mitral valve disease (Chapter 75), especially stenosis, must be excluded. Alcohol abuse should also be considered. β-Blockers and digoxin are given to control the ventricular rate. The patient should be supervised closely after the initiation of these treatments because underlying sinus node dysfunction may raise the risk for bradycardia. Unless the patient presents emergently with symptoms or signs of heart failure, myocardial ischemia, or hypertension, there is little or no evidence to support a strategy of restoring sinus rhythm rather than controlling the ventricular rate in most patients with heart failure (Chapter 64).[5] Atrioventricular node ablation and pacing may be required to control ventricular rate (Chapter 66). Catheter ablation can cure atrial fibrillation in some patients with heart failure, but the rate of success and its long-term benefits remain uncertain (Chapter 66). There is a strong indication for thromboembolism prophylaxis with warfarin in patients with heart failure and atrial fibrillation (Chapter 64).

Asthma and Reversible Airways Obstruction

Asthma is a contraindication for use of a β-blocker, but most patients with chronic obstructive pulmonary disease (Chapter 88) can tolerate a β-blocker. Pulmonary congestion can mimic chronic obstructive pulmonary disease. Systemic administration of a corticosteroid to treat reversible airways obstruction may cause sodium and water retention and exacerbate heart failure, whereas inhalation therapy is better tolerated.

Diabetes Mellitus

Diabetes mellitus is discussed in detail elsewhere (Chapter 229). The prevalence and incidence of diabetes mellitus are high in heart failure, and the risk for development of type 2 diabetes may be reduced by ACE inhibitors and ARBs. β-Blocker treatment is not contraindicated and is of benefit in patients with diabetes and heart failure. Thiazolidinediones cause sodium and water retention, may lead to decompensation, and are not recommended in patients with or at risk for heart failure. Metformin may cause lactic acidosis and is not recommended in patients with severe heart failure. Saxagliptin increases the risk for developing heart failure in patients with diabetes.

Abnormal Thyroid Function

Both thyrotoxicosis and hypothyroidism can cause heart failure (and thyrotoxicosis can cause atrial fibrillation, which may precipitate heart failure). Amiodarone can also induce both hypothyroidism and hyperthyroidism, the latter being particularly difficult to diagnose. The risk for thyroid dysfunction may be less with the related antiarrhythmic agent dronedarone, but dronedarone increases mortality in severe heart failure and should be avoided in patients with stage C or D heart failure or recently decompensated heart failure.

Gout

Hyperuricemia and gout (Chapter 273) are common in heart failure and can be caused or aggravated by diuretic treatment. Allopurinol may prevent gout, and acute attacks are better treated with colchicine, oral steroids, or intra-articular steroids rather than by an NSAID.

Renal Dysfunction

Most patients with heart failure have a reduced glomerular filtration rate. ACE inhibitors, ARBs, and mineralocorticoid receptor antagonists often cause a further small reduction in glomerular filtration rate and rise in serum blood urea nitrogen and creatinine levels, which, if limited, should not lead to discontinuation of treatment. Marked increases in blood urea nitrogen and creatinine, however, should prompt consideration of underlying renal artery stenosis (Chapter 125). Renal dysfunction may also be caused by sodium and water depletion, leading to relative hypovolemia (e.g., due to excessive diuresis, diarrhea, and vomiting) or hypotension. Nephrotoxic agents such as NSAIDs are also a common cause of renal dysfunction in heart failure.

Prostatic Obstruction

For prostatic disease (Chapter 129), a 5α-reductase inhibitor may be preferable to an α-adrenoceptor antagonist, which can cause hypotension and salt and water retention. A cyclic guanine monophosphate phosphodiesterase type 5 inhibitor is an alternative, but it cannot be used in patients taking nitrates. Prostatic obstruction should also be considered in male patients with deteriorating renal function.

Anemia

A normocytic, normochromic anemia (Chapter 158) is also common in heart failure, in part because of the high prevalence of renal dysfunction. Malnutrition and blood loss may also contribute. Intravenous iron treatment with 200 mg of ferric carboxymaltose improves quality of life and reduces symptoms in patients with NYHA class II to III heart failure, a reduced ejection fraction, and iron deficiency without adverse effects. By comparison, erythropoiesis-stimulating agent darbepoetin is not of benefit.

Depression

Depression (Chapter 397) is common in patients with heart failure, perhaps partly owing to disturbance of the hypothalamic-pituitary axis and other neurochemical pathways but also as a result of social isolation and the adjustment to chronic disease. Depression is associated with worse functional status, reduced adherence to treatment, and poor clinical outcomes. Both psychosocial interventions and pharmacologic treatment are helpful. Selective serotonin reuptake inhibitors are believed to be the best tolerated pharmacologic agents, whereas tricyclic antidepressants should be avoided because of their anticholinergic actions and potential to cause arrhythmias.

Cancer

Many anticancer drugs, particularly anthracyclines, cyclophosphamide, and trastuzumab (Herceptin), can cause myocardial damage and heart failure, as can mediastinal radiotherapy. Pericardial constriction can be a result of previous radiotherapy, and malignant pericardial involvement can cause effusion and tamponade (Chapter 77).

Devices and Surgery

Implantable Cardioverter-Defibrillators

About half of patients with heart failure die suddenly, mainly as the result of a ventricular arrhythmia. The relative risk for sudden death, as opposed to death from progressive heart failure, is greatest in patients with milder symptoms. In patients with more advanced heart failure, progressive pump failure deaths are relatively more common. Antiarrhythmic drugs have not been shown to improve survival in heart failure, but ICDs (Chapter 66) reduce the risk for death in selected patients after myocardial infarction (Chapter 73) and improve survival in patients with class II to III heart failure and systolic dysfunction who were otherwise treated with optimal medical therapy. All patients with class II or III heart failure, irrespective of etiology, and a left

ventricular ejection fraction remaining at 35% or less despite at least 3 months of treatment with disease-modifying therapy (i.e., an ACE inhibitor [or ARB], β-blocker, and mineralocorticoid receptor antagonist [or ARB]) should be considered for an ICD provided they have no other conditions greatly limiting life expectancy (i.e., have an anticipated survival of at least a year) or quality of life.

Cardiac Resynchronization Therapy

About 30% of patients with heart failure have substantial prolongation of the QRS duration on the surface electrocardiogram, which is a marker of abnormal electrical activation of the left ventricle causing dyssynchronous contraction, less efficient ventricular emptying, and, often, mitral regurgitation. Atrioventricular coupling may also be abnormal, as reflected by a prolonged PR interval, as may interventricular synchrony. Cardiac resynchronization therapy (CRT) with atrial-biventricular or multisite pacing optimizes atrioventricular timing and improves synchronization of cardiac contraction. In symptomatic patients (NYHA class II to IV) who are in sinus rhythm, have marked systolic dysfunction (left ventricular ejection fraction ≤35%), and have a wide QRS, the addition of CRT to optimal medical therapy and an ICD improves pump function, reduces mitral regurgitation, relieves symptoms, and significantly prolongs exercise capacity. CRT also substantially reduces the risk for death and for hospital admission for worsening heart failure in such patients (see E-Table 59-1). Many other outcome measures, including quality of life, are also improved. The greatest benefit appears to be in patients with a left bundle branch block (LBBB) morphology and in patients with mild symptoms (NYHA class II). Whether CRT is beneficial in patients in atrial fibrillation or with non-LBBB QRS widening is uncertain. All patients in sinus rhythm with persistent symptoms (NYHA class II to IV) and an ejection fraction of 35% or less despite optimal disease-modifying medical therapy (ACE inhibitor, β-blocker, and mineralocorticoid receptor antagonist) should be considered for CRT if they have a QRS duration of 130 msec or longer, especially if they have LBBB.[6]

Surgery

With the exception of cardiac transplantation and ventricular assist devices, there are no generally accepted criteria for surgical intervention. Use of operative procedures is variable among centers and greatly dependent on local experience and expertise. Expert imaging and detailed hemodynamic and functional assessments are usually required when any patient with heart failure is considered for surgery, and close liaison between the relevant experts in these fields is essential. The collective expertise in surgical centers is often used to make highly individualized decisions about whether to operate and what procedures will be attempted. "Established" operative treatments for patients with heart failure include coronary artery bypass grafting, surgery for mitral valve incompetence, surgery or percutaneous interventions for aortic valve stenosis (Chapter 75), implantation of ventricular assist devices, and heart transplantation.

A recent trial showed no benefit (in symptoms or rates of death or hospitalization for cardiac causes) of surgical ventricular reconstruction. Cardiomyoplasty and partial left ventriculectomy are other operations for heart failure now thought to be without benefit.

Percutaneous Coronary Intervention or Coronary Artery Bypass Grafting

Percutaneous coronary intervention or coronary artery bypass grafting (Chapter 74), as appropriate, is indicated for relief of angina. The extent of ischemia and residual myocardial viability can be determined by noninvasive assessments such as dobutamine echocardiography (Chapter 55), magnetic resonance imaging (Chapter 56), and positron emission tomographic scanning (Chapter 56) in patients with impaired left ventricular ejection fraction. Coronary artery bypass grafting may reduce the risk for cardiovascular death and hospitalization (including heart failure hospitalization) in selected heart failure patients with angina, a reduced ejection fraction, and two-vessel or three-vessel coronary disease. Coronary artery bypass grafting is therefore recommended in patients who have symptomatic heart failure (NYHA class II to III), angina pectoris, and suitable coronary anatomy and who are otherwise fit for surgery.

Whether coronary artery bypass grafting is beneficial in patients who have coronary artery disease but who do not have angina is less certain. Many physicians and surgeons consider revascularization in such asymptomatic patients only if ischemia affects a substantial area of myocardium as documented by noninvasive imaging. However, the hypothesis that improvement of coronary blood flow to viable but noncontracting ("hibernating") myocardium can improve ventricular function and clinical outcomes remains to be proved.

Cardiac Transplantation

Cardiac transplantation (Chapter 82) remains the most accepted surgical intervention in end-stage heart failure. Selection criteria usually focus on patients with refractory heart failure, that is, those with severe symptoms and functional limitations (peak oxygen consumption of less than 12 mL/kg per minute), as well as a particularly worrisome clinical course and prognosis attributed to their cardiac condition. These patients are often dependent on intravenous inotropic agents and mechanical support.

Mechanical Circulatory Support

Given the scarcity of organ donors, mechanical circulatory support using a left ventricular (or biventricular) assist device may be used as a "bridge to transplantation" or even as a permanent, definitive, procedure ("destination therapy") for some patients with advanced or end-stage heart failure. A pulsatile volume-displacement left ventricular assist device can provide a short but significant prolongation of survival in patients who have end-stage heart failure and are ineligible for transplantation (see E-Table 59-1), but the rates of bleeding, infective and thrombotic complications, and mechanical dysfunction necessitating repeat surgery were high with this older device. In patients with end-stage heart failure ineligible for transplantation, a newer continuous-flow device is significantly better than this older device in terms of 2-year survival without repeat device surgery or disabling stroke (46% versus 11%). Because not every hospital could or should be expected to offer all these levels of support for patients with advanced heart failure, there is a general recognition that such services should be concentrated in a limited number of tertiary centers.[7] Centers implanting these devices use criteria such as persistent (>2 months) severe symptoms despite optimal drug and device therapy and other features placing patients at high-risk for death (e.g., left ventricular ejection fraction <25%, three or more heart failure hospitalizations in the prior 12 months, peak oxygen consumption <12 mL/kg per minute, dependence on intravenous inotropic therapy, progressive end-organ dysfunction, and deteriorating right ventricular function) to decide who should considered for mechanical circulatory support.

Heart Failure with Preserved Left Ventricular Ejection Fraction (Diastolic Dysfunction)

Although all patients with symptomatic heart failure share a constellation of signs and symptoms, impaired physical capacity, and reduced quality of life, some have a preserved left ventricular ejection fraction (generally >40 or 50%), and many are thought to have diastolic dysfunction (Chapter 58).[8] Diastolic heart failure often has a cause different from that of systolic heart failure and a better survival rate (Chapters 53, 58, and 60), but sometimes it is an early manifestation of what will evolve into heart failure with a reduced left ventricular ejection fraction. The distinction is important, however, because most of the RCTs that generated the evidence for treatment of heart failure included only patients with reduced left ventricular ejection fractions (see E-Table 59-1). Treatment of the underlying cardiovascular and other disorders that contribute to symptomatic stage C and stage D of heart failure with preserved left ventricular ejection fraction, such as hypertension, myocardial ischemia, and diabetes, is critical and is as for stages A and B (see earlier). In patients with atrial fibrillation, control of the ventricular rate with a β-blocker or a rate-limiting calcium-channel blocker (or restoration of sinus rhythm) may be particularly important (Chapter 64). Diuretics are used empirically to treat sodium and water retention, according to the same principles as in heart failure with reduced left ventricular ejection fraction. In one trial of patients with a left ventricular ejection fraction greater than 40% (mean 54%), treatment with the ARB candesartan decreased the risk for hospital admission for heart failure but did not improve survival or the composite outcome of cardiovascular death or hospital admission for worsening heart failure. In a subsequent study of patients with a left ventricular ejection fraction of 45% or greater (mean 60%), the ARB irbesartan had no beneficial effect, raising the possibility that the benefit of candesartan in the earlier trial was largely in patients with borderline systolic dysfunction (i.e., a left ventricular ejection fraction of 40 to 50%). In a more recent study of patients with a left ventricular ejection fraction or 45% or greater (mean 56%), treatment with the mineralocorticoid receptor antagonist spironolactone decreased the risk for hospital admission for heart failure but did not improve survival or the composite outcome of cardiovascular death or hospital admission for worsening heart failure. Smaller studies in patients in sinus rhythm have shown that the calcium-channel blocker verapamil may improve symptoms and exercise capacity in patients with heart failure and preserved left ventricular ejection fraction, possibly by reducing heart rate and thereby increasing the duration of diastolic left ventricular filling as well as by directly enhancing myocardial relaxation. There are, however, no current RCTs in which this drug decisively reduced mortality or morbidity in patients with heart failure and preserved left ventricular ejection fraction, so treatment currently is aimed at relieving symptoms.

Heart Failure Due to Valvular Heart Disease

Heart failure also can arise as a result of regurgitant and stenotic valve disease (Chapter 75). It can sometimes be difficult to determine whether mitral regurgitation is primary or secondary in a patient with heart failure and left ventricular dilation, although a prior history of known valve disease or rheumatic fever may suggest a primary valve problem. The objective of treatment of primary valve disease is the prevention of heart failure by surgical repair or replacement of the diseased valve or valves (Chapter 75). The development of overt heart failure is an ominous sign, sometimes requiring

emergent valve replacement (e.g., aortic stenosis) but sometimes indicating that valve replacement may not be possible (e.g., because of severe pulmonary hypertension).

Aortic Stenosis

Evaluation of the aortic valve (Chapter 75) can be difficult in patients with poor left ventricular systolic function. Such patients may have insufficient cardiac output to generate a high gradient across even a severely stenotic valve. Conversely, a calcified and degenerate but nonstenotic aortic valve may appear stenosed simply because it does not open normally in patients with very low cardiac output. A calculated valve area provides a better assessment of the severity of aortic stenosis in these patients. Stress echocardiography (Chapter 55) may help assess the potential for ventricular recovery after relief of aortic stenosis. Consideration should be given as to whether concomitant myocardial ischemia from coronary artery disease may also be contributing to a reversible depression of systolic function. Transcatheter valve replacement is a valuable technique for patients who have aortic stenosis but are at very high risk for open valve replacement.

Mitral Regurgitation

Mitral regurgitation can be a primary cause or a secondary manifestation in a patient with heart failure and left ventricular dilation (Chapter 75). Surgery sometimes will result in clinical improvement, but some patients with advanced left ventricular dysfunction will not achieve substantial benefit (e.g., mitral valve surgery in a patient with long-standing severe mitral regurgitation). Valve repair or annuloplasty may, however, be beneficial in carefully selected patients with secondary mitral regurgitation caused by or exacerbated by left ventricular dilation. It is not known whether valve repair is preferable to valve replacement. The role of percutaneous mitral valve repair is still uncertain and is under investigation.

Heart Failure Due to Nonischemic Dilated Cardiomyopathy

Patients with heart failure and normal coronary arteries should be evaluated for possible reversible causes. Untreated hypertension is now an unusual cause of dilated cardiomyopathy in the United States, but hypertension was once a leading cause in the United States and still remains a major consideration in many parts of the world. Infiltrative cardiomyopathies (e.g., hemochromatosis, amyloid, sarcoid) and arteritides sometimes have specific recommended therapies (Chapters 60, 95, 188, and 212). Chagas disease (Chapter 347) must be considered in patients from endemic areas. Alcohol and other toxins (e.g., chemotherapeutic agents) are other recognized causes of dilated cardiomyopathy. Dilated cardiomyopathy can also develop in the peripartum period. Most cases of nonischemic dilated cardiomyopathy are usually labeled "idiopathic" (i.e., no specific etiology can be determined), although many may have a genetic origin, especially if there is a positive family history. Irrespective of etiology, nonischemic dilated cardiomyopathy should be treated in the same way as dilated ischemic cardiomyopathy.

Heart Failure Due to Hypertrophic Cardiomyopathy

Heart failure can arise in patients with hypertrophic cardiomyopathy because of predominant diastolic dysfunction, associated mitral incompetence, or the development of systolic dysfunction. The management of hypertrophic cardiomyopathy and its complications is often very different from the management of dilated cardiomyopathy (Chapter 60), thereby underscoring the value of echocardiography in the evaluation of the patient with heart failure.

Acute Decompensated Heart Failure and Pulmonary Edema

Patients presenting with acute heart failure include those who develop heart failure de novo as a consequence of another cardiac event, usually a myocardial infarction, and those who present for the first time with decompensation of previously asymptomatic and often unrecognized cardiac dysfunction (patients previously in stage B, a transition with profound prognostic implications).[9] However, because of frequent recurrences, most episodes of acute decompensation occur in patients with established, chronic heart failure that has worsened as a result of the unavoidable natural progression of the syndrome, with an intercurrent cardiac (e.g., arrhythmia) or noncardiac (e.g., pneumonia) event, or as a consequence of an avoidable reason, such as nonadherence with treatment or use of an agent that can alter renal function. Although it is not always identified, searching for a reversible precipitant is an important aspect of the initial therapy plan (Table 59-6).

Most patients with acute heart failure require admission to the hospital, especially if pulmonary edema is present. In contrast to chronic heart failure, data from RCTs generally are not available to guide effective therapy for patients with acute decompensated heart failure. The principal goals of management of this heterogeneous group of patients are to relieve symptoms, the most important of which is extreme dyspnea, and to maintain or to restore vital organ perfusion. An intravenous bolus or infusion of a loop diuretic and, in hypoxemic patients, oxygen are the key first-line treatments.

TABLE 59-6 SOME COMMON PRECIPITATING CAUSES OF HEART FAILURE

Myocardial ischemia or infarction
Atrial fibrillation or other supraventricular tachycardias
Uncontrolled hypertension
Valvar disease
Ventricular tachycardia
Pulmonary embolism
Pericardial disease
Sepsis
Anemia
Poor dietary or medical adherence
Adverse drug effects
Hyperthyroidism or hypothyroidism

From Kimmelstiel CD, DeNofrio D, Konstam MA. Heart failure. In: Wachter RM, Goldman L, Hollander H, eds. *Hospital Medicine.* 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2005:360.

A small RCT suggested that high-dose diuretic (more than 2.5 times previous oral dose) resulted in greater relief of dyspnea and congestion compared with low-dose diuretic (same intravenous dose as prior oral dose), but at the expense of more, albeit transient, renal dysfunction. An intravenous opiate may also be given in selected patients to relieve anxiety and distress. Noninvasive ventilation using a tight-fitting mask to provide positive-pressure ventilation reduces respiratory distress and metabolic disturbances more rapidly than standard oxygen therapy but has not reduced short-term mortality. Intravenous infusion of a nitrate (e.g., continuous intravenous infusion of 20 to 200 μg/mm of nitroglycerin, titrated according to the symptomatic response and hemodynamic measurements, particularly arterial blood pressure) may also be valuable in patients with hypertension or myocardial ischemia (Fig. 59-3). Intravenous nesiritide (human B-type natriuretic peptide as a 2.0-μg/kg intravenous bolus followed by a continuous intravenous infusion of 0.01 to 0.03 μg/kg/mm titrated according to the symptomatic response and hemodynamic measurements, particularly arterial blood pressure) can reduce the pulmonary capillary wedge pressure more promptly than intravenous nitroglycerin but has minimal effect on dyspnea and does not improve other clinical outcomes. In volume-overloaded patients with severe heart failure unresponsive to diuretics, ultrafiltration is an option at specialized centers, although it was not superior to intensified pharmacologic therapy in a recent trial.

In patients with marked hypotension or other evidence of organ hypoperfusion, an inotropic agent such as dobutamine (continuous intravenous infusion of 2.5 to 25 μg/kg/min, titrated according to hemodynamic and heart rate response and induction of arrhythmias or myocardial ischemia) or a phosphodiesterase inhibitor (e.g., milrinone) should be considered, although neither treatment has ever been shown to reduce in-hospital deaths. In some countries, the calcium sensitizer levosimendan is also available for use in these patients. In general, potent inotropic agents should be used in a cardiac monitored setting at the lowest clinically effective dose and for the shortest duration possible (Chapter 107). Although low-dose dopamine (intravenous infusion of 2.5 μg/kg/mm) is often administered in an attempt to improve diuresis and renal function, such benefits were not confirmed in a recent RCT. In one trial of patients admitted to the hospital for acute heart failure and systolic blood pressure of 125 mm Hg or greater, serelaxin (recombinant human relaxin-2, at 30 μg/kg per day intravenously for 48 hours) improved dyspnea and reduced death at 6 months by one third, but this therapy has not been approved for use in the United States or the European Union. Patients with severe hyponatremia may benefit from the arginine vasopressin antagonist tolvapan (15 to 60 mg orally once daily).

In more critically ill patients, mechanical circulatory support (e.g., with an intra-aortic balloon pump) may also be considered (Chapter 107). The aim of treatment is to support the patient's circulation and vital organ function until either the patient's own heart recovers or a definitive operative procedure can be performed (e.g., transplantation or implantation of a long-term ventricular assist device).

In patients admitted to the hospital, discharge planning and subsequent management to reduce the risk for readmission are important.[10,11] Ideally, an effective oral diuretic regimen should have been identified, and fluid-volume and biochemical stability should have been achieved. This optimization of volume status and development of a stable oral regimen before discharge is thought to reduce the risk for early readmission. Treatment with an ACE inhibitor (or ARB), β-blocker, and mineralocorticoid receptor antagonist (or ARB), as appropriate, should also be started and titrated in the stabilized patient before discharge. Outpatient follow-up should be arranged to ensure that any of those treatments that have not been started before discharge are initiated

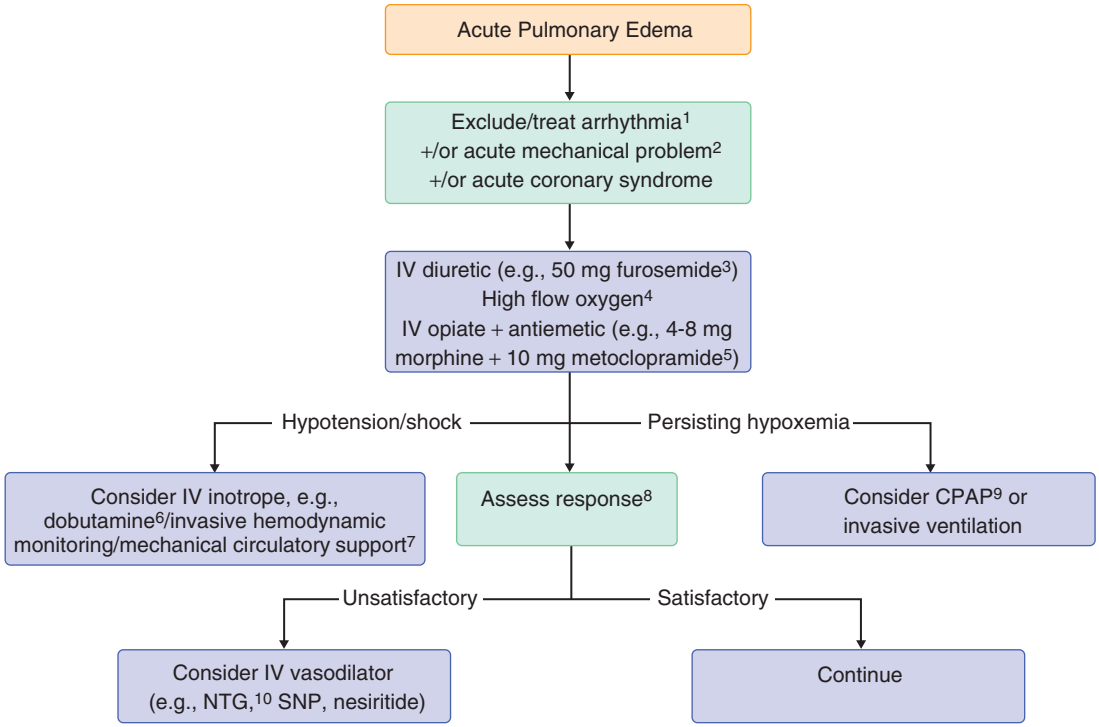

[1] Causal arrhythmia (e.g., ventricular tachycardia). It can be difficult to determine whether atrial fibrillation is a primary cause of acute pulmonary edema or secondary to it. An ECG is an essential investigation.

[2] Acute mechanical problems include ventricular septal rupture and mitral valve papillary muscle rupture. Mechanical support (e.g., an intra-aortic balloon pump) and urgent surgery should be considered. An echocardiogram should be performed as soon as possible, especially in a patient without a prior diagnosis of heart failure/other relevant heart disease (e.g., prior myocardial infarction or valve disease).

[3] Dose of diuretic depends on prior diuretic use and renal function—a lower dose may suffice if preserved renal function and no prior diuretic use.

[4] Oxygen causes an increase in systemic vascular resistance and a reduction in heart rate and cardiac output and should only be administered to patients with hypoxemia.

[5] Consider if patient agitated/distressed/in pain; may cause respiratory depression and dose should be reduced in very elderly.

[6] An intravenous infusion of dobutamine may be started at a dose of 2.5 µg/kg/min, doubling every 15 minutes according to response and tolerability (dose titration usually limited by excessive tachycardia, arrhythmias, or ischemia). A dose above 20 µg/kg/min is rarely needed.

[7] E.g., an intra-aortic balloon pump

[8] Improvement in symptoms and peripheral perfusion and adequate urine output—patient should be monitored closely, and usually a response will occur within 30 minutes. Bladder catheterization may help in monitoring urine output.

[9] Continuous positive airways pressure (CPAP) is valuable in severe pulmonary edema, especially if associated with hypoxemia. Endotracheal intubation and invasive mechanical ventilation should be considered in patients with persisting hypoxemia and physical ventilatory exhaustion.

[10] If systolic blood pressure is adequate (>100 mm Hg), an intravenous infusion of nitroglycerin (NTG) can be considered. Start at a dose of 10 µg/min and double every 10 minutes according to response and tolerability (usually dose up-titration is limited by hypotension). A dose of more than 100 µg/min is rarely needed.

FIGURE 59-3. Approach to the patient with acute pulmonary edema. SNP = sodium nitroprusside.

after discharge and that the dose of each drug is increased, as tolerated, to the appropriate target.

Outpatient Follow-Up

The key to successful follow-up is the careful tracking of clinical symptoms and the patient's weight, which often involves interviewing not only the patient but also family members, who may be more aware of changes in status than the patient is (see previous section, Organization of Care). Continuity of care and seamless transitions from the inpatient to the outpatient setting are crucial aspects of optimal management. Patients with advanced heart failure and patients requiring frequent hospitalization require special attention. Programs that provide telephone-based tracking of daily weights and symptoms can detect deterioration in time to intervene before the need for hospitalization. Although these programs may be costly, several evaluations have found them to be cost effective. Because the care of these patients requires considerable experience and expertise, specialized disease management programs and clinics have been developed and may provide additional benefit compared with traditional care.

PROGNOSIS

The prognosis of patients with heart failure is poor despite advances in therapy. Of patients who survive the acute onset of heart failure, only 35% of men and 50% of women are alive after 5 years. Although it is difficult to predict prognosis in individual patients, patients with symptoms at rest (class IV) have a 30 to 50% annual mortality rate, patients who are symptomatic with mild activity (class III) have mortality rates of 10 to 20% annually, and patients with symptoms only with moderate activity (class II) have a 5 to 10% annual mortality rate. Mortality rates are higher in older patients, men, and patients with a reduced left ventricular ejection fraction or underlying coronary heart disease.

End-of-Life Considerations

Although predicting the trajectory of illness in patients with advanced heart failure is notoriously difficult, it is often apparent when a patient has progressed to end-stage heart failure, commonly associated with concomitant renal failure. In these circumstances, the expertise of the palliative care team

may be especially helpful (Chapter 3).[12] Useful websites providing information on palliative care relevant to heart failure are available (http://www.goldstandardsframework.nhs.uk/ and http://www.palliativecarescotland.org.uk/content/publications/HF-final-document.pdf). Medications such as parenteral opiates (with an antiemetic) and benzodiazepines may be particularly helpful in relieving dyspnea, anxiety, and pain that arises from ascites, hepatic congestion, lower limb edema, and pressure points. At this stage in the patient's illness, it may be appropriate to discuss withdrawal of conventional treatment, deactivation of an ICD to avoid undesired and unpleasant electrical discharges, and a do-not-resuscitate order if the patient and others involved in the patient's care agree that comfort care is appropriate. Hospice care may be chosen by some at this point.

FUTURE DIRECTIONS

Multiple experimental approaches employing regenerative biology are under investigation, but none has yet generated sufficient data to warrant clinical use. Device therapy is likely to advance, and potential development of the total artificial heart remains a long-term goal.

Grade A References

A1. Sciarretta S, Palano F, Tocci G, et al. Antihypertensive treatment and development of heart failure in hypertension: a Bayesian network meta-analysis of studies in patients with hypertension and high cardiovascular risk. *Arch Intern Med*. 2011;171:384-394.

A2. Pfeffer MA, McMurray JJ, Velazquez EJ, et al. Valsartan, captopril, or both in myocardial infarction complicated by heart failure, left ventricular dysfunction, or both. *N Engl J Med*. 2003;349:1893-1906.

A3. Dargie HJ. Effect of carvedilol on outcome after myocardial infarction in patients with left-ventricular dysfunction: the CAPRICORN randomised trial. *Lancet*. 2001;357:1385-1390.

A4. Zannad F, McMurray JJ, Krum H, et al. Eplerenone in patients with systolic heart failure and mild symptoms. *N Engl J Med*. 2011;364:11-21.

A5. McMurray JJ, Packer M, Desai AS, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. *N Engl J Med*. 2014;371:993-1004.

A6. Swedberg K, Komajda M, Bohm M, et al. Ivabradine and outcomes in chronic heart failure (SHIFT): a randomised placebo-controlled study. *Lancet*. 2010;376:875-885.

A7. Taylor AL, Ziesche S, Yancy C, et al. Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. *N Engl J Med*. 2004;351:2049-2057.

A8. Rizos EC, Ntzani EE, Bika E, et al. Association between omega-3 fatty acid supplementation and risk of major cardiovascular disease events: a systematic review and meta-analysis. *JAMA*. 2012;308:1024-1033.

A9. Anderson JL, Halperin JL, Albert NM, et al. Management of patients with atrial fibrillation compilation of 2006 ACCF/AHA/ESC and 2011 ACCF/AHA/HRS recommendations: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2013;61:1935-1944.

A10. Camm AJ, Lip GY, De Caterina R, et al. 2012 Focused update of the ESC guidelines for the management of atrial fibrillation: an update of the 2010 ESC guidelines for the management of atrial fibrillation. Developed with the special contribution of the European Heart Rhythm Association. *Eur Heart J*. 2012;33:2719-2747.

A11. Scirica BM, Bhatt DL, Braunwald E, et al. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *N Engl J Med*. 2013;369:1317-1326.

A12. Konstam MA. Home monitoring should be the central element in an effective program of heart failure disease management. *Circulation*. 2012;125:820-827.

A13. Desai AS. Home monitoring heart failure care does not improve patient outcomes: looking beyond telephone-based disease management. *Circulation*. 2012;125:828-836.

A14. Abraham WT, Adamson PB, Bourge RC, et al. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. *Lancet*. 2011;377:658-666.

A15. Ponikowski P, van Veldhuisen DJ, Comin-Colet J, et al. Beneficial effects of long-term intravenous iron therapy with ferric carboxymaltose in patients with symptomatic heart failure and iron deficiencydagger. *Eur Heart J*. 2014; [Epub ahead of print].

A16. Swedberg K, Young JB, Anand IS, et al. Treatment of anemia with darbepoetin alfa in systolic heart failure. *N Engl J Med*. 2013;368:1210-1219.

A17. Cleland JG, Abraham WT, Linde C, et al. An individual patient meta-analysis of five randomized trials assessing the effects of cardiac resynchronization therapy on morbidity and mortality in patients with symptomatic heart failure. *Eur Heart J*. 2013;34:3547-3556.

A18. Goldenberg I, Kutyifa V, Klein HU, et al. Survival with cardiac-resynchronization therapy in mild heart failure. *N Engl J Med*. 2014;370:1694-1701.

A19. Ruschitzka F, Abraham WT, Singh JP, et al. Cardiac-resynchronization therapy in heart failure with a narrow QRS complex. *N Engl J Med*. 2013;369:1395-1405.

A20. Velazquez EJ, Lee KL, Deja MA, et al. Coronary-artery bypass surgery in patients with left ventricular dysfunction. *N Engl J Med*. 2011;364:1607-1616.

A21. Slaughter MS, Rogers JG, Milano CA, et al. Advanced heart failure treated with continuous-flow left ventricular assist device. *N Engl J Med*. 2009;361:2241-2251.

A22. Pitt B, Pfeffer MA, Assmann SF, et al. Spironolactone for heart failure with preserved ejection fraction. *N Engl J Med*. 2014;370:1383-1392.

A23. Acker MA, Parides MK, Perrault LP, et al. Mitral-valve repair versus replacement for severe ischemic mitral regurgitation. *N Engl J Med*. 2014;370:23-32.

A24. Felker GM, Lee KL, Bull DA, et al. Diuretic strategies in patients with acute decompensated heart failure. *N Engl J Med*. 2011;364:797-805.

A25. O'Connor CM, Starling RC, Hernandez AF, et al. Effect of nesiritide in patients with acute decompensated heart failure. *N Engl J Med*. 2011;365:32-43.

A26. Chen HH, Anstrom KJ, Givertz MM, et al. Low-dose dopamine or low-dose nesiritide in acute heart failure with renal dysfunction: the ROSE acute heart failure randomized trial. *JAMA*. 2013;310:2533-2543.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. *Circulation*. 2013;128:e240-e327.
2. McMurray JJ, Adamopoulos S, Anker SD, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur Heart J*. 2012;33:1787-1847.
3. Feltner C, Jones CD, Cene CW, et al. Transitional care interventions to prevent readmissions for persons with heart failure: a systematic review and meta-analysis. *Ann Intern Med*. 2014;160:774-784.
4. Damman K, Tang WH, Felker GM, et al. Current evidence on treatment of patients with chronic systolic heart failure and renal insufficiency: practical considerations from published data. *J Am Coll Cardiol*. 2014;63:853-871.
5. Trulock KM, Narayan SM, Piccini JP. Rhythm control in heart failure patients with atrial fibrillation: contemporary challenges including the role of ablation. *J Am Coll Cardiol*. 2014;64:710-721.
6. Prinzen FW, Vernooy K, Auricchio A. Cardiac resynchronization therapy: state-of-the-art of current applications, guidelines, ongoing trials, and areas of controversy. *Circulation*. 2013;128:2407-2418.
7. Jorde UP, Kushwaha SS, Tatooles AJ, et al. Results of the destination therapy post-food and drug administration approval study with a continuous flow left ventricular assist device: a prospective study using the INTERMACS registry (Interagency Registry for Mechanically Assisted Circulatory Support). *J Am Coll Cardiol*. 2014;63:1751-1757.
8. Basuray A, French B, Ky B, et al. Heart failure with recovered ejection fraction: clinical description, biomarkers, and outcomes. *Circulation*. 2014;129:2380-2387.
9. Dworzynski K, Roberts E, Ludman A, et al. Diagnosing and managing acute heart failure in adults: summary of NICE guidance. *BMJ*. 2014;349:g5695.
10. Lainscak M, Blue L, Clark AL, et al. Self-care management of heart failure: practical recommendations from the Patient Care Committee of the Heart Failure Association of the European Society of Cardiology. *Eur J Heart Fail*. 2011;13:115-126.
11. McDonagh TA, Blue L, Clark AL, et al. European Society of Cardiology Heart Failure Association Standards for delivering heart failure care. *Eur J Heart Fail*. 2011;13:235-241.
12. Whellan DJ, Goodlin SJ, Dickinson MG, et al. End-of-life care in patients with heart failure. *J Card Fail*. 2014;20:121-134.

60

DISEASES OF THE MYOCARDIUM AND ENDOCARDIUM

WILLIAM J. MCKENNA AND PERRY ELLIOTT

MYOCARDIAL DISEASE

A substantial minority of cases of heart failure result from familial (genetic) or nonfamilial (acquired) disorders, which can be confined to the heart or be multisystem disorders. The term *cardiomyopathy* refers to myocardial disorders in which the heart muscle is structurally and functionally abnormal in the absence of coronary artery disease (Chapter 73), hypertension (Chapter 67), valvular disease (Chapter 75), or congenital heart disease (Chapter 69) sufficient to cause the observed myocardial abnormality.[1] Cardiomyopathies are classified according to ventricular morphology and pathophysiology into four major types: dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and arrhythmogenic right ventricular cardiomyopathy (ARVC) (Table 60-1 and Fig. 60-1). Diseases that do not fit into these groups (such as endocardial fibroelastosis and left ventricular noncompaction) are termed unclassified cardiomyopathies. Mixed phenotypes can exist; for example, patients with hypertrophic and dilated cardiomyopathies frequently have a restrictive left ventricular physiology or develop ventricular dilation.

Hypertrophic Cardiomyopathy

DEFINITION AND EPIDEMIOLOGY

Hypertrophic cardiomyopathy is defined as unexplained left ventricular hypertrophy in the absence of abnormal loading conditions (valve disease, hypertension, congenital heart defects) sufficient to explain the degree of hypertrophy.[2] The disease occurs in all racial groups, with a prevalence of between 0.2 and 0.5%.

PATHOBIOLOGY

Hypertrophic cardiomyopathy is usually familial with autosomal dominant inheritance. Mutations in sarcomeric contractile protein genes (Table 60-2) account for approximately 50 to 60% of cases. More than 1400 different mutations have been identified, with marked variation in disease penetrance and clinical expression. A similar clinical phenotype is seen in association with other uncommon genetic disorders, including Noonan syndrome (Chapter 69), Friedreich ataxia (Chapter 421), neurofibromatosis (Chapter 417), hereditary spherocytosis (Chapter 161), respiratory chain disorders, glycogen storage diseases (Chapter 207), and lysosomal storage disorders (Chapter 208) (see Table 60-2).

Pathology

In the common form of autosomal dominant hypertrophic cardiomyopathy, myocardial hypertrophy usually affects the interventricular septum more than other regions of the left ventricle. Other patterns, including concentric, mid-ventricular (sometimes associated with a left ventricular apical diverticulum), and apical, also occur. Coexistent right ventricular hypertrophy is present in up to 44% of cases. The papillary muscles are often poorly developed and may be displaced anteriorly, thereby contributing to systolic anterior motion of the anterior mitral valve leaflet in 25% of patients and of the posterior leaflets in 10% of cases in the resting state. Often, the mitral valve is structurally abnormal, with elongation of the anterior leaflet and occasional direct insertion of the papillary muscle into the anterior leaflet. The histologic hallmark of hypertrophic cardiomyopathy is a triad of myocyte hypertrophy, myocyte disarray, and interstitial fibrosis. Myocyte disarray refers to architectural disorganization of the myocardium, with adjacent myocytes aligned obliquely or perpendicular to each other in association with increased inter-

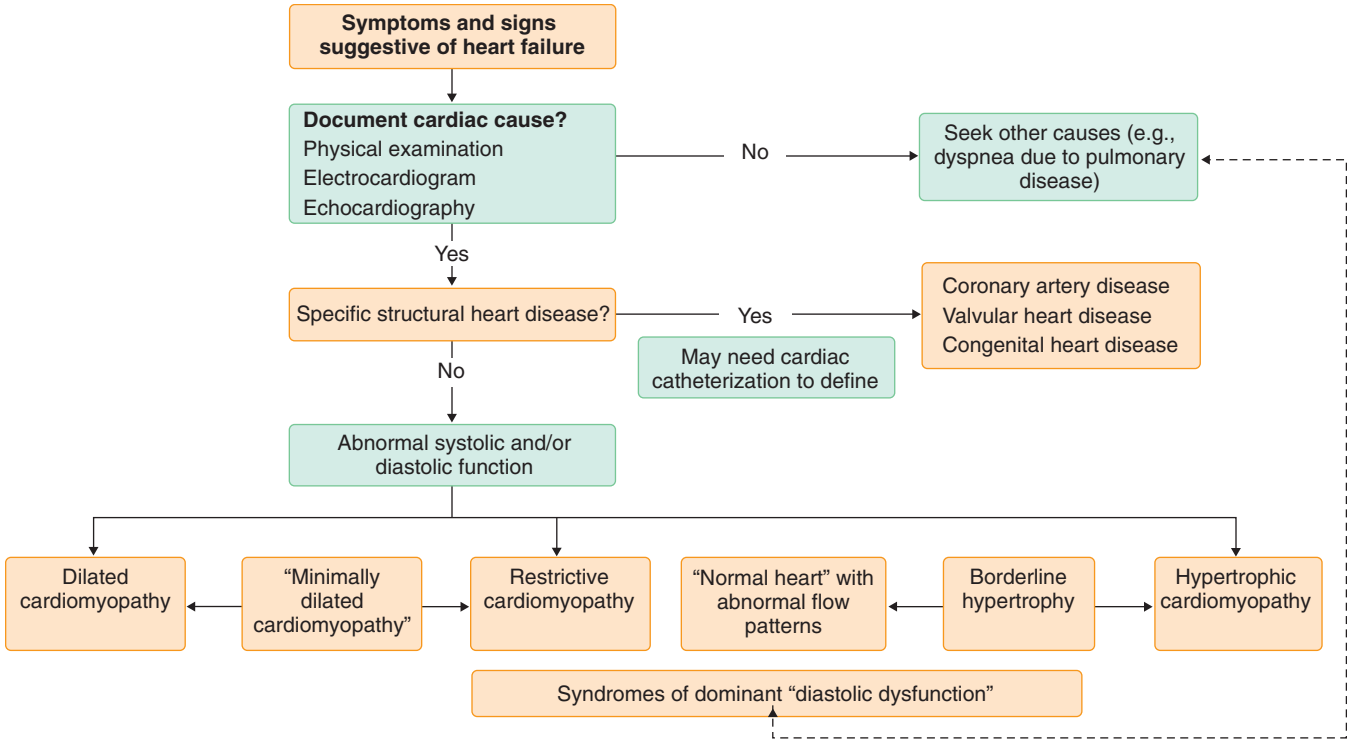

FIGURE 60-1. Initial approach to classification of cardiomyopathy. The evaluation of symptoms or signs consistent with heart failure first includes confirmation that they can be attributed to a cardiac cause. Although this conclusion is often apparent from routine physical examination and electrocardiography, echocardiography serves to confirm cardiac disease and provides clues to the presence of other cardiac diseases, such as focal abnormalities suggesting primary valve disease or congenital heart disease. Having excluded these conditions, cardiomyopathy is generally considered to be dilated, restrictive, or hypertrophic, as shown in Table 60-1. Patients with apparently normal cardiac structure and contraction are occasionally found to demonstrate abnormal intracardiac flow patterns consistent with diastolic dysfunction but should also be evaluated carefully for other causes of their symptoms. Most patients with so-called diastolic dysfunction also demonstrate at least borderline criteria for left ventricular hypertrophy, frequently in the setting of chronic hypertension and diabetes. A moderately decreased ejection fraction without marked dilation or a pattern of restrictive cardiomyopathy is sometimes referred to as minimally dilated cardiomyopathy, which may represent either a distinct entity or a transition between acute and chronic disease.

TABLE 60-1 PROFILES OF MYOCARDIAL DISEASE

	HYPERTROPHIC	DILATED	RESTRICTIVE	ARVC
Causes	Genetic (see Table 60-2)	Myocarditis (see Table 60-4) Metabolic/endocrine Genetic (see Table 60-2)	Infiltrative or storage diseases (see Table 60-8) Endomyocardial (e.g., Löffler, carcinoid) Genetic (see Table 60-2)	Genetic (see Table 60-2)
Ejection fraction	Increased	Reduced	25-50%	Normal until end stage 30% regional LV disease
LV end-diastolic dimension	Usually decreased	Increased	Normal	Normal until end stage Right ventricle dilated
LV wall thickness	Increased	Normal	Normal or mildly increased	Normal
Atrial size	Increased	Increased	Increased; may be massive	Left atrium normal; right dilated in severe disease
Valvular disease	Mitral regurgitation (SAM)	Mitral (functional); tricuspid regurgitation in late stages	Mitral and tricuspid regurgitation, rarely severe	Tricuspid regurgitation in severe disease
Common symptoms	Dyspnea; chest pain, syncope Late: orthopnea, PND	Dyspnea, fatigue Late: orthopnea, PND	Dyspnea Late: orthopnea, PND, right-sided heart failure	Palpitations, syncope Late: right-sided heart failure
Arrhythmia	Atrial fibrillation, ventricular tachycardia; conduction block in *PRKAG2*, mitochondrial; Fabry disease	Ventricular tachyarrhythmias; heart block in Chagas disease, giant cell myocarditis, laminopathies	Atrial fibrillation; conduction block in sarcoid, amyloidosis, desminopathy	Ventricular ectopy and tachycardia

ARVC = arrhythmogenic right ventricular cardiomyopathy; LV = left ventricular; PRKAG2 = protein kinase, AMP-activated, gamma 2 non-catalytic subunit mutation; SAM = systolic anterior motion of mitral valve; PND = paroxysmal nocturnal dyspnea.

stitial collagen. The myofibrillar architecture within the myocyte is also disorganized. Although myocyte disarray occurs in aortic stenosis, long-standing hypertension, and some forms of congenital heart disease, the presence of extensive disarray (more than 10% of ventricular septal myocytes) is thought to be a highly specific marker for hypertrophic cardiomyopathy. Small intramural coronary arteries are often dysplastic and narrowed because of wall thickening by smooth muscle cell hyperplasia.

PATHOPHYSIOLOGY

Abnormal ventricular geometry, wall thickening, myocyte hypertrophy, myocyte and myofibrillar disarray, and myocardial fibrosis all contribute to impairment of left ventricular diastolic function. The net result is elevation of left ventricular end-diastolic pressures, symptoms of heart failure, and reduced exercise tolerance. Global measures of left ventricular systolic

TABLE 60-2 GENETIC CAUSES OF CARDIOMYOPATHY

GENE	SYMBOL	INHERITANCE	PHENOTYPES	ESTIMATED FREQUENCY
SARCOMERIC PROTEINS				
Cardiac β-myosin heavy chain	MYH7	AD	Variable: moderate to severe prognosis; HCM; LVNC; DCM; Laing distal myopathy	HCM 30-40%; DCM 4-6%
Cardiac myosin-binding protein C	MYBPC3	AD	Late onset of HCM described; cases of children with a severe hypertrophy also reported; DCM	HCM 30-40%; DCM ~1%
Cardiac troponin T	TNNT2	AD	HCM: possible high incidence of sudden death; DCM	HCM 10-15%; DCM 3-5%
Cardiac troponin I	TNNI3	HCM: AD DCM: AD, AR	RCM; HCM; DCM	HCM 2-5%; DCM <1%
α-Tropomyosin	TPM1	AD	HCM; DCM	HCM ~1-2%; DCM <1%
Regulatory myosin light chain	MYL2	AD	HCM; DCM	HCM ~1%; DCM rare
Cardiac actin	ACTC	AD	DCM; LVNC; HCM	DCM ~1%; HCM ~1%
Essential myosin light chain	MYL3	AD	HCM; DCM	Rare
Cardiac α-myosin heavy chain	MYH6	AD	DCM; HCM	HCM <1%; DCM rare
Cardiac troponin C	TNNC	AD	HCM; DCM	Rare
SARCOMERE AND Z-DISC–RELATED PROTEINS				
Titin	TTN	AD	DCM; HCM; ARVC	HCM rare; DCM 15-25%; ARVC rare
BCL2-associated athanogene 3	BAG3	AD	DCM	2-4%
Cypher/ZASP	LDB3	AD	LVNC; DCM	DCM <1%
Titin-cap or telethonin	TCAP	AD	HCM; DCM	HCM <1%; DCM rare
α-Actinin-2	ACTN2	AD	HCM; DCM	Rare
Ankyrin repeat domain-containing protein 1	ANKRD1	AD	DCM	Rare
Cysteine and glycine-rich protein 3 (cardiac LIM protein)	CSRP3	AD	DCM; HCM	Rare
Four-and-a-half LIM protein 1	FHL1	AD	DCM	Rare
Four-and-a-half LIM protein 2	FHL2	AD	DCM	Rare
Myozenin 2	MYOZ2	AD	HCM	Rare
Myopalladin	MYPN	AD	DCM	Rare
Nexilin	NEXN	AD	HCM; DCM	Rare
Nebulette	NEBL	AD	DCM	Rare
PDZ and LIM domain protein 3	PDLIM3	AD	DCM; HCM	Rare
Metavinculin	VCL	AD	DCM; HCM	Rare
CYTOSKELETAL PROTEINS				
Desmin	DES	AD, AR	DCM; desminopathies; DCM with clinical features usually associated with ARVC	DCM <1%
Dystrophin	DMD	XL	DCM in Duchenne muscular dystrophy (DMD); Becker muscular dystrophy (BMD)	DCM <1%
Caveolin-3	CAV3	AD	HCM	Rare
α-B crystallin	CRYAB	AD	DCM; myofibrillar myopathies	Rare
α-, β-, γ-, and δ-sarcoglycans	SGCA, SGCB, SGCG, SGCD	SGCD: AD	DCM	Rare
NUCLEAR PROTEINS				
Lamin A/C	LMNA	LVNC, DCM: AD EMD2: AD EMD3: AR LGMD1B: AD ARVC: AD	LVNC; DCM; DCM in Emery-Dreifuss muscular dystrophy types 2 and 3 (EMD2 and EMD3); DCM in limb girdle muscular dystrophy; DCM with clinical features usually associated with ARVC	DCM 4-8%
Dystrobrevin	α-DTNA	AD	DCM; LVNC	Rare
Emerin	EMD	XL	DCM, Emery-Dreifuss muscular dystrophy	Rare
PR domain-containing protein 16	PRDM16	AD	LVNC; DCM	Rare
Syntrophin	SNTA1	AD	DCM	Rare
Spectrin repeat containing, nuclear envelope 1	SYNE1	AD	DCM	Rare
Spectrin repeat containing, nuclear envelope 2	SYNE2	AD	DCM	Rare
Transmembrane protein 43	TMEM43	AD	ARVC	Rare
Thymopoietin	TMPO	AD	DCM	Rare

TABLE 60-2 GENETIC CAUSES OF CARDIOMYOPATHY—cont'd

GENE	SYMBOL	INHERITANCE	PHENOTYPES	ESTIMATED FREQUENCY
ION CHANNEL AND ION CHANNEL RELATED				
Cardiac sodium channel	SCN5A	AD	DCM; LVNC	DCM 1-2%
Regulatory SUR2A subunit of the cardiac K(ATP) channel	ABCC9	AD	DCM	Rare
DESMOSOMAL PROTEINS				
Plakophilin 2	PKP2	AD, AR	ARVC	AD 30-40%; AR rare
Desmoglein 2	DSG2	AD	ARVC	12-40%
Desmoplakin	DSP	ARVC: AD Carvajal syndrome: AR	ARVC; DCM in Carvajal syndrome	ARVC 6-16%
Desmocollin 2	DSC2	AD	ARVC	Rare
Plakoglobin	JUP	ARVC: AD Naxos disease: AR	ARVC; Naxos disease	Rare
CALCIUM-HANDLING PROTEINS				
Phospholamban	PLN	AD	DCM; HCM; ARVC	DCM <1%; HCM rare; ARVC rare
Calsequestrin 2 (cardiac muscle)	CASQ2	AD	LVNC; CPVT	Rare
Junctophilin 2	JPH2	AD	HCM	Rare
Cardiac ryanodine receptor	RYR2	AD	ARVC; CPVT	Rare
METABOLIC PROTEINS				
Amylo-1,6-glucosidase	AGL	AR	Cardiomyopathy in Forbes disease	?
Acid α-1,4-glucosidase	GAA	AR	Cardiomyopathy in Pompe disease	?
α-Galactosidase A	GLA	XL	HCM in Anderson-Fabry disease	?
Lysosomal-associated membrane protein 2	LAMP2	XL	HCM in Danon disease	?
Protein kinase, AMP-activated, γ_2 noncatalytic subunit	PRKAG2	AD	HCM in Wolff-Parkinson-White syndrome	?
Frataxin	FRDA	AR	HCM in Friedreich ataxia	?
OTHERS				
RNA-binding protein 20	RBM20	AD	DCM	3-5%
M_2 muscarinic receptor	CHRM2	AD	DCM	Rare
Cardiotrophin 1	CTF1	AD	DCM	Rare
αT-catenin	CTNNA3	AD	ARVC	Rare
Dolichol kinase	DOLK	AD	DCM	Rare
Eyes absent 4	EYA4	AD	DCM	Rare
Fukutin-related protein	FKRP	AD, AR	DCM as part of limb-girdle muscular dystrophy 2I	Rare
Fukutin	FKTN	AD, AR	DCM; Limb-girdle muscular dystrophy	Rare
GATA zinc finger domain-containing protein 1	GATAD1	AR	DCM	Rare
Hereditary hemochromatosis	HFE	AR	DCM and RCM in hereditary hemochromatosis	Rare
Laminin α2	LAMA2	AD	DCM	Rare
Laminin α4	LAMA4	AD	DCM	Rare
Integrin-linked kinase	ILK	AD	DCM	Rare
Genes encoding mitochondrial components	MTTG, MTTY, MTND5, others	AD, maternal	HCM in MELAS, MERRF, LHON syndromes	?
Muscle-related coiled-coil protein	MURC	AD	DCM	Rare
Myosin light chain kinase 2	MYLK2	AD	HCM	Rare
Myomesin 1	MYOM1	AD	HCM	Rare
Myomesin 2	MYOM2	AD	HCM	Rare
Myotilin	MYOT	AD	DCM	Rare
Presenilin 1	PSEN1	AD	DCM	Rare
Presenilin 2	PSEN2	AD	DCM	Rare
RAS-MAPK pathway genes	PTPN11, RAF1, SOS1, KRAS, HRAS, BRAF, MEK1-2, others	AD	HCM in Noonan syndrome and LEOPARD syndrome	?
Tafazzin	TAZ	XL	DCM; LVNC; Barth syndrome	Rare
Transcription factor TBX20	TBX20	AD	DCM with developmental anomalies	Rare
Transforming growth factor β3	TGFB3	AD	ARVC	Rare
Muscle RING Finger 1 (MuRF1)	TRIM63	AD	HCM	Rare
Hereditary amyloidosis	TTR	AD	HCM and RCM in hereditary amyloidosis	Rare

AD = autosomal dominant; AR = autosomal recessive; ARVC = arrhythmogenic right ventricular cardiomyopathy; CPVT = catecholaminergic polymorphic ventricular tachycardia; DCM = dilated cardiomyopathy; HCM = hypertrophic cardiomyopathy; LVNC = left ventricular noncompaction; RCM = restrictive cardiomyopathy; XL = X-linked.

function are often normal, but regional myocardial dysfunction and progressive systolic impairment are relatively common.

Approximately 25% of patients have left ventricular outflow tract obstruction at rest caused by contact between the anterior leaflet of the mitral valve and the interventricular septum during ventricular systole. Many patients without outflow obstruction at rest develop it during physiologic and pharmacologic interventions that reduce left ventricular end-diastolic volume or increase left ventricular contractility.

CLINICAL MANIFESTATIONS

Most patients are asymptomatic or have only mild or intermittent symptoms. Symptomatic progression is usually slow, age related, and associated with a gradual deterioration in left ventricular function during decades. Less than 5% of patients may have rapid, symptomatic deterioration. Symptoms can develop at any age, even many years after the appearance of electrocardiographic (ECG) or echocardiographic manifestations of left ventricular hypertrophy. On occasion, sudden death may be the initial presentation. However, most individuals with hypertrophic cardiomyopathy have few if any symptoms, and the diagnosis is often made as a result of family screening or the incidental detection of a heart murmur or ECG abnormality.

Approximately 20 to 30% of adults develop chest pain (Chapters 51 and 71), which may occur on exertion, at rest, or nocturnally. Postprandial angina associated with mild exertion is typical. Mild to moderate dyspnea on exertion is relatively common, and some patients develop paroxysmal nocturnal dyspnea that may be caused by transient myocardial ischemia or arrhythmia. Approximately 20% of patients experience syncope (Chapters 51 and 62), and a similar proportion complain of presyncope. Palpitations (Chapter 62) are frequent and are usually attributable to supraventricular or ventricular ectopy or to forceful cardiac contraction. Sustained palpitations are usually caused by supraventricular tachyarrhythmias, but initial presentation with a symptomatic arrhythmia is uncommon. Patients with distal or apical hypertrophy have fewer symptoms and arrhythmias, better exercise capacity, and good prognosis. On occasion, however, patients with distal or apical hypertrophy may have severe refractory chest pain or may present with troublesome supraventricular arrhythmias.

DIAGNOSIS

A three- to four-generation family history, which should be obtained in all patients with a new diagnosis of cardiomyopathy, helps determine the probability of familial disease and its mode of inheritance. The initial diagnostic evaluation includes a family history focusing on premature cardiac disease or death, a comprehensive medical history focusing on cardiovascular symptoms, a careful physical examination, a 12-lead electrocardiogram, and a two-dimensional echocardiogram.

The general evaluation may provide diagnostic clues in patients whose hypertrophic cardiomyopathy is associated with syndromes or metabolic disorders. For example, Noonan syndrome is characterized by short stature, developmental delay, cutaneous abnormalities (cafe au lait spots), hypertelorism, ptosis, low-set posteriorly rotated ears, and webbed neck. These features are shared with the less common LEOPARD syndrome. Angiokeratomas, anhidrosis, Raynaud-like symptoms with neuropathy, cornea verticillata, retinal vascular dilation, tinnitus, diarrhea, and proteinuria are typical features of Fabry disease (Chapter 208).

Clinical examination of the cardiovascular system is often normal. In the presence of left ventricular outflow tract obstruction, the arterial pulse has a rapid upstroke and downstroke (sometimes with a bisferiens character), the apex beat is sustained or double (reflecting a palpable atrial impulse followed by left ventricular contraction), and auscultation will demonstrate a systolic ejection murmur that is heard loudest at the left sternal edge and that radiates to the right upper sternal edge and apex (Chapter 51). Most patients with left ventricular outflow tract obstruction also have the murmur of mitral regurgitation, which results from failure of the mitral valve leaflets to coapt due to the systolic anterior motion of the mitral valve. Physiologic and pharmacologic maneuvers that decrease afterload or venous return (e.g., standing, Valsalva maneuver, inhalation of amyl nitrite) or increase contractility (e.g., a post-extrasystole beat) will increase the intensity of the murmur, whereas interventions that increase afterload and venous return (e.g., squatting or handgrip) will reduce it (see Table 51-8). In contrast, physical signs in most patients who do not have left ventricular outflow tract obstruction are subtle and are limited to features that reflect the hyperdynamic contraction (rapid upstroke pulse) and poorly compliant right (prominent *a* wave in jugular venous pressure) and left (S_4 gallop, double-apex beat) ventricles (Chapter 51).

Diagnostic Testing

More than 95% of patients have abnormal ECG findings, but no changes are disease specific. The most common abnormalities are increased QRS voltage consistent with left ventricular hypertrophy, left axis deviation (15 to 20%), abnormal Q waves (25 to 30%, most commonly in inferolateral leads), and ST segment or T wave changes (>50%). An isolated increase in the QRS voltage without ST segment changes or T wave inversion is rare in hypertrophic cardiomyopathy. The presence of predominantly distal or apical thickening is associated with giant negative T wave inversion, which is maximal in leads V_3 and V_4.

Two-dimensional echocardiography (Chapter 55) is the mainstay of diagnostic imaging, but magnetic resonance imaging (Chapter 56) and computed tomography (Chapter 56) provide alternatives if the echocardiogram is of poor quality. In most patients, the hypertrophy is asymmetrical and involves the anterior and posterior intraventricular septum (Fig. 60-2). The hypertrophy, however, may be more generalized and involve the free wall of the left ventricle, or it may be localized and confined to areas other than the septum, such as the lateral or posterior wall of the left ventricle. The echocardiogram

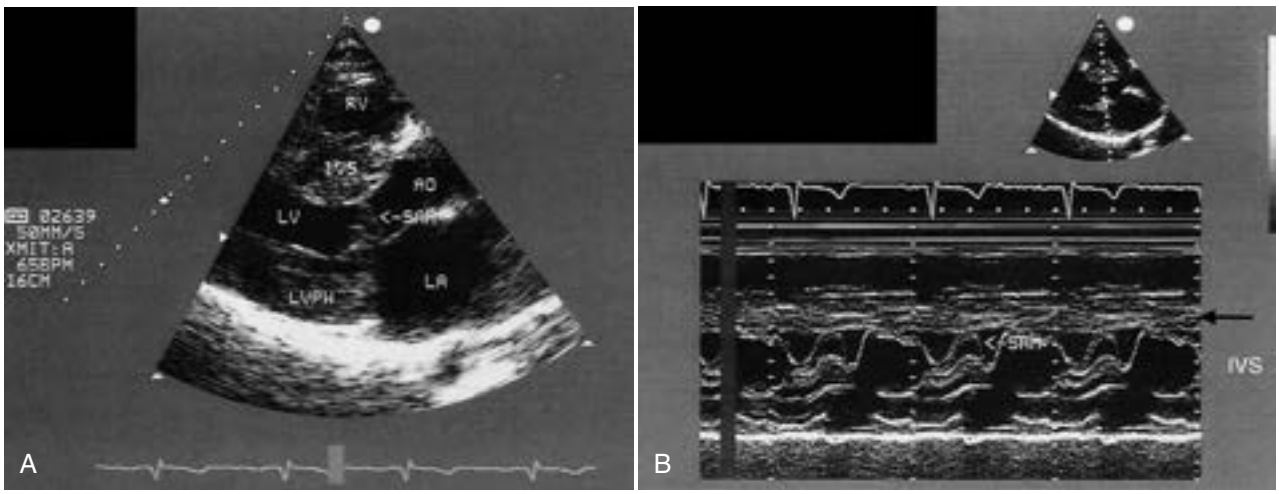

FIGURE 60-2. Hypertrophic obstructive cardiomyopathy. A, The two-dimensional long axis parasternal view shows the chambers of the heart. The left ventricle posterior wall (LVPW) is thickened, and the most striking abnormality is the hypertrophy of the interventricular septum (IVS). Another characteristic feature is a Venturi effect: as blood leaves the left ventricle (LV), it sucks the anterior leaflet of the mitral valve forward, a phenomenon called systolic anterior motion (SAM). **B,** This phenomenon is more clearly shown in the parasternal long axis M-mode echocardiogram. The massive thickening of the septum is also obvious in the M-mode image (IVS). AO = aorta; LA = left atrium; RV = right ventricle. (From Forbes CD, Jackson WF. Color Atlas and Text of Clinical Medicine. 3rd ed. London: Mosby; 2003.)

can measure left ventricular outflow tract obstruction, both at rest and after provocative maneuvers. Patients with an outflow tract gradient of 30 mm Hg or more typically have systolic anterior motion of the mitral valve, with contact of either the anterior or (less commonly) the posterior mitral leaflet with the intraventricular septum during systole, in association with a posteriorly directed jet of mitral regurgitation, the severity of which is proportionate to the severity of the obstruction. Most patients with hypertrophic cardiomyopathy have left atrial enlargement as well as echocardiographic evidence of diastolic dysfunction. Magnetic resonance imaging, although not needed for the diagnosis, readily demonstrates the characteristic abnormalities (E-Figs. 60-1 to 60-4).

When it is available, cardiopulmonary exercise testing with metabolic gas exchange measurements provides an accurate and reproducible assessment of exercise capacity, which can be followed serially. Cardiac catheterization is rarely required for diagnosis or management, but it may be indicated when measurement of intracardiac pressures is required to guide therapeutic decisions (e.g., in patients with severe mitral regurgitation) and for the exclusion of coexistent coronary artery disease in patients with chest pain.

Diagnostic Criteria

A wall thickness of more than 2 standard deviations above the mean, corrected for age, gender, and height, is generally accepted as diagnostic. In adults, this value is typically 1.5 cm or more in men and 1.3 cm or more in women. In the presence of other causes of left ventricular hypertrophy, such as long-standing systemic hypertension or aortic stenosis, the diagnosis of hypertrophic cardiomyopathy may be problematic. However, secondary hypertrophy from other causes rarely exceeds 1.8 cm. Hypertrophy in the highly trained athlete is usually less than 1.6 cm in men and 1.4 cm in women and typically occurs in association with an increased left ventricular end-diastolic dimension and stroke volume. An ECG tracing showing Q waves or inferolateral repolarization changes in an athlete favors the diagnosis of hypertrophic cardiomyopathy.

Given the 50% probability of disease in first-degree relatives of a patient with hypertrophic cardiomyopathy, modified diagnostic criteria (Table 60-3) consider the high probability that their otherwise unexplained ECG and echocardiographic findings reflect incomplete disease expression, with the corresponding risks for complications and for passing the gene to their children.

TABLE 60-3 DIAGNOSTIC CRITERIA FOR HYPERTROPHIC CARDIOMYOPATHY IN FIRST-DEGREE RELATIVES OF AFFECTED PATIENTS*

MAJOR CRITERIA	MINOR CRITERIA
ECHOCARDIOGRAPHY	
Left ventricular wall thickness ≥13 mm in the anterior septum or posterior wall or ≥15 mm in the posterior septum or free wall	Left ventricular wall thickness of 12 mm in the anterior septum or posterior wall or of 14 mm in the posterior septum or free wall
Severe SAM of the mitral valve (septal-leaflet contact)	Moderate SAM of the mitral valve (no mitral leaflet–septal contact)
	Redundant mitral valve leaflets
ELECTROCARDIOGRAPHY	
Left ventricular hypertrophy with repolarization changes (Romhilt and Estes)	Complete bundle branch block or (minor) interventricular conduction defects (in left ventricular leads)
T wave inversion in leads I and aVL (≥3 mm with QRS-T wave axis difference ≥30 degrees), V_3-V_6 (≥3 mm), or II and III and aVF (≥5 mm)	Minor repolarization changes in left ventricular leads Deep S wave in lead V_2 (>25 mm)
Abnormal Q waves (>40 msec or >25% R wave) in at least two leads from II, III, aVF (in absence of left anterior hemiblock), and V_1-V_4; or I, aVL, V_5-V_6	Unexplained chest pain, dyspnea, or syncope

*The diagnosis of hypertrophic cardiomyopathy in first-degree relatives of patients with the disease is based on the presence of one major criterion, two minor echocardiographic criteria, or one minor echocardiographic criterion and two minor electrocardiographic criteria.
aVF = augmented voltage unipolar left foot lead; aVL = augmented voltage unipolar left arm lead; SAM = systolic anterior motion.
Modified from McKenna WJ, Spirito P, Desnos M, et al. Experience in clinical genetics in hypertrophic cardiomyopathy. *Heart*. 1997;77:130-132.

TREATMENT

Clinical management is based mainly on symptoms (Fig. 60-3).[3,4] Exceptions include specific therapies for lysosomal storage diseases, such as Pompe disease (Chapter 207) and Fabry disease (E-Fig. 60-5; Chapter 208), and for Friedreich ataxia (Chapter 421). The treatment of the remaining patients with hypertrophic cardiomyopathy focuses on the counseling of family members, the management of symptoms, and the prevention of disease-related complications.

Family Evaluation

All patients with hypertrophic cardiomyopathy should be counseled on the implications of the diagnosis for their families. Careful pedigree analysis can reassure relatives who are not at risk for inheriting the disease. For those who are at risk, current guidelines recommend screening with a 12-lead electrocardiogram and echocardiogram at intervals of 12 to 18 months, usually starting at the age of 12 years (unless there is a "malignant" family history of premature sudden death, the child is symptomatic or a competitive athlete, or there is a clinical suspicion of left ventricular hypertrophy) until full growth and maturation are achieved (usually by the age of 18 to 21 years). Thereafter, if there are no signs of disease expression, screening approximately every 5 years is advised because the onset of left ventricular hypertrophy may be delayed until well into adulthood in some families. Modified diagnostic criteria (see Table 60-3) consider the high probability that otherwise unexplained ECG and echocardiographic findings in first-degree relatives reflect incomplete disease expression.

When it is available, genetic testing can identify a disease-causing mutation in an index case and thereby provide presymptomatic diagnosis of family members. Whenever genetic testing is considered, individuals should be informed about the purpose of the test, the most probable mode of inheritance, and the potential hazards and limitations of genetic testing.

Symptom Management

Medical Therapy

Therapeutic options in patients *without* left ventricular outflow gradients are limited predominantly to pharmacologic therapy. β-Blockade may improve chest pain and dyspnea. The dose (starting at a dose equivalent to propranolol 120 mg/day) should be titrated to achieve a target heart rate of 50 to 70 beats per minute at rest and 130 to 140 beats per minute at peak exercise. Calcium antagonists such as verapamil (starting at a dose of 120 mg/day) and diltiazem (starting at a dose of 180 mg/day) are useful alternatives, particularly in patients with refractory chest pain, but high doses (e.g., verapamil >480 mg/day, diltiazem >360 mg/day) may be required. In patients with paroxysmal nocturnal dyspnea and no evidence of ventricular outflow obstruction, a transient mechanism such as myocardial ischemia or arrhythmia may be implicated, although investigations usually fail to identify the precise cause. Such patients as well as those with chronically raised pulmonary pressures may require diuretics (e.g., furosemide, 20 to 40 mg orally as needed, followed by 20 mg/day if required). The dose and duration of diuretic therapy should be minimized because injudicious use of these drugs can be dangerous, particularly in patients with severe diastolic impairment or labile obstruction.

In patients with symptoms caused by left ventricular outflow tract obstruction, the main aim of treatment is to reduce the outflow tract gradient. Options include negative inotropic drugs, surgery,[5] atrioventricular sequential pacing, and percutaneous alcohol ablation of the interventricular septum. Approximately 60 to 70% of patients improve with β-blockers, although high doses (equivalent to propranolol at 480 mg/day) are frequently required, and side effects are often limiting. When β-blockade alone is ineffective, disopyramide, titrated to the maximal tolerated dose (usually between 400 and 600 mg/day), may be effective in up to two thirds of patients, but side effects related to the anticholinergic effects (e.g., dry eyes and mouth, urinary retention) limit its use. Disopyramide should be given concomitantly with a small to medium dose of a β-blocker (e.g., propranolol, 120 to 240 mg/day), which will slow the heart rate and also blunt rapid atrioventricular nodal conduction should supraventricular arrhythmias develop. In patients who have left ventricular outflow tract obstruction and are taking a β-blocker and disopyramide, other antiarrhythmic drugs that alter repolarization (e.g., sotalol or amiodarone) must be avoided because of the potential proarrhythmic effect. In patients with outflow tract gradients, verapamil can be effective, but caution is required in patients with severe obstruction or elevated pulmonary pressures.

Interventional Therapy

Surgery should be considered for significant outflow obstruction (gradient >50 mm Hg) in patients who have symptoms refractory to medical therapy. The most commonly performed surgical procedure, ventricular septal myectomy, either abolishes or substantially reduces the gradient in 95% of cases, reduces mitral regurgitation, and improves exercise capacity and symptoms. Surgery should be performed in an experienced center, where mortality rates should be less than 1% for isolated myomectomy. The main complications (atrioventricular block, ventricular septal defects) are uncommon (2 to 5%).

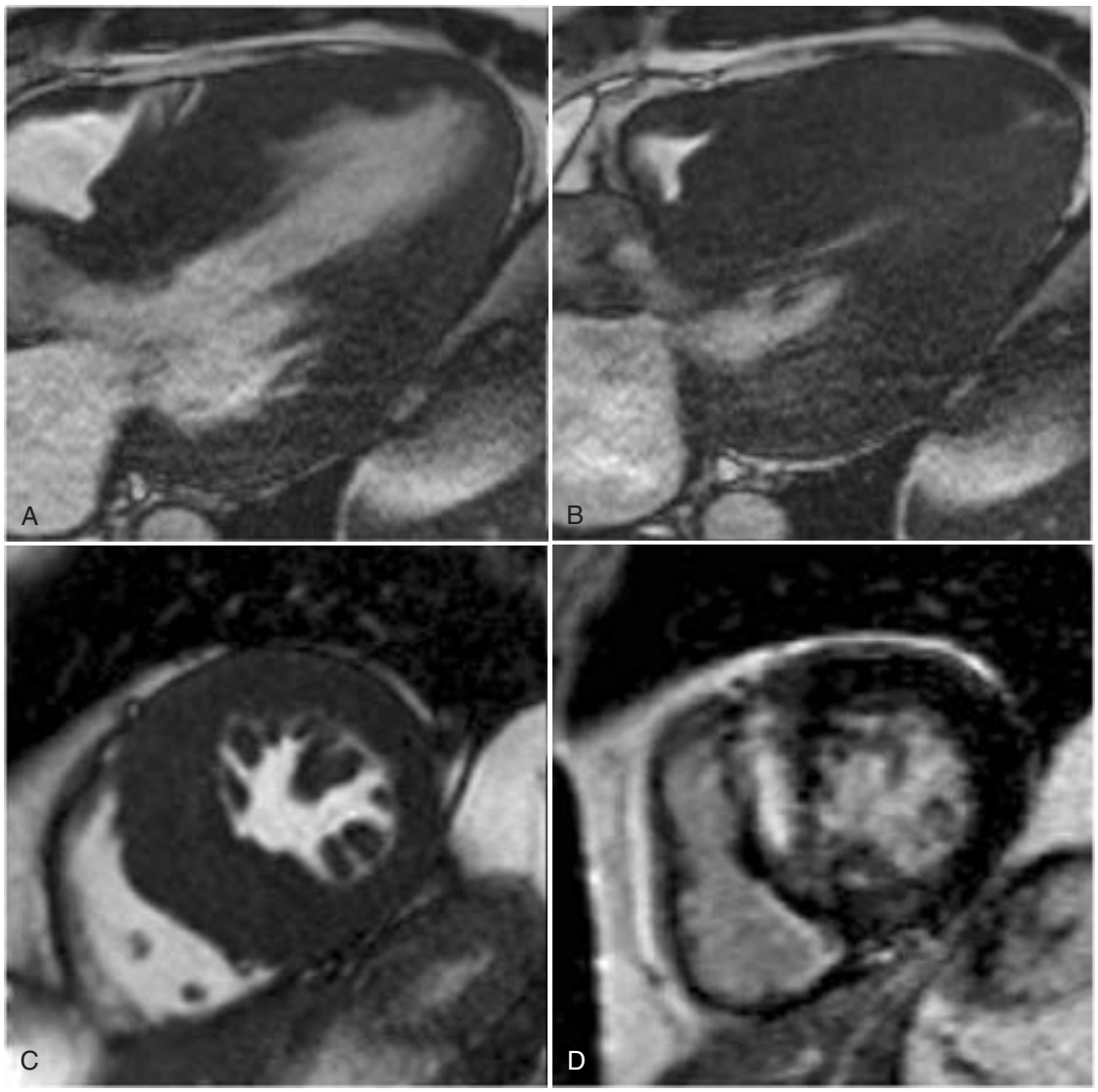

E-FIGURE 60-1. Cardiac magnetic resonance image in a 40-year-old man with hypertrophic cardiomyopathy. Three-chamber cine in diastole (**A**) and systole (**B**). There is mid-cavity obliteration and left ventricular outflow tract obstruction with complete systolic anterior motion of the mitral valve. The basal short axis cine (**C**) in diastole measures 30 mm, and there is mid-myocardial scar, visualized as late gadolinium enhancement (**D**).

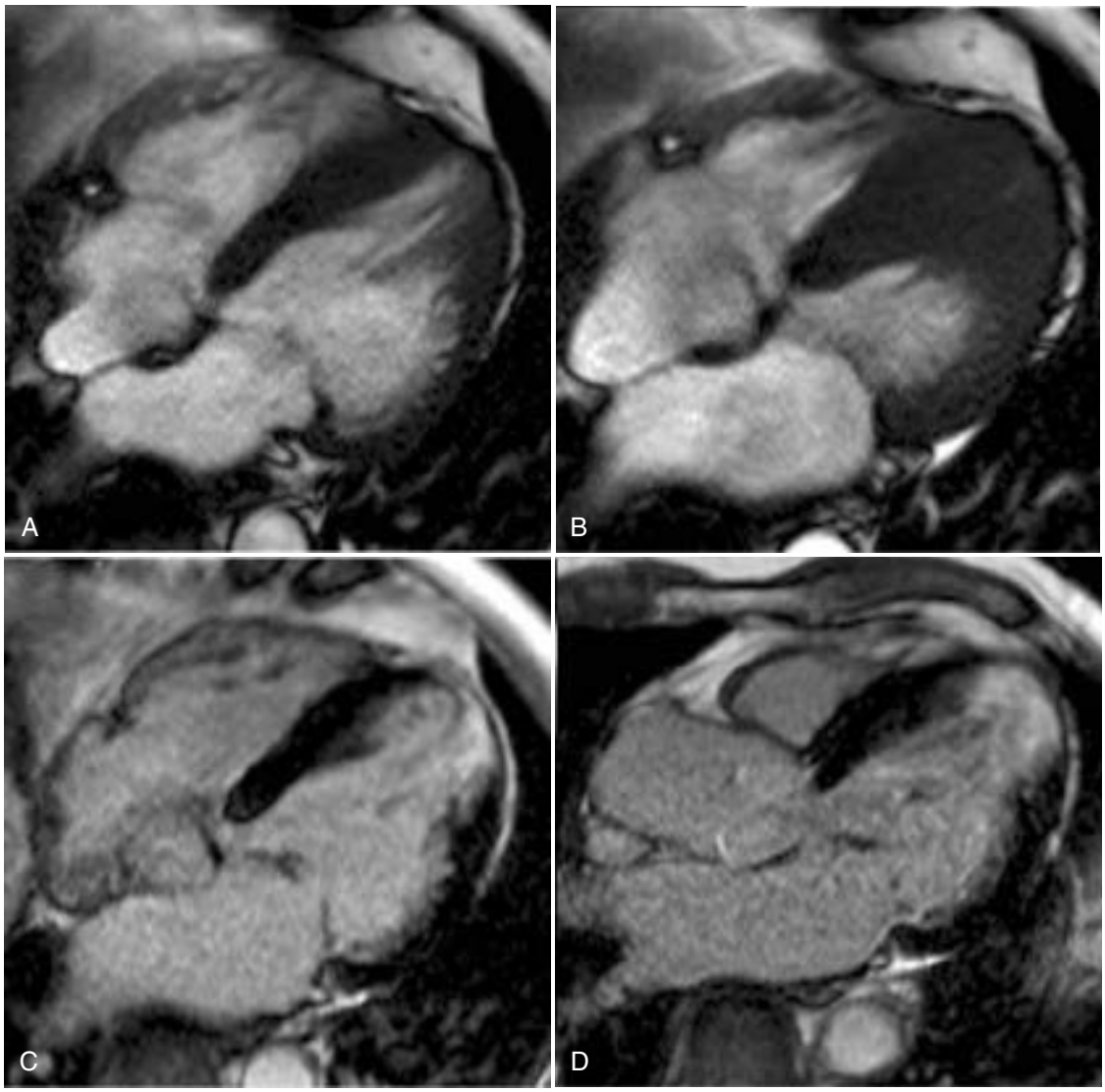

E-FIGURE 60-2. A 70-year-old man with apical hypertrophic cardiomyopathy. Four-chamber cine in diastole (**A**) and systole (**B**). There is apical hypertrophy (16 mm) with 3 cm of apical cavity obliteration in systole. Note the papillary muscle insertion distally. Apical scar is present (**C, D**) without an overt apical aneurysm.

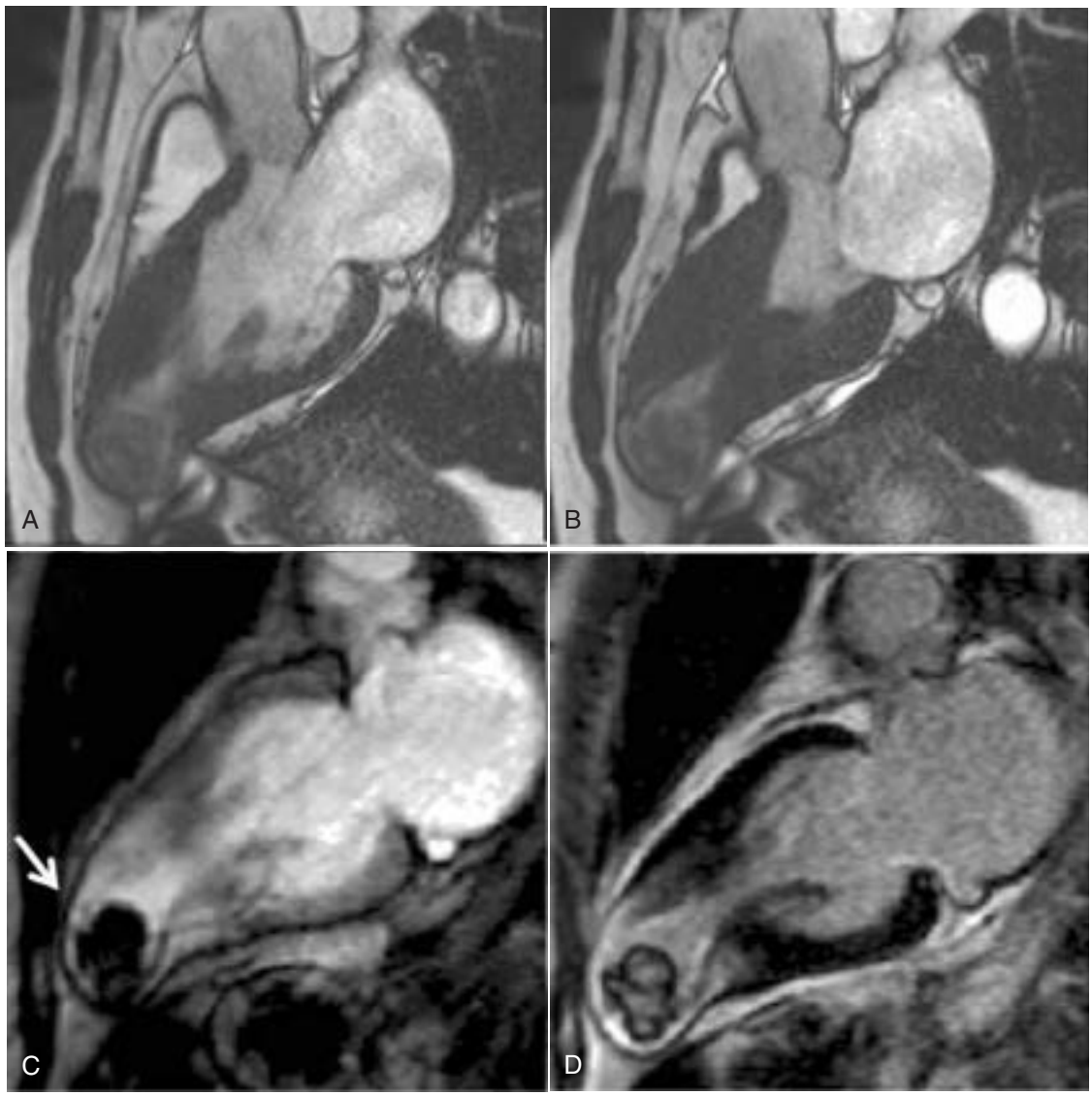

E-FIGURE 60-3. Cardiac magnetic resonance images in a 70-year-old man with an electrocardiographic abnormality, previous ventricular tachycardia (nonsustained), and recent ischemic stroke. Cardiac magnetic resonance images (**A**, diastole; **B**, systole) show mid-cavity hypertrophy and obstruction with an apical aneurysm containing a filling defect, shown to be avascular on early gadolinium-enhanced imaging (**C**, *arrow*, thrombus), with extensive apical scar (**D**). This is apical hypertrophic cardiomyopathy with apical aneurysm formation.

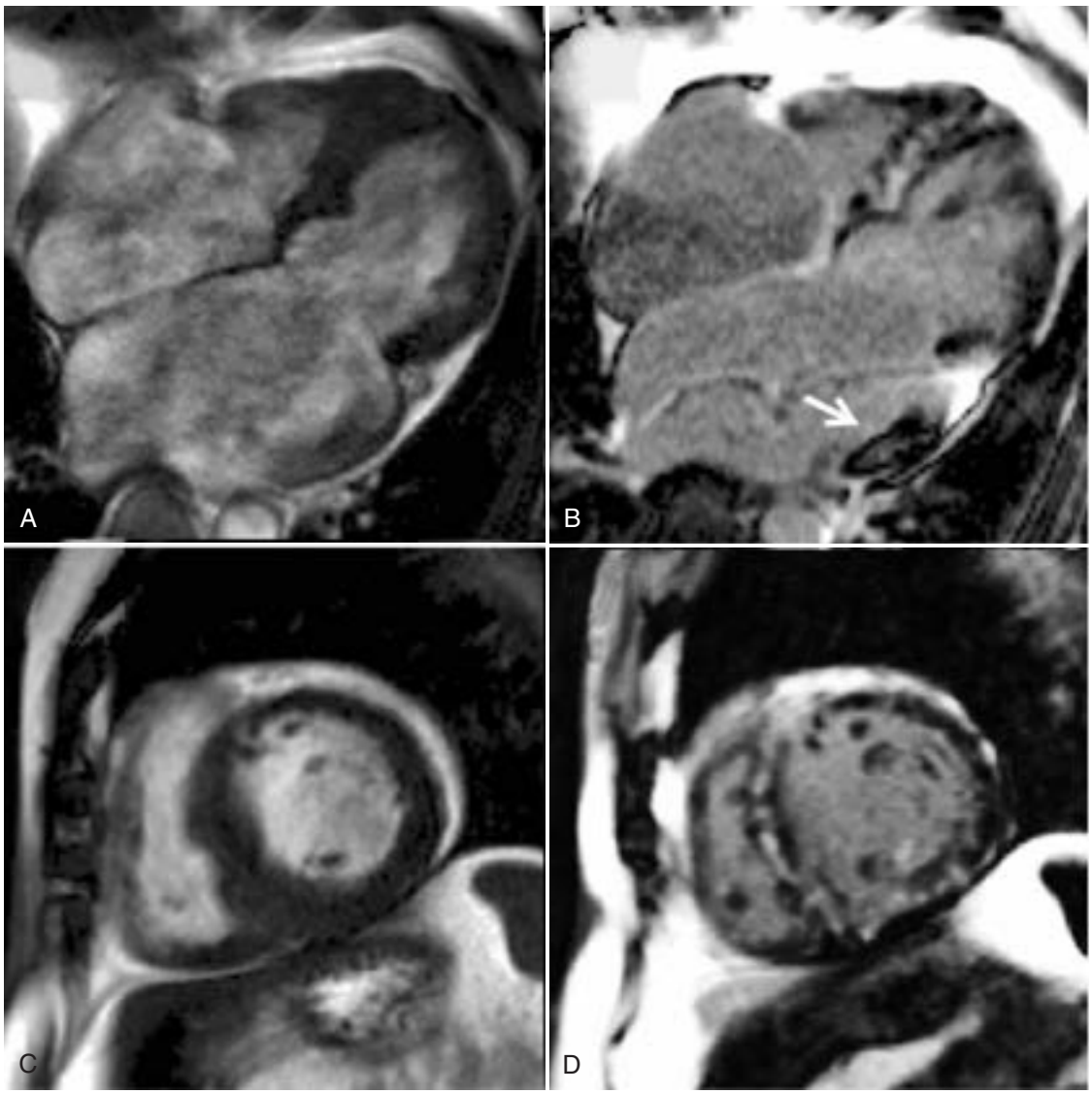

E-FIGURE 60-4. A 50-year-old man with hypertrophic cardiomyopathy in a progressive (dilated) disease phase. The left ventricular ejection fraction is 35% with probably restrictive physiology, although the patient is in atrial fibrillation. There is left ventricular dilation, particularly for hypertrophic cardiomyopathy, and severe atrial dilation (**A**, diastolic frame). The maximal wall thickness is known to have reduced during recent follow-up (**C**, was 23 mm, now 18 mm). There is extensive, non-ischemic myocardial scar estimated at 40% of the myocardium (**B, D**) and left atrial thrombus (*arrow*, **B**).

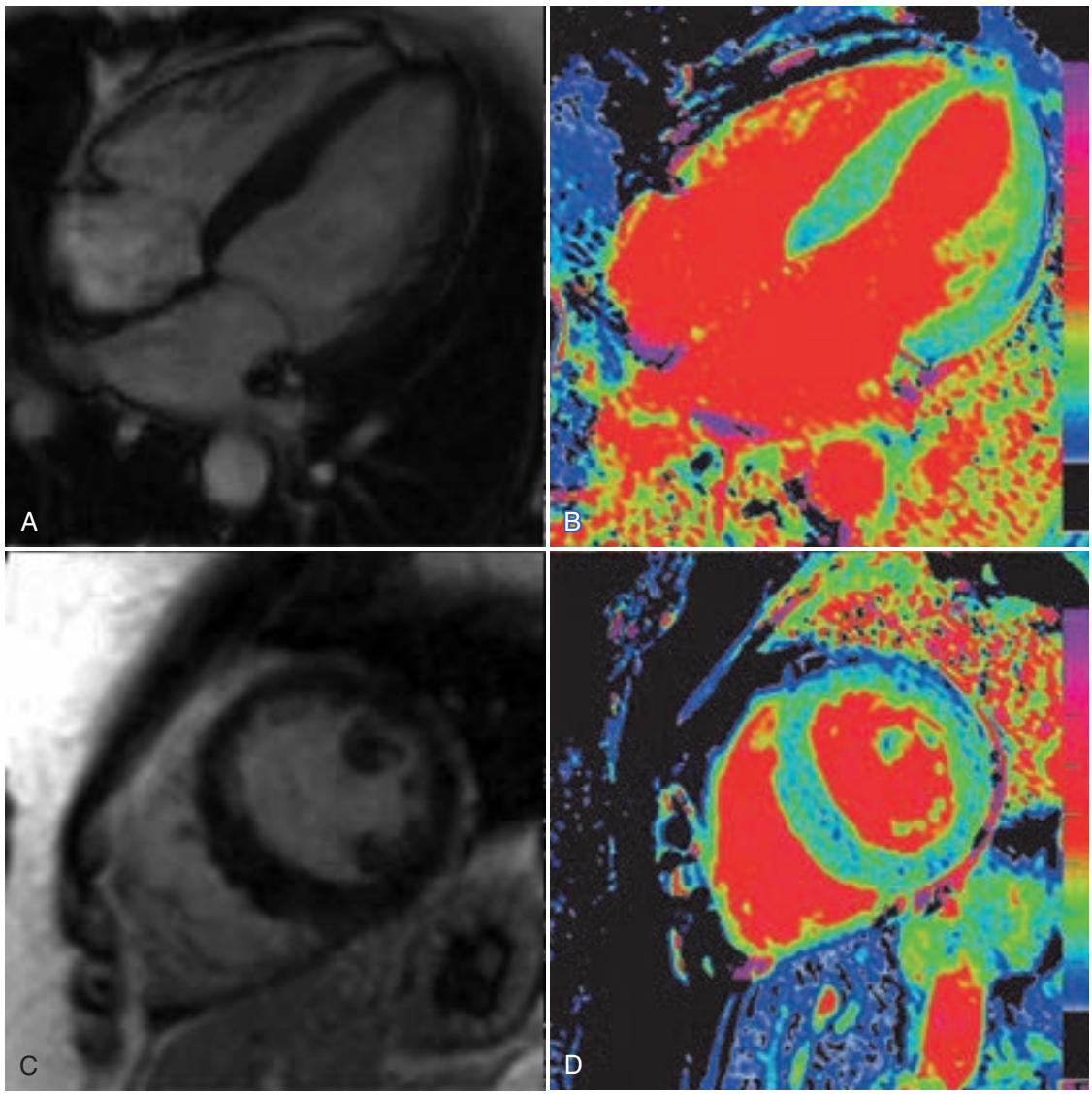

E-FIGURE 60-5. A 38-year-old man with Fabry disease and cardiac involvement. Cardiac magnetic resonance image shows a mild hypertrophy in the four-chamber (**A**) and short axis end-diastolic (**C**) views. The non–contrast-enhanced T1 map shows blue areas (T1 lowering) present circumferentially at the basal segments (**B, D**). This finding is in keeping with accumulation of glycosphingolipid in myocytes. This case does not have high T1 in the basal inferolateral wall (in many such patients, red signal indicating fibrosis on this color scale is seen here in an area matched by scar after contrast enhancement).

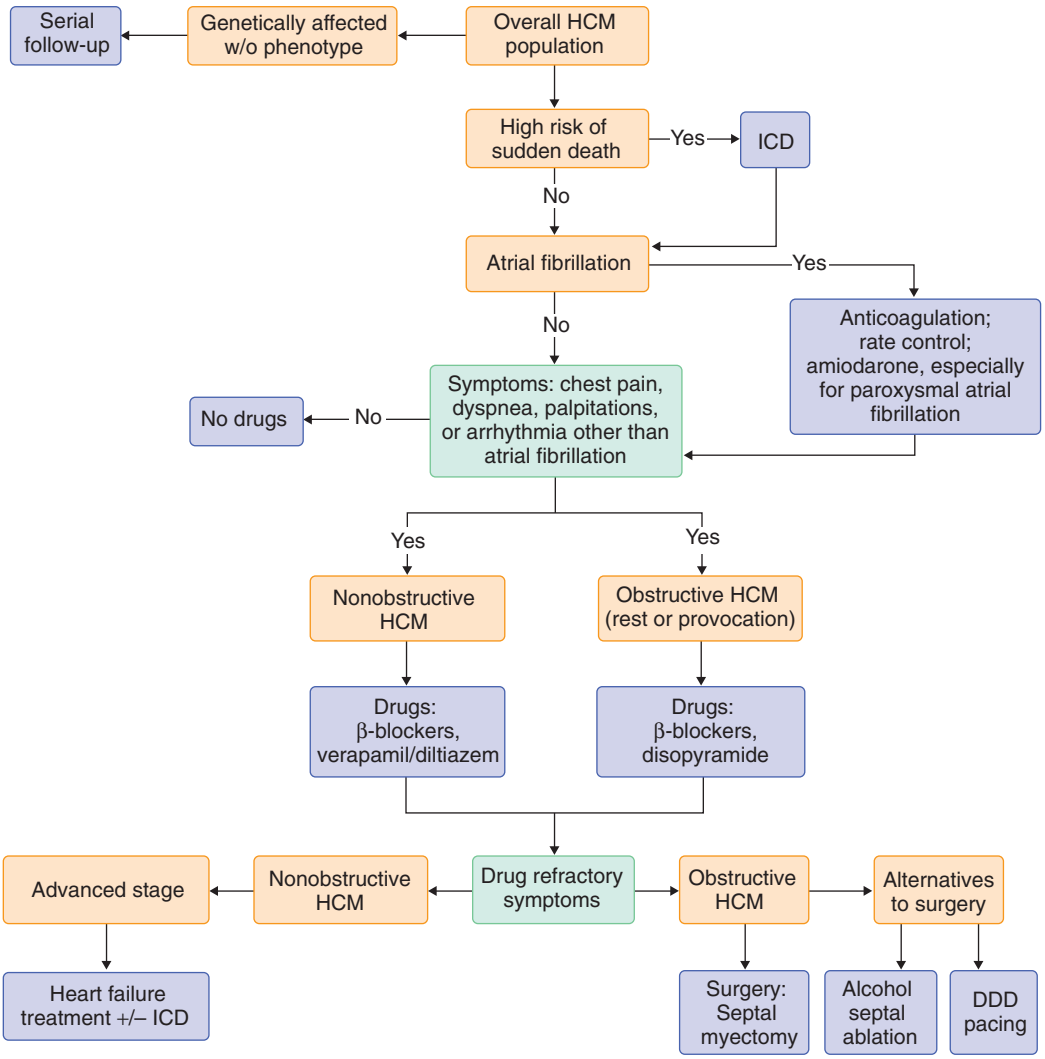

FIGURE 60-3. Approach to the management of hypertrophic cardiomyopathy (HCM). DDD = dual chamber, ICD = implantable cardioverter-defibrillator. (Modified from Maron BJ, McKenna WJ, Danielson GK, et al. American College of Cardiology/European Society of Cardiology clinical expert consensus document on hypertrophic cardiomyopathy. *J Am Coll Cardiol*. 2003;42:1687-1713.)

When concomitant procedures (e.g., mitral valve repair or replacement, coronary artery bypass grafting) are required or when other significant comorbidities are present, perioperative mortality rates are higher (4 to 5%).

In experienced centers, the selective injection of alcohol into a septal perforator branch of the left anterior descending coronary artery to create a localized septal scar yields outcomes similar to surgery. The main nonfatal complication is atrioventricular block requiring a pacemaker in 5 to 20% of patients.

Dual-chamber pacing with a short programmed atrial ventricular delay to produce maximal preexcitation while maintaining effective atrial transport can reduce the outflow gradient by 30 to 50% but provides little objective improvement in exercise capacity in most patients. Outcomes (gradient reduction, improved symptoms) are best in older patients with angulated septa and localized upper septal hypertrophy.

Supraventricular Arrhythmia

Atrial fibrillation in hypertrophic cardiomyopathy is associated with a high risk for systemic embolization, so anticoagulation (international normalized ratio in the range of 2.0 to 3.0) should be considered in all patients with sustained or paroxysmal atrial fibrillation (Chapter 64). Treatment with low-dose amiodarone, 1000 to 1400 mg/week, is effective in maintaining sinus rhythm and in controlling the ventricular response during breakthrough episodes. The addition of a low-dose β-blocker, verapamil, or diltiazem may be required for rate control. Serious side effects with low-dose amiodarone are uncommon. β-Blockers, particularly those with class III action (e.g., sotalol), are less effective alternatives. In general, the principles of managing atrial fibrillation in patients with hypertrophic cardiomyopathy are similar to those in other conditions (Chapter 64), with the provision that the threshold to use anticoagulation should be low because of the significant embolic risk.

Prevention of Sudden Death

The overall risk for sudden death in children and adults with hypertrophic cardiomyopathy is approximately 0.5 to 1% per year, but a minority of individuals have a much greater risk for ventricular arrhythmia and sudden death. The most powerful predictor of sudden cardiac death in hypertrophic cardiomyopathy is a history of previous cardiac arrest. In patients without such a history, the most useful markers of risk are a family history of premature (<40 years of age) sudden cardiac death, unexplained syncope (unrelated to neurocardiogenic mechanisms), flat or hypotensive blood pressure response to upright exercise, nonsustained ventricular tachycardia on ambulatory ECG monitoring or during exercise, and severe left ventricular hypertrophy on echocardiography (defined as a maximal left ventricular wall thickness of 30 mm or more). A clinical risk prediction model for sudden death (available online at http://www.hcmrisk.org/) can estimate an individual patient's absolute 5-year risk of sudden death.[5] Patients with an annual mortality rate of 4 to 6% based on the online risk predictor or on the presence of two or more of these markers should be considered for an implantable cardioverter-defibrillator (ICD) (Chapter 66). All patients with hypertrophic cardiomyopathy should be advised to avoid competitive sports and intense physical exertion. Patients without any risk factors do not warrant an ICD. For patients with one risk factor, decisions about an ICD should be individualized on the basis of the patient's age and severity of disease and level of risk that is acceptable to the patient.

PROGNOSIS

Most patients with hypertrophic cardiomyopathy follow a stable and benign course with a low risk for adverse events and a survival similar to that

of age- and gender-matched normal populations, but many experience progressive symptoms caused by atrial arrhythmia and gradual deterioration in left ventricular systolic and diastolic function. Between 0.5 and 1% of affected individuals die suddenly each year. The annual incidence of stroke varies from 0.56 to 0.8% per year, rising to 1.9% in patients older than 60 years, and 23% of strokes are fatal. The development of severe systolic heart failure is associated with a poor prognosis, with an overall mortality rate of up to 11% per year. The incidence of infective endocarditis is 1.4 per 1000 person-years overall but 3.8 per 1000 person-years in patients with obstruction.

Myocarditis

DEFINITION AND EPIDEMIOLOGY

Myocarditis, which is an inflammatory process involving the myocardium, can be caused by infections, immune-mediated damage, or toxins (Table 60-4). The incidence and prevalence of myocarditis are difficult to estimate because the clinical presentation varies from asymptomatic ECG abnormalities to hemodynamic collapse and sudden death. Population estimates of the prevalence of myocarditis range from 1 in 100,000 to 1 in 10,000, whereas postmortem studies report myocarditis in up to 12% of young victims of sudden cardiac death.

Worldwide, the most common infective myocarditis is Chagas disease, caused by *Trypanosoma cruzi*, a protozoan organism endemic in rural areas of South and Central America (Chapter 347). In the Western world, viral myocarditis is the most common cause of inflammatory heart disease. Human immunodeficiency virus (HIV) infection (Chapter 384) is associated with lymphocytic myocarditis and is a strong predictor of poor prognosis. Smallpox vaccination (Chapter 18) causes myopericarditis with a reported incidence of 7.8 cases per 100,000 vaccine administrations. Other rare myocarditides include giant cell myocarditis, myocarditis complicating autoimmune disorders such as systemic lupus erythematosus (Chapter 266), and cocaine abuse (Chapter 34).

TABLE 60-4 CAUSES OF MYOCARDITIS

INFECTION
Viral
Coxsackievirus, human immunodeficiency virus, echovirus, adenovirus, influenza, measles, mumps, parvovirus, poliovirus, rubella, varicella-zoster virus, herpes simplex virus, cytomegalovirus, hepatitis C virus, rabies virus, respiratory syncytial virus, vaccine virus, dengue virus, yellow fever virus
Protozoal
Trypanosoma cruzi, Toxoplasma gondii
Bacterial
Brucella, Corynebacterium diphtheriae, Salmonella, Haemophilus influenzae, Mycoplasma pneumoniae, Neisseria meningitidis (meningococcus), *Streptococcus pneumoniae, Staphylococcus, Mycobacterium, Neisseria gonorrhoeae* (gonococcus), *Vibrio cholerae*
Spirochetal
Treponema pallidum, Borrelia, Leptospira
Fungal
Aspergillus, Candida, Cryptococcus, Actinomyces, Blastomyces, Histoplasma, Coccidioides
Rickettsial
Coxiella burnetii, Rickettsia rickettsii, Rickettsia tsutsugamushi
Parasitic
Trichinella spiralis, Echinococcus granulosus, Taenia solium
IMMUNE-MEDIATED DISORDERS
Alloantigens
Heart transplant rejection
Autoantigens
Churg-Strauss syndrome, celiac disease, Whipple disease, giant cell myocarditis, Kawasaki disease, systemic lupus erythematosus, systemic sclerosis, sarcoidosis, scleroderma, polymyositis, thrombocytopenic purpura
Allergens (Drugs)
Penicillin, sulfonamides, tetracycline, methyldopa, streptomycin, tricyclic antidepressants, thiazide diuretics, dobutamine, indomethacin
TOXIC CAUSES
Drugs
Anthracyclines, catecholamines, amphetamines, cocaine, cyclophosphamide, 5-fluorouracil, trastuzumab, interferon, interleukin-2
Physical Agents
Electric shock, radiation, hyperpyrexia
Heavy Metals
Copper, iron, lead
Others
Arsenic, snake bite, scorpion bite, wasp and spider stings, phosphorus, carbon monoxide
GENETIC DISORDERS
Inherited cardiomyopathies with immune-mediated pathogenesis (dilated and right ventricular cardiomyopathy)

PATHOBIOLOGY

Myocarditis is defined histologically by the presence of myocyte injury, with degeneration or necrosis, and an inflammatory infiltrate not due to ischemia. Four patterns are recognized: *active myocarditis*, with myocyte degeneration or necrosis and definite cellular infiltrate with or without fibrosis; *borderline myocarditis*, with a definite cellular infiltrate without evidence of myocardial cellular injury; *persistent myocarditis*, with continued active myocarditis on repeated biopsy; and *resolving or resolved myocarditis*, characterized by a diminished or absent infiltrate with evidence of connective tissue healing on repeated biopsy. Despite their widespread use, the so-called Dallas criteria have low specificity and sensitivity, with a diagnostic yield as low as 10 to 20% in some series. Therefore, newer virology techniques used in conjunction with conventional light microscopy include nested polymerase chain reaction or reverse transcription polymerase chain reaction on RNA and DNA extracted from endomyocardial biopsy specimens and immunohistochemical staining for subtypes of infiltrating lymphocytes and abnormal expression of cellular adhesion molecules on interstitial or endothelial cells.

Viral Myocarditis

Most data on the pathology of viral myocarditis come from murine models. Initially, there is direct invasion of the myocardium by cardiotropic viruses, which enter the cardiomyocyte through receptor-mediated endocytosis. The viral genome, which translated intracellularly to produce viral protein or is incorporated into the host cell genome, may contribute to myocyte dysfunction by cleaving dystrophin. In the second phase, activation of the host immune system, including recruitment of natural killer cells and macrophages, increases the expression of proinflammatory cytokines such as interleukin-1 and tumor necrosis factor. Activation of $CD4^+$ T lymphocytes promotes clonal expansion of B lymphocytes, thereby resulting in further myocardial cell damage, inflammation, and the production of circulating antiheart antibodies directed against contractile, structural, and mitochondrial proteins. This autoimmune response may result in long-term ventricular remodeling by direct effects on myocardial structural components or alterations in the extracellular matrix.

CLINICAL MANIFESTATIONS

Some patients report prodromal symptoms of viremia, including fever, myalgia, coryzal symptoms, and gastroenteritis, but many individuals with myocarditis are asymptomatic and manifest only transient ECG abnormalities, such as nonspecific ST segment and T wave abnormalities, pathologic Q waves, and low QRS voltages. Less common presentations include acute myocardial infarction with angiographically normal coronary arteries (Chapter 73), atrioventricular block (Chapter 64), and ventricular arrhythmias (Chapter 65). Patients with impairment of left ventricular function may present with symptoms and signs of fulminant cardiogenic shock (Chapter 107) with acute cardiovascular collapse. In some cases, sudden cardiac death is the first presentation.

DIAGNOSIS

The diagnosis of myocarditis requires a high index of suspicion because it may mimic other common conditions. There are no typical features on echocardiography, but impaired left or right ventricular systolic performance (with or without ventricular dilation), regional wall motion abnormalities, left (or right) ventricular thrombus, diastolic impairment, and pericardial effusions may be present. Cardiac magnetic resonance imaging can detect myocardial inflammation and myocyte injury, with pericellular and cellular edema.

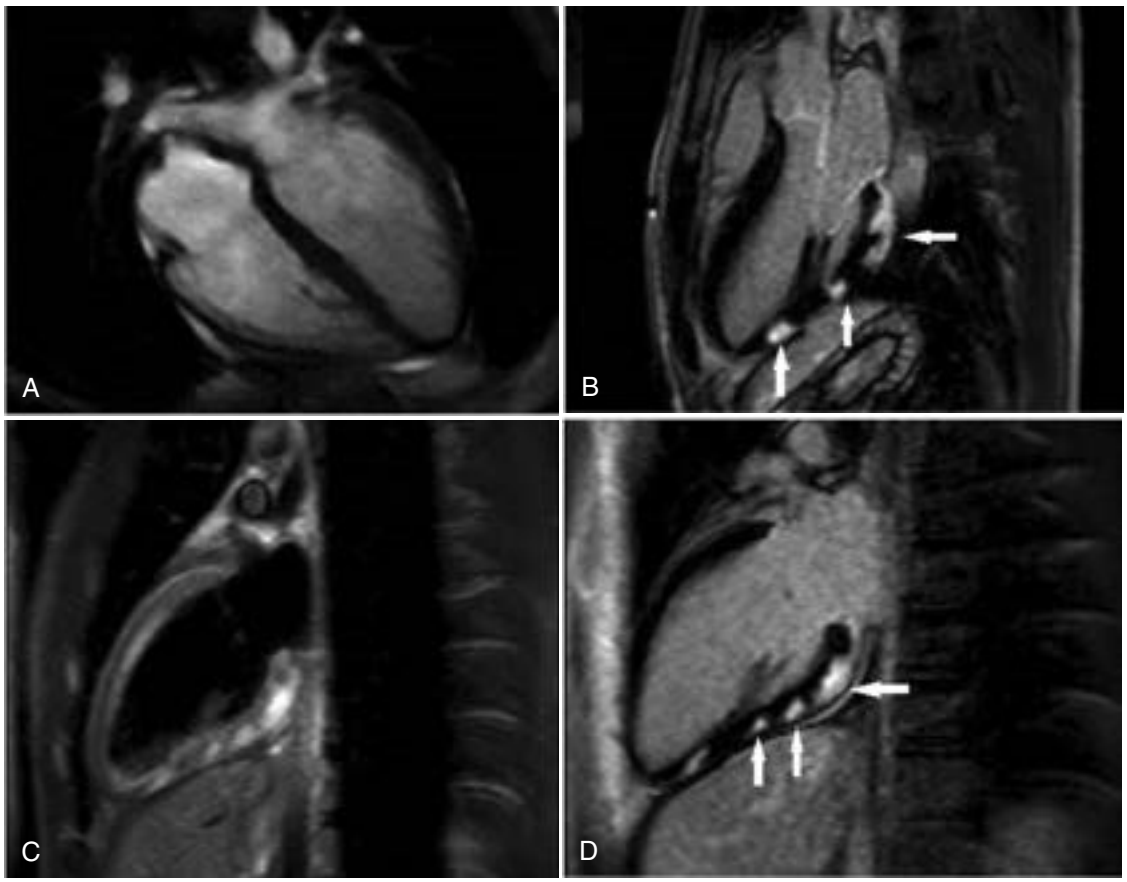

FIGURE 60-4. A 17-year-old man presenting with apparent acute myopericarditis. The four-chamber end-diastolic view (**A**) presents a mildly dilated left ventricle. There is patchy late gadolinium enhancement involving the inferior and lateral walls (**B, D,** *arrows*). Edema is seen on T1 imaging (**C**), and there is extensive late enhancement. The differential diagnosis of this appearance also includes sarcoid and giant cell myocarditis.

Routine blood tests, such as full blood count and erythrocyte sedimentation rate, are usually unhelpful. Serum markers of myocardial injury, such as troponins T and I, may be elevated, but myocarditis may be proven by biopsy even in the absence of elevated serum troponin levels. Creatine kinase and its cardiac isoform CK-MB are less sensitive and specific than troponin. Increased levels of autoantibodies against myocardial proteins (such as myosin and the adenine nucleotide translocator protein) are biomarkers of autoimmune myocarditis and correlate with progressive worsening of ventricular function.

Cardiac magnetic resonance imaging with gadolinium enhancement can show evidence of myocarditis (Fig. 60-4), but cardiac catheterization with right ventricular endomyocardial biopsy remains the "gold standard" diagnostic test. In the United States, biopsy has generally been reserved for patients with heart failure refractory to standard management, features suggestive of systemic disease (e.g., connective tissue disease [Chapter 260], amyloidosis [Chapter 188], hemochromatosis [Chapter 212], sarcoidosis [Chapter 95]), or suspicion of giant cell myocarditis because of new-onset heart failure associated with tachyarrhythmias or conduction disease. By comparison, a European consensus statement recommends endomyocardial biopsy to achieve an etiologic diagnosis and to guide potential novel treatment options in patients with "clinically suspected myocarditis," [6] despite the absence of definitive outcome data to support this recommendation.

Specific Causes

Viral myocarditis may be suspected from the clinical picture of recent febrile illness, often with prominent myalgias, followed by angina-like chest pain, dyspnea, or arrhythmias. Elevated troponin levels support the diagnosis, and increasing viral titers (to coxsackievirus, echovirus, adenovirus, or influenza virus) are consistent with recent infection. The correlation with biopsy-proven myocarditis is strongest with HIV and Lyme disease. Clinical cardiomyopathy occurs in 10 to 40% of patients infected with HIV due to the HIV infection itself or to coinfection with cytomegalovirus.

Giant cell myocarditis, which accounts for 10 to 20% of biopsy-positive cases of myocarditis, is manifested with the rapid onset of chest pain, fever, and hemodynamic compromise, often with ventricular tachycardia or atrioventricular block. When ventricular tachyarrhythmias or progressive heart failure are major features of clinically suspected myocarditis, particularly in a young person, endomyocardial biopsy is recommended to determine whether giant cell myocarditis is present.

Toxoplasmosis (Chapter 349) *myocarditis*, due to intermittent rupture of cysts in the myocardium, can cause atypical chest pain, arrhythmias, pericarditis, and symptomatic heart failure. Diagnosis is made from antibody titers. Lyme carditis (Chapter 321) is classically manifested with conduction system abnormalities resulting from infection with *Borrelia burgdorferi*, which is diagnosed serologically.

Immune-mediated myocarditis can be associated with polymyositis (Chapter 269) or systemic lupus erythematosus (Chapter 266), although pericarditis and coronary artery vasculitis are more common. Hypersensitivity reactions, especially to drugs (Chapter 254), can cause myocarditis that is often associated with peripheral eosinophilia and can be confirmed by endomyocardial biopsy.

TREATMENT

Rx

The first-line treatment of myocarditis is supportive with afterload reduction and diuresis (Chapter 59). Patients with fulminant acute myocarditis may require inotropic support, mechanical assist devices, or extracorporeal membrane oxygenation (Chapter 107). After initial stabilization, patients with symptoms and signs of heart failure should receive angiotensin-converting enzyme inhibitors, diuretics, β-blockers, and anticoagulants in accordance with standard guidelines (Chapter 59). Patients with intractable and deteriorating heart failure may require cardiac transplantation (Chapter 82).

The role of immunosuppression is uncertain. In one randomized, placebo-controlled trial of 111 adults with biopsy-proven myocarditis, there was no difference in mortality or improvement in left ventricular function in patients treated with prednisolone plus either cyclosporine or azathioprine. Conversely, in a randomized trial of patients who had major histocompatibility

complex expression on endomyocardial biopsy samples and who were randomized to prednisolone (1 mg/kg/day tapering to a maintenance dose of 0.2 mg/kg/day for a total of 90 days) and azathioprine (1 mg/kg/day for a total of 100 days) versus placebo, left ventricular ejection fraction improved in the immunosuppressed group, but no difference was observed in mortality or rates of transplantation or rehospitalization during a 2-year follow-up period.[6] Immunosuppression has been reserved for patients with giant cell myocarditis,[6] but recent recommendations now extend immunosuppression to other biopsy-proven, infection-negative immune-mediated forms of myocarditis including eosinophilic myocarditis, cardiac sarcoidosis, and lymphocytic myocarditis that is refractory to conventional heart failure therapy. The optimal regimens remain to be determined, and intravenous immune globulin therapy is not helpful.

PROGNOSIS

Patients with acute myocarditis with mild heart failure or symptoms suggestive of myocardial ischemia or infarction typically improve within weeks without sequelae. An acute presentation of myocarditis with advanced heart failure (ejection fraction <35%) may resolve but can lead to chronic left ventricular dysfunction (dilated cardiomyopathy) or progress to death or cardiac transplantation. Patients who present with acute fulminant myocarditis, however, may have an excellent prognosis, with survival rates of more than 90%. Overall, however, biopsy-proven viral myocarditis is associated with a long-term mortality of almost 20% at 4.7 years, and the presence of biventricular dysfunction at presentation is the best predictor of all-cause mortality. Giant cell myocarditis is usually fatal without heart transplantation, but it can be stabilized by early diagnosis and prompt introduction of immunosuppression.[7]

Dilated Cardiomyopathy

DEFINITION AND EPIDEMIOLOGY

Dilated cardiomyopathy is a heart muscle disorder defined by dilation and impaired systolic function of the left ventricle or both ventricles, in the absence of coronary artery disease, valvular abnormalities, or pericardial disease. In adults, prevalence estimates range from 14 to 36 per 100,000. In children, dilated cardiomyopathy is the most common cardiomyopathy, accounting for up to 58% of cases. Overall, males and females are approximately equally affected, except for dilated cardiomyopathy associated with neuromuscular disorders or inborn errors of metabolism, for which there is male predominance because some of these conditions have an X-linked inheritance.

PATHOBIOLOGY

A number of conditions are associated with dilated cardiomyopathy, including neuromuscular disorders, inborn errors of metabolism, and malformation syndromes. In most patients, no identifiable cause is found, and the disease is termed idiopathic dilated cardiomyopathy.

Genetic Dilated Cardiomyopathy

Between 20 and 50% of individuals with dilated cardiomyopathy have evidence of familial disease (see Table 60-2).[8] Autosomal dominant inheritance, which accounts for 68% of familial cases, has two major forms: isolated dilated cardiomyopathy, which is manifested with a clinical picture of heart failure; and dilated cardiomyopathy, in which an associated arrhythmia (i.e., conduction system disease, ventricular tachycardia or fibrillation) is usually the initial manifestation. The latter patients may also have an associated skeletal myopathy. Genes implicated in isolated dilated cardiomyopathy include cytoskeletal and sarcomeric protein genes. Truncating mutations in *TTN*, the gene encoding the sarcomere protein titin, occur in approximately 25% of familial cases of idiopathic dilated cardiomyopathy and in 18% of sporadic cases. Mutations in the lamin A/C gene, which encodes a nuclear envelope protein, cause atrial and ventricular arrhythmia and progressive atrioventricular conduction disease, which may precede the development of dilated cardiomyopathy.

X-linked inheritance accounts for between 2 and 5% of familial cases of dilated cardiomyopathy. Neuromuscular disorders account for 26% of cases, 90% of which are Duchenne, Becker, and Emery-Dreifuss muscular dystrophies (Chapter 421). Isolated X-linked dilated cardiomyopathy, also caused by mutations in the dystrophin gene, is characterized by raised serum creatine kinase muscle isoforms but does not result in clinical signs or symptoms of skeletal muscular dystrophy.

Acquired Dilated Cardiomyopathy

Common acquired causes of dilated cardiomyopathy include infectious myocarditis, chemotherapy (Chapter 179), radiation therapy (Chapter 20), alcohol (Chapter 33), cocaine (Chapter 34), nutritional deficiencies (Chapter 215), iron overload (Chapter 212), inflammatory and autoimmune disorders (Chapters 266 and 270), endocrinopathies (Chapter 226), and pregnancy (Chapter 239). Tachycardia-mediated cardiomyopathy (tachycardiomyopathy) is rare and usually reverses once the tachycardia is controlled.

CLINICAL MANIFESTATIONS

The symptoms and signs associated with dilated cardiomyopathy depend on the age of the patient and the degree of left ventricular dysfunction. Although the first presentation may be with sudden death or a thromboembolic event, most patients present with symptoms of high pulmonary venous pressure or low cardiac output (Chapter 58), which can be acute, sometimes precipitated by intercurrent illness or arrhythmia, or chronic. Increasingly, dilated cardiomyopathy is diagnosed incidentally in asymptomatic individuals during family screening.

Adults initially present with reduced exercise tolerance and dyspnea on exertion. With worsening left ventricular function, patients may develop dyspnea at rest, orthopnea, paroxysmal nocturnal dyspnea, peripheral edema, and ascites. Symptoms related to mesenteric ischemia, such as abdominal pain after meals, nausea, vomiting, and anorexia, may dominate, especially in children. Arrhythmia symptoms, such as palpitations, presyncope, and syncope, can occur at any age.

In advanced disease, features of low cardiac output include sinus tachycardia, weak peripheral pulses, and hypotension. The jugular venous pressure may be elevated, and the apical impulse is displaced. Peripheral edema, hepatomegaly, and ascites are common in patients with heart failure. Auscultation of the chest typically reveals basal crackles. Auscultation of the heart may reveal the presence of a third (and sometimes also a fourth) heart sound. In patients with functional mitral regurgitation, a pansystolic murmur may be heard at the apex and radiate to the axilla, but frequently no murmurs are heard, even in the presence of mitral incompetence, especially if cardiac output is very low.

DIAGNOSIS

The electrocardiogram may be normal but more typically shows sinus tachycardia, nonspecific ST segment and T wave changes (most commonly in the inferior and lateral leads), atrial enlargement, and voltage criteria for ventricular hypertrophy. Atrioventricular block raises the possibility of mutations in the lamin A/C gene. Supraventricular and ventricular arrhythmias are common. The chest radiograph is usually abnormal, with an increased cardiothoracic ratio (>0.5) reflecting left ventricular and left atrial dilation. Patients with pulmonary edema have increased pulmonary vascular markings and pleural effusion.

On echocardiography, the presence of ventricular end-diastolic dimensions greater than 2 standard deviations above body surface area–corrected means (or greater than 112% of predicted dimension) and fractional shortening less than 25% are sufficient to make the diagnosis. Other common features include functional mitral and tricuspid regurgitation and abnormalities of diastolic left ventricular function. Cardiac magnetic resonance imaging may show areas of myocardial fibrosis (E-Fig. 60-6).

Other recommended tests (Table 60-5) include a complete blood count and tests of renal, thyroid, and hepatic function. Levels of serum creatine kinase should be measured in all patients with dilated cardiomyopathy because this may provide important clues to the etiology. For example, dystrophin-linked dilated cardiomyopathy has been diagnosed in up to 8% of men with dilated cardiomyopathy and should be considered in men with increased serum creatine kinase levels and an X-linked family history. Other cardiac biomarkers, such as troponin I and troponin T, can be elevated. Plasma B-type natriuretic peptide levels predict survival, hospitalization rates, and listing for cardiac transplantation. Symptom-limited exercise testing, combined with respiratory gas analysis, is a useful technique to assess functional limitation and disease progression in patients with stable dilated cardiomyopathy.

Cardiac catheterization is rarely needed except perhaps to exclude severe coronary artery disease or to provide more precise information about possible valvular heart disease. Endomyocardial biopsy may be diagnostic for myocarditis and for some metabolic or mitochondrial disorders but is rarely advised. Hemodynamic assessment of left ventricular end-diastolic and pulmonary artery pressures may be necessary before transplantation.

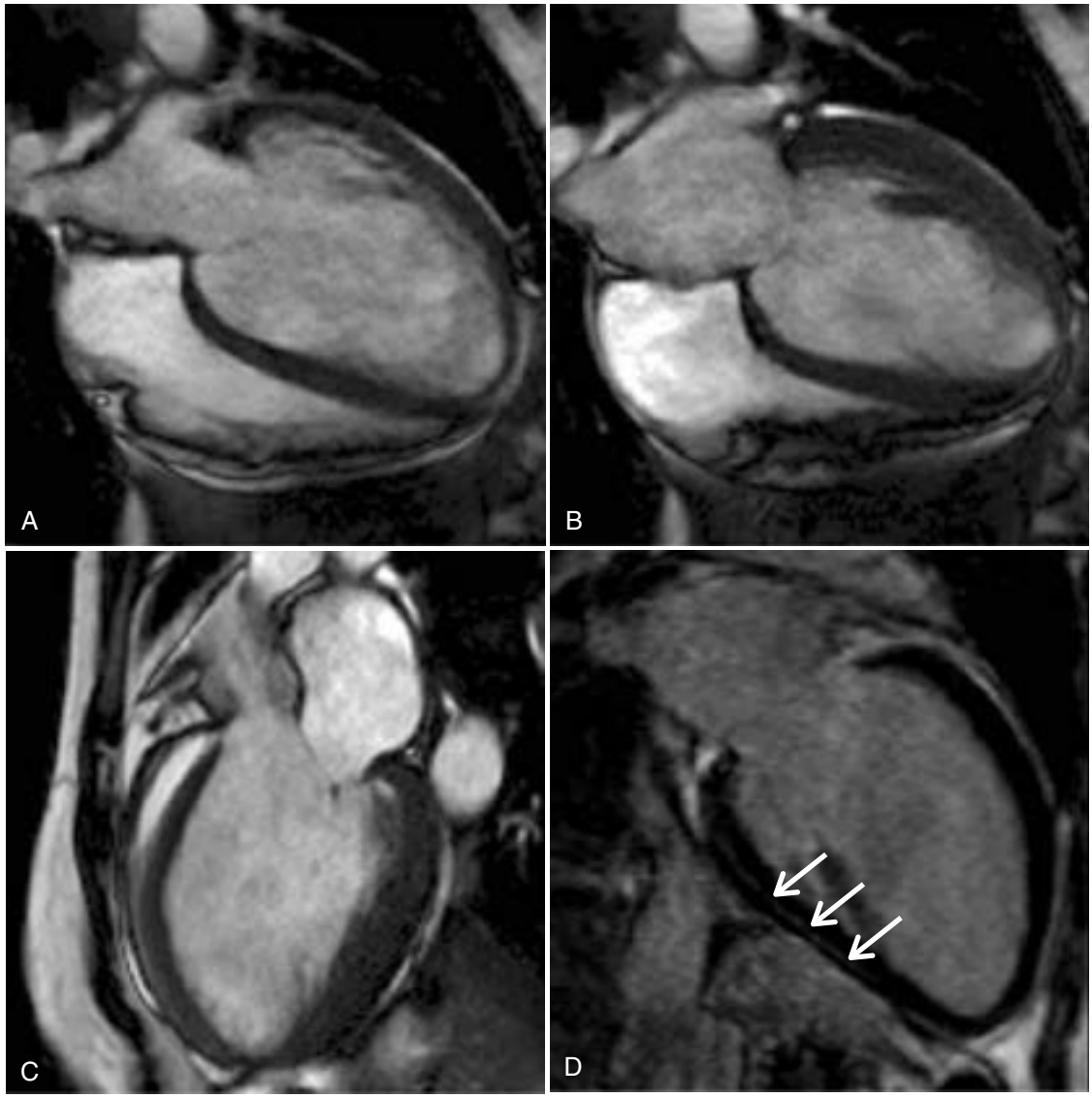

E-FIGURE 60-6. A 45-year-old woman with dilated cardiomyopathy, left bundle branch block, and heart failure. The left ventricular ejection fraction is 27%, and the left ventricle is severely dilated (diastolic volume of 288 mL; **A**, diastolic; **B, C**, systolic frames). There is a rim of mid-myocardial late gadolinium enhancement at the inferior wall (*arrows*, **D**).

TABLE 60-5 LABORATORY EVALUATION OF CARDIOMYOPATHY

CLINICAL EVALUATION
History and physical examination to identify cardiac and noncardiac disorders*
Assessment of ability to perform routine and desired activities*
Assessment of volume status*
LABORATORY EVALUATION
Electrocardiogram*
Chest radiograph*
Two-dimensional and Doppler echocardiogram*
Chemistry
Serum sodium,* potassium,* glucose, creatinine,* blood urea nitrogen,* calcium,* magnesium*
Albumin,* total protein,* liver function tests,* serum iron, ferritin
Urinalysis
Creatine kinase
Thyroid-stimulating hormone*
Hematology
Hemoglobin/hematocrit*
White blood cell count with differential,* including eosinophils
Erythrocyte sedimentation rate
INITIAL EVALUATION IN SELECTED PATIENTS ONLY
Titers for suspected infection
Acute viral (coxsackievirus, echovirus, influenza virus)
Human immunodeficiency virus, Epstein-Barr virus
Lyme disease, toxoplasmosis
Chagas disease
Catheterization with coronary angiography in patients with angina who are candidates for intervention*
Serologic studies for active rheumatologic disease
Endomyocardial biopsy

*Level I recommendations from Hunt SA, Abraham WT, Chin MH, et al. ACC/AHA 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult. *Circulation.* 2005;112:e154-e235.

TREATMENT

Rx

Supportive therapy includes sodium and fluid restriction, avoidance of alcohol and other toxins, and use of established heart failure medications (Chapter 59). Although older recommendations emphasized rest and avoidance of exercise, this advice should be limited to patients with myocarditis or peripartum cardiomyopathy; for other patients, a submaximal exercise regimen is desirable to sustain mobility, to avoid deconditioning, and to maintain physical and psychological health. Patients with atrial fibrillation or with echocardiographic evidence of a left atrial or left ventricular mural thrombosis should be anticoagulated to an international normalized ratio of 2.0 to 3.0. An ICD is preferred to medication for ventricular arrhythmias,■ and some patients require management for advanced heart failure (Chapter 59) with biventricular pacing, inotropic medications, ventricular assist devices, and cardiac transplantation (Chapter 82).

Family Screening

Familial evaluation of first-degree relatives by history and physical examination and with 12-lead ECG and two-dimensional echocardiographic studies is warranted at the time of diagnosis and serially thereafter. Precise algorithms to guide the interval of evaluation remain to be determined; because disease progression is usually slow, evaluation about every 5 years until the age of 50 years appears appropriate. The detection of early disease in a family member offers an opportunity to initiate treatment, usually with an angiotensin-converting enzyme inhibitor or β-blocker, but the efficacy of such therapy remains to be proved.

PROGNOSIS

The prognosis of idiopathic and genetically determined dilated cardiomyopathy is related to the severity of disease at the time of presentation and the response to treatment. Most patients improve with treatment, but 5-year survival is less than 50% in patients who present with severe disease (e.g., ejection fraction <25%, left ventricular end-diastolic dimension >65 mm, peak oxygen consumption <12 mL/kg/minute).[9]

Specific Causes of Dilated Cardiomyopathy

Alcoholic Cardiomyopathy

In the United States, excess alcohol consumption (Chapter 33) contributes to more than 10% of cases of heart failure. Alcohol and its metabolite, acetaldehyde, are cardiotoxins. Myocardial depression is initially reversible but, if alcohol consumption is sustained, can lead to myocyte vacuolization, mitochondrial abnormalities, and myocardial fibrosis. Even in chronic stages, however, the heart failure represents a sum of both reversible and irreversible myocardial dysfunction. The amount of alcohol necessary to produce symptomatic cardiomyopathy in susceptible individuals is not known but has been estimated to be six drinks (~4 oz of pure ethanol) a day for 5 to 10 years. Frequent binging without heavy daily consumption may also be sufficient. Alcoholic cardiomyopathy can develop in patients without social evidence of an alcohol problem. Abstinence leads to improvement in at least 50% of patients with severe symptoms, some of whom normalize their left ventricular ejection fractions. Patients with other causes of heart failure also should limit alcohol consumption.

Chemotherapy

Anthracycline (doxorubicin, daunorubicin, epirubicin) cardiotoxicity (Chapter 179) causes characteristic histologic changes on endomyocardial biopsy with overt heart failure in 5 to 10% of patients who receive doses of 450 mg/m^2 of body surface area or more. In adults who receive these drugs, combined treatment with enalapril (starting at 1.25 mg or 2.5 mg twice daily, increasing to 10 mg twice daily as tolerated with systolic blood pressure ≥90 mm Hg) and carvedilol (starting at 6.25 mg twice daily and increasing to 25 mg twice daily if there is no heart failure, bradycardia, or atrioventricular block) can significantly reduce the risk of left ventricular dilation and heart failure.■ Patients who have received anthracyclines in the prepubertal period without apparent cardiotoxicity may develop cardiac failure in young adulthood. The risk is higher in patients who have lower baseline ejection fractions, concomitant radiation therapy, or higher doses of anthracycline. *Cyclophosphamide* and *ifosfamide* can cause acute severe heart failure and malignant ventricular arrhythmias. Some *tyrosine kinase inhibitors* (e.g., sunitinib) cause a reduction in systolic function, especially in the presence of coronary artery disease, but there is good response to withdrawal and conventional medical therapy (Chapter 184). *5-Fluorouracil* can cause coronary artery spasm and depressed left ventricular contractility. Up to 11% of patients who receive *trastuzumab* (Chapter 198), a recombinant monoclonal antibody that binds to human epidermal growth factor type 2, develop dilated cardiomyopathy, which is reversible after withdrawal and conventional drug treatment. The risk for cardiotoxicity increases with previous anthracycline and radiation treatment. *Interferon alfa* may be associated with hypotension and arrhythmias in up to 10% of patients, and *interleukin-2* rarely has been associated with cardiotoxicity.

Metabolic and Endocrine Disease

Excess catecholamines, as in *pheochromocytoma* (Chapter 228), may injure the heart by compromising the coronary microcirculation or by direct toxic effects on myocytes. *Cocaine* (Chapter 34) increases synaptic concentrations of catecholamines by inhibiting reuptake at nerve terminals; the result may be an acute coronary syndrome or chronic cardiomyopathy.

Thiamine deficiency from poor nutrition or alcoholism (Chapter 218) can cause beriberi heart disease, with vasodilation and high cardiac output followed by low output. *Calcium deficiency* resulting from hypoparathyroidism, gastrointestinal abnormalities, or chelation directly compromises myocardial contractility.

Hypophosphatemia (Chapter 119), which may occur in alcoholism, during recovery from malnutrition, and in hyperalimentation, also reduces myocardial contractility. Patients with *magnesium depletion* due to impaired absorption or increased renal excretion (Chapter 119) also may present with left ventricular dysfunction.

Hypothyroidism (Chapter 226) depresses contractility and conduction and may cause pericardial effusions, whereas *hyperthyroidism* increases cardiac output, can worsen underlying heart failure, and may rarely be the sole cause of heart failure.

The presenting sign of *diabetes* (Chapter 229) can be cardiomyopathy, especially with diastolic dysfunction, independent of epicardial coronary atherosclerosis, for which it is a major risk factor.

Obesity (Chapter 220) can cause cardiomyopathy with increased ventricular mass and decreased contractility, which improve after weight loss, or it can aggravate underlying heart failure from other causes.

Peripartum Cardiomyopathy

Peripartum cardiomyopathy appears in the last month of pregnancy or in the first 5 months after delivery in the absence of preexisting cardiac disease (Chapter 239). The incidence is between 1 in 3000 and 1 in 15,000 deliveries, with increased risk in older mothers or in the setting of twins, malnutrition, tocolytic therapy, toxemia, or hypertension. Lymphocytic myocarditis, found in 30 to 50% of biopsy specimens, suggests an immune component, perhaps cross-reactivity between uterine and cardiac myocyte proteins or an enhanced susceptibility to viral myocarditis. More recently, it has been suggested that enhanced oxidative stress triggers activation of cathepsin D, an ubiquitous lysosomal enzyme that cleaves serum prolactin in its antiangiogenic and proapoptotic 16-kD form, which appears to promote endothelial inflammation and impair cardiomyocyte metabolism and contraction. Presentation is usually with orthopnea and dyspnea on minimal exertion, most often within the first weeks after delivery when the excess volume of pregnancy would normally be mobilized. Preexisting cardiac disease must be excluded. Diuretics facilitate postpartum diuresis, and angiotensin-converting enzyme inhibitors improve symptoms (Chapter 59). In a small randomized trial, oral bromocriptine (2.5 mg twice daily for 2 weeks, then daily for 6 weeks) significantly improved recovery of left ventricular function and may reduce deaths.■ The prognosis is improvement to normal or near-normal ejection fraction during the next 6 months in more than 50% of patients. About 4% require heart transplantation, and about 9% die suddenly or from complications of heart transplantation.

Overlap with Restrictive Cardiomyopathy

Diseases causing primarily restrictive cardiomyopathies can occasionally overlap to cause a picture consistent with dilated cardiomyopathy. For example, *hemochromatosis* (Chapter 212) and *sarcoidosis* (Chapter 95) should be considered in evaluating any patient with a cardiomyopathy, although these conditions are more often considered with the restrictive diseases. *Amyloidosis* (Chapter 188) is less commonly confused with dilated than with hypertrophic cardiomyopathy but should be considered in a patient with a thick-walled ventricle with moderately depressed contractile function.

Arrhythmogenic Right Ventricular Cardiomyopathy

DEFINITION AND EPIDEMIOLOGY

ARVC (Chapter 65) is a genetically determined heart muscle disorder characterized histologically by loss of cardiomyocytes with replacement by fibrous or fibrofatty tissue in the right ventricular myocardium; clinically by ventricular arrhythmias, heart failure, and sudden death; and histologically by cardiomyocyte loss and replacement. The disease is seen in patients of European, African, and Asian descent, with an estimated prevalence between 1 in 1000 and 1 in 5000 adults.

PATHOBIOLOGY

ARVC is inherited as an autosomal dominant disease with incomplete penetrance, although recessive forms with cutaneous manifestations are recognized (see Table 60-2). Most cases are caused by heterozygous mutations in genes encoding components of the desmosomal junction of cardiomyocytes. The most common occur in plakophilin 2, desmocollin 2, desmoplakin, and desmoglein 2. Homozygous mutations in plakoglobin and desmoplakin are responsible for the rare autosomal recessive forms (i.e., Naxos disease and Carvajal syndrome). A highly penetrant and lethal mutation in the transmembrane cytoplasmic protein 43 (TMEM43) has been described in families from Newfoundland. Two other nondesmosomal genes, the cardiac ryanodine receptor and transforming growth factor-β3, have been linked with ARVC but are probably not important in most patients.

Pathology

The main pathologic feature is progressive loss of right ventricular myocardium, which is replaced by adipose and fibrous tissue. These changes begin in the inflow, outflow, and apical regions of the right ventricle. Aneurysm formation in these areas is typical. Progressive myocardial involvement may lead to global right ventricular dilation. Severe right ventricular disease is often associated with fibrofatty substitution of the left ventricular myocardium, with the posterolateral wall preferentially affected.

Mutations in desmosomal protein genes may increase the susceptibility of the myocardium to the damaging effects of mechanical stress, thereby predisposing to cardiomyocyte detachment, death, and eventual replacement with fibrofatty tissue. The acute phase of myocardial injury may be accompanied by inflammation. The predilection for the right ventricle has been explained by its thin wall and greater distensibility. As desmosomal proteins interact with many other proteins, including components of the cellular cytoskeleton and intermediate filaments, it is possible that ventricular dysfunction occurs as the result of reduced cytoskeletal integrity and impaired force transduction. Some desmosomal proteins, in particular plakoglobin, are also important signaling molecules that regulate the transcription of many other genes. Finally, a reduction in the number and size of gap junctions may result in a electrical coupling defect, thereby increasing the propensity to arrhythmia without significant morphologic changes.

CLINICAL MANIFESTATIONS

By convention, the natural history of ARVC is divided into phases, but it is not inevitable that patients will progress through all phases. In the early phase, patients are usually asymptomatic, but resuscitated cardiac arrest and sudden death may be the initial manifestations, particularly in adolescents and young adults. The overt arrhythmic phase usually begins in adolescents and young adults, when patients note palpitations or syncope. Symptomatic sustained arrhythmias are usually accompanied by ECG, morphologic, and functional abnormalities of the right ventricle sufficient to fulfill diagnostic criteria for ARVC. A small proportion of patients progress to a more advanced phase, which is characterized by diffuse right or left ventricular impairment that requires conventional treatment for heart failure (Chapter 59).

DIAGNOSIS

Clinical evaluation includes inquiry for symptoms of arrhythmia (syncope, presyncope, sustained palpitation); a family history of premature cardiac symptoms or sudden death; 12-lead, 24-hour, and maximal exercise ECG testing; and two-dimensional echocardiography with specific right ventricular views. Contrast echocardiography may be required to obtain better endocardial definition of the right ventricular myocardium and apex of the left ventricle. Magnetic resonance imaging (Chapter 56) may provide accurate assessment of ventricular volumes as well as noninvasive characterization of the characteristic fibrous tissue and fat that establishes the diagnosis and provides prognostic information (Fig. 60-5).[10]

Ventricular arrhythmias with a left bundle branch block morphology, consistent with a right ventricular origin, are characteristic. However, the ECG and arrhythmic manifestations are not specific to ARVC and overlap with many other disease states, so standard criteria are recommended for diagnosis (Table 60-6). Because these criteria are highly specific but lack sensitivity for detection of early disease, more sensitive criteria are recommended for first-degree relatives of known cases (Table 60-7). The diagnosis of ARVC in a proband also raises the possibility of mutation analysis in the family to identify those at risk and in need of serial evaluation as well as those who need no specific follow-up.

Differential Diagnosis

The differential diagnosis includes other inherited cardiomyopathies, the inherited arrhythmia syndromes (long QT syndrome, Brugada syndrome, and catecholaminergic polymorphic ventricular tachycardia; Chapter 65), cardiac sarcoidosis, myocarditis, and causes of right ventricular dilation such as intracardiac or extracardiac shunts (Chapter 69). The differentiation from so-called benign right ventricular outflow tract tachycardia may be problematic, although the 12-lead ECG and right ventricular imaging studies are typically normal, and no familial disease is present. Some patients with desmosomal protein gene mutations demonstrate left ventricular involvement early in the disease, and a minority may have a predominant left ventricular dilated cardiomyopathy phenotype.

TREATMENT

Pharmacologic treatment is the first-line therapy for patients with well-tolerated, non–life-threatening ventricular arrhythmias, such as frequent ventricular extrasystoles. Treatment of patients with symptomatic ventricular arrhythmias is with an ICD, with supplemental sotalol (160 to 240 mg/day) or even amiodarone (maintenance dose of 200 mg/day). Catheter ablation (Chapter 66) may be required in patients with drug-refractory incessant ventricular arrhythmia or frequent recurrences of ventricular tachycardia after implantation of an ICD, although recurrence is common.

Retrospective analyses of clinical and pathologic series have identified a number of possible predictors of adverse outcome in probands, including an early age at onset of symptoms, competitive sporting activity, severe right ventricular dilation, left ventricular involvement, syncope, episodes of complex

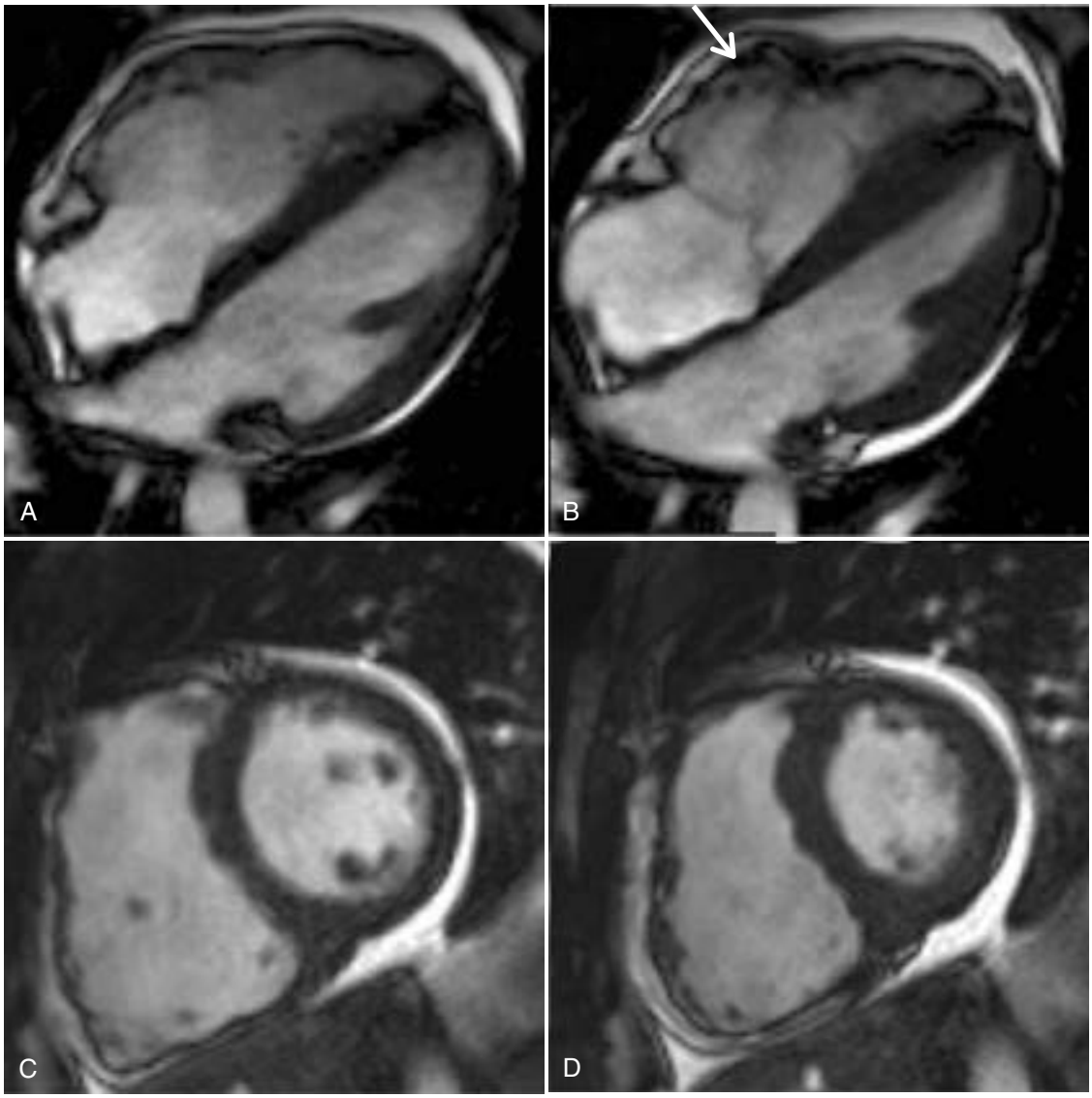

FIGURE 60-5. A 21-year-old man with arrhythmogenic right ventricular cardiomyopathy. End-diastolic (**A**) and end-systolic (**B**) frames from a four-chamber view, with a dilated and impaired right ventricle with a basal wall motion abnormality (*arrow*), confirmed on end-diastolic (**C**) and end-systolic (**D**) short axis views. This individual had no scar in the left ventricle with well preserved systolic function; scar imaging in the thin right ventricle (with adjacent fat and small effusion) was equivocal.

ventricular arrhythmias or VT, and increased QRS dispersion on the 12-lead electrocardiogram. ICD implantation is recommended for the prevention of sudden cardiac death in patients with documented sustained ventricular tachycardia or ventricular fibrillation and a reasonable expectation of survival with a good functional status for longer than 1 year. ICD implantation may also be appropriate in patients with extensive disease, including those with left ventricular involvement, or undiagnosed syncope when ventricular tachycardia or ventricular fibrillation has not been excluded as the cause.

Standard heart failure therapy, including diuretics, angiotensin-converting enzyme inhibitors, and β-blockers, is indicated in patients in whom ARVC has progressed to severe heart failure or biventricular systolic dysfunction (Chapter 59). Anticoagulation should be considered in the presence of atrial fibrillation (Chapter 64), marked ventricular dilation, or ventricular aneurysms. In patients in whom heart failure is refractory, cardiac transplantation (Chapter 82) should be considered.

PROGNOSIS

Most data on prognosis in ARVC are derived from small, high-risk populations. By the age of 40 years, event-free survival is 50 to 60% in patients with Naxos disease and some autosomal dominant forms. In patients who have syncope or sustained ventricular arrhythmias and are treated with an ICD, freedom from appropriate shock therapy is about 75% at 48 months after implantation, with 96% of patients alive. Risk factors for sudden cardiac death include severe right ventricular disease, left ventricular involvement, and a history of unexplained syncope.

Restrictive Cardiomyopathy

DEFINITION AND EPIDEMIOLOGY

The incidence and prevalence of restrictive cardiomyopathy in adults are unknown. Restrictive cardiomyopathies (Table 60-8) are characterized by stiffness, impaired filling, elevated left ventricular diastolic pressures, and reduced diastolic volume of the left or right ventricle despite normal or near-normal systolic function and wall thickness. Primary forms are uncommon, whereas secondary forms, in which the heart is affected as part of a multisystem disorder, usually present at the advanced stage of an infiltrative disease (e.g., amyloidosis or sarcoidosis) or a systemic storage disease (e.g., hemochromatosis). Idiopathic restrictive cardiomyopathy affects both male and female patients and may be manifested in children and young adults.

PATHOBIOLOGY

Approximately 30% of patients with idiopathic restrictive cardiomyopathy have familial disease, and most of these patients will have mutations in the cardiac sarcomere protein genes, particularly troponin I and β-myosin heavy chain. Mutations in the gene encoding desmin (an intermediate filament) cause restrictive cardiomyopathy associated with skeletal myopathy and cardiac conduction system abnormalities.

The macroscopic features of restrictive cardiomyopathy include biatrial dilation and small ventricular cavities. In many hearts, there is thrombus in the atrial appendages and patchy endocardial fibrosis. The histologic features of idiopathic restrictive cardiomyopathy are typically nonspecific with patchy interstitial fibrosis, but myocyte disarray is not uncommon in patients with

TABLE 60-6 REVISED TASK FORCE CRITERIA FOR ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY IN PROBANDS*

CRITERIA	
MAJOR	**MINOR**
I. GLOBAL OR REGIONAL DYSFUNCTION AND STRUCTURAL ALTERATIONS*	
By Two-Dimensional Echo • Regional RV akinesia, dyskinesia, or aneurysm • *and* 1 of the following (end diastole): PLAX RVOT ≥32 mm (corrected for body size [PLAX/BSA] ≥19 mm/m^2) PSAX RVOT ≥36 mm (corrected for body size [PLAX/BSA] ≥21 mm/m^2) *or* fractional area change ≤33%	**By Two-Dimensional Echo** • Regional RV akinesia or dyskinesia • *and* 1 of the following (end diastole): PLAX RVOT ≥29 to <32 mm (corrected for body size [PLAX/BSA] ≥16 to <19 mm/m^2) PSAX RVOT ≥32 to <36 mm (corrected for body size [PSAX/BSA] ≥18 to <21 mm/m^2) *or* fractional area change >33% to ≤40%
By MRI • Regional RV akinesia or dyskinesia or dyssynchronous RV contraction • *and* 1 of the following: RV end-diastolic volume indexed to BSA ≥110 mL/m^2 (male) or ≥100 mL/m^2 (female) *or* RV ejection fraction ≤40%	**By MRI** • Regional RV akinesia or dyskinesia or dyssynchronous RV contraction • *and* 1 of the following: RV end-diastolic volume indexed to BSA ≥100 to <110 mL/m^2 (male) or ≥90 to <100 mL/m^2 (female) *or* RV ejection fraction >40% to ≤45%
By RV angiography • Regional RV akinesia, dyskinesia, or aneurysm	
II. TISSUE CHARACTERIZATION OF WALL	
• Residual myocytes <60% by morphometric analysis (or <50% if estimated), with fibrous replacement of the RV free wall myocardium in ≥1 sample, with or without fatty replacement of tissue on endomyocardial biopsy	• Residual myocytes 60% to 75% by morphometric analysis (or 50% to 65% if estimated), with fibrous replacement of the RV free wall myocardium in ≥1 sample, with or without fatty replacement of tissue on endomyocardial biopsy
III. REPOLARIZATION ABNORMALITIES	
• Inverted T waves in right precordial leads (V_1,V_2, and V_3) or beyond in individuals >14 years of age (in the absence of complete right bundle branch block QRS ≥120 msec)	• Inverted T waves in leads V_1 and V_2 in individuals >14 years of age (in the absence of complete right bundle branch block) or in V_4, V_5, or V_6 • Inverted T waves in leads V_1, V_2, V_3, and V_4 in individuals >14 years
IV. DEPOLARIZATION/CONDUCTION ABNORMALITIES	
• Epsilon wave (reproducible low-amplitude signals between end of QRS complex to onset of the T wave) in the right precordial leads (V_1 to V_3)	• Late potentials by SAECG in ≥1 of 3 parameters in the absence of a QRS duration ≥110 msec on the standard ECG • Filtered QRS duration (fQRS) ≥114 msec • Duration of terminal QRS <40 μV (low-amplitude signal duration) ≥38 msec • Root-mean-square voltage of terminal 40 msec ≤20 μV • Terminal activation duration of QRS ≥55 msec measured to the end of the QRS, including R′, in V_1, V_2, or V_3, in the absence of complete right bundle branch block
V. ARRHYTHMIAS	
• Nonsustained or sustained ventricular tachycardia of left bundle branch morphology with superior axis (negative or indeterminate QRS in leads II, III, and aVF and positive in lead aVL)	• Nonsustained or sustained ventricular tachycardia of RV outflow configuration, left bundle branch block morphology with inferior axis (positive QRS in leads II, III, and aVF and negative in lead aVL) or of unknown axis • >500 ventricular extrasystoles per 24 hours (Holter)
VI. FAMILY HISTORY	
• ARVC confirmed in a first-degree relative who meets current Task Force criteria • ARVC confirmed pathologically at autopsy or surgery in a first-degree relative • Identification of a pathogenic mutation† categorized as associated or probably associated with ARVC in the patient under evaluation	• History of ARVC in a first-degree relative in whom it is not possible or practical to determine whether the family member meets current Task Force criteria • Premature sudden death (<35 years of age) due to suspected ARVC in a first-degree relative

*Hypokinesis is not included in this or subsequent definitions of right ventricular (RV) regional wall motion abnormalities for the proposed modified criteria.
†A pathogenic mutation is a DNA alteration associated with ARVC that alters or is expected to alter the encoded protein, is unobserved or rare in a large non-ARVC control population, and either alters or is predicted to alter the structure or function of the protein or has demonstrated linkage to the disease phenotype in a conclusive pedigree.
ARVC = arrhythmogenic right ventricular cardiomyopathy; aVF = augmented voltage unipolar left foot lead; aVL = augmented voltage unipolar left arm lead; BSA = body surface area; ECG = electrocardiogram; MRI = magnetic resonance imaging; PLAX = parasternal long axis view; PSAX = parasternal short axis view; RVOT = right ventricular outflow tract; SAECG = signal-averaged electrocardiogram.
Diagnostic terminology for original criteria: this diagnosis is fulfilled by the presence of 2 major, 1 major plus 2 minor, or 4 minor criteria from different groups.
Diagnostic terminology for revised criteria: definite diagnosis: 2 major, 1 major and 2 minor, or 4 minor criteria from different categories; borderline: 1 major and 1 minor or 3 minor criteria from different categories; possible: 1 major or 2 minor criteria from different categories.
From Marcus FI, McKenna WJ, Sherrill D, et al. Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: proposed modification of the task force criteria. *Circulation*. 2010;121:1533-1541.

pure restrictive cardiomyopathy. Amyloidosis, hemochromatosis, and sarcoidosis are among the systemic diseases that cause restrictive cardiomyopathy (see later).

CLINICAL MANIFESTATIONS

Most patients present with symptoms and signs of heart failure and arrhythmia. Common symptoms include dyspnea on exertion, recurrent respiratory tract infections, general fatigue, and weakness. Symptoms may progress rapidly to dyspnea at rest, orthopnea, paroxysmal nocturnal dyspnea, and abdominal discomfort due to hepatic engorgement. Many patients complain of chest pain and palpitation. Syncope is a presenting symptom in 10% of children. Rarely, sudden death is the initial manifestation of the disease.

Physical examination typically reveals an elevated jugular venous pressure, which has a prominent *y* descent and fails to fall (or rises) during inspiration (Kussmaul sign). On cardiac auscultation, the pulmonary component of the second heart sound may be loud if pulmonary vascular resistance is high. A third heart sound and occasionally a fourth heart sound commonly produce a gallop rhythm. Peripheral edema, ascites, and hepatomegaly are common.

DIAGNOSIS

The most frequent ECG abnormalities include P mitrale and P pulmonale, nonspecific ST segment and T wave abnormalities, ST segment depression, and T wave inversion, usually in the inferolateral leads. Voltage criteria for left and right ventricular hypertrophy may be present, although patients with

TABLE 60-7 ARRHYTHMOGENIC RIGHT VENTRICULAR CARDIOMYOPATHY: CRITERIA FOR DIAGNOSIS OF FIRST-DEGREE RELATIVES WHO DO NOT FULFILL CRITERIA AS PROBANDS*

ARVC in a first-degree relative plus one of the following:	
ECG	T wave inversion in right precordial leads (V_2 and V_3)
Signal-averaged ECG	Late potentials seen on signal-averaged ECG
Arrhythmia	Left bundle branch block–type ventricular tachycardia on ECG, on Holter monitoring, or during exercise testing; >200 extrasystoles during a 24-hour period
Structural or functional abnormality of the right ventricle	Mild global right ventricular dilation or reduction in ejection fraction with normal left ventricle; mild segmental dilation of the right ventricle; regional right ventricular hypokinesia

*Any one criterion is adequate for the diagnosis.
ARVC = arrhythmogenic right ventricular cardiomyopathy; ECG = electrocardiogram.
From Hamid MS, Norman M, Quraishi A, et al: Prospective evaluation of relatives for familial arrhythmogenic right ventricular cardiomyopathy reveals a need to broaden diagnostic criteria. *J Am Coll Cardiol*. 2002;40:1445-1450.

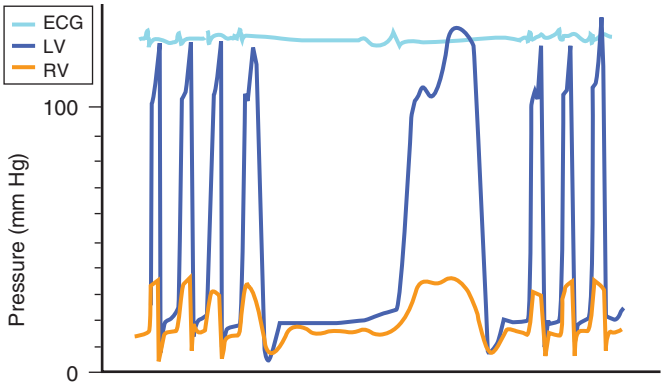

FIGURE 60-6. Idiopathic restrictive cardiomyopathy. Right ventricular (RV) and left ventricular (LV) pressure electrocardiographic (ECG) tracings in a patient with idiopathic restrictive cardiomyopathy. A dip-and-plateau pattern is seen in both ventricles, and diastolic filling pressures are elevated. The plateaus occur at different pressures, approximately 16 mm Hg for the RV tracing compared with 20 mm Hg for the LV tracing. The diagnosis of restrictive disease was confirmed by thoracotomy. (Redrawn from Benofti JR, Grossman W, Cohn PF. The clinical profile of restrictive cardiomyopathy. *Circulation*. 1980;61:1206.)

TABLE 60-8 CAUSES OF RESTRICTIVE CARDIOMYOPATHIES

INFILTRATIVE DISORDERS
Amyloidosis
Sarcoidosis
STORAGE DISORDERS
Hemochromatosis
Fabry disease
Glycogen storage diseases
FIBROTIC DISORDERS
Radiation
Scleroderma
Drugs (e.g., doxorubicin, serotonin, ergotamine)
METABOLIC DISORDERS
Carnitine deficiency
Defects in fatty acid metabolism
ENDOMYOCARDIAL DISORDERS
Endomyocardial fibrosis
Hypereosinophilic syndrome (Löffler endocarditis)
MISCELLANEOUS CAUSES
Carcinoid syndrome

amyloidosis have low-voltage QRS complexes. Conduction abnormalities include intraventricular conduction delay and abnormal Q waves.

On cardiac imaging, both atria are markedly dilated and can dwarf the size of the ventricles in patients with normal global systolic function and a nonhypertrophied, nondilated left ventricle. Pulsed-wave Doppler velocities typically show increased early diastolic filling velocity, decreased atrial filling velocity, increased ratio of early diastolic filling to atrial filling, decreased E wave deceleration time, and decreased isovolumic relaxation time. Pulmonary vein and hepatic vein pulsed-wave Doppler velocities demonstrate higher diastolic than systolic velocities, increased atrial reversal velocities, and atrial reversal duration greater than mitral atrial filling duration. Tissue Doppler imaging usually shows reduced diastolic annular velocities and an increased ratio of early diastolic tissue Doppler annular velocity to mitral early diastolic filling velocity, reflecting elevated left ventricular end-diastolic pressures.

The characteristic hemodynamic feature on cardiac catheterization is a deep and rapid early decline in ventricular pressure at the onset of diastole, with a rapid rise to a plateau in early diastole ("dip-and-plateau" or "square root sign") (Fig. 60-6). Left ventricular end-diastolic, left atrial, and pulmonary capillary wedge pressures are markedly elevated, usually 5 mm Hg or more above right atrial and right ventricular end-diastolic pressures. Volume loading and exercise accentuate the difference between left-sided and right-sided pressures.

The diagnostic evaluation aims to exclude potentially reversible conditions. In such cases, the cardiac manifestations may provide the clues, but definitive diagnosis relies on the demonstration of disease-specific features, such as amyloid protein in amyloidosis (Chapter 188), noncaseating granulomas in sarcoidosis (Chapter 95), abnormal iron studies in hemochromatosis (Chapter 212), or reduced α-galactosidase A levels in Fabry disease (Chapter 208). Endomyocardial biopsy is rarely required to make these diagnoses.

TREATMENT

Rx

Diuretics are the main therapy for heart failure symptoms (Chapter 59), but they must be carefully administered so as not to reduce left ventricular filling pressures to the point of hypotension. Angiotensin-converting enzyme inhibitors and β-blockers are commonly recommended despite few data on their benefit. In patients with secondary restrictive cardiomyopathies, specific treatment of the underlying systemic disease is often appropriate (see later). Referral for transplant assessment should be considered early because pulmonary hypertension may develop and necessitate heart and lung transplantation.

PROGNOSIS

In adults with restrictive cardiomyopathy, the clinical course is usually slow and protracted. Survival from the time of diagnosis is often 10 years or more, except for AL amyloidosis, which progresses much more rapidly. Symptoms of heart failure are generally progressive and respond poorly to treatments for heart failure.

Specific Clinical Syndromes

SARCOIDOSIS

The frequency of myocardial involvement in patients with sarcoidosis (Chapter 95) is difficult to determine because it is frequently subclinical and patchy in nature. Postmortem studies suggest that the heart is involved in at least 25% of patients, but clinical cardiac involvement occurs in less than 10% of patients. Clinical manifestations of sarcoid include heart failure, conduction abnormalities, atrial and ventricular arrhythmias, pericardial effusion, valvular dysfunction, and, rarely, sudden cardiac death.[11] Right-sided heart failure secondary to pulmonary hypertension may occur in patients with extensive fibrotic lung disease. Myocardial infiltration by sarcoid granulomas results in restrictive or dilated cardiomyopathy. The most common site is in the lateral wall of the left ventricle. Papillary muscle involvement is responsible for the most common valvulopathy, mitral regurgitation. Granuloma formation in the basal interventricular septum may cause conduction abnormalities. Ventricular arrhythmias are also frequent. Biopsy of extracardiac sites is usually adequate for the diagnosis, but imaging with a gallium scan, T2 magnetic resonance imaging, or positron emission tomography/computed tomography often demonstrates cardiac inflammation. A myocardial biopsy may show granulomas but, because of the focal distribution of the lesions, may be nondiagnostic. Corticosteroid therapy may improve arrhythmias, but heart failure may worsen despite such therapy. An ICD is generally indicated for ventricular arrhythmias.

AMYLOIDOSIS

EPIDEMIOLOGY AND PATHOBIOLOGY

Amyloidosis can result in deposition of amyloid protein in the atria, ventricles, coronary vessels, conduction system, and valves. The degree of cardiac involvement varies among subtypes.[12] Hematologic disorders (Chapter 187) associated with excessive light chain (AL) immunoglobulin production are the most common cause of cardiac amyloid. Familial forms caused by the accumulation of mutant proteins (transthyretin or apolipoprotein A) (Chapter 188) have variable cardiac involvement. Secondary amyloidosis, due to deposition of serum amyloid A protein in chronic inflammatory diseases, rarely affects the heart. In senile systemic amyloidosis, cardiomyopathy is caused by deposition of normal wild-type transthyretin; this disease nearly always affects elderly persons (>70 years), with a clinical course that is considerably slower than with other types of amyloid.

DIAGNOSIS

The ECG tracing in most forms of cardiac amyloid characteristically shows decreased voltage despite increased wall thickness on echocardiography. Characteristic two-dimensional echocardiographic findings in advanced cardiac amyloidosis are biventricular hypertrophy, thickened valves and interatrial septum, dilated atria, and a small pericardial effusion. The myocardium has a hyperreflective granular texture (Fig. 60-7), best seen on digital image analysis. Echo Doppler in advanced disease demonstrates a restrictive left ventricular filling pattern. Cardiac magnetic resonance imaging may show subendocardial late gadolinium enhancement with abnormal gadolinium kinetics (Fig. 60-8). Nuclear scans with ^{123}I-labeled serum amyloid P component are highly specific. In hereditary transthyretin-related amyloidosis, abnormalities usually can be detected on ^{99m}Tc-DPD scintigraphy before the appearance of echocardiographic changes.

A definitive diagnosis of amyloidosis requires a tissue biopsy specimen, which can be obtained from other sites. For example, fine-needle aspiration of abdominal fat is positive for amyloid deposits in more than 70% of patients with AL amyloidosis. If the result is negative, endomyocardial biopsy has a very high sensitivity.

TREATMENT AND PROGNOSIS

Rx

Specific therapies to impede precursor protein production and fibril formation should be implemented whenever possible (Chapter 188). Diuretics, often in high doses (e.g., furosemide, 40 to 80 mg daily), are the mainstay of the palliative heart failure regimen. Angiotensin-converting enzyme or angiotensin II inhibitors should be used cautiously because they are often poorly tolerated and of unproven efficacy in cardiac amyloid. Aldosterone inhibitors might be helpful in advanced cases. Patients may be hypersensitive to digoxin because of enhanced drug binding with amyloid fibrils. Patients with atrial fibrillation in AL amyloidosis should receive anticoagulation with warfarin (Chapter 38) because of a high rate of thromboembolism. Cardiac transplantation remains controversial, but heart transplantation (Chapter 82) with high-dose chemotherapy and with stem cell transplantation (Chapter 178) has been used in patients with AL amyloidosis.

Patients with amyloidosis with heart failure have a median survival time of less than 1 year and a 5-year survival rate of less than 5%. Most deaths occur suddenly. Patients with familial amyloidosis have a slower course than that of patients with a monoclonal gammopathy.

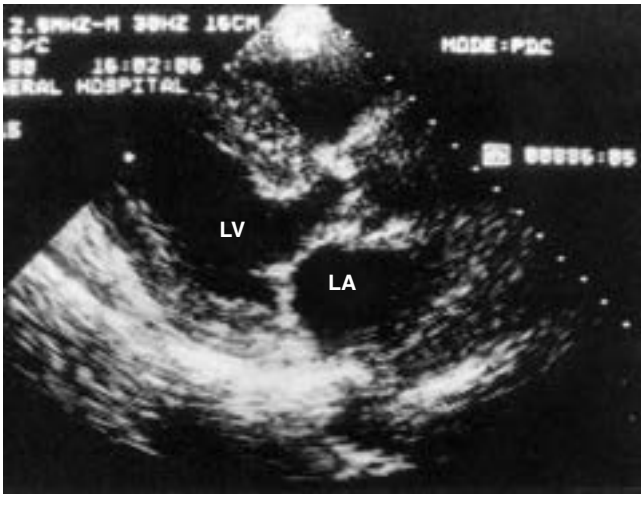

FIGURE 60-7. Amyloidosis. An apical four-chamber echocardiographic image demonstrates biventricular hypertrophy in a patient with biopsy-proven amyloidosis. LA = left atrium; LV = left ventricle. (From Levine RA. Echocardiographic assessment of the cardiomyopathies. In: Weyman AE, ed. Principles and Practice of Echocardiography. 2nd ed. Philadelphia: Lea & Febiger; 1994:810.)

HEREDITARY HEMOCHROMATOSIS

Hereditary hemochromatosis (Chapter 212) is an autosomal recessive disorder caused by excessive iron deposition in various organs, including the liver, spleen, pancreas, endocrine glands, and heart. In whites, its prevalence is between 1 in 200 and 1 in 500, with an even higher prevalence in the Irish population. The most common form is caused by mutations in the *HFE* gene, with two missense mutations accounting for most cases (C282Y and H63D).

Most patients with classic disease present between the ages of 40 and 60 years with hyperpigmentation, diabetes mellitus, and hepatomegaly. Up to 35% of patients with hemochromatosis experience heart failure, and 36% develop arrhythmias. Restrictive physiologic features dominate early in the disease, followed by ventricular dilation. The diagnosis is generally made from the clinical picture, an elevated serum iron level, and a high transferrin saturation. Genetic testing is helpful, and the diagnosis can be confirmed by endomyocardial biopsy. Phlebotomy and iron chelation therapy with deferoxamine (Chapter 212) may improve cardiac function before cell injury becomes irreversible. Standard heart failure treatment (Chapter 59) is generally recommended. Death from hemochromatosis results more often from cirrhosis and liver carcinoma than from cardiac disease.

Unclassified Cardiomyopathies

LEFT VENTRICULAR NONCOMPACTION

Failure of the trabecular or spongiform layer of the myocardium to compact may occur with congenital heart disease, including atrial and ventricular septal defects and coarctation of the aorta (Chapter 69), and with the rare X-linked multisystem disorder Barth syndrome.[13] With recent improvements in imaging technology, it has also been recognized in patients with hypertrophic and dilated cardiomyopathy. The prevalence of localized areas of noncompaction is unknown, but clinically significant isolated left ventricular noncompaction in the absence of other cardiac abnormalities is uncommon.

Areas of noncompacted myocardium may be best delineated from normal myocardium by the demonstration of flow within the myocardium by Doppler or contrast echocardiography or cardiac magnetic resonance imaging (Fig. 60-9). When extensive areas are involved, systolic performance may be impaired, and there is a risk of ventricular arrhythmias and systemic emboli. Treatment, when necessary, is for associated heart failure (Chapter 59), arrhythmias (Chapters 64 and 65), and the risk of emboli (Chapter 59). Natural history and prognosis are not well established.

TAKOTSUBO CARDIOMYOPATHY

Takotsubo cardiomyopathy is a syndrome of transient apical left ventricular dysfunction that mimics myocardial infarction (Chapter 73).[14] Postulated mechanisms include coronary artery spasm, myocarditis, a hyperadrenergic syndrome, and dynamic mid-cavity obstruction.

The clinical syndrome classically includes chest pain, ST segment elevation, and raised cardiac biomarkers in association with emotional or physical stress. Coronary arteriography reveals normal epicardial vessels. Conservative treatment with rehydration and removal of the determinants of stress usually results in rapid resolution within hours of the symptoms, ECG changes, and wall motion abnormalities. Of the approximately 12,000 patients who develop takotsubo cardiomyopathy each year in the United States, in-hospital mortality is 4.2%, mostly in people who have another underlying critical illness.

DISEASES OF THE ENDOCARDIUM

Endocardial fibrosis, fibroelastosis, and thrombosis are subclassified into endomyocardial diseases with hypereosinophilia (hypereosinophilic syndromes) and endomyocardial disease without hypereosinophilia (e.g., endomyocardial fibrosis) (see Table 60-8).

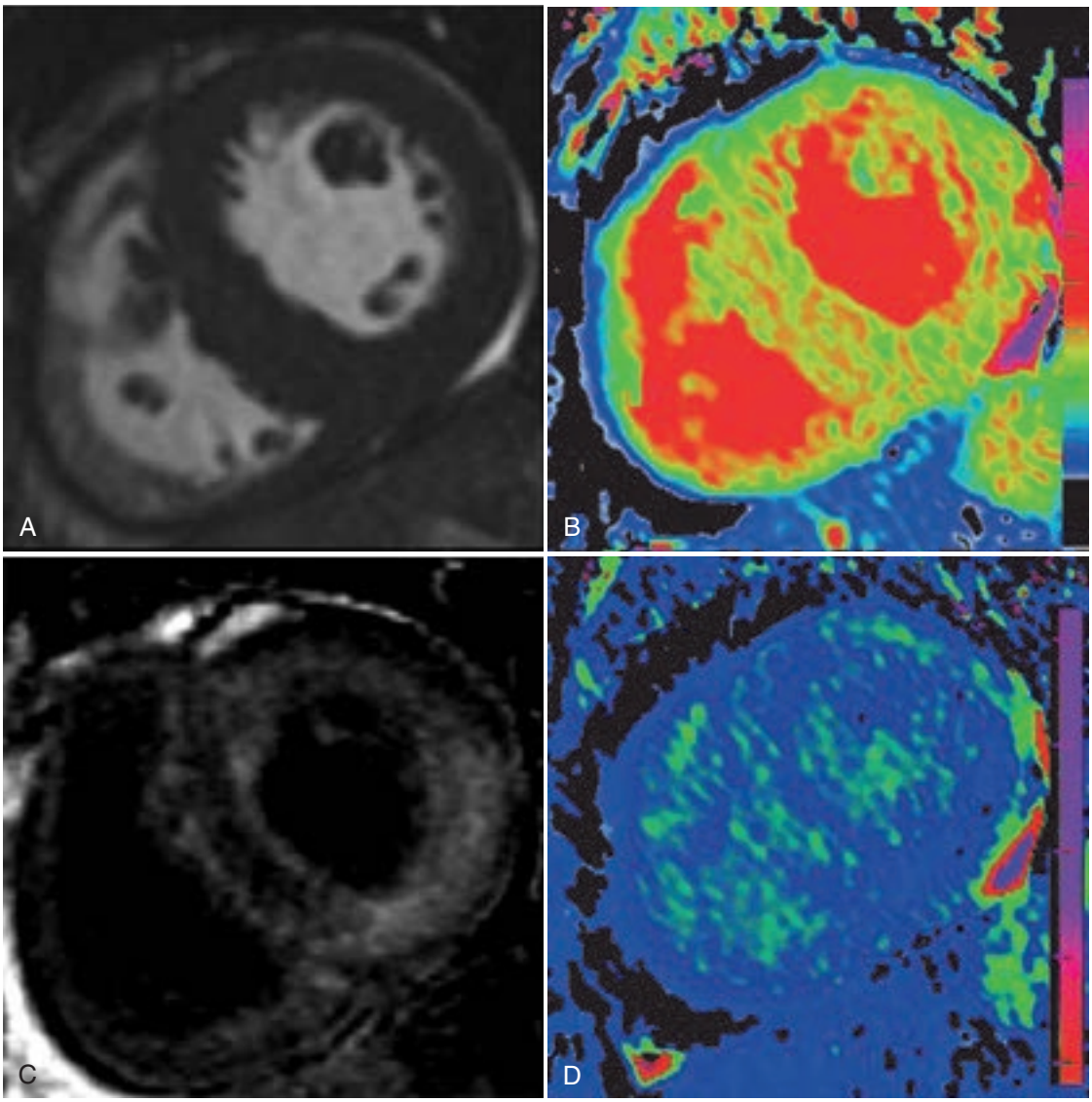

FIGURE 60-8. A 66-year-old woman with AL amyloidosis and cardiac involvement. Note the concentric hypertrophy in the short axis (**A**) and the presence of red signal in the myocardium suggestive of high "native" T1 (**B**). There is a transmural and circumferential late gadolinium pattern (**C**) that is typical of cardiac amyloidosis. On the postcontrast image (**D**), the dark blue areas inside the myocardium are characterized by presence of gadolinium and lower T1 than in the other parts of the myocardium and the blood (green). These findings suggest high extracellular volume (amyloid fibrils substitution) inside the myocardium.

Hypereosinophilic Syndrome

Hypereosinophilic syndromes are a rare and heterogeneous group of disorders defined as persistent blood eosinophilia ($>1.5 \times 10^9$/L) for more than 6 consecutive months, associated with evidence of eosinophil-induced organ damage in the absence of causes of hypereosinophilia, such as allergic, parasitic, and malignant disorders (Chapter 170). Pathogenic mechanisms include stem cell mutations that lead to expression of PDGFRA-containing fusion genes, mainly the *FIP1L1-PDGFRA* fusion gene, with constitutive tyrosine kinase activity and sustained overproduction of interleukin-5 by activated T-cell subsets. Clinically, hypereosinophilic syndrome can be classified into chronic eosinophilic leukemia, lymphocytic hypereosinophilic syndrome, myeloproliferative hypereosinophilic syndrome, and idiopathic hypereosinophilic syndrome. The term *organ-restricted eosinophilic disease*, such as eosinophilic gastroenteritis, dermatitis, or pneumonia, is used when a specific organ or tissue is the exclusive target of eosinophilic infiltration and damage. The term *Löffler fibroplastic endocarditis* with eosinophilia has been used to describe cardiac damage caused by direct toxicity of circulating eosinophils in patients with persistent hypereosinophilia, but its use is now discouraged.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Hypereosinophilic syndrome is a rare disorder that tends to occur in patients 20 to 50 years of age, but all age groups are affected. Cardiac involvement generally evolves in three phases: an early necrotic stage that involves the endomyocardium, which is usually asymptomatic but can be manifested as acute heart failure; a thrombotic stage, in which thrombi develop on the ventricular endocardium, sometimes causing peripheral emboli; and the final fibrotic stage, endomyocardial fibrosis, which causes restrictive cardiomyopathy and damage to atrioventricular valves. Chest pain, cough, dyspnea or orthopnea, and edema of the lower extremities are typical symptoms. Some patients may develop arrhythmias.

The characteristic two-dimensional echocardiographic findings include endocardial thickening, apical obliteration of one or both ventricles by an echogenic material, hyperdynamic contraction of the spared ventricular walls with bilateral atrial enlargement, and a restrictive pattern on echo Doppler.

TREATMENT

Rx

Patients with the F/P fusion gene chromosomal rearrangement should be treated with the tyrosine kinase inhibitor imatinib (100 mg daily for 1 week, increasing by 100 mg each week to 400 mg as guided by toxicity and hematologic response); the duration of therapy is still under investigation. Because some patients develop severe congestive heart failure within days after initiation of therapy, pretreatment with corticosteroids is recommended by some authorities. For patients without the F/P fusion gene, corticosteroids (median maximal daily dose of prednisone of 40 mg [range, 5 to 60 mg] for a duration of 2 months to 20 years; median maintenance dose of 10 mg daily [range, 1 to 40 mg/day]) are the most common first-line therapy. Steroid-sparing and second-line drugs include hydroxyurea (median maximal daily dose of 1000 mg [range, 500 to 2000 mg], adjusted to response), interferon alfa (median maximal dose of 14 million units per week [range, 3 to 40 million units per week], adjusted to response), and imatinib (as before).

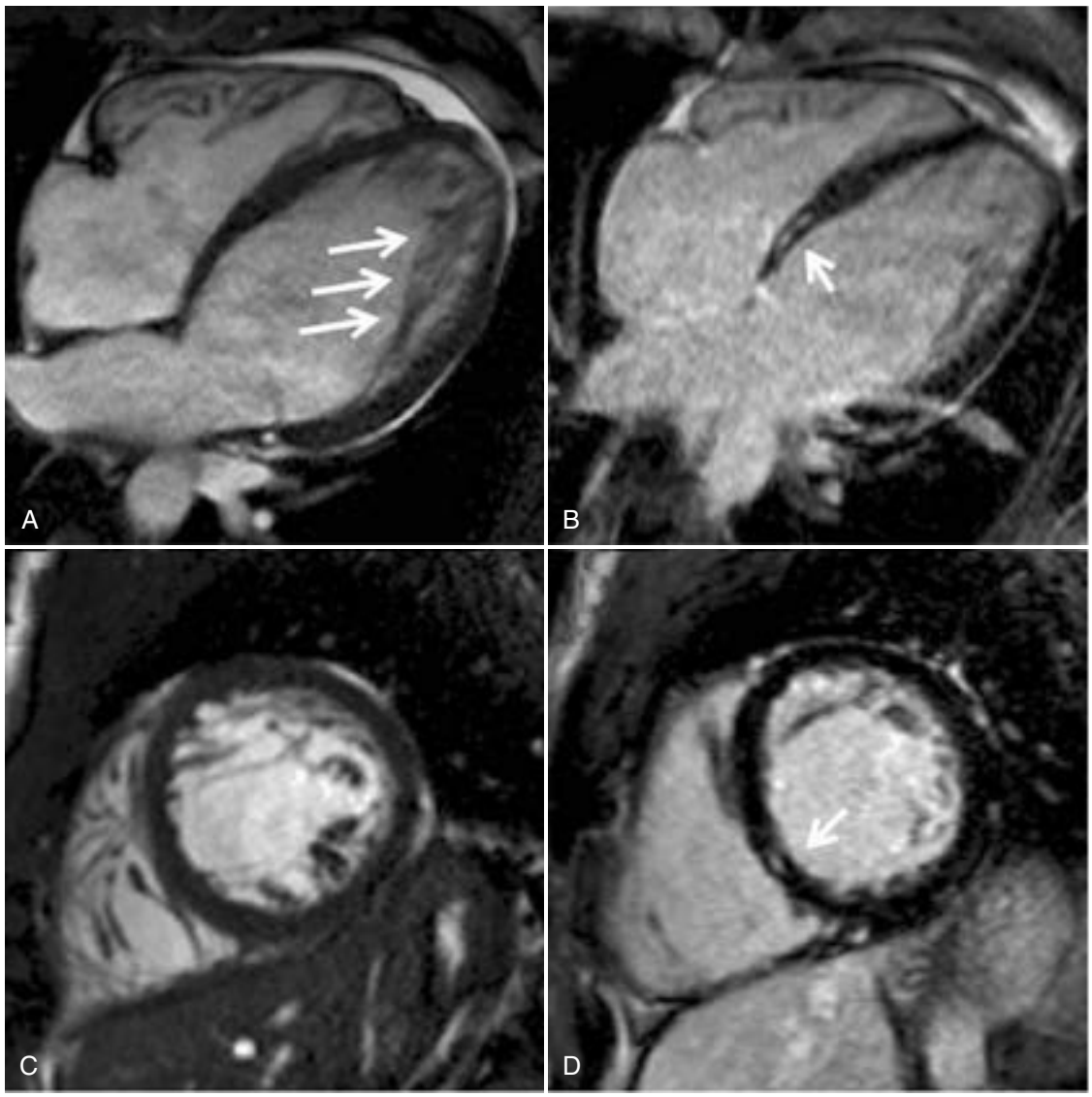

FIGURE 60-9. A 23-year-old white man with left ventricular noncompaction. Four-chamber end-diastolic (**A**) and short axis (**C**) views show left ventricular dilation, prominent trabeculae (*long arrows*), and poorly formed papillary muscles. The ejection fraction was 55%. There is limited mid-myocardial septal late gadolinium enhancement (**B, D,** *short arrows*).

Tropical Endomyocardial Fibrosis

Tropical endomyocardial fibrosis is probably the most common type of restrictive cardiomyopathy worldwide. The disease occurs predominantly within the tropics and affects mostly children and adolescents, usually from low socioeconomic backgrounds. Its cause is unknown, but potential contributors include infection, autoimmunity, genetic predisposition, ethnicity, diet, climate, and poverty.

Severe hypereosinophilia is found in some patients early in the initial stage of the illness; it is characterized by febrile illness, pancarditis, facial and periorbital swelling, pruritus, urticaria, and neurologic symptoms. This phase is followed by ventricular thrombosis that affects the apices and the subvalvular apparatus and then evolves to endocardial fibrosis. The final stage is characterized by restrictive physiology, atrioventricular valve regurgitation, and marked atrial dilation. Death results from complications of chronic heart failure but can occur suddenly from thromboembolism or arrhythmia.

Atrial fibrillation is common at presentation. In advanced disease, the electrocardiogram shows low-voltage QRS complexes, nonspecific ST-T wave changes, and conduction abnormalities. Echocardiography demonstrates apical obliteration, reduction of ventricular cavity size, and tethering or retraction of mitral or tricuspid leaflets or both. There is no specific laboratory test, and hypereosinophilia is present only early in the disease.

There is no specific treatment for endomyocardial fibrosis. Medical treatment is used to control the heart failure (Chapter 59) and arrhythmias (Chapters 64 and 65). Surgical endocardial resection, combined with valve repair or replacement, has an early postoperative mortality between 15 and 30%. The overall prognosis is poor, with a 44% mortality rate at 1 year, increasing to nearly 90% at 3 years.

Carcinoid Syndrome

EPIDEMIOLOGY AND PATHOBIOLOGY

Carcinoid tumors are rare (1 in 100,000) neuroendocrine malignant neoplasms originating mostly from enterochromaffin cells in the gastrointestinal tract (Chapter 232). Carcinoid syndrome, with flushing, diarrhea, and bronchospasm, occurs after tumor cells metastasize to the liver and the vasoactive substances produced by the tumors enter the systemic circulation through the hepatic vein. Carcinoid heart disease occurs in up to 70% of cases of carcinoid syndrome.

The typical cardiac lesion is the carcinoid plaque, which is composed of smooth muscle cells, myofibroblasts, and elastic tissue that forms a fibrous layer on the endocardial surface of the right ventricle and atrium, the valve leaflets, and the subvalvular apparatus, including the chordae and papillary muscles. The tricuspid valve plaques tend to develop on the ventricular side of the leaflets, where they adhere to the mural endocardium and cause valvular regurgitation. On the pulmonary valve, the predominant lesion is stenosis. In patients with a patent foramen ovale, left-sided valvular involvement can occur. Occasional patients may have concomitant myocardial metastases and pericardial effusions from direct tumor invasion.

The most common presentation is dyspnea with signs and symptoms of right-sided heart failure. The electrocardiogram and radiograph are nonspecific. Echocardiography shows thickening of the tricuspid valve, the subvalvular apparatus, and the pulmonary valve. In severe disease, the tricuspid

leaflets are retracted and fixed, with loss of normal coaptation. Similar findings can be seen on cardiac magnetic resonance imaging.

TREATMENT AND PROGNOSIS

Rx

Treatment of the underlying carcinoid with a somatostatin analogue can improve systemic symptoms (Chapter 232). Valve replacement now has an operative mortality of less than 10% (Chapter 232). Without treatment, patients with carcinoid heart disease have a mean life expectancy of 1.6 years. In one series, cardiac surgery for valve disease was associated with about a 50% risk reduction.

Nonbacterial Thrombotic (Marantic) Endocarditis

EPIDEMIOLOGY AND PATHOBIOLOGY

Platelet-fiber masses that are adherent to the mitral or aortic valves are seen in about 20% of patients with malignant tumors, especially mucin-producing adenocarcinomas, melanomas, leukemias, and lymphomas. The lesions are sterile, commonly verruciform, and without accompanying inflammation.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Nonbacterial thrombotic endocarditis is virtually always asymptomatic but occasionally is a source of systemic emboli. Because of the small size of many of the emboli, the first presentation is often with cerebral symptoms. Larger lesions are detectable by echocardiography, but even transesophageal echocardiography is not sufficiently sensitive to identify lesions that may be found at autopsy and that may have been the source of systemic emboli.

TREATMENT

Rx

No treatment has been proved efficacious. However, systemic anticoagulation similar to that used in patients with tumor-associated deep venous thrombosis is often tried (Chapters 81 and 179).

CARDIAC TUMORS

Myocardial Tumors

Most primary cardiac tumors (Table 60-9) are benign. However, all tumors that extend from other tissues into the heart are malignant, as are metastatic lesions.

TABLE 60-9 CARDIAC TUMORS

PRIMARY
Benign
Myxoma
Lipoma
Fibroma
Rhabdomyoma
Fibroelastoma
Malignant
Sarcoma
Mesothelioma
Lymphoma
SECONDARY
Direct Extension
Lung cancer
Breast cancer
Mediastinal tumors
Metastatic Tumors
Malignant melanoma
Leukemia
Lymphoma
Venous Extension
Renal cell cancer
Adrenal cancer
Liver cancer

EPIDEMIOLOGY AND PATHOBIOLOGY

Primary tumors of the heart are unusual, with a prevalence of 1 in 2000 to 1 in 4000 in autopsy series. Nearly all these primary tumors are benign myxomas, although fibromas, lipomas, and fibroelastomas also occur. Rhabdomyomas are seen in children, especially with tuberous sclerosis (Chapter 417). The rare primary malignant tumors include sarcomas, especially angiosarcomas (see Table 60-9). Rarely, a primary mesothelioma or lymphoma may originate in the heart.

Up to 20% of advanced cancers may involve the pericardium, epicardium, or cardiac chambers either by direct extension of the primary tumor or by metastatic disease. Direct extension occurs principally from cancers of the lung, breast, esophagus, and mediastinum. Extension through the inferior vena cava to the right atrium and even to the right ventricle occurs with cancers of the kidney, adrenal gland, and liver. Metastatic spread is most common with melanomas or lymphomas.

Pericardial Tumors

CLINICAL MANIFESTATIONS

Pericardial tumors almost always result from direct extension of tumors, principally lung and breast, which produce a pericardial effusion that can progress to cardiac tamponade (Chapter 77). Patients typically are asymptomatic or minimally symptomatic in terms of the cardiac involvement until the effusion is large, although they often may be very ill because of progressive tumor elsewhere.

DIAGNOSIS

The diagnosis is often suspected in a patient with advanced malignant disease on the basis of evidence of heart failure, hypertension, or arrhythmia and is confirmed by echocardiography. The differentiation between pericardial involvement by tumor and postradiation pericarditis depends on pericardiocentesis, often guided by echocardiography, and cytologic examination.

TREATMENT

Rx

Cardiac tamponade must be treated with urgent pericardiocentesis, preferably under echocardiographic or radiologic guidance (Chapter 77). Although such a procedure can be life-saving and provide short-term to intermediate-term palliation, control of the effusion often requires prolonged drainage, administration of intrapericardial chemotherapeutic agents, or limited or full pericardiectomy (Chapter 77). Some patients with pericardial tumors may respond to aggressive systemic chemotherapy, but recurrent accumulation of fluid is sufficiently likely that creation of a pericardial window should be considered before hospital discharge.

PROGNOSIS

In many cases, a tumor that is causing pericarditis has extended or will eventually extend through the pericardial space and into the myocardium, so no therapy is likely to be successful. The prognosis is very poor, except in unusual cases in which the tumor responds dramatically to systemic therapy.

Intracavitary Tumors

MYXOMA

DEFINITION AND EPIDEMIOLOGY

A myxoma is a benign polypoid neoplasm that originates from endocardial cells and is attached to the interatrial septum, usually protruding into the left atrium but occasionally into the right atrium and rarely into the ventricles. Myxomas are more common in women, especially between the ages of 30 and 60 years, than in men. These tumors can be familial and are rarely associated with other systemic abnormalities.

CLINICAL MANIFESTATIONS

Myxomas are slow growing and usually do not produce symptoms or signs until they enlarge. The typical presentation is with a tumor embolus, whereby usually small portions of the myxoma break loose and cause a single embolism or a shower of emboli.[15] However, a large embolism from a myxoma can be of sufficient size to obstruct a medium-sized artery. Some patients have systemic symptoms, including fever, malaise, and arthralgias, as part of a clinical syndrome that may be confused with bacterial endocarditis (Chapter 76) or a collagen vascular disease. Large myxomas can prolapse into the mitral valve orifice during diastole, or they may obstruct blood flow from the left atrium to the left ventricle and mimic rheumatic mitral stenosis.

DIAGNOSIS

A myxoma large enough to obstruct the mitral orifice can produce an audible "tumor plop" when the myxoma prolapses and obstructs blood flow during diastole, at the same time that the opening snap of mitral stenosis would typically be heard. If obstruction is incomplete, the tumor plop may be followed by a diastolic rumble. As obstruction becomes more severe, cardiac output may fall precipitously. Echocardiography (Chapter 55) is usually definitive; transesophageal echocardiography provides a higher sensitivity than does transthoracic echocardiography, and magnetic resonance imaging can be helpful.

TREATMENT

Rx

Surgical removal is generally curative, although myxomas can be multiple or recur in about 5% of cases. Follow-up postoperative echocardiography is generally recommended. However, the optimal frequency and duration for follow-up screening are uncertain.

OTHER PRIMARY INTRACAVITARY TUMORS

Papillary fibroelastomas are rare, typically frondlike tumors that may arise from a cardiac valve, often the mitral valve, and are generally detected incidentally by echocardiography. However, like myxomas, they can be manifested with systemic or even coronary emboli. Surgical excision is usually successful.

Angiosarcomas, which are more frequent in men than in women, typically involve the pericardium and right atrium. They cause obstruction with clinical signs and symptoms of right-sided heart failure. These sarcomas are generally not amenable to therapy.

EXTENSION OF TUMOR INTO THE CARDIAC CAVITIES

Direct extension of tumor up the inferior vena cava into the right atrium can be seen with renal cell carcinomas and less commonly with liver and adrenal cancers. In some cases, tumor extension is accompanied by adherent clot, and either the tumor or the clot may cause obstruction or pulmonary emboli (Chapter 98). No treatments are generally successful, and the prognosis is grim.

Intramyocardial Tumors

Benign tumors in the myocardium include lipomas, fibromas, and rhabdomyomas. Primary malignant tumors include sarcomas, lymphomas, and mesotheliomas. Metastatic tumors include melanomas, lymphomas, and leukemias. The tumors may be clinically silent, or they may produce arrhythmias or even impinge on coronary arteries, thereby causing ischemic syndromes. Large tumors may protrude into the cardiac chamber and cause obstruction. Therapies are not successful, except for occasional patients whose metastatic tumors may respond to systemic chemotherapy or whose primary tumors have been cured by heart transplantation.

Grade A References

A1. Mason JW, O'Connell JB, Herskowitz A, et al. A clinical trial of immunosuppressive therapy for myocarditis: the Myocarditis Treatment Trial Investigators. *N Engl J Med*. 1995;333:269-275.

A2. Wojnicz R, Nowalany-Kozielska E, Wojciechowska C, et al. Randomized, placebo-controlled study for immunosuppressive treatment of inflammatory dilated cardiomyopathy: two-year follow-up results. *Circulation*. 2001;104:39-45.

A3. Cooper LT Jr, Hare JM, Tazelaar HD, et al. Usefulness of immunosuppression for giant cell myocarditis. *Am J Cardiol*. 2008;102:1535-1539.

A4. Kadish A, Dyer A, Daubert JP, et al. Prophylactic defibrillator implantation in patients with nonischemic dilated cardiomyopathy. *N Engl J Med*. 2004;350:2151-2158.

A5. Bosch X, Rovira M, Sitges M, et al. Enalapril and carvedilol for preventing chemotherapy-induced left ventricular systolic dysfunction in patients with malignant hemopathies: the OVERCOME trial (preventiOn of left Ventricular dysfunction with Enalapril and caRvedilol in patients submitted to intensive ChemOtherapy for the treatment of Malignant hEmopathies). *J Am Coll Cardiol*. 2013;61:2355-2362.

A6. Sliwa K, Blauwet L, Tibazarwa K, et al. Evaluation of bromocriptine in the treatment of acute severe peripartum cardiomyopathy: a proof-of-concept pilot study. *Circulation*. 2010;121:1465-1473.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Watkins H, Ashrafian H, Redwood C. Inherited cardiomyopathies. *N Engl J Med*. 2011;364:1643-1656.
2. Maron BJ, Maron MS. Hypertrophic cardiomyopathy. *Lancet*. 2013;381:242-255.
3. Gersh BJ, Maron BJ, Bonow RO, et al. 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2011;58:2703-2736.
4. Elliott PM, Anastasakis A, Borger MA, et al. 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: The Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). *Eur Heart J*. 2014;35:2733-2779.
5. O'Mahony C, Jichi F, Pavlou M, et al. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). *Eur Heart J*. 2014;35:2010-2020.
6. Caforio AL, Pankuweit S, Arbustini E, et al. Current state of knowledge on aetiology, diagnosis, management, and therapy of myocarditis: a position statement of the European Society of Cardiology Working Group on Myocardial and Pericardial Diseases. *Eur Heart J*. 2013;34:2636-2648, 2648a-2648d.
7. Kandolin R, Lehtonen J, Salmenkivi K, et al. Diagnosis, treatment, and outcome of giant-cell myocarditis in the era of combined immunosuppression. *Circ Heart Fail*. 2013;6:15-22.
8. Maron BJ, Ommen SR, Semsarian C, et al. Hypertrophic cardiomyopathy: present and future, with translation into contemporary cardiovascular medicine. *J Am Coll Cardiol*. 2014;64:83-99.
9. Gulati A, Jabbour A, Ismail TF, et al. Association of fibrosis with mortality and sudden cardiac death in patients with nonischemic dilated cardiomyopathy. *JAMA*. 2013;309:896-908.
10. te Riele ASJM, Bhonsale A, James CA, et al. Incremental value of cardiac magnetic resonance imaging in arrhythmic risk stratification of arrhythmogenic right ventricular dysplasia/cardiomyopathy-associated desmosomal mutation carriers. *J Am Coll Cardiol*. 2013;62:1761-1769.
11. Chapelon-Abric C. Cardiac sarcoidosis. *Curr Opin Pulm Med*. 2013;19:493-502.
12. Banypersad SM, Moon JC, Whelan C, et al. Updates in cardiac amyloidosis: a review. *J Am Heart Assoc*. 2012;1:e000364.
13. Thavendiranathan P, Dahiya A, Phelan D, et al. Isolated left ventricular non-compaction controversies in diagnostic criteria, adverse outcomes and management. *Heart*. 2013;99:681-689.
14. Sharkey SW. Takotsubo cardiomyopathy: natural history. *Heart Fail Clin*. 2013;9:123-136.
15. Gošev I, Paic F, Duric Z, et al. Cardiac myxoma the great imitators: comprehensive histopathological and molecular approach. *Int J Cardiol*. 2013;164:7-20.

REVIEW QUESTIONS

1. A 35-year-old man without vascular disease risk factors is found to have an abnormality on the electrocardiogram (ECG) with left axis deviation (−45 degrees) and a prolonged PR interval (260 msec) in association with mild left ventricular dilation and a mildly reduced ejection fraction (56%). Physical examination is unremarkable. A diagnosis of dilated cardiomyopathy with conduction disease is made. Which of the following is the most appropriate recommendation?

A. Orchestrate a three-generation family history focusing on premature cardiac disease.
B. Perform myocardial perfusion studies to determine the need for coronary angiography.
C. Perform computed tomography or invasive coronary angiography to exclude significant coronary artery disease.
D. Initiate and up-titrate treatment with a β-blocker and an angiotensin-converting enzyme inhibitor.
E. Refer to a geneticist for mutation analysis.

Answer: A A clinical presentation with unexplained conduction disease and left ventricular dysfunction usually results in a diagnosis of dilated cardiomyopathy (in Europe) or nonischemic cardiomyopathy (in the United States). When conduction disease is part of the early presentation of a dilated or nonischemic cardiomyopathy, it is a "red flag" for an arrhythmic form of dilated cardiomyopathy. A family history focusing on relatives who required pacemakers, had ventricular arrhythmias, died suddenly, or had associated phenotypes (e.g., skeletal myopathy, retinitis pigmentosa, sensory neural deafness) may provide clues to the ultimate diagnosis, requirements for genetic testing, and potential need for an implantable cardioverter-defibrillator. (Rapezzi C, Arbustini E, Caforio AL, et al. Diagnostic work-up in cardiomyopathies: bridging the gap between clinical phenotypes and final diagnosis. A position statement from the ESC Working Group on Myocardial and Pericardial Diseases. *Eur Heart J.* 2013;34:1448-1458.)

2. An asymptomatic 70-year-old man is found to have an abnormality on the 12-lead ECG during a preoperative clinical evaluation. A two-dimensional echocardiogram reveals 2-cm asymmetrical septal hypertrophy without features of a left ventricular outflow tract gradient. Familial evaluation for hypertrophic cardiomyopathy reveals that his 43-year-old son has an abnormality on the ECG but normal findings on echocardiography, whereas an 18-year-old asymptomatic grandson has 2.2-cm asymmetric septal hypertrophy without left ventricular outflow tract obstruction. A 24-hour ECG reveals three beats of nonsustained ventricular tachycardia at 130 beats per minute. The estimated absolute 5-year risk for sudden cardiac death in the grandfather is 3.4%; in the son, 2.2%; and in the grandson, 5.2%. Which of the following recommendations is most appropriate?

A. The grandson should receive an implantable cardioverter-defibrillator.
B. Mutation analysis should be performed to enable genetic cascade screening of the family.
C. Septal reduction therapies should be considered if the grandfather's exercise capacity is less than 75% of predicted during objective testing.
D. Both the patient and his grandson should receive an implantable cardioverter-defibrillator.
E. No interventions are warranted in the patient, his son, or his grandson, but serial evaluations should continue to assess risk of atrial fibrillation, emboli, and serious ventricular arrhythmia with a view to early introduction of prophylactic therapy.

Answer: A Nonsustained ventricular tachycardia recorded during ambulatory ECG monitoring is usually asymptomatic, brief, and at rates of 140 to 160 beats per minute. It is, however, associated with increased risk of sudden death, particularly in the young, in whom it carries a relative risk of approximately 5 and an absolute 5-year sudden cardiac death event rate of 5 to 6% in a 20-year-old compared with a relative risk of 2 or 3 and absolute 5-year risk of 4% in a 70-year-old. Ultimately, the decision to implant an implantable cardioverter-defibrillator is individualized on the basis of the physician's assessment of the level of risk and the level of risk acceptable to the patient. The estimated events risks quoted here are excessive in an adolescent but only mildly increased in someone in the seventh to eighth decade. (O'Mahony C, Lambiase PD, Quarta G, et al. The long-term survival and the risks and benefits of implantable cardioverter defibrillators in patients with hypertrophic cardiomyopathy. *Heart.* 2012;98:116-125. O'Mahony C, Tome-Esteban M, Lambiase PD, et al. A validation study of the 2003 American College of Cardiology/European Society of Cardiology and 2011 American College of Cardiology Foundation/American Heart Association risk stratification and treatment algorithms for sudden cardiac death in patients with hypertrophic cardiomyopathy. *Heart.* 2013;99:534-541. O'Mahony C, Jichi F, Pavlou M, et al. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy [HCM Risk-SCD]. *Eur Heart J.* 2014;35:2010-2020.)

3. An 18-year-old African American female marathon runner presents with postexertional syncope. Her 12-lead ECG tracing is abnormal, with a biphasic T wave in leads V_2 to V_4. Family history reveals a maternal nephew who died suddenly at the age of 23 years while dancing in a discotheque. Her two-dimensional echocardiogram is unremarkable. Which of the following is incorrect?

A. An implantable cardioverter-defibrillator should be recommended, given the family history and the postexertional syncopal episode.
B. An abnormality on the ECG is often the earliest manifestation of hypertrophic cardiomyopathy, arrhythmogenic right ventricular cardiomyopathy, and cardiac sarcoidosis.
C. Imaging with two-dimensional echocardiography or cardiac magnetic resonance imaging provides the definitive diagnosis for hypertrophic cardiomyopathy.
D. Familial evaluation is paramount to determine a diagnosis and potential risk to the patient.
E. A biphasic T wave in V_2 and V_3 is a recognized normal variant in African American athletes.

Answer: A Electrocardiographic abnormalities in heart muscle disease are rarely diagnostic but are often the earliest manifestation of disease and warrant further diagnostic evaluation. Inverted or biphasic T waves in V_1 to V_3 are seen in African American athletes, but changes that extend to V_4 or beyond are usually associated with myocardial disease. New-onset conduction disease and ventricular arrhythmia are common presenting features of Chagas disease and cardiac sarcoidosis, whereas conduction disease is rarely seen early in arrhythmogenic right ventricular cardiomyopathy or hypertrophic cardiomyopathy. Demonstration of left ventricular hypertrophy with echocardiographic or cardiac magnetic resonance imaging is usually diagnostic of hypertrophic cardiomyopathy, but concern should be raised about the accuracy of the measurements when the ECG recording is normal. Imaging abnormalities often provide the definitive diagnosis but are late manifestations of arrhythmogenic right ventricular cardiomyopathy because they reflect myocyte cell death with replacement scar. The onset of dilated cardiomyopathy in adults is usually insidious, with slow progression of left ventricular dilation and impaired contraction. The implantable cardioverter-defibrillator may be life-saving, but it is associated with lifelong consequences and should not be recommended without a diagnosis and clear estimate of sudden death risk. (Sheikh N, Papadakis M, Ghani S, et al. Comparison of electrocardiographic criteria for the detection of cardiac abnormalities in elite black and white athletes. *Circulation.* 2014;129:1637-1649.)

61

PRINCIPLES OF ELECTROPHYSIOLOGY

GLENN I. FISHMAN

The rhythmic beating of the heart reflects the tightly regulated and exquisitely integrated activity of numerous protein complexes that control the flow of ions across cell membranes, including channels, transporters, exchangers, and gap junction channels.[1] The human heart beats almost 3 billion times during a normal lifespan, and even brief periods of dysfunction may lead to life-threatening consequences. Thus, the failure rate of cardiac rhythmicity is exceptionally low. Nonetheless, inherited syndromes, as well as acquired heart disease, may affect cardiac rhythmicity, and these disorders lead to substantial morbidity and mortality, including sudden cardiac death (Chapter 63). This chapter reviews the molecular, cellular, and organ-level determinants of cardiac rhythmicity and relates these principles to fundamental mechanisms responsible for clinically important arrhythmias.

BASIC CONCEPTS

The function of the heart as a highly dynamic pump is intricately entwined with the tightly regulated electrical activation of its constituent cardiomyocytes. During each cardiac cycle, an electrical impulse known as an *action potential* is spontaneously generated by a relatively small number of pacemaker cells in the sinoatrial node and then propagated to neighboring cardiac myocytes through arrays of intercellular channels known as gap junctions. Subpopulations of myocytes within the heart have unique electrical properties that reflect regional specialization. Myocytes within the sinoatrial and atrioventricular nodes produce spontaneous action potentials that reflect their pace-making function. Cells within the His-Purkinje network are optimized to rapidly deliver excitatory current to the large mass of ventricular myocardium, whereas ventricular myocytes display action potentials optimized to facilitate excitation-contraction coupling, that is, to trigger the release of calcium ions from the sarcoplasmic reticulum and to promote the actomyosin cross-bridge formation that underlies cardiac contraction (Chapter 53). Abnormalities in cardiac electrophysiology, whether the result of congenital syndromes or acquired heart disease, can lead to disturbances in the initiation, propagation, or conduction of electrical impulses and, as a result, to a wide variety of arrhythmic syndromes.

IONIC BASIS OF CARDIAC ELECTROPHYSIOLOGY

The Cardiac Action Potential

The cardiac action potential (Fig. 61-1) is a recording of a cell's membrane potential, V_m, versus time. During each cardiac cycle, ions move back and forth across the cardiomyocyte cell membrane, thereby changing V_m. The cardiac action potential, which reflects the integrated behavior of numerous individual ionic currents, is largely dominated by the movement of Na^+, Ca^{+2}, and K^+ ions. These ions traverse the cell membrane through ion-selective pores formed by assemblies of integral membrane-spanning proteins and accessory proteins. The behavior of these ionic pathways is highly regulated, and permeation of specific ions is influenced by multiple factors, the most prominent of which are changes in membrane potential (i.e., voltage gating), ligand binding, second messengers such as cyclic adenosine monophosphate, and post-translational modification. Channel function and, by extension, action potential behavior are dynamically tuned in response to normal physiologic factors, especially heart rate. However, a number of pathologic stressors influence channel activity, including acquired syndromes that are associated with cardiac hypertrophy and failure, as well as an ever-growing number of congenital diseases. Regardless of the underlying pathology, the effects on action potential behavior may trigger arrhythmic activity.

The cardiac action potential is divided into phases, each reflecting the major ionic movements that take place. In working cardiomyocytes, such as ventricular or atrial myocytes, the *resting membrane potential* during diastole, or phase 4 of the cardiac action potential, is determined by the baseline ionic and charge gradients that exist across the sarcolemmal membrane. These gradients are generated by pumps and transporters, the most important of which is the Na^+, K^+-ATPase. This energy-requiring electrogenic pump, which is the major target of ouabain-like compounds such as digoxin, extrudes three Na^+ ions from the intracellular compartment in exchange for two K^+ ions,

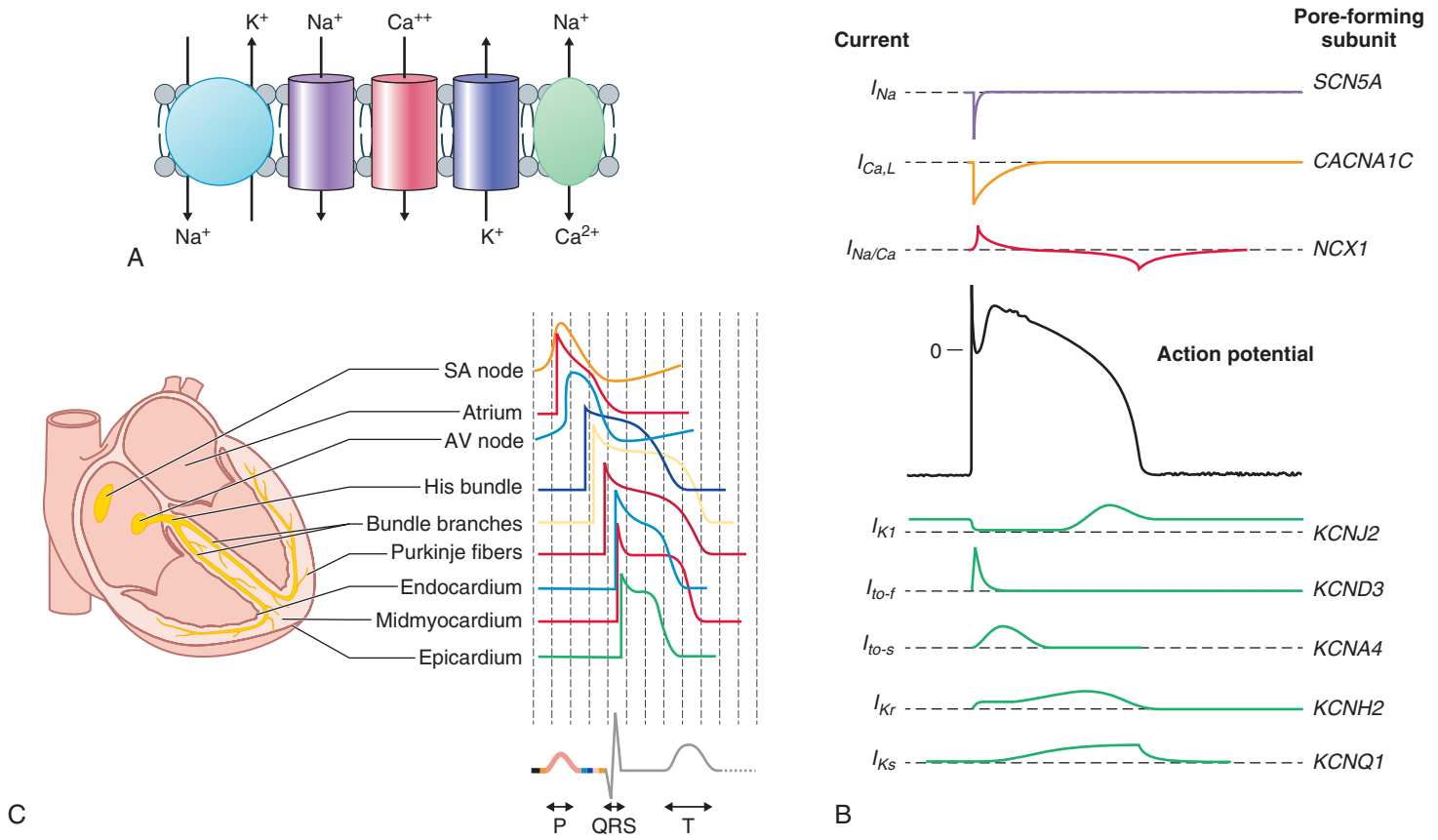

FIGURE 61-1 **Ion channels and the cardiac action potential. A,** Key channels involved in cardiac excitability and generation of the cardiac action potential. Inward currents are carried by Na^+ channels *(purple)* and Ca^{2+} channels *(red)*. Repolarizing currents are primarily carried by K^+ channels *(blue)*. The Na^+, K^+-ATPase is an energy-requiring exchanger that pumps K^+ out of the cell in exchange for Na^+ and is essential for establishing resting ionic gradients and the resting membrane potential. **B,** Time course and relative magnitude of ionic currents active during the cardiac action potential. Inward currents are represented by downward deflections and outward currents by upward deflections. **C,** Action potentials from different regions of the heart and their relationship to the surface electrocardiogram are indicated. AV = atrioventricular. (Adapted from Marbán E. Cardiac channelopathies. *Nature.* 2002;415:213-218.)

thereby resulting in directionally opposite gradients of Na^+ ions (outside > inside) and K^+ ions (inside > outside). Under resting conditions, a subset of membrane channels highly permeable to K^+ is open, but those that allow for the passage of other ions such as Na^+ or Ca^{2+} are only minimally permeable. As a consequence, the concentration gradient promotes the movement of potassium ions from inside to outside of the cell, until the resulting excess of negative charge within the cell balances the diffusional forces and an electrochemical equilibrium is established. The equilibrium potential for a given ion is calculated by the *Nernst equation*, where E_{eq} is the equilibrium potential, R is the universal gas constant, T is the absolute temperature, z is the valence of the ionic species, and F is Faraday constant:

$$E_{eq} = \frac{RT}{zF} ln\left(\frac{[X]_{out}}{[X]_{in}}\right)$$

If the cell membrane were *only* permeable to K^+ ions, at the measured concentrations of intracellular and extracellular K^+, the resting membrane potential would be approximately −100 mV. However, because of the slight but measurable permeability to other ionic species, which have Nerst potentials that are less negative than that for K^+, the actual resting membrane potential in a typical ventricular cardiac myocyte is closer to −85 mV.

When the cardiac cell is depolarized to its excitatory threshold, an action potential is triggered through a series of highly regulated time-dependent changes in ionic conductances (see Fig. 61-1B). The fast sodium current is activated and very rapidly depolarizes the membrane during phase 0 of the action potential. The sodium current is inactivated at the peak of depolarization, which is approximately +40 mV. The increase in V_m during phase 0 activates several additional voltage-gated currents. A transient outward potassium current, or I_{to}, partially *re*polarizes the cell, thereby producing a small notch in the action potential, denoted as phase 1. The increase in V_m during phase 0 also activates, albeit more slowly, the inward L-type calcium current, $I_{Ca\text{-}L}$. It is this trigger for Ca^{2+} that is responsible for Ca^{2+}-induced release from the sarcoplasmic reticulum and is integral to the process of excitation-contraction coupling (Chapter 53). The inward Ca^{2+} current is balanced by several outward repolarizing currents, including the rapid component of the delayed rectifier potassium current I_{Kr}, the slow component of the delayed rectifier potassium current I_{Ks}, and the electrogenic Na^+-Ca^{+2} exchanger, thereby resulting in a plateau in the action potential known as phase 2. When the outward potassium currents increase and the calcium current decreases at the end of phase 2, the action potential progresses to phase 3, which is the phase of rapid repolarization. The inward rectifier potassium current I_{K1} contributes significantly to this final phase of repolarization and brings the action potential back to its resting membrane potential, or phase 4, at which point the cell is primed for another action potential.

Action potential recordings from atrial cardiomyocytes and from cells of the His-Purkinje system are qualitatively similar to those described previously, but with some notable differences that primarily reflect the differential expression of repolarizing potassium currents that tend to abbreviate (in the case of atrial cells) or lengthen (for Purkinje cells) the action potential duration (see Fig. 61-1C).

The relatively small populations of cells in the sinoatrial node (SAN) and atrioventricular node (AVN) express unique complements of ionic currents that are responsible for spontaneous depolarization during phase 4 and the triggering of action potentials. Pacemaker cells express substantially less I_{K1} compared with ventricular myocytes, and as a consequence, their minimum V_m is −65 mV, and they do not repolarize to the same extent as working ventricular cardiomyocytes. In addition, pacemaker cells display what is known as the funny current, I_f, which is activated by *hyper*polarization and carried by sodium. Activation of I_f during phase 4 slowly depolarizes the cell membrane. In addition, a subsarcolemmal calcium clock contributes to diastolic depolarization through the spontaneous, rhythmic release of Ca^{2+} from the sarcoplasmic reticulum, a process that is linked to the voltage clock through the activity of the sodium-calcium exhange current, I_{NCX}. Inasmuch as there is minimal fast inward I_{Na} expressed in nodal cells, the action potential triggered by this spontaneous phase 4 depolarization is due to activation of calcium currents carried by $I_{Ca,L}$ and $I_{Ca,T}$. The magnitude of I_f and I_{Ca}, and hence the slope of phase 4 depolarization as well as the upstroke velocity of the action potential in pacemaker cells, is augmented by adrenergic stimulation, which produces a chronotropic response.

Owing to considerable regional heterogeneity in the density of individual ionic currents, even within distinct compartments of the heart such as the ventricular myocardium, not all ventricular action potentials are identical (see Fig. 61-1C). Much of this heterogeneity is due to differences in the magnitude of various repolarizing K^+ currents. For example, although electrotonic coupling through gap junction channels mitigates this intrinsic heterogeneity, action potentials recorded from epicardial, mid-myocardial, and endocardial cells show substantial differences in morphology, both at rest and especially in response to provocative stimuli such as changes in rate or pharmacologic agents. Moreover, action potential morphology is not static; it varies in response to changes in physiologic state. *Action potential duration adaption*, which reflects the normal shortening of the action potential observed during increased heart rate, provides a mechanism to preserve adequate time for ventricular filling during diastole. Action potential duration shortening in this setting is due to a net increase in repolarizing currents, primarily from increased I_{Ks} and reverse-mode I_{NaCa}. This adaptation is regulated, at least in part, by the kinetics of activation and inactivation of the channels that are responsible for these currents, as well as their modulation by various signaling cascades, such as those regulated by the autonomic nervous system. However, maladaptive regulation of ionic currents is a frequent manifestation of acquired forms of heart disease,[2] and this *pathologic electrical remodeling* may amplify intrinsic heterogeneities in cardiac electrophysiology and form a substrate that promotes arrhythmic behavior.

Impulse Propagation

In the intact heart, action potentials not only must be generated but also must propagate from cell to cell as a wave of excitation throughout the atrial and ventricular myocardium. For successful propagation, the upstream excited cell must provide sufficient charge to bring the membrane potential, V_m, of downstream cells up to their excitation threshold potential. Gap junctions, which comprise arrays of intercellular channels, provide the structural basis for this electrotonic flow of current from cell to cell. For propagation to be successful, the ratio of the charge generated to charge consumed during the excitation cycle, known as the *safety factor*, must be greater than 1.

Unlike nerves, the action potential duration of human cardiomyocytes is quite long, on the order of 200 msec. This longer action potential duration is required so that each myocyte has sufficient time for contraction and relaxation before the next heartbeat. Impulses that arrive too early in the cardiac cycle will not produce normal action potentials. If the impulse occurs during the upstroke or plateau phase (phases 0 to 2), the sodium channels will not have had sufficient time to recover from fast inactivation, and the cell displays *absolute refractoriness*. If the impulse occurs somewhat later, during phase 3 of the action potential, a supranormal stimulus is required to overcome the potassium currents that remain active during the terminal portion of the action potential, a phenomenon known as *relative refractoriness*. Moreover, because not all of the sodium channels will have recovered from inactivation, the rate of rise of voltage during phase 0 of the premature beat may be diminished.

MOLECULAR BASIS OF CARDIAC ELECTROPHYSIOLOGY

The individual currents that are responsible for cardiac excitability reflect the integrated behavior of various protein complexes that are assembled into ion-specific channels, transporters, and exchangers.[3] At the molecular level, ion channels comprise multi-subunit glycoproteins, including a pore-forming major, or α-subunit, and one or several accessory proteins (E-Figure 61-1). The latter influence a range of channel properties including trafficking of the major subunit to the sarcolemmal membrane as well as regulation of channel biophysical properties, that is, the opening and closing of the channel in response to various factors such as membrane voltage, ligands, mechanical stimuli, second messengers, or post-translational modification.

Sodium Channels

Voltage-gated sodium channels are responsible for the activation and propagation of the cardiac action potential. Not surprisingly, acquired and inherited syndromes that affect the function of voltage-gated sodium channels in the heart are responsible for a broad range of arrhythmic phenotypes. Cardiac sodium channels activate extremely rapidly, within 1 msec, and begin to inactivate almost completely within several milliseconds. The very small proportion of channels that remain active for several hundred milliseconds results in the persistent or late Na^+ current, I_{NaL}.

The most abundant cardiac sodium channel comprises a pore-forming α-subunit known as $Na_v1.5$ and several smaller accessory subunits, or β-subunits, designated $Na_vβ1$ to 4. The α-subunit is an approximately 260-kD protein that consists of four homologous domains, each comprising six transmembrane segments. Substantial experimental work has identified the key regions of the protein that regulate channel properties, including voltage dependence, activation, and inactivation, as well as the binding sites for pharmacologic agents such as local anesthetics, antiarrhythmic drugs, and neurotoxins. The β-subunits are single-membrane-spanning proteins that associate with $Na_v1.5$ through their extracellular immunoglobulin-fold domains. They serve to increase the delivery of the α-subunit to the sarcolemmal membrane and to influence channel function. The β-subunits also enhance the subcellular localization of the channel and its interactions with various adaptors, signaling molecules, and cytoskeletal proteins. In addition to $Na_v1.5$, several "neuronal" α-subunits are expressed at low levels in the heart and likely contribute to regional heterogeneity in sodium channel function. Mutations in $Na_v1.5$, in specific β-subunits, and in several interacting regulatory and scaffolding proteins all may influence the behavior of the sodium current in the heart and thereby produce a range of arrhythmic syndromes, including long QT syndrome type 3 (LQT3) and Brugada syndrome (Chapter 65).[4]

The I_f current contributes to phase 4 depolarization in pacemaker cells and is a reflection of the activity of hyperpolarization-activated cyclic nucleotide gated, or *HCN* channels. The full channels are composed of dimers of *HCN* proteins, each of which has six transmembrane domains. *HCN4* and *HCN1* are the predominant isoforms found in the nodes, whereas *HCN2* is found throughout the conduction system. Binding of cyclic adenosine monophosphate, a key second messenger in the adrenergic signaling cascade, to the carboxy terminus of the channel shifts activation positively, thereby increasing the slope of phase 4 depolarization and linking autonomic tone to heart rate.

Calcium Channels

Voltage-gated calcium channels are important for generating the action potential in the sinoatrial and atrioventricular nodes and for excitation-contraction coupling in virtually all contractile cardiomyocytes. The dominant forms expressed in the heart are the L-type (large and long-lasting) and the T-type (tiny and transient) calcium channels, both of which include pore-forming α-subunits, similar in overall structure to that of voltage-gated sodium channels.

L-type Ca^{2+} channels comprise an $α_1$ subunit, a noncovalently bound β accessory subunit ($Ca_vβ1$-4), and an alternatively spliced $α_2$-δ-subunit that is post-translationally processed through cleavage and disulfide bond formation. The dominant $α_1$-subunit in the heart is $Ca_v1.2$, whereas $Ca_v1.3$ is restricted primarily to nodal and atrial cells. Both $α_1$-subunits are alternatively spliced to produce variants that are uniquely regulated.

The related T-type calcium channels are also found in the heart but display distinct biophysical properties compared with L-type Ca^{2+} channels; they activate at more negative voltages (−70 mV) and inactivate more rapidly. The major isoform expressed in the heart is heart is $Ca_v3.1$ and, to a lesser extent, $Ca_v3.2$. These channels are normally restricted to the nodes, Purkinje cells, and atrial myocytes. In pace-making cells, the T-type currents contribute to phase 4 depolarization. Mutations in Ca^{2+} channel subunits are also responsible for a number of arrhythmic syndromes, including Timothy syndrome (LQT8) and a subset of individuals with Brugada syndrome.

Potassium Channels

Numerous classes of potassium channels are expressed in the heart, where they contribute to repolarization and maintenance of the resting membrane potential. The heterogeneous expression of potassium channels in different regions and cell types is largely responsible for the variable action potential morphologies that are observed. As with sodium and calcium channels, potassium channels are formed from the assembly of pore-forming subunits along with various accessory β-subunits. However, α-subunits of potassium channels include between two and six transmembrane domains, and the full channel is formed as a dimer or tetramer, depending on the specific subfamily. Dysregulation of expression and function of potassium channels is quite common in many acquired forms of heart disease. The resulting loss of repolarizing currents leads to action potential duration prolongation and acquired long QT syndrome. In addition, heritable mutations that diminish potassium currents are responsible for several forms of inherited long QT syndrome.

Voltage-gated potassium channels, or *Kv* channels, are activated by membrane depolarization. Numerous classes of *Kv* channels have been identified in the heart. The α-subunits are six transmembrane domain proteins. Unlike sodium and calcium channels, functional potassium channels are formed by the assembly of four such subunits and various β-subunits. The transient outward current is composed of two components, $I_{to,fast}$ and $I_{to,slow}$; both are rapidly activated and contribute to phase 1 of the cardiac action potential,

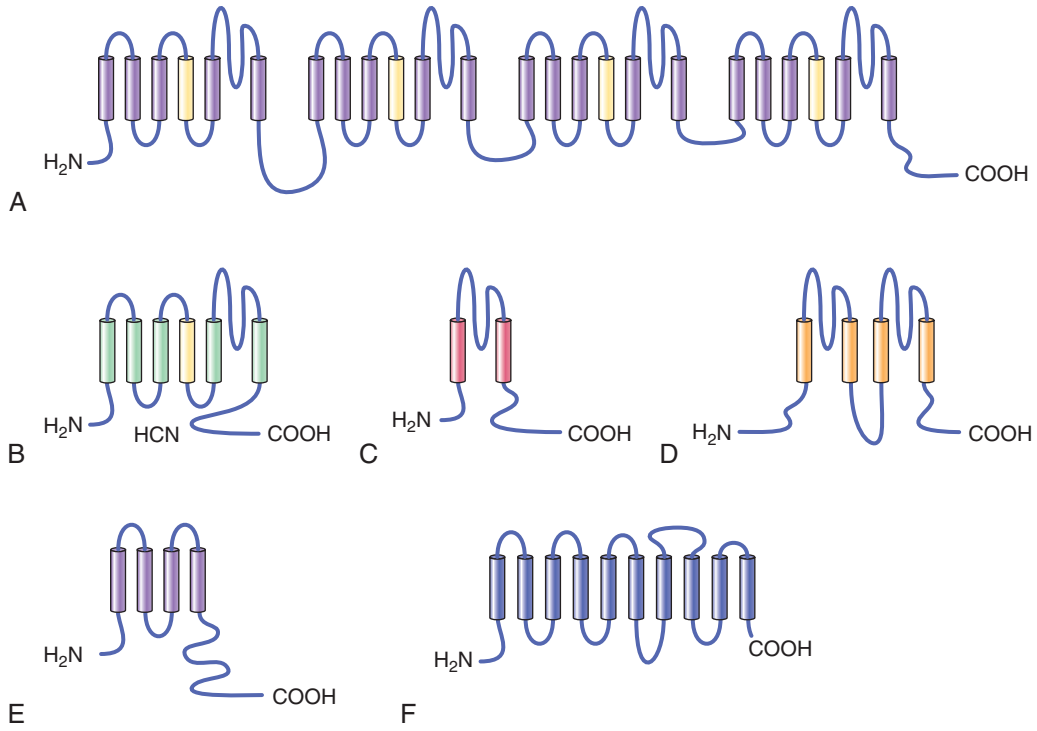

E-FIGURE 61-1. Structure of the pore-forming subunits of ion channels expressed in the heart. Individual barrels represent transmembrane domains. In all cases, the amino- and carboxy-terminal domains reside within the cytoplasmic compartment. **A,** Voltage-gated sodium and calcium channels. **B,** Voltage-gated potassium channel. **C,** Inwardly rectifying potassium channel. **D,** Two-pore domain potassium channel. **E,** Gap junction channel. **F,** Sodium-potassium adenosine triphosphatase.

but their recovery kinetics differ. $I_{to,fast}$ is particularly prominent in the epicardial layer of the ventricles, especially in the right ventricle. This differential expression is thought to contribute to the pathology of J wave syndromes, including Brugada syndrome.

The other major class of voltage-gated potassium channels in the heart is responsible for the delayed rectifier currents, broadly classified as I_K currents. These channels include the ultra-rapidly activating I_{Kur}, which is restricted to atrial myocytes, and the delayed rectifier currents, I_{Kr} and I_{Ks}, both of which contribute to phase 3 repolarization of the cardiac action potential. I_{Kr} activates and inactivates rapidly and shows strong inward rectification; that is, current moves more easily (but not exclusively) in the inward direction than in the outward direction, although it is the outward current that is physiologically relevant. I_{Ks} activates slowly and does not display inward rectification. Both currents also show marked regional heterogeneity. Numerous cardiac and noncardiac medications, as well as heritable syndromes that reduce the magnitude of these currents (particularly I_{Kr}), result in action potential duration prolongation and acquired or heritable long QT syndrome.

The second major class of potassium currents in the heart are the *Kir* currents carried by inwardly rectifying potassium channels. I_{K1} is observed in both atrial and ventricular cardiomyocytes. Conductance through these channels is high at negative membrane potentials, so this current is critical for terminal repolarization (phase 3) and for setting the resting membrane potential (phase 4). Another inwardly rectifying potassium current is carried by I_{KATP} channels. The full channels include not only the pore-forming subunit but also auxiliary *SUR* subunits, which are targets for channel inhibition by the sulfonylurea class of drugs. Because I_{KATP} currents are inhibited by intracellular adenosine triphosphate, they are activated in the setting of ischemia. The augmented outward current shortens action potential duration and abbreviates systole, thereby diminishing energetic requirements. Thus I_{KATP} channels provide a link between metabolic state and membrane excitability. Importantly, the resulting action potential duration shortening diminishes refractoriness, which may enhance the risk for re-entrant arrhythmias. The last major class of inward rectifiers includes the acetylcholine- and adenosine-activated potassium channels, which are encoded by *K_{ir}3.1* and *K_{ir}3.4*. These channels, which are enriched in nodal and atrial cardiac myocytes, are activated when ligands bind to muscarinic or purinergic G protein–coupled receptors, which facilitate the uncoupling of $G_{\beta\gamma}$ from G_α and the activation of the K_{ir} channels by the released $G_{\beta\gamma}$.

Gap Junction Channels

Gap junction channels, which are responsible for the electrotonic coupling of cardiac myocytes, are essential for normal impulse propagation throughout the myocardium. The channels are formed by the hexameric assembly of connexin monomers, each of which is a tetramembrane spanning protein. Connexin 43 is the dominant isoform expressed in ventricular and atrial myocardium, whereas connexin 40 is also abundantly expressed in the atrium. The nodes express variable amounts of connexin 45 and connexin 30.2, and the bundle branches and Purkinje fibers express significant levels of connexin 40. Gap junctions in the node integrate the intrinsic beating rate of each nodal cell into a single functional unit. Abnormalities in connexin expression and function, a process known as pathologic gap junction remodeling, are observed in atrial and ventricular myocardium in many acquired forms of heart disease. The remodeling contributes to aberrant impulse propagation and predisposes to arrhythmic behavior. In addition, germline or somatic mutations in cardiac connexin genes are associated with arrhythmic syndromes, especially atrial fibrillation.

MECHANISMS OF ARRYHTHMOGENESIS

Cardiac arrhythmias, which are disturbances in the rate or rhythm of the heartbeat, are a reflection of abnormal impulse formation or conduction. Inasmuch as cardiac myocytes reside within a complex multicellular environment and are electrotonically coupled by gap junction channels, arrhythmic syndromes almost always reflect a complex interplay of individual, or *cell autonomous*, properties within a multicellular network.[5] Most clinically important arrhythmias arise in the setting of acquired heart disease, in which pathologic *electrical remodeling*, resulting from dysregulation of ion channel expression or function, accompanies *structural remodeling*. However, many arrhythmic syndromes result from, or are exacerbated by, genetic variations, including disease-causing alterations in coding regions that directly affect the function of proteins, which regulate cardiac electrophysiology, as well as sequence variants in regulatory or other noncoding genome regions, which appear to regulate transcriptional and post-transcriptional behavior.

Disorders of Impulse Formation

In the healthy heart, the sinus node, which is located at the junction of the right atrium and the superior vena cava, is the predominant pacemaker. Secondary pacemakers with intrinsically slower pacing rates are found further downstream in the specialized conduction system within the atrioventricular node and the His-Purkinje system. The firing rate of pacemaker cells is regulated primarily by autonomic tone: sympathetic stimulation increases the slope of phase 4 depolarization, whereas parasympathetic stimulation decreases the slope by augmenting repolarizing currents. Nodal suppression may result from pharmacologic agents, such as β-adrenergic blockers, calcium channel blockers, or digitalis, as well as from fibrotic diseases. Moreover, mutations in several genes that affect the voltage clock (*SCN5A* and *HCN4*), the calcium clock (*RYR2* and *CASQ2*), or both (*ANKB*) may cause familial sinus node dysfunction.

Conversely, under pathologic conditions, myocardial cells outside the specialized conduction system may exhibit spontaneous activity, a phenomenon termed *abnormal automaticity*. Abnormal automaticity is most often seen with ischemia or reperfusion, in which maximum diastolic potentials are reduced to approximately −60 mV to −50 mV, a level at which Na^+ or Ca^{2+} channels may reach their activation threshold and trigger action potentials.

Afterdepolarizations and Triggered Activity

During cardiac repolarization, a number of inward and outward currents are active, and small changes in conductance of individual channels can markedly affect the trajectory of repolarization. *Afterdepolarizations*, which are abnormal oscillations in membrane potential, occur either during (early afterdepolarizations) or after (delayed afterdepolarizations) an action potential. Afterdepolarizations of sufficient magnitude to evoke an action potential produce triggered activity. Early afterdepolarizations are almost always observed in the setting of abnormal action potential duration prolongation, which provides sufficient time for *re*-activation of L-type Ca^{2+} channels during the plateau phase of the action potential. Thus, congenital syndromes, as well as bradycardia, hypokalemia, hypomagnesemia, antiarrhythmic medications, and many noncardiac drugs, are associated with QT prolongation and promote early afterdepolarizations. Conversely, rapid pacing and drugs that shorten the action potential duration tend to suppress early afterdepolarizations. Early afterdepolarizations in the setting of action potential duration prolongation often trigger *torsades de pointes* (Chapter 65), a polymorphic ventricular tachycardia, especially when there is increased dispersion of repolarization. Delayed afterdepolarizations, in contrast, are usually the result of intracellular Ca^{2+} overload and are typically seen in the setting of catecholamine excess, ischemia, toxic concentrations of digitalis-like agents, and some congenital syndromes, including catecholaminergic polymorphic ventricular tachycardia. The excessive Ca^{2+} load activates the electrogenic Na^+-Ca^{2+} exchanger, producing a depolarizing transient inward current, I_{TI}.

Disorders of Impulse Conduction

During each cardiac cycle, impulses must be generated in pacemaker cells within the sinus node, and a wave of excitation must propagate throughout the atria, travel down the specialized conduction system (including the atrioventricular node and His-Purkinje network), and then activate the ventricular myocardium. Processes that diminish intercellular coupling, such as fibrosis or calcification of the specialized conduction system, can diminish the safety factor for conduction and produce varying degrees of heart block. Inherited defects in conduction have been observed with mutations in sodium channel subunits *SCN5A* and *SCN1B*, which affect phase 0 of the cardiac action potential; in *KCNJ2*, which affects terminal repolarization and the resting membrane potential; and in a number of developmental disorders that affect the cardiac conduction system, such as Holt-Oram syndrome, Emery-Dreifuss muscular dystrophy, and myotonic dystrophy type 1. Conduction block may also be seen with the secondary electrical remodeling that is associated with structural heart disease and with many cardioactive drugs.

Re-entry

Re-entry is considered the most common mechanism responsible for clinically significant cardiac arrhythmias, including both supraventricular and ventricular disorders. Fundamentally, re-entry involves self-perpetuating waves of excitation that circulate around an inexcitable obstacle. Depending on the number of re-entrant waves within a tissue (one or multiple), their size, and their spatial stability, the surface electrocardiogram may reveal a relatively organized rhythm, such as atrial flutter or monomorphic ventricular

tachycardia, or a seemingly disorganized rhythm, such as atrial fibrillation or polymorphic ventricular tachycardia. Re-entry normally requires the presence of unidirectional block within a "fast" conducting pathway around an obstacle, combined with recirculation of the impulse from a second "slow" pathway in the retrograde direction, as might be the case at a bifurcating Purkinje-ventricular junction or around scar tissue of a healed myocardial infarction (Fig. 61-2). However, the "obstacle" may also be viable myocardium that is inexcitable owing to its intrinsic electrophysiologic properties, such as cellular uncoupling or refractoriness, a phenomenon referred to as *functional block*. Because refractoriness is critically dependent on the action potential duration, areas of myocardium with prolongation of the action potential duration may form a suitable substrate for functional re-entry.

Heterogeneity in action potential duration and the concomitant *dispersion of refractoriness* also play critical roles in the maintenance of arrhythmic behavior, especially through a phenomenon known as *phase 2 re-entry*. This term refers to the flow of current during phase 2 of the cardiac action potential from a depolarized cell to neighboring cells that are more fully repolarized and not refractory to reexcitation. This principle is best characterized in the J wave syndromes, especially Brugada syndrome, in which loss of function of inward currents (I_{Na} or I_{Ca}) or gain of function of outward currents (I_{to}, $I_{K\text{-}ATP}$) causes loss of the action potential dome during phase 2 and an abbreviated action potential duration in a subset of cardiac myocytes. Current can then flow into these cells from neighboring cells in which the action potential dome is maintained, thereby causing local reexcitation, a closely coupled extrasystole, and the initiation of re-entry. In Brugada syndrome, this process is thought to arise in the right ventricular outflow tract, where the transient outward current density is significantly greater in the epicardium compared with the endocardium (Fig. 61-3).

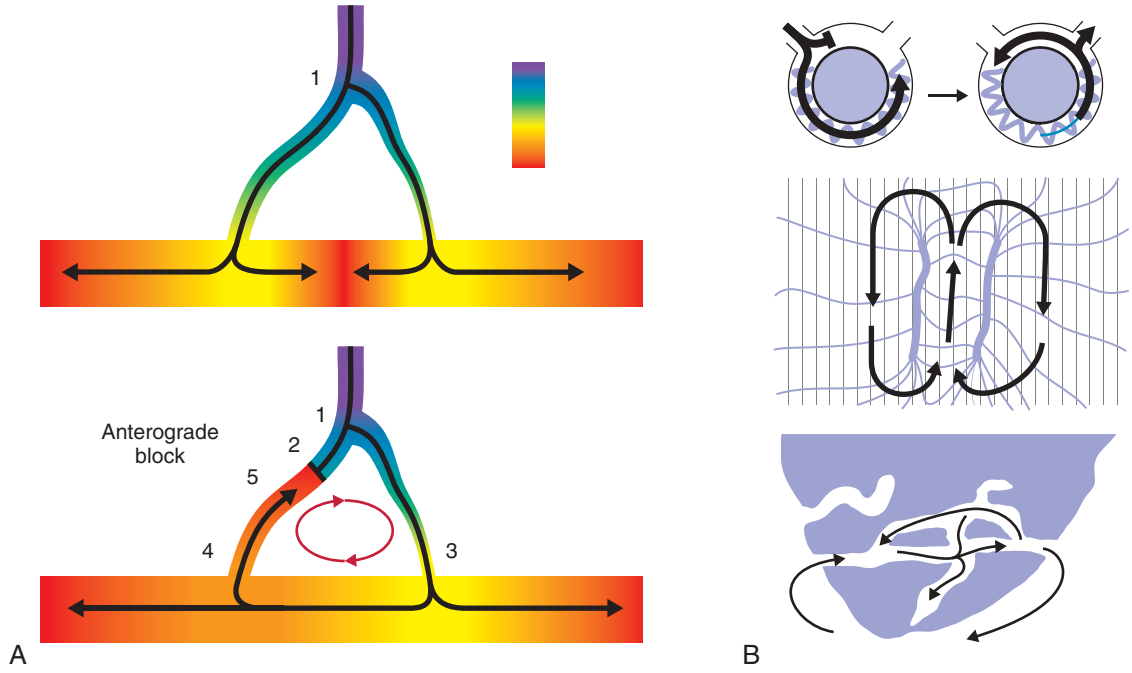

FIGURE 61-2. Re-entrant cardiac arrhythmias. A, Re-entry at the Purkinje-ventricular junction. *Upper panel:* Normally an impulse propagates along a Purkinje fiber and divides into two pathways (1), and together they activate the underlying ventricular myocardium. *Lower panel:* The impulse propagates along the right pathway (3) but is blocked within the left pathway (2). The original impulse travels within the ventricular myocardium, reenters the left pathway in the retrograde direction (4), and successfully propagates through the area with block (5). Continued propagation throughout this circuit *(red circle)* would produce re-entrant ventricular tachycardia. **B,** Re-entry associated with myocardium scar. *Upper panel:* Diagram representing a single circuit of re-entry that initiates with unidirectional block. The circuit length must be longer than the longest refractory period in the circuit. *Middle panel:* A figure 8, in which re-entry is established due to dispersion of refractoriness during tachycardia. *Lower panel:* Anatomic labyrinth circuit, created by strands of viable myocardium within the scar, with potential for multiple re-entry circuits. (The image to the right is reproduced from Benito B, Josephson ME. Ventricular tachycardia in coronary artery disease. *Rev Esp Cardiol [Engl Ed].* 2012;65:939-955.)

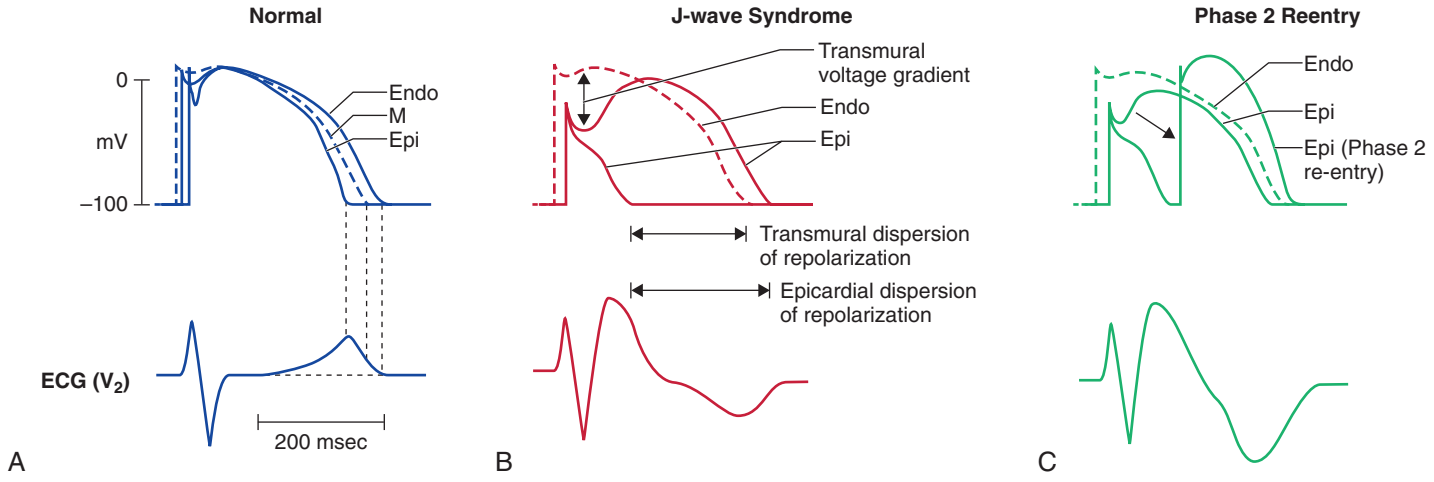

FIGURE 61-3. Cellular basis for J wave syndromes and phase 2 re-entry. Under normal conditions, the ST segment is isoelectric because of the absence of transmural voltage gradients at the level of the action potential plateau. Accentuation of the phase I notch under pathophysiologic conditions, such as loss of function of I_{Na} or gain of function of I_{to}, leads to loss of the action potential dome at some epicardial sites but not others, producing ST and T wave changes typically observed in Brugada syndrome. Loss of the action potential dome in epicardium but not endocardium results in the development of a marked transmural dispersion of repolarization and conduction of the action potential dome, from sites at which it is maintained to sites at which it is lost, thereby causing local reexcitation through a phase 2 re-entry mechanism. (Modified from Antzelevitch C, Brugada P, Brugada J, et al. Brugada syndrome: from cell to bedside. *Curr Probl Cardiol.* 2005;30:9-54.)

SUMMARY

The concepts of cardiac electrophysiology, including the genesis of the action potential, the molecular basis of cardiac excitability, and the mechanisms that are responsible for abnormal cardiac rhythms, provide a foundation for understanding and identifying clinically relevant arrhythmias. Insights into the biophysical basis of congenital arrhythmic syndromes and into the pathologic remodeling observed in acquired arrhythmic syndromes have already resulted in several targeted new therapies informed by the expression, function, and regulation of ion channels.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Park DS, Fishman GI. The cardiac conduction system. *Circulation*. 2011;123:904-915.
2. Nattel S, Harada M. Atrial remodeling and atrial fibrillation: recent advances and translational perspectives. *J Am Coll Cardiol*. 2014;63:2335-2345.
3. Webster G, Berul CI. An update on channelopathies: from mechanisms to management. *Circulation*. 2013;127:126-140.
4. George AL Jr. Molecular and genetic basis of sudden cardiac death. *J Clin Invest*. 2013;123:75-83.
5. Roberts BN, Yang PC, Behrens SB, et al. Computational approaches to understand cardiac electrophysiology and arrhythmias. *Am J Physiol Heart Circ Physiol*. 2012;303:H766-H783.

62

APPROACH TO THE PATIENT WITH SUSPECTED ARRHYTHMIA

JEFFREY E. OLGIN

CLINICAL MANIFESTATIONS

Patients with suspected arrhythmias can present in a variety of ways. Typical symptoms include palpitations, syncope, and presyncope (dizziness). On occasion, arrhythmias can manifest more subtly as exercise intolerance, lethargy, and vague complaints of malaise or without any symptoms at all. Conversely, arrhythmias occasionally manifest as aborted sudden cardiac death (cardiac arrest) (Chapter 63). The specific differential diagnosis, prognosis, and treatment of these symptoms are determined by the severity of the symptom (i.e., whether it results in syncope) and whether the patient has underlying structural heart disease. In general, the likelihood of a life-threatening arrhythmia, such as ventricular tachycardia or ventricular fibrillation, in a patient with symptoms of palpitations or syncope is significantly greater in a patient who has structural heart disease. Therefore, the determination of whether structural heart disease is present is a key step in the diagnosis and prognosis of patients with suspected arrhythmias.

Palpitations

Palpitations, defined as an awareness of an irregular or rapid heartbeat, are most commonly due to ectopic beats—namely, premature atrial contractions (PACs; Chapter 64) and premature ventricular contractions (PVCs; Chapter 65)—or to tachyarrhythmias. A careful history can often distinguish benign palpitations from those that need further evaluation. It can be useful to have the patient tap out with a finger what the palpitations feel like. An irregularly irregular pattern suggests atrial fibrillation, whereas a more regular, rapid pattern suggests a sustained tachycardia. A reliable symptom suggesting that palpitations are caused by a tachyarrhythmia, particularly a supraventricular tachycardia, is the sensation of a regular, rapid-pounding sensation in the neck. Conversely, most patients who complain of symptoms from PACs or PVCs are often more aware of the post-extrasystolic pause or the accentuated output of the post-extrasystolic beat than of the actual premature beat itself. Most patients who have symptoms suggestive of premature beats but not of sustained tachycardia do not require further evaluation if they have no other symptoms and no evidence of structural heart disease—that is, an otherwise normal cardiac history, physical examination, and electrocardiogram (ECG) (see Table 51-4). If, however, the symptoms are not due to a single occasional extrasystole or are accompanied by presyncope or syncope, further evaluation is required (Fig. 62-1). Antiarrhythmic therapy is usually not necessary to treat PACs or PVCs unless the symptoms are frequent or severe. β-Blockers (e.g., metoprolol 25 mg/day or atenolol 25 mg/day) are first-line therapy in highly symptomatic patients with documented PACs or PVCs.

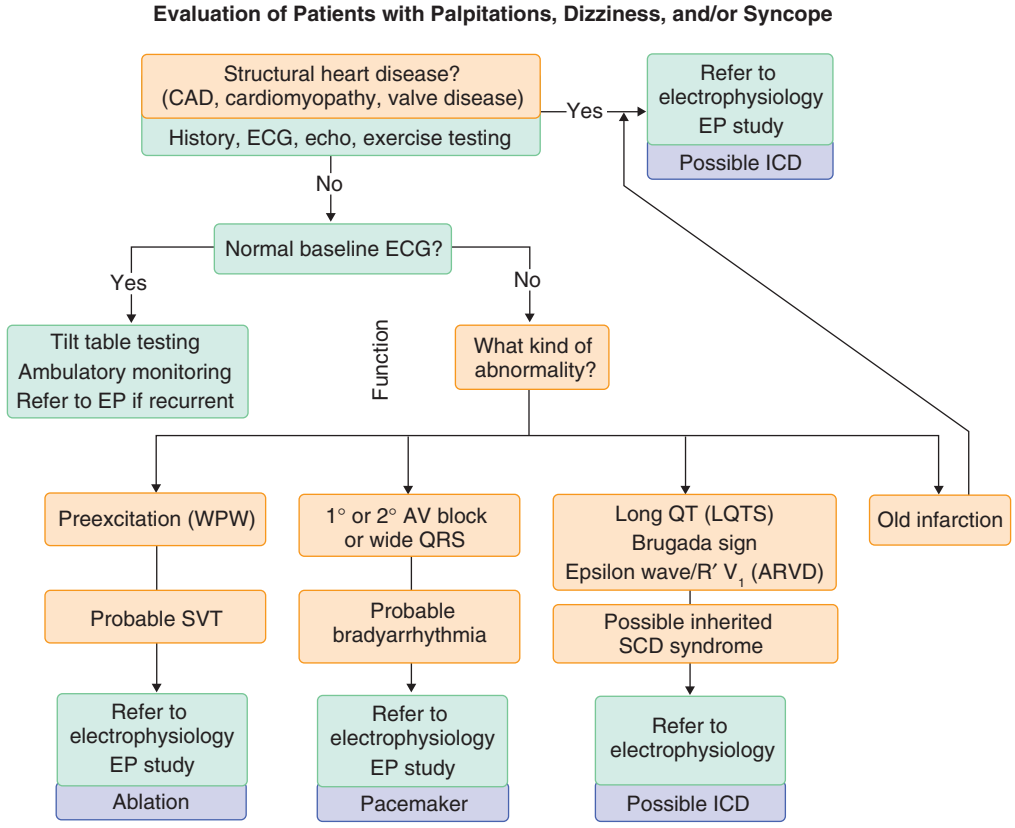

FIGURE 62-1. Algorithm for evaluating patients with symptoms of palpitation, dizziness, or syncope. ARVD = arrhythmogenic right ventricular dysplasia; AV = atrioventricular; CAD = coronary artery disease; ECG = electrocardiogram; echo = echocardiogram; EP = electrophysiology; ICD = implantable cardioverter-defibrillator; LQTS = long QT syndrome; SCD = sudden cardiac death; SVT = supraventricular tachycardia; WPW = Wolff-Parkinson-White syndrome.

Palpitations are the most common presentation of tachyarrhythmias. Most tachyarrhythmias in patients without structural heart disease are due to supraventricular tachycardias (Chapter 64) that resolve spontaneously within several seconds. When the tachyarrhythmia is more prolonged, it often resolves with simple interventions. Patients themselves can cough several times, perform the Valsalva maneuver, exhale forcefully against a closed glottis for several seconds, or even rub gently on their eyeballs. A physician can use carotid sinus massage (Chapter 64), performed by pressing and rubbing the carotid pulse just below the angle of the mandible for 5 to 15 seconds. This maneuver should be avoided in elderly patients and in patients with a history of cerebrovascular accident, known carotid artery stenosis, or carotid bruit on auscultation. In patients with structural heart disease, palpitations may signify ventricular tachycardia (Chapter 65), particularly if they occur with syncope or presyncope. Rarely do bradyarrhythmias manifest as palpitations.

Presyncope and Syncope

Syncope, defined as a sudden loss of consciousness, and presyncope, or lightheadedness, are caused by global impairment of blood flow to the brain (Table 62-1). Syncope can be a manifestation of tachyarrhythmias, bradyarrhythmias, or neurocardiogenic syncope, or it can be unrelated to any arrhythmia. A careful history and physical examination are necessary to exclude other cardiac causes (e.g., acute ischemia, aortic stenosis) or neurologic causes. Important historical features that suggest an arrhythmic cause are an association with palpitations and the lack of any neurologic deficits preceding or following the event. Important differential diagnoses include conditions other than lightheadedness that may be termed dizziness by the patient. Vertigo (Chapter 428), a sense of imbalance or of the "room spinning," and ataxia (Chapter 410) can usually be distinguished by the history and physical examination. The possibility of seizures (Chapter 403) must also be evaluated; syncope from an arrhythmia or neurocardiogenic syncope occasionally results in seizure-like activity, and seizures can sometimes be confused with syncope. The most important distinguishing feature is that postictal symptoms, a key feature of seizure disorders, are absent when syncope is the result of an arrhythmia. Patients with syncope from an arrhythmia usually awaken without any neurologic residual, unless the patient experienced a cardiac arrest with prolonged hypoxia and required resuscitation.

Because most spells of episodic loss of consciousness occur outside medical observation, the history is the most critical part of the evaluation (Table 62-2). Each syncopal episode should be reviewed in detail, with special attention to symptoms preceding the episode, events during unconsciousness, and the symptoms and time course of regaining orientation after consciousness is restored. Information from a witness can be essential to the evaluation.

The patient's presymptomatic activity and positioning, as well as symptoms when the syncopal episode began, are important clues to diagnosis. Seizures or cardiac arrhythmias can occur in any body position, but recumbent patients rarely develop neurocardiogenic (vasovagal) syncope and never have orthostatic hypotension. Prodromal lightheadedness, dizziness (but uncommonly vertigo), bilateral tinnitus, nausea, diffuse weakness, and dimming of vision are symptoms of cerebral hypoperfusion and support the diagnosis of syncope, which may be from a cardiac, orthostatic, or neurocardiogenic cause. Loss of consciousness so rapid that a prodrome is absent may occur with seizures and with some cardiac arrhythmias such as asystole, which typically causes loss of consciousness within 4 to 8 seconds in the upright position but usually requires 12 to 15 seconds in the recumbent position. Palpitations during the prodrome suggest a tachyarrhythmia. The activity of the patient immediately before the onset of symptoms may also provide clues. Syncope associated with the cessation of exertion or with anxiety or pain suggests neurocardiogenic syncope, whereas symptoms during exertion suggest an arrhythmia. Syncope associated with a change in posture suggests orthostatic causes, whereas syncope while straining at urination suggests situational neurocardiogenic syncope.

A witness's description of the events during the episode of unconsciousness is very helpful. Although body stiffening and limb jerking occur with generalized seizures, similar movements can result from cerebral hypoperfusion, especially if perfusion is not restored rapidly. Such muscle jerking is often

TABLE 62-1 CAUSES OF SYNCOPE AND THEIR PREVALENCE

NEUROCARDIOGENIC CAUSES
Vasovagal (8-41% of patients)
Situational (1-8% of patients)
Micturition
Defecation
Swallow
Cough
Carotid sinus syncope (0.4% of patients)
Neuralgias
Psychiatric disorders
Medications, exercise
ORTHOSTATIC HYPOTENSION (4-10% OF PATIENTS)
DECREASED CARDIAC OUTPUT
Obstruction to flow (1-8% of patients)
Obstruction to left ventricular outflow or inflow: aortic stenosis, hypertrophic obstructive cardiomyopathy, mitral stenosis, myxoma
Obstruction to right ventricular outflow or inflow: pulmonic stenosis, pulmonary embolism, pulmonary hypertension, myxoma
Other heart disease
Pump failure, myocardial infarction, coronary artery disease, coronary spasm, tamponade, aortic dissection
ARRHYTHMIAS (4-38% OF PATIENTS)
Bradyarrhythmias: sinus node disease, second- and third-degree atrioventricular block, pacemaker malfunction, drug-induced bradyarrhythmias
Tachyarrhythmias: ventricular tachycardia, torsades de pointes (e.g., associated with congenital long QT syndrome or acquired QT prolongation), supraventricular tachycardia
NEUROLOGIC AND PSYCHIATRIC DISEASES (3-32% OF PATIENTS)
Migraine
Transient ischemic attacks
UNKNOWN (13-41% OF PATIENTS)

Adapted from Kapoor W. Approach to the patient with syncope. In: Braunwald E, Goldman L, eds. *Primary Cardiology*, 2nd ed. Philadelphia: Saunders; 2003.

TABLE 62-2 CLINICAL FEATURES SUGGESTING SPECIFIC CAUSES

DIAGNOSTIC CONSIDERATION
Neurocardiogenic
Symptoms after prolonged motionless standing, sudden unexpected pain, fear, or unpleasant sight, sound, or smell
Syncope in a well-trained athlete after exertion (without heart disease)
Situational syncope during or immediately after micturition, cough, swallowing, or defecation
Syncope with throat or facial pain (glossopharyngeal or trigeminal neuralgia)
Organic Heart Disease (e.g., coronary artery disease, aortic stenosis, primary arrhythmia, obstructive hypertrophic cardiomyopathy, pulmonary hypertension)
Brief loss of consciousness, no prodrome, history of heart disease
Syncope with exertion
Family history of sudden death
Neurological
Seizures: confusion for >5 min after regaining consciousness
Transient ischemic attack, subclavian steal, basilar migraine: syncope associated with vertigo, dysarthria, diplopia, arm exercise
Migraine: syncope associated with antecedent headaches
Other Vascular
Carotid sinus: syncope with head rotation or pressure on the carotid sinus (as in tumors, shaving, tight collars)
Orthostatic hypotension: syncope immediately on standing
Subclavian steal or aortic dissection: differences in blood pressure or pulse between the two arms
Drug-Induced
Patient is taking a medication that may lead to long QT syndrome, orthostasis, or bradycardia
Psychiatric Illness
Frequent syncope, somatic complaints, no heart disease

Adapted from Kapoor WN. Syncope. *N Engl J Med.* 2000;343:1856-1862.

multifocal and can be synchronous or asynchronous. In contrast to epileptic seizures, which generally produce tonic-clonic activity for at least 1 to 2 minutes, muscle jerking in syncope rarely persists for longer than 30 seconds. If an arrhythmia continues or the patient is physically maintained upright, tonic stiffening of the body followed by jerking movements of the limbs can occur. Occasionally, motor movements identical to a tonic-clonic seizure occur, and a mistaken diagnosis of epilepsy can be made. Urinary incontinence during the spell is frequently used to support or refute a diagnosis of epilepsy; however, fainting with a full bladder can result in incontinence, whereas seizures with an empty bladder will not. Tongue biting favors seizures.

The time frame over which consciousness and orientation are regained is perhaps the most important clue in differentiating seizures from syncope. Recovery of orientation after neurocardiogenic syncope occurs within seconds of regaining consciousness. Recovery of orientation after self-reversible arrhythmia-associated syncope is usually proportional to the duration of the unconsciousness and is usually rapid (0 to 10 seconds). Life-threatening arrhythmias (e.g., prolonged asystole or ventricular fibrillation) usually do not resolve without resuscitation, and the confusion after regaining consciousness may be permanent owing to ischemic brain injury (Chapter 63). By comparison, the period of confusion after seizures, often accompanied by agitation, continues for 2 to 20 minutes after recovery of consciousness.

DIAGNOSIS

Arrhythmias are generally categorized as bradyarrhythmias (slow heart rates), tachyarrhythmias (fast heart rates), or premature beats (single extrasystoles from the atrium or the ventricle—PACs [see Fig. 64-10] or PVCs [see Fig. 65-1], respectively) (see Table 62-1). Although not a primary arrhythmia, neurocardiogenic syncope is a related diagnostic and management issue because its symptoms are frequently similar to those of arrhythmias and because neurocardiogenic syncope secondarily results in bradycardia (see later).[1] A systematic approach can optimize the likelihood of identifying the cause of transient loss of consciousness (Table 62-3).

Bradyarrhythmias

Bradyarrhythmias (Chapter 64) can be due to dysfunction in the sinoatrial node, atrioventricular (AV) node, or His-Purkinje system (below the AV node). Sinus bradycardia manifests as a slow atrial (sinus) rate and can occur at rest or as an inappropriately slow rate during exercise (chronotropic incompetence). Sinus arrest can be intermittent, when transient loss of sinus activity (loss of the P wave on the ECG) causes brief sinus pauses, or persistent, with prolonged loss of atrial activation. The sinus rate and even the presence of sinus pauses are influenced by autonomic tone. Therefore, healthy individuals—particularly younger patients and well-trained athletes (with high vagal tone)—have occasional sinus slowing, often during sleep. A sinus pause of more than 3 seconds is considered pathologic if it is associated with symptoms while a patient is awake. Sinus bradycardia and sinus arrest can also be the result of medications, typically β-blockers and calcium-channel blockers. When not "physiologic" or due to medications, sinus bradycardia and sinus arrest are the result of intrinsic conduction system disease. Sinus bradycardia, especially if it is intermittent, can also signify disease of the right coronary artery.

Bradyarrhythmias from AV nodal disease result from the failure of impulse conduction from the atrium to the ventricle. Like the sinus node, the AV node is dramatically affected by autonomic tone. Mobitz type I second-degree AV block (Wenckebach block; see Fig. 64-6) can be seen during periods of high vagal tone (such as while sleeping) and is not necessarily pathologic; for example, it does not progress to complete heart block and is not associated with a widened QRS. Many drugs, such as β-blockers and calcium-channel blockers, commonly cause first-degree AV block and should be considered a potential cause of any degree of AV block. Mobitz type II block (see Fig. 64-7) signifies that the level of AV block is below the AV node in the His-Purkinje system, which is not sensitive to autonomic tone; the resulting QRS is widened, and there is a high likelihood of progression to complete heart block (third-degree AV block; see Figs. 64-8 and 64-9). Idiopathic paroxysmal atrioventricular block, detected by continuous ECG monitoring, can also cause syncope. Intermittent complete heart block, which can result in drop attacks or Stokes-Adams attacks, is usually preceded by abnormal baseline findings on the ECG, such as a bundle branch block or second-degree AV block. The treatment of choice for patients with symptomatic bradyarrhythmias or those likely to progress to complete heart block is implantation of a permanent pacemaker (Chapter 66).

TABLE 62-3 SUMMARY OF CLINICAL RECOMMENDATIONS FOR TRANSIENT LOSS OF CONSCIOUSNESS

TOPIC	RECOMMENDATIONS
Initial assessment	Detailed history, especially from witnesses Full clinical examination 12-lead ECG
Uncomplicated faints	Suggestive features include: Posture: occurrence during prolonged standing or similar previous episodes avoided by lying down Provoking factors, such as pain or a medical procedure Prodromal symptoms, such as sweating or feeling warm or hot before TLoC Further investigation and specialist referral are not needed.
Epilepsy	Suggestive features are a bitten tongue; head turning to one side during TLoC; no memory of abnormal behavior that occurred before, during, or after TLoC; unusual posturing; prolonged limb jerking (brief seizure-like activity often occurs during syncope, including uncomplicated faints); confusion after the event; or prodromal déjà vu or jamais vu. If features of epilepsy are present, arrange for early review by an epilepsy specialist. Do not arrange for EEG before neurologic assessment. Note that brief seizure-like activity often occurs during syncope, including uncomplicated faints. Do not suspect epilepsy unless suggestive features are present. Arrange for cardiovascular assessment if the cause of TLoC is unclear.
Urgent specialist referral	Give immediate treatment for clinically urgent problems (such as complete AV block or severe bleeding). Arrange for urgent specialist cardiovascular assessment for patients at risk for a severe adverse event (such as those with long QT interval, cardiac arrhythmia, or structural heart disease).
Further cardiovascular assessment	Focus on specific disorders that may cause TLoC, such as orthostatic hypotension, the carotid sinus syndrome, structural heart disease, or cardiac arrhythmia. Assessment should include repeated history, clinical examination, and 12-lead ECG. For suspected cardiac arrhythmia or unexplained TLoC, use ambulatory ECG for further assessment: Very frequent episodes: use 24- to 48-hour Holter monitoring. Moderately frequent episodes: use external event monitoring. Infrequent episodes: use an implantable event recorder.

AV = atrioventricular; ECG = electrocardiography; EEG = electroencephalogram; TLoC = transient loss of consciousness.
Adapted from Cooper PN, et al. Synopsis of the National Institute for Health and Clinical Excellence Guideline for management of transient loss of consciousness. *Ann Intern Med.* 2011;155:543-549.

Tachyarrhythmias

Tachyarrhythmias can arise from the atrium or AV node (supraventricular tachycardia) or from the ventricle (ventricular tachycardia). Supraventricular tachyarrhythmias that may be associated with palpitations, presyncope, or syncope include atrial tachycardia (see Fig. 64-16), AV nodal re-entrant tachycardia (see Fig. 64-15), AV junctional tachycardia (see Fig. 64-18), atrial flutter (see Fig. 64-21), and atrial fibrillation (see Fig. 64-22), sometimes in association with accessory conduction pathways that facilitate the re-entry needed to sustain the arrhythmia. Ventricular tachyarrhythmias include the various forms of ventricular tachycardia (see Figs. 65-2 through 65-4). Treatment is guided by the specific tachyarrhythmia and its underlying cause (Table 62-4; see Tables 64-5 and 64-6) (Chapters 63 through 66).

Neurocardiogenic Syncope and Related Syndromes

Neurocardiogenic syncope is the sudden onset of lightheadedness or loss of consciousness as a result of autonomic reflexes and is more common in younger patients (teenage to third decade of life). It is sometimes called a vasovagal episode, a common faint, or situational syncope if it is clearly induced by a particular activity (e.g., micturition syncope). Some families have autosomal dominant vasovagal syncope, which is genetically heterogeneous but seems to be linked to chromosome 15q26.[2]

In this form of neurocardiogenic syncope, heightened parasympathetic output, either due to direct stimulation (e.g., micturition, defecation,

TABLE 62-4 INDICATIONS FOR INITIAL OBSERVATION AND RAPID EVALUATION OF SYNCOPE

EUROPEAN SOCIETY OF CARDIOLOGY*	CANADIAN CARDIOVASCULAR SOCIETY†
Known coronary or structural heart disease, heart failure, or prior arrhythmia	Heart failure or history of ischemic, arrhythmic, obstructive, or valvular heart disease
ECG showing nonsustained ventricular tachycardia, bifascicular block, sinus bradycardia <50 beats/min, sinoatrial block, preexcitation, or evidence of an inherited disease	Abnormal ECG: arrhythmia, conduction disease, new ischemia, or evidence of prior myocardial infarction
Syncope during exertion or when supine, palpitations preceding syncope, family history of sudden cardiac death	Systolic blood pressure <90 mm Hg
Important comorbidities (e.g., severe anemia, electrolyte disturbance)	Comorbid conditions: age >60 years, dyspnea, hematocrit <30%, hypertension, cerebrovascular disease, family history of sudden death before age 50 years, syncope while supine, syncope during exercise, syncope with no prodromal symptoms

ECG = electrocardiogram.
*Moya A, Sutton R, Ammirati F, et al. Guidelines for the diagnosis and management of syncope (version 2009): the Task Force for the Diagnosis and Management of Syncope of the European Society of Cardiology (ESC). *Eur Heart J.* 2009;30:2631-2671.
†Sheldon RS, Morillo CA, Krahn AD, et al. Standardized approaches to the investigation of syncope: Canadian Cardiovascular Society position paper. *Can J Cardiol.* 2011;27:246-253.

abdominal pain, or other gastrointestinal conditions) or as a reflex in response to sympathetic stimulation (e.g., seeing blood, abrupt cessation of exercise), results in arterial dilation (called the vasodilatory response) and an inhibition of sinus and AV node activity (the cardioinhibitory response). The result is a transient decrease in blood pressure, often manifested as lightheadedness or syncope. Because they are associated with parasympathetic (vagal) output, episodes are frequently accompanied by nausea, diaphoresis, and salivation. Twin analyses provide strong evidence for genetic factors in vasovagal syncope.

Treatment of this form of syncope can be challenging. The most effective therapies are behavioral (avoidance of triggers), wearing of compression stockings, and maintenance of adequate hydration and salt intake. Lying down with the feet elevated and performing isometric hand exercises may abort an acute episode. Medical therapy, including β-blockers (pindolol 5 to 15 mg twice daily), mineralocorticoids (fludrocortisone 0.1 mg/day), paroxetine (10 to 20 mg/day), and midodrine (an α-adrenergic agonist and vasoconstrictor; 5 to 10 mg three times daily), has shown some efficacy in reducing recurrence rates, although the efficacy of β-blockers for reducing syncopal episodes has been inconsistent.

Rarely, situational syncope is associated with swallowing or coughing. Swallowing can trigger brain stem reflexes that lead to vagally induced bradyarrhythmias, with resultant syncope. This phenomenon may or may not be associated with severe pain in the tonsillar pillar, which may radiate to the ear (i.e., glossopharyngeal neuralgia; Chapter 398). The pain can usually be prevented by carbamazepine (400 to 1000 mg/day total in divided doses of 2-3 times per day orally); in refractory cases, 300 mg/day in divided doses of 1-4 times per day of phenytoin can be added. Cough-related syncope can occur with severe, repeated coughing, which may increase thoracic pressure and result in increased vagal tone or a transient reduction in outflow from the intracranial veins, followed by a transient increase in intracranial pressure and impaired blood flow.

A related cause of syncope is carotid body hypersensitivity, in which vagal tone is increased by direct stimulation of the carotid body. This condition is frequently seen in older patients (particularly men older than 60 years), in whom episodes are associated with mechanical stimulation of the neck (e.g., turning the head, shaving, wearing a tight collar or necktie). Use of β-blockers, calcium-channel blockers, and digitalis can exacerbate or predispose to this condition. This form of syncope is diagnosed by documenting pauses longer than 3 seconds in response to carotid sinus massage and is curable with a pacemaker because carotid body stimulation does not cause significant vasodilation.

Postural or orthostatic hypotension can result in recurrent syncope. The history confirms that the patient is in the upright posture during spells, that the prodromal symptoms are those of cerebral hypoperfusion, and that the symptoms are relieved with recumbency. The diagnosis is supported by detecting a decrease of 30 mm Hg or greater in systolic blood pressure or a decrease of 10 mm Hg or greater in diastolic blood pressure between recumbent and upright postures. The many causes include drugs, polyneuropathies (Chapter 420), and neurodegenerative disorders (Chapter 409).

Cerebrovascular syncope results from cerebral hypoperfusion due to vascular phenomena, as opposed to generalized hypotension caused by arrhythmias or neurocardiogenic reflexes. Loss of consciousness can be a component of a basilar artery transient ischemic attack, but other brain stem symptoms nearly always precede or accompany the unconsciousness. Vertigo is most frequent, but diplopia or visual field disturbances, hemifacial or perioral numbness, and dysarthria or ataxia are also common. Recovery of consciousness may require 30 to 60 minutes. Although the diagnosis is suggested by the history and clinical presentation, imaging studies can be useful to confirm the diagnosis. Carotid Doppler studies may show various degrees of stenosis, especially in older patients. However, unconsciousness requires bihemispheric dysfunction; thus, unilateral carotid stenosis alone does not cause syncope. Transcranial Doppler studies or magnetic resonance angiography of the basilar artery is indicated only if brain stem ischemic symptoms are present in addition to loss of consciousness; false-positive tests are common, especially with increasing age. These patients, who are at risk for basilar artery stroke, should be treated with aspirin and should be considered for other treatments (e.g., surgery, stent placement) appropriate for their symptoms and anatomy (Chapter 407).

Other syndromes that can cause syncope include subclavian artery stenosis, which may result in retrograde blood flow from the vertebral artery to one arm, with resultant brain stem hypoperfusion (i.e., subclavian steal syndrome). Asymmetry in upper extremity systolic blood pressure, typically averaging 45 mm Hg, is nearly always present. Brain stem symptoms are similar to those in basilar transient ischemic attacks, including loss of consciousness, but a subsequent stroke from subclavian steal is rare. Repair of the stenosis is the treatment of choice. Syncope may also occur in up to 10% of patients with basilar artery migraine (Chapter 398). It can have a postural (orthostatic) manifestation or be associated with other basilar artery symptoms.

Neuropsychiatric syncope is a diagnosis of exclusion but is suggested by young age, frequent spells, multiple symptoms (e.g., dizziness, vertigo, lightheadedness, numbness), and duplication of the patient's symptoms by hyperventilation with the mouth open for 2 to 3 minutes. Whereas syncope and seizures occur with the eyes open, often with gaze deviation, psychogenic events frequently begin with eye closing.

Seizures (Chapter 403) can cause loss of consciousness and occasionally present clinically as syncope. However, seizures usually have a characteristic presentation and include a postictal phase, whereas most patients experiencing a syncopal episode quickly regain consciousness, except when cerebral perfusion is so compromised as to cause a secondary seizure or persistent anoxia and brain damage.

Diagnostic Tests

Electrocardiography

The baseline ECG is critical in the evaluation of a patient with palpitations or syncope. The presence of ventricular preexcitation, as manifested by a short PR interval and a delta wave (see Fig. 64-19), establishes the likely diagnosis of Wolff-Parkinson-White syndrome in a patient with palpitations and AV reciprocating tachycardia (Chapter 64); it can also be used to determine the location of the responsible accessory pathway. The baseline ECG provides useful predictive information about the likelihood of conduction system abnormalities being responsible for bradyarrhythmias (e.g., sinus bradycardia suggests sinus node dysfunction, a prolonged PR interval suggests AV nodal disease, and a widened QRS suggests disease below the AV node). The ECG is also useful in diagnosing prior myocardial infarction (i.e., pathologic Q waves), which raises the likelihood of ventricular tachycardia as a potential cause of syncope or palpitations. Abnormalities such as a prolonged QT interval in a patient with syncope and a family history of syncope or sudden death suggest one of the congenital long QT syndromes (Chapter 65). An incomplete right bundle branch block with coved ST segment elevation in ECG lead V_1 or V_2 in a patient with syncope or palpitations suggests Brugada syndrome, whereas an epsilon wave, incomplete right bundle branch block, and inverted T waves in V_1 are suggestive of right ventricular dysplasia (Chapter 65). All these syndromes carry an increased risk for recurrent syncope and sudden death if untreated (Chapters 63 through 65). The short QT syndrome also predisposes to ventricular arrhythmias, but currently there is no clear definition of a pathologically short QT duration.

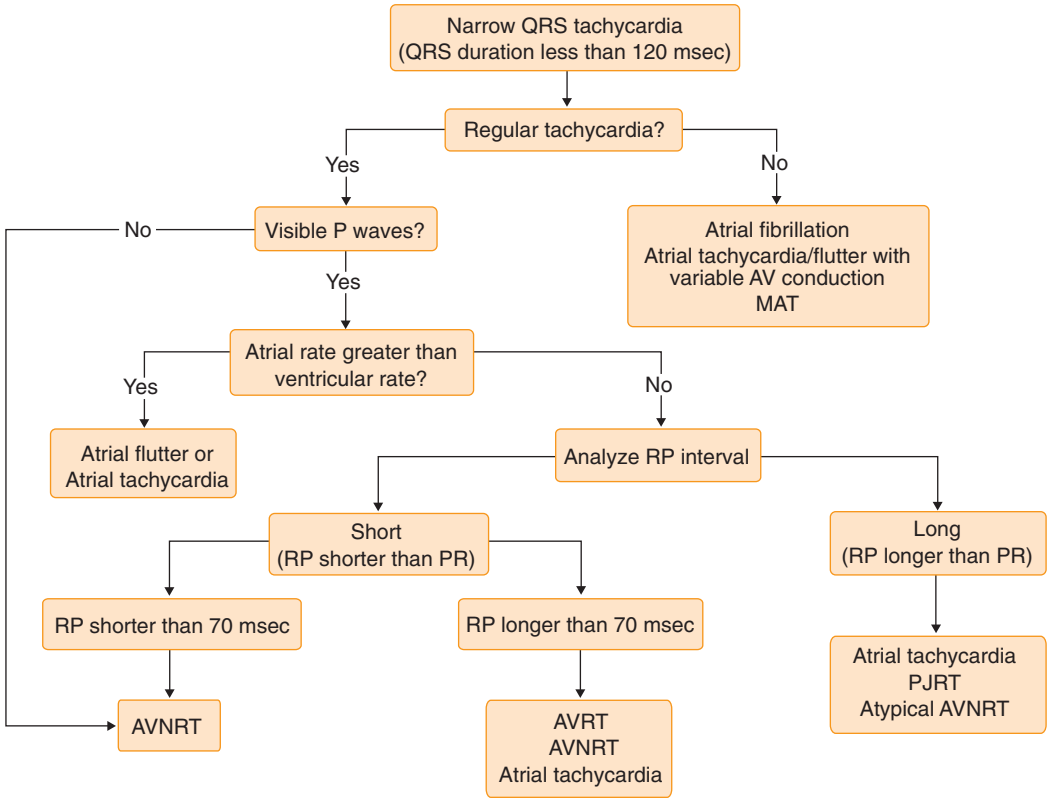

FIGURE 62-2. ECG algorithm for diagnosis of narrow-complex tachycardias. AVNRT = atrioventricular nodal reciprocating tachycardia; AVRT = atrioventricular reciprocating tachycardia; MAT = multifocal atrial tachycardia; PJRT = permanent form of junctional reciprocating tachycardia. (From Blomstrom-Lundqvist C, Scheinman MM, Aliot EM, et al. ACC/AHA/ESC guidelines for the management of patients with supraventricular arrhythmias—executive summary. *Circulation.* 2003;108:1871-1909.)

Performing an ECG during an episode of palpitations is extremely useful in making a definitive diagnosis. For narrow–QRS complex tachycardias, the specific supraventricular tachycardia can often be surmised from the 12-lead ECG obtained during symptoms (Fig. 62-2). Moreover, for wide–QRS complex tachycardias, the 12-lead ECG is useful in distinguishing a supraventricular tachycardia (with aberrancy) from a ventricular tachycardia (Fig. 62-3). The presence of fusion beats or AV dissociation during a wide–QRS complex tachycardia leads to the diagnosis of ventricular tachycardia. For ventricular tachycardias, the morphology of the QRS complex is useful in determining the location of the ventricular tachycardia focus and in identifying idiopathic ventricular tachycardia (right ventricular outflow tract or fascicular), which has a much more benign course than ventricular tachycardia in the setting of coronary disease (Chapter 65).

The effect of carotid sinus massage, vagal maneuvers, or adenosine (given as a rapid intravenous bolus of 6 mg and repeated at a dose of 12 mg if the initial dose is ineffective) is also useful in narrowing the differential diagnosis of a tachycardia. These maneuvers slow conduction through the AV node. Therefore, tachycardias that terminate with either maneuver are likely to involve the AV node as a critical component of the re-entrant circuit (AV nodal re-entrant tachycardia or AV re-entrant tachycardia). If the maneuver induces AV block but does not terminate the arrhythmia, likely causes are atrial fibrillation, atrial flutter, and atrial tachycardias (or occasionally ventricular tachycardia if the QRS is wide). On rare occasions, atrial tachycardias and some idiopathic ventricular tachycardias terminate in response to adenosine. Important clues to the specific mechanism can be obtained at the onset or termination of tachycardia, so obtaining a continuous 12-lead ECG during carotid sinus massage or the administration of adenosine is very useful.

During bradycardias, the ECG is useful in determining the level of the conduction system (sinus node, AV node, or His bundle) responsible for the bradycardia. Sinus bradycardia is diagnosed when a slow (<50/minute at rest) atrial rate (P wave) conducts to the ventricle. Sinus arrest or sinus pauses are diagnosed by absent or dropped P waves. First-degree AV block (see Fig. 64-5) is defined as a prolonged PR interval (>200 msec), and second-degree AV block is defined by P waves that occasionally do not conduct to the ventricle (P wave without an ensuing QRS); Mobitz type I second-degree AV block (also known as Wenckebach block; see Fig. 64-6) is characterized by progressive lengthening of the PR interval until one P wave does not conduct to the ventricle. This form of AV block is often seen in younger patients, is usually benign, and rarely progresses to complete AV (third-degree) block. Mobitz type II second-degree AV block (see Fig. 64-7), which is characterized by the sudden, unexpected loss of conduction of a P wave to the ventricle (dropped QRS), signifies disease of the His-Purkinje system and often progresses to complete heart block. Complete heart block or third-degree AV block (see Figs. 64-8 and 64-9) is diagnosed by the dissociation of P waves from QRS complexes, with an atrial rate faster than the ventricular rate.

Ambulatory Monitoring

For intermittent symptoms such as palpitations, dizziness, or syncope, it is often difficult to obtain a 12-lead ECG while the symptoms are occurring. Therefore, ambulatory monitoring, which allows ECG monitoring over long periods, is a vital diagnostic tool. There are currently three types of ambulatory monitors: Holter monitors, which continuously record the ECG for 24 to 48 hours; event recorders, which are wearable loop recorders that record only during specific events (when the patient activates the recorder because of symptoms or the recorder detects a heart rate above or below a specified threshold) and can be worn for 1 month or more; and implantable loop recorders, which function similarly to event recorders but can be used for up to 14 months. In addition, home telemetry units can allow patients to undergo prolonged continuous remote monitoring by wireless or Internet connections. The choice of ambulatory monitoring method is largely determined by the frequency of the symptoms and the likelihood of capturing an episode during a given monitoring period.

Ambulatory monitoring is diagnostic only if abnormalities occur during symptoms or if the patient has typical symptoms without any concurrent abnormalities. A "normal" monitoring record is nondiagnostic if the patient does not have symptoms during the period.

Holter Monitors

Holter monitors use either a tape (in older devices) or digital media (in newer devices) to record a 3-, 5-, or 12-lead surface ECG continuously, usually for 24 to 48 hours but for 3 weeks or more when indicated. Processing, printing, and analysis of the recordings are performed offline with commercial systems.

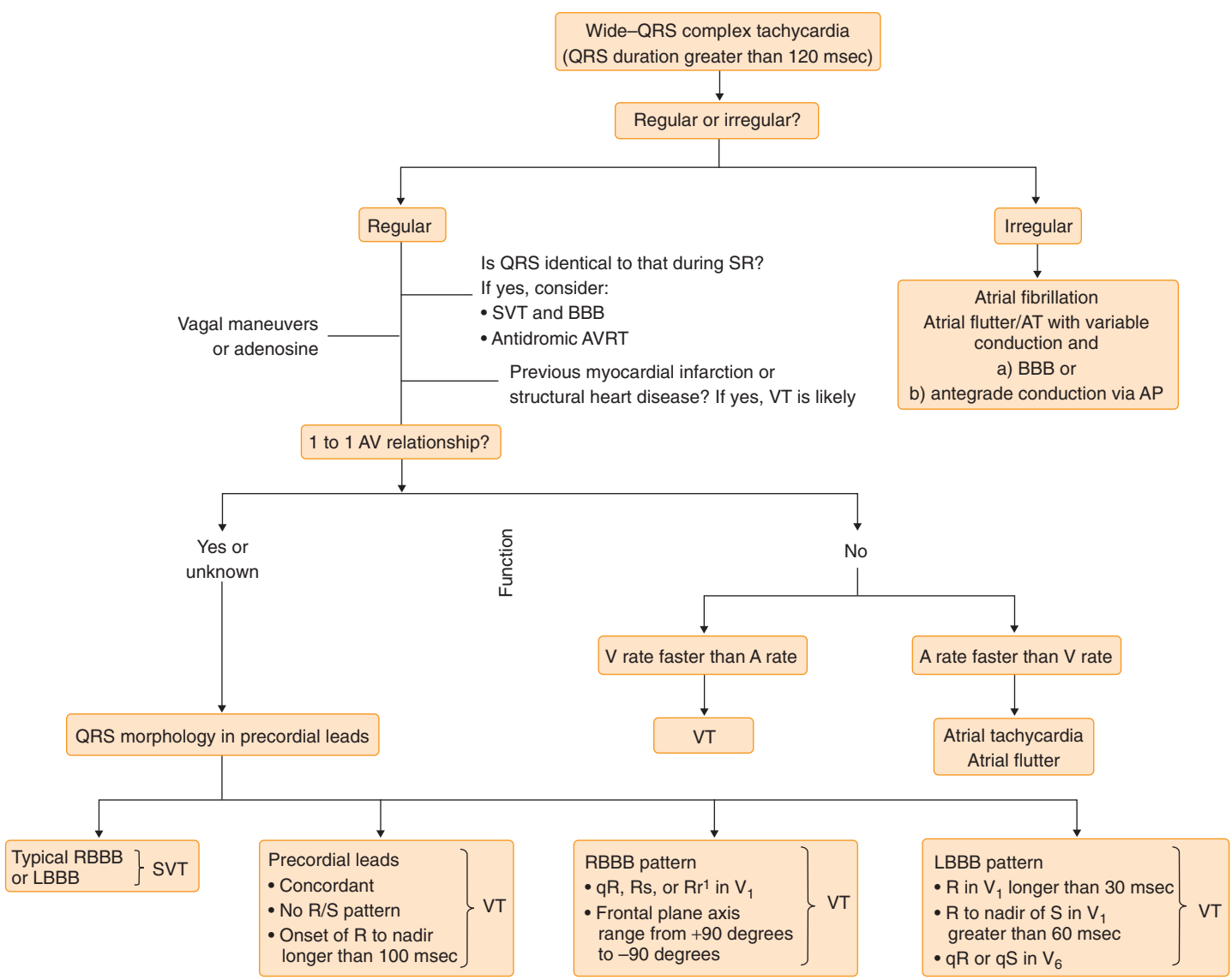

FIGURE 62-3. ECG algorithm for diagnosis of wide-complex tachycardias. A = atrial; AP = accessory pathway; AT = atrial tachycardia; AV = atrioventricular; AVRT = atrioventricular reciprocating tachycardia; BBB = bundle branch block; LBBB = left bundle branch block; RBBB = right bundle branch block; SR = sinus rhythm; SVT = supraventricular tachycardia; V = ventricular; VT = ventricular tachycardia. (From Blomstrom-Lundqvist C, Scheinman MM, Aliot EM, et al. ACC/AHA/ESC guidelines for the management of patients with supraventricular arrhythmias—executive summary. *Circulation.* 2003;108:1871-1909.)

In addition to recording the rhythm, analyses of heart rate variability and ST segment changes and accurate counts of PACs and PVCs can be automated. Some systems allow extrapolation to produce a "virtual" 12-lead recording at any time during the monitoring period. Holter monitoring is useful for detecting symptoms that are frequent (multiple times daily) and for diagnosing sinus node dysfunction (sinus node arrest, sick sinus syndrome) or intermittent AV block. It can also be useful to assess the adequacy of ventricular rate control in a patient with atrial fibrillation.

Event Monitors

Event monitors, also known as loop recorders, are designed to record intermittent episodes during long periods (weeks to months) and are thus useful for patients with less frequent symptoms. The system records the ECG into a loop buffer that is continuously updated and overwritten. The duration of memory varies from a few seconds to a few minutes and is usually programmable. When activated, the information is "locked" into memory and continues to record forward for a preprogrammed amount of time. Newer systems allow both patient-activated (when symptoms occur) and event-triggered (when the heart rate is above or below a preset threshold) recording. Some recorders have algorithms to detect and record atrial fibrillation automatically, regardless of the heart rate. After episodes have been recorded, the patient transmits the recording over the telephone to centralized receivers. Newer systems use cell-phone technology to transmit the data automatically. Some event monitors require leads similar to Holter monitors, whereas others are worn on the wrist or are put into small credit card–sized devices that are placed on the chest during symptoms. The latter type is useful only in patients whose symptoms last for several minutes and who do not have syncope.

Implantable Loop Recorders

Implantable loop recorders are small devices with integrated leads that are implanted in a small subcutaneous pocket during a simple surgery, usually performed in the electrophysiology laboratory. They function similarly to event recorders in terms of recording ECGs. Patients can activate the device with a small transmitter, or the device can be autotriggered on the basis of preprogrammed heart rates. The device can be interrogated by a computer, similar to the way pacemakers are interrogated to program the device's parameters and to retrieve ECGs that have been recorded. In patients with recurrent, difficult-to-diagnose syncope, an implantable loop recorder is better than the combination of tilt testing, an external loop recorder, and electrophysiologic testing.

Tilt Table Testing

Tilt table testing is used to confirm the diagnosis of neurocardiogenic syncope. The test involves continuous heart rate and blood pressure monitoring during head-up tilting. After baseline measurements in the supine position, the patient is tilted head-up at 60 to 80 degrees for 60 minutes. Some laboratories use isoproterenol or nitroglycerin as additional provocation. A positive result is a sudden and precipitous fall in blood pressure and heart

rate, with concurrent reproducibility of symptoms (syncope). Because there is an appreciable false-positive rate, the test is best used as a confirmatory test in patients with a history suggestive of neurocardiogenic syncope or in patients with syncope in whom structural heart disease and other causes of syncope have been excluded.

Electrophysiologic Studies

Electrophysiologic studies involve the placement of several transvenous catheters in the heart to make temporary measurements of intracardiac electrograms and to perform pacing. Electrophysiologic studies are useful to identify the precise mechanism of tachyarrhythmias and are a necessary prelude to curative ablation (Chapter 66). Most arrhythmias, especially those with re-entrant mechanisms, can be readily induced during electrophysiologic studies. In addition, the existence and characteristics of accessory AV pathways (i.e., those responsible for Wolff-Parkinson-White syndrome or other re-entrant tachyarrhythmias) can be readily assessed by an electrophysiologic study. In patients with previous myocardial infarction, electrophysiologic studies are useful in determining the existence of a substrate for ventricular arrhythmias (Chapter 65), which may be treated with ablation or implantable defibrillators (Chapter 66). Electrophysiologic studies are also useful to determine the integrity of the conduction system and the precise mechanism of bradyarrhythmias that may be causing syncope. Therefore, electrophysiologic studies are indicated in patients with documented or suspected tachyarrhythmias as a prelude to curative ablation in patients with documented or suspected supraventricular tachycardia or idiopathic ventricular tachycardia; in patients with a previous myocardial infarction and syncope, presyncope, or palpitations to exclude ventricular tachycardia; and in patients with severe or prolonged symptoms and no apparent diagnosis by history or ambulatory monitoring, especially in the setting of an abnormal ECG.

Other Tests

Echocardiography

Echocardiography (Chapter 55) can be useful to ensure that a patient does not have underlying structural heart disease, which can be an important prognostic factor in patients with ventricular tachycardia or syncope. Echocardiography should be performed in patients who present with syncope that is not obviously neurocardiogenic to ensure that there is no valvular or myocardial cause.

Exercise Testing

Exercise testing (Chapters 51 and 71) can be useful to assess arrhythmias, particularly in patients whose symptoms are exercise related. Exercise testing can also be useful in the evaluation of patients with bradyarrhythmias to diagnose chronotropic incompetence, and it can differentiate AV block due to autonomic tone (improves with exercise) from intrinsic conduction disease (generally worsens with an increasing rate).

Neurologic Testing

Routine electroencephalography (Chapter 396) is not helpful because a single study may be normal, even in epileptic patients. Structural brain diseases rarely cause episodic loss of consciousness, and routine brain imaging studies are indicated only in patients with focal neurologic findings. Carotid Doppler (Chapter 407) studies can document stenosis, but unconsciousness requires bihemispheric dysfunction. Transcranial Doppler or magnetic resonance angiography of the basilar artery is indicated only in patients with symptoms suggestive of brain stem ischemia.

TREATMENT

Rx

Treatment of syncope depends on the underlying cause.[3,4] Proximate to the syncopal episode, hospital admission (e.g., observation in a chest pain unit, syncope unit, or the equivalent) is recommended when the cause of syncope is unclear, especially in elderly patients, otherwise fragile or worrisome patients, or those suspected of having a cardiac or cerebrovascular cause, or if the syncope resulted in significant injury (see Table 62-4). Patients at highest risk have a systolic blood pressure below 90 mm Hg, a history of myocardial infarction or heart failure, a complaint of shortness of breath, an abnormal initial ECG, or a hematocrit less than 30%.

Until the cause of the syncope is determined and treated, patients should be instructed to avoid situations that may cause injury as a result of the syncope, especially if there is no prodrome and episodes are frequent. Careful consideration should be given to driving restrictions, which may be mandatory depending on local laws, and restrictions on dangerous work-related activity (e.g., for pilots, heavy machine operators, bus drivers) until definitive therapy is given.

In patients with a cardiac cause of syncope, targeted treatments include valve replacement for aortic stenosis (Chapter 75); medications for hypertrophic cardiomyopathy (Chapter 60); a pacemaker for bradyarrhythmias (Chapters 64 and 66); cardioversion, an implantable cardioverter-defibrillator, ablation, or medications for tachyarrhythmias (Table 62-5; Chapters 63 through 65); and fluid repletion for orthostatic hypotension.

In patients with neurocardiogenic syncope, behavioral guidance should encourage an increased intake of fluid and salt, as well as the avoidance of situations that precipitate symptoms. Patients should also be taught how to tense their arms and legs and grip their hands during prodromal symptoms to increase peripheral resistance and systemic blood pressure.■ If neurocardiogenic syncope recurs despite education and lifestyle changes, fludrocortisone (0.1 mg/day, starting dose) can expand intravascular volume but has not been proved to prevent syncope. Midodrine (usually 5-10 mg three times daily), an α_1-receptor agonist and vasoconstrictor, has shown potential benefit,■ but other α-agonists have not. Paroxetine (20 mg/day), a selective serotonin re-uptake inhibitor, reduced recurrent neurocardiogenic syncope in one trial of very symptomatic patients but otherwise has been disappointing.■ In randomized trials, β-blockers have not been useful. Pacemakers reduce recurrent neurocardiogenic syncope in select patients with primarily a cardioinhibitory component or severely asystolic neutrally mediated syncope.■■ For example, in patients older than 40 years with frequent syncopal episodes (at least three episodes in 2 years) and demonstrated bradycardia (asystole or AV block) during an event (≥3 seconds during a syncopal episode or ≥6 seconds during a presyncopal episode), dual-chamber pacing reduces syncope by 32%.

TABLE 62-5 ARRHYTHMIC CAUSES OF PALPITATIONS AND SYNCOPE

ETIOLOGY	SPECIFIC ARRHYTHMIA	SYMPTOMS			Treatment	Comments
		Palpitations	*Dizziness*	*Syncope*		
BRADYARRHYTHMIAS						
Sinus node dysfunction	Sinus bradycardia	No	Occasional	Rare	Pacemaker (if symptoms)	Can be seen in association with neurocardiogenic syncope
	Sinus arrest	Occasional	Yes	Occasional	Pacemaker	Pause >3 sec
	Sick sinus syndrome	Occasional	Yes	Occasional	Pacemaker	
AV nodal disease	First-degree AV block	No	No	No	None	
	Type I second-degree AV block	Occasional	No	No	None	Can be seen in association with neurocardiogenic syncope
	Type II second-degree AV block	Occasional	Rare	No	Pacemaker if severe	Can progress to complete heart block
	Third-degree AV block	Yes	Yes	Yes	Pacemaker	
Tachy-brady syndrome		Yes	Yes	Occasional	Treat tachycardia if possible Pacemaker	Can also be manifestation of sick sinus syndrome

TABLE 62-5 ARRHYTHMIC CAUSES OF PALPITATIONS AND SYNCOPE—cont'd

ETIOLOGY	SPECIFIC ARRHYTHMIA	SYMPTOMS			Treatment	Comments
		Palpitations	*Dizziness*	*Syncope*		
TACHYARRHYTHMIAS						
SVT	Atrial tachycardia	Yes	Occasional	Rare	Ablation β-Blockers (e.g., metoprolol, atenolol)* Calcium-channel blockers (e.g., diltiazem)*	
	Atrial flutter	Yes	Occasional	Rare	Ablation Antiarrhythmic drugs (e.g., amiodarone)* Cardioversion (acute episode)	Often difficult to control rate
	Atrial fibrillation	Yes	Occasional	Rare	Ventricular rate control Warfarin Antiarrhythmic drugs (e.g., amiodarone)* Cardioversion (acute episode) Ablation	
	AV nodal re-entrant tachycardia	Yes	Yes	Rare	Ablation β-Blockers (e.g., metoprolol, atenolol)* Calcium-channel blockers (e.g., diltiazem)*	
	AV re-entrant tachycardia (WPW)	Yes	Yes	Rare	Ablation Antiarrhythmic drugs*	
VT	Idiopathic (RV outflow tract, fascicular)	Yes	Yes	Occasional	Ablation	Absence of structural heart disease Low risk for sudden death
	VT secondary to CAD, cardiomyopathy	Yes	Yes	Yes	ICD Amiodarone (400 mg qd)* Ablation	Increased incidence of sudden death
	Bundle branch re-entry	Yes	Yes	Yes	Ablation	Usually in the setting of LV dysfunction and baseline intraventricular conduction delay
	Genetic syndromes (e.g., long QT syndrome, Brugada, arrhythmic right ventricular dysplasia)	Occasional	Yes	Yes	ICD	Not always a clear family history Increased incidence of sudden death
Ectopy	PACs	Occasional	No	No	None β-Blockers (e.g., atenolol, metoprolol) if symptomatic*	
	PVCs	Occasional	No	No	None β-Blockers (e.g., atenolol, metoprolol) if symptomatic*	Benign in absence of structural heart disease
NEUROCARDIOGENIC SYNCOPE		No	Yes	Yes	Behavioral (hydration, avoid triggers, abort episodes) Midodrine (10 mg tid)	

*See Table 64-5 for drug doses.
AV = atrioventricular; CAD = coronary artery disease; ICD = implantable cardioverter-defibrillator; LV = left ventricle; PACs = premature atrial contractions; PVCs = premature ventricular contractions; RV = right ventricle; SVT = supraventricular tachycardia; VT = ventricular tachycardia; WPW = Wolff-Parkinson-White syndrome.

PROGNOSIS

One syncopal event predicts a substantial risk for recurrent syncope. Although syncope itself does not appear to increase the risk for death, patients with cardiac or cerebrovascular causes have higher mortality rates than patients with definable noncardiac causes or those without a definable cause. For otherwise healthy individuals discharged with a primary diagnosis of syncope, the subsequent risk for all-cause mortality is increased by 6%, with stroke increased by 35% and a cardiovascular hospitalization by 75%.[5] Among patients who come to an emergency department, the overall death rate is about 7.5% at 1 year. In patients with inherited arrhythmias, such as the long QT syndrome (Chapter 65), syncope worsens prognosis. Compared with other patients with supraventricular tachycardia, syncope per se does not increase mortality but does increase the likelihood of needing medical or ablation therapy (Chapter 66).

In patients with arrhythmias, a key issue is whether they should be allowed to drive a motor vehicle. Consensus recommendations vary depending on the arrhythmia and its treatment (Table 62-6).[6]

TABLE 62-6 DRIVING IN PATIENTS WITH ARRHYTHMIAS

CARDIOVASCULAR DISORDER	DRIVING RESTRICTION
SVT: atrial fibrillation, atrial flutter, narrow complex SVT, wide complex SVT	No driving if symptomatic Can drive if asymptomatic for 1 month (3-6 months for wide complex SVT)
VT, VF	No driving for 6 months
Bradyarrhythmias	No restriction if asymptomatic; no driving if syncope occurs
After successful catheter ablation	Can drive after recovery from procedure
After pacemaker implantation	No driving for 1 week (4 weeks for commercial drivers)
After ICD implantation	No driving for 6 months (barred from commercial driving)

ICD = implantable cardioverter-defibrillator; SVT = supraventricular tachycardia; VF = ventricular fibrillator; VT = ventricular tachycardia.
Adapted from Banning AS, Ng GA. Driving and arrhythmia: a review of scientific basis for international guidelines. *Eur Heart J.* 2013;34:236-244.

Grade A References

A1. Krahn AD, Klein GJ, Yee R, et al. Randomized assessment of syncope trial: conventional diagnostic testing versus a prolonged monitoring strategy. *Circulation*. 2001;104:46-54.

A2. van Dijk N, Quartieri F, Blanc JJ, et al. Effectiveness of physical counterpressure maneuvers in preventing vasovagal syncope: the Physical Counterpressure Manoeuvres Trial (PC-Trial). *J Am Coll Cardiol*. 2006;48:1652-1657.

A3. Izcovich A, Gonzalez Malla C, Manzotti M, et al. Midodrine for orthostatic hypotension and recurrent reflex syncope: A systematic review. *Neurology*. 2014;83:1170-1177.

A4. Moya A, Sutton R, Ammirati F, et al. Guidelines for the diagnosis and management of syncope (version 2009): the Task Force for the Diagnosis and Management of Syncope of the European Society of Cardiology (ESC). *Eur Heart J*. 2009;30:2631-2671.

A5. Brignole M, Menozzi C, Moya A, et al. Pacemaker therapy in patients with neurally mediated syncope and documented asystole. Third International Study on Syncope of Uncertain Etiology (ISSUE-3): a randomized trial. *Circulation*. 2012;125:2566-2571.

A6. Connelly SJ, Sheldon R, Thorpe KE, et al. Pacemaker therapy for prevention of syncope in patients with recurrent severe vasovagal syncope. Second Vasovagal Pacemaker Study (VPS II): a randomized trial. *JAMA*. 2003;289:2224-2229.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hatoum T, Sheldon R. A practical approach to investigation of syncope. *Can J Cardiol*. 2014; 30:671-674.
2. Klein KM, Bromhead CJ, Smith KR, et al. Autosomal dominant vasovagal syncope: clinical features and linkage to chromosome 15q26. *Neurology*. 2013;80:1485-1493.
3. Cooper PN, Westby M, Pitcher DW, et al. Synopsis of the National Institute for Health and Clinical Excellence Guideline for management of transient loss of consciousness. *Ann Intern Med*. 2011; 155:543-549.
4. Saklani P, Krahn A, Klein G. Syncope. *Circulation*. 2013;127:1330-1339.
5. Ruwald MH, Hansen ML, Lamberts M, et al. Prognosis among healthy individuals discharged with a primary diagnosis of syncope. *J Am Coll Cardiol*. 2013;61:325-332.
6. Banning AS, Ng GA. Driving and arrhythmia: a review of scientific basis for international guidelines. *Eur Heart J*. 2013;34:236-244.

REVIEW QUESTIONS

1. The most import factor in determining the prognosis in patients with suspected arrhythmias is which of the following?

A. The existence of underlying structural heart disease (e.g., coronary artery disease, prior myocardial infarction, or ventricular dysfunction)
B. The duration of the symptoms
C. The frequency of the symptoms
D. Family history of palpitations
E. Responsiveness to Valsalva maneuvers

Answer: A See Clinical Manifestations. In general, the most important factor determining the prognosis of arrhythmias is whether they occur in the setting of underlying heart disease. Specifically, patients with heart failure, prior myocardial infarction, aortic stenosis, and hypertrophic cardiomyopathy have a much higher likelihood of having ventricular tachycardia or ventricular fibrillation that may lead to a cardiac arrest than do patients with a normal heart. In addition, other nonlethal arrhythmias, such as atrial fibrillation, also are more prevalent in these patients. Evaluation and treatment of symptoms of a suspected arrhythmia are therefore different in patients with and without structural heart disease.

2. A 25-year-old patient without any known prior cardiac disease presents with a 3-month history of intermittent palpitations, which occur every few weeks, persist for up to about 5 minutes, and spontaneously resolve. On occasion they are accompanied by dizziness, and on a recent episode he had a syncopal episode. What would be the most appropriate *first* test to obtain for him in the evaluation of his symptoms?

A. A cardiac catheterization
B. A head-up tilt table test
C. A 12-lead electrocardiogram (ECG)
D. A 24-hour Holter monitor
E. A complete blood count and electrolyte panel

Answer: C See Diagnostic Tests: Electrocardiography. Even though the patient is not having palpitations at this time, there may be clues on the ECG to indicate their cause and to guide further evaluation. For example, ventricular preexcitation makes supraventricular tachycardia (SVT) and Wolff-Parkinson-White syndrome the likely diagnosis. In addition, other syndromes such as the long QT syndrome, hypertrophic cardiomyopathy, or other abnormalities may be seen on a baseline ECG. The evaluation would not end with an ECG and would likely require further ambulatory monitoring. However, because the symptoms do not occur every day, a 24-hour Holter monitor is unlikely to capture the rhythm during his symptoms.

3. What type of ambulatory monitoring would be most appropriate for a patient who has intermittent palpitations that occur every 1 to 2 weeks?

A. A 48-hour Holter monitor
B. An implantable loop recorder
C. Ambulatory blood pressure monitor
D. A wearable event monitor
E. A wearable defibrillator vest

Answer: D See Diagnostic Tests: Ambulatory Monitoring. Because the symptoms occur less frequently than daily, it is very unlikely that a 24- or 48-hour Holter monitor will detect an event. A wearable event monitor, which can be worn for 3 to 4 weeks, is likely to capture an event. Although an implantable loop recorder would likely capture an event, it is inappropriate in this setting because of its expense and invasive nature. An ambulatory blood pressure monitor is unable to detect an arrhythmia. A wearable defibrillator vest is not necessary because the patient has not had a cardiac arrest. Assuming a baseline ECG was obtained as part of the initial evaluation (see previous question), an ECG following an episode when the patient is asymptomatic is unlikely to be helpful.

4. When a patient presents in a narrow complex tachycardia, which of the following is the *most* useful diagnostic test?

A. Carotid sinus massage
B. A 12-lead ECG of the tachycardia
C. Administration of adenosine while recording a continuous 12-lead ECG
D. An echocardiogram while still in the tachycardia
E. A nuclear stress test

Answer: C See Diagnostic Tests. Adenosine is useful for narrow complex tachycardias both diagnostically and therapeutically. For example, if the tachycardia continues with flutter waves or P waves during atrioventricular (AV) block, then a diagnosis of atrial flutter or atrial tachycardia is established. Likewise, the termination of the tachycardia with a P wave that is not conducted to the ventricle is much more likely to be an AV or AV nodal re-entrant tachycardia. In addition, adenosine after termination of the tachycardia can bring out ventricular preexcitation suggestive of Wolff-Parkinson-White syndrome that may be otherwise subtle on the baseline ECG. Comparison of the narrow complex tachycardia to the baseline ECG can also be very useful in identifying subtle differences in the QRS complexes that can be important clues to the mechanism of tachycardia. For example, in orthodromic AV re-entrant circus tachycardia (with the antegrade limb being the AV node and retrograde limb being an AV accessory pathway), preexcitation will be lost during tachycardia but may be present during normal sinus rhythm (though with concealed accessory pathways in which there is no antegrade conduction, there will not ever be ventricular preexcitation). Carotid massage can also terminate SVTs but is less useful diagnostically if not performed during simultaneous 12-lead ECG. A tilt table test is not a useful test in patients with palpitations or clearly documented tachycardia, even with syncope; it is useful only in neurocardiogenic syncope. A 12-lead ECG during the tachycardia is very useful and should be obtained whenever possible; however, more information will be obtained from a 12-lead ECG during adenosine administration. An echocardiogram might be useful in excluding valvular or structural heart disease, depending on the history and physical examination, but it does not need to be performed during the tachycardia. Similarly, a nuclear stress test is indicated only if there is suspicion of coronary artery disease, which is rarely the primary cause of an SVT.

5. What features from the history are most important in distinguishing neurocardiogenic syncope from seizure disorder?

A. Post-episode confusion lasting more than 5 minutes
B. Jerking movements of limbs during the episode
C. Triggered by pain
D. A prodrome of nausea and sweatiness
E. A, C, and D

Answer: E See Table 62-2 and Neurocardiogenic Syncope and Related Syndromes. Because neurocardiogenic syncope commonly presents in young, otherwise healthy individuals, the differential diagnosis often involves a new seizure disorder. Several important historical features and the appearance of the patient during the syncopal episode (when observed by a bystander) can be very useful in distinguishing neurocardiogenic syncope from a seizure disorder. In seizure disorders, a postictal confusion period lasting more than 5 minutes is common; in neurocardiogenic syncope, by comparison, disorientation rapidly resolves within a 1 or 2 minutes upon regaining consciousness. In a patient who has had a cardiac arrest (distinct from syncope), the period of confusion can last very long if the patient has suffered anoxic brain injury; however, anoxic brain injury is very uncommon in neurocardiogenic syncope. Jerking movements of the limbs can be confused with myoclonic movements that are common in neurocardiogenic syncope, so these movements are not helpful in distinguishing between seizure disorder and neurocardiogenic syncope. Neurocardiogenic syncope is often triggered by pain or prolonged standing; commonly one can elicit of history of syncope or presyncope with abdominal cramping, phlebotomy, or other painful stimuli. Seizures are not usually triggered by pain. Although seizure disorders can sometimes be associated with prodromes, they commonly involve sensory auras. Neurocardiogenic syncope often is preceded by reproducible prodromes of nausea and sweatiness, which are triggered by the heightened parasympathetic tone that causes the acute drop in blood pressure and heart rate and which can persist after the episode.

63

APPROACH TO CARDIAC ARREST AND LIFE-THREATENING ARRHYTHMIAS

ROBERT J. MYERBURG

Sudden cardiac arrest is characterized by an abrupt loss of consciousness because of absence of blood flow owing to loss of cardiac pumping action. If not treated promptly, it will lead to central nervous system injury or death within minutes. Sudden cardiac arrest is often forewarned by a change in cardiovascular status, as indicated by the onset or worsening of symptoms related to transient arrhythmias, such as palpitations, lightheadedness, or near-syncope or syncope (Chapter 62). Other forewarnings may include new or worsening chest pain, dyspnea, or weakness. In individual patients, however, these warning symptoms have limited sensitivity and predictive power for sudden cardiac arrest because they also predict the acute coronary syndrome (Chapter 72) and acute myocardial infarction (MI) (Chapter 73).

EPIDEMIOLOGY

Patients with advanced ischemic and nonischemic cardiomyopathies (Chapter 60), heart failure (Chapters 58 and 59), and certain acquired and inherited arrhythmia syndromes (Chapters 64 and 65) have an increased risk for sudden cardiac arrest. However, most sudden cardiac arrests occur either as a first cardiac event in an apparently healthy individual with unrecognized disease or in patients known to be low risk. The incidence in the general population over the age of 35 years ranges from 1 to 2 per 1000 per year. Among adolescents and young adults, it is 1 per 100,000 per year. In addition, competitive and high-intensity recreational athletes have a low but finite increase in risk for sudden cardiac arrest during training or competition,[1] with estimates ranging from 1 in 75,000 to 1 in 200,000. These risks are higher in males and in association with specific sports, such as basketball and football in the United States and cycling, jogging, and soccer in Europe. In adolescents and younger adults in the United States, hypertrophic cardiomyopathy (Chapter 60) is the most commonly identified structural cause in competitive athletes;[2] but beyond the age of 30 to 35 years, coronary artery disease is more common (Chapter 73).[3] Offspring in families in which sudden cardiac arrest was the initial manifestation of heart disease are themselves at high risk for sudden cardiac arrest as the initial manifestation of heart disease, thereby emphasizing the importance of a careful family history for assessing risk.[4]

PATHOBIOLOGY

In the past, ventricular fibrillation (VF) and pulseless ventricular tachycardia (VT) were the most common electrical mechanisms of sudden cardiac arrest (Chapter 65), largely in association with acute MI and with chronic ischemic and nonischemic cardiomyopathies. Over the past two decades, however, asystole and pulseless electrical activity have now become the first recorded rhythm in the majority of both in-hospital and out-of-hospital cases. These rhythms may also follow deterioration or active termination of prolonged VF by electrical cardioversion. Pulseless electrical activity is defined as primary when it is the initial rhythm noted in patients with predisposing cardiac disorders and as secondary when it occurs in the setting of noncardiac predisposing factors, such as hypoxia, metabolic disorders, massive pulmonary embolism, or blood loss.[5]

Premature ventricular contractions (PVCs) and short runs of nonsustained ventricular tachycardia may forewarn a long-term risk for sudden cardiac arrest, primarily when associated with advanced structural heart disease, but there is no evidence that suppression of chronic PVCs is protective. In contrast, sustained wide QRS tachycardias are of greater concern and should be considered of ventricular origin, because of their potentially high immediate risk, until determined otherwise (Chapter 62). Most wide-QRS tachycardias are initially approached as a medical urgency or emergency, whereas most narrow-QRS tachycardias of supraventricular origin are approached with less urgency (Chapters 64 and 65).

CLINICAL MANIFESTATIONS

The absence of a pulse in conjunction with no respiratory efforts or only gasping or agonal respirations is diagnostic of cardiac arrest. Although the absence of a carotid or femoral pulse is a primary diagnostic criterion for the health care professional, palpation for a pulse is no longer recommended for lay responders. The absence of respiratory efforts or severe stridor with *persistence of a pulse* suggests a primary respiratory arrest that may lead to cardiac arrest in a short time; skin color may be pale or intensely cyanotic. In the latter circumstance, initial efforts should include oropharyngeal exploration in search of a foreign body and the Heimlich maneuver, which entails wrapping the arms around the victim from the back and delivering a sharp thrust to the upper part of the abdomen with a closed fist, particularly in a setting in which aspiration is likely (e.g., collapse in a restaurant).

DIAGNOSIS

Distinguishing Supraventricular from Ventricular Tachycardias

Differentiating supraventricular tachycardia (SVT) (Chapter 64) with either narrow or wide QRS complexes from VT is an important clinical challenge for both risk prediction and therapy. Although it is generally assumed that narrow-QRS tachycardias are SVTs, VT occasionally has a narrow QRS complex on a one- or two-lead rhythm strip, thereby mimicking SVT. Whenever possible, the classification of a tachycardia as SVT or VT should be based on a 12-lead electrocardiogram (ECG). However, a standard ECG will not always suffice because patients with intraventricular conduction abnormalities (such as a left or right bundle branch block) will have wide-complex tachycardias during SVTs, usually with a QRS vector similar to that seen in normal sinus rhythm. In addition, when an SVT is very rapid, a functional bundle branch block may transiently prolong the QRS duration and shift the QRS axis. In both examples, the wide QRS may mimic VT, and it may be necessary to perform an electrophysiologic study to determine the diagnosis (Chapter 62).

When wide-complex SVT is suspected clinically, transient vagal stimulation by carotid sinus massage or an atrioventricular nodal blocking agent, such as intravenous adenosine (see Table 64-5), may be useful for transiently slowing the ventricular rate or terminating an SVT. A continuous rhythm strip should be recorded during administration of adenosine or performance of vagal maneuvers because characterizing transient changes on a monitor screen may be unreliable. Intravenous calcium-blocking agents generally should not be used for the diagnosis or treatment of wide-QRS tachycardias, especially in the presence of structural heart disease, because of their myocardial depressant effects. The exception is when it is known with certainty that the tachycardia is an SVT in a patient with normal or near-normal left ventricular function.

Sustained VT occurs most commonly in the presence of structural heart disease and must be interpreted as a forewarning of fatal arrhythmia in that setting. It is characterized by QRS complexes that are usually longer than 0.12 second, with a mean vector that is markedly different from the QRS vector of normally conducted impulses. The rate of most VTs is between 140 and 200 impulses per minute, but rates may be slower or faster. VT may be electrically stable (such as monomorphic VT patterns at relatively slow rates; Fig. 63-1A) or unstable (such as polymorphic VTs or monomorphic VTs at rates exceeding 190 to 200 per minute; see Fig. 63-1B) (Chapter 65).

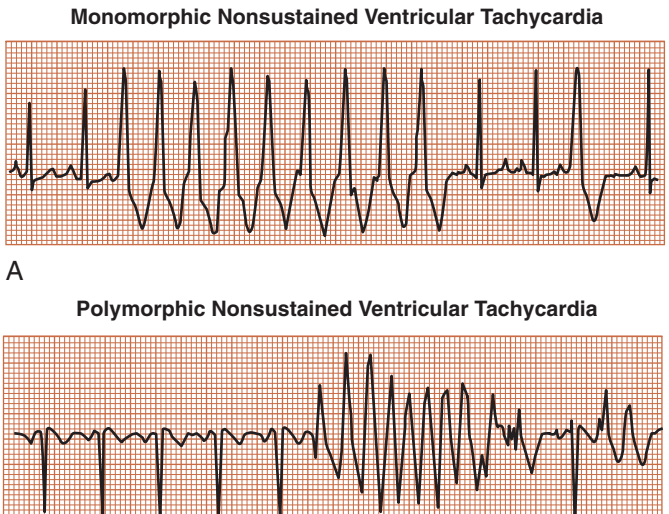

FIGURE 63-1. Nonsustained ventricular tachycardia. Monomorphic patterns (A) are characterized by a slower and more stable electrical pattern than polymorphic patterns (B). Both have long-term prognostic implications in patients with advanced structural heart disease, but monomorphic patterns tend to be more stable over the short term.

TREATMENT

The management of a patient in cardiac arrest involves artificial maintenance of blood flow to preserve viability of the central nervous system, heart, and other vital organs, while striving to restore spontaneous circulation as quickly as possible. These goals are accomplished by assessing the patient and contacting an emergency response team; initiating basic life support (BLS); early defibrillation for those with VF or pulseless VT; advanced life support (ACLS) as needed; and post–cardiac arrest care, the latter now formally incorporated into the concept of the post–cardiac arrest syndrome. In a randomized trial, allowing family presence during cardiopulmonary resuscitation (CPR) was associated with better psychological health among family members and did not interfere with medical efforts, increase stress in the health care team, or result in medicolegal conflicts.

Basic Life Support

The first action in BLS is to confirm that the collapse is the result of a cardiac arrest. After an initial evaluation for response to voice or tactile stimulation, observation for respiratory movements and skin color, and simultaneous palpation of major arteries for the presence of a pulse, the determination that a life-threatening incident is in progress should immediately prompt a call to an emergency medical rescue system (911).

After confirming the cardiac arrest, the goal of BLS is to re-establish perfusion as quickly as possible using CPR or the newly proposed concept of cardiocerebral resuscitation (see later). The previous "ABC" algorithm of basic life support (airway-breathing-compression) has been changed to "CAB" (compression-airway-breathing), based on the recognition that compression alone is the primary maneuver because the patient is better perfused by minimizing interruptions between compressions and can be harmed by excessive ventilation.[5]

A precordial thump may be attempted by a properly trained rescuer as part of an initial response, although its added benefit is questionable. The technique involves one or two blows delivered firmly to the junction of the middle and lower thirds of the sternum from a height of 8 to 10 inches. A thump should not be used in an unmonitored patient with a perceptible rapid tachycardia or without complete loss of consciousness because of concern about converting cardiac electrical activity into VF, and the effort should be abandoned if a spontaneous pulse does not appear immediately.

Prompt initiation of CPR, which can be performed by professional and paraprofessional personnel, by experienced emergency medical technicians, and by trained laypersons, is the key element for successful resuscitation. The delay between diagnosis and preparatory efforts in the initial response and institution of CPR should be minimal. If only one witness is present, the only activity that should precede BLS is telephone contact (911) of emergency personnel.

Clearing the airway includes tilting the head backward and lifting the chin, in addition to exploring the airway for foreign bodies—including dentures—and removing them. The Heimlich maneuver should be performed if there is reason to suspect a foreign body lodged in the oropharynx, as suggested by severe respiratory stridor rather than by slow agonal respirations or apnea. When the person at the scene has insufficient physical strength to perform the maneuver, mechanical dislodgement of a foreign body can sometimes be achieved by abdominal thrusts with the unconscious patient in a supine position. If there is suspicion that respiratory arrest precipitated the cardiac arrest, particularly in the presence of a mechanical airway obstruction, a second precordial thump should be delivered after the airway has been cleared.

With the head properly positioned and the oropharynx clear, mouth-to-mouth respiration can be initiated, but bystander compression-only CPR is as good as, if not better than, compression plus rescue breathing. With the exception of Heimlich maneuvers, ventilation strategies are now reserved for emergency medical responders and medical professionals, rather than bystander responders. Devices available for establishing ventilation include plastic oropharyngeal airways, esophageal obturators for establishing ventilation, a masked Ambu bag, and endotracheal tubes. Intubation is the preferred procedure, but time should not be sacrificed, even in the in-hospital setting, while awaiting an endotracheal tube or a person trained to insert it quickly and properly. Temporary support with Ambu bag ventilation is the usual method in the hospital until endotracheal intubation can be accomplished. When ventilatory support is provided by emergency responders in the out-of-hospital setting, the lungs should be inflated twice in succession after every 30 chest compressions.[5]

Circulatory support, which is the primary element of BLS, is intended to maintain blood flow until definitive steps can be taken. The rationale is based on the hypothesis that chest compression maintains an externally driven pump function by sequential emptying and filling of its chambers, with competent valves favoring the forward direction of flow. The palm of one hand is placed over the lower part of the sternum while the heel of the other rests on the dorsum of the lower hand. The sternum is then depressed with the resuscitator's arms straight at the elbows to provide a less tiring and more forceful fulcrum at the junction of the shoulders and back. With this technique, sufficient force is applied to depress the sternum at least 2 inches (>5 cm), with abrupt relaxation. The cycle is carried out at a rate of about 100 compressions per minute. In current guidelines for emergency cardiac care, the integration of respiratory and compression actions was changed to a compression-ventilation ratio of 30:2 for single responders to victims from infancy (excluding newborns) through adulthood, and for two responders to adult victims.[6] For two-rescuer CPR for infants and children, the compression-ventilation ratio is 15:2. Another recently suggested modification is the "hands-only" (cardiac-only, compression-only) technique,[5] which uses 200 successive compressions without interruption.[7] This variation, which may be more effective than compression-ventilation sequences, encourages more bystander CPR by untrained or remotely trained bystanders who lack confidence and also allays concerns about mouth-to-mouth ventilation of unknown victims in the absence of mechanical airway devices.

Intermediate Life Support: Automated External Defibrillators

Despite the temporizing benefit of BLS, time to defibrillation is the major determinant of survival. Because ACLS strategies are generally implemented by in-hospital personnel or out-of-hospital emergency medical rescue system responders, an intermediate strategy is for nonconventional first responders to use automated external defibrillators (AEDs). Referred to as public access defibrillation or lay first-responder systems, the strategy relies on devices that prompt the user to deliver a defibrillation shock when deemed appropriate by a computerized rhythm detection system in the device. The operators can be trained police officers, security guards, airline personnel, or trained (or even untrained) lay responders (Table 63-1). A number of studies have suggested improved survival rates when such strategies are deployed in public sites,[8] but an initial study of a home deployment strategy was disappointing. Further study is warranted because most out-of-hospital cardiac arrests occur at home. AED programs are not a replacement for ACLS (see later), but rather are an intermediate supplement to the BLS-ACLS sequence that is intended to attempt earlier defibrillation while awaiting the arrival of ACLS-trained emergency rescue personnel.

Advanced Cardiac Life Support

ACLS methods, other than those directly related to control of tachyarrhythmias, are guided by comprehensive protocols to aid responders over a broad expanse of clinical circumstances and mechanisms of cardiac arrest ranging from transient clinical events to end-stage multisystem disease. The general goals of ACLS are to restore a hemodynamically effective cardiac rhythm, optimize ventilation, and maintain and support the restored circulation. During ACLS, the patient's cardiac rhythm is promptly cardioverted or defibrillated as the first priority, if appropriate equipment is immediately available. If cardiac arrest has lasted for 4 to 5 minutes before the availability of a defibrillator, a short period of closed-chest cardiac compression immediately before defibrillation increases the probability of survival.

After the initial attempt to restore a hemodynamically effective rhythm, the patient is intubated and oxygenated, if needed. Electrical pacing of the heart

TABLE 63-1 AUTOMATED EXTERNAL DEFIBRILLATOR STRATEGIES FOR RAPID RESPONSE TO CARDIAC ARRESTS CAUSED BY VENTRICULAR FIBRILLATION

DEPLOYMENT	EXAMPLES	RESCUERS	ADVANTAGES	LIMITATIONS
Emergency vehicles	Police cars Fire engines Ambulances	Trained emergency personnel	Experienced users Broad deployment Objectivity	Deployment time Arrival delays Community variations
Public access sites	Public buildings Stadiums, malls Airports Airliners	Security personnel Designated rescuers Random laypersons	Population density Shorter delays Lay and emergency personnel access	Low event rates Inexperienced users Panic and confusion
Multifamily dwellings	Apartments Condominiums Hotels	Security personnel Designated rescuers Family members	Familiar locations Defined personnel Shorter delays	Infrequent use Low event rates Geographic factors
Single-family dwellings	Private homes Apartments Neighborhood "Heart Watch"	Family members	Immediate access Familiar setting	Acceptance Victim may be alone One-time user; panic

should be attempted if a severe bradyarrhythmia or asystole is present (Chapter 66). An intravenous line is established to deliver medications. After intubation, the goal of ventilation is to reverse hypoxemia and not merely to achieve a high alveolar P_{O_2}. When available, oxygen rather than room air should be used to ventilate the patient, and arterial O_2 saturation should be monitored, when possible. In the out-of-hospital setting, a face mask or an Ambu bag by means of an endotracheal tube is generally used.

Approach to Specific Arrhythmias

Tachyarrhythmic Cardiac Arrest

Slow, well-tolerated monomorphic VTs, especially in the absence of structural heart disease, can usually be treated with antiarrhythmic drugs or β-adrenergic blocking agents in some circumstances (see Table 64-5). In contrast, when rapid VT or VF is identified on a monitor or by telemetry, defibrillation should be performed immediately (Fig. 63-2).[5,6] When a reversible cause, such as an acute ischemic syndrome or electrolyte disturbance, is the mechanism, normal rhythm can be successfully restored in up to 90% of VF victims weighing up to 90 kg with a DC monophasic shock of up to 360 J, or a with a biphasic shock of up to 200 J, delivered within 2 to 3 minutes. Failure of the initial shock to restore an effective rhythm is a poor prognostic sign. Although some previous algorithms suggested a succession of monophasic shock energies from 200 to 360 J, or biphasic waveforms from 100 to 200 Joules, during a sequence of attempts to defibrillate, there is little to be gained from beginning with energies less than 300 J monophasic or less than 150 J biphasic during a cardiac arrest response.

After a single shock using a 150 or 200 J biphasic waveform or a 300 or 360 J monophasic waveform, the patient should be checked immediately for restoration of a spontaneous pulse; CPR should be continued for five cycles if a pulse remains absent. Subsequently, a second shock should be delivered, followed by epinephrine, 1 mg intravenously (IV). If a pulse is still absent, CPR is repeated for five cycles before the next shock. Epinephrine may be repeated at 3- to 5-minute intervals with defibrillator shocks in between, but high-dose epinephrine does not appear to provide added benefit. Vasopressin, 40 U given IV once, is an equally good alternative to epinephrine,[■] but the combination does not appear to be better than either one alone.[■]

SVTs can precipitate cardiac arrest in two circumstances. One is in patients with high-grade coronary artery stenoses, in whom rapid heart rates can cause myocardial ischemia because of the dependence of coronary blood flow on the diastolic interval. In this setting, the arrhythmia should be treated urgently by restoring sinus rhythm or slowing the heart rate, either by medical therapy (e.g., intravenous adenosine, β-blockers, or Ca^{2+} blockers) (Chapter 64) or by electrical direct current (DC) cardioversion (Chapter 66). The second mechanism of concern is atrial fibrillation in patients with Wolff-Parkinson-White syndrome, who may have ventricular rates greater than 300 beats per minute when the accessory pathway has a short refractory period (see Fig. 64-19). This pathophysiology can cause hypotensive VT or VF and requires prompt therapy (Chapter 64).

Pharmacotherapy for Resistant Arrhythmias

For a patient who continues in VF or pulseless VT despite multiple attempts at DC cardioversion after epinephrine, or who has recurrent episodes of VF or VT after cardioversion, electrical stability may be achieved by administering intravenous antiarrhythmic agents while continuing resuscitative efforts (see Fig. 63-2). Amiodarone (150 mg IV over a 10-minute period, followed by 1 mg/minute for up to 6 hours and 0.5 mg/minute thereafter) is the initial treatment of choice.[■] Additional bolus dosing, to a maximum of 500 mg, can be tried if the initial bolus is unsuccessful. Amiodarone need not be given as a routine to individuals who respond to initial defibrillation with a persistently stable

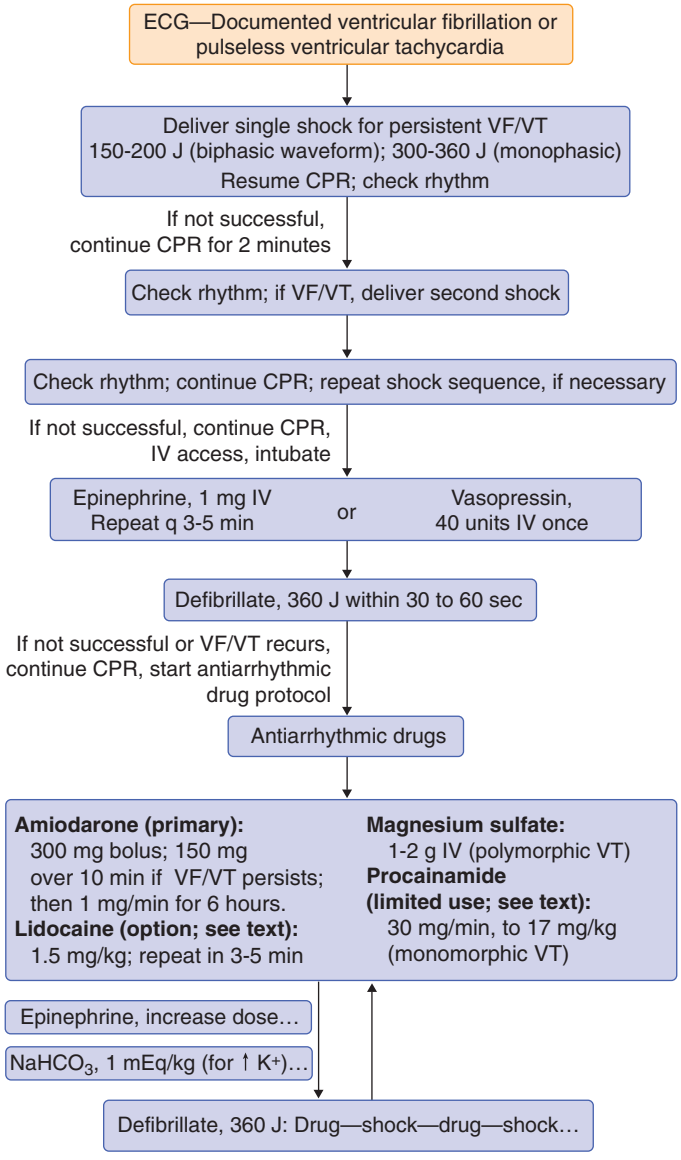

FIGURE 63-2. General algorithm for advanced cardiac life support (ACLS) response to ventricular fibrillation (VF) or pulseless ventricular tachycardia (VT). For more detail, see the ACLS guidelines in the Grade A References. *Note:* In a 2008 advisory, 200 compression-only sequences were suggested as an alternative to standard CPR cycles between shocks, and this approach is under consideration for future guidelines. CPR = cardiopulmonary resuscitation; ECG = electrocardiogram.

rhythm, but it is preferred for those who have recurrent episodes of VT or VF after initial defibrillation and oxygenation.

If there is sufficient clinical evidence that the cardiac arrest was heralded by the onset of an acute coronary syndrome, lidocaine (1.0- to 1.5-mg/kg bolus given IV, with the dose repeated in 2 minutes) may be used instead of amiodarone, or if amiodarone has failed. When acute or intermittent ischemia is not thought to be the mechanism, intravenous amiodarone is the preferred initial drug, but lidocaine may be tried if amiodarone fails. Intravenous procainamide (loading infusion of 100 mg/5 minutes to a total dose of 500 to 800 mg, followed by a continuous infusion at 2 to 5 mg/minute) is now rarely used but may be tried in those with persisting, hemodynamically unstable arrhythmias.

In patients with acute hyperkalemia as the triggering event for resistant VF, hypocalcemia, or arrest potentially caused by excess doses of calcium-blocking drugs, 10% calcium gluconate (5 to 20 mL infused at a rate of 2 to 4 mL/minute) may be helpful. Otherwise, calcium should not be used routinely during resuscitation, even though ionized Ca^{2+} levels may be low during resuscitation from cardiac arrest.

Resistant forms of polymorphic VT (torsades de pointes), rapid monomorphic VT, ventricular flutter (rate > 260/minute), or resistant VF may respond to $MgSO_4$ (1 to 2 g given IV over a 1- to 2-minute period) or to β-blocker therapy (propranolol, 1-mg boluses IV to a total dose of up to 15 to 20 mg; or metoprolol, 5 mg IV, up to 20 mg). $MgSO_4$ is specifically indicated for polymorphic VTs due to inherited or acquired (drug-induced) long QT patterns (Chapter 65). This VT pattern also occurs with marked hypokalemia, so 20 mEq/hour of intravenous potassium chloride should be included in the treatment of patients who have a serum K^+ of less than 3 mEq/L and whose polymorphic VT is resistant to other therapies. However, hypokalemia also may follow the acid-base and electrolyte shifts associated with prolonged arrests and should not be considered a primary cause of the cardiac arrest in that circumstance.

Asystole, Bradyarrhythmias, and Pulseless Electrical Activity

The approach to a patient with bradyarrhythmia, asystolic arrest, or pulseless electrical activity differs from the approach to patients with tachyarrhythmic events (VT/VF).[9] Effective CPR is critical because no electrical strategies are effective for restoring circulation. As soon as this form of cardiac arrest is recognized, efforts should focus on continuing CPR, intubation, and establishing intravenous access. Possible reversible causes, including hypovolemia, hypoxia, cardiac tamponade, tension pneumothorax, preexisting acidosis, drug overdose, hypothermia, and hyperkalemia, must be identified and treated immediately (Fig. 63-3). Respiratory causes of pulseless electrical activity or asystole may respond promptly to appropriate interventions, as do tamponade and hypovolemic causes. Epinephrine (1.0 mg IV every 3 to 5 minutes) or isoproterenol (up to 15 to 20 µg/minute IV), which are commonly used in an attempt to elicit spontaneous electrical activity or increase the rate of a bradycardia, have only limited success. In one observational study, prehospital epinephrine increased the chance of return of spontaneous circulation before hospital arrival but decreased the chance of survival and good functional outcomes 1 month after the event.[10] In a randomized trial of patients with cardiac arrest requiring vasopressors, combined vasopressin-epinephrine and methylprednisolone during CPR and stress-dose hydrocortisone in post-resuscitation shock improved survival to hospital discharge with favorable neurologic status compared with epinephrine/saline placebo.[■] In the absence of an intravenous line, epinephrine (1 mg, i.e., 10 mL of a 1:10,000 solution) may be given by the intracardiac route, but there is danger of coronary or myocardial laceration. Sodium bicarbonate, 1 mEq/kg, may be tried for known or strongly suspected preexisting hyperkalemia or bicarbonate-responsive acidosis but is no longer recommended for routine use. Atropine is no longer recommended for initial management of bradyarrhythmic cardiac arrests because of lack of efficacy.

External pacing (Chapter 66) should be tried for out-of-hospital bradycardic or asystolic arrest, although existing data suggest little influence on outcome. In the hospital setting, external pacing is generally used during the initial response to a bradycardic or asystolic arrest, but it should be superseded by transvenous pacing if the arrest is prolonged, if continuous pacing is needed, or if the external device fails to pace. Unfortunately, an *asystolic* patient continues to have a very poor prognosis despite available techniques.

Post-Resuscitation Care

After return of spontaneous circulation, particularly after a prolonged resuscitation, attention shifts to the elements of injury caused by cardiac arrest. The four components of the post–cardiac arrest syndrome include brain injury, myocardial dysfunction, systemic ischemia-reperfusion responses, and control of persistent precipitating factors.[4] The therapeutic goal is to maintain a stable electrical, hemodynamic, and central nervous system status. Specific therapy is determined by the clinical circumstances.[11] The most pressing issue is the presence of anoxic encephalopathy, which is a strong predictor of in-hospital death and post-arrest disability.[12] To prevent post-arrest encephalopathy, therapeutic hypothermia to 33°C or 34°C is no better than cooling to 36°C.[■■]

During or after therapy targeted to restoration of an electrically stable cardiac rhythm, the patient's general metabolic state should be addressed by improving oxygenation and reversing acidosis. Intravenous sodium bicarbonate (1 mEq/kg), with up to 50% of this dose repeated every 10 to 15 minutes during the course of CPR, is recommended for patients with known or suspected preexisting bicarbonate-responsive causes of acidosis, for certain drug overdoses (Chapter 110), and after prolonged and unsuccessful attempts at resuscitation. Caution must be exercised, however, because excessive quantities of sodium bicarbonate can be deleterious by causing alkalosis, hypernatremia, and hyperosmolality. When possible, arterial pH, P_{O_2}, and P_{CO_2} should be monitored during the resuscitation. Myocardial injury (Chapter 73) and hemodynamic dysfunction (Chapter 107) are managed by standard techniques.

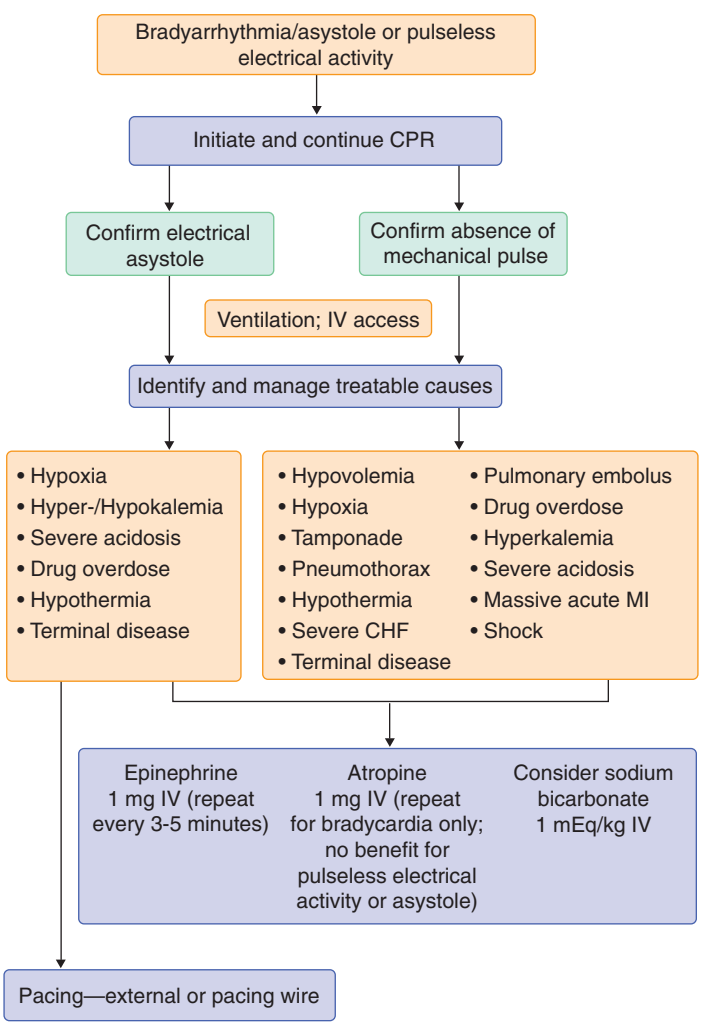

FIGURE 63-3. General algorithm for advanced cardiac life support response to bradycardic or asystolic cardiac arrest or pulseless electrical activity. For more detail, see the Grade A References. CHF = congestive heart failure; CPR = cardiopulmonary resuscitation; MI = myocardial infarction.

PROGNOSIS

The probability of survival after a prompt intervention is about 25% to 30% for VF/VT compared with about 15% for pulseless electrical activity and less than 5% for asystole.[13,14] In the hospital, the probability of survival is determined by the specific patient category (acute syndromes better than end-stage diseases), the mechanism of cardiac arrest (better for tachyarrhythmias than for bradyarrhythmias, asystole, or pulseless electrical activity), and the hospital site (better in intensive care units or other monitored settings than on an unmonitored general care unit). Immediate defibrillation in highly protected settings, such as a cardiac catheterization laboratory where response times of less than 60 seconds are the norm, is associated with greater than a 90% survival rate for VF in the absence of pathophysiologic conditions that tend to perpetuate the potentially fatal arrhythmia. In many acute care settings, including patients with acute coronary syndromes (Chapters 72 and 73), outcomes also can be excellent. For other in-hospital settings and most out-of-hospital settings, the absolute number and proportion of survivors

remain low, except in unique out-of-hospital settings that can provide an extraordinarily rapid response time to victims in VF or VT.

Survival rates also vary considerably based on the response time and the site of the arrest, with some public locations achieving survival rates of 50% or higher. If 3 to 4 minutes have elapsed from the onset of cardiac arrest to attempted defibrillation, the survival probability falls below 50% in most in-hospital and out-of-hospital circumstances. Survival rates continue to fall rapidly thereafter, decreasing to 25% or less by 4 to 6 minutes and to less than 10% by 10 minutes. Although immediate defibrillation is the preferred method within the first few minutes after the onset of cardiac arrest, a brief period of CPR to provide oxygenation of the victim improves survivability when the time to defibrillation exceeds 4 to 5 minutes.

Long-term prognosis after discharge is determined by a number of factors, including pre-event ventricular function, history of heart failure, and the severity of residual neurologic injury. Patients whose cardiac arrest is due to controllable transient ischemia or electrolyte disturbances, who have preserved left ventricular function, and who have minimal if any neurologic defects, have a good prognosis. In contrast, patients with advanced heart disease have annual mortality rates that range from 10 to 50%. Survivors of a cardiac arrest that is not due to transient factors remain at high risk for recurrent cardiac arrest and sudden cardiac death. In such patients, an implantable defibrillator (Chapter 66) achieves a relative risk reduction of 25 to 30%, with absolute risk decreasing from 21 to 25% to 15 to 18% over a 2-year follow-up.

Grade A References

A1. Jabre P, Belpomme V, Azoulay E, et al. Family presence during cardiopulmonary resuscitation. *N Engl J Med*. 2013;368:1008-1018.

A2. Rea TD, Fahrenbruch C, Culley L, et al. CPR with chest compression alone or with rescue breathing. *N Engl J Med*. 2010;363:426-433.

A3. Wik L, Hansen TB, Fylling F, et al. Delaying defibrillation to give basic cardiopulmonary resuscitation to patients with out-of-hospital ventricular fibrillation: a randomized trial. *JAMA*. 2003;289:1389-1395.

A4. Aung K, Htay T. Vasopressin for cardiac arrest: a systematic review and meta-analysis. *Arch Intern Med*. 2005;165:17-24.

A5. Gueugniaud PY, David JS, Chanzy E, et al. Vasopressin and epinephrine vs. epinephrine alone in cardiopulmonary resuscitation. *N Engl J Med*. 2008;359:21-30.

A6. Dorian P, Cass D, Schwartz B, et al. Amiodarone as compared with lidocaine for shock-resistant ventricular fibrillation. *N Engl J Med*. 2002;346:884-890.

A7. Mentzelopoulos SD, Malachias S, Chamos C, et al. Vasopressin, steroids, and epinephrine and neurologically favorable survival after in-hospital cardiac arrest: a randomized clinical trial. *JAMA*. 2013;310:270-279.

A8. Nielsen N, Wetterslev J, Cronberg T, et al. Targeted temperature management at 33 degrees C versus 36 degrees C after cardiac arrest. *N Engl J Med*. 2013;369:2197-2206.

A9. Kim F, Nichol G, Maynard C, et al. Effect of prehospital induction of mild hypothermia on survival and neurological status among adults with cardiac arrest: a randomized clinical trial. *JAMA*. 2014;311:45-52.

A10. Connolly SJ, Hallstrom AP, Cappato R, et al. Meta-analysis of the implantable cardioverter defibrillator secondary prevention trials. AVID, CASH and CIDS studies. Antiarrhythmics vs Implantable Defibrillator study. Cardiac Arrest Study Hamburg. Canadian Implantable Defibrillator study. *Eur Heart J*. 2000;21:2071-2078.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Marijon E, Tafflet M, Celermajer DS, et al. Sports-related sudden death in the general population. *Circulation*. 2011;124:672-681.
2. Maron BJ, Haas TS, Murphy CJ, et al. Incidence and causes of sudden death in U.S. college athletes. *J Am Coll Cardiol*. 2014;63:1636-1643.
3. Eckart RE, Shry EA, Burke AP, et al. Sudden death in young adults an autopsy-based series of a population undergoing active surveillance. *J Am Coll Cardiol*. 2011;58:1254-1261.
4. Myerburg RJ, Junttila MJ. Sudden cardiac death caused by coronary heart disease. *Circulation*. 2012;125:1043-1052.
5. Field JM, Hazinski MF, Sayre MR, et al. Part 1: executive summary. 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circulation*. 2010;122:S640-S656.
6. Meaney PA, Bobrow BJ, Mancini ME, et al. Cardiopulmonary resuscitation quality: improving cardiac resuscitation outcomes both inside and outside the hospital: a consensus statement from the American Heart Association. *Circulation*. 2013;128:417-435.
7. Bobrow BJ, Spaite DW, Berg RA, et al. Chest compression-only CPR by lay rescuers and survival from out-of-hospital cardiac arrest. *JAMA*. 2010;304:1447-1454.
8. Murakami Y, Iwami T, Kitamura T, et al. Outcomes of out-of-hospital cardiac arrest by public location in the public-access defibrillation era. *J Am Heart Assoc*. 2014;3:e000533.
9. Myerburg RJ, Halperin H, Egan D, et al. Pulseless electrical activity: definition, causes, mechanisms, management, and research priorities for the next decade. Report from a National Heart, Lung, and Blood Institute Workshop. *Circulation*. 2013;128:2532-2541.
10. Hagihara A, Hasegawa M, Abe T, et al. Prehospital epinephrine use and survival among patients with out-of-hospital cardiac arrest. *JAMA*. 2012;307:1161-1168.
11. Morrison LJ, Neumar RW, Zimmerman JL, et al. Strategies for improving survival after in-hospital cardiac arrest in the United States: 2013 consensus recommendations: a consensus statement from the American Heart Association. *Circulation*. 2013;127:1538-1563.
12. Scirica BM. Therapeutic hypothermia after cardiac arrest. *Circulation*. 2013;127:244-250.
13. Mader TJ, Nathanson BH, Millay S, et al. Out-of-hospital cardiac arrest outcomes stratified by rhythm analysis. *Resuscitation*. 2012;83:1358-1362.
14. Kudenchuk PJ, Redshaw JD, Stubbs BA, et al. Impact of changes in resuscitation practice on survival and neurological outcome after out-of-hospital cardiac arrest resulting from nonshockable arrhythmias. *Circulation*. 2012;125:1787-1794.

REVIEW QUESTIONS

1. Which of the following electrocardiogram (ECG) patterns at first contact with a cardiac arrest victim has the lowest probability of survival?

A. Asystole
B. Polymorphic ventricular tachycardia
C. Pulseless ventricular tachycardia
D. Pulseless electrical activity
E. Ventricular fibrillation

Answer: A Cardiac arrest victims with shockable rhythms (B, C, and E) have higher probabilities of survival than pulseless electrical activity or asystole (A and D). Asystole may appear after delayed or prolonged treatment of ventricular tachyarrhythmias or fibrillation, or as a primary mechanism in patients with preexisting advanced heart disease or terminal noncardiac diseases. In either situation, survival is low. Pulseless electrical activity also has a low probability of survival because it generally takes longer to restore circulation in nonshockable rhythms, but survival is better than for asystole.

2. A 35-year-old man suffered a cardiac arrest while running a marathon and was found to be in ventricular fibrillation by emergency rescue personnel. He had had no prior history of chest pain or palpitations. A structural cause for his cardiac arrest is identified at postmortem examination. Which of the following disorders is the most likely cause of his cardiac arrest?

A. Aortic stenosis
B. Congenital long QT interval syndrome
C. Coronary atherosclerosis
D. Hypertrophic cardiomyopathy
E. Nonischemic dilated cardiomyopathy

Answer: C The less common cardiac disorders such as hypertrophic cardiomyopathy and the inherited arrhythmia syndromes are more common causes of sudden cardiac arrest than coronary artery disease in the adolescent and younger adult population. However, coronary atherosclerosis emerges as the most common cause in people older than 30 to 35 years.

3. A 45-year-old man comes to the emergency department because of the sudden onset of a tachycardia associated with mild lightheadedness. Initial ECG shows a regular wide QRS tachycardia at a rate of 165 beats per minute; blood pressure is 126/82 mm Hg. He denies a prior history of heart disease, and no information is available on his left ventricular function. Which of the following drugs should not be used as initial management of this arrhythmia?

A. Adenosine
B. Digoxin
C. Metoprolol
D. Procainamide
E. Verapamil

Answer: E Verapamil given intravenously causes clinically relevant myocardial depression in patients with left ventricular dysfunction due to structural heart disease. Therefore, regardless of the mechanism of a tachycardia, intravenous verapamil should not be used in the absence of prior knowledge that left ventricular function is normal.

4. A 72-year-old woman with a prior myocardial infarction and an ejection fraction of 40% collapsed in a shopping mall. A bystander found her to be pulseless and began "compression-only" cardiopulmonary resuscitation (CPR). An automated external defibrillator was deployed approximately 3 minutes after collapse and demonstrated ventricular fibrillation. An initial 200-J biphasic shock failed to restore sinus rhythm. Which of the following should be the next step in management?

A. Deliver a second 200-J biphasic shock immediately.
B. Inflate the lungs with mouth-to-mouth respirations twice and then deliver a second shock.
C. Resume CPR for five cycles of 15 : 2 compressions/ventilations before delivering another shock.
D. Deliver a precordial thump, and recheck the rhythm.
E. Resume CPR immediately and deliver another 200-J biphasic shock after 2 minutes of chest compressions.

Answer: E Based on the concept of maximizing chest compressions for preserving organ viability, the prior strategy of delivering three successive shocks before resuming CPR has been replaced by the guideline that CPR should be resumed after a single shock that fails to restore circulation. CPR should be continued using the compression-only technique for 2 minutes before the second shock is delivered.

5. For a cardiac arrest patient who is in ventricular fibrillation and has failed multiple attempts to restore circulation despite CPR, attempted defibrillation, and epinephrine, which of the following intravenous medications should be tried next?

A. Amiodarone
B. Lidocaine
C. Magnesium sulfate
D. Procainamide
E. Propranolol

Answer: A After failure to restore sinus rhythm by 3 cycles of defibrillator shocks and chest compressions, including the administration of epinephrine or vasopressin, a randomized clinical trial has demonstrated that amiodarone intravenously is more effective than lidocaine for restoring circulation, usually with subsequent shocks. Lidocaine is an acceptable alternative if the ventricular fibrillation (VF) event is due to an acute ischemic syndrome, and magnesium sulfate is preferable when VF is due to prolonged QT interval syndrome. Procainamide can be tried after amiodarone and/or lidocaine has failed, but it is unlikely to be of benefit.

64

CARDIAC ARRHYTHMIAS WITH SUPRAVENTRICULAR ORIGIN

PETER ZIMETBAUM

Supraventricular arrhythmias are divided into bradyarrhythmias and tachyarrhythmias. Any rhythm that originates above where the His bundle bifurcates into the right and left bundle branches is considered to be supraventricular in origin.

ANATOMY AND NORMAL ELECTROPHYSIOLOGY

The normal cardiac impulse begins in the sinus node complex, which is located at the junction of the right atrium and the superior vena cava. It then travels through the right atrium and primarily activates the left atrium through the coronary sinus. The time it takes to activate the atria is represented by the P wave on the electrocardiogram (ECG). After depolarizing the atria, the impulse enters the atrioventricular (AV) node, which is located in the inferior septal region of the right atrium, where a delay occurs. This delay allows time for the atria to contract and fill the ventricles. In most individuals, the impulse travels through the AV node over a uniform functional pathway or route. Some people have two or more functional pathways called dual AV nodal pathways (fast and slow pathways). The delay in the AV node represents most of the isoelectric portion of the PR interval on the ECG (E-Fig. 64-1). The usual duration of atrial activation including delay in the AV node is up to 140 msec and can be measured directly as the atrial-His interval. The impulse then travels into the specialized infranodal conducting system, that is, through the His bundle, then right and left bundle branches, and into the Purkinje network. The Purkinje network extends or fans out throughout the ventricular endocardium. Impulses conduct rapidly through the Purkinje network, thereby allowing nearly simultaneous activation of the ventricles. A small portion of the isoelectric segment of the PR interval represents infranodal conduction. This infranodal conduction through the His-Purkinje system can also be measured directly (His-ventricle interval) and should take between 40 and 60 msec. Once out of the Purkinje network, the impulse proceeds relatively slowly from the endocardial to epicardial surface of the ventricles. The QRS complex on ECG represents depolarization of the bundle branches and ventricular myocardium. Patients who have ventricular preexcitation activate the ventricles through an alternative route to the normal ventricular conduction system. These patients have bypass tracts or accessory pathways over which the ventricular myocardium can be activated directly rather than traveling over the AV node and His-Purkinje network. These pathways, which develop as a failure of the normal fibrous separation of the atria and ventricles, are located in proximity to the tricuspid and mitral valves. Direct activation of the ventricular myocardium without the usual delay in the AV node results in a slurred upstroke of the QRS complex called a delta (δ) wave (E-Fig. 64-2).

The normal heart rate is generated by tissues or pacemaker cells with intrinsic automaticity. The sinus node cells produce the greatest rate (60 to 100 beats per minute) of automaticity and suppress other potentially automatic (AV junctional, 40 to 55 beats per minute; His-Purkinje cells, 15 to 40 beats per minute) tissues with slower rates of depolarization. The sinus node and AV node are heavily influenced by the parasympathetic (vagal) and sympathetic (adrenergic) nervous system. At rest, the parasympathetic system controls sinus node automaticity. With exertion or emotional or physical stress, a withdrawal of parasympathetic tone and an increase in heart rate, which is then perpetuated by sympathetic tone, further increase heart rate. Sinus arrhythmia refers to the normal variation in heart rate with inspiration and expiration. With inspiration, a withdrawal of vagal tone increases heart rate; by comparison, expiration is associated with a drop in heart rate (Fig. 64-1). During sleep, a dominance of vagal tone slows heart rate.

BRADYARRHYTHMIAS

Bradyarrhythmias may be caused by sinus node, AV node, or His-Purkinje dysfunction (Table 64-1).

Sinus Bradycardia and Sinus Node Dysfunction

Sinus bradycardia (Fig. 64-2) is generally defined as a sinus rate of less than 60 beats per minute. It should be noted, however, that sinus rates as low as 45 to 50 beats per minute, particularly at rest, can be physiologically normal. Sinus node dysfunction encompasses a group of disorders including sinus bradycardia, sinoatrial (SA) exit block, sinus arrest (pause of >2 to 3 seconds) during sinus rhythm, chronotropic incompetence, and tachycardia-bradycardia (tachy-brady) syndrome. Sinus node dysfunction in combination with symptoms such as fatigue, dizziness, near or complete syncope (Chapters 51 and 62), or worsening of heart failure (Chapter 58) is called sick sinus syndrome. The tachy-brady syndrome is often identified by a prolonged delay in sinus node recovery following the termination of atrial fibrillation (AF) (Fig. 64-3). SA exit block refers to the electrophysiologic phenomenon of sinus node firing with delay or block of the impulse as it travels from the sinus node to the surrounding atrial tissue (Fig. 64-4). SA exit block can be first degree, second degree (type 1 or 2), and third degree. First-degree SA block is difficult to diagnose from the surface ECG. Second-degree SA exit block type 1 is manifest by progressive PP shortening preceding the sinus pause. The PP interval following the pause must be greater than twice the PP interval that preceded the pause. Second-degree SA exit block type 2 is characterized by a pause equaling an exact multiple of the sinus rate

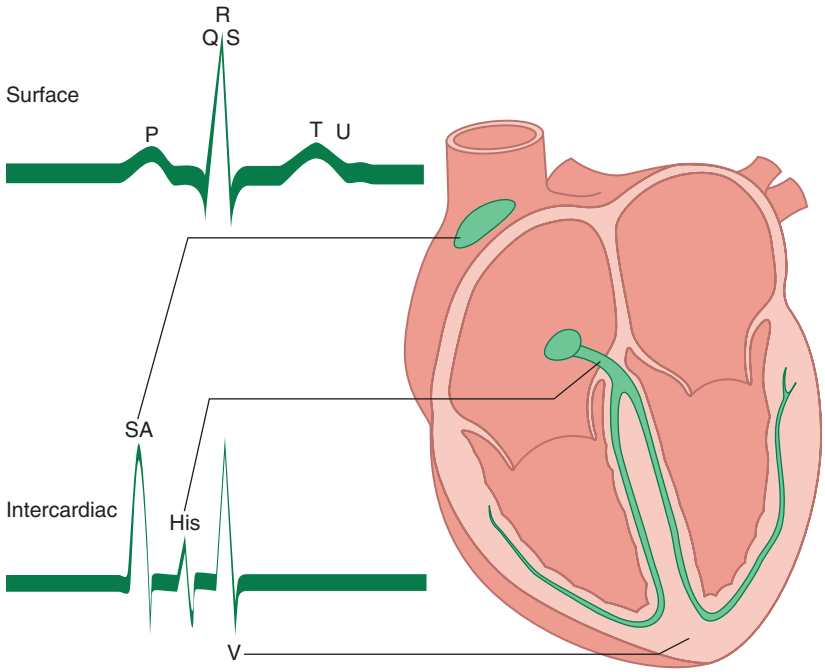

E-FIGURE 64-1. Diagram of the surface electrocardiogram (P, QRS, T, and U waves) and the related intracardiac electrograms (SA, His, and V). The atrial (SA) electrogram represents intra-atrial conduction. The His bundle electrogram (His) represents the recording of a His potential, and the ventricular (V) electrogram represents activation of the ventricles. The SA-His interval corresponds to intra-atrial conduction and delay in the atrioventricular node. The His-V interval represents conduction between the His bundle and the ventricles (infranodal conduction).

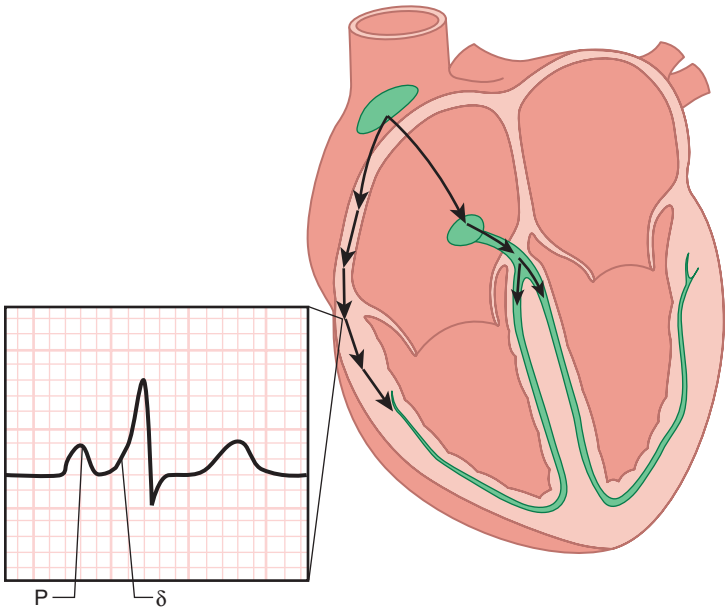

E-FIGURE 64-2. Conduction over an accessory pathway resulting in a delta (δ) wave.

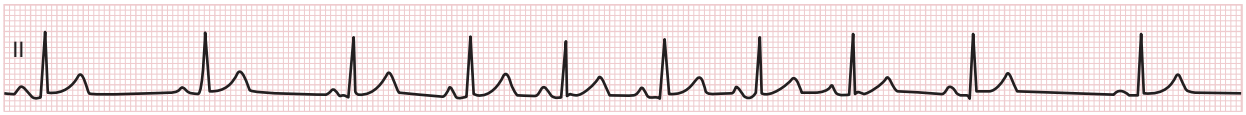

FIGURE 64-1. Sinus arrhythmia. Note the variation in sinus rates, which fluctuate with normal variations in autonomic tone.

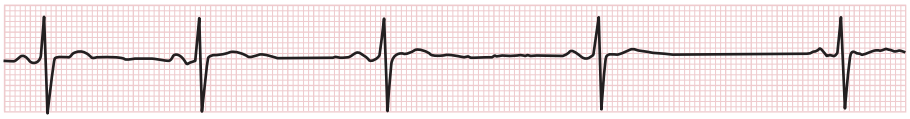

FIGURE 64-2. Sinus bradycardia. Progressive sinus bradycardia—in this case related to heightened vagal tone while sleeping.

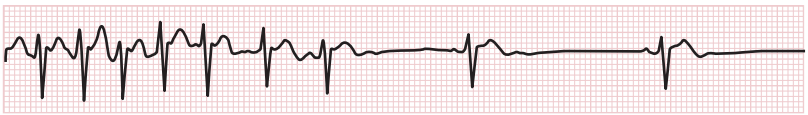

FIGURE 64-3. Electrocardiographic evidence of tachy-brady syndrome. Atrial fibrillation with a tachycardic ventricular response followed by conversion to sinus bradycardia.

FIGURE 64-4. Sinoatrial block. Sinoatrial exit block, probably type 2, is characterized by a pause equaling an exact multiple of the sinus rate.

TABLE 64-1 BRADYCARDIAS

SINUS NODE DYSFUNCTION
Sinus bradycardia < 45 beats/min
Sinoatrial exit block
First-degree
Second-degree
Third-degree
Sinus arrest
Bradycardia-tachycardia syndrome
ATRIOVENTRICULAR BLOCK
First-degree
Second-degree
Mobitz type I (Wenckebach phenomenon)
Mobitz type II
Higher degree (e.g., 2 : 1, 3 : 1)
Third-degree
Atrioventricular node
His-Purkinje system

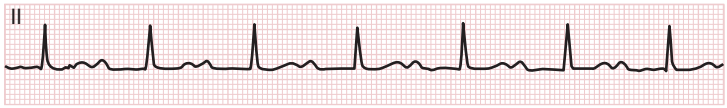

FIGURE 64-5. First-degree atrioventricular (AV) block. Note the prolonged (>200 msec) AV conduction.

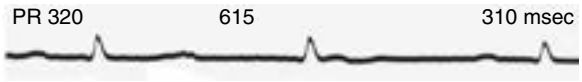

FIGURE 64-6. Mobitz I block. Progressive PR prolongation from 320 to 615 msec, followed by a blocked P wave. The subsequent conducted PR interval is less than the PR interval before the dropped P wave.

(i.e., constant PP interval before and after the pause). High-degree SA exit block refers to the absence of multiple P waves with a pause still corresponding to an absolute multiple of the underlying PP intervals, and third-degree AV block results in a complete absence of sinus P waves.

Chronotropic incompetence refers to the inability to increase the sinus rate appropriately in response to exercise or other physiologic demand. In most patients, chronotropic incompetence is manifest by a maximal heart rate of less than 100 beats per minute.

ATRIOVENTRICULAR CONDUCTION DISTURBANCES

AV conduction disturbances refer to abnormal conduction in the AV node or in the His-Purkinje system (HPS) below the AV node. Electrical transmission through the AV conduction system is primarily limited by the AV node, which conducts in a decremental fashion to prevent excessively rapid conduction to the ventricles. The normal AV node rarely conducts faster than 200 beats per minute and slows with aging. The AV node is heavily influenced by autonomic tone and may conduct more than 200 beats per minute in the presence of heightened sympathetic and withdrawal of parasympathetic tone. Conduction through the HPS system is faster and nondecremental.

The AV blocks are classified as first, second, high, and third degree. First-degree AV block is a misnomer because nothing is actually blocked—rather, there is delay, usually in the AV node, manifest by a prolonged PR interval (Fig. 64-5). Second-degree AV block is divided into Mobitz type I (Wenckebach) or Mobitz type II. Mobitz type I is defined by progressive PR prolongation with eventual block after a P wave (Fig. 64-6). The initial PR prolongation is longest, the subsequent RR intervals shorten, and the PR interval following the blocked P wave is shorter than the last conducted PR interval before the blocked P wave. Mobitz type I usually occurs in the AV node. Mobitz type II, which is characterized by the abrupt failure of conduction after a P wave without preceding PR prolongation, usually represents conduction disease below the AV node. In patients with 2 : 1 block (two P waves for every QRS), it can be difficult to determine whether the block is Mobitz I in the AV node or Mobitz II below the AV node. Clues to conduction disease in the AV node include a prolonged PR interval (i.e., more than 300 msec) and a narrow QRS duration. Clues to conduction disease below the node include a normal PR interval but with bundle branch block (Fig. 64-7).

High-degree or advanced AV block, which is a form of second-degree block with multiple or successive nonconducted P waves, or both (Fig. 64-8), frequently recurs or persists. Third-degree block (or complete heart block) refers to a rhythm in which the atrial and ventricular activity occur independently, and the atrial rate usually exceeds the ventricular rate. Third-degree heart block can be seen with sinus rhythm or any atrial tachyarrhythmia with a regular escape rhythm in the AV junction or below (Fig. 64-9). Sometimes there is no escape rhythm, and heart block results in asystole. Complete heart block, particularly when it is acute and accompanied by an escape rhythm, can be associated with marked QT prolongation, which signifies a risk for

torsades de pointes (Chapter 65). For this reason, patients who undergo AV node ablation have their pacemakers set at 80 beats per minute for at least 6 weeks after the procedure to prevent QT prolongation and torsades de pointes. Eventually, the pacing rate can be reduced to more physiologic levels without the ongoing risk for ventricular arrhythmia. Complete heart block and other types of severe bradyarrhythmia may present with syncope if there is a prolonged pause before an escape rhythm develops. More often, these rhythms present with fatigue and dyspnea. The blood pressure is often elevated owing to peripheral vasoconstriction, and there may be renal insufficiency secondary to reduced cardiac output.

The term AV dissociation refers to any rhythm in which the atria and ventricles beat independently of one another. If the atrial rate is faster than the ventricular rate, it is called complete heart block or third-degree AV block. AV dissociation with an atrial rate slower than the subsidiary pacemaker is usually seen with junctional or ventricular tachycardias.

CLINICAL MANIFESTATIONS

Sinus bradycardia and various degrees of AV nodal blocks can occur asymptomatically during sleep in healthy individuals. Asymptomatic first- and second-degree AV block, particularly when partially or completely reversed by exercise, is usually benign. Persistent second-degree and third-degree AV nodal block is abnormal and is often associated with dizziness, fatigue, exertional dyspnea, worsening of heart failure, near-syncope, or syncope. Third-degree AV block with a good junctional escape mechanism that accelerates during exercise, as often noted in patients with congenital AV block, may remain asymptomatic. Patients with congenital heart block may not appreciate their potential for a more active lifestyle because of the lack of a reference point but often feel much better when an appropriate heart rate acceleration can be achieved after pacemaker therapy.

DIAGNOSIS

Bradycardias are typically diagnosed by the ECG. In symptomatic patients with symptoms suggestive of bradyarrhythmia, 24-hour Holter monitoring or prolonged loop monitoring usually can make the diagnosis, but some patients may require formal electrophysiologic testing (Chapter 62). Any bradycardias, including sinus node dysfunction (Table 64-2) and AV nodal block (Table 64-3), can be caused at least in part by vagal influences, such as vasovagal episodes, vomiting, abdominal surgery, and upper and lower gastrointestinal invasive procedures. Syncope, sometimes caused by the bradycardia and sometimes by vasodepression with hypotension, may result (Chapters 51 and 62). Medications, infiltrative diseases, fibrocalcific degeneration, and a variety of other causes must be considered. Lyme disease also is a common cause of reversible complete heart block, usually localized to the AV node.

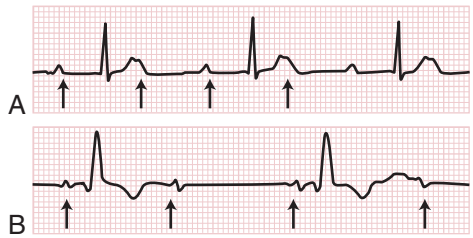

FIGURE 64-7. Two-to-one conduction in the atrioventricular (AV) node and below the AV node. A, Two-to-one conduction, with the conducted beats demonstrating a prolonged PR interval (>300 msec) and a narrow QRS, indicating Mobitz I block in the AV node. **B,** Normal-duration conducted PR interval and a wide QRS duration favoring Mobitz II block into an infranodal site of block.

FIGURE 64-8. High-degree atrioventricular block. Periods of complete heart block with three nonconducted P waves (the first of which is buried in the first QRS complex) followed by two P waves, the second of which is conducted to the ventricle with a long PR interval. The subsequent three P waves are once again nonconducted, with the last P wave buried in the last QRS complex.

TREATMENT

Rx

The treatment of sinus node dysfunction and AV blocks consists of first removing any medications that may precipitate dysfunction (see Table 64-2). Although some patients will recover normal conduction, the susceptibility to medications usually indicates an underlying conduction abnormality that may worsen over time.

Asymptomatic sinus node dysfunction requires no therapy. Asymptomatic AV nodal block, particularly if the QRS escape complex is narrow, can also be managed without intervention unless the QT interval is markedly prolonged, in which case a pacemaker should be considered. For symptomatic sinus node dysfunction and second- and third-degree AV block, acute management includes intravenous atropine (1 mg) or isoproterenol (usually 1 to 2 µg/minute infusion) to increase the heart rate. Temporary cardiac pacing (Chapter 66) may be required. If sinus node dysfunction or AV block is due to transient

TABLE 64-2 CAUSES OF SINUS NODE DYSFUNCTION

INTRINSIC
Hypothyroidism
Fibrocalcific degeneration
Increased vagal tone, especially in sleep apnea
Congenital mutations
Scleroderma
Amyloidosis
Chagas disease
EXTRINSIC
Trauma, including cardiac surgery
Drugs
Calcium-channel blockers
β-Blockers
Digoxin
Antiarrhythmic medications (amiodarone, dronedarone, sotalol, flecainide, propafenone)
Lithium

TABLE 64-3 CAUSES OF ATRIOVENTRICULAR BLOCK

All causes of sinus node dysfunction listed in Table 64-2 and also:
Lyme disease
Bacterial endocarditis with abscess formation
Cardiac sarcoidosis with granuloma
Inferior myocardial infarction
Anterior myocardial infarction (less common and often associated with cardiogenic shock)
Congenital mutations (possibly associated with maternal lupus erythematosus and transmission of anti-Ro and La antibodies)
Corrected transposition of the great vessels
Chagas disease
Some neurologic conditions (especially myotonic dystrophy)
Drugs as in Table 64-2

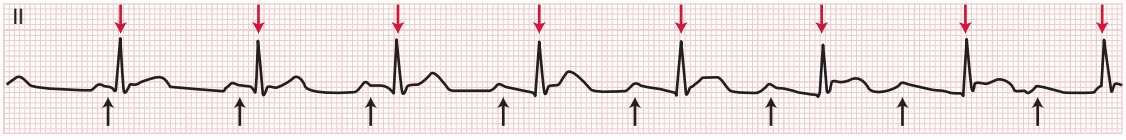

FIGURE 64-9. Complete heart block with the atrial beats *(black arrows)* dissociated from the ventricular beats *(red arrows)*. Note the ST elevation in leads III and aVF indicating acute inferior myocardial infarction as the cause of complete heart block.

abnormalities, such as drug-induced, acute ischemic syndromes, temporary pacing is usually sufficient; however, when intra-His or infra-His block is suspected (e.g., exercise-induced AV block or asymptomatic Mobitz type II block) and the site can be documented with His bundle recording, permanent pacing (see Tables 66-1 and 66-2) is the only effective chronic therapy, and consensus recommendations for pacemaker implantation should guide its use. For all forms of persistent symptomatic sinus node dysfunction or second- or third-degree AV block, permanent pacing is the therapy of choice (Chapter 66).[1,2] In patients with atrioventricular block and systolic dysfunction, biventricular pacing is preferred. The only exception is Lyme disease, in which most cases of heart block resolve within a week of antibiotic therapy (Chapter 321).

Supraventricular Rhythms with a Normal Rate

Atrial premature beats (APBs) can arise from the right or left atrium or the pulmonary veins. The P wave, which differs from the P wave of sinus rhythm unless the APB originates near the sinus node, always precedes the QRS complex (Fig. 64-10). If the P wave is blocked, however, it is not followed by a QRS complex. A blocked premature atrial contraction may be confused with second-degree AV block, unless its prematurity is recognized, or with sinus node dysfunction, if it is inconspicuous. Altered appearance of the ST-T segment is often a clue to the presence of a P wave. A premature QRS complex with the morphology of the underlying sinus rhythm in the absence of a premature P wave represents an AV junctional beat (Fig. 64-11).

An ectopic atrial rhythm refers to a nonsinus atrial rhythm from a single focus with a single P wave morphology (Fig. 64-12). Wandering atrial pacemaker refers to an ectopic atrial rhythm with at least three distinct P wave morphologies at rates between 50 and 100 beats per minute (Fig. 64-13).

CLINICAL MANIFESTATIONS

APBs are most often asymptomatic but can occasionally be experienced as palpitations (Chapter 62). Similarly, ectopic atrial rhythms, including a wandering atrial pacemaker, are virtually always asymptomatic. Rarely, ectopic atrial rhythm can be very slow and associated with symptoms of fatigue.

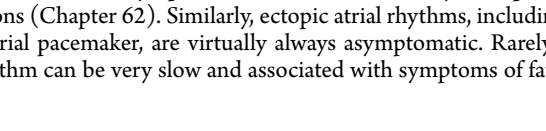

FIGURE 64-10. Atrial premature beat. Sinus rhythm with an atrial premature beat as the third complex in the rhythm strip. The P wave is buried in the preceding T wave.

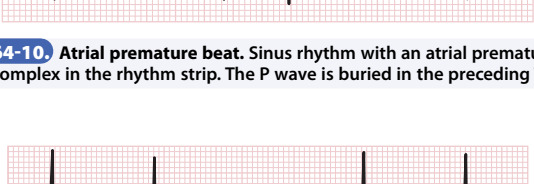

FIGURE 64-11. Junctional premature beat. The premature beat is narrow but slightly different in morphology compared with the surrounding sinus beats. This premature beat is from the atrioventricular junction or slightly more distal in the fascicles. The inverted P wave following the QRS represents retrograde activation of the atria.

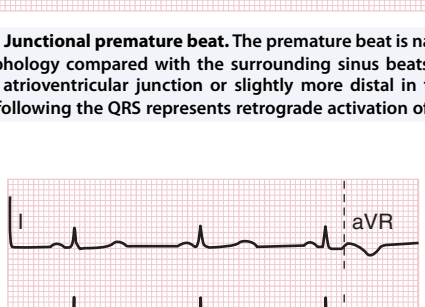

FIGURE 64-12. Ectopic atrial rhythm. The inverted P waves in leads II, III, and aVF indicate a nonsinus P wave originating in the low right atrium.

DIAGNOSIS

The diagnoses of atrial premature beats, ectopic atrial rhythms, and wandering atrial pacemakers are all made by ECG or on an ambulatory Holter monitor or loop event recorder (Chapter 62).

TREATMENT

Rx

Premature atrial contractions do not generally require treatment unless they are associated with significant symptoms. Treatment consists primarily of β-blockers (e.g., atenolol, 25 to 100 mg daily) or calcium-channel blockers (e.g., long-acting diltiazem, 180 to 300 mg daily). Ectopic atrial rhythms and wandering atrial pacemakers are rarely symptomatic and are not treated with medications. Rarely, a slow ectopic rhythm associated with fatigue can be treated with atrial pacing at a rate faster than the ectopic atrial rhythm.

Supraventricular Tachyarrhythmias

The supraventricular tachycardias (SVTs) are defined as arrhythmias with three or more beats at a rate of greater than 100 beats per minute (Table 64-4). The beats, which can be regular or irregular, are usually narrow complex but can be wide complex when associated with bundle branch block (aberration) or conduction over an accessory pathway. Atrial dilation, acute myocardial infarction, pulmonary embolism, acute or chronic inflammatory states, or scars from prior surgery involving atrial myocardium or pericardium are among the causes of atrial tachyarrhythmias.

The most important diagnostic clues for diagnosing the type of SVT include the ventricular response rate, regularity, and, if known, the suddenness of the onset of tachycardia. The QRS is regular during sinus tachycardia, atrial flutter, AV nodal re-entrant tachycardia, atrioventricular reciprocating tachycardia, and atrial tachycardia; it is irregular with atrial AF, atrial flutter with variable AV block, multiple atrial premature contractions, and multifocal atrial tachycardia. Sudden onset and termination suggest acute AF, atrial flutter, AV nodal re-entrant tachycardia, atrioventricular reciprocating tachycardia, and atrial tachycardia. Gradual onset and recession suggest sinus tachycardia, chronic AF, atrial premature contractions and multifocal atrial tachycardia.

SINUS TACHYCARDIA

Sinus tachycardia (Fig. 64-14) is an arrhythmia that is almost always a physiologic response to an emotional or physical stress such as anxiety, exercise, anemia, hypotension, hypoxemia, fever, thyrotoxicosis, or heart failure. It is characterized by a gradual increase and decrease in heart rate and rarely exceeds 180 beats per minute. In rare cases, it may be a nonphysiologic condition called inappropriate sinus tachycardia, which is characterized by sinus tachycardia that develops in response to minimal stress and that continues beyond the time when the normal response would have slowed.

Sinus node re-entry is a rare form of SVT caused by re-entry in the sinus node. As opposed to physiologic ST, it begins abruptly, is often triggered by a premature atrial beat, and ends abruptly. The P wave morphology is identical to sinus rhythm, and β-blockers are the treatment of choice if required for symptoms.

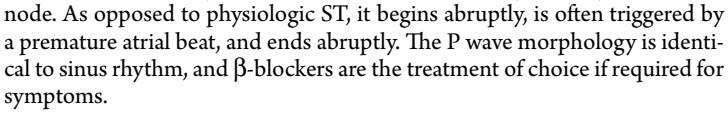

FIGURE 64-13. Wandering atrial pacemaker. Rhythm with three distinct P wave morphologies.

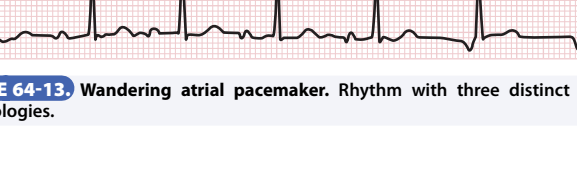

FIGURE 64-14. Sinus tachycardia at 120 beats/minute.

TABLE 64-4 SUPRAVENTRICULAR TACHYCARDIAS

	R-R REGULARITY	P WAVE MORPHOLOGY
ATRIAL TACHYCARDIAS		
Sinus tachycardia	Regular	Positive in II, III, aVF; negative in aVR
Sinus node re-entry	Regular	Positive in II, III, aVF; negative in aVR
Atrial tachycardia, unifocal	Regular	P different from sinus
Atrial tachycardia, multifocal	Irregular	Three or more different P wave morphologies
Atrial flutter, common: counterclockwise	Regular; irregular if variable AV block	Sawtooth flutter waves; regular waveform; negative in II, III, aVF, positive in V_1, negative in V_6
clockwise		Positive in II, III, aVF; negative in V_1; positive in V_6
Atrial flutter: uncommon	Regular; irregular if variable AV block	Pattern different than common atrial flutter (counterclockwise or clockwise)
Atrial fibrillation	Irregularly irregular	Irregular fibrillation waves
AV JUNCTIONAL TACHYCARDIAS		
AV re-entry (using accessory pathways)		
Orthodromic	Regular	Retrograde P in ST-T wave
Antidromic	Regular preexcited, except in irregular preexcited atrial fibrillation	Retrograde P, short RP
Slow conducting	Regular	Retrograde P at end of T wave or later (long RP)
Atriofascicular (antidromic)	Regular preexcited	Retrograde P, short RP
AV nodal re-entry		
Common (slow-fast)	Regular	Retrograde P obscured by QRS or alters the end of QRS (short RP)
Uncommon (fast-slow)	Regular	Retrograde P at end of T wave or later (long RP)
Others (slow-slow)	Regular	PR-RP approximately equal
Nonparoxysmal junctional tachycardia*	Regular, slow rate	AV dissociation
Automatic junctional tachycardia*	Regular	AV dissociation

AV = atrioventricular.
*Site of origin is usually infranodal.

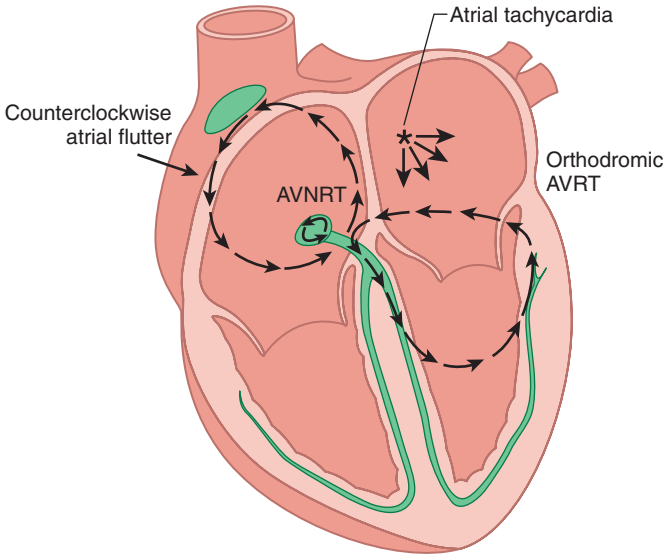

FIGURE 64-15. Diagram of the site and mechanism of common forms of supraventricular tachycardia. AVNRT = atrioventricular nodal re-entrant tachycardia; AVRT = atrioventricular re-entrant tachycardia.

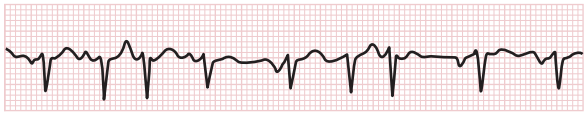

FIGURE 64-16. Multifocal atrial tachycardia. An atrial tachycardia with three different P wave morphologies.

ATRIAL TACHYCARDIA

The term atrial tachycardia refers to a group of SVTs that originate from focal anatomic locations in the atria and propagate in a centrifugal pattern (Fig. 64-15). These locations include the pulmonary veins, crista terminalis in the right atrium, tricuspid or mitral annulus, coronary sinus, atrial septa, left atrial appendage, aortomitral continuity, or regions of scar tissue from previous cardiac surgery. Atrial tachycardias, which are usually regular rhythms that rarely exceed 200 beats per minute, may present in young age but more often develop later in life. Atrial tachycardias can be re-entrant, triggered, or automatic in mechanism. Some forms of atrial tachycardia are incessant and can predispose to a tachycardia-related cardiomyopathy (Chapter 60). Other forms of atrial tachycardia are paroxysmal and may remit spontaneously without treatment. When persistent, atrial tachycardias can be managed with medications (calcium-channel or β-blockers or other antiarrhythmic drugs). Electrical mapping and ablation (Chapter 66) of atrial tachycardia are highly effective and are increasingly offered as first-line therapy. Multifocal atrial tachycardia (Fig. 64-16), which is a unique form of atrial tachycardia characterized by three or more P wave morphologies, is distinguished from wandering atrial pacemaker when the rate is 100 beats per minute or more rather than less than 100 beats per minute. It is an irregular rhythm and occurs almost exclusively in patients with advanced pulmonary disease.

ATRIOVENTRICULAR NODAL RE-ENTRANT TACHYCARDIA

Atrioventricular nodal re-entrant tachycardia (AVNRT), which is a common form of SVT in all age groups, presents most often in young adulthood and frequently occurs after a change in position. It is more common in women than men and may develop or be exacerbated by pregnancy or certain phases of the menstrual cycle. Patients with AVNRT have two functional AV nodal pathways (dual pathways). Typically, an atrial premature beat blocks in one pathway (fast pathway) and conducts slowly over the other pathway (slow pathway). If this beat conducts back up the fast pathway and then re-enters the slow pathway, AV nodal re-entry occurs (Fig. 64-17). This re-entrant circuit is confined to the AV node, and the result is that both the atria and ventricles are activated nearly simultaneously. P waves may not be visible on the ECG because they are buried in the QRS complex. When P waves are present, they usually are seen just after the QRS and have a negative or superior axis (negative in leads II, III, and aVF; short RP tachycardia) (see Fig. 64-15). Atypical AVNRT is rare and occurs with conduction down the fast pathway and up the slow pathway, with a long interval between the preceding R wave and subsequent negative P wave (e.g., long RP tachycardia).

JUNCTIONAL TACHYCARDIA

Junctional tachycardia refers to a focal (non-re-entrant) tachycardia originating in the AV junction (Fig. 64-18). A number of different forms of junctional tachycardia have differing clinical patterns. Nonparoxysmal junctional tachycardia is a benign arrhythmia that rarely exceeds 120 beats per minute and typically exhibits a "warm-up" and "cool-down" pattern. A more rapid paroxysmal form of junctional tachycardia occurs in young adults and is often associated with exercise. Congenital junctional ectopic tachycardia, which occurs in the pediatric population, is associated with very rapid rates and a risk for tachycardia-related cardiomyopathy.

ACCESSORY PATHWAY TACHYCARDIAS

Tachycardias associated with an accessory pathway or bypass tract are called atrioventricular re-entrant tachycardia (AVRT). These tachycardias most often conduct down the normal AV conducting system and back up

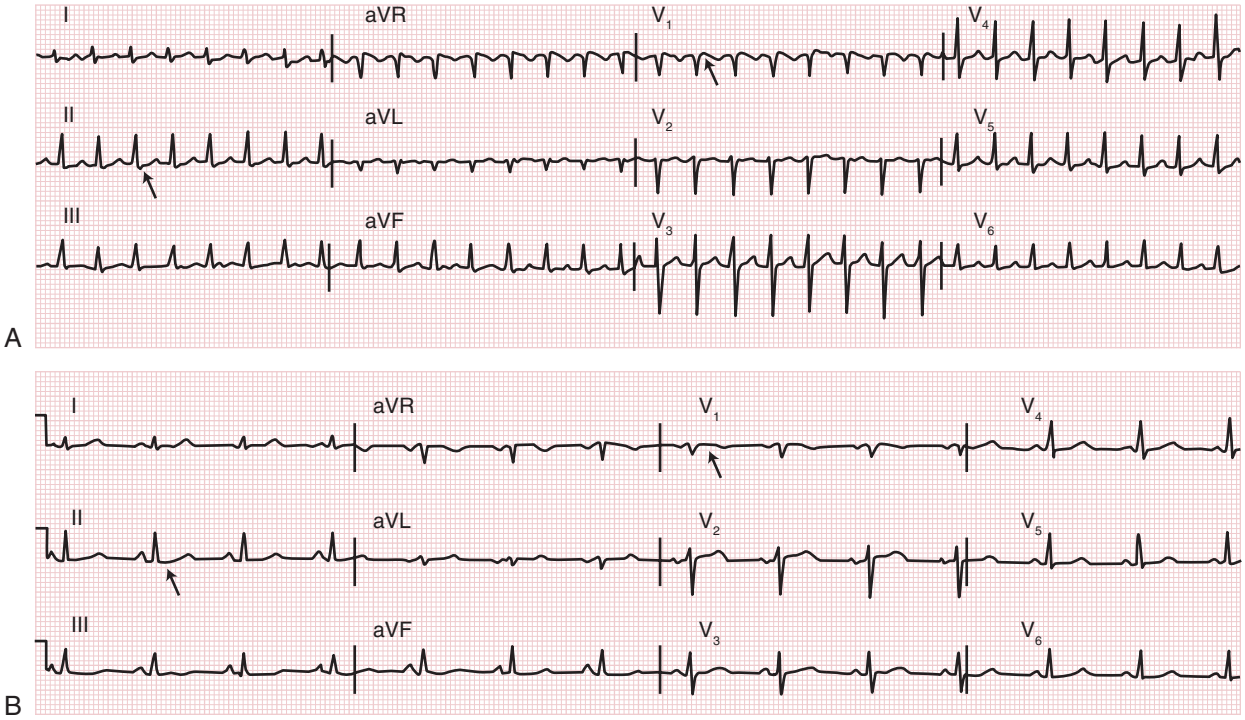

FIGURE 64-17. Electrocardiograms demonstrating different rhythms in the same patient. A, Atrioventricular nodal re-entrant tachycardia, with *arrows* indicating retrograde P waves (negative in lead II, positive in V_1). **B,** Same patient in sinus rhythm. The absence of a negative deflection at the end of the QRS in lead II and positive deflection in V_1 confirms that those findings in panel **A** were retrograde P waves.

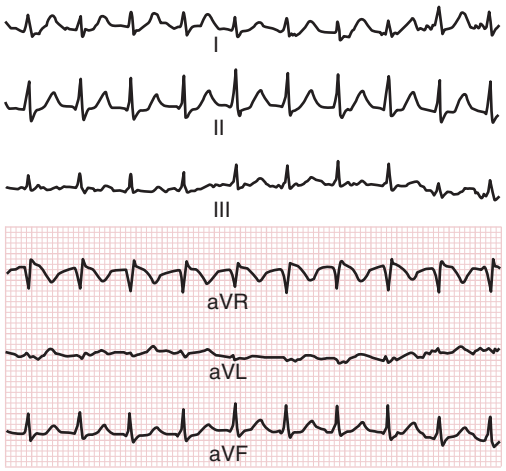

FIGURE 64-18. Junctional tachycardia. A tachycardia without evident P waves. The negative deflections at the end of leads II, III, and aVF represent retrograde P waves. This rhythm is impossible to distinguish from atrioventricular nodal re-entrant tachycardia on the surface electrocardiogram.

the accessory pathway (orthodromic AVRT) (see Fig. 64-15). Antidromic tachycardia is rare and represents conduction down the accessory pathway and back up the normal conducting system or, less commonly, a second accessory pathway. Wolff-Parkinson-White syndrome is defined by the presence of delta waves or preexcitation on the ECG in sinus rhythm (Fig. 64-19), in combination with a history of AVRT. It is important to recognize, however, that many patients with ventricular preexcitation never get AVRT. In addition, up to 40% of patients with accessory pathways have AF. Because accessory pathway tissue, unlike the AV node, is usually nondecremental, a rapidly conducting accessory pathway may allow AF to conduct to the ventricles at excessive rates and precipitate ventricular fibrillation. Drugs like digoxin, which accelerate accessory pathway conduction, should be avoided in these patients.

The ECG during orthodromic AVRT can have a narrow complex, but a wide QRS complex will be seen if there is aberration from a bundle branch block, especially at more rapid rates. The retrograde P wave is negative in leads I and aVL if it conducts over a left lateral accessory pathway and negative in leads II, III, and aVF if it conducts over a septal or posteroseptal accessory pathway (Fig. 64-20). The P wave occurs later (after the preceding R wave) compared with AVNRT because in AVRT, the impulse must conduct beyond the AV node and into the ventricle before returning to the atria over an accessory pathway. In antidromic AVRT, the QRS complex is wide with a slurred upstroke and resembles that seen with ventricular tachycardia as opposed to a bundle branch block. Antidromic conduction with AF is often fast, broad, and irregular (see Fig. 64-19).

ATRIAL FLUTTER

Atrial flutter is an arrhythmia with an atrial rate of approximately 300 beats per minute and a ventricular response of 150 (2:1), 100 (3:1), or slower multiples. Typical atrial flutter is a macro-re-entrant circuit that circulates in the right atrium in a clockwise or counterclockwise loop (Fig. 64-21). The inferior portion of typical flutter uses the narrow region between the inferior vena cava and tricuspid annulus as a critical isthmus of conduction. Typical atrial flutter almost always occurs in patients with underlying cardiovascular or pulmonary disease. It can also develop in patients who receive antiarrhythmic drugs (e.g., sodium-channel blocking drugs) for AF.[3] Atypical atrial flutter refers to macro-re-entrant atrial circuits that do not use this critical cavotricuspid isthmus of tissue. These atypical forms of atrial flutter often occur in the left atrium after mitral valve surgery or catheter ablation of AF. Typical atrial flutter can be recognized by a "sawtooth" P wave morphology, which is predominantly negative in leads II, III, and aVF and positive in V_1 (counterclockwise atrial flutter) or positive in leads II, III, and aVF and negative in V_1 (clockwise atrial flutter).

ATRIAL FIBRILLATION

AF is identified by the absence of P waves and the presence of an irregularly irregular ventricular rate. Coarse AF, which describes the presence of residual atrial activity on the ECG, is generally best seen in lead V_1 with the absence of P wave–like activity in other leads (Fig. 64-22). AF can be present with a regular ventricular rate when there is concomitant AV dissociation (e.g., complete heart block or ventricular tachycardia).

AF, which is the most common arrhythmia in clinical practice, afflicts more than 2 million Americans. The likelihood of developing AF increases with aging, with an anticipated three-fold increase in prevalence as the U.S. population ages in the next 20 years. AF is almost always a recurrent

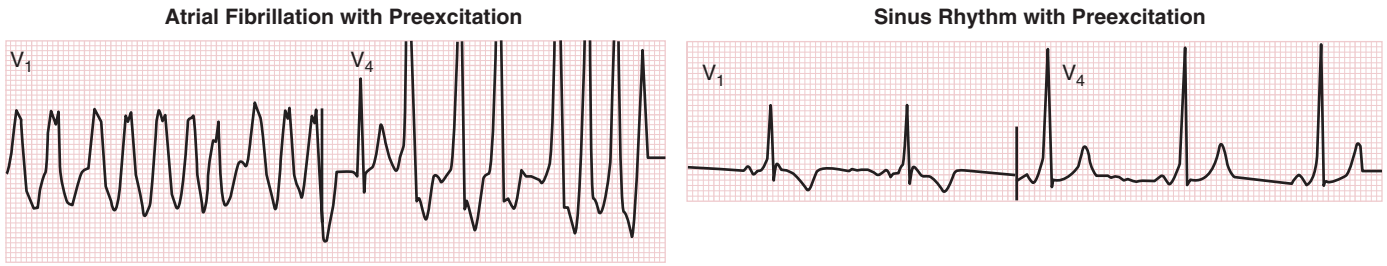

FIGURE 64-19. Two panels of the same patient with ventricular preexcitation. The rhythm on the left is fast, broad, and irregular, indicating atrial fibrillation with conduction over an accessory pathway. The rhythm on the right is sinus with conduction over an accessory pathway.

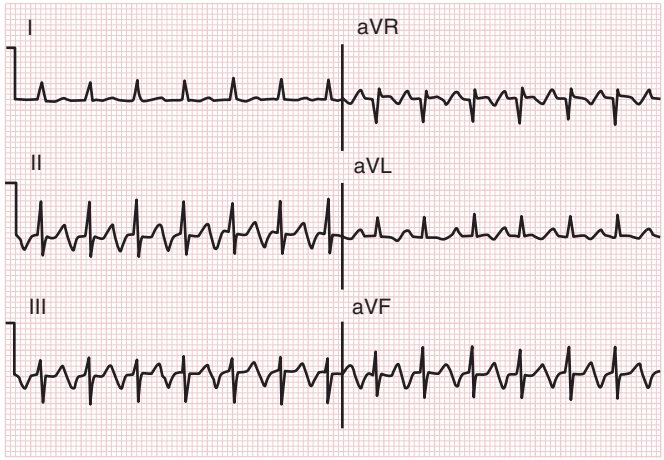

FIGURE 64-20. Atrioventricular re-entrant tachycardia. Negative P waves are seen in leads II, III, and aVF indicative of retrograde conduction to the atria over a septal accessory pathway.

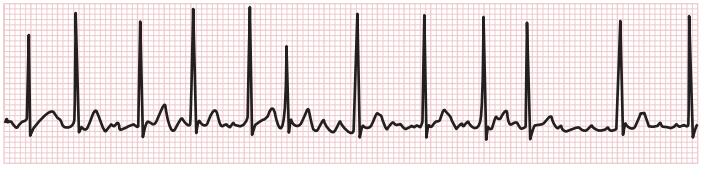

FIGURE 64-22. Coarse atrial fibrillation. The wavy baseline is suggestive of atrial activity and can be misinterpreted as P waves or flutter waves. The irregularity of the QRS complexes indicates that this is atrial fibrillation.

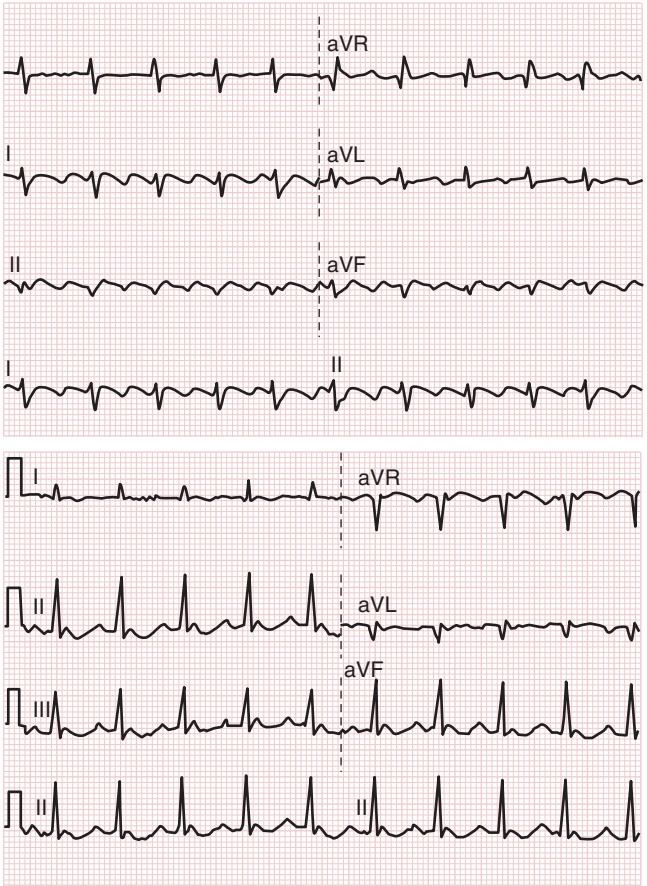

FIGURE 64-21. Two examples of atrial flutter. A, Typical counterclockwise right atrial flutter with classic sawtooth flutter waves seen in leads II, III, and aVF. **B,** Clockwise right atrial flutter.

disorder, with the possible exception of AF that develops in association with hyperthyroidism (Chapter 226) and surgery (Chapter 433). AF can be paroxysmal (terminates spontaneously) or persistent (persists for at least 7 days or until cardioverted). Paroxysmal AF can occur as self-remitting arrhythmia for decades or can progress to permanent AF. Patients with persistent AF generally progress to permanent AF unless sinus rhythm is restored with cardioversion.

In addition to its association with aging, AF frequently occurs in association with hypertension (Chapter 67) or other comorbid conditions such as diabetes mellitus, thyrotoxicosis, heart failure, coronary artery disease, valvular heart disease, or lung disease such as chronic obstructive pulmonary disease or obstructive sleep apnea. About 20% of patients have no associated comorbidity and have what may be termed *lone AF*. Some patients develop AF after binges of alcohol use ("holiday heart") or after parasympathetic surges such as after vigorous exercise or a large meal (vagally induced AF). Excessive caffeine intake is an often cited but very rare cause of AF. AF requires a trigger in the form of atrial premature depolarizations, which often originate in the pulmonary veins and which, in the susceptible individual, result in AF. The susceptibility to AF may relate to changes in the electrical function of the left atrium. Prolonged periods of AF can lead to electrical and structural remodeling of the atria and promote the further perpetuation of AF ("AF begets AF").

AF results in the loss of the atrial contribution to ventricular filling. This so-called loss of atrial kick is generally well tolerated in normal individuals, in whom only about 15% of ventricular filling is the result of atrial contraction. In patients who have stiff, noncompliant ventricles (e.g., patients with aortic stenosis, hypertrophic or restrictive cardiomyopathy, or long-standing hypertension), however, up to 40% of ventricular filling may be related to atrial contraction, so stroke volume may fall noticeably in such patients.

The primary morbidity associated with AF is thromboembolism. Thromboembolism in patients with nonvalvular AF typically results from thrombus formation in and dislodgement from the left atrial appendage. The risk for thromboembolism is related to the presence of underlying vascular disease but not to the pattern of AF (paroxysmal or persistent).[4] For example, even subclinical AF is associated with a 2.5-fold increased risk for ischemic stroke or systemic embolism. However, the risk for embolization is further increased in the first 3 to 4 weeks after cardioversion, when the gradual return of atrial mechanical function can result in a particularly high risk for thromboembolism. The risk for thromboembolism in AF increases with age, diabetes mellitus, hypertension, previous embolic episodes, vascular disease, and heart failure. The lowest incidence (<1% annually) is in patients younger than 65 years with lone AF. AF is associated with about a 1.4-fold higher risk for cognitive impairment and dementia, with or without a history of clinical stroke.

CLINICAL MANIFESTATIONS

SVTs can produce symptoms specifically related to the rhythm itself, including fatigue, palpitations (Chapter 62), dizziness, shortness of breath, chest discomfort, presyncope, and syncope. These symptoms are mostly related to the rapidity of the ventricular response and are most prevalent at the onset of the arrhythmia. In contrast to other forms of SVT, asymptomatic episodes of AF are common in patients who also have symptomatic AF.

Incessant SVT and uncontrolled ventricular rates can cause tachycardia-related cardiomyopathy, which is reversible with control of these arrhythmias. In some situations, however, the clinical presentation may be dominated by the underlying condition that precipitates the arrhythmia, such as fever, physical stress, hypovolemia, heart failure (Chapter 58), hypoxia, sympathomimetic or parasympatholytic medications, thyrotoxicosis (Chapter 226), and pheochromocytoma (Chapter 228).

Re-entrant tachycardias begin and end abruptly, whether without treatment or when terminated with vagal maneuvers or intravenous medications. An accessory pathway with ventricular preexcitation can be undiagnosed into adulthood, when it can mimic myocardial infarction or right ventricular hypertrophy on the electrocardiogram.

DIAGNOSIS

The first key to diagnosis is the 12-lead ECG. A Holter monitor is useful if the arrhythmia is likely to be detected by 24 to 48 hours of monitoring, whereas a continuous loop event recorder, which can be worn for up to a month and activated by the patient for symptoms, is preferred if the arrhythmia is less frequent (Chapter 62). Another alternative is a device that records data continuously and transmits to a central station, capturing both symptomatic and asymptomatic arrhythmias over a period of up to 1 month.

On physical examination, patients with a blocked atrial beat or AVNRT may have large (cannon) "a" waves detected in their jugular veins (see Fig. 51-4). Carotid sinus massage, which has a vagal effect on the AV node, can terminate SVTs that depend on the AV node as part of the circuit (AVNRT and AVRT). Carotid sinus massage also can occasionally terminate atrial tachycardia but more often will slow the pulse by reducing the number of atrial inputs conducted through the AV node. Similarly, carotid sinus massage will slow but not terminate atrial flutter and AF.

Adenosine (6 mg rapidly intravenously in 1 to 3 seconds) slows the heart rate in sinus tachycardia, atrial tachycardia, AF, or atrial flutter, and can terminate AV nodal re-entrant tachycardia, AV reciprocating tachycardia, and some atrial tachycardias. It should not be used in patients with wide, irregular QRS because it may exacerbate the AF of Wolff-Parkinson-White syndrome.

Electrophysiologic testing (Chapter 66) is the definitive way to distinguish SVT from ventricular tachycardia (see Table 65-2). The SVT can often be initiated by paced premature beats, after which the mechanism and location of the arrhythmia can be determined.

TREATMENT

Rx

Acute Therapy

Sinus, Atrial, Atrioventricular Nodal Re-entrant, and Junctional Tachycardias

Sinus tachycardia rarely should be treated directly. Instead, treatment should focus on identifying and treating any precipitating underlying conditions, especially heart failure, pulmonary disease, fever, anemia, and thyroid disease. In the occasional patients who require treatment for symptoms associated with inappropriate sinus tachycardia, ivabradine 5 mg twice daily provides significant improvement and completely eliminates symptoms in approximately half of the patients.

Multifocal atrial tachycardia is relatively unresponsive to medical therapy, is very difficult to ablate, and is best treated by management of the underlying pulmonary disorder. Nondihydropyridine calcium-channel blocking drugs (diltiazem, verapamil; Table 64-5) are most often used if a therapy is absolutely required.

Sustained or repeated episodes of nonsustained SVT, however, generally require effective therapy. If rapid control is desired (e.g., in patients with myocardial ischemia or hypotension), cardioversion is the best solution (Chapter 66). Atrial tachycardias, including atrial flutter or AF, may also convert spontaneously or convert after treatment of an underlying cause, such as hypoxia or heart failure, or after cessation of precipitating medications.

TABLE 64-5 ANTIARRHYTHMIC DRUGS: DOSES AND SIDE EFFECTS

ANTIARRHYTHMIC DRUG AND COMMON USE	DOSE/METABOLISM	SIDE EFFECTS AND REQUIRED MONITORING	SELECTED DRUG INTERACTIONS
Ivabradine (not available in the U.S.)	2.5 mg starting dose, 5 mg bid for 1 month followed by increase to 7.5 bid if needed	Sinus bradycardia Headaches Visual brightness	Azole antifungals, macrolide antibiotics, nefazadone, nelfinavir, ritonavir
Quinidine	Hepatic CYP 3A4 (70%), renal (30%) Dose: sulfate—600 mg tid, gluconate—324 to 648 mg q8h Dose reduced for renal failure	Thrombocytopenia Cinchonism Pruritus, rash QT prolongation/torsades de pointes	↑ Digoxin and amiodarone concentrations Quinidine inhibits CYP 2D6 and may increase drugs metabolized by this enzyme, e.g., ↑ effect of tricyclic antidepressants, haloperidol, some β-blockers, fluoxetine, narcotics Quinidine metabolism inhibited by cimetidine Quinidine metabolism increased by phenobarbital, phenytoin, and rifampicin
Procainamide	Mostly hepatic—rapid acetylators produce more NAPA; NAPA renally cleared PO dose: 50 mg/kg/24 hr IV dose: 1 g over 25 min, then 20-60 μg/kg/min infusion Reduce dose for renal dysfunction or low cardiac output	Rash, fever, arthralgias, drug-induced lupus, particularly in slow acetylators Agranulocytosis QT prolongation/torsades de pointes	Procainamide clearance reduced by trimethoprim, cimetidine, and ranitidine
Disopyramide	Renal, hepatic (CYP 3A4) Dose: 100-400 mg q8-12h; max dose, 800 mg/24 hr Reduce dose for renal or hepatic dysfunction	Anticholinergic (contraindicated for narrow-angle glaucoma): dry mouth, urinary retention, constipation, blurry vision QT prolongation/torsades de pointes	None
Propafenone	Hepatic: 150-300 mg q8h or sustained release 225-425 mg bid	Metallic taste, dizziness, SIADH Atrial flutter, ventricular tachycardia	May decrease the metabolism of warfarin Increase digoxin levels
Flecainide	Renal, hepatic CYP 2D6 50-100 mg bid; max dose, 300-400 mg/day	Dizziness, headache, visual blurring Atrial flutter, ventricular tachycardia	May increase digoxin levels Flecainide levels increased by amiodarone, haloperidol, quinidine, cimetidine, and fluoxetine

TABLE 64-5 ANTIARRHYTHMIC DRUGS: DOSES AND SIDE EFFECTS—cont'd

ANTIARRHYTHMIC DRUG AND COMMON USE	DOSE/METABOLISM	SIDE EFFECTS AND REQUIRED MONITORING	SELECTED DRUG INTERACTIONS
β-Blockers (selected)	Hepatic, renal Only renal (atenolol, nadolol) IV esmolol: 250-500 µg over 1 min, then 50-300 µg/kg/min over 4 min Acebutolol, 200-600 mg bid; atenolol, 25-100 mg qd; carvedilol, 3.125-50 mg bid; metoprolol, 25-150 mg bid; nadolol, 20-120 mg qd; nebivolol, 5-40 mg qd; propranolol, 10-120 mg bid	Fatigue, depression, bronchospasm, impotence	Minimal, except for carvedilol and metoprolol, whose levels may be increased by amiodarone, propafenone, quinidine, fluoxetine, haloperidol, paroxetine, and cimetidine
Sotalol	Renal: 80-120 mg bid Max dose, 240 mg bid	Bronchospasm QT prolongation/torsades de pointes	No significant interactions
Dofetilide	Renal, hepatic CYP 3A4 CrCl > 60 (500 µg bid), CrCl 40-60 (250 µg bid), CrCl 20-39 (125 µg bid)	QT prolongation and torsades de pointes Three days of in-hospital monitoring is required during drug initiation	Contraindicated with verapamil, ketoconazole, cimetidine, megestrol, prochlorperazine, and trimethoprim Hydrochlorothiazide increases dofetilide levels Must discontinue amiodarone at least 3 mo before dofetilide initiation
Ibutilide	Hepatic CYP 3A4 1 mg IV over 10 min, repeat after 10 min if necessary	Nausea QT prolongation and torsades de pointes Must monitor for 4 hr after drug initiation	None
Amiodarone	Hepatic half-life 50 days PO load 10 g over 7-10 days, then 400 mg for 3 wk, then 200 mg/day for atrial fibrillation Maintenance dose of 400 mg/day for VT Dose reduce load for bradycardia or QT prolongation IV: 150-300 mg bolus, then 1 mg/min infusion for 6 hr, followed by 0.5 mg/min thereafter	Pulmonary (acute hypersensitivity pneumonitis, chronic interstitial infiltrates), hepatitis Thyroid (hypo- or hyperthyroidism) Photosensitivity, blue-gray discoloration with chronic high dose, nausea, ataxia, tremor, alopecia Avoid if identified thyroid nodule LFTs two to three times a year, TFTs twice yearly, PFTs and CXR at initiation and CXR yearly thereafter. QT prolongation expected; reduce dose if exceeds 500 msec	Inhibits CYP 450 enzymes—increases concentrations of warfarin, digoxin, cyclosporine, alprazolam, carbamazepine, HMG-CoA inhibitors, phenytoin, and quinidine
Dronedarone	Hepatic CYP 3A4 half-life 30 hr PO 400 mg bid Improved absorption with food	Reduces the secretion of creatinine without a reduction in GFR Hepatic failure Avoid in heart failure	Increases digoxin levels (dose reduce digoxin by half) May increase myositis with simvastatin Avoid grapefruit
Calcium-channel blocker (nondihydropyridine)	Hepatic Inhibit CYP 3A4 IV diltiazem, 20 mg bolus over 2 min, then 5-15 mg per hour maintenance infusion Verapamil long-acting 120-480 mg qd Diltiazem long-acting 180-300 mg qd	Constipation, rash, peripheral edema	Inhibits CYP 3A4—will increase levels of alprazolam, carbamazepine, dihydropyridine, cyclosporine, HMG-CoA inhibitors. Verapamil (but not diltiazem) increases digoxin levels
Adenosine	Erythrocyte, endothelial cell 6-mg IV push, followed if necessary by 12 mg after 1-2 min	Nausea, headache, flushing, chest pain, bronchospasm (contraindicated if asthma)	Methylxanthines compete for adenosine receptors with adenosine Dipyridamole decreases the metabolism of adenosine
Digoxin	Renal, hepatic, gastrointestinal, 0.125-375 mg/day	Anorexia, nausea, fatigue, confusion, altered vision with green-yellow halos	Levels of or sensitivity to digoxin increased by hypokalemia, quinidine, verapamil, amiodarone, propafenone, renal failure, hypoxia, decreased muscle mass Levels of or sensitivity to digoxin decreased by malabsorption, hyperkalemia, hypocalcemia

CXR = chest x-ray; CYP = cytochrome P-450; GFR = glomerular filtration rate; HMG-CoA = 3-hydroxy-3-methylglutaryl coenzyme A; IV = intravenous administration; LFT = liver function test; NAPA = *N*-acetyl procainamide; PFT = pulmonary function test; PO = oral administration; SIADH = syndrome of inappropriate diuretic hormone; TFT = thyroid function test; VT = ventricular tachycardia.

An acute episode of AVNRT can often be terminated with vagal maneuvers, such as carotid massage. In most atrial tachycardias, adenosine or vagal stimulation (or both) produces enough AV block to unmask the atrial origin of the tachycardia. However, some atrial tachycardias and most episodes of AVNRT terminate after administration of adenosine. Intravenous β-blockers or calcium-channel blockers (see Table 64-5) can be used for the same purpose. For sustained control of the ventricular rate during atrial tachycardia, intravenous esmolol and diltiazem are effective.

Accessory Pathway Tachycardia

The acute management of SVTs that use an accessory pathway (AVRT) depends on the mechanism of the rhythm. If the tachycardia uses the AV node as one limb of the circuit, it may terminate with vagal maneuvers such as carotid sinus massage. If the tachycardia uses an accessory pathway as the antegrade limb (antidromic AVRT), it is important to avoid measures that may accelerate the conduction properties of the accessory pathway. For example, digoxin and epinephrine will increase conduction over the accessory pathway and potentially accelerate the tachycardia. Patients may have an adrenergic response to a drop in blood pressure induced by the vasodilating properties of β-blockers and calcium-channel blockers, particularly when AF conducts over an accessory pathway. As a result of these potential complications, adenosine (see Table 64-5) is the drug of choice for the acute termination of antidromic AVRT, but a true antiarrhythmic drug, such as intravenous amiodarone, or cardioversion should be considered for antidromically conducted AF. In

orthodromic AVRT, in which the AV node is the antegrade limb of the tachycardia, β-blockers or calcium-channel blockers can be used as well as adenosine.

Atrial Flutter

It can be very difficult to control the rate of acute atrial flutter, which generally must be treated by restoring sinus rhythm. Intravenous ibutilide is approximately 60% effective for converting atrial flutter to sinus rhythm. Direct current electrical cardioversion, which is highly (>95%) effective for restoring sinus rhythm, should not be performed unless the episode of atrial flutter is believed to be less than 48 hours in duration, or transesophageal echocardiogram has excluded clot in the left atrial appendage, or until the risk for stroke has been minimized by achieving a therapeutic international normalized ratio (INR) of 2 to 3 with warfarin or administration of a therapeutic dose of dabigatran, rivaroxaban, or apixaban (Chapter 38) for the prior 4 weeks.

Atrial Fibrillation

For acute AF without hypotension, rate control is crucial and can be accomplished with esmolol, metoprolol, verapamil, or diltiazem (see Table 64-5); digoxin is usually a third-line agent (Fig. 64-23). Most patients who have new-onset AF who are considered good candidates for cardioversion should be anticoagulated with heparin, warfarin, or a therapeutic dose of dabigatran, rivaroxaban, or apixaban (Chapter 38).[5,6] One half of patients with acute-onset AF will spontaneously revert to sinus rhythm within for the first 48 to 96 hours. The decision of whether to restore and maintain sinus rhythm or allow recurrences or progression to permanent AF is a fundamental component of AF management. Trials of a strategy of rate control compared with a strategy of rhythm control with antiarrhythmic drugs have demonstrated no difference in total or arrhythmic mortality associated with these two approaches,[■] even in patients with a reduced ejection fraction.[■] Because a significant component of the adverse events experienced in the rhythm control arms of these studies was due to stroke in un-anticoagulated patients and toxicity from antiarrhythmic drugs, advances in antiarrhythmic and anticoagulant medications as well as the advent of reliable ablative therapies for AF may necessitate a reevaluation of this question.

If a strategy of rate control is chosen, it is important to confirm a heart rate of 80 to 110 beats per minute at rest and less than 140 beats per minute with exercise, preferably by monitoring the heart rate during exercise on an exercise treadmill test or with an ambulatory monitor. More strict rate control is not beneficial.[■] Failure to confirm rate control can result in the development of tachycardia-induced cardiomyopathy. First-line therapy for rate control includes β-blockers or calcium-channel blockers; digoxin can also be used but is generally less effective. Patients commonly require a combination of mediations to achieve goal heart rates.[7]

If a strategy of rhythm control is chosen, many patients will first require cardioversion, either pharmacologic or electrical (Chapter 66). The risk for clot formation must be mitigated before cardioversion in all patients with AF of more than 48 hours' duration. The first step generally is to perform transesophageal echocardiography (TEE) (Chapter 55). If TEE shows no evidence of a left atrial clot, cardioversion can be undertaken without systemic anticoagulation; if the patient has risk factors for stroke in association with AF, however, most clinicians administer anticoagulation during the cardioversion and for the subsequent 4 weeks. If the TEE shows evidence of clot, 4 consecutive weeks of warfarin anticoagulation with an INR of at least 2, or equivalent anticoagulation with therapeutic doses of dabigatran, rivaroxaban, or apixaban is required; anticoagulation must be maintained for at least 3 to 4 weeks after cardioversion. Electrical cardioversion, which should be performed with a minimum of 200 joules, is successful in more than 90% of cases. Pharmacologic cardioversion can be performed with intravenous drugs such as ibutilide, which is more successful for atrial flutter (60% efficacy) than AF (50%). Oral medications can also be used as a "pill in the pocket" strategy. Patients can take a single dose of propafenone (600 mg) or flecainide (300 mg) with a conversion rate for recent-onset AF (<48 hours' duration) of approximately 50% without the need for a screening TEE. Oral amiodarone (typically loaded with 10 g over the first week, followed by 400 to 600 mg a day for the next 3

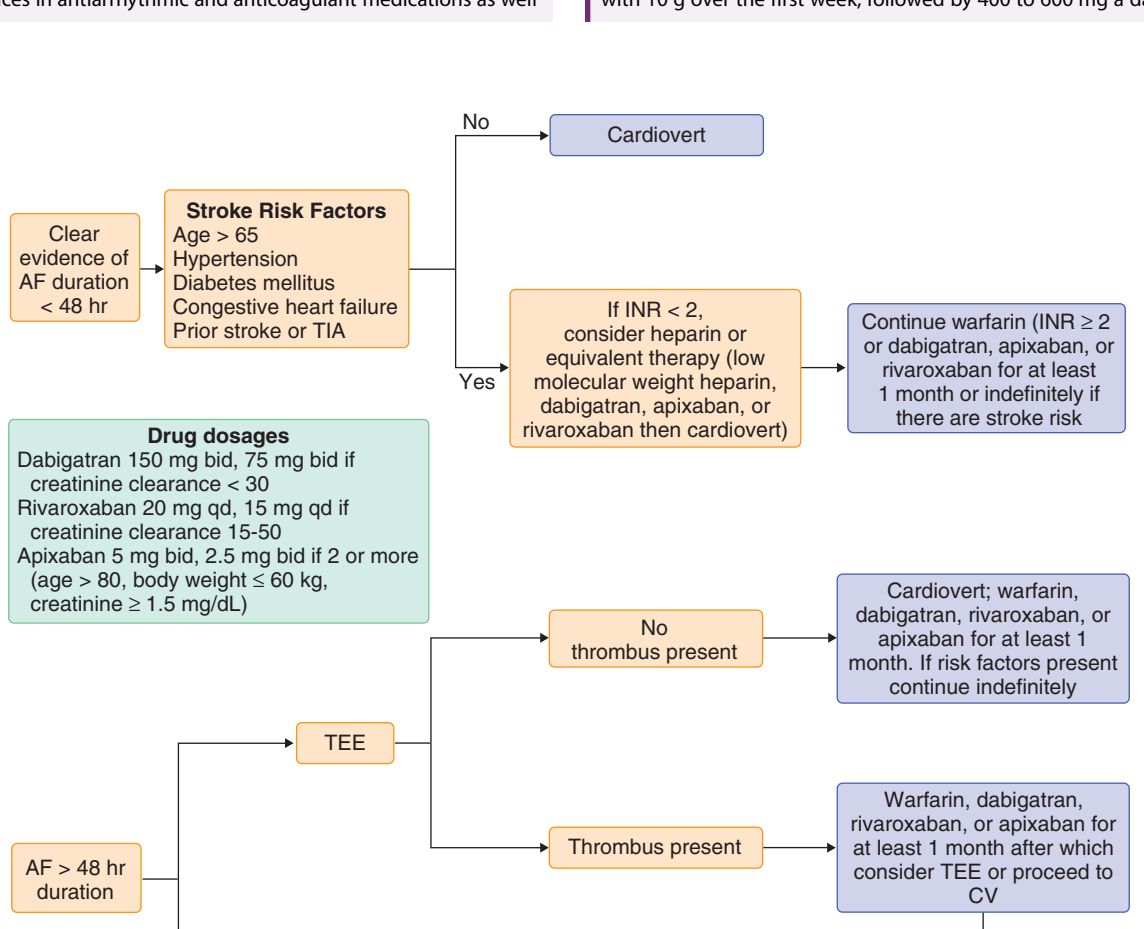

FIGURE 64-23. Management of recent-onset atrial fibrillation (AF). CV = cardioversion; INR = international normalized ratio; TEE = transesophageal echocardiography; TIA, transient ischemic attack.

weeks) can also be used for cardioversion and is successful in approximately 50% of patients with both recent and more prolonged AF.

Long-Term Management

AVNRT, AVRT, Atrial Tachycardias, and Atrial Flutter

Chronic therapy for AVNRT is guided by the frequency and severity of symptoms.[8] Many patients are able to live with this rhythm with infrequent recurrences, which terminate spontaneously or with adenosine. If chronic therapy is required, β-blockers or calcium-channel blockers and less commonly digoxin are used. Ablation of AVNRT (Chapter 66) is highly effective and should be considered before using sodium- or potassium-channel blocking drugs.

Most patients with symptomatic AVRT are treated with catheter ablation (Chapter 66). Ablation of an accessory pathway located near the AV node or His bundle carries a 1% risk for complete heart block, whereas ablation of accessory pathways on the left side of the heart and distant from the AV node and His bundle region is not associated with a risk for heart block but carries a small risk for stroke. At present, it is not standard of care to ablate accessory pathways in patients without symptomatic arrhythmias.[9,10]

The long-term management of atrial tachycardia depends on symptoms. If the rhythm is highly symptomatic, it is generally managed with a β-blocker or calcium-channel blocker. If these medications are unsuccessful or not tolerated, ablation is frequently recommended, but antiarrhythmic medications are an alternative.

In patients with atrial flutter, ventricular rate control is possible by achieving AV nodal block with β-blockers, calcium-channel blockers, and digitalis. However, radiofrequency ablation, which is curative, is now the preferred choice for most patients with atrial flutter (Chapter 66), especially recurrent atrial flutter. Because atrial flutter carries a 3% per year risk for thromboembolism, patients with atrial flutter should also receive long-term anticoagulation similar to what is recommended for AF (see later). If atrial flutter is successful, the risk for recurrence is very small, and long-term anticoagulation is not necessary.

Atrial Fibrillation

Therapies for the chronic maintenance of sinus rhythm in patients with AF include pharmacologic and procedural approaches. The procedural approaches include catheter-based ablation inside the left atrium with the goal of electrically isolating the pulmonary veins from the left atrium. Similarly, a minimally invasive surgical approach can electrically isolate the pulmonary veins from the external surface of the heart with the additional resection of the left atrial appendage. Both these procedures have become standard options for AF, especially in patients who have recurrent AF despite at least one antiarrhythmic drug.

The catheter approach carries a small risk for cardiac perforation, including pericardial tamponade and atrioesophageal fistula formation, and a 1% risk for stroke. There is also a small risk for pulmonary vein stenosis, which has been reduced by newer technologies. A second procedure typically is offered to patients who have recurrent AF following a first catheter-based procedure.

In randomized trials of patients with paroxysmal AF, the cumulative burden of AF over a period of 2 years appeared to be slightly lower with initial radiofrequency catheter ablation therapy compared with antiarrhythmic medications, but at the expense of procedural risks and without any differences in patient-reported quality of life. Catheter ablation may, however, be preferred as initial therapy in patients with persistent AF and symptomatic heart failure. For patients with long-standing persistent AF, 5-year success rates are 20% for a single ablation procedure and 45% for multiple ablation procedures.

The surgical approach carries a higher risk for cardiac bleeding, particularly during the resection of the left atrial appendage, and is associated with a significantly longer recovery time than the percutaneous approach. However, there should be no stroke risk associated with the surgical procedure because it is performed completely from the epicardial surface of the heart. A more extensive surgical operation, called the maze procedure, requires a full thoracotomy and is most often performed concomitantly as part of open coronary artery bypass surgery or an open valve operation. In this procedure, electrical lines of block are created in the left atrium to interrupt the perpetuation of AF, the pulmonary veins are isolated, and the left atrial appendage is resected. Success rates for this procedure, which should be reserved for refractory, symptomatic AF, exceed 80%.

The pharmacologic options for the treatment of AF work by blocking sodium, potassium, or a combination of cardiac channels. Blockade of these channels results in slowing of cardiac conduction (sodium channels) and prolongation in cardiac repolarization (potassium channels) as well as additional effects from modulation of the autonomic nervous system. The choice of antiarrhythmic drug is based on the patient's underlying clinical condition (Table 64-6).

Amiodarone is the most widely used medication for AF with an efficacy of 60 to 70% at 1 year. It is associated with a number of drug interactions, most notably with warfarin and digoxin. Its associated risk for thyroid, liver, and lung toxicities, related in part to the iodine moieties on this compound, necessitate careful follow-up. For prevention of recurrent AF, oral amiodarone is significantly more effective than propafenone, flecainide, dofetilide, or sotalol, which are the recommended alternatives. Dronedarone (400 mg twice daily), which is related to amiodarone but has no iodine and a 24-hour half-life, is well tolerated in terms of noncardiovascular side effects but has been associated with an increased risk for heart failure, stroke, and death in patients with permanent AF. As a result, it should be discontinued in patients in whom sinus rhythm is not well maintained.

Quinidine, procainamide, and disopyramide are predominantly sodium-channel blocking drugs that also block potassium channels at slow heart rates. Each of these drugs is moderately successful in AF, with about 50% of treated patients in sinus rhythm at 1 year, but each also has idiosyncratic noncardiovascular toxicities that can significantly limit their utility (see Table 64-5). Propafenone and flecainide are also sodium-channel blockers that are widely used for the maintenance of sinus rhythm. These drugs are moderately effective, with a 50% rate of sinus rhythm at 1 year, and are generally well tolerated but must be avoided in patients with structural heart disease, particularly with a history of prior myocardial infarction and impaired left ventricular function, because of a risk for drug-induced ventricular arrhythmia. Dofetilide is a potassium-channel blocking medication that is moderately effective for suppressing AF but carries a dose-dependent risk for QT prolongation and torsades de pointes.

TABLE 64-6 SELECTION OF ANTIARRHYTHMIC DRUGS

PATIENT CHARACTERISTICS	ANTIARRHYTHMIC DRUG CHOICES
No structural heart disease	*First line:* flecainide, propafenone, dronedarone, sotalol *Second line:* amiodarone, dofetilide
Depressed left ventricular ejection fraction with heart failure	*First line:* amiodarone, dofetilide *Avoid:* dronedarone, flecainide, propafenone
Coronary artery disease without congestive heart failure	*First line:* sotalol, dronedarone, dofetilide, amiodarone *Avoid:* flecainide, propafenone
Hypertrophic cardiomyopathy	*First line:* amiodarone, sotalol, dronedarone *Second line:* disopyramide

TABLE 64-7 CURRENT RECOMMENDATIONS FOR THROMBOEMBOLIC PROPHYLAXIS FOR PATIENTS WITH ATRIAL FIBRILLATION BASED ON RISK FACTORS FOR STROKE

RISK FACTORS*	RECOMMENDATIONS
Heart failure (1 point) Hypertension (1 point) Age ≥65 (1 point), ≥75 (2 points) Diabetes (1 point) Stroke/TIA (2 points) Vascular disease (1 point) Female gender (1 point)	2 or more points: anticoagulation with warfarin or a new oral anticoagulant 1 point: anticoagulation or no therapy depending on the preference of the patient and treating physician 0 points: no therapy

*Based on the CHA_2DS_2-VASc risk stratification scoring system.
TIA = transient ischemic attack.
Data from January CT, Wann S, Alpert JS, et al. 2014 AHA/ACC/HRS Guideline for the Management of Patients with Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. *J Am Coll Cardiol* 2014;64:e1-e76.

Anticoagulation

The presence or absence of associated conditions helps determine which patients with AF require chronic anticoagulation with warfarin or other systematic coagulants (Table 64-7). Long-term anticoagulation therapy with warfarin, dabigatran, rivaroxaban, or apixaban is generally recommended in all patients who have persistent or paroxysmal AF, who are older than 65 years, and who have no contraindications to anticoagulation.[4] Anticoagulation also should be maintained for 6 months after both catheter and surgical procedures in patients without clinical risk factors for stroke and chronically in patients with risk factors. Catheter-based procedures directed at excluding the left atrial appendage from the systemic blood stream may become options in patients who have a high risk for stroke and who cannot tolerate systemic anticoagulation owing to an excessive risk for bleeding.

Warfarin alone is superior to aspirin or the combination of clopidogrel and aspirin, with meta-analysis showing that adjusted-dose warfarin and

antiplatelet agents reduce stroke by approximately 60% and 20%, respectively.[A] Although there is some protective effect at an INR as low as 1.8, the target INR for chronic anticoagulation with warfarin should be 2 to 3 to avoid INRs less than 1.8. Guidelines no longer recommend aspirin or other antiplatelet agents in patients without an indication for warfarin or the newer anticoagulants.

New oral anticoagulant medications have the potential to replace warfarin as more effective and safer (except for gastrointestinal bleeding)[A] primary therapy to prevent systemic emboli in patients with AF. In a randomized trial of patients with nonvalvular atrial fibrillation, rivaroxaban (an oral factor Xa inhibitor at 20 mg per day) was better than warfarin at preventing stroke or systemic embolization, with significantly less intracranial and fatal bleeding.[A] In another randomized trial of patients with atrial fibrillation, apixaban (an oral factor Xa inhibitor at 5 mg twice daily) prevented more strokes and systemic emboli than warfarin, with less bleeding from all causes and fewer deaths.[A] Dabigatran, a direct thrombin inhibitor (150 mg twice daily), is superior to warfarin for preventing thromboembolism, with a lower risk for intracranial bleeding but a slightly higher risk for extracranial bleeding.[A] All three drugs are eliminated by the kidney (apixaban 25%, rivaroxaban 65%, and dabigatran 85%), so they are not recommended in patients with substantial renal dysfunction, and the doses should be reduced in patients with moderate renal dysfunction (Chapter 38). A reasonable approach is to use apixaban in patients at the highest risk for bleeding, to use rivaroxaban in patients who prefer once-daily dosing, and to avoid dabigatran in patients older than 80 years because of increased bleeding risk. The addition of aspirin to moderate-intensity warfarin (INR 2 to 3) or to dabigatran, rivaroxaban, or apixaban can decrease vascular events and is recommended, despite its increased risk of causing bleeding, in some AF patients with concomitant risk factors, such as coronary artery disease or a prior stroke that is attributed to vascular disease rather than to AF.

Grade A References

A1. Curtis AB, Worley SJ, Adamson PB, et al. Biventricular pacing for atrioventricular block and systolic dysfunction. *N Engl J Med*. 2013;368:1585-1593.

A2. Cappato R, Castelvecchio S, Ricci C, et al. Clinical efficacy of ivabradine in patients with inappropriate sinus tachycardia: a prospective, randomized, placebo-controlled, double-blind, crossover evaluation. *J Am Coll Cardiol*. 2012;60:1323-1329.

A3. Al-Khatib SM, Allen LaPointe NM, Chatterjee R, et al. Rate- and rhythm-control therapies in patients with atrial fibrillation: a systematic review. *Ann Intern Med*. 2014;160:760-773.

A4. Roy D, Talajic M, Nattel S, et al. Rhythm control versus rate control for atrial fibrillation and heart failure. *N Engl J Med*. 2008;358:2667-2677.

A5. Van Gelder I, Groenveld H, Crijns H. Lenient versus strict rate control in patients with atrial fibrillation. *N Engl J Med*. 2010;362:1363-1373.

A6. Wilber DJ, Pappone C, Neuzil P, et al. Comparison of antiarrhythmic drug therapy and radiofrequency catheter ablation in patients with paroxysmal atrial fibrillation: a randomized controlled trial. *JAMA*. 2010;303:333-340.

A7. Morillo CA, Verma A, Connolly SJ, et al. Radiofrequency ablation vs antiarrhythmic drugs as first-line treatment of paroxysmal atrial fibrillation (RAAFT-2): a randomized trial. *JAMA*. 2014;311:692-700.

A8. Cosedis Nielsen J, Johannessen A, Raatikainen P, et al. Radiofrequency ablation as initial therapy in paroxysmal atrial fibrillation. *N Engl J Med*. 2012;367:1587-1595.

A9. Jones DG, Haldar SK, Hussain W, et al. A randomized trial to assess catheter ablation versus rate control in the management of persistent atrial fibrillation in heart failure. *J Am Coll Cardiol*. 2013;61:1894-1903.

A10. Reddy VY, Möbius-Winkler S, Miller MA, et al. Left atrial appendage closure with the Watchman device in patients with a contraindication for oral anticoagulation: the ASAP study (ASA Plavix Feasibility Study With Watchman Left Atrial Appendage Closure Technology). *J Am Coll Cardiol*. 2013;61:2551-2556.

A11. Hart RG, Pearce LA, Aguilar MI, et al. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Ann Intern Med*. 2007;146:857-867.

A12. Ruff CT, Giugliano RP, Braunwald E, et al. Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. *Lancet*. 2014;383:955-962.

A13. Patel MR, Mahaffey KW, Garg J, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med*. 2011;365:883-891.

A14. Granger CB, Alexander JH, McMurray JJ, et al. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2011;365:981-992.

A15. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2009;361:1139-1151.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Brignole M, Auricchio A, Baron-Esquivias G, et al. 2013 ESC guidelines on cardiac pacing and cardiac resynchronization therapy: the task force on cardiac pacing and resynchronization therapy of the European Society of Cardiology (ESC). Developed in collaboration with the European Heart Rhythm Association (EHRA). *Europace*. 2013;15:1070-1118.
2. Epstein AE, DiMarco JP, Ellenbogen KA, et al. 2012 ACCF/AHA/HRS focused update incorporated into the ACCF/AHA/HRS 2008 guidelines for device-based therapy of cardiac rhythm abnormalities: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. *J Am Coll Cardiol*. 2013;61:e6-e75.
3. Zimetbaum P. Antiarrhythmic drug therapy for atrial fibrillation. *Circulation*. 2012;125:381-389.
4. Wann LS, Curtis AB, Ellenbogen KA, et al. Management of patients with atrial fibrillation (compilation of 2006 ACCF/AHA/ESC and 2011 ACCF/AHA/HRS recommendations): a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *Circulation*. 2013;127:1916-1926.
5. January CT, Wann LS, Alpert JS, et al. 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. *J Am Coll Cardiol*. 2014;64:e1-e76.
6. Heidbuchel H, Verhamme P, Alings M, et al. EHRA practical guide on the use of new oral anticoagulants in patients with non-valvular atrial fibrillation: executive summary. *Eur Heart J*. 2013;34:2094-2106.
7. Verma A, Cairns JA, Mitchell LB, et al. 2014 focused update of the Canadian Cardiovascular Society guidelines for the management of atrial fibrillation. *Can J Cardiol*. 2014;30:1114-1130.
8. Link MS. Evaluation and initial treatment of supraventricular tachycardia. *N Engl J Med*. 2012;367:1438-1448.
9. Cohen MI, Triedman JK, Cannon BC, et al. PACES/HRS expert consensus statement on the management of the asymptomatic young patient with a Wolff-Parkinson-White (WPW, ventricular preexcitation) electrocardiographic pattern: developed in partnership between the Pediatric and Congenital Electrophysiology Society (PACES) and the Heart Rhythm Society (HRS). Endorsed by the governing bodies of PACES, HRS, the American College of Cardiology Foundation (ACCF), the American Heart Association (AHA), the American Academy of Pediatrics (AAP), and the Canadian Heart Rhythm Society (CHRS). *Heart Rhythm*. 2012;9:1006-1024.
10. Pappone C, Vicedomini G, Manguso F, et al. Wolff-Parkinson-White syndrome in the era of catheter ablation: insights from a registry study of 2169 patients. *Circulation*. 2014;130:811-819.

REVIEW QUESTIONS

1. Supraventricular tachyarrhythmias. A 25-year-old woman presents to the emergency department with palpitations and the attached electrocardiogram. Which of the following statements is false?

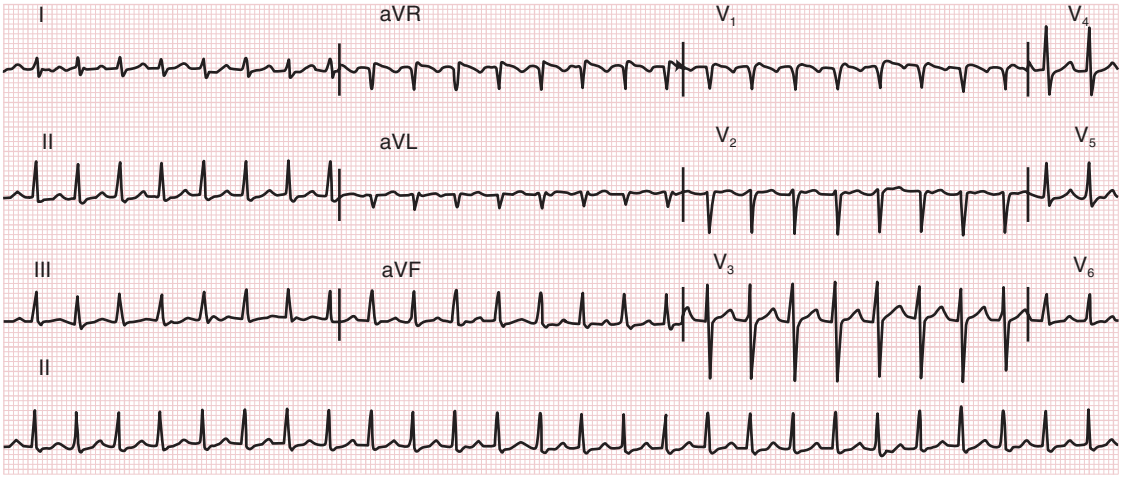

A. The rhythm can be treated with intravenous diltiazem.
B. The rhythm is best treated with adenosine.
C. The rhythm is curable with radiofrequency ablation.
D. The rhythm is most consistent with atrial flutter.
E. The rhythm is most consistent with atrioventricular (AV) nodal re-entrant tachycardia.

Answer: D The rhythm is AV nodal re-entrant tachycardia given the absence of evident P waves. This rhythm is typical in women in this age group. It can be treated acutely with adenosine or diltiazem, and long-term treatment can include a curative ablation procedure. The absence of flutter waves makes atrial flutter unlikely.

2. Supraventricular tachyarrhythmias. A 54-year-old man presents to the emergency department with the new onset of palpitations and shortness of breath. The accompanying electrocardiogram is obtained. What is the diagnosis?

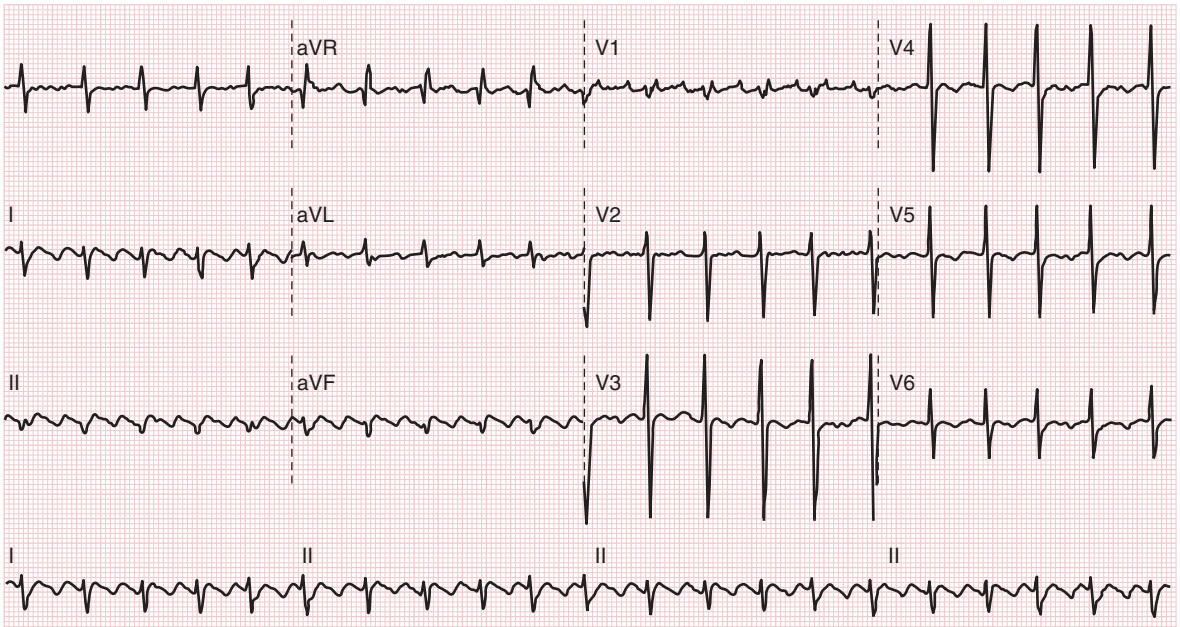

A. AV nodal re-entrant tachycardia
B. AV re-entrant tachycardia
C. Atrial flutter
D. Atrial fibrillation
E. Atrial tachycardia

Answer: C The electrocardiogram demonstrates negative ("sawtooth") waves in leads 2, 3, and aVF, positive flutter waves in V_1 and negative in V_6. There are two flutter waves for every QRS complex, consistent with typical atrial flutter with 2 : 1 conduction.

3. Atrioventricular conduction disturbances. An 18-year-old man presents to the hospital with fatigue for the prior week. It is late August and he has spent the summer on Nantucket as a life guard. His pulse is 40 beats per minute, and the accompanying electrocardiogram is obtained. What is the most likely diagnosis?

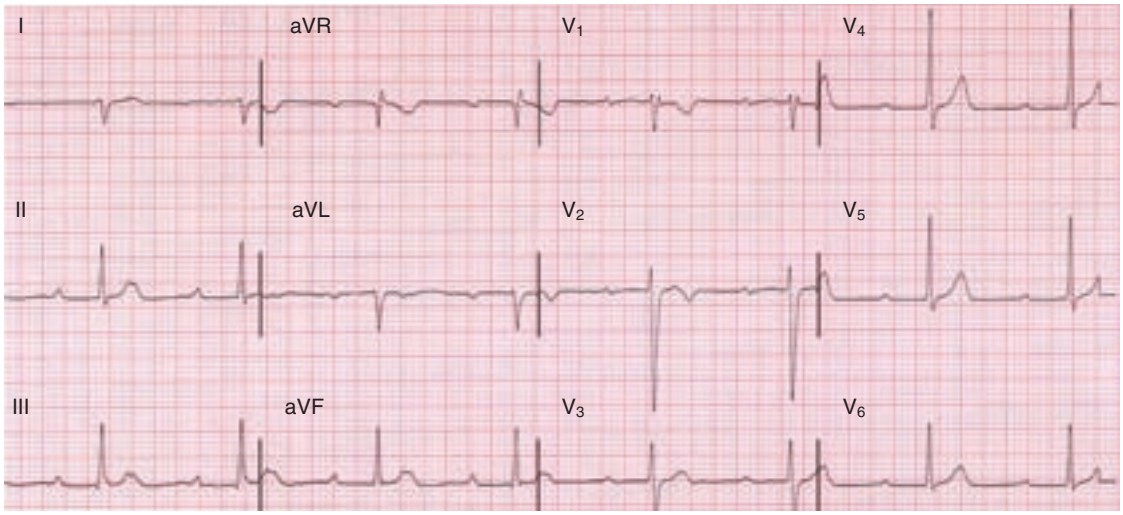

A. Sinus bradycardia
B. AV Wenckebach
C. Complete heart block
D. 2 : 1 AV block
E. Junctional rhythm

Answer: D The rhythm shows 2 : 1 AV block with a long PR interval associated with the conducted P wave. The clinical history is consistent AV block secondary to Lyme disease. It should be treated with antibiotics, and in most cases there will be resolution of conduction disease.

4. Supraventricular tachycardia. An 18-year-old man presents with a syncopal episode while playing basketball. He is brought to the emergency department where the following electrocardiogram is recorded. His blood pressure is stable. What is the diagnosis?

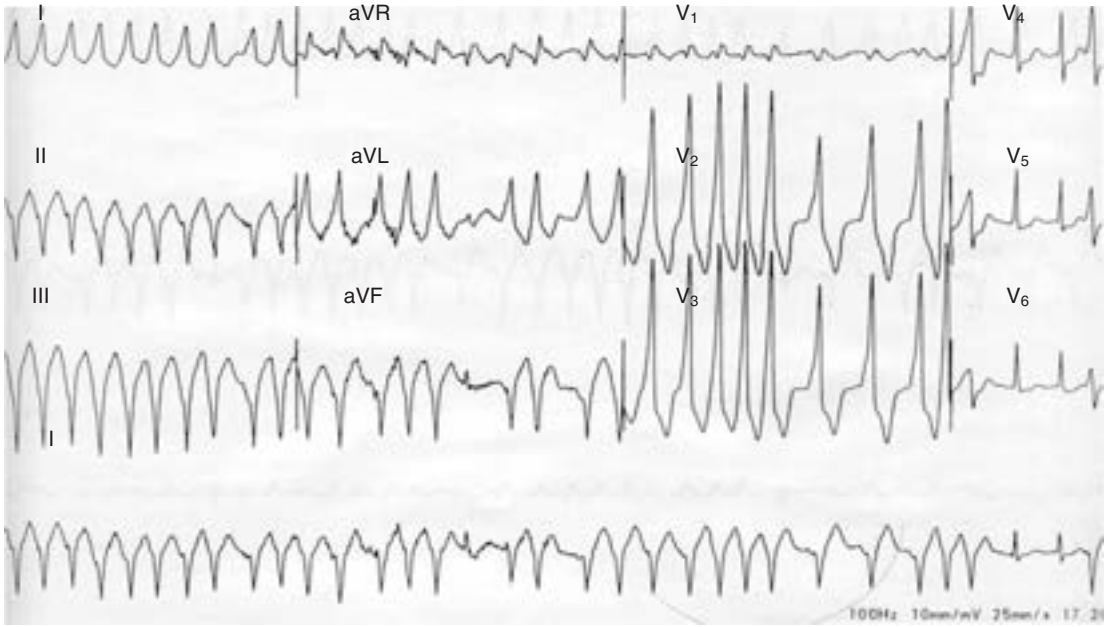

A. Ventricular tachycardia
B. AV nodal re-entrant tachycardia with aberration
C. Atrioventricular tachycardia with aberration
D. Atrioventricular tachycardia without aberration
E. Atrial fibrillation with conduction over an accessory pathway (bypass tract)

Answer: E The rhythm is irregularly irregular, consistent with atrial fibrillation, and the variable QRS durations are consistent with varying amounts of conduction over an accessory pathway. This classic fast, broad, and irregular pattern is consistent with atrial fibrillation conducted over an accessory pathway.

5. Bradyarrhythmias. A 45-year-old man complains of daytime somnolence. Evaluation is notable for morbid obesity and hypertension. He takes 20 mg of lisinopril for hypertension. The following telemetry strip occurred during sleep and was not associated with symptoms. During daytime hours, his heart rate never dropped below 70 beats per minute. Which statement is true?

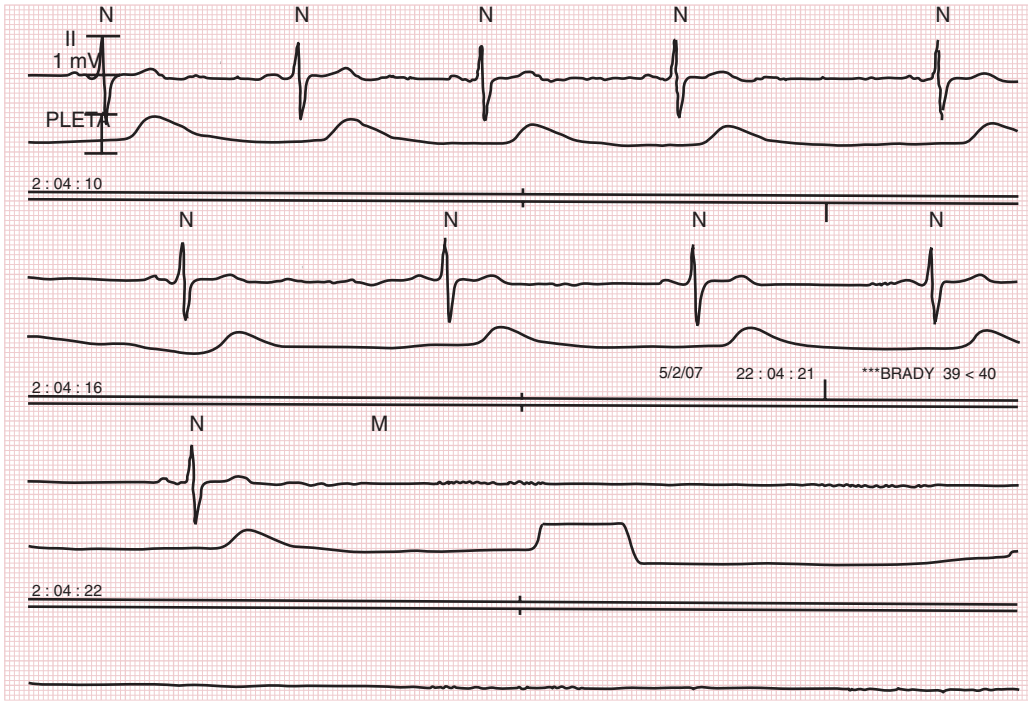

A. The telemetry strip shows complete heart block.
B. The telemetry strip shows sinus node dysfunction, which requires a pacemaker.
C. The telemetry strip shows sinus bradycardia progressing to sinus arrest, consistent with a vagal mechanism.
D. Sinus bradycardia is due to metoprolol.
E. The telemetry strip shows sinus arrhythmia.

Answer: C The telemetry strip shows sinus bradycardia progressing to a sinus pause. This finding is most consistent with a vagal mechanism, probably associated with obstructive sleep apnea. The best approach is to treat the sleep apnea. In the absence of symptoms such as syncope, a pacemaker is not indicated.

65

VENTRICULAR ARRHYTHMIAS

HASAN GARAN

DEFINITIONS

Ventricular arrhythmias are cardiac rhythms that originate in the ventricular myocardium or in the His-Purkinje tissue. They include a wide spectrum of arrhythmias, from the most innocuous isolated premature ventricular contraction (PVC) to the most malignant and life-threatening ventricular arrhythmia (Fig. 65-1).

Two consecutive PVCs are termed a *couplet*, whereas *ventricular tachycardia* (VT) is arbitrarily defined as three or more ventricular contractions in a row at a rate faster than 100 beats per minute. The definition of *sustained* VT—a continuous ventricular rhythm, at a rate faster than 100 beats per minute, with no interruption for 30 seconds or longer—is equally arbitrary. However, most if not all sustained VTs are much faster than 100 beats per minute, persist for more than 30 seconds, and cause a substantial decrease in ventricular function and cardiac output, especially in patients with underlying organic heart disease. These abrupt physiologic changes may result in acute heart failure, hypotension, syncope, or even circulatory collapse within several seconds to minutes after the onset of VT.

Monomorphic VT is electrocardiographically defined as a wide-complex tachycardia with no change in QRS configuration, frontal axis, or horizontal axis from one beat to the next (see Fig. 65-1C). Monomorphic ventricular tachycardia (Fig. 65-2A) at a very rapid (>250 beats per minute) rate is sometimes called ventricular flutter, but there is no consensus for a definite rate cutoff, and it is not possible to separate the QRS clearly from the T waves when the rate exceeds 250 beats per minute. Polymorphic VT is characterized by beat-to-beat changes in the QRS morphology and axis, and very fast polymorphic VT may be difficult to distinguish from *ventricular fibrillation* (VF) (Fig. 65-2B). VF is a grossly irregular ventricular rhythm, usually at a rate faster than 300 beats per minute and with markedly variable low amplitude in the QRS morphology, during which there is no cardiac output. Torsades de pointes and bidirectional polymorphic VT are two distinct subtypes of polymorphic VT. To avoid confusion, the term *pleomorphic* VT should be used rather than the term *polymorphic* VT to describe the phenomenon of multiple clinical monomorphic VTs, each with distinct QRS configurations and axis observed at different times in the same patient.

EPIDEMIOLOGY

The prevalence of PVCs is a function of sampling method and duration, and PVCs may be seen in 50% of apparently healthy individuals if the monitoring time is 24 hours or longer. Nonsustained VT may be recorded in up to 3% of apparently healthy individuals with no identifiable heart disease. The prevalence of PVCs and nonsustained VT increases with age, but also with the presence and severity of an underlying heart disease. Therefore, the finding of nonsustained VT often leads to a cardiac evaluation to exclude organic heart disease, even if it is incidentally discovered in an asymptomatic patient. The prevalence of nonsustained VT rises to 7 to 12% in the late phase of myocardial infarction (MI) and may be as high as 80% in patients with heart failure owing to dilated cardiomyopathy (Chapter 60).

Approximately 10% of patients with documented sustained VT have no identifiable heart disease, in which case idiopathic VT is diagnosed. Idiopathic VF is exceedingly rare. Sudden cardiac death (Chapter 63) owing to ventricular arrhythmias accounts for an estimated 50% of all annual cardiovascular deaths in the United States.[1]

The nature of the underlying heart disease in patients dying of VT or VF is age dependent. Before 30 years of age, the organic heart disease most commonly associated with VT and VF is genetic cardiomyopathy (Chapter 60), whereas acute MI and chronic ischemic cardiomyopathy are the most common underlying heart diseases in individuals older than 40 years. In about one third of cases of sudden cardiac death without obvious underlying organic heart disease at autopsy, post-mortem genetic analysis may identify a deleterious mutation in an ion channel—a so-called channelopathy that predisposes to VT and VF.

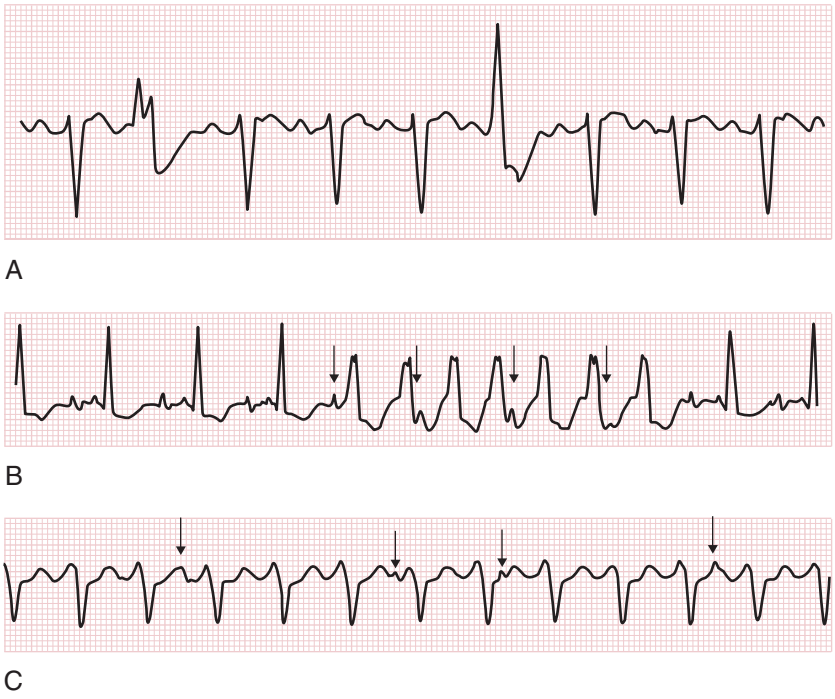

FIGURE 65-1. Ventricular arrhythmias. A, Multifocal premature ventricular beats. **B,** Nonsustained monomorphic ventricular tachycardia. Note dissociated P waves indicated by *arrows*. **C,** Sustained monomorphic ventricular tachycardia. Dissociated P waves are indicated by *arrows*.

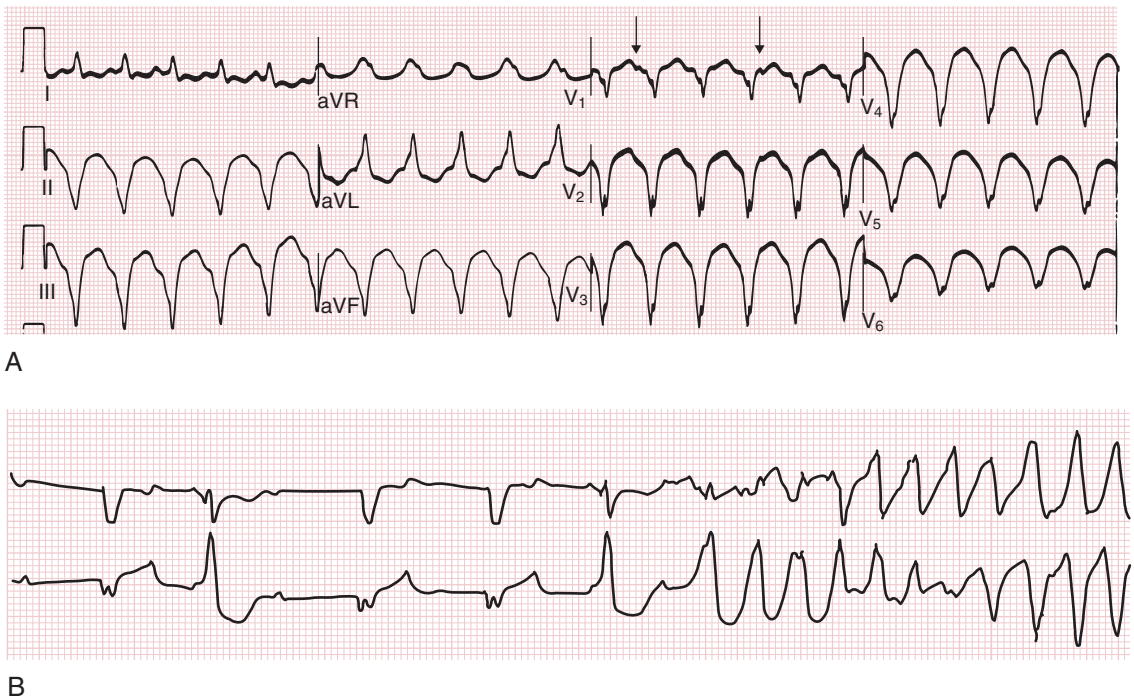

FIGURE 65-2. A, Monomorphic ventricular tachycardia (VT) in a patient with chronic myocardial infarction. The *arrows* identify P waves in lead V_1, showing atrioventricular dissociation. No R wave is recorded in any of the precordial leads V_1 to V_6 during VT. **B,** Polymorphic VT in a patient with chronic ischemic cardiomyopathy and marked first-degree atrioventricular block. There is no QT prolongation before the onset of the polymorphic VT.

PATHOBIOLOGY

Based on their underlying mechanisms, ventricular arrhythmias are classified as re-entrant, triggered, or automatic (Chapter 61). Re-entry, which results from activation in pathways sharing a common isthmus, is initiated by the simultaneous presence of conduction block in one limb and abnormally slow conduction in an adjacent limb, thereby allowing recovery of excitability in the former (E-Fig. 65-1A). One type of triggered activity results from early afterdepolarizations, which are oscillatory depolarizations occurring during the late phase of the action potential (E-Fig. 65-1B). Another type of triggered activity results from delayed afterdepolarizations, which are transient depolarizations that occur immediately after the termination of the action potential and may reach activation threshold. Automatic arrhythmias arise from accelerated pacemaker activity (E-Fig. 65-1C).

Sustained re-entrant activation in the myocardium, which is the most common cause of monomorphic VT, usually arises from subendocardial scarring, which is the result of prior ischemic injury and which creates an electrophysiologically abnormal substrate that results in re-entry. Other pathologic conditions capable of creating a substrate for re-entry include inflammation, granuloma (e.g., cardiac sarcoidosis), fibrofatty infiltration (e.g. arrhythmogenic right ventricular cardiomyopathy [ARVC]), genetically caused sarcomeric disarray (e.g., hypertrophic cardiomyopathy), and iatrogenic scar or

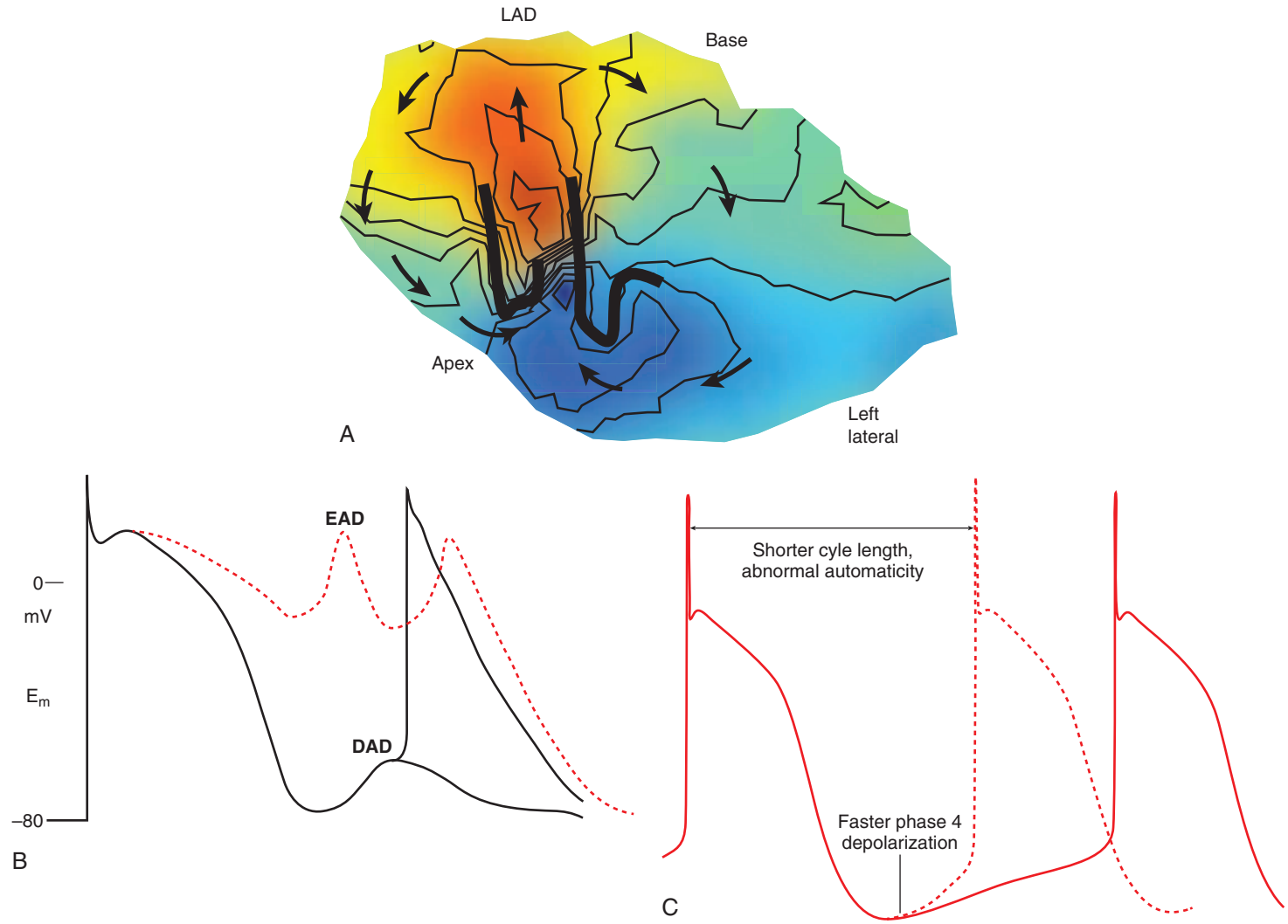

E-FIGURE 65-1. **A,** Re-entry within the myocardial infarction zone in an experimental canine model of ventricular tachycardia (VT). A central region of slow abnormal conduction, commonly referred to as the isthmus, is characterized by narrow crowded isochrones flanked by arcs of bidirectional conduction block depicted by *dark lines,* isolating the isthmus and enabling the maintenance of re-entry. The *arrows* indicate the spread of the wave of depolarization outside the central isthmus, in the shape of a figure of 8, with the red zone as the early breakthrough of activation and the dark blue area as the late activation in the VT cycle, which is also the point of re-entry into the isthmus. LAD = left anterior descending artery. **B,** Schematic depiction of the cardiac action potential with early afterdepolarizations (EAD) during phase 3 of a prolonged action potential *(dotted lines)* and delayed afterdepolarizations (DAD) reaching threshold and resulting in a premature action potential at the end of the phase 3 and the very start of the phase 4. E_m = membrane potential. **C,** Schematic depiction of cardiac action potential with an increased slope of depolarization toward the threshold during phase 4, at a site of automatic tachycardia.

patch (e.g., surgical repair of tetralogy of Fallot). These substrates may also result in polymorphic VT and VF by more than one mechanism.

The mechanism of ventricular arrhythmias in Brugada syndrome is not completely understood. One proposed mechanism is based on intraventricular phase 2 re-entry owing to an exaggerated endocardial-to-epicardial gradient in membrane potential due to differences in transient outward current. Other evidence suggests abnormal conduction in the epicardium of the right ventricular outflow tract.

Triggered activity, which results from adenosine-sensitive delayed afterdepolarizations rather than re-entry, is thought to be the underlying mechanism for idiopathic monomorphic VT of outflow tract origin. Idiopathic VT from re-entry in the fascicles of the left bundle branch has a relatively narrow QRS complex that always manifests right bundle branch block mimicry, most commonly with left, but rarely with right, frontal axis deviation.

Torsades de pointes is caused by early afterdepolarizations that arise during an abnormally prolonged action potential owing to a delayed repolarization process in the setting of genetic long QT syndromes or acquired long QT during therapy with QT-prolonging drugs. The cause may be either diminished outflowing potassium currents or enhanced inflowing sodium or calcium currents. Although many episodes terminate spontaneously, the rates are usually very fast, and a torsade episode, if long enough, can transform into VF.

Bundle branch re-entry, which results from re-entrant activation incorporating the right and the left bundle branches distally joined by the slowly conducting septal myocardium, may cause one or two nonsustained ventricular beats in a normal heart. However, sustained bundle branch re-entry occurs when myocardial disease causes chamber enlargement and bundle branch elongation and/or disease in the conduction system causes abnormal slow conduction, thereby creating the scenario for sustained bundle branch re-entry. The common type of bundle branch re-entry has anterograde activation over the right bundle and uses the left bundle retrogradely, thereby resulting in a left bunch branch block (LBBB) pattern on surface electrocardiogram (ECG), but the reverse direction with right bundle branch block (RBBB) may also occur rarely.

Accelerated pacemaker activity in an ectopic location, with rates exceeding the underlying sinus rhythm rate, may arise in settings such as transient inflammation, excess digoxin levels, intracellular calcium loading, electrolyte imbalance, and coronary reperfusion following thrombotic occlusion. Bidirectional VT is thought to result from calcium overload of the myocytes owing to congenitally acquired abnormal calcium release from the ryanodine receptor or digitalis toxicity.

Finally, there is no consensus regarding the mechanisms underlying VF. Theoretically, VF may be initiated when early or delayed afterdepolarizations fall in the vulnerable period of the action potential, thereby precipitating a re-entrant wave that breaks into sister wavelets and results in high-frequency electrical activity. In fact, VF may be regarded as an end stage for a variety of severe electrophysiologic abnormalities that result in chaotic activation.

CLINICAL MANIFESTATIONS

Ventricular arrhythmias can present in a variety of clinical settings (Table 65-1). Often, ventricular arrhythmias are asymptomatic and are detected by an irregular pulse on a physical examination, on a routine ECG, on an exercise test, or on routine inpatient monitoring. In other patients, symptomatic ventricular arrhythmias can present as palpitations, dizziness, syncope (Chapters 51 and 62), shortness of breath, or sudden cardiac arrest (Chapter 63). The diagnosis usually can be confirmed on an ECG, but ambulatory monitoring (Chapter 62) is often needed because the arrhythmia may be intermittent. Ambulatory monitoring can also help correlate arrhythmias with any potentially related symptoms. In some patients, exercise testing can be helpful, especially in patients with exercise-induced symptoms.

On the ECG, the QRS complex duration will typically be more than 0.12 seconds. In monomorphic VT (Fig. 65-3), the QRS complexes are the same from beat to beat, whereas polymorphic VT has multiple and changing QRS morphologies (Fig. 65-4). In VF, the ECG shows continuous irregular activation without any discrete QRS complexes (Fig. 65-5). Although underlying structural heart disease is usually present, these arrhythmias do not require a fixed structural substrate.

Acute Myocardial Infarction

VT and VF may arise as early as minutes to hours after the onset of symptoms during acute myocardial infarction (MI), and prehospital VT and VF during acute MI are responsible for a large proportion of out-of-hospital sudden cardiac deaths (Chapter 63). The incidence of peri-infarction VF has declined over the past two decades, presumably related to the widespread practice of coronary revascularization (Chapter 74) during acute MI. Among patients with ST elevation MI who now reach the hospital, about 3 to 4% develop VT, mostly during the acute phase. The incidence of VT in patients with non-ST elevation MI (Chapter 72) is lower, about 1%. Accelerated idioventricular rhythm (AIVR) is an automatic ventricular rhythm that is faster than the sinus rate but usually less than 120 beats per minute. It may occur in the setting of acute MI and is commonly observed immediately after coronary reperfusion. AIVR rates are slower than those of the fast and malignant VT and VF of acute MI, and this arrhythmia typically terminates spontaneously without causing hemodynamic instability.

TABLE 65-1 VENTRICULAR TACHYCARDIA AND CARDIAC DIAGNOSIS

STRUCTURAL HEART DISEASE
Acquired heart disease
Acute myocardial infarction
Chronic myocardial infarction, ischemic heart disease
Nonischemic dilated cardiomyopathy
Hypertensive heart disease
Valvar heart disease
Cardiac sarcoidosis
Cardiac amyloidosis
Other infiltrative diseases (e.g., Chagas disease)
Cardiac tumors
Congenital heart disease
Arrhythmogenic right ventricular cardiomyopathy
Hypertrophic cardiomyopathy
Genetic dilated cardiomyopathies
Iatrogenic
Surgically repaired congenital heart disease
Left ventricular assist devices
NO STRUCTURAL HEART DISEASE
Idiopathic ventricular tachycardia
Right and left ventricular outflow tract tachycardias
Left intrafascicular re-entry
Papillary muscle tachycardias
Idiopathic ventricular fibrillation
Ion channel mutations
Long QT syndromes
Catecholaminergic polymorphic ventricular tachycardia
Short QT syndrome
Mixed etiology
Brugada syndrome

DIAGNOSIS

Not every wide-complex tachycardia is VT. The diagnosis is straightforward from the His bundle electrogram recorded at the time of a wide-complex tachycardia during a cardiac electrophysiology study, but diagnosis on a standard 12-lead ECG may be challenging (Table 65-2). The differential diagnosis of a sustained regular-rate wide-complex tachycardia includes any type of supraventricular tachycardia with aberrant conduction (Chapter 64), supraventricular tachycardia with ventricular preexcitation, bundle branch re-entry (which is a specific type of VT), and myocardial VT. The clinical setting and the patient's background (e.g., history of previous MI or cardiomyopathy) play a major role in making an accurate diagnosis. New-onset wide-complex tachycardia in a young and otherwise healthy individual with no structural heart disease is most likely supraventricular tachycardia (SVT) with aberration, an SVT with preexcitation, or idiopathic VT.

The most reliable observation in favor of VT is evidence of AV dissociation, that is, absence of any relationship between the atrial and ventricular rate, with the ventricular rate faster than the atrial (see Fig. 65-2A), or a regular wide-complex tachycardia with the atria fibrillating. However, the absence of atrioventricular (AV) dissociation does not exclude VT because ventriculoatrial conduction is present in about 25% of VTs. Fusion beats (which occur when an occasional sinus beat conducts through the AV node and reaches the His-Purkinje system at the same time as the VT source activates the myocardium, thereby resulting in a beat with a morphology that is the hybrid of a conducted QRS complex and the VT complex) confirm AV

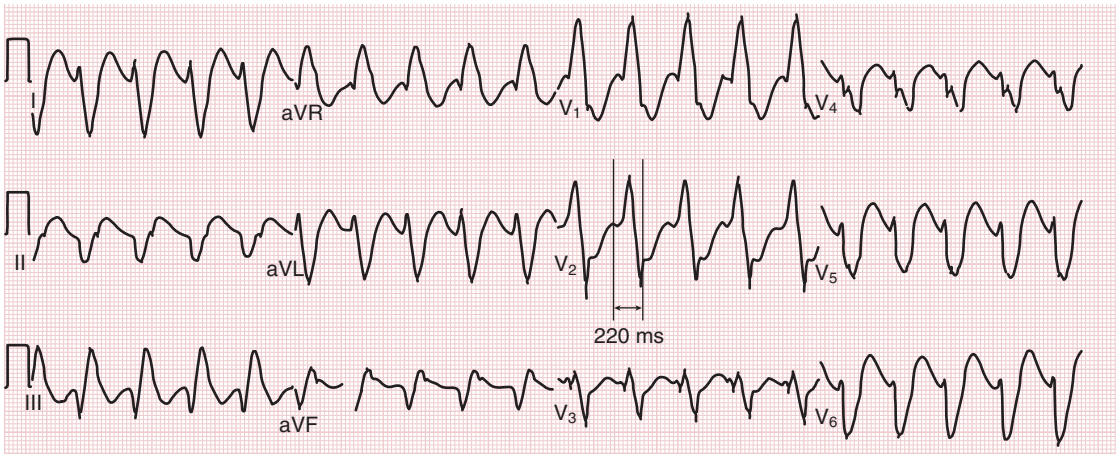

FIGURE 65-3. **Monomorphic ventricular tachycardia in a patient with chronic ischemic cardiomyopathy.** In lead V_2, the duration from the onset of the R wave to the nadir of the S wave is more than 200 msec. See text for further explanation.

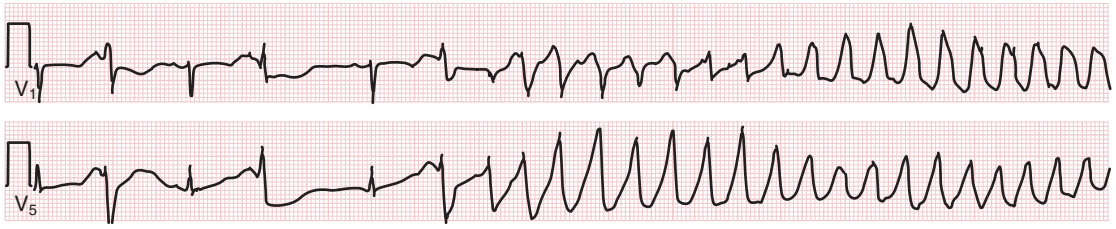

FIGURE 65-4. Torsades de pointes (TdP) in a patient with a markedly prolonged QT interval. A premature ventricular beat just after the peak of the T wave initiates TdP. As the tachycardia progresses, the rotation or the "twist" in the QRS axis is clearly observed in lead V_1, with the polarity of the signal changing gradually from negative to positive.

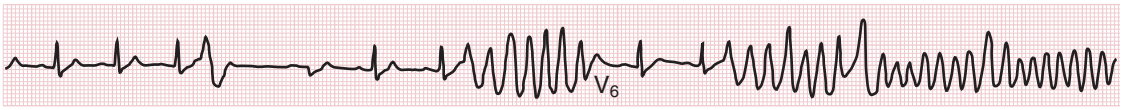

FIGURE 65-5. This electrocardiogram in a patient with idiopathic ventricular fibrillation (VF) shows recurrent closely coupled premature ventricular contractions (PVCs) and the initiation of VF by one of these closely coupled PVCs.

TABLE 65-2 DISTINGUISHING VENTRICULAR TACHYCARDIA FROM SUPRAVENTRICULAR TACHYCARDIA WITH ABERRANT CONDUCTION

VENTRICULAR TACHYCARDIA	SUPRAVENTRICULAR TACHYCARDIA
AV dissociation	Same QRS morphology as preexisting bundle branch block in sinus rhythm
aVR: initial R > S or initial R or Q > 40 msec	V_1: rsR′
Absence of any R wave in V_1 to V_6	
V_1 to V_6: onset of R to S > 100 msec in any lead	
QRS duration >160 msec	
Initial R wave in aVR	

AV = atrioventricular.

dissociation but are observed only when VT rates are relatively slow. Other findings that favor VT include a QRS duration longer than 160 msec, or longer than 140 msec with an RBBB pattern. One approach, based on the QRS configuration on the ECG, uses the absence of RS complex in all precordial leads or an interval of more than 100 msec from the onset of R to the nadir of S wave as observations strongly favoring VT (see Figs. 65-2A and 65-3). The absence of any R waves in the QRS complexes recorded from all six precordial leads, described as negative concordance, strongly suggests VT, but unfortunately is not a common finding. Prominent R waves observed in all six precordial ECG leads, termed *positive concordance,* may be seen in SVT with left ventricular preexcitation but otherwise also suggests VT with a basal site of origin. In the absence of preexcitation, a slow rate of rise in the voltage during the first 40 to 60 msec of the QRS onset suggests VT, as does the presence of initial R wave in lead aVR. A wide-complex tachycardia with a QRS morphology identical to that of aberrantly conducted beats manifesting bundle branch block (BBB) on a previously recorded ECG in the same patient should raise suspicion of bundle branch re-entry VT if AV dissociation is present.

If AV dissociation is not present, the differential diagnosis includes SVT with aberrant conduction, but the rare condition of preexcitation with an atriofascicular accessory pathway should also be considered in a patient with LBBB aberration. A monomorphic wide-complex tachycardia with an irregular rate, manifested by more than 60-msec difference in cycle length from one beat to the next, is likely to be atrial fibrillation (AF) or atrial flutter, with variable AV block and aberrant conduction or with preexcitation. It is important to emphasize that electrolyte imbalances or the use of antiarrhythmic drugs diminishes the predictive accuracy of all of these diagnostic clues.

Sustained polymorphic wide-complex tachycardia with marked beat-to-beat changes in the QRS morphology is always ventricular and either terminates spontaneously or transforms into VF. Torsades de pointes, a specific type of polymorphic VT, derives its name from the "twisting" or rotating of the QRS axis as the tachycardia progresses. It occurs in genetic or acquired long QT syndrome and is frequently pause dependent—typically starting when a premature beat falls on the prolonged T wave of the beat following a long RR interval (see Fig. 65-4). Finally, bidirectional VT manifesting a unique feature of beat-by-beat axis alternans may occur with digitalis toxicity or in the congenital catecholaminergic polymorphic ventricular tachycardia syndrome.

Several different algorithms based on the configurations of the QRS complexes have high sensitivity, high specificity, and acceptable predictive accuracy for distinguishing epicardial VT from endocardial VT (Table 65-3).[2] All

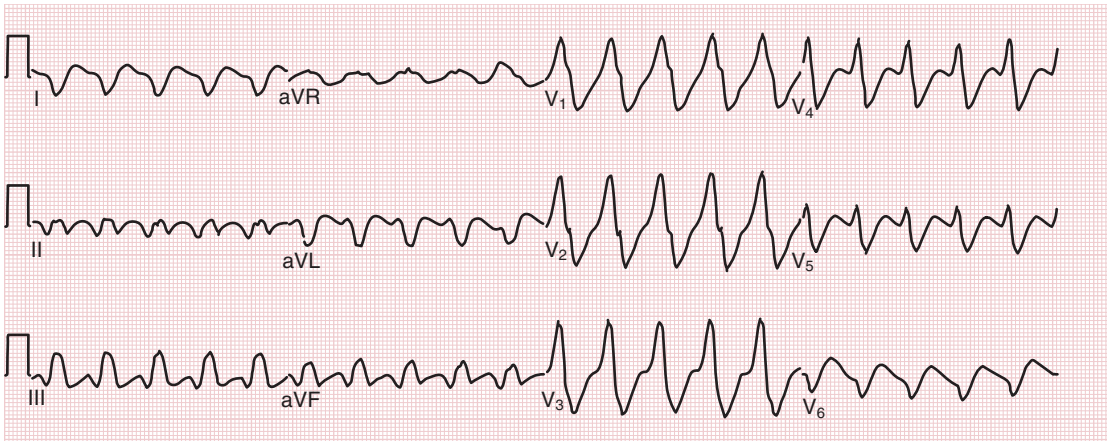

FIGURE 65-6. Monomorphic epicardial ventricular tachycardia in a patient with nonischemic dilated cardiomyopathy. The positive polarity pseudo-delta wave is prominent in the right precordial leads and the negative polarity pseudo-delta wave is prominent in the inferior limb leads.

TABLE 65-3 ELECTROCARDIOGRAPHIC PARAMETERS USED TO PREDICT AN EPICARDIAL ORIGIN OF VENTRICULAR TACHYCARDIA

PARAMETER	CRITERIA
Pseudo-delta wave	>75 msec favors epicardial site
Intrinsicoid deflection time	>85 msec favors epicardial site
Onset of R to nadir of S in precordial leads	>120 msec favors epicardial site
QRS duration	Epicardial longer
Q waves during VT in lead I	Favors epicardial site
Q waves during VT in II-III-aVF	Favors endocardial site
aVR/aVL amplitude ratio	Epicardial higher

VT = ventricular tachycardia.

are based on ECG criteria for whether the initial activation likely starts at an epicardial site. If so, the rapidly conducting His-Purkinje system is not available immediately, and the intramyocardial conduction delay produces a slurred initial component of the QRS complex, often called a *pseudo-delta wave*, which is manifested as a slow rate of rise of voltage before it reaches the intrinsicoid deflection (Fig. 65-6). Early recognition of ECG findings suggesting an epicardial origin of VT is important in planning and preparing a patient before a catheter ablation procedure (Chapter 66) because the epicardial approach requires a special technique in the cardiac electrophysiology laboratory.

Cardiac electrophysiology testing (Chapter 62) may be indicated in patients who have organic heart disease and recurrent syncope but in whom the history, physical examination, ECG, echocardiogram, and ambulatory cardiac rhythm monitoring fail to clarify the cause, especially if the patient has a history of myocardial infarction or cardiomyopathy, either of which increases the probability that VT may be the cause of syncope. A second diagnostic indication is to identify the mechanism underlying a documented wide-complex tachycardia before the consideration of catheter ablation therapy (Chapter 66).

Identifying the Underlying Cause of Ventricular Arrhythmias

In patients with a diagnosed ventricular arrhythmia, the next step is to conduct a careful evaluation to exclude any underlying structural heart disease. This evaluation must include a comprehensive history and physical examination (Chapter 51), echocardiography (Chapter 55), and stress testing (Chapter 71). The family history may provide clues to guide genetic testing for an inherited cardiomyopathy (Chapter 60). Cardiac magnetic resonance imaging (Chapter 56) is indicated in selected patients to exclude conditions such as sarcoidosis and ARVC.

Despite a comprehensive evaluation, about 10 to 15% of patients will have PVCs or VT with no identifiable structural or genetically identifiable cause. Most of the idiopathic monomorphic VTs are in one of two categories, defined by ECG morphology. VTs that arise in the right or left ventricular outflow tract typically manifest an inferiorly directed frontal axis and are markedly positive in inferior leads (E-Fig. 65-2); the QRS configuration observed in the right precordial leads may further discriminate the sites of origin as the right or the left ventricular outflow tract or one of the sinuses of Valsalva. By comparison, idiopathic left ventricular tachycardia usually manifests RBBB mimicry and left axis deviation, but there may also be right axis deviation. The QRS complexes typically are not very wide because the involved region is His-Purkinje tissue adjacent to the interventricular septum. The differential diagnosis includes idiopathic VT arising in one of the left ventricular papillary muscles (E-Fig. 65-3). When either of these typical patterns is observed in a patient with no structural heart disease, the physician should suspect idiopathic VT. Conversely, sustained VT that does not fall into either of these two broad categories should always raise a high index of suspicion that organic heart disease may be present.

Chronic Ischemic Heart Disease and Post–Myocardial Infarction Ventricular Tachycardia

In survivors of ST elevation MI, the prevalence of sustained VT by 6 weeks is about 1%, and VT may occur as late as 15 to 20 years after the acute MI without any intervening event. VT commonly, but not invariably, reflects poor left ventricular function, especially a dyskinetic left ventricular wall segment. The electrophysiologic substrate is the surviving but electrophysiologically abnormal tissue embedded in the infarcted zone, which creates the conditions for re-entry. The areas that harbor pathways underlying re-entry can be identified by low-amplitude fractionated local electrograms recorded from the endocardium. Up to 16% of the patients have VT of epicardial origin. The same substrate may cause polymorphic VT and VF, which do not depend on a long QT interval and are different than torsades de pointes seen with repolarization abnormalities.

Nonischemic Dilated Cardiomyopathy

The most common cause of sustained monomorphic VT in nonischemic cardiomyopathy (Chapter 60) is also re-entry within the myocardium, but it differs from the post-infarction VT of chronic ischemic heart disease. The pathologic substrate, such as fibrosis, may be hard to identify. The abnormal, low-voltage, fractionated local electrograms tend to be located in basal, lateral, and often perivalvar left ventricular areas, which may correlate with the location of intramyocardial or subepicardial scarring identified by cardiac magnetic resonance imaging. The proportion of monomorphic VTs due to bundle branch re-entry is higher in nonischemic dilated cardiomyopathy compared with chronic ischemic heart disease, and VT with a focal rather than re-entrant mechanism rarely may be observed. Also, VT of nonischemic dilated cardiomyopathy is more likely to have an epicardial origin—as high as 22 to 35% in many series—and reaching 70% in Chagas disease.[3] Ventricular tachycardia resulting from bundle branch re-entry also is more common in nonischemic dilated cardiomyopathy.

Heart Failure

The failing heart from any underlying cause (Chapter 58) is highly vulnerable to ventricular arrhythmias, and 40 to 60% of the deaths in patients with

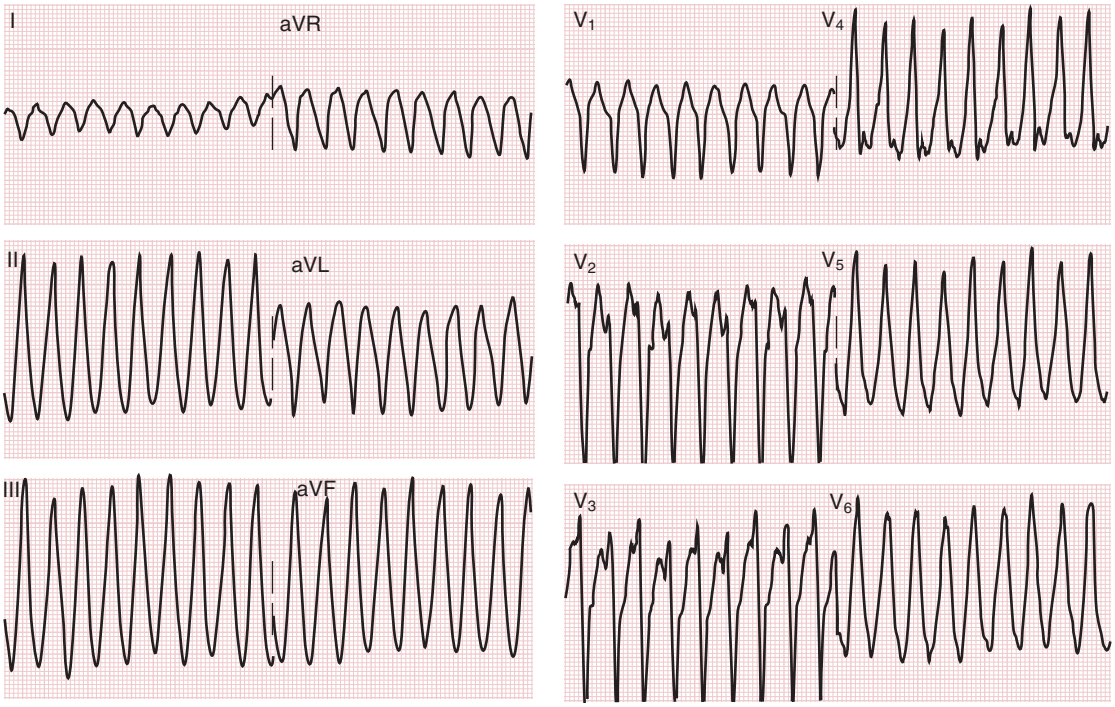

E-FIGURE 65-2. Electrocardiogram recorded during idiopathic ventricular tachycardia originating in the right ventricular outflow tract and manifesting a deeply inferior frontal axis and left bundle branch block mimicry in the precordial leads.

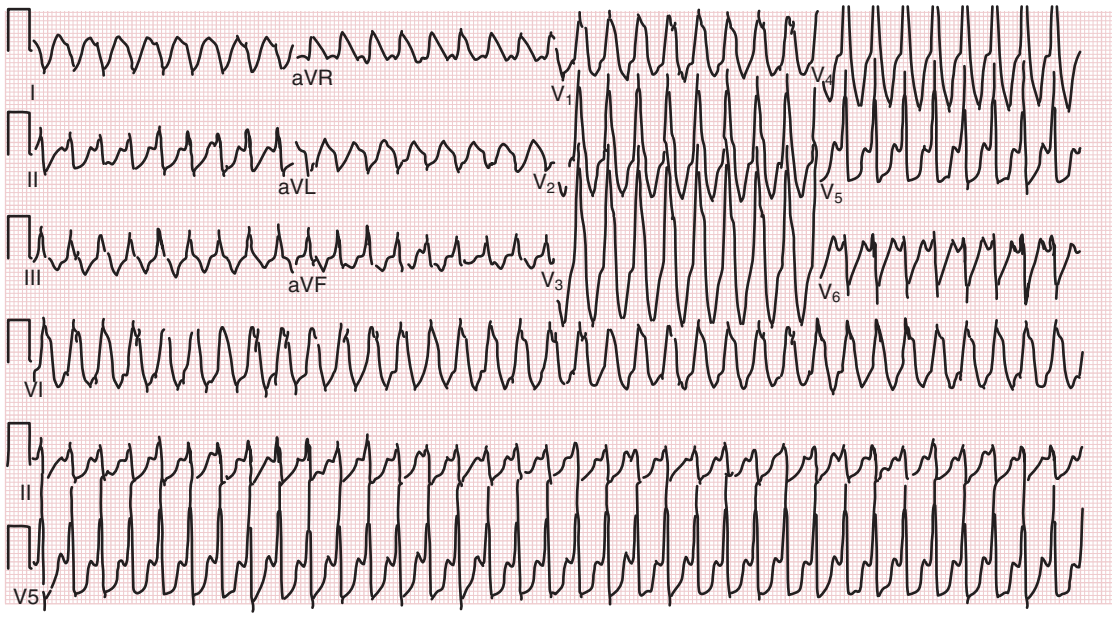

E-FIGURE 65-3. Ventricular tachycardia (VT) originating in the anterolateral papillary muscle. Note the right bundle branch mimicry of the QRS and the right axis deviation, similar to the electrocardiographic configuration of an interfascicular re-entrant VT.

heart failure are sudden and commonly from VT and VF. Re-entrant VT is common especially in patients whose heart failure is due to advanced ischemic heart disease, but triggered activity resulting from derangements of calcium homeostasis may also play a prominent role. In addition, hormonal factors, electrolyte abnormalities, and changes in autonomic nervous system activity also increase the vulnerability of the failing heart to ventricular arrhythmias.

Inflammatory and Infiltrative Disease

Among patients with sarcoidosis (Chapter 95), about 40 to 50% have cardiac involvement, which may first manifest as progressive AV block and VT. Although the true prevalence of VT in sarcoidosis is not known, in the selected patients who have received implantable cardiac defibrillators (ICDs) for cardiac sarcoidosis diagnosed by endomyocardial biopsy, cardiac magnetic resonance imaging, or cardiac positron emission tomographic scans, about 15% per year have appropriate ICD discharges for sustained VT.[4] Patients with other infiltrative heart diseases such as amyloidosis (Chapter 188) also have an elevated risk for VT and life-threatening ventricular arrhythmias.[5]

Adult Congenital Heart Disease

VT may occur in the setting of any adult congenital heart disease when there is a ventricular surgical scar or patch, as is seen after repair of tetralogy of Fallot or a ventricular septal defect closure, or a failing ventricle such as after a Mustard or Senning procedure to palliate transposition of great arteries (Chapter 69). In patients with surgically repaired tetralogy of Fallot, the prevalence of VT is about 5%, and about 2% have sudden cardiac death.

Genetically Inherited Cardiomyopathies

Hypertrophic cardiomyopathy (Chapter 60) is responsible for more than one third of sudden cardiac deaths in patients younger than age 25 years (Chapter 63), and mortality in young hypertrophic cardiomyopathy patients is almost exclusively due to VT and VF. Neither genetic testing nor a cardiac EP study can definitively identify patients at high risk for VT and VF, and the risk is determined based on findings such as a history of syncope, documented nonsustained VT especially in a young patient, a markedly thickened (>3 cm) interventricular septum, and a paradoxical decrease in blood pressure during exercise.[6]

ARVC is a congenital cardiomyopathy (Chapter 60), usually with an autosomal dominant inheritance. The fibrofatty infiltration of the right ventricular myocardium, which may also involve the interventricular septum and the left ventricle, results in progressive histologic change and marked electrophysiologic abnormalities, which may be manifest on the surface ECG as an epsilon wave (Fig. 65-7). The markedly altered conduction characteristics are conducive to re-entry. The incidence of VT in ARVC is related to the severity of the pathologic myocardial changes and ranges from 25 to 100%, depending on the penetrance and the expressivity of the disease. VT typically is initiated by exercise and demonstrates LBBB mimicry in the precordial ECG leads. However, unlike idiopathic right ventricular outflow tract VT, the frontal axis may be variable and not always inferiorly directed, and the site of origin may be epicardial in about 40% of cases.

Genetically Inherited "Channelopathies"

Several genetically acquired syndromes, including the long QT syndromes, Brugada syndrome, and catecholaminergic polymorphic VT increase the risk for sudden cardiac death due to ventricular tachyarrhythmias. Despite the remarkable heterogeneity of the long QT syndrome, most of the cases (LQT1, LQT2, LQT3) result from mutations in the genes coding for one of the potassium channels or the sodium channel.[7] The other genetic mutations are extremely rare. The VT of long QT syndrome is torsades de pointes, and both bradycardia and pauses increase its probability in patients who are predisposed. The incidence of torsades de pointes is influenced by multiple factors, including age, gender, the particular genetic mutation, and the magnitude of QT prolongation (Fig. 65-8A).

The electrocardiographic hallmark of Brugada syndrome, which also predisposes to ventricular tachyarrhythmias and sudden cardiac death, is the coved ST segment elevations in the right precordial leads (Fig. 65-8B). In some cases, this pattern may not be present except when the patient is febrile. The inheritance is autosomal dominant, a sodium channel mutation is present in 20 to 30% of the cases, but the genetics are heterogenous. Catecholaminergic polymorphic VT is a rare genetic condition resulting from abnormal calcium homeostasis. It is characterized by exercise-induced, wide-complex tachycardia manifesting alternating ECG axes from one beat to the next. This condition also predisposes the patient to exercise-induced VF. In addition, a chromosomal haplotype causing overexpression of dipeptidyl peptidase-like protein-6 has been described in one type of familial idiopathic VF, a rare but challenging subset of inheritable arrhythmia syndromes causing sudden cardiac death.

Iatrogenic Ventricular Tachycardia and Ventricular Fibrillation

QT-prolonging drugs, including class III antiarrhythmic drugs (see www.sads.org.uk), may precipitate torsades de pointes and VF in genetically predisposed individuals even if the baseline QTc is normal or borderline. Class IC antiarrhythmic drugs may cause life-threatening VT in patients with ischemic or any other organic heart disease and in patients with Brugada syndrome. Ventricular scarring owing to aneurysmectomy, tetralogy of Fallot repair, ventricular septal defect repair, alcohol ablation of the interventricular septum to relieve dynamic outflow tract obstruction in hypertrophic cardiomyopathy, or implantation of a left ventricular assist device may create a substrate for re-entry and VT.

TREATMENT

Rx

Premature Ventricular Contractions and Nonsustained Ventricular Tachycardia

In the absence of structural heart disease, there is no convincing evidence that ventricular ectopic activity influences survival. Therefore PVCs do not need treatment in asymptomatic patients. If ventricular ectopy results in symptoms that substantially decrease quality of life, a cardioselective β-blocker (e.g., metoprolol 50 mg twice daily or atenolol 50 mg once daily) is safe but

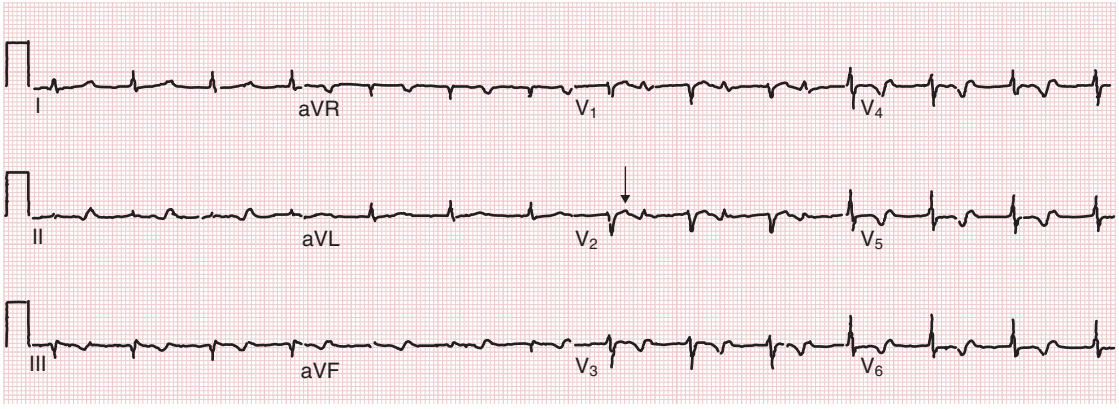

FIGURE 65-7. This electrocardiogram was recorded in a patient with arrhythmogenic right ventricular cardiomyopathy, marked first-degree heart block, and recurrent ventricular tachycardia. Epsilon waves, marked by the *arrow,* are visible in the right precordial leads.

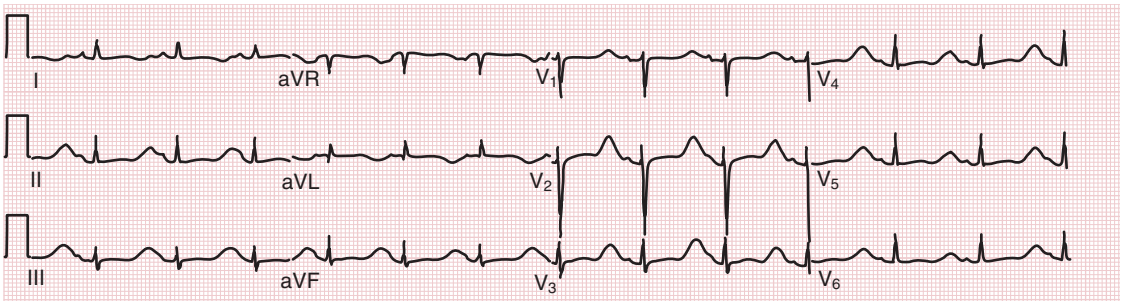

A

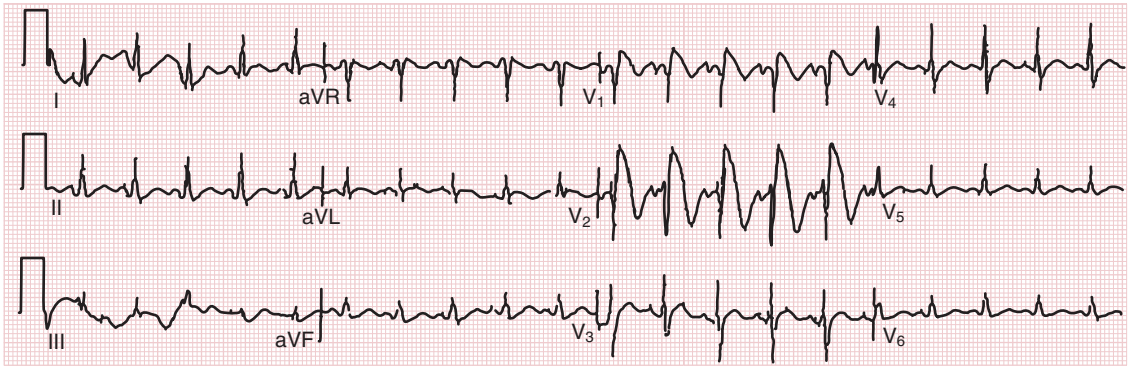

B

FIGURE 65-8. **A, Electrocardiogram showing a QT interval of 640 msec in a woman with LQT1 syndrome, with the terminal portion of the T wave merging with the P wave. B, Electrocardiogram of a man with Brugada syndrome, showing the typical "coved" ST elevation in lead V_1.**

not a very effective first-choice therapy to eradicate PVCs. PVCs may be more effectively suppressed using class IC antiarrhythmic agents such as flecainide (50 to 100 mg twice daily) or propafenone (150 to 225 three times daily), which are safe in the absence of organic heart disease but are contraindicated in organic heart disease, especially ischemic heart disease.[8]

In some patients, the frequency of PVCs and nonsustained VT reaches a critical level that results in decreased systolic ventricular function. These patients should be treated aggressively, including catheter mapping and ablation (Chapter 66) to avoid the potential adverse effects of antiarrhythmic drugs.

Acute Management of Ventricular Tachycardia and Ventricular Fibrillation

The management of VT with hemodynamic instability and VF should conform to the guidelines for advanced cardiac life support (Chapter 63), with an emphasis on defibrillation. For patients who have sustained VT with modest hypotension and normal mental status, intravenous drug therapy with lidocaine (given as a 50-mg bolus) or amiodarone (150 mg infused intravenously over 10 minutes) may be tried. Lidocaine, which works best at rapid heart rates, is an effective drug to terminate VT, which invariably occurs at a high rate, but not to prevent recurrences, except in the setting of acute ischemia. By comparison, amiodarone is more effective at slower heart rates and therefore is better for preventing recurrent VT after sinus rhythm is restored. Intravenous calcium channel blocker therapy should not be given unless the mechanism is known with certainty to be verapamil-sensitive idiopathic left ventricular tachycardia.

The most important factor in preventing early recurrence is the prompt identification and reversal of any precipitating causes. Examples include hypokalemia and other electrolyte imbalances, low oxygen saturation, intravenous β-agonist agents, and acute heart failure (Chapter 59) or myocardial ischemia (Chapter 72 and 73). Heart failure should be rigorously treated (Chapter 59). VF suggests the presence of residual ischemia; the feasibility of coronary revascularization (Chapter 74) should be addressed, but even then recurrences are common.

Treatment of Electrical Storm

Electrical storm is a term used to describe frequently recurrent VT or VF requiring repeated defibrillations. Electrical storm rarely occurs in nonischemic cardiomyopathy or in genetically acquired ventricular arrhythmias. When this condition is encountered in the early phase of acute MI, relief of ischemia is of paramount importance. If the electrical storm continues even after coronary reperfusion, insertion of an intra-aortic balloon pump and use of an intravenous β-blocker therapy, preferably with a short half-life drug (e.g., esmolol 50 to 300 μg/kg per minute by intravenous infusion) should be considered. Intravenous lidocaine (2 to 4 mg per minute) or intravenous amiodarone (0.5 to 1.0 mg per minute) may also be used if esmolol is ineffective.

Idiopathic Ventricular Tachycardia and Ventricular Fibrillation

Although cardiac arrest resulting from transformation of idiopathic VT to VF is exceedingly rare, sustained VT at a rate faster than 200 beats per minute commonly causes cardiopulmonary symptoms and even syncope (Chapter 62). Idiopathic VTs of outflow tract origin may respond to β-blocker therapy (e.g., metoprolol 50 mg every 12 hours, or atenolol 50 mg daily), and a few may respond to an empirical trial of calcium-channel blockers (e.g., sustained-release diltiazem 120 to 240 mg daily, or sustained-release verapamil 120 to 240 mg daily). The so-called idiopathic left ventricular tachycardia resulting from left fascicular re-entry frequently responds to verapamil (e.g., sustained-release 180 to 360 mg daily), but papillary muscle VT may not.

Catheter ablation therapy (Chapter 66), which may be curative for idiopathic VTs because of their focal origin in sites such as the right or left ventricular outflow tracts, epicardium, or papillary muscle, should be considered as a preferred alternative to long-term antiarrhythmic drug therapy.[9] The long-term success rate of catheter ablation for these focal sites can be about 85% or even higher. Studies with smaller groups of patients have reported even higher rates of success for idiopathic VT arising in the sinuses of Valsalva.

By comparison to the relatively benign prognosis of idiopathic sustained VT, idiopathic VF accounts for 5 to 10% of all cases of sudden cardiac death (Chapter 63). The appropriate treatment for survivors of idiopathic VF is no different than for any other survivor of VF (i.e., ICD therapy). The ECG may show recurrent PVCs with short coupling intervals. If these PVCs are monomorphic, they may be amenable to catheter ablation. Catheter ablation may decrease the risk for recurrence but still does not obviate the need for ICD protection.

Ventricular Tachycardia and Ventricular Fibrillation with Structural Heart Disease

Sustained VT and VF in patients with organic heart disease have become a common indication for ICD therapy. Three randomized trials comparing ICD therapy to antiarrhythmic drug therapy all showed significant survival benefit with ICD over drug therapy. After an acute MI, an ICD reduces mortality in patients who have survived more than 40 days and have a left ventricular ejection fraction of 30% or less or who have symptomatic heart failure and an ejection fraction of less than 0.35%; and patients more than 5 days after MI who have a reduced ejection fraction, nonsustained VT, and inducible sustained VT or VF on electrophysiologic testing. By comparison, ICDs do not

reduce mortality when routinely implanted soon after MI or in patients after recent coronary artery revascularization.

In patients with a chronically depressed ejection fraction of less than 30%, insertion of an ICD reduces the mortality rate by 20%, from 36% to 29%, over the next 5 years. By comparison, amiodarone suppresses ventricular ectopy and reduces sudden death but does not appear to improve survival.

However, there are no randomized placebo-controlled trials of antiarrhythmic drug therapy or catheter ablation for the secondary prevention of recurrent VT in patients with organic heart disease, probably because treatment of a potentially lethal arrhythmia with placebo has been considered unacceptable. As a result, antiarrhythmic drugs and catheter ablation currently serve as palliative treatments to modify the course of VT in patients who receive too many ICD shocks because of frequently recurrent or nearly incessant VT. If medications are chosen, amiodarone (100 to 400 mg daily) is more successful than β-blockers or sotalol for palliative therapy of VT in patients with ICD. However, sustained VT can be reproducibly initiated by focal electrophysiologic stimulation in 65% of the patients with chronic ischemic heart disease, and the identified site is usually amenable to treatment with catheter ablation. Among such patients who have an ICD for ischemic VT but experience recurrent VT, an average of about 50% of patients treated with catheter ablation will be free of VT for 2 years, although some studies report even higher success rates. These procedures should be undertaken only at institutions with the highest level of expertise and experience.

In patients who are clinically unstable because of incessant VT despite an ICD, reversible precipitating factors should be sought and, if present, corrected promptly. Intravenous β-blockers (e.g., esmolol 50 to 300 μg/kg per minute) and amiodarone (0.5 to 1.0 mg per minute) constitute the first line of therapy. Intravenous lidocaine (e.g., 2 to 4 mg per minute) can be added, especially if myocardial ischemia is present, and emergent catheter ablation may be considered.

Torsades de Pointes

The first line of therapy for torsades de pointes in patients with long QT syndrome is β-blocker therapy (e.g., metoprolol 50 to 100 mg daily, atenolol 50 mg daily, or nadolol 40 mg daily, and titrated as tolerated), but its success is influenced by gender and the magnitude of QT prolongation, as well as by the specific genotype.[10] In an acute setting with frequently recurrent torsades de pointes, magnesium sulfate (1 to 2 g intravenous infusion over 10 to 30 minutes) may be effective. An ICD is recommended if patients who are on β-blocker therapy develop recurrent syncope or documented torsades de pointes. Whether nonpharmacologic therapy, such as sympathetic ganglionic denervation, will prove effective in some cases is uncertain. β-Blockers are also the drugs of choice for catecholaminergic polymorphic VT, with ICD therapy recommended for patients with recurrent syncope or documented VT while on β-blocker therapy. Whether oral flecainide can obviate the need for ICD protection in such patients is uncertain.

Treatment of Genetically Acquired Ventricular Tachycardia and Ventricular Fibrillation

Most of the VTs observed in patients with ARVC may be reproducibly induced by programmed electrophysiologic stimulation and are amenable to catheter ablation. Recent clinical studies suggest that nearly half of these VTs have an epicardial site of origin, and simultaneous endocardial and epicardial catheter mapping and ablation may be the most effective method of treatment.[11] However, catheter ablation cannot substitute for ICD therapy. In patients with hypertrophic cardiomyopathy, ICD therapy is routinely recommended in high-risk patients (Chapter 60). Case reports suggest a benefit of catheter ablation in highly selected patients, and amiodarone may sometimes be useful.

For Brugada syndrome, catheter ablation of the electrophysiologically abnormal substrate in the right ventricular outflow tract epicardium can lead to eventual disappearance of the pathognomonic ST elevation on the ECG.[12] Quinidine (600 to 900 mg daily in three or four divided doses) is the only antiarrhythmic drug that appears to be useful to treat this syndrome, but whether catheter ablation or quinidine will be a substitute for an ICD in the highest risk patients, who have unprovoked Brugada pattern in their resting ECG and have a history of syncope, is unproved.

Iatrogenic Ventricular Tachycardia

For sustained monomorphic VT in patients with surgically repaired congenital heart disease, catheter mapping and ablation are recommended. The techniques are similar to those used for catheter ablation of ischemic VT.

Although sustained VT or VF during the course of LVAD therapy is usually well tolerated acutely, recurrent intractable VT or VF can result in right heart failure and in frequent ICD shocks, both of which may carry significant morbidity. Amiodarone (200 to 400 mg daily) and β-blocker therapy (e.g., metoprolol 100 to 200 mg daily) may be effective in at least rendering VT no longer incessant, and catheter ablation therapy has occasionally been tried in refractory VT with modest success.

PROGNOSIS

VT or VF in the early minutes or hours of acute MI has not been shown to affect the long-term prognosis in patients who survive to hospital discharge. Sustained monomorphic VT occurring after the hyperacute phase but within the next few days of an anterior wall MI portends poor prognosis, with about a seven-fold increase in subsequent mortality, because it usually occurs after the necrosis of a large amount of myocardium.

The modern natural course of untreated VT and VF perhaps can best be assessed from prospective studies of patients receiving ICD therapy. Among patients who have had documented sustained VT, appropriate ICD therapy for VT or VF occurs in about 70% of patients within 2 years and 85% of patients within 3 years. For patients who have survived cardiac arrest, the rate is about 70% at 3 years. These figures underscore the high rate of recurrence of VT and VF in patients with organic heart disease presenting with sustained ventricular arrhythmias.

The prognosis of patients with PVCs or nonsustained VT is less well known but is critically dependent on the underlying heart disease. No study to date has shown any survival benefit of treating PVCs or nonsustained VT in patients who do not have underlying organic heart disease and who do not develop a cardiomyopathy as a result of their ventricular ectopic activity. In very high-risk patients with ischemic heart disease, an ejection fraction of less than 40%, and electrically induced sustained VT, the 2-year rate of cardiac arrest or death from ventricular arrhythmia is more than 30%.

Grade A References

A1. Goldenberg I, Gillespie J, Moss AJ, et al. Long-term benefit of primary prevention with an implantable cardioverter-defibrillator: an extended 8-year follow-up study of the Multicenter Automatic Defibrillator Implantation Trial II. *Circulation*. 2010;122:1265-1271.

A2. Steinbeck G, Andresen D, Seidl K, et al. Defibrillator implantation early after myocardial infarction. *N Engl J Med*. 2009;361:1427-1436.

A3. Piccini JP, Berger JS, O'Connor CM. Amiodarone for the prevention of sudden cardiac death: a meta-analysis of randomized controlled trials. *Eur Heart J*. 2009;30:1245-1253.

A4. Bardy GH, Lee KL, Mark DB, et al. Amiodarone or an implantable cardioverter-defibrillator for congestive heart failure. *N Engl J Med*. 2005;352:225-237.

A5. Connolly SJ, Dorian P, Roberts RS, et al. Comparison of beta-blockers, amiodarone plus beta-blockers, or sotalol for prevention of shocks from implantable cardioverter defibrillators: the OPTIC Study: a randomized trial. *JAMA*. 2006;295:165-171.

A6. Kuck KH, Schaumann A, Eckardt L, et al. Catheter ablation of stable ventricular tachycardia before defibrillator implantation in patients with coronary heart disease (VTACH): a multicentre randomised controlled trial. *Lancet*. 2010;375:31-40.

A7. Reddy VY, Reynolds MR, Neuzil P, et al. Prophylactic catheter ablation for the prevention of defibrillator therapy. *N Engl J Med*. 2007;357:2657-2665.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Goldberger JJ, Basu A, Boineau R, et al. Risk stratification for sudden cardiac death: a plan for the future. *Circulation*. 2014;129:516-526.
2. Vallès E, Bazan V, Marchlinski FE. ECG criteria to identify epicardial ventricular tachycardia in nonischemic cardiomyopathy. *Circ Arrhythm Electrophysiol*. 2010;3:63-71.
3. Della Bella P, Brugada J, Zeppenfeld K, et al. Epicardial ablation for ventricular tachycardia: a European multicenter study. *Circ Arrhythm Electrophysiol*. 2011;4:653-659.
4. Betensky BP, Tschabrunn CM, Zado ES, et al. Long-term follow-up of patients with cardiac sarcoidosis and implantable cardioverter-defibrillators. *Heart Rhythm*. 2012;9:884-891.
5. Lin G, Dispenzieri A, Kyle R, et al. Implantable cardioverter defibrillators in patients with cardiac amyloidosis. *J Cardiovasc Electrophysiol*. 2013;24:793-798.
6. O'Mahony C, Jichi F, Pavlou M, et al. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM Risk-SCD). *Eur Heart J*. 2014;35:2010-2020.
7. Abrams DJ, Macrae CA. Long QT syndrome. *Circulation*. 2014;129:1524-1529.
8. Pedersen CT, Kay GN, Kalman J, et al. EHRA/HRS/APHRS expert consensus on ventricular arrhythmias. *Europace*. 2014;16:1257-1283.
9. Stevenson WG. Current treatment of ventricular arrhythmias: state of the art. *Heart Rhythm*. 2013;10:1919-1926.
10. Napolitano C, Bloise R, Monteforte N, et al. Sudden cardiac death and genetic ion channelopathies: long QT, Brugada, short QT, catecholaminergic polymorphic ventricular tachycardia, and idiopathic ventricular fibrillation. *Circulation*. 2012;125:2027-2034.
11. Bai R, Di Biase L, Shivkumar K, et al. Ablation of ventricular arrhythmias in arrhythmogenic right ventricular dysplasia/cardiomyopathy: arrhythmia-free survival after endo-epicardial substrate based mapping and ablation. *Circ Arrhythm Electrophysiol*. 2011;4:478-485.
12. Brugada P, Brugada J, Roy D. Brugada syndrome 1992-2012: 20 years of scientific excitement, and more. *Eur Heart J*. 2013;34:3610-3615.

66

ELECTROPHYSIOLOGIC INTERVENTIONAL PROCEDURES AND SURGERY

DAVID J. WILBER

PACEMAKERS

Temporary Pacemaking

In emergencies such as asystolic cardiac arrest (Chapter 63), transcutaneous pacing with electrode pads applied to the chest wall occasionally can be life-saving. Usually, however, time allows for temporary pacemaker leads to be inserted percutaneously, through an internal jugular or subclavian vein, and to be positioned and gently embedded in the right ventricular apex under fluoroscopic guidance. The lead is then attached to an external generator. A temporary pacemaker is often required as urgent therapy in a patient who has an indication for a permanent pacemaker and is awaiting that definitive procedure. Another indication for temporary pacing is the treatment of a

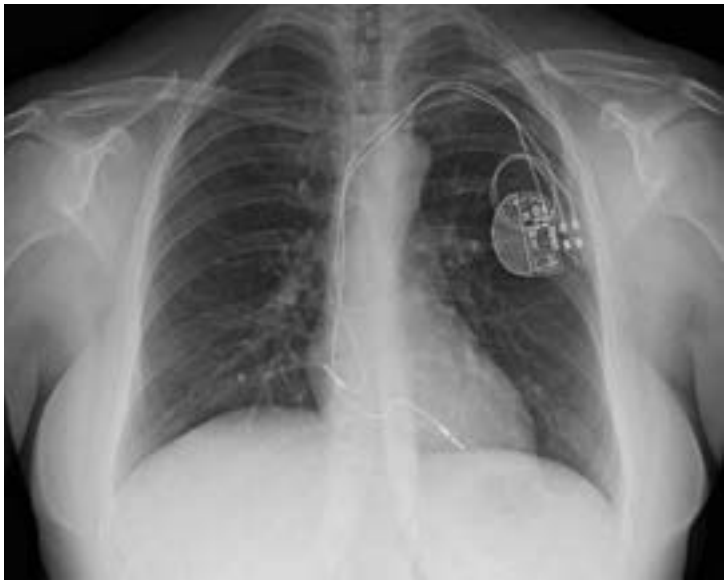

FIGURE 66-1. Typical lead placement for a dual-chamber permanent pacemaker.

transient symptomatic bradycardia, such as may be caused by drug toxicity or a metabolic perturbation, or to maintain a rate of 85 to 100 beats per minute in order to suppress torsades de pointes (Chapter 65) until the causative factor, especially an offending drug, has been eliminated. Prophylactic temporary pacing is used in patients who have high-degree atrioventricular (AV) block in the setting of an acute myocardial infarction (Chapter 73) and for patients who are at a high risk for developing symptomatic bradycardia during an interventional or surgical cardiac procedure.

The most common complication of temporary pacing is infection owing to inadequate sterile techniques at the time of implantation or to suboptimal antisepsis afterward. The risk can be minimized by limiting temporary pacing to 48 hours or by replacing the lead at that time under optimal sterile conditions.

Permanent Pacing

Permanent pacemaker leads can be inserted during cardiac surgery, but they much more frequently are inserted percutaneously through the subclavian vein or by cutdown through a cephalic vein. Ventricular leads are typically positioned in the right ventricular apex or, alternatively, higher on the right ventricular septum or outflow tract, and then are secured in place with a screw mechanism. Atrial leads are usually placed in the right atrial appendage (Fig. 66-1).

Lithium iodide pacemaker batteries, which have a 7- to 8-year lifespan and weigh less than 30 g, typically are implanted subcutaneously in the infraclavicular region (E-Fig. 66-1). The programmability of many different parameters has become standard, as has the ability of the pacemaker to provide diagnostic and telemetric data.

Indications for Permanent Pacemaking

Pacemakers are implanted either to alleviate symptoms caused by bradycardia or to prevent severe symptoms in patients in whom symptomatic bradycardia is likely to develop (Tables 66-1 and 66-2).[1-3] The most common bradycardia-induced symptoms are dizziness or lightheadedness, syncope or near-syncope (Chapters 51 and 62), exercise intolerance, and heart failure. Because these symptoms are nonspecific, documentation of an association between symptoms and bradycardia should be obtained before a pacemaker is recommended. If the bradycardia is persistent, such as in a patient with a complete AV block, a simple electrocardiogram may be sufficient to document the need for a pacemaker. If the bradycardia is intermittent, other diagnostic testing, such as 24-hour ambulatory monitoring, a continuous loop recorder, an implantable event monitor, or an electrophysiology test (Chapter 62), may be needed to document a relationship between symptoms and bradycardia.

Even after a symptomatic bradycardia has been documented, however, a correctable cause for the bradycardia (Chapter 64) should be excluded before a pacemaker is implanted. Correctable causes of symptomatic bradycardia include hypothyroidism (Chapter 226), an overdose with drugs such as

TABLE 66-1 CLASS I INDICATIONS* FOR IMPLANTATION OF A PERMANENT PACEMAKER

SINUS NODE DYSFUNCTION
Symptomatic sinus bradycardia
Symptomatic chronotropic incompetence
Symptomatic sinus bradycardia resulting from required drug therapy
ATRIOVENTRICULAR (AV) BLOCK
Third-degree and advanced second-degree AV block associated with symptomatic bradycardia
Third-degree and advanced second-degree AV block in an awake patient with asystole of >3 seconds or an escape rate of <40 beats per minute or with an infranodal escape rhythm
Atrial fibrillation with a pause of ≥5 seconds
Third-degree and advanced second-degree AV block due to postoperative AV block that is not expected to resolve
Third-degree and advanced second-degree AV block with neuromuscular diseases such as myotonic muscular dystrophy, Kearns-Sayre syndrome, and Erb dystrophy
Asymptomatic third-degree AV block if cardiomegaly or left ventricular dysfunction is present, or if the block is below the AV node
CHRONIC BIFASCICULAR BLOCK
Advanced second-degree or intermittent third-degree AV block
Type II second-degree AV block
Alternating bundle branch block
AFTER ACUTE PHASE OF MYOCARDIAL INFARCTION
Second-degree infranodal AV block with alternating bundle branch block
Third-degree infranodal AV block
Transient advanced second-degree or third-degree infranodal AV block and associated bundle branch block
Persistent symptomatic second-degree or third-degree AV block
CAROTID SINUS SYNDROME
Recurrent syncope caused by spontaneous carotid sinus stimulation or carotid sinus pressure that induces asystole of >3 seconds in duration

*Class I indications are conditions for which a pacemaker is indicated.
Adapted from Gillis AM, Russo AM, Ellenbogen KA, et al. HRS/ACCF expert consensus statement on pacemaker device and mode selection. *J Am Coll Cardiol.* 2012;60:682-703.

TABLE 66-2 CLASS IIA INDICATIONS* FOR IMPLANTATION OF A PERMANENT PACEMAKER

SINUS NODE DYSFUNCTION
Heart rate of <40 beats per minute when a clear association between symptoms consistent with bradycardia and the actual presence of bradycardia has not been demonstrated
Syncope of unclear etiology when sinus node dysfunction is demonstrated by electrophysiologic testing
ATRIOVENTRICULAR (AV) BLOCK
Persistent third-degree AV block with an escape rate of >40 beats per minute in an asymptomatic adult without cardiomegaly
Asymptomatic second-degree infranodal AV block
First-degree or second-degree AV block associated with symptoms similar to pacemaker syndrome
Asymptomatic type II second-degree AV block with a narrow QRS
CHRONIC BIFASCICULAR BLOCK
Syncope, when other potential causes of syncope have been excluded
An HV interval of ≥100 msec
Pathologic pacing-induced infranodal AV block during electrophysiologic testing
CAROTID SINUS SYNDROME
Syncope without clear provocative events and with asystole of >3 seconds during carotid sinus pressure

HIV = His-ventricle.
*Class IIA indications are conditions for which a pacemaker is reasonable.
Adapted from Tracy CM, Epstein AE, Darbar D, et al. 2012 ACCF/AHA/HRS focused update of the 2008 guidelines for device-based therapy of cardiac rhythm abnormalities: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2012;60:1297-1313.

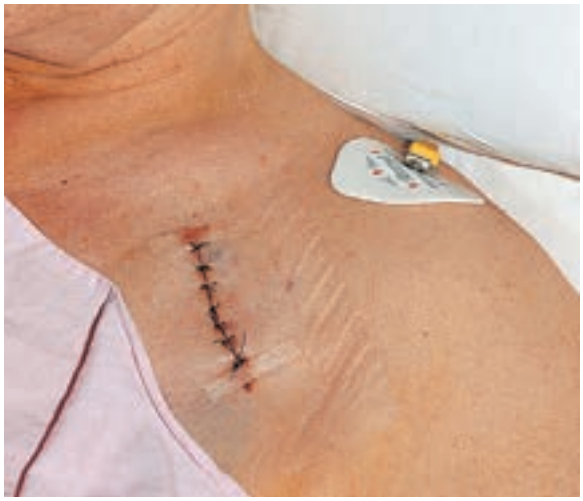

E-FIGURE 66-1. Site of implantation of a permanent pacemaker battery and generator. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine.* 3rd ed. London: Mosby; 2003.)

digitalis, electrolyte disturbances, and medications such as β-adrenergic blocking agents (administered either orally or in the form of eye-drops for glaucoma), calcium-channel blocking agents, and antiarrhythmic medications (Chapter 64). At times, a pacemaker is necessary to allow continued treatment with a medication that is responsible for the bradycardia, such as in a patient in whom symptomatic sinus bradycardia develops after initiation of therapy with a β-adrenergic blocking agent for paroxysmal atrial fibrillation (AF) associated with a rapid ventricular response.

Pacing Modes

Pacing modes are described by a simple code. The first letter represents the chamber being paced (A for atrium, V for ventricle, D for dual chamber). The second letter identifies the chamber whose depolarizations are being sensed by the pacemaker (A, V, D, or O for no sensing). The third letter indicates whether the pacemaker is functioning in an inhibited (I) mode, a tracking (T) mode, in both modes (D), or asynchronously (O). The fourth letter designates whether the pacemaker can modulate the heart rate on its own, independent of the patient's intrinsic atrial activity. An additional fifth letter may be used to define the pacemaker's ability to provide antitachycardia pacing (P), to deliver shocks (S), or both (D).

The most appropriate pacing mode must be determined for each individual. By far the most common permanent pacing modes now used in the United States are DDD (pacing and sensing of the atrium and ventricle in both inhibited and tracking fashion) and DDDR (with the additional ability to adjust the atrial rate independently in patients with a poor intrinsic heart rate response to exercise).

In patients with sinus node dysfunction, atrial or dual-chamber pacing significantly reduces the risk for AF and improves quality of life compared with ventricular pacing. Although atrial pacing alone is an option for younger active patients with normal AV conduction, the high risk that these patients will develop symptomatic AV block makes initial dual-chamber pacing attractive. In patients with AV block, dual-chamber pacing improves quality of life, reduces the risk for developing AF, and avoids the 25% risk for developing the pacemaker syndrome, which consists of symptoms of weakness, lightheadedness, exercise intolerance, or palpitations owing to the absence of AV synchrony during ventricular pacing. This syndrome is treated by restoring AV synchrony with atrial-based pacing modes, which would require an additional procedure to implant an atrial lead if a dual-chamber pacemaker were not originally placed. For these reasons, current consensus guidelines recommend dual-chamber pacing for most patients with sinus node dysfunction or AV block.

In patients who have paroxysmal AF and dual-chamber pacemakers, the ventricular rate will attempt to track the rapid atrial rates during the arrhythmia. Mode-switching pacemakers can pace in the DDD mode during sinus rhythm and automatically switch to rate-responsive ventricular pacing during AF or other supraventricular arrhythmias (Fig. 66-2). In patients who have long-standing persistent AF and in whom further attempts to restore sinus rhythm are not planned, there is no indication for atrial pacing or for the placement of an atrial lead.

An exception to the recommendation for dual-chamber pacing is in patients who have chronic AF with occasional symptomatic pauses. VVIR pacing is recommended to protect against the pauses and to provide a normal rate response to exercise if needed.

Another option for patients with AV block is biventricular pacing. In patients with AV block, class I to III heart failure, and left ventricular systolic dysfunction, biventricular pacing is superior to conventional right ventricular pacing with an insignificant 17% reduction in death but a larger reduction in severe heart failure.

Complications of Pacemakers

About 1 to 2% of patients develop complications from the implantation procedure itself, including pocket hematoma, pneumothorax, perforation of the atrium or ventricle, lead dislodgement, subclavian vein thrombosis, and infection. A strategy of continued warfarin treatment at the time of implantation of a pacemaker or an implantable cardioverter-defibrillator (ICD) markedly reduces the incidence of clinically significant device-pocket hematoma compared with bridging therapy with heparin. The subcutaneous pocket may develop a hematoma or local infection.[4] Pacemaker infections typically involve primarily the subcutaneous pacemaker pocket, but long-term resolution of infection generally requires removal of both the pulse generator and leads as well as long-term antibiotic therapy (Chapter 76). If the device becomes infected, nearly 40% of patients have coexisting valve involvement, predominantly tricuspid valve infection (Chapter 76), with mortality rates as high as 15% in the hospital and 20 to 25% at 1 year.[4]

During long-term follow-up after pacemaker implantation, potential problems include failure to pace, failure to capture, and changes in the pacing rate. These problems may be a manifestation of suboptimal programming, fracture of a lead or a break in its insulation, generator malfunction, or battery depletion.

Ventricular pacing, particularly from the right ventricular apex, is associated with a delayed and abnormal activation sequence, and interventricular and intraventricular mechanical dyssynchrony. When more than 40% of heart beats are the result of ventricular pacing, even in dual-chamber pacing modes, patients can develop adverse ventricular remodeling with ventricular dilation, systolic dysfunction, altered myocardial metabolism, and functional mitral regurgitation. Clinically, such patients are at greater risk for developing AF and heart failure. To avoid those complications, every attempt should be made to minimize the amount of ventricular pacing, including programming longer AV delays (220 to 250 msec) or implanting pacemakers with algorithms that minimize the cumulative percentage of ventricular pacing. In patients with bradycardia and a left ventricular ejection fraction of 35% or less, cardiac resynchronization therapy should be considered if significant right ventricular pacing (>40% of heart beats) is anticipated.

TRANSTHORACIC CARDIOVERSION AND DEFIBRILLATION

Techniques

Defibrillators generate and then discharge an electrical current across two paddle electrodes. The resulting shock simultaneously depolarizes large portions of the atria or ventricles, thereby terminating re-entrant circuits and extinguishing re-entrant arrhythmias that rely on such circuitry (Chapters 61, 64, and 65). Synchronization with the QRS complex (*cardioversion*) is always advised in patients with either a supraventricular tachycardia (SVT) or ventricular tachycardia (VT) because a nonsynchronized shock coincident with the T wave may precipitate ventricular fibrillation (VF). If a shock is needed

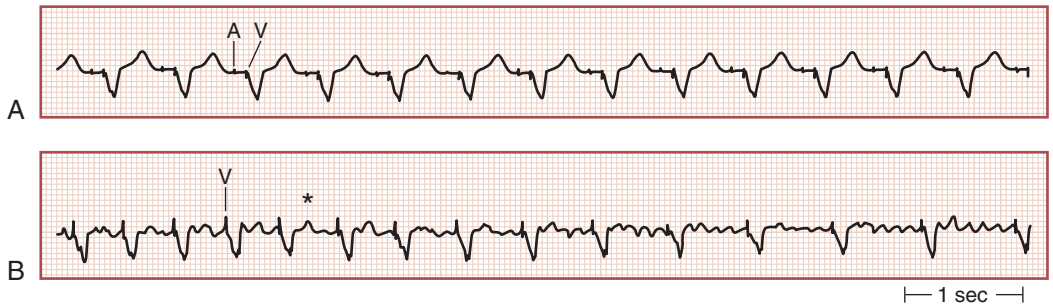

FIGURE 66-2. **Rhythm strips from a Holter monitor in a patient with complete atrioventricular block, sinus bradycardia, paroxysmal atrial fibrillation, and a rate-responsive dual-chamber pacemaker with mode-switching capability. A,** When the patient is in sinus rhythm, the pacemaker functions in a DDDR mode, with synchronized atrial and ventricular pacing at 105 beats per minute while the patient is walking. **B,** At the onset of an episode of atrial fibrillation, there is tracking of the atrium that results in ventricular pacing at 140 beats per minute, which is the upper rate limit of the pacemaker. Within 2 seconds *(asterisk),* the mode-switch feature results in VVIR pacing, and the ventricular pacing rate gradually falls to 70 beats per minute, which is the lowest rate limit of the pacemaker. A = atrial stimulus; V = ventricular stimulus. (Courtesy of Dr. Fred Morady.)

to terminate VF, however, this *defibrillation* does not require synchronization to the QRS complex.

The success of cardioversion or defibrillation is affected by the shock waveform and shock strength. Biphasic shock waveforms are recommended because they are significantly more effective than monophasic waveforms at equivalent energies. Other technique-dependent variables that maximize the energy delivered to the heart include increasing paddle pressure, delivery of the shock during expiration, and repetitive shocks. Patient-related factors that may decrease the probability of successful cardioversion and defibrillation include metabolic disturbances, a longer duration of arrhythmia, and higher body weight.

Indications and Technique

The most common arrhythmias treated by cardioversion and defibrillation are VF, VT, AF, and atrial flutter (Chapters 63, 64, and 65). Treatment of VF is always an emergency: a 200-J defibrillation shock should be delivered emergently, followed by one or more 360-J shocks if necessary. Cardioversion of VT may be a life-saving emergency procedure, similar to defibrillation for VF, or an urgent but controlled procedure with an initial shock strength of 50 to 100 J followed by higher energy shocks if needed. For AF, cardioversion is usually an elective procedure, with an initial shock of 200 J in adults, followed by shocks of 300 to 360 J if necessary. For cardioversion of atrial flutter, an initial shock of 50 to 100 J is appropriate. Regardless of the underlying arrhythmia, the energy required is a probability function and not a discrete value, so subsequent shocks may be effective for successful cardioversion or defibrillation even if the first 360-J shock is not effective.

Elective cardioversion requires fasting for at least 8 hours, a reliable catheter in a peripheral vein, oxygen, suction, and equipment for potential emergency airway management. Patients are premedicated (Chapter 432), usually with propofol. In the anteroposterior configuration, which may be more effective for initial cardioversion of AF, one electrode is positioned to the left of the sternum at the fourth intercostal space, with the second electrode placed posteriorly, to the left of the spine, at the same level as the anterior electrode. In the anteroapical configuration, one electrode is placed to the right of the sternum at the level of the second intercostal space, and the second electrode is placed at the mid-axillary line, lateral to the apical impulse.

Precautions and Complications

Cardioversion of AF (Chapter 64) may be complicated by thromboembolism. If no atrial thrombi are seen on a transesophageal echocardiogram, preprocedure anticoagulation is not necessary. Otherwise, anticoagulation is necessary for 3 weeks before elective cardioversion. All patients should be anticoagulated for 1 month after cardioversion if AF has been present for 48 hours or longer.[5]

VF may rarely occur even when shocks are synchronized to the QRS complex. The risk for post-shock ventricular arrhythmias is increased in patients with electrolyte disturbances and digitalis toxicity, so elective cardioversion should be delayed in such patients. Many patients develop elevations of serum troponin levels, sometimes with transient ST segment elevation, after cardioversion, especially if higher energies were delivered in a short period of time, but clinical myocardial dysfunction is rare.

Post-shock bradycardia or asystole, which may occur because of vagal discharge or an underlying sick sinus syndrome, sometimes can require atropine or emergency transcutaneous pacing. If a patient has a pacemaker or ICD, the shocking electrodes should be placed as far away from the generator as possible, and both the generator and pacing threshold should be checked after the procedure.

OTHER IMPLANTABLE DEVICES: CARDIOVERTER-DEFIBRILLATORS AND CARDIAC RESYNCHRONIZATION THERAPY

Implantable Cardioverter-Defibrillators

ICD Pulse Generators and Leads

The procedures for implanting ICDs are analogous to those used for permanent pacemakers. The 60-g pulse generators are similarly implanted subcutaneously in the infraclavicular area. ICDs deliver biphasic shocks at strengths of less than 1 to 42 J while recording the electrogram during the arrhythmia and its treatment. They also can provide antitachycardia overdrive pacing as well as dual-chamber antibradycardia pacing.

A developing alternative is the totally subcutaneous, implantable ICD.[6] With this device, defibrillation is achieved by current flowing between a pulse generator implanted in the axilla and a subcutaneous coil implanted parallel and just lateral to the sternum (Fig. 66-3). The pulse generator must deliver higher stored energy (80 J) compared with transvenous defibrillation to ensure an adequate margin of safety for defibrillation, but initial data suggest that the device's effectiveness for sensing and terminating VF is comparable to transvenous systems. The device is not suitable for patients with concomitant bradycardia, in those with indications for cardiac resynchronization therapy, or in patients in whom frequent antitachycardia pacing is needed or anticipated. However, it is likely to become an attractive option for patients who have had prior complications from transvenous systems (e.g., vein thrombosis, infection) and for primary prevention in patients in whom frequent shocks are not anticipated.

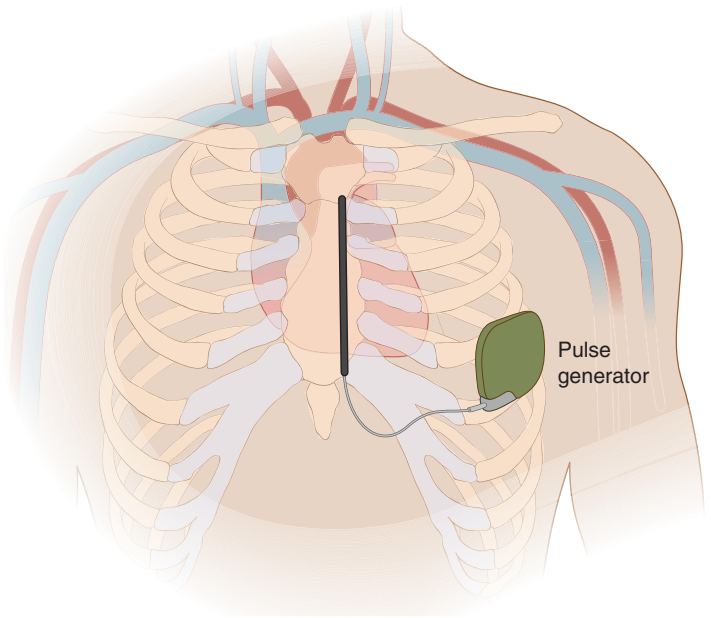

FIGURE 66-3. Schematic of pulse generator and lead position for the subcutaneous defibrillator.

Indications

ICD therapy initially evolved as secondary prevention to treat patients who had survived an episode of ventricular fibrillation or hemodynamically unstable ventricular tachycardia and were at high risk for subsequent death owing to recurrent ventricular arrhythmias.[6] Based on multiple randomized trials and careful risk-to-benefit analyses, ICDs now are also implanted as primary prevention in individuals at high risk for a first cardiac arrest (Table 66-3), including patients who have idiopathic dilated cardiomyopathy (Chapter 60) and unexplained episodes of syncope; patients who have dilated ischemic or nonischemic cardiomyopathy with an ejection fraction of 35% or less and class II or III heart failure; patients who have coronary artery disease, an ejection fraction or 35% or less, spontaneous episodes of nonsustained VT, and inducible sustained VT in the electrophysiology laboratory (Chapter 65); patients who have had a previous myocardial infarction and now have an ejection fraction of less than 30% (Chapter 73); and selected patients with conditions such as hypertrophic cardiomyopathy, Brugada syndrome, and long-QT syndrome (Chapters 60 and 65).[7,8] Further refinements in these criteria are to be expected over the coming years as data from ongoing and planned trials become available.

Programming of ICDs

A fundamental goal of ICD implantation and lead configuration is to deliver sufficient energy to the ventricular myocardium to ensure reliable defibrillation. The most common pathway for this energy is from the implanted pectoral pulse generator to a coil electrode on the distal portion of the transvenous lead that is positioned in the right ventricle. The energy required for successful defibrillation is probabilistic, and the relationship between successful defibrillation and energy of the shock is a sigmoidal curve, with an intermediate zone in which the success rate of any single defibrillation attempt is variable. Most transvenous ICD pulse generators can store and deliver 30 to 40 J, which exceeds the typical defibrillation energy threshold of 10 to 15 J, thereby allowing a substantial margin of safety for defibrillation. If the

standard lead configuration is insufficient for reliable defibrillation, alternative configurations can be tested at the time of implantation.

ICDs can perform a variety of functions beyond defibrillation. A substantial proportion of ICD recipients have sustained monomorphic VT that can be painlessly terminated by appropriately timed, overdrive antitachycardia pacing. ICDs can be programmed to deliver different combinations of antitachycardia pacing rates and shocks depending on the rate of the spontaneous arrhythmia. ICDs also incorporate a variety of programmable detection algorithms designed to withhold unnecessary therapies for brief episodes of rapid supraventricular arrhythmias, especially sinus tachycardia or atrial fibrillation (Fig. 66-4). In addition, many ventricular arrhythmias, particularly at lower rates, will self-terminate within seconds, so algorithms designed to delay cardioversion for a few seconds can reduce unnecessary shocks. Optimal programming reduces the patient's discomfort, maximizes the life of the pulse generator battery, and reduces the potential adverse effects of unneeded shocks on left ventricular function. For example, optimized programming (incorporating higher rate cutoffs, detection delay, rhythm discriminators, and antitachycardia overdrive pacing can reduce inappropriate device shocks by 50% and mortality by 30% compared with standard ICD programming, without increasing the risk for syncope.[■■] Many athletes with ICDs can engage in vigorous and competitive sports without physical injury or failure to terminate the arrhythmia.[9]

TABLE 66-3 INDICATIONS FOR IMPLANTABLE CARDIOVERTER-DEFIBRILLATOR IMPLANTATION

CLASS I*: SECONDARY PREVENTION
Cardiac arrest survivor (VF or unstable sustained VT not associated with a completely reversible cause)
Spontaneous sustained VT (irrespective of stability) and structural heart disease
Syncope of unknown origin associated with clinically relevant and hemodynamically significant sustained VT or VT induced at electrophysiologic testing
CLASS I*: PRIMARY PREVENTION
Prior myocardial infarction (≥40 days), LVEF < 35%, NYHA class II or III
Prior myocardial infarction (≥40 days), LVEF < 30%, NYHA class I
Nonischemic dilated cardiomyopathy, LVEF ≤ 35%, NYHA class II or III
Prior myocardial infarction, LVEF < 40%, spontaneous NSVT, inducible sustained VT or VF at electrophysiologic testing
CLASS IIA†
Nonischemic dilated cardiomyopathy, significant LV dysfunction, and syncope
Sustained VT and normal or near-normal ventricular function
Hypertrophic cardiomyopathy and one or more risk factors for sudden death
Arrhythmogenic right ventricular cardiomyopathy and one or more risk factors for sudden death
Long QT syndrome with syncope or VT despite β-blocker therapy
Nonhospitalized patients awaiting cardiac transplantation
Brugada syndrome and syncope or VT
Catecholaminergic polymorphic VT and syncope or VT on β-blocker therapy
Cardiac sarcoidosis, giant cell myocarditis, or Chagas disease

*Class I indications are conditions for which an implantable cardioverter-defibrillator is indicated.
†Class IIA indications are conditions for which an implantable cardioverter-defibrillator is reasonable.
LVEF = left ventricular ejection fraction; NSVT = nonsustained ventricular tachycardia, VF = ventricular fibrillation; VT = ventricular tachycardia.
Adapted from Tracy CM, Epstein AE, Darbar D, et al. 2012 ACCF/AHA/HRS focused update of the 2008 guidelines for device-based therapy of cardiac rhythm abnormalities: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2012;60:1297-1313.

Cardiac Resynchronization Therapy

Cardiac resynchronization therapy (CRT) requires the transvenous placement of electrodes into the right ventricle and into the coronary venous system for synchronous pacing of both ventricles (Fig. 66-5).

Indications for CRT

Abnormal and prolonged ventricular activation, as indicated by QRS prolongation of more than 120 msec on the electrocardiogram (ECG), contributes to poorly coordinated dyschronous activation of the left ventricle, thereby resulting in spontaneously reduced systolic function as would occur iatrogenically with right ventricular pacing. Over time, these abnormal electrical patterns can lead to progressive ventricular remodeling with dilation, further impairment of systolic function, and new or worsening heart failure (Chapter 58). Implantation of an additional ventricular lead, typically into a posterolateral coronary vein, permits pacing of the region of latest left ventricular activation. Synchronizing the timing between pacing of the right ventricular septum and the left ventricular free wall promotes more synchronous ventricular contraction and results in "reverse remodeling" with a reduction in ventricular volumes, improved systolic function, and clinical improvement in heart failure (Chapter 59).[10] In clinical trials of patients with reduced systolic function, prolonged QRS duration, and class III or ambulatory class IV heart failure, the addition of CRT to guideline-directed medical therapy is associated with an approximately 30% reduction in hospitalizations and a 24 to 36% reduction in total mortality.[■] Functional improvement in heart failure symptoms and quality of life is seen in a majority of patients, but 30 to 40% of patients may not improve symptomatically. Superior outcomes are associated with longer QRS delays and a left bundle branch block QRS configuration.[2] Although a prolonged QRS duration is typically associated with mechanical dyssynchrony as demonstrated by cardiac imaging, there is no evidence that imaging-identified mechanical dyssynchrony alone, in the absence of QRS prolongation, identifies patients who will improve with CRT.

Clinical trials have also established that the benefit of CRT extends to patients with less severe heart failure symptoms (class I or II), in whom CRT

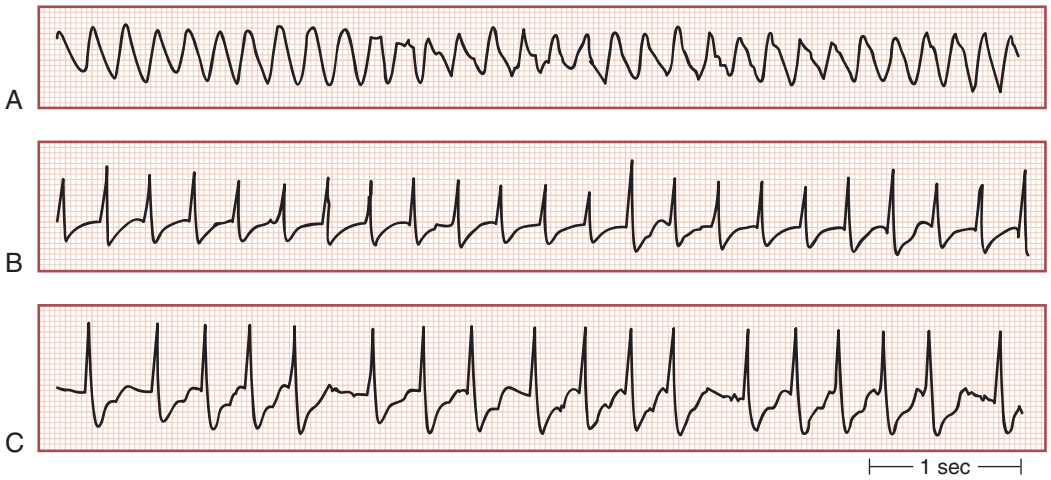

FIGURE 66-4. Examples of stored electrograms obtained several hours after three different patients had experienced a flurry of shocks from an implantable cardioverter-defibrillator and showing the rhythm recorded by the device immediately before a shock was delivered. A, In this patient, the stored electrogram demonstrates ventricular tachycardia at a rate of 300 beats per minute, thus indicating that the shock was appropriate. He was treated with amiodarone to reduce the frequency of episodes of ventricular tachycardia. **B,** This patient received shocks because of paroxysmal supraventricular tachycardia at a rate of 206 beats per minute, which exceeded the programmed rate cutoff of 170 beats per minute. He underwent radio frequency ablation of the paroxysmal supraventricular tachycardia and received no further inappropriate shocks. **C,** The stored electrograms in this patient indicate that the patient received inappropriate shocks that were triggered by atrial fibrillation at a rate of 180 beats per minute. The rate cutoff of the device in this patient was 150 beats per minute. This patient was treated with a β-blocker to keep the ventricular rate less than 150 beats per minute during atrial fibrillation. (Courtesy of Dr. Fred Morady.)

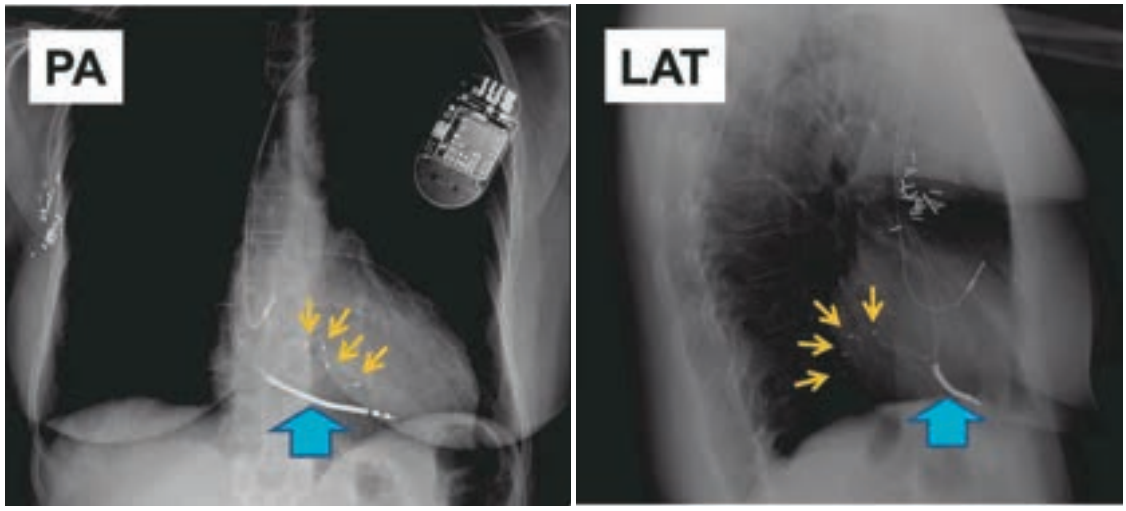

FIGURE 66-5 **Typical position of the left ventricular pacing lead in a posterolateral branch of the coronary sinus in a patient with a cardiac resynchronization device.** Note the presence of a defibrillator coil on the distal right ventricular lead, indicating that the device is also capable of defibrillation (CRT-D).

appears to delay the onset of symptomatic heart failure and significantly reduce heart failure events over the next 1 to 7 years if the ejection fraction is 30% or less and the QRS duration is more than 130 msec, especially in patients with left bundle branch block. Current indications for CRT therapy (Table 66-4) likely will continue to evolve as additional evidence accrues. In patients who meet criteria for both CRT and ICD implantation, some evidence suggests that implantation of devices with both functions (CRT-D) may provide additional mortality benefit.[3] However, CRT alone may be appropriate for some patients with more advanced heart failure, extensive comorbidity, and limited life expectancy, in whom the primary objective is symptomatic improvement.

Complications

Many of the complications related to ICD and CRT implantation procedures are somewhat more frequent but are similar to those associated with pacemakers. Major procedure-related complications include pneumothorax, myocardial perforation, and infection, all of which should have an incidence of less than 1%. The approach to infection is similar as for permanent pacemakers. Overall, major complications occur in 2 to 3% of new ICD implants and are more common during pulse generator replacement (5 to 6%). Long-term complications are primarily related to infection and lead failure. Given their larger size and complexity, ICD leads are more likely to fail (1 to 4% annually) than are pacemaker leads (<0.5% annually). Lead failure is most often due to insulation failure or fracture of the conductor wires. Survival of contemporary ICD leads has been estimated at 80 to 98% at 5 years, with lead failure rates accelerating thereafter. Lead failure is highly dependent on the lead's design and on the materials used. CRT devices have the highest likelihood of procedure-related complications, predominantly owing to dislodgement of the coronary sinus lead, and reoperation is required in 4 to 8% of patients.

Although an occasional therapeutic ICD discharge is common, flurries of discharges require urgent evaluation to determine the cause (see Fig. 66-3). Such causes can range from flurries of VT or VF, which may have a correctable precipitant (e.g., an electrolyte imbalance, drug toxicity), to AF or another SVT with a rapid ventricular response, lead fracture, or insulation failure. In some cases, antiarrhythmic drug therapy (see Tables 64-5 and 64-6), catheter ablation, or both may reduce or eliminate the arrhythmias associated with frequent shocks.

TABLE 66-4 INDICATIONS FOR IMPLANTATION OF CRT DEVICE

CLASS I INDICATIONS*
LVEF ≤ 35%, sinus rhythm, LBBB (≥150 msec), NYHA class II, III, or ambulatory class IV on guideline-directed medical therapy
CLASS IIA INDICATIONS†
LVEF ≤ 35%, sinus rhythm, LBBB (120-149 msec), NYHA class II, III, or ambulatory class IV on guideline-directed medical therapy
LVEF < 35%, sinus rhythm, non-LBBB (≥150 msec), NYHA class III or ambulatory class IV on guideline-directed medical therapy
LVEF ≤ 35, atrial fibrillation, on guideline-directed medical therapy, and both: • Requires ventricular pacing or otherwise meets CRT criteria • Near 100% pacing with CRT can be achieved by either atrioventricular node ablation or pharmacologic rate control
LVEF ≤ 35% on guideline-directed medical therapy and are undergoing new or replacement device implantation with anticipated requirement for significant (>40%) ventricular pacing

*Class I indications are conditions for which CRT is indicated.
†Class IIA indications are conditions for which CRT is reasonable.
CRT = cardiac resychronization therapy; LBBB = left bundle branch block; LVEF = left ventricular ejection fraction; NYHA=New York Heart Association (functional class).
Adapted from Brignole M, Auricchio A, Baron-Esquivias G, et al. 2013 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: the Task Force on cardiac pacing and resynchronization therapy of the European Society of Cardiology (ESC). Developed in collaboration with the European Heart Rhythm Association (EHRA). *Eur Heart J.* 2013;34:2281-2329; and Tracy CM, Epstein AE, Darbar D, et al. 2012 ACCF/AHA/HRS focused update of the 2008 guidelines for device-based therapy of cardiac rhythm abnormalities: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2012;60:1297-1313.

CATHETER ABLATION

Catheter-based ablation techniques are based on the concept that each arrhythmia requires a critical anatomic region or regions to initiate and maintain the formation and propagation of the abnormal impulse. Selective destruction of myocardial tissue in these areas can eliminate the arrhythmia. Depending on the type of arrhythmia, target sites for ablation are selected by recording the electrical activation sequence during an episode of sustained tachycardia, with a goal of identifying a discrete site of origin or a critical component of a larger reentrant circuit. For these arrhythmias, reproduction of the clinical arrhythmia during diagnostic electrophysiologic testing (Chapter 62) is critical and can be facilitated by careful selection of pacing sites, judicious use of sedation, and intravenous infusion of catecholamines. Alternatively, the target site can sometimes be selected based on specific anatomic landmarks or tissue characteristics identified during sinus rhythm.

Tissue Effects of Applied Energy

Radio frequency current, typically in the range of 300 to 750 kHz, is the most common form of energy used in catheter ablation. The energy is applied in a unipolar fashion between a small electrode in contact with the targeted myocardium and a large dispersive cutaneous patch electrode placed on the back. The small electrode area at the myocardial interface results in a high-density current and rapid resistive heating in the myocardium that is immediately subjacent to the electrode, with slower conductive heating of deeper myocardial layers. A tissue temperature greater than 60° C is required for irreversible myocyte injury. Saline irrigation of the electrode reduces heating at the tissue-electrode interface, moves the zone of maximal heating deeper into the tissue, and results in larger, deeper lesions. Among other energy sources,

cryoablation is the most widely used, whereas microwave, laser, and ultrasound have limited applications at present.

Radio Frequency Ablation of Supraventricular Tachycardias

Radio frequency ablation is the recommended first-line treatment for paroxysmal SVT, Wolff-Parkinson-White (WPW) syndrome, or type 1 (typical) atrial flutter that is symptomatic enough to warrant therapy (Chapter 64). For atrial flutter other than type 1 and for inappropriate sinus tachycardia, an ablation procedure is recommended only in patients who have significant symptoms and recurrences despite antiarrhythmic medications. AV nodal re-entrant tachycardia (Chapter 64), which is the most common type of paroxysmal SVT, is successfully eliminated in 98% of cases (with a <1% risk for high-degree AV block) by radio frequency ablation of the "slow" limb of the re-entry circuit, usually at the posteroseptal aspect of the right atrium, near the ostium of the coronary sinus. Cryoablation, which is associated with a lower risk for AV block but also a lower long-term success rate, may be considered for patients at higher risk for AV block, such as small children.

Left-sided accessory pathways are ablated by using either a retrograde aortic or a transseptal approach, whereas right-sided and septal lesions are ablated with a venous approach. Detailed mapping is essential to identify the optimal site for ablation, usually on either the atrial or the ventricular aspect of the mitral or tricuspid annulus. For the ablation of an accessory pathway, the success rate is 90 to 98%, with an overall complication rate of 2 to 3% and a less than 0.1% risk for a fatal complication. The most common serious complications are cardiac tamponade, owing to mechanical perforation of the heart by an electrode catheter, and high-degree AV block when the accessory pathway is near the AV node. Cryoablation is a reasonable alternative when the accessory pathway is near the AV node.

Most atrial tachycardias arise in the right atrium and are mapped using a venous approach, but left atrial tachycardias require a transseptal approach. Approximately 10 to 20% of patients with atrial tachycardia have more than one focus, and this rhythm may also be observed in association with other forms of SVT, particularly AV nodal re-entrant tachycardia. If the atrial tachycardia originates from a single site, ablation has about a 90% success rate, and complications are rare. For patients with multiple sources of atrial tachycardia, long-term success rates are lower.

Type 1 atrial flutter (Chapter 64), which arises in the right atrium, can be eliminated by ablation directed at a critical isthmus located in the low right atrium, between the tricuspid annulus and the inferior vena cava. The long-term success rate is more than 90%, with a less than 1% risk for serious complications.

Ablation of Atrial Fibrillation

Strategies for catheter ablation of AF (Chapter 64) are evolving.[11] At present, the primary aim of AF ablation is to reduce symptoms caused by recurrent episodes and thereby to improve quality of life. For patients with recurrent paroxysmal AF, a primary driver of the arrhythmia is from focal sources in the regions surrounding the proximal portions of the pulmonary vein and other thoracic veins as they insert into atrial myocardium. However, focally triggered arrhythmias may arise from other right and left atrial sites in 10 to 20% of patients.

For patients with recurrent symptomatic paroxysmal AF despite at least one antiarrhythmic drug trial, radio frequency or cryoballon catheter ablation reduces the risk for recurrent atrial arrhythmias by 50 to 70% and significantly improves quality of life compared with continued attempts to control the arrhythmia with alternative drug therapy.[12,13] Clinical trial data also demonstrate improved outcomes for reducing recurrent AF in selected patients with paroxysmal AF in whom catheter ablation is used as first-line treatment,[A1,A2] although rates of recurrence remain in the 50% range at 2 years even in patients who underwent ablation. Because AF is a progressive disease, however, the window of opportunity for intervention to prevent progression from paroxysmal to permanent AF or to prevent stroke, heart failure, or death may be limited. Ongoing large clinical trials are likely to provide more definitive information on long-term outcomes.

For patients with continuous persistent AF of greater than 1-year duration, catheter ablation is less successful. In patients with symptomatic heart failure, a left ventricular ejection fraction of 35% or less, and persistent AF, ablation can improve symptoms and neurohormonal status compared with rate control therapy.[A3] However, isolation of the pulmonary veins alone may be insufficient to maintain sinus rhythm. Adjunctive or alternative strategies, guided by anatomic considerations or mapping during AF, are the subject of ongoing investigation, and their long-term success is unclear.

Overall, the 1-year success rate of catheter ablation is 75 to 85% for paroxysmal AF and 60 to 75% for longstanding persistent AF. The most serious complications of AF ablation are atrial perforation, thromboembolism, and atrioesophageal fistula, with an overall risk of about 2%. Other complications are phrenic nerve injury or vascular access issues. Pulmonary vein stenosis is a potentially serious complication that can be avoided by not delivering energy within the tubular portion of the pulmonary veins.

In patients with refractory AF associated with an uncontrolled ventricular rate despite pharmacologic AV node blockade, ablation of the AV node can improve symptoms, functional capacity, and left ventricular function. In AV node ablation, third-degree AV block is intentionally induced with a success rate that approaches 100%. When the ablation lesions are placed sufficiently proximal in the AV junction, a junctional escape rhythm usually can be preserved. All patients require a permanent pacemaker to provide adequate rate response to physical activity.

Ablation of Ventricular Arrhythmias

Radio frequency ablation has an 85 to 100% success rate for the treatment of focal idiopathic VT (Chapter 65), whether it arises in the outflow tract of the right ventricle with a left bundle branch block configuration and superior axis, or arises from other sites, including the AV valve annuli, the sinuses of Valsalva, the left ventricular septum, or the papillary muscles. Given these outcomes, ablation is recommended in symptomatic patients, either after failure of initial drug therapy or as first-line therapy depending on a patient's preference.[14] Complications have been rare, and patients can avoid medications or an ICD.

By comparison, VT in patients with coronary artery disease usually arises in diseased tissue adjacent to an area of previous infarction in the left ventricle. Radio frequency ablation is not usually curative because the disease process is diffuse and the VT may originate from multiple sites. However, radio frequency ablation can be used as adjunctive therapy to reduce the number of ICD discharges, with a success rate of 65 to 95% and serious complications in about 5% of patients.

ARRHYTHMIA SURGERY

Ventricular Tachycardia

Subendocardial resection or cryoablation of the scar tissue that triggers monomorphic VT (Chapter 65) in patients with a prior myocardial infarction can eliminate VT in selected patients, with a success rate of 85 to 90% but an operative mortality rate of 5 to 15% even in experienced centers. Because of this high mortality rate, the procedure is limited to patients who have recurrent VT and other indications for cardiac surgery, such as large aneurysms associated with heart failure or the need to implant a left ventricular assist device.

Atrial Fibrillation

In the Maze procedure, a series of incisions or linear lesions or both are created by cryoablation or radio frequency ablation in the specific regions of the left and right atria to subdivide the atria into parts too small to sustain AF.[15] The success rate for eliminating AF is about 90%, and the operative mortality rate is less than 2%. The most common indication for the Maze procedure currently is for treatment of symptomatic AF in patients undergoing other cardiac surgical procedures, such as coronary revascularization or valve replacement or repair. A variety of simpler and even minimally invasive operative procedures have been developed for AF, but their long-term efficacy and their role in the treatment of AF remain unclear.

Grade A References

A1. Sweeney MO, Bank AJ, Nsah E, et al. Minimizing ventricular pacing to reduce atrial fibrillation in sinus-node disease. *N Engl J Med*. 2007;357:1000-1008.

A2. Lamas GA, Lee KL, Sweeney MO, et al. Ventricular pacing or dual-chamber pacing for sinus-node dysfunction. *N Engl J Med*. 2002;346:1854-1862.

A3. Curtis AB, Worley SJ, Adamson PB, et al. Biventricular pacing for atrioventricular block and systolic dysfunction. *N Engl J Med*. 2013;368:1585-1593.

A4. Birnie DH, Healey JS, Wells GA, et al. Pacemaker or defibrillator surgery without interruption of anticoagulation. *N Engl J Med*. 2013;368:2084-2093.

A5. Moss AJ, Schuger C, Beck CA, et al. Reduction in inappropriate therapy and mortality through ICD programming. *N Engl J Med*. 2012;367:2275-2283.

A6. Tan VH, Wilton SB, Kuriachan V, et al. Impact of programming strategies aimed at reducing nonessential implantable cardioverter defibrillator therapies on mortality: a systematic review and meta-analysis. *Circ Arrhythm Electrophysiol*. 2014;7:164-170.

A7. Cleland JG, Abraham WT, Linde C, et al. An individual patient meta-analysis of five randomized trials assessing the effects of cardiac resynchronization therapy on morbidity and mortality in patients with symptomatic heart failure. *Eur Heart J*. 2013;34:3547-3556.
A8. Goldenberg I, Kutyifa V, Klein HU, et al. Survival with cardiac-resynchronization therapy in mild heart failure. *N Engl J Med*. 2014;370:1694-1701.
A9. Cosedis Nielsen J, Johannessen A, Raatikainen P, et al. Radiofrequency ablation as initial therapy in paroxysmal atrial fibrillation. *N Engl J Med*. 2012;367:1587-1595.
A10. Morillo CA, Verma A, Connolly SJ, et al. Radiofrequency ablation vs antiarrhythmic drugs as first-line treatment of paroxysmal atrial fibrillation (RAAFT-2): a randomized trial. *JAMA*. 2014;311:692-700.
A11. Jones DG, Haldar SK, Hussain W, et al. A randomized trial to assess catheter ablation versus rate control in the management of persistent atrial fibrillation in heart failure. *J Am Coll Cardiol*. 2013;61:1894-1903.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Gillis AM, Russo AM, Ellenbogen KA, et al. HRS/ACCF expert consensus statement on pacemaker device and mode selection. *J Am Coll Cardiol*. 2012;60:682-703.
2. Brignole M, Auricchio A, Baron-Esquivias G, et al. 2013 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: the Task Force on cardiac pacing and resynchronization therapy of the European Society of Cardiology (ESC). Developed in collaboration with the European Heart Rhythm Association (EHRA). *Eur Heart J*. 2013;34:2281-2329.
3. Tracy CM, Epstein AE, Darbar D, et al. 2012 ACCF/AHA/HRS focused update of the 2008 guidelines for device-based therapy of cardiac rhythm abnormalities: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2012;60:1297-1313.
4. Mulpuru SK, Pretorius VG, Birgersdotter-Green UM. Device infections: management and indications for lead extraction. *Circulation*. 2013;128:1031-1038.
5. Wann LS, Curtis AB, Ellenbogen KA, et al. Management of patients with atrial fibrillation (compilation of 2006 ACCF/AHA/ESC and 2011 ACCF/AHA/HRS recommendations): a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *Circulation*. 2013;127:1916-1926.
6. Aziz S, Leon AR, El-Chami MF. The subcutaneous defibrillator: a review of the literature. *J Am Coll Cardiol*. 2014;63:1473-1479.
7. Russo AM, Stainback RF, Bailey SR, et al. ACCF/HRS/AHA/ASE/HFSA/SCAI/SCCT/SCMR 2013 appropriate use criteria for implantable cardioverter-defibrillators and cardiac resynchronization therapy: a report of the American College of Cardiology Foundation appropriate use criteria task force, Heart Rhythm Society, American Heart Association, American Society of Echocardiography, Heart Failure Society of America, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance. *Heart Rhythm*. 2013;10:e11-e58.
8. Poole JE. Present guidelines for device implantation: clinical considerations and clinical challenges from pacing, implantable cardiac defibrillator, and cardiac resynchronization therapy. *Circulation*. 2014;129:383-394.
9. Lampert R, Olshansky B, Heidbuchel H, et al. Safety of sports for athletes with implantable cardioverter-defibrillators: results of a prospective, multinational registry. *Circulation*. 2013;127:2021-2030.
10. Neubauer S, Redwood C. New mechanisms and concepts for cardiac-resynchronization therapy. *N Engl J Med*. 2014;370:1164-1166.
11. Camm AJ, Lip GY, De Caterina R, et al. 2012 focused update of the ESC Guidelines for the management of atrial fibrillation: an update of the 2010 ESC Guidelines for the management of atrial fibrillation. Developed with the special contribution of the European Heart Rhythm Association. *Eur Heart J*. 2012;33:2719-2747.
12. Calkins H, Kuck KH, Cappato R, et al. 2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design. *Heart Rhythm*. 2012;9:632-696.
13. January CT, Wann LS, Alpert JS, et al. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. *J Am Coll Cardiol*. 2014;64:e1-e76.
14. Aliot EM, Stevenson WG, Almendral-Garrote JM, et al. EHRA/HRS expert consensus on catheter ablation of ventricular arrhythmias: developed in a partnership with the European Heart Rhythm Association (EHRA), a Registered Branch of the European Society of Cardiology (ESC), and the Heart Rhythm Society (HRS); in collaboration with the American College of Cardiology (ACC) and the American Heart Association (AHA). *Heart Rhythm*. 2009;6:886-933.
15. Robertson JO, Lawrance CP, Maniar HS, et al. Surgical techniques used for the treatment of atrial fibrillation. *Circ J*. 2013;77:1941-1951.

67 ARTERIAL HYPERTENSION

RONALD G. VICTOR

DEFINITION

Hypertension is defined as a usual office blood pressure of 140/90 mm Hg or higher (Table 67-1), blood pressure levels for which the benefits of drug treatment have been shown in randomized controlled trials. However, epidemiologic data show continuous positive relationships between the risk for death from coronary artery disease (CAD) and stroke with systolic or diastolic blood pressure values as low as 115/75 mm Hg (Fig. 67-1). The artificial dichotomy between "hypertension" and "normotension" may delay medical treatment until vascular health has been irreversibly compromised by elevated blood pressure values that were previously considered normal. As a result, guideline committees continue to debate how far to lower blood pressure with antihypertensive medication and whether to recommend drug therapy for high-risk patients with blood pressure in the "prehypertensive" range of 120 to 139/80 to 89 mm Hg.[1-4]

EPIDEMIOLOGY

Affecting one fourth of the adult population (78 million adults in the United States and more than 1 billion people worldwide), arterial hypertension is the leading cause of death in the world and the most common cause for an outpatient visit to a physician; it is the most easily recognized treatable risk factor for stroke (Chapters 406, 407, and 408), myocardial infarction (Chapters 72 and 73), heart failure (Chapters 58 and 59), peripheral vascular disease (Chapter 79), aortic dissection (Chapter 78), atrial fibrillation (Chapter 64), and end-stage kidney disease (Chapter 130). Because of increasing rates of obesity and aging of the population, hypertension is projected to affect 1.5 billion persons, one third of the world's population, by the year 2025. Presently, about 54% of strokes and 47% of ischemic heart disease worldwide is attributable to high blood pressure. Half of this disease burden is in people who meet the definition of hypertension, and the remainder is in people with lesser degrees of high blood pressure (*prehypertension*).

The asymptomatic nature of hypertension and the inherent variability in blood pressure delay diagnosis. Effective treatment requires frequent medical checkups and continuity of care by a knowledgeable clinician, both of which are less common in men and in members of low-income minority groups. Most cases of hypertension are multifactorial, and management remains empirical, often requiring three or more drugs with complementary mechanisms of action, in addition to any other medications that may be needed for concomitant medical conditions. Pill burden, prescription drug costs, medication side effects, and insufficient time for patient education contribute to nonadherence with medications. Busy primary care physicians often undertreat hypertension. Lifestyle modification (particularly diet and exercise) can lower blood pressure somewhat, but the reduction rarely is enough to eliminate the need for medication. For all these reasons, blood pressure remains elevated—140/90 mm Hg or higher—in more than half of affected individuals in the United States, with marked racial, ethnic, and gender disparities.[5] The resultant annual cost to the U.S. health care system exceeds $73 billion.

TABLE 67-1 STAGING OF OFFICE BLOOD PRESSURE*

BLOOD PRESSURE STAGE	SYSTOLIC BLOOD PRESSURE (mm Hg)	DIASTOLIC BLOOD PRESSURE (mm Hg)
Normal	<120	<80
Prehypertension	120-139	80-89
Stage 1 hypertension	140-159	90-99
Stage 2 hypertension	≥160	≥100

*Calculation of seated blood pressure is based on the mean of two or more readings on two separate office visits.
From Chobanian A, Bakris G, Black H, et al. The Seventh Report of the Joint National Committee on the Prevention, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA*. 2003;289:2560-2572.

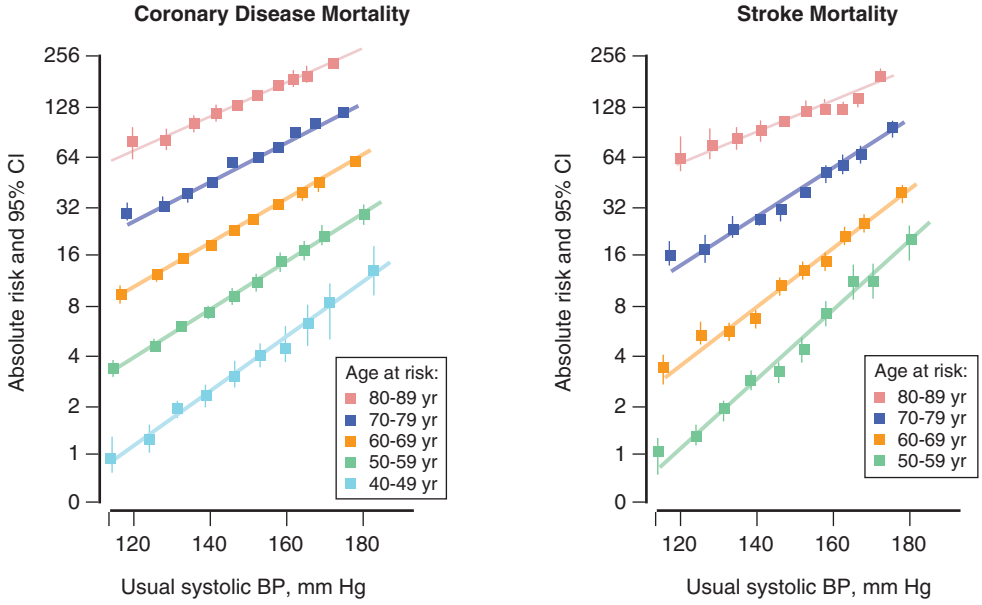

FIGURE 67-1. Absolute risk for coronary artery disease and stroke mortality by usual systolic blood pressure (BP) levels. CI = confidence interval. (From Lewington S, Clarke R, Qizilbash N, et al, for the Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*. 2002;360:1903-1913.)

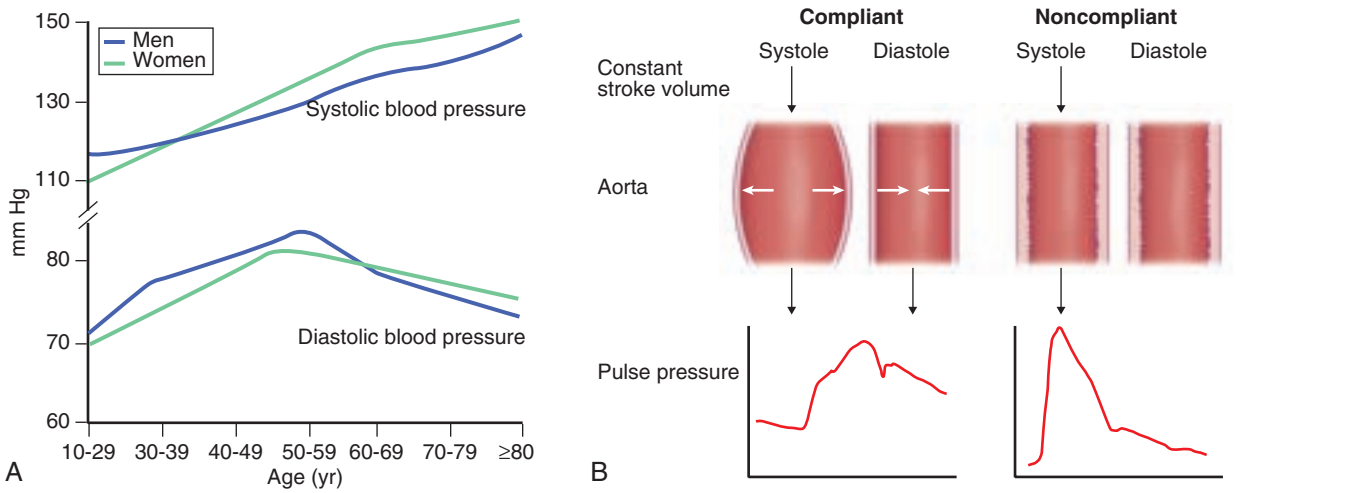

FIGURE 67-2. Aging and pulse pressure. A, Age-dependent changes in systolic and diastolic blood pressure in the United States. **B,** Schematic diagram showing the relation between aortic compliance and pulse pressure. **(A,** From Burt V, Whelton P, Rocella EJ, et al. Prevalence of hypertension in the U.S. adult population: results from the Third National Health and Nutrition Examination Survey, 1988-1991. *Hypertension.* 1995;25:305-313; **B,** Courtesy of Dr. Stanley Franklin, University of California at Irvine.)

Aging and Pulse Pressure

Patients often ask: is systolic or diastolic blood pressure more important? The answer is: both. In industrialized societies, systolic pressure rises progressively with age; if individuals live long enough, almost all (>90%) will develop hypertension. This age-dependent rise in blood pressure is not an essential part of human biology. In less developed countries where consumption of calories and salt is low, blood pressures remain low and do not rise with age. In developed countries, diastolic pressure rises until the age of 50 years and decreases thereafter, producing a progressive rise in pulse pressure (systolic pressure minus diastolic pressure) (Fig. 67-2).

Different hemodynamic faults underlie hypertension in younger and older persons. Patients who develop hypertension before the age of 50 years typically have *combined systolic and diastolic hypertension:* systolic pressure above 140 mm Hg *and* diastolic pressure above 90 mm Hg. The main hemodynamic fault is vasoconstriction at the level of the resistance arterioles. In contrast, most patients who develop hypertension after the age of 50 years have *isolated systolic hypertension:* systolic pressure above 140 mm Hg but diastolic pressure below 90 mm Hg (often below 80 mm Hg). In isolated systolic hypertension, the primary hemodynamic fault is decreased distensibility of the large conduit arteries. Collagen replaces elastin in the elastic lamina of the aorta, a process that is accelerated by both aging and hypertension. When pulse wave velocity increases sufficiently, the rapid return of the arterial pulse wave from the periphery augments central systolic (rather than diastolic) pressure. The augmented systolic load on the left ventricle increases myocardial oxygen demands, whereas the rapid diastolic runoff compromises myocardial perfusion. As the population ages, most uncontrolled hypertension occurs in older patients with isolated systolic hypertension.

Gender and Race/Ethnicity

Before the age of 50 years, hypertension is less common in women than men, suggesting a protective effect of estrogen. After menopause, hypertension is more common in women than men.

Forty percent of African American adults in the United States have hypertension, compared with 25% of white and Mexican American individuals.[5] In African Americans, hypertension also starts at a younger age, is more severe, and causes more target organ damage, leading to premature disability and death. In the Bogalusa Heart Study, African American children already had higher blood pressures than white children by grade school. African Americans also have a more rapid progression from prehypertension to hypertension. However, hypertension is more prevalent in the white populations of several European countries (e.g., Finland, Germany, and Spain) than in African Americans and is less prevalent among Africans living in Africa (Fig. 67-3), although hypertension is increasing in developing countries undergoing Westernization. These international data emphasize the importance of environmental factors.

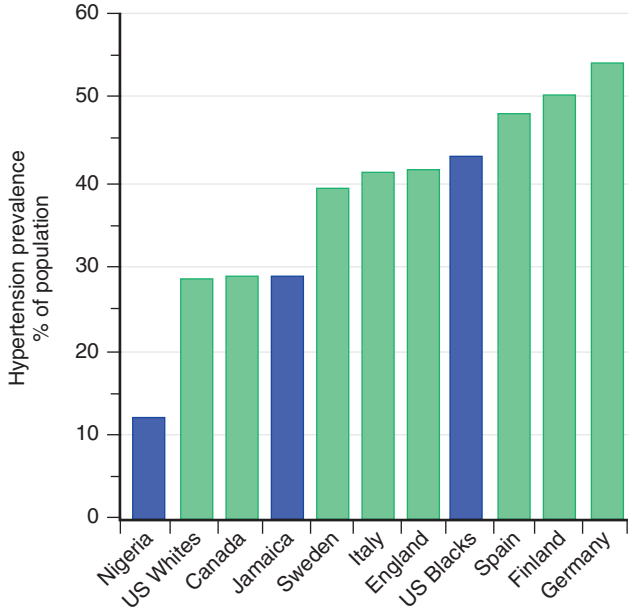

FIGURE 67-3. Geographic variation in hypertension prevalence in populations of African and European ancestries. (From Cooper RS, Wolf-Maier K, Luke A, et al. An international comparative study of blood pressure in populations of European vs. African descent. *BMC Med.* 2005;3:1-8.)

PATHOBIOLOGY

In 90 to 95% of hypertensive patients, a single reversible cause of the elevated blood pressure cannot be identified, hence the term *primary hypertension.* However, in most patients with primary hypertension, readily identifiable behaviors—habitually excessive consumption of calories, salt, or alcohol—contribute to the elevated blood pressure. In the remaining 5 to 10%, a more discrete mechanism can be identified, and the condition is termed *secondary or identifiable hypertension.* At the organ-system level, hypertension results from a gain in function of pathways that promote vasoconstriction and renal sodium retention or a loss in function of pathways that promote vasodilation and renal sodium excretion. Neural, hormonal, renal, and vascular mechanisms are involved. There is increasing evidence that neurohormonal activation contributes to the early pathogenesis by compromising vascular function (e.g., endothelium-dependent vasodilation) and structure (e.g., inward remodeling) that precede hypertension.

Behavioral Determinants of Human Blood Pressure Variation

The most important behavioral determinants of blood pressure are related to dietary consumption of calories and salt. Across populations, the prevalence of hypertension increases linearly with average body mass index. With the obesity epidemic in both developed and developing societies, increasing attention is being paid to the *metabolic syndrome* (Chapter 229) that often accompanies hypertension. The metabolic syndrome refers to the frequent clustering of elevated blood pressure with abdominal ("male pattern") adiposity, insulin resistance with glucose intolerance, and a dyslipidemic pattern consisting typically of elevated plasma triglyceride and low high-density lipoprotein cholesterol levels. In the Framingham Heart Study, obesity has been estimated to account for as much as 60% of the new cases of hypertension. The underlying mechanisms by which weight gain leads to hypertension are incompletely understood, but there is mounting evidence for an expanded plasma volume plus sympathetic overactivity. The sympathetic overactivity is thought to be a compensatory attempt to burn fat but at the expense of peripheral vasoconstriction, renal salt and water retention, and hypertension. In some obese individuals, sleep apnea (Chapter 100) is an important cause of hypertension. Repeated arterial desaturation sensitizes the carotid body chemoreceptors, causing sustained sympathetic overactivity even during waking hours.

Dietary sodium intake is another key behavioral determinant of human hypertension. In the INTERSALT study of 52 locations around the world, the risk for development of hypertension during three decades of adult life was linearly and tightly related to dietary sodium intake. Interindividual variability in blood pressure responses to dietary sodium loading and sodium restriction indicates an important genetic underpinning.

Genetic Determinants of Human Blood Pressure Variation

Concordance of blood pressures is greater within families than in unrelated individuals, greater between monozygotic twins than between dizygotic twins, and greater between biologic siblings than between adoptive siblings living in the same household. As much as 70% of the familial aggregation of blood pressure is attributed to shared genes rather than to shared environment.

The complex regulation of blood pressure has thwarted the genetic dissection of primary human hypertension, with the first positive genome-wide association studies suggesting multiple risk alleles, each having very small effects. Mutations in 20 salt-handling genes cause ultra-rare syndromes of severe, early-onset hypotension (salt-wasting syndromes) or hypertension (all inherited as mendelian traits). The clinical relevance of these mutations to common primary hypertension has been limited, although recent data indicate that heterozygous mutations in genes underlying the pediatric salt-wasting syndromes (Bartter and Gitelman) are present in 1 to 2% of the general adult population and confer resistance against primary hypertension.

CLINICAL MANIFESTATIONS

Hypertension is called the silent killer—an asymptomatic chronic disorder that silently damages the blood vessels, heart, brain, and kidneys if it is undetected and untreated. Although headaches (Chapter 398) are common in patients with mild to moderate hypertension, episodes of headaches do not correlate with fluctuations in blood pressure. Rather, they correlate with a person's awareness of his or her diagnosis.

DIAGNOSIS

Initial Evaluation for Hypertension

The initial evaluation for hypertension should accomplish three goals: (1) stage the blood pressure, (2) assess the patient's additional cardiovascular risk factors, and (3) detect clues of secondary hypertension that require further evaluation.

Goal 1: Accurate Assessment of Blood Pressure

Office Blood Pressure

Traditionally, blood pressure has been staged as normal, prehypertension, or hypertension based on the average of two or more readings taken at two or more office visits (see Table 67-1). The blood pressure should be measured at least twice after 5 minutes of rest with the patient seated, the back supported, and the arm bare and at heart level. A large adult-sized cuff should be used to measure blood pressure in overweight adults because the standard-sized cuff will cause falsely elevated readings. Tobacco and caffeine should be avoided for at least 30 minutes. Blood pressure should be measured in both arms, to exclude coarctation of the aorta, and after 5 minutes of standing, to exclude a significant postural fall, particularly in older patients and patients with diabetes or other conditions (e.g., Parkinson disease) that predispose to autonomic insufficiency.

Home and Ambulatory Blood Pressure Monitoring

A person's blood pressure varies so much throughout a 24-hour period that it is impossible to characterize it accurately except by repeated measurements under various conditions. Out-of-office readings are the only way to obtain a clear picture of a person's usual blood pressure for accurate diagnosis and management. These readings are more predictive of cardiovascular events than office readings and overcome many of the pitfalls of office measurement, including physician errors and "white coat" reactions. Home blood pressure monitoring can improve medication adherence by actively involving patients in their own medical care.

New recommendations include the following: (1) home blood pressure monitoring should become a routine part of the clinical management of patients with known or suspected hypertension the same way that home blood glucose monitoring is essential to the management of patients with diabetes; (2) two or three readings should be taken in the morning and at night for 4 days; the first day's readings should be discarded as being artificially high, and all the other readings should be averaged; and (3) the target treatment goal is an average home blood pressure of less than 135/85 mm Hg for most patients and less than 130/80 mm Hg for patients with proteinuric chronic kidney disease and possibly for some other high-risk patients.

A validated electronic oscillometric monitor with an arm cuff should be chosen from the dabl Educational Web site (www.dableducational.org). Each patient's monitor should be checked in the physician's office for its accuracy and appropriate cuff size. Patients must be taught correct measurement techniques and to avoid reporting bias. Wrist monitors are inaccurate and not recommended. The oscillometric method may not work well in patients with atrial fibrillation or frequent extrasystoles, for whom manual sphygmomanometry is required. Some patients will become obsessed about taking their blood pressure and need to stop.

Ambulatory blood pressure monitoring (Table 67-2) provides automated measurements of blood pressure during a 24-hour period while patients are engaged in their usual activities, including sleep (Fig. 67-4). Ambulatory blood pressure measurement is superior to standard office measurement in predicting fatal and nonfatal myocardial infarction and stroke. Recommended normal values are an average daytime blood pressure below 135/85 mm Hg, nighttime blood pressure below 120/70 mm Hg, and 24-hour blood pressure below 130/80 mm Hg (Table 67-3) Some experts have recommended a lower cutoff value of 130/80 mm Hg as a more stringent definition of normal daytime blood pressure.

TABLE 67-2 RECOMMENDED CLINICAL INDICATIONS FOR AMBULATORY BLOOD PRESSURE MONITORING

- Suspicion of white-coat hypertension
  - High office blood pressure in untreated patients with no target organ damage and low cardiovascular risk
- Suspicion of masked hypertension
  - Normal office blood pressure in untreated or treated patients with target organ damage or high cardiovascular risk
- Differentiation of pseudoresistant from truly resistant hypertension
- Labile blood pressure
  - Suspicion of dysautonomia: orthostatic hypotension ± supine hypertension, postprandial hypotension, baroreflex failure
  - Suspicion of drug-induced hypotension
- Suspicion of nocturnal hypertension in patients with sleep apnea, chronic kidney disease, or diabetes
- Assessment of hypertension in elderly people, children, and adolescents, and pregnant women.

Adapted from O'Brien E, Parati G, Stergiou G, et al. European society of hypertension position paper on ambulatory blood pressure monitoring. *J Hypertens.* 2013;31:1731-1768; and Mancia G, Fagard R, Narkiewicz K, et al. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J Hypertens.* 2013;31:1281-1357.

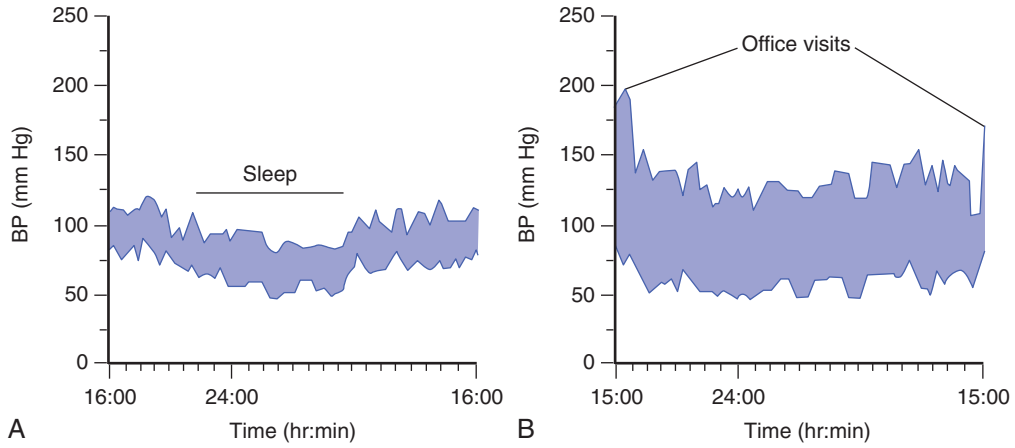

FIGURE 67-4. **The 24-hour ambulatory blood pressure (BP) monitor tracings of two different patients. A,** Optimal blood pressure in a healthy 37-year-old woman. Note the normal variability in blood pressure, the nocturnal dip in blood pressure during sleep, and the sharp increase in blood pressure on awakening. **B,** Pronounced white coat effect in an 80-year-old woman referred for evaluation of medically refractory hypertension. Documentation of the white coat effect prevented overtreatment of the patient's isolated systolic hypertension. **(A,** Provided by Ronald G. Victor, MD, Hypertension Center, Cedars-Sinai Heart Institute, Los Angeles, California; **B,** Courtesy of Wanpen Vongpatanasin, MD, Hypertension Division, Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas.)

TABLE 67-3 DEFINITIONS OF HYPERTENSION BY OFFICE AND OUT-OF-OFFICE BLOOD PRESSURE LEVELS

CATEGORY	SYSTOLIC (mm Hg)		DIASTOLIC (mm Hg)
Office blood pressure	≥140	and/or	≥90
Home blood pressure	≥135	and/or	≥85
Ambulatory blood pressure			
• Daytime (or awake)	≥135	and/or	≥85
• Nighttime (or sleep)	≥120	and/or	≥70
• 24 hour	≥130	and/or	≥80

Adapted from Mancia G, Fagard R, Narkiewicz K, et al. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J Hypertens.* 2013;31:1281-1357.

About 20% of patients with elevated office blood pressures have normal home or ambulatory blood pressures. If the daytime blood pressure is below 135/85 mm Hg (or preferably below 130/80 mm Hg) and there is no target organ damage despite consistently elevated office readings, the patient has "office-only" or "white coat" hypertension, caused by a transient adrenergic response to the measurement of blood pressure in the physician's office. If there are no other risk factors (such as metabolic syndrome), the cardiovascular risk is similar to that in persons with consistently normal blood pressure. Many patients do not have pure white coat hypertension but rather white coat aggravation, a white coat reaction superimposed on a milder level of out-of-office hypertension that nevertheless needs treatment. For example, up to 30% of treated patients who have persistently elevated office blood pressure readings will be shown by ambulatory monitoring to have adequate or even excessive control of their hypertension, thereby eliminating overtreatment (see Fig. 67-4). In other patients, office readings underestimate ambulatory blood pressures, presumably because of sympathetic overactivity (e.g., owing to job stress, home stress, or tobacco smoke) that dissipates when the patient comes to the office. Such "masked hypertension" carries the same increased cardiovascular risk as sustained office and home hypertension and is particularly common in men, elderly patients, and patients with diabetes or chronic kidney disease (E-Fig. 67-1). Both white coat and masked hypertension are so common in elderly patients that office blood pressure measurement alone, without home or ambulatory monitoring, would lead to either overtreatment or undertreatment in three out of every four patients.[6] Current U.K. and European guidelines place far greater emphasis than U.S. guidelines on home and ambulatory blood pressure monitoring for clinical decision making.

Ambulatory monitoring is the only way to detect hypertension during sleep. Blood pressure normally dips during sleep and increases sharply when a person awakens and becomes active. Nocturnal hypertension increases the aggregate blood pressure burden on the cardiovascular system and is a stronger predictor of cardiovascular outcomes than daytime ambulatory blood pressure or office measurements. Nocturnal hypertension is particularly common in patients with chronic kidney disease (Chapter 130), presumably because of their sustained sympathetic overactivity, which does not shut down during sleep, and centralization of blood volume with nocturnal recumbence. Nocturnal hypertension also is prevalent in African Americans, in whom the normal nocturnal dipping of blood pressure is often impaired.

Goal 2: Cardiovascular Risk Stratification

Although cardiovascular risk increases with increasing blood pressure, it also increases if the patient has hypertensive target organ damage or additional cardiovascular risk factors (Table 67-4). More than 75% of hypertensive patients will benefit from lipid-lowering statins (Chapter 206), and 25% have diabetes. Thus, the minimal laboratory testing required for the initial evaluation of hypertension is determination of blood electrolyte, fasting glucose, and serum creatinine levels (with calculated glomerular filtration rate [GFR]), a fasting lipid panel, hematocrit, spot urinalysis (including urine albumin-to-creatinine ratio), and a resting 12-lead electrocardiogram (ECG).

The gradient of increasing levels of blood pressure with cardiovascular risk steepens as additional risk factors are added. The patient's global cardiovascular risk should be estimated from the 2013 ACC/AHA pooled atherosclerotic cardiovascular disease (ASCVD) risk calculator (http://www.cardiosource.org/Science-And-Quality/Practice-Guidelines-and-Quality-Standards/2013-Prevention-Guideline-Tools.aspx). Decisions regarding treatment thresholds and treatment targets, however, still depend largely on specific blood pressure values rather than on an individual's global cardiovascular risk. Higher risk hypertensive patients are more likely to be treated with blood pressure medications but less likely to have their office blood pressure controlled to less than 140/90 mm Hg.

Goal 3: Identification and Treatment of Secondary (Identifiable) Causes of Hypertension

The third goal of the initial evaluation is to screen for identifiable causes of hypertension (Table 67-5), in the hope of finding a surgical cure. A thorough search for secondary causes, which is not cost effective in most patients with hypertension, becomes critically important in two circumstances: (1) when there is a compelling finding on the initial evaluation and (2) when the hypertensive process is so severe that it either is refractory to intensive multiple-drug therapy or requires hospitalization.

RENAL PARENCHYMAL HYPERTENSION

Chronic kidney disease (Chapter 130) is the most common cause of secondary hypertension. Hypertension is present in more than 85% of patients with chronic kidney disease and is a major factor causing their increased cardiovascular morbidity and mortality. The mechanisms causing the hypertension include an expanded plasma volume and peripheral vasoconstriction; the peripheral vasoconstriction is caused by both activation of vasoconstrictor

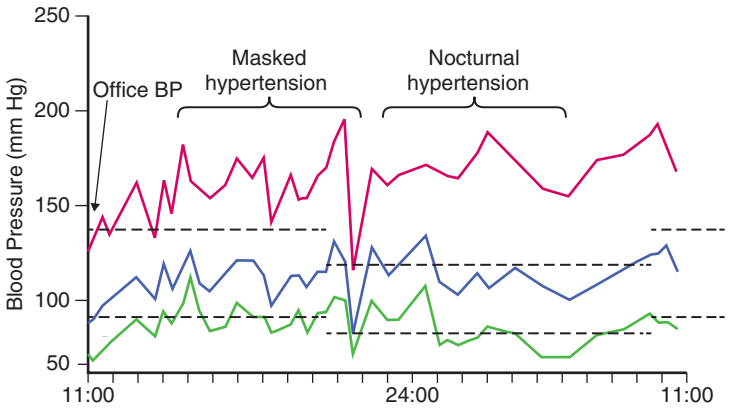

E-FIGURE 67-1. The 24-hour ambulatory blood pressure (BP) recording of a patient with apparently normal office blood pressure but with masked hypertension and nocturnal hypertension. The patient had left ventricular hypertrophy by electrocardiogram and nondiabetic chronic kidney disease, which is a common cause of masked and nocturnal hypertension. Antihypertensive medication was intensified to manage the out-of-office blood pressure. (Provided Ronald G. Victor, MD, Hypertension Center, Cedars-Sinai Heart Institute, Los Angeles, California.)

pathways (renin-angiotensin and sympathetic nervous systems) and inhibition of vasodilator pathways (nitric oxide).

Measurement of serum creatinine alone is an inadequate screening test for renal insufficiency. A spot urine specimen should be obtained to screen for microalbuminuria, which is defined as a urine albumin-to-urine creatinine ratio of 30 to 300 mg/g (equivalent to excretion of 30 to 300 mg of albumin per 24 hours); higher levels of albuminuria indicate more advanced kidney disease. Using the spot urine specimen, creatinine clearance should be calculated (www.nephron.com) (Chapter 114) to screen for an estimated GFR below 60 mL/minute per 1.73 m^2.

TABLE 67-4 FACTORS OTHER THAN BLOOD PRESSURE LEVEL THAT INFLUENCE GLOBAL CARDIOVASCULAR RISK IN PATIENTS WITH HYPERTENSION

RISK FACTORS

- Male
- Age (men ≥55 yr, women ≥65 yr)
- Smoking
- Dyslipidemia
- Impaired fasting glucose (100-125 mg/dL)
- Obesity (BMI ≥ 30 kg/m^2 or waist circumference: men, ≥102 cm, women, ≥88 cm)
- Family history or premature cardiovascular disease (men aged < 55 yr, women aged < 65 yr)

ASYMPTOMATIC TARGET ORGAN DAMAGE

- Left ventricular hypertrophy by ECG or transthoracic echocardiography
- Chronic kidney disease (eGFR ≤ 60 mL/min/1.73 m^2)
- Microalbuminuria (albumin-to-creatinine ratio, 30-300 mg/g)
- Ankle-brachial index < 0.9
- Pulse wave velocity > 10 m/sec

DIABETES MELLITUS

(fasting plasma glucose ≥126 mg/dL × 2; or hemoglobin A_{1C} ≥ 7%; or postload plasma glucose > 198 mg/dL)

ESTABLISHED CARDIOVASCULAR OR RENAL DISEASE

- Stroke or TIA
- CAD: myocardial infarction, angina, myocardial revascularization
- Heart failure (with decreased or preserved ejection fraction)
- Intermittent claudication (symptomatic peripheral artery disease)
- Chronic kidney disease with eGFR < 30 mL/min/1.73m^2
- Advanced retinopathy: hemorrhages or exudates, papilledema

BMI = body mass index; CAD = coronary artery disease; ECG = electrocardiogram; eGFR = estimated glomerular filtration rate; TIA = transient ischemic attack.
Adapted from Mancia G, Fagard R, Narkiewicz K, et al. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J Hypertens.* 2013;31:1281-1357.

In patients with mild (stage 2: GFR of 60 to 90 mL/minute per 1.73 m^2) or moderate (stage 3: GFR of 30 to 60 mL/minute per 1.73 m^2) proteinuric chronic kidney disease, stringent blood pressure control is important both to slow the progression to end-stage renal disease and to reduce the excessive cardiovascular risk. In patients with severe chronic kidney disease, hypertension often becomes difficult to treat and may require either (1) intensive medical treatment with loop diuretics, potent vasodilators (e.g., minoxidil), high-dose β-adrenergic blockers, and central sympatholytics; or (2) initiation of chronic hemodialysis as the only effective way to reduce plasma volume. In chronic hemodialysis patients, the challenge is to control interdialytic hypertension without exacerbating dialysis-induced hypotension. The annual mortality rate in the hemodialysis population is 25%; half of this excessive mortality is caused by cardiovascular events that are related, at least in part, to hypertension.

RENOVASCULAR HYPERTENSION

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

The two main causes of renal artery stenosis (Chapter 125) are atherosclerosis (85% of cases), typically in older persons with other clinical manifestations of systemic atherosclerosis, and fibromuscular dysplasia (15% of cases), typically in young women who are otherwise healthy. Although renal artery stenosis and hypertension frequently coexist, the presence of a renal artery stenosis proves neither that the patient's hypertension is renovascular in origin nor that revascularization will improve renal perfusion and blood pressure.

Unilateral renal artery stenosis can lead to underperfusion of the juxtaglomerular cells, thereby causing renin-dependent hypertension even though the contralateral kidney is able to maintain normal blood volume. In contrast, bilateral renal artery stenosis (or unilateral stenosis with a solitary kidney) constitutes a potentially reversible cause of progressive renal failure and volume-dependent hypertension. The following clinical clues increase the suspicion of renovascular hypertension: any hospitalization for urgent or emergent hypertension; recurrent "flash" pulmonary edema; recent worsening of long-standing, previously well-controlled hypertension; severe hypertension in a young adult or after the age of 50 years; precipitous and progressive worsening of renal function in response to angiotensin-converting enzyme (ACE) inhibition or angiotensin II receptor blockade; unilateral small kidney by any radiographic study; extensive peripheral arteriosclerosis; and a flank bruit.

DIAGNOSIS

Contrast-enhanced computed tomography (CT) and magnetic resonance angiography are the preferred screening tests for renal artery stenosis, but gadolinium-enhanced magnetic resonance imaging (MRI) is contraindicated in patients with advanced chronic kidney disease to avoid potentially fatal gadolinium-induced nephrogenic systemic fibrosis (Chapter 267). Fibromuscular dysplasia classically causes a "string of beads" lesion in the midportion of a renal artery (Fig. 67-5A), whereas atherosclerotic renal artery lesions

TABLE 67-5 GUIDE TO EVALUATION OF IDENTIFIABLE CAUSES OF HYPERTENSION

SUSPECTED DIAGNOSIS	CLINICAL CLUES	DIAGNOSTIC TESTING
Chronic kidney disease	Estimated GFR < 60 mL/min/1.73 m^2 Urine albumin-to-creatinine ratio ≥ 30 mg/g	Renal sonography
Renovascular disease	New elevation in serum creatinine, marked elevation in serum creatinine with ACE inhibitor or ARB, drug-resistant hypertension, flash pulmonary edema, abdominal or flank bruit	Renal sonography (atrophic kidney), CT or MR angiography, invasive angiography
Coarctation of the aorta	Arm pulses > leg pulses, arm BP > leg BP, chest bruits, rib notching on chest radiography	MR angiography, TEE, invasive angiography
Primary aldosteronism	Hypokalemia, drug-resistant hypertension	Plasma renin and aldosterone, 24-hour urine aldosterone and potassium after oral salt loading, adrenal vein sampling
Cushing syndrome	Truncal obesity, wide and blanching purple striae, muscle weakness	1 mg dexamethasone-suppression test, urinary cortisol after dexamethasone, adrenal CT
Pheochromocytoma	Paroxysms of hypertension, palpitations, perspiration, and pallor; diabetes	Plasma metanephrines, 24-hour urinary metanephrines and catecholamines, abdominal CT or MR imaging
Obstructive sleep apnea	Loud snoring, large neck, obesity, somnolence	Polysonography

ACE = angiotensin-converting enzyme; ARB = angiotensin receptor blocker; BP = blood pressure; CT = computed tomography; GFR = glomerular filtration rate; MR, magnetic resonance; TEE = transesophageal echocardiography.

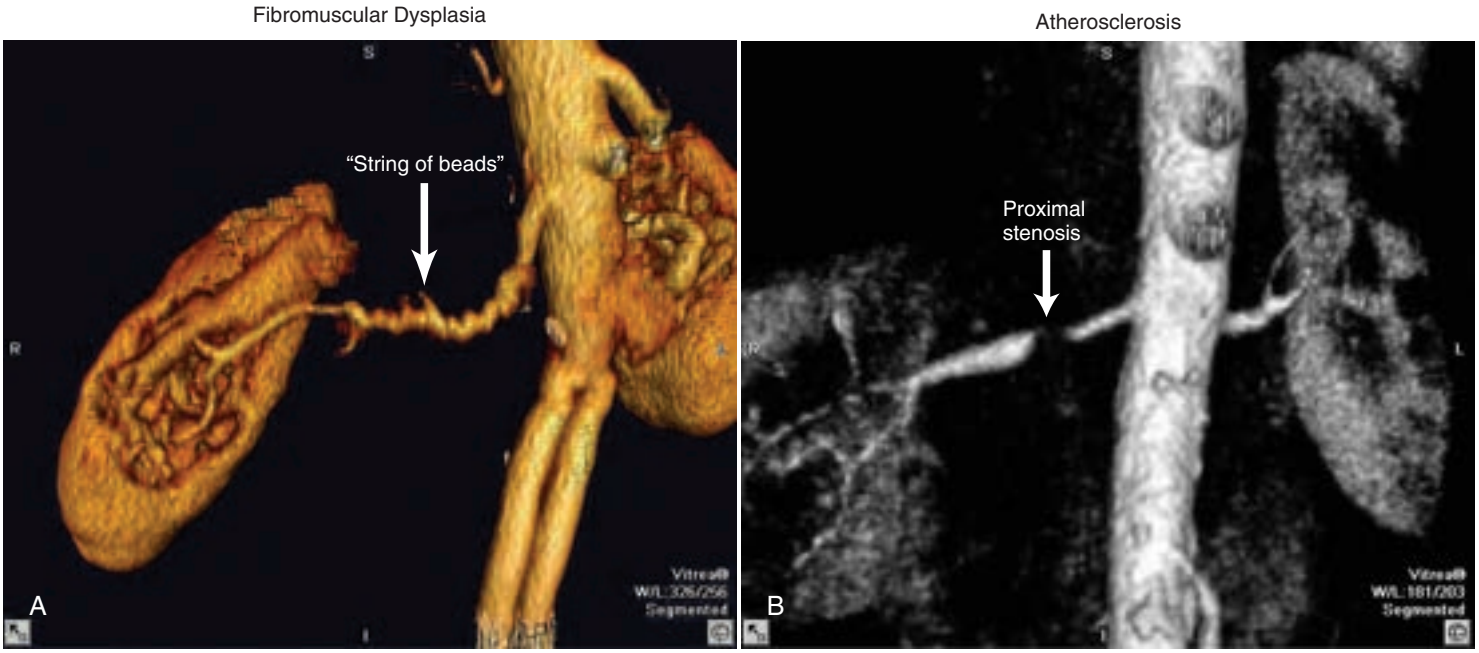

FIGURE 67-5 **Computed tomographic angiogram with three-dimensional reconstruction. A,** The classic "string of beads" lesion of fibromuscular dysplasia (bilateral in this patient). **B,** A severe proximal atherosclerotic stenosis of the right renal artery and mild stenosis of the left renal artery. (Images courtesy of Bart Domatch, MD, Radiology Department, University of Texas Southwestern Medical Center, Dallas, Texas.)

are proximal and discrete (Fig. 67-5B). Invasive renal angiography is the gold standard for confirming the diagnosis of renal artery stenosis.

TREATMENT Rx

Balloon angioplasty is the treatment of choice for fibromuscular dysplasia, with overall favorable outcomes and at least one third of patients no longer needing any antihypertensive medications.[7] In contrast, most atherosclerotic renal artery lesions do not cause hypertension or progressive renal failure, and most patients will not benefit from revascularization (balloon angioplasty or stenting), which carries substantial risks for serious complications. As a result, medical management of hypertension (with a regimen that includes a renin-angiotensin system inhibitor) and the associated atherosclerotic risk factors is first-line treatment for atherosclerotic renal artery stenosis, except that patients with truly drug-resistant hypertension, a progressive decline in renal function (ischemic nephropathy), or recurrent acute ("flash") pulmonary edema may benefit from stent-based revascularization (Chapter 125).

MINERALOCORTICOID-INDUCED HYPERTENSION DUE TO PRIMARY ALDOSTERONISM

PATHOBIOLOGY

The most common causes of primary aldosteronism (Chapter 227) are a unilateral aldosterone-producing adenoma and bilateral adrenal hyperplasia. Because aldosterone is the principal ligand for the mineralocorticoid receptor in the distal nephron, excessive aldosterone production causes excessive renal Na^+-K^+ exchange, often resulting in hypokalemia.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The diagnosis should always be suspected when hypertension is accompanied by either unprovoked hypokalemia (serum potassium concentration below 3.5 mmol/L in the absence of diuretic therapy) or a tendency to develop excessive hypokalemia during diuretic therapy (serum potassium concentration below 3.0 mmol/L). However, more than one third of patients do not have hypokalemia on initial presentation, and the diagnosis should also be considered in any patient with resistant hypertension.

Screening for hyperaldosteronism should be restricted to the small fraction of hypertensive patients with hypokalemia or severe drug-resistant hypertension. If such patients have a positive screening test—a high serum aldosterone level and a suppressed plasma renin activity level—and want to consider laparoscopic adrenalectomy, the patient should be referred to an experienced center for further evaluation: salt-loading to test for nonsuppressible aldosteronism and, if present, adrenal vein sampling to test for lateralization.

TREATMENT Rx

Both CT and MRI have too many false-positive and false-negative results to be used as a noninvasive alternative to invasive adrenal vein sampling. Laparoscopic adrenalectomy and mineralocorticoid receptor blockade with eplerenone (50 to 100 mg per day) constitute highly effective therapeutic options that target the disease-causing mechanism with a favorable risk-to-benefit ratio.

MENDELIAN FORMS OF MINERALOCORTICOID-INDUCED HYPERTENSION

Almost all the rare mendelian forms of hypertension are mineralocorticoid induced and involve excessive activation of the epithelial Na^+ channel (ENaC), the final common pathway for reabsorption of sodium from the distal nephron (E-Fig. 67-2). Thus, salt-dependent hypertension can be caused both by gain-of-function mutations of ENaC or the mineralocorticoid receptor and by increased production or decreased clearance of mineralocorticoid receptor ligands, which are aldosterone, deoxycorticosterone, and cortisol.

PHEOCHROMOCYTOMAS AND PARAGANGLIOMAS

Pheochromocytomas are rare catecholamine-producing tumors of the adrenal chromaffin cells, whereas paragangliomas are even rarer tumors of the extra-adrenal chromaffin cells (Chapter 228). The diagnosis should be suspected when hypertension is drug resistant or paroxysmal, particularly when accompanied by paroxysms of headache, palpitations, pallor, or diaphoresis. In some patients, pheochromocytoma is misdiagnosed as panic disorder. A family history of early-onset hypertension may suggest pheochromocytoma as part of the multiple endocrine neoplasia syndromes (Chapter 231). An increasing number of pheochromocytomas are being detected incidentally on abdominal imaging studies for nonadrenal indications. If the diagnosis is missed, outpouring of catecholamines from the tumor can cause unsuspected hypertensive crisis during unrelated surgical procedures, in which case mortality rates exceed 80%.

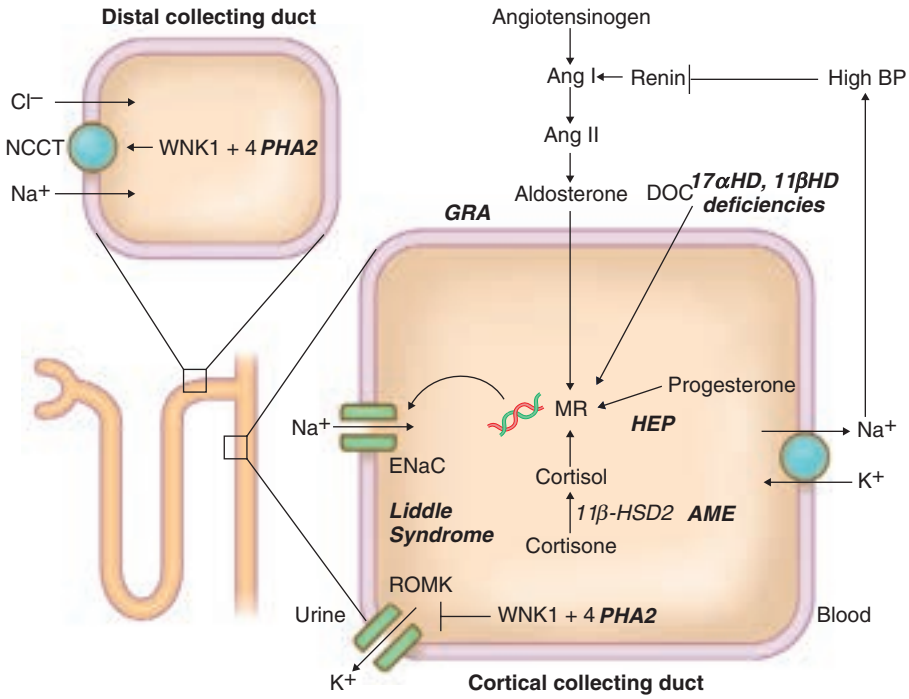

E-FIGURE 67-2. Mendelian forms of hypertension that cause mineralocorticoid-induced hypertension. AME = apparent mineralocorticoid excess; Ang = angiotensin; BP = blood pressure; GRA = glucocorticoid-remediable aldosteronism; 17αHD and 11βHD = 17α- and 11β-hydroxylase deficiency; 11β-HSD2 = 11β-hydroxysteroid dehydrogenase type 2; DOC = deoxycorticosterone; ENaC = epithelial Na^+ channel; HEP = hypertension exacerbated by pregnancy; MR = mineralocorticoid receptor; NCCT = sodium-chloride cotransporter; PHA2 = pseudohypoaldosteronism type 2; ROMK = rectifying outer medullary K^+ channel; WNK = with no lysine kinases. See text for explanation. (Modified from Lifton RP, Gharavi AG, Geller DS. Molecular mechanisms of human hypertension. *Cell.* 2001;104:545-556.)

OTHER NEUROGENIC CAUSES

Other causes of neurogenic hypertension that can be confused with pheochromocytoma include sympathomimetic agents (cocaine, methamphetamine; Chapter 34), baroreflex failure, and obstructive sleep apnea (Chapter 100). A history of surgery and radiation therapy for head and neck tumors (Chapter 190) raises suspicion of baroreceptor damage. Snoring and somnolence suggest sleep apnea, but continuous positive airway pressure to treat the sleep apnea rarely improves blood pressure substantially (Chapter 100).

OTHER CAUSES OF SECONDARY HYPERTENSION

Coarctation of the aorta typically occurs just distal to the origin of the left subclavian artery, so the blood pressure is lower in the legs than in the arms (opposite of the normal situation) (see Fig. 69-7). The clue is that the pulses are weaker in the lower than in the upper extremities, indicating the need to measure blood pressure in the legs as well as in both arms. Intercostal collaterals can produce bruits on examination and rib notching on the chest radiograph. Coarctations can be cured with surgery or angioplasty.

Hyperthyroidism tends to cause systolic hypertension with a wide pulse pressure, whereas hypothyroidism tends to cause mainly diastolic hypertension. Treatment is for the underlying disease. Hyperparathyroidism (Chapter 245) also has been associated with hypertension. Cyclosporine and tacrolimus are important causes of secondary hypertension in transplant recipients, apparently by inhibition of calcineurin, the calcium-dependent phosphatase that is expressed not only in lymphoid tissue but also in neural, vascular, and renal tissue. In the absence of outcomes data, nondihydropyridine calcium-channel blockers (CCBs) have become the drugs of first choice, but they increase cyclosporine blood levels. Combination therapy with diuretics, CCBs, and central sympatholytics often is required.

PREVENTION AND TREATMENT OF HYPERTENSION

Rx

At the population level, the primary prevention of hypertension requires large-scale societal changes, including further efforts to influence the food industry to reduce salt in processed foods, efforts to increase exercise, and the availability of fresh fruits and vegetables. After a person's blood pressure rises to hypertensive or even prehypertensive levels, lifestyle modifications alone are almost never enough to return blood pressure to normal, and recidivism is typical.

Although short-term pharmacologic therapy with a low-dose angiotensin receptor blocker (ARB) may prevent the conversion from prehypertension to full-blown hypertension, blood pressure quickly rises again if the ARB is discontinued. Thus, lifelong prescription medication is the cornerstone of effective therapy for primary hypertension, with lifestyle modification serving as a very important adjunct but not as an alternative. The objective is to reduce the blood pressure and associated metabolic abnormalities sufficiently to reduce the risk for cardiovascular events and end-stage renal disease without compromising the patient's quality of life.

Randomized trials have proved beyond any doubt that antihypertensive drug therapy reduces cardiovascular risk, with benefits that are proportional to the reduction in blood pressure achieved (E-Fig. 67-3). However, in practice, most treated patients do not achieve the same low risk levels of truly normotensive persons because their blood pressures remain higher than optimal owing to the threshold levels of guidelines, hesitation of practicing physicians to start and intensify drug treatment, costs of medications, and medication noncompliance despite the declining costs of generic medications. This residual risk also may be attributable to the cardiovascular damage sustained before starting drug therapy.

Multidrug regimens with two, three, or even more medications of different drug classes are almost always required to achieve currently recommended blood pressure goals. Low-dose drug combinations exert synergistic effects on blood pressure while minimizing dose-dependent side effects. For most patients with hypertension, lipid-lowering therapy (Chapter 206) is indicated as part of a comprehensive cardiovascular risk-reduction strategy (Chapter 52).

Lifestyle Modification

Lifestyle modification (Table 67-6) should be part of every antihypertensive regimen. However, dietary and exercise interventions are difficult to sustain long term. For example, short-term trials have proved that individuals with prehypertension or stage 1 hypertension can lower their blood pressures on average by 6/3 mm Hg even without restricting calorie or sodium intake if they adhere to a diet rich in fresh fruits and vegetables and low-fat dairy products (www.nhlbi.nih.gov/files/docs/public/heart/dash_brief.pdf). Modest dietary sodium restriction produces a further reduction in blood pressure and decreases cardiovascular disease risk.[8] Sodium reduction is particularly effective in black hypertensive patients. Most dietary sodium comes from processed foods, and patients should be taught to read food labels (6 g of NaCl = 2.4 g of sodium = 100 mmol of sodium).

TABLE 67-6 LIFESTYLE RECOMMENDATIONS TO LOWER BLOOD PRESSURE IN ADULTS WITH HYPERTENSION OR PRE-HYPERTENSION

DIET
1. Adopt a diet that is: • *High* in vegetables, nuts, fruits, grains, low-fat dairy products, fish, poultry, etc. • *Low* in sweets, sugar-sweetened beverages, and red meats Adapt this dietary pattern to calorie requirements, personal/cultural food preferences, and medical conditions such as diabetes. 2. Lower sodium intake
PHYSICAL ACTIVITY
1. Engage in three to four 40-minute sessions of moderate-to-intense aerobic physical activity per week.

Adapted from Eckel RH, Jakicic JM, Ard JD, et al. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2014;63:2960-2984.

Moderately intense aerobic exercise programs can lower blood pressure by 2 to 5/1 to 2 mm Hg. The larger reductions in blood pressure are seen immediately after a bout of aerobic exercise (Chapter 16), but smaller reductions can persist for several hours.

Relaxation and stress management techniques (e.g., meditation, biofeedback, breathing exercises) can decrease blood pressure transiently but generally produce little if any demonstrable effect on ambulatory blood pressure (Chapter 39). However, some individuals with overwhelming home or job strain or anger can benefit from cognitive behavior therapy and anxiolytics (Chapter 397).

Blood pressure increases transiently by 10 to 15 mm Hg after each cigarette, so smokers of more than 20 cigarettes per day often have higher blood pressures out of the office than in the smoke-free medical office. Smokers should be counseled to quit tobacco (Chapter 32) not only because it raises blood pressure but also because it is such a potent risk factor for coronary heart disease, stroke, and the progression of hypertensive kidney disease.

Blood pressure increases by up to 10 to 15 mm Hg with the first morning cup of coffee, but the pressor response to caffeine often habituates throughout the day. Thus, caffeine consumption need not be totally eliminated but may need to be reduced.

Moderate alcohol (Chapter 33) consumption (one or two drinks per day) does not seem to increase the risk for hypertension in Western populations; but in Japanese populations, hypertension is more common in men who are moderate drinkers than in men who cannot drink because of a loss-of-function mutation in the alcohol dehydrogenase gene. In all populations, heavy drinking (three or more standard-sized drinks per day) and especially binge drinking activate the sympathetic nervous system the next day during withdrawal and are associated with an increased incidence and severity of hypertension, which is reversible if alcohol consumption decreases.

Antihypertensive Drugs

Although every hypertensive patient should adopt sensible lifestyle modifications, almost all will require medication to optimize outcomes. Lowering blood pressure with medication reduces but does not eliminate the risks for cardiovascular events, renal failure, and death.

Classes of Oral Antihypertensive Drugs

Multiple classes of oral antihypertensive drugs are approved by the U.S. Food and Drug Administration, although all have specific contraindications (Tables 67-7 and 67-8).

First-Line Drugs for Hypertension

Multiple practice guidelines[1-4] recommend initiating drug treatment with one or more of three classes of first-line drugs (Fig. 67-6), which have additive or synergistic effects when used in combination: (1) CCBs, (2) renin-angiotensin system blockers—either ACE inhibitors or ARBs, and (3) thiazide-like diuretics.

Calcium-Channel Blockers

Mechanism of Action. The CCBs block the opening of voltage-gated (L-type) Ca^{2+} channels in cardiac myocytes and vascular smooth muscle cells. The resultant decrease in the cytosolic Ca^{2+} signal decreases heart rate and ventricular contractility and relaxes vascular smooth muscle. Blood pressure lowering is related mainly to peripheral arterial vasodilation, with the rank

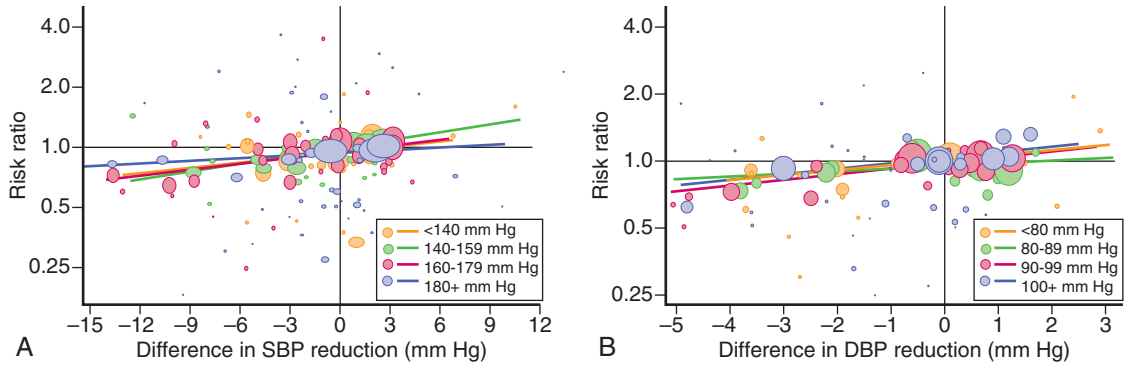

E-FIGURE 67-3. **Comparison of the associations between blood pressure change and the risk ratio reduction in total major cardiovascular events according to categories of (A) systolic blood pressure (SBP) and (B) diastolic blood pressure (DBP) of 201,566 participants in 32 randomized controlled trials. The area of each circle is proportional to inverse variance of log odds ratio. Fitted line represents summary meta-regression for total major cardiovascular events. Each regression line is for a different level of baseline blood pressure. (From Czernichow S, Zanchetti A, Turnbull F, et al. The effects of blood pressure reduction and of different blood pressure-lowering regimens on major cardiovascular events according to baseline blood pressure: meta-analysis of randomized trials. *J Hypertens.* 2011;29:4-16.)**

TABLE 67-7 SELECTED ORAL ANTIHYPERTENSIVE AGENTS

DRUG	DOSE RANGE, TOTAL, MG/DAY (DOSES PER DAY)	USUAL STARTING DOSE, MG/DAY (DOSES PER DAY)
DIURETICS		
Thiazide and Thiazide-Like Diuretics		
Chlorthalidone	6.25-50 (1)	6.25 (1)
HCTZ	6.25-50 (1)	12.5 (1)
Indapamide	1.25-5 (1)	1.25 (1)
Metolazone	2.5-5 (1)	2.5 (1)
Loop Diuretics		
Bumetanide	0.5-2 (2)	1 (2)
Ethacrynic acid	25-100 (2)	25 (2)
Furosemide	20-160 (2)	20 (2)
Torsemide	2.5-20 (1-2)	5 (2)
Potassium Sparing		
Amiloride	5-20 (1)	10 (2)
Eplerenone	25-100 (1-2)	25 (1)
Spironolactone	6.25-400 (1-2)	12.5 (1)
Triamterene	25-100 (1)	37.5 (1)
β-BLOCKERS		
Acebutolol	200-800 (2)	200 (2)
Atenolol	25-100 (1)	25 (1)
Betaxolol	5-20 (1)	5 (1)
Bisoprolol	2.5-20 (1)	2.5 (1)
Carteolol	2.5-10 (1)	2.5 (1)
Metoprolol	50-450 (2)	50 (2)
Metoprolol XL	50-200 (1-2)	50 (1)
Nadolol	20-320 (1)	40 (1)
Penbutolol	10-80 (1)	10 (1)
Pindolol	10-60 (2)	10 (1)
Propranolol	40-180 (2)	40 (2)
Propranolol LA	60-180 (1-2)	60 (1)
Timolol	20-60 (2)	20 (2)
VASODILATING β-BLOCKERS		
Carvedilol	6.25-50 (2)	6.25 (2)
Carvedilol CR	10-80 (1)	20 (1)
Labetalol	100-2400 (2)	200 (2)
Nebivolol	2.5-40 (1)	5 (1)
CALCIUM-CHANNEL BLOCKERS		
Dihydropyridines		
Amlodipine	2.5-10 (1)	2.5 (1)
Felodipine	2.5-20 (1-2)	2.5 (2)
Isradipine CR	2.5-20 (2)	2.5 (2)
Nicardipine SR	30-120 (2)	30 (2)
Nifedipine XL	30-120 (1)	30 (1)
Nisoldipine	10-40 (1-2)	10 (2)
Nondihydropyridines		
Diltiazem CD	120-540 (1-2)	180 (1)
Verapamil HS	120-480 (1)	180 (1)
ANGIOTENSIN-CONVERTING ENZYME INHIBITORS		
Benazepril	10-80 (1-2)	20 (1)
Captopril	25-150 (2)	25 (2)
Enalapril	2.5-40 (2)	5 (2)
Fosinopril	10-80 (1-2)	20 (2)
Lisinopril	5-80 (1-2)	40 (2)
Moexipril	7.5-30 (1)	7.5 (1)
Perindopril	4-16 (1)	4 (1)
Quinapril	5-80 (1-2)	40 (2)
Ramipril	2.5-20 (1)	2.5 (1)
Trandolapril	1-8 (1)	2 (1)
ANGIOTENSIN RECEPTOR BLOCKERS		
Candesartan	8-32 (1)	8 (1)
Eprosartan	400-800 (1-2)	400 (1)
Irbesartan	150-300 (1)	150 (1)
Losartan	25-100 (2)	50 (1)
Olmesartan	5-40 (1)	20 (1)
Telmisartan	20-80 (1)	40 (1)
Valsartan	80-320 (1-2)	160 (2)
DIRECT RENIN INHIBITOR		
Aliskiren	150-300 (1)	150 (1)
α-BLOCKERS		
Doxazosin	1-16 (1)	1 (1)
Prazosin	1-40 (2-3)	1 (2)
Terazosin	1-20 (1)	1 (1)
Phenoxybenzamine for pheochromocytoma	20-120 (2)	20 (2)
CENTRAL SYMPATHOLYTICS		
Clonidine	0.3-1.2 (3)	0.3 (3)
Clonidine patch	0.1-0.6 (weekly)	0.1 (weekly)
Guanabenz	2-32 (2)	2 (2)
Guanfacine	1-3 (1) (qhs)	1 (1)
Methyldopa	250-1000 (2)	250 (2)
Reserpine	0.05-0.25 (1)	0.05 (1)
DIRECT VASODILATORS		
Hydralazine	10-200 (2)	20 (2)
Minoxidil	2.5-100 (1)	2.5 (1)
FIXED-DOSE COMBINATIONS		
Aliskiren/HCTZ	150/12.5-300/25 (1)	150/12.5 (1)
Amiloride/HCTZ	5/50 (1)	5/50 (1)
Amlodipine/benazepril	2.5-5/10-20 (1)	2.5/10 (1)
Amlodipine/olmesartan	5-10/20-40 (1)	5/20 (1)
Amlodipine/telmisartan	5/20-10/80 (1)	5/20 (1)
Amlodipine/valsartan	5/160-10/320 (1)	5/160 (1)
Atenolol/chlorthalidone	50-100/25 (1)	50/25 (1)
Benazepril/HCTZ	5-20/6.25-25 (1)	20/6.25 (1)
Bisoprolol/HCTZ	2.5-10/6.25 (1)	2.5/6.25 (1)
Candesartan/HCTZ	16-32/12.5-25 (1)	16/12.5 (1)
Enalapril/HCTZ	5-10/25 (1-2)	5/25 (1)
Eprosartan/HCTZ	600/12.5-25 (1)	600/12.5 (1)
Fosinopril/HCTZ	10-20/12.5 (1)	10/12.5 (1)
Irbesartan/HCTZ	150-300/12.5-25 (1)	150/12.5 (1)
Losartan/HCTZ	50-100/12.5-25 (1)	50/12.5 (1)
Olmesartan/HCTZ	20-40/12.5 (1)	20/12.5 (1)
Spironolactone/HCTZ	25/25 (½-1)	25/25 (1/2)
Telmisartan/HCTZ	40-80/12.5-25 (1)	40/12.5 (1)
Trandolapril/verapamil	2-4/180-240 (1)	2/180 (1)
Triamterene/HCTZ	37.5/25 (½-1)	37.5/25 (1/2)
Valsartan/HCTZ	80-160/12.5-25 (1)	160/12.5 (1)
Valsartan/Amlodipine/HCTZ	80-160/5-10/12.5-25 (1)	160/5/12.5 (1)

HCTZ = hydrochlorothiazide.

TABLE 67-8 MAJOR CONTRAINDICATIONS AND SIDE EFFECTS OF ANTIHYPERTENSIVE DRUGS

DRUG CLASS	MAJOR CONTRAINDICATIONS	SIDE EFFECTS
Diuretics		
Thiazides	Gout	Insulin resistance, new-onset type 2 diabetes Hypokalemia, hyponatremia Hypertriglyceridemia Hyperuricemia, precipitation of gout Erectile dysfunction (more than other drug classes) Potentiate nondepolarizing muscle relaxants Photosensitivity dermatitis
Loop diuretics	Hepatic coma	Interstitial nephritis Hypokalemia Potentiate succinylcholine Potentiate aminoglycoside ototoxicity
Potassium-sparing diuretics	Serum potassium concentration > 5.5 mEq/L GFR < 30 mg/mL/1.73 m^2	Hyperkalemia
ACE inhibitors	Pregnancy Bilateral renal artery stenosis Hyperkalemia	Cough Hyperkalemia Angioedema Leukopenia Fetal toxicity Cholestatic jaundice (rare fulminant hepatic necrosis if the drug is not discontinued)
Dihydropyridine CCBs	As monotherapy in chronic kidney disease with proteinuria	Headaches Flushing Ankle edema Heart failure Gingival hyperplasia Esophageal reflux
Nondihydropyridine CCBs	Heart block Systolic heart failure	Bradycardia, AV block (especially with verapamil) Constipation (often severe with verapamil) Worsening of systolic function, heart failure Gingival edema or hypertrophy Increase cyclosporine blood levels Esophageal reflux
ARBs, DRI	Pregnancy Bilateral renal artery stenosis Hyperkalemia	Hyperkalemia Angioedema (very rare) Fetal toxicity
β-Adrenergic blockers	Heart block Asthma Depression Cocaine and methamphetamine abuse	New-onset type 2 diabetes (especially in combination with a thiazide) Heart block, acute decompensated heart failure Bronchospasm Depression, nightmares, fatigue Cold extremities, claudication (β_2 effect) Stevens-Johnson syndrome Agranulocytosis
α-Adrenergic blockers	Orthostatic hypotension Systolic heart failure Left ventricular dysfunction	Orthostatic hypotension Drug tolerance (in the absence of diuretic therapy) Ankle edema Heart failure First-dose effect (acute hypotension) Potentiate hypotension with PDE-5 inhibitors (e.g., sildenafil)
Central sympatholytics	Orthostatic hypotension	Depression, dry mouth, lethargy Erectile dysfunction (dose dependent) Rebound hypertension with clonidine withdrawal Coombs test–positive hemolytic anemia and elevated liver enzymes with α-methyldopa
Direct vasodilators	Orthostatic hypotension	Reflex tachycardia Fluid retention Hirsutism, pericardial effusion with minoxidil Lupus with hydralazine

ACE = angiotensin-converting enzyme; ARBs = angiotensin receptor blockers; AV = atrioventricular; CCBs = calcium-channel blockers; DRI = direct renin inhibitor; GFR = glomerular filtration rate; NSAIDs = nonsteroidal anti-inflammatory drugs; PDE-5 = phosphodiesterase-5.

order of potency being dihydropyridines > diltiazem ≫ verapamil. In contrast, for negative chronotropic and inotropic effects, the rank order of potency is verapamil ≫ diltiazem > dihydropyridines.

Therapeutic Principles. The most recommended CCBs are amlodipine followed by diltiazem. Amlodipine's long half-life permits once-daily dosing, and its costs are low since it became generic. Amlodipine is equivalent to a potent diuretic or lisinopril in protecting against nonfatal coronary events, stroke, and death, but it provides less protection against heart failure.[1] Unlike diuretics, ARBs, and ACE inhibitors, a high-salt diet or concurrent nonsteroidal anti-inflammatory drug (NSAID) therapy does not compromise the effectiveness of dihydropyridine CCBs. The CCBs have some diuretic action because they dilate the afferent renal arteriole and may reduce requirements for additional diuretic therapy in mild hypertension. Amlodipine and other dihydropyridine CCBs are less renoprotective than ACE inhibitors or ARBs in patients with proteinuric chronic kidney disease. Such patients should not receive amlodipine as first-line therapy, but it may be useful as adjunctive therapy after initiation of appropriate first-line therapy with either an ACE inhibitor or ARB, as well as a diuretic.

Diltiazem is a usually well-tolerated alternative in patients who cannot tolerate amlodipine or would benefit from its other effects. Verapamil is not recommended because it is a weak antihypertensive medication and causes constipation.

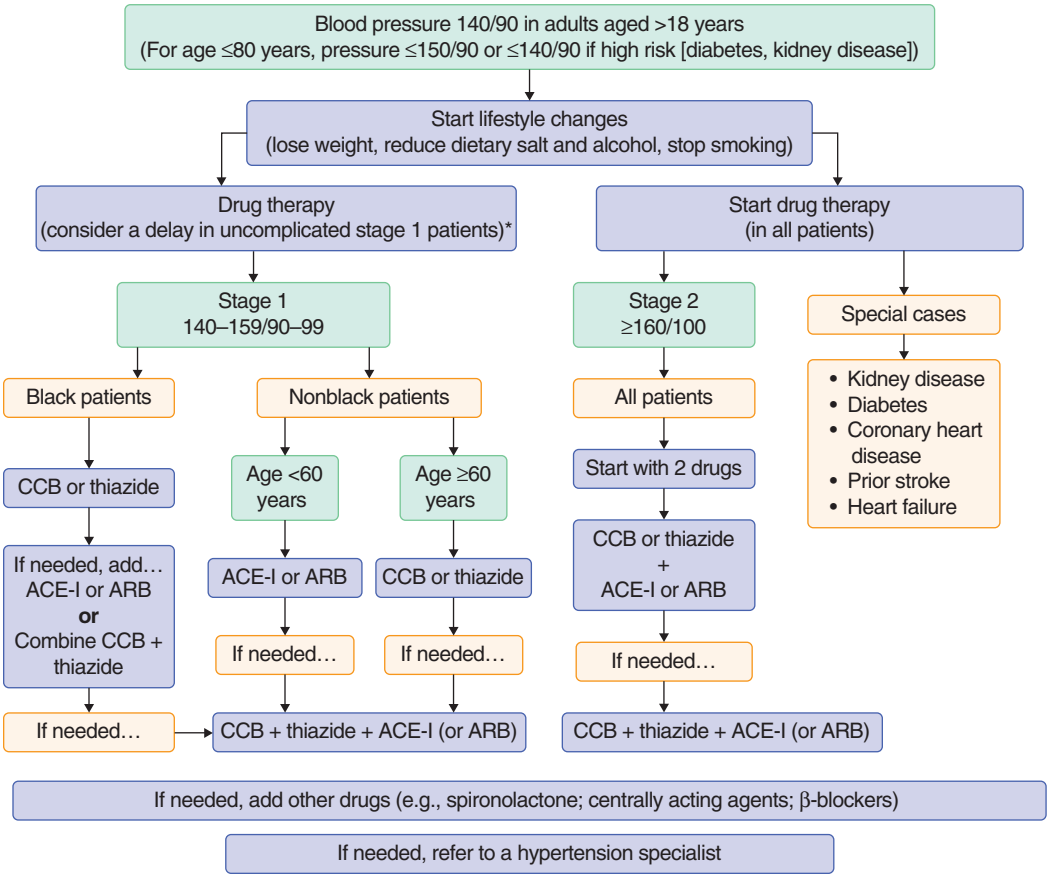

FIGURE 67-6. 2014 Hypertension Management Algorithm Recommended by the American Society of Hypertension and the International Society of Hypertension. At any stage, it is entirely appropriate to seek help from a hypertension expert if treatment is proving difficult. *In patients with stage 1 hypertension and no history of a prior cardiovascular, stroke, or renal event, no evidence of end organ damage, and no diabetes or other major risk factors, drug therapy can be delayed for a short trial of lifestyle modification; however, most patients will require medication to achieve recommended blood pressure targets. In all other patients (including those with stage 2 hypertension), drug therapy should be started as soon as the diagnosis of hypertension is made. ACE-I = angiotensin-converting enzyme inhibitors; ARB = angiotensin receptor blocker; CCB = calcium channel blocker; thiazide = thiazide or thiazide-like diuretics. Blood pressure values are in mm Hg. (From Weber MA, Schiffrin EL, White WB, et al. Clinical practice guidelines for the management of hypertension in the community: a statement by the American Society of Hypertension and the International Society of Hypertension. *J Clin Hypertens.* 2014;16:14-26.)

Side Effects. *Short-acting dihydropyridines should not be used to treat hypertension.* By triggering an abrupt fall in blood pressure with reflex sympathetic activation, these rapidly acting arterial vasodilators can precipitate myocardial ischemia/infarction and death. The principal side effect of long-acting dihydropyridines is dose-dependent ankle edema, which is far more common with 10 mg of amlodipine than with lower doses. Because this vasogenic edema is caused by selective arterial dilation, it may be improved by adding an ACE inhibitor or ARB that produces balanced arterial and venous dilation. The long-acting dihydropyridines rarely cause flushing and headache. All CCBs can cause gingival hyperplasia, a rare side effect that is reversible if detected early but can otherwise lead to several dental problems. Dihydropyridines relax the smooth muscle of the distal esophagus and can exacerbate gastroesophageal reflux disease. Diltiazem can impair cardiac conduction, especially in older patients also receiving digoxin, β-blockers, or central sympatholytic agents.

Renin-Angiotensin System Inhibitors: ACE Inhibitors, ARBs, and Direct Renin Inhibitors

Mechanisms of Action. ACE inhibitors block conversion of the inactive precursor angiotensin I (AT1) to angiotensin 2 (AT2). ARBs block the action of AT2 on the type 1 angiotensin receptor. The direct renin inhibitor aliskiren blocks the conversion of pro-renin to renin, thereby blocking renin-angiotensin system activation at its origin. High levels of circulating pro-renin may stimulate AT1 receptor-independent signaling pathways that are both potentially beneficial and potentially harmful.

Clinical Use. ACE inhibitors are easy to use and have a rather flat dose-response curve. Lisinopril monotherapy is equivalent to amlodipine or chlorthalidone monotherapy except for producing a smaller reduction in blood pressure and thus less stroke protection for black hypertensive patients and older patients with low-renin hypertension. Even in these patients, however, ACE inhibitors are quite effective when combined with a diuretic or CCB. The ARBs confer the same benefits as ACE inhibitors in treating hypertension, while avoiding the ACE inhibitor cough (see later). There are no randomized trials of aliskiren monotherapy.

ACE inhibitors and ARBs are standard first-line therapy for patients with chronic kidney disease, especially if they have proteinuria. ACE inhibitors and ARBs have comparable effects on renal function, but ARBs produce more regression of left ventricular hypertrophy than do other antihypertensives.

Side Effects. All renin-angiotensin system inhibitors are contraindicated in pregnancy because they cause fetal renal agenesis and other birth defects. ACE inhibitors block the degradation of bradykinin, which activates nociceptive sensory fibers in the lungs that trigger a dry cough, which is more common in black patients and even more common in Asian patients. Bradykinin also may underlie the much less common ACE inhibitor-induced angioedema. If a cough develops in a patient who is on an ACE inhibitor and who needs renin-angiotensin system blockade, an ARB should be substituted.

The ACE inhibitors and ARBs can provoke hyperkalemia in the setting of chronic kidney disease or diabetes with type 4 renal tubular acidosis (Chapter 118). Serum potassium and creatinine levels must be monitored closely in all patients. In patients with stage 3 chronic kidney disease with proteinuria, initiation of ACE inhibitor or ARB therapy often causes a small transient increase in the serum creatinine level, but these drugs should be decreased in dose or temporarily discontinued only if the elevation is more than 30%.

Diuretics

Mechanism of Action. With initiation of diuretic therapy, contraction of blood volume causes the initial fall in blood pressure. With continued therapy, blood volume is partially restored, and vasodilator mechanisms (e.g., opening of adenosine triphosphate [ATP]-sensitive K^+ channels) sustain the antihypertensive action. Loop diuretics block Na^+-K^+-$2Cl^-$ transport in the thick ascending loop of Henle, where a large portion of the filtered sodium is reabsorbed. Thiazide diuretics and the indoline derivative indapamide block Na^+-Cl^- cotransporter in the distal convoluted tubule, where a smaller portion of the filtered sodium is reabsorbed.

Therapeutic Principles. Diuretics are still among the most effective antihypertensive medications. The thiazide-type diuretic chlorthalidone is at least as effective as (and in some circumstances more effective than) lisinopril or amlodipine in lowering blood pressure and preventing the attendant cardiovascular complications in all subgroups of patients. Combined with other

classes of antihypertensive medications, diuretics exert a synergistic effect on blood pressure.

Despite the long-standing popularity of hydrochlorothiazide in clinical practice, chlorthalidone has a much longer duration of action and appears to be more efficacious in lowering blood pressure. A 25-mg dose of chlorthalidone is roughly equivalent in potency to a 50-mg dose of hydrochlorothiazide.

Loop diuretics are less effective blood pressure–lowering agents and should be reserved for treating hypertension in the setting of advanced chronic kidney disease (stage 3 or higher). Chlorthalidone also may be effective in stage 3 chronic kidney disease.

Side Effects. Thiazides and thiazide-like diuretics can aggravate glucose intolerance (particularly in higher doses and when used in combination with a β-blocker), cause hypokalemia and hypomagnesemia, precipitate gout, and elevate serum lipid levels, especially triglyceride levels. They sometimes cause photosensitive dermatitis and are more likely than any other antihypertensive drugs to cause erectile dysfunction. Though less well recognized than thiazide-induced hypokalemia, thiazide-induced hyponatremia (Chapter 116) is a common reason that some elderly hypertensive patients simply cannot tolerate even low-dose thiazides. In hypertensive patients with chronic kidney disease, high doses of loop diuretics may precipitate acute renal decompensation, especially if combined with a high-dose ACE inhibitor or ARB; the medications can be restarted carefully at low doses after the GFR has returned to baseline.

Add-on Drug for Difficult Hypertension

Mineralocorticoid Receptor Antagonists and ENaC Antagonists ("Potassium-Sparing Diuretics")

Mechanism of Action. Eplerenone and spironolactone prevent circulating aldosterone and other mineralocorticoids from activating the mineralocorticoid receptor in the distal nephron, thereby inhibiting the downstream activation of ENaC. By comparison, triamterene and amiloride block ENaC directly. Because less sodium is presented to the Na^+, K^+-ATPase on the vascular side of the collecting duct cells, less potassium is excreted in the urine than with thiazide or loop diuretics.

Therapeutic Principles. Eplerenone (25-100 mg daily) or low-dose spironolactone (12.5-100 mg daily) is widely recommended as a highly effective add-on drug for difficult cases of hypertension. Eplerenone is a much more specific antagonist that avoids the infrequent sexual side effects of low-dose spironolactone (painful gynecomastia, erectile dysfunction, nonmenstrual uterine bleeding). Hyperkalemia must be avoided when using these agents in patients with kidney disease.

β-Adrenergic Blockers

Vasodilating β-blockers (labetalol, carvedilol, and nebivolol) also are highly effective add-on drugs for difficult hypertension. Standard β-blockers (e.g., metoprolol, atenolol) are not.

Mechanism of Action. With the initiation of standard β-blocker therapy, blood pressure changes little at first because a compensatory increase in peripheral resistance offsets the fall in cardiac output. Over time, blood pressure falls progressively as the peripheral vasculature relaxes. Thus, the antihypertensive effect of β-blockade involves decreases in cardiac output (β_1-receptors), renin release (β_1-receptors), and norepinephrine release (prejunctional β_2-receptors). The prototype β-blocker propranolol nonselectively blocks both β_1- and β_2-receptors. Other standard β-blockers (metoprolol, atenolol, acebutolol, and bisoprolol) are relatively cardioselective. In low doses, they exert a greater inhibitory effect on β_1- than on β_2-receptors, but selectivity is lost at high doses. Vasodilating β-blockers such as labetalol or carvedilol also block α-adrenergic receptors, whereas nebivolol stimulates endogenous production of nitric oxide.

Therapeutic Principles and Side Effects. Although standard β-blockers are first-line medical therapy for ischemic heart disease (Chapters 72 and 73) and heart failure (Chapter 59), they are no longer first-line or even second-line agents for uncomplicated hypertension. They predispose to diabetes, particularly when combined with a thiazide, and offer less stroke protection than other antihypertensive drugs; they provide modest protection against cardiovascular events but do not reduce all-cause mortality. In addition to being rather weak antihypertensives, they are less effective than other agents in lowering central aortic blood pressure because bradycardia allows more time for wave reflection and thus central pressure augmentation.

By contrast, vasodilating β-blockers are much more potent antihypertensive agents and do not adversely affect glucose tolerance, but they have not been studied in large randomized trials. Labetalol is effective treatment for hypertensive urgencies but is too short acting to be recommended for chronic hypertension management. Common side effects such as fatigability cause high discontinuation rates for all β-blockers. β-Blockers also can impair cardiac conduction, precipitate acute bronchospasm, and promote weight gain.

α-Adrenergic Blockers

Mechanism of Action. By blocking the interaction of norepinephrine on vascular α-adrenergic receptors, these drugs cause peripheral vasodilation, thereby lowering blood pressure. By increasing skeletal muscle blood flow, they increase insulin sensitivity. By dilating urethral smooth muscle, they improve symptoms of prostatism. Prazosin, doxazosin, terazosin, and intravenous phentolamine selectively block α_1-adrenoreceptors; phenoxybenzamine blocks both α_1- and α_2-receptors.

Therapeutic Principles and Side Effects. Phenoxybenzamine is the drug of choice for preoperative management of pheochromocytoma (Chapter 228); after α-blockade is achieved, a β-blocker should be added to block an otherwise excessive reflex tachycardia. The selective α_1-blockers are not first-line agents and should not be used as monotherapy because their propensity to cause fluid retention can lead to tachyphylaxis and unmask or exacerbate heart failure. When used in a combination regimen that includes a diuretic, however, they are effective add-on therapy for difficult hypertension and are particularly useful in older men with prostatism. Although marketed specifically for prostatism and not as an antihypertensive agent, the selective α_{1A}-blocker tamsulosin lowers blood pressure in some men.

Central Sympatholytics

Mechanism of Action. Stimulation of postsynaptic α_2-adrenergic receptors and imidazoline receptors in the central nervous system lowers central sympathetic outflow, while stimulation of presynaptic α_2-receptors causes feedback inhibition of norepinephrine release from peripheral sympathetic nerve terminals. These combined actions reduce adrenergic drive to the heart and peripheral circulation.

Therapeutic Principles and Side Effects. The central sympatholytics are best reserved for short-term oral treatment of hypertensive urgency. They are potent antihypertensive agents that may be needed as add-on therapy for very difficult hypertension, but their troublesome central nervous system side effects reduce quality of life. To avoid rebound hypertension between doses, short-acting clonidine must be given every 6 to 8 hours or, whenever possible, discontinued using a gradual tapering schedule. Rebound hypertension is less of a problem with longer acting preparations (guanfacine, clonidine patch). α-Methyldopa remains a useful drug for management of hypertension in pregnancy (Chapter 239) but is no longer first-line therapy.

Direct Vasodilators

Mechanism of Action. Minoxidil and hydralazine are potent hyperpolarizing arterial vasodilators that work by opening vascular ATP-sensitive K^+ channels.

Therapeutic Principles and Side Effects. By causing selective and rapid arterial dilation, both drugs cause profound reflex sympathetic activation and tachycardia. Hydralazine is useful for the treatment of preeclampsia (Chapter 239). A combination of hydralazine plus nitrates is useful for the treatment of heart failure specifically in African American patients, in whom hypertensive heart disease causes heart failure most commonly (Chapter 59) Severe hypertension accompanying advanced chronic kidney disease is the main indication for minoxidil, which must be combined with aβ-blocker to prevent excessive reflex tachycardia and with a loop diuretic to prevent excessive fluid retention. Institution of hemodialysis is usually a more effective means of controlling hypertension in this setting.

Antihypertensive Drug Interactions

By inhibiting the kidney's ability to excrete sodium, NSAIDs can negate the antihypertensive action of diuretics and renin-angiotensin system inhibitors, but they do not interfere with CCBs. The risk for increased blood pressure and associated cardiovascular events is lowest with acetaminophen and low-dose aspirin (81 mg daily), intermediate with non-COX-2 selective NSAIDs including high-dose aspirin, and highest with COX-2-selective NSAIDs. Even a single glass of grapefruit juice increases the bioavailability of dihydropyridine CCBs by inhibiting the intestinal cytochrome P-450 3A4 (CYP34A) system, which is responsible for the first-pass metabolism of many oral medications. Diltiazem and verapamil, which are potent CYP3A4 inhibitors, should be avoided in patients undergoing anti-VEGF cancer chemotherapy with either sunitinib or sorafenib and should be used with caution in patients receiving digoxin or immunosuppressive therapy with either cyclosporine or tacrolimus because blood levels will increase and must be monitored. Carvedilol should be taken after meals to optimize its absorption, whereas a high-fat meal will impair the absorption of aliskiren.

Which Blood Pressure Goals and Which Drugs for Which Patients?

Despite the large evidence base for the medical treatment of hypertension, important gaps remain on (1) how far to lower blood pressure and (2) with which drugs for which patients. Current guidelines[1-4] recommend blood pressure goals of less than 150/90 mm Hg for "elderly" patients and less than 140/90 mm Hg for essentially all other patients. However, some experts recommend the goal of less than 140/90 mm Hg in patients older than 60 years if they are not frail (Chapter 25) and are able to tolerate such treatment without side effects.[9] The new guidelines place more emphasis on combination therapy with any two or all three of the first-line drug classes (CCBs, renin-angiotensin system blockers, thiazide diuretics; Tables 67-9 and 67-10). However, the preferred antihypertensive drug classes differ for specific types of patients (Table 67-11).

TABLE 67-9 BLOOD PRESSURE TREATMENT RECOMMENDATIONS SHAPED BY ALL OF OUR CURRENT MAJOR GUIDELINES

	2011-2014 GUIDELINES
Blood pressure treatment goals	<150/90 mm Hg for "elderly" patients <140/90 mm Hg for "nonelderly" patients and patients with diabetes or chronic kidney disease
Preferred first-line treatment	Three choices: calcium-channel blocker, ACE inhibitor or ARB, or thiazide diuretic (chlorthalidone preferred)
Combination treatment	Good option for stage 1 hypertension ACE inhibitor + calcium-channel blocker preferred over an ACE inhibitor + thiazide-type diuretic

Adapted from James PA, Oparil S, Carter BL, et al. 2014 Evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014;311:507-520; Weber MA, Schiffrin EL, White WB, et al. Clinical practice guidelines for the management of hypertension in the community: a statement by the American Society of Hypertension and the International Society of Hypertension. *J Hypertens*. 2014;32:3-15; Mancia G, Fagard R, Narkiewicz K, et al. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J Hypertens*. 2013;31:1281-1357; National Institute for Health and Clinical Excellence (NICE). Clinical guideline 127. Hypertension: clinical management of primary hypertension in adults. 2011. www.nice.org.uk/guidance/cg127.
ACE = angiotensin-converting enzyme; ARB = angiotensin receptor blocker.

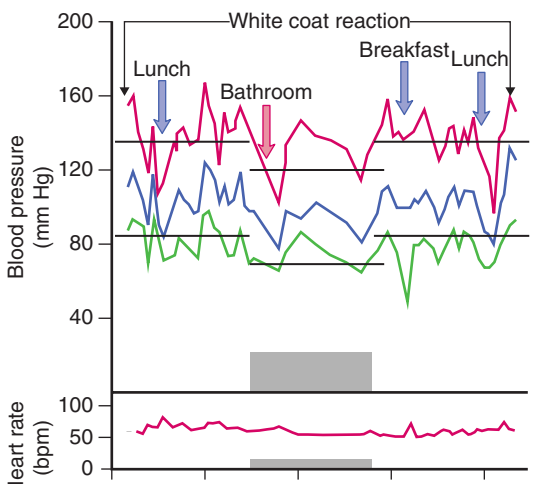

FIGURE 67-7. The 24-hour blood pressure monitor tracing of a patient with postprandial and orthostatic hypotension. This frail 70-year-old woman was referred for evaluation of labile hypertension and dizziness. *Blue arrows* show repeated episodes of postprandial hypotension. The *red arrow* shows an episode of orthostatic hypotension when the patient walked to the bathroom 90 minutes after going to sleep. White coat reactions are also seen when the patient came to clinic both to have the monitor placed and then to have it removed. (Provided by Ronald G. Victor, MD, Hypertension Center, Cedars-Sinai Heart Institute, Los Angeles, California.)

Hypertensive Patients in General

Current recommendations are based on hypertension trials, in which the active treatment group achieved a final mean systolic blood pressure below 140 mm Hg but never below 130 mm Hg. By comparison, some data indicate benefits of lowering blood pressure to less than 140/90 mm Hg for older as well as younger patients (see E-Fig. 67-3), and meta-regression analysis suggests—but does not prove—that additional lowering of blood pressure will provide additional cardiovascular protection even when the baseline blood pressure is 140/90 mm Hg or lower.

In randomized trials, differences in systolic blood pressure reduction rather than drug class explain the benefits of treatment with three caveats. First, β-blockers provide less stroke protection and CCBs more stroke protection than other drugs. Second, the combination of an ACE inhibitor and CCB may be an excellent initial option because it prevents more cardiovascular events than either the combination of a β-blocker with a thiazide diuretic or an ACE inhibitor with a thiazide diuretic (E-Fig. 67-4). CCBs also are better tolerated than diuretics and avoid their metabolic side effects. Third, dual renin-angiotensin system blockade with both an ACE inhibitor and an ARB has no advantage over monotherapy with either drug but results in much more symptomatic hypotension and more renal impairment (E-Fig. 67-5).

Systolic Hypertension in Elderly Patients

Most hypertensive patients are now older than 65 years, and most have isolated systolic hypertension. Placebo-controlled trials provide unequivocal proof that any blood pressure–lowering regimen reduces coronary events, strokes, heart failure events, and deaths in elderly hypertensive patients, even patients older than 80 years.[1] Although antihypertensive drugs are just as effective in preventing cardiovascular events in older patients, the intensity of their blood pressure reduction must be weighed against increased risks for hypotension, which can precipitate falls and ischemic cardiac events. The lowest mean systolic blood pressure reached in these trials in elderly patients is 145 mm Hg, so evidence supports an office systolic blood pressure treatment target below 150 mm Hg. Whether there are additional benefits to reducing office systolic blood pressure in active otherwise healthy elderly patients to below 140 mm Hg remains uncertain at this time.

Ambulatory monitoring is key for detecting postprandial hypotension and orthostatic hypotension, which are common in hypertensive elderly patients (Fig. 67-7). Although the management of postprandial hypotension is challenging, useful strategies include frequent small low-carbohydrate meals, caffeine with meals, and liberalized salt intake. If these nondrug strategies prove insufficient, fludrocortisone (0.1 to 0.2 mg daily) can be added; it often worsens supine hypertension, which can be managed with elevation of the head of the bed (with 6-inch cinder blocks producing a 30-degree head-up tilt) and a low-dose short-acting ARB (e.g., losartan 25-50 mg) at bedtime.[10]

Although most elderly patients will require combination therapy with two or three drugs to manage their hypertension, it is important to titrate medications more slowly in elderly patients and to check frequently for orthostatic hypotension and adverse drug reactions, such as thiazide-induced hyponatremia. Nonadherence and potential drug-drug interactions are key concerns. Therapy should be simplified and individualized, based more on the patient's overall health or frailty than on chronologic age. For example, a seated home blood pressure of 155 mm Hg may be an appropriate treatment target for a frail 70-year-old patient with marked orthostatic and postprandial hypotension, whereas a home seated blood pressure of 130 mm Hg may be an appropriate treatment target for a vigorous healthy 85-year-old patient whose chief concern is to avoid a disabling stroke.

Hypertension with Left Ventricular Hypertrophy

More than one third of hypertensive patients have electrocardiographic left ventricular hypertrophy by the time of diagnosis, a finding that places them at increased risk for hypertensive complications including heart failure, stroke, and atrial fibrillation. Meta-analyses consistently show the superiority of ARBs for the regression of left ventricular hypertrophy. In patients with stage 2 hypertension and left ventricular hypertrophy on their ECG, an ARB-based regimen is more effective than a β-blocker-based regimen for reducing cardiovascular events, especially stroke.

Hypertension in Patients with Diabetic Nephropathy or Nondiabetic Chronic Kidney Disease

Diabetic nephropathy (Chapter 124) is accompanied by proteinuria, loss of renal autoregulation, hypertension, progression to end-stage renal disease, and a high incidence of cardiovascular events. Because the addition of an ARB but not amlodipine to background antihypertensive therapy slows progression of nephropathy in patients with type 2 diabetes, type 2 diabetes with nephropathy is an indication for an ARB. Evidence suggests an office blood pressure goal of less than 140/90 mm Hg for patients with type 2 diabetic nephropathy and even a goal of less than 130/80 mm Hg in patients with significant proteinuria (urine plasma albumin-to-creatinine ratio of more than 30 mg/g (corresponding to >30 mg of urinary albumin excretion in 24 hours).

Similar goals are recommended for patients with proteinuric nondiabetic chronic kidney disease, in whom ramipil appears to be more renoprotective than either amlodipine or metoprolol. Aliskiren should not be added to background therapy with an ACE-I or ARB because such dual renin-angiotensin system blockade produces hyperkalemia and hypotension while producing no added cardiovascular benefit.

Blood Pressure Reduction in Patients with Prior Coronary Events or Strokes

Evidence from randomized trials is insufficient to determine an optimal blood pressure treatment target for the secondary prevention of coronary events in patients with preexisting coronary disease. Current recommendations include goal office blood pressure of less than 140/90 mm Hg and using two or more medications if this office goal is not achieved. β-Blockers and CCBs are preferred drugs because they are both antihypertensive and antianginal. Overtreatment of diastolic blood pressure can theoretically impair coronary perfusion, worsen myocardial ischemia, and provoke coronary events in patients with underlying coronary disease, but prospective data have not yet defined a critical lower limit of on-treatment diastolic blood pressure. Nevertheless, it is probably prudent not to treat a diastolic blood pressure below 60 mm Hg.

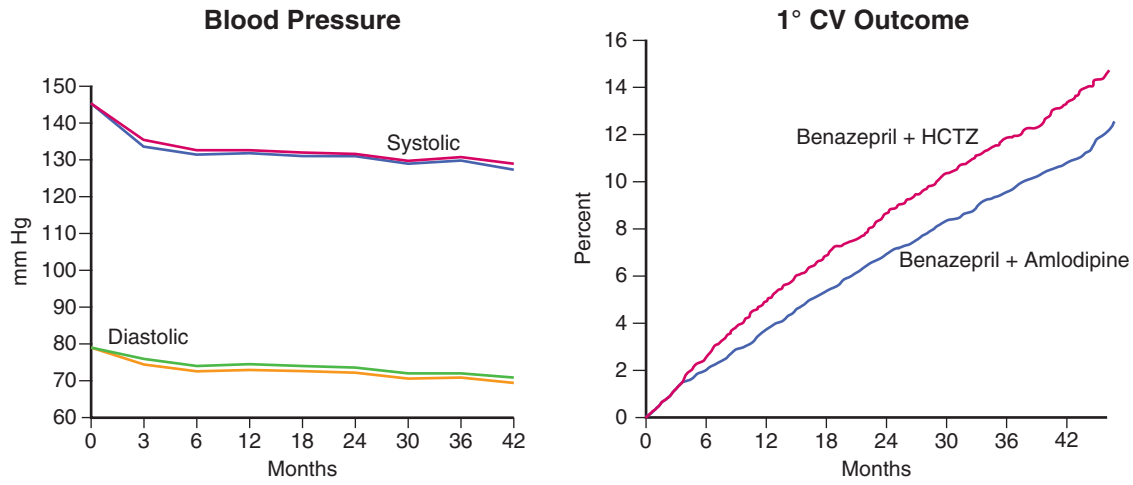

E-FIGURE 67-4. Major outcomes of the Avoiding Cardiovascular Events through Combination Therapy in Patients Living with Systolic Hypertension (ACCOMPLISH) Trial. Systolic and diastolic levels achieved and corresponding Kaplan-Meier analysis of the primary outcome (death from cardiovascular causes, nonfatal myocardial infarction, nonfatal stroke, hospitalization for angina, resuscitation after sudden cardiac arrest, and coronary revascularization) are shown for patients randomized to benazepril plus amlodipine or benazepril plus HCTZ. CV = cardiovascular event. (From Jamerson K, Weber MA, Bakris GL, et al. Benazepril plus amlodipine or hydrochlorothiazide for hypertension in high-risk patients. *N Engl J Med.* 2008;359:2417-2428.)

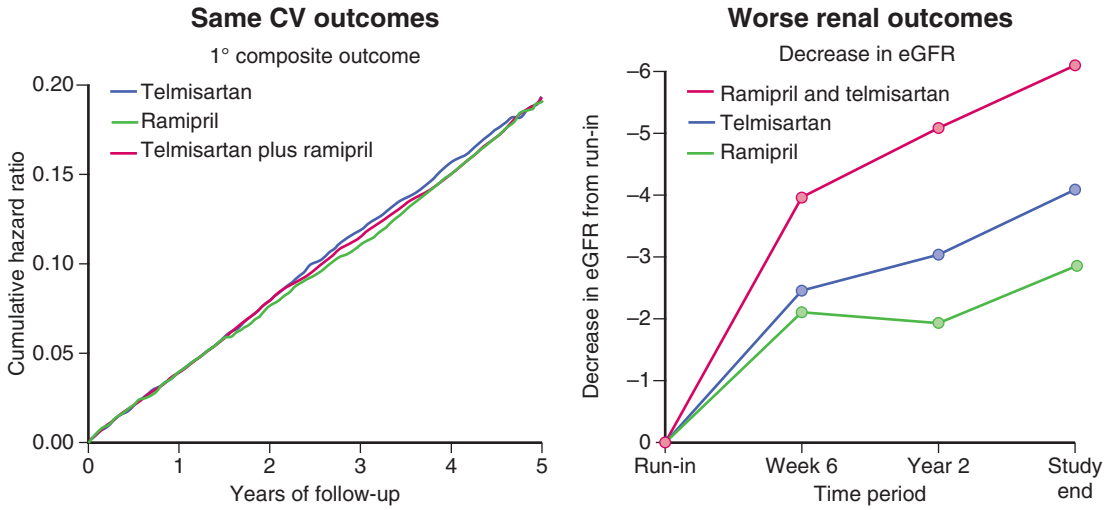

E-FIGURE 67-5. Major cardiovascular and renal outcomes of the Ongoing Telmisartan Alone and in Combination with Ramipril Global Endpoint Trial (ONTARGET). ***Left panel,*** **The cumulative hazard ratio for the primary composite cardiovascular event (CV) outcome (death from cardiovascular causes, myocardial infarction, stroke, or hospitalization for heart failure) was indistinguishable for patients randomized to telmisartan alone (T), ramipril alone (R), or combination treatment with telmisartan plus ramipril (R+T). (From Yusuf S, Teo KK, Pogue J, et al. Telmisartan, ramipril, or both in patients at high risk for vascular events.** ***N Engl J Med.*** **2008;358:1547-1559.)** ***Right panel,*** **The decrease in estimated glomerular filtration rate (eGFR) was best with ramipril alone, intermediate with telmisartan alone, and worst when telmisartan and ramipril were combined. (From Mann JF, Schmieder RE, McQueen M, et al. Renal outcomes with telmisartan, ramipril, or both, in people at high vascular risk (the ONTARGET study): a multicentre, randomised, double-blind, controlled trial.** ***Lancet.*** **2008;372:547-553.)**

TABLE 67-10 DIFFERENCES AMONG CURRENT TREATMENT GUIDELINES FOR ADULTS WITH HYPERTENSION

GUIDELINE	POPULATION	GOAL BLOOD PRESSURE (mm Hg)	INITIAL DRUG TREATMENT OPTIONS
2014 JNC 8 Committee[1]	General ≥60 yr	<150/90	Nonblack: thiazide, ACE-I or ARB, CCB
	General <60 yr	<140/90	Black: thiazide, CCB
	Diabetes	<140/90	Thiazide, ACE-I or ARB, CCB
	CKD	<140/90	ACE-I or ARB
2014 ASH/ISH[2]	General ≥80 yr	<150/90	Nonblack/stage 1: thiazide, ACE-I or ARB, CCB
	General <80 yr	<140/90	Black/stage 1: thiazide, CCB
			Stage 2: CCB or thiazide *plus* ACE-I or ARB
	Diabetes	<140/90	ACE-I or ARB
	CKD	<140/90	ACE-I or ARB
2013 AHA/ACC/CDC[3]	General	<140/90	Stage 1: thiazide for most or ACE-I or ARB, CCB
			Stage 2: thiazide *plus* ACE-I or ARB *or* thiazide *plus* CCB *or* ACE-I or ARB *plus* CCB
2013 ESH/ESC[4]	General ≥80 yr	<150/90	BB, thiazide, CCB, ACE-I or ARB
	General 60-79 yr	<150/90 *or* <140/90	ARB
	General ≤60 yr	<140/90	ARB
	Diabetes	<140/85	ACE-I or ARB
	CKD no proteinuria	<140/90	ACE-I or ARB
	CKD + proteinuria	<130/90	ACE-I or ARB
2013 CHEP[5]	General ≥80 yr	<150/90	Thiazide, BB (<60 yr), ACE-I or ARB (nonblack)
	General <80 yr	<140/90	Thiazide, BB (<60 yr), ACE-I or ARB (nonblack)
	Diabetes	<130/80	ACE-I or ARB (+ additional CVD risk); ACE-I or ARB, thiazide, CCB (– additional CVD risk)
	CKD	<140/90	ACE-I or ARB
2013 ADA[6]	Diabetes	<140/80	ACE-I or ARB
2012 KDIGO[7]	CKD no proteinuria	≤140/90	ACE-I or ARB
	CKD + proteinuria	≤130/80	ACE-I or ARB
2011 UK NICE[8]	General ≥80 yr	<150/90	≥55 yr or black: CCB, thiazide
	General <80 yr	<140/90	<55 yr: ACE-I or ARB
2011ACCF/AHA: Elderly hypertensives[9]	General ≥80 yr	SBP ≤140 or 145	ACE-I or ARB, CCB, thiazide
	General <80 yr	SBP ≤140	
2010 ISHIB[10]	Black	<135/85	Thiazide, CCB
	Black + target organ disease or CVD risk	<130/80	

ACE-I = angiotensin-converting enzyme inhibitor; ARB = angiotensin receptor blocker; CCB = calcium-channel blocker; CVD = cardiovascular disease.

[1]James PA, Oparil S, Carter BL, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA.* 2014;311:507-520.
[2]Weber MA, Schiffrin EL, White WB, et al. Clinical practice guidelines for the management of hypertension in the community a statement by the American Society of Hypertension and the International Society of Hypertension. *J Hypertens.* 2014;32:3-15.
[3]Go AS, Bauman M, Coleman King SM, et al. An effective approach to high blood pressure control: a science advisory from the American Heart Association, the American College of Cardiology, and the Centers for Disease Control and Prevention. *Hypertension.* 2014;63:1230-1238.
[4]Mancia G, Fagard R, Narkiewicz K, et al. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J Hypertens.* 2013;31:1281-1357.
[5]Canadian Hypertension Education Program (CHEP) 2013 Recommendations. Retrieved August 4, 2014 from http://www.hypertension.ca/chep.
[6]American Diabetes Association. Standards of medical care in diabetes: 2013. *Diabetes Care.* 2013;36(Suppl 1):S11-66.
[7]Kidney Disease: Improving Global Outcomes (KDIGO) Blood Pressure Work Group. KDIGO clinical practice guideline for the management of blood pressure in chronic kidney disease. *Kidney Int Suppl.* 2012;2:337-414.
[8]National Institute for Health and Clinical Excellence (NICE). Clinical guideline 127. Hypertension: clinical management of primary hypertension in adults. 2011. www.nice.org.uk/guidance/cg127.
[9]Aronow WS, Fleg JL, Pepine CJ, et al. ACCF/AHA 2011 expert consensus document on hypertension in the elderly: a report of the American College of Cardiology Foundation Task Force on Clinical Expert Consensus Documents. *Circulation.* 2011;123:2434-2506.
[10]Flack JM, Sica DA, Bakris G, et al. Management of high blood pressure in blacks: an update of the International Society on Hypertension in Blacks consensus statement. *Hypertension.* 2010;56:780-800.

Stroke survivors are at high risk for recurrent stroke, further disability, and death. Reduction in systolic blood pressure to below 130 mm Hg with thiazide and ACE inhibitor combination therapy can reduce these risks.[11]

Hypertension in Minority Populations

Mexican Americans have the lowest rate of control of hypertension of all U.S. racial/ethnic groups but also have the highest risk for diabetes. Thus, antihypertensive regimens should be tailored to avoid causing more new cases of diabetes (e.g., by avoiding high-dose chlorthalidone). In African Americans, hypertension not only is more prevalent than in the general population but also starts at a younger age, is less well controlled, and causes disproportionate and premature disability and death.[5] African Americans commonly have lower plasma renin levels and have less response to monotherapy with an ACE inhibitor or ARB than do white hypertensive patients. African American participants have a higher risk for fatal stroke when taking an ACE inhibitor alone than when taking a diuretic alone, but racial/ethnic differences disappear when high doses of an ACE inhibitor or ARB are used in combination with a diuretic or CCB. ACE inhibitors or ARBs can help achieve excellent control of hypertension in African American patients when used as part of an appropriate multidrug regimen.

Hypertension Associated with Oral Contraceptives and Estrogen Replacement

Oral contraceptives, particularly current low-dose estrogen preparations, cause a small increase in blood pressure in most women but rarely cause a large increase into the hypertensive range. The mechanism is unknown, but women older than 35 years and those who smoke or are overweight appear to be at increased risk. If hypertension develops, oral contraceptive therapy

TABLE 67-11 PREFERRED ORAL ANTIHYPERTENSIVE DRUG CLASSES IN SPECIFIC CONDITIONS

CONDITION	DRUG CLASS(ES)
Prehypertension	ARB
Hypertension in general	CCB, ACE-I or ARB, D
Hypertension in elderly patients	CCB, ACE-I or ARB, D
Hypertension with left ventricular hypertrophy	ARB, D, CCB
Hypertension in patients with diabetes mellitus	CCB, ACE-I or ARB, D
Hypertension in patients with diabetic nephropathy	ARB, D
Hypertension in nondiabetic chronic kidney disease	ACE-I or ARB, BB, D
BP reduction for secondary prevention of coronary events	ACE-I, CCB, BB, D
BP reduction for secondary prevention of stroke	ACE-I + D, CCB
BP management for patient with heart failure	D, BB, ACE-I or ARB, aldosterone antagonist
Gestational hypertension (stage 2, without preeclampsia)	Labetalol, nifedipine, methyldopa
Thoracic aortic aneurysm	BB, ACE-I or ARB, D
Atrial fibrillation (ventricular rate control)	BB, nondihydropyridine CCB

ACE-I = angiotensin-converting enzyme inhibitor; ARB = angiotensin receptor blocker; BB = β-blocker; CCB = calcium-channel blocker; D, diuretic (thiazide-like such as chlorthalidone is preferred).
Adapted from Mancia G, Fagard R, Narkiewicz K, et al. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J Hypertens.* 2013;31:1281-1357.

should be discontinued in favor of other methods of contraception. Oral estrogen replacement therapy after menopause appears to cause a small increase in blood pressure, whereas transdermal estrogen (which bypasses first-pass hepatic metabolism) appears to cause a small decrease in blood pressure.

Hypertension in Pregnancy

Hypertension, the most common nonobstetric complication of pregnancy, is present in about 10% of all pregnancies (Chapter 239). About one third of cases are caused by chronic hypertension and two thirds by gestational hypertension or preeclampsia, the latter defined as an increase in blood pressure to 160/110 mm Hg or higher after the 20th week of gestation, accompanied by proteinuria and pathologic edema. Preeclampsia sometimes also is accompanied by seizures (eclampsia) and the multisystem HELLP syndrome (Chapter 239) of hemolysis, elevated liver enzymes, and low platelets. Preeclampsia is the most common cause of maternal mortality and perinatal mortality.

Current guidelines recommend low-dose aspirin, beginning in the first trimester, to reduce the risk for recurrent preeclampsia in women with a past history of preeclampsia.[12] Women with gestational hypertension or chronic hypertension should be monitored twice weekly with measurements of blood pressure and weekly assessment of platelet counts, liver enzymes, and proteinuria.

Antihypertensive medication of mild maternal hypertension does not improve perinatal outcome and may be associated with fetal growth retardation, so medications are not recommended for uncomplicated stage 1 gestational hypertension but rather are reserved for stage 2 hypertension (blood pressure > 160/110 mm Hg). For pregnant women with stage 2 hypertension but without severe preeclampsia/eclampsia, oral drug therapy should be initiated with any one of three preferred drugs: labetalol (400 to 2400 mg daily), nifedipine XL (30 to 120 mg daily), or methyldopa (500 to 3000 mg daily). Combination therapy is rarely needed, and excessive reductions in blood pressure must be avoided. All renin-angiotensin system blockers must be discontinued. The definitive cure of preeclampsia is termination of pregnancy.

Intravenous labetalol (0.5 to 2 mg/minute up to a cumulative dose of 300 mg) has replaced hydralazine as the drug of choice to treat severe preeclampsia/eclampsia. Intravenous magnesium sulfate is not a reliable antihypertensive agent but is effective in treating or preventing seizures (eclampsia). Delivery soon after maternal stabilization is recommended irrespective of gestational age for women with superimposed preeclampsia and any of the following: uncontrollable severe hypertension, eclampsia, pulmonary edema, abruptio placentae, disseminated intravascular coagulation, or fetal distress. Intravenous nitroglycerin (10 to 100 μg/minute) is the treatment of choice when pulmonary edema accompanies preeclampsia.

For women with preeclampsia or even gestational hypertension, blood pressure should be monitored closely in the hospital for 72 hours postpartum and again as an outpatient 7 to 10 days after delivery. All blood pressure drugs are secreted into human breast milk, but only propranolol and nifedipine are secreted in high enough concentrations that they should be avoided in mothers who are breast-feeding. Women whose preeclampsia caused preterm delivery have an almost 10-fold increased risk for cardiovascular disease in later life and are candidates for aggressive risk factor modification.

TABLE 67-12 CAUSES OF PSEUDORESISTANT AND TRULY RESISTANT HYPERTENSION

PSEUDORESISTANT HYPERTENSION	TRULY RESISTANT HYPERTENSION
Inadequate medical regimen Pressor substances (e.g., nonsteroidal anti-inflammatory drugs [NSAIDs], calcineurin inhibitors such as cyclosporine or tacrolimus, or sympathomimetics such as cocaine or methamphetamine) White coat reaction, improper blood pressure measurement Medication nonadherence	Chronic kidney disease Primary aldosteronism Other secondary hypertension (e.g., pheochromocytoma, Cushing syndrome, atherosclerotic renal artery stenosis, fibromuscular renal artery stenosis, Takayasu arteritis, coarctation of the aorta, hyperthyroidism, hypothyroidism, hyperparathyroidism) Difficult primary hypertension

Drug-Resistant Hypertension

Up to one in five hypertensive patients has *resistant hypertension*, defined as high blood pressure uncontrolled with three antihypertensive drugs, including a diuretic, or controlled on four or more drugs.[13] More than half of such patients have *pseudoresistance* because of improper blood pressure measurement techniques, white coat reactions, medication nonadherence, intake of drugs that raise blood pressure (e.g., NSAIDs, excessive alcohol, psychiatric drugs), or an inadequate blood pressure regimen (Table 67-12). Common correctable issues are clonidine rebound (especially with as-needed dosing), inadequate diuretic therapy, inappropriate use of a loop diuretic in a patient with normal renal function, infrequent dosing with a short-acting loop diuretic (e.g., once-daily furosemide), and use of a low-dose thiazide in a patient with chronic kidney disease.

Truly drug-resistant patients are at high risk because of their severe hypertension and target organ damage. Patients should be screened for secondary hypertension, especially chronic kidney disease, obstructive sleep apnea (Chapter 100), primary aldosteronism (Chapter 227), and pheochromocytoma (Chapter 228). In the absence of an identifiable cause for the hypertension, a mineralocorticoid receptor antagonist or a vasodilating β-blocker can serve as highly effective add-on therapies. Low-dose eplerenone or spironolactone can be remarkably effective for resistant hypertension—even when the serum aldosterone is within the normal range.

Despite maximally tolerated doses of five or more different antihypertensive medications, some patients still have uncontrolled hypertension. Unfortunately, neither catheter-based renal denervation[14] nor an implantable carotid baroreceptor pacemaker have been shown to be reliably beneficial in such patients.

Acute Severe Hypertension

Twenty-five percent of all emergency department patients present with an elevated blood pressure (Chapter 8). *Hypertensive emergencies* are acute, often severe elevations in blood pressure, accompanied by rapidly progressive target organ dysfunction, such as myocardial or cerebral ischemia or infarction, pulmonary edema, or renal failure. The blood pressure typically is 220/130 mm Hg or higher, but it may be much lower in women who have preeclampsia in the absence of preexisting hypertension. The full-blown clinical picture of a hypertensive emergency is a critically ill patient who presents with a blood pressure typically above 220/130 mm Hg, headaches, confusion, blurred vision, nausea and vomiting, seizures, pulmonary edema, oliguria, and grade 3 or grade 4 hypertensive retinopathy (Fig. 67-8). *Hypertensive emergencies* require immediate reduction of blood pressure with intravenous medication (Table 67-13) and intra-arterial monitoring in an intensive care unit (ICU). In contrast, *hypertensive urgency* denotes severe uncontrolled hypertension sometimes with vague symptoms (such as headache, malaise, anxiety) but without objective evidence of acute target organ damage. In the absence of acute target organ damage, a patient with a blood pressure of 220/130 mm Hg or higher should be treated with short-acting oral medication. *Severe hypertension,* defined as a blood pressure between 180/110 and 219/129 mm Hg without symptoms or acute target organ damage, almost always occurs in patients who have chronic hypertension and who ran out of or stopped taking their blood pressure medication. Long-acting oral medication simply can be restarted. Patients with a hypertensive urgency or severe hypertension require outpatient follow-up within 24 to 72 hours with either a primary care physician or hypertension specialist.

The most common hypertensive cardiac emergencies include acute aortic dissection (Chapter 78), hypertension after cardiac surgery (Chapter 74), acute

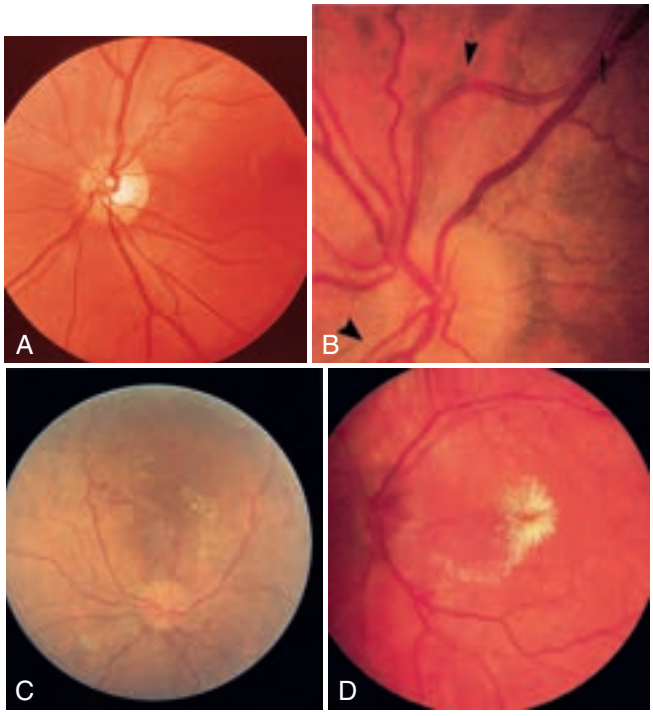

FIGURE 67-8. **Hypertensive retinopathy is traditionally divided into four grades. A, Grade 1 shows early and minor changes in a young patient. Increased tortuosity of a retinal vessel and increased reflectiveness (silver wiring) of a retinal artery are seen at 1 o'clock in this view. Otherwise, the fundus is completely normal. B, Grade 2 also shows increased tortuosity and silver wiring *(arrowheads)*. In addition, there is "nipping" of the venules at arteriovenous crossings. C, Grade 3 shows the same changes as grade 2 plus flame-shaped retinal hemorrhages and soft "cotton-wool" exudates. D, In grade 4, there is swelling of the optic disc (papilledema), retinal edema is present, and hard exudates may collect around the fovea, producing a typical "macular star." (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine.* 3rd ed. London: Mosby; 2003.)**

myocardial infarction (Chapter 73), and unstable angina (Chapter 72). Other hypertensive emergencies include cocaine-induced sympathetic crisis, eclampsia (Chapter 239), head trauma (Chapter 399), severe body burns (Chapter 111), postoperative bleeding from vascular suture lines, and epistaxis that cannot be controlled with anterior and posterior nasal packing. Neurologic emergencies—acute ischemic stroke, hemorrhagic stroke, subarachnoid hemorrhage, and hypertensive encephalopathy—can be difficult to distinguish from one another (Chapters 406 to 408). Hypertensive encephalopathy (Chapter 408) is characterized by severe hypertensive retinopathy (retinal hemorrhages and exudates, with or without papilledema) and a posterior leukoencephalopathy (affecting mainly the white matter of the parieto-occipital regions) seen on cerebral MRI or CT. A new focal neurologic deficit suggests a stroke in evolution, which demands a much more conservative approach to the elevated blood pressure (Chapter 407).

In most other hypertensive emergencies, the goal of parenteral therapy is to achieve a controlled and gradual lowering of blood pressure. A good rule of thumb is to lower the initially elevated arterial pressure by 10% in the first hour and by an additional 15% during the next 3 to 12 hours to a blood pressure of no less than 160/110 mm Hg. Blood pressure can be reduced further during the next 48 hours. Exceptions to this rule are aortic dissection (Chapter 78) and postoperative bleeding from vascular suture lines, two situations that demand much more rapid normalization of blood pressure. In most other cases, unnecessarily rapid correction of the elevated blood pressure to completely normal values places the patient at high risk for worsening cerebral, cardiac, and renal ischemia. In chronic hypertension, cerebral autoregulation is reset to higher than normal blood pressures. This compensatory adjustment prevents tissue overperfusion (increased intracranial pressure) at very high blood pressures, but it also predisposes to tissue underperfusion (cerebral ischemia) when an elevated blood pressure is lowered too quickly (Chapter 407). In patients with coronary disease, overly rapid or excessive reduction in diastolic blood pressure in the ICU can precipitate an acute myocardial ischemia or infarction.

Intravenous Drugs for Hypertensive Emergencies

First-line drug options for hypertensive emergencies (Table 67-14) are intravenous labetalol (a combined α- and β-blocker), nitroprusside, nicardipine (a dihydropyridine CCB), or urapadil (a new central sympatholytic that acts on central serotonergic pathways and also selectively blocks peripheral α_1-adrenergic receptors). In patients with impaired cerebral autoregulation (see

TABLE 67-13 RECOMMENDED TREATMENT OF HYPERTENSIVE EMERGENCIES BY END ORGAN INVOLVED

TYPE OF EMERGENCY	TIME-LINE, TARGET BP	FIRST-LINE THERAPY	ALTERNATIVE THERAPY
Hypertensive crisis with retinopathy, microangiopathy, or acute renal insufficiency	Several hours, MAP −20 to −25%	Labetalol	Nitroprusside Nicardipine Urapadil
Hypertensive encephalopathy	Immediate, MAP −20 to −25%	Labetalol	Nicardipine Nitroprusside
Acute aortic dissection	Immediate, Systolic BP < 110 mm Hg	Nitroprusside plus metoprolol	Labetalol
Acute pulmonary edema	Immediate, MAP 60 to 100 mm Hg	Nitroprusside with loop diuretic	Nitroglycerine Urapadil with loop diuretic
Acute coronary syndrome	Immediate, MAP 60 to 100 mm Hg	Nitroglycerine	Labetalol
Acute ischemic stroke and BP > 220/120 mm Hg	1 hr, MAP −15%	Labetalol	Nicardipine Nitroprusside
Cerebral hemorrhage and Systolic BP > 180 mm Hg or MAP > 130 mm Hg	1 hr, Systolic BP < 180 mm Hg and MAP < 130 mm Hg	Labetalol	Nicardipine Nitroprusside
Acute ischemic stroke with indication for thrombolytic therapy and BP > 185/110 mm Hg	1 hr, MAP < −15%	Labetalol	Nicardipine Nitroprusside
Cocaine/XTC intoxication	Several hours, Systolic BP < 140 mm Hg	Phentolamine (after benzodiazepines)	Nitroprusside
Pheochromocytoma crisis	Immediate	Phentolamine	Nitroprusside Urapadil
Perioperative hypertension during or after CABG	Immediate	Nicardipine	Urapadil Nitroglycerine
During or after craniotomy	Immediate	Nicardipine	Labetalol
Severe preeclampsia/eclampsia	Immediate, BP < 160/105 mm Hg	Labetalol (plus $MgSO_4$ and oral antihypertensive medication such as nifedipine with or without methyldopa)	Ketanserin Nicardipine

Adapted from van den Born BJ, Beutler JJ, Gaillard CA, et al. Dutch guideline for the management of hypertensive crisis: 2010 revision. *Neth J Med.* 2011;69:248-255.
BP = blood pressure; CABG = coronary artery bypass grafting; MAP = mean arterial pressure; MgSO4 = magnesium sulfate; XTC = ecstasy.

TABLE 67-14 INTRAVENOUS DRUGS FOR HYPERTENSIVE EMERGENCIES

DRUG	ONSET OF ACTION	HALF-LIFE	DOSE	CONTRAINDICATIONS AND SIDE EFFECTS
Labetalol	5-10 min	3-6 hr	0.25-0.5 mg/kg; 2-4 mg/min until goal BP is reached, thereafter 5-20 mg/hr	Second- or third-degree AV block; systolic heart failure, COPD (relative); bradycardia
Nicardipine	5-15 min	30-40 min	5-15 mg/hr as continuous infusion, starting dose 5 mg/hr, increase every 15-30 min with 2.5 mg until goal BP, thereafter decrease to 3 mg/hr	Liver failure
Nitroprusside	Immediate	1-2 min	0.3-10 μg/kg/min, increase by 0.5 μg/kg/min every 5 min until goal BP	Liver/kidney failure (relative), cyanide toxicity
Nitroglycerine	1-5 min	3-5 min	5-200 μg/min, 5 μg/min increase every 5 min	
Urapadil	3-5 min	4-6 hr	12.5-25 mg as bolus injections; 5-40 mg/hr as continuous infusion	
Esmolol	1-2 min	10-30 min	0.5-1.0 mg/kg as bolus; 50-300 μg/kg/min as continuous infusion	Second- or third-degree AV block, systolic heart failure, COPD (relative); bradycardia
Phentolamine	1-2 min	3-5 min	1-5 mg, repeat after 5-15 min until goal BP is reached; 0.5-1 mg/hr as continuous infusion	Tachyarrhythmia, angina pectoris

AV = atrioventricular; BP = blood pressure; COPD = chronic obstructive pulmonary disease.
Adapted from van den Born BJ, Beutler JJ, Gaillard CA, et al. Dutch guideline for the management of hypertensive crisis: 2010 revision. *Neth J Med.* 2011;69:248-255.

later), labetalol causes a smaller adverse fall in cerebral blood flow than nitroprusside but has a longer half-life, thereby leading to more adverse episodes of systemic hypotension.[15] Intravenous nicardipine appears to produce a more predictable and consistent reduction in blood pressure than labetalol with a similar safety profile; however, physicians and hospital pharmacies are less familiar with nicardipine.[16]

After the blood pressure has been brought under acute control, oral labetalol and dihydropyridine CCBs are particularly useful agents in weaning patients from parenteral therapy so that they can be transferred from the ICU. A few doses of intravenous furosemide are often needed to overcome drug resistance due to secondary volume expansion resulting from parenteral vasodilator therapy.

Secondary hypertension should be suspected in patients admitted to the ICU with hypertensive crisis. Normal 24-hour urinary catecholamine values or normal plasma normetanephrine and metanephrine values collected when the blood pressure is the highest (first 24 hours in ICU) effectively rule out pheochromocytoma (Chapter 228). Bilateral renal artery stenosis (Chapter 125) and other secondary causes should be excluded after the patient has been transferred out of the ICU but before discharge from the hospital.

Oral Medications for Hypertensive Urgencies

Labetalol is effective in a dose of 200 to 300 mg, which can be repeated in 2 to 3 hours and then prescribed in twice-daily dosing. If a β-blocker is contraindicated, clonidine is effective in an initial dose of 0.1 or 0.2 mg followed by additional hourly doses of 0.1 mg. Captopril, a short-acting ACE inhibitor, lowers blood pressure within 15 to 30 minutes of oral dosing. A small test dose of 6.25 mg should be used to avoid an excessive fall in blood pressure in hypovolemic patients; then, the full oral dose is 25 mg, which can be repeated in 1 to 2 hours and prescribed as 25 to 75 mg twice daily.

Incidental Blood Pressure Elevation in the Emergency Department

Blood pressures above 160/110 mm Hg are a common incidental finding among patients who present to emergency departments and other acute care settings for urgent medical or surgical care of symptoms that are unrelated to blood pressure (e.g., musculoskeletal pain, orthopedic injury). The elevated blood pressure more often is the first indication of chronic hypertension than a simple physiologic stress reaction, so there is an important opportunity to initiate primary care referral for formal evaluation of possible chronic hypertension. Home or ambulatory blood pressure monitoring is indicated to determine whether the patient's blood pressure normalizes completely after the acute illness has resolved.

PROGNOSIS

The prognosis of the patient with hypertension is related both to the severity of the blood pressure elevation and the presence of additional cardiovascular risk factors. Undertreatment of hypertension and underuse of combination drug therapy, both of which are common in busy outpatient practices,[17] worsens outcomes, whereas management protocols with fixed-dose/once-daily combination pills, proactive follow-up, and access to walk-in blood pressure checks can improve hypertension control rates to as high as 80% (Table 67-15). In many states, pharmacists can work with patients and can implement a preset medication intensification protocol under a collaborative practice agreement with physician oversight. Long-term continuation rates are best for the ARBs, intermediate for ACE inhibitors and CCBs, and worst for diuretics and β-blockers.[18] Patients with drug-resistant hypertension should be referred to a hypertension specialist (www.ash-us.org/HTN-Specialist.aspx).

TABLE 67-15 STRATEGIES TO OPTIMIZE HYPERTENSION MANAGEMENT

HEALTH SYSTEM
- Standardized medication intensification protocol
- Team-based approach involving clinical pharmacists
- Pay providers for performance

DRUG TREATMENT
- Low-dose combination therapy
- Best-tolerated drug classes
- Fixed-dose single pill combinations
- Long-acting once daily drugs
- Low-cost generics

PATIENT ENGAGEMENT
- Shared goals
- Medication reconciliation/education
- Out-of-office blood pressure monitoring
- Social support from family and peers

Grade A References

A1. Cooper CJ, Murphy TP, Cutlip DE, et al. Stenting and medical therapy for atherosclerotic renal-artery stenosis. *N Engl J Med.* 2014;370:13-22.

A2. Wheatley K, Ives N, Gray R, et al. Revascularization versus medical therapy for renal-artery stenosis. *N Engl J Med.* 2009;361:1953-1962.

A3. Sundstrom J, Arima H, Woodward M, et al. Blood pressure-lowering treatment based on cardiovascular risk: a meta-analysis of individual patient data. *Lancet.* 2014;384:591-598.

A4. Eckel RH, Jakicic JM, Ard JD, et al. 2013 AHA/ACC Guideline on Lifestyle Management to Reduce Cardiovascular Risk: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2014;63:2960-2984.

A5. ALLHAT Officers and Coordinators. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA.* 2002;288:2981-2997.

A6. Jamerson K, Weber MA, Bakris GL, et al. Benazepril plus amlodipine or hydrochlorothiazide for hypertension in high-risk patients. *N Engl J Med.* 2008;359:2417-2428.

A7. Mann JF, Schmieder RE, McQueen M, et al. Renal outcomes with telmisartan, ramipril, or both, in people at high vascular risk (the ONTARGET study): a multicentre, randomised, double-blind, controlled trial. *Lancet.* 2008;372:547-553.

A8. Fagard RH, Celis H, Thijs L, et al. Regression of left ventricular mass by antihypertensive treatment: a meta-analysis of randomized comparative studies. *Hypertension.* 2009;54:1084-1091.

A9. Kidney Disease: Improving Global Outcomes (KDIGO) Blood Pressure Work Group. KDIGO clinical practice guideline for the management of blood pressure in chronic kidney disease. *Kidney Int Suppl*. 2012;2:337-414.
A10. Bhatt DL, Kandzari DE, O'Neill WW, et al. A controlled trial of renal denervation for resistant hypertension. *N Engl J Med*. 2014;370:1393-1401.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. James PA, Oparil S, Carter BL, et al. 2014 Evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014;311:507-520.
2. Weber MA, Schiffrin EL, White WB, et al. Clinical practice guidelines for the management of hypertension in the community: a statement by the American Society of Hypertension and the International Society of Hypertension. *J Hypertens*. 2014;32:3-15.
3. Mancia G, Fagard R, Narkiewicz K, et al. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J Hypertens*. 2013;31:1281-1357.
4. National Institute for Health and Clinical Excellence (NICE). Clinical guideline 127. Hypertension: clinical management of primary hypertension in adults. www.nice.org.uk/guidance/cg127; 2011. Accessed February 18, 2015.
5. Centers for Disease Control and Prevention. Racial/ethnic disparities in the awareness, treatment, and control of hypertension: United States, 2003-2010. *MMWR Morb Mortal Wkly Rep*. 2013;62:351-355.
6. Franklin SS, Thijs L, Hansen TW, et al. Significance of white-coat hypertension in older persons with isolated systolic hypertension: a meta-analysis using the International Database on Ambulatory Blood Pressure Monitoring in Relation to Cardiovascular Outcomes population. *Hypertension*. 2012;59:564-571.
7. Trinquart L, Mounier-Vehier C, Sapoval M, et al. Efficacy of revascularization for renal artery stenosis caused by fibromuscular dysplasia: a systematic review and meta-analysis. *Hypertension*. 2010;56:525-532.
8. Cook NR, Appel LJ, Whelton PK. Lower levels of sodium intake and reduced cardiovascular risk. *Circulation*. 2014;129:981-989.
9. Wright JT Jr, Fine LJ, Lackland DT, et al. Evidence supporting a systolic blood pressure goal of less than 150 mm Hg in patients aged 60 years or older: the minority view. *Ann Intern Med*. 2014;160:499-503.
10. Arnold AC, Okamoto LE, Gamboa A, et al. Angiotensin II, independent of plasma renin activity, contributes to the hypertension of autonomic failure. *Hypertension*. 2013;61:701-706.
11. Davis SM, Donnan GA. Clinical practice. Secondary prevention after ischemic stroke or transient ischemic attack. *N Engl J Med*. 2012;366:1914-1922.
12. Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy. *Obstet Gynecol*. 2013;122:1122-1131.
13. Vongpatanasin W. Resistant hypertension: a review of diagnosis and management. *JAMA*. 2014;311:2216-2224.
14. Mahfoud F, Luscher TF, Andersson B, et al. Expert consensus document from the European Society of Cardiology on catheter-based renal denervation. *Eur Heart J*. 2013;34:2149-2157.
15. Immink RV, van den Born BJ, van Montfrans GA, et al. Cerebral hemodynamics during treatment with sodium nitroprusside versus labetalol in malignant hypertension. *Hypertension*. 2008;52:236-240.
16. Peacock WF 4th, Hilleman DE, Levy PD, et al. A systematic review of nicardipine vs labetalol for the management of hypertensive crises. *Am J Emerg Med*. 2012;30:981-993.
17. Khanna RR, Victor RG, Bibbins-Domingo K, et al. Missed opportunities for treatment of uncontrolled hypertension at physician office visits in the United States, 2005 through 2009. *Arch Intern Med*. 2012;172:1344-1345.
18. Jaffe MG, Lee GA, Young JD, et al. Improved blood pressure control associated with a large-scale hypertension program. *JAMA*. 2013;310:699-705.

REVIEW QUESTIONS

1. An 80-year-old woman makes an appointment for initial evaluation of hypertension, which was detected by her orthopedist. She has enjoyed excellent health her entire life and remains very active. She has never been treated for hypertension. Her only other medical issue is occasional sciatica from lumbar degenerative disc disease. Her only medication is vitamin D with calcium. In your office, her seated blood pressure averages 200/90 mm Hg and heart rate is 77 beats per minute. There are no significant orthostatic changes in her blood pressure or heart rate. Her complete metabolic panel and resting electrocardiogram are normal. Which of the following is the most appropriate course of action?

A. Initiate antihypertensive therapy by prescribing a thiazide-type diuretic.
B. Schedule a follow-up visit in 2 weeks to repeat the blood pressure measurement to confirm the diagnosis of hypertension before prescribing medication.
C. Order an ambulatory blood pressure monitor.
D. Send the patient to the emergency department for treatment of hypertensive urgency.
E. Initiate antihypertensive therapy with an angiotensin-converting enzyme (ACE) inhibitor plus a thiazide because the patient has stage 2 hypertension.

Answer: C White coat hypertension and masked hypertension are so common in elderly patients that the office blood pressure will lead to overtreatment or undertreatment of blood pressure in 3 out of 4 patients. Because this patient has no evidence of symptomatic or asymptomatic target organ damage, you suspect white coat hypertension; the best option is to send the patient home with a 24-hour ambulatory blood pressure monitor, which she should wear until she returns to your office the next day. Her out-of-office daytime blood pressure averages 141/55 mm Hg, and her sleeping blood pressure averages 143/57 mm Hg. Normal awake daytime blood pressure is less than 135/85 mm Hg, and normal nighttime sleeping blood pressure is less than 120/70 mm Hg. Thus, this patient has a brisk white coat reaction superimposed on very mild daytime hypertension. She also has nocturnal hypertension (no dip), which merits gentle medication therapy. There is no indication that the patient has a hypertensive urgency. It is not appropriate to institute medication therapy before confirming the diagnosis of hypertension, but a second office visit is almost certainly going to evoke another white coat reaction. (Franklin SS, Thijs L, Hansen TW, et al. Significance of white-coat hypertension in older persons with isolated systolic hypertension: a meta-analysis using the International Database on Ambulatory Blood Pressure Monitoring in Relation to Cardiovascular Outcomes population. *Hypertension.* 2012;59:564-571.)

2. A 52-year-old man has chronic hypertension and has had type 2 diabetes for 8 years. He is treated with diet and metformin. You are treating his hypertension with lisinopril/hydrochlorothiazide 40/25 mg daily and amlodipine 5 mg daily. On this regimen, he has a mild cough. His office blood pressure averages 145/91 mm Hg, and his home blood pressures (which he recorded as you requested) are between 133/83 and 151/92 mm Hg, with an average of 140/87 mm Hg. His basic metabolic panel includes: Na 133 mEq/L, K 3.7 mEq/L, CO_2 25 mmol/L, Cl 100 mmol/L, creatinine 1.2 mg/dL, BUN 22 mg/dL. His eGFR is 65 mL/min/1.73 m^2. His fasting glucose level is 165 mg/dL with a hemoglobin A_{1C} of 7.5%. A spot morning urine albumin-to-creatinine ratio is 350 mg/g. Which of the following is the most appropriate next step?

A. Add an angiotensin receptor blocker (ARB) to the current regimen to reduce the patient's proteinuria and slow the progression of his diabetic nephropathy.
B. Add the new direct renin inhibitor aliskiren to the current regimen to reduce the patient's proteinuria and slow the progression of diabetic nephropathy.
C. Add metoprolol to the current regimen to improve the control of his hypertension.
D. Replace lisinopril/hydrochlorothiazide with an ARB and start controlled-release carvedilol 20 mg daily.
E. Maintain the current regimen.

Answer: D The patient has early diabetic nephropathy with proteinuria and also has an angiotensin-converting enzyme (ACE) inhibitor cough. He is on a good blood pressure regimen with three first-line drugs. However, his blood pressure is not yet at goal (office blood pressure either less than 140/90 mm Hg or less than 130/80 mm Hg by different current guidelines; home blood pressure less than 135/85 mm Hg), and his hemoglobin A_{1C} level is elevated. The best option is to replace the ACE inhibitor with an ARB and to add a vasodilating β-blocker (such as carvedilol), which will not exacerbate his hyperglycemia. His hemoglobin A_{1C} level may improve if his hypertension can be controlled without a thiazide diuretic; if not, low-dose chlorthalidone (6.25 mg daily to start and titrate) would be the best option. Dual renin-angiotensin system blockade with either any combination of an ARB plus ACE inhibitor, or an ACE inhibitor plus aliskiren, increases complications without providing benefit compared with either an ACE inhibitor alone or an ARB alone and should not be used. (Mann JF, Schmieder RE, McQueen M, et al. Renal outcomes with telmisartan, ramipril, or both, in people at high vascular risk [the ONTARGET study]: a multicentre, randomised, double-blind, controlled trial. *Lancet.* 2008;372:547-553; Parving HH, Brenner BM, McMurray JJ, et al. Cardiorenal end points in a trial of aliskiren for type 2 diabetes. *N Engl J Med.* 2012;367:2204-2213.)

3. A 60-year-old male bus driver transfers to your care with a chief complaint of resistant and labile hypertension and an inability to return to work until you provide medical clearance. His current regimen is hydrochlorothiazide 25 mg daily, metoprolol XL 100 mg daily, and clonidine 0.2 mg twice daily, with an extra dose of 0.2 mg if his systolic blood pressure is above 180 mm Hg. His office blood pressure is 189/88 mm Hg with a heart rate of 50 seated, and 195/92 mm Hg with a heart rate of 50 standing. The patient's home blood pressure readings range anywhere from 85/55 to 225/125 mm Hg. He is generally lethargic, often dizzy, and feels anxious when his blood pressure spikes. What is the most likely issue and best course of action?

A. Medication nonadherence: draw drug blood levels and confront the patient if they are low.
B. Clonidine rebound: taper clonidine gradually while starting a calcium-channel blocker (amlodipine 5 mg daily) and an angiotensin receptor blocker (telmisartan), and replace hydrochlorothiazide with chlorthalidone 25 mg daily.
C. Pheochromocytoma: measure plasma metaneprhines and start an α-blocker (doxazosin 8 mg twice daily).
D. Dysautonomia: refer the patient to a neurologist for autonomic function testing
E. Resistant hypertension: refer the patient to a hypertension specialist for consideration of renal denervation

Answer: B This patient likely has pseudoresistant hypertension owing to an inadequate medical regimen with clonidine rebound. His blood pressure is rebounding in your office (the appointment is 8 hours after his morning dose of clonidine). The appropriate course of action is to reduce the clonidine dose to 0.1 mg given 4 times daily, so as to prevent rebound between doses, for 1 or 2 weeks, next to reduce the dose to 0.05 mg four times daily for another week, and finally to stop clonidine altogether and never restart it. At the same time, the patient should be started on the three first-line drugs for hypertension recommended by almost all current guidelines: a long-acting calcium-channel blocker, an angiotensin-converting enzyme inhibitor or angiotensin receptor blocker, and a thiazide-type diuretic (chlorthalidone). The patient's heart rate of 50 beats per minute indicates that he is taking his metoprolol, which is not a first-line drug for hypertension. Metoprolol should be tapered gradually later if he has no indication, such as coronary artery disease or heart failure, for a β-blocker. There is nothing in his history to suggest a pheochromocytoma, which is far less common than clonidine rebound, which itself can cause false-positive metanephrine screening. There is no orthostatic hypotension to suggest autonomic failure. Renal denervation is an exciting potential percutaneous intervention for true drug-resistant hypertension, which he does not have.

4. A 57-year-old African American man has a progressive increase in his serum creatinine from 1.3 to 1.8 mg/dL over the past 2 years, accompanied by microalbuminuria and left ventricular hypertrophy with repolarization abnormalities ("strain") by electrocardiogram despite having office blood pressures consistently averaging 130/65 mm Hg on the following regimen: furosemide 40 mg daily, atenolol 50 mg daily, and amlodipine 5 mg daily. He has no diabetes and quit smoking 15 years ago. What would you recommend?

A. Order a 24-hour ambulatory blood pressure monitor.
B. Continue current medications and refer the patient to a nephrologist for renal biopsy.
C. Replace furosemide with chlorthalidone 25 mg daily, taper off atenolol, and add an angiotensin receptor blocker (ARB).
D. A, B, and C
E. A and C

Answer: E This patient likely has masked and nocturnal hypertension, which can be documented by ambulatory blood pressure monitoring. Masked hypertension and nocturnal hypertension are particularly common in chronic kidney disease, presumably owing to sympathetic overactivity, and they increase the risk for left ventricular hypertrophy and cardiovascular death. His current regimen is not appropriate. An ARB is indicated for renal protection in chronic kidney disease with microalbuminuria and for regression of left ventricular hypertrophy. Once-daily furosemide is not an effective antihypertensive medication because of its short half-life; chlorthalidone is a much better diuretic for 24-hour blood pressure control and should be effective in stage 2 or 3 chronic kidney disease. Furthermore, amlodipine should not be given without an ARB or angiotensin-converting enzyme inhibitor in the setting of chronic kidney disease. (Fagard RH, Celis H, Thijs L, et al. Regression of left ventricular mass by antihypertensive treatment: a meta-analysis of randomized comparative studies. *Hypertension.* 2009;54: 1084-1091.)

5. A 59-year-old woman with difficult-to-control hypertension just had a nuclear medicine stress test as part of her annual executive physical elsewhere. However, because of her uncontrolled hypertension, the cardiologist considered renovascular hypertension and ordered a renal computed tomographic angiogram, which showed a 75% stenosis of the proximal left renal artery and a small plaque in her right renal artery. Both of her kidneys were normal in size and function. Her office blood pressure is now averaging 150/70 mm Hg on hydrochlorothiazide 25 mg daily, lisinopril 40 mg daily, metoprolol XL 25 mg daily, and diltiazem CD 360 mg daily. She has a low-density lipoprotein cholesterol level of 139 mg/dL but does not want to take a statin because she is concerned about potential side effects. She is overweight (body mass index of 29) and does not exercise regularly. The patient trusts your judgment and wants to know if she should undergo renal angioplasty for renovascular hypertension. What would you recommend?

A. Renal angiography and balloon angioplasty with stenting
B. Intensify her blood pressure by replacing hydrochlorothiazide with chlorthalidone, metoprolol with nebivolol, and diltiazem CD with amlodipine.
C. A low sodium diet, high in fruits and vegetables, combined with a regular moderately intense aerobic exercise regimen at her company's gym for 40 minutes per session, 4 to 5 days per week
D. Strongly encourage statin therapy.
E. B, C, and D

Answer: E Stenting is no more effective than intensive medical therapy for renal artery stenosis and is accompanied by significant complications. Because her kidneys are of equal and normal size on computed tomographic images, the unilateral renal artery stenosis is unlikely to be severe enough to cause substantial renal ischemia. The best approach to prevent or delay progression of her atherosclerotic renal artery stenosis is to encourage lifestyle recommendations, intensify her antihypertensive regimen with stronger and longer lasting medication, and start statin therapy. (Cooper CJ, Murphy TP, Cutlip DE, et al. Stenting and medical therapy for atherosclerotic renal-artery stenosis. *N Engl J Med* 2014;370:13-22; Wheatley K, Ives N, Gray R, et al. Revascularization versus medical therapy for renal-artery stenosis. *N Engl J Med.* 2009;361:1953-1962.)

68

PULMONARY HYPERTENSION

VALLERIE MCLAUGHLIN

DEFINITION

The normal pulmonary vasculature is a low-pressure system, with less than one tenth the resistance to flow observed in the systemic vasculature. Pulmonary hypertension refers to the hemodynamic state in which the pressure in the pulmonary artery is elevated above a mean of 25 mm Hg. A specific type of pulmonary hypertension, pulmonary arterial hypertension, also requires that the left-sided heart filling pressure (pulmonary capillary wedge pressure, left ventricular end-diastolic pressure, or left atrial pressure) be 15 mm Hg or less and that the calculated pulmonary vascular resistance be greater than 3 Wood units (Wood unit = [pulmonary artery pressure minus mean pulmonary capillary wedge pressure] divided by cardiac output).[1] The syndrome of pulmonary arterial hypertension (Table 68-1)[2] results when blood flow through the pulmonary circulation is restricted, thereby leading to pathologic increases in pulmonary vascular resistance and, ultimately, to right ventricular failure. Pulmonary hypertension may also be a consequence of many other chronic diseases, including left-sided heart failure (Chapter 58), a variety of parenchymal lung diseases, and thromboembolic disease (Chapter 98).

EPIDEMIOLOGY

Normal pulmonary blood pressure is 20/10 (mean, 15) mm Hg at rest at sea level, rising to 30/13 (mean, 20) mm Hg with mild exercise. Pressures rise with altitude, and at an altitude of about 15,000 feet, normal resting pulmonary artery pressures are about 38/14 (mean, 20) mm Hg. Pulmonary arterial systolic pressure rises gradually with age, and each increase of 10 mm Hg is associated with a 2.7-fold greater risk for mortality.

Idiopathic pulmonary arterial hypertension, formerly called primary pulmonary hypertension, is the prototype of group 1 pulmonary arterial hypertension. This disease affects women more than men in a 2 : 1 ratio. It may be manifested at any age, with a mean age at onset of 37 years. The prevalence of pulmonary arterial hypertension is between 15 and 26 per million persons. Heritable pulmonary arterial hypertension occurs in a familial context, most often (70%) due to a mutation in the bone morphogenetic protein receptor type 2. (See Pathobiology.)

Drug- and toxin-induced pulmonary arterial hypertension has been most clearly linked to anorexigens, including aminorex, fenfluramine, and dexfenfluramine. Although these agents are no longer used, observational studies link amphetamines, methamphetamines, and L-tryptophan to pulmonary arterial hypertension. The tyrosine kinase inhibitor dasatinib also has been associated with the development of pulmonary arterial hypertension.

One of the most common types of group 1 pulmonary arterial hypertension occurs in the setting of connective tissue diseases. For example, the prevalence of pulmonary arterial hypertension in patients with scleroderma (Chapter 267) is in the range of 7 to 12%. It is less common in systemic lupus erythematosus (Chapter 266), rheumatoid arthritis (Chapter 264), and other systemic vasculitides (Chapter 270). Pulmonary arterial hypertension is a rare but well-established complication of human immunodeficiency virus (HIV) infection (Chapter 391), and its prevalence of 0.5% in such patients has not changed with the widespread use of highly active antiretroviral therapy. Prospective hemodynamic studies show that 2 to 6% of patients with portal hypertension (Chapter 153) develop pulmonary arterial hypertension, although the reason for the association is not clear.

TABLE 68-1 CLINICAL CLASSIFICATION OF PULMONARY HYPERTENSION

GROUP 1
Pulmonary arterial hypertension
Idiopathic pulmonary arterial hypertension
Heritable pulmonary arterial hypertension
BMPR2
ALK1, endoglin, SMAD9, CAV1, KCNK3
Unknown
Drug- and toxin-induced pulmonary hypertension
Associated with
Connective tissue diseases
HIV infection
Portal hypertension
Congenital heart diseases
Schistosomiasis
Pulmonary veno-occlusive disease and/or pulmonary capillary hemangiomatosis
Persistent pulmonary hypertension of the newborn
GROUP 2
Pulmonary hypertension due to left heart disease
Systolic dysfunction
Diastolic dysfunction
Valvular disease
Congenital or acquired left heart inflow or outflow tract obstruction
GROUP 3
Pulmonary hypertension due to lung diseases and/or hypoxia
Chronic obstructive pulmonary disease
Interstitial lung disease
Other pulmonary diseases with mixed restrictive and obstructive pattern
Sleep-disordered breathing
Alveolar hypoventilation disorders
Chronic exposure to high altitude
Developmental lung diseases
Congenital diaphragmatic hernia
Bronchopulmonary dysplasia
GROUP 4
Chronic thromboembolic pulmonary hypertension
GROUP 5
Pulmonary hypertension with unclear multifactorial mechanisms
Hematologic disorders: chronic hemolytic anemias, myeloproliferative disorders, splenectomy
Systemic disorders: sarcoidosis, pulmonary Langerhans cell histiocytosis, lymphangioleiomyomatosis, neurofibromatosis, vasculitis
Metabolic disorders: glycogen storage disease, Gaucher disease, thyroid disorders
Others: segmental pulmonary arterial hypertension, tumoral obstruction, fibrosing mediastinitis, chronic renal failure

ALK1 = activin receptor-like kinase type 1; BMPR2 = bone morphogenetic protein receptor type 2; HIV = human immunodeficiency virus.
From Simonneau G, Gatzoulis MA, Adatia I, et al. Updated clinical classification of pulmonary hypertension. *J Am Coll Cardiol.* 2013;62:D34-D41.

A significant proportion of patients with untreated systemic-to-pulmonary shunts, commonly due to congenital heart disease (Chapter 69), develop pulmonary arterial hypertension. Persistent exposure of the pulmonary vasculature to increased blood flow and pressure leads to an elevated pulmonary vascular resistance. In some cases, Eisenmenger syndrome (Chapter 69), with a reversal of flow across the defect, results in right-to-left shunting. Pulmonary veno-occlusive disease (Chapter 98) and pulmonary capillary hemangiomatosis are rare disorders that directly affect the pulmonary vasculature. The presentation of each is often similar to pulmonary arterial hypertension, but the prognosis is particularly poor.

Pulmonary hypertension due to left-sided heart disease probably represents the most frequent cause of pulmonary hypertension seen in practice (group 2 patients). Left-sided ventricular (Chapter 58) or valvular (Chapter 75) disease may increase left atrial pressure, which then is transmitted back to the pulmonary vasculature. Often, the transpulmonary gradient and pulmonary vascular resistance are normal. In such cases, optimal treatment of the left-sided heart disease results in reduction of the left-sided heart filling pressures and, consequently, a reduction in the pulmonary artery pressures. On occasion, patients with left-sided heart disease have an elevation of pulmonary artery pressure greater than expected on the basis of the elevation of

left-sided heart filling pressures, with a transpulmonary gradient of more than 12 mm Hg and a pulmonary vascular resistance of more than 3 Wood units. This difference may be due to an increase in pulmonary artery vasomotor tone or pulmonary vascular remodeling in the setting of persistently elevated left-sided heart filling pressures.

Group 3 patients have pulmonary hypertension due to lung diseases or hypoxia. Any disorder that results in hypoxemia (e.g., chronic obstructive lung disease [Chapter 88], interstitial lung disease [Chapter 92], sleep-disordered breathing [Chapter 100]) may result in pulmonary hypertension, although the pressure elevation tends to be modest, with a mean pulmonary artery pressure of 25 to 35 mm Hg. Echocardiography-based observations have suggested that up to 80% of patients with chronic obstructive lung disease and idiopathic pulmonary fibrosis have elevated pulmonary artery pressures. In patients who have more advanced parenchymal lung disease and undergo evaluation for lung volume reduction surgery or lung transplantation, 40 to 50% have pulmonary hypertension at the time of right-sided heart catheterization. Most often, the elevations in pulmonary artery pressures are modest, but a small proportion of patients have more substantial elevations.

Group 4 patients have chronic thromboembolic pulmonary hypertension (Chapter 98), which must be differentiated from the other groups because the treatment is different. Approximately 4% of patients who have suffered an acute pulmonary embolism progress to development of chronic thromboembolic pulmonary hypertension. Approximately half of those ultimately diagnosed with chronic thromboembolic pulmonary hypertension do not have a known history of an acute pulmonary embolism. Chronic thromboembolic pulmonary hypertension occurs equally in both genders. All age groups can be affected, with a median age of 63 years.

PATHOBIOLOGY

The pathobiology of pulmonary arterial hypertension is complex and incompletely elucidated (E-Fig. 68-1). The pulmonary arterial hypertension phenotype is characterized by endothelial dysfunction, a decreased ratio of apoptosis to proliferation in pulmonary artery smooth muscle cells, and a thickened, disordered adventitia in which adventitial metalloproteases are excessively activated. The evolution of pulmonary vascular disease frequently originates with the interaction of a predisposing state and one or more inciting stimuli, a concept referred to as the multiple-hit hypothesis.

In group 1 pulmonary arterial hypertension, patients have a panvasculopathy predominantly affecting the small pulmonary arterioles. It is characterized by a variety of arterial abnormalities, including intimal hyperplasia, medial hypertrophy, adventitial proliferation, thrombosis in situ, varying degrees of inflammation, and plexiform lesions. An individual patient may manifest all or some of these lesions, and the distribution of the lesions may be diffuse or focal.

The genetic defect best characterized in heritable pulmonary arterial hypertension is that of the bone morphogenetic protein receptor type 2, a member of the transforming growth factor-β (TGF-β) signaling family. Mutations in activin receptor–like kinase type 1, or endoglin, have also been identified, usually in families with coexistent hereditary hemorrhagic telangiectasia. Less commonly, mutations in activin receptor–like kinase type 1, or endoglin, have been identified in patients with pulmonary arterial hypertension, predominantly with coexistent hereditary hemorrhagic telangiectasia (Chapter 173). Mutations in other genes (i.e., BMPR1B, caveolin-1, and SMAD9), all of which are involved in the TGF-β signaling pathway, are considerably less common. A novel channelopathy of KCNK3, which has been identified in familial and idiopathic cases of pulmonary arterial hypertension, is the first indication that the disease may involve factors apparently independent of the TGF-β signaling pathway.

The imbalance in the production or metabolism of vasoactive mediators in the pulmonary vasculature includes a reduction in prostacyclin and nitric oxide, which have vasodilator and antiproliferative properties, and an increase in thromboxane and endothelin, which are vasoconstrictors as well as mitogens. The reduction in nitric oxide synthase in pulmonary arterial hypertension diminishes nitric oxide and, subsequently, cyclic guanosine monophosphate production. Endothelin-1 is a potent vasoconstrictor and smooth muscle mitogen that may contribute to the development of pulmonary arterial hypertension. Prostacyclin synthase is reduced in pulmonary arterial hypertension, resulting in an inadequate production of prostacyclin, which is a vasodilator with potent antiproliferative effects. Other aberrations include those of the voltage-dependent potassium channels and serotonin pathways. Disorders of inflammatory and coagulation pathways have also been described.

Chronic changes in the pulmonary vasculature also occur as a result of other types of pulmonary hypertension. Chronic elevation of left-sided heart filling pressures causes a backward transmission of pressure to the pulmonary venous system and triggers vasoconstriction in the pulmonary arterial bed. On histologic evaluation, the veins are thickened abnormally, and a neointima is formed. As secondary features, medial hypertrophy and thickening of the neointima on the arterial side of the pulmonary circulation occur. These changes can be reversed with therapies that result in chronic reduction of left-sided heart filling pressures.

In parenchymal lung disease, changes in the distal pulmonary arterial vessels are related to hypoxia. Hypoxia induces muscularization of the distal vessels and medial hypertrophy of the more proximal vessels. Neither neointima formation nor the development of plexiform lesions is observed.

The pathologic process of chronic thromboembolic pulmonary hypertension is often distinct from idiopathic pulmonary arterial hypertension. The lesions are frequently more variable, with some arterial pathways that appear relatively unaffected and others that show recanalized vascular thromboses. However, the involvement of distal microvessels, particularly when thromboses have occurred in subsegmental arteries, can resemble idiopathic pulmonary arterial hypertension with the formation of plexiform lesions.

Pathophysiology

The normal pulmonary vasculature bed has a remarkable capacity to dilate and recruit unused vasculature to accommodate increases in pulmonary blood flow. In pulmonary hypertension, the pulmonary artery pressure and pulmonary vascular resistance are increased at rest and further increase with exertion. In response to this increased afterload, the normally very thin right ventricle hypertrophies and eventually dilates. Early in the process, the right ventricle may be capable of maintaining normal cardiac output at rest, although it may fail to augment cardiac output with exercise, thereby leading to exertional dyspnea. As the disease progresses, the right ventricular dysfunction may progress to the point that resting cardiac output is impaired. Right ventricular function is a major determinant of functional capacity and prognosis in pulmonary arterial hypertension. Although the left ventricle is not affected by pulmonary vascular disease itself, progressive right ventricular dilation can impair left ventricular filling and lead to mildly increased left-sided heart filling pressure. The pathophysiologic mechanism of pulmonary hypertension related to left-sided heart and lung disease is further complicated by those underlying disorders.

The two most frequent mechanisms of death are progressive right ventricular failure and sudden death. Right ventricular failure, as evidenced by elevated jugular venous pressure, lower extremity edema, and occasionally ascites, may also be accompanied by evidence of poor forward flow due to inadequate filling of the left ventricle. Hypotension, hypoperfusion, and renal insufficiency may result. Other potential causes of death include pneumonia, sepsis, and pulmonary embolism.

CLINICAL MANIFESTATIONS

History

Dyspnea, which is the most common symptom of pulmonary hypertension, initially may be attributed to underlying disorders such as heart failure or obstructive lung disease, but the dyspnea of pulmonary hypertension typically is insidious in progression and reproducible. Dyspnea is classified by the World Health Organization (WHO) system, which is similar to the New York Heart Association classification system for angina and heart failure (see Table 51-5), and may progress to dyspnea at rest.

Other common symptoms of pulmonary hypertension include fatigue, lightheadedness, chest pain (Chapter 51), and palpitations (Chapters 51 and 62). Syncope (Chapter 62), which is an ominous finding, is often exertional in nature; it signifies the inability of the right ventricle to augment cardiac output as needed for physical activity. Symptoms of right-sided heart failure, including edema and ascites, signify advanced disease.

The nonspecific symptoms of pulmonary hypertension often explain its delayed recognition. In various reviews, the delay from onset of symptoms to diagnosis can be as long as 2 years.[3]

Patients often have symptoms associated with their underlying disease, which typically is far advanced by the time pulmonary arterial hypertension develops. For example, patients with pulmonary hypertension associated with left-sided heart disease (group 2) often have paroxysmal nocturnal dyspnea and orthopnea. Patients with pulmonary hypertension related to hypoxic lung disease (group 3) may have cough, sputum production, or wheezing. Clinical symptoms of chronic thromboembolic pulmonary

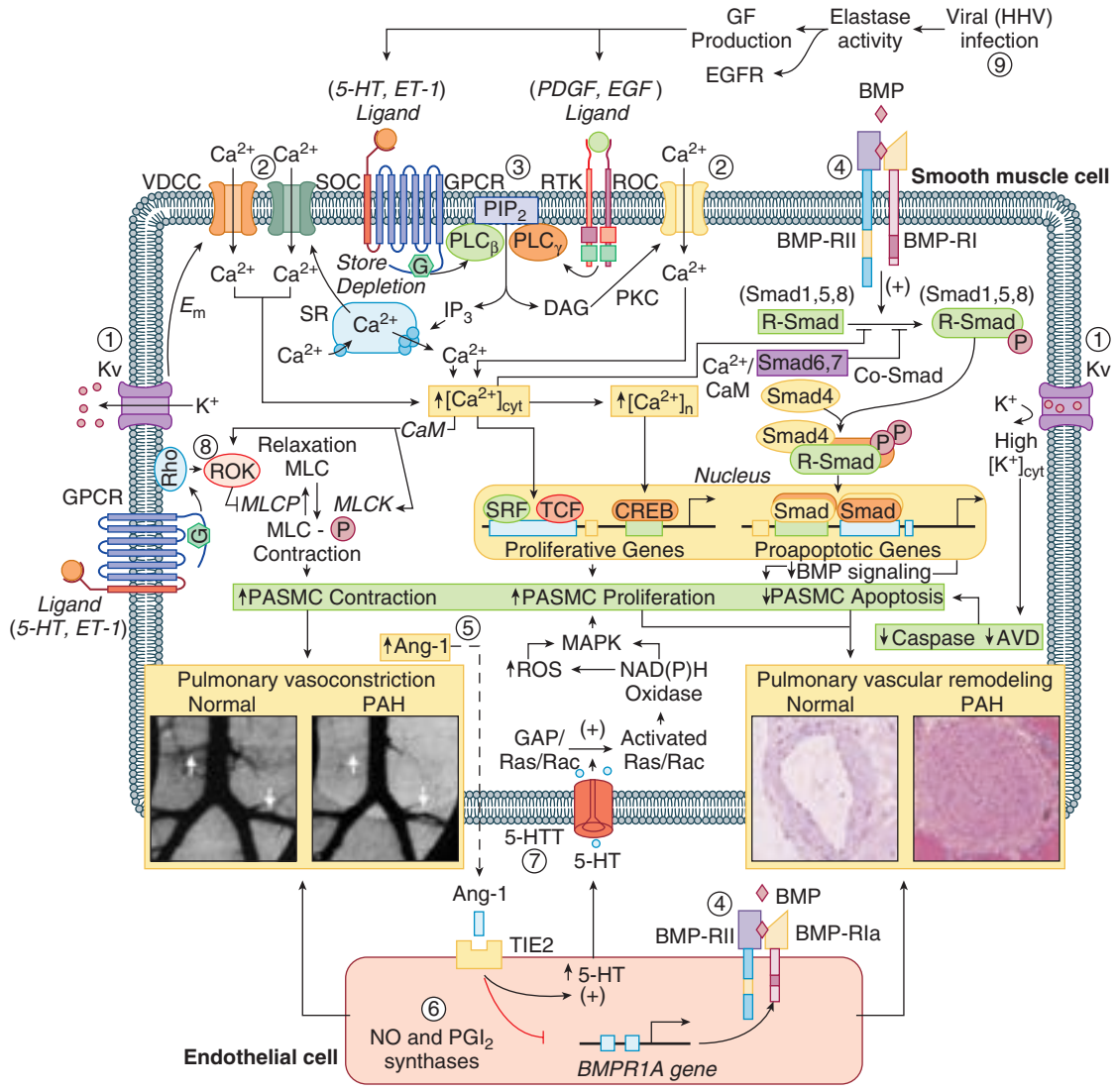

E-FIGURE 68-1. Potential mechanisms involved in the development of pulmonary arterial hypertension. Ang = angiopoietin; AVD = apoptotic volume decrease; BMP = bone morphogenetic protein; BMPR = bone morphogenetic protein receptor; CaM = calmodulin; CREB = cAMP response element–binding protein; DAG = diacylglycerol; E_m = membrane potential; EGF = epidermal growth factor; EGFR = epidermal growth factor receptor; ET = endothelin; GAP = GTPase-activating protein; GPCR = G protein–coupled receptor; HHV = human herpesvirus; HT = hydroxytryptamine (serotonin); HTT = hydroxytryptamine (serotonin) transporter; IP_3 = inositol 1,4,5-trisphosphate; Kv = voltage-gated K^+; MAPK = mitogen-activated protein kinase; MLC = myosin light chain; MLCK = myosin light chain kinase; NA(D)PH = nicotinamide adenine dinucleotide phosphate; NO = nitric oxide; PASMC = pulmonary artery smooth muscle cell; PDGF = platelet-derived growth factor; PGI_2 = prostacyclin; PKC = protein kinase C; PLC = phospholipase C; ROC = receptor-operated Ca^{2+} channels; ROS = reactive oxygen species; RTK = receptor tyrosine kinase; SR = sarcoplasmic reticulum; SRF = serum response factor; TCF = T-cell factor; TIE = endothelial-specific tyrosine kinase; VDCC = voltage-dependent calcium channel. (Modified from Morrell NW, Adnot S, Archer SL, et al. Cellular and molecular basis of pulmonary hypertension. *J Am Coll Cardiol.* 2009;54:S20-S31.)

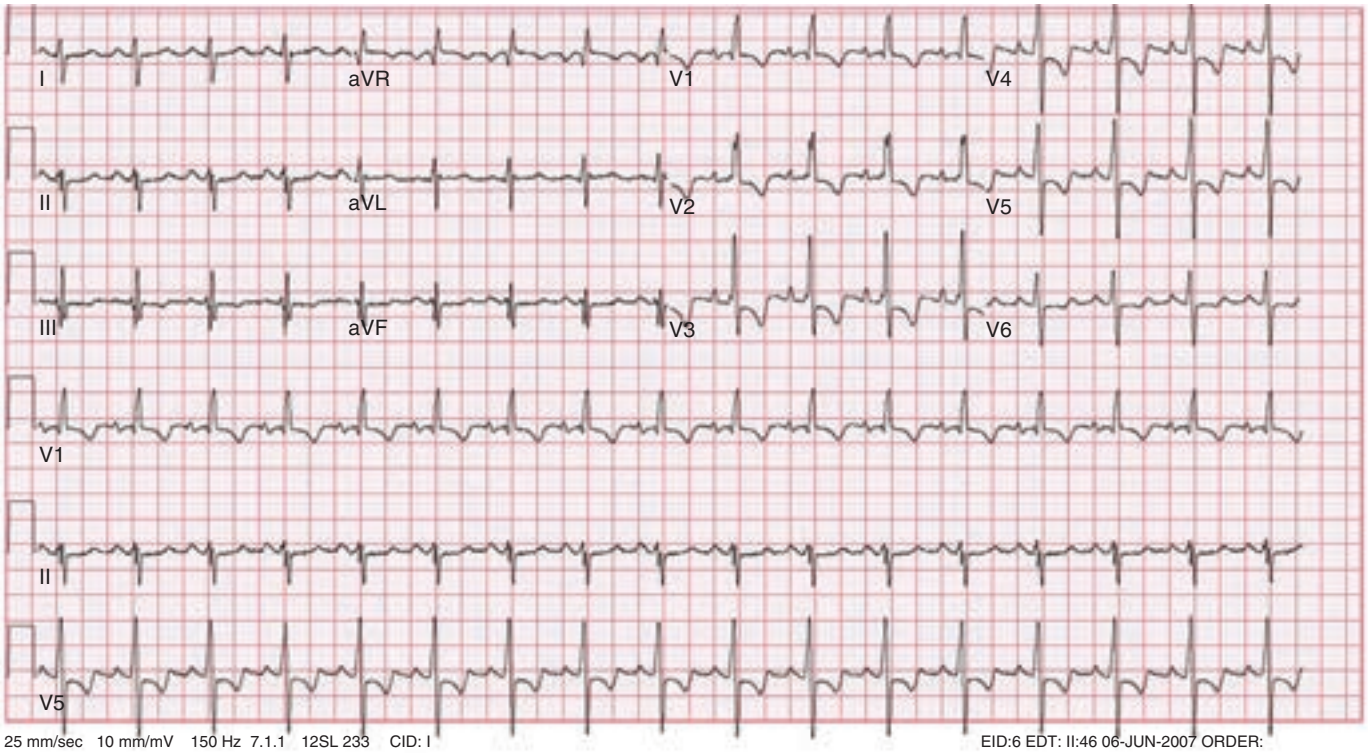

FIGURE 68-1. **Electrocardiogram demonstrating sinus rhythm, right axis deviation, and right ventricular hypertrophy with a strain pattern.**

hypertension resemble those of idiopathic pulmonary arterial hypertension, except that edema and hemoptysis occur more often in chronic thromboembolic pulmonary hypertension, whereas syncope is more common in idiopathic pulmonary arterial hypertension.

Physical Examination

Distention of jugular veins (see Fig. 51-3) may signify right ventricular failure, and prominent *v* waves (see Fig. 51-4) may be a result of tricuspid regurgitation. The amplitude of the carotid upstroke may give some insight into the cardiac output. The classic physical examination finding in pulmonary hypertension is a loud pulmonic component to the second heart sound, which reflects high pulmonary pressures that increase the force of the pulmonic valve closure. Palpation of the sternum often reveals a parasternal lift as the hypertrophied, pressure-overloaded right ventricle obliterates the retrosternal air space. A right ventricular fourth heart sound reflects diastolic filling of the hypertrophied, noncompliant right ventricle, akin to the left-sided fourth heart sound in a patient with systemic hypertension and left ventricular hypertrophy. The murmur of tricuspid regurgitation, which is holosystolic, located at the left lower sternal border, and augments with inspiration, is common in patients with moderate to severe pulmonary hypertension. Other findings on auscultation may include an early systolic click and the murmur of pulmonic regurgitation. A right ventricular third heart sound often signifies advanced disease and right-sided heart failure. Other signs consistent with right ventricular failure include hepatomegaly, peripheral edema (see Fig. 51-7), ascites, hypotension, diminished pulse pressure, and cool extremities.

Other physical examination findings may give some insight into the etiology of the pulmonary hypertension. For example, central cyanosis and clubbing may suggest an intracardiac shunt and Eisenmenger physiology. Sclerodactyly, telangiectasias (see Fig. 267-3), arthritis, Raynaud phenomenon (see Fig. 80-7), and rashes may increase the suspicion of an underlying connective tissue disease. Splenomegaly, spider angioma, palmar erythema (see Fig. 146-2), icterus (see Fig. 146-1), and caput medusae may suggest portal hypertension as an etiology. Signs of left-sided heart disease, such as pulmonary congestion, left-sided third heart sound, or findings of mitral or aortic valve disease on auscultation, may signify pulmonary hypertension as a result of left-sided heart disease. Fine rales, accessory muscle use, wheezing, protracted expiration, and productive cough may denote group 3 pulmonary hypertension as a result of hypoxic lung disease. Pulmonary vascular bruits suggest chronic thromboembolic pulmonary hypertension.

DIAGNOSIS

Initial assessments include an electrocardiogram and chest radiograph.[4] The electrocardiogram may show right axis deviation, right ventricular enlargement, right atrial enlargement, and ST and T wave changes across the anterior precordium that reflect right ventricular strain (Fig. 68-1). The chest radiograph may demonstrate enlarged proximal pulmonary arteries (see Fig. 56-6) with peripheral tapering or pruning of the pulmonary vasculature (Fig. 68-2A). The lateral radiograph may reveal the reduction in retrosternal air space as a result of right ventricular enlargement (Fig. 68-2B).

If, on the basis of history, physical examination, electrocardiogram, and chest radiograph, there is a reasonable suspicion for pulmonary hypertension, a series of diagnostic evaluations should follow (Fig. 68-3), usually beginning with an echocardiogram and with further testing guided by the patient's subtype. The echocardiogram gives insight not only into the presence of pulmonary hypertension but also into the presence of common disorders of the left side of the heart that may result in pulmonary hypertension. Two-dimensional echocardiographic findings reflective of elevated pulmonary artery pressures include right atrial enlargement, right ventricular enlargement, flattening of the intraventricular septum, and underfilled left ventricle (Fig. 68-4). The right ventricular systolic pressure may be estimated on the basis of the velocity of the tricuspid regurgitant jet by the modified Bernoulli equation (Chapter 55) (Fig. 68-5), although a reliable estimate of right ventricular systolic pressure is not always obtainable, and this measurement is prone to error, particularly in patients with parenchymal lung disease.

The echocardiogram is also useful to assess for left-sided heart causes of pulmonary hypertension, such as systolic dysfunction, diastolic dysfunction, and valvular heart disease. On occasion, a previously unknown congenital heart defect is discovered during this evaluation. In approximately 25% of patients, a previously trivial patent foramen ovale may shunt blood from the right atrium to the left atrium because of the high pulmonary vascular resistance and thereby worsen systemic oxygenation.

In a patient with unexplained dyspnea and evidence of pulmonary hypertension on echocardiography, chronic thromboembolic pulmonary hypertension must be excluded.[5] The study of choice for this assessment is the ventilation-perfusion (V/Q) scan (see Fig. 98-4), which often shows multiple perfusion defects that are not matched on ventilation. Although spiral computed tomography is excellent for the assessment of acute pulmonary embolus, it sometimes fails to detect surgically accessible chronic thromboembolic disease. If either of these studies detects an abnormality, further evaluation with pulmonary angiography may be required to determine

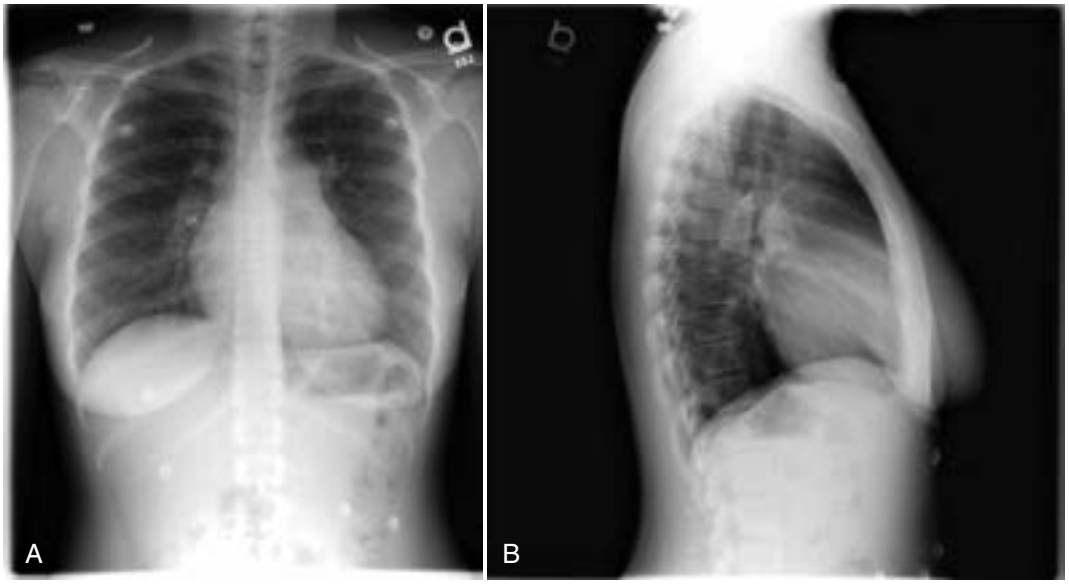

FIGURE 68-2. **Posterior-anterior (A) and lateral (B) chest radiographs demonstrating enlarged proximal pulmonary arteries and right ventricular enlargement.**

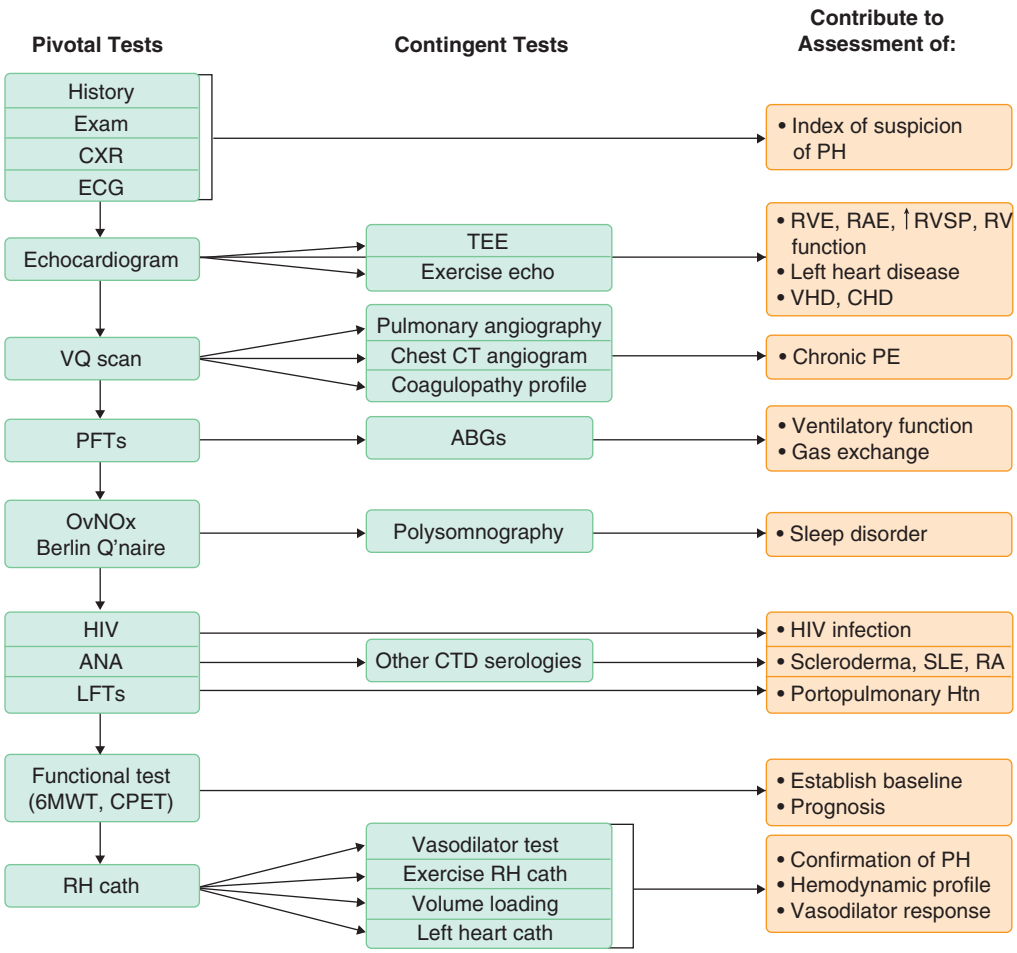

FIGURE 68-3. **Diagnostic approach to pulmonary arterial hypertension.** Because the suspicion of pulmonary arterial hypertension (PH) may arise in various ways, the sequence of tests may vary. However, the diagnosis of PH requires that certain data support a specific diagnosis. In addition, the diagnosis of idiopathic pulmonary arterial hypertension (IPAH) is one of excluding all other reasonable possibilities. *Pivotal tests* are those that are essential to establishing a diagnosis of any type of PH by either identification of criteria of associated disease or exclusion of diagnoses other than IPAH. All pivotal tests are required for a definitive diagnosis and baseline characterization. An abnormality of one assessment (such as obstructive pulmonary disease on pulmonary function tests) does not preclude that another abnormality (chronic thromboembolic disease on V/Q scan and pulmonary angiogram) is contributing or predominant. *Contingent tests* are recommended to elucidate or to confirm results of the pivotal tests and need be performed only in the appropriate clinical context. The *combination* of pivotal and appropriate contingent tests contributes to assessment of the differential diagnoses in the right-hand column. Definitive diagnosis may require additional specific evaluations not necessarily included in this general guideline. ABGs = arterial blood gases; ANA = antinuclear antibody serology; CHD = congenital heart disease; CPET = cardiopulmonary exercise test; CT = computed tomography; CTD = connective tissue disease; CXR = chest x-ray; ECG = electrocardiogram; HIV = human immunodeficiency virus screening; Htn = hypertension; LFTs = liver function tests; 6MWT = 6-minute walk test; OvNOx = overnight oximetry; PE = pulmonary embolism; PFTs = pulmonary function tests; RA = rheumatoid arthritis; RAE = right atrial enlargement; RH cath = right-sided heart catheterization; RV, right ventricle; RVE = right ventricular enlargement; RVSP = right ventricular systolic pressure; SLE = systemic lupus erythematosus; TEE = transesophageal echocardiography; VHD = valvular heart disease; V/Q scan = lung ventilation-perfusion scintigram. (From McLaughlin VV, Archer SL, Badesch DB, et al. ACCF/AHA 2009 expert consensus document on pulmonary hypertension: a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents and the American Heart Association developed in collaboration with the American College of Chest Physicians; American Thoracic Society, Inc.; and the Pulmonary Hypertension Association. *J Am Coll Cardiol.* 2009;53:1573-1619.)

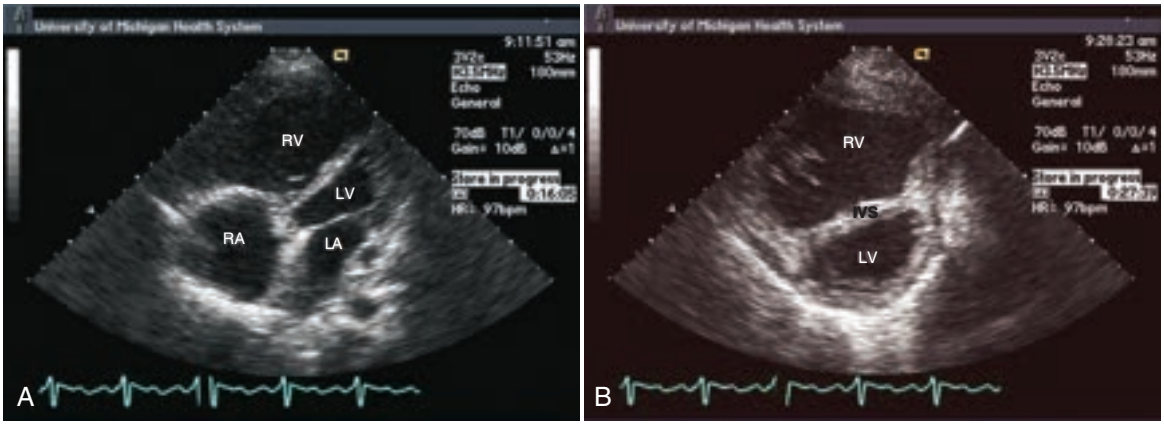

FIGURE 68-4. Echocardiographic images of the heart. A, Four-chamber view. Right atrial (RA) enlargement, right ventricular (RV) enlargement. The left atrium (LA) and left ventricle (LV) are small and underfilled. **B,** Short axis view. RV enlargement is present. Flattening of the intraventricular septum (IVS) results from pressure and volume overload of the RV.

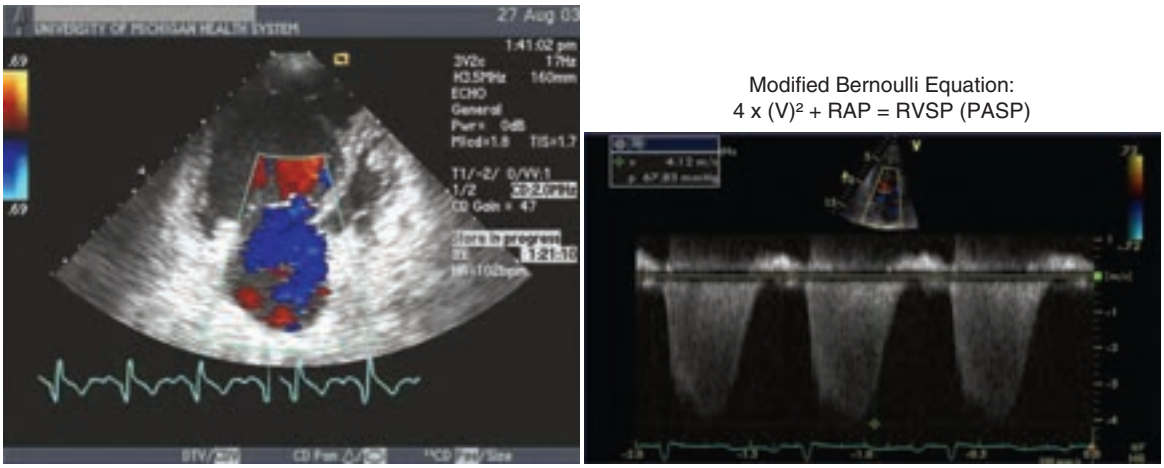

FIGURE 68-5. Calculation of estimated pulmonary artery pressure based on the velocity of the tricuspid regurgitant jet. RAP = right atrial pressure; RVSP = right ventricular systolic pressure; PASP = pulmonary artery systolic pressure; V = tricuspid jet velocity (m/sec).

whether chronic thrombolic disease is the diagnosis and, if so, whether it is surgically accessible.

Pulmonary function tests in patients with pulmonary arterial hypertension may show mild restrictive disease and a mildly reduced diffusing capacity for carbon monoxide. Patients with the scleroderma spectrum of diseases tend to have more substantial reductions in the diffusing capacity for carbon monoxide, which may even precede the development of pulmonary hypertension. Pulmonary function tests may disclose evidence of obstructive or restrictive lung disease; further evaluation with chest computed tomography may be necessary. Overnight oximetry is useful to screen for obstructive sleep apnea (Chapter 100). If indicated, formal polysomnography may be required.

Recommended serologic testing, given the known associations, includes an antinuclear antibody test and HIV serology in addition to liver function tests to assess for chronic liver disease. A study of functional capacity, most commonly the 6-minute walk test, is useful to assess the severity of disease, to determine the potential need for oxygen, and to establish a baseline against which to assess subsequent changes in exercise capacity as a result of medical interventions. The high prevalence of pulmonary arterial hypertension in patients with scleroderma serves as an opportunity for screening of this high-risk population with echocardiography to make an early diagnosis.[6]

If pulmonary arterial hypertension is suspected on the basis of the noninvasive evaluation, the diagnosis must be confirmed with a right-sided heart catheterization that measures right atrial pressure, right ventricular pressure, pulmonary artery (systolic, diastolic, and mean) pressures, pulmonary arterial wedge pressure (reflective of left ventricular end-diastolic pressure or left atrial pressure), cardiac output and index, heart rate, systemic blood pressure, and oxygen saturations in the superior vena cava, inferior vena cava, pulmonary artery, and a systemic artery. From this information, pulmonary vascular resistance and systemic vascular resistance may be calculated, right ventricular performance can be ensured, and an intracardiac or intrapulmonary shunt can be confirmed or excluded.

In patients with left-sided heart or parenchymal lung disease, however, optimal management of the underlying condition is often undertaken before right-sided heart catheterization is considered. Patients with suspected chronic thromboembolic disease often undergo both right-sided heart catheterization and pulmonary angiography (Chapter 98) to assess surgical candidacy and operative risk.

Measurement of the wedge pressure, a surrogate for left atrial pressure in the absence of pulmonary vein obstruction, is useful to exclude pulmonary hypertension caused by left-sided heart disease or, in rare cases, pulmonary veno-occlusive disease. If an optimal wedge pressure tracing cannot be obtained, or if there is any question about the accuracy of the wedge pressure tracing, a left ventricular end-diastolic pressure should be obtained.

Acute vasodilator testing is often performed at the time of the initial right-sided heart catheterization, not only for its prognostic implications but also to identify patients who might be candidates for therapy with calcium-channel blockers. Although the data regarding vasodilator testing and treatment with calcium-channel blockers are largely restricted to patients with idiopathic pulmonary arterial hypertension, vasodilator testing is often performed in patients with other types of pulmonary arterial hypertension. However, acute vasodilator testing is not indicated and may be harmful in patients with significantly elevated left-sided heart filling pressures because pulmonary edema may ensue.

The three agents most commonly used for acute vasodilator testing in the cardiac catheterization laboratory are inhaled nitric oxide, intravenous epoprostenol, and intravenous adenosine. A positive response to an acute vasodilator is a decrease in mean pulmonary artery pressure by at least 10 mm Hg to a mean pulmonary artery pressure of less than 40 mm Hg, without a decrease in cardiac output. If a patient meets these criteria, it is reasonable to administer a trial of oral calcium-channel blockers.

Unfortunately, adherence to the published algorithms for the diagnosis of pulmonary arterial hypertension is dismal. Studies that mistakenly remain

unperformed include the V/Q scan (57%), HIV serology (29%), and serologies for connective tissue diseases (50%).[7] Ten percent of patients are given the diagnosis of pulmonary arterial hypertension without a right-sided heart catheterization confirmation, and only a minority of patients treated with calcium-channel blockers fulfilled the criteria for likely being an acute responder.

TREATMENT

General Measures

Basic counseling and education are important components in the care of patients with pulmonary arterial hypertension. Patients are encouraged to engage in low-level graded aerobic exercise, such as walking, as tolerated, and to enroll in an intensive pulmonary rehabilitation program. Patients should avoid heavy physical exertion or isometric exercise, both of which may provoke exertional syncope. Exposure to high altitudes may contribute to hypoxic pulmonary vasoconstriction and may not be well tolerated. A sodium-restricted diet (<2400 mg/day) is advised and is particularly important to manage volume status in patients with right ventricular failure. Routine immunizations, such as those against influenza and pneumococcal pneumonia (Chapter 18), are advised. Because hypoxia is a potent pulmonary vasoconstrictor, supplemental oxygen is recommended to maintain saturations above 92% at rest, with exertion, and during sleep. In patients with intracardiac shunting and Eisenmenger physiology (Chapter 69), this goal may not be possible.

Women with pulmonary arterial hypertension have been advised to avoid pregnancy because the hemodynamic fluctuations of pregnancy, labor, delivery, and the postpartum period are potentially life-threatening. In a recent series of 26 women whose pulmonary arterial hypertension was well controlled and whose pregnancies were managed at highly specialized centers, three died, another developed refractory right-sided heart failure that required heart-lung transplantation, two had spontaneous abortions, and six had induced abortions.[8] Overall, 62% of the pregnancies resulted in a healthy baby without maternal complications. Current guidelines continue to recommend that pregnancy be avoided or terminated early in women with pulmonary arterial hypertension, although referral to a specialized center can be considered before termination. Women of childbearing potential should be counseled on contraception options at the time of diagnosis.

Background Therapy

On the basis of uncontrolled observational series in patients with primarily idiopathic pulmonary arterial hypertension, consensus recommendations advocate the use of warfarin titrated to an international normalized ratio of 1.5 to 2.5 in patients with idiopathic pulmonary arterial hypertension. Diuretics (e.g., furosemide, initiated at 20 mg and titrated as needed) are indicated to manage right ventricular volume overload; in some patients, intravenous diuretics may be necessary. Serum electrolytes and renal function must be closely monitored. Digoxin, 0.125 to 0.25 mg/day, is rarely used in patients with right ventricular failure and a low cardiac output and in patients with atrial arrhythmias, despite the paucity of data; if the patient experiences any evidence of digoxin toxicity, the drug should be discontinued because of the unfavorable risk-to-benefit ratio.

Vasodilator Therapy

Treatment of pulmonary arterial hypertension has evolved considerably, in part because of advances in understanding of the disease process and the availability of agents that target known pathobiologic derangements (Fig. 68-6).

Calcium-Channel Blockers

Approximately 7% of adult patients with idiopathic pulmonary arterial hypertension have a favorable response to acute vasodilator testing and excellent prognosis with calcium-channel blockers. Long-acting nifedipine (90 to 180 mg daily), diltiazem (360 to 720 mg daily), and amlodipine (10 to 20 mg daily) are the most commonly used calcium-channel blockers. Because of its potential negative inotropic effects, verapamil should be avoided. Patients must be observed closely for both the safety and efficacy of this therapy. If a patient who meets the definition of an acute response does not improve to WHO functional class I or II with calcium-channel blocker therapy, the patient should not be considered a chronic responder; alternative or additional pulmonary arterial hypertension–specific therapy should be instituted.

Targeted Therapies

In clinical trials, intravenous epoprostenol improves functional class, exercise endurance, hemodynamics, and survival in patients with idiopathic pulmonary arterial hypertension, and it also improves exercise tolerance and hemodynamics in patients with pulmonary arterial hypertension related to the scleroderma spectrum of diseases. Uncontrolled studies have also reported favorable effects with intravenous epoprostenol in patients with numerous forms of associated pulmonary arterial hypertension. Observational series suggest a long-term survival benefit with intravenous epoprostenol compared with historical controls. Epoprostenol must be delivered by continuous intravenous infusion, commonly initiated in the hospital at a dose of 2 ng/kg/minute, with the dose titrated up on the basis of symptoms of pulmonary arterial hypertension and side effects of the therapy. Each patient must learn the techniques of sterile preparation of the medication, operation of the ambulatory infusion pump, and care of the central venous catheter. Although dosing must be highly individualized, maintenance doses in the range of 25 to 40 ng/kg/minute are typically needed for patients receiving monotherapy. Common side effects include headache, jaw pain, flushing, nausea, diarrhea, rash, and musculoskeletal pain. Infections and infusion interruption can be life-threatening.

Subcutaneous treprostinil can provide a modest but statistically significant improvement in exercise tolerance. The main limitation of this therapy is pain and erythema at the site of the subcutaneous infusion, a complication that occurs in 85% of patients. Oral treprostinil (starting at 0.25 mg with meals twice daily and increased to as high as 12 mg twice daily as tolerated) also is modestly effective as monotherapy. For subcutaneous or oral therapy, other prostanoid-type side effects, including headache, diarrhea, rash, and nausea, also occur. Subcutaneous treprostinil is often started in the home, with the dose titrated up on the basis of symptoms of pulmonary arterial hypertension and drug side effects. Treprostinil is less potent than epoprostenol, and higher doses are required to achieve the desired efficacy. Inhaled treprostinil four times daily and inhaled iloprost six to nine times daily also are effective for improving exercise capacity. However, cough is an additional side effect with this method of administration.

The endothelin receptor antagonist bosentan (initiated orally at 62.5 mg twice daily and titrated up to 125 mg twice daily after 1 month) improves hemodynamics, exercise capacity, and the clinical course of pulmonary arterial hypertension. Liver enzymes must be monitored on a monthly basis; the dose should be reduced if liver enzymes rise to more than three to five times the upper limits of normal and discontinued if they rise to five times the upper limit of normal. Ambrisentan (administered orally at doses of either 5 mg or 10 mg once daily) has similar benefits. Other side effects include lower extremity edema, headache, and nasal congestion. Macitentan, a dual endothelin-receptor antagonist, significantly reduces the composite end point of death, atrial septostomy, lung transplantation, initiation of treatment with parenteral prostanoids, or worsening pulmonary arterial hypertension by 30% when it is given as 3 mg daily and by 46% if it is given as 10 mg daily. The most frequent adverse events are headache, nasopharyngitis, and anemia, without any increased rate of peripheral edema or elevated liver enzymes.

Riociguat is a first in class soluble guanylate cyclase stimulator, which directly stimulates soluble guanylate cyclase independent of nitric oxide and increases the sensitivity of soluble guanylate cyclase to nitric oxide. In a randomized controlled trial that included some patients who previously had been treated with endothelin receptor antagonists or nonparenteral prostanoids, riociguat (at a dose of 1.0 mg up to 2.5 mg three times daily) significantly improved the primary end point of 6-minute walking distance as well as pulmonary vascular resistance, brain natriuretic peptide levels, functional class, and time to clinical worsening. The most common adverse events included headache, dyspepsia, peripheral edema, and hypotension. Riociguat should not be used concurrently with phosphodiesterase type 5 inhibitors.

The chronic administration of inhaled nitric oxide is cumbersome and not clinically useful. However, the phosphodiesterase type 5 antagonists sildenafil and tadalafil are effective and useful for pulmonary arterial hypertension. Sildenafil is approved at a dose of 20 mg three times daily and tadalafil at a dose of 40 mg once daily. The most common side effects of the inhibitors are headache, flushing, dyspepsia, and epistaxis.

Given the availability of therapies that target different pathologic processes, combination therapy is an attractive theoretical option in pulmonary arterial hypertension. Emerging data support the incremental benefit of combining more than one targeted therapy under careful observation, usually in a specialized center.

Invasive Therapies

Despite advances in medical therapies for pulmonary arterial hypertension, many patients experience progressive functional decline, largely related to worsening right-sided heart failure. In carefully selected patients, atrial septostomy may improve symptoms. Atrial septostomy creates a right-to-left interatrial shunt, thereby decreasing right-sided heart filling pressures and improving right-sided heart function and left-sided heart filling. Although the right-to-left shunting decreases systemic arterial oxygen saturation, it is anticipated that the improvement in cardiac output will result in overall augmentation in systemic oxygen delivery. Contraindications to performing atrial septostomy include severe right ventricular failure on cardiorespiratory support, mean right atrial pressure of more than 20 mm Hg, pulmonary vascular resistance index of more than 55 U/m^2, resting oxygen saturation of less than 90% on room air, and left ventricular end-diastolic pressure of more than 18 mm Hg. Because of the high morbidity and mortality associated with this procedure, it should be performed only by experienced operators in specialized centers.

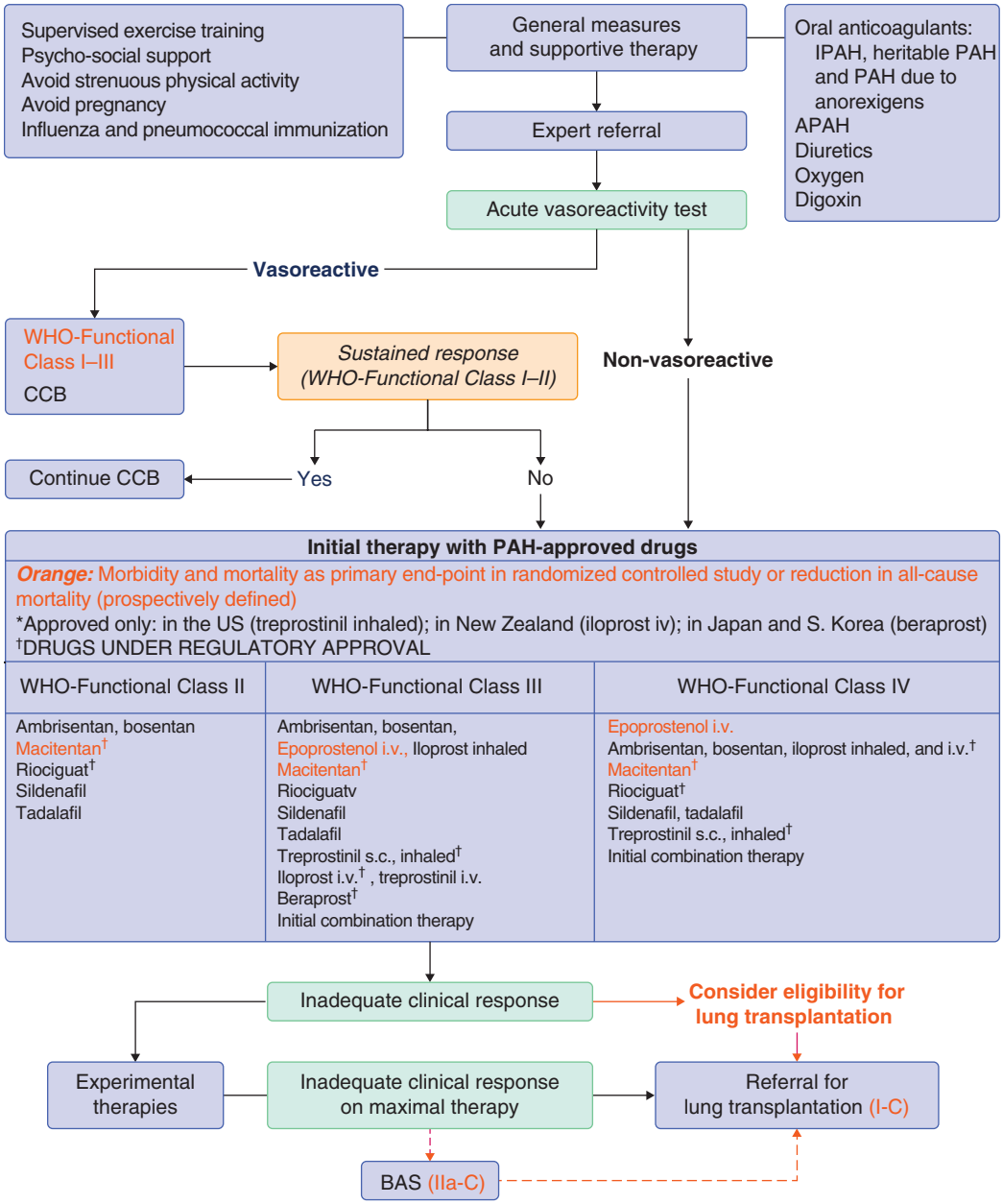

FIGURE 68-6. Pulmonary hypertension evidence-based treatment algorithm. APAH = associated pulmonary arterial hypertension; BAS = balloon atrial septostomy; CCB = calcium-channel blockers; IPAH = idiopathic pulmonary arterial hypertension; i.v. = intravenous; PAH = pulmonary arterial hypertension; s.c. = subcutaneous; WHO = World Health Organization. (Modified from Galie N, Corris PA, Frost A, et al. Updated treatment algorithm of pulmonary arterial hypertension. *J Am Coll Cardiol.* 2013;62:D60-D72.)

Bilateral lung (Chapter 101) or heart-lung (Chapter 82) transplantation is the final option for selected patients with pulmonary arterial hypertension when medical therapy fails. The 1-, 3-, 5-, and 10-year survival rates are 66%, 57%, 47%, and 27%, respectively, in patients with idiopathic pulmonary arterial hypertension who undergo transplantation. Transplantation as a potential therapeutic option should be discussed with selected patients at the time of diagnosis, although timing of referral is challenging. The International Society for Heart and Lung Transplantation recommends that patients with pulmonary arterial hypertension be referred for transplantation evaluation if they have persistent functional class III or IV symptoms despite treatment with pulmonary arterial hypertension–specific therapies, including prostanoids. Patients who are otherwise good transplant candidates should be referred when they have an unacceptable response to medical therapies.

Special Populations

Group 2: Pulmonary Venous Hypertension

No specific therapy is currently approved for the treatment of pulmonary venous hypertension that causes secondary pulmonary arterial hypertension. For example, a multicenter randomized controlled trial found no difference in peak oxygen consumption or secondary clinical end points when patients with stable heart failure and a preserved ejection fraction were treated with sildenafil compared with placebo.

Group 3: Primary Lung Disease

For these disorders, treatment of underlying lung disease is indicated.

Group 4: Chronic Thromboembolic Hypertension

For patients with chronic thromboembolic pulmonary hypertension and a significant clot burden, pulmonary endarterectomy (E-Fig. 68-2) is the treatment of choice and is a potentially curative procedure. Riociguat (a soluble guanylate cyclase stimulator at a dose of 1.0 up to 2.5 mg three times daily) improves 6-minute walking distance, pulmonary vascular resistance, NT-pro–brain natriuretic peptide level, and functional class in patients with either inoperable chronic thromboembolic pulmonary hypertension or persistent pulmonary hypertension after pulmonary endarterectomy. Warfarin anticoagulation is also recommended.

Group 5: Other Causes

Pulmonary hypertension also consists of several forms for which the etiology is unclear or multifactorial. Among these conditions are a number of

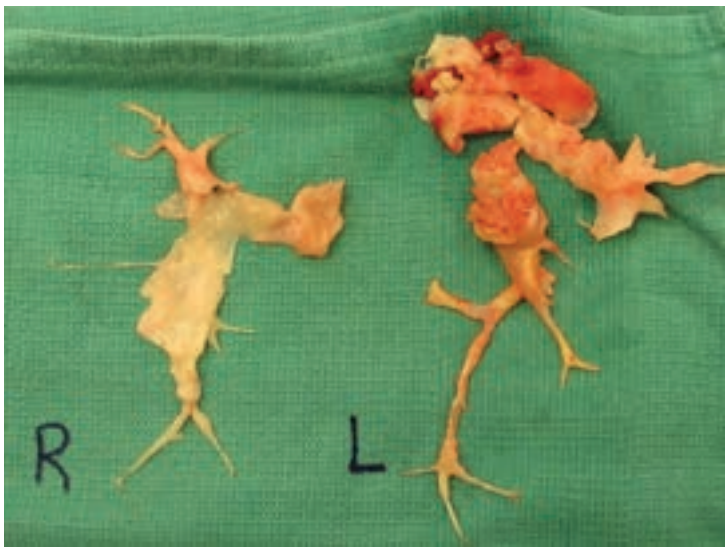

E-FIGURE 68-2. Chronic thromboembolic material endarterectomized from a patient. (Image courtesy Dr. Jonathan Haft.)

TABLE 68-2 LONGITUDINAL EVALUATION OF THE PATIENT WITH PULMONARY ARTERIAL HYPERTENSION

CLINICAL PARAMETER	STABLE SYMPTOMS, WELL COMPENSATED	UNSTABLE IN SYMPTOMS OR DECOMPENSATION
DEFINITION		
Physical examination	No evidence of right-sided heart failure	Signs of right-sided heart failure
WHO functional class	I/II	IV
6-minute walk distance	>400 m	<300 m
Echocardiography	RV size/function normal	RV enlargement/ dysfunction
Hemodynamics	RA pressure normal; CI normal	RA pressure high; CI low
BNP	Nearly normal, remaining stable or decreasing	Elevated or increasing
EVALUATION AND TREATMENT		
Frequency of visits	Every 3 to 6 months	Every 1 to 3 months
Functional class assessment	Every visit	Every visit
6-minute walk test	Every visit	Every visit
Echocardiography	Every 12 months or center dependent	Every 6 to 12 months or center dependent
BNP	Center dependent	Center dependent
RHC	Clinical deterioration and center dependent	Every 6 to 12 months or with clinical deterioration
Oral therapy	Treatment	IV prostacyclin or combination treatment

BNP = brain natriuretic peptide; CI = cardiac index; RA = right atrial; RHC = right-sided heart catheterization; RV = right ventricle; WHO = World Health Organization.

TABLE 68-3 PULMONARY ARTERIAL HYPERTENSION: DETERMINANTS OF PROGNOSIS*

DETERMINANTS OF RISK	LOWER RISK (GOOD PROGNOSIS)	HIGHER RISK (POOR PROGNOSIS)
Clinical evidence of RV failure	No	Yes
Progression of symptoms	Gradual	Rapid
WHO class†	II, III	IV
6-minute walk test‡	Longer (>400 m)	Shorter (<300 m)
Cardiopulmonary exercise testing	Peak Vo_2 > 10.4 mL/kg/min	Peak Vo_2 < 10.4 mL/kg/min
Echocardiography	Minimal RV dysfunction	Pericardial effusion, significant RV enlargement or dysfunction, right atrial enlargement
Hemodynamics	RA pressure < 10 mm Hg CI > 2.5 L/min/m^2	RA pressure > 20 mm Hg CI < 2.0 L/min/m^2
BNP level§	Minimally elevated	Significantly elevated

*Most data available pertain to idiopathic pulmonary arterial hypertension. Few data are available for other forms of pulmonary arterial hypertension. One should not rely on any single factor to make risk predictions.
†WHO class is the functional classification for pulmonary arterial hypertension and is similar to the New York Heart Association functional class, except that patients with syncope are defined as class IV.
‡The 6-minute walk test is also influenced by age, gender, and height.
§Because there are currently limited data regarding the influence of BNP on prognosis, and many factors including renal function, weight, age, and gender may influence BNP, absolute numbers are not given for this variable.
BNP = brain natriuretic peptide; CI = cardiac index; peak Vo_2 = average peak oxygen uptake during exercise; RA = right atrial; RV = right ventricle; WHO = World Health Organization.
Modified from McLaughlin VV, Gaine SP, Howard LS, et al. Treatment goals of pulmonary hypertension. *J Am Coll Cardiol.* 2013;62:D73-D81; and McLaughlin VV, Archer SL, Badesch DB, et al. ACCF/AHA 2009 expert consensus document on pulmonary hypertension: a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents and the American Heart Association developed in collaboration with the American College of Chest Physicians; American Thoracic Society, Inc.; and the Pulmonary Hypertension Association. *J Am Coll Cardiol.* 2009;53:1573-1619.

hematologic, systemic, and metabolic disorders (see Table 68-1). No treatments are of proven value.

Assessing Response to Therapy

Given the complexity of the disease, the variable response to therapy, and the goal of optimizing and individualizing care, patients with pulmonary arterial hypertension must be observed closely (Table 68-2). Consensus recommendations rely on the routine assessment of important prognostic indicators, such as WHO functional class, 6-minute walking distance, and echocardiographic and hemodynamic parameters (Table 68-3).[9] Patients who achieve these parameters, no matter which specific therapy or approach is used, seem to have a better prognosis than those who do not achieve these goals.

PROGNOSIS

Several clinical factors are correlated with prognosis (see Table 68-3). The natural history of symptomatic idiopathic pulmonary arterial hypertension is a median survival of 2.8 years with 1-, 3-, and 5-year survival rates of 68%, 48%, and 34%, respectively. In the era of targeted therapies, survival has improved but still remains suboptimal, with 1-, 2-, and 3-year survival of 86%, 70%, and 55% for incident cases.[10] Pulmonary hypertension itself is the direct cause of death in about 50% of patients and contributes to but does not directly cause death in the other 50%.[11]

Patients with pulmonary arterial hypertension related to the scleroderma spectrum of diseases tend to have a poorer prognosis than that of those with idiopathic pulmonary arterial hypertension, whereas patients with pulmonary arterial hypertension related to congenital heart disease tend to have a better prognosis, perhaps because they have better right ventricular function. Two large registries have shed light on the prognosis of patients with pulmonary arterial hypertension. Important predictors of poorer survival include male gender, worse functional class, reduced exercise tolerance as measured by the 6-minute walking distance, elevated right atrial pressures, and lower cardiac output. The natural history of patients with groups 2, 3, and 4 pulmonary hypertension is influenced by their left-sided heart and lung disease. In most cases, the presence of pulmonary hypertension in addition to the underlying disease portends a poor prognosis.

Grade A References

A1. Galie N, Corris PA, Frost A, et al. Updated treatment algorithm of pulmonary arterial hypertension. *J Am Coll Cardiol.* 2013;62:D60-D72.
A2. Jing ZC, Parikh K, Pulido T, et al. Efficacy and safety of oral treprostinil monotherapy for the treatment of pulmonary arterial hypertension: a randomized, controlled trial. *Circulation.* 2013;127:624-633.
A3. Rubin LJ, Badesch DB, Barst RJ, et al. Bosentan therapy for pulmonary arterial hypertension. *N Engl J Med.* 2002;346:896-903.
A4. Galie N, Olschewski H, Oudiz RJ, et al. Ambrisentan for the treatment of pulmonary arterial hypertension: results of the ambrisentan in pulmonary arterial hypertension, randomized, double-blind, placebo-controlled, multicenter, efficacy (ARIES) study 1 and 2. *Circulation.* 2008;117:3010-3019.
A5. Pulido T, Adzerikho I, Channick RN, et al. Macitentan and morbidity and mortality in pulmonary arterial hypertension. *N Engl J Med.* 2013;369:809-818.
A6. Ghofrani HA, Galie N, Grimminger F, et al. Riociguat for the treatment of pulmonary arterial hypertension. *N Engl J Med.* 2013;369:330-340.
A7. Galie N, Brundage BH, Ghofrani HA, et al. Tadalafil therapy for pulmonary arterial hypertension. *Circulation.* 2009;119:2894-2903.
A8. Redfield MM, Chen HH, Borlaug BA, et al. Effect of phosphodiesterase-5 inhibition on exercise capacity and clinical status in heart failure with preserved ejection fraction: a randomized clinical trial. *JAMA.* 2013;309:1268-1277.
A9. Ghofrani HA, D'Armini AM, Grimminger F, et al. Riociguat for the treatment of chronic thromboembolic pulmonary hypertension. *N Engl J Med.* 2013;369:319-329.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hoeper MM, Bogaard HJ, Condliffe R, et al. Definitions and diagnosis of pulmonary hypertension. *J Am Coll Cardiol*. 2013;62:D42-D50.
2. Dweik RA, Rounds S, Erzurum SC, et al. An official American Thoracic Society Statement: pulmonary hypertension phenotypes. *Am J Respir Crit Care Med*. 2014;189:345-355.
3. Brown LM, Chen H, Halpern S, et al. Delay in recognition of pulmonary arterial hypertension: factors identified from the REVEAL registry. *Chest*. 2011;140:19-26.
4. Rich JD, Rich S. Clinical diagnosis of pulmonary hypertension. *Circulation*. 2014;130:1820-1830.
5. Lang IM, Madani M. Update on chronic thromboembolic pulmonary hypertension. *Circulation*. 2014;130:508-518.
6. Coghlan JG, Denton CP, Grunig E, et al. Evidence-based detection of pulmonary arterial hypertension in systemic sclerosis: the DETECT study. *Ann Rheum Dis*. 2014;73:1340-1349.
7. McLaughlin VV, Langer A, Tan M, et al. Contemporary trends in the diagnosis and management of pulmonary arterial hypertension: an initiative to close the care gap. *Chest*. 2013;143:324-332.
8. Jais X, Olsson KM, Barbera JA, et al. Pregnancy outcomes in pulmonary arterial hypertension in the modern management era. *Eur Respir J*. 2012;40:881-885.
9. McLaughlin VV, Gaine SP, Howard LS, et al. Treatment goals of pulmonary hypertension. *J Am Coll Cardiol*. 2013;62:D73-D81.
10. Humbert M, Sitbon O, Chaouat A, et al. Survival in patients with idiopathic, familial, and anorexigen-associated pulmonary arterial hypertension in the modern management era. *Circulation*. 2010;122:156-163.
11. Tonelli AR, Arelli V, Minai OA, et al. Causes and circumstances of death in pulmonary arterial hypertension. *Am J Respir Crit Care Med*. 2013;188:365-369.

REVIEW QUESTIONS

1. A 63-year-old man presents with a 3-year history of progressive dyspnea on exertion. He is now unable to walk more than 100 feet without becoming dyspneic. On physical examination, he has a parasternal lift, a loud pulmonic component to the second heart sound, a tricuspid regurgitant murmur, a pulmonary vascular bruit, and 1+ lower extremity edema. His echocardiogram demonstrates normal left ventricular function, moderate right ventricular enlargement and dysfunction, and an estimated right ventricular systolic pressure of 75 mm Hg. Which is the most likely diagnosis?

A. Heart failure with preserved ejection fraction
B. Chronic thromboembolic pulmonary hypertension
C. Sleep-disordered breathing
D. Idiopathic pulmonary arterial hypertension
E. Pulmonary arterial hypertension due to an atrial septal defect

Answer: B The presence of a pulmonary vascular bruit on physical examination is specific to chronic thromboembolic pulmonary hypertension. Pulmonary vascular bruits arise from turbulent flow through partially obstructed pulmonary arteries. A history of a prior acute pulmonary embolism is absent in about 50% of patients who are diagnosed with chronic thromboembolic pulmonary hypertension.

2. A 35-year-old woman presents with a 1-year history of progressive dyspnea. She has been treated with bronchodilators but has not improved. She has symptoms of dyspnea and chest discomfort with modest exertion. Her physical examination reveals a right ventricular heave, a loud pulmonic component to her second heart sound, and a 2/6 tricuspid regurgitant murmur. Her electrocardiogram demonstrates right axis deviation. What is the most appropriate next study in her evaluation?

A. Exercise treadmill test
B. Echocardiogram
C. Pulmonary function tests
D. High-resolution chest computed tomography
E. Right-sided heart catheterization

Answer: B This patient has typical symptoms of pulmonary hypertension, and her physical examination is consistent with that diagnosis. The most appropriate next study is an echocardiogram to assess the size and function of her right-sided heart chambers, to estimate her right ventricular systolic pressure, and to evaluate her for other potential cardiac abnormalities that may contribute to pulmonary hypertension. Although she has exertional chest discomfort, coronary artery disease is unlikely to be the cause on the basis of her age and the findings on physical examination. Further evaluation of parenchymal lung disease may be appropriate at some point, but on the basis of her symptoms and physical examination findings, pulmonary hypertension is more likely.

3. A 32-year-old woman is diagnosed with idiopathic pulmonary arterial hypertension. On presentation, she had dyspnea with minimal exertion, exertional lightheadedness, and evidence of right-sided heart failure with 2+ lower extremity edema. Her right-sided heart catheterization revealed a mean pulmonary artery pressure of 60 mm Hg, right atrial pressure of 21 mm Hg, pulmonary capillary wedge pressure of 7 mm Hg, and cardiac index of 1.8 L/min/m^2. What is the most appropriate initial therapy for her?

A. Macitentan
B. Sildenafil
C. Riociguat
D. Intravenous epoprostenol
E. Inhaled iloprost

Answer: D This patient has advanced symptoms (functional class IV), signs of right-sided heart failure, and hemodynamics that portend a poor prognosis. Intravenous epoprostenol is the only pulmonary arterial hypertension therapy that has a grade A recommendation for patients with functional class IV symptoms. The other agents may be appropriate therapy for patients with less advanced symptom.

4. Which of the following are accepted treatment goals for patients with pulmonary arterial hypertension?

A. Functional class I or II symptoms
B. Cardiac index > 2.5 L/min/m^2
C. Nearly normal right ventricular function on imaging study
D. Normal brain natriuretic peptide
E. All of the above

Answer: E Treatment goals of pulmonary arterial hypertension have been established primarily from observational studies. Improving right ventricular function and consequently symptoms and exercise tolerance portends a better prognosis, regardless of which therapies have been used to obtain the goal. Patients with pulmonary arterial hypertension should be observed with reassessment of symptoms, exercise tolerance, and right ventricular function on a periodic basis. Combination therapy may be required to achieve treatment goals.

69

CONGENITAL HEART DISEASE IN ADULTS

ARIANE J. MARELLI

The convergence of major progress in medicine, pediatrics, and cardiovascular surgery has resulted in the survival of an increasingly large number of adult patients with congenital heart disease. Adult physicians are becoming increasingly responsible for these patients, commonly in concert with a cardiologist and a tertiary care facility.

DEFINITIONS

Patients can be divided into three categories according to the surgical status: unoperated, surgically palliated, or physiologically repaired. Congenital heart lesions can be classified as *acyanotic* or *cyanotic*. *Cyanosis* refers to a blue discoloration of the mucous membranes resulting from an increased amount of reduced hemoglobin. Central cyanosis occurs when the circulation is mixed because of a right-to-left shunt.

A *native lesion* refers to an anatomic lesion present at birth. Acquired lesions, naturally occurring or as a result of surgery, are superimposed on the native anatomy. *Palliative* interventions are performed in patients with cyanotic lesions and are defined as operations that either increase or decrease pulmonary blood flow while allowing a mixed circulation and cyanosis to persist (Table 69-1). *Physiologic repair* applies to procedures that provide total or nearly total anatomic and physiologic separation of the pulmonary and systemic circulations in complex cyanotic lesions and results in patients who are acyanotic.

Eisenmenger complex refers to flow reversal across a ventricular septal defect (VSD) when pulmonary vascular resistance exceeds systemic levels. *Eisenmenger physiology* designates the physiologic response in a broader category of shunt lesions in which a right-to-left shunt occurs in response to an elevation in pulmonary vascular resistance. *Eisenmenger syndrome* is a term applied to common clinical features shared by patients with Eisenmenger physiology.

Each congenital lesion can influence the course of another. For example, the physiologic consequences of a VSD are different if it occurs in isolation or in combination with pulmonary stenosis. A *simple lesion* is defined as either a shunt lesion or an obstructive lesion of the right or left side of the heart occurring in isolation. A *complex lesion* is a combination of two or more abnormalities.

EPIDEMIOLOGY

Genetic Determinants

About 20% of congenital heart defects are associated with a syndrome or chromosomal anomaly, and the most common such chromosomal anomaly is Down syndrome (trisomy 21), in which about 50% of patients have defects of the endocardial cushions and the ventricular septum. VSDs also occur in 90% of patients with trisomy 13 and trisomy 18. The most frequently observed defects in patients with Turner syndrome (45,X) are aortic coarctation, aortic stenosis, and atrial septal defect (ASD). About 15% of patients with tetralogy of Fallot have a deletion on chromosome 22q11; prevalence is higher in those with a right aortic arch. Abnormalities involving the chromosomal band 22q11 can also result in a group of syndromes, the most common of which is DiGeorge syndrome. The shared phenotypic features are designated CATCH-22 syndromes, that is, a combination of *c*ardiac defects, *a*bnormal facies, *t*hymic hypoplasia, *c*left palate, and *h*ypocalcemia. For families with a child who carries a congenital cardiac malformation due to a chromosomal anomaly, the risk of the cardiac malformation in future children is related to the risk of recurrence of the chromosomal anomaly itself.

Typically, single mutant genes are also associated with syndromes of cardiovascular malformations, although not every patient with the syndrome has the characteristic cardiac anomaly. Examples include osteogenesis imperfecta (autosomal recessive; Chapter 260), associated with aortic valve disease; Jervell and Lange-Nielsen syndrome (autosomal recessive) and Romano-Ward syndrome (autosomal dominant), associated with a prolonged QT interval and sudden death; and Holt-Oram syndrome (autosomal dominant), in which an ASD occurs with a range of other skeletal anomalies. Osler-Weber-Rendu telangiectasias (Chapter 173) are associated with pulmonary arteriovenous fistulas. Williams syndrome (Chapters 40 and 41) occurs with supravalvular aortic stenosis in most cases. Noonan syndrome is associated with pulmonary stenosis, ASD, and hypertrophic cardiomyopathy. Although autosomal dominant inheritance has been implicated for both, most cases are sporadic. Deletion at chromosome 7q11.23 has been identified in patients with Williams syndrome, and a gene defect has been mapped to 12q22-qter in patients with Noonan syndrome (Chapter 60).

The risk for recurrence when the mother carries a sporadically occurring congenital lesion varies from 2.5 to 18%, depending on the lesion. Obstructive lesions of the left ventricular outflow tract have the highest recurrence rates in offspring. When the father carries the lesion, 1.5 to 3% of the offspring are affected. When a sibling has a congenital cardiac anomaly, the risk for recurrence in another sibling varies from 1 to 3%.

PREVENTION

Genetic screening for the 22q11.2 microdeletion should be considered if patients with tetralogy of Fallot plan to have children. Without the 22q11 deletion, the risk of congenital heart disease in the fetus is 4 to 6%.

Incidence and Prevalence

Congenital heart defects are diagnosed in approximately 1% of births in the United States. The prevalence of congenital heart disease has increased in the general population, with the steepest rise observed in adults with severe or complex lesions. An overall prevalence of 6 per 1000 adults has been documented. The median age of patients with severe lesions has increased from childhood to late adolescence. Currently, more than 1 million adults are expected to be alive in the United States with congenital heart disease.[1] Advances in medical and surgical therapy have increased the survival of patients with congenital heart defects, thereby emphasizing that congenital heart disease is a lifelong condition that influences health care utilization and costs across the lifespan.

Bicuspid aortic valve occurs in about 2% of the general population, is the most common congenital cardiac anomaly encountered in adult populations, and accounts for up to half of surgical cases of aortic stenosis in adults (Chapter 75). ASDs constitute 30 to 40% of cases of congenital heart disease in adults, with ostium secundum ASD accounting for 7% of all congenital lesions. A solitary VSD represents 15 to 20% of all congenital lesions and is the most common congenital cardiac lesion observed in children; its high spontaneous closure rates explain the lesser prevalence in adults. Patent ductus arteriosus (PDA) accounts for 5 to 10% of all congenital cardiac lesions in infants with a normal birthweight. Pulmonary stenosis and coarctation of the aorta represent 3 to 10% of all congenital lesions.

Tetralogy of Fallot is the most common cyanotic congenital anomaly observed in adults. Together with complete transposition of the great arteries, these lesions account for 5 to 12% of congenital heart disease in infants. More complex lesions, such as tricuspid atresia, univentricular heart, congenitally corrected transposition of the great arteries, Ebstein anomaly, and double-outlet right ventricle, account for 2.5% or less of all congenital heart disease.

TABLE 69-1 PALLIATIVE SURGICAL SHUNTS FOR CONGENITAL HEART LESIONS

PALLIATIVE SHUNT	ANASTOMOSIS
SYSTEMIC ARTERIAL TO PULMONARY ARTERY SHUNTS	
Classic Blalock-Taussig	Subclavian artery to PA
Modified Blalock-Taussig	Subclavian artery to PA (prosthetic graft)
Potts anastomosis	Descending aorta to left PA
Waterston shunt	Ascending aorta to right PA
SYSTEMIC VENOUS TO PULMONARY ARTERY SHUNTS	
Classic Glenn	SVC to right PA
Bidirectional Glenn	SVC to right and left PA
Bilateral Glenn	Right and left SVC to right and left PA

PA = pulmonary artery; SVC = superior vena cava.
From Marelli A, Mullen M. Palliative surgical shunts for congenital heart lesions. *Clin Paediatr.* 1996;4:189.

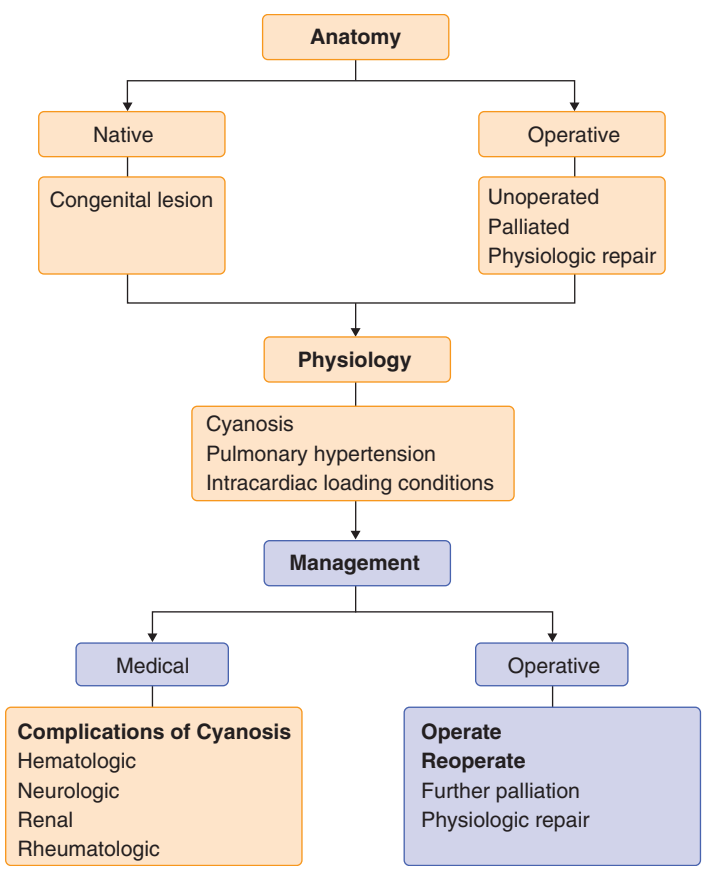

FIGURE 69-1. **The goals of complete clinical assessment in congenital heart disease are to define the anatomy and physiology to determine appropriate management.**

CLINICAL MANIFESTATIONS

Congenital heart disease is a lifelong condition during which the patient and the lesion evolve concurrently. A patient may have been monitored for many years because of an erroneous diagnosis made in infancy or childhood when diagnostic techniques were more limited. The differential diagnosis of native and surgical anatomy in the adult with an unknown diagnosis depends on whether the patient is cyanotic or acyanotic. On completion of the evaluation, the following questions should be answered (Fig. 69-1): What is the native anatomy? Has this patient undergone surgery for the condition? What is the physiology? What can and should be done for this patient both medically and surgically, and importantly, who should do it?

If the patient has not undergone surgery, the question is, Why not? If the patient is palliated, has the degree of cyanosis progressed as evidenced by a drop in systemic saturation or a rise in hemoglobin? If the patient has undergone a physiologic repair, what procedure was performed? Are residual lesions present, and have new lesions developed as a consequence of surgery? The patient's physiology is determined by the presence or absence of cyanosis, pulmonary hypertension, adequate filling of the cardiac chambers, and any resulting medical complications.

A clinical assessment, 12-lead electrocardiogram (ECG), chest radiograph, and baseline oxygen saturation should be part of every initial assessment. Two-dimensional transthoracic echocardiography (Chapter 55) and Doppler and color flow imaging are used to establish the diagnosis and to monitor the evolution of documented hemodynamic complications. Transesophageal echocardiographic examination is particularly useful in adults and is increasingly important during interventional catheter-guided therapy and surgery. Magnetic resonance imaging (Chapter 56) and computed tomography (Chapter 56) are useful adjuncts. Cardiac catheterization for congenital heart disease has shifted from pure diagnosis to include intervention. Coronary arteriography is recommended for adults older than 40 years in whom surgical intervention is contemplated.

Pulmonary Hypertension and Its Complications

Pulmonary hypertension secondary to structural disease of the heart or circulation can occur with or without an increase in pulmonary vascular resistance. Pulmonary vascular obstructive disease occurs when pulmonary vascular resistance rises and becomes fixed and irreversible. In the most common congenital anomalies, pulmonary hypertension is a result of increased pulmonary blood flow because of a native left-to-right shunt. Examples include ASD, a moderately sized VSD, PDA, and a variety of complex lesions. The rate at which pulmonary hypertension progresses to become pulmonary vascular obstructive disease varies from one lesion to another and depends at least in part on the source of pulmonary blood flow. Pulmonary hypertension typically develops in patients with an ASD after the fourth decade; Eisenmenger syndrome is a late complication seen in only 5 to 10% of cases. In contrast, in patients with a large VSD or persistent PDA, progressive elevation in pulmonary vascular resistance occurs rapidly because the pulmonary vascular bed is exposed not only to the excess volume of the left-to-right shunt but also to systemic arterial pressures. As a result, Eisenmenger complex develops in approximately 10% of patients with a large VSD during the first decade. Surgical pulmonary artery banding is a palliative measure aimed at decreasing pulmonary blood flow and protecting the pulmonary vascular bed against the development of early pulmonary vascular obstructive disease.

If forward flow from the right side of the heart is insufficient, native collaterals or surgical shunts provide an alternative source of pulmonary blood flow (see Table 69-1). With large surgical shunts, however, direct exposure of the pulmonary vascular bed to the high pressures of the systemic circulation causes pulmonary vascular obstructive disease. As a result, systemic to pulmonary arterial shunts are currently less favored in neonates and infants, in whom systemic venous to pulmonary arterial shunts are now preferred.

The term *Eisenmenger's syndrome* should be reserved for patients in whom pulmonary vascular obstructive disease is present and pulmonary vascular resistance is fixed and irreversible. These findings, in combination with the absence of left-to-right shunting, render the patient inoperable.

The clinical manifestations of Eisenmenger's syndrome include dyspnea on exertion, syncope, chest pain, congestive heart failure, and symptoms related to erythrocytosis and hyperviscosity. On physical examination, central cyanosis and digital clubbing are hallmark findings. Systemic oxygen saturations typically vary between 75 and 85%. The pulse pressure narrows as the cardiac output falls. Examination of jugular venous pressure can reveal a dominant *a* wave reflecting a noncompliant right ventricle until tricuspid insufficiency is severe enough to generate a large *v* wave. A prominent right ventricular impulse is felt in the left parasternal border in end expiration or in the subcostal area in end inspiration. A palpable pulmonary artery is commonly felt. The pulmonary component of the second heart sound is increased and can be felt in most cases. Pulmonary ejection sounds are common when the pulmonary artery is dilated with a structurally normal valve. Right atrial gallop is heard more frequently when the *a* wave is dominant. A murmur of tricuspid insufficiency is common, but the inspiratory increase in the murmur (Carvallo's sign) disappears when right ventricular failure occurs. In diastole, a pulmonary insufficiency murmur is often heard. The 12-lead ECG shows evidence of right atrial enlargement, right ventricular hypertrophy, and right axis deviation. Chest radiographic findings include a dilated pulmonary artery segment, cardiac enlargement, and diminished pulmonary vascular markings. Echocardiography confirms the right-sided pressure overload and pulmonary artery enlargement as well as the tricuspid and pulmonary insufficiency. Cardiac catheterization is indicated if doubt exists about the potential reversibility of the elevated pulmonary vascular resistance in a patient who might otherwise benefit from surgery.

Cyanosis occurs when persistent venous to arterial mixing results in hypoxemia. Adaptive mechanisms to increase oxygen delivery include an increase in oxygen content, a rightward shift in the oxyhemoglobin dissociation curve, a higher hematocrit, and an increase in cardiac output. When cyanosis is not relieved, chronic hypoxemia and erythrocytosis result in hematologic, neurologic, renal, and rheumatic complications.

Hematologic complications of chronic hypoxemia include erythrocytosis, iron deficiency, and bleeding diathesis. Hemoglobin and hematocrit levels as well as red blood cell indices should be checked regularly and correlated with systemic oxygen saturation levels. Symptoms of hyperviscosity include headaches, faintness, dizziness, fatigue, altered mentation, visual disturbances, paresthesias, tinnitus, and myalgia. Symptoms are classified as mild to moderate when they interfere with only some activities, or they can be marked to severe and interfere with most or all activities. Patients with compensated erythrocytosis establish an equilibrium hematocrit at higher levels in an

iron-replete state with minimal symptoms. Patients with decompensated erythrocytosis manifest unstable, rising hematocrit levels and experience severe hyperviscosity symptoms.

Hemostatic abnormalities can occur in up to 20% of cyanotic patients with erythrocytosis. Bleeding is usually mild and superficial and leads to easy bruising, skin petechiae, or mucosal bleeding, but epistaxis, hemoptysis, or even life-threatening postoperative bleeding can occur. A variety of clotting factor deficiencies and qualitative and quantitative platelet disorders have been described.

Neurologic complications, including cerebral hemorrhage, can be caused by hemostatic defects and are most often seen after inappropriate use of anticoagulant therapy. Patients with right-to-left shunts may be at risk for paradoxical cerebral emboli. Focal brain injury may provide a nidus for brain abscess if bacteremia supervenes. Attention should be paid to the use of air filters in peripheral intravenous lines to avoid paradoxical emboli through a right-to-left shunt.

Prophylactic phlebotomy has no place in the prevention of cerebral arterial thrombosis. Indications for phlebotomy are the occurrence of symptomatic hyperviscosity in an iron-replete patient and prevention of excessive bleeding perioperatively.

Pulmonary complications include massive pulmonary hemorrhage and in situ arterial thrombosis. A rapid clinical deterioration associated with progressive hypoxemia often marks the terminal stage of disease. No clear benefits are observed with the use of anticoagulants (systemic or intrapulmonary) because of the risk for prolonged bleeding due to the underlying coagulopathy. The chronic disease process and high mortality prohibit pulmonary endarterectomy.

Renal dysfunction can be manifested as proteinuria, hyperuricemia, or renal failure. Focal interstitial fibrosis, tubular atrophy, and hyalinization of afferent and efferent arterioles can be seen on renal biopsy. Increased blood viscosity and arteriolar vasoconstriction can lead to renal hypoperfusion with progressive glomerulosclerosis. Hyperuricemia is commonly seen in patients with cyanotic congenital heart disease and is thought to be due mainly to the decreased reabsorption of uric acid rather than to overproduction from erythrocytosis. Asymptomatic hyperuricemia need not be treated because lowering of uric acid levels has not been shown to prevent renal disease or gout.

Rheumatologic complications include gout and hypertrophic osteoarthropathy, which is thought to be responsible for the arthralgias affecting up to one third of patients with cyanotic congenital heart disease. In patients with right-to-left shunting, megakaryocytes released from the bone marrow bypass the lung and are entrapped in systemic arterioles and capillaries, where they release platelet-derived growth factor, which promotes local cell proliferation. Digital clubbing and new osseous formation with periostitis occur and cause the symptoms of arthralgia. Symptomatic hyperuricemia and gouty arthritis can be treated as necessary with colchicine, probenecid, or allopurinol; nonsteroidal anti-inflammatory drugs are best avoided, given the baseline hemostatic anomalies in these patients.

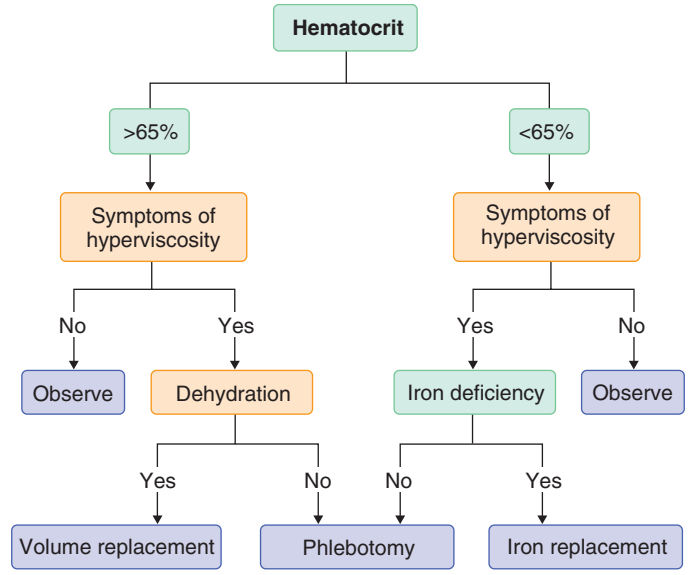

FIGURE 69-2. **Treatment algorithm for erythrocytosis of cyanotic congenital heart disease.**

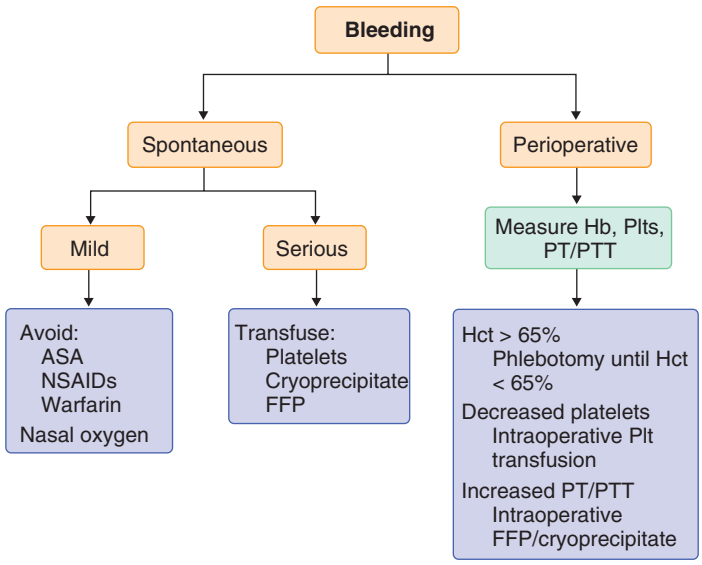

FIGURE 69-3. **Treatment algorithm for bleeding diathesis of cyanotic congenital heart disease.** ASA = acetylsalicylic acid; FFP = fresh-frozen plasma; Hb = hemoglobin; Hct = hematocrit; NSAIDs = nonsteroidal anti-inflammatory drugs; Plts = platelets; PT = prothrombin time; PTT = partial thromboplastin time.

TREATMENT

Rx

In patients with Eisenmenger syndrome, bosentan (e.g., 62.5 mg twice daily for 4 weeks, then 125 mg twice daily) may improve hemodynamics and exercise capacity after 4 months of use. Chronic oxygen therapy is unlikely to benefit hypoxemia caused by right-to-left shunting in the setting of a fixed pulmonary vascular resistance. Oxygen therapy may be considered for cyanotic patients during long-distance flights.

In the iron-replete state, moderate to severe hyperviscosity symptoms typically occur when hematocrit levels exceed 65%. If no evidence of dehydration is present, removal of 500 mL of blood during a 30- to 45-minute period should be followed by quantitative volume replacement with normal saline or dextran (Fig. 69-2). The procedure may be repeated every 24 hours until symptomatic improvement occurs.

Treatment of spontaneous bleeding is dictated by its severity and the abnormal hemostatic parameters (Fig. 69-3). For severe bleeding, platelet transfusions, fresh-frozen plasma, vitamin K, cryoprecipitate, and desmopressin have been used. Reduction in erythrocyte mass also improves hemostasis, so cyanotic patients undergoing surgery should have prophylactic phlebotomy if the hematocrit is greater than 65%.

Iron deficiency is common in cyanotic adult patients because of excessive bleeding or phlebotomy. In contrast to normocytic erythrocytosis, which is rarely symptomatic at hematocrit levels less than 65%, iron deficiency may be manifested with hyperviscosity symptoms at hematocrit levels well below 65%. The treatment of choice is not phlebotomy but oral iron repletion until a rise in hematocrit is detected, typically within 1 week.

SIMPLE LESIONS

Isolated Shunt Lesions

Hemodynamic complications of significant shunts relate to volume overload and chamber dilation of the primary chamber receiving the excess left-to-right shunt and to secondary complications of valvular dysfunction and damage to the pulmonary vascular bed. The size and duration of the shunt determine the clinical course and therefore the indications for closure. The degree of shunting is a function of both the size of the communication and, depending on its location, biventricular compliance or pulmonary and systemic vascular resistance. Clinically apparent hemodynamic sequelae of shunts are typically apparent or can be expected to occur when pulmonary to systemic flow ratios exceed 1.5 : 1.

Shunt size can be inferred and measured with cardiac ultrasonography. Secondary enlargement of the cardiac chambers receiving excess shunt flow in diastole occurs as the shunt size becomes hemodynamically significant; in addition, the pulmonary artery becomes enlarged as pulmonary pressure rises. When tricuspid insufficiency occurs primarily from right ventricular dilation or secondary to pulmonary hypertension, the regurgitant jet can be used to estimate the pulmonary pressure as another indicator of shunt significance. When the pulmonary to systemic flow ($\dot{Q}:\dot{Q}$) exceeds 2 : 1, the volume of blood in both circulations can be estimated by comparing the stroke volume at the pulmonary and aortic valves. Shunt detection and quantification can also be obtained by a first-pass radionuclide study. As a bolus of radioactive substance is injected into the systemic circulation, the rise and fall of radionuclide activity can be measured in the lungs. When a shunt is significant, the rate of persistent activity in the lungs over time can be used to calculate the shunt fraction. For both echocardiographic and radionuclide quantification of shunt size, sources of error are multiple. The most predictable results are obtained only in experienced laboratories. Uncertainty about the physiologic significance of a borderline shunt can be minimized by integrating serial determinations from multiple clinical and relevant diagnostic sources rather than basing management decisions on a single calculated shunt value.

Atrial Septal Defect

Classification of ASDs is based on anatomic location. Most commonly, an ostium secundum ASD occurs in the central portion of the interatrial septum as a result of an enlarged foramen ovale or excessive resorption of the septum primum (Video 69-1). The combination of a secundum ASD and acquired mitral stenosis is known as *Lutembacher syndrome*, the pathophysiology of which is determined by the relative severity of each. Abnormal development of the embryologic endocardial cushions results in a variety of atrioventricular canal defects, the most common of which consists of a defect in the lower part of the atrial septum in the ostium primum location, typically accompanied by a cleft mitral valve and mitral regurgitation. The sinus venosus defect, which accounts for 2 to 3% of all interatrial communications, is located superiorly at the junction of the superior vena cava and right atrium and is generally associated with anomalous drainage of the right-sided pulmonary veins into the superior vena cava or right atrium. Less commonly, interatrial communications can be seen at the site of the coronary sinus, typically associated with an anomalous left superior vena cava.

The pathophysiology is determined by the effects of the shunt on the heart and pulmonary circulation. Right atrial and right ventricular dilation occurs as shunt size increases with pulmonary to systemic flow ratios greater than 1.5 : 1. Superimposed systemic hypertension and coronary artery disease modify left ventricular compliance and favor left-to-right shunting. Mitral valve disease can occur in up to 15% of patients older than 50 years. Right-sided heart failure, atrial fibrillation, or atrial flutter can occur as a result of chronic right-sided volume overload and progressive ventricular and atrial dilation. Stroke can result from paradoxical emboli, atrial arrhythmias, or both. A rise in pulmonary pressure occurs because of the increased pulmonary blood flow. Pulmonary hypertension is unusual before 20 years of age but is seen in 50% of patients older than 40 years. The overall incidence of pulmonary vascular obstructive disease is 15 to 20% in patients with ASD. Eisenmenger disease with reverse shunting, a late and rare complication of isolated secundum ASD, is reported in 5 to 10% of patients.

DIAGNOSIS

Although most patients are minimally symptomatic in the first three decades, more than 70% become impaired by the fifth decade. Initial symptoms include exercise intolerance, dyspnea on exertion, and fatigue caused most commonly by right-sided heart failure and pulmonary hypertension.[2] Palpitations, syncope, and stroke can occur with the development of atrial arrhythmias.

On physical examination, most adults have a normal general physical appearance. When Holt-Oram syndrome is present, the thumb may have a third phalanx or may be rudimentary or absent. With an uncomplicated nonrestrictive communication between both atria, the *a* and *v* waves are equal in amplitude. Precordial palpation typically discloses a normal left ventricular impulse unless mitral valve disease occurs. Characteristically, if the shunt is significant, a right ventricular impulse can be felt in the left parasternal area in end expiration or in the subxiphoid area in end inspiration. A dilated pulmonary artery can sometimes be felt in the second left intercostal space. On auscultation, the hallmark of an ASD is the wide and fixed splitting of the second heart sound. Pulmonary valve closure, as reflected by P_2, is delayed

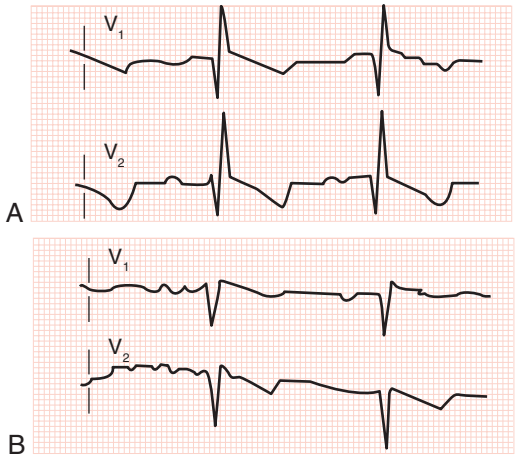

FIGURE 69-4. Electrocardiographic hallmark in atrial septal defect. Right precordial leads V_1 and V_2 illustrate two variants of an incomplete right bundle branch block pattern, the rSrT pattern (A) and the rsR′ pattern (B).

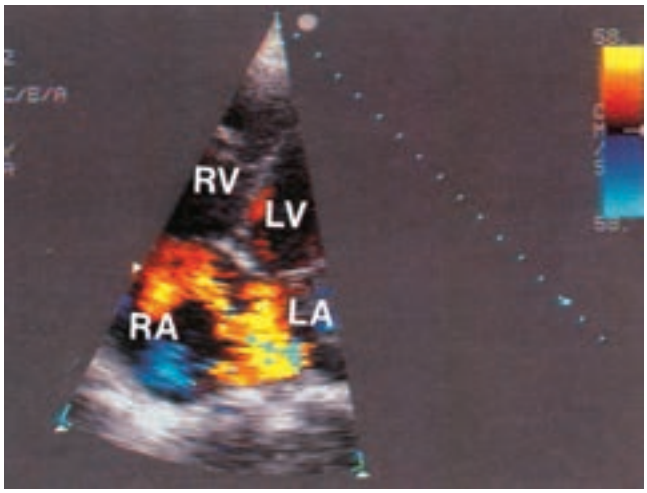

FIGURE 69-5. Color flow Doppler apical four-chamber view showing blood flow from the left atrium (LA) to the right atrium (RA) through a moderately sized atrial septal defect. LV = left ventricle; RV = right ventricle. (From Forbes CD, Jackson WF. Color Atlas and Text of Clinical Medicine. 3rd ed. London: Mosby; 2003.)

because of right ventricular overload and the increased capacitance of the pulmonary vascular bed. The A_2-P_2 interval is fixed because the increase in venous return elevates the right atrial pressure during inspiration, thereby decreasing the degree of left-to-right shunting and offsetting the usual phasic respiratory changes. In addition, compliance of the pulmonary circulation is reduced from the high flow, thus making the vascular compartment less susceptible to any further increase in blood flow. A soft midsystolic murmur generated by the increased flow across the pulmonary valve is usually heard in the second left interspace. In the presence of a high left-to-right shunt volume, increased flow across the tricuspid valve is heard as a mid-diastolic murmur at the lower left sternal border. With advanced right-sided heart failure, evidence of systemic venous congestion is present.

The ECG characteristically shows an incomplete right bundle branch block pattern (Fig. 69-4). Right axis deviation and atrial abnormalities, including a prolonged PR interval, atrial fibrillation, and flutter, are also seen. Typically, the chest radiograph shows pulmonary vascular plethora with increased markings in both lung fields consistent with increased pulmonary blood flow (see Fig. 56-6). The main pulmonary artery and both its branches are dilated. Right atrial and right ventricular dilation can be seen. Cardiac ultrasonography is diagnostic and provides important prognostic information (Fig. 69-5). Ostium primum and secundum ASDs are easily identifiable with transthoracic imaging, but a sinus venosus ASD can be missed unless it is specifically sought. For more accurate visualization of the superior interatrial septum and localization of the pulmonary veins, transesophageal echocardiography is useful. With Doppler study, pulmonary artery pressures can be quantified, and the $\dot{Q}:\dot{Q}$ can be measured.

VIDEO 69-1. Secundum Atrial Septal Defect

TREATMENT

Rx

Closure of an ASD either percutaneously or surgically is indicated in the presence of right-sided heart enlargement, with or without symptoms. Centrally located defects measuring up to 3.5 cm can be occluded by transcatheter techniques in a cardiac catheterization laboratory. Advantages of this approach include the avoidance of sternotomy and cardiopulmonary bypass. Uncomplicated secundum ASDs may be closed surgically in children and adults with minimal operative mortality, although surgical closure is usually reserved for patients in whom concomitant repair of associated valvular anomalies is required, anomalous pulmonary veins are present, or device closure is not technically feasible.

In patients older than 40 years with symptoms and significant shunts, closure improves functional status and survival.■ In the presence of a significant shunt, closure of an ASD before 25 years of age without evidence of pulmonary hypertension results in a long-term outcome that is similar to that of age- and sex-matched controls. Advanced age (60 years) is not a contraindication to ASD closure in the presence of a significant shunt because a significant number of patients will show evidence of symptomatic improvement. Preoperative pulmonary artery pressure and the presence or absence of pulmonary vascular disease are important predictors of successful interventional outcome.

Patent Foramen Ovale

Integrity of the fetal circulation depends on the patency of the foramen ovale. In most cases, the fall in pulmonary vascular resistance at birth induces the foramen to become sealed. Necropsy studies have revealed that the foramen ovale remains patent beyond the first year of life in about 30% of individuals, and clinical studies have demonstrated that the prevalence of patent foramen ovale is three times higher in patients with cryptogenic stroke (Chapter 407), particularly before the age of 55 years, because of right-to-left shunting and paradoxical embolization of material from the venous circulation. Cardiac investigation of the patient with cryptogenic stroke includes transesophageal echocardiography with agitated saline injection to visualize the presence of a right-to-left shunt (Chapter 55). Patent foramen ovale most likely to result in future paradoxical embolization is found in patients younger than 55 years with a prior cryptogenic stroke, in association with a hypermobile septum with aneurysm formation, and when a significant amount of right-to-left shunting is present at rest without provocative maneuvers.

TREATMENT

Rx

Current information does not support closure of a patent foramen ovale for primary prevention of a first stroke in a patient in whom it is fortuitously diagnosed on routine echocardiography.[3] Among patients with a prior cryptogenic stroke, however, randomized trials provide some guidance. In one randomized trial of patients with cryptogenic stroke or transient ischemic attack and a patent foramen ovale, device closure was no better than medical therapy alone for preventing a recurrent stroke or transient ischemic attack.■ In a second trial, closure of a patent foramen ovale for secondary prevention of cryptogenic embolism did not significantly reduce recurrent embolic events or death (3.4%) after 4 years compared with medical therapy (5.2%).■ In a third randomized trial of adults who had had a cryptogenic ischemic stroke, there also was no significant benefit from closure of a patent foramen ovale, but closure was superior to medical therapy alone in prespecified analyses limited to patients who actually received and adhered to the original treatment (0.66 vs. 1.39 events per 100 person-years).■ These findings together suggest a probable small benefit of closure compared with medical therapy, which is warfarin to an international normalized ratio of 2.0 to 3.0.[4] Primary closure of a patent foramen ovale is clearly indicated when a patient has contraindications to medical therapy, if medical therapy has failed, or in the presence of a hypercoagulable state not treatable by medical therapy.

Ventricular Septal Defect

For anatomic classification of VSDs, the interventricular septum can be divided into four regions. Defects of the membranous septum, or infracristal VSDs, are located in a small translucent area beneath the aortic valve and account for up to 80% of VSDs. These VSDs typically show a variable degree of extension into the inlet or outlet septum, hence their designation as perimembranous (Video 69-2). Infundibular defects or supracristal outlet VSDs occur in the conal septum above the crista supraventricularis and below the pulmonary valve. Inlet defects are identified at the crux of the heart between the tricuspid and mitral valves and are usually associated with other anomalies of the atrioventricular canal. Defects of the trabecular or muscular septum can be multiple and occur distal to the septal attachment of the tricuspid valve and toward the apex.

The pathophysiology and clinical course of VSDs depend on the size of the defect, the status of the pulmonary vascular bed, and the effects of shunt size on intracardiac hemodynamics. Unlike ASDs, a VSD may decrease in size with time. Approximately half of all native VSDs are small, and more than half of them close spontaneously; moderate or even large VSDs may also close in 10% or less of cases. The highest closure rates are observed in the first decade of life; spontaneous closure in adult life is unusual.[5]

Patients who have a small defect with trivial or mild shunts are defined as those with a $\dot{Q}:\dot{Q}$ of less than 1.5 and normal pulmonary artery pressure and vascular resistance. Patients with moderate defects have a $\dot{Q}:\dot{Q}$ ratio of greater than 1.2 and elevated pulmonary artery pressure but not elevated pulmonary vascular resistance. Patients with a large and severe defect have an elevated $\dot{Q}:\dot{Q}$ ratio with high pulmonary pressure and elevated pulmonary vascular resistance. Eisenmenger complex develops in about 10% of patients with VSDs, usually when there is no resistance to flow at the level of the defect, which can be as large as the aorta. When a systolic pressure gradient is present between the ventricles, the physiologic severity may be trivial or mild but can also be moderate or severe.

Minimal or mild defects usually cause no significant hemodynamic or physiologic abnormality. A moderate or severe defect causes left atrial and ventricular dilation consistent with the degree of left-to-right shunting. Shunting across the ventricular septum occurs predominantly during systole when left ventricular pressure exceeds that on the right; diastolic filling abnormalities occur in the left atrium. With moderate or severe defects, the right side of the heart becomes affected as a function of the rise in pulmonary pressure and pulmonary blood flow.

DIAGNOSIS

An adult with a VSD most commonly has a small restrictive lesion that either was small at birth or has undergone some degree of spontaneous closure. A second group of patients consists of those with large, nonrestrictive VSDs that have not been operated on; these patients have had Eisenmenger complex for most of their lives. Patients with a moderately sized defect are typically symptomatic as children and are therefore more likely to have repair at a young age.

Patients with a trivial or mild shunt across a small, restrictive VSD are usually asymptomatic. Physical examination discloses no evidence of systemic or pulmonary venous congestion, and jugular venous pressure is normal. A thrill may be palpable at the left sternal border. Auscultation reveals normal S_1 and S_2 without gallops. A grade 4 or louder, widely radiating, high-frequency, pansystolic murmur is heard maximally in the third or fourth intercostal space and reflects the high-pressure gradient between the left and right ventricles throughout systole. The striking contrast between a loud murmur and otherwise normal findings on cardiac examination is an important diagnostic clue. The ECG and chest radiograph are also normal in patients with small VSDs.

At the other end of the spectrum are patients with Eisenmenger complex (see earlier). Between these two extremes are patients with a moderate defect, whose pathologic process reflects a combination of pulmonary hypertension and left-sided volume overload resulting from a significant left-to-right shunt. In adults, shortness of breath on exertion can be the result of both pulmonary venous congestion and elevated pulmonary pressure. On physical examination, a diffuse palpable left ventricular impulse occurs with a variable degree of right ventricular hypertrophy and an accentuated second heart sound. A systolic murmur persists as long as pulmonary vascular resistance is below systemic resistance. The ECG commonly shows left atrial enlargement and left ventricular hypertrophy. The chest radiograph shows shunt vascularity with an enlarged left atrium and ventricle. The degree of pulmonary hypertension determines the size of the pulmonary artery trunk.

Echocardiography can identify the defect and determine the significance of the shunt by assessing left atrial and ventricular size, pulmonary artery pressure, and presence or absence of right ventricular hypertrophy. Cardiac catheterization is reserved for those in whom surgery is considered. Adults with a small defect of no physiologic significance need not be studied invasively. Those with Eisenmenger complex have severe pulmonary vascular disease and are not surgical candidates. Patients who have a moderately sized shunt that appears hemodynamically significant and in whom pulmonary pressures are elevated are most likely to benefit from direct measurements of pulmonary vascular resistance and reactivity.

VIDEO 69-2. Perimembranous Ventricular Septal Defect

TREATMENT Rx

Patients with Eisenmenger complex have pulmonary vascular resistance that is prohibitive to surgery. For this group of patients, management centers on the medical complications of cyanosis (see earlier). In a few patients with small defects, complications can relate to progressive tricuspid insufficiency caused by septal aneurysm formation or to acquired aortic insufficiency when an aortic cusp becomes engaged in the high-velocity jet flow generated by the defect. The intermediate group of patients with a defect of moderate physiologic significance should have their VSDs closed unless closure is contraindicated by high pulmonary vascular resistance. For a perimembranous VSD, transcatheter closure is as effective as open surgical closure and has a lower rate of minor adverse events.

Late results after operative closure of isolated VSDs include residual patency in up to 20% of patients, only about 5% of whom need a reoperation. Rhythm disturbances after surgical closure of VSDs include tachyarrhythmias and conduction disturbances. Right bundle branch block occurs in one third to two thirds of patients, whereas first-degree atrioventricular block and complete heart block occur in less than 10%. Sudden cardiac death after surgical repair of VSD occurs in 2% of patients.

Patent Ductus Arteriosus

The ductus arteriosus connects the descending aorta to the main pulmonary trunk near the origin of the left subclavian artery (Fig. 69-6). Normal postnatal closure results in fibrosis and degenerative changes in the ductal lumen, leaving in its place the residual ligamentum arteriosum, which rarely can become part of an abnormal vascular ring. When the duct persists, significant calcification of the aortic ductal end is observed.

The physiologic consequences of a PDA are determined by its size and length as well as by the ratio of pressure and resistance of the pulmonary and aortic circulations on either end of the duct. If systolic and diastolic pressure in the aorta exceeds that in the pulmonary artery, aortic blood flows continuously down a pressure gradient into the pulmonary artery and then returns to the left atrium. The left atrium and subsequently the left ventricle dilate, whereas the right side of the heart becomes progressively affected as pulmonary hypertension develops.

A small PDA has continuous flow throughout the entire cardiac cycle without left-sided heart dilation, pulmonary hypertension, or symptoms. Patients with a small PDA, although protected from hemodynamic complications of a significant left-to-right shunt, remain at risk for infectious endarteritis, which usually develops on the pulmonary side of the duct and occurs at a rate of about 0.45% per year after the second decade. Because endarteritis accounts for up to one third of the total mortality in patients with PDA, ductal closure should be considered even when the PDA is small.

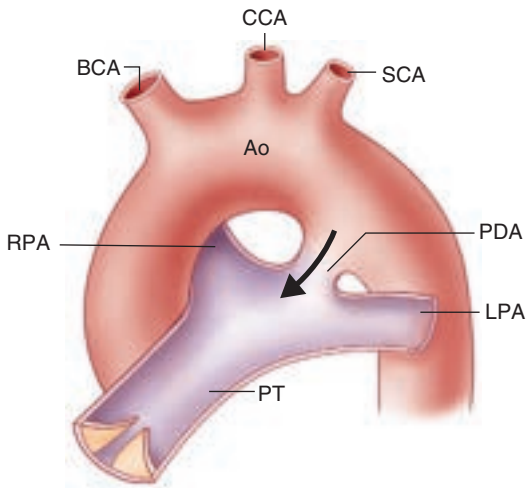

FIGURE 69-6. Anatomy of a patent ductus arteriosus. Note the relationships of the position of the ductus, left subclavian artery, and pulmonary artery bifurcation. Ao = aorta; BCA = brachiocephalic; CCA = common carotid artery; LPA = left pulmonary artery; PDA = patent ductus arteriosus; PT = pulmonary trunk; RPA = right pulmonary artery; SCA = subclavian artery. (From Perloff JK, ed. Clinical Recognition of Congenital Heart Disease. 4th ed. Philadelphia: WB Saunders; 1994:510.)

A PDA is of moderate or large size but still restrictive when a left-to-right shunt occurs throughout systole and diastole is of variable duration. Left atrial or ventricular dilation and pulmonary hypertension will vary with the quantity of left-to-right shunting as well as with the secondary effects on the pulmonary vascular bed. Symptoms generally increase by the second and third decades and include dyspnea, palpitations, and exercise intolerance. As heart failure, pulmonary hypertension, or endarteritis develops, mortality rises to 3 to 4% per year by the fourth decade, and two thirds of patients die by 60 years of age. Eisenmenger physiology with systemic or suprasystemic pulmonary pressure and a right-to-left shunt develops in 5% of patients with an isolated PDA.

DIAGNOSIS

In patients with Eisenmenger physiology, a right-to-left shunt from the pulmonary artery to the descending aorta results in decreased oxygen saturation in the lower extremities compared with the upper extremities. This difference in cyanosis and clubbing is most prominent in the toes; the left arm is variably affected through the left subclavian artery, and the right arm is typically spared. With a large left-to-right shunt, the pulse pressure widens as diastolic flow into the pulmonary artery lowers systemic diastolic pressure. The arterial pulse becomes bounding as a result of increased stroke volume. Precordial palpation discloses variable left and right ventricular impulses as determined by the relative degree of left-sided volume overload and pulmonary hypertension. In the presence of a continuous aortopulmonary gradient, the classic "machinery" murmur of a PDA can be heard at the first or second left intercostal space below the left clavicle. As the pulmonary pressure rises, the diastolic component of the murmur becomes progressively shorter. With the development of Eisenmenger physiology and equalization of aortic and pulmonary pressure, the entire murmur may disappear, and the clinical findings are dominated by pulmonary hypertension.

In adult patients with a significant left-to-right shunt, the ECG shows a bifid P wave in at least one limb lead consistent with left atrial enlargement and a variable degree of left ventricular hypertrophy. The PR interval is prolonged in about 20% of patients. In older patients, the chest radiograph shows calcification at the location of the PDA. Characteristically, the ascending aorta and pulmonary artery are dilated, and the left-sided chambers are enlarged. Echocardiography may not directly visualize the PDA but can accurately identify it by a Doppler signal that often parallels the length of the murmur. Left-sided heart dilation and pulmonary hypertension can be quantified and monitored. Cardiac catheterization to assess pulmonary vascular resistance is commonly indicated before closure.

TREATMENT Rx

After ligation of a PDA in infancy or early childhood, cardiac function is commonly normal, and no special follow-up is required. If pulmonary artery pressure and pulmonary vascular resistance are substantially elevated, preoperative evaluation should assess the degree of reversibility. With Eisenmenger disease, closure is contraindicated. Closure of a PDA either percutaneously or surgically is indicated in the presence of left-sided heart enlargement or if prior endarteritis has occurred. Reported operative mortality rates vary from less than 1 to 8%, depending on the presence of calcification and the degree of pulmonary hypertension. Transcatheter or coil occlusion is an accepted procedure in adults. Residual shunt rates vary from 0.5 to 8%, depending on the device used. Small residual defects that are detected by echocardiography but are not associated with an audible murmur or hemodynamic findings do not appear to carry a significant risk for endarteritis.

Aortopulmonary Window

An aortopulmonary window is typically a large defect across the adjacent segments of both great vessels above their respective valves and below the pulmonary artery bifurcation. The pathophysiologic mechanism is similar to that of a PDA. The shunt is usually large, so pulmonary vascular resistance rises rapidly and abolishes the aortopulmonary gradient in diastole. The murmur is usually best heard at the third left intercostal space. With a right-to-left shunt, differential cyanosis never occurs because the shunt is proximal to the brachiocephalic vessels. Differentiation of an aortopulmonary window from a PDA can usually be confirmed with echocardiography; the left-to-right shunt is seen in the main pulmonary artery in the aortopulmonary window compared with the left pulmonary artery bifurcation in PDA. Cardiac catheterization confirms the diagnosis and hemodynamics. Surgical

repair is necessary unless pulmonary vascular obstructive disease precludes closure.

Pulmonary Arteriovenous Fistulas

Pulmonary arteriovenous fistulas can occur as isolated congenital disorders or as part of generalized hereditary hemorrhagic telangiectasia (Osler-Weber-Rendu syndrome; Chapter 173). These fistulas typically occur in the lower lobes or the right middle lobe and can be small or large, single or multiple. The arterial supply usually comes from a dilated, tortuous branch of the pulmonary artery.

The most common finding is that of abnormal opacity on a chest radiograph in a patient with buccal ruby patches or in an otherwise healthy adult who has mild cyanosis. Shunting between deoxygenated pulmonary arterial blood and the oxygenated pulmonary venous blood results in a physiologic right-to-left shunt. The degree of shunting is typically small and not significant enough to result in dilation of the left atrium and ventricle. Heart failure is unusual. Hemoptysis can result if a fistula ruptures into a bronchus. In patients with hereditary hemorrhagic telangiectasia, angiomas occur on the lips and mouth as well as in the gastrointestinal tract and on pleural, liver, and vaginal surfaces. Epistaxis is most common, but cerebrovascular accidents can also occur. Patients with hereditary hemorrhagic telangiectasia can have symptoms that resemble those of a transient ischemic attack even in the absence of right-to-left shunting. On physical examination, cyanosis and clubbing can be notable or barely detectable. Auscultation can disclose soft systolic or continuous noncardiac murmurs on the chest wall adjacent to the fistula. The murmur typically increases with inspiration. The ECG is usually normal. The chest radiograph shows one or more densities, typically in the lower lobes or in the right middle lobe. An echocardiogram can confirm the presence of the fistula by showing early opacification of the left atrium in the absence of any other intracardiac communication when saline is injected into a peripheral vein. The absence of a hemodynamically significant shunt can be confirmed by documenting normal cardiac chamber size.

If the hypoxemia is progressive or if a neurologic complication is documented to have occurred because of paradoxical emboli, fistula closure should be considered. Options include percutaneous catheter techniques if the fistula is small and accessible or a pulmonary wedge resection or lobectomy if the fistula is large. Multiple or recurrent fistulas create a major therapeutic challenge.

Isolated Obstructive Lesions of the Right and Left Ventricular Outflow Tract

Complications of obstructive lesions of the outflow tract relate to the secondary effects of exposure to pressure overload in the chamber proximal to the obstruction. The inability to increase systemic or pulmonary blood flow in the face of a fixed obstruction can cause exercise intolerance, inadequate myocardial perfusion, ventricular arrhythmias, and sudden death.

RIGHT VENTRICULAR OUTFLOW TRACT OBSTRUCTION

Obstruction of the right ventricular outflow tract can occur at the level of the pulmonary valve (see later), above it in the main pulmonary artery or its branches, or below it in the right ventricle itself. Supravalvular and branch pulmonary artery stenoses are important and common complications in patients with tetralogy of Fallot (see later). Residual supravalvular pulmonary stenosis is sometimes seen after palliative pulmonary artery banding to decrease pulmonary blood flow in patients with large left-to-right shunts. Congenital branch pulmonary artery stenosis can occur in isolation or with valvular pulmonary stenosis, shunt lesions, or a variety of syndromes. Patients with Noonan syndrome have a characteristic phenotypic facial appearance, short stature, and webbed neck; cardiac lesions may include a dysplastic pulmonary valve, left ventricular hypertrophic cardiomyopathy, and peripheral pulmonary artery stenosis. Supravalvular pulmonary stenosis can be seen with supravalvular aortic stenosis in Williams (elfin facies) syndrome.

Pulmonary atresia refers to an absent, imperforate, or closed pulmonary valve, which typically occurs in conjunction with other malformations. Pulmonary atresia with a nonrestrictive VSD is a complex cyanotic malformation that is discussed later.

Primary infundibular stenosis with an intact ventricular septum can result from a fibrous band just below the infundibulum. In a double-chambered right ventricle, obstruction is caused by anomalous muscle bundles that divide the right ventricle into a high-pressure chamber below the hypertrophied muscle bundles and a low-pressure chamber above the bundles and below the valve. The clinical features vary according to the presence or absence of other lesions, such as pulmonary valvular stenosis or VSD.

VALVULAR PULMONARY STENOSIS

Isolated congenital valvular pulmonary stenosis (Chapter 75) is a common lesion due to a bicuspid valve in 20% of cases, a dysplastic valve caused by myxomatous changes and severe thickening in 10% of cases, and an abnormal trileaflet valve in most of the remaining cases. Fusion of the leaflets results in a variable degree of thickening and calcification in older patients.

The 25-year survival of patients with valvular pulmonary stenosis is greater than 95% but is worse in those with severe stenosis and peak systolic gradients greater than 80 mm Hg. For patients with mild (<50 mm Hg gradients) and moderate (50 to 80 mm Hg gradients) pulmonary stenosis, bacterial endocarditis, complex ventricular arrhythmias, and progression of the stenosis are uncommon.

DIAGNOSIS

A patient with moderate or even severe pulmonary stenosis may be asymptomatic. With severe stenosis, exercise intolerance can be associated with presyncope and ventricular arrhythmias. Progressive right-sided heart failure is the most common cause of death. On physical examination of patients with significant pulmonary stenosis, jugular venous pressure has a dominant *a* wave, reflecting a noncompliant right ventricle. Palpation discloses a sustained parasternal lift of right ventricular hypertrophy. An expiratory systolic ejection click is characteristic if the leaflets are still mobile. In moderate or severe stenosis, a grade 3 or louder systolic murmur can be heard and felt in the second left interspace. The length of the murmur increases as it peaks progressively later in systole with an increasing degree of obstruction. If right-sided heart failure occurs, tricuspid insufficiency and systemic venous congestion develop. The ECG can show right axis deviation and tall, peaked right atrial P waves in lead II. With more than mild stenosis, the R wave exceeds the S wave in lead V_1. On chest radiography, the main pulmonary artery can be dilated even if the stenosis is mild. Characteristically, the left pulmonary artery is more dilated than the right because of the leftward direction of the high-velocity jet. A variable degree of right ventricular hypertrophy is manifested as right-sided chamber enlargement. Echocardiography can establish the diagnosis and determine the severity by Doppler ultrasound examination. Patients with valvular pulmonary stenosis do not require cardiac catheterization. The mean gradient at echocardiography correlates well with the gradient measured at cardiac catheterization and should be used for therapeutic decisions because peak instantaneous Doppler gradients tend to overestimate the severity of disease.

TREATMENT

Rx

Depending on symptoms, percutaneous balloon valvotomy should be considered for patients with isolated valvular pulmonary stenosis and mean Doppler gradients of 30 mm Hg or greater unless there is moderate or severe pulmonary regurgitation. In the presence of a doming valve, pulmonary angioplasty is the procedure of choice for adults, who achieve persistently good results at 10-year follow-up. For patients with hypoplastic pulmonary arteries or subvalvular stenosis (double-chambered right ventricle), surgical resection of right ventricular muscle bands can be performed.

LEFT VENTRICULAR OUTFLOW TRACT OBSTRUCTION

Stenosis of the left ventricular outflow tract can occur at, below, or above the aortic valve. Discrete subaortic stenosis, most commonly caused by a fibromuscular ring just below the valve, accounts for 15 to 20% of all cases of congenital obstruction of the left ventricular outflow tract. Concomitant aortic insufficiency occurs in 50% of cases. Supravalvular aortic stenosis results from thickened media and intima above the aortic sinuses; early coronary atherosclerosis or even ostial coronary obstruction can occur.

CONGENITAL VALVULAR AORTIC STENOSIS

The normal aortic valve has three cusps and commissures. A unicuspid aortic valve accounts for most cases of severe aortic stenosis in infants (Chapter 75). A bicuspid aortic valve, which is the most common congenital cardiac malformation, functions normally at birth but often becomes gradually obstructed as calcific and fibrous changes occur; prolapse of one or both cusps can cause aortic insufficiency.

The pathophysiologic mechanism of aortic stenosis depends not only on its severity but also on the age at diagnosis. When a bicuspid aortic valve becomes stenotic in adulthood because of degenerative changes, criteria for diagnosis and intervention parallel those for other forms of acquired aortic stenosis (Chapter 75). The estimated overall 25-year survival rate for patients with congenital valvular aortic stenosis diagnosed in childhood is 85%. Children with initial peak cardiac catheterization gradients of less than 50 mm Hg have long-term survival rates of higher than 90%, as opposed to survival rates of 80% in those with gradients of 50 mm Hg or greater.

DIAGNOSIS

Symptoms include angina, exertional dyspnea, presyncope, and syncope and may progress to heart failure. The auscultatory hallmark of a bicuspid aortic valve is an audible systolic ejection click that is typically of a higher pitch than the first heart sound and is best heard not at the cardiac base but at the apex. The sound is caused by sudden movement of the stenotic valve as it moves superiorly in systole and is followed by the typical aortic stenosis murmur (Chapter 75). When significant calcification of the valve results in reduced mobility, the ejection sound is no longer heard. The diagnosis is easily confirmed by two-dimensional echocardiography, with which the number and orientation of aortic cusps can readily be identified.

TREATMENT Rx

Conservative management is generally indicated for mild stenosis with a peak gradient of less than 25 mm Hg, but close supervision is required because 20% of these patients require an intervention during long-term follow-up. Unlimited athletic participation is allowed only for asymptomatic patients with peak gradients of less than 20 to 25 mm Hg, a normal ECG, and a normal exercise test result. For children who are symptomatic or have gradients greater than 30 mm Hg but do not have significant aortic insufficiency, transcatheter aortic valvotomy is preferred. Aortic valvuloplasty can be considered in young adults, but calcification limits its success, and valve replacement is usually required (Chapter 75). For adults, treatment decisions are similar to those for aortic stenosis from other causes. For patients with subvalvular aortic stenosis, surgical intervention is indicated in the presence of peak gradients above 50 mm Hg, symptoms, or progressive aortic insufficiency.

COARCTATION OF THE AORTA

Aortic coarctation typically occurs just distal to the left subclavian artery at the site of the aortic ductal attachment or its residual ligamentum arteriosum. Less commonly, the coarctation ridge lies proximal to the left subclavian. A bicuspid aortic valve is the most common coexisting anomaly, but VSDs and PDAs are also seen. Pseudocoarctation refers to buckling or kinking of the aortic arch without the presence of a significant gradient.

The most common complications of aortic coarctation are systemic hypertension (Chapter 67) and secondary left ventricular hypertrophy with heart failure. Systemic hypertension is caused by decreased vascular compliance in the proximal aorta and activation of the renin-angiotensin system in response to renal artery hypoperfusion below the obstruction. Left ventricular hypertrophy occurs in response to chronic pressure overload. Congestive heart failure occurs most commonly in infants and then after 40 years of age. The high pressure proximal to the obstruction stimulates the growth of collateral vessels from the internal mammary, scapular, and superior intercostal arteries to the intercostals of the descending aorta. Collateral circulation increases with age and contributes to perfusion of the lower extremities and the spinal cord. This mechanism, although adaptive in a patient who has not undergone surgery, accounts for significant morbidity during surgery when the motor impairment results from inadequate protection of spinal perfusion. Aneurysms occur most notably in the ascending aorta and in the circle of Willis. Premature coronary disease is thought to be related to the resulting hypertension. Complications, including bacterial endarteritis at the coarctation site or, more commonly, endocarditis at the site of a bicuspid aortic valve, cerebrovascular complications, myocardial infarction, heart failure, and aortic dissection, occur in 2 to 6% of patients, more frequently in those with advancing age who have not undergone surgery.

DIAGNOSIS

Young adults may be asymptomatic with incidental systemic hypertension and decreased lower extremity pulses. Coarctation should always be considered in adolescents and young adult men with unexplained upper extremity hypertension. The pressure differential can cause epistaxis, headaches, leg fatigue, or claudication. Older patients have angina, symptoms of heart failure, and vascular complications.

On physical examination, the lower half of the body is typically slightly less developed than the upper half. The hips are narrow and the legs are short, in contrast to broad shoulders and long arms. Blood pressure measurements should be obtained in each arm and one leg; an abnormal measurement is an increase of less than 10 mm Hg in popliteal systolic blood pressure compared with arm systolic blood pressure. The diastolic pressure should be the same in the upper and lower extremities. A pressure differential of more than 30 mm Hg between the right and the left arms is consistent with compromised flow in the left subclavian artery. Right brachial palpation characteristically reveals a strong or even bounding pulse compared with a slowly rising or absent femoral, popliteal, or pedal pulse. Examination of the eyegrounds can reveal tortuous or corkscrew retinal arteries. Precordial palpation is consistent with left ventricular pressure overload. On auscultation, a systolic ejection sound reflecting the presence of a bicuspid aortic valve should be sought. The coarctation itself generates a systolic murmur heard posteriorly, in the midthoracic region, the length of which correlates with the severity of the coarctation. Over the anterior of the chest, systolic murmurs reflecting increased collateral flow can be heard in the infraclavicular areas and the sternal edge or in the axillae.

In adult coarctation, the most common finding on the ECG is left ventricular hypertrophy. Chest radiographic findings are diagnostic. Location of the coarctation segment between the dilated left subclavian artery above and the leftward convexity of the descending aorta below results in the "3 sign" (Fig. 69-7). Bilateral rib notching as a result of dilation of the posterior intercostal arteries is seen on the posterior of the third to eighth ribs when the coarctation is below the left subclavian. Unilateral rib notching sparing the left ribs is observed when the coarctation occurs proximal to the left subclavian artery. Transthoracic echocardiography documents the gradient in the descending aorta and determines the presence of left ventricular hypertrophy. Magnetic resonance imaging (Chapter 56) is the best modality for visualizing the anatomy of the descending aorta. Cardiac catheterization should measure pressures and assess collaterals when surgery is contemplated.

TREATMENT Rx

Intervention is recommended in patients who have gradients of 20 mm Hg or more on cardiac catheterization (Chapter 57) or who have evidence of significant collateral flow on imaging studies. The choice between catheter intervention and surgical intervention, which should be made in conjunction with a specialist, depends on the associated anomalies and the anatomy of the coarctation segment. Fifty percent of patients repaired when they are older

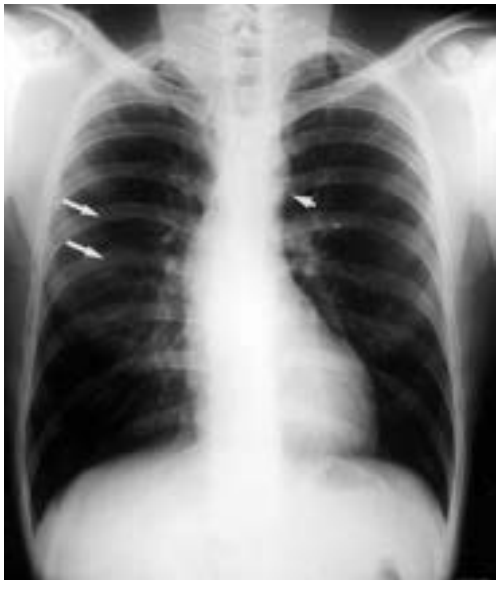

FIGURE 69-7. Chest radiograph of a patient with coarctation of the aorta. The radiographic *3* formed by the dilated subclavian artery above and the dilated aorta below (*short arrow*) is shown. Note the notching, best seen at the level of the seventh and eighth ribs (*long arrows*). The dilated ascending aortic segment can also be seen.

than 40 years have residual hypertension, whereas those who have undergone surgery between the ages of 1 and 5 years have a less than 10% prevalence of hypertension on long-term follow-up. Actuarial survival rates are 94%, 86%, and 74% at 10, 20, and 30 years, respectively, after initial surgical repair.[6]

Balloon angioplasty is the treatment of choice for focal recoarctation in patients who have previously been operated on. The incidence of incomplete relief and restenosis is decreased in adults by endovascular stent placement. Focal complications include aortic aneurysms and, rarely, aortic rupture.

Anomalies of the Sinuses of Valsalva and Coronary Arteries

SINUS OF VALSALVA ANEURYSMS

At the base of the aortic root, the aortic valve cusps are attached to the aortic wall, above which three small pouches, or sinuses, are seated. The right coronary artery originates from one sinus and the left main coronary artery from a second; the third is called the *noncoronary sinus.* A weakness in the wall of the sinus can result in aneurysm formation with or without rupture. In more than 90% of cases, the aneurysm involves the right or noncoronary cusp. Rupture typically occurs into the right side of the heart at the right atrial or ventricular level with a resulting large left-to-right shunt driven by the high aortic pressure.

A previously asymptomatic young man typically has chest pain and rapidly progressing shortness of breath sometimes after physical strain. The physical examination is consistent with significant heart failure. Even if the communication is between the aorta and the right side of the heart, biventricular failure is not unusual. The classic murmur is loud and continuous, often with a thrill. A murmur of aortic insufficiency secondary to damage to the adjacent aortic valve may be superimposed. The chest radiograph shows volume overload of both ventricles with evidence of shunt vascularity and pulmonary venous congestion. The echocardiogram is diagnostic. Cardiac catheterization can verify the integrity of the coronary artery adjacent to the ruptured aneurysm.

Even though symptoms may abate as the heart dilates, progressive cardiac decompensation typically results in death within 1 year of the rupture. A ruptured sinus of Valsalva aneurysm therefore requires urgent surgical repair.

CORONARY ARTERY FISTULAS

Fistulas arise from the right or left coronary arteries and in 90% of cases drain into the right ventricle, the right atrium, or the pulmonary artery in order of decreasing frequency. Typically, young patients are asymptomatic, but supraventricular arrhythmias are seen with progressive dilation of the intracardiac chambers. Angina can occur as the fistula creates a coronary steal by diverting blood away from the myocardium. Heart failure is seen with large fistulas. A continuous murmur heard in a young, otherwise normal acyanotic, asymptomatic patient should suggest the diagnosis. Most fistulas are associated with a small shunt, and hence the murmur is often less than grade 3 and is heard in the precordial area. Unless the shunt is large, the ECG is normal, as is the chest radiograph. The echocardiogram, especially the transesophageal echocardiogram, is diagnostic. Percutaneous transcatheter closure with coil embolization is preferred, but surgical ligation is also an alternative.

ANOMALOUS ORIGIN OF THE CORONARY ARTERIES

The left main coronary artery normally arises from the left sinus of Valsalva and courses leftward, posterior to the right ventricular outflow tract. The right coronary artery arises from the right sinus of Valsalva and courses rightward to the right ventricle. Isolated ectopic or anomalous origins of the coronary arteries (see Fig. 57-7) are seen in 0.6 to 1.5% of patients undergoing coronary angiography.

The most common anomaly is ectopic origin of the left circumflex artery from the right sinus of Valsalva, followed by anomalous origin of the right coronary artery from the left sinus and anomalous origin of the left main coronary artery from the right sinus. If the anomalous coronary artery does not course between the pulmonary artery and aorta, the prognosis is favorable. Risks of ischemia, myocardial infarction, and death are greatest when the left main coronary artery courses between both great vessels.

Coronary arteries can also originate from the pulmonary trunk. If both the right and left arteries originate from the pulmonary trunk, death usually occurs in the neonatal period. If only the left anterior descending coronary artery originates from the pulmonary trunk, the rate of survival to adulthood is approximately 10%, depending on the development of collateral retrograde flow to the anomalous artery from a normal coronary artery. This collateral flow may cause a continuous murmur along the left sternal border, congestive heart failure from the large shunt, and a coronary steal syndrome as blood is diverted away from the normal artery.

A single coronary ostium can provide a single coronary artery that branches into right and left coronary arteries, the left then giving rise to the circumflex and the anterior descending arteries. The ostium can originate from the right or left aortic sinus. The coronary circulation is functionally normal unless one of the branches passes between the aorta and the pulmonary artery.

Diagnostic procedures include angiography, magnetic resonance imaging, and transesophageal echocardiography. For an anomalous coronary artery that originates from the pulmonary artery, surgical reimplantation into the aorta is preferred. For an anomalous artery that courses between the pulmonary artery and aorta, a bypass graft to the distal vessel is preferred.

SPECIFIC COMPLEX LESIONS

Tetralogy of Fallot

Tetralogy of Fallot, the most common cyanotic malformation, is characterized by superior and anterior displacement of the subpulmonary infundibular septum, which causes the tetrad of pulmonary stenosis, VSD, aortic override, and right ventricular hypertrophy. The VSD is perimembranous in 80% of cases. Additional cardiac anomalies include a right-sided aortic arch in up to 25% of patients. An anomalous left anterior descending artery originating from the right coronary cusp and crossing over the right ventricular outflow tract is seen in 10% of cases. Other associated anomalies include ASD, left superior vena cava, defects of the atrioventricular canal, and aortic insufficiency. With pulmonary atresia, pulmonary blood flow occurs through aortic to pulmonary collaterals. Life expectancy is limited unless staged reconstructive surgery is performed.

The physiology in unrepaired tetralogy of Fallot is determined by the severity and location of the pulmonic outflow obstruction and by the interaction of pulmonary and systemic vascular resistance across a nonrestrictive VSD. Because the pulmonary stenosis results in a relatively fixed pulmonary resistance, a drop in systemic vascular resistance as occurs with exercise is associated with increased right-to-left shunting and increasing cyanosis. A child who squats after running is attempting to reverse the process by increasing systemic vascular resistance by crouching with bent knees. Native pulmonary blood flow is typically insufficient. Unless a PDA has remained open, a cyanotic adult will typically have undergone a palliative procedure to increase pulmonary blood flow.

Examination of unrepaired patients reveals central cyanosis and clubbing. The right ventricular impulse is prominent. The second heart sound is single and represents the aortic closure sound with an absent or inconspicuous P_2. Typically, little or no systolic murmur is heard across the pulmonary valve because the more severe the obstruction, the more right-to-left shunting occurs and the less blood flows across a diminutive right ventricular outflow tract. A diastolic murmur of aortic insufficiency is often heard in adults. In the presence of a palliative systemic arterial to pulmonary artery shunt, the high-pressure gradient generates a loud continuous murmur. In a patient who has not undergone surgery, progressive infundibular stenosis and cyanosis occur. Before the advent of palliative surgery, mortality rates were 50% in the first few years of life, and survival past the third decade was unusual.

Complete surgical repair consists of patch closure of the VSD and relief of the right ventricular outflow tract obstruction. Adequate pulmonary blood flow is ensured by reconstruction of the distal pulmonary artery bed. Previous palliative shunts are usually taken down. Complete repair in childhood yields a 90 to 95% 10-year survival rate with good functional results, and 30-year survival rates may be as high as 85%. Total correction with low mortality and a favorable long-term follow-up is possible even in adulthood.

After repair, residual pulmonary stenosis, proximal or distal, with a right ventricular pressure greater than 50% of systemic occurs in up to 25% of patients. Some degree of pulmonary insufficiency is common, particularly if a patch has been inserted at the level of the pulmonary valve or if a pulmonary valvotomy has been performed. Residual VSDs can be found in up to 20% of patients. Patients may be asymptomatic or may have symptoms related to long-term complications after surgical repair. Symptoms can reflect residual right ventricular pressure or volume overload or arrhythmias at rest or with exercise. Angina can occur in a young patient if surgical repair has damaged an anomalous left anterior descending artery as it courses across the right ventricular outflow tract. In acyanotic adults, clubbing commonly regresses.

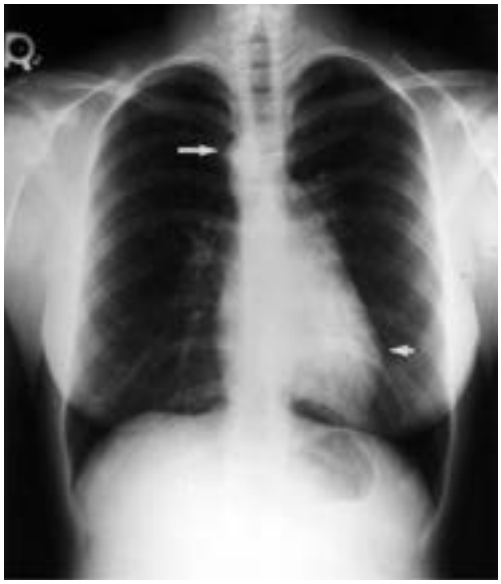

FIGURE 69-8. Chest radiograph of an adult after tetralogy of Fallot repair. A right aortic arch with rightward indentation of the trachea (*long arrow*) can be seen. The right ventricular apex remains upturned (*short arrow*). Note the sternal wires consistent with intracardiac repair, clarifying the fullness of the pulmonary artery segment often seen after extensive enlargement of the right ventricular outflow tract.

A right ventricular impulse is often felt as a result of residual pulmonary insufficiency or stenosis. A systolic murmur can represent residual pulmonary stenosis, residual VSD, or tricuspid insufficiency. A diastolic murmur can reflect aortic or pulmonary insufficiency. Ventricular arrhythmias are common after repair, with an incidence of sudden death as high as 5%.

The ECG in unrepaired tetralogy of Fallot shows right axis deviation, right atrial enlargement, and dominant right ventricular forces over the precordial leads. The most common finding after repair is complete right bundle branch block, which is seen in 80 to 90% of patients. The chest radiograph typically shows an upturned apex with a concave pulmonary artery segment giving the classic appearance of a boot-shaped heart. Figure 69-8 demonstrates the findings in an adult after repair. The apex is persistently upturned, although the pulmonary artery segment is no longer concave. Echocardiography can confirm the diagnosis and document intracardiac complications in repaired and unrepaired patients. Shunt patency can be determined by Doppler examination. Magnetic resonance imaging can accurately document stenosis in the distal pulmonary artery bed. Cardiac catheterization is reserved for patients in whom operative or reoperative treatment is contemplated or in whom the integrity of the coronary circulation needs to be verified.

Patients with a change in exercise tolerance, angina, or evidence of heart failure and those with symptomatic arrhythmias or syncope should be referred for complete evaluation. Surgical reintervention is generally considered when right ventricular pressure is more than two thirds as high as systemic pressure because of residual right ventricular outflow tract obstruction, free pulmonary regurgitation occurs with right ventricular dysfunction or sustained arrhythmias, or a residual VSD causes a significant shunt.

Patients should be seen yearly by an adult congenital heart specialist for assessment of right ventricular function, pulmonary valve dysfunction, and arrhythmia. Surveillance should include a history, physical examination, and 12-lead ECG. Sudden cardiac death can occur in 3 to 6% of patients observed between 20 and 30 years, sometimes despite favorable hemodynamics.[7]

Complete Transposition of the Great Arteries

Complete transposition of the great arteries is the second most common cyanotic lesion, and surgically corrected adults are increasingly common. In simple transposition of the great arteries, the atria and ventricles are in their normal positions, but the aorta arises from the right ventricle, and the pulmonary artery arises from the left ventricle. When the aorta is anterior and rightward with respect to the pulmonary artery, as is most common, D-transposition is present. The native anatomy has the pulmonary and systemic circulations in parallel, with deoxygenated blood recirculating between the right side of the heart and the systemic circulation, whereas oxygenated blood recirculates from the left side of the heart to the lungs. The condition is incompatible with life unless a VSD, PDA, or ASD is present or an ASD is created; a hemodynamically significant VSD is present in 15% of cases. Subpulmonary obstruction of the left ventricular outflow tract occurs in 10 to 25% of cases.

The Senning or Mustard atrial baffle repairs, which were the first corrective procedures, redirect oxygenated blood from the left atrium to the right ventricle so that it may be ejected into the aorta while deoxygenated blood detours the right atrium and heads for the left ventricle and into the pulmonary artery. Although this operation results in acyanotic physiology, the right ventricle assumes a permanent position under the aorta and pumps against systemic pressures, a lifelong task for which it was not designed. When the subpulmonary obstruction is significant, the Rastelli procedure reroutes blood at the ventricular level by tunneling the left ventricle to the aorta inside the heart through a VSD. A conduit is then inserted outside the heart between the left ventricle and aorta. The best current option, which is the arterial switch operation, transects the aorta and pulmonary artery above their respective valves and switches them to become realigned with their physiologic outflow tracts and appropriate ventricles. The proximal coronary arteries are translocated from the sinuses of the native aorta to the neoaorta (native pulmonary artery). In this operation, each ventricle reassumes the role that it was embryologically destined to fulfill. Long-term complications include neoaortic regurgitation, supravalvular stenosis, and chronotropic incompetence. In addition, surveillance is required for possible coronary artery complications.[8]

At present, adults with transposition of the great arteries most commonly have undergone an atrial baffle repair, with an expected 15-year survival rate of 75% and a 20-year survival rate of 70%. For patients with an atrial baffle procedure, symptoms include exercise intolerance, palpitations caused by bradyarrhythmias or atrial flutter, and right ventricular failure. The patient is typically acyanotic unless a baffle leak exists. Reoperation is required in approximately 20% of patients for baffle-related complications, progressive left ventricular outflow tract stenosis, or severe tricuspid regurgitation.

If an adult patient is cyanotic and has a native intracardiac shunt or a palliative shunt, referral to an appropriate facility should be undertaken to explore the possibility of intracardiac repair. Comprehensive echocardiographic imaging should be performed in a specialized adult congenital center. Echocardiography can confirm the diagnosis and explore related abnormalities.

Congenitally Corrected Transposition of the Great Arteries

In congenitally corrected transposition of the great arteries, the great arteries are transposed, the ventricles are inverted, but the atria remain in their normal position. The systemic circulation (left atrium, morphologic right ventricle, and aorta) and pulmonary circulation (right atrium, morphologic left ventricle, and pulmonary artery) are in series. The patient is therefore acyanotic unless an intracardiac shunt is also present. The right ventricle is aligned with the aorta and performs lifelong systemic work, which accounts in part for its eventual failure. Associated lesions include a VSD, pulmonary stenosis, and Ebstein malformation of the left-sided tricuspid valve. Complete heart block develops at a rate of 2% per year. Patients with congenitally corrected transposition of the great arteries and no other associated defects can remain free of symptoms until the sixth decade, at which time significant atrioventricular valve regurgitation, failure of the right (systemic) ventricle, supraventricular arrhythmias, and heart block occur.

Right-Sided Ebstein Anomaly

The septal and posterior cusps of the tricuspid valve are largely derived from the right ventricle as it liberates a layer of muscle that skirts away from the cavity to become valve tissue. When this process occurs abnormally, the posterior and septal cusps of the tricuspid valve remain tethered to the muscle and adhere to the right ventricular surface—hence the diagnostic hallmark of Ebstein anomaly, apical displacement of the septal tricuspid leaflet.

In right-sided Ebstein anomaly of the tricuspid valve, the right side of the heart consists of three anatomic components: the right atrium proper, the true right ventricle, and the atrialized portion of the right ventricle between the two. The displaced septal and posterior tricuspid leaflets lie between the atrialized right ventricle and the true right ventricle. In mild Ebstein anomaly, the degree of tricuspid leaflet tethering is only mild, the anterior leaflet retains mobility, and the size of the true right ventricle is only mildly reduced. Severe Ebstein anomaly is associated with severe tethering of the tricuspid leaflet tissue and a diminutive, hypocontractile true right ventricle. Functionally, the valve is regurgitant because it is unable to appose its three leaflets during ventricular contraction. Valvular regurgitation and asynchronous, abnormal

right ventricular function cause the dilation and right-sided heart failure observed in the more severe forms of the lesion. The wide spectrum of severity of the anomaly is based on the degree of tricuspid leaflet tethering and the relative proportion of atrialized and true right ventricle. The most common associated cardiac defect, a secundum ASD or patent foramen ovale, is reported in more than 50% of patients. On physical examination, a clicking "sail sound" is heard as the second component of S_1 when tricuspid valve closure becomes loud and delayed.

The 12-lead ECG typically shows highly peaked P waves with a wide, often bizarre-looking QRS complex. Preexcitation occurs in 20% of patients; supraventricular tachyarrhythmias, atrial fibrillation, and atrial flutter occur in 30 to 40% of patients and constitute the most common findings in adolescents and adults with right-sided Ebstein anomaly.

When patients of all ages are taken together, the predicted mortality is approximately 50% by the fourth or fifth decade. Complications include atrial arrhythmias due to severe right atrial enlargement and cyanosis caused by a right-to-left atrial shunt as tricuspid insufficiency increases and the right ventricle fails. Atrial arrhythmias, cyanosis, and the presence of an intra-atrial communication also increase the risk for stroke.

Intervention is considered when functional status or cyanosis worsens, significant atrial arrhythmias are documented, or a cerebrovascular accident occurs. Surgical options include replacement or repair of the tricuspid valve and closure of the ASD. The feasibility of tricuspid valvuloplasty depends on the size and mobility of the anterior tricuspid leaflet, which is used to construct a unicuspid right-sided valve.

Atrioventricular Canal Defect

Embryologic septation of the atrioventricular canal results in closure of the inferior portion of the interatrial septum and the superior portion of the interventricular septum. Septation is achieved with the growth of endocardial cushions, which also contribute to development of the mitral and tricuspid valves. Hence, the nomenclature *atrioventricular canal defect* or *endocardial cushion defect* is used to designate this group of anomalies.

A partial atrioventricular canal defect refers to an ostium primum ASD with a cleft mitral valve. The anomaly is manifested as a hemodynamic combination of an ASD with a variable degree of mitral regurgitation. The 12-lead ECG shows the typical findings of left axis deviation with a Q wave in leads I and aVL and a prolonged PR interval. The echocardiogram shows a defect in the inferior portion of the interatrial septum and a cleft mitral valve.

A complete atrioventricular canal defect is an uncommon defect consisting of a primum ASD, an inlet VSD that usually extends to the membranous interventricular septum, and a common atrioventricular valve. Adults who have not been operated on usually have Eisenmenger syndrome unless concomitant pulmonary stenosis has protected the pulmonary vascular bed or the VSD has undergone spontaneous closure, in which case the physiologic consequences are similar to those of a partial atrioventricular canal.

Surgical repair of an atrioventricular defect consists of closing the interatrial or interventricular communication with reconstruction of the common atrioventricular valve or closure of the cleft in the mitral valve. An adult who has undergone repair may have significant residual regurgitation of the mitral or tricuspid valve. Even after surgery, acquired subaortic obstruction can occur in the long left ventricular outflow tract, which has a classic gooseneck deformity on cardiac angiography.

Univentricular Heart and Tricuspid Atresia

The terms *single ventricle*, *common ventricle*, and *univentricular heart* have been used interchangeably to describe the double-inlet ventricle, in which one ventricular chamber receives flow from both the tricuspid and mitral valves. In 75 to 90% of cases, the single ventricle is a morphologic left ventricle. Obstruction of one of the great arteries is common, and life expectancy is short without an operation. The patients most likely to survive to adulthood palliated or, rarely, without surgery have a single ventricle of the left morphologic type, with pulmonary stenosis protecting the pulmonary vascular bed.

In tricuspid atresia, no orifice is found between the right atrium and right ventricle, and an underdeveloped or hypoplastic right ventricle is present. The morphologic left ventricle is consistently normally developed and therefore becomes the single functional ventricle. Typically, blood flows into the right atrium, then through an obligatory ASD and to the left atrium, where it then proceeds to the left ventricle. Variable features include a VSD, the abnormal position of the great arteries, and the relative degree of pulmonary stenosis, all of which are used to classify tricuspid atresia. Without surgery, 50% of patients die in the first 6 months and 90% in the first decade.

Adult patients rarely have not been operated on. They may be acyanotic after the Fontan operation; if cyanotic and palliated, the patient may benefit from further palliation or may be eligible for the Fontan operation. With the Glenn shunt or the Fontan operation, a direct anastomosis is created between the systemic venous and pulmonary circulations. Venous blood flows passively from the systemic veins to the pulmonary circulation and returns oxygenated to a left-sided atrium and into the single functional ventricle, which then pumps oxygenated blood into the systemic circulation. The Glenn anastomosis diverts part of the systemic venous return to the lungs, whereas the Fontan procedure makes the patient acyanotic by diverting the entire systemic venous circulation to the pulmonary vascular bed. For optimal results, a successful Fontan operation requires low pulmonary vascular resistance, preserved single ventricular function, and unobstructed anastomosis between the systemic veins and the pulmonary arteries. At 5-year follow-up, 80% or more of Fontan survivors are in New York Heart Association functional class I or II, with successful pregnancy reported in a small number of patients. When patients of all ages are considered together, 10-year survival rates vary from 60 to 70%. Late deaths are due to reoperation, arrhythmia, ventricular failure, protein-losing enteropathy, and liver dysfunction. Yearly follow-up with specialized imaging is recommended.

Vascular Malformations

AORTIC ARCH ANOMALIES

Vascular Rings and Other Arch Anomalies

One of the most frequent developmental errors of the aortic arch is an aberrant right subclavian artery originating distal to the left subclavian and coursing rightward behind the esophagus at the level of the third thoracic vertebra. Although the finding is frequent, symptoms are uncommon. When symptoms occur, the term *dysphagia lusoria* has been used in reference to swallowing difficulties that result from esophageal compression. Abnormal development of the brachial arches and dorsal aorta can result in a variety of anomalies that lead to the formation of vascular rings around the trachea and esophagus. The outcome is often benign, but symptoms of respiratory compromise or dysphagia warrant surgery. When the left pulmonary artery arises from the right and passes leftward between the trachea and esophagus, a pulmonary artery sling occurs. Symptoms of tracheal compression warrant correction.

A right aortic arch occurs when the aortic arch courses toward the right instead of the left. Mirror-image branching is the most common anatomic variant. In most cases, this anomaly coexists with other congenital lesions, notably tetralogy of Fallot.

ANOMALOUS VENOUS CONNECTIONS

Anomalies of Systemic Venous Return

A persistent left superior vena cava can be fortuitously diagnosed on chest radiography or on echocardiography. Its clinical relevance depends on development of the coronary sinus. If the coronary sinus is normally formed, typically the left superior vena cava drains into the right atrium through the coronary sinus. If the coronary sinus is not normally developed, the persistent left superior vena cava drains into the left atrium, and cyanosis results from the obligatory right-to-left shunt; this commonly occurs with an ASD or a complex cardiac anomaly.

Venous return above the renal veins can be abnormal with inferior vena cava interruption and azygos or hemiazygos continuation. In the former, inferior vena cava flow above the renal veins continues into the azygos vein, which courses normally up the right of the spine to empty into the junction between the superior vena cava and right atrium. In a less common anatomic arrangement, the caval flow empties into a hemiazygos vein, which empties into a persistent left superior vena cava. The finding rarely occurs in isolation but can be seen in patients with associated simple or complex malformations.

Anomalies of Pulmonary Venous Return

In partial anomalous pulmonary venous return, one or more but not all four pulmonary veins are not connected to the left atrium. The most common pattern has the right pulmonary veins connected to the superior vena cava, usually with a sinus venosus ASD. Anomalous connection of the right pulmonary veins to the inferior vena cava results in a chest radiographic shadow that resembles a Turkish sword, hence the designation *scimitar syndrome*. Associated anomalies include hypoplasia of the right lung, anomalies of the

bronchial system, hypoplasia of the right pulmonary artery, and dextroposition of the heart. Partial anomalous pulmonary venous return results in a left-to-right shunt physiology similar to that of an ASD.

In total anomalous pulmonary venous return, all the pulmonary veins connect abnormally to either the right atrium or one of the systemic veins above or below the diaphragm. Concurrent obstruction of the pulmonary veins is present when drainage occurs below the diaphragm and variable when drainage occurs above it. An ASD is essential to sustain life. One third of cases occur with major complex cardiac malformations.

In cor triatriatum, the pulmonary veins drain into an accessory chamber that is usually connected to the left atrium through an opening of variable size. The hemodynamic consequences are determined by the size of this opening and are similar to those of mitral stenosis. If symptoms of pulmonary venous hypertension occur, surgical treatment is indicated.

CARDIAC MALPOSITIONS

The normal heart is left sided and hence the designation *levocardia*. Cardiac malpositions are defined in terms of the intrathoracic position of the heart in relation to the position of the viscera (visceral situs), which are usually concordant with the position of the atria. That is, when the liver is on the right and the stomach is on the left, the atrium receiving systemic venous blood (right atrium) is right sided and the atrium receiving pulmonary venous blood (left atrium) is left sided. Asplenia and polysplenia syndromes are associated with a variety of complex cardiovascular malformations.

Dextrocardia and Mesocardia

In dextrocardia, the heart is on the right side of the thorax with or without situs inversus. When the heart is right sided with inverted atria, the stomach is right sided, and the liver is left sided, the combination is dextrocardia with situs inversus. In this arrangement, also called *mirror-image dextrocardia*, the ventricles are inverted, but so are the viscera and therefore the atria. The heart usually functions normally, and the diagnosis is often fortuitous. The heart sounds are louder on the right side of the chest, and the liver is palpable on the left. The chest radiograph shows a right-sided cardiac apex with a lower left hemidiaphragm and a right-sided stomach bubble. The ECG shows an inverted P and T wave in lead I with a negative QRS deflection and a reverse pattern between aVR and aVL. A mirror-image progression is seen from V_1 to a right-sided V_6 lead. An echocardiogram should be performed to ensure that intracardiac anatomy is normal.

When dextrocardia with situs solitus occurs, the ventricles are inverted but not the viscera and therefore not the atria. Associated severe cardiac malformations are typical.

In mesocardia, the heart is centrally located in the chest with normal atrial and visceral anatomy. The apex is central or rightward displaced on the chest radiograph. Typically, no associated cardiac malformations are present.

SPECIALIZED ISSUES

Endocarditis Prophylaxis

Prolonged survival of patients with complex congenital heart disease has resulted in a population at increased risk for infective endocarditis (Chapter 76). Adults with congenital heart disease should be informed about the risks of endocarditis. Any unexplained fever requires blood cultures to be drawn before antibiotics are initiated. Thorough transthoracic and transesophageal echocardiograms should be performed to assess the presence of vegetations. If infection of prosthetic material is suspected, early consultation with a specialist who has access to a congenital heart surgeon should be initiated because of the potential for rapid deterioration. The risk of endocarditis is highest in patients with cyanotic congenital heart disease and next highest in patients with endocardial cushion defects.[9]

Antibiotic prophylaxis before dental procedures that involve manipulation of the gingiva, periapical regions of the teeth, or mucosal tissue is indicated in patients with previous infective endocarditis, unrepaired cyanotic lesions, palliative shunts or conduits, prosthetic valves or prosthetic materials used for valve repair, repaired congenital heart disease with prosthetic material or transcatheter device within 6 months of intervention, and repaired congenital heart disease with residual lesions at or adjacent to the site of a prosthetic patch or device (Chapter 76). It is also reasonable to consider prophylaxis against endocarditis before vaginal delivery at the time of membrane rupture in such patients. Prophylaxis is not indicated for nondental procedures in the absence of active infection.

Exercise

The goal of exercise evaluation is to assess the functional results of therapeutic interventions and to provide guidelines for exercise prescriptions.[10]

TABLE 69-2 EXERCISE RECOMMENDATIONS IN ADULTS WITH CONGENITAL HEART DISEASE

CONDITION	UNRESTRICTED	LOW-MODERATE INTENSITY*	PROHIBITED
ASD†	No PHT; no arrhythmia; normal ventricular function	PA pressure >40 mm Hg *with* normal ETT; no arrhythmia	Eisenmenger
VSD†	Small; no PHT; no arrhythmia; normal ventricular function	Moderate VSD	Eisenmenger
PDA†	Small; no PHT; no arrhythmia; normal ventricular function	PA pressure >40 mm Hg *with* normal ETT; no arrhythmia	Eisenmenger
Coarctation‡	Gradient ≤20 mm Hg arm to leg; normal BP at rest and exercise	Gradient ≥20 mm Hg arm to leg *with* normal BP and normal ETT	Gradient ≥50 mm Hg arm to leg *or* aortic aneurysm
PS	Gradient <40 mm Hg; no arrhythmia; normal ventricular function	Gradient 40-60 mm Hg	Gradient ≥70 mm Hg *or* ventricular arrhythmia
AS	Gradient, <30 mm Hg; normal ECG; normal ETT; asymptomatic	Gradient 30-50 mm Hg *with* normal ECG, normal ETT; asymptomatic	Gradient >50 mm Hg *or* ventricular arrhythmia
TOF after repair	Normal RV pressure; no shunt; no arrhythmia	Increased RV pressure *or* moderate PR *or* SVT	RV pressure ≥65% systemic *or* ventricular arrhythmia on ETT *or* severe PR
Mustard or Senning		No cardiomegaly, arrhythmia, or syncope; normal ETT	Cardiomegaly *or* arrhythmia at rest or exercise
c-TGA unoperated	No cardiomegaly; mild TR; no arrhythmia; normal ETT	Moderate RV dysfunction, moderate TR; no arrhythmia	Severe TR *or* uncontrolled arrhythmia
Ebstein anomaly	Mild Ebstein; no arrhythmia; operated with mild TR	Moderate TR *with* no arrhythmia	Severe Ebstein *or* uncontrolled arrhythmia
Fontan		Normal O_2 saturation *with* near-normal ETT and ventricular function	Moderate-severe MR or TR *or* uncontrolled arrhythmia

*Based on peak dynamic and static components of exercise during competition for individual sports (see credit line).
†Unoperated or 6 months after surgery.
‡Unoperated or 1 year after surgery.
AS = aortic stenosis; ASD = atrial septal defect; BP = blood pressure; c-TGA = corrected transposition of the great arteries; ECG = electrocardiogram; ETT = exercise tolerance test; MR = mitral regurgitation; PA = pulmonary artery; PDA = patent ductus arteriosus; PHT = pulmonary hypertension; PR = pulmonary regurgitation; PS = pulmonary stenosis; RV = right ventricle; SVT = supraventricular tachyarrhythmia; TOF = tetralogy of Fallot; TR = tricuspid regurgitation; VSD = ventricular septal defect.
Based on guidelines recommended in Graham TP, Driscoll DJ, Gersony WM, et al. Task Force 2: Congenital heart disease. *J Am Coll Cardiol.* 2005;45:1326-1333.

Patients with residual hemodynamic lesions or unrepaired congenital cardiac anomalies should be evaluated on an annual basis with a physical examination, an ECG, and a cardiac ultrasonographic examination if indicated. Pertinent additional tests may include Holter monitoring and exercise testing. Attention should be directed to the detection of pulmonary hypertension, arrhythmias, myocardial dysfunction, and symptoms such as exercise-induced dizziness, syncope, dyspnea, or chest pain.

A series of exercise guidelines have been proposed for major groups of congenital heart defects (Table 69-2). Patients beyond 6 months after repair of a single shunt lesion without pulmonary hypertension, arrhythmias, or evidence of myocardial dysfunction can participate in all sports. With residual shunts, if the peak pulmonary artery pressure is less than 40 mm Hg in the absence of ventricular dysfunction or significant arrhythmias, patients can enjoy a free range of activity. Patients with elevated pulmonary vascular resistance are at risk of sudden death during intense exercise; although most self-limit their activity, participation in competitive sports is contraindicated. Patients with aortic and pulmonary stenosis should be counseled as recommended earlier, according to gradient severity. For patients with uncomplicated aortic coarctation, athletic participation is permitted if the arm-leg blood pressure gradient is 20 mm Hg or less at rest and the peak systolic blood pressure during exercise is normal. For patients after tetralogy of Fallot repair, repair of transposition of the great arteries, and the Fontan operation, exercise recommendations vary according to residual ventricular function and the presence or absence of arrhythmias. For such complex patients, care in a specialized center is associated with better outcomes.[11]

Grade A References

A1. Gatzoulis MA, Beghetti M, Galiè N, et al. Longer-term bosentan therapy improves functional capacity in Eisenmenger syndrome: results of the BREATHE-5 open-label extension study. *Int J Cardiol*. 2008;127:27-32.

A2. Attie F, Rosas M, Granados N, et al. Surgical treatment for secundum atrial septal defects in patients >40 years old: a randomized clinical trial. *J Am Coll Cardiol*. 2001;38:2035-2042.

A3. Furlan AJ, Reisman M, Massaro J, et al. Closure or medical therapy for cryptogenic stroke with patent foramen ovale. *N Engl J Med*. 2012;366:991-999.

A4. Meier B, Kalesan B, Mattle HP, et al. Percutaneous closure of patent foramen ovale in cryptogenic embolism. *N Engl J Med*. 2013;368:1083-1091.

A5. Carroll JD, Saver JL, Thaler DE, et al. Closure of patent foramen ovale versus medical therapy after cryptogenic stroke. *N Engl J Med*. 2013;368:1092-1100.

A6. Yang J, Yang L, Yu S, et al. Transcatheter versus surgical closure of perimembranous ventricular septal defects in children: a randomized controlled trial. *J Am Coll Cardiol*. 2014;63:1159-1168.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Marelli AJ, Ionescu-Ittu R, Mackie AS, et al. Lifetime prevalence of congenital heart disease in the general population from 2000 to 2010. *Circulation*. 2014;130:749-756.
2. Geva T, Martins JD, Wald RM. Atrial septal defects. *Lancet*. 2014;383:1921-1932.
3. Di Tullio MR, Jin Z, Russo C, et al. Patent foramen ovale, subclinical cerebrovascular disease, and ischemic stroke in a population-based cohort. *J Am Coll Cardiol*. 2013;62:35-41.
4. Rengifo-Moreno P, Palacios IF, Junpaparp P, et al. Patent foramen ovale transcatheter closure vs. medical therapy on recurrent vascular events: a systematic review and meta-analysis of randomized controlled trials. *Eur Heart J*. 2013;34:3342-3352.
5. Penny DJ, Vick GW 3rd. Ventricular septal defect. *Lancet*. 2011;377:1103-1112.
6. Brown ML, Burkhart HM, Connolly HM, et al. Coarctation of the aorta: lifelong surveillance is mandatory following surgical repair. *J Am Coll Cardiol*. 2013;62:1020-1025.
7. Wu M, Lu C, Chen H, et al. Arrhythmic burdens in patients with tetralogy of Fallot: A national database study. *Heart Rhythm*. 2015;12:604-609.
8. Khairy P, Clair M, Fernandes SM, et al. Cardiovascular outcomes after the arterial switch operation for D-transposition of the great arteries. *Circulation*. 2013;127:331-339.
9. Rushani D, Kaufman JS, Ionescu-Ittu R, et al. Infective endocarditis in children with congenital heart disease: cumulative incidence and predictors. *Circulation*. 2013;128:1412-1419.
10. Budts W, Börjesson M, Chessa M, et al. Physical activity in adolescents and adults with congenital heart defects: individualized exercise prescription. *Eur Heart J*. 2013;34:3669-3674.
11. Mylotte D, Pilote L, Ionescu-Ittu R, et al. Specialized adult congenital heart disease care: the impact of policy on mortality. *Circulation*. 2014;129:1804-1812.

REVIEW QUESTIONS

1. A 43-year-old man presents to the emergency department with a history of tetralogy of Fallot and syncope. He currently feels well. He has no chest pain. His 12-lead electrocardiogram (ECG) shows a right bundle branch block and three premature ventricular ectopic beats. What is your plan with respect to this patient's disposition?

A. You admit him to a monitored bed and consult a cardiologist with expertise in adult congenital heart disease.
B. You check his old ECG and find that this one is unchanged, so you send him home to follow up with his cardiologist as an outpatient.
C. You consult the neurologist because you are worried that given his history, he may have had a seizure.
D. You admit him for a cardiac catheterization because you know that tetralogy of Fallot may be associated with an anomalous coronary artery.
E. None of the above

Answer: A Patients with tetralogy of Fallot have up to a 5% risk of sudden cardiac death. Therefore, the presentation of syncope should always prompt an admission for monitoring to exclude a malignant ventricular arrhythmia. In addition, consultation with a congenital heart disease expert should always be initiated to exclude any new or progressive hemodynamic abnormality, such as pulmonary regurgitation that may require intervention. Although patients with tetralogy of Fallot may have coronary artery anomalies, syncope always warrants monitoring, and unexplained syncope may warrant an electrophysiology study.

2. A 50-year-old woman presents with shortness of breath to your office. One year ago, she had a moderate-sized myocardial infarction, for which she underwent successful dilation of a midcircumflex lesion with no complications. She denies angina. You examine her and find that her blood pressure is 140/90 mm Hg and her heart rate is 80 beats per minute. She has no evidence of systemic or pulmonary venous congestion. She has a positive hepatojugular reflux and a widely split second heart sound. She has an incomplete right bundle branch block on the ECG with no ischemic changes. You order an echocardiogram and find that she has an enlarged right side of the heart, which was not seen on her last echocardiogram 1 year ago. What investigations are you interested in?

A. You review the echocardiogram looking for evidence of a right-to-left shunt at the atrial level.
B. You order a cardiac catheterization to exclude disease of the right coronary artery.
C. You review the echocardiogram looking for evidence of a left-to-right shunt at the atrial level.
D. You order an exercise echocardiogram to see if there are new wall motion anomalies.
E. None of the above

Answer: C This is a classic presentation of an atrial septal defect that becomes symptomatic when diastolic dysfunction from acquired cardiovascular disease reduces left ventricular compliance, increases left ventricular end-diastolic pressure and left atrial pressure, and increases left-to-right shunting. Patients who have an atrial septal defect may be asymptomatic into their fourth and fifth decade until a condition such as hypertension, mitral regurgitation, aortic valve disease, cardiomyopathy, or myocardial infarction decreases left ventricular compliance and increases left-to-right shunting. Atrial septal defects can be easily missed on echocardiography if they are not specifically sought. The patient is not cyanotic, so an increase in right-to-left shunting may not be found. In addition, although the patient has coronary disease, an echocardiogram would still be the best initial test to evaluate ventricular function.

3. A 20-year-old man is referred to you for unexplained hypertension. He otherwise feels well. You notice he has a well-developed torso, and he tells you he lifts weights. His blood pressure is 150/90 mm Hg in both arms. He has a sustained apical impulse and a systolic ejection click. His ECG shows left ventricular hypertrophy. What is in your management plan?

A. You order an echocardiogram to exclude a bicuspid aortic valve and initiate β-blocker therapy.
B. You order an evaluation for renal artery stenosis and initiate therapy with an angiotensin-converting enzyme inhibitor.
C. You are unsure of what to do so you refer him to a local cardiologist.
D. You order an echocardiogram to exclude aortic coarctation, you order a 24-hour blood pressure monitor to confirm his hypertension, and you give him a follow-up appointment in 2 weeks with the intention of starting antihypertensive medication.
E. None of the above

Answer: D The finding of unexplained hypertension in a young man should always prompt a careful clinical examination to detect a possible brachial-to-femoral pulse delay and an echocardiogram to exclude aortic coarctation. Because blood flow is decreased in the lower extremities, these patients often have small hips relative to a well-developed torso. A significant portion of patients also have a bicuspid aortic valve, but it does not explain the hypertension. Because aortic coarctation is a reversible cause of hypertension, it is important to look for it and to address its complications.

4. A 45-year-old woman is referred to you from another general internal medicine practice where the internist is retiring. She is in a wheelchair with an oxygen tank and nasal prongs. She has bluish discoloration of her skin and clubbing of her fingernails. She explains that she has a heart condition and is here to have her monthly phlebotomy. What should you do?

A. You order the phlebotomy, and you have her come back in 1 month.
B. You review her medical record, take a history for symptoms of hyperviscosity, exclude dehydration, and order serum iron studies.
C. You make sure she does not smoke and consult a pulmonologist to determine if she needs the oxygen.
D. You go through the medical record and wonder why she has not had an operation for her cyanosis. You consult the nearest adult congenital heart specialist.
E. All of the above

Answer: B Patients with cyanotic congenital heart disease often undergo unnecessary phlebotomy. This clinical scenario requires a careful history to establish the presence or absence of hyperviscosity, which is an indication for phlebotomy in the absence of iron deficiency, and a discussion of the need or lack thereof for phlebotomy. Patients are often resistant initially; but after their hemoglobin levels stabilize and in the absence of iron deficiency, they often feel much better.

5. A 25-year-old woman with a bicuspid aortic valve and mild aortic regurgitation consults you before visiting her dentist. She requests a prescription for antibiotic prophylaxis before undergoing extensive dental cleaning. How do you respond?

A. You give her the prescription as she requests.
B. You call the dentist, who confirms the request, and you give her the prescription.
C. You take a careful history to be sure she is not allergic to penicillin.
D. You explain that antibiotic prophylaxis is no longer required for dental procedures and write her dentist a confirmatory letter.
E. None of the above

Answer: D Bacterial endocarditis prophylaxis is no longer recommended for isolated native valve lesions, such as a bicuspid aortic valve, in the absence of surgery or a previous episode of infective endocarditis. It is important to understand the rationale for the change in guidelines. It is not that infective endocarditis does not occur in patients with isolated lesions; rather, antibiotic prophylaxis cannot be shown to reduce the risk of this complication, especially because bacteremias after routine dental procedures are no worse than after brushing one's teeth. This situation requires careful discussion and accurate transmission of information to the patient.

70

ATHEROSCLEROSIS, THROMBOSIS, AND VASCULAR BIOLOGY

GÖRAN K. HANSSON AND ANDERS HAMSTEN

Atherosclerosis is the underlying cause of most cases of myocardial infarction, ischemic stroke, and peripheral arterial disease.[1] It is also a major cause of chronic heart failure and vascular dementia. Atherosclerosis, which is a chronic inflammatory response to the accumulation of lipid in the artery wall, is characterized by clinically silent intimal plaques that develop in arteries for years and even decades.[2] Fissuring or erosion of atherosclerotic plaques triggers the formation of a thrombus that accumulates during seconds to minutes to cause acute ischemia of the end organ. This ischemia, in turn, results in the dramatic clinical manifestations. It is estimated that approximately 90% of cases of myocardial infarction (Chapter 73), 60% of strokes (Chapter 407), most cases of heart failure (Chapter 58), and up to one third of all cases of dementia (Chapter 402) are due to atherosclerosis.

RISK FACTORS FOR ATHEROSCLEROSIS

The major risk factors that promote the development of atherosclerosis are an elevated low-density lipoprotein (LDL) cholesterol level, cigarette smoking, type 2 diabetes (Chapter 229), hypertension (Chapter 67), and a family history of coronary heart disease, ischemic stroke, or peripheral arterial disease. Other conditions that increase the risk of atherosclerotic disease or events include a low high-density lipoprotein (HDL) level (Chapter 206), abdominal obesity, hypertriglyceridemia, high plasma levels of lipoprotein (a), hyperfibrinogenemia, the inflammatory marker C-reactive protein, and physical inactivity (Chapter 52). Other emerging risk factors, including uric acid, psychosocial stress encompassing external stressors (e.g., job stress, life events, and financial problems), and reactions to stress (e.g., depression [Chapter 397], anxiety, psychosocial distress, and sleep disturbances [Chapters 100 and 405]), also appear to contribute. Elevation of plasma total homocysteine is also associated with increased cardiovascular risk, but it is possible that chronic renal dysfunction accounts for at least some of the vascular disease seen in hyperhomocysteinemia.

An atherogenic lipoprotein phenotype has been defined as the presence of a predominance of small, dense LDL particles, hypertriglyceridemia, and low plasma HDL cholesterol concentration. This lipoprotein phenotype, which is strongly linked to obesity, insulin resistance, hypertension, and abnormalities in postprandial lipoprotein metabolism, is similar to the so-called metabolic syndrome in that both are associated with a cluster of atherogenic and thrombotic risk factors—raised plasma levels of fibrinogen, plasminogen activator inhibitor-1, and coagulation factor VII as well as platelet hyperactivity. The inflammatory biomarker C-reactive protein, which is a predictor of cardiovascular events, is not causatively related to atherosclerosis but reflects ongoing inflammation, in atherosclerotic lesions or elsewhere in the body, that may accelerate the atherosclerotic process.

FORMATION OF ATHEROSCLEROTIC LESIONS

Atherosclerosis is thought to be initiated when apolipoprotein B–containing lipoproteins, predominantly LDL, accumulate in the vascular intima, the innermost layer of the artery (Fig. 70-1). Small, dense LDL particles are particularly prone to accumulate in the intima, where they associate with proteoglycans of the extracellular matrix. Lipoprotein lipase produced locally in the artery can bridge LDL to the extracellular matrix, and phospholipase and sphingomyelinase actions may contribute to the entrapment of LDL. Once trapped in the artery wall, LDL particles can be attacked by enzymes such as myeloperoxidase and NADPH oxidases; they may also be modified by nonenzymatic oxidation. During oxidative modification of LDL, certain biologically active oxidized phospholipid species are released and activate endothelial cells and macrophages. Such activation leads to production of chemokines and expression of leukocyte adhesion molecules that together instigate recruitment of monocytes and T cells to the intima. Local growth factors induce recruited monocytes to develop into macrophages.

In the intima, macrophages take up oxidized LDL through their scavenger receptors, start to accumulate cholesterol, and are gradually transformed into cholesterol-laden foam cells.[3] Some macrophages in the intima produce proinflammatory mediators, including tumor necrosis factor (TNF), interleukin-1, proinflammatory eicosanoids, radical oxygen and nitrogen species, and prothrombotic factors. At least some of this inflammatory activity may be instigated when intracellular cholesterol microcrystals in the macrophage activate the inflammasome machinery that generates the inflammatory mediator interleukin-1β.

T cells that are stimulated to enter the intima may recognize antigens presented by macrophages.[4] These antigens include components of LDL, other endogenous proteins, and possibly microbial antigens. Activated intimal T cells produce T_H1-type cytokines, such as interferon-γ, TNF, and lymphotoxin, all of which are strongly proatherogenic. For example, interferon-γ release also inhibits collagen fiber formation and smooth muscle proliferation. With the entry and activation of T cells and macrophages, the accumulation of lipid in the intima leads to the chronic inflammatory disease process of atherosclerosis.

Although adaptive immunity is believed to exert a net proatherogenic effect, antiatherogenic immune responses against LDL involve activation of regulatory T cells, secretion of the anti-inflammatory cytokines interleukin-10 and transforming growth factor-β, and production of anti-LDL antibodies.[5] In addition to T cells and macrophages, atheroma formation is also stimulated by dendritic cells that take up and present antigen and by mast cells that secrete enzymes and bioactive mediators.

Triglyceride-rich lipoprotein remnant particles, which have adverse effects on endothelial function, penetrate into the subendothelial space of normal intima and atherosclerotic plaques, where they are retained. Inflammation

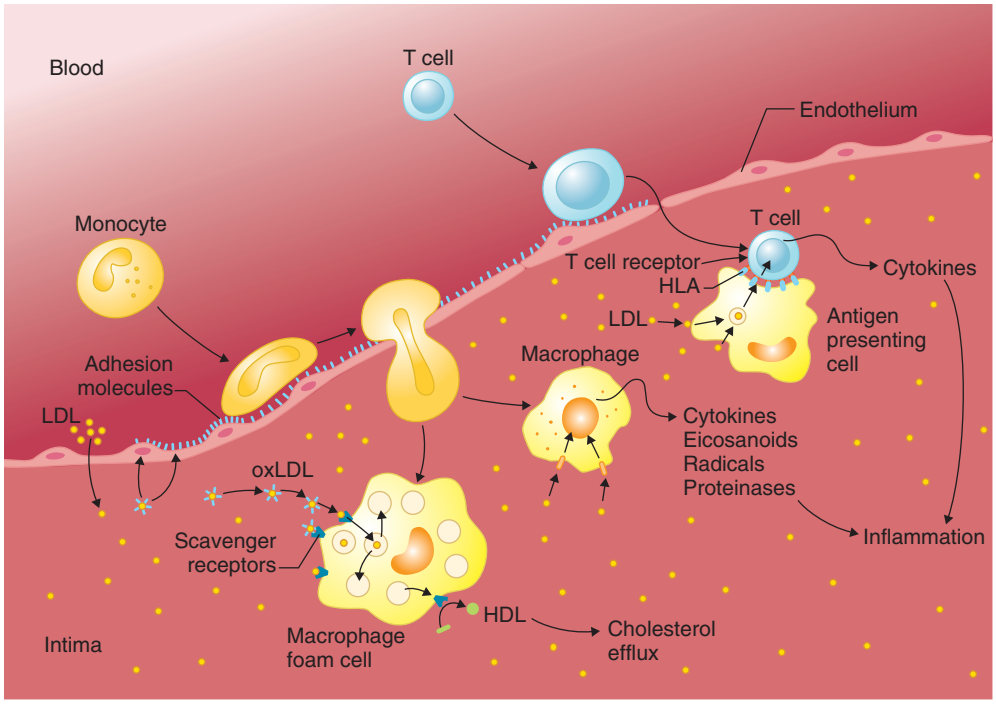

FIGURE 70-1. Formation of atherosclerotic plaques. Low-density lipoproteins (LDL) transit from the blood stream to the arterial intima and accumulate under the endothelial cell layer. LDL particles undergo oxidative modification in the intima (denoted by spikes on LDL particles), thereby leading to their binding to scavenger receptors and uptake by macrophages, which accumulate cholesterol and develop into foam cells. Cholesterol efflux to high-density lipoprotein (HDL) counteracts the tendency to foam cell formation. Molecules released from oxidatively modified LDL (oxLDL) activate endothelial cells to express leukocyte adhesion molecules that promote binding of monocytes and T cells to the surface of the artery. Chemokines stimulate monocytes and T cells to migrate into the intima, where the monocytes differentiate into macrophages. Although many macrophages develop into foam cells, some are activated, thereby leading to release of proinflammatory cytokines, eicosanoids, radicals, and proteases. T cells entering through mechanisms similar to those of monocytes can recognize local antigens, such as LDL components, which are presented to them by antigen-presenting cells (dendritic cells and macrophages) that express human leukocyte antigen (HLA) molecules. T cells whose receptors can recognize local antigens are activated, thereby leading to release of a host of cytokines that can activate macrophages and enhance vascular inflammation. (Modified from Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med.* 2005;352: 1685-1695.)

may increase the levels of such lipoprotein particles in blood by inhibiting their clearance. The LDL-like lipoprotein (a) particle exerts both proatherogenic and prothrombotic actions.

Conversely, antiatherogenic HDL particles counteract the formation of atherosclerotic lesions in model systems.[6] These particles mediate cholesterol efflux from cells by acting as acceptors of cholesterol delivered from specific transport proteins termed adenosine triphosphate–binding cassette (ABC) A1 and G1. In addition, HDL particles carry anti-inflammatory and antioxidant proteins.

GROWTH, DEATH, AND THE PROGRESSION OF DISEASE

Early atherosclerotic lesions grow by the accumulation of cholesterol; the infiltration of inflammatory cells; the activation, proliferation, and death of such cells; and the gradual development of a core that contains cellular debris and lipids. As a tissue response to this process, smooth muscle cells form a subendothelial cap structure dominated by collagen fibers that are produced by these cells. The collagen cap mechanically stabilizes the plaque and creates a barrier between the hemostatic components of the blood and the thrombogenic material of the plaque. Until the plaque is far advanced, compensatory enlargement ("remodeling") of the arterial wall prevents it from significantly protruding into the arterial lumen. However, after the plaque has enlarged to a sufficient size, the lumen narrows as the plaque grows, and the artery remodels inward, often accompanied by exaggerated or paradoxical vasoconstriction.

PLAQUE ACTIVATION, THROMBOSIS, AND INFARCTION

The atherosclerotic process typically is silent for months, years, and even decades, and it may never result in clinical manifestations. However, if the plaque's surface is damaged, thrombotic occlusion of the artery may ensue.[7] Surface continuity may be damaged by fissuring (so-called plaque rupture, observed in 60 to 80% of cases of acute coronary syndrome) or surface erosion (present in 20 to 40% of cases with coronary thrombosis, especially in women and young victims of sudden coronary death). Recent studies suggest that the proportion of infarctions caused by rupture vs erosion is changing, with more cases due to erosion and fewer to overt plaque rupture.[8] Fissures and erosions trigger atherothrombosis by exposing thrombogenic material inside the plaque, such as phospholipids, tissue factor, and matrix molecules, to platelets and coagulation factors (Fig. 70-2). Platelet aggregates that form on these exposed surfaces are stabilized by a fibrin network. Tissue factor, expressed in vascular smooth muscle cells and macrophages of the atherosclerotic plaque, is the primary cellular initiator of the blood coagulation cascade that leads to fibrin formation. Atherothrombi expand rapidly and can fill the lumen within minutes, thereby leading to ischemia and infarction.

A range of factors may contribute to atherothrombosis.[9,10] Disturbance of the balance between prothrombotic and fibrinolytic activity on the plaque surface probably plays an important role for precipitating the thrombotic event, but the precise sequence of events that operate in vivo is not yet known.

The cause of plaque rupture also remains unclear. Clinical studies have associated ischemic atherothrombotic events such as myocardial infarction (Chapter 73) and stroke (Chapter 407) with infections and stressful events. Histopathologic analysis shows increased inflammation with infiltration of macrophages, activated T cells, dendritic cells, and mast cells as well as reduced thickness of the fibrous cap and increased neovascularity at sites of plaque rupture and thrombosis. Matrix metalloproteinases and cysteine proteinases, which are produced by macrophages, are found at sites of plaque rupture and have been implicated in rupture, but their effects on the composition and size of lesions are complex. Cell death alone may be an important trigger of plaque rupture. Apoptotic cells contained in the plaque are usually removed by efferocytosis. If this process fails, secondary necrosis ensues, thereby leading to reduced mechanical integrity and accumulation of prothrombotic material from dead cells.

Ruptured plaques also tend to have a large necrotic lipid core. In contrast, plaques underlying erosions do not have a large lipid core and show less inflammation compared with ruptured plaques. Plaque rupture frequently occurs without clinical manifestations, possibly reflecting variation in the

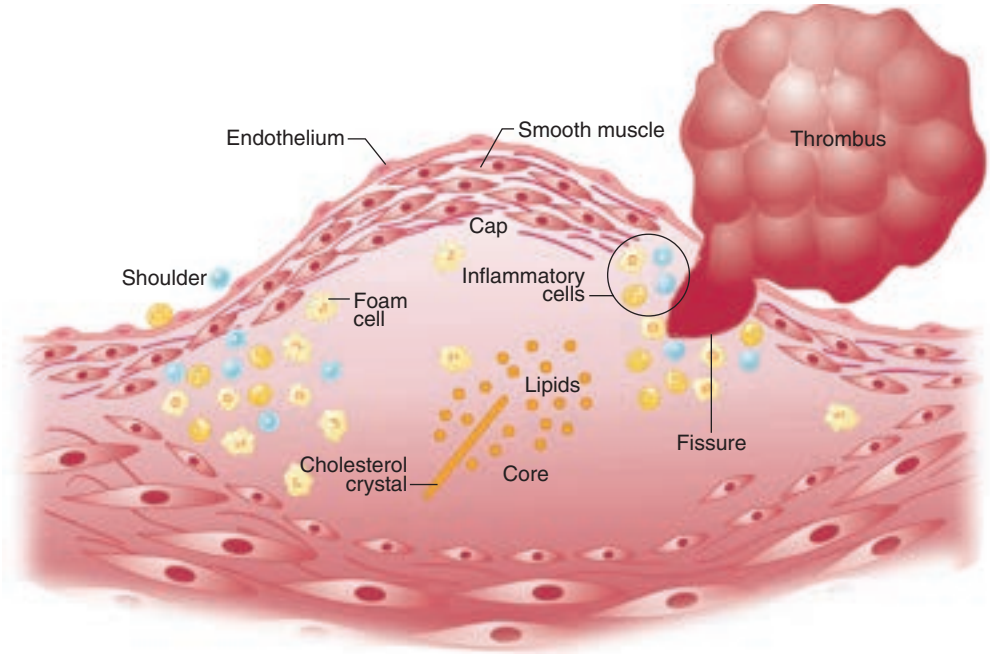

FIGURE 70-2. Plaque rupture and atherothrombosis. The advanced atherosclerotic plaque has a central core with lipids (especially cholesterol), live and dead cells, necrotic material from dead foam cells, and calcium salts. The plaque is overlaid by a fibrous cap that consists of smooth muscle cells and collagen (produced by the muscle cells) and covered by an intact layer of endothelial cells. Inflammatory cells (macrophages, T cells, mast cells, dendritic cells, and occasional B cells) are interspersed with these components and are particularly abundant in the shoulder regions of plaques, where fissures (also called ruptures) may expose thrombogenic core material (e.g., lipids, collagen, tissue factor) to blood components. This event triggers platelet aggregation and humoral coagulation, thereby leading to thrombus formation at the site of fissuring. Thrombi may expand locally to obstruct blood flow or they may detach to cause embolization. (Modified from Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med.* 2005;352:1685-1695.)

thrombotic response depending on the thrombogenicity of exposed plaque constituents, local hemorrheology, shear-induced platelet activation, systemic clotting activity, fibrinolytic function, and sensitivity of the end organ to ischemia.

PRINCIPLES OF ANTIATHEROSCLEROTIC THERAPY

Current treatment of atherosclerosis aims to control risk factors and to maintain or to restore perfusion in affected arteries. However, progress in understanding the pathogenesis of atherosclerosis is expected to result in more direct approaches. To date, firmly established interventions include smoking cessation, dietary and pharmacologic reduction of LDL cholesterol (Chapter 206), and management of blood pressure (Chapter 67). Available data also strongly support intervention directed toward hyperglycemia (Chapter 229), hypertriglyceridemia (Chapter 206), obesity (Chapter 220), and physical inactivity (Chapter 16).

Cholesterol-lowering statins clearly reduce atherosclerotic lesions and inhibit their progression.[A1] Statins also prevent nitroglycerin-induced endothelial dysfunction and nitrate tolerance, and they inhibit immune activity and inflammation.[A2] Aspirin and other inhibitors of platelet aggregation, β-adrenergic receptor blockers, and angiotensin-converting enzyme inhibitors or angiotensin II antagonists are also part of the routine secondary prevention of coronary heart disease (Chapters 71, 72, and 73). Inhibitors of platelet aggregation are widely used for secondary prevention of atherosclerotic cardiovascular disease. Aspirin inhibits formation of proaggregatory prostaglandins, whereas other inhibitors of platelet aggregation modulate expression of platelet adhesion molecules. Nitroglycerin and similar compounds that mimic the action of endogenous nitric oxide remain the most important vasodilators used in secondary prevention (Chapter 71). Eicosapentaenoic acid treatment has also shown promising results in secondary prevention.

FUTURE DIRECTIONS

Novel therapeutic approaches include new lipid-lowering treatment, immunosuppressive and anti-inflammatory compounds, and vaccination with disease-related antigens. Investigational agents targeting atherogenic lipoproteins include proprotein convertase subtilisin kexin type 9 (PCSK9),[A3] inhibitors of squalene synthase, microsomal triglyceride transfer protein, and antisense oligonucleotides to apolipoprotein B. In contrast, therapies to raise HDL levels have been disappointing (Chapter 206). Peroxisome proliferator-activated receptor agonists (Chapter 229), besides their beneficial effects on lipid and blood glucose levels, have shown direct antiatherosclerotic effects in experimental studies. However, inhibitors of secreted and lipoprotein-associated phospholipase A_2 have not reduced cardiovascular events in clinical trials.[A4]

Members of the interleukin-1 and TNF families of proinflammatory proteins, eicosanoids, and cell surface proteins promoting antigen-specific T-cell activation are particularly promising targets of anti-inflammatory therapy, whereas stimulation of anti-inflammatory signaling pathways represents a different potential antiatherosclerotic therapy. Patients whose asthma is treated with the leukotriene receptor blocker montelukast have a reduced risk for ischemic stroke and myocardial infarction, but whether it or other anti-inflammatory agents will be clinically useful for reducing cardiovascular events is uncertain.

Vaccination against immunogenic epitopes in the protein and lipid moieties of LDL may induce anti-inflammatory regulatory immunity and also reduce LDL uptake in cells of the atherosclerotic lesion. Identification of genes that increase the risk of coronary artery disease by genome-wide association studies is likely to generate a new set of potential targets.[11]

Grade A References

A1. Taylor F, Huffman MD, Macedo AF, et al. Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev*. 2013;1:CD004816.
A2. Liuni A, Luca MC, Di Stolfo G, et al. Coadministration of atorvastatin prevents nitroglycerin-induced endothelial dysfunction and nitrate tolerance in healthy humans. *J Am Coll Cardiol*. 2011;57:93-98.
A3. Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med*. 2008;359:2195-2207.
A4. Blom DJ, Hala T, Bolognese M, et al. A 52-week placebo-controlled trial of evolocumab in hyperlipidemia. *N Engl J Med*. 2014;370:1809-1819.
A5. White HD, Held C, Stewart R, et al. Darapladib for preventing ischemic events in stable coronary heart disease. *N Engl J Med*. 2014;370:1702-1711.
A6. Nicholls SJ, Kastelein JJ, Schwartz GG, et al. Varespladib and cardiovascular events in patients with an acute coronary syndrome: the VISTA-16 randomized clinical trial. *JAMA*. 2014;311:252-262.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Libby P, Ridker PM, Hansson GK. Progress and challenges in translating the biology of atherosclerosis. *Nature*. 2011;473:317-325.
2. Libby P, Lichtman AH, Hansson GK. Immune effector mechanisms implicated in atherosclerosis: from mice to humans. *Immunity*. 2013;38:1092-1104.
3. Moore KJ, Tabas I. Macrophages in the pathogenesis of atherosclerosis. *Cell*. 2011;145:341-355.
4. Duewell P, Kono H, Rayner KJ, et al. NLRP3 inflammasomes are required for atherogenesis and activated by cholesterol crystals. *Nature*. 2010;464:1357-1361.
5. Hansson GK, Hermansson A. The immune system in atherosclerosis. *Nat Immunol*. 2011;12:204-212.
6. van Capelleveen JC, Brewer HB, Kastelein JJ, et al. Novel therapies focused on the high-density lipoprotein particle. *Circ Res*. 2014;114:193-204.
7. Libby P. Mechanisms of acute coronary syndromes and their implications for therapy. *N Engl J Med*. 2013;368:2004-2013.
8. Giugliano RP, Braunwald E. The year in acute coronary syndrome. *J Am Coll Cardiol*. 2014;63:201-214.
9. Puri R, Nicholls SJ, Ellis SG, et al. High-risk coronary atheroma: the interplay between ischemia, plaque burden, and disease progression. *J Am Coll Cardiol*. 2014;63:1134-1140.
10. Malarstig A, Hamsten A. Genetics of atherothrombosis and thrombophilia. *Curr Atheroscler Rep*. 2010;12:159-166.
11. Deloukas P, Kanoni S, Willenborg C, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45:25-33.

71

ANGINA PECTORIS AND STABLE ISCHEMIC HEART DISEASE

WILLIAM E. BODEN

DEFINITION

Ischemic heart disease is most commonly caused by obstruction or stenosis of one or more of the epicardial coronary arteries by atheromatous plaque (Chapter 70). Obstruction can result in myocardial ischemia and may culminate in infarction (Chapters 72 and 73) with associated symptoms of angina, dyspnea, heart failure (Chapter 58), arrhythmic complications (Chapters 63, 64, and 65), and ultimately death.

Angina pectoris is generally a consequence of a supply-demand imbalance: an activity increases cardiac workload or "demand," thereby resulting in an increase in heart rate, blood pressure, and contractility, leading to an increase in left ventricular (LV) wall tension; but stenotic coronary arteries (Chapter 57) are unable to augment antegrade flow or "supply" adequately in response to this increase in demand. Such an imbalance classically results in chest discomfort (Chapter 51) of varying intensity and duration. Angina pectoris is generally defined as a discomfort in the chest or adjacent areas caused by myocardial ischemia. Often, angina is described incorrectly as "chest pain." The term *angina*, however, derives from a neologism of two Latin words, *angor animi*, which literally translates into "fear of life being extinguished ('from the breast')," according to Heberden original description in 1768. Had Heberden been trying to convey the literal term for chest pain, he would more likely have used the Latin term *dolor pectoris*.

Grading of Angina Pectoris

The Canadian Cardiovascular Society (CCS) angina grading scale is a widely used four-point ordinal scale that classifies angina pectoris from mild (class I: angina occurring only during strenuous or prolonged physical activity) to severe (class IV: inability to perform any activity without angina, or angina at rest) and includes the full spectrum of angina from chronic stable to unstable (see Table 51-5). Operationally, the CCS angina scale permits clinicians to categorize patients as mild or stable (generally CCS classes I and II) versus severe or unstable (typically CCS classes III and IV). For classification purposes, subjects with stable ischemic heart disease generally exhibit CCS class I-II symptoms, which are typically provoked by exertion. Other grading systems include a specific activity scale, which is based on the metabolic cost of specific activities, and an anginal score, which integrates the clinical features and tempo of angina with electrocardiographic changes and offers independent prognostic information beyond that provided by age, gender, ventricular function, and coronary anatomy.

EPIDEMIOLOGY

An estimated 15 million Americans currently have coronary heart disease (Chapter 51).[1] Among individuals aged 45 years and older, the incidence of stable angina pectoris is about 500,000 per year, of whom about 65% are men. Nearly 8 million Americans have prevalent angina pectoris. Although deaths attributable to coronary disease have declined in the United States during the past several decades, ischemic heart disease is now the leading cause of death worldwide, and it is expected that this rate of rise will continue to accelerate in the coming decade as a consequence of the epidemic rise in obesity (Chapter 220), type 2 diabetes (Chapter 229), and the metabolic syndrome, which may give rise to an increasing risk for development of premature coronary artery disease in younger generations. An estimated 7 million patients went to emergency departments in 2010 for chest pain, and approximately 1.5 million of them were hospitalized with the acute coronary syndrome (ACS; Chapter 72).[2]

PATHOBIOLOGY

Angina is the most frequent clinical expression of myocardial ischemia. Ischemia, which rapidly develops when a mismatch arises between myocardial oxygen needs and myocardial oxygen supply, can be manifested clinically in many different ways besides angina, from no symptoms (e.g., silent myocardial ischemia) to unstable angina, myocardial infarction (MI), or sudden cardiac death. It may remain stable for many years in selected patients or may be rapidly progressive with an abrupt change in frequency and tempo during days to weeks. Conversely, atherosclerosis, which is the most common cause of myocardial ischemia, may evolve for years without any manifestations of ischemia.

In contrast to the inherent pathogenetic complexity mediated by differing mechanisms associated with abrupt plaque rupture, fissuring, or erosion in patients with ACS (Chapters 70 and 72), the pathogenesis of chronic stable angina is, by comparison, seemingly less complicated and heterogeneous because it is a consequence of a myocardial supply-demand mismatch. In most patients with stable ischemic heart disease, atherosclerosis involves a fundamentally different histopathologic process (small necrotic lipid core with an overlying thick or very thick fibrous cap and a low proclivity to plaque rupture) compared with ACS, in which the principal histopathologic picture is that of a large lipid core subtended by a thin-capped fibroatheroma, which harbors the high-risk or vulnerable plaque with a high proclivity for rupture (Chapter 70).

Two major pathogenetic mechanisms may result in myocardial ischemia and angina in the chronic setting: so-called *demand angina*, which is caused by an increase in myocardial oxygen requirements and workload; and *supply angina*, which is caused by diminished oxygen delivery to myocardial tissue. Demand angina is a consequence of the increased myocardial oxygen requirements that occur with increased physical activity, emotion, or stress. In a patient with chronic, restricted oxygen delivery due to atherosclerotic narrowing of a coronary artery, this increased demand may precipitate angina. Other extracardiac precipitants of angina include the excessive metabolic demands imposed by fever, thyrotoxicosis (Chapter 226), severe anemia (Chapter 158) from blood loss, tachycardia from any cause (Chapters 62, 63, and 64), hypoglycemia (Chapter 230), and pain.

By contrast, supply angina may occur in patients with either unstable angina (Chapter 72) or chronic stable angina by transient reductions in myocardial oxygen delivery as a consequence of coronary vasoconstriction with resulting dynamic coronary stenosis. In the presence of coronary luminal narrowing due to atherosclerosis, superimposed platelet thrombi and leukocytes may elaborate vasoconstrictor substances, such as serotonin and thromboxane A_2, whereas endothelial damage in diseased coronary arteries may decrease production of vasodilator substances such as nitric oxide and adenosine. The result is an abnormal physiologic vasoconstrictor response to exercise and other stimuli, such as exogenously administered adenosine, or the paradoxical vasoconstrictor response to the typical flow-mediated reactive hyperemia associated with brachial artery compression. In some clinical settings, patients who have normal coronary arteries or non–flow-limiting stenoses may exhibit dynamic obstruction alone, which can cause myocardial ischemia and result in angina at rest (Prinzmetal's [variant] angina). Conversely, in patients with severe fixed obstruction to coronary blood flow, only a minor increase in dynamic obstruction can reduce blood flow below a critical level and cause myocardial ischemia.

The pathophysiologic basis for angina and ischemia in patients with stable ischemic heart disease has important implications for the selection of antiischemic agents. The greater the contribution from increased myocardial oxygen requirements to the imbalance between supply and demand, the greater the likelihood that agents that reduce heart rate and wall tension, such as β-blockers or non-dihydropyridine calcium antagonists, will provide clinical benefit, whereas nitrates and calcium antagonists with more potent vasodilatory properties (particularly the dihydropyridines) will be more beneficial to alleviate angina and ischemia mediated by coronary vasoconstriction.

Although the most common cause of ischemic heart disease is atherosclerotic narrowing of the coronary arteries resulting in flow-limiting obstruction to epicardial blood flow, obstructive coronary artery disease may also have nonatherosclerotic causes, such as congenital abnormalities of the coronary arteries (Chapter 69), vasospasm, myocardial bridging, coronary arteritis in association with systemic vasculitides (Chapter 270), and radiation-induced coronary disease (Chapter 20). Radiation therapy for thoracic or mediastinal malignant neoplasms, particularly breast cancer and Hodgkin lymphoma, can cause long-term coronary microangiopathy and macroangiopathy. There is a four- to seven-fold increase in clinically significant, high-grade coronary artery stenosis of the mid and distal left anterior descending coronary artery in women with irradiated left-sided breast cancer compared with those treated with radiation for right-sided breast cancer. Myocardial ischemia and angina pectoris may also occur in the absence of obstructive coronary artery disease, as in the case of aortic valve disease (Chapter 75), hypertrophic

cardiomyopathy (Chapter 60), and dilated cardiomyopathy. Moreover, ischemic heart disease may coexist with these other forms of heart disease.

CLINICAL MANIFESTATIONS

History

It is important to recognize that there are many causes of chest discomfort (see Table 51-2), that angina-like chest pain may not represent ischemic heart disease (Table 71-1), that ischemic heart disease causes symptoms other than anginal pain (Table 71-2), and that nonatherosclerotic coronary artery abnormalities may cause ischemic chest pain (Table 71-3).

Angina pectoris has four cardinal clinical features: the character of the discomfort, its site and distribution, its provocation, and its duration. The character of anginal discomfort is typically described as a pressure sensation that conveys a feeling of strangling and anxiety (Chapter 51), but patients may also use descriptors such as heaviness, squeezing, constricting, viselike, suffocating, and crushing. In some patients, especially women and the elderly, the quality of the sensation is more vague and atypical. Some patients may describe the discomfort as a burning sensation in the mid-epigastrium or as an uncomfortable, numb sensation. *Anginal equivalents* (i.e., symptoms of myocardial ischemia other than angina), such as dyspnea, fatigue, lightheadedness or dizziness, and gastric eructations, also may be manifestations of ischemic heart disease.

TABLE 71-1 PROBABILITY (%) OF CORONARY ARTERY DISEASE BY AGE, GENDER, AND SYMPTOMS

GENDER	AGE (yr)	DEFINITE ANGINA	ATYPICAL ANGINA	NONCARDIAC CHEST PAIN
Men	30-39	83	46	3
	40-49	88	57	12
	50-59	94	71	18
	60-69	95	78	31
	≥70	97	94	63
Women	30-39	—	20	4
	40-49	56	31	4
	50-59	68	30	6
	60-69	81	48	10
	≥70	96	56	—

From Chaitman BR, Bourassa MG, Davis K, et al. Angiographic prevalence of high-risk coronary artery disease in patient subsets (CASS). *Circulation*. 1981;64:360-367.

TABLE 71-2 NON–CHEST PAIN SYMPTOMS OF CHRONIC ISCHEMIC HEART DISEASE

DYSPNEA
Dyspnea on exertion
Dyspnea at rest
Paroxysmal nocturnal dyspnea
Temporal change of increasing exertional dyspnea with declining effort tolerance
NON–CHEST LOCATIONS OF DISCOMFORT (EITHER EXERTIONAL OR AT REST)
Neck or mandibular discomfort or pain
Throat tightness
Shoulder discomfort
Upper arm or forearm discomfort (more often left-sided)
Interscapular or infrascapular discomfort
MID-EPIGASTRIC OR ABDOMINAL
Mid-epigastric burning, often postprandially
Sharp abdominal pain (atypical, but more common in women)
Right upper quadrant discomfort (may mimic gallbladder disease or pancreatitis)
Nausea or vomiting (often associated with increased vagal tone secondary to inferior myocardial ischemia or infarction)
DIAPHORESIS
EXCESSIVE FATIGUE AND WEAKNESS
Often a discernible prodrome of increasing fatigue with declining effort tolerance
DIZZINESS AND SYNCOPE
Uncommon, unless precipitated or exacerbated by alterations in heart rate or rhythm (e.g., bradyarrhythmia, tachyarrhythmia, heart block), blood pressure (e.g., hypotension), or cardiac output (e.g., decreased cerebral perfusion)

The site and distribution of anginal discomfort are predominantly midsternal or retrosternal but can be precordial. Radiation is common, usually to the left side of the neck and shoulder and down the ulnar surface of the left arm; radiation to the right arm is less common. Discomfort radiating to the jaw is common and must be distinguished from dental pain. Epigastric discomfort alone or in association with chest pressure may occur. Provocation of angina is classically caused by physical exertion or activity, emotional stress, exposure to the cold, sexual intercourse, or eating a large meal. Angina that occurs at rest or nocturnally often heralds a change in the pattern from stable to unstable and may indicate that there is an incipient plaque rupture leading to ACS. Vasospastic (or Prinzmetal's) angina may occur spontaneously at rest or nocturnally without provocation.

The typical duration of an episode of angina pectoris is brief. An episode usually begins gradually and reaches its maximal intensity during a period of minutes before abating. It is unusual for angina pectoris to peak and trough in less than a minute, and it is common that patients with exertional angina usually prefer to rest, to sit, or to stop walking during episodes that may be precipitated by the offending activity. Chest discomfort that persists for more than 15 to 20 minutes, especially at rest or nocturnally, is likely to represent ACS or MI. By contrast, features that suggest a noncardiac etiology of angina pectoris include pleuritic pain, pain reproduced by movement or palpation of the chest wall or arms, sharp or constant pain lasting for many hours, pain or discomfort that a patient can localize to the chest wall with the tip of one finger, and very brief episodes of pain lasting seconds (Chapter 51). Typical angina pectoris is generally relieved within minutes by rest or the use of sublingual, oral, or cutaneous nitroglycerin. The response to sublingual nitroglycerin is often a helpful diagnostic tool, although some noncardiac pain (e.g., esophageal spasm) may also respond to nitroglycerin.

Although chest discomfort is usually the predominant symptom in ischemic heart disease, chest discomfort may be absent, atypical, or not prominent in some patients. Patients with stable ischemic heart disease may complain predominantly or exclusively of dyspnea, diminishing exercise tolerance, fatigue, or weakness. Others will first present with an abnormal exercise test result or other evidence of myocardial ischemia without any symptoms. Some patients may present with cardiac arrhythmias or even sudden cardiac death.

Physical Examination

Many patients with stable ischemic heart disease present with normal physical findings, but a diligent physical examination may reveal findings that represent either the consequences of myocardial ischemia or evidence of risk factors for coronary artery disease. Inspection of the eyes may reveal a corneal arcus, and examination of the skin may show xanthomas (see Fig. 51-12). Retinal arteriolar changes are common in patients with coronary artery disease who have hypertension or diabetes mellitus (see Figs. 423-26 and 423-24).

The cardiac examination is generally of limited benefit in evaluating patients with chest pain or establishing a diagnosis of ischemic heart disease.

TABLE 71-3 NONATHEROSCLEROTIC CAUSES OF ISCHEMIC CHEST PAIN

PRIMARY CARDIAC CAUSE
Coronary artery abnormalities
Coronary spasm
Coronary arteritis
Coronary dissection
Coronary artery anomalies
Radiation-induced coronary disease
Myocardial bridging
Aortic stenosis
Hypertrophic cardiomyopathy
Dilated cardiomyopathy
Tachycardia
PRIMARY NONCARDIAC CAUSE
Anemia
Sickle cell disease
Hypoxemia
Carbon monoxide poisoning
Hyperviscosity (e.g., polycythemia)
Hyperthyroidism
Pheochromocytoma

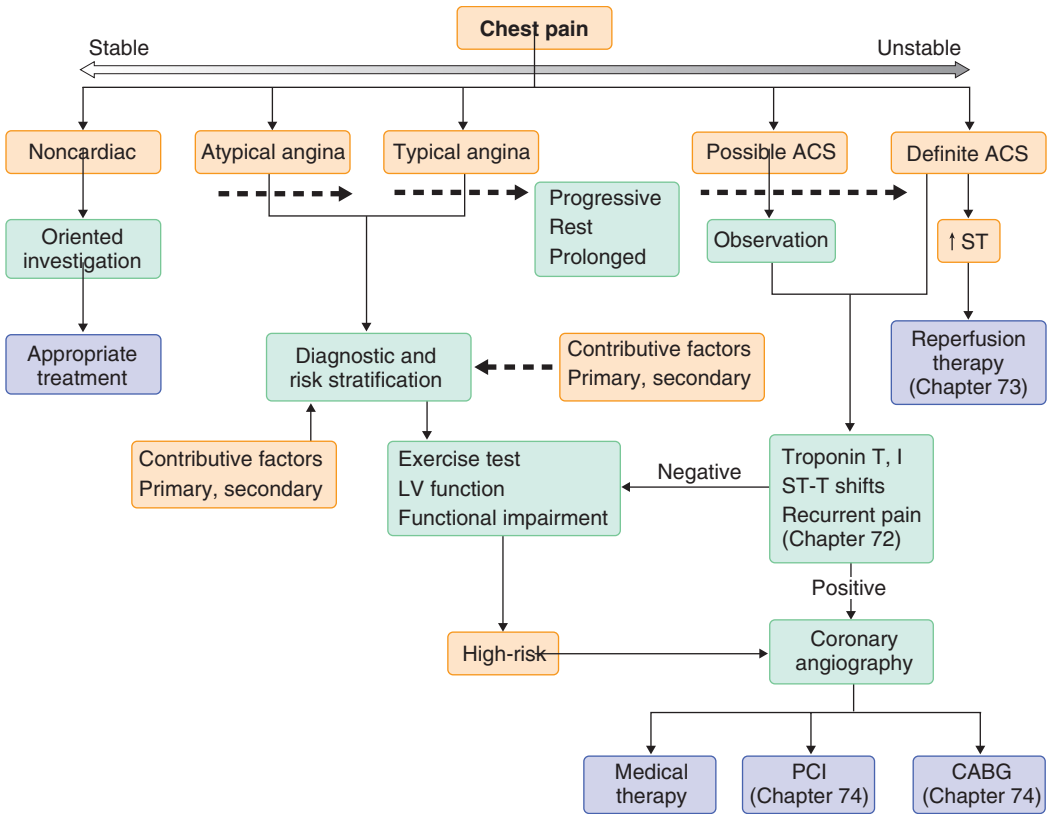

FIGURE 71-1. Evaluation of chest pain. ACS = acute coronary syndrome; CABG = coronary artery bypass graft; LV = left ventricular; PCI = percutaneous coronary intervention. (Modified from Théroux P. Angina pectoris. In: Goldman L, Ausiello DA, eds. Cecil Textbook of Medicine. 23rd ed. Philadelphia: Saunders Elsevier; 2008.)

During an episode of chest discomfort, myocardial ischemia may produce either a third or fourth heart sound.

Myocardial ischemia also can cause a transient holosystolic or mid-late systolic apical murmur due to reversible papillary muscle dysfunction that results in mitral regurgitation. These murmurs are more prevalent in patients with extensive coronary artery disease, especially with inferior or inferoposterior ischemia due to right coronary artery disease. It is important to distinguish such a murmur from the murmur of aortic stenosis or obstructive hypertrophic cardiomyopathy (see Tables 51-7 and 51-8). A displaced LV apical impulse, particularly if dyskinetic, is a sign of significant LV systolic dysfunction.

If patients have coexisting heart failure, an elevated jugular venous pressure, pulmonary rales, and peripheral edema may be present (Chapter 58). The physical examination may reveal other implicating or contributing conditions, such as thyroid enlargement (Chapter 226) or severe anemia (Chapter 158).

DIAGNOSIS

In addition to a careful history and physical examination, assessment of patients with stable ischemic heart disease includes the 12-lead electrocardiogram (ECG), measurement of biochemical and inflammatory markers, and noninvasive diagnostic testing.[3,4] The first goal is to assess the patient's probability of ischemia so that an appropriate evaluation can expedite effective therapy (Fig. 71-1).

Resting Electrocardiogram

Although there may be focal, diagnostic findings of ST segment depression or T wave inversions (Fig. 71-2) on the resting ECG in chronic ischemic heart disease, even patients with extensive anatomic coronary artery disease may have a normal tracing at rest. In addition to myocardial ischemia, other conditions that can produce ST-T wave abnormalities include LV hypertrophy and dilation due to long-standing hypertension and valvular heart disease (e.g., aortic stenosis, hypertrophic cardiomyopathy), electrolyte abnormalities, neurogenic effects, and antiarrhythmic drugs. The presence of new ST-T wave abnormalities on the resting ECG, however, can be helpful in the diagnosis of coronary artery disease and may correlate with the severity of the underlying heart disease.

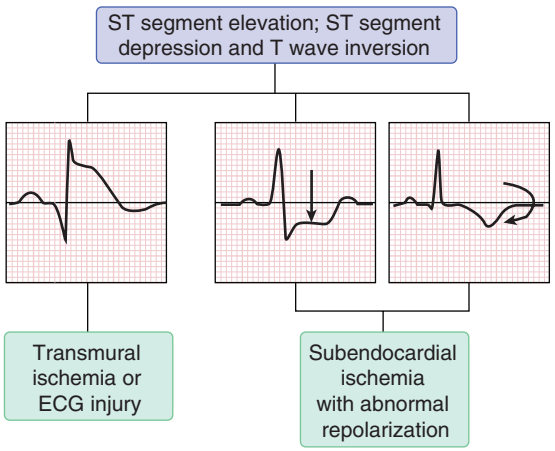

FIGURE 71-2. Ischemic ST segment shifts and repolarization changes on electrocardiogram (ECG).

In addition to focal ST-T wave abnormalities, the ECG may reveal various conduction disturbances, most frequently left bundle branch block and left anterior fascicular block (Chapter 54). The finding of abnormal Q waves is relatively specific for the presence of previous MI but may not help determine when such an event occurred. Arrhythmias, especially ventricular premature beats (Chapter 65), may be present on the ECG but have a low sensitivity and specificity for coronary artery disease.

During a spontaneous episode of angina pectoris or during exertion or stress, the ECG becomes abnormal in 50% or more of patients with normal resting ECGs. The most common abnormality observed is focal ST segment depression, usually in one or more ECG lead groups, which signifies the presence of subendocardial ischemia. On occasion, transient but diminutive ST segment elevation and normalization of previous resting ST-T wave depression or inversion (pseudonormalization) may develop during chronic angina and ischemia, although ST segment elevation is far more commonly observed in ACS patients with plaque rupture.

TABLE 71-4 BLOOD TESTS TO OBTAIN ROUTINELY (OR SELECTIVELY*) IN PATIENTS WITH CHRONIC STABLE ISCHEMIC HEART DISEASE

LIPID LEVELS
Low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol
Triglyceride level
*LDL electrophoresis (especially apolipoprotein B and small dense LDL)
*Lipoprotein (a)
*Lipoprotein-associated phospholipase A_2
METABOLIC EVALUATION
Fasting plasma glucose concentration
Serum creatinine level
Thyroxine level
*Hemoglobin A_{1c} in patients with known or suspected diabetes
MARKERS OF INFLAMMATION OR CARDIAC FUNCTION
*High-sensitivity C-reactive protein
*Brain natriuretic peptide
PROTHROMBOTIC ASSESSMENT
Plasma fibrinogen
Platelet count
*Factor V Leiden
*D-dimer
*Plasminogen activator inhibitor type 1
TO ASSESS OTHER POTENTIAL CARDIAC RISK FACTORS
*Serum homocysteine level

Laboratory Testing

In patients with new-onset or worsening symptoms, serial troponin measurements can distinguish MI and ACS from stable ischemic heart disease (Chapters 72 and 73). The plasma level of brain natriuretic peptide, which increases in response to spontaneous or provoked ischemia, does not reliably distinguish stable from unstable ischemic heart disease but is associated with the risk of future cardiovascular events in patients who are at risk for coronary disease. High-sensitivity C-reactive protein, an acute phase reactant of inflammation, has a strong and consistent relationship to the risk of future cardiovascular events, and an elevated level may warrant more aggressive diagnostic evaluation and therapy.

All patients with chronic angina should have biochemical evaluation of total cholesterol, low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, triglyceride, serum creatinine (estimated glomerular filtration), and fasting blood glucose levels (Table 71-4). Other biochemical markers that are not routinely recommended but are associated with higher risk of future cardiovascular events include lipoprotein (a), apolipoprotein B, small dense LDL cholesterol, and lipoprotein-associated phospholipase A_2. Homocysteine levels correlate with the risk for development of coronary heart disease, but randomized trials have failed to demonstrate a reduction of clinical events when elevated homocysteine levels are reduced; as a result, routine screening for an elevated homocysteine level is not recommended.

Noninvasive Testing

Noninvasive stress testing with a standard electrocardiographic treadmill or bicycle exercise, radionuclide imaging (Chapter 56), stress echocardiography (Chapter 55), or newer diagnostic modalities such as cardiac magnetic resonance (CMR; Chapter 56) and positron emission tomography (PET; Chapter 56) is a useful and clinically important approach to establishing the diagnosis and prognosis in patients with stable ischemic heart disease (Fig. 71-3). The predictive accuracy of these tests is defined not only by their sensitivity and specificity but also by the prevalence of disease (or pretest probability) in the population under study. Noninvasive testing should be performed only if the incremental information is likely to alter the planned management strategy. Thus, the value of noninvasive stress testing is greatest when the pretest likelihood is intermediate because the test result is likely to have the greatest effect on the post-test probability of coronary artery disease and, hence, on clinical decision making.

Each noninvasive test has a sensitivity and specificity (Table 71-5) that, when combined with a patient's pretest probability (see Table 71-1), can yield a post-test probability for coronary artery disease (Fig. 71-4). The choice among tests depends on the patient's characteristics (Table 71-6).

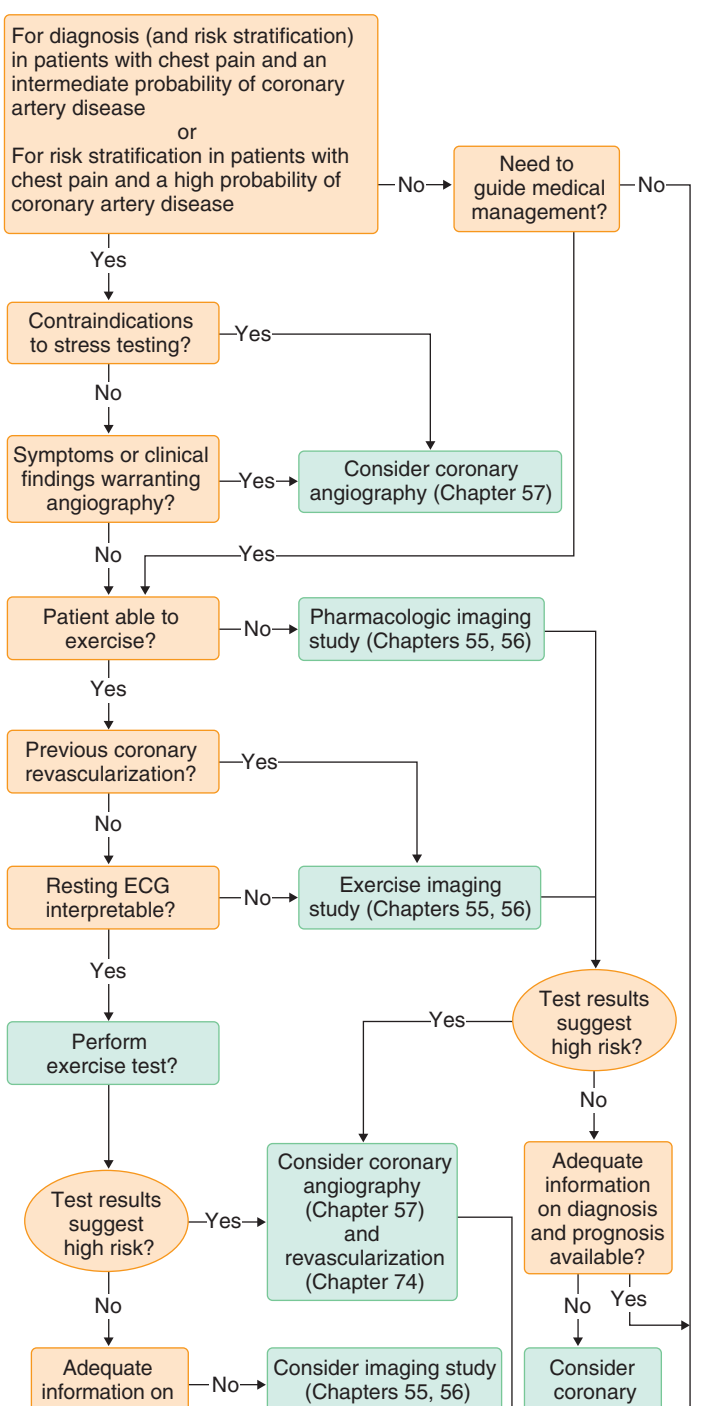

FIGURE 71-3 Approach to the use of stress testing and angiography for the evaluation of chronic stable angina. ECG = electrocardiogram. (Modified from American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Management of Patients with Chronic Stable Angina. ACC/AHA/ACP-ASIM Pocket Guidelines. Philadelphia: Elsevier Science; 2000.)

Exercise Electrocardiography

An exercise ECG is the preferred test in patients who have suspected angina pectoris and are considered to have a moderate probability of coronary artery disease if the resting ECG is normal (i.e., ST segments are not obscured by structural heart disease or medication), provided subjects are capable of achieving an adequate workload. Interpretation of the exercise ECG should include the exercise capacity achieved (duration and metabolic equivalents of the external workload; see Table 51-3), the magnitude and extent of ST

TABLE 71-5 APPROXIMATE SENSITIVITY AND SPECIFICITY OF COMMON TESTS TO DIAGNOSE CORONARY ARTERY DISEASE

	SENSITIVITY	SPECIFICITY
EXERCISE ELECTROCARDIOGRAPHY		
>1 mm ST depression	0.70	0.75
>2 mm ST depression	0.33	0.97
>3 mm ST depression	0.20	0.99
PERFUSION SCINTIGRAPHY		
Exercise SPECT	0.88	0.72
Pharmacologic SPECT	0.90	0.82
ECHOCARDIOGRAPHY		
Exercise	0.85	0.81
Pharmacologic stress	0.81	0.79
PET	0.95	0.95

PET = positron emission tomography; SPECT = single-photon emission computed tomography.
From Gibbons RJ, Abrams J, Chatterjee K, et al. ACC/AHA 2002 guideline update for the management of patients with chronic stable angina—summary article: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Patients with Chronic Stable Angina). *Circulation*. 2003;107:149-158.

TABLE 71-6 SUGGESTED NONINVASIVE TESTS IN DIFFERENT TYPES OF PATIENTS WITH STABLE ANGINA

Exertional angina, mixed angina, walk-through angina, postprandial angina with or without prior myocardial infarction	
Normal resting ECG	Treadmill exercise ECG test
Abnormal, uninterpretable resting ECG	Exercise myocardial perfusion scintigraphy (^{201}Tl, ^{99m}Tc-sestamibi) or exercise echocardiography
Unsuitable for exercise	Dipyridamole, adenosine, or regadenoson myocardial perfusion scintigraphy; dobutamine stress echocardiography
Atypical chest pain with normal or borderline abnormal resting ECG or with nondiagnostic stress ECG, particularly in women	Exercise myocardial perfusion scintigraphy, exercise echocardiography
Vasospastic angina	ECG during chest pain, ST segment ambulatory ECG, exercise test
Dilated ischemic cardiomyopathy with typical angina or for assessment of hibernating or stunned myocardium	Regional and global ejection fraction by radionuclide ventriculography or two-dimensional echocardiography, radionuclide myocardial perfusion scintigraphy; in selected patients, flow and metabolic studies with positron emission tomography
Syndrome X	Treadmill exercise stress ECG, coronary blood flow by positron emission tomography, Doppler probe
Known severe aortic stenosis or severe hypertrophic cardiomyopathy with stable angina	Exercise stress tests contraindicated; dipyridamole, adenosine, or regadenoson myocardial perfusion scintigraphy in selected patients; coronary angiography preferred
Mild aortic valvular disease or hypertrophic cardiomyopathy with typical exertional angina	"Prudent" treadmill myocardial perfusion scintigraphy; dipyridamole, adenosine, or regadenoson myocardial perfusion scintigraphy

ECG = 12-lead electrocardiogram.
Modified from Braunwald E, Goldman L, eds. Primary Care Cardiology. 2nd ed. Philadelphia: WB Saunders; 2003.

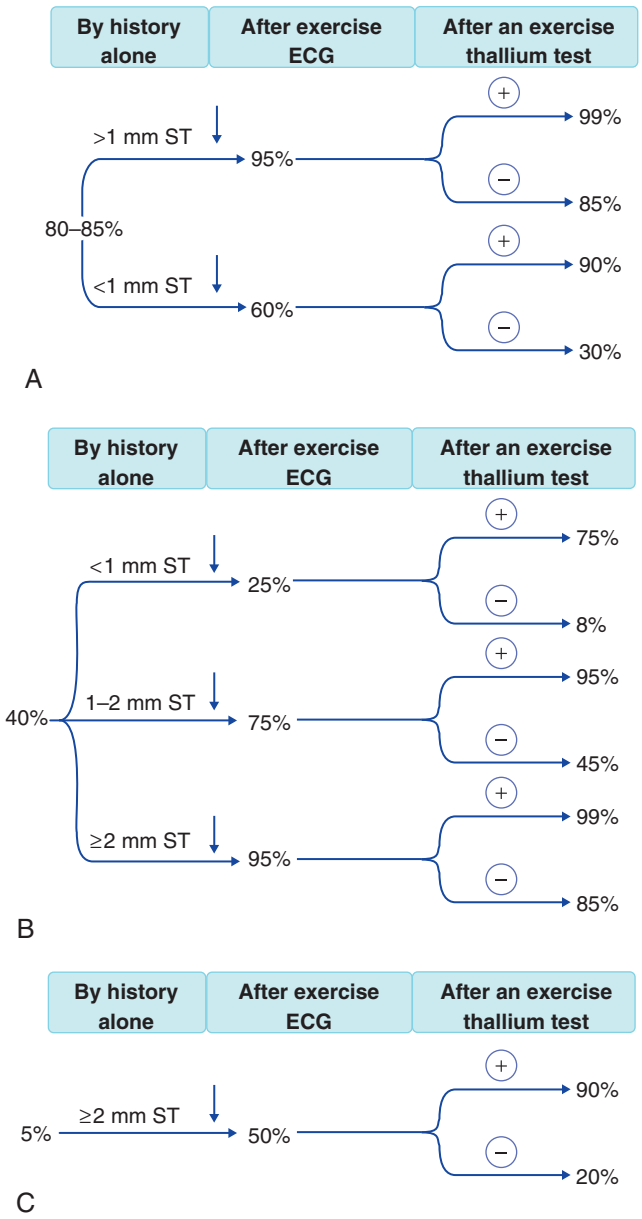

FIGURE 71-4. Approximate probabilities of coronary artery disease in different patient groups. A, Approximate probability of coronary artery disease before and after noninvasive testing in a patient with typical angina pectoris. These percentages demonstrate how the sequential use of an electrocardiogram (ECG) and an exercise thallium test may affect the probability of coronary artery disease in a patient with typical angina pectoris. **B,** Approximate probability of coronary artery disease before and after noninvasive testing in a patient with atypical angina symptoms. **C,** Approximate probability of coronary artery disease before and after noninvasive testing in an asymptomatic subject in the coronary artery disease age range. (Redrawn from Branch WB Jr, ed. Office Practice of Medicine. 3rd ed. Philadelphia: WB Saunders; 1994:45.)

segment deviation, the clinical and hemodynamic responses to exercise, and the rapidity with which the heart rate returns to normal after exercise.

The exercise test protocol is usually adjusted to a patient's tolerance, aiming for 6 to 12 minutes of exercise time (i.e., Bruce protocol stages II to IV) to achieve maximal oxygen consumption and to elicit objective evidence of inducible ischemia, if present. Exercise stress testing is generally safe, with death or MI occurring in less than one case per 2500 tests, when such provocative testing is avoided in patients with ACS, severe aortic stenosis, severe hypertension, or uncontrolled heart failure. Other contraindications are acute MI, symptomatic arrhythmias, acute pulmonary embolism, and suspected acute aortic dissection. Relative contraindications are hypertension above 200 mm Hg systolic or 110 mm Hg diastolic, hypertrophic cardiomyopathy, and high-degree atrioventricular block.

Concomitant antianginal therapy (notably the use of β-blockers) reduces the sensitivity of exercise testing as a screening tool. If the purpose of the exercise test is to diagnose ischemia, it should be performed,

whenever possible, before β-blockers are initiated or 2 to 3 days after their discontinuation.

Nuclear Cardiology Imaging

Stress myocardial perfusion imaging (Chapter 56) with single-photon emission computed tomography (SPECT) with simultaneous electrocardiographic testing (see Fig. 56-9) is particularly helpful in the diagnosis of coronary artery disease in patients with abnormal resting ECGs and among those in whom ST segment responses cannot be interpreted accurately, such as patients with repolarization abnormalities caused by LV hypertrophy, those with left bundle branch block, and those receiving digitalis. Its sensitivity and specificity are superior to exercise electrocardiography alone in detecting coronary artery disease (especially multivessel disease), in identifying regional perfusion defects that may localize to and correlate with diseased vessels, and in delineating the magnitude and extent of ischemic and infarcted myocardium. Treadmill testing is preferred for patients who are capable of performing such physical activity because of the additional diagnostic and prognostic information achieved with graded exercise. In the 40 to 50% of patients who are unable to exercise adequately, however, pharmacologic vasodilator stress with dipyridamole, adenosine, or regadenoson may be the preferred approach to noninvasive testing.

Stress Echocardiography (Chapter 55)

Stress two-dimensional echocardiography with exercise or pharmacologic stress can detect regional ischemia by identifying new wall motion abnormalities (see Fig. 55-7). Additional clinical information about associated structural heart disease, chamber dimensions, and valve function can be readily obtained. Exercise echocardiography can detect the presence of coronary artery disease with an accuracy similar to that achieved with stress myocardial perfusion imaging and is useful for localizing and quantifying ischemic myocardial segments. Pharmacologic stress is usually performed with dobutamine in patients who are unable to exercise and in those unable to achieve adequate heart rates with exercise.

Ambulatory Ischemic Monitoring

Patients with symptomatic myocardial ischemia have episodes of silent ischemia that occur with the activities of daily living and are detectable on ambulatory monitoring but go unrecognized clinically because of the absence of angina or anginal equivalents. Although 24-hour ambulatory electrocardiography may detect such "silent myocardial ischemia" and may provide a quantitative estimate of the frequency and duration of ischemic episodes, its sensitivity for detection of coronary artery disease is much less reliable than that of exercise electrocardiography.

Stress Cardiac Magnetic Resonance Imaging (Chapter 56)

Pharmacologic stress perfusion with CMR imaging is becoming increasingly available in many centers and may provide additional diagnostic capability in detecting the presence of structural heart disease, in addition to suspected coronary artery disease. CMR with gadolinium enhancement is the most accurate way to diagnose a scar from a prior MI (see Fig. 56-17).

Chest Radiography

Unless there is a history of prior MI, heart failure, or structural heart disease, the chest radiograph (Chapter 56) is usually normal in patients with chronic angina or stable ischemic heart disease. If an enlarged cardiac silhouette is present, it is generally indicative of a previous MI with LV dilation and cardiac remodeling. Other causes of cardiomegaly include long-standing hypertension, concomitant valvular heart disease, pericardial effusion, and nonischemic cardiomyopathy.

Cardiac Computed Tomographic Angiography

Cardiac computed tomographic angiography (CCTA; Chapter 56) is a highly sensitive method to detect coronary calcification (see Fig. 56-10), which is strongly associated with coronary atherosclerosis, and can also provide noninvasive angiography of the proximal coronary arteries. Although coronary calcification is a highly sensitive (approximately 90%) finding in patients with coronary artery disease, the specificity for identifying patients with obstructive coronary artery disease is much lower (approximately 50%). Because of the potential unnecessary testing from false-positive results, CCTA is currently not recommended as a routine screening approach for suspected obstructive coronary artery disease in individuals at low risk (<10% 10-year estimated risk of coronary events). By contrast, selective screening of intermediate-risk patients may be reasonable because a high calcium score may reclassify such individuals at higher risk and thereby lead to more intense risk factor modification.[5] CCTA can also be coupled with PET imaging in a hybrid PET and computed tomography scanner, which can provide a quantitative assessment of coronary anatomy along with regional myocardial blood flow and cardiac metabolism.

Diagnostic Coronary Angiography

Despite the continued evolution of noninvasive diagnostic testing, invasive coronary angiography (Chapter 57) remains the "gold standard" for anatomic definition of coronary artery disease. Among patients with the clinical diagnosis of stable ischemic heart disease referred for coronary angiography, a 70% or more luminal diameter narrowing is found in one (about 25%), two (about 25%), or all three (about 25%) epicardial coronary arteries in about 75% of cases; another 5 to 10% of patients have obstruction of the left main coronary artery; and the remaining 15 to 20% have no flow-limiting coronary obstructions. These data emphasize the persisting role of coronary angiography for diagnostic purposes (Table 71-7), but angiography is also helpful for

TABLE 71-7 CORONARY ANGIOGRAPHY FOR DIAGNOSIS AND RISK STRATIFICATION IN PATIENTS WITH CHRONIC ANGINA AND STABLE ISCHEMIC HEART DISEASE

FOR INITIAL DIAGNOSTIC INDICATION
Recommended on the Basis of Evidence or General Consensus
Patients with suspected angina and evidence of intermediate-high risk, moderate-severe ischemia on noninvasive testing, or a changing angina pattern who have survived sudden cardiac death or serious ventricular arrhythmia
Weight of Evidence or Opinion is in Favor
Uncertain diagnosis after noninvasive testing, and the benefit of a more certain diagnosis outweighs the risk and cost of coronary angiography
Inability to undergo noninvasive testing because of disability, illness, or morbid obesity
Occupational requirement for a definitive diagnosis
Suspected nonatherosclerotic cause of myocardial ischemia
Suspicion of a coronary spasm
High pretest probability of left main or three-vessel disease
Recurrent hospitalization for chest pain in the absence of definitive diagnosis
Overriding desire for a definitive diagnosis and a greater than low probability of CAD
Not Recommended
Significant comorbidity in patients in whom the risk of coronary arteriography outweighs the benefit of the procedure
Overriding personal desire for a definitive diagnosis and a low probability of CAD
FOR INITIAL RISK STRATIFICATION OR TREATMENT INDICATION
Recommended on the Basis of Evidence or General Consensus
With disabling (CCS class III and class IV) chronic stable angina despite medical therapy
With high-risk criteria on noninvasive testing regardless of anginal severity
Patients with angina who have survived sudden cardiac death or serious ventricular arrhythmia
Angina and symptoms and signs of congestive heart failure
Clinical characteristics that indicate a high likelihood of severe CAD
Weight of Evidence or Opinion is in Favor
Significant left ventricular dysfunction (EF < 45%), CCS class I or class II angina, and demonstrable ischemia but less than high-risk criteria on noninvasive testing
High-risk criteria suggesting ischemia on noninvasive testing
Inadequate prognostic information after noninvasive testing
Clinical characteristics that indicate a high likelihood of severe CAD
CCS class I or class II angina, preserved left ventricular function (EF > 45%), and less than high-risk criteria on noninvasive testing
CCS class III or class IV angina that improves to class I or class II with medical therapy
CCS class I or class II angina but intolerance (unacceptable side effects) to adequate medical therapy
Not Recommended
CCS class I or class II angina in patients who respond to medical therapy and who have no evidence of ischemia on noninvasive testing
Patients who prefer to avoid revascularization after adequate explanation

CAD = coronary artery disease; CCS = Canadian Cardiovascular Society; EF = ejection fraction.
Modified from Gibbons RJ, Abrams J, Chatterjee K, et al. ACC/AHA 2002 guideline update for the management of patients with chronic stable angina—summary article: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Patients with Chronic Stable Angina). *Circulation*. 2003;107:149-158.

risk stratification in patients with clear-cut angina and ischemic heart disease. In patients with less severe coronary stenoses (i.e., 50 to 70% on angiography), coronary intravascular ultrasonography (see Fig. 57-6) can substantially enhance the quantification of obstruction and vulnerability of the coronary atheroma to future instability. Alternatively, an invasive physiologic approach with use of a pressure wire positioned proximal and distal to a coronary stenosis can measure the severity of the stenosis and determine whether functionally significant flow reduction (i.e., a decreased fractional flow reserve [FFR] < 0.8) is present. This technique may be useful when there is a borderline (50 to 60%) visual coronary stenosis at angiography, particularly if such a stenosis subtends an ischemic myocardial segment observed on noninvasive testing. An FFR below 0.8 is generally considered to represent a sufficiently important reduction in coronary flow to justify proceeding to percutaneous coronary intervention (Chapter 74) to reduce symptoms and the subsequent need for urgent revascularization.[1,2] By contrast, an FFR of 0.8 or higher would indicate that myocardial revascularization of the stenotic coronary artery would be of little benefit clinically.

Differential Diagnosis of Angina

Many common noncardiac disorders may be manifested with clinical features that can be confused with angina pectoris (see Table 51-2). In some instances, symptoms may be indistinguishable from ischemic heart disease. For example, many patients with angina have coexisting esophageal disorders (Chapter 138), and both angina and esophageal discomfort may be relieved by nitroglycerin (Chapter 51). A distinguishing feature from angina is that esophageal discomfort is often relieved by antacids, proton pump inhibitors, or food.

Costochondritis can mimic angina but can typically be distinguished by the presence of well-localized pain on palpation. However, pressure, if it is applied too firmly to the anterior chest wall during examination of a patient with suspected angina pectoris, may elicit symptoms of discomfort even in normal subjects. Cervical radiculopathy may cause pain radiating to the shoulders, neck, or upper arms and can be confused with angina. However, this condition typically causes a constant ache that is often exacerbated by neck movement or rotation and may be accompanied by a focal sensory deficit or radiculopathy.

Pulmonary hypertension (Chapter 68) can cause exertional chest discomfort that may share many of the characteristics of angina pectoris. It is believed that right ventricular ischemia during physical exertion may cause this discomfort along with associated symptoms of exertional dyspnea, dizziness, and syncope. Findings on physical examination typically include a parasternal lift, a loud (and sometimes palpable) pulmonary component of the second heart sound, and findings of right ventricular hypertrophy on electrocardiography.

Chest pain may also be an important presenting clinical feature of pulmonary embolism (Chapter 98). Physical findings typically include tachycardia and tachypnea, an accentuated pulmonic component of the second heart sound, and occasionally a right-sided S_4 gallop. Pleuritic discomfort suggests pulmonary infarction, whereas a history of pain exacerbated by inspiration or deep breathing, along with a pleural friction rub, usually helps distinguish it from angina pectoris.

Acute pericarditis (Chapter 77) may be confused with the discomfort of angina pectoris, but pericarditis tends to cause chest pain that is generally sharp, is not relieved by rest or nitroglycerin, is exacerbated by movement or deep breathing, and is associated with a pericardial friction rub that may be evanescent. Aortic dissection (Chapter 78), which may be manifested with acute, severe chest pain, may be confused with an acute MI but generally not with angina.

Risk Stratification

Clinical and noninvasive criteria can be used in a complementary fashion to refine the estimate of risk for the individual patient with stable ischemic heart disease (Table 71-8). Clinical characteristics that include age, male sex, diabetes mellitus, previous MI, and symptoms typical of angina are predictive of the presence of coronary artery disease. Heart failure and LV dysfunction (generally defined by an ejection fraction < 50%), the severity and extent of angina, and associated symptoms such as dyspnea are also important predictors of outcome in patients with stable ischemic heart disease.

The simple classification of disease into single-, double-, or triple-vessel or left main coronary artery disease remains the most widely used approach (Table 71-9). Additional prognostic information is provided by the severity and extent of coronary luminal narrowing and its location. For example, high-grade lesions of the left main coronary artery or its equivalent, as defined by severe proximal left anterior descending and proximal left circumflex coronary artery disease, are particularly life-threatening. The SYNTAX trial has permitted the identification of subsets of coronary artery disease patients into three tertiles of low-, moderate-, and high-risk anatomic findings based on

TABLE 71-8 USING THE RESULTS OF NONINVASIVE RISK STRATIFICATION TO GUIDE CLINICAL DECISION MAKING

HIGH RISK (>3% ANNUAL MORTALITY RATE)
Severe resting left ventricular dysfunction (LVEF < 35%)
High-risk treadmill score (≤−11)*
Severe exercise left ventricular dysfunction (exercise LVEF < 35%)
Stress-induced large perfusion defect (particularly if anterior)
Stress-induced multiple perfusion defects of moderate size
Large, fixed perfusion defect with left ventricular dilation or increased lung uptake (^{201}Tl)
Stress-induced moderate perfusion defect with left ventricular dilation or increased lung uptake (^{201}Tl)
Echocardiographic wall motion abnormality (involving more than two segments) developing at low dose of dobutamine or at a low heart rate (<120 beats/min)
Stress echocardiographic evidence of extensive ischemia
INTERMEDIATE RISK (1-3% ANNUAL MORTALITY RATE)
Mild to moderate resting left ventricular dysfunction (LVEF = 35-49%)
Intermediate-risk treadmill score (−11 < score < 5)*
Stress-induced moderate perfusion defect without left ventricular dilation or increased lung intake (^{201}Tl)
Limited stress echocardiographic ischemia with a wall motion abnormality only at higher doses of dobutamine involving two segments or less
LOW RISK (<1% ANNUAL MORTALITY RATE)
Low-risk treadmill score (≥ 5)*
Normal or small myocardial perfusion defect at rest or with stress†
Normal stress echocardiographic wall motion or no change of limited resting wall motion abnormalities during stress†

*Score = (duration of exercise in minutes) − (5 × mm of ST segment depression) − (4 × angina score), where 0 = no angina, 1 = nonlimiting angina, and 2 = angina that causes discontinuation of the test.
†Although the published data are limited, patients with these findings will probably not be at low risk in the presence of either a high-risk treadmill score or severe resting left ventricular dysfunction (LVEF < 35%).
LVEF = left ventricular ejection fraction.
From Gibbons RJ, Abrams J, Chatterjee K, et al. ACC/AHA 2002 guideline update for the management of patients with chronic stable angina—summary article: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines (Committee on the Management of Patients with Chronic Stable Angina). *J Am Coll Cardiol.* 2003;41:159-168.

TABLE 71-9 CORONARY ARTERY DISEASE PROGNOSTIC INDEX

EXTENT OF CORONARY ARTERY DISEASE	5-YEAR MORTALITY RATE (%)*
1-vessel disease, 75%	7
>1-vessel disease, 50-74%	7
1-vessel disease, ≥95%	9
2-vessel disease	12
2-vessel disease, both ≥95%	14
1-vessel disease, ≥95% proximal LAD	17
2-vessel disease, ≥95% LAD	17
2-vessel disease, ≥95% proximal LAD	21
3-vessel disease	21
3-vessel disease, ≥95% in at least 1	27
3-vessel disease, 75% proximal LAD	33
3-vessel disease, ≥95% proximal LAD	41

*Assuming medical treatment only.
LAD = left anterior descending coronary artery.
From Califf RM, Armstrong PW, Carver JR, et al. Task Force 5: stratification of patients into high, medium and low risk subgroups for purposes of risk factor management. *J Am Coll Cardiol.* 1996;27:1007-1019.

multiple lesional characteristics (see Tables 71-8 and 71-9). However, the plaque causing the most severe chronic stenosis is not necessarily the one that will subsequently rupture to cause ACS or acute MI.

TREATMENT

Rx

Comprehensive management of angina and stable ischemic heart disease (Fig. 71-5) entails multiple therapeutic approaches to the identification and treatment of associated diseases that can precipitate or worsen angina and ischemia (Table 71-10): cardiac risk factor identification and intervention; application of pharmacologic and nonpharmacologic interventions for secondary prevention; pharmacologic and symptomatic management of angina and ischemia; and myocardial revascularization with percutaneous coronary intervention (PCI) or coronary artery bypass graft (CABG) surgery, when indicated (Table 71-11). A multidimensional management approach integrates all of these considerations, often simultaneously, in each patient. Among pharmacotherapies, three drug classes are classified as being "disease modifying" in that they have been demonstrated to reduce mortality and morbidity in patients with stable ischemic heart disease and preserved LV function: antiplatelet agents such as aspirin; inhibitors of the renin-angiotensin-aldosterone system, especially angiotensin-converting enzyme (ACE) inhibitors; and effective lipid-lowering agents, principally statins. Other therapies, such as nitrates, β-blockers, calcium antagonists, and ranolazine, can reduce myocardial ischemia,

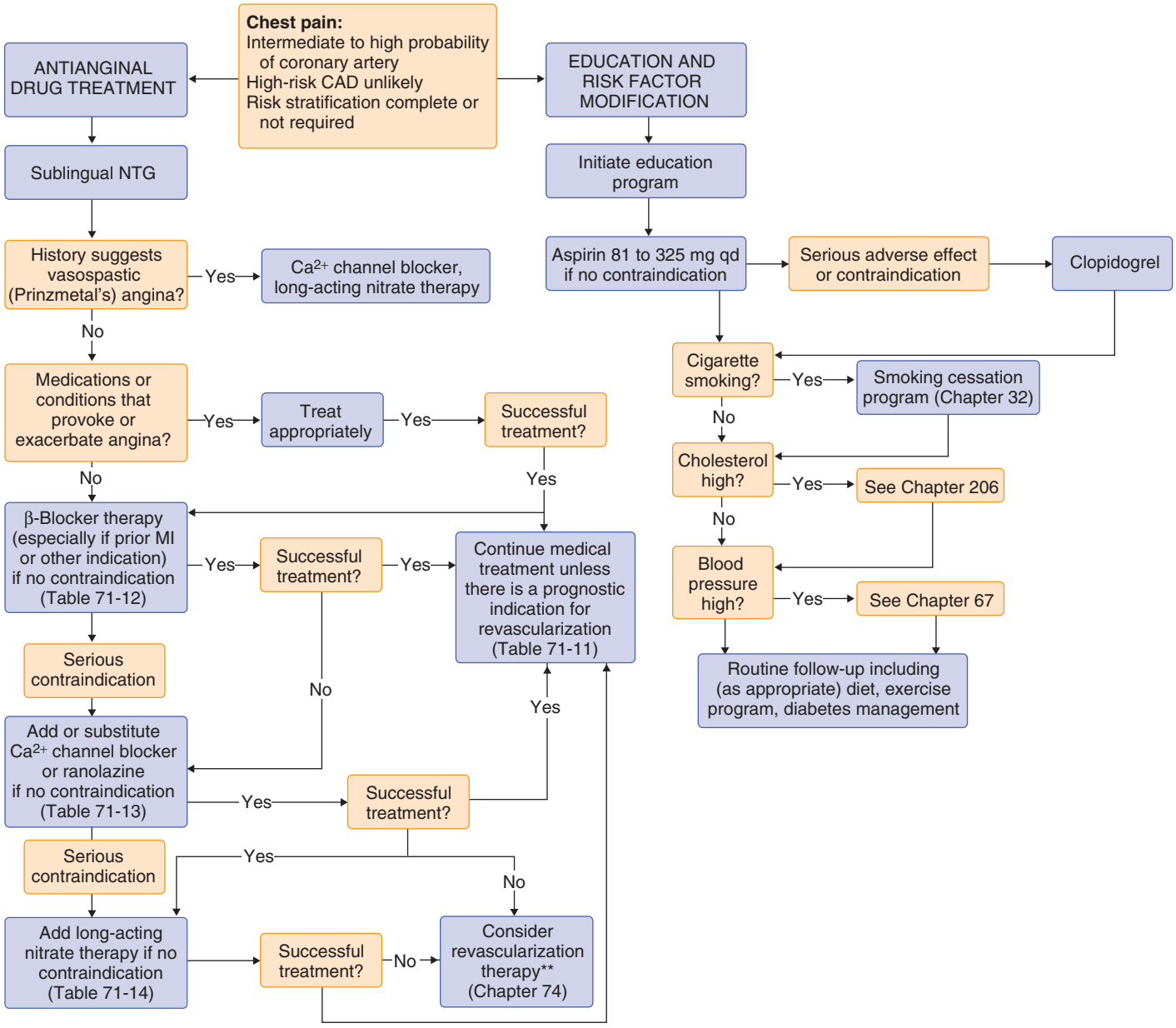

*Conditions that exacerbate or provoke angina:

Medications:	Other medical problems:	Other cardiac problems:
Vasodilators	Profound anemia	Tachyarrythmias
Excessive thyroid replacement	Uncontrolled hypertension	Bradyarrythmias
Vasoconstrictors	Hyperthyroidism	Valvular heart disease (espec. AS)
	Hypoxemia	Hypertrophic cardiomyopathy

**At any point in this process, based on coronary anatomy, severity of anginal symptoms, and patient preferences, it is reasonable to consider evaluation for coronary revascularization. Unless a patient is documented to have left main, three-vessel, or two-vessel CAD with significant stenosis of the proximal left anterior descending coronary artery, there is no demonstrated survival advantage associated with revascularization in low-risk patients with chronic stable angina; thus, medical therapy should be attempted in most patients before considering PTCA or CABG.

FIGURE 71-5. Algorithm for the treatment of stable angina. AS = aortic stenosis; CABG = coronary artery bypass graft; CAD = coronary artery disease; MI = myocardial infarction; NTG = nitroglycerin; PTCA = percutaneous transluminal coronary angioplasty. (Modified from American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Management of Patients with Chronic Stable Angina. ACC/AHA/ACP-ASIM Pocket Guidelines. Philadelphia: Elsevier Science; 2000.)

TABLE 71-10 TREATMENT OF PATIENTS WITH STABLE ANGINA

GENERAL MEASURES
Rule out and control aggravating conditions
Associated noncardiac diseases
Associated cardiac disease
Use of drugs aggravating angina
Smoking cessation
Dietary counseling for body weight and lipid control
Exercise prescription
Treat to targets
Hypertension
Blood lipids
Diabetes
PHARMACOLOGIC THERAPY: RECOMMENDATIONS FOR PHARMACOTHERAPY TO PREVENT MI AND DEATH AND TO REDUCE SYMPTOMS
Recommended on the Basis of Evidence or General Consensus
Aspirin in the absence of contraindications
β-Blockers as initial therapy in the absence of contraindications in patients with prior MI or without prior MI
Angiotensin-converting enzyme inhibitor in all patients with CAD who also have diabetes or left ventricular systolic dysfunction
Low-density lipoprotein–lowering therapy in patients with documented or suspected CAD and LDL cholesterol greater than 130 mg/dL, with a target LDL of less than 100 mg/dL
Sublingual nitroglycerin or nitroglycerin spray for the immediate relief of angina
Calcium-channel antagonists or long-acting nitrates as initial therapy for reduction of symptoms when β-blockers are contraindicated
Calcium-channel antagonists or long-acting nitrates in combination with β-blockers when initial treatment with β-blockers is not successful
Calcium-channel antagonists and long-acting nitrates as a substitute for β-blockers if initial treatment with β-blockers leads to unacceptable side effects
Weight of Evidence or Opinion is in Favor
Clopidogrel when aspirin is contraindicated
Long-acting non-dihydropyridine calcium-channel antagonists instead of β-blockers as initial therapy
In patients with documented or suspected CAD and LDL cholesterol level of 100 to 129 mg/dL, several therapeutic options are available (Level of Evidence: B)
Lifestyle and/or drug therapies to lower LDL to less than 100 mg/dL
Weight reduction and increased physical activity in persons with the metabolic syndrome
Institution of treatment of other lipid or nonlipid risk factors; consider use of nicotinic acid or fibric acid for elevated triglycerides or low HDL cholesterol
Angiotensin-converting enzyme inhibitor in patients with CAD or other vascular disease
Usefulness Unclear
Low-intensity anticoagulation with warfarin in addition to aspirin
Not Recommended
Dipyridamole
Chelation therapy

CAD = coronary artery disease; HDL = high-density lipoprotein; LDL = low-density lipoprotein; MI = myocardial infarction.

From Fihn SD, Blankenship JC, Alexander KP, et al. 2014 ACC/AHA/AATS/PCNA/SCAI/STS focused update of the guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines, and the American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *J Am Coll Cardiol*. 2014;64:1929-1949.

TABLE 71-11 CURRENT RECOMMENDATIONS FOR MYOCARDIAL REVASCULARIZATION IN PATIENTS WITH CHRONIC STABLE ANGINA

CABG SURGERY VERSUS MEDICAL THERAPY
Among patients with medically refractory angina pectoris, CABG surgery is indicated for symptom improvement.
Among patients with medically stable angina pectoris, CABG surgery is indicated to prolong life in left main coronary artery disease or three-vessel disease (regardless of left ventricular function) and, possibly, to help symptoms.
CABG surgery may be indicated for prolongation of life if the proximal left anterior descending coronary artery is involved (regardless of the number of diseased vessels).
CABG surgery may reduce the composite end point of death, myocardial infarction, or stroke in diabetic patients with extensive multivessel (two- to three-vessel) coronary artery disease compared with medical therapy.
PCI VERSUS MEDICAL THERAPY
For the initial management of patients with stable ischemic heart disease, PCI does not reduce the risk of death, myocardial infarction, or other major cardiovascular events when it is added to optimal medical therapy.
Among patients with medically refractory angina pectoris, PCI is indicated for symptom improvement.
PCI may be indicated in the presence of severe myocardial ischemia, regardless of symptoms. PCI does not appear to improve survival compared with medical treatment among patients with one- or two-vessel disease.
In the absence of symptoms or myocardial ischemia, PCI is not indicated (merely for the presence of an anatomic stenosis).
PCI VERSUS CABG SURGERY
For single-vessel disease, PCI and CABG surgery provide excellent symptom relief, but repeated revascularization procedures are required more frequently after PCI. Intracoronary stenting is preferred to regular PCI, but direct comparison with CABG surgery is limited.
For treated diabetic patients with two- or three-vessel disease, CABG surgery is the treatment of choice.
For nondiabetic patients, multivessel PCI and CABG surgery are acceptable alternatives. The choice of PCI or CABG surgery for initial treatment depends primarily on local expertise and the patient's and physician's preferences.
In general, PCI is preferred for patients at low risk and CABG surgery for patients at high risk.

CABG = coronary artery bypass graft; PCI = percutaneous coronary intervention.

ameliorate or prevent angina, and improve exercise performance, but they have not been shown to reduce mortality in patients with stable ischemic heart disease.

Disease-Modifying Therapies

Careful attention to lifestyle and the management of coronary risk factors is essential (Fig. 71-6). Such secondary prevention strategies can reduce the risk of progressive coronary disease, morbidity, and mortality.

Drugs That Alter Lipid Metabolism

Each 1% increase in the LDL cholesterol level results in a 2 to 3% increase in risk for coronary events (Chapter 52). Large, randomized clinical trials in patients with ischemic heart disease have shown a consistent and significant reduction in mortality and cardiac events with statin therapy (Chapter 206). Patients with stable angina should routinely be treated with statins[3,6-8] (see Table 206-5). Recent U.S. guidelines recommend high-dose statin therapy (e.g., atorvastatin up to 80 mg per day) without monitoring of LDL levels.[8] By comparison, European guidelines[7] generally recommend lowering of LDL levels at least to a target of less than 100 mg/dL and often to a target of less than 70 mg/dL.

Despite the epidemiologic link of LDL with coronary disease events, other lipid-lowering agents have not been proved to provide the clinical benefit expected from their lipid-lowering effects. For example, bile acid sequestrants are difficult for patients to take chronically and have not been proved to provide clinical benefits consistent with their lipid lowering. Nevertheless, they are often used in patients who are intolerant of statins or fail to reach their LDL goals with statins alone (Chapter 206), especially if they have high-risk features of prior ACS, MI, peripheral arterial disease, or cerebral arterial disease. Gemfibrozil, a fibrate at 1200 mg daily, can reduce fatal and nonfatal MI in men who have coronary heart disease and normal levels of LDL cholesterol but who also have low levels of HDL cholesterol and elevated triglycerides. However, fibrates have not reduced cardiac events in diabetic patients when added to statins and have not been shown to reduce overall mortality despite about a 10% reduction in major cardiovascular events. Similarly, ezetimibe lowers LDL cholesterol levels but has not been shown to reduce cardiac events or mortality.

Epidemiologically, each decline of 1 mg/dL in HDL cholesterol is associated with a 2 to 3% increase in the risk of MI and death from cardiac causes. However, drug therapy that elevates HDL levels with either niacin or cholesteryl ester transfer protein inhibitors has been universally disappointing.

Angiotensin-Converting Enzyme Inhibitors

ACE inhibitor administration (see Table 67-7) reduces cardiac events, cardiovascular mortality, and all-cause mortality in patients with risk factors for or with previously diagnosed coronary artery disease, including high-risk patients with vascular disease or diabetes and patients with stable coronary artery disease and no clinical evidence of heart failure. By contrast, ACE inhibitors do not appear to prevent future cardiac events in post-MI patients with

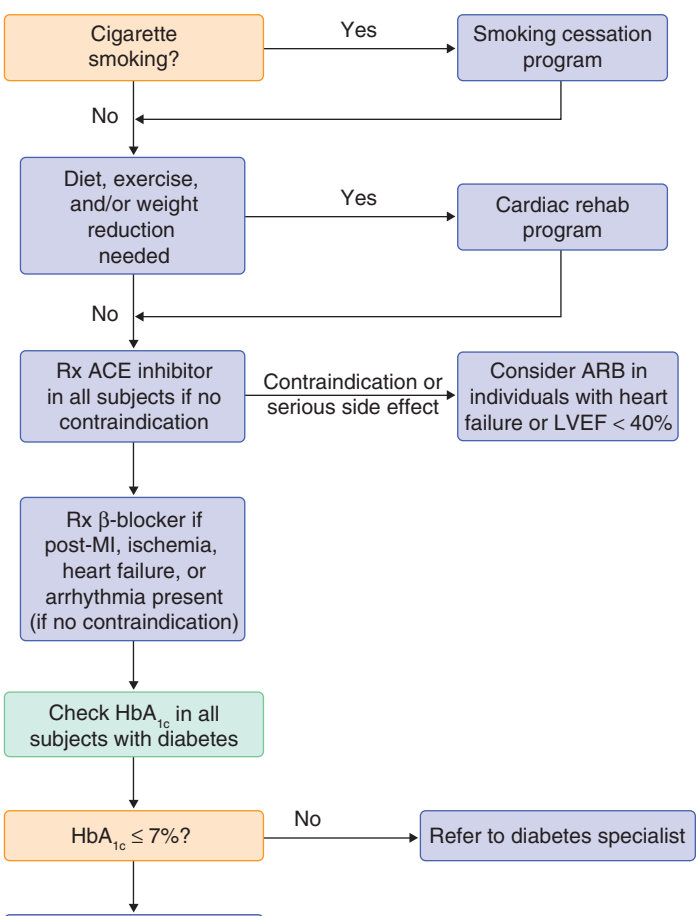

FIGURE 71-6. Approach to lifestyle interventions and pharmacotherapy. ACE = angiotensin-converting enzyme; ARB = angiotensin receptor blocker; HbA_{1c} = hemoglobin A_{1c}; LVEF = left ventricular ejection fraction; MI = myocardial infarction; Rx = prescribe.

preserved LV function. Whether angiotensin receptor blockers have similar benefits in patients with chronic angina and stable ischemic heart disease is not yet known.

Antiplatelet Agents and Anticoagulants

Aspirin reduces the risk of adverse cardiovascular events by 33% in patients with stable angina. The reduction in vascular events is comparable for doses of 75 to 150 mg daily and 160 to 325 mg daily, but daily doses of less than 75 mg have less benefit.[1] Therefore, in the absence of contraindications, aspirin, 75 to 325 mg once daily, should be administered routinely in all patients with angina and stable ischemic heart disease.

Clopidogrel (Chapter 38), which is the most widely used thienopyridine in the treatment of patients with coronary artery disease, is of proven benefit when it is combined with aspirin to reduce the composite end point of death, MI, or stroke in ACS patients and in patients who undergo PCI, especially with drug-eluting stents (Chapters 72 and 74). The standard regimen is an initial loading dose of 300 to 600 mg orally, followed by a maintenance dose of 75 mg daily. In patients with angina and stable ischemic heart disease, however, adding clopidogrel to low-dose (75 to 162 mg/day) aspirin does not reduce the primary composite end point of MI, stroke, or death from cardiovascular causes,[1] so clopidogrel should be reserved for patients who cannot tolerate aspirin or who have had an acute coronary event (Chapter 72) or a stent implantation (Chapter 74). Both prasugrel and ticagrelor are approved in patients with ACS and in those who undergo PCI with stent placement, but neither agent has yet been studied in the medical management of patients with chronic angina and stable coronary artery disease.

Warfarin is generally as effective as aspirin for preventing coronary events in patients with angina and is preferred to aspirin for patients with concomitant atrial fibrillation (Chapter 64), but it is associated with a higher risk of bleeding. Combination therapy with warfarin plus aspirin is superior to aspirin alone if the international normalized ratio is maintained above 2.0, but the benefit of the combination must be weighed against a 1 per 100 patient-years risk of bleeding.[1] Factor Xa inhibitors (dabigatran, rivaroxaban, and apixaban) approved for the treatment of venous thromboembolic disease and for the prevention of systemic embolism (Chapter 38) in patients with nonvalvular atrial fibrillation have not yet been tested in patients with stable ischemic heart disease.

Therapeutic Agents to Reduce Angina and Ischemia

The goal of antianginal therapy is to reduce symptoms of cardiac ischemia and to improve quality of life. β-Blockers, which prevent the binding of catecholamines to the β-adrenergic receptor, lower heart rate and myocardial contractility, thereby reducing myocardial workload, myocardial oxygen demand, and ischemia and anginal symptoms. β-Blockers raise the ischemic threshold and delay or prevent the onset of angina with exercise. β-Blockers also reduce the rate of secondary cardiac events and sudden cardiac death in post-MI patients, but there have been no placebo-controlled outcome trials in angina patients. All β-blockers appear to be equally effective in patients with chronic stable angina (Table 71-12). The β-blocker dose should be titrated to a target resting heart rate of 50 to 60 beats per minute as tolerated by the patient.

Calcium-channel blockers (Table 71-13) reduce afterload by their peripheral vasodilatory effects and thus lower myocardial workload and myocardial oxygen demand. Calcium-channel blockers also reduce coronary vascular resistance and inhibit coronary vasospasm by preventing coronary arterial smooth muscle contraction. This favorable reduction in myocardial oxygen demand, coupled with an increase in myocardial oxygen supply, results in a reduction in angina and ischemia. Non-dihydropyridine calcium-channel blockers, such as verapamil and diltiazem, also reduce heart rate. Conversely, dihydropyridine calcium-channel antagonists, such as amlodipine, have greater effect on vascular smooth muscle, are better peripheral and coronary vasodilators, and hence may have advantages for use in the hypertensive patient with angina. In randomized clinical trials, calcium-channel blockers and β-blockers are generally equally effective in relieving angina, improving time to onset of angina, and improving time to ischemic ST depression during exercise. Because calcium-channel blockers have not been shown to reduce death or MI in patients with stable or previously unstable ischemic heart disease, these agents are usually used in patients who cannot tolerate β-blockers or who require additional pharmacotherapy to control their symptoms. When calcium-channel blockers are used with β-blockers, care must be taken not to cause symptomatic bradycardia with verapamil and diltiazem. When calcium-channel blockers are used alone, diltiazem is often preferred because the dihydropyridine calcium-channel blockers can increase the heart rate.

Nitrates (Table 71-14) continue to be widely prescribed for antianginal treatment and are effective when they are administered sublingually, orally, or topically. They act as vasodilators by entering vascular smooth muscle, where they are metabolized to nitric oxide, which relaxes vascular smooth muscle, including in coronary arteries. These effects reduce angina by improving coronary blood flow. Nitrates also lower preload because of their venodilatory effects, with a resulting reduction in LV end-diastolic pressure and wall tension, which in turn lowers subendocardial oxygen demand. When nitrates are used in patients with stable angina, they improve exercise tolerance, time to onset of angina, and ST segment depression during treadmill exercise testing. Long-acting nitrates, which are frequently combined with β-blockers and calcium-channel blockers, have additive antianginal and anti-ischemic effects in patients with stable ischemic heart disease. Sublingual nitroglycerin or oral spray can terminate an angina attack and can be used as prophylaxis to prevent exertional angina. Long-acting nitrates administered orally or transdermally are used to prevent angina and to improve exercise tolerance. For avoidance of nitrate tolerance or tachyphylaxis, an 8- to 12-hour nitrate-free interval daily is recommended. Nitroglycerin and nitrates can cause vasodilation-induced headache, a decrease in blood pressure, and, more rarely, severe hypotension with bradycardia due to activation of the vagal Bezold-Jarisch reflex. Because the vasodilation by nitroglycerin is markedly exaggerated and prolonged in the presence of the phosphodiesterase inhibitors sildenafil (Viagra), vardenafil (Levitra), and tadalafil (Cialis), these agents and nitrates should not be used concurrently.

Ranolazine (initiated at a dose of 500 mg twice daily and titrated up to a maximal dose of 1000 mg twice daily) acts by reducing intracellular calcium overload in ischemic myocytes by inhibiting late inward sodium current entry. The net effect of reduced late inward sodium current is a reduction in LV wall tension and myocardial oxygen demand, thereby reducing angina and ischemia. Ranolazine increases exercise tolerance in patients with stable angina, reduces episodes of recurrent ischemia, and provides additional antianginal benefit in patients who are already receiving intensive antianginal therapy with β-blockers and calcium-channel blockers.[1] Whether ranolazine can reduce death, MI, or recurrent ischemia compared with placebo in patients with chronic angina is unproven.[2,6,9]

Nonpharmacologic Treatment

Enhanced external counterpulsation (EECP) is an alternative treatment for patients with refractory angina. EECP is generally administered as 35 sequential treatments (1 hour daily; 5 days/week) during 7 weeks. EECP does not reduce ischemia on myocardial perfusion imaging, and the mechanisms

TABLE 71-12 CLINICAL USE OF B-BLOCKERS

COMPOUND BY RECEPTOR ACTIVITY	INTRINSIC SYMPATHOMIMETIC ACTIVITY*	MEMBRANE STABILITY EFFECT	HALF-LIFE (hr)	EXCRETION	USE
β_1 AND β_2					
Propranolol	–	++	1-6	Hepatic	20-80 mg bid-tid
Propranolol long-acting	–	++	8-11	Hepatic	80-360 mg/day
Nadolol	–	–	40-80	Renal	40-80 mg/day
Pindolol	+	+	3-4	Renal	2.5-7.5 mg tid
Sotalol	–	–	7-18	Renal	40-160 mg bid
Timolol	–	–	4-5	Hepatic-renal	10-15 mg bid
β_1 SELECTIVE					
Acebutolol	+	+	3-4	Hepatic	200-600 mg bid
Atenolol	–	–	6-9	Renal	50-200 mg/day
Bisoprolol	–	–	9-12	50% renal	5-20 mg/day
Metoprolol	–	–	3-7	Hepatic	50-200 mg bid
Metoprolol long-acting	–	–	14-25	Hepatic	100-400 mg
Esmolol	–	–	4.5 min	Esterases in red cells	Bolus 500 µg/kg 50-300 µg/kg/min IV
β_1, β_2, α_2					
Labetalol	+	–	6	Hepatic	200-600 mg bid
Carvedilol	–	+	6-10	Hepatic	2.5-25 mg bid

*Presence commonly associated with maintenance of or increase in heart rate; absence associated with decrease in heart rate.
From Théroux P. Angina pectoris. In: Goldman L, Ausiello DA, eds. Cecil Textbook of Medicine. 23rd ed. Philadelphia: Saunders Elsevier; 2008.

TABLE 71-13 PROPERTIES OF CALCIUM-CHANNEL BLOCKING DRUGS IN CLINICAL USE

DRUGS	USUAL DOSE	ELIMINATION HALF-LIFE (hr)	*HEMODYNAMIC EFFECT* HR	*HEMODYNAMIC EFFECT* PVR	SIDE EFFECTS
DIHYDROPYRIDINES					
Nifedipine PA*	10-40 mg bid	10	↑↑	↓↓↓	Hypotension, dizziness, flushing, edema, constipation
Nifedipine XL*	30-120 mg/day	24	↑	↓↓	
Amlodipine	2.5-10 mg/day	30-50	=	↓↓↓	Headache, edema
Felodipine	2.5-10 mg/day	11-16	↑	↓↓↓	Headache, dizziness
Isradipine	2.5-10 mg bid	8	=	↓↓↓	Headache, fatigue
Nicardipine	20-40 mg tid	2-4	↑	↓↓↓	
Nicardipine SR*	30-60 mg bid	8-10	↑	↓↓	Headache, dizziness, flushing, edema
Nisoldipine	10-40 mg/day	7-12	=	↓↓↓	As for nifedipine
Nitrendipine	20 mg/day or bid	5-12	↑	↓↓↓	As for nifedipine
OTHERS					
Bepridil	200-400 mg/day	24-40	↓	↓	Arrhythmias, dizziness, nausea
Diltiazem	30-90 mg tid	4-6	↓	↓	Hypotension, dizziness, bradycardia, edema
Diltiazem CD*	120-540 mg/day	—	↓	↓	
Verapamil	80-160 mg tid	3-8	↓	↓↓	
Verapamil SR*	120-480 mg/day	—	↓	↓↓	Hypotension, heart failure, edema, bradycardia

*PA, XL, SR, CD: long acting.
HR = heart rate; PVR = peripheral vascular resistance.
From Théroux P. Angina pectoris. In: Goldman L, Ausiello DA, eds. Cecil Textbook of Medicine. 23rd ed. Philadelphia: Saunders Elsevier; 2008.

underlying its effects are poorly understood. Possible mechanisms include durable hemodynamic changes that reduce myocardial oxygen demand, improvement in myocardial perfusion by diastolic augmentation of retrograde coronary flow, and improved endothelial function. Although EECP increased the time to ST segment depression during exercise testing, reduced angina, and improved health-related quality of life for at least 1 year in one randomized, double-blind study of patients with chronic stable angina, its role in the treatment of angina remains unclear.

Myocardial Revascularization

Coronary revascularization with either PCI or CABG (Chapter 74) prolongs life, reduces major cardiovascular events, and improves health status, quality

TABLE 71-14 CLINICAL USE OF NITROGLYCERIN AND NITRATES

	DOSE	DURATION OF ACTION	INDICATION
NITROGLYCERIN			
Sublingual or buccal spray	0.15-1.5 mg	Relief of angina	Before or at onset of pain
Ointment	7.5-40 mg	8-12 hr	Prophylaxis of angina
Transdermal	0.2-0.8 mg/hr	8-16 hr	Prophylaxis of angina
Intravenous	5-400 µg/hr	Ongoing; increasing doses as needed	Recurrent chest pain, systemic hypertension, left-sided heart failure
ISOSORBIDE DINITRATE			
Oral	5-40 mg tid	6-8 hr	Prophylaxis of angina
ISOSORBIDE-5-MONONITRATE			
Oral	20 mg bid	8-12 hr	Prophylaxis of angina
Oral, slow release	30-240 mg/day	12-20 hr	Prophylaxis of angina

From Théroux P. Angina pectoris. In: Goldman L, Ausiello DA, eds. Cecil Textbook of Medicine. 23rd ed. Philadelphia: Saunders Elsevier; 2008.

of life, and functional capacity in selected patients with chronic, stable ischemic heart disease who meet certain anatomic criteria: the presence of significant left main coronary artery disease, three-vessel coronary artery disease, or multivessel coronary artery disease with an LV ejection fraction below 50%. For other patients, however, the data are more mixed.

Comparisons of PCI with Optimal Medical Therapy

Numerous randomized clinical trials have compared PCI with medical therapy in patients who have stable coronary heart disease and do have anatomic criteria as noted before. In these patients, PCI improves angina symptoms but does not reduce the risk of death, MI, or other major cardiovascular events when it is added to optimal medical therapy as an initial management strategy in patients with stable ischemic heart disease. As a result, a trial of optimal medical therapy to control symptoms is justifiable and more cost-effective[10] for patients whose stable coronary disease is not associated with anatomic features for which revascularization has been shown to prolong life. For patients with stable angina and coronary lesions with reduced FFR below 0.8 in one or more visually stenotic arteries (≥50% stenosis), initial PCI significantly reduces the need for hospitalization for urgent revascularization but has not been shown to reduce the composite end point of MI or death. Because PCI as an initial management strategy does not reduce long-term death, MI, or other major cardiovascular events when it is added to optimal medical therapy in patients with stable coronary artery disease but without specific anatomic criteria, preventive pharmacotherapy and lifestyle modification for secondary prevention of major cardiovascular events must be paramount in such patients.

Comparisons of CABG with Medical Therapy

Randomized trials comparing CABG with medical therapy indicate that a greater severity of ischemia, a greater extent of disease, and the presence of LV dysfunction favor a greater magnitude of survival benefit of CABG over medical therapy. CABG prolongs survival in patients with significant left main coronary artery disease irrespective of symptoms, in patients with multivessel coronary artery disease and impaired LV function (ejection fraction < 50%), and in patients with three-vessel coronary artery disease that includes the proximal left anterior descending coronary artery. Patients with extensive multivessel coronary artery disease appear to benefit more from CABG surgery, particularly if they also have diabetes, whereas PCI is most appropriate for patients with one- or two-vessel coronary artery disease. Of note is that these randomized trials found little difference in mortality between CABG surgery and medical therapy at 1 year, but the benefits of CABG surgery over medical therapy steadily emerged during the next 3 to 5 years. Few patients in these early trials received arterial grafts (which would likely improve the long-term results of CABG surgery), aspirin to prevent graft occlusion, lipid-lowering drugs to mitigate late graft disease progression, or inhibitors of the renin-angiotensin-aldosterone system. With improvements in operative techniques, more common use of antiplatelet agents, more widespread use of disease-modifying therapies, and more aggressive risk factor management during the past decade, the benefits of modern-day CABG surgery compared with contemporary medical therapy have likely changed.

Comparisons of PCI with CABG Surgery for Multivessel Coronary Artery Disease

In randomized trials that have compared PCI with CABG in patients with multivessel coronary artery disease, most excluded patients with significant left main coronary artery disease and were conducted in an era before the advent of stents and other advances in PCI technology, including newer adjunctive medical therapies that are increasingly in widespread clinical use. In patients with severe three-vessel disease or left main disease randomized to either CABG surgery or PCI with a paclitaxel drug-eluting stent, CABG significantly reduced the end point of cardiac death, recurrent MI, and repeated revascularization in the CABG-treated patients with multivessel disease, especially in patients with diabetes. In a large randomized trial of diabetic patients with multivessel but stable ischemic heart disease, CABG significantly reduced both death and MI compared with PCI.

In summary, among patients who remain symptomatic despite intensive treatment or who have substantial ischemia or extensive coronary artery disease, revascularization with either PCI or CABG is appropriate, depending on the anatomic complexity of disease (see Table 71-11; Chapter 73). CABG surgery is clearly superior to PCI in symptomatic patients with three-vessel or left main coronary artery disease and in diabetic patients with stable ischemic heart disease. Although PCI appears to provide equivalent survival outcomes in lower risk patients and in patients with one- or two-vessel disease, repeated procedures are more often needed.

OTHER ANGINAL SYNDROMES

Variant Angina or Prinzmetal's Angina

The diagnosis of variant or Prinzmetal's angina is based on the documentation of transient ST segment elevation during an episode of chest pain in the absence of a severe, fixed coronary stenosis. Prinzmetal's variant angina typically is caused by an occlusive spasm superimposed on a coronary artery stenosis that otherwise does not limit blood flow significantly. In some patients, however, no underlying stenoses are seen. Associated Raynaud phenomenon and migraine headache have been described in some patients, suggesting that the syndrome may be part of a more generalized vasospastic disorder.

The chest discomfort occurs predominantly at rest, although approximately one third of patients may also experience pain during exercise. There is a predilection for the pain to wake the patient in the early morning hours when sympathetic activity is increasing. The syndrome is often cyclical; periods of exacerbation with repetitive episodes of chest pain may persist only for seconds or be more prolonged and severe, alternating with periods with few or no symptoms. The symptoms are typically relieved by nitroglycerin. The ST segment elevation accompanying the pain signifies transmural ischemia due to total abrupt occlusion of a nonsignificant stenosis in the absence of adequate collateral circulation. The subsequent rapid reperfusion may explain the high prevalence of severe life-threatening arrhythmias.

Coronary angiography, with a provocative test for spasm such as injection of acetylcholine into the affected coronary artery, usually precipitates the syndrome. Such testing is useful to establish the diagnosis and to assess the response to therapy, especially in patients with normal or nearly normal coronary angiograms, in whom the diagnosis is otherwise unclear. Dihydropyridine calcium-channel blockers (e.g., amlodipine, 5 to 10 mg orally per day) are preferred in patients with Prinzmetal's angina.

Microvascular Angina with Normal Coronary Angiography

Angina can occur despite normal coronary arteries, even after an acetylcholine challenge. More detailed testing may reveal increased coronary resistance and an inability to increase coronary resistance and to increase coronary flow in response to stimuli such as exercise, adenosine, dipyridamole, and atrial pacing.

Symptoms occur most frequently at rest and often in relation to emotional stress. Periods of exacerbation commonly alternate with symptom-free

periods. The syndrome is more frequent in women, and some patients have an altered perception of pain or hypersensitivity to certain stimuli.

The diagnosis requires objective documentation of ischemia based on ST-T segment changes, a metabolic abnormality, a transient regional perfusion defect or new wall motion abnormality on echocardiography, or endothelial dysfunction that limits blood flow reserve. Abnormal endothelium-dependent vasoreactivity can be associated with regional myocardial perfusion defects on SPECT and PET imaging.

β-Blockers may be useful, particularly when a relative tachycardia, hypertension, or decreased heart rate variability on Holter monitoring is present. Nitroglycerin can relieve symptoms in approximately 50% of patients, and long-acting nitrates or calcium antagonists are sometimes helpful.

The prognosis is generally favorable and not different from that of a general age-matched population in the absence of coronary artery disease. However, some studies have indicated that an ischemic response to exercise is associated with increased mortality.

Silent Myocardial Ischemia

Up to 20% of patients with ischemic heart disease may not present with angina. Such patients are often described as having silent myocardial ischemia. Some patients are totally asymptomatic despite obstructive coronary artery disease, which may be severe. Others have silent ischemia after a prior documented MI. The third and most common form occurs in patients who may also exhibit the usual forms of chronic stable angina, unstable angina, and Prinzmetal's angina. When monitored, these patients, who typically have silent ischemia episodes in addition to symptomatic ischemia, are sometimes referred to as having *mixed angina*. This mixed angina is estimated to be present in approximately one third of all treated patients with angina, although an even higher prevalence has been reported in diabetic patients. In these patients, about 85% of ambulant ischemic episodes occur without chest pain, and 66% of angina episodes are not accompanied by ST segment depression, suggesting that overt angina pectoris is merely the "tip of the ischemic iceberg." Pharmacologic agents that reduce or abolish episodes of symptomatic ischemia also reduce or abolish episodes of silent ischemia.

PROGNOSIS WITH OPTIMAL MANAGEMENT

Modern treatments have improved the prognosis of patients with stable ischemic heart disease to an annual mortality rate of 1 to 3% and a 1 to 2% rate of major ischemic events. The 1-year rate of cardiovascular death is now 1.9% (95% confidence interval [CI], 1.7 to 2.1), with a 2.9% (95% CI, 2.6 to 3.2) rate of all-cause mortality and a 4.5% (95% CI, 4.2 to 4.8) rate of the combined outcome of cardiovascular death, MI, or stroke. Even in patients who have refractory angina and are not candidates for revascularization, modern medical treatment is associated with a 1-year mortality of only 4% and a 9-year mortality of 30%.[11]

Recurrent angina is a common subsequent complaint in many patients with stable ischemic heart disease, even those who were initially treated successfully with PCI. About 30% of patients continue to experience angina one or more times per week, with associated greater physical limitation and worse quality of life, and almost 80% of patients who underwent initially successful PCI for chronic angina still take one or more antianginal agents at 1 year.

Selection of optimal treatment requires a full understanding of the potential risks and benefits of each treatment approach. In patients with stable symptoms who have not had an adequate trial of medical therapy (e.g., 8 to 12 weeks of multifaceted medical therapy and lifestyle intervention), such an initial approach of aggressive medical therapy is recommended. For patients whose angina or quality of life is not adequately controlled with optimal medical therapy, revascularization with either PCI or CABG surgery should be considered. Recent clinical trials data now show convincing benefit of CABG surgery over PCI for patients with left main or three-vessel coronary artery disease and for diabetic patients with multivessel disease. Although the results of randomized trials must be individualized for specific patients, a multidisciplinary approach to clinical decision making can ensure that all therapeutic options are fully and transparently discussed so that patients are offered the most appropriate evidence-based treatment recommendations that are tailored to the level of ischemic risk and coronary anatomic findings.

Grade A References

A1. De Bruyne B, Pijls NH, Kalesan B, et al. Fractional flow reserve–guided PCI versus medical therapy in stable coronary disease. *N Engl J Med*. 2012;367:991-1001.

A2. De Bruyne B, Fearon WF, Pijls NH, et al. Fractional flow reserve-guided PCI for stable coronary artery disease. *N Engl J Med*. 2014;371:1208-1217.

A3. Cannon CP, Braunwald E, McCabe CH, et al. Intensive versus moderate lipid lowering with statins after acute coronary syndromes. *N Engl J Med*. 2004;350:1495-1504.

A4. Ginsberg HN, Elam MB, Lovato LC, et al. Effects of combination lipid therapy in type 2 diabetes mellitus. *N Engl J Med*. 2010;362:1563-1574.

A5. Jun M, Foote C, Lv J, et al. Effects of fibrates on cardiovascular outcomes: a systematic review and meta-analysis. *Lancet*. 2010;375:1875-1884.

A6. Boden WE, Probstfield JL, Anderson T, et al. Niacin in patients with low HDL cholesterol levels receiving intensive statin therapy. *N Engl J Med*. 2011;365:2255-2267.

A7. Landray MJ, Haynes R, Hopewell JC, et al. Effects of extended-release niacin with laropiprant in high-risk patients. *N Engl J Med*. 2014;371:203-212.

A8. Dagenais GR, Pogue J, Fox K, et al. Angiotensin-converting enzyme inhibitors in stable vascular disease without left ventricular systolic dysfunction or heart failure: a combined analysis of three trials. *Lancet*. 2006;368:581-588.

A9. Lièvre M, Cucherat M. Aspirin in the secondary prevention of cardiovascular disease: an update of the APTC meta-analysis. *Fundam Clin Pharmacol*. 2010;24:385-391.

A10. Bhatt DC, Fox KA, Hacke W, et al. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. *N Engl J Med*. 2006;354:1706-1717.

A11. Anand SS, Yusuf S. Oral anticoagulants in patients with coronary artery disease. *J Am Coll Cardiol*. 2003;41:62S-69S.

A12. Kosiborod M, Arnold SV, Spertus JA, et al. Evaluation of ranolazine in patients with type 2 diabetes mellitus and chronic stable angina: results from the TERISA randomized clinical trial (Type 2 Diabetes Evaluation of Ranolazine in Subjects With Chronic Stable Angina). *J Am Coll Cardiol*. 2013;61:2038-2045.

A13. McKenna C, McDaid C, Suekarran S, et al. Enhanced external counterpulsation for the treatment of stable angina and heart failure: a systematic review and economic analysis. *Health Technol Assess*. 2009;13:1-90.

A14. Stergiopoulos K, Boden WE, Hartigan P, et al. Percutaneous coronary intervention outcomes in patients with stable obstructive coronary artery disease and myocardial ischemia: a collaborative meta-analysis of contemporary randomized clinical trials. *JAMA Intern Med*. 2014;174:232-240.

A15. Boden WE, O'Rourke RA, Teo KK, et al. Optimal medical therapy with or without PCI for stable coronary disease. *N Engl J Med*. 2007;356:1503-1516.

A16. BARI Investigators. The final 10-year follow-up results from the BARI randomized trial. *J Am Coll Cardiol*. 2007;49:1600-1606.

A17. Frye R, August P, Brooks M, et al. BARI 2D: a randomized clinical trial of treatment strategies for type 2 diabetes and coronary artery disease. *N Engl J Med*. 2009;360:2503-2515.

A18. Mohr FW, Morice MC, Kappetein AP, et al. Coronary artery bypass graft surgery versus percutaneous coronary intervention in patients with three-vessel disease and left main coronary disease: 5-year follow-up of the randomised, clinical SYNTAX trial. *Lancet*. 2013;381:629-638.

A19. Farkouh ME, Domanski M, Sleeper LA, et al. Strategies for multivessel revascularization in patients with diabetes. *N Engl J Med*. 2012;367:2375-2384.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ward BW, Schiller JS, Goodman RA. Multiple chronic conditions among US adults: a 2012 update. *Prev Chronic Dis*. 2014;11:E62.
2. Ambulatory health care data. Centers for Disease Control and Prevention, 2013. http://www.cdc.gov/nchs/index.htm; Accessed December 3, 2014.
3. Fihn SD, Blankenship JC, Alexander KP, et al. 2014 ACC/AHA/AATS/PCNA/SCAI/STS focused update of the guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines, and the American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *J Am Coll Cardiol*. 2014;64:1929-1949.
4. Qaseem A, Fihn SD, Williams S, et al. Diagnosis of stable ischemic heart disease: summary of a clinical practice guideline from the American College of Physicians/American College of Cardiology Foundation/American Heart Association/American Association for Thoracic Surgery/Preventive Cardiovascular Nurses Association/Society of Thoracic Surgeons. *Ann Intern Med*. 2012;157:729-734.
5. Mark DB, Berman DS, Budoff MJ, et al. ACCF/ACR/AHA/NASCI/SAIP/SCAI/SCCT 2010 expert consensus document on coronary computed tomographic angiography: a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents. *Circulation*. 2010;121:2509-2543.
6. Qaseem A, Fihn SD, Dallas P, et al. Management of stable ischemic heart disease: summary of a clinical practice guideline from the American College of Physicians/American College of Cardiology Foundation/American Heart Association/American Association for Thoracic Surgery/Preventive Cardiovascular Nurses Association/Society of Thoracic Surgeons. *Ann Intern Med*. 2012;157:735-743.
7. Montalescot G, Sechtem U, Achenbach S, et al. 2013 ESC guidelines on the management of stable coronary artery disease: the Task Force on the management of stable coronary artery disease of the European Society of Cardiology. *Eur Heart J*. 2013;34:2949-3003.
8. Stone NJ, Robinson J, Lichtenstein AH, et al. 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129:S1-S45.
9. Belsey J, Savelieva I, Mugelli A, et al. Relative efficacy of antianginal drugs used as add-on therapy in patients with stable angina: a systematic review and meta-analysis. *Eur J Prev Cardiol*. 2014 [Epub ahead of print].
10. Caruba T, Katsahian S, Schramm C, et al. Treatment for stable coronary artery disease: a network meta-analysis of cost-effectiveness studies. *PLoS ONE*. 2014;9:e98371.
11. Henry TD, Satran D, Hodges JS, et al. Long-term survival in patients with refractory angina. *Eur Heart J*. 2013;34:2683-2688.

72

ACUTE CORONARY SYNDROME: UNSTABLE ANGINA AND NON-ST ELEVATION MYOCARDIAL INFARCTION

RICHARD A. LANGE AND L. DAVID HILLIS

DEFINITION

The term *acute coronary syndrome* (ACS) is used to describe the continuum of myocardial ischemia (unstable angina pectoris) or infarction (with or without concomitant ST segment elevation). The patient with *unstable angina* has cardiac chest pain that is new, worsening (i.e., more severe, prolonged, or frequent than previous episodes of angina), or occurring at rest, *without* serologic evidence of myocyte necrosis—that is, no elevation of serum concentrations of troponin or the MB isoenzyme of creatine kinase (CK-MB). The patient with cardiac chest pain *with* serologic evidence of myonecrosis and without ST segment elevation is said to have a *non–ST segment elevation myocardial infarction* (MI). Because unstable angina and non–ST segment elevation MI are characterized by the absence of ST segment elevation, they are collectively termed *non–ST segment elevation ACS*, or NSTE ACS (Fig. 72-1). The patient with acute-onset cardiac chest pain, serologic evidence of myonecrosis, and persistent (>20 minutes) ST segment elevation is said to have an *ST segment elevation MI* (Chapter 73).

EPIDEMIOLOGY

Almost 1.2 million individuals in the United States are hospitalized annually with ACS, of whom approximately two thirds have NSTE ACS. More than

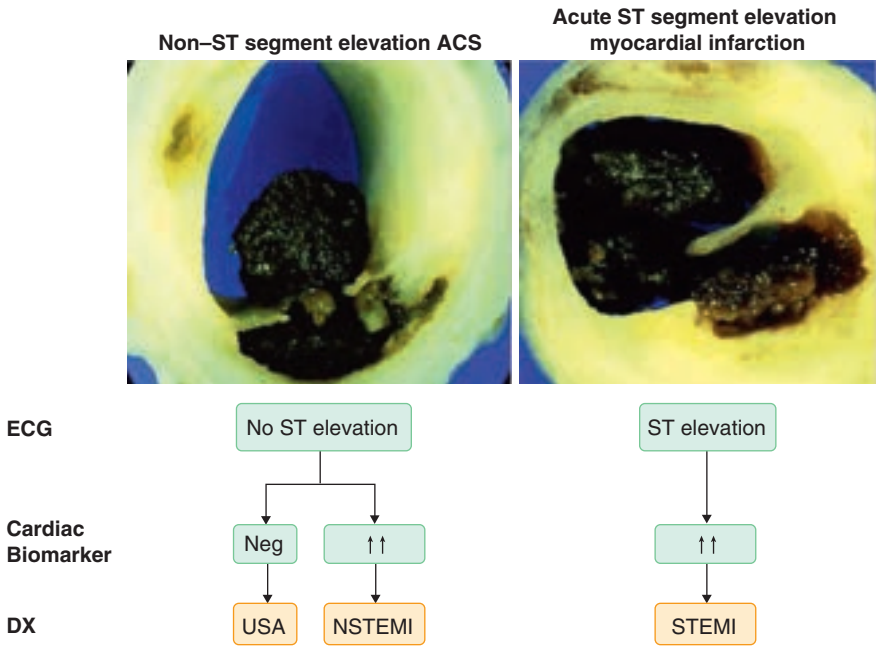

FIGURE 72-1. Acute coronary syndrome (ACS). Symptomatic, morphologic, electrocardiographic, and serologic findings in patients with various kinds of ACS. Subjects with ACS usually complain of chest pain. If the involved coronary artery is totally occluded by fresh thrombus (shown on the *right*), the patient's electrocardiogram (ECG) reveals ST segment elevation, cardiac biomarkers subsequently are elevated, and the patient is diagnosed with an ST segment elevation myocardial infarction (STEMI). If the involved coronary artery is partially occluded by fresh thrombus (shown on the *left*), the patient's ECG does not show ST segment elevation. If cardiac biomarkers are not elevated, the patient is diagnosed with unstable angina (USA). If cardiac biomarkers are elevated, the patient is diagnosed with a non–ST segment elevation MI (NSTEMI). Dx = diagnosis.

half of those with NSTE ACS are older than 65 years, and almost half are women. NSTE ACS is more common in individuals with one or more risk factors for atherosclerosis (Chapter 52), peripheral vascular disease, or a chronic inflammatory disorder, such as rheumatoid arthritis, psoriasis, or infection.

Most subjects with ACS have so-called primary ACS, which is precipitated by rupture of a coronary arterial atherosclerotic plaque, with subsequent platelet aggregation and thrombus formation, leading in turn to diminished blood flow in the involved artery. An occasional individual has so-called secondary ACS, which is caused by a transient or sustained marked imbalance between myocardial oxygen supply and demand. Substantial reductions in oxygen supply, for example, can be caused by severe systemic arterial hypotension, anemia, or hypoxemia; dramatic increases in oxygen demand can be caused by tachycardia, severe systemic arterial hypertension, or thyrotoxicosis. In the subject thought to have secondary ACS, therapy should be directed at the underlying cause.

PATHOBIOLOGY

The precipitating event in almost all subjects with NSTE ACS is coronary arterial atherosclerotic plaque rupture or erosion, with subsequent platelet aggregation and thrombus formation, leading to subtotal occlusion of the involved artery.[1] In an occasional patient, total thrombotic occlusion of the artery leads to NSTE ACS rather than to ST segment elevation MI when extensive collateral blood supply perfuses the region of myocardium that is distal to the occluded artery.

Rarely, intense vasospasm of a segment of an epicardial coronary artery, due to focal endothelial dysfunction (i.e., Prinzmetal's angina) or drug ingestion (caused, for example, by cocaine, chemotherapeutic agents, or one of the serotonin receptor agonist "triptans"), causes a transient or sustained compromise of coronary arterial blood flow, with resultant NSTE ACS. Spontaneous coronary arterial dissection, which occurs most often in peripartum women and patients with vasculitis, may result in NSTE ACS.

Plaque Rupture

Coronary arterial atherosclerotic plaque rupture (Chapter 70) or erosion is the initiating event in most patients with NSTE ACS. Several factors may play a role in the deterioration of the protective fibrous cap that separates the atheroma in the vessel wall from the coronary arterial lumen. Local and systemic inflammation, mechanical features, and anatomic changes contribute to the transformation of a stable atherosclerotic plaque to a so-called vulnerable plaque, the rupture of which triggers platelet adherence, activation, and aggregation, with subsequent thrombus formation.

The deposition of oxidized low-density lipoprotein in the coronary arterial wall stimulates an inflammatory response, which results in the accumulation of macrophages and T lymphocytes at the plaque border. These inflammatory cells secrete cytokines (e.g., tissue necrosis factor, interleukin-1, interferon-γ), which inhibit collagen synthesis and deposition, as well as enzymes (e.g., matrix metalloproteinases and cathepsins), which promote collagen and elastin degradation, thereby rendering the overlying fibrous cap vulnerable to rupture.

Systemic inflammation may play a role in plaque rupture, as evidenced by the predisposition to development of ACS in individuals with chronic gingivitis, rheumatoid arthritis, and chronic or acute infection. Angiographic and angioscopic studies of the coronary arteries of ACS patients often demonstrate plaque ulceration and thrombosis at more than one site, thereby suggesting that a systemic and diffuse inflammatory process is present.

The mechanical characteristics and location of coronary arterial plaques appear to influence their stability. For example, thin fibrous caps are more likely to erode or to rupture than are thick ones. Sites of low shear stress, such as vessel bifurcations, have reduced production of endothelial vasodilator substances (i.e., nitric oxide and prostacyclin), accelerated accumulation of lipids and inflammatory cells, increased degradation of the extracellular matrix, and thinning of the fibrous cap, all of which contribute to plaque instability.

Detailed histologic examination of evolving atherosclerotic plaques reveals a rich neovascularization that is the result of angiogenic peptides, such as fibroblast growth factors, vascular endothelial growth factor, placental growth factor, oncostatin M, and hypoxia-inducible factor, which are secreted by smooth muscle cells, inflammatory cells, and platelets. This neovascularization contributes to the growth of atheroma and to leukocyte trafficking, plaque hemorrhage, and destabilization.

Thrombus Formation

Platelets play a pivotal role in the pathobiology of ACS. After erosion or rupture of a vulnerable plaque, circulating platelets *adhere* to the exposed subendothelial proteins, after which they are *activated*. With activation, the platelets change shape from discoid to stellate, thereby increasing the surface area on which thrombin formation can occur. The platelets then release the contents of their intracellular granules (i.e., thromboxane, serotonin, adenosine diphosphate [ADP], von Willebrand factor, fibrinogen) into the immediate environment and promote focal vasoconstriction of the adjacent arterial segment and activation of nearby platelets. Platelets also increase the number of glycoprotein IIb/IIIa receptors on their surface and the affinity of these receptors to bind circulating fibrinogen. The result is platelet *aggregation*,

which occurs as fibrinogen binds to the glycoprotein IIb/IIIa receptors of adjacent platelets, thereby creating a "platelet plug."

With formation of the platelet plug, the coagulation system activates and generates thrombin, which is a powerful stimulator of further platelet activation and aggregation. Thrombin also converts fibrinogen to fibrin, which is incorporated into the thrombus. Subtotal coronary arterial occlusion by this platelet-rich thrombus compromises blood flow in the involved artery, thereby resulting in an imbalance of oxygen supply and demand of the myocytes perfused by the artery. Distal embolization of platelet-rich thrombi from the site of a ruptured plaque may contribute to the compromise in blood flow. If the supply-demand imbalance is transient, the involved myocytes become ischemic but do not die because the ischemia is of insufficient duration to cause necrosis. The patient typically complains of cardiac chest pain at rest, but serologic evidence of myonecrosis, as evidenced by elevated serum concentrations of troponin or CK-MB, is absent; a diagnosis of *unstable angina* is made. In contrast, if the supply-demand imbalance is sustained, ischemic myocytes begin to die, and infarction occurs. The patient typically complains of cardiac chest pain at rest, and serologic evidence of myonecrosis confirms the diagnosis of *non–ST segment elevation MI*.

CLINICAL MANIFESTATIONS

Symptoms

The patient with NSTE ACS typically complains of retrosternal pressure, squeezing, or heaviness that may be intermittent and recurrent or persistent (Chapter 51). If the episodes are intermittent and recurrent, the duration of each episode may range from only a few minutes to several hours. The chest pain may radiate to the left arm, neck, or jaw, and it may be accompanied by diaphoresis, nausea, abdominal pain, dyspnea, or syncope.

Atypical presentations of NSTE ACS are not uncommon and may include aching or vague chest discomfort, epigastric pain, acute-onset indigestion, unexplained fatigue, or dyspnea. Such atypical complaints are often observed in younger (25 to 40 years of age) and older (>75 years of age) patients, women, and patients with diabetes mellitus, chronic renal insufficiency, or dementia.

Physical Examination

The patient with NSTE ACS often has normal findings on physical examination. On occasion, evidence of left ventricular dysfunction (Chapter 58), such as basilar rales or a ventricular gallop, hypotension, or peripheral hypoperfusion, may accompany an episode of NSTE ACS or appear shortly thereafter. An important goal of the physical examination is to exclude other potential causes of the patient's symptoms, including both noncardiac causes (i.e., costochondritis, pneumothorax, pulmonary embolism, pneumonia) and other cardiac disorders not attributable to myocardial ischemia (i.e., aortic dissection, pericarditis, severe systemic arterial hypertension, arrhythmias; Chapter 51). Accordingly, differences in blood pressure between the upper and lower limbs, decreased lung sounds, friction rubs, and pain on sternal palpation suggest a diagnosis other than NSTE ACS. Other findings on physical examination—such as an elevated blood pressure, tachycardia, pallor, or increased sweating or tremor—point toward precipitating conditions, such as uncontrolled hypertension (Chapter 67), arrhythmias (Chapters 64 and 65), anemia (Chapter 158), or thyrotoxicosis (Chapter 226).

DIAGNOSIS

The patient with suspected ACS should be evaluated promptly because an expedient and accurate diagnosis permits the timely initiation of appropriate therapy, which can reduce the rate of complications. The initial assessment should be directed at determining whether the subject's symptoms are likely to be caused by myocardial ischemia, MI, or some other disorder. The likelihood of ACS can be estimated from the history, physical examination, and electrocardiogram (ECG) (Table 72-1). In the acute setting, the presence or absence of traditional risk factors for atherosclerosis is less important for determining the presence or absence of ACS than are the patient's symptoms, ECG findings, and serologic evidence of myonecrosis. As a result, these long-term risk factors are not integral in determining whether an individual should be evaluated, hospitalized, or treated for ACS.

Conditions that increase the likelihood that the symptomatic patient is experiencing myocardial ischemia or MI include older age, male gender, diabetes mellitus, extracardiac vascular disease, and chest pain radiating to the left arm, neck, or jaw as the presenting symptom. Myocardial ischemia is highly likely if anginal symptoms are accompanied by ECG abnormalities (i.e., Q waves, ST segment depression or elevation ≥1 mm in magnitude, or T wave inversion in multiple precordial leads) or elevated serum concentrations of troponin or CK-MB.

TABLE 72-1 LIKELIHOOD THAT SYMPTOMS AND SIGNS REPRESENT AN ACUTE CORONARY SYNDROME CAUSED BY CORONARY ARTERIAL PLAQUE RUPTURE

HIGH LIKELIHOOD
Any of the Following Features
Chest or left arm pain as the main symptom, similar in nature to previously noted angina
Known coronary artery disease
Evidence on physical examination of transient murmur of mitral regurgitation, hypotension, diaphoresis, or pulmonary edema
New or transient ST segment deviation (≥1 mm) or T wave inversion in multiple precordial leads
Elevated serum troponin or CK-MB concentration
INTERMEDIATE LIKELIHOOD
Absence of High-Likelihood Features and any of the Following
Chest or left arm discomfort as main symptom
Age > 70 years
Male gender
Diabetes mellitus
Extracardiac vascular disease
Q waves, ST segment depression (0.5-1 mm), or T wave inversion (>1 mm) in leads with dominant R waves
Normal cardiac troponin or CK-MB
LOW LIKELIHOOD
Absence of High- or Intermediate-Likelihood Features, but May Have
Probable ischemic symptoms in the absence of any of the intermediate likelihood characteristics
Recent cocaine use
Chest discomfort reproduced by palpation
T wave flattening or inversion < 1 mm in leads with dominant R waves
Normal electrocardiogram
Normal serum troponin or CK-MB concentration

Modified from Anderson JL, Adams CD, Antman EM, et al. ACC/AHA 2007 guidelines for the management of patients with unstable angina/non-ST-elevation myocardial infarction. *Circulation.* 2007;116:e148-e304.

In a patient with known coronary artery disease, typical symptoms are likely to be caused by myocardial ischemia or MI rather than by another condition, particularly if the patient confirms that the symptoms are similar to previous anginal episodes. Conversely, a young individual who has a normal ECG and no risk factors for atherosclerosis is unlikely to be having ACS even when complaining of chest pain with features consistent with ischemia or infarction.

It is important to inquire about the use of cocaine and methamphetamines in the patient with suspected ACS, especially in those who are younger than 40 years or have few traditional risk factors for atherosclerosis. These drugs can increase myocardial oxygen demand and concomitantly decrease oxygen supply by causing vasospasm and thrombosis. A urine toxicologic analysis should be considered when substance abuse is suspected as a cause of ACS.

Electrocardiogram

An ECG, which should be obtained and examined promptly in the patient with suspected ACS, is particularly valuable if it is obtained during a symptomatic episode. If the patient has persistent (>20 minutes) ST segment elevation, prompt reperfusion therapy should be initiated (Chapter 73). Transient ST segment abnormalities that develop during a symptomatic episode at rest and resolve when the patient is asymptomatic strongly suggest NSTE ACS. ST segment depression (or transient ST segment elevation) and T wave abnormalities occur in up to 50% of NSTE ACS patients.

A completely normal ECG does not exclude the possibility of NSTE ACS; in fact, about 5% of patients who are discharged from the emergency department and ultimately diagnosed with ACS have a normal ECG. Ischemia or infarction in the territory of the left circumflex coronary artery often escapes detection with a standard 12-lead ECG, but it may be detected with right-sided leads (V_4R and V_3R) or posterior leads (V_7 to V_9). In the patient whose initial ECG is normal, subsequent ECGs should be obtained in the first 24 hours and during symptomatic episodes, and they should be compared with previous tracings to identify new ST segment or T wave abnormalities. Deep

(>2 mm), symmetrical T wave inversion in the anterior chest leads is often associated with a hemodynamically significant stenosis of the left main or proximal left anterior descending coronary artery.

Serum Biomarkers

With modern high-sensitivity assays, troponin (Chapter 73) is detectable in the blood within 2 hours of the onset of symptoms in patients with non–ST segment elevation MI.[2] An undetectable high-sensitivity troponin level at presentation to the hospital reduces the probability of acute MI to less than 1%.[3] With conventional assays, troponin elevations are generally detectable within 4 hours of the onset of myonecrosis, but detection may be delayed for up to 8 hours in some individuals. Because a single normal serum troponin measurement is insufficient to exclude MI in a patient with recent symptoms, patients with suspected MI are generally observed, either in the emergency department or in a chest pain evaluation unit, with a repeated troponin measurement (and a repeated ECG) 2 to 6 hours later or whenever chest pain recurs.[4] Serum troponin levels also help in acute risk stratification of all ACS patients at the time of the patient's arrival to the hospital.

Serum troponin concentrations can be measured with point-of-care instruments at the patient's bedside by desktop devices or handheld rapid qualitative assays. The advantage of point-of-care systems for avoiding delays must be weighed against their higher costs and the need for stringent quality control. In addition, point-of-care assays are qualitative or semiquantitative and observer dependent, whereas the central laboratory provides more accurate quantitative information concerning biomarker concentrations.

Up to one third of ACS patients whose serum CK-MB concentrations are normal have detectable serum concentrations of troponin T and I, indicating that myonecrosis has occurred and establishing the diagnosis of non–ST segment elevation MI. Current recommendations call for the use of the serum troponin concentration for acute risk stratification at the time of the patient's arrival to the hospital.

Noninvasive Testing

The patient considered to have a low likelihood for ACS (on the basis of the history, physical examination, ECG, and serum biomarkers) should undergo timely stress testing (Chapter 51). Although stress testing does not absolutely establish or exclude the presence of coronary artery disease, it has the advantage of also defining a patient's exercise tolerance, which helps tailor therapeutic decisions. Alternatively, multidetector coronary computed tomographic angiography, which has a high (>98%) negative predictive value to exclude coronary artery disease when it is performed and interpreted at experienced centers, can help reduce hospital stay when the findings are normal in emergency department patients at low to intermediate risk of possible ACS. Conversely, the patient who is believed to be at higher risk for ACS or who continues to have typical ischemic chest pain with ECG abnormalities or elevated cardiac biomarkers should not undergo stress testing or coronary computed tomographic angiography but rather should either undergo coronary angiography or be rendered symptom free with medical therapy before stress testing.

An echocardiogram may be helpful in the patient with chest pain if the ECG is nondiagnostic (i.e., minimal ST segment or T wave abnormalities). If left ventricular hypokinesis or akinesis is observed during an episode of chest pain and then improves when symptoms resolve, myocardial ischemia is likely. In the patient with anterior T wave inversion of uncertain etiology, hypokinesis of the left ventricular anterior wall suggests that the observed T wave abnormality is due to a severe stenosis of the left anterior descending coronary artery. Because echocardiography can help evaluate and identify alternative causes for the patient's chest pain (i.e., myocarditis [Chapter 60], aortic dissection [Chapter 78], or pulmonary embolism [Chapter 98]), it is recommended in patients whose diagnosis is uncertain.

Coronary Angiography

Coronary angiography (Chapter 57) should be performed in patients who are thought to be at high risk for death, MI, or recurrent ischemia in the ensuing days, weeks, and months (see later); in patients who have spontaneous or inducible myocardial ischemia despite appropriate medical therapy; and in patients who have a confusing or difficult clinical presentation and a subsequent inconclusive noninvasive evaluation. The results of angiography help determine whether revascularization is appropriate and, if so, whether it should be attempted by coronary artery bypass grafting or percutaneous coronary intervention (PCI) (Chapter 74).

In patients with NSTE ACS, coronary angiography demonstrates significant stenosis of the left main coronary artery in about 15% of patients, of all three major epicardial coronary arteries in about 30 to 35% of patients, of two of the three epicardial arteries in about 20 to 30% of patients, and of one major epicardial artery in 20 to 30% of patients. About 15% of patients have no coronary arterial narrowing of hemodynamic significance. Women with NSTE ACS are likely to have less extensive coronary artery disease than men have, and patients with non–ST segment elevation MI usually have more extensive disease than those with unstable angina.

On angiography, the coronary arterial lesion responsible for NSTE ACS (the so-called culprit lesion) typically is asymmetrical or eccentric, with scalloped or overhanging edges and a narrow base or neck, features that reflect underlying plaque disruption and thrombus formation. Although obvious thrombus is visible by angiography in only one third of patients with NSTE ACS, coronary angioscopy shows plaque rupture with overlying thrombus in the majority. Interestingly, the lesion that is the nidus for ACS often is not severely stenotic when it is assessed on recently performed angiograms; in fact, two thirds of culprit lesions previously had less than 50% luminal diameter narrowing (and therefore would not have been considered appropriate for revascularization).

Risk Assessment and Triage

The initial evaluation of the patient with possible or suspected ACS should focus on an assessment of the patient's risk of acutely sustaining a cardiac ischemic event (death, MI, or recurrent ischemia). Patients considered to be at low risk for a cardiac ischemic event may be observed in a chest pain evaluation unit for several hours, with repeated troponin level and ECG. If the findings of that brief evaluation are normal, the patient should be discharged home, with further evaluation performed on an outpatient basis. Conversely, patients not at low risk should be hospitalized for further evaluation and treatment (Fig. 72-2; see Fig. 51-1).

After the initial triage decision is made, therapeutic interventions are based on the risk of adverse events in the ensuing hours, days, weeks, and months—estimated by either the Thrombolysis in Myocardial Infarction (TIMI) or Global Registry of Acute Coronary Events (GRACE) risk algorithm—balanced against the risk of a bleeding complication from intensive medical therapy (Table 72-2) or an adverse event from an invasive cardiac procedure. On the basis of this initial assessment, the patient's therapy should be tailored to minimize the likelihood of adverse events.

Although serum markers of myonecrosis represent only one of the TIMI or GRACE risk variables, the presence of this variable alone identifies the patient as being at high risk. However, although elevated serum markers indicate myonecrosis, they provide no insight into its cause because myonecrosis can occur with disease entities other than coronary artery disease (e.g., pulmonary embolism, decompensated heart failure, severe hypertension or tachycardia, anemia, sepsis). Thus, in evaluation of the patient with possible ACS, the presence of elevated serum markers should be assessed in conjunction with other variables.

Increasing age is associated with a higher incidence of both ACS-related cardiac ischemic events and complications from intensive medical therapy and invasive cardiac procedures. Even though elderly individuals are at increased risk of treatment-related complications, they nonetheless derive a greater absolute and relative benefit from such intensive therapy compared with younger individuals. Apart from this initial risk assessment, the ACS patient's general medical and cognitive status, anticipated life expectancy, and, most important, personal preferences should be evaluated and considered.

Once the risk status of the ACS patient is established, therapy is initiated and tailored to the patient's risk of sustaining a subsequent ischemic cardiac event or a treatment-related complication (Table 72-3).[5] For example, the patient considered to be at low risk of a subsequent ischemic event does not benefit from intensive antithrombotic therapy or routine coronary angiography and revascularization. Conversely, in patients considered to be at high risk of sustaining an ischemic event, optimal therapy—including coronary angiography and revascularization (if appropriate)—results in a 20 to 40% decrease in the risk of recurrent ischemia and MI and an approximately 10% reduction in mortality.

Differential Diagnoses

Several cardiac and noncardiac conditions, some of which are potentially life-threatening, may mimic NSTE ACS. The patient with a pulmonary embolism (Chapter 98) will often complain of chest pain and dyspnea and may have ECG abnormalities and an elevated serum troponin concentration. Aortic dissection (Chapter 78) should be considered and excluded because the therapies for NSTE ACS are contraindicated in patients with this

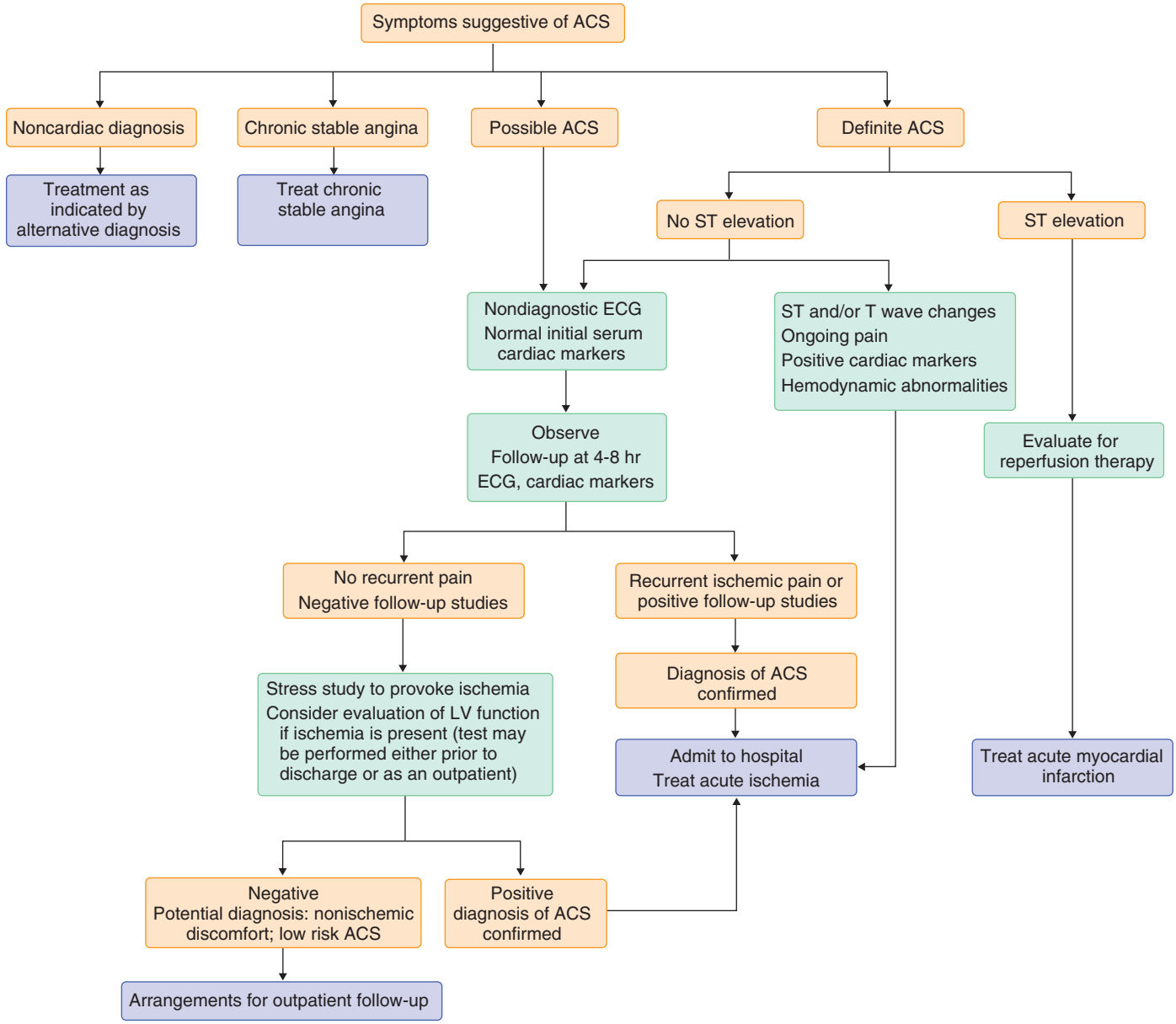

FIGURE 72-2. **Initial triage for patients with symptoms suggestive of an acute coronary syndrome (ACS).** ECG = electrocardiogram; LV = left ventricular. (Modified from Anderson JL, Adams CD, Antman EM, et al. ACC/AHA 2007 guidelines for the management of patients with unstable angina and non-ST-segment elevation myocardial infarction. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation.* 2007;116:e148-e304.)

condition. Stroke (Chapter 407) and subarachnoid hemorrhage (Chapter 408) may be accompanied by ECG abnormalities, left ventricular segmental wall motion abnormalities, and elevated serum biomarker concentrations. Underlying chronic cardiac conditions, such as valvular heart disease (i.e., aortic stenosis, aortic regurgitation) and hypertrophic cardiomyopathy (Chapter 60), may be associated with symptoms similar to those of NSTE ACS, elevated serum biomarker concentrations, and ECG abnormalities. Myocarditis (Chapter 60), pericarditis (Chapter 77), and myopericarditis often cause chest pain that resembles angina, ECG abnormalities, and elevated serum biomarker concentrations; an influenza-like or upper respiratory tract infection often precedes or accompanies these conditions. Patients with "stress cardiomyopathy" (takotsubo syndrome) typically have chest pain, ST segment abnormalities and deeply inverted T waves, and mildly elevated serum biomarker concentrations (Chapter 60).

TREATMENT

Rx

The goals of treatment of the subject with NSTE ACS are to prevent recurrent ischemia (by correcting the imbalance between myocardial oxygen supply and demand), to prevent thrombus propagation, and to stabilize the vulnerable plaque. Antianginal medications, such as nitroglycerin (see Table 71-14), β-adrenergic blockers (see Table 71-12), and calcium-channel blockers (see Table 71-13), favorably affect myocardial oxygen supply and demand, thereby preventing recurrent ischemia. Antiplatelet and antithrombotic agents retard thrombus propagation, and statins promote plaque stabilization. Once the risk status of the ACS patient is established, treatment is initiated (see Table 72-3).

Every NSTE ACS patient, regardless of the level of risk, should promptly receive antianginal medications, antiplatelet therapy, and a statin, unless contraindicated. A patient considered to be at low risk may receive unfractionated heparin, but more intensive anticoagulant therapy is not necessary because such therapy increases the risk of bleeding without further reducing the risk of an ischemic cardiac event. Routine coronary angiography and revascularization are not beneficial and should be reserved for the patient with recurrent ischemia despite intensive medical therapy.

Conversely, the high-risk patient should receive antianginal medications, antiplatelet therapy, a statin, intensive anticoagulant therapy, and coronary angiography followed by revascularization (if indicated). In the patient whose coronary anatomy is suitable, revascularization reduces the incidence of ischemia and recurrent MI, and it also improves survival in certain patients (see later).

Antianginal Therapy

Nitroglycerin

Nitroglycerin (see Table 71-14), which is a venodilator at low doses and an arteriolar dilator at higher doses, may prevent recurrent ischemia in patients with unstable angina, but no studies of sufficient statistical power have

TABLE 72-2 RISK VARIABLES FOR ISCHEMIC EVENTS AND BLEEDING COMPLICATIONS

RISK VARIABLES PREDICTIVE OF DEATH, MYOCARDIAL INFARCTION, OR RECURRENT ISCHEMIA
Thrombolysis in Myocardial Infarction (TIMI) Score*
Age > 65 years
Three or more risk factors for atherosclerosis
Known coronary artery disease (previous coronary arteriography or myocardial infarction)
Two or more episodes of anginal chest pain at rest in the 24 hours before hospitalization
Use of aspirin in the 7 days before hospitalization
ST segment deviation ≥ 0.5 mV
Elevated serum concentrations of troponin or CK-MB
Global Registry of Acute Coronary Events (GRACE)†
Age
Heart failure class
Heart rate
Systolic blood pressure
ST segment deviation
Cardiac arrest during presentation
Serum creatinine concentration
Elevated serum markers of myonecrosis
RISK FACTORS FOR BLEEDING COMPLICATIONS WITH INTENSIVE THERAPY‡
Female gender
Older age
Renal insufficiency
Low body weight
Tachycardia
Systolic arterial pressure (high or low)
Anemia
Diabetes mellitus

*Individuals with three or more of these variables are considered to be at high risk, whereas those with none, one, or two are considered to be at low risk. (From Diez JG, Cohen M. Balancing myocardial ischemic and bleeding risks in patients with non-ST-segment elevation myocardial infarction. *Am J Cardiol.* 2009;103:1396-1402.)

†Each variable is assigned a numerical score on the basis of its specific value, and the eight scores are summed to yield a total score, which is applied to a reference nomogram to determine the patient's risk. The GRACE application tool is available online at www.outcomes-umassmed.org/grace. (From Brieger D, Fox KA, Fitzgerald G, et al. Predicting freedom from clinical events in non-ST-elevation acute coronary syndromes: the Global Registry of Acute Coronary Events. *Heart.* 2009;95:888-894.)

‡The patient's bleeding risk can be estimated with the tool available at www.crusadebleedingscore.org. (From Subherwal S, Bach RG, Chen AY, et al. Baseline risk of major bleeding in non-ST-segment-elevation myocardial infarction: the CRUSADE [Can Rapid risk stratification of Unstable angina patients Suppress ADverse outcomes with Early implementation of the ACC/AHA Guidelines] Bleeding Score. *Circulation.* 2009;119:1873-1882.)

determined whether it reduces the risk of MI in this population of patients. In patients who complain of recurrent symptoms, nitroglycerin should be given sublingually or by buccal spray (0.3 to 0.6 mg). Patients with ongoing or recurrent chest pain should receive intravenous nitroglycerin (5 to 10 µg/minute with use of nonabsorbable tubing), with escalation of the dose in increments of 10 µg/minute until symptoms resolve or adverse effects develop. Nitroglycerin's most common adverse effects are headache, nausea, dizziness, hypotension, and reflex tachycardia.

Nitrate tolerance can be avoided by periodically providing the patient with a nitrate-free period (i.e., a brief cessation of drug administration). Nitroglycerin should not be given to patients who have received a phosphodiesterase-5 inhibitor (i.e., sildenafil, tadalafil, or vardenafil) within the previous 24 to 48 hours as severe hypotension may ensue.

β-Adrenergic Blockers

β-Adrenergic blockers diminish symptoms and the risk of MI in ACS patients who are not already taking a β-blocker at the time of hospitalization. In the normotensive patient without ongoing chest pain or tachycardia, metoprolol should be initiated at 50 mg orally every 6 to 8 hours, with the dose increased (to 100 mg twice daily) as necessary to control heart rate, blood pressure, and symptoms. In high-risk patients and in patients with tachycardia or elevated systemic arterial pressure, metoprolol should be administered intravenously (three boluses of 5 mg each given 5 minutes apart) initially, after which an oral dose should be initiated. A reasonable target heart rate is 50 to 60 beats per minute at rest.

β-Blockers should not be administered to patients with decompensated heart failure, hypotension, hemodynamic instability, or advanced atrioventricular block. Because most patients with chronic obstructive pulmonary disease or peripheral vascular disease tolerate β-blockers without difficulty, these conditions should not preclude their use.

Calcium-Channel Blockers

Calcium-channel blockers, which cause arterial vasodilation, increase coronary arterial blood flow and lower systemic arterial pressure. The nondihydropyridine calcium-channel blockers diltiazem and verapamil slow heart rate and are recommended for the patient with a contraindication to a β-adrenergic blocker or persistent or recurrent symptoms despite treatment with nitroglycerin or a β-blocker. Oral diltiazem (30 to 90 mg four times daily of the short-acting preparation or up to 360 mg once daily of the long-acting preparation) is the preferred calcium-channel blocker because it reduces the incidence of myocardial ischemia and recurrent MI in patients with NSTE ACS. Diltiazem is contraindicated in patients with left ventricular systolic dysfunction or pulmonary vascular congestion. Caution should be exercised when combining a β-blocker with diltiazem because the two drugs may act synergistically to depress left ventricular systolic function as well as sinus and atrioventricular nodal conduction. Patients with ACS should not be prescribed short-acting nifedipine unless they are already receiving a β-blocker because it may increase the risk of death. The risks and benefits of long-acting dihydropyridines in patients with NSTE ACS are undefined.

Antiplatelet Agents

ACS patients should receive dual antiplatelet therapy (aspirin and an ADP receptor inhibitor) acutely and for at least 1 year unless the patient has an aspirin allergy or active bleeding. In patients with NSTE ACS, aspirin (Chapter 37) reduces the risk of death or MI by about 50%.[■] The recommended dose is 75 to 162 mg daily, continued indefinitely. The choice of which ADP receptor antagonist to use in combination with aspirin is determined by each patient's characteristics (i.e., risk of bleeding), medication costs, and pharmacologic properties of the agent (see following details). The patient who is allergic to or intolerant of aspirin should be treated with an ADP receptor inhibitor (clopidogrel, ticagrelor, or prasugrel [if PCI treated]) alone.[5]

Clopidogrel (Chapter 38) is a thienopyridine that blocks the $P2Y_{12}$ ADP receptor, thereby diminishing ADP-mediated platelet activation. Its antiplatelet activity is synergistic with aspirin because the two agents inhibit different platelet-activating pathways. Clopidogrel is a prodrug that must be metabolized by the cytochrome P-450 system to the active form. Polymorphisms in the cytochrome P-450 isoform CYP2C19, which are present in 15 to 20% of individuals, slow metabolism of the prodrug to the active form, thereby reducing the magnitude of platelet inhibition. Drugs that are potent inhibitors of the CYP2C19 enzyme (e.g., omeprazole, esomeprazole, cimetidine, fluconazole, ketoconazole, voriconazole, etravirine, felbamate, fluoxetine, and fluvoxamine) should not be administered with clopidogrel because they affect the metabolism to its active form and reduce its antiplatelet effects.

In subjects with NSTE ACS, the addition of clopidogrel (a loading dose of 300 to 600 mg, then 75 mg daily for at least 1 year) to aspirin reduces the composite end point of cardiovascular death, nonfatal MI, or stroke by 20% (2.1% reduction in absolute risk) compared with treatment with aspirin alone.[■] The benefit of an aspirin-clopidogrel combination is seen as early as 24 hours after drug initiation and persisted for the 12 months of the study, despite an increase in minor bleeding.

Prasugrel (Chapter 38), another thienopyridine, has a greater antiplatelet effect and a more rapid onset of action than clopidogrel. In patients with ACS who are referred for PCI, prasugrel in combination with aspirin reduces ischemic events (i.e., a combination of cardiovascular death, nonfatal MI, and stroke) by 20% compared with concomitant clopidogrel and aspirin (2.2% absolute risk reduction) therapy.[■] However, this benefit is obtained at a 0.5% increased risk of life-threatening bleeding and a 0.3% increased risk of fatal bleeding. At present, prasugrel is approved for use in the ACS patient who is referred for PCI. In combination with aspirin, it is administered as a 60-mg oral loading dose followed by a 10-mg daily maintenance dose. Because prasugrel-associated bleeding complications are highest in patients with a previous stroke or transient ischemic attack, age older than 75 years, or a body weight of less than 60 kg, it should not be used in patients with any of these features.

Ticagrelor (Chapter 38), a thienopyridine that does not require hepatic activation, has more rapid onset and more pronounced platelet inhibition than clopidogrel. It is a reversible inhibitor of the $P2Y_{12}$ receptor, so platelet function returns more rapidly after discontinuation than with clopidogrel. In a randomized trial in NSTE ACS patients, the addition of ticagrelor to aspirin reduced the composite end point of vascular death, nonfatal MI, or stroke by about 15% compared with treatment with clopidogrel and aspirin but increased non–procedure-related bleeding by an absolute 0.7%.[■] In combination with aspirin, ticagrelor is administered as a 180-mg oral loading dose, followed by a 90-mg twice-daily maintenance dose. In patients who receive ticagrelor, the daily aspirin maintenance dose should be 100 mg or less, and ticagrelor should not be used in patients with a history of intracranial hemorrhage.

Glycoprotein IIb/IIIa inhibitors (Chapter 38) block platelet aggregation in response to all potential agonists, so they are the most potent antiplatelet agents available. Three glycoprotein IIb/IIIa inhibitors, each of which must be

TABLE 72-3 MANAGEMENT STRATEGIES FOR PATIENTS WITH ACUTE CORONARY SYNDROME

THERAPY	INITIATION	DURATION	DOSE, ROUTE, AND DURATION	BENEFIT VS. PLACEBO (REDUCED INCIDENCE OF …)
LOW-RISK PATIENT				
Antianginal				
β-Blocker*	Immediately	Hospitalization ± indefinitely	Metoprolol, 5-mg IV boluses (three given 2 to 5 minutes apart), then 50 mg orally twice daily, titrated up to 100 mg twice daily; or atenolol, 5 to 10 mg IV bolus, then 100 mg orally daily	Recurrent ischemia
Nitroglycerin	Immediately	Hospitalization ± indefinitely	0.3-0.6 mg sublingually or 5-10 μg/min IV initially and increased by 10 μg/min every 5 minutes	Not studied
Diltiazem or verapamil*	Immediately	Hospitalization ± indefinitely	30-90 mg orally four times daily or up to 360 mg of long-acting preparation orally daily	MI, recurrent ischemia
Lipid Lowering				
Statin	Before hospital discharge	Indefinitely	Atorvastatin, up to 80 mg orally daily	Recurrent ischemia
Antiplatelet				
Aspirin	Immediately	Indefinitely	162-325 mg orally initial dose, then 81 mg orally daily	Death, MI
Clopidogrel	Immediately	1-12 months	300 mg orally initial dose, then 75 mg orally daily	MI, recurrent ischemia
Anticoagulant				
Unfractionated heparin	Immediately	2 to 5 days	IV bolus of 60 U/kg, then 12 U/kg IV adjusted to achieve an aPTT of 50 to 70 seconds	Death or MI (combined)
HIGH-RISK PATIENT				
Antianginal				
β-Blocker*	Immediately	Hospitalization ± indefinitely	Metoprolol, 5-mg IV boluses (three given 2 to 5 minutes apart), then 50 mg orally twice daily titrated up to 100 mg twice daily; or atenolol, 5- to 10-mg IV bolus, then 100 mg orally daily	Death, MI, recurrent ischemia
Nitroglycerin	Immediately	Hospitalization ± indefinitely	0.3-0.6 mg sublingually or 5-10 μg/min IV initially and increased by 10 μg/min every 5 minutes	Not studied
Diltiazem or verapamil*	Immediately	Hospitalization ± indefinitely	30-90 mg orally four times daily or up to 360 mg of long-acting preparation orally daily	MI, recurrent ischemia
Lipid Lowering				
Statin	Before hospital discharge	Indefinitely	Atorvastatin, up to 80 mg orally daily	Recurrent ischemia
Antiplatelet				
Aspirin *and* Clopidogrel *or* Prasugrel *or* Ticagrelor	Immediately Immediately At time of PCI At time of PCI	Indefinitely ≥12 months 15 months 12 months	162-325 mg orally initial dose, then 81 mg orally 300 mg orally initial dose, then 75 mg orally daily 60 mg orally initial dose, then 10 mg orally daily 180 mg orally initially, then 90 mg twice daily	Death, MI MI, recurrent ischemia Cardiovascular death, MI or stroke (combined)† Vascular death, MI or stroke (combined)†
Glycoprotein IIb/IIIa inhibitor (eptifibatide, tirofiban, or abciximab)	At time of PCI	12-24 hours after PCI	Abciximab, IV bolus of 0.25 mg/kg, then 0.125 μg/kg/min IV (max. 10 μg/min) for 12 hours; or eptifibatide, IV bolus of 180 μg/kg, then 2.0 μg/kg/min IV for 18-24 hours; or tirofiban, 0.4 μg/kg/min IV for 30 minutes, then 0.1 μg/kg/min IV for 12 to 24 hours	MI
Anticoagulants				
Unfractionated heparin *or* Enoxaparin *or* Bivalirudin *or* Fondaparinux	Immediately Immediately Immediately Immediately	2 to 5 days; discontinue after successful PCI Duration of hospitalization (up to 8 days); discontinue after successful PCI Up to 72 hours; discontinue 4 hours after PCI Duration of hospitalization (up to 8 days); if used during PCI, it must be coadministered with another anticoagulant with factor IIa activity	IV bolus of 60 U/kg, then 12 U/kg IV adjusted to achieve an aPTT of 50 to 70 seconds 1 mg/kg subcutaneously twice daily IV bolus of 0.75 mg/kg, then 1.75 mg/kg/hr IV 2.5-mg subcutaneous injection once daily	Death or MI (combined) MI, recurrent ischemia‡ Bleeding§ Bleeding¶

TABLE 72-3 MANAGEMENT STRATEGIES FOR PATIENTS WITH ACUTE CORONARY SYNDROME—cont'd

THERAPY	INITIATION	DURATION	DOSE, ROUTE, AND DURATION	BENEFIT VS. PLACEBO (REDUCED INCIDENCE OF …)
Invasive Management				
Coronary angiography followed by revascularization (if appropriate)	Up to 36-80 hours after hospitalization; within 24 hours in "very high risk" patients			MI, recurrent ischemia

*Avoid in the patient with decompensated heart failure, hypotension, or hemodynamic instability.
†Compared with clopidogrel.
‡Compared with unfractionated heparin.
§As monotherapy compared with heparin and glycoprotein IIb/IIIa inhibitor combination.
¶Compared with enoxaparin.
aPTT = activated partial thromboplastin time; MI = myocardial infarction; PCI = percutaneous coronary intervention.
Modified from Lange RA, Hillis LD. Optimal management of acute coronary syndromes. *N Engl J Med.* 2009;260:2237-2240.

administered parenterally, are available: abciximab is the Fab fragment of a monoclonal antibody to the receptor; eptifibatide is a peptide; and tirofiban is a peptidomimetic molecule.

Glycoprotein IIb/IIIa inhibitors reduce the incidence of recurrent ischemic events in patients with NSTE ACS who undergo PCI but not in patients who are managed with medical therapy alone. When a glycoprotein IIb/IIIa inhibitor is administered to PCI patients, it should be initiated at the time of angiography because its routine administration beforehand carries an increased bleeding risk and no improvement in outcomes. The glycoprotein IIb/IIIa inhibitor infusion (see Table 72-3) typically is continued for 12 to 24 hours after PCI.

Anticoagulants

Anticoagulant therapy should be administered to all patients with ACS unless a contraindication, such as active bleeding, is present. For the patient in whom a noninvasive, ischemia-guided management strategy is selected, treatment with unfractionated heparin, low-molecular-weight heparin (LMWH), or fondaparinux is appropriate, with fondaparinux recommended for the patient at increased risk of bleeding. For the patient in whom an invasive management strategy is selected, unfractionated heparin and LMWH are the agents of choice. Although bivalirudin may be preferred in patients undergoing PCI, it is not used in the initial management of the patient with ACS.

Heparin

Unfractionated heparin (Chapter 38) exerts its anticoagulant effect by accelerating the action of circulating antithrombin; it prevents thrombus propagation but does not lyse existing thrombi. In the patient with NSTE ACS, the addition of heparin to aspirin reduces the rate of in-hospital ischemic events (i.e., death or MI) by 33%.

Unfractionated heparin should be initiated with an intravenous bolus of 60 U/kg, followed by a continuous infusion of approximately 12 U/kg/hour (maximum, 1000 U/hour), adjusted to maintain the activated partial thromboplastin time (aPTT) at 1.5 to 2.5 times control (i.e., 50 to 70 seconds) or a heparin concentration at 0.3 to 0.7 U/mL (by anti–factor Xa determinations). The infusion should be continued for 48 hours or until revascularization is performed, whichever occurs sooner. Frequent monitoring of the aPTT or heparin concentration is necessary because the anticoagulant response to a standard dose of unfractionated heparin varies widely among individuals; even when a weight-based nomogram (see Table 81-4) is followed, the aPTT is outside the therapeutic range more than one third of the time.

Mild thrombocytopenia occurs in 10 to 20% of patients treated with unfractionated heparin. In 1 to 5% of patients, a more severe form of thrombocytopenia develops. This antibody-mediated response usually occurs 4 to 14 days after the initiation of treatment (although it may appear more quickly in patients who received heparin within the preceding 6 months) and is associated with thromboembolic sequelae in 30 to 80% of subjects (Chapter 172).

Low-Molecular-Weight Heparin

LMWHs (Chapter 38), which are fragments of unfractionated heparin, exert a more predictable anticoagulant effect, have a longer half-life, and are less likely to cause thrombocytopenia compared with unfractionated heparin. Because they provide predictable and sustained anticoagulation with once- or twice-daily subcutaneous administration, monitoring of their anticoagulant effect is not required.

LMWH is superior to unfractionated heparin in preventing MI or death during hospitalization in NSTE ACS patients who have elevated serum cardiac biomarkers as well as in those considered to be at high risk for recurrent ischemia (see Table 72-2). In the low-risk subject, unfractionated heparin and LMWH have similar efficacy.

Two LMWHs, enoxaparin and dalteparin, are approved for the treatment of the patient with NSTE ACS. The dose of enoxaparin is 1 mg/kg subcutaneously twice daily, and the dose of dalteparin is 120 IU/kg (maximum, 10,000 IU) subcutaneously twice daily. Therapy should be continued for the duration of the hospitalization, up to 8 days, or until revascularization is performed (whichever occurs first). In obese (>120 kg), thin (<60 kg), or renally impaired (creatinine clearance < 30 mL/minute) patients, the LMWH dose should be adjusted to achieve an anti–factor Xa concentration of 0.5 to 1.5 IU/mL 4 to 6 hours after drug administration. LMWH should be avoided in the patient with a history of heparin-induced thrombocytopenia. In the patient with renal failure, treatment with LMWH has been associated with the development of hyperkalemia.

Fondaparinux

Fondaparinux (Chapter 38), which is a selective factor Xa inhibitor, does not require dose adjustment and monitoring. Fondaparinux does not cause thrombocytopenia. Fondaparinux is as effective as enoxaparin in preventing ischemic cardiac events but with 50% fewer major bleeding episodes (2.2% vs 4.1%). Because an increased incidence of catheter-related thrombosis has been reported after fondaparinux treatment, it is not recommended for patients who are likely to undergo coronary angiography. Fondaparinux is a desirable anticoagulant for the ACS patients who are managed in an ischemia-guided fashion, especially patients at higher risk of a bleeding complication with anticoagulant therapy, but not for other patients.

For the patient with ACS, fondaparinux is administered as a 2.5-mg subcutaneous injection once daily for up to 5 days or until hospital discharge. Its use is contraindicated in patients with severe renal impairment and in those who weigh 50 kg or less, and it should not be used as the sole anticoagulant during a PCI.

Bivalirudin

Bivalirudin, a direct thrombin inhibitor, is currently recommended as an alternative anticoagulant for patients undergoing PCI. Because it has not been tested in patients whose ACS is managed with an ischemia-guided strategy, its administration in a setting other than the cardiac catheterization laboratory is not recommended. In the patient undergoing PCI, bivalirudin (0.75 mg/kg intravenous bolus followed by an infusion of 1.75 mg/kg/hour for up to 4 hours after the PCI) is as effective as combination heparin and glycoprotein IIb/IIIa inhibitor therapy in preventing ischemic events, but it causes fewer major bleeding episodes. Bivalirudin is the anticoagulant of choice for the patient with ACS who has heparin-induced thrombocytopenia.

Statins

Prompt initiation of statin therapy is recommended in all patients with NSTE ACS to promote plaque stabilization and to restore endothelial function. Moreover, when statin therapy is initiated during the patient's hospitalization (rather than at hospital discharge), long-term medical compliance is substantially improved. In the absence of contraindications, high-dose atorvastatin (80 mg daily) should be given orally to the patient with NSTE ACS, regardless of the baseline serum low-density lipoprotein cholesterol concentration; a lower dose is not as effective in reducing ischemic events.

Recurrent or Refractory Unstable Angina

In most patients hospitalized with NSTE ACS, symptoms do not recur after the institution of appropriate antianginal therapy. The occasional patient with continued or recurrent chest pain despite optimal medical therapy is at high risk for an MI. For the patient with refractory myocardial ischemia or hemodynamic instability despite optimal medical therapy, intra-aortic balloon counterpulsation can reduce the incidence of ischemic episodes until revascularization can be performed. Intra-aortic balloon function is

synchronized with the patient's ECG so that it inflates during diastole and deflates during systole, thereby augmenting coronary arterial blood flow and reducing myocardial oxygen demand by decreasing afterload. Intra-aortic balloon counterpulsation causes lower limb ischemia in approximately 3% of patients in whom the device is placed, but this complication usually resolves with its removal.

Coronary Revascularization

Coronary revascularization is performed to relieve angina that is persistent or recurrent despite optimal medical therapy, to prevent recurrent ischemia or MI in patients at high risk for a subsequent ischemic event, and to improve survival in patients with suitable coronary arterial anatomy.

Coronary revascularization is successful in relieving symptoms in 90% of the patients with angina refractory to medical therapy. Whether coronary bypass surgery or PCI is the more appropriate method of revascularization is determined by the location and severity of coronary arterial stenoses and the presence of comorbid medical conditions that may affect the performance or safety of the revascularization procedure.[6,7]

The patient who has been rendered symptom free with optimal medical therapy should undergo an assessment to determine whether he or she is at high risk or relatively low risk of sustaining a cardiac ischemic event (death, MI, or recurrent ischemia) in the ensuing days, weeks, and months and a bleeding complication from intensive medical therapy or an invasive cardiac procedure. Patients who are at low risk of having a subsequent ischemic event (those with a normal serum troponin concentration, age younger than 75 years, and zero, one, or two TIMI risk variables) should be evaluated noninvasively for inducible ischemia before hospital discharge. If the patient has spontaneous or provocable ischemia, coronary angiography and, if appropriate, revascularization should be performed.

The ACS patient with a detectable serum troponin concentration, age older than 75 years, or three or more TIMI risk variables is considered to be at high risk for a subsequent event and should be referred for routine coronary angiography and revascularization (if appropriate) during the hospitalization because this management strategy reduces the incidence of subsequent ischemic cardiac events. In most subjects, early (within 24 hours of hospitalization) invasive therapy is no better at preventing death, MI, or stroke than somewhat delayed (median, 50 hours) invasive management, although it is associated with a modest decrease in the occurrence of recurrent ischemia. In contrast, in the one third of subjects considered to be at very high risk (GRACE risk score of >140, corresponding to an incidence of in-hospital death or MI of >20%), an early invasive management strategy is superior to a delayed strategy in reducing the incidence of death, MI, or stroke.

The patient with clinical features or noninvasive test results suggestive of severe coronary artery disease (i.e., left ventricular dysfunction, hemodynamic instability, life-threatening ventricular arrhythmias, or extensive inducible ischemia) should be referred for coronary angiography (Table 72-4) to determine if left main or three-vessel coronary arterial disease is present because patients with these coronary anatomic findings derive a survival benefit with coronary revascularization compared with medical therapy (Chapter 74). In patients who are taking an ADP receptor inhibitor and in whom coronary artery bypass grafting can be delayed, the drug should be discontinued (5 days for clopidogrel or ticagrelor and at least 7 days for prasugrel) to allow dissipation of the antiplatelet effect.

Complications

Patients with NSTE ACS can develop recurrent ischemic events or any of the complications associated with ST segment elevation MI, including arrhythmias, heart failure, and mechanical complications (Chapter 73). However, the acute complications other than recurrent ischemia occur less often in subjects with NSTE ACS because the amount of myocardial damage usually is less.

Because intensive medical therapy in conjunction with invasive management can lead to life-threatening bleeding complications, the patient's risk of such should be assessed before these therapies are instituted. Female gender, older age, renal insufficiency, low body weight, tachycardia, systolic arterial pressure, hematocrit, and diabetes mellitus predict an increased risk of major bleeding, often due to excessive dosing of antiplatelet or anticoagulant agents. The bleeding risk can be estimated with the tool available at www.crusadebleedingscore.org.

Integrated Approach to Treatment

Although the treatment of the subject with NSTE ACS should be individualized, taking into account the specific features of the disease and the particular circumstances of the patient, algorithms nonetheless provide a useful framework (Fig. 72-3). Smoking cessation (Chapter 32), cholesterol lowering (Chapter 206), and control of blood pressure (Chapter 67), obesity, and diabetes mellitus (Chapter 229) are important long-term prevention strategies.[8-11] Maintaining compliance long term with medical therapy appears to reduce the risk of a future ischemic event by up to 80%.

TABLE 72-4 SELECTION OF INITIAL TREATMENT STRATEGY: INVASIVE VERSUS CONSERVATIVE

GENERALLY PREFERRED STRATEGY	PATIENT CHARACTERISTICS
Invasive	Recurrent angina or ischemia at rest or with low-level activities despite intensive medical therapy Elevated cardiac biomarkers (troponin) New or presumably new ST segment depression Signs or symptoms of heart failure or new or worsening mitral regurgitation High-risk findings from noninvasive testing Hemodynamic instability Sustained ventricular tachycardia PCI within 6 months Previous CABG High-risk score (e.g., TIMI, GRACE) Mild to moderate renal dysfunction Diabetes mellitus Reduced left ventricular systolic function (EF < 40%)
Conservative	Low-risk score (e.g., TIMI, GRACE) Patient or physician preference in the absence of high-risk features

CABG = coronary artery bypass grafting; EF = left ventricular ejection fraction; GRACE = Global Registry of Acute Coronary Events; HF = heart failure; PCI = percutaneous coronary intervention; TIMI = Thrombolysis in Myocardial Infarction.

PROGNOSIS

Because the number of ECG leads demonstrating ST segment depression and the magnitude of such depression are indicative of the extent and severity of myocardial ischemia and MI, it is not surprising that ST segment depression correlates with the patient's prognosis. Compared with subjects without ST segment depression, the patient with NSTE ACS who has ST segment depression of 1 mm or greater in two or more leads is almost four times as likely to die within 1 year, and the patient with ST segment depression of 2 mm or greater in magnitude is almost six times as likely to die within 1 year. If ST segment depression of 2 mm or greater is present in more than one region of the ECG, the mortality is increased 10-fold. Even the 20% of patients with ACS who have only 0.5 to 1 mm of ST segment depression have an adverse prognosis. Patients with ST segment depression also have a higher risk for subsequent cardiac events compared with patients with only T wave inversions (>1 mm).

The magnitude of the serum troponin concentration predicts short-term (30 days) and long-term (1 year) risks of recurrent MI and death, independent of ECG abnormalities or markers of inflammatory activity. C-reactive protein measured with a highly sensitive assay, which is a widely used marker of inflammation, has no role in the diagnosis of ACS but is predictive of long-term (6 months) mortality among patients with troponin-negative NSTE ACS. Elevated serum concentrations of natriuretic peptides (B-type natriuretic peptide [BNP] or its N-terminal prohormone [NT-pro-BNP]) are associated with a three- to five-fold increased mortality in patients with NSTE ACS, although they have limited value for diagnosis, initial risk stratification, and selection of an initial management strategy. Natriuretic peptide concentrations measured a few days after the onset of symptoms have better predictive value than those measured at the time of hospitalization. In patients with NSTE ACS, a simultaneous assessment of troponin, high-sensitivity C-reactive protein, and BNP is superior to a single biomarker assessment at predicting short-term outcome.

In contrast to patients with ST segment elevation MI, in whom most events occur before or shortly after presentation to the hospital, patients with NSTE ACS continue to be at high risk for such events during the ensuing days, weeks, and months. Although in-hospital mortality is higher in patients with ST segment elevation MI than among those with NSTE ACS (7 vs. 5%, respectively), the mortality rates at 6 months are similar for the two conditions (12 vs. 13%, respectively). During long-term follow-up of patients hospitalized with ACS, rates of death are actually higher in those with NSTE ACS than in those with ST segment elevation MI, with a two-fold difference after 4 years. As a result, treatment strategies for NSTE ACS should address the issues related to both the acute event and longer-term treatment.

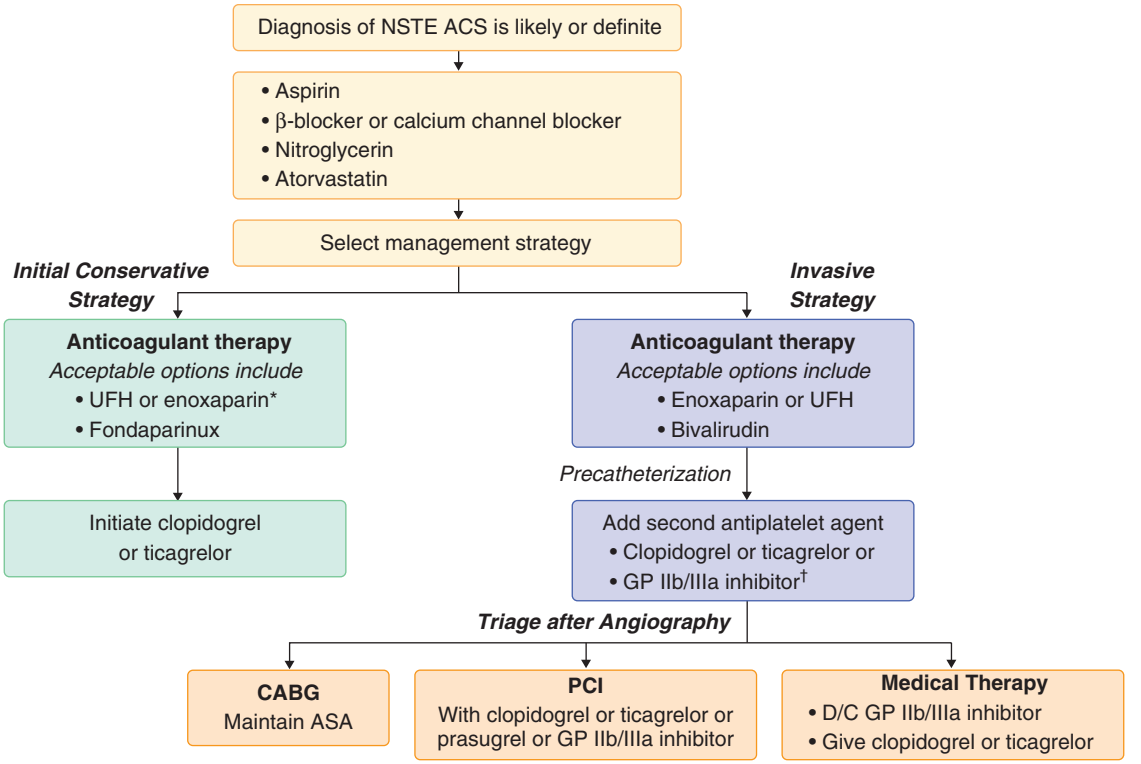

FIGURE 72-3. Approach to the patient with non–ST segment elevation acute coronary syndrome (NSTE ACS). *Enoxaparin or fondaparinux is preferred to unfractionated heparin (UFH). †Intravenous eptifibatide or tirofiban is preferred. ASA = aspirin; CABG = coronary artery bypass grafting; D/C = discontinue; GP = glycoprotein; PCI = percutaneous coronary intervention.

Grade A References

A1. Litt HI, Gatsonis C, Snyder B, et al. CT angiography for safe discharge of patients with possible acute coronary syndromes. *N Engl J Med*. 2012;366:1393-1403.

A2. Mehta SR, Granger CB, Boden WE, et al. Early versus delayed invasive intervention in acute coronary syndromes. *N Engl J Med*. 2009;360:2165-2175.

A3. O'Donoghue ML, Vaidya A, Afsal R, et al. An invasive or conservative strategy in patients with diabetes mellitus and non-ST-segment elevation acute coronary syndromes: a collaborative meta-analysis of randomized trials. *J Am Coll Cardiol*. 2012;60:106-111.

A4. Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ*. 2002;324:71-86.

A5. Yusuf S, Zhao F, Mehta SR, et al. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. *N Engl J Med*. 2001;345:494-502.

A6. Wiviott SD, Braunwald E, McCabe CH, et al. Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med*. 2007;357:2001-2015.

A7. Kohli P, Wallentin L, Reyes E, et al. Reduction in first and recurrent cardiovascular events with ticagrelor compared with clopidogrel in the PLATO Study. *Circulation*. 2013;127:673-680.

A8. Eikelboom JW, Anand SS, Malmberg K, et al. Unfractionated heparin and low-molecular-weight heparin in acute coronary syndrome without ST elevation: a meta-analysis. *Lancet*. 2000;355:1936-1942.

A9. Yusuf S, Mehta SR, Chrolavicius S, et al. Comparison of fondaparinux and enoxaparin in acute coronary syndromes. *N Engl J Med*. 2006;354:1464-1476.

A10. Cannon CP, Braunwald E, McCabe CH, et al. Intensive versus moderate lipid lowering with statins after acute coronary syndromes. *N Engl J Med*. 2004;350:1495-1504.

A11. O'Donoghue M, Boden WE, Braunwald E, et al. Early invasive vs conservative treatment strategies in women and men with unstable angina and non-ST-segment elevation myocardial infarction: a meta-analysis. *JAMA*. 2008;300:71-80.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Crea F, Liuzzo G. Pathogenesis of acute coronary syndromes. *J Am Coll Cardiol*. 2013;61:1-11.
2. de Lemos JA. Increasingly sensitive assays for cardiac troponins: a review. *JAMA*. 2013;309: 2262-2269.
3. Body R, Carley S, McDowell G, et al. Rapid exclusion of acute myocardial infarction in patients with undetectable troponin using a high-sensitivity assay. *J Am Coll Cardiol*. 2011;58:1332-1339.
4. Cullen L, Mueller C, Parsonage WA, et al. Validation of high-sensitivity troponin I in a 2-hour diagnostic strategy to assess 30-day outcomes in emergency department patients with possible acute coronary syndrome. *J Am Coll Cardiol*. 2013;62:1242-1249.
5. Jneid H, Anderson JL, Wright RS, et al. 2012 ACCF/AHA focused update of the guideline for the management of patients with unstable angina/non-ST-elevation myocardial infarction (updating the 2007 guideline and replacing the 2011 focused update): a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2012;60:645-681.
6. Levine GN, Bates ER, Blankenship JC, et al. 2011 ACCF/AHA/SCAI Guideline for Percutaneous Coronary Intervention: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines and the Society for Cardiovascular Angiography and Interventions. *Catheter Cardiovasc Interv*. 2012;79:453-495.
7. Hillis LD, Smith PK, Anderson JL, et al. 2011 ACCF/AHA guideline for coronary artery bypass graft surgery: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Thorac Cardiovasc Surg*. 2012; 143:4-34.
8. Stone NJ, Robinson J, Lichtenstein AH. 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129(suppl 2):S1-S45.
9. Eckel RH, Jakicic JM, Ard JD, et al. 2013 AHA/ACC Guideline on Lifestyle Management to Reduce Cardiovascular Risk: a report of the American College of Cardiology American/Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129(suppl 2):S76-S99.
10. Jensen MD, Ryan DH, Apovian CM, et al. 2013 AHA/ACC/TOS Guideline for the Management of Overweight and Obesity in Adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. *Circulation*. 2014; 129(suppl 2):S102-S138.
11. Go AS, Bauman M, Coleman King SM, et al. An effective approach to high blood pressure control: a science advisory from the American Heart Association, the American College of Cardiology, and the Centers for Disease Control and Prevention. *Hypertension*. 2014;63:878-885.

REVIEW QUESTIONS

1. In comparison to patients with ST segment elevation myocardial infarction (STEMI), the mortality rates for subjects with non–ST segment elevation acute coronary syndrome (NSTE ACS) are

A. Lower in-hospital, similar at 6 months, and higher long term
B. Similar in-hospital, at 6 months, and long term
C. Higher in-hospital, 6 months later, and long term
D. Lower in-hospital, 6 months later, and long term
E. Higher in-hospital, then similar at 6 months as well as long term

Answer: A Large observational cohort studies have demonstrated that mortality rates with NSTE ACS are lower in-hospital, similar at 6 months, and actually higher long term, such that 4 years after the event, the mortality for those with NSTE ACS is twice that of survivors of STEMI. As a result, the chosen management strategy for patients with NSTE ACS is designed to minimize the chance of recurrent ACS or death, both acutely and chronically.

2. The acute coronary syndrome (ACS) patient with the following criteria is considered to be at high risk and, unless it is contraindicated, should be managed with aggressive medical therapy as well as coronary angiography and revascularization if it is anatomically appropriate.

A. Age younger than 65 years, diabetes mellitus, and a history of smoking
B. Three or more Thrombolysis in Myocardial Infarction (TIMI) risk variables, age older than 75 years, and a positive serum troponin concentration
C. ST segment depression, positive serum troponin concentration, and male gender
D. Previous coronary artery bypass grafting or percutaneous coronary intervention, diabetes mellitus, and hypertension
E. Heart rate above 100 beats per minute, T wave inversions in multiple leads, and more than three TIMI risk variables

Answer: B Meta-analyses of all randomized trials of management of subjects with NSTE ACS have demonstrated that those with more than three TIMI risk variables, age older than 75 years, and a positive serum troponin concentration are at high risk of a recurrent ischemic event (myocardial infarction, recurrent ischemia, or death) and obtain a substantial benefit when they are managed invasively (intensive medical therapy and coronary angiography/revascularization within 72 hours of hospitalization).

3. To promote plaque stabilization, all patients with NSTE ACS should receive the following before and after hospital discharge.

A. Rosuvastatin, 10 mg daily
B. Simvastatin, 40 mg daily
C. Atorvastatin, 40 mg daily
D. Pravastatin, 40 mg daily
E. Atorvastatin, 80 mg daily

Answer: E Randomized trials have shown that high-dose atorvastatin (80 mg daily), begun in-hospital and continued indefinitely, effectively reduces the incidence of recurrent ischemic events. Lower doses of the same drug and rosuvastatin have not been assessed in this population of patients. Simvastatin 40 mg is no more efficacious than placebo and pravastatin 40 mg is inferior to high-dose atorvastatin in NSTE ACS patients.

4. The anticoagulant of choice for the subject with heparin-induced thrombocytopenia who is referred for coronary arteriography is

A. Low-molecular-weight heparin
B. Dextran
C. Fondaparinux
D. Bivalirudin
E. High-dose aspirin and clopidogrel

Answer: D In the subject with heparin-induced thrombocytopenia, any form of heparin is contraindicated (hence, low-molecular-weight heparin is an incorrect choice). Intensive high-dose antiplatelet therapy (aspirin and clopidogrel) is of no additional benefit. Dextran is of no benefit at all. Because an increased incidence of catheter-related thrombosis has been reported after fondaparinux treatment, it is not recommended for the patient who is likely to undergo coronary angiography.

73

ST SEGMENT ELEVATION ACUTE MYOCARDIAL INFARCTION AND COMPLICATIONS OF MYOCARDIAL INFARCTION

JEFFREY L. ANDERSON

DEFINITION

Conceptually, myocardial infarction (MI) is myocardial necrosis caused by ischemia. Practically, MI can be diagnosed and evaluated by clinical, electrocardiographic, biochemical, radiologic, and pathologic methods. Technologic advances in detecting much smaller amounts of myocardial necrosis than previously possible (e.g., by high-sensitivity troponin determinations) have required a redefinition of MI. Given these developments, MI now also should be qualified with regard to size, precipitating circumstance, and timing. This chapter focuses on acute MI associated with ST segment elevation on the electrocardiogram (ECG). This category of acute MI is characterized by profound ("transmural") acute myocardial ischemia affecting relatively large areas of myocardium. The underlying cause is essentially *complete* interruption of regional myocardial blood flow (resulting from coronary occlusion, usually atherothrombotic; Chapter 70). This clinical syndrome should be distinguished from non–ST segment elevation MI, in which the blockage of coronary flow is incomplete and for which different acute therapies are appropriate (Chapter 72).

EPIDEMIOLOGY

The risk of cardiovascular disease and MI has declined in recent years in the United States and the Western world, but the burden of disease remains high. Cardiovascular disease is responsible for almost half of all noncommunicable deaths worldwide (Chapter 52), and coronary heart disease causes about 1 of every 6 deaths in the United States or about 380,000 deaths per year. Each

TABLE 73-1 CONDITIONS OTHER THAN CORONARY ATHEROSCLEROSIS THAT CAN CAUSE ACUTE MYOCARDIAL INFARCTION

Coronary emboli	Causes include aortic or mitral valve lesions, left atrial or ventricular thrombi, prosthetic valves, fat emboli, intracardiac neoplasms, infective endocarditis, and paradoxical emboli
Thrombotic coronary artery disease	Can occur with oral contraceptive use, sickle cell anemia and other hemoglobinopathies, polycythemia vera, thrombocytosis, thrombotic thrombocytopenic purpura, disseminated intravascular coagulation, antithrombin III deficiency and other hypercoagulable states, macroglobulinemia and other hyperviscosity states, multiple myeloma, leukemia, malaria, and fibrinolytic system shutdown secondary to impaired plasminogen activation or excessive inhibition
Coronary vasculitis	Seen with Takayasu disease, Kawasaki disease, polyarteritis nodosa, lupus erythematosus, scleroderma, rheumatoid arthritis, and immune-mediated vascular degeneration in cardiac allografts
Coronary vasospasm	Can be associated with variant angina, nitrate withdrawal, cocaine or amphetamine abuse, and angina with "normal" coronary arteries
Infiltrative and degenerative coronary vascular disease	Can result from amyloidosis, connective tissue disorders (e.g., pseudoxanthoma elasticum), lipid storage disorders and mucopolysaccharidoses, homocystinuria, diabetes mellitus, collagen vascular disease, muscular dystrophies, and Friedreich ataxia
Spontaneous coronary dissection	Eighty percent of cases occur in women; the most common causes are extreme exertion in men and postpartum status in women; the presentation is ST elevation myocardial infarction in 50% of cases, and the estimated 10-year rate of subsequent major adverse cardiac events is about 50%
Coronary ostial occlusion	Associated with aortic dissection, luetic aortitis, aortic stenosis, and ankylosing spondylitis syndromes
Congenital coronary anomalies	Including Bland-White-Garland syndrome of anomalous origin of the left coronary artery from the pulmonary artery, left coronary artery origin from the anterior sinus of Valsalva, coronary arteriovenous fistula or aneurysms, and myocardial bridging with secondary vascular degeneration
Trauma	Associated with and responsible for coronary dissection, laceration, or thrombosis (with endothelial damage secondary to trauma such as angioplasty) and with radiation and cardiac contusion
Augmented myocardial oxygen requirements exceeding oxygen delivery	Encountered with aortic stenosis, aortic insufficiency, hypertension with severe left ventricular hypertrophy, pheochromocytoma, thyrotoxicosis, methemoglobinemia, carbon monoxide poisoning, shock, and hyperviscosity syndromes

year, about 635,000 Americans are diagnosed with a first MI, about 230,000 have a recurrent MI, and about 150,000 more have a silent first MI.[1] An American suffers a coronary event approximately every 30 seconds, and one dies from one every minute.

More than 5 million people visit emergency departments in the United States each year for evaluation of chest pain and related symptoms (Chapter 51), about 680,000 of whom are diagnosed with an acute coronary syndrome (non–ST segment elevation MI/unstable angina; Chapter 72). The presence of ST segment elevation or new left bundle branch block (LBBB) on the ECG distinguishes patients with acute MI who require consideration of immediate reperfusion (recanalization) therapy from other patients with an acute coronary syndrome. Changing demographics, lifestyles, and medical therapies have led to a decrease in the ratio of ST segment elevation MI (STEMI) to non–ST segment elevation acute coronary syndromes during the past 15 to 20 years, so STEMI now accounts for about 30% of all MIs.

PATHOBIOLOGY

Erosion, fissuring, or rupture of vulnerable atherosclerotic plaques has been determined to be the initiating mechanism of coronary thrombotic occlusion, thereby precipitating intraplaque hemorrhage, coronary spasm, and occlusive luminal thrombosis (Chapter 70). Plaque rupture most frequently occurs in lipid-laden plaques with an endothelial cap weakened by internal collagenase (metalloproteinase) activity derived primarily from macrophages. These macrophages are recruited to the plaque from blood monocytes responding to inflammatory mediators and adhesion molecules.

With plaque rupture, elements of the blood stream are exposed to the highly thrombogenic plaque core and matrix containing lipid, tissue factor, and collagen. Platelets adhere, become activated, and aggregate; vasoconstrictive and thrombogenic mediators are secreted; vasospasm occurs; thrombin is generated and fibrin formed; and a partially or totally occlusive platelet- and fibrin-rich thrombus is generated. When coronary flow is occluded, electrocardiographic ST segment elevation occurs, resulting in STEMI. Partial occlusion, occlusion in the presence of collateral circulation, and distal coronary embolization result in unstable angina or non–ST segment elevation MI (Chapter 72). Ischemia from impaired myocardial perfusion causes myocardial cell injury or death, ventricular dysfunction, and cardiac arrhythmias.

Although most MIs are caused by atherosclerosis, occasional patients can develop complete coronary occlusions due to coronary dissections,[2] emboli, in situ thrombosis, vasculitis, primary vasospasm, infiltrative or degenerative diseases, diseases of the aorta, congenital anomalies of a coronary artery, or trauma (Table 73-1). In a canine model of coronary occlusion and reperfusion, myocardial cell death begins within 15 minutes of occlusion and proceeds rapidly in a wave front from endocardium to epicardium. Partial myocardial salvage can be achieved by releasing the occlusion within 3 to 6 hours; the degree of salvage is inversely proportional to the duration of ischemia and occurs in a reverse wave front from epicardium to endocardium. The extent of myocardial necrosis can also be altered by modification of metabolic demands and collateral blood supply. The temporal dynamic of infarction in human disease, although more complex, is generally similar.

Susceptibility to coronary artery disease and subsequent MI is estimated to be 40% genetic, with the balance being environmental. A large international collaborative meta-analysis identified or validated 23 common genetic susceptibility loci for coronary artery disease, many unrelated to traditional risk factors, but these loci account for only 10% of genetic variance. Interestingly, among patients with coronary artery disease, only a variant in the glycosyltransferase gene associated with the ABO blood group O phenotype has been shown to protect against MI.

CLINICAL MANIFESTATIONS

The diagnosis of acute MI has traditionally rested on the triad of ischemic-type chest discomfort, electrocardiographic abnormalities, and elevated serum cardiac biomarkers of necrosis. Acute MI was considered present when at least two of the three were present. With their increasing sensitivity and specificity, serum cardiac biomarkers (i.e., troponin I [TnI] and troponin T [TnT]) have assumed a dominant role in confirming the diagnosis of acute MI in patients with suggestive clinical or electrocardiographic features.

History

Ischemic-type chest discomfort is the most prominent clinical symptom in most patients with acute MI (see Table 51-1). The discomfort is characterized by its quality, location, duration, radiation, and precipitating and relieving factors. The discomfort associated with acute MI is qualitatively similar to that of angina pectoris but more severe. It often is perceived as heavy, pressing, crushing, squeezing, bandlike, viselike, strangling, constricting, aching, or burning; it rarely is perceived as sharp pain and generally not as stabbing pain (Chapters 51 and 71).

The primary location of typical ischemic pain is most consistently retrosternal, but it also can present left parasternally, left precordially, or across the anterior chest (Chapter 51). On occasion, discomfort is predominantly perceived in the anterior neck, jaw, arms, or epigastrium. It generally is somewhat diffuse; highly localized pain (finger point) is rarely angina or acute MI. The most characteristic pattern of radiation is to the left arm, but the right arm or both arms can be involved. The shoulders, neck, jaw, teeth, epigastrium, and interscapular areas also are sites of radiation. Discomfort above the jaws or

below the umbilicus is not typical of acute MI. Associated symptoms often include nausea, vomiting, diaphoresis, weakness, dyspnea, restlessness, and apprehension.

The discomfort of acute MI is more severe and lasts longer (typically 20 minutes to several hours) than angina, and it is not reliably relieved by rest or nitroglycerin. The onset of acute MI usually is unrelated to exercise or other apparent precipitating factors. Nevertheless, acute MI begins during physical or emotional stress and within a few hours of arising more frequently than is explained by chance.

It is estimated that at least 20% of acute MIs are painless ("silent") or atypical (unrecognized). Elderly patients, especially women, and patients with diabetes are particularly prone to painless or atypical MI, which is the presentation of MI in as many as one third to one half of such patients. Because the prognosis is worse in elderly patients and in those patients with diabetes, diagnostic vigilance is required. In these patients, acute MI can be manifested as sudden dyspnea (which can progress to pulmonary edema), weakness, lightheadedness, nausea, and vomiting. Confusional states, sudden loss of consciousness, a new rhythm disorder, and an unexplained fall in blood pressure are other uncommon presentations. The differential diagnosis of ischemic chest discomfort also should include gastrointestinal disorders (e.g., reflux esophagitis; Chapter 138), musculoskeletal pain (e.g., costochondritis), anxiety or panic attacks, pleurisy or pulmonary embolism (Chapter 98), and acute aortic dissection (see Table 51-2 and Chapter 78).

Physical Examination

No physical findings are diagnostic or pathognomonic of acute MI. The physical examination findings can be entirely normal or may reveal only nonspecific abnormalities. An S_4 gallop frequently is found if it is carefully sought. Blood pressure often is initially elevated, but it may be normal or low. Signs of sympathetic hyperactivity (tachycardia, hypertension, or both) often accompany anterior wall MI, whereas parasympathetic hyperactivity (bradycardia, hypotension, or both) is more common with inferior wall MI.

The examination is best focused on an overall assessment of cardiac function. Adequacy of vital signs and peripheral perfusion should be noted. Signs of cardiac failure, both left and right sided (e.g., S_3 gallop, pulmonary congestion, elevated neck veins) should be sought, and observation for arrhythmias and mechanical complications (e.g., new murmurs) is essential. If hypoperfusion is present, determination of its primary cause (e.g., hypovolemia, right-sided heart failure, left-sided heart failure) is critical to management.

DIAGNOSIS

Electrocardiogram

In patients with a possible acute MI, an ECG must be obtained immediately. Although the initial ECG is neither perfectly specific nor perfectly sensitive in all patients who develop acute STEMI, it plays a critical role in initial stratification, triage, and management (Chapter 51). In an appropriate clinical setting, a pattern of ST segment elevation of 2 mm (0.2 mV) or more at the J point in V_2 to V_3 in men or 1.5 mm (0.15 mV) or more in women in the absence of left ventricular (LV) hypertrophy or 1 mm (0.1 mV) or more in two or more other contiguous chest or limb leads suggests coronary occlusion causing marked myocardial ischemia.[1] In such patients, emergency reperfusion (primary angioplasty or fibrinolysis) should be performed unless it is contraindicated. Hyperacute T wave changes may suggest the diagnosis in the early phase of STEMI before the onset of ST elevation. A new or presumably new LBBB, which may obscure ST elevation analysis, may suggest a STEMI equivalent in the appropriate clinical setting. ST depression in two or more precordial leads V_1 to V_4 may indicate transmural posterior injury due to occlusion of the left circumflex coronary artery; extending ECG analysis to leads V_7 to V_9 can help confirm this diagnosis. Other ECG patterns (ST segment depression, T wave inversion, nonspecific changes, normal ECG) in association with ischemic chest discomfort are consistent with a non–ST segment elevation acute coronary syndrome and are treated with different triage and initial management strategies (Chapter 72).

Electrocardiographic Evolution

Serial ECG tracings improve the sensitivity and specificity of the ECG for the diagnosis of acute MI and assist in assessing the outcomes of therapy. When typical ST segment elevation persists for hours and is followed within hours to days by T wave inversions and Q waves, the diagnosis of acute MI can be made with virtual certainty. The ECG changes in acute STEMI evolve through three overlapping phases: hyperacute or early acute, evolved acute, and chronic (stabilized).

Early Acute Phase

This earliest phase begins within minutes, persists, and evolves during hours. T waves increase in amplitude and widen over the area of injury (hyperacute pattern). ST segments evolve from concave to a straightened to a convex upward pattern (acute pattern). When prominent, the acute injury pattern of blended ST-T waves can take on a tombstone appearance (Figs. 73-1 and 73-2). ST segment depressions that occur in leads opposite those with ST segment elevation are known as reciprocal changes and are associated with larger areas of injury and a worse prognosis but also with greater benefits from reperfusion therapy.

Other causes of ST segment elevation must be considered and excluded. These conditions include pericarditis (Chapter 77), LV hypertrophy with J point elevation, and normal variant early repolarization (Chapter 54). Pericarditis (or perimyocarditis) is of particular concern because it can mimic acute MI clinically, but fibrinolytic therapy is *not* indicated and can be hazardous.

Evolved Acute Phase

During the second phase, ST segment elevation begins to regress, T waves in leads with ST segment elevation become inverted, and pathologic Q or QS waves become fully developed (>0.03-second duration or depth >30% of R wave amplitude, or both).

Chronic Phase

Resolution of ST segment elevation is variable. Resolution is usually complete within 2 weeks of inferior MI, but it can be delayed further after anterior MI. Persistent ST segment elevation, often seen with a large anterior MI, is indicative of a large area of akinesis, dyskinesis, or ventricular aneurysm. Symmetrical T wave inversions can resolve during weeks to months or can persist for an indefinite period; hence, the age of an MI in the presence of T wave inversions is often termed indeterminate. Q waves usually do not resolve after anterior MI but often disappear after inferior wall MI.

Early reperfusion therapy accelerates the time course of ECG changes to minutes or hours instead of days to weeks. ST segments recede rapidly, T wave inversions and loss of R waves occur earlier, and Q waves may not develop or progress and occasionally may regress. Indeed, failure of ST segment elevation to resolve by more than 50 to 70% within 1 to 2 hours

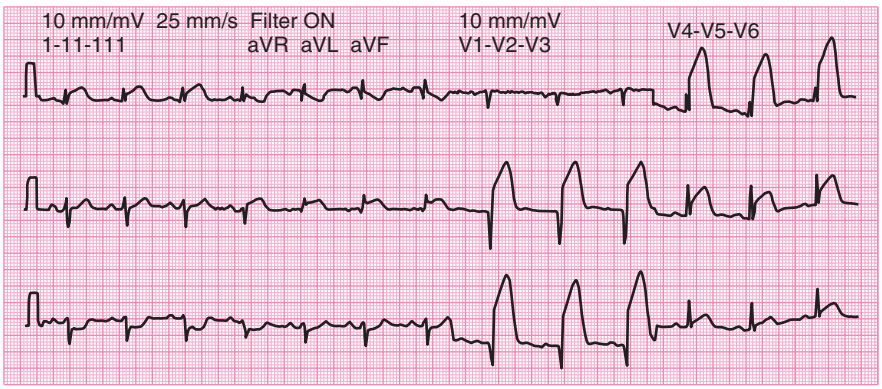

FIGURE 73-1. Electrocardiographic tracing shows an acute anterolateral myocardial infarction. Note ST segment elevation in leads I, aVL, and V_1 to V_6 with Q waves in V_1 to V_4.

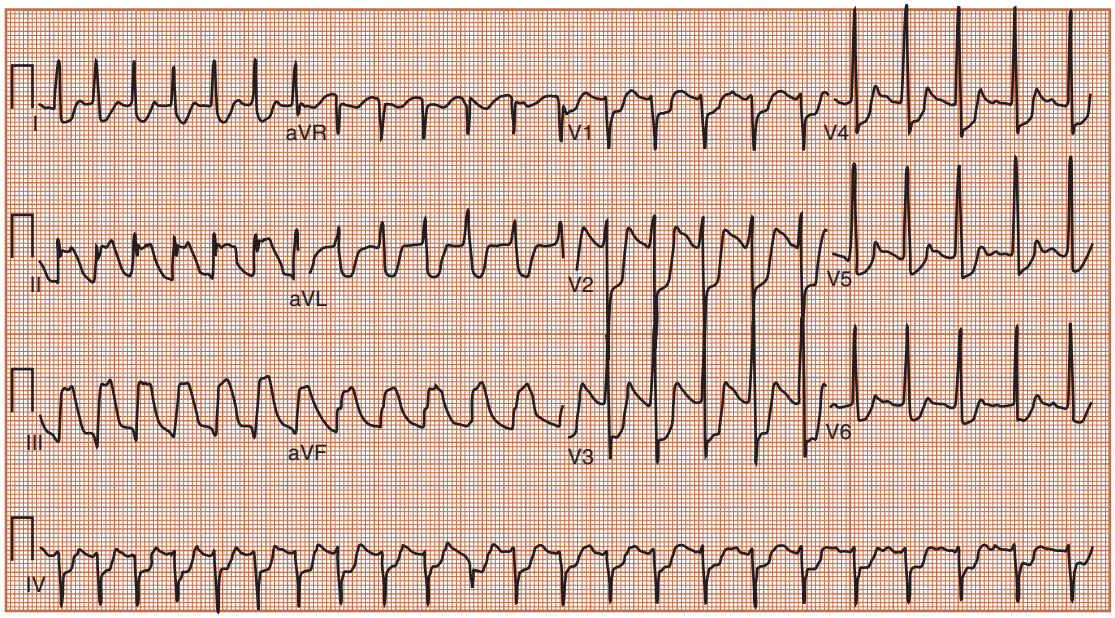

FIGURE 73-2. Electrocardiographic tracing shows an acute inferoposterior myocardial infarction.

suggests failure of fibrinolysis and should prompt referral for urgent angiography and consideration of "rescue angioplasty."

True Posterior Myocardial Infarction and Left Circumflex Myocardial Infarction Patterns

"True posterior" MI presents a mirror-image pattern of ECG injury in leads V_1 to V_2 to V_4 (Fig. 73-2). The location of injury of true posterior MI by magnetic resonance imaging actually involves portions of the *lateral* LV wall and is typically caused by occlusion of a nondominant left circumflex artery. In the precordial leads, the acute phase is characterized by ST segment depression rather than by ST segment elevation. The evolved and chronic phases show increased R wave amplitude and widening instead of Q waves. Recognition of a true posterior acute MI pattern is challenging but important because the diagnosis should lead to an immediate reperfusion strategy. Extending the ECG to measure left posterior leads V_7 to V_9 increases sensitivity for detection of acute left circumflex–related injury patterns (i.e., ST segment elevation) with excellent specificity (Chapter 54). Other causes of prominent upright anteroseptal forces include right ventricular (RV) hypertrophy, ventricular preexcitation variants (Wolff-Parkinson-White syndrome; Chapter 64), and normal variants with early R wave progression. New appearance of these changes or the association with an acute or evolving inferior MI usually allows the diagnosis to be made.

Right Ventricular Infarction

Proximal occlusion of the right coronary artery before the acute marginal branch can cause RV infarction as well as acute inferior MI in about 30% of cases. Because the prognosis and treatment of acute inferior MI differ in the presence of RV infarction, it is important to make this diagnosis. The diagnosis is assisted by obtaining right precordial ECG leads, which are routinely indicated for inferior acute MI (Chapter 54). Acute ST segment elevation of at least 1 mm (0.1 mV) in one or more leads V_4R to V_6R is both sensitive and specific (>90%) for identifying acute RV injury, and Q or QS waves effectively identify RV infarction.

Diagnosis in the Presence of Bundle Branch Block

The presence of LBBB often obscures ST segment analysis in patients with suspected acute MI. The presence of a new (or presumed new) LBBB in association with clinical (and laboratory) findings suggesting acute MI is associated with high mortality; patients with new-onset LBBB benefit substantially from reperfusion therapy and should undergo triage and treatment in the same way as patients with STEMI do. Certain ECG patterns, although relatively insensitive, suggest acute MI if they are present in the setting of LBBB: Q waves in two of leads I, aVL, V_5, and V_6; R wave regression from V_1 to V_4; ST segment elevation of 1 mm or more in leads with a positive QRS complex; ST segment depression of 1 mm or more in leads V_1, V_2, or V_3; and ST segment elevation of 5 mm or more associated with a negative QRS complex. The presence of right bundle branch block (RBBB) usually does not mask typical ST-T wave or Q wave changes except in rare cases of isolated true posterior acute MI, which are characterized by tall right precordial R waves and ST segment depressions.

TABLE 73-2 CONDITIONS THAT CAN MIMIC ST SEGMENT ELEVATION MYOCARDIAL INFARCTION

Early repolarization with noncoronary chest pain
Myocarditis
Pericarditis
Takotsubo ("stress") cardiomyopathy

Differential Diagnosis

Although STEMI is often an easy diagnosis to make on the basis of the presentation and test results (see later), other considerations include acute pericarditis (Chapter 77), acute myocarditis (Chapter 60), stress-induced (takotsubo) syndrome (Chapter 60), and early repolarization (Table 73-2). All but early repolarization can be associated with abnormal biomarkers, but none are associated with a coronary occlusion. Early coronary angiography is advised when the cause of ST segment elevation is unclear (see Table 73-1).

Serum Cardiac Biomarkers of Necrosis

Cardiac-derived TnI (cTnI) and TnT (cTnT), which are proteins of the sarcomere, are not normally present in the blood with standard sensitivity assays and have amino acid sequences distinct from their skeletal muscle isoforms. The troponins generally are first detectable 1 to 4 hours after the onset of acute MI,[3] are maximally sensitive at 8 to 12 hours, peak at 10 to 24 hours, and persist for 5 to 14 days. Their long persistence has allowed them to replace other markers for the diagnosis of acute MI in patients presenting late (>1 to 2 days) after symptoms. However, this persistence can obscure the diagnosis of an early recurrent MI, for which more rapidly cleared markers (i.e., the MB isoenzyme of creatine kinase [CK-MB]) may be selectively useful.

The sensitivity and specificity of cardiac-specific TnI and TnT make them the "gold standard" for detection of myocardial necrosis (see later). However, because of the 1- to 12-hour delay after the onset of symptoms before markers become detectable or diagnostic across the spectrum of acute coronary syndromes, the decision to proceed with urgent reperfusion (primary angioplasty or fibrinolysis) in STEMI must be based on the patient's clinical history and initial ECG (Chapter 51).

Clinically, cTnI and cTnT appear to be of approximately equivalent utility, except that renal failure is more likely to be associated with false-positive elevations of cTnT than of cTnI. Because troponins also may be present in low concentration in a number of nonischemic cardiovascular conditions, the

TABLE 73-3 NON–MYOCARDIAL INFARCTION CAUSES OF AN ELEVATED TROPONIN LEVEL

OTHER CARDIAC CAUSES
Myocardial injury: cardiac contusion, surgery, ablation, shocks
Myocardial inflammation: myocarditis, pericarditis
Heart failure
Cardiomyopathies: infiltrative, stress, hypertensive, hypertrophic
Aortic dissection
Severe aortic stenosis
Tachycardias
PULMONARY CAUSES
Pulmonary embolism
Pulmonary hypertension
Respiratory failure
NEUROLOGIC CAUSES
Stroke
Intracranial hemorrhage
OTHER
Shock: septic, hypovolemic, cardiogenic
Renal failure

Modified from Thygesen K, Alpert JS, Jaffe AS, et al. Third universal definition of myocardial infarction. *Circulation.* 2012;126:2020-2035.

clinician also must consider the clinical context and the temporal rise and fall of troponin levels (Table 73-3).

Other Laboratory Tests

On admission, routine assessment of complete blood count and platelet count, standard blood chemistry studies, a lipid panel, and coagulation tests (prothrombin time, partial thromboplastin time) is useful. Results assist in assessing comorbid conditions and prognosis and in guiding therapy. Hematologic tests provide a useful baseline before initiation of antiplatelet, anticoagulant, and fibrinolytic therapy or coronary angiography or angioplasty. Myocardial injury precipitates polymorphonuclear leukocytosis, commonly resulting in an elevation of white blood cell count of up to 12,000 to 15,000/μL, which appears within a few hours and peaks at 2 to 4 days. The metabolic panel provides a useful check on electrolytes, glucose, and renal function. On hospital admission or the next morning, a fasting lipid panel is recommended as a baseline for lipid-lowering (statin) therapy (Chapter 206). Unless carbon dioxide retention is suspected, finger oximetry is adequate to titrate oxygen therapy. The C-reactive protein level increases with acute MI, but its incremental prognostic value in the acute setting has not been established. B-type natriuretic peptide, which increases with ventricular wall stress and relative circulatory fluid overload, may provide useful incremental prognostic information in the setting of acute MI.

Imaging

A chest radiograph is the only imaging test *routinely* obtained on admission for acute MI. Although the chest radiograph is often normal, findings of pulmonary venous congestion, cardiomegaly, or widened mediastinum can contribute importantly to diagnosis and management decisions. For example, a history of severe, "tearing" chest and back pain in association with a widened mediastinum should raise the question of a dissecting aortic aneurysm (Chapter 78). In such cases, fibrinolytic therapy must be withheld pending more definitive diagnostic imaging of the aorta. Other noninvasive imaging (e.g., echocardiography [Chapter 55], cardiac nuclear scanning [Chapter 56], and other testing) is performed for evaluation of specific clinical issues, including suspected complications of acute MI. Coronary angiography (Chapter 57) is performed urgently as part of an interventional strategy for acute MI or later for risk stratification in higher-risk patients who are managed medically.

Echocardiography

Two-dimensional transthoracic echocardiography with color flow Doppler imaging is the most generally useful noninvasive test obtained on admission or early in the hospital course (Chapter 55). Echocardiography efficiently assesses global and regional cardiac function and enables the clinician to evaluate suspected complications of acute MI. The sensitivity and specificity of echocardiography for regional wall motion assessment are high (>90%), although the age of the abnormality (new vs. old) must be distinguished clinically or by electrocardiography. Echocardiography is helpful in determining the cause of circulatory failure with hypotension (relative hypovolemia, LV failure, RV failure, or mechanical complication of acute MI). Echocardiography also can assist in differentiating pericarditis and perimyocarditis from acute MI. Doppler echocardiography is indicated to evaluate a new murmur and other suspected mechanical complications of acute MI (papillary muscle dysfunction or rupture, acute ventricular septal defect, LV free wall rupture with tamponade or pseudoaneurysm). Later in the course of acute MI, echocardiography may be used to assess the degree of recovery of stunned myocardium after reperfusion therapy, the degree of residual cardiac dysfunction and indications for angiotensin-converting enzyme (ACE) inhibitors and other therapies for heart failure, and the presence of LV aneurysm and mural thrombus (requiring oral anticoagulants).

Radionuclide, Magnetic Resonance, and Other Imaging Studies

Radionuclide techniques generally are too time-consuming and cumbersome for routine use in the acute setting of definite or probable acute MI. More commonly, they are used in risk stratification before or after hospital discharge to augment exercise or pharmacologic stress testing (Chapter 56). Thallium Tl 201 or technetium Tc 99m sestamibi nuclear scans or rubidium Rb 82 positron emission tomography scans can assess myocardial perfusion and viability as well as infarct size. Cardiac magnetic resonance imaging (Chapter 56) with late gadolinium enhancement also can assess infarct size as well as myocardial function during the convalescent phase. Computed tomography and magnetic resonance imaging also can be useful to evaluate patients with a suspected dissecting aortic aneurysm (Chapter 78). When a nonatherosclerotic cause of myocardial necrosis is raised (e.g., perimyocarditis simulating acute MI), contemporary multislice (e.g., 64-slice) coronary computed tomography can assess coronary artery disease qualitatively and semiquantitatively as well as distinguish other causes of chest pain syndromes (Chapters 51 and 56).

TREATMENT

Assessment and Management

Prehospital Phase

More than half of deaths related to acute MI occur within 1 hour of onset of symptoms and before the patient reaches a hospital emergency department. Most of these deaths are caused by ischemia-related ventricular fibrillation (VF) and can be reversed by defibrillation (Chapters 63 and 66). Rapid defibrillation allows resuscitation in 60% of patients when treatment is delivered by a bystander using an on-site automatic external defibrillator or by a first-responding medical rescuer (Chapter 63). Moreover, the first hour represents the best opportunity for myocardial salvage with reperfusion therapy. Thus, the three goals of prehospital care are to recognize symptoms promptly and seek medical attention; to deploy an emergency medical system team capable of cardiac monitoring, defibrillation and resuscitation, and emergency medical therapy; and to transport the patient expeditiously to a medical care facility staffed with personnel capable of providing expert coronary care, including reperfusion therapy (primary angioplasty or fibrinolysis).

The greatest time lag to reperfusion therapy is the patient's delay in calling for help. Public education efforts have yielded mixed results, and innovative approaches are needed. Emergency medical personnel should perform a 12-lead ECG at the site of first medical contact in patients with symptoms consistent with STEMI.[1,4] In coordinated systems and when transportation delays are substantial (i.e., >90 to 120 minutes), fibrinolytic and other antithrombotic therapy ideally is administered in the field, thereby shortening the time to reperfusion.

Hospital Phases

Emergency Department

The goals of emergency department care are to identify patients with acute myocardial ischemia rapidly, to stratify them into acute STEMI compared with other acute coronary syndromes (see Fig. 72-1 and Fig. 73-1), to initiate a reperfusion strategy and other appropriate medical care in qualifying patients with acute STEMI, and to prioritize by triage rapidly to inpatient care (cardiac intensive care unit, step-down unit, observation unit) or outpatient care (patients without suspected ischemia) (see Fig. 72-2).

The evaluation of patients with chest pain and other suspected acute coronary syndromes begins with a 12-lead ECG even as the physician is beginning a focused history, including contraindications to fibrinolysis and angiography, and a targeted physical examination. Continuous ECG monitoring should be started, an intravenous line should be established, and admission blood tests should be performed (including cardiac biomarkers such as cTnI or cTnT). As

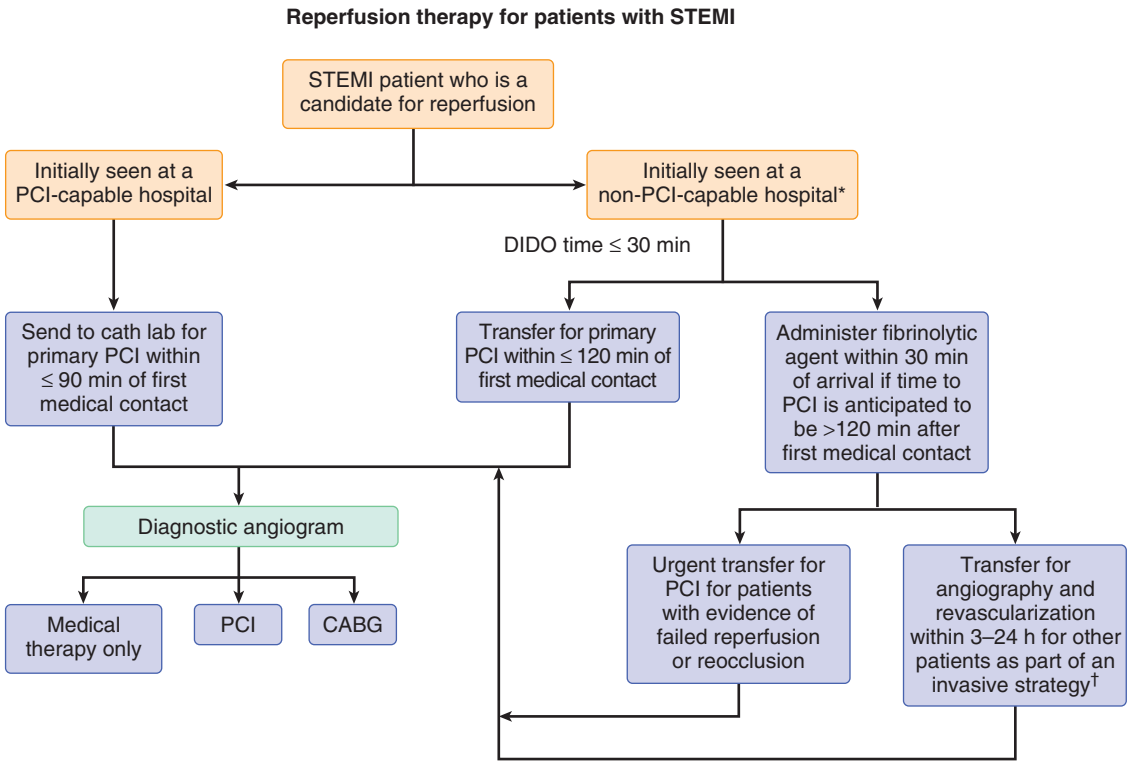

FIGURE 73-3. **Reperfusion therapy for patients with ST segment elevation myocardial infarction (STEMI). *Patients with cardiogenic shock or severe heart failure initially seen at a non–PCI-capable hospital should be transferred for cardiac catheterization and revascularization as soon as possible, irrespective of time delay from MI onset (Class I, Level of Evidence: B). †Angiography and revascularization should not be performed within the first 2 to 3 hours after administration of fibrinolytic therapy. CABG = coronary artery bypass grafting; DIDO = door in–door out; PCI = percutaneous coronary intervention. (Modified from O'Gara PT, Kushner FG, Ascheim DD, et al. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation.* 2013;127:e362-e425.)**

rapidly as possible, the patient should be stratified as having a probable acute STEMI, a non–ST segment elevation acute MI, probable or possible unstable angina, or likely noncardiac chest pain.

In patients with acute STEMI by clinical and electrocardiographic criteria, a reperfusion strategy must be selected. Alternative choices are primary percutaneous coronary intervention (primary PCI; the patient is transferred directly to the cardiac catheterization laboratory with a systems goal of first medical contact to device time of less than 90 minutes) and fibrinolysis (begun immediately in the emergency department with a goal of door to needle time of less than 30 minutes)[1] (Fig. 73-3). In patients who present to a non–PCI-capable hospital, the goal is a first medical contact to device time of 120 minutes or less; otherwise, fibrinolytic therapy is indicated.

Specific Therapeutic Measures

Reperfusion Therapy

Early reperfusion of ischemic, infarcting myocardium represents the most important conceptual and practical advance for acute STEMI and is the primary therapeutic goal. Coronary reperfusion is accomplished by primary PCI with angioplasty and stenting or by fibrinolytic (thrombolytic) therapy. During the past decade, the application of reperfusion therapy has remained relatively constant in the United States and other Western countries at 70 to 75% of eligible patients with acute MI. The percentage of patients undergoing primary PCI has increased substantially over time, although fibrinolytic therapy continues to be commonly applied in developing countries.[5]

With broad application of reperfusion therapy, 30-day mortality rates from acute STEMI have progressively declined during the past three decades (from 20 to 30% to 5 to 10%). Each community should develop and follow an optimal STEMI system of care within the resources available (Fig. 73-3).

Primary Percutaneous Coronary Intervention

Prompt PCI by the femoral or, increasingly, the radial artery approach is the reperfusion strategy of choice in patients with STEMI and ischemic symptoms of less than 12 hours in duration at PCI-capable hospitals (Table 73-4). The relative benefits of primary PCI over fibrinolysis include a significantly lower acute mortality rate, lower rates of nonfatal reinfarction, and lower risks of intracerebral hemorrhage. PCI generally includes a bare metal or drug-eluting (e.g., sirolimus, paclitaxel, everolimus, zotarolimus) stent (Chapter 74).

Currently, a primary PCI strategy begins with initiation of a $P2Y_{12}$ inhibitor in the emergency department, together with aspirin and an anticoagulant (e.g., heparin or bivalirudin), followed by rapid PCI with stenting. Augmented antiplatelet therapy with a glycoprotein IIb/IIIa (GPIIb-IIIa) inhibitor may be given in selected patients, generally at the time of catheterization. The addition of a reduced dose of a plasminogen activator to GPIIb-IIIa therapy in the field or emergency department may further improve outcomes in selected patients who undergo early PCI, but this approach is generally not recommended. Pharmacologically facilitated PCI, whereby patients at hospitals without PCI capabilities are given adjusted doses of fibrinolytic or GPIIb-IIIa inhibitors, or both, and then are transferred to other hospitals for emergent (i.e., within 1 to 2 hours) PCI, overall appears to be no better than rapid transfer for primary PCI within 1 to 2 hours.

Primary PCI also is recommended in patients with STEMI and cardiogenic shock or acute, severe heart failure emergently and irrespective of time delay

TABLE 73-4 INDICATIONS FOR PRIMARY ANGIOPLASTY AND COMPARISON WITH FIBRINOLYTIC THERAPY

INDICATIONS
A preferred reperfusion strategy for ST segment elevation or LBBB acute MI within 12 hours of symptom onset (or >12 hours if symptoms persist)
Cardiogenic shock developing within 36 hours of ST segment elevation/Q wave acute MI or LBBB acute MI in patients <75 years old who can be revascularized within 18 hours of shock onset
Recommended only at centers performing >200 PCIs/year with backup cardiac surgery and for operators performing >75 PCIs/year
ADVANTAGES OF PRIMARY PCI
Higher initial reperfusion rates
Reduced risk of intracerebral hemorrhage
Less residual stenosis; less recurrent ischemia or infarction
Usefulness when fibrinolysis is contraindicated
Improvement in outcomes with cardiogenic shock
DISADVANTAGES OF PRIMARY PCI (COMPARED WITH FIBRINOLYTIC THERAPY)
Access, advantages restricted to high-volume centers and operators
Longer average time to treatment
Greater dependence on operators for results
Higher system complexity and costs

LBBB = left bundle branch block; MI = myocardial infarction; PCI = percutaneous coronary intervention (includes balloon angioplasty, stenting).

TABLE 73-5 CHARACTERISTICS OF INTRAVENOUS FIBRINOLYTIC AGENTS APPROVED BY THE U.S. FOOD AND DRUG ADMINISTRATION

	STREPTOKINASE (SK)	ALTEPLASE (t-PA)	RETEPLASE (r-PA)	TENECTEPLASE (TNK–t-PA)
Dose	1.5 MU in 30-60 minutes	100 mg in 90 minutes*	10 U + 10 U, 30 minutes apart	30-50 mg† during 5 seconds
Circulating half-life (minutes)	≅20	≅4	≅18	≅20
Antigenic	Yes	No	No	No
Allergic reactions	Yes	No	No	No
Systemic fibrinogen depletion	Severe	Mild to moderate	Moderate	Minimal
Intracerebral hemorrhage	≅0.4%	≅0.7%	≅0.8%	≅0.7%
Patency (TIMI 2/3) rate, 90 minutes‡	≅51%	≅73-84%	≅83%	≅77-88%
Lives saved per 100 treated	≅3§	≅4‖	≅4	≅4
Cost per dose (approximate U.S. dollars)	300	1800	2200	2200

*Accelerated t-PA given as follows: 15-mg bolus, then 0.75 mg/kg during 30 minutes (maximum, 50 mg), then 0.50 mg/kg during 60 minutes (maximum, 35 mg).
†TNK–t-PA is dosed by weight (supplied in 5-mg/mL vials): <60 kg = 6 mL; 61-70 kg = 7 mL; 71-80 kg = 8 mL; 81-90 kg = 9 mL; >90 kg = 10 mL.
‡TIMI = Thrombolysis in Myocardial Infarction. Data from Granger CB, Califf RM, Topol EJ. Thrombolytic therapy for acute myocardial infarction: a review. *Drugs.* 1992;44:293-325; and Bode C, Smalling RW, Berg G, et al. Randomized comparison of coronary thrombolysis achieved with double-bolus reteplase (recombinant plasminogen activator) and front-loaded, accelerated alteplase (recombinant tissue plasminogen activator) in patients with acute myocardial infarction: the RAPID II Investigators. *Circulation.* 1996;94:891-898.
§Patients with ST segment elevation or bundle branch block, treated <6 hours.
‖Based on the finding from the GUSTO trial that t-PA saves one more additional life per 100 treated than does SK. Data from The GUSTO Investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *N Engl J Med.* 1993;329:673-682; and Simes RJ, Topol EJ, Holmes DR Jr, et al. Link between the angiographic substudy and mortality outcomes in a large randomized trial of myocardial reperfusion: importance of early and complete infarct artery reperfusion. GUSTO-I Investigators. *Circulation.* 1995;91:1923-1928.

from onset of MI symptoms (Chapter 107). Primary PCI also is reasonable in patients with STEMI who have clinical or electrocardiographic evidence of ongoing ischemia between 12 and 24 hours after the onset of symptoms. However, delayed PCI of a totally occluded infarct-related artery after 24 hours is not recommended in asymptomatic, stable patients with one- or two-vessel disease. PCI performed in a stenotic but noninfarcted artery at the time of primary PCI in STEMI patients who otherwise are hemodynamically stable has recently been shown to be superior to medical therapy alone; staged PCI of such arteries (i.e., at a somewhat delayed time after primary PCI) is a common current practice whose safety is supported by registry data. Increasing positive experience with PCI of the left main coronary artery with stents, especially drug-eluting stents, suggests that it may be an alternative to coronary artery bypass grafting (CABG) in STEMI patients with a culprit left main coronary artery with compromised flow when PCI can be performed more rapidly and safely than CABG. Routine application of mechanical thrombus aspiration at the time of angiography probably does not provide additional benefit over PCI alone, but selected use may be beneficial in patients with a large burden of thrombus.

Fibrinolytic Therapy

Various fibrinolytic agents (Table 73-5) are useful in patients with STEMI or new or presumed new LBBB who present for treatment within 12 hours of the onset of symptoms and who have no contraindications to their use (Table 73-6). Benefit declines from about 40 lives or more saved per 1000 within the first hour, to 20 to 30 lives saved per 1000 for hours 2 to 12, to a nonsignificant 7 lives saved per 1000 for hours 13 to 24. An accelerated regimen of tissue plasminogen activator (t-PA plus intravenous heparin) is preferred to streptokinase because the patency rate of the infarct-related artery at 90 minutes is higher and mortality is lower. Longer-acting variants of t-PA, given by single-bolus (tenecteplase) or double-bolus (reteplase) injections, are now in widespread clinical use because they are more convenient to give, but they have not led to further improvements in survival. A nonimmunogenic fibrinolytic agent is preferred for patients with a history of prior streptokinase use.

The major risk of fibrinolytic therapy is bleeding. Intracranial hemorrhage is the most serious and frequently fatal complication; its incidence rate is 0.5 to 1% with currently approved regimens. Older age (>70 to 75 years), female gender, hypertension, and higher relative doses of t-PA and heparin increase the risk for intracranial hemorrhage. The risk-to-benefit ratio should be assessed in each patient when fibrinolysis is considered and specific regimens are selected.

For failed fibrinolysis, rescue PCI is more effective than repeated fibrinolysis. After fibrinolysis, regardless of its apparent success, a preferred strategy is to transfer all STEMI patients with high-risk features rapidly to a hospital with PCI facilities to undergo angiography rather than to transfer only selected patients in whom fibrinolysis has failed or recurrent ischemia has developed. This early transfer and angiography strategy at a median of 3 hours after fibrinolysis reduces the risk for recurrent ischemia, reinfarction, heart failure, cardiogenic shock, or death by 36%.

Selecting a Reperfusion Regimen

Whether to use PCI or fibrinolytic therapy depends on local resources and experience as well as on patient factors. Primary PCI is feasible in community hospitals without surgical capability, but operator and team experience as well as organizational and transfer issues are critical to success. In general, primary PCI with stenting is preferred in experienced facilities that are able to mobilize and to treat patients quickly (<90 minutes door to device system time). PCI is particularly preferred for patients at higher risk for mortality (including shock), for later presentations (>3 hours), and for patients with greater risk of intracranial hemorrhage (age >70 years, female gender, therapy with hypertensive agents).

In non–PCI-capable hospitals and in other situations in which PCI is not feasible or would be significantly delayed (e.g., by a long transfer time to a PCI-capable facility) to more than 120 minutes after first medical contact, fibrinolytic therapy should be given to patients with STEMI within 12 hours of the onset of symptoms unless it is contraindicated. Fibrinolytic therapy is reasonable within 12 to 24 hours of the onset of symptoms in the setting of a large MI or hemodynamic instability. Prehospital fibrinolysis followed by a routine *emergent* (i.e., within 1 to 2 hours) invasive strategy on hospital arrival causes a *higher* rate of in-hospital mortality, cardiac ischemic events, and strokes compared with primary PCI alone or by a more delayed invasive approach after fibrinolysis in stabilized patients and cannot be recommended.

TABLE 73-6 INDICATIONS FOR AND CONTRAINDICATIONS TO FIBRINOLYTIC THERAPY

INDICATIONS
Ischemic-type chest discomfort or equivalent for 30 minutes–12 hours with new or presumed new ST segment elevation in two contiguous leads of ≥2 mm (≥0.2 mV) in leads V_1, V_2, or V_3 or ≥1 mm in other leads
New or presumed new left bundle branch block with symptoms consistent with myocardial infarction
Absence of contraindications
CONTRAINDICATIONS, ABSOLUTE
Active bleeding or bleeding diathesis (menses excluded)
Prior hemorrhagic stroke, ischemic stroke within 3 months, except acute ischemic stroke within 3-4.5 hours
Intracranial or spinal cord neoplasm or arteriovenous malformation
Suspected or known aortic dissection
Closed head or facial trauma within 3 months
CONTRAINDICATIONS, RELATIVE
Severe, uncontrolled hypertension by history or on presentation (>180/110 mm Hg)
Anticoagulation with therapeutic or elevated international normalized ratio (>2-3)
Old ischemic stroke (>3 months ago); intracerebral disease other than above
Recent (<3 weeks) major trauma/surgery or prolonged (>10 minutes) cardiopulmonary resuscitation or internal bleeding
Active peptic ulcer
Recent noncompressible vascular punctures
Pregnancy
For streptokinase/anistreplase: prior exposure (especially if >5 days ago) or allergic reaction

Modified from Kushner FG, Hand M, Smith SC Jr, et al. 2009 Focused updates: ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction and the ACC/AHA/SCAI guidelines on percutaneous coronary intervention. *Circulation.* 2009;120:2271-2306.

However, immediate transfer to a PCI-capable hospital for coronary angiography is recommended if such patients develop cardiogenic shock or severe heart failure and is reasonable in patients with suspected failed reperfusion or reocclusion after fibrinolytic therapy and even in hemodynamically stable patients with apparently successful reperfusion, provided angiography is delayed for at least 2 to 3 hours after fibrinolytic therapy.

Ancillary and Other Therapies

Initial Medical Management

Aspirin (162 to 325 mg) should be given on presentation to all patients unless it is contraindicated (Fig. 73-4). A loading dose of an adenosine diphosphate receptor ($P2Y_{12}$) inhibitor (i.e., clopidogrel, 600 mg, or prasugrel, 60 mg, or ticagrelor, 180 mg) also should be given as early as possible or at the time of primary PCI to STEMI patients for whom an invasive approach is planned. In addition, it is reasonable to start treatment with a GPIIb-IIIa receptor antagonist—abciximab (IV bolus of 0.25 mg/kg, then 0.125 µg/kg/minute [maximum, 10 µg/minute] for up to 12 hours), tirofiban (IV bolus of 25 µg/kg, then 0.15 µg/kg/minute for up to 12 to 18 hours; reduce infusion rate by 50% for estimated creatinine clearance < 30 mL/minute), or eptifibatide (IV bolus of 180 µg/kg, second bolus after 10 minutes, then 2.0 µg/kg/minute for up to 18 hours; reduce infusion by 50% for estimated creatinine clearance < 50 mL/minute)—at the time of primary PCI for STEMI in selected patients, such as those with a large burden of thrombus or those who have not received an adequate loading dose of a $P2Y_{12}$ inhibitor. The value of starting a GPIIb-IIIa receptor antagonist before arrival in the catheterization laboratory is less certain.

Anticoagulant therapy should be initiated on presentation. Options include intravenous heparin (initial bolus of 60 IU/kg [maximum, 4000 IU], then 12 IU/kg/hour [maximum, 1000 IU/hour] for patients >70 kg, adjusted to maintain activated partial thromboplastin time 1.5 to 2 times the control value), low-molecular-weight heparin (LMWH; e.g., enoxaparin, IV bolus of 30 mg, then 1 mg/kg subcutaneously twice daily for patients <75 years old without renal insufficiency with fibrinolysis or IV bolus of 0.5 mg/kg with primary PCI), and bivalirudin (with a primary PCI strategy, bolus of 0.75 mg/kg, then infusion of 1.75 mg/kg/hour). In STEMI patients who are undergoing PCI and who are at higher risk for bleeding, evidence supports use of bivalirudin anticoagulation with a $P2Y_{12}$ inhibitor but without a GPIIb-IIIa receptor antagonist. Fondaparinux may be used as adjunctive anticoagulant therapy with fibrinolysis but not as the sole anticoagulant with primary PCI.

Patients with chest pain should be given sublingual nitroglycerin (0.4 mg every 5 minutes for up to three doses), after which an assessment should be made of the need for intravenous nitroglycerin. Persistent ischemic pain may be treated with titrated intravenous doses of morphine (i.e., 2 to 4 mg intravenously, repeated every 5 to 15 minutes to relieve pain). Initiation of β-blocker therapy is usually indicated, especially in patients with hypertension, tachycardia, and ongoing pain; however, decompensated heart failure is a contraindication to the acute initiation of β-blocker therapy, particularly by the intravenous route. Oxygen should be used, if needed, in doses sufficient to avoid hypoxemia (e.g., initially at 2 to 4 L/minute by nasal cannula; fingertip oximetry may be used to monitor effect). The ideal systolic blood pressure is 100 to 140 mm Hg. Excessive hypertension usually responds to titrated nitroglycerin, β-blocker therapy, and morphine (also given for pain). Relative hypotension could require discontinuation of these medications, fluid administration, or other measures as appropriate to the hemodynamic subset (Table 73-7). Atropine (0.5 to 1.5 mg intravenously) should be available to treat symptomatic bradycardia and hypotension related to excessive vagotonia. Direct transfer to the catheterization laboratory or fibrinolysis followed by transfer to the cardiac intensive care unit should occur as expeditiously as possible.

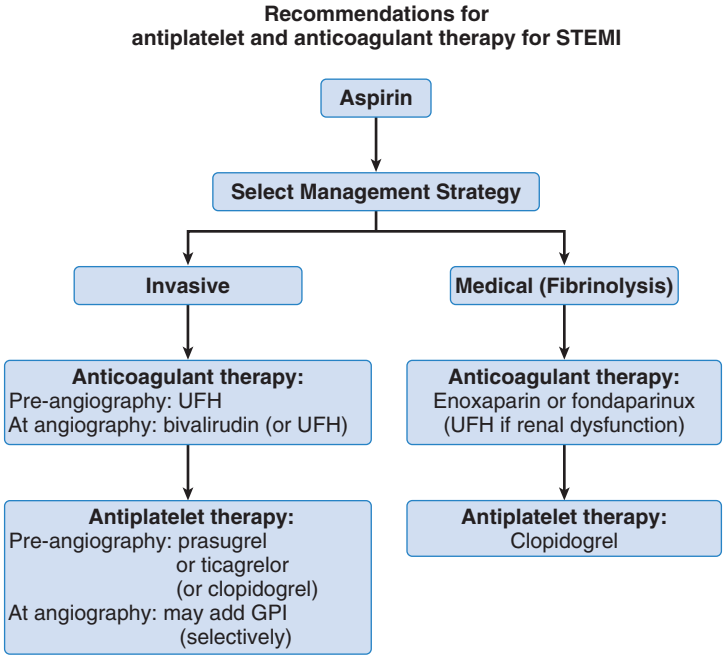

FIGURE 73-4. Recommendations for antiplatelet and anticoagulant therapy for ST segment elevation myocardial infarction (STEMI). See text for doses. GPI = glycoprotein IIb/IIIa inhibitor; UFH = unfractionated heparin.

Early Hospital Phase: Coronary Intensive Care

Coronary intensive care for early hospital management of acute MI has reduced in-hospital mortality by more than 50%. The goals of such care include continuous electrocardiographic monitoring and antiarrhythmic therapy for serious arrhythmias (i.e., rapid defibrillation of VF), initiation or continuation of a coronary reperfusion strategy to achieve myocardial reperfusion, initiation or continuation of other acute medical therapies, hemodynamic monitoring and appropriate medical interventions for different hemodynamic subsets of patients, and diagnosis and treatment of mechanical and physiologic complications of acute MI (Table 73-8).

TABLE 73-7 HEMODYNAMIC SUBSETS OF ACUTE MYOCARDIAL INFARCTION

	BLOOD PRESSURE (RELATIVE)	TYPICAL PHYSICAL FINDINGS	CARDIAC INDEX (L/min/m²)	PA WEDGE PRESSURE (mm Hg)	SUGGESTED INTERVENTIONS
Normal	Normal	$\pm S_4$	>2.5	≤12	None required
Hyperdynamic	Normal or high	Anxious	>3	<12	Control pain, anxiety; β-blocker; treat SBP to <140 mm Hg
Hypovolemia	Low	Dry	≤2.7	≤9	Add fluids to maintain normal pressure; can develop pulmonary edema if hypotension caused by unrecognized LV failure
Mild LV failure	Low to high	Rales, $\pm S_3$	2-2.5	>15	Diuresis; nitrates, ACE inhibitor; consider low-dose β-blocker
Severe LV failure	Low to normal	Above $+S_3$, ± ↑ JVP, ± edema	<2	>20	Diuresis; nitrates; low-dose ACE inhibitor; avoid β-blockers; consider inotropes, urgent revascularization
Cardiogenic shock	Very low	Above + cool, clammy; ↓ mental or renal function	≤1.5	>25	Avoid hypotensive agents; place intra-aortic balloon pump; urgent revascularization if possible
RV infarct	Very low	↑ JVP with clear lungs	<2.5	≤12	Give IV fluids; avoid nitrates and hypotensive agents; dobutamine if refractory to fluids

↑ = increased; ↓ = decreased; ACE = angiotensin-converting enzyme; IV = intravenous; JVP = jugular venous pressure; LV = left ventricle; PA = pulmonary artery; RV = right ventricle; SBP = systolic blood pressure.

Modified from Forrester JS, Diamond G, Chatterjee K, et al. Medical therapy of acute myocardial infarction by application of hemodynamic subsets (second of two parts). *N Engl J Med.* 1976;295:1404-1413.

TABLE 73-8 SAMPLE ADMISSION ORDERS FOR ST SEGMENT ELEVATION ACUTE MYOCARDIAL INFARCTION

Diagnosis	Acute ST segment elevation myocardial infarction
Admit	Coronary care unit with telemetry
Condition	Serious
Vital signs	q½h until stable, then q1-4h and PRN; pulse oximetry × 24 hr; notify if heart rate <50 or >100; respiratory rate <8 or >20; SBP <90 or >150 mm Hg; O_2 saturation <90%
Activity	Bedrest × 12 hr with bedside commode; thereafter, light activity if stable
Diet	NPO except for sips of water until pain free and stable; then 2 g sodium, heart-healthy diet as tolerated, unless on call for catheterization (or other test requiring NPO)
Laboratory tests*	Troponin I or T at 2 hours, then q8h × 3; comprehensive blood chemistry, magnesium, CBC with platelets; PT/INR, aPTT; BNP; lipid profile (fasting in morning); portable CXR
IV therapy	D_5W or NS to keep vein open (increase fluids for relative hypovolemia); second IV if IV medication given
Reperfusion therapy*	Emergency primary coronary angioplasty or fibrinolysis (if appropriate) 1. Primary angioplasty (preferred if available within 90 minutes) 2. Tenecteplase, alteplase, reteplase, or streptokinase (see Table 73-5 for doses), if primary PCI is unavailable within 90-120 minutes
Medications	1. Nasal O_2 at 2 L/min with or at risk of hypoxemia, titrated to keep O_2 saturation >90% 2. Aspirin 162-325 mg chewed on admission, then 81-162 mg PO daily 3. IV heparin, 60-U/kg bolus (maximum, 4000 U) and 12 U/kg/hr (maximum, 1000 U/hr), titrate to target aPTT 1.5-2.0 × control (about 50-70 seconds); *or* enoxaparin (preferred with fibrinolytic), 30 mg IV, then 1 mg/kg SC q12h (maximum SC doses, 100 mg on day 1; reduce to 0.75 mg/kg for age ≥75 years, increase interval to q24h for CrCl <30 mL/min); *or* bivalirudin (with primary PCI), 0.75-mg/kg IV bolus, then 1.75 mg/kg/hr (delay 30 minutes if heparin given) 4. Metoprolol, 12.5 PO q6h, incremented to 25-50 mg q6h as tolerated (hold for SBP < 100 mm Hg, pulse < 50 beats/min, asthma, heart failure); may consider IV metoprolol if immediate effect required (tachyarrhythmia, severe hypertension, unrelieved pain) in the absence of heart failure 5. Consider IV nitroglycerin drip × 24-48 hr (titrated to SBP 100-140 mm Hg) 6. Morphine sulfate, 2-4 mg IV and increment at 5-15 minutes PRN for unrelieved pain 7. Stool softener 8. Anxiolytic or hypnotic if needed 9. ACE inhibitor for hypertension, anterior acute MI, or LV dysfunction, in low oral dose (e.g., captopril, 6.25 mg q8h), begun within 24 hours or when stable (SBP > 100 mm Hg) and adjusted upward 10. Lipid-lowering therapy (i.e., high-intensity statin) regardless of LDL: target LDL reduction of ≥50%; give atorvastatin 80 mg PO on admission (pre-cath) and continue daily 11. Antiplatelet therapy: clopidogrel, 600 mg PO on admission for invasive strategy, 300 mg PO with fibrinolytic, then 75 mg PO daily; *or* ticagrelor, 180 mg PO on admission, then 90 mg bid (with primary PCI strategy); *or* prasugrel, 60 mg PO on admission, then 10 mg PO daily (with primary PCI strategy). For specific treatments for hemodynamic subgroups, see Table 73-7.

*If not ordered in the emergency department.
ACE = angiotensin-converting enzyme; aPTT = activated partial thromboplastin time; BNP = brain natriuretic peptide; CABG = coronary artery bypass graft surgery; CBC = complete blood count; CrCl = creatinine clearance; CXR = chest radiograph; D_5W = 5% dextrose in water; INR = international normalized ratio; IV = intravenous; LDL = low-density lipoprotein; LV = left ventricle; MI = myocardial infarction; NPO = nothing by mouth; NS = normal saline; PCI = percutaneous coronary intervention; PO = orally; PRN = as needed; PT = prothrombin time; daily = once daily; SBP = systolic blood pressure; SC = subcutaneous.
Modified from Kushner FG, Hand M, Smith SC Jr, et al. 2009 Focused updates: ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction and the ACC/AHA/SCAI guidelines on percutaneous coronary intervention. *Circulation.* 2009;120:2271-306.

General care measures include attention to activity, diet and bowels, education, reassurance, and sedation. Bedrest is encouraged for the first 12 hours. In the absence of complications, dangling, bed-to-chair, and self-care activities can begin within 24 hours or earlier after successful reperfusion therapy. When stabilization has occurred, usually within 1 to 3 days, patients may be transferred to a step-down unit where progressive reambulation occurs. The risk for emesis and aspiration or the anticipation of angiography or other procedures usually dictates nothing by mouth or clear liquids for the first 4 to 12 hours. Thereafter, a heart-healthy diet in small portions is recommended. In patients at high risk for bleeding gastric stress ulcers, a proton pump inhibitor or an H_2-receptor antagonist is recommended. Many patients benefit from an analgesic (e.g., morphine sulfate, in 2- to 4-mg increments) to relieve ongoing pain and an anxiolytic or sedative during the acute phase. A benzodiazepine is frequently selected, but routine use of anxiolytics is neither necessary nor recommended. Sedatives should not substitute for education and reassurance from concerned caregivers to relieve emotional distress and improve behavior. Constipation often occurs with bedrest and narcotics; stool softeners and a bedside commode are advised.

The ECG should be monitored continuously during the entire hospital course to detect serious arrhythmias and to guide therapy. Measures to limit infarct size (i.e., coronary reperfusion) and to optimize hemodynamics also stabilize the heart electrically. Routine antiarrhythmic prophylaxis (e.g., with lidocaine or amiodarone) is not indicated, but specific arrhythmias require treatment (see later text).

Hemodynamic evaluation is helpful in assessing prognosis and in guiding therapy (see Table 73-7). Clinical and noninvasive evaluation of vital signs is adequate for normotensive patients without pulmonary congestion. Patients with pulmonary venous congestion alone can usually be managed conservatively. Invasive monitoring is appropriate when the cause of circulatory failure is uncertain and when titration of intravenous therapies depends on hemodynamic measurements (e.g., pulmonary capillary wedge pressure and cardiac output). Similarly, an arterial line is not necessary in all patients and may be associated with local bleeding after fibrinolysis or potent antiplatelet and anticoagulant therapy. Arterial catheters are appropriate and useful in clinically unstable, hypotensive patients who do not respond to intravenous fluids to replete or to expand intravascular volume (see the later discussion of complications).

Later Hospital Phase

Transfer from intensive care to the step-down unit usually occurs within 1 to 3 days, when the cardiac rhythm and hemodynamics are stable. The duration of this late phase of hospital care is usually an additional 1 to 3 days in uncomplicated cases. Activity levels should be increased progressively under continuous electrocardiographic monitoring. Medical therapy should progress from parenteral and short-acting agents to oral medications appropriate and convenient for long-term outpatient use.

Risk stratification and functional evaluations are critical to assess prognosis and to guide therapy as the time for discharge approaches. Functional evaluation also can be extended to the early period after hospital discharge. Education must be provided about diet, activity, smoking, and other risk factors (e.g., lipids, hypertension, and diabetes).

Antiplatelet Therapy

Aspirin

Aspirin (Chapters 37 and 38) reduces the relative risk of vascular death in patients with acute MI and is strongly recommended at a dose of 162 to 325 mg, preferably chewed, both before primary PCI and with fibrinolytic therapy. Aspirin should be continued throughout hospitalization and then indefinitely on an outpatient basis at 81 mg/day; 325 mg/day may be preferable for the first month after PCI in patients who are taking clopidogrel.

Adenosine Diphosphate ($P2Y_{12}$) Receptor Antagonists

In addition to aspirin, *clopidogrel* (300 mg followed by 75 mg/day) is recommended in STEMI patients 75 years of age or younger who are treated with a fibrinolytic, in whom it reduces predischarge occlusion rates of infarct-related

arteries (by 41%) and reduces ischemic complications at 30 days (by 20%) without increasing rates of intracerebral hemorrhage. It is also recommended without a loading dose in STEMI patients older than 75 years who receive fibrinolytic therapy. Clopidogrel also is indicated in STEMI patients undergoing primary PCI with stenting (loading dose, 600 mg); it is less effective but has a lower bleeding risk than prasugrel and ticagrelor (see later). Clopidogrel increases the risk of bleeding and blood transfusions with CABG; if CABG is planned, clopidogrel should be withheld for at least 5 days unless the urgency of surgery outweighs the risk of excessive bleeding. Common genetic variants (i.e., in CYP2C19) may reduce activation of clopidogrel, but dosing guided by genetic testing has not been shown to improve clinical outcomes. Similarly, proton pump inhibitors, especially omeprazole, impair clopidogrel's antiplatelet activity, but coadministration has not been associated with any adverse clinical consequences.

Prasugrel, a more potent thienopyridine, is indicated as adjunctive therapy (60 mg loading dose, 10 mg/day maintenance dose), on admission or at the time of angiography, in patients who receive primary PCI but not in patients who receive fibrinolytic therapy. However, prasugrel is contraindicated in patients with a prior history of stroke or transient ischemia attack, should be used with caution (or in reduced doses) in older (≥75 years) and smaller (<60 kg) patients, and should be withheld for at least 7 days before nonemergent CABG.

Ticagrelor, a direct-acting and non-thienopyridine $P2Y_{12}$ inhibitor, also is more effective than clopidogrel and is indicated for adjunctive therapy (loading dose of 180 mg, maintenance dose of 90 mg twice daily) on admission or at angiography with primary PCI but not with fibrinolytic therapy. Ticagrelor is contraindicated in patients with a history of intracranial hemorrhage.

The recommended duration of $P2Y_{12}$ inhibitor therapy is 1 year after PCI for STEMI with stenting to prevent stent thrombosis and recurrent ischemic events. Platelet function testing can be used to guide dosing of $P2Y_{12}$ inhibitors, but studies to date have not demonstrated sufficient benefit to warrant their routine use. Platelet function testing can be used to document adherence and may prove useful in high-risk patients with drug-eluting stents.

Glycoprotein IIb/IIIa Receptor Antagonists

High-risk patients with non–ST segment elevation acute coronary syndrome benefit from antagonists of the platelet membrane GPIIb-IIIa receptor, either on admission or after PCI (Chapters 38 and 72). The benefit is smaller in STEMI with routine stenting, but it is reasonable to administer a GPIIb-IIIa receptor antagonist at the time of primary PCI in STEMI patients who have a large burden of thrombus or who have not received adequate loading with a $P2Y_{12}$ inhibitor. If early CABG is a possibility after angiography, a shorter-acting inhibitor (eptifibatide, tirofiban) may impart a lower perioperative risk for bleeding than abciximab. Earlier ("upstream") glycoprotein inhibition before hospital admission or in the emergency department (precatheterization) can improve coronary patency at the time of emergency angiography, but incremental benefit on clinical outcomes has not been established.

Anticoagulant Therapy

Low-Molecular-Weight Heparins

In patients with acute STEMI who are treated with fibrinolytic therapy, LMWH can reduce reinfarction rates by 25% and mortality by about 10% compared with unfractionated heparin. Enoxaparin is dosed according to age and weight (i.e., age <75 years: IV bolus of 30 mg, then 1 mg/kg SC every 12 hours [maximum, 100 mg for first two doses]; age ≥75 years: no bolus, 0.75 mg/kg SC every12 hours [maximum, 75 mg, first two doses]) as well as by creatinine clearance (if the creatinine clearance is <30 mL/minute: 1 mg/kg SC every 24 hours) until hospital discharge or up to 8 days. Enoxaparin is more complicated to use with a primary PCI strategy; it is an option in this setting in current European[6] but not in American[1] guidelines.

Unfractionated Heparin

Unfractionated heparin can benefit patients treated with primary PCI or fibrinolytic agents (see Fig. 73-3). When it is given with a fibrin-specific fibrinolytic agent, intravenous heparin is begun concurrently and continued for 48 hours, beginning with a bolus of 60 U/kg (maximum, 4000 U), followed initially by an infusion of 12 U/kg/hour (maximum, 1000 U/hour), with adjustment after 3 hours based on the activated partial thromboplastin time (target of 50 to 70 seconds, 1.5 to 2 times control). Experimental regimens including a GPIIb-IIIa inhibitor and a fibrinolytic agent have used even lower heparin doses. During primary PCI, high-dose heparin (bolus of 70 to 100 U/kg) is used to achieve an activated clotting time of 250 to 300 seconds. Given together with a GPIIb-IIIa inhibitor during PCI, the dose of heparin is adjusted (bolus of 50 to 70 U/kg) to achieve a lower activated clotting time range (200 to 250 seconds). Additional boluses are given as needed to maintain therapeutic activated time levels.

Factor Xa Inhibitors

Selective factor Xa inhibition (i.e., with fondaparinux, 2.5 mg, initial dose intravenously, then subcutaneously once daily for up to 8 days during index hospitalization) may reduce death or reinfarction at 30 days by 18 to 23% independent of heparin use in patients who receive fibrinolysis or no reperfusion therapy, but it is not of benefit in patients who have undergone PCI. As a result, it may be a preferred alternative to unfractionated heparin or no heparin (e.g., in patients who present later and in patients treated with streptokinase) in patients with STEMI who are not undergoing a primary PCI strategy. Fondaparinux is contraindicated if the creatinine clearance is less than 30 mL/minute.

Direct Antithrombins

Bivalirudin, a synthetic hirudin analogue with direct antithrombin activity (Chapter 38), combined with early administration of high-dose clopidogrel is better than the combination of heparin plus a GPIIb-IIIa inhibitor for STEMI patients who undergo primary PCI. The recommended dose is an IV bolus of 0.75 mg/kg, followed by an infusion of 1.75 mg/kg/hour; an additional bolus of 0.3 mg/kg can be given if needed. Bivalirudin has not been tested and is not indicated in patients receiving fibrinolytic therapy.

Other Pharmacologic Therapies

Nitrates

Nitroglycerin and other organic nitrates (isosorbide dinitrate and isosorbide mononitrate) reduce excessive cardiac preload and afterload, increase coronary caliber in responsive areas of stenosis, reverse distal small coronary arterial vasoconstriction, improve coronary collateral flow to ischemic myocardium, and inhibit platelet aggregation in acute MI (Chapter 71). The results are improved oxygen delivery and reduced oxygen consumption.

Nitroglycerin is useful for the first 24 to 48 hours for patients with acute MI and pulmonary congestion, large anterior MI, persistent ischemia, or hypertension. In patients treated with fibrinolytic therapy and aspirin, nitrates provide a modest relative survival benefit of about 4 lives saved per 1000 patients treated. In patients undergoing primary PCI, however, no such benefit has been shown.

When it is used in the setting of an acute MI, intravenous nitroglycerin should begin with a bolus injection of 12.5 to 25 μg followed by an infusion of 10 to 20 μg/minute. The infusion rate is increased by 5 to 10 μg/minute every 5 to 10 minutes up to about 200 μg/minute during hemodynamic monitoring until clinical symptoms are controlled or blood pressure targets are reached (blood pressure decreased by 10% in normotensive patients or by 30% in hypertensive patients but not to less than 80 mm Hg mean or 90 mm Hg systolic). Nitrates should be avoided in patients with hypotension, marked bradycardia or tachycardia, or RV infarction, and they are not indicated routinely during the convalescent phase of STEMI.

β-Blockers

β-Adrenoceptor blockers reduce heart rate, blood pressure, and myocardial contractility, and they stabilize the heart electrically. These actions provide clinical benefit to most patients with acute MI by limiting myocardial oxygen consumption, relieving ischemia, reducing infarct size, and preventing serious arrhythmias.

Early (first-day) β-blockade is generally recommended for STEMI patients who do not have signs of heart failure, evidence of a low-output state or increased risk for cardiogenic shock, or other contraindications, such as heart block, severe bradycardia, and active reactive airways disease, regardless of concomitant fibrinolysis or PCI. Oral β-blocker therapy may be titrated to tolerance or goal (e.g., metoprolol, 25 to 100 mg twice daily; atenolol, 50 to 100 mg/day; or carvedilol, 6.25 to 25 mg twice daily). Intravenous β-blockade has been reserved for STEMI patients who have no contraindications to its use and who are hypertensive or have ongoing ischemia. β-Blocker therapy should be continued after hospitalization for all STEMI patients without contraindications, but it also may benefit patients with anterior STEMI if given before reperfusion.

Renin-Angiotensin-Aldosterone System Inhibitors

The renin-angiotensin-aldosterone system is activated in acute MI and heart failure. Use of an ACE inhibitor has been shown to improve remodeling after acute MI (especially after large anterior MI). ACE inhibitors also have demonstrated efficacy in heart failure, wherein they prevent disease progression, hospitalization, and death (Chapter 59). A meta-analysis of three major trials and 11 smaller ones involving more than 100,000 patients showed an overall mortality reduction of 6.5%, representing about 5 lives saved per 1000 patients treated. Benefit is concentrated and greater in higher-risk patients with large or anterior MI and with LV dysfunction or heart failure, although patients with lesser degrees of LV dysfunction and only moderate cardiovascular risk can also benefit in the long term.

Oral *ACE inhibitor therapy* should begin within the first 24 hours in patients with anterior infarction, heart failure, or low ejection fraction (≤0.40) in the absence of hypotension (systolic pressure <100 mm Hg or >30 mm Hg less than usual baseline) or other contraindications. An *angiotensin receptor blocker* (ARB) should be given to otherwise qualifying patients who are intolerant of ACE inhibitors. An ACE inhibitor or an ARB also is reasonable for other patients with STEMI, especially those with a relative indication (e.g., hypertension, diabetes, or mild renal insufficiency), and in those with the expectation of a smaller but worthwhile benefit.

All patients without contraindications or intolerance to initial ACE inhibitor or ARB therapy also should receive these drugs during the in-hospital

convalescent phase. ACE inhibitor therapy should begin with low oral doses and should be progressively advanced to a full or maximally tolerated dose. For example, the short-acting agent captopril may be started in a dose of 6.25 mg or less and adjusted during 1 to 2 days to 50 mg twice daily. Before discharge, a transition may be made in graded dose schedules to longer-acting agents such as ramipril (2.5 mg titrated to 10 mg/day), lisinopril (2.5 to 5 mg titrated to 10 mg/day), or enalapril (2.5 mg titrated to up to 20 mg twice daily). In patients who cannot tolerate ACE inhibitors (e.g., because of cough), graded doses of an ARB may be substituted (e.g., valsartan, 80 to 160 mg twice daily, or losartan, 50 to 100 mg/day).

Selective *aldosterone receptor blockade* with eplerenone (25 to 50 mg/day) reduces total and cardiovascular mortality (including sudden death) as well as cardiovascular hospitalizations in post-MI patients with an ejection fraction of 0.40 or less and heart failure or diabetes and who are already receiving other optimal therapies, including ACE inhibitors.[■] Spironolactone also benefits patients with advanced heart failure, including those in whom it is caused by a remote MI. Hence, aldosterone receptor blockade should be added to other standard therapies (i.e., ACE inhibitors and β-blockers) during convalescence in STEMI patients with an ejection fraction of 0.40 or less and either symptomatic heart failure or diabetes mellitus. Hyperkalemia, the most common side effect, requires monitoring (Chapter 59).

Antiarrhythmic Agents and Implantable Cardioverter-Defibrillators

Antiarrhythmic therapy is reserved for treatment of or short-term prevention after symptomatic or life-threatening atrial fibrillation (AF) or ventricular arrhythmias, together with other appropriate measures (cardioversion, treatment of ischemia and metabolic disturbances). When medications are indicated, amiodarone usually is the most appropriate agent.

An implantable cardioverter-defibrillator (ICD) is indicated in patients with VF or hemodynamically significant sustained ventricular tachycardia (VT) occurring more than 2 days after STEMI and not due to a transient or reversible cause (e.g., ischemia, reinfarction, metabolic abnormalities). An ICD also may be considered for patients with severe LV dysfunction (ejection fraction ≤0.30) at least 40 days after STEMI and 3 months after CABG without spontaneous or induced VT or VF. These differences reflect an apparent time dependence, in which the benefit of an ICD appears to be delayed until the early post-MI and post-revascularization periods (Chapter 66). By comparison, early ICD implantation is not beneficial in a broader group of patients because its usefulness in preventing similar deaths is offset by the high rate of nonsudden deaths.[■]

Inotropes

Digitalis and intravenous inotropes can increase oxygen demand, provoke serious arrhythmias, and extend infarction. Current opinion supports the use of digoxin in selected patients recovering from acute MI who develop supraventricular tachyarrhythmias (e.g., AF) or heart failure refractory to ACE inhibitors and diuretics. Intravenous inotropes (e.g., dobutamine, dopamine, milrinone, and norepinephrine) are reserved for temporary support of patients with hypotension and circulatory failure that is unresponsive to volume replacement (Chapters 59 and 107).

Lipid-Lowering Therapy

Lipid lowering, particularly with statins, reduces event rates in patients with coronary disease (Chapter 206). Initiation and continuation of high-intensity statin therapy (e.g., atorvastatin, 80 mg daily) is recommended in the setting of MI,[7,■] independent of fasting lipid levels, although it is reasonable to obtain a fasting lipid profile within 24 hours of admission.

Other Medical Therapies

Calcium-channel blockers, although anti-ischemic, also are negatively inotropic and have not been shown to reduce mortality after acute STEMI. Furthermore, certain agents may cause harm in some patients. Verapamil or diltiazem (heart rate–slowing drugs) may be given to patients in whom β-blockers are ineffective or contraindicated for control of rapid ventricular response with AF or relief of ongoing ischemia in the absence of heart failure, LV dysfunction, or atrioventricular (AV) block (Chapter 64).

In comatose STEMI patients resuscitated from out-of-hospital cardiac arrest due to VF or pulseless VT, *therapeutic hypothermia* to a targeted temperature of 36°C is beneficial and should be started as soon as possible after hospital arrival (Chapter 63). Prehospital cooling does not provide incremental benefit.

Glucose-insulin-potassium affords no benefit on mortality, cardiac arrest, or cardiogenic shock when this combination is added to usual care in patients with acute STEMI. However, *glucose control*, by an insulin-based regimen to achieve and to maintain glucose levels of less than 180 mg/dL while avoiding hypoglycemia, is recommended in the acute phase of STEMI. After the acute phase, individualized treatment is indicated with agents or combinations of agents that best achieve glycemic control and are well tolerated (Chapter 229).

Magnesium is of no benefit in patients with acute MI who are treated with fibrinolysis. Supplementation is recommended if the magnesium level is below normal or in patients with torsades de pointes–type VT associated with a prolonged QT interval. Intracoronary infusion of *autologous bone marrow mononuclear cells* is not effective.

Management of Complications

Recurrent Chest Pain

When chest pain recurs after acute MI, the diagnostic possibilities include post-infarction ischemia, pericarditis, infarct extension, and infarct expansion. Characterization of the pain, physical examination, electrocardiography, echocardiography, and cardiac marker determinations assist in the differential diagnosis. CK-MB may discriminate reinfarction better than cTnI or cTnT.

Post-infarction angina developing spontaneously during hospitalization for acute MI despite medical therapy usually merits coronary angiography. β-Blockers (intravenously, then orally) and nitroglycerin (intravenously, then orally or topically) are recommended medical therapies. Pain with recurrent ST segment elevation or recurrent elevation of cardiac markers may be treated with readministration of t-PA or, possibly, a GPIIb-IIIa inhibitor, together with nitroglycerin, β-blockade, and heparin. Streptokinase, which induces neutralizing antibodies, generally should not be readministered after the first few days. If facilities for angiography, PCI, and surgery are available, an invasive approach is recommended to relieve discomfort occurring hours to days after an acute MI that is associated with objective signs of ischemia. Radionuclide perfusion stress testing can be helpful in patients with discomfort that is transient or of uncertain ischemic origin. For lesions with questionable degrees of stenosis at angiography, coronary pressure (fractional flow reserve) or Doppler velocimetry or intracoronary ultrasound can determine whether PCI is warranted.

Infarct expansion implies circumferential slippage with thinning of the infarcted myocardium. Infarct expansion can be associated with chest pain but without recurrent elevation of cardiac markers. Expansive remodeling can lead to an LV aneurysm. The risk for remodeling is reduced with early reperfusion therapy and administration of ACE inhibitors.

Acute pericarditis most commonly is manifested on days 2 to 4 in association with large, transmural infarctions causing pericardial inflammation. On occasion, hemorrhagic effusion with tamponade develops; thus, excessive anticoagulation should be avoided. Pericarditis developing later (2 to 10 weeks) after acute MI could represent Dressler syndrome, which is believed to be immune mediated. The incidence of this post-MI syndrome has decreased dramatically in the modern reperfusion era. Pericardial pain after STEMI is treated with aspirin, but colchicine, acetaminophen, or narcotic analgesics are reasonable if aspirin (even in high doses) is not effective. Glucocorticoids and nonsteroidal anti-inflammatory drugs are potentially harmful after STEMI and should be avoided.

Rhythm Disturbances

Ventricular Arrhythmias

Acute MI is associated with a proarrhythmic environment that includes heterogeneous myocardial ischemia, heightened adrenergic tone, intracellular electrolyte disturbance, lipolysis and free fatty acid production, and oxygen free radical production on reperfusion. Arrhythmias thus are common early during acute MI. Micro-re-entry is likely the most common electrophysiologic mechanism of early-phase arrhythmias, although enhanced automaticity and triggered activity also are observed in experimental models.

Primary VF, the most serious MI-related arrhythmia, contributes importantly to mortality within the first 24 hours. It occurs with an incidence of 3 to 5% during the first 4 hours and then declines rapidly during 24 to 48 hours. Polymorphic VT and, less commonly, monomorphic VT are associated life-threatening arrhythmias that can occur in this setting. Clinical features (including warning arrhythmias) are not adequately specific or sensitive to identify patients at risk for sustained ventricular tachyarrhythmias, so all patients should be continuously monitored. Prophylactic lidocaine, which reduces primary VF but does not decrease (and may increase) mortality, is not recommended. Primary VF is associated with a higher rate of in-hospital and short-term mortality, but longer-term risk of recurrent VT or VF and mortality is largely unaffected in survivors.

Accelerated idioventricular rhythm (60 to 100 beats per minute) frequently occurs within the first 12 hours and is generally benign (i.e., is not a risk factor for VF). Indeed, accelerated idioventricular rhythm frequently heralds reperfusion after fibrinolytic therapy. Antiarrhythmic therapy is not indicated except for sustained, hemodynamically compromising accelerated idioventricular rhythm.

Late VF, which is defined as VF developing more than 48 hours after the onset of acute MI, often occurs in patients with larger MIs or heart failure, portends a worse prognosis for survival, and is an indication for aggressive measures (e.g., consideration of an ICD). Monomorphic VT resulting from re-entry in the context of a recent or old MI also can appear late after MI, and patients may require long-term therapy (e.g., an ICD, see earlier).

Electrical cardioversion is required for VF and sustained polymorphic VT (unsynchronized shock) and for sustained monomorphic VT that causes hemodynamic compromise (synchronized shock) (Chapters 65 and 66). Brief intravenous sedation is given to conscious, "stable" patients. For slower, stable VT and nonsustained VT requiring therapy, intravenous amiodarone or intravenous lidocaine is commonly considered. After episodes of VT or VF, infusions of antiarrhythmic drugs may be given for 6 to 24 hours; the ongoing risk for

arrhythmia then is reassessed. Electrolyte and acid-base imbalance and hypoxia should be corrected. β-Blockade is useful in patients with frequent polymorphic VT associated with adrenergic activation ("electrical storm"). Additional, aggressive measures should be considered to reduce cardiac ischemia (e.g., emergency PCI or CABG) and to address LV dysfunction (e.g., intra-aortic balloon pump) in patients with recurrent polymorphic VT despite the use of β-blockers or amiodarone, or both.

Patients with sustained VT or VF occurring late in the hospital course should be considered for long-term prevention and therapy. When indicated, an ICD provides greater survival benefit than antiarrhythmic drugs in patients with ventricular tachyarrhythmias and can improve survival after convalescence from acute MI for patients with an ejection fraction of 30% or less, regardless of their rhythm status (see earlier).

Atrial Fibrillation and Other Supraventricular Tachyarrhythmias

AF now occurs in 5 to 10% of patients with an acute MI, usually within the first 24 hours (Chapter 64). The incidence of atrial flutter or another supraventricular tachycardia is much lower. The risk for AF increases with age, larger MIs, heart failure, pericarditis, atrial infarction, hypokalemia, hypomagnesemia, hypoxia, pulmonary disease, and hyperadrenergic states. The incidence of AF is reduced by effective early reperfusion. Hemodynamic compromise with rapid rates and systemic embolism (in ~2%) are adverse consequences of AF. Systemic embolism can occur on the first day, so prompt anticoagulation with heparin is indicated.

Recommendations for management of AF include electrical cardioversion for patients with hemodynamic compromise or ischemia; rate control with intravenous digoxin for patients with ventricular dysfunction (i.e., give 1.0 mg, half initially and half in 4 hours), with an intravenous β-blocker (e.g., metoprolol, 5 mg during 2 minutes to a total of 15 mg during 10 to 15 minutes) in those without clinical ventricular dysfunction, or with intravenous diltiazem or verapamil in hemodynamically compensated patients with a contraindication to β-blockers; and anticoagulation with heparin (or LMWH). Amiodarone, which is generally reserved for patients with or at high risk for recurrence, may be started and continued for 6 weeks if sinus rhythm is restored and maintained.

Bradycardias, Conduction Delays, and Heart Block

Sinus and AV nodal dysfunction is common during acute MI. Sinus bradycardia, a result of increased parasympathetic tone often in association with inferior acute MI, occurs in 30 to 40% of patients. Sinus bradycardia is particularly common during the first hour of acute MI and with reperfusion of the right coronary artery (Bezold-Jarisch reflex). Vagally mediated AV block also can occur in this setting. Anticholinergic therapy (atropine, 0.5 to 1.5 mg IV) is indicated for *symptomatic* sinus bradycardia (heart rate generally <50 beats per minute associated with hypotension, ischemia, or escape ventricular arrhythmia), including ventricular asystole, and *symptomatic* second-degree (Wenckebach) or third-degree block at the AV nodal level (narrow QRS complex escape rhythm). Atropine is not indicated and can worsen infranodal AV block (anterior MI, wide-complex escape rhythm).

New-onset infranodal AV block and intraventricular conduction delays or bundle branch blocks predict substantially increased in-hospital mortality. Fortunately, their incidence has declined in the reperfusion era (from 10 to 20% to ~4%). Mortality is related more to extensive myocardial damage than to heart block itself, so cardiac pacing only modestly improves survival. Prophylactic placement of multifunctional patch electrodes, which allow immediate transcutaneous pacing (and defibrillation) if needed, is indicated for symptomatic sinus bradycardia refractory to drug therapy, infranodal second-degree (Mobitz II) or third-degree AV block, and new or indeterminate-age bifascicular block (LBBB; RBBB with left anterior or left posterior fascicular block) or trifascicular block (bilateral or alternating bundle branch block [any age], bundle branch block with first-degree AV block). Transcutaneous pacing is uncomfortable and is intended for prophylactic and temporary use only. Temporary pacing is indicated in the setting of STEMI for *symptomatic* bradyarrhythmias unresponsive to medical therapy. Patients who require a pacemaker to maintain an adequate heart rate or who are at very high risk (>30%) of requiring pacing (including patients with alternating, bilateral bundle branch block, with new or indeterminate-age bifascicular block with first-degree AV block, and with infranodal second-degree AV block) should have a transvenous pacing electrode inserted as soon as possible for temporary pacing.

Indications for permanent pacing after acute MI depend on the prognosis of the AV block and not solely on symptoms. Indications include even transient second- or third-degree AV block in association with bundle branch block and ongoing *symptomatic* AV block at any level. However, block at the AV nodal level (Wenckebach) rarely is persistent or symptomatic enough to warrant permanent pacing.

Heart Failure and Other Low-Output States

Cardiac pump failure is the leading cause of circulatory failure and in-hospital death from acute MI. Manifestations of circulatory failure can include a weak pulse, low blood pressure, cool extremities, a third heart sound, pulmonary congestion, oliguria, and obtundation. However, several distinct mechanisms, hemodynamic patterns, and clinical syndromes characterize the spectrum of circulatory failure in acute MI. Each requires a specific approach to diagnosis, monitoring, and therapy (see Table 73-7).

Left Ventricular Dysfunction

The degree of LV dysfunction correlates well with the extent of acute ischemia and infarction. Hemodynamic compromise becomes evident when impairment involves 20 to 25% of the left ventricle, and cardiogenic shock or death occurs with involvement of 40% or more (Chapter 107). Pulmonary congestion and S_3 and S_4 gallops are the most common physical findings. Early reperfusion (with fibrinolytic agents, PCI, or CABG) is the most effective therapy to reduce infarct size, ventricular dysfunction, and associated heart failure. Medical treatment of heart failure related to the ventricular dysfunction of acute MI is otherwise generally similar to that of heart failure in other settings (Chapter 59) and includes adequate oxygenation and diuresis (begun early, blood pressure permitting, and continued on a long-term basis if needed). Morphine sulfate (i.e., 2 to 4 mg intravenously, with increments as needed after 5 to 15 minutes or more) is useful for patients with pulmonary congestion. Nitroglycerin also reduces preload and effectively relieves congestive symptoms. Titrated oral ACE inhibitor therapy (e.g., captopril, incremented from 3.125 to 6.25 mg three times daily to 50 mg twice daily as tolerated, or lisinopril, 1.25 to 2.5 mg one or twice daily, titrated as tolerated up to 10 to 20 mg twice daily) also is indicated for heart failure and pulmonary edema unless excessive hypotension (systolic blood pressure <100 mm Hg) is present. Treatment can be begun sublingually (0.4 mg every 5 minutes three times), and then the transition can be made to intravenous therapy (initially 5 to 10 μg/minute, incrementing by 5 to 20 μg/minute until symptoms are relieved or until mean arterial pressure falls by 10% in normotensive patients or 30% in hypertensive patients but not <90 mm Hg or >30 mm Hg lower than baseline).

Intravenous vasodilator therapy to reduce preload and afterload (as blood pressure permits), inotropic support, intra-aortic balloon counterpulsation (IABP), and LV assist devices, together with urgent reperfusion, are indicated in cardiogenic shock (see later and Chapter 107).

Volume Depletion

Relative or absolute hypovolemia is a frequent cause of hypotension and circulatory failure and is easily corrected if it is recognized and treated promptly. Poor hydration, vomiting, diuresis, and disease- or drug-induced peripheral vasodilation can contribute to this condition. Hypovolemia should be identified and corrected with intravenous fluids before more aggressive therapies are considered. An empirical fluid challenge may be tried in the appropriate clinical setting (e.g., for hypotension in the absence of congestion, for inferior or RV infarction, and for hypervagotonia). If filling pressures are measured, cautious fluid administration to a pulmonary capillary wedge pressure of up to about 18 mm Hg may optimize cardiac output and blood pressure without impairing oxygenation.

Right Ventricular Infarction

RV ischemia and infarction occur with proximal occlusion of the right coronary artery (before the take-off of the RV branches). Ten percent to 15% of inferior acute STEMIs show classic hemodynamic features, and these patients form the highest-risk inferior MI subgroup for morbidity and mortality (25 to 30% vs. <6% hospital mortality). Improvement in RV function commonly occurs over time, a finding suggesting reversal of ischemic stunning and other favorable accommodations if short-term management is successful.

Hypotension in patients with clear lung fields and elevated jugular venous pressure in the setting of inferior or inferoposterior acute MI should raise the suspicion of RV infarction. Kussmaul sign (distention of the jugular vein on inspiration) is relatively specific and sensitive in this setting. ST segment elevation in lead V_1 and in right precordial lead V_4R (Chapter 54), particularly in the first 24 hours, is the most sensitive electrocardiographic marker of RV infarction. Echocardiography is helpful in confirming the diagnosis (RV dilation and dysfunction are observed). When right-sided heart pressures are measured, a right atrial pressure of 10 mm Hg or greater and 80% or more of the pulmonary capillary wedge pressure are relatively sensitive and specific for RV ischemic dysfunction.

Management of RV infarction consists of early maintenance of RV preload with intravenous fluids, reduction of RV afterload (i.e., afterload-only reducing drugs as for LV dysfunction; consider intra-aortic balloon pump), early reperfusion, short-term inotropic support if needed, and avoidance of venodilators (e.g., nitrates) and diuretics used for LV failure (they may cause marked hypotension). Volume loading with normal saline solution alone is often effective. If the cardiac output fails to improve after 0.5 to 1 L of fluid, inotropic support with intravenous dobutamine (starting at 2 μg/kg/minute and titrating to hemodynamic effect or tolerance, up to 20 μg/kg/minute) is recommended. High-grade AV block is common, and restoration of AV synchrony with temporary AV sequential pacing can lead to substantial improvement in cardiac output. Because the onset of AF (in up to one third of RV infarcts) can cause severe hemodynamic compromise, it requires prompt cardioversion. Early coronary reperfusion with fibrinolysis or PCI markedly improves outcomes.

Cardiogenic Shock

Cardiogenic shock (Chapter 107) is a form of severe LV failure characterized by marked hypotension (systolic pressures <80 mm Hg) and reductions in cardiac index (to <1.8 L/minute/m^2) despite high LV filling pressure (pulmonary capillary wedge pressure >18 mm Hg). The cause is loss of a critical

functional mass (>40%) of the left ventricle. Cardiogenic shock is associated historically with mortality rates of more than 70 to 80% despite aggressive medical therapy. Risk factors include age, large (usually anterior) acute MI, previous MI, and diabetes. In patients with suspected shock, hemodynamic monitoring and IABP are indicated. Intubation often is necessary. Vasopressors are often needed.

Emergency revascularization with either PCI or CABG is recommended in STEMI patients with cardiogenic shock irrespective of the time delay from MI onset (Chapter 107). *IABP* and LV assist devices can be useful for patients with medically refractory unstable ischemic syndromes and cardiogenic shock. Primary IABP therapy for cardiogenic shock associated with acute MI provides temporary stabilization but does not reduce mortality. IABP is currently recommended in the setting of acute MI as a stabilizing measure for patients undergoing angiography and subsequent PCI or surgery for cardiogenic shock, mechanical complications (acute mitral regurgitation, acute ventricular septal defect), refractory post-MI ischemia, or recurrent intractable VT or VF associated with hemodynamic instability. In patients with acute anterior MI without shock, however, IABP plus PCI is no better than PCI alone for reducing infarct size. IABP also is not useful in patients with significant aortic insufficiency or severe peripheral vascular disease. LV assist devices may provide superior hemodynamic support, but clinical trials evidence is still limited.

Mechanical Complications

Mechanical complications usually occur within the first days to weeks and account for approximately 15% of MI-related deaths. Such complications include acute mitral valve regurgitation, ventricular septal defect, free wall rupture, and LV aneurysm. Suspicion and investigation of a mechanical defect should be prompted by a new murmur or sudden, progressive hemodynamic deterioration with pulmonary edema or a low-output state. Transthoracic or transesophageal Doppler echocardiography usually establishes the diagnosis. A balloon flotation catheter can be helpful in confirming the diagnosis. Arteriography to identify correctable coronary artery disease is usually warranted, and interim support with IABP may be useful. However, surgical consultation should be requested promptly, and urgent repair is usually indicated.

Acute mitral valve regurgitation (Chapter 75) results from infarct-related rupture or dysfunction of a papillary muscle. Total rupture leads to death in 75% of patients within 24 hours. Medical therapy is initiated with nitroprusside (beginning with 0.1 μg/kg/minute and titrating upward every 3 to 5 minutes to the desired effect, as tolerated by blood pressure response, up to 5 μg/kg/minute), to lower preload and to improve peripheral perfusion, and inotropic support (e.g., dobutamine, titrated from 2 up to 20 μg/kg/minute in normotensive patients; dopamine, titrated from 2 up to 20 μg/kg/minute in hypotensive patients; or combined dobutamine and dopamine). IABP is used to maintain hemodynamic stability. For papillary muscle rupture, emergency surgical repair (if possible) or replacement (more commonly) is then undertaken. Surgery is associated with high mortality (≥20%), but it leads to better functional and survival outcomes than medical therapy alone. For patients with ischemic (functional) mitral regurgitation but intact anatomy, key interventions include early reperfusion, diuretics, and afterload reduction. If residual mitral regurgitation is severe, however, surgery with mitral valve repair, mitral ring annuloplasty, or often mitral valve replacement may be needed.

Post-infarction septal rupture with *ventricular septal defect*, which occurs with increased frequency in elderly patients, in patients with hypertension, and possibly after fibrinolysis, also warrants urgent surgical repair. Because a small post-MI ventricular septal defect can suddenly enlarge and cause rapid hemodynamic collapse, all septal perforations should be repaired. On diagnosis, invasive monitoring is recommended, together with vasodilators (e.g., nitroprusside, initially 0.1 μg/kg/minute, titrated upward every 3 to 5 minutes to desired effect, as tolerated by blood pressure response, up to 5 μg/kg/minute) and, if needed, judicious use of inotropic agents (e.g., dobutamine, titrated from 2 up to 20 μg/kg/minute in normotensive patients; dopamine, titrated from 2 up to 20 μg/kg/minute in hypotensive patients; or combined dobutamine and dopamine). An intra-aortic balloon pump should be inserted, a surgical consultation promptly obtained, and the surgical repair undertaken as soon as feasible. Percutaneous closure is an appealing,[8] less invasive option for initial emergent treatment, but experience is limited and residual shunts are common.

LV free wall rupture usually causes acute cardiac tamponade with sudden death. In a small percentage of cases, however, resealing or localized containment ("pseudoaneurysm") can allow medical stabilization, usually with inotropic support or IABP, followed by emergency surgical repair.

An *LV aneurysm* can develop after a large, most commonly anterior, acute MI. If refractory heart failure, VT, or systemic embolization occurs despite medical therapy and PCI, aneurysmectomy with CABG is indicated.

Thromboembolic Complications

The risk of thromboembolism has declined markedly with contemporary therapy for STEMI. Systemic arterial emboli (including cerebrovascular emboli) typically arise from an LV mural thrombus, whereas pulmonary emboli commonly arise from thrombi in leg veins. Arterial embolism can cause dramatic clinical events, such as hemiparesis, loss of a pulse, ischemic bowel, or sudden hypertension, depending on the regional circulation involved.

Mural thrombosis with embolism typically occurs in the setting of a large (especially anterior) acute STEMI and heart failure. The risk for embolism is particularly high when a mural thrombus is detected by echocardiography. Thus, in patients with anterior acute STEMI and in other high-risk patients, echocardiography should be performed during hospitalization; if results are positive, anticoagulation should be started (with an anticoagulant), if not already initiated, and continued (with warfarin) for 6 months.

Deep venous thrombosis can be prevented by lower extremity compression therapy, by limiting the duration of bedrest, and by the use of subcutaneous unfractionated heparin or LMWH (in patients at risk not receiving intravenous heparin) until patients are fully ambulatory (Chapter 81). Patients with pulmonary embolism are treated with intravenous heparin and then oral anticoagulation for 6 months (Chapter 98).

Adding an anticoagulant (e.g., warfarin) to aspirin and a $P2Y_{12}$ receptor inhibitor for the prevention of venous thromboembolic disease (Chapters 176 and 98) or the prevention of systemic embolization in patients with AF (Chapter 64) should be restricted to situations in which the benefit of preventing a thromboembolic event exceeds that of the increased risk of bleeding. Anticoagulant therapy is indicated in STEMI patients with AF and a high risk of systemic emboli, mechanical heart valves, recent venous thromboembolism, or a hypercoagulable disorder.[1] The duration of triple antithrombotic therapy should be limited to the shortest time possible to minimize the risk of bleeding. Anticoagulant therapy also is reasonable for patients with STEMI and an asymptomatic LV mural thrombus and might be considered for at-risk patients with anterior akinesis or dyskinesis, with the duration of anticoagulation being limited to 3 months. In the absence of data on the new anti–factor Xa agents, warfarin, usually to an international normalized ratio of 2.0 to 2.5, is recommended. To minimize bleeding in at-risk patients, the physician must pay careful attention to the appropriate selection of antithrombotic regimens and to their doses based on age, weight, renal function, and other comorbidities.

Risk Stratification after Myocardial Infarction

Risk stratification is a continuous process that begins on admission and continues through hospital discharge. The goals of risk stratification before and early after discharge for acute MI are to assess ventricular and clinical function, latent ischemia, and arrhythmic risk; to use this information for patient education and prognostic assessment; and to guide therapeutic strategies. Formal risk stratification tools (e.g., Thrombolysis in Myocardial Infarction, Global Registry of Acute Coronary Events) have been developed and validated for acute coronary syndrome patients for use on admission and at discharge (E-Fig. 73-1).

Cardiac Catheterization and Noninvasive Stress Testing

Risk stratification generally involves functional assessment by one of three strategies: cardiac catheterization, submaximal exercise stress electrocardiography before discharge (at 4 to 6 days), or symptom-limited stress testing at 2 to 6 weeks after discharge. Many or most patients with acute STEMI undergo invasive evaluation for primary PCI or after fibrinolytic therapy. Catheterization generally is performed during hospitalization for patients at high risk. In others, predischarge submaximal exercise testing (to peak heart rate of 120 to 130 beats per minute or 70% of the predicted maximum) appears safe when it is performed in patients who are ambulating without symptoms; it should be avoided within 2 to 3 days of acute MI and in patients with unstable post-MI angina, uncompensated heart failure, or serious cardiac arrhythmias. Alternatively or in addition, patients may undergo symptom-limited stress testing at 2 to 6 weeks before they return to work or resume other increased physical activities. Abnormal test results include not only ST segment depression but also low functional capacity, exertional hypotension, and serious arrhythmias. Patients with positive test results should be considered for coronary angiography.

The sensitivity of stress testing can be augmented with radionuclide perfusion imaging (Chapter 56) or echocardiography (Chapter 55). Supplemental imaging also can quantify the LV ejection fraction and size the area of infarction or ischemia (e.g., by cardiac magnetic resonance imaging; Chapter 56). For patients with ST segment or QRS changes that preclude accurate interpretation of the ECG, an imaging study is recommended with initial stress testing. In others, an imaging study may be performed selectively for those in whom the exercise electrocardiography test result is positive or equivocal. For patients unable to exercise, pharmacologic stress testing can be performed with adenosine, a long-acting bolus analogue of adenosine (e.g., regadenoson), or dipyridamole scintigraphy or by dobutamine echocardiography.

To assess LV function, ejection fraction should be measured in all patients with STEMI before discharge. Patients who initially have a reduced LV ejection fraction and who are possible candidates for ICD therapy should have the LV ejection fraction reevaluated 40 days or more after discharge.

Electrocardiographic Monitoring

Modern telemetry systems capture complete rhythm information during hospital observations and allow identification of patients with serious arrhythmias, so routine 24- to 48-hour ambulatory electrocardiographic (Holter) monitoring before or after hospital discharge is not recommended. Patients

Risk calculator for 6-month postdischarge mortality after hospitalization for acute coronary syndrome

Record the points for each variable at the bottom left and sum the points to calculate the total risk score. Find the total score on the x-axis of the nomogram plot. The corresponding probability on the y-axis is the estimated probability of all-cause mortality from hospital discharge to 6 months.

Medical history

① Age in years — Points

≤29	0
30–39	0
40–49	18
50–59	36
60–69	55
70–79	73
80–89	91
≥90	100

② History of congestive heart failure

③ History of myocardial infarction

Findings at initial hospital presentation

④ Resting heart rate, beats/min — Points

≤49.9	0
50–69.9	3
70–89.9	9
90–109.9	14
110–149.9	23
150–199.9	35
≥200	43

⑤ Systolic blood pressure, mm Hg

≤79.9	24
80–99.9	22
100–199.9	18
120–139.9	14
140–159.9	10
160–199.9	4
≥200	0

1

⑥ ST-Segment depression 11

Findings during hospitalization

⑦ Initial serum creatinine, mg/dL — Points

0–0.39	1
0.4–0.79	3
0.8–1.19	5
1.2–1.59	7
1.6–1.99	9
2–3.99	15
≥4	20

⑧ Elevated cardiac enzymes 15

⑨ No in-hospital percutaneous coronary invervention 14

Points

① _____
② _____
③ _____
④ _____
⑤ _____
⑥ _____
⑦ _____
⑧ _____
⑨ _____

Total risk score _____ **(Sum of points)**

Mortality risk _____ **(From plot)**

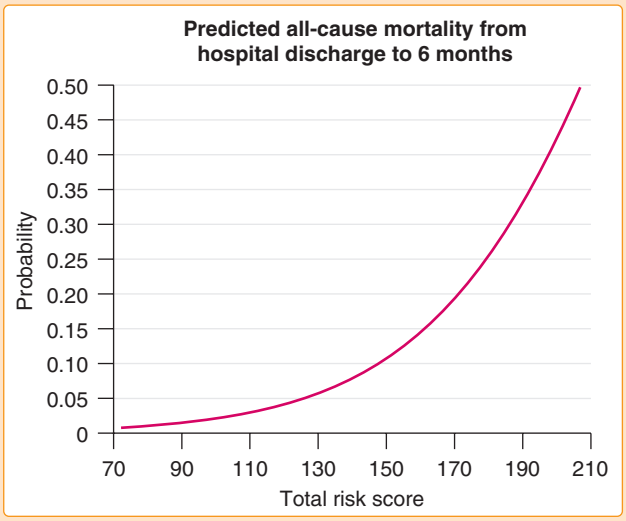

E-FIGURE 73-1. Global Registry of Acute Coronary Events (GRACE) risk score calculator for all-cause mortality from discharge after acute coronary syndrome to 6 months. (Reprinted with permission from Eagle KA, Lim MJ, Dabbous OH, et al. A validated prediction model for all forms of acute coronary syndrome: estimating the risk of 6-month post-discharge death in an international registry. *JAMA.* 2004;291:2727-2733.)

with sustained VT or VF occurring late during hospitalization or provoked during electrophysiologic study with nonsustained VT on monitoring are candidates for an ICD, especially if the ejection fraction is less than 40% (Fig. 73-5) (Chapters 65 and 66). Prophylactic ICD placement at least 1 month after acute MI prevents sudden death for patients with severely depressed function (ejection fraction ≤ 0.30) regardless of the rhythm status.

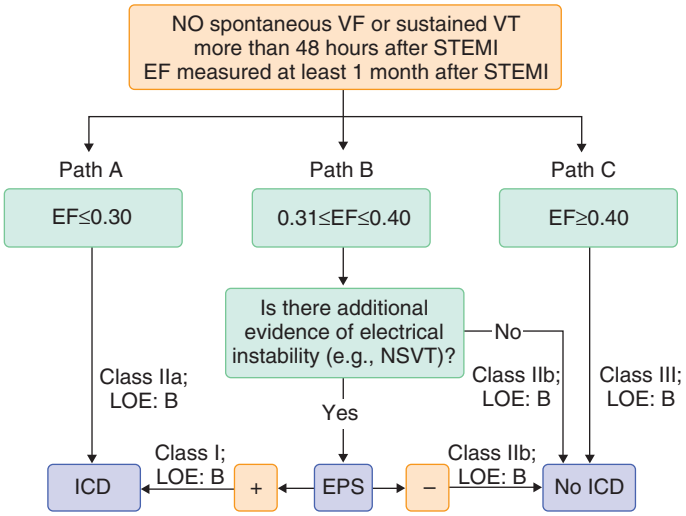

FIGURE 73-5 **Algorithm to aid in selection of implantable cardioverter-defibrillator (ICD) in patients with ST segment elevation myocardial infarction (STEMI) and diminished ejection fraction (EF).** The appropriate management path is selected on the basis of left ventricular EF measured at least 1 month after STEMI. All patients, whether an ICD is implanted or not, should receive medical therapy. EPS = electrophysiologic studies; LOE = level of evidence; NSVT = nonsustained ventricular tachycardia; VF = ventricular fibrillation; VT = ventricular tachycardia. (Modified from Antman EM, Anbe DT, Armstrong PW, et al. 2004 Update: ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction—executive summary. A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines [Writing Committee to Revise the 1999 Guidelines for the Management of Patients with Acute Myocardial Infarction]. *Circulation.* 2004;110:588-636.)

Post-Hospital Care: Secondary Prevention, Patient Education, and Rehabilitation

Before discharge, a detailed, evidence-based plan of care should emphasize the importance of a healthy lifestyle, including diet, exercise, smoking cessation, and compliance with medications. Timely outpatient health care follow-up should be scheduled and a rehabilitation program referral provided.

Secondary Prevention

Advances in secondary prevention have resulted in increasingly effective measures to reduce recurrent MI and cardiovascular death. Secondary prevention should be conscientiously applied after acute MI (Table 73-9).

A fasting *lipid profile* is recommended within 24 hours of admission, and high-intensity *lipid-lowering therapy* with a statin should start in the hospital, preferably on admission, and be continued as outpatient therapy in all patients with STEMI and no contraindication to its use (e.g., atorvastatin, 80 mg daily; Chapter 206).

Continued smoking doubles the subsequent mortality risk after acute MI, and *smoking cessation* reduces the risk for reinfarction and death within 1 year (Chapter 32). An individualized smoking cessation plan should be formulated, including pharmacologic aids (nicotine gum and patches, bupropion, or varenicline).

Antiplatelet therapy (Chapter 38; Fig. 73-6) should consist of aspirin, given on a long-term basis to all patients without contraindications (recommended maintenance dose, 81 mg/day). Clopidogrel (75 mg/day) or prasugrel (10 mg/day) or ticagrelor (90 mg twice daily) is given to patients who received PCI with stenting, and clopidogrel is also appropriate for other patients at higher risk for recurrent vascular events. Therapy should continue for 1 year after stent placement, although a shorter duration may be considered for patients experiencing or at high risk of major bleeding, especially if they have a bare metal stent or a second-generation drug-eluting stent.

Anticoagulant therapy (i.e., warfarin, with an international normalized ratio goal of 2.0 to 3.0) is indicated after acute MI for patients unable to take antiplatelet therapy, for patients with persistent or paroxysmal AF, for patients with LV thrombus, and for patients who have suffered a systemic or pulmonary embolism. Anticoagulants also may be considered for patients with extensive wall motion abnormalities and markedly depressed ejection fraction with or without heart failure. Data on the benefits and risks of warfarin added to antiplatelet therapy (Fig. 73-6) are sparse.

ACE inhibitor therapy can prevent adverse myocardial remodeling after acute MI and can reduce heart failure and death; it is clearly indicated for long-term use in patients with anterior acute MI or an LV ejection fraction of less

TABLE 73-9 DISCHARGE MEDICATION CHECKLIST AFTER MYOCARDIAL INFARCTION*

MEDICATION	DOSES	REASONS NOT TO USE	COMMENTS
Aspirin	Initial: 162-325 mg Maintenance: 75-162 mg daily	High bleeding risk	Reduces mortality, reinfarction, and stroke
Clopidogrel *or*	Initial dose: 300-600 mg (75-150 mg after fibrinolysis in patients >75 years) Maintenance: 75 mg daily	High bleeding risk; suboptimal antiplatelet response	Indicated after PCI for at least 1 year (shorter time for BMS if high bleeding risk); also reduces vascular events when added to aspirin in non–ST segment elevation acute MI (also useful on the basis of clinical trials after ST segment elevation acute MI) Genetic variants (CYP2C19) may reduce response Controversial interaction with proton pump inhibitors (e.g., omeprazole)
Prasugrel *or*	Initial dose: 60 mg Maintenance: 10 mg daily	High bleeding risk	Avoid with history of prior stroke or TIA
Ticagrelor	Initial dose: 180 mg Maintenance: 90 mg bid		Consider 5 mg daily in patients >75 years or <60 kg Avoid if history of intracranial bleed Limit daily aspirin dose to 81 mg
β-Blocker (e.g., metoprolol, carvedilol)	Metoprolol: 25-200 mg daily Carvedilol: 6.25-25 mg bid	Asthma, bradycardia, heart failure	Reduces mortality, reinfarction, sudden death, arrhythmia, hypertension, angina, atherosclerosis progression
ACE inhibitor (e.g., ramipril, lisinopril) or ARB (e.g., valsartan, losartan)	Ramipril: 2.5-10 mg daily Lisinopril: 5-10 mg daily Valsartan: 80-160 mg daily-bid Losartan: 50-100 mg daily	Hypotension, allergy, hyperkalemia	Reduces mortality, reinfarction, stroke, heart failure, diabetes, atherosclerosis progression
Lipid-lowering agent (i.e., a high-intensity statin; e.g., atorvastatin, rosuvastatin)	Atorvastatin: 80 mg daily Rosuvastatin: 20-40 mg daily	Myopathy, rhabdomyolysis, hepatitis	Goal = LDL ≥50% reduction (statins also can benefit patients with lower LDL†)
Nitroglycerin sublingual	0.4 mg SL PRN for angina	Aortic stenosis; sildenafil (Viagra) use	Instruct on PRN use and appropriate need for medical attention

*Medications given at hospital discharge improve long-term compliance.
†Heart Protection Study (*Lancet.* 2002;360:7); and PROVE-IT study (*N Engl J Med.* 2004;350:1495).
ACE = angiotensin-converting enzyme; ARB = angiotensin receptor blocker; BMS = bare metal stent; LDL = low-density lipoprotein; MI = myocardial infarction; PCI = percutaneous coronary intervention; PRN = as needed; daily = once daily; SL = sublingual; TIA = transient ischemic attack.
Modified from Kushner FG, Hand M, Antman EM, et al. 2009 Focused updates: ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction and the ACC/AHA/SCAI guidelines on percutaneous coronary intervention. *Circulation.* 2009;120:2271-2306.

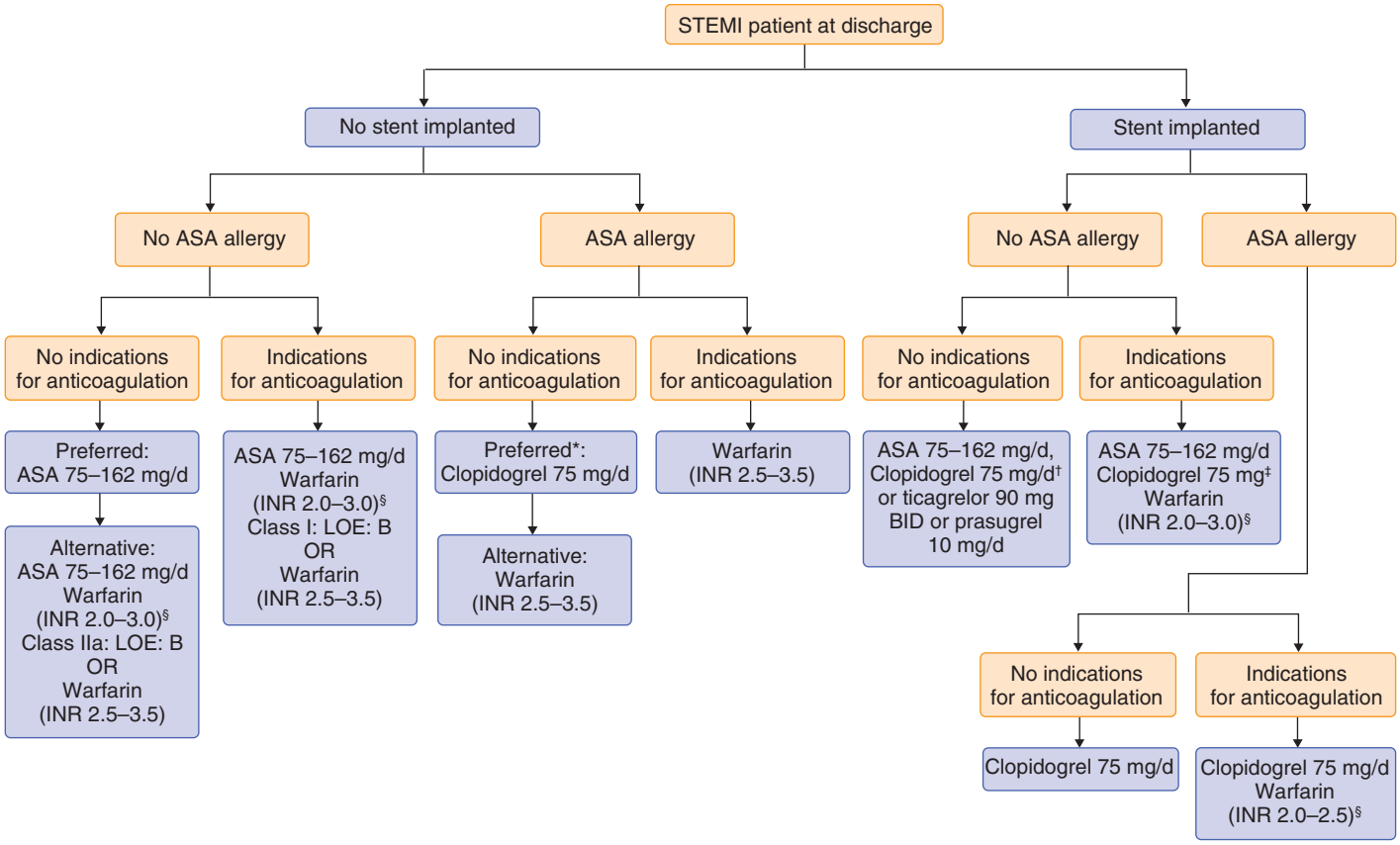

FIGURE 73-6. Long-term antithrombotic therapy at hospital discharge after ST segment elevation myocardial infarction (STEMI). *Clopidogrel is preferred to warfarin because of increased risk of bleeding and low patient compliance in warfarin trials. †For 12 months. ‡Discontinue clopidogrel 1 month after implantation of a bare metal stent or several months after implantation of a drug-eluting stent (3 months after sirolimus and 6 months after paclitaxel) because of the potentially increased risk of bleeding with warfarin and two antiplatelet agents. Continue aspirin (ASA) and warfarin on a long-term basis if warfarin is indicated for other reasons, such as atrial fibrillation, left ventricular thrombus, cerebral emboli, or extensive regional wall motion abnormality. §An international normalized ratio (INR) of 2.0 to 3.0 is acceptable with tight control, but the lower end of this range is preferable. The combination of antiplatelet therapy and warfarin may be considered in patients younger than 75 years who have a low bleeding risk and who can be monitored reliably. LOE = level of evidence. (Modified from Antman EM, Anbe DT, Armstrong PW, et al. ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction—executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines [Writing Committee to Revise the 1999 Guidelines for the Management of Patients with Acute Myocardial Infarction]. *Circulation.* 2004;110:588-636.)

than 40%. ACE inhibitors also reduce recurrent MI in higher-risk patients with an ejection fraction greater than 40%. In contrast, ACE inhibition, when added to other contemporary therapies, provides little additional benefit in reducing cardiovascular events in patients who have stable coronary disease and a low risk (<5%/year) for a coronary event. These data suggest a rationale for the long-term use of ACE inhibitors (e.g., ramipril, 2.5 mg titrated to 10 mg/day, or lisinopril, 2.5 to 5 mg titrated to 10 mg/day) in most patients after MI, except perhaps those at lowest risk (i.e., without heart failure, hypertension, glucose intolerance, or reduced ejection fraction).[1] An ARB (e.g., valsartan, 80 to 160 mg twice daily, or losartan, 50 to 100 mg/day) should be substituted in patients who cannot tolerate an ACE inhibitor; in patients with advanced heart failure, both an ACE inhibitor and an ARB may be complementary (Chapter 59).[1] An aldosterone receptor blocker (e.g., eplerenone, 25 mg/day orally, increased to 50 mg/day after 4 weeks if tolerated, with monitoring of serum potassium levels) also should be added to the ACE inhibitor or ARB (but not both) regimen on a long-term basis in patients with depressed ejection fraction (≤0.40) and clinical heart failure or diabetes, unless this approach is contraindicated.[1]

Long-term β-blocker therapy is strongly recommended for all MI survivors without uncompensated heart failure or other contraindications. Options include metoprolol, 20 to 200 mg/day, and carvedilol, 6.25 to 25 mg twice daily. Long-term therapy in patients at low risk (normal ventricular function, successful reperfusion, absence of arrhythmias) is reasonable but not mandatory.

Nitroglycerin (0.4 mg) is prescribed routinely for sublingual or buccal administration for acute anginal attacks. Longer-acting oral therapy (isosorbide mononitrate, 30 to 60 mg orally every morning, or dinitrate, 10 to 40 mg orally two or three times daily) or topical nitroglycerin (e.g., start with 0.5 inch; can be titrated up to 2 inches, every 6 hours for 2 days) may be added to treatment regimens for angina or heart failure in selected patients.

Calcium-channel blockers are negatively inotropic and are *not* routinely given on a long-term basis. However, they may be given to selected patients without LV dysfunction (ejection fraction >0.40) who are intolerant of β-blockers and who require these drugs for antianginal therapy (e.g., amlodipine, 5 to 10 mg/day orally, or diltiazem, 120 to 480 mg/day orally, as sustained release or in divided doses) or for control of heart rate in AF (e.g., diltiazem, 120 to 480 mg/day orally, or verapamil, 180 to 480 mg/day orally, as sustained release or in divided doses). Short-acting nifedipine should be avoided.

Hormone therapy with estrogen with or without progestin is not begun after an acute MI because it increases thromboembolic risk and does not prevent reinfarction. For women already receiving hormone replacement, therapy should be discontinued unless it is being given for a compelling indication.

Hypertension (Chapter 67) and diabetes mellitus (Chapter 229) must be assessed and tightly controlled in patients after acute MI. ACE inhibitors or β-blockers as described earlier are usually the first-choice therapies for hypertension, with ARBs indicated when ACE inhibitors are not tolerated. ACE inhibitors and ARBs also can reduce the long-term complications of diabetes.

Antioxidant supplementation (e.g., vitamin E, vitamin C) does not benefit patients after acute MI and is *not* recommended. Folate therapy reduces homocyst(e)ine levels but does not reduce clinical events. Routine fish oil supplements also are not supported by accumulating evidence.

Antiarrhythmic drugs are *not* generally recommended after acute MI, and class I antiarrhythmic agents can increase the risk for sudden death. Class III drugs (amiodarone, sotalol, dofetilide) may be used as part of the management strategy for specific arrhythmias (e.g., AF, VT) (Chapters 64 and 65).

Patient Education and Rehabilitation

The hospital stay provides an important opportunity to educate patients about their MI and its treatment, coronary risk factors, and behavioral modification. Education should begin on admission and should continue after discharge. However, the time before hospital discharge is particularly opportune. Many hospitals use case managers and prevention specialists to augment physicians and nurses, to provide educational materials, to review important

concepts, to assist in formulating and actualizing individual risk reduction plans, and to ensure proper and timely outpatient follow-up. This follow-up should include early return appointments with the patient's physician (within a few weeks). Instructions on activities also should be given before discharge. Many hospitals have cardiac rehabilitation programs that provide supervised, progressive exercise. Exercise-based cardiac rehabilitation reduces reinfarction by 47% and overall mortality by 26%.

PROGNOSIS

Both in-hospital and postdischarge prognosis of STEMI has improved during the past several decades with progressive improvement in management, including early reperfusion therapy. STEMI currently is associated with in-hospital and 1-year mortality rates of approximately 5 to 6% and 7 to 18%, respectively.[1] Six-month mortality rates have declined from about 30% in the 1970s to 9% by 2005, and this reduction is directly related to the number of evidence-based therapies given.

Grade A References

A1. Bernat I, Horak D, Stasek J, et al. ST-segment elevation myocardial infarction treated by radial or femoral approach in a multicenter randomized clinical trial: the STEMI-RADIAL trial. *J Am Coll Cardiol*. 2014;63:964-972.

A2. Keeley EC, Boura JA, Grines CL. Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials. *Lancet*. 2003;361:13-20.

A3. Ellis SG, Tendera M, de Belder MA, et al. Facilitated PCI in patients with ST-elevation myocardial infarction. *N Engl J Med*. 2008;358:2205-2217.

A4. Jeger RV, Urban P, Harkness SM, et al. Early revascularization is beneficial across all ages and a wide spectrum of cardiogenic shock severity: a pooled analysis of trials. *Acute Card Care*. 2011;13:14-20.

A5. Hochman JS, Reynolds HR, Dzavik V, et al. Long-term effects of percutaneous coronary intervention of the totally occluded infarct-related artery in the subacute phase after myocardial infarction. *Circulation*. 2011;124:2320-2328.

A6. Wald DS, Morris JK, Wald NJ, et al. Randomized trial of preventive angioplasty in myocardial infarction. *N Engl J Med*. 2013;369:1115-1123.

A7. Wijeysundera HC, Vijayaraghavan R, Nallamothu BK, et al. Rescue angioplasty or repeat fibrinolysis after failed fibrinolytic therapy for ST-segment myocardial infarction: a meta-analysis of randomized trials. *J Am Coll Cardiol*. 2007;49:422-430.

A8. Cantor WJ, Fitchett D, Borgundvaag B, et al. Routine early angioplasty after fibrinolysis for acute myocardial infarction. *N Engl J Med*. 2009;360:2705-2718.

A9. Aversano T, Lemmon CC, Liu L. Outcomes of PCI at hospitals with or without on-site cardiac surgery. *N Engl J Med*. 2012;366:1792-1802.

A10. Sabatine MS, Cannon CP, Gibson CM, et al. Addition of clopidogrel to aspirin and fibrinolytic therapy for myocardial infarction with ST-segment elevation. *N Engl J Med*. 2005;352:1179-1189.

A11. Chen ZM, Jiang LX, Chen YP, et al. Addition of clopidogrel to aspirin in 45,852 patients with acute myocardial infarction: randomised placebo-controlled trial. *Lancet*. 2005;366:1607-1621.

A12. Wiviott SD, Braunwald E, McCabe CH, et al. Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med*. 2007;357:2001-2015.

A13. Wallentin L, Becker RC, Budaj A, et al. Ticagrelor versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med*. 2009;361:1045-1057.

A14. Antman EM, Morrow DA, McCabe CH, et al. Enoxaparin versus unfractionated heparin with fibrinolysis for ST-elevation myocardial infarction. *N Engl J Med*. 2006;354:1477-1488.

A15. Mehran R, Lansky AJ, Witzenbichler B, et al. Bivalirudin in patients undergoing primary angioplasty for acute myocardial infarction (HORIZONS-AMI): 1-year results of a randomised controlled trial. *Lancet*. 2009;374:1149-1159.

A16. Pizarro G, Fernandez-Friera L, Fuster V, et al. Long-term benefit of early pre-reperfusion metoprolol administration in patients with acute myocardial infarction: results from the METOCARD-CNIC trial (Effect of Metoprolol in Cardioprotection During an Acute Myocardial Infarction). *J Am Coll Cardiol*. 2014;63:2356-2362.

A17. Pfeffer MA, McMurray JJ, Velazquez EJ, et al. Valsartan, captopril, or both in myocardial infarction complicated by heart failure, left ventricular dysfunction, or both. *N Engl J Med*. 2003;349:1893-1906.

A18. Pitt B, Remme W, Zannad F, et al. Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction. *N Engl J Med*. 2003;348:1309-1321.

A19. Steinbeck G, Andresen D, Seidl K, et al. Defibrillator implantation early after myocardial infarction. *N Engl J Med*. 2009;361:1427-1436.

A20. Cannon CP, Steinberg BA, Murphy SA, et al. Meta-analysis of cardiovascular outcomes trials comparing intensive versus moderate statin therapy. *J Am Coll Cardiol*. 2006;48:438-445.

A21. Thiele H, Zeymer U, Neumann FJ, et al. Intraaortic balloon support for myocardial infarction with cardiogenic shock. *N Engl J Med*. 2012;367:1287-1296.

A22. Braunwald E, Domanski MJ, Fowler SE, et al. Angiotensin-converting-enzyme inhibition in stable coronary artery disease. *N Engl J Med*. 2004;351:2058-2068.

A23. Lawler PR, Filion KB, Eisenberg MJ. Efficacy of exercise-based cardiac rehabilitation post-myocardial infarction: a systematic review and meta-analysis of randomized controlled trials. *Am Heart J*. 2011;162:571-584.e2.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. O'Gara PT, Kushner FG, Ascheim DD, et al. 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2013;127:e362-e425.
2. Tweet MS, Hayes SN, Pitta SR, et al. Clinical features, management, and prognosis of spontaneous coronary artery dissection. *Circulation*. 2012;126:579-588.
3. de Lemos JA. Increasingly sensitive assays for cardiac troponins: a review. *JAMA*. 2013;309:2262-2269.
4. Quinn T, Johnsen S, Gale CP, et al. Effects of prehospital 12-lead ECG on processes of care and mortality in acute coronary syndrome: a linked cohort study from the Myocardial Ischaemia National Audit Project. *Heart*. 2014;100:944-950.
5. Kristensen SD, Laut KG, Fajadet J, et al. Reperfusion therapy for ST elevation acute myocardial infarction 2010/2011: current status in 37 ESC countries. *Eur Heart J*. 2014;35:1957-1970.
6. Steg PG, James SK, Atar D, et al. ESC guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. *Eur Heart J*. 2012;33:2569-2619.
7. Stone NJ, Robinson J, Lichtenstein AH, et al. 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014;129(suppl 2):S1-S45.
8. Calvert PA, Cockburn J, Wynne D, et al. Percutaneous closure of postinfarction ventricular septal defect: in-hospital outcomes and long-term follow-up of UK experience. *Circulation*. 2014;129:2395-2402.

REVIEW QUESTIONS

1. The recommended duration of antiplatelet therapy with a $P2Y_{12}$ inhibitor after drug-eluting stent placement for ST segment elevation myocardial infarction (STEMI) is

A. 1 month
B. 3 months
C. 6 months
D. 1 year
E. Indefinitely

Answer: D Early discontinuation of $P2Y_{12}$ inhibition is associated with an increased risk of stent thrombosis. A favorable benefit-to-risk ratio of longer-term/indefinite therapy has not been proved.

2. Which of the following scenarios should *not* be considered for immediate reperfusion therapy (e.g., primary percutaneous coronary intervention) in a patient presenting with suspected acute coronary syndrome?

A. ST elevation >2 mm in precordial leads or >1 mm in limb leads
B. New-onset left bundle branch block
C. Resuscitated ventricular fibrillation (VF) arrest
D. ST depression in precordial leads and ST elevation in supplemental leads V_7 to V_9
E. None of the above (i.e., all of the above are correct)

Answer: E A and B are classic indications. VF in the setting of suspected acute coronary syndrome/STEMI is also an indication for rapid percutaneous coronary intervention. D represents a posterior myocardial infarction, which typically is caused by occlusion of the left circumflex coronary artery and should be treated as STEMI.

3. Newly available high-sensitivity, cardiac-specific troponin assays promise the following diagnostic advantages *except*

A. Increased diagnostic sensitivity
B. Increased diagnostic specificity
C. Reduced time to diagnostic rise in troponin levels after the onset of pain
D. All of the above
E. None of the above

Answer: B There are many non–myocardial infarction causes of troponin elevation; the newer, high-sensitivity assays can detect circulating troponin at baseline in a higher percentage of all patients, including those without acute coronary syndromes. As a result, the assays increase sensitivity, reduce time to diagnosis, and reduce specificity, thereby necessitating improved clinical judgment in their interpretation.

4. An implantable cardioverter-defibrillator (ICD) should be considered before discharge for which of the following STEMI patients?

A. STEMI presenting with VF (successfully resuscitated)
B. STEMI with sustained ventricular tachycardia (VT)/VF occurring more than 48 hours after presentation
C. STEMI with ongoing runs of nonsustained VT on telemetry
D. STEMI with predischarge ejection fraction of 30% or less, regardless of rhythm
E. All of the above

Answer: B The risk of recurrent VT/VF is generally low for early (ischemic) VF, given current reperfusion and antithrombotic therapies. In contrast, convalescent presentation (at >48 hours) of sustained VT/VF carries a high risk of recurrence, and early ICD implantation should be considered. For patients with low ejection fractions, clinical trials (and Medicare reimbursement rules) support a waiting period of at least 40 days after myocardial infarction before considering an ICD. For nonsustained VT, implantation of an ICD is generally not indicated except in patients who have a reduced left ventricular ejection fraction and in whom it is induced on an electrophysiologic study.

74

INTERVENTIONAL AND SURGICAL TREATMENT OF CORONARY ARTERY DISEASE

PAUL S. TEIRSTEIN AND BRUCE W. LYTLE

Percutaneous coronary intervention (PCI) and coronary artery bypass graft (CABG) surgery represent alternative and sometimes complementary approaches to coronary revascularization. Each has its relative indications, advantages, disadvantages, and contraindications.

PERCUTANEOUS CORONARY INTERVENTION

Percutaneous coronary intervention is applicable to most forms of coronary artery disease, including multivessel disease, total occlusions, saphenous vein graft disease, unstable angina (Chapter 72), and acute myocardial infarction (MI) (Chapter 73). An estimated 2 million PCIs are performed worldwide each year, making it one of the most widely used medical procedures. Its popularity is based largely on its simplicity, the need for only local anesthesia, a short (≈1 day) hospitalization, and negligible postprocedure recovery time.

Mechanisms and Technical Considerations

Under local anesthesia, a hollow-bore needle is inserted percutaneously into a peripheral artery (usually the femoral or radial artery). A guidewire (≈0.038 inch) is placed through this needle and advanced into the aorta. The needle is removed, leaving the guidewire, over which a small-caliber (≈3 mm), specially shaped catheter (called a guiding catheter) is advanced under fluoroscope guidance into the ostium of the obstructed coronary artery. By use of radiographic contrast injections that provide fluoroscopic visualization of the coronary artery lumen, a thin (≈0.014 inch), highly steerable guidewire is directed down the coronary artery and across the stenotic lesion. This guidewire becomes a "rail" over which therapeutic tools such as inflatable balloons, stents, and atherectomy catheters are passed to the diseased segment (Fig. 74-1).

Balloon catheters (Fig. 74-2) typically have two lumens, one to allow passage over the guidewire and another to carry a mixture of saline and radiographic contrast material to inflate a balloon at the distal catheter tip. Under fluoroscopy, the balloon is centered across the lesion and inflated to 3 to 20 atmospheres of pressure. Balloon inflation widens the narrowed lumen by stretching the vessel and, in most cases, causing a tear (a therapeutic dissection) at the edges of the plaque, where the atheroma meets the nondiseased media. Atherectomy catheters, which also are passed over a guidewire to the

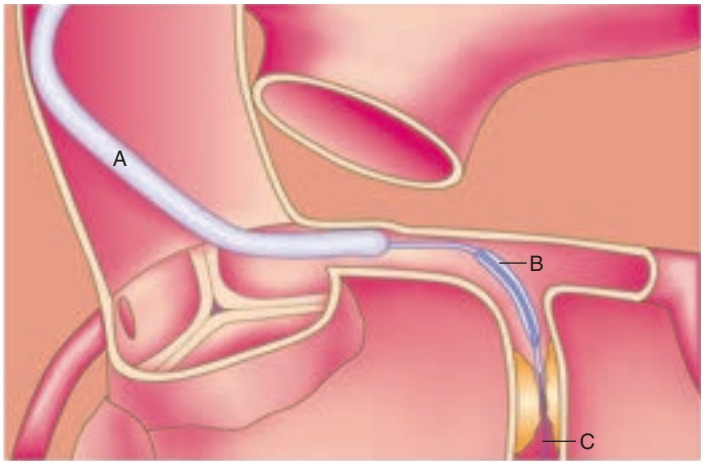

FIGURE 74-1. Schematic view of coronary angioplasty technique. A guide catheter (A) is inserted into the orifice of the coronary artery (in this figure, the left main artery), and a balloon catheter (B) is advanced over a thin guidewire (C) into the lesion. Balloon inflation dilates the stenotic region. (Modified from Baim DS. Percutaneous balloon angioplasty and general coronary intervention. In: Baim DS, ed. *Grossman's Cardiac Catheterization, Angiography, and Intervention.* 7th ed. Philadelphia: Lippincott Williams & Wilkins; 2005.)

diseased segment, remove plaque by a shaving, grinding, slicing, or suction mechanism. Coronary stents are metallic or polymeric scaffolding devices that are crimped onto a deflated balloon catheter before insertion into the diseased vessel (Videos 74-1 to 74-3). During balloon inflation, the collapsed stent expands to support the vessel lumen (Fig. 74-3 and Video 74-4). While balloons and atherectomy devices create an adequate, albeit rough channel through diseased arteries, the supporting structure of the stent can widen the lumen to near its predisease dimensions. With a stent, tissue flaps are "pinned" against the wall, and recoil is limited (Video 74-5). Most stents are designed so that the metallic struts comprise only about 20% of the surface area to allow endothelialization to reduce the risk of thrombosis. Drug-eluting stents release local medications that further reduce the risk of restenosis.[1]

Filters deployed within a coronary vessel beyond the target lesion to limit distal embolization of plaque, platelet aggregates, and other "debris" can further reduce ischemic complications in selected, high-risk patients. Stents composed of polymers or metals that provide a temporary scaffolding and then biodegrade (usually over several years) are currently undergoing intense clinical study.

During the PCI procedure, the interventional cardiologist is able to assess the target vessel fluoroscopically by injections of contrast material through the guiding catheter (Fig. 74-4). When the coronary artery has been opened successfully, all catheters are withdrawn, and the arterial access site is sealed by mechanical pressure, an absorbable plug, or a remote suturing device. Patients without comorbidity ambulate in 3 to 6 hours. Discharge from the hospital usually occurs on the morning after the procedure after stability of the arterial access site, cardiac biomarkers, and electrocardiogram is confirmed. Increasingly, selected patients are treated as outpatients and released 6 to 12 hours after the procedure without an overnight stay.

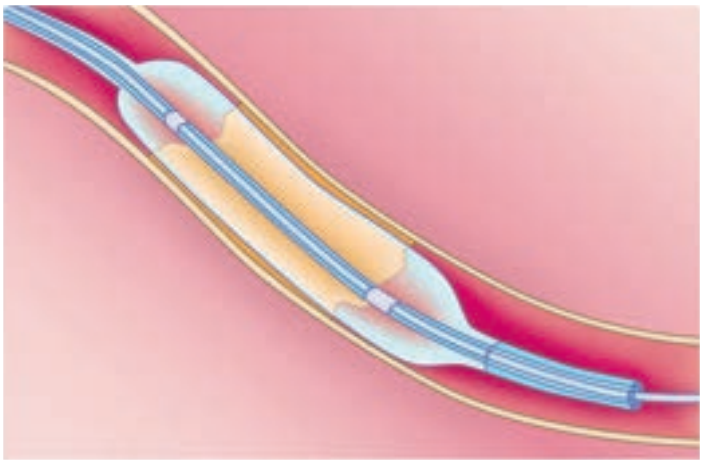

FIGURE 74-2. Balloon angioplasty catheter. The catheter consists of two lumens, an inflation lumen and a guidewire lumen. Two radiopaque markers, indicating the lateral balloon margins, aid in positioning of the balloon before inflation.

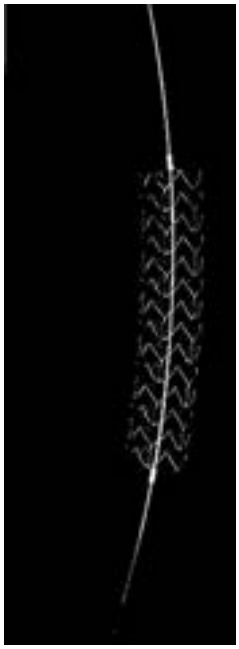

FIGURE 74-3. Balloon-expandable coronary stent. The stainless steel stent is crimped onto a balloon catheter to allow low-profile passage through the coronary artery. When it is positioned across the lesion, the balloon is inflated, expanding the stent. After balloon deflation and removal, the stent remains, providing a scaffold that supports the vessel lumen.

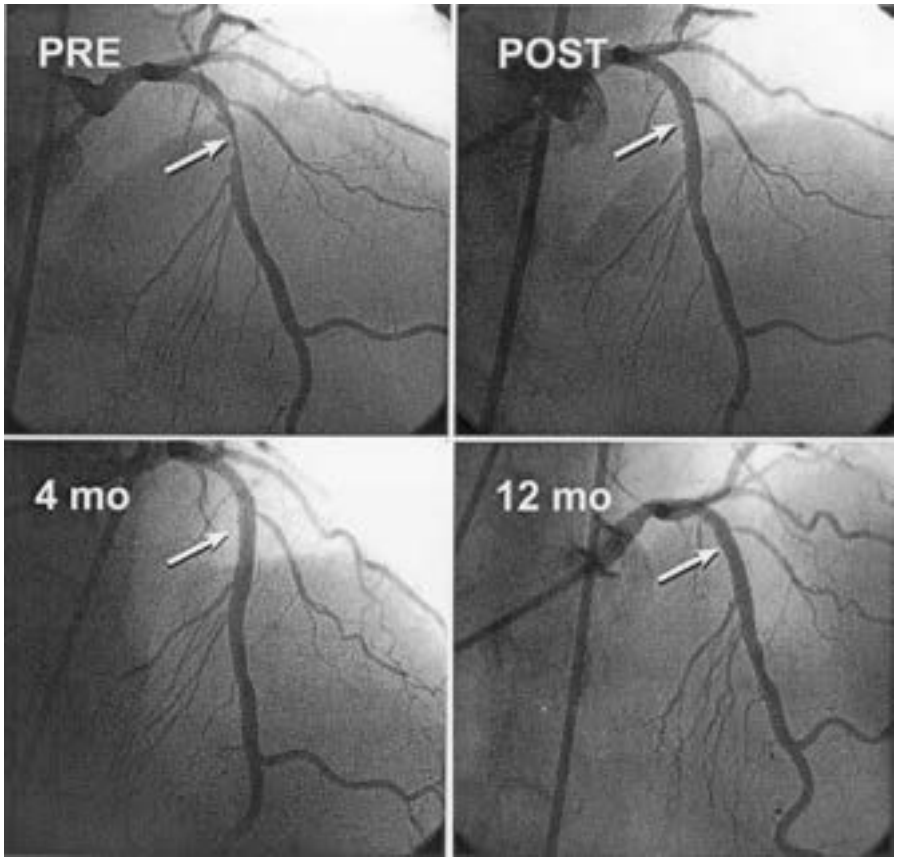

FIGURE 74-4. Angiographic images before, after, and at late follow-up after placement of a sirolimus-eluting stent. The left anterior descending artery contains a tight stenosis *(arrow, upper left panel)*. After stent implantation *(upper right panel)*, the stenosis is abolished *(arrow)*. Follow-up at 4 and 12 months *(bottom panels)* reveals a completely open lumen with no evidence of restenosis *(arrows)*.

VIDEO 74-1. Coronary stent placement.

VIDEO 74-2. Guidewire passage.

VIDEO 74-3. Delivering the stent.

VIDEO 74-4. Inflating the stent.

VIDEO 74-5. Final result.

Selection of Patients for Percutaneous Coronary Intervention

Any decision to perform PCI must include a review of the coronary angiogram (Chapter 57) by an experienced interventional cardiologist to assess the lesion's technical suitability for the procedure. The disease must narrow the coronary artery lumen by at least 60%, and the quantity of myocardium subtended by the vessel should not be trivial. The pressure drop across a stenosis, called the *fractional flow reserve* (FFR), during maximum hyperemia (usually induced by intravenous adenosine) reflects its hemodynamic severity. Whereas a pressure drop of more than 20% (which corresponds to a FFR <0.80) predicts a clinical benefit from PCI, an FFR greater than 0.80 has been correlated with clinical harm.■■ High-risk lesion characteristics, such as longer lesion length, vessel tortuosity, lesion calcification, or the presence of thrombus, must be taken into consideration. Subtle angiographic findings, such as the presence of collateral vessels that supply a different myocardial territory and that originate distal to the target, should be appreciated. For each patient, the benefits of PCI must be weighed against the procedural risk. Characteristics of the patient conveying increased risk include advanced age (i.e., >75 years), diabetes, smaller vessels that are often found in women, prior MI, significant impairment of left ventricular function, and renal insufficiency.

Procedural Success and Complications

With use of modern techniques in appropriately selected patients, most PCI procedures have a greater than 95% success rate. The single exception is a chronic total coronary occlusion (100% obstruction of the lumen), in which the interventional cardiologist's ability to negotiate a guidewire through the blockage is only about 50% to 90% and varies substantially with the operator's expertise. With the increased use of coronary stents and adjunctive antiplatelet agents (thienopyridines and platelet glycoprotein IIb/IIIa inhibitors), abrupt coronary artery closure is rarely encountered. When PCI is performed by an experienced interventional cardiologist in appropriately selected patients, the risk of in-hospital death is less than 1%; MI (usually small, non–ST segment elevation MI) is approximately 5%, the need for urgent or emergent CABG surgery is less than 1%, the risk of stroke is less than 0.1%, the chance of coronary perforation is less than 1%, and morbidity at the arterial access site (i.e., hematoma, pseudoaneurysm, or arteriovenous fistula) occurs in fewer than 5% of patients. As a result of these low complication rates, PCI without the availability of onsite surgery is not associated with higher mortality or rate of emergency CABG surgery.■■

Restenosis and Thrombosis

Restenosis is a renarrowing of an artery after a PCI procedure, usually resulting from one of two mechanisms. The first mechanism, unfavorable remodeling and elastic recoil, is a mechanical renarrowing caused by adventitial constriction and shrinkage of the vessel lumen. The second mechanism, neointimal hyperplasia, is caused by the proliferation of smooth muscle cells and matrix in response to the injury caused by balloons, stents, or atherectomy devices. Restenosis occurs in 10% to 50% of PCI patients after balloon angioplasty without stenting, usually within the first 6 months after the procedure. Characteristics associated with higher risk of restenosis include longer lesions, small-diameter vessels, diabetes, and multivessel disease. Treatment with either balloon angioplasty or atherectomy devices results in similar rates of restenosis. Coronary stents, which provide a semirigid scaffolding within the lumen and reduce restenosis by eliminating the mechanical renarrowing caused by unfavorable remodeling and elastic recoil, reduce restenosis by about one third compared with balloon angioplasty alone.[1]

Although bare metal stents eliminate the mechanical component of restenosis, the proliferative component is enhanced. Smooth muscle cell division and matrix formation can migrate through the stent struts to renarrow the vessel lumen. Drug-eluting stents, which contain and release antiproliferative drugs (e.g., sirolimus, everolimus, zotarolimus, and paclitaxel) reduce the need for early repeat PCI to less than 5%.■■ However, drug-eluting stents may carry a slightly higher risk of thrombosis later (>1 year), especially if patients cannot or do not continue aspirin plus a thienopyridine.

Choices Related to Stenting

Based on long-term follow-up data, the choice between a bare metal stent and a drug-eluting stent remains controversial. Patients receiving drug-eluting stents must continue dual antiplatelet therapy with aspirin and a thienopyridine for a minimum of 12 months. If a patient is unlikely to be able to adhere to such therapy because of bleeding risks or the need for an invasive or surgical procedure, a bare metal stent is preferable. Biodegradable scaffolds, which completely dissolve over 2 to 3 years,[2] are approved in some countries but not currently in the U.S. Atherectomy is rarely used, with the exception of the Rotablator (Boston Scientific, Maple Grove, MN), which pulverizes plaque into microparticles that pass through the coronary microcirculation and is particularly helpful for the treatment of heavily calcified lesions. After plaque debulking with the Rotablator, a drug-eluting stent is usually implanted.

Discharge Issues

Discharge planning after PCI represents an important opportunity to emphasize evidence-based medical treatment of atherothrombotic disease and coronary risk factor modification. All patients should receive aspirin (81-325 mg/day) indefinitely. For patients receiving bare metal stents, a minimum 2-week course of a thienopyridine (e.g., clopidogrel 75 mg/day, prasugrel 10 mg/day, or ticagrelor 90 mg/day) is mandatory. If a drug-eluting stent is deployed, this dual antiplatelet therapy generally should be extended for at least 12 months, but aspirin alone may be sufficient in some lower-risk patients.[3] Extending dual antiplatelet therapy to 36 months reduces cardiovascular events, increases bleeding, and probably has no benefit for overall survival.■ Prolonged use of aspirin, clopidogrel, angiotensin-converting enzyme inhibitors, β-blockers, and lipid-lowering agents should be considered on the basis of randomized trials showing improved long-term outcome, particularly in patients who present with unstable angina syndromes (Chapters 71, 72, and 73). Smoking cessation (Chapter 32), blood pressure control (Chapter 67), stress management, exercise, weight loss, changes in dietary habits, and strict blood glucose control for patients with diabetes (Chapter 229) also are important elements of the discharge plan.

Activity restrictions after PCI are modest. If the femoral artery was instrumented, heavy lifting is discouraged for several days. Intense aerobic exercise is usually discouraged for 2 to 4 weeks (especially after stent implantation) because exercise can activate platelets and lead to formation of thrombus at the angioplasty site. Patients may return to work 1 or 2 days after the procedure if their occupation does not include heavy lifting or excessive physical exercise. There is usually no restriction on driving an automobile. With modern management, mortality after stenting is more likely to be from noncardiac than from cardiac causes.[4]

CORONARY ANGIOPLASTY VERSUS MEDICAL THERAPY

Percutaneous coronary intervention reduces angina and commonly leads to better treadmill exercise performance and improved quality-of-life measurements. However, PCI has not been shown to reduce the risks of death, MI, or other major cardiovascular events compared with modern optimal medical therapy in patients with stable angina (Chapter 71).■ PCI reduces symptoms for the first 24 months, especially in patients with more severe angina, but symptoms were similarly improved over baseline with both intensive medical therapy and PCI at 36 months.■ In truly asymptomatic patients, significant ischemia first should be documented by functional testing, or a large quantity of myocardium should be supplied by the stenotic coronary artery. Patients experiencing acute ST segment elevation MI (Chapter 73) represent an important subgroup in whom PCI has proved beneficial compared with medical therapy (Fig. 74-5). In ST segment elevation MI, randomized trials consistently have reported a reduction in mortality, stroke, subsequent MI, and recurrent ischemia with immediate PCI compared with thrombolytic therapy■■ or with initial thrombolytic therapy followed by rescue PCI as needed■■ even if immediate PCI requires transfer to another hospital. For non–ST segment elevation MI and many patients with unstable angina, an early aggressive approach that includes either PCI or CABG in angiographically suitable patients is generally preferable to a conservative strategy■■■ except in lower risk patients (Chapter 72).

CORONARY ARTERY SURGERY

Coronary artery bypass graft surgery is based on the premise that the morbidity and mortality associated with coronary atherosclerosis are largely related to atherosclerotic coronary stenoses that can be demonstrated by coronary angiography (Chapter 57) and that if grafts are constructed to route blood flow around these stenoses, myocardial blood supply can be improved or preserved, cardiac symptoms relieved, cardiac events diminished, and survival prolonged. Over time, the fundamentals of that concept have been shown to be correct.

The most common types of grafts for coronary artery bypass have been reversed segments of saphenous vein and the internal thoracic arteries. Saphenous vein grafts are anastomosed to the aorta (proximal anastomosis) and to the coronary artery distal to the major obstruction (Fig. 74-6). Saphenous vein grafts have the advantages of availability, larger size than most coronary arteries, and favorable handling characteristics. With time,

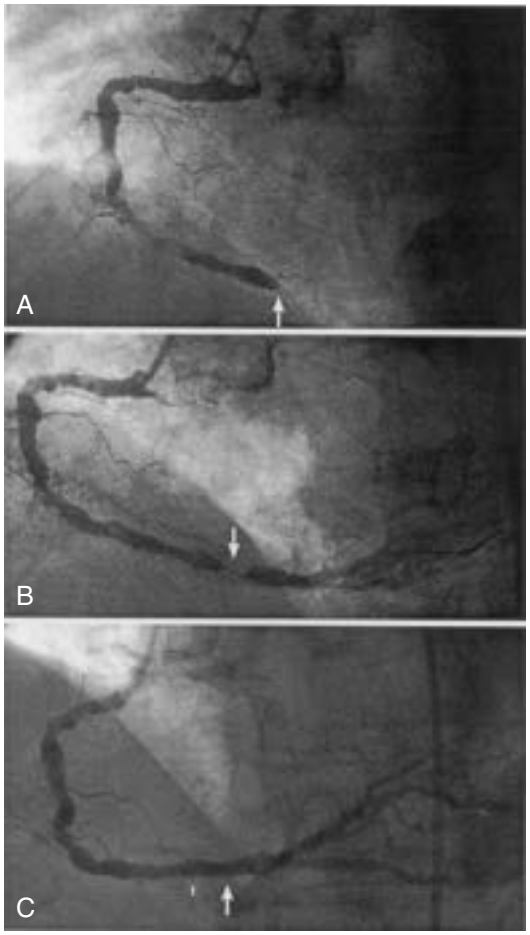

FIGURE 74-5. Primary coronary angioplasty for acute myocardial infarction. This 50-year-old man presented at midnight with 70 minutes of crushing substernal chest pressure accompanied by inferior ST segment elevation. Emergency angiography performed 45 minutes after arrival found 100% occlusion of the right coronary artery (**A**, *arrow*). Within 10 minutes, a guidewire was negotiated through the obstruction (presumably caused by fresh thrombus), allowing perfusion into the distal vessel and uncovering a high-grade stenotic lesion (**B**, *arrow*). After deployment of a coronary stent (**C**, *arrow*), the stenosis was abolished and significant myocardial damage aborted.

however, saphenous vein grafts may develop intrinsic pathologic changes, intimal fibroplasia, and vein graft atherosclerosis, each of which may lead to narrowing or occlusions. Modern treatment with platelet inhibitors and statins (Chapter 206) decreases the risk of vein graft failure but cannot eliminate it. Internal thoracic artery grafts, on the other hand, are resistant to the development of late atherosclerosis. When used as an in situ (subclavian origin intact) graft to the left anterior descending (LAD) coronary artery, the left internal thoracic artery graft has a more than 90% patency rate up to 20 years after operation. As a result, patients who receive a left internal thoracic artery to LAD graft, with or without saphenous vein grafts, have a better long-term survival rate, fewer reoperations, and fewer cardiac events compared with patients receiving only saphenous vein grafts. The right internal thoracic artery may also be used for revascularization as an in situ graft, as an aorta to coronary artery graft, or as a composite arterial graft from the left internal thoracic artery to a coronary artery. Use of both internal thoracic arteries as grafts provides incremental benefit over a single internal thoracic artery graft strategy and produces an improved survival with a lower risk for reoperation.[5] Although the experience is more limited, radial artery grafts also appear to function better than saphenous vein grafts.

Most CABG operations are performed with a full median sternotomy incision, historically with the aid of cardiopulmonary bypass, aortic cross-clamping, and cardioplegic solution—techniques that allow exposure and arrest of the heart such that detailed microsurgical anastomoses can be constructed while myocardial function is effectively protected. By comparison, operations performed through smaller incisions (minimally invasive surgery) have had limited application for coronary revascularizations. In randomized trials, beating heart ("off-pump") surgery without cardiopulmonary bypass has not consistently reduced morbidity and has often been associated with a lower rate of long-term graft patency; as a result, it is usually reserved for carefully selected patients.

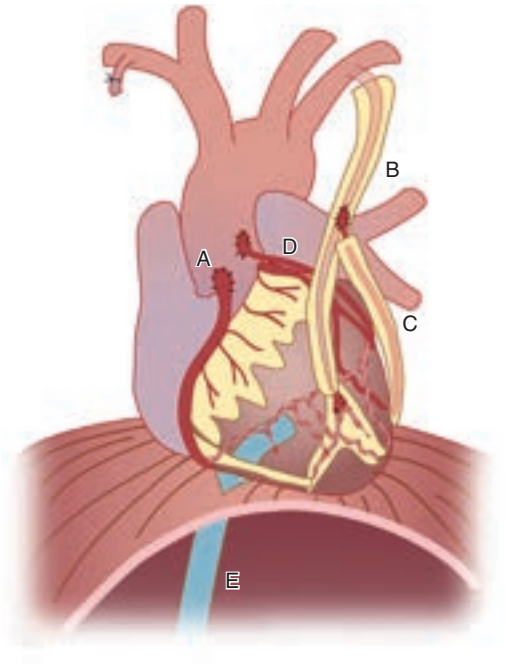

FIGURE 74-6. Types of bypass grafts. Bypass grafts include reversed saphenous vein graft from the aorta to the right coronary artery (**A**), in situ left internal mammary artery graft to the anterior descending coronary artery (**B**), Y graft of the right internal mammary artery from the left internal mammary artery to the circumflex coronary artery (**C**), radial artery graft from the aorta to the circumflex coronary artery (**D**), and in situ gastroepiploic graft to the posterior descending branch of the right coronary artery (**E**).

Perioperative Risks

The risk of mortality associated with CABG correlates with ischemia at the time of operation, left ventricular function, extent of coronary stenoses, noncardiac atherosclerosis, and comorbid conditions, as well as with the experience, skill, and judgment of the surgeon. Effective myocardial protection has diminished much of the incremental risk based on the severity of cardiac disease. For patients younger than 70 years without serious comorbid conditions, the mortality risk of primary CABG surgery is less than 1% in experienced hands. Nevertheless, CABG surgery in the presence of ongoing myocardial ischemia caused by acute MI, unstable angina, or acute vessel closure after PCI is still associated with increased risk. Noncardiac comorbid conditions (aortic atherosclerosis, renal function, chronic obstructive pulmonary disease, and coagulation system disorders) increase perioperative risk when these conditions are severe.

The most serious postoperative morbidity after CABG is stroke, often related to aortic or cerebrovascular atherosclerosis and atherosclerotic embolization. Heightened awareness of the importance of aortic and carotid atherosclerosis and improved management strategies appear to have decreased the risk of focal stroke in patients previously at high risk. Serious wound complications of median sternotomy are uncommon (1%-2%).

Late Outcomes

The late outcomes after CABG are related to age, severity of cardiac disease before operation, noncardiac comorbid conditions, progression of atherosclerosis, and the operation itself. CABG tends to diminish but not eliminate long-term survival differences based on the number of diseased coronary vessels, left main stenosis, and left ventricular function. The achievement of complete revascularization (bypass grafts to all stenotic coronary vessels) and the use of internal thoracic artery grafts improve long-term survival rates and symptomatic status.

For subgroups of patients with severe coronary artery disease, CABG prolongs life expectancy (see later discussion). More than 80% of patients are alive more than 10 years after operation. Over the long term, control of the progression of atherosclerosis by lifestyle modifications, pharmacologic treatment of hypertension (Chapter 67) and lipids (Chapter 206), and platelet inhibitors (Chapter 38) appear to extend the benefits of CABG.

INDICATIONS FOR BYPASS SURGERY

The goals of CABG are to relieve symptoms and to prolong life expectancy. On the basis of randomized trials and the emergence of alternative medical

treatments and PCI, the surgical population has evolved toward patients who benefit most from CABG relative to other treatments: those with complex conditions, often involving left main or triple-vessel disease, diffuse coronary stenoses, totally obstructed vessels, abnormal left ventricular function, and diabetes. Surgically treated patients with single-vessel disease usually have LAD stenoses or have failed alternative treatments.

Symptom Relief

If patients who experience angina have severe stenoses in graftable coronary arteries that supply areas of myocardium ischemic at rest or with stress, CABG will reliably relieve angina. Randomized trials have shown that the relief of angina after CABG is more consistent than that achieved with alternative treatments. When intermittent heart failure symptoms represent an "anginal equivalent" that is also caused by ischemia, such symptoms also respond well to relief of that ischemia by CABG. For patients with symptoms of heart failure at rest, dobutamine echocardiography (Chapter 55) and positron emission tomography (Chapter 56) can identify segments of viable but hibernating myocardium (ischemic at rest) that may improve with bypass grafting, thus reducing symptoms of heart failure.

Survival

Chronic Stable Angina

In randomized trials of patients with mild to moderate chronic stable angina, an improved survival rate has been documented for patients treated with initial CABG compared with those treated with initial medical treatment in the presence of a left main stenosis of more than 50% of the diameter, triple-vessel disease, double-vessel disease with a proximal LAD lesion, abnormal left ventricular function, or a strongly positive exercise test result (Chapter 71). Meta-analysis of these randomized trials also suggests a survival benefit of CABG for any patient with a proximal LAD lesion and myocardial ischemia. These are subgroups of patients for whom bypass surgery should be strongly considered even in the absence of severe symptoms. During these trials, patients with severe angina were not randomized but were included in observational studies that noted improved survival rates with CABG for patients with double- and triple-vessel disease and normal or abnormal left ventricular function. Medical, interventional, and surgical treatments have all advanced substantially since these trials were completed.

Unstable Angina or Non–ST Segment Elevation Myocardial Infarction

Current data suggest an aggressive strategy, including CABG when indicated, in patients with unstable angina or non–ST segment elevation acute MI (Chapter 72).

Ischemic Syndromes without Randomized Trials

ST Segment Elevation Acute Myocardial Infarction

For patients with ST segment elevation acute MI, CABG may be indicated in the acute setting when thrombolytic therapy or PCI has not been effective, ischemia is ongoing, and large areas of myocardium remain jeopardized. CABG after a completed MI may be indicated in patients in whom persistent ischemia in noninfarcted areas of myocardium produces postinfarction angina or hemodynamic instability. Mechanical complications of myocardial necrosis, including papillary muscle rupture, ventricular septal rupture, and myocardial free wall rupture, are acute life-threatening situations that usually require urgent operation for repair of the defect, often combined with CABG (Chapter 73).

Failed Percutaneous Coronary Intervention

The availability of intracoronary stents has decreased the need for emergency CABG to treat acute failure of PCIs. Current indications for emergency CABG include closure or threatened closure of a vessel supplying a significant amount of myocardium.

Coronary Bypass Reoperations

Patients in whom new stenoses develop in native arteries or in bypass grafts may have recurrent ischemic syndromes. Severe vein graft atherosclerosis is an unstable lesion that often leads to serious cardiac events, particularly if the LAD or multiple vessels are jeopardized; reoperation appears to improve the survival rate of these patients. Conversely, if the LAD coronary artery is supplied by a patent internal thoracic artery graft, reoperation does not appear to improve survival, although it may improve symptoms. Reoperations are more difficult and dangerous than primary procedures, but the risk now approaches that for primary procedures in institutions performing a large number of reoperations. PCI is sometimes an alternative for the treatment of vein graft disease.

Coexisting Cardiac Disease

During cardiac operations performed for valvular (Chapter 75) or aortic (Chapter 78) disease, the standard treatment is to perform bypass grafts to major coronary arteries with angiographic stenoses of more than 50% of the luminal diameter. No randomized trials have addressed this issue, and these indications, although logical given the natural history of atherosclerosis, remain practice patterns based on consensus but not on definitive data.

PERCUTANEOUS CORONARY INTERVENTION VERSUS CORONARY ARTERY BYPASS GRAFTING

The decision between PCI and CABG surgery is largely determined by clinical status and anatomic features, but some areas remain controversial. For patients with acute coronary syndromes, PCI is the preferred initial approach and is known to improve the survival rate of patients with ST segment elevation MI. CABG is reserved for patients with failed acute PCI or residual myocardial ischemia. For patients with chronic coronary syndromes, there are currently no data to confirm that PCI prolongs life regardless of anatomy, but randomized trials, albeit older trials, demonstrate that CABG prolongs the life expectancy of patients with severe coronary artery disease, particularly patients with left ventricular dysfunction or severe ischemia. Surgery, therefore, is often the initial approach to these patient subsets.

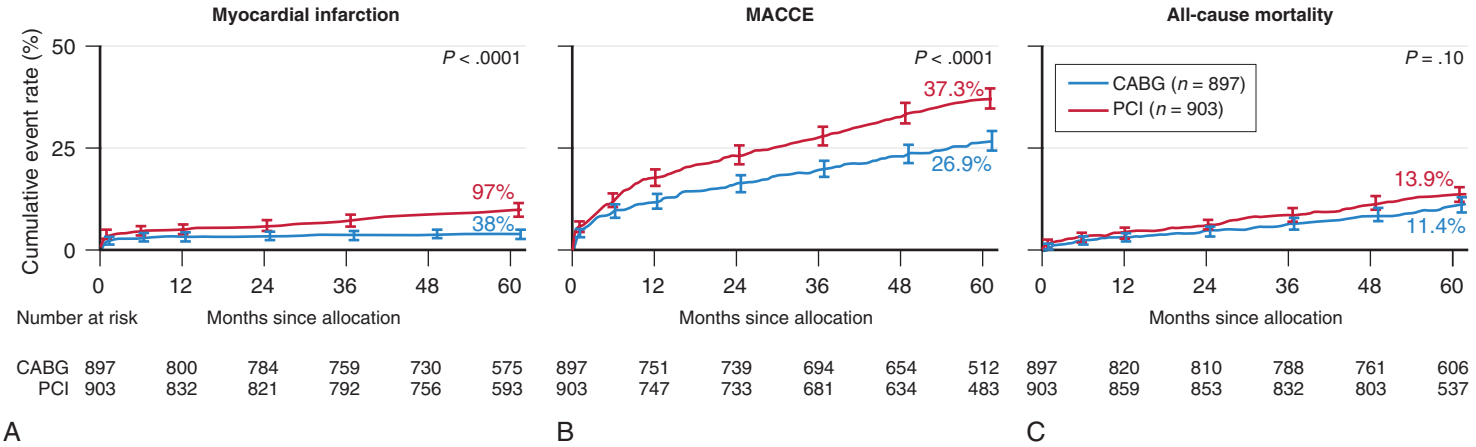

FIGURE 74-7. **The SYNTAX trial randomized patients with left main vessel or three-vessel coronary artery disease to coronary artery bypass grafting (CABG) versus percutaneous coronary intervention (PCI). At 5 years, patients randomized to CABG surgery had fewer myocardial infarctions, fewer major acute coronary or cerebrovascular events (MACCE = death, myocardial infarction, stroke, and repeat revascularization), and a trend toward lower mortality. (Data from More F, Morice M, Serruys P, et al. Coronary artery bypass graft surgery versus percutaneous coronary intervention with three vessel disease and left main coronary disease: 5-year follow-up of the randomized, clinical SYNTAX Trial. *Lancet*. 2013;381:629-638.)**

Randomized trials have provided information aiding in the selection of therapy. The SYNTAX trial randomized patients with triple vessel or left main disease to PCI or CABG and grouped patients based on scores reflecting the anatomic complexity and extent of coronary disease into low-, medium-, and high-risk subgroups. At 5-year follow-up, the entire group of patients randomized to bypass surgery had fewer MIs (9.7% vs. 3.8%; $P < .001$), fewer repeat revascularization (25.9% vs. 13.7%; $P < .001$), and a trend toward lower mortality rates (13.9% vs. 11.4%; $P = 0.10$) (Fig. 74-7). Patients with more complex coronary artery disease benefitted most from CABG, including a clear improvement in survival rate, but patients with more limited coronary disease had equivalent survival outcomes with PCI. Randomized trials also show that diabetic patients with multivessel coronary artery disease have better survival rates with CABG compared with PCI.

Grade A References

A1. De Bruyne B, Pijls N, Fearon W, et al. Fractional flow reserve-guided PCI versus medical therapy in stable coronary artery disease. *N Engl J Med*. 2012;367:991-1001.

A2. De Bruyne B, Fearon WF, Pijls NH, et al. Fractional flow reserve-guided PCI for stable coronary artery disease. *N Engl J Med*. 2014;371:1208-1217.

A3. Jacobs AK, Normand SL, Massaro JM, et al. Nonemergency PCI at hospitals with or without on-site cardiac surgery. *N Engl J Med*. 2013;368:1498-1508.

A4. Aversano T, Lemmon CC, Liu L, Atlantic CPORT Investigators. Outcomes of PCI at hospitals with or without on-site cardiac surgery. *N Engl J Med*. 2012;366:1792-1802.

A5. Palmerini T, Biondi-Zoccai G, Della Riva D, et al. Clinical outcomes with drug-eluting and bare-metal stents in patients with ST-segment elevation myocardial infarction: evidence from a comprehensive network meta-analysis. *J Am Coll Cardiol*. 2013;62:496-504.

A6. Bangalore S, Kumar S, Fusaro M, et al. Short- and long-term outcomes with drug-eluting and bare-metal coronary stents: a mixed-treatment comparison analysis of 117 762 patient-years of follow-up from randomized trials. *Circulation*. 2012;125:2873-2891.

A7. Mauri L, Kereiakes DJ, Yeh RW, et al. Twelve or 30 months of dual antiplatelet therapy after drug-eluting stents. *N Engl J Med*. 2014;371:2155-2166.

A8. Stergiopoulos K, Boden WE, Hartigan P, et al. Percutaneous coronary intervention outcomes in patients with stable obstructive coronary artery disease and myocardial ischemia: a collaborative meta-analysis of contemporary randomized clinical trials. *JAMA Intern Med*. 2014;174:232-240.

A9. Weintraub WS, Spertus JA, Kolm P, et al. Effect of PCI on quality of life in patients with stable coronary artery disease. *N Engl J Med*. 2008;359:677-687.

A10. Keeley EC, Boura JA, Grines CL. Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials. *Lancet*. 2003;361:13-20.

A11. Di Mario C, Dudek D, Piscione F, et al. Immediate angioplasty versus standard therapy with rescue angioplasty after thrombolysis in the Combined Abciximab REteplase Stent Study in Acute Myocardial Infarction (CARESS-in-AMI): an open, prospective, randomised, multicentre trial. *Lancet*. 2008;371:559-568.

A12. Mehta SR, Granger CB, Boden WE, et al., for the TIMACS Investigators. Early versus delayed invasive intervention in acute coronary syndromes. *N Engl J Med*. 2009;360:2165-2175.

A13. O'Donoghue M, Boden WE, Braunwald E, et al. Early invasive vs. conservative treatment strategies in women and men with unstable angina and non-ST-segment elevation myocardial infarction: a meta-analysis. *JAMA*. 2008;300:71-80.

A14. Deb S, Cohen EA, Singh SK, et al. Radial artery and saphenous vein patency more than 5 years after coronary artery bypass surgery: results from RAPS (Radial Artery Patency Study). *J Am Coll Cardiol*. 2012;60:28-35.

A15. Houlind K, Kjeldsen BJ, Madsen SN, et al. On-pump versus off-pump coronary artery bypass surgery in elderly patients: results from the Danish On-Pump Versus Off-Pump Randomization Study. *Circulation*. 2012;125:2431-2439.

A16. Frye RL, August P, Brooks MM, et al. A randomized trial of therapies for type 2 diabetes and coronary artery disease. *N Engl J Med*. 2009;360:2503-2515.

A17. Mohr F, Morice M, Serruys P, et al. Coronary artery bypass graft surgery versus percutaneous coronary intervention with three vessel disease and left main coronary disease: 5-year follow-up of the randomized, clinical SYNTAX Trial. *Lancet*. 2013;381:629-638.

A18. Deb S, Wijeysundera HC, Ko DT, et al. Coronary artery bypass graft surgery vs percutaneous interventions in coronary revascularization: a systematic review. *JAMA*. 2013;310:2086-2095.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Stefanini GG, Holmes DR Jr. Drug-eluting coronary-artery stents. *N Engl J Med*. 2013;368:254-265.
2. Kang SH, Park KW, Kang DY, et al. Biodegradable-polymer drug-eluting stents vs. bare metal stents vs. durable-polymer drug-eluting stents: a systematic review and Bayesian approach network meta-analysis. *Eur Heart J*. 2014;35:1147-1158.
3. Feres F, Costa RA, Abizaid A, et al. Three vs twelve months of dual antiplatelet therapy after zotarolimus-eluting stents: the OPTIMIZE randomized trial. *JAMA*. 2013;310:2510-2522.
4. Spoon DB, Psaltis PJ, Singh M, et al. Trends in cause of death after percutaneous coronary intervention. *Circulation*. 2014;129:1286-1294.
5. Vallely MP, Edelman JJ, Wilson MK. Bilateral internal mammary arteries: evidence and technical considerations. *Ann Cardiothorac Surg*. 2013;2:570-577.

REVIEW QUESTIONS

1. Bare metal coronary stents reduce restenosis by

A. providing a semirigid scaffold that resists vessel recoil and remodeling.
B. inhibiting intimal proliferation.
C. reducing thrombus formation.
D. increasing blood flow within the coronary artery.
E. reducing vessel spasm.

Answer: A Bare metal stents use mechanical force to hold a vessel open. They do not reduce vessel spasm or thrombus formation. Stents increase intimal proliferation. Although stents increase vessel blood flow, the increased blood flow does not influence restenosis rates.

2. Compared with medical therapy, percutaneous coronary intervention

A. reduces the mortality rate in patients with stable angina.
B. reduces costs associated with medical care.
C. reduces the mortality rate in patients experiencing an acute myocardial infarction.
D. decreases myocardial infarction in patients with stable coronary artery disease.
E. reduces hospitalization rates.

Answer: C In randomized trials, coronary stents have been shown to reduce mortality rates compared with medical therapy alone in patients experiencing an acute myocardial infarction (MI). In patients with stable coronary disease, stents reduce angina but do not reduce death, MI, hospitalization rates, or costs.

3. Randomized trials comparing bypass surgery to coronary stents

A. found an increased late stroke rate in patients receiving bypass surgery.
B. found less angina in patients who underwent bypass surgery.
C. found higher MI rates in patients undergoing stenting.
D. demonstrated improved late outcomes in patients with diabetes undergoing stenting.
E. demonstrate improved survival in patients with extensive complex coronary disease

Answer: E In randomized trials, bypass surgery has been shown to reduce mortality rates compared with coronary stenting in patients with three-vessel, complex coronary disease. Stroke rates are higher in bypass patients early after surgery, but stroke rates are similar at later follow-up. Rates of MI and angina are equivalent on long-term follow-up. Patients with diabetes who have multivessel disease have better long-term outcomes with bypass surgery.

4. What is an important component of the coronary bypass surgery operation for increasing life expectancy?

A. Saphenous vein graft to the left anterior descending (LAD) coronary artery
B. Radial graft to the right coronary artery
C. Minimally invasive surgery
D. Left internal mammary artery graft to the LAD coronary artery
E. The right gastroepiploic graft to the right coronary artery

Answer: D Although no randomized trials have examined the impact of the left internal mammary artery to LAD graft, many observational comparative studies have established the long-term survival benefit of this surgical strategy.

5. What is the most common indication for recommending bypass surgery for patients with stable angina in preference to other treatment options is?

A. To prevent myocardial infarction
B. To increase exercise capacity
C. To improve life expectancy
D. To decrease the side effects of optimal medical therapy

Answer: C Advances in medical therapy and in percutaneous interventions have commonly been successful in controlling symptoms of angina, so coronary bypass surgery is rarely needed to treat symptoms that are unresponsive to alternative management strategies. However, for patients with severe coronary artery disease, bypass surgery is the form of therapy that has been most clearly shown to improve life expectancy. The most common indications for CABG are in anatomic and clinical situations in which it offers a survival benefit.

75

VALVULAR HEART DISEASE

BLASE A. CARABELLO

The cardiac valves permit unobstructed forward blood flow through the heart when they are open while preventing backward flow when they are closed. Most valvular heart diseases cause either valvular stenosis with obstruction to forward flow or valvular regurgitation with backward flow. Valvular stenosis imparts a pressure overload on the left or right ventricle (RV) because these chambers must generate higher than normal pressure to overcome the obstruction to pump blood forward. Valvular regurgitation imparts a volume overload on the heart, which now must pump additional volume to compensate for what is regurgitated. When valve disease is severe, these hemodynamic burdens can lead to ventricular dysfunction, heart failure, and sudden death (Table 75-1). In almost every instance, definitive therapy for severe valvular heart disease is mechanical restoration of valve function.

AORTIC STENOSIS

EPIDEMIOLOGY

Bicuspid and Other Congenitally Abnormal Aortic Valves

Approximately 1% of the population is born with a bicuspid aortic valve, with a male preponderance (Chapter 69). Although this abnormality does not usually cause a hemodynamic disturbance at birth, bicuspid aortic valves tend to deteriorate with age. Approximately one third of these valves become stenotic, another third become regurgitant, and the remainder cause only minor hemodynamic abnormalities. When stenosis develops, it usually occurs when patients are in their 40s, 50s, and 60s.

Sometimes congenital aortic stenosis from a unicuspid, bicuspid, or even abnormal tricuspid valve causes symptoms during childhood and requires correction by adolescence. Occasionally, these congenitally stenotic aortic valves escape detection until adulthood.

Tricuspid Aortic Valve Stenosis

In some patients born with apparently normal tricuspid aortic valves, thickening and calcification develop similar to what occurs in bicuspid valves. When aortic stenosis develops in previously normal tricuspid aortic valves, it usually does so in the 60s to 80s. Although stenosis and calcifications of bicuspid and tricuspid aortic valves were formerly considered to be degenerative processes, it is clear that this type of aortic stenosis arises from an active inflammatory process similar to that of coronary heart disease. This concept is supported by many pieces of evidence. First, the initial lesion of aortic stenosis is similar to the plaque of coronary disease. Second, both diseases have hypertension and hyperlipidemia, including elevated levels of lipoprotein(a), as risk factors. Third, there is excellent correlation between calcification of the aortic valve and calcification of the coronary arteries. Fourth, patients with the most severe aortic stenosis have the highest levels of C-reactive protein. However, many diseased aortic valves demonstrate actual bone formation, not merely calcification. Future targets for controlling or preventing the disease likely will aim at bone-forming pathways.

Rheumatic Valvular Heart Disease

Rheumatic valve disease is now a rare cause of aortic stenosis in developed countries, but rheumatic fever and its sequelae remain an important issue in many developing countries (Chapter 290). In virtually every case, the mitral valve is also detectably abnormal.

PATHOBIOLOGY

The normal aortic valve area is 3 to 4 cm^2, and little hemodynamic disturbance occurs until the orifice is reduced to about one third of normal, at which point a systolic gradient develops between the left ventricle (LV) and aorta.[1] LV and aortic pressures are normally nearly equal during systole. In aortic stenosis, intracavitary LV pressure must increase above aortic pressure, however, to produce forward flow across the stenotic valve and to achieve acceptable downstream pressure (see Fig. 57-2 in Chapter 57). Each of the processes that cause aortic stenosis has a typical pathoanatomic configuration (Fig. 75-1). There is a geometric progression in the magnitude of the gradient as the valve area narrows. Given a normal cardiac output, the gradient rises rapidly from 10 to 15 mm Hg at valve areas of 1.5 to 1.3 cm^2 to about 25 mm Hg at 1.0 cm^2, 50 mm Hg at 0.8 cm^2, 70 mm Hg at 0.6 cm^2, and 100 mm Hg at 0.5 cm^2. The rate of progression of aortic stenosis varies widely from patient to patient; it may remain stable for many years or increase by more than 15 mm Hg per year. A major compensatory response to the increased LV pressure associated with aortic stenosis is the development of concentric LV hypertrophy (LVH). The Laplace equation—tress (s) = Pressure (p) × Radius (r)/2 Thickness (th)—indicates that the systolic force on any unit of LV myocardium (afterload) varies directly with ventricular pressure and radius and inversely with wall thickness. As pressure increases, it can be offset by increased LV wall thickness (concentric hypertrophy). The

TABLE 75-1 SUMMARY OF SEVERE VALVAR HEART DISEASE

	AORTIC STENOSIS	MITRAL STENOSIS	MITRAL REGURGITATION	AORTIC REGURGITATION
Etiology	Idiopathic calcification of a bicuspid or tricuspid valve Congenital Rheumatic	Rheumatic fever Annular calcification	Mitral valve prolapse Ruptured chordae Endocarditis Ischemic papillary muscle dysfunction or rupture Collagen vascular diseases and syndromes Secondary to LV myocardial diseases	Annuloaortic ectasia Hypertension Endocarditis Marfan syndrome Ankylosing spondylitis Aortic dissection Syphilis Collagen vascular disease
Pathophysiology	Pressure overload on the LV with compensation by LVH As disease advances, reduced coronary flow reserve causes angina. Hypertrophy and afterload excess lead to systolic and diastolic LV dysfunction.	Obstruction to LV inflow increases LA pressure and limits cardiac output, thus mimicking LV failure. Mitral valve obstruction increases the pressure work of the RV Right ventricular pressure overload is augmented further when pulmonary hypertension develops.	Places volume overload on the LV. Ventricle responds with eccentric hypertrophy and dilation, which allow increased ventricular stroke volume. Eventually, however, LV dysfunction develops if volume overload is uncorrected.	*Chronic:* Total stroke volume causes hyperdynamic circulation, induces systolic hypertension, and causes pressure and volume overload. Compensation is by concentric and eccentric hypertrophy. *Acute:* Because cardiac dilation has not developed, hyperdynamic findings are absent. High diastolic LV pressure causes mitral valve preclosure and potentiates LV ischemia and failure.
Symptoms	Angina Syncope Heart failure	Dyspnea Orthopnea PND Hemoptysis Hoarseness Edema Ascites	Dyspnea Orthopnea PND	Dyspnea Orthopnea PND Angina Syncope
Signs	Systolic ejection murmur radiating to the neck Delayed carotid upstroke S_4, soft or paradoxical S_2	Diastolic rumble after an opening snap Loud S_1 RV lift Loud P_2	Holosystolic apical murmur radiating to the axilla, S_3 Displaced PMI	*Chronic:* Diastolic blowing murmur Hyperdynamic circulation Displaced PMI Quincke pulse de Musset sign Austin Flint murmur *Acute:* Short diastolic blowing murmur Soft S_1
Electrocardiogram	LAA LVH	LAA RVH	LAA LVH	LAA LVH
Chest radiograph	Boot-shaped heart Aortic valve calcification on lateral view	Straightening of left heart border Double density at right heart border Kerley B lines Enlarged pulmonary arteries	Cardiac enlargement	*Chronic:* Cardiac enlargement Uncoiling of the aorta *Acute:* Pulmonary congestion with normal heart size
Echocardiographic findings	Concentric LVH Reduced aortic valve cusp separation Doppler shows mean gradient ≥40 mm Hg in most severe cases	Restricted mitral leaflet motion Valve area ≤1.5 cm^2 in most severe cases Tricuspid Doppler may reveal pulmonary hypertension	LV and LAA in chronic severe disease Doppler: large regurgitant jet	*Chronic:* LV enlargement Large Doppler jet PHT <400 msec *Acute:* Small LV Mitral valve preclosure
Catheterization findings	Increased LVEDP Transaortic gradient 40 mm Hg AVA ≤0.8 in most severe cases	Elevated pulmonary capillary wedge pressure Transmitral gradient usually >5 mm Hg in severe cases MVA <1.5 cm^2	Elevated pulmonary capillary wedge pressure Ventriculography shows regurgitation of dye into LV	Wide pulse pressure Aortography shows regurgitation of dye into LV Usually unnecessary
Medical therapy	Avoid vasodilators Digitalis, diuretics, and nitroglycerin in inoperable cases	Diuretics for mild symptoms Anticoagulation in atrial fibrillation Digitalis, β-blockers, verapamil, or diltiazem for rate control	Vasodilators in acute disease No proven therapy in chronic disease	*Chronic:* Vasodilators in chronic asymptomatic disease with hypertension even if LV function is normal. *Acute:* Vasodilators
Indications for surgery	Appearance of symptoms in patients with severe disease (see text)	Appearance of more than mild symptoms Development of pulmonary hypertension Appearance of persistent atrial fibrillation	Appearance of symptoms EF <0.60 ESD ≥40 mm	*Chronic:* Appearance of symptoms EF <0.50 ESD ≥50 mm *Acute:* Even mild heart failure Mitral valve preclosure

AVA = aortic valve area; EF = ejection fraction; ESD = end-systolic diameter; LA = left atrium; LAA = left atrial enlargement; LV = left ventricle; LVEDP = left ventricular end-diastolic pressure; LVH = left ventricular hypertrophy; MVA = mitral valve area; PHT = pressure half-time; PMI = point of maximal impulse; PND = paroxysmal nocturnal dyspnea; RV = right ventricle; RVH = right ventricular hypertrophy.

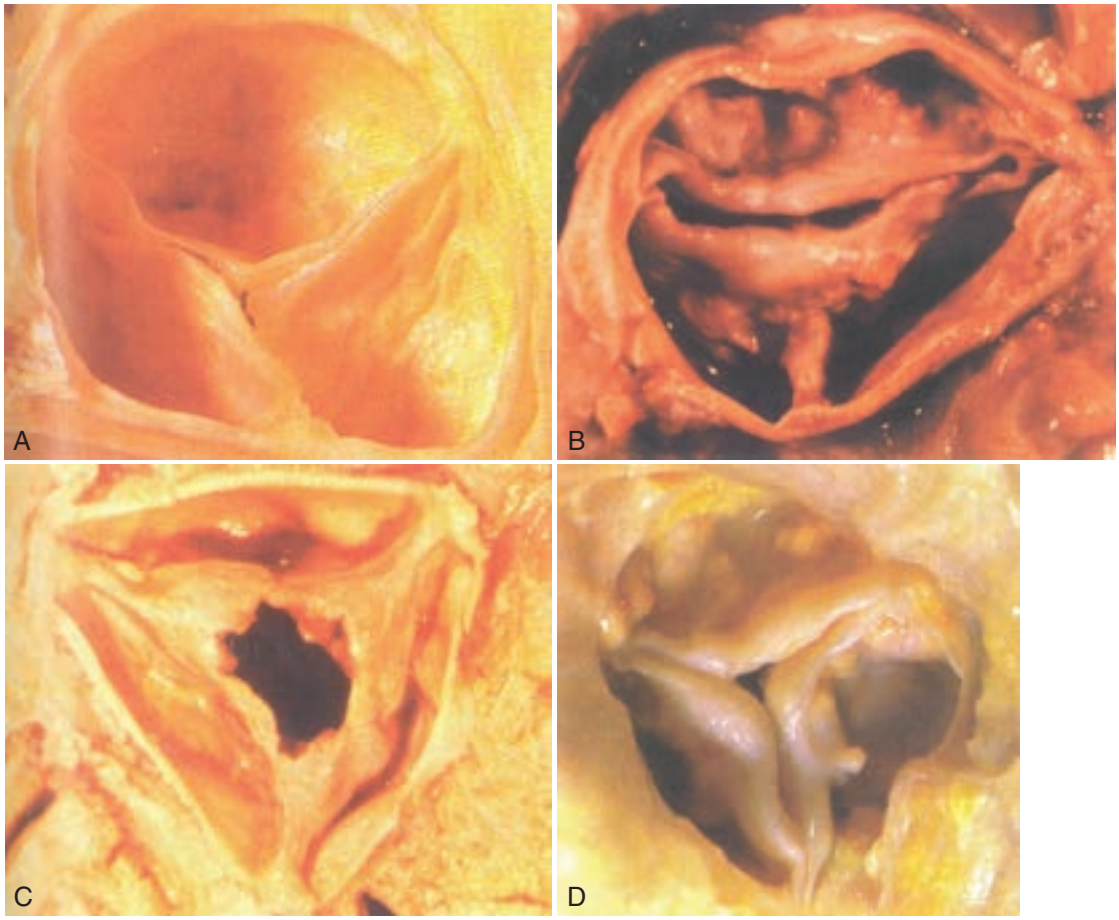

FIGURE 75-1. Pathology of aortic stenosis. A, A normal aortic valve. **B,** A stenotic congenital bicuspid valve. **C,** Rheumatic aortic stenosis. **D,** A stenotic calcified tricuspid aortic valve. (From Bonow RO, Braunwald E. Valvular heart disease. In: Zipes DP, Libby P, Bonow RO, Braunwald E, eds. *Heart Disease: A Textbook of Cardiovascular Medicine.* 7th ed. Philadelphia: WB Saunders; 2005:1583.)

determinants of LV ejection fraction are contractility, preload, and afterload. By normalizing afterload, the development of concentric hypertrophy helps preserve ejection fraction and cardiac output despite the pressure overload. Although hypertrophy clearly serves a compensatory function, it also has a pathologic role and is in part responsible for the classic symptoms and the poor outcome of untreated symptomatic aortic stenosis.

Angina

In general, angina (Chapter 71) results from myocardial ischemia when LV oxygen (and other nutrient) demand exceeds supply, which is predicated on coronary blood flow. In normal subjects, coronary blood flow can increase five- to eightfold under maximum metabolic demand, but in patients with aortic stenosis, this reserve is limited. Reduced coronary blood flow reserve may be caused by a relative diminution in capillary ingrowth to serve the needs of the hypertrophied LV or by a reduced transcoronary gradient for coronary blood flow because of the elevated LV end-diastolic pressure. Restricted coronary blood flow reserve is in part responsible for angina in many patients who have aortic stenosis despite normal epicardial coronary arteries, but many patients with restricted flow do not develop angina. In other patients, angina is caused by increased oxygen demand when inadequate hypertrophy allows wall stress, a key determinant of myocardial oxygen consumption, to increase.

Syncope

Syncope (Chapters 51 and 62) generally occurs because of inadequate cerebral perfusion. In aortic stenosis, syncope is usually related to exertion. It may result when exertion causes a fall in total peripheral resistance that cannot be compensated for by increased cardiac output because output is limited by the obstruction to LV outflow; this combination reduces systemic blood pressure and cerebral perfusion. In addition, high LV pressure during exercise may trigger a systemic vasodepressor response that lowers blood pressure and produces syncope. Cardiac arrhythmias, possibly caused by exertional ischemia, also cause hypotension and syncope.

Heart Failure

In aortic stenosis, contractile dysfunction (systolic failure) and failure of normal relaxation (diastolic failure) occur and cause symptoms (Chapter 58). The extent of ventricular contraction is governed by contractility and afterload. In aortic stenosis, contractility (the ability to generate force) is often reduced. The mechanisms of contractile dysfunction may include abnormal calcium handling, microtubular hyperpolymerization causing an internal viscous load on the myocyte, and myocardial ischemia. In some cases, contractile function is normal, but the hypertrophy is inadequate to normalize wall stress and excessive afterload results. Excessive afterload inhibits ejection, reduces forward output, and leads to heart failure.

The increased wall thickness that helps normalize stress increases diastolic stiffness. Even if muscle properties remain normal, higher filling pressure is required to distend a thicker ventricle. As aortic stenosis advances, collagen deposition also stiffens the myocardium and adds to the diastolic dysfunction.

CLINICAL MANIFESTATIONS

The diagnosis of aortic stenosis is usually first suspected when the classic systolic ejection murmur is heard during physical examination (Chapter 51). The murmur is loudest in the aortic area and radiates to the neck. In some cases, the murmur may disappear over the sternum and reappear over the LV apex, thereby giving the false impression that a murmur of mitral regurgitation is also present (Gallavardin phenomenon). The intensity of the murmur increases with cycle length because longer cycles are associated with greater aortic flow. In mild disease, the murmur peaks in intensity in early systole or midsystole. As the severity of stenosis worsens, the murmur peaks progressively later in systole. Perhaps the most helpful clue to the severity of aortic stenosis by physical examination is the characteristic delay in the carotid pulse with a diminution in its volume (see Fig. 51-5 in Chapter 51); in elderly patients, however, increasing carotid stiffness may pseudonormalize the carotid upstrokes. The LV apical impulse in aortic stenosis is not displaced but is enlarged and forceful. The simultaneous palpation of a forceful LV apex

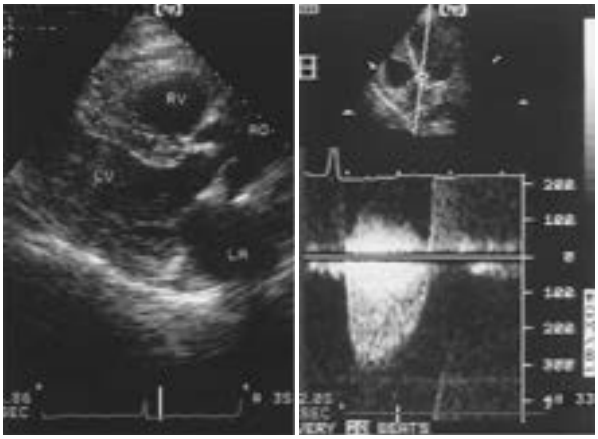

FIGURE 75-2. **Doppler echocardiogram from a patient with aortic stenosis. The *left panel* shows thickened aortic valve leaflets that dome into the aorta with restricted opening in systole. The *top right panel* shows a miniaturized apical four-chamber view at the top with a Doppler cursor through the aorta, and the *bottom right panel* shows a continuous-wave spectral Doppler signal with a peak velocity of 3 m/sec. The peak valve gradient can be calculated as 4×3^2, or 36 mm Hg. AO = aorta; LA = left atrium; LV = left ventricle; RV = right ventricle. (Courtesy of Dr. Anthony DeMaria.)**

beat and a delayed and weakened carotid pulse is a persuasive clue that severe aortic stenosis is present. S_1 in aortic stenosis is generally normal. In congenital aortic stenosis when the valve is not calcified, S_1 may be followed by a systolic ejection click. In calcific disease, S_2 may be single and soft when the aortic component is lost because the valve neither opens nor closes well. In some cases, delayed LV emptying secondary to LV dysfunction may create paradoxical splitting of S_2. An S_4 gallop is common. In advanced disease, pulmonary hypertension and signs of right-sided failure are common.

Because of the dire consequences of missing the diagnosis of aortic stenosis, the physician must have a low threshold for obtaining an echocardiogram whenever aortic stenosis cannot be excluded by physical examination, especially in patients with a history of angina, syncope, or heart failure. In asymptomatic patients with suspicious murmurs, early diagnosis allows the patient and physician to be more vigilant regarding possible early signs and symptoms.

DIAGNOSIS

The electrocardiogram (ECG) in patients with aortic stenosis usually shows LVH (Chapter 54). In some cases of even severe aortic stenosis, however, LVH is absent on the ECG, possibly because of the lack of LV dilation. Left atrial (LA) abnormality is common because the stiff LV increases LA afterload and causes the LA to dilate.

The chest radiograph in aortic stenosis is generally nondiagnostic. The cardiac silhouette is not usually enlarged but may assume a boot-shaped configuration. In advanced cases, there may be signs of cardiomegaly and pulmonary congestion; aortic valve calcification may be seen in the lateral view.

Echocardiography (Chapter 55) is indispensable to assess the extent of LVH, systolic ejection performance, and aortic valve anatomy (Fig. 75-2). Doppler interrogation of the aortic valve makes use of the modified Bernoulli equation (Gradient = $4 \times$ Velocity^2) to assess the severity of the stenosis (Chapter 55). As blood flows from the body of the LV across the stenotic valve, the flow rate must accelerate for the volume to remain constant. Doppler interrogation of the valve can be performed to detect this increase in velocity for estimation of the valve gradient and valve area. The peak aortic flow velocity in patients with preserved LV systolic function is a useful clinical guide to prognosis. In patients with a flow velocity of 3.0 m/second or less, symptoms are unlikely to develop in the next 5 years; by comparison, in patients with a flow velocity of 4.0 m/second or greater, symptoms usually develop within 2 years; and when the velocity exceeds 5 m/second, symptoms are likely to develop within 1 year.

Although exercise testing is contraindicated in symptomatic patients with aortic stenosis because of the high risk for complications, cautious exercise testing is gaining favor in asymptomatic patients. Such testing often reveals latent symptoms or hemodynamic instability that has gone unrecognized during the patient's normal daily activities. Exercise-induced hypotension or symptoms are indications for aortic valve replacement in patients with severe aortic stenosis; in patients with mild to moderate aortic stenosis, another source of exercise limitation should be sought.

Brain natriuretic peptide (BNP) levels may be higher in patients who will become symptomatic in a short time span. Levels exceeding 550 pg/mL portend a poor prognosis, and rising BNP levels on repeated measurements should be a cause for concern. However, it is still premature to rely on this biomarker to indicate the need for valve replacement.

Cardiac catheterization for performance of coronary arteriography is usually undertaken before surgery because most patients with aortic stenosis are of the age at which coronary disease is common. When echocardiography shows severe aortic stenosis and the patient has one or more of the classic symptoms of the disease, formal invasive documentation of the severity of the stenosis is not necessary, and coronary angiography need not be performed in young adults. When the hemodynamic diagnosis is unclear, however, right- and left-sided heart catheterization should be performed to determine the transaortic valvular pressure gradient and cardiac output, which are used to calculate the aortic valve area by the Gorlin formula:

$$A = \frac{CO/SEP \times HR}{44.3\sqrt{h}}$$

where CO is cardiac output (mL/min), SEP is the systolic ejection period (seconds), HR is the heart rate, and h is the mean gradient.

Low Flow, Low Gradient Aortic Stenosis

Two types of pathophysiology can reduce cardiac output and reduce the aortic gradient, thereby misleading the clinician into underestimating the severity and clinical gravity of aortic stenosis.[2] The first type occurs in patients who have developed systolic dysfunction, either from long-standing neglected disease or from a coexisting myocardial infarction. In such patients, the ejection fraction and stroke volume are reduced, as is the transvalvular gradient. Although these patients have a poorer prognosis than do patients with preserved LV function, many still benefit from aortic valve replacement. The second type is patients who have severe LVH, which reduces LV volume. Although the ejection fraction is normal, stroke volume and the valve gradient are reduced. Such patients also usually benefit from valve replacement.

TREATMENT Rx

Medical Therapy

In asymptomatic patients, close follow-up is very important,[3] but no treatment is indicated, nor is any known to be beneficial. Even statins are not useful despite the similar pathobiology between aortic stenosis and coronary disease.

There also is no accepted effective medical therapy for symptomatic aortic stenosis. In patients with heart failure awaiting surgery, diuretics can be used cautiously to relieve pulmonary congestion. Nitrates may also be used cautiously to treat angina pectoris. Although vasodilators, especially angiotensin-converting enzyme (ACE) inhibitors, have become a cornerstone of therapy for heart failure, they are not recommended for aortic stenosis. With fixed valvular obstruction to outflow, vasodilation reduces pressure distal to the obstruction without increasing cardiac output and may cause syncope. When surgery and valvuloplasty are unsuccessful or impossible, diuretics can be used to improve symptoms with the understanding that they will not improve life expectancy.

Invasive Therapy

Valve Replacement Surgery

The only proven effective therapy for aortic stenosis is aortic valve replacement.[4-6] Even octogenarians benefit from valve replacement unless other comorbid factors preclude surgery, so aortic valve replacement should not be denied simply on the basis of age. Valve replacement should also not be denied because the ejection fraction is reduced; the excess afterload imposed by the stenotic valve is relieved with valve replacement, and a depressed ejection fraction usually improves dramatically after surgery. The exception to this rule is a severely reduced ejection fraction in the face of only a small aortic valve gradient; in this case, the severity of the aortic stenosis may be overestimated because the failing LV has difficulty opening a mildly to moderately stenotic valve. In such patients, LV muscle dysfunction either has another cause or is often so severe that it does not recover after valve replacement. Evidence indicates, however, that even some well-selected patients in this category, such as patients who demonstrate increased cardiac output during dobutamine infusion, may benefit from aortic valve replacement.

Percutaneous Aortic Valve Replacement

Percutaneous transcatheter aortic valve implantation reduces the 1-year mortality rate by 45% (from 51% to 31%) in patients who have severe aortic stenosis and are too ill to undergo surgery. This 1-year survival advantage

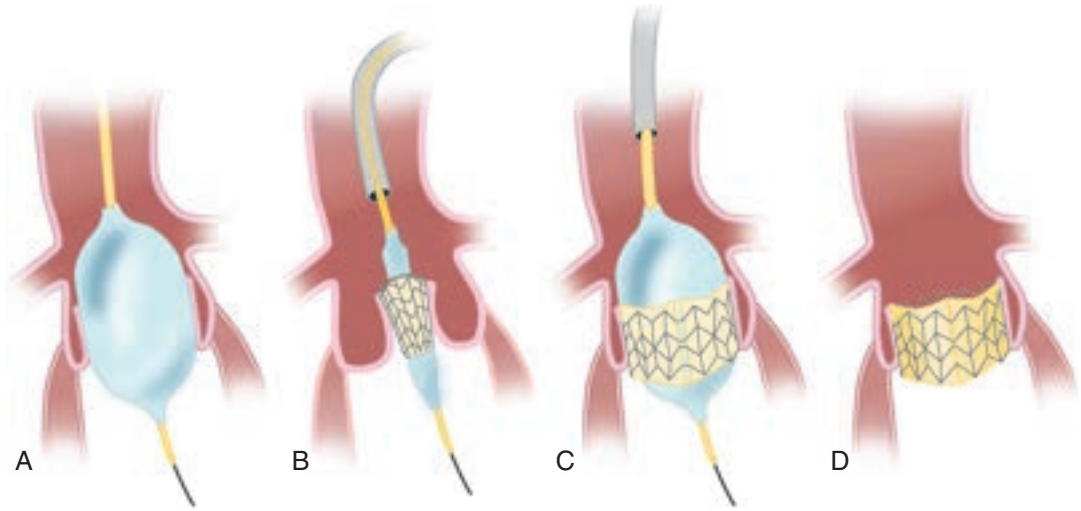

FIGURE 75-3. Steps for transcatheter aortic valve replacement. A, Dilation of the native stenotic valve. **B,** A crimped stented valve has been inserted over a guidewire into the aortic annulus. **C,** Inflation of the balloon deploys the valve. **D,** The balloon is deflated and removed with the aortic valve replaced. (Modified from Cleveland Clinic. Heart valve disease—percutaneous interventions. http://my.clevelandclinic.org/heart/percutaneous/percutaneousValve.aspx.)

remains significant (a subsequent 42% reduction in mortality) among patients who survive beyond the first year, but the mortality benefit may be limited to patients who do not have extensive coexisting conditions. Percutaneous valve implantation also is as good as standard valve replacement in high-risk adults and may be even better with the newer self-expanding prosthesis. Even in the high-risk patients who undergo the procedure, national statistics show a 5.5% in-hospital mortality rate and a 2% stroke rate.[7] Percutaneous valve implantation also benefits patients with severe low-flow aortic stenosis.[8]

Two types of valves are available throughout much of the world, including the United States. One is balloon expandable, and the other uses a self-expanding platform. In the first type, the native valve is dilated, and then a stented valve is inserted over a balloon into the aortic annulus (Fig. 75-3). The balloon is expanded to secure the valve and its stent, which is intended to help prevent restenosis. In the second type, the bimetallic frame expands when it contacts body heat. Paravalvular regurgitation, seen in more than 60% of patients, is associated with late morbidity in proportion to its severity, and engineering efforts are underway to limit paravalvular leak.

Balloon Aortic Valvotomy

In acquired calcific aortic stenosis, leaflet restriction results from heavy calcium deposition in the leaflets themselves and is not caused by commissural fusion. Balloon aortic valvotomy is relatively ineffective in improving aortic stenosis; it generally results in a residual gradient of 30 to 50 mm Hg and a valve area of 1.0 cm^2. The mortality rate after this procedure is similar to that in untreated patients. Balloon aortic valvotomy has generally been abandoned in centers that can perform percutaneous aortic valve replacement but still may be used in other settings when immediate temporary relief is required, either because of the demands of other noncardiac conditions or to improve cardiac status temporarily as a bridge to definitive aortic valve replacement.

PROGNOSIS

In asymptomatic patients with normally functioning or minimally dysfunctional bicuspid aortic valves, the survival rate is similar to age-matched control participants, and sudden death is rare, occurring in fewer than 1% of asymptomatic patients. However, 27% of patients require surgery by 20 years after diagnosis. In adults with asymptomatic but hemodynamically significant aortic stenosis, symptoms typically develop within 5 years.[9] A higher peak aortic jet velocity, heavy valve calcification, a positive exercise test result, severe LV dysfunction, and high β-type natriuretic peptide levels predict a worse prognosis and may warrant consideration of valve replacement in asymptomatic patients with severe disease.

The progression of mild to moderate aortic stenosis to severe disease is the key to the natural history of the disease and is quite variable.[10] Aortic stenosis may remain mild for a decade or more in some patients, but in others, it may progress to severe disease in as little as 5 years.

When symptoms develop, survival declines precipitously. Approximately 35% of patients with aortic stenosis are initially evaluated for angina. Of these, 50% are dead in 5 years unless aortic valve replacement is performed. Approximately 15% have syncope; of these, 50% are dead in only 3 years unless the aortic valve is replaced. Of the 50% with symptoms of heart failure, 50% are dead in 2 years without aortic valve replacement. In all, only 25% of patients with symptomatic aortic stenosis survive 3 years in the absence of valve replacement, and the annual risk for sudden death ranges from 10% in patients with angina to 15% with syncope to 25% with heart failure. After valve replacement surgery, prognosis improves to near normal, especially for patients older than 65 years at the time of valve implantation, presumably because older patients have fewer years at risk for valve-related complications.

MITRAL STENOSIS

EPIDEMIOLOGY

In almost all cases of acquired mitral stenosis, the cause is rheumatic heart disease. As the population ages, however, severe calcification of the mitral annulus is increasingly causing mitral stenosis in elderly adults in the absence of rheumatic involvement. Rheumatic mitral stenosis is three times more common in women and usually develops in the 40s and 50s. Although the disease has become rare in developed countries because of the waning incidence of rheumatic fever, mitral stenosis is still prevalent in developing nations, where rheumatic fever is common (Chapter 290).

PATHOBIOLOGY

At the beginning of diastole, a transient gradient between the LA and LV normally initiates LV filling. After early filling, LA and LV pressures equilibrate. In mitral stenosis, obstruction to LV filling increases LA pressure and produces a persistent gradient between the LA and the LV (see Fig. 57-2 in Chapter 57). The combination of elevated LA pressure (and pulmonary venous pressure) and restriction of inflow into the LV limits cardiac output. Although myocardial involvement from the rheumatic process occasionally affects LV muscle function, the muscle itself is normal in most patients with mitral stenosis. However, in approximately one third of patients with mitral stenosis, LV ejection performance is reduced despite normal muscle function because of reduced preload (from inflow obstruction) and increased afterload as a result of reflex vasoconstriction caused by reduced cardiac output.

Because the RV generates most of the force that propels blood across the mitral valve, the RV incurs the pressure overload of the transmitral gradient. In addition, secondary but reversible pulmonary vasoconstriction develops, thus further increasing pulmonary artery pressure and the burden on the RV. As mitral stenosis worsens, RV failure develops.

CLINICAL MANIFESTATIONS

Patients with mitral stenosis usually remain asymptomatic until the valve area is reduced to about one third its normal size of 4 to 5 cm^2. Then the symptoms typical of left-sided failure—dyspnea on exertion, orthopnea, and paroxysmal nocturnal dyspnea—develop. As the disease progresses and RV failure occurs, ascites and edema are common. Hemoptysis, which is common in mitral stenosis but uncommon in other causes of LA hypertension, develops when high LA pressure ruptures the anastomoses of small bronchial veins. In some cases, a large LA may impinge on the left recurrent laryngeal

nerve and cause hoarseness (Ortner syndrome) or may impinge on the esophagus and cause dysphagia.

Physical Examination

Although mitral stenosis produces typical and diagnostic findings on physical examination, the diagnosis is missed frequently because the auscultatory findings may be subtle. Palpation of the precordium finds a quiet apical impulse. If pulmonary hypertension and RV hypertrophy (RVH) have developed, the examiner notes a parasternal lift. S_1 is typically loud and may be the most prominent physical finding of the disease. A loud S_1 is present because the transmitral gradient holds the mitral valve open throughout diastole until ventricular systole closes the fully opened valve with a loud closing sound. In far-advanced disease, the mitral valve may be so damaged, however, that it neither opens nor closes well, so S_1 may become soft. S_2 is normally split; the pulmonic component is increased in intensity if pulmonary hypertension has developed. Left-sided S_3 and S_4 gallop sounds, which represent the ventricular and atrial components of rapid LV filling, are exceedingly rare in mitral stenosis because obstruction at the mitral valve prevents rapid filling. S_2 is usually followed by an opening snap. The distance between S_2 and the opening snap provides a reasonable estimation of LA pressure and the severity of the mitral stenosis. The higher the LA pressure, the sooner the LA pressure and the falling LV pressure of early ventricular relaxation equilibrate. At this equilibration point, the mitral valve opens, and the opening snap occurs. When LA pressure is high, the opening snap closely (0.06 second) follows S_2. Conversely, when LA pressure is relatively normal, the snap occurs later (0.12 second) and may mimic the cadence of an S_3 gallop. The opening snap is followed by the classic low-pitched early diastolic mitral stenosis rumble, which increases in length as the mitral stenosis worsens. This murmur may be inaudible if the patient has a relatively low resting cardiac output. Modest exercise, such as isometric handgrip, may accentuate the murmur's intensity. If the patient is in sinus rhythm, atrial systole may produce a presystolic accentuation of the murmur. If pulmonary hypertension has developed, the pulmonic component of S_2 increases in intensity to become as loud or louder than the aortic component. With pulmonary hypertension, a diastolic blowing murmur of pulmonary insufficiency (Graham Steell murmur) is often heard, although in many cases, a coexistent murmur of mild aortic insufficiency is mistaken for this murmur. Neck vein elevation, ascites, and edema are present if RV failure has developed.

DIAGNOSIS

Atrial fibrillation is common, but LA abnormality is generally present on the ECG if the patient is in sinus rhythm. If pulmonary hypertension has developed, there is often evidence of RVH.

On the chest radiograph, LA enlargement produces straightening of the left heart border and a double density at the right heart border as a result of the combined silhouettes of the RA and LA. Pulmonary venous hypertension produces increased vascularity. Kerley B lines, which represent thickening of the pulmonary septa secondary to chronic venous engorgement, may also be seen.

The echocardiogram produces excellent images of the mitral valve and is the most important diagnostic tool in confirming the diagnosis (Fig. 75-4). Transthoracic echocardiography or, if necessary, transesophageal echocardiography makes the diagnosis in nearly 100% of cases and accurately assesses severity. Mitral stenosis, similar to aortic stenosis, can be quantified by assessing the transvalvular gradient with the modified Bernoulli principle. The stenosis is considered severe when the area is smaller than 1.5 cm^2 and very severe when valve area is smaller than 1.0 cm^2.

During echocardiography, the suitability of the valve for balloon valvotomy can also be assessed (see later). If even mild tricuspid regurgitation is present, the systolic gradient across the tricuspid valve can be used to gauge pulmonary artery pressure, which is an important prognostic factor in mitral stenosis because the prognosis worsens as pulmonary pressure increases.

Invasive Evaluation

Cardiac Catheterization

Cardiac catheterization is usually unnecessary to assess the severity of mitral stenosis. Because many patients with mitral stenosis are of an age when coronary disease might be present, however, coronary arteriography is generally performed if cardiac surgery is anticipated or if the patient has coexistent angina. In these cases, it is common to perform left- and right-sided heart catheterization to confirm the transmitral gradient and to calculate the valve area from the Gorlin formula (see earlier).

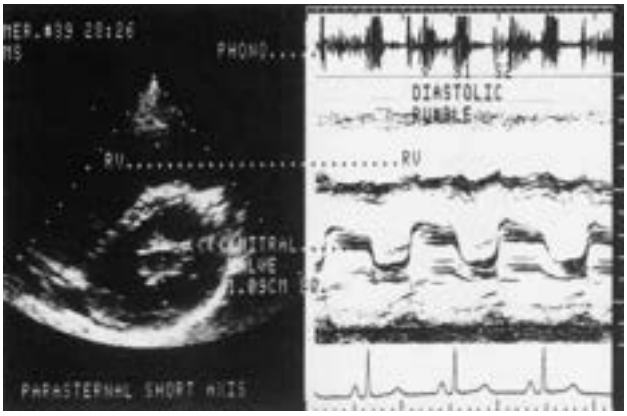

FIGURE 75-4. Mitral stenosis. An en face view of a stenotic mitral valve in the short-axis view of the left ventricle is shown on the *left*. Planimetry for the mitral valve orifice yielded an area of 1.09 cm^2. The M-mode echocardiogram on the *right* has been aligned with the appropriate structures on the *left*. It shows the restricted opening of the mitral valve in diastole associated with the classic diastolic rumbling murmur. RV = right ventricle. (From Assey ME, Usher BW, Carabello BA. The patient with valvular heart disease. In: Pepine CJ, Hill JA, Lambert CR, eds. *Diagnostic and Therapeutic Cardiac Catheterization.* 3rd ed. Baltimore: Williams & Wilkins; 1998:709.)

PREVENTION, TREATMENT, AND PROGNOSIS

Rx

Mitral stenosis can be prevented by appropriate antibiotic treatment of β-hemolytic streptococcal infections (Chapter 290).

Medical Therapy

Asymptomatic patients with mitral stenosis and sinus rhythm require no therapy. Symptoms of mild dyspnea and orthopnea can be treated with diuretics alone. When symptoms worsen to more than mild or if pulmonary hypertension develops, mechanical correction of the stenosis is preferable to medical therapy because it improves longevity in severely symptomatic patients.[11]

Patients with mitral stenosis in whom atrial fibrillation develops usually decompensate because the rapid heart rate reduces diastolic filling time, increases LA pressure, and decreases cardiac output. The heart rate must be controlled promptly, preferably with an infusion of diltiazem, amiodarone, or esmolol for acute atrial fibrillation or with a β-blocker, a calcium channel blocker, or oral digoxin in chronic atrial fibrillation (Chapter 64).

Conversion to sinus rhythm is routinely recommended either pharmacologically or with direct-current countershock (Chapter 64) after anticoagulation is therapeutic. It should be noted that patients with rheumatic atrial fibrillation have been excluded from trials of echocardiogram-guided cardioversion without anticoagulation and trials of rate control versus rhythm control for the chronic management of atrial fibrillation. If sinus rhythm cannot be maintained, mechanical therapy for the mitral stenosis is generally recommended in the hope that sinus rhythm can be restored after the obstruction to atrial outflow is corrected. However, the cause of atrial fibrillation in patients with mitral stenosis probably includes atrial rheumatic inflammation, so restoration of sinus rhythm is unpredictable even after mechanical intervention.

Because patients with concomitant mitral stenosis and atrial fibrillation have an extraordinarily high risk for systemic embolism, they should undergo chronic anticoagulation with warfarin at an international normalized ratio (INR) target of 2.5 to 3.5. Anticoagulation is warranted in all patients unless there is a serious contraindication to its use.

Mechanical Therapy

When symptoms progress past early functional class II, that is, symptoms with more than ordinary activity, or if pulmonary hypertension develops, the prognosis is worse unless the mitral stenosis is relieved. In most instances, an excellent result can be achieved with percutaneous balloon valvotomy. In contrast to aortic stenosis, in mitral stenosis, there is fusion of the valve leaflets at the commissures. Balloon dilation produces a commissurotomy and a substantial increase in valve area that appears to persist for at least a decade and provides improvement comparable to that of closed or open commissurotomy in suitable patients. Nonrheumatic mitral stenosis caused by mitral annular calcification does not respond to balloon valvotomy. The only effective mechanical therapy for this condition is surgical debridement of the mitral annulus followed by mitral valve replacement. Suitability for balloon valvotomy is determined partially during echocardiography. Patients with pliable valves, little valvular calcification, little involvement of the subvalvular

apparatus, and less than moderate mitral regurgitation are ideal candidates. Even when valve anatomy is not ideal, however, valvotomy may be attempted in patients with advanced age or in situations in which comorbid risk factors increase surgical risk. In otherwise healthy patients with unfavorable valve anatomy, surgery to perform an open commissurotomy or valve replacement is undertaken. Even at 20 years, 30% of patients have durable functional benefit after a percutaneous mitral commissurotomy.[12]

PRIMARY (ORGANIC) MITRAL REGURGITATION

EPIDEMIOLOGY

The mitral valve is composed of the mitral annulus, the leaflets, the chordae tendineae, and the papillary muscles. Abnormalities in any of these structures may lead to mitral regurgitation. In primary mitral regurgitation, valvular abnormalities cause the valve to leak; the resulting hemodynamic overload, if prolonged and severe, causes LV damage, heart failure, and eventual death if untreated. This condition must be distinguished from secondary or functional mitral regurgitation, wherein disease of the LV causes the valve to leak; this secondary mitral regurgitation is discussed later in this chapter.

The most common cause of primary mitral regurgitation in the United States is mitral valve prolapse, which is responsible for approximately 90% of all cases and comprises many diseases, including myxomatous degeneration of the valve (E-Fig. 75-E1). Annular calcification, endocarditis, papillary muscle dysfunction or infarction, collagen vascular disease, and rheumatic heart disease are less common causes. Use of the weight loss agents dexfenfluramine and fenfluramine has been implicated in causing valve damage in a few patients who received these drugs.

Primary mitral regurgitation can be subdivided on the basis of chronicity. Common causes of severe acute mitral regurgitation include ruptured chordae tendineae and infective endocarditis. Chronic severe mitral regurgitation is more likely to be caused by myxomatous degeneration of the valve, rheumatic heart disease, or annular calcification.

PATHOBIOLOGY

The pathophysiology of mitral regurgitation can be divided into three phases (Fig. 75-5). In acute mitral regurgitation of any cause, the sudden option for ejection of blood into the LA "wastes" a portion of the LV stroke volume as backward rather than forward flow. The combined regurgitant and forward flow causes volume overload of the LV and stretches the existing sarcomeres toward their maximum length. Use of the Frank-Starling mechanism is maximized, and end-diastolic volume increases concomitantly. The regurgitant pathway unloads the LV in systole because it allows ejection into the relatively low-impedance LA and thereby reduces end-systolic volume. Although increased end-diastolic volume and decreased end-systolic volume act in concert to increase total stroke volume, forward stroke volume is subnormal because a large portion of the total stroke volume is regurgitated into the LA. This regurgitant volume increases LA pressure, so the patient experiences heart failure with low cardiac output and pulmonary congestion despite normal LV contractile function.

In many cases, severe acute mitral regurgitation necessitates emergency surgical correction. By comparison, patients who can be managed through the acute phase or in whom the valve abnormalities develop more slowly may enter the phase of hemodynamic compensation. In this phase, eccentric LVH and increased end-diastolic volume, combined with normal contractile function, allow ejection of a sufficiently large total stroke volume to permit forward stroke volume to return toward normal. LA enlargement allows accommodation of the regurgitant volume at a lower filling pressure. In this phase, the patient may be relatively asymptomatic even during strenuous exercise.

Although severe mitral regurgitation may be tolerated for many years, the lesion often causes LV dysfunction, atrial fibrillation, or heart failure within 5 years of the detection of severe mitral regurgitation. The now damaged ventricle has impaired ejection performance, and end-systolic volume increases. Greater LV residual volume at end systole increases end-diastolic volume and end-diastolic pressure, and the symptoms of pulmonary congestion may reappear. Additional LV dilation may worsen the amount of regurgitation by causing further enlargement of the mitral annulus and malalignment of the papillary muscles. Although there is substantial contractile dysfunction, the increased preload and the presence of the regurgitant pathway, which tends to normalize afterload despite ventricular enlargement, augment the ejection fraction and may maintain it in a relatively normal range.

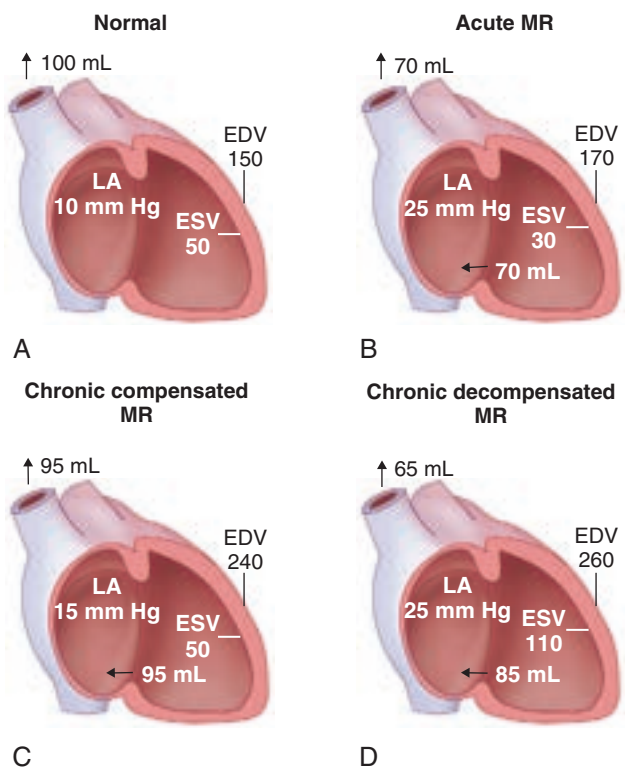

	Preload SL (μ)	Afterload ESS (kdyne/cm²)	CF	EF	RF	FSV (mL)
N	2.07	90	N	.67	.0	100
AMR	2.25	60	N	.82	.50	70
CCMR	2.19	90	N	.79	.5	95
CDMR	2.19	120	↓	.58	.57	65

FIGURE 75-5. Mitral regurgitation. Normal physiology (N) **(A)** is compared with the physiology of acute mitral regurgitation (AMR) **(B)**. Acutely, the volume overload increases preload (sarcomere length [SL]), and end-diastolic volume (EDV) increases from 150 to 170 mL. Unloading of the left ventricle by the presence of the regurgitant pathway decreases afterload (end-systolic stress [ESS]), and end-systolic volume (ESV) falls from 50 to 30 mL. These changes result in an increase in the ejection fraction (EF). Because 50% of the total left ventricular (LV) stroke volume (regurgitant fraction [RF]) is ejected into the left atrium (LA), however, forward stroke volume (FSV) falls from 100 to 70 mL. At this stage, contractile function (CF) is normal. **C,** Chronic compensated mitral regurgitation (CCMR). In CCMR, eccentric cardiac hypertrophy has developed, and EDV has increased substantially. Increased EDV, combined with normal contractile function, permits ejection of a larger total stroke volume and a larger forward stroke volume than in the acute phase. Left atrial enlargement permits lower left atrial pressure. Because the radius term in the Laplace equation has increased with increasing LV volume, afterload and ESV return to normal. **D,** Chronic decompensated mitral regurgitation (CDMR). In this stage, contractile dysfunction causes a large increase in ESV with a fall in total and forward stroke volume. Additional LV enlargement leads to worsening mitral regurgitation. The relatively favorable loading conditions in this phase still permit a normal EF, however, despite contractile dysfunction. (From Carabello BA. Mitral regurgitation: basic pathophysiologic principles. *Mod Concepts Cardiovasc Dis.* 1988;57:53-57.)

The causes of LV contractile dysfunction in patients with mitral regurgitation may relate to loss of contractile proteins and abnormalities in calcium handling. In at least some cases, contractile dysfunction is reversible by timely mitral valve surgery

CLINICAL MANIFESTATIONS

In the medical history, the standard symptoms of left-sided heart failure should be sought (Chapter 58). An attempt to discover potential causes should be made by questioning for a prior history of a heart murmur or abnormal findings on cardiac examination (Chapter 51), rheumatic heart disease, endocarditis (Chapter 76), or the use of anorexigenic drugs.

Volume overload of the LV displaces the apical impulse downward and to the left. S_1 may be reduced in intensity, whereas S_2 is usually physiologically

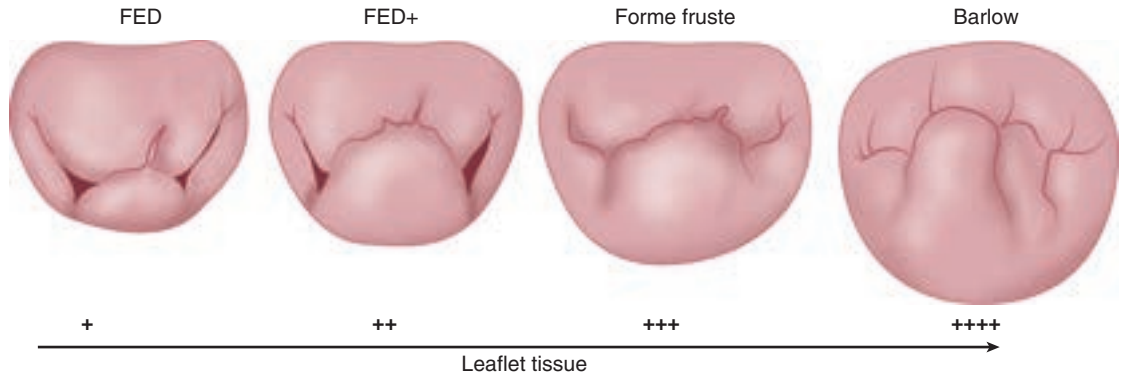

E-FIGURE 75-1. Pathology specimens of myxomatous degeneration of the mitral valve showing the progression of disease from its simplest form (fibroelastic deficiency, FED) to more extensive FED (FED+) to the early stages of a Barlow type valve (Forme Fruste) to the highly redundant Barlow type valve. (From Adams DH, Rosenhek R, Falk V. Degenerative mitral valve regurgitation: best practice revolution. *Eur Heart J.* 2010;31:1958-1966.)

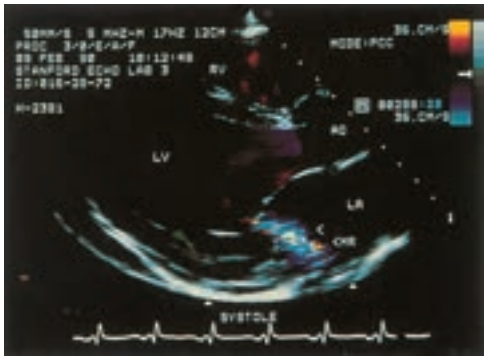

FIGURE 75-6. Two-dimensional echocardiogram of mitral regurgitation with Doppler flow mapping superimposed on a portion of the image. The color information is represented in the sector of the imaging plane extending from the apex of the triangular plane to the *two small arrows* at the *bottom* of the image plane. Mitral regurgitation (MR) is indicated *(open arrows)* and extends from the mitral valve leaflets toward the posterior aspect of the left atrium (LA) during systole. The mosaic of colors representing the mitral regurgitant signal is typical of high-velocity turbulent flow. The low-intensity *orange-brown* signal represents flow directed away from the transducer on the chest wall, and the *blue* shades represent blood in the left ventricular outflow tract moving toward the transducer. AO = aorta; LV = left ventricle; RV = right ventricle.

split. In severe mitral regurgitation, S_2 is followed by S_3, which does not indicate heart failure but reflects rapid filling of the LV by the large volume of blood stored in the LA during systole. The typical murmur of mitral regurgitation is a holosystolic apical murmur that often radiates toward the axilla (Chapter 51). There is a rough correlation between the intensity of the murmur and the severity of the disease, but this correlation is too weak to use in clinical decision making because the murmur may be soft when cardiac output is low. In contrast to aortic stenosis, murmur intensity does not usually vary with the RR interval. In acute mitral regurgitation, the presence of a large v wave may produce rapid equilibration of LA and LV pressure, thereby reducing the driving gradient and shortening the murmur. Pulmonary hypertension may develop and produce right-sided signs; including an RV lift, an increased P_2; and if RV dysfunction has developed, signs of right-sided heart failure.

DIAGNOSIS

The ECG usually shows LVH and LA abnormality. The chest radiograph typically shows cardiomegaly; the absence of cardiomegaly indicates either that the mitral regurgitation is mild or that it has not been chronic enough to allow cardiac dilation to occur.

Echocardiography shows the extent of LA and LV enlargement (Chapter 55). Ultrasonic imaging of the mitral valve is excellent and offers clues to the mitral valve abnormalities responsible for the regurgitation. In some patients, three-dimensional echocardiography can add pathoanatomic information of potential use in aiding surgical repair of the valve. Color-flow Doppler interrogation of the valve (Fig. 75-6) helps assess the severity of regurgitation, but because this technique images flow velocity rather than actual flow, it is subject to errors in interpretation. The Doppler technique is excellent for excluding the presence of mitral regurgitation and for distinguishing between mild and severe degrees. Newer techniques may quantify regurgitation more precisely but are not applicable in every patient, and standard color-flow Doppler examination may not be sufficient for exact quantification of mitral regurgitation or to determine whether the severity of the lesion is sufficient to cause eventual LV dysfunction. Cardiac magnetic resonance imaging (MRI) (Chapter 56), when available, can more precisely quantify the severity of the regurgitation. When the severity of mitral regurgitation is in doubt or if mitral valve surgery is being contemplated, cardiac catheterization (Chapter 57) is helpful in resolving the severity of the lesion; coronary arteriography should be included in patients older than 40 years or with symptoms suggesting coronary disease (Chapter 71).

TREATMENT AND PROGNOSIS

Rx

Medical Therapy

Severe Acute Mitral Regurgitation

In severe acute mitral regurgitation, the patient is usually symptomatic with heart failure or even shock. The goal of medical therapy is to increase forward cardiac output while concomitantly reducing regurgitant volume (Chapter 59). Arterial vasodilators reduce systemic resistance to flow and preferentially increase aortic outflow and simultaneously decrease the amount of mitral regurgitation and LA hypertension. If hypotension already exists, vasodilators such as nitroprusside lower blood pressure further and cannot be used. In these cases, intraaortic balloon counterpulsation (Chapter 107) is preferred if the aortic valve is competent. Counterpulsation increases forward cardiac output by lowering ventricular afterload while augmenting systemic diastolic pressure.

Chronic Symptomatic Mitral Regurgitation

In patients with *symptomatic* mitral regurgitation, ACE inhibitors (e.g., lisinopril, 20 mg/day) reduce LV volume and improve symptoms. Observational evidence suggests that such patients may also benefit from administration of β-blockers. Mitral valve surgery rather than medical therapy is generally preferred, however, in most symptomatic patients with mitral regurgitation. When atrial fibrillation is present, long-term anticoagulation should achieve the same INR goal as for mitral stenosis.

Chronic Asymptomatic Mitral Regurgitation

Vasodilators have had little effect in reducing LV volume or improving normal exercise tolerance in patients with mitral regurgitation, perhaps because afterload is not usually increased in those with chronic asymptomatic mitral regurgitation. There is no definitive indication to begin afterload reduction before symptoms appear because no large randomized trials have been performed, and smaller trials have generally shown no benefit from these therapies.

Surgical Therapy

The timing of mitral valve surgery must weigh the risks of the operation and placement of a prosthesis, if one is inserted, against the risk for irreversible LV dysfunction if surgery is delayed unwisely.[13] For most other types of valve diseases, surgical correction usually requires placement of a prosthetic valve, but in patients with mitral regurgitation, the native valve can often be repaired. Because conservation of the native valve obviates the risks associated with a prosthesis, the option of mitral valve repair should influence the patient and physician toward earlier surgery.

Types of Mitral Valve Surgery

Mitral Valve Repair

When feasible, mitral valve repair (Fig. 75-7) is the preferred operation. Repair restores valve competence, maintains the functional aspects of the apparatus, and avoids the insertion of a prosthesis. Repair is most applicable

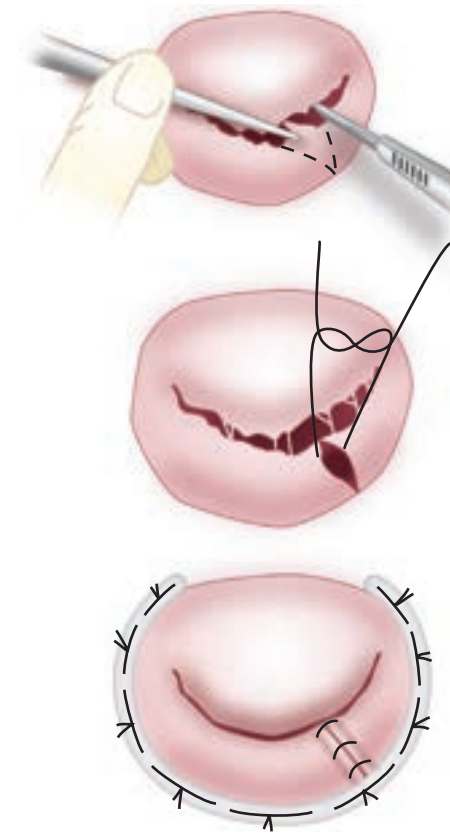

FIGURE 75-7. The stages of mitral valve repair. (Modified from Cleveland Clinic. Mitral valve repair. http://my.clevelandclinic.org/heart/disorders/valve/mvrepair.aspx.)

in cases of posterior chordal rupture; anterior involvement and rheumatic involvement make repair more difficult. Currently, the percentage of mitral valve surgeries that are valve repair varies from 0% to 95% at different hospital centers, averaging about 70% across the United States overall. For severe ischemic mitral regurgitation, mitral valve repair and chordal-sparing mitral valve replacement provide equivalent clinical outcomes at 1 year, although replacement results in less residual mitral regurgitation. For moderate ischemic mitral regurgitation, mitral valve repair does not clearly add benefit above and beyond the benefits of coronary artery bypass surgery.

Percutaneous mitral valve repair with implantation of a clip device is less invasive than conventional mitral valve repair but is also substantially less effective than open surgery for reducing the amount of mitral regurgitation. It is approved in the United States for use in symptomatic inoperable patients. In all cases, the feasibility of repair depends on the pathoanatomy that is causing the mitral regurgitation and the skill and experience of the operating surgeon.

Mitral Valve Replacement with Preservation of the Mitral Apparatus

In this procedure, a prosthetic valve is inserted, but continuity between the native leaflets and the papillary muscles is maintained. This procedure has the advantage of ensuring mitral valve competence while preserving the LV functional aspects of the mitral apparatus. Even if only the posterior leaflets and chordae are preserved, the patient benefits from improved postoperative ventricular function and possibly better survival. In many cases, it is possible to preserve the anterior and posterior chordal attachments, although anterior continuity can be associated with LV outflow tract obstruction. Although the patient benefits from restored mitral valve competence and maintenance of LV function, insertion of a prosthesis still carries all prosthesis-associated risks. Operative mortality with all mitral valve replacement operations is at least twice as high as with mitral valve repair.

Mitral Valve Replacement without Preservation of the Mitral Apparatus

When the native valve cannot be repaired or the chordae preserved, such as in severe rheumatic deformity, the mitral valve leaflets and its apparatus are removed, and a prosthetic valve is inserted. Although this operation almost guarantees mitral valve competence, the mitral valve apparatus is responsible for coordinating LV contraction and for helping maintain the efficient prolate ellipsoid shape of the LV. Destruction of the apparatus leads to a sudden fall in LV function and a decline in postoperative ejection fraction that is often permanent.

Timing of Surgery

Symptomatic Patients

Most patients with symptoms of dyspnea, orthopnea, or fatigue should undergo surgery regardless of which operation is performed because they already have lifestyle limitations from their disease. The mere presence of symptoms worsens the prognosis despite relatively well-preserved LV function. The onset or worsening of symptoms is a summary of the patient's pathophysiology and may give a broader view of cardiovascular integrity than possible with any single measurement of pressure or function. For patients with acute mitral regurgitation owing to flail mitral valve leaflets, earlier surgery appears to be better than prolonged medical management, but randomized trials have not been performed.[14]

Asymptomatic Patients with Normal Left Ventricular Function

Surgery has increasingly been considered in asymptomatic patients who have normal LV function but echocardiographic findings indicating that valve *repair* is likely to be successful. Although these patients are at low risk without surgery, the risk associated with valve repair is less than 1%, and this approach reduces the risks of subsequent LV dysfunction or atrial fibrillation, which may occur if the valvular disease progresses. Furthermore, life span can be normal after successful repair before LV dysfunction has developed. Valve repair obviates the need for protracted, expensive follow-up and provides a durable correction of the lesion. This approach is sensible, however, only if it is certain that valve repair can be performed because insertion of a prosthesis carries unacceptable risk in this low-risk group.

Asymptomatic Patients with Left Ventricular Dysfunction

The onset of LV dysfunction in patients with mitral regurgitation may occur without causing symptoms. Early surgery is warranted to prevent the muscle dysfunction from becoming severe or irreversible. Regardless of whether valve repair or replacement is eventually performed, survival is prolonged if surgery is performed before the ejection fraction declines to less than 0.60 or before the LV is unable to contract to an end-systolic dimension of 40 mm. Patients with severe mitral regurgitation should be monitored yearly with a history, physical examination, and echocardiographic evaluation of LV function. When the patient reports symptoms or echocardiography shows the onset of LV dysfunction, surgery should be undertaken.

Asymptomatic Elderly Patients

Among Medicare recipients, the operative mortality rate is 4% for patients undergoing mitral valve repair and 9% for patients undergoing replacement. The 1-, 5-, and 10-year survival estimates are 91%, 77%, and 54%, respectively, for patients undergoing repair compared with 83%, 65%, and 37% for patients undergoing replacement, respectively.[15] Patients older than 75 years of age may have poorer surgical results than younger patients, especially if coronary disease is present or if mitral valve replacement rather than repair must be performed. However, results of surgery in older patients with mitral regurgitation have steadily improved during the past decade, and elderly patients with symptoms refractory to medical therapy may benefit from surgery. Nevertheless, there is little compelling reason to commit elderly *asymptomatic* patients to a mitral valve operation.

MITRAL VALVE PROLAPSE

DEFINITION

Mitral valve prolapse occurs when one or both of the mitral valve leaflets prolapse into the LA superior to the mitral valve annular plane during systole.[16] The importance of mitral valve prolapse varies from patient to patient. In some cases, prolapse is simply a consequence of normal LV physiology without significant medical impact, such as in situations that produce a small LV (e.g., the Valsalva maneuver or an atrial septal defect), in which reduction of ventricular volume causes relative lengthening of the chordae tendineae and subsequent mitral valve prolapse. At the other end of the spectrum, severe redundancy and deformity of the valve, which occurs in myxomatous valve degeneration, increases the risk for stroke, arrhythmia, endocarditis, and progression to severe mitral regurgitation.

DIAGNOSIS

History

Most patients with mitral valve prolapse are asymptomatic. In some cases, however, mitral valve prolapse is associated with symptoms, including palpitations, syncope, and chest pain. In some cases, chest pain is associated with a positive thallium scintigram indicating the presence of true ischemia despite normal epicardial coronary arteries, perhaps because excessive tension on the papillary muscles increases oxygen consumption and causes ischemia. Palpitations, syncope, and presyncope, when present, are linked to autonomic dysfunction (Chapters 51, 62, and 418), which appears to be more prevalent in patients with mitral valve prolapse.

Physical Examination

On physical examination, the mitral valve prolapse syndrome produces the characteristic findings of a midsystolic click and a late systolic murmur. The click occurs when the chordae tendineae are stretched taut by the prolapsing mitral valve in midsystole. As this occurs, the mitral leaflets move past their coaptation point, permit mitral regurgitation, and cause the late systolic murmur (see Table 51-7 in Chapter 51). Maneuvers that make the LV smaller, such as the Valsalva maneuver, cause the click to appear earlier and the murmur to be more holosystolic and often louder (see Table 51-8 in Chapter 51). In some cases of echocardiographically proven mitral valve prolapse, neither the click nor the murmur is present; in other cases, only one of these findings is present.

Noninvasive Evaluation

Echocardiography is useful to prove that prolapse is present, to image the amount of regurgitation and its physiologic effects, and to discern the pathoanatomy of the mitral valve. Although an echocardiogram is not necessary to diagnose prolapse in patients with the classic physical findings, the echocardiogram adds significant prognostic information because it can detect patients who have specifically abnormal valve morphology and in whom most of the complications of the disease occur. Prolapse shown in the four-chamber echocardiographic view should be confirmed in the parasternal long-axis view.

TREATMENT

Because most cases of mitral valve prolapse are asymptomatic, therapy is usually unnecessary. Although prophylaxis against infective endocarditis was previously recommended in these patients, guidelines no longer recommend antibiotic prophylaxis based on available data (Chapter 76). In patients with palpitations and autonomic dysfunction, β-blockers are often effective in relieving symptoms. Low-dose aspirin therapy has been recommended for patients with redundant leaflets because these patients have a slightly increased risk for stroke. No data from large studies are available to support

this contention, however. If severe mitral regurgitation or a flail mitral leaflet develops, however, the therapy is the same as for other causes of mitral regurgitation.

PROGNOSIS

Most patients with mitral valve prolapse have a benign clinical course; even for complication-prone patients with redundant and misshapen mitral leaflets, complications are relatively rare. Approximately 10% of patients with thickened leaflets experience infective endocarditis, stroke, progression to severe mitral regurgitation, or sudden death. The progression to severe mitral regurgitation varies with gender and age, and men are approximately twice as likely to progress as women. By 50 years of age, only approximately one in 200 men requires surgery to correct mitral regurgitation. By the age of 70 years, the risk increases to approximately 3%.

SECONDARY (FUNCTIONAL) MITRAL REGURGITATION

DEFINITION

Secondary or functional mitral regurgitation is a very different disease than is primary mitral regurgitation. In primary mitral regurgitation, treating the mitral regurgitation cures the patient. Conversely, secondary mitral regurgitation is a consequence of LV myocardial dysfunction caused either by myocardial infarction or dilated cardiomyopathy. Because treating secondary mitral regurgitation cannot reverse those entities, the role of valve therapy is often unclear.

EPIDEMIOLOGY

An estimated 5 million Americans have heart failure (Chapters 58 and 59), and about half of them have heart failure with a reduced ejection fraction. About 75% of these latter patients also have some degree of secondary mitral regurgitation, which is severe in about 20% of them.

PATHOBIOLOGY

In secondary mitral regurgitation, the valve itself is normal. Regurgitation occurs because ventricular damage leads to dilation, which causes displacement of the papillary muscles, which in turn prevents a normal valve from reaching its coaptation point. Mitral closing is further impaired by mitral annular dilation and reduced closing force from the weakened myocardium.

CLINICAL MANIFESTATIONS

Because virtually all patients with secondary mitral regurgitation have heart failure, almost all complain of the symptoms of heart failure. Some may note a worsening of symptoms when more than mild secondary mitral regurgitation develops. Probably because the mitral valve itself is normal, as well as because of the weakened force of contraction, the murmur of functional mitral regurgitation may be unimpressive or even inaudible.

DIAGNOSIS

Echocardiography is the mainstay of the diagnosis. In addition to evaluating the severity of the mitral regurgitation itself, the echocardiogram is helpful in establishing the extent of the dilated cardiomyopathy or of the prior myocardial infarction that is responsible for causing the secondary mitral regurgitation.

THERAPY AND PROGNOSIS

Rx

The presence of mitral regurgitation in patients with systolic dysfunction is associated with worsened prognosis, which in turn probably reflects both poorer LV function as well as the imposition of an extra volume overload on an already weakened LV.

Because all patients with secondary mitral regurgitation have heart failure, they should receive standard treatment for heart failure (Chapter 59). Many patients with secondary mitral regurgitation also have conducting system abnormalities, and patients with left bundle branch block may benefit from cardiac resynchronization therapy (Chapters 59 and 66), which improves systolic function while often reducing secondary mitral regurgitation, sometimes even eliminating it. When the above therapies fail to relieve symptoms, surgical treatment can be considered. Unlike in primary mitral regurgitation, it is unclear whether valve repair is superior to valve replacement, and it is also unclear whether mitral surgery prolongs life. Indeed, recent data from a randomized trial found no difference in overall outcome between repair and replacement for secondary mitral regurgitation. The average survival rate at 5 years is about 50%, and this figure has not changed much over the past several decades.

AORTIC REGURGITATION

DEFINITION

Aortic regurgitation is caused either by abnormalities of the aortic leaflets or by abnormalities of the proximal aortic root. Leaflet abnormalities causing aortic regurgitation include a bicuspid aortic valve,[17] infective endocarditis, and rheumatic heart disease; anorexigenic drugs have also been implicated. Common aortic root abnormalities that cause aortic regurgitation include Marfan syndrome (Chapter 260), hypertension-induced annuloaortic ectasia, aortic dissection (Chapter 78), syphilis (Chapter 319), ankylosing spondylitis (Chapter 265), and psoriatic arthritis (Chapter 265). Acute aortic regurgitation is usually caused by infective endocarditis (Chapter 76) or aortic dissection.

PATHOBIOLOGY

As with mitral regurgitation, aortic regurgitation imparts a volume overload on the LV because the LV must pump the forward flow entering from the LA and the regurgitant volume returning through the incompetent aortic valve. Also as with mitral regurgitation, the volume overload is compensated for by the development of eccentric cardiac hypertrophy, which increases chamber size and allows the ventricle to pump a greater total stroke volume and a greater forward stroke volume. Ventricular enlargement also allows the LV to accommodate the volume overload at a lower filling pressure. In contrast to mitral regurgitation, the entire stroke volume is ejected into the aorta in aortic regurgitation. Because pulse pressure is proportional to stroke volume and elastance of the aorta, the increased stroke volume increases systolic pressure. Systolic hypertension leads to afterload excess, which does not generally occur in mitral regurgitation. Accordingly, ventricular geometry also differs between mitral and aortic regurgitation because the afterload excess in aortic regurgitation causes a modest element of concentric hypertrophy, as well as severe eccentric hypertrophy.

In acute aortic insufficiency, such as might occur in infective endocarditis, severe volume overload of the previously unprepared LV results in a sudden fall in forward output while precipitously increasing LV filling pressure. It is probably this combination of pathophysiologic factors that leads to rapid decompensation, presumably because the severely diminished gradient for coronary blood flow causes ischemia and progressive deterioration in LV function. In acute aortic insufficiency, reflex vasoconstriction increases peripheral vascular resistance. In compensated chronic aortic insufficiency, vasoconstriction is absent, and vascular resistance may be reduced and contribute to the hyperdynamic circulation observed in these patients.

CLINICAL MANIFESTATIONS

The most common symptoms from chronic aortic regurgitation are those of left-sided heart failure, that is, dyspnea on exertion, orthopnea, and fatigue. In acute aortic regurgitation, cardiac output and shock may develop rapidly. The onset of symptoms in patients with chronic aortic regurgitation usually heralds the onset of LV systolic dysfunction. Some patients with symptoms have apparently normal systolic function, however, and the symptoms may be attributed to diastolic dysfunction. Other patients may have ventricular dysfunction yet remain asymptomatic.

Angina may also occur in patients with aortic insufficiency but less commonly than in those with aortic stenosis. The cause of angina in aortic regurgitation is probably multifactorial. Coronary blood flow reserve is reduced in some patients because diastolic runoff into the LV lowers aortic diastolic pressure while increasing LV diastolic pressure; these two influences lower the driving pressure gradient for flow across the coronary bed. When angina occurs in aortic regurgitation, it may be accompanied by flushing. Other symptoms include carotid artery pain and an unpleasant awareness of the heartbeat.

DIAGNOSIS

Physical Examination

Aortic regurgitation produces a myriad of signs because a hyperdynamic, enlarged LV ejects a large stroke volume at high pressure into the systemic

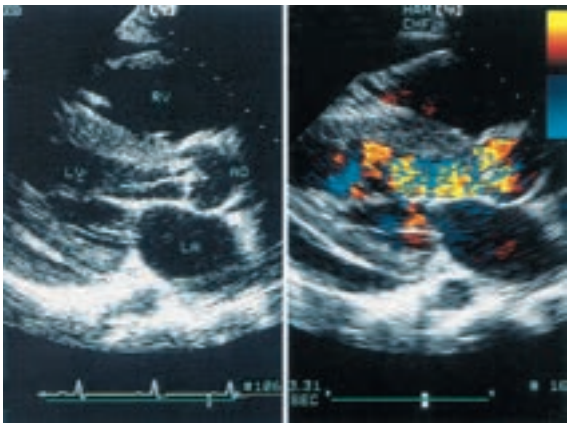

FIGURE 75-8. Echocardiogram of a patient with aortic regurgitation caused by infective endocarditis. The *left panel* shows a linear vegetation *(arrow)* prolapsing into the left ventricular (LV) outflow tract from the aortic valve leaflet in diastole. The *right panel* is a color-flow Doppler image exhibiting turbulent blood flow filling the LV tract during diastole. AO = aorta; LA = left atrium; RV = right ventricle. (Courtesy of Dr. Anthony DeMaria.)

circulation. Palpation of the precordium finds a hyperactive apical impulse displaced downward and to the left. S_1 and S_2 are usually normal. S_2 is followed by a diastolic blowing murmur heard best along the left sternal border with the patient sitting upright. In mild disease, the murmur may be short and heard only in the beginning of diastole when the gradient between the aorta and the LV is highest. As the disease worsens, the murmur may persist throughout diastole. A second murmur, a mitral valve rumble, is heard at the LV apex in patients with severe aortic insufficiency. Although the cause is still debated, this Austin Flint murmur is probably produced as the regurgitant jet impinges on the mitral valve and causes it to vibrate.

In chronic aortic regurgitation, the high stroke volume and reduced systemic arterial resistance result in a wide pulse pressure, which may generate a number of signs, including Corrigan pulse (sharp upstroke and rapid decline of the carotid pulse), de Musset sign (head bobbing), Duroziez sign (combined systolic and diastolic bruits created by compression of the femoral artery with the stethoscope), and Quincke pulse (systolic plethora and diastolic blanching in the nail bed when gentle traction is placed on the nail). Perhaps the most reliable of physical signs indicating severe aortic regurgitation is Hill sign, an increase in femoral systolic pressure of 40 mm Hg or more compared with systolic pressure in the brachial artery.

In contrast to chronic aortic insufficiency with its myriad clinical signs, acute aortic insufficiency may have a subtle manifestation. The eccentric hypertrophy, which compensates for chronic aortic insufficiency, has not yet had time to develop, and the large total stroke volume responsible for most of the signs of chronic aortic insufficiency is absent. The only clues to the presence of acute aortic insufficiency may be a short diastolic blowing murmur and reduced intensity of S_1. This latter sign occurs because high diastolic LV pressure closes the mitral valve early in diastole (mitral valve preclosure) so that when ventricular systole occurs, only the tricuspid component of S_1 is heard.

Noninvasive Evaluation

The ECG in patients with aortic insufficiency is nonspecific but almost always demonstrates LVH. The chest radiograph shows an enlarged heart, often with uncoiling and enlargement of the aortic root.

Echocardiography (Chapter 55) is the most important noninvasive tool for assessing the severity of aortic insufficiency and its impact on LV geometry and function (Fig. 75-8). During echocardiography, the LV end-diastolic dimension, end-systolic dimension, and fractional shortening are determined. Aortic valve anatomy and aortic root anatomy can be assessed and the cause of the aortic regurgitation can often be determined. Color-flow Doppler examination of the aortic valve helps quantify the severity of aortic regurgitation by assessing the depth and width to which the diastolic jet penetrates the LV. Another way to assess the severity of aortic regurgitation is the pressure half-time method: continuous-wave Doppler interrogation of the aortic valve displays the decay of the velocity of retrograde flow across the valve. In mild aortic insufficiency, the gradient across the valve is high throughout diastole, and its rate of decay is slow, with production of a long Doppler half-time (the time that it takes the velocity to decay from its peak to that value divided by the square root of 2). In severe aortic regurgitation, there is rapid equilibration between pressure in the aorta and pressure in the LV, and the Doppler half-time is short. If mitral valve preclosure is detected in acute aortic insufficiency, urgent surgery is necessary. In cases in which the severity of aortic insufficiency is in doubt, MRI can quantify regurgitant flow or catheterization with aortography can visualize regurgitant flow to resolve the issue.

TREATMENT AND PROGNOSIS

Medical Therapy

Asymptomatic Patients with Normal Left Ventricular Function

Because aortic regurgitation increases LV afterload, which decreases cardiac efficiency, afterload reduction with nifedipine and other vasodilators, including ACE inhibitors and hydralazine, improves hemodynamics in the short term. Although initial data suggested that such therapy could delay or reduce the need for aortic valve surgery without any adverse effects when surgery is finally performed, more recent data suggest no benefit from such therapy. These discrepant results from relatively small trials preclude firm recommendations. When hypertension accompanies aortic regurgitation, it should be treated according to standard guidelines (Chapter 67).

Symptomatic Patients or Patients with Left Ventricular Dysfunction

Patients who have symptoms or manifest LV dysfunction should not be treated medically, except for short-term stabilization, but should undergo aortic valve surgery as soon as feasible.

Surgical Therapy

Acute Aortic Regurgitation

When any of the symptoms or signs of heart failure develop, even if mild, the medical mortality rate is high and approaches 75%. Echocardiographic evidence of preclosure of the mitral valve from high diastolic intracavitary LV pressure is an especially ominous sign. Therapy with vasodilators, such as nitroprusside, may temporarily improve the patient's condition before surgery but is never a substitute for surgery. In patients with acute aortic regurgitation caused by bacterial endocarditis (Chapter 76), surgery may be delayed to permit a full or partial course of antibiotics, but persistent, severe aortic regurgitation requires emergency valve replacement. Even when blood cultures have been positive recently and antibiotic therapy has been of brief duration, the valve reinfection rate is low, 0% to 10%, with valve replacement or valve repair. Emergency surgery should not be withheld simply because the duration of antibiotic therapy has been brief.

Chronic Aortic Regurgitation

Asymptomatic patients who manifest evidence of LV dysfunction benefit from surgery. Because loading conditions differ between aortic and mitral regurgitation, the objective markers for the presence of LV dysfunction also differ. In aortic regurgitation, when the ejection fraction is less than 0.50 or the end-systolic dimension is greater than 50 mm, postoperative outcome is impaired, presumably because these markers indicate that LV dysfunction has developed. Surgery should be performed before these benchmarks are reached. A calculated regurgitant orifice above 30 mm^2 portends a poorer prognosis and may warrant surgery.

Patients with advanced symptoms are at increased risk for a suboptimal surgical outcome regardless of whether they have evidence of LV dysfunction. Patients should undergo aortic valve replacement before symptoms impair lifestyle.

Although some patients may be able to undergo successful aortic valve repair to restore aortic valve competence, most patients require insertion of an aortic valve prosthesis.

TRICUSPID REGURGITATION

DEFINITION

Tricuspid regurgitation is usually secondary to a hemodynamic load on the RV rather than a structural valve deformity. Diseases that cause pulmonary hypertension, such as chronic obstructive airway disease or intracardiac shunts, lead to RV dilation and subsequent tricuspid regurgitation. Because most of the force that is needed to fill the LV is provided by the RV, LV dysfunction leading to elevated LV filling pressure also places the RV under a hemodynamic load and can eventually lead to RV failure and tricuspid

regurgitation. In some instances, tricuspid regurgitation may be caused by pathology of the valve itself. The most common cause of primary tricuspid regurgitation is infective endocarditis, usually stemming from drug abuse and unsterile injections. Other causes include trauma (especially from hitting the steering wheel or dashboard in motor vehicle accidents), carcinoid syndrome, rheumatic involvement of the tricuspid valve, myxomatous degeneration, RV infarction, and mishaps during endomyocardial biopsy.

DIAGNOSIS

The symptoms of tricuspid regurgitation are those of right-sided heart failure and include ascites, edema, and occasionally right upper quadrant pain. On physical examination, tricuspid regurgitation produces jugular venous distention accentuated by a large v wave as blood is regurgitated into the RA during systole. Regurgitation into the hepatic veins causes hepatic enlargement and liver pulsation. RV enlargement is detected as a parasternal lift. Ascites and edema are common. The murmur of tricuspid regurgitation is a holosystolic murmur heard along the left sternal border, often increasing with inspiration. The murmur may be faint, and it usually can be heard only under the best auscultatory conditions.

The definitive diagnosis of tricuspid regurgitation is made during echocardiography. Doppler interrogation of the tricuspid valve shows systolic disturbance of the right atrial blood pool. Echocardiography (Chapter 55) can also be used to determine the severity of pulmonary hypertension, measure RV dilation, and assess whether the valve itself is intrinsically normal or abnormal.

TREATMENT AND PROGNOSIS

Rx

Therapy for secondary tricuspid regurgitation is generally aimed at the cause of the lesion. If LV failure has been responsible for RV failure and tricuspid regurgitation, the standard therapy for improving LV failure (Chapter 59) lowers LV filling pressure, reduces secondary pulmonary hypertension, relieves some of the hemodynamic burden of the RV, and partially restores tricuspid valve competence. If pulmonary disease is the primary cause, therapy is directed toward improving lung function. Medical therapy directed at tricuspid regurgitation is usually limited to diuretics because the vasodilators that are so useful in the treatment of left-sided heart failure are often ineffective in treating pulmonary hypertension. However, a number of options can improve late-stage symptoms that are primarily caused by the pulmonary hypertension itself (Chapter 68).

Surgical intervention for the tricuspid valve is rarely entertained in isolation. However, if other cardiac surgery is planned in a patient with severe tricuspid regurgitation, concomitant ring annuloplasty or tricuspid valve repair is frequently attempted to ensure postoperative tricuspid competence. Because a second operation to address residual tricuspid regurgitation after successful left-sided valve surgery carries an unacceptably high mortality rate, concomitant tricuspid annuloplasty is now entertained for even mild to moderate tricuspid regurgitation during left-sided valve surgery. Tricuspid valve replacement is often not well tolerated and is rarely performed except when severe deformity, as is often seen in endocarditis or carcinoid disease, precludes valve repair.

PULMONIC STENOSIS

DEFINITION

Pulmonic stenosis is a congenital disease resulting from fusion of the pulmonic valve cusps (Chapter 69). It is usually detected and corrected during childhood, but occasionally cases are diagnosed for the first time in adulthood. Symptoms of pulmonic stenosis include angina and syncope. Occasionally, symptoms of right-sided heart failure develop. During physical examination, the uncalcified valve in pulmonic stenosis produces an early systolic ejection click on opening. During inspiration, the click diminishes or disappears because increased flow into the right side of the heart during inspiration partially opens the pulmonic valve in diastole so that systole causes less of an opening sound. The click is followed by a systolic ejection murmur that radiates to the base of the heart. If the transvalvular gradient is severe, RVH develops and produces a parasternal lift.

The diagnosis of pulmonic stenosis is confirmed by echocardiography, which quantifies the transvalvular gradient and the degree of RVH and dysfunction.

TREATMENT AND PROGNOSIS

Rx

In asymptomatic patients with a gradient less than 25 mm Hg, no therapy is required. If symptoms develop or the gradient exceeds 50 mm Hg, balloon commissurotomy is effective in reducing the gradient and relieving symptoms. Although long-term prognosis is not yet established, 90% of patients do not require reintervention 10 years after balloon therapy.

Postoperative Care of Patients with Substitute Heart Valves

After a prosthetic valve has been inserted, a baseline echocardiogram should be obtained to provide a reference point in the event that valve dysfunction is suspected at a later date. Echocardiography does not need to be repeated unless there is a change in clinical status or physical findings. The major causes of mechanical valve dysfunction are infective endocarditis and thrombus formation. For bioprostheses, endocarditis is also a risk, but valve degeneration, which can cause either stenosis or regurgitation, is a major concern.

Whenever a patient with a prosthetic heart valve has a temperature higher than 100° F, endocarditis must be excluded by blood culture; for fever with signs of sepsis, broad-spectrum antibiotics must be begun while awaiting culture results. For patients with bioprosthetic valves, mechanical prostheses, and homografts, endocarditis prophylaxis should be instituted at the time of procedures that are associated with a high risk for bacteremia (Chapter 76). Whether prophylaxis is necessary for pulmonary autografts is currently unclear, but physicians usually prescribe prophylaxis for these patients.

All patients with a mechanical heart valve require anticoagulation.[18] Recommended INR values range from 2.0 for a young normotensive patient in sinus rhythm with an aortic valve prosthesis to 3.5 for a patient with atrial fibrillation and a mitral valve prosthesis. Aspirin, 325 mg, is recommended in addition to warfarin to reduce the risk for valve thrombosis in patients who have mechanical prosthetic valves that are at higher risk for thromboembolic complications. Newer anticoagulants such as dabigatran are associated with more thrombosis and bleeding in patients with mechanical prosthetic valves and should not be used. When thrombosis occurs on a prosthetic valve, about 60% of patients can have restored valve function after intravenous infusion of a thrombolytic agent.[19] In patients with periprosthetic paravalvular leaks, a percutaneous procedure can successfully close about 85% of the leaks.[20]

Choices among Prosthetic Valves

Different types of prosthetic valves (Fig. 75-9) have different advantages and disadvantages (Table 75-2). At long-term follow-up, primary valve failure with resulting need for reoperation is much more common with bioprosthetic valves, and bleeding is generally more common with mechanical valves. For mitral valves, long-term survival is similar with bioprosthetic and mechanical valves. Survival after aortic valve replacement is probably better with mechanical valves. Another alternative is to use the patient's own pulmonic valve to replace the diseased aortic valve and then to implant a prosthetic pulmonic valve (the Ross procedure). In a randomized trial, this procedure was superior to homograft valve and root replacement, with a 10-year survival rate of 97% compared with 83%.

In general, a tissue valve is recommended in patients who have a life expectancy of less than 15 years or who are unable or unwilling to maintain warfarin anticoagulation. A mechanical valve is preferred in patients who already have another indication for anticoagulation or who have a longer life expectancy and want to minimize the risk of reoperation. Because many patients prefer the risk of reoperation to that of anticoagulation, the age at which bioprostheses are implanted has steadily declined, and many 60-year-old patients now request and receive a bioprosthetic valve.

TABLE 75-2 ADVANTAGES AND DISADVANTAGES OF SUBSTITUTE CARDIAC VALVES

TYPE OF VALVE	ADVANTAGES	DISADVANTAGES
Bioprosthesis (Carpentier-Edwards, Hancock)	Avoids anticoagulation in patients with sinus rhythm	Durability limited to 10-15 yr Relatively stenotic
Mechanical valves (St. Jude, Medtronic-Hall, Starr-Edwards)	Good flow characteristics in small sizes Durable	Require anticoagulation
Homografts and autografts	Anticoagulation not required Durability increased over that of bioprostheses	Surgical implantation technically demanding

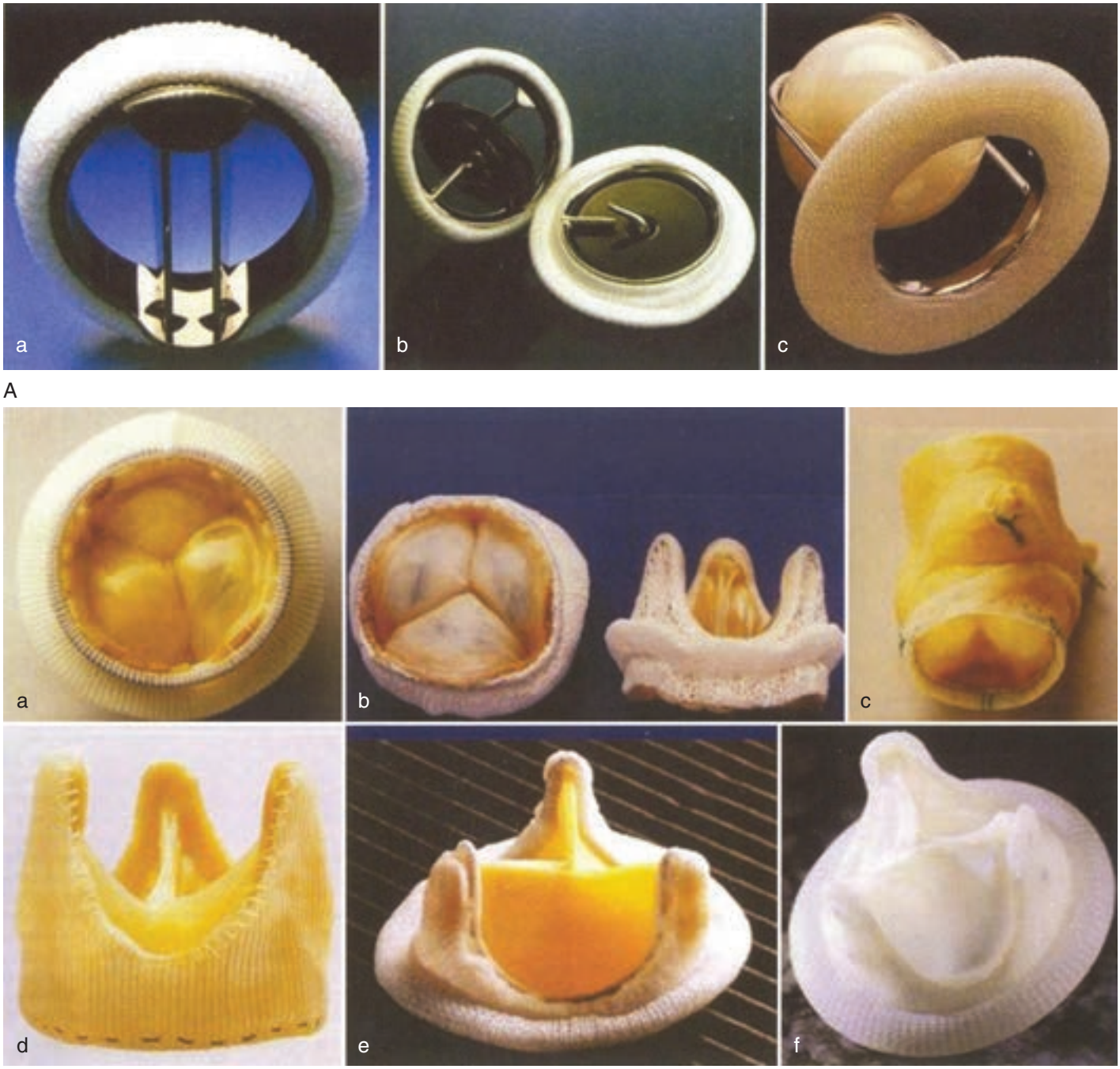

FIGURE 75-9. **A, Common mechanical valves: *a,* a bileaflet St. Jude Medical valve; *b,* a Medtronic-Hall tilting disc valve; *c,* a Starr-Edwards ball cage valve (no longer manufactured but still in use). (From Antunes MJ, Burke AP, Carabello B. Valvular heart disease. In: Braunwald E, Rahimtoola SH, eds. *Essential Atlas of Heart Diseases*. 3rd ed. Philadelphia: Current Medicine Group; 2005:296-297.) B, Common bioprostheses: *a,* Hancock modified orifice stented valve; *b,* Carpentier-Edwards stented porcine valve; *c,* Medtronic free style stentless valve; *d,* St. Jude Medical Toronto SPV stentless valve; *e,* Carpentier-Edwards pericardial valve; *f,* Autologous pericardial valve. (From Grunkemeier GL, Rahimtoola SH, Starr A. Prosthetic heart valves. In: Rahimtoola SH, ed. *Valvular Heart Disease*. Philadelphia: Current Medicine Group; 1997:13.9-13.11.)**

Grade A References

A1. Rossebø AB, Pedersen TR, Boman K, et al. Intensive lipid lowering with simvastatin and ezetimibe in aortic stenosis. *N Engl J Med*. 2008;359:1343-1356.

A2. Leon MB, Smith CR, Mack M, et al. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med*. 2010;363:1597-1607.

A3. Makkar RR, Fontana GP, Jilaihawi H, et al. Transcatheter aortic-valve replacement for inoperable severe aortic stenosis. *N Engl J Med*. 2012;366:1696-1704.

A4. Kodali SK, Williams MR, Smith CR, et al. Two-year outcomes after transcatheter or surgical aortic-valve replacement. *N Engl J Med*. 2012;366:1686-1695.

A5. Adams DH, Popma JJ, Reardon MJ, et al. Transcatheter aortic-valve replacement with a self-expanding prosthesis. *N Engl J Med*. 2014;370:1790-1798.

A6. Acker MA, Parides MK, Perrault LP, et al. Mitral-valve repair versus replacement for severe ischemic mitral regurgitation. *N Engl J Med*. 2014;370:23-32.

A7. Smith PK, Puskas JD, Ascheim DD, et al. Surgical treatment of moderate ischemic mitral regurgitation. *N Engl J Med*. 2014;371:2178-2188.

A8. Mauri L, Foster E, Glower DD, et al. 4-year results of a randomized controlled trial of percutaneous repair versus surgery for mitral regurgitation. *J Am Coll Cardiol*. 2013;62:317-328.

A9. Eikelboom JW, Connolly SJ, Brueckmann M, et al. Dabigatran versus warfarin in patients with mechanical heart valves. *N Engl J Med*. 2013;369:1206-1214.

A10. Hammermeister K, Sethi GK, Henderson WG, et al. Outcomes 15 years after valve replacement with a mechanical versus a bioprosthetic valve: final report of the Veterans Affairs randomized trial. *J Am Coll Cardiol*. 2000;36:1152-1158.

A11. Stassano P, Di Tommaso L, Monaco M, et al. Aortic valve replacement: a prospective randomized evaluation of mechanical versus biological valves in patients ages 55 to 70 years. *J Am Coll Cardiol*. 2009;54:1862-1868.

A12. El-Hamamsy I, Eryigit Z, Stevens LM, et al. Long-term outcomes after autograft versus homograft aortic root replacement in adults with aortic valve disease: a randomised controlled trial. *Lancet*. 2010;376:524-531.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Carabello BA. Introduction to aortic stenosis. *Circ Res*. 2013;113:179-185.
2. Saikrishnan N, Kumar G, Sawaya FJ, et al. Accurate assessment of aortic stenosis: a review of diagnostic modalities and hemodynamics. *Circulation*. 2014;129:244-253.
3. Manning WJ. Asymptomatic aortic stenosis in the elderly: a clinical review. *JAMA*. 2013;310:1490-1497.
4. Lindman BR, Bonow RO, Otto CM. Current management of calcific aortic stenosis. *Circ Res*. 2013;113:223-237.
5. Erbel R, Aboyans V, Boileau C, et al. 2014 ESC Guidelines on the diagnosis and treatment of aortic diseases: document covering acute and chronic aortic diseases of the thoracic and abdominal aorta of the adult. The Task Force for the Diagnosis and Treatment of Aortic Diseases of the European Society of Cardiology (ESC). *Eur Heart J*. 2014;35:2873-2926.
6. Nishimura RA, Otto CM, Bonow RO, et al. 2014 AHA/ACC Guideline for the management of patients with valvular heart disease: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129:2440-2492.
7. Mack MJ, Brennan JM, Brindis R, et al. Outcomes following transcatheter aortic valve replacement in the United States. *JAMA*. 2013;310:2069-2077.
8. Herrmann HC, Pibarot P, Hueter I, et al. Predictors of mortality and outcomes of therapy in low-flow severe aortic stenosis: a Placement of Aortic Transcatheter Valves (PARTNER) trial analysis. *Circulation*. 2013;127:2316-2326.
9. Otto CM, Prendergast B. Aortic-valve stenosis—from patients at risk to severe valve obstruction. *N Engl J Med*. 2014;371:744-756.
10. Coffey S, Cox B, Williams MJ. The prevalence, incidence, progression, and risks of aortic valve sclerosis: a systematic review and meta-analysis. *J Am Coll Cardiol*. 2014;63:2852-2861.
11. Vahanian A, Alfieri O, Andreotti F, et al. Guidelines on the management of valvular heart disease (version 2012). *Eur Heart J*. 2012;33:2451-2496.
12. Bouleti C, Iung B, Laouénan C, et al. Late results of percutaneous mitral commissurotomy up to 20 years: development and validation of a risk score predicting late functional results from a series of 912 patients. *Circulation*. 2012;125:2119-2127.
13. Bonow RO. Chronic mitral regurgitation and aortic regurgitation: have indications for surgery changed? *J Am Coll Cardiol*. 2013;61:693-701.
14. Suri RM, Vanoverschelde JL, Grigioni F, et al. Association between early surgical intervention vs watchful waiting and outcomes for mitral regurgitation due to flail mitral valve leaflets. *JAMA*. 2013;310:609-616.
15. Vassileva CM, Mishkel G, McNeely C, et al. Long-term survival of patients undergoing mitral valve repair and replacement: a longitudinal analysis of Medicare fee-for-service beneficiaries. *Circulation*. 2013;127:1870-1876.
16. Delling FN, Vasan RS. Epidemiology and pathophysiology of mitral valve prolapse: new insights into disease progression, genetics, and molecular basis. *Circulation*. 2014;129:2158-2170.
17. Verma S, Siu SC. Aortic dilatation in patients with bicuspid aortic valve. *N Engl J Med*. 2014;370:1920-1929.
18. Iung B, Rodes-Cabau J. The optimal management of anti-thrombotic therapy after valve replacement: certainties and uncertainties. *Eur Heart J*. 2014;35:2942-2949.
19. Huang G, Schaff HV, Sundt TM, et al. Treatment of obstructive thrombosed prosthetic heart valve. *J Am Coll Cardiol*. 2013;62:1731-1736.
20. Ruiz CE, Jelnin V, Kronzon I, et al. Clinical outcomes in patients undergoing percutaneous closure of periprosthetic paravalvular leaks. *J Am Coll Cardiol*. 2011;58:2210-2217.

REVIEW QUESTIONS

1. Each of the following findings indicates severe valvular heart disease except

A. in mitral stenosis, the S_2 opening snap interval is 0.06 seconds.
B. in aortic stenosis, there is a harsh systolic murmur peaking in mid-systole.
C. in aortic stenosis, there is a soft single second sound.
D. in aortic regurgitation, an Austin Flint murmur is heard.
E. in mitral regurgitation, an S_3 is heard.

Answer: B The murmur of severe aortic stenosis peaks late in systole, not in midsystole. An S_2-OS interval of 0.06 sec is short, indicating high left atrial pressure and severe mitral stenosis. Severe aortic regurgitation impinges on the mitral valve, thereby causing an Austin Flint murmur. A severely calcified and stenotic aortic valve produces little sound when it closes, so S_2 will be soft and single.

2. Pick the best choice regarding the hemodynamic overload of valvular heart disease.

A. Aortic regurgitation places a pure volume overload on the left ventricle.
B. Aortic stenosis produces a combined pressure and volume overload on the left ventricle.
C. Aortic regurgitation places a combined pressure and volume overload on the left ventricle.
D. Mitral regurgitation places a combined pressure and volume overload on the left ventricle.
E. Tricuspid regurgitation places a combined pressure and volume overload on the right ventricle.

Answer: C The increased total stroke volume in aortic regurgitation increases systolic blood pressure, thereby placing a combined pressure and volume overload on the left ventricle. Whereas aortic stenosis causes a pure left ventricle pressure overload, mitral and tricuspid regurgitation place a pure volume overload on the left ventricle and right ventricle respectively.

3. Indications for valve surgery or other valve intervention include all but which of the following?

A. A patient with syncope and an aortic valve area of 0.9 cm^2.
B. An asymptomatic patient with severe mitral regurgitation and an ejection fraction of 0.55.
C. An asymptomatic patient with mitral stenosis and a pulmonary artery pressure of 60 mm Hg.
D. An asymptomatic patient with an Austin Flint murmur and an ejection fraction that has decreased from 0.60 to 0.45.
E. A patient with functional mitral regurgitation, class II symptoms, and a left bundle branch block.

Answer: E A patient with functional mitral regurgitation should undergo evaluation for cardiac resynchronization therapy, which might substantially improve his or her mitral regurgitation. A patient with aortic stenosis has classic symptoms that require aortic valve replacement. Pulmonary hypertension in mitral stenosis worsens prognosis, so further delay is contraindicated. The other two patients have developed left ventricular dysfunction requiring prompt surgical intervention to prevent further deterioration.

4. The best defined need for medical therapy in valve disease is

A. the use of vasodilators in chronic aortic regurgitation.
B. the use of statins in aortic stenosis.
C. the use of vasodilators in asymptomatic aortic regurgitation.
D. the use of vasodilators in asymptomatic mitral regurgitation.
E. the use of warfarin in patients with mitral stenosis, atrial fibrillation, and a remote history of peptic ulcer disease.

Answer: E The risk of stroke in the patient with mitral stenosis far outweighs the small risk of gastrointestinal bleeding. Statins in aortic stenosis and vasodilators in asymptomatic regurgitant disease have not been proven to be effective.

5. The least desirable option for the treatment of severe valvular heart disease is

A. mitral repair for mitral regurgitation.
B. balloon valvotomy for mitral stenosis.
C. mitral valve replacement for a patient with posterior leaflet prolapse.
D. a pulmonary autograft for a young patient with aortic stenosis.
E. a tricuspid valve repair for tricuspid regurgitation.

Answer: C Mitral valve replacement for a patient with simple posterior leaflet prolapse that can easily be repaired is an error in management because mitral valve repair is far superior to mitral valve replacement. Balloon valvotomy is preferred over surgery in most cases of mitral stenosis. A pulmonary autograft is a good choice in a young patient in whom lifelong anticoagulation with a mechanical valve may be undesirable.

76

INFECTIVE ENDOCARDITIS

VANCE G. FOWLER, JR., ARNOLD S. BAYER, AND LARRY M. BADDOUR

DEFINITION

Infective endocarditis is defined as an infection, usually bacterial, of the endocardial surface of the heart. Infective endocarditis affects primarily the cardiac valves, although in some cases, the septa between the chambers, the mural endocardium, or cardiovascular implantable electronic devices (e.g., pacemakers) may be involved. Traditionally, infective endocarditis was categorized as "acute" or "subacute," based on the duration of symptoms before presentation. Typically, acute infective endocarditis was caused by *Staphylococcus aureus*, and subacute infective endocarditis was caused by viridans group streptococci. However, these categories have proven to be unreliable. A classification that considers the causative organism and the involved valve is much more clinically relevant.

EPIDEMIOLOGY

The true incidence of infective endocarditis is difficult to determine because of the different criteria for diagnosis and methods of reporting. An analysis based on strict case definitions reveals that only a relatively small proportion (≈20%) of clinically diagnosed cases are categorized as "definite." In 10 large surveys, infective endocarditis accounted for approximately one case per 1000 U.S. hospital admissions, with a range of 0.16 to 5.4 cases per 1000 admissions. Estimates from the American Heart Association (AHA) place the incidence of infective endocarditis in the United States at 10,000 to 20,000 new cases per year. Recent data also suggest that rates of *S. aureus* infective endocarditis have increased significantly.

Men are more commonly affected than women (mean male : female ratio, 1.7 : 1 in 18 large series). However, in patients younger than 35 years, more cases occur in women. More than 50% of patients with infective endocarditis in the U.S. are now older than 50 years of age, due in part to the low incidence of acute rheumatic heart disease (Chapter 290), the low subsequent development of rheumatic heart disease, and a simultaneous rise in the prevalence of degenerative valvular heart disease in the aging population.

Although some patients have no clearly definable risk factor for endocarditis, cardiac conditions that cause turbulent flow at the endocardial surface or across a valve (Chapter 75) predispose patients to infective endocarditis (Table 76-1). The most commonly affected valves in descending order of prevalence are the mitral valve only, the aortic valve, the mitral and aortic valves together, the tricuspid valve, mixed right- and left-sided infection, and the pulmonic valve.

Historically, rheumatic heart disease with valvular dysfunction was the most common underlying condition, although its contribution has diminished in the antibiotic era, especially in developed countries. Degenerative valvular disease is also associated with infective endocarditis, particularly in elderly patients; the increasing relevance of senile calcification as a risk factor is reflected in the increasing proportion of aortic valve involvement in infective endocarditis. Most significant congenital heart defects (Chapter 69) confer an increased risk of infective endocarditis, particularly complex cyanotic disease such as single-ventricle states, transposition of the great vessels, and tetralogy of Fallot. Similarly, surgically constructed pulmonary–systemic shunts and ventricular septal defects are risk factors associated with infective endocarditis.

Mitral valve prolapse is currently the most common underlying cardiac condition in patients with infective endocarditis, a statistic that reflects its prevalence in the general population (4%). Notably, mitral valve prolapse is a risk only in patients with thickened mitral leaflets or significant regurgitation, in which case the risk of endocarditis increases by about 10-fold over that of the general population. In addition, patients with hypertrophic cardiomyopathy are at increased risk of infective endocarditis, particularly in the presence of outflow obstruction. Finally, previous endocarditis is among the highest risk factors for subsequent infective endocarditis cases.

Prosthetic cardiac valves represent an important risk factor for infective endocarditis. More than 150,000 heart valves are implanted annually worldwide, and prosthetic valve infective endocarditis develops in 1% to 4% of prosthetic valve recipients in the first year after valve replacement and in approximately 0.8% of recipients annually thereafter. Mechanical prosthetic valves may initially be more susceptible to infective endocarditis, but bioprosthetic valves are more likely to develop infective endocarditis after 1 year; overall, the rate is similar with either type of valve.

The incidence of infective endocarditis in injection drug users (Chapter 34) may be 30 times higher than in the general population and four times higher than in adults with rheumatic heart disease. In some areas of the United States, injection drug use is the most common predisposing cause of infective endocarditis in patients younger than 40 years. *S. aureus* is the predominant organism, and tricuspid valve involvement is noted in 78% of cases, mitral involvement in 24%, and aortic involvement in 8%. More than one valve is infected in approximately 20% of cases, and some of these infections are polymicrobial.

Health care–associated infective endocarditis arises primarily as a consequence of invasive therapies, including intravenous (IV) catheters, hyperalimentation lines, pacemakers, other cardiovascular implantable electronic devices, and hemodialysis devices.[1] In a recent prospective, multinational cohort study of more than 1600 non–drug-using patients with native valve endocarditis, more than one third of patients had health care–associated endocarditis, many of which originated in the community (e.g., in patients on outpatient hemodialysis). The emerging importance of health care–associated infective endocarditis in industrialized nations has also influenced the microbiology of the disease, with an increasing prevalence of *S. aureus* and a decreasing prevalence of viridans group streptococci in much of the industrialized world.

Cardiovascular electronic devices can become infected at the time they are implanted, particularly if patients develop complications, such as a hematoma, at the incision site or need the device to be revised or replaced. Other factors that increase the risk of device infection include older age, comorbid conditions (particularly dialysis-dependent renal failure), and a larger number of device leads.[1,2]

Systemic medical conditions predispose patients to the development of infective endocarditis. For example, HIV infection is an independent risk factor for the development of infective endocarditis in injection drug users, with the risk increasing as the CD4 count decreases. Vascular catheter–related bacteremia is an important risk factor for nosocomial infective endocarditis. Patients with end-stage renal disease, particularly those receiving long-term hemodialysis, and patients with diabetes mellitus are also at increased risk, presumably because of the recurrent vascular access infections associated with the former and the low-level immunosuppression associated with both conditions.

TABLE 76-1 PREDISPOSING CONDITIONS ASSOCIATED WITH INCREASED RISK OF ENDOCARDITIS

MORE COMMON	LESS COMMON
Mitral valve prolapse	Rheumatic heart disease†
Degenerative valvular disease	Idiopathic hypertrophic subaortic stenosis
Intravenous drug use*	Pulmonary–systemic shunts*
Prosthetic valve*	Coarctation of the aorta
Congenital (valvular heart disease or ventricular septal defect)	Previous endocarditis*
	Complex cyanotic congenital heart disease*

*Indicates conditions with highest risk for endocarditis.
†Still common in developing countries.

PATHOBIOLOGY

Experimental models of infective endocarditis have demonstrated that the disease follows a predictable sequence: endocardial damage, aggregation of platelets and fibrin to create a sterile vegetation, transient bacteremia resulting in seeding of the vegetation, microbial proliferation on and invasion of the endocardial surface, and metastatic infection to visceral organs (e.g., kidneys, spleen) and brain.

Most cases of infective endocarditis begin with a damaged endocardial surface. Damage to the endocardium may be caused by a number of factors, ranging from inflammatory (e.g., rheumatic fever) to congenital (e.g., mitral valve prolapse) to senile degeneration and calcification; indeed, any excessive

turbulence or high-pressure gradient can cause injury to the nearby endocardium. Next, fibrin-platelet aggregates develop at the site of damage to form sterile vegetations, also termed *nonbacterial thrombotic endocarditis*. Nonbacterial thrombotic endocarditis may occur spontaneously in patients with systemic illnesses (e.g., the marantic endocarditis of malignancy [Chapter 60] or other wasting diseases or Libman-Sacks endocarditis in systemic lupus erythematosus [Chapter 266]). When transient bacteremia occurs—for example, as a result of distant infection or gingival manipulations—the previously sterile vegetation may be seeded. Some bacterial species, such as staphylococci and streptococci, are more avidly adherent to vegetations and better able to evade innate endovascular host defenses, so they more frequently cause endocarditis. The bacteria then proliferate within the vegetation and may ultimately achieve an organism load of 10^9 to 10^{11} colony-forming units per gram of tissue. Last, the surfaces of cardiac valves and vegetations are avascular, thereby making antibiotic therapy and healing difficult. Implantable cardiac devices disrupt the endothelial surface and predispose to infection and the formation of biofilms, particularly early after their placement and before endothelialization and fibrosis occur.

CLINICAL MANIFESTATIONS

History

The initial presentation of infective endocarditis varies enormously from patient to patient. Some cases develop acutely, with symptoms progressing rapidly over several days. Other cases develop insidiously and present with progressive but nonspecific symptoms for weeks or months. In patients suspected of having infective endocarditis, the initial history should include a complete review of systems, a travel history, and a thorough discussion of health-related behaviors such as illicit drug use and sexual activity. Most patients complain of fever and nonspecific constitutional symptoms such as fatigue, malaise, or weight loss. Nearly 50% of patients complain of musculoskeletal symptoms ranging from frank arthritis to diffuse myalgias. In about 5% to 10% of patients, low back pain is the chief complaint, even in the absence of osteomyelitis or epidural abscess. In IV drug users with tricuspid valve infective endocarditis and patients with venous or right heart implantable devices, endocarditis can present as pleuritic chest pain and multilobar pneumonia. Health care–associated infective endocarditis is more likely to be clinically occult and requires a high index of suspicion.

Physical Examination

A thorough physical examination should include a search for the peripheral stigmata (Table 76-2), which are very helpful when present but are less frequent now than in the past. Although fever is present in up to 90% of patients, it is less common in elderly patients and in patients with renal or heart failure. A widened pulse pressure should alert the clinician to the possibility of acute aortic insufficiency (Chapter 75). The skin and nails should be carefully examined for suggestive but nonspecific embolic phenomena such as petechiae (Fig. 76-1), Osler nodes, Janeway lesions, and splinter hemorrhages. Petechiae are most often found on the conjunctiva, palate, and extremities. Osler nodes are small, painful nodules found most often on the palmar surfaces of the fingers and toes; they frequently wax and wane (Fig. 76-2). Classically considered to be an immunologic phenomenon, Osler nodes may have an immune complex–mediated component but are most likely initiated by microemboli. Janeway lesions (Fig. 51-11 in Chapter 51) are hemorrhagic, nonpainful macules also found primarily on the palms and soles; they are embolic in origin and are less frequently noted than the other cutaneous stigmata. Splinter hemorrhages (see Fig. 51-11 in Chapter 51) are nonblanching, linear, brownish red lesions in the nail beds parallel to the direction of nail growth; they are nonspecific and may also be found in a significant percentage of hospitalized patients without infective endocarditis.

Funduscopic examination should be performed to look for Roth spots (see Fig. 423-28 in Chapter 423), chorioretinitis, or endophthalmitis; the latter two are present in a substantial proportion of cases of fungal endocarditis. A careful cardiac examination should be performed to detect any systolic or diastolic murmurs, especially new murmurs, or evidence of heart failure, which is an ominous sign. Of note, patients with health care–associated infective endocarditis are less likely than others to have a pathologic murmur on initial presentation. The abdomen should be examined for evidence of splenomegaly (Chapter 168), which is more common in patients with subacute endocarditis. Finally, a thorough neurologic examination should be performed, both to assess for any focal neurologic deficits and to serve as a baseline during the patient's hospital stay. The neurologic examination may demonstrate evidence of major vessel embolism, cranial nerve palsies, visual field defects, or generalized toxic-metabolic encephalopathy with altered mental status. Up to 15% to 20% of patients with endocarditis have a stroke before presentation or during the course of their disease. Patients with device-related infective endocarditis often have local signs and symptoms of infection at the site of implantation.

TABLE 76-2 PHYSICAL EXAMINATION AND LABORATORY FINDINGS IN INFECTIVE ENDOCARDITIS

FINDING	% OF CASES
Fever	96
Worsening of previous murmur	20
New murmur	48
Vascular embolic event	17
Splenomegaly	11
Splinter hemorrhages	8
Osler nodes	3
Janeway lesions	5
Roth spots	2
Elevated ESR	61
Hematuria	26
Positive rheumatoid factor	5
Abnormal chest radiography findings (effusion, infiltrate, septic emboli)	67-85 (right-sided infective endocarditis)

ESR = erythrocyte sedimentation rate.
Adapted from Murdoch DR, Corey GR, Hoen B, et al. Clinical presentation, etiology, and outcome of infective endocarditis in the 21st century: International Collaboration on Endocarditis—Prospective Cohort Study. *Arch Intern Med.* 2009;169:463-473.

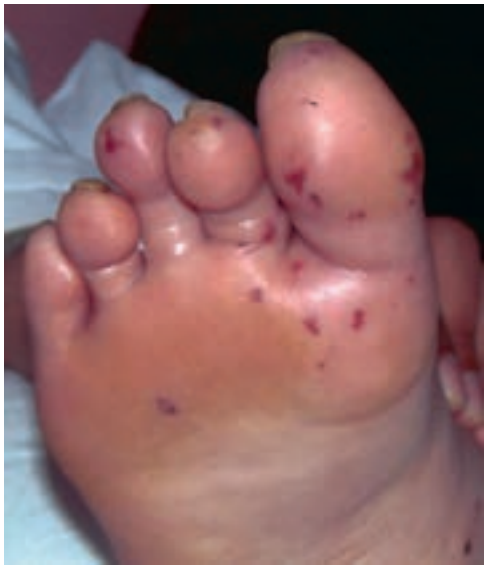

FIGURE 76-1. Petechiae in infective endocarditis.

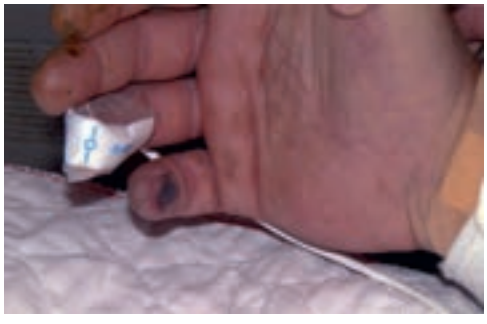

FIGURE 76-2. Osler node in infective endocarditis.

DIAGNOSIS

The "gold standard" for the diagnosis of infective endocarditis is culture of a pathologic organism from a valve or other endocardial surface. However, unless the patient undergoes valve replacement or postmortem examination, the diagnosis is made clinically. The most widely accepted clinical criteria are the modified Duke criteria (Table 76-3), which rely heavily on blood culture results and echocardiographic data and which have an estimated 76% to 100% sensitivity and 88% to 100% specificity, with a negative predictive value of at least 92%.[3,4]

Microbiology

At least three sets of blood cultures, each set consisting of one aerobic and one anaerobic bottle, should be obtained from separate sites, with careful attention to aseptic technique. Ideally, these sets should be collected at least 1 hour apart to document continuous bacteremia; however, when patients are critically ill, this approach may not be feasible.

Causative Organisms

About 90% of community-acquired, native valve infective endocarditis is caused by staphylococci, streptococci, or enterococci, which are normal inhabitants of the skin, oropharynx, and urogenital tract, respectively, and which have frequent access to the blood stream. These organisms express specific receptors for attachment and adherence to damaged endothelial surfaces. Streptococcal species (Chapter 290) are the most common cause of community-acquired infective endocarditis in patients with no history of injection drug use or health care contact. In patients with either of these latter epidemiologic risk factors, *S. aureus* (Chapter 288) is the predominant cause of infective endocarditis. Because of the emergence of health care contact as the predominant risk factor for blood stream infections, *S. aureus* is now the most common cause of infective endocarditis in most industrialized regions of the world.

Viridans group streptococci (Chapter 290) are the most common streptococci implicated in native valve endocarditis. This group of organisms, which normally inhabit the oropharynx, includes species such as *Streptococcus sanguis, Streptococcus mutans,* and *Streptococcus mitis*. Group B streptococci, β-hemolytic organisms that are also normal oropharyngeal and urogenital flora most frequently cause infective endocarditis in patients with cirrhosis or diabetes mellitus, as well as in injection drug users. In contrast, group A streptococci, although also β-hemolytic, rarely cause infective endocarditis. *Streptococcus gallolyticus,* a group D streptococcus (previously known as *Streptococcus bovis*), is now a leading cause of infective endocarditis in some parts of the world; for example, its incidence in France has increased significantly in recent years. Its presence in blood cultures should prompt endoscopic evaluation for adenocarcinoma of the colon or other malignant lesions of the gastrointestinal tract.

Pneumococcal endocarditis is decreasing in incidence but is quite fulminant when it occurs. It is associated with high morbidity and mortality rates, especially when it occurs as part of Austrian (or Osler) pneumococcal endocarditis triad of bacteremia, pneumonia, and meningitis.

S. aureus (Chapter 288) is the pathogen of primary concern among injection drug users and patients with health care contact. The clinical course of *S. aureus* endocarditis is typically acute, with a rapid progression over the course of days. Because about 10% to 15% of patients with *S. aureus* bacteremia have echocardiographic evidence of endocarditis even in the absence of classic stigmata, possible cardiac involvement should always be considered in any patient with *S. aureus* bacteremia but especially in patients with relapsing or persistent bacteremia or fever, community-acquired infection, or an implantable cardiac device. Coagulase-negative staphylococci are a relatively uncommon cause of native valve endocarditis but are important pathogens in prosthetic valve endocarditis.

Enterococcal bacteremia is far more common, particularly in hospitalized patients, than enterococcal endocarditis. However, enterococci are responsible for a significant number of cases of both community-acquired and nosocomial endocarditis. In most cases, the source of the bacteremia is thought to be the genitourinary tract, and the presentation is usually subacute. Enterococcal endocarditis, as opposed to enterococcal bacteremia, is suggested by community acquisition of infection, the absence of a clear source of infection, preexistent valvular heart disease, and the absence of polymicrobial bacteremia. As in most enterococcal infections, the overwhelming majority of cases (>90%) are caused by *Enterococcus faecalis*.

The HACEK group of gram-negative organisms (*Haemophilus* spp., *Aggregatibacter* spp. [formerly *Actinobacillus actinomycetemcomitans*], *Cardiobacterium hominis, Eikenella corrodens,* and *Kingella* spp.) accounts for about 5% of endocarditis cases. Because these fastidious organisms usually grow in blood cultures within 7 days using current methods, prolonged incubation is no longer required to isolate HACEK strains. Many other gram-negative bacilli have been reported to cause infective endocarditis but are even more unusual. Traditionally, injection drug use has been regarded as the primary risk factor for enteric gram-negative bacterial endocarditis. However, recent experience from large multinational studies shows that health care contact, not injection drug use, is the most common risk factor for enteric gram-negative endocarditis.

Fungal endocarditis is often difficult to diagnose and treat; it is most commonly found in patients with a history of injection drug use, recent cardiac valve surgery, or prolonged use of indwelling vascular catheters, especially those used for total parenteral nutrition. The most common fungi found in infective endocarditis are *Aspergillus* and *Candida* spp. *Aspergillus* (Chapter 339) rarely grows in blood cultures and must usually be cultured from a pathologic specimen (either an embolic site or vegetation); by contrast, *Candida* spp. (Chapter 338) frequently grows in blood cultures. Mortality is very high, and valve replacement surgery is usually necessary for fungal endocarditis.

TABLE 76-3 MODIFIED DUKE CRITERIA FOR THE DIAGNOSIS OF INFECTIVE ENDOCARDITIS

MAJOR CRITERIA

1. Blood culture positive
   a. Typical organism (α-hemolytic streptococcus, *Streptococcus bovis*, HACEK organisms, or community-acquired *Staphylococcus aureus* or enterococcus without a primary focus) from 2 separate blood cultures
   Or
   b. Persistent bacteremia with any organism (two positive cultures >12 hr apart or three positive cultures or a majority of ≥4 cultures positive >1 hr apart)
   Or
   c. Bacteremia with *S. aureus*, regardless of whether the bacteremia was nosocomially acquired or whether a removable focus of infection is found
2. Evidence of endocardial involvement
   a. Echocardiographic findings: mobile mass attached to valve or valve apparatus, abscess, or new partial dehiscence of prosthetic valve
   b. New valvular regurgitation
3. Serology: single positive blood culture for *Coxiella burnetii* or antiphase 1 IgG antibody titer >1 : 800

MINOR CRITERIA

1. Predisposing condition: IV drug use or predisposing cardiac condition
2. Fever ≥38° C
3. Vascular phenomena: arterial embolism, septic pulmonary emboli, mycotic aneurysm, intracranial hemorrhage, conjunctival hemorrhages, Janeway lesions
4. Immunologic phenomena: glomerulonephritis, Osler nodes, Roth spots, rheumatoid factor
5. Echocardiogram findings consistent with endocarditis but not meeting major criteria
6. Microbiologic evidence: positive blood cultures not meeting major criteria or serologic evidence of active infection consistent with endocarditis

DEFINITIVE INFECTIVE ENDOCARDITIS

1. Pathologically proven infective endocarditis
   Or
2. Clinical criteria meeting
   a. Two major criteria *or*
   b. One major and one minor criteria *or*
   c. Three minor criteria

POSSIBLE INFECTIVE ENDOCARDITIS

Findings that fall short of definitive infective endocarditis but do not reject it

REJECTED INFECTIVE ENDOCARDITIS

1. Firm alternative diagnosis *or*
2. Resolution of infective endocarditis syndrome with antibiotic therapy for ≤4 days
   or
3. No pathologic evidence of infective endocarditis at surgery or autopsy with antibiotic therapy for ≤4 days

HACEK = *Haemophilus* spp., *Aggregatibacter* spp. (formerly *Actinobacillus actinomycetemcomitans*), *Cardiobacterium hominis, Eikenella corrodens,* and *Kingella* spp.; IgG = immunoglobulin G; IV = intravenous.
Adapted from Li JS, Sexton DJ, Mick N, et al. Proposed modifications to the Duke criteria for the diagnosis of infective endocarditis. *Clin Infect Dis.* 2000;30:633-638.

Prosthetic valve endocarditis can be classified into one of two groups based on the time between valve surgery and disease onset: *early* (<2 months after surgery) and *late* (>2 months) (Table 76-4). Staphylococci, particularly *S. aureus*, predominate during the early period, when most episodes of infective endocarditis are thought to be related to perioperative infection. In the late period, the spectrum of organisms becomes more akin to that of community-acquired native valve disease, in which *S. aureus* and viridans group streptococci predominate. Of note, among the coagulase-negative staphylococci, oxacillin resistance can be seen in these late cases.

Staphylococcal species account for the large majority (≥70%) of implantable cardiac device infections. The prevalence of oxacillin resistance among *S. aureus* strains varies from study to study but is generally in the 30% to 50% range.

Endocarditis with Negative Blood Cultures

In most patients with infective endocarditis who have not received previous antibiotic therapy, every blood culture is positive because the bacteremia of endocarditis is continuous. Blood cultures are truly negative in fewer than 5% of cases of endocarditis; however, prior antibiotic administration may decrease the yield of blood cultures by up to 35%. Accordingly, most "culture-negative" cases of endocarditis occur in patients who have recently received antimicrobial agents. These cases are probably caused by the same organisms responsible for most native valve endocarditis; viridans group streptococci and the HACEK organisms are the most likely suspects because they are much more fastidious than staphylococci and enterococci and are therefore more likely to be affected by previous antibiotic administration. Ultimately, however, when blood cultures are negative and endocarditis is suspected, especially when a history of recent antimicrobial treatment is lacking, consideration should be given to fastidious organisms, fungi, and noncultivatable organisms (Table 76-5), particularly when the patient's history suggests exposure to farm animals or unpasteurized milk (*Coxiella burnetii, Brucella* spp.), cats (*Bartonella henselae*), body lice (*Bartonella quintana*), or birds (*Chlamydia psittaci*). It is important to notify the microbiology laboratory that endocarditis is suspected because special culture techniques can increase the yield for the HACEK species, nutritionally variant streptococci (*Abiotrophia* and *Granulicatella* spp.), *Brucella* spp., *Legionella* spp., and some fungi. The traditional practice of holding blood cultures for 2 to 4 weeks is no longer required routinely. Specific serologic tests can diagnose endocarditis related to *C. burnetii* (the agent of Q fever), *Brucella* spp., *Bartonella* spp., and *C. psittaci*. *Tropheryma whippelii*, the etiologic agent in Whipple disease, and multiple other organisms may be diagnosed by polymerase chain reaction. Histopathologic features of resected tissue also can provide clues in the etiologic diagnosis of culture-negative endocarditis. If the search for a causative organism is fruitless, noninfectious causes such as marantic or Libman-Sacks endocarditis and atrial myxoma (Chapter 60) should be considered.

TABLE 76-4 CAUSES OF PROSTHETIC VALVE ENDOCARDITIS*

EARLY (<2 mo POSTOPERATIVELY)	LATE (>2 mo POSTOPERATIVELY)
Staphylococcus aureus	Coagulase-negative staphylococci
Coagulase-negative staphylococci	*Staphylococcus aureus*
Gram-negative bacilli	Viridans group streptococci
Enterococci	Enterococci
Fungi	
Diphtheroids	

*Listed in order of relative frequency.
Adapted from Wang A, Athan E, Pappas PA, et al. Contemporary clinical profile and outcome of prosthetic valve endocarditis. *JAMA*. 2007; 297:1354-1361.

Laboratory Findings

Initial laboratory tests should include a complete blood count with differential, serum electrolytes, measurement of renal function, and urinalysis. Most patients with subacute infective endocarditis have the serum iron profile of anemia of chronic disease (Chapter 159). The white blood cell count is frequently elevated in acute infective endocarditis, particularly if *S. aureus* is the causative organism, but may not be elevated in more subacute forms. Microscopic hematuria is common, as is proteinuria.

The chest radiograph is abnormal—demonstrating consolidation, atelectasis, pleural effusion, or clear septic emboli—in the overwhelming majority of patients with right-sided endocarditis. In others, it may show evidence of heart failure. The electrocardiogram (ECG) should be carefully examined for evidence of atrioventricular conduction blocks, especially a prolonged PR interval (see Figs. 64-5 through 64-9 in Chapter 64), suggestive of an aortic ring abscess or frank myocardial infarction (see Figs. 73-1 and 73-2 in Chapter 73).

Rheumatoid factor, which is an ancillary test that has been included in the modified Duke criteria as a "minor criterion" in the category of "immunologic phenomenon," may be positive in subacute or chronic endocarditis. Other ancillary tests, such as the erythrocyte sedimentation rate, the C-reactive

TABLE 76-5 ORGANISMS CAUSING "CULTURE-NEGATIVE" ENDOCARDITIS*

ORGANISM	EPIDEMIOLOGY	DIAGNOSTIC TESTS
HACEK spp.	Mostly oral flora, so often history of periodontal disease	May require up to 7 days to grow
Nutritionally variant streptococci	Slow and indolent course	Supplemented culture media or growth as satellite colonies around *Staphylococcus aureus* streak
Coxiella burnetii (Q fever)	Worldwide; exposure to raw milk, farm environment, or rural areas	Serologic tests (high titers of antibody to both phase 1 and phase 2 antigens); also PCR on blood or valve tissue
Brucella spp.	Ingestion of contaminated milk or milk products; close contact with infected livestock	Bulky vegetations usually seen on echocardiography; blood cultures positive in 80% of cases with incubation time of 4-6 wk; lysis-centrifugation technique may expedite growth; serologic tests are available
Bartonella spp.	*Bartonella henselae:* transmitted by cat scratch or bite or by cat fleas *Bartonella quintana:* transmitted by human body louse; predisposing factors include homelessness and alcohol abuse	Serologic testing (may cross-react with *Chlamydia* spp.); PCR of valve or emboli is best test; lysis-centrifugation technique may be useful
Chlamydia psittaci	Exposure to birds	Serologic tests available, but must exclude *Bartonella* spp. because of cross-reactivity; monoclonal antibody direct stains on tissue may be useful; PCR now available
Tropheryma whippelii (Whipple disease)	Systemic symptoms include arthralgias, diarrhea, abdominal pain, lymphadenopathy, weight loss, CNS involvement; however, endocarditis may be present without systemic symptoms	Histologic examination of valve with PAS stain; valve cultures may be done using fibroblast cell lines; PCR on vegetation material
Legionella spp.	Contaminated water distribution systems; often nosocomial outbreaks; usually prosthetic valves	Lysis-centrifugation technique; also periodic subcultures onto buffered charcoal yeast extract medium; serologic tests and PCR available
Aspergillus and other noncandidal fungi	Prosthetic valve	Lysis-centrifugation technique; also culture and direct examination of any emboli

*Listed in approximate order of relative frequency.
CNS = central nervous system; HACEK = *Haemophilus* spp., *Aggregatibacter* spp., *Cardiobacterium hominis, Eikenella corrodens,* and *Kingella* spp.; PAS = periodic acid–Schiff; PCR = polymerase chain reaction.

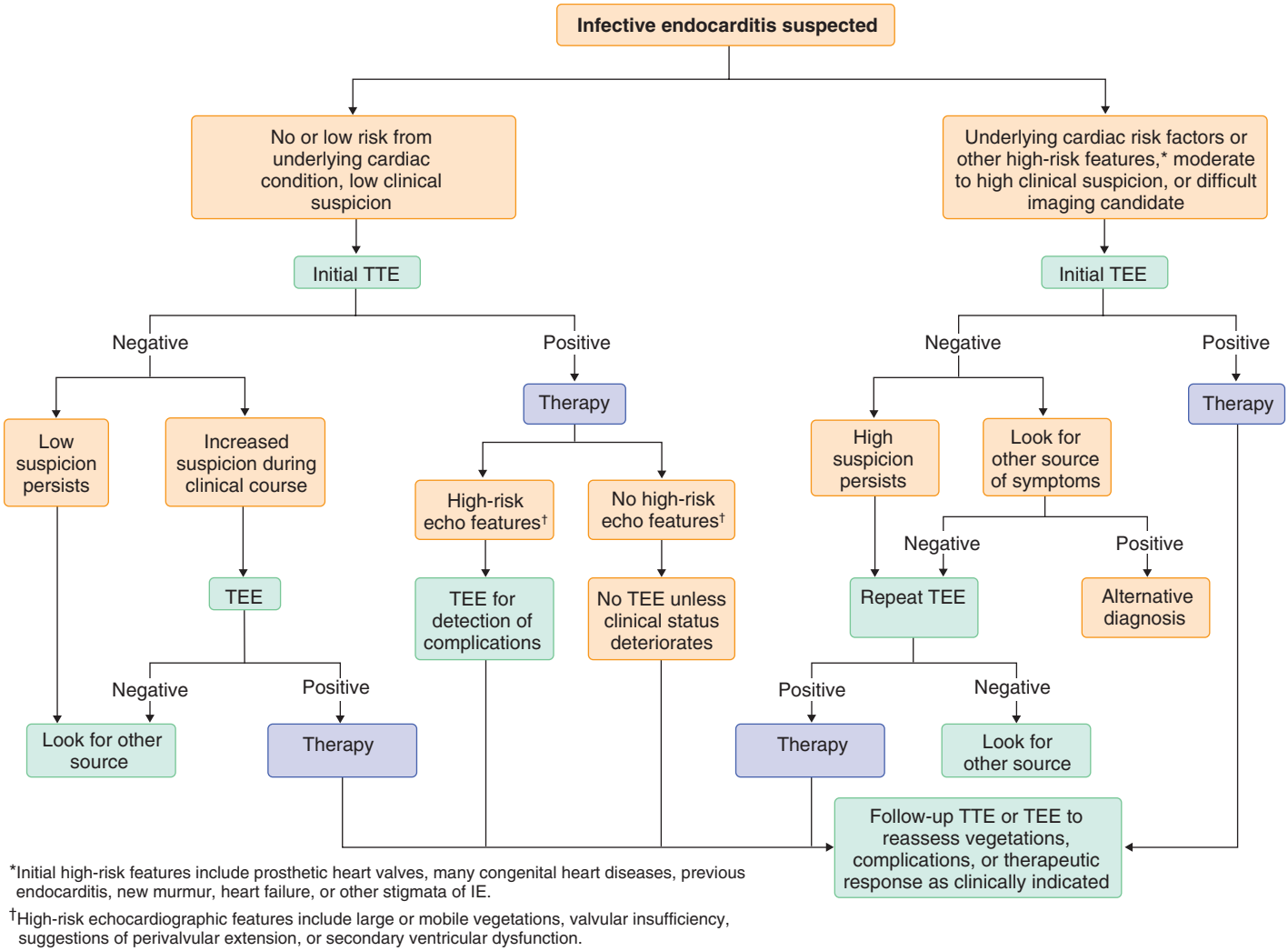

FIGURE 76-3. **Algorithm for the diagnostic use of echocardiography (echo) in suspected cases of infective endocarditis (IE). TEE = transesophageal echocardiography; TTE = transthoracic echocardiography. (Adapted from Bayer AS, Bolger AF, Taubert KA, et al. Diagnosis and management of infective endocarditis and its complications. *Circulation.* 1998; 98:2936-2948.)**

protein level, and the procalcitonin level, are generally not helpful in establishing an endocarditis diagnosis.

Echocardiography

Both transthoracic echocardiography (TTE) and transesophageal echocardiography (TEE) (Chapter 55) are highly specific tests (≈98%) when used as part of the diagnostic evaluation of suspected endocarditis. By contrast, TEE has a much higher sensitivity (90%-95%) in this setting than TTE (48%-63%). In most cases in which endocarditis is a serious diagnostic consideration, the evaluation should begin with TEE because negative TTE findings are not sensitive enough to exclude endocarditis (Fig. 76-3). Because TEE is the only relatively noninvasive means of detecting perivalvular extension of infection, any patient with a new conduction system abnormality or persistent fever—clinical predictors of perivalvular extension—should be evaluated with TEE. Likewise, TEE is strongly preferred in the evaluation of suspected prosthetic valve– or device-related endocarditis, although bland clots can occur on the leads of 5% to 10% of patients with intracardiac devices, and the finding of a "vegetation" on a lead is not specific for infection. The high sensitivity of TEE in detecting valvular vegetations on native valves also may be used in combination with clinical parameters (e.g., prompt resolution of bacteremia and defervescence) to support the clinical decision to abbreviate therapy in patients with vascular catheter–associated *S. aureus* bacteremia.

A negative TEE result has a negative predictive value of about 95%. Nevertheless, when clinical suspicion of endocarditis is high and the initial TEE is negative, repeat TEE in 7 to 10 days may reveal the diagnosis. If TEE is unavailable, technically impossible, or considered too invasive by the patient, it is reasonable to begin with TTE.

TREATMENT

Rx

Definitive antibiotic treatment of infective endocarditis (Table 76-6) is guided by antimicrobial susceptibility testing of the responsible pathogen isolated from clinical cultures. Although it is often advisable to begin empirical treatment before definitive culture results are available, not all patients who are admitted because of possible endocarditis necessarily need to be treated empirically. Patients who are clinically stable, with a subacute presentation syndrome, and without evidence of heart failure or other end-organ complications, can be closely observed without antibiotics so that serial blood cultures can be obtained. Likewise, such stable patients who were started on empiric antibiotics before hospitalization and before blood was drawn for cultures can discontinue antibiotics so that blood cultures can be obtained, preferably as long as possible after stopping the antibiotics. By contrast, acutely ill patients, patients with evidence of complications of endocarditis, and patients who are at high risk for endocarditis (e.g., prosthetic valve recipients) should be treated empirically with antibiotics pending culture results. In most cases of infective endocarditis, an infectious diseases specialist can assist in guiding the diagnostic evaluation and designing an appropriate antibiotic regimen.

Either of two regimens provides appropriate empirical coverage for patients with suspected native valve endocarditis: nafcillin (or oxacillin)–penicillin–gentamicin or vancomycin–gentamicin (Table 76-7). Nafcillin–penicillin–gentamicin is suitable in most cases of suspected native valve endocarditis because it provides optimal coverage for viridans group streptococci, methicillin-sensitive staphylococci, enterococci, and HACEK organisms. Some experts recommend a regimen of nafcillin–ceftriaxone–penicillin–gentamicin to cover for HACEK isolates that produce β-lactamase. If methicillin-resistant *S. aureus* (MRSA) is an important consideration, as in injection drug users and patients with health care contact, empirical therapy should consist of vancomycin–ceftriaxone–gentamicin. This regimen is also acceptable for

TABLE 76-6 DEFINITIVE THERAPY OF BACTERIAL ENDOCARDITIS

ORGANISM AND REGIMEN*	COMMENTS
PCN-SUSCEPTIBLE VIRIDANS STREPTOCOCCI (MIC ≤0.1 μg/mL) AND *STREPTOCOCCUS GALLOLYTICUS (formerly S. bovis)*	
1. PCN 2-3 million units IV q4h × 4 wk	1. Also effective for other PCN-susceptible nonviridans streptococci
2. Ceftriaxone 2 g IV qd × 4 wk	2. Uncomplicated infection with viridans streptococci in a candidate for outpatient therapy; also for those with PCN allergy
3. PCN 2-3 million units IV q4h × 2 wk plus gentamicin 1 mg/kg IV q8h × 2 wk	3. Uncomplicated infection with none of the following features: renal insufficiency, eighth cranial nerve deficit, prosthetic valve infection, CNS complications, severe heart failure, age >65 yr; also not acceptable for nutritionally variant streptococci
4. PCN 2-4 million units IV q4h × 4 wk plus gentamicin 1 mg/kg IV q8h for at least 2 wk with ID input	4. Nutritionally variant strain; for prosthetic valve, give 6 wk of PCN
5. Vancomycin 15-20 mg/kg IV q8-12h × 4 wk	5. For PCN allergy; goal trough level of 15-20 mg/L
RELATIVELY PCN-RESISTANT VIRIDANS STREPTOCOCCI (MIC 0.12-<0.5 μg/mL)	
1. PCN 4 million units IV q4h × 4 wk plus gentamicin 1 mg/kg IV q8h × 2 wk	—
2. Vancomycin 15-20 mg/kg IV q8-12h × 4 wk	2. For PCN allergy or to avoid gentamicin; goal trough level of 15-20 mg/L
ENTEROCOCCI† AND PCN-RESISTANT VIRIDANS STREPTOCOCCI (PCN MIC >0.5 μg/mL)	
1. PCN‡ 18-30 million units IV per day in divided doses × 4-6 wk or ampicillin 12 g/24 hr IV in 6 equally divided doses plus gentamicin 1 mg/kg IV q8h × 4-6 wk	1. Increase duration of both drugs to 6 wk for prosthetic valve infection or symptoms >3 mo in enterococcal infection
2. Vancomycin 15-20 mg/kg IV q8-12h × 6 wk plus gentamicin 1 mg/kg q8h × 6 wk§	2. For PCN allergy; PCN desensitization is also an option; high risk of nephrotoxicity with this regimen
3. Ampicillin 12 g/24 h IV in 6 equally divided doses plus ceftriaxone 2 g IV q12h	3. PCN-susceptible, aminoglycoside-resistant enterococci or patients who have significant underlying renal disease
STAPHYLOCOCCUS AUREUS	
1. Nafcillin 2 g IV q4h × 4-6 wk	1. Methicillin-susceptible strain; omit gentamicin if significant renal insufficiency
2. Vancomycin 15-20 mg/kg IV q8-12h × 6 wk	2. PCN allergy (immediate hypersensitivity or anaphylaxis) or MRSA
3. Nafcillin 2 g IV q4h × 2 wk plus gentamicin 1 mg/kg IV q8h × 2 wk	3. Methicillin-susceptible strain; 2-wk regimen only for use in IV drug abusers with only tricuspid valve infection, no renal insufficiency, and no extrapulmonary infection
4. Nafcillin 2 g IV q4h × >6 wk plus gentamicin 1 mg/kg IV q8h × 2 wk plus rifampin 300 mg PO/IV q8h × ≥6 wk	4. Prosthetic valve infection with methicillin-susceptible strain; use vancomycin instead of nafcillin for MRSA
5. Cefazolin 2 g IV q8h × 4-6 wk	5. PCN allergy other than immediate hypersensitivity
6. Daptomycin 6 mg/kg IV qd × 14-42 days	Daptomycin is FDA-approved for treatment of right-sided *S. aureus* infective endocarditis; for adults, some experts recommend 8-10 mg/kg IV
COAGULASE-NEGATIVE STAPHYLOCOCCI, PROSTHETIC VALVE INFECTION	
Vancomycin 15-20 mg/kg IV q8-12h × >6 wk plus gentamicin 1 mg/kg IV q8h × 2 wk plus rifampin 300 mg PO/IV q8h × >6 wk	Can substitute nafcillin in above doses for vancomycin if isolate is methicillin sensitive
HACEK STRAINS	
1. Ceftriaxone 2 g IV qd × 4 wk; 6 wk for prosthetic valves	—
2. Ampicillin–sulbactam 3 g IV q6h × 4 wk; 6 wk for prosthetic valves	2. HACEK strains increasingly may produce β-lactamase
NON-HACEK GRAM-NEGATIVE BACILLI	
Enterobacteriaceae	
Extended-spectrum PCN or cephalosporin plus aminoglycosides for susceptible strains	Treat for a minimum of 6-8 wk; some species exhibit inducible resistance to third-generation cephalosporins; valve surgery is required for most patients with left-sided endocarditis caused by gram-negative bacilli; consultation with a specialist in infectious diseases is recommended
Pseudomonas aeruginosa	
High-dose tobramycin (8 mg/kg/day IV or IM in once-daily doses) with maintenance of peak and trough concentrations of 15 to 20 μg/mL and ≤2 μg/mL, respectively, in combination with an extended-spectrum PCN (e.g., ticarcillin, piperacillin, azlocillin); ceftazidime, cefepime, or imipenem in full doses; or imipenem	Treat for a minimum of 6-8 wk; early valve surgery usually required for left-sided *Pseudomonas* endocarditis; consultation with a specialist in infectious diseases is recommended
Fungi	
Treatment with a parenteral antifungal agent (usually a lipid-containing amphotericin B product, 3-5 mg/kg/day IV for at least 6 weeks) and valve replacement; Fluconazole, 400 mg daily PO is an alternative for susceptible yeasts; other azoles, such as voriconazole, may be required for resistant yeasts or molds.	Long-term or lifelong suppressive therapy with PO antifungal agents often required; consultation with a specialist in infectious diseases is recommended

*Dosages are for patients with normal renal function; for those with renal insufficiency, adjustments must be made for all drugs except nafcillin, rifampin, and ceftriaxone. Gentamicin doses should be adjusted to achieve a peak serum concentration of approximately 3 μg/mL 30 min after dosing and a trough gentamicin level of <1 μg/mL.
†Enterococci must be tested for antimicrobial susceptibility. These recommendations are for enterococci sensitive to PCN, gentamicin, and vancomycin.
‡Ampicillin 12 g/day can be used instead of PCN.
§The need to add an aminoglycoside has not been demonstrated for PCN-resistant streptococci.
HACEK = *Haemophilus* spp., *Aggregatibacter* spp. (formerly *Actinobacillus actinomycetemcomitans*), *Cardiobacterium hominis, Eikenella corrodens,* and *Kingella* spp.; IM = intramuscular; IV = intravenous; MIC = minimum inhibitory concentration; MRSA = methicillin-resistant *Staphylococcus aureus*; PCN = penicillin; PO = oral; q = every; qd = every day.
Adapted from Baddour LM, Wilson WR, Bayer AS, et al. Infective endocarditis: diagnosis, antimicrobial therapy, and management of complications. *Circulation.* 2005;111:e394-e433.

TABLE 76-7 EMPIRICAL TREATMENT OF ENDOCARDITIS

CHARACTERISTICS OF PATIENTS	TREATMENT REGIMEN*
Native valve, community-acquired infection, MRSA unlikely	Nafcillin 2 g IV q4h plus penicillin 4 million units IV q4h plus gentamicin 1 mg/kg IV q8h
Any of the following: health care–associated infection or other reason to suspect MRSA; severe penicillin allergy	Vancomycin 15-20 mg/kg IV q8-12h† plus gentamicin 1 mg/kg IV q8h
Prosthetic valve	Vancomycin 15-20 mg/kg IV q8-12h[†] plus gentamicin 1 mg/kg IV q8h plus rifampin 300 mg PO/IV q8h

IV = intravenous; MRSA = methicillin-resistant *Staphylococcus aureus*; PO = oral; q = every.
*Dosages are for patients with normal renal function; for those with renal insufficiency, adjustments must be made for all drugs except nafcillin.
†Goal is trough level of 15-20 mg/L.

patients with a serious penicillin allergy. Patients with prosthetic valves should be empirically treated with vancomycin–gentamicin–rifampin for adequate coverage of the most important pathogens in this setting (MRSA, methicillin-sensitive staphylococci, and coagulase-negative staphylococci). Nearly 40% of patients with infective endocarditis related to implantable cardiovascular devices have concomitant valve involvement, predominantly tricuspid valve infection, with in-hospital and 1-year mortality rates of 15% and 23%, respectively. Device removal appears to reduce the mortality rate by about 50% (from about 40% to about 20%).

Treatment of Specific Organisms

When the organism is definitively identified, antibiotic treatment must be narrowed accordingly, and validated regimens should be followed[5] (see Table 76-6). More controversy exists over the treatment of unusual organisms, and consultation with infectious disease specialists is advisable in such circumstances.

Many regimens recommend consideration of low-dose gentamicin to provide antibacterial synergy with a low risk of toxicity. However, aminoglycoside toxicity is a significant risk in elderly patients and in patients with preexisting renal disease or hearing impairment; even low-dose gentamicin increases the likelihood of a decrease in creatinine clearance by about threefold. Among the organisms listed in Table 76-6, gentamicin is critical for cure only in enterococcal endocarditis. As a result of these risks and the minimal data supporting its benefit, initial low-dose gentamicin should not be routinely used.

In uncomplicated viridans group streptococcal endocarditis, outpatient therapy with once-daily ceftriaxone is as effective as more complex regimens, provided the patient has been observed in the hospital for the development of complications. The decision to administer antimicrobial therapy in the outpatient setting must, of course, take into account the patient's social situation, likelihood of compliance, and other risks involved with either an indwelling IV line or recurrent peripheral IV line placements.

Standard therapy for infective endocarditis caused by fully susceptible enterococci includes penicillin or ampicillin plus gentamicin. Although gentamicin is preferred over streptomycin, the choice of a specific aminoglycoside should be based on in vitro susceptibility testing. Nonrandomized data suggest that the duration of aminoglycoside therapy can be limited to 2 to 3 weeks in combination with either penicillin, ampicillin, or vancomycin or that aminoglycoside therapy can be avoided in favor of combination therapy with ampicillin plus high-dose (2 g IV every 12 hours) ceftriaxone. Optimal therapy for enterococci that are resistant to aminoglycosides or vancomycin is not well defined. Endocarditis caused by vancomycin-resistant enterococci may be treated with daptomycin, quinupristin–dalfopristin (7.5 mg/kg IV every 8 hours), or linezolid (600 mg orally or IV twice daily); however, clinical experience with these agents is limited. In this situation, relapse or failure rates are likely to be high, and many cases require surgical intervention (discussed later).

Semisynthetic penicillins, such as nafcillin, are advocated for endocarditis caused by methicillin-susceptible *S. aureus*. Cefazolin represents an alternative to semisynthetic penicillins in cases in which the latter are not tolerated or feasible to administer. Although vancomycin is recommended in patients who are allergic to β-lactams, the microbiologic and clinical cure rates are less than that of β-lactam therapy. In a recent randomized trial, daptomycin (6 mg/kg/day for 10 to 42 days, depending on the severity of infection) was as effective as either a semisynthetic antistaphylococcal penicillin or vancomycin for the treatment of *S. aureus* bacteremia and right-sided infective endocarditis caused by methicillin-susceptible *S. aureus* and MRSA, and this agent is now approved by the Food and Drug Administration for these indications. More recently, ceftaroline, a fifth-generation cephalosporin, has been successful in case series of patients with MRSA bacteremia and endocarditis.[6]

Rifampin or gentamicin can be added to either nafcillin or vancomycin for the treatment of prosthetic valve infection caused by methicillin-susceptible *S. aureus* or to MRSA, respectively. Gentamicin is administered for 2 weeks, and rifampin is given for the duration of either nafcillin or vancomycin therapy. Rifampin is never used as monotherapy because of the rapid development of resistance.

Fungal endocarditis is usually a consequence of extensive health care contact. Traditionally, fungal endocarditis was regarded as a primary indication for valvular surgery, and amphotericin B (Chapter 331) was considered the adjunctive treatment of choice. However, many patients with *Candida* endocarditis can be treated medically with azole-containing antimicrobial agents, with or without amphotericin. The management of fungal endocarditis should always involve the collaboration of an experienced infectious diseases specialist.

Zoonotic endocarditis is usually culture negative and most commonly caused by *Bartonella* spp. (Chapter 315), *C. burnetii* (Chapter 327), or *Brucella* species (Chapter 310). The treatments of choice for these fastidious pathogens are based on limited data, but documented *Bartonella* endocarditis is treated with doxycycline for 6 weeks plus gentamicin for the first 2 weeks.

In cases of presumed culture-negative endocarditis in which unusual organisms (see Table 76-5) and other infections have been reasonably excluded, an empirical course of treatment may be undertaken. In this situation, most authorities recommend a 4- to 6-week regimen of ceftriaxone alone, vancomycin–ceftriaxone–gentamicin, or vancomycin–gentamicin (if the clinical setting suggests a risk for enterococcal endocarditis). The vancomycin–ceftriaxone–gentamicin regimen provides optimal coverage for HACEK and *Abiotrophia* and *Granulicatella* spp. (formerly known as "nutritionally variant streptococci"), which are the two most common causes of nonzoonotic, culture-negative infective endocarditis.

Continuing Care of the Patient with Endocarditis

In addition to antibiotics, appropriate inpatient care includes surveillance for the development of complications. Widening of the pulse pressure should alert the clinician to the possible development of acute aortic insufficiency (Chapter 75). A careful cardiac examination should be performed on a daily basis to assess for new regurgitant murmurs.

Repeat echocardiography is recommended during therapy for patients with persistent fever, recurrent embolic events, a new murmur, widening of the pulse pressure, or signs or symptoms of heart failure. It is also recommended to screen for periannular complications, especially in prosthetic valve endocarditis. By comparison, repeat echocardiography is not routinely recommended if patients respond adequately to antimicrobial therapy, although serial echocardiography is usually suggested over the ensuing years to screen for long-term valvular dysfunction.

Routine serial ECGs are not recommended. ECG-documented conduction abnormalities are a late sign of perivalvular infections in patients with endocarditis; TEE is the screening method of choice if this complication is suspected.

Any new neurologic findings should prompt a search for evidence of central nervous system (CNS) complications such as embolic events, cerebral hemorrhage, mycotic aneurysm, or brain abscess. Renal function should be closely monitored so that antibiotic doses can be adjusted if necessary. If gentamicin is used for more than a few days, the patient should be alerted to watch for the signs and symptoms of vestibular or otic toxicity. Audiometric testing at baseline and periodically thereafter should be considered in patients at high risk for aminoglycoside-induced ototoxicity, including elderly patients, patients with preexisting renal dysfunction or hearing damage, patients receiving prolonged courses of gentamicin, and patients who also receive other potentially nephrotoxic agents. Serum gentamicin trough concentrations should also be assayed at regular intervals (e.g., twice weekly and more often if renal function is changing) and should be targeted for 1 to 3 μg/mL or less; higher concentrations should prompt either lower or less frequent dosing or both.

Follow-up blood cultures may be indicated toward the end of the first week of therapy in patients whose infective endocarditis is caused by organisms that commonly fail first-line treatment, such as *S. aureus* or aerobic gram-negative bacilli. Positive cultures in this setting might suggest the need to change therapy, search for metastatic abscesses, or repeat echocardiography, but negative cultures are reassuring.

Patients with infective endocarditis may continue to be febrile for some time after the institution of appropriate antibiotic treatment. About 50% of patients defervesce within 3 days of starting antibiotics, 75% by 1 week, and 90% by 2 weeks. Patients whose endocarditis is caused by *S. aureus,* aerobic gram-negative organisms, or fungi tend to defervesce more slowly than patients infected with other organisms. Prolonged fever (>1 week after the institution of appropriate antibiotics) should prompt repeat blood cultures. If such cultures are negative, several possibilities should be considered: myocardial abscess, extracardiac infection (e.g., mycotic aneurysm, psoas or splenic abscess, vertebral osteomyelitis, septic arthritis), immune complex–mediated tissue damage, or a complication of hospitalization and therapy (e.g., drug fever, nosocomial superinfection, pulmonary embolism). Appropriate studies might include TEE, computed tomography (CT) scan of the abdomen, bone

scan, and urinalysis with microscopy (to elicit evidence of interstitial nephritis). IV line sites should be carefully examined for evidence of infection, and indwelling central lines should be changed according to published guidelines.

Anticoagulation in individuals with infective endocarditis is controversial. Although new anticoagulation in the setting of native valve endocarditis does not appear to provide a benefit, continuing ongoing anticoagulation may be advisable. Some authorities recommend continuing anticoagulation in patients with mechanical prosthetic valve endocarditis. However, discontinuation of all anticoagulation for at least the first 2 weeks of antibiotic therapy is generally advised in patients with *S. aureus* prosthetic valve endocarditis who have experienced a recent CNS embolic event; this approach allows the thrombus to organize and potentially prevents the acute hemorrhagic transformation of embolic lesions. Reintroduction of anticoagulation in these patients must be cautious, and the international normalized ratio must be monitored carefully. The best option for patients with other indications for anticoagulation, such as deep vein thrombosis, major vessel embolization, or atrial fibrillation, is less clear and should be decided in a multidisciplinary fashion that balances the risks and benefits for each individual patient.

High-dose (325 mg/day) aspirin does not prevent embolic events and tends to increase the incidence of bleeding in patients with infective endocarditis. Whether a patient should remain on chronic, low-dose (81 mg) aspirin if they develop subsequent infective endocarditis is uncertain.

Complications

The complications of infective endocarditis can be divided into four groups for ease of classification: direct valvular damage and consequences of local invasion, embolic complications, metastatic infections from bacteremia, and immunologic phenomena. Local damage to the endocardium or myocardium may directly erode through the involved cardiac valve or adjacent myocardial wall, resulting in hemodynamically significant valvular perforations or intra- or extracardiac fistulae. Such local complications typically present clinically with the acute onset of heart failure and carry a poor prognosis, even with prompt cardiac surgery. Valve ring abscesses also require surgical intervention and are more frequent in patients with prosthetic valves. Although a conduction defect on ECG may suggest the diagnosis, TEE is the diagnostic technique of choice for detecting paravalvular abscess, valve perforation, or intracardiac fistulae. Frank myocardial abscesses are found in up to 20% of cases on autopsy, and *Aspergillus* endocarditis invades the myocardium in more than 50% of cases. Pericarditis is rare and is usually associated with myocardial abscess. Myocardial infarction (MI), thought to be caused by embolism of vegetative material into the coronary arteries, is seen in 40% to 60% of cases on autopsy, although most cases are clinically silent and lack characteristic ECG changes. However, up to 15% of elderly patients may present with clinical evidence of acute MI, with potentially disastrous complications if the MI is thought to be the primary event and the patient is given thrombolytic therapy. Heart failure is the leading cause of death in infective endocarditis, usually related to direct valvular damage.

Embolic events are less common now than in the preantibiotic era, but about 35% of patients have at least one clinically evident embolic event. In fungal endocarditis, the majority of patients have at least one embolic event, frequently with a large embolus. The presence of large (>10 mm), mobile vegetations on the echocardiogram, particularly when the anterior mitral valve leaflet is involved, predicts a high risk of embolic complications. In addition, patients may have frank infarction of cutaneous tissue from emboli. In addition to the skin, systemic emboli most commonly lodge in the kidneys, spleen, large blood vessels, or CNS. Vegetations of right-sided endocarditis usually embolize to the lungs and cause abnormalities on the chest radiograph, although occasionally such emboli reach the left-sided circulation via a patent foramen ovale.

Renal abscesses are rare in infective endocarditis; however, bland renal infarction is a frequent asymptomatic finding on abdominal CT scanning, seen in more than 50% of cases at autopsy. Similarly, splenic infarction occurs in up to 44% of autopsy-confirmed cases. Such emboli may be asymptomatic but also can cause left upper quadrant pain radiating to the left shoulder, sometimes as the presenting symptom of infective endocarditis. A splenic infarction that progresses to form an abscess can cause persistent fever or bacteremia, so such patients should undergo abdominal CT to search for this complication. Mycotic vascular aneurysms, which frequently occur at bifurcation points, may be clinically silent until they rupture (which may be months to years after apparently successful antibiotic treatment of infective endocarditis) and have been found in 10% to 15% of cases at autopsy. Whereas peripheral mycotic aneurysms require surgical resection, intracerebral aneurysms should be resected or managed with intravascular techniques (e.g., coils) if they bleed or if they are causing a mass effect.

Many patients may have evidence of cerebrovascular emboli, which have a predilection for the middle cerebral artery distribution and may be devastating. Most emboli to the CNS occur early in the course of the disease and are evident at the time of presentation or shortly thereafter. Embolic strokes may undergo hemorrhagic transformation, with a sudden worsening of the patient's neurologic status. Many patients with fungal endocarditis present with an embolic stroke or large emboli that occlude major vessels.

Some complications of infective endocarditis result when bacteremic seeding causes metastatic infection at a distant site. Patients may present with or develop osteomyelitis, septic arthritis, or epidural abscess. Purulent meningitis (Chapter 412) is a rare complication except in pneumococcal endocarditis, although many patients with *S. aureus* infective endocarditis who undergo lumbar puncture have a pleocytosis. Intracranial abscesses are uncommon in bacterial endocarditis but frequent in *Aspergillus* endocarditis; such a finding in the setting of culture-negative endocarditis should prompt the consideration of *Aspergillus* as an etiologic agent. Importantly, the finding of one metastatic complication of infective endocarditis does not exclude the possibility of additional sites of hematogenous infection, particularly in *S. aureus* endocarditis. Thus, the need for additional diagnostic evaluations should be guided by the patient's clinical course.

The immunologic phenomena of infective endocarditis are often directly related to high levels of circulating immune complexes. Renal biopsy results nearly always are abnormal in the setting of active infective endocarditis, which classically causes a hypocomplementemic glomerulonephritis (Chapter 121). Histopathologically, the glomerular changes may be focal, diffuse, or membranoproliferative, or they may be akin to the immune complex disease found in systemic lupus erythematosus. In addition, many of the musculoskeletal conditions associated with infective endocarditis, including monoarticular and oligoarticular arthritides, are probably immune mediated. These immunologic phenomena usually abate with successful antimicrobial therapy.

Surgery

Some patients with infective endocarditis require surgical treatment, either to cure the infection or to avoid its complications[7,8] (Table 76-8). Most patients with evidence of direct extension of infection to myocardial structures,

TABLE 76-8 INDICATIONS FOR SURGERY IN ENDOCARDITIS

INDICATION	CLASS*
NATIVE VALVE ENDOCARDITIS	
Acute aortic insufficiency or mitral regurgitation with heart failure	I
Acute aortic insufficiency with tachycardia and early closure of the mitral valve on echocardiogram	I
Fungal endocarditis	I
Evidence of annular or aortic abscess, sinus or aortic true or false aneurysm, valvular dehiscence, rupture, perforation, or fistula	I
Evidence of valve dysfunction and persistent infection after a prolonged period (7-10 days) of appropriate therapy, provided there are no noncardiac causes of infection	I
Recurrent emboli after appropriate antibiotic therapy	I
Infection with enteric gram-negative organisms or organisms with a poor response to antibiotics in patients with evidence of valve dysfunction	I
Anterior mitral leaflet vegetation (especially with size >10 mm) or persistent vegetation after systemic embolization	IIa
Increase in vegetation size despite appropriate antimicrobial therapy	IIb
Early infections of the mitral valve that can probably be repaired, especially in the presence of large vegetations and/or recurrent emboli	III
Persistent fever and leukocytosis with negative blood cultures	III
PROSTHETIC VALVE ENDOCARDITIS	
Early prosthetic valve endocarditis (<2 mo after surgery)	I
Heart failure with prosthetic valve dysfunction	I
Nonstreptococcal endocarditis	I
Evidence of perivalvular leak, annular or aortic abscess, sinus or aortic true or false aneurysm, fistula formation, or new-onset conduction disturbances	I
Persistent bacteremia after 7-10 days of appropriate antibiotic therapy, with noncardiac causes for bacteremia excluded	IIa
Recurrent peripheral embolus despite therapy	IIa
Vegetation of any size seen on or near the prosthesis	IIb

*Class I = conditions for which there is evidence or general agreement that a given procedure or treatment is useful and effective; class II = conditions for which there is conflicting evidence or a divergence of opinion about the usefulness or efficacy of a procedure or treatment; class IIa = weight of evidence or opinion is in favor of usefulness or efficacy; class IIb = usefulness or efficacy is less well established by evidence or opinion; class III = conditions for which there is evidence or general agreement that the procedure or treatment is not useful and in some cases may be harmful. Adapted with permission from Bonow RO, Carabello B, de Leon AC, et al. Guidelines for the management of patients with valvular heart disease. *Circulation.* 1998;98:1949-1984.

prosthetic valve dysfunction, or heart failure from endocarditis-induced valvular damage should undergo surgery. In addition, many cases of endocarditis caused by fungi, by aerobic gram-negative bacilli or multidrug-resistant organisms (e.g., vancomycin- or gentamicin-resistant enterococci) require surgical management. Progression of disease or persistence of fever and bacteremia for more than 7 to 10 days in the presence of appropriate antibiotic therapy may indicate the need for surgery; however, a thorough search must first be conducted to exclude other metastatic foci of infection. In a randomized trial of patients with left-sided infective endocarditis, severe valve disease, and large vegetations (>10 mm), early surgery did not significantly reduce all-cause mortality at 6 months but markedly decreased the risk of systemic embolism, including stroke and MI. Surgical management should also be considered for patients with recurrent (two or more) embolic events or those with large vegetations (>10 mm) on echocardiography and one embolic event, although the data in these situations are less convincing. The presence of *S. aureus* endocarditis involving the anterior mitral valve leaflet and large vegetations (>10 mm) may be a special circumstance calling for early surgical intervention to reduce the high risk of CNS emboli, especially when mitral valve repair, rather than valve replacement, can be accomplished. Unfortunately, only about 15% to 20% of these latter patients end up being good candidates for valve repair.

Delaying surgery in patients with deteriorating cardiac function in an attempt to sterilize the affected valve is ill advised because the risk of progressive heart failure or further complications usually outweighs the relatively small risk of recurrent infective endocarditis after prosthetic valve implantation. Relative contraindications to valve replacement include recent large CNS emboli (>2 cm) or bleed (because of the risk of bleeding in the perioperative period, when systemic anticoagulation is required), multiple prior valve replacements (because of the difficulty of sewing a new valve into tissue already weakened from previous surgeries), and ongoing injection drug use. On occasion, patients have both a compelling indication for valve replacement (e.g., acute heart failure) and a recent CNS embolic event. The risk of hemorrhagic transformation of such lesions during cardiac bypass–associated anticoagulation is controversial. However, it appears that the greatest risk of such transformation events is in larger (>2 cm) emboli, especially those that have exhibited a hemorrhagic component. In these latter scenarios, it is prudent to try to delay surgery for at least 2 to 4 weeks to allow organization and resolution of such emboli. However, there appears to be no survival benefit in delaying indicated valve replacement surgery (>7 days) after an ischemic stroke.

After definitive surgical treatment, most patients should receive further antibiotic therapy unless a full course of antibiotics was administered before surgery and there is no evidence of ongoing infection. If the patient received antibiotics for less than 1 week before surgery or the culture from the operative site is positive, the patient should receive the equivalent of a full initial course of antibiotics appropriate for the organism. If the patient received antibiotics for 2 weeks or more and the culture result from the operative site is negative (regardless of whether valve histopathology shows inflammation or a positive Gram stain result), the patient should receive whatever remains of the originally planned course of appropriate antibiotic therapy.

In patients with infective endocarditis related to implanted cardiovascular devices, complete device removal is mandatory, regardless of the pathogen, if the goal is to cure the infection. If a replacement device needs to be implanted, the optimal timing for such a procedure is unclear. However, blood culture results should be negative, and any concomitant local or pocket site infection should be completely resolved.

The duration of antimicrobial therapy after device extraction depends on the device and the infection.[9] For lead-related infective endocarditis, which is usually associated with bloodstream infection, 2 weeks of therapy is recommended if there are no infection complications. For infection caused by *S. aureus*, therapy should be extended for up to 4 weeks. In patients with valve infection, 4 to 6 weeks of therapy is recommended.

PREVENTION

Despite a lack of definitive data for dental procedures,[10] prophylactic antibiotics are recommended to prevent infective endocarditis (Table 76-9) when patients with the highest risk of adverse outcomes from endocarditis undergo dental procedures that involve manipulation of gingival tissue or the periapical region of teeth or perforation of the oral mucosa; an invasive procedure of the respiratory tract, with incision or biopsy of the respiratory mucosa, such as tonsillectomy and adenoidectomy; or invasive procedures involving infected skin, skin structures, or musculoskeletal tissue (Table 76-10).[11] Other consensus guidelines have also narrowed the indications for antimicrobial prophylaxis. In the United Kingdom, for example, no prophylaxis is advised for any dental patient, regardless of underlying cardiac valvular conditions. In contrast, French and other European guidelines are largely consistent with current AHA guidelines. Since the recent publications of these more limited recommendations from the AHA, France, and the United Kingdom, follow-up surveys in these countries have shown no appreciable increase in the incidence of viridans group streptococcal infective endocarditis.[12,13]

TABLE 76-9 HIGH-RISK CARDIAC CONDITIONS FOR WHICH ENDOCARDITIS PROPHYLAXIS WITH DENTAL PROCEDURES IS REASONABLE

Prosthetic cardiac valve or prosthetic material used for cardiac valve repair
Previous endocarditis
Complex congenital heart disease involving unrepaired cyanotic congenital heart disease (including palliative shunts and conduits), completely repaired congenital heart disease with prosthetic material within 6 mo of the procedure, or repaired congenital heart disease with residual defects at the site or adjacent to the site of prosthetic material
Cardiac transplantation recipients who develop cardiac valvuloplasty

Adapted from Wilson W, Taubert KA, Gewitz M, et al. Prevention of infective endocarditis guidelines from the American Heart Association: a guideline from the American Heart Association Rheumatic Fever, Endocarditis, and Kawasaki Disease Committee, Council on Cardiovascular Disease in the Young, and the Council on Clinical Cardiology, Council on Cardiovascular Surgery and Anesthesia, and the Quality of Care and Outcomes Research Interdisciplinary Working Group. *Circulation*. 2007;116:1736-1754.

TABLE 76-10 RECOMMENDATIONS FOR ENDOCARDITIS PROPHYLAXIS

PROPHYLAXIS IS RECOMMENDED*
Dental: all dental procedures involving manipulation of gingival tissue or the periapical region of teeth or perforation of the oral mucosa
Respiratory: procedures involving incision or biopsy of the respiratory mucosa, such as tonsillectomy and adenoidectomy
Other: procedures involving infected skin, skin structures, or musculoskeletal tissue prior to incision and drainage
PROPHYLAXIS IS NOT RECOMMENDED
Dental: routine anesthetic injections through noninfected tissue, dental radiographs, placement of removable prosthodontic or orthodontic appliances, adjustment of orthodontic appliances, placement of orthodontic brackets, shedding of deciduous teeth, bleeding from trauma to the lips or oral mucosa
Respiratory: procedures not involving incision or biopsy of the respiratory mucosa, including bronchoscopy (unless the procedure involves incision of the respiratory tract mucosa)
Genitourinary: antibiotic prophylaxis solely to prevent infective endocarditis is not recommended
Gastrointestinal: antibiotic prophylaxis solely to prevent infective endocarditis is not recommended

*Only in patients with underlying cardiac conditions associated with the highest risk for adverse outcome from endocarditis (listed in Table 76-9).
Adapted from Wilson W, Taubert KA, Gewitz M, et al. Prevention of infective endocarditis guidelines from the American Heart Association: a guideline from the American Heart Association Rheumatic Fever, Endocarditis, and Kawasaki Disease Committee, Council on Cardiovascular Disease in the Young, and the Council on Clinical Cardiology, Council on Cardiovascular Surgery and Anesthesia, and the Quality of Care and Outcomes Research Interdisciplinary Working Group. *Circulation*. 2007;116:1736-1754.

The antibiotics chosen for preprocedure prophylaxis should be active against the organisms most likely to be released into the blood stream by the procedure of interest (Table 76-11). Thus, antibiotics that cover primarily oral flora are recommended for dental and upper respiratory procedures. For patients with the conditions listed in Table 76-9 who undergo a procedure for infected skin, skin structure, or musculoskeletal tissue, the therapeutic regimen should contain an agent active against staphylococci and β-hemolytic streptococci.

Patients with implanted cardiac devices do not require antibiotic prophylaxis for dental or other invasive procedures. However, such patients require surgical site prophylaxis at the time of device placement.[2] The recommended regimens generally include a β-lactam (commonly cefazolin, 1 g IV 1 hour before device placement), regardless of whether a new device is being placed or a device is being revised.

PROGNOSIS

Untreated infective endocarditis is uniformly fatal. Aggressive medical and surgical management dramatically improves the outcome. The overall mortality rate from both native and prosthetic valve endocarditis remains fairly high, ranging from 17% to 36%. Whereas certain subgroups, such as patients with viridans group streptococcal endocarditis, have a lower risk of death,

TABLE 76-11 SUGGESTED ANTIBIOTICS FOR ENDOCARDITIS PROPHYLAXIS FOR DENTAL OR RESPIRATORY TRACT PROCEDURES* IN PATIENTS WITH HIGH-RISK CARDIAC CONDITIONS†

PATIENT CHARACTERISTICS	REGIMEN‡
Able to take oral medications	Amoxicillin 2 g PO
Unable to take oral medications	Ampicillin 2 g IV or IM; or cefazolin or ceftriaxone 1 g IM or IV
Allergic to penicillin or ampicillin and able to take oral medications	Cephalexin 2 g PO (or other first- or second-generation oral cephalosporin in equivalent adult doses); clindamycin 600 mg PO; azithromycin 500 mg PO; or clarithromycin 500 mg PO Cephalosporins should not be used in an individual with a history of anaphylaxis, angioedema, or urticaria with penicillin or ampicillin
Allergic to penicillin or ampicillin and unable to take oral medications	Cefazolin or ceftriaxone 1 g IM or IV; or clindamycin 600 mg IM or IV

*For the applicable procedures, see Table 76-10.
†For the applicable conditions, see Table 76-9.
‡All regimens consist of a single dose 30-60 min before the procedure.
IM = intramuscular; IV = intravenous; PO = oral.
Adapted from Wilson W, Taubert KA, Gewitz M, et al. Prevention of infective endocarditis guidelines from the American Heart Association: a guideline from the American Heart Association Rheumatic Fever, Endocarditis, and Kawasaki Disease Committee, Council on Cardiovascular Disease in the Young, and the Council on Clinical Cardiology, Council on Cardiovascular Surgery and Anesthesia, and the Quality of Care and Outcomes Research Interdisciplinary Working Group. *Circulation*. 2007;116:1736-1754.

patients with *S. aureus*, fungal, and zoonotic endocarditis have higher mortality rates. Heart failure and CNS events are the most frequent causes of death.

Endocarditis recurs in about 12% to 16% of patients and is more common in injection drug users, elderly people, and patients with prosthetic valves. The rate of relapse also varies depending on the causative organism. Easily treated infections, such as those with viridans group streptococci, have a low rate of relapse (5%), but more difficult-to-eradicate organisms may have significantly higher rates.

FUTURE DIRECTIONS

As cardiac imaging technology continues to improve, the duration of treatment of endocarditis may be dictated in part by the characteristics of visualized vegetations. In addition, now that large vegetations have been demonstrated to cause more embolic events, earlier interventions to remove vegetations from functionally competent valves or to introduce agents that prevent the formation or promote the dissolution of vegetations may be feasible. Moreover, advances in imaging may allow routine screening of patients for subclinical infections. Novel therapeutic and prophylactic approaches, such as antibacterial antibodies, targeted bacterial vaccines, and cell wall–specific enzymes that can act as adjuncts to antibiotics in facilitating bacteriologic clearance, are currently in development.

Grade A References

A1. Cosgrove SE, Vigliani GA, Fowler VG Jr, et al. Initial low-dose gentamicin for *Staphylococcus aureus* bacteremia and endocarditis is nephrotoxic. *Clin Infect Dis*. 2009;48:713-721.
A2. Sexton DJ, Tenenbaum MJ, Wilson WR, et al. Ceftriaxone once daily for four weeks compared with ceftriaxone plus gentamicin once daily for two weeks for treatment of endocarditis due to penicillin-susceptible streptococci. Endocarditis Treatment Consortium Group. *Clin Infect Dis*. 1998;27:1470-1474.
A3. Fowler VG Jr, Boucher HW, Corey GR, et al. Daptomycin versus standard therapy for bacteremia and endocarditis caused by *Staphylococcus aureus*. *N Engl J Med*. 2006;355:653-665.
A4. Chan KL, Dumesnil JG, Cujec B, et al. A randomized trial of aspirin on the risk of embolic events in patients with infective endocarditis. *J Am Coll Cardiol*. 2003;42:775-780.
A5. Kang DH, Kim YJ, Kim SH, et al. Early surgery versus conventional treatment for infective endocarditis. *N Engl J Med*. 2012;366:2466-2473.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Baddour LM, Cha YM, Wilson WR. Clinical practice. Infections of cardiovascular implantable electronic devices. *N Engl J Med*. 2012;367:842-849.
2. Baddour LM, Epstein AE, Erickson CC, et al. Update on cardiovascular implantable electronic device infections and their management: a scientific statement from the American Heart Association. *Circulation*. 2010;121:458-477.
3. Hoen B, Duval X. Clinical practice. Infective endocarditis. *N Engl J Med*. 2013;368:1425-1433.
4. Thuny F, Grisoli D, Cautela J, et al. Infective endocarditis: prevention, diagnosis, and management. *Can J Cardiol*. 2014;30:1046-1057.
5. Baddour LM, Wilson WR, Bayer AS, et al. Infective endocarditis: diagnosis, antimicrobial therapy, and management of complications: a statement for healthcare professionals from the Committee on Rheumatic Fever, Endocarditis, and Kawasaki Disease, Council on Cardiovascular Disease in the Young, and the Councils on Clinical Cardiology, Stroke, and Cardiovascular Surgery and Anesthesia, American Heart Association: endorsed by the Infectious Diseases Society of America. *Circulation*. 2005;111:e394-e434.
6. Tattevin P1, Boutoille D2, Vitrat V, et al. Salvage treatment of methicillin-resistant staphylococcal endocarditis with ceftaroline: a multicentre observational study. *J Antimicrob Chemother*. 2014;69: 2010-2013.
7. Barsic B, Dickerman S, Krajinovic V, et al. Influence of the timing of cardiac surgery on the outcome of patients with infective endocarditis and stroke. *Clin Infect Dis*. 2013;56:209-217.
8. Lalani T, Cabell CH, Benjamin DK, et al. Analysis of the impact of early surgery on in-hospital mortality of native valve endocarditis: use of propensity score and instrumental variable methods to adjust for treatment selection bias. *Circulation*. 2010;121:1005-1013.
9. Sandoe JA, Barlow G, Chambers JB, et al. Guidelines for the diagnosis, prevention and management of implantable cardiac electronic device infection. Report of a joint Working Party project on behalf of the British Society for Antimicrobial Chemotherapy (BSAC, host organization), British Heart Rhythm Society (BHRS), British Cardiovascular Society (BCS), British Heart Valve Society (BHVS) and British Society for Echocardiography (BSE). *J Antimicrob Chemother*. 2015;70: 325-359.
10. Glenny AM, Oliver R, Roberts GJ, et al. Antibiotics for the prophylaxis of bacterial endocarditis in dentistry. *Cochrane Database Syst Rev*. 2013;10:CD003813.
11. Wilson W, Taubert KA, Gewitz M, et al. Prevention of infective endocarditis: guidelines from the American Heart Association: a guideline from the American Heart Association Rheumatic Fever, Endocarditis, and Kawasaki Disease Committee, Council on Cardiovascular Disease in the Young, and the Council on Clinical Cardiology, Council on Cardiovascular Surgery and Anesthesia, and the Quality of Care and Outcomes Research Interdisciplinary Working Group. *Circulation*. 2007; 116:1736-1754.
12. Desimone DC, Tleyjeh IM, Correa de Sa DD, et al. Incidence of infective endocarditis caused by viridans group streptococci before and after publication of the 2007 American Heart Association's endocarditis prevention guidelines. *Circulation*. 2012;126:60-64.
13. Duval X, Delahaye F, Alla F, et al. Temporal trends in infective endocarditis in the context of prophylaxis guideline modifications: three successive population-based surveys. *J Am Coll Cardiol*. 2012;59:1968-1976.

REVIEW QUESTIONS

1. A 76-year-old man with known rheumatic valvular heart disease underwent elective mitral valve replacement with a Saint Jude prosthetic valve. His dentist calls you for advice regarding choice of antibiotic prophylaxis before dental extraction because the patient had developed bronchospasm and a diffuse urticarial rash after amoxicillin administration before a dental cleaning approximately 6 months ago. Which antibiotic should be administered?

A. Clindamycin 600 mg orally 1 hour before the procedure
B. Cefuroxime axetil 500 mg orally before the procedure
C. Amoxicillin 2 g intravenously 1 hour before the procedure with corticosteroid and antihistamine coverage
D. Nafcillin sodium 2 g IV 1 hour before the procedure
E. Gentamicin sulfate 1 mg/kg IV 1 hour before the procedure

Answer: A The current (2007) AHA guidelines recommend clindamycin in patients who have a history of an immediate type hypersensitivity reaction to β-lactam antibiotics, which this patient demonstrated. Therefore, amoxicillin, cephalosporins, and nafcillin should be avoided. Levofloxacin and gentamicin are not recommended in these guidelines. (Wilson W, Taubert KA, Gewitz M, et al. Prevention of infective endocarditis: guidelines from the American Heart Association: a guideline from the American Heart Association Rheumatic Fever, Endocarditis, and Kawasaki Disease Committee, Council on Cardiovascular Disease in the Young, and the Council on Clinical Cardiology, Council on Cardiovascular Surgery and Anesthesia, and the Quality of Care and Outcomes Research Interdisciplinary Working Group. *Circulation*. 2007;116:1736-1754.)

2. A 68-year-old man with diabetes on chronic hemodialysis developed the acute onset of fever, chills, and left-sided abdominal pain. He had an ICD implanted 3 years ago. Blood cultures grew *Staphylococcus aureus*, and transesophageal echocardiography demonstrated vegetations on the mitral valve. Splenic infarctions were seen on computed tomographic scanning. Which one of the following is *true* regarding this presentation?

A. Health care–associated infection is accounting for an increasing number of infective endocarditis cases in this country.
B. *Escherichia coli* is a common cause of infective endocarditis in the hemodialysis population.
C. Pending susceptibility testing results, gentamicin should be administered.
D. To reduce health care costs, transthoracic echocardiography should have been performed instead of transesophageal echocardiography.
E. For chronic hemodialysis, a tunneled catheter has a lower risk of blood stream infection compared with an arteriovenous fistula.

Answer: A Health care exposure accounts for an increasing number of cases of infective endocarditis in developed countries. *Staphylococcus aureus*, including methicillin-resistant strains, is a common cause of these infections. Empiric vancomycin should be administered until susceptibility results are known. (Athan E, Chu VH, Tattevin P, et al. Clinical characteristics and outcome of infective endocarditis involving implantable cardiac devices. *JAMA*. 2012;307:1727-1735.)

3. A 55-year-old veterinarian presents with several months of low-grade fever and night sweats. On an echocardiograph, he has evidence of endocarditis. Despite no recent antibiotic therapy, three sets of blood cultures remain negative for 7 days. Which one of the following is the most likely pathogen?

A. Coagulase-negative staphylococcus
B. *Coxiella burnetii*
C. *Enterococcus faecium*
D. *Escherichia coli*
E. Orf virus

Answer: B Exposure to animals, particularly sheep and goats, is a risk factor for *Coxiella* infection and a well-known cause of culture-negative endocarditis. Orf virus does not cause endocarditis. The other choices should result in positive blood culture results in a patient who has infective endocarditis and who has not received antibiotic therapy recently. (Baddour LM, Wilson WR, Bayer AS, et al. Infective endocarditis: diagnosis, antimicrobial therapy, and management of complications: a statement for healthcare professionals from the Committee on Rheumatic Fever, Endocarditis, and Kawasaki Disease, Council on Cardiovascular Disease in the Young, and the Councils on Clinical Cardiology, Stroke, and Cardiovascular Surgery and Anesthesia, American Heart Association: endorsed by the Infectious Diseases Society of America. *Circulation*. 2005;111:e394-434.)

4. A 78-year-old woman presents with left upper chest pain, swelling, and purulent drainage at the site where a permanent pacemaker generator was implanted. She has had no fever or chills, and her white blood cell count is normal. She underwent generator exchange 4 months ago and underwent a dental cleaning without prophylaxis 3 months ago. Which one of the following is *true* regarding cardiac implantable electronic device (CIED) infections?

A. Surgical site prophylaxis has not been shown to reduce the risk of a CIED site infection.
B. Prophylactic antibiotic should have been given before her dental cleaning.
C. Device manipulation is a risk factor for device infection.
D. The most likely cause of this infection is a HACEK organism.
E. Antibiotic therapy for 4 weeks will likely cure the device infection without the device being removed.

Answer: C Manipulation of an implantable cardiac device is associated with the development of acute infection. Antibiotic prophylaxis before manipulation of the surgical site is beneficial, but dental prophylaxis is not. The most likely cause is *Staphylococcus* spp., and removal of the device is required for cure of the infection. (Baddour LM, Cha YM, Wilson WR. Clinical practice. Infections of cardiovascular implantable electronic devices. *N Engl J Med.* 2012;367:842-849.)

5. A 25-year-old morbidly obese man who injects heroin and cocaine presents with fever and blood cultures that grow *Staphylococcus aureus*. He had a past history of prior *S. aureus* blood stream infection 2 years ago, when he had an allergic reaction to vancomycin. At that time, he had evidence of tricuspid valve endocarditis. Which one of the following is *true*?

A. A transthoracic echocardiography should be obtained.
B. Initial empiric therapy should include daptomycin until susceptibility results are known.
C. His mortality risk is high (>50%).
D. Two weeks of antibiotic therapy should be curative.
E. Rifampin should be administered.

Answer: B Daptomycin should be administered in case the blood culture isolate is methicillin-resistant *Staphylococcus aureus*. Transesophageal, rather than transthoracic, echocardiography should be obtained to evaluate for both right-sided and left-sided endocarditis. Cure rates with active antibiotic therapy for 2 weeks are high, provided there is no evidence of left-sided endocarditis or of metastatic foci of infection. Rifampin is not routinely recommended in native valve infections caused by staphylococci. (Fowler VG Jr, Boucher HW, Corey GR, et al. Daptomycin versus standard therapy for bacteremia and endocarditis caused by *Staphylococcus aureus*. *N Engl J Med.* 2006;355:653-665.)

77

PERICARDIAL DISEASES

WILLIAM C. LITTLE AND JAE K. OH

The pericardium, which is a relatively avascular fibrous sac that surrounds the heart, has two layers, the visceral and parietal pericardia. The potential space between the two layers normally contains only 10 to 50 mL of fluid, which is an ultrafiltrate of plasma. The pericardium is well innervated, so pericardial inflammation may produce severe pain and trigger vagally mediated reflexes.

As a result of its relatively inelastic physical properties, the pericardium limits acute cardiac dilation and enhances mechanical interactions of the cardiac chambers. In response to long-standing stress, the pericardium dilates to allow a slowly accumulating pericardial effusion to become quite large without compressing the cardiac chambers and to allow left ventricular remodeling to occur without pericardial constriction. Conversely, a scarred or thickened pericardium can limit the filling of the heart, resulting in pericardial constriction. Despite the important functions of the normal pericardium, congenital absence or surgical resection of the pericardium does not appear to have any major untoward effects.

ACUTE PERICARDITIS

EPIDEMIOLOGY AND PATHOBIOLOGY

Acute inflammation of the pericardium, with or without an associated pericardial effusion, can occur as an isolated clinical problem or as a manifestation of systemic disease. Although about 85% of isolated cases of acute pericarditis are idiopathic or viral, the list of other potential causes is quite extensive (Table 77-1). Patients with fever greater than 38° C or a subacute course or who fail to respond promptly to therapy are most likely to have pericarditis caused by a systemic autoimmune disease, malignancy, or viral or bacterial infection.

Pericarditis can occur after an acute myocardial infarction (MI). It occurs 1 to 3 days after a transmural MI, presumably owing to the interaction between the healing necrotic epicardium and the overlying pericardium. Dressler syndrome, which is another form of pericarditis associated with MI, typically occurs weeks to months after MI. It is similar to the pericarditis that can occur days to months after traumatic pericardial injury, surgical manipulation of the pericardium, or pulmonary infarction. This syndrome is presumed to be mediated by an autoimmune mechanism and is associated with signs of systemic inflammation, including fever and polyserositis.

CLINICAL MANIFESTATIONS

Most patients with acute pericarditis experience sharp retrosternal chest pain (see Table 51-2 in Chapter 51), which can be quite severe and debilitating. In some cases, however, pericarditis is asymptomatic, such as when it accompanies rheumatoid arthritis. Pericardial pain is usually worse with inspiration and when supine, and it is generally relieved by sitting and leaning forward. Typically, pericardial pain is referred to the scapular ridge, presumably owing to irritation of the phrenic nerves, which pass adjacent to the pericardium. The chest pain of acute pericarditis must be differentiated from pulmonary embolism and myocardial ischemia or infarction (Table 77-2).

The pericardial friction rub is the classic finding in patients with acute pericarditis. A friction rub is a high-pitched, scratchy sound that can have one, two, or three components occurring when the cardiac volumes are most rapidly changing: during ventricular ejection, during rapid ventricular filling in early diastole, and during atrial systole. A pericardial rub, which is differentiated from a murmur by its scratchy quality, is sometimes localized to a small area on the chest wall and may come and go spontaneously or with changes in position. To hear a rub, it may be necessary to auscultate the heart with the patient in multiple positions, especially using the diaphragm with the patient learning forward and not breathing after full expiration. The pericardial friction rub must be differentiated from a pleural rub, which is absent during suspended respiration, but the pericardial rub is unaffected.

DIAGNOSIS

Early in the course of acute pericarditis, the electrocardiogram (ECG) typically displays diffuse ST elevation in association with PR depression (Fig. 77-1). The ST elevation is usually present in all leads except for aVR,

TABLE 77-1 CAUSES OF PERICARDITIS: INFECTIOUS AND NONINFECTIOUS

INFECTIOUS PERICARDITIS (⅔ OF CASES)

Viral (most common): echovirus and coxsackievirus (usual), influenza, EBV, CMV, adenovirus, varicella, rubella, mumps, HBV, HCV, HIV, parvovirus B19, human herpesvirus 6 (increasing reports)
Bacterial: tuberculosis (4%-5%)* and *Coxiella burnetii* (most common); other bacterial causes (rare) include pneumococcosis, meningococcosis, gonococcosis, *Haemophilus*, staphylococci, Chlamydia, *Mycoplasma, Legionella, Leptospira, Listeria*
Fungal (rare): histoplasmosis more likely in immunocompetent patients; aspergillosis, blastomycosis, candidiasis more likely in immunosuppressed patients
Parasitic (very rare): *Echinococcus, Toxoplasma*

NONINFECTIOUS PERICARDITIS (⅓ OF CASES)

Autoimmune Pericarditis (<10%)*
Pericardial injury syndromes: post–myocardial infarction syndrome; postpericardiotomy syndrome; posttraumatic pericarditis, including iatrogenic pericarditis (e.g., after percutaneous coronary interventions, pacemaker insertion, ablation)
Pericarditis in systemic autoimmune and autoinflammatory diseases: more common in systemic lupus erythematosus, Sjögren syndrome, rheumatoid arthritis, systemic sclerosis, systemic vasculitides, Behçet syndrome, sarcoidosis, familial Mediterranean fever
Autoreactive pericarditis*†
Neoplastic Pericarditis (5%-7%)*
Primary tumors (rare): pericardial mesothelioma
Secondary metastatic tumors (common): lung and breast cancer, lymphoma
Metabolic pericarditis: uremia, myxedema (common); others rare
Traumatic Pericarditis (rare)
Direct injury: penetrating thoracic injury, esophageal perforation, iatrogenic
Indirect injury: nonpenetrating thoracic injury, radiation injury
Drug-related pericarditis (rare): procainamide, hydralazine, isoniazid, and phenytoin (lupus-like syndrome), penicillins (hypersensitivity pericarditis with eosinophilia), doxorubicin, and daunorubicin (often associated with cardiomyopathy; may cause pericardiopathy)

*Percentages refer to unselected cases.
†The diagnosis of autoreactive pericarditis is established using the following criteria: (1) increased number of lymphocytes and mononuclear cells >5000/mm^3 (autoreactive lymphocytic) or the presence of antibodies against heart muscle tissue (antisarcolemmal) in the pericardial fluid (autoreactive antibody mediated); (2) signs of myocarditis on epicardial or endomyocardial biopsies by ≥14 cells/mm^2; and (3) exclusion of infections, neoplasia, and systemic and metabolic disorders.
CMV = cytomegalovirus; EBV = Epstein-Barr virus; HBV = hepatitis B virus; HCV = hepatitis C virus; HIV = human immunodeficiency virus.
From Imazio M, Spodick DH, Brucato A, et al. Controversial issues in the management of pericardial diseases. *Circulation.* 2010;121:916-928.

TABLE 77-2 DIFFERENTIATION OF PERICARDITIS FROM MYOCARDIAL ISCHEMIA OR INFARCTION AND PULMONARY EMBOLISM

FINDINGS	MYOCARDIAL ISCHEMIA OR INFARCTION	PERICARDITIS	PULMONARY EMBOLISM
CHEST PAIN			
Character	Pressure-like heavy, squeezing	Sharp, stabbing, occasionally dull	Sharp, stabbing
Change with respiration	No	Worsened with inspiration	In phase with respiration (absent when the patient is apneic)
Change with position	No	Worse when supine; improved when sitting up or leaning forward	No
Duration	Minutes (ischemia); hours (infarction)	Hours to days	Hours to days
Response to nitroglycerin	Improved	No change	No change
PHYSICAL EXAMINATION			
Friction rub	Absent (unless pericarditis is present)	Present in most patients	Pleural friction rub may occur
ELECTROCARDIOGRAM			
ST segment elevation	Localized convex	Widespread concave	Limited to leads III, aVF, and V_1
PR segment depression	Rare	Frequent	None

Modified from Little WC, Freeman GL. Pericardial disease. *Circulation.* 2006;113:1622-1632.

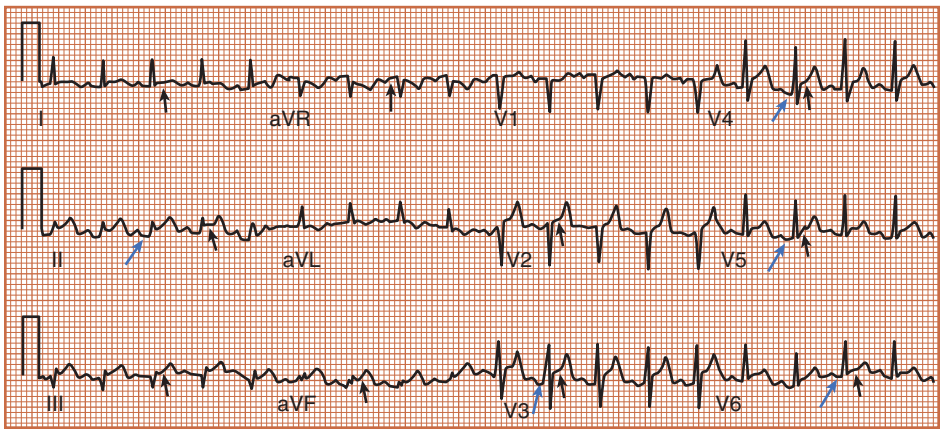

FIGURE 77-1 Electrocardiogram demonstrating typical features of acute pericarditis on presentation. There are diffuse ST elevation and PR depression except in aVR, where there is ST depression and PR elevation.

but the changes may be more localized in post-MI pericarditis. Classically, the ECG changes of acute pericarditis evolve over several days; resolution of the ST elevation is followed by widespread T-wave inversion that subsequently normalizes. Uremic pericarditis usually occurs without the typical ECG abnormalities.

Patients with acute pericarditis usually have evidence of systemic inflammation, including leukocytosis, an elevated erythrocyte sedimentation rate (ESR), and increased C-reactive protein (CRP) level. A low-grade fever is common, but a temperature greater than 38° C is unusual and suggests the possibility of bacterial pericarditis.

About 85% of cases of acute pericarditis are idiopathic or viral. Viral causes include echoviruses and group B coxsackieviruses, but obtaining specific viral titers does not alter patient management. About 6% of cases are neoplastic in origin, about 4% are caused by tuberculosis, about 3% are caused by other bacterial or fungal infections, and about 2% are caused by collagen vascular disease. A targeted evaluation (Table 77-3) can help identify the various causes (Table 77-4).

Troponin levels typically are minimally elevated in acute pericarditis owing to some involvement of the epicardium by the inflammatory process. An elevated troponin level in acute pericarditis usually returns to normal within 1 to 2 weeks and is not associated with a worse prognosis. Although the elevated troponin level may lead to the misdiagnosis of an ST elevation MI (Chapter 73), most patients with elevated troponin levels and acute pericarditis have normal coronary angiograms. An echocardiogram (Chapter 55) can help avoid a misdiagnosis of MI. Interestingly, patients with myopericarditis and elevated troponin levels tend to have a lower recurrence rate than do patients with pure pericarditis and normal troponin levels.[1]

Echocardiography may demonstrate a small pericardial effusion in the presence of acute pericarditis, but normal echocardiogram results do not exclude the diagnosis of acute pericarditis. An echocardiogram is critical, however, in excluding the diagnosis of cardiac tamponade (see later). When the diagnosis of acute pericarditis is unclear, cardiac magnetic resonance imaging (MRI) can demonstrate pericardial inflammation as delayed enhancement of the pericardium (Fig. 77-2). Diagnostic pericardiocentesis is indicated in suspected purulent tuberculosis or malignant pericarditis or if the patient has cardiac tamponade.

TABLE 77-3 SELECTED DIAGNOSTIC TESTS IN ACUTE PERICARDITIS

IN ALL PATIENTS
Tuberculin skin test (plus control skin test to exclude anergy)
BUN and creatinine to exclude uremia
Erythrocyte sedimentation rate
Electrocardiogram
Chest radiograph
Echocardiogram
IN SELECTED PATIENTS
Cardiac magnetic resonance imaging
ANA and rheumatoid factor to exclude SLE or rheumatoid arthritis in patients with acute arthritis or pleural effusion
TSH and T_4 to exclude hypothyroidism in patients with clinical findings suggestive of hypothyroidism and in asymptomatic patients with unexplained pericardial effusion
HIV test to exclude AIDS in patients with risk factors for HIV disease or a compatible clinical syndrome
Blood cultures in febrile patients to exclude infective endocarditis and bacteremia
Fungal serologic tests in patients from endemic areas or in immunocompromised patients
ASO titer in children or teenagers with suspected rheumatic fever
Heterophil antibody test to exclude mononucleosis in young or middle-aged patients with a compatible clinical syndrome or acute fever, weakness, and lymphadenopathy

AIDS = acquired immunodeficiency virus; ANA = antinuclear antibody; ASO = antistreptolysin O; BUN = blood urea nitrogen; HIV = human immunodeficiency virus; SLE = systemic lupus erythematosus; T_4 = thyroxine; TSH = thyroid-stimulating hormone.
Modified from Nishimura RA, Kidd KR. Recognition and management of patients with pericardial disease. In: Braunwald E, Goldman L, eds. *Primary Cardiology*. 2nd ed. Philadelphia: WB Saunders; 2003:625.

TABLE 77-4 PRESENTATION AND TREATMENT OF THE MOST COMMON CAUSES OF PERICARDITIS

TYPE	PATHOGENESIS OR ETIOLOGY	DIAGNOSIS	TREATMENT	COMPLICATIONS	COMMENTS
Viral	Coxsackievirus B Echovirus type 8 Epstein-Barr virus	Leukocytosis Elevated ESR Mild cardiac biomarker elevation	Symptomatic relief, NSAIDs, colchicine	Tamponade Relapsing pericarditis	Peaks in spring and fall
Tuberculous	*Mycobacterium tuberculosis*	Isolation of organism from biopsy fluid Granulomas not specific	Triple-drug antituberculosis regimen Pericardial drainage followed by early (4-6 wk) pericardiectomy if signs of tamponade or constriction develop	Tamponade Constrictive pericarditis	1%-8% of patients with tuberculosis pneumonia; rule out HIV infection
Bacterial	Group A streptococcus *Staphylococcus aureus* *Streptococcus pneumoniae*	Leukocytosis with marked left shift Purulent pericardial fluid	Pericardial drainage by catheter or surgery Systemic antibiotics Pericardiectomy if constrictive physiology develops	Tamponade in one third of patients	Very high mortality rate if not recognized early
Post–myocardial infarction	12 hr–10 days after infarction	Fever Pericardial friction rub Echo: effusion	Aspirin Prednisone	Tamponade rare	More frequent in large Q wave infarctions Anterior > inferior
Uremic	Untreated renal failure: 50% Chronic dialysis: 20%	Pericardial rub: 90%	Intensive dialysis Indomethacin: probably ineffective Catheter drainage Surgical drainage	Tamponade Hemodynamic instability on dialysis	Avoid NSAIDs ≈50% respond to intensive dialysis
Neoplastic	In order of frequency: lung cancer, breast cancer, leukemia and lymphoma, others	Chest pain, dyspnea Echo: effusion CT, MRI: tumor metastases to pericardium Cytologic examination of fluid positive in 85%	Catheter drainage Subxiphoid pericardiectomy Chemotherapy directed at underlying malignant neoplasm	Tamponade Constriction	

CT = computed tomography; ESR = erythrocyte sedimentation rate; HIV = human immunodeficiency virus; MRI = magnetic resonance imaging; NSAIDs = nonsteroidal anti-inflammatory drugs.
Modified from Malik F, Foster E. Pericardial disease. In: Wachter RM, Goldman L, Hollander H, eds. *Hospital Medicine*. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2005:449.

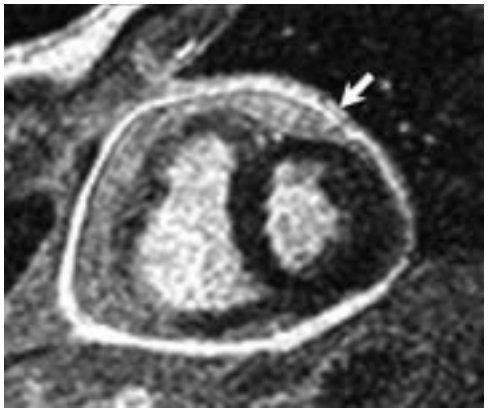

FIGURE 77-2. Cardiac magnetic resonance image of a patient with acute pericarditis shows late gadolinium hyperenhancement of the pericardium and epicardium.

TREATMENT

Rx

Acutely ill patients with fever should be hospitalized, as should patients with suspected acute MI (Chapter 73), large effusions, evidence of impending hemodynamic compromise, or a cause other than viral or idiopathic pericarditis because of the risk of a rapidly accumulating effusion with potential tamponade. Patients without effusions can usually be followed as outpatients (Fig. 77-3).

If acute pericarditis is a manifestation of an underlying disease, it often responds to treatment of the primary condition. Most cases of acute idiopathic or viral pericarditis are self-limited and respond to treatment with aspirin (650 mg every 6 hours) or another nonsteroidal antiinflammatory drug (NSAID) such as ibuprofen (300 to 800 mg every 6 to 8 hours). The dose of NSAID should be tapered after symptoms and any pericardial effusion have resolved, but the medication should be taken for at least 3 to 4 weeks to minimize the risk of recurrent pericarditis.

In addition, colchicine (0.6 to 1.2 mg/day for 3 months) should be started in all patients with acute pericarditis to reduce the rate of persistent symptoms at 72 hours, reduce the likelihood of recurrent pericarditis from 55% to 24% at 18 months, and reduce the rate of subsequent hospitalization.[■■] The major side effect of colchicine is diarrhea. The lower dose of colchicine should be used in patients who weigh less than 70 kg or who have side effects with the higher dose. Colchicine should be avoided in patients with abnormal renal or hepatic function and in patients being treated with macrolide antibiotics, which alter its metabolism.

A proton pump inhibitor, such as omeprazole (20 mg/day), should be considered to improve the gastric tolerability of NSAIDs. Warfarin and heparin should be avoided to minimize the risk of hemopericardium, but anticoagulation may be required if the patient is in atrial fibrillation or has a prosthetic heart valve. It is prudent to avoid exercise until after the chest pain completely resolves. If pericarditis recurs, the patient can be reloaded with colchicine and intravenous ketorolac (20 mg) and then continued on an oral NSAID and colchicine for at least 3 months.

Although acute pericarditis usually responds dramatically to systemic corticosteroids, observational studies strongly suggest that the use of steroids increases the probability of relapse in patients treated with colchicine. Except when needed to treat an underlying inflammatory disease, every effort should be made to avoid the use of steroids, reserving low-dose steroids for patients who cannot tolerate aspirin and other NSAIDs or whose recurrence is not responsive to colchicine and intravenous NSAIDs. If steroids are used, low-dose prednisone (0.2 to 0.5 mg/kg) appears to be as effective as higher doses and is less likely to be associated with recurrence. Steroids should be continued for at least 1 month before slow tapering, which can be guided by return of the CRP level to the normal range. Pericardiocentesis is not recommended unless purulent or tuberculous pericarditis is clinically suspected or the patient fails to respond to 2 to 3 weeks of NSAID therapy.

PROGNOSIS

The course of viral and idiopathic pericarditis is usually self-limited, and most patients recover completely.[2] About 25% of patients, however, have recurrent pericarditis weeks to months later, probably caused by an immune response, and some patients may have multiple debilitating episodes. In patients whose acute pericarditis is accompanied by myocarditis, as evidenced by elevation of serum troponin levels, the recurrence rate is closer to 10%.[1] Recurrent pericarditis is more common in patients treated with steroids for the acute episode, especially during a rapid steroid taper. In these patients, prolonged high-dose NSAID treatment (e.g., ibuprofen 300 to 600 mg three times a day) plus colchicine (0.5 to 0.6 mg twice daily, declining to once daily after 3 to 6 months) is effective.[■] In patients who cannot tolerate colchicine or who have recurrent episodes despite colchicine and high-dose NSAID treatment (e.g., indomethacin 50 mg three times a day or ibuprofen 800 mg four times a day), oral steroids (e.g., prednisone 0.2 to 0.5 mg/kg/day for 2 to 4 weeks; then slowly tapered over several months) are generally recommended. In patients with refractory, recurrent pericarditis, surgical pericardiectomy can be considered.[3] Patients who have tuberculous or purulent pericarditis or have recurrent episodes of pericarditis have the highest risk for progression to constrictive pericarditis (see later).

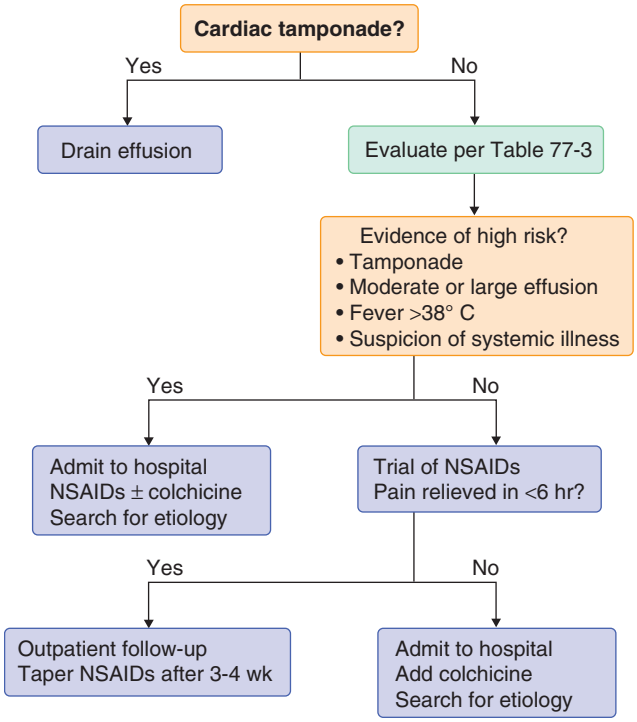

FIGURE 77-3. Initial management of patients with pericarditis. NSAID = nonsteroidal antiinflammatory drug.

TABLE 77-5 CAUSES OF MODERATE TO LARGE ASYMPTOMATIC PERICARDIAL EFFUSIONS

CAUSE	CASES (%)
Idiopathic or viral	37
Neoplastic	19
Iatrogenic or trauma	13
Tuberculous or purulent	6
Acute myocardial infarction	6
Collagen vascular disease	4
Heart failure	4
Uremia	4
Radiation induced	2
Aortic dissection	2
Hypothyroidism	1
Other	2

Modified from Nishimura RA, Kidd KR. Recognition and management of patients with pericardial disease. In: Braunwald E, Goldman L, eds. *Primary Cardiology.* 2nd ed. Philadelphia: WB Saunders; 2003:625.

CARDIAC EFFUSION AND TAMPONADE

EPIDEMIOLOGY

A pericardial effusion can be caused by any disease that causes acute pericarditis (see Table 77-1), but a majority of cases are caused by conditions other than viral or idiopathic pericarditis (Table 77-5). For example, tamponade

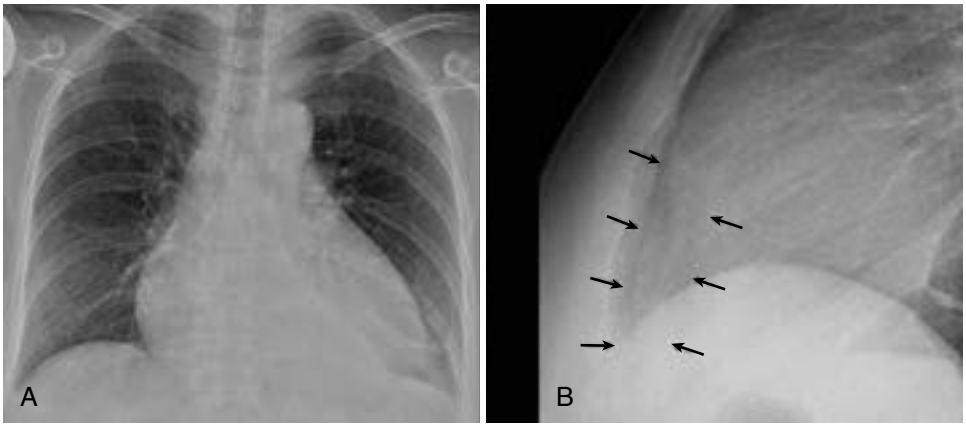

FIGURE 77-4. Chest radiographs in a patient with a large pericardial effusion. The cardiac silhouette on the posteroanterior view (**A**) is enlarged with a "water bag" configuration. The lateral view (**B**) shows a separation between the pericardial and epicardial fat stripes *(arrows)*.

occurs in about 10% to 15% of patients with idiopathic pericarditis, but it develops in more than 50% of patients with malignant, tuberculous, or purulent pericarditis. Tuberculosis and neoplastic disease are typically associated with serosanguineous effusions, but such effusions can also be seen with typical viral or idiopathic pericarditis, with uremia, and after mediastinal irradiation. Hemopericardium is seen most commonly with trauma, myocardial rupture after MI, catheter-induced myocardial or epicardial coronary artery rupture, aortic dissection with rupture into the pericardial space, or primary hemorrhage in patients receiving anticoagulant therapy, often after cardiac valve surgery. Chylopericardium is rare and results from leakage or injury to the thoracic duct.

PATHOBIOLOGY

Under normal conditions, the space between the parietal and visceral pericardium can accommodate only a small amount of fluid before the development of tamponade physiology. The clinical consequences of a pericardial effusion depend on the rate of increase. A rapidly accumulating effusion, as in hemopericardium caused by trauma or aortic dissection, may result in tamponade physiology with just 100 to 200 mL of fluid. It is not surprising, therefore, that cardiac perforation quickly results in tamponade. By comparison, a more slowly developing effusion, as is typical with uremia and hypothyroidism, may allow the gradual stretching of the pericardium, with asymptomatic or minimally symptomatic effusions of 1500 mL or more.

Tamponade physiology occurs when fluid accumulation in the intrapericardial space is sufficient to compress the heart, resulting in impaired cardiac filling. The increased pericardial pressure in cardiac tamponade accentuates the interdependence among the cardiac chambers as the total cardiac volume is limited by the pericardial effusion. With normal inspiration, right ventricular filling is enhanced, the intraventricular septum is displaced toward the left ventricle, and left ventricular filling and the resulting stroke volume are reduced. Because of its lower pressures, the right ventricle is most vulnerable to compression by a pericardial effusion, and abnormal right heart filling is the earliest sign of a hemodynamically significant pericardial effusion. In tamponade, left heart filling occurs preferentially during expiration, when there is less filling of the right heart. The small normal respiratory increase in right ventricular volume, with a concomitant decrease in left ventricular stroke volume and systolic arterial pressure is markedly accentuated in cardiac tamponade and results in the clinical finding of "paradoxical pulse." A small (<10 mm Hg) pulsus paradoxus, which is a decline in systemic blood pressure during inspiration, is normal and is related to the ventricles being confined within the pericardium and sharing a common septum. In cardiac tamponade, this phenomenon is exaggerated, and systemic blood pressure falls by more than 10 mm Hg during inspiration. A pulsus paradoxus also may be present with hypovolemic shock, chronic obstructive pulmonary disease, and bronchospasm.

CLINICAL MANIFESTATIONS

A slowly accumulating, isolated pericardial effusion is often completely asymptomatic. The physical examination results may be normal, but the heart sounds may be muffled. The diagnosis is usually suggested by a chest radiograph that shows cardiomegaly with a globular heart (Fig. 77-4; see also Fig. 56-5) or by an echocardiogram, computed tomography (CT) scan, or MRI performed for another indication. Patients with hypothyroidism, uremia, or collagen vascular disease may have asymptomatic effusions discovered during comprehensive evaluations.

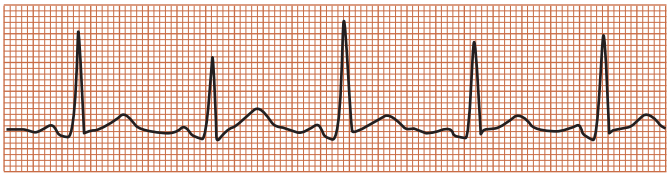

FIGURE 77-5. Electrical alternans. Lead V_5 rhythm strip from a patient with a large pericardial effusion and tamponade physiology. Note the relatively low voltage and electrical alternans.

Patients with impending or early tamponade are usually anxious and tachycardic, and they may complain of dyspnea, orthopnea, and chest pain. The increased venous pressure is usually apparent as jugular venous distention. The x descent (during ventricular systole) is typically the dominant jugular venous wave, with little or no y descent (during early diastole) (Chapter 51). The heart sounds are classically soft or muffled, especially if there is a large pericardial effusion. In rapidly developing cardiac tamponade, especially hemorrhagic cardiac tamponade, the jugular veins may not be distended because the time course has been insufficient for a compensatory increase in venous pressure. Such "low-pressure" tamponade may also occur with uremic pericarditis in volume-depleted patients. The patient may have signs of right heart failure, with peripheral edema, right upper quadrant pain caused by hepatic congestion, or abnormal liver enzymes and serum bilirubin level.

The hallmark of cardiac tamponade is a paradoxical pulse, which is defined as more than a 10–mm Hg drop in systolic arterial pressure during inspiration. When severe, the paradoxical pulse may be apparent as the absence of a palpable brachial or radial pulse during inspiration. A paradoxical pulse can also occur when there are wide swings in intrathoracic pressure, pulmonary embolism (Chapter 98), or hypovolemic shock (Chapter 106). A paradoxical pulse may be difficult to recognize in the presence of severe shock.

DIAGNOSIS

Cardiac tamponade, which is a treatable cause of shock (Chapter 107), can be rapidly fatal if unrecognized. As such, cardiac tamponade should be considered in the differential diagnosis of any patient with shock or pulseless electrical activity.

Cardiac tamponade is usually suspected based on jugular venous distention, sinus tachycardia with hypotension, narrow pulse pressure, elevated (>10 mm Hg) pulsus paradoxus, and distant heart sounds. Pulsus paradoxus may be obvious by palpation; it is more accurately measured with a sphygmomanometer during slow inspiration.

The ECG often shows low voltage and sometimes electrical alternans (Fig. 77-5) when the heart swings within a large pericardial effusion. The chest radiograph shows a globular, enlarged cardiac shadow (see Fig. 56-5 and Fig. 77-4) without pulmonary venous congestion.

Echocardiography, which is the key diagnostic test for cardiac tamponade, must be performed without delay in any patient suspected of having this condition. Echocardiography visualizes pericardial effusions as an echo-free space around the heart (Fig. 77-6), demonstrates the presence and size of the

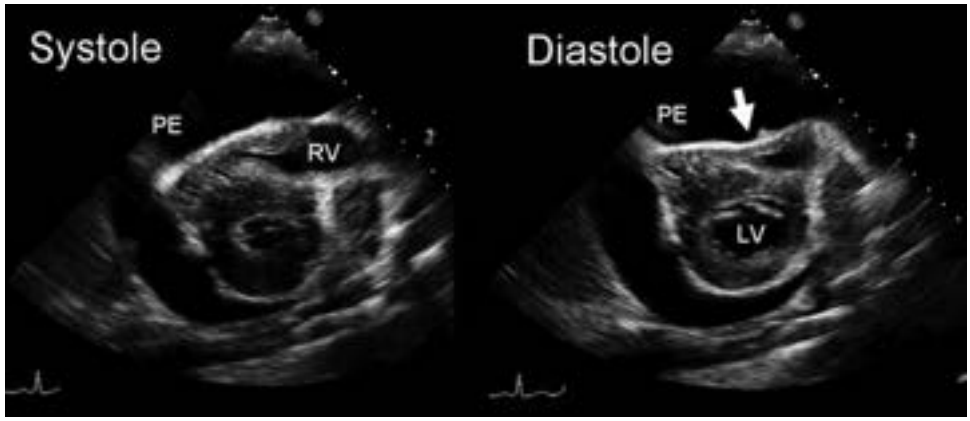

FIGURE 77-6. Two-dimensional echocardiogram from a patient with cardiac tamponade. A large pericardial effusion (PE) is apparent as an echo-free space surrounding the left ventricle (LV) and right ventricle (RV). In diastole, there is collapse of the right ventricle *(arrow)*.

pericardial effusion, and reflects its hemodynamic consequences. The inferior vena cava is almost always enlarged, right atrial and right ventricular collapse indicates cardiac compression, and enhanced respiratory variation of ventricular filling is a manifestation of increased ventricular interdependence. Right ventricular collapse is more specific for tamponade than is right atrial collapse, but the right-sided chambers may not collapse when tamponade occurs in patients with pulmonary hypertension. Cardiac tamponade can result from a loculated pericardial effusion after cardiac surgery or trauma. A loculated effusion may not be apparent on transthoracic echocardiography, but transesophageal echocardiography and thoracic CT or MRI can delineate loculated pericardial effusions.

On Doppler study, mitral inflow velocity (especially early diastolic velocity, designated as E velocity) normally increases with expiration and decreases with inspiration; the opposite respiratory variation is seen in tricuspid inflow velocity. Doppler findings for tamponade, which are more sensitive than two-dimensional echocardiography, include augmented respiratory variation of mitral and tricuspid inflow E velocities as a function of ventricular interdependence. These changes may be seen even before frank hemodynamic compromise caused by pericardial effusion. Although Doppler echocardiography provides important information, it must be emphasized that cardiac tamponade is ultimately a clinical diagnosis.

The routine evaluation should include an assessment of renal function, a thyroid-stimulating hormone level, a complete blood count with differential, a platelet count, coagulation parameters, and a tuberculin skin test. Common medications that can cause a pericardial effusion include cromolyn, isoniazid, and phenytoin; hydralazine, procainamide, and reserpine are others. Blood cultures are indicated if an infectious cause is suspected. Complement levels, antinuclear antibodies, and the ESR can suggest systemic lupus erythematosus (Chapter 266), which rarely presents initially as an isolated pericardial effusion.

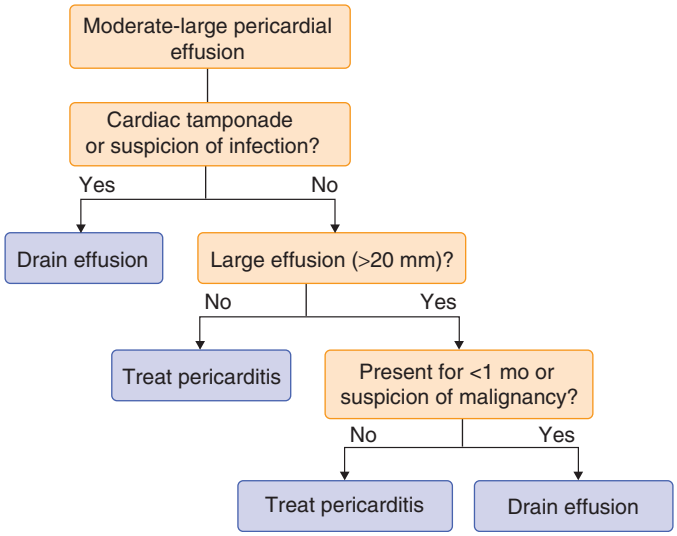

FIGURE 77-7. Algorithm for managing patients with moderate to large pericardial effusions. (Modified from Little WC, Freeman GL. Pericardial disease. *Circulation.* 2006;113:1622-1632.)

TREATMENT

Pericardial Effusion without Tamponade

Acute pericarditis is often accompanied by a small pericardial effusion that does not produce tamponade. If there is no hemodynamic compromise and the diagnosis can be established by other means, pericardiocentesis is not necessary (Fig. 77-7). Even small pericardial effusions may be related to underlying systemic illnesses such as systemic lupus erythematosus (Chapter 266), cardiac amyloid (Chapter 188), scleroderma (Chapter 267), hypothyroidism (Chapter 226), or AIDS, so it is important to consider and treat associated illnesses. Chylous pericardial effusion, which is usually related to obstruction of the thoracic duct, may require a surgical procedure for relief.

For suspected pericardial effusion, transthoracic echocardiography is the initial test of choice, although loculated effusions may be identified better by CT or MRI. If a small (0.5 to 1 cm) echolucent or "organized" pericardial effusion is observed, a follow-up echocardiogram in 1 to 2 weeks, or sooner if the patient deteriorates, is recommended. If the effusion is getting smaller, subsequent echocardiograms are not necessary unless the patient's clinical condition changes.

For moderate (1 to 2 cm) or large (>2 cm) effusions in patients who are hemodynamically stable and in whom tamponade is not suspected, a follow-up echocardiogram should be performed in 7 days and then every month until the effusion is minimal.[4] If bacterial or malignant pericarditis is suspected, diagnostic pericardiocentesis should be performed immediately even in the absence of clinical instability or suggestion of tamponade; tuberculous pericarditis is diagnosed best by pericardial biopsy. Anticoagulation with heparin or warfarin should be discontinued unless the patient has a mechanical heart valve or atrial fibrillation.

In hypothyroidism (Chapter 226), the effusion and the coexistent cardiomyopathy respond to hormone replacement, sometimes over several months. Uremic pericardial effusions often respond to initiation of dialysis or more intensive dialysis (Chapter 131).

Cardiac Tamponade

The treatment of cardiac tamponade is urgent drainage of the pericardial effusion, especially when there is hemodynamic compromise. Fluid resuscitation may be of transient benefit if the patient is volume depleted (hypovolemic cardiac tamponade), but inotropic agents are usually ineffective because there is already intense endogenous adrenergic stimulation. The initiation of mechanical ventilation in a patient with tamponade may produce a sudden drop in blood pressure because the positive intrathoracic pressure further impairs cardiac filling.

Echocardiographic-guided percutaneous pericardiocentesis, which can be performed at the bedside by experienced operators (Fig. 77-8), is indicated if a patient is in dire circumstances and at least 1 cm of fluid is seen anterior to the mid-right ventricular free wall throughout diastole. The ideal entry site (usually the apex) is defined using echocardiography as the minimal distance from the skin to pericardial fluid without intervening structures. The pericardial space is entered with a needle and then drained through a catheter. As much fluid as possible should be removed. The pericardial fluid should be sent for pH, glucose, lactate dehydrogenase, protein, cell count, and cytology as well as staining and culture for bacteria, fungi, and tuberculosis. Continued drainage of the pericardial fluid through an indwelling catheter minimizes the risk of recurrent effusion. For hemodynamically significant effusions of

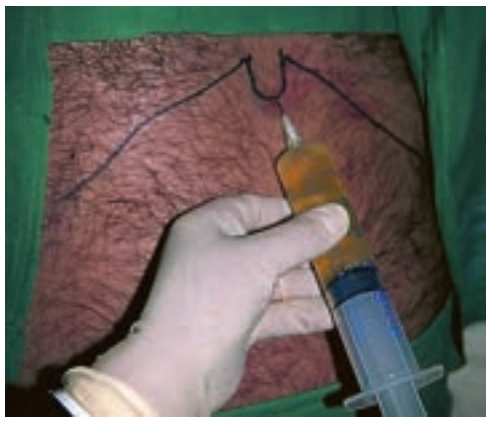

FIGURE 77-8. Aspiration of pericardial fluid is indicated in cardiac tamponade or to obtain fluid for diagnostic purposes. A wide-bore needle is inserted in the epigastrium below the xiphoid process and advanced in the direction of the medial third of the right clavicle. An alternative site is over the left ventricular apex. The procedure should be performed under echocardiographic guidance, but it may need to be performed emergently for life-saving purposes in other settings. Complications of the procedure include puncture of the heart, arrhythmias, vasovagal attack, and pneumothorax. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine.* 3rd ed. London: Mosby; 2003.)

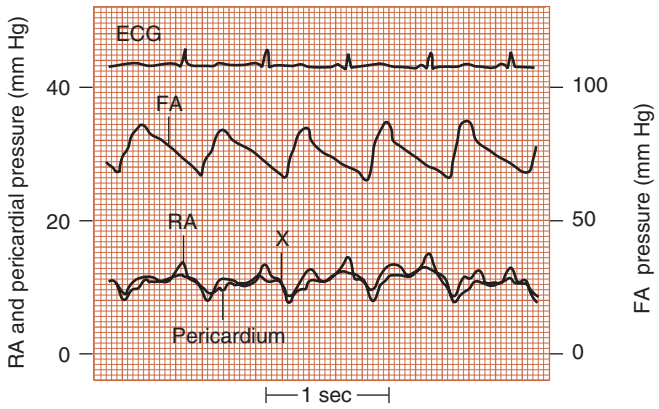

FIGURE 77-9. Right atrial (RA) pressure recording from a patient with constrictive pericarditis. Note the elevation in pressure and the prominent y descent, corresponding to rapid, early diastolic right atrial emptying. ECG = electrocardiogram; FA = femoral artery. (From Lorell BH. Profiles in constriction, restriction and tamponade. In: Baim DS, Grossman W, eds. *Cardiac Catheterization, Angiography, and Intervention.* 6th ed. Philadelphia: Williams Wilkins; 2000: 832.)

less than 1 cm, organized or multiloculated effusions, and focal effusions, a limited thoracotomy-mediastinoscopy and creation of a pericardial window are advised.

Surgical drainage may be the preferred treatment if pericardial tissue is required for diagnosis or in the case of recurrent effusions or bacterial pericarditis. Malignant pericardial effusions frequently reoccur and, similar to other recurrent pericardial effusions, may necessitate the surgical creation of a pericardial window that allows the effusion to drain into the pleural space, preventing reoccurrence of cardiac tamponade. An attractive alternative in these patients, especially if their overall prognosis is poor from the malignancy, is the percutaneous creation of a pericardial window by balloon dilation. Hemorrhagic effusions related to cardiac trauma or aortic dissection are best managed by emergency surgery.

PROGNOSIS

A pericardial effusion may recur or persist. Symptoms are usually weight loss, fatigue, dyspnea on exertion, and whatever symptoms are associated with the specific cause. Treatment of chronic or recurrent idiopathic effusions is similar to the treatment of recurrent pericarditis. If medical therapy is unsuccessful, creation of a pericardial window is indicated.

A large idiopathic, asymptomatic effusion that persists for 6 months or longer can unpredictably result in tamponade in as many as 30% of patients over long-term follow-up; diagnostic pericardiocentesis occasionally detects a neoplastic or tuberculous cause. Pericardiocentesis with prolonged drainage resolves many chronic large pericardial effusions, but pericardiectomy is often required. The long-term prognosis depends on the cause of the effusion. With pericardial tamponade, the in-hospital mortality rate is less than 10%, but the subsequent mortality rate is about 75% with a malignant effusion compared with only a 3% to 5% subsequent annual mortality rate for other causes.

PERICARDIAL CONSTRICTION

EPIDEMIOLOGY AND PATHOBIOLOGY

Pericardial constriction, which is usually the result of long-standing pericardial inflammation, occurs when a scarred, thickened, or calcified pericardium impairs cardiac filling, thereby limiting the total cardiac volume. The most frequent causes in the developed world are previous cardiac surgery, chronic idiopathic or viral pericarditis, and mediastinal radiation. Constriction may follow cardiac surgery by several weeks to months and may occur decades after chest wall irradiation. In developing countries, tuberculous pericarditis is a more common cause of constrictive pericarditis. Other less common causes include malignant disease, especially lung cancer, breast cancer, or lymphoma; histoplasmosis; rheumatoid arthritis; and uremia. However, a specific cause may not be identified in many patients.

With chronic constriction, the pericardium may thicken from its normal 2 mm or less, calcify, and adhere to the epicardium. In a subset of the patients with constriction, the pericardium may be only minimally thickened and less calcified. Fibrous scarring and adhesions of both pericardial layers obliterate the pericardial cavity. The ventricles are unable to fill because of physical constraints imposed by a thickened, rigid, and sometimes calcified pericardium. The pathophysiologic hallmark of pericardial constriction is the exaggerated interventricular dependence and differential ventricular filling with respiration.

Although both cardiac tamponade and pericardial constriction impair diastolic ventricular filling and elevate venous pressure, the impairment in ventricular filling with constriction is minimal in early diastole until cardiac volume reaches the anatomic limit set by the noncompliant pericardium, at which time diastolic pressure rises abruptly and remains elevated until the onset of systole. This prominent y descent with an elevated plateau of ventricular pressure, which has been termed the "square root" sign (Fig. 77-9), differentiates constriction from tamponade, in which the y descent is absent. Stroke volume and cardiac output are reduced because of impaired filling, but the intrinsic systolic function of the ventricles can be normal.

CLINICAL MANIFESTATIONS

Patients with pericardial constriction typically present with manifestations of elevated systemic venous pressures and low cardiac output. Because there is equalization of all cardiac pressures (including right and left atrial pressures), systemic congestion is much more marked than pulmonary congestion. Typically, patients develop marked jugular venous distention, hepatic congestion, ascites, and peripheral edema, but their lungs remain clear. The limited cardiac output typically presents as exercise intolerance and may progress to cardiac cachexia with muscle wasting. In long-standing pericardial constriction, pleural effusions, ascites, and hepatic dysfunction may be prominent clinical features. Patients with pericardial constriction are much more likely to have left-sided or bilateral pleural effusions than right-sided effusions. Because of the prominent clinical symptoms of ascites and liver enzyme abnormalities, patients may be evaluated for hepatic disease before constrictive pericarditis is recognized.

The jugular veins are distended with prominent x and y descents. The normal inspiratory drop in jugular venous distention may be replaced by a rise in venous pressure (Kussmaul sign). The classic auscultatory finding of pericardial constriction is a pericardial knock (Chapter 51), which is a high-pitched sound early in diastole when there is the sudden cessation of rapid ventricular diastolic filling, coinciding with the nadir of the y descent.

DIAGNOSIS

Pericardial constriction should be considered in any patient with unexplained systemic venous congestion. Pericardial calcification, seen best on the lateral plain chest radiograph, is a classic finding but is present in only 25% of patients with constrictive pericarditis, mostly in those with long-standing constriction. Similarly, most patients with pericardial constriction have a thickened pericardium (>2 mm) that can be imaged by echocardiography, CT, and MRI (Fig. 77-10). It is important to recognize, however, that pericardial constriction can be present without pericardial calcification and, in about 20% of patients, without any obvious pericardial thickening.

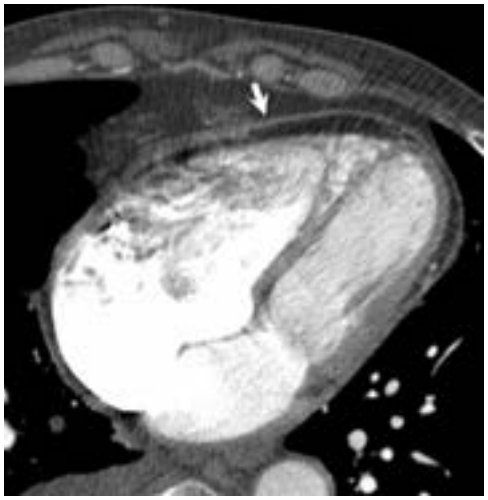

FIGURE 77-10. Computed tomography in a patient with constrictive pericarditis shows a thickened pericardium *(arrow)*.

Transesophageal Doppler echocardiography may demonstrate pericardial thickening and calcification, but increased pericardial thickness can be missed on a transthoracic echocardiogram. Echocardiography also differentiates pericardial constriction from right heart failure caused by tricuspid valve disease or associated pulmonary hypertension.

Differential Diagnosis

The most difficult differentiation is between pericardial constriction and restrictive cardiomyopathy (Chapter 60), the clinical manifestations of which may be very similar to those of pericardial constriction (Table 77-6). Doppler echocardiography is the most useful method to distinguish constriction from restriction.[5] Whereas patients with pericardial constriction usually have pronounced respiratory variation (>25%) of mitral inflow E velocity, patients with restrictive cardiomyopathies do not. In some patients with pericardial constriction and markedly elevated venous pressures, the respiratory variation may be present only after head-up tilt. The tissue Doppler measurement of early diastolic septal mitral annular velocity (e′) is almost always reduced in patients with myocardial restriction, but it remains normal or increased in patients with pericardial constriction. In addition, lateral e′, which is higher than septal or medial e′ velocity in normal and restrictive cardiomyopathy, is lower than septal e′ in most patients with constrictive pericarditis. Whereas a prominent diastolic reversal of hepatic vein flow velocity during expiration is characteristic of constriction, the reversal flow velocity occurs during inspiration in patients with right heart failure from other causes. Patients with pericardial constriction usually have only minimally elevated (<200 pg/mL) brain natriuretic peptide (BNP), but BNP levels are typically markedly increased (<600 pg/mL) in patients with restrictive cardiomyopathy.

Confirmation of the diagnosis of constriction may require cardiac catheterization in patients whose noninvasive evaluation is not clear cut. Traditional invasive hemodynamic findings of equalized end-diastolic pressures in the right and left ventricles and the "dip and plateau" pattern of left ventricular diastolic pressure do not reliably differentiate constriction from restrictive cardiomyopathy. More specific invasive hemodynamic features of constriction and restriction are based on the respiratory variation in ventricular filling; the simultaneous measurement of left and right ventricular pressures demonstrates discordant changes in their systolic pressures with respiration in constrictive pericarditis. By comparison, the direction of these pressures is concordant (both left and right sides increase with expiration and decrease with inspiration) in restrictive cardiomyopathy.

All patients with documented but otherwise unexplained pericardial constriction should be evaluated for potential tuberculosis.

TREATMENT AND PROGNOSIS

Rx

In some patients with pericardial constriction of less than 3 months' duration, the symptoms and constriction may resolve over several weeks with medical therapy consisting of NSAIDs (e.g., ibuprofen 300-800 mg every 6-8 hours), colchicine (0.6 mg once or twice daily), and the cautious use of diuretics. If the patient is severely compromised hemodynamically, steroid therapy (e.g., 0.5-1.0 mg/kg up to a maximum dose of 60 mg tapered slowly over 3 months as guided by the clinical response and normalization of the CRP) can be very effective.[6] The patients with constriction who respond to medical therapy usually have high inflammatory biomarkers (ESR and CRP) and intense inflammation of the pericardium on cardiac MRI.[7] For more chronic pericardial constriction or cases that do not respond to medical therapy, the definitive treatment is surgical pericardial decortication, with a wide resection of both the visceral and parietal pericardium. This operation is a major undertaking with substantial risk (≈10% mortality rate even in the most experienced centers).[8] In many patients, surgery does not immediately restore normal cardiac function; it may take weeks after removal of the constricting pericardium to return to normal. Empirical treatment of tuberculosis (Chapter 324) may be required in patients with constriction and a high suspicion of tuberculosis, even without a definitive diagnosis.

TABLE 77-6 DIFFERENTIATION OF PERICARDIAL CONSTRICTION FROM RESTRICTIVE CARDIOMYOPATHY

FINDINGS	PERICARDIAL CONSTRICTION	RESTRICTIVE CARDIOMYOPATHY
PHYSICAL EXAMINATION		
Pulmonary congestion	Usually absent	Usually present
Early diastolic sound	Pericardial knock	S_3 (low pitched)
ECHO/DOPPLER		
Respiratory variation in E wave (%)	>25	<20
Mitral septal annular early diastolic velocity (cm/sec)	>7	<7
CT/MRI		
Pericardial thickness	>2 mm (but <2 mm in 20%)	<2 mm
BIOMARKER		
B-type natriuretic peptide (pg/mL)	<200	>600
HEMODYNAMICS		
PA systolic pressure (mm Hg)	<60	>60
PCW-LV diastolic pressure	Respiratory variation with reduction in inspiration	No variation
Respiratory variation in RV/LV peak systolic pressure	Discordant	Concordant

CT = computed tomography; LV = left ventricular; MRI = magnetic resonance imaging; PA = pulmonary artery; PCW = pulmonary capillary wedge; RV = right ventricular.
Modified from Little WC, Freeman GL. Pericardial disease. *Circulation.* 2006;113:1622-1632.

Effusive-Constrictive Pericarditis

In some patients (<10%) who present with cardiac tamponade, the elevated right atrial pressure and jugular venous distention do not resolve after removal of the pericardial fluid. In these patients, pericardiocentesis converts the hemodynamics from those typical of tamponade to those of constriction. Thus, the restriction of cardiac filling is not only attributable to pericardial effusion but also to pericardial constriction, predominantly involving the visceral pericardium. Effusive-constrictive pericarditis[9] most likely represents an intermediate transition from acute pericarditis with pericardial effusion to pericardial constriction. Frequently, the symptoms resolve after several weeks of treatment with an NSAID (e.g., ibuprofen 300-800 mg every 6-8 hours).

SPECIFIC FORMS OF PERICARDIAL DISEASE

Postcardiotomy Syndrome

Postcardiotomy syndrome is acute pericarditis occurring weeks to months after open heart surgery. Patients have typical symptoms of acute pericarditis, which is associated with antimyocardial antibodies. A similar clinical picture is seen with the postperfusion syndrome caused by cytomegalovirus (CMV) infection (Chapter 376) in patients who were previously uninfected but were exposed to CMV-positive blood during cardiopulmonary bypass or

transfusions. Atypical lymphocytes and elevated liver enzymes are often seen. Treatment of postcardiotomy syndrome is similar to the treatment of acute pericarditis. One month of colchicine therapy (0.5 mg twice daily for patients weighing 70 kg or more; half that dose for smaller patients or patients with side effects) started 48 to 72 hours prior to cardiac surgery appears to reduce the rate of the postcardiotomy syndrome from about 30% to less than 20% but did not reduce the occurrence of atrial fibrillation.

Some patients who have undergone cardiac surgery develop late pericardial constriction without preceding acute pericarditis owing to bleeding in and around the open pericardium followed by inflammation, scarring, and fibrosis. If active inflammation is present, a trial of antiinflammatory agents may be instituted; however, surgical removal of the blood clot, usually accompanied by more extensive pericardiectomy, is typically required.

Post–Myocardial Infarction Pericarditis

Acute pericarditis can develop several days after an acute MI (Chapter 73), usually because of transmural extension of the infarction to the pericardial surface. This syndrome is uncommon in the reperfusion era. Anticoagulation should be temporarily withheld to avoid a bloody effusion that might progress to cardiac tamponade.

A late autoimmune pericarditis, termed *Dressler syndrome*, can develop weeks to months after a Q-wave MI, but this syndrome is very uncommon since the advent of reperfusion therapy. Diagnosis and treatment are as for acute pericarditis.

Uremic Pericarditis

Pericardial effusions develop in patients with severe renal failure, especially those on dialysis (Chapter 131). Aggressive dialysis may decrease the pericardial effusion, but sometimes the effusion persists and requires drainage or even pericardiectomy.

Infectious Pericarditis

BACTERIAL PERICARDITIS

Purulent bacterial pericarditis may result from direct extension of bacterial pneumonia; direct extension of pleural empyema; or, rarely, peritonitis or a subphrenic abscess. Most patients are acutely ill with systemic sepsis and develop acute tamponade. The most common organisms are streptococci, pneumococci, and staphylococci. Urgent pericardiocentesis, which is required for both diagnosis and therapy, shows leukocytosis and frank pus, and the fluid glucose level is markedly depressed. Persistent or recurrent drainage with an indwelling catheter or repeated taps combined with antimicrobial therapy lead to a high survival rate,[10] but late constrictive pericarditis requiring pericardiectomy develops in 30% to 40% of patients.

TUBERCULOUS PERICARDITIS

Tuberculous pericarditis is common in developing countries but accounts for less than 5% of cases of acute pericarditis in developed countries, usually in patients who are immunosuppressed, including patients infected with the human immunodeficiency virus.[11] Symptoms are often nonspecific, and acute painful pericarditis is rare. Most patients have an effusive-constrictive physiology or pericardial constriction. The chest radiograph suggests active pulmonary tuberculosis is about 30% of cases, and a pleural effusion is present in 40% to 60% of cases. The echocardiogram typically shows fibrinous strands in the pericardial effusion, with multiple echo densities adherent to the pericardial surface. Pericardiocentesis is mandatory if tuberculous pericarditis is suspected. About 75% of patients have a positive culture, and a pericardial fluid adenosine deaminase level of 40 U/L or greater is seen in 75% of patients. Polymerase chain reaction testing and pericardial biopsy are recommended when tuberculous pericarditis is strongly suspected but not otherwise confirmed.

Aggressive antituberculosis therapy (Chapter 324) yields a cure rate of about 85% to 90%. Neither corticosteroids nor adjunctive immunotherapy improves outcome. Empirical treatment may be required in patients with a consistent clinical picture but without a confirmed diagnosis.

Even with prompt treatment, 30% to 60% of patients develop constrictive pericarditis, for which surgical pericardiectomy is the treatment of choice. In cases of suspected tuberculous constriction, antituberculosis therapy (Chapter 324) should be administered before and after pericardial surgery.

FUNGAL PERICARDITIS

The most common fungal pericarditis is histoplasmosis, which usually resolves in several weeks and can be treated successfully with NSAIDs. Specific antifungal therapy (Chapter 331) is recommended only in patients with disseminated histoplasmosis (Chapter 332).

Malignant Pericarditis

About 6% of cases of acute pericarditis that initially have no obvious cause, and about 20% of cases of moderate to large pericardial effusions are related to malignant diseases; in patients with cardiac tamponade, the percentage is even higher. About 80% of malignant pericarditis is linked to breast cancer (Chapter 198), lymphoma (Chapters 185 and 186), and leukemia (Chapters 183 and 184). Melanoma (Chapter 203) is an uncommon cause of malignant pericarditis, but a large proportion of patients with melanoma have pericardial involvement.

Most patients have direct tumor extension from an adjacent malignant lesion or a tumor from hematogenous or lymphatic spread. Pericardial effusions also may be caused by pericardial irritation or compromised lymphatic drainage in patients with mediastinal lymphoma.

Pericardiocentesis is key to diagnosis and management. Fluid cytology results are positive in about 85% of patients. Complete drainage with an indwelling catheter for 2 or 3 days is the treatment of choice; if the effusion does not resolve, pericardiectomy is recommended.[12] The prognosis depends on the treatment of the underlying malignancy, but the 1-year mortality rate is 80% or higher.

Postradiation Pericarditis

Postradiation pericarditis develops in about 2% of patients after mantle radiation for Hodgkin disease (Chapter 186) and in 0.4% to 5% of patients after irradiation for breast cancer. Pericardial injury occasionally manifests during treatment, but it more commonly appears months or even a decade later. The initial pericarditis and effusion may resolve spontaneously, but constrictive pericarditis, adjacent myocarditis, and even coronary artery damage can develop. Pericardiocentesis is critical to distinguish postradiation pericarditis from malignant pericardial disease. Pericardiectomy is recommended for recurrent pericarditis and for large recurrent pericardial effusions.

Autoimmune Pericarditis

Up to 50% of patients with systemic lupus erythematosus (Chapter 266) have pericarditis, usually during an acute flare. Patients usually present with acute pericarditis or an asymptomatic effusion; cardiac tamponade is uncommon, and constrictive pericarditis is rare. If purulent pericarditis is not suspected, pericardiocentesis is not usually required. The underlying disease should be treated aggressively.

Acute pericarditis or asymptomatic pericardial effusions can develop in patients with advanced rheumatoid arthritis (Chapter 264), scleroderma (Chapter 267), or mixed connective tissue disease (Chapter 267). The process is usually self-limited or responds to aggressive treatment of the underlying disease, although cardiac tamponade can develop.

Myopericarditis

Concomitant myopericarditis (Chapter 60) may develop in patients with pericarditis, probably owing to direct extension of the inflammatory process. It is often manifested by ECG conduction delays, ventricular arrhythmias, and elevated troponin levels. No specific treatment is available.

Acute ventricular dilation and sudden or progressive heart failure develop in some patients after pericardiectomy for constrictive pericarditis or even after drainage of a large pericardial effusion. The syndrome may be underlying myocarditis.

Congenital Abnormalities

Congenital total absence of the pericardium is asymptomatic and clinically unimportant. However, partial or localized absence of the pericardium around the left atrium can cause focal herniation and lead to strangulation. CT or MRI can establish the diagnosis. Patients may present with atypical chest pain or sudden death. Surgical repair is often recommended for a partial pericardial defect.

Benign Cysts

Benign pericardial cysts are rare and are usually asymptomatic, but they can be associated with chest pain. They are typically seen as rounded or lobulated structures adjacent to the heart on the chest radiograph or adjacent to the right atrium on transthoracic echocardiography (Chapter 55). Thoracic CT and MRI are useful for the diagnosis. Cysts rarely rupture and do not require treatment unless they become symptomatic with chest pain. In this

situation, the cyst can be removed surgically or drained percutaneously or with a thoracoscope.

Grade A References

A1. Imazio M, Brucato A, Cemin R, et al. Colchicine for recurrent pericarditis (CORP): a randomized trial. *Ann Intern Med*. 2011;155:409-414.

A2. Imazio M, Brucato A, Cemin R, et al. A randomized trial of colchicine for acute pericarditis. *N Engl J Med*. 2013;369:1522-1528.

A3. Imazio M, Belli R, Brucato A, et al. Efficacy and safety of colchicine for treatment of multiple recurrences of pericarditis (CORP-2): a multicentre, double-blind, placebo-controlled, randomised trial. *Lancet*. 2014;383:2232-2237.

A4. Imazio M, Brucato A, Ferrazzi P, et al. Colchicine for prevention of postpericardiotomy syndrome and postoperative atrial fibrillation: the COPPS-2 randomized clinical trial. *JAMA*. 2014;312:1016-1023.

A5. Mayosi BM, Ntsekhe M, Bosch J, et al. Prednisolone and *Mycobacterium indicus pranii* in tuberculous pericarditis. *N Engl J Med*. 2014;371:1121-1130.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Imazio M, Brucato A, Barbieri A, et al. Good prognosis for pericarditis with and without myocardial involvement: results from a multicenter, prospective cohort study. *Circulation*. 2013;128:42-49.
2. Lilly LS. Treatment of acute and recurrent idiopathic pericarditis. *Circulation*. 2013;127: 1723-1726.
3. Khandaker MH, Schaff HV, Greason KL, et al. Pericardiectomy vs medical management in patients with relapsing pericarditis. *Mayo Clin Proc*. 2012;87:1062-1070.
4. Imazio M, Adler Y. Management of pericardial effusion. *Eur Heart J*. 2013;34:1186-1197.
5. Welch TD, Ling LH, Espinosa RE, et al. Echocardiographic diagnosis of constrictive pericarditis: Mayo Clinic criteria. *Circ Cardiovasc Imaging*. 2014;7:526-534.
6. Syed FF, Schaff HV, Oh JK. Constrictive pericarditis—a curable diastolic heart failure. *Nat Rev Cardiol*. 2014;11:530-544.
7. Feng D, Glockner J, Kim K, et al. Cardiac magnetic resonance imaging pericardial late gadolinium enhancement and elevated inflammatory markers can predict the reversibility of constrictive pericarditis after antiinflammatory medical therapy: a pilot study. *Circulation*. 2011;124:1830-1837.
8. Tokuda Y, Miyata H, Motomura N, et al. Outcome of pericardiectomy for constrictive pericarditis in Japan: a nationwide outcome study. *Ann Thorac Surg*. 2013;96:571-576.
9. Syed FF, Ntsekhe M, Mayosi BM, et al. Effusive-constrictive pericarditis. *Heart Fail Rev*. 2013;18:277-287.
10. Imazio M, Brucato A, Mayosi BM, et al. Medical therapy of pericardial diseases: part I: idiopathic and infectious pericarditis. *J Cardiovasc Med (Hagerstown)*. 2010;11:712-722.
11. Ntsekhe M, Mayosi BM. Tuberculous pericarditis with and without HIV. *Heart Fail Rev*. 2013;18: 367-373.
12. Burazor I, Imazio M, Markel G, et al. Malignant pericardial effusion. *Cardiology*. 2013;124: 224-232.

78 DISEASES OF THE AORTA

FRANK A. LEDERLE

The ascending aorta, which is located in the anterior mediastinum, is about 3 cm in diameter and 5 cm long. The aortic root, which is just above the aortic valve, is composed of the three sinuses of Valsalva. The ascending aorta meets the aortic arch in the superior mediastinum, where the brachiocephalic arteries branch off it. The descending thoracic aorta, which is about 2.5 cm in diameter and 20 cm long, courses posteriorly, crosses the diaphragm, and becomes the abdominal aorta, which is normally 2 cm in diameter and extends for 15 cm until it bifurcates into two common iliac arteries.

The aorta itself is composed of three layers. The thin inner layer, the *intima*, is lined with endothelial cells. In the thick middle layer, the *media*, sheets of elastic tissue provide the tensile strength to withstand needed systolic pressures. The outer *adventitia*, which is composed mostly of collagen, provides arterial and venous blood supply to the aorta itself.

AORTIC ANEURYSMS

DEFINITION

An aneurysm is a pathologic dilation of the artery, often defined as a 50% increase over the expected diameter. An aneurysm can be defined based on its cause, location, shape, and size. In terms of shape, whereas a fusiform aneurysm is a symmetrical dilation of the aorta, a saccular aneurysm involves dilation mainly of one wall. A false aneurysm or pseudoaneurysm occurs when the aorta is enlarged because of dilation of only the outer layers of the vessel wall, as may occur with a contained rupture of the aortic wall.

EPIDEMIOLOGY

Although an aneurysm can develop in any part of the aorta, abdominal aortic aneurysms are much more common than thoracic aneurysms. Thoracic aortic aneurysms are most common in the ascending aorta followed in frequency by the descending aorta and the aortic arch. When an aneurysm in the descending thoracic aorta extends into the abdominal aorta, it is termed a *thoracoabdominal aortic aneurysm*.

For abdominal aortic aneurysms, risk factors include increased age, male gender, smoking, a family history of the disease, and occlusive atherosclerotic disease. Diabetes and black race are associated with a reduced risk.

Ascending thoracic aortic aneurysms are often associated with genetic mutations as in Marfan or Ehlers-Danlos syndromes (Chapter 260). Risk factors for descending thoracic and thoracoabdominal aortic aneurysms include age, smoking, and chronic obstructive pulmonary disease. Compared with abdominal aortic aneurysms, thoracic aneurysms have a stronger familial component and no gender predilection.

Deaths from abdominal aortic aneurysm increased markedly from 1950 to 1970 but have been declining in the United States since the 1990s (Fig. 78-1). The prevalence of an asymptomatic abdominal aortic aneurysm, defined as diameter larger than 3 cm, detected by screening, has also declined sharply, from more than 5% in older men in 1990 to less than 2% by 2010.[1] The most likely explanation for these changes is the increase and then decrease in U.S. rates of smoking, which accounts for about three fourths of all abdominal aortic aneurysms. By comparison, global death rates for abdominal aortic aneurysm have not declined because smoking rates and levels of other cardiovascular risk factors have not declined.[2] In the United States, deaths from rupture of thoracic aortic aneurysms also decreased by 50% from 1997 to 2010.

PATHOBIOLOGY

Atherosclerosis was once considered the most common underlying cause of abdominal aortic aneurysms, but this relationship has been called into question by a variety of observations. For example, abdominal aortic aneurysm is less common in patients with diabetes and is much more strongly associated with smoking and male gender than is atherosclerosis.[1,2] Although aortic atherosclerosis may contribute to the process, the pathogenesis of abdominal aortic aneurysms appears to include genetic, environmental, hemodynamic, and immunologic factors. The strength of the aortic wall depends on elastin and collagen within the extracellular matrix of its media. The degradation of these structural proteins by matrix metalloproteinases and inflammatory infiltrates, especially macrophages and T lymphocytes, weakens the aortic wall and allows aneurysms to develop. However, attempts to retard the

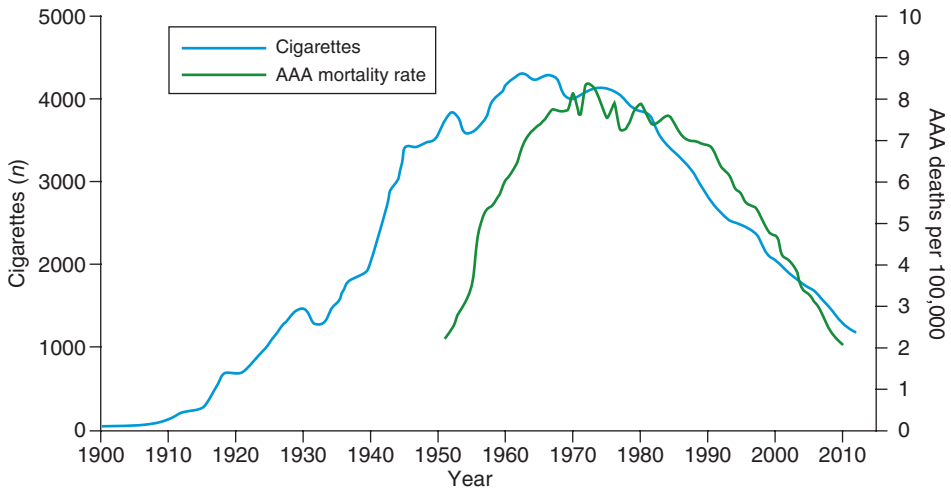

FIGURE 78-1. U.S. annual adult per capita cigarette consumption and U.S. age-adjusted abdominal aortic aneurysm (AAA) mortality rate per 100,000 white men by year. Rates include abdominal, thoracoabdominal, and unspecified aortic aneurysm and exclude thoracic aneurysm and dissection. (Adapted from Lederle FA. The rise and fall of abdominal aortic aneurysm. *Circulation.* 2011;124:1097-1099.)

process using doxycycline, which suppresses matrix metalloproteinase, have not been successful.

In the ascending thoracic aorta, the most important cause of aneurysms is cystic medial degeneration, with necrosis of smooth muscle cells and degeneration of elastic layers within the media, often related to a genetic mutation such as Marfan or Ehlers-Danlos syndrome. Almost all patients with Marfan syndrome, who are at very high risk for developing thoracic aortic aneurysms, have underlying cystic medial degeneration. In patients who develop ascending aortic aneurysms without overt evidence of connective tissue disease, a bicuspid aortic valve and familial thoracic aortic aneurysm syndrome are important congenital causes. Syphilis (Chapter 319) was formerly a common cause of thoracic aortic aneurysms in the United States but is rarely implicated now. Other uncommon causes of ascending thoracic aortic aneurysms include infectious aortitis, great vessel arteritis, aortic trauma, and aortic dissection. Descending thoracic and thoracoabdominal aortic aneurysms can also result from previous dissection, but they are usually idiopathic and degenerative.

CLINICAL MANIFESTATIONS

The majority of aortic aneurysms are asymptomatic and are discovered incidentally on an imaging study or on routine abdominal palpation. Symptoms resulting from progressive enlargement of unruptured aneurysms, such as vertebral erosion or nerve compression, are very rare.

Abdominal aortic aneurysm rupture can cause sudden death from cardiovascular collapse. Other patients present with pain in the hypogastrium, flank, lower back, or hips. The pain is often severe and frightening to the patient and may be accompanied by abdominal tenderness. Symptoms of bloating, constipation, or urinary retention may be caused by a hematoma that compresses the bowel or the urinary tract or by impaired blood supply to these organs. These symptoms typically persist for hours to a few days before death ensues or the aneurysm is repaired, although a few cases of chronic contained rupture persisting for weeks or even months have been reported.

Thoracic aortic aneurysm can expand and compress adjacent mediastinal structures. Symptoms include coughing, wheezing, dyspnea, hoarseness, recurrent pneumonia, and dysphagia. A ruptured thoracic aneurysm may present with chest or back pain. Vascular complications include aortic insufficiency, sometimes with secondary heart failure, hemoptysis, and arterial thromboembolism.

DIAGNOSIS

Abdominal aortic aneurysms may be detectable on deep abdominal palpation as a pulsatile mass, although obesity can obscure even large aneurysms. The examiner first feels deeply for the aortic pulsation, usually found a few centimeters cephalad of the umbilicus (the umbilicus marks the level of the aortic bifurcation) and slightly to the left of midline. The examiner then positions both hands on the abdomen with the palms down and an index finger on either side of the pulsation, both to confirm that it is the aorta (each systole should move the two fingers apart) and to measure the aortic width. Sufficient abdominal skin should be included between the two index fingers, and time should be taken to allow the abdominal muscles to relax. On examination, the width, not the intensity, of the pulsation guides the diagnosis. A pulsatile abdominal mass should be confirmed by ultrasonography, which provides an accurate measure of diameter; a width of 3 cm or larger establishes the diagnosis and warrants ongoing surveillance.

Clinical diagnosis of a rupturing abdominal aortic aneurysm can be challenging. The classic triad of abdominal pain, hypotension, and a pulsatile abdominal mass is insensitive because the blood pressure may be normal or near normal at presentation, and palpation may be difficult owing to guarding or bloating. Furthermore, accompanying bowel and bladder symptoms can lead to misdiagnoses. In patients with ruptured aneurysms, the white blood cell count is usually elevated, but the hematocrit can initially be normal because hemodilution does not occur acutely. The physician should have a low threshold for obtaining appropriate imaging, particularly in older men with a history of smoking.

Abdominal aortic aneurysms can be detected and measured by either abdominal ultrasonography or computed tomography (CT) (Fig. 78-2). However, CT should be obtained when rupture is considered because rupture is not reliably diagnosed by ultrasonography.

Thoracic aortic aneurysms usually cannot be palpated even when they are very large. As a result, thoracic aortic aneurysms are frequently recognized on chest radiographs, where they often are diagnosed by a widened mediastinal silhouette, enlarged aortic knob, or a trachea that is displaced from the midline. CT is accurate for detecting and measuring thoracic aneurysms and monitoring their diameter over time. Transthoracic echocardiography, which generally visualizes the aortic root and ascending aorta, is useful for screening patients with Marfan syndrome (Chapter 260), who are at particular risk for aneurysms involving this portion of the aorta.

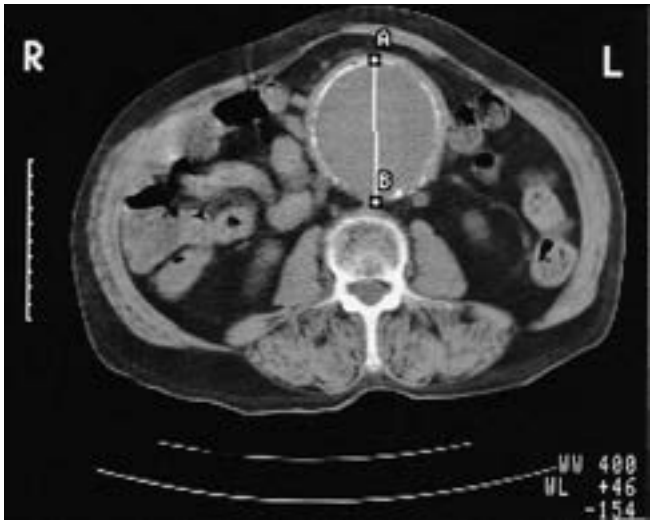

FIGURE 78-2. Abdominal aortic aneurysm on computed tomography. This sensitive imaging method allows precise measurement of size (*point A* to *point B*) and demonstrates the thickened wall of the aneurysm. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine*. 3rd ed. London: Mosby; 2003.)

TREATMENT

Rx

The morbidity and mortality of aortic aneurysms result from both rupture and elective repair. A judicious strategy for intervention is therefore essential. Open surgical repair consists of insertion of a synthetic prosthetic tube graft. When aneurysms involve branch vessels, such as the renal or mesenteric arteries, the vessels must be reimplanted into the graft. Similarly, when a dilated aortic root must be replaced in the repair of an ascending thoracic aortic aneurysm, the coronary arteries must be reimplanted. An alternative approach to repair abdominal aortic aneurysms and some descending thoracic aneurysms is the percutaneous placement of an expandable endovascular stent graft inside the aneurysm.

In a patient with a ruptured aortic aneurysm, emergent repair by either the open or endovascular method is required. Randomized trials in patients with stab wounds and upper gastrointestinal bleeding and observational studies of ruptured abdominal aortic aneurysms have raised concerns that excessive volume expansion and transfusion before control of bleeding may increase the mortality rate.

For patients with asymptomatic abdominal aortic aneurysms smaller than 5.5 cm in diameter, elective repair, by either the open or endovascular method, does not reduce the overall mortality rate. The benefit of elective repair of larger abdominal aortic aneurysms has been demonstrated indirectly through randomized trials of ultrasound screening, which reduces both aneurysm-related and total mortality. Elective endovascular repair has a lower postoperative mortality rate than open repair (1.5% vs. 4%), but excess late deaths after endovascular repair result in similar survival rates after 3 to 5 years. Patients who have large aneurysms and who are medically unfit for open repair have a high rupture rate[3] but may not benefit from endovascular repair because of their burdens of comorbid conditions. Randomized trial data are insufficient for women, who have both a higher rate of rupture and a higher operative mortality rate.

Small aortic aneurysms should be monitored periodically to detect progressive enlargement, which may indicate the need for surgical repair. Whereas ultrasonography is the preferred modality for monitoring abdominal aneurysms, CT is used for thoracic aneurysms. Surveillance intervals are based on the probability of exceeding the operative threshold.

For abdominal aortic aneurysms, proposed screening intervals are every 3 years for aneurysms of 3.0 to 3.9 cm in diameter, 2 years for 4.0 to 4.4 cm in diameter, and yearly for 4.5 to 5.4 cm in diameter. Elective repair should be considered when the diameter exceeds 5.5 cm.[4] High-volume surgeons and hospitals have better outcomes from elective repair.

For asymptomatic thoracic aneurysms, management is less certain. No randomized trials have addressed when thoracic aortic aneurysms should be repaired or which method is preferable, and data on natural history are limited.

Endovascular repair of unruptured descending thoracic aortic aneurysms is associated with lower perioperative mortality rate than open repair (5%-6% vs. 7%-12%), but it does not improve the adjusted 5-year survival rate.[5] Normal aortic diameter is 1.5 to 2 times greater, and the operative mortality rate is two to four times higher in the thoracic aorta compared with the abdominal aorta, and the risk of rupture is low for thoracic aneurysms less than 6.0 cm in diameter. These comparisons suggest that diameter thresholds for elective repair should probably be higher than the 5.5 cm established for abdominal aortic aneurysms. An exception may be patients with Marfan syndrome, in whom the risk of dissection or rupture is higher and operative mortality rate is lower, perhaps because of their younger age. Patients with Marfan syndrome with aortic diameters less than 5 cm are at low risk, but natural history data are inadequate for larger diameters.[6] Randomized trials in patients with Marfan syndrome, which mostly involve small numbers of patients with normal aortic diameters, have suggested that several drugs (e.g., losartan) can reduce subsequent aortic root enlargement, but no differences have yet been demonstrated in the progression to aneurysm or in clinical outcomes.

PREVENTION

Ultrasonography is the safest and most practical screening method for abdominal aortic aneurysms. The U.S. Preventive Services Task Force recommends one-time ultrasound screening for men age 65 to 75 years who have ever smoked.[7] No medical therapy has been shown to reduce abdominal aortic aneurysm enlargement, but smoking cessation is important both for prevention and to slow progression.

PROGNOSIS

Most aneurysms expand over time, and the risk of rupture increases with diameter. Abdominal aortic aneurysms with diameters less than 5.5 cm have an annual rupture risk of 1% or less. Rupture rates are unknown for good operative candidates with larger aneurysms but are 10% or more per year in poor surgical candidates with large abdominal aortic aneurysms. For thoracic aortic aneurysms, a population-based study reported a 5-year rupture risk of 0% for aneurysms less than 4.0 cm in diameter, 16% for aneurysms 4.0 to 6.0 cm in diameter, and 31% for aneurysms larger than 6.0 cm in diameter.

The overall mortality rate in patients with a ruptured aortic aneurysm remains about 75%, including about a 40% mortality rate even among those who are able to undergo emergent intervention.[8] Lifelong imaging surveillance is recommended after endovascular aneurysm repair. Otherwise, the prognosis after aneurysm repair is determined largely by risk factors and comorbidities rather than the aneurysm itself. Observational data suggest that survival after repair may be improved by treatment with statins,[9] probably because it reduces the risk of death from concurrent coronary artery disease.

INTRAMURAL AORTIC HEMATOMA AND AORTIC DISSECTION

DEFINITION

An intramural aortic hematoma develops when blood accumulates within the aortic media because of either bleeding from the vasa vasorum or a tear in the intima. An aortic dissection occurs when the media of the artery becomes longitudinally cleaved, thereby forming a false lumen that communicates with the true lumen.

About two thirds of aortic dissections are classified as type A (involving the ascending aorta), and the other third are classified as type B (not involving the ascending aorta) (Fig. 78-3). Hematomas and dissections are classified as acute if they developed within the prior 2 weeks and chronic thereafter. Hematomas and dissections that are acute and involve the ascending aorta are more likely to rupture and cause severe disability or death.[10]

Aortic dissection causes about 3000 deaths per year in the United States. In patients without Marfan syndrome, the peak incidence of aortic dissection is in individuals between 60 and 80 years of age, with men affected 1.5 times more frequently than women. Hypertension and especially uncontrolled hypertension is the dominant risk factor, although patients with a bicuspid aortic valve are also at increased risk. Some patients are known to have preexisting thoracic aortic aneurysms. Aortic dissections are a rare complication in young women during the peripartum period (Chapter 239). Intraaortic catheterization procedures and cardiac surgery are iatrogenic causes of aortic dissection.

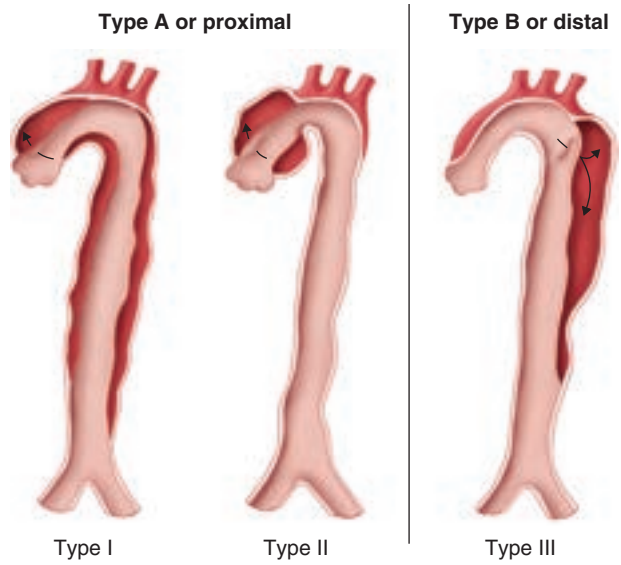

FIGURE 78-3. Classification systems for aortic dissection. (From Isselbacher EM. Diseases of the aorta. In: Braunwald E, Zipes DP, Libby P, Bonow RO, eds. *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine.* 7th ed. Philadelphia: Saunders; 2004:1416.)

PATHOBIOLOGY

The most common predisposing factor for aortic dissection is degeneration of the collagen and elastin in the aortic media. Classic cystic medial degeneration in patients with Marfan syndrome explains the particularly high risk for aortic dissection at a relatively young age.

Aortic dissection typically begins either when a tear in the aortic intima exposes the diseased medial layer to the systemic pressure of intraluminal blood or when a leaking vasa vasorum creates an intramural hematoma. This hematoma may remain relatively localized, or, alternatively may propagate longitudinally along a variable length of the aorta and rupture through the intima and into the aortic lumen. If such communication occurs, the result of an initially intramural hematoma becomes no different than a dissection that began with an intimal tear. With dissection, the media cleaves longitudinally into two layers, thereby producing a blood-filled false lumen that propagates, usually distally but sometimes retrograde, within the aortic wall for a variable distance from the site of the intimal tear. The abdominal aorta rarely dissects except as an extension of a dissecting thoracic aorta.

CLINICAL MANIFESTATIONS

Pain, which is the most common initial symptom, occurs in 96% of cases of both aortic intramural hematoma and dissection. The pain is typically severe and is usually felt in the chest or back, although it may involve the abdomen. Pain commonly begins suddenly and is worst at the start. It is often described as sharp or ripping. An isolated hematoma rarely causes symptoms other than pain. A dissection, however, can cause signs and symptoms related to its propagation, such as acute aortic insufficiency, right coronary artery occlusion, hemopericardium, cerebrovascular accident, mesenteric ischemia, or ischemic peripheral neuropathy. Syncope (Chapter 62) is associated with proximal dissection and worse outcomes.

True hypotension, which is present in more than 25% of patients, augurs a poor prognosis. Pseudohypotension occurs when measured upper extremity blood pressure is falsely low because the subclavian artery is involved in the dissection. Pulse delays or deficits, which help identify which blood vessels are compromised by the dissection, are evident on physical examination in patients whose dissections involve the subclavian, carotid, or femoral arteries.

More than one third of patients with a proximal aortic dissection develop acute aortic valve insufficiency. In some patients with acute aortic insufficiency, the murmur may be undetectable or unimpressive until the finding of a widened pulse pressure leads to more careful examinations. When aortic dissection compromises a coronary artery, especially the right coronary artery, myocardial ischemia or infarction (Chapter 73) develops. Retrograde dissection can also cause acute and even fatal hemopericardium and tamponade (Chapter 71). Whereas compromise of flow through the brachiocephalic arteries may produce stroke (Chapter 407) or coma, involvement of the

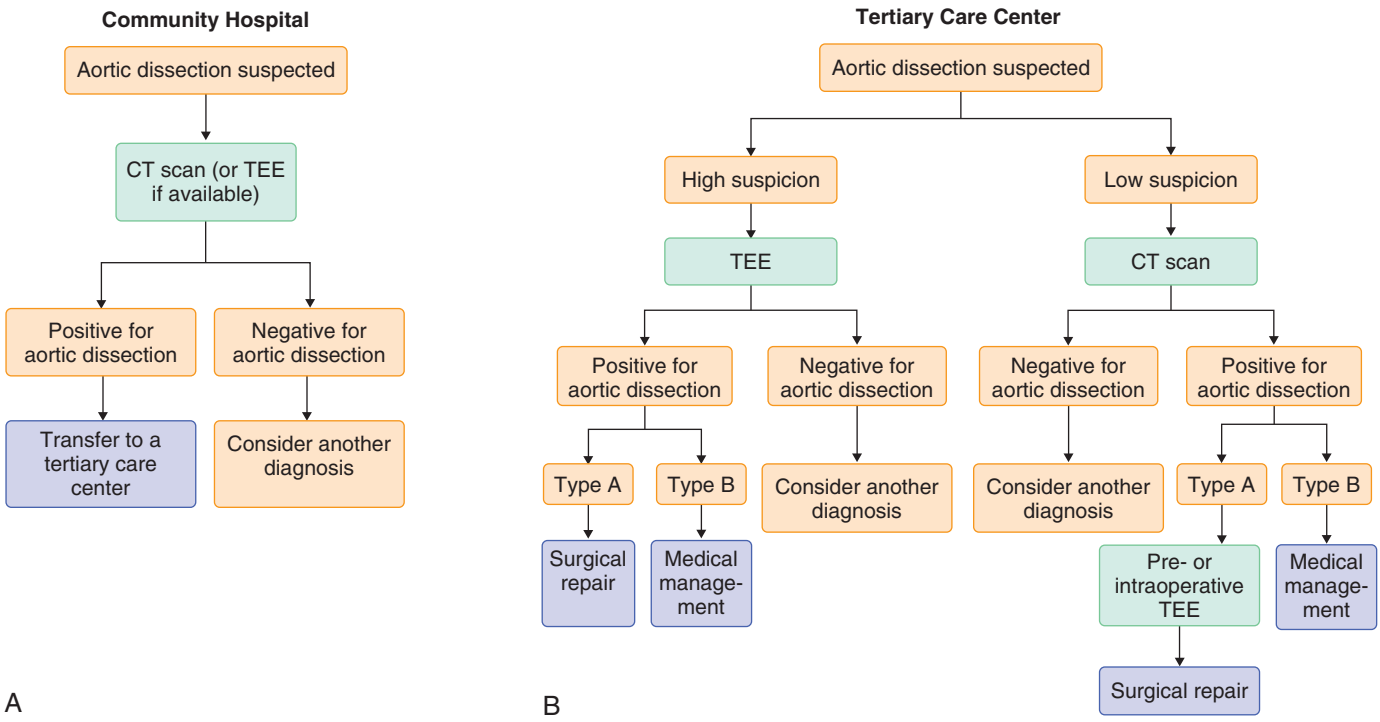

FIGURE 78-4. Algorithms for the evaluation of suspected acute aortic dissection. A, This approach is used in many community hospitals where cardiac surgery is not performed. **B,** This approach is used in many tertiary care centers where transesophageal echocardiography (TEE) and cardiac surgery are available. CT = computed tomography. (Courtesy of Eric M. Isselbacher, MD.)

spinal arteries may produce paraplegia. When a dissection extends into the abdominal aorta, compromised flow to one or both renal arteries may result in acute renal failure (Chapter 120). Mesenteric ischemia or infarction (Chapter 143) may manifest as abdominal pain and bloody diarrhea. Dissection distal to the aortic bifurcation can compromise or occlude one or both common iliac arteries, with a resulting femoral pulse deficit and lower extremity ischemia (Chapter 79).

DIAGNOSIS

Aortic dissection may be suspected based on a typical history or when pulse deficits are noted on physical examination. Often, however, an enlarged mediastinal silhouette on the chest radiograph of a patient who is being evaluated for chest pain (Chapter 51) and possible myocardial ischemia raises concerns even though mediastinal widening may be seen in only two thirds of patients in whom dissection is ultimately diagnosed. If the descending thoracic aorta is involved, a left pleural effusion is commonly seen, sometimes because of frank blood but often because of a small exudate from the inflamed aortic wall. Electrocardiographic findings in aortic dissection are nonspecific. Blood tests are not very helpful, but an acute aortic dissection is unlikely if the plasma D-dimer level is below 500 ng/mL.

When the clinical suspicion of aortic dissection is high, the diagnosis must be confirmed or excluded emergently with an imaging study (Fig. 78-4). Transesophageal echocardiography (Fig. 78-5) is the most rapid means of providing sufficient detail to proceed directly to the operating room. CT (Fig. 78-6) provides complementary information. Magnetic resonance (MR) imaging provides even better anatomical detail, but it is not as easy to obtain emergently, and it can be difficult to provide needed acute care in an MR suite. On cross-sectional imaging, an isolated intramural hematoma appears as a crescentic thickening around the aortic wall rather than true and false lumens separated by an intimal flap.

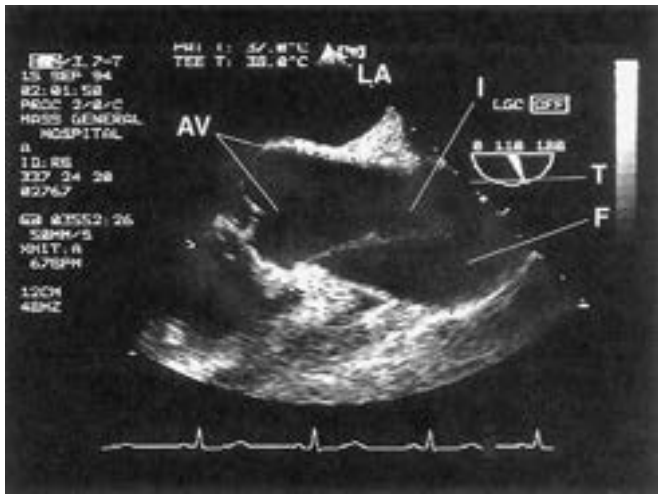

FIGURE 78-5. Transesophageal echocardiogram of the ascending aorta in the long axis in a patient with type A aortic dissection. The aortic valve (AV) is on the *left,* and the ascending aorta extends to the *right*. Within the aorta is an intimal flap (I) that originates at the level of the sinotubular junction. The true (T) and the false (F) lumens are separated by the intimal flap. LA = left atrium. (From Isselbacher EM. Diseases of the aorta. In: Braunwald E, Zipes DP, Libby P, Bonow RO, eds. *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*. 7th ed. Philadelphia: Saunders; 2004:1423.)

TREATMENT

Rx

Whenever aortic intramural hematoma or dissection is suspected, therapy should be instituted immediately even while imaging studies are ordered rather than waiting until the diagnosis is confirmed. The goal of initial medical therapy, which is to halt further progression and reduce the risk for rupture, should be the same for an isolated hematoma as for true dissection because of the risk that the hematoma will propagate.[11]

Despite the absence of randomized trials, current guidelines recommend emergent medical therapy directed at lowering the blood pressure, usually to below 120 mm Hg, and heart rate, usually to below 60 beats/min, while maintaining perfusion to the brain, heart, kidneys, and any other organs whose arterial supply may be jeopardized by dissection.[12] The most common option is intravenous (IV) labetalol (a combined α- and β-blocker, initially at 20 mg administered over a 2-minute period followed by additional doses of 20 to 80 mg every 10 to 15 minutes, up to a maximum total dose of 300 mg, and then a continuous infusion at 2 to 8 mg/min). An alternative is a pure β-blocker (e.g., IV propranolol at 1-mg boluses every 3 to 5 minutes to start followed by a continuous infusion at rates up to 20 mg/hr) combined with IV nitroprusside

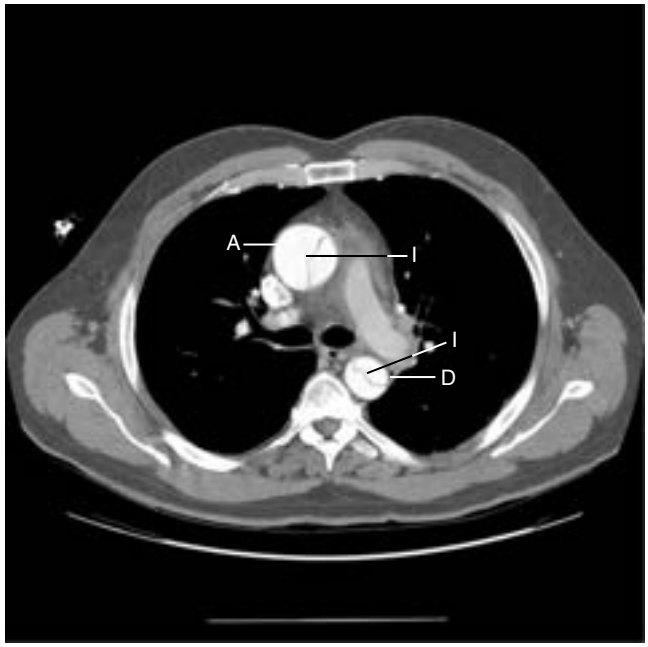

FIGURE 78-6. Aortic dissection. Contrast-enhanced computed tomography scan of the chest at the level of the pulmonary artery shows an intimal flap (I) separating the two lumens of the ascending (A) and descending (D) thoracic aorta in a type A aortic dissection. (Courtesy of Eric M. Isselbacher, MD.)

(0.5 to 8 μg/kg/min) to titrate blood pressure minute by minute as needed. If β-blockers are contraindicated, calcium channel blockers (e.g., IV diltiazem with an initial bolus of 20 mg over a 2-minute period followed by a continuous infusion of 5 to 15 mg/hr) may be useful.

For acute type A dissection, urgent surgical repair is recommended to reduce the risk of life-threatening complications such as rupture, cardiac tamponade, severe aortic insufficiency, or stroke. When patients have significant hypotension, pseudohypotension should be excluded. True hypotension may be caused by acute myocardial infarction (Chapter 73) owing to a compromise of the right coronary artery or by hemopericardium and cardiac tamponade (Chapter 77) as a result of rupture of the dissection into the pericardium. Patients with tamponade should be treated with volume expansion and taken to surgery as quickly as possible because early death is extremely high; pericardiocentesis should be performed only as a last resort because it may precipitate hemodynamic collapse and death. By comparison, patients with chronic type A dissections can often be managed medically because they have already survived the early period of high mortality associated with acute proximal dissections.

Patients with acute type B dissection are at much lower risk for life-threatening complications and are usually managed with medical therapy[13] because the addition of routine endovascular repair to medical treatment has not improved survival in small randomized trials of acute or chronic type B dissection. If, however, a type B dissection is associated with a serious complication, such as end-organ ischemia, intervention is indicated with either endovascular techniques or open surgery.

For an isolated intramural hematoma, the likelihood of progressive dissection or other complications is lower than in patients who initially have dissection, especially if the hematoma is small and aortic dimensions are normal.[14] In one series of East Asian patients, most intramural hematomas regressed, and surgery usually was not necessary. In a Western series, an estimated 10% of hematomas regressed spontaneously, and 25% to 50% progressed during the follow-up period. Hematomas in the ascending aorta have the highest risk, and surgery is often recommended for them. For distal intramural hematomas, management is generally the same as for distal dissection, with surgery reserved for patients with progressive disease (Fig. 78-7).

Before discharge, medically managed patients should be started on a regimen that will control hypertension and reduce ventricular contractility. Recommendations emphasize β-blockers (e.g., metoprolol 25 to 200 mg twice daily or atenolol 25 to 200 mg/day) as the drugs of choice in this setting, but a second (e.g., lisinopril 5 to 40 mg/day) or third (amlodipine 2.5 to 10 mg/day or hydrochlorothiazide 12.5 to 50 mg/day) agent usually is required to achieve the goal of systolic blood pressure less than 120 mm Hg (Chapter 67), although little evidence supports treatment goals different from those used for the general population.

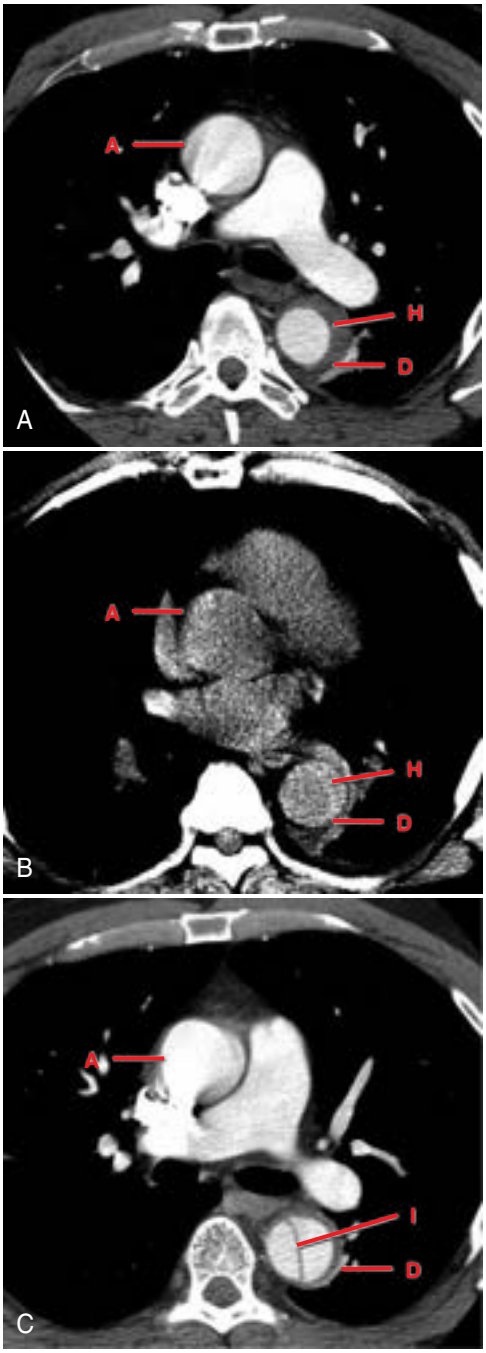

FIGURE 78-7. Intramural aortic hematoma. A, Contrast-enhanced computed tomography (CT) scan of the chest at the level of the pulmonary artery demonstrates an intramural hematoma (H) of the descending thoracic aorta (D). The hematoma appears as a crescentic thickening of the aortic wall that does not enhance from the contrast within the aortic lumen. The ascending thoracic aorta (A) is unaffected. **B,** Corresponding image from a non–contrast-enhanced CT scan in which the intramural hematoma appears as a bright crescentic thickening of the aortic wall because the density of the hematoma is greater than that of the blood within the aortic lumen. **C,** Corresponding image from a contrast-enhanced CT scan performed 1 week later for surveillance in which the intramural hematoma has evolved into a classic aortic dissection with an intimal flap (I) and contrast evident within a patent false lumen. (Courtesy of Eric M. Isselbacher, MD.)

PROGNOSIS

The acute mortality rate from untreated type A dissection is about 1% per hour, and most deaths with either type occur within 7 days of the onset of symptoms.[15] Patients are at highest risk for complications during the first 2 years, but patients who survive the initial hospitalization generally do well thereafter, regardless of whether they were treated medically or surgically. However, continued blood pressure control is critical to reduce the long-term risk of complications such as aortic insufficiency, recurrent dissection, aneurysm formation, and aneurysm rupture. Some progressive aortic expansion

typically occurs without symptoms, so medically treated patients must be observed closely with serial aortic imaging at 6-month intervals for the first 2 years and annually thereafter, provided that their anatomy is stable.

TAKAYASU ARTERITIS

Takayasu arteritis is a chronic inflammatory disease of unknown cause. The median age at onset is 29 years, and it is eight times more frequent in women than in men. It occurs more often in Asia and Africa than in Europe or North America. The early stage is characterized by active inflammation involving the aorta and its branches. The disease often progresses at a variable rate to a sclerotic stage with intimal hyperplasia, medial degeneration, and obliterative changes. Most of the resulting arterial lesions are stenotic, but aneurysms may also occur.

Takayasu arteritis involves the aorta and its branches, and it occasionally involves the pulmonary artery as well. The disease tends to be most pronounced at branch points in the aorta, especially the aortic arch, brachiocephalic vessels, and abdominal aorta. Takayasu arteritis can be diffuse or patchy, and affected areas may be separated by variable lengths of normal aorta.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Most patients initially develop symptoms of a systemic inflammatory process, including fatigue, headache, fever, night sweats, arthralgia, and weight loss.[16] By the time of diagnosis, however, 90% of patients have symptoms of vascular insufficiency, typically with claudication in the upper extremities and sometimes in the lower extremities.

Physical examination commonly reveals absent pulses and diminished blood pressure in the upper extremities, consistent with why Takayasu arteritis has sometimes been called *pulseless disease*. Bruits may be heard in over the affected arteries. More than 50% of patients have significant hypertension because of renal artery involvement, but hypertension may be difficult to diagnose on routine examination because of the diminished pulses. Proximal aortic involvement can cause aortic valve insufficiency. Involvement of the ostia of the coronary arteries may cause angina or myocardial infarction, and patients may develop heart failure owing to myocardial infarction, hypertension, or aortic insufficiency. Carotid artery involvement may cause cerebral ischemia or stroke. Abdominal angina may result from compromise of the mesenteric circulation.

Laboratory abnormalities during the acute phase include elevated C-reactive protein (CRP) level, an elevated erythrocyte sedimentation rate (ESR), anemia, mild leukocytosis, and elevated immunoglobulin levels. The diagnosis is best made by aortography, CT angiography, or MR angiography, which reveal stenosis of the aorta and stenosis or occlusion of its branch vessels, often with poststenotic dilation or associated aneurysms.

TREATMENT AND PROGNOSIS

Corticosteroids (e.g., prednisone 60 to 100 mg/day, often continued for months and tapered only when symptoms or evidence of inflammation subside) are the primary therapy for the acute inflammatory stage and may be effective in improving constitutional symptoms, lowering the ESR, and slowing the progression of the disease. Additional immunosuppressive medications that are used when corticosteroid therapy alone is ineffective include methotrexate (15 to 25 mg/wk). It is unknown whether medical therapy reduces the risk for major complications or prolongs life.

Percutaneous balloon angioplasty can effectively dilate short stenotic lesions of the aorta and its branch arteries, although restenosis is common. Surgery may be necessary to bypass or reconstruct key segments, such as the coronary, carotid, or renal arteries, or to treat aortic insufficiency. Ideally, surgery should not be performed during the inflammatory phase.

The overall 15-year survival rate in patients diagnosed with Takayasu arteritis is about 85%, about 65% in patients with major complications, and about 95% in patients without major complications. Most deaths result from stroke, myocardial infarction, or heart failure.

GIANT CELL ARTERITIS

Giant cell arteritis (Chapter 271) typically affects medium-sized arteries, but in 15% of cases, it involves the aorta and branches of the aortic arch.[17] Narrowing of the aorta is rare, but weakening of the wall of the ascending aorta may lead to localized thoracic aortic aneurysms and secondary aortic valve insufficiency. If branches of the aortic arch are narrowed, symptoms will be similar to those seen in Takayasu arteritis. The arteriographic lesions of the aorta and its primary branches are generally similar to Takayasu arteritis, thereby suggesting that Takayasu arteritis and giant cell arteritis may be part of a spectrum of the same disease. Giant cell arteritis is associated with increased risks for atherosclerotic events, such as myocardial infarction and stroke.

The mean age of onset is about 67 years. Symptoms include headaches, visual disturbances, polymyalgia rheumatic, jaw claudication, and fever (Chapter 271). An elevated ESR is a universal finding; the serum CRP level is typically elevated, and anemia is common. Because the temporal artery is commonly involved, the diagnosis is usually made by temporal artery biopsy, including contralateral biopsy if the first biopsy is negative. In patients with visual symptoms, scheduling of temporal artery biopsy should not delay treatment because of the risk of sudden blindness and because biopsy specimens remain diagnostic for weeks after the onset of therapy. Management is high-dose corticosteroid therapy (e.g., prednisone 60 to 100 mg/day, often for months and tapered only after symptoms or evidence of inflammation subside), to which the disease is usually responsive. Initial use of an IV corticosteroid pulse dose or of methotrexate does not add benefit.

Grade A References

A1. Powell JT, Sweeting MJ, Thompson MM, et al. Endovascular or open repair strategy for ruptured abdominal aortic aneurysm: 30 day outcomes from IMPROVE randomised trial. *BMJ*. 2014;348:f7661.

A2. Powell JT, Brown LC, Forbes JF, et al. Final 12-year follow-up of surgery versus surveillance in the UK Small Aneurysm Trial. *Br J Surg*. 2007;94:702-708.

A3. Lederle FA, Wilson SE, Johnson GR, et al. Immediate repair compared with surveillance of small abdominal aortic aneurysms. *N Engl J Med*. 2002;346:1437-1444.

A4. Cao P, De Rango P, Verzini F, et al. Comparison of surveillance versus aortic endografting for small aneurysm repair (CAESAR): results from a randomised trial. *Eur J Vasc Endovasc Surg*. 2011;41:13-25.

A5. Ouriel K, Clair DG, Kent KC, et al. Endovascular repair compared with surveillance for patients with small abdominal aortic aneurysms. *J Vasc Surg*. 2010;51:1081-1087.

A6. Thompson SG, Ashton HA, Gao L, et al. Final follow-up of the Multicentre Aneurysm Screening Study (MASS) randomized trial of abdominal aortic aneurysm screening. *Br J Surg*. 2012;99:1649-1656.

A7. Takagi H, Niwa M, Mizuno Y, et al. The Last Judgment upon abdominal aortic aneurysm screening. *Int J Cardiol*. 2013;167:2331-2332.

A8. Greenhalgh RM, Brown LC, Powell JT, et al. Endovascular versus open repair of abdominal aortic aneurysm. *N Engl J Med*. 2010;362:1863-1871.

A9. Lederle FA, Freischlag JA, Kyriakides TC, et al. Long-term comparison of endovascular and open repair of abdominal aortic aneurysm. *N Engl J Med*. 2012;367:1988-1997.

A10. Greenhalgh RM, Brown LC, Powell JT, et al. Endovascular repair of aortic aneurysm in patients physically ineligible for open repair. *N Engl J Med*. 2010;362:1872-1880.

A11. Brunkwall J, Kasprzak P, Verhoeven E, et al. Endovascular repair of acute uncomplicated aortic type B dissection promotes aortic remodelling: 1 year results of the ADSORB trial. *Eur J Vasc Endovasc Surg*. 2014;48:285-291.

A12. Nienaber CA, Rousseau H, Eggebrecht H, et al. Randomized comparison of strategies for type B aortic dissection: the INvestigation of STEnt Grafts in Aortic Dissection (INSTEAD) trial. *Circulation*. 2009;120:2519-2528.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

REFERENCES

1. Lederle FA. The rise and fall of abdominal aortic aneurysm. *Circulation*. 2011;124:1097-1099.
2. Sidloff D, Stather P, Dattani N, et al. Aneurysm global epidemiology study: public health measures can further reduce abdominal aortic aneurysm mortality. *Circulation*. 2014;129:747-753.
3. Lederle FA, Johnson GR, Wilson SE, et al. Rupture rate of large abdominal aortic aneurysms in patients refusing or unfit for elective repair. *JAMA*. 2002;287:2968-2972.
4. Bown MJ, Sweeting MJ, Brown LC, et al. Surveillance intervals for small abdominal aortic aneurysms: a meta-analysis. *JAMA*. 2013;309:806-813.
5. Goodney PP, Travis L, Lucas FL, et al. Survival after open versus endovascular thoracic aortic aneurysm repair in an observational study of the Medicare population. *Circulation*. 2011;124:2661-2669.
6. Jondeau G, Detaint D, Tubach F, et al. Aortic event rate in the Marfan population: a cohort study. *Circulation*. 2012;125:226-232.
7. Guirguis-Blake JM, Beil TL, Senger CA, et al. Ultrasonography screening for abdominal aortic aneurysms: a systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014;160:321-329.
8. Karthikesalingam A, Holt PJ, Vidal-Diez A, et al. Mortality from ruptured abdominal aortic aneurysms: clinical lessons from a comparison of outcomes in England and the USA. *Lancet*. 2014;383:963-969.
9. de Bruin JL, Baas AF, Heymans MW, et al. Statin therapy is associated with improved survival after endovascular and open aneurysm repair. *J Vasc Surg*. 2014;59:39-44.
10. Goldfinger JZ, Halperin JL, Marin ML, et al. Thoracic aortic aneurysm and dissection. *J Am Coll Cardiol*. 2014;64:1725-1739.
11. Lederle FA, Powell JT, Nienaber CA. Does intensive medical treatment improve outcomes in aortic dissection? *BMJ*. 2014;349:g5288.
12. Hiratzka LF, Bakris GL, Beckman JA, et al. 2010 ACCF/AHA/AATS/ACR/ASA/SCA/SCAI/SIR/STS/SVM guidelines for the diagnosis and management of patients with Thoracic Aortic Disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, American Association for Thoracic Surgery, American College of Radiology, American Stroke Association, Society of Cardiovascular Anesthesiologists, Society for Cardiovascular Angiography and Interventions, Society of Interventional Radiology, Society of Thoracic Surgeons, and Society for Vascular Medicine. *Circulation*. 2010;121:e266-e369.
13. Fattori R, Cao P, De Rango P, et al. Interdisciplinary expert consensus document on management of type B aortic dissection. *J Am Coll Cardiol*. 2013;61:1661-1678.
14. Alomari IB, Hamirani YS, Madera G, et al. Aortic intramural hematoma and its complications. *Circulation*. 2014;129:711-716.
15. Booher AM, Isselbacher EM, Nienaber CA, et al. The IRAD classification system for characterizing survival after aortic dissection. *Am J Med*. 2013;126:730.
16. Schmidt J, Kermani TA, Bacani AK, et al. Diagnostic features, treatment, and outcomes of Takayasu arteritis in a US cohort of 126 patients. *Mayo Clin Proc*. 2013;88:822-830.
17. Waldman CW, Waldman SD, Waldman RA. Giant cell arteritis. *Med Clin North Am*. 2013;97:329-335.

79

ATHEROSCLEROTIC PERIPHERAL ARTERIAL DISEASE

CHRISTOPHER J. WHITE

DEFINITION

Lower extremity atherosclerotic peripheral arterial disease is one subset of a larger group of peripheral vascular diseases that includes all noncoronary vascular disorders that may affect the arterial, venous (Chapters 80 and 81), or lymphatic circulation. Atherosclerotic arterial diseases are characterized by arterial narrowing or occlusion caused by the accumulation of atherosclerotic plaque elements in the vessel wall (Chapter 70). Atherosclerotic vascular disease can also lead to aneurysm formation, which is the pathologic enlargement of arterial segments, and may result in rupture, dissection, or thromboembolism (Chapter 78).

How to Perform and Calculate the ABI

PARTNERS Program ABI Interpretation
Above 0.90—Normal
0.71-0.90—Mild obstruction
0.41-0.70—Moderate obstruction
0.00-0.40—Severe obstruction

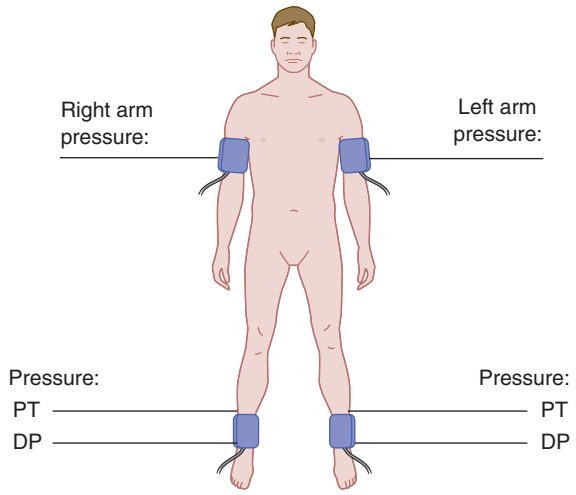

RIGHT ABI **LEFT ABI**

$$\frac{\text{Higher right ankle pressure}}{\text{Higher arm pressure}} = \_\_\_\_\_ \qquad \frac{\text{Higher left ankle pressure}}{\text{Higher arm pressure}} = \_\_\_\_\_$$

EXAMPLE

$$\frac{\text{Higher ankle pressure}}{\text{Higher arm pressure}} = \frac{92 \text{ mm Hg}}{164 \text{ mm Hg}} = 0.56$$ (See ABI chart for interpretation)

FIGURE 79-1. Performing pressure measurements and calculating the ankle-brachial index (ABI). To calculate the ABI, systolic pressures are determined in both arms and both ankles with the use of a handheld Doppler instrument. The higher reading in either the dorsalis pedis (DP) or the posterior tibial (PT) arteries of the right foot and the left foot is divided by the higher systolic blood pressure reading in either the right or left arm to calculate the index. PAD = peripheral arterial disease; PARTNERS = PAD Awareness, Risk, and Treatment: New Resources for Survival.

EPIDEMIOLOGY

The prevalence of lower extremity peripheral arterial disease in the United States, Europe, and Asia continues to increase as the population ages and is exposed to atherosclerotic risk factors (Chapter 52). The presence of peripheral arterial disease is defined as an ankle-brachial index (ABI)—the ratio of the highest systolic blood pressure in the ankle divided by the highest systolic blood pressure in the arm (Fig. 79-1)—less than 0.90. Among individuals age 40 years and older, the prevalence is 4.3% (95% confidence interval [CI], 3.1%-5.5%), but the prevalence in individuals with diabetes ranges from 20% to 30%.[1]

Risk factors for atherosclerosis (Chapter 52) increase the likelihood of developing lower extremity peripheral arterial disease. More than 95% of individuals with peripheral arterial disease have at least one traditional cardiovascular risk factor, and most have multiple risk factors. More than one third of patients with peripheral arterial disease have significant coronary disease, and up to one quarter have carotid artery disease. As a result, the risk of heart attack, stroke, and death are increased severalfold in patients with peripheral arterial disease.

Among conventional atherosclerotic risk factors, cigarette smoking is two to three times more likely to cause lower extremity peripheral arterial disease than to cause coronary artery disease.

Hypertension is also associated with lower extremity peripheral arterial disease. The development of peripheral arterial disease is more likely in patients with lipid abnormalities (elevated total and low-density lipoprotein [LDL] cholesterol, decreased high-density lipoprotein [HDL] cholesterol, and hypertriglyceridemia), and the risk increases by 5% to 10% for each 10–mg/dL increase in total cholesterol. Increased homocysteine levels are associated with a two- to threefold increased risk for developing atherosclerotic peripheral arterial disease. Diabetes mellitus increases the risk of lower extremity peripheral arterial disease by two- to fourfold, and the risk of developing lower extremity peripheral arterial disease is proportional to the severity and duration of diabetes. Tight control of diabetes is important because the risk of developing peripheral arterial disease increases by 28% for every 1% increase in glycosylated hemoglobin (Chapter 229). Patients with diabetes who have lower extremity peripheral arterial disease are seven- to 15-fold more likely to undergo a major amputation than nondiabetics with lower extremity peripheral arterial disease.

Peripheral arterial disease disproportionately affects older individuals (Fig. 79-2), non-Hispanic blacks, current smokers, individuals with diabetes, and those with abnormal renal function. The overall prevalence of peripheral arterial disease in the United States among persons age 70 years and older is 14.5% (95% CI, 10.8%-18.2%), which corresponds to approximately 4 million individuals.

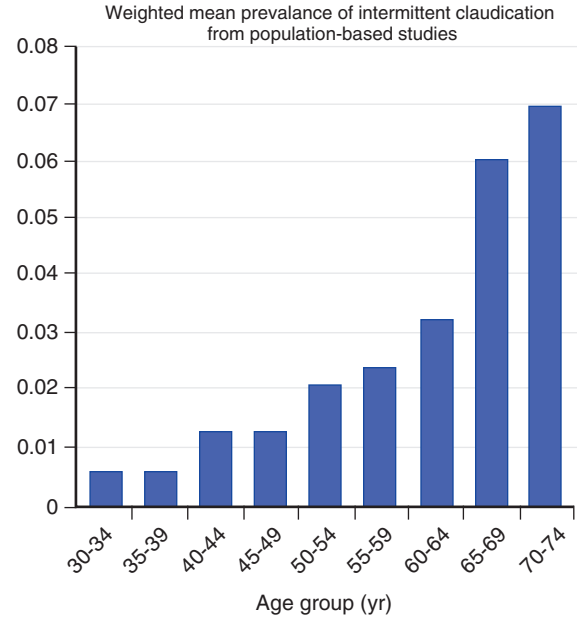

FIGURE 79-2. Weighted mean prevalence of intermittent claudication. (Modified from Dormandy JA, Rutherford RB. Management of peripheral arterial disease. TransAtlantic Inter-Society Consensus (TASC) Working Group. *J Vasc Surg.* 2000;31(suppl):S1-S296.)

PATHOBIOLOGY

Acute Limb Ischemia

Acute limb ischemia occurs when blood flow to an extremity is abruptly halted or markedly diminished, resulting in hypoperfusion that threatens the viability of the limb. Acute limb ischemia is most commonly caused by either thrombosis or embolism. Most emboli originate from the heart as a mural

TABLE 79-1 CLINICAL CATEGORIES OF ACUTE LIMB ISCHEMIA

CATEGORY	DESCRIPTION	SENSORY LOSS	MUSCLE WEAKNESS	ARTERIAL DOPPLER	VENOUS DOPPLER
I. Viable	Not immediately threatened	None	None	Audible	Audible
IIa. Threatened marginally	Salvageable if promptly treated	Minimal or none	None	Inaudible	Audible
IIb. Threatened immediately	Salvageable if immediately treated	More than toes, associated with rest pain	Mild to moderate	Inaudible	Audible
III. Irreversible	Major tissue loss inevitable	Profound, anesthetic	Profound, paralysis, rigor	Inaudible	Inaudible

Adapted from Rutherford RB, Baker JD, Ernst C, et al. Recommended standards for reports dealing with lower extremity ischemia: revised version. *J Vasc Surg.* 1997;26:517-538.

thrombus from a recent myocardial infarction (MI) (Chapter 73) or from the atrial appendage in a patient with atrial fibrillation (Chapter 64). A less common cause of lower extremity emboli is an abdominal aortic aneurysm (Chapter 78) that serves as the source of cholesterol emboli (Chapter 80).

Arterial in situ thrombosis as a result of plaque rupture usually represents the final stage of a chronically diseased artery, most commonly the femoral or popliteal artery. If native artery thrombosis occurs in the absence of a preexisting stenosis, a thorough search for a hypercoagulable state should be undertaken. Thrombosis of a popliteal aneurysm may present as acute limb ischemia.

Chronic Limb Ischemia

Lower extremity peripheral artery disease may be manifest as either chronic stable disease or as critical limb ischemia. Peripheral arterial disease is most commonly caused by atherosclerosis, but it also may be caused by thromboembolism, inflammatory disease, trauma, aneurysmal disease, adventitial cysts, entrapment syndromes, or congenital abnormalities. Aneurysms may be associated with atherosclerosis, or they may be caused by underlying hereditary (familial) or acquired (e.g., smoking or trauma) causes.

Patients with critical limb ischemia have inadequate blood flow to sustain viability in the distal tissue bed. Critical limb ischemia is most often caused by atherosclerosis, but it can also be caused by atheroembolic or thromboembolic disease, vasculitis, in situ thrombosis related to hypercoagulable states, thromboangiitis obliterans, cystic adventitial disease, popliteal entrapment, or trauma. Patients presenting with critical limb ischemia typically have multisegment disease along the length of the limb.

Inflammation plays a fundamental role in the development and progression of atherosclerosis (Chapter 70). Elevated levels of C-reactive protein (CRP) are strongly associated with the development of peripheral arterial disease. Markers of inflammation such as interleukin-6, tumor necrosis factor-α, CRP, and platelet activation are increased compared with normal subjects.

No specific genetic markers have been confirmed for peripheral arterial disease, although one study identified a linkage on chromosome 1p. The proportion of low ABIs attributable to heritability is estimated at 20%. Concordance rates among twins are about 33% for monozygotic pairs and about 31% for dizygotic pairs, suggesting a limited role for heritability. Taken together, these data suggest a modest but significant heritability factor for peripheral arterial disease.

CLINICAL MANIFESTATIONS

Acute Limb Ischemia

A patient with acute limb ischemia presents with a cool, painful extremity (Table 79-1).[2] Typically, the major muscle groups below the level of obstruction are symptomatic. The absence of a pulse may help localize the site of occlusion, but pulses may be normal in cases of microemboli or cholesterol emboli (Chapter 80). Venous and capillary filling is an indicator of the severity of acute limb ischemia. The leg should be carefully examined for color and temperature abnormalities. Pallor is seen early on, but with time, cyanosis is common. Poikilothermia, or coolness, is an important finding, particularly if the opposite limb is warm. A transition level for color and temperature changes, which is often clinically obvious, should be correlated with the pulses and denoted as a baseline reference at the initial examination for comparison with subsequent examinations.

Sensory changes include numbness and paresthesias. Paralysis indicates that ischemia had advanced to jeopardize the survival of the limb unless the patient undergoes urgent revascularization. Any motor deficit indicates tissue anoxia and implies a very poor prognosis. Motor deficits progress from distal to more proximal muscle groups, so early motor weakness is seen in the intrinsic foot muscles. Complete motor paralysis is a late symptom that suggests irreversible injury. With irreversible ischemia, paralysis progresses to rigor.

TABLE 79-2 FONTAINE'S AND RUTHERFORD'S CLINICAL CLASSIFICATIONS OF CHRONIC LOWER LIMB ISCHEMIA

FONTAINE		RUTHERFORD		
STAGE	CLINICAL	GRADE	CATEGORY	CLINICAL
I	Asymptomatic	0	0	Asymptomatic
IIa	Mild claudication	I	1	Mild claudication
IIb	Moderate to severe claudication	I	2	Moderate claudication
		I	3	Severe claudication
III	Rest pain	II	4	Rest pain
IV	Ulceration or gangrene	III	5	Minor tissue loss
		IV	6	Ulceration or gangrene

Adapted from Norgren L, Hiatt WR, Dormandy JA, et al. Inter-society consensus for the management of peripheral arterial disease (TASC II). *Eur J Vasc Endovasc Surg.* 2007;33(suppl):S1-S75.

Chronic Stable Lower Limb Ischemia

Among patients with chronic stable lower extremity peripheral arterial disease, 50% of patients are asymptomatic despite abnormal pulse examinations, the presence of a vascular bruit, or an abnormal ABI (see Fig. 79-1). About 40% of patients have atypical symptoms (e.g., leg tiredness or fatigue), and only about 10% of patients have classic symptoms of intermittent claudication.

Claudication is defined as exertional discomfort, relieved with rest, in specific muscle groups at risk for ischemia during exercise (Table 79-2). Symptoms usually begin one segment below the level of the arterial narrowing. For example, whereas vascular obstructions (occlusions or stenoses) of the iliac vessels typically cause hip, thigh, and calf pain, femoral and popliteal artery obstructions typically cause symptoms in the calf and foot muscles. Claudication, which is a specific vascular syndrome, must be distinguished from other conditions that cause exertional leg pain, which have been termed *pseudoclaudication* (Table 79-3).

Symptoms in individual patients are remarkably variable despite similar degrees of vascular stenosis, in part owing to collateral vessel formation. A patient with superficial femoral artery occlusion but robust collateral formation via the deep femoral artery and geniculate collaterals, which supply blood to the infrapopliteal vessels, may have minimal or no symptoms. Another patient with similar anatomy but poor collaterals may have severe functional limitation.

Chronic Critical Lower Limb Ischemia

Critical limb ischemia, which develops in about 10% of all patients with peripheral arterial disease, presents as resting limb pain, nonhealing lower extremity ulcers, or gangrene (see Table 79-2). These patients' limbs are in jeopardy, and even the most minor trauma from a poorly fitting shoe or a carelessly clipped toenail may cause a nonhealing wound or infection that leads to amputation. Critical limb ischemia can be exacerbated by conditions that reduce blood flow to the microvascular bed, such as diabetes; severe low cardiac output states; and, rarely, vasospastic diseases.

TABLE 79-3 DIFFERENTIATION OF TRUE CLAUDICATION FROM PSEUDOCLAUDICATION

	INTERMITTENT CLAUDICATION	SPINAL STENOSIS	ARTHRITIS	VENOUS CONGESTION	COMPARTMENT SYNDROME
Character of discomfort	Cramping, tightness, or tiredness	Same as claudication or tingling, weakness, clumsiness	Aching	Tightness, bursting pain	Tightness, bursting pain
Location of discomfort	Buttock, hip, thigh, calf, foot	Buttock, hip, thigh	Hip, knee	Groin, thigh	Calf
Exercise-induced discomfort	Yes	Variable	Variable	After walking	Excessive exercise
Walking distance to discomfort	Reproducible	Variable	Variable	Variable	Excessive exercise
Occurs with standing	No	Yes	Yes, but positional	Yes, but positional	Yes, but positional
Relief of discomfort	Rapid relief with rest	Relief with sitting or changing position	Slow relief with avoidance of weight bearing	Slow relief with leg elevation	Slow relief with leg elevation
Other	Associated with atherosclerosis and decreased pulses	History of lower back problems	Discomfort at joint	History of deep vein thrombosis, signs of venous congestion	Typical in athletes

From White C. Intermittent claudication. *N Engl J Med.* 2007;356:1241-1250.

DIAGNOSIS

Acute Limb Ischemia

In patients with acute limb ischemia, the history and physical examination are the most important steps, not only for assessing the cause and severity of ischemia but also for determining the diagnostic and therapeutic path. Upon completion of the history and physical examination, the physician should be able to answer the following questions about the severity of acute limb ischemia: Is the limb viable? Is the limb's viability immediately threatened? Are there already irreversible changes that may preclude salvage of the limb? Three findings that help differentiate "threatened" from "viable" extremities are the presence of persistent pain, sensory loss, and muscle weakness.

Chronic Limb Ischemia

The clinical severity of chronic stable and chronic critical limb ischemia can be semiquantitatively assessed using either the Fontaine or the Rutherford classification (see Table 79-2). The clinician must distinguish intermittent claudication from nonvascular causes that may mimic claudication (i.e., pseudoclaudication), such as neurogenic pain from spinal stenosis or nerve root compression (Chapter 400), musculoskeletal or arthritic pain, or discomfort from venous congestion or a compartment syndrome (see Table 79-3). A typical history of claudication has a low sensitivity but a high specificity for peripheral arterial disease.

Patients presenting with peripheral arterial disease should be assessed for atherosclerotic risk factors (Table 79-4) and undergo a complete vascular physical examination with their shoes and socks removed, including measurement of the ABI (see Fig. 79-1).[3] ABI results should be uniformly reported with noncompressible values defined as greater than 1.40, normal values defined as 1.00 to 1.40, borderline values as 0.91 to 0.99, and abnormal values as 0.90 or less.

Imaging

Duplex imaging combines ultrasound imaging and Doppler blood velocity measurements to localize vascular obstructions and estimate lesion severity. The sensitivity and specificity of duplex ultrasonography for the diagnosis of a 50% or greater stenosis in the lower extremity is 90% or higher.[4]

Computed tomography angiography (CTA) and magnetic resonance angiography (MRA) offer cross-sectional images that can be reconstructed into a three-dimensional angiogram (Fig. 79-3). Whereas CTA requires iodinated intravenous contrast material and ionizing radiation, MRA uses gadolinium contrast and does not expose the patient to ionizing radiation but cannot be used in patients with ferromagnetic metallic implants such as pacemakers or defibrillators. The major toxicity of gadolinium is an uncommon but potentially lethal systemic disorder called nephrogenic systemic fibrosis or nephrogenic sclerosing dermopathy (Chapter 267); a glomerular filtration rate of 60 mL/minute or less is the major risk factor.

An advantage of CTA over MRA is its ability to visualize metallic stents and stent grafts (Fig. 79-4). CTA requires only about one fourth the radiation required for invasive digital angiography, and it can be performed more quickly and with less pretreatment planning than MRA. Vascular calcification may be a source of artifact with CTA but is inconsequential with MRA. MRA tends to overestimate lesions at the ostia of arteries owing to turbulent flow.

TABLE 79-4 INITIAL LABORATORY EVALUATION OF A PATIENT WITH PERIPHERAL ARTERIAL DISEASE

Serum electrolytes, including fasting serum glucose
Renal function (serum creatinine, blood urea nitrogen, estimated glomerular filtration rate)
Complete blood count
Fasting lipid profile
High-sensitivity C-reactive protein
Ankle-brachial index

Adapted from Rutherford RB, Baker JD, Ernst C, et al. Recommended standards for reports dealing with lower extremity ischemia: revised version. *J Vasc Surg.* 1997;26:517-538.

In a randomized trial comparing MRA with CTA for initial imaging in peripheral arterial disease, there was no difference between the two techniques in terms of ease, clinical utility, or patient outcome, but CTA reduced total diagnostic costs.

Invasive digital angiography remains the "gold standard" for the diagnosis and evaluation of peripheral arterial disease (see Fig. 79-3) despite the need for iodinated contrast material and the exposure to ionizing radiation. Invasive angiographic procedures (Chapter 57) are associated with a relatively small but nontrivial rate of complications, including severe contrast allergy in 0.1%, access-related bleeding complications, and contrast-induced nephropathy (Chapter 57).

TREATMENT Rx

The treatment of patients with atherosclerotic lower extremity peripheral arterial disease is directed at reducing the patient's risk for life-threatening cardiovascular complications of atherosclerosis, improving walking distance, and salvaging the limb. Revascularization strategies have shifted from open surgical approaches to percutaneous, catheter-based endovascular treatments because of the relative safety, success, and durability of stenting. As a result, amputation rates for peripheral arterial disease have fallen by more than 25% in the past decade or so.

Acute Limb Ischemia

Patients with irreversible limb ischemia (nonviable limb) should not undergo angiography and should be scheduled for amputation (Fig. 79-5). All other patients with acute limb ischemia should undergo emergent angiography to define the culprit lesion and to treat it endovascularly if possible. Before angiography, anticoagulation to prevent thrombus propagation or embolization (e.g., unfractionated heparin 5000 IU bolus; then 1000-IU/hr infusion) and analgesia (e.g., morphine 2 mg intravenously, as needed) should be initiated, and any underlying medical comorbidities (e.g., heart failure) should be managed aggressively to stabilize the patient.

The options for prompt revascularization include (1) endovascular therapies, with intraarterial thrombolysis (e.g., recombinant tissue plasminogen activator 0.5 mg/hr intraarterially) and/or catheter-based thrombectomy and stenting or (2) open surgical thrombectomy with or without arterial bypass. In general, endovascular therapy is recommended for patients with an onset of acute symptoms less than 14 days earlier, and surgical revascularization is

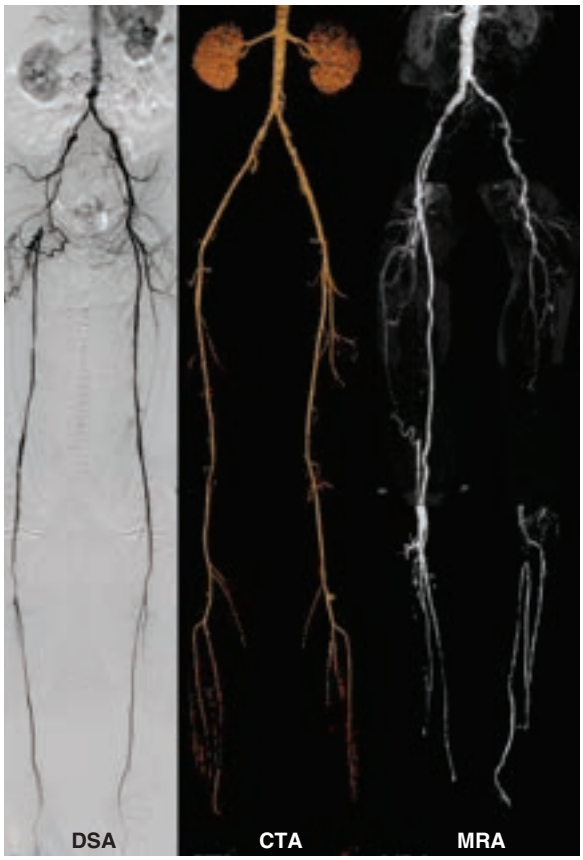

FIGURE 79-3. **Aortography with distal run-off in three different patients by three different methods: digital subtraction angiography (DSA), computed tomography angiography (CTA), and magnetic resonance angiography (MRA). (From White C. Intermittent claudication. *N Engl J Med.* 2007;356:1241-1250.)**

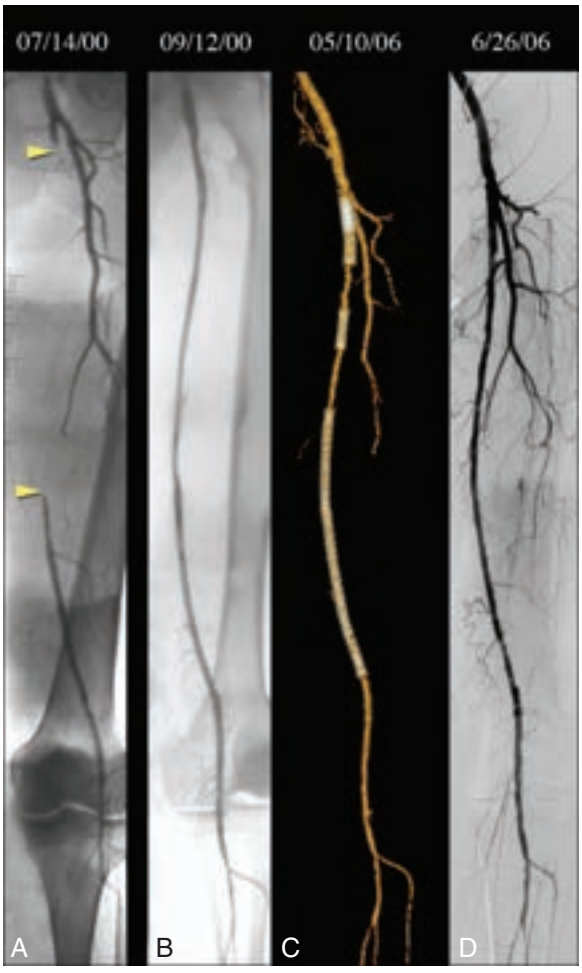

FIGURE 79-4. **A, Baseline angiogram of left leg showing occlusion *(arrowheads)* of the femoral-popliteal segment. B, Post-treatment angiogram after balloon angioplasty and stent placement. C, More than 5 years later, the patient returns with claudication and a reduced ankle-brachial index. Follow-up computed tomography angiogram shows narrowing of the superficial femoral artery between the two stents. D, Final angiogram following balloon angioplasty and stent placement.**

recommended for patients with a duration of symptoms longer than 14 days. These treatments are often complementary, each with its own advantages and limitations. The less invasive endovascular therapy allows simultaneous angiography to guide reperfusion therapy. Surgery offers definitive restoration of blood flow with thrombectomy or bypass surgery, but it carries higher short-term risks.

Chronic Lower Limb Ischemia

The treatment of patients with claudication is directed at both improving the walking distance and, perhaps more important, reducing the patient's risk for life-threatening cardiovascular complications of atherosclerosis. Risk factor modification (Chapter 52), an exercise prescription, antiplatelet agents (Chapter 38), and medical therapy to reduce claudication are the basic elements of treating chronic stable peripheral arterial disease.

Risk Factor Modification

Patients with known peripheral arterial disease should be encouraged to modify or eliminate atherosclerotic risk factors such as diabetes (Chapter 229), tobacco use (Chapter 32), hyperlipidemia (Chapter 206), and hypertension (Chapter 67) and to exercise regularly (Chapter 16). Specific goals include: blood pressure below 140/90 mmHg and below 130/80 mmHg for those with diabetes, LDL cholesterol below 100 mg/dL and below 70 mg/dL for those at high risk of ischemic events, HDL cholesterol above 40 mg/dL, and hemoglobin A_{1C} level below 7%.

Patients who are smokers or former smokers should be asked about their status of tobacco use at every visit. Smokers should receive counseling, usually be offered pharmacotherapy (e.g., nicotine replacement therapy, sometimes with varenicline or bupropion), and be considered for referral to a smoking cessation program (Chapter 32).[5] For patients with severe peripheral arterial disease and elevated CRP levels, statin therapy substantially improves overall survival.[■] This finding in patients with peripheral arterial disease is consistent with findings in other studies of patients with a history of vascular disease, with or without peripheral arterial disease, and with either elevated LDL cholesterol levels or elevated CRP levels. Blood pressure control is very important, especially in patients with coexisting diabetes. β-Blockers are effective antihypertensive therapy and are not contraindicated in peripheral arterial disease. Angiotensin-converting enzyme inhibitors (e.g., ramipril 10 mg/day) improve pain-free and maximum walking distance[■] and reduce the risk of cardiovascular death in patients with peripheral arterial disease.[■]

Exercise Therapy

Patients should be reassured that exercising, even though it may precipitate their claudication symptoms, is not harmful and is the preferred initial treatment. A meta-analysis comparing supervised with unsupervised exercise therapy demonstrated superior improvement for claudication with supervised exercise.[■] Improvement in walking distance can often be achieved with pharmacologic management; discontinuation of tobacco use; and a regular, supervised exercise program.[■] For example, a randomized trial of patients with aortoiliac peripheral arterial disease who underwent a supervised exercise program showed greater improvement in walking performance compared with those who underwent primary stent therapy or home walking with cilostazol.[■]

Supervised exercise commonly involves walking on a treadmill, with the initial workload set to elicit symptoms within 3 to 5 minutes of walking. The patient is permitted to rest until the symptoms resolve but then resumes exercising. Maximal benefits are associated with programs that require the patient to continue walking until the pain is near maximal and with sessions that last more than 30 minutes, occur three or more times per week, and continue for more than 6 months. It typically takes 1 to 2 months for the patient to begin to notice benefits, which gradually increase over several months. Recent evidence supports the benefit of home-based walking programs with group-mediated cognitive behavior intervention to improve functional performance when compared with a control arm of health education alone. These findings have implications for a large number of patients who are unable or unwilling to participate in supervised exercise programs.[■]

Antiplatelet and Antithrombotic Therapy

Antiplatelet therapy with aspirin (usually 75 to 325 mg/day) has an established role in the secondary prevention of cardiovascular events in all patients with peripheral arterial disease because they are at high risk of MI, stroke, and

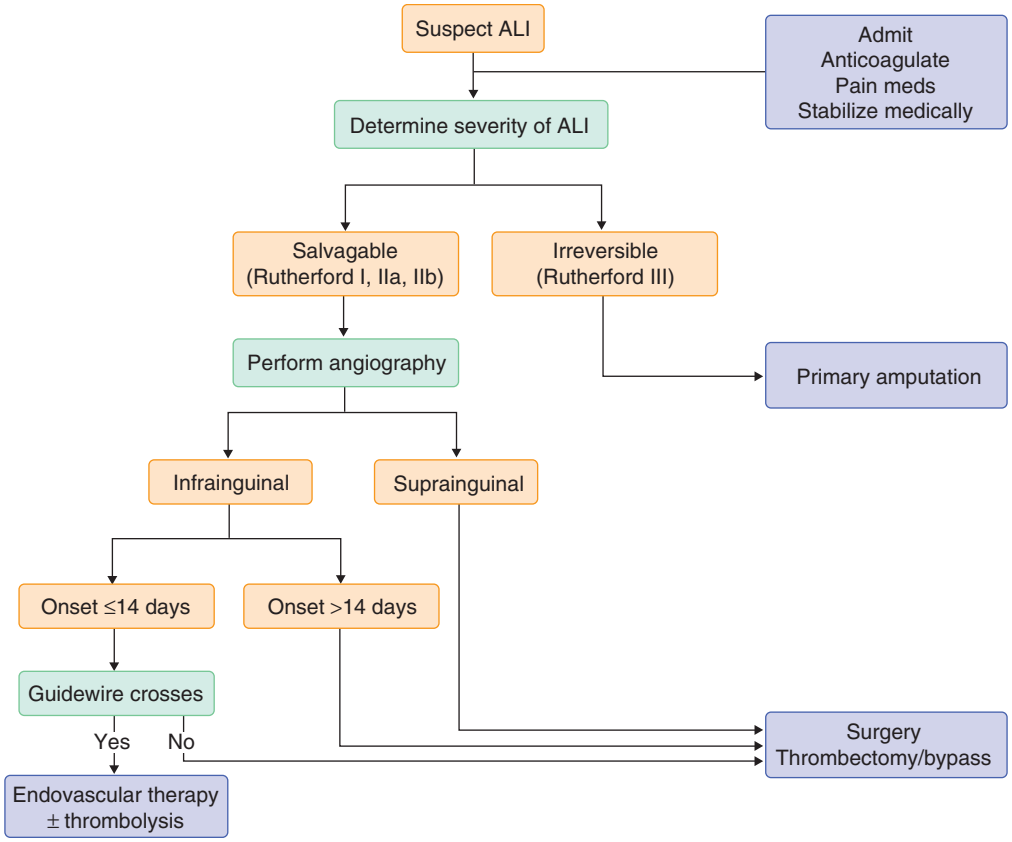

FIGURE 79-5. Treatment algorithm for acute limb ischemia (ALI). (Modified from Gray BH, Conte MS, Dake MD, et al. Atherosclerotic Peripheral Vascular Disease Symposium II: lower-extremity revascularization: state of the art. *Circulation.* 2008;118:2864-2872.)

vascular death, whether they are being managed medically, have had prior endovascular or surgical revascularization, or have had prior amputation. The thienopyridines, such as clopidogrel (usually 75 mg/day), are indicated only if aspirin is not tolerated, based on clopidogrel's efficacy compared with aspirin among patients with peripheral arterial disease. Another future possibility may be the novel protease-activated receptor-1 antagonist voraxapar, which can significantly reduce the incidence acute limb ischemia and the need for revascularization in patients with peripheral arterial disease.

Pharmacology

Currently, two medications are approved in the United States for the symptomatic treatment of intermittent claudication: pentoxifylline and cilostazol. Cilostazol, a phosphodiesterase inhibitor, at 50 to 100 mg twice daily improves maximal walking distance by 40% to 50% compared with placebo and has been shown in a randomized trial to reduce femoral artery stent restenosis. Cilostazol is contraindicated in patients with heart failure. By comparison, pentoxifylline has not consistently improved treadmill walking distance in randomized trials, so it cannot be recommended. Oral vasodilating prostaglandins, vitamin E, and chelation therapy with EDTA (ethylenediaminetetraacetic acid) have not been effective in improving either symptoms or walking distance.

Revascularization

The decision to perform a percutaneous or surgical revascularization procedure for the relief of claudication in a patient with chronic stable lower limb ischemia is based upon a risk-to-benefit assessment that weighs the patient's disability and discomfort against the estimated short- and long- term benefits and risks of the procedure. Because very few patients with claudication are in danger of losing their limbs, the primary goal of revascularization is long-lasting relief of symptoms, not limb salvage. Patients selected for revascularization to relieve symptoms of intermittent claudication should have significant lifestyle limitations or be unable to work and should have failed to respond to pharmacologic and exercise therapy.[6]

Superficial femoral and popliteal artery stenosis or occlusion (Fig. 79-6) are commonly associated with calf claudication. Revascularization with surgery or percutaneous transluminal angioplasty, with or without stenting, is indicated for the relief of vocational or lifestyle-limiting claudication in patients who have failed exercise and pharmacologic therapy. Angioplasty is preferred when possible in patients younger than 50 years because they have a higher risk of surgical graft failure than do older patients.

In patients with claudication, supervised exercise therapy plus balloon angioplasty improves the ABI at 1 year compared with supervised exercise therapy alone but does not improve quality-of-life measures. Randomized trials comparing surgery with angioplasty show similar rates of mortality, amputation, and patency at 4 years in patients with lower extremity ischemia. Percutaneous transluminal angioplasty with or without stenting is preferred in amenable lesions because of its lower periprocedural mortality and morbidity. Angioplasty is more cost effective than surgery if the expected 5-year patency rate of the treated vessel is 30% or greater.

To obtain a durable benefit after angioplasty, primary stenting with nitinol self-expanding stents or paclitaxel coated self-expanding stents is recommended for longer femoral lesions (7-10 cm) to reduce restenosis and to improve ABI and walking distance. In more discrete femoral lesions (mean, 4.5 cm), a strategy of balloon angioplasty first, with stenting only for bailout if the primary angioplasty is unsuccessful, is as good as routine stenting.

Adjunctive angioplasty devices such as atherectomy, cryotherapy, and the cutting balloon have not been meaningfully tested in any population, and there are few data to support their use versus less expensive, more conventional therapies. In randomized trials, laser angioplasty is not superior to conventional percutaneous transluminal angioplasty or stent placement in the superficial femoral artery. Given the substantial additional expense associated with these devices, more evidence of their efficacy is needed before widespread adoption can be justified. Small randomized trials have shown superior patency when femoral angioplasty is performed using drug-coated balloons compared with standard balloon angioplasty alone, but it is too soon to know whether this approach will substitute for femoropopliteal stenting in the future.

For chronic critical lower limb ischemia, endovascular therapy has reduced the rate of amputation. For patients with critical limb ischemia and salvageable limbs, the optimal treatment is urgent revascularization. The therapeutic goal is to reestablish pulsatile, straight-line flow to the distal extremity. Establishment of uninterrupted flow to at least one infrapopliteal vessel (i.e., the anterior or posterior tibial or peroneal arteries) is a prerequisite for wound healing.

In patients with critical limb ischemia caused by infrainguinal disease, percutaneous transluminal angioplasty and surgery are comparable as first-line therapies, but endovascular therapy is less costly and is associated with lower morbidity. It is reasonable to attempt a percutaneous therapy first if a patient is a candidate for either surgery or angioplasty, particularly if the patient's life expectancy is less than 2 years.[5] The use of coronary drug-eluting stents in tibial arteries appears to be more effective than conventional therapy.

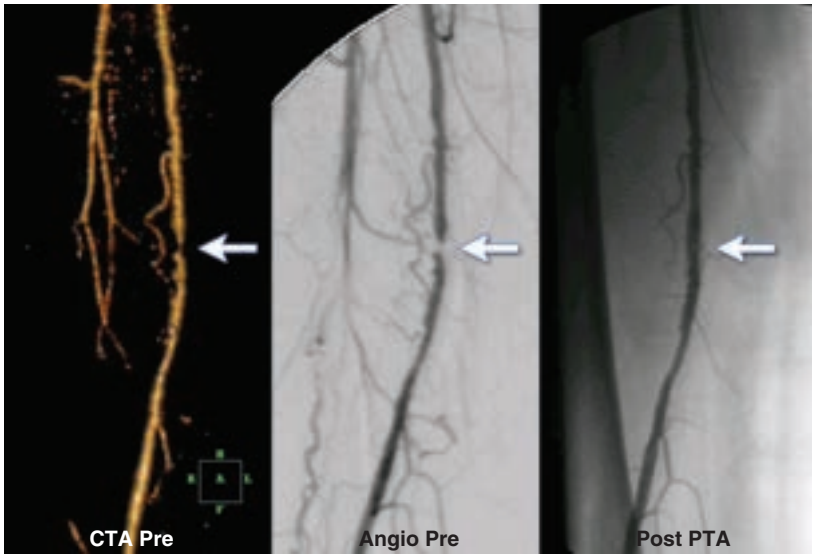

FIGURE 79-6. **Discrete stenosis** *(arrows)* **of the left superficial femoral artery seen on computed tomography angiography (CTA; *left*), digital subtraction angiography *(middle)*, and digital angiography after percutaneous transluminal angioplasty (PTA; *right*). (From White C. Intermittent claudication. *N Engl J Med.* 2007;356:1241-1250.)**

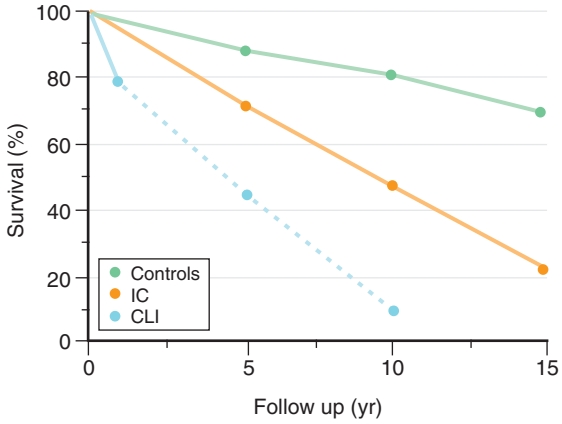

FIGURE 79-7. **Survival curve of patients with intermittent claudication (IC) versus critical limb ischemia (CLI). (Modified from Norgren L, Hiatt WR, Dormandy JA, et al. Inter-society consensus for the management of peripheral arterial disease (TASC II). *Eur J Vasc Endovasc Surg.* 2007;33(suppl):S1-S75.)**

Nevertheless, primary amputation is indicated in patients who have extensive necrosis or infectious gangrene with rest pain, are not ambulatory, and are not candidates for revascularization. Cellular therapies and growth factor therapy are being investigated for limb salvage, but none have yet proven clinically effective.

PROGNOSIS

Peripheral arterial disease is a major cause of acute and chronic illness associated with impaired functional capacity, reduced quality of life, limb loss, and increased risk of death (Fig. 79-7). Approximately two thirds of patients with peripheral arterial disease have at least one severely diseased coronary artery, and up to one quarter of patients have significant carotid artery stenosis. Consequently, patients with peripheral arterial disease face an increased risk of cardiovascular ischemic events such as MI, ischemic stroke, and death. It is estimated that coronary and cerebrovascular adverse events occur two- to fourfold more commonly than do limb adverse events in patients with peripheral arterial disease. Outcomes are improved when guideline-recommended therapy is followed.[7]

The annual mortality rate for patients with peripheral arterial disease is about 5%, but it is higher for patients with severe disease. For example, the estimated 1-year mortality rate for patients with critical limb ischemia is 25%, but this can climb to 45% for patients requiring amputation; in contrast, annual mortality rate for patients with intermittent claudication is only 1% to 2%. For patients presenting with acute limb ischemia, the 30-day amputation rate is as high as 40%, and mortality rates up to 30% have been reported.

Patients with lower extremity claudication should be reassured that the risk of limb loss is low. Moreover, a history of claudication by itself only slightly increases the risk of amputation after 10 years. However, a reduced ABI and diabetes mellitus are associated with the development of ischemic rest pain and ischemic ulceration, which may lead to limb loss. Among patients with peripheral arterial disease, those who have diabetes are 15 times more likely to have an amputation than are patients without diabetes, whose annual amputation rate is less than 1%.

Grade A References

A1. Ouwendijk R, de Vries M, Pattynama PM, et al. Imaging peripheral arterial disease: a randomized controlled trial comparing contrast-enhanced MR angiography and multi-detector row CT angiography. *Radiology*. 2005;236:1094-1103.

A2. Schillinger M, Exner M, Mlekusch W, et al. Statin therapy improves cardiovascular outcome of patients with peripheral artery disease. *Eur Heart J*. 2004;25:742-748.

A3. Ahimastos AA, Walker PJ, Askew C, et al. Effect of ramipril on walking times and quality of life among patients with peripheral artery disease and intermittent claudication: a randomized controlled trial. *JAMA*. 2013;309:453-460.

A4. Yusuf S, Sleight P, Pogue J, et al. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *N Engl J Med*. 2000;342:145-153.

A5. Fokkenrood HJ, Bendermacher BL, Lauret GJ, et al. Supervised exercise therapy versus non-supervised exercise therapy for intermittent claudication. *Cochrane Database Syst Rev*. 2013;8:CD005263.

A6. Lane R, Ellis B, Watson L, et al. Exercise for intermittent claudication. *Cochrane Database Syst Rev*. 2014;7:CD000990.

A7. Murphy TP, Cutlip DE, Regensteiner JG, et al. Supervised exercise versus primary stenting for claudication resulting from aortoiliac peripheral artery disease: six-month outcomes from the Claudication: Exercise Versus Endoluminal Revascularization (CLEVER) study. *Circulation*. 2012;125:130-139.

A8. McDermott MM, Liu K, Guralnik JM, et al. Home-based walking exercise intervention in peripheral artery disease: A randomized clinical trial. *JAMA*. 2013;310:57-65.

A9. Bedenis R, Stewart M, Cleanthis M, et al. Cilostazol for intermittent claudication. *Cochrane Database Syst Rev*. 2014;10:CD003748.

A10. Iida O, Yokoi H, Soga Y, et al. Cilostazol reduces angiographic restenosis after endovascular therapy for femoropopliteal lesions in the Sufficient Treatment of Peripheral Intervention by Cilostazol study. *Circulation*. 2013;127:2307-2315.

A11. Ahimastos AA, Pappas EP, Buttner PG, et al. A meta-analysis of the outcome of endovascular and noninvasive therapies in the treatment of intermittent claudication. *J Vasc Surg*. 2011;54:1511-1521.

A12. Schillinger M, Sabeti S, Loewe C, et al. Balloon angioplasty versus implantation of nitinol stents in the superficial femoral artery. *N Engl J Med*. 2006;354:1879-1888.

A13. Dake MD, Ansel GM, Jaff MR, et al. Sustained safety and effectiveness of paclitaxel-eluting stents for femoropopliteal lesions: 2-year follow-up from the Zilver PTX randomized and single-arm clinical studies. *J Am Coll Cardiol*. 2013;61:2417-2427.

A14. Werk M, Albrecht T, Meyer DR, et al. Paclitaxel-coated balloons reduce restenosis after femoropopliteal angioplasty: evidence from the randomized PACIFIER trial. *Circ Cardiovasc Interv*. 2012;5:831-840.

A15. Feiring AJ, Krahn M, Nelson L, et al. Preventing leg amputations in critical limb ischemia with below-the-knee drug-eluting stents: the PaRADISE (PReventing Amputations using Drug eluting StEnts) trial. *J Am Coll Cardiol*. 2010;55:1580-1589.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics—2014 update: a report from the American Heart Association. *Circulation*. 2014;129:e28-e292.
2. Creager MA, Kaufman JA, Conte MS. Clinical practice. Acute limb ischemia. *N Engl J Med*. 2012; 366:2198-2206.
3. Wennberg PW. Approach to the patient with peripheral arterial disease. *Circulation*. 2013;128: 2241-2250.
4. Mohler ER 3rd, Gornik HL, Gerhard-Herman M, et al. ACCF/ACR/AIUM/ASE/ASN/ICAVL/ SCAI/SCCT/SIR/SVM/SVS 2012 Appropriate Use Criteria for Peripheral Vascular Ultrasound and Physiological Testing Part I: Arterial Ultrasound and Physiological Testing: A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, American College of Radiology, American Institute of Ultrasound in Medicine, American Society of Echocardiography, American Society of Nephrology, Intersocietal Commission for the Accreditation of Vascular Laboratories, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, Society for Interventional Radiology, Society for Vascular Medicine, and Society for Vascular Surgery. *J Am Coll Cardiol*. 2012;60:242-276.
5. Anderson JL, Halperin JL, Albert NM, et al. Management of patients with peripheral artery disease (compilation of 2005 and 2011 ACCF/AHA guideline recommendations): a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2013;127:1425-1443.
6. Friedell ML, Stark KR, Kujath SW, et al. Current status of lower-extremity revascularization. *Curr Probl Surg*. 2014;51:254-290.
7. Armstrong EJ, Chen DC, Westin GG, et al. Adherence to guideline-recommended therapy is associated with decreased major adverse cardiovascular events and major adverse limb events among patients with peripheral arterial disease. *J Am Heart Assoc*. 2014;3:e000697.

REVIEW QUESTIONS

1. With the following pressures, calculate the patient's ankle-brachial index: right arm, 150/80 mm Hg; left arm 120/98 mm Hg; right dorsalis pedis, 100/50 mm Hg; right posterior tibial, 60/40 mm Hg; left dorsalis pedis, 75/50 mm Hg; and left posterior tibial, 40/20 mm Hg.

A. Right, 0.66; left, 0.50
B. Right, 0.50; left, 0.33
C. Right, 0.40; left, 0.33
D. Right, 0.66; left, 0.625

Answer: A The highest arm systolic pressure is used as the denominator (150 mm Hg) with the highest ankle systolic pressure on the right (100 mm Hg) and left (75 mm Hg) as the numerator.

2. Which of the following has the strongest effect on late patency for femoropopliteal stents?

A. Stent type: stainless steel vs. nitinol.
B. Stent length: ≤7 cm vs. longer.
C. Lesion length: ≤7 cm vs. longer.
D. Multiple vs. single stents placed.

Answer: C Lesion length is the strongest predictor of long-term stent patency in the femoropopliteal artery.

3. Which new technology holds the most promise for improving femoropopliteal artery endovascular results?

A. Laser angioplasty
B. Drug-eluting stents
C. Directional atherectomy
D. Cryoplasty

Answer: B Randomized trial evidence documents the improved outcomes with drug-eluting stents compared with primary angioplasty or provisional bare metal stents. Laser angioplasty, atherectomy, and cryoplasty have not demonstrated improved outcomes compared with conventional therapy.

4. Which of the following serious complications is most likely to occur with magnetic resonance angiography imaging of the lower extremities?

A. Contrast-induced nephropathy
B. Anaphylaxis to gadolinium contrast
C. Nephrogenic systemic fibrosis
D. Hemolytic anemia

Answer: C The major toxicity of gadolinium is an uncommon but potentially lethal systemic disorder called nephrogenic systemic fibrosis or nephrogenic sclerosing dermopathy (Chapter 267). A glomerular filtration rate of 60 mL/min or less is the major risk factor.

5. The major advantage for magnetic resonance angiography (MRA) over computed tomography angiography (CTA) for lower extremity imaging is

A. lower total diagnostic costs.
B. better imaging of calcified arteries.
C. better imaging of metallic stents.
D. more rapid imaging.

Answer: B Calcifications can cause artifacts that may be misdiagnosed as stenosis in CTA but not magnetic resonance angiography. An advantage of computed tomographic angiography over MRA is its ability to visualize metallic stents and stent grafts (see Fig. 79-4). CTA can be performed more quickly and with less pretreatment planning than MRA. In a randomized trial comparing MRA with CTA for initial imaging in peripheral arterial disease, there was no difference between the two techniques in terms of ease of use, clinical utility, or patient outcome, but CTA reduced total diagnostic costs.

80

OTHER PERIPHERAL ARTERIAL DISEASES

MICHAEL R. JAFF AND JOHN R. BARTHOLOMEW

POPLITEAL ARTERY ENTRAPMENT SYNDROME

Popliteal artery entrapment syndrome, which is a cause of intermittent claudication in young and often athletic patients, results from extrinsic compression on the popliteal artery by muscles or ligaments within or surrounding the popliteal fossa. The artery may be entrapped by various muscle components within the popliteal fossa. The popliteal vein also can be involved, rarely, and sometimes the symptoms are functional owing to hypertrophy of the medial head of the gastrocnemius muscle.

Anatomic evidence of entrapment can be found on physical examination or imaging performed for unrelated reasons in up to 4% of apparently normal individuals and imaging evidence for popliteal artery compression with provocative maneuvers in as many of 80% of asymptomatic individuals. Clinically evident popliteal artery entrapment syndrome is rare, with one study showing a prevalence of approximately 0.16% among military recruits.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Symptoms of popliteal artery entrapment syndrome include exertional limb pain, paresthesias and cold feet after exercise, ischemic rest pain, and even tissue necrosis if undiagnosed cases progress to arterial degeneration with thromboembolization. Entrapment of the popliteal vein results in leg swelling, heaviness, varicosities, nocturnal calf cramping, and even deep vein thrombosis (Chapter 81).

The diagnosis of popliteal artery entrapment syndrome is suggested by demonstration of popliteal artery compression on active pedal plantar flexion against resistance. This compression is manifested by a decrease in the intensity of the pedal arterial pulse on physical examination and loss of the continuous wave Doppler signal over the pedal arteries. Pulse volume recordings and segmental limb pressures should be measured at rest with the knee extended and the ankle in the neutral, dorsiflexed, and plantarflexed positions. Exercise treadmill studies also may demonstrate diminished limb arterial pressure after exercise in the symptomatic limb. Arterial duplex ultrasonography, dynamic computerized tomographic (CT) arteriography, or magnetic resonance (MR) arteriography may confirm the diagnosis, and CT or MR arteriography also define the structures that cause the entrapment.

TREATMENT AND PROGNOSIS

Rx

Treatment of structural popliteal artery entrapment is largely surgical with relief of the entrapment by resection or translocation of the compressing elements.[1] If popliteal artery entrapment syndrome is not identified until after the artery has been injured, arterial reconstruction or surgical bypass, ideally with a venous autologous graft, may be required. In functional popliteal artery entrapment, resection or translocation of the medial head of the gastrocnemius muscle is performed. If therapy is offered early in the course of the syndrome, patients may expect full recovery with complete resolution of symptoms.

CYSTIC ADVENTITIAL DISEASE OF THE LOWER EXTREMITY ARTERIES

In cystic adventitial disease, a mucinous cyst develops within the adventitial layers of the arterial wall and encroaches on the arterial lumen.[2,3] Cystic adventitial disease most commonly affects the popliteal artery, where it results in intermittent claudication, classically in middle-aged men. It also has been described in the external iliac, femoral, radial, and ulnar arteries.

The exact cause of cystic adventitial disease is unknown. Theories include a systemic disorder, repetitive trauma, and a persistent embryonic synovial track.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Intermittent claudication is the most common manifesting symptom of cystic adventitial disease. However, limb discomfort may persist for up to 20 minutes after the cessation of activity, unlike the classic relief of limb pain on cessation of activity in atherosclerotic peripheral artery disease (Chapter 79). The limb symptoms of cystic adventitial disease also tend to be inconsistent, often reappearing without obvious precipitants and resolving without an obvious explanation. Acute limb ischemia (Chapter 79) secondary to arterial compression and thrombosis also can occur.

The diagnosis is suspected when pedal pulses disappear with passive knee flexion and occasionally with exercise. However, cystic adventitial disease is most often confirmed with MR angiography. By comparison, arteriography may demonstrate only compression of the arterial lumen without identifying the cyst.

TREATMENT AND PROGNOSIS

Rx

Image-guided aspiration of the cyst may be attempted, but recurrence is likely. Balloon angioplasty is unlikely to be an effective and durable therapy and is not advised. Surgical resection of the cyst, potentially with placement of a venous interposition graft, is the primary therapy. When this condition is identified early, patients may anticipate complete resolution of claudication after resection of the cyst.

ENDOFIBROSIS OF THE ILIAC ARTERY

Endofibrosis of the iliac artery represents stenosis of the external iliac artery, thought to be due to repetitive trauma in highly functioning and competitive athletes.[3] This condition commonly occurs in cyclists or runners 30 to 50 years of age.

Endofibrosis results from intimal hyperplasia and fibrosis of the arterial wall. Smooth muscle cell proliferation and medial and intimal proliferation also may occur.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Symptoms include exercise-interfering intermittent claudication along with a sensation of swelling or paresthesia in the proximal lower limb at the time of maximal exercise. Physical examination may be normal at rest, although a bruit may be heard over the ipsilateral pelvic fossa or inguinal region.

Diagnostic imaging includes preexercise and postexercise Doppler ankle pressure measurements at the time of maximal, symptom-limiting treadmill exercise. Ultrasound imaging and contrast angiography, preferably when the leg is flexed at the hip in the cycling position, will reveal concentric stenosis, often with lengthening of the affected iliac artery. Intravascular ultrasound and intra-arterial translesional pressure gradients may be helpful.

TREATMENT AND PROGNOSIS

Rx

Surgical revascularization with patch angioplasty or interposition grafts has been the mainstay of treatment. However, percutaneous transluminal angioplasty and stent deployment may represent the optimal initial strategy, particularly among patients who agree to change their exercise routines. In patients who continue with high-intensity exercise, however, the long-term durability of a metallic stent within an artery prone to repetitive trauma is an ongoing concern.

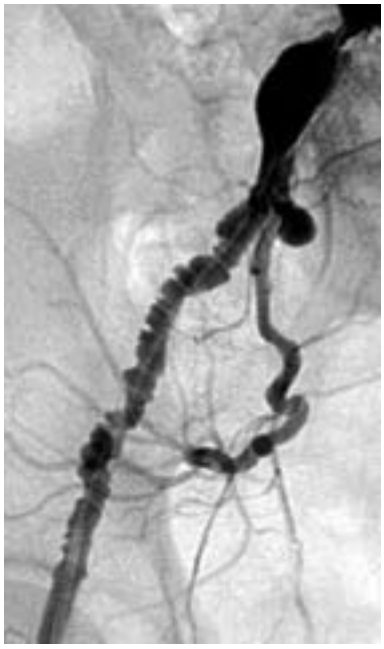

FIGURE 80-1. Angiogram showing a typical "string of beads" pattern, typical of the medial type in fibromuscular dysplasia, in the external iliac artery. Also note the aneurysm proximal to the area of dysplasia. Courtesy Dr. Jeffrey W. Olin.

After early detection, surgical or endovascular revascularization usually ameliorates the symptoms. With continued trauma to the arterial segment, however, careful surveillance for recurrence warranted.

FIBROMUSCULAR DYSPLASIA

Fibromuscular dysplasia[4] (Chapter 125) is a noninflammatory, nonatherosclerotic arterial disease that most commonly involves the medial layer of the artery wall and can affect any artery in the body. It is most commonly seen in women 20 to 60 years of age, with a mean age at onset of 52 years, but a range of 5 to 83 years.[5] Its true prevalence is unknown, but a series of potential renal donors showed a prevalence as high as almost 4%.

Medial fibroplasia, which is the most common type of fibromuscular dysplasia, results in the classic "string of beads" appearance on contrast arteriography (Fig. 80-1). Pathology shows alternating segments of thinning and thickening of the arterial media, but the adventitia or intima may be predominantly affected, thereby resulting in an angiographic appearance different from that for medial fibroplasia. Other types of fibromuscular dysplasia include an intimal variant that causes a single, focal, weblike stenosis within the affected artery.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The clinical presentation of patients with fibromuscular dysplasia depends on the arteries involved, and sometimes it may be incidentally discovered without any symptoms when imaging is performed for other reasons. The renal arteries are affected in almost 80% of cases (Chapter 125), and the extracranial carotid arteries are involved in almost 75% of cases. The third most commonly affected artery is the vertebral artery, which is involved in nearly 40% of patients. Not surprisingly, hypertension (Chapter 67) and headache (Chapter 398) represent the two most common manifesting signs and symptoms. Less common symptoms include dizziness, cervical bruits, and pulsatile tinnitus. Fibromuscular dysplasia may occasionally affect the iliac, femoral, or popliteal arteries.

The diagnosis of fibromuscular dysplasia relies on a combination of clinical and imaging findings. Doppler ultrasound, MR angiography, and CT angiography can identify fibromuscular dysplasia and assess possible aneurysms, but contrast angiography and intravascular ultrasound are the best diagnostic tests. Because fibromuscular dysplasia may involve several vascular beds, patients should initially undergo comprehensive MR or CT imaging of their intracranial arteries, the extracranial carotid arteries, the thoracic and abdominal aorta and its branches, and the renal arteries. If arterial aneurysms or dissections are identified without clear evidence of fibromuscular dysplasia, a genetic collagen disorder should be considered.

TREATMENT AND PROGNOSIS

Rx

Treatment of fibromuscular dysplasia depends on the clinical presentation and extent of arterial involvement. Asymptomatic patients are empirically prescribed aspirin. In young patients with sudden onset of hypertension and renal artery fibromuscular dysplasia, percutaneous transluminal angioplasty can cure the hypertension that is often found (Chapter 125).[6] Renal and intracranial artery aneurysms should be followed closely for expansion and may require surgical repair. The decision about repair is often based on the size of the aneurysm, its rate of expansion, and the presence of associated symptoms. Carotid artery fibromuscular dysplasia is generally treated with antiplatelet therapy (e.g., aspirin 81 to 325 mg/day) or 3 to 6 months of warfarin anticoagulation in the case of carotid artery dissection; percutaneous transluminal angioplasty is reserved for patients with hemispheric cerebral ischemic symptoms and progressive dissection. Peripheral artery percutaneous transluminal angioplasty is recommended only for patients with symptoms of intermittent claudication.

In general, patients with fibromuscular dysplasia have an excellent outcome. After percutaneous transluminal angioplasty of the renal arteries, approximately one third of patients with hypertension will have normal blood pressures and most others will have substantial improvement. Optimal antiplatelet therapy (e.g., aspirin 81 to 325 mg/day) is recommended for the majority of patients with fibromuscular dysplasia. The most feared complications, which occur in less than 5% of patients, are stroke owing to progressive carotid dissection, rupture of an intracranial arterial aneurysm, or rupture of a visceral artery aneurysm.

THROMBOANGIITIS OBLITERANS (BUERGER DISEASE)

Thromboangiitis obliterans is a nonatherosclerotic, segmental inflammatory disorder that affects small and medium-sized arteries, veins, and nerves. Thromboangiitis obliterans is typically found in patients who are under the age of 50 years and who smoke or chew tobacco.[7]

Although rarely performed, biopsy of affected digits demonstrates highly cellular inflammatory thrombi with sparing of the artery wall. Thromboangiitis obliterans is one of the few vascular disorders that affects both arteries and veins, so superficial thrombophlebitis can also develop.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients classically present with digital or extremity pain with digital ischemia, distal limb claudication (such as in the arch of the foot or arm), and digital ulcerations (Fig. 80-2). More than 40% of patients have Raynaud phenomenon.

Clinicians must maintain a high degree of clinical suspicion in young smokers with distal extremity ischemia and pain without evidence of atherosclerosis. Many patients will demonstrate an abnormal Allen test, in which one of two of the major arteries to the hand (radial or ulnar) does not fill after the examiner compresses the other artery, as well as absent peripheral arterial pulses. The ankle brachial-index (Chapter 79) is classically abnormal, even though plethysmographic studies demonstrate essentially normal proximal artery circulation. Serologic results show no evidence for inflammatory vasculitis (Chapter 270), and proximal sources of arterial emboli are not found. Contrast arteriography is often needed to demonstrate distal arterial involvement with the absence of atherosclerosis. Segmental arterial occlusions with corkscrew collaterals are commonly found but are not pathognomonic.

TREATMENT AND PROGNOSIS

Rx

The cornerstone of treatment is complete abstinence from tobacco use (Chapter 32). Although less effective, antiplatelet therapy (e.g., aspirin 81 to 325/day) is commonly prescribed. A 28-day course of the prostacyclin analogue iloprost (1 ng/kg/minute for 6 hours daily) can substantially improve pain and accelerate wound healing compared with aspirin or lumbar sympathectomy. There are no data demonstrating benefit with vasodilators such as calcium channel antagonists, peripheral α-blockers, or sildenafil. Digital sympathectomy, bosentan, spinal cord stimulators, and intermittent pneumatic compression have all been tried with variable success.

Patients with thromboangiitis obliterans must discontinue all exposure to tobacco. However, unsuccessful patients commonly have unremitting pain, recurrent tissue necrosis, and the need for progressive limb amputation.

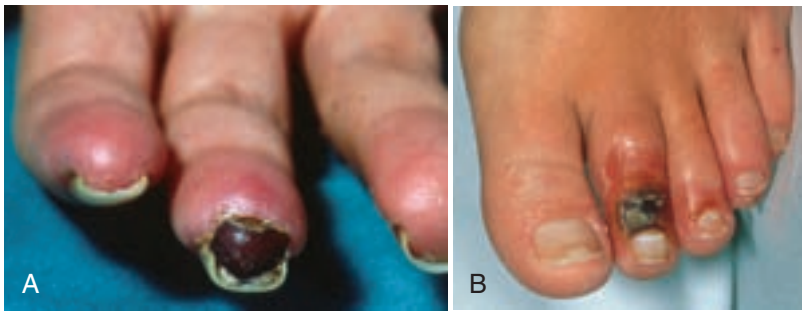

FIGURE 80-2. Buerger disease. Ischemic finger of a young male patient (**A**) and ischemic toe of a 28-year-old woman (**B**) with Buerger disease. Courtesy Dr. Jeffrey W. Olin.

LIVEDO RETICULARIS

Livedo reticularis is a vasospastic disorder that appears as a violaceous or bluish netlike discoloration surrounding a pale central area of skin. It normally involves the lower extremities.

Livedo reticularis is commonly found in healthy women during their second to fifth decade of life and strongly associated with antiphospholipid antibodies (Chapter 176). It is rare in males.

Livedo reticularis is an ischemic dermopathy caused by an increased prominence of venous beds owing to obstruction of arterial inflow, venous dilatation, or blockage of venous outflow.[8] Primary livedo reticularis is a benign condition exacerbated by cold, tobacco, or emotional upset, whereas the secondary form, referred to as livedo racemosa, is a pathologic variant seen in association with a number of disorders (Table 80-1).

TABLE 80-1 DISORDERS ASSOCIATED WITH LIVEDO RACEMOSA

Polyarteritis nodosa, systemic lupus erythematosus, rheumatoid arthritis, Sjögren syndrome, scleroderma
Atheromatous embolization, atrial myxoma
Antiphospholipid syndrome, Sneddon syndrome
Infections
Polycythemia vera, essential thrombocythemia, cryoglobulinemia, cryofibrinogenemia, cold agglutinin disease
Calciphylaxis, hyperoxaluria
Medications: Amantadine, amphetamines, ergotamines, vasopressors (epinephrine, norepinephrine, dopamine), heparin, minocycline

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The legs and to a lesser extent arms are most commonly involved in primary livedo reticularis, whereas the face, trunk, buttocks, and extremities may be involved with livedo racemosa. The discoloration of livedo reticularis is worsened by dependency and improves with elevation. It is important to differentiate the usually painless, symmetrical, unbroken vessel network of primary livedo reticularis from the frequently painful, irregular, asymmetrical, and broken pattern observed in livedo racemosa.

A thorough history should focus on exacerbating conditions. Laboratory investigation is usually not necessary for the primary form, but specific testing for diagnosing the secondary forms includes antiphospholipid antibody levels (including circulating lupus anticoagulant; anticardiolipin antibodies; P and C antineutrophil cytoplasmic antibodies) should be performed. A skin biopsy may be necessary when serologic findings are indeterminate and the diagnosis remains unclear. The differential diagnosis includes erythema ab igne, livedoid vasculopathy (a rare subtype of livedo racemosa), and acrocyanosis.

TREATMENT AND PROGNOSIS

Rx

Cold temperatures, tobacco, and stressful conditions should be avoided. Primary livedo reticularis does not require treatment, although calcium channel blockers (e.g., nifedipine 10 to 20 mg PO every 6 hours or amlodipine 2.5 to 10 mg/day PO) may be beneficial for patients who are uncomfortable with their appearance. Treatment for livedo racemosa should focus on the underlying disorder. Anticoagulants, including warfarin to provide an international normalized ratio of 2.0 to 3.0, antiplatelet therapy (e.g., aspirin 81 to 325/day), or both are recommended for livedo racemosa associated with the antiphospholipid syndrome, cerebrovascular disease (Sneddon syndrome), or both. The prognosis is excellent for primary livedo reticularis, but depends on the underlying disorder for patients with livedo racemosa.

ATHEROMATOUS EMBOLIZATION

Atheromatous embolization, also known as cholesterol crystal embolization, atheroembolic renal disease, or the blue or purple toe syndrome, refers to the showering of multiple small cholesterol crystals or fibrin-platelet aggregates to the extremities or any organ. The exact incidence of atheromatous embolization is unknown, in part because it is often poorly recognized. In unselected autopsy series, the prevalence ranges from 0.18% to 2%, although it is reported to be much higher in individuals who have undergone aortic manipulation, arteriography, or cardiac or vascular procedures.[9] As many as 5% to 10% of all cases of acute renal failure may result from atheroembolism (Chapter 125). Atheromatous embolization is more common in elderly individuals with advanced atherosclerosis but is less commonly diagnosed in blacks, largely owing to a failure to recognize the classic dermatologic feature because of their skin pigmentation.

Atheromatous embolization usually originates from ulcerated or stenotic atherosclerotic plaques or aneurysms of both large and small arteries. Precipitating factors include arteriography, endovascular procedures (cerebral, coronary, or peripheral), surgery, trauma, or anticoagulation. Atheroembolism also may occur spontaneously. Light microscopy demonstrates multiple biconvex, needle-shaped cholesterol crystals that lodge in arterioles and result in a foreign body reaction in which polymorphonuclear leukocytes, macrophages, and multinucleated giant cells are observed days to weeks after the initiating event. This process eventually leads to end-organ damage owing to intraluminal obliteration, ischemia, and even often infarction.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Dermatologic findings involving the lower extremities are the most common clinical presentation (Fig. 80-3), but clinical manifestations may be seen in multiple organs (Table 80-2).[10] Atheromatous embolization is frequently overlooked or misdiagnosed because the signs and symptoms are nonspecific and diverse. Once suspected, the diagnosis usually can be made on clinical grounds alone in a patient who has had a precipitating event, acute or subacute renal failure, difficult-to-control hypertension, or evidence of peripheral embolization.[11] In some cases, the definitive diagnosis may require biopsy of the skin, kidney, or gastrointestinal tract. Patients may have elevated markers of inflammation (erythrocyte sedimentation rate, C-reactive protein) or transient eosinophilia. Elevations in amylase, hepatic aminotransferase, blood urea nitrogen, serum creatinine, or serum creatinine kinase levels may be seen with involvement of the pancreas, liver, kidney, or muscle, respectively. The urine is generally nonspecific, but eosinophiluria can be seen.

Invasive diagnostic angiographic procedures should be avoided. Noninvasive procedures including multidetector CT angiography, MR angiography, or transesophageal echocardiography may be helpful if they reveal a markedly irregular and shaggy aorta.

The differential diagnosis, which depends on the end organ involved, includes contrast-induced nephropathy (Chapters 57 and 120), polyarteritis nodosa (Chapter 270), leukocytoclastic vasculitis (Chapters 270 and 439), cryoglobulinemia (Chapter 187), the antiphospholipid syndrome

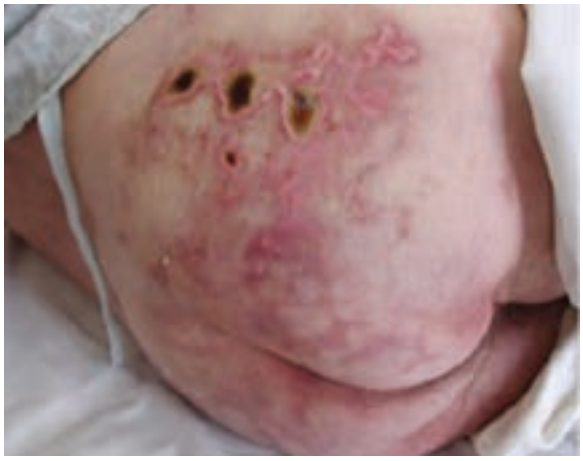

FIGURE 80-3. Livedo racemosa with ischemic ulcerations and violaceous, netlike, and broken patterns in a patient who developed atheromatous embolization after cardiac catheterization.

TABLE 80-2 CLINICAL MANIFESTATIONS OF ATHEROMATOUS EMBOLIZATION

Skin	Purple or blue toes Gangrenous digits Livedo reticularis or livedo racemosa Petechiae Ulcers, nodules Splinter hemorrhages
Kidney	Uncontrolled hypertension Advanced renal disease End-stage renal disease
Neurologic	Amaurosis fugax Hollenhorst plaque Transient ischemic attack or stroke Confusion, organic brain syndrome Spinal cord infarction
Gastrointestinal	Abdominal pain Diarrhea Gastrointestinal bleeding Ischemic bowel Acute pancreatitis Acute gangrenous cholecystitis
Cardiac	Angina pectoris Myocardial infarction
Constitutional symptoms	Fever Weight loss Malaise, myalgias Anorexia, nausea, vomiting

(Chapter 176), and thrombotic thrombocytopenia purpura (Chapter 172). An underlying malignancy is part of the differential diagnosis in patients who present with constitutional symptoms such as anorexia and weight loss. Cardiac sources such as nonbacterial thrombotic endocarditis (Chapter 76), infective endocarditis (Chapter 76), or atrial myxoma (Chapter 60) should be excluded.

TREATMENT AND PROGNOSIS

Rx

The most important aspect of treatment is prevention. The prevention of recurrent atheroembolism is essential. Smoking cessation (Chapter 32) and aggressive control of hypertension (Chapter 67), diabetes (Chapter 229), and hyperlipidemia (Chapter 206) should be instituted to prevent progression of the disease. In the absence of data from randomized trials, therapy is directed toward avoiding recurrent embolization, removing the source of the atheroemboli, and providing symptomatic care of the end organ(s) involved.

Patients with ischemic ulcers require pain control and local wound care. Statins (e.g., atorvastatin 10 to 80 mg/day or rosuvastatin 10 to 40 mg/day to obtain optimal low-density lipoprotein cholesterol levels [Chapter 206]) have been reported to be beneficial, likely a result of their plaque-stabilizing activity. Other pharmacologic approaches—including antiplatelet agents (aspirin or clopidogrel), calcium channel blockers (e.g., nifedipine 10 to 20 mg PO every 6 hours or amlodipine 2.5 to 10 mg/day PO), cilostazol (100 mg PO every 12 hours), pentoxifylline (400 mg PO every 8 hours), and intravenous prostaglandins—have been tried with varying degrees of success. The use of anticoagulants is controversial because of concerns that they may lead to plaque instability, and data do not support the use of corticosteroids.

Nonpharmacologic approaches including chemical or surgical sympathectomy, a spinal cord stimulator, and arterial flow pumps may help for pain control. Endovascular therapies (angioplasty, stent placement, atherectomy, or covered-stent grafts, and the use of embolic protective devices) may help prevent future embolic events, but studies are limited and further clinical evaluation is needed. Surgical therapies including thromboendarterectomy, aortobiiliac or aortobifemoral bypass grafting, or extra-anatomic reconstruction may be necessary to eliminate the embolic source and reduce further embolization.

Patients with advanced atherosclerosis and atheromatous embolization have a poor prognosis. The estimated 1-year mortality rate is approximately 15% even with optimal medical care.

THERMAL DISORDERS

A variety of clinical syndromes can be caused by heat or cold (Fig. 80-4).

Erythromelalgia

The term *erythromelalgia* is derived from the Greek words *erythros* ("red"), *melos* ("extremities"), and *algos* ("pain"). It is characterized by episodic periods of erythema, increased warmth, and intense burning pain of the extremities.

Erythromelalgia is an uncommon disorder that affects younger to middle-aged women, adolescents, and children. The exact incidence is not known but has been reported at 1.3 per 100,000 living in Olmsted County, Minnesota. Erythromelalgia also occurs in nearly 30% of patients with polycythemia vera (Chapter 166).

Both primary (familial and sporadic/idiopathic) and secondary forms occur. The primary trigger is heat exposure or increased ambient temperature. The familial form is autosomal dominant, with symptoms most often beginning in childhood. The pathophysiology is believed to be due to a small fiber neuropathy and vasculopathy. Although the exact cause for primary erythromelalgia is unknown, it is a neuropathic disorder with neuronal hyperexcitability owing to sodium channel abnormalities because of a mutation of the gene *(SCN9A)* that encodes the Nav1.7v sodium channel (located on 2q).[12] The causes of secondary erythromelalgia (Table 80-3) are not as well understood but are thought to be due to neuropathologic and microvascular functional changes caused by the underlying condition.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

A triad of clinical findings including erythema, increased warmth, and intense burning pain of the extremities is usually observed (Fig. 80-5). The lower limbs (soles of the feet) are more often affected than the hands. Involvement of the knees, elbows, ears, and face has been reported. Symptoms are generally bilateral and paroxysmal. An attack, which may last for several minutes or hours to days, is generally aggravated by dependency, alcohol, warm rooms, summer heat, exercise, or simply wearing shoes and socks or gloves. Erythromelalgia is often a disabling condition, and ulceration or even gangrene can occur in secondary forms.

The history and physical examination are keys to the diagnosis. The physical examination is usually normal unless the patient is examined during an attack. A thorough vascular and neurologic examination should include demonstration of color changes and measurement of elevated skin temperatures during an attack. A complete blood count is essential to exclude an underlying myeloproliferative disorder. Other tests including electromyography and nerve conduction studies, autonomic and small fiber nerve testing, and vascular studies may help exclude other disorders. The quantitative sudomotor axon reflex test is useful to assess for small fiber neuropathy. Genetic testing is helpful for diagnosing primary erythromelalgia.

The differential diagnosis includes complex regional pain syndrome (Chapter 30), cellulitis (Chapter 441), peripheral neuropathy (Chapter 420), osteomyelitis (Chapter 272), Raynaud syndrome, acrocyanosis, peripheral arterial disease (Chapter 79), and gout (Chapter 273).

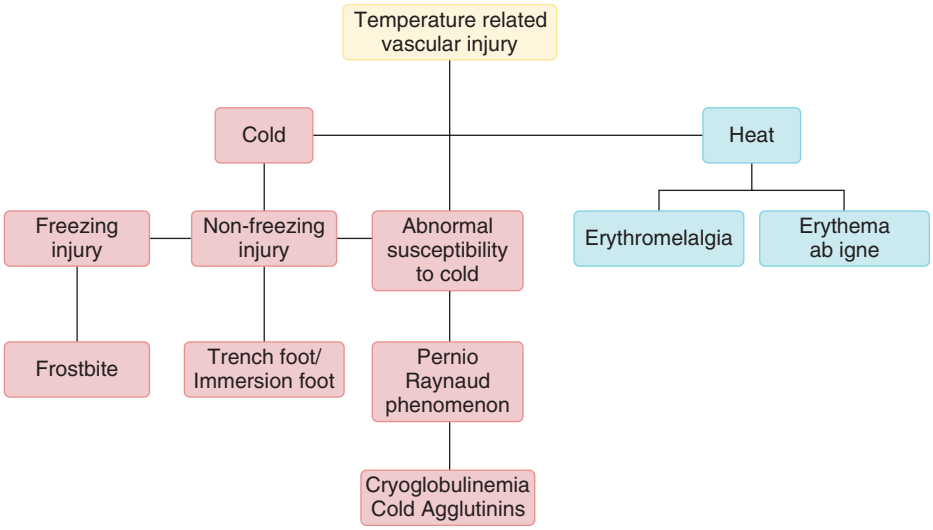

FIGURE 80-4. Clinical syndromes caused by heat or cold.

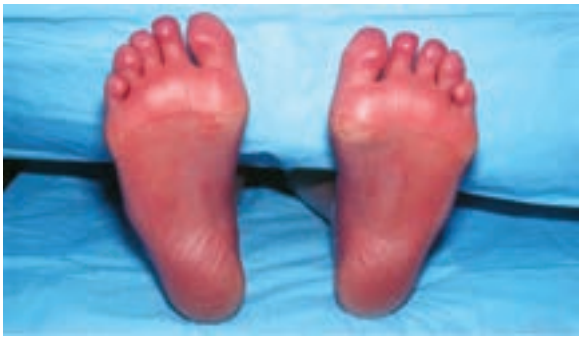

FIGURE 80-5. Erythromelalgia. Note the erythema of the feet. The patient also had pain and increased warmth on physical examination.

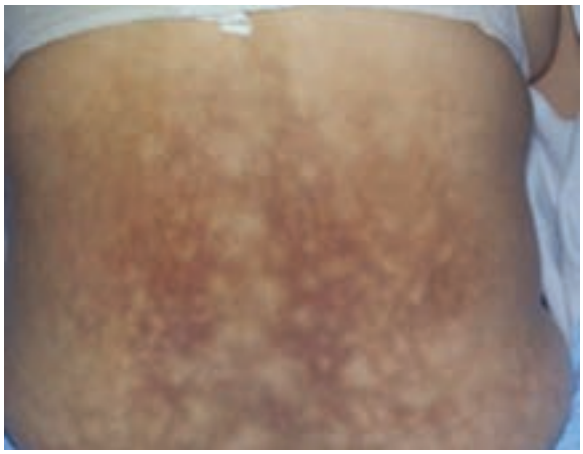

FIGURE 80-6. Erythema ab igne. Note the hyperpigmentation and livedo reticularis pattern in a patient who used a heating pad for back pain.

TABLE 80-3 CAUSES OF SECONDARY ERYTHROMELALGIA

Myeloproliferative neoplasms: Essential thrombocythemia, polycythemia vera, chronic myelogenous leukemia, myelodysplastic syndrome
Drugs: Calcium channel blockers, cyclosporine, bromocriptine, pergolide
Infectious diseases: Human immunodeficiency virus, hepatitis B vaccine, influenza vaccine, infectious mononucleosis, varicella virus
Connective tissue diseases: Systemic lupus erythematosus, rheumatoid arthritis
Neuropathic: Diabetic neuropathy, peripheral neuropathies, neurofibromatosis, Riley-Day syndrome, multiple sclerosis, spinal cord disease
Neoplastic: Paraneoplastic syndrome, astrocytoma, malignant thymoma, colorectal cancer, lung cancer, and thyroid cancer
Others: Mushroom ingestion, mercury poisoning

TREATMENT AND PROGNOSIS

Rx

Patients must avoid aggravating conditions, such as exercise, tight shoes, and alcohol intake, as well as learn to cool their involved areas without causing tissue damage. Patients seek relief by cooling the affected area with a fan, cold towels, cooling blankets, or immersion in ice water. Elevation of the feet may help. Medications aimed at treatment of both the neuropathy and vasculopathy are often tried. In one small randomized trial, intravenous iloprost (at varying doses for up to 6 hours per day on 3 consecutive days) significantly reduced symptoms and sympathetic dysfunction. In the genetic form, a small trial showed benefit from orally administered XEN402 (an inhibitor of $Na_v1.7$ that is designated as an orphan drug by the U.S. Food and Drug Administration). Aspirin can help patients with an underlying myeloproliferative disorder. Tricyclic antidepressants (e.g., nortriptyline 25 to 100 mg/day PO), anticonvulsants (e.g., pregabalin 50 to 100 mg PO three times daily), mexiletine (10 mg/kg/day PO), topical lidocaine patches, and opioids (e.g., hydromorphone 2 to 8 mg PO every 4 to 6 hours as needed) have been used with benefit in some patients. A topical gel of 1% amitriptyline and 0.5% ketamine has demonstrated promise without systemic side effects. Other methods, including a pain rehabilitation program, biofeedback, sympathectomy, and epidural blocks, have been tried with varying degrees of success. Patients have a reduced quality of life, and life expectancy is also reduced, primarily owing to suicide.

Erythema Ab Igne

Erythema ab igne is a hyperpigmented skin condition that results from repeated or chronic exposure to a heating source or infrared radiation that is not warm enough to burn the skin. The incidence is unknown, but women are more frequently affected than men.

Skin manifestations result from damage to the dermis and venous plexus system. Dysplastic changes can predispose the patient to actinic keratosis and squamous cell carcinomas (Chapter 203). Erythema ab igne is an occupational hazard for persons whose arms are repeatedly exposed to fire in bakeries, foundries, or kitchens. It can also result from repeated application of a hot water bottle, heating pads, or electric blankets or may be seen in persons who sit too close to space heaters, wood burning stoves, or even car heaters. It also has been reported involving the anterior thighs of persons using laptop computers.[13]

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients generally have no symptoms, although some individuals report a slight burning or itching sensation. The skin discoloration is described as reticular, erythematous, and brownish hyperpigmentation (Fig. 80-6). Ulceration and bullous lesions have been reported in chronic cases.

The diagnosis is made clinically based solely on dermatologic findings. A history of exposure to a heating source or infrared radiation should be pursued because no laboratory tests are helpful. A biopsy must be obtained if there are signs of a malignant transformation. The differential diagnosis includes livedo reticularis and livedo racemosa.

TREATMENT AND PROGNOSIS

Rx

Removing the offending heat source is essential for treatment, and patients should be advised to avoid prolonged exposure to any form of infrared heat. Topical therapies including topical tretinoin or hydroquinone have been used to reduce the hyperpigmentation.

The prognosis is favorable once the source is removed, but the condition can be chronic and progressive if prolonged and repeated exposure continues. Follow-up examinations are recommended because of the potential for malignant conversion.

Raynaud Phenomenon

Raynaud phenomenon is defined as episodic attacks of discoloration of the digits brought on by cold or emotional stimuli and resulting in a characteristic triphasic color change from white to blue to red. Raynaud phenomenon affects 3 to 5% of the U.S. population. It is seen more often in young women in whom it is reported to have a prevalence as high as 5 to 15%. The prevalence is higher in cooler northern climates and in smokers. Family history, estrogen exposure, and emotional stress are commonly associated with Raynaud phenomenon in women.[14] The hand-arm vibration syndrome is more common in men, especially men who use pneumatic hammers, chain saws, sanders, and grinders. It also has been reported in typists, pianists, meat cutters, and sewing machine operators.

The pathophysiology of the vasoconstriction in Raynaud phenomenon is not well understood. It includes abnormalities of the blood vessel wall, neural control mechanisms, and intravascular factors, including platelet activation and oxidative stress. As the arterial vasoconstriction subsides, postcapillary venular constriction leads to deoxygenation of the blood and the cyanotic appearance. On rewarming, blood flow increases as a result of vasodilation, which leads to the red or hyperemic appearance of the digits.

Primary Raynaud phenomenon is a benign vasospastic disorder, whereas a number of conditions are associated with the secondary form (Table 80-4). Several clinical features can help distinguish between these two forms (Table 80-5).

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Raynaud phenomenon is characterized by triphasic color change of the digits, and pallor, cyanosis, and rubor after exposure to cold or stressful stimuli (Fig. 80-7).[15] All three color changes are not seen in most individuals, and pallor may be the only finding. The middle and ring fingers are most commonly involved, whereas the thumb may be entirely spared. Raynaud phenomenon also can affect the toes, nose, ears, tongue, knees, or nipples. Patients may experience paresthesias and clumsiness of the hand during an attack. Some patients develop numbness, intense ischemic pain, and even necrosis.

TABLE 80-4 UNDERLYING CONDITIONS ASSOCIATED WITH SECONDARY RAYNAUD PHENOMENON

Rheumatologic: Scleroderma, systemic lupus erythematosus, rheumatoid arthritis, Sjögren syndrome, mixed connective tissue disorders
Obstructive arterial disease: Atherosclerosis, thromboangiitis obliterans, arterial embolism
Occupational/environmental disorders: Hypothenar hammer syndrome, hand-arm vibration syndrome, frostbite)
Endocrine: Hypothyroidism
Hematologic: Polycythemia vera, multiple myeloma, cryoglobulinemia, cryofibrinogenemia, cold agglutinins
Drugs: Amphetamines, cocaine, β-blockers, clonidine, ergot preparations, oral contraceptives, cyclosporine, certain anti-neoplastic agents
Infections: Hepatitis B and C antigenemia
Thoracic outlet syndrome, subclavian artery aneurysm
Complex regional pain syndrome
Arteriovenous fistula
Lead and arsenic poisoning

The evaluation should start with a thorough history. The physical examination is normal in primary Raynaud unless there is an ongoing attack. Findings in secondary Raynaud may include ulceration of the fingertips or an abnormal Allen test. Routine laboratory testing including a complete blood count, erythrocyte sedimentation rate and C-reactive protein, urinalysis, thyroid function tests, antinuclear antibody, serum protein electrophoresis, and chest radiograph should be performed to evaluate for secondary Raynaud phenomenon. If the antinuclear antibody is positive or if the patient's history and physical examination indicate an underlying rheumatologic disorder, specific autoantibodies should be ordered (Chapter 257). Other tests including cryoglobulin levels, cryofibrinogen, and cold agglutinins may be helpful, depending on the clinical presentation (Chapter 256).

A number of noninvasive tests may help differentiate primary and secondary Raynaud phenomenon and evaluate the extent of vasospasm, including photoplethysmography, pulse volume recordings, laser Doppler flux, duplex ultrasonography, and nail-fold capillary microscopy, often performed after a cold stress challenge, such as immersion of the hands in a bath of ice water. MR angiography or contrast angiography may be necessary, and the latter can help determine the cause of ischemia.

TREATMENT AND PROGNOSIS

Rx

Lifestyle modifications, including avoidance of cold exposure and known stressful stimuli, with an emphasis on keeping the body core temperature and extremities warm, are essential for preventing attacks. Raynaud phenomenon can be treated with both pharmacologic and nonpharmacologic approaches (Table 80-6).[15]

Calcium channel blockers can reduce attacks and the severity of symptoms in primary Raynaud.[■] Aggressive treatment of the underlying condition is essential for patients with secondary Raynaud phenomenon. Calcium channel blockers (e.g., nifedipine 10 to 20 mg PO every 6 hours or amlodipine 2.5 to 10 mg/day PO) and iloprost (1 ng/kg/minute for 6 hours daily) are clearly beneficial for secondary Raynaud[■] as is dual endothelin receptor blockade (e.g., bosentan 62.5 to 125 mg twice daily).[■] Phosphodiesterase-5 inhibitors also have moderate benefits.[■] Pain control with opioids, chemical or surgical sympathectomy, and spinal cord stimulation may be necessary in the most severe cases, particularly in situations of nonhealing digital ulceration and tissue loss.

The prognosis for primary Raynaud phenomenon is excellent, whereas the prognosis for secondary Raynaud phenomenon depends on the underlying condition.

TABLE 80-5 FEATURES SUGGESTIVE OF PRIMARY OR SECONDARY RAYNAUD PHENOMENON

CLINICAL FEATURE	PRIMARY	SECONDARY
Sex	Female	Female or male
Age	<40 yr	≥ 40 yr
Involvement	Bilateral	Unilateral or bilateral
Ischemic digits or ulcerations	Absent	±
Underlying cause	Absent	Present
Systemic complaints	Absent	±

FIGURE 80-7. Unilateral Raynaud phenomenon. From Forbes CD, Jackson WF. Color Atlas and Text of Clinical Medicine, 3rd ed. London: Mosby; 2003.

TABLE 80-6 PHARMACOLOGIC AND NONPHARMACOLOGIC THERAPIES FOR RAYNAUD PHENOMENON

NONPHARMACOLOGIC THERAPIES
Educate and reassure patients about their condition
Avoid cold exposure or other triggering factors
Use warm clothing to maintain core body temperature (cover the entire body and wear a hat and scarf)
Teach how to terminate attacks: Exit from cold, warming techniques
Avoid nicotine
Biofeedback
PHARMACOLOGIC THERAPIES
Calcium channel blockers (nifedipine 10-20 mg q6h and amlodipine 2.5-10 mg q25 hr
Phosphodiesterase-5 inhibitors (sildenafil 25-50 mg TID, tadalafil 5-20 mg TID) and phosphodiesterase-3 inhibitors (cilostazol 100 mg BID)
Nitroglycerin topical 1 inch q6 hr
Angiotensin-converting enzyme inhibitors/angiotensin-receptor blockers (especially for scleroderma-associated Raynaud)
Others: Hydralazine 10-50 mg QID; reserpine 01.-0.25 mg/day PO; bosentan 125 mg PO BID
Prostaglandins: Iloprost, epoprostenol, alprostadil, beraprost
Antithrombotics/anticoagulants: (aspirin 81-325 mg/day, dipyridamole 75 mg TID, heparin [via weight-based nomogram; see Table 81-4 in Chapter 81], low-molecular-weight heparin [i.e., enoxaparin 1 mg/kg q12h; see Table 38-2 in Chapter 38])
Botulinum toxin A

Pernio

Pernio, also known as chilblains, is a cold-induced vasospastic disorder that affects the skin after exposure to nonfreezing temperatures or damp climates.[16] Pernio is seen more commonly in the northern United States and northwestern Europe. Although it can occur in children and older individuals, it is most common in young women between the ages of 15 to 30 years and in individuals with a low body mass.

The cause is unknown but is likely a result of cold-induced vasoconstriction that induces inflammation and ischemia of vessels and surrounding tissue. The histopathologic findings include dermal edema, keratinocyte necrosis, and a deep dermal lymphocytic infiltrate.

Pernio can be classified as acute or chronic. Acute pernio develops a few hours after exposure, whereas chronic pernio develops after repeated exposures to nonfreezing cold or damp conditions.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Pernio occurs most frequently in late fall to early spring in wet or nonfreezing cold environments. Acute pernio is characterized by intense itching, numbness, or a burning sensation that develops shortly after exposure to cold or damp conditions and disappears within a few weeks. Pernio is generally symmetrical. It usually involves the toes and fingers and less commonly the nose, ears, cheeks, or thighs. Pernio is associated with single or multiple erythematous, brownish or purple-blue skin lesions (macules, papules, or plaques) that may progress to blisters or ulcers (Fig. 80-8).

Chronic pernio develops after repeated cold exposure and results in cyanotic papules, macules, or nodules. Patients often report a history of similar episodes that develop each year during the cold months and typically resolve with warmer temperatures.

The diagnosis is based on the history and physical examination. Patients generally have a normal arterial examination. Pulse volume recordings may reveal vasoconstriction, but capillaroscopy is usually normal. A skin biopsy may be necessary to differentiate pernio from other disorders, such as Raynaud phenomenon, frostbite, acrocyanosis, atheromatous embolization, erythema nodosum (Chapter 440), erythema induratum (Chapter 440), lupus erythematosus (Chapter 266), sarcoidosis (Chapter 95), or atherosclerosis (Chapter 79). Laboratory testing may be necessary to exclude an underlying collagen vascular disease (Chapter 256).

TREATMENT AND PROGNOSIS

Rx

Treatment may include the use of calcium channel blockers (nifedipine 20 to 60 mg/day or amlodipine 2.5 to 10 mg/day PO) to alleviate symptoms. Nifedipine may also be given in a topical gel form. Pentoxifylline (400 mg three times daily) and capsaicin have also been reported to be helpful.

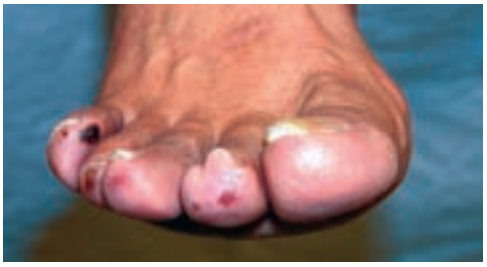

FIGURE 80-8. Pernio on the toes of the right foot. The lesions on the second, third, and fourth toes are the typical red, brown, and yellow scaling lesions. The lesion on the fifth toe can be confused with atheromatous embolization. Courtesy Dr. Jeffrey W. Olin.

PREVENTION

Rx

Patients susceptible to pernio should be advised to avoid cold exposure. If they must go outside in cold or damp weather, they should dress appropriately with layered outdoor clothing, insulated footgear, gloves, scarf, and hat.

Pernio is usually self-limiting in the acute state. Chronic pernio can lead to scarring, atrophy, and chronic occlusive vascular disease.

Frostbite

Frostbite is a local cold-induced injury (Chapter 109) that occurs when persons are exposed to temperatures below the freezing point of intact skin, or in above-freezing temperatures in association with wet environments, high altitudes, and strong winds.

Frostbite, once considered primarily a military problem, is now seen in individuals 30 to 49 years of age who participate in winter outdoor sports, are homeless, have psychiatric illness (Chapter 397), consume excess alcohol (Chapter 33), use illegal drugs (Chapter 34), or survive outdoor trauma (Chapter 111). Patients with peripheral arterial disease (Chapter 79), a smoking history (Chapter 32), younger or older age, or diabetes (Chapter 229) are also at increased risk.[17] Frostbite injury involves three pathophysiologic components: tissue injury from extracellular and intracellular ice crystal formation, intracellular dehydration, and ischemia.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The severity of frostbite relates to the absolute temperature and to the duration of exposure. The digits of the hands and feet account for most injuries, although the ears, nose, and cheeks may be affected. Patients complain of numbness or paresthesias and report clumsiness and lack of fine coordination if the hands are involved. The numbness may persist even after rewarming. The skin may be pale, waxy, and cool to touch, and the patient may have mild to extensive swelling depending on the severity of the frostbite. Large, clear, blisters often appear (Fig. 80-9), followed by black scabs. Damage to the muscles, tendons, cartilage, joints, and bones occur in severe frostbite. Pain is often severe during the rewarming process.

The diagnosis of frostbite is made by the exposure history and physical examination. The differential diagnosis includes peripheral arterial disease, pernio, trench foot, or thermal burns.

TREATMENT AND PROGNOSIS

Rx

Proper recognition of frostbite and removal from cold exposure is essential. Local rewarming should begin only if refreezing will not occur while the patient is being transferred to a hospital. Uninterrupted rapid rewarming in a water bath of between 40 and 42 degrees centigrade for 15 to 30 minutes is critical to help minimize tissue loss. Removal of wet or damp clothing is important, but rubbing or massage should be avoided. Splinting and elevation of the affected limb can help minimize swelling and improve perfusion. Patients should receive tetanus toxoid, analgesics (e.g., hydromorphone 2 to 8 mg either PO or parenterally every 3 to 4 hours as needed for pain), and broad-spectrum antibiotics if secondary tissue infection is present. Daily cleansing in a warm whirlpool bath and physical therapy are important. Clear blisters should be left alone, but ruptured blisters should be covered with a topical antibiotic, such as neomycin/bacitracin/polymyxin B ointment. A combination

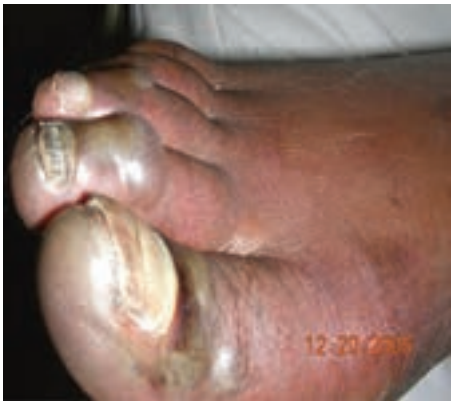

FIGURE 80-9. Blisters in a patient with frostbite.

of aspirin (250 mg), prostacyclin (0.5 to 2 ng/kg IV for 6 hours on 8 consecutive days), and tissue plasminogen activator (100 mg on day 1) has shown promise in markedly reducing the amputation rate in patients with severe frostbite, compared with aspirin plus buflomedil.[A10] If possible, surgical debridement and amputation should be avoided until complete demarcation occurs.

Patients subjected to frostbite are more susceptible to future cold injuries. They are susceptible to chronic pain, complex regional pain syndrome (Chapter 30), cold hypersensitivity, and a reduced sensitivity to touch.

Acrocyanosis

DEFINITION

Acrocyanosis is a poorly defined and often misunderstood clinical condition that manifests as painless, symmetrical, bluish or cyanotic discoloration affecting the hands, the feet, or both. Acrocyanosis occurs in primary and secondary forms. Primary acrocyanosis is generally a benign condition seen most often in young women during their second to fourth decades of life. It may be more common in cooler temperatures, and a familial predisposition has been reported. The overall incidence is not known, but its prevalence has been reported to be approximately 20 to 40% in persons with anorexia nervosa (Chapter 219) and up to 25% of all patients with cancer.

Acrocyanosis was originally thought to be a vasospastic disorder that develops when small cutaneous arteries and arterioles constrict and reduce blood flow, dilation, and oxygen desaturation in the venules. More recent data suggest that low pressures and sluggish flow result in capillary constriction.[18]

The underlying cause of primary acrocyanosis is unknown. Secondary acrocyanosis can be associated with a number of conditions, including Ehlers-Danlos syndrome (Chapter 260), hypoxemia (Chapter 104), cryoglobulins (Chapter 187), cryofibrinogens, cold agglutinins, antiphospholipid antibodies (Chapter 174), malignancy, spinal cord injury (Chapter 399), arsenic poisoning (Chapter 22), starvation, and some medications. It is also seen with the "puffy hand syndrome," a finding unique to intravenous drug abusers (Chapter 34) who inject their hands or fingers.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Persistent, painless, symmetrical bluish discoloration commonly involves the hands and feet but also can affect the forearms, nose, ears, and even nipples. Acrocyanosis may be exacerbated by cold exposure, emotional stress, or dependency of the limbs. It improves with elevation. Patients also report clamminess and hyperhidrosis of their hands and feet. Secondary acrocyanosis may be asymmetrical and associated with pain, ulceration, or tissue loss or gangrene.

The diagnosis of acrocyanosis is based on the history and physical examination. Laboratory evaluation should include a complete blood count, metabolic profile, and levels of antiphospholipid antibodies, cold agglutinins, cryofibrinogens, and cryoglobulins, as well as testing for connective tissue disorders (Chapter 256).

The differential diagnosis includes Raynaud phenomenon, pernio, erythromelalgia, and peripheral cyanosis. Acrocyanosis can be differentiated from peripheral cyanosis by the presence of cyanosis on the mucous membranes and hypoxia on an arterial blood sample.

TREATMENT AND PROGNOSIS

No treatment is necessary for primary acrocyanosis other than reassurance and avoidance of cold and damp exposures. A trial of α-adrenergic blocking agents such as prazosin may be considered. Calcium channel blockers, including amlodipine or nifedipine, are generally not helpful. Bioflavonoids and nicotinic acid derivatives have been reported beneficial in some cases, and sympathetic nerve block and sympathectomy may be tried for more severe cases. Treatment for secondary acrocyanosis depends on the underlying cause.

The prognosis of primary acrocyanosis is excellent. The prognosis of secondary acrocyanosis prognosis depends on the underlying cause.

Grade A References

A1. Fiessinger JN, Schafer M. Trial of iloprost versus aspirin treatment for critical limb ischaemia of thromboangiitis obliterans. The TAO Study. *Lancet*. 1990;335:555-557.

A2. Bozkurt AK, Koksal C, Demirbas MY, et al. A randomized trial of intravenous iloprost (a stable prostacyclin analogue) versus lumbar sympathectomy in the management of Buerger's disease. *Int Angiol*. 2006;25:162-168.

A3. Kalgaard OM, Mork C, Kvernebo K. Prostacyclin reduces symptoms and sympathetic dysfunction in erythromelalgia in a double-blind randomized pilot study. *Acta Derm Venereol*. 2003;83:442-444.

A4. Goldberg YP, Price N, Namdari R, et al. Treatment of Na(v)1.7-mediated pain in inherited erythromelalgia using a novel sodium channel blocker. *Pain*. 2012;153:80-85.

A5. Ennis H, Anderson ME, Wilkinson J, et al. Calcium channel blockers for primary Raynaud's phenomenon. *Cochrane Database Syst Rev*. 2014;1:CD002069.

A6. Huisstede BM, Hoogvliet P, Paulis WD, et al. Effectiveness of interventions for secondary Raynaud's phenomenon: a systematic review. *Arch Phys Med Rehabil*. 2011;92:1166-1180.

A7. Nguyen VA, Eisendle K, Gruber I, et al. Effect of the dual endothelin receptor antagonist bosentan on Raynaud's phenomenon secondary to systemic sclerosis: a double-blind prospective, randomized, placebo-controlled pilot study. *Rheumatology (Oxford)*. 2010;49:583-587.

A8. Roustit M, Blaise S, Allanore Y, et al. Phosphodiesterase-5 inhibitors for the treatment of secondary Raynaud's phenomenon: systematic review and meta-analysis of randomised trials. *Ann Rheum Dis*. 2013;72:1696-1699.

A9. Noaimi AA, Fadheel BM. Treatment of perniosis with oral pentoxyfylline in comparison with oral prednisolone plus topical clobetasol ointment in Iraqi patients. *Saudi Med J*. 2008;29:1762-1764.

A10. Cauchy E, Cheguillaume B, Chetaille E. A controlled trial of a prostacyclin and rt-PA in the treatment of severe frostbite. *N Engl J Med*. 2011;364:189-190.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Liu Y, Sun Y, He X, et al. Imaging diagnosis and surgical treatment of popliteal artery entrapment syndrome: a single-center experience. *Ann Vasc Surg*. 2014;28:330-337.
2. Wu X, Lun Y, Jiang H, et al. Cystic adventitial disease of the common femoral vessels: report of 2 cases and literature review. *Vasc Endovascular Surg*. 2014;48:325-328.
3. Weinberg I, Jaff MR. Nonatherosclerotic arterial disorders of the lower extremities. *Circulation*. 2012;126:213-222.
4. Olin JW, Gornik HL, Bacharach JM, et al. Fibromuscular dysplasia: state of the science and critical unanswered questions: a scientific statement from the American Heart Association. *Circulation*. 2014;129:1048-1078.
5. Olin JW, Froehlich J, Gu X, et al. The United States Registry for Fibromuscular Dysplasia: results in the first 447 patients. *Circulation*. 2012;125:3182-3190.
6. Trinquart L, Mounier-Vehier C, Sapoval M, et al. Efficacy of revascularization for renal artery stenosis caused by fibromuscular dysplasia: a systematic review and meta-analysis. *Hypertension*. 2010;56:525-532.
7. Piazza G, Creager MA. Thromboangiitis obliterans. *Circulation*. 2010;121:1858-1861.
8. Dean SM. Livedo reticularis and related disorders. *Curr Treat Options Cardiovasc Med*. 2011;13:179-191.
9. Fries C, Roos M, Gaspert A, et al. Atheroembolic disease: a frequently missed diagnosis: results of a 12-year matched-pair autopsy study. *Medicine (Baltimore)*. 2010;89:126-132.
10. Kronzon I, Saric M. Cholesterol embolization syndrome. *Circulation*. 2010;122:631-641.
11. Scolari F, Ravani P. Atheroembolic renal disease. *Lancet*. 2010;375:1650-1660.
12. Skeik N, Rooke TW, Davis MD, et al. Severe case and literature review of primary erythromelalgia: novel SCN9A gene mutation. *Vasc Med*. 2012;17:44-49.
13. Arnold AW, Itin PH. Laptop computer-induced erythema ab igne in a child and review of the literature. *Pediatrics*. 2010;126:e1227-e1230.
14. Maverakis E, Patel F, Kronenberg DG, et al. International consensus criteria for the diagnosis of Raynaud's phenomenon. *J Autoimmun*. 2014;48-49:60-65.
15. Goundry B, Bell L, Langtree M, et al. Diagnosis and management of Raynaud's phenomenon. *BMJ*. 2012;344:e289.
16. Cappel JA, Wetter DA. Clinical characteristics, etiologic associations, laboratory findings, treatment, and proposal of diagnostic criteria of pernio (chilblains) in a series of 104 patients at Mayo Clinic, 2000 to 2011. *Mayo Clin Proc*. 2014;89:207-215.
17. Zafren K. Frostbite: prevention and initial management. *High Alt Med Biol*. 2013;14:9-12.
18. Kurklinsky AK, Miller VM, Rooke TW. Acrocyanosis: the Flying Dutchman. *Vasc Med*. 2011;16:288-301.

REVIEW QUESTIONS

1. Suspicion of popliteal artery entrapment syndrome should occur in which of the following?

A. A 66-year-old woman with hypertension, tobacco use, and exertional limb discomfort
B. A 27-year-old tobacco-using man with ischemic rest pain of the right third digit and superficial thrombophlebitis
C. A 52-year-old with red discoloration of the feet with heat exposure
D. A 22-year-old male athlete who has bilateral calf claudication with long-distance cycling
E. A 31-year-old woman with hypertension and right femoral artery bruit

Answer: D Popliteal artery entrapment should be considered as a cause of intermittent claudication in young patients without evidence of atherosclerosis or other systemic inflammatory disorders.

2. Fibromuscular dysplasia is classically represented by which of the following?

A. Arterial calcification and stenosis
B. New-onset hypertension and a cervical bruit in a healthy 34-year-old woman
C. Cyanosis with ischemic rest pain of the toes
D. A rapidly expanding ascending thoracic aortic aneurysm
E. Carotid artery dissection in a 61-year-old man after a motor vehicle accident

Answer: B Fibromuscular dysplasia is often found incidentally in patients undergoing imaging tests for other reasons. Patients are more commonly younger and female, with the most common finding being hypertension.

3. The most common clinical manifestations of erythromelalgia include which of the following?

A. Presentation in late fall or early winter
B. Intense itching in the hands and feet
C. Increased warmth, erythema, and a burning sensation in the feet
D. History of trauma as the inciting event
E. Little effect on patient's quality of life

Answer: C Erythromelalgia is a cold-induced vascular disorder that predominantly causes erythema, heat, and burning in the feet.

4. What are the most common clinical characteristics of pernio?

A. Intense itching and burning pain, more commonly seen in females
B. Edema and burning pain that is exacerbated by warmer temperatures
C. May result in amputation if not recognized
D. Stasis ulceration is common
E. Anticoagulation is the treatment of choice

Answer: A Pernio, also known as chilblains, is a cold-induced vascular disorder characterized by seasonal itching, burning, and discoloration of the toes, more often in women than men. It does not routinely ulcerate, does not result in amputation, and requires only aggressive thermal protection during the winter months.

81

PERIPHERAL VENOUS DISEASE

JEFFREY S. GINSBERG

DEEP VEIN THROMBOSIS

DEFINITION

Deep vein thrombosis (DVT), which is the most important disease affecting the peripheral veins, has an estimated annual incidence of 0.1% in whites. Most pulmonary emboli (Chapter 98) arise from DVT of the legs. In fact, DVT and pulmonary embolism are usually considered different clinical manifestations of one disease, venous thromboembolism (VTE), because up to 50% of patients who present with symptomatic proximal (popliteal vein or more proximal) DVT have imaging evidence of clinically silent pulmonary emboli, whereas up to 90% of patients with proved pulmonary emboli have DVT, even though only 15% of them have leg symptoms. For the most part, the cornerstones of management of DVT and pulmonary embolism are the same—long-term (>3 months of) anticoagulation.

Superficial thrombophlebitis consists of thrombosis and inflammation of one or more superficial veins. Provided the associated thrombus has not extended into the deep veins, affected patients have a negligible risk for development of pulmonary emboli and often can be effectively managed conservatively with ice, elevation, and anti-inflammatory medication.

EPIDEMIOLOGY

In nonpregnant individuals, DVT usually originates in one of the distal veins or calf veins, where it has little or no potential to cause clinically important pulmonary emboli. The true incidence of calf vein thrombosis is not known because many affected patients remain asymptomatic while the thrombus forms and spontaneously resolves. On the basis of results of studies of *symptomatic* patients with suspected DVT, approximately 10 to 25% actually have a diagnosable DVT, of whom approximately 15% have isolated calf DVT. Approximately one fourth of these thrombi that are initially isolated to a calf vein subsequently extend into the proximal veins, usually within 1 week of manifestation, where they *then* have the potential to cause pulmonary emboli.

In pregnancy, most (~90%) thrombi occur in the deep veins of the left leg and frequently involve the ileofemoral veins but not the calf or popliteal veins. These findings suggest an anatomic predisposition to left leg ileofemoral DVT, which may be a result of compression of the left iliac vein by the fetus, an exaggeration of the "obstruction" that occurs where the right iliac artery crosses the left iliac vein, and an increase in venous webs at the left iliac vein (May-Thurner syndrome). These observations strongly suggest that most, if not all, of the increase in VTE during pregnancy is attributable to the increase in left iliac DVT.

Significant triggers of hospitalization for VTE include major surgeries, fractures, immobility, and cancer, with or without chemotherapy. From a clinical perspective, risk factors can be subdivided by duration, that is, transient and finite duration (e.g., fractured fibula treated with plaster immobilization) compared with permanent or long-term duration (e.g., congenital antithrombin deficiency, metastatic cancer), and according to the magnitude of the risk, that is, major (hip or knee replacement surgery) or minor (long-distance air travel, use of oral contraceptives). Classification of patients according to the presence or absence and type of risk factor is predictive of the risk for recurrence after a prolonged (≥3 months) course of anticoagulant therapy and provides key information that helps determine the optimal duration of anticoagulant therapy. Patients in whom DVT develops in association with a major risk factor that has resolved have a much lower risk for recurrence after a 3-month course of anticoagulants than do patients whose DVT was apparently idiopathic or associated with an ongoing risk factor. Patients whose DVT was associated with a transient minor risk factor that has resolved have an intermediate risk for recurrence.

PATHOBIOLOGY

Virchow triad of hypercoagulability, venous stasis, and injury to the vessel wall provides a model for understanding many of the risk factors that lead to the formation of thrombosis. For example, in patients who have total hip or knee replacement surgery, venous endothelial injury is caused by surgery, venous stasis resulting from perioperative immobilization, and hypercoagulability as a result of postoperative fibrinolytic shutdown. In other patients, an identifiable "thrombophilia" or "tendency to clot," such as congenital antithrombin (formerly antithrombin III) deficiency or the presence of factor V Leiden (Chapter 176),[1] combined with use of oral contraceptives results in DVT in women of childbearing age. However, a relatively high proportion of patients have unexplained DVT without "clinical" risk factors that cause endothelial damage or venous stasis or identifiable thrombophilias that cause hypercoagulability. Undoubtedly, some of these patients have yet to be determined to have thrombophilias, but the DVT currently is labeled idiopathic or unprovoked.

CLINICAL MANIFESTATIONS

The clinical features of lower extremity DVT include leg pain, tenderness, swelling (Fig. 81-1), palpable cord, discoloration (red for inflammation and purplish for venous stasis), as well as dilation and prominence of the superficial veins. These signs and symptoms are nonspecific, so accurate diagnostic imaging is required for a definitive diagnosis. In patients who present with symptoms suspicious for DVT, DVT is confirmed in only 10 to 30% of cases. Moreover, patients with relatively minor symptoms and signs of DVT may have extensive DVT with or without pulmonary embolism. Conversely, approximately one third of patients with findings highly suspicious for DVT will not have it (e.g., patients with a ruptured Baker cyst).

DIAGNOSIS

By itself, clinical diagnosis of DVT is inaccurate because no individual symptom or sign is sufficiently predictive for the diagnosis to be made or excluded. Clinical assessment can categorize patients according to their pretest probability of DVT with reasonable accuracy, but should almost never be the only test used to exclude or make a diagnosis of DVT. By combining a validated prediction rule (Table 81-1) to assess pretest probability with the results of noninvasive tests, diagnostic accuracy can be improved, thereby often limiting or eliminating the need for further investigation (Fig. 81-2).

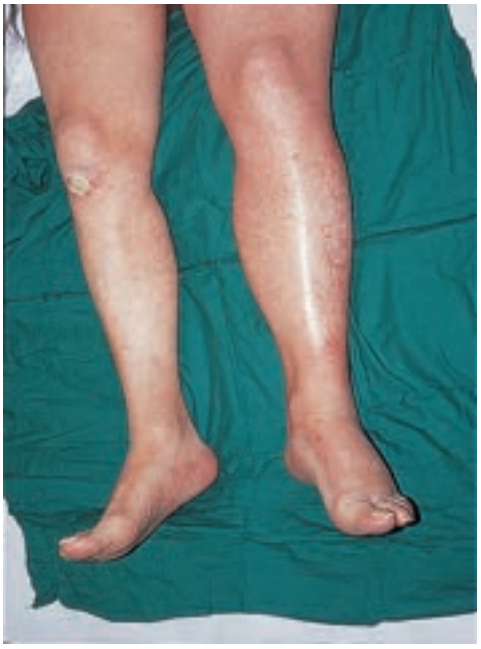

FIGURE 81-1. Deep vein thrombosis (DVT) manifesting as an acutely swollen left leg. Note the dilation of the superficial veins. The leg was hot to the touch, and palpation along the line of the left popliteal and femoral veins caused pain. Less than 50% of DVTs manifest in this way, and other conditions may mimic DVT, so further investigation is always indicated. Note the coincidental psoriatic lesion below the patient's right knee. (From Forbes CD, Jackson WF. Color Atlas and Text of Clinical Medicine, 3rd ed. London: Mosby; 2003.)

TABLE 81-1 PREDICTION RULE FOR DEEP VEIN THROMBOSIS

CLINICAL CHARACTERISTIC	SCORE*
Active cancer (treatment ongoing within previous 6 mo or palliative)	1
Paralysis, paresis, or recent plaster immobilization of the lower extremities	1
Recent bedrest of >3 days or major surgery within 3 mo requiring anesthesia	1
Localized tenderness of the deep veins of the leg	1
Entire leg swollen	1
Calf swelling of >3 cm larger than asymptomatic side measured 10 cm below tibial tuberosity	1
Pitting edema confined to the symptomatic leg	1
Collateral superficial veins (not varicosed)	1
Previously documented deep vein thrombosis	1
Alternative diagnosis as likely as or more likely than deep vein thrombosis	–2

*A score of 0 or less indicates low probability, 1 or 2 indicates moderate probability, and 3 or more indicates high probability.
Modified from Wells PS, Anderson DR, Bormanis J, et al. Value of assessment of pretest probability of deep-vein thrombosis in clinical management. *Lancet.* 1997;350:1795-1798.

Imaging

Compression Ultrasonography

Compression venous ultrasonography with or without Doppler imaging is the most widely used noninvasive test for suspected DVT because of its accuracy in detection of thrombus involving the popliteal or more proximal veins.[2] Noncompressibility (Fig. 81-3) of the proximal leg veins on

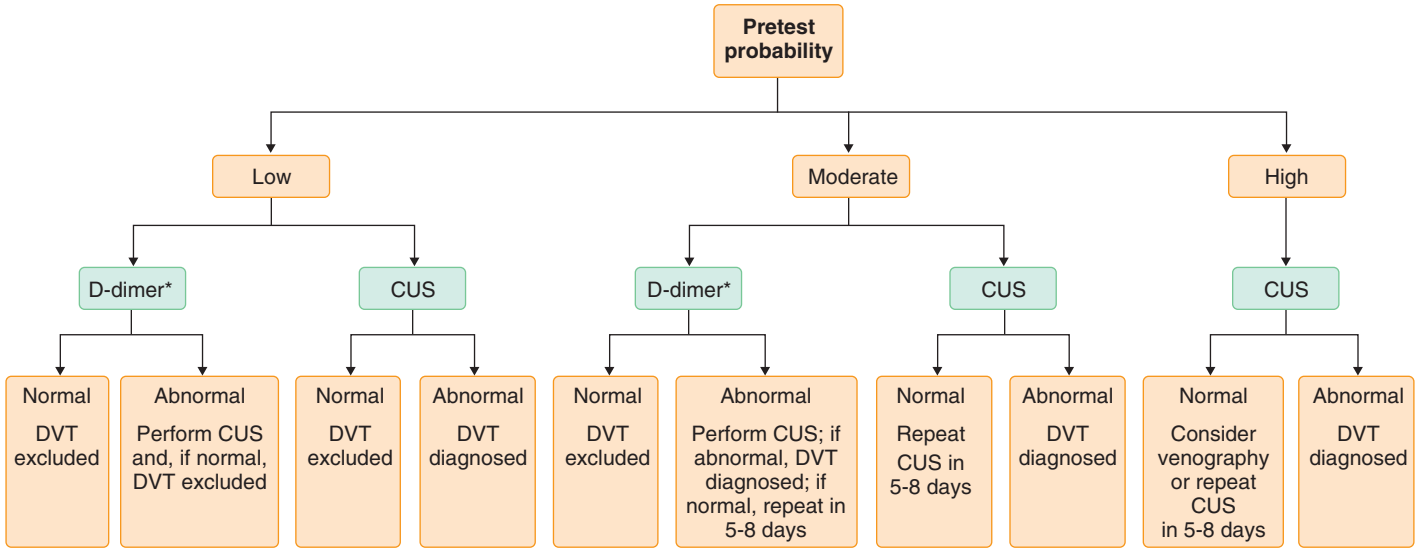

FIGURE 81-2. Diagnostic algorithm for suspected deep vein thrombosis. This algorithm uses evaluation of pretest probability based on a clinical prediction rule (see Table 81-1) and D-dimer testing to complement compression ultrasonography (CUS). The *asterisk* indicates use of a highly sensitive (>95%) D-dimer.

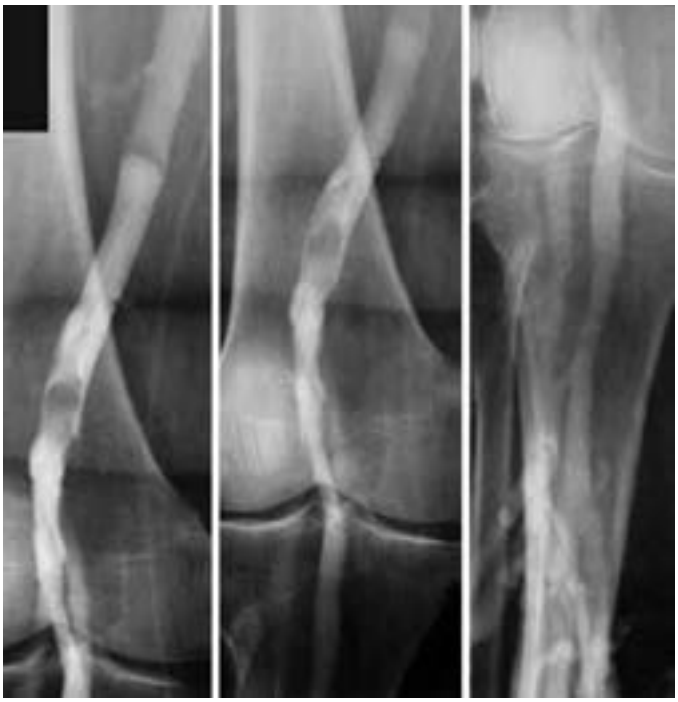

FIGURE 81-3. Abnormal venogram demonstrates a persistent (two or more different views) intraluminal filling defect in the popliteal vein.

ultrasonography is diagnostic of DVT in symptomatic patients and is an indication for treatment. Of patients with symptoms suggestive of DVT but with normal findings on initial ultrasound examination of the proximal veins, approximately 15% will have undetected isolated calf DVT; progression into the proximal veins occurs in a minority of patients, usually within a week of presentation. Isolated calf DVT that does not extend into the proximal veins is rarely if ever associated with clinically important pulmonary embolus. The sensitivity of ultrasonography for calf DVT is well below 90%, with a wide range of accuracies reported for different populations of patients.

Imaging of the calf veins is time-consuming and potentially inaccurate. Rather, two-point (common femoral and popliteal) or three-point (two-point plus the calf "trifurcation") compression ultrasonography should be performed. If two-point compression is normal, the test should be repeated about 1 week after the initial examination. This approach will identify the 20 to 25% of patients who have had proximal extension of distal clot in the calf veins. If the repeated ultrasound examination 1 week later also is normal, further investigation and therapy can be safely withheld. In centers with highly skilled operators, one normal ultrasound of the proximal veins and the calf veins near the popliteal vein at presentation is sufficiently accurate to exclude clinically important DVT and eliminate the need for follow-up testing. In patients with either a normal D-dimer test result or a low clinical pretest probability, normal two-point compression ultrasonography excludes DVT.

Magnetic Resonance Venography

Magnetic resonance venography (MRV), which uses the difference in magnetic resonance signals between flowing blood and stationary clot, has a high sensitivity and specificity for proximal DVT. Recent interest has focused on magnetic resonance for direct imaging of the thrombus because a thrombus produces a positive image without the use of contrast material, owing to its methemoglobin content. Although MRV is accurate in diagnosing and excluding DVT, it is expensive and not readily available in most centers outside of the United States.

Contrast Venography

Ascending contrast venography remains the gold standard for diagnosis, but because of its expense, discomfort to the patient, and potential for adverse experiences, venography is currently indicated in symptomatic patients only when diagnostic uncertainty persists after noninvasive testing or if noninvasive testing is unavailable. A constant intraluminal filling defect is diagnostic of acute thrombosis (Fig. 81-4), and DVT can be excluded in patients who have a normal, adequately performed venogram. Minor side effects of local pain, nausea, and vomiting are not uncommon, whereas more serious adverse reactions, such as anaphylaxis or other allergic manifestations, are rare. Venography also can induce DVT.

Laboratory Findings

D-Dimer

D-Dimer is a plasma protein specifically produced after lysis of cross-linked fibrin by plasmin. Levels are almost invariably elevated in the presence of acute VTE, so measurement of D-dimer levels is a sensitive test for recent DVT and pulmonary embolism. Unfortunately, numerous nonthrombotic conditions, including sepsis, pregnancy, surgery, and cardiac or renal failure, also can cause elevated levels. As a result of this nonspecificity, the role of D-dimer assays is limited to helping exclude VTE when levels are not raised.

Laboratory tests for D-dimer use enzyme-linked immunosorbent assay or agglutination techniques, both involving specific monoclonal antibodies. Sensitivity and cut points vary among assays, so results cannot be generalized. Highly sensitive tests, consisting of new rapid ELISA or immunoturbidimetric assays, have sensitivities of 95 to 100% for acute VTE but in general have low specificities (20 to 50%). Highly sensitive D-dimer assays can be

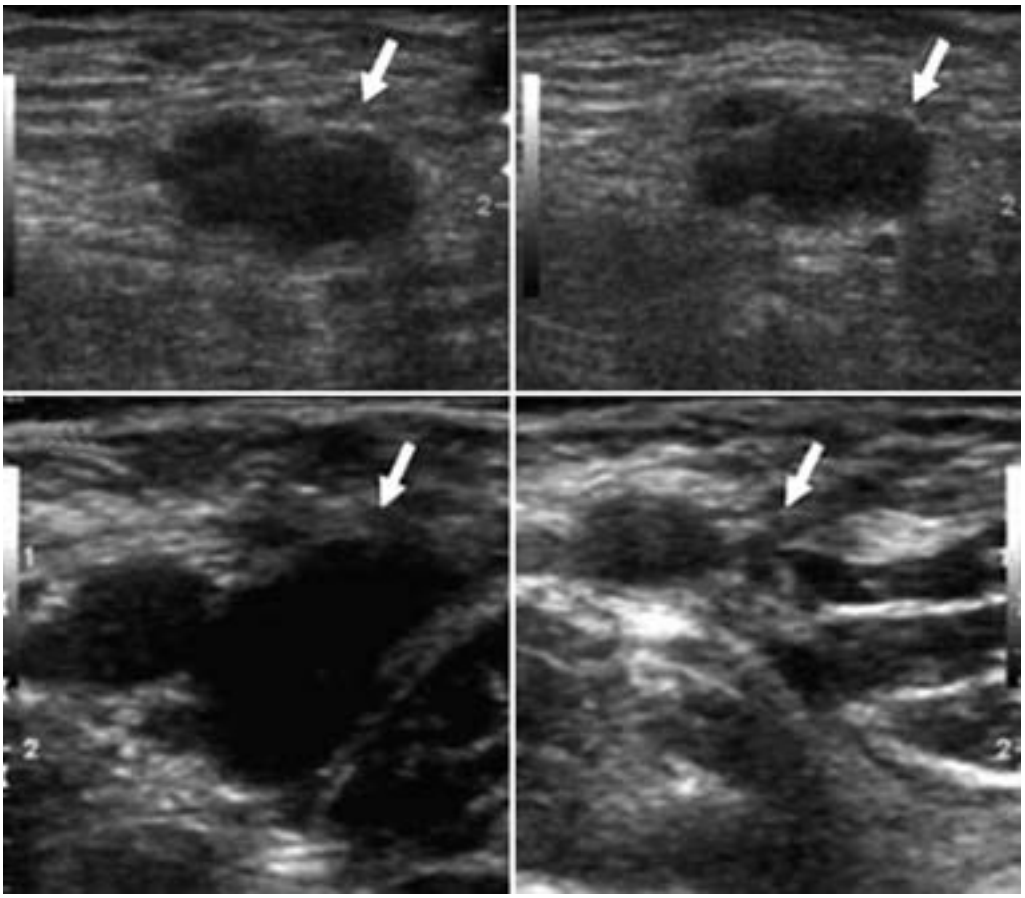

FIGURE 81-4. Compression venous ultrasonography demonstrates thrombosis of the popliteal vein. The sonograms in the *top row* demonstrate examination without *(left side)* and with *(right side)* gentle probe compression of the skin overlying the popliteal vein. The lack of compressibility is diagnostic of deep vein thrombosis. The *bottom row* shows analogous views of the femoral vein, which shows partial compressibility.

employed as stand-alone tests for exclusion of DVT, but clinicians must be aware of the accuracy of the assay in their institution before using the D-dimer assay to make management decisions. D-Dimer measured after a 3-month (or longer) initial treatment with warfarin also appears to be predictive of recurrent DVT. In addition, an elevated D-dimer level 1 month after stopping warfarin predicts a clinically and statistically significant higher recurrence rate than is seen in patients in whom the D-dimer levels were normal or low.

Algorithms for Diagnosis of Deep Venous Thrombosis and Their Risk for Recurrence

A number of diagnostic algorithms have been tested in prospective management trials (see Fig. 81-2).

Clinical Assessment and Venous Ultrasonography

It is safe to perform only a single ultrasound examination in patients with a low pretest probability by a validated clinical prediction rule (Table 81-2). Other patients require serial ultrasonographic testing if only clinical assessment and ultrasonography are used. Venography should be considered in patients with a high pretest probability and normal compression ultrasonography because the probability of DVT is still approximately 20% in such patients.

Clinical Assessment, D-Dimer Testing, and Venous Ultrasonography

Diagnostic imaging and treatment can be safely withheld in patients who have (1) a low pretest probability based on a validated clinical prediction rule and a negative value on a moderately sensitive D-dimer assay or (2) a low or intermediate pretest probability and a negative value on a highly sensitive D-dimer assay. Patients with a high pretest probability require ultrasonography regardless of the D-dimer result. A normal D-dimer result with use of either a moderately or highly sensitive assay can safely obviate the need for repeated imaging in patients with normal findings on the initial ultrasound examination. Algorithms for predicting the recurrence of an initially unprovoked DVT after the cessation of anticoagulant therapy are undergoing validation in prospective trials.

TABLE 81-2 ALTERNATIVE DIAGNOSES IN 87 CONSECUTIVE PATIENTS WITH CLINICALLY SUSPECTED VENOUS THROMBOSIS AND NORMAL VENOGRAMS*

DIAGNOSIS	PATIENTS (%)
Muscle strain	24
Direct twisting injury to the leg	10
Leg swelling in paralyzed limb	9
Lymphangitis, lymphatic obstruction	7
Venous reflux	7
Muscle tear	6
Baker cyst	5
Cellulitis	3
Internal abnormality of the knee	2
Unknown	26

*The diagnosis was made once venous thrombosis was excluded by venography.

Differential Diagnosis

A number of conditions can mimic DVT (see Table 81-2), but DVT often can be excluded only by accurate diagnostic testing. In some patients, however, the cause of pain, tenderness, and swelling remains uncertain.

Suspected Recurrent Deep Venous Thrombosis

Approximately 10% of patients with unprovoked VTE will experience recurrent thromboembolism in the first year after ceasing anticoagulant therapy. In addition, many patients will have positional leg swelling and pain early during treatment as a result of venous outflow obstruction or later (≥6 months after diagnosis) because of the post-thrombotic syndrome after endogenous thombolysis has maximized removal of the thrombus and venous valvular incompetence manifests. These and other nonthrombotic

disorders can produce symptoms that are similar to those of acute recurrent DVT, so accurate diagnostic testing to confirm recurrence is mandatory. However, residual venous abnormalities are common after an initial event; persistent abnormalities are seen on compression ultrasonography in approximately 80% of patients at 3 months and 50% of patients at 1 year after a documented proximal DVT. Therefore comparison with previous ultrasound images is required in patients with suspected recurrence. Although an increase in diameter of 4 mm or more in the compressed vein strongly suggests recurrent DVT, a new noncompressible proximal venous segment is the most reliable criterion for the diagnosis of recurrence. When compression ultrasonography is inconclusive, venography should be considered; a new intraluminal filling defect is diagnostic of acute DVT, and the absence of a filling defect excludes the diagnosis. Nonfilling of venous segments may mask recurrent DVT and is considered a nondiagnostic finding. A normal D-dimer test result is useful in excluding recurrent DVT.

Pregnancy

Symptoms of leg pain or swelling, shortness of breath, and atypical chest pain are common during pregnancy, so objective testing is needed to diagnose VTE. As in nonpregnant patients, compression ultrasonography is the initial test of choice. A normal D-dimer test is also reassuring in excluding DVT. Because isolated iliac and iliofemoral DVT is more common in pregnancy and has the potential to be missed by ultrasonography, efforts should be made to image the iliac veins to detect such thrombi. MRV, which is sensitive for pelvic DVT, may be useful when the clinical suspicion is high or if Doppler imaging of the iliac vein is inconclusive.

TREATMENT

Rx

The large majority of patients with acute DVT can now be treated on an outpatient basis, regardless of their treatment regimen (Fig. 81-5).[3] The principal indications for admission are clinical instability, the inability to adhere to outpatient therapy, or the need to use intravenous heparin for extensive iliofemoral thrombosis.

Initial Treatment

Low-molecular-weight heparin (LMWH) preparations (Chapter 38) are administered subcutaneously using weight-based dosing to provide reliable outpatient management of DVT without the need for routine laboratory monitoring. Dosage regimens differ for the various LMWH formulations (Table 81-3), but once-daily administration of LMWH is thought to be as safe and effective as twice-daily administration.[4]

Anti–factor Xa monitoring should be considered for three populations of patients: (1) patients with renal insufficiency (calculated creatinine clearance of less than 30 mL/minute); (2) obese patients, in whom the volume of distribution of LMWH might be different, so weight-adjusted dosing might not be appropriate; and (3) pregnant women, in whom it is unclear whether the dose should be adjusted according to the woman's weight change. Levels are usually determined on blood samples drawn 4 hours after subcutaneous injection; therapeutic ranges of 0.6 to 1.0 U/mL for twice-daily administration and 1.0 to 2.0 U/mL for once-daily treatment have been proposed.

Fixed-dose subcutaneous injection of LMWH is at least as effective and safe as adjusted-dose intravenous administration of unfractionated heparin for the treatment of acute DVT, with a trend toward a significant difference in mortality benefit favoring LMWH, probably because of improved survival in patients with malignant disease.

However, patients with extensive iliofemoral DVT have often been excluded from trials of LMWH, and extended-duration (i.e., >5 days) intravenous unfractionated heparin therapy is often administered to such patients. Unfractionated heparin is usually administered by continuous intravenous infusion (Table 81-4), with either fixed initial dosing or dosing according to a patient's weight, results in more rapid achievement of therapeutic activated partial thromboplastin time (aPTT) levels. The initial aPTT level should be measured 6 hours after therapy is commenced. Up to 25% of patients with acute VTE have resistance to heparin, defined as a requirement for greater than expected doses of unfractionated heparin to achieve a "therapeutic" aPTT. If it is available, anti–factor Xa monitoring is recommended in patients with heparin resistance.

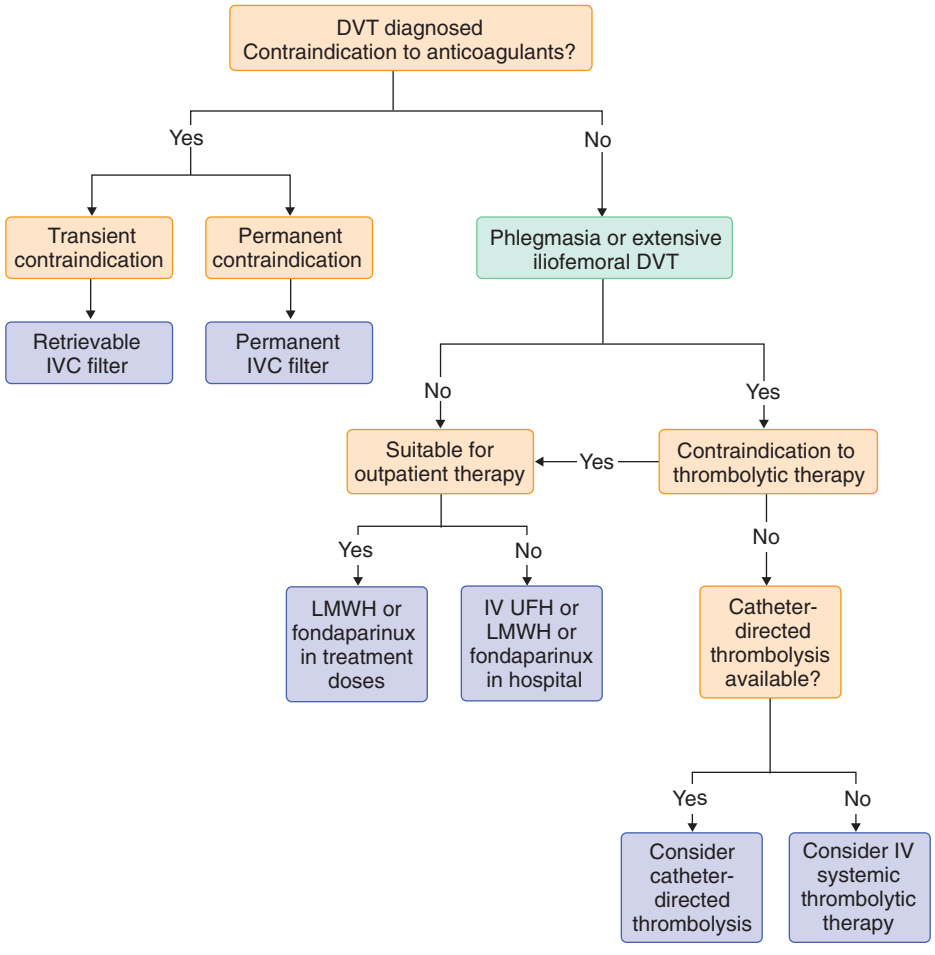

FIGURE 81-5. Guidelines for treatment of deep vein thrombosis (DVT). IVC = inferior vena cava; IV = intravenous; LMWH = low-molecular-weight heparin; UFH = unfractionated heparin.

TABLE 81-3 GUIDELINES FOR ANTICOAGULATION WITH LOW-MOLECULAR-WEIGHT HEPARIN AND FONDAPARINUX

INDICATIONS	GUIDELINES
VTE suspected	Obtain baseline aPTT, PT, CBC Check for contraindication to heparin therapy Order imaging study; consider giving IV unfractionated heparin (5000 IU) or LMWH
VTE confirmed	Give LMWH (dalteparin,* enoxaparin,† nadroparin,‡ tinzaparin,§ fondaparinux¶) Start warfarin therapy on day 1 at 5 mg and adjust the subsequent daily dose according to INR Check platelet count between days 3 and 5 Stop LMWH therapy after at least 4 or 5 days of combined therapy when the INR is > 2 Anticoagulate with warfarin for at least 3 months at an INR of 2.5, range of 2-3 (See text for alternatives to warfarin: "Oral Direct Thrombin and Factor Xa Inhibitors")

Modified from Hyers TM, Agnelli G, Hull RD, et al. Antithrombotic therapy for venous thromboembolic disease. *Chest.* 2001;119:176S-193S.
*Dalteparin sodium, 200 anti-Xa IU/kg/day SC. A single dose should not exceed 18,000 IU (approved in Canada).
†Enoxaparin sodium, 1 mg/kg q12h SC, or enoxaparin sodium, 1.5 mg/kg/day SC. A single daily dose should not exceed 180 mg (approved in both the United States and Canada).
‡Nadroparin calcium, 86 anti-Xa IU/kg two times daily SC for 10 days (approved in Canada), or nadroparin calcium, 171 anti-Xa IU/kg SC daily. A single dose should not exceed 17,100 anti-Xa IU.
§Tinzaparin sodium, 175 anti-Xa IU kg/day SC daily (approved in Canada and the United States).
¶Fondaparinux according to weight: <50 kg, 5 mg /day SC; 50-100 kg, 7.5 mg SC; and > 100 kg, 10 mg SC.
aPTT = activated partial thromboplastin time; CBC = complete blood count; INR = international normalized ratio; LMWH = low-molecular-weight heparin; PT = prothrombin time; VTE = venous thromboembolism.

TABLE 81-4 WEIGHT-BASED NOMOGRAM FOR INITIAL INTRAVENOUS HEPARIN THERAPY

aPTT	DOSE (IU/kg)
Initial dose	80 bolus, then 18/hr
<35 sec (<1.2×)*	80 bolus, then 4/hr
35-45 sec (1.2-1.5×)	40 bolus, then 2/hr
46-70 sec (1.5-2.3×)	No change
71-90 sec (2.3-3×)	Decrease infusion rate by 2/hr
>90 sec (>3×)	Hold infusion 1 hr, then decrease infusion rate by 3/hr

Modified from Raschke RA, Reilly BM, Guidry JR, et al. The weight-based heparin dosing nomogram compared with a "standard care" nomogram: a randomized controlled trial. Ann Intern Med. 1993;119:874-881.
*Figures in parentheses show comparison with control.
aPTT = activated partial thromboplastin time. In general, with contemporary aPTT reagents, the target therapeutic range is more than 1.2 to 2.3 times control.

Fondaparinux

Fondaparinux is a synthetic analogue of the critical pentasaccharide sequence required for binding of heparin molecules to antithrombin (Chapter 38). Given subcutaneously, fondaparinux demonstrates 100% bioavailability, with peak plasma concentrations occurring 1.7 hours after dosing. Once-daily subcutaneous administration of fondaparinux (5 mg/day if weight is < 50 kg; 7.5 mg/day if weight is 50 to 100 kg; 10 mg/day if weight is > 100 kg) is an effective and safe alternative to LMWH for the initial 5 to 10 days of treatment of DVT. Clearance is predominantly renal, with approximately 70% of the initial dose recovered in the urine in an unchanged form. Patients with reduced creatinine clearance, such as elderly patients, have higher peak drug levels and longer drug half-life, so their dose may need to be adjusted downward.

Transition to Oral Treatment: Coumarin Derivatives (Warfarin)

Warfarin is a vitamin K antagonist that inhibits the production of clotting factors II (prothrombin), VII, IX, and X, as well as the naturally occurring anticoagulants protein C and protein S. In patients with DVT, the drug should be started within 24 to 48 hours of initiation of heparin with a goal of achieving international normalized ratio (INR) results between 2.0 and 3.0 (Chapter 38). A higher target INR of 3.0 to 4.0 is associated with more bleeding but no better efficacy, even in patients with the antiphospholipid antibody syndrome (Chapter 176), and lower intensity warfarin therapy (target INR, 1.5 to 1.9) is significantly less effective at preventing recurrent VTE, despite similar rates of major bleeding.

The dose is empirical, but a starting dose of 5 to 10 mg is suitable for most patients. Warfarin doses are adjusted according to the prothrombin time, expressed as the INR, performed daily or every other day until the results are in the therapeutic range for at least 24 hours. After initial dosing, warfarin can be monitored two or three times per week for 1 to 2 weeks and then less frequently, depending on the stability of INR results, up to intervals as long as 4 to 6 weeks. If dose adjustment is needed, such as when medications that can interact with warfarin are introduced, the cycle of more frequent monitoring is repeated until a stable dose response is again achieved.

It is now clear that pharmacogenetics have a large impact on the relatively wide range of warfarin dose requirements among different populations and the variability of warfarin requirements over time in any individual patient.[5] Polymorphisms in the gene encoding cytochrome P-450 2C9 enzyme, the enzyme that primarily clears the S-enantiomer of warfarin, contribute to variable responses to warfarin. Vitamin K epoxide reductase (VKORC1) recycles vitamin K epoxide to the reduced form of vitamin K and is the target of warfarin. Genotyping for *CYP2C9*2*, *CYP2C9*3*, *VKORC1* can help guide warfarin dosing and increase the amount of time patients are in the the therapeutic INR range. Polymorphisms are associated with a need for lower doses of warfarin during long-term therapy. Routine pharmacogenetic testing may ultimately be recommended in candidates for long-term (>3 months) warfarin therapy to identify individuals who are likely to require higher or lower warfarin doses.

Long-Term Treatment

The preferred long-term treatment of DVT for most patients is warfarin or another coumarin derivative (e.g., acenocoumarol), continued until the benefits of treatment for reducing recurrent VTE no longer outweigh its risks for major bleeding. The decision to prolong or to stop anticoagulation should be individualized, and a patient's preferences should be considered.[6]

Patients with symptomatic proximal DVT or pulmonary emboli should be treated for at least 3 months, even if the VTE was associated with a transient risk factor, but the optimal duration of treatment for patients whose VTE is not associated with a transient risk factor is controversial. Three months of treatment is associated with a 10 to 27% risk for a recurrence during the 12 months after anticoagulant therapy is stopped, whereas 6 months of anticoagulant therapy reduces the risk for recurrence in the first year after stopping to approximately 10%. In patients whose VTE developed in association with minor risk factors (e.g., air travel, pregnancy, within 6 weeks of estrogen therapy, after leg injury or immobilization), the risk for recurrence is probably lower than 10%. Continuation of treatment beyond 6 months reduces the risk for recurrent VTE during the course of therapy, but the benefit is lost after warfarin is discontinued.

Current guidelines recommend 3 months of therapy for a first proximal DVT, pulmonary embolism, or both that is provoked by surgery or by a nonsurgical risk factor. For unprovoked VTE, the recommendation is also 3 months if the bleeding risk is high but extended therapy if the bleeding risk is low or moderate. For patients VTE associated with active cancer, extended therapy is recommended using LMWH (see later) rather than warfarin.

The most convincing association of thrombophilia with the risk for recurrent VTE is the antiphospholipid antibody (lupus anticoagulant or anticardiolipin antibody [Chapter 176]), which is associated with a two-fold increase in the risk for recurrence. Homozygous factor V Leiden, and deficiencies of antithrombin, protein C, and protein S also have been associated with an increased risk for recurrence in some reports, but other data suggest that testing for heritable thrombophilia does not predict recurrent VTE in the first 2 years after anticoagulant therapy is stopped. In the absence of randomized trials to assess different durations of anticoagulation in patients with VTE and thrombophilia, routine testing for thrombophilias need not be performed but should be considered in young (<50 years) patients, patients with venous thrombosis in unusual sites, and patients with a strong family history of VTE (i.e., one or more first-degree relatives with a history of VTE).

The decision to extend anticoagulant therapy beyond 3 months must balance the risk for recurrent VTE with the risk for bleeding. The annual risk for major bleeding when warfarin is adjusted to achieve a target INR of 2.0 to 3.0 is 1 to 3%, with a case-fatality rate of 10% when major bleeding occurs in patients who received treatment for more than 3 months. By comparison, the case-fatality rate for recurrent VTE is approximately 5%. In patients whose VTE was associated with a transient risk factor or who are at high risk for bleeding, treatment for 3 months is generally adequate because the risk for fatal recurrent VTE is lower than the risk for fatal bleeding if warfarin treatment is prolonged. Among patients without a reversible or transient cause, however, prolonged warfarin therapy for more than 6 months can be considered because the risk for fatal hemorrhage is counterbalanced by the risk for fatal recurrence. The argument to prolong therapy is stronger in patients with high-risk thrombophilia (e.g., homozygous factor V Leiden; antiphospholipid antibody; deficiency of antithrombin, protein C, or protein S; or combined heterozygous state

for factor V Leiden and the prothrombin gene mutation). Indefinite therapy (preferably with LMWH) should be considered in patients with cancer-related VTE (Chapter 179) if the risk for bleeding is not high because the risk for recurrent VTE is more than 10% in the first year after anticoagulation is stopped. In motivated and capable patients, self-management of warfarin therapy is better than management by a physician or nurse.

Alternatives to Coumarin Derivatives

For patients in whom warfarin is impractical or contraindicated and for those who have recurrent VTE while being treated with appropriate doses of oral anticoagulants, therapeutic doses of LMWH are as effective as warfarin. For patients with cancer-related VTE (DVT, pulmonary embolus, or both), weight-based LMWH that is decreased to 75% of the initial dose after 1 month of treatment reduces the risk for recurrent VTE compared with warfarin, with similar bleeding rates. For patients who have had unprovoked VTE and who have discontinued anticoagulant treatment, aspirin reduces the risk for recurrent VTE with no apparent increase in the risk for major bleeding.

Oral Direct Thrombin and Factor Xa Inhibitors

Two stiochiometric or direct oral factor Xa inhibitors (rivaroxaban and apixaban) and one direct oral thrombin inhibitor (dabigatran) have been extensively evaluated (and approved by the U.S. Food and Drug Administration [FDA] and various regulatory agencies) for the prevention of stroke and systemic embolism in patients with nonvalvular atrial fibrillation (Chapter 64), and these agents also have been evaluated for the treatment of acute VTE and its secondary prevention. Currently, however, only rivaroxaban had been approved by the FDA for VTE treatment. Rivaroxaban, 15 mg twice daily for 3 weeks followed by 20 mg/day, is not inferior to enoxaparin followed by warfarin for the first 6 months of therapy and is more efficacious (recurrent VTE rates of 1.3% vs. 7.1%) than placebo when used for extended therapy for 6 to 12 months after an idiopathic DVT.

Apixaban 10 mg twice daily for 7 days followed by 5 mg twice daily for 6 months is as efficacious but safer than standard therapy with enoxaparin followed by warfarin for the treatment of acute DVT. It also is more efficacious than placebo without a significantly increased bleeding risk when used at 2.5 mg or 5 mg twice daily or placebo for another 12 months in patients who have completed 6 to 12 months of antigoagulant therapy.

In patients who have completed initial therapy with LMWH or unfractionated heparin, extended dabigatran (150 mg twice daily) is not inferior to warfarin for both efficacy and safety. Continued dabigatran at 150 mg twice daily after at least 3 months of anticoagulant therapy is as effective and safer than warfarin but causes more bleeding than placebo.

These agents are a reasonable alternative to LMWH followed by warfarin for the treatment of VTE, but their precise role has yet to be determined. They offer the advantage of fixed dosing, without a need for monitoring. Whether any are truly more effective or safer than warfarin in clinical practice and whether the lack of monitoring overrides the increase in cost remain to be determined. Another challenge is the absence of an easy antidote when patients have major bleeding (Chapter 38).[7]

Thrombolytic Therapy

Although thrombolytic therapy results in increased rates of early patency of leg veins after DVT, it has not been conclusively shown to decrease the subsequent rate of post-thrombotic syndrome or pulmonary emboli. Except for patients who have life-threatening limb ischemia as a result of massive thrombosis, thrombolysis is not recommended in patients with DVT. Ongoing trials of catheter-directed thrombolysis versus "standard therapy" should provide the definitive evidence about the relative efficacy of catheter-directed thrombolysis for the prevention of post-thrombotic syndrome.

Side Effects of Anticoagulants

Bleeding is the most common side effect of anticoagulant therapy. Major bleeding (e.g., intracranial [Chapter 408], gastrointestinal [Chapter 135], or retroperitoneal) leading to hospitalization, transfusion, or death occurs in approximately 2% of patients treated with intravenous unfractionated heparin for acute VTE. Factors such as recent surgery, trauma, and concurrent aspirin or thrombolytic therapy increase the risk for bleeding.

The risk for major bleeding with warfarin in doses adjusted to achieve a target INR of 2.0 to 3.0 ranges from 1 to 3% per year and appears to be highest soon after treatment is started or if anticoagulation is difficult to control. Risks are somewhat lower for direct thrombin and factor Xa inhibitors, but their actions cannot yet be reversed pharmaceutically (Chapter 38). The risk for major bleeding increases according to individual characteristics, such as older age, the presence of comorbid conditions (e.g., diabetes, hypertension, renal insufficiency, previous gastrointestinal bleeding, or cancer) and the use of concomitant drugs, in particular antiplatelet therapy.

Heparin-induced thrombocytopenia, which is a relatively common nonhemorrhagic complication of therapy with unfractionated heparin and a very uncommon complication of LMWH, is manifested typically with thrombocytopenia and new thrombosis (Chapter 38). Monitoring of the platelet count is recommended every other day until day 14 in patients receiving therapeutic unfractionated heparin but is not routinely recommended with LMWH or fondaparinux because of the extremely low risk with these newer medications.

When Medications Fail or Are Contraindicated

Therapeutic strategies to manage patients in whom symptomatic VTE recurs while they are receiving conventional-intensity warfarin or direct thrombin or factor Xa inhibitors include LMWH, higher-intensity warfarin (e.g., INR range of 3.0 to 4.0), and insertion of a vena caval filter. However, the optimal management of such patients is unknown because no randomized studies have been performed.

Inferior vena caval filters should be used in patients who have contraindications to anticoagulant therapy or develop major bleeding while receiving it, as well as in patients who develop recurrent VTE while receiving appropriate anticoagulation. Retrievable or removable inferior vena caval filters can be retrieved and removed within 14 days to several weeks after insertion or can be left in permanently. These filters are ideal for a patient who has a reversible cause of, or the potential for, major bleeding (e.g., DVT after craniotomy, DVT late in pregnancy).

Deep Vein Thrombosis in Pregnancy

The management of pregnant women with DVT (Chapter 239) is problematic because all coumarin derivatives cross the placenta and have the potential to cause warfarin embryopathy, consisting of nasal hypoplasia and epiphyseal stippling, if the newborn is exposed to warfarin between 6 and 12 weeks of gestation. Consequently, parenteral unfractionated heparin and LMWH, which do not cross the placenta and are safe for the fetus, are the agents of choice. The easiest approach is to initiate therapy with weight-adjusted "treatment" doses of LMWH (see Table 81-3), continued for the duration of the pregnancy. Although not proved, it is likely that the dose of LMWH can be safely decreased to approximately 80% of the therapeutic dose after 3 months of therapy. As pregnancy progresses, women normally gain weight and generally require higher doses of LMWH to achieve an anti–factor Xa level similar to that achieved at the time of diagnosis. The adequacy of the dose can be assessed by measuring a 4-hour postinjection anti–factor Xa level and targeting the dose to achieve a level of 0.5 to 1.0 U/mL for twice-daily LMWH and 0.8 to 1.5 U/mL for once-daily LMWH. Alternatively, the dose of LMWH can simply be adjusted periodically on the basis of the woman's weight.

Unfractionated heparin is less attractive than LMWH because it is associated with a greater reduction of bone density and a higher risk for heparin-induced thrombocytopenia. Unfractionated heparin can be initiated either by continuous intravenous infusion in doses adjusted to maintain an aPTT in the therapeutic range, followed by 12-hourly subcutaneous injections, or simply with 12-hourly subcutaneous injections throughout the course of pregnancy. The dose should be adjusted to target a mid-interval (6-hours after) aPTT in the therapeutic range.

Pregnant women with a DVT should probably be treated for the duration of pregnancy and for at least 6 weeks postpartum. If the DVT occurred early in pregnancy, elective induction of delivery at approximately 37 weeks with discontinuation of the heparin 24 hours earlier is recommended. If the DVT occurs in the latter part of the third trimester, intravenous heparin should be administered by continuous infusion until approximately 6 hours before the expected time of delivery. Intravenous unfractionated heparin or subcutaneous LMWH should be started postpartum as soon as hemostasis has been achieved. Maternal warfarin therapy is safe for the breast-fed infant because warfarin and its metabolites are not secreted into breast milk in doses sufficient to cause an anticoagulant effect. Consequently, warfarin (with bridging LMWH or unfractionated heparin until the INR is 2.0 or higher) can be used after delivery.

PREVENTION

Despite the plethora of large randomized trials demonstrating the efficacy and safety of mechanical and pharmacologic measures in reducing the risk for VTE in a wide range of hospitalized populations of patients, prophylaxis remains grossly underused. Factors that increase the risk for DVT include surgery (particularly major hip and knee surgery, as well as neurosurgery [Chapters 431 and 433]), major trauma (Chapter 111), prolonged bedrest or immobilization, previous episodes of VTE, presence of malignant disease, paralysis, morbid obesity, and increasing age.

Comprehensive consensus guidelines have been developed for the prevention of VTE in different populations of patients (Chapter 38). In general, mechanical prophylaxis (antiembolic stockings and intermittent pneumatic compression) should be used as an adjunct to pharmacologic prophylaxis or in patients with a high risk for bleeding. For general medical patients admitted to the hospital with a major illness and in whom mobility is likely to be reduced for 72 hours or longer, low-dose unfractionated heparin or LMWH (see Table 38-2 in Chapter 38) should be considered. In patients who

undergo major hip or knee surgery, warfarin (to an INR of 2.0 to 3.0), subcutaneous LMWH, subcutaneous fondaparinux (2.5 mg/day), or oral rivaroxaban (10 mg/day) should be used for at least 7 to 14 days postoperatively. In patients with continued immobility, prophylaxis should be considered until the patient regains preoperative mobility. In one randomized trial, the ultra-low-molecular-weight heparin semuloparin (20 mg/day SC) reduced the risk for VTE from 3.4% to 1.2% over 3.5 months without increasing major bleeding.

VENOUS THROMBOSIS OF THE UPPER EXTREMITIES

DVT of the upper extremities (including the arm and the axillary, subclavian, and internal jugular veins, as well as the superior vena cava) is much less common than DVT of the legs, but it is not rare,[8] especially among critically ill adults in intensive care units.[9] Factors associated with upper extremity DVT include central venous catheters, acquired or hereditary thrombophilias, and anatomic (cervical rib) and physiologic (muscular individuals) impingement of the vein. The incidence of clinically important post-thrombotic syndrome is not high if patients are treated with anticoagulants alone.

Contrast venography is the gold standard for the diagnosis of upper extremity DVT, but venous ultrasonography is accurate and less invasive. Because it is not feasible to test for compression of the subclavian vein, a diagnosis of subclavian DVT by ultrasonography is based on flow abnormalities or direct visualization of thrombus by B-mode ultrasonography. Upper extremity DVT can cause pulmonary emboli, although the exact frequency is not known.

Considerable controversy exists about the management of patients in whom DVT develops in association with a central venous catheter. If the line is not necessary or is nonfunctional, some recommend simply removing the line without subsequent anticoagulant therapy, whereas others treat with full-dose anticoagulants (a heparin-related compound, followed by 1 to 3 months of warfarin). If the line is functional and must stay in place (e.g., no alternative venous access), full-dose anticoagulants should be given. Otherwise, anticoagulant therapy should be given in all patients with upper extremity DVT, with medications, doses, regimens, and durations identical to those for treatment of DVT of the leg.

SUPERFICIAL THROMBOPHLEBITIS

Superficial thrombophlebitis usually manifests with pain, swelling, redness, and tenderness of superficial veins. Varicose veins[10] (Fig. 81-6) can be red, warm, and clustered in a circumscribed area. When superficial thrombophlebitis occurs in the short or long saphenous veins, usually redness, tenderness, and often linear induration follow the course of the involved vein (medial calf or thigh). Superficial thrombophlebitis also can occur at the insertion site of an intravenous catheter. Invariably, superficial thrombophlebitis is associated with thrombosis of the corresponding vein; particularly when the long saphenous vein is involved, venous ultrasonography should be performed to exclude extension into the deep veins, which occurs in up to 19% of patients. Approximately 25% of patients who have superficial venous thrombosis have concurrent DVT at the time of presentation, and approximately 3% of the others will subsequently develop DVT or pulmonary embolism in the next 3 months.[11]

Nonsteroidal anti-inflammatory drugs (NSAIDs) and either moderate or full doses of LMWH are each approximately 70% better than placebo for treating superficial thrombophlebitis.[12] LMWH relieves symptoms more quickly and prevents growth of thrombus more effectively than do NSAIDs. Fondaparinux, 2.5 mg/day for 45 days, reduces the risk for DVT or pulmonary embolism from approximately 1.3% to 0.2% without adverse effects. Thus, it is reasonable to use moderate doses of LMWH or fondaparinux for the initial treatment of acute, symptomatic superficial thrombophlebitis, particularly for patients with severe symptoms, proximal saphenous vein thrombosis, recurrent disease, or evidence of thrombophilia. Alternatively, and particularly for intravenous catheter–induced superficial thrombophlebitis, an NSAID can be tried. For varicose veins, laser and surgical treatments appear to be superior to foam sclerotherapy.[13]

POST-THROMBOTIC SYNDROME

The initial pain and swelling in many patients with DVT are due to the venous obstruction or the inflammatory process mediated by the acute thrombus. Once anticoagulant therapy is initiated, the acute obstruction usually resolves during a period of several months as recanalization occurs and collateral venous channels develop, thereby leading to initial improvement in pain and swelling. However, in the long term, probably because of venous valvular incompetence produced when the thrombosed venous segments recanalize and because of residual chronic obstruction, venous hypertension and sometimes pain and swelling can recur.

This post-thrombotic syndrome develops in up to 50% of patients with proximal DVT, usually within the first 1 to 2 years after DVT.[14] The syndrome is often a chronic, progressive disease with pain, swelling, and occasionally ulceration of the leg in patients with previous DVT. In a randomized trial of patients with acute iliofemoral DVT, the addition of catheter-directed thrombolysis using alteplase reduced the development of post-thrombotic syndrome from 56% to 41% at 24 months, but at the expense of an 8% risk for clinically relevant or major bleeding.

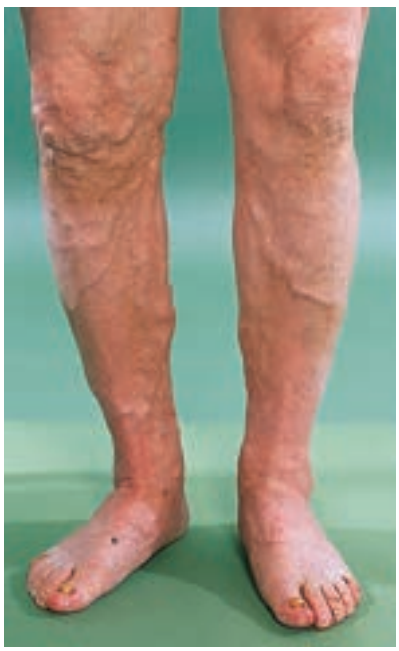

FIGURE 81-6. Varicose veins are a risk factor for deep vein thrombosis and may result from it. (From Forbes CD, Jackson WF. Color Atlas and Text of Clinical Medicine, 3rd ed. London: Mosby; 2003.)

PREVENTION AND TREATMENT

Rx

Despite earlier enthusiasm, a large, randomized trial showed compression stockings were not beneficial in preventing post-thrombotic syndrome. Consequently, it would seem reasonable to wait until the acute inflammatory process and acute outflow obstruction have subsided (usually up to 6 months) and then prescribe stockings if the patient's symptoms persist at that time.

Simple lifestyle alteration (such as frequent leg elevation, avoidance of prolonged standing or sitting, and occasional use of analgesics) relieve symptoms in many patients. If symptoms are severe, it is usually because extensive thrombus is causing massive edema. In such patients, a lightweight stocking (such as support hose) can be helpful until the edema improves. If symptoms persist or worsen despite these measures, or if ulceration seems imminent (as evidenced by severe skin changes), a full-strength stocking (30 to 40 mm Hg of pressure at the ankle) can be prescribed. However, if symptoms subside and the patient remains asymptomatic or has only trivial persistent signs or symptoms with little or no effect on quality of life, stockings can be avoided and the patient can be observed for clinically important signs and symptoms of post-thrombotic syndrome.

VENOUS ULCERS

Venous ulcers, which are the most severe complication of post-thrombotic syndrome, typically occur in the perimalleolar area of the leg. The best management is prevention by application of graduated compression stockings either at the time of diagnosis of DVT or, at the latest, when skin changes

develop in association with leg swelling. When an ulcer occurs, treatment with an emollient and regular wrapping should be commenced. Once the ulcer heals, the patient should be prescribed graduated compression stockings and watched for recurrent ulceration.[15] Surgical closure or removal of the incompetent saphenous veins plus dressing management in patients with chronic venous ulceration does not reduce healing time of the acute ulcer compared with dressing management alone but significantly reduces the rate of recurrent ulceration for at least the next 4 years.

Grade A References

A1. Hull RD, Pineo GF, Brant RF, et al. Long-term low-molecular-weight heparin versus usual care in proximal-vein thrombosis patients with cancer. *Am J Med*. 2006;119:1062-1072.

A2. Buller HR, Davidson BL, Decousus H, et al. Fondaparinux or enoxaparin for the initial treatment of symptomatic deep venous thrombosis: a randomized trial. *Ann Intern Med*. 2004;140:867-873.

A3. Kearon C, Ginsberg JS, Kovacs MJ, et al. Comparison of low-intensity warfarin therapy with conventional-intensity warfarin therapy for long-term prevention of recurrent venous thromboembolism. *N Engl J Med*. 2003;349:631-639.

A4. Pirmohamed M, Burnside G, Eriksson N, et al. A randomized trial of genotype-guided dosing of warfarin. *N Engl J Med*. 2013;369:2294-2303.

A5. Kearon C, Ginsberg JS, Anderson DR, et al. Comparison of 1 month of anticoagulation with 3 months of anticoagulation for a first episode of venous thromboembolism provoked by a transient risk factor. *J Thromb Haemost*. 2003;2:743-749.

A6. Kearon C, Akl EA, Comerota AJ, et al. Antithrombotic therapy for VTE disease: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2012;141:e419S-e494S.

A7. Bloomfield HE, Krause A, Greer N, et al. Meta-analysis: effect of patient self-testing and self-management of long-term anticoagulation on major clinical outcomes. *Ann Intern Med*. 2011;154:472-482.

A8. Lee AY, Levine MN, Baker RI, et al. Randomized comparison of low-molecular-weight heparin versus oral anticoagulant therapy for the prevention of recurrent venous thromboembolism in patients with cancer. *N Engl J Med*. 2003;349:109-111.

A9. Becattini C, Agnelli G, Schenone A, et al. Aspirin for preventing the recurrence of venous thromboembolism. *N Engl J Med*. 2012;366:1959-1967.

A10. EINSTEIN Investigators. Oral rivaroxaban for symptomatic venous thromboembolism. *N Engl J Med*. 2010;363:2499-2510.

A11. AMPLIFY Investigators. Oral apixaban for the treatment of acute venous thromboembolism. *N Engl J Med*. 2013;369:799-808.

A12. AMPLIFY-EXT Investigators. Apixaban for extended treatment of venous thromboembolism. *N Engl J Med*. 2013;368:699-708.

A13. RE-COVER Study Group. Dabigatran versus warfarin in the treatment of acute venous thromboembolism. *N Engl J Med*. 2009;361:2342-2352.

A14. RE-SONATE and RE-MEDY Study Groups. Extended use of dabigatran, warfarin, or placebo in venous thromboembolism. *N Engl J Med*. 2013;368:709-718.

A15. Decousus H, Prandoni P, Mismetti P, et al, for the CALISTO Study Group. Fondaparinux in the treatment of lower-limb superficial-vein thrombosis. *N Engl J Med*. 2010;363:1222-1232.

A16. Enden T, Haig Y, Kløw NE, et al. Long-term outcome after additional catheter-directed thrombolysis versus standard treatment for acute iliofemoral deep vein thrombosis (the CaVenT study): a randomised controlled trial. *Lancet*. 2012;379:31-38.

A17. Kahn SR, Shapiro S, Wells PS, et al. Compression stockings to prevent post-thrombotic syndrome: a randomised placebo-controlled trial. *Lancet*. 2014;383:880-888.

A18. O'Meara S, Richardson R, Lipsky BA. Topical and systemic antimicrobial therapy for venous leg ulcers. *JAMA*. 2014;311:2534-2535.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Martinelli I, De Stefano V, Mannucci PM. Inherited risk factors for venous thromboembolism. *Nat Rev Cardiol*. 2014;11:140-156.
2. Gornik HL, Sharma AM. Duplex ultrasound in the diagnosis of lower-extremity deep venous thrombosis. *Circulation*. 2014;129:917-921.
3. Lozano F, Trujillo-Santos J, Barron M, et al. Home versus in-hospital treatment of outpatients with acute deep venous thrombosis of the lower limbs. *J Vasc Surg*. 2014;59:1362-1367.
4. Wells PS, Forgie MA, Rodger MA. Treatment of venous thromboembolism. *JAMA*. 2014;311:717-728.
5. Roth JA, Boudreau D, Fujii MM, et al. Genetic risk factors for major bleeding in patients treated with warfarin in a community setting. *Clin Pharmacol Ther*. 2014;95:636-643.
6. Kearon C, Akl EA. Duration of anticoagulant therapy for deep vein thrombosis and pulmonary embolism. *Blood*. 2014;123:1794-1801.
7. Siegal DM, Crowther MA. Acute management of bleeding in patients on novel oral anticoagulants. *Eur Heart J*. 2013;34:489-498b.
8. Engelberger RP, Kucher N. Management of deep vein thrombosis of the upper extremity. *Circulation*. 2012;126:768-773.
9. Lamontagne F, McIntyre L, Dodek P, et al. Nonleg venous thrombosis in critically ill adults: a nested prospective cohort study. *JAMA Intern Med*. 2014;174:689-696.
10. Piazza G. Varicose veins. *Circulation*. 2014;130:582-587.
11. Decousus H, Quéré I, Presles E, et al. Superficial venous thrombosis and venous thromboembolism: a large, prospective epidemiologic study. *Ann Intern Med*. 2010;152:218-224.
12. Hamdan A. Management of varicose veins and venous insufficiency. *JAMA*. 2012;308:2612-2621.
13. Brittenden J, Cotton SC, Elders A, et al. A randomized trial comparing treatments for varicose veins. *N Engl J Med*. 2014;371:1218-1227.
14. Eberhardt RT, Raffetto JD. Chronic venous insufficiency. *Circulation*. 2014;130:333-346.
15. Zenilman J, Valle MF, Malas MB, et al. AHRQ Comparative Effectiveness Reviews. Chronic venous ulcers: a comparative effectiveness review of treatment modalities. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013.

REVIEW QUESTIONS

1. Which of the following statements about D-dimer testing is true?

A. When abnormal, the test is diagnostic of VTE.
B. The test can be used to diagnose pulmonary embolism when the pretest probability is high or intermediate.
C. When the results are normal in combination with a low pretest probability in suspected DVT, DVT can be excluded.
D. The degree of elevation accurately estimates clot size.
E. All of the above.

Answer: C D-Dimer is a sensitive but not specific test for diagnosing VTE. As a result, a negative test in a low-risk patient effectively excludes the diagnosis.

2. Which of the statements about VTE in pregnancy is untrue?

A. In at least 80% of pregnant women with DVT, the left leg is involved.
B. Warfarin is contraindicated in pregnant women between 6 and 12 weeks of gestation because of the potential to cause warfarin embryopathy.
C. When bleeding occurs at injection sites of LMWH, it is worthwhile to review the injection technique and to check the hemoglobin, hematocrit, platelet count, and anti–factor Xa level.
D. When heparin-induced thrombocytopenia is suspected in pregnant women treated with unfractionated heparin, LMWH is a reasonable substitute because there is minimal cross-reactivity of the causative antibody with LMWH.
E. None of the above.

Answer: D LMWH frequently cross-reacts with the heparin-induced thrombocytopenia antibody and can perpetuate the associated thrombotic complications.

3. Which of these statements about post-thrombotic syndrome is incorrect?

A. It inevitably leads to skin ulcers.
B. There is no gold standard test to diagnose it.
C. There is some doubt about whether elastic stockings prevent it.
D. It can be confused with recurrent DVT.
E. None of the above.

Answer: A Only a small proportion of subjects who develop post-thrombotic syndrome will progress to develop skin ulcers.

82

CARDIAC TRANSPLANTATION

DONNA MANCINI AND YOSHIFUMI NAKA

Heart failure is a progressive disease that now affects over 5 million patients in the United States (Chapter 58). Recent estimates suggest that 5 to 10% of all patients with heart failure have advanced, or stage D, disease, which is associated with a very high mortality rate and very poor quality of life (Chapter 59). Heart transplantation and mechanical assist devices are the only therapies that improve quality of life and survival in patients with stage D disease. With the improving results of cardiac transplantation—that is, 1- and 5-year survival rates approaching 90 and 72%, respectively—more patients are referred for transplant evaluation. Moreover, increasing experience and excellent outcomes have resulted in an expansion of eligibility criteria (Table 82-1) to patients whose comorbid conditions would have made them ineligible in prior decades.

TABLE 82-1 INDICATIONS FOR HEART TRANSPLANTATION

1. Refractory cardiogenic shock requiring continuous intravenous inotropic support or mechanical circulatory support with an intra-aortic balloon pump, venoarterial extracorporeal membrane oxygenation, or left ventricular assist device
2. Persistent class IV New York Heart Association congestive heart failure symptoms refractory to maximal medical therapy (peak oxygen consumption < 10-12 mL/kg/min)
3. Intractable or severe symptoms of ischemia in patients with coronary artery disease not amenable to percutaneous or surgical revascularization
4. Recurrent life-threatening arrhythmias refractory to medical therapy, catheter ablation, implantation of intracardiac defibrillator, or a combination of these
5. Congenital heart disease with severe ventricular dysfunction or that cannot be corrected or palliated by either surgical or medical treatment

Based on data that the 127 active heart transplant centers in the United States report to the United Network of Organ Sharing and the Scientific Registry for Transplant Recipients, approximately 2200 to 2400 heart transplant procedures are performed annually in the United States (Fig. 82-1). The major limitation to the growth of cardiac transplantation continues to be the scarcity of donor organs. With the increasing number of potential heart transplant recipients but a relatively constant number of donors, the wait list for transplantation and time to transplantation have continued to increase.[1-3]

SELECTION CRITERIA FOR CARDIAC TRANSPLANTATION

When a patient with advanced heart failure is referred to a transplantation center, the initial evaluation requires an assessment of the severity of heart failure, the identification of any potentially reversible factors, and an assessment of the adequacy of current medical therapy. If no reversible causes are identified and therapy is optimal, the evaluation process begins by determining whether the patient meets criteria for transplantation (see Table 82-1).[4,5]

The current United Network of Organ Sharing allocation policy takes into account the intensity of therapy used to support the patient (parenteral inotropic or mechanical support), time accrued on the wait list, blood type compatibility, and geographic distance.[6] Unfortunately, these policies result in substantial regional differences in wait times.[7] In patients who depend on parenteral inotropic support because of refractory cardiogenic shock or are deteriorating on parenteral inotropic agents, cardiac replacement therapy is the only option for long-term survival. The transplant evaluation must proceed expeditiously because these patients often require several days to weeks of circulatory support with percutaneous or implantable devices before the decision can be made as to whether to transition to long-term device support, proceed directly to cardiac transplantation, or withdraw care. The proportion of patients who require mechanical support before cardiac transplantation grew from 6% in 1998 to 24% in 2011 and now exceeds 30% in some regions of the United States.

Ambulatory patients with New York Heart Association (NYHA) class IIIB/IV symptoms comprise the largest number of referrals for cardiac transplant evaluation. Patients with a preserved exercise capacity, defined as a peak oxygen consumption (VO_2) greater than 14 mL/kg/minute, have a 1-year survival that is comparable to the expected survival in newly transplanted patients. By comparison, a peak VO_2 less than 10 mL/kg/minute is an absolute indication for transplantation. For patients with a peak VO_2 of 11 to 14 mL/kg/minute, decisions regarding transplantation must be individualized. Multivariable risk models can help predict whether survival would likely be better with or without the procedure based on factors such as the resting heart rate, mean arterial blood pressure, left ventricular ejection fraction, presence or absence of an intraventricular conduction defect, and serum sodium level.

Contraindications to Cardiac Transplantation

Absolute and relative exclusion criteria for heart transplantation (Table 82-2), which have evolved over time, consist of factors that increase perioperative risk, impair patients' ability to care for themselves, or affect long-term survival. Although no absolute age cutoff for heart transplantation exists and some centers will perform transplants in patients up to age 72 years, long-term survival is clearly decreased in older patients.

The major hemodynamic factor excluding cardiac transplantation is a nonreversible pulmonary vascular resistance (PVR) greater than 6 Wood units,

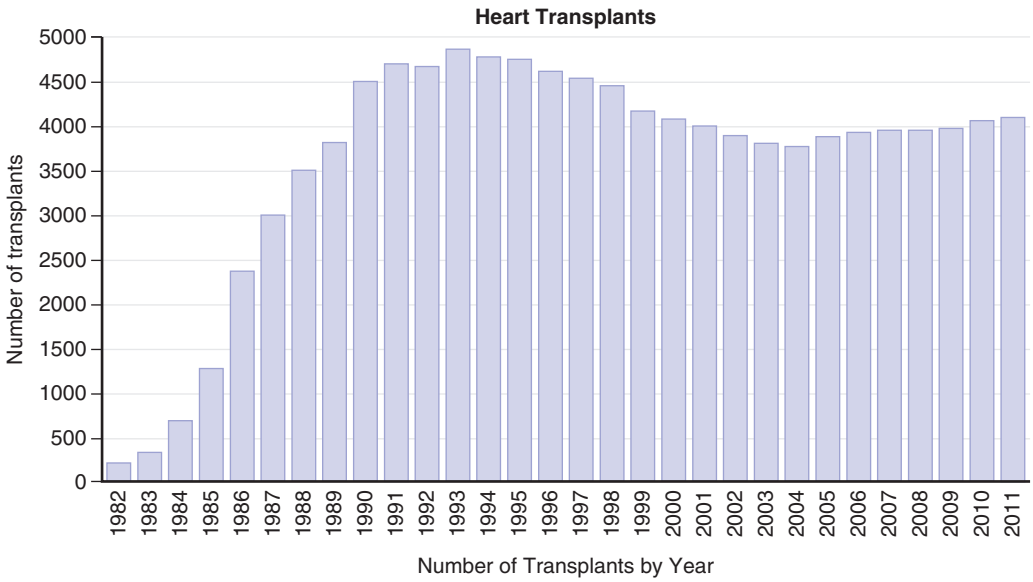

FIGURE 82-1. Heart transplant volume by year. (From Lund LH, Edwards LB, Kucheryavaya AY, et al. The Registry of the International Society for Heart and Lung Transplantation: thirtieth official adult heart transplant report—2013; focus theme: age. *J Heart Lung Transplant.* 2013;32:951-964.)

TABLE 82-2 CARDIAC TRANSPLANTATION CONTRAINDICATION CRITERIA

I. Absolute Contraindications
1. Systemic illness with a limited life expectancy despite heart transplant including:
   a. Active or recent solid organ or blood malignancy
   b. Irreversible renal or hepatic dysfunction in patients considered for heart-only transplantation
   c. Severe obstructive pulmonary disease (FEV_1 < 1 L/min)
   d. Active multisystem diseases
2. Fixed pulmonary hypertension with a mean transpulmonary gradient > 5 mm Hg or pulmonary vascular resistance > 6 Wood units not reduced with vasodilators, parenteral inotropic agents, phosphodiesterase type V inhibitors, endothelin receptor antagonists, or a mechanical assist device.

II. Relative contraindications
1. Age > 72 yr
2. Any active infection (with exception of device-related infection in ventricular assist device recipients)
3. Active peptic ulcer disease
4. Diabetes mellitus with moderate end-organ involvement (neuropathy, nephropathy, or retinopathy)
5. Severe peripheral vascular or cerebrovascular disease
6. Morbid obesity (BMI > 35) or cachexia (BMI < 18)
7. Significant chronic renal impairment with creatinine > 2.5 mg/dL or creatinine clearance < 25 mL/min*
8. Significant hepatic impairment with bilirubin > 2.5 mg/dL, serum transaminase levels > 3 times normal, INR > 1.5 off warfarin
9. Severe pulmonary dysfunction with FEV_1 < 40% normal
10. Recent pulmonary infarction within 6-8 wk
11. Irreversible neurologic or neuromuscular disorder
12. Active mental illness or psychosocial instability
13. Drug, tobacco, or alcohol abuse within 6 mo
14. Significant coagulopathies

*May be suitable for cardiac transplantation if inotropic support and hemodynamic management produce a creatinine < 2 mg/dL and creatinine clearance > 50 mL/min. Transplantation may also be advisable as combined heart-kidney transplant.
BMI = body mass index; FEV_1 = forced expiratory volume in one second; INR = international normalized ratio.

which increases the risk for immediate postoperative right ventricular failure and the 30-day mortality rate. In most patients with advanced heart failure, however, pulmonary hypertension is reversible with vasodilators (Chapter 68) or after implantation of a left ventricular assist device.

Although diabetes mellitus with evidence of significant end-organ damage (e.g., neuropathy or nephropathy) is a relative contraindication to heart transplantation, carefully selected patients with diabetes can undergo successful transplantation with morbidity and mortality similar to that in patients without diabetes. In patients with diabetes who have renal dysfunction (Chapter 124), combined heart and kidney transplantation (Chapter 131) can be considered, with a survival rate comparable to that of heart transplantation alone.

Patients with an active or recent malignancy may be offered mechanical support either before or after cancer treatment as a way to bridge them to transplant. However, any patient with a history of malignancy has an increased risk for developing a second malignancy owing to immunosuppression after the transplantation (Chapter 49).

Combined heart and stem cell transplantation (Chapter 178) is an option in patients with primary amyloid light-chain amyloidosis (Chapter 188), but survival rates are lower than in other transplant patients because of the frequent recurrence of amyloidosis in the transplanted heart. In contrast, survival of patients with familial amyloidosis caused by a mutant form of the protein transthyretin is comparable to that in other transplant recipients.

Repeat transplantation now accounts for 3% of U.S. heart transplants, usually in patients who have developed chronic allograft dysfunction because of severe transplant coronary artery disease, often with a left ventricular ejection fraction less than 45% or with restrictive cardiomyopathy, but without any other significant comorbid conditions. However, repeat transplantation is associated with greater risk for infection and malignancies, owing to the heightened immunosuppression, and a poorer long-term survival.

Currently, 3% of adults undergoing cardiac transplantation have complex congenital heart disease (Chapter 69) as the cause of their heart failure. As more patients with complex congenital heart disease survive into adulthood, however, an estimated 10 to 20% of such patients will become candidates for heart or combined heart-lung transplantation at some time during their lives. In such patients, the short-term post-transplant survival is significantly lower compared with patients who have ischemic or dilated cardiomyopathies owing to their higher rate of intraoperative and post-operative bleeding. If a patient with congenital heart disease survives the surgery, however, 10-year survival post-transplant is excellent.

Active bacterial infection is a temporary absolute contraindication to heart transplantation, except in the setting of mechanical device infection, in which transplant is felt to be curative. Patients who are positive for human immunodeficiency virus (HIV) and have end-stage cardiomyopathy can be considered for transplant, with good short-term outcome in carefully screened patients who have low or undetectable viral loads and no recent significant bacterial infections. Patients with chronic hepatitis B or C (Chapter 149) have an increased incidence of postoperative liver disease, but their post-transplant survival is not reduced.

Organ allocation

Donor Criteria

Donors and recipients are matched for ABO blood compatibility and size. Weight matching is generally within 25% of recipient body weight, though

donors of equal size or larger are preferred for recipients with high PVR. Height mismatches greater than 6 inches are currently not recommended. Males under age 40 and females under age 45 are suitable donors, provided echocardiography shows no evidence of preexisting heart disease or impaired myocardial function. Older individuals also may be suitable donors if coronary atherosclerotic lesions can be excluded, optimally by cardiac catheterization. Donors with serologic findings positive for HIV, hepatitis B and C, and nonprimary brain malignancies are generally not accepted. Organs procured with ischemic times in excess of 4 hours are associated with a higher rate of primary graft failure.

Matching Donors and Recipients

Approximately 10% of transplant candidates have human leukocyte antigen (HLA) antibodies that could lead to a positive crossmatch. The sensitized candidate is typically a multiparous woman, a patient who has received multiple prior transfusions, or patients supported with a mechanical assist device. Patients with high antibody levels require a donor-specific T-cell crossmatch before transplantation to exclude the presence of lymphocytotoxic immunoglobulin G antibodies against donor HLA class I antigens, a situation that can cause hyperacute rejection. Sensitized patients are also at risk for acute humoral rejection and an earlier onset of accelerated coronary artery disease.

A donor-specific T-cell crossmatch is a contraindication to transplantation; thus, sensitized candidates have longer wait times before receiving a cardiac allograft. With technologies using solid-phase assays and flow cytometric techniques that can rapidly identify class I and II antibodies, long-distance donors can be screened for unacceptable antigens—a "virtual crossmatch" that enlarges the potential donor pool for sensitized patients.

Surgical Technique

Orthotopic cardiac transplantation can be performed by using a *biatrial* anastomosis and reconnecting the pulmonary artery and aorta above the semilunar valves. Increasingly, however, atrial function is preserved by performing a bi*caval* anastomosis, which results in improved atrial geometry, better right ventricular function, less frequent atrioventricular valve regurgitation, and less sinus node dysfunction.

Immunosuppression

Most transplant centers use a triple drug therapy regimen including a calcineurin inhibitor (cyclosporine or tacrolimus), an antiproliferative agent (usually mycophenolate mofetil), and steroids (Chapter 49). Some centers, however, use only a single agent such as tacrolimus.[1] By the third month after surgery, most patients receive only prednisone 5 mg/day and approximately 25 to 60% of patients can tolerate total withdrawal of steroids by the end of the first year.

Acute rejection occurs most frequently in the first 3 months after transplant. Some centers use selective induction agents, such as basiliximab, that target the activated interleukin-2 (IL-2) receptor on T cells, or potent nonselective immunosuppressive agents, such as thymoglobulin, in the perioperative period to decrease early allograft rejection. However, this intensification of immunosuppression can predispose patients to more frequent opportunistic infections. Currently, approximately 50% of transplant centers use induction therapy.

For calcineurin inhibitors, prospective randomized trials have demonstrated a decreased incidence of allograft rejection, less hypertension, lower lipid levels, less hirsuitism, and less gingival hyperplasia using tacrolimus compared with cyclosporine, although most studies demonstrating no difference in survival.[2,3] Mycophenolate mofetil, a selective de novo purine inhibitor, is the preferred antiproliferative agent to reduce rejection and the development of transplant vasculopathy.[4] Although everolimus may be better than mycophenolate mofetil in preventing early transplant vasculopathy, it also is associated with more side effects,[5] so mycophenolate mofetil remains the current agent of choice.

Rejection

Allograft rejection, which can be antibody-mediated or cell-mediated, occurs most frequently in the first 6 months after the transplant. Antibody-mediated, or humoral, rejection generally occurs very early after the transplant, particularly in a previously sensitized recipient. It is characterized histologically by immunoglobulin and complement deposition in the absence of cellular rejection, and often it is associated with hemodynamic compromise. T-cell mediated rejection, triggered by the recognition of foreign antigens on the surface of engrafted cells, accounts for more than 90% of rejection episodes, usually within the first 6 months.

TABLE 82-3 HISTOLOGIC GRADING OF CELLULAR REJECTION*

Grade 0R	No rejection
Grade 1R (mild)	Interstitial, perivasicular, or both infiltrate with up to 1 focus of myocyte damage
Grade 2R (moderate)	≥2 foci of infiltrate with myocyte damage
Grade 3R (severe)	Diffuse infiltrate with multifocal myocyte damage, edema, hemorrhage, vasculitis

*Modified from Stewart S, Winters GL, Fishbein MC, et al. Revision of the 1990 working formulation for the standardization of nomenclature in the diagnosis of heart rejection. *J Heart Lung Transplant.* 2005;24:1710-1720.

TABLE 82-4 2013 INTERNATIONAL SOCIETY FOR HEART & LUNG TRANSPLANTATION CLASSIFICATION FOR DIAGNOSIS OF CARDIAC ANTIBODY-MEDIATED REJECTION (AMR)*

GRADE	DEFINITION
pAMR 0	No rejection
pAMR 1 (H+)	Histopathologic changes are present without immunopathologic findings.
pAMR 1 (I+)	Immunopathologic findings are positive without histologic findings.
pAMR 2	Both immunologic and histologic findings are present.
pAMR 3	Severe pathologic antibody-mediated rejection with interstitial hemorrhage, capillary fragmentation, mixed inflammatory infiltrates, endothelial cell pyknosis, karyorrhexis, or a combination of these, and marked edema *and* immunopathologic findings are present. These cases may be associated with profound hemodynamic dysfunction and poor clinical outcomes.

*Modified from Berry GJ, Burke MM, Andersen C, et al. The 2013 International Society for Heart and Lung Transplantation Working Formulation for the standardization of nomenclature in the pathologic diagnosis of antibody-mediated rejection in heart transplantation. *J Heart Lung Transplant.* 2013;32:1147-1162.

To diagnose allograft rejection, endomyocardial biopsies are usually performed by the transjugular approach weekly for the first month, then every other week for 2 months, then every 1 to 2 months for the first year. Biopsy grading of cellular rejection is based on the severity of lymphocyte infiltration and myocyte necrosis on hematoxylin and eosin staining (Table 82-3), whereas antibody-mediated rejection also includes immunologic staining (Table 82-4). Although the presence of donor-specific antibodies is not a requirement for the diagnosis of antibody-mediated rejection, patients suspected of having antibody-mediated rejection are frequently tested and serially monitored for donor-specific antibody. In asymptomatic patients on low doses of steroids, gene expression profiling of peripheral blood samples can reduce the number of biopsies while providing equivalent clinical outcomes.[6]

Treatment of Rejection

The treatment of allograft rejection is determined by the presence of symptoms, the degree of left ventricular dysfunction, the time since transplant, and the pathologic grade of the biopsy. Most episodes of cellular rejection are easily treated with high-dose oral or intravenous steroids. Patients with hemodynamically significant rejection will require rescue therapy with thymoglobulin. Humoral rejection therapy includes modalities that both clear and reduce the production of the antibody, such as plasmapheresis, intravenous immunoglobulin, thymoglobulin, high-dose steroids, and B cell–specific monoclonal antibodies such as rituximab[8] (Chapter 49).

Transplant Vasculopathy

Transplant vasculopathy, which is primarily a form of chronic rejection, occurs at an annual incidence rate of 5 to 10% and remains one of the main causes of late death after cardiac transplantation. Risk factors for transplant

vasculopathy include an increased number of HLA mismatches, increased number of acute rejection episodes, older donor age, prior cytomegalovirus (CMV) infection, ischemia-reperfusion injury, and the classic risk factors for atherosclerotic disease—age, smoking, obesity, diabetes, dyslipidemia, and hypertension. Histologic examination shows subendothelial accumulation of primarily T cells, myointimal proliferation of smooth muscle cells, lipid-laden foam cells, and perivascular fibrosis.

Patients rarely experience exertional angina and are more likely to present with severe fatigue, heart failure, myocardial infarction, ventricular arrhythmia, or sudden death. Patients should be screened annually either by angiography, with or without intravascular ultrasound, or by dobutamine stress echocardiography. An increase in intimal thickness of at least 0.5 mm, as detected by intravascular ultrasound, is a reliable indicator of both cardiac allograft vasculopathy and 5-year mortality.

In contrast to routine coronary artery disease, cardiac allograft vasculopathy is usually manifested by concentric narrowing, owing to neointimal proliferation of vascular smooth muscle cells throughout the length of the vessel and angiographic evidence of rapid tapering, pruning, and obliteration of vessels. Some patients may have focal coronary lesions that are amenable to stent placement, but generally the disease is diffuse and not amenable to percutaneous coronary interventions or bypass grafting.

Treatment has been predominantly aimed at prevention. Use of high-dose or low-dose statins (see Table 206-6 in Chapter 206) can reduce the development of cardiac vasculopathy.[■] Sirolimus and everolimus also can reduce the incidence of cardiac vasculopathy and slow its progression. However, the only definitive treatment of transplant vasculopathy is repeat transplantation, and the survival of patients undergoing repeat transplantation for severe vasculopathy is comparable to that of de novo heart transplant recipients.

Malignancy

With the improved survival of heart transplant recipients and longer exposure to immunosuppressive drugs, malignancies, especially lymphomas (Chapter 185), are now almost equal to transplant vasculopathy as the leading cause of long-term mortality. Post-transplant lymphoproliferative disease (Chapters 49 and 185) includes a spectrum of predominantly B-cell lymphomas (~90%) frequently associated with Epstein-Barr virus (Chapter 377). T-cell lymphomas are much less frequent (10%) and often more difficult to treat (Chapter 185). Skin cancers (Chapter 203) are also more frequent in post-transplant patients, but the incidence of solid organ malignancies is not much different from that in the general population.

Infection

Infections (Chapter 281) account for approximately 20% of deaths within the first year after transplant surgery and continue to be a common cause of morbidity and mortality throughout the recipient's life. In the waiting period, careful attention should be given to updating immunizations against pneumococcal pneumonia, hepatitis B, and herpes zoster (Chapter 18). Young women should receive vaccination against the human papilloma virus. Patients with positive tuberculosis skin tests should be treated with isoniazid and pyridoxine (Chapter 324), and patients with latent syphilis should be treated with penicillin (Chapter 319).

Infections early after transplant are predominantly bacterial from hospital-acquired organisms (Chapter 282), catheters, the surgical site, a prior left ventricular assist device, or occasionally from donor-transmitted disease. Any infection early after transplantation increases the risk for a subsequent fatal CMV infection (Chapter 376). Any CMV-positive patient or CMV-negative patient receiving a CMV-positive organ should receive prophylactic ganciclovir followed by valganciclovir.

Fungal and viral infections are usually more frequent starting a month or so after transplant. Antibiotic prophylaxis includes perioperative antibacterial agents, such as cefazolin, and initiation of prophylaxis against CMV infection, *Pneumocystis jiroveci* pneumonia, herpes simplex virus infection, and oral candidiasis (Chapter 338). The prophylactic use of one single-strength trimethoprim-sulfamethoxazole tablet daily, typically for the first year after transplantation, has virtually eliminated *P. jiroveci* (Chapter 341) and also prevents nocardial infections (Chapter 330) and toxoplasmosis (Chapter 349). Aspergillosis (Chapter 339) and candidiasis (Chapter 338) are the most common fungal infections after heart transplantation; oral nystatin solution or clotrimazole troches are routinely used in the first 3 to 6 months (Chapter 331).

Comorbid Conditions

Within 5 years, hypertension (Chapter 67) occurs in over 92% of transplant recipients, primarily owing to the side effects of calcineurin inhibitors. Diabetes (Chapter 229) is observed in approximately 40% of patients owing to treatment with corticosteroids or tacrolimus. Hyperlipidemia is found in approximately 90% of patients and is treated with statins, with the same approach as for patients with known coronary disease (Chapter 206). Osteoporosis (Chapter 243) is also common because of chronic steroid treatment as well as the use of calcineurin inhibitors. Significant renal insufficiency (serum creatinine > 2.5 mg/dL) occurs in 20% of patients by 10 years post-transplant.

PROGNOSIS

During the first year after transplantation, early causes of death are graft failure, infection, multiorgan failure, and allograft rejection. Overall survival, which is approximately 85% at 1 year and 75% at 5 years, has been improving over time (Fig. 82-2). After 5 years, cardiac allograft vasculopathy (14%), late graft failure (18%), malignant disease (25%), and non-CMV infection (10%) are the most prominent causes of death. Of the patients, 50% will survive for more than 11 years and many survive for 20 to 30 years, often with good ventricular function.[9]

Functional capacity is usually excellent, with more than 90% of 1-year survivors reporting no functional limitations. Exercise capacity improves but remains reduced compared with that of age- and gender-matched controls, owing to denervation of the heart, side effects of immunosuppressive drugs,

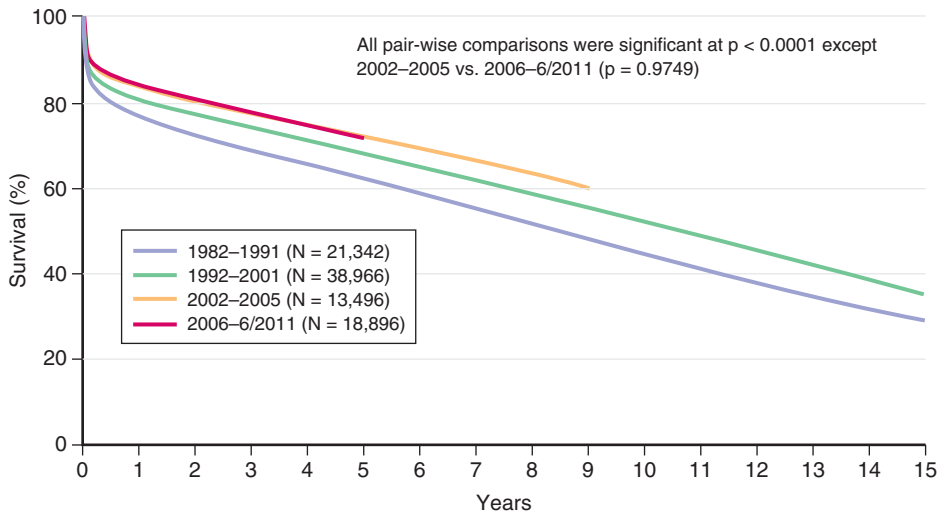

FIGURE 82-2. Survival by transplant era. Adult heart transplants: Kaplan-Meier survival by era, January 1982-June 2011. (From Lund LH, Edwards LB, Kucheryavaya AY, et al. The Registry of the International Society for Heart and Lung Transplantation: thirtieth official adult heart transplant report—2013; focus theme: age. *J Heart Lung Transplant.* 2013;32:951-964.

and donor-recipient size mismatch. Many patients return to full-time employment, travel extensively, and participate in vigorous sports such as skiing, running, and hiking. Some young patients may bear children of their own, though genetic counseling is recommended in patients with familial cardiomyopathies (Chapter 60). In female recipients, pregnancy, which requires modification of immunosuppression and close monitoring for allograft rejection, is recommended only in 1-year survivors with normal graft function.

FUTURE DIRECTIONS

With the continuing scarcity of donor organs, physicians must use this scarce resource wisely by selecting candidates who would benefit the greatest over time. In the future, warm preservation techniques may extend harvest time as well and enable the resuscitation of donor organs whose function may initially appear marginal. The field of mechanical circulatory support continues to evolve rapidly, and devices (Chapter 59) rather than transplantation will probably offer a greater chance for long-term survival for most patients with advanced heart failure.

Grade A References

A1. Baran DA, Zucker MJ, Arroyo LH, et al. A prospective, randomized trial of single-drug versus dual-drug immunosuppression in heart transplantation: the tacrolimus in combination, tacrolimus alone compared (TICTAC) trial. *Circ Heart Fail*. 2011;4:129-137.

A2. Ye F, Ying-Bin X, Yu-Guo W, et al. Tacrolimus versus cyclosporine microemulsion for heart transplant recipients: a meta-analysis. *J Heart Lung Transplant*. 2009;28:58-66.

A3. Penninga L, Moller CH, Gustafsson F, et al. Tacrolimus versus cyclosporine as primary immunosuppression after heart transplantation: systematic review with meta-analyses and trial sequential analyses of randomised trials. *Eur J Clin Pharmacol*. 2010;66:1177-1187.

A4. Sanchez-Lazaro IJ, Almenar L, Martinez-Dolz L, et al. A prospective randomized study comparing cyclosporine versus tacrolimus combined with daclizumab, mycophenolate mofetil, and steroids in heart transplantation. *Clin Transplant*. 2011;25:606-613.

A5. Guethoff S, Meiser BM, Groetzner J, et al. Ten-year results of a randomized trial comparing tacrolimus versus cyclosporine a in combination with mycophenolate mofetil after heart transplantation. *Transplantation*. 2013;95:629-634.

A6. Eisen HJ, Kobashigawa J, Keogh A, et al. Three-year results of a randomized, double-blind, controlled trial of mycophenolate mofetil versus azathioprine in cardiac transplant recipients. *J Heart Lung Transplant*. 2005;24:517-525.

A7. Eisen HJ, Kobashigawa J, Starling RC, et al. Everolimus versus mycophenolate mofetil in heart transplantation: a randomized, multicenter trial. *Am J Transplant*. 2013;13:1203-1216.

A8. Pham MX, Teuteberg JJ, Kfoury AG, et al. Gene-expression profiling for rejection surveillance after cardiac transplantation. *N Engl J Med*. 2010;362:1890-1900.

A9. Som R, Morris PJ, Knight SR. Graft vessel disease following heart transplantation: a systemic review of the role of statin therapy. *World J Surg*. 2014;38:2324-2334.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Lund LH, Edwards LB, Kucheryavaya AY, et al. The Registry of the International Society for Heart and Lung Transplantation: thirtieth official adult heart transplant report—2013; focus theme: age. *J Heart Lung Transplant*. 2013;32:951-964.
2. Kirklin JK, Naftel DC, Kormos RL, et al. Fifth INTERMACS annual report: risk factor analysis from more than 6,000 mechanical circulatory support patients. *J Heart Lung Transplant*. 2013;32:141-156.
3. Colvin-Adams M, Smithy JM, Heubner BM, et al. OPTN/SRTR 2012 Annual Data Report: heart. *Am J Transplant*. 2014;14(suppl 1):113-138.
4. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. *Circulation*. 2013;128:1810-1852.
5. Mancini D, Lietz K. Selection of cardiac transplantation candidates in 2010. *Circulation*. 2010;122:173-183.
6. Colvin-Adams M, Valapour M, Hertz M, et al. Lung and heart allocation in the United States. *Am J Transplant*. 2012;12:3213-3234.
7. Schulze PC, Kitada S, Clerkin K, et al. Regional differences in recipient waitlist time and pre- and post-transplant mortality after the 2006 United Network for Organ Sharing policy changes in the donor heart allocation algorithm. *JACC Heart Fail*. 2014;2:166-177.
8. Kobashigawa J, Crespo-Leiro MG, Ensminger SM, et al. Report from a consensus conference on antibody-mediated rejection in heart transplantation. *J Heart Lung Transplant*. 2011;30:252-269.
9. Galeone A, Kirsch M, Barreda E, et al. Clinical outcome and quality of life of patients surviving 20 years or longer after heart transplantation. *Transpl Int*. 2014;26:576-582.

83 APPROACH TO THE PATIENT WITH RESPIRATORY DISEASE

MONICA KRAFT

Respiratory symptoms, which are among the most common reasons why patients seek medical care, are responsible for approximately 20% of office visits to a primary care physician. In addition to a careful history, a systematic physical examination is critical for accurate diagnosis.

A careful pulmonary examination complements the cardiac physical examination (Chapter 51). Inspection may reveal an elevated jugular pressure, indicative of right heart failure owing to cor pulmonale (Chapter 68). Cervical or supraclavicular adenopathy (Chapter 168) may be the first clue to suggest a thoracic malignancy (Chapter 191) or mycobacterial infection (Chapter 324). Unilateral arm swelling can be caused by venous thrombosis (Chapter 81), whereas venous engorgement of the head and neck can be caused by a tumor that results in superior vena cava syndrome (see Fig. 99-8 in Chapter 99). On the cardiac examination, a loud pulmonic second heart sound is suggestive of pulmonary hypertension, which also can result in a murmur of tricuspid (see Table 51-7 in Chapter 51) or pulmonic valve insufficiency.

Inspection of the chest may show hyperinflation and reduced diaphragmatic excursion, typical of chronic obstructive pulmonary disease (COPD; Chapter 88), chest wall abnormalities such as kyphoscoliosis (Chapter 99), or diaphragmatic muscle wall weakness as in many hypoventilation syndromes (Chapter 86). Percussion may reveal dullness in patients with pleural effusions or with lung that has been consolidated by pneumonia.

Auscultation of the lungs[1] includes listening at both apices and over both upper and lower lobes, anteriorly and posteriorly, and during inspiration and respiration. Normal lung sounds are heard during inspiration and early expiration as soft and non-musical sounds (Table 83-1). *Bronchial breath sounds*, which sound similar to but often somewhat harsher than normal lung sounds, are heard throughout expiration as well as inspiration, similar to what would be heard by placing a stethoscope over the trachea.

The term *rales* is no longer used and has been replaced by the term *crackles*. Fine crackles are non-musical and heard typically in late inspiration; they are most commonly a sign of heart failure (Chapter 58) or interstitial lung disease (Chapter 92). By comparison, coarse crackles, which unlike fine crackles tend to be transmitted through the mouth and cleared by coughing, are typical of bronchitis (Chapter 96) and COPD (Chapter 88). *Wheezes* are high-pitched, musical sounds heard during expiration and sometimes inspiration, most commonly in asthma (Chapter 87) and sometimes in COPD (Chapter 88). When these diseases are severe, however, the degree of airflow may be insufficient to produce wheezes. A *rhonchus* is a musical, low-pitched sound typically heard in expiration and sometimes during inspiration; it often resolves with coughing. Like coarse crackles, rhonchi are common in bronchitis (Chapter 96) and COPD (Chapter 88). A *pleural friction rub*, which classically occurs during inspiration but sometimes also during expiration, is heard in patients with inflammatory diseases or malignancies involving the pleura (Chapters 99 and 191). *Stridor* is a musical, high-pitched sound that may be audible without a stethoscope and that indicates upper airway obstruction, such as found with acute inflammatory or chronic degenerative diseases of the larynx (Chapter 429) or obstruction of the trachea, as may be caused by intrathoracic malignant diseases (Chapter 191). An absence of breath sounds would be noted if the lung is not ventilated because of a complete bronchial obstruction or if it is displaced by a pleural effusion.

Tactile fremitus, which is a vibratory sensation noted during breathing, is increased in patients who have consolidated lung from pneumonia, because the vibratory sensation conducts better through such lung tissue and is diminished in patients with pleural effusion. *Egophony,* by which a patient's recitation of the long E sound is heard on auscultation as a long A sound, is another indication of consolidation typical of pneumonia.

Evaluation of the abdomen may show a readily palpable liver, sometimes mistaken for hepatomegaly, in patients with COPD and low diaphragm. Examination of the extremities may reveal cyanosis in patients who are hypoxemic, usually with a partial pressure of oxygen less than 55 mm Hg, although it also may be observed in patients with methemoglobinemia (Chapter 158). Clubbing (Chapter 51) is indicative of chronic hypoxemia, as seen in patients with chronic right-to-left-shunting from congenital heart disease (Chapter 69) or other causes of long-standing hypoxemia (Chapters 88 and 92), but it also may be indicative of pleural-based diseases (Chapter 99) as part of the syndrome of hypertrophic pulmonary osteoarthropathy (Chapters 179 and 275).

In patients with suspected hypoxemia, careful analyses of arterial blood gases can help determine its severity and guide therapy (Chapter 103). In patients in whom it is difficult to distinguish heart failure from a pulmonary cause of hypoxemia, an elevated brain natriuretic peptide level may point to a cardiac cause (Chapter 58). Chest imaging (Chapter 84) is a crucial part of the evaluation of many potential pulmonary complaints, and pulmonary function testing (Chapter 85) can be extremely helpful in distinguishing among causes of acute and chronic lung disease.

Among the most common respiratory complaints are cough, wheezing, dyspnea, and hemoptysis. Each can and should be approached in a systematic way.

TABLE 83-1 DIAGNOSTIC UTILITY OF LUNG AUSCULTATION

AUSCULTATORY FINDING	CLINICAL CORRELATION
Bronchial breathing	Pneumonia or interstitial lung disease
Fine crackle	Heart failure, interstitial lung disease, alveolar filling disorders
Coarse crackle	Bronchitis
Wheeze	Asthma, COPD
Rhonchus	Bronchitis, COPD
Stridor	Upper-airway obstruction from laryngeal or tracheal inflammation, mass lesions, or external compression
Pleural friction rub	Pleural inflammation or tumors

COPD = chronic obstructive pulmonary disease.

APPROACH TO THE PATIENT WITH COUGH

Cough is the single most common respiratory complaint for which patients seek care. Referrals of patients with persistently troublesome chronic cough of unknown cause account for 10 to 38% of outpatient visits to respiratory specialists.

For acute cough, defined as coughing that has been present for less than 8 weeks, a careful medical history and physical examination will usually reveal the diagnosis (Table 83-2). Although most acute coughs are of minor

TABLE 83-2 SPECTRUM OF CAUSES AND FREQUENCIES OF COUGH IN IMMUNOCOMPETENT ADULTS

COMMON	LESS COMMON
ACUTE COUGH	
Common cold	Asthma
Acute bacterial sinusitis	Pneumonia
Pertussis	Heart failure
Exacerbations of COPD	Aspiration syndromes
Allergic rhinitis	Pulmonary embolism
Environmental irritant rhinitis	Exacerbation of bronchiectasis
CHRONIC COUGH	
Rhinosinus conditions/UACS	Bronchogenic carcinoma
Asthma	Chronic interstitial pneumonia
Gastroesophageal reflux	Sarcoidosis
Chronic bronchitis	Left heart failure
Eosinophilic bronchitis	Obstructive sleep apnea
Bronchiectasis	Chronic tonsillar enlargement
ACE inhibitors	
Postinfection	

ACE = angiotensin-converting enzyme; COPD = chronic obstructive pulmonary disease; UACS = upper airway cough syndrome.

consequence, cough can occasionally be a sign of a potentially life-threatening illness, such as pulmonary embolism (Chapter 98), pneumonia (Chapter 97), or heart failure (Chapter 58).

Up to 98% of all cases of chronic cough, defined as a cough that persists for more than 8 weeks, in immunocompetent adults are caused by eight common conditions: postnasal drip syndrome from a variety of rhinosinus conditions (Chapter 251), asthma (Chapter 87), gastroesophageal reflux disease (GERD) (Chapter 138), chronic bronchitis (Chapter 88), eosinophilic bronchitis, bronchiectasis (Chapter 90), use of angiotensin-converting enzyme (ACE) inhibitors, and postinfectious cough. Postinfectious cough is usually nonproductive and lasts for 3 to 8 weeks after an upper respiratory tract infection; patients have a normal chest radiograph. Uncommon causes of chronic cough include bronchogenic carcinoma (Chapter 191), chronic interstitial pneumonia (Chapter 92), sarcoidosis (Chapter 95), left ventricular failure (Chapter 58), and aspiration (Chapter 94).

DIAGNOSIS

In chronic cough (Fig. 83-1), the character and timing are not of diagnostic help. A chest radiograph should be obtained in all patients, but other tests should not be ordered in current smokers or patients taking ACE inhibitors until the response to smoking cessation or discontinuation of the drug for at least 4 weeks can be assessed. Sinus radiographs, barium esophagography, methacholine challenge, esophageal pH, and bronchoscopy can be ordered as part of the initial evaluation, depending on the history and physical examination (Table 83-3; see Fig. 83-1). If a test points toward a possible diagnosis, a trial of treatment for that condition is needed to confirm the diagnosis.[2]

TREATMENT Rx

The specific cause of cough can be diagnosed and treated successfully 84 to 98% of the time, so nonspecific therapy[3] aimed to suppress the cough per se is rarely indicated. There is no strong evidence that nonspecific therapies such as antitussives, mucolytics, decongestants, or antihistamine-decongestant

TABLE 83-3 TESTING CHARACTERISTICS OF DIAGNOSTIC PROTOCOL FOR EVALUATION OF CHRONIC COUGH

TESTS	DIAGNOSIS	POSITIVE PREDICTIVE VALUE, %	NEGATIVE PREDICTIVE VALUE, %
Sinus radiograph	Sinusitis	57-81	95-100
Methacholine inhalation challenge	Asthma	60-82	100
Modified barium esophagography	GERD, esophageal stricture	38-63	63-93
Esophageal pH*	GERD	89-100	
Bronchoscopy	Endobronchial mass/lesion	50-89	100

*24-Hour esophageal pH monitoring.
GERD = gastroesophageal reflux disease.

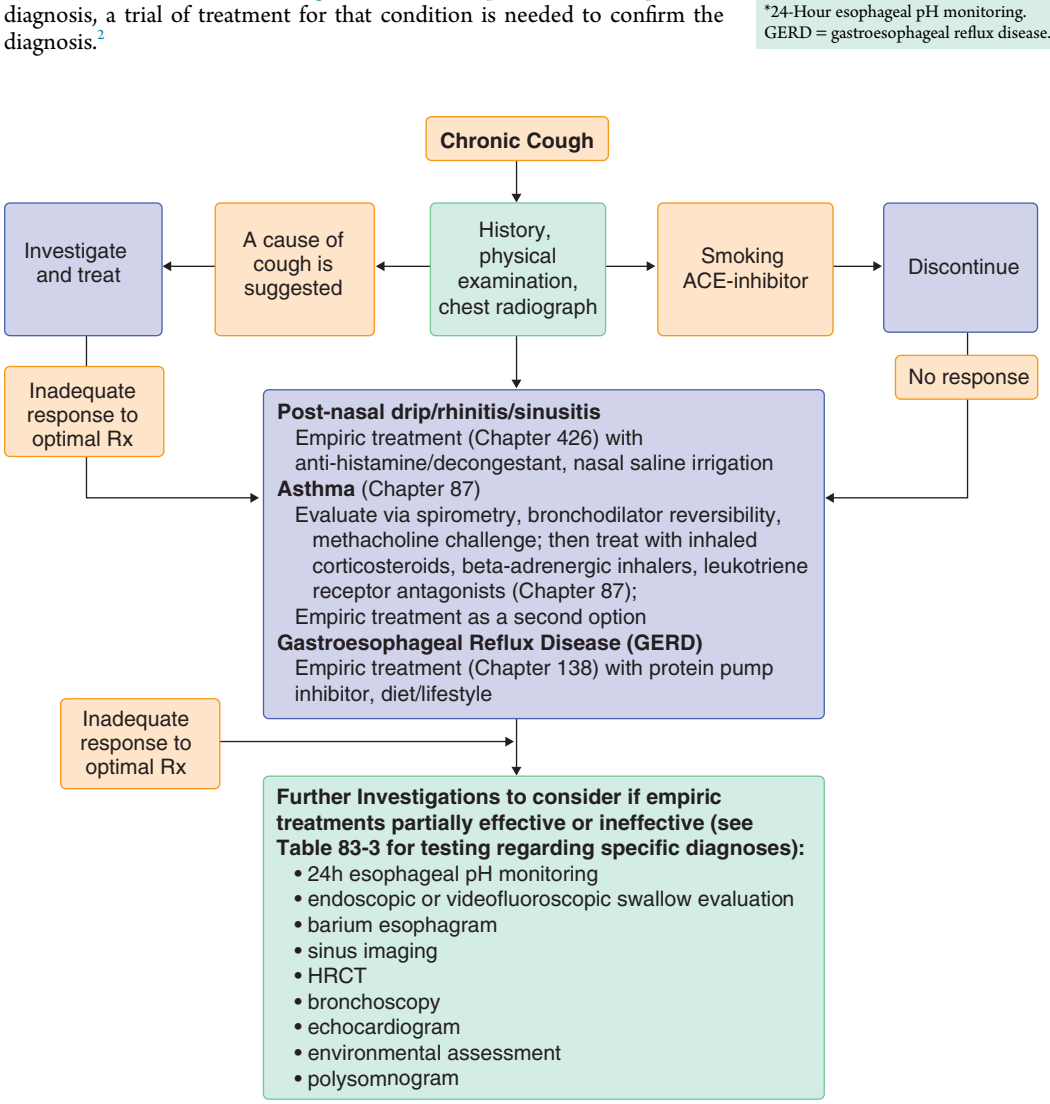

FIGURE 83-1. Algorithm for the management of chronic cough lasting longer than 8 weeks. ACE = angiotensin-converting enzyme; HRCT = high-resolution computed tomography; Rx = prescription.

combinations are efficacious for acute cough in the setting of an upper respiratory tract infection.[1] For nonspecific persistent cough, effective treatment of chronic gastroesophageal reflux disease with a proton pump inhibitor (Chapter 138) provides no more than modest benefit, with approximately one in five patients improving.[2] Inhaled corticosteroids can reduce cough but should be used only after evaluation by chest radiography and often spirometry.[3] Dextromethorphan and codeine-containing cough suppressants can reduce chronic cough by approximately 40%. In adults with refractory chronic cough without active respiratory disease or infection, gabapentin (up to a maximum daily dose of 1800 mg) significantly improves cough-specific quality of life compared with placebo.[4] For chronic refractory cough despite comprehensive evaluation and opioid therapy, a combination of education, breathing exercises, cough suppression techniques, and counseling can significantly reduce cough and its negative impact on quality of life.[5] Coughing can also be reduced by training patients to focus externally rather than internally.[6]

APPROACH TO THE PATIENT WITH WHEEZING

Wheeze is a continuous musical sound that lasts longer than 80 to 100 msec, likely generated by flow through critically narrowed collapsible bronchi. Although expiratory wheezing is a common physical finding in asthma (Chapter 87), the many causes of wheezing (Table 83-4) (e.g., COPD [Chapter 88], pulmonary edema [Chapter 58], bronchiolitis [Chapter 92], bronchiectasis [Chapter 90], and less common entities such as carcinoid [Chapter 232] and parasitic infections) often can be distinguished based on the history, physical examination, and pulmonary function testing (Chapter 85).[4]

DIAGNOSIS

On pulmonary function testing, the shape of inspiratory and expiratory flow-volume loops provide key information about the presence of airway obstruction and whether the obstruction is extrathoracic or intrathoracic (Fig. 83-2). An important cause of extrathoracic obstruction is vocal cord lesions (Chapter 190). Variable intrathoracic obstruction can be caused by tracheomalacia, whereas fixed upper airway obstruction can be caused by a proximal tracheal tumor.

TREATMENT Rx

Treatment of the specific cause will usually lead to complete or at least partial resolution of wheezing. However, treatment of associated asymptomatic or minimally symptomatic gastroesophageal reflux disease is not beneficial.[7]

APPROACH TO THE PATIENT WITH DYSPNEA

Dyspnea is the sensation of difficult, labored, or unpleasant breathing. The word *unpleasant* is very important to this definition because the labored or difficult breathing encountered by healthy individuals while exercising does

TABLE 83-4 DIAGNOSIS OF SELECTED WHEEZING ILLNESSES OTHER THAN ASTHMA

DISEASES	DISTINGUISHING FEATURES
UPPER AIRWAY DISEASES	
Postnasal drip syndrome	History of postnasal drip, throat clearing, nasal discharge; physical examination shows oropharyngeal secretions or cobblestone appearance to mucosa.
Epiglottis	History of sore throat out of proportion to pharyngitis. Evidence of supraglottitis on endoscopy or lateral neck radiographs.
Vocal cord dysfunction syndrome	Lack of symptomatic response to bronchodilators, presence of stridor plus wheeze in absence of increased $P(A\text{-}a)O_2$; extrathoracic variable obstruction on flow-volume loops; paradoxical inspiratory, and/or early expiratory adduction of vocal cords on laryngoscopy during wheezing. This syndrome can masquerade as asthma, be provoked by exercise, and often coexists with asthma.
Retropharyngeal abscess	History of stiff neck, sore throat, fever, trauma to posterior pharynx; swelling noted by lateral neck or CT radiographs.
Laryngotracheal injury due to tracheal cannulation	History of cannulation of trachea by endotracheal or tracheostomy tube; evidence of intrathoracic or extrathoracic variable obstruction on flow-volume loops, neck and chest radiographs, laryngoscopy, or bronchoscopy.
Neoplasms	Bronchogenic carcinoma, adenoma, or carcinoid tumor is suspected when there is hemoptysis, unilateral wheeze, or evidence of lobar collapse on chest radiograph or combinations of these; diagnosis is confirmed by bronchoscopy.
Anaphylaxis	Abrupt onset of wheezing with urticaria, angioedema, nausea, diarrhea, and hypotension, especially after insect bite, in association with other signs of anaphylaxis such as hypotension or hives, or administration of drug or IV contrast, or family history.
LOWER AIRWAY DISEASES	
COPD	History of dyspnea on exertion and productive cough in cigarette smoker. Because productive cough is nonspecific, it should only be ascribed to COPD when other cough-phlegm syndromes have been excluded, forced expiratory time to empty more than 80% of vital capacity >4 sec, and there is decreased breath sound intensity, unforced wheezing during auscultation, and irreversible, expiratory airflow obstruction on spirometry.
Pulmonary edema	History and physical examination consistent with passive congestion of the lungs, ARDS, impaired lung lymphatics; abnormal chest radiograph, echocardiogram, radionuclide ventriculography, cardiac catheterization, or combinations of these.
Aspiration	History of risk for pharyngeal dysfunction or gastroesophageal reflux disease; abnormal modified barium swallow, 24-hr esophageal pH monitoring, or both.
Pulmonary embolism	History of risk for thromboembolic disease, positive confirmatory tests.
Bronchiolitis	History of respiratory infection, connective tissue disease, transplantation, ulcerative colitis, development of chronic airway obstruction over months to a few years rather than over many years in a nonsmoker; mixed obstructive and restrictive pattern on PFTs and hyperinflation; may be accompanied by fine nodular infiltrates on chest radiograph.
Cystic fibrosis	Combination of productive cough, digital clubbing, bronchiectasis, progressive COPD with *Pseudomonas* sp colonization and infection, obstructive azoospermia, family history, pancreatic insufficiency, and two sweat chloride determinations of > 60 mEq/L; some patients are not diagnosed until adulthood, in one instance as late as age 69 yr; when sweat test is occasionally normal, definitive diagnosis may require nasal transepithelial voltage measurements and genotyping.
Carcinoid syndrome	History of episodes of flushing and watery diarrhea; elevated 5-hydroxyindoleactic acid level in 24-hr urine specimen.
Bronchiectasis	History of episodes of productive cough, fever, or recurrent pneumonias; suggestive chest radiographs or typical chest CT findings; ABPA should be considered when bronchiectasis is central.
Lymphangitic carcinomatosis	History of dyspnea or prior malignancy; reticulonodular infiltrates with or without pleural effusions; suggestive high-resolution chest CT scan; confirmed by bronchoscopy with biopsies.
Parasitic infections	Consider in a nonasthmatic patient who has traveled to an endemic area and complains of fatigue, weight loss, fever; peripheral blood eosinophilia; infiltrates on chest radiograph; stools for ova and parasites for nonfilarial causes; blood serologic studies for filarial causes.

ABPA = allergic bronchopulmonary aspergillosis; ARDS = acute respiratory distress syndrome; COPD = chronic obstructive pulmonary disease; CT = computed tomography; IV = intravenous; $P(A\text{-}a)O_2$ = alveolar-arterial oxygen tension gradient; PFTs = pulmonary function tests.

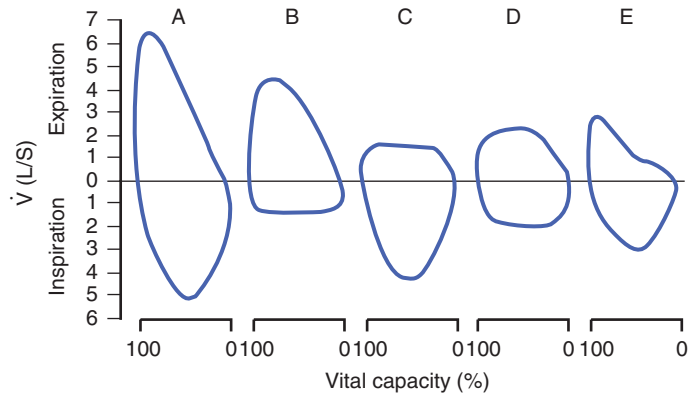

FIGURE 83-2. Schematic flow-volume loop configurations in a spectrum of airway lesions. ***A*** **is normal;** ***B*** **is variable extrathoracic upper airway obstruction;** ***C*** **is variable intrathoracic upper airway lesion;** ***D*** **is fixed upper airway obstruction; and** ***E*** **is small airway obstruction. L/S = liters per second; = ventilation.**

not qualify as dyspnea because it is at the level expected for the degree of exertion. The sensation of dyspnea is often poorly or vaguely described by the patient. The physiology of dyspnea remains unclear, but multiple neural pathways can be involved in processes that lead to dyspnea.

In acute dyspnea, or shortness of breath of sudden onset, the history, physical examination, and laboratory testing must first focus on potential life-threatening conditions, including pulmonary embolism (Chapter 98), pulmonary edema (Chapters 58 and 59), acute airway obstruction from anaphylaxis or foreign bodies, pneumothorax (Chapter 99), or pneumonia (Chapter 97). For chronic dyspnea, specific conditions to consider include COPD (Chapter 88), asthma (Chapter 87), interstitial lung disease (Chapter 92), heart failure (Chapter 58), cardiomyopathy (Chapter 60), GERD (Chapter 138), other respiratory diseases, or hyperventilation syndrome (Table 83-5).

DIAGNOSIS

A chest radiograph, electrocardiogram (ECG), pulmonary function testing, and an exercise test with electrocardiographic monitoring and pulse oximetry at rest and during exercise are key tests to assess patients with unexplained dyspnea (Fig. 83-3).[5] For acute dyspnea, B-type natriuretic peptide testing can be extremely helpful in distinguishing heart failure from other causes. The utility of more detailed pulmonary testing with maximal inspiratory and expiratory pressures, flow-volume loops, with or without methacholine challenge, computed tomographic screening of the chest, and echocardiography depends on history and physical examination and the results of these tests. When GERD is a suspected cause of dyspnea, a modified barium esophagogram or 24-hour esophageal pH monitoring, or both, should be considered (Chapter 138). Other more invasive tests such as cardiac catheterization or lung biopsy may be indicated when the results of less invasive tests have not been conclusive.

TREATMENT

Rx

Whenever possible, the final determination of the cause of dyspnea is made by observing which specific therapy eliminates it. Because dyspnea may be simultaneously the result of more than one condition, it may be necessary to treat more than one condition.

APPROACH TO THE PATIENT WITH HEMOPTYSIS

Hemoptysis is the expectoration of blood from the lung parenchyma or airways.[6] Hemoptysis may be scant, with just the appearance of streaks of bright red blood in the sputum, or massive, with the expectoration of a large volume of blood. Massive hemoptysis, which is defined as the expectoration of at least 600 mL of blood in 24 to 48 hours, may occur in 3 to 10% of patients with hemoptysis. Dark red clots also may be expectorated when the blood has been present in the lungs for days.

Pseudohemoptysis, which is the expectoration of blood from a source other than the lower respiratory tract, may cause diagnostic confusion when patients cannot clearly describe the source of the bleeding. Pseudohemoptysis can occur when blood from the oral cavity, nares, pharynx, or tongue clings to the back of the throat and initiates the cough reflex, or when patients who have hematemesis aspirate into the lower respiratory tract. When the oropharynx is colonized with *Serratia marcescens*, a red-pigment–producing aerobic gram-negative rod, the sputum can also be red and be confused with hemoptysis.

Hemoptysis can be caused by a wide variety of disorders. Virtually all causes of hemoptysis (Table 83-6) may result in massive hemoptysis, but massive hemoptysis is most frequently caused by infection (e.g., tuberculosis [Chapter 324], bronchiectasis and lung abscess [Chapter 90], and cancer [Chapter 191]). Infections with aspergilloma (Chapter 339) and in patients with cystic fibrosis (Chapter 89) also are associated with massive hemoptysis. Iatrogenic causes of massive hemoptysis include rupture of a pulmonary artery after less than 0.2% of cases of balloon-guided flotation catheterization and tracheal artery fistula as a complication of tracheostomy.

In nonmassive hemoptysis, the cause is bronchitis in more than one third of cases (Chapter 96), bronchogenic carcinoma (Chapter 191) in one fifth of cases, tuberculosis (Chapter 324) in 7%, pneumonia (Chapter 97) in 5%, and bronchiectasis in 1% (Chapter 90). Using a systematic diagnostic approach (see later), the cause of hemoptysis can be found in 68 to 98% of cases. The remaining 2 to 32% have idiopathic or central hemoptysis, which occurs most commonly in men between 30 and 50 years of age. Prolonged follow-up of idiopathic hemoptysis almost always fails to reveal the source of bleeding, even though 10% continue to have occasional episodes of hemoptysis.

DIAGNOSIS

The diagnostic evaluation for hemoptysis begins with a detailed medical history and a complete physical examination. Information on the amount of

TABLE 83-5 DISEASES THAT CAUSE DYSPNEA GROUPED BY PHYSIOLOGIC MECHANISMS OF ACTION*

INCREASED RESPIRATORY DRIVE
Stimulation of Chemoreceptors
Conditions leading to acute hypoxemia
Impaired gas exchanger (e.g., asthma, pulmonary embolism, pneumonia, congestive heart failure†)
Environmental hypoxia (e.g., altitude, contained space with fire)
Conditions leading to increased dead space, acute hypercapnia
Impaired gas exchanger (e.g., acute, severe asthma; exacerbation of COPD; severe pulmonary edema)
Impaired ventilator pump (e.g., muscle weakness, airflow obstruction)
Metabolic acidosis
Renal disease (e.g., renal failure, renal tubular acidosis)
Decreased oxygen carrying capacity (e.g., anemia)
Decreased release of oxygen to tissues (e.g., hemoglobinopathy)
Decreased cardiac output
Stimulation of Pulmonary Receptors (irritant, mechanical, vascular)‡
Interstitial lung disease
Pleural effusion (compression atelectasis)
Pulmonary vascular disease (e.g., thromboembolism, idiopathic pulmonary hypertension)
Heart failure
Mild asthma
Behavioral Factors
Hyperventilation syndrome, anxiety disorders, panic attacks
VENTILATORY PUMP: INCREASED EFFORT OR WORK OF BREATHING
Muscle Weakness
Myasthenia gravis, Guillain-Barré syndrome, spinal cord injury, myopathy, postpoliomyelitis syndrome
Decreased compliance of the chest wall
Severe kyphoscoliosis, obesity, pleural effusion
Airflow Obstruction (including increased resistive load from narrowing of the airways and increased elastic load from hyperinflation)
Asthma, COPD, laryngospasm, aspiration of foreign body, bronchitis

*Some diseases appear in more than one category, because they act via several physiologic mechanisms.
†Heart failure includes both systolic and diastolic dysfunction. Systolic dysfunction may produce dyspnea at rest and with activity. Diastolic dysfunction typically leads to symptoms primarily with exercise. In addition to the mechanisms noted above, systolic heart failure may also produce dyspnea via metaboreceptors, which are postulated to exist in muscles and be stimulated by changes in the metabolic milieu when oxygen delivery does not meet oxygen demand.
‡These conditions probably produce dyspnea by a combination of increased ventilator drive and primary sensory input from the receptors.
COPD = chronic obstructive pulmonary disease.

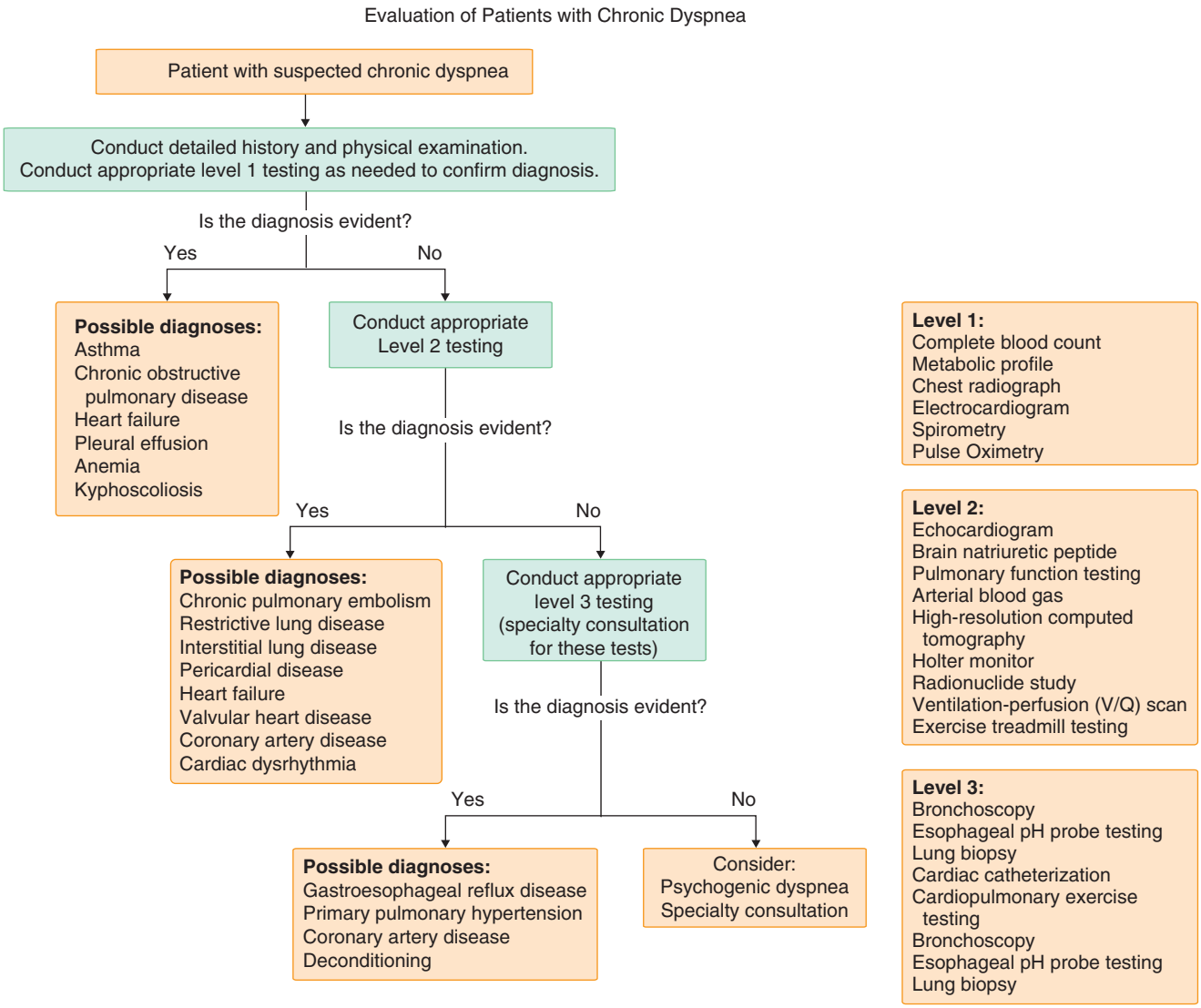

FIGURE 83-3. **Algorithm outlining the approach to chronic dyspnea.** (Modified from Karnani NG, Reisfield GM, Wilson GR. Evaluation of chronic dyspnea. *Am Fam Phys.* 2005;71:1529-1537.)

TABLE 83-6 COMMON CAUSES OF MASSIVE HEMOPTYSIS

Cardiovascular
Arterial bronchial fistula
Heart failure, especially from mitral stenosis
Pulmonary arteriovenous fistula
Diffuse intrapulmonary hemorrhage
Diffuse parenchymal disease
Iatrogenic
Malposition of chest tube
Pulmonary artery rupture following pulmonary arterial catheterization
Tracheoarterial fistula
Infections
Aspergilloma
Bronchiectasis
Bronchitis
Cystic fibrosis
Lung abscess
Sporotrichosis
Tuberculosis
Malignancies
Bronchogenic carcinoma
Leukemia
Metastatic cancer
Trauma
Drugs and toxins
Penicillamine
Solvents
Crack cocaine
Trimelletic anhydride
Bevacizumab

bleeding should be obtained, as well as details about the frequency, timing, and duration of hemoptysis. For example, repeated episodes of hemoptysis occurring over a period of months to years suggest a bronchial adenoma or bronchiectasis as the cause, whereas small amounts of hemoptysis occurring every day for weeks are more likely to be caused by bronchogenic carcinoma. A travel history can suggest coccidioidomycosis (Chapter 333) and histoplasmosis (Chapter 332) in the United States, paragonimiasis and ascariasis (Chapter 358) in the Far East, and schistosomiasis (Chapter 355) in South America. Orthopnea and paroxysmal nocturnal dyspnea suggest heart failure (Chapter 58), especially from mitral stenosis (Chapter 75). In patients who have occupational exposure to trimellitic anhydride, which occurs when heated metal surfaces are sprayed with a corrosion-resistant epoxy resin, hemoptysis can be part of the postexposure syndrome. In a patient with the triad of upper airway disease, lower airway disease, and renal disease, granulomatosis with polyangiitis (Chapter 270) should be suspected. Pulmonary hemorrhage also may be a presenting manifestation of systemic lupus erythematosus (Chapter 266). Goodpasture syndrome, which typically occurs in young men, is also associated with renal disease (Chapter 121). Diffuse alveolar hemorrhage occurs in 20% of cases during autologous bone marrow transplantation (Chapter 178) and should be suspected in patients who have undergone recent bone marrow transplantation when they present with cough, dyspnea, hypoxemia, and diffuse pulmonary infiltrates.

On physical examination, inspection of the skin and mucous membranes may show telangiectasias suggesting hereditary hemorrhagic telangiectasia (Chapter 173) or ecchymoses and petechiae, suggesting a hematologic abnormality (Chapter 172). Pulsations transmitted to a tracheostomy cannula should heighten suspicion of a tracheal artery fistula. Inspection of the thorax should show evidence of recent or old chest trauma, and unilateral

wheeze or crackles may herald localized disease such as a bronchial adenoma or carcinoma. Although pulmonary embolism (Chapter 98) cannot be definitively diagnosed on physical examination, tachypnea, phlebitis, and pleural friction rub suggest this disorder. If crackles are heard on the chest examination, heart failure as well as other diseases causing diffuse pulmonary hemorrhage (see earlier) or idiopathic pulmonary hemosiderosis (Chapter 92) should be considered. Careful cardiovascular examination may help diagnose mitral stenosis (Chapter 75), pulmonary artery fistulas, or pulmonary hypertension (Chapter 68).

Routine laboratory studies should include a complete blood count, urinalysis, and coagulation studies. The complete blood count may suggest an infection, hematologic disorder, or chronic blood loss. Urinalysis may reveal hematuria and suggest the presence of a systemic disease (e.g., Wegener granulomatosis, Goodpasture syndrome, systemic lupus erythematosus) associated with renal disease. Coagulation studies may uncover a hematologic disorder that is primarily responsible for hemoptysis or that contributes to excessive bleeding from another disease. The ECG may help suggest the presence of a cardiovascular disorder. Although as many as 30% of patients with hemoptysis have a normal chest radiograph, routine chest radiographs may be diagnostically valuable.

Bronchoscopy can localize the bleeding site in up to 93% of patients by fiberoptic bronchoscopy and in up to 86% with rigid bronchoscopy.[7] It may establish sites of bleeding different from those suggested by the chest radiograph. The best results are obtained when bronchoscopy is performed during or within 24 hours of active bleeding, and rates of diagnosis fall to approximately 50% by 48 hours after bleeding. When there is no active bleeding, bronchoscopy with bronchoalveolar lavage can be helpful in patients thought to have diffuse intrapulmonary hemorrhage. Typical findings include bright red or blood-tinged lavage fluid from multiple lobes in both lungs or a substantial number of hemosiderin-laden macrophages (i.e., at least 20% of the total number of alveolar macrophages).

Depending on the results of the initial evaluation and the likely categories of hemoptysis, additional diagnostic tests can be helpful (Table 83-7). Bronchoscopy may not be needed in patients who have stable chronic bronchitis (Chapter 88) with one episode of blood streaking or who have acute tracheobronchitis (Chapter 88). Bronchoscopy also may not be needed with obvious cardiovascular causes of hemoptysis, such as heart failure and pulmonary embolism.

TABLE 83-7 EXAMPLES OF SPECIAL EVALUATIONS FOR HEMOPTYSIS ACCORDING TO CATEGORY OF DISEASE*

TRACHEOBRONCHIAL DISORDERS
Expectorated sputum for TB, parasites, fungi, and cytology Bronchoscopy (if not done) High-resolution chest CT scan
LOCALIZED PARENCHYMAL DISEASES
Expectorated sputum for TB, parasites, fungi, and cytology Chest CT scan Lung biopsy with special stains
DIFFUSE PARENCHYMAL DISEASES
Expectorated sputum for cytology Blood for BUN, creatinine, ANA, RF, complement, cryoglobulins, ANCA, anti-GBM antibody Lung or kidney biopsy with special stains
CARDIOVASCULAR DISORDERS
Echocardiogram Arterial blood gas on 21% and 100% oxygen Ventilation-perfusion scans Pulmonary arteriogram Aortogram, contrast-enhanced CT scan
HEMATOLOGIC DISORDERS
Coagulation studies Bone marrow

*This table is not meant to be all inclusive.
ANA = antinuclear antibody; ANCA = antineutrophil cytoplasmic antibody; BUN = blood urea nitrogen; CT = computed tomography; GBM = glomerular basement membrane; RF = rheumatoid factor; TB = tuberculosis.

TREATMENT

Rx

Treatment is targeted toward the cause of hemoptysis. Bronchoscopic approaches (Chapter 101) are increasingly used for endobronchial lesions.

Grade A References

A1. Smith SM, Schroeder K, Fahey T. Over-the-counter (OTC) medications for acute cough in children and adults in community settings. *Cochrane Database Syst Rev*. 2014;11:CD001831.
A2. Shaheen NJ, Crockett SD, Bright SD, et al. Randomised clinical trial: high-dose acid suppression for chronic cough—a double-blind, placebo-controlled study. *Aliment Pharmacol Ther*. 2011;33:225-234.
A3. Johnstone KJ, Chang AB, Fong KM, et al. Inhaled corticosteroids for subacute and chronic cough in adults. *Cochrane Database Syst Rev*. 2013;3:CD009305.
A4. Ryan NM, Birring SS, Gibson PG. Gabapentin for refractory chronic cough: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2012;380:1583-1589.
A5. Chamberlain S, Birring SS, Garrod R. Nonpharmacological interventions for refractory chronic cough patients: systematic review. *Lung*. 2014;192:75-85.
A6. Janssens T, Silva M, Davenport PW, et al. Attentional modulation of reflex cough. *Chest*. 2014;146:135-141.
A7. Mastronarde JG, Anthonisen NR, Castro M, et al. Efficacy of esomeprazole for treatment of poorly controlled asthma. *N Engl J Med*. 2009;360:1487-1499.
A8. Lam LL, Cameron PA, Schneider HG, et al. Meta-analysis: effect of B-type natriuretic peptide testing on clinical outcomes in patients with acute dyspnea in the emergency setting. *Ann Intern Med*. 2010;153:728-735.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bohadana A, Izbicki G, Kraman SS. Fundamentals of lung auscultation. *N Engl J Med*. 2014;370:744-751.
2. Goldsobel AB, Kelkar PS. The adult with chronic cough. *J Allergy Clin Immunol*. 2012;130:825-825.e6.
3. Dicpinigaitis PV, Morice AH, Birring SS, et al. Antitussive drugs: past, present, and future. *Pharmacol Rev*. 2014;66:468-512.
4. Busse WW. What is the best pulmonary diagnostic approach for wheezing patients with normal spirometry? *Respir Care*. 2012;57:39-46.
5. Parshall MB, Schwartzstein RM, Adams L, et al. An official American Thoracic Society statement: update on the mechanisms, assessment, and management of dyspnea. *Am J Respir Crit Care Med*. 2012;185:435-452.
6. Hurt K, Bilton D. Haemoptysis: diagnosis and treatment. *Acute Med*. 2012;11:39-45.
7. Sakr L, Dutau H. Massive hemoptysis: an update on the role of bronchoscopy in diagnosis and management. *Respiration*. 2010;80:38-58.

REVIEW QUESTIONS

1. Which of the following is true about cough variant asthma?

A. The diagnosis requires an improvement in the cough with traditional asthma medications.
B. It does not reverse with a bronchodilator.
C. Sputum eosinophils are absent.
D. The cough is not responsive to inhaled corticosteroids.
E. Hemoptysis has been present on two occasions.

Answer: A Cough variant asthma responds to asthma medications, in contrast to other causes of cough, which may be either partially or not affected by traditional asthma medications.

2. A 53-year-old woman returns for reevaluation of cough of 1 year duration. The cough is occasionally productive, occurs during the day and night, and is triggered by talking, laughing, and cold air. Her pulmonary function tests are normal. The cough has not responded to appropriate therapy for asthma with inhaled corticosteroids and short-acting β-agonists. She denies any history of lung disease, cigarette smoking, atopy, rhinitis, gastroesophageal reflux disease, or obstructive sleep apnea. The physical examination is unrevealing. Further evaluation, including esophageal impedance and manometry, were unremarkable. What is the appropriate next step in this patient's management?

A. High-resolution computed tomography of the chest
B. Computed tomography of the sinuses
C. Bronchoscopy
D. Echocardiogram
E. Video laryngeal swallowing study

Answer: A With a cough that is productive, even occasionally, an assessment of airway structure via computed tomography should be performed first before bronchoscopy to assess for bronchiectasis, sarcoidosis, or occult interstitial lung disease. The patient has no history of atopy or rhinitis, so sinus computed tomography is not indicated. A cough due to cardiac causes is less likely to be productive, so an echocardiogram is not indicated. The likely diagnosis is mild bronchiectasis causing cough.

3. A 55-year-old man presents with dyspnea on exertion beginning approximately 2 months ago. He denies chest pain, cough, wheezing, chest tightness, or a history of cardiovascular disease, lung disease, diabetes, or cigarette smoking. He takes medication daily for hypertension and gastroesophageal reflux disease. He does not exercise regularly. His BMI is 36 kg/m^2. On physical examination his blood pressure is 135/70, and his other vital signs and examination are normal. Pulmonary function tests, chest radiograph, complete blood count, and brain natiuretic peptide levels are normal. What is the next appropriate step in this patient's evaluation?

A. Echocardiogram
B. Pulse oximetry with ambulation (6-minute walk test)
C. Exercise treadmill test
D. High-resolution chest computed tomography
E. 24-hour Holter monitoring

Answer: C This patient likely has coronary artery disease manifesting as dyspnea on exertion. His risk factors include hypertension, obesity, and lack of exercise. With a normal chest radiograph and pulmonary function tests, primary lung diseases are less likely. A normal brain natiuretic peptide level makes reduced cardiac function less likely, and anemia is excluded by a normal complete blood count. Cardiovascular disease is the most likely diagnosis, and additional testing likely will be required if the exercise treadmill test is unrevealing.

4. A 40-year-old woman with a history of atopy, rhinitis, and gastroesophageal reflux disease (GERD) presents with wheezing of 3 months duration. The wheezing occurs primarily during the day, is triggered by cold air and exercise, and has not responded to a 6-week course of an inhaled corticosteroid and short-acting β-agonist. She denies chest tightness, chest pain, or a history of lung or cardiovascular disease. Her history is notable for allergies to dust, multiple weeds, grasses, and animal dander. She has persistent rhinitis for which she performs nasal saline rinses and uses intranasal corticosteroids daily. She also takes a proton pump inhibitor once daily for treatment of GERD, and these symptoms are well controlled. Her physical examination reveals normal vital signs and erythema of the upper airway without nasal polyps. The remainder of her physical examination is normal without wheezing. Pulmonary function tests, chest radiograph, brain natriuretic peptide, and methacholine challenge testing are normal. What is the next step in this patient's evaluation?

A. Sinus computed tomography
B. High-resolution chest computed tomography
C. Direct laryngoscopy after exercise
D. Echocardiogram
E. Bronchoscopy with biopsy

Answer: C This presentation suggests vocal cord dysfunction presenting as wheezing. Wheezing in the setting of rhinitis and GERD suggests an airway process as the most likely diagnosis. Because the patient has not responded to therapy for asthma and a methacholine challenge is negative, asthma is highly unlikely. A normal BNP makes cardiac dysfunction less likely. In the setting of rhinitis and GERD, evaluation of the upper airway to exclude vocal cord dysfunction during an episode of wheezing is the next appropriate step. An echocardiogram should be obtained if direct laryngoscopy is negative.

84

IMAGING IN PULMONARY DISEASE

PAUL STARK

IMAGING OF THE LUNGS, MEDIASTINUM, AND CHEST WALL

EPIDEMIOLOGY

Worldwide, chest radiography is the most commonly performed imaging procedure; more than 75 million chest radiographs are performed every year in the United States alone. Chest radiographs provide useful information about the patient's anatomy and disease at a minimal monetary cost and with radiation exposure that most experts agree is negligible (0.05 to 0.1 mSv) (Chapter 20). Although many novel imaging techniques are available, the conventional chest radiograph remains invaluable in the initial assessment of disorders of the lung, pleura, mediastinum, and chest wall.

Imaging Techniques

The standard chest radiograph is performed at 2 m from the x-ray tube focal spot to the image detector, in frontal and lateral projections. If possible, the radiographs should be obtained with the patient inhaling to total lung capacity. These images, which provide views of the lungs, mediastinum, and chest wall simultaneously, are typically acquired, stored, and distributed digitally.

Bedside Radiography

Although bedside radiography accounts for a large number of chest radiographs, especially in the intensive care unit (ICU), the images obtained are generally of lower technical quality, cost more, and are more difficult to interpret. Lung volumes are low, thereby leading to crowding of vascular structures, and the low kilovoltage technique required for the mobile equipment yields radiographs with overexposed lungs and an underpenetrated mediastinum. The anteroposterior projection and the slightly lordotic angulation of the x-ray beam combine to distort the basal lung structures and magnify

the cardiac silhouette. Recumbent studies also make recognition of pleural effusions or pneumothoraces more difficult. In the ICU, chest radiography can be ordered selectively rather than as a daily routine, without compromising care.[1]

Computed Tomography

Computed tomography (CT) has multiple advantages over conventional radiography. It displays cross-sectional anatomy free of superimposition, with a 10-fold higher contrast resolution. Multislice CT scanners acquire a continuous, volumetric, near-isotropic data set with possibilities for high-quality two-dimensional or three-dimensional reformatting (volume rendering) in any plane. High-resolution CT of the lung parenchyma is an important application; narrow collimation of the beam combined with an edge-enhancing high spatial frequency algorithm results in exquisite detail of normal and abnormal lungs, and correlation with pathologic anatomy is high.

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) depends on the magnetic properties of hydrogen atoms. Magnetic coils and radio frequency coils lead to induction, excitation, and eventual readout of magnetized protons. The molecular environment of hydrogen atoms will affect the rate at which they release energy; this energy yields a spatial distribution of signals that is converted into an image by computer algorithms, similar to CT. Because of its soft tissue specificity, MRI has applications in the assessment of chest wall invasion, mediastinal infiltration, and diaphragmatic involvement by lung cancer or malignant mesothelioma.

Positron Emission Tomography

Fluorodeoxyglucose positron emission tomography (FDG-PET) uses labeled fluorodeoxyglucose to image the glycolytic pathway of tumor cells or other metabolically active tissues with affinity for glucose. This technique has proved helpful in studying intrathoracic tumors and has facilitated the work-up of solitary pulmonary nodules. Integrated PET-CT scans have improved the diagnosis and staging of intrathoracic tumors.[2]

Ultrasonography

Outside the heart, ultrasonography plays only a limited role in thoracic imaging. Its primary use is to localize pleural effusions and guide their drainage (Chapter 99). In the intensive care setting, ultrasound also may help with the diagnosis of pneumothorax and diffuse alveolar damage.

Evaluation of Chest Images

Images of the chest are best evaluated by examining regions of the lung for specific findings and relating these findings to known diagnostic groups. A number of critical radiographic features should be considered, with an appreciation for the known causes of these changes.

Diffuse Lung Disease

Diffuse lung disease is an overall term for a number of related abnormal parenchymal radiographic patterns. Although radiologists have attempted to separate alveolar from interstitial lung disease radiographically, this distinction is no longer recommended because the correlation between the radiographic localization to a compartment and the actual histopathologic findings is relatively poor.[3] For example, nodular patterns can be produced by either interstitial or alveolar disease. Conversely, so-called alveolar disease processes can induce an interstitial reaction. Ground-glass opacities can be induced by either alveolar or interstitial disease. Air bronchograms, the presumed paradigm of air space disease, can be identified in a small percentage of patients with predominantly interstitial lung disease, such as sarcoidosis, pulmonary lymphoma, and pulmonary calcinosis.

Because of such limitations, a graphically descriptive approach that combines analysis of predominant opacities, assessment of lung expansion, and distribution and profusion of disease yields a differential diagnosis. The term *infiltrate* should be avoided; instead, the term *pulmonary opacities* should be used, with opacities further classified as large (i.e., >1 cm in largest dimension) or small (i.e., <1 cm in diameter).

Large Opacities

Large opacities (Table 84-1) are characterized according to their distribution. Diffuse homogeneous opacities are typical for diffuse alveolar damage (Fig. 84-1), increased permeability (noncardiogenic) pulmonary edema, diffuse viral pneumonia (see Fig. 84-1B), or *Pneumocystis jirovecii* pneumonia. Multifocal patchy opacities (see Fig. 84-1C) are found in multifocal bronchopneumonia, recurrent aspiration, or vasculitis (Fig. 84-E1). Lobar opacities without atelectasis are typically seen in lobar pneumonia (Figs. 84-2, 84-E2, and 84-E3). Lobar opacities with atelectasis often result from obstruction of a lobar bronchus by foreign bodies, tumors, or mucous plugs. Perihilar opacities are seen in hydrostatic pulmonary edema as a result of left-sided heart failure (Figs. 84-3A and B; 84-E4), renal failure, volume overload, or pulmonary hemorrhage.

TABLE 84-1 CLASSIFICATION OF LARGE PULMONARY OPACITIES

Diffuse homogeneous
Multifocal patchy
Lobar without atelectasis
Lobar with atelectasis
Perihilar
Peripheral

TABLE 84-2 PATTERNS OF SMALL PULMONARY OPACITIES

Micronodular
Acinar
Linear
Reticular
Bronchial
Arterial
Destructive

Small Opacities

In contrast to the large pulmonary opacities, a number of radiographic patterns characterize small pulmonary opacities in diffuse lung disease. It is helpful to differentiate small nodular, linear, reticular, or combined patterns (Table 84-2).[4]

Nodular Patterns

Micronodular opacities, which include nodules 1 mm and smaller in diameter, can result from talc granulomatosis in intravenous drug abusers (Chapter 34), alveolar microlithiasis, rare cases of silicosis, talcosis, coal workers' pneumoconiosis (Chapter 93), and beryllium-induced lung diseases (Chapter 93), as well as from occasional cases of sarcoidosis (Chapter 95) or hemosiderosis. The nodular pattern includes nodules up to 1 cm in diameter. Frequent causes include infections or inflammatory granulomata such as miliary tuberculosis (Chapter 324), sarcoidosis (Chapter 95), fungal diseases, hypersensitivity pneumonia, and Langerhans cell histiocytosis (Chapter 92).

Linear Patterns

Linear patterns, also called Kerley lines, are mostly a reflection of thickened interlobular septa. Kerley A lines, which radiate 2 to 4 cm from the hilum toward the pulmonary periphery and particularly toward the upper lobes (Fig. 84-4), reflect thickening of the axial interstitial compartment and can be a feature of left ventricular failure or allergic reactions. Kerley B lines, which reflect thickening of the subpleural interstitial compartment, typically are approximately 1 cm in length and 1 mm in thickness and usually found in the periphery of the lower lobes, abutting the pleura. The B lines are characteristic of subacute and chronic left ventricular failure (Chapter 58), mitral valve disease (Chapter 75), lymphangitic carcinomatosis, viral pneumonia, and pulmonary fibrosis (Chapter 92). Kerley C lines, which are rarely diagnosed by radiologists, result from thickening of the lung parenchymal interstitium and form a reticular pattern on chest radiographs.

Reticular Patterns

Reticular patterns are small polygonal, irregular, or curvilinear opacities on chest radiographs (Fig. 84-5). The differential diagnosis varies according to the timeline of the pathologic change. Acute onset of a reticular pattern can occur in interstitial pulmonary edema (e.g., due to left-sided heart failure), atypical pneumonias (e.g., viral or mycoplasmal pneumonia), early exudative changes in a connective tissue disorder (e.g., systemic lupus

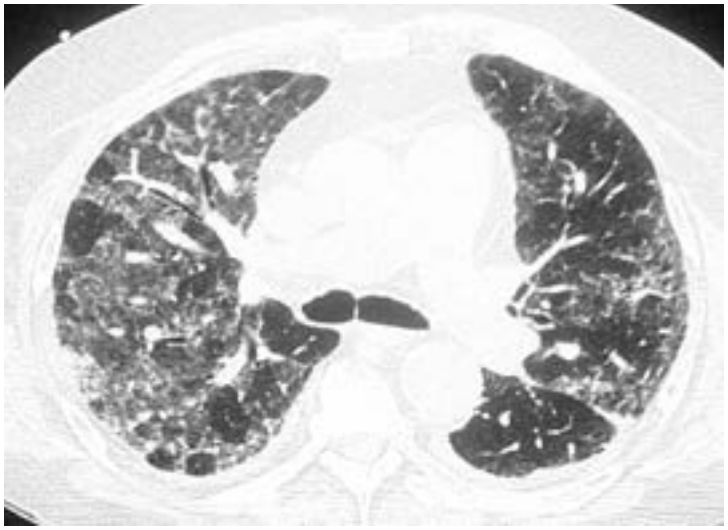

E-FIGURE 84-1. Microscopic polyangitis with pulmonary hemorrhage. Computed tomographic scan shows heterogenous ground-glass opacification with mosaic attenuation.

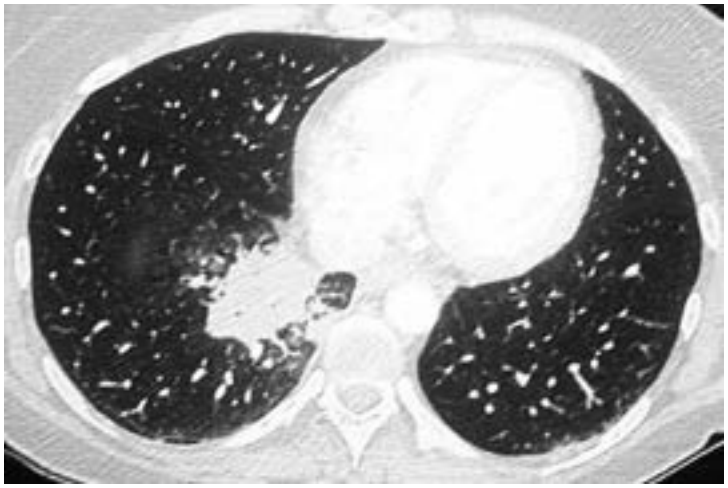

E-FIGURE 84-2. Primary coccidioidomycosis pneumonia. Computed tomographic scan shows segmental consolidation of the mediobasal segment of the right lower lobe.

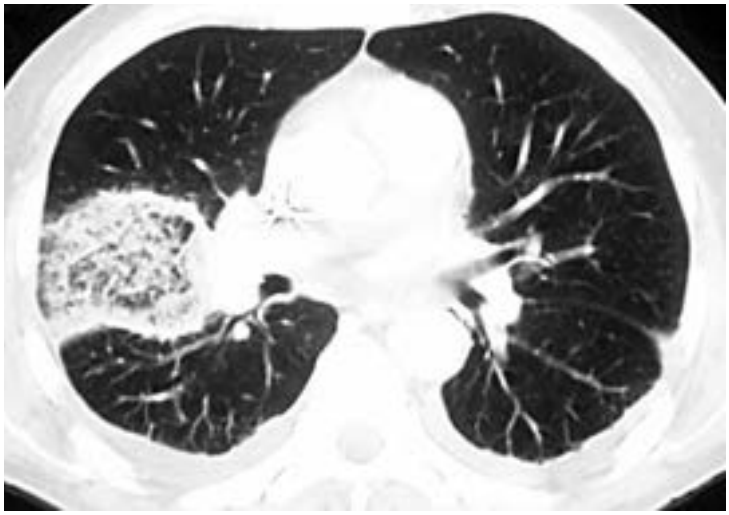

E-FIGURE 84-3. Mucormycosis in an immunocompromised patient. Computed tomographic scan shows a focal right upper lobe opacity with an "inverse halo sign," typical of either focal cryptogenic pneumonia or mucormycosis.

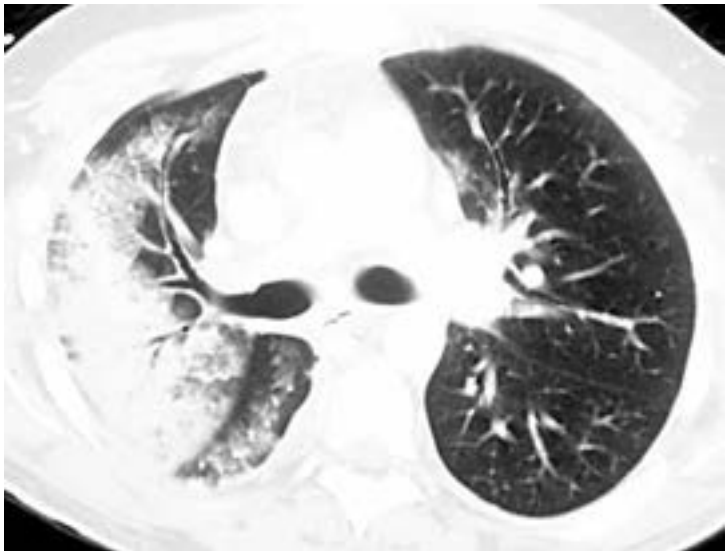

E-FIGURE 84-4. Unilateral reexpansion pulmonary edema after evacuation of right pneumothorax. Computed tomographic scan shows unilateral right-sided consolidation.

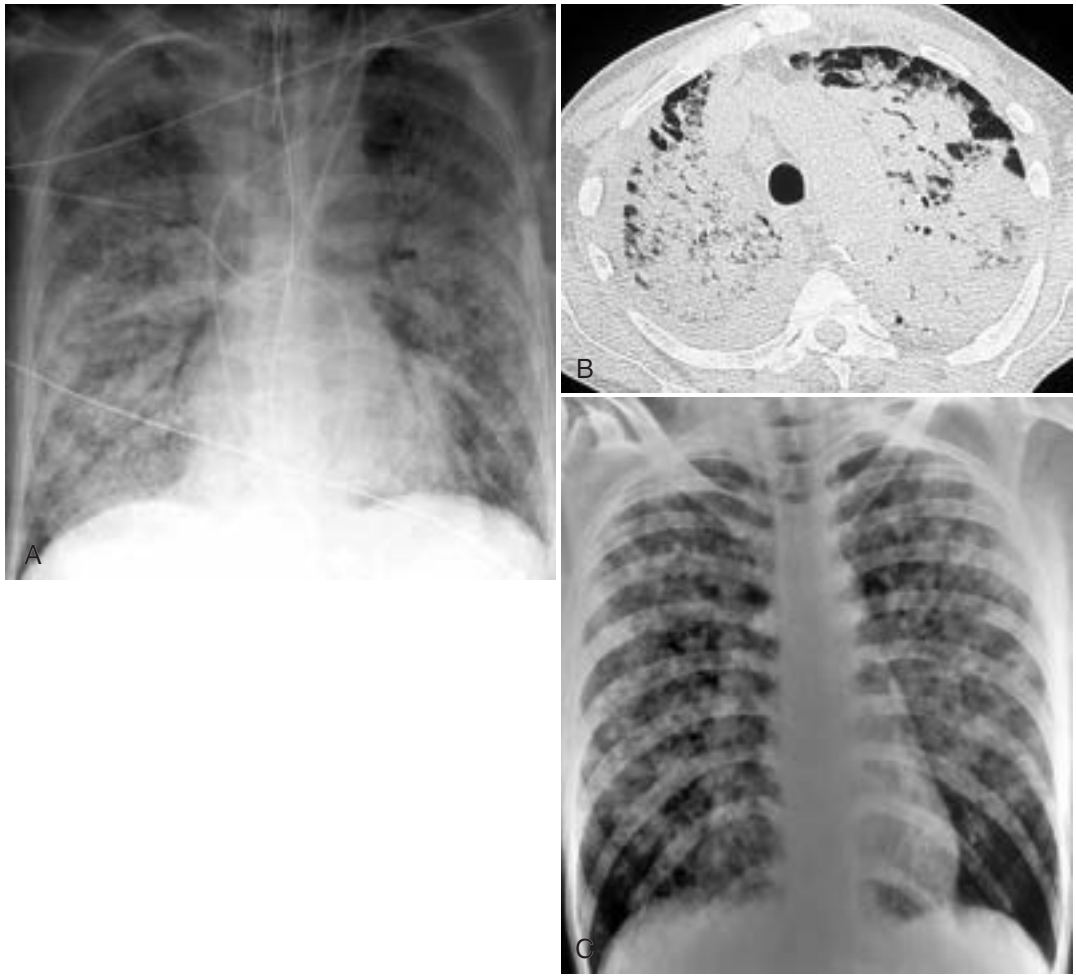

FIGURE 84-1. Diffuse alveolar damage. A, Chest radiograph shows diffuse homogeneous opacification of both lungs with clearly visible air bronchograms. **B,** Computed tomographic scan demonstrates diffuse consolidation with air-bronchograms extending to the periphery of both lungs in a patient with community-acquired pneumonia. **C,** Patient with acute varicella pneumonia. Chest radiograph demonstrates multiple acinar nodules with tendency for confluence, yielding multifocal patchy parenchymal opacification.

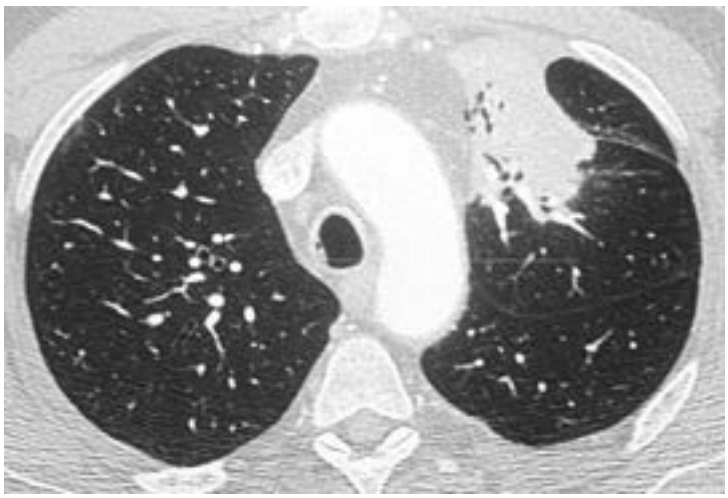

FIGURE 84-2. Bacterial pneumonia in the left upper lobe. Computed tomographic scan demonstrates segmental consolidation of the anterior segment of the left upper lobe with air-bronchograms.

erythematosus; Chapter 266), and acute allergic reactions (e.g., transfusion reactions [Chapter 177] or reactions to *Hymenoptera* stings). The common chronic processes resulting in a reticular pattern are idiopathic interstitial pneumonias (Chapter 92), connective tissue diseases (particularly scleroderma and rheumatoid lung), asbestosis (Chapter 93), radiation fibrosis (Chapter 92), end-stage hypersensitivity pneumonia (Chapters 92 and 93), drug reactions, lymphangitic spread of cancer, end-stage granulomatous infection, lymphoma in its bronchovascular form, Kaposi sarcoma in its bronchovascular manifestation, and sarcoidosis.

Honeycombing

Honeycombing, which is an indication of end-stage interstitial lung disease (Chapter 92), reflects a restructuring of pulmonary anatomy accompanied by bronchiolectasis. Honeycombs (Fig. 84-6) form a multilayer of small subpleural spaces between 3 and 10 mm in diameter. They can be distinguished from paraseptal emphysema by their thicker wall and multiple layers.

Alveolar Pattern

An alveolar (Chapter 91) or air space pattern is characterized by acinar nodules, 0.6 to 1 cm in diameter. These nodules encompass more than the acinus, in the strict anatomic sense, with surrounding peribronchiolar lung tissue. Other patterns include ground-glass opacities (a reflection of incomplete alveolar filling), coalescent large opacities, consolidation involving whole lobes or segments, opacification in a bronchocentric distribution, air bronchograms, and air alveolograms. These radiographic features are helpful in placing a disease into a particular radiologic category, but the radiographic pattern called *alveolar* does not simply correspond to exclusive histologic alveolar filling because the interstitial compartment is involved as well in most cases. A more accurate description is parenchymal rather than alveolar opacification or consolidation.

Bronchial Patterns

Bronchial patterns, as best depicted by diffuse bronchiectasis (Chapter 90), are seen on conventional radiographs as linear, tubular, or cystic lucencies and opacities that follow the expected path of bronchi, so-called tramlines because they resemble tram tracks. Mucoid impaction, as seen in patients

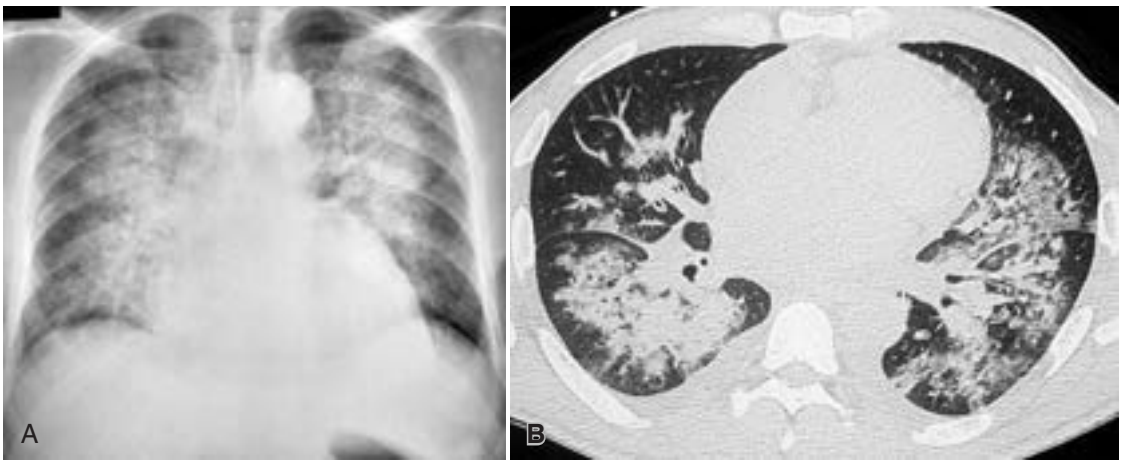

FIGURE 84-3. A, Pulmonary edema. Chest frontal radiograph demonstrates classic "batwing" distribution of hydrostatic pulmonary edema. **B,** Computed tomographic scan shows bilateral perihilar consolidations and a small right pleural effusion in a patient with pulmonary edema.

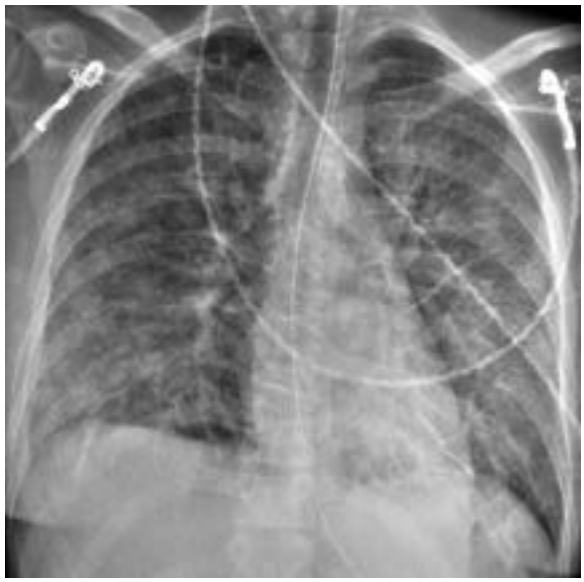

FIGURE 84-4. Patient with known transfusion reaction. Chest radiograph displays ground-glass opacification of both lungs and bilateral Kerley A lines, presenting as long linear structures extending from the hilar regions into the pulmonary periphery.

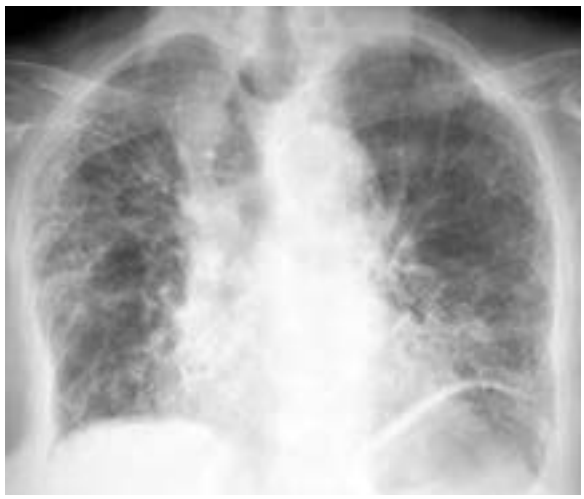

FIGURE 84-5. Diffuse reticular lung disease. Chest radiograph in a 94-year-old patient with diffuse reticular opacities due to idiopathic pulmonary fibrosis with honeycombing and traction bronchiectases. The lung volumes are typically reduced by a decreased pulmonary compliance.

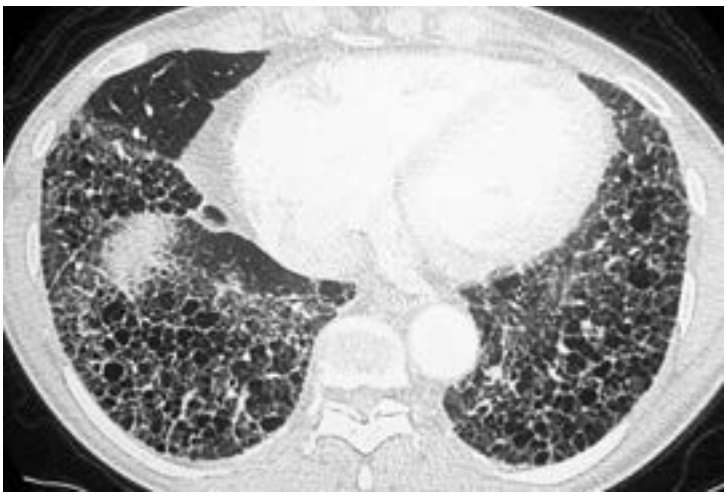

FIGURE 84-6. Honeycombing in a patient with idiopathic pulmonary fibrosis and usual interstitial pneumonia. Computed tomographic scan shows multiple bibasal reticular opacities with honeycombing and traction bronchiectases.

with asthma, allergic bronchopulmonary aspergillosis, or plastic bronchitis, leads to opacities described as toothpaste, cluster of grapes, or finger-in-glove. The "dirty lung" pattern seen in smokers with chronic bronchitis (Chapter 88) results from bronchial wall thickening, peribronchial fibrosis, respiratory bronchiolitis, and pulmonary arterial hypertension.

Vascular Patterns

Arterial patterns reflect changes in pulmonary perfusion. The term *caudalization* reflects the normal blood flow distribution pattern in an upright person in which the basal pulmonary vessels are two to three times wider than the upper lobe vasculature. *Cephalization,* in which the ratios of diameters of vessels are reversed, is frequently seen in recumbent persons, in whom it may be considered normal; however, when it is present in individuals imaged in the upright position, it indicates left ventricular failure, mitral valve disease, or basal emphysema (Fig. 84-7). Equalization, or balanced flow with well-demonstrated vessels to upper and lower lung zones, is found in hyperkinetic circulation due to anemia, obesity, pregnancy, hyperthyroidism, or left-to-right shunts. Equalization or balanced flow with oligemia can be seen in hypovolemia, diffuse emphysema, or right-to-left shunts. Centralization reflects dilation of central pulmonary vessels, with accompanying normal or diminished peripheral circulation. Typically, it is seen in pulmonary arterial hypertension (Fig. 84-8). Lateralization of flow, favoring one lung over the other, also called *asymmetrical perfusion*, is visible with unilateral emphysema, unilateral bronchiolitis obliterans (Swyer-James-McLeod syndrome), or unilateral obstruction of the pulmonary artery. Locally dilated pulmonary vessels occur adjacent to affected oligemic lung regions in patchy emphysema, multiple pulmonary emboli, arteriovenous malformations, and

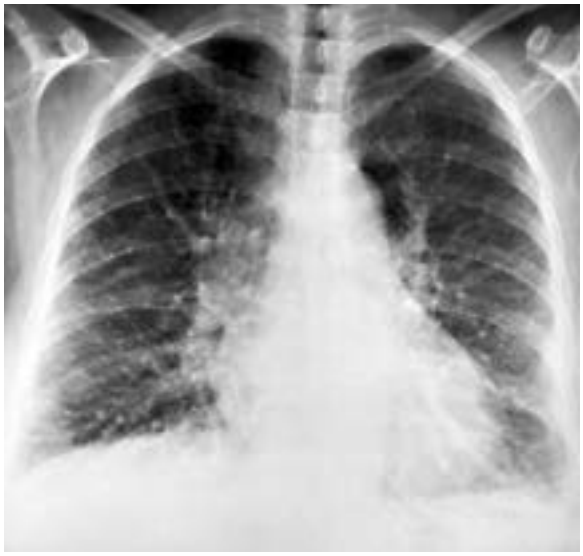

FIGURE 84-7. Patient with left ventricular failure. Chest frontal radiograph shows cephalization of pulmonary blood flow.

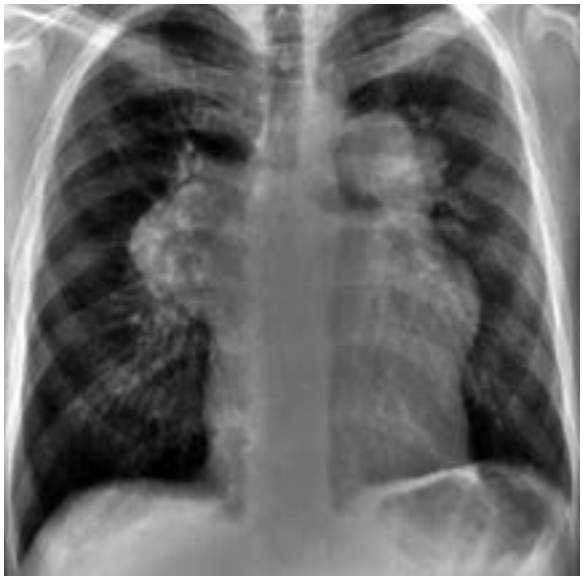

FIGURE 84-8. Patient with primary pulmonary arterial hypertension. Chest frontal radiograph shows centralization of flow with pulmonary artery aneurysms and peripheral pulmonary oligemia.

nonuniform bronchiolitis obliterans. This pattern produces mosaic attenuation on high-resolution CT scanning. Focal oligemia with vascular deficiency is characteristically seen in emphysema. Centrilobular emphysema, paraseptal emphysema, and bullous lung disease have a predilection for the upper lung regions, whereas panlobular emphysema induces basal oligemia with vascular deficiency.

Lung Volume

Conventional radiographs and CT scans are performed during a breath hold at full inspiration and total lung capacity. Low lung volumes are inferred by the high position of the diaphragm and the crowding of basal vascular structures (Table 84-3). Lung volumes larger than expected are commonly found in patients with diffuse emphysema (Fig. 84-9) (Chapter 88), chronic asthma (Chapter 87), or diffuse bronchiolitis obliterans and in highly trained athletes. With a few rare exceptions, chronic diffuse infiltrative lung diseases (Chapter 92) lead to loss of volume.

Anatomic Distribution

The anatomic distribution of disease can significantly facilitate the approach to diagnosis (Table 84-4 and Fig. 84-10). Upper zone lung disease

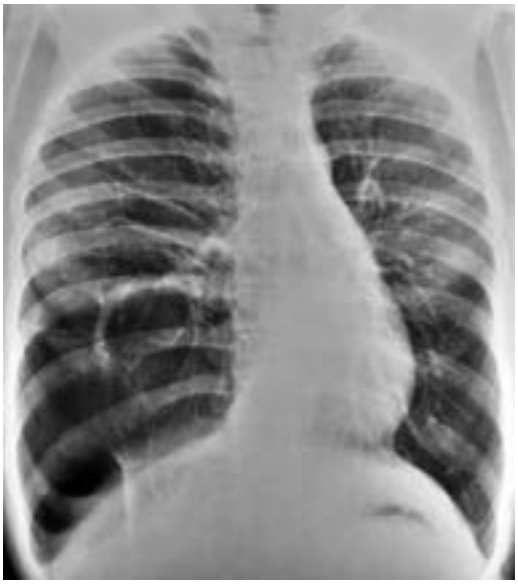

FIGURE 84-9. Patient with severe emphysema. Chest radiograph shows hyperexpansion of both lungs with bullous changes at the right lung base and leftward mediastinal shift.

TABLE 84-3 CONDITIONS ASSOCIATED WITH VARIOUS LUNG VOLUMES IN PATIENTS WITH AN UNDERLYING DIFFUSE LUNG DISEASE PATTERN

LARGE LUNG VOLUMES
Emphysema
Chronic asthma
Diffuse bronchiolitis obliterans
Highly trained athletes
Lymphangioleiomyomatosis
SMALL LUNG VOLUMES
End-stage lung fibrosis
Bilateral diaphragmatic paralysis
Massive ascites
NORMAL LUNG VOLUMES
Sarcoidosis
Langerhans cell histiocytosis
Neurofibromatosis
Combined pulmonary fibrosis and emphysema

TABLE 84-4 CONDITIONS ASSOCIATED WITH DISEASE DISTRIBUTION PATTERNS

UPPER ZONE LUNG DISEASE
Bullous lung disease
Centrilobular and paraseptal emphysema
Tuberculosis
Fungal disease
Sarcoidosis
Pneumoconioses
Langerhans cell histiocytosis
Cystic fibrosis
End-stage hypersensitivity pneumonia
Ankylosing spondylitis
Radiation pneumonia
BASAL LUNG DISEASE
Panlobular emphysema
Bronchiectasis
Aspiration
Drug reactions
Interstitial pulmonary fibrosis, nonspecific interstitial pneumonia, desquamative interstitial pneumonia, cryptogenic organizing pneumonia also called bronchiolitis obliterans with organizing pneumonia
Asbestosis
Scleroderma

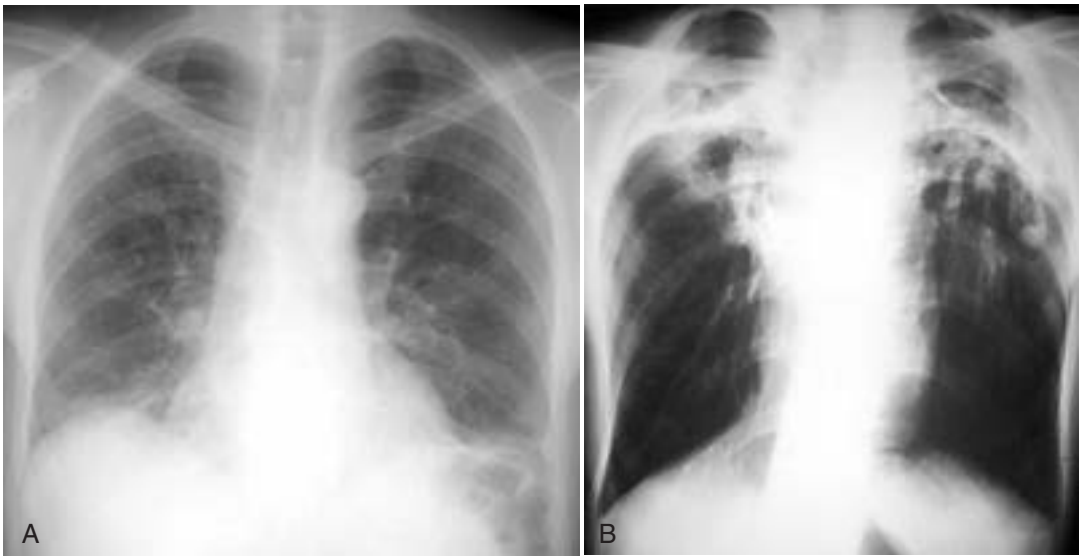

FIGURE 84-10. A, Basal pulmonary disease. Chest radiograph in a 48-year-old patient with known scleroderma. Bibasal fine reticular opacities and parenchymal bands are visible in both lower lobes. **B, Apical lung disease.** Chest radiograph in a 42-year-old patient with ankylosing spondylitis. Severe architectural distortion with cicatrizing atelectasis of both upper lobes, retraction of both pulmonary arteries cephalad, and bilateral bulla formation containing fungus balls are evident.

predominates in tuberculosis, fungal disease, sarcoidosis, pneumoconiosis (except asbestosis), Langerhans cell histiocytosis, ankylosing spondylitis, cystic fibrosis, cystic *P. jirovecii* pneumonia, radiation pneumonia, and end-stage hypersensitivity pneumonia. Basal lung disease is preferentially found in bronchiectases, aspiration, desquamative interstitial pneumonia, nonspecific interstitial pneumonia, usual interstitial pneumonia, drug reactions, asbestosis, scleroderma, and rheumatoid arthritis. Peripheral lung disease can be seen in eosinophilic pneumonia, cryptogenic organizing pneumonia, usual interstitial pneumonia, bronchioloalveolar cell carcinoma, adenocarcinoma in situ or minimally invasive adenocarcinoma (Chapter 191), and in occasional patients with so-called alveolar sarcoidosis (Table 84-5). However, any diffuse lung process will eventually progress to involve both lungs irrespective of zonal boundaries.

Lymph Nodes

Enlarged lymph nodes that are visible on chest CT scans and, when larger, on chest radiographs can provide diagnostic information (Table 84-6). The following entities can be associated with diffuse lung disease and concurrent enlarged lymph nodes: sarcoidosis (Chapter 95); lymphoma; fungal disease; tuberculosis (Chapter 324); pneumoconioses (Chapter 93), particularly silicosis and beryllium-associated lung disease; lung cancer; and metastatic malignant disease other than lung cancer.

Pulmonary Nodules

Solitary pulmonary nodules are covered in Chapter 191. Most patients with multiple pulmonary nodules larger than 1 cm in diameter have metastatic disease from primary cancers either within or outside the lung (Fig. 84-11). These lesions have a predilection for subpleural lung regions, including the lung subtending the interlobar fissures. In patients with human immunodeficiency virus infection, Kaposi sarcoma and lymphoma can induce the formation of such nodules. Infectious processes that present with multiple nodules include multiple abscesses from recurrent aspiration (Chapter 94) or septic emboli (Chapter 76); tuberculous and nontuberculous mycobacterial granulomata (Chapters 324 and 325); fungal processes, including histoplasmosis (Chapter 332), coccidioidomycosis (Chapter 333), and cryptococcosis (Chapter 336); and infection with flukes, such as *Paragonimus westermani* (Chapter 356). Noninfectious inflammatory conditions that can present with multiple pulmonary nodules include granulomatosis with angiitis (previously called Wegener granulomatosis, Chapter 270), rheumatoid nodules (Chapter 264), sarcoidosis (Chapter 95), and amyloidosis (Chapter 188).

Pleural Disease

Abnormalities of the pleural space (Chapter 99) can be displayed effectively by conventional radiographic methods supplemented by CT scanning. The volume of a pleural effusion (see Figs. 99-3 and 99-4 in Chapter 99) can be

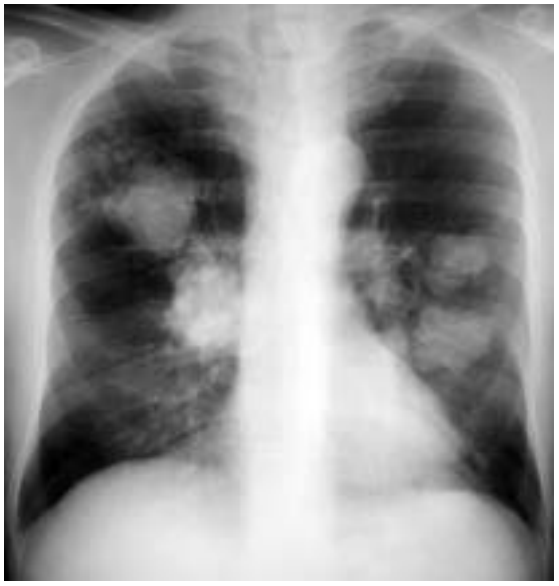

FIGURE 84-11. Multifocal pulmonary opacities. Chest radiograph in a 70-year-old patient with known carcinoma of the thyroid gland widening the superior mediastinum and displacing the cervical trachea to the right. Bilateral large and small pulmonary nodules and masses due to metastatic tumor are present.

TABLE 84-5 DISEASES AFFECTING THE LUNG PERIPHERY

Chronic eosinophilic pneumonia
Cryptogenic organizing pneumonia
Idiopathic interstitial fibrosis (usual interstitial pneumonia)
Bronchioloalveolar cell carcinoma (rare)
Pseudo-alveolar sarcoidosis (rare)

TABLE 84-6 CONDITIONS ASSOCIATED WITH HILAR AND MEDIASTINAL LYMPH NODE ENLARGEMENT

Sarcoidosis
Lymphoma
Fungal disease
Tuberculosis
Metastatic cancer
Silicosis, coal worker's pneumoconiosis, beryllium lung

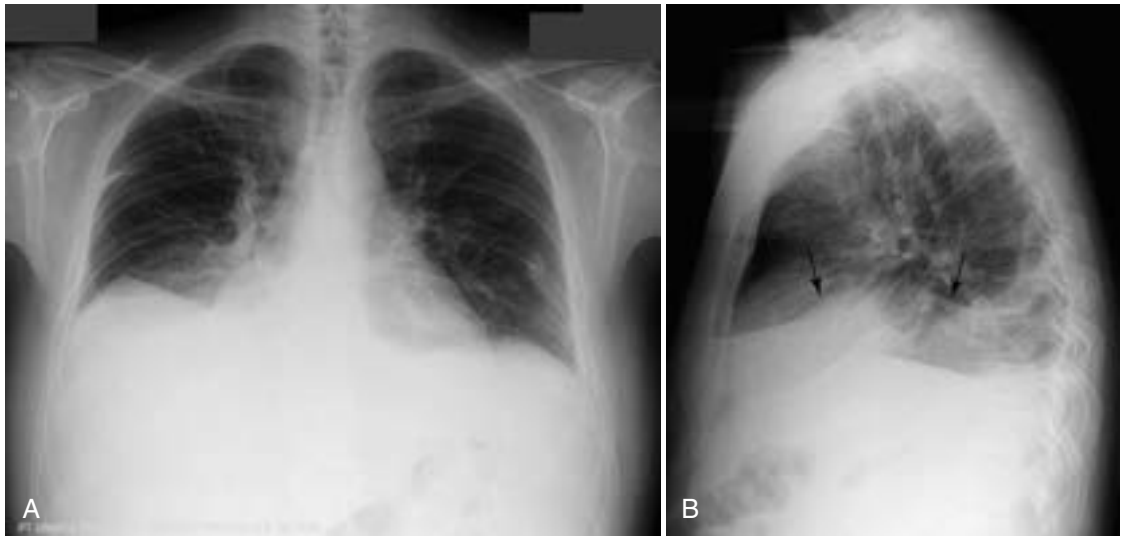

FIGURE 84-12. Bilateral subpulmonic pleural effusions. Chest frontal (**A**) and lateral (**B**) views. Both lung bases are elevated, with lateralization of the lung base curvature that mimics an elevated diaphragm. In the lateral projection, the configuration of the interface between the effusion and the aerated lung mimics the shape of the rock of Gibraltar (arrows).

reliably estimated on standard frontal upright chest radiograph: 75 mL obscures the posterior costophrenic sulcus; 150 mL obscures the lateral costophrenic sulcus; 200 mL produces a rind of 1 cm in thickness on decubitus films; 500 mL obscures the diaphragm and is visible on supine radiographs; and 1000-mL effusions reach the level of the fourth anterior rib on upright chest radiographs. An effusion of 200 mL or more can be sampled by thoracentesis. The smallest amount visible on decubitus radiographs is 10 mL. With care, as little as 175 mL of effusion can be detected on supine images. Free layering pleural effusions produce a veil of opacity or filter effect superimposed on the aerated lung; pulmonary vessels are clearly visible through the added opacity generated by the effusion, and air bronchograms are absent.

Subpulmonic and Loculated Pleural Effusions

Subpulmonic pleural effusions elevate the lung base, mimicking a high-riding hemidiaphragm. The highest curvature point of the pseudodiaphragm is shifted laterally with an abrupt lateral descent, the so-called Rock of Gibraltar sign; it also can be seen in the lateral projection (Fig. 84-12). Large pleural effusions can lead to diaphragmatic inversion. Separation of the lung base from the gas-containing stomach is indicative of a subpulmonic effusion, particularly when the stomach gas bubble is displaced inferomedially.

Loculated pleural effusions suggest the presence of pleural adhesions or indicate an underlying parenchymal lung abnormality with a focal decrease in adjacent pleural pressure. Such encapsulated collections have obtuse angles of interface with the chest wall, tapered borders that are incomplete towards the chest wall, and a sharply defined contour with the adjacent lung (Fig. 84-13).

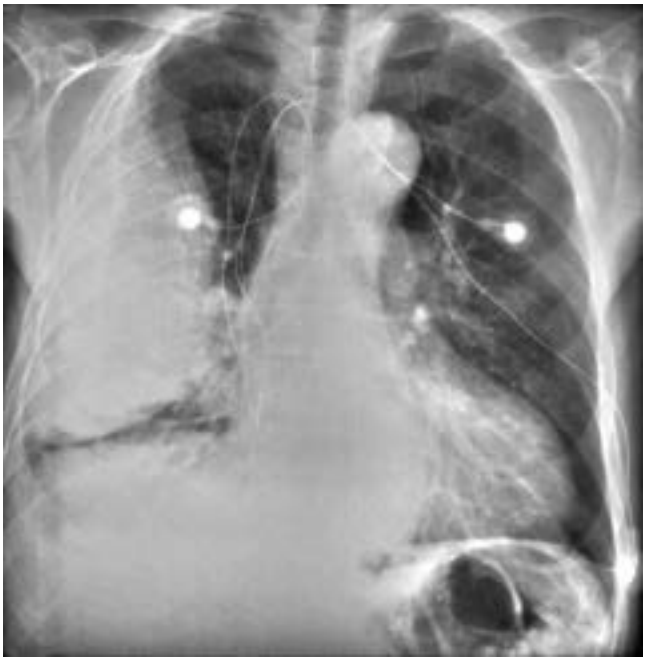

FIGURE 84-13. Loculated right-sided pleural effusion. Chest frontal view. A right-sided peripheral pleural mass is seen with characteristic obtuse angles of interface toward the chest wall, with tapered borders and sharp contour towards the lung. These findings localize the mass to the pleural or extrapleural compartment and not to the lung parenchyma. This loculated effusion proved to be an empyema.

Pleural Plaques

Pleural plaques result from parietal pleural accumulation of hyalinized collagen fibers (Fig. 84-14); their presence suggests asbestos exposure (Chapter 93). Plaques preferentially involve the parietal pleura adjacent to ribs six through nine and the diaphragm. They are less pronounced in the intercostal spaces and spare the costophrenic sulci as well as the apices. Calcifications are visible on chest radiographs in 20% and on CT scans in 50% of individuals with pleural plaques. Imaged in profile, pleural plaques produce focal areas of apparent pleural thickening. Over the diaphragm, they appear as curvilinear calcifications or scalloping. Pleural plaques viewed en face can simulate lung disease. Their appearance has been likened to holly leaves, sunburst patterns, or geographic patterns, and, when calcified, to a dripping candle, rolled margins, or stippled or irregular structures. Rare visceral pleural plaques that occur in interlobar fissures can mimic pulmonary nodules.

Diffuse Pleural Thickening

Diffuse pleural thickening is a response observed after exposure to any of a number of stimuli, including infection, inflammation, trauma, tumor, thromboembolism, radiation, and asbestos. Severe involvement results in formation of a generalized pleural peel with smooth margins, usually less than 2 cm in thickness. Radiologically diffuse pleural thickening is characterized by a smooth, noninterrupted pleural opacity involving at least one fourth of the chest wall circumference, obliterating the costophrenic sulci and encompassing also the apices. The CT criteria for diffuse pleural thickening include a thickness of at least 3 mm.

Malignant Disease

Malignant tumors of the pleura are more common than benign ones, and metastatic disease is more frequent than primary pleural mesothelioma. Primary tumors originate from pleural membranes. Pleural invasion by lung cancer, subpleural plaques in lymphoma, hematogenous dissemination to the pleura, and direct pleural seeding are other mechanisms of pleural involvement by tumor. Benign pleural tumors include lipomas, fibrous tumors, and neurogenic tumors. Lipomas are most common; their diagnosis is facilitated by CT scanning. Fibrous tumors of the pleura originate from pluripotent mesenchymal cells found in the visceral pleura or, less commonly, in the

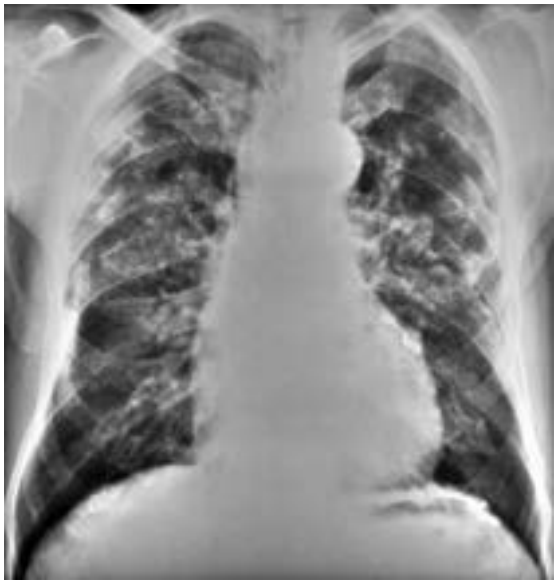

FIGURE 84-14. Patient with known prior occupational asbestos exposure. Chest radiograph shows extensive bilateral calcified plaques seen en face, in profile, and along the diaphragmatic contour.

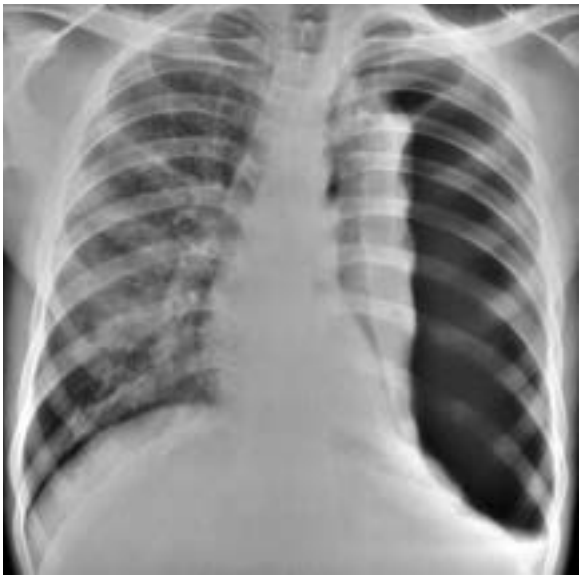

FIGURE 84-15. Patient with spontaneous tension hydropneumothorax. Chest radiograph shows complete atelectasis of the left lung with a large pneumothorax and a left basal gas-liquid level. The patient had primary tuberculosis.

parietal pleura. They can induce paraneoplastic syndromes such as hypertrophic osteoarthropathy (Chapter 179) or hypoglycemia and only rarely invade or metastasize. In nearly half of these patients, the tumor can be on a pedicle and be mobile as a patient changes position.

Pneumothorax

Pneumothorax means gas in the pleural space (Chapter 99). The most important radiologic feature of a pneumothorax is a visceral pleural line or edge that is convex or straight toward the chest wall and produces a lucent separation of the visceral and parietal pleura (Fig. 84-15). In most cases, no pulmonary vascular structures are visible beyond the visceral pleura. On upright chest radiographs, gas is primarily found in the apicolateral pleural space. Expiratory chest radiographs are not necessary for the detection of small pneumothoraces because all pneumothoraces are visible on inspiratory studies. On supine chest radiographs, pleural gas accumulates in a subpulmonic location; it outlines the costophrenic sulcus, forming the deep sulcus sign. A tension pneumothorax leads to a marked shift of the mediastinum to the contralateral side and to flattening or inversion of the ipsilateral hemidiaphragm. In supine patients with a hydropneumothorax, a veil of opacity can be seen with a gradient of decreasing attenuation toward the apex of the affected hemithorax.

TABLE 84-7 CLASSIFICATION OF MEDIASTINAL COMPARTMENTS

ANTERIOR MEDIASTINUM
Retrosternal
MIDDLE MEDIASTINUM: VISCERAL COMPARTMENT
Subcarinal space Paratracheal region Retrotracheal space Aortopulmonic window Retrocardiac space
POSTERIOR MEDIASTINUM
Paraspinal region

IMAGING OF THE MEDIASTINUM

The mediastinum encompasses midline thoracic structures that are delineated by mediastinal pleura, the diaphragm, the sternum, the spine, and the thoracic inlet. The mediastinum is commonly divided into an anterior compartment, a visceral middle compartment, and a paraspinal, posterior mediastinal compartment (Table 84-7). Each compartment contains specific pathologic entities.

Imaging Techniques

On well-penetrated chest radiographs, the anterior junction line, the posterior-superior junction line, the azygoesophageal stripe, the pleuroesophageal stripe, the paratracheal stripe, and the para-aortic and the paraspinal stripes or lines should be assessed (Fig. 84-E5). Mediastinal masses need to be detected and localized first. Their obtuse angles of interface with the mediastinal pleura, incomplete border towards the mediastinum, sharp contour towards the lung, and extension into both hemithoraces indicate the mediastinal origin of such lesions.

CT facilitates localization of a mass to a specific mediastinal compartment. When it is known whether the mass is predominantly fatty or cystic, contains soft tissue, or is calcified, the differential diagnosis can be limited. MRI of the mediastinum has a role in diagnosis of vertebral disease or neurogenic tumors with extension into the spinal canal. It is equivalent to CT in diagnosis of aortic aneurysms and dissections (Chapter 78).

Mediastinal Compartments

The anterior mediastinum is actually a potential space that may contain the fatty replaced thymus and small normal lymph nodes. Space-occupying lesions in this compartment typically include thymomas, lymphomas, teratomas and other germ cell tumors, substernal thyroid goiters, lipomas, and other connective tissue tumors, as well as hemangiomas or lymphangiomas (Fig. 84-16A).

The middle mediastinum is subdivided into the subcarinal space, paratracheal region, retrotracheal region, aortopulmonic window region, and retrocardiac space. Characteristic lesions are enlarged lymph nodes and bronchopulmonary foregut malformations (see Fig. 84-16B).

In the retrotracheal region, aberrant right subclavian arteries, posterior descending goiters, esophageal tumors, diverticula, or thoracic duct cysts can be found. In the aortopulmonic window, enlarged lymph nodes, ductus diverticula, bronchopulmonary foregut malformations, and aortic or pulmonic artery aneurysms can form compartment-specific space-occupying lesions.

The paraspinal region is considered radiologically to belong to the posterior mediastinum. Important masses in that space include neurogenic tumors that originate from the sympathetic chain or from segmental nerve roots (see Fig. 84-16C). Extramedullary hematopoiesis in patients with severe anemia can result in paravertebral masses formed by hypertrophied bone marrow that extrudes from ribs or vertebral bodies. Enlarged lymph nodes due to lymphoma or metastatic disease are occasionally seen in a paraspinal location. Vertebral disease, including bacterial or tuberculous spondylitis, tumors, and post-traumatic hematomas, can widen the paraspinal region and produce contour abnormalities.

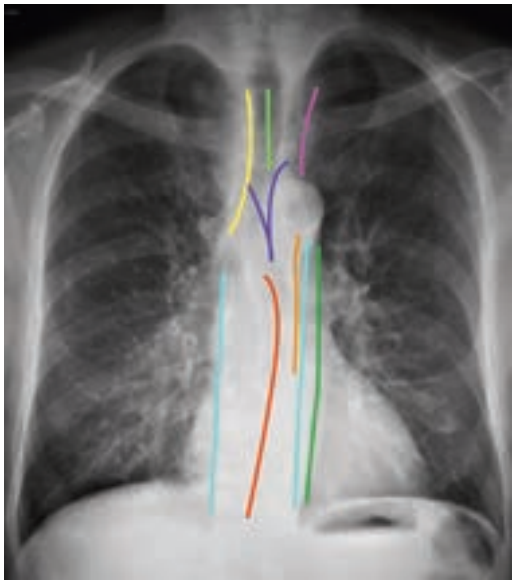

E-FIGURE 84-5. Chest radiograph with superimposed mediastinal stripes. *Yellow:* Right paratracheal stripe. *Light blue:* Right and left paraspinal stripes. *Red:* Azygoesophageal stripe. *Brown:* Pleuroesophageal stripe. *Purple:* Anterior junction line complex. *Pink:* Left subclavian artery border. *Light green:* Posterior-superior junction line. *Dark green:* Para-aortic line.

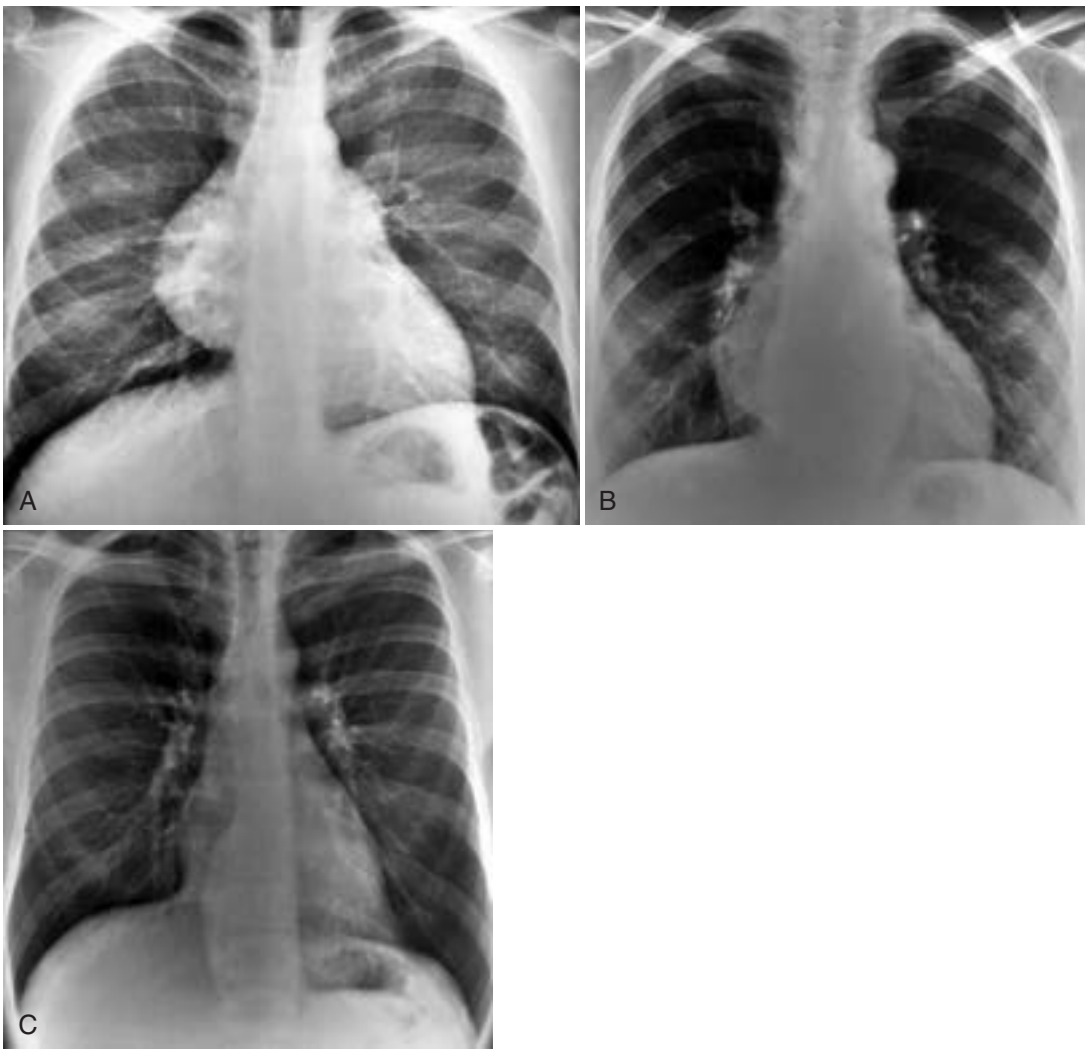

FIGURE 84-16. A, Patient with anterior mediastinal teratoma. Chest radiograph shows a mediastinal contour abnormality due to projection of the mass into the right hemithorax. Note the obtuse angle of interface formed by the pleura covering the mass with the mediastinum. **B,** Patient with Castleman giant lymph node hyperplasia. Chest frontal radiograph shows large subcarinal middle mediastinal mass that projects lateral to the right atrium. **C,** Patient with paraspinal ganglioneuroma. Chest radiograph shows right lower paraspinal contour abnormality widening the right paraspinal region and encompassing the height of three thoracic vertebrae.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Oba Y, Zaza T. Abandoning daily routine chest radiography in the intensive care unit: meta-analysis. *Radiology*. 2010;255:386-395.
2. Kim HS, Lee KS, Ohno Y, et al. PET/CT versus MRI for diagnosis, staging, and follow-up of lung cancer. *J Magn Reson Imaging*. 2014; [Epub ahead of print].
3. Hansell DM. Classification of diffuse lung diseases: why and how. *Radiology*. 2013;268:628-640.
4. Jin GY, Lynch D, Chawla A, et al. Interstitial lung abnormalities in a CT lung cancer screening population: prevalence and progression rate. *Radiology*. 2013;268:563-571.

REVIEW QUESTIONS

1. Concerning chest radiography, which of the following statements is incorrect?

A. Chest radiography is the most commonly performed imaging procedure.
B. Chest radiography provides useful information at modest monetary cost with very low radiation exposure.
C. Chest radiographs were previously obtained by digital imaging but are preferentially performed with cassettes and radiographic film.
D. The standard chest radiograph is performed in frontal and lateral projections.
E. The chest radiographs should be obtained at total lung capacity.

Answer: C Chest radiographs were previously obtained with film and cassette and now are acquired digitally.

2. Concerning the radiologic distinction between alveolar and interstitial patterns, which of the following statements is incorrect?

A. Correlation between radiographic localization to an anatomic compartment and the actual histopathologic finding is poor.
B. Nodular patterns can be seen in either interstitial or alveolar disease.
C. Alveolar diseases can induce an interstitial reaction.
D. Air-bronchograms are seen only in airspace disease.
E. Ground-glass opacities can be induced by either interstitial or alveolar disease.

Answer: D Air bronchograms are visualized whenever the lung parenchyma, including the alveolar and interstitial compartments that surround gas-containing bronchi, becomes opacified. The resulting contrast between soft tissues and gas yields the air-bronchograms.

3. The size of pleural effusions can be estimated as noted below, except:

A. The minimum visible on a lateral decubitus examination is 200 mL.
B. On a lateral chest radiograph, 75 mL can be identified.
C. On the frontal chest radiograph, 120 mL can be identified.
D. Five hundred milliliters obscures the ipsilateral hemidiaphragm.
E. Up to 1000 mL are present if the meniscoid arc reaches the level of the fourth anterior rib.

Answer: A The smallest amount of pleural effusion visible on decubitus radiographs is 10 mL.

85

RESPIRATORY FUNCTION: MECHANISMS AND TESTING

PAUL D. SCANLON

Pulmonary function testing has been used in the medical evaluation of patients with respiratory issues since Hutchinson 1846 demonstration that vital capacity, the largest volume of air that can be exhaled, is an important measure of health. Data from epidemiologic studies show that lung function is one of the most important predictors of all-cause mortality.[1]

Measures of lung function include assessments of respiratory mechanics, for example, the volume of gas that is contained by the lung in various circumstances, the inspiratory and expiratory flow rates across the vital capacity, the pressures that can be generated by inspiratory and expiratory efforts, and the resistance to airflow, as well as calculations of gas exchange. Some lung function tests can be self-performed by a patient, but most are performed in pulmonary function laboratories. The analysis of bronchoalveolar lavage fluid, obtained by bronchoscopy, can also provide important insights into pulmonary disease.

SPIROMETRY

The simplest, most commonly performed, and most clinically useful pulmonary function test is spirometry (Table 85-1).[2] This test measures the forced vital capacity (FVC), which is the amount of air that can be forcefully expelled from the lungs as a function of time, beginning at maximal inhalation (termed total lung capacity) and ending when the lungs are emptied to their minimal volume (residual volume). Many secondary measures are derived from the FVC maneuver, including volume exhaled in a given time, termed the forced expiratory volume (FEV), with a subscript indicating the number of seconds during which this measurement is made (e.g., FEV_1, FEV_3). The ratio of the FEV_1/FVC is the proportion of the total vital capacity that can be expelled in the first second of a maximal expiratory effort; a low FEV_1/FVC ratio is a commonly used indicator of obstructive lung disease. In addition, expiratory flow can be measured at specific portions of exhaled vital capacity, termed forced expiratory flow (FEF), followed by a number to represent the percentage of the FVC at which the flow was measured (e.g., FEF_{75}, FEF_{50}, FEF_{25}). Expiratory flow can also be measured over a given volume range (e.g., FEF_{25-75}). Spirometry requires relatively simple equipment, modest technician training, and modest patient effort. Results are highly reproducible within a test session and over time between test sessions, thereby allowing meaningful comparisons over time for clinical evaluation and as an important outcome in research studies.

The data from a maximal exhalation can be displayed in a flow-volume curve (Fig. 85-1A), which depicts exhaled flow at any given volume as a function of exhaled volume. The flow-volume curve, which has a unique shape for

TABLE 85-1 PULMONARY FUNCTION TESTS

LUNG VOLUME	
TLC	Total lung capacity (volume of gas in lungs at the end of maximal inspiration)
FRC	Functional residual capacity (volume of gas in the lungs at relaxation point, when elastic inward pull of lungs is balanced by outward pull of the chest wall and diaphragm)
ERV	Expiratory reserve volume (volume of gas expired from FRC to maximal expiration)
RV	Residual volume (FRC − ERV, volume of gas left in lungs after maximal exhalation)
EXPIRATORY FLOW	
FEV_1	Forced expiratory volume (in 1 second)
FVC	Forced vital capacity
$FEV_{1\%}$	FEV_1/FVC ratio (expressed as percentage)
DIFFUSING CAPACITY	
D_{LCO}	Diffusing capacity of lungs for carbon monoxide
ARTERIAL BLOOD GASES	
Pa_{O_2}	Partial pressure of oxygen in arterial blood
Pa_{CO_2}	Partial pressure of carbon dioxide in arterial blood
pH	Negative log of hydrogen ion concentration

any individual at any given time, is often displayed as a graphic image. It can reflect airway obstruction (Fig. 85-1B, C), such as is seen in asthma (Chapter 87) and chronic obstructive lung disease (COPD; Chapter 88), or lung restriction (Fig. 85-1D, E), such as is seen in many interstitial lung diseases (Chapter 92). It also may demonstrate less common abnormalities, including variable extrathoracic (Fig. 85-1F) or intrathoracic (Fig. 85-1G) obstruction, tracheal stenosis (Fig. 85-1H), or severe muscle weakness (Chapter 86; Fig. 85-1I). A flow-volume curve does not display time explicitly, so the FEV_1 cannot be measured directly unless the curves are marked with the point of 1 second of exhalation.

The FVC, FEV_1, FEV_1/FVC ratio, and shape of the flow-volume curves are highly reproducible measures of lung function if the tested individual makes an expiratory effort above a certain, easy-to-obtain level in a pulmonary function laboratory with modern spirometry equipment and trained staff. In the absence of quality assurance, however, measurements of lung function are neither accurate nor reproducible; poor-quality pulmonary function testing results are biased to lower values, thereby giving an incorrect impression of disease where none may exist. One way to assess a patient's performance is to compare multiple efforts and to document that the two best measures of both FVC and FEV_1 are within 150 mL of each other, a standard that most people can meet without difficulty. Poor performance should be noted by the interpreter to avoid erroneous diagnoses.

Spirometry is recommended for diagnosis of airflow obstruction in symptomatic patients. It is not recommended as a screening test for asymptomatic persons thought to be at risk for development of lung disease, such as current or former smokers,[3] but it may be part of comprehensive workplace respiratory health programs in certain occupational settings.[4] An abnormal spirometric result has not been shown to improve the likelihood that such at-risk individuals will quit smoking, and a normal screening spirometry test result might be misinterpreted as an indicator that smokers can continue smoking without risk.

Spirometry is often performed before and after administration of an inhaled bronchodilator, either a β-agonist (e.g., albuterol) or a muscarinic antagonist (e.g., ipratropium) or both, especially if it shows changes consistent with airway obstruction. Dosing may use two or four puffs from a metered dose inhaler or nebulized aerosols. The degree of improvement after bronchodilator administration indicates the degree of airway reactivity, which is generally more in asthma (Chapter 87) and less in COPD (Chapter 88). Response to bronchodilators varies with dosage, is poorly repeatable from test to test, and is not a good predictor of the clinical response to bronchodilator therapy in an individual patient.

OTHER TESTS OF VENTILATION

Maximal voluntary ventilation (MVV), which is an indication of the maximal ventilation a patient can perform, is measured during 12 seconds, using the best 6 seconds and expressed in liters per minute. MVV estimates a person's upper limit of ventilatory capacity. Reductions in MVV may be due to inspiratory obstruction, muscle weakness, or poor performance. Because MVV is effort dependent, it may be a better predictor of postoperative respiratory complications (Chapter 433) than is FEV_1. Inspiratory flows, which are not part of routine spirometry, may be useful for patients in whom there is a suspicion of upper airway disease, such as a patient referred by an otorhinolaryngologist, a patient who has stridor, or a patient whose MVV is reduced out of proportion to the FEV_1.

REFERENCE EQUATIONS

Because lung size varies substantially from person to person, the values obtained from pulmonary function testing are compared with those of normal individuals. Because test results vary as a function of sex, height, age, and ethnicity, they are usually expressed as a percentage of a reference value calculated with those four factors. Normal values in African American individuals are about 12% lower than values from white persons of the same sex, age, and height. Normal values are also about 6 to 15% smaller in Asians. Genetic markers of ethnic ancestry are predictive of lung function and, in the future, might be used to improve on traditional methods of adjusting for racial or ethnic factors.

LUNG VOLUMES

The volume of air in the lung at any given time can be partitioned (Fig. 85-2). The air that remains in the lung after a maximal expiratory effort is the residual volume. The amount of air in the lungs at the relaxation point, when muscle effort is minimized and the inward recoil of the lung is balanced by the outward recoil of the chest wall, is the functional residual capacity (FRC). The difference between FRC and residual volume is the expiratory reserve volume. The volume exhaled in a normal breath is the tidal volume. The volume that can be inhaled above tidal volume is the inspiratory reserve volume.

A series of *capacities* consist of the sum of two or more different volumes. FRC is the sum of expiratory reserve volume plus residual volume. Inspiratory capacity is the sum of tidal volume plus inspiratory reserve volume. Vital capacity is the sum of tidal volume plus inspiratory reserve volume plus expiratory reserve volume. Total lung capacity is the sum of residual volume plus expiratory reserve volume plus tidal volume plus inspiratory reserve volume.

Three of the volumes (tidal volume, inspiratory reserve volume, expiratory reserve volume) are simply volumes of exhaled gas and can be measured with a spirometer. Measurement of residual volume or any of the capacities that include it, so-called absolute lung volumes, requires more sophisticated methods, such as body plethysmography, the inert gas dilution technique, or the nitrogen washout technique.

Body Plethysmography

Body plethysmography, which is the preferred method for measuring lung volumes, is based on Boyle's law: at a given temperature, the product of the pressure and volume of a quantity of gas at one time will be equal to the product of the pressure and volume of the gas at another time ($P1 \times V1 = P2 \times V2$). The process of measuring lung volume by plethysmography consists of panting against a closed shutter to compress and to rarify gas in the chest. The body plethysmograph, a sealed box in which the patient sits, measures the changes in lung volume during panting; pressure measured at the mouth represents the pressure changes within the lung during these volume changes. A similar panting maneuver with the shutter open is used to calculate airway resistance. The clinical utility of this measurement is largely limited to instances in which airway obstruction is suspected but the FEV_1/FVC ratio is normal. Although body plethysmography is generally the most accurate method for measurement of lung volumes, particularly in patients with airway obstruction, it can overestimate lung volumes if panting is too rapid. A plethysmographic total lung capacity more than 150% of the reference value should be viewed with suspicion.

Inert Gas Dilution Technique

Lung volumes also can be measured by having the patient rebreathe from a device containing a known volume and concentration of inert gas (e.g., helium or less commonly neon or argon, which do not react with elements in the blood or tissues) until equilibrium is achieved. The final concentration of helium equals the initial helium concentration times the initial volume of the device divided by the final volume of the lungs plus the device, correcting

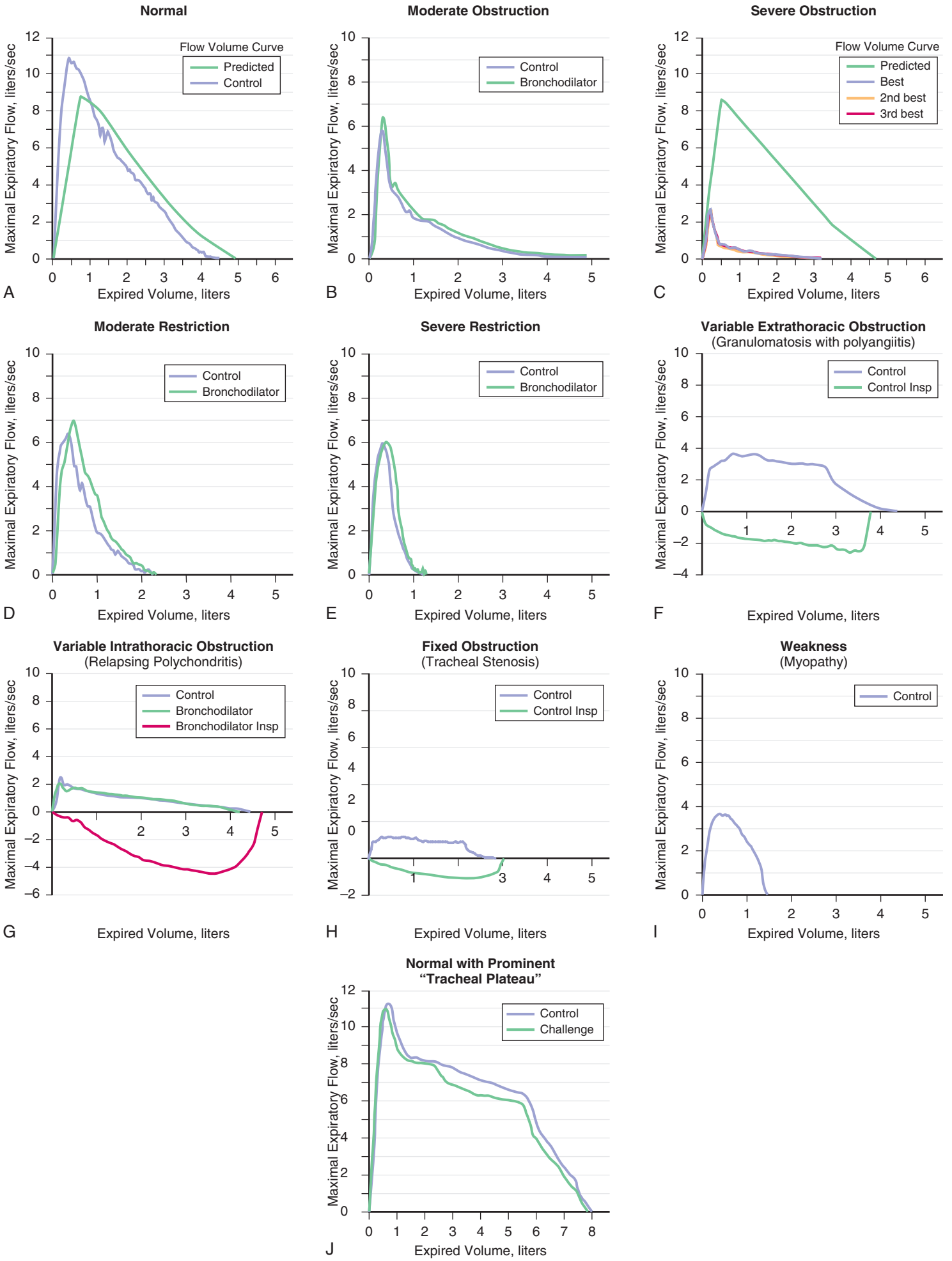

FIGURE 85-1. **Common patterns of flow-volume curve. A,** Normal. **B,** Moderate obstruction. **C,** Severe obstruction. **D,** Moderate restriction. **E,** Severe restriction. **F,** Variable extrathoracic obstruction (granulomatosis with polyangiitis). **G,** Variable intrathoracic obstruction (relapsing polychondritis). **H,** Fixed obstruction (tracheal stenosis). **I,** Weak effort (myopathy). **J,** Normal but with a prominent tracheal plateau. FET = forced expiratory time.

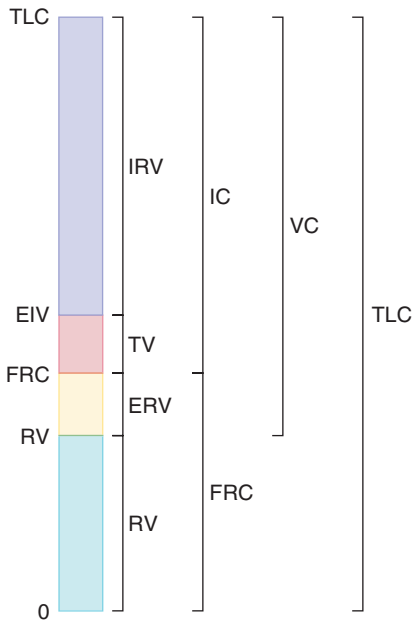

FIGURE 85-2. A schematic diagram showing lung volume partitions as measured in lung function tests. EIV = end-inspiratory volume; ERV = expiratory reserve volume; FRC = functional residual capacity; IC = inspiratory capacity; IRV = inspiratory reserve volume; RV = residual volume; TLC = total lung capacity; TV = tidal volume; VC = vital capacity.

for oxygen consumption and carbon dioxide production during the test. The equation can be solved for lung volume. This method underestimates lung volumes when portions of the lung communicate poorly with the central airways, as in patients with emphysematous bullae.

Nitrogen Washout Technique

The air that we breathe consists of approximately 21% oxygen, 1% argon, 0.04% carbon dioxide, and a variable amount of water vapor. The remainder is nitrogen. Exhaled air contains a lower concentration of oxygen, usually 14 to 16%, plus 3 to 5% carbon dioxide and water. For the nitrogen washout technique, the test subject inhales 100% oxygen beginning at the FRC. All exhaled gas is collected until the concentration of nitrogen reaches a plateau. Knowing that the initial concentration of nitrogen is approximately 78% and measuring the final concentration and volume of gases collected, the initial volume of gas in the lungs at FRC can be calculated. This method also underestimates lung volumes in patients with poorly communicating air spaces.

Lung volumes can also be measured from chest radiographs and computed tomography scans. The correlation among the measurement techniques is very good for people with reasonably normal lungs. In the presence of lung disease, however, each of the methods has limitations.

Absolute lung volumes as determined by body plethysmography or one of the gas dilution methods can be used to refine the spirometric evaluation of both obstructive and restrictive disorders. In obstructive disorders, air trapping or hyperinflation can be inferred from an increased residual volume, total lung capacity, or residual volume/total lung capacity ratio. If the total lung capacity is greater than 125% or 130% of predicted, hyperinflation is present. A residual volume greater than the upper limit of normal suggests air trapping. In subjects with chest wall limitation or neuromuscular weakness, residual volume may be increased—not because of true airway trapping but because of limitation to chest wall movement, so the term *air trapping* should be used with caution.

Reduced lung volumes are the sine qua non of the diagnosis of restriction. However, about half of patients whose spirometry suggests restriction (reduced vital capacity with normal FEV_1/FVC ratio) have a normal total lung capacity, so they do not have true restriction but rather what is called the nonspecific pattern.[5] Patients with the nonspecific pattern commonly have evidence of an obstructive disorder, not restriction, as evidenced either by increased airway resistance or by response to a bronchodilator. Some patients, however, do not have evidence of airway obstruction but have chest wall limitations, neuromuscular weakness, poor performance, heart failure, or any of a variety of other conditions. The nonspecific pattern occurs in 9 to 10% of all complete pulmonary function tests and is approximately as frequent as true restriction.

DIFFUSING CAPACITY

The single-breath diffusing capacity for carbon monoxide (DLCO) is the most common clinically used measure of the gas exchange capacity of the lungs. The maneuver for measurement of DLCO requires breathing out to the residual volume and then quickly inhaling a mixture of gas with a known concentration of an inert gas (e.g., helium or neon) plus a small concentration of carbon monoxide. After inhaling to total lung capacity, the patient holds his or her breath for 10 seconds, during which time the helium or other tracer gas mixes with other gases occupying the total lung capacity while the carbon monoxide is absorbed from the alveolar spaces because of the strong affinity of hemoglobin for carbon monoxide. After a 10-second breath-hold, the remaining gas mixture is exhaled. The concentration of inert tracer is used to calculate the volume of the lungs (alveolar volume); the concentration of carbon monoxide is used to calculate the absorption of carbon monoxide in volume per minute per unit of pressure (mL/min • mm Hg).

A normal value for DLCO is generally interpreted as indicative of normal gas exchange, which requires a normal pulmonary gas-exchanging surface, normal capillary blood volume, and relatively homogeneous regional ventilation-perfusion relationships. A low DLCO is indicative of impaired gas exchange. In obstructive disorders (Chapters 87 and 88), impaired gas exchange occurs most commonly in patients with emphysema as opposed to asthma. In restrictive disorders, it is seen most commonly in the presence of interstitial disorders. Patients with pulmonary vascular disorders may have restriction or normal lung mechanics (Chapter 92). An isolated reduction in DLCO (i.e., in association with normal total lung capacity, vital capacity, and FEV_1) can indicate a pulmonary vascular disorder but is more commonly seen in association with pulmonary fibrosis (Chapter 92), emphysema (Chapter 88), or a combination of the two.

An increased DLCO is relatively uncommon but can be seen most often in individuals with asthma (Chapter 87) or obesity (Chapter 220). It can also be seen in association with polycythemia (Chapter 166), with a left-to-right intracardiac shunt (Chapter 69), with acute pulmonary hemorrhage (Chapter 91), or during exercise.

MAXIMAL RESPIRATORY PRESSURES

Maximal respiratory pressures help identify muscle weakness, which can cause a restrictive disorder, a nonspecific pattern, or an isolated reduction in MVV relative to FEV_1. Maximal respiratory pressures do not distinguish muscle weakness from poor test performance. They are not a routine test in most laboratories but can be added to evaluate specific abnormalities (Fig. 85-3).

INTERPRETATION OF LUNG FUNCTION TESTS

The interpretation of pulmonary function tests uses the information from the measures noted to make a physiologic diagnosis, that is, to categorize the nature and magnitude of the mechanical impairments to lung function (Table 85-2). The four broad categories of physiologic abnormalities that can be gleaned from these tests are obstructive disease, such as asthma and COPD (Chapters 87 and 88); restrictive disease of the lung, such as pulmonary fibrosis (Chapter 92), or restriction due to factors outside the lung, such as chest wall limitation due to obesity, pleural disease, or musculoskeletal disorders; weak chest wall (Chapter 99), such as Guillain-Barré syndrome (Chapter 420); and impaired gas exchange in the presence of normal mechanical function, such as pulmonary embolism (Chapter 98). Some patients have mixed physiologic defects, such as a combined restrictive and obstructive defect (Fig. 85-3), or more than one cause of restriction (e.g., pulmonary fibrosis plus obesity or heart failure).

Spirometry screening of the U.S. adult population shows evidence of airflow obstruction in about 13.5% of individuals and evidence of restriction in about 6.5%.[6] Of individuals with spirometric evidence of restriction, about 50% have true restriction when lung volumes are measured, whereas the other 50% have a nonspecific pattern of pulmonary function abnormality.

The first step in interpretation of a set of pulmonary function measurements is to inspect the numerical data, the spirogram, and the flow-volume curve to assess the quality of the test. A poor-quality test result, whether it is due to poor performance by the patient or poor coaching by the technician, may have an irregular flow-volume curve or poor reproducibility of results from one effort to another. Once good quality is affirmed, the presence of an abnormal pattern (e.g., obstruction or restriction) can be determined. If so, attention then turns to assessing gradations of severity, subtleties of the flow-volume curve, and other physiologic data (e.g., total lung capacity, residual

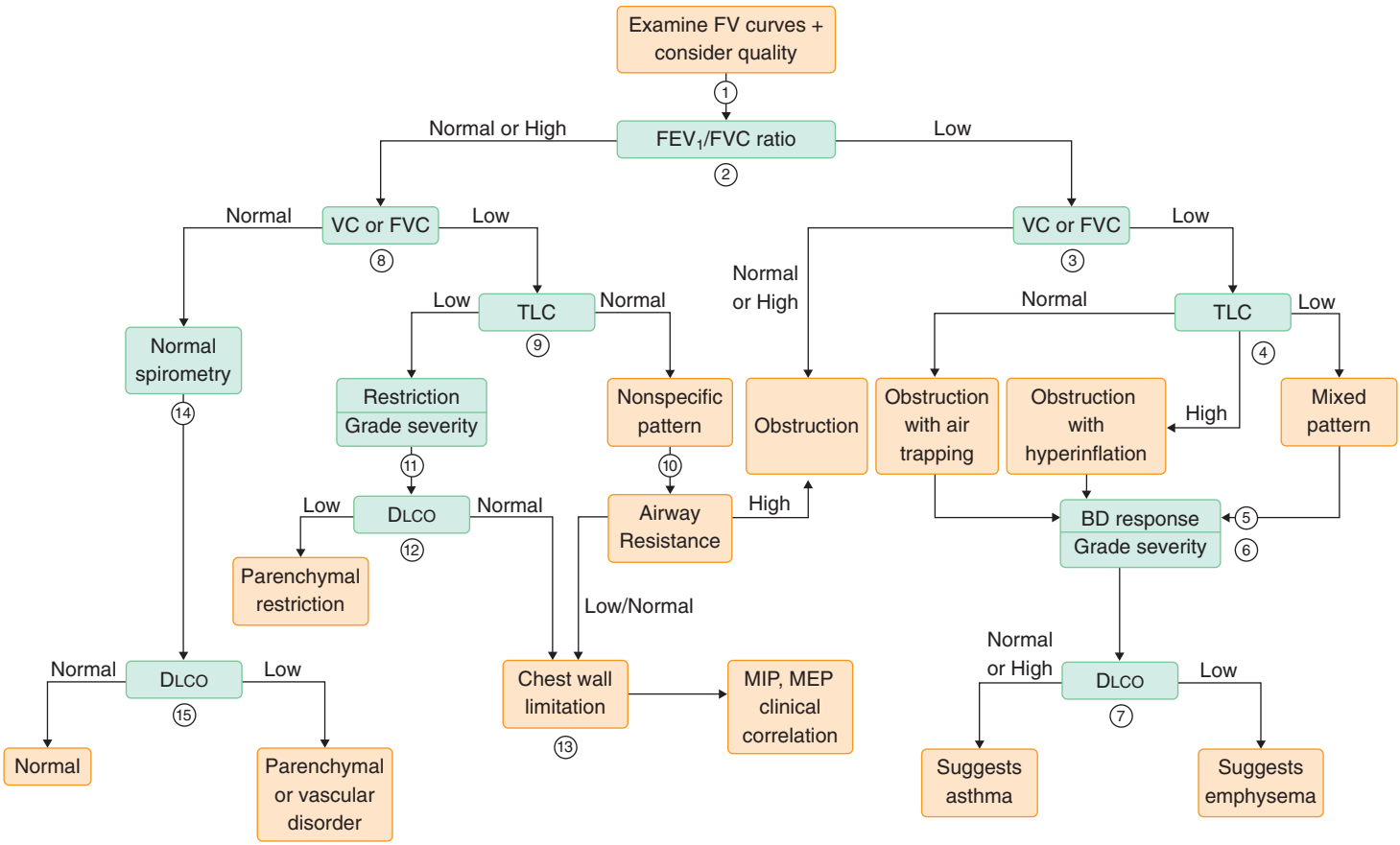

FIGURE 85-3. An algorithm for interpreting pulmonary function tests in which spirometry, lung volumes, and DLCO are measured. If only spirometry is available, interpretation is more limited. The legend keys refer to numbered branch points in the algorithm. BD = bronchodilator; DLCO = diffusing capacity of the lung for carbon monoxide; FEV_1 = forced expiratory volume in one second; FVC = forced vital capacity; MEP = maximal expiratory pressure; MIP = maximal inspiratory pressure; NI = normal; TLC = total lung capacity; VC = vital capacity.

1. The algorithm begins at the top. Inspect the data and flow-volume (FV) curve to assess test quality and then consider the basic type of abnormality (e.g., obstruction vs. restriction).
2. A reduced FEV_1/FVC ratio suggests obstruction, and the obstruction algorithm on the right side should be followed. If the FEV_1/FVC ratio is normal or high, the restriction side (left side) of the algorithm should be followed.
3. If FEV_1/FVC is low and the VC or FVC is normal or high, simple obstruction is present. If VC or FVC is low, TLC should be checked.
4. If TLC is normal, simple obstruction is present. If TLC is high, obstruction with hyperinflation is present. If TLC is low, a mixed obstructive/restrictive pattern is present. (Note that the inert gas dilution and nitrogen washout methods commonly underestimate TLC in the presence of obstruction and can give a false impression of a mixed disorder.)
5. The response to a bronchodilator (BD) should be assessed to determine whether FEV_1 or FVC meets criteria for a positive response (i.e., a ≥12% improvement and at least a 200 mL absolute increase) and to determine the degree of positivity.
6. The severity of obstruction should be graded. Some algorithms grade severity based on post-bronchodilator values.
7. In current or former smokers with obstruction, a low DLCO suggests emphysema or other pulmonary parenchymal or vascular disorders. A normal DLCO may suggest asthma or bronchitis.
8. If FEV_1/FVC is normal or high, the restriction (left) side of the algorithm is followed. If VC or FVC is normal, spirometry is generally normal (occasional patients have an isolated abnormality of FEV_1 of uncertain significance). If VC or FVC is low, TLC should be checked.
9. If TLC is low, a restrictive disorder is present. If TLC is normal, the "nonspecific pattern" is present.
10. If the nonspecific pattern is identified, airway resistance (Raw) can be measured. An increased Raw suggests obstruction. A normal Raw suggests an alternative cause (See #13).
11. If true restriction is present, grade severity on the basis of the reduction in TLC percentage predicted.
12. If restriction is demonstrated, DLCO should be measured next. If abnormal, it indicates a pulmonary parenchymal restrictive process. If normal, it suggests an extraparenchymal or nonpulmonary cause of restriction.
13. Restriction with a normal DLCO or a nonspecific pattern with normal Raw suggests an alternative cause (chest wall limitation, weakness, heart failure, poor performance). Consider measurement of maximal respiratory pressures and review the study for test performance.
14. If spirometry is normal, lung volumes are rarely useful, but DLCO is sometimes helpful.
15. If DLCO is normal, pulmonary function is normal. An isolated reduction in DLCO is seen most often in patients with emphysema or pulmonary fibrosis or both. It less commonly indicates a pulmonary vascular disorder, such as primary pulmonary hypertension, or an obliterative vasculopathy, as sometimes seen in Sjögren syndrome.

TABLE 85-2 COMMON CHANGES ASSOCIATED WITH PATTERNS OF LUNG FUNCTION ABNORMALITY

	FORCED EXPIRATORY VOLUME IN 1 SECOND (FEV_1)	FORCED VITAL CAPACITY (FVC)	FEV_1/FVC RATIO	RESIDUAL VOLUME	TOTAL LUNG CAPACITY	MAXIMAL INSPIRATORY AND EXPIRATORY PRESSURE RATIOS
Normal	Normal*	Normal	Normal	Normal	Normal	Normal
Obstructive	↓	Normal to ↓	↓	↑ to ↑↑	Normal to ↑↑	Normal
Restrictive	↓	↓ to ↓↓	Normal or ↑	Normal or ↓	↓ to ↓↓	Normal
Weak chest wall	↓	↓ to ↓↓	Normal or ↑	↑	Normal or ↓	↓

*Normal or abnormal values are determined by comparing the measured values with those predicted from regression equations based on the patient's sex, age, height, and race. The normal range for FEV_1/FVC also varies, mainly with age, ranging from 0.70 to 0.80 among 25-year-olds to 0.63 to 0.68 among 65-year-olds.

volume, MVV, and DLCO) that either support or supplement the initial impression.

The normal flow-volume curve is roughly triangular (Fig. 85-1A). A *tracheal plateau*, which is a normal variant (Fig. 85-1J) usually seen in younger subjects, is caused by flow limitation in the trachea in the absence of peripheral airway obstruction.

Patients with obstructive disorders typically have a reduced FEV_1/FVC ratio and a flow-volume curve with a "scooped out" appearance (Fig. 85-1B, C). Atypical patients may have unusually shaped flow-volume curves or unusual patterns of obstruction (e.g., abnormal airway resistance despite a normally shaped flow-volume curve and normal FEV_1/FVC ratio). If a patient has what appears to be obstruction, the next step is to determine whether the obstruction can be reversed by the administration of a bronchodilator. If it is clinically indicated, hyperinflation or air trapping can be diagnosed on the basis of an increased total lung capacity or residual volume, and the adequacy of gas exchange can be assessed by measuring DLCO and oximetry. In individuals with normal spirometry and alveolar volume from the DLCO measurement, measurement of total lung capacity can be avoided to save unnecessary expense.

In obstructive diseases, the degree of impairment of pulmonary functions can be classified on the basis of the FEV_1. An FEV_1 less than the lower limit of normal but greater than 70% is mild, 60 to 69% is moderate, 50 to 59% is moderately severe, 35 to 49% is severe, and less than 35% is very severe.

Patients with restrictive disorders have reduced lung volumes and typically have a flow-volume curve with a "witch's hat" shape—a tall peaked curve (Fig. 85-1D, E). Restriction may be due to either reduced lung compliance or mechanical changes to the chest wall and tissues surrounding the lungs (e.g., muscle weakness, chest wall deformity, obesity, pregnancy, pleural effusion, or heart failure). For restrictive diseases, the severity of the restriction can be graded with use of the total lung capacity as a percentage of the predicted value. Changes on serial testing help predict prognosis.[7]

In patients with restriction caused by interstitial disease, the total lung capacity and the vital capacity or FVC are usually reduced by a similar proportion. In some patients with restriction, the total lung capacity as a percentage of predicted and the vital capacity percentage of predicted are quite different (>10% difference). The usual cause is the presence of more than one restrictive process, such as a parenchymal restrictive disorder plus obesity, respiratory muscle weakness, or heart failure.

Some patients have a mixed disorder with evidence of both obstruction and restriction. Common causes include cystic fibrosis (Chapter 89), sarcoidosis (Chapter 95), and heart failure (Chapters 58 and 59) as well as cases in which the cause of the obstructive disorder and the restrictive disorder are unrelated.

Disorders of the central airways can cause characteristic patterns of abnormality. In a "fixed airway obstruction" such as tracheal stenosis (Fig. 85-1H), flow is typically reduced on both inspiration and expiration. In contrast, in a variable extrathoracic upper airway obstruction (Fig. 85-1F), inspiration is disproportionately reduced; however, expiration is often abnormal, merely less so. Likewise, in variable intrathoracic obstruction (e.g., relapsing polychondritis, tracheomalacia, or a dynamic intrathoracic tracheal tumor), the expiratory flow-volume curve is reduced but in a pattern unlike that seen in asthma or COPD (Fig. 85-1G). These central airway obstructive patterns may signify a locally treatable cause of obstruction.

PROVOCATIVE TESTING

Assessing Airway Responsiveness

Hyperresponsiveness of airways to the smooth muscle–contracting effect of pharmacologic agents such as methacholine, as well as to cold air, dry air, and other physical stimuli, is characteristic of asthma (Chapter 87). It is also observed in COPD and other obstructive airway diseases. Bronchoprovocation studies, in which graded doses of a stimulus are used to elicit airway constriction, are performed to measure airway responsiveness. A responsive airway, that is, one in which a small stimulus leads to a fall in FEV_1, may be used to confirm the diagnosis of asthma (Chapter 87).

Exhaled nitric oxide is a marker of eosinophilic airway inflammation and can be used to predict the likelihood that airway obstruction will improve with corticosteroid treatment. However, the utility of exhaled nitric oxide levels for asthma management is controversial.

CARDIOPULMONARY EXERCISE TESTS

Some patients have dyspnea (Chapter 83) or exercise limitation that is not adequately explained by the clinical examination, standard pulmonary function testing, and chest imaging. For such patients, laboratory testing of physiologic performance during exercise can be enlightening. Cardiopulmonary exercise testing, which is usually performed on a cycle ergometer or treadmill, includes monitoring of the heart rate, electrocardiography, and pulse oximetry as well as breath-by-breath measurement of tidal volume, breathing rate, oxygen consumption, and carbon dioxide production. Optional measurements include arterial blood gases and noninvasive cardiac output. Outcomes include maximal oxygen uptake ($\dot{V}O_2max$), maximal workload, maximal heart rate, ventilation parameters during exercise, and measurements of gas exchange. Results are analyzed to determine if anaerobic metabolism occurs when the study subject reaches maximal effort and to determine what limits the ability of a patient to exercise—a gas exchange abnormality, ventilatory limitation, cardiac limitation, or deconditioning. Simple tests of exercise performance, such as the 6-minute walk test, can quantify and serially assess exercise performance.[8]

BRONCHOALVEOLAR LAVAGE

Bronchoalveolar lavage can be useful for evaluation of opportunistic infections in immunocompromised hosts (Chapter 281), but its utility in the evaluation of interstitial lung disease is more controversial.[9] The procedure is generally safe, although provision must be made for the transient deterioration in gas exchange after the procedure. Oxygen supplementation is usually necessary, and intubation and mechanical ventilation are sometimes needed.

A normal bronchoalveolar lavage specimen includes 85% macrophages or more, 10 to 15% lymphocytes, 3% neutrophils or less, 1% eosinophils or less, 1% mast cells or less, and less than 5% squamous epithelial cells (which are an indicator of contamination from the upper airway). Smokers may have higher cell counts and a higher percentage of neutrophils. Increased lymphocyte counts are seen in sarcoidosis (Chapter 95), hypersensitivity pneumonitis (Chapter 94), nonspecific interstitial pneumonitis (Chapter 92), collagen vascular diseases (Chapter 92), radiation pneumonitis (Chapter 94), cryptogenic organizing pneumonia (Chapter 92), and lymphoproliferative disorders. Increased neutrophil counts are seen in idiopathic pulmonary fibrosis (Chapter 92), collagen vascular diseases (Chapter 92), infectious pneumonia (Chapter 97), aspiration pneumonia (Chapter 97), acute respiratory distress syndrome (Chapter 104), diffuse alveolar damage (Chapter 91), acute interstitial pneumonia (Chapter 92), and asbestosis (Chapter 93). Increased eosinophils can be seen in asthma (Chapter 87), bronchitis (Chapter 96), allergic bronchopulmonary aspergillosis (Chapter 339), Churg-Strauss vasculitis (Chapter 270), Hodgkin lymphoma (Chapter 186), and drug-induced lung disease (Chapter 94). If eosinophils are more than 25%, eosinophilic pneumonia is likely (Chapter 170). If lymphocytes are increased and the clinical differential diagnosis includes sarcoidosis or hypersensitivity pneumonitis, analysis of T-cell populations may be helpful; the CD4:CD8 ratio is typically increased in sarcoidosis but reduced in hypersensitivity pneumonitis. If more than 20% of macrophages stain positive for hemosiderin, diffuse alveolar hemorrhage is considered likely (Chapter 91), particularly if lavage fluid is progressively bloody in successive aliquots of lavage fluid.

Cellular constituents of bronchoalveolar lavage are usually stained for cytologic analysis for malignant cells and viral inclusions. If Langerhans cell histiocytosis (Chapter 92) is considered possible, 5% or more CD1a–positive cells supports the diagnosis. If chronic beryllium disease or beryllium sensitization is possible, a lymphocyte proliferation test in response to exposure to beryllium salts can be helpful (Chapter 93). Staining of solid material from the bronchoalveolar lavage with periodic acid–Schiff (PAS) stain for the presence of PAS-positive material is essential to the diagnosis of pulmonary alveolar proteinosis (Chapter 91). A diagnosis of lipoid pneumonia (Chapter 94), caused by the aspiration of oil, can be confirmed by an excess of lipid-laden macrophages from bronchoalveolar lavage. The presence of asbestos bodies or silica is not diagnostic of lung disease related to these substances (Chapter 93) but does indicate significant exposure.

PULMONARY FUNCTION IN OBESITY

The epidemic of obesity is manifested in many organ systems. Dyspnea, exercise limitation, and respiratory failure are more common in obese persons than in the nonobese. Asthma is more common and more severe in obese patients.[10] The effects of obesity on lung function are usually relatively modest among ambulatory patients with a body mass index (BMI) of less than 40. The most commonly observed effect of obesity on lung function is a reduction in expiratory reserve volume (the amount of air exhaled between FRC and residual volume), which is substantially reduced even in persons who are

overweight (BMI 25-30) or mildly obese (BMI 30-35). Vital capacity is reduced in obesity, but the effect is modest and highly variable. In large studies, on average, for each unit increase in BMI above 25, vital capacity or FVC is reduced by 0.5 to 0.8%. Effects of obesity on total lung capacity and FEV_1 are somewhat smaller. The FEV_1/FVC ratio and DLCO actually increase slightly with increase in BMI.

In exercise studies, the effects of obesity among ambulatory outpatients are likewise modest. Such patients have an increased work of breathing, but maximal oxygen uptake is often normal.

FUTURE DIRECTIONS

New test methods are likely to evolve, such as using "electronic nose" devices to identify volatile compounds in exhaled gases. Such efforts are encouraged by reports of dogs that can be trained to identify persons with malignant neoplasms and other conditions. For example, cancer-specific volatile carbonyl aldehydes and ketones can be identified in the exhaled breath condensate of patients with lung cancer, and these concentrations may return to normal after surgery.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Burney PG, Hooper R. Forced vital capacity, airway obstruction and survival in a general population sample from the USA. *Thorax*. 2011;66:49-54.
2. Garcia-Rio F, Calle M, Burgos F, et al. Spirometry. *Arch Bronconeumol*. 2013;49:388-401.
3. Qaseem A, Wilt TJ, Weinberger SE, et al. Diagnosis and management of stable chronic obstructive pulmonary disease: a clinical practice guideline update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. *Ann Intern Med*. 2011;155:179-191.
4. Redlich CA, Tarlo SM, Hankinson JL, et al. Official American Thoracic Society Technical Standards: spirometry in the occupational setting. *Am J Respir Crit Care Med*. 2014;189:983-993.
5. Iyer VN, Schroeder DR, Parker KO, et al. The nonspecific pulmonary function test: longitudinal follow-up and outcomes. *Chest*. 2011;139:878-886.
6. Ford ES, Mannino DM, Wheaton AG, et al. Trends in the prevalence of obstructive and restrictive lung function among adults in the United States: findings from the National Health and Nutrition Examination surveys from 1988-1994 to 2007-2010. *Chest*. 2013;143:1395-1406.
7. Schmidt SL, Tayob N, Han MK, et al. Predicting pulmonary fibrosis disease course from past trends in pulmonary function. *Chest*. 2014;145:579-585.
8. Casanova C, Celli BR, Barria P, et al. The 6-min walk distance in healthy subjects: reference standards from seven countries. *Eur Respir J*. 2011;37:150-156.
9. Meyer KC, Raghu G, Baughman RP, et al. An official American Thoracic Society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. *Am J Respir Crit Care Med*. 2012;185:1004-1014.
10. Gibeon D, Batuwita K, Osmond M, et al. Obesity-associated severe asthma represents a distinct clinical phenotype: analysis of the British Thoracic Society Difficult Asthma Registry Patient cohort according to BMI. *Chest*. 2013;143:406-414.

REVIEW QUESTIONS

1. A 58-year-old man with exercise-related cough and a body mass index of 42 has a hemoglobin level of 14 g/dL and a diffusing capacity (DLCO) that is 144% of the reference value. The most appropriate next test is

A. Bronchoalveolar lavage for hemosiderin-laden macrophages
B. Echocardiography to exclude intracardiac shunt
C. Quantitative assay for *JAK2* mutation
D. Methacholine challenge
E. Measurement of hemoglobin P_{50}

Answer: D Obesity and asthma are the most likely causes of an increased DLCO, and a search for rare causes of an increased DLCO usually is not indicated. Bronchoalveolar lavage can be useful if clinical information suggests pulmonary hemorrhage. An echocardiogram may demonstrate a left to right shunt, which is a rare cause of an increased DLCO. *JAK2* mutations are associated with polycythemia vera, but the patient is not polycythemic.

2. A 61-year-old male former smoker (40 pack-years) complains of dyspnea and cough. Pulmonary function testing shows normal spirometry and lung volumes; there is an isolated reduction in diffusing capacity (DLCO). The most useful next test is

A. Echocardiography
B. Right-sided heart catheterization
C. High-resolution computed tomography of the chest
D. Maximal respiratory pressures
E. Bronchoalveolar lavage for hemosiderin-laden macrophages

Answer: C An isolated reduction in DLCO, which is most often associated with emphysema or fibrosis or both, is seen best on computed tomography. An isolated reduction in DLCO is less often due to pulmonary vascular disorders such as pulmonary hypertension, so echocardiography and right-sided heart catheterization would have lower yields. Muscle weakness can reduce the DLCO, but it also reduces lung volumes. More than 20% hemosiderin-laden macrophages on bronchoalveolar lavage is suggestive of diffuse alveolar hemorrhage, which is a rare cause of an increased DLCO.

3. A 53-year-old never-smoker with a saddle nose deformity has severe dyspnea and dry cough. His pulmonary function test results are as follows:

FVC	4.51	93%
FEV_1	1.52	40%
FEV_1/FVC	33.7	
FEF max	2.7	31%
FIFmax	4.5	
MVV	55	36%

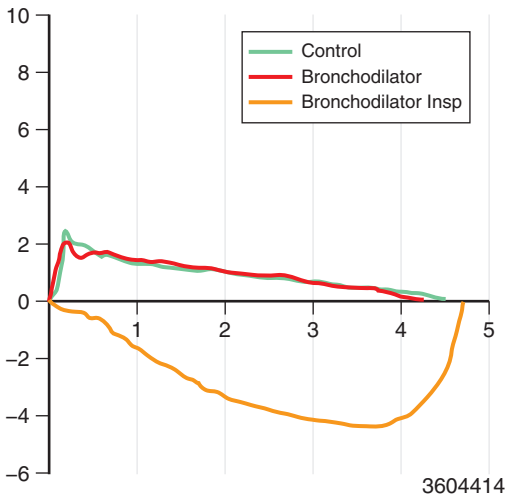

He reports episodes of ear pain and erythema, refractory to antibiotics but responsive to steroids. What is the next most appropriate test?

A. Methacholine challenge
B. Maximal respiratory pressures
C. Airway resistance
D. Imaging of the central airways (computed tomography or bronchoscopy)
E. Measurement of exhaled nitric oxide

Answer: D He has relapsing polychondritis. His main respiratory issue is central airway collapse due to chondromalacia of the tracheal and bronchial cartilage. The flow-volume curve shows characteristic flattening, as opposed to the "scooped out" pattern of asthma and COPD. He does not have a disorder of airway reactivity, so methacholine challenge adds little useful information and may not be safe with this degree of obstruction. Maximal respiratory pressures are not likely to be abnormal. Airway resistance will be abnormal but will add nothing diagnostically. Exhaled nitric oxide is abnormal in patients with eosinophilic airway inflammation and would not be expected to be abnormal in this case.

4. A patient with mild obstruction on spirometry has a maximal voluntary ventilation that is reduced out of proportion to the FEV_1. Which of the following is *least* likely to be helpful?

A. Maximal respiratory pressures
B. Inspiratory flow-volume curve
C. Cardiopulmonary exercise challenge
D. Airway resistance measurement
E. Careful scrutiny of test for repeatability of measures and technician comments on patient performance

Answer: C A disproportionate reduction in maximal voluntary ventilation may be due to inspiratory obstruction, muscle weakness, or poor performance. Cardiopulmonary exercise testing is likely to be abnormal regardless of the cause of the abnormality. The other four options would yield more specific diagnostic information.

5. A 43-year-old woman is being evaluated for dyspnea and lack of energy. Results are as follows: TLC, 79% predicted; FVC, 46%; FEV_1, 54%; FEV_1/FVC, 0.99; DLCO, 81%. The expiratory flow-volume curve is as shown:

TLC	3.56	79%
RV	1.92	145%
RV/TLC	53.9	183%
FVC	1.47	46%
FEV_1	1.45	54%
FEV_1/FVC	98.6%	
FEF max	3.7	64%
DLCO	18	81%
SPO_2	94% →	91%

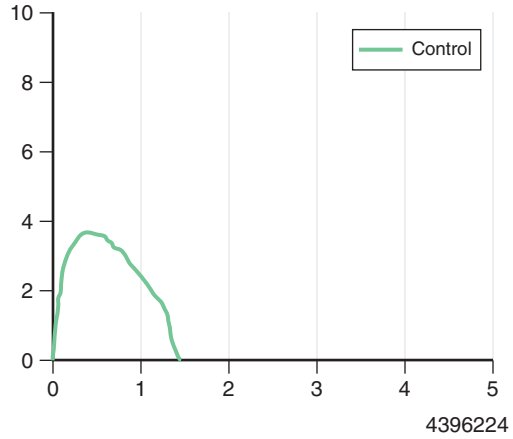

What test is likely to be most helpful?

A. Maximal respiratory pressures
B. Airway resistance
C. Methacholine challenge
D. Cardiopulmonary exercise test
E. Arterial blood gases

Answer: A The convex shape of the flow-volume curve in an adult suggests muscle weakness or poor performance. In a patient with restriction, the disproportionate reduction in FVC compared with TLC suggests muscle weakness. The patient probably has a myopathy. Alternative possibilities include chest wall limitation, poor performance, and occult airflow obstruction. The most helpful measurements on this patient will be maximal respiratory pressures, which will likely result in referral to a neurologist. Airway resistance is unlikely to be abnormal. There is little to suggest asthma, and an exercise study is likely to be abnormal but may not reveal the cause of the abnormality. Arterial blood gases are usually normal in neuromuscular disorders until the FEV_1 and FVC are severely reduced, after which hypercapnia develops as a manifestation of end-stage respiratory failure.

6. Case FX-1

50 yo M Ht = 178 cm Wt = 79 kg BMI = 25		
Dx: Tracheal stenosis, never-smoker		
	CONTROL	**%PRED**
TLC	5.73	83
RV	2.32	124
RV/TLC	0.41	148
FVC	3.40	68
FEV_1	1.51	38
FEV_1/FVC	44.5	56
MVV	11	7
FEF_{50}/FIF_{50}	1.1	110
DLCO (hb adj)	18.58	61
SpO_2	98	

%PRED = percentage of predicted value.

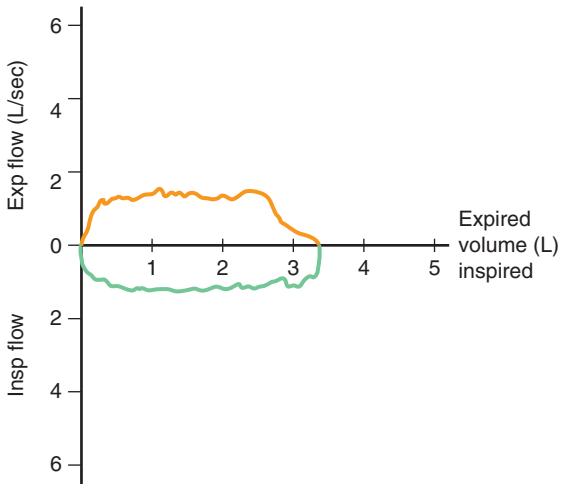

Interpretation: Abnormal. Severe fixed airway obstruction is indicated by the reduced FEV_1 and MVV and shape of the inspiratory and expiratory flow-volume curves. There is no immediate response to bronchodilator. DLCO is mildly reduced, consistent with a pulmonary parenchymal or vascular process. Lung volumes and oxygen saturations are normal.

86 DISORDERS OF VENTILATORY CONTROL

ATUL MALHOTRA AND FRANK POWELL

DEFINITIONS AND PATHOGENESIS

Ventilatory Control

Ventilation is controlled by complex interactions between central chemoreceptors, which predominantly are responsive to carbon dioxide tensions in arterial blood, and peripheral chemoreceptors, which primarily respond to carbon dioxide and oxygen tensions (Table 86-1). Disorders of ventilatory control are caused by derangements in these control systems.

HYPOVENTILATION SYNDROMES

Hypoventilation syndromes are defined by a lack of adequate alveolar ventilation to maintain a normal arterial carbon dioxide tension of 40 mm Hg. The two most common clinical settings that result in chronic hypoventilation are severe chronic obstructive pulmonary disease (COPD; Chapter 88) and morbid obesity (Chapters 100 and 220); less common causes are chronic opiate therapy, neuromuscular weakness (Chapters 421 and 422), and severe kyphoscoliosis (Chapter 99). The epidemiology of these hypoventilation syndromes is poorly studied, but about 15% of patients with severe COPD or morbid obesity have an elevated $Paco_2$. Regardless of the cause, patients with hypoventilation frequently have further worsening of their ventilation at the onset of sleep due to loss of the wakefulness stimulus, which is the normal drive to breathe while awake, and some degree of upper airway collapse after the onset of sleep (Chapter 100).

Patients with central sleep apnea (Chapter 100), which is a group of conditions in which cessation of airflow occurs because of a lack of respiratory effort, are classified into those with inadequate ventilatory drive and those with excessive drive.[1] The apparent paradox of how excessive drive leads to central apnea is explained by the concept of loop gain. A negative feedback control system with a high loop gain is prone to instability that leads to periods of excessive breathing followed by periods of apnea (Table 86-2). The prototype of a condition with high loop gain is periodic breathing or Cheyne-Stokes breathing (Fig. 86-1).

Cheyne-Stokes Breathing

Cheyne-Stokes breathing is a waxing and waning pattern of breathing, which is classically described as crescendo-decrescendo and often includes periods of central apnea. Cheyne-Stokes is seen most commonly during sleep in patients with heart failure.

EPIDEMIOLOGY

Cheyne-Stokes breathing is a form of ventilatory instability that occurs in 30 to 40% of patients with left ventricular systolic dysfunction.[2] Male sex, advanced age, low baseline $Paco_2$, and atrial fibrillation are risk factors for Cheyne-Stokes breathing among patients with heart failure. Controversy remains regarding whether this breathing pattern itself is deleterious or whether it is simply a marker of the underlying severity of cardiac disease. Cheyne-Stokes breathing represents about 5 to 10% of all cases of sleep apnea (Chapter 100) and is uncommon among patients who do not have heart failure.

PATHOBIOLOGY

Individuals with Cheyne-Stokes breathing have robust chemosensitivity as evidenced by marked increases in ventilation with small increases in $Paco_2$. The drive to breathe may be further increased by neural reflexes that are triggered by extravascular lung fluid and an elevated left atrial pressure. Intermittent hypoxemia and catecholamine surges, which are frequent in these patients, contribute to oxidative stress and neuroendocrine activation, both of which are thought to contribute to worsening of the underlying heart failure.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients with Cheyne-Stokes breathing can sometimes be diagnosed at the bedside by careful observation of their breathing pattern. During sleep or exercise, breathing becomes dependent primarily on metabolic stimuli. Patients may complain of fatigue or sleepiness because arousals from sleep tend to occur during the hyperpneic phase. Paroxysmal nocturnal dyspnea, a classic symptom of heart failure (Chapter 58), most commonly reflects

TABLE 86-1 CLASSIFICATION OF CENTRAL SLEEP APNEA

CENTRAL SLEEP APNEA SYNDROME	MECHANISM	THERAPY
Sleep transition apneas	Carbon dioxide fluctuations during transitions from sleep to wake to sleep	Reassurance, occasionally hypnotics or oxygen
Chronic narcotic therapy	Lack of central drive	Reduce narcotic dose Consider positive-pressure device
Cheyne-Stokes breathing	High loop gain from robust chemosensitivity and ventilatory drive	Optimize medical therapy for heart failure, consider PAP devices
Idiopathic central apnea	Unknown	Supportive, bilevel PAP; consider ventilatory stimulants
Treatment of emergent central apnea or "complex apnea"	Lowering upper airway resistance at CPAP initiation improves efficiency of carbon dioxide excretion	Reassurance, generally resolves spontaneously
Sleep hypoventilation syndromes	Fall in drive with loss of wakefulness stimulus, loss of accessory muscle activity during REM sleep	Noninvasive ventilation

CPAP = continuous positive airway pressure; PAP = positive airway pressure; REM = rapid eye movement.

TABLE 86-2 CLASSIFICATION OF HYPERCAPNIC DISEASES

HYPERCAPNIC DISEASE	MECHANISM	DIAGNOSIS	TREATMENT
Narcotic overdose	Reduced central drive	History, narcotized pupils, toxicology	Supportive care, naloxone
Acute severe asthma	Severe airflow obstruction, high dead space	Typical history, wheezing on examination, low FEV_1/FVC	Bronchodilators, anti-inflammatories, mechanical ventilation (usually invasive)
Acute exacerbation of COPD	Airflow obstruction, high dead space	History, cigarette smoking, low FEV_1/FVC, infectious etiology	Bronchodilators, anti-inflammatories, noninvasive ventilation
Obesity-hypoventilation syndrome	Low respiratory system compliance, high upper airway resistance, low central drive	High BMI, lack of other diagnoses; blunted carbon dioxide response	Weight loss, nocturnal bilevel positive airway pressure
Central congenital hypoventilation syndrome	*PHOX2B* mutation, lack of central drive	Genetic testing	Supportive care, mechanical ventilation (usually noninvasive)
Neuromuscular disease (e.g., myasthenia gravis, ALS, polymyositis, GBS/AIDP)	Lack of respiratory muscle force	Immediate orthopnea, low VC, low MIPs/MEPs	Underlying cause; nocturnal noninvasive ventilation; supportive care
Severe parenchymal lung disease, e.g., COPD	Lack of alveolar surface area; high pulmonary dead space and work of breathing	Typical history, smoking, low FEV_1 and FEV_1/FVC	Bronchodilator, anti-inflammatory therapy, possible nocturnal noninvasive ventilation, smoking cessation
Kyphoscoliosis	Low respiratory system compliance	Physical examination	Supportive care, noninvasive ventilation

AIDP = acute inflammatory demyelinating polyneuropathy; ALS = amyotrophic lateral sclerosis; BMI = body mass index; COPD = chronic obstructive pulmonary disease; FEV_1 = forced expiratory volume in 1 second; FVC = forced vital capacity; GBS = Guillain-Barré syndrome; MEPs = maximal expiratory pressures; MIPs = maximal inspiratory pressures; VC = vital capacity.

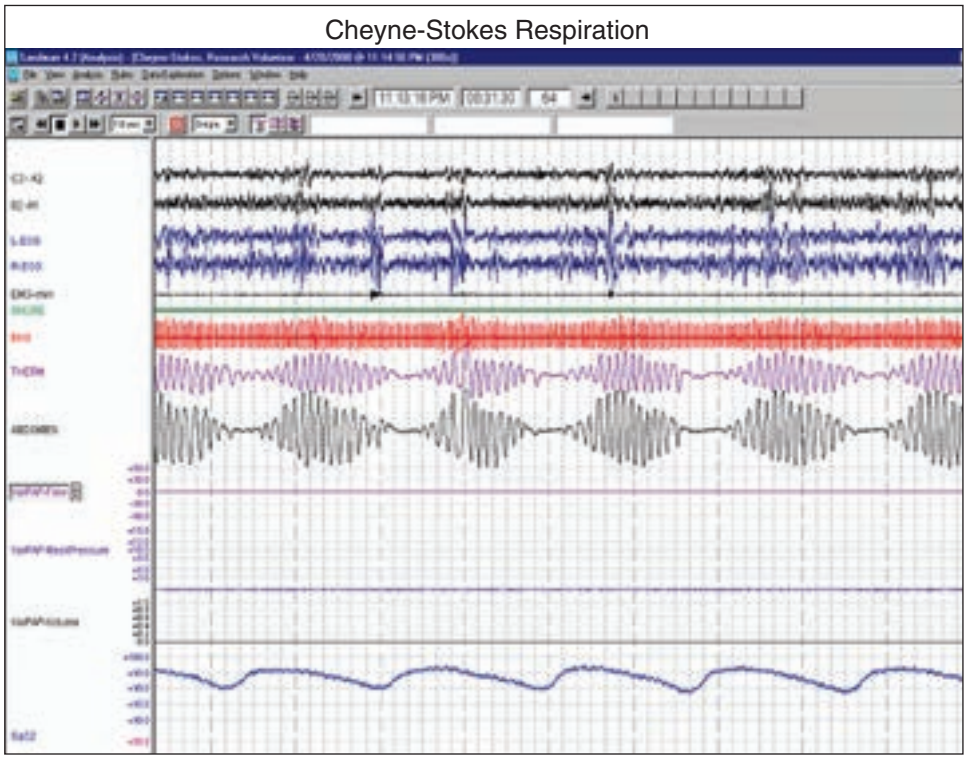

FIGURE 86-1. Cheyne-Stokes breathing with crescendo-decrescendo pattern of breathing. The thermistor detects air temperature changes at the mouth and nose. Note absences in airflow without respiratory effort seen in the abdominal belts. This breathing pattern leads to intermittent desaturations, arousals from sleep, and bursts of tachycardia. The loop gain concept can be understood by considering the thermostat analogy in which a control system is working to regulate a stable room temperature (e.g., 20°C). By analogy, the respiratory control system is working primarily to maintain a stable $Paco_2$ of 40 mm Hg and stable pH. Situations in which marked fluctuations in room temperature might occur would include one in which the thermostat is excessively sensitive (i.e., furnace turns on if room temperature falls to 19.999°C); if the furnace is too powerful, a marked overshoot in room temperature will be followed by a prolonged period when the furnace does not run. In the analogy to Cheyne-Stokes breathing, carbon dioxide is equated to room temperature and would be predicted to be unstable if chemosensitivity (i.e., the thermostat) were excessively robust (i.e., a marked increase in ventilation for a small change in carbon dioxide) or if the efficiency of carbon dioxide excretion were high (i.e., marked fall in $Paco_2$ with increased ventilation). Situations that increase the propensity for carbon dioxide fluctuations lead to elevated loop gain and thus increase the risk for Cheyne-Stokes breathing.

underlying Cheyne-Stokes breathing. Patients often are diagnosed in the sleep laboratory while undergoing investigation for possible obstructive sleep apnea.

The diagnosis of Cheyne-Stokes breathing, if it is not readily apparent, can be made during overnight polysomnography, when the typical oscillatory pattern of tidal volume is seen in the absence of ventilatory efforts during the apneic periods. In evaluating such recordings, and in contrast to obstructive apnea, it is important to note that Cheyne-Stokes breathing usually resolves during rapid eye movement (REM) sleep, that arousals on the electroencephalogram typically occur during the hyperpneic phase, and that Cheyne-Stokes breathing generally does not resolve immediately when nasal continuous positive airway pressure (CPAP) is applied.

TREATMENT

Medical management of Cheyne-Stokes breathing most often is treatment of the underlying heart failure (Chapter 59). After optimization of medical management, the Cheyne-Stokes breathing pattern frequently resolves. CPAP can improve breathing indices but is no better than standard medical therapy from the standpoint of mortality. Newer approaches to non-invasive ventilation show promise but require further evaluation.[3]

Central Congenital Hypoventilation Syndrome

DEFINITION AND EPIDEMIOLOGY

Central congenital hypoventilation syndrome is a rare congenital condition, previously referred to as Ondine curse, characterized by a diminished ventilatory response to carbon dioxide.[4] The central congenital hypoventilation syndrome was traditionally diagnosed in neonates, but more subtle forms of disease are increasingly noted in older children and adults.

PATHOBIOLOGY

The syndrome is now defined by a mutation in the *PHOX2B* gene, located on chromosome 4p12.[5] The *PHOX2B* gene is a highly conserved homeobox gene that is expressed mainly in the afferent and efferent pathways of respiratory, cardiovascular, and digestive reflexes. Deletion of the gene in mice causes irregular breathing, a reduced hypercapnic ventilatory response, and death from central apnea. These mice have neuronal loss in the retrotrapezoid nucleus and parafacial region of the brain stem, thereby suggesting the importance of this medullary region in normal breathing. Abnormalities in *PHOX2B* genes have also been associated with Hirschsprung disease (Chapter 136), neural crest tumors, cardiac asystole (Chapter 63), and other abnormalities of the autonomic nervous system (Chapter 418).

Because most parents of affected children with the central congenital hypoventilation syndrome do not carry a *PHOX2B* mutation, the mutations are de novo. About 90% of patients are heterozygous for a polyalanine repeat expansion mutation, in which the affected allele has 24 to 33 alanines rather than the normal 20 alanines. The remaining 10% of central congenital hypoventilation syndrome patients have missense, nonsense, or frameshift mutations in the *PHOX2B* gene.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Neonates can present with cyanosis at birth, recurrent central apneas, or both. Adults can present with idiopathic central sleep apnea, unexplained hypercapnia, or autonomic abnormalities (Chapter 418). Confirmation of the diagnosis requires the demonstration of an abnormality in the *PHOX2B* gene.

TREATMENT

There are currently no specific therapies for central congenital hypoventilation syndrome beyond supportive care. Genetic counseling is required for afflicted individuals and their families, given the autosomal dominant pattern of inheritance. Patients must be cautioned against the use of sedatives, which could precipitate respiratory failure. Mechanical ventilation during sleep either invasively (through tracheostomy) or noninvasively (through bilevel positive airway pressure support [Chapter 100]) is required in most patients. Some patients remain fully ventilator dependent. Alternative treatments, such as ventilatory stimulants and diaphragmatic pacing, are generally ineffective.

Acquired Hypoventilation Syndromes

DEFINITION AND EPIDEMIOLOGY

Patients with hypoventilation syndromes cannot maintain adequate minute ventilation to keep their $Paco_2$ at 40 mm Hg. Patients can be classified into those who lack central ventilatory drive and those who have a pulmonary mechanical or neuromuscular abnormality that prevents adequate gas exchange. The case frequency is unknown, but hypercapnic respiratory failure is one of the more common admission diagnoses in intensive care units.

PATHOBIOLOGY

Patients with conditions characterized by the lack of central drive have reasonably normal lungs and respiratory muscle function but lack adequate response to carbon dioxide and hypoxia. In contrast, most patients with mechanical or neuromuscular abnormalities have a larger work of breathing compared with normal individuals; the most common underlying conditions are severe COPD (Chapter 88) and morbid obesity (Chapter 220) with the obesity-hypoventilation syndrome. Such individuals have diminished but not absent chemoresponsiveness. Another cause of inadequate gas exchange is neuromuscular disease; common causes include disorders of neuromuscular transmission (Chapter 422), severe muscle weakness (Chapter 421), the residua from poliovirus infection (Chapter 379), Guillain-Barré syndrome (Chapter 420), and acute poisoning (Chapter 110).

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients with hypoventilation have myriad presentations ranging from asymptomatic abnormalities in laboratory testing (e.g., elevated $Paco_2$, unexplained low Sao_2, or elevated serum bicarbonate level) to respiratory failure in the intensive care unit (e.g., respiratory infection with laboratory evidence of chronic abnormalities, such as acute-on-chronic respiratory acidosis). Patients who acutely overdose on sedative-hypnotic or narcotic agents may present with acute respiratory acidosis and loss of consciousness. Patients who take chronic narcotics may present with central sleep apnea-hypopnea or otherwise unexplained oxygen desaturation at night.

Once it is suspected, the diagnosis of hypoventilation is confirmed by the finding of $Paco_2$ higher than 42 mm Hg on analysis of an arterial blood sample. If the increase in $Paco_2$ is of short duration so that renal compensation has not yet occurred (Chapter 118), the serum bicarbonate level is increased by 1 mEq/L for every rise of 10 mm Hg in $Paco_2$. By comparison, if the respiratory acidosis is of sufficient duration for renal compensation to occur, the serum bicarbonate level will be increased by 4 mEq for every rise of 10 mm Hg in $Paco_2$ (Fig. 86-2).

Once an elevated $Paco_2$ is established, it is appropriate to distinguish patients who "can't breathe" from those who "won't breathe." "Can't breathe" implies that a respiratory mechanical problem or neuromuscular weakness is causing the elevation in $Paco_2$. Abnormalities in pulmonary function testing (e.g., a very low vital capacity) suggest a parenchymal or chest wall disorder. Ultrasound can identify phrenic neuropathy causing diaphragmatic dysfunction.[6] Patients who "won't breathe" have central nervous system abnormalities that affect central drive, chemosensitivity, or both.

TREATMENT AND PROGNOSIS

The treatment of hypoventilation should focus on the underlying cause. Acute poisonings can be managed supportively or, in some cases, with specific antidotes (Chapter 110). Chronic conditions can be treated by addressing the underlying cause, such as weight loss in obesity-hypoventilation syndrome or cholinesterase inhibitors in myasthenia gravis (Chapter 422). For parenchymal lung disease, treatment is directed at the underlying cause, if possible (Chapters 88 and 92).

Sedative medications should be used cautiously because they can occasionally precipitate acute respiratory failure. Although profound hypoxemia can clearly be deleterious, oxygen occasionally can precipitate severe acute respiratory acidosis, particularly in patients with acute exacerbations of COPD (Chapter 88). As a result, hypoventilating patients with COPD require cautious management including the careful administration of supplemental oxygen, which should be titrated to an arterial oxygen saturation of 90% or an arterial oxygen tension of 60 mm Hg.

Severe hypoventilation requires mechanical ventilation (Chapter 105), such as noninvasive ventilation for an acute exacerbation of COPD. For other presentations in which the $Paco_2$ is believed to be acutely elevated, endotracheal intubation and mechanical ventilation are frequently used, especially in patients with impaired consciousness. For chronic hypoventilation in hypercapnic COPD, noninvasive bilevel positive airway pressure through a face mask during sleep can maintain alveolar ventilation, but there is no definitive evidence that noninvasive positive-pressure ventilation can prolong life or reduce hospitalizations in patients with COPD and chronic respiratory failure. In addition, the considerable difficulty of adhering to nocturnal bilevel therapy in COPD emphasizes the need for discussions with patients and families regarding its risks and benefits.

Other chronic hypoventilation syndromes are also commonly treated with bilevel positive airway pressure, although data are not compelling. In some chronic conditions, such as motor neuron disease (Chapter 419), tracheostomy should be discussed, although the impact of such interventions on quality of life should be carefully considered. Regardless of the underlying cause, an elevation in the $Paco_2$ level is considered a poor prognostic sign. End-of-life discussions are also important in such cases because the prognosis of patients with chronic respiratory failure is generally poor.

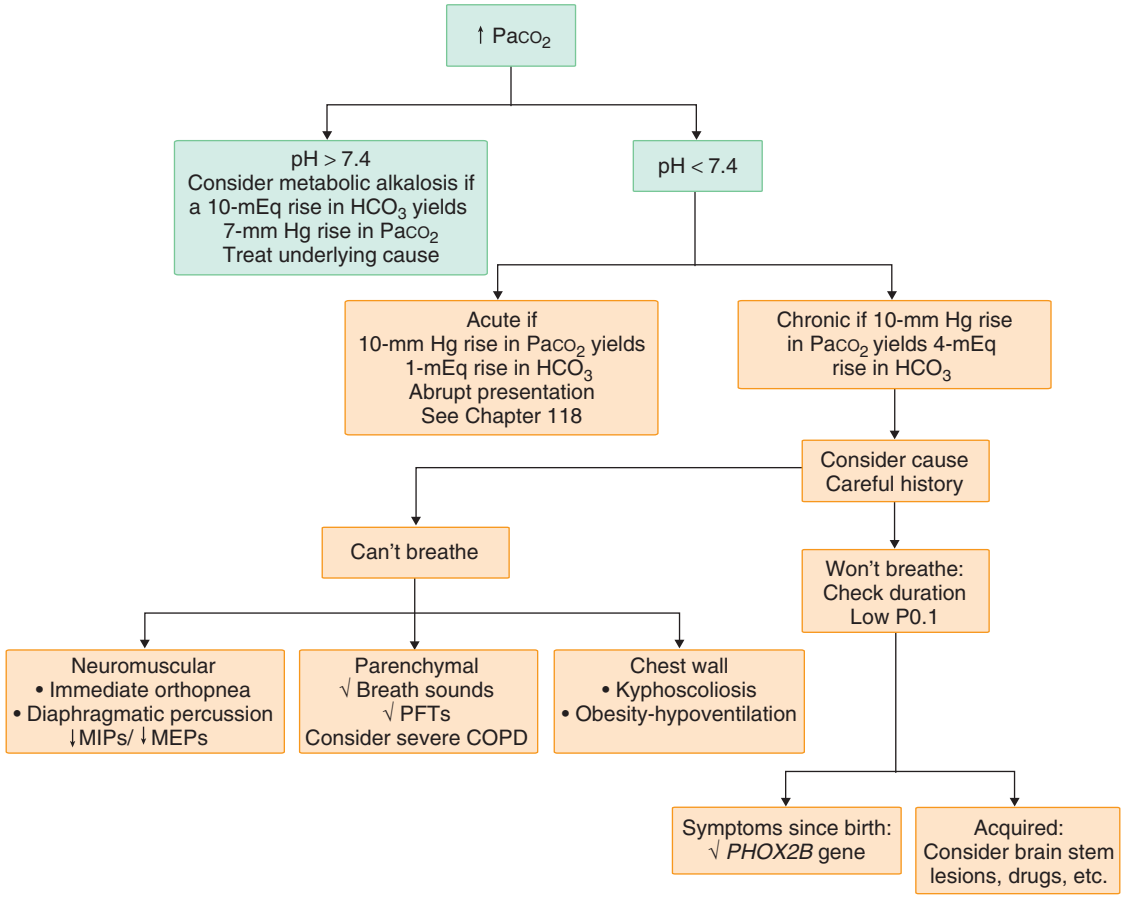

FIGURE 86-2. A flow chart of a systematic approach to hypercapnia and various causes of hypoventilation. The change in pH can help determine the cause and chronicity. A careful history and physical examination, coupled with pulmonary function testing, can help classify patients into those who "can't breathe" because of neuromuscular or mechanical abnormalities of the respiratory system compared with those who "won't breathe" because of central nervous system disease. COPD = chronic obstructive pulmonary disease; MEPs = maximal expiratory pressures; MIPs = maximal inspiratory pressures; P0.1 = the negative mouth pressure generated during the first 100 msec of an occluded inspiration; PFTs = pulmonary function tests.

Grade A References

A1. Bradley TD, Logan AG, Kimoff RJ, et al. Continuous positive airway pressure for central sleep apnea and heart failure. *N Engl J Med*. 2005;353:2025-2033.

A2. Kohnlein T, Windisch W, Kohler D, et al. Non-invasive positive pressure ventilation for the treatment of severe stable chronic obstructive pulmonary disease: a prospective, multicentre, randomised, controlled clinical trial. *Lancet Respir Med*. 2014;2:698-705.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Pack AI. Central sleep apnea. *Handb Clin Neurol*. 2011;98:411-419.
2. McGee S. Cheyne-stokes breathing and reduced ejection fraction. *Am J Med*. 2013;126:536-540.
3. Combs D, Shetty S, Parthasarathy S. Advances in positive airway pressure treatment modalities for hypoventilation syndromes. *Sleep Med Clin*. 2014;9:315-325.
4. Weese-Mayer DE, Berry-Kravis EM, Ceccherini I, et al. An official ATS clinical policy statement. Congenital central hypoventilation syndrome: genetic basis, diagnosis, and management. *Am J Respir Crit Care Med*. 2010;181:626-644.
5. Rand CM, Yu M, Jennings LJ, et al. Germline mosaicism of PHOX2B mutation accounts for familial recurrence of congenital central hypoventilation syndrome (CCHS). *Am J Med Genet A*. 2012;158A:2297-2301.
6. Boon AJ, Sekiguchi H, Harper CJ, et al. Sensitivity and specificity of diagnostic ultrasound in the diagnosis of phrenic neuropathy. *Neurology*. 2014;83:1264-1270.

REVIEW QUESTIONS

1. Which of the following is currently the treatment of choice for Cheyne-Stokes breathing in congestive heart failure?

A. Optimize medical therapy
B. Nasal continuous positive airway pressure
C. Nasal bilevel positive airway pressure
D. Oral appliance therapy
E. Uvulopalatopharyngoplasty

Answer: A The treatment of choice for Cheyne-Stokes breathing is currently optimization of medical therapy for the underlying heart failure. Trials of nasal continuous positive airway pressure have failed to improve outcome compared with usual care. Bilevel therapy has not been rigorously studied but may make the situation worse. Although the upper airway can sometimes narrow or collapse in central apnea, there is no role for oral appliance therapy or uvulopalatopharyngoplasty in the absence of obstructive sleep apnea.

2. Which of the following best defines the measurement of loop gain in control of breathing?

A. A measure of cardiac output
B. A measure of extravascular lung fluid
C. A measure of hemoglobin concentration
D. A measure of the tendency toward instability in the ventilatory control system
E. A measure of partial pressure of oxygen in the arterial circulation

Answer: D Loop gain refers to the overall instability in a negative feedback control system, such as the ventilatory control system, whose role is to maintain $Paco_2$ levels. The other factors listed can all influence control of breathing in some way but are not defining loop gain.

3. Which of the following is the gene associated with the central congenital hypoventilation syndrome?

A. *CFTR*
B. *PHOX2B*
C. *A1AT*
D. *RET*
E. *BMP*

Answer: B The *PHOX2B* gene is the hallmark for the diagnosis of central congenital hypoventilation syndrome. The other genes have been associated with various respiratory conditions but not with the central congenital hypoventilation syndrome.

4. Which of the following is true regarding obesity-hypoventilation syndrome?

A. Serum bicarbonate is a useful screening test.
B. Leptin deficiency is generally seen in afflicted humans.
C. Obstructive sleep apnea is uncommon in afflicted patients.
D. It is present in roughly 50% of patients with obstructive sleep apnea.
E. Hypercapnia usually persists despite major weight loss.

Answer: A Serum bicarbonate is a useful screening test because it is elevated in the majority of patients with chronic hypoventilation. Leptin is deficient in some animal models but rare in human obesity. Obstructive sleep apnea is common in the obesity-hypoventilation syndrome, and obesity-hypoventilation syndrome is seen in roughly 10% of obstructive sleep apnea. Weight loss usually improves gas exchange in these patients.

5. Which of the following is true regarding sleep stages in sleep apnea?

A. Obstructive sleep apnea is generally worst during slow wave sleep.
B. Cheyne-Stokes breathing is generally worst during REM sleep.
C. Periodic breathing at altitude usually resolves during REM sleep.
D. Obstructive sleep apnea is rare during REM sleep.
E. Arousal in obstructive sleep apnea has no major effect on breathing.

Answer: C Both Cheyne-Stokes and periodic breathing at altitude are improved in REM compared with NREM sleep. Slow wave sleep is associated with improvement in obstructive sleep apnea. Arousal in obstructive sleep apnea can serve to restore pharyngeal airway patency and allow the resumption of effective tidal breathing.

87

ASTHMA

JEFFREY M. DRAZEN

DEFINITION

Asthma is a clinical syndrome of unknown etiology characterized by three distinct components: (1) recurrent episodes of airway obstruction that resolve spontaneously or as a result of treatment; (2) exaggerated bronchoconstrictor responses to stimuli that have little or no effect in nonasthmatic subjects, a phenomenon known as airway hyperresponsiveness; and (3) inflammation of the airways as defined by a variety of criteria. Although airway obstruction is largely reversible, it is currently thought that changes in the asthmatic airway may be irreversible in some settings.

EPIDEMIOLOGY

Asthma is an extremely common disorder affecting boys more commonly than girls and, after puberty, women slightly more commonly than men; approximately 8% of the adult population of the United States has signs and symptoms consistent with a diagnosis of asthma. Although most cases begin before the age of 25 years, new-onset asthma may develop at any time throughout life.

The worldwide prevalence of asthma has increased more than 45% since the late 1970s. In the last decade alone, the prevalence of wheezing in children has increased by about 0.1% per year.

During the past four decades, the greatest increases in the prevalence of asthma have occurred in countries that adopted an "industrialized" lifestyle. For example, epidemiologic data suggest that being raised in a farming environment is associated with a much lower risk of asthma, independent of genetic factors, and this difference may be attributable to exposure to a greater diversity of environmental microbes early in life.[1]

Asthma is among the most common reasons to seek medical treatment. In the United States, it is responsible for about 15 million annual outpatient visits to physicians and for nearly 2 million annual inpatient hospital days of treatment. The estimated yearly direct and indirect costs of asthma care in the United States are more than $55 billion.

PATHOBIOLOGY

Genetics

In twin studies, asthma has about 60% heritability, indicating that both genetic and environmental factors are important in its etiology. A region on chromosome 17q21, at or near the locus for *ORMDL3*, a member of a gene family that encodes endoplasmic reticulum transmembrane proteins, has been repeatedly associated with childhood-onset asthma. Although the exact

functional variant in this region has not been identified, the isolation of a locus for childhood-onset asthma supports the clinical observation that adult- and childhood-onset asthma appear to be distinct disorders. Genetic variants that influence the response to treatment also have been identified and widely replicated.

Pathology

The pathology of mild asthma, as delineated by bronchoscopic and biopsy studies, is characterized by edema and hyperemia of the mucosa and by infiltration of the mucosa with mast cells, eosinophils, and lymphocytes bearing the T_H2 phenotype. Controlled trials using antibodies against interleukin-5 or the interleukin-4 receptor α chain in patients with persistent asthma symptoms and eosinophilia, despite treatment with corticosteroids, provide solid evidence for the pathobiologic role of these inflammatory cytokines in asthma. As a result of these inflammatory stimuli coupled with the mechanical deformation of the epithelium from airway, smooth muscle constriction,[2] the airway wall is thickened by the deposition of type III and type V collagen below the true basement membrane. In addition, in severe chronic asthma, there is hypertrophy and hyperplasia of airway glands and of both surface and glandular secretory cells as well as hyperplasia of airway smooth muscle. Morphometric studies of airways from asthmatic subjects have demonstrated airway wall thickening of sufficient magnitude to increase airflow resistance and to enhance airway responsiveness. During a severe asthmatic event, the airway wall is thickened markedly; in addition, patchy airway occlusion occurs by a mixture of hyperviscous mucus and clusters of shed airway epithelial cells.

The episodic airway obstruction that constitutes an asthma attack results from narrowing of the airway lumen to airflow. Although it is now well established that asthma is associated with infiltration of the airway by inflammatory cells, the links between the presence of these cells and the pathobiologic processes that account for asthmatic airway obstruction are just beginning to be delineated. Three possible but not mutually exclusive links have been postulated: the constriction of airway smooth muscle, the thickening of airway epithelium, and the presence of liquids within the confines of the airway lumen. Among these mechanisms, the constriction of airway smooth muscle due to the local release of bioactive mediators or neurotransmitters is the most widely accepted explanation for the acute reversible airway obstruction in asthma attacks. Several bronchoactive mediators are thought to be the agents that initiate the airway obstruction characteristic of asthma. Moreover, the chronic airway narrowing, termed airway wall remodeling, that occurs in many patients with asthma likely results from the actions of inflammatory cells in the asthmatic airway.

Mediators of the Acute Asthmatic Response

Acetylcholine

Acetylcholine released from intrapulmonary motor nerves causes constriction of airway smooth muscle through direct stimulation of muscarinic receptors of the M_3 subtype. The potential role for acetylcholine in the bronchoconstriction of asthma primarily derives from the observation that tiotropium bromide, a muscarinic antagonist, can reduce bronchoconstriction.

Histamine

Histamine, or β-imidazolylethylamine, was identified as a potent endogenous bronchoactive agent more than 100 years ago. Mast cells, which are prominent in airway tissues obtained from patients with asthma, constitute the major pulmonary source of histamine. Clinical trials with novel potent antihistamines indicate only a minor role for histamine as a mediator of airway obstruction in asthma.

Leukotrienes and Lipoxins

The cysteinyl leukotrienes, namely, LTC_4, LTD_4, and LTE_4, as well as the dihydroxy leukotriene LTB_4 are derived by the lipoxygenation of arachidonic acid released from target cell membrane phospholipids during cellular activation. 5-Lipoxygenase, the 5-lipoxygenase–activating protein, and LTC_4 synthase make up the cellular protein and enzyme content needed to produce the cysteinyl leukotrienes. The production of LTB_4 requires 5-lipoxygenase, the 5-lipoxygenase–activating protein, and LTA_4 epoxide hydrolase. Mast cells, eosinophils, and alveolar macrophages have the enzymatic capability to produce cysteinyl leukotrienes from their membrane phospholipids, whereas polymorphonuclear leukocytes produce exclusively LTB_4, which is predominantly a chemoattractant molecule; LTC_4 and LTD_4 are among the most potent contractile agonists ever identified for human airway smooth muscle. The efficacy of a leukotriene receptor antagonist (i.e., pranlukast, zafirlukast, and montelukast) or a synthesis inhibitor (i.e., zileuton) in the treatment of chronic persistent asthma has led to the conclusion that the leukotrienes are important but not exclusive mediators of the asthmatic response. Lipoxins, which are double lipoxygenase products of arachidonic acid metabolism, have been shown to be endogenous downregulators of the inflammatory response. The amounts of lipoxins are decreased in the airways of patients with severe asthma.

Nitric Oxide

Nitric oxide (NO•) is produced enzymatically by airway epithelial cells and by inflammatory cells found in the asthmatic lung. Free NO• has a half-life on the order of seconds in the airway and is stabilized by conjugation to thiols to form RS-NO, where *R* designates any one of a number of molecular entities that can support this chemical linkage. Both NO• and RS-NO have bronchodilator actions and may play a homeostatic role in the airway. Paradoxically, high levels of NO•, when it is coavailable with superoxide anion, may form toxic oxidation products, such as peroxynitrite ($OONO^-$), which could damage the airway. Patients with asthma have higher than normal levels of NO• in their expired air, and these levels decrease consistently after treatment with corticosteroids.

Physiological Changes in Asthma

An increased resistance to airflow is the consequence of the airway obstruction induced by smooth muscle constriction, thickening of the airway epithelium, or free liquid within the airway lumen. Obstruction to airflow is manifested by increased airway resistance and decreased flow rates throughout the vital capacity. At the onset of an asthma attack, obstruction occurs at all airway levels; as the attack resolves, these changes are reversed—first in the large airways (i.e., mainstem, lobar, segmental, and subsegmental bronchi) and then in the more peripheral airways. This anatomic sequence of onset and reversal is reflected in the physiological changes observed during resolution of an asthmatic episode. Specifically, as an asthma attack resolves, flow rates first normalize at volumes high in the vital capacity and only later at volumes low in the vital capacity. Because asthma is an airway disease, not an air space disease, no primary changes occur in the static pressure-volume curve of the lungs. However, during an acute attack of asthma, airway narrowing may be so severe as to result in airway closure, with individual lung units closing at a volume that is near their maximal volume. This closure results in a change of the pressure-volume curve such that for a given

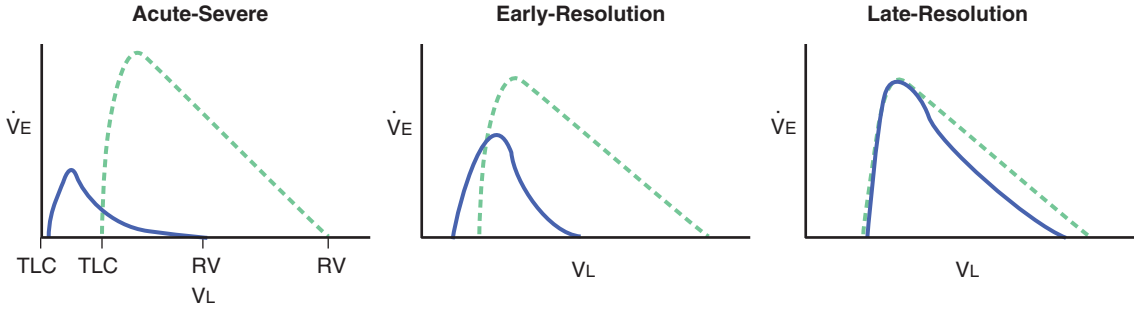

FIGURE 87-1. Schematic flow-volume curves in various stages of asthma. In each figure, the *dashed line* depicts the normal flow-volume curve. Predicted and observed total lung capacity (TLC) and residual volume (RV) are shown at the extremes of each curve. $\dot{V}_E$ = expiratory flow rate; V_L = lung volume.

contained gas volume within the thorax, elastic recoil is decreased, which in turn further depresses expiratory flow rates.

Additional factors influence the mechanical behavior of the lungs during an acute attack of asthma. During inspiration in an asthma attack, the maximal inspiratory pleural pressure becomes more negative than the subatmospheric pressure of 4 to 6 cm H_2O usually required for tidal airflow. The expiratory phase of respiration also becomes active as the patient tries to force air from the lungs. As a consequence, peak pleural pressures during expiration, which normally are, at most, only a few centimeters of water above atmospheric pressure, may be as high as 20 to 30 cm H_2O above atmospheric pressure. The low pleural pressures during inspiration tend to dilate airways, whereas the high pleural pressures during expiration tend to narrow airways. During an asthma attack, the wide pressure swings, coupled with alterations in the mechanical properties of the airway wall, lead to a much higher resistance to expiratory airflow than to inspiratory airflow.

The respiratory rate is usually rapid during an acute asthmatic attack. This tachypnea is driven not by abnormalities in arterial blood gas composition but rather by stimulation of intrapulmonary receptors with subsequent effects on central respiratory centers. One consequence of the combination of airway narrowing and rapid airflow rates is a heightened mechanical load on the ventilatory pump. During a severe attack, the load can increase the work of breathing by a factor of 10 or more and can predispose to fatigue of the ventilatory muscles. With respect to gas exchange, the patchy nature of asthmatic airway narrowing results in a maldistribution of ventilation (V) relative to pulmonary perfusion (Q). A shift occurs from the normal preponderance of V/Q units, with a ratio of near unity, to a distribution with a large number of alveolar-capillary units, with a V/Q ratio of less than unity. The net effect is to induce arterial hypoxemia. In addition, the hyperpnea of asthma is reflected as hyperventilation with a low arterial P_{CO_2}.

CLINICAL MANIFESTATIONS

History

During an acute asthma attack, patients seek medical attention for shortness of breath accompanied by cough, wheezing, and anxiety. The degree of breathlessness experienced by the patient is not closely related to the degree of airflow obstruction but is often influenced by the acuteness of the attack. Dyspnea may occur only with exercise (exercise-induced asthma),[3] after aspirin ingestion (aspirin-exacerbated respiratory disease),[4] after exposure to a specific known allergen (extrinsic asthma), or for no identifiable reason (intrinsic asthma). Variants of asthma exist in which cough, hoarseness, or inability to sleep through the night is the only symptom. Identification of a provoking stimulus through careful questioning helps establish the diagnosis of asthma and may be therapeutically useful if the stimulus can be avoided. Most patients with asthma complain of shortness of breath when they are exposed to rapid changes in the temperature and humidity of inspired air. For example, during the winter months in less temperate climates, patients commonly become short of breath on leaving a heated house; in warm humid climates, patients may complain of shortness of breath on entering a cold dry room, such as an air-conditioned theater.

An important factor to consider in taking a history from a patient with asthma is the potential for occupational exposures in asthma (Chapter 93). Asthma that is brought on by occupational exposures is termed occupational asthma; preexisting asthma that is exacerbated by workplace exposures is termed workplace-exacerbated asthma. In reactive airway dysfunction syndrome, a single large exposure leads to a persistent asthma-like phenotype in a previously normal individual.[5]

Physical Examination

Vital Signs

Common features noted during an acute attack of asthma include a rapid respiratory rate (often 25 to 40 breaths per minute), tachycardia, and pulsus paradoxus (an exaggerated inspiratory decrease in the systolic pressure). The magnitude of the pulsus is related to the severity of the attack; a value greater than 15 mm Hg indicates an attack of moderate severity. Pulse oximetry, with the patient respiring ambient air, commonly reveals an oxygen saturation near 90%.

Thoracic Examination

Inspection may reveal that patients experiencing acute attacks of asthma are using their accessory muscles of ventilation; if so, the skin over the thorax may be retracted into the intercostal spaces during inspiration. The chest is usually hyperinflated, and the expiratory phase is prolonged relative to the inspiratory phase. Percussion of the thorax demonstrates hyperresonance, with loss of the normal variation in dullness due to diaphragmatic movement; tactile fremitus is diminished. Auscultation reveals wheezing, which is the cardinal physical finding in asthma but does not establish the diagnosis (Chapter 83). Wheezing, commonly louder during expiration but heard during inspiration as well, is characterized as polyphonic in that more than one pitch may be heard simultaneously (Video 87-1). Accompanying adventitious sounds may include rhonchi, which are suggestive of free secretions in the airway lumen, or rales, which should raise the suspicion of an alternative diagnosis and are indicative of localized infection or heart failure. The loss of intensity or the absence of breath sounds in a patient with asthma is an indication of severe airflow obstruction.

DIAGNOSIS

Laboratory Findings

Pulmonary Function Findings

A decrease in airflow rates throughout the vital capacity is the cardinal pulmonary function abnormality during an asthmatic episode.[6,7] The peak expiratory flow rate (PEFR), the forced expiratory volume in the first second (FEV_1), and the maximal mid-expiratory flow rate (MMEFR) are all decreased in asthma (Chapter 85). In severe asthma, dyspnea may be so severe as to prevent the patient from performing a complete spirogram. In this case, if 2 seconds of forced expiration can be recorded, useful values for PEFR and FEV_1 can be obtained. Gradation of attack severity (Table 87-1) *must* be assessed by objective measures of airflow; no other methods yield accurate and reproducible results. As the attack resolves, the PEFR and the FEV_1 increase toward normal together while the MMEFR remains substantially depressed; as the attack resolves further, the FEV_1 and the PEFR may normalize while the MMEFR remains depressed (see Fig. 87-1). Even when the attack has fully resolved clinically, residual depression of the MMEFR is not uncommon; this depression may resolve during a prolonged course of treatment. If the patient is able to cooperate such that more complete measurements of lung function can be made, lung volume measurements made during an attack demonstrate an increase in both total lung capacity and residual volume; the changes in total lung capacity and residual volume resolve with treatment. Because of the extra cooperation needed for this testing, it is not advised during an acute asthmatic event but is indicated before discharge in a patient hospitalized for the treatment of asthma or between episodes of asthma.

Pulmonary function testing obtained when the patient is relatively stable usually demonstrates airway obstruction, as indicated by low FEV_1 (as a percentage of the patient's predicted value), low forced vital capacity, and slightly elevated total lung capacity and residual volume values. These results may fully normalize after administration of a bronchodilator, but a "bronchodilator response" is canonically defined as a 12% increase in the FEV_1, provided it is at least 200 mL (E-Fig. 87-1).

TABLE 87-1 RELATIVE SEVERITY OF AN ASTHMATIC ATTACK AS INDICATED BY PEFR, FEV_1, AND MMEFR

TEST	PREDICTED VALUE (%)	SEVERITY OF ASTHMA
PEFR	>80	No spirometric abnormalities
FEV_1	>80	
MMEFR	>80	
PEFR	>80	Mild asthma
FEV_1	>70	
MMEFR	55-75	
PEFR	>60	Moderate asthma
FEV_1	45-70	
MMEFR	30-50	
PEFR	<50	Severe asthma
FEV_1	<50	
MMEFR	10-30	

FEV_1 = forced expiratory volume in the first second; MMEFR = maximal mid-expiratory flow rate; PEFR = peak expiratory flow rate.

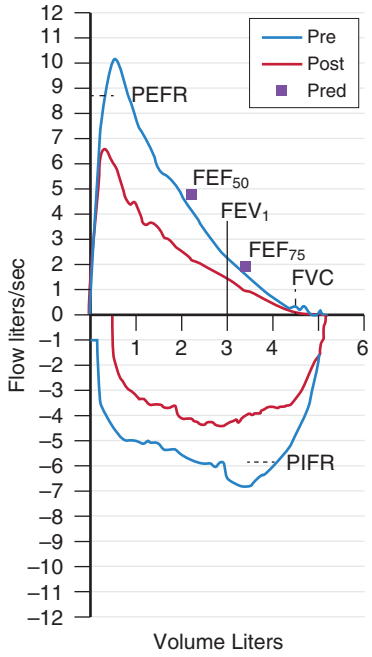

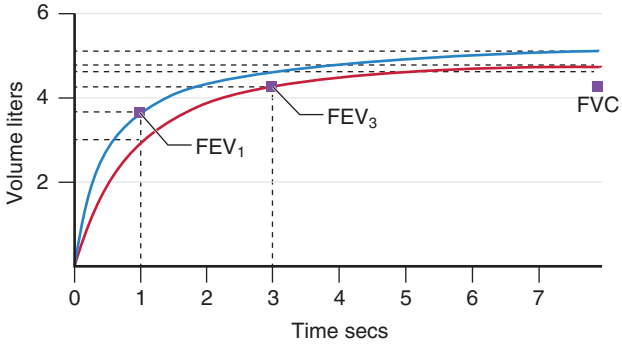

Spirometry (BTPS)

		Pre Bronchodilator			*Predicted Range*		Post Bronchodilator	*Percent*
		Actual	%Pred	Mean	95% CI	Actual	%Pred	Change
FVC	(L)	4.78	105	4.51	3.46	5.12	113	7
FEV_1	(L)	3.01	80	3.75	2.89	3.70	98	22
FEV_1/FVC	(%)	63.00	75	83.00	75.00	72.00	86	14
FEF_{25-75}	(L/s)	1.87	46	4.03		2.64	65	41
PEFR	(L/s)	6.64	75	8.74		10.17	116	53
FET	(sec)	7.89				7.89		

E-FIGURE 87-1. **Lung function test results of a patient with asthma before and after treatment with inhaled albuterol.**

Exhaled NO•

The fraction of NO• in the exhaled air (Fe_{NO}) is elevated in patients with asthma. Although the exact concentration considered "elevated" will vary with the details of the technique used to obtain the gas sample, a concentration of 15 parts per billion is a convenient and reliable level that can be used to distinguish people without asthma from patients with untreated asthma. However, the measurement of exhaled nitric oxide has not been shown to be of value in the day-to-day management of asthma.

Arterial Blood Gases

Blood gas analysis need not be undertaken in individuals with mild asthma. If the asthma is of sufficient severity to merit prolonged observation, however, blood gas analysis is indicated; in such cases, hypoxemia and hypocapnia are the rule. With the subject breathing ambient air, the Pa_{O_2} is usually between 55 and 70 mm Hg and the Pa_{CO_2} between 25 and 35 mm Hg. At the onset of the attack, an appropriate pure respiratory alkalemia is usually evident; with attacks of prolonged duration, the pH returns toward normal as a result of a compensatory metabolic acidemia. A normal Pa_{CO_2} in a patient with moderate to severe airflow obstruction is reason for concern because it may indicate that the mechanical load on the respiratory system is greater than can be sustained by the ventilatory muscles and that respiratory failure is imminent. When the Pa_{CO_2} increases in such settings, the pH decreases quickly because the bicarbonate stores have become depleted as a result of renal compensation for the prolonged preceding respiratory alkalemia. Because this chain of events can take place rapidly, close observation is indicated for asthmatic patients with "normal" Pa_{CO_2} levels and moderate to severe airflow obstruction.

Other Blood Findings

Asthmatic subjects are frequently atopic; thus, blood eosinophilia is common but not universal. In addition, elevated serum levels of immunoglobulin E (IgE) are often documented; epidemiologic studies indicate that asthma is unusual in subjects with low IgE levels. If indicated by the patient's history, specific immunosorbent tests, which measure IgE directed against specific offending antigens, can be conducted. In rare instances during severe asthma attacks, serum concentrations of aminotransferases, lactate dehydrogenase, muscle creatine kinase, ornithine transcarbamylase, and antidiuretic hormone may be elevated.

Radiographic Findings

The chest radiograph of a subject with asthma is often normal. Severe asthma is associated with hyperinflation, as indicated by depression of the diaphragm and abnormally lucent lung fields. Complications of severe asthma, including subcutaneous emphysema, pneumomediastinum (E-Fig. 87-2), and pneumothorax, may be detected radiographically. In mild to moderate asthma without adventitious sounds other than wheezing, a chest radiograph need not be obtained; if the asthma is of sufficient severity to merit hospital admission, a chest radiograph is advised.

Electrocardiographic Findings

The electrocardiogram, except for sinus tachycardia, is usually normal in acute asthma. However, right axis deviation, right bundle branch block, "P pulmonale," or even ST-T wave abnormalities may arise during severe asthma and resolve as the attack resolves.

Sputum Findings

The sputum of the asthmatic patient may be either clear or opaque with a green or yellow tinge. The presence of color does not invariably indicate infection, and examination of a Gram-stained and Wright-stained sputum smear is indicated. The sputum often contains eosinophils, Charcot-Leyden crystals (crystallized eosinophil lysophospholipase), Curschmann spirals (bronchiolar casts composed of mucus and cells), or Creola bodies (clusters of airway epithelial cells with identifiable cilia that, in fresh samples, can often be seen to beat), which can affect color without the presence of infection.

DIAGNOSIS

Differential Diagnosis

Asthma is easy to recognize in a young patient without comorbid medical conditions who has exacerbating and remitting airway obstruction accompanied by blood eosinophilia. A rapid response to bronchodilator treatment is usually all that is needed to establish the diagnosis. However, in the patient with cryptic episodic shortness of breath, an elevated Fe_{NO} can help establish a diagnosis of asthma. However, in the absence of an elevated Fe_{NO}, other causes of wheezing (see Table 83-3) should be investigated.

PREVENTION AND TREATMENT

Rx

There is currently no way to prevent a patient from developing an asthmatic diathesis. For example, trials of allergen avoidance in childhood have not been successful. If a patient has such a diathesis with an allergic component, avoidance of allergens can reduce the frequency of asthma attacks. For example, removal of indoor mold can improve symptoms by 25% and reduce medication use by 50%.

The treatment of asthma is directed at two distinct facets of the disease: the control of symptoms and the prevention of exacerbations. Symptomatic control is measured by the severity and frequency of asthma symptoms during the day, including limitations of activities of daily life, the need to use "rescue" β-agonist inhalers, and asthma symptoms that wake the patient from sleep. The prevention of exacerbations is less linked to symptoms than to levels of lung function, so management must include objective measures of lung function. The best measure is FEV_1, but measures of PEFR can be substituted. Inexpensive and easy-to-use peak flowmeters make the measurement feasible in virtually all cases.

Treatment of asthma has two components. The first is the use of acute reliever (rescue) agents (i.e., bronchodilators) for acute asthmatic airway obstruction. The second is the use of controller treatments, which modify the asthmatic airway environment so that acute airway narrowing, requiring rescue treatments, occurs much less frequently.

In a given individual, the intensity of asthma treatment is adjusted, for the most part, to achieve five goals:

1. to allow the patient to pursue the activities of his or her daily life without excessive interference from asthma;
2. to allow the patient to sleep without awakening because of asthmatic symptoms;
3. to minimize the use of rescue bronchodilator treatment;
4. to prevent the need for unscheduled medical care; and
5. to maintain lung function reasonably near normal.

A patient who meets these standards on the basis of a careful history, chest examination, and measurement of lung function is said to be "in control," whereas a patient whose disease activities prevent these goals from being met is said to be "out of control." Patients who are not in control should have their treatment stepped up, whereas patients whose asthma is in good control for 3 months should attempt to have their treatment stepped down (Fig. 87-2).[8]

Rescue Treatments

All patients with asthma should be prescribed a rapid-acting β-agonist rescue inhaler to use if acute asthmatic airway obstruction develops. Patients should be shown how to use the inhaler (Video 87-2) and tested for their ability to use it correctly. All albuterol inhalers now contain hydrofluoroalkane propellants. Aerosol "spacers" can help patients who have difficulty in coordinating their inspiratory effort and inhaler actuation.

β-Adrenergic Agents

Short-acting β-adrenergic agents given by inhalation are the mainstay of bronchodilator treatment of asthma.[9] Constricted airway smooth muscle relaxes in response to stimulation of β_2-adrenergic receptors. β-Adrenergic agonists with varying degrees of β_2-selectivity are available for use in inhaled (by nebulizer or metered-dose inhaler; Fig. 87-3), oral, or parenteral preparations. Most patients with mild intermittent asthma should be treated with a short-acting β_2-selective inhaler (such as albuterol) on an as-needed basis. Regardless of the specific type of medication used, rescue treatment should consist of two "puffs" from the inhaler, with the first and second puffs separated by a 3- to 5-minute interval, which is thought to allow enough time for the first puff to dilate narrowed airways, thus giving the agent better access to affected areas of the lung. Patients should be instructed to exhale to a comfortable volume, to breathe in very slowly (such as they would when sipping hot soup), and to actuate the inhaler as they inspire. Inspiration to near total lung capacity is followed by holding the breath for 5 seconds to allow the deposition of smaller aerosol particles in more peripheral airways. This treatment can be repeated every 4 to 6 hours; patients should be instructed to "advance" their asthma treatment as noted in Figure 87-2 if they need to use more than 12 puffs of a β-agonist in a 72-hour period.

Randomized trials document that the regular use (i.e., two puffs four times a day) of inhaled albuterol is not associated with adverse events. Asthma patients may notice a difference in the inhaled "feel" of albuterol inhalers compared with inhaled glucocorticoid inhalers because all albuterol inhalers are powered by hydrofluoroalkanes and the puff has a lower velocity. Randomized trials show therapeutic equivalence with chlorofluorocarbon-powered inhalers.

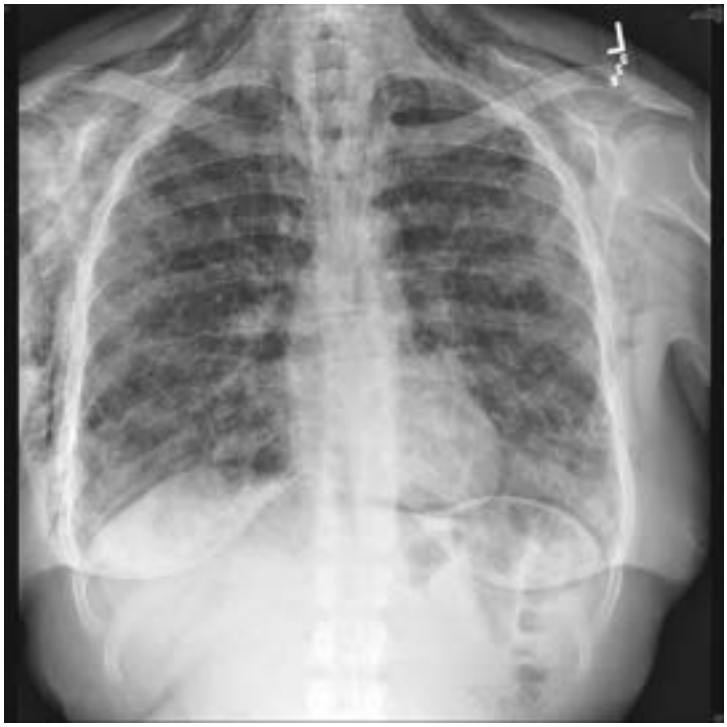

E-FIGURE 87-2. Chest radiograph showing substantial subcutaneous emphysema and pneumomediastinum in a patient with a severe asthma attack. (Radiograph courtesy Christopher Fanta, MD.)

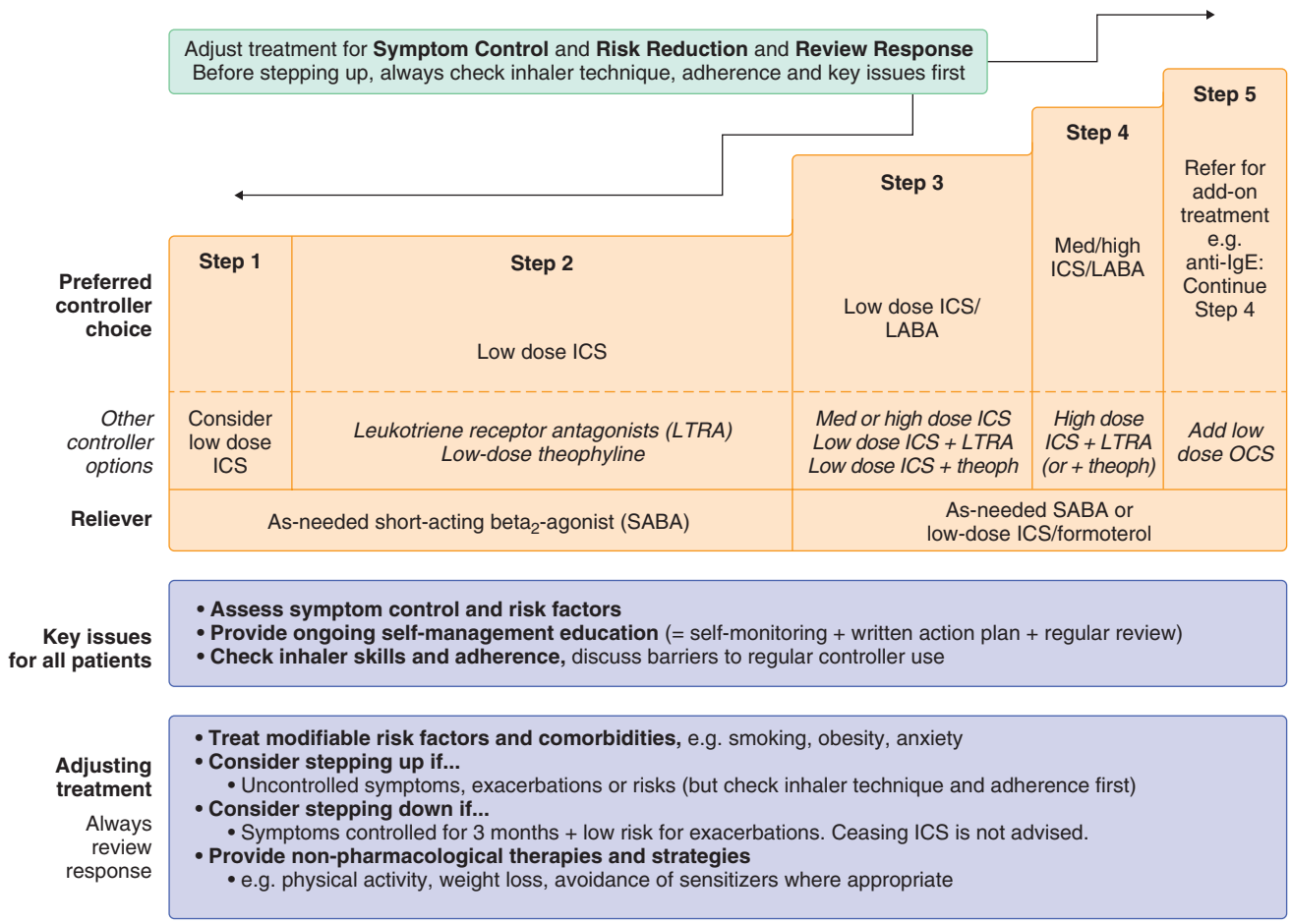

FIGURE 87-2. Asthma treatment algorithm modified from the Global Initiative for Asthma (GINA 2014). There are five steps to the algorithm. First determine your patient's asthma treatment regimen and locate it on the algorithm, that is, step 1 to step 5. Next determine the level of control of your asthma patient. Your patient is in "symptom control" if he or she is able to participate in the activities of daily life without interference from asthma, if there are no nocturnal awakenings from asthma, if there have been no unscheduled visits for asthma care, and if the rescue inhaler is used minimally (twice daily at most). If the patient is in symptom control, the physician must determine whether the asthma is in functional control, that is, if lung function is normal or nearly normal. If the asthma is in symptomatic and functional control, leave the patient at the current step or step down as advised in the figure. If the asthma is not in control, step up from the current level of control. ICS = inhaled glucocorticosteroids; LABA = long-acting β-agonists; OCS = oral corticosteroids; theoph = theophylline.

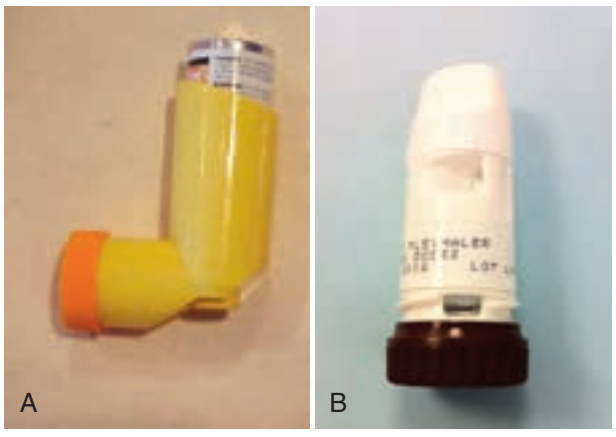

FIGURE 87-3. Commonly used inhalers. A, A pressurized metered-dose inhaler for a branded form of albuterol. Such inhalers propel the medication by means of a pressurized gas; many inhalers use propellants that do not harm the ozone layer. **B,** One of many types of dry powder inhalers; the one shown is a Flexhaler and dispenses budesonide. When this type of inhaler is activated, the active agent is released as a dry powder into a chamber. The patient creates the energy for airflow by means of an inspiratory effort that is directed through the device and that entrains medication into the inhaled airway.

Anticholinergics

Atropinic agents inhibit the effects of acetylcholine released from the intrapulmonary motor nerves that run in the vagus nerve and innervate airway smooth muscle. Ipratropium bromide, the atropinic agent used therapeutically in asthma, is available in a metered-dose inhaler; the recommended dose is two puffs from a metered-dose inhaler every 4 to 6 hours. Although anticholinergic inhalers are useful asthma treatments,[10] none of the inhalers marketed in the United States have a specific label indication for the treatment of asthma. In adults with uncontrolled asthma, inhaled tiotropium (18 µg every morning) added to an inhaled glucocorticoid is more effective than a double dose of glucocorticoids and similar in effect to salmeterol.

Controller Treatments

Inhaled Corticosteroids

Inhaled corticosteroids (see Fig. 87-3), which have less systemic impact than systemic steroids for a given level of therapeutic effect, are effective controller treatments for improving lung function and preventing asthmatic exacerbations in patients with persistent asthma. A *GLCCI1* polymorphism is associated with a reduced response to inhaled glucocorticoids in patients with asthma, but routine screening for this polymorphism is not currently recommended. Patients whose disease can be categorized as "mild persistent asthma" can be treated with an inhaled corticosteroid only when they have increased asthma symptoms rather than requiring an inhaled corticosteroid on a regularly scheduled basis. However, inhaled corticosteroids do not change the natural history of asthma. A wide variety of inhaled corticosteroid products are on the market (Table 87-2). All available products are effective treatments of persistent asthma but differ in terms of cost, the magnitude of adrenal suppression, and the potential for systemic effects, including growth retardation in children, loss of bone mineralization, cataracts, and glaucoma. Overall, no convincing data are available to suggest that there is reason to prefer one corticosteroid to the others. Adverse effects common to all inhaled corticosteroids are oral thrush and hoarseness (attributed to myopathy of the laryngeal muscles); the risk and severity can be reduced by aerosol spacers and good oropharyngeal hygiene (i.e., rinsing out the mouth by gargling after dosing).

Antileukotrienes

Agents with the capacity to inhibit the synthesis of the leukotrienes (zileuton, 600 mg four times a day, or controlled-release [Zyflo CR], 1200 mg twice

TABLE 87-2 ESTIMATED EQUIPOTENT DAILY DOSE FOR INHALED CORTICOSTEROIDS FOR ADULTS

DRUG	LOW DAILY DOSE (μg)*	MEDIAN DAILY DOSE (μg)	HIGH DAILY DOSE (μg)
Beclomethasone dipropionate (QVAR)†	200-500	500-1000	>1000-2000
Budesonide (Pulmicort)†	200-400	400-800	>800-1600
Ciclesonide (Alvesco)†	80-160	160-320	>320-1280
Flunisolide (AeroBid/AeroBid-M)†	500-1000	1000-2000	>2000
Fluticasone (Flovent)†	100-250	250-500	>500-1000
Mometasone furoate (Asmanex)†	200-400	400-800	>800-1200
Triamcinolone acetonide (Azmacort)†	400-1000	1000-2000	>2000

*Once-a-day dosing is acceptable for low daily dose.
†Trade name in the United States.
Note: Some doses may be outside package labeling. Metered-dose inhaler doses are expressed as the amount of drug leaving the valve, not all of which is available to the patient. Dry powder inhaler doses are expressed as the amount of drug in the inhaler after activation.
Modified from 2012 Global Initiative for Asthma guidelines. www.ginasthma.com.

daily) or the action of leukotrienes at the $CysLT_1$ receptor (montelukast [Singulair], 10 mg once a day; pranlukast [Onon, Ultair], 225 mg twice a day, not available in the United States; and zafirlukast [Accolate], 20 mg twice a day) are effective oral controller medications for patients with mild or moderate persistent asthma.[11] In patients treated with zileuton, alanine aminotransferase levels should be monitored for the first 3 to 6 months of treatment; if levels rise to more than three times the upper limit of normal, the drug should be stopped. Theophylline metabolism is slowed by zileuton, so monitoring of levels is indicated if both are prescribed. These treatments can be used on their own for mild persistent asthma or in combination with inhaled steroids for more severe asthma.

Long-Acting β-Agonists

In contrast to short- to medium-acting β-agonists, long-acting β-agonists currently available in the United States (salmeterol [Serevent, 42 μg per puff; the same dose is labeled 50 μg per puff outside of the United States; one or two puffs should be delivered every 12 hours], formoterol [Foradil, 12 μg through a proprietary dry powder inhaler every 12 hours], and bambuterol [Bambec and Oxeol, 10 to 20 mg orally each evening]) have a duration of action of nearly 12 hours and are considered controller agents rather than acute bronchodilator agents. Indacaterol (trade names: Arcapta in the United States and Onbrez in Europe), which is an ultralong-lasting β-agonist delivered by a dry powder inhaler, is used only once a day; its label indications are for chronic obstructive lung disease, not asthma, but it may be used in patients whose asthma is also being treated concomitantly with an inhaled corticosteroid.

Randomized controlled trials demonstrate that long-acting β-agonists should not be used as a sole controller agent. Other trials have shown that there are excess asthma deaths (about one for every 650 patient-years of treatment) when long-acting β-agonists are used. Therefore, long-acting β-agonists should be used in patients with asthma only when they are given in concert with inhaled corticosteroids.

A number of combination products contain both inhaled steroids and long-acting β-agonists in the same aerosol device. These products prevent patients with asthma from using inhaled long-acting β-agonists without inhaled corticosteroids. When prescribing a combination inhaler, the physician should determine the inhaled dose of corticosteroids (fluticasone, budesonide, beclomethasone, mometasone) that the patient requires and then choose a combination product that will deliver a dose of long-acting β-agonist with the inhaled corticosteroid when it is given as two puffs twice per day. The dose of long-acting β-agonist varies with brand and type of inhaler used.

Theophylline

Theophylline and its more water-soluble congener aminophylline are moderately potent bronchodilators that are useful in both inpatient and outpatient management of asthma. Treatment with theophylline is recommended only for patients who have moderate or severe persistent asthma and who are receiving controller medications, such as inhaled steroids or antileukotrienes, but whose asthma is not adequately controlled despite these treatments.

The mechanism by which theophylline exerts its effects has not been established with certainty but is probably related to the inhibition of certain forms of phosphodiesterase. Theophylline is not widely used because of its toxicity and the wide variations in the rate of its metabolism, both in a single individual over time and among individuals in a population. Because blood levels need to be monitored for optimal dosing, most physicians have reserved theophylline for third- or fourth-line therapy. For most preparations, the starting dose should be about 300 mg/day; the frequency will depend on the preparation used.

Acceptable plasma levels for therapeutic effects are between 10 and 20 μg/mL; higher levels are associated with gastrointestinal, cardiac, and central nervous system toxicity, including anxiety, headache, nausea, vomiting, diarrhea, cardiac arrhythmias, and seizures. These last catastrophic complications may occur without antecedent mild side effects when plasma levels exceed 20 μg/mL. Because of these potentially life-threatening complications of treatment, plasma levels need to be measured with great frequency in hospitalized patients receiving intravenous aminophylline and less frequently in stable outpatients receiving one of the long-acting theophylline preparations. Most asthma care providers use dosing amounts and intervals to achieve steady-state theophylline levels of 10 to 14 μg/mL, thereby avoiding the toxicity associated with decrements in metabolism.

Systemic Corticosteroids

Systemic corticosteroids are effective for the treatment of moderate to severe persistent asthma as well as for occasional severe exacerbations of asthma in a patient with otherwise mild asthma, but the mechanism of their therapeutic effect has not been established. No consensus has been reached on the specific type, dose, or duration of corticosteroid to be used in the treatment of asthma. In nonhospitalized patients with asthma refractory to standard therapy, a steroid "pulse" with initial doses of prednisone on the order of 40 to 60 mg/day, tapered to zero during 7 to 14 days, is recommended. For patients who cannot stop taking steroids without having recurrent uncontrolled bronchospasm despite the addition of multiple other controller treatments, alternate-day administration of oral steroids is preferable to daily treatment. For patients whose asthma requires in-hospital treatment but is not considered life-threatening, an initial intravenous bolus of 2 mg/kg of hydrocortisone, followed by continuous infusion of 0.5 mg/kg/hour, has been shown to be beneficial within 12 hours. In attacks of asthma that are considered life-threatening, the use of intravenous methylprednisolone (125 mg every 6 hours) has been advocated. In each case, as the patient improves, oral steroids are substituted for intravenous steroids, and the oral dose is tapered during 1 to 3 weeks; addition of inhaled steroids to the regimen is strongly recommended when oral steroids are started.

Monoclonal Antibody Treatment

Omalizumab

Subcutaneous administration of omalizumab, a humanized murine monoclonal antibody that binds circulating IgE, is associated with decreased serum free (not total) IgE levels. In patients who have moderate to severe allergic asthma with elevated levels of serum IgE and who are receiving inhaled corticosteroids, omalizumab treatment improves asthma control even as doses of inhaled steroids are decreased. Dosing is guided by weight and by pretreatment IgE levels: a monthly subcutaneous dose of 0.016 mg × body weight (kg) × IgE level (IU/mL). For example, in a patient weighing 70 kg with a pretreatment total IgE level of 300 IU/mL, 336 mg of omalizumab would be administered monthly by subcutaneous injection. Dosing calculators can be found online (e.g., http://www.xolairhcp.com/hcp/determining-the-dose.html). Anti-IgE antibodies can reduce exacerbations and improve quality of life in patients with severe allergic asthma, but their place in treatment schema has not been established. Because of the potential for anaphylaxis, all patients need to be monitored after injection; the duration of the monitoring period is not specified by the U.S. Food and Drug Administration (FDA), but most physicians monitor for 30 to 60 minutes.

Other Monoclonal Antibodies

In a randomized trial, lebrikizumab (a monoclonal antibody against interleukin-13 at 250 mg subcutaneously once per month for 6 months) enhanced lung function in adults with asthma, especially in patients with low pretreatment serum periostin levels.[■] Mepolizumab is a monoclonal antibody directed against interleukin-5. In randomized trials using 75 to 100 mg daily, it reduced asthma exacerbations by about 50% among relatively rare patients with moderately severe asthma who still had sputum eosinophilia despite treatment with oral and inhaled corticosteroids.[■■] Among patients with more conventional asthma, however, mepolizumab treatment did not have a salutary effect. A trial of a monoclonal antibody against the α subunit of the shared interleukin-4 and interleukin-13 receptor, dupilumab, in patients whose asthma and eosinophilia was not controlled with conventional doses of inhaled corticosteroids and long-acting β-agonists showed that both the inhaled long-acting β-agonists and inhaled corticosteroids could be withdrawn without losing asthma control when the monoclonal antibody was administered.[■] None of these drugs is currently approved by the U.S. FDA.

Other Controller Drugs

Cromolyn sodium (two to four times a day by nebulizer using 20-mg nebules) is a nonsteroid inhaled treatment used in the management of mild to moderate persistent asthma. It appears to be most useful in pediatric populations or when an identifiable stimulus (such as exercise or allergen exposure) elicits an asthmatic response.

The use of systemic gold (as in rheumatoid arthritis), methotrexate, or cyclosporine has been suggested as adjunctive treatment of patients with severe chronic asthma who cannot otherwise discontinue high-dose corticosteroid treatment. However, these agents are experimental, and their routine use is not advocated. Despite initial encouraging trials, agents that inhibit the action of tumor necrosis factor-α do not benefit patients with asthma and should not be used.

Based on the concern that asthma could be caused by silent gastroesophageal reflux disease, treatment with a proton pump inhibitor has been advocated in patients with mild to moderate asthma even in the absence of gastrointestinal symptoms. Adequately powered clinical studies suggest that this approach provides no benefit for asthma control.

Vaccination for Seasonal Influenza and Pneumococcal Disease

Vaccination of patients for seasonal influenza is safe and not associated with enhanced asthma exacerbations. Vaccination against seasonal influenza and pneumococcal disease is recommended in patients with asthma.

Radio Frequency Ablation of Airway Smooth Muscle

A proprietary system to ablate airway smooth muscle by delivery of radio frequency energy through a bronchoscopically placed probe has reduced asthma exacerbations in sham-controlled trials among patients whose asthma remained out of control despite the use of multiple controller medications. Although a device for such treatment has been approved by the FDA, the long-term impacts of this treatment on airway or lung function are not known.

Control-Driven Asthma Therapy

Because all current asthma treatment is symptomatic (i.e., no current treatment changes the disease history), the approach to the management of asthma is to titrate treatment to achieve an adequate level of control. If a patient's asthma is well controlled, treatment can be continued or stepped down (see Fig. 87-2). If a patient's asthma is poorly controlled, treatment intensity should be stepped up. At the mild end of the spectrum, a patient who has rare limitations in activities of daily life, has nearly normal lung function, and sleeps without interruption from asthma can be prescribed nothing more than inhaled rescue treatment on an as-needed basis. In general, if a patient can control his or her asthma with the use of a single metered-dose inhaler of rescue treatment dispensed every 7 to 8 weeks or less frequently, there is no need for background controller treatment. If a patient has a requirement for more rescue treatment, has symptoms that interfere with sleeping through the night, or has moderately deranged lung function, controller therapy should be added.

Single-agent controller therapy should consist of an inhaled corticosteroid or an antileukotriene. If control is not achieved with one of these agents, the patient can be switched to the other or have a second agent added. The best studied two-agent combination is inhaled corticosteroids and a long-acting inhaled β_2-adrenergic agonist, available in a single inhaler under the trade names of Symbicort, Advair, and Dulera in the United States; trade names vary in other parts of the world. These combinations provide excellent disease control and often allow a reduction in the dose of inhaled corticosteroids. Data indicate that another combination, an antileukotriene and inhaled steroid, is more effective than either treatment alone, but this regimen does not have as substantial an evidence base as the combination of inhaled corticosteroids and a long-acting β-agonist.

Specific Treatment Scenarios

Concurrent Pulmonary Infection

In some patients, acute exacerbations of asthma may be due to concurrent infection, which requires targeted therapy (Chapters 88, 90, and 97).

Aspirin-Exacerbated Respiratory Disease (Previously Termed Aspirin-Induced Asthma)

Approximately 5% of patients with moderate to severe persistent asthma develop asthma when they ingest agents that inhibit cyclooxygenase, such as aspirin and other nonsteroidal anti-inflammatory drugs (Chapter 37). Inhibitors of cyclooxygenase 2 are less likely to cause these reactions, but aspirin-type reactions have been reported in sensitive patients treated with selective cyclooxygenase 2 inhibitors. Although the physiologic manifestations of laboratory-based aspirin challenge can be blocked by leukotriene pathway inhibitors, these agents do not prevent clinical aspirin-exacerbated respiratory disease. Thus, patients with this form of asthma must avoid aspirin and other nonsteroidal anti-inflammatory drugs.

Asthma in the Emergency Department

When a patient with asthma presents for acute emergency care, objective measures of the severity of the attack, including quantification of pulsus paradoxus and measurement of airflow rates (PEFR or FEV_1), should be evaluated in addition to the usual vital signs. If the attack has been prolonged and failed to respond to treatment with bronchodilators (e.g., albuterol by metered-dose inhaler, two puffs every 2 to 3 hours) and high-dose inhaled steroids (e.g., more than 2000 μg/day of beclomethasone or half that amount of fluticasone) before arrival at the emergency department, intravenous steroids (40 to 60 mg of methylprednisolone or its equivalent) should be administered. If the patient has not been receiving treatment with a leukotriene receptor antagonist, such agents should be administered (10 mg of montelukast or 20 mg of zafirlukast) as soon as possible. Treatment with inhaled β-agonists (either nebulized albuterol, 0.5 mL of a 0.5% solution repeated at 20- to 30-minute intervals, or albuterol by metered-dose inhaler, two puffs every 30 minutes) should be used until the PEFR or FEV_1 increases to greater than 40% of the predicted values. If this point is not reached within 2 hours, admission to the hospital for further treatment is strongly advocated.

When patients have PEFR and FEV_1 values that are greater than 60% of their predicted value on arrival in the emergency department, treatment with inhaled β_2-agonists alone, albuterol (0.5 mL of an albuterol 0.083% solution) or equivalent, is likely to result in an objective improvement in airflow rates. If significant improvement takes place in the emergency department, such patients can usually be treated as outpatients with inhaled β_2-agonists and a controller agent (see Fig. 87-2). A good strategy is to add inhaled corticosteroids if the patient has not been receiving this treatment or has been using a single controller therapy.

For patients whose PEFR and FEV_1 values are between 40% and 60% of the values predicted at the time of initial evaluation in the emergency care setting, a plan of treatment varying in intensity between these two plans is indicated. Failure to respond to treatment by objective criteria (PEFR or FEV_1) within 2 hours of arrival at the emergency department is an indication for the use of systemic corticosteroids.

Status Asthmaticus

The asthmatic subject whose PEFR or FEV_1 does not increase to greater than 40% of the predicted value with treatment, whose $PaCO_2$ increases without improvement of indices of airflow obstruction, or who develops major complications such as pneumothorax or pneumomediastinum should be admitted to the hospital for close monitoring. Frequent treatments with inhaled β-agonists (0.5 mL of an albuterol 0.083% solution every 2 hours), intravenous aminophylline (at doses to yield maximal acceptable plasma levels, that is, 15 to 20 μg/mL; 500- to 1000-mg loading dose given during an hour followed by an infusion of 30 to 60 mg/hour), and high-dose intravenous steroids (methylprednisolone, 40 to 60 mg every 4 to 6 hours) are indicated. Oxygen should be administered by face mask or nasal cannula in amounts sufficient to achieve SaO_2 values between 92% and 94%; a higher FIO_2 promotes absorption atelectasis and provides no therapeutic benefit. If objective evidence of an infection is present, appropriate treatment should be given for that infection. If no improvement is seen with treatment and if respiratory failure appears imminent, bronchodilator treatment should be intensified to the maximum tolerated by the patient as indicated by the maximum tolerated heart rate, usually 130 to 140 beats per minute. If indicated, intubation of the trachea and mechanical ventilation can be instituted; in this case, the goal should be to provide a level of ventilation just adequate to sustain life but *not sufficient to normalize arterial blood gases*. For example, a $PaCO_2$ of 60 to 70 mm Hg, or even higher, is acceptable for a patient in status asthmaticus.

Asthma in Pregnancy

Asthma may be exacerbated, remain unchanged, or remit during pregnancy (Chapter 239). There need not be substantial departures from the ordinary management of asthma during pregnancy, although one randomized trial suggests that unlike in other settings, measurement of the fraction of exhaled nitric oxide can improve the management of asthma during pregnancy. However, no unnecessary medications should be administered; systemic steroids should be used sparingly to avert fetal complications, and certain drugs should be avoided, including tetracycline (as a treatment of intercurrent infection), ipratropium bromide (which may cause fetal tachycardia), terbutaline (which is contraindicated during active labor because of its tocolytic effects), and iodine-containing mucolytics (such as saturated solution of potassium iodide). Moreover, use of prostaglandin $F_{2\alpha}$ as an abortifacient should be avoided in asthmatic patients.

PROGNOSIS

Asthma is a chronic relapsing disorder. Most patients have recurrent attacks without a major loss in lung function for many years. A minority of patients experience a significant irreversible loss in lung function over and above the normal pulmonary senescence. Methods to distinguish these various clinical phenotypes have not been developed.

Grade A References

A1. Petsky HL, Cates CJ, Lasserson TJ, et al. A systematic review and meta-analysis: tailoring asthma treatment on eosinophilic markers (exhaled nitric oxide or sputum eosinophils). *Thorax*. 2012;67:199-208.

A2. Peters SP, Kunselman SJ, Icitovic N, et al. Tiotropium bromide step-up therapy for adults with uncontrolled asthma. *N Engl J Med*. 2010;363:1715-1726.

A3. Busse WW, Pedersen S, Pauwels RA, et al. START Investigators Group. The Inhaled Steroid Treatment As Regular Therapy in Early Asthma (START) study 5-year follow-up: effectiveness of early intervention with budesonide in mild persistent asthma. *J Allergy Clin Immunol*. 2008;121:1167-1174.

A4. Calhoun WJ, Ameredes BT, King TS, et al. Asthma Clinical Research Network of the National Heart, Lung, and Blood Institute. Comparison of physician-, biomarker- and symptom-based strategies for adjustment of inhaled corticosteroid therapy in adults with asthma: the BASALT randomized controlled trial. *JAMA*. 2012;308:987-997.

A5. Corren J, Lemanske RF, Hanania NA, et al. Lebrikizumab treatment in adults with asthma. *N Engl J Med*. 2011;365:1088-1098.

A6. Bel EH, Wenzel SE, Thompson PJ, et al. Oral glucocorticoid-sparing effect of mepolizumab in eosinophilic asthma. *N Engl J Med*. 2014;371:1189-1197.

A7. Ortega HG, Liu MC, Pavord ID, et al. Mepolizumab treatment in patients with severe eosinophilic asthma. *N Engl J Med*. 2014;371:1198-1207.

A8. Wenzel S, Ford L, Pearlman D, et al. Dupilumab in persistent asthma with elevated eosinophil levels. *N Engl J Med*. 2013;368:2455-2466.

A9. Peters SP, Anthonisen N, Castro M, et al. American Lung Association Asthma Clinical Research Centers. Randomized comparison of strategies for reducing treatment in mild persistent asthma. *N Engl J Med*. 2007;356:2027-2039.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ege MJ, Mayer M, Normand AC, et al. Exposure to environmental microorganisms and childhood asthma. *N Engl J Med*. 2011;364:701-709.
2. Grainge CL, Lau LC, Ward JA, et al. Effect of bronchoconstriction on airway remodeling in asthma. *N Engl J Med*. 2011;364:2006-2015.
3. Anderson SD, Kippelen P. Assessment and prevention of exercise-induced bronchoconstriction. *Br J Sports Med*. 2012;46:391-396.
4. Kowalski ML, Makowska JS, Blanca M, et al. Hypersensitivity to nonsteroidal anti-inflammatory drugs (NSAIDs)—classification, diagnosis and management: review of the EAACI/ENDA and GA2LEN/HANNA. *Allergy*. 2011;66:818-829.
5. Tarlo SM, Lemiere C. Occupational asthma. *N Engl J Med*. 2014;370:640-649.
6. Bel EH. Clinical practice. Mild asthma. *N Engl J Med*. 2013;369:549-557.
7. Martinez FD, Vercelli D. Asthma. *Lancet*. 2013;382:1360-1372.
8. *Global Initiative for Asthma*. http://www.ginasthma.org. Accessed January 7, 2015.
9. Cazzola M, Page CP, Rogliani P, et al. β_2-Agonist therapy in lung disease. *Am J Respir Crit Care Med*. 2013;187:690-696.
10. Rogers L, Hanania NA. Role of anticholinergics in asthma management: recent evidence and future needs. *Curr Opin Pulm Med*. 2015;21:103-108.
11. Price D, Musgrave SD, Shepstone L, et al. Leukotriene antagonists as first-line or add-on asthma-controller therapy. *N Engl J Med*. 2011;364:1695-1707.

REVIEW QUESTIONS

1. A 25-year-old woman who works in an office complains of shortness of breath during the recovery period from her usual aerobic exercise routine. She had a history of asthma as a child, but it went into remission when she was in junior high school. What single test would be the best to order to easily confirm that she is now having a return of her asthma?

A. Blood eosinophils
B. Measurement of forced vital capacity before and after bronchodilator
C. Measurement of the forced expiratory volume in the first second (FEV_1) before and after bronchodilator
D. Testing for airway hyperresponsiveness with inhalation of cold air
E. Measurement of the diffusion capacity for carbon monoxide

Answer: C If the FEV_1, measured before and after bronchodilator treatment, shows an improvement of 12% or more (assuming an absolute increase in the FEV_1 of at least 200 mL), this finding would be diagnostic of asthma in this setting. The same is not true of the forced vital capacity. Measurement of airway hyperresponsiveness would make the diagnosis but is complicated and time-consuming. Measurements of blood eosinophils or diffusion capacity are not indicated in this setting.

2. For the same patient as in question 1, you take a medical history to determine if her asthma is in control. Which of the following questions would you *not* need to ask?

A. Has she been able to sleep through the night without asthma symptoms?
B. How often has she needed to use her rescue inhaler to control her asthma symptoms?
C. Has she needed to seek unscheduled medical care for her asthma?
D. Can she fulfill the activities of her daily life without asthma symptoms?
E. Is she able to scuba dive without difficulty?

Answer: E Answers A through D are part of the information that should be sought from all asthma patients. Patients with asthma should not scuba dive.

3. A 35-year-old woman with asthma comes for her yearly checkup. For the past year, she has been treated with an albuterol inhaler that she uses on an as-needed basis. During this time, she has used two inhalers (each with 200 puffs of treatment) and no other asthma medications. She is able to sleep through the night without asthmatic symptoms. She is able to participate in all her desired activities, including a regular exercise routine, without asthmatic symptoms, except during the colder months of the year, when she pretreats herself with albuterol before running. She has not had any unscheduled visits for asthma care. Her lung function test results are essentially normal. At this point you should

A. Add an inhaled corticosteroid at low dose to her asthma regimen
B. Add a leukotriene modifier to her asthma regimen
C. Add a combination inhaler (long-acting β-agonist and inhaled corticosteroid)
D. Leave her regimen as it is
E. Remove her albuterol before she becomes addicted to it

Answer: D She is well controlled while she is treated with albuterol alone. There is no need to add or to subtract other therapies.

4. A 25-year-old woman whose asthma is in good control comes to your office indicating that she and her husband are trying to have a child. Her only treatment has been inhaled albuterol on an as-needed basis, and she uses about four inhalers (each contains 200 puffs of medication) a year. What is the single statement most likely to be true about her asthma treatment?

A. There is no need to change her treatment while she is trying to conceive or becomes pregnant.
B. She should stop the albuterol inhaler and simply "suffer through" any asthma events.
C. She should continue the albuterol inhaler and start an inhaled corticosteroid.
D. She should consult an asthma specialist now before she attempts to become pregnant.
E. She should begin to use her albuterol inhaler on a regularly scheduled basis, two puffs four times a day.

Answer: A There is no need to change her asthma therapy at this time because albuterol is safe to use during pregnancy. The other answers are, by elimination, incorrect.

88

CHRONIC OBSTRUCTIVE PULMONARY DISEASE

DENNIS E. NIEWOEHNER

DEFINITIONS

Chronic obstructive pulmonary disease (COPD) is now the preferred term for a condition that is characterized by progressive, largely irreversible airflow obstruction, usually with clinical onset in middle-aged or elderly persons with a history of cigarette smoking, and that cannot be attributed to another specific disease, such as bronchiectasis (Chapter 90) or asthma (Chapter 87). Commonly used terms for this condition in the past included chronic bronchitis and emphysema. That terminology is outdated because nearly all patients with a clinical diagnosis of COPD have both air space destruction (i.e., emphysema) and pathologic changes of the conducting airways consistent with chronic bronchitis.

Emphysema is defined pathologically by abnormal enlargement of the air spaces due to destruction and deformation of alveolar walls. The severity of emphysema may vary widely in COPD patients with similar degrees of airflow obstruction. Chronic bronchitis is defined clinically as persistent cough and sputum production and pathologically as abnormal enlargement of the mucous glands within the central cartilaginous airways. Chronic bronchitis was once thought to be a key element in the pathogenesis of chronic airflow obstruction, but it is now known that increased airflow resistance in COPD can be attributed principally to a variety of pathologic changes within the distal airways of the lung ("small airways disease").

EPIDEMIOLOGY

COPD represents a growing global public health problem, although prevalence estimates vary widely according to the definition used. Cigarette smoking (Chapter 32) is the principal risk factor for COPD, so prevalence tends to reflect societal smoking habits with a lag phase of 20 to 30 years. Cigarette consumption has leveled off or decreased in large segments of North America and Europe, but the prevalence of COPD may continue to increase as exposed populations age. A greater future burden of COPD may be anticipated in Asia and other regions of the world because of rapidly increasing cigarette consumption.

More than 10% of the population older than 45 years in the United States has airflow obstruction of at least moderate severity as judged by spirometric criteria. COPD is the third leading cause of death in the United States, and mortality from COPD has increased during the past 30 years in both men and women.[1] Worldwide, COPD also is the third leading cause of death globally and the fifth leading cause of years lived with disability.[2,3] Medical costs and lost productivity attributable to COPD exceed $40 billion annually in the United States. Direct medical costs rise precipitously as COPD becomes more severe, with hospitalization for exacerbations accounting for more than half of the total.

Cigarette Smoking

Cigarette smoking is the principal cause of COPD, but the relationship is complex and COPD may develop without a smoking history.[4] Airflow obstruction is the sentinel physiologic disturbance in COPD, and the forced expiratory volume in the first second (FEV_1) is the single best indicator of severity. Cigarette smoking causes declines in lung function that exceed those expected from aging alone, and the magnitude of loss is dependent on both the intensity and duration of exposure to cigarette smoke. Thus, the cumulative effects of smoking largely account for the increasing prevalence of COPD with advancing age.

Individual losses of lung function vary widely, even after adjustment for smoking intensity. After the age of 30 years, everyone loses lung function on a yearly basis, but smoking further affects the rate of lung function loss. The mean annual reduction in the FEV_1 (Chapter 85) in normal nonsmoking white men is about 25 mL per year, but the loss increases to an average of about 40 mL per year among smokers (Fig. 88-1). A small minority of smokers, "susceptible smokers," suffer annual FEV_1 losses of 100 mL or more and may develop clinically significant airflow obstruction in the fourth and fifth decades of life. Factors that distinguish the susceptible smoker from the average smoker remain largely unknown.

Adverse effects of cigarette smoke on lung function may extend as far back as fetal development. Maternal smoking during pregnancy, secondhand cigarette smoke exposure during early childhood, and active smoking during adolescence impair lung growth. As a consequence, the lower lung function in early adulthood constitutes a significant risk factor for COPD later in life.

Other Environmental Exposures

Workers exposed to dust in certain workplace environments, such as mines, cotton mills, and grain-handling facilities, commonly develop symptoms of cough and sputum and may suffer permanent loss of lung function (Chapter 93). In some regions of the world, repeated exposure to biomass combustion in confined living quarters causes airflow obstruction. Current urban air

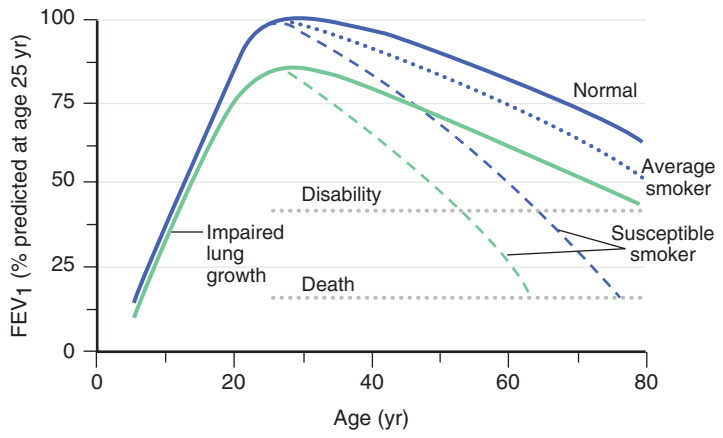

FIGURE 88-1. **Lung growth occurs during childhood and adolescence, with the forced expiratory volume in 1 second (FEV_1) reaching a maximum at about 25 years of age. Thereafter, the FEV_1 steadily declines owing to normal aging effects. Lung function declines more rapidly in smokers, but the average effect is so small that clinically significant airflow obstruction would never develop. However, a proportion of "susceptible smokers" lose lung function much faster than the average, so they develop disabling chronic obstructive pulmonary disorder (COPD). If lung growth is impaired, lung function reserve is less as a young adult, and a susceptible smoker will develop disabling COPD at an earlier age.**

pollution in economically advanced countries appears to have little effect on the prevalence of airflow obstruction, but this factor may be more important in heavily polluted urban centers in industrializing countries.

Respiratory Infections

Recurrent respiratory infections were once thought to be a major factor in the development of airflow obstruction, but longitudinal cohort studies have yielded inconclusive findings. An effect, if present, appears weak relative to cigarette smoking. Whether childhood respiratory infections leave residual effects on adult lung function is similarly unclear.

Airway Responsiveness

Acute bronchoconstriction after inhalation of dilute concentrations of methacholine or histamine, termed bronchial hyperresponsiveness (Chapter 87), is a defining feature of asthma but is also present in many COPD patients. Bronchial hyperresponsiveness independently predicts accelerated loss of lung function in persons with mild to moderate COPD, especially among persons who continue to smoke.

Genetic Factors

A severe deficiency of α_1-antitrypsin is the only genetic risk factor proven to have a major impact on the development of COPD.[5] This deficiency is found in about 1 to 2% of patients with an established diagnosis of COPD. α_1-Antitrypsin, which is a serine protease inhibitor that is secreted into the circulation from the liver, is thought to protect lung tissue against digestion by neutrophil elastase and related serine proteinases that have been implicated in the pathogenesis of human emphysema. The most common allele at the α_1-antitrypsin genetic locus is M, and MM homozygotes have what are considered normal levels of α_1-antitrypsin (100 to 300 mg/dL). Numerous variant alleles have been identified, but severe deficiency is most commonly found in persons who are homozygous for the Z allele, in whom serum levels are generally less than 20 to 30% of the lower range of normal. Affected persons are very susceptible to cigarette smoke–induced damage and may develop severe COPD at a relatively early age. The risk for clinically important emphysema appears to be much less if patients with the risk alleles do not smoke. Emphysema associated with severe α_1-antitrypsin deficiency is characteristically of the panacinar type with a predominant basal distribution.

About 2 to 3% of northern European populations possess the MZ heterozygote serum and have α_1-antitrypsin levels about half of normal. These individuals may be at greater risk for development of chronic airflow obstruction, but the magnitude of the effect, if present, appears to be quite small. Studies of family aggregations and of molecular genetics suggest some heritable risks beyond those associated with α_1-antitrypsin deficiency. Women with severe COPD appear to have relatively more airway disease and less emphysema compared with men with similar airflow obstruction.

PATHOBIOLOGY

Pathology

Emphysema

Emphysema is characterized by abnormal enlargement of the air spaces distal to the terminal bronchiole, with destruction of the alveolar walls but without obvious fibrosis. The terminal bronchiole, which is the most distal nonalveolated airway within the bronchial tree, supplies ventilation to a lung unit that is termed the acinus. Distal to the terminal bronchiole are two or three generations of partially alveolated respiratory bronchioles, and then the alveolar zone, where most gas exchange occurs. Air spaces may enlarge throughout the alveolated zone owing to destruction or rearrangement of their walls.

Human emphysema consists of two major subtypes. Centriacinar emphysema localizes to the respiratory bronchioles just distal to the terminal bronchiole, whereas the remainder of the acinus is largely spared. Individual lesions, which may be up to 10 mm in diameter, tend to be more prominent in the upper lobe. Severe centriacinar emphysema is almost always related to cigarette smoking, but mild centriacinar emphysema can occur from other environmental exposures. Focal areas of inflammation, fibrosis, and carbonaceous pigment are commonly present in adjacent alveolar and bronchiolar walls.

In panacinar emphysema, alveolar ducts are diffusely enlarged; adjacent alveoli may become effaced to the extent that individual units can no longer be identified. With progression of the disease, individual lesions can coalesce to form large bullae. Panacinar emphysema, which is typical of severe α_1-antitrypsin deficiency, also commonly occurs in patients in whom the major risk factor for COPD is cigarette smoking. Most patients with severe COPD appear to have mixed elements of centriacinar and panacinar emphysema, and individual subtypes cannot be reliably distinguished in advanced disease.

Chronic Bronchitis and Bronchiolitis

Mucous glands, located between the epithelial basement membrane and the cartilage plates within the central bronchial tree, and goblet cells in the airway epithelium secrete mucus into the bronchial lumen to aid in host defenses. Enlargement of the bronchial mucous glands and expansion of the epithelial goblet cell population, which occur commonly in COPD, are correlated with clinical symptoms of cough and excess sputum production but not with airflow obstruction. A low-grade inflammatory response, consisting of neutrophils, macrophages, and $CD8^+$ T lymphocytes, may also be seen in the cartilaginous airways of COPD patients.

The principal sites of increased airflow resistance in COPD are the small distal airways that have an internal diameter near the lung's functional residual capacity of less than 2 mm. The earliest pathologic changes identified in young cigarette smokers consist of focal collections of brown-pigmented macrophages in the respiratory bronchiole and a sparse infiltrate of neutrophils and lymphocytes in the walls of the terminal bronchiole. In older patients with established COPD, the inflammatory response is more intense, but still with a similar mix of neutrophils, macrophages, and lymphocytes. Other pathologic changes in the distal airways include fibrosis, goblet cell and squamous cell metaplasia of the lining epithelium, smooth muscle enlargement within the airway walls, and scattered regions of mucous plugging. Compared with normal subjects, distal airways in patients with COPD have thicker airway walls and smaller lumens.

Pulmonary Vasculature

Hypoxemia causes vasoconstriction in small pulmonary arteries and a consequent increase in pulmonary vascular resistance. Vascular remodeling in response to chronic hypoxemia results in irreversible pulmonary hypertension (Chapter 68). Medial smooth muscle enlargement and intimal fibrosis in small pulmonary arteries are the most important vascular changes. In addition, a substantial portion of the capillary bed may be destroyed by severe emphysema.

Pathogenesis

Emphysema appears to be caused by an elastase-antielastase imbalance in the lung due to either elastase excess or antielastase deficiency. Human lungs contain a rich network of elastin-containing fibers and other matrix proteins that confer structural integrity and elasticity to alveolar walls. Intratracheal instillation of proteinases, particularly those capable of hydrolyzing native elastin, induces lesions with morphologic and functional features of human emphysema in experimental animals. Chronic inflammation induced from cigarette smoke increases the burden of inflammatory cell–derived proteinases within lung parenchyma. Severe deficiency of α_1-antitrypsin, a potent inhibitor of neutrophil elastase and other serine proteinases, is associated with development of severe panacinar emphysema in humans. In addition to neutrophil elastase, other neutrophil-derived serine proteinases, such as proteinase 3 and cathepsin G, and matrix metalloproteinases degrade elastin and other matrix components, including collagen, proteoglycans, and fibronectin. A macrophage-derived metalloproteinase, MMP-12, is essential to the development of cigarette smoke–induced emphysema in an animal model, and genetic studies show that a single-nucleotide polymorphism in the promoter region of the MMP-12 gene is associated with reduced risk of COPD in adult smokers.

Relatively less is understood about the pathogenesis of distal airways disease. Particulate matter and toxic gases from inhaled cigarette smoke initiate an inflammatory response composed primarily of macrophages and neutrophils. This early inflammatory response may be mediated by the innate defense system as a response to cell injury. In more advanced disease, inflammation persists even after the patient has stopped smoking. At this stage, humoral and cellular components of the adaptive immune system may predominate, possibly in response to infection or specific antigens from other sources. Infiltration of airway walls with $CD4^+$, $CD8^+$, and B lymphocytes is a prominent feature of more advanced COPD. Repair from either type of immune response might cause airway remodeling by stimulating connective tissue matrix synthesis and smooth muscle formation and by increasing the proportion of mucus-secreting goblet cells within the epithelial layer.

Lung and Heart Mechanics

Elastic recoil refers to the lung's intrinsic tendency to deflate after inflation. A dense labyrinth of elastic fibers and other matrix elements within the lung parenchyma, along with surface tension at the alveolar air-liquid interface, confers this important mechanical property. Elastic recoil maintains the patency of small airways through radial alveolar attachments, similar to the way a tent is held up by its guy ropes, and provides a portion of the driving pressure during expiration. Age-related loss of lung elasticity largely explains the normal decline in FEV_1 with advancing age. In emphysema, loss of lung elastic recoil results from damage to elastic fibers and loss of alveolar surface area, with consequent airflow obstruction.

An increase in bronchial airflow resistance is another sentinel feature of lung mechanics in COPD. The increased resistance in COPD is due primarily to narrowing and loss in airways of less than 2 mm in diameter, known as small airways, even before emphysematous destruction occurs. As a result, peripheral airflow resistance of COPD is higher than in normal lungs by an order of magnitude or more. In contrast, airflow resistance in the central airways of lungs from COPD patients differs little from that of normal lungs. One of the key physiologic aspects of COPD is limitation of expiratory airflow (Fig. 88-2) due to loss of lung elastic recoil and increased viscous resistance to airflow in the small airways (Chapter 85). The severity of emphysema and airflow obstruction is directly related to impaired left ventricular filling, reduced stroke volume, and lower cardiac output without reducing the ejection fraction.

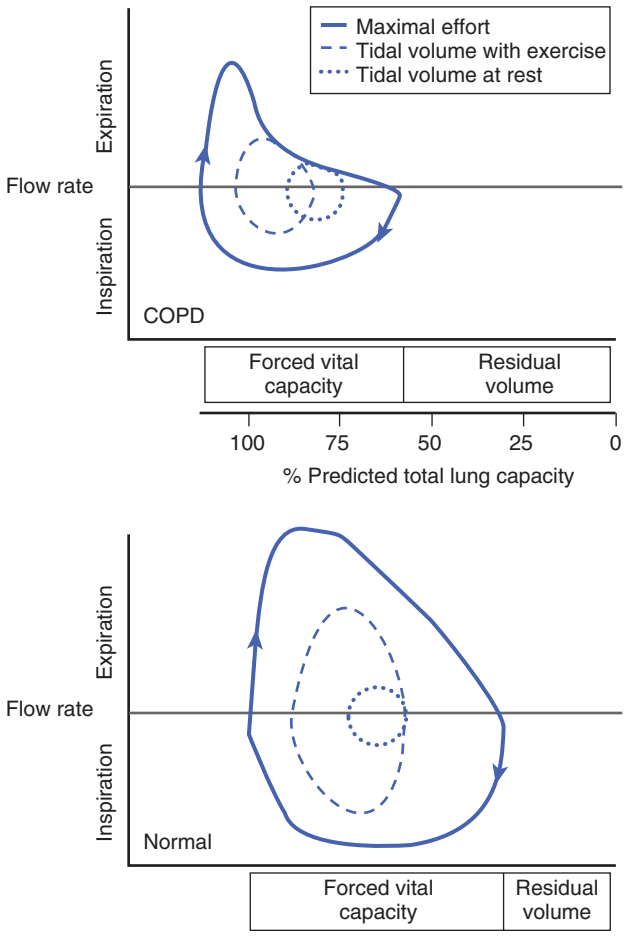

FIGURE 88-2. Inspiratory and expiratory flow-volume loops at rest, with exercise, and with maximal effort in a normal subject are compared with those in a patient with chronic obstructive pulmonary disorder (COPD). The normal subject can easily increase both tidal volume and breathing frequency to match the metabolic requirements of vigorous exercise. In contrast, the COPD patient exhibits maximal expiratory flow limitation even at rest and must breathe at larger lung volumes to optimize expiratory airflow. Lung hyperinflation requires greater respiratory work because the lung and chest wall become stiffer at larger volumes. This effect is accentuated during exercise, which causes end-expiratory lung volume to increase further. This phenomenon is described as *dynamic hyperinflation* and is an important mechanism in limiting exercise and causing dyspnea.

Gas Exchange

Mild hypoxemia may be detected in the early stages of COPD, and hypoxemia often becomes more prominent as airflow obstruction worsens. Hypercapnia usually appears only with severe COPD but is sometimes absent even in late-stage disease. Ventilation-perfusion mismatching, due to changes in both the airways and pulmonary vessels, is largely responsible for hypoxemia, with uneven ventilation being the primary event. Gas exchange is most efficient when the ratio of ventilation to perfusion is uniform in all lung regions. In COPD, there is "wasted ventilation" because some lung regions have inadequate pulmonary blood flow for the ventilation. The calculated A-a gradient for oxygen is larger than anticipated for the patient's age (Chapters 85 and 103). Thus, in most cases of COPD, modest increases in the fraction of inspired oxygen result in a resolution of clinical hypoxemia.

CLINICAL MANIFESTATIONS

History and Physical Examination

COPD should be suspected in all adults who complain of chronic respiratory symptoms, particularly dyspnea (Chapter 83) that limits activities of daily living.[6,7] Clinical features that increase the likelihood of COPD include older age, current or past cigarette use, insidious onset of dyspnea with slow progression, history of acute bronchitis for which medical care is sought, and symptoms of chronic cough, sputum production, or wheezing. Symptoms of cough and sputum may antedate dyspnea by many years. Some patients date the onset of dyspnea to a respiratory infection, but careful questioning usually elicits some history of impaired exercise tolerance before that event. Absence of cigarette smoking does not preclude a diagnosis of COPD because a few persons develop severe irreversible airflow obstruction without smoking history and even without known genetic predispositions. Some nonsmokers may relate a history of occupational dust or noxious gas exposure (Chapters 93 and 94), but in others no putative cause can be discerned.

The physical examination findings are usually normal in patients with mild to moderate disease, and characteristic physical signs may be absent even in severe disease. Physical examination findings commonly present in severe COPD include the appearance of a barrel-shaped chest, low diaphragm detected by percussion, prolonged expiratory phase, and use of accessory muscles of respiration. Heart sounds are usually distant, and auscultation of the chest may reveal diminished breath sounds or a variety of rhonchi, wheezes, and rattles. Auscultatory wheezes may be prominent, particularly during exacerbations, but this physical sign does not reliably differentiate COPD from asthma. With severe hypoxemia, cyanosis may be clinically evident. Clubbing is not associated with COPD, and its presence should suggest another diagnosis. Pedal edema, distended jugular veins, and hepatic congestion are signs of pulmonary hypertension and cor pulmonale (Chapter 68). Patients with advanced COPD may be cachectic, with loss of muscle mass and subcutaneous fat.

Clinical Phenotypes

One of the more enduring efforts to categorize COPD into subtypes is the description of the patients as either "pink puffers" or "blue bloaters."[8] The pink puffer is described as a cachectic individual with unrelenting dyspnea, clinical and radiographic signs of severe lung hyperinflation, and normal or near-normal arterial blood gases at rest. Salient features of the blue bloater are a stout body habitus, chronic cough and sputum, less troubling dyspnea, and severe hypoxemia and hypercapnia resulting in polycythemia and signs of cor pulmonale. In the original description of these phenotypes, the pink puffer phenotype was equated with severe emphysema, whereas the blue bloater was thought to have predominant chronic bronchitis.

Selected COPD patients do fit one or the other of these clinical subtypes, but most cannot be simply categorized. Limited information from clinical and pathologic correlative studies fails to show a consistent association of either clinical subtype with distinguishing pathologic features in lung parenchyma or airways. The blue bloater phenotype may now be less common, possibly because hypoxemia is recognized and treated earlier or because some COPD patients once described as blue bloaters may have had coexisting obstructive sleep apnea (Chapter 100).

COPD patients with similar degrees of airflow obstruction vary greatly with respect to severity of dyspnea, impairment of exercise tolerance, frequency of exacerbations, body habitus, and severity of arterial blood gas disturbances. There is limited understanding about the mechanisms underlying these clinical characteristics.

DIAGNOSIS

Pulmonary Function Tests

Airflow obstruction can be determined best by spirometry. If the ratio of FEV_1 to forced vital capacity (FEV_1/FVC) is less than 0.70 (Chapter 85) after administration of an inhaled bronchodilator, an obstructive defect is present (Table 88-1). A large improvement of perhaps 30 to 40% in FEV_1 after treatment with an inhaled bronchodilator may help identify a patient with predominant asthma, but the test otherwise has little clinical utility and cannot reliably identify patients who will benefit from any particular form of therapy.

Lung volume measurements may help distinguish obstructive and restrictive lung diseases in selected patients, but they are unnecessary in most COPD patients. The diffusing capacity for carbon monoxide (DLCO; Chapter 85) measures the uptake of carbon monoxide between inspired air and the blood stream. Decreases in the DLCO reflect the loss of alveolar surface area that is available for gas transfer and roughly correspond to the severity of emphysema. However, the test provides information of no practical value in the customary management of COPD.

After the diagnosis is established, follow-up spirometry may help determine whether worsening breathlessness is due to COPD, as indicated by a decrease in FEV_1, or to another cause, such as heart failure (Chapter 58). However, repeated spirometry should not be used as a guide to drug therapy because the background variability of the measurement is large relative to treatment effects.

TABLE 88-1 SEVERITY OF AIRFLOW OBSTRUCTION IN COPD ACCORDING TO POSTBRONCHODILATOR SPIROMETRY

STAGE AND SEVERITY	DEFINITION
I: Mild	$FEV_1/FVC < 0.70$, $FEV_1 \geq 80\%$ of predicted
II: Moderate	$FEV_1/FVC < 0.70$, $50\% \leq FEV_1 < 80\%$ of predicted
III: Severe	$FEV_1/FVC < 0.70$, $30\% \leq FEV_1 < 50\%$ of predicted
IV: Very severe	$FEV_1/FVC < 0.70$, $FEV_1 < 30\%$ of predicted or $FEV_1 < 50\%$ of predicted plus chronic respiratory failure

COPD = chronic obstructive pulmonary disease; FEV_1 = forced expiratory volume in the first second; FVC = forced vital capacity.
Data from *Global Initiative for Chronic Obstructive Lung Disease*. http://www.goldcopd.com.

Oximetry

Hypoxemia and hypercapnia become increasingly common as COPD worsens. Because treatment with supplemental oxygen improves mortality, patients with severe COPD should be tested for hypoxemia at regular intervals. Hypoxemia can be detected and quantified by oximetry or arterial blood gases. Oximetry is generally preferred because it is simpler, cheaper, and causes no discomfort. The added information from a set of arterial blood gases (Chapter 103) is most helpful in COPD patients with severe exacerbations.

Radiographic Studies

Common signs of severe COPD on a chest radiograph include hyperinflated lungs, flattened diaphragms, and increased retrosternal clear space (Fig. 88-3). The walls of large emphysematous bullae may be visualized as thin curvilinear lines, and severe emphysema may appear as regions of relative hyperlucency. Chest radiographs are usually normal in mild to moderate COPD and sometimes in severe COPD. Hence, a chest radiograph is not an adequate diagnostic test for COPD, and it is used mostly to exclude other pulmonary diseases. Chest computed tomography (CT), which is a superior imaging modality to assess the magnitude and distribution of emphysema (Fig. 88-4), is not helpful in the usual management of COPD.

Other Studies

Measurement of the serum level of α_1-antitrypsin deficiency may be considered, particularly if the patient has a strong family history of COPD or if the onset of airflow obstruction occurs at an early age. If the α_1-antitrypsin level is less than 20 to 30% of normal, further testing with specialized phenotyping and genotyping studies is required to confirm the diagnosis.

Differential Diagnosis

COPD is most commonly confused with asthma (Chapter 87), particularly in older patients. Clinical features that favor asthma over COPD include onset of disease at an early age, presence of atopy, lack of a smoking history, substantial variability of symptoms over time, and largely reversible airflow obstruction. However, new onset of asthma may occur in elderly people, some asthmatics smoke, an atopic history is not a requisite for development of asthma, and airflow obstruction may become fixed in patients with severe, long-standing asthma. Because treatment is much the same, distinguishing asthma from COPD may not be so important.

Bronchiectasis (Chapter 90) is characterized by chronic inflammation and abnormal dilation of airways associated with chronic cough and expectora-

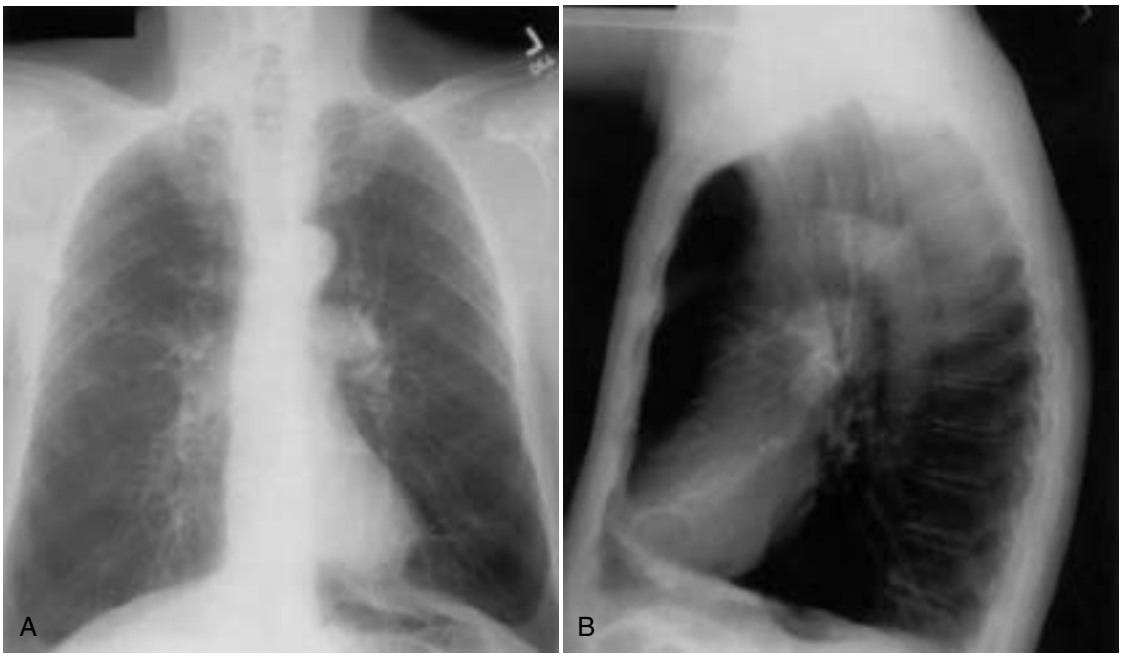

FIGURE 88-3. Posteroanterior (A) and lateral (B) radiographs of the thorax in a patient with emphysema. The most obvious abnormalities are those associated with increased lung volume. The lungs appear dark because of their increased air relative to tissue. The diaphragms are caudal to their normal position and appear flatter than normal. The heart is oriented more vertically than normal because of caudal displacement of the diaphragm, and the transverse diameter of the rib cage is increased; as a result, the width of the heart relative to the rib cage on the posteroanterior view is decreased. The space between the sternum and heart and great vessels is increased on the lateral view.

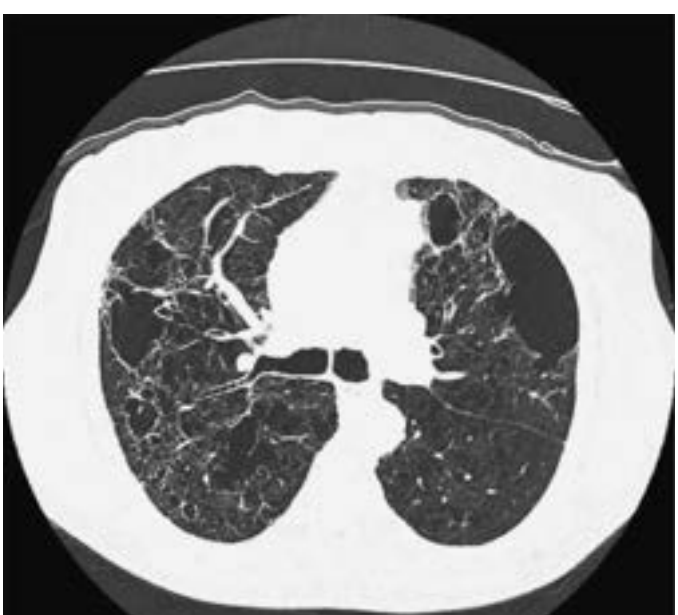

FIGURE 88-4. High-resolution axial computed tomography scan of a 1-mm section of the thorax of a patient with emphysema at the level of the tracheal carina. The right lung is on the left. Multiple large bullae—black holes—are evident. Many smaller areas of similar tissue destruction are also present in both lungs. The right upper lobe bronchus is seen entering the lung; its walls are thickened, suggesting chronic inflammation. (Courtesy Dr. Bruce Maycher.)

tion of purulent sputum. It can be distinguished from COPD with a predominant bronchitis component by chest CT imaging.

Bronchiolitis obliterans is characterized by cicatricial narrowing of the distal airways with severe irreversible airflow obstruction. The condition may occur in association with collagen vascular diseases (Chapters 264 and 266) and is commonly seen after lung transplantation (Chapter 101). A similar disorder has been described with certain industrial inhalants, such as diacetyl, a butter-like flavoring manufactured for use with microwavable popcorn (Chapter 93). In nonsmokers, the diagnosis of bronchiolitis obliterans can be reliably inferred from the history, the presence of irreversible airflow obstruction, and the absence of emphysema or other explanatory conditions on chest CT images. Attribution of cause in smokers is more difficult because the airway disease with diacetyl exposure is similar to that found with cigarette smoke.

TREATMENT

Rx

Stable Disease

Smoking Cessation

Smoking cessation (Chapter 32) reduces symptoms of cough and sputum production in many patients with COPD, but it improves lung function to only a small extent. Most important, about a decade after smoking cessation, the rate of decline of FEV_1 in patients with mild to moderate disease reverts to that seen in lifelong nonsmokers, thereby making it unlikely that these former smokers will ever develop severe COPD. Smoking cessation in patients with mild to moderate COPD also improves long-term mortality by reducing both respiratory and cardiovascular deaths. Smoking cessation probably also slows the decline of lung function in patients with more severe COPD.

Limited information indicates that counseling and pharmacotherapy achieve the same low success rates for smoking cessation in COPD patients as in the general population (Chapter 32). COPD patients tend to quit smoking as the disease progresses, possibly because they have greater awareness of their disease or because cigarette smoke makes their respiratory symptoms worse. Sharing information about abnormal spirometry has not been shown to motivate patients to quit smoking.

Bronchodilators

Both β_2-adrenergic agonists and anticholinergics are widely used to treat COPD (Table 88-2). Short-acting β_2-adrenergic agonists, such as albuterol and the short-acting anticholinergic ipratropium bromide, can be administered either by oral inhaler devices or by nebulization, with little objective superiority of one delivery device over the other if a spacer is used with oral inhaler devices. Longer-acting bronchodilators have largely replaced shorter-acting drugs, but a short-acting bronchodilator, such as albuterol, is still recommended for "rescue" or "as-needed" use in patients who experience bothersome dyspnea (Table 88-3).

Inhaled long-acting β_2-adrenergic agonist bronchodilators widely used for COPD include salmeterol and formoterol, administered by one inhalation twice daily, and indacaterol, administered by one inhalation once daily (Table 88-2). Inhaled anticholinergics include tiotropium (administered by one inhalation once daily) and aclidinium (administered by one inhalation twice daily). Tiotropium (18 µg once daily) appears to be superior to salmeterol (50 µg twice daily) for reducing exacerbations,[1] and combining a long-acting β_2-adrenergic agonist inhaler with a separate long-acting anticholinergic inhaler appears to be more effective than either agent alone.[2] Once-daily combination inhalers likely will soon be commercially available.

Compared with placebo, each class of long-acting bronchodilators reduces exacerbation rates by about 15 to 20% in relative terms. Because the average patient with severe COPD has about one serious exacerbation per year, the number of patients that need to be treated to prevent one exacerbation is about six. Adverse symptomatic events of both classes of long-acting bronchodilators in COPD patients are generally minor.

Theophylline is a poor bronchodilator that largely has been replaced with inhaled drugs, but its effect is additive when it is given along with inhaled bronchodilators. Theophylline may also reduce exacerbations. To be used effectively and safely, it should be started with an oral daily dose of between 150 and 300 mg and titrated to achieve serum levels of 8 to 12 µg/mL. Higher levels are poorly tolerated, especially in older patients. Theophylline interacts with numerous other drugs (e.g., allopurinol, diazepam, cimetidine, ciprofloxacin), and conditions such as heart failure and liver disease may reduce its elimination rates. Patients' drug levels must be monitored on a regular basis, and toxic levels can develop even in a patient receiving a stable dose. Oral roflumilast, a phosphodiesterase 4 inhibitor at 500 µg once daily, can increase FEV_1 by 50 mL and reduce moderate to severe exacerbations in patients with COPD and chronic bronchitis, even in patients already treated with tiotropium.

Corticosteroids

Inhaled corticosteroids produce marginal improvements in lung function and respiratory health status in COPD patients, and they reduce COPD exacerbation rates by about 15 to 20% in relative terms.[3] Inhaled corticosteroids combined with an inhaled long-acting β_2-agonist provide added benefit over that seen with either monotherapy, but added benefit appears quite small when added to both a long-acting β_2-agonist and a long-acting anticholinergic.[4] (Table 88-3). Multiple large trials have found little effect of inhaled corticosteroids in reducing FEV_1 loss during periods of several years.

The most common adverse effects of inhaled corticosteroids are dysphonia and upper airway thrush. Less commonly, they may predispose patients to pneumonia. Observational studies suggest that long-term inhaled corticosteroids may cause osteoporosis (Chapter 243) and cataracts (Chapter 423).

A few COPD patients are prescribed systemic corticosteroids on a regular basis, usually in doses of 10 to 15 mg/day of prednisone or its equivalent. These patients are sometimes considered "prednisone dependent" because it is frequently difficult to wean them completely off drug. There are no proven benefits of chronic, low-dose prednisone in COPD, and adverse effects involving bone, eyes, and other organs are well documented (Chapter 35). Consequently, efforts should be made to reduce or to discontinue chronic systemic corticosteroids while optimizing other treatment.

Oxygen

Chronic hypoxemia in patients with COPD can induce irreversible pulmonary hypertension and cor pulmonale (Chapter 68). Long-term oxygen therapy extends life in patients who are persistently hypoxemic. The principal qualifying criteria are an arterial Pao_2 of less than 56 mm Hg and an arterial oxygen saturation of less than 89%, both while breathing ambient air at rest in a stable clinical state. Patients should also be considered for home oxygen if their Pao_2 is less than 60 mm Hg in the presence of right-sided heart failure or polycythemia. Treatment should consist of home oxygen to be used for at least 18 hours daily, to include sleep time. To determine an appropriate prescription, the oxygen flow rate should be adjusted in 1-L/minute increments at 15-minute intervals until the resting oxygen saturation remains above 90%. In qualifying patients, long-term oxygen may also decrease polycythemia and pulmonary hypertension and improve neuropsychiatric function. Oxygen does not improve mortality in patients with similarly severe airflow obstruction but milder hypoxemia.

Many patients with severe COPD may be normoxemic at rest while breathing ambient air but exhibit oxygen desaturation with exercise. Physicians commonly prescribe ambulatory oxygen in this setting, with the expectation that it will improve exercise tolerance and increase daily activity. Ambulatory oxygen modestly improves exercise endurance for such patients in a laboratory setting, but efforts to show benefit during activities of daily living have

TABLE 88-2 COMMONLY USED MEDICATIONS FOR STABLE CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD)

	MODE OF DELIVERY	DOSE AND FREQUENCY	POSSIBLE ADVERSE REACTIONS
SHORT-ACTING INHALED BRONCHODILATORS			
Albuterol (β_2-adrenergic agonist)	Inhaler	100 µg per inhalation; 1-2 inhalations every 4-6 hours, as needed	Palpitations, tachycardia, tremor, hypersensitivity reaction
	Nebulizer	2.5 mg; every 4-6 hours, as needed	
Ipratropium (anticholinergic)	Inhaler	17 µg per inhalation; 2 inhalations 4 times daily, up to 12 inhalations a day	Dry mouth, cough, blurred vision, hypersensitivity reaction
	Nebulizer	0.5 mg; every 6-8 hours	
Albuterol/ipratropium	Inhaler device	90 µg/18 µg per inhalation; 2 inhalations 4 times daily, up to 12 inhalations per day	All those occurring with either albuterol or ipratropium
	Nebulizer	2.5 mg/0.5 mg; 4 times daily, up to 2 additional doses daily	
LONG-ACTING INHALED BRONCHODILATORS			
Formoterol (β_2-adrenergic agonist)	Inhaler	12 µg; 1 inhalation twice daily	Dizziness, tremor, throat irritation, hypersensitivity reaction
	Nebulizer	20 µg; twice daily	
Salmeterol (β_2-adrenergic agonist)	Inhaler	50 µg; 1 inhalation twice daily	Headache, tremor, throat irritation, hypersensitivity reaction
Indacaterol (β_2-adrenergic agonist)	Inhaler	75 µg; 1 inhalation daily	Cough, oropharyngeal pain, nasopharyngitis, headache, nausea, hypersensitivity reaction
Tiotropium (anticholinergic)	Inhaler	18 µg; 1 inhalation each morning	Dry mouth, urinary retention, symptoms of narrow-angle glaucoma, hypersensitivity reaction
Aclidinium (anticholinergic)	Inhaler	400 µg; 1 inhalation twice daily	Same as for tiotropium
INHALED CORTICOSTEROIDS			
Fluticasone powder	Inhaler	250 µg; 1-2 inhalations twice daily	Sore throat, dysphonia, headache, hypersensitivity reaction
Budesonide	Inhaler	160 µg; 1-2 inhalations twice daily	Nasopharyngitis, thrush, hypersensitivity reactions
COMBINATION INHALERS			
Fluticasone/salmeterol	Inhaler	250 µg/50 µg; 1 inhalation twice daily	All those occurring with either fluticasone or salmeterol
Budesonide/formoterol	Inhaler	160 µg/4.5 µg; 2 inhalations twice daily	All those occurring with either budesonide or formoterol
ORAL DRUGS			
Theophylline (24-hour sustained release)	Pill	200-800 mg, once daily; start with daily dose of 150-300 mg and titrate to blood level of 8-12 µg/mL	Nausea and vomiting, seizures, tremor, insomnia, multifocal atrial tachyarrhythmia, hypersensitivity reaction
Roflumilast	Pill	500 µg, once daily	Depression, suicidal thought, insomnia, loss of appetite, weight loss, diarrhea

TABLE 88-3 GUIDELINE RECOMMENDATIONS FOR DIAGNOSIS AND MANAGEMENT OF STABLE COPD

Spirometry should be obtained to diagnose airflow obstruction in patients with respiratory symptoms. Spirometry should not be used to screen for airflow obstruction in individuals without respiratory symptoms.

For stable COPD patients with respiratory symptoms and FEV_1 between 60% and 80% of predicted, treatment with inhaled bronchodilators may be used.

Stable COPD patients with respiratory symptoms and $FEV_1 < 60\%$ should be treated with inhaled bronchodilators.

Clinicians should prescribe monotherapy with either long-acting inhaled anticholinergics or long-acting inhaled β-agonists for symptomatic patients with COPD and $FEV_1 < 60\%$ predicted. Clinicians should base the choice of specific monotherapy on the patient's preference, the cost, and the adverse effect profile.

Clinicians may administer combination inhaled therapies (long-acting inhaled anticholinergics, long-acting inhaled β-agonists, or inhaled corticosteroids) for symptomatic patients with stable COPD and $FEV_1 < 60\%$ predicted.

Clinicians should prescribe pulmonary rehabilitation for symptomatic patients with an $FEV_1 < 50\%$ predicted. Clinicians may consider pulmonary rehabilitation for symptomatic or exercise-limited patients with an $FEV_1 > 50\%$ predicted.

Clinicians should prescribe continuous oxygen therapy in patients with COPD who have severe resting hypoxemia (arterial oxygen partial pressure ≤ 55 mm Hg or arterial oxygen saturation ≤ 88%)

COPD = chronic obstructive pulmonary disease; FEV_1 = forced expiratory volume in 1 second.
Modified from Qaseem A, Wilt TJ, Weinberger SE, et al. Diagnosis and management of stable chronic obstructive pulmonary disease: a clinical practice guideline update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. *Ann Intern Med.* 2011;155:179-191.

been mostly unsuccessful. Even normoxemic COPD patients with isolated nocturnal hypoxemia have not been shown to benefit from oxygen therapy.

Immunizations and Prophylactic Antibiotics

An annual influenza vaccination (Chapter 18) is recommended for all patients with COPD, although few trials have targeted this population of patients. Observational studies suggest that influenza vaccination substantially reduces hospitalization and mortality rates in COPD patients. Polysaccharide pneumococcal vaccination (Chapter 18) is also recommended, although supporting evidence is weak. Chronic prophylactic macrolide use (e.g., azithromycin 250 mg daily) in addition to regular treatment can decrease exacerbations and improve quality of life in patients with COPD, but such treatment may cause hearing loss and may not be safe if patients have a prolonged QTc interval or if they are taking other drugs known to prolong the QTc interval.

Pulmonary Rehabilitation

COPD patients become increasingly sedentary as their disease progresses. Lack of physical activity causes muscle and cardiovascular deconditioning, which further complicates the ability to perform routine tasks. The principal goal of pulmonary rehabilitation is to reverse this process with a program of exercise endurance training. Educational and behavior modification elements are usually included in an effort to improve coping skills and psychological functioning. Most programs are hospital based and consist of 3- to 4-hour sessions, three times a week, during a 6- to 12-week period. Patients who become breathless with minimal activity or who have exercise-limiting comorbidities are not suitable candidates.

Numerous randomized, controlled trials have shown that pulmonary rehabilitation confers substantial improvements in respiratory health status and in walking distance and possibly in reduction of health care use. Unfortunately, the benefits of pulmonary rehabilitation erode rapidly in the

absence of a continuation plan after completion of the initial program. Pulmonary rehabilitation is also not accessible to most patients for a variety of reasons.

Surgical Options

In lung volume reduction surgery (Chapter 101), severely emphysematous tissue is resected from the upper lobes of both lungs to permit less diseased portions of unresected lung to expand and to function more normally. Patients who are severely disabled from COPD and who have no other major comorbid conditions may be candidates for this procedure if CT imaging shows that severe emphysema is mostly localized to the upper lobes. Compared with controls receiving no surgical treatment, lung volume reduction surgery improves lung function, exercise capacity, and respiratory health status in COPD patients with severe emphysema, but there is no mortality benefit from this procedure, with a possible exception in the subset of patients with both predominant upper lobe emphysema and low exercise capacity. Procedures that deflate severely emphysematous regions of the lung by endoscopic placement of one-way bronchial valves or of transbronchial stents or that completely ablate bronchi that subtend regions of severe emphysema, with either thermal injury or biologic sealants, have not yielded clear clinical benefits and are not recommended (Chapter 101).

Lung transplantation is an option for patients who are severely incapacitated from COPD and have no major comorbid conditions (Chapter 101). Median survival after lung transplantation is only about 5 years, primarily because of the development of bronchiolitis obliterans, a form of chronic graft rejection causing severe airflow obstruction in the peripheral airways. It is unclear whether lung transplantation extends survival in patients with COPD, but patients who are fortunate enough to avoid complications are able to resume normal daily activities.

Exacerbations

Exacerbations represent an important element in the natural history of COPD.[9] An exacerbation is defined as some combination of dyspnea, cough, and productive sputum, each of which has worsened from the stable state or has newly appeared. Exacerbations may also be associated with symptoms of rhinorrhea, sore throat, fever, and chest congestion. A symptom-based clinical event, as described previously, coupled with administration of an antibiotic or a systemic corticosteroid or admission to a hospital, is a definition that has been widely used in clinical trials.

Patients with severe COPD experience an average of about one such exacerbation per year along with additional milder exacerbations that meet the symptomatic definition but do not require a medical intervention. Exacerbations are acute in onset, but recovery may require several weeks. Severe exacerbations have a major adverse impact on health status and may cause permanent loss of lung function. Hospitalization for exacerbations consumes more than half of total medical costs for COPD. For poorly understood reasons, some patients suffer frequent exacerbations, whereas others have very few, despite similar degrees of airflow obstruction. Independent risk factors include low lung function, older age, history of frequent exacerbations, elevated blood levels of inflammatory biomarkers,[10] and prior hospitalizations as well as the presence of a productive cough, gastroesophageal reflux, and cardiovascular comorbidities.

Respiratory infections are thought to cause most exacerbations, although many of these implicated microorganisms may be recovered from sputum during periods of stable disease. Bacteria commonly implicated include *Haemophilus influenzae* (Chapter 300), *Streptococcus pneumoniae* (Chapter 289), and *Moraxella catarrhalis* (Chapter 300). *Pseudomonas aeruginosa* (Chapter 306) and enteric gram-negative bacilli (Chapters 304 and 305) are less common but are seen in patients with very severe COPD who were recently hospitalized or intubated. Putative viral pathogens include rhinoviruses (Chapter 361), influenza (Chapter 364), parainfluenza (Chapter 363), and respiratory syncytial virus (Chapter 362). Periods of increased airborne pollution with diesel particulates, sulfur dioxide, ozone, and nitrogen dioxide are associated with more COPD hospitalizations, but no cause can be assigned to many exacerbations.

Evaluation and management of a patient with a suspected exacerbation vary according to severity. Mild exacerbations encountered in an office setting can be diagnosed and treated on the basis of a brief history and physical examination. Patients seen in emergency department or hospital settings generally are sicker and require a more extensive evaluation (Table 88-4). A chest radiograph should be obtained to look for signs of pneumonia (Chapter 97), pneumothorax (Chapter 99), and heart failure (Chapter 58). If pulmonary embolism (Chapter 98) is suspected, spiral CT of the chest is the test of choice. Arterial blood gases should be measured if there is any suspicion of hypercapnia because this information influences subsequent therapy. Sputum cultures need not be done routinely because they are unproven guides to antibiotic therapy. During seasonal outbreaks of influenza (Chapter 364), type A and B viruses can be identified with rapid commercially available polymerase chain reaction assays having a sensitivity of greater than 90%. These tests should not be relied on to withhold antiviral therapy if the patient is severely ill or if there is a strong clinical suspicion of influenza.

Cardiac disease is a common comorbidity in COPD patients, and distinguishing a COPD exacerbation from left ventricular failure (Chapter 58) by history and physical examination alone is often problematic. Dyspnea (Chapter 83) is common to both conditions. Peripheral edema (Chapter 51) and elevated jugular venous pressure may occur with either left ventricular failure or cor pulmonale secondary to COPD. Echocardiography (Chapter 55) and serum brain natriuretic peptide (BNP) levels (Chapter 58) are useful in this clinical setting, although echocardiography is more difficult to perform in patients

TABLE 88-4 GUIDELINE RECOMMENDATIONS FOR HOSPITAL MANAGEMENT OF COPD EXACERBATIONS

	GLOBAL INITIATIVE FOR CHRONIC OBSTRUCTIVE LUNG DISEASE*	NATIONAL INSTITUTE FOR CLINICAL EXCELLENCE†
Date of statement	2013	2010
Diagnostic testing	Chest radiograph, oximetry, ABGs, and ECG Other testing as warranted by clinical indication	Chest radiograph, ABGs, ECG, complete blood count, sputum smear and culture, blood cultures if febrile
Bronchodilator therapy	Inhaled short-acting β_2-agonist is recommended Consider ipratropium if inadequate clinical response Consider theophylline or aminophylline as second-line intravenous therapy	Administer inhaled drugs by nebulizer or hand-held inhaler with spacer device Specific agents and dosing regimens not specified Consider theophylline if inadequate response to inhaled bronchodilators
Antibiotics (see text for dosing)	Recommended if (1) increases in dyspnea, sputum volume, and sputum purulence all are present; (2) increase in sputum purulence along with increase in either dyspnea or sputum volume; or (3) need for assisted ventilation Initial empirical therapy with aminopenicillin with or without clavulanic acid, macrolide, or tetracycline, based on local bacterial resistance patterns Subsequent therapy based on sputum and blood cultures	Administer only if the patient has a history of purulent sputum Initiate with an aminopenicillin, a macrolide, or a tetracycline, taking into account local bacterial resistance patterns Adjust therapy according to sputum and blood cultures
Systemic corticosteroids	Daily prednisolone 30-40 mg (or its equivalent) for 10-14 days	Daily prednisolone 30 mg (or its equivalent) orally for 7-14 days
Supplemental oxygen	Maintain oxygen saturation 88-92% Monitor ABGs for hypercapnia and acidosis	Maintain oxygen saturation within the individualized target range Monitor ABGs
Assisted ventilation	Indications for NPPV include respiratory acidemia (arterial pH ≤ 7.35) or severe dyspnea with clinical signs of respiratory muscle fatigue or increased work of breathing	NPPV is the treatment of choice for persistent hypercapnic respiratory failure Consider functional status, body mass index, home oxygen, comorbidities, prior ICU admissions, age, and FEV_1 when assessing suitability for intubation and ventilation

*Data from http://www.goldcopd.com.
†Data from http://www.nice.org.uk.
ABGs = arterial blood gases; COPD = chronic obstructive pulmonary disease; ECG = electrocardiogram; ICU = intensive care unit; NPPV = noninvasive positive-pressure ventilation.

with severe COPD. BNP levels may be modestly elevated in both stable and exacerbated COPD in the absence of left ventricular dysfunction. A normal BNP level excludes a diagnosis of left-sided heart failure with a high level of confidence, but an elevated level does not confirm its presence unless it is markedly elevated.

Decisions about the need for hospitalization rely mostly on clinical judgment because there are no well-validated guidelines. Clinical assessment should consider intensity of dyspnea, use of accessory muscles of respiration, arterial blood gas disturbances, hemodynamic stability, and mental alertness.

Guideline recommendations (see Table 88-4) for treatment of patients hospitalized for COPD exacerbations emphasize that antibiotics hasten recovery.[A] Antibiotics are most effective when cough and purulent sputum are present, but there are no well-validated methods for determining which patients should be treated. If patients are sufficiently ill to seek medical attention for an exacerbation, most should probably receive an antibiotic.

Most randomized placebo-controlled trials evaluated first-generation antibiotics, such as amoxicillin, trimethoprim-sulfamethoxazole, and tetracyclines, and it is unclear whether newer classes of antibiotics, such as macrolides and fluoroquinolones, are more effective. Choice of an antibiotic should be made with considerations to cost, safety, and local patterns of antibiotic resistance among the bacterial species commonly isolated from sputa during exacerbations. Doxycycline, 100 mg twice daily for 7 to 10 days, or trimethoprim-sulfamethoxazole, 160/800 mg twice daily for 7 to 10 days, would be reasonable choices for initial therapy in many locales.

Systemic corticosteroids improve lung function, shorten the recovery period, and prevent relapse when given to patients who are hospitalized or present to an emergency department with a COPD exacerbation. Severely symptomatic patients seen in other clinical settings are also likely to benefit. Prednisone, 40 mg once daily for 5 days, is appropriate for most patients.[A] Longer courses of systemic corticosteroid therapy are strongly discouraged because they are no more effective and they increase the likelihood of adverse effects. Parenteral corticosteroids should be given only if gastrointestinal absorption is thought to be impaired. The major adverse effect of systemic corticosteroids is transient hyperglycemia, which may require treatment, particularly in patients with known diabetes mellitus (Chapter 229).

Patients should be encouraged to increase their use of short-acting bronchodilators during outpatient treatment of an exacerbation. For hospitalized patients, a short-acting bronchodilator should be administered on a regular schedule, every 4 to 6 hours and more frequently as needed. Anticholinergic and β_2-agonist agents are similarly effective, and a few small trials found no significant additive effect during exacerbations. Some patients express a preference for a nebulizer delivery system, although equivalent objective results can be achieved when inhalers are used with a spacer.

Sufficient oxygen should be provided to maintain arterial oxygen saturations just above 90%, usually with oxygen flow rates of 2 to 3 L/minute delivered through a nasal cannula. Even at low flow rates, oxygen therapy can be expected to increase $PaCO_2$ by an average of about 5 to 10 mm Hg in patients with chronic hypercapnia. It is prudent to use the lowest flow of oxygen that achieves the desired result. If oxygen is prescribed for hypoxemia during an exacerbation, it is important to retest the patient several weeks later after recovery to determine when long-term oxygen is needed.

The introduction of noninvasive positive-pressure ventilation (NIPPV) has significantly improved the care of patients with severe COPD exacerbations who have respiratory failure.[A] With NIPPV, the patient wears a tightly fitting nasal or full facial mask that is attached to a positive-pressure ventilator, avoiding the need for an endotracheal tube or a tracheostomy (Chapter 105). Compared with usual care, treatment with NIPPV is associated with fewer intubations, a shorter hospital stay, and improved all-cause mortality.

PROGNOSIS

About two-thirds of patients have progressive disease.[11] Severe COPD is associated with excess mortality, and lung function, usually expressed as the percentage of predicted FEV_1, is the single strongest predictor of death. Patients with COPD have variable rates of decline in FEV_1, with more rapid average rates in smokers than in former smokers, but spirometry repeated at intervals of 1 year or more provides only limited information about prognosis. Only about half of patients with an FEV_1 that is about 40% of predicted will survive 5 years.

The severity of emphysema by CT or carbon monoxide diffusion is independently associated with a rapid annual decline in FEV_1. Additional risk factors include the severity of dyspnea, weight loss, limited walking distance, hospitalization for exacerbation, hypoxemia, hypercapnia, and impaired quality of life. The development of bronchiectasis is independently associated with an increased risk of all-cause mortality in patients with moderate to severe COPD.[12] The only interventions shown to reduce mortality are smoking cessation in patients with mild to moderate COPD, long-term oxygen therapy for the subset of patients with chronic hypoxemia, and NIPPV in selected patients who are hospitalized for respiratory failure.

Grade A References

A1. Vogelmeier C, Hederer B, Glaab T, et al. Tiotropium versus salmeterol for the prevention of exacerbations of COPD. *N Engl J Med*. 2011;364:1093-1103.

A2. Wedzicha JA, Decramer M, Ficker JH, et al. Analysis of chronic obstructive pulmonary disease exacerbations with the dual bronchodilator QVA149 compared with glycopyrronium and tiotropium (SPARK): a randomised, double-blind, parallel-group study. *Lancet Respir Med*. 2013;1:199-209.

A3. Yang IA, Clarke MS, Sim EHA, et al. Inhaled corticosteroids for stable chronic obstructive pulmonary disease. *Cochrane Database Syst Rev*. 2012;7:CD002991.

A4. Magnussen H, Disse B, Rodriguez-Roisin R, et al. Withdrawal of inhaled glucocorticoids and exacerbations of COPD. *N Engl J Med*. 2014;371:1285-1294.

A5. Abernethy AP, McDonald CF, Frith PA, et al. Effect of palliative oxygen versus room air in relief of breathlessness in patients with refractory dyspnoea: a double-blind, randomised controlled trial. *Lancet*. 2010;376:784-793.

A6. Herath SC, Poole P. Prophylactic antibiotic therapy in chronic obstructive pulmonary disease. *JAMA*. 2014;311:2225-2226.

A7. COPD Working Group. Pulmonary rehabilitation for patients with chronic pulmonary disease (COPD): an evidence based analysis. *Ont Health Technol Assess Ser*. 2012;12:1-75.

A8. Shah PL, Slebos DJ, Cardoso PF, et al. Bronchoscopic lung-volume reduction with Exhale airway stents for emphysema (EASE trial): randomised, sham-controlled, multicentre trial. *Lancet*. 2011;378:997-1005.

A9. Vollenweider DJ, Jarrett H, Steurer-Stey CA, et al. Antibiotics for exacerbations of chronic obstructive pulmonary disease. *Cochrane Database Syst Rev*. 2012;12:CD010257.

A10. Leuppi JD, Schuetz P, Bingisser R, et al. Short-term vs conventional glucocorticoid therapy in acute exacerbations of chronic obstructive pulmonary disease. The REDUCE randomized clinical trial. *JAMA*. 2013;309:2223-2231.

A11. McCurdy BR. Noninvasive positive pressure ventilation for acute respiratory failure patients with chronic obstructive pulmonary disease (COPD): an evidence-based analysis. *Ont Health Technol Assess Ser*. 2012;12:1-102.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Thun MJ, Carter BD, Feskanich D, et al. 50-year trends in smoking-related mortality in the United States. *N Engl J Med*. 2013;368:351-364.
2. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2095-2128.
3. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2163-2196.
4. Gan WQ, FitzGerald JM, Carlsten C, et al. Associations of ambient air pollution with chronic obstructive pulmonary disease hospitalization and mortality. *Am J Respir Crit Care Med*. 2013;187:721-727.
5. Stoller JK, Aboussouan LS. A review of α_1-antitrypsin deficiency. *Am J Respir Crit Care Med*. 2012;185:246-259.
6. Niewoehner DE. Clinical practice: outpatient management of severe COPD. *N Engl J Med*. 2010;362:1407-1416.
7. Vestbo J, Hurd SS, Agusti AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187:347-365.
8. Kim V, Criner GJ. Chronic bronchitis and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2013;187:228-237.
9. Criner GJ, Bourbeau J, Diekemper RL, et al. Prevention of acute exacerbations of chronic obstructive pulmonary disease: American College of Chest Physicians and Canadian Thoracic Society Guideline. *Chest*. 2014; [Epub ahead of print].
10. Thomsen M, Ingebrigtsen TS, Marott JL, et al. Inflammatory biomarkers and exacerbations in chronic obstructive pulmonary disease. *JAMA*. 2013;309:2353-2363.
11. Vestbo J, Agusti A, Wouters EF, et al. Should we view chronic obstructive pulmonary disease differently after ECLIPSE? A clinical perspective from the study team. *Am J Respir Crit Care Med*. 2014;189:1022-1030.
12. Martinez-Garcia MA, de la Rosa Carrillo D, Soler-Cataluna JJ, et al. Prognostic value of bronchiectasis in patients with moderate-to-severe chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2013;187:823-831.

REVIEW QUESTIONS

1. A 54-year-old man seeks medical attention because of increasing exercise intolerance. He works as a construction worker and finds it increasingly difficult to complete the assigned work. He started smoking at the age of 18 years and currently smokes about one pack per day. In addition to dyspnea, he also has daily cough and sputum and admits to frequent "chest colds." A diagnosis of COPD is strongly suspected. Which of the following would constitute the most appropriate initial assessment?

A. History and physical examination; spirometry; chest computed tomography; oximetry
B. History and physical examination; spirometry; electrocardiogram; oximetry
C. History and physical examination; spirometry; diffusing capacity for carbon monoxide; arterial blood gases
D. History and physical examination; spirometry; chest radiograph; oximetry
E. History and physical examination; lung volumes; chest radiograph; arterial blood gases

Answer: D On the basis of the history, a diagnosis of COPD is likely in this patient. Clinical suspicion must always be confirmed by spirometry to demonstrate airflow obstruction and, if it is present, to obtain information about severity. Measurements of lung volumes or the diffusing capacity for carbon monoxide are more complex than spirometry and provide little additional clinically useful information. Chest computed tomography is superior to a chest radiograph in defining the presence and severity of emphysema, but this information presently has no practical implications for managing most COPD patients. Assessment of need for supplemental oxygen can be done with oximetry or arterial blood gases. For patients with stable disease, oximetry is preferred because it is simple and noninvasive.

2. A 63-year-old woman with severe COPD confirmed by spirometry (FEV_1/FVC, 0.53; FEV_1, 37% of predicted) was admitted to the hospital with an exacerbation. At time of discharge, her arterial oxygen saturation was 81% while breathing ambient air at rest. She was given a prescription for home oxygen. At a follow-up visit 3 months later, the patient stated that she had fully recovered from the exacerbation and was able to resume all of her usual activities. Her arterial oxygen saturation at this time was 86%, again while breathing ambient air at rest. Which of the following is the most appropriate recommendation?

A. Discontinue home oxygen as the patient has fully recovered clinically
B. Continue oxygen long term, titrate oxygen flow rate to ensure an arterial oxygen saturation of 90% or higher, and instruct the patient to use oxygen at least 18 hours daily
C. Continue home oxygen with instructions that it be used only if the patient feels more breathless
D. Use oxygen only during sleep and with exercise
E. Prescribe oxygen at a flow rate of 2 L/min with instructions to use it at least 12 hours daily

Answer: B This patient has severe COPD with persistent hypoxemia after recovery from a severe exacerbation. Long-term oxygen should be prescribed because it improves all-cause mortality in such patients, and 18 hours of daily use is better than 12 hours. Long-term oxygen may confer other benefits, such as improved neurocognitive function and reduced risk of pulmonary hypertension. When long-term oxygen is first prescribed, the flow rate should be titrated to achieve an arterial oxygen saturation of at least 90%. In patients who do not qualify for long-term oxygen but experience modest desaturation with exercise, oxygen therapy has not been shown to decrease breathlessness or to increase daily activity.

3. A 68-year-old man with long-standing, severe COPD presents to an emergency department with a several-day history of increasing breathlessness, coupled with worsening cough and purulent sputum. After a brief evaluation, it is determined that hospitalization is required. The admitting physician obtains a history and physical examination, which reveals a severely breathless patient with audible expiratory wheezing. A chest radiograph shows hyperinflation and a focal pneumonic infiltrate. An electrocardiogram is unremarkable except for a sinus tachycardia. Routine laboratory test results are normal, and arterial blood gas analysis on ambient air shows a P_{O_2} of 48 mm Hg, a P_{CO_2} of 64 mm Hg, and a pH of 7.18. The patient is capable of swallowing oral medications. Which of the following would constitute the most appropriate treatment?

A. Respiratory antibiotic appropriate to the locale; prednisone given orally (40 mg daily for 5 days); short-acting bronchodilator by nebulizer every 4 to 6 hours; supplemental oxygen sufficient to raise saturation to 90%; noninvasive positive-pressure ventilation
B. Respiratory antibiotic appropriate to the locale; prednisone given orally (40 mg daily for 5 days); short-acting bronchodilator by nebulizer every 4 to 6 hours; supplemental oxygen sufficient to raise saturation to 90%; intubation and mechanical ventilation
C. Respiratory antibiotic appropriate to the locale; prednisone given orally (40 mg daily for 5 days); short-acting bronchodilator by nebulizer every 4 to 6 hours; intravenous theophylline; supplemental oxygen sufficient to raise saturation to 90%; noninvasive positive-pressure ventilation
D. Respiratory antibiotic pending results of sputum culture; prednisone given orally (40 mg daily for 5 days); short-acting bronchodilator by nebulizer every 4 to 6 hours; supplemental oxygen sufficient to raise saturation to 90%; noninvasive positive-pressure ventilation
E. Respiratory antibiotic appropriate to the locale; methylprednisolone given intravenously (30 mg four times daily for 10 days); short-acting bronchodilator by nebulizer every 4 to 6 hours; supplemental oxygen sufficient to raise saturation to 90%; intubation and mechanical ventilation

Answer: A Although never tested in rigorous randomized clinical trials, there is a consensus that regularly scheduled short-acting bronchodilators and controlled supplemental oxygen are beneficial for hospitalized patients with COPD. Antibiotics appear to be most beneficial for COPD exacerbations that result in hospitalization. No single antibiotic is proven to be superior to any other, and the choice should be based primarily on local bacterial resistance patterns. There is very good evidence that systemic corticosteroids improve clinical outcomes in this setting as well as good evidence that short courses of oral prednisone are as effective as longer courses. Several small, randomized clinical trials failed to show any benefit from theophylline in hospitalized COPD patients, and gastrointestinal side effects can be troublesome. Strong evidence supports the use of noninvasive positive-pressure ventilation, rather than intubation and mechanical ventilation, for COPD patients with mild to moderate respiratory failure.

4. A 53-year-old woman with severe COPD confirmed by spirometry is seen in the office because of progressive dyspnea with exertion to the point that she could walk only one block on the level before having to stop and catch her breath. She has regularly smoked cigarettes continuously since she was 18 years old and still smokes half a pack per day. She made two visits to emergency departments in the previous year for "bronchitis," and on each visit she had been given an antibiotic but no other respiratory medications. She has no known history of cardiac disease. Examination of the chest reveals signs of lung hyperinflation with decreased breath sounds. The chest radiograph shows signs of hyperinflation. Spirometry results include an FEV_1/FVC of 0.53 and an FEV_1 of 37% of predicted, thus confirming the diagnosis of severe COPD. Oxygen saturation at rest on ambient air is 90%. Which of the following would constitute the most appropriate management plan for this woman with newly diagnosed severe, exacerbation-prone COPD?

A. Smoking cessation; short-acting bronchodilator; long-acting β_2-adrenergic bronchodilator or long-acting anticholinergic bronchodilator
B. Smoking cessation; short-acting bronchodilator; long-acting β_2-adrenergic or long-acting anticholinergic bronchodilator; theophylline; pulmonary rehabilitation
C. Smoking cessation; short-acting bronchodilator; combination inhaled therapy with a long-acting β_2-adrenergic bronchodilator, a long-acting anticholinergic bronchodilator, or a glucocorticosteroid; low-dose prednisone (10-15 mg daily)
D. Smoking cessation; short-acting bronchodilator; combination inhaled therapy with at least two of the following: long-acting β_2-adrenergic bronchodilator, long-acting anticholinergic bronchodilator, inhaled glucocorticosteroid; pulmonary rehabilitation
E. Smoking cessation; short-acting bronchodilator; long-acting β_2-adrenergic or anticholinergic bronchodilator; daily azithromycin; ambulatory oxygen

Answer: D This woman with newly diagnosed COPD has severe airflow obstruction by spirometric criteria, a pronounced limitation of exercise, and a history of two exacerbations in the past year. Smoking cessation is the most important element in her long-term care. The severity of her disease would justify combination therapy including at least two of these three classes of inhaled drugs: long-acting β_2-adrenergic bronchodilators, long-acting anticholinergic bronchodilators, or inhaled glucocorticosteroids. Theophylline and daily azithromycin are not considered first-line treatment of COPD. Chronic prednisone is not known to be clinically effective in patients with stable COPD but is known to cause harm. Assuming that the patient is motivated and that a program is available, pulmonary rehabilitation is recommended for all patients who have exercise-limiting COPD.

5. A 58-year-old woman with severe COPD has become increasingly incapacitated from her lung disease to the point that she is able to walk only about a half-block. She was once very physically active but has had to curtail most of these activities. She is compliant with her medications (including an inhaled short-acting bronchodilator, an inhaled long-acting bronchodilator, an inhaled corticosteroid, and theophylline), but she finds that these medications have had only modest effect on her exercise tolerance. She quit smoking 10 years ago and has no comorbid conditions that should significantly limit her physical activities. Her arterial oxygen saturation is 92% while at rest and breathing ambient air, but it decreases to 86% after a brisk walk down the hallway. Which of the following is the most appropriate intervention at this stage of the disease?

A. Ambulatory oxygen
B. Lung volume reduction surgery
C. Lung transplantation
D. A pulmonary vasodilating drug
E. Pulmonary rehabilitation

Answer: E Assuming that the patient is motivated, has no other severe disabling comorbid conditions, and has access to a program, pulmonary rehabilitation has been shown to increase walking distance and respiratory health status substantially. Most rehabilitation programs are only 6 to 12 weeks in length, and the clinical benefits that accrue during that period erode during the ensuing year unless a longer-term program is put in place. Lung volume reduction surgery or a lung transplant may be considered at a later stage of her disease if all other interventions fail. The best available evidence raises serious doubts that ambulatory oxygen confers significant clinical benefits when it is prescribed for isolated exercise-induced hypoxemia. Studies of limited size have failed to show that pulmonary vasodilators are of any clinical benefit in COPD.

89

CYSTIC FIBROSIS

FRANK J. ACCURSO

DEFINITION

Cystic fibrosis is an autosomal recessive disease caused by mutations in the gene that encodes the cystic fibrosis transmembrane conductance regulator (CFTR) protein, which is a membrane protein that regulates ion flux at epithelial surfaces. Cystic fibrosis affects the lungs, pancreas, intestines, liver, sweat glands, sinuses, and vas deferens, thereby resulting in substantial morbidity and premature mortality. Progressive lung disease is the cause of death in 80% of patients.

EPIDEMIOLOGY

The incidence of cystic fibrosis in the United States, Europe, and Australia is one in 3000 to 5000 births. Cystic fibrosis is most common in the non-Hispanic white population but also occurs in significant numbers in Hispanics (one in 7000), African Americans (one in 12,000), and some Native American populations. It also occurs rarely in individuals of Asian origin.

Approximately 30,000 persons in the United States have cystic fibrosis, for an estimated prevalence of approximately one in 10,000. Worldwide, an estimated 100,000 individuals are affected. Intensive daily care and exacerbations, particularly those that require hospitalization, are associated with enormous social and monetary costs.

PATHOBIOLOGY

Lung and Sinus

The pathobiology of cystic fibrosis is based on the ion transport activities of the CFTR, which is a membrane glycoprotein that functions as a chloride channel but is also involved in the regulation of transepithelial sodium and bicarbonate transport. In the airway, CFTR dysfunction reduces chloride

secretion from the epithelial lining cell into the airway lumen. In addition, sodium absorption from the lumen into the cell is markedly increased. The net effect is a thinning of the airway surface's liquid lining layer, thereby crucially impairing mucociliary clearance. The subsequent chronic infection leads to an intense neutrophil-dominated inflammatory response. Neutrophil products, including proteolytic enzymes and oxidants, are thought to mediate the pathologic changes in the airway, including bronchiectasis, bronchiolectasis, bronchial stenosis, and fibrosis. Mucus plugging of airways, likely owing to chronic infection and inflammation as well as to CFTR dysfunction in mucus glands, is another prominent feature of airway disease (Fig. 89-1).

The origin of sinus disease is believed to be similar to that in the lung. Impaired mucociliary clearance leads to chronic infection and inflammation. Nasal and sinus polyps are common, but their cause is poorly understood.

Pancreas

Pathologic studies of the pancreas in infants demonstrate ductal obstruction and dilation as well as acinar dilation. The CFTR is expressed in ductal tissue, suggesting that impairment of chloride and bicarbonate secretion into the lumen of the ducts leads to the viscous secretions that obstruct the ducts and cause acinar dilation. The exposure of pancreatic tissue to proteolytic enzymes of acinar origin leads to a cystic and fibrotic pancreas in the first few years of life. Unlike the lung, injury to the exocrine pancreas does not involve infection. Almost complete exocrine pancreatic insufficiency is seen in 85% of patients and is related to genotype.

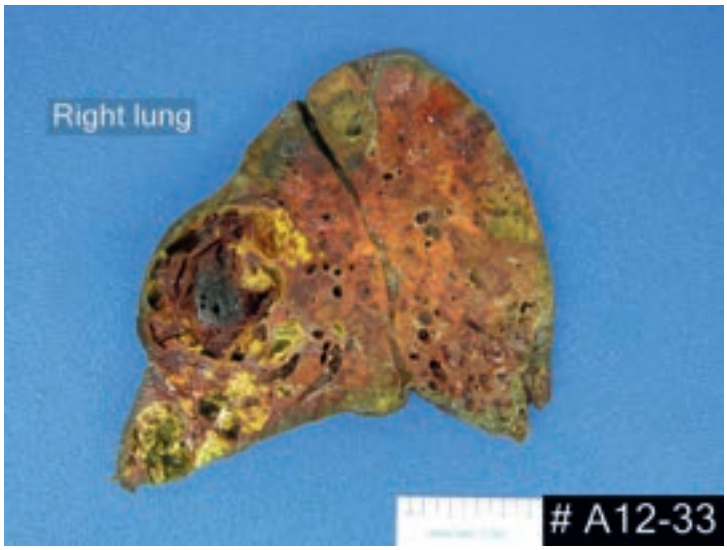

FIGURE 89-1. Section through the right lung from a 13-year-old young woman with cystic fibrosis demonstrating the gross appearance of cavity formation, bronchiectasis, and purulent mucus plugging.

Intestine and Liver

CFTR is expressed throughout the intestine. In approximately 15% of cases, cystic fibrosis is accompanied by meconium ileus as a manifestation of severe intestinal obstruction at birth. The incidence of jejunal and ileal stenoses and atresias is greatly increased compared with normal individuals. It is unclear how these severe abnormalities arise, but mucus obstruction, which is frequently seen in intestinal crypts at birth, suggests that abnormalities in CFTR lead to viscous meconium that interferes with normal intestinal development.

In the liver, bile duct obstruction is the first pathologic change noted. Focal areas of sclerosis ensue, probably owing to obstructed bile ducts. Infection is not involved in hepatic injury.

Sweat Gland

In the sweat gland, CFTR dysfunction leads to a failure of chloride absorption from the lumen into the sweat ductal lining cell. In contrast, the abnormality in the lung involves chloride secretion. The failure to absorb chloride and, by electroneutrality, sodium, results in marked elevations in the chloride and sodium content of sweat. This abnormality is not accompanied by tissue destruction.

Male Reproductive Tract

The vas deferens appears to be the organ that is most sensitive to CFTR dysfunction. It often becomes obstructed in fetuses or infants. Resorption of the vas deferens occurs very early in life, and the vas is ultimately not identifiable in most males.

Other Organ Involvement

The primary abnormalities in cystic fibrosis result in secondary involvement of a number of other systems. Diabetes (Chapter 229), which is increasingly common in adolescents and adults, has historically been attributed to the extension of scarring from the exocrine pancreas into the islets of Langerhans. Recent evidence, however, suggests that functional β-cell abnormalities related to an abnormal CFTR. Osteopenia and osteoporosis (Chapter 243), which are common in adults, result from a combination of malnutrition and chronic infection.[1] Delayed puberty (Chapters 234 and 235) is also common. Patients can experience recurrent vasculitis or arthralgias that are believed to be caused by the host response to chronic infection. Exocrine pancreatic insufficiency leads to impaired growth and to a multitude of potential nutritional complications, including deficiencies in fat-soluble vitamins and trace elements (Chapter 218).[2]

Genetics

The gene that encodes the CFTR spans more than 250,000 base pairs on the long arm of chromosome 7. The CFTR (ABCC7), which is a protein of 1480 amino acids, belongs to the adenosine triphosphate–binding cassette transporter family. More than 1500 mutations of five different classes have been described (Table 89-1).[3] In the United States, only five mutations are present in more than 1% of cases. The F508δ mutation is by far the most

TABLE 89-1 CLASSES OF CFTR MUTATIONS

CLASS	MECHANISM	GENETIC AND MOLECULAR ABNORMALITIES	REPRESENTATIVE GENOTYPE
I	Defective protein production	Unstable mRNA Truncated protein Premature stop mutations Frameshift Splicing variants	 W1282X Del394TT 1717-1G to A
II	Defective protein processing	Trafficking abnormality Protein degraded in proteasome Deletion	 F508del
III	Defective channel regulation	Protein at membrane Failure of gating Amino acid substitution	 G551D
IV	Defective channel conductance	Protein at membrane Decreased gating Amino acid substitution	 R117H
V	Decreased active CFTR	CFTR has normal activity at membrane but is decreased in amount Splice variant Substitution	 3849+10kb C to T A455E

CFTR = cystic fibrosis transmembrane conductance regulator.

common and is present in approximately 90% of patients in the United States. The next most common mutation, G542X, is present in only 5% of patients.

Class 1 mutations are nonsense mutations that result in essentially no expression of the CFTR protein. Class II mutations lead to defective protein processing; in the case of 508δF, protein trafficking to the cell membrane is disrupted because the protein is recognized as defective by cellular quality control mechanisms, which direct it to the proteasome for degradation. In class III mutations, a protein is produced and processed correctly, but the channel remains closed in response to physiologic stimuli. In class IV mutations, the channel is present in the membrane but opens only partially in response to stimuli. In class V mutations, normal CFTR is produced but in reduced amounts because of defective splicing.

Different mutations lead to differing levels of CFTR dysfunction.[4] Whereas severe mutations (classes 1-3) may reduce CFTR activity to 1% to 3% of normal, mild mutations (classes 4 and 5) may be associated with CFTR activity that is 10% to 20% of normal. An important clinical correlation of CFTR activity is in the exocrine pancreas: patients with severe mutations almost always have pancreatic insufficiency, but some patients with milder mutations may retain pancreatic sufficiency. Patients with mild mutations tend, on average, to have less severe lung disease as well.

The clinical course of cystic fibrosis is variable even after controlling for the type of mutation in CFTR, suggesting additional heritable and environmental influences. Genes that code for transforming growth factor-β, mannose-binding lectin, and interferon-related developmental regulator 1 are among the identified modifiers of the severity and course of cystic fibrosis. Most modifiers have to do with the host response to infection or the development of fibrosis rather than the ion transport function of CFTR.

CLINICAL MANIFESTATIONS

Without specific supportive care, most patients succumb in infancy or early childhood because of malnutrition or lung disease. With the use of pancreatic enzyme replacement therapy, better pulmonary care, and the establishment of specialized centers of expertise, most patients live into their fourth or fifth decade.

Lung Disease

Cough, often persistent after viral infections, is the most prominent early feature of the disease. Viral infection may require more frequent hospitalizations in children with cystic fibrosis than in normal children.

Although lung disease begins in infancy, pulmonary function is often preserved until adolescence, when a steep decline frequently begins; at this time, pulmonary exacerbations become common. Most patients with cystic fibrosis have a daily productive cough by late adolescence or young adulthood.

Cystic fibrosis causes obstructive lung disease, initially with decreased flows at low lung volumes. Forced expiratory volume in 1 second (FEV_1) (Chapter 85) is the best correlate of outcome and starts to differ markedly from normal during late adolescence. The rate of decline in FEV_1 often predicts the clinical course.

Early in the disease, the chest radiograph demonstrates hyperinflation and peribronchial thickening. Computed tomography (Fig. 89-2) can demonstrate bronchiectasis (Chapter 90) early in the course of the disease, even before pulmonary function abnormalities are notable.

Airway infection, which is the key clinical manifestation, can be detected by culture of sputum or bronchoalveolar lavage fluid. *Pseudomonas aeruginosa* (Chapter 306) is the primary pathogen, although its prevalence is decreasing in the United States, likely owing to improved treatment. *Staphylococcus aureus* (Chapter 288), which is another prominent pathogen, can be methicillin resistant and exist in a small-colony variant form that makes antibiotic treatment difficult.

Most infections remain endobronchial and rarely cause invasive disease. An exception is *Burkholderia* infection, which can result in sepsis that leads to death. *Burkholderia* infection can also lead to an accelerated decline in lung function and result in death over months to years. Nontuberculous mycobacterial infection can cause granulomatous disease in the airway. *Aspergillus* (Chapter 339) and other fungal species, which are often identified in sputum samples, can cause allergic bronchopulmonary mycoses, but whether they contribute to endobronchitis apart from allergy is unknown.

The polymicrobial nature of airway disease is increasingly appreciated. *Stenotrophomonas maltophilia, Achromobacter xylosoxidans,* and *Inquilinus limosus* are frequently identified serially in airway cultures. Anaerobic infection may also be important.

Individuals with cystic fibrosis are subject to acute exacerbations characterized by cough, dyspnea, decreased exercise tolerance, fatigue, increased sputum production, and change in sputum color that may last days to weeks.[5] Frequently, crackles are increased on physical examination, and both the resting oxygen saturation and lung function may decline. Increasing evidence suggests that the permanent loss of lung function is accelerated during periods of exacerbation.

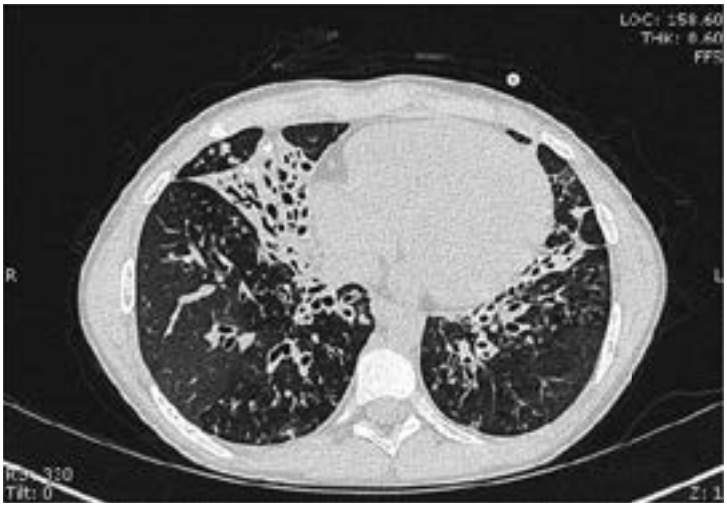

FIGURE 89-2. A computed tomography image of a 13-year-old young woman with cystic fibrosis demonstrating bronchiectasis in several different regions of the lung, right middle lobe collapse, partial lingular collapse, patchy tree-in-bud opacities, and mild hypoattenuation.

Pulmonary complications can also include pneumothorax (Chapter 99), hemoptysis (Chapter 83), and pulmonary hypertension (Chapter 68).[6] Some patients with more advanced disease can develop acute ventilatory failure (Chapter 104) with their exacerbations.

Gastrointestinal Disease

Exocrine pancreatic insufficiency, which is apparent in the first year of life in most patients, results in impaired growth and lifelong difficulty in maintaining a normal weight. Patients at all ages may exhibit signs of malabsorption, including bulky, foul-smelling stools and flatulence. Fat-soluble vitamin and trace element deficiencies are common and are difficult to diagnose without regular laboratory monitoring.

About 15% of patients retain exocrine pancreatic sufficiency, most of whom have mild mutations associated with 10% to 20% of CFTR function. About one sixth of these patients are subject to recurrent episodes of pancreatitis (Chapter 144) that can lead to pancreatic pseudocysts or ultimately result in exocrine pancreatic insufficiency.

Intestinal obstruction can occur at any age. Frequently, the blockage is at the ileocecal valve, but generalized chronic constipation (Chapter 136) is even more common. Intussusception of the appendix can also occur. Inflammatory bowel disease (Chapter 141) and gastrointestinal malignancies (Chapters 192 and 193) appear to be more common than in the general population. Chronic abdominal pain can occur at any time of life and is often difficult to treat.

Most patients who develop liver disease do so in childhood or adolescence. Liver abnormalities are often first appreciated when physical examination reveals splenomegaly or a palpable, firm liver. Occasionally, hematemesis leads to the identification of esophageal or gastric varices that are indicative of portal hypertension. Splenic sequestration can lead to neutropenia or thrombocytopenia. Decreased hepatic production of clotting factors can also contribute to bleeding. Occasionally, jaundice is a presenting sign of hepatobiliary disease. Except for γ-glutamyl transpeptidase (GGT) levels, liver enzymes are frequently normal, even in patients with advanced disease. Gallstones (Chapter 155) are common and may or may not lead to symptoms. The hepatopulmonary syndrome (Chapter 153) can occur.

Other Organ Involvement

Although most patients have radiographic evidence of sinus changes, acute or chronic sinusitis occurs in only a minority of individuals. Sinusitis can be accompanied by debilitating headache and anosmia. Nasal or sinus polyposis can lead to obstructed breathing during sleep.

Hypoelectrolytemia from sweat losses can occur at any age. Symptoms range from nausea, vomiting, and decreased appetite to seizures and

circulatory collapse with fatal consequences. Almost all men are sterile because of the changes in the vas deferens. Spermatogenesis is normal, however.

Cystic fibrosis–related diabetes (Chapter 229) increases in frequency with age.[7] By 30 years of age, approximately one third of patients have diabetes. Although patients rarely develop ketoacidosis, the microvascular and macrovascular complications of diabetes can occur. In addition, patients with diabetes appear to have an accelerated decline in lung function. Osteoporosis (Chapter 243), osteopenia, and increased fractures also increase in frequency with age. Vasculitis accompanied by rash or arthralgia can occur at any time of life. Chronic pain and depression are other important complications that increase with age.

DIAGNOSIS

Newborn Screening and Diagnosis

In the United States, all 50 states require newborn screening for cystic fibrosis to allow early diagnosis and immediate treatment. All newborn screening programs currently measure immunoreactive trypsinogen, a marker of pancreatic injury, from a dried blood spot taken during the first few days of life as the first step in the screening process. This biochemical screen identifies a large number of infants with abnormalities, only a fraction of whom have cystic fibrosis. Most programs perform genetic mutation analysis as the next step. Sweat testing is required to establish the diagnosis if suspected patients carry only one identifiable mutation, but most programs perform confirmatory sweat testing even if two mutations are present.

Sweat testing measures the chloride concentration in sweat that is stimulated by pilocarpine iontophoresis. The result is considered abnormal in adults and children when the concentration of chloride in the sweat is greater than 60 mmol/L; in infants, a concentration greater than 40 mmol/L is considered diagnostic. Patients with milder mutations may have normal sweat chloride values. A family history of cystic fibrosis also provides supportive evidence.

Diagnosis in Adulthood

Five percent of patients are diagnosed after 18 years of age, mostly on the basis of recurrent pancreatitis, nasal polyposis, chronic sinusitis, bronchiectasis, male infertility, allergic bronchopulmonary mycoses, and nontuberculous mycobacterial infection (Table 89-2). If the predominant symptoms are respiratory, the differential diagnosis includes primary ciliary dyskinesia, immune deficiency, or postinfectious bronchiectasis (Chapter 90). If the predominant symptom is recurrent pancreatitis (Chapter 144), the differential diagnosis includes hereditary pancreatitis with abnormalities in the *SPINK* gene. Transepithelial potential differences are altered in cystic fibrosis because of abnormal transport of sodium and chloride. The measurement of nasal potential difference, therefore, can sometimes be used as a diagnostic tool, particularly in adults.

It is increasingly recognized that some patients appear to have cystic fibrosis on clinical grounds but do not meet the criteria for diagnosis, usually because their sweat test results are in the normal range or two genetic mutations cannot be identified. These patients are sometimes diagnosed as having atypical cystic fibrosis, nonclassical cystic fibrosis, or variant cystic fibrosis. Full analysis of the CFTR coding and flanking regions may be helpful in making the diagnosis. Such patients should be followed at a cystic fibrosis center so that their lung disease can be treated and they can be monitored for other complications of cystic fibrosis.

TABLE 89-2 APPROACH TO DIAGNOSIS OF CYSTIC FIBROSIS IN ADULT PATIENTS

CONDITIONS SUGGESTING THE DIAGNOSIS OF CYSTIC FIBROSIS IN ADULTS
Recurrent pancreatitis
Male infertility
Chronic sinusitis
Nasal polyposis
Nontuberculous mycobacterial infection
Allergic bronchopulmonary mycosis
Bronchiectasis
RECOMMENDED DIAGNOSTIC STUDIES
Sweat electrolyte determination
Extended CFTR mutation analysis
Nasal potential difference
High-resolution CT scan to identify bronchiectasis
CT scan of sinuses for polyposis
Sputum induction or bronchoalveolar lavage to identify bacterial and fungal pathogens

CFTR = cystic fibrosis transmembrane conductance regulator; CT = computed tomography.

TREATMENT

The general consensus is that treatment is best conducted at specialized centers that use a team approach. Much of their success is based on the education of patients and families regarding symptoms, complications, the need for daily treatment, the importance of close monitoring of pulmonary function, and the potential benefits of rapid intervention for any detected abnormalities.[8]

Much research has addressed the possibility of treating this genetic disease with gene therapy, but such approaches have not yet yielded positive results. As a result, interest has shifted to improving CFTR function by using druglike molecules.[9,10] In a randomized trial, ivacaftor, a CFTR modulator (150 mg twice daily for 48 weeks), significantly improved lung function, weight, and sweat chloride in cystic fibrosis patients with the G551D mutation. This drug was rapidly approved for clinical use by the U.S. Food and Drug Administration.

Pulmonary Infections

Pulmonary infections can be treated with oral, inhaled, or intravenous antibiotics. An increase in cough or other respiratory symptoms should be addressed with the introduction of antibiotics or a change in antibiotics within a few days. Nebulized antibiotics (4 weeks of either aztreonam 75 mg two or three times a day or tobramycin 300 mg twice daily), alone or in combination with oral antibiotics, improve lung function and decrease exacerbations in patients with chronic *Pseudomonas* infection. These same antibiotic strategies have also been increasingly successful for eradicating *Pseudomonas* infection. Chronic oral macrolide treatment (e.g., azithromycin 5-15 mg/kg/day, 500 mg three times per week) can reduce exacerbations for up to 6 months. It is not yet clear whether the chronic use of antibiotics in this setting leads to the development of more resistant organisms.

More severe changes in symptoms or an acute fall in lung function requires intravenous antibiotics aimed at the cultured pathogen (Chapter 97). Nontuberculous mycobacterial infections are treated for 6 months or longer using multiple antibiotic agents (Chapter 325). Allergic bronchopulmonary mycoses are treated with corticosteroids and antifungal agents (Chapter 331).

Several agents known to decrease the viscosity of mucus have proven to be of clinical benefit in cystic fibrosis. Daily use of inhaled rhDNase (2.5 mg) is associated with improvement in lung function and fewer exacerbations. Inhaled hypertonic (7%) saline can increase pulmonary function and reduce exacerbations, and adding inhaled mannitol (400 mg twice daily) to standard therapy can produce a sustained improvement in pulmonary function for 26 to 52 weeks.

Many patients have hyperreactive airways and may benefit from inhaled bronchodilators (Chapter 87). Inhaled corticosteroids are controversial and do not have proven benefit. Oral corticosteroid "bursts" (e.g., 5 days of prednisone, 1 mg/kg twice a day in children and 60 mg a day in adults) are often useful, but chronic administration of oral corticosteroids can result in severe complications, including diabetes and stunted growth. Most patients perform physical means of airway secretion clearance one or more times a day.

Even passive smoke exposure is deleterious. Oxygen therapy is often required to maintain saturation and prevent the development of pulmonary hypertension. Noninvasive ventilation is used mainly in patients with more advanced disease. Pneumothorax almost always requires pleurodesis. Persistent or recurrent hemoptysis is treated with bronchial artery embolization. Occasionally, lobectomy is required. Patients in acute ventilatory failure should receive mechanical ventilation unless they have decided against such treatment. In patients who have advanced disease, the possible need for ventilation should be addressed before the need actually arises.

Lung transplantation (Chapter 101) is an option for many patients. Individuals with cystic fibrosis have survival rates after transplantation comparable to or better than those of other patients.

Gastrointestinal Diseases

Pancreatic enzyme replacement (Chapter 144) is the mainstay of treatment for exocrine pancreatic insufficiency. Because gastric acid decreases enzyme activity, H_2-blockers (e.g., ranitidine 150 mg twice daily in children weighing >30 kg and in adults) or proton pump inhibitors (e.g., lansoprazole 30 mg orally once daily in children weighing >30 kg and in adults) are often used. Children and adolescents frequently use multiple nutritional supplements every day to maintain weight. Fat-soluble vitamin replacement therapy is necessary in most patients. Between 10% and 20% of patients may require gastrostomy feeding to aid growth or maintain weight.

To prevent intestinal obstruction, dietary fiber should be increased, and polyethylene glycol at varying doses (e.g., 17 g orally with 8 oz of water one to three times per day) is frequently used on a daily basis. Acute obstructions can be treated with more intensive use of polyethylene glycol or Gastrografin

enema. Occasionally, refractory constipation (Chapter 136) requires surgical approaches that can result in loss of intestine.

Other Organ Systems

A combination of nasal rinses and topically applied corticosteroids and antibiotics is used to treat sinus disease (Chapter 426). Surgery is often required, however, especially for polyps.

Many pediatric patients receive daily salt supplementation. Adults should be counseled on the symptoms of salt depletion and encouraged to increase the amount of salt in their diets if there are no medical contraindications to doing so.

Regular screening for the onset of impaired glucose homeostasis or frank diabetes is required in all patients older than 10 years. Diabetes is treated with insulin (Chapter 229) because the safety and efficacy of oral antihyperglycemic agents have not been demonstrated in those with cystic fibrosis. Bone health is addressed through vitamin D supplementation, calcium supplementation, and oral bisphosphonate therapy (Chapter 243). Delayed puberty and short stature require consultation with endocrinologists and sometimes hormonal administration. Most clinicians believe that both aerobic exercise and strength training can have beneficial effects, although the implementation of exercise programs has been difficult in clinical practice. Men with cystic fibrosis can father children through the use of epididymal aspiration to retrieve sperm followed by in vitro fertilization.

General Care

Given all the pulmonary, nutritional, and other therapies prescribed for individuals with cystic fibrosis, their care amounts to several hours a day. The transition from pediatric care to adult care can be challenging and requires diligent planning and execution.[11] This burden has a major influence on the quality of life in patients and their families and may contribute to the increasing incidence of depression observed in this population.

End-of-life care encompasses many complex issues. Patients are often depressed and experience chronic pain. They are asked to perform increasingly intense therapeutic regimens. They may have changed locations to await transplantation. Family, medical, and professional relationships are disrupted. Excellent communication with caregivers about advance directives and other planning is necessary.

PREVENTION

Prenatal carrier screening, which is offered in many countries, can decrease the incidence of cystic fibrosis by approximately 25%. Newborn screening programs may also decrease the incidence by influencing the future reproductive decisions of parents of an affected child.

PROGNOSIS

The median expected survival time for cystic fibrosis patients at birth in the United States is 37 years. However, the peak age at death is 26 years, demonstrating that some patients are particularly vulnerable to devastating lung disease. Late adolescence and early adulthood are high-risk times for pulmonary insufficiency. Patients who survive to their 30s and beyond are often more stable, have milder CFTR mutations, and have a very slow decline in lung function. The success of ivacaftor in treating patients with the G551D mutation has spurred research for small molecules aimed at improving CFTR function for other mutations.

Grade A References

A1. Lee TW, Southern KW. Topical cystic fibrosis transmembrane conductance regulator gene replacement for cystic fibrosis-related lung disease. *Cochrane Database Syst Rev*. 2013;11:CD005599.

A2. Ramsey BW, Davies J, McElvaney NG, et al. A CFTR potentiator in patients with cystic fibrosis and the G551D mutation. *N Engl J Med*. 2011;365:1663-1672.

A3. Ramsey BW, Pepe MS, Quan JM, et al. Intermittent administration of inhaled tobramycin in patients with cystic fibrosis. *N Engl J Med*. 1999;340:23-30.

A4. McCoy KS, Quittner Al, Oermann CM, et al. Inhaled aztreonam lysine for chronic airway *Pseudomonas aeruginosa* in cystic fibrosis. *Am J Respir Crit Care Med*. 2008;178:921-928.

A5. Saiman L, Marshall BC, Mayer-Hamblett N, et al. Azithromycin in patients with cystic fibrosis chronically infected with Pseudomonas aeruginosa: a randomized controlled trial. *JAMA*. 2003; 290:1749-1756.

A6. Elkins MR, Robinson M, Rose BR, et al. A controlled trial of long-term inhaled hypertonic saline in patients with cystic fibrosis. *N Engl J Med*. 2006;354:229-240.

A7. Aitken ML, Bellon G, De Boeck K, et al. Long-term inhaled dry powder mannitol in cystic fibrosis: an international randomized study. *Am J Respir Crit Care Med*. 2012;185:645-652.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Sermet-Gaudelus I, Bianchi ML, Garabedian M, et al. European cystic fibrosis bone mineralisation guidelines. *J Cyst Fibros*. 2011;10(suppl 2):S16-S23.
2. Culhane S, George C, Pearo B, et al. Malnutrition in cystic fibrosis: a review. *Nutr Clin Pract*. 2013;28:676-683.
3. Tsui LC, Dorfman R. The cystic fibrosis gene: a molecular genetic perspective. *Cold Spring Harb Perspect Med*. 2013;3:a009472.
4. Sosnay PR, Siklosi KR, Van Goor F, et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet*. 2013;45:1160-1167.
5. Mogayzel PJ Jr, Naureckas ET, Robinson KA, et al. Cystic fibrosis pulmonary guidelines. Chronic medications for maintenance of lung health. *Am J Respir Crit Care Med*. 2013;187:680-689.
6. Flume PA, Mogayzel PJ Jr, Robinson KA, et al. Cystic fibrosis pulmonary guidelines: pulmonary complications: hemoptysis and pneumothorax. *Am J Respir Crit Care Med*. 2010;182:298-306.
7. Kelly A, Moran A. Update on cystic fibrosis-related diabetes. *J Cyst Fibros*. 2013;12:318-331.
8. Smyth AR, Bell SC, Bojcin S, et al. European Cystic Fibrosis Society Standards of Care: best practice guidelines. *J Cyst Fibros*. 2014;13(suppl 1):S23-S42.
9. Boyle MP, Bell SC, Konstan MW, et al. A CFTR corrector (lumacaftor) and a CFTR potentiator (ivacaftor) for treatment of patients with cystic fibrosis who have a phe508del CFTR mutation: a phase 2 randomised controlled trial. *Lancet Respir Med*. 2014;2:527-538.
10. Van Goor F, Yu H, Burton B, et al. Effect of ivacaftor on CFTR forms with missense mutations associated with defects in protein processing or function. *J Cyst Fibros*. 2014;13:29-36.
11. Nazareth D, Walshaw M. Coming of age in cystic fibrosis—transition from paediatric to adult care. *Clin Med*. 2013;13:482-486.

REVIEW QUESTIONS

1. A 27-year-old male with a nagging, productive cough is known to be infertile. Which diagnostic approach is NOT likely to be helpful in determining whether he has cystic fibrosis?

A. Ultrasonography of the kidneys
B. Sweat test
C. Extended CFTR genotype analysis
D. Sputum culture
E. Exploration of family history looking for deaths in childhood

Answer: A The kidneys are not usually affected in cystic fibrosis, at least to the extent that can be appreciated on ultrasonography. The other tests can all reveal supportive evidence of the diagnosis of cystic fibrosis.

2. Considering the same 27-year-old man as in question 1, the sweat test result comes back abnormal at 78 mmol/L. Which of the following statements regarding this patient's care is INCORRECT?

A. You tell him that it is unlikely that he will need follow-up at a cystic fibrosis center because he was diagnosed at such a late age.
B. You should order a mutation analysis to define his genotype.
C. You should order lung function testing.
D. You should order a high-resolution computed tomography scan of his chest looking for bronchiectasis.
E. You should order a sputum culture.

Answer: A All patients with a diagnosis of cystic fibrosis should have follow-up at a cystic fibrosis care center.

3. Considering the same 27-year-old man as in questions 1 and 2, his lung function testing reveals a forced expiratory volume in 1 second of 65% predicted with no reversibility after a bronchodilator. His computed tomography scan shows definite bronchiectasis. His sputum culture grows *Pseudomonas aeruginosa*. His genotype is G551D/A455E. Which of the following treatments is NOT indicated?

A. Inhaled tobramycin or inhaled aztreonam daily, one month on and one month off
B. Inhaled dornase alfa daily
C. Ivacaftor 150 mg PO twice daily
D. Oral corticosteroid treatment at 30 mg every other day in the morning.
E. Daily airway secretion clearance

Answer: D Corticosteroids are not part of the first-time care of cystic fibrosis, and they also can cause complications, especially by making patients more susceptible to infection.

90

BRONCHIECTASIS, ATELECTASIS, CYSTS, AND LOCALIZED LUNG DISORDERS

ANNE E. O'DONNELL

BRONCHIECTASIS

DEFINITION

Bronchiectasis is an abnormal permanent dilation of the bronchi and bronchioles caused by repeated cycles of airway infection and inflammation. The distal airways become thickened; the mucosal surfaces develop edema, inflammation, and suppuration; an ultimately, there is neovascularization of the adjacent bronchial arterioles. Bronchiectasis, which can be focal or diffuse, is triggered by a variety of genetic, anatomic, and systemic processes. Abnormalities of cilia, mucus clearance, mucus rheology, airway drainage, and host defenses can result in bronchiectasis. Regardless of the cause, patients with bronchiectasis develop chronic infections, which may lead to progressive lung destruction.

EPIDEMIOLOGY

Based on insurance claims reviews, an estimated 110,000 or more patients in the United States are receiving treatment for bronchiectasis that is not related to cystic fibrosis (Chapter 89),[1] and these numbers appear to be increasing.[2] The prevalence in the United States has been reported as 4.2 per 100,000 persons age 18 to 34 years and 272 per 100,000 among those older than 75 years. In the older age category, women are disproportionally represented. Other epidemiologic surveys suggest that there is increased risk for the development of bronchiectasis in individuals with reduced access to health care and higher rates of pulmonary infection in childhood.

PATHOBIOLOGY

In up to one third of cases, the cause of bronchiectasis is not identified. Other cases are related to pulmonary infections, genetic causes, anatomic abnormalities, and immune and autoimmune diseases.[3]

Pulmonary Infections

Approximately one third of patients with bronchiectasis have an infectious trigger, usually years before the onset of the disease. Childhood viral infections, such as pertussis (Chapter 313) and bacterial infection, can cause permanent damage to the airways, leading to bronchiectasis years after the initial infection. Mycobacterial tuberculosis with its resultant granulomatous inflammation of the airway, lung parenchyma, and lymph nodes can cause subsequent bronchiectasis (Chapter 324), and nontuberculous mycobacterial infections have been recognized as an increasing cause and complication of bronchiectasis, particularly in white women older than 55 years (Chapter 325). Nontuberculous mycobacterial-related bronchiectasis typically involves the right middle lobe and lingula and can be associated with the "tree-in-bud" pattern of bronchiolar infection as well.

Genetics

Cystic fibrosis (Chapter 89) is characterized by bilateral diffuse bronchiectasis. Although many cystic fibrosis patients are diagnosed in childhood with multisystem disease, older patients may present with only pulmonary or pulmonary and sinus manifestations. Some patients with bronchiectasis may have subtle defects in the cystic fibrosis transmembrane conductance regulator channel without a clear-cut diagnosis of cystic fibrosis.[4]

In primary ciliary dyskinesia, abnormalities in the dynein arms prevent normal ciliary beating. Patients with primary ciliary dyskinesia generally have significant sinopulmonary disease and infertility, and approximately half of these patients have Kartagener syndrome with situs inversus (Chapter 69). Patients with α_1-antitrypsin deficiency also may develop bronchiectasis.

Anatomic Causes

Patients with chronic abnormalities of their swallowing mechanism or with esophageal dysfunction may develop focal or diffuse bronchiectasis with lower lobe predominance (Chapter 138). Direct lung injury caused by acid

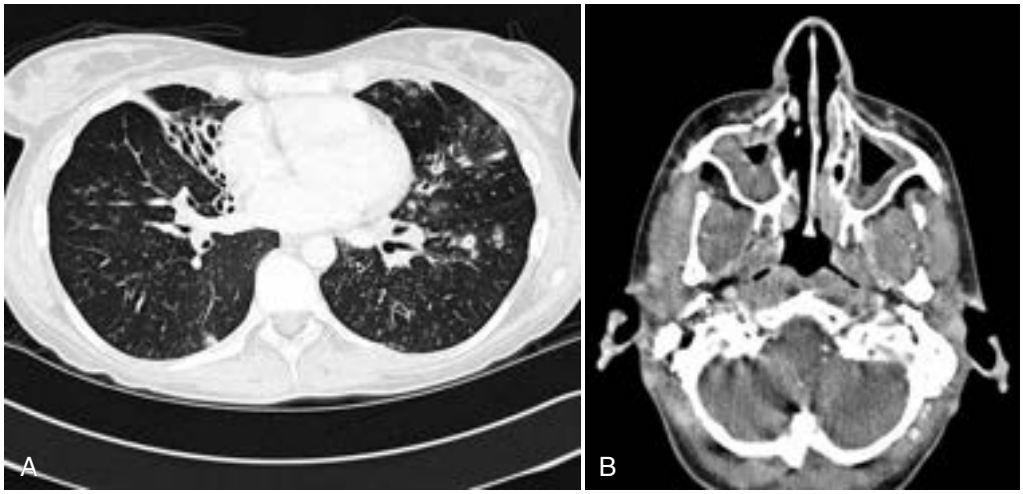

FIGURE 90-1. A and B, High-resolution computed tomographic images of bilateral bronchiectasis in a patient with primary ciliary dyskinesia.

or particulate matter aspiration or recurrent pneumonia may lead to bronchiectasis.

Chronic obstructive pulmonary disease (COPD) is sometimes complicated by bronchiectasis (Chapter 88). Patients with chronic lower airway bacterial colonization and increased airway inflammation may develop areas of bronchiectasis. Rarely, patients with asthma (Chapter 87) have been found to have bronchiectasis. Allergic bronchopulmonary aspergillosis (Chapter 339) can cause a distinct "finger-in-glove" central bronchiectasis owing to chronic inflammation and mucous plugging. Airway abnormalities such as endobronchial tumors (Chapter 191), extrinsic compression by lymph nodes (right middle lobe syndrome), and foreign bodies are also rare causes of focal bronchiectasis. Tracheobronchomegaly (Mounier-Kuhn syndrome) is associated with distal bronchiectasis.

Immune and Autoimmune Diseases

Primary hypogammaglobulinemia (Chapter 250) leads to recurrent pulmonary infections that may result in bronchiectasis. Patients with immunoglobulin G subclass deficiencies may develop bronchiectasis if the deficiency leads to reduction in antibody production. Defects of neutrophil adhesion and chemotaxis (Chapter 169) have been found to cause bronchiectasis. Patients with human immunodeficiency virus infection (Chapter 391) have a higher prevalence of bronchiectasis than individuals with a normally functioning immune system.

Bronchiectasis is an increasingly recognized complication of collagen vascular diseases, particularly rheumatoid arthritis (Chapter 264) and Sjögren syndrome (Chapter 268). The airway injury is likely attributable to chronic inflammation or esophageal dysfunction. Inflammatory bowel disease (Chapter 141) also causes bronchiectasis by undetermined mechanisms.

CLINICAL MANIFESTATIONS

Patients present with chronic cough and usually have mucopurulent or purulent sputum production. Occasionally, a dry nonproductive cough is the primary manifestation. Other symptoms include dyspnea, intermittent hemoptysis, and pleuritic chest pain. Weight loss, malaise, and fatigue sometimes develop. When patients have infectious exacerbations, they may develop fever as well as an increase in their baseline symptoms. Physical findings in patients with bronchiectasis are nonspecific and include an abnormal chest examination with wheezing, crackles, or both. Clubbing of the digits is rare.

The clinical course of patients with bronchiectasis is variable. Some patients have few to no symptoms, others have daily cough with sputum production, and some patients have occasional to frequent exacerbations. A slow decline in pulmonary function is seen with bronchiectasis; the decline is more rapid in patients infected with *Pseudomonas aeruginosa* (Chapter 306) and in patients who have more frequent exacerbations.

DIAGNOSIS

Imaging Studies

Although the diagnosis may be suspected by plain chest radiography, high-resolution computed tomography (HRCT) is the current "gold standard" for confirming bronchiectasis. The characteristic computed tomography (CT) findings are lack of bronchial tapering, bronchi visible in the peripheral 1 cm of the lungs, and an internal bronchial diameter greater than the diameter of the accompanying bronchial artery. Other associated HRCT findings are cysts off the end of a bronchus, tree-in-bud irregular branching lines (E-Fig. 90-E1) indicating mucus impaction (E-Figs. 90-E2 and 90-E3), volume loss (E-Fig. 90-E4), and occasionally associated consolidation (Fig. 90-1). The location of the bronchiectatic airways may suggest the cause: upper lobe predominance is seen in cystic fibrosis and lower lobe predominance in aspiration syndromes (E-Fig. 90-E5). Whereas right middle lobe and lingula involvement suggests the presence of nontuberculous mycobacterial infection (Fig. 90-2 and E-Figs. 90-E6A and 90-E6B), central bronchiectasis is seen with allergic bronchopulmonary aspergillosis (Fig. 90-3).

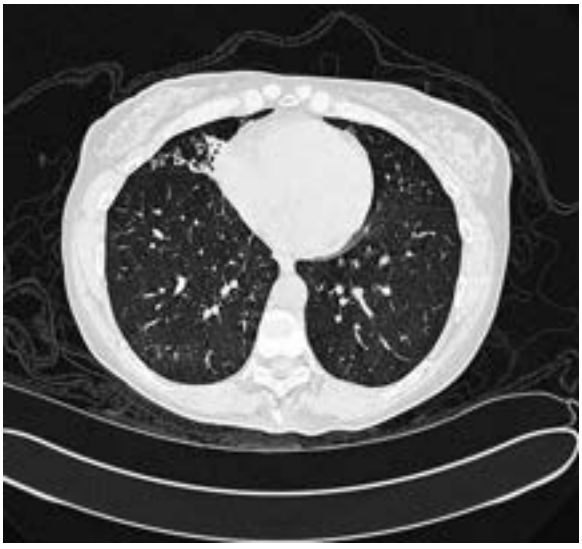

FIGURE 90-2. High-resolution computed tomography image of nodular bronchiectasis caused by a nontuberculous mycobacterium infection.

Pulmonary function testing, which should be performed on all patients with suspected bronchiectasis, usually shows airflow obstruction as measured by the ratio between the forced expiratory volume in 1 second (FEV_1) and forced vital capacity (FVC) (Chapter 85). The severity of the airflow obstruction and the rate of decline correlate with radiographic extent of disease and frequency of exacerbation. Bronchoscopy will detect airway abnormalities, including tumors, structural deformities, and foreign bodies, and hence should be considered in the evaluation of localized bronchiectasis.

Cultures of sputum and of bronchoalveolar lavage when expectorated sputum is not available have an important role in assessing the infectious complications of bronchiectasis. Molecular techniques have recently demonstrated diverse polymicrobial communities in the lungs of patients with bronchiectasis, both when they are clinically stable and also during exacerbations.[5]

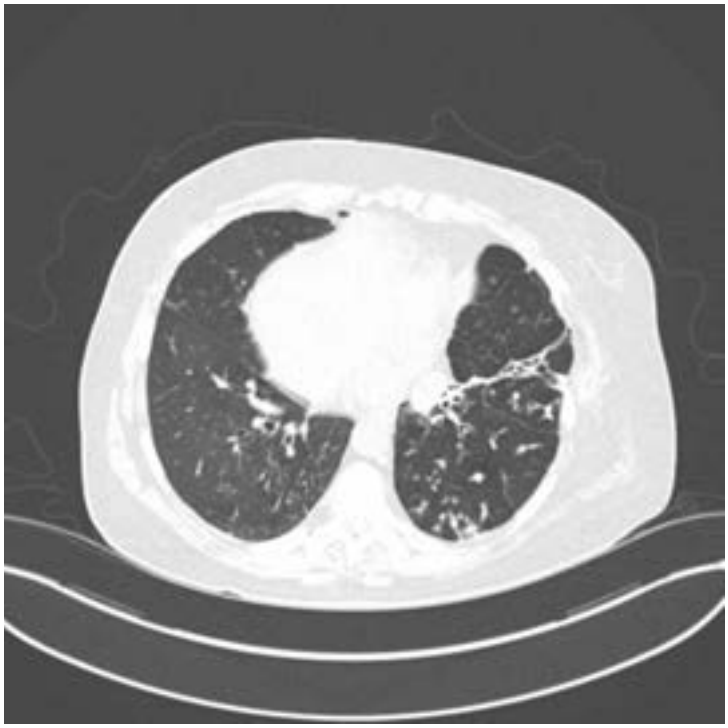

E-FIGURE 90-1. A thin-slice high-resolution computed tomography image showing bronchiectasis and volume loss on the left and areas of mild cylindrical bronchiectasis and "tree-in-bud" bronchiolitis in the right lower lobe.

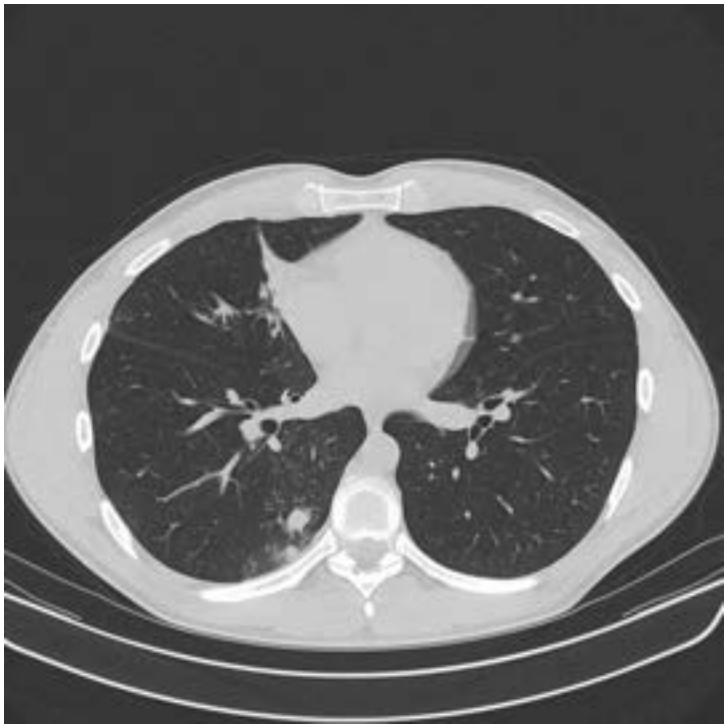

E-FIGURE 90-2. A thin-slice high-resolution computed tomography image through the middle chest; there are mild bronchiectasis in the right middle lobe adjacent to the heart and mucous plugging in areas of bronchiectasis at the right base.

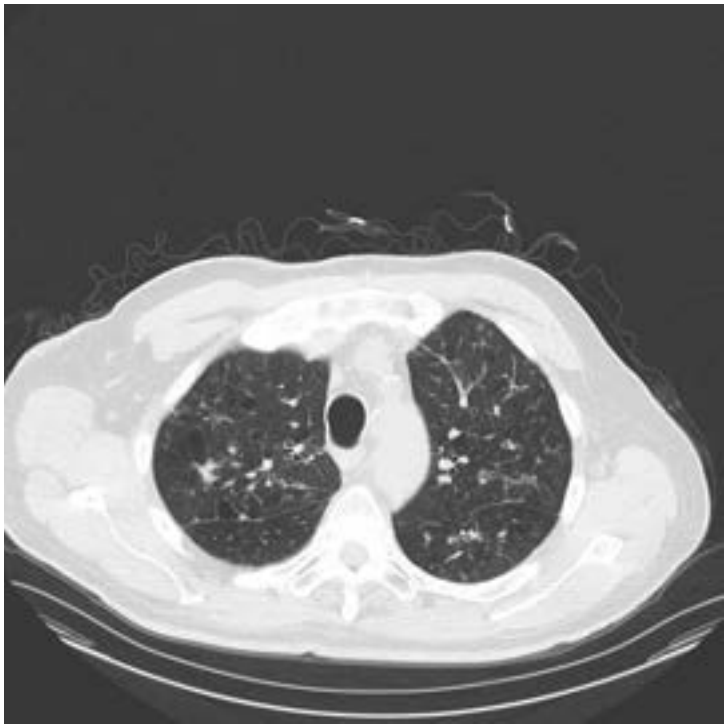

E-FIGURE 90-3. Bronchiectatic changes and mucous plugging superimposed on diffuse emphysematous changes in the lungs.

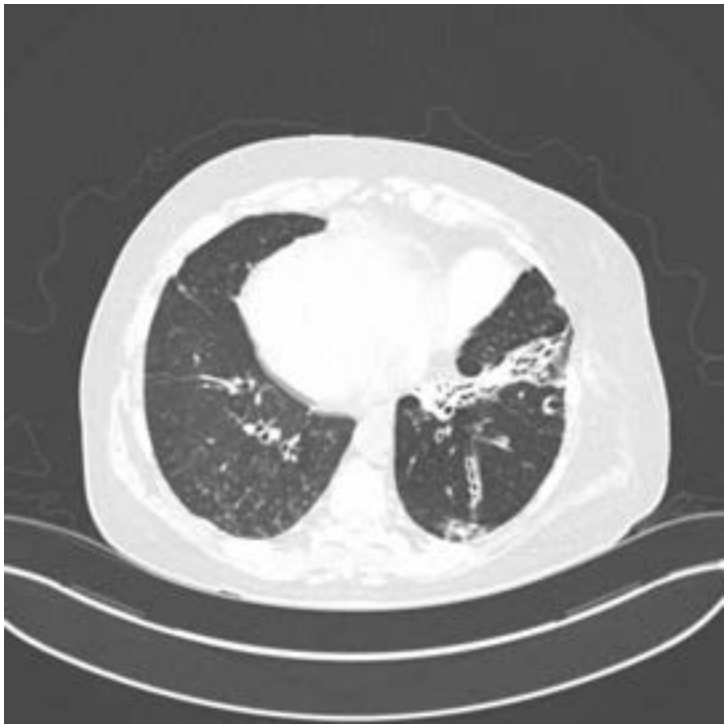

E-FIGURE 90-4. Extensive bronchiectasis in lower cuts through the right lung with associated volume loss.

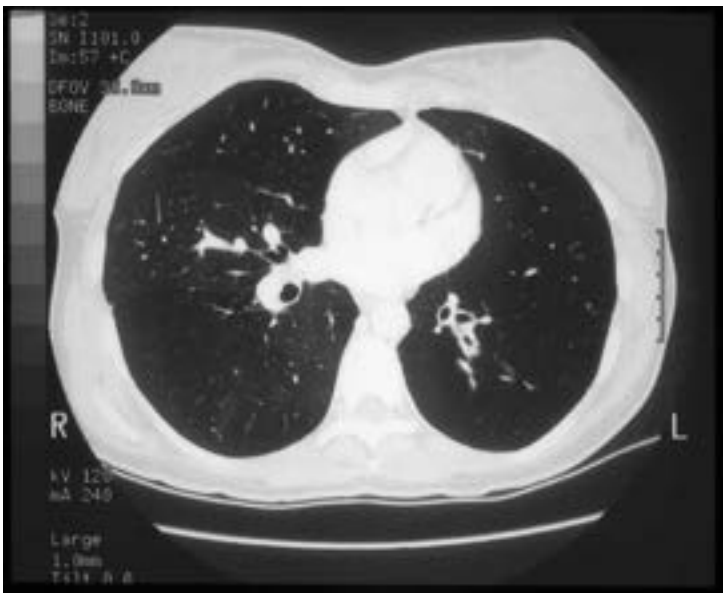

E-FIGURE 90-5. Computed tomography slice showing a "signet ring" abnormality consistent with a localized area of bronchiectasis in the right lower lobe.

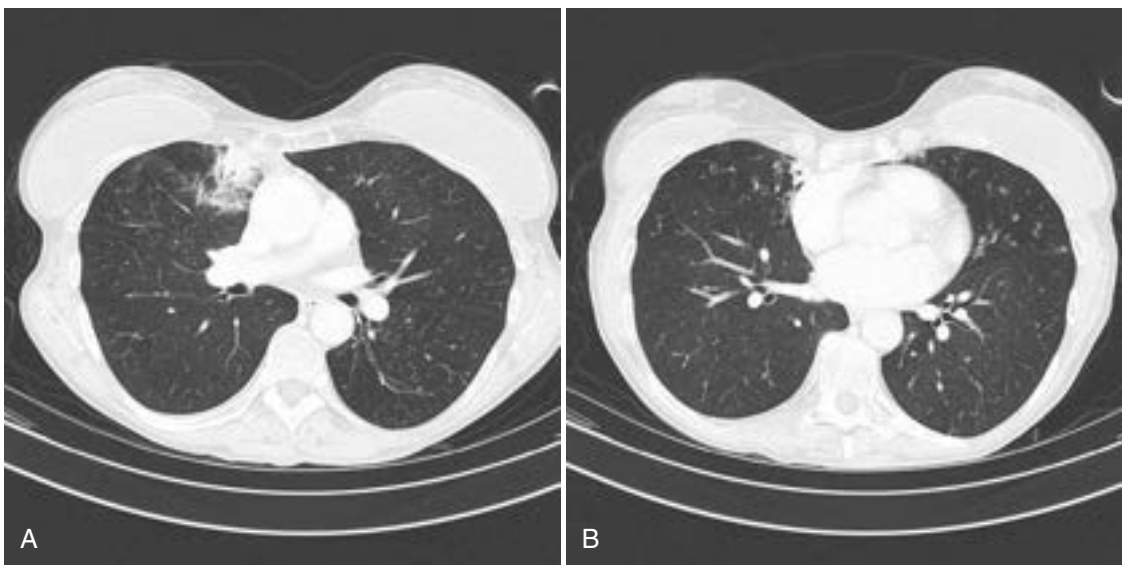

E-FIGURE 90-6. **A,** A pneumonic infiltrate superimposed on an area of bronchiectasis in the right middle lobe. Bilateral breast implants are also visible. The patient had a history of chronic mycobacterial infection but was acutely ill with a superimposed bacterial pneumonia when this study was performed. **B,** Follow-up imaging after treatment of the acute infection. A residual minor area of right middle lobe bronchiectasis is seen.

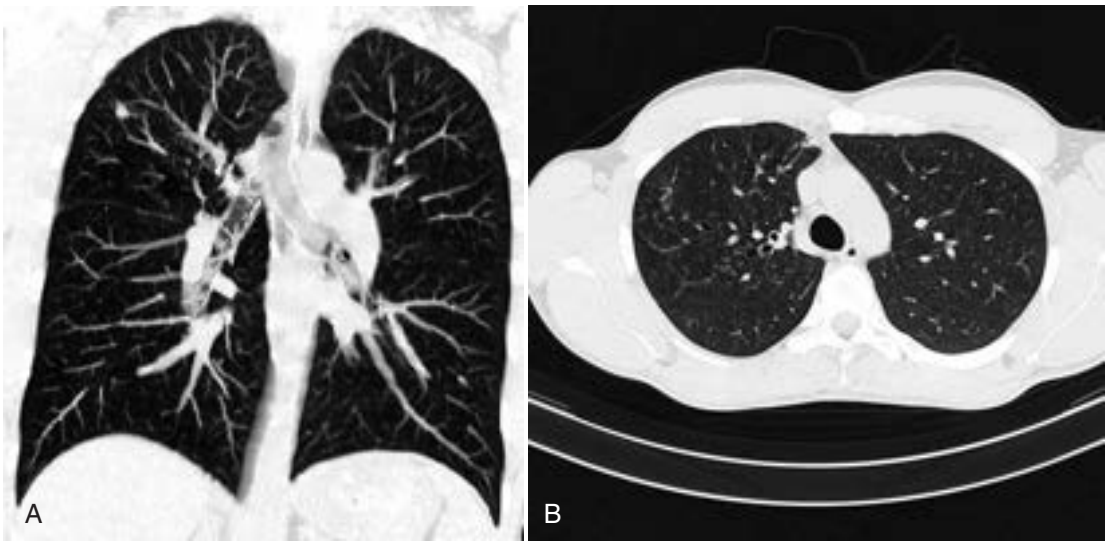

FIGURE 90-3. A and B, High-resolution computed tomography images of finger-in-glove central bronchiectasis caused by allergic bronchopulmonary aspergillosis.

The dominant organisms are *P. aeruginosa* and *Haemophilus influenzae*, but anaerobic organisms may also be detected. The presence of *P. aeruginosa* portends a worse prognosis and more frequent exacerbations. Patients with no identifiable pathogens have the mildest disease. *Staphylococcus aureus* in the airway may suggest cystic fibrosis as the cause of the bronchiectasis. Nontuberculous mycobacteria are found with increasing frequency in the airways of patients with bronchiectasis, usually as a complication of preexisting bronchiectasis but occasionally as its primary cause. The laboratory evaluation of patients with bronchiectasis should be individualized. All patients should have sputum cultures for bacterial and mycobacterial testing. Other tests that should be considered include measurement of serum immunoglobulin levels and screening for genetic diseases, particularly in patients with diffuse bronchiectasis. Cystic fibrosis (Chapter 89) is diagnosed by elevated sweat chloride levels and by genetic testing. Primary ciliary dyskinesia can be evaluated by measurement of nasal nitric oxide levels, ciliary beat frequency and pattern testing, and electron microscopy studies.[6] α_1-Antitrypsin deficiency is diagnosed by measuring levels and performing phenotyping (Chapter 88). Screening for rheumatoid arthritis (Chapter 264) or Sjögren syndrome (Chapter 268) also may be reasonable in patients with diffuse bronchiectasis.

TREATMENT

Rx

The goals of treatment are to reduce the frequency of exacerbations and potentially to improve quality of life, reduce symptoms, and alter the natural history of the disease (Table 90-1). Multimodality maintenance treatment[7] for patients with more advanced disease or three or more exacerbations may include airway clearance and anti-inflammatory therapies, as well as short and long-term antibiotic therapy, which reduces markers of airways and systemic inflammation. Exacerbations are treated based on clinical acuity. Because patients are heterogeneous and therapeutic trials are few, therapy is commonly individualized, especially because no therapies are currently approved by the U.S. Food and Drug Administration for non–cystic fibrosis bronchiectasis and because the proven treatments for cystic fibrosis are often not effective.

Preventing Exacerbations

The 23-valent pneumococcal vaccination (Chapter 18) is recommended for patients with bronchiectasis. Routine seasonal influenza vaccination is also standard. At present, no vaccines are available for prevention of the other infectious complications of bronchiectasis.

Treatment of the Underlying Etiology

For treatable conditions, such as immunoglobulin deficiency, replacement therapy (Chapter 250) should be considered even though there are few data on whether that alters the natural history of the lung disease. Patients with allergic bronchopulmonary aspergillosis (Chapter 339) should be treated with steroids to mitigate the inflammatory process that leads to the bronchiectasis.

TABLE 90-1 POTENTIAL THERAPIES FOR BRONCHIECTASIS

Treat underlying condition, if possible
Mobilization of secretions
Pharmacologic
Mechanical
Anti-inflammatory therapy
Inhaled steroids
Macrolides
Antimicrobial therapy
Pathogen specific
Surgery
Localized or refractory disease
Transplantation
End-stage disease

Adapted from O'Donnell A. Bronchiectasis. *Chest.* 2008;134:815-823.

Airway Clearance

Chest physiotherapy and the use of devices to aid mucociliary clearance appear to be beneficial in non–cystic fibrosis bronchiectasis. In a randomized trial, for example, twice-daily use of an oscillatory positive expiratory pressure device (Acapella) improved sputum volume and quality of life end points compared with no routine physiotherapy.■ Other techniques that may also have a role for airway clearance include traditional chest physical therapy with postural drainage and the use of chest wall oscillator vests.[8] Formal pulmonary rehabilitation and exercise also likely provide benefit to patients with bronchiectasis.

Inhaled therapy with nebulized hypertonic saline (7%) may enhance airway clearance, decrease exacerbations, and improve lung function as well as quality of life.■ Chronic inhalation of dry powder mannitol improves sputum clearance but does not reduce the frequency of exacerbations. Although recombinant human DNase is efficacious in cystic fibrosis bronchiectasis, a large clinical trial showed it had deleterious effects when given a maintenance therapy in patients with non–cystic fibrosis bronchiectasis, so it should not be used. Other mucolytic agents are of unproven benefit.■

No randomized trials support the use of routine short-acting β-agonist or anticholinergic bronchodilators in bronchiectasis. However, a subset of patients with airway reactivity likely benefits from use of these agents (Chapter 87).

Reduction of Airway Inflammation

One clinical trial demonstrated that inhaled medium-dose budesonide, when combined with formoterol, is safe and more effective than high-dose budesonide in treating patients with non–cystic fibrosis bronchiectasis.■ Oral steroids, although occasionally used in patients with bronchiectasis, have never been evaluated in a clinical trial.

Antimicrobial Therapy

At present, there is no firm evidence to support the use of routine maintenance antibiotics, although such therapy may be considered in patients with frequent exacerbations and progressive lung destruction. When

mycobacterial species are cultured from patients with bronchiectasis, decisions regarding whether to treat and which antimicrobial agents to use are based on published guidelines (Chapters 324 and 325).

Chronic low-dose oral macrolide therapy (e.g., azithromycin 500 mg three times per week or 250 mg/day or erythromycin ethylsuccinate 400 mg twice daily) can reduce exacerbations in patients with non–cystic fibrosis bronchiectasis.[■■] Whether patients will develop resistant organisms is of concern, and macrolide therapy alone should not be used in patients co-infected with nontuberculous mycobacteria.

Targeted inhaled antimicrobial therapies are also an option, particularly in patients infected with *Pseudomonas* spp.[■] For example, nebulized gentamicin (80 mg twice daily) for 12 months can provide sustained bacteriologic and clinical benefit. Clinical trials have demonstrated microbiologic benefits with inhaled tobramycin, 300 mg twice per day as a 4-week trial for one cycle and a 2-week-on, 2-week-off trial for three cycles, but clinical benefit was not firmly established, and some patients experienced unacceptable respiratory side effects. Antimicrobial resistance is also a concern. Inhaled colistin, 1 million IU twice daily delivered by nebulizer, also was recently shown to be safe and possibly effective in adherent patients with bronchiectasis and pseudomonas aeruginosa infection.[■] Additional inhaled antibiotics currently being evaluated in clinical trials include dry powder ciprofloxacin, nebulized liposomal ciprofloxacin, and dry powder tobramycin. Other off-label antibiotic strategies include prolonged intravenous antibiotics targeted at the cultured pathogens.

Surgery and Transplantation

Resectional surgery may have a role for patients who have focal disease or for patients who have hemoptysis that cannot be controlled by embolization of the bleeding vessels (Chapter 101). Surgical resection can also benefit some patients who have diffuse bronchiectasis unresponsive to conventional therapy and some patients infected with nontuberculous mycobacteria. Double-lung transplantation (Chapter 101) has been successfully performed in patients with end-stage lung disease caused by non–cystic fibrosis bronchiectasis, and the clinical outcomes parallel those seen with transplantation for other end-stage lung diseases.

Treatment of Acute Exacerbations of Bronchiectasis

When a patient with bronchiectasis experiences an acute exacerbation, antimicrobial treatments should be aimed at the known infecting organisms. Mild to moderate exacerbations can be treated with oral antibiotics, targeted to the results of the sputum culture, for 2 to 3 weeks. More severe exacerbations or exacerbations caused by resistant organisms generally require intravenous antibiotics administered in hospital or at home. No benefit has yet been demonstrated by adding an inhaled antibiotic to systemic therapy for an acute exacerbation. Patients experiencing an acute exacerbation likely benefit from airway clearance modalities and the other nonantibiotic therapies discussed previously.

PROGNOSIS

Non–cystic fibrosis bronchiectasis is a heterogeneous disease with a widely variable prognosis. Patients with more severe obstructive and restrictive findings on pulmonary function tests, poor gas transfer, and chronic pseudomonal infection have the worst prognosis.[9] Independent predictors of future hospitalization include prior hospital admissions, advanced dyspnea, FEV_1 less than 30% predicted, *P. aeruginosa* colonization, colonization with other pathogenic organisms, and three or more lobes involved on HRCT.[10] Bronchiectasis in patients with moderate to severe COPD is an independent risk factor for all-cause mortality. Radiographic extent of disease, hypoxemia, hypercapnia, and evidence of right heart failure are also predictors of outcome. Bronchiectasis patients who are admitted to an intensive care unit for respiratory failure have been reported to have a 60% 4-year survival rate.

ATELECTASIS

DEFINITION

Atelectasis, or collapse, is caused by hypoventilation of lung units. Atelectasis may involve an entire lung or a lobe, segment, or subsegment. Atelectasis can be caused by intrinsic obstruction of an airway or external compression from lymph nodes, parenchymal masses, or other entities. When lung units are atelectatic, ventilation-perfusion mismatch leads to hypoxemia. Infection may result from sustained atelectasis.

EPIDEMIOLOGY AND PATHOBIOLOGY

The lung bases and posterior segments are vulnerable to dependent atelectasis, which is caused by inadequate ventilation, particularly in an immobilized or postoperative patient. Patchy atelectasis is caused by alveolar filling processes, such as hemorrhage and edema (Chapter 91). Passive, relaxation, or compression atelectasis occurs when the lung recoils to a smaller volume because of fluid or air in the adjacent pleural space.

Obstructive or resorptive atelectasis is caused by bronchial block to the entry of air, with resultant retractile consolidation. Intrinsic airway obstruction may be caused by mucous plugs, foreign bodies, or tumors in the airway. Extrinsic airway obstruction results from compression of the airway owing to peribronchial lymph node enlargement or other masses impinging on the airway.

Rounded atelectasis is caused by pleural thickening that invaginates and traps adjacent lung. Any chronic pleural disease can cause rounded atelectasis, particularly asbestos-related pleural disease.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Atelectasis is typically asymptomatic and diagnosed on chest imaging, but it may cause dyspnea and tachypnea and result in hypoxemia. In postoperative patients, atelectasis may be a cause of low-grade fever. Plain chest radiography shows loss of lung volume and the displacement of the lobar fissure, mediastinum, or diaphragm toward the involved lung unit (Figs. 90-4 and 90-5). Platelike or discoid atelectasis manifests as horizontal or curvilinear lines on plain chest radiography. Rounded atelectasis is an ovoid masslike density abutting the pleura. The type and cause of atelectasis can sometimes be

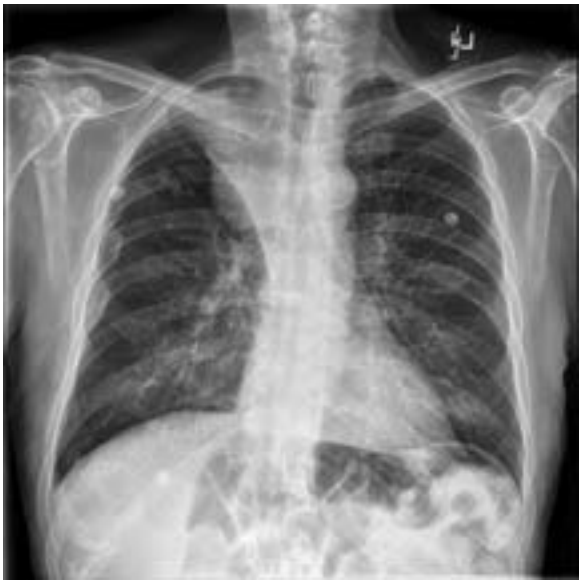

FIGURE 90-4. Plain chest radiograph demonstrating right upper lobe atelectasis (caused by an endobronchial tumor).

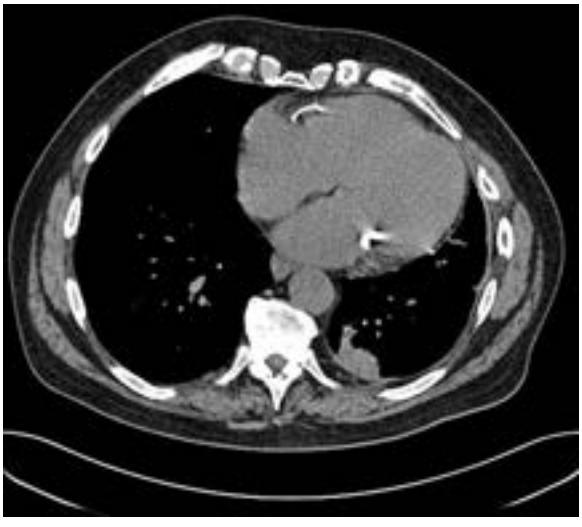

FIGURE 90-5. Computed tomography image of rounded atelectasis.

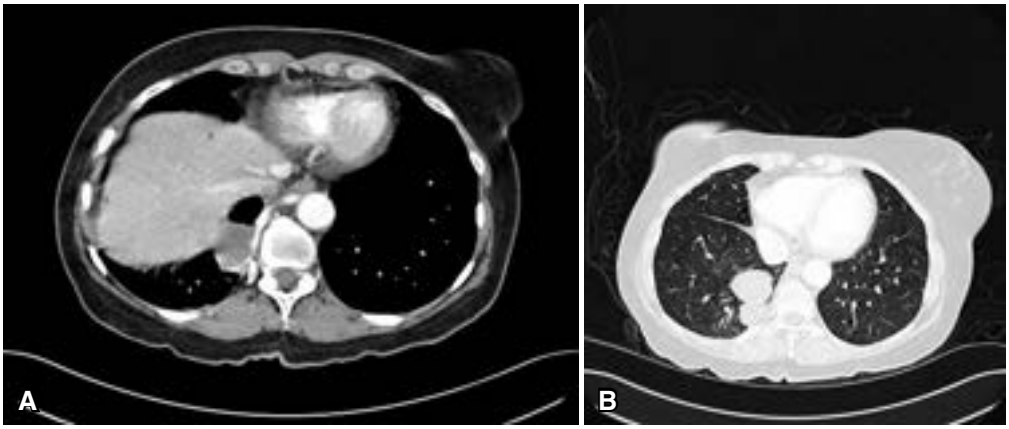

FIGURE 90-6. **Pulmonary sequestration. A,** Computed tomography image of pulmonary sequestration in right lower lobe. **B,** Feeding vessel visible arising from the aorta.

elucidated by CT or ultrasonography. Bronchoscopy is required to confirm intrinsic versus extrinsic compression in obstructive-resorptive atelectasis and to determine the exact pathology of the obstruction. An oxygen saturation concentration can help assess the severity of the atelectasis and overall lung dysfunction.

PREVENTION AND TREATMENT

Rx

Incentive spirometry is commonly prescribed to prevent or treat atelectasis in patients with limited mobility because of recent surgery, neuromuscular weakness, or any prolonged immobilization, no randomized controlled trials have proven its effectiveness. Preoperative inspiratory muscle training may reduce atelectasis in patients undergoing upper abdominal surgery,[11] and prophylactic use of noninvasive ventilation may reduce pulmonary dysfunction after lung resection surgery. Other modalities such as positive expiratory pressure devices and high-frequency chest wall oscillation airway clearance are of uncertain benefit.

Patchy atelectasis is treated by addressing the underlying disease process in the lung parenchyma. Compression atelectasis is treated by alleviating the pleural space process.

Obstructive or resorptive atelectasis often requires bronchoscopy for diagnosis and treatment. In patients with obstruction owing to retained secretions, multiple bronchoscopies are sometimes required, but the mucus often rapidly reaccumulates and will resolve only when the patient's overall status improves.

Rounded atelectasis does not require treatment. CT is helpful in distinguishing rounded atelectasis from parenchymal tumor.

CONGENITAL CYSTIC DISEASES OF THE THORAX

Thoracic cysts, which are exceedingly rare, develop because of abnormal development or branching of the foregut. Cysts may develop in the mediastinum at an early stage of gestation or in the lung parenchyma at a later stage. Abnormalities include bronchogenic cysts (mediastinal and parenchymal), congenital pulmonary airway malformation, and pulmonary sequestrations. The cysts are lined with airway and alveolar epithelium but do not communicate in a normal fashion with the airways or lung tissue.

Most patients with thoracic cysts present in childhood, but the cysts can remain asymptomatic and unnoticed until adulthood. In the absence of symptoms, these cystic lesions sometimes present as an incidental finding on chest imaging performed for another indication. Congenital cystic diseases can cause recurrent pneumonia, hemoptysis, or compression of normal structures.

Computed tomography scanning with CT angiography can usually detect congenital cystic lesions of the thorax, but pulmonary or bronchial angiography is sometimes necessary to define the blood flow to the lesion.

Bronchogenic cysts are usually found in the right paratracheal or subcarinal areas of the mediastinum but are occasionally seen in the lung parenchyma.[12] These cysts are often asymptomatic, but they can cause wheezing, dyspnea, and cough when they compress adjacent structures. Secondary infection may develop in the cysts, and there are a few case reports of malignant transformation. Complete surgical resection is generally recommended, but partial excision with de-epithelization of the cysts has also been performed. Observation is also an option when the cysts are asymptomatic.

Congenital pulmonary airway malformation, previously called *congenital cystic adenomatoid malformation of the lung*, is an exceedingly rare abnormality with reported incidence of one in every 25,000 to 35,000 pregnancies. The abnormality is caused by arrested development of the bronchial tree. Most patients are diagnosed prenatally by ultrasonography, but a few adults have first presented with complications, including pneumothorax and air embolism. Treatment is anatomic surgical resection.

Pulmonary sequestrations are areas of nonfunctioning pulmonary parenchyma with no communication to the tracheobronchial tree and abnormal arterial supply and venous drainage (Fig. 90-6). Intralobar sequestration, which accounts for about 75% of cases, does not have visceral pleura and is generally found in a lower lobe, the left more frequently than the right. Extralobar sequestrations have their own visceral pleura, are separate from the normal lobes, and may even be found below the diaphragms. Sequestrations usually have a feeding vessel that arises from the aorta. Patients with sequestrations may be asymptomatic but sometimes develop recurrent infections and or hemoptysis. Surgical excision with special care for the management of the feeding vessel is curative. Embolization of the feeding vessel is sometimes a successful treatment option.

Hyperlucent lungs are diagnosed by a paucity of vascular and interstitial markings noted on chest imaging. Lung parenchymal air collections can be caused by congenital parenchymal cysts, congenital lobar emphysema (almost exclusively diagnosed in infancy), giant bullous emphysema (vanishing lung syndrome), or Swyer-James syndrome. Lung parenchymal cysts may be a bullous alveolar type or may contain bronchial wall elements such as cartilage, smooth muscle, and glands. They may become infected and may rupture to cause pneumothorax. Surgical resection is generally recommended unless the lesions are small. Congenital lobar emphysema, otherwise known as *congenital large hyperlucent lobe*, may cause severe respiratory distress in infants owing to compression of surrounding lung tissue. Giant bullous emphysema is a rare condition that usually affects the upper lobes of young male smokers. Compression of normal lung parenchyma from these overdistended lobes may require surgical resection.

Swyer-James-Macleod syndrome, which is characterized by unilateral lucency of an entire lung, is caused by childhood bronchiolitis obliterans owing to viral or bacterial infection or toxic inhalation. CT shows air trapping and hyperlucency of the affected lung, with a normal contralateral lung. No therapy is required.

Grade A References

A1. Patterson JE, Hewitt O, Kent L, et al. Acapella versus "usual airway clearance" during acute exacerbation in bronchiectasis: a randomized crossover trial. *Chron Respir Dis*. 2007;4:67-74.
A2. Kellett F, Robert NM. Nebulised 7% hypertonic saline improves lung function and quality of life in bronchiectasis. *Respir Med*. 2011;105:1831-1835.
A3. Wilkinson M, Sugumar K, Milan SJ, et al. Mucolytics for bronchiectasis. *Cochrane Database Syst Rev*. 2014;CD001289.
A4. Martinez-Garcia MA, Soler-Cataluna JJ, Catalan-Serra P, et al. Clinical efficacy and safety of budesonide-formoterol in non-cystic fibrosis bronchiectasis. *Chest*. 2012;141:461-468.
A5. Altenburg J, de Graaff CS, Stienstra Y, et al. Effect of azithromycin maintenance treatment on infectious exacerbations among patients with non-cystic fibrosis bronchiectasis: the BAT randomized controlled trial. *JAMA*. 2013;309:1251-1259.

A6. Serisier DJ, Martin ML, McGuckin MA, et al. Effect of long-term, low-dose erythromycin on pulmonary exacerbations among patients with non-cystic fibrosis bronchiectasis: the BLESS randomized controlled trial. *JAMA*. 2013;309:1260-1267.
A7. Brodt AM, Stovold E, Zhang L. Inhaled antibiotics for stable non-cystic fibrosis bronchiectasis: a systematic review. *Eur Respir J*. 2014;44:382-393.
A8. Haworth CS, Foweraker JE, Wilkinson P, et al. Inhaled colistin in patients with bronchiectasis and chronic Pseudomonas aeruginosa infection. *Am J Respir Crit Care Med*. 2014;189:975-982.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. McShane PJ, Naureckas ET, Tino G, et al. Non-cystic fibrosis bronchiectasis. *Am J Respir Crit Care Med*. 2013;188:647-656.
2. Seitz AE, Olivier KN, Adjemian J, et al. Trends in bronchiectasis among Medicare beneficiaries in the United States, 2000 to 2007. *Chest*. 2012;142:432-439.
3. Moulton BC, Barker AF. Pathogenesis of bronchiectasis. *Clin Chest Med*. 2012;33:211-217.
4. Bienvenu T, Sermet-Gaudelus I, Burgel PR, et al. Cystic fibrosis transmembrane conductance regulator channel dysfunction in non-cystic fibrosis bronchiectasis. *Am J Respir Crit Care Med*. 2010; 181:1078-1084.
5. Tunney MM, Einarsson GG, Wei L, et al. Lung microbiota and bacterial abundance in patients with bronchiectasis when clinically stable and during exacerbation. *Am J Respir Crit Care Med*. 2013; 187:1118-1126.
6. Knowles MR, Daniels LA, Davis SD, et al. Primary ciliary dyskinesia. Recent advances in diagnostics, genetics, and characterization of clinical disease. *Am J Respir Crit Care Med*. 2013;188: 913-922.
7. Pasteur MC, Bilton D, Hill AT. British Thoracic Society guideline for non-CF bronchiectasis. *Thorax*. 2010;65(suppl 1):i1-i58.
8. Flude LJ, Agent P, Bilton D. Chest physiotherapy techniques in bronchiectasis. *Clin Chest Med*. 2012;33:351-361.
9. Loebinger MR, Wells AU, Hansell DM, et al. Mortality in bronchiectasis: a long-term study assessing the factors influencing survival. *Eur Respir J*. 2009;34:843-849.
10. Chalmers JD, Goeminne P, Aliberti S, et al. The bronchiectasis severity index. An international derivation and validation study. *Am J Respir Crit Care Med*. 2014;189:576-585.
11. Restrepo RD, Wettstein R, Wittnebel L, et al. Incentive spirometry: 2011. *Respir Care*. 2011; 56:1600-1604.
12. Cilleruelo Ramos A, Ovelar Arribas Y, Garcia Yuste M. Cervical bronchogenic cyst in adults. case report and literature review. *Arch Bronconeumol*. 2015;51:95-96.

REVIEW QUESTIONS

1. Of the following, which organism is associated with the worst prognosis in non–cystic fibrosis bronchiectasis?

A. Staphylococcus aureus
B. *Escherichia coli*
C. Streptococcus pneumoniae
D. *Pseudomonas aeruginosa*
E. *Moraxella catarrhalis*

Answer: D *Pseudomonas aeruginosa* is the correct answer. Molecular techniques have demonstrated that diverse polymicrobial communities are present in the lungs of patients with bronchiectasis, both when they are clinically stable as well as during exacerbations. The dominant organisms are *P. aeruginosa* and *Haemophilus influenzae*, but anaerobic organisms may also be detected.

2. Of the following, which organism found in your patient with bronchiectasis would raise the suspicion for undiagnosed cystic fibrosis?

A. *Pseudomonas aeruginosa*
B. *Staphylococcus aureus*
C. *Mycobacterium avium* complex
D. *Streptococcus pneumoniae*
E. *Acinetobacter baumannii*

Answer: B The presence of *Staphylococcus aureus* raises the suspicion for cystic fibrosis, and further testing to establish or exclude this diagnosis is indicated.

3. Which imaging technique is used to confirm the diagnosis of bronchiectasis?

A. Routine chest radiography
B. Noncontrast high-resolution CT scan
C. CT of the chest with contrast
D. MRI of thorax
E. Bronchography

Answer: B High-resolution computed tomography of the chest without the need for contrast is the imaging test of choice for confirming bronchiectasis

4. Which of the following is most predictive of poor prognosis in patients with bronchiectasis?

A. Chronic infection with *Mycobacterium avium* complex
B. Total lung capacity less than 70% predicted
C. Admission to the ICU for respiratory failure
D. Chronic infection with *Staphylococcus aureus*
E. Dependence on supplemental oxygen

Answer: C When a patient with bronchiectasis has developed respiratory failure, the long-term prognosis is poor. More than 60% of patients admitted to a medical intensive care unit for bronchiectasis will die during that hospitalization.

91 ALVEOLAR FILLING DISORDERS

STEPHANIE M. LEVINE

DEFINITION

Alveolar filling disorders (Table 91-1) are characterized by chest radiographic findings of alveolar involvement ranging from a ground-glass appearance to consolidation; the pathologic process shows primary involvement of the alveolar air spaces distal to the terminal bronchioles. For example, in pulmonary alveolar proteinosis, the alveoli are filled by proteinaceous fluid. By comparison, the alveolar walls are lined by adenocarcinoma cells in invasive mucinous adenocarcinoma and lepidic predominant nonmucinous adenocarcinoma, formerly called bronchioloalveolar cell cancer. In acute interstitial pneumonia, exudative organizing fibroproliferative infiltrates fill the alveolar space; in the alveolar hemorrhage disorders, blood fills the alveolar space. Alveolar spaces filled with acute inflammatory cells, as in bacterial pneumonia (Chapter 97); water, as in cardiogenic or hydrostatic pulmonary edema (Chapter 58); or high-protein fluid, as in noncardiogenic or increased permeability pulmonary edema (Chapter 104), are also part of the radiographic differential diagnosis of alveolar filling disorders and must be excluded.

A general approach to these suspected alveolar filling diseases (Fig. 91-1) can be stratified by the time elapsed since the onset of symptoms. The typical patient may present with the onset of cough (usually dry) and dyspnea of variable duration, depending on the disease process. Hemoptysis is a frequent presenting symptom in the alveolar hemorrhagic disorders. With the exception of acute interstitial pneumonia, symptoms suggesting an acute infectious process such as fever, leukocytosis, and productive cough are usually absent. If the initial chest radiograph or chest computed tomography (CT) is consistent with a possible alveolar filling process (Chapter 84) and acute pneumonia and pulmonary edema are excluded, bronchoscopy with bronchoalveolar lavage (BAL) (Chapter 85) and transbronchial biopsy should be performed, particularly if pulmonary alveolar proteinosis, invasive mucinous adenocarcinoma, lepidic predominant nonmucinous adenocarcinoma (Chapter 191), or alveolar hemorrhage is suspected. When these tests are nondiagnostic, and in many cases of suspected acute interstitial pneumonia, a surgical lung biopsy obtained by thoracoscopy or an open surgical procedure may be indicated.

PULMONARY ALVEOLAR PROTEINOSIS

EPIDEMIOLOGY

Pulmonary alveolar proteinosis is a rare alveolar filling disease caused by the accumulation of phospholipoproteinaceous surfactant material in the alveoli. The incidence is estimated to be 3.7 cases per million people. Pulmonary

TABLE 91-1 ALVEOLAR FILLING DISORDERS

DISEASES	PATHOPHYSIOLOGY	RADIOGRAPHIC FINDINGS
Pulmonary alveolar proteinosis	Impaired processing of surfactant by alveolar macrophages caused by defects in GM-CSF signaling	Bilateral alveolar opacities with "crazy paving" and diffuse areas of ground-glass attenuation on CT scan
Acute interstitial pneumonia	Diffuse alveolar damage with temporal uniformity	Diffuse alveolar filling process similar to the acute respiratory distress syndrome
Diffuse alveolar hemorrhage	Bleeding from the pulmonary microcirculation, usually from the capillaries	Acute development of bilateral alveolar opacities
Invasive mucinous adenocarcinoma and lepidic predominant nonmucinous adenocarcinoma (formerly called bronchioloalveolar cell carcinoma)	Cancer cells growing along the alveolar septa	Pneumonic opacities, consolidation with air bronchograms, ground-glass opacities (either solitary or multiple)

CT = computed tomography; GM-CSF = granulocyte-macrophage colony-stimulating factor.

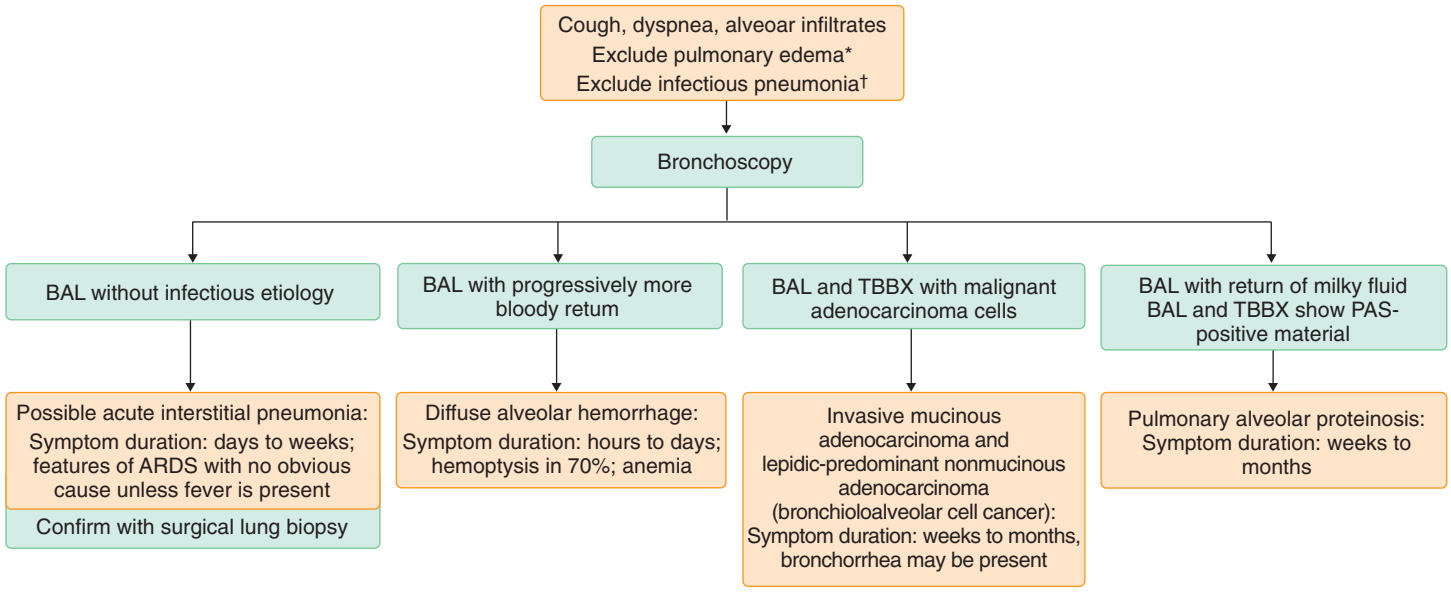

FIGURE 91-1. A general approach to the alveolar filling disorders. *See Chapter 58. †See Chapter 97. ARDS = acute respiratory distress syndrome; BAL = bronchoalveolar lavage; PAS = periodic acid–Schiff; TBBX = transbronchial biopsy.

alveolar proteinosis is a primary acquired, autoimmune disorder in more than 90% of cases, but similar histopathologic features may be found with identifiable secondary causes, such as acute silicosis (silicoproteinosis; Chapter 93), aluminum dust exposure (Chapter 93), indium dust exposure, immunodeficiency disorders (e.g., immunoglobulin G monoclonal gammopathy and severe combined immunodeficiency syndrome), hematologic malignant neoplasms (particularly myeloid leukemias; Chapters 183 and 184), and certain infections (e.g., *Pneumocystis jiroveci* pneumonia). Pulmonary alveolar proteinosis has also been described after bone marrow transplantation (Chapter 178). A congenital form also usually presents in infancy.

PATHOBIOLOGY

The pathogenesis of pulmonary alveolar proteinosis is related to impaired processing of surfactant by alveolar macrophages caused by defects in granulocyte-macrophage colony-stimulating factor (GM-CSF) signaling. This impairment may be caused by autoantibodies against GM-CSF or GM-CSF receptor gene mutations, which are found in 90% of cases. The secondary forms are caused by a relative GM-CSF deficiency, leading to macrophage dysfunction and reduced surfactant clearance. The autosomal recessive congenital form of pulmonary alveolar proteinosis, caused by a mutation in the genes encoding surfactant protein B or C, or the GM-CSF receptor results in abnormal surfactant function and severe respiratory distress in homozygous infants. The result of this impairment is accumulation of surfactant-rich material and progressive dysfunction in phagocytosis caused by excessive production or diminished clearance of surfactant by alveolar macrophages.

Histologic examination in pulmonary alveolar proteinosis reveals alveoli filled with lipoproteinaceous material that stains pink (positive reaction) with periodic acid–Schiff stain. Classically, there is no destruction of alveolar architecture. Electron microscopy reveals lamellar (phospholipid-containing) myelin bodies.

CLINICAL MANIFESTATIONS

Pulmonary alveolar proteinosis presents in patients in the third to fourth decade of life with a 2:1 male predominance. Most patients (72%) are smokers. Patients present with the insidious onset of dyspnea and cough, which may be dry or occasionally productive of grayish material. The duration of symptoms before diagnosis is typically 6 weeks to 6 to 8 months. Low-grade fevers, malaise, and weight loss may also be present. Hemoptysis is unusual. On physical examination, rales are present in 50% of cases. Clubbing is an unusual finding until later stages of disease.

DIAGNOSIS

Mildly elevated leukocyte counts and mildly to moderately elevated lactate dehydrogenase (LDH) levels may be found in more than 80% of patients; LDH levels may correlate with the severity of disease. Polycythemia and hypergammaglobulinemia may also be present. The chest radiograph (Fig. 91-2) and chest CT scans demonstrate a diffuse symmetrical alveolar filling process with predominance in the lower two thirds of the lung fields; the radiographic appearance may mimic pulmonary edema. The characteristic CT pattern is often described as "crazy paving," which is attributable to scattered or diffuse areas of ground-glass attenuation with thickening of intralobular structures and interlobular septa in polygonal shapes (Fig. 91-3 and E-Fig. 91-E1).[1] This radiographic pattern is not specific for this disorder and can be seen with acute respiratory distress syndrome (ARDS; Chapter 104), *P. jiroveci* pneumonia (Chapter 341), adenocarcinomas formerly described as bronchioloalveolar carcinoma (Chapter 191), lipoid pneumonia (Chapter 94), sarcoidosis (Chapter 95), organizing pneumonia, drug reactions, and pulmonary hemorrhage as well as with cardiogenic pulmonary edema (Chapter 59) and acute interstitial pneumonias. Pulmonary function tests often, but not always, show a restrictive pattern with a reduced diffusing capacity. Arterial blood gas analyses reveal hypoxemia.

Bronchoscopy should be the initial procedure when pulmonary alveolar proteinosis is suspected. The diagnosis of pulmonary alveolar proteinosis can be established in most cases by the recovery of milky white to sandy-colored or light brown fluid on BAL. When it is subjected to cytologic analysis, the BAL fluid has a positive reaction on periodic acid–Schiff staining and reveals alveolar macrophages filled with positive staining material. Transbronchial biopsy or thoracoscopic biopsy can confirm the diagnosis by providing tissue that has similar staining characteristics. Serologic analysis for GM-CSF antibodies is now often performed to support the diagnosis.

TREATMENT Rx

About 8% to 30% of cases of pulmonary alveolar proteinosis resolve spontaneously, and smoking cessation may contribute to spontaneous resolution. A second group of patients will progress to respiratory failure. The remainder will have stable disease. Superinfection with *Nocardia* spp., atypical mycobacteria, fungi, and other opportunistic organisms can occur in more than 15% of patients as a result of macrocyte phagocytic dysfunction.[2]

Therapy depends on the severity of symptoms.[3] Treatment of severely symptomatic patients with dyspnea and hypoxemia begins with multistage or sequential whole-lung lavage performed under general anesthesia with a double-lumen endotracheal tube. The recovered fluid initially has an opaque, sandy-colored appearance (E-Fig. 91-E2). This procedure may have to be repeated at variable intervals. Asymptomatic patients should be observed and followed closely. Mildly symptomatic patients should receive supportive therapy with supplemental oxygen as needed. Subcutaneous or inhaled GM-CSF can improve quality of life, oxygenation, pulmonary function, and exercise capacity in about half of treated patients.[4,5] Treatment with GM-CSF is usually reserved for patients who cannot undergo whole-lung lavage or who have failed standard treatment with whole-lung lavage. Isolated case series have reported variable responses to CD-20 monoclonal antibody, rituximab, or plasmapheresis. Corticosteroids are not routinely used. Lung transplantation can be performed, but recurrent pulmonary alveolar proteinosis has been reported.

Survival rates at 5 years approach 75%.

ACUTE INTERSTITIAL PNEUMONIA

DEFINITION

Acute interstitial pneumonia, also referred to as the *Hamman-Rich syndrome*, is a rare and often fatal disease that mimics ARDS (Chapter 104). The etiology is unknown, and acute interstitial pneumonia is sometimes defined as

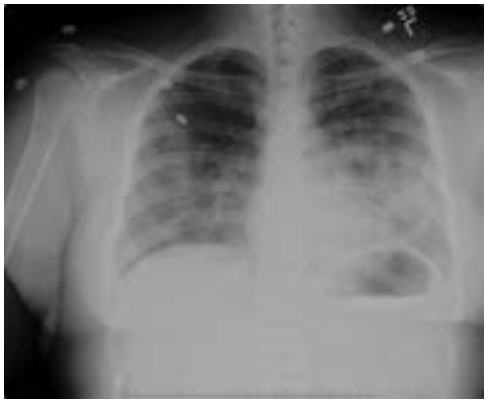

FIGURE 91-2. A chest radiograph showing bilateral alveolar opacities in a patient with pulmonary alveolar proteinosis.

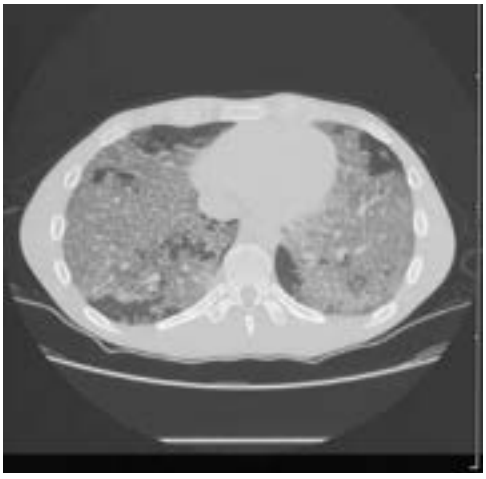

FIGURE 91-3. A chest computed tomography scan showing the "crazy paving" pattern characteristic of pulmonary alveolar proteinosis.

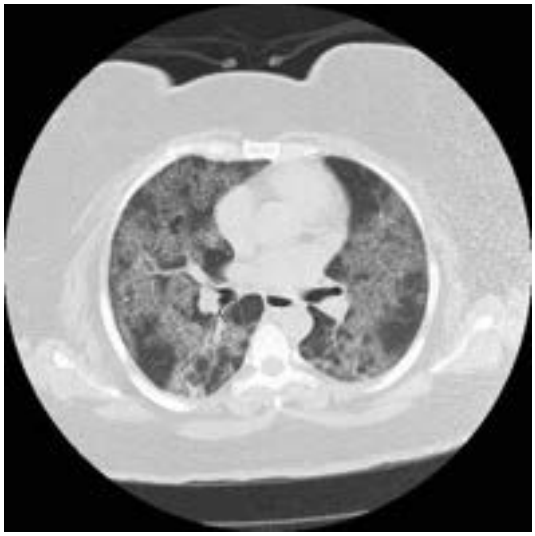

E-FIGURE 91-1. A chest computed tomography scan showing the "crazy paving" pattern characteristic of pulmonary alveolar proteinosis.

E-FIGURE 91-2. Whole-lung lavage from a patient with pulmonary alveolar proteinosis revealing sandy-colored fluid.

the development of ARDS in the absence of known triggers.[6] A similar acute presentation may be seen in patients with an acute exacerbation of idiopathic pulmonary fibrosis (Chapter 92), but most investigators believe that acute interstitial pneumonia is a separate disease process with no histologic evidence of underlying usual interstitial pneumonia.

PATHOBIOLOGY

The pathogenesis of acute interstitial pneumonia is damage to the epithelium of the alveolar membranes by a neutrophil-mediated mechanism; the result is pouring of exudate into the air space in the initial exudative phase of disease. Histologic examination reveals diffuse alveolar damage with intraalveolar hyaline membrane formation, interstitial and intraalveolar edema, acute inflammation, and epithelial cell necrosis with a nonspecific distribution and temporal uniformity. This process progresses to the organizing phase, characterized by alveolar septal thickening, type II pneumocyte hyperplasia, and fibroblast proliferation along the interstitium and alveolar spaces. In situ thrombi of small pulmonary arteries may be present. Finally, a fibrotic phase occurs with alveolar septal thickening from organizing fibrosis. One of the key pathologic findings in acute interstitial pneumonia is the temporal uniformity of the diffuse alveolar damage and of organizing and proliferating connective tissue. This uniformity supports a single acute injury at a particular point in time. Long-standing fibrosis is not a typical pathologic finding in acute interstitial pneumonia.

CLINICAL MANIFESTATIONS

Acute interstitial pneumonia manifests with equal frequency in men and women, typically in previously healthy individuals in the 50- to 55-year age range. It develops acutely to subacutely during a few days to a few weeks. The mean duration of symptoms is 15 days. Dry cough, shortness of breath, malaise, and fever (in 50% of patients) are typical clinical findings. A virus-like prodrome period has been described. Pulmonary rales are heard on physical examination, and hypoxemia is characteristic. Clubbing is rare. Acute interstitial pneumonia often progresses to hypoxemic ventilatory failure, and intensive care unit admission with mechanical ventilation is usually required. Early mortality rates are high. Radiographic features of acute interstitial pneumonia are diffuse alveolar opacities and air space consolidation similar to the appearance of ARDS (Fig. 91-4); CT scans reveal bilateral air space consolidation with areas of ground-glass opacities with little honeycombing. Septal thickening and a subpleural distribution of the opacities may also be present.

DIAGNOSIS

The diagnosis of acute interstitial pneumonia is made in the appropriate clinical setting in a patient who has a clinical presentation compatible with ARDS but without a clear etiology. The differential diagnosis histologically and clinically includes other causes of ARDS (Chapter 104), such as severe infection, trauma, and sepsis, as well as other causes of acute lung injury (Chapter 94), such as drug toxicity, inhalation injury, and collagen vascular diseases. The presentation is clinically and radiographically similar to that of diffuse alveolar hemorrhage, acute hypersensitivity pneumonitis, acute exacerbation of pulmonary fibrosis, acute eosinophilic pneumonia, and cryptogenic organizing pneumonia. Bronchoscopy with BAL is often performed to exclude alveolar hemorrhage, eosinophilic pneumonias, and infectious causes of lung injury. In a small number of cases, transbronchial biopsy may yield the diagnosis, but definitive diagnosis in most cases of acute interstitial pneumonia requires a surgical lung biopsy revealing diffuse alveolar damage.

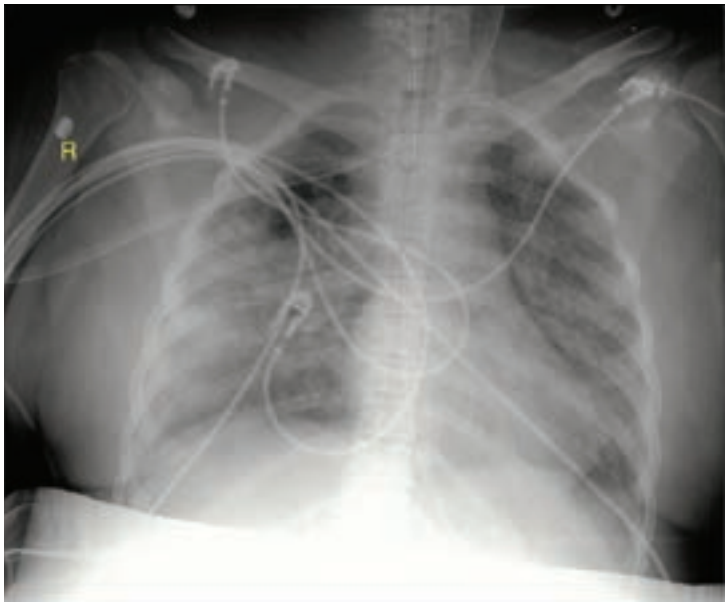

FIGURE 91-4. A chest radiograph showing bilateral alveolar opacities in a patient with acute interstitial pneumonia.

TREATMENT

Rx

Treatment includes supportive intensive care unit management. In small case series, corticosteroids at doses of 1 to 2 g of methylprednisolone in divided doses intravenously per day for 3 consecutive days followed by prednisone or equivalent at 1 mg/kg/day with a taper during several weeks to months, with or without cyclophosphamide, may be of benefit, but the mortality rate remains higher than 70%. Patients also can have recurrences in months to years. Some cases of acute interstitial pneumonia may resolve without sequelae, but in some series, more than 50% of survivors may be left with residual fibrosis.

DIFFUSE ALVEOLAR HEMORRHAGE

DEFINITION

The alveolar hemorrhage syndromes cause alveolar filling disease, usually with an acute onset and often with life-threatening severity. They can be associated with ANCA- (antineutrophil cytoplasmic antibody–) associated vasculitides,[7] such as microscopic polyangiitis (Chapter 270) and granulomatosis with polyangiitis (Chapter 270); immunologic diseases, such as Goodpasture syndrome (anti–glomerular basement membrane antibody disease; Chapter 121); collagen vascular diseases, such as systemic lupus erythematosus (Chapter 266); cocaine inhalation (Chapter 34); drugs (including penicillamine, mitomycin C, trimellitic anhydride, all-*trans* retinoic acid, propylthiouracil, and isocyanates); bone marrow transplantation (Chapter 178); coagulopathy (Chapter 174); and mitral stenosis (Chapter 75). A small percentage of idiopathic and recurrent cases are termed *idiopathic pulmonary hemosiderosis*. In Goodpasture syndrome, there is a strong association with tobacco use and a male predominance, with young men most frequently affected. A viral syndrome and exposure to hydrocarbons may simulate Goodpasture disease. Idiopathic pulmonary hemosiderosis most often occurs in children and young adults.

PATHOBIOLOGY

Alveolar hemorrhage is caused by bleeding from the pulmonary microcirculation, including the capillaries, arterioles, and venules. It may be associated with injury or neutrophilic inflammation of the alveolar walls and adjacent interstitial capillaries or with a capillaritis, usually when it is associated with collagen vascular or vasculitic processes. In Goodpasture syndrome, for example, the circulating anti–glomerular basement membrane antibodies are directed against the α_3 chain of type IV collagen in the glomerular basement membrane, where they cause glomerulonephritis; these core antibodies can cross-react with the alveolar capillary basement membranes, resulting in alveolar hemorrhage. Alternatively, alveolar hemorrhage may be associated with relatively bland pathologic changes with red blood cells in the alveolar spaces. Idiopathic pulmonary hemosiderosis is an example of bland hemorrhage.

CLINICAL MANIFESTATIONS

Patients present acutely (usually in hours to a week) with dyspnea, shortness of breath, hemoptysis (which may not be present in all patients), and cough. Some patients also have low-grade fever. Lung examination reveals rales.

Laboratory examination may reveal anemia, and arterial blood gases reveal hypoxemia. In Goodpasture syndrome and the ANCA-associated vasculitides, hematuria and renal insufficiency caused by glomerulonephritis are typically present.

Radiographic features include the acute development of bilateral alveolar filling disease similar to pulmonary edema but without cardiomegaly or

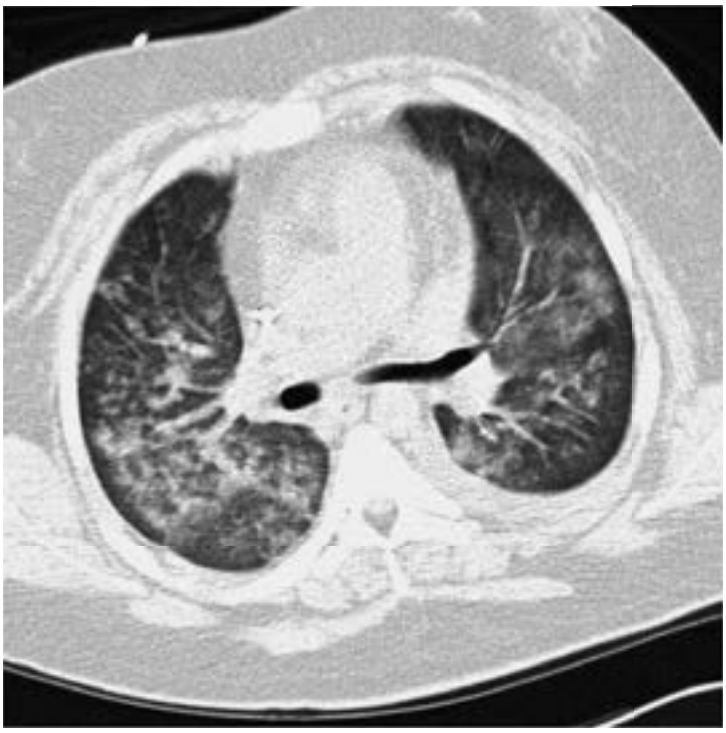

FIGURE 91-5. A chest computed tomography scan showing alveolar opacities in a patient with diffuse alveolar hemorrhage.

pleural effusions (E-Fig. 91-E3). Rapid remission and recurrences are seen with repeated episodes of bleeding, which also may result in chronic interstitial changes on the chest radiograph. Pulmonary function testing may reveal an increase in the diffusion capacity for carbon monoxide because of the presence of hemoglobin in the alveolar spaces.

DIAGNOSIS

The diagnosis of alveolar hemorrhage is usually made in the appropriate clinical setting by the triad of diffuse alveolar opacities (Fig. 91-5),[8] hemoptysis (in two thirds of patients), and anemia. BAL typically demonstrates the return of progressively more bloody aliquots of fluid (E-Fig. 91-E4), and cytologic analysis reveals that more than 20% of the macrophages are hemosiderin laden. Goodpasture syndrome is diagnosed by circulating anti–glomerular basement membrane antibodies, which are present in more than 90% of patients, or by the demonstration of linear deposition of immunoglobulin G antibodies along the alveolar or renal capillary basement membrane tissue when it is viewed by direct immunofluorescence. c-ANCA- (cytoplasmic antineutrophil cytoplasmic antibody–) associated vasculitis causes a focal, segmental, necrotizing glomerulonephritis and is associated with the presence of proteinase 3 antineutrophilic cytoplasmic antibodies in 90% of active cases of granulomatosis with polyangiitis (Chapter 270). Necrotizing granulomatous inflammation is often found in the upper airway in addition to the lungs and kidneys. A perinuclear myeloperoxidase antineutrophilic antibody is often present in association with microscopic polyarteritis (Chapter 270). Patients with systemic lupus erythematosus usually have antinuclear antibodies (Chapter 266). Patients using illicit drugs such as cocaine may have a positive drug screen. Idiopathic pulmonary hemosiderosis is a diagnosis of exclusion after other causes of diffuse alveolar hemorrhage have been eliminated.

TREATMENT

Rx

Treatment of alveolar hemorrhage varies according to its underlying cause. Massive hemoptysis (Chapter 83) from any cause of alveolar hemorrhage should be managed as needed. In the case of anticoagulation-, drug-, or toxin-related alveolar hemorrhage, the offending agent should be withdrawn, and supportive care is indicated. In Goodpasture syndrome, the ANCA-associated vasculitides, and other vasculitides (Chapter 270), treatment typically includes immunosuppressant agents such as corticosteroids (methylprednisolone, 500-2000 mg/day in divided doses for 3-5 days followed by a prednisone taper beginning at 1 mg/kg/day during the next 6 to 9 months) and cyclophosphamide (2 mg/kg/day orally or 0.5 g/m^2-0.75 g/m^2 intravenously for one dose) and then if needed, every 2 weeks or change to oral therapy. Plasmapheresis (3-14 exchanges) to remove the offending circulating antibody is a mainstay therapy for Goodpasture syndrome and may also be used in some cases of alveolar hemorrhage from ANCA-associated vasculitis and systemic lupus erythematosus. In patients with ANCA-associated vasculitis, rituximab (375 mg/m^2 once a week for 4 weeks) is at least as good as cyclophosphamide followed by azathioprine for 12 to 15 months for inducing and maintaining remission (≈50% at 12 months and about 40% at 18 months),[■] and rituximab (at 500 mg on days 0 and 14 and at 6, 12, and 18 months) is better than azathioprine for maintaining remission at 28 months.[■]

PROGNOSIS

Alveolar hemorrhage can result in acute respiratory failure and death. Recurrent alveolar hemorrhage from any cause, such as idiopathic pulmonary hemosiderosis, can be associated with the development of pulmonary fibrosis. Alveolar hemorrhage related to collagen vascular disease, vasculitides, and idiopathic pulmonary hemosiderosis can have mortality rates ranging from 25% to 50%. With Goodpasture syndrome, renal failure is common, and the degree of renal impairment may correlate with outcome.

INVASIVE MUCINOUS ADENOCARCINOMA AND LEPIDIC PREDOMINANT NONMUCINOUS ADENOCARCINOMA (FORMERLY CALLED BRONCHIOLOALVEOLAR CELL CARCINOMA)

DEFINITION

In 2011, a multidisciplinary committee eliminated the term *bronchioloalveolar cell carcinoma* and divided pulmonary adenocarcinomas into five types: adenocarcinoma in situ, minimally invasive adenocarcinoma, lepidic predominant nonmucinous adenocarcinoma, invasive mucinous adenocarcinoma, and invasive adenocarcinoma and its subtypes.[9] The types that would most likely be correlated with an alveolar filling appearance pathologically and on chest imaging are invasive mucinous adenocarcinoma, in which consolidation and air-bronchograms may be present, and lepidic-predominant nonmucinous adenocarcinoma, in which a ground-glass appearance is characteristic. Both of these types of adenocarcinoma are among those formally characterized as bronchioloalveolar cell carcinoma. In general, these tumors are characterized by malignant cells lining the alveolar cell wall (Chapter 191). Among bronchogenic carcinomas, these subtypes are the least associated with tobacco use, and patients with these cancers are more likely to be nonsmokers. Unlike other non–small cell lung cancers, the sex ratio approaches 1 : 1 or may be slightly female predominant, and younger patients may be affected.

PATHOBIOLOGY

These types of adenocarcinoma usually arise in the periphery of the lung and may be characterized by lepidic growth, which means contiguous growth along the intact alveolar septa, with varying degrees of stromal, pleural, vascular, or lymphatic invasion and without a known primary adenocarcinoma elsewhere. The mucinous type is thought to derive from respiratory goblet cells and columnar cells, and the nonmucinous type from type II pneumocytes or Clara cells.

CLINICAL MANIFESTATIONS

Patients present with a gradual onset of shortness of breath and cough. The duration of symptoms is usually several months. Constitutional symptoms such as malaise and weight loss may be present. Hemoptysis may occur. An unusual but unique clinical finding is bronchorrhea, with patients reporting the production of copious amounts of clear sputum daily. This finding is more common in the invasive mucinous form of the disease.

DIAGNOSIS

Radiographic patterns vary[10] and can include localized disease with peripheral solitary or multiple nodules or masses in 60% of cases or a persistent pneumonic pattern in 40% of cases (E-Fig. 91-E5). The radiographic findings of consolidation with air bronchograms are often initially thought to be consistent with acute pneumonia, but the typical clinical presentation is that of a nonresolving peripheral density on chest radiograph.[11] In addition, CT may show areas of ground-glass attenuation. Positron emission tomography may be normal because of the low glucose uptake of these tumors. The diagnosis

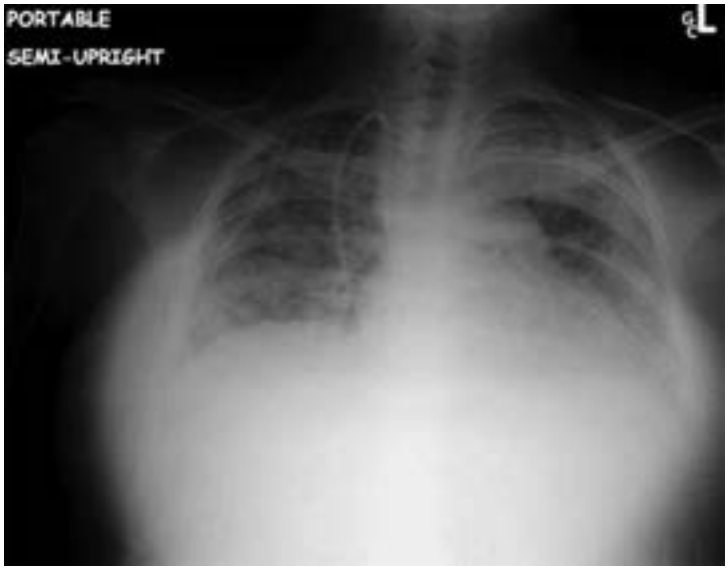

E-FIGURE 91-3. A chest radiograph showing bilateral alveolar opacities in a patient with diffuse alveolar hemorrhage.

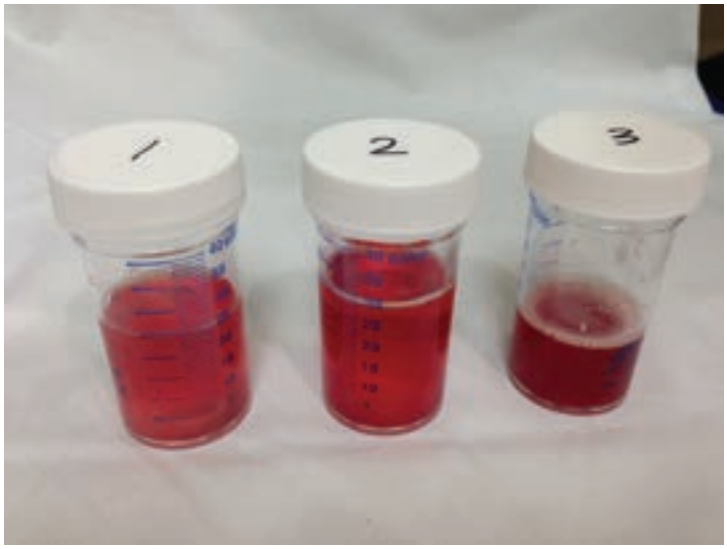

E-FIGURE 91-4. Bronchoalveolar lavage fluid revealing a progressively increasingly bloody lavage consistent with diffuse alveolar hemorrhage

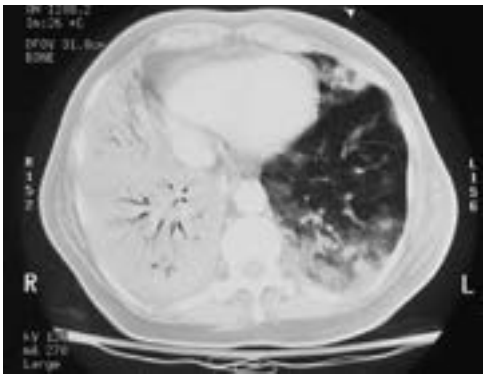

E-FIGURE 91-5. A chest computed tomography scan in a patient with invasive mucinous adenocarcinoma.

of invasive mucinous adenocarcinoma and lepidic predominant nonmucinous adenocarcinoma is most often made by bronchoscopy with transbronchial biopsy.

TREATMENT

Rx

For staging and treatment, these types of adenocarcinoma are approached like other types of non–small cell lung cancers (Chapter 191). Testing for epidermal growth factor receptor (EGFR) mutations should be performed, and chemotherapy planned accordingly. In general, the invasive mucinous adenocarcinomas are KRAS positive, and EGFR negative. The lepidic predominant nonmucinous type tends to be EGFR positive. Bilateral lung transplantation has been performed, but recurrence in the transplanted lungs has been reported.

PROGNOSIS

Prognosis correlates with disease stage and with the histologic and radiographic patterns. Patients who undergo surgical resection for adenocarcinoma in situ or minimally invasive adenocarcinoma with a single focus of disease have a better prognosis than patients with other adenocarcinomas of like stage, with the 5-year survival rate approaching 100%. More advanced forms likely have a prognosis similar to that of other adenocarcinomas.

Grade A References

A1. Specks U, Merkel PA, Seo P, et al. Efficacy of remission-induction regimens for ANCA-associated vasculitis. *N Engl J Med*. 2013;369:417-427.

A2. Guillevin L, Pagnoux C, Karras A, et al. Rituximab versus azathioprine for maintenance in ANCA-associated vasculitis. *N Engl J Med*. 2014;371:1771-1780.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ben-Dov I, Segel MJ. Autoimmune pulmonary alveolar proteinosis: clinical course and diagnostic criteria. *Autoimmun Rev*. 2014;13:513-517.
2. Punatar AD, Kusne S, Blair JE, et al. Opportunistic infections in patients with pulmonary alveolar proteinosis. *J Infect*. 2012;65:173-179.
3. Leth S, Bendstrup E, Vestergaard H, et al. Autoimmune pulmonary alveolar proteinosis: treatment options in year 2013. *Respirology*. 2013;18:82-91.
4. Khan A, Agarwal R, Aggarwal AN. Effectiveness of granulocyte-macrophage colony-stimulating factor therapy in autoimmune pulmonary alveolar proteinosis: a meta-analysis of observational studies. *Chest*. 2012;141:1273-1283.
5. Tazawa R, Inoue Y, Arai T, et al. Duration of benefit in patients with autoimmune pulmonary alveolar proteinosis after inhaled granulocyte-macrophage colony-stimulating factor therapy. *Chest*. 2014;145:729-737.
6. Mukhopadhyay S, Parambil JG. Acute interstitial pneumonia (AIP): relationship to Hamman-Rich syndrome, diffuse alveolar damage (DAD), and acute respiratory distress syndrome (ARDS). *Semin Respir Crit Care Med*. 2012;33:476-485.
7. West S, Arulkumaran N, Ind PW, et al. Diffuse alveolar haemorrhage in ANCA-associated vasculitis. *Intern Med*. 2013;52:5-13.
8. Lichtenberger JP 3rd, Digumarthy SR, Abbott GF, et al. Diffuse pulmonary hemorrhage: clues to the diagnosis. *Curr Probl Diagn Radiol*. 2014;43:128-139.
9. Travis WD, Brambilla E, Noguchi M, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society: international multidisciplinary classification of lung adenocarcinoma: executive summary. *Proc Am Thorac Soc*. 2011;8:381-385.
10. Austin JH, Garg K, Aberle D, et al. Radiologic implications of the 2011 classification of adenocarcinoma of the lung. *Radiology*. 2013;266:62-71.
11. Lee HJ, Lee CH, Jeong YJ, et al. IASLC/ATS/ERS International Multidisciplinary Classification of Lung Adenocarcinoma: novel concepts and radiologic implications. *J Thorac Imaging*. 2012;27:340-353.

REVIEW QUESTIONS

1. The correct treatment for patient with pulmonary alveolar proteinosis and hypoxemia is

A. corticosteroids.
B. lung transplantation.
C. plasmapheresis.
D. whole-lung lavage.
E. azathioprine.

Answer: D Patients with pulmonary alveolar proteinosis and significant clinical findings, such as dyspnea and hypoxemia, should be treated with whole-lung lavage, which is the current standard therapy.

2. Bronchorrhea is a finding sometimes associated with

A. diffuse alveolar hemorrhage.
B. pulmonary alveolar proteinosis.
C. invasive mucinous adenocarcinoma.
D. cardiogenic pulmonary edema.
E. ARDS.

Answer: C Bronchorrhea, which is the production of copious amounts of sputum that is usually clear and sometimes salty tasting, is seen in some cases of invasive mucinous adenocarcinoma.

3. Which of the following is best characterized radiographically as an alveolar filling disorder?

A. Idiopathic pulmonary fibrosis
B. Lymphangioleiomyomatosis
C. Acute interstitial pneumonitis
D. Lymphocytic interstitial pneumonitis
E. Langerhans cell histiocytosis

Answer: C Acute interstitial pneumonia is characterized radiographically as an alveolar filling process. The other options have an interstitial pattern on chest imaging.

4. The most common type of pulmonary alveolar proteinosis is

A. acquired or autoimmune.
B. congenital
C. associated with hematologic disease.
D. associated with metal dust exposure.
E. associated with respiratory infection.

Answer: A Autoimmune processes account for 90% of adult cases of pulmonary alveolar proteinosis, usually caused by GM-CSF antibodies. It also can be congenital or secondary to dust exposure or hematologic malignancy and is the most common form overall.

5. Plasmapheresis is part of the standard treatment for diffuse alveolar hemorrhage caused by which of the following diseases?

A. Idiopathic pulmonary hemosiderosis
B. Goodpasture syndrome
C. Cocaine inhalation
D. Autologous bone marrow transplantation
E. Pneumococcal pneumonia

Answer: B Plasmapheresis along with corticosteroids and cyclophosphamide is part of the standard treatment of diffuse alveolar hemorrhage related to Goodpasture syndrome.

92

INTERSTITIAL LUNG DISEASE

GANESH RAGHU

DEFINITION

In an apparently immunocompetent host, *interstitial lung disease* (ILD) is a clinical term for a heterogenous group of acute and chronic lower respiratory tract disorders with many potential causes. However, clinical and physiological features common to all ILDs include exertional dyspnea, a restrictive pattern on pulmonary function testing (Chapter 85), coexistent airflow obstruction, decreased diffusing capacity (D_{LCO}), increased alveolar-arterial oxygen difference (P_{AO_2}-P_{aO_2}) (Chapter 103) at rest or during exertion, and absence of pulmonary infection or neoplasm. ILDs comprise several acute and chronic lung disorders with variable degrees of pulmonary fibrosis (Table 92-1). The term *interstitial* is a misnomer because the pathologic processes are not restricted to the interstitium, which is the microscopic space bounded by the basement membranes of epithelial and endothelial cells. Rather, all of the several cellular and soluble constituents that make up the gas exchange units (alveolar wall, capillaries, alveolar space, and acini) and the bronchiolar lumen, terminal bronchioles, and pulmonary parenchyma beyond the gas exchange units (as well as the pleura and lymphatics and sometimes the lymph nodes) are involved in the pathogenesis and manifestations of ILD.

TABLE 92-1 CLINICAL CLASSIFICATION OF INTERSTITIAL LUNG DISEASE

IDIOPATHIC INTERSTITIAL PNEUMONIAS
Chronic Fibrosing Interstitial Pneumonias
Idiopathic pulmonary fibrosis Nonspecific interstitial pneumonia
Smoking-Related Interstitial Pneumonias
Respiratory bronchiolitis–interstitial lung disease Desquamative interstitial pneumonia
Acute or Subacute Idiopathic Interstitial Pneumonias
Cryptogenic organizing pneumonia Acute interstitial pneumonia Lymphoid and lymphocytic interstitial pneumonia
Rare interstitial Pneumonias
Histologic pattern of acute fibrinous and organizing pneumonia Histological pattern of interstitial pneumonias with a bronchiolocentric distribution Pleuroparenchymal fibroelastosis
INTERSTITIAL LUNG DISEASE ASSOCIATED WITH CONNECTIVE TISSUE DISEASE
Progressive systemic sclerosis Rheumatoid arthritis Systemic lupus erythematosus Dermatomyositis and polymyositis Sjögren syndrome Mixed connective tissue disease Ankylosing spondylitis
HYPERSENSITIVITY PNEUMONITIS
Occupational and environmental factors (e.g., farmer's lung; bird fancier's lung) Iatrogenic
DRUG-INDUCED AND IATROGENIC INTERSTITIAL LUNG DISEASE
See Table 92-2
ALVEOLAR FILLING DISORDERS (Chapter 91)
Goodpasture syndrome Pulmonary alveolar proteinosis Pulmonary hemosiderosis Alveolar hemorrhage syndromes Chronic eosinophilic pneumonia
INTERSTITIAL LUNG DISEASE ASSOCIATED WITH PULMONARY VASCULITIS
Pulmonary capillaritis Granulomatosis with polyangiitis (formerly known as Wegener granulomatosis) Churg-Strauss syndrome
OTHER SPECIFIC FORMS OF INTERSTITIAL LUNG DISEASE
Sarcoidosis Langerhans cell histiocytosis (histiocytosis X) Lymphangioleiomyomatosis
INHERITED FORMS OF INTERSTITIAL LUNG DISEASE
Familial idiopathic pulmonary fibrosis Familial pulmonary fibrosis or interstitial pneumonia Tuberous sclerosis Neurofibromatosis Gaucher disease Niemann-Pick disease Hermansky-Pudlak syndrome

EPIDEMIOLOGY

Among persons 18 years or older, the prevalence of all ILDs in the United States is about 81 per 100,000 men and 67 per 100,000 women. The overall incidence is also higher in men (31.5 per 100,000 per year) than in women (26.1 per 100,000 per year). Moreover, the prevalence of undiagnosed or early ILD is estimated to be 10 times that of clinically recognized disease; as physicians' awareness of these entities increases, it is expected that the frequency of the diagnosis of ILD will rise. Among the ILDs, the most common is idiopathic pulmonary fibrosis, which represents at least 30% of incident cases. In the United States, the annual incidence of idiopathic pulmonary fibrosis is 94 per 100,000 in the Medicare population, with a mean age of onset of 79 years.[1] Recent data suggest that more than 5000 new cases are diagnosed each year in the United Kingdom.

PATHOBIOLOGY

Interstitial lung diseases are thought to result from an unknown tissue injury and attempted repair in the lung of a genetically predisposed person. Genetic variants within the *hTERT* or *hTR* components of the telomerase gene and surfactant protein gene have been associated in a subset of familial pulmonary fibrosis and in some sporadic cases.[2] An *MUC5B* promotor polymorphism is

associated with familial interstitial pneumonia and idiopathic pulmonary fibrosis.[3]

In *idiopathic pulmonary fibrosis,* varying degrees of acute, subacute, and chronic fibroproliferation are present in the lungs at the time of diagnosis. Ultimately, progressive fibrosis results in honeycombing, an end-stage finding that is often associated with increased pulmonary vascular resistance and secondary pulmonary hypertension. As a reflection of these dynamic processes, histopathologic examination of lung tissue often reveals highly heterogeneous findings; for example, a single biopsy specimen may show normal alveoli adjacent to abnormal areas of inflammation and fibrosis, with or without granulomas, vasculitis, or secondary vascular changes within the pulmonary parenchyma.

CLINICAL MANIFESTATIONS

Interstitial lung diseases are typically characterized by progressive dyspnea. Nonproductive cough and fatigue are also common complaints. Pleuritic chest pain may occur with certain connective tissue or drug-induced ILDs, and acute pleuritic chest pain with dyspnea may represent a spontaneous pneumothorax (Chapter 99) in association with lymphangioleiomyomatosis, tuberous sclerosis (Chapter 417), neurofibromatosis, or Langerhans cell histiocytosis. Hemoptysis suggests a diffuse alveolar hemorrhagic syndrome, systemic lupus erythematosus (SLE) (Chapter 266), lymphangioleiomyomatosis, granulomatosis with polyangiitis (Chapter 270), or Goodpasture syndrome (Chapter 121); it is rare in other ILDs. In patients with existing ILD, new hemoptysis should prompt consideration of a superimposed malignancy, pulmonary embolus, or infection such as aspergillosis.

In some patients, the first and the only clue to the presence of an ILD may be the finding of coarse rales (crackles) on auscultation of the lungs. These coarse crackles must be distinguished from the finer rales typical of heart failure (Chapter 58) or noncardiogenic pulmonary edema (Chapter 104). Unlike patients with obstructive lung disease, wheezes are not common. A history of wheezing suggests the coexistence of occult hyperactive airways and airflow obstruction and raises the possibility of allergic bronchopulmonary aspergillosis (Chapter 339), Churg-Strauss syndrome (Chapter 270), chronic eosinophilic pneumonia (see later), or parasitic infection (Chapter 344). In some patients, the initial presentation may be with peripheral cyanosis, clubbing, or the signs and symptoms of an underlying systemic disease (see later).

DIAGNOSIS

The first key in patients with an ILD is to establish the syndromic diagnosis and then pursue the differential diagnosis of its specific cause (Fig. 92-1). However, a conclusive cause often may not be identified despite an exhaustive medical history and invasive diagnostic interventions, including bronchoalveolar lavage (BAL)[4] and sufficiently large and multiple lung biopsy specimens. Thus, the cause of several of the ILDs, even when diagnosed as specific entities, remains unknown.

History

The patient's age, sex, and cigarette smoking history may provide useful clues to the diagnosis. Idiopathic pulmonary fibrosis is an adult disorder that usually occurs in patients older than 50 years. Pulmonary sarcoidosis (Chapter 95), in contrast, is more common in young adults and middle-aged persons. Whereas pulmonary Langerhans cell histiocytosis (previously known as pulmonary histiocytosis X or eosinophilic granuloma) characteristically occurs in young cigarette-smoking men, lymphangioleiomyomatosis occurs exclusively in women of childbearing age. Respiratory bronchiolitis–associated ILD is seen almost exclusively in cigarette smokers but occurs in both men and women of all ages.

The medical history also should focus on environmental factors, especially changes in environmental exposures (including domestic, recreational, hot tub, whirlpool baths, indoor swimming pool, ventilation system at home, automobiles, and workplaces), occupational exposure, medications, and drug use (Chapters 93 and 94). A family medical history should address possible familial ILD. Environmental risk factors that may suggest the diagnosis of hypersensitivity pneumonitis include farming or exposure to overt or occult avian antigens at or in the close vicinity of home, bird droppings, feather duvets ("bird fancier's lung" or "pigeon breeder's lung"), visible molds, unkempt dusty homes, unchanged filters in furnaces, unkempt ventilatory systems, water leaks, or humidifiers in the domestic environment (hypersensitivity to thermophilic actinomycetes, *Aureobasidium pullulans*). At-risk occupations include mining (pneumoconioses), machine tool grinding, sandblasting and working with granite (silicosis), welding and working in a shipyard (asbestosis), and working in the aerospace or electronic industries (berylliosis) (Chapters 93 and 94). Because of the long interval between the exposure and the onset of symptoms in many occupations associated with ILD, it is important to take a lifelong occupational history

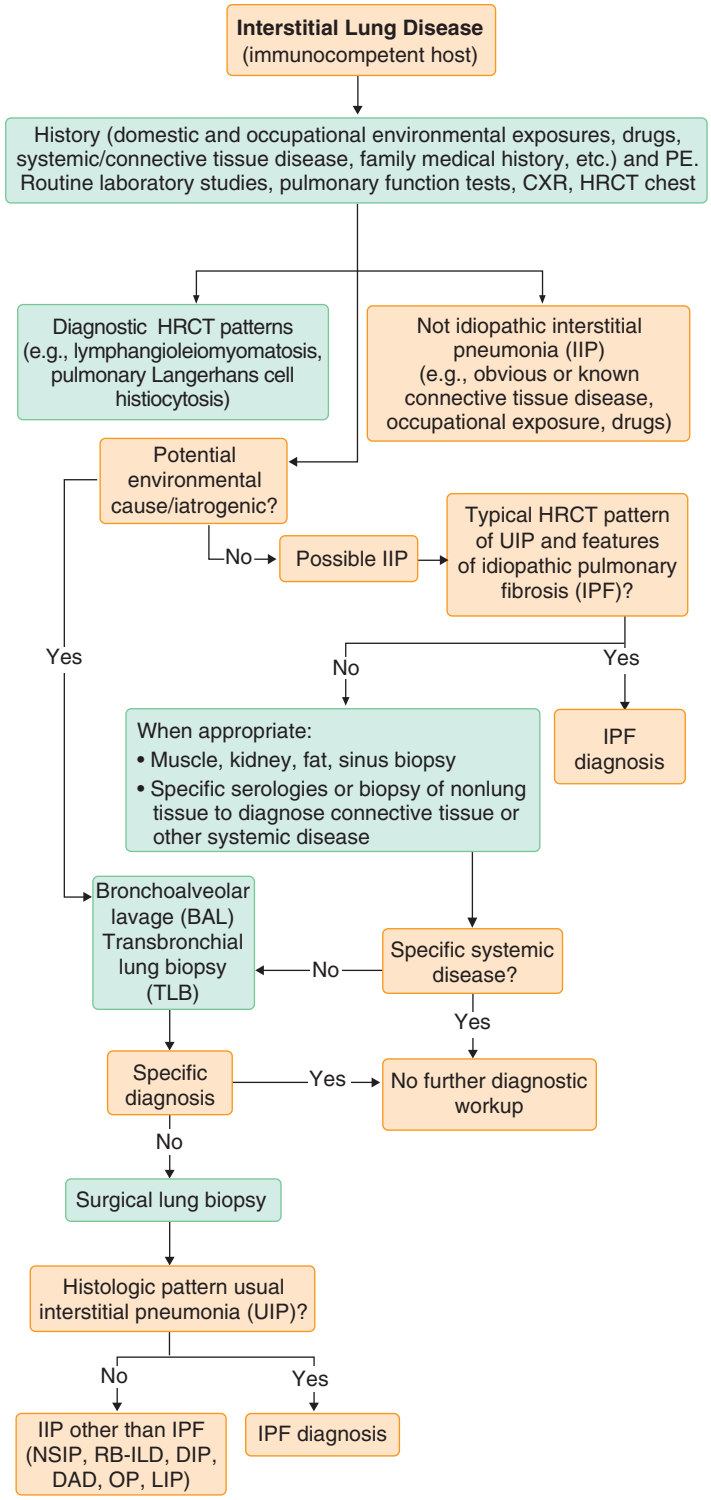

FIGURE 92-1. An approach to interstitial lung disease. DAD = diffuse alveolar damage; DIP = desquamative interstitial pneumonia; HRCT = high-resolution computed tomography; IIP = idiopathic interstitial pneumonia; IPF = idiopathic pulmonary fibrosis; LIP = lymphoid interstitial pneumonia; NSIP = nonspecific interstitial pneumonia; OP = organizing pneumonia; PE = physical examination; PFT = pulmonary function test; RB-ILD = respiratory bronchiolitis–associated interstitial lung disease; TLB = transbronchial lung biopsy; UIP = usual interstitial pneumonia. (Adapted from American Thoracic Society/European Respiratory Society: International multidisciplinary consensus classification of idiopathic interstitial pneumonias. *Am J Respir Crit Care Med.* 2002;165:277-304 and Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med.* 2011;183:788-824.)

TABLE 92-2 DRUG-INDUCED AND IATROGENIC INTERSTITIAL LUNG DISEASE*

ANTIMICROBIAL AGENTS
Cephalosporins
Isoniazid
Nitrofurantoin
Penicillins
Sulfonamides
ANTI-INFLAMMATORY AGENTS
Aspirin
Gold
Methotrexate
Nonsteroidal anti-inflammatory agents
Penicillamine
Phenylbutazone
Zafirlukast
CARDIOVASCULAR DRUGS
Amiodarone
Angiotensin-converting enzyme inhibitors
β-Blockers
Hydralazine
Hydrochlorothiazide
Procainamide
Protamine sulfate
Tocainide
ANTINEOPLASTIC AND CHEMOTHERAPEUTIC AGENTS
Bleomycin
Busulfan
Chlorambucil
Cyclophosphamide
Erlotinib
Gefitinib
Gemcitabine
Imatinib
Melphalan
Mercaptopurine
Methotrexate
Mitomycin
Nitrosoureas
Procarbazine
CENTRAL NERVOUS SYSTEM DRUGS
Carbamazepine
Chlorpromazine
Imipramine
Phenytoin
ORAL HYPOGLYCEMIC AGENTS
Chlorpropamide
Tolazamide
Tolbutamide
ILLICIT DRUGS
Cocaine
Heroin
Methadone
Propoxyphene
OTHER AGENTS
Antithymocyte globulin
All-*trans*-retinoic acid
Colony-stimulating factors
Interferon-α and -β
Irradiation
Mycophenolate mofetil
Tumor necrosis factor-α modulating agents
High fraction of inspired oxygen (F_{IO_2}) with mechanical ventilation

*This list contains examples only and is not meant to be exhaustive.

(Chapter 19) as well as to establish the interval between exposure and the onset of symptoms. Because the list of medications known to cause ILD is long and continues to grow (Table 92-2), a careful history regarding recent use of prescription and over-the-counter products is essential. Risk factors for immunosuppression, including infection with human immunodeficiency virus, raise the possibility of opportunistic lung infections (Chapter 391), neoplasm (Chapter 191), and transplant-related pulmonary complications.

Particular attention should be paid to the onset and duration of symptoms; the rate of disease progression; and association with hemoptysis, fever, or extrathoracic symptoms. Symptoms that persist 4 weeks or less and the presence of fever suggest cryptogenic organizing pneumonia, drug-induced pulmonary injury, or acute hypersensitivity pneumonitis, BUT idiopathic pulmonary fibrosis, ILD associated with connective tissue diseases, and Langerhans cell histiocytosis tend to have a more subacute onset. Extrapulmonary symptoms suggest that the ILD may be associated with systemic disorders (e.g., sarcoidosis; Chapter 95), and symptoms such as dysphagia, dry eyes or mouth, skin rashes, or arthritis may suggest a connective tissue disorder (Chapters 266 to 270). Proximal muscle aches or weakness suggests the possibility of polymyositis or dermatomyositis (Chapter 269), and recurrent sinusitis suggests granulomatosis with polyangiitis (Chapter 270). Extrathoracic manifestations present in tuberous sclerosis (Chapter 417) include hematuria, epilepsy, and mental retardation.

Physical Examination

Physical examination of the respiratory system is rarely helpful in the diagnostic evaluation of ILD because findings such rhonchi and rales on auscultation and digital clubbing are nonspecific. Findings on cardiac examination, such as an accentuated P_2, a right ventricular heave, or tricuspid insufficiency, are suggestive of pulmonary hypertension (Chapter 68) and cor pulmonale in patients with advanced lung disease. However, extrathoracic findings such as skin abnormalities, peripheral lymphadenopathy, and hepatosplenomegaly may be more specifically associated with underlying sarcoidosis (Chapter 95); muscle tenderness and proximal muscle weakness may point to coexisting polymyositis (Chapter 269); and signs of arthritis may indicate connective tissue disease (Chapters 264, 266, and 270) or sarcoidosis (Chapter 95). Characteristic rashes occur in several connective tissue diseases, disseminated Langerhans cell histiocytosis, tuberous sclerosis, and neurofibromatosis. Ophthalmologic findings (Chapter 423) such as iridocyclitis, uveitis, or conjunctivitis, may be a clue to the diagnosis of sarcoidosis or a connective tissue disease, AND central nervous system abnormalities may be present in sarcoidosis, SLE, Langerhans cell histiocytosis, or tuberous sclerosis.

Laboratory Testing

Routine laboratory testing should include a complete blood count, leukocyte differential, erythrocyte sedimentation rate, chemistry panel (calcium, liver enzymes, electrolytes, creatinine), and urinalysis. Although these data rarely yield a specific diagnosis, they may provide helpful clues. Routine serology for occult connective tissue diseases (Chapter 257) may reveal typical findings of SLE (e.g., antinuclear antibodies), rheumatoid arthritis (rheumatoid factor, anticitrullinated peptide antibody), scleroderma (ScL 70), dermatomyositis or polymyositis (creatine kinase, aldolase, and anti–Jo-1 antibody), granulomatosis with polyangiitis (antineutrophil cytoplasmic antibodies), and Goodpasture syndrome (anti–basement membrane antibodies).

Mild hypoxemia is typically present on arterial blood gas analysis because of abnormal ventilation-perfusion ratios, especially in moderate to severe cases of ILD. However, carbon dioxide retention is rare and suggests possible coexisting emphysema (Chapter 88) or a hypoventilatory disorder (Chapter 86).

Noninvasive Evaluation

Chest Radiograph

The distribution and appearance of radiographic abnormalities (Chapter 84) may prove useful in differentiating the clinicopathologic syndromes in patients with ILD (Table 92-3). Comparison of previous chest radiographs with the current one is important in establishing the rate of progression of the patient's disease. A diffuse ground-glass pattern is often observed early in the course of ILD followed by progression to reticular (linear) infiltrates with nodules (reticulonodular infiltrates) or, in the case of alveolar filling disorders, ill-defined nodules (acinar rosettes) with air bronchograms. Most ILDs cause infiltrates in the lower lung zones, but upper lobe predominance is typically present in sarcoidosis, berylliosis, Langerhans cell histiocytosis, silicosis, chronic hypersensitivity pneumonitis, cystic fibrosis, and ankylosing spondylitis. The middle and lower lung zones show the most prominent abnormalities in lymphangitic carcinomatosis, idiopathic pulmonary fibrosis, subacute eosinophilic pneumonia, asbestosis, and pulmonary fibrosis caused by rheumatoid arthritis or progressive systemic sclerosis. Hilar adenopathy and mediastinal adenopathy are not common in ILDs; their presence should suggest sarcoidosis, berylliosis, silicosis, lymphocytic interstitial pneumonia (LIP), amyloidosis, or Gaucher disease. A pattern of peripherally located

TABLE 92-3 CHARACTERISTIC CHEST RADIOGRAPHIC PATTERNS IN PATIENTS WITH INTERSTITIAL LUNG DISEASE

PATTERN	SUGGESTED DIAGNOSES*
Decreased lung volumes	Idiopathic pulmonary fibrosis, nonspecific interstitial pneumonia, desquamative interstitial pneumonia, connective tissue disease, chronic eosinophilic pneumonia, asbestosis, chronic hypersensitivity pneumonitis, or drug-induced interstitial lung disease (ILD)
Increased or preserved lung volumes	Idiopathic pulmonary fibrosis with emphysema, respiratory bronchiolitis–associated ILD, cryptogenic organizing pneumonia, hypersensitivity pneumonitis, lymphangioleiomyomatosis, Langerhans cell histiocytosis, sarcoidosis, neurofibromatosis, tuberous sclerosis
Micronodules	Infection, hypersensitivity pneumonitis, sarcoidosis, respiratory bronchiolitis–associated ILD
Septal thickening	Malignancy, infection, chronic congestive heart failure, pulmonary veno-occlusive disease
Honeycombing	Idiopathic pulmonary fibrosis, fibrotic nonspecific interstitial pneumonia, connective tissue disease, asbestosis, chronic hypersensitivity pneumonitis, sarcoidosis
Recurrent infiltrates	Cryptogenic organizing pneumonia, chronic eosinophilic pneumonia, drug- or radiation-induced ILD
Migratory or fleeting infiltrates	Cryptogenic organizing pneumonia, hypersensitivity pneumonitis, Churg-Strauss syndrome, Löffler syndrome, allergic bronchopulmonary aspergillosis
Pleural disease	Connective tissue disease, asbestosis, malignancy, radiation-induced ILD, amyloidosis, sarcoidosis, lymphangioleiomyomatosis, nitrofurantoin-induced ILD
Pneumothorax	Langerhans cell histiocytosis, lymphangioleiomyomatosis, tuberous sclerosis, neurofibromatosis
Mediastinal or hilar adenopathy	Lymphocytic interstitial pneumonia, connective tissue disease, silicosis, chronic berylliosis, malignancy, infection, sarcoidosis, amyloidosis, Gaucher disease
Normal (rare)	Cellular nonspecific interstitial pneumonia, respiratory bronchiolitis–associated interstitial lung disease, connective tissue disease, hypersensitivity pneumonitis, sarcoidosis
LOCATION OF RADIOGRAPHIC ABNORMALITY	**SUGGESTED DIAGNOSES***
Mid to upper lung zone	Hypersensitivity pneumonitis, chronic berylliosis, ankylosing spondylitis, silicosis, Langerhans cell histiocytosis, sarcoidosis, pleuroparenchymal fibroelastosis, cystic fibrosis
Lower lung zone	Idiopathic pulmonary fibrosis, nonspecific interstitial pneumonia (fibrotic), connective tissue disease, asbestosis, chronic hypersensitivity pneumonitis
Peripheral	Idiopathic pulmonary fibrosis, nonspecific interstitial pneumonia (fibrotic), cryptogenic organizing pneumonia, chronic eosinophilic pneumonia

*This list is not intended to be comprehensive.
Adapted from Raghu G, Brown K. Clinical issues: patient evaluation. In: Baughman RP, du Bois RM, eds. *Diffuse Lung Disease: A Practical Approach*. New York: Oxford University Press; 2004.

TABLE 92-4 RADIOGRAPHIC FEATURES OF IDIOPATHIC INTERSTITIAL PNEUMONIAS

CLINICAL DIAGNOSIS	USUAL RADIOGRAPHIC FEATURES	TYPICAL FINDINGS ON HRCT
Idiopathic pulmonary fibrosis	Basal-predominant reticulation abnormality with volume loss	Pattern of usual interstitial pneumonia; peripheral, basal, subpleural reticulation; honeycombing, traction bronchiectasis
Nonspecific interstitial pneumonia	Ground-glass and reticular opacification	Peripheral, basal, subpleural, symmetrical ground-glass attenuation with irregular lines and consolidation; subpleural sparing
Cryptogenic organizing pneumonia	Patchy bilateral consolidation	Subpleural or peribronchial patchy consolidation or nodules
Acute interstitial pneumonia	Diffuse ground-glass density or consolidation	Diffuse consolidation and ground-glass opacification, often with lobular sparing and late traction bronchiectasis
Desquamative interstitial pneumonia	Ground-glass opacity	Peripheral, lower lung zone ground-glass attenuation with reticulation and/or small cysts
Respiratory bronchiolitis–associated interstitial lung disease	Bronchial wall thickening, ground-glass opacification	Diffuse bronchial wall thickening with poorly defined centrilobular nodules and patchy ground-glass opacification
Lymphocytic interstitial pneumonia	Reticular opacities and nodules	Diffuse centrilobular nodules, ground-glass attenuation, septal and bronchovascular wall thickening, and thin-walled cysts

HRCT = high-resolution computed tomography.
Adapted from American Thoracic Society/European Respiratory Society. International multidisciplinary revised classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med*. 2002;165:277-304; Travis WD, Costabel U, Hansell DM, et al. An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med*. 2013;188:733-748; and Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med*. 2011;183:788-824.

pulmonary infiltrates in the upper and middle lung zones with relatively clear perihilar and central zones is a clue to chronic eosinophilic pneumonia. Recurrent infiltrates raise the possibility of cryptogenic organizing pneumonia, chronic eosinophilic pneumonia, or drug- or radiation-induced pneumonitis, and fleeting or migratory infiltrates may occur in Churg-Strauss syndrome (allergic angiitis), allergic bronchopulmonary aspergillosis, tropical eosinophilic pneumonia, or Löffler syndrome. Whereas localized pleural plaques may indicate asbestosis, diffuse pleural thickening can result from asbestosis, rheumatoid arthritis, progressive systemic sclerosis, radiation pneumonitis, nitrofurantoin, or malignancy. In the absence of left ventricular failure, the presence of a pleural effusion (Chapter 99) raises the possibility of rheumatoid arthritis, SLE, acute hypersensitivity pneumonitis, sarcoidosis, asbestosis, amyloidosis, lymphangioleiomyomatosis, or lymphangitic carcinomatosis. A reduction of lung volumes is typical in most ILDs; the presence of preserved lung volumes or hyperinflation should raise suspicion for chronic hypersensitivity pneumonitis, Langerhans cell histiocytosis, lymphangioleiomyomatosis, neurofibromatosis, sarcoidosis, or tuberous sclerosis. However, plain chest radiographs may be normal in about 10% of patients with ILD.

High-Resolution Computed Tomography

Because of its increased sensitivity and ability to distinguish ground-glass changes, which are generally considered to be reversible areas of lung disease, from irreversible fibrotic and honeycomb changes, high-resolution computed tomography (HRCT) is essential in both the diagnosis and staging of ILD. Although microscopic ILD cannot be excluded by a normal HRCT result, HRCT allows recognition of abnormalities not apparent in plain chest radiographs and may lead to an earlier diagnosis, help narrow the differential diagnosis patterns (Table 92-4), aid in selecting the site or sites for BAL and lung biopsy, and assist in choosing among therapeutic options and in estimating the response to treatment. Whereas normal HRCT excludes the diagnosis of pulmonary fibrosis, the presence of patchy subpleural reticular and basilar septal fibrosis, traction bronchiectasis, and honeycombing increases the level

of diagnostic confidence for the pattern of usual interstitial pattern, which is characteristic of idiopathic pulmonary fibrosis. The finding of bilateral cysts, including their size, configuration, distribution, and appearance, helps differentiate among lymphangioleiomyomatosis, tuberous sclerosis, and pulmonary Langerhans cell histiocytosis. HRCT can detect ILD despite normal chest radiographs in patients with asbestosis, silicosis, sarcoidosis, and scleroderma. Patients with respiratory bronchiolitis–associated ILD typically have patchy ground-glass attenuation on HRCT in concert with bilateral interstitial prominence, fine nodular radiographic infiltrates, and normal lung volumes. Images obtained in the supine and prone positions and on deep inspiration and exhalation sometimes help to differentiate fibrosis from atelectasis.

Pulmonary Function Tests

The most characteristic physiologic abnormalities in patients with ILD, regardless of etiology, are a restrictive lung defect and decreased DLCO (see Table 85-2 in Chapter 85). Forced expiratory volume in 1 second (FEV_1) and forced vital capacity (FVC) are decreased proportionally such that the ratio of the two remains normal or may even be increased. Both total lung capacity (TLC) and lung volumes measured by body plethysmography are reduced. Pulmonary function tests (PFTs) may be useful in monitoring the progression of disease and prognosis; significant changes in FVC, DLCO (corrected to hemoglobin), and physiological measurements (FVC, DLCO) at 1 year portend a worse survival in patients with idiopathic pulmonary fibrosis.

Certain PFT findings may also aid in the differential diagnosis. A mixed obstructive–restrictive pattern occurs in patients with Churg-Strauss syndrome, allergic bronchopulmonary aspergillosis, endobronchial sarcoidosis, hypersensitivity pneumonitis, cryptogenic organizing pneumonia, tropical pulmonary interstitial eosinophilia, coexisting chronic obstructive pulmonary disease or asthma, or secondary bronchiectasis. Diseases associated with respiratory muscle weakness, such as polymyositis, progressive systemic sclerosis, and SLE, may exhibit a decrease in maximal voluntary ventilation and increased residual volume out of proportion to the decrease in FEV_1.

Exercise Testing

The magnitude of the increase in P_{AO_2}-P_{aO_2} on exercise correlates well with the severity of disease and the degree of pulmonary fibrosis in patients with idiopathic pulmonary fibrosis. Other exercise-induced physiologic abnormalities in ILD include a decrease in work rate and maximal oxygen consumption, abnormally high minute ventilation at submaximal work rates, decreased peak minute ventilation, and failure of tidal volumes to increase at submaximal levels of work while the respiratory rate increases disproportionately. The 6-minute walk test, performed on a flat surface, can provide quantitative data on exercise capacity and on oxygen desaturation with exercise and can justify use of supplemental oxygen based on clinical and physiological needs.

Invasive Evaluation

A collegial interaction and multidisciplinary discussions among the pulmonary clinician, chest radiologist, thoracic surgeon, and pathologist can help determine the best diagnostic approach for an individual patient (see Fig. 92-1).

Findings on BAL can be diagnostic in some patients with ILD and can narrow the differential diagnosis in others (Chapter 85). For example, a lymphocyte-predominant cellular pattern raises the possibility of sarcoidosis or hypersensitivity pneumonitis in the appropriate clinical setting. Eosinophils are seen in pulmonary Langerhans cell granulomatosis, an asbestos body count greater than 1 fiber per milliliter of BAL fluid is seen in asbestosis, and specially staining surfactant material is seen in pulmonary alveolar proteinosis. A transbronchial lung biopsy may reveal noncaseating granulomas in sarcoidosis, "loose" noncaseating granulomas in hypersensitivity pneumonitis, giant cell granulomas in hard metal pneumoconiosis, or smooth muscle proliferation in lymphangioleiomyomatosis. However, failure to establish a diagnosis on BAL and transbronchial lung biopsy does not exclude these entities.

Video-assisted thoracoscopic biopsy (Chapter 101) or open lung biopsy may be required to obtain an adequate sample for histologic evaluation of a patient with unexplained signs and symptoms when other studies have failed to establish a diagnosis, but most patients with idiopathic pulmonary fibrosis do not need to have a biopsy to confirm the diagnosis. The mortality rate for the procedure is generally less than 1%, and the morbidity rate is less than 3%.

TABLE 92-5 INTERSTITIAL LUNG DISEASE: CLINICAL RESPONSE TO SYSTEMIC CORTICOSTEROIDS ALONE*

GENERALLY RESPONSIVE	UNRESPONSIVE†
Sarcoidosis	Idiopathic interstitial pneumonia
Acute hypersensitivity pneumonitis	Idiopathic pulmonary fibrosis (usual interstitial pneumonia)
Drug induced	Desquamative interstitial pneumonia (subset)
Environmental causes (some)	Chronic secondary and advanced pulmonary fibrosis
Idiopathic interstitial pneumonia	Chronic hypersensitivity pneumonitis (subset)
Cryptogenic organizing pneumonia	Chronic radiation fibrosis
Nonspecific interstitial pneumonia (cellular)	Cryptogenic organizing pneumonia (subset)
Respiratory bronchiolitis–associated ILD	Acute interstitial pneumonia (?)
Lymphocytic interstitial pneumonia	Chronic pulmonary hemorrhage syndromes
Desquamative interstitial pneumonia (subset)	Pulmonary veno-occlusive disease
Acute interstitial pneumonia (?)	Environmental (e.g., asbestosis, pneumoconiosis)
Acute pulmonary capillaritis	End-stage ILDs, pulmonary fibrosis coexisting or associated with pulmonary hypertension
Eosinophilic pneumonia (acute and chronic)	Pulmonary Langerhans cell granulomatosis
Acute radiation pneumonitis‡	Lymphangioleiomyomatosis
Organizing pneumonia associated with connective tissue diseases	ILD in inherited disorders (?)

*The dosage plus duration of corticosteroids used is variable and based on anecdotal experience, individual expert opinion, clinical judgment, and response as judged by objective measurements (clinical, radiologic, or physiologic). Oral prednisone or prednisolone is the most common corticosteroid used. Most patients who respond during the first few weeks of 20 to 60 mg of prednisone per day require maintenance low-dose oral prednisone at 5 to 10 mg/day beyond 6 months. Some patients who require maintenance of oral prednisone doses higher than 20 mg/day beyond 4 to 6 months may tolerate lower doses of prednisone if other immune-modulating agents (e.g., azathioprine, mycophenolate) are used in combination. There is no evidence to recommend a specific regimen. Patients should be monitored carefully and regularly for known side effects of corticosteroid use (e.g., osteoporosis, glucose intolerance), and preventive and therapeutic measures must be undertaken appropriately.
†Some patients unresponsive to oral corticosteroids alone may respond to combined treatment with corticosteroids and other immune-modulating drugs (e.g., azathioprine, mycophenolate).
‡Although most patients respond to modest doses of oral prednisone (initially, 40-60 mg/day), it is important to taper the prednisone very slowly to reach a maintenance dose of 5 to 10 mg/day beyond 6 months; rapid taper of oral prednisone has been associated with "rebound," which is an exaggerated lung injury beyond the irradiated segment of the lung and in the contralateral lung.
ILD = interstitial lung disease.

TREATMENT

When the cause of the ILD is clearly known (e.g., acute or subacute hypersensitivity pneumonitis, occupational ILD, iatrogenic), further avoidance of the inciting agent or agents is essential (Chapter 93). Although systemic corticosteroids are generally indicated and are associated with a favorable response in some ILDs, the dosage and duration are unclear and essentially based on anecdotal experience (Table 92-5).

Supportive oxygen supplementation is dictated by clinical needs. For selected patients with end-stage ILDs, such as those associated with significant pulmonary fibrosis and pulmonary hypertension, lung transplantation (Chapter 101) may be a feasible and viable option. Treatments for pulmonary hypertension associated with ILDs (Chapter 68) are indicated in patients with connective tissue diseases, but their clinical benefit for patients with other ILDs have been disappointing.

SPECIFIC TYPES OF INTERSTITIAL LUNG DISEASE

Idiopathic Interstitial Pneumonias

Idiopathic interstitial pneumonias, which are a subset of acute or chronic ILDs of unknown etiology, are characterized by the presence of varying degrees of interstitial and alveolar inflammation and fibrosis. Distinct clinicopathologic forms of idiopathic interstitial pneumonia include chronic

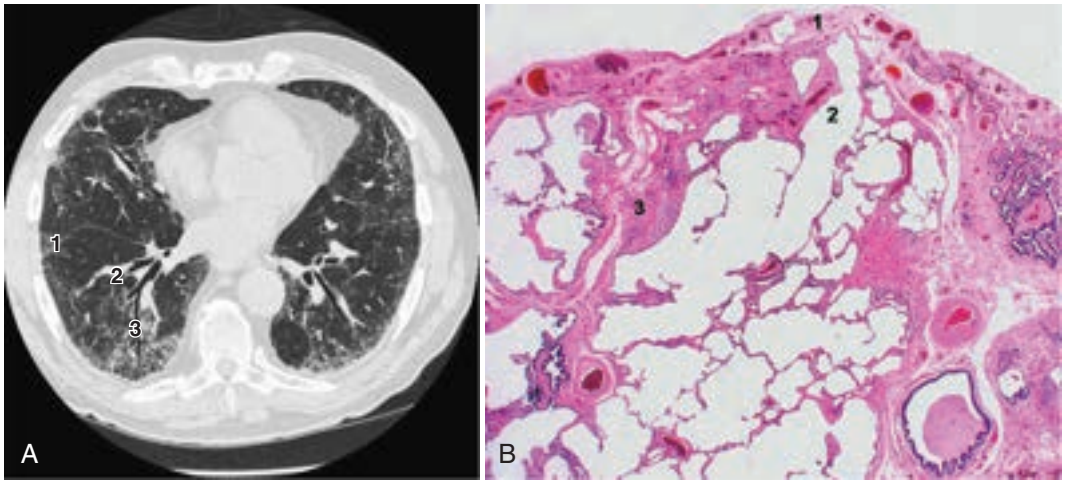

FIGURE 92-2. Diagnosis of idiopathic pulmonary fibrosis. A, The usual interstitial pneumonia pattern of idiopathic pulmonary fibrosis in the lower lobes on high-resolution computed tomography consists of (1) subpleural fibrotic changes with (2) traction bronchiectasis and (3) honeycomb cysts in the lower lobes. **B,** Usual interstitial pneumonia pattern of idiopathic pulmonary fibrosis. Note the presence of (1) subpleural fibrosis with (2) traction emphysema, (3) fibroblastic foci, and temporal heterogeneity of microscopic abnormalities at low magnification. (Courtesy of Dr. Kevin Leslie.)

fibrosing interstitial pneumonias (idiopathic pulmonary fibrosis and nonspecific interstitial pneumonia), smoking-related interstitial pneumonias (respiratory bronchiolitis–ILD and desquamative interstitial pneumonia), and acute or subacute idiopathic interstitial pneumonias (cryptogenic organizing pneumonia and acute interstitial pneumonia). Rare histologic patterns of acute fibrinous and organizing pneumonia and interstitial pneumonias with a bronchiolocentric distribution and pleuroparenchymal fibroelastosis have been recently recognized.[5] In some patients, mixed histopathologic features are evident in different segments of the same lung. When the distinct pathological forms are not evident, the diagnosis of unclassifiable interstitial pneumonia has been recently recognized. The accuracy of the diagnosis of idiopathic interstitial pneumonias is increased by multidisciplinary discussions among expert pulmonologists, radiologists, and pathologists familiar with interstitial lung diseases and idiopathic interstitial pneumonias.

Although the clinical severity may vary, the idiopathic interstitial pneumonias tend to manifest as an insidious onset of exertional dyspnea and a nonproductive cough. Chest pain and systemic symptoms such as weight loss and fatigue may be present. Bibasilar end-inspiratory crackles are often heard on auscultation. Clubbing, although not specific, is found in 25% to 50% of patients with idiopathic pulmonary fibrosis. Findings on the chest radiograph are most often nonspecific, and the presence of normal lung markings on the chest radiograph does not exclude ILD. On HRCT, many pathologic entities have characteristic image patterns that have greatly aided diagnosis (see Table 92-4). The clinical course of idiopathic pulmonary fibrosis is heterogeneous.

CHRONIC FIBROSING INTERSTITIAL PNEUMONIA

Idiopathic Pulmonary Fibrosis

Idiopathic pulmonary fibrosis accounts for 50% to 60% of all idiopathic interstitial pneumonias. Idiopathic pulmonary fibrosis occurs in adult men and women with a mean age at onset of 62 years. Some patients have familial disease, likely as an autosomal dominant with variable penetrance. The best validated genetic risk factor is a polymorphism in the promoter of the gene encoding mucin-5B (*MUCSB*), which is associated with both familial and sporadic forms.[6] Variants in the gene encoding surfactant protein C have been strongly associated with familial idiopathic pulmonary fibrosis, and mutations in the gene encoding surfactant protein A2 have been associated with familial pulmonary fibrosis and lung cancer. Telomere shortening caused by genetic variants within the human telomerase RNA or human telomerase reverse transcriptase has been associated with both familial and sporadic idiopathic pulmonary fibrosis.

Idiopathic pulmonary fibrosis is limited to the lungs in adults, usually older than 60 years, and it generally occurs in men with a history of cigarette smoking. Most often patients have otherwise been in good health and have no known connective tissue disease or exposure to drugs or environmental factors known to cause pulmonary fibrosis, although patients with significant cigarette smoking history may have coexisting emphysema. Typical clinical manifestations include a gradual onset and progression of exertional dyspnea, restrictive abnormalities on PFTs (Chapter 85), and a distinct pattern of bilateral pulmonary fibrosis on HRCT.

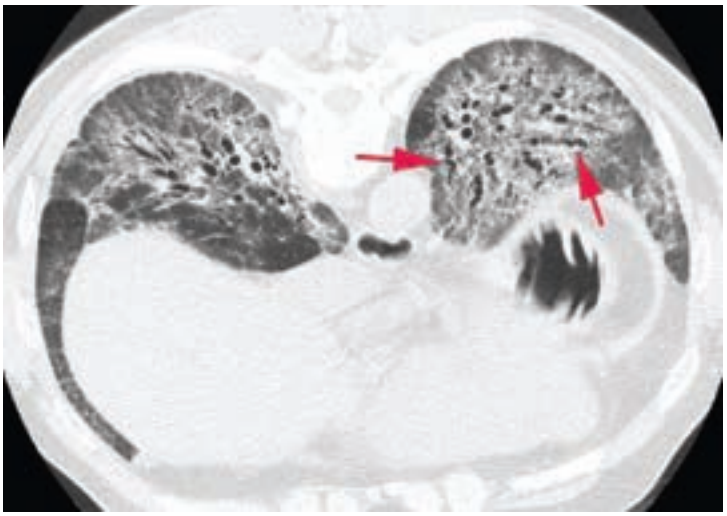

FIGURE 92-3. Computed tomography scan showing traction bronchiectasis *(arrows)*.

DIAGNOSIS

Chest radiographs typically show basal-predominant reticular abnormalities with low lung volumes.[7] The diagnostic features on HRCT are peripheral, predominantly basilar patchy intralobular reticulation, often with subpleural honeycomb cysts, traction bronchiectasis, and traction bronchiolectasis as the disease becomes more advanced (Fig. 92-2, *A*). Reticulation may progress to honeycombing, although neither alveolar consolidation nor parenchymal nodules are present. Compared with the other idiopathic interstitial pneumonias, the HRCT appearance of idiopathic pulmonary fibrosis is distinguished by the presence of fibrotic abnormalities, predominantly in the bases of the lower lobes (E-Fig. 92-E1), by subpleural reticulations (E-Fig. 92-E2), and by its hallmark honeycombing (E-Fig. 92-E3) and traction bronchiectasis (Fig. 92-3). There is a notable absence of extensive ground-glass opacification, micronodules, cysts, consolidation, significant air trapping in multiple lobes, pleural plaques, pleural effusion, and extensive mediastinal adenopathy, all of which are inconsistent with the radiographic pattern of usual interstitial pneumonia.

Pulmonary function tests usually show a progressive restrictive pattern. However, patients with milder disease may have normal lung volumes and a small decrease in DLCO; rarely, PFT results may be normal.

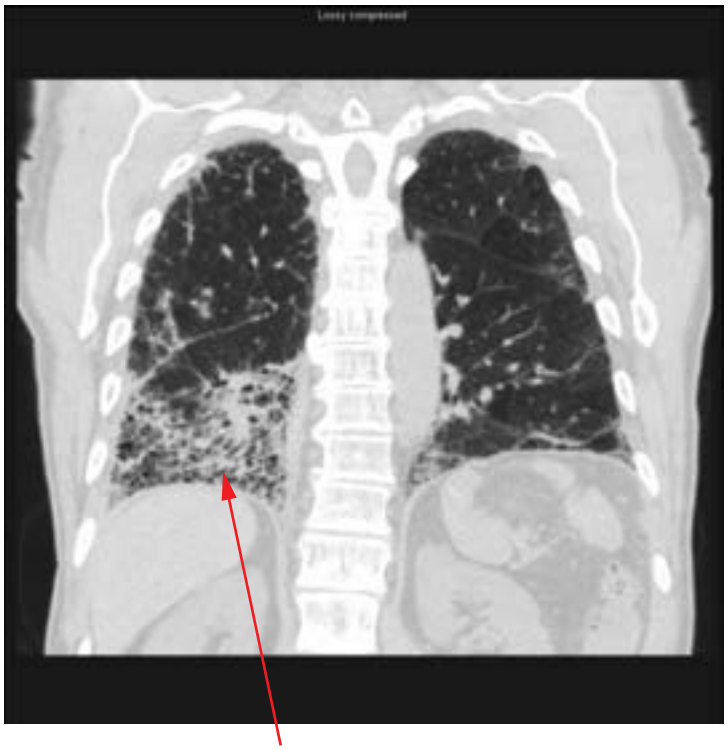

E-FIGURE 92-1. Computed tomography scan demonstrating lower lobe predominant fibrosis in a patient with idiopathic pulmonary fibrosis.

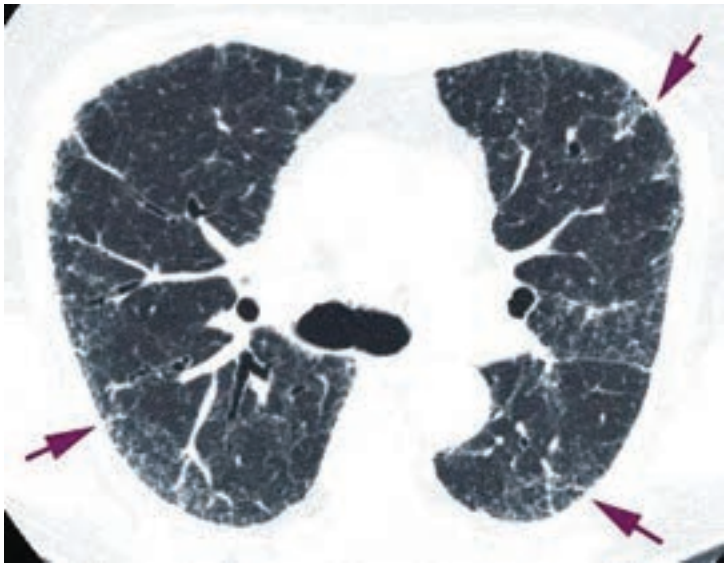

E-FIGURE 92-2. Computed tomography scan showing subpleural reticulations *(arrows)* in a patient with idiopathic pulmonary fibrosis (A) The cysts are interspersed within areas of ground glass attenuation (B and C).

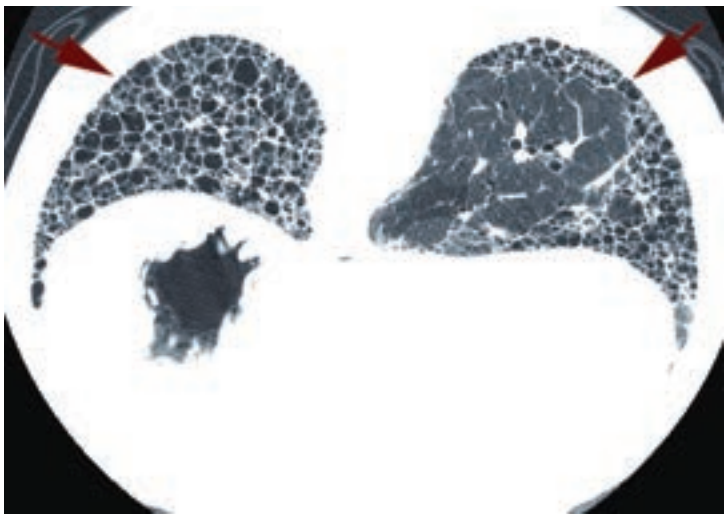

E-FIGURE 92-3. Computed tomography scan showing honeycombing *(arrows)* in a patient with idiopathic pulmonary fibrosis.

TABLE 92-6 DIAGNOSIS CRITERIA FOR IDIOPATHIC PULMONARY FIBROSIS

The diagnosis of idiopathic pulmonary fibrosis requires the presence of usual interstitial pneumonia (UIP) in the absence of other causes of interstitial lung disease (e.g., domestic, occupational, and environmental exposures, connective tissue disease, and drug toxicity) AND

a. The presence of a UIP pattern on chest HRCT in the absence of a lung biopsy or

b. Specific combinations* of chest HRCT patterns (UIP, possible UIP, inconsistent with UIP) and histopathologic features (UIP, probable UIP, possible UIP, not UIP) on surgical lung biopsy

HRCT FEATURES OF UIP	HISTOPATHOLOGIC FEATURES OF UIP
• Subpleural, basal predominance • Reticular abnormality • Honeycombing with or without traction bronchiectasis • Absence of peribronchovascular predominance, extensive ground-glass abnormality, diffuse micronodules, discrete cysts, diffuse mosaic attenuation, or consolidation	• Marked fibrosis/architectural distortion, +/– honeycombing in a predominantly subpleural/paraseptal distribution • Patchy parenchymal lung fibrosis • Fibroblast foci • No features suggesting an alternate diagnosis*

*Based on data from Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med.* 2011;183:788-824.
HRCT = high-resolution computed tomography.

The cellular pattern in BAL fluid, which is nonspecific, is marked by an excess of neutrophils in proportion to the extent of reticular change on HRCT; the percentage of eosinophils may be mildly increased. The histopathologic pattern of usual interstitial pneumonia consists of patchy interstitial changes alternating with zones of honeycombing, fibrosis, minimal inflammatory cells, collagen deposition, and normal lung (Fig. 92-2, *B*). Subepithelial fibroblastic foci, small aggregates of myofibroblasts, and fibroblasts within myxoid matrix are invariably present and represent areas of active fibrosis. The presence of temporal heterogeneity, or areas at different stages of fibrosis transitioning with normal areas and honeycomb cysts, along with fibrotic foci within the lung, is an essential feature of usual interstitial pneumonia that distinguishes it from other processes such as nonspecific interstitial pneumonia. Interstitial cellular inflammation is minimal in usual interstitial pneumonia. Although usual interstitial pneumonia characterizes the microscopic abnormality in idiopathic pulmonary fibrosis, the same histologic and radiologic pattern can also be seen in patients with rheumatologic lung diseases, chronic hypersensitivity pneumonitis, and asbestosis (Chapter 93). In the appropriate clinical setting (and after definitive exclusion of other known clinical conditions associated with ILD) (see later), a definitive diagnosis of idiopathic pulmonary fibrosis is based on the presence of a pattern of usual interstitial pneumonia on HRCT or surgical lung biopsy (Table 92-6).

TREATMENT

Rx

Pirfenidone (1800 mg/day for 1 year) decreases the rate of decline in forced vital capacity in clinical trials of patients who have idiopathic pulmonary fibrosis and mild to moderate impairment in pulmonary function,[1] and pooled data suggest an improvement in survival.[2] Pirfenidone is approved for the treatment of idiopathic pulmonary fibrosis in the U.S., Japan, and Europe.[3] Treatment with nintedanib (a tyrosine kinase inhibitor at 150 mg orally twice daily) also decreases the rate of disease progression as measured by FVC over 52 weeks in patients with idiopathic pulmonary fibrosis and mild to moderate impairment in pulmonary function;[4] it is now approved in the U.S. and Europe. Sildenafil (a phosphodiesterase inhibitor at 20 mg orally three times a day) has shown small benefits in terms of dyspnea, oxygenation, and quality of life but not exercise capacity in patients with idiopathic pulmonary fibrosis and severe impairment in pulmonary function.[5] Abnormal acid gastroesophageal reflux (Chapter 138) is very common in patients with idiopathic pulmonary fibrosis, and treatment with standard doses of proton pump inhibitors, H2 receptor antagonists, or both as used for gastroesophageal reflux disease (Chapter 138) can slow the rate of progression of idiopathic pulmonary fibrosis.[6]

By comparison, *N*-acetylcysteine alone[7] or as part of triple-therapy combined with prednisone plus azathioprine[8] is not beneficial. Warfarin increases respiratory hospitalizations and death in patients with idiopathic pulmonary fibrosis.[9] Interferon-γ1b, cyclophosphamide, colchicine, D-penicillamine, dual and selective endothelin receptor antagonists, and oral corticosteroids as monotherapy or in combination with immunosuppressive agents are not beneficial.

Despite the absence of data, patients who require hospitalization and intensive care for an acute exacerbation with loss of respiratory function in the absence of infection or other complications are usually treated with empirical intravenous corticosteroids (e.g., methylprednisolone 1.0 g intravenously as a pulse dose once a day for 3 days and followed by hydrocortisone, 125 mg every 6 hours for another 3 to 5 days), with further dosing dependent on the clinical response. Ancillary treatment measures, including supplemental oxygen (based on clinical and physiologic needs); prompt detection and treatment of respiratory tract infections and pulmonary embolism (Chapter 98); pulmonary rehabilitation; and immunization for influenza, herpes zoster, and pneumococcus, are all appropriate. Pulmonary hypertension, if present, may be treated (Chapter 68), but there is no evidence that such treatment will be beneficial.[10] Lung transplantation (Chapter 101) is indicated in selected patients, but about two thirds of patients with idiopathic pulmonary fibrosis are older than 60 to 65 years, which is a relative contraindication to lung transplantation. It is important to initiate discussion of palliative care measures before patients reach the terminal stages of the disease.

PROGNOSIS

The natural course of idiopathic pulmonary fibrosis is heterogeneous. Most patients exhibit a slow and steady decline, with a mortality rate of about 7% at 1 year and 14% at 2 years after diagnosis.[9] A small subset of patients declines at a rapid rate over several months, but another subset of patients remains stable over several years before declining. Progressive impairment of lung function and gas exchange ultimately is fatal unless the patient undergoes lung transplantation. Patients who survive longer generally have less fibrosis on HRCT, less functional impairment, no evidence of pulmonary hypertension, and no significant oxygen desaturation during a modified version of the 6-minute walk test. Patients with coexisting emphysema, pulmonary hypertension, or episodes of acute exacerbation have even shorter survival times. By comparison, patients who have a polymorphism in the gene encoding *MUC5B* may have better survival times.

Nonspecific Interstitial Pneumonia

Nonspecific interstitial pneumonia is often associated with connective tissue diseases, but idiopathic nonspecific interstitial pneumonia is also recognized as a distinct clinical entity. It typically occurs in middle-aged, nonsmoking women with an average age at diagnosis of about 50 years. The prevalence of nonspecific interstitial pneumonia has been estimated at one to nine per 100,000.

Two subgroups have been described, cellular and fibrotic. Because the average age at onset is about 10 years earlier in nonspecific interstitial pneumonia than in idiopathic pulmonary fibrosis and because the clinical features of idiopathic fibrotic nonspecific interstitial pneumonia are very similar to early cases of idiopathic pulmonary fibrosis, questions persist as to whether idiopathic fibrotic nonspecific interstitial pneumonia is a separate clinical entity or represents an early form of idiopathic pulmonary fibrosis.

DIAGNOSIS

Chest radiographs show bilateral patchy pulmonary infiltrates with a lower lung zone predominance in all forms of nonspecific interstitial pneumonia. HRCT reveals a predominant ground-glass pattern of attenuation, usually bilateral and often associated with subpleural reticulation (Fig. 92-4), and loss of volume in the lower lobe. In cellular nonspecific interstitial pneumonia, HRCT shows ground-glass opacification, consolidation, or both, but the biopsy shows mild to moderate lymphoplasmacytic interstitial chronic inflammation. The major differential diagnosis to consider as an alternative to cellular nonspecific interstitial pneumonia is acute or subacute hypersensitivity pneumonitis, so a thorough history regarding environmental exposures is crucial. In contrast, fibrotic nonspecific interstitial pneumonia has a bilateral lower lobe distribution with architectural derangement on HRCT; histopathologically, it has uniformly dense interstitial fibrosis and may sometimes be difficult to distinguish from idiopathic pulmonary fibrosis and usual interstitial pneumonia in the early clinical stages. In these circumstances, the diagnosis of fibrotic nonspecific interstitial pneumonia can be ascertained only by the histologic features in a surgical lung biopsy specimen.

TREATMENT AND PROGNOSIS

Rx

Patients with cellular nonspecific interstitial pneumonia usually respond to treatment with corticosteroids (see Table 92-5), and their prognosis is generally better than that of patients with idiopathic pulmonary fibrosis. Nonetheless, some patients progress over several years, and some manifest acute exacerbations similar to patients with idiopathic pulmonary fibrosis. Immunemodulating drugs, including prednisone, azathioprine, and mycophenolate, have been used empirically, with their doses based on clinical response as assessed by clinicians and not on evidence with randomized clinical trials.

SMOKING-RELATED INTERSTITIAL PNEUMONIAS

Respiratory Bronchiolitis–Associated Interstitial Lung Disease

This ILD is almost invariably associated with chronic and current cigarette smoking, and it usually manifests clinically during the fourth or fifth decade of life. However, it may also be detected incidentally on radiographs in relatively younger and asymptomatic persons with a previous history of cigarette smoking or in people passively exposed to chronic cigarette smoke Respiratory bronchiolitis–associated ILD is always associated with chronic exposure to cigarette smoke.

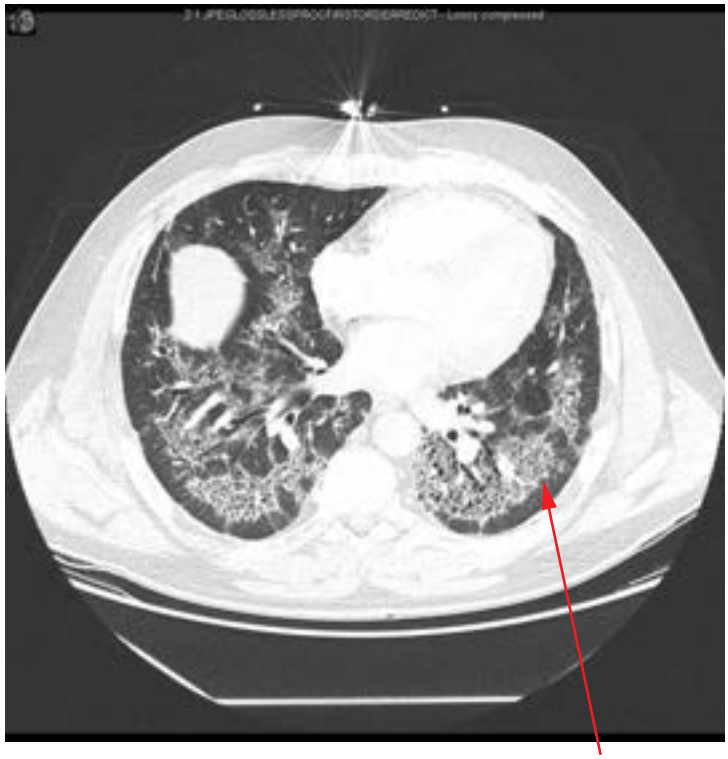

FIGURE 92-4. Nonspecific interstitial pneumonia. Computed tomography scan showing characteristic subpleural sparing.

DIAGNOSIS

Pulmonary function tests show varying degrees of airway obstruction, mildly decreased or preserved TLC, and decreased DLCO. The chest radiograph typically reveals bronchial wall thickening and areas of ground-glass attenuation. HRCT reveals centrilobular nodules with an upper lobe predominance, patchy ground-glass attenuation, and peribronchial alveolar septal thickening (Fig. 92-5, *A*). Areas of hypoattenuation (mosaic attenuation) represent air trapping as a result of small airways disease. The characteristic finding on BAL is numerous brown-pigmented alveolar macrophages, often with a modest increase in neutrophils. Lung biopsy is rarely needed, but its hallmark histopathologic feature is the accumulation of pigmented alveolar macrophages with glassy eosinophilic cytoplasm and granular pigmentation within respiratory bronchioles, typically with a chronic inflammatory cell infiltrate in the bronchioles and surrounding alveolar walls (Fig. 92-5, *B*). Fibroblastic foci and honeycomb change are not present, but centrilobular emphysema is frequent.

TREATMENT AND PROGNOSIS

Rx

Progression to honeycomb lung and end-stage fibrosis seldom occurs, and the prognosis is good with cessation of smoking. Discontinuation of cigarette smoking is essential, and patients may benefit from low-dose corticosteroids (e.g., prednisone, 10 to 20 mg/day) for a few months.

Desquamative Interstitial Pneumonia

Desquamative interstitial pneumonia is a rare entity (<3% of all ILDs) that may represent a form of respiratory bronchiolitis–associated ILD extending into the alveolar spaces and alveolar walls. Although most affected individuals are cigarette smokers, the histologic pattern of desquamative interstitial pneumonia may also occur in pneumoconiosis, rheumatologic disease, and drug-associated ILD. Patients are often initially seen with advanced disease and striking hypoxemia. The histopathology features of desquamative interstitial pneumonia are characterized by accumulation of pigmented alveolar macrophages within the alveoli. Histologic changes within the respiratory bronchioles and within the alveolar spaces can coexist and represent a histopathologic spectrum of alveolar macrophage accumulation.

DIAGNOSIS

Pulmonary function tests reveal a restrictive lung defect and decreased DLCO with or without coexisting airway obstruction. The chest radiograph shows patchy basal consolidation with a lower lobe and peripheral predominance (Fig. 92-6). HRCT shows bilateral symmetrical ground-glass opacities with a predominantly basal and peripheral distribution as well as diffuse alveolar septal thickening (Fig. 92-7, *A*). Irregular linear opacities, typically associated

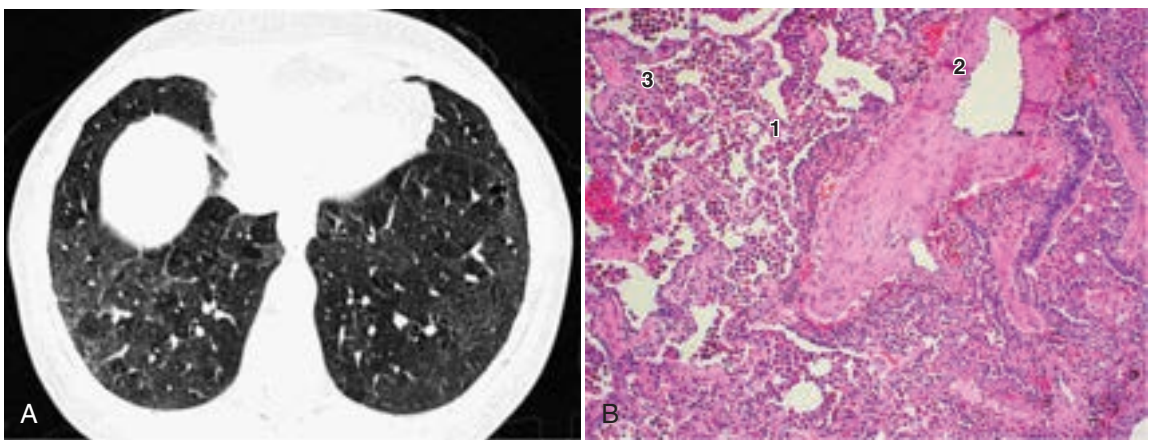

FIGURE 92-5. Respiratory bronchiolitis–associated interstitial lung disease. **A,** Ground-glass attenuation with a mosaic pattern on high-resolution computed tomography. **B,** Note the dense aggregates of (1) pigmented macrophages present in the air spaces around the terminal airways with (2) variable bronchiolar metaplasia and (3) interstitial fibrosis.

with traction bronchiectasis, may be noted. The finding of small discrete cysts, believed to represent trapped air in dilated bronchioles, within areas of ground-glass changes (E-Fig. 92-E4) and intervening normal lung parenchyma is highly suggestive of desquamative interstitial pneumonia. Fluid recovered from BAL quite often shows increased numbers of pigmented alveolar macrophages, frequently with increased neutrophils. Histopathologic findings on biopsy include diffuse alveolar septal thickening, hyperplasia of type II pneumocytes, and intense accumulation of intra-alveolar granular pigmented macrophages in a uniform manner (Fig. 92-7, *B*); fibrosis is minimal.

TREATMENT AND PROGNOSIS

Rx

With cessation of smoking and administration of oral corticosteroid therapy (see Table 92-5), outcomes are generally good, with an estimated overall survival rate of 70% at 10 years. However, a subset of patients may progress despite cessation of cigarette smoking, and a trial of corticosteroid therapy and lung transplantation is an appropriate consideration for selected patients.

ACUTE OR SUBACUTE IDIOPATHIC INTERSTITIAL PNEUMONIA

Acute Interstitial Pneumonia

Acute interstitial pneumonia is seen in otherwise healthy persons usually after an apparent acute viral upper respiratory infection (Fig. 92-8). The syndrome, historically known as Hamman-Rich syndrome (Chapter 91), mimics acute respiratory distress syndrome (Chapter 104). Acute interstitial pneumonia is a rare and fulminant idiopathic interstitial pneumonia that presents with acute symptoms and leads to respiratory distress or failure.

Cryptogenic Organizing Pneumonia

Cryptogenic organizing pneumonia, previously referred to as bronchiolitis obliterans organizing pneumonia of unknown cause (BOOP), is an idiopathic form of organizing pneumonia. Organizing pneumonia affects the small airways, including the distal bronchioles, respiratory bronchioles, alveolar ducts, and alveolar walls. Although the incidence and prevalence of cryptogenic organizing pneumonia are unknown, the estimated annual incidence in the United States is six to seven cases per 100,000. The mean age at presentation is about 60 years, and there is no sex predominance.

DIAGNOSIS

Cryptogenic organizing pneumonia most commonly manifests as a flulike illness with a nonproductive cough followed by exertional dyspnea. PFTs

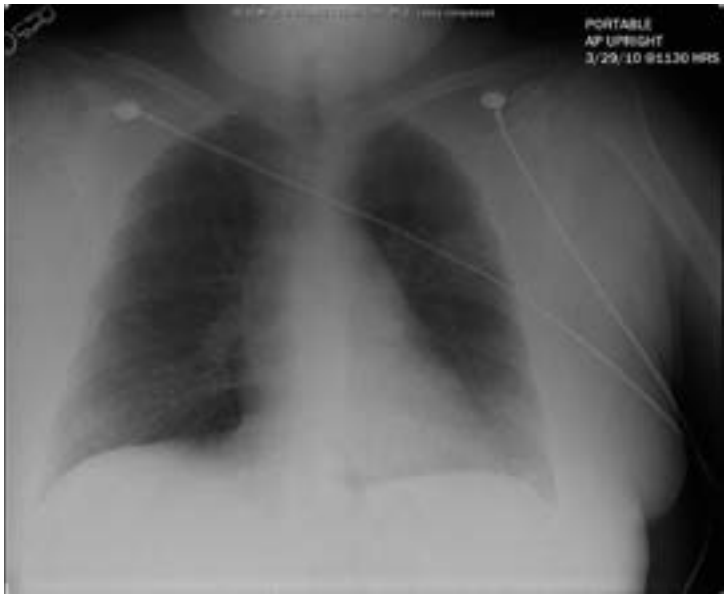

FIGURE 92-6. Anteroposterior chest radiograph showing patchy ground-glass infiltrates typical of desquamative interstitial pneumonia.

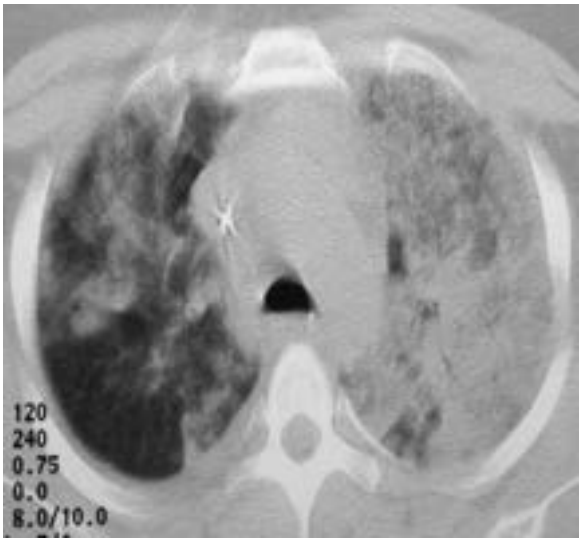

FIGURE 92-8. Acute interstitial pneumonia with diffuse alveolar damage histologically. Note the dense air space consolidation.

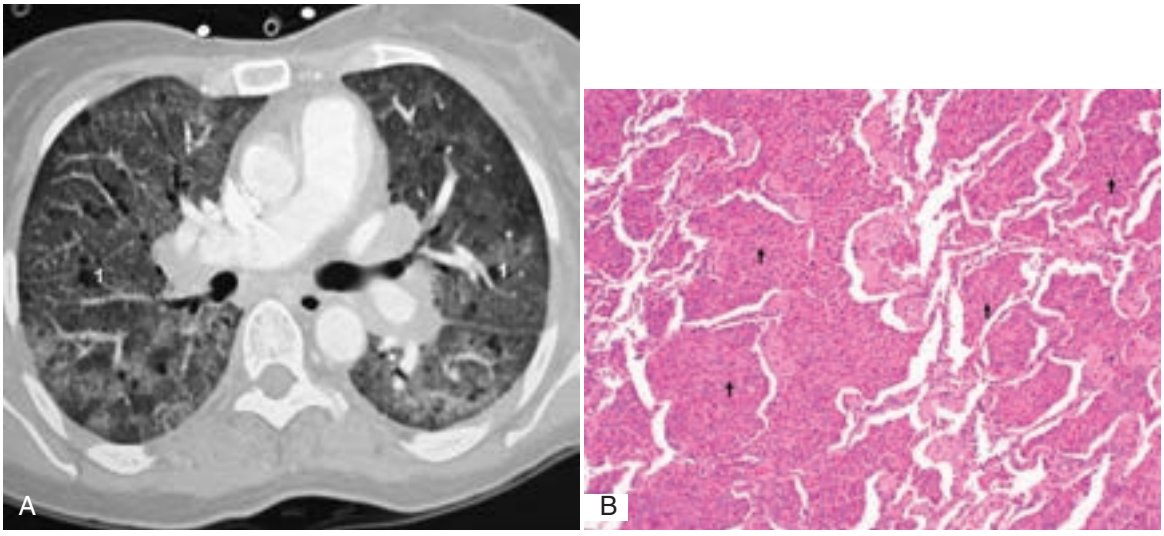

FIGURE 92-7. Desquamative interstitial pneumonia. **A,** Ground-glass attenuation with cystic spaces on high-resolution computed tomography. **B,** Note that the alveolar spaces are densely filled with macrophages *(arrowheads).*

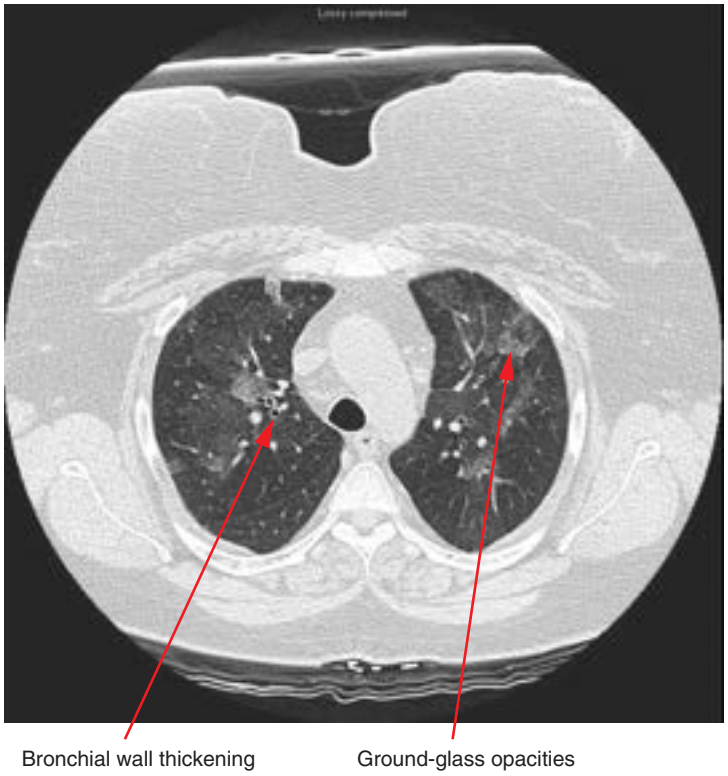

E-FIGURE 92-4. Computed tomography scan showing ground-glass opacities and bronchial wall thickening in a patient with desquamative interstitial pneumonia.

show a restrictive defect, but 20% of patients, most of whom are current or past smokers, also have an obstructive defect. Chest radiography reveals patchy unilateral or bilateral alveolar opacities that may be peripheral or migratory; small nodular opacities are seen in 10% to 50% of cases (Fig. 92-9). In about 90% of patients, HRCT shows areas of air space consolidation with lower lung zone predominance, frequently in a subpleural or peribronchial distribution (Fig. 92-10); other features include small nodules along bronchovascular bundles and ground-glass attenuation. BAL is nonspecific; increased lymphocytes, neutrophils, and eosinophils may be seen. On biopsy, key histologic features are excessive proliferation of granulation tissue within the small airways and alveolar ducts as well as chronic inflammation in the surrounding alveoli.

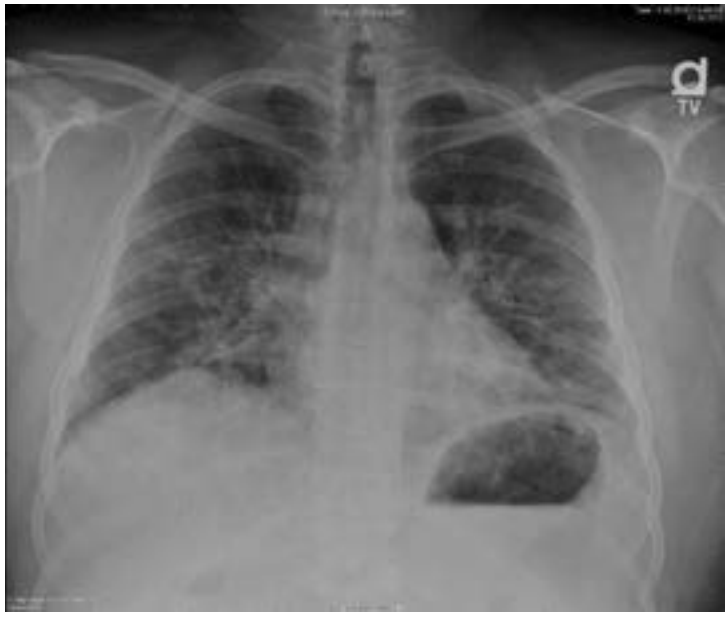

FIGURE 92-9. Chest radiograph showing cryptogenic organizing pneumonia. Note the bilateral patchy air space opacities.

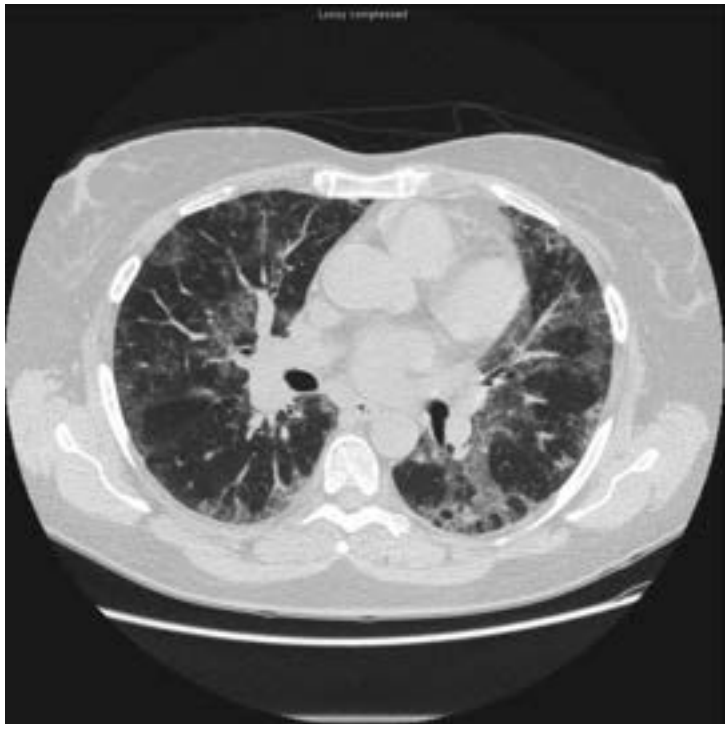

FIGURE 92-10. Peripheral ground-glass opacities in a patient with cryptogenic organizing pneumonia.

TREATMENT AND PROGNOSIS

Rx

Most patients recover rapidly and completely when treated with oral corticosteroids (see Table 92-5) for 6 months but may relapse after discontinuation and require oral corticosteroids for longer periods and sometimes indefinitely, often with adjunct immunosuppressive agents such as azathioprine. A small subset of patients in whom pulmonary fibrosis develops despite corticosteroids and azathioprine behave similarly to patients with idiopathic pulmonary fibrosis. Spontaneous remissions are known to occur.

Lymphoid and Lymphocytic Interstitial Pneumonia

This condition is more common in women, especially in the fifth decade of life, but it may occur at any age. Patients should be evaluated for concurrent connective tissue disease, an autoimmune disorder (especially Sjögren syndrome; Chapter 268), or common variable immunoglobulin deficiency (Chapter 250) because idiopathic LIP is very rare. Symptoms are nonspecific and include a gradual onset of cough and exertional dyspnea. Lymphoid and LIP is within the spectrum of benign lymphoproliferative disorders, and some patients manifest pseudolymphoma or lymphoma as a complication.

DIAGNOSIS

Chest radiographs show a reticular or reticulonodular pattern predominantly involving the lower lung zones. HRCT reveals bilateral ground-glass attenuation, small or large nodules, and scattered cysts; perivascular honeycombing and reticular abnormalities may also be seen (E-Fig. 92-E5). Increased numbers of lymphocytes are found on BAL, and biopsy reveals a dense interstitial lymphocytic infiltrate.

TREATMENT AND PROGNOSIS

Rx

Some patients respond to or stabilize with oral corticosteroids (see Table 92-5). The prognosis is variable, with more than one third of patients progressing to diffuse pulmonary fibrosis.

Interstitial Lung Disease Associated with Connective Tissue Disease

Many of the connective tissue diseases, including progressive systemic sclerosis (Chapter 267), rheumatoid arthritis (Chapter 264), SLE (Chapter 266), dermatomyositis and polymyositis (Chapter 269), Sjögren syndrome (Chapter 268), and mixed connective tissue disorder (Chapter 267), may have ILD as one of their manifestations. In fact, up to 20% of patients with connective tissue diseases may initially be thought to have an ILD alone. Therefore, these diagnoses must be considered in patients with ILD, even in the absence of extrathoracic findings. Conversely, because pulmonary involvement is a major cause of death in patients with connective tissue diseases, the presence of ILD should be carefully sought in affected patients. All forms of idiopathic interstitial pneumonia can occur in patients with connective tissue diseases. The natural history of ILD complicating connective tissue diseases is variable, especially because coexisting pulmonary vascular disease or nonparenchymal pulmonary involvement may be present.

PROGRESSIVE SYSTEMIC SCLEROSIS

Of the connective tissue diseases, progressive systemic sclerosis is most frequently associated with ILD. Pulmonary symptoms may antedate cutaneous or digital manifestations of the disease by several years. Most patients affected have nonspecific interstitial pneumonia, with a minority having a usual interstitial pneumonia pattern, and D_{LCO} levels correlate with mortality. Pulmonary hypertension, which can occur in the absence of pulmonary fibrosis, may result in cor pulmonale. Patients with chronic pulmonary fibrosis also have an increased risk for bronchogenic carcinoma, usually either bronchoalveolar cell carcinoma or adenocarcinoma. In a controlled trial, treatment with cyclophosphamide (50-100 mg/day orally) for 1 year stabilized the PFT findings and improved health-related quality of life in patients with scleroderma-related ILD.

RHEUMATOID ARTHRITIS

Although rheumatoid arthritis is more common in women (2:1 to 4:1 ratio), ILD associated with rheumatoid arthritis is more common in men

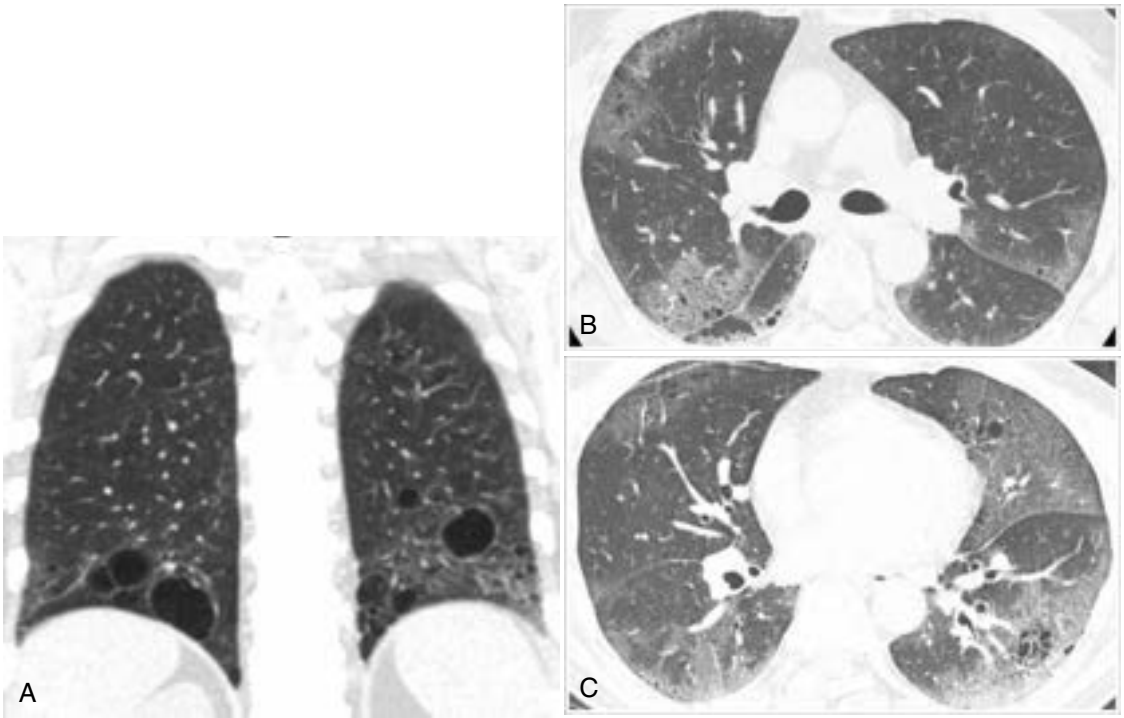

E-FIGURE 92-5. Lymphoid interstitial pneumonia. Note the cysts within the patchy ground-glass attenuation in lower lobes (A). The cysts are interspersed within areas of ground glass attenuation (B and C).

(3 : 1 ratio). Most cases occur at 50 to 60 years of age, and pulmonary symptoms follow the onset of arthritis in about 75% of cases. Lung involvement in rheumatoid arthritis may take many forms, but bronchiectasis, bronchiolitis, idiopathic interstitial pneumonias, and pleural effusions or pleural thickening are some of the most common. Early in the course, the histologic changes are similar to those of idiopathic interstitial pneumonias, including pulmonary fibrosis, but are distinguished by a prominent lymphocytic infiltrate that may contain germinal follicles adjacent to vessels and airways. As the disease progresses, the infiltration becomes less pronounced and is replaced by fibrous tissue, honeycomb changes, or both. Other pulmonary manifestations include pulmonary nodules, vasculitis, pulmonary hypertension, and Caplan syndrome (progressive upper lobe nodular pulmonary fibrosis in a coal miner with rheumatoid arthritis) but are relatively rare. Treatment is directed at the underlying rheumatoid arthritis (Chapter 264).

SYSTEMIC LUPUS ERYTHEMATOSUS

Pulmonary abnormalities complicating SLE (Chapter 266) may vary greatly. Pleural disease or pleural effusions (or both) are commonly present in lung disease that complicates SLE. Acute lupus pneumonitis may mimic acute interstitial pneumonia, with widespread ground-glass attenuation admixed with consolidation, or it may manifest as diffuse alveolar hemorrhage. Chronic ILD may also occur. Infection must always be considered in acutely ill patients who have received steroids or other immunosuppressive therapy. Rarely, the restrictive lung defect, which may be predominantly a result of diaphragmatic weakness, leads to a chest radiographic pattern of small-appearing lungs that may look progressively smaller over time. This so-called "shrinking lung" is generally resistant to corticosteroids or other immunosuppressive agents used to treat SLE. Otherwise, treatment of the ILD is similar to treatment of the underlying SLE.

DERMATOMYOSITIS AND POLYMYOSITIS

In contrast to progressive systemic sclerosis, the pattern of lung involvement in dermatomyositis and polymyositis is more heterogeneous. Usual interstitial pneumonia, nonspecific interstitial pneumonia, and organizing pneumonia have all been reported. Most patients have anti–Jo-1 antibody, and the disease is typically progressive over time. An acute interstitial pneumonia–like syndrome occurs in a subset of patients and is associated with high mortality rate despite aggressive immunosuppressive agents and high-dose corticosteroids. ILD may precede the muscular manifestations by months to years or be superimposed on established muscle disease. The severity of the muscular disease does not correlate with that of the ILD. Treatment is directed at the underlying disease (Chapter 269).

SJÖGREN SYNDROME

Interstitial lung disease is seen in patients with Sjögren syndrome, particularly those with the primary form of the disease. LIP is the most frequent subtype, but cryptogenic organizing pneumonia may also be present. Respiratory infections and bronchiectasis are common in advanced stages, perhaps because of inspissated mucus. Response to corticosteroid or immunosuppressive therapy is usually good (Chapter 268).

MIXED CONNECTIVE TISSUE DISEASE

This overlap syndrome (Chapter 267) combines features of progressive systemic sclerosis, SLE, rheumatoid arthritis, and polymyositis or dermatomyositis. Pulmonary disease is common, but it is most often subclinical and identified only radiographically. Treatment includes corticosteroids and other immune-modulating agents for the underlying disease.

ANKYLOSING SPONDYLITIS

The most common pulmonary manifestation of ankylosing spondylitis (Chapter 265) is upper lobe, bilateral reticulonodular infiltrates with cyst formation as a result of parenchymal destruction. There is no known effective therapy for this apical fibrobullous disease.

HYPERSENSITIVITY PNEUMONITIS

PATHOBIOLOGY

Hypersensitivity pneumonitis, also known as extrinsic allergic alveolitis, is a syndrome caused by repeated inhalation of specific antigens from occupational or environmental exposure (Chapters 93 and 94) in sensitized individuals. Within a short period after inhalation of an inciting agent, patients develop a nonspecific diffuse pneumonitis with inflammatory cell infiltration of the bronchioles, alveoli, and interstitium, sometimes associated with a pleural effusion. In the subacute and chronic stages, loosely formed, noncaseating, epithelioid cell granulomas and a mononuclear infiltrate may be dispersed in the interstitium. Hypersensitivity pneumonitis can occur with exposure to a wide range of inhaled antigens (Chapters 93 and 94). Some of the more common exposures are farmer's lung, bird fancier's lung, parakeet keeper's lung, and pigeon breeder's lung. The exposure to an avian antigen may be occult and related to bird droppings or bird nests. Hobbies (woodworker's lung) and recreational activities (sauna taker's lung; hot tub) may be implicated as well as occupations.

CLINICAL MANIFESTATIONS

The clinical features and severity of symptoms vary according to the frequency and intensity of exposure. A history of exposure to potential agents or changes in the domestic and other environments (or both) is essential to diagnosis and treatment (Chapters 19 and 93). The interval between exposure to the antigen and the clinical manifestations of lung disease is unknown, although symptoms can occur as soon as 4 to 12 hours after exposure. In such cases, fever and chills are common symptoms, are often temporally related to the workplace or to hobbies, and may actually disappear on vacations or during absence from the site of exposure, only to recur when exposure is resumed. In more chronic and low-level exposures, however, the onset is insidious. Some patients with chronic hypersensitivity pneumonitis owing to many years of exposure may manifest clinical features similar to those of patients with idiopathic pulmonary fibrosis.

DIAGNOSIS

Findings on chest radiography are diverse, with focal patchy consolidation or a diffuse ground-glass appearance in acute hypersensitivity pneumonitis (E-Fig. 92-E6); micronodular and reticular shadowing in subacute forms; and diffuse, predominantly upper lung zone reticulation with honeycombing in the chronic form. Chest radiograph results may be normal in up to 30% of patients with significant physiologic abnormalities.

On HRCT, small centrilobular ill-defined nodules of ground-glass densities are seen, along with evidence of mosaic attenuation (trapped air) as a result of concomitant bronchiolitis and upper lobe predominance of the parenchymal abnormalities (E-Fig. 92-E7). Chronically, findings of lung fibrosis may be indistinguishable from the patterns seen in usual interstitial pneumonia and idiopathic pulmonary fibrosis (E-Fig. 92-E8).

Precipitating serum antibodies for potential causes of hypersensitivity pneumonitis confirm exposure but not cause and effect, and the absence of antibodies does not exclude hypersensitivity pneumonitis. In some cases, a thorough investigation of the patient's home and workplace by an industrial hygienist may reveal occult molds, spores, *Thermoactinomycetes* spp., *Aureobasidium pullulans*, and other precipitating causes. When such exposures are evident and suspected to be the cause of the manifested ILD, further diagnostic interventions such as BAL may be quite helpful by showing the most marked increase in T lymphocytes and an increased number of plasma cells. The characteristic histologic triad in hypersensitivity pneumonitis is cellular nonspecific interstitial pneumonia, cellular bronchiolitis, and granulomatous inflammation; however, this triad is seen in no more than 75% of affected patients. Differentiation from cellular nonspecific interstitial pneumonia may be challenging. Prompted by an elicited history of environmental exposure for an attributable antigen, bronchochallenge testing with the suspected antigen in experienced laboratories may provide further clues for the diagnosis of hypersensitivity pneumonitis in a patient otherwise suspected to have idiopathic pulmonary fibrosis. A subset of patients who meet the diagnostic criteria for idiopathic pulmonary fibrosis may be subsequently diagnosed as having chronic hypersensitivity pneumonitis after careful evaluation by experts who are familiar with hypersensitivity pneumonitis.[10]

TREATMENT AND PROGNOSIS

Rx

A thorough investigation must be undertaken to identify the antigen in the patient's environment. Sometimes an industrial hygienist is needed to obtain samples for culture from potential sources in the patient's domestic or workplace environment. The identified antigen must be eradicated from the patient's environment. Avoidance of exposure to the identified antigen or antigens and treatment with corticosteroids (see Table 92-5) are important if improvement is to be obtained. However, despite thorough searches, the antigen may remain undetected in a substantial number of patients who have hypersensitivity pneumonitis confirmed by lung biopsy. Continued exposure

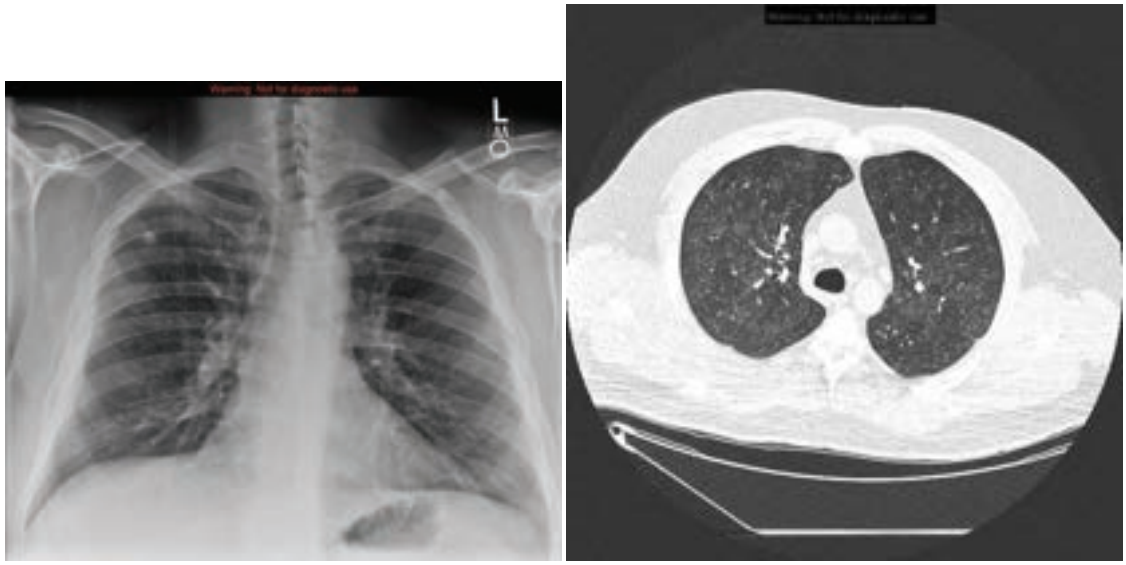

E-FIGURE 92-6. Chest radiograph (left) and computed tomography scan (right) on the same patient with hypersensitivity pneumonitis. Note the centrilobular nodules and ground-glass opacities on the computed tomography scan.

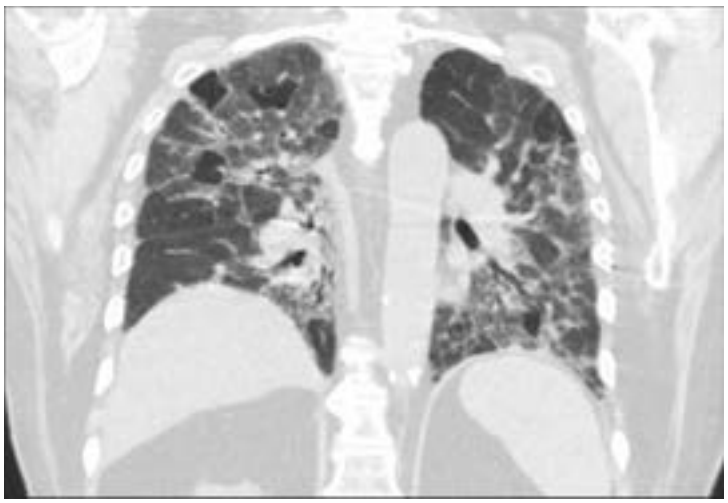

E-FIGURE 92-7. Upper lobe air trapping and mosaic attenuation in a patient with hypersensitivity pneumonitis.

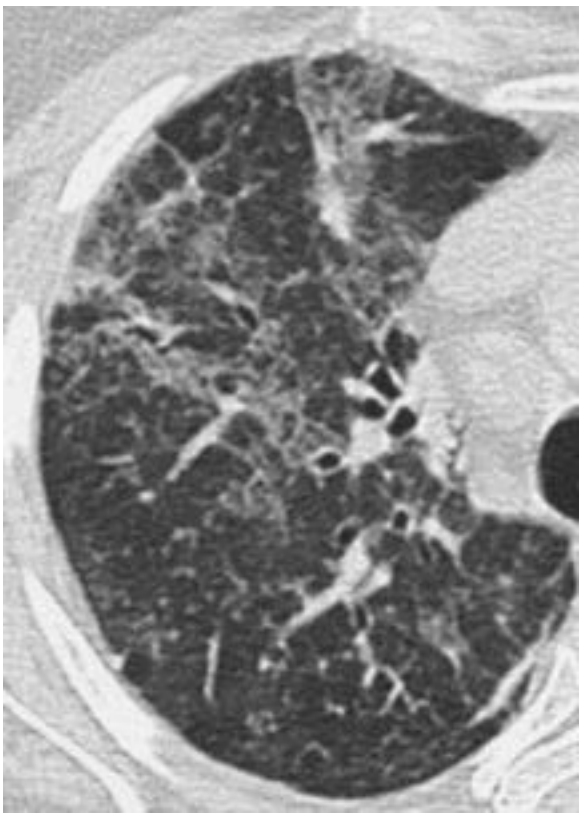

E-FIGURE 92-8. Upper lobe nodular and patchy fibrosis in a patient with chronic hypersensitivity pneumonitis.

to the unidentifiable antigens, prolonged exposure to antigens, or both can lead to chronic hypersensitivity pneumonitis and irreversible fibrosis that may not respond to any treatment regimen. In the fibrotic stages, the prognosis and clinical course may be similar to those of idiopathic pulmonary fibrosis.

OCCUPATIONAL INTERSTITIAL LUNG DISEASES

Interstitial lung diseases associated with specific occupations generally involve the inhalation and deposition of dust in the lungs followed by a tissue reaction that ultimately results in fibrosis. Examples include silicosis (inhalation of silica in crystalline form or silicon dioxide as quartz, cristobalite, or tridymite; at-risk occupations include sandblasting and working with granite), coal workers' pneumoconiosis (inhalation of coal dust), asbestosis (deposition of fibers during mining, milling, or other handling of asbestos; welding and working in a shipyard are two at-risk occupations), berylliosis (seen in aerospace workers and in electronic industries), and hard metal disease (Chapter 93). Radiographic features vary depending on the inciting inhalant. Cessation of the exposure is important, but the fibrosis is generally irreversible.

DRUG-INDUCED INTERSTITIAL LUNG DISEASE

More than 300 drugs, biomolecules, or homeopathic remedies (see Table 92-2) can cause acute, subacute, or chronic ILD, and the list continues to increase as new medications are introduced. The clinical and radiographic manifestations are quite varied. Examples of known syndromes include chronic nitrofurantoin-induced ILD that mimics idiopathic pulmonary fibrosis (and is fatal in ≈8% of cases), granulomatous pneumonitis secondary to methotrexate (<5%), sarcoid-like granulomatous ILD induced by interferon-α and tumor necrosis factor modulating agents, nonspecific bilateral alveolar and interstitial inflammatory and fibrotic abnormalities caused by bleomycin and other chemotherapeutic agents, and alveolar and interstitial abnormalities and nodular densities in acute and chronic amiodarone pulmonary toxicity. Most drug-induced ILD is reversible if recognized early and if use of the responsible drug is discontinued. In addition to discontinuing the implicated drug, treatment with corticosteroids (see Table 92-5) is indicated in patients with moderate to severe functional impairment.

Alveolar Filling Disorders

In alveolar filling disorders (Chapter 91), air spaces distal to the terminal bronchioles are filled with blood, lipid, protein, water, or inflammatory cells. The radiographic appearance is that of an alveolar infiltrate with small nodular densities and ill-defined margins; hence, the radiographic picture is similar to an ILD, and virtually all the alveolar filling disorders may result in ILD, including Goodpasture syndrome, pulmonary alveolar proteinosis (primary and secondary), alveolar hemorrhage syndromes (Chapter 91), acute interstitial pneumonia, and bronchoalveolar cell carcinoma (Chapter 191).

IDIOPATHIC PULMONARY HEMOSIDEROSIS

This rare disorder of children and young adults is characterized by intermittent, diffuse alveolar hemorrhage without evidence of vasculitis, inflammation, granulomas, or necrosis. The etiology is poorly understood. Anemia and hepatosplenomegaly may be present. Hemosiderin-laden macrophages in BAL fluid and lung tissue are part of the diagnostic picture. The chest radiograph reveals diffuse, bilateral alveolar infiltrates. A chronic interstitial infiltrate may develop after repeated episodes, infrequently with hilar and mediastinal adenopathy. Systemic corticosteroids (see Table 92-5) may be beneficial in treating acute disease.

CHRONIC EOSINOPHILIC PNEUMONIA

The clinical manifestation of chronic eosinophilic pneumonia varies over a wide spectrum, from asymptomatic to respiratory failure. The disease often occurs in women in the second to fourth decades of life; such women often manifest constitutional symptoms of fevers, sweats, weight loss, fatigue, dyspnea, and cough. Peripheral blood eosinophilia (Chapter 170), usually at levels of 10% to 40%, is common but may be absent in up to one third of affected patients at initial evaluation. On chest radiography and HRCT, the cardinal feature is peripheral multifocal consolidation, predominantly in the upper and mid lung zones. These dense peripheral infiltrates, which have sometimes been called the "photographic negative of pulmonary edema," often resolve dramatically after treatment with corticosteroids. Ground-glass attenuation commonly accompanies the consolidation. BAL fluid may show greater than 40% eosinophils during exacerbations. Treatment with corticosteroids (see Table 92-5) results in a rapid response, frequently within hours; in fact, such a dramatic resolution of symptoms with radiographic clearance of infiltrates shortly after initiation of corticosteroid therapy is considered "diagnostic." However, the rate of relapse is high, so most patients require prolonged treatment with low-dose corticosteroids (prednisone, 5-10 mg/day) to stay in remission.

Interstitial Lung Disease Associated with Pulmonary Vasculitides

GRANULOMATOSIS WITH POLYANGIITIS (FORMERLY KNOWN AS WEGENER GRANULOMATOSIS)

Granulomatosis with polyangiitis (Chapter 270) is the most common form of vasculitis that involves the lung. The systemic necrotizing granulomatous inflammation and small-vessel vasculitis are often manifested first in the upper respiratory tract as chronic rhinitis or sinusitis (or both), epistaxis, oropharyngeal ulcerations, gingival hyperplasia with clefting, or serous otitis media. Destruction of the nasal cartilage may lead to septal perforation or a saddle nose deformity. Ulcerative lesions of the tracheobronchial tree, cavitating nodules within the lung parenchyma, and diffuse alveolar hemorrhage caused by pulmonary capillaritis are lower respiratory tract manifestations. Focal segmental necrotizing glomerulonephritis is the most common extrathoracic manifestation, although pulmonary involvement may occur without renal disease. Chest radiography usually reveals multiple nodular or cavitating infiltrates, but single nodules may be found as well. The diagnosis is most commonly made serologically, with demonstration of antineutrophil cytoplasmic antibodies, although a negative test result does not exclude the disease. Treatment is usually with cyclophosphamide (50-100 mg/day or 2 mg/kg of ideal body weight per day but not more than 150 mg/day) in conjunction with oral corticosteroids (prednisone, 10-40 mg/day). Initial remission occurs in more than 90% of patients, but most patients require treatment for several years. Relapses may occur in up to 30% of patients, especially when treatment is tapered; such patients may need treatment indefinitely. Rituximab is an alternative if cyclophosphamide is unsuccessful or not tolerated (Chapter 270). Prophylaxis for *Pneumocystis jiroveci* infection is indicated in patients receiving chronic treatment.

CHURG-STRAUSS SYNDROME (ALLERGIC ANGIITIS)

This systemic necrotizing vasculitis (Chapter 270) affects both the upper and lower respiratory tracts and is almost invariably preceded by allergic disorders such as asthma, allergic rhinitis, sinusitis, or a drug reaction. Peripheral and lung eosinophilia, bronchospasm, increased immunoglobulin E levels, and rashes are common manifestations. The pulmonary radiographic findings are bilateral patchy, fleeting infiltrates; diffuse nodular infiltrates; or diffuse reticulonodular disease. Histopathologic examination of lung tissue is generally diagnostic with features of granulomatous angiitis or vasculitis. Although treatment with corticosteroids is indicated, the dosage and duration are unclear (see Table 92-5).

IDIOPATHIC PULMONARY CAPILLARITIS

Idiopathic pulmonary capillaritis may involve the pulmonary vasculature within the alveolar walls and be manifested as ILD. Patients may also have subclinical alveolar hemorrhage, often associated with the presence of perinuclear antineutrophilic cytoplasmic antibodies. Corticosteroids are the mainstay of treatment, but the doses and duration of treatment are unclear. Frequently, patients need adjunctive treatment with cyclophosphamide or rituximab, similar to patients with vasculitis and granulomatosis with polyangiitis (Chapter 270).

Other Forms of Interstitial Lung Disease

SARCOIDOSIS

See Chapter 95.

PULMONARY LANGERHANS CELL HISTIOCYTOSIS

This condition, previously known as pulmonary histiocytosis X or eosinophilic granuloma of the lung, is an idiopathic, granulomatous ILD that typically occurs in the second or third decade of life; there is a male preponderance. The currently accepted term is *Langerhans cell histiocytosis*. Pulmonary Langerhans cell histiocytosis is rare, with an estimated incidence of two to five cases per million population. The large majority (~90%) of affected individuals are male smokers, and current evidence suggests that the disorder results from an abnormal immune response to a component or derivative of cigarette smoke.

CLINICAL MANIFESTATIONS

The clinical findings are variable and range from an abnormal chest radiograph in an asymptomatic patient to progressive dyspnea with a nonproductive cough. Systemic symptoms of malaise, fever, and weight loss may be present. Hemoptysis is rare. Spontaneous pneumothorax, which occurs in approximately 25% of patients and is caused by rupture of subpleural cysts, may be an initial finding. Langerhans cell histiocytosis may be confined to the lung or may be a component of a multisystem disease that includes painful cystic bone lesions and diabetes insipidus (Chapter 225).

DIAGNOSIS

The chest radiograph shows diffuse symmetrical reticulonodular opacities superimposed on multiple small cysts in the upper and mid lung zones. HRCT reveals subpleural nodules; scattered ground-glass densities; and irregular cysts of varying number, size, and configuration in both lungs, with sparing of the lung bases (E-Fig. 92-E9). In the appropriate clinical setting, this pattern may be pathognomonic. As the disease progresses, the increase in fibrosis and cysts may lead to honeycombing of the lung. PFTs are characterized by a mixed restrictive and obstructive pattern, including a reduction in diffusing capacity. Vital capacity is disproportionately reduced compared with TLC because of air trapping within the cysts; the result is an increased residual volume. BAL reveals Langerhans cells (atypical histiocytes) that have the characteristic "x body" (i.e., Birbeck granule) on electron microscopy; immunostaining shows CD1 antigen on the cell surface and S-100 protein in the cytoplasm. However, the absence of these findings does not exclude the diagnosis. The diagnosis is usually made by transbronchial biopsy or open lung biopsy, which reveals interstitial and peribronchiolar collections of histiocytes, eosinophils, and lymphocytes; peribronchiolar nodules; and cysts with areas of central stellate fibrosis.

TREATMENT AND PROGNOSIS

Rx

Although definitive regression after discontinuation of smoking has not been proved, small series report improvement, so patients should be encouraged to discontinue smoking. The prognosis in pulmonary Langerhans cell histiocytosis is usually favorable, with approximately 75% of patients improving or stabilizing, especially with cessation of smoking; some patients, however, may progress to end-stage lung disease. In patients with progressive disease, corticosteroids (see Table 92-5) with or without vincristine, arabinoside, cyclosporine, cyclophosphamide, and azathioprine have been used with some anecdotal reports of success. Lung transplantation has been performed, but recurrent disease has been reported in the allograft.

LYMPHANGIOLEIOMYOMATOSIS

This rare ILD is limited to women, primarily of childbearing age. Proliferation of abnormal smooth muscle around bronchioles leads to bilateral small cysts, which give an appearance of ILD on chest radiographs, and progressive impairment of lung function. Hemoptysis, pneumothorax (from rupture of subpleural cysts), and chylothorax (from lymphatic obstruction) may be initial symptoms that distinguish this disorder from other diffuse lung diseases. Although lymphangioleiomyomatosis is usually limited to the lungs, an association with angiomyolipomas of the mediastinal and retroperitoneal lymph nodes and kidney has been described, so the disease may mimic the manifestations of tuberous sclerosis (Chapter 417). Coarse reticulonodular infiltrates, often with cysts or bullae, are typically seen on chest radiographs. In contrast to most other ILDs, increased lung volumes may be present and should prompt consideration of this diagnosis in a nonsmoking woman of reproductive age. HRCT shows characteristic diffuse thin-walled cysts, generally less than 2 cm in diameter. BAL may show occult alveolar hemorrhage (E-Fig. 92-E10). Lung biopsy reveals abnormal smooth muscle cells lining the airways, lymphatics, and blood vessels, with concurrent airflow obstruction and replacement of the lung parenchyma with cysts.

TREATMENT AND PROGNOSIS

Rx

In a randomized trial, sirolimus, an inhibitor of rapamycin signaling, initially at 2 mg/day and titrated to maintain trough levels between 5 and 15 ng/mL, was safe and stabilized lung function.[A] Treatment with progesterone or tamoxifen has been tried, but no randomized trials support the use of interventions to alter the estrogen–progesterone balance. Although lung transplantation is indicated as the patient reaches severe functional impairment, the disease may recur in the transplanted lung. Most patients currently die of respiratory failure about 10 years after the onset of symptoms.

Inherited Disorders

Several rare genetic disorders are associated with ILD and pulmonary fibrosis. Inheritance is autosomal dominant with variable penetrance for most cases of familial idiopathic pulmonary fibrosis and familial idiopathic interstitial pneumonia and for tuberous sclerosis (Chapter 417), neurofibromatosis (Chapter 417), and familial hypocalciuric hypercalcemia (Chapter 245). Inheritance is autosomal recessive for Gaucher disease (Chapter 208), Niemann-Pick disease (Chapter 208), and Hermansky-Pudlak syndrome (Chapter 208). The congenital form of the alveolar filling disorder pulmonary alveolar proteinosis also is inherited in an autosomal recessive manner.

Tuberous sclerosis (Chapter 417), an autosomal dominant disease of variable penetrance, is characterized pathologically by the presence of hamartomas in multiple organs. The most well-known clinical manifestations include epilepsy, mental retardation, adenoma sebaceum, and renal angiomyolipomas. ILD occurs in only 1% of patients with tuberous sclerosis, usually in women older than 30 years with little or no mental retardation. The pulmonary involvement, which is indistinguishable from lymphangioleiomyomatosis both radiographically and histopathologically, may be manifested as exertional dyspnea, recurrent pneumothorax, and hemoptysis. HRCT reveals thin-walled cysts and a diffuse reticulonodular infiltrate. Recurrent parenchymal hemorrhage may lead to hemosiderin deposition and interstitial pulmonary fibrosis. There is no cure, and treatment is supportive.

Neurofibromatosis (Chapter 417) may affect all age groups and both sexes. Type 1 (von Recklinghausen disease) is characterized by cafe au lait spots, neurofibromas, optic glioma, and bony lesions; the more rare type 2 is associated with bilateral acoustic neuromas. Diffuse ILD is manifested as bilateral lower lobe fibrosis as well as bullae or cystic changes. Interstitial fibrosis and alveolitis with thickening of the alveolar septa, accompanied by a cellular infiltrate, are seen on lung biopsy. Management is as for idiopathic pulmonary fibrosis.

Gaucher disease (Chapter 208), a lysosomal glycolipid storage disorder, has a predilection for the Ashkenazi Jewish population. Pulmonary manifestations, which occur most frequently in type 2 disease, may be a result of interstitial infiltration by Gaucher cells with fibrosis, alveolar consolidation, and filling of alveolar spaces; capillary plugging by Gaucher cells may cause secondary pulmonary hypertension. Treatment is as for the systemic disease, and the general approaches for idiopathic interstitial pneumonias and pulmonary arterial hypertension may be followed for these patients.

Niemann-Pick disease (Chapter 208) is a rare lipid storage disease that may cause infiltration of the characteristic "foam cell" throughout the pulmonary lymphatics, the pulmonary arteries, and the pulmonary alveoli. Patients with type B may survive into adulthood. Treatment is as for the systemic disease.

Hermansky-Pudlak syndrome (Chapter 208) is characterized by oculocutaneous albinism, a bleeding diathesis, and ceroid inclusions in macrophages. Most patients are of Puerto Rican ancestry, and women are affected more frequently than men. Pulmonary fibrosis, with onset in the third or fourth decade, is slowly progressive. Treatment principles and interventions are largely supportive and extrapolated from other related conditions, especially idiopathic pulmonary fibrosis.

Grade A References

A1. Corte TJ, Keir GJ, Dimopoulos K, et al. Bosentan in pulmonary hypertension associated with fibrotic idiopathic interstitial pneumonia. *Am J Respir Crit Care Med*. 2014;190:208-217.

A2. Taniguichi H, Ebina M, Kondoh Y, et al. Pirfenidone in idiopathic pulmonary fibrosis. *Eur Respir J*. 2010;35:821-829.

A3. Noble PW, Albera C, Bradford WZ, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet*. 2011;377:1760-1769.

A4. King TE Jr, Bradford WZ, Castro-Bernardini S, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med*. 2014;370:2083-2092.

A5. Xaubet A, Serrano-Mollar A, Ancochea J. Pirfenidone for the treatment of idiopathic pulmonary fibrosis. *Expert Opin Pharmacother*. 2014;15:275-281.

A6. Richeldi L, du Bois RM, Raghu G, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med*. 2014;370:2071-2082.

A7. Zisman DA, Schwarz M, Anstrom KJ, et al. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N Engl J Med*. 2010;363:620-628.

A8. Lee JS, Collard HR, Anstrom KJ, et al. Anti-acid treatment and disease progression in idiopathic pulmonary fibrosis: an analysis of data from three randomised controlled trials. *Lancet Respir Med*. 2013;1:369-376.

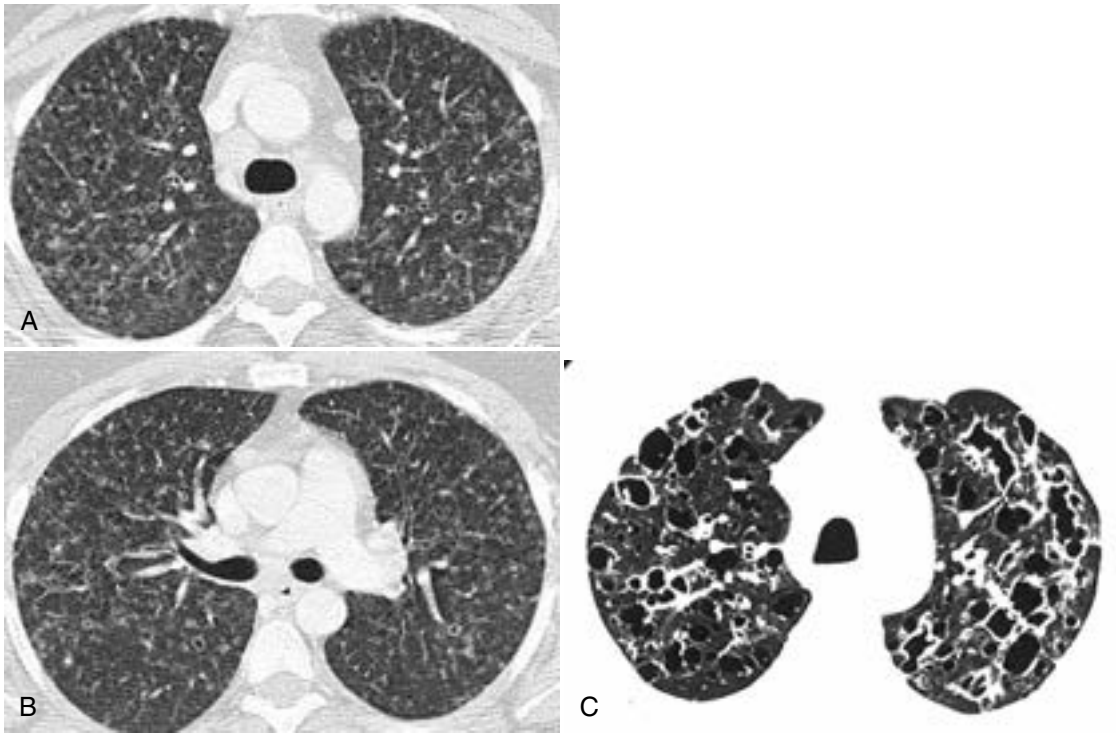

E-FIGURE 92-9. Pulmonary Langerhans cell histiocytosis. High-resolution computed tomography reveals subpleural nodules, scattered ground-glass densities (**A** and **B**), and irregular cysts of varying number and size (**C**).

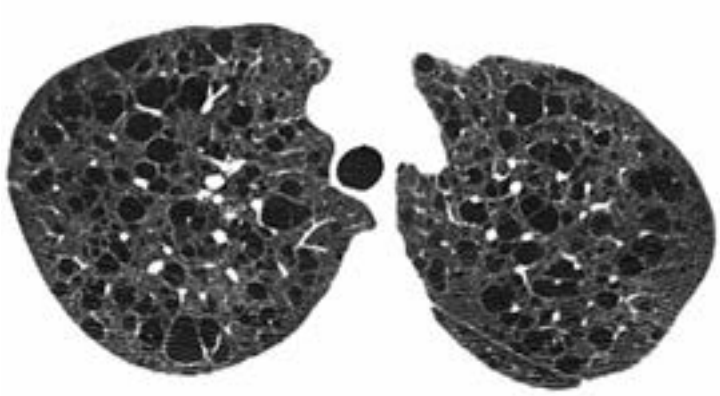

E-FIGURE 92-10. Bilateral thin wall cysts in a patient with lymphangioleiomyomatosis.

A9. Martinez FJ, de Andrade JA, Anstrom KJ, et al. Randomized trial of acetylcysteine in idiopathic pulmonary fibrosis. *N Engl J Med*. 2014;370:2093-2101.
A10. Raghu G, Anstrom KJ, King TE Jr, et al. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N Engl J Med*. 2012;366:1968-1977.
A11. Noth I, Anstrom KJ, Calvert SB, et al. A placebo-controlled randomized trial of warfarin in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2012;186:88-95.
A12. Raghu G, Behr J, Brown KK, et al. Treatment of idiopathic pulmonary fibrosis with ambrisentan: a parallel, randomized trial. *Ann Intern Med*. 2013;158:641-649.
A13. Tashkin DP, Elashoff R, Clements PJ, et al. Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med*. 2006;354:2655-2666.
A14. McCormack FX, Inoue Y, Moss J, et al. Efficacy and safety of sirolimus in lymphangioleiomyomatosis. *N Engl J Med*. 2011;364:1595-1606.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Raghu G, Chen SY, Yeh WS, et al. Idiopathic pulmonary fibrosis in US Medicare beneficiaries aged 65 years and older: incidence, prevalence, and survival, 2001-11. *Lancet Respir Med*. 2014;2:566-572.
2. Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet*. 2013;45:613-620.
3. Hunninghake GM, Hatabu H, Okajima Y, et al. MUC5B promoter polymorphism and interstitial lung abnormalities. *N Engl J Med*. 2013;368:2192-2200.
4. Meyer KC, Raghu G, Baughman RP, et al. An official American Thoracic Society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. *Am J Respir Crit Care Med*. 2012;185:1004-1014.
5. Travis WD, Costabel U, Hansell DM, et al. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med*. 2013;188:733-748.
6. Stock CJ, Sato H, Fonseca C, et al. Mucin 5B promoter polymorphism is associated with idiopathic pulmonary fibrosis but not with development of lung fibrosis in systemic sclerosis or sarcoidosis. *Thorax*. 2013;68:436-441.
7. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med*. 2011;183:788-824.
8. Xaubet A, Serrano-Mollar A, Ancochea J. Pirfenidone for the treatment of idiopathic pulmonary fibrosis. *Expert Opin Pharmacother*. 2014;15:275-281.
9. King TE Jr, Albera C, Bradford WZ, et al. All-cause mortality rate in patients with idiopathic pulmonary fibrosis. Implications for the design and execution of clinical trials. *Am J Respir Crit Care Med*. 2014;189:825-831.
10. Morell F, Villar A, Montero MA, et al. Chronic hypersensitivity pneumonitis in patients diagnosed with idiopathic pulmonary fibrosis: a prospective case-cohort study. *Lancet Respir Med*. 2013;1:685-694.

REVIEW QUESTIONS

1. The diagnosis of idiopathic pulmonary fibrosis is ascertained by

A. bronchoalveolar lavage (BAL) cellular analyses.
B. the mere pattern of usual interstitial pneumonia (UIP) in the surgical lung biopsy.
C. the mere pattern of usual interstitial pneumonia (UIP) pattern in HRCT image of the chest.
D. exclusion of environmental and other clinical conditions of interstitial lung disease (ILD).
E. all of the above.

Answer: D It is imperative to exclude all possible factors known to cause or be associated with ILD and pulmonary fibrosis, including the patient's environment (work and domestic), where the patient spends time, connective tissue diseases, and the use of licit and illicit drugs. The clinical features of idiopathic pulmonary fibrosis, including HRCT findings and histopathologic patterns of usual interstitial pneumonia, are nonspecific. A recent prospective study in one center with experienced ILD experts documented that 43% of patients who were diagnosed with idiopathic pulmonary fibrosis based on current 2011 guidelines turned out to have chronic hypersensitivity pneumonitis.

2. Cigarette smoking is *not* known to be a risk factor for

A. idiopathic pulmonary fibrosis.
B. respiratory bronchiolitis.
C. pulmonary Langerhans cell histiocytosis.
D. desquamative interstitial pneumonia.
E. acute interstitial pneumonia.

Answer: E Acute interstitial pneumonia, which is a rare interstitial pneumonia of unknown etiology, occurs acutely in otherwise healthy adults and rapidly progresses to respiratory failure. The radiologic and histopathologic features are of diffuse alveolar damage, similar to patients with acute respiratory distress syndrome. There are no known risk factors for acute interstitial pneumonia, not even cigarette smoking.

93 OCCUPATIONAL LUNG DISEASE

SUSAN M. TARLO

Occupational lung diseases include a wide spectrum of respiratory disorders with symptoms, signs, and diagnostic test results that often present with features similar to nonoccupational diseases (Table 93-1). For example, adult-onset asthma (Chapter 87) may be occupational asthma, presumed sarcoidosis (Chapter 95) may actually be chronic beryllium disease, apparent idiopathic pulmonary fibrosis may be asbestosis, or a suspected viral pneumonia (Chapter 97) may be hypersensitivity pneumonitis from an occupational cause such as contaminated metal-working fluid.

When evaluating any respiratory disease, the clinician should consider the possibility of an occupational cause or contribution (see E-Fig. 19-1; see Table 93-1). The onset of disease after an occupational exposure may occur with a short latency period, as for an acute toxic inhalation injury, or over a period of months to years, as for occupational asthma or hypersensitivity pneumonitis. Latency can be 20 years or more in chronic beryllium disease or lung cancer from chromium, asbestos, or other carcinogens. The most relevant job and occupational exposure history will therefore depend in part on the type of lung disease: for acute syndromes, the recent job exposure is most relevant; for asthma or hypersensitivity pneumonitis, the exposures at the onset of symptoms and ongoing exposures are most relevant; but for chronic diseases or diseases that may result from a long latency exposure, a full working history is essential. Consideration also should be given to potentially relevant exposures that may be related to a patient's hobbies or avocations (e.g., woodworking, model building, or insect collecting).

The clinical relevance of a correct occupational attribution is most apparent for diseases with a close temporal relationship between exposure and the onset of symptoms because intervention to reduce or remove exposure may reverse the disease or prevent progression. In addition, interventions in the workplace may reduce or prevent disease in other workers. However, even for diseases with a potential long latency, such as chronic beryllium disease, identification of disease in one worker should be regarded as a sentinel event that can lead to investigation of the workplace exposures and introduction of preventive measures. With the assistance of their physicians, workers with occupational lung diseases also often can qualify for workers' compensation.

EPIDEMIOLOGY

No reliable figures exist for the total incidence or prevalence of occupational lung diseases, and regional variation in occupations and exposures is substantial. Work-related asthma has become the most common chronic occupational lung disease in developed countries, where occupational asthma (asthma caused by work) accounts for about 15% of all adult-onset asthma, and work-exacerbated asthma occurs in 25 to 52% of asthmatic workers. The occupational contribution of workplace dusts, fumes, and gases to chronic obstructive pulmonary disease (COPD) is estimated at 15%.

In contrast, pneumoconiosis from silica or coal dust, although still important in developing countries, has declined in incidence in developed countries (E-Fig. 93-1A and B)[1] as a result of occupational hygiene measures. For example, approximately 100,000 Americans received benefits from the Federal Black Lung Program in 2005, compared with about 500,000 in 1980, and the percentage of coal miners with pneumoconiosis has fallen from 11% in the mid-1970s down to 3%. In some states, however, mortality rates have started to rise again, especially in smaller mines. Newer exposures that can cause silicosis include the textile industry's use of jet silica blasting of denim jeans, the use of artificial stone for kitchen counters, and hydraulic fracturing (fracking). Newly recognized asbestos-related diseases continue to occur, owing to the long latency period between exposure and clinical disease, despite the declining exposure to asbestos in developed countries. Although annual deaths from asbestosis have now reached a plateau in North America (E-Fig. 93-1C) and will likely decline, new cases of mesothelioma, which has a latency of up to 35 years or more, are not estimated to plateau until 2020.

Chronic beryllium disease declined in frequency and severity after the elimination of beryllium from fluorescent light bulbs in the 1950s, but then increased because of the increasing use of beryllium in nuclear facilities, aerospace, electronics, dental ceramics, metal alloys, recycling of metals, and products such as golf clubs and bicycles. The beryllium lymphocyte proliferation test can identify beryllium sensitization, which can be found in up to 10% of exposed workers and facilitates earlier diagnosis of chronic beryllium disease.

SPECIFIC OCCUPATIONAL LUNG DISORDERS

Occupational lung diseases are often misdiagnosed as other common nonoccupational diseases, but a careful history and appropriate investigations can lead to a correct diagnosis. For many occupational lung diseases, the diagnosis can significantly improve prognosis and lead to measures to prevent illness in other workers.

Work-Related Asthma

Work-related asthma includes both occupational asthma that is caused by work[2] and asthma that is not caused by work but is exacerbated by work exposures.

SENSITIZER-INDUCED OCCUPATIONAL ASTHMA

EPIDEMIOLOGY

Occupational asthma is most commonly associated with a specific immune response to a high- or low-molecular-weight sensitizer (Table 93-2). Sensitizer-induced occupational asthma usually affects no more than 5 to 10% of workers exposed to the sensitizing agent, but exposure to complex platinum salts or detergent enzymes may result in symptoms in about 50% of highly exposed workers. In most studies, higher levels of exposure are associated with higher rates of sensitization in the exposed populations, but there is no clear threshold exposure below which all workers are protected from the risk for sensitization.

PATHOBIOLOGY

Genetic factors increase the risk for sensitization, but the risks appear to be polygenic and may differ for different allergens and sensitizers. Underlying atopy, as exemplified by a history of allergy or positive skin tests to common environmental allergens (Chapter 249), carries an increased risk for sensitization to the high-molecular-weight allergens, and smoking (Chapter 32) has been reported as a risk factor for sensitization to complex platinum salts. Currently, no host factors are sufficiently specific to justify exclusion of workers from settings with exposure to potential sensitizers.

Occupational asthma from a high-molecular-weight allergen is associated with specific immunoglobulin E (IgE) antibody production. Low-molecular-weight sensitizers may act as haptens or may induce neoantigens by reacting with proteins in vivo, but specific IgE antibodies have been demonstrated with only a few low-molecular-weight sensitizers, such as complex platinum salts and acid anhydrides used in epoxy compounds.

CLINICAL MANIFESTATIONS

Sensitizer-induced occupational asthma has a latency period ranging from weeks to several years before it develops, but most patients develop symptoms within the first few years of exposure. After a patient has become sensitized and has developed asthma, even very small subsequent exposures can trigger asthma, sometimes including exposures that may be below the limit of measurable detection. Pulmonary function and histologic changes are similar to those in nonoccupational asthma (Chapter 87). Sensitizer-induced occupational asthma from a high-molecular-weight agent sensitizer typically causes a prompt asthmatic response within minutes after exposure with or

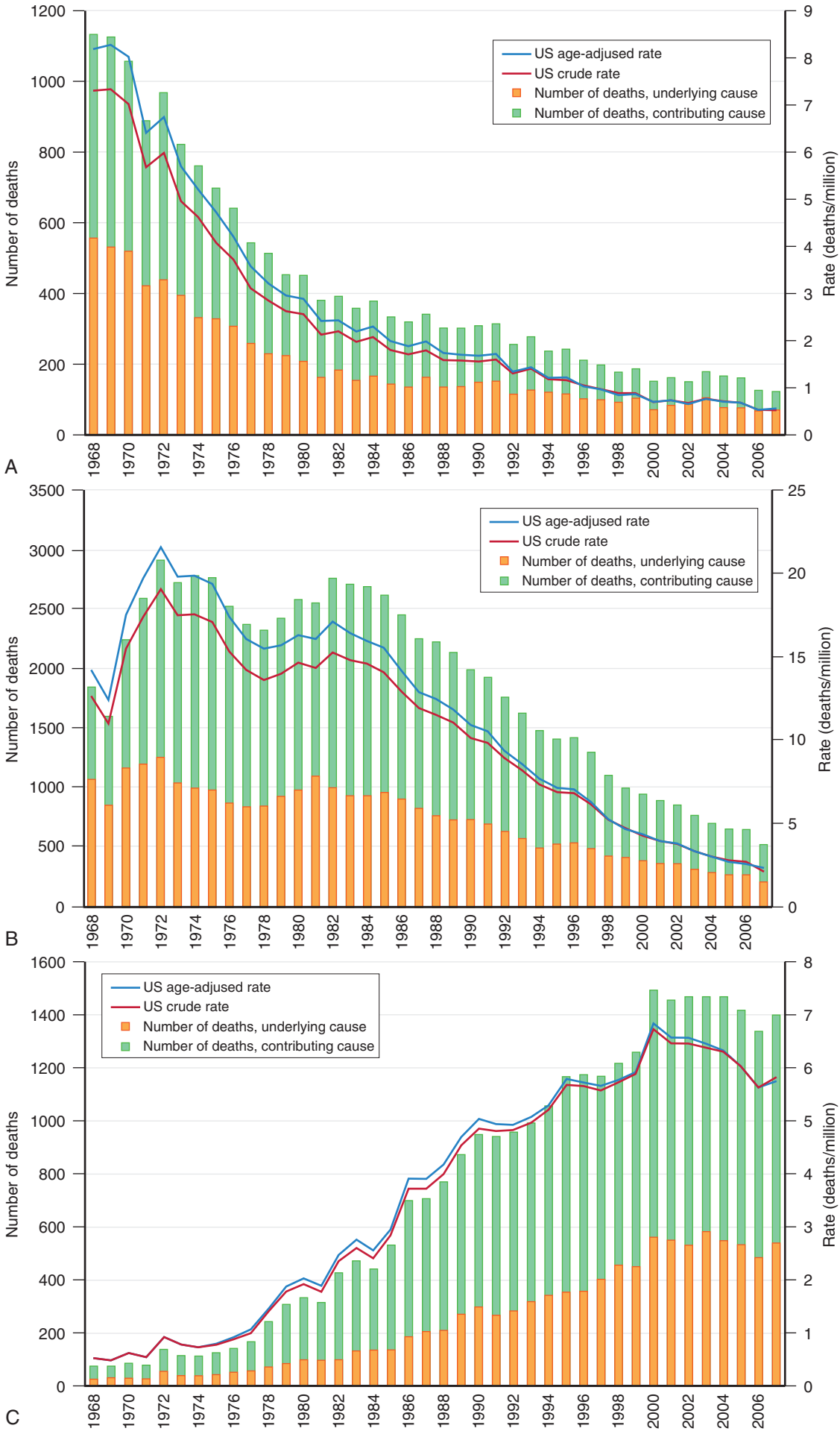

E-FIGURE 93-1. Changing patterns of mortality from silicosis (**A**), coal workers' pneumoconiosis (**B**), and asbestosis (**C**)—United States, 1968-2007. NIOSH [2012]. Work-Related Lung Disease Surveillance System (eWoRLD). [Figure 2012F01-01, F02-01, F03-01, Sections 1-3] U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Division of Respiratory Disease Studies, Surveillance Branch. Retrieved August 5, 2014, from http://www2a.cdc.gov/drds/WorldReportData/.

TABLE 93-1 EXAMPLES OF OCCUPATIONAL RESPIRATORY DISEASES THAT COULD BE MISDIAGNOSED AS COMMON NONOCCUPATIONAL RESPIRATORY DISEASE

DISEASE THAT IS MIMICKED	POSSIBLE OCCUPATIONAL DISEASE	EXAMPLES OF SUGGESTIVE FEATURES LEADING TO A CORRECT DIAGNOSIS
Asthma	Occupational asthma from a work sensitizer	Asthma symptoms begin and are worse during a working period, with some improvement on days or weeks off work. Exposure to a high- or low-molecular-weight workplace sensitizer
	Occupational asthma—irritant induced, including reactive airways dysfunction syndrome	Asthma begins within days after a high-level (accidental) workplace exposure
	Work-exacerbated asthma	Asthma usually began before starting the job or exposure, but severity is worse on days of work, or work exposures to expected asthma triggers or common allergens at work.
COPD	Occupational COPD	Prolonged exposure at work to dusts, fumes, or gases
Pneumonia	Acute hypersensitivity pneumonitis	Symptoms typically resolve within days and recur on re-exposure to the same work trigger (e.g., metal-working fluid, moldy hay, humidifiers)
Acute viral respiratory illness or pneumonia	Humidifier fever, organic dust toxic syndrome, metal fume fever, polymer fume fever, cotton dust fever	Exposure triggers the episodes
Sarcoidosis	Chronic beryllium disease	History of exposure to beryllium dust or fumes up to 30 years or more before onset of disease
	Silicosis	History of exposure; typical radiographic findings of rounded opacities with upper lobe predominance and progressive massive fibrosis, biopsy
Idiopathic pulmonary fibrosis	Asbestosis	History of moderate or high previous asbestos exposure and appropriate latency period, often with other markers of asbestos exposure, such as radiographic evidence of pleural plaques
	Chronic hypersensitivity pneumonitis	± Work exposure to a known trigger, ± improvement during periods away from exposure
	Flock-worker's lung	Lymphocytic bronchiolitis and interstitial lung disease from nylon/synthetic textile microfibers
Idiopathic pulmonary fibrosis or alveolar proteinosis	Indium lung	Exposure to indium-tin oxide in making of flat screens
Idiopathic pulmonary fibrosis or hypersensitivity pneumonitis	Hard metal disease	History of exposure to hard metal (tungsten, cobalt), and histologic findings of giant cell pneumonitis on lung biopsy
Chest infections	Occupational causes of chest infections, e.g., SARS or TB in health care workers, histoplasmosis in construction workers, anthrax in wool workers or farmers	History of occupation and exposures
Pleural effusion	Asbestos-related benign pleural effusion	Previous asbestos exposure with appropriate latency; pleural plaques commonly present
Incidental pulmonary nodule	Rounded atelectasis from asbestos	Previous asbestos exposure with appropriate latency; pleural plaques commonly present
Multiple nodules	Silicosis or pneumoconiosis	History of exposure, distribution of nodules, presence of progressive massive fibrosis
Lung cancer	Occupational lung cancer	History of exposure to carcinogens at work, with an appropriate latency period (e.g., asbestos, radon, chromium)
Bronchiolitis obliterans	Popcorn lung Bronchiolitis in military personnel	History of working with microwave popcorn or flavorings History of deployment in South-East Asia with exposure to burn-pits and other irritants

COPD = chronic obstructive pulmonary disease; SARS = severe acute respiratory syndrome; TB = tuberculosis.

without a late asthmatic response starting 4 to 6 hours after exposure. By comparison, responses to low-molecular-weight sensitizers typically start 4 to 6 hours after exposure.

DIAGNOSIS

The diagnosis of sensitizer-induced occupation asthma is clinically suspected by history and should be considered in all cases of new-onset asthma in patients who work. Supportive features include symptomatic improvement when away from work, such as weekends off work or holidays, but not necessarily in the evenings after a work shift, when symptoms from a late asthmatic response may occur. In patients who are exposed to high-molecular-weight sensitizers, allergic rhinitis or conjunctivitis associated with work frequently appears before the development of asthma. A detailed occupational history (Chapter 19) or review of material safety data sheets or occupational hygiene reports may reveal a known occupational sensitizer. However, more than 300 occupational respiratory sensitizers are currently known, and new agents or exposures are reported each year; as a result, the absence of a recognized sensitizer does not exclude occupational asthma.

Although the history can be very helpful, the evaluation should always include objective pulmonary function testing (Chapter 85) to confirm asthma, either when the patient has symptoms or within 24 hours of the typical suspected work exposure (Fig. 93-1). Allergy skin-prick tests, blood samples, or both should be obtained to test for specific IgE antibodies to any relevant sensitizer if feasible. Serial monitoring of peak expiratory flow rates, symptom diaries, or use of rescue inhalers can provide supportive information. The results of a methacholine challenge test (Chapter 87) toward the end of a typical work week can help when compared with results after 10 days or more without exposure. A comparison of eosinophil counts in induced sputum at work and after a period away from exposure, showing higher levels when exposed, provides supportive diagnostic information. If the diagnosis is still in doubt, a carefully controlled specific inhalation challenge with the suspected workplace sensitizer can be performed. Each investigation can be falsely positive or negative, so a combination of investigations is advised while the patient continues to work until the diagnosis is confirmed. Given the specialized nature of many of these studies, consultation with a specialist is recommended. The main differential diagnosis for patients with confirmed asthma is the coincidental onset of asthma with subsequent work-exacerbated asthma. Other conditions, such as vocal cord dysfunction, may explain symptoms or may coexist with asthma and confound the diagnosis.

TABLE 93-2 COMMON CAUSES OF SENSITIZER-INDUCED OCCUPATIONAL ASTHMA

OCCUPATION	ALLERGEN BY SETTING
Bakers	Wheat, rye, fungal amylase in flour
Laboratory workers	Animal allergens, e.g., proteins in rat urine, mouse or rabbit dander
Detergent-making, medical instrument cleaning, pharmaceuticals or laboratory workers	Enzymes: e.g., *Bacillus subtilis*, pancreatic enzymes
Farmers	Grains, plant, and animal allergens; mites
Greenhouse workers and florists	Pollen, fungi, mites
Food workers	Airborne food allergens, e.g., powdered milk or eggs and vegetables
Some office workers	Fungal allergens in moldy or "sick" buildings
Health care workers	Latex allergens from gloves, glutaraldehyde, orthophthaldehyde, aerosolized medications
Factory or other industrial workers	Chemicals in spray paints, glues, polyurethane, coatings and spray insulation, adhesives
Electronic workers	Soldering flux with colophony

TREATMENT AND PREVENTION

The optimal management of patients with sensitizer-induced occupational asthma includes complete removal from further exposure to the sensitizer and cross-reacting agents combined with the usual pharmacologic approach to the treatment of asthma (Chapter 87) and advocacy for appropriate workers' compensation.[3] Consideration of the other exposed workers typically includes communication with the workplace or public health officials, in the hope that measures may be instituted to protect other workers from similar exposures and symptoms. Recommendations for primary prevention have been to reduce exposures to occupational sensitizers as far as possible, removing unnecessary sensitizing agents (e.g., removing high-powdered and high-protein latex gloves) and limiting exposures to sensitizers with occupational hygiene measures. Ongoing medical surveillance measures in the workplace may also be of value.

PROGNOSIS

Outcome is best if an early diagnosis results in removal from further exposure while asthma is relatively mild. Improvement may continue to occur up to 10 years after removal from exposure, but asthma does not completely resolve in most patients. For patients with occupational asthma from natural rubber latex, use of powder-free, low-protein latex gloves by coworkers and direct avoidance of natural rubber products by the sensitized worker result in improvement that is similar to the improvement in patients who are completely removed from work.

WORK-EXACERBATED ASTHMA

EPIDEMIOLOGY AND CLINICAL MANIFESTATIONS

Work-exacerbated asthma is defined as asthma that is not caused by work but is aggravated or exacerbated by work conditions.[4] Asthma may have been present before starting employment or may begin coincidentally during employment, but it is not caused by work. Work exposures that commonly exacerbate asthma include extreme temperature or humidity, exertion, dusts, fumes, and gases. Patients may be exposed at work to common environmental allergens (e.g., fungal allergens in an office setting or dust mites or animals in domestic settings) that exacerbate asthma in patients who are sensitized to these allergens. Symptoms of work-exacerbated asthma may occur transiently with an unusual work exposure (e.g., during renovation in a work building) or may occur on a daily basis (e.g., daily exposure to fumes while performing physical exertion in an industrial setting).

DIAGNOSIS AND TREATMENT

Transient work-exacerbated asthma is commonly diagnosed on the basis of the history of work exposures and the associated increase in asthmatic symptoms, medication requirements, or unscheduled physician visits. The recommended evaluation of patients with daily or frequent work exacerbations of asthma is similar to that for patients with suspected occupational asthma. Work-related changes in serial peak flow recordings mimic those seen in occupational asthma, but sputum eosinophil counts typically show less of a work-related increase than is observed with occupational asthma. If the workplace exposure includes a potential work sensitizer, immunologic testing or a controlled challenge exposure may confirm whether respiratory sensitization has occurred.

Management includes the same pharmacologic measures as for non-work-related exacerbations, including the optimization of pharmacologic asthma management (Chapter 87) and, when needed, adjusting the work exposures to avoid ongoing exacerbations. Occupational hygiene measures can reduce exposures, but some patients require a change in job description or work area. Workers' compensation may be available for some patients who miss work owing to work-exacerbation of asthma.

Irritant Exposure and Reactive Airways Dysfunction Syndrome

A high level of usually accidental exposure to an irritant agent can cause asthma. Although the clinical manifestations can be dramatic, irritant-induced occupational asthma represents a relatively small proportion of all occupational asthma. The most definitive criteria for this condition are those applied to the term *reactive airways dysfunction syndrome*: the onset of asthma symptoms within 24 hours of the exposure, generally severe enough to lead to an unscheduled physician visit; exposure to a single high-level irritant; asthma symptoms that persist for at least 3 months; pulmonary function testing that confirms asthma with a significant beneficial response to bronchodilators or a bronchoconstrictor response to a methacholine challenge; and the lack of preexisting lung disease or other conditions to explain the symptoms. When these criteria are not completely met (e.g., symptoms start later than 24 hours after exposure or resolve within weeks after exposure), the term *irritant-induced asthma* is commonly applied, recognizing that this diagnosis is less certain than reactive airways dysfunction syndrome. Chronic low-level exposures to cleaning products and other irritants also may precipitate asthma.

Irritant-induced asthma and reactive airways dysfunction syndrome may clear after weeks or months. Management is the same as for other causes of asthma (Chapter 87), although these patients are often less responsive to the usual pharmacologic treatment.

Occupational hygiene measures at the workplace should be improved to prevent similar future exposures. Affected patients may need a modified work environment to prevent subsequent exacerbations of asthma.

OCCUPATIONAL CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Chronic exposure to dusts, fumes, and gases can cause occupationally induced COPD, with pathophysiologic changes essentially identical to those seen in COPD that is related to smoking[5] (Chapter 88). Symptoms of chronic bronchitis, including chronic cough and sputum production, may occur with or without changes on pulmonary function testing. Causes include mineral dusts such as silica and organic dust exposures such as those of farmers and woodworkers; particulate matter in diesel exhaust fumes; and nitrogen oxides, ozone, and ultrafine particles in welding fumes.[6] Occupational exposures such as diacetyl in artificial flavorings and styrene can cause constrictive bronchiolitis.[7]

No specific diagnostic tests distinguish an occupational from a nonoccupational cause of COPD. The history of exposure, with objective documentation, is helpful. Confirmation of the absence of a smoking history can assist in determining probability of an occupational cause. However, a positive smoking history does not exclude an occupational contribution because the two can be synergistic.

Management is the same as for patients with nonoccupational COPD (Chapter 88). In addition, however, further exposure to dusts, fumes, and gases that are likely to worsen disease should be minimized.

HYPERSENSITIVITY PNEUMONITIS

Many exposures that lead to hypersensitivity pneumonitis (Chapters 92 and 94) occur in the workplace, and several bear the name of the occupation or

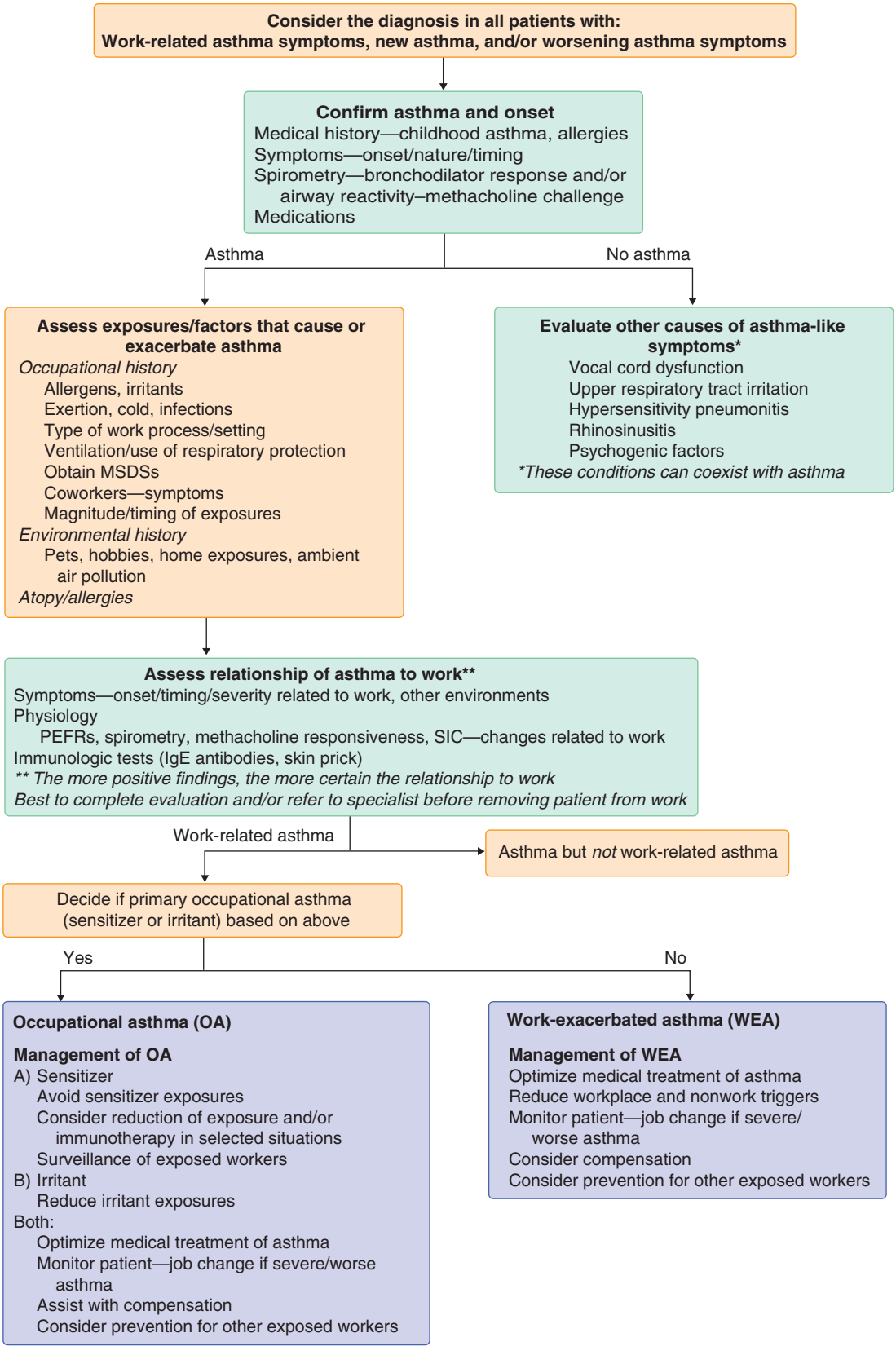

FIGURE 93-1. Clinical evaluation and management of work-related asthma. IgE = immunoglobulin E; MSDS = Material Safety Data Sheet; PEFR = peak expiratory flow rates; SIC = specific inhalation challenge. (From American College of Chest Physicians. Consensus statement: diagnosis and management of work-related asthma. *Chest.* 2008;134:1S-41S.)

job associated with them: farmer's lung; maple bark stripper's lung; cheese washer's lung; thatcher's lung; mushroom worker's lung, and metal-working fluid hypersensitivity pneumonitis (Table 93-3).[8] Occupational causes also include exposure to contaminated humidifiers (with protozoa or fungi) in factories or office buildings (humidifier lung), as well as lifeguards exposed to sprays of water from pool fountains contaminated with microorganisms. Metal workers or machinists can develop hypersensitivity pneumonitis from recirculated coolants that can become contaminated with gram-negative bacteria or atypical mycobacteria, which then are aerosolized in the coolant mist and inhaled.

PATHOBIOLOGY

Most antigenic exposures that lead to hypersensitivity pneumonitis are organic, especially thermophilic actinomycetes (Chapter 329), fungi,

TABLE 93-3 EXAMPLES OF OCCUPATIONAL CAUSES OF HYPERSENSITIVITY PNEUMONITIS

OCCUPATION	CAUSE
Farmer	Thermophilic actinomycetes in moldy hay
Metal worker	Contamination of metal-working fluids with microorganisms such as *Mycobacteria immunogens* or fungi
Worker exposed to humidifiers	Contamination with microorganisms such as protozoa or fungi
Sugarcane worker	Moldy sugarcane (bagassosis)
Maple bark stripper	Fungi
Chicken or turkey worker	Avian proteins
Pharmaceutical worker	Penicillin
Food handler	Soybeans
Office worker	Microorganisms contaminating air conditioners or humidifiers
Swimming pool attendant	Fungal contamination in sprays around pool area
Animal worker	Rat proteins
Mushroom worker	Fungi
Wheat farmer or handler	Weevil-infested flour
Greenhouse worker	Fungi
Workers spraying urethane paint or adhesives/sealants (or less often, other workers using diisocyanate)	Methylene diphenyl diisocyanate, hexamethylene diisocyanate, toluene diisocyanates
Chemical worker using plastics, resins, paints	Trimellitic anhydride

atypical mycobacteria (Chapter 325), and protozoa. Other common antigens include avian and rat proteins. Less commonly, hypersensitivity pneumonitis can be induced by low-molecular-weight chemical antigens, such as penicillin or methylene diphenyl diisocyanate (MDI), which is used as a sealant or binder. Small particles, commonly 3 to 5 µm in mass median aerodynamic diameter, reach the small airways and alveoli, where the immune response leads to hypersensitivity pneumonitis. This immune response is associated with specific IgG antibodies and T lymphocytes, and it recurs with repeated exposures.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The acute form of disease manifests as cough, dyspnea, chills, and malaise, typically occurring 4 to 8 hours after exposure and clearing by 12 to 24 hours. On examination, patients typically are febrile and tachypneic, with reduced chest expansion and basal crackles. Neutrophilia is common, and the chest radiograph shows acute infiltrates. Pulmonary function testing may show a restrictive pattern, with a reduced diffusing capacity, and arterial blood gases may show hypoxemia owing to ventilation-perfusion mismatch.

Chronic hypersensitivity pneumonitis may follow repeat acute episodes or start de novo. It causes a chronic dry cough, progressive dyspnea, and often, significant weight loss. The physical examination typically reveals reduced chest expansion and basal crackles. Results on pulmonary function testing and radiographic findings may be similar to nonspecific idiopathic pulmonary fibrosis (Chapter 92), and ground-glass opacities are often seen on a computed tomography (CT) scan of the chest. Bronchoalveolar fluid typically shows an increase in the lymphocyte count, and there may be a predominance of CD8 T lymphocytes (Chapter 85).

The specific occupational cause for hypersensitivity pneumonitis may be suspected from a temporal relationship to work exposures. The differential diagnosis in the chronic form includes idiopathic pulmonary fibrosis, although clubbing is less common in hypersensitivity pneumonitis. Radiographic and pulmonary function test findings may also mimic idiopathic pulmonary fibrosis, but a distinguishing finding is often a bronchoalveolar lavage that shows lymphocytes as high as 60 to 80% of the cells, usually with a predominance of $CD8^+$ T lymphocytes but sometimes with $CD4^+$ cells in chronic forms of disease.

Laboratory investigations include determining the presence of serum IgG antibodies to the suspected antigen. However, IgG antibodies may also be present in exposed individuals who do not have disease and are therefore not specific to the diagnosis. Conversely, failure to demonstrate specific antibodies is not uncommon in hypersensitivity pneumonitis because the limited number of antigens used for testing may not include the relevant occupational antigen. Specific challenge with the suspected antigen in a laboratory setting is occasionally needed if the diagnosis is in doubt.

Some patients can safely undergo "work challenge" that monitors changes in symptoms, fever, blood neutrophil count, radiographic findings, and pulmonary function with and without exposure to the suspected agent. Lung biopsy, if performed, may show granulomas and foreign body giant cells. If other findings are supportive of hypersensitivity pneumonitis, however, open biopsy and challenges usually are not needed.

TABLE 93-4 POTENTIAL EXPOSURES TO BERYLLIUM

OCCUPATIONAL EXPOSURES
Metal and alloy production (alloys of aluminum, copper, and nickel; recently includes golf clubs and metal pen clips)
Ceramic manufacturing
Metal casting, including dental technicians (crowns, bridges)
Electronics, including computer components, transistors, microwave and x-ray windows, heat sinks, telecommunications
Aerospace and atomic engineering (rocket fuels, heat shields, nose cones, and metal parts)
Aircraft manufacture and repair
Nuclear reactors, nuclear weapons, and defense industry
Coating of cathode ray tubes for radar and similar installations
Laboratories
Extraction from ore
Metal reclamation and recycling
NONOCCUPATIONAL EXPOSURES
Family members exposed to dust from workers' clothing
Breakage of old fluorescent lamps (made before 1950 in North America)
Downwind exposure from industrial accidents (e.g., from a nuclear processing plant in Kazakhstan, in the former Soviet Union in 1990)

From Tarlo SM, Rhee K, Powell E, et al. Marked tachypnoea in siblings with chronic beryllium disease due to copper-beryllium alloy. *Chest.* 2001;119:647-650.

TREATMENT AND PROGNOSIS

Rx

Treatment principles are the same as for nonoccupational hypersensitivity pneumonitis (Chapter 94). Removal from exposure to the causative agent is the primary treatment measure. As with occupational asthma from a sensitizer, the removal must be complete and often requires a change in work if the causative agent cannot be removed. Reduction of exposure by use of respiratory protective devices is generally not practical and not effective, with the exception of air-supplied helmet respirators for occasional short-term exposures. Patients with acute hypersensitivity pneumonitis may not require any medications in addition to removal from antigen exposure, but if acute episodes are severe, they may need supportive measures, including corticosteroids (e.g., 20 to 60 mg of prednisone orally per day), supplemental oxygen, and intensive care (Chapter 94). Chronic hypersensitivity pneumonitis may require additional oral corticosteroid treatment (e.g., 5 to 10 mg of prednisone orally per day) as for nonoccupational chronic hypersensitivity pneumonitis, and severe end-stage fibrosis may lead to need for lung transplantation. Prognosis is better with early diagnosis and complete removal from exposure to the causative agent. Preventive measures include occupational hygiene measures to avoid contamination of aerosolized fluid or dusts with bio-organisms and use of appropriate respiratory protective devices.

CHRONIC BERYLLIUM DISEASE

EPIDEMIOLOGY AND PATHOBIOLOGY

Acute toxic pneumonitis was described in workers who had high exposure to beryllium in the manufacture of fluorescent light bulbs in the 1940s, and a hypersensitivity response causing chronic beryllium disease was described in the 1950s. Acute toxic effects are now rare, but chronic beryllium disease remains a problem because of the expanded use of beryllium (Table 93-4) and better recognition of sensitization by development of an immunologic blood test.

Chronic beryllium disease is a hypersensitivity disease with a strong genetic association with HLA-DPB1 gene variants that code for Glu69 and

that have been identified in 83 to 97% of patients with disease. However, this gene variant occurs in 30 to 48% of the general population and, as a result, is not useful as a screening test.

CLINICAL MANIFESTATIONS

The pulmonary clinical features of chronic beryllium disease are similar to those of sarcoidosis (Chapter 95), ranging from asymptomatic histologic or radiographic findings, to potential progression, to severe granulomatous restrictive lung disease. Onset can occur up to 20 years or more after exposure to beryllium, even if the patient no longer is exposed. The clinical history in all patients with apparent sarcoidosis must include inquiry about possible beryllium exposure, even many years ago.

DIAGNOSIS AND TREATMENT

Rx

The chest radiograph shows changes that appear identical to sarcoidosis with enlarged hilar or mediastinal lymph nodes or multiple lung nodules, or both (Fig. 93-2). Sensitization to beryllium can be detected by a beryllium lymphocyte proliferation test that demonstrates the presence of sensitized lymphocytes in blood or bronchoalveolar lavage fluid. This test also can detect sensitization to beryllium among asymptomatic exposed workers, who can then be evaluated to assess possible chronic beryllium disease and provided with advice for reducing or eliminating further work exposures.

After disease develops, removal from exposure is advised, but the disease may still worsen. Progressive deterioration in lung function is treated similarly to sarcoidosis (Chapter 95), with oral corticosteroids and supportive measures.

ASBESTOS-RELATED DISEASES

Although the use of asbestos has declined, and better protective equipment has been mandated, asbestos-related disease has continued to occur owing to the long latency between exposure and disease. Chrysotile asbestos has less effect on the lungs than other forms of asbestos, that is, amphiboles. Effects of exposure include benign and malignant disease.

Benign asbestos disease is often asymptomatic and identified on chest imaging. Pleural thickening and pleural plaques, commonly with calcification, can occur 20 to 30 years after first exposure and may initially appear on the chest radiograph as calcified linear opacities over the hemidiaphragms and cardiac border (see Fig. 84-14). If extensive, it may be difficult to exclude intrapulmonary opacities except by CT scan. Pleural plaques are a marker of asbestos exposure but do not occur in all workers with significant asbestos exposure. They generally do not cause significant changes in lung function, except that diffuse pleural thickening may result in exertional dyspnea and extrapulmonary restrictive lung disease. Pleural thickening may cause rounded atelectasis (Chapter 90) when encasement of a portion of the peripheral lung tissue by thickened pleura causes an apparent lung nodule, typically with a "comet sign" showing the thickened pleura. Benign pleural effusion can develop, typically about 10 to 15 years after asbestos exposure. It requires further investigation because the differential diagnosis includes malignant pleural effusion (Chapter 99).

Asbestosis is the term for interstitial lung disease caused by asbestos. The clinical presentation is usually with dry cough and dyspnea on exertion. Physical examination usually reveals digital clubbing and basal crackles on lung auscultation. Chest imaging shows basal interstitial lung disease, with or without additional pleural changes as described earlier. Pulmonary function testing shows restrictive lung disease (Chapter 85), and histologic findings are the same as in usual interstitial pneumonia (Chapter 92). Findings supporting the diagnosis of asbestosis rather than usual interstitial pneumonia include a significant duration and level of exposure to asbestos, an appropriate latency of usually 20 to 40 years after first exposure, and the finding of ferruginous asbestos bodies in sputum or lung tissue (Fig. 93-3). Unfortunately, pharmacologic treatment is not effective, and the lung disease may progress to end-stage fibrosis. Management is supportive, including supplemental oxygen and consideration for lung transplantation (Chapter 101). As with the other diseases of long latency, preventing exposure is paramount.

Mesothelioma (Chapter 191), a malignant tumor of the pleura, peritoneum, or both, is the one complication of asbestos exposure that can occur after even relatively minor exposure, such as second-hand exposure from dust on clothing in the families of those working with exposure. It typically occurs 30 to 40 years after exposure to asbestos and may present incidentally on chest imaging or with chest pain or weight loss. Radiographs show pleural thickening, and a pleural effusion may be present. Mesothelioma often is difficult to distinguish from benign pleural thickening without a biopsy. No treatment has proved effective (Chapter 191), so routine screening to detect mesothelioma in exposed persons is not currently recommended. The risk for lung cancer (Chapter 191) increases after significant exposure to asbestos, with a usual latency period of 20 to 30 years. Smoking and asbestos exposure have additive effects, whereas smoking and asbestosis have even greater effects on the risk for lung cancer.

SILICOSIS AND OTHER PNEUMOCONIOSES

The incidences of silicosis and other inorganic dust diseases of the lungs (Table 93-5) have declined substantially in recent decades owing to better worksite protection in mines, sandblasting, and other settings. There is an association between silicosis and the development of collagen vascular disease, especially rheumatoid arthritis. Patients with pneumoconiosis and rheumatoid arthritis may be at higher risk for developing rheumatoid nodules in the lung, so-called Caplan syndrome, and mycobacterial infections.

Patients may initially be identified incidentally during a medical surveillance program or by a chest radiograph that shows multiple small lung nodules, often with enlarged mediastinal lymph nodes that can mimic sarcoidosis (Fig. 93-4).[9] Nodules can coalesce and lead to progressive massive fibrosis, especially in the upper lungs, sometimes mimicking malignancy, and can be positive on PET scanning due to their metabolic activity. There can be compensatory emphysema in the lower lung fields. On chest imaging,

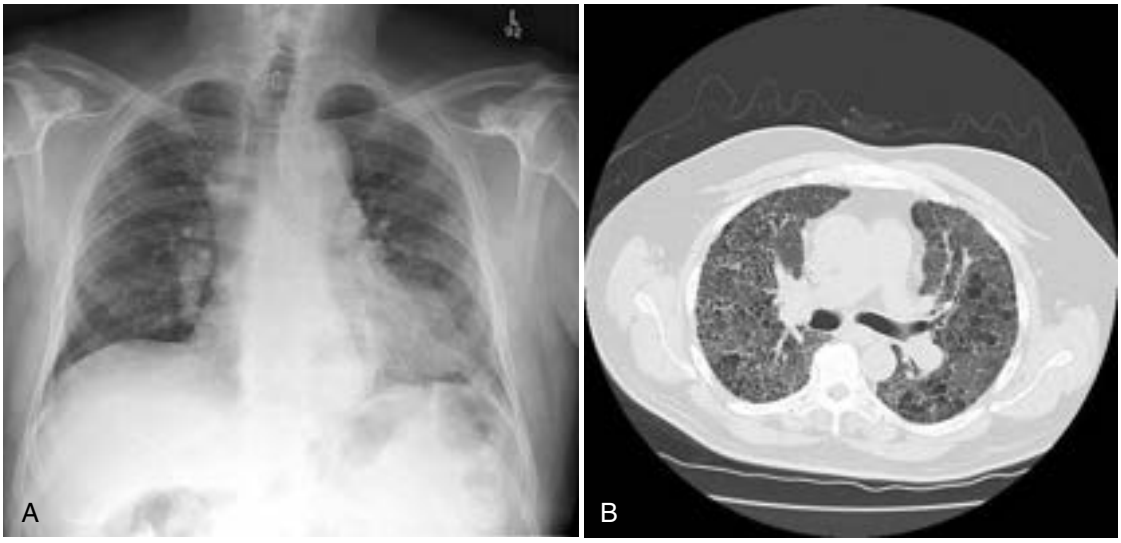

FIGURE 93-2. Posteroanterior chest radiograph (**A**) and high-resolution computed tomography scan (**B**) from patients with chronic beryllium disease. The chest radiograph demonstrates hilar adenopathy and infiltrates, and the scan shows air space destruction and infiltrates.

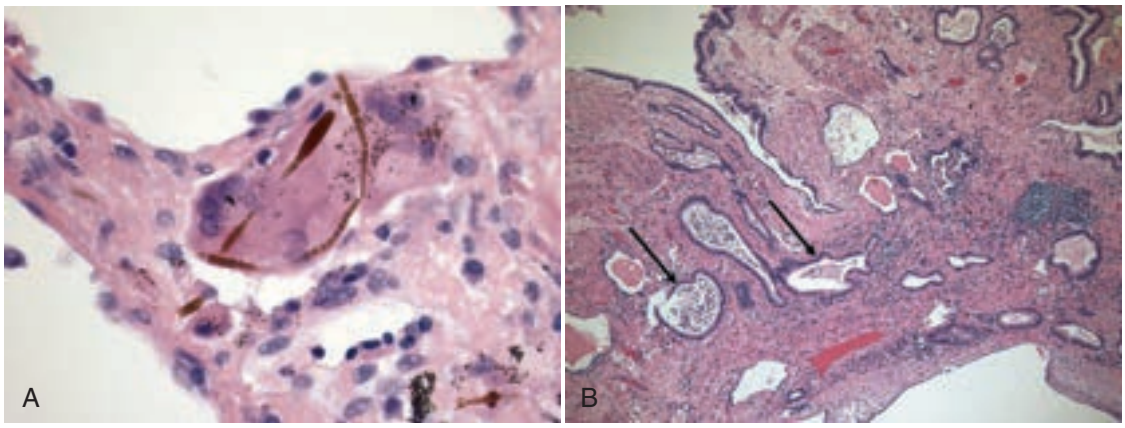

FIGURE 93-3. Histology from a lung biopsy showing asbestos bodies. Ferruginous bodies consisting of asbestos fibers coated by iron-protein-mucopolysaccharide material with typical golden-brown, beaded appearance. The two longest asbestos bodies at the center of the figure are present within a multinucleated giant cell. (Hematoxylin and eosin stain, ×400). (Courtesy of Dr. David Hwang, Toronto General Hospital.)

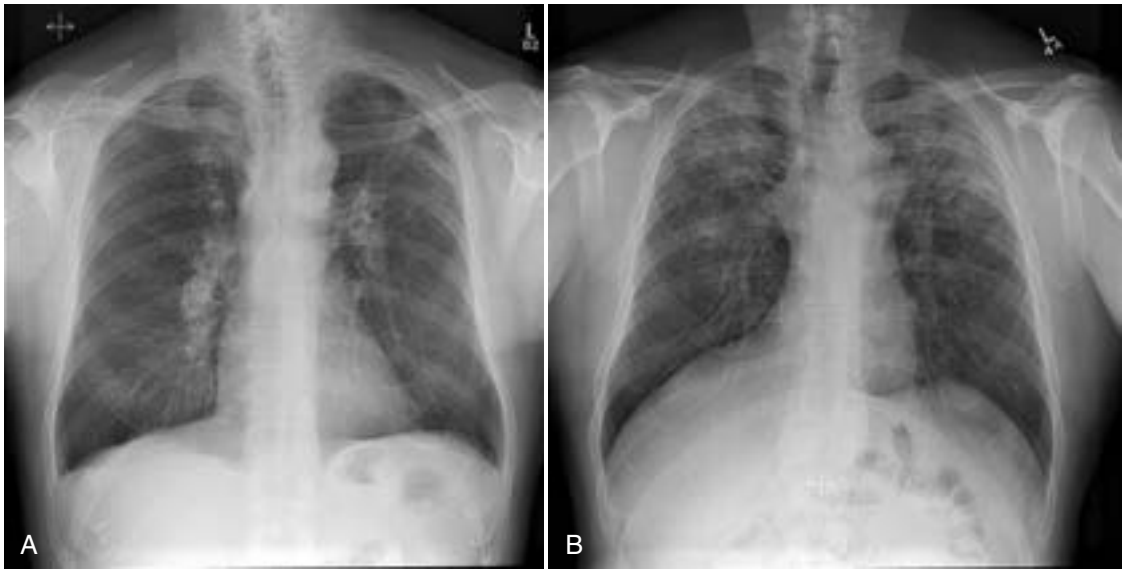

FIGURE 93-4. Posteroanterior chest radiographs from two patients with silicosis. A, Small nodules and eggshell calcification of hilar lymph nodes. **B,** Progressive massive fibrosis of the upper lung zones with compensatory emphysema.

TABLE 93-5 JOBS THAT CAN LEAD TO SILICOSIS

Mining: surface or underground mining (tunneling)
Milling: ground silica for abrasives and filler
Quarrying
Sandblasting: e.g., of buildings, preparing steel for painting
Pottery; ceramic or clay work
Grinding, polishing using silica wheels
Stone work
Foundry work: grinding, molding, chipping
Refractory brick work
Glass making: to polish and as an abrasive
Boiler work: cleaning boilers
Manufacture of abrasives

TABLE 93-6 OCCUPATIONAL CAUSES OF AN ACUTE FEBRILE SYNDROME

SYNDROME	CAUSE
Polymer fume fever or Teflon fever	Polytetrafluoroethylene and other fluorocarbon polymer fumes
Metal fume fever	Zinc fumes from welding of galvanized steel, less commonly other metal fumes
Cotton mill fever	Dust and endotoxins from bacterial contamination of unprocessed cotton, flax, and hemp
Humidifier fever	Microorganisms found in reservoirs, e.g., humidifiers, air conditioners, aquariums
Organic dust toxic syndrome	Grain dust, moldy wood chips

mediastinal lymph nodes may have a characteristic "eggshell" calcification in silicosis. Treatment is supportive. Patients with exposure to silica or coal dust may develop COPD from the dust exposure or dust-related diffuse fibrosis.[10] Patients who develop end-stage lung disease may be considered for lung transplantation.

ACUTE FEBRILE SYNDROMES

A variety of occupational exposures can cause acute febrile respiratory syndromes that may mimic acute viral respiratory illnesses (Table 93-6). The mechanism of these syndromes is incompletely understood, but they are associated with systemic neutrophilia and cytokine activation, often with increased interleukin-6 (IL-6) and IL-8.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Typically, chills, fever, malaise, dry cough, and chest tightness start about 6 to 8 hours after onset of an exposure at work and generally resolve by the next day. Occasionally, shortness of breath and other respiratory symptoms are severe enough for patients to seek emergency medical attention. Infiltrates

on the chest radiograph can occur with neutrophilia and hypoxemia that can mimic acute pneumonia or acute hypersensitivity pneumonitis. Symptoms and signs generally resolve in 24 to 48 hours without antibiotics and recur with further exposures, although the clinical manifestations generally become milder with repeated daily exposures (e.g., Monday morning fever in cotton mill workers). Workers are often familiar with the syndrome because it commonly affects up to 30% of exposed workers. If the diagnosis is not provided by the patient, however, careful elicitation of potential work exposures is needed.

TREATMENT

Rx

Treatment is supportive. If the causative exposure can be removed (e.g., cleaning a contaminated humidifier), symptoms can be prevented. If the cause cannot be removed and symptoms are severe, the patient may need reduction or change of the work exposure.

OCCUPATIONAL LUNG CANCER

A significant duration and level of exposure to a recognized carcinogen such as asbestos, hexavalent chromium (as in chromate production and the pigment industry), soluble radon compounds or radon gas, polycyclic aromatic hydrocarbons, chloromethyl ethers, arsenic, or silica can increase the risk for lung cancer (Chapter 191). Such a history should be elicited in all patients, and exposure to these agents represents a risk factor when considering whether to recommend patients for CT screening for lung cancer. The International Agency for Research on Cancer provides a listing of occupational lung carcinogens and the likelihood of their association with cancer.[11]

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. National Institute for Occupational Safety and Health Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Division of Respiratory Disease Studies, Surveillance Branch. Work-Related Lung Disease Surveillance System (eWoRLD). [Section 1-3]. 2013. http://www2a.cdc.gov/drds/WorldReportData/. Accessed February 3, 2015.
2. Tarlo SM, Lemiere C. Occupational asthma. *N Engl J Med*. 2014;370:640-649.
3. Baur X, Sigsgaard T, Aasen TB, et al. Guidelines for the management of work-related asthma. *Eur Respir J*. 2012;39:529-545.
4. Henneberger PK, Redlich CA, Callahan DB, et al. An official American Thoracic Society statement: work-exacerbated asthma. *Am J Respir Crit Care Med*. 2011;184:368-378.
5. Mehta AJ, Miedinger D, Keidel D, et al. Occupational exposure to dusts, gases, and fumes and incidence of chronic obstructive pulmonary disease in the Swiss Cohort Study on Air Pollution and Lung and Heart Diseases in Adults. *Am J Respir Crit Care Med*. 2012;185:1292-1300.
6. Omland O, Wurtz ET, Aasen TB, et al. Occupational chronic obstructive pulmonary disease: a systematic literature review. *Scand J Work Environ Health*. 2014;40:19-35.
7. Kreiss K. Occupational causes of constrictive bronchiolitis. *Curr Opin Allergy Clin Immunol*. 2013;13:167-172.
8. Lacasse Y, Girard M, Cormier Y. Recent advances in hypersensitivity pneumonitis. *Chest*. 2012;142:208-217.
9. Leung CC, Yu IT, Chen W. Silicosis. *Lancet*. 2012;379:2008-2018.
10. Petsonk EL, Rose C, Cohen R. Coal mine dust lung disease. New lessons from old exposure. *Am J Respir Crit Care Med*. 2013;187:1178-1185.
11. World Health Organization. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. 2013. http://monographs.iarc.fr/ENG/Monographs/PDFs/index.php. Accessed February 3, 2015.

REVIEW QUESTIONS

1. Which one of the following statements is *not* correct?

A. Occupational asthma is defined as the new-onset of asthma caused by work.
B. Occupational asthma is defined as asthma that is transiently worsened by exposures at work.
C. Irritant-induced asthma is a form of work-related asthma that includes reactive airways dysfunction syndrome as the most definitive subtype.
D. Asthma that is worsened but not caused by work is termed work-exacerbated asthma.
E. Occupational asthma can be associated with immunoglobulin E (IgE) antibodies to the causative exposure agent

Answer: B Occupational asthma is defined as asthma caused by work. Asthma that is transiently worsened at work is termed work-exacerbated asthma (Tarlo SM, Balmes J, Balkissoon R, et al. Diagnosis and management of work-related asthma: American College of Chest Physicians Consensus Statement. *Chest.* 2008;134:1S-41S.). Occupational asthma can be caused by an IgE-mediated response to a workplace sensitizer, but the exact mechanism is unclear for many chemical causes of occupational asthma. High-level irritant exposures can also cause occupational asthma. See Work-Related Asthma section.

2. A former soldier who had been deployed to Afghanistan on three occasions complains of shortness of breath on exertion. His pulmonary function tests and chest radiograph are normal. Which one of the following work-related diagnoses should be considered and further investigated?

A. Pulmonary hypertension
B. Constrictive bronchiolitis
C. Chronic obstructive pulmonary disease (COPD)
D. Asbestosis
E. Irritable larynx syndrome

Answer: B Lung biopsy findings of constrictive bronchiolitis have been reported among military personnel with dyspnea on exertion after return from deployment in Asia, often without significant abnormalities on pulmonary function testing or chest imaging. The probable cause is fumes from burn-pits, desert storms, or other inhaled toxic exposures (Kreiss K. Occupational causes of constrictive bronchiolitis. *Curr Opin Allergy Clin Immunol.* 2013;13:167-172).

3. A 40-year-old man has findings of interstitial pulmonary fibrosis. He has worked in a facility making screens for smart phones. Which of the following tests is *not* likely to be helpful in determining whether his lung disease has resulted from his work?

A. History of alveolar proteinosis among coworkers
B. Review of material safety data sheets to determine possible exposure to indium-tin oxide
C. Blood indium assay
D. Review of lung biopsy for characteristic changes
E. Pulmonary function tests

Answer: E Indium-tin oxide has caused alveolar proteinosis and pulmonary fibrosis in a subset of exposed workers. Lung biopsy findings have shown alveolar proteinosis or cholesterol clefts. See Table 93-1. (Cummings KJ, Nakano M, Omae K, et al. Indium lung disease. *Chest.* 2012;141:1512-1521).

4. A 30-year-old man comes to the emergency department with acute fever, malaise, dry cough, chest tightness, and neutrophilia. The chest radiograph is normal. Which of the following is *not* a cause for an acute febrile syndrome?

A. Stainless steel welding
B. Cotton dust
C. Grain handling
D. Baking
E. Polymer fume work

Answer: D The other exposures can all cause an acute febrile illness, typically occurring after several hours of exposure and typically worse early in the work week but improving later in the week. See Table 93-6.

94 PHYSICAL AND CHEMICAL INJURIES OF THE LUNG

DAVID C. CHRISTIANI

SUBMERSION INCIDENTS: DROWNING

DEFINITION

Drowning is defined as the process of experiencing respiratory impairment from submersion/immersion in liquid. The term *near-drowning* was previously used to describe individuals who survived a submersion incident, at least temporarily, but it has been abandoned on the basis of recommendations of the First World Congress of Drowning in Amsterdam in 2002.

EPIDEMIOLOGY

The estimated annual number of deaths worldwide due to drowning is 500,000. About 4200 persons are treated per year for nonfatal drowning in U.S. emergency departments, and about another 3400 suffer fatal drowning. Alcohol use, age younger than 4 years, and male gender are associated with increased rates of both nonfatal and fatal drowning.

PATHOBIOLOGY

The initial response to submersion/immersion is apnea, followed almost invariably by aspiration.[1] Laryngospasm may result in aspiration of a variable quantity of liquid medium into the lungs. Hypoxemia, hypercapnia, and acidemia develop acutely. Aspiration of either fresh or salt water results in occlusion of the airway, reduced surfactant activity, direct alveolar injury, and bronchospasm. Acute lung injury or the acute respiratory distress syndrome (ARDS)—associated with noncardiogenic pulmonary edema, respiratory failure, and severe hypoxemia—may develop hours or days after the incident. Acute renal failure may also occur. Alcohol consumption also increases the risk for hypothermia. Changes to serum electrolytes with drowning in either fresh water or salt water are not clinically significant.

The most serious secondary consequence of hypoxemia is anoxic brain injury. Fortunately, a reduction of brain temperature by 10° C during drowning decreases adenosine triphosphate (ATP) consumption by approximately 50%, thereby doubling the duration of time that the brain can survive. Mortality is primarily due to the cardiovascular sequelae of severe early or late hypoxemia.

CLINICAL MANIFESTATIONS

The initial presentation of a drowning victim varies widely. Hypothermia, which is common in drowning victims, may be associated with bradycardia or cardiac arrest due to asystole or ventricular fibrillation. Tachypnea, tachycardia, and low-grade fever are typical in nonhypothermic patients. Cyanosis may be present, and a coughing patient may produce pink frothy sputum. Neurologic evaluation may reveal agitation with or without intoxication or coma. The patient should be examined carefully for signs of associated trauma.

Expected laboratory findings include mild electrolyte abnormalities independent of whether submersion occurs in salt water or fresh water, moderate leukocytosis, and slight decrease in hematocrit in the first 24 hours or slight increase in free hemoglobin with a stable hematocrit in fresh water submersion due to hemolysis, severe hypoxemia, and metabolic acidosis. Evidence of disseminated intravascular coagulation (DIC) may occur. Initial electrocardiographic changes include sinus tachycardia and nonspecific ST segment and T wave changes, which revert to normal within hours. Life-threatening ventricular arrhythmias, complete heart block, or evidence of myocardial infarction can occur early or late in the course. Chest radiographs may initially be normal, despite severe respiratory impairment. Bilateral patchy alveolar infiltrates, which indicate progression to ARDS, may develop.

DIAGNOSIS

The diagnosis of drowning is made on clinical history of submersion in liquid medium with resulting respiratory impairment. Patients with unusual presenting circumstances should be carefully examined for evidence of trauma or assault.

PREVENTION

Drowning incidents are largely preventable, particularly in children. Pool fencing is a proven, effective strategy to prevent drowning. The primary cause of drowning of infants and toddlers is lack of adult supervision, and supervision of all young children near any form of water is strongly recommended. The role of alcohol in teenage and adult drowning incidents is substantial, and all individuals participating in water-based activities should restrict alcohol intake. The use of personal flotation devices is recommended for children and adults.

TREATMENT

Rx

When the victim has been recovered from submersion, treatment should focus on basic life support, including notification of emergency response personnel, establishment of an adequate airway, and cardiopulmonary resuscitation, if necessary (Chapter 63). If the victim is apneic, rescue breathing should occur immediately, even before removal from the water. Cervical spine stabilization is needed if there is a history of diving, use of a water slide, signs of injury, or signs of alcohol intoxication. Spinal cord injury (Chapter 399) is otherwise unlikely, and cervical spine stabilization techniques and equipment may impede timely and effective treatment. Attempts to remove water from the airway are unnecessary. Cardiac arrhythmias should be treated with Advanced Cardiac Life Support protocols, including the use of automated external defibrillators when appropriate (Chapter 63). A majority of drowning victims who receive cardiopulmonary resuscitation or rescue breathing will vomit; if vomiting occurs, the head should be turned to the side, and any visible vomitus remaining in the oral cavity should be removed with a finger. When vomiting occurs in patients who may have spinal cord injury, log-rolling techniques are recommended for turning the patient to the side.

All victims of a submersion incident should be transported to a hospital for further evaluation, treatment of potential respiratory failure (Chapter 104), and monitoring for up to 24 hours. Bronchoscopy may be required to evaluate localized wheezing or persistent atelectasis. Prophylactic antibiotics are not useful, but evidence of pneumonia (Chapter 97) should be treated with

appropriate antibiotics. Because unusual microorganisms may be isolated from the lower airways, efforts should be made to identify specific microbial flora pertinent to the locus of the drowning incident. Randomized controlled trials of specific ventilator strategies have not been conducted, but lung-protective ventilation (i.e., a tidal volume of 4 to 6 mL/kg ideal body weight) with sufficient positive end expiratory pressure of 5 to 10 cm H_2O to avoid atelectasis is commonly recommended.

Treatment of neurologic injury is focused on supportive care while the extent of cerebral edema is minimized. To decrease cerebral edema and intracranial pressure, intravenous hypertonic saline (7.5 to 23% to target serum sodium values of 145 to 155 mmol/L) or mannitol (1g/kg bolus of 20% mannitol, with repeat dosing every 6 to 8 hours as necessary to target serum osmolality of less than 300-320 mOsm/kg) may be used, although their benefit in this setting is unproved. Hyperventilation to a $Paco_2$ of 34 to 36 mm Hg may be helpful. Intracranial pressure may increase in response to shivering or purposeless movements, which should be reduced. Induced therapeutic hypothermia to 36°C improves neurologic outcome after cardiac arrest (Chapter 109), and current recommendations for drowning victims who remain comatose after rescue are to avoid rewarming to core or tympanic temperatures above 34°C and to maintain temperature of 32° to 34°C for 24 to 48 hours. Hyperthermia should be avoided at all times.

PROGNOSIS

The mortality rate for drowning victims who present alive to an emergency department is about 25%. Long-term neurologic deficits persist in approximately 6% of nonfatal drowning victims. Prolonged duration of submersion is associated with a worse prognosis, and the risk for death or severe permanent neurologic deficits increases from 10% after less than 5 minutes of submersion, to about 55% with 6 to 10 minutes of submersion, nearly 90% with 11 to 25 minutes of submersion, and nearly 150% with more than 25 minutes of submersion.[1] However, young children who are hypothermic when they are rescued after submersion times of up to 60 minutes have recovered without neurologic damage. Other factors associated with poor prognosis include hypotension, persistent apnea, coma, more than a 10-minute delay in receiving basic life support, and duration of resuscitation of more than 25 minutes.

DISEASES OF HIGH ALTITUDE

DEFINITION

Neurologic and pulmonary disturbances, primarily due to direct tissue effects of hypoxia, occur in individuals who either ascend to or reside at altitudes of 7000 feet (2133 meters) or more (Table 94-1).[2]

EPIDEMIOLOGY

Acute mountain sickness is the most common high-altitude syndrome. It occurs in approximately 20% of individuals who ascend to altitudes of 7000 to 9000 feet, 40% at 10,000 to 14,000 feet, and more than 50% above 14,000 feet. The incidence of chronic mountain sickness, also known as Monge disease, is thought to be between 5 and 18%. More severe neurologic disturbances due to high-altitude cerebral edema are rare, occurring in approximately 1 to 2% of individuals who ascend to altitudes above 15,000 feet. High-altitude pulmonary edema occurs in approximately 2 to 6% of otherwise healthy individuals who ascend to altitudes of 8000 to 15,000 feet. However, the incidence in individuals with a prior history of high-altitude pulmonary edema may be as high as 60%, or higher during rapid ascents. The occurrence of high-altitude retinal hemorrhage is approximately 33% among individuals who ascend to very high altitudes (up to 19,000 feet) and is thought to be common at lower altitudes as well. High-altitude retinal hemorrhage is not associated with high-altitude cerebral edema or long-term visual consequences.

TABLE 94-1 HIGH-ALTITUDE SYNDROMES

SYNDROME	CLINICAL DESCRIPTION
Acute mountain sickness	Common after recent ascent to altitudes above 7,000 feet; symptoms include headache, anorexia, and malaise
Chronic mountain sickness	Occurs in 5-18% of people dwelling above 10,000 feet; symptoms include headache, fatigue, dyspnea, and digestive disturbances
High-altitude pulmonary edema	Occurs in 2-6% of people above 9500 feet; symptoms include dyspnea, cough, and tachycardia
High-altitude retinal hemorrhage	Common above 15,000 feet; asymptomatic or reversible vision changes
High-altitude cerebral edema	Occurs in 1-2% of people above 15,000 feet; symptoms include confusion, ataxia, hallucinations, coma, or death

PATHOBIOLOGY

Clinically significant hypoxemia is the underlying factor in all high-altitude diseases. The decrease in barometric pressure during an ascent to altitude causes a decrease in the alveolar pressure of oxygen (PAo_2). For example, PAo_2 drops from 105 mm Hg at sea level to 60 mm Hg at 10,000 feet and to 40 mm Hg at 18,000 feet. Below 60 mm Hg, oxygen dissociates from hemoglobin more readily (see Fig. 158-2), thereby decreasing oxygen saturation and oxygen delivery to tissues. The effect is even more noticeable in patients with impaired diffusion capacity, such as occurs in emphysema, interstitial lung diseases, or heart failure. Furthermore, increased ventilatory drive induces an acute respiratory alkalosis. In the brain, tissue hypoxia causes cerebral vasodilation, whereas hypobaria causes cerebral vasoconstriction. In severe hypoxemia, vasodilation is the likely cause of cerebral edema in susceptible individuals. The response to hypoxemia in the lungs is primarily increased pulmonary arterial pressures due to hypoxic pulmonary vasoconstriction, which results in reversible injury to pulmonary capillaries, increased capillary permeability, and eventually pulmonary edema. Reduced oxygen consumption also occurs, perhaps because of impaired mitochondrial function. Sleep disordered breathing has also been noted but has minimal if any clinical significance.

CLINICAL MANIFESTATIONS

Symptoms of acute mountain sickness begin 2 to 3 hours after ascent and include breathlessness, lightheadedness, fatigue, nausea, anorexia, headache, and insomnia. Most symptoms resolve within 2 to 3 days, although insomnia may persist. Chronic symptoms of headache, fatigue, sleep disturbances, dyspnea, and digestive complaints are seen with chronic mountain sickness in individuals residing at higher elevations. Chronic mountain sickness may be associated with polycythemia (hemoglobin concentrations above 21 g/dL). Severe neurologic symptoms with high-altitude cerebral edema include ataxia and confusion that may progress to coma or death. Symptoms of high-altitude pulmonary edema that usually begin 2 to 4 days after ascent to higher altitudes include dyspnea, cough, and tachycardia. Fundoscopic changes of flame-shaped hemorrhages are seen with high-altitude retinal hemorrhage.

DIAGNOSIS

Diagnosis of most high-altitude disease is made on the basis of clinical manifestations at high altitude. Diagnosis of chronic mountain sickness, and a milder form often termed subacute mountain sickness, is more challenging because it may mimic other cardiopulmonary, neurologic, or psychiatric disease. Individuals with chronic mountain sickness typically have higher hemoglobin concentrations, higher serum erythropoietin levels, higher nocturnal heart rates, lower nocturnal oxygen saturation, and higher systolic and diastolic arterial pressure than do normal individuals living at similar altitude.

PREVENTION

Prevention of high-altitude disease can be achieved by avoidance in high-risk individuals, such as young children and persons with a history of high-altitude disease. Gradual ascent and acclimatization are crucial to prevention of high-altitude illness, particularly at extreme altitudes. At altitudes up to 10,000 feet, 2 to 3 days or more may be needed for adjustment to the effects of hypoxemia. For mountaineers, current recommendations are to ascend no more than approximately 984 feet (300 meters) per day at altitudes higher than 9843 feet (3000 meters).

When rapid ascent is unavoidable, such as flights to high-altitude locales, the carbonic anhydrase inhibitor acetazolamide (125 mg orally twice daily) provides effective prophylaxis of acute mountain sickness, and 125 mg at night may improve sleep (Table 94-2). Other medications shown to be effective in the prevention of acute mountain sickness in randomized, controlled trials include ibuprofen (600 mg three times daily), sumatriptan

TABLE 94-2 RECOMMENDED MEDICATION FOR THE PREVENTION AND TREATMENT OF ALTITUDE SICKNESS

MEDICATION	INDICATION	ROUTE	DOSE
Acetazolamide	Prevention of AMS, HACE	Oral	125 mg twice per day Pediatrics: 2.5 mg/kg every 12 hr
	Treatment of AMS*	Oral	250 mg twice per day Pediatrics: 2.5 mg/kg every 12 hr
Dexamethasone	Prevention of AMS, HACE	Oral	2 mg every 6 hr or 4 mg every 12 hr Pediatrics: should not be used for prophylaxis
	Treatment of AMS, HACE	Oral, IV, IM	AMS: 4 mg every 6 hr HACE: 8 mg once then 4 mg every 6 hr Pediatrics: 0.15 mg/kg/dose every 6 hr
Nifedipine	Prevention of HAPE	Oral	30 mg SR version every 12 hr, *or* 20 mg of SR version every 8 hr
	Treatment of HAPE	Oral	30 mg SR version every 12 hr, *or* 20 mg of SR version every 8 hr
Tadalafil	Prevention of HAPE	Oral	10 mg twice per day
Salmeterol	Prevention of HAPE	Inhaled	125 µg twice per day†

*Acetazolamide can also be used at this dose as an *adjunct* to dexamethasone in HACE treatment, but dexamethasone remains the primary treatment for that disorder.
†Should not be used as monotherapy and should only be used in conjunction with oral medications.
AMS = acute mountain sickness; HACE = high altitude cerebral edema; HAPE = high altitude pulmonary edema; SR = sustained release; IV = intravenous; IM, intramuscular.

(50 mg orally once after ascent),[■] dexamethasone (2 mg every 6 hours or 4 mg every 12 hours),[■] and prednisolone (20 mg daily). For the prevention of high-altitude pulmonary edema in susceptible individuals, options include the long-acting β-adrenergic agonist salmeterol (125 µg inhaled twice daily), the phosphodiesterase inhibitor tadalafil (10 mg twice daily), the calcium-channel blocker nifedipine (20 mg of slow-release formulation every 8 hours, or 30 mg of slow-release formulation every 12 hours), and dexamethasone (8 mg twice daily).

TREATMENT

Rx

For acute, life-threatening symptoms such as high-altitude pulmonary edema and high-altitude cerebral edema, the best treatment is immediate descent, if possible, combined with supplemental oxygen therapy and, if needed, use of a portable hyperbaric chamber. For individuals who develop less severe symptoms, sildenafil (50 mg, one time dose) can increase exercise capacity at altitude by 10 to 35%.[■] Increasing inspired oxygen concentrations in high-altitude working facilities also improves productivity and quality of sleep. Sustained-release theophylline (300 mg daily) significantly reduces the symptoms of acute mountain sickness compared with placebo, and acetazolamide (250 mg twice daily) is useful in treating symptoms of acute as well as chronic mountain sickness.[■] Milder forms of acute mountain sickness, such as headache, can be treated with typical doses of nonsteroidal anti-inflammatory medications, including aspirin or acetaminophen.

PROGNOSIS

Symptoms of high-altitude disease respond rapidly to immediate descent. However, high-altitude cerebral edema and high-altitude pulmonary edema can be fatal, particularly at extreme altitudes and weather when descent may be impossible.

DECOMPRESSION ILLNESS: DECOMPRESSION SICKNESS, BAROTRAUMA, AND ARTERIAL GAS EMBOLISM

DEFINITION

Exposures to changes in ambient pressure cause a spectrum of illness, either by increasing or decreasing the volume of gas in air-filled body cavities or by causing the release of inert gas bubbles from solution in tissues or blood vessels. Symptoms associated with decreasing ambient pressure, which occur most commonly with ascent from depth during recreational or occupational diving, are known as decompression illness. The most common form of decompression illness is decompression sickness, which is classified as either type I (mild symptoms, such as general fatigue or joint pain) or type II (more severe neurologic or cardiopulmonary disturbances). Life-threatening forms of decompression illness include pulmonary barotrauma and arterial gas embolism syndromes. During the descent of a dive, increasing ambient pressure may cause mild symptoms of facial or sinus pain, often called "the squeezes."

EPIDEMIOLOGY

In addition to the approximately 9 million recreational divers in the United States, aviators, astronauts, and compressed air workers are also exposed to changes in ambient pressure that may cause decompression illness. Among recreational divers, the annual incidence rate for either type I or type II decompression sickness is estimated at 1 case per 5000 to 10,000 dives. Approximately 1000 episodes of decompression illness severe enough to warrant recompression therapy occur each year, up to 10% of which are fatal. Well-recognized risk factors for decompression illness include long duration of dives, deep dives, repetitive dives, heavy exertion at depth, cold water, and rapid ascent. Additional risk occurs in individuals who experience further decreases in ambient pressure after the dive, such as on commercial or private aircraft or driving over mountainous areas.[3]

PATHOBIOLOGY

The principles of Boyle's law and Henry's law describe the properties of gases during changes in ambient pressure. Boyle's law states that the volume of a gas varies inversely to changes in pressure, $P_1V_1 = P_2V_2$. During the descent of a dive, pain due to the "squeezes" is caused by increasing ambient pressures that are not equalized by a compensatory increase in gas volume. The resulting negative pressure causes a vacuum effect in the mask associated with engorgement of the blood vessels in adjacent tissues, such as periorbital and ocular vessels, and may result in swelling, pain, and subconjunctival hemorrhages. Facial sinuses, the middle ear, and the external auditory canals may also be affected.

Barotrauma in sinus, otic, or pulmonary tissues may be due to changes in ambient pressures and the resulting increase or decrease in the volume of gas. During descent, the decreasing volume of gas causes vascular engorgement in the sinuses and otic compartments and may result in rupture of the tympanic or inner ear membranes. During ascent, breath holding, particularly with compressed air devices (SCUBA diving), and the presence of obstructive lung disease with delayed exhalation times and air trapping impair equilibration and increase the risk for pulmonary barotrauma (E-Figure 94-1). If the expanding volume of gas causes a pressure gradient between the alveoli and pulmonary interstitium that exceeds the compliance of the lung, alveolar rupture will lead to pulmonary interstitial emphysema. Further extension of gas along pulmonary tissues may cause additional barotrauma, leading to pneumothorax, mediastinal emphysema, pneumopericardium, and soft tissue emphysema.

Arterial gas embolism, which is a serious consequence of pulmonary barotrauma, results in the development of free gas in the pulmonary arterial circulation. The resulting bubbles may then enter the systemic circulation by overwhelming the filtering mechanism of the pulmonary capillaries or through a right-to-left intracardiac shunt (Chapter 69), such as a patent foramen ovale. Bubbles may then migrate to the brain, spinal cord, heart, lung, or kidney and lead to tissue ischemia or infarct.

Henry's law states that the solubility of a gas in liquid is proportional to the partial pressure of that gas above the liquid. An increase in partial pressure of gases during descent will therefore cause the amount of gas dissolved in the pulmonary capillaries to increase. Dissolved oxygen is used during normal body metabolism; however, inert nitrogen, which is abundant in inspired air, becomes dissolved in the blood and tissues, particularly in fat, where it is five-fold more soluble than in water. During ascent, decreasing ambient pressure causes tissues to become supersaturated with nitrogen, and nitrogen is subsequently released into blood vessels and tissues as gas bubbles. Decompression-induced gas bubbles cause decompression illness by either mechanical compression of tissues or embolization through blood vessels to end organs. Bubbles that obstruct capillaries or venules damage the endothelium and cause tissue ischemia, which leads to activation of inflammatory mediators or tissue reperfusion injury. Although not well understood, toxic

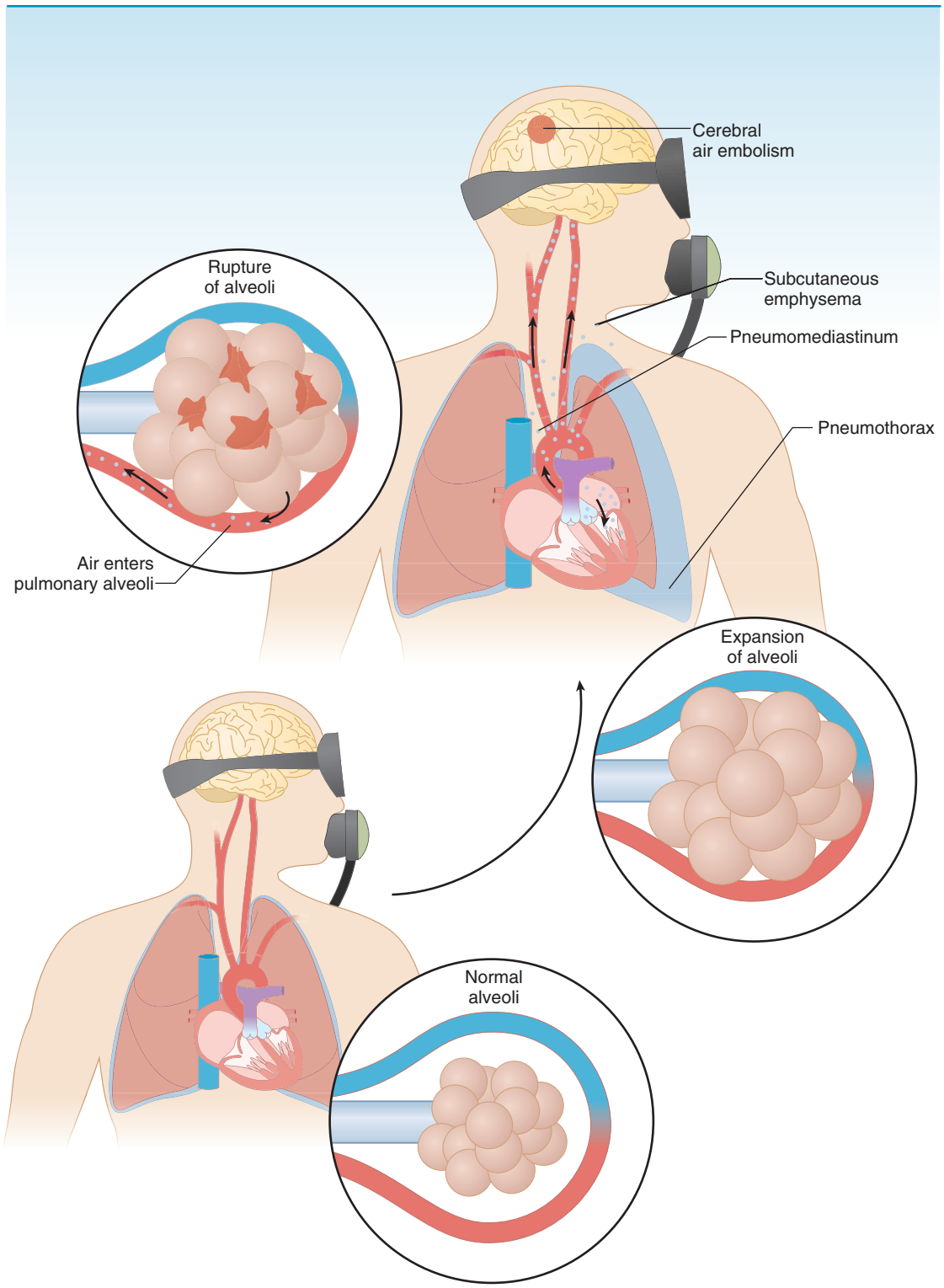

E-FIGURE 94-1. Pulmonary barotrauma in a diver during breath-holding ascent. (From Vann RD, Butler FK, Mitchell SJ, et al. Decompression illness. *Lancet.* 2011;377:153-164, Fig. 1, p. 154.)

effects due to increased partial pressure of gases also are likely to contribute to symptoms of decompression illness, possibly by denaturing of proteins and release of fatty acids from cell membranes.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Symptoms of decompression illness can occur within minutes and up to 24 hours or more after exposure to changes in ambient pressure associated with dives of 20 feet in depth or more. The severity of symptoms depends on the rate and the magnitude of the change of ambient pressure and can vary among individuals. Diagnosis is based on clinical manifestations, which can be classified according to whether they are caused by formation of inert nitrogen gas bubbles or the localized toxic effects of gas (associated with decompression sickness), barotrauma associated with descent (sinus or otic barotrauma), barotrauma associated with ascent (pulmonary barotrauma), or more severe arterial gas embolism syndromes.

Symptoms vary according to location of bubble formation. For example, type I decompression sickness, also known as the bends or caisson disease, is typically associated with pain in the joints, from mild to severe, and numbness of the extremities. Rashes and lymphedema may also occur. Symptoms of type II decompression sickness may be systemic (fatigue, hypovolemic shock), cardiopulmonary (cough, substernal chest pain, tachypnea, asphyxia), otic (vertigo, hearing loss), or neurologic (ataxia, aphasia, speech disturbances, incontinence, confusion, personality changes, depression, paralysis, and loss of consciousness).

Otic barotrauma, which typically occurs during descent, can affect the external, middle, or inner ear (Chapter 426). External ear symptoms, such as a sensation of ear fullness or otalgia, are caused by a blockage of the canal, for example, with the use of ear plugs or presence of cerumen. Middle ear symptoms of otalgia, vertigo, tinnitus, transient conductive hearing loss, and facial nerve palsy occur when inadequate equalization of pressures results from blocked eustachian tubes, typically in association with allergic rhinitis or upper respiratory infections. Inner ear barotrauma, which is a more serious form of otic barotrauma, is associated with elevated intracranial pressure and rupture of the inner ear membrane. Inner ear barotrauma causes symptoms of sensorineural deafness, tinnitus, vertigo, nausea, and vomiting. Sinus barotrauma typically occurs during descent, is associated with facial pain and epistaxis, and occurs more frequently in individuals with mucosal inflammation from allergies or infection.

Pulmonary barotrauma, which is the second leading cause of death among divers, should be suspected in postdive individuals, particularly at-risk individuals, with symptoms of sudden pleuritic pain, dyspnea, or coughing. Physical examination findings include tachypnea, subcutaneous emphysema, and dullness to percussion or decreased breath sounds over a pneumothorax. Development of tension pneumothorax (Chapter 99) or severe pneumomediastinum may lead to decreased venous return of systemic blood and reduced cardiac preload, a situation that is characterized by hypotension and may lead to refractory shock or cardiac arrest. Chest and neck radiographs are recommended for diagnosis, particularly because pneumothoraces must be treated with chest tube thoracostomy before recompression therapy.

Because arterial gas embolism syndromes are caused by pulmonary barotrauma, careful neurologic assessment is critical. The neurologic findings are similar to those of an acute stroke (Chapter 407), with manifestations of focal or unilateral motor deficits, visual disturbances, sensory deficits, speech difficulties, and cognitive disturbances, including loss of consciousness. Symptoms typically occur within 10 minutes after ascent. Delayed neurologic symptoms are more likely to be due to type II decompression sickness.

PREVENTION

Education is the most effective method of preventing decompression illness.[4] Before participation in diving-related activities, all individuals should undergo a thorough and intensive training program. Instruction of proper pressure equalization techniques is critical in the prevention of decompression illness. Persons with asthma who wish to dive should be assessed by a physician (preferably knowledgeable in the field of diving medicine), have no wheezing on physical examination, and have normal spirometry before and after exercise. The presence of structural lung disease (e.g., lung cysts or bullae) is associated with a significant increase in the risk for pneumothorax and is a contraindication to diving. Presence of a known right-to-left intracardiac shunt, such as a patent foramen ovale, is not an absolute contraindication to diving, although conservative diving is recommended, and patients should be cautioned that they are at increased risk for decompression illness.

TREATMENT

Rx

Symptoms of decompression illness at altitude should be treated with supplemental oxygen and return to the lowest attainable altitude. Serious decompression illness associated with diving requires immediate medical evaluation by emergency personnel, including basic and advanced life support (Chapter 63) when hemodynamic instability is present. Pneumothorax (Chapter 99) should be treated immediately with needle decompression or chest tube thoracostomy. Symptoms that persist for more than 2 hours or increase in intensity require recompression therapy, preferably with 100% oxygen or transfer to a facility with a hyperbaric chamber, where standard protocols should be followed. The only two randomized controlled trials to evaluate recompression therapy found that the addition of tenoxicam (20 mg daily, not available in the United States) or a helium-oxygen mixture (rather than pure oxygen) decreases the number of recompression treatments needed in divers with decompression illness, but neither improved the overall effectiveness of recompression treatment.

PROGNOSIS

Survival of patients with decompression illness depends on prompt medical evaluation and treatment. Immediate hyperbaric oxygen therapy per standard protocols is associated with resolution of symptoms in 95% of cases. However, symptoms of decompression illness, even neurologic deficits, may respond to recompression therapy after delays of 24 hours or more.

INHALATION INJURIES

Smoke Inhalation and Thermal Injury

DEFINITION

Pulmonary complications, largely caused by smoke inhalation, occur in a large proportion of burn victims (Chapter 111) and account for a substantial number of deaths in these patients. Even patients who do not sustain surface burns in a fire can inhale sufficient smoke to result in injury to the lungs or airways.

EPIDEMIOLOGY

Modern building codes and the widespread presence of firefighting personnel in communities have decreased the importance of fire as a cause of death in the United States. However, fire continues to cause several thousand deaths annually. Also, larger scale fires with mass casualties and wildfires that affect large geographic areas still occur occasionally.

PATHOGENESIS

Smoke loses heat rapidly as it traverses the upper airway, so direct thermal injury is often limited to the mucosa of the supraglottic airway. A notable exception is steam inhalation, which can produce thermal injury throughout the airways. Smoke inhalation injury affects the entire respiratory tract. The pathogenesis of smoke inhalation is complicated by the wide variety of pulmonary irritants in smoke, many of which are directly toxic to respiratory epithelial or alveolar cells: aldehydes such as acrolein, acetaldehyde, and formaldehyde; acids such as hydrochloric, hydrofluoric, and hydrocyanic acid; and ammonia, nitrogen oxides, and phosgene.

Irritants can rapidly induce intense neutrophilic inflammation, which evolves during 12 to 24 hours after injury and is characterized by mucosal edema and ulceration, abnormally increased permeability of pulmonary capillaries with resultant capillary leak, and epithelial, alveolar, and immune cellular dysfunction. Bronchospasm or bronchorrhea may occur, and the processes may result in ARDS. In addition, because oxygen is consumed in fires, breathing of hypoxic air for prolonged periods may potentiate other injuries or cause clinically significant hypoxemia in its own right.

CLINICAL MANIFESTATIONS

Thermal injury to upper airway mucosa can cause airway compromise, particularly due to laryngeal edema, sometimes rapidly and sometimes over the first 12 to 24 hours. Burns to the face, mouth, and neck can externally damage and distort structures of the upper airway and cause airway compromise, both subacutely and late in the course. Inhalation injury manifests primarily with bronchospasm and bronchorrhea, which cause cough, dyspnea, or wheezing and may progress rapidly to respiratory failure. Accumulation of secretions, failure of mucociliary clearance and immune mechanisms, and epithelial necrosis predispose to pulmonary infection, particularly 3 to 5 days

after injury. Late pulmonary complications can also be caused indirectly by eschar formation and restriction of thoracic motion.

DIAGNOSIS

Patients with apparent or suspected burn injuries (Chapter 111) should be assessed emergently for airway patency. Head or neck burns, respiratory distress, stridor, or visibly erythematous or edematous oral mucosa should prompt immediate laryngoscopic evaluation of the oropharynx and supraglottic airway. Hypoxemia may develop and may be severe enough to meet criteria for ARDS. Chest radiography should be performed serially to detect the evolution of lung injury or superinfection.

TREATMENT Rx

If airway patency is threatened, endotracheal intubation should be performed immediately. Delay can result in increased edema and greater technical difficulty of intubation. Patients who cannot be intubated should have emergent tracheostomy performed surgically. Because of the risk for ARDS, mechanical ventilation with a goal tidal volume of 4 to 6 mL/kg of ideal body weight should be considered. All patients should receive supplemental oxygen with the goal of providing a high fractional concentration of inspired oxygen (FIO_2) to reverse the effects of hypoxemia and carbon monoxide inhalation (see later). Preliminary data suggest a role for inhaled anticoagulants to mitigate the development of acute lung injury, although well-designed prospective trials are currently lacking.[5] Pulmonary toilet is essential to clear secretions in the face of bronchorrhea and epithelial sloughing. Because of the risk for superinfection, surveillance for infection should be vigilant, including diagnostic bronchoscopy if ventilator-associated pneumonia is suspected.

PROGNOSIS

Patients who survive burns and recover generally do not have long-term pulmonary sequelae. Tracheostomies placed at the time of injury can usually be removed later, unless airway structures are damaged or distorted. Impaired pulmonary function is uncommon but may be manifested as airway hyperresponsiveness that has been termed reactive airway dysfunction syndrome.

Carbon Monoxide Poisoning

DEFINITION AND EPIDEMIOLOGY

Carbon monoxide is a colorless, odorless gas produced by the combustion of carbon-based fuels. Because of the ubiquity of these substances, carbon monoxide inhalation is often coincident with smoke inhalation in fires or may occur accidentally in association with malfunctioning equipment or improper venting of emissions from heaters, stoves, combustion motors, or other similar devices. In addition, intentional inhalation of carbon monoxide is a method commonly used in suicide attempts. Carbon monoxide inhalation is the leading cause of death from poisoning (Chapter 110) worldwide.

PATHOBIOLOGY

Carbon monoxide readily diffuses across the alveolar-capillary interface and binds to hemoglobin with extremely high affinity. When the resulting carboxyhemoglobin molecule undergoes an allosteric change at oxygen-binding sites, the ability of bound oxygen to dissociate and to be delivered to peripheral tissues is greatly reduced. This tissue hypoxia can cause severe functional impairment and ischemic injury of oxygen-sensitive tissues, particularly in the brain and heart.

CLINICAL MANIFESTATIONS

Mild carbon monoxide intoxication may go unrecognized because the symptoms are nonspecific and may include headache, nausea, malaise, fatigue, and dizziness. With more severe intoxications, neuropsychiatric symptoms may range from minor disturbances in attention and cognition to agitation, confusion, hallucination, or, in the worst intoxications, seizures or frank coma. Physical findings, which are generally nonspecific, can include tachycardia or hyperthermia. The classic "cherry-red" skin thought to be associated with carbon monoxide intoxication is rarely seen. Other manifestations of severe intoxications may include lactic acidosis, cardiac dysfunction with arrhythmia or ischemia, pulmonary edema, and rhabdomyolysis.

DIAGNOSIS

A high index of suspicion is required for diagnosis because clinical findings are nonspecific. All patients known to have been involved in fires, suicide attempts, or other scenarios compatible with exposures should have arterial carboxyhemoglobin levels checked by co-oximetry. Although levels do not correlate well with clinical findings or risk for complications, symptoms generally occur at carboxyhemoglobin concentrations of 10% or higher.

TREATMENT Rx

All patients should be treated with 100% supplemental oxygen, which competes with carbon monoxide for hemoglobin-binding sites and gradually eliminates it from the blood. If patients require mechanical ventilation because of a depressed neurologic status or respiratory problems, 100% oxygen should be administered by endotracheal tube. Rigorously conducted randomized controlled trials are lacking, but most expert guidelines endorse treatment with hyperbaric oxygen at a pressure of 2.5 to 3.0 atm, which can increase the dissolved oxygen content of blood by more than 10-fold.[6] At least one treatment of approximately 2 hours should be considered in severe cases to reverse the effects of acute intoxication; three hyperbaric oxygen treatments within 24 hours of diagnosis reduce neurocognitive sequelae.

PROGNOSIS

The mortality rate is highly variable according to the severity of intoxication but can approach 30% in severe cases. Approximately two thirds of patients who survive acute intoxication will recover without sequelae. Many of the remainder will suffer from long-term neuropsychiatric symptoms, including cognitive dysfunction, abnormal mood or affect, memory disturbances, and other motor or sensory abnormalities, which can often occur within the first month but may be delayed for up to 6 to 9 months.

Cyanide and Other Gases

PATHOBIOLOGY

In addition to carbon monoxide and pulmonary irritants, cyanide gas may be formed when a number of commonly found substances, particularly plastics and textiles, are combusted. This gas is highly toxic and can rapidly cause morbidity and death by binding to cytochrome enzymes and inhibiting cellular respiration.

Other inhaled gases that can injure the lungs in occupational settings include ammonia, chlorine, nitrogen dioxide, organic dust, paraquat, phosgene (which has also been used as a chemical weapon), sulfur dioxide, and toxic metal fumes such as cadmium and mercury.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Cyanide intoxication typically includes shock, lactic acidosis, and coma; it can rapidly lead to death before results of laboratory studies are available. In the setting of possible exposure, an elevated venous oxygen saturation indicates that cyanide is preventing cells from extracting oxygen from arterial blood.

Other inhaled gases that can produce potent irritant responses include ammonia, chlorine, and nitrogen dioxide ("silo-filler's lung"). Phosgene is notable for its propensity to cause delayed symptoms, up to 24 hours after exposure. Other inhalants may produce an acute chemical pneumonitis with respiratory distress (Chapter 93). A diverse group of inhaled toxins can cause syndromes of inhalational fever, including heavy metal fumes, polymer fumes, and organic dust aerosols that contain thermophilic bacteria, gram-negative bacteria and their associated endotoxins, and fungal elements. These inhalations are characterized by fever and malaise with mild respiratory symptoms and are also notable for tachyphylaxis with repeated exposure (and thus referred to as Monday morning fever in some occupational settings).

TREATMENT Rx

Cyanide intoxication is treated by use of a Taylor Cyanide Antidote Package (Taylor Pharmaceuticals) that contains amyl nitrate gas ampules (one 0.3-mL ampule each minute until sodium nitrate infusion begins) for inhalation. This treatment is followed by intravenous administration of sodium nitrite (300 mg one-time dose; can give an additional 150 mg if symptoms return), which converts hemoglobin to methemoglobin by attracting bound and unbound cyanide. Finally, patients are treated with intravenous sodium thiosulfate (12.5 g one-time dose; can repeat half the original dose if symptoms return), which converts cyanide to less harmful thiocyanate ions. If carboxyhemoglobinemia is present or the patient has heart disease or lung disease, sodium

thiosulfate should be used alone because of the additive toxicity of methemoglobinemia.

Hydroxocobalamin (5 mg given intravenously once, can be repeated for a total of 10 mg), which directly binds cyanide, can be used in conjunction with sodium thiosulfate. It is thought to be safer than the amyl nitrate gas ampules because it does not cause methemoglobinemia and is better suited to prehospital care.

The mainstay for treatment of other irritating inhalations is to remove the patient immediately from the toxic environment and to provide supportive care for respiratory injury. Depending on the intensity and duration of exposure, most patients will recover completely without sequelae.

OXYGEN TOXICITY

DEFINITION

Hypoxemic respiratory failure often requires treatment with supplemental oxygen to maintain tissue oxygenation. In some settings, such as ARDS, patients may require high F_{IO_2} for prolonged periods to combat severe hypoxia. However, it has long been recognized that oxygen may be toxic to the lungs when it is present in concentrations higher than those found in ambient air.[7]

PATHOBIOLOGY

When the concentration of oxygen in the airways is high, formation of reactive oxygen species and free radicals is increased. Under normal circumstances, innate antioxidant mechanisms in airway epithelia and alveoli are sufficient to abrogate the effect of these molecules. However, under conditions of critical illness, prolonged exposure to increased concentrations of these toxins may overwhelm these defenses. Superoxide, hydrogen peroxide, and hydroxyl radicals may directly oxidize cellular components. Cellular damage potentiates inflammation and may be synergistic with inflammatory processes already underway in the diseased lung; the result can be alveolar edema, formation of hyaline membranes, hypoxemia, and progression to fibrosis and obliteration of alveolar and capillary structures. In addition, washout of nitrogen from air spaces can result in absorptive atelectasis if oxygen is removed by the circulation faster than it can be replenished by ventilation (especially in the setting of ventilation-perfusion mismatch). Hyperoxia can also worsen hypercapnia through multiple mechanisms, as occurs in patients with chronic obstructive pulmonary disease who suffer from carbon dioxide retention.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Although the exact levels of hyperoxia that cause lung injury are unclear, it appears to occur with exposure to F_{IO_2} of 50 to 60% after exposures as short as 6 hours in duration. Because of the high flow of supplemental oxygen required to deliver this F_{IO_2}, oxygen toxicity is observed primarily in mechanically ventilated patients being treated for hypoxemic respiratory failure. This level of exposure can cause a clinically detectable tracheobronchitis, demonstrable by symptoms of cough and dyspnea, as well as airway erythema that is visible macroscopically on bronchoscopy. This syndrome may impair mucociliary clearance and result in impaction of secretions, especially in conjunction with absorptive atelectasis.

Patients who may be susceptible to oxygen toxicity generally already have a significant degree of parenchymal injury from other processes. Thus, although some patients may appear to display a syndrome of worsening air space disease, atelectasis, consolidation, hypoxia, and diffuse alveolar damage, it is not clear when these changes are related to oxygen therapy or merely occur as part of the acute lung injury from other causes.

TREATMENT

Rx

Because the threshold level for oxygen toxicity is unknown, a general guideline for treatment of hypoxemic respiratory failure (Chapter 104) is that patients be ventilated with the lowest possible F_{IO_2} that is required to restore an acceptable oxygen saturation. A Sa_{O_2} of 90%, corresponding to Pa_{O_2} of 55 to 60 mm Hg, is generally considered the minimum acceptable level. Unfortunately, under conditions of severe hypoxia, such as ARDS, patients often require an F_{IO_2} approaching 100% to achieve this oxygen level. Maneuvers to improve oxygenation without increasing F_{IO_2} include paralysis with continuous infusions of neuromuscular blocking agents; inhaled pulmonary vasodilators; red blood cell transfusions to improve the delivery of oxygen; alternative ventilatory strategies, such as high-frequency oscillatory ventilation, airway pressure-release ventilation, inverse-ratio ventilation, and prone positioning; and alveolar recruitment maneuvers using positive end-expiratory pressure or transiently increased inflation pressures (Chapter 105).

PROGNOSIS

Patients who sustain oxygen toxicity in the setting of prior bleomycin exposure may be left with residual pulmonary fibrosis. In other patients, the incremental impact of oxygen toxicity on prognosis is unknown.

LUNG INJURY

Radiation Lung Injury

DEFINITION

Accidental or occupational radiation exposures (Chapter 20) are generally characterized by systemic toxicity that outweighs any injury to the lungs. As such, radiation lung injury refers to a pneumonitis that can progress to pulmonary fibrosis and that results from therapeutic use of ionizing radiation, usually in the treatment of malignant neoplasms.[8]

EPIDEMIOLOGY

As many as 50% of patients receiving thoracic radiation will display radiographic abnormalities after treatment; duration and dose of therapy affect the odds for development of lung injury. However, most of these patients will never have clinically significant radiation lung injury. For unclear reasons, the incidence of lung injury appears to vary by the type of underlying malignant disease and modality of treatment. The highest frequency is in lung cancer (10 to 20%).

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

The pathogenesis of radiation lung injury is often divided into three or four phases on the basis of time course. Typically, the early phase occurs immediately after exposure and is characterized by injury to alveolar cells, resulting in mild alveolitis, recruitment of inflammatory cells, capillary leak, and pulmonary edema. These changes are usually asymptomatic; patients do not usually come to clinical attention, although the chest radiograph will be abnormal if it is performed. In most patients, these changes resolve without progression within 1 to 3 months.

A minority of patients will progress to the next phase, in which alveolar cells desquamate and the air spaces fill with protein-rich fluid. In this phase, referred to as radiation pneumonitis, patients will complain of cough, dyspnea, and occasionally fever or pleuritic chest pain. Severe cases may present with hypoxemic respiratory failure. This phase generally resolves within 3 to 6 months after exposure and is followed by an organizing phase, in which alveolar edema resolves and the damaged alveoli heal. Clinically, patients generally show improvement in symptoms during this period. However, this phase is also characterized by fibroblast proliferation and deposition of collagen in the lung. In a minority of patients, this process will proceed unchecked and result in clinically significant fibrosis, with progressive loss of alveolar-capillary surface and development of restrictive lung disease.

DIAGNOSIS

Clinical history and radiographic evaluation are often sufficient for diagnosis of radiation lung injury as patients typically present with respiratory symptoms and opacities on chest radiography after undergoing radiation therapy. Radiography may show air space disease with alveolar filling or consolidation during the pneumonitis phase, which may progress to an interstitial pattern and eventual honeycombing with parenchymal distortion in the chronic phase. Because radiation characteristically causes injury only within directly affected lung tissue, radiography may show opacities that are well delineated and form straight lines that cross anatomically distinct regions of lung. This finding is rarely if ever seen in other conditions.

The differential diagnosis may include pneumonia, recurrence, and metastatic malignant disease. On occasion, invasive evaluation with bronchoalveolar lavage or even biopsy is required to exclude these possibilities if the clinical history and imaging do not provide a diagnosis.

TREATMENT AND PREVENTION

Rx

Corticosteroids, such as prednisone 1 mg/kg body weight per day for 2 to 3 weeks followed by a slow taper over several weeks or months, are the mainstay of therapy. Patients often show a dramatic response to treatment, and symptoms may recur after treatment is discontinued. Other immunosuppressive agents have been used successfully in case reports of patients who failed to respond to corticosteroids. Prophylactic amifostine, a cytoprotective agent, significantly reduces the risk for radiation pneumonitis but does not improve survival in patients who receive radiation treatment for lung cancer.[1] Ambroxol, a free-radical scavenger given at 90 mg three times daily for 3 months, decreases the production of harmful cytokines and lessens the decrement in the diffusion capacity for carbon monoxide (DLCO), a marker of interstitial lung damage, in patients with locally advanced lung cancer.[2] Dixiong decoction, a Chinese herbal preparation, has also been reported to decrease the incidence and severity of radiation pneumonitis in patients with non–small cell lung cancer.

PROGNOSIS

Within 2 years after initial exposure, progression will usually slow, and symptoms and lung function will stabilize or improve. After this time, further improvement or worsening is uncommon. In severe cases that become chronic, patients may develop features of advanced interstitial lung disease, including pulmonary hypertension and hypoxic respiratory failure.

Aspiration Injury

DEFINITION

Aspiration, which is defined as the inhalation of any nongaseous foreign substance into the lungs, generally refers specifically to the inhalation of gastric contents or secretions from the oropharynx. Aspiration is a common occurrence, and in most cases it resolves spontaneously without clinical manifestations. Clinically significant aspiration can range from acute pneumonitis and respiratory failure caused by a single massive aspiration to chronic symptoms of respiratory disease caused by recurrent small-scale aspiration. These syndromes may overlap with pneumonia that occurs when the lungs are exposed to bacteria from the gastrointestinal tract (Chapter 97).

PATHOBIOLOGY

The common element of clinically significant aspiration is impairment of normal airway protective mechanisms. Under normal circumstances, the airway is protected by the normal swallowing mechanism, the cough reflex, and the anatomy of the supraglottic airway. However, even healthy individuals experience microaspiration despite having functional protective mechanisms. These secretions are handled by normal pulmonary clearance mechanisms.

Any disturbance of these protective mechanisms can result in aspiration injury to the lungs. An altered level of consciousness can impair normal swallowing and suppress the cough reflex. Even in patients who are alert, neurologic injury can result in dysphagia and concomitant aspiration, as in patients who have bulbar neurologic deficits in association with ischemic stroke. Patients with altered airway or oropharyngeal anatomy, such as patients who have received surgical or radiation therapy for head and neck malignant neoplasms, may also be highly susceptible to aspiration of oral secretions.

The nature of the aspirated material is also important in determining whether an injury occurs. Materials with a pH lower than 2.5, such as acidic gastric contents, are much more likely to cause a significant chemical pneumonitis. Particulate matter also increases the likelihood for development of clinically significant inflammation. A large-volume aspiration with distribution throughout the lungs is more likely to produce an acute, severe pneumonitis.

When material has been aspirated into the lungs, the injury that occurs is similar to a chemical burn. Acid rapidly injures airway epithelial and alveolar cells; within hours, cells become dysfunctional and capillary leak occurs, resulting in profound noncardiogenic pulmonary edema. In severe cases, diffuse alveolar damage may result.

CLINICAL MANIFESTATIONS

The classic eponym applied to aspiration pneumonia is Mendelsohn syndrome, which refers to a single, large-volume aspiration of gastric contents followed by rapidly progressive hypoxemic respiratory failure that develops within hours. Patients may suffer from cough, dyspnea, fever, and respiratory distress. Physical examination may reveal diffuse crackles, wheezing, cyanosis, and hypotension. Chest radiography may show a pattern of alveolar filling with diffuse bilateral involvement or involvement of dependent regions, particularly the right lower lobe if the patient is upright at the time of aspiration. In many patients, this period of acute deterioration is followed by stabilization and resolution within 2 or 3 days. In other patients, deterioration may continue, and patients may meet clinical criteria for ARDS. If the volume of aspirated material is large enough, the initial aspiration may be sufficient to cause tracheal obstruction and asphyxiation.

In patients who initially improve, a small percentage will show further deterioration after 2 or 3 days. This deterioration should prompt an investigation for bacterial superinfection.

DIAGNOSIS

The clinical history and presentation are generally sufficient for diagnosis of aspiration pneumonitis. Bacterial pneumonia and other causes of ARDS also should be considered, as should cardiogenic pulmonary edema. Airway erythema and edema on bronchoscopy can be suggestive of aspiration. Bronchoalveolar lavage may help evaluate for the presence of bacterial infection.

PREVENTION

Prevention should focus on identification of patients who are at risk for aspiration and then use of strategies to minimize the risk. Patients with swallowing dysfunction or airway abnormalities can work with speech pathologists to learn effective strategies for swallowing. Patients who are unsuccessful or not suitable for this approach may benefit from tracheostomy or enteral tube feedings, which do not prevent microaspiration but can prevent large-volume aspiration. However, the risk for aspiration pneumonia is no different among patients with nasogastric tubes compared with patients who have percutaneous endoscopy gastrostomy tubes.[3] In hospitalized patients, particularly patients with an altered mental status due to illness or sedation, simple strategies such as avoidance of oral feeding and semirecumbent positioning can effectively reduce the risk for aspiration. Use of histamine-2 blockers or proton pump inhibitors (Chapter 139) can alter gastric pH to reduce the risk for injury from acidic secretions.

TREATMENT

Rx

Because of the acuity and severity of aspiration pneumonitis, immediate attention should be paid to maintaining a patent airway. The oropharynx and trachea should be suctioned to clear any potentially obstructing material, and endotracheal intubation should be performed if necessary (Chapter 105). Bronchoscopy is often performed to clear residual particulate or solid matter, but it cannot remove acidic secretions, which damage airways and parenchyma quickly and then are rendered neutral. Oxygen supplementation should be provided as needed for hypoxia. Corticosteroids have not been shown to be beneficial. Antibiotics should be reserved for patients who appear to have developed bacterial superinfection (Chapter 97).

PROGNOSIS

For patients with severe respiratory failure or ARDS, mortality can be high. In others, improvement should be expected within days. If the underlying factor that led to aspiration is irreversible, patients have an increased likelihood of recurrent episodes.

Lipoid Pneumonia

DEFINITION

Lipoid pneumonia is a chronic inflammatory reaction of the lungs to the presence of lipid substances. Exogenous lipoid pneumonia results from the aspiration of vegetable, animal, or (most commonly) mineral oils.

PATHOBIOLOGY

The most frequently implicated agent is mineral oil used as a laxative and to reduce dysphagia, either in clear liquid form or as petroleum jelly. Mineral oil is bland and, when introduced into the pharynx, can enter the bronchial tree without eliciting the cough reflex. It also mechanically impedes the ciliary

action of the airway epithelium. The risk for mineral oil aspiration is increased in debilitated or senile patients, in those with neurologic disease that interferes with deglutition, and in patients with esophageal disease. Mineral oil taken as nose drops to relieve nasal dryness can also cause lipoid pneumonia. Inhalation of mineral oil mist by airplane and automobile mechanics has also been implicated as a cause.

Mineral oils, which cannot be hydrolyzed in the body, provoke a chronic inflammatory reaction that may not become clinically overt until years later. In the alveolar spaces, macrophages accumulate and phagocytose the emulsified oil. Some macrophages disintegrate, releasing their lysosomal enzymes and oil. The alveolar septa become thickened and edematous, containing lymphocytes and lipid-laden macrophages. Oil droplets are seen in the pulmonary lymphatics and hilar nodes. Later, fibrosis develops, and the normal lung architecture is effaced. A single pathologic specimen may include both the early inflammatory and the later fibrotic picture, in keeping with repetitive aspirations during many months or years. Nodular lesions may grossly resemble tumor and be called paraffinomas.

CLINICAL MANIFESTATIONS

Most patients are asymptomatic and come to the physician's attention because of an abnormal chest radiograph. When patients are symptomatic, cough and exertional dyspnea are the most frequent complaints. Chest pain (sometimes pleuritic), hemoptysis, fever (usually low grade), chills, night sweats, and weight loss may occur. Findings on physical examination may be completely normal, but fever, tachypnea, dullness on percussion of the chest, bronchial or bronchovesicular breath sounds, rales, and rhonchi may be found. Clubbing and cor pulmonale are rare.

DIAGNOSIS

In mild lipoid pneumonia, arterial blood gas values may be normal with the patient at rest but may show hypoxemia after exercise. In more severe disease, resting hypoxemia, hypocapnia, and mild respiratory alkalosis develop. Pulmonary function testing reveals a restrictive ventilatory defect; lung compliance is decreased. The only specific laboratory finding is the presence in sputum of macrophages with clusters of vacuoles that are 5 to 50 µm in diameter and that stain deep orange with Sudan IV; extracellular droplets may stain similarly.

On radiographic examination, the earliest abnormalities are air space infiltrates, most often in the dependent portions of the lung. The infiltrates may be unilateral or bilateral, localized or diffuse. Air bronchograms may be seen. Hilar adenopathy and pleural reaction are rare. As fibrosis develops, volume loss occurs, and linear and nodular infiltrates appear. A solid lesion that closely resembles bronchogenic carcinoma may develop. High-resolution computed tomography usually shows consolidated areas of low attenuation and "crazy paving" (Fig. 94-1).

The differential diagnosis is extensive, particularly in the late phase, when multiple other causes of pulmonary fibrosis must be considered. The key to the correct diagnosis before biopsy is the history of chronic oral or intranasal use of an oil- or a lipid-based product or an occupational exposure to oil mists. The presence of lipid-laden macrophages in sputum or bronchoalveolar lavage fluid also can be used to confirm the diagnosis, particularly in conjunction with typical findings on high-resolution computed tomography.

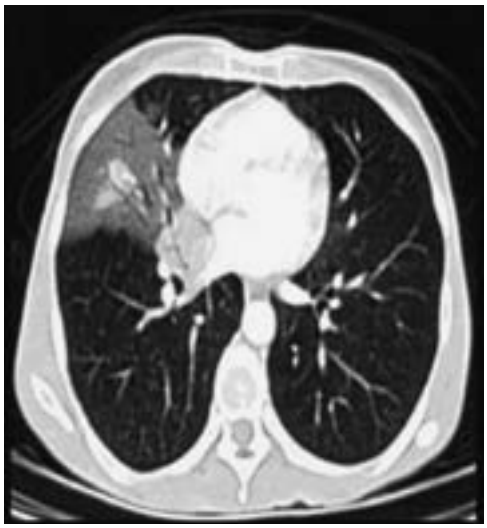

FIGURE 94-1. Lipoid pneumonia on a computed tomographic scan.

TREATMENT AND PREVENTION

Rx

When the diagnosis has been made and the aspiration stopped, the subsequent course is variable. Because the only way the lung can dispose of mineral oil is by expectoration, the patient should be instructed in coughing exercises to be performed many times each day for months. Expectorants have not been shown to help. In some uncontrolled case reports, systemic corticosteroids have been used successfully at varying doses and length, but the literature suggests that they cannot be routinely recommended for treatment.

Transfusion-Related Acute Lung Injury

DEFINITION

The syndrome of transfusion-related acute lung injury (TRALI; Chapter 177) involves the rapid onset of respiratory distress within minutes to hours after the transfusion of blood products (fresh-frozen plasma, platelets, and red blood cells).[9] The initial clinical picture is indistinguishable from acute lung injury or ARDS due to other causes, such as sepsis, multiple trauma, and lung injury. TRALI may similarly be confused with pulmonary edema due to volume overload (Chapter 58).

EPIDEMIOLOGY

The true incidence of TRALI is unknown; incidence rates are underestimated because of the difficulty in distinguishing TRALI from other causes of acute respiratory failure and the labor-intensive and costly diagnostic evaluation required. Reported incidences range from 1 in 1000 to 1 in 100,000 units of blood products transfused. The risk for TRALI varies according to the type of blood product transfused, with pooled products associated with a higher incidence. Only 8 to 21 TRALI-related deaths are reported to the U.S. Food and Drug Administration annually, and even liberal estimates accounting for underreporting suggest the number may be only as high as 300 per year of an estimated 25 million transfusions in the United States.

PATHOBIOLOGY

The physiologic manifestations of TRALI are caused by alveolar filling with fluid and protein. This alveolar process is the result of increased microvascular permeability due to pulmonary endothelial damage mediated by either leukocyte antibodies or the priming and activation of neutrophils in the pulmonary circulation by bioactive substances.

TRALI most commonly occurs when human leukocyte antigen (HLA) type I or II or neutrophil-specific antigen antibodies from the donor attach to the recipient's leukocytes, leading to the release of injurious oxidative and nonoxidative products. Development of HLA antibodies occurs commonly in women during pregnancy, and increasing parity in female blood donors is associated with an increased risk for TRALI.

Episodes of TRALI occurring in patients without HLA or neutrophil-specific antigen antibodies in either the donor or recipient are thought to be caused by a two-hit process of neutrophil priming and activation. Neutrophils are primed and sequestered in the lung by conditions that often occur in patients requiring blood products, such as multiple trauma, surgery, or sepsis. Primed neutrophils are then activated by bioactive lipids and cytokines stored in the blood products, thereby leading to lung injury and alveolar damage. Levels of these bioactive lipids or cytokines may increase after prolonged storage of blood products.

CLINICAL MANIFESTATIONS

Although most cases of TRALI present within 1 to 2 hours after the transfusion of blood products, tachypnea, hypoxemia, cyanosis, dyspnea, and fever can develop during the transfusion or up to 6 hours later. Hypertension or hypotension commonly occurs, depending on the severity of the reaction. Copious amounts of pink, frothy edema fluid may be present. Lung auscultation generally reveals bilateral crackles and decreased breath sounds in dependent lung zones.

DIAGNOSIS

Bilateral patchy infiltrates consistent with alveolar edema are found on plain chest radiographs, typically without effusions. Arterial blood gas analysis demonstrates reduced Po_2, and further laboratory testing may reveal thrombocytopenia or a transient leukopenia. Diagnosis of TRALI requires the presence of the following: acute onset of hypoxemia with Pao_2/Fio_2 of less than 300 mm Hg or room air oxygen saturation of less than 90% during or within 6 hours after a transfusion; bilateral infiltrates on the chest radiograph; no evidence of left atrial hypertension; no preexisting acute lung injury before transfusion; and no temporal relationship to an alternative risk factor for acute lung injury.

The diagnosis of "possible TRALI" is made in patients who have a concurrent diagnosis of another risk factor for acute lung injury, including direct lung injury due to aspiration, pneumonia, toxic inhalation, lung contusion, or nonfatal drowning, and indirect lung injury due to severe sepsis, shock, multiple trauma, burn injury, acute pancreatitis, cardiopulmonary bypass, or drug overdose. Absolute confirmation of the diagnosis requires testing for HLA and neutrophil-specific antigen antibodies, usually performed first in female donors, then in male donors, and finally in the recipient.

TREATMENT

Rx

Most cases are self-limited and resolve within hours to days with supplemental oxygen and supportive care. Volume resuscitation, with or without vasopressors, is required for hypotension. Mechanical ventilation should be managed as for any other case of acute lung injury, with the implementation of a low tidal volume ventilation strategy to prevent further ventilator-induced lung injury. Diuresis should be attempted cautiously and may even be detrimental because intravascular filling pressures are often low.

PROGNOSIS

The mortality rate of TRALI is approximately 5%. If an implicated donor can be identified, the recipient should not receive any further transfusions from that donor, but patients are not at increased risk for further episodes of TRALI from nonimplicated donor transfusions.

Grade A References

A1. Low EV, Avery AJ, Gupta V, et al. Identifying the lowest effective dose of acetazolamide for the prophylaxis of acute mountain sickness: systematic review and meta-analysis. *BMJ*. 2012;345:e6779.

A2. Lipman GS, Kanaan NC, Holck PS, et al. Ibuprofen prevents altitude illness: a randomized controlled trial for prevention of altitude illness with nonsteroidal anti-inflammatories. *Ann Emerg Med*. 2012;59:484-490.

A3. Jafarian S, Gorouhi F, Salimi S, et al. Sumatriptan for prevention of acute mountain sickness: randomized clinical trial. *Ann Neurol*. 2007;62:273-277.

A4. Tang E, Chen Y, Luo Y. Dexamethasone for the prevention of acute mountain sickness: systematic review and meta-analysis. *Int J Cardiol*. 2014;173:133-138.

A5. Xu Y, Liu Y, Liu J, et al. Meta-analysis of clinical efficacy of sildenafil, a phosphodiesterase type-5 inhibitor on high altitude hypoxia and its complications. *High Alt Med Biol*. 2014;15:46-51.

A6. Richalet JP, Rivera-Ch M, Maignan M, et al. Acetazolamide for Monge's disease: efficiency and tolerance of 6-month treatment. *Am J Respir Crit Care Med*. 2008;177:1370-1376.

A7. Bennett MH, Lehm JP, Mitchell SJ, et al. Recompression and adjunctive therapy for decompression illness. *Cochrane Database Syst Rev*. 2012;5:CD005277.

A8. Buckley NA, Juurlink DN, Isbister G, et al. Hyperbaric oxygen for carbon monoxide poisoning. *Cochrane Database Syst Rev*. 2011;4:CD002041.

A9. Bourhis J, Blanchard P, Maillard E, et al. Effect of amifostine on survival among patients treated with radiotherapy: a meta-analysis of individual patient data. *J Clin Oncol*. 2011;29:2590-2597.

A10. Xia DH, Xi L, Xv C, et al. The protective effects of ambroxol on radiation lung injury and influence on production of transforming growth factor beta1 and tumor necrosis factor alpha. *Med Oncol*. 2010;27:697-701.

A11. Gomes CA Jr, Lustosa SA, Matos D, et al. Percutaneous endoscopic gastrostomy versus nasogastric tube feeding for adults with swallowing disturbances. *Cochrane Database Syst Rev*. 2012;3:CD008096.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Szpilman D, Bierens JJ, Handley AJ, et al. Drowning. *N Engl J Med*. 2012;366:2102-2110.
2. Bartsch P, Swenson ER. Clinical practice: acute high-altitude illnesses. *N Engl J Med*. 2013;368:2294-2302.
3. Vann RD, Butler FK, Mitchell SJ, et al. Decompression illness. *Lancet*. 2011;377:153-164.
4. Divers Alert Network. From: http://www.diversalertnetwork.org/; Accessed January 2, 2015.
5. Miller AC, Elamin EM, Suffredini AF. Inhaled anticoagulation regimens for the treatment of smoke inhalation-associated acute lung injury: a systematic review. *Crit Care Med*. 2014;42:413-419.
6. Hampson NB, Piantadosi CA, Thom SR, et al. Practice recommendations in the diagnosis, management, and prevention of carbon monoxide poisoning. *Am J Respir Crit Care Med*. 2012;186:1095-1101.
7. Kallet RH, Matthay MA. Hyperoxic acute lung injury. *Respir Care*. 2013;58:123-141.
8. Benveniste MF, Welsh J, Godoy MC, et al. New era of radiotherapy: an update in radiation-induced lung disease. *Clin Radiol*. 2013;68:e275-e290.
9. Vlaar AP, Juffermans NP. Transfusion-related acute lung injury: a clinical review. *Lancet*. 2013;382:984-994.

REVIEW QUESTIONS

1. Which one of the following is *not* part of the pathophysiologic response to drowning?

A. Reduced surfactant levels and function
B. Laryngospasm
C. Increased PaO_2
D. Apnea
E. Metabolic acidosis

Answer: C Aspiration, apnea, and laryngospasm all lead to hypoxemia. See Drowning: Pathobiology.

2. Which one of the following is *not* true regarding acute mountain sickness?

A. The best treatment for life-threatening symptoms is immediate descent, if possible, combined with supplemental oxygen therapy.
B. Acetazolamide 125 mg orally twice daily, has been shown to be effective in preventing acute mountain sickness.
C. If descent is not possible, high-altitude pulmonary edema and high-altitude cerebral edema can be fatal.
D. Individuals who have previously suffered from acute mountain sickness or other diseases of altitude are protected from further episodes during their next ascent.
E. The pathophysiology centers on the decreased partial pressure of oxygen at altitudes above 7000 feet, leading to decreased oxygen delivery to tissues.

Answer: D People who have developed acute mountain sickness, high-altitude pulmonary edema, or high-altitude cerebral edema are at significantly increased risk for subsequent episodes, especially during rapid ascents. See Diseases of High Altitude.

3. Which one of the following statements is *not* true about decompression sickness?

A. The symptoms of decompression sickness are related to exposure to increasing ambient pressures.
B. Breath holding while ascending from a dive leads to pulmonary barotrauma and potentially arterial gas embolism.
C. Risk factors for decompression sickness include long duration of dives, repetitive dives, heavy exertion at depth, cold water, and rapid ascent.
D. Bullous or cystic lung disease is a contraindication to diving.
E. Recompression therapy with 100% oxygen in a hyperbaric chamber is the standard of care for severe or persistent decompression illness associated with diving.

Answer: A Decompression sickness is caused by decreasing ambient pressures, as the partial pressure of nitrogen decreases and tissues that have been supersaturated with nitrogen during descent begin releasing nitrogen into the surrounding tissues and bloodstream in the form of gas bubbles. These gas bubbles then cause damage either by mechanical compression of tissues or embolization through blood vessels to end organs. See Decompression Illness: Decompression Sickness, Barotrauma, and Arterial Gas Embolism.

4. Which on of the following is *not* an appropriate treatment for cyanide poisoning?

A. Sodium nitrate, 300 mg IV
B. At least one treatment with hyperbaric oxygen lasting approximately 2 hours; can be repeated a second time if symptoms persist
C. Hydroxocobalamin, 5 mg IV
D. Amyl nitrate gas ampules, one 0.3-mL gas ampule each minute
E. Sodium thiosulfate, 12.5g IV

Answer: B Hyperbaric oxygen therapy, which is central to treatment of decompression sickness and carbon monoxide poisoning, plays no role in cyanide poisoning. See Smoke Inhalation: Cyanide and Other Gases.

5. Which one of the following is *not* a correct statement regarding transfusion-related acute lung injury (TRALI)?

A. TRALI is a clinical syndrome that can be indistinguishable from acute lung injury or acute respiratory distress syndrome (ARDS).
B. All blood products are associated with a similar incidence of TRALI.
C. The causative antibodies are HLA type I or type II or neutrophil-specific antigen antibodies.
D. Transfusions from women of increasing parity are associated with an increased risk for developing TRALI.
E. Symptoms of TRALI (tachypnea, hypoxemia, cyanosis, dyspnea and fever) can develop up to 6 hours after the transfusion of blood products.

Answer: B Different blood products carry different risks for causing TRALI, with pooled products carrying the highest risk. See Transfusion-Related Acute Lung Injury.

95

SARCOIDOSIS

MICHAEL C. IANNUZZI

DEFINITION

Sarcoidosis, a systemic granulomatous disease of unknown cause, is characterized by a variable clinical presentation and course. More than 90% of patients exhibit thoracic involvement with mediastinal and hilar lymph node enlargement or parenchymal lung disease, but any organ may be involved. The presentation and course vary from asymptomatic disease with spontaneous resolution to organ system failure and even death.

EPIDEMIOLOGY

Sarcoidosis occurs worldwide, affects people of all racial and ethnic groups, and may present at any age. Sarcoidosis usually develops before the age of 50 years, with the incidence peaking at 20 to 39 years. The incidence of sarcoidosis fluctuates throughout the world, most likely because of differences in the presentation of the disease and the surveillance methods used. The annual incidence of sarcoidosis is highest in northern European countries, where it is 5 to 40 cases per 100,000 people. The incidence in Japan is about 1 to 2 cases per 100,000 people. In the United States, the adjusted annual incidence among black Americans is about 3.5 times higher than among white Americans (35.5 cases per 100,000 compared with 10.9 per 100,000). Regardless of ethnic or racial group, sarcoidosis affects women more often. The reportedly low incidence in certain regions such as Africa, China, India, and Russia may be due to decreased access to health care, minimal surveillance, and misdiagnosis of sarcoidosis as tuberculosis or leprosy.

Black women in the United States have the highest lifetime risk for developing sarcoidosis (2.7%). In both black men and women, sarcoidosis occurs later in life, peaks in the fourth decade, and is more likely to be chronic and fatal. Socioeconomic status does not affect the incidence of sarcoidosis, but low income and other financial barriers to care are associated with more severe sarcoidosis, even after adjustment for age, sex, and race or ethnic group.

Sarcoidosis clusters in families, and monozygotic twins are more often concordant for disease than are dizygotic twins. Familial sarcoidosis occurs in 10% of cases from the Netherlands, 7.5% from Germany, 6% from the United Kingdom, 4.7% from Finland, and 0.8% from Spain. In the United States, sarcoidosis patients are five times as likely to have siblings or parents with sarcoidosis as control subjects. However, less than 1% of the first-degree relatives of patients with sarcoidosis are affected, so screening for disease in asymptomatic relatives is ineffective.

PATHOBIOLOGY

The development and accumulation of granulomas represent the basic pathologic abnormality in sarcoidosis (E-Fig. 95-1). Sarcoidal granulomas are tightly organized collections of macrophages and macrophage-derived epithelioid cells encircled by lymphocytes. Fused epithelioid cells, which become multinucleated giant cells, are often found scattered throughout the granuloma. This appearance suggests that the granuloma is meant to contain an inciting agent, but no such agent has been identified.

Granuloma formation begins with HLA-mediated processing of antigens by macrophages. T-cell activation follows, with oligoclonal expansion of $CD4^+$ (helper-inducer) lymphocytes. These $CD4^+$ cells, primarily of the T_H1 phenotype, produce interleukin-2 (IL-2) and interferon-γ (IFN-γ). A complex interplay of cytokines, along with macrophage-derived tumor necrosis factor-α, organizes the inflammatory cells into granulomas. In some cases, fibroblasts and collagen encase the granulomas with subsequent fibrosis. Fibrosis has been associated with a shift in the involved lymphocytes from the T_H1 (IL-2 and IFN-γ) to the T_H2 phenotype (IL-4, -10, and -13). This fibrotic process irreversibly alters organ architecture and function. Increased 1-α hydroxylase activity in macrophages within granulomas and the alveoli converts 25-hydroxyvitamin D to the biologically active form 1,25-dihydroxyvitamin D (calcitriol), thereby resulting in increased intestinal absorption of calcium.

In sarcoidosis, an immune paradox exists: patients often develop anergy as indicated by a suppressed response to tuberculin, despite active

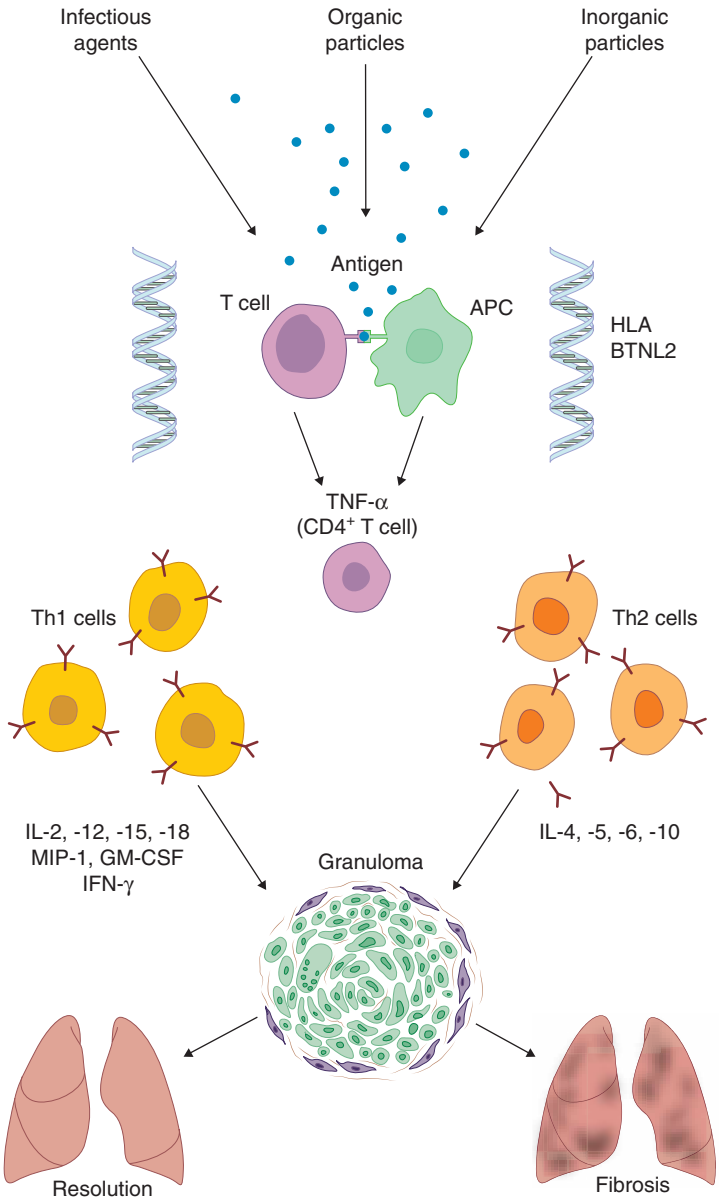

E-FIGURE 95-1. Immunopathogenesis of sarcoidosis. APC = antigen-presenting cell; GM-CSF = granulocyte-macrophage colony-stimulating factor; IFN-γ = interferon-γ; IL = interleukin; MIP-1 = macrophage inflammatory protein-1; TNF-α = tumor necrosis factor-α.

granulomatous inflammation. This anergy results in part from an expansion of $CD25^{bright}$ regulatory T cells, which are a subgroup of $CD4^+$ T lymphocytes and which can abolish production of IL-2 and inhibit T-cell proliferation. Although sarcoidosis is predominantly a T-cell-driven disease, the presence of polyclonal hyperglobulinemia indicates that B lymphocytes may also play a role.

Multiple inciting agents rather than a single pathogen are likely to cause sarcoidosis. Airborne antigens are suspected because the lungs, followed by the eyes and skin, are the most commonly involved organs. A microbial origin of the antigens is favored because sarcoidosis occurs sporadically in clusters and is worldwide. Furthermore, sarcoidosis may recur in transplanted organs and can develop in recipients of tissues from donors with sarcoidosis.

Several environmental exposures are modestly associated with the risk for sarcoidosis, each with odds ratios of about 1.5: mold or mildew, musty odors at work, agricultural employment, and pesticide-using industries. Highly sensitive methods for detecting microbial DNA and proteins, such as polymerase chain reaction and mass spectrometry, have yielded potential etiologic agents. One promising candidate of microbial origin is the mycobacterial KatG protein.

Major histocompatibility complex (MHC) genes and non-MHC genes located on the short arm of chromosome 6p have been implicated as genetic risk factors for sarcoidosis.[1] HLA-DQB1*0201 and HLA-DRB1*0301 alleles are strongly associated with acute disease and good prognosis. The HLA-DRB1*1501/DQB1*0602 haplotype predicts a chronic course and severe pulmonary sarcoidosis. Genome-wide scans have identified the gene candidates *BTNL2* (butyrophilin-like 2 gene) and *ANXA11* (annexin A11).[2]

CLINICAL MANIFESTATIONS

Patients may present with any organ involved, with or without concomitant intrathoracic involvement.[3,4] The severity of symptoms and organ dysfunction also varies, and as many as 30 to 50% of patients have no symptoms at the time of diagnosis. Clinical manifestations also vary by ethnicity, race, and sex as well as by the particular organ system predominantly affected.[5]

Sarcoidosis often first comes to attention during routine screening when enlarged mediastinal and hilar lymph nodes are detected on a chest radiograph (Figs. 95-1 and 95-2). Symptomatic individuals commonly experience fatigue, night sweats, and weight loss. Fatigue often remains a prominent problem and can lead to impaired quality of life.

Symptomatic sarcoidosis may present insidiously or as an acute illness. Löfgren syndrome, an acute form of the disease, consists of erythema nodosum (see Fig. 440-24), arthritis, and bilateral hilar adenopathy. Fever and uveitis may also accompany Löfgren syndrome. Erythema nodosum occurs predominantly in women, whereas arthritis predominates in men.

Respiratory System Disease

More than 90% of patients have respiratory tract involvement, sometimes with symptoms and sometimes with asymptomatic radiographic abnormalities. The most common respiratory symptoms are dry cough and dyspnea. Chest pain, when present, is vague and nonspecific.

Upper respiratory tract involvement occurs in 2 to 6% of patients, with most having nasal mucosal involvement. Severe upper respiratory tract disease may lead to anosmia, erosion of septal cartilage, and nasal deformity. Supraglottic and glottic involvement can lead to stridor, dysphonia, cough, and dysphagia.

Examination of the chest often reveals few or no findings. Localized wheezing suggests endobronchial granulomatous disease, whereas diffuse wheezing suggests hyperreactive airways disease. Crackles may be heard when bronchiectasis and fibrosis are present. Hemoptysis may occur with bronchiectasis or cavitary disease.

The most common radiographic finding is intrathoracic lymph node enlargement with or without parenchymal lung involvement. Mediastinal lymphadenopathy without hilar adenopathy is extremely rare. Hilar lymph node enlargement is usually symmetrical, and less than 3% of patients have unilateral enlargement. Concurrent enlargement of right paratracheal and aortic-pulmonary window lymph nodes is common.

The chest radiograph is traditionally categorized into five stages (Table 95-1), but these stages do not necessarily denote the severity or chronologic progression of disease, particularly when extrathoracic involvement is present. Furthermore, 50 to 94% of patients will have hilar or mediastinal lymphadenopathy on computed tomography (CT) scans irrespective of their stage on a plain chest radiograph.

Chest CT may demonstrate nodules, ground-glass opacities, bronchiectasis, cysts, and thickening of the pleural surface (Fig. 95-3). In more advanced cases, patients develop fibrotic changes (Fig. 95-4) and evidence of pulmonary hypertension (Fig. 95-5). CT, however, generally adds little to diagnosis and management, particularly for those with a stage I chest radiograph. CT can be justified when atypical clinical and chest radiographic findings are present, a normal chest radiograph is found during the evaluation of suspected extrathoracic disease, or complications are suspected.

TABLE 95-1 CHEST RADIOGRAPHIC STAGING

Stage 0: Normal
Stage 1: Bilateral hilar adenopathy, often with right paratracheal adenopathy
Stage 2: Bilateral hilar adenopathy and parenchymal infiltration
Stage 3: Parenchymal infiltration without lymphadenopathy
Stage 4: Advanced parenchymal disease demonstrating fibrosis and possibly including honeycombing, cysts, bullae, and traction bronchiectasis

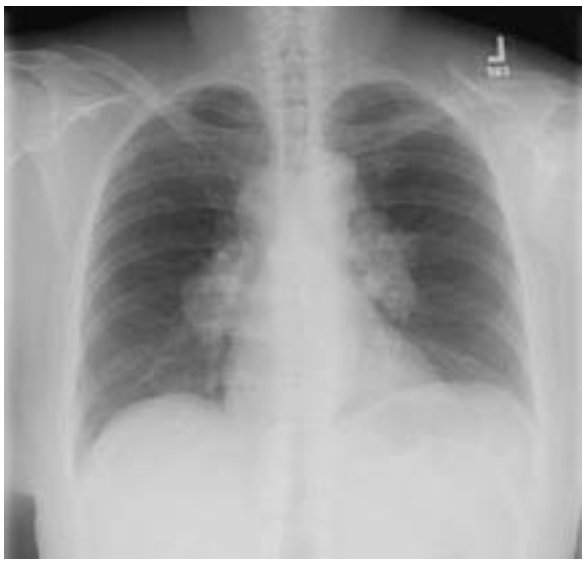

FIGURE 95-1. Chest radiograph demonstrating a stage I disease with enlarged mediastinal and hilar lymph nodes.

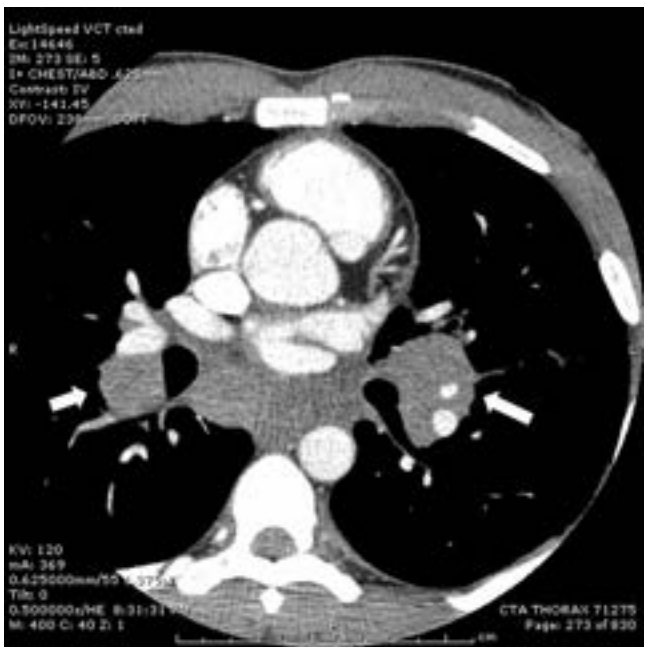

FIGURE 95-2. Chest computed tomographic scan showing typical hilar adenopathy, which correlates with stage I.

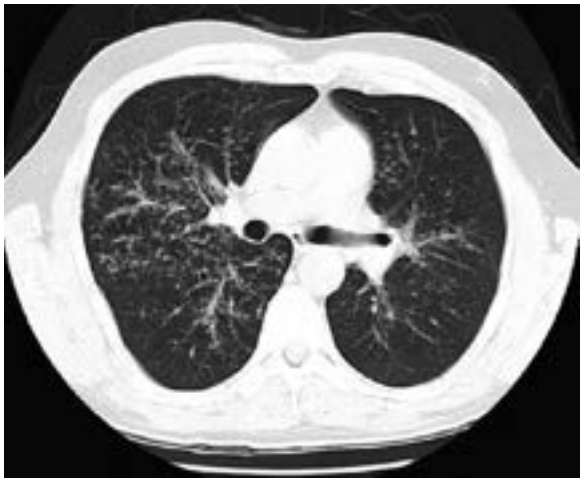

FIGURE 95-3. High-resolution chest computed tomographic scan demonstrating numerous small nodules in a predominantly bronchovascular distribution.

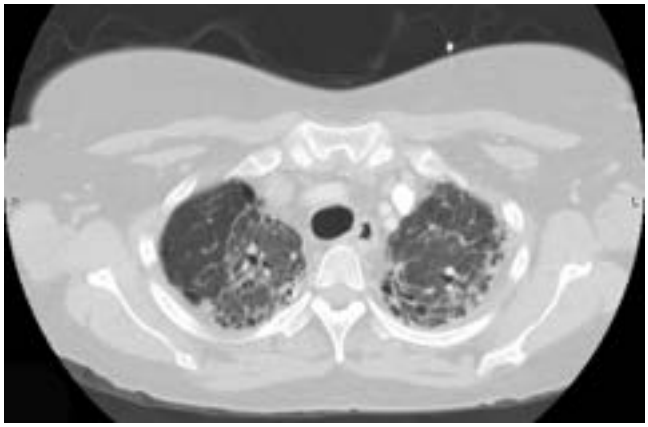

FIGURE 95-4. Chest computed tomographic scan showing fibrotic changes in the upper lobes with bronchiectasis.

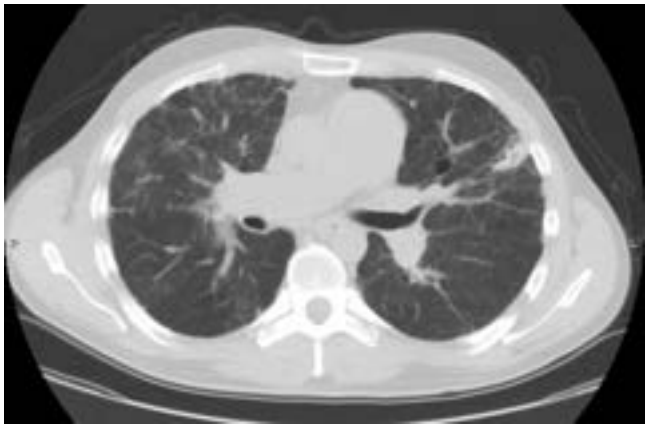

FIGURE 95-5. Chest computed tomographic scan showing diffuse bilateral fibrosis along with severe pulmonary artery enlargement consistent with pulmonary hypertension.

Endobronchial sarcoidosis may lead to bronchial stenosis and recurrent obstructive pneumonias. Pleural effusions on plain radiography are uncommon (1 to 3% of patients). Pulmonary hypertension (Chapter 68) may complicate sarcoidosis, particularly when pulmonary fibrosis is present.

Skin Disease

Skin involvement occurs in 25 to 35% of patients, but lesions are commonly misdiagnosed because they can present as macules, papules, plaques, subcutaneous lesions, areas of increased or decreased pigmentation, and ulcerations. Lesions are commonly found on the upper back, the nape of the neck, and extremities (Fig. 95-6) and have a predilection for scars and tattoos. The most common skin manifestation is a maculopapular eruption, consisting of firm, flesh-colored to violaceous lesions that have a predilection for the eyelids and perioral area. Scalp involvement with alopecia may occur. Skin lesions in black American patients frequently leave hyperpigmented scars and pale or depigmented areas.

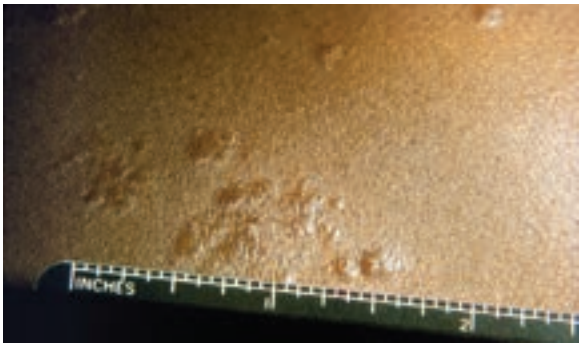

FIGURE 95-6. **Skin lesions of sarcoidosis.** Sarcoidal lesions may occur at any site. This patient's papules have a waxy appearance and are located on the upper part of the back.

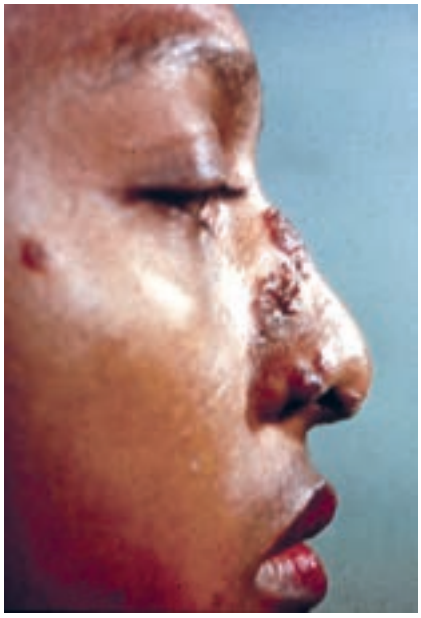

FIGURE 95-7. *Lupus pernio* is the term used to describe infiltrative skin lesions affecting the nose, cheeks, and ears in chronic sarcoidosis.

Erythema nodosum occurs in about 10% of patients and commonly occurs as part of Löfgren syndrome. It presents as raised, red, hot, tender subcutaneous nodular lesions, most commonly on the shins, but sometimes also on the arms and buttocks (see Fig. 440-24). Lesions persist for 1 to 3 weeks and may recur. The lesions of erythema nodosum typically show nonspecific septal panniculitis without sarcoidal granulomas, so biopsy is not usually helpful. Erythema nodosum portends a good prognosis, with up to 85% of patients resolving their sarcoidosis within 2 years.

Lupus pernio (Fig. 95-7) consists of chronic, lumpy, violaceous, indurated plaques and nodules distributed around the nose, cheeks, lips, and ears; it is specific to sarcoidosis. Patients with lupus pernio more commonly have a chronic course, fibrotic lung disease, and upper respiratory tract involvement.

Eye Disease

More than 25% of patients have ocular involvement (Chapter 423), and any part of the eye and adnexa may be involved. Lacrimal gland enlargement and conjunctival involvement are common. Conjunctival involvement consists of pale yellow nodules, which demonstrate granulomas on biopsy. Corneal involvement is rare. Acute anterior uveitis presents with pain, photophobia,

TABLE 95-2 NEUROSARCOIDOSIS

Cranial neuropathies
VII—Facial nerve palsy (unilateral, bilateral)
II—Optic nerve (unilateral and bilateral)
VIII—Hearing loss
Other cranial neuropathies
Myelopathy (sensory and motor)
Seizure
Basilar meningitis
Central diabetes insipidus
Panhypopituitarism
Intraparenchymal mass
Hydrocephalus
Peripheral neuropathy

lacrimation, and redness. Chronic anterior uveitis, which is more common than acute uveitis, may have minimal symptoms. Posterior uveitis, which occurs in about 30% of patients with ocular sarcoidosis, is frequently accompanied by central nervous system involvement. Choroidal lesions may occur anywhere in the fundus and may be multifocal. In 10 to 15% of patients with uveitis, both the anterior and posterior segments are affected.

Uveitis can herald the nonocular signs of sarcoidosis and may precede the diagnosis of sarcoidosis by decades. Chronic anterior uveitis may lead to cataracts and glaucoma. Heerfordt syndrome consists of anterior uveitis accompanied by parotid gland enlargement and fever.

Cardiac Disease

Cardiac granulomas are found in about 25% of sarcoidosis patients who are examined at autopsy, but cardiac sarcoidosis is suspected in only about 5% of patients. In cardiac sarcoidosis (Chapter 60), granulomas most often infiltrate the left ventricular free wall. Next most frequently, sarcoidal granulomas infiltrate the intraventricular septum, where they often involve the conduction system. Cardiac manifestations include atrioventricular block, ventricular arrhythmias, left ventricular dysfunction, and sudden death.[6]

Neurologic Disease

Nervous system granulomas are found in up to 25% of sarcoidosis patients who undergo autopsy, but only 10% of patients present with neurologic symptoms (Table 95-2). In patients with neurologic involvement, the neurologic signs or symptoms precede the diagnosis of sarcoidosis in up to 75% of patients and may be the only manifestation of sarcoidosis.

Neurosarcoidosis has a predilection for the base of the brain, hypothalamus, and pituitary gland.[7] Myelopathy may occur anywhere in the spinal cord and carries a poor prognosis. Peripheral neuropathy (Chapter 420) may manifest as mononeuropathy or polyneuropathy.

Liver and Spleen

Liver involvement is twice as common in black Americans as in white Americans. Symptoms due to liver disease are infrequent, but abdominal pain and pruritus are the most common symptoms. Fever, weight loss, and jaundice are present in less than 5% of those with liver involvement. About 20% of patients will have hepatomegaly on physical examination, and 35% will have an elevated serum aminotransferase or alkaline phosphatase level. Sarcoidosis can rarely (<1%) cause progressive liver disease that leads to portal hypertension with variceal bleeding; the hepatopulmonary syndrome, with refractory hypoxemia, cirrhosis, and liver failure, may also occur.

Splenomegaly is found on physical examination in 5 to 15% of patients with sarcoidosis. Massive splenomegaly is rare.

Bone and Joint

Most patients complain of arthralgias, but only about 35% develop arthritis. Acute sarcoid arthritis (Chapter 275) consists of large joint periarthritis, particularly involving the ankles and knees, and commonly occurs with Löfgren syndrome. These patients often report difficulty walking related to joint pain. Acute sarcoid arthritis usually persists for up to 3 months but is self-limited. Chronic sarcoid arthritis with direct granulomatous synovial infiltration is rare. Osseous sarcoidosis usually does not produce symptoms.

Calcium Metabolism

Aberrant calcium and vitamin D metabolism occurs in up to 50% of patients and may result in renal stones (Chapter 126), nephrocalcinosis with renal insufficiency, and hypercalciuria (urine calcium > 300 mg/24 hours) with or without hypercalcemia (Chapter 245). In chronic sarcoidosis, 10 to 14% of patients have at least one symptomatic renal stone.

Renal Involvement

Renal involvement, other than as a result of dysregulated calcium metabolism, occurs in less than 1% of patients. Renal disease may include granulomatous interstitial nephritis, glomerular disease, renal tubular dysfunction, renal vascular disease, and obstructive uropathy. Granulomatous interstitial nephritis is more common in white men.

DIAGNOSIS

The diagnosis of sarcoidosis should be based on compatible clinical and radiographic findings supported by histologic evidence of noncaseating granulomas in one or more organs in the absence of any foreign particles or organisms.[8] An occupational history consistent with beryllium exposure (Chapter 93), such as employment in the aerospace, automotive, ceramic, or computer industries, requires further investigation because chronic beryllium disease cannot be clinically or histologically distinguished from sarcoidosis.

A diagnosis of sarcoidosis is reasonably certain even without histologic confirmation in patients who present with Löfgren syndrome. In all other cases, a biopsy specimen from an involved organ should be obtained. The organ for which biopsy is safest, such as the skin, peripheral lymph nodes, lacrimal glands, or conjunctiva, should be sampled.

If diagnosis requires intrathoracic sampling, fine-needle aspiration of enlarged intrathoracic lymph nodes guided by endobronchial ultrasound has become the preferred procedure,■ with a diagnostic yield of 80% or higher.[9] Endobronchial ultrasound also visualizes mediastinal structures, including the subcarinal region, the paraesophageal space, and the aortopulmonary window. Bronchoscopy and endobronchial ultrasound-guided fine-needle aspiration have generally replaced more invasive approaches, such as mediastinoscopy or video-assisted thoracoscopic surgical biopsies.

Bronchoalveolar lavage sampling demonstrates lymphocytosis with normal or low granulocyte counts in more than 85% of patients (Chapter 85). The CD4/CD8 ratio is increased in bronchoalveolar lavage in about 50% of patients with sarcoidosis, but bronchoalveolar lavage findings are nonspecific and should not be used to diagnose sarcoidosis. Bronchoalveolar lavage findings also cannot predict prognosis or responsiveness to corticosteroid therapy.

Sarcoidal granulomas produce angiotensin-converting enzyme (ACE). Although serum ACE levels are elevated in 60% of patients with sarcoidosis, the value of a serum ACE level in diagnosing sarcoidosis remains limited because positive and negative predictive values are only about 84% and 74%, respectively.

Patients should routinely undergo pulmonary function testing, which often does not correlate with the chest radiographic stage. For example, pulmonary function tests may be abnormal even in patients with a normal chest radiograph. Diffusion capacity for carbon monoxide (DLCO) is usually the first abnormality detected and the last to normalize on remission. A DLCO of less than 50% of predicted is associated with exercise-induced oxygen desaturation and should prompt formal oxygen saturation testing with exercise.

About two thirds of patients have airflow limitation at presentation. Spirometry usually indicates restrictive ventilatory dysfunction with a reduced forced vital capacity (FVC) and reduced forced expiratory volume in 1 second (FEV_1). At least 50% of patients have concurrent obstructive airways disease. Airway hyperreactivity, as measured by increased responsiveness to methacholine, occurs in up to 83% of patients. In patients with abnormal spirometric findings at diagnosis, spirometry returns to normal in 80%. Alterations in cardiopulmonary exercise testing have been reported in nearly 50% of patients with sarcoidosis. Abnormalities, including a ventilatory limitation with exercise and a widened alveolar-arterial O_2 gradient, may be present. FVC is the single best test for following respiratory involvement and correlates well with FEV_1, total lung capacity, and DLCO.

Because ocular involvement is common and vision loss may occur, complete ocular evaluation with slit lamp and funduscopic examinations should be routinely performed during the initial evaluation and then annually in patients with active systemic disease.

An electrocardiogram (ECG) should be performed at the initial encounter. Screening for cardiac symptoms (palpitations, dizziness, and syncope) should be performed during the initial evaluation and routinely during follow-up visits if the disease is active. Any cardiac symptoms or abnormalities on the ECG should prompt further cardiac event monitoring (Chapter 62) and testing by echocardiography (Chapter 55). Endomyocardial biopsy

has less than 20% diagnostic yield because cardiac involvement is patchy and is usually most dense in the left ventricle and basal ventricular septum where endomyocardial biopsies are avoided. If cardiac involvement is suspected, positron emission tomography (PET) or cardiac magnetic resonance imaging (MRI) with contrast should be performed.[10] In patients who have arrhythmias, decreased ventricular function, or septal involvement detected on MRI or PET, electrophysiologic studies should be considered, and the need for an automatic implantable cardioverter-defibrillator (AICD) should be evaluated (Chapters 65 and 66).

The criteria for the diagnosis of neurosarcoidosis in the absence of histologic confirmation are not established. Compatible MRI findings, a characteristic presentation, and exclusion of other neurologic diseases are required. Histologic confirmation of disease elsewhere supports the diagnosis of neurosarcoidosis. Cerebrospinal fluid (CSF) analysis typically demonstrates nonspecific lymphocytic inflammation (Chapter 412). The CSF glucose can be as low as 14 mg/dL, with a white blood cell count as high as 350 cells/μL and a protein concentration as high as 670 mg/dL. CSF ACE levels are neither sensitive nor specific. CSF oligoclonal immunoglobulin bands are elevated in one third of patients, thereby making it difficult to differentiate sarcoidosis from multiple sclerosis (Chapter 411).

Up to 65% of patients will have granulomas on liver biopsy, but hepatic granulomas (Chapter 151) occur commonly with other disorders such as infection and drug-induced hepatitis. As a result, the diagnosis of sarcoidosis should not be based solely on detection of hepatic granulomas.

On plain radiographic images, most skeletal lesions of sarcoidosis are seen in the small bones of the hands and feet. MRI and PET scanning show much greater involvement of the axial skeleton and long bones, and osseous sarcoidosis may appear similar to metastatic disease on a PET scan.

Sarcoidosis Complicating Type 1 Interferon Therapy

Type 1 interferons, IFN-α and IFN-β, used to treat viral hepatitis (Chapter 149), multiple sclerosis (Chapter 411), and autoimmune and malignant disease may increase the T_H1 cytokines, IFN-γ, and IL-2 and rarely (<1 to 5%) result in sarcoidosis. Most reported cases of interferon-induced sarcoidosis occur within 6 months of therapy and manifest primarily with lung and skin involvement.

TREATMENT

Rx

Because the inciting agent has not yet been identified nor genetic risk factors firmly established, no means exist to prevent sarcoidosis. Most patients with sarcoidosis are not disabled by their illness, so the decision to recommend treatment should weigh the risks of using corticosteroids (Chapter 35), which are the most common treatment, against potential benefits (Table 95-3). Hypercalcemia, cardiac disease, and neurologic disease are indications for treatment, and immediate treatment is appropriate whenever organ function is threatened or when symptoms are severe. Detection of granulomatous inflammation on physical examination, biopsy, or imaging or the presence of an elevated serum ACE level is not a mandate to provide treatment.

Treatment is, however, recommended for patients who have a progressive decline in their pulmonary function, which sometimes is accompanied by progressive changes on chest imaging.[11] Oral prednisone at a dose of 20 to 40 mg per day for 3 months is usually the initial recommended therapy.[3] Response to treatment is indicated by improvement in symptoms or objective measures, such as decreased size and number of skin lesions, increased FVC, or a reduction of detectable cardiac or brain lesions. For pulmonary involvement, steroids improve symptoms, pulmonary function test results, and chest radiographic findings. If a response is noted at 3 months, the prednisone dose should be tapered to 10 to 15 mg per day for an additional 6 to 9 months and then tapered off. Because recurrence is possible, patients should be followed closely for 1 to 2 years after discontinuing treatment. The indications for restarting treatment are the same as those used to begin treatment. Lack of response after a 3-month trial suggests nonadherence to therapy, an inadequate dose of prednisone, or irreversible fibrotic disease. Inhaled corticosteroids should be used only in patients with bronchial hyperreactivity or persistent cough.

Despite the absence of definitive randomized trials, cytotoxic and immunosuppressive drugs have been used to treat patients who do not respond to corticosteroid therapy or who cannot tolerate it (see Table 95-3), and both methotrexate and azathioprine appear to have significant steroid-sparing effects.[12,13] Generally, 3 to 6 months of cytotoxic and immunosuppressive treatment is required to determine whether a response has occurred. For patients who respond, the duration of treatment should be 9 to 12 months. In patients who are steroid dependent but have ongoing extrapulmonary disease, the addition of infliximab, 3 to 5 mg/kg for 24 weeks, can be efficacious.

TABLE 95-3 TREATMENT

ORGAN	CLINICAL FINDINGS	TREATMENT
Respiratory	Dyspnea, persistent cough, FVC < 70%	Prednisone, 20-40 mg/day
	Mild cough, wheezing	Inhaled corticosteroid
Skin	Lupus pernio	Prednisone, 20-40 mg/day Hydroxychloroquine, 400 mg/day Methotrexate, 10-15 mg/wk Infliximab, 3-5 mg/kg every 2-4 wk
	Maculopapular eruptions	Topical corticosteroid Hydroxychloroquine, 400 mg/day Prednisone, 20-40 mg/day
	Erythema nodosum	NSAID*
Eyes	Anterior uveitis Posterior uveitis	Topical corticosteroid Prednisone, 20-40 mg/day
Cardiac	Complete heart block, ventricular arrhythmias	AICD, prednisone, 20-40 mg/day
	Decreased LVEF (<35%)	AICD, prednisone, 20-40 mg/day
Central nervous system	Cranial nerve palsies	Prednisone, 20-40 mg/day
	Myelopathy	Prednisone, 40-60 mg/day, and azathioprine, 150 mg/day (or mycophenolate mofetil, or cyclophosphamide)
	Intracerebral involvement	Prednisone, 40-60 mg/day, and azathioprine, 150 mg/day (or mycophenolate mofetil, 1000-2000 mg/day, or cyclophosphamide, 1-5 mg/kg/day)
Liver	Cholestatic hepatitis	Prednisone, 20-40 mg/day
Bone and joint	Arthralgias	NSAID
	Granulomatous arthritis	Prednisone, 20-40 mg/day Methotrexate, 10-15 mg/wk
	Bone destruction/pain	Prednisone, 20-40 mg/day Methotrexate, 10-15 mg/wk
Hypercalciuria and hypercalcemia	Kidney stones	Prednisone, 20-40 mg/day Hydroxychloroquine, 400 mg/day

*For example, ibuprofen, 200-800 mg three times a day.
AICD = automatic implantable cardioverter-defibrillator; FVC = forced vital capacity; LVEF = left ventricular ejection fraction; NSAID = nonsteroidal anti-inflammatory drug.

Bosentan (up to 125 mg daily) can significantly reduce pulmonary hypertension but has not been shown to improve exercise tolerance in patients with sarcoidosis. Adalimumab may be effective for treating skin lesions.

Moderate doses of corticosteroids (15 to 20 mg prednisone) reduce the elevated serum and urinary calcium within a few days of starting treatment. Failure of serum calcium to normalize within 2 weeks should alert the clinician to an alternative diagnosis such as hyperparathyroidism or malignancy (Chapter 245). With treatment, most patients with granulomatous interstitial nephritis regain renal function but are often left with chronic kidney disease of varying severity.

Less than 1% of heart and liver and about 3% of lung transplantations are performed in patients with sarcoidosis. The 1- and 5-year graft survival rates for lung, liver, and heart transplantations in patients with sarcoidosis compare favorably with the results obtained for patients with other disorders.

PROGNOSIS

Patients with Löfgren syndrome have a good prognosis characterized by spontaneous resolution with few consequences. Overall, spontaneous remission with few or no consequences occurs in more than 50% of patients within 3 years of diagnosis and in two thirds of patients within a decade. After one or more years of remission without treatment, recurrence occurs in fewer than 5% of patients. Up to one third of patients have unrelenting disease that leads to significant organ impairment, but less than 5% of patients die from sarcoidosis. Black Americans tend to have a worse prognosis with more chronic disease and more extrathoracic involvement of the eyes, liver, bone marrow, extrathoracic lymph nodes, and skin. In the absence of extrathoracic disease, patients with stage I radiographs generally have the best prognosis. Pulmonary function testing is more reliable in determining prognosis than is radiograph staging. Patients with neurologic and cardiac involvement have a poorer prognosis.

Grade A References

A1. von Bartheld MB, Dekkers OM, Szlubowski A, et al. Endosonography vs conventional bronchoscopy for the diagnosis of sarcoidosis: the GRANULOMA randomized clinical trial. *JAMA*. 2013;309:2457-2464.

A2. Paramothayan S, Jones PW. Corticosteroid therapy in pulmonary sarcoidosis: a systematic review. *JAMA*. 2002;287:1301-1307.

A3. Judson MA, Baughman RP, Costabel U, et al. Efficacy of infliximab in extrapulmonary sarcoidosis: results from a randomized trial. *Eur Respir J*. 2008;31:1189-1196.

A4. Baughman RP, Culver DA, Cordova FC, et al. Bosentan for sarcoidosis-associated pulmonary hypertension: a double-blind placebo controlled randomized trial. *Chest*. 2014;145:810-817.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Spagnolo P, Grunewald J. Recent advances in the genetics of sarcoidosis. *J Med Genet*. 2013;50:290-297.
2. Levin AM, Iannuzzi MC, Montgomery CG, et al. Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun*. 2013;14:13-18.
3. Valeyre D, Prasse A, Nunes H, et al. Sarcoidosis. *Lancet*. 2014;383:1155-1167.
4. Iannuzzi MC, Fontana JR. Sarcoidosis: clinical presentation, immunopathogenesis, and therapeutics. *JAMA*. 2011;305:391-399.
5. Pereira CA, Dornfeld MC, Baughman R, et al. Clinical phenotypes in sarcoidosis. *Curr Opin Pulm Med*. 2014;20:496-502.
6. Kron J, Ellenbogen KA. Cardiac sarcoidosis: contemporary review. *J Cardiovasc Electrophysiol*. 2015;26:104-109.
7. Segal BM. Neurosarcoidosis: diagnostic approaches and therapeutic strategies. *Curr Opin Neurol*. 2013;26:307-313.
8. Israel-Biet D, Valeyre D. Diagnosis of pulmonary sarcoidosis. *Curr Opin Pulm Med*. 2013;19:510-515.
9. Gindesgaard CB, Schousboe LP, Christensen RK. Endobronchial ultrasound-guided transbronchial needle aspiration in an unselected cohort. *J Bronchology Interv Pulmonol*. 2013;20:140-146.
10. Mc Ardle BA, Leung E, Ohira H, et al. The role of F(18)-fluorodeoxyglucose positron emission tomography in guiding diagnosis and management in patients with known or suspected cardiac sarcoidosis. *J Nucl Cardiol*. 2013;20:297-306.
11. Patterson KC, Strek ME. Pulmonary fibrosis in sarcoidosis: clinical features and outcomes. *Ann Am Thorac Soc*. 2013;10:362-370.
12. Korsten P, Mirsaeidi M, Sweiss NJ. Nonsteroidal therapy of sarcoidosis. *Curr Opin Pulm Med*. 2013;19:516-523.
13. Vorselaars AD, Cremers JP, Grutters JC, et al. Cytotoxic agents in sarcoidosis: which one should we choose? *Curr Opin Pulm Med*. 2014;20:479-487.

REVIEW QUESTIONS

1. Which of the following environmental exposures has been associated with risk for sarcoidosis?

A. Mold or mildew
B. Musty odor at work
C. Agricultural employment
D. Pesticide using industries
E. All of the above

Answer: E Several environmental exposures are modestly associated with the risk for sarcoidosis, each with odds ratios of about 1.5: mold or mildew, musty odor at work, agricultural employment, and pesticide-using industries. Highly sensitive methods for detecting microbial DNA and proteins, such as polymerase chain reaction and mass spectrometry, have yielded potential etiologic agents. One promising candidate of microbial origin is the mycobacterial KatG protein.

2. Which of the following is the most common manifestation of ocular sarcoidosis?

A. Acute anterior uveitis
B. Chronic anterior uveitis
C. Acute posterior uveitis
D. Chronic posterior uveitis
E. Pan uveitis

Answer: B Chronic anterior uveitis, which is more common than acute uveitis, may have minimal symptoms. Uveitis can herald the nonocular signs of sarcoidosis and may precede the diagnosis of sarcoidosis by decades. Chronic anterior uveitis may lead to cataracts and glaucoma. Heerfordt syndrome consists of anterior uveitis accompanied by parotid gland enlargement and fever.

3. Which of the following best describes the role of transplantation in sarcoidosis?

A. Transplantation should not be used because of the high frequency of disease recurrence.
B. One- and 5-year graft survival rates for lung, liver, and heart are much less than for other diseases.
C. One- and 5-year graft survival rates for lung, liver, and heart compare favorably with those of other diseases.
D. Sarcoidosis accounts for more than 10% of all transplantations.

Answer: C Less than 1% of heart and liver and about 3% of lung transplantations are performed in patients with sarcoidosis. The 1- and 5-year graft survival rates for lung, liver, and heart transplantations in patients with sarcoidosis compare favorably with the results obtained for patients with other disorders.

96 ACUTE BRONCHITIS AND TRACHEITIS

RICHARD P. WENZEL

DEFINITION

The term *acute bronchitis and tracheitis* defines a self-limited (1 to 3 weeks) inflammation of the large airways of the lung that extends to the tertiary bronchi (Fig. 96-1).[1] In patients with a primary symptom of cough (Chapter 83), the diagnosis is made if there is no clinical or radiologic evidence of pneumonia. At the bedside, the absence of criteria for systemic inflammatory response syndrome (SIRS) (Chapter 108) suggests bronchitis and tracheitis and makes a diagnosis of pneumonia (Chapter 97) unlikely. The SIRS criteria are met if the patient has more than two of the following: temperature lower than 36° C or higher than 38° C, pulse greater than 90 beats per minute, respiratory rate higher than 20 breaths per minute, or white blood cell count less than 4000 cells/mm^3 or higher than 12,000 cells/mm^3 or with greater than 10% bands.

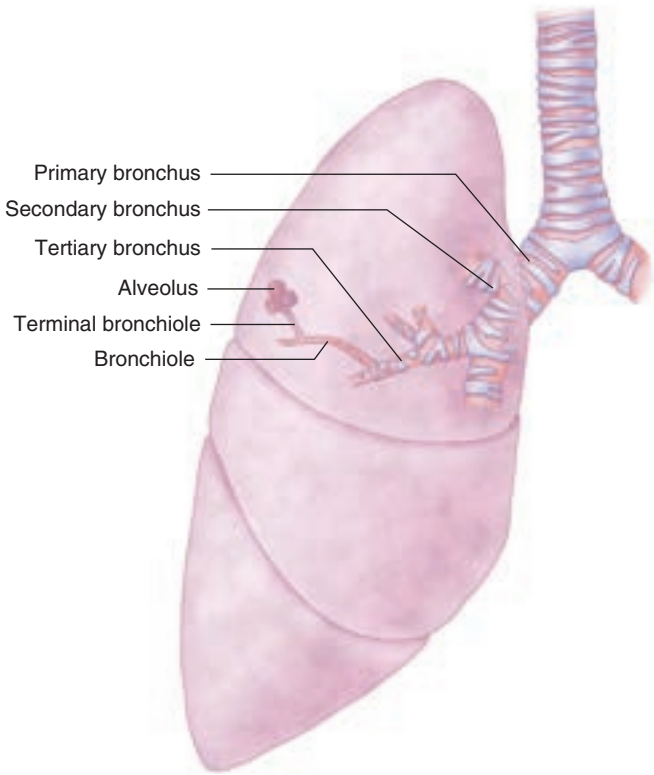

FIGURE 96-1. Many infecting agents that cause bronchitis and tracheitis can infect both large and small airways of the lung and occasionally the alveoli. Not surprisingly, a wide spectrum of signs and symptoms are associated with bronchitis, including cough, wheezing, and shortness of breath.

The definition of acute bronchitis and tracheitis also seeks to differentiate the illness from acute inflammation of the small airways (bronchiolitis), even though the accompanying symptoms with the former may include sputum production, wheezing, and shortness of breath.[1,2] Among patients with primarily small airways disease, some might be expected to have prominently decreased breath sounds in the areas involved. Acute bronchitis and tracheitis is also different from bronchiectasis (Chapter 90), which is associated with permanent dilation of bronchi and a chronic cough. Furthermore, a diagnosis of chronic bronchitis (Chapter 88) is reserved for patients who have prolonged cough and sputum production: at least 3 months of the year for 2 consecutive years.

EPIDEMIOLOGY

Occurring at a rate of 44 per 1000 adults per year, acute bronchitis and tracheitis affects approximately 5% of adults annually. A higher incidence is observed in the winter and fall than in the summer and spring. In the United States, acute bronchitis and tracheitis is the ninth most common illness in outpatients as reported by physicians.

The disorder is thought to be viral in origin almost all the time. However, viruses have been isolated in only 8 to 37% of patients. Thus, the true causes of the illness are unknown in most cases. Nevertheless, at least 70% of patients with acute bronchitis and tracheitis in the United States receive antibacterial antibiotics after visiting a physician. Importantly, although the same bacteria that are commonly implicated in community-acquired pneumonia are also isolated from the sputum in half the patients, their role in the pathobiology of acute bronchitis and tracheitis or its attendant symptoms is unclear, and bronchial biopsies have not shown bacterial invasion.

PATHOBIOLOGY

Infections of the epithelium of the bronchi and trachea are thought to incite an inflammatory response. Pathologically, there is an accompanying microscopic thickening of bronchial and tracheal mucosa corresponding to the inflamed areas. Such pathologic findings are also consistent with the occasional case report of upper airway inflammation confined to the bronchi and trachea detected by ^{18}F-labeled fluorodeoxyglucose positron emission tomography (FDG-PET).

In various studies, specific pathogens have been found in more than half of patients with acute bronchitis.[1] These pathogens include common respiratory viruses, typical respiratory bacteria, and atypical bacteria. At least three variables can influence the yield of specific pathogens: the presence of epidemics, the season of the year, and the population's influenza vaccination status. Furthermore, there are probably wide variations in the anatomic distribution of all pathogens causing acute bronchitis and tracheitis, extending from the nasal mucosa to the bronchiolar epithelium.

The viruses implicated in acute bronchitis and tracheitis include influenza A and B (Chapter 364), parainfluenza (Chapter 363), respiratory syncytial virus (Chapter 362), coronavirus (Chapter 366), adenovirus (Chapter 365), and rhinoviruses (Chapter 361), usually in this order from the most to the least common. Human metapneumovirus (Chapter 361) also has been identified as an etiologic agent. In children, bronchiolitis has been associated with respiratory syncytial virus, influenza virus, parainfluenza virus, and metapneumovirus.

Typical bacteria implicated in acute bronchitis include *Haemophilus influenzae* (Chapter 300) and *Moraxella catarrhalis* (Chapter 300). Severe bronchiolitis has also been reported with *Mycoplasma pneumoniae* (Chapter 317), even though this pathogen is usually associated with either pneumonia or acute bronchitis and tracheitis in adults. Up to 25% of cases of acute bronchitis and tracheitis may be due to "atypical" bacteria: *Bordetella pertussis* (Chapter 313), *Chlamydophila (Chlamydia) pneumoniae* (Chapter 318), and *Mycoplasma pneumoniae* (Chapter 317).

CLINICAL MANIFESTATIONS

The cardinal clinical symptom is cough (Chapter 83) of recent onset. Because most upper respiratory infections resolve within 1 week, a more extended period of cough is useful for considering the diagnosis of acute bronchitis. Patients usually seek care from their physician after 4 to 7 days of coughing that is not resolving. With acute bronchitis, there is often a continued cough and sometimes a worsened cough that lasts an initial 1 to 3 weeks. Associated

symptoms vary and include sputum production, fever, malaise, wheezing, and dyspnea. Adults with pertussis may exhibit paroxysms of coughing, whooping, or vomiting, although less commonly than seen in children with this infection.

DIAGNOSIS

Acutely ill patients may not be able to distinguish their early symptoms from those accompanying very mild upper respiratory infections. However, with acute bronchitis and tracheitis, a protracted phase of coughing persists beyond 1 to 5 days, during which time pulmonary function tests may become abnormal. A substantial proportion of patients will have significant declines in forced expiratory volume in 1 second (FEV_1) (Chapter 85).

When "atypical" bacteria are identified by culture or serology, patients tend to be seen later in the course of their illness than patients with viral causes and more often have wheezing. In some studies, 12 to 32% of patients with coughing that persists for longer than 1 week have pertussis. In other studies, however, pertussis has been confirmed in only 1% of such patients.

Because a major issue is whether symptoms are caused by an organism that could be responsive to antibiotics, the differentiation between bacterial infection compared with viral infection is often more important than determining the precise organism responsible for the infection. Measurement of serum levels of procalcitonin, which is the prohormone of calcitonin, can help detect the presence of bacterial infection and thereby guide the initiation or discontinuation of antibacterial antibiotics.[3] A level lower than 0.1 ng/L makes bacterial infection highly unlikely, and limiting treatment to patients with levels higher than 0.25 ng/L can markedly reduce the use of antibiotics without adverse clinical outcomes.[A1,A2] If the test is available, the finding of very low levels may help the clinician decide against prescribing antibiotics.

In the occasional instance in which it is important to diagnose specific viral cause, rapid diagnostic tests exist for most viruses linked to acute bronchitis and tracheitis. However, their value lies in identifying a virus for which there is therapy or avoiding antibacterial antibiotics if any virus is identified. Not all rapid tests are widely available, and they are expensive and rarely cost-effective in an outpatient setting.

For diagnosing specific bacterial causes, polymerase chain reaction (PCR) testing of nasopharyngeal swabs or aspirates is the easiest and most sensitive way to diagnose infections by *B. pertussis, M. pneumoniae,* and *C. pneumoniae;* most experts recommend calcium alginate swabs for pertussis because cotton inhibits growth. Dacron swabs with aluminum handles are preferred for specimens used to diagnose *Chlamydophila* because cotton, calcium alginate, and the wooden shaft can all inhibit growth of the organism. Cultures for *M. pneumoniae* are slow and insensitive. In general, testing for atypical organisms should not be done because of the cost of PCR and both the insensitivity and slowness of cultures. However, if the clinician suspects an outbreak in the community or the likelihood of pertussis, rapid testing with PCR may be quite beneficial.

TREATMENT

Rx

Although fewer than 50% of cases of acute bronchitis are probably caused by bacterial pathogens, antibiotics are used in 50 to 85% of cases worldwide. In a meta-analysis of randomized trials of antibacterial antibiotics for acute bronchitis, patients receiving antibiotics were significantly less likely to have a cough, but their 0.58 fewer days with cough, 0.46 fewer days of productive cough, and 0.64 fewer days of feeling ill were not statistically significant and were generally offset by the adverse effects of the antibiotics themselves.[A1] When antibiotics are prescribed, azithromycin (500 mg daily for three days) is probably better than other alternatives.[A1] Another option is to delay the initiation of antibiotics and to use them only in patients with persistent symptoms. This strategy slightly reduces patient satisfaction but has had no detectable adverse clinical implications.[A1] Of note is that discolored sputum per se does not appear to predict responsiveness to antibiotics.[A1]

Antibiotics may be used in patients with known atypical pathogens, but even then their effect on outcomes is not clear except to limit the spread of pertussis, especially during a defined outbreak. In adults suspected of having pertussis, erythromycin, 500 mg four times a day for 14 days, is thought to be most effective. However, many patients cannot tolerate erythromycin, and either doxycycline, 100 mg every 12 hours, or a newer macrolide such as azithromycin, 500 mg on day 1 and 250 mg/day thereafter, is effective. The latter two drugs are also active against *C. pneumoniae* and *M. pneumoniae,* although the optimal duration of therapy for acute bronchitis is unknown. A useful range is 5 to 14 days.

During influenza season, anti-influenza agents may be useful in decreasing symptoms by approximately 1 day and may lead to a 0.5-day earlier return to normal activity in patients with influenza.[A1] The first-generation drugs amantadine and rimantadine are ineffective against H3N2 influenza A viruses and are not recommended. Second-generation drugs such as zanamivir (two inhalations of 5 mg each, twice a day) or oseltamivir (75 mg twice a day) can be given for 5 days. Up to 20% of patients given oseltamivir will have nausea or vomiting.

Antihistamines and over-the-counter antitussives and expectorants have no apparent value. Although some subsets of patients with bronchial hyperresponsiveness may benefit from β-agonists, overall they appear to confer no benefit.[A1] No data support the routine use of inhaled steroids. Mucolytic agents may be of small benefit.[A1]

In experimental rhinovirus colds, nonsteroidal drugs, alone or in combination with antihistamines, reduce the severity of symptoms, including cough. However, the widespread use of either type of drug alone or as a combination in naturally occurring, community-acquired bronchitis and tracheitis has not been evaluated. In patients with acute bronchitis, various alternative medicines include *Pelargonium sidoides* (a herbaceous perennial widely used in Europe), extracts of thyme herb and primrose root, and a phytomedicine from essential oils; these medicines have shown some benefit compared with placebo in acute bronchitis, but corroborating studies are needed before they can be routinely recommended.

Because antibiotics are overused and probably lead to community-wide alterations in antibiotic resistance, interventions to reduce overuse could have substantial long-term benefits while also reducing health care costs. Unfortunately, such interventions have had only very modest success.[4,5]

PROGNOSIS

Coughing usually lasts 10 to 14 days, during which time the illness causes significant transient decrements in vitality and social functioning. Limited data on short- and long-term outcomes show that up to 20% of patients have persistent or recurrent symptoms for a month. Antibiotics may reduce symptoms by a fraction of a day, but side effects, the emergence of antibiotic resistance, and cost must be weighed against their modest benefits. The mean duration of an office visit for adults in the United States with upper respiratory tract infections is 14.2 minutes when patients are prescribed antibiotics versus 15.2 minutes without prescription of antibiotics, but antibiotic use is not an independent predictor of visit length. Future widespread use of rapid diagnostic tests for specific bacterial and viral pathogens will be useful in targeting effective therapies.

Grade A References

A1. Albrich WC, Dusemund F, Bucher B, et al. Effectiveness and safety of procalcitonin-guided antibiotic therapy in lower respiratory tract infections in "real life": an international, multicenter poststudy survey (ProREAL). *Arch Intern Med*. 2012;172:715-722.

A2. Schuetz P, Muller B, Christ-Crain M, et al. Procalcitonin to initiate or discontinue antibiotics in acute respiratory tract infections. *Cochrane Database Syst Rev*. 2012;9:CD007498.

A3. Smith SM, Fahey T, Smucny J, et al. Antibiotics for acute bronchitis. *Cochrane Database Syst Rev*. 2014;3:CD000245.

A4. Panpanich R, Lerttrakarnnon P, Laopaiboon M. Azithromycin for acute lower respiratory tract infections. *Cochrane Database Syst Rev*. 2008;1:CD001954.

A5. Spurling GK, Del Mar CB, Dooley L, et al. Delayed antibiotics for respiratory infections. *Cochrane Database Syst Rev*. 2013;4:CD004417.

A6. Llor C, Moragas A, Bayona C, et al. Efficacy of anti-inflammatory or antibiotic treatment in patients with non-complicated acute bronchitis and discoloured sputum: randomised placebo controlled trial. *BMJ*. 2013;347:f5762.

A7. Jefferson T, Jones M, Doshi P, et al. Neuraminidase inhibitors for preventing and treating influenza in healthy adults: systematic review and meta-analysis. *BMJ*. 2009;339:b5106.

A8. Becker LA, Hom J, Villasis-Keever M, et al. Beta2-agonists for acute bronchitis. *Cochrane Database Syst Rev*. 2011;7:CD001726.

A9. Poole P, Black PN, Cates CJ. Mucolytic agents for chronic bronchitis or chronic obstructive pulmonary disease. *Cochrane Database Syst Rev*. 2012;8:CD001287.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Wark P. Bronchitis (acute). *Clin Evid (Online)*. 2011;06:1508. from: http://www.clinicalevidence.bmj.com/x/pdf/clinical-evidence/en-gb/systematic-review/1508.pdf; Accessed January 21, 2015.
2. Wunderink RG, Niederman MS. Update in respiratory infections 2011. *Am J Respir Crit Care Med*. 2012;185:1261-1265.
3. Certain L, Schuetz P. The role of procalcitonin in respiratory infections. *Curr Infect Dis Rep*. 2012;14:308-316.
4. Little P, Stuart B, Francis N, et al. Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: a multinational, cluster, randomised, factorial, controlled trial. *Lancet*. 2013;382:1175-1182.
5. Gonzales R, Anderer T, McCulloch CE, et al. A cluster randomized trial of decision support strategies for reducing antibiotic use in acute bronchitis. *JAMA Intern Med*. 2013;173:267-273.

REVIEW QUESTIONS

1. Which one the following best characterizes the clinical picture of acute bronchitis and tracheitis?

A. Absence of criteria for systemic inflammatory response syndrome
B. Cough that lasts 1 to 3 weeks
C. A syndrome that affects 5% of adults annually
D. Most known causes are viral, yet antibiotics are prescribed in 70% of cases
E. All of the above

Answer: E See Definitions, Epidemiology, and Clinical Manifestations sections. Acute bronchitis is common, with a cough lasting longer than most patients appreciate, and is usually caused by viruses. Too often, antibacterial antibiotics are prescribed.

2. Which one of the following statements is correct?

A. Rapid diagnostic tests in general are relatively inexpensive and appear cost effective for outpatient use.
B. Patients with acute bronchitis have significant declines of FEV_1.
C. Patients with atypical bacteria linked to acute bronchitis, such as *Mycoplasma* species, are seen earlier rather than later in their illness compared with those with viral etiologies.
D. A procalcitonin level of less than 1 ng/L indicates a likely bacterial infection.
E. Antibiotics have good efficacy in treating atypical organisms, such as *Mycoplasma* species, causing acute bronchitis.

Answer: B See Diagnosis section. In patients with acute bronchitis, significant declines in FEV_1 cause many of their symptoms. Rapid diagnostic tests are expensive, and antibiotics are usually not helpful because most infections are caused by viruses. Polymerase chain reaction in fact has good reliability for identifying atypical infections, most of which are linked to patients seeking help later in their illnesses than those with viral infections. A low procalcitonin of less than 1 ng/L suggests infection caused by viruses.

3. Which one of the following statements is correct?

A. Inflammation of the small airways is called bronchiolitis.
B. Patients with acute bronchitis can have wheezing.
C. Bronchiectasis is associated with permanent dilation of bronchi and chronic cough.
D. Chronic bronchitis describes patients who have prolonged cough and sputum production for at least 3 months of the year for two consecutive years.
E. All of the above

Answer: E See Definition section. This question is designed to distinguish the terms bronchiectasis and chronic bronchitis from acute bronchitis. Acute bronchitis can involve the small airways and cause wheezing. Isolated inflammation of the small airways is called bronchiolitis.

97

OVERVIEW OF PNEUMONIA

DANIEL M. MUSHER

Pneumonia, which has been a major cause of death throughout recorded history, affects people of all ages without regard to social class or economic status. It remains a major cause of morbidity and mortality in both the developed and the developing world.

DEFINITION

Pneumonia occurs when an infection of the lung parenchyma causes lower respiratory symptoms, and a pulmonary infiltrate is detected on a chest radiograph. Pneumonias are commonly classified as community acquired,[1] health care associated, and health care acquired. Health care–associated pneumonia occurs in patients who live in a skilled nursing facility, have been hospitalized for more than 2 days in the preceding 90 days, have repeated exposure to a medical facility (e.g., for hemodialysis, wound care, or intravenous antibiotics or chemotherapy) or are immunosuppressed. Hospital-acquired pneumonia either appears 48 hours or more after admission in a patient who did not already have or was incubating pneumonia at the time of admission or develops soon after discharge from a hospital. All other pneumonias are considered to be community acquired.

EPIDEMIOLOGY

Pneumonia is the most common potentially lethal acute infection in the United States, each year affecting 1% of the population and causing more than 1.25 million hospitalizations. The incidence of pneumonia is high among infants and toddlers, declines greatly in childhood, remains relatively uncommon among young adults, but begins to increase after 50 years of age and especially after 65 years of age (E-Fig. 97-1). Not only is bacterial pneumonia more prevalent in elderly patients, it is also more severe, with the risk for death rising steadily with increasing age. Factors predisposing older adults to pneumonia include diminished gag and cough reflexes, poor glottal function, diminished toll-like receptor responses, and less robust antibody responses. These factors are far more prominent in persons who are bedridden—whether at home, in a nursing facility, or in a hospital—than they are in healthy aging persons.

Most adults of any age who develop bacterial pneumonia are likely to have one or more underlying predisposing conditions (Table 97-1). Most common is an antecedent viral respiratory infection, which increases adherence of bacteria to respiratory epithelial cells and damages clearance mechanisms by interfering with ciliary action. Influenza virus (Chapter 364) greatly increases the susceptibility to pneumonia caused by *Streptococcus pneumoniae* (Chapter 289), *Haemophilus influenzae* (Chapter 300), or *Staphylococcus aureus* (Chapter 288), and recent studies have confirmed that most of the deaths attributed to influenza virus during the great pandemic of 1918-1919 were due to bacterial superinfection. Bacterial pneumonia generally does not affect perfectly healthy young adults. Even when outbreaks of pneumococcal pneumonia occur among seemingly healthy adults, such as in military recruits, concurrent viral infection, physical exhaustion, and stress are all thought to play a contributory role. *Haemophilus* and *Moraxella* pneumonia almost always occur in persons with chronic obstructive pulmonary disease (COPD; Chapter 88), and *Pseudomonas* pneumonia (Chapter 306) occurs in patients with COPD, cystic fibrosis (Chapter 89), bronchiectasis (Chapter 90), or other structural abnormalities of the lung, especially if they are also taking corticosteroids.

Other predisposing factors include malnutrition (Chapter 215), which weakens the immune system; excessive alcohol intake (Chapter 33), which suppresses the cough reflex and affects migration of white blood cells (WBCs); cigarette smoking, which increases pulmonary secretions and damages ciliary action; hepatic or renal disease, which decrease antibody formation and WBC function; diabetes mellitus, which decreases WBC function; and immunoglobulin deficiencies of any cause. The nearly 100-fold increase in bacterial pneumonia in young adults who have AIDS (Chapter 391) is thought to be related largely to defective antibody production.

In contrast to bacterial pneumonia, viral or mycoplasmal pneumonia occurs when organisms are transmitted to immunologically naïve hosts. The presence or absence of preexisting immunity and the competence of the immune system itself appear to be principal determinants of whether infection occurs. Immune compromise contributes greatly to the severity of pneumonia due to respiratory syncytial virus (Chapter 362), influenza virus (Chapter 364), and parainfluenza virus (Chapter 363), and pregnancy predisposes to severe pneumonia due to influenza virus or measles virus.

TABLE 97-1 CONDITIONS SUGGESTING VARIOUS CAUSES OF PNEUMONIA

UNDERLYING CONDITION	ASSOCIATED MICROORGANISM
Active smoking/chronic obstructive lung disease	*Streptococcus pneumoniae, Haemophilus influenzae, Moraxella catarrhalis, Pseudomonas aeruginosa*
Nursing home resident	*S. pneumoniae*, gram-negative bacilli, *H. influenzae, Staphylococcus aureus,* microaerophilic and anaerobic mouth flora
Alcoholism	*S. pneumoniae*, gram-negative bacilli, microaerophilic and anaerobic mouth flora, *Mycobacterium tuberculosis,*
Gross aspiration/poor dentition	Microaerophilic and anaerobic mouth flora
Travel to southwestern United States	*Coccidioides immitis*
Residence in Mississippi River basins, exposure to bats	*Histoplasma capsulatum*
Exposure to birds	*Cryptococcus neoformans, Chlamydia psittaci, H. capsulatum*
Exposure to sick psittacine birds	*Chlamydia psittaci*
Exposure to rabbits	*Francisella tularensis*
Exposure to farm animals	*Coxiella burnetii* (Q fever)
Influenza active in community	Influenza virus, *S. aureus, S. pneumoniae, H. influenzae, Streptococcus pyogenes*
Bronchiectasis, cystic fibrosis	*Pseudomonas aeruginosa, Burkholderia cepacia, S. aureus, Aspergillus* species, nontuberculous mycobacteria
Cavitary lung lesion	Microaerophilic and anaerobic mouth flora, *S. aureus,* tuberculous and nontuberculous mycobacteria, endemic fungi
Intravenous drug use	*S. aureus, M. tuberculosis, S. pneumoniae*
Endobronchial obstruction	Microaerophilic and anaerobic mouth flora, gram-negative bacilli, *S. aureus,*
Recent antibiotic therapy	Drug-resistant *S. pneumoniae*
HIV (early)	*S. pneumoniae, H. influenzae, M. tuberculosis*
HIV (late)	The pathogens listed for early HIV infection, plus *Pneumocystis jirovecii, Cryptococcus, Histoplasma, Aspergillus, Mycobacterium kansasii, Mycobacterium avium* complex, *P. aeruginosa,*
Travel to Middle East	Middle East respiratory syndrome (MERS) coronavirus
In the context of bioterrorism	*Bacillus anthracis* (anthrax), *Yersinia pestis* (plague), *Francisella tularensis* (tularemia)

CA-MRSA = community-acquired methicillin-resistant *Staphylococcus aureus*; HIV = human immunodeficiency virus.
Data from Infectious Diseases Society of America/American Thoracic Society. Consensus guidelines on the management of community acquired pneumonia in adults. *Clin Infect Dis.* 2007;44:S27-S72.

PATHOBIOLOGY

Infecting microorganisms may reach alveoli and trigger pneumonia because of aspiration of small amounts of nasopharyngeal secretions or mouth contents, especially during sleep. Some degree of aspiration is a normal occurrence. More substantial aspiration, however, is prominent in older persons, especially those who are frail or bedridden, in whom it can be documented by applying radiopaque material to the posterior pharynx at bedtime and then documenting its presence in the bronchi and lungs by a plain chest radiograph the following morning. Aspiration is the usual cause of *S. pneumoniae* (Chapter 289) or *H. influenzae* (Chapter 300) bacterial pneumonia,

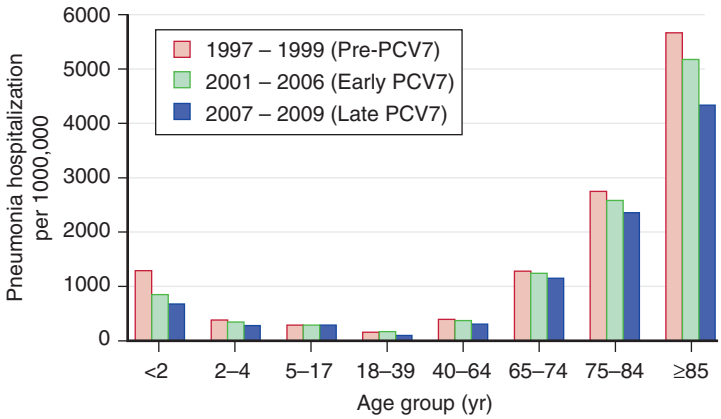

E-FIGURE 97-1. The incidence of community-acquired pneumonia rises at extremes of age. (Data from Griffin MR, Zhu Y, Moore MR, et al. U.S. hospitalizations for pneumonia after a decade of pneumococcal vaccination. *N Engl J Med.* 2013;369:155-163.)

in which the upper airways are colonized with potentially infective bacteria. Aspiration probably is also responsible for the large number of cases in which no causative organism is found but which may be due to mixed microaerophilic and anaerobic organisms of the mouth and upper respiratory tract.

This aspiration of small amounts of secretions should be distinguished from gross aspiration (Chapter 94). Gross aspiration occurs, for example, in persons who have seizures, in persons who choke while vomiting, or in people whose gag reflex is markedly suppressed by alcohol, drugs, or neurologic diseases. In such patients, the clinical syndrome of aspiration pneumonia includes the effects of the aspirated microorganisms and material as well as the accompanying gastric acid.

Another cause of pneumonia is the direct inhalation of aerosolized material into the lungs. This process is uncommon for most bacterial pneumonias but is characteristic of pneumonia due to *Mycobacterium tuberculosis* (Chapter 324) and *Bacillus anthracis* (Chapter 294). This mechanism also explains infection by some viruses, such as influenza virus or respiratory syncytial virus. Some viruses that infect the lung—such as influenza virus (Chapter 364), respiratory syncytial virus (Chapter 362), and human metapneumovirus (Chapter 361)—replicate and spread along cells that line the lower respiratory tract, usually but not always after first gaining entry by inhalation.

Bacteria also may reach the lungs through the blood stream, be filtered by normal host clearance mechanisms, but then escape and cause pneumonia. *S. aureus* pneumonia (Chapter 288) is commonly caused by hematogenous spread, especially if an endovascular infection such as endocarditis is present. *Escherichia coli* (Chapter 304) and other gram-negative rods also may precipitate pneumonia through hematogenous spread.

An array of host defense factors protects the lower respiratory tract against the entry of infectious organisms. The configuration of the upper airways ensures that a thin, laminar flow of air passes over hairs and sticky surfaces that can trap potentially infectious particles. Secretory immunoglobulin A (IgA), which constitutes 10% of the protein in nasal secretions, neutralizes viruses. Immunoglobulins also inhibit bacterial colonization. Closure of the epiglottis prevents food particles from passing into the trachea during swallowing. The larynx prevents the passage of secretions into the trachea and allows the generation of intrapulmonic pressure needed for an effective cough. When microorganisms bypass these mechanisms, ciliary action of the epithelial cells moves them steadily upward toward the larynx, and the cough reflex propels them more rapidly in the same direction. Tracheobronchial secretions maintain moist surfaces, and pulmonary surfactant probably helps to prevent atelectasis, which might interfere with distal clearance.

When potentially infective organisms reach the alveoli, innate and specific defenses come into play. Cells that line the respiratory tract produce substances that inhibit or kill microorganisms, including lysozyme, lactoferrin, β-defensins, and surfactant. Bacterial cell wall components, such as lipopolysaccharide in gram-negative bacteria and peptidoglycan in gram-positive bacteria, activate the alternative complement cascade, leading to opsonization or killing of bacteria. They also upregulate toll-like receptors, with subsequent enhancement of humoral and cellular immune mechanisms. Antibodies to surface-expressed bacterial components greatly enhance the host defense response, and serotype-specific antibodies against bacterial capsular polysaccharide are especially important in protecting against pneumococcal infection.

If these defense mechanisms fail, bacteria may replicate in the alveoli, where they stimulate the local production of cytokines and cause capillary leakage and the accumulation of plasma and inflammatory cells. This sequence initiates a vicious cycle, in which additional inflammatory cells are attracted, and further cytokine release is stimulated. In bacterial pneumonia, the host inflammatory response is responsible for most of the manifestations of disease.

Pathology

Pathologically, pneumonia results from the replication and spread of microorganisms through the pulmonary interstitium and alveoli. The presence of microorganisms and the resulting inflammatory response, characterized by the accumulation of plasma and WBCs in alveoli, explains most of the clinical manifestations of pneumonia. This progressive inflammatory exudate, which is detected radiographically as pneumonia, causes a ventilation-perfusion mismatch and hypoxemia.

Unlike bacterial pneumonia, influenza virus directly invades columnar epithelium cells, thereby resulting in pathologic changes that range from vacuolization of some respiratory epithelial cells to desquamation of the entire epithelial layer. These widespread changes, which lead to a diffuse interstitial pattern on the chest radiograph, also predispose to secondary bacterial invasion. *Chlamydophila pneumoniae* (Chapter 318) adheres to specific receptors and replicates within cells, thereby producing microcolonies that stimulate an inflammatory response and result in focal pneumonia. Mycoplasma also damages respiratory epithelial cells (Chapter 317) but, rather than invading cells, adheres to the cell surface, where it impairs ciliary activity and generates toxic substances. Secondary bacterial pneumonia is uncommon in *Mycoplasma* pneumonia, perhaps because these organisms, unlike the influenza virus, do not adversely affect phagocytic cells.

CLINICAL MANIFESTATIONS

Pneumonia is generally characterized by an acute onset of fever, a cough often with sputum production, and a newly recognized pulmonary infiltrate detected on a chest radiograph. However, patients with pneumonia may not cough, do not always produce sputum, can be afebrile when first evaluated, and may not have obvious radiographic infiltrates, especially if they are high-risk adults with chronic lung disease, are obese, or are evaluated with only a portable chest radiograph. In young adults who have bacterial pneumonia, acute severe malaise and subjective fever are common, often with chills, cough, and sputum production, or at least a sensation of needing to produce sputum. Some patients also have pleuritic chest pain. When sputum is produced, it may be tinged with blood. With advancing age, patients are increasingly likely to have only some or even few of these specific manifestations of pneumonia, in part because they produce lower levels of cytokines and may exhibit a less vigorous response to them. As a result, elderly patients with pneumonia may present with disorientation, confusion, fatigue, or more subtle changes in mental status. Diarrhea, which can be a prominent manifestation of *Legionella* pneumonia (Chapter 314), is also frequent in pneumococcal and probably in other bacterial pneumonias, likely because of a nonspecific gastrointestinal response to circulating cytokines.

Viral pneumonia is more likely to present with upper respiratory symptoms such as rhinorrhea or a sore throat and a dry cough. Patients often recall being exposed to someone with a respiratory infection.

DIAGNOSIS

Physical Examination

Younger patients with bacterial or influenza pneumonia appear acutely ill, in contrast to elderly and frail persons, who may simply appear listless. Younger adults with noninfluenzal viral, *Mycoplasma*, or *Chlamydophila* pneumonia (Chapters 317 and 318) often do not look as acutely ill. Patients with tuberculosis or other more chronic forms of pneumonia may appear chronically ill or may look relatively well.

A respiratory rate more than 20 breaths per minute is distinctly abnormal, and a rate of more than 25 breaths per minute should cause serious concern. An oxygen saturation (Sao_2) of less than 92% is likely to indicate a very low partial pressure of oxygen, and a low saturation together with a rapid respiratory rate suggests serious respiratory compromise. In a patient with pneumonia and tachypnea, an Sao_2 of less than 90% should raise concern about impending respiratory distress (Chapter 104).

In bacterial pneumonia, crackles or rales are generally present over the affected area. Bronchial breath sounds and egophony strongly suggest pneumonia when present but are not sensitive for diagnosis. Dullness to percussion over the affected area may be detected in about one half of cases. Increased tactile fremitus is often present and is especially useful in distinguishing a pulmonary infiltrate from a pleural effusion, in which fremitus is diminished or absent. The failure to detect excursion of the diaphragm by percussion suggests an effusion. Unfortunately, the overall sensitivity and specificity of the physical examination for pneumonia is fairly low. As a result, the diagnosis of pneumonia requires radiographic validation.

Radiographic Findings

Pneumonia is usually diagnosed by the presence of an infiltrate on a chest radiograph or excluded by the absence of an infiltrate. A dense consolidation that involves a segment or a lobe of the lung is very likely to reflect an acute bacterial infection. Many patients with bacterial pneumonia, however, have radiographic infiltrates that are not clearly segmental. Small areas of alveolar consolidation may be missed by a chest radiograph, especially an anterior-posterior portable radiograph, but be detected by the far more sensitive computed tomography (CT) scan.[2] However, small areas of consolidation (ground-glass appearance) are often described on the CT scan of patients who do not have pneumonia.

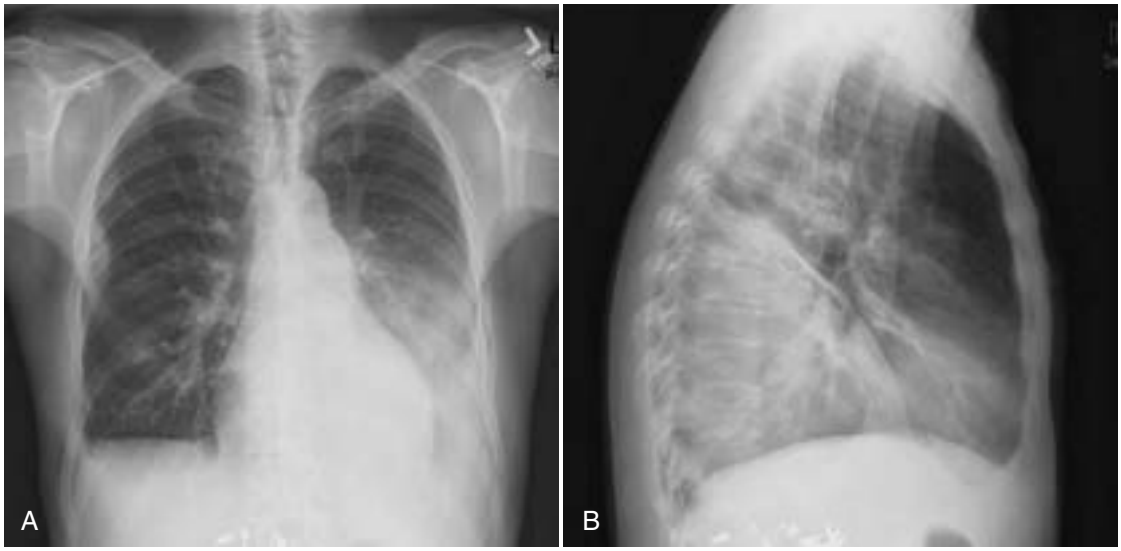

FIGURE 97-1. Pneumococcal left lower lobe pneumonia as seen in posterior-anterior (**A**) and lateral (**B**) views.

Although the presence of an infiltrate is the key to making a diagnosis of pneumonia, its radiographic appearance provides very little insight into the etiology. Dense consolidation of a segment or lobe is usually bacterial (Fig. 97-1), especially pneumococcal, but other bacteria, including *Legionella* (Fig. 97-2), may cause a similar picture, but many bacterial pneumonias do not cause segmental or lobar infiltrates.

Aerogenous *S. aureus* pneumonia (Chapter 288) presents as a segmental or lobar pneumonia, whereas hematogenous *S. aureus* pneumonia and gram-negative bacilli often cause necrotizing pneumonia, which is defined as a cavitary lesion in a pneumonic infiltrate. However, pneumococcal pneumonia also causes lung necrosis that is visible on a chest radiograph in 2% of cases and on a CT scan in 11% of cases (see Fig. 97-2).

Hematogenous *S. aureus* pneumonia (Chapter 288), especially when seen with endocarditis or an infected intravascular source (Chapter 76), can present with a distinctive radiographic appearance of several 1- to 3-cm round lesions, which are likely to cavitate (Fig. 97-3A and B). Subsegmental or "patchy" pneumonia (Fig. 97-4) may be due to bacteria, viruses, *Mycoplasma*, or *Chlamydophila* (Chapters 317 and 318). *Pneumocystis jiroveci* (Chapter 341) causes a diffuse interstitial infiltrate that may, in its earlier clinical stages, be mistaken for prominent pulmonary markings. Aspiration of mixed anaerobic, microaerophilic, and facultative bacteria from the mouth may cause pneumonia but may also lead to a lung abscess (Chapter 90) with a thick wall, a fluid level, and surrounding consolidation, especially in the superior segments of the lower lobes or posterior segments of the upper lobes. A cavitary lesion of an upper lobe without a fluid level, especially if confined to the posterior segment, suggests tuberculosis (Chapter 324). Occasionally, more acute presentations of tuberculosis may mimic acute bacterial pneumonia. *Aspergillus* (Chapter 339) can grow as a mass within a cavity, causing the distinctive appearance of an intracavitary mycetoma (fungus ball) surrounded by an arc or halo of air (E-Fig. 97-2).

Rapidly progressive pneumonia of any cause may result in diffuse pulmonary infiltrates consistent with the acute respiratory distress syndrome (Chapter 104). Although the appearance or enlargement of infiltrates after hospitalization is often attributed to fluid repletion, such progression more likely reflects the ongoing inflammatory response. A chest CT scan may help clarify the nature of an infiltrate and determine whether an effusion or mass is present, but it is usually not necessary at hospital admission for patients in whom a good-quality chest radiograph can be obtained.

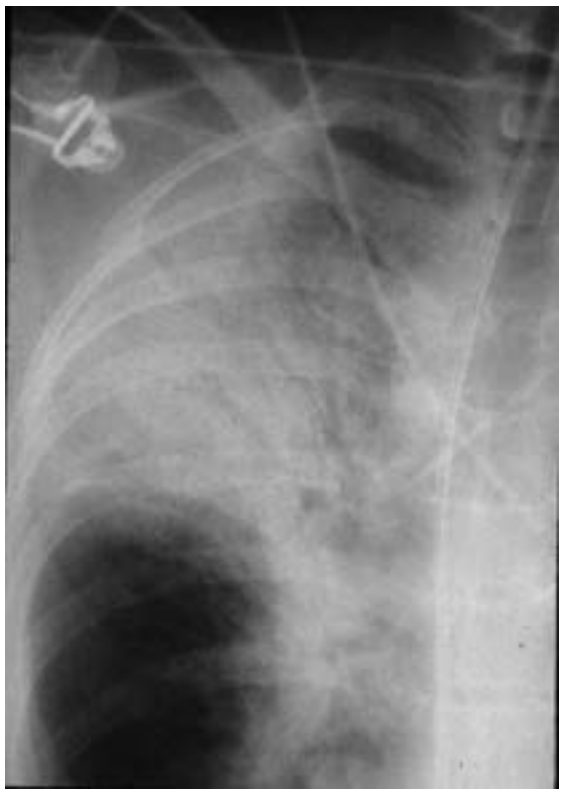

FIGURE 97-2. Dense lobar consolidation with air bronchograms in a patient with proven *Legionella* pneumonia.

Laboratory Findings

Most patients with bacterial pneumonia have a WBC count higher than 11,500/μL at the time of admission to hospital, and about one third have a WBC count higher than 15,000 WBC/μL. A low WBC count should not be interpreted as reassuring because WBC counts of 6000/μL or less may be seen in overwhelming bacterial infection. When overwhelming bacterial infection suppresses the WBC count, immature (band) cells are almost always elevated. About 40% of uninfected patients who present to the hospital with a syndrome consistent with community-acquired pneumonia, such as patients with pulmonary edema or lung cancer, also have WBC counts higher than 11,500/μL, so an elevated WBC count is by no means specific for pneumonia. Nevertheless, WBC counts higher than 20,000/μL are unusual in acute pulmonary conditions other than bacterial pneumonia. Mild nonspecific elevations in the serum bilirubin level, aminotransferase levels, and lactate dehydrogenase (LDH) level are often noted. Marked elevations of the LDH level can be seen in *Pneumocystis* and *Histoplasma* pneumonia (Chapters 332 and 341) in AIDS patients.

An elevated serum procalcitonin level increases the likelihood of a bacterial infection, whereas a low level opposes such a diagnosis. In randomized trials of patients who present with lower respiratory tract symptoms, even rather severe symptoms, treatment guided by a procalcitonin level

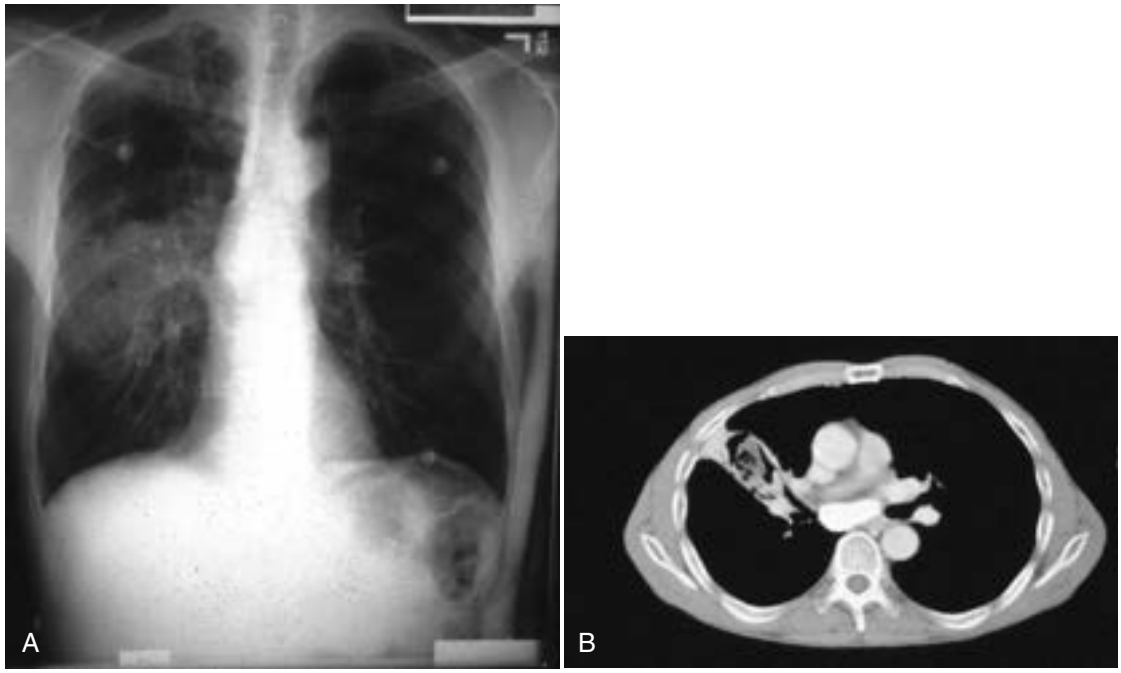

E-FIGURE 97-2. A pulmonary aspergilloma, sometimes called an intracavitary fungus ball, on a plain chest radiograph (**A**) and chest tomography (**B**), which shows the fungal growth within a cavity and attached to the wall by a stalk.

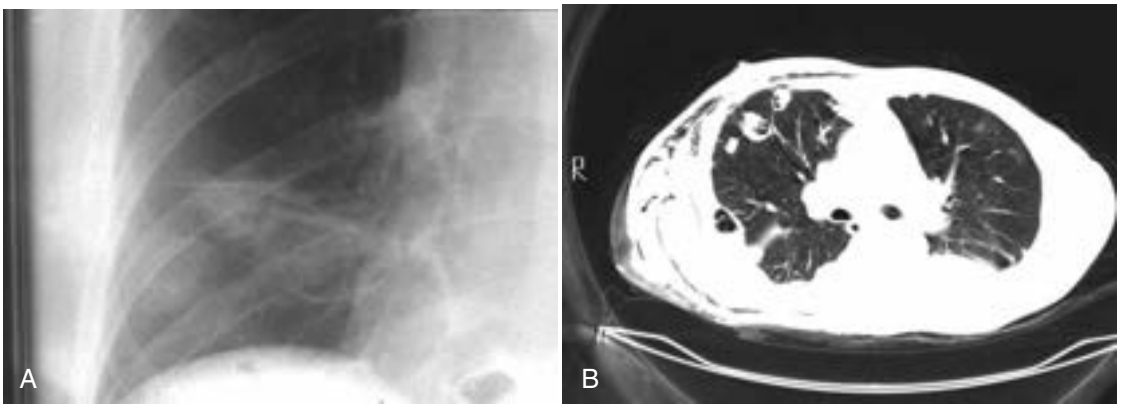

FIGURE 97-3. Hematogenous *Staphylococcus aureus* pneumonia. **A,** The chest radiograph shows the characteristic round lesions with cavitation. **B,** The same lesions confirmed by computed tomographic scan.

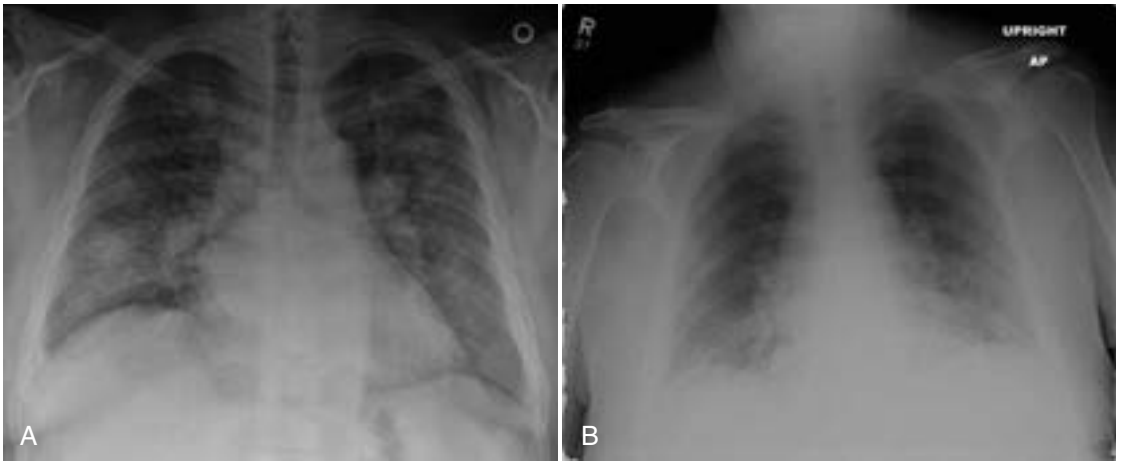

FIGURE 97-4. **A,** Bilateral "patchy" infiltrates in a patient with coronavirus pneumonia. **B,** Chest radiograph of community-acquired pneumonia ultimately proved to be caused by human metapneumovirus infection.

(antibiotics discouraged if the level is ≤0.25 μg/L and strongly discouraged if <0.1 μg/L) has resulted in less use of antibiotics without any adverse overall effects.■■ However, up to 25% of patients with bacterial pneumonia have a normal procalcitonin level, and about 25% of patients with a pneumonia syndrome but no evidence for bacterial infection have an elevated procalcitonin level, so this test cannot be used alone to determine therapeutic decisions.

Microbiologic Diagnosis

The respiratory tract clears inflammatory exudate by the ciliary action of cells that line the bronchi and trachea as well as by the cough reflex. Sputum is composed of this exudate—plasma, white blood cells, and bacteria—with a greater or lesser admixture of saliva. The presence of large numbers of a single type of bacterium in an inflammatory specimen that is relatively free of contaminating epithelial cells strongly suggests this organism as the etiologic agent of the pneumonia (Fig. 97-5). *Haemophilus* (Chapter 300), *Moraxella* (Chapter 300), and gram-negative rods (Chapters 305 and 306) are even more distinctive in their microscopic appearance. If antibiotics have not already been administered, the absence of visible organisms in an inflammatory specimen suggests that the cause of pneumonia is either a bacterium that does not readily accept Gram stain (e.g., *Legionella* or *Mycobacteria*), or another kind of organism such as *Mycoplasma*, *Chlamydophila*, or a virus that lacks typical bacterial cell walls and does not take up Gram stain. Reports of the poor sensitivity of sputum to detect bacteria largely reflect the inclusion of specimens that are inadequate or have been obtained after antibiotic therapy has been begun. In patients with bacteremic pneumococcal pneumonia who cough up a valid specimen and have not received antibiotics, the sensitivity of Gram stain or culture is each about 90% for detecting pneumococci. Because the sensitivity of these standard microbiologic tests falls dramatically after 18 hours of antibiotic treatment, specimens are useful diagnostically only if they are collected in a timely fashion.

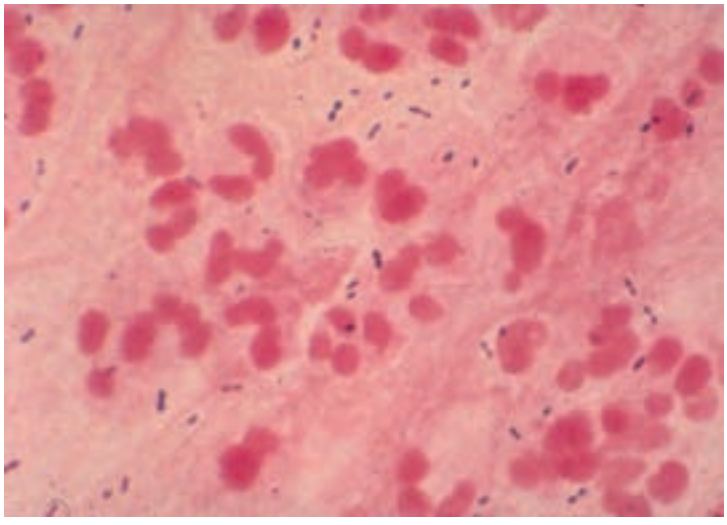

FIGURE 97-5. Gram stain of sputum from a patient with pneumococcal pneumonia shows large numbers of polymorphonuclear leukocytes and many lancet-shaped gram-positive cocci with no epithelial cells, indicating that this specimen originated in the lower airways. Such a specimen is diagnostic of pneumococcal pneumonia, although it cannot exclude coexisting infection by an organism such as a virus that is not seen by Gram stain.

Examination of a Gram-stained sputum specimen is also useful if it does not show bacteria. If antibiotics have not already been administered, the absence of visible organisms in an inflammatory specimen suggests that the cause of pneumonia is either a bacterium that does not readily accept Gram stain (e.g., *Legionella* or *Mycobacteria*), an organism such as *Mycoplasma* or

Chlamydophila that lacks typical bacterial cell walls, or a virus that does not take up Gram stain. In the past, diagnosis of infection due to these organisms relied upon unreliable serologic techniques. The availability of modern diagnostic techniques is likely to redefine the role of these agents in causing community-acquired pneumonia.

Bacterial cultures will readily yield *Haemophilus, Moraxella, S. aureus*, or gram-negative bacilli when these organism cause pneumonia. However, finding *S. aureus* or gram-negative rods by culture when they have not been seen microscopically in a good-quality sputum suggests that these are contaminating mouth flora. Detection of pneumococcus on a sputum culture may be more difficult because of the prevalence of other α-hemolytic streptococci in saliva.

Enzyme-linked immunosorbent assay (ELISA) can detect pneumococcal cell wall or capsular polysaccharide in the urine of 60 to 80% of patients with bacteremic pneumococcal pneumonia and a smaller proportion of those with nonbacteremic disease. An ELISA for urinary *Legionella* antigen detects only the most common *Legionella* serotype but is positive in about 70% of cases of *Legionella* pneumonia (Chapter 314), with higher sensitivity in more severe disease. *Histoplasma* (Chapter 332) urine and *Cryptococcus* (Chapter 336) serum antigen tests are positive in patients with disseminated disease but are less likely to be positive in patients with discrete pulmonary infiltrates.

Polymerase chain reaction (PCR) testing of sputum is a highly sensitive technique that may be nonspecific because it can detect colonization rather than infection. When used in African patients with AIDS and suspected pneumonia, quantitative PCR testing on a nasopharyngeal swab reliably identified pneumococcal pneumonia.[3] The generalizability of this method to patients in developed countries remains to be determined. For organisms that do not normally colonize the upper airways, PCR is specific as well as sensitive. PCR on a throat swab can reliably detect *Chlamydophila* and *Mycoplasma* as well as 15 respiratory viruses (including influenza, parainfluenza, respiratory syncytial virus, human metapneumovirus, coronavirus, and adenovirus) with very high sensitivity and specificity. As a result, PCR has generally replaced viral culture as the gold standard for diagnosing influenza virus infection (Chapter 364). Sputum PCR also can detect *M. tuberculosis* (Chapter 324) and is now part of the recommended evaluation in patients suspected of having this diagnosis.

Bacteremia is documented in about 10% of patients who are hospitalized for community-acquired pneumonia, including in about 25% of patients with pneumococcal pneumonia, 10 to 15% of patients who are hospitalized for aerogenous pneumonia due to *S. aureus* or gram-negative rods, and a lower proportion of patients with nontypable *H. influenzae* pneumonia, and only rarely in pneumonia caused by *Moraxella catarrhalis*. By comparison, patients with hematogenous *S. aureus* pneumonia virtually always have positive blood cultures.

Differential Diagnosis

Streptococcus pneumoniae (Chapter 289), which is the most commonly identified infectious cause of community-acquired pneumonia, causes up to 20% of cases in hospitalized patients in the United States[4] but a greater proportion of cases in Europe, perhaps because of higher rates of smoking in Europe and higher rate of pneumococcal vaccination in the U.S. *H. influenzae* (Chapter 300), *S. aureus* (Chapter 288), *P. aeruginosa* (Chapter 306), and other gram-negative bacilli (Chapters 304 and 305) are the next most common bacterial causes of community-acquired pneumonia. *Haemophilus* generally causes pneumonia only in persons who have preexisting bronchopulmonary disease. *Pseudomonas* and other gram-negative rods generally cause pneumonia in patients who have structural lung disease and who are receiving corticosteroids or are otherwise immunocompromised.

When influenza is active in the community, this virus is identified in a substantial proportion of patients admitted to an intensive care unit because of community-acquired pneumonia.[5] Identification of influenza virus in a patient with pneumonia should lead to appropriate antiviral treatment (Chapter 360), even if more than 48 hours have passed since the onset of symptoms. Evidence for viral infection is found in 20 to 30% of all adults hospitalized for community-acquired pneumonia. Rhinovirus (Chapter 361) is most commonly recognized, but its role as a cause of pneumonia is not well established. Respiratory syncytial virus (Chapter 362), coronavirus (Chapter 366), and human metapneumovirus (Chapter 361) are clearly implicated as causes of pneumonia and are the next most commonly found viruses in adults hospitalized for pneumonia. Outbreaks of adenovirus (Chapter 365) pneumonia may occur in military recruits. Because coinfecting bacteria are also detected in about half of cases of documented viral pneumonia,[6] identifying a virus by PCR in a patient with pneumonia does not prove that the virus is the cause or especially the sole cause of illness. Other features, such as the history, severity of illness, sputum production, nature of the pulmonary infiltrate, WBC count, and procalcitonin level may help to establish or refute a diagnosis of viral pneumonia.

Although *Mycoplasma* and *C. pneumoniae* (Chapters 317 and 318) may be common causes of pneumonia in the ambulatory setting, they less frequently cause disease that requires hospitalization. Pneumonia caused by these organisms is characterized by prolonged nonproductive cough, low-grade fever, and scattered pulmonary infiltrates. PCR is preferred over serologies for diagnosing *Chlamydophila* or *Mycoplasma* infections.

Other Causes of a Pneumonia Syndrome

Fever, cough, and sputum production without an infiltrate is called acute bronchitis (Chapter 96). In persons who do not have chronic lung disease, acute bronchitis is generally a self-limited viral illness, but bacterial infection, for example with *H. influenzae* (Chapter 300), may cause bronchitis in patients who have COPD. Persistent cough (Chapter 83) of several weeks' duration without fever or sputum suggests pertussis (Chapter 313) or a postviral infection syndrome due, for example, to adenovirus.

Epidemiologic clues may suggest specific infectious causes of pneumonia (Table 97-2). *Coccidioides immitis* (Chapter 333), found in arid regions of the Americas, or *Histoplasma capsulatum* (Chapter 332), found worldwide but especially in river basins of North America, cause a variable proportion of community-acquired pneumonia in endemic regions. Exposure to livestock or late summer residence in a hot and dry ranching area suggests *Coxiella burnetii* (Q fever) (Chapter 327), especially if patients have a severe headache and abnormal liver enzymes. Exposure to sick psittacine birds raises concern for *Chlamydia psittaci* (Chapter 318).

Tuberculosis (Chapter 324) should be suspected in persons who have lived in endemic areas, patients who have served time in prison or been homeless, and patients who are immunocompromised, especially patients with AIDS. *Mycobacterium kansasii* (Chapter 325) may cause an identical syndrome in patients with none of those risk factors. *Mycobacterium avium,* often called *Mycobacterium avium-intracellulare* (MAI) or *Mycobacterium avium* complex (MAC), causes diffuse bilateral pneumonia in immunocompromised patients (Chapter 325). *Mycobacterium intracellulare* also is a well-known cause of pneumonia in adults, usually in men with bronchiectasis (Chapter 90) or extensive lung scarring owing to emphysema (Chapter 88) or previously treated tuberculosis. This organism also causes disease in

TABLE 97-2 COMMON CAUSES OF PNEUMONIA SYNDROME

REQUIRING HOSPITAL ADMISSION	Uncommon
Common	*Nocardia*
Streptococcus pneumoniae	*Legionella*†§
Haemophilus influenzae	*Chlamydophila*§
Staphylococcus aureus	*Mycoplasma*§
Influenza virus,† other respiratory viruses	Anaerobic bacteria
Lung cancer	Cryptogenic organizing, eosinophilic, and other noninfectious pneumonias
Pulmonary edema	Sarcoidosis
Mycobacterium tuberculosis	Kaposi sarcoma
Pneumocystis jiroveci	Q fever‡
Less Common	*Coccidioides immitis*‡
Moraxella catarrhalis	**OUTPATIENT TREATMENT MAY BE ADEQUATE**
Pseudomonas	*Streptococcus pneumoniae*
Klebsiella	*Mycoplasma pneumoniae*
Nontuberculous mycobacteria	*Haemophilus influenzae*
Histoplasma‡	*Chlamydophila pneumoniae*
Cryptococcus	Respiratory viruses*
Pulmonary infarction	

*Routine use of polymerase chain reaction (PCR) technology substantially increases the recognition of these agents.
†Likely to be associated with secondary bacterial infection.
‡Strong dependence on geographic exposure.
§True incidence of pneumonia caused by these organisms will be clarified by routine use of PCR technology.

middle-aged women who lack these risk factors; in such patients, subtle infiltrates may be missed by routine chest radiography.

Patients with HIV infection are susceptible to a variety of pulmonary infections depending on how immunocompromised they are (see Table 97-2 and Chapter 391). These opportunistic infections include typical and atypical tuberculous and nontuberculous mycobacteria, *Pneumocystis* (Chapter 341), *Histoplasma* (Chapter 332), and *Cryptococcus* (Chapter 336). AIDS patients also have a 50- to 100-fold increased risk for developing pneumococcal disease.

Even after an exhaustive evaluation, no causative organism is identified in about half of patients who are hospitalized for symptoms, physical examination findings, laboratory abnormalities, and radiographic changes consistent with community-acquired pneumonia. The presumption is that bacterial infection is responsible for most of these pneumonias because they generally respond to antibiotic therapy.

Noninfectious Considerations

Many noninfectious conditions cause patients to present with a syndrome consistent with acute or subacute pneumonia (see Table 97-1). Cryptogenic organizing pneumonia (Chapter 91), acute interstitial pneumonia, eosinophilic pneumonia, and other interstitial pneumonias (Chapter 92) are uncommon conditions that almost always are initially misdiagnosed as community-acquired pneumonia. Pulmonary hemorrhage and vasculitis may also cause pulmonary infiltrates and fever. In ANCA-associated granulomatous vasculitis, these infiltrates may also be associated with cavitary lesions. Attention to the patient's history may reveal a longer history of symptoms, and careful review of earlier chest radiographs may reveal prior radiographic abnormalities, consistent with a chronic, noninfectious process. Pulmonary embolus with infarction can cause pleuritic chest pain and pulmonary infiltrates, with sputum that contains neutrophils but few or no bacteria. Patients with septic pulmonary emboli should be assessed for other foci of infection, such as an infected heart valve or intravascular device.

Pulmonary edema (Chapter 58) is the most common noninfectious cause of a community-acquired pneumonia-like syndrome in middle-aged and older patients. The diagnosis should be made based on history, physical examination, and radiographic findings, supported by elevated B-natriuretic peptide levels. Patients with lung cancer (Chapter 191) commonly present with fever and a pulmonary infiltrate, which sometimes is attributed to a postobstructive pneumonia. Acute respiratory distress syndrome (Chapter 104) in response to a serious nonpulmonary infection is often indistinguishable from pneumonia because it commonly presents with fever, lung crackles, an elevated WBC count, and pulmonary infiltrates. In light of all these possibilities, at least brief consideration should be given to other infectious and noninfectious causes before treating presumptive community-acquired pneumonia with guideline-recommended empirical therapy.

TREATMENT

Hospital Admission

Scoring systems can help determine whether a patient with community-acquired pneumonia requires hospitalization (Tables 97-3 and 97-4). A corollary decision, whether a patient should be admitted to an intensive care unit (ICU), can be guided by the respiratory rate, heart rate, systolic blood pressure, oxygenation, mental status, extent of pulmonary involvement, serum albumin level, and arterial pH (Table 97-5).[7] This decision aid is 92% sensitive in detecting patients who will require ICU transfer for intensive respiratory or vasopressor support. Other indicators of overwhelming infection that indicate a need for ICU admission include a WBC count of 6000/μL or lower in bacterial pneumonia (especially if increased band forms are present), thrombocytopenia, and hypothermia.

Supportive therapy should include fluid replacement to maintain blood pressure (e.g., an average of 4.5 liters of fluids with electrolytes will be needed in a patient with pneumonia and septic shock [Chapter 108]), oxygen for hypoxemia, and mechanical ventilation (Chapter 105).

Antibiotic Therapy

Based on good evidence that outcomes are worse with prolonged delays, initial antibiotics should be given as soon as the diagnosis of pneumonia is considered likely, regardless of whether that be in a physician's office or an emergency department. Guidelines for the empirical antibiotic therapy of community-acquired pneumonia (Table 97-6)[8,9] focus on common infectious causes of pneumonia and are generally successful because a specific etiologic agent is infrequently determined in outpatients who have pneumonia. However, this approach should not discourage the careful consideration of possible noninfectious causes of fever and pulmonary infiltrates or discourage

TABLE 97-3 PNEUMONIA OUTCOMES RESEARCH TRIAL (PORT) SEVERITY INDEX

	POINTS ASSIGNED FOR EACH CRITERION
VITAL SIGNS	
Pulse >125/min	10
Systolic BP <90 mm Hg	20
Temp <35 or >40°C	15
Respiratory rate >30/min	20
HISTORY OF CO-MORBID CONDITIONS	
Neoplasm (active, not skin)	30
Cirrhosis or chronic hepatitis	20
Heart failure, stroke, chronic renal insufficiency	10
Altered mental status	20
DEMOGRAPHY	
Age	age (subtract 10 for women)
Nursing home resident	10
LABORATORY DATA	
Arterial pH <7.35	30
BUN >30 mg/dL	20
Serum sodium <130 mEq/L	20
Glucose >250 mg/dL	10
Hematocrit <30%	10
pO2 <60 mm Hg or O2 saturation <90%	10
Pleural effusion on chest radiograph	10

For mortality associated with various PORT scores, see Table 97-4.
Data from Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med.* 1997;336:243-250.

TABLE 97-4 PORT SEVERITY INDEX (PSI) AND MORTALITY AT 30 DAYS

		MORTALITY		
POINT SCORE	**CLASS**	***Community-Acquired Pneumonia****	***Community-Acquired Pneumonia**†*	***S. pneumoniae**‡*
≤70	II	<1%	3%	—
71-90	III	3%	4%	3%
91-130	IV	8%	8%	21%
>130	V	29%	22%	35%

*Original calculation in patients with community-acquired pneumonia report of PORT score (Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med.* 1997;336:243-250.)
†Mortality in patients admitted for pneumonia during a 1-year period, Veterans Affairs Medical Center, Houston (patients with noninfectious causes were excluded) (Musher DM, Roig IL, Cazares G, et al. Can an etiologic agent be identified in adults who are hospitalized for community-acquired pneumonia: results of a one-year study. *J Infect.* 2013;67:11-18.)
‡Results in patients with proven pneumococcal pneumonia (Musher DM, Alexandraki I, Graviss EA, et al. Bacteremic and nonbacteremic pneumococcal pneumonia. A prospective study. *Medicine [Baltimore].* 2000;79:210-221.)

reasonable attempts to establish a specific infectious diagnosis, especially in patients who do not respond promptly.

Outpatient Antibiotic Regimens

For empirical outpatient therapy, guidelines from the Infectious Diseases Society of America and the American Thoracic Society[8] recommend a macrolide, doxycycline, a "respiratory" quinolone (levofloxacin or moxifloxacin, but not ciprofloxacin, which is thought to be slightly less effective against pneumococci), or a β-lactam together with a macrolide. These recommendations are based on a desire to provide therapy effective for common bacterial causes of pneumonia such as *S. pneumoniae* (Chapter 289), *H. influenzae* (Chapter 364), *M. catarrhalis* (Chapter 300), and *Legionella* (Chapter 314), as well as infections due to *Mycoplasma* (Chapter 317) or *Chlamydophila* (Chapter 318).

In contrast, the Swedish Society of Infectious Diseases[9] recommends outpatient treatment for acute pneumonia with oral penicillin or amoxicillin. The rationale for this approach is that pneumococcus, which is the most likely potentially dangerous cause of community-acquired pneumonia, is much better treated by penicillin or amoxicillin than by doxycycline or macrolides, whereas a patient who fails to respond to penicillin or amoxicillin within a few days can be switched to a macrolide or doxycycline to treat potential *Mycoplasma* and *Chlamydophila*. *Legionella* pneumonia presents with an acute syndrome that is indistinguishable from pneumococcal pneumonia but, in the absence of a specific history of exposure, only rarely causes pneumonia in an outpatient setting. In the United States, one third of *Haemophilus* and the majority of *Moraxella* produce β-lactamase, which would make amoxicillin plus clavulanic acid a better choice in patients who have underlying lung disease. Patients with pneumonia and a history of low-grade fever and cough for more than 5 to 6 days should be treated with a macrolide or doxycycline.

In-Hospital Antibiotic Regimens

In a patient who is sick enough to be hospitalized, the physician should make a conscientious effort to determine an etiologic agent; initial therapy may be empirical, but antibiotics should be tailored to an identified causative organism. Recommended initial empirical antibiotic treatment of pneumonia that does not require ICU care (see Table 97-6) includes a respiratory fluoroquinolone (levofloxacin or moxifloxacin, but not ciprofloxacin) or a β-lactam (cefotaxime, ceftriaxone, ampicillin, or ampicillin-sulbactam), together with a macrolide. Either of these regimens will treat pneumonia due to *S. pneumoniae*, and they will also be effective against *Haemophilus, Moraxella, Legionella, Mycoplasma*, and *Chlamydophila*, as well as some less common organisms. However, this approach risks both overtreatment and undertreatment.

To avoid overtreatment,[10] every effort should be made to identify the responsible organism and narrow treatment accordingly. However, the

TABLE 97-5 SMART-COP SCORING SYSTEM*

	POINTS
Low **S**ystolic blood pressure (<90 mm Hg)	2
Multilobar involvement (on chest radiograph)	1
Low **A**lbumin (<3.5 g/dL)	1
High **R**espiratory rate (≥25 if <50 years old, ≥30 if >50 years old)	1
Tachycardia (heart rate >125 beats/min)	1
New-onset **C**onfusion	1
Poor **O**xygenation (PaO_2 <70 mm Hg if <50 years old, <60 mm Hg if >50 years old)	2
Low arterial **p**H (<7.35)	2

*The risk for needing intensive respiratory or ventilatory support is low if ≤2 points, about 10-15% if 3-4 points, about 35% if 5-6 points, and about 65% if ≥7 points.
Data from Chalmers JD, Singanayagam A, Hill AT. Predicting the need for mechanical ventilation and/or inotropic support for young adults admitted to the hospital with community-acquired pneumonia. *Clin Infect Dis.* 2008;47:1571-1574; and Charles PG, Wolfe R, Whitby M, et al. SMART-COP: a tool for predicting the need for intensive respiratory or vasopressor support in community-acquired pneumonia. *Clin Infect Dis.* 2008;47:375-384.

TABLE 97-6 EMPIRICAL TREATMENT FOR COMMUNITY-ACQUIRED PNEUMONIA

OUTPATIENTS*

For syndromes suggesting "typical" bacterial pneumonia (acute onset of cough, sputum, high fever, high white cell count, elevated procalcitonin level):
- Amoxicillin-clavulanic acid (500-125 mg every 6 hr for 5-7 days); add azithromycin (500 mg orally on day 1, followed by 250 mg orally each day on days 2-5) if *Legionella* is a consideration, *or*
- Levofloxacin (750 mg daily), moxifloxacin (400 mg daily), or gatifloxacin (320 mg daily) for 5 days

For syndromes suggesting influenza pneumonia:
- Oseltamivir (75 mg twice daily for 5 days); observe for secondary bacterial infection

For syndromes suggesting viral pneumonia other than influenza (exposure to someone with viral infection, upper respiratory tract symptoms, patient doesn't look very ill, WBC <10,500, procalcitonin level not elevated):
- Symptomatic therapy

For subacute syndromes suggesting *Mycoplasma* or *Chlamydophila* pneumonia:
- Azithromycin (500 mg on day 1, followed by 250 mg daily for 5 days) or doxycycline (100 mg twice daily for 7 days)

INPATIENTS

Patients hospitalized for pneumonia are sufficiently likely to have a bacterial infection that antibacterial agents are nearly always prescribed unless an alternate diagnosis is strongly suspected. In every hospitalized patient, all reasonable efforts should be made to determine an etiologic diagnosis. For initial empiric therapy:
- A beta-lactam (ceftriaxone 1gm daily, cefotaxime 1 gm every 6 hr, or ceftaroline 600 mg every 12 hr) AND a macrolide (azithromycin 500 mg)
- *or*
- A quinolone (levofloxacin 750 mg, moxifloxacin 400 mg or gatifloxacin 320 mg daily)
- Initial therapy is IV until patient is clinically stable; thereafter switch to oral therapy, tailoring therapy based on culture results or selecting agent(s) that provide similar coverage. Total duration of antibiotic therapy generally 5-8 days (see text).

If influenza is likely:
- Oseltamivir (75 mg twice daily for 5 days) with vigilant observation for possible secondary bacterial infection

If influenza is complicated by secondary bacterial pneumonia, add to oseltamivir:
- Ceftaroline, ceftriaxone or cefotaxime plus either vancomycin or linezolid (ceftaroline alone may suffice if it is licensed to treat *S. aureus* pneumonia)

If *S. aureus* is likely, add to the usual empiric antibacterial regimen:
- Vancomycin (15-20 mg/kg every 8-12 hr, monitoring to maintain trough serum levels at 15-20 mcg/ml or linezolid (600 mg every 12 hr); see above comment regarding ceftaroline. Duration of treatment for proven *S. aureus pneumonia* is 10-14 days.

If *Pseudomonas* or other Gram negative organism is likely:
- Antipseudomonal beta-lactam: piperacillin-tazobactam (4.5 gm every 6 hr), cefepime 1-2 gm every 6-8 hr
- *or*
- A carbapenem (meropenem 500mg IV every 6 hr or 1 gm every 8 hr or imipenem-cilastatin 500mg IV every 6 hr or 1 gm every 8 hr; extended infusions of carbapenems may be preferable), plus azithromycin as above†

Adapted from Musher DM, Thorner AT. Community-acquired pneumonia. *N Engl J Med.* 2014;371:1619-1628; and from Infectious Diseases Society of America/American Thoracic Society. Consensus guidelines on the management of community acquired pneumonia in adults. *Clin Infect Dis.* 2007;44:S27-S72.
*A decision to treat pneumonia as an outpatient should be made after assessing need for hospitalization and only if follow-up contact is planned.
†Add an aminoglycoside in patients with severe community-acquired pneumonia in whom *P. aeruginosa* or an antibiotic-resistant Gram-negative organism is likely because susceptibility is difficult to predict. Narrow therapy to one agent with activity against gram-negative bacilli once susceptibility results are available.
IM = intramuscularly; IV = intravenously.

addition of a macrolide to a β-lactam may improve outcomes in pneumococcal pneumonia.

To avoid undertreatment, possible *S. aureus* infection is an important consideration, especially in patients who may have influenza or a history of injection drug use, chronic renal failure, prior corticosteroid therapy, or progression despite outpatient antibiotics.[11] In such patients, the prevalence of methicillin-resistant *S. aureus* (MRSA) in the community mandates further consideration of vancomycin or linezolid; ceftaroline (not yet approved for this purpose) also may prove to be useful in the future. The absence of nasal carriage of MRSA on PCR testing reduces the likelihood that this organism is causing pneumonia.

ICU Antibiotic Regimens

For patients who require ICU admission, a β-lactam (cefotaxime, ceftriaxone, or ampicillin-sulbactam) should be given in combination with either azithromycin or a respiratory fluoroquinolone. When *Pseudomonas* is a consideration (e.g., in patients who have predominantly gram-negative rods on sputum examination; patients with COPD, bronchiectasis, or other structural lung disease; or patients who have been treated with glucocorticoids or other immunosuppressive drugs), an antipseudomonal β-lactam or carbapenem (piperacillin-tazobactam, cefepime, imipenem, or meropenem) should be selected. Whether adding a second antipseudomonal drug, such as a quinolone or aminoglycoside, is helpful is less clear. A more important principle is that treatment of gram-negative pneumonia should focus on using the optimal dose of an appropriate antibiotic based on susceptibility testing. Whether low-dose corticosteroid treatment is beneficial in patients with community-acquired pneumonia, especially with more severe disease, is uncertain.[12]

Hospital Course

The value of aggressive attempts to determine the cause of pneumonia becomes obvious during a patient's hospital course. The expected response to therapy includes defervescence, return of the WBC count to normal, and disappearance of the systemic signs of acute infection within a few days after antibiotics have been begun. Cough may persist for weeks, and fatigue may persist for months, especially in elderly persons.

Patients may not respond or may even deteriorate during the first day or two despite correct antibiotic therapy, and physicians need diagnostic information to avoid the temptation to simply add antibiotics that may provide no further benefit and may have deleterious side effects. For patients who do respond, identification of the causative organism allows broad-spectrum antibiotics to be replaced by a simpler regimen, thereby potentially shortening the hospitalization, reducing the risk for complications such as *Clostridium difficile* colitis (Chapter 296), and avoiding uncertainty about the culprit medication should an adverse drug reaction occur.

Failure to Respond to Antibiotic Therapy

The failure to respond to antibiotic therapy raises a number of concerns (Table 97-7).[13] The initial diagnosis of community-acquired pneumonia may be incorrect, and one of the many other causes of a pulmonary infiltrate, cough, and fever (e.g., a *Mycobacterium* or a fungus) may be responsible. Alternatively, an entirely different class of disease such as lung cancer or an inflammatory lung condition may be present. If the patient simply does not improve, the antibiotic therapy may not be appropriate for the infecting organism. The first step should be to review culture results and antibiotic susceptibilities to ensure that the patient has received an adequate dose of an appropriate antibiotic. The antimicrobial therapy may have been correct, and the causative organism may be susceptible, but there may be a loculated infection such as empyema (Chapter 99), especially in patients who initially improve but then have persistent low-grade fever and leukocytosis.

When antibiotics have been given empirically and cultures have not been obtained, the failure to respond promptly creates a difficult therapeutic dilemma. In patients with a partial or inadequate response to initial therapy, additional diagnostic measures—such as additional cultures, a chest CT scan, a thoracentesis if an effusion is present, bronchoscopy with bronchoalveolar lavage, and possibly transbronchial biopsy—should be aggressively considered rather than simply adding antibiotics.

TABLE 97-7 REASONS FOR FAILURE OF ANTIMICROBIAL THERAPY IN TREATING PNEUMONIA

Correct organism, inappropriate antibiotic choice or dose
Organism not susceptible
Wrong dosage (e.g., morbidly obese or fluid-overloaded patient)
Antibiotic not given
Correct organism and antibiotic, but infection is loculated
Empyema (the most common)
Obstruction (e.g., lung cancer, foreign body)
Unidentified causative organism responsible
Noninfectious cause
Malignancy
Inflammatory infiltrate

Duration of Treatment

The optimal duration of therapy is uncertain. In general, outpatients with community-acquired pneumonia should be treated for 5 to 7 days. Patients who are hospitalized should receive parenteral therapy until they are hemodynamically stable and able to ingest and absorb oral antibiotics, after which oral antibiotics can be given. Three to 5 days of parenteral therapy and a final few days of oral treatment after the patient has become afebrile (temperature <99°F) may be the best approach for pneumococcal pneumonia. Treatment for community-acquired pneumonia of undetermined etiology should generally not exceed a total of 7 to 8 days,[■] and 3 days of treatment is as effective as 8 days for mild to moderately severe pneumonia.[■] In contrast, pneumonia due to *S. aureus* (Chapter 288) or gram-negative bacilli (Chapters 304, 305, and 306) probably requires 10 to 14 days of treatment, whereas bacteremic *S. aureus* pneumonia requires 4 weeks of treatment because of concerns regarding endocarditis, either as a cause or as a result of the pneumonia. For documented *Legionella* pneumonia (Chapter 314), the recommendation is 5 to 10 days of treatment with azithromycin, 14 days with a fluoroquinolone, or 3 weeks with either regimen if the patient is immunocompromised.

Patients may be discharged when they are clinically stable, have no other medical problems requiring continued hospitalization, and have a suitable discharge environment. Reported markers of clinical stability include temperature of 37.8°C or lower, heart rate of 100 beats per minute or lower, respiratory rate of 24 breaths per minute or lower, systolic blood pressure of 90 mm Hg or greater, oxygen saturation of 90% or higher, or Po_2 of 60 mm Hg or higher on room air (for patients not previously dependent on supplemental oxygen), and mental status at baseline. More rigorous criteria are for temperature to be less than 99°F, respiratory rate to be normal or nearly normal, and oxygen saturation and blood pressure to be back to baseline. Patients are commonly observed in an inpatient setting for up to 24 hours after switching from intravenous to oral therapy, but there is no evidence to support this practice, and it is not necessary for patients who are otherwise stable.

Infectious Complications

Empyema (Chapter 99), which is the most common infectious complication of pneumonia, should be considered in patients who have persisting fever and leukocytosis after 4 to 5 days of appropriate antibiotic therapy for pneumonia. A repeat chest radiograph and a CT scan are important diagnostic tools. Other extrapulmonary infections occur when bacteria are carried by the blood stream to bones or joints (especially intervertebral spaces), peritoneal cavity (if peritoneal fluid was present when bacteremia occurred), meninges, heart valves, or even large muscle groups. Such infections will usually declare themselves by causing symptoms, to which the physician must remain attuned in patients whose recovery is not a rapid as was expected.

Noninfectious Complications

Myocardial infarction (Chapter 73) and new arrhythmias, especially atrial fibrillation (Chapter 64), are seen in 7 to 10% of patients admitted for community-acquired pneumonia, and worsening of heart failure is even more frequent. These cardiac events are associated with substantial increases in morbidity and mortality.

PREVENTION

Maintaining good general health, smoking cessation, avoidance of excessive alcohol ingestion, and control of blood sugar in diabetic patients are good general measures to reduce the risk for bacterial pneumonia. Vaccination against influenza (Chapters 18 and 364) reduces the risk not only of influenza but also of all causes of pneumonia because influenza infection predisposes to secondary bacterial pulmonary infection. Two currently available pneumococcal vaccines[14] specifically reduce the risk for pneumococcal pneumonia (Chapter 18): a polysaccharide vaccine that contains capsular polysaccharide from 23 different pneumococcal serotypes (marketed in the United States as Pneumovax 23 or Pnu-Imune), and a pneumococcal conjugate vaccine, in which capsular polysaccharides are conjugated to an immunogenic protein (originally marketed in the United States as Prevnar7 and now as Prevnar13. Unlike polysaccharide vaccine, conjugate vaccine also eliminates detectable nasopharyngeal carriage of pneumococci and, as a result, the spread of pneumococci to unvaccinated individuals. Pneumococcal infection caused by strains contained in the 7-valent vaccine has fallen in children by 95% and in adults by 85% since widespread vaccination of

children has taken place. This same decline has already begun to be observed for strains contained in the 13-valent vaccine.

Pneumococcal polysaccharide vaccine is recommended for all persons 65 years of age or older and for all persons 19 to 64 years of age who are immunocompromised (e.g., congenital or acquired immunodeficiency, HIV infection, chronic renal failure, nephrotic syndrome, hematologic malignancies, iatrogenic immunosuppression, generalized malignancy, and organ transplantation) or have conditions that predispose to pneumococcal infection or that put an individual at particularly high risk for complications (e.g., asplenia, cigarette smoking, asthma, chronic lung disease, heart failure, diabetes, alcoholism, chronic liver disease, cerebrospinal fluid leak, and cochlear implant).

Patients who received pneumococcal polysaccharide vaccine before age 65 years should receive another dose at or after age 65 years, provided at least 5 years have passed since the last dose. Multiple revaccinations after age 65 years are not recommended.

Pneumococcal conjugate vaccine is recommended for all immunocompromised adults 19 years of age and older (see earlier), and individuals who have had a splenectomy or who have cerebrospinal fluid leak or cochlear implant should receive a single lifetime dose of conjugate vaccine. If they have not previously received any pneumococcal vaccine, they should receive conjugate vaccine, followed at least 8 weeks later by the 23-valent polysaccharide vaccine. There is no recommendation to administer more than one dose of conjugate vaccine to an adult.

PROGNOSIS

In developed countries, the mortality among outpatients with pneumonia is less than 2%, but it exceeds 10% in patients who are hospitalized and approaches 40% in patients who require admission to an ICU. In patients with pneumococcal pneumonia, a WBC count of less than 6000/μL is associated with greater than 65% mortality. In community-acquired pneumonia, a serum sodium level of less than 130 mEq/L, a newly elevated serum creatinine level, or a serum glucose level of greater than 250 mg/dL in a nondiabetic patient is also associated with a poor prognosis.

Patients often recuperate only slowly, with residual fatigue and weakness persisting for months. After recovery from bacterial pneumonia requiring hospitalization, mortality is substantially increased at 1 year and, at least in the case of pneumococcal pneumonia, is still significantly increased 3 to 5 years later,[15] presumably because the pneumonia has served as a marker for comorbid conditions that limit lifespan. There is a good correlation between the severity of the pneumonia and the later risk for death.

ASPIRATION PNEUMONIA

EPIDEMIOLOGY AND PATHOBIOLOGY

Although microaspiration underlies most cases of pneumonia, some patients experience repeated gross aspiration of oropharyngeal contents. In such patients, predisposing social factors include alcoholism, cigarette smoking, poor dental hygiene, and homelessness. Predisposing medical conditions include acute or chronic mental status changes, neuromuscular disease, esophageal obstruction, and severe esophageal reflux (Chapter 138). Aspiration pneumonia may not be infectious if it results from damage produced by gastric acid or as a response to gastric contents other than bacteria. However, in practice, a distinction between infectious and noninfectious aspiration pneumonia cannot be made, and aspiration initially should be treated as if it is due to infection.

CLINICAL MANIFESTATIONS

Patients with aspiration pneumonia can present acutely because of the aspiration of food or the acute irritant effects of gastric acid. In most cases, however, clinical deterioration, decreased oxygenation, fever, dyspnea, purulent sputum, and leukocytosis evolve over a number of days.[16] In cases of lung abscess, patients state that their sputum has a foul taste and odor. Physical examination generally reveals malnutrition, findings of chronic comorbid conditions, poor dentition, signs of chronic lung disease, and coarse rhonchi in the lower lobes or dependent lung regions.

DIAGNOSIS

On chest radiography, aspiration pneumonia is most commonly seen as a parenchymal bronchopneumonia process in the superior segments of the right lower lobe and the posterior segments of the upper lobes,[17] but aspiration can involve any part of the lung depending on the patient's position during aspiration. The finding of a thick-walled abscess (Chapter 90) with a fluid level provides strong confirmatory evidence.

Microbiology

Because oropharyngeal secretions contain massive numbers of aerobic and anaerobic organisms, aspiration pneumonia is usually a polymicrobial infection. The diagnosis of aspiration pneumonia can be made by examination of gram-stained sputum that shows many WBCs, few or no epithelial cells, and profuse numbers of mixed bacteria. Because the usual pathogens are mixed normal flora, cultures of sputum are not often helpful, although they might demonstrate other pathogenic bacteria such as *S. aureus* or a multidrug-resistant gram-negative rod that requires therapy. Cultures of resected lung abscess or of transtracheal aspirates in patients with lung abscess typically yield expected mouth organisms, including microaerophilic streptococci and staphylococci, *Bacteroides* species, *Fusobacteria,* and *Prevotella* species. *S. pneumoniae, S. aureus,* and *H. influenzae* may also be present. Because the oropharynx of hospitalized patients and residents of long-term care facilities is regularly colonized by facultative gram-negative bacteria, *P. aeruginosa*, and *S. aureus*, these organisms are also likely to be implicated in aspiration pneumonia.

TREATMENT

Rx

Patients who are admitted from the community with aspiration pneumonia or lung abscess should be treated initially with parenteral ampicillin-sulbactam (1.5 to 3 g intravenously every 6 hours) or clindamycin (600 mg intravenously every 8 hours) for at least 5 days. Results of cultures might suggest that other antibiotics be used as well, but the prominence of microaerophilic and anaerobic organisms that will not be identified by routine cultures mandates that one of these drugs be continued. When the patient is stable, treatment can be switched to oral therapy (e.g., clindamycin 600 mg three times daily or ampicillin-sulbactam 750 mg three times daily). Aspiration pneumonia is treated for 7 to 10 days unless cavitation is present, in which case treatment is continued for several weeks or even until the cavity is no longer detectable. In elderly, bedridden patients, especially patients in nursing units or hospitals, intravenous piperacillin-tazobactam (3.375 g intravenously every 6 hours), meropenem (1 g intravenously every 8 hours), or imipenem (1 g intravenously every 6 to 8 hours) for at least 5 days is probably more appropriate initial therapy because of the likelihood of gram-negative bacteria, especially multidrug-resistant organisms that may produce extended-spectrum β-lactamases or a carbapenemase. If MRSA is suspected or is documented by culture, appropriate therapy needs to be added for it as well (Chapter 288).

The most common, serious complication of anaerobic pneumonia or a lung abscess is the development of an empyema. Insertion of one or more chest tubes may control the disease, but thoracotomy with pleural stripping may be the only way to remove the infected material.[18] Unfortunately, patients who develop this complication are often unable to undergo such an aggressive procedure, thereby creating a major therapeutic dilemma. In patients with underlying neurologic disease or malignancy, a gastrostomy or jejunostomy feeding tube can be inserted to provide palliative nutrition, fluids, and medications.

PROGNOSIS

Unless an empyema has developed, the prognosis for aspiration pneumonia or lung abscess is largely determined by the comorbid conditions that led to its occurrence rather than a failure of the pneumonia or abscess to respond.

HOSPITAL-ACQUIRED PNEUMONIA, VENTILATOR-ASSOCIATED PNEUMONIA, AND HEALTH CARE–ASSOCIATED PNEUMONIA

EPIDEMIOLOGY

Hospital-acquired pneumonia, ventilator-associated pneumonia, and health care–associated pneumonia represent the second most common nosocomial infections (Chapter 282) in the United States. Hospital-acquired pneumonia, which increases costs and the length of hospital stay, is responsible for up to 25% of all ICU infections. Hospital-acquired pneumonia and ventilator-associated pneumonia occurring within the first 4 hospital days tend to be caused by antibiotic-susceptible bacteria, whereas late-onset infections are more frequently caused by multidrug-resistant organisms (Table 97-8).

TABLE 97-8 EMPIRICAL ANTIBIOTIC TREATMENT OF HOSPITAL-ACQUIRED PNEUMONIA, VENTILATOR-ASSOCIATED PNEUMONIA, AND HEALTH CARE–ASSOCIATED PNEUMONIA

GROUP A: PATIENTS WITH EITHER HOSPITAL-ACQUIRED PNEUMONIA OR VENTILATOR-ASSOCIATED PNEUMONIA, WITHOUT RISK FACTORS FOR MULTIDRUG-RESISTANT PATHOGENS, AND WITH EARLY-ONSET PNEUMONIA	
POTENTIAL PATHOGENS	**RECOMMENDED THERAPY**
Streptococcus pneumoniae *Haemophilus influenzae* Methicillin-sensitive *Staphylococcus aureus* Antibiotic-sensitive enteric gram-negative bacilli *Escherichia coli* *Klebsiella pneumoniae* *Enterobacter* species *Proteus* species *Serratia marcescens*	Ceftriaxone, 1-2 g IV/IM every 12-24 hr, maximum of 4 g/day, with duration dependent on clinical response and individualized, as discussed in text *or* Levofloxacin, 500-750 mg IV every day, with duration dependent on clinical response and individualized; or ciprofloxacin, 400 mg IV every 8 hr, with duration dependent on clinical response and individualized; or moxifloxacin, 400 mg IV or orally every 24 hr, with duration dependent on clinical response and individualized *or* Ampicillin-sulbactam, 1.5-3 g (1-2 g ampicillin and 0.5-1 g sulbactam) IV/IM every 6 hr, maximum of 4 g sulbactam/day, depending on type and severity of infection, with duration dependent on clinical response and individualized *or* Ertapenem, 1 g IV/IM once a day, with duration dependent on clinical response and individualized
GROUP B: PATIENTS WITH HOSPITAL-ACQUIRED PNEUMONIA, VENTILATOR-ASSOCIATED PNEUMONIA, OR HEALTH CARE–ASSOCIATED PNEUMONIA AND WITH LATE-ONSET PNEUMONIA OR WITH RISK FACTORS FOR MULTIDRUG-RESISTANT PATHOGENS	
ORGANISMS	**THERAPY**
S. pneumoniae *H. influenzae* Methicillin-sensitive *S. aureus* Antibiotic-sensitive enteric gram-negative bacilli *E. coli* *K. pneumoniae* *Enterobacter* species *Proteus* species *S. marcescens* Multidrug-resistant pathogens *Pseudomonas aeruginosa* *K. pneumoniae* (extended spectrum β-lactamase producing) *Acinetobacter* species Methicillin-resistant *S. aureus* Legionella pneumophila	Antipseudomonal cephalosporin (ceftazidime, 2 g IV every 8 hr, or cefepime, 1-2 g every 8-12 hr, with duration dependent on clinical response and individualized) *or* Antipseudomonal carbapenems (meropenem, 1 g every 8 hr, or imipenem, 500 mg every 6 hr or 1 g every 8 hr, with duration dependent on clinical response and individualized) *or* β-Lactam/β-lactamase inhibitor (piperacillin-tazobactam, 4.5 g IV every 6 hr, with duration dependent on clinical response and individualized) *plus* Antipseudomonal fluoroquinolone (levofloxacin, 750 mg IV every day, or ciprofloxacin, 400 mg IV every 8 hr, with duration dependent on clinical response and individualized) *or* Aminoglycoside (amikacin, 15-20 mg/kg/day, divided every 8-12 hr, with monitoring to maintain trough lower than 4-5 μg/mL; or gentamicin, 7 mg/kg/day as a single daily dose, with monitoring to maintain trough levels lower than 1 μg/mL; or tobramycin, 4-7 mg/kg/day as a single daily dose, with monitoring to maintain trough levels lower than 1 μg/mL and duration dependent on clinical response and individualized) *plus* Vancomycin (15 mg/kg IV every 12 hr, with monitoring to maintain trough at 10-15 μg/mL and duration dependent on clinical response and individualized) or linezolid (600 mg IV every 12 hr, with duration dependent on clinical response and individualized)

Data from American Thoracic Society. Guidelines for management of adults with hospital-acquired ventilator associated, and healthcare-associated pneumonia. *Am J Respir Crit Care Med.* 2005:171:388-416.

PATHOBIOLOGY

When pneumonia occurs in the first few days after hospitalization, including the 3 to 4 days after an elective surgical procedure, the most likely organisms include microaerophilic and anaerobic bacteria of the mouth and bacteria that normally cause community-acquired pneumonia, such as *S. pneumoniae* and *H. influenzae.* Thereafter, *S. aureus* and facultative gram-negative bacilli, (e.g., *P. aeruginosa, E. coli, Klebsiella pneumoniae,* and *Acinetobacter* species) become increasingly more common. Many cases are polymicrobial and include gram-positive agents such as *S. aureus,* particularly MRSA strains, especially in patients with severe underlying chronic disease. *P. aeruginosa, Acinetobacter* species, *Stenotrophomonas maltophilia,* and *Burkholderia cepacia* complex rapidly become resistant to multiple classes of antibiotics, so routine local surveillance and monitoring are critical to help predict drug susceptibilities.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Hospital-acquired pneumonia, ventilator-associated pneumonia,[19] and health care–associated pneumonia can present with typical signs, such as fever, leukocytosis, and purulent sputum or as increased tracheal secretions in an intubated patient. Chest radiographs are often difficult to interpret, but they may show new or worsening pulmonary infiltrates. Oxygen levels fall, and acute respiratory distress (Chapter 104) may result.

If microscopic examination of a gram-stained sputum or tracheal secretions does not show many inflammatory cells with a single organism, cultures of specimens obtained by bronchoscopy, using a protected brush with quantitative analysis, may be needed to determine the causative organism. Sterile culture of lower respiratory tract secretions in the absence of a new antibiotic in the past 72 hours essentially excludes most bacterial pneumonias, although *Legionella* and viral infection are still possible in this situation.

TREATMENT

If the patient is unstable, or if there is a high suspicion for hospital-acquired pneumonia, ventilator-associated pneumonia, or health care–associated pneumonia, prompt empirical antibiotic therapy (see Table 97-8) is required because delays in antimicrobial therapy increase mortality. For early-onset disease in the first 4 days of hospitalization, options include either ceftriaxone, an intravenous fluoroquinolone (levofloxacin, moxifloxacin, or ciprofloxacin), ampicillin-sulbactam, or ertapenem. These options will cover *S. pneumoniae, H. influenzae,* MSSA, and most antibiotic-sensitive gram-negative bacilli, including *E. coli, K. pneumoniae, Proteus* species, *Enterobacter* species, and *Serratia marcescens.*

With late-onset hospital-acquired pneumonia, ventilator-associated pneumonia, or health care–associated pneumonia, or when risk factors for multidrug-resistant infection have been identified, the initial choice of antibiotics should be guided by knowledge of the antibiotic susceptibility of commonly isolated organisms in the facility in question. Because such prediction can be very difficult, multiagent regimens are recommended, but this recommendation emphasizes the importance of obtaining good culture specimens so that therapy can later be narrowed. Options include an antipseudomonal cephalosporin such as ceftazidime or cefepime, an antipseudomonal carbapenem (meropenem or imipenem), or a β-lactam/β-lactamase inhibitor agent

such as piperacillin-tazobactam; in addition, either an antipseudomonal fluoroquinolone (ciprofloxacin) or an aminoglycoside such as amikacin, gentamicin, or tobramycin should be considered. If *Legionella* is strongly suspected, the fluoroquinolone should suffice; otherwise a macrolide such as azithromycin should be added. Finally, either vancomycin or linezolid should be added for coverage of MRSA unless the presence of this organism can be excluded.

If the patient improves over the first 48 to 72 hours, strong consideration should be given to de-escalating antibiotic therapy based on culture results. If lower respiratory cultures remain negative but the patient has not improved, an extrapulmonary site of infection should be considered, and additional radiographic studies or cultures may be helpful. In general, aminoglycosides should be limited to 5 to 7 days; overall antibiotic therapy can be as short as 7 days if the patient has improved, but some patients require 14 to 21 days of therapy.

PREVENTION

Prevention centers first on staff education and compliance with alcohol-based hand disinfection. Patients with documented multidrug-resistant organisms should be isolated or, if isolation is not possible, cohorted, in order to reduce the risk for patient cross-contamination. Ventilator-associated pneumonia can be reduced by elevation of the head of the patient's bed, regular aspiration of subglottic secretions, daily "sedation vacations," and daily assessment of the patient's readiness for extubation.

PROGNOSIS

The overall mortality attributed to hospital-acquired pneumonia may be as high as 30 to 50%. Mortality is due, in part, to the severity of the pneumonia and the difficulty of providing adequate antibiotic coverage for some gram-negative bacilli but also to the underlying health of the patient.

Grade A References

A1. Schuetz P, Christ-Crain M, Thomann R, et al. Effect of procalcitonin-based guidelines vs standard guidelines on antibiotic use in lower respiratory tract infections: the ProHOSP randomized controlled trial. *JAMA*. 2009;302:1059-1066.

A2. Schuetz P, Muller B, Christ-Crain M, et al. Procalcitonin to initiate or discontinue antibiotics in acute respiratory tract infections. *Cochrane Database Syst Rev*. 2012;9:CD007498.

A3. Zhanel GG, Wolter KD, Calciu C, et al. Clinical cure rates in subjects treated with azithromycin for community-acquired respiratory tract infections caused by azithromycin-susceptible or azithromycin-resistant *Streptococcus pneumoniae*: analysis of Phase 3 clinical trial data. *J Antimicrob Chemother*. 2014;69:2835-2840.

A4. Li JZ, Winston LG, Moore DH, et al. Efficacy of short-course antibiotic regimens for community-acquired pneumonia: a meta-analysis. *Am J Med*. 2007;120:783-790.

A5. el Moussaoui R, de Borgie CA, van den Broek P, et al. Effectiveness of discontinuing antibiotic treatment after three days versus eight days in mild to moderate-severe community acquired pneumonia: randomised, double blind study. *BMJ*. 2006;332:1355.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Musher DM, Thorner AT. Community-acquired pneumonia. *N Engl J Med*. 2014;371:1619-1628.
2. Self WH, Courtney DM, McNaughton CD, et al. High discordance of chest x-ray and computed tomography for detection of pulmonary opacities in ED patients: implications for diagnosing pneumonia. *Am J Emerg Med*. 2013;31:401-405.
3. Albrich WC, Madhi SA, Adrian PV, et al. Use of a rapid test of pneumococcal colonization density to diagnose pneumococcal pneumonia. *Clin Infect Dis*. 2012;54:601-609.
4. Musher DM, Roig IL, Cazares G, et al. Can an etiologic agent be identified in adults who are hospitalized for community-acquired pneumonia? Results of a one-year study. *J Infect*. 2013;67:118-119.
5. Wiemken T, Peyrani P, Bryant K, et al. Incidence of respiratory viruses in patients with community-acquired pneumonia admitted to the intensive care unit: results from the Severe Influenza Pneumonia Surveillance (SIPS) project. *Eur J Clin Microbiol Infect Dis*. 2013;32:705-710.
6. Falsey AR, Becker KL, Swinburne AJ, et al. Bacterial complications of respiratory tract viral illness: a comprehensive evaluation. *J Infect Dis*. 2013;208:432-441.
7. Charles PG, Wolfe R, Whitby M, et al. SMART-COP: a tool for predicting the need for intensive respiratory or vasopressor support in community-acquired pneumonia. *Clin Infect Dis*. 2008;47:375-384.
8. Mandell LA, Wunderink RG, Anzueto A, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis*. 2007;44:S27-S72.
9. Spindler C, Stralin K, Eriksson L, et al. Swedish guidelines on the management of community-acquired pneumonia in immunocompetent adults: Swedish Society of Infectious Diseases 2012. *Scand J Infect Dis*. 2012;44:885-902.
10. Nussenblatt V, Avdic E, Cosgrove S. What is the role of antimicrobial stewardship in improving outcomes of patients with CAP? *Infect Dis Clin North Am*. 2013;27:211-228.
11. Minejima E, Lou M, Nieberg P, et al. Patients presenting to the hospital with MRSA pneumonia: differentiating characteristics and outcomes with empiric treatment. *BMC Infect Dis*. 2014;14:252.
12. Shafiq M, Mansoor MS, Khan AA, et al. Adjuvant steroid therapy in community-acquired pneumonia: a systematic review and meta-analysis. *J Hosp Med*. 2013;8:68-75.
13. Sialer S, Liapikou A, Torres A. What is the best approach to the nonresponding patient with community-acquired pneumonia? *Infect Dis Clin North Am*. 2013;27:189-203.
14. Tomczyk S, Bennett NM, Stoecker C, et al. Use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine among adults aged >/=65 years: recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb Mortal Wkly Rep*. 2014;63:822-825.
15. Sandvall B, Rueda AM, Musher DM. Long-term survival following pneumococcal pneumonia. *Clin Infect Dis*. 2013;56:1145-1146.
16. Lanspa MJ, Jones BE, Brown SM, et al. Mortality, morbidity, and disease severity of patients with aspiration pneumonia. *J Hosp Med*. 2013;8:83-90.
17. Prather AD, Smith TR, Poletto DM, et al. Aspiration-related lung diseases. *J Thorac Imaging*. 2014;29:304-309.
18. Schweigert M, Solymosi N, Dubecz A, et al. Surgery for parapneumonic pleural empyema: what influence does the rising prevalence of multimorbidity and advanced age has on the current outcome? *Surgeon*. 2014 [Epub ahead of print].
19. Barbier F, Andremont A, Wolff M, et al. Hospital-acquired pneumonia and ventilator-associated pneumonia: recent advances in epidemiology and management. *Curr Opin Pulm Med*. 2013;19:216-228.

REVIEW QUESTIONS

1. For which of the following groups is pneumococcal polysaccharide vaccine not routinely recommended?

A. Otherwise healthy adults aged 50 to 65 years
B. Adults aged 19 to 65 years who have immunocompromising conditions
C. Cigarette smokers
D. Adults with asthma
E. Alcoholics, especially if liver disease is present

Answer: A The incidence of pneumococcal pneumonia is highest in persons younger than 2 years of age or older than 65 years of age. Thus, pneumococcal polysaccharide vaccine is not recommended for otherwise healthy adults aged 50 to 65 years. This vaccine is, however, recommended for persons who have immunocompromising conditions or lung disease (such as chronic obstructive pulmonary disease or asthma) and for persons who are cigarette smokers and heavy consumers of alcohol.

2. Which of the following is the least likely to be found as a cause for a syndrome consistent with community-acquired pneumonia?

A. Pulmonary edema
B. Pulmonary fibrosis
C. Lung cancer
D. *Nocardia* infection
E. Influenza virus infection

Answer: D Pulmonary edema, pulmonary fibrosis, and lung cancer, although not infectious in origin, are frequent causes of a new pulmonary infiltrate, cough, and sputum production, sometimes accompanied by fever. In well-documented large case series, these symptoms and signs are only slightly more common in patients with bacterial pneumonia than in noninfectious conditions such as pulmonary edema, pulmonary fibrosis, or lung cancer. Whereas influenza virus is a major cause of hospitalization for pneumonia (often with secondary bacterial infection), *Nocardia* is an infrequent cause.

3. Which of the following contributes to a need for intensive unit care in a patient who is hospitalized for bacterial pneumonia?

A. Confusion
B. Serum creatinine elevated to twice baseline
C. $Pao_2 = 90$ mm Hg
D. White blood cell count < 6000 with 15% band forms
E. All of the above

Answer: E Confusion, a rising serum creatinine, and a falling O_2 saturation indicate severe disease. In a patient with a bacteremia pneumonia, a low white blood cell count with increased band forms is an ominous sign that is associated with a 65% mortality in pneumococcal infections.

4. Which of the following conditions is not associated with an increased incidence or severity of pneumococcal pneumonia?

A. Poorly controlled hypertension
B. Diabetes mellitus
C. Renal insufficiency
D. Cirrhosis of the liver
E. Multiple myeloma

Answer: A In multiple myeloma, the inability to produce normal immunoglobulins to new antigenic stimuli can result in pneumonia. The well-documented increased incidence of pneumonia in patients who have diabetes mellitus, renal insufficiency, and cirrhosis of the liver is multifactorial, with poor leukocyte function (migration, ingestion, and bacterial killing) as well as poor antibody responses contributing. In contrast, there is no relationship between hypertension and pneumonia.

98

PULMONARY EMBOLISM

JEFFREY I. WEITZ

DEFINITIONS

Pulmonary embolism (PE) refers to an obstruction of a pulmonary artery by material that has traveled to the lungs from elsewhere in the body through the blood stream.[1] Thrombus from the deep veins of the legs or arms represents the most common type of material to embolize to the lungs—a process known as venous thromboembolism (VTE).[2] In addition to thrombotic pulmonary emboli, nonthrombotic material also can embolize to the lungs. Such material includes fat, air, amniotic fluid, tumor cells, talc in intravenous drug users, and various medical devices. Regardless of the type of embolic material, blockage of blood flow through the lungs and the resultant increased pressure in the right ventricle are responsible for the symptoms and signs of PE.

THROMBOTIC PULMONARY EMBOLISM

EPIDEMIOLOGY

VTE, which includes deep vein thrombosis (DVT; Chapter 81) and PE, represents the third most common cause of cardiovascular death after myocardial infarction and stroke. A first episode of VTE occurs in about 1 to 2 persons per 1000 each year in the United States. The incidence rises exponentially with age, with 5 cases per 1000 persons per year by the age of 80 years. Although men and women are affected equally, the incidence is higher in whites and African Americans than in Hispanic persons and Asian Pacific Islanders.

Approximately one third of patients with symptomatic VTE present with PE; the remainder present with DVT alone but have clinically silent PE in 10 to 15% of cases.[3] Up to half of the patients with a first episode of VTE have no identifiable risk factors and are described as having unprovoked or idiopathic VTE. The remainder develop VTE secondary to well-recognized, transient risk factors, such as surgery or immobilization. PE accounts for an estimated 15% of deaths in hospitalized patients, with at least 100,000 deaths from PE each year in the United States.

PATHOBIOLOGY

PE and DVT are part of the spectrum of VTE and share the same genetic and acquired risk factors, which determine the intrinsic risk for VTE for each individual (E-Fig. 98-1). Genetic risk factors include abnormalities associated with hypercoagulability of the blood (Chapter 176), the most common of which are factor V Leiden and the prothrombin 20210 gene mutation. Acquired risk factors include advanced age, history of previous VTE, obesity, and active cancer, all of which limit mobility and may be associated with hypercoagulability. Superimposed on this background risk, VTE often occurs in the presence of triggering factors, which increase the risk above the critical threshold. The triggering factors, including surgery and pregnancy or estrogen therapy, lead to endothelial cell activation, stasis, and hypercoagulability, which are the components of Virchow triad.

In at least 90% of patients, PE originates from DVT in the lower limbs, and up to 70% of patients with proven PE still have demonstrable DVT on presentation. Thrombi usually start in the calf veins. About 20% of these calf vein thrombi then extend into the popliteal and more proximal veins of the leg, from which they are more likely to embolize. Although often asymptomatic, PE can be detected in about 50% of patients with proximal DVT (Chapter 81). Upper extremity DVT involving the axillary or subclavian veins also can give rise to PE, but only 10 to 15% of such patients develop PE. Upper extremity DVT most often occurs in patients with cancer (Chapter 179), particularly those with indwelling central venous catheters. Unprovoked upper extremity DVT, usually involving the dominant arm, can occur with strenuous effort—the so-called Paget-Schroetter syndrome.

PE often involves both lungs, with the lower lobes affected more frequently than the upper. Larger emboli tend to lodge in the main pulmonary artery or its branches, whereas smaller emboli occlude more peripheral arteries. Peripheral PE can lead to pulmonary infarction, which is characterized by intra-alveolar hemorrhage and necrosis that may be pleura based. Because the circulation to the lungs arises from bronchial as well as pulmonary arteries, pulmonary infarction occurs in only about 10% of patients without underlying cardiopulmonary disease. In contrast, pulmonary infarction occurs in up to 30% of patients who have compromised oxygenation of the affected areas of lung because of preexisting disorders, such as airways disease or increased pulmonary venous pressure because of left ventricular dysfunction.

The clinical impact of PE depends on the extent of reduction in pulmonary blood flow, the time frame over which vascular obstruction occurs, and the absence or presence of underlying cardiopulmonary disease. With acute PE, most patients develop tachypnea and some degree of hypoxemia. Stimulation of irritant receptors in the lungs likely accounts for the increase in respiratory rate. Obstruction of pulmonary arteries contributes to the hypoxemia and the increase in alveolar-arterial oxygen tension gradient, which reflects inefficient oxygen transfer across the lungs. These abnormalities result mainly from the increase in alveolar dead space that occurs because ventilation to alveoli exceeds blood flow in parts of the lung affected by PE. Other contributing factors to the hypoxemia include ventilation-perfusion mismatch because of relative overperfusion of normal areas of the lung, and shunting of blood through nonventilated atelectatic or collapsed areas of lung that retain at least some perfusion.

Pulmonary vascular resistance increases with PE because of vascular occlusion by thrombi. In addition, humoral mediators, such as serotonin and

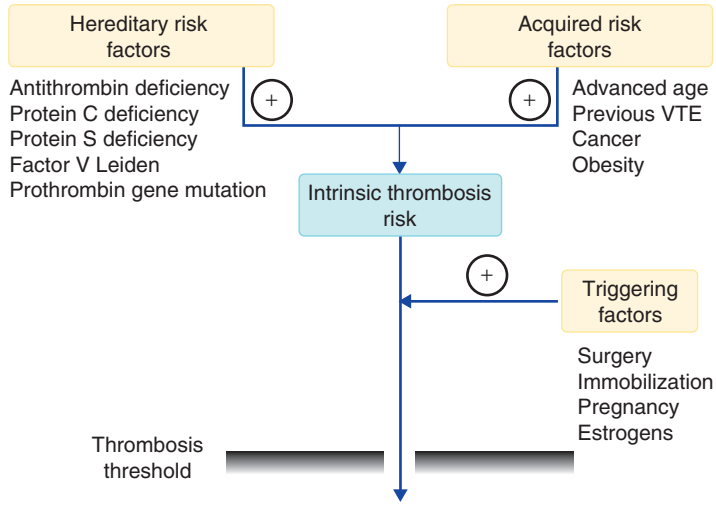

E-FIGURE 98-1. Thrombosis threshold. Hereditary and acquired risk factors combine to create an intrinsic risk of thrombosis for each individual. This risk is increased by extrinsic triggering factors. If the intrinsic and extrinsic forces exceed a critical threshold where thrombin generation overwhelms protective mechanisms, thrombosis occurs. VTE = venous thromboembolism.

thromboxane, are released from activated platelets and may trigger vasoconstriction in unaffected areas of lung. Consequently, the increase in pulmonary vascular resistance may be disproportionate to the extent of pulmonary vascular occlusion. With obstruction of less than 50% of the pulmonary vascular bed, the mean pulmonary artery pressure rarely exceeds 25 mm Hg. Under these circumstances, the right ventricle maintains its output, so cardiac output and systemic blood pressure remain normal.

With acute occlusion of more than 50% of the pulmonary circulation, the pulmonary artery systolic pressure increases, thereby increasing right ventricular afterload. The pulmonary artery systolic pressure rarely exceeds 55 mm Hg with sudden occlusion because of insufficient time for right ventricular hypertrophy to occur. If the thin-walled right ventricle fails to maintain output in the face of the increased pulmonary artery pressure, it dilates, and right heart failure ensues. Right ventricular end-diastolic and right atrial pressures increase as the right ventricle fails. Dilation of the right ventricle may result in tricuspid regurgitation, which can compromise left ventricular filling and lead to reduced cardiac output and subsequent hypotension. Rightward bulging of the interventricular septum may also contribute to left ventricular diastolic dysfunction. The decrease in aortic pressure, together with the increase in right ventricular pressure, can produce right ventricular ischemia because of decreased perfusion of the right coronary artery despite increased demand by the dilated right ventricle. If this process occurs over a rapid time frame (i.e., minutes to hours), syncope or sudden death, often associated with electromechanical dissociation, may be the first manifestation of severe PE.

With multiple pulmonary emboli over an extended period of time, the right ventricle has an opportunity to adapt to the increased pulmonary vascular resistance. The subsequent increase in right ventricular systolic pressure results in less right heart failure than occurs with acute large PE. Patients with multiple smaller PEs over an extended period of time often have increasing dyspnea with progressively decreasing exercise tolerance. With maintained cardiac output, hypotension does not develop. Additional emboli, however, may convert the clinical picture to one of severe acute PE.

Patients who have underlying cardiopulmonary disease or who are elderly, frail, and debilitated will be more sensitive to the effects of PE than patients who were previously healthy. Consequently, even a small PE may be fatal in patients with limited reserve.

CLINICAL MANIFESTATIONS

Patients with PE most often present with a history of dyspnea, which may be sudden in onset and tends to progress in severity over time. Dyspnea may be associated with pleuritic chest pain, cough, and hemoptysis, particularly in patients with pulmonary infarction. Although the symptoms and signs of PE can be nonspecific, the diagnosis should be suspected in patients with risk factors for VTE, such as prolonged immobility, recent surgery, or active malignancy. Patients with associated DVT (Chapter 81) may present with recent onset of leg pain or with swelling and tenderness along the course of the deep veins. The superficial veins of the leg may be dilated, and the affected leg may be warm to touch with skin that is red or dusky blue in color.

Most patients with PE have tachypnea and tachycardia associated with hypoxemia, but these findings also can occur with disorders such as heart failure, pneumonia, or chronic obstructive pulmonary disease. Other nonspecific symptoms include palpitations, anxiety, and lightheadedness.

Patients with acute severe PE often complain of dyspnea at rest or with minimal exertion, and they may present with syncope (Chapters 51 and 62) because of hypoxemia and low cardiac output. The combination of hypotension, hypoxemia, and increased cardiac workload may trigger angina (Chapter 71) or overt myocardial infarction (Chapter 73).

Central and peripheral cyanosis can occur, and a gallop rhythm may develop as a consequence of heart failure. The jugular veins may be distended if right heart failure develops. The second heart sound can be widely split and the pulmonic component may be loud because of delayed emptying of the right ventricle. A right ventricular heave may be present with massive PE and acute pulmonary hypertension.

DIAGNOSIS

Most patients with PE will have one or more of the following clinical features: dyspnea, often of sudden onset; tachypnea, with a respiratory rate of more than 20 breaths per minute; and chest discomfort, which is usually substernal and often pleuritic in nature. When patients present with these features, the differential diagnosis includes pulmonary disorders, such as pneumonia (Chapter 97), an exacerbation of chronic obstructive lung disease (Chapter 88), or asthma (Chapter 87); pleurisy secondary to connective tissue disease (Chapter 99); cardiac disorders, such as heart failure (Chapter 58), acute coronary syndrome (Chapter 72), or pericarditis (Chapter 77); and musculoskeletal disorders, such as rib fracture.

TABLE 98-1 WELLS' CLINICAL PREDICTION RULE FOR LIKELIHOOD OF PULMONARY EMBOLISM

VARIABLE	POINTS
PREDISPOSING FACTORS	
Previous VTE	1.5
Recent surgery or immobilization	1.5
Cancer	1
SYMPTOMS	
Hemoptysis	1
SIGNS	
Heart rate > 100 beats/min	1.5
Clinical signs of DVT	3
CLINICAL JUDGMENT	
Alternative diagnosis less likely than PE	3
CLINICAL PROBABILITY	**TOTAL POINTS**
Low	<2
Moderate	2-6
High	>6

DVT = deep vein thrombosis; PE = pulmonary embolism; VTE = venous thromboembolism.
Adapted from Wells PS, Ginsberg JS, Anderson DR, et al. Use of a clinical model for safe management of patients with suspected pulmonary embolism. *Ann Intern Med.* 1998;129:997-1005.

Because the clinical features are nonspecific, the diagnosis of PE requires objective testing. Patients who require such testing can be identified by their pretest likelihood of PE using validated clinical prediction rules (Table 98-1) that include components of the clinical assessment, presence of risk factors for VTE, and absence of an alternative diagnosis to explain the symptoms and signs. Some clinical prediction rules also include the results of simple tests, such as the electrocardiogram (ECG) and the chest radiograph.

Based on the results of such an assessment, the pretest likelihood of PE can be designated as low, moderate, or high, and this likelihood then guides the subsequent selection of blood tests, such as the D-dimer assay, and noninvasive or invasive tests for diagnosis of PE or DVT (Fig. 98-1) (Chapter 81). Tests for diagnosis of DVT are relevant because a diagnosis of DVT in a patient with suspected PE provides sufficient grounds for initiation of treatment, and the treatment of DVT and PE is usually the same. Noninvasive tests include computed tomography (CT) pulmonary angiography or ventilation-perfusion lung scanning for diagnosis of PE and venous compression ultrasound for diagnosis of DVT. These tests have largely replaced pulmonary angiography to diagnose PE and venography to diagnose DVT.

Diagnostic Tests

D-Dimer

A plasmin-derived degradation product of cross-linked fibrin, D-dimer can be measured in whole blood or plasma to provide an indirect index of ongoing activation of the coagulation system. An elevated D-dimer level has an 85% to 98% sensitivity for the diagnosis of PE, but all available D-dimer assays have low specificities.[4] False-positive D-dimer elevations can occur with advanced age, chronic inflammatory conditions, and malignancy. In addition, hospitalized patients are more likely to have an elevated D-dimer level than outpatients. Because of this lack of specificity, the value of the D-dimer assay resides with its high negative predictive value and the ability of a normal D-dimer to reduce the probability of PE sufficiently to avoid further diagnostic testing in patients with a low or moderate pretest likelihood, who represent up to 30% of patients with suspected VTE.

Computed Tomography Pulmonary Angiography

Multidetector CT pulmonary angiography has largely replaced ventilation-perfusion lung scanning for PE diagnosis because of its wide availability and the rapidity of its results. In contrast to lung scanning, CT pulmonary angiography not only permits direct visualization of thrombi in the pulmonary arteries of patients with PE (Fig. 98-2) but also provides an alternative

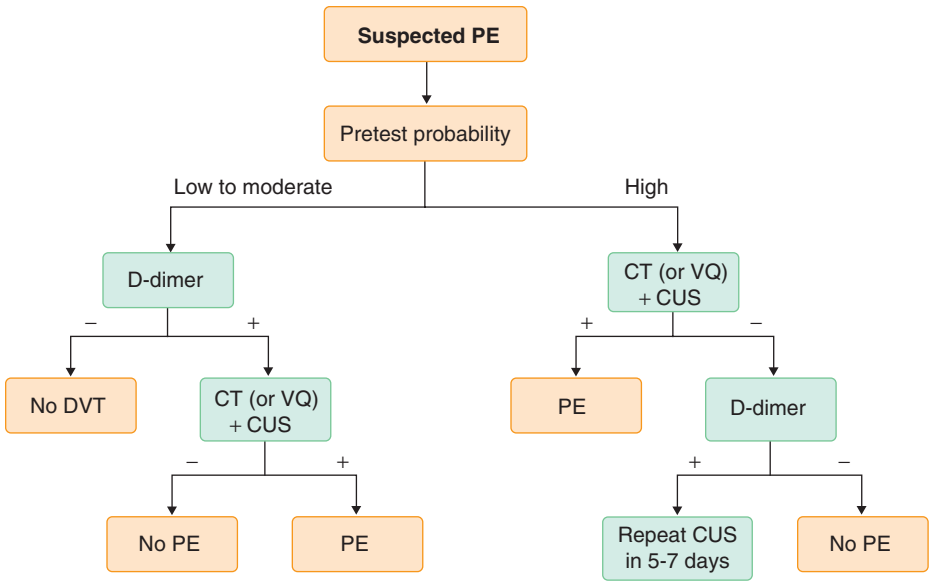

FIGURE 98-1. Clinical approach to patients with suspected pulmonary embolism (PE). CT = computed tomography; CUS = venous compression ultrasound; DVT = deep vein thrombosis; VQ = ventilation-perfusion lung scan.

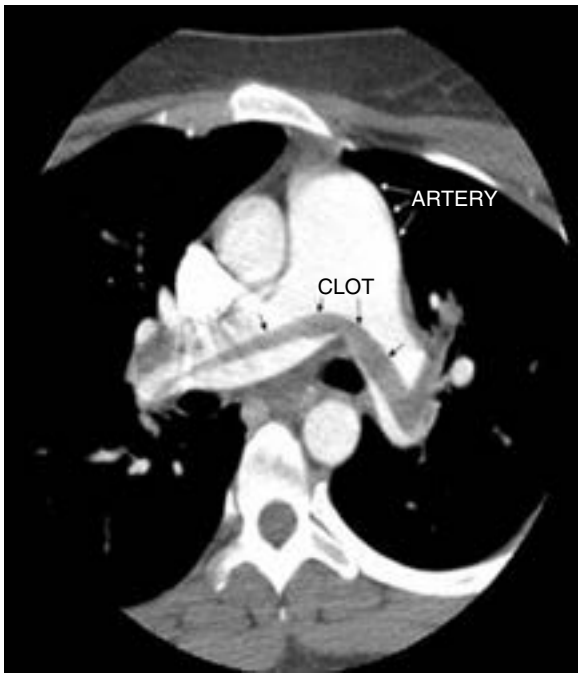

FIGURE 98-2. Computed tomographic pulmonary arteriogram demonstrating a pulmonary embolus.

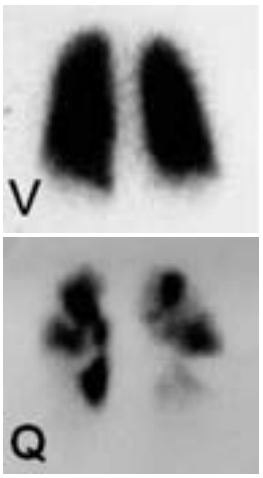

FIGURE 98-3. Ventilation-perfusion lung scan demonstrating pulmonary emboli. The ventilation scan (V) is normal, whereas the perfusion scan (Q) shows multiple defects.

diagnosis in many of those who prove not to have PE. With the evolution from single-detector to multidetector CT scanners, the sensitivity and specificity of CT pulmonary angiography are sufficient for its use as a stand-alone test: a CT pulmonary angiogram showing thrombus in pulmonary arteries up to the segmental level provides evidence of PE, whereas a negative CT pulmonary angiogram excludes PE and is associated with subsequent clinical outcomes that are at least as good as in patients with negative lung scanning. Compared with lung scanning, CT pulmonary angiography detects more isolated subsegmental thrombi in about 1 to 5% of patients, but the importance of such thrombi is unclear because patients with isolated subsegmental defects appear to have uneventful outcomes without treatment. A negative D-dimer and/or a normal venous compression ultrasound examination may help to exclude the possibility of VTE in this setting.

CT pulmonary angiography can be combined with CT venography so that the diagnosis of PE and DVT can be established with a single test and only one injection of contrast dye. Compared with CT pulmonary angiography alone, the combination of CT pulmonary angiography and CT venography increases the sensitivity for diagnosing PE from 83% to 90%, but the specificity remains unchanged, thereby resulting in only a modest increase in negative predictive value. CT venography adds significant radiation exposure and only marginally increases the overall detection rate, so venous compression ultrasound is preferred for the diagnosis of DVT because it provides the same information without exposing patients to ionizing radiation. Standard venography is not recommended in patients with suspected PE.

Ventilation-Perfusion Lung Scanning

This two-part test consists of a ventilation phase and a perfusion phase. The ventilation phase involves inhalation of an aerosol form of radioactive isotopes of xenon or technetium to assess air delivery to the various parts of the lung. In contrast, the perfusion phase involves the intravenous injection of technetium-labeled macroaggregates of albumin, which enables assessment of blood flow within the lungs after the aggregates lodge in the pulmonary microcirculation. Images for both parts of the test are acquired using a gamma counter. Areas of lung affected by PE do not light up on the perfusion scan because the macroaggregates of albumin fail to reach sites where the pulmonary arteries are occluded (Fig. 98-3). By comparison, areas with abnormal perfusion because of PE will ventilate normally, yielding a ventilation-perfusion mismatch.

A normal ventilation-perfusion lung scan effectively excludes the diagnosis of PE, but only 25% of patients with suspected PE have a normal scan. One or more segmental or larger perfusion defects that ventilate normally characterize a high probability lung scan, which establishes the diagnosis of PE. However, even with single-photon emission CT technology, only about 10% of patients with suspected PE have a high-probability lung scan. The

remaining 65% of lung scans exhibit smaller areas of mismatch or matched defects and fall within the non-high-probability category. Because up to 40% of patients with non-high-probability scans have PE, such patients require additional investigations to exclude the diagnosis.

Although CT pulmonary angiography has largely replaced lung scanning, the lung scan is the diagnostic test of choice for patients with renal impairment or a history of allergy to angiographic contrast media and for women younger than 40 years to reduce radiation exposure to the breasts. For diagnosis of PE in pregnancy, lung scanning produces less maternal and fetal radiation exposure and is preferred over CT pulmonary angiography.

Magnetic Resonance Imaging

In contrast to CT pulmonary angiography, gadolinium-enhanced magnetic resonance imaging (MRI) does not subject patients to ionizing radiation, and gadolinium can safely be given to patients with a history of allergy to contrast dye. Although originally promoted as the test of choice to diagnose PE in patients with renal impairment, the emergence of nephrogenic systemic fibrosis (Chapter 267) as a complication of gadolinium administration in patients with renal impairment has tempered enthusiasm for the test. The accuracy of MRI for the detection of PE appears to be similar to that of CT pulmonary angiography, but magnetic resonance (MR) pulmonary angiography is more technically demanding. The combination of MR pulmonary angiography with MR venography has a higher sensitivity than MR pulmonary angiography alone for the diagnosis of PE. With these limitations, MRI should be used to diagnose PE only in centers with experience with the test and in patients for whom standard tests are contraindicated.

Pulmonary Angiography

Pulmonary angiography requires direct contrast injection into the pulmonary arteries, followed by imaging using digital subtraction technology to provide high-quality images. Presence of a thrombus, which appears as a filling defect or as a sudden cutoff of blood flow in a pulmonary arterial branch, establishes the diagnosis of PE. Although direct angiography allows visualization of small thrombi in subsegmental pulmonary arteries, high interobserver variability in the interpretation of isolated filling defects at this level limits the specificity of this finding. As an invasive test, the mortality rate associated with pulmonary angiography is 0.2%, with deaths usually occurring in patients with hemodynamic compromise or respiratory failure. Because of its associated risks and because CT pulmonary angiography offers similar or better information, direct pulmonary angiography is now rarely performed, except in patients who may undergo pulmonary embolectomy.

Other Tests

Routine blood tests in PE patients should include a complete blood count, including a platelet count, and a baseline international normalized ratio (INR), and activated partial thromboplastin time (aPTT). A serum creatinine and blood urea nitrogen level are needed to help guide the choice of anticoagulant therapy, and serum electrolytes and liver enzymes provide useful baseline information.

Blood markers of right ventricular dysfunction, which occurs with extensive PE, include brain natriuretic peptide (BNP) or its precursor, N-terminal proBNP, which are released in response to myocardial stretching (Chapter 58). Elevated levels of troponin I or T provide evidence of myocardial injury (Chapter 72). Although these markers are associated with a worse prognosis in patients with PE, their positive predictive value for making the diagnosis is low.

The ECG may show new changes suggestive of right ventricular strain, such as T wave inversion in leads V_1 to V_4; the classic S1, Q3, T3 pattern (Fig. 98-4); and complete or incomplete right bundle branch block. However, these ECG changes have limited sensitivity and are seen predominantly in patients with more severe pulmonary emboli.

Echocardiographic findings suggestive of right ventricular dysfunction include right ventricular dilation or hypokinesis, increased right ventricular diameter relative to that of the left ventricle, and increased velocity of the jet of tricuspid regurgitation. Echocardiography can also detect a right-to-left shunt through a patent foramen ovale and may provide evidence of right ventricular thrombus, both of which are associated with increased mortality in patients with PE. With no universal criteria for the diagnosis of right ventricular dysfunction, however, the utility of routine echocardiography remains uncertain. In general, routine echocardiography is recommended only in patients with severe hypoxemia or other evidence to suggest hemodynamic compromise.

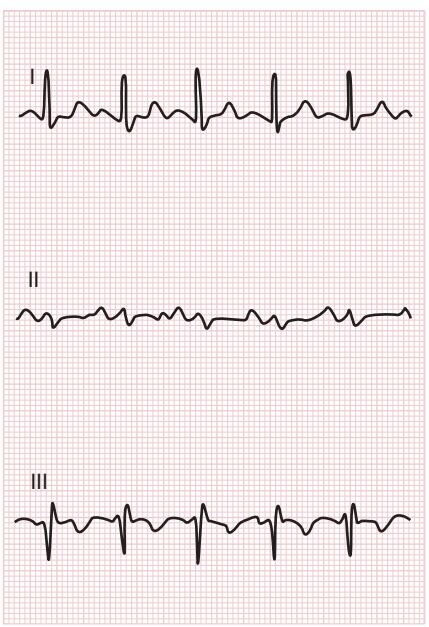

FIGURE 98-4 Classic electrocardiogram of pulmonary embolus. Note the "S1, Q3, T3" pattern.

Diagnostic Strategies

Patients with a low to moderate pretest likelihood of PE should undergo D-dimer testing (see Fig. 98-2). A negative D-dimer test excludes the diagnosis of PE in these patients, whereas a positive test should prompt multidetector CT pulmonary angiography. Patients with a high pretest likelihood of PE should be sent directly for multidetector CT pulmonary angiography. A positive CT pulmonary angiogram establishes the diagnosis of PE, whereas a negative test excludes it.

The role of venous compression ultrasound, which is used to establish the diagnosis of DVT (Chapter 81), remains controversial. Because of the suboptimal sensitivity of single-detector CT pulmonary angiography for the diagnosis of PE, bilateral venous compression ultrasound should also be performed if multidetector technology is not available. Although a negative multidetector CT pulmonary angiogram safely excludes PE, at least in patients with low to moderate pretest likelihood, bilateral venous compression ultrasound can still be helpful. For example, the finding of proximal DVT in the lower or upper extremities, which obviates the need for further testing, can be particularly helpful in patients who are poor candidates for CT pulmonary angiography, such as those with renal impairment or a history of allergy to contrast dye. Patients with a non-high-probability lung scan or with equivocal CT pulmonary angiographic findings should undergo serial venous compression ultrasound examination of the lower extremities to exclude the possibility of calf DVT, which then extends into the proximal veins.

TREATMENT

Rx

Although anticoagulant therapy remains the mainstay of treatment of PE, patients with severe hemodynamic compromise may require reperfusion therapy or surgical thrombectomy to restore blood flow rapidly to the pulmonary arteries and to reduce pulmonary artery pressure. By comparison, outpatient treatment is as safe as routine inpatient treatment in low-risk patients with acute PE. In PE patients at intermediate risk (Table 98-2), outpatient management can be considered, but brief admission to hospital may be a safer approach. Therefore, rapid risk stratification is crucial to help guide treatment. PE patients who have contraindications to anticoagulant therapy may require insertion of an inferior vena cava filter.

Severe Pulmonary Embolism

High-risk patients can be identified at the bedside (see Table 98-2) based on the presence or absence of hemodynamic compromise. Such patients also have right ventricular dysfunction and have elevated levels of biomarkers. The most common cause of death in patients with severe PE is acute right ventricular failure, which causes low systemic output. To prevent this complication, patients with severe PE require hemodynamic and respiratory support and may benefit from reperfusion therapy. Patients with right ventricular failure often require modest fluid expansion and may need inotropic agents,

TABLE 98-2 RISK STRATIFICATION OF PATIENTS WITH PULMONARY EMBOLISM AND TREATMENTS IN ADDITION TO ANTICOAGULATION

CLASSIFICATION	HEMODYNAMIC COMPROMISE	RIGHT VENTRICULAR DYSFUNCTION	INCREASED TROPONIN AND/OR BNP LEVELS*	TREATMENT
Severe	Yes	Yes	Yes	Fibrinolytic therapy
Moderate	No	Yes	Yes	May consider fibrinolytic therapy if very symptomatic
Mild	No	No	No	Consider outpatient treatment

*Blood markers include troponin, brain natriuretic peptide (BNP), and N-terminal proBNP.

such as dobutamine (starting at a dose of 0.5 to 1.0 µg/kg/minute and then titrated according to the blood pressure), dopamine (starting at a dose of 5 µg/kg/minute and then increased gradually in 5- to 10-µg/kg/minute increments, according to the blood pressure, up to 20 to 50 µg/kg/minute) or norepinephrine (starting at a dose of 2 to 4 µg/minute and then titrated according to the blood pressure) for severe hypotension or shock. Emerging data raise the possibility that endothelin antagonists and phosphodiesterase-5 inhibitors may attenuate the pulmonary hypertension in patients with severe PE, but these drugs are not currently approved for this indication.

Patients with PE frequently have hypoxemia and hypocapnia. Hypoxemia can usually be reversed with nasal oxygen. Measures to reduce fever and agitation with acetaminophen and mild sedation may help minimize oxygen consumption. In patients who have severe PE and who require mechanical ventilation, low tidal volumes should be used, and positive end-expiratory pressure should be applied with caution because it can reduce venous return and worsen right ventricular failure (Chapter 105).

Reperfusion Therapy

Patients with severe PE associated with hypotension or shock may benefit from pharmacologic[5], mechanical, or surgical reperfusion[6] therapy. Pharmacologic reperfusion therapy involves the systemic administration of a fibrinolytic agent (Table 98-3), preferably within 48 hours of the onset of symptoms, but later treatment may still be of benefit. Up to 13% of patients who receive fibrinolytic therapy experience a major bleed, and the rate of intracranial or fatal bleeding can reach 1.8%. Consequently, fibrinolytic therapy currently is justified only in patients who have severe PE and no contraindications (Table 98-4). In patients at intermediate risk, fibrolytic therapy prevents hemodynamic deterioration but increases bleeding and stroke.

Mechanical reperfusion includes percutaneous catheter embolectomy with thrombus fragmentation, an approach that avoids the need for fibrinolytic drugs altogether, or catheter-directed fibrinolytic therapy, which requires lower doses of fibrinolytic agents than are used for systemic administration. In some centers, surgical pulmonary embolectomy may be an option for patients who have severe PE, such as a saddle embolus occluding the main pulmonary artery (Fig. 98-5), and who are at high risk for bleeding with systemic fibrinolytic therapy or who have failed such treatment. Mechanical techniques require skilled operators and should only be performed in high-volume centers.

Anticoagulation Therapy

Anticoagulation therapy is the cornerstone of PE treatment and should be initiated immediately, even while patients with suspected PE are awaiting the results of confirmatory tests. Treatment starts with a rapidly acting parenteral anticoagulant (e.g., heparin, low-molecular-weight heparin (LMWH), fondaparinux, or with a direct orally active factor Xa inhibitor (e.g., rivaroxaban 15 mg twice daily). Patients with severe PE should be treated with heparin because the other agents have not been extensively evaluated in this setting. In addition, the short half-life of heparin is beneficial should reperfusion therapy be necessary. Heparin should also be used in patients with severe renal impairment (creatinine clearance <30 mL/minute) because LMWH, fondaparinux, and rivaroxaban are cleared by the kidneys. If LMWH or fondaparinux is used in patients with moderate renal impairment (creatinine clearance of 30 to 50 mL/minute), lower doses should be considered and anti–factor Xa levels should be measured at trough to ensure there is no accumulation.

Heparin should be administered by continuous intravenous infusion and dosed using weight-based nomograms (see Table 81-4). Typically, an 80 U/kg bolus is followed by an infusion at the rate of 18 U/kg per hour, and subsequent doses are adjusted based on the results of the aPTT. Rapid achievement and maintenance of a therapeutic aPTT are important to reduce the risk for recurrent PE. In addition to monitoring the aPTT, the platelet count should be measured at least two to three times per week because of the risk for heparin-induced thrombocytopenia (Chapter 172).

Subcutaneous LMWH or fondaparinux or oral rivaroxaban can be used for intermediate- or low-risk PE patients (see Table 98-2) using the regimens illustrated in Table 98-5. Unlike heparin, these agents do not require coagulation monitoring. The risk for heparin-induced thrombocytopenia is lower with LMWH than with heparin, is minimal with fondaparinux, and is nonexistent with rivaroxaban.

TABLE 98-3 APPROVED REGIMENS FOR FIBRINOLYTIC THERAPY FOR TREATMENT OF SEVERE PULMONARY EMBOLISM

AGENTS	RECOMMENDED REGIMENS
Streptokinase	250,000 IU as a loading dose over 30 min, followed by 100,000 IU/hr over 12 to 24 hr
Tissue plasminogen activator	100 mg over 2 hr, or 0.6 mg/kg over 10 to 15 min (maximal dose of 50 mg)

TABLE 98-4 CONTRAINDICATIONS TO FIBRINOLYTIC THERAPY

ABSOLUTE CONTRAINDICATIONS
Any hemorrhagic stroke or stroke of unknown origin
Central nervous system damage or neoplasm
Major trauma, surgery, or head injury in past 3 weeks
Gastrointestinal bleeding in past month
Significant ongoing bleeding
RELATIVE CONTRAINDICATIONS
Ischemic stroke or transient ischemic attack in past 6 months
Treatment with a vitamin K antagonist
Pregnancy or within 1 week of delivery
Noncompressible puncture site
Traumatic resuscitation
Advanced liver disease
Infective endocarditis
Active peptic ulcer disease

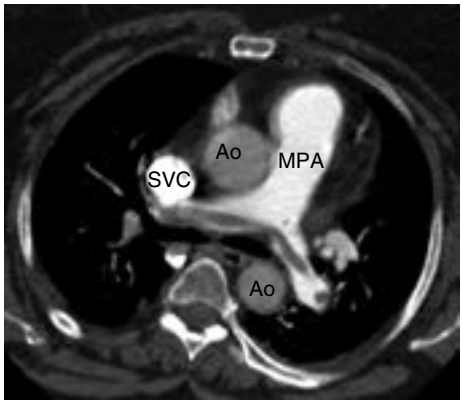

FIGURE 98-5 Thrombotic pulmonary embolism. Computed tomographic pulmonary angiogram revealing a saddle embolism in the main pulmonary artery (MPA). Ao = aorta; SVC = superior vena cava.

After initial treatment with a parenteral anticoagulant, patients with PE require long-term therapy with a vitamin K antagonist, such as warfarin (Chapter 81), or with an oral factor Xa inhibitor to prevent recurrent VTE. In PE patients at low or intermediate risk (see Table 98-2), warfarin can be started on the same day that parenteral anticoagulant therapy is initiated. Parenteral anticoagulant therapy should be continued for at least 5 days and should only be stopped when the INR has been within the therapeutic range of 2 to 3, which is the target range for long-term therapy, for at least 24 hours. Initiation of warfarin therapy should be delayed in patients with severe PE; such patients should receive heparin until they have stabilized.

TABLE 98-5 LOW-MOLECULAR-WEIGHT HEPARIN, FONDAPARINUX, RIVAROXABAN, AND APIXABAN REGIMENS FOR TREATMENT OF PULMONARY EMBOLISM

AGENT	DOSE	INTERVAL
Enoxaparin	1 mg/kg 1.5 mg/kg	Twice daily Once daily
Dalteparin	100 U/kg 200 U/kg	Twice daily Once daily
Tinzaparin	175 U/kg	Once daily
Fondaparinux	5 mg (weight < 50 kg) 7.5 mg (weight 50-100 kg) 10 mg (weight > 100 kg)	Once daily Once daily Once daily
Rivaroxaban	15 mg 20 mg	Twice daily × 3 weeks Once daily thereafter
Apixaban	10 mg 5 mg	Twice daily × 7 days Twice daily thereafter

Oral rivaroxaban (15 mg twice daily for 3 weeks, followed by 20 mg once daily thereafter) is as efficacious as enoxaparin followed by warfarin for the initial treatment of acute PE and causes less major bleeding. Alternatives include oral apixaban (10 mg twice daily for 7 days followed by 5 mg twice daily), or enoxaparin for at least 5 days followed by dabigatran (150 mg twice daily), and edoxaban (60 mg once daily or 30 mg once daily for a creatinine clearance 30 to 50 mL/minute or weight below 60 kg).[■■■] Edoxaban is not yet licensed for this indication.

Duration of Anticoagulant Therapy

Patients who develop PE as a complication of a reversible risk factor, such as surgery, trauma, or medical illness, have a low risk for recurrence when anticoagulant therapy is stopped. Consequently, a 3-month course of warfarin therapy represents adequate treatment in such patients provided that their risk factors have resolved.[7] Women who develop PE with estrogen therapy also can be treated for 3 months, provided that hormonal treatment is withdrawn. In contrast, patients with unprovoked PE have a higher rate of recurrent VTE when anticoagulant therapy stops and require longer treatment, perhaps indefinitely provided that the risk for bleeding remains low. An elevated D-dimer level 1 month after stopping anticoagulant therapy may help to identify such patients.[■] After a minimum 3-month course of usual-intensity warfarin (target INR between 2 and 3), a lower-intensity regimen (target INR between 1.5 and 2.0) may simplify management by decreasing the frequency of INR monitoring and reducing the risk for bleeding, but the risk for recurrent VTE is slightly higher with this lower intensity warfarin regimen. Rivaroxaban (20 mg once daily), apixaban (2.5 mg twice daily), or dabigatran (150 mg twice daily) are long-term alternatives. These agents produce less bleeding than extended warfarin and no difference in the risk for recurrent VTE. There may be a higher risk for myocardial infarction with dabigatran than with warfarin.[■]

Inferior Vena Cava Filters

Inferior vena cava filters, which are inserted percutaneously, are usually placed below the level of the renal veins but can be placed higher if thrombus extends into the inferior vena cava. Both permanent and retrievable filters reduce the risk for recurrent PE but have not been shown to prolong survival, in part because permanent filters can be associated with long-term complications, including inferior vena cava occlusion because of thrombus, recurrent DVT, and post-thrombotic syndrome. Retrievable filters, designed to be removed within 2 to 4 weeks of implantation, can circumvent these long-term complications, but device migration or thrombosis occurs in up to 10% of patients with temporary filters because most are not removed.[8] Because of these potential problems, vena cava filters should be restricted to patients who have high risk for recurrent PE and an absolute contraindication for anticoagulation, such as patients who develop a PE after major surgery, patients who experience major bleeding with anticoagulant therapy, and pregnant women who suffer a PE shortly before delivery. Retrievable filters should be used in these cases, and the devices should be removed as soon as anticoagulant therapy can safely be administered. Permanent filters are suitable for patients who have ongoing contraindications to anticoagulation.

Specific Patient Subgroups

Patients with PE in the setting of active cancer, women who suffer a PE during pregnancy, and patients with chronic thromboembolic pulmonary hypertension (Chapter 68) require special treatment.

Cancer

Active cancer and its treatment with chemotherapy, radiation therapy, and growth factors or other biologic agents increase the risk for VTE (Chapter 179). Patients with advanced cancer often have limited mobility, which adds to their risk for VTE.[9] In addition, indwelling percutaneously inserted or central venous access catheters can trigger upper extremity DVT, which can lead to PE. Therefore, the index of suspicion should be high in cancer patients who present with symptoms and signs suggestive of PE or DVT, or both. With advances in diagnostic imaging, incidental PE may be discovered on CT scans performed for staging purposes or for monitoring response to treatment. Although 20% of patients with VTE have an underlying malignancy, patients with PE should not undergo routine extensive screening for cancer.

Like patients without cancer, initial treatment of PE in cancer patients involves administration of a rapidly acting parenteral anticoagulant. For extended treatment, however, LMWH reduces the risk for recurrent VTE to a greater extent than warfarin. In addition, in the face of poor nutritional intake, severe nausea and vomiting, transient thrombocytopenia, or invasive procedures, LMWH is easier to manage than warfarin. The role of the new oral anticoagulants in cancer patients with PE is uncertain.

Cancer patients who develop PE after curative surgery or with adjuvant chemotherapy for limited-stage disease should be treated for at least 3 months or until they have completed their chemotherapy. Those with PE on the background of advanced cancer have a risk for recurrence of at least 20% in the first year after stopping anticoagulant therapy, so they often require extended treatment.[10]

Pregnancy

Treatment of PE in pregnancy (Chapter 239) centers mainly on heparin or LMWH because, unlike warfarin or the new oral anticoagulants, these agents do not cross the placenta.[11] Although both heparin and LMWH can be given subcutaneously, weight-adjusted LMWH is preferred over heparin because it can be given once daily without routine monitoring and because the risks of heparin-induced thrombocytopenia and osteoporosis are lower with LMWH than with heparin. Anti–factor Xa monitoring of LMWH should be considered in women at extremes of body weight and in those with renal impairment. Fondaparinux should only be considered for pregnant women who have a history of heparin-induced thrombocytopenia or who develop injection-site reactions to heparin or LMWH.

LMWH should be continued throughout pregnancy. Warfarin should be avoided because it crosses the placenta and can cause bone and central nervous system abnormalities, fetal hemorrhage, or placental abruption. During labor and delivery, epidural analgesia should be avoided unless prophylactic LMWH has been stopped at least 12 hours before insertion of the epidural catheter and therapeutic LMWH has been stopped at least 24 hours before. Treatment can be resumed within 6 hours of epidural catheter withdrawal.[12] After delivery, anticoagulation therapy should be continued for at least 3 months; warfarin can be used in place of LMWH because it does not appear in breast milk.

Fibrinolytic agents have been used successfully for treatment of severe PE in pregnancy but can cause bleeding, usually from the urogenital tract. If PE develops late in pregnancy, a retrievable filter may prevent recurrence during delivery when anticoagulant therapy must be withheld.

Chronic Thromboembolic Pulmonary Hypertension

A rare complication of PE, chronic thromboembolic pulmonary hypertension develops in 0.5 to 5% of patients over the course of months or years when emboli in major pulmonary arteries are replaced by fibrous tissue that becomes incorporated into the vessel wall, thereby narrowing or obstructing it.[13] Chronic obstruction of the pulmonary vascular bed increases pulmonary arterial resistance and can lead to right heart failure. Although patients initially may be asymptomatic, they experience increasing dyspnea on exertion and hypoxemia as the disease progresses. Chronic thromboembolic pulmonary hypertension should be suspected in patients with pulmonary hypertension (Chapter 68), and the diagnosis can be established with a combination of echocardiography and lung scanning or CT pulmonary angiography.

Medical therapy focuses on treatment of right heart failure and the use of prostacyclin, endothelin receptor antagonists, or phosphodiesterase-5 inhibitors, or a combination of these, to lower pulmonary artery pressure (see Table 68-2 and Fig. 68-6). Riociguat, a soluble guanylate cyclase stimulator, is a new option.[■] These agents may be of limited utility, however, because of the fibrotic nature of the obstructing material. Definitive treatment involves surgical thromboendarterectomy to remove the occluding material from the pulmonary arteries. This procedure is associated with a perioperative mortality rate that can be as high as 4%, depending on the severity of the disease, and a 3-year survival rate of about 80%.

PREVENTION

At least half of the outpatients with newly diagnosed VTE have a history of recent hospitalization, and most failed to receive thromboprophylaxis during their hospital stay; as a result, PE is the most common preventable cause of death in hospitalized patients in the United States. Guidelines for

primary prophylaxis are available and should be followed (see Tables 38-2 and 38-3).

PROGNOSIS

With the diagnosis established and adequate anticoagulant therapy initiated, most patients with PE survive. Case-fatality rates 1 month after diagnosis of DVT or PE are 6% and 12%, respectively. Patients with severe PE who present with shock have the highest mortality rate, and many die within an hour of presentation. Although the case-fatality rate in patients with PE is twice that in those with DVT, many of the deaths are the result of comorbid conditions rather than the PE itself. Factors associated with early mortality after VTE include presentation as PE, advanced age, cancer, and underlying cardiovascular disease. The most serious long-term complication of PE is chronic thromboembolic pulmonary hypertension (Chapter 68).

Despite anticoagulant therapy, recurrent VTE occurs in up to 6% of patients during the first 6 months. While on anticoagulation treatment, patients with provoked and unprovoked VTE have similar risks for recurrence. In contrast, when anticoagulant therapy is stopped, patients with unprovoked VTE have a risk for recurrence of 10% at 1 year and 30% at 5 years, whereas those with provoked VTE have recurrence rates of 3% at 1 year and 10% at 5 years. Recurrent events often mirror the index events; after an initial PE, about 60% of recurrences are PE. Because of the high risk for recurrence in patients with unprovoked VTE, many experts recommend indefinite anticoagulant therapy for such patients. In contrast, because of the lower risk for recurrence in patients with provoked VTE, anticoagulation therapy can be stopped after 3 months provided that transient risk factors for VTE have resolved.

In patients with unprovoked VTE, anticoagulants reduce the risk for recurrence by 80 to 90%, whereas aspirin reduces the risk by only about 35%. Therefore, anticoagulation treatment is preferred over aspirin. Extended warfarin therapy can be cumbersome, and emerging data indicate that the new oral anticoagulants, such as dabigatran, rivaroxaban and apixaban, are effective for long-term secondary VTE prevention.

NONTHROMBOTIC PULMONARY EMBOLISM

DEFINITION

Nonthrombotic material that can embolize to the lungs includes fat, air, amniotic fluid, tumor cells, talc in intravenous drug abusers, and medical devices.

Fat Embolism Syndrome

EPIDEMIOLOGY

Fat embolism syndrome usually occurs in the setting of trauma, particularly after fracture of long bones or the pelvis. The risk increases with the number of fractured bones, and the syndrome occurs more often with closed fractures than with open ones. Fat embolism also can complicate orthopedic surgery or trauma to tissues rich in fat, such as may occur with liposuction.

PATHOBIOLOGY

Characterized by a combination of respiratory, neurologic, hematologic, and cutaneous manifestations, fat embolism syndrome reflects a combination of vascular obstruction by fat globules (E-Fig. 98-2) as well as the deleterious effects of free fatty acids released from these fat globules by the action of lipoprotein lipases. These free fatty acids increase vascular permeability, induce a capillary leak syndrome, and can trigger platelet aggregation.

CLINICAL MANIFESTATIONS

Symptoms typically develop 24 to 72 hours after trauma or surgery. Patients often complain of vague chest pain and shortness of breath. Tachypnea and fever associated with disproportionate tachycardia are common. The syndrome can rapidly progress to severe hypoxemia that requires mechanical ventilation. Neurologic manifestations, which often start after the respiratory distress, include drowsiness, confusion, decreased level of consciousness, and seizures. Patients may have petechiae, particularly involving the conjunctiva, oral mucosa, and upper half of the body.

DIAGNOSIS

Fat embolism syndrome should be suspected when respiratory distress occurs a day or more after major trauma or orthopedic surgery, particularly when there are associated neurologic defects and petechiae. The chest radiograph may reveal diffuse alveolar infiltrates. Although fat droplets may be found in bronchoalveolar lavage fluid, this finding lacks specificity for the fat embolism syndrome.

PREVENTION, TREATMENT, AND PROGNOSIS

Rx

Early stabilization of long bone fractures reduces the risk for fat embolization. Supportive treatment should be provided, including oxygen and mechanical ventilation. The utility of corticosteroids remains controversial.

Although mortality rates as high as 10% have been reported, the prognosis is generally good.

Venous Air Embolism

EPIDEMIOLOGY

Venous air embolism, which involves entrapment of environmental air or exogenous gas in the venous system, requires direct communication between the air and a vein, as well as a pressure gradient that favors entry of the air into the vein. Air can be introduced through indwelling central venous catheters as a consequence of invasive surgical or medical procedures or after barotrauma.

PATHOBIOLOGY

Large venous air emboli obstruct the right ventricular pulmonary outflow tract, whereas mixtures of air bubbles and fibrin thrombi can obstruct pulmonary arterioles. In either case, right ventricular failure can result. With a patent foramen ovale (Chapter 69), venous air emboli can enter the coronary, cerebral, or systemic circulation.

CLINICAL MANIFESTATIONS

Symptoms and signs depend on the volume of air and the rapidity of its entry into the circulation. Large, rapid boluses of air are tolerated less well than slow entry of smaller amounts. Small air emboli may be asymptomatic. With larger emboli, patients often complain of dyspnea and retrosternal chest discomfort, and they may feel lightheaded. Physical findings include tachypnea, tachycardia, and evidence of respiratory distress. Patients may have signs of right heart failure. A continuous, drum-like, mill-wheel murmur, which reflects air in the right ventricle, may be heard.

DIAGNOSIS

Patients may present with ECG evidence of right ventricular dysfunction associated with elevated levels of troponin, indicative of myocardial injury. Echocardiography or chest CT may reveal air in the right ventricle. Patients may have hypoxemia and hypercapnia, and the platelet count may be low.

PREVENTION AND TREATMENT

Rx

All catheters should be removed using techniques that minimize air embolism, air should be removed from syringes before injection, and care should be taken during surgery to ensure that air bubbles do not form in blood vessels. To avoid air embolism associated with barotrauma, divers require training in how to dive and surface safely (Chapter 94).

The source of any air embolism should be identified so that further embolism can be prevented. Left lateral decubitus positioning may benefit patients who have a large air bubble trapped in the right ventricular outflow tract; such positioning places the outflow tract below the right ventricular cavity, thereby allowing the air bubble to migrate into a nonobstructing position. Aspiration of the right ventricle through a central venous catheter may also be of benefit. Patients should receive high-flow supplemental oxygen, and hyperbaric oxygenation should be considered for patients with cardiac or neurologic dysfunction.

PROGNOSIS

The outcome depends on the extent of air embolism. With good supportive care, the mortality rate can be less than 10%, even in patients with major air emboli. However, residual neurologic defects often persist.

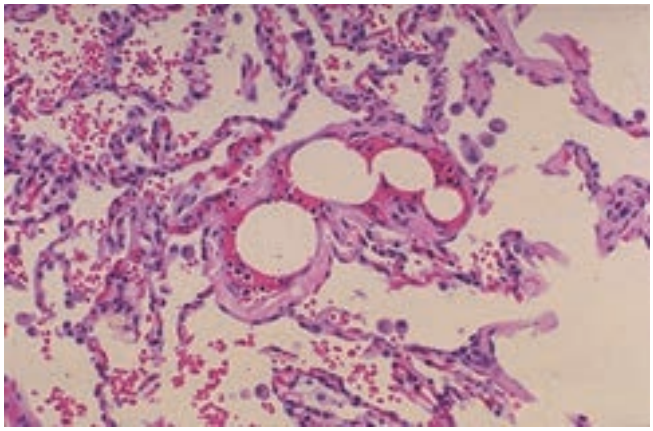

E-FIGURE 98-2. Fat embolism. Fat globules in small vessels of the lung in a patient who died 2 days after traumatic fracture of the femur.

Amniotic Fluid Embolism

EPIDEMIOLOGY AND PATHOBIOLOGY

Amniotic fluid embolism is a rare but catastrophic complication of pregnancy, occurring in about 1 in 8000 to 1 in 80,000 pregnancies. The syndrome develops when amniotic fluid and fetal cells enter the maternal blood stream through small tears in the uterine veins during labor. Emboli to the heart and lungs (E-Fig. 98-E3) cause cardiac dysfunction and respiratory distress. In addition, amniotic fluid and other debris activate the coagulation system, and the resultant thrombin then triggers fibrin formation and platelet activation to induce disseminated intravascular coagulation (Chapter 175).

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The syndrome often starts with the abrupt onset of dyspnea, cyanosis, and hypotension that can rapidly progress to cardiovascular collapse and death. Patients who survive this stage often develop manifestations of disseminated intravascular coagulation (Chapter 175) characterized by diffuse bleeding, petechiae, and ecchymoses.

The diagnosis should be suspected in women late in pregnancy, often in labor, who present with sudden onset of respiratory distress followed by cyanosis, hypotension, and shock. These findings are often associated with confusion or reduced level of consciousness, seizures, and evidence of a consumptive coagulopathy.

TREATMENT

Rx

Supportive measures include oxygen, mechanical ventilation, and hemodynamic support. Fresh-frozen plasma, cryoprecipitate, and platelets transfusion can be given to replace consumed clotting factors and platelets. Heparin, often in low therapeutic doses, may be useful in some cases. If amniotic fluid embolism occurs before or during delivery, the fetus often has a poor outcome. As soon as the mother stabilizes, therefore, every attempt should be made to deliver the fetus.

PROGNOSIS

Although rare, amniotic fluid embolism remains the leading cause of maternal death during labor and the first few hours after delivery. Despite advances in critical care management, maternal and fetal mortality rates continue to be about 60% and 20%, respectively, with up to half of the survivors, both mother and baby, suffering from permanent hypoxia-induced neurologic dysfunction.

Other Embolic Material

Many substances, such as talc, starch, and cellulose are used as fillers in the manufacture of illicit drugs. Some of these drugs are ground up by drug users (Chapter 34), mixed in liquids, and then injected intravenously. The filler particles can then be trapped in the pulmonary vasculature, where they can induce granuloma formation.

Tumor emboli in the lung can mimic pneumonia, tuberculosis, or interstitial lung disease on the chest radiograph. Cancers of the prostate and breast are the most common sources of such emboli, followed by hepatocellular cancer and cancers of the stomach and pancreas. Although found in up to 26% of autopsies in patients with advanced cancer, tumor emboli are infrequently identified before death.

Various types of intravascular devices can embolize to the lungs, including vena cava filters, broken catheter tips, guidewires, stent fragments, and coils used for embolization. Many of these devices lodge in the right atrium, right ventricle, or pulmonary arteries. Intravascular retrieval can recover most of these devices; open surgery may be required for the remainder.

Grade A References

A1. Anderson DR, Kahn SR, Rodger MA, et al. Computed tomographic pulmonary angiography vs ventilation-perfusion lung scanning in patients with suspected pulmonary embolism: a randomized controlled trial. *JAMA*. 2007;298:2743-2753.

A2. Aujesky D, Roy PM, Verschuren F, et al. Outpatient versus inpatient treatment for patients with acute pulmonary embolism: an international, open-label, randomised, non-inferiority trial. *Lancet*. 2011;378:41-48.

A3. Zondag W, Kooiman J, Klok FA, et al. Outpatient versus inpatient treatment in patients with pulmonary embolism: a meta-analysis. *Eur Respir J*. 2013;42:134-144.

A4. Meyer G, Vicaut E, Danays T, et al. Fibrinolysis for patients with intermediate-risk pulmonary embolism. *N Engl J Med*. 2014;370:1402-1411.

A5. Schulman S, Kearon C, Kakkar AK, et al. Dabigatran versus warfarin in the treatment of acute venous thromboembolism. *N Engl J Med*. 2009;361:2342-2352.

A6. Agnelli G, Buller HR, Cohen A, et al. Oral apixaban for the treatment of acute venous thromboembolism. *N Engl J Med*. 2013;369:799-808.

A7. Büller HR, Décousus H, Grosso MA, et al. Edoxaban versus warfarin for the treatment of symptomatic venous thromboembolism. *N Engl J Med*. 2013;369:1406-1415.

A8. Prandoni P, Prins MH, Lensing AW, et al. for the AESOPUS Investigators. Residual thrombosis on ultrasonography to guide the duration of anticoagulation in patients with deep venous thrombosis: a randomized trial. *Ann Intern Med*. 2009;150:577-585.

A9. Schulman S, Kearon C, Kakkar AK, et al. Extended use of dabigatran, warfarin, or placebo in venous thromboembolism. *N Engl J Med*. 2013;368:709-718.

A10. Ghofrani HA, D'Armini AM, Grimminger F, et al. Riociguat for the treatment of chronic thromboembolic pulmonary hypertension. *N Engl J Med*. 2013;369:319-329.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

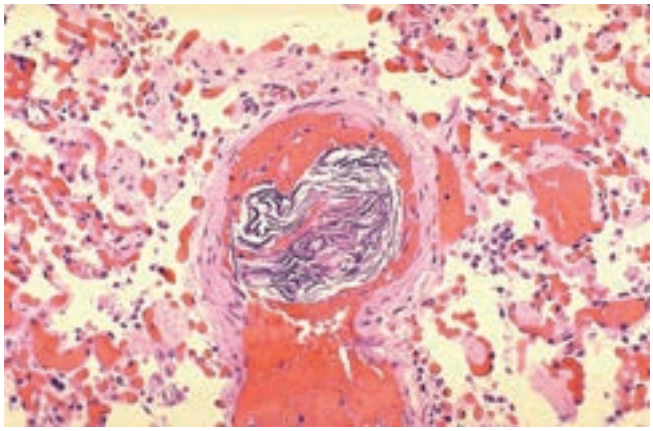

E-FIGURE 98-3. Amniotic fluid embolism. Fetal squamous cells, hair and mucin in the pulmonary vasculature of a woman who developed acute shortness of breath, circulatory collapse, and coagulopathy shortly after delivery.

GENERAL REFERENCES

1. Agnelli G, Becattini C. Acute pulmonary embolism. *N Engl J Med*. 2010;363:266-274.
2. Goldhaber SZ, Bounameaux H. Pulmonary embolism and deep vein thrombosis. *Lancet*. 2012; 379:1835-1846.
3. Hughes MJ, Stein PD, Matta F. Silent pulmonary embolism in patients with distal deep venous thrombosis: Systematic review. *Thromb Res*. 2014;134:1182-1185.
4. Huisman MV, Klok FA. Diagnostic management of acute deep vein thrombosis and pulmonary embolism. *J Thromb Haemost*. 2013;11:412-422.
5. Chatterjee S, Chakraborty A, Weinberg I, et al. Thrombolysis for pulmonary embolism and risk of all-cause mortality, major bleeding, and intracranial hemorrhage: a meta-analysis. *JAMA*. 2014;311: 2414-2421.
6. Worku B, Gulkarov I, Girardi LN, et al. Pulmonary embolectomy in the treatment of submassive and massive pulmonary embolism. *Cardiology*. 2014;129:106-110.
7. Kearon C, Akl EA, Comerota AJ, et al. Antithrombotic therapy for VTE disease. Antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2012;141:e419S-e494S.
8. Sarosiek S, Crowther M, Sloan JM. Indications, complications, and management of inferior vena cava filters: the experience in 952 patients at an academic hospital with a level 1 trauma center. *JAMA Intern Med*. 2013;173:513-517.
9. Lyman GH, Khorana AA, Kuderer NM, et al. Venous thromboembolism prophylaxis and treatment in patients with cancer: American Society of Clinical Oncology clinical practice guideline update. *J Clin Oncol*. 2013;31:2189-2204.
10. Connors JM. Prophylaxis against venous thromboembolism in ambulatory patients with cancer. *N Engl J Med*. 2014;370:2515-2519.
11. Conti E, Zezza L, Ralli E, et al. Pulmonary embolism in pregnancy. *J Thromb Thrombolysis*. 2013; 1-20.
12. McLintock C, Brighton T, Chunilal S, et al. Recommendations for the diagnosis and treatment of deep venous thrombosis and pulmonary embolism in pregnancy and the postpartum period. *Aust N Z J Obstet Gynecol*. 2012;52:14-22.
13. Lang IM, Madani M. Update on chronic thromboembolic pulmonary hypertension. *Circulation*. 2014;130:508-518.

REVIEW QUESTIONS

1. A 20-year-old woman presents with a 2-day history of pleuritic chest pain and shortness of breath on exertion. She has no medical history, but she started on an oral contraceptive 6 weeks ago. Her chest radiograph is normal. Which of the following tests would you order?

A. D-dimer
B. Ventilation-perfusion lung scan
C. Computed tomographic pulmonary angiogram
D. Bilateral compression ultrasonography of the lower extremities
E. Electrocardiogram

Answer: B The pretest probability of pulmonary embolism is high, so diagnostic testing is required to evaluate this possibility. With a normal chest radiograph, ventilation-perfusion lung scanning is likely to establish the diagnosis of pulmonary embolism and is preferred over computed tomographic pulmonary angiography in young women because there will be less radiation exposure to the breasts. A D-dimer test is not helpful in patients with a high pretest probability of pulmonary embolism because additional testing is needed regardless of whether the D-dimer is positive or negative. Although most pulmonary emboli arise from deep vein thrombosis in the lower extremities, compression ultrasonography is negative in at least 50% of patients with confirmed pulmonary embolism. The echocardiogram is not diagnostic in patients with pulmonary embolism (den Exter PL, Klok FA, Huisman MV. Diagnosis of pulmonary embolism: advances and pitfalls. *Best Pract Res Clin Haematol.* 2012;25:295-302; Komissarova M, Chong S, Frey K, et al. Imaging of acute pulmonary embolism. *Emerg Radiol.* 2013;20:89-101).

2. A 68-year-old man underwent left knee arthroplasty 6 days ago and is receiving rivaroxaban (10 mg once daily) for thromboprophylaxis. He complains of shortness of breath while ambulating, and his oxygen saturation during these episodes decreases to 92%. A computed tomographic pulmonary angiogram reveals several segmental pulmonary emboli. Which of the following treatment strategies would facilitate his discharge home?

A. Stop rivaroxaban and start warfarin; discharge him when his international normalized ratio (INR) is 2 or higher.
B. Stop rivaroxaban, give enoxaparin in therapeutic doses, and start warfarin; discharge him when the INR is 2 or higher.
C. Increase the dose of rivaroxaban to 15 mg twice daily and discharge him when he is clinically stable.
D. Insert a retrievable filter and start warfarin; remove the filter when the INR is 2 or higher.

Answer: C This patient has suffered a pulmonary embolism despite thromboprophylaxis with rivaroxaban. With an episode of provoked pulmonary embolism, he requires 3 months of anticoagulant therapy. Increasing the rivaroxaban dose from the prophylactic dose of 10 mg once daily to the treatment dose of 15 mg twice daily offers the simplest therapy. After 3 weeks, the rivaroxaban dose can be decreased to 20 mg once daily. Warfarin alone is inadequate because of its delayed onset of action. An overlap of enoxaparin with warfarin is an alternative option, but this strategy would require subcutaneous injection of enoxaparin for at least 5 days, and the warfarin would require INR monitoring and dose adjustments (Agnelli G, Becattini C. Acute pulmonary embolism. *N Engl J Med.* 2010;363:366-274; Vanassche T, Verhamme P. Rivaroxaban for the treatment of pulmonary embolism. *Adv Ther.* 2013;30:589-606).

3. A 60-year-old woman is receiving chemotherapy for metastatic breast cancer. A staging computed tomographic scan reveals multiple pulmonary emboli in segmental arteries. She is asymptomatic. Which of the following treatments is the best?

A. Give dalteparin and transition her to warfarin.
B. Give dalteparin and continue it at least as long as she is receiving chemotherapy.
C. Give rivaroxaban.
D. Give warfarin.
E. No therapy.

Answer: B Patients who have cancer and who develop multiple incidental pulmonary emboli require treatment. Low-molecular-weight heparin is better than warfarin for the prevention of recurrence in such patients. Although rivaroxaban and apixaban appear to be as effective as conventional anticoagulation for the treatment of patients with pulmonary embolism, they have not been extensively evaluated in patients with cancer, nor have they been compared with low-molecular-weight heparin in cancer patients (Lyman GH, Khorana AA, Kuderer AM, et al. Venous thromboembolism prophylaxis and treatment in patients with cancer: American Society of Clinical Oncology Clinical Practice Guideline Update. *J Clin Oncol.* 2013;31:2189-2204; Farge D, Debourdeau P, Beckers M, et al. International clinical practice guidelines for the treatment and prophylaxis of venous thromboembolism in patients with cancer. *J Thromb Haemost.* 2013;11:56-70).

4. A 32-year-old woman has been diagnosed with pulmonary embolism in her second trimester of pregnancy. Which of the following treatments would you prescribe?

A. Warfarin
B. Rivaroxaban
C. Unfractionated heparin
D. Low-molecular-weight heparin
E. Fondaparinux

Answer: D Warfarin and rivaroxaban cross the placenta and can cause fetal anomalies and/or fetal bleeding. Therefore, these agents should be avoided in pregnancy. Neither unfractionated nor low-molecular-weight heparin crosses the placenta, but low-molecular-weight heparin is preferred because it can be given once daily and is associated with a lower risk for heparin-induced thrombocytopenia and osteoporosis than is unfractionated heparin. Because there is less information about the safety of fondaparinux in pregnancy, it should be reserved for patients with a history of heparin-induced thrombocytopenia or an allergy to low-molecular-weight heparin (Fogerty AE, Connors JM. Treating venous thromboembolism in pregnancy. *Hematol Oncol Clin North Am.* 2011;25:379-391; Arya R. How I manage venous thromboembolism in pregnancy. *Br J Haematol.* 2011;153:698-708).

5. A 38-year-old previously healthy woman presents to the emergency department with a 7-day history of low-grade fever, shortness of breath on exertion, and nonproductive cough. Her heart rate is 90 beats/min, her oxygen saturation on room air is 98%, and her chest radiograph is normal. Which of the following tests would help to exclude the diagnosis of pulmonary embolism?

A. Bilateral compression ultrasound examination of the lower extremities
B. D-dimer
C. Electrocardiogram
D. Computed tomographic pulmonary angiogram
E. Ventilation-perfusion lung scan

Answer: B This patient has no risk factors for pulmonary embolism, so her pretest probability is low. Therefore, a negative D-dimer gest will help to exclude the diagnosis and will obviate the need for further investigation (Agnelli G, Becattini C. Acute pulmonary embolism. *N Engl J Med.* 2010;363:366-274; den Exter PL, Klok FA, Huisman MV. Diagnosis of pulmonary embolism: advances and pitfalls. *Best Pract Res Clin Haematol.* 2012;25:295-302).

99

DISEASES OF THE DIAPHRAGM, CHEST WALL, PLEURA, AND MEDIASTINUM

F. DENNIS MCCOOL

DIAPHRAGM

The diaphragm is a dome-shaped structure that separates the thorax from the abdomen. It consists of a central tendon and a peripheral muscular component that inserts into the rib cage laterally along the inner surface of the lower six ribs, the costal cartilages anteromedially, and the upper three lumbar vertebral bodies posteriorly. The diaphragm is innervated by the phrenic nerve, which originates from cervical nerve roots 3 through 5.

Diaphragmatic Weakness and Paralysis

EPIDEMIOLOGY AND PATHOBIOLOGY

When the diaphragm is activated, it contracts and descends caudally. The downward descent of the diaphragm increases abdominal pressure, expands the lower rib cage, and lowers pleural pressure, thereby resulting in lung inflation. It is the major muscle of inspiration, and its action accounts for approximately 70% of the inspired tidal volume in the normal individual. Diaphragm function can be impaired by disorders that affect the brain (Chapter 404), spinal cord (Chapter 400), phrenic nerve (Chapter 420), neuromuscular junction (Chapter 422), and muscle itself (Chapter 421). The incidence and prevalence of diaphragm paralysis and weakness are unknown.

CLINICAL MANIFESTATIONS

Diaphragmatic weakness or paralysis can involve either one or both hemidiaphragms.[1] With unilateral diaphragmatic paralysis, patients are generally asymptomatic at rest but may have dyspnea with exertion or when supine, especially with comorbid conditions such as obesity. If they are asymptomatic, the abnormality may be discovered as an incidental finding of an elevated hemidiaphragm on chest radiography.

Bilateral diaphragmatic paralysis is not as common as unilateral paralysis. Generally, the disability with bilateral paralysis is more dramatic than with unilateral paralysis. Orthopnea is an especially prominent symptom, and patients are often unable to sleep in the supine position. These individuals, who also experience significant dyspnea with exertion when they are lifting objects or bending, are at an increased risk to hypoventilate during sleep, especially during rapid eye movement (REM) sleep. Consequently, initial symptoms of individuals with bilateral or unilateral diaphragm weakness or paralysis may be related to nocturnal hypoventilation and include frequent nocturnal awakenings, nocturia, vivid nightmares, night sweats, daytime hypersomnolence, depression, and morning headaches.

With bilateral diaphragmatic paralysis, the physical examination is remarkable for use of accessory muscles of inspiration and paradoxical inward motion of the abdominal wall during inspiration, which is especially

noticeable when these individuals are asked to lie flat. Percussion during inspiration and expiration can detect the absence of diaphragmatic movement.

DIAGNOSIS

Disorders that can cause unilateral diaphragmatic weakness or paralysis (Table 99-1) include traumatic phrenic nerve injury, herpes zoster (Chapter 375), cervical spinal disease (Chapter 400), compressive tumors, and phrenic nerve injury related to cardiac or thoracic surgery, or mechanical ventilation. Diagnosis is often suggested by elevation of a hemidiaphragm on a chest radiograph (Fig. 99-1). The diagnosis can be confirmed by performing a "sniff test" using fluoroscopy or ultrasound, in which paradoxical (cephalad) movement of the hemidiaphragm dome occurs during a sniff maneuver.

The presence of bilateral diaphragmatic paralysis can be much more difficult to ascertain than unilateral paralysis. Chest radiography, which typically shows elevation of both hemidiaphragms, may be interpreted as a "poor inspiratory effort" or "low lung volumes." Tests that can support or refute the diagnosis include pulmonary function testing (Chapter 85), which typically shows a moderate to severe reduction of vital capacity (VC) and total lung capacity (TLC) (30 to 60% predicted). The restriction becomes more severe (10 to 30% further decrease for unilateral and 30 to 50% further decrease for bilateral paralysis) when the individual assumes the supine position. Maximal static inspiratory pressure measured at the airway opening (PImax) is reduced to 20 to 30% of predicted in individuals with bilateral diaphragmatic paralysis. The diagnosis of bilateral diaphragmatic paralysis can be confirmed if measurements of transdiaphragmatic pressure (the pressure difference between the thoracic and abdominal cavity) do not change with inspiration. Diaphragm electromyography and phrenic nerve conduction studies may be useful to distinguish neuropathy or myopathy. Unlike unilateral diaphragmatic paralysis, a sniff test is not helpful in individuals with bilateral diaphragmatic paralysis because it can yield both false-negative and false-positive results. Using two-dimensional ultrasound, the normal thickening of the diaphragm during inspiration will not be observed. If the diagnosis of bilateral diaphragmatic paralysis is confirmed, an evaluation for nocturnal hypoventilation (Chapter 86) should be undertaken. Computed tomography (CT) of the chest may be needed to exclude a mediastinal mass, and magnetic resonance imaging (MRI) of the neck may be necessary to evaluate the spinal cord and nerve roots (E-Fig. 99-1).

TABLE 99-1 CAUSES OF DIAPHRAGMATIC WEAKNESS AND PARALYSIS

TRAUMA
Cardiac surgery with cold cardioplegia
Blunt trauma
Spinal cord injury
Cervical manipulation
Scalene and brachial nerve block
TUMOR COMPRESSION
Lung cancer
Metastatic mediastinal tumor
METABOLIC
Diabetes
Vitamin deficiency (B_6, B_{12}, folate)
Hypothyroidism
Acid maltase deficiency
INFLAMMATORY NEURITIS
Neuralgic amyotrophy (Parsonage-Turner)
Mononeuritis multiplex
Vasculitis
Paraneoplastic
MUSCULAR DYSTROPHIES
Limb-girdle
Duchenne and Becker
MISCELLANEOUS
Amyloidosis
Malnutrition
Radiation injury
Cervical spondylosis
Poliomyelitis
Amyotrophic lateral sclerosis
IDIOPATHIC

TREATMENT AND PROGNOSIS

Rx

Bilateral diaphragmatic paralysis may not be reversible unless the underlying cause is treatable. For example, myopathies (Chapter 421) related to metabolic disturbances may be improved by correcting electrolyte imbalances or replacing thyroid hormone. Toxic or metabolic disturbances related to diabetes, alcohol, or viral infections may resolve with treatment of the underlying disease. Idiopathic diaphragmatic paralysis or paralysis due to neuralgic amyotrophy (brachial plexus neuritis) may spontaneously improve or resolve completely in approximately 60% of individuals, but recovery can take 18 months to 3 years. The phrenic nerve courses through the pericardium and can be injured during

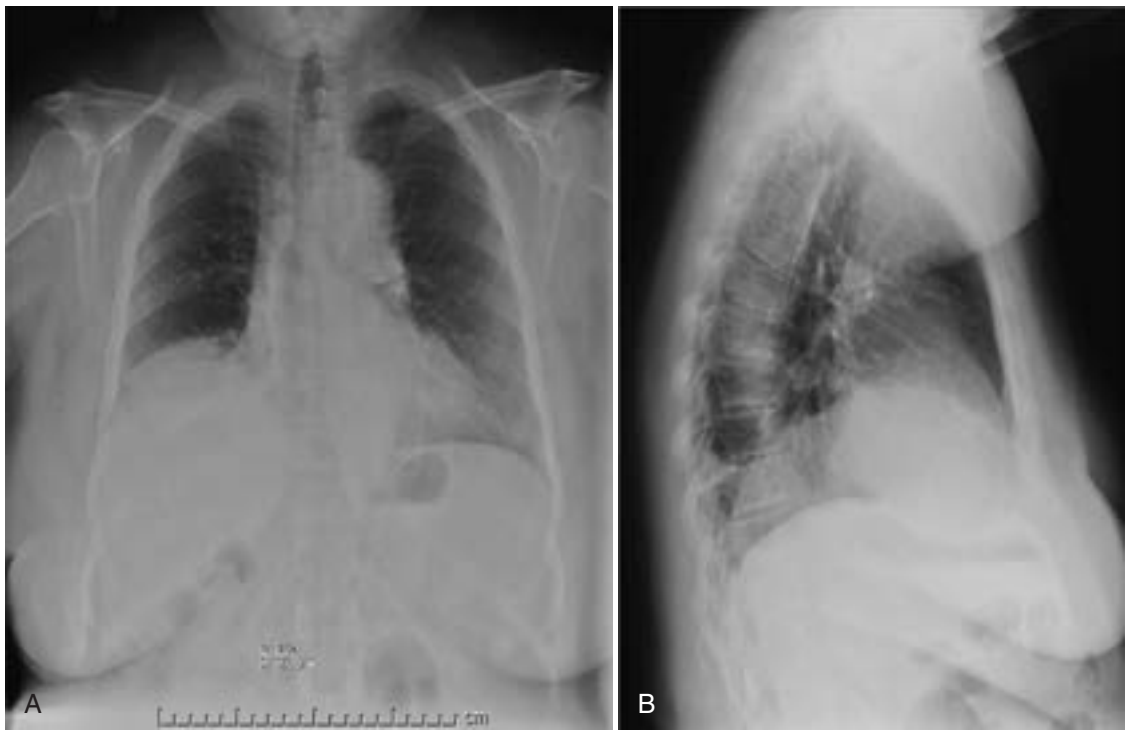

FIGURE 99-1. Patient with paralysis of the right hemidiaphragm as seen on the posteroanterior radiograph of the chest (**A**) and lateral radiograph of the chest (**B**).

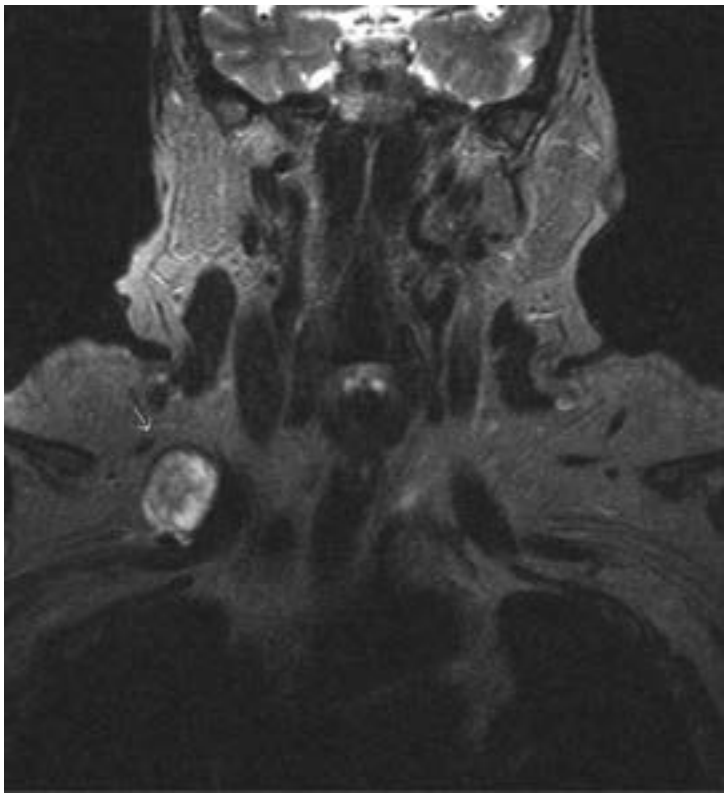

E-FIGURE 99-1. Magnetic resonance imaging of the neck in the same patient as in Figure 99-1. There is a hypervascular mass at the apex of the right axilla and right supraclavicular fossa that was a schwannoma. The right phrenic nerve was involved, resulting in right hemidiaphragm paralysis.

TABLE 99-2 THERAPEUTIC BENEFITS OF NONINVASIVE MECHANICAL VENTILATION IN PATIENTS WITH CHEST WALL AND NEUROMUSCULAR DISORDERS*

GAS EXCHANGE INDICES	
PaO_2	Increase
$PaCO_2$	Decrease
Bicarbonate	Decrease
RESPIRATORY MECHANICS	
MIP, MEP	No change or slight increase
HEMODYNAMIC PARAMETERS	
PAP	Decrease
VENTILATORY CONTROL	
Hypercapnic ventilatory response	Increase
SLEEP	
Epworth sleepiness score	Decrease
OTHER PARAMETERS	
Quality of life	Improvement
Survival	Increase

*Efficacy data derived from mostly nonrandomized, noncontrolled studies.
MEP = maximal expiratory pressure; MIP = maximal inspiratory pressure; PAP = pulmonary artery pressure.

thoracic surgery. Phrenic nerve damage related to cardiac surgery usually resolves spontaneously but may persist if the phrenic nerve is transected. For a high spinal cord injury, in which the phrenic nerve roots remain intact (injury above C3), phrenic nerve pacing can provide ventilation.

As in sleep-disordered breathing (Chapter 100), noninvasive positive-pressure ventilation (NPPV) is the preferred method of treatment for patients with diaphragmatic paralysis because it can improve both symptoms and physiologic derangements.

Nocturnal positive-pressure ventilation is associated with a number of benefits in patients with neuromuscular disease (Table 99-2). The improvement in ventilation may be related to resting the diaphragm during periods of mechanical ventilation; this reduced work appears to reverse chronic respiratory muscle fatigue and improve daytime function. Other benefits may be related to changes in the control of breathing by reversing the brain's adaptation to high levels of CO_2 by "resetting" the central controller toward normal. When patients with unilateral paralysis have severe symptoms, surgical plication of the paralyzed hemidiaphragm may improve vital capacity, but this intervention has no role in bilateral diaphragm paralysis.

Miscellaneous Diaphragmatic Disorders

Diaphragmatic eventration results from localized atrophy of the diaphragm muscle or from part of the diaphragm being replaced with fibroelastic tissue. Eventration most often results in an elevation of the right anteromedial portion of the diaphragm. Metastatic tumors to the diaphragm usually are related to direct extension of lung cancer. Primary tumors of the diaphragm are very rare. Lipomas are the most common benign tumor, and fibrosarcomas are the most common malignant neoplasm.

CHEST WALL

The chest wall is a key component of the "inspiratory pump" and allows for maintenance of normal alveolar ventilation. It consists of the bony structures of the rib cage, the articulations between the ribs and the vertebrae, the diaphragm, intercostal muscles, and the abdomen. Disorders that affect any of the components of the chest wall can result in impaired breathing.

Kyphoscoliosis

EPIDEMIOLOGY AND PATHOBIOLOGY

Kyphoscoliosis, which is a common spinal disorder, affects approximately 1 in 1000 individuals, and about 1 in 10,000 affected individuals has a severe spinal deformity.[2] Deformities include excessive spinal curvature in the coronal (scoliosis) and sagittal (kyphosis) planes as well as rotation of the spinal axis. Kyphoscoliosis can be idiopathic or can be secondary (paralytic) and associated with neuromuscular diseases, such as muscular dystrophy and polio (Chapters 421 and 379). Kyphoscoliosis also may be associated with congenital vertebral malformations.

Idiopathic kyphoscoliosis, in which there may be a familial predominance, usually manifests in late childhood or early adolescence and involves females more than males with a ratio of 4 : 1. Although a defect in the chromatin-remodeling gene family (*CHD7*) has been associated with idiopathic kyphoscoliosis, other genes have also been identified.

Kyphoscoliosis produces one of the most severe restrictive impairments of all the chest wall diseases. TLC and VC may be reduced to as low as 30% of predicted values, which are based on arm span rather than height. This restrictive pathology becomes most severe as the degree of spinal angulation increases. The patient's age, degree of spinal rotation, presence of respiratory muscle weakness, and involvement of the thoracic vertebrae are all factors that promote the restrictive process. Respiratory failure is a common cause of morbidity and mortality in patients with kyphoscoliosis.

CLINICAL MANIFESTATIONS

Individuals with mild to moderate kyphoscoliosis may have complaints of back pain and have psychosocial problems as a result of their deformity. Kyphoscoliosis may be classified as mild, moderate, or severe based on the angle of spinal deformity. Adolescents with mild idiopathic kyphoscoliosis usually have normal exercise capacity, whereas those with moderate idiopathic kyphoscoliosis have reduced exercise capacity with additional exercise limitation due to deconditioning. With severe deformities, patients may experience dyspnea with minimal exertion or at rest.

Severe kyphoscoliosis can be readily diagnosed on physical examination. Typical findings are the dorsal hump, which is due to the angulated ribs and shoulder asymmetry, as well as the hip tilt that is related to the spinal rotation. In younger individuals with milder spinal deformities, the initial changes may be subtle. The Adams forward bend test, in which the examiner observes for thoracic or lumbar region asymmetry while the patient bends forward at the waist until the spine becomes parallel to the floor, can help detect minor deformities. With severe kyphoscoliosis, signs of right heart failure (Chapter 58) may be present, such as cyanosis, distended neck veins, peripheral edema, and hepatomegaly.

Individuals with kyphoscoliosis are particularly prone to hypoventilation during sleep, especially REM sleep. Because sleep-related abnormalities and their effects on cardiorespiratory function are potentially treatable, individuals with kyphoscoliosis should be evaluated for nocturnal hypoventilation well in advance of the development of daytime hypercapnia.

DIAGNOSIS

Although spinal deformity is often readily apparent on physical examination, the degree of spinal deformity should be assessed by calculation of the angle of spinal curvature (the Cobb angle) from radiographs. This angle is formed by the intersection of lines parallel to the top and bottom vertebrae of the scoliotic or kyphotic curves (Fig. 99-2). Angles more than 100 degrees are severe and usually associated with respiratory symptoms such as dyspnea. Angles more than 120 degrees can be associated with respiratory failure. Factors associated with progression of the spinal deformity include inspiratory muscle weakness, a large spinal curvature at the time of presentation, skeletal immaturity, and a thoracic location of the curve apex. Individuals with inspiratory muscle weakness and kyphoscoliosis are more prone to develop respiratory failure than those with normal inspiratory muscle strength.

TREATMENT

Rx

Patients should be encouraged to remain physically active to minimize peripheral muscle deconditioning. In addition, general supportive measures including immunizations against influenza and pneumococci (Chapter 18), smoking cessation (Chapter 32), maintenance of a normal body weight (Chapter 220), and treatment of respiratory infections in a timely fashion should be instituted. Patients with severe kyphoscoliosis and Cobb angles of more than 100 degrees should be monitored closely for respiratory complications and nocturnal hypoventilation. Respiratory failure may be precipitated by respiratory infections or by medications that produce central nervous system depression.

Nocturnal hypoventilation, which typically precedes findings of daytime hypercapnia and hypoxemia, should be treated with NPPV. This is typically delivered through a nasal or full face mask. Indications for instituting NPPV include symptoms of nocturnal hypoventilation or signs of cor pulmonale (Chapter 68) with either an elevated daytime $PaCO_2$ or nocturnal oxygen saturation of less than 89% for 5 consecutive minutes. Supplemental oxygen will be needed if hypoxemia persists despite correction of hypoventilation. NPPV

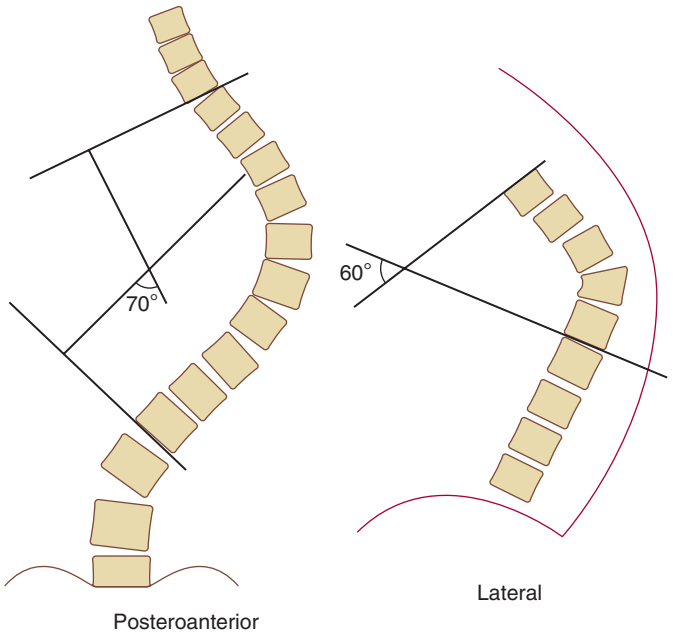

FIGURE 99-2. Schematic drawings of the spine illustrating the lines constructed to measure the Cobb angle of scoliosis and kyphosis. The angle can be calculated either from the intersection of the lines parallel to the vertebrae (as shown for kyphosis, on the *right*) or from the intersection of lines perpendicular to these lines (as shown for scoliosis, on the *left*).

can reduce the number and duration of hospitalizations and improve gas exchange, daytime blood gases, quality of life, and survival (see Table 99-2).

Surgical and nonsurgical (back-brace) treatments have been used in skeletally immature patients with idiopathic kyphoscoliosis in an effort to correct or prevent progression of the spinal deformity. Braces have been used for growing children with Cobb angles between 25 and 40 degrees, whereas surgery has been used for adolescents with a Cobb angle of more than 45 degrees. Surgical techniques have improved since the introduction of Harrington rods in the 1960s, but the overall role of surgical management in restoring pulmonary function to that of scoliotic individuals and minimizing the possibility of respiratory failure is not clear.[3]

PROGNOSIS

Idiopathic kyphoscoliosis has a better prognosis than kyphoscoliosis secondary to neuromuscular diseases. In general, individuals with mild idiopathic kyphoscoliosis have an overall benign course. Patients with moderate or severe deformities are at higher risk for developing respiratory complications.

In secondary kyphoscoliosis, early age of onset, rapid curve progression during growth, progression of scoliosis after skeletal maturity, large curves at the time of presentation, and a thoracic rather than a thoracolumbar or lumbar location of the curve apex are risk factors for respiratory complications. Respiratory failure may occur in individuals with mild or moderate kyphoscoliosis owing to concurrent respiratory muscle dysfunction. Muscle strength should be evaluated in individuals with respiratory failure and Cobb angles of less than 100 degrees. When cor pulmonale develops (Chapter 68), the prognosis is poor, and death may occur within 1 year without therapy.

Pectus Excavatum

EPIDEMIOLOGY AND PATHOBIOLOGY

Pectus excavatum, a common congenital chest wall deformity that occurs in approximately 0.5 to 2% of the population, is characterized by excessive depression of the sternum and its adjacent costal cartilages. The ratio of affected males to females is 4:1, and a family history is common. Pectus excavatum produces minimal functional impairment of the respiratory system. Occasionally, a restrictive defect will be present with mild reductions in VC and TLC. Individuals with the most severe pectus deformities may exhibit a mild reduction in maximal exercise capacity.

CLINICAL MANIFESTATIONS

Cosmetic concerns are the usual reason for seeking medical attention. Dyspnea with activity or exercise may be present but is usually out of proportion to any measurable abnormality in cardiopulmonary function. On physical examination, sternal depression is readily apparent, and there is normal excursion of the rib cage during inspiration. A mild degree of scoliosis may be present in 40 to 60% of individuals.

DIAGNOSIS

Pectus excavatum, which is diagnosed by inspecting the rib cage, usually becomes noticeable before or at puberty. Chest CT is the best means of assessing the sternal deformity. The anteroposterior diameter of the rib cage and the transverse diameter of the rib cage are measured at the level of the deepest sternal depression. Normally, the transverse-to-anteroposterior diameter ratio is 2.5. A ratio of greater than 3.5 signifies a significant pectus deformity.

TREATMENT AND PROGNOSIS

Rx

Surgical correction of the deformity is considered for patients with a CT ratio of more than 3.5 in conjunction with symptoms of dyspnea or laboratory evidence of cardiac or pulmonary restriction. However, there is no convincing evidence that correction of the deformity improves either cardiopulmonary function or exercise capacity. Invasive surgical approaches include resecting costal cartilage and repositioning the sternum. Sternal necrosis and infection may complicate invasive surgical procedures, and sternal osteotomy should be avoided in early childhood because it may be complicated by arrested growth of the rib cage. Minimally invasive approaches insert curved metal rods into the sternum through small incisions on each side of the rib cage; over time, the rod is rotated to force the sternum outward. After approximately 2 years, the rods are removed. Complications do not differ between these surgical approaches. The prognosis is excellent in individuals who have only mild deformity and patients who have undergone minimally invasive surgical correction of more severe deformities.

Flail Chest

EPIDEMIOLOGY AND PATHOBIOLOGY

Flail chest occurs following trauma in which there have been either double fractures of three or more contiguous ribs or combined sternal and rib fractures. The result is an unstable segment of the chest wall, which moves paradoxically inward with inspiration and outward with expiration. When multiple single rib fractures produce an unstable segment of the chest wall, this condition is referred to as *nonintegrated chest wall* rather than *flail chest*. In adults, flail chest is most commonly a consequence of blunt chest wall trauma owing to automobile accidents or falls (Chapter 111). In children, the chest wall is more compliant, and flail chest is infrequent.

With a flail chest, inspiratory capacity is limited, and VC may be reduced by 50% or more. Reductions in lung compliance, owing to concomitant pulmonary contusion or microatelectasis, will worsen the restrictive dysfunction and impair gas exchange. The most common location for a flail segment is the lateral chest wall, and this location is generally associated with more clinical derangement. By contrast, a posterior flail segment has less impact on the respiratory system because of splinting provided by the paraspinal muscles. Patients with flail chest may have concomitant lung contusion, pneumothorax, or hemothorax; appropriate diagnostic studies should be undertaken (Chapter 111).

Hypoventilation and flail-induced changes in respiratory muscle function are key factors in the development of respiratory failure in individuals with flail chest (E-Fig. 99-2). Excessive shortening of the inspiratory muscles related to the flail segment and excessive pressure requirements related to lung stiffness increase the work of breathing. The increased work, in concert with reduced oxygen supplies due to pulmonary contusion or hypoventilation, predisposes these individuals to developing respiratory muscle fatigue. Pain associated with the rib fractures and the disordered motion of the flail segment during expiration impair cough. Flail chest and its associated pain promote atelectasis, impair cough, and may lead to respiratory failure.

DIAGNOSIS

Bedside inspection and gentle palpation of the rib cage and abdomen can reveal the characteristic paradoxical movement of the flail segment during spontaneous breathing. Chest radiographs will confirm the presence of multiple rib fractures, but a three-dimensional reconstruction of a thoracic CT can provide better visualization of thoracic injuries, including lung contusion.

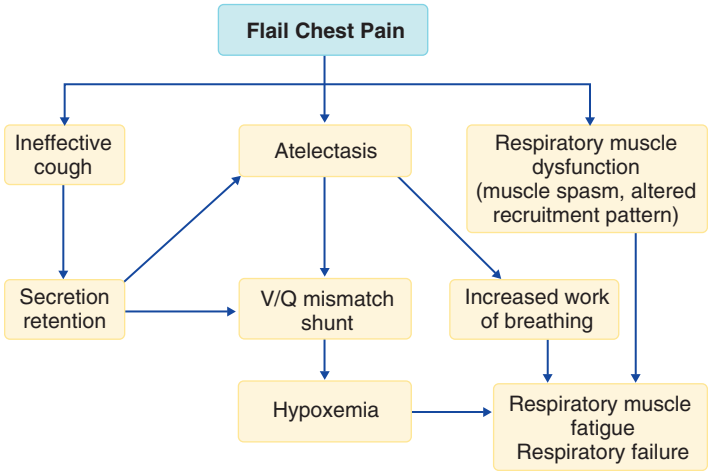

E-FIGURE 99-2. Factors involved in the pathophysiology of respiratory failure in flail chest.

The diagnosis may be less apparent in a sedated mechanically ventilated patient, in whom paradoxical motion of a segment of the rib cage may not occur because the positive alveolar and pleural pressures act as a "pneumatic splint" and allow for uniform inflation of the chest wall. However, after withdrawal of sedation, spontaneous breathing should reveal the flail segment.

TREATMENT

Rx

Nonsurgical management of flail chest consists of adequate analgesia, clearance of bronchial secretions, and mechanical ventilatory assistance if needed. Pain relief (Chapter 30) can be accomplished by oral medications, patient-controlled analgesic pumps, intercostal nerve blocks, or epidural anesthesia. Pain control is crucial in averting atelectasis and achieving an effective cough. These interventions often produce a successful outcome with avoidance of respiratory failure and mechanical ventilation, and as the rib cage heals, function is restored. If respiratory failure ensues, not only will mechanical ventilation enhance gas exchange, but also the positive pleural pressure provided by the ventilator during inspiration provides a pneumatic splint, which will stabilize the flail segment. Ventilation delivered by a nasal or face mask to patients who are spontaneously breathing improves gas exchange and may allow for early mobilization and access to physical therapy.

Surgical techniques, none of which is supported by large randomized clinical trial evidence, include open thoracotomy with stabilization of the fractured ribs.[4] The indications for operative fixation are not fully defined, but patients who are unable to wean from mechanical ventilation owing to chest wall instability or patients who are undergoing thoracotomy for concomitant injuries may be candidates for surgical fixation.

PROGNOSIS

A flail chest is a marker of increased mortality, both in patients with isolated chest wall trauma and in patients with multiple sites of trauma. The overall mortality from chest wall trauma is high, ranging from 7 to 16%.[5] If concomitant lung contusion and flail are present, the mortality rate increases further and may be as high as 70%. The associated high mortality rate is also partly due to the coexistence of other injuries, such as fractures of the long bones, head trauma, or rupture of major vascular structures (Chapter 111). Older age is another poor prognostic factor.

In patients with flail chest without lung contusion, the restrictive impairment can resolve, and the VC can return to baseline values within 6 months after the acute injury. By contrast, individuals with flail chest and concomitant lung contusion often have impairment of pulmonary function that may persist for up to 4 years after injury. Operative fixation of the chest wall may reduce long-term respiratory dysfunction in these individuals.

Ankylosing Spondylitis

EPIDEMIOLOGY AND PATHOBIOLOGY

Ankylosing spondylitis (Chapter 265) is a chronic progressive inflammatory disease that involves the ligamentous structures of the spine, sacroiliac, and large peripheral joints. It affects men more than women, with the most common age of onset between 15 and 25 years. There is a genetic predisposition, with the HLA-B27 antigen present in 95% of whites with ankylosing spondylitis. Chronic inflammation of the spine and peripheral joints may cause fibrosis and ossification of structures adjacent to the spine. Fusion of the costovertebral and sternoclavicular joints produces relative fixation of the rib cage in an inspiratory position, with limited motion of the rib cage during inspiration and a mild restrictive respiratory impairment. The degree of restriction is proportional to disease activity and duration as well as to the degree of spinal and rib cage immobility. Concomitant kyphosis, which may occur in advanced disease or be secondary to osteoporosis, will further impair respiratory function. Because the rib cage is less distensible, the diaphragm shortens to a greater degree for a given tidal volume, thereby increasing the work performed by the diaphragm. Intercostal muscle atrophy secondary to decreased rib cage mobility may cause inspiratory muscle weakness.

About 1 to 4% of patients with ankylosing spondylitis develop fibrobullous upper lobe disease. The cause of this apical disease is unknown, but its presence may impair gas exchange. Occasionally, individuals with ankylosing spondylitis develop nonapical interstitial lung disease (Chapter 92). When these complications occur, lung compliance, which otherwise is usually normal despite a pronounced reduction in rib cage mobility, may be reduced.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Ankylosing spondylitis is diagnosed when there is a history of low back pain and stiffness for more than 3 months, limited chest wall expansion, limited lumbar spine motion in the sagittal and frontal planes, and radiographic evidence of sacroiliitis. Radiographic findings include calcifications and ossifications of the vertebrae and paravertebral tissues.

Low back pain and spinal stiffness are the most common presenting symptoms. Dyspnea may be present as the disease progresses and chest wall expansion becomes more restricted. Physical examination may reveal loss of lateral flexion of the lumbar spine, tenderness over the sacroiliac joints, and kyphosis. Respiratory function may be further impaired and kyphosis worsened by fractures involving the rigid spine. Endotracheal intubation should be performed with caution because hyperextension of a rigid cervical spine may lead to fracture because of frequent involvement of the cricoarytenoid joint.

TREATMENT AND PROGNOSIS

Rx

Exercise and physiotherapy programs can enhance cardiorespiratory fitness and spinal mobility. When ankylosing spondylitis is treated with nonsteroidal anti-inflammatory agents and biologic agents, such as tumor necrosis factor antagonists (Chapter 265), rib cage expansion can also improve. A high index of suspicion for possible reactivation of latent tuberculosis is required for patients treated with anti–tumor necrosis factor-α (anti-TNF-α) agents. Corticosteroids do not prevent the progression of fibrobullous disease, which can be complicated by major hemoptysis. Thoracic surgery may be indicated for hemoptysis, but surgical intervention carries major risk, with 50 to 60% of patients developing bronchopleural fistulas. Ankylosing spondylitis rarely leads to pulmonary disability unless individuals develop fibrobullous disease.

Obesity

PATHOBIOLOGY

The degree of obesity (Chapter 220) can be assessed by measuring the body mass index (BMI), which is the ratio of body weight (BW) in kilograms to the square of the height (Ht) in meters (BW/Ht^2). Individuals with a BMI between 18.5 and 24.9 kg/m^2 are normal, whereas those with a BMI greater than 40 kg/m^2 are considered morbidly obese.

Obesity can be accompanied by changes in pulmonary function or alterations in respiratory control. Obesity characteristically reduces end-expiratory lung volume or functional reserve capacity (FRC) and expiratory reserve volume (ERV). TLC and VC may be normal or only mildly reduced, except in the most severe cases. Obesity can be accompanied by alterations in respiratory control, resulting in elevated levels of P_{ACO_2} (Chapter 86), because chronic hypoxemia or hypercapnia resets the central chemoreceptors.

Dyspnea and exercise intolerance, which are the most frequent respiratory complaints of obese individuals, may be related to disordered chest wall mechanics, altered respiratory control, or the presence of inflammatory mediators, such as those linking obesity to airway hyper-responsiveness. The increase in intra-abdominal pressure when supine will further reduce FRC, may result in expiratory flow limitation during tidal breathing, and can cause orthopnea. The worsening of respiratory mechanics in the supine position must be considered when evaluating the risk for general anesthesia in obese patients (Chapter 431). Weight loss (Chapter 220) is the optimal therapy, and it improves lung volumes, respiratory muscle performance, gas exchange, dyspnea, and sleep apnea. In individuals with obesity-hypoventilation syndrome (Chapter 100), nocturnal positive-pressure ventilation can reverse gas exchange abnormalities.

PLEURA

The pleura, which is a thin membrane that covers the inner surfaces of the thoracic cavity, consists of a layer of mesothelial cells supported by a network of connective and fibroelastic tissue. The visceral pleura lines the lung, whereas the parietal pleura lines the rib cage, diaphragm, and mediastinal structures. The closed space between the visceral and parietal pleura is referred to as the pleural space. The vascular supply of the parietal pleural surface is from the systemic circulation, and it contains sensory nerves and lymphatics. By contrast, the visceral pleura is supplied with blood vessels from the pulmonary circulation and has no sensory nerves.

Pleural Effusion

EPIDEMIOLOGY AND PATHOBIOLOGY

The overall frequency of pleural effusion on a chest radiograph ranges from 0.3 to 1% but varies widely depending on the underlying disease. Pleural effusions occur most frequently in patients with pneumonia (Chapter 97) or heart failure (Chapter 58).

Normally, a small amount of fluid in the pleural space forms a thin layer between the visceral and parietal pleural surfaces and acts as a lubricant to minimize friction between the chest wall and lung as they move against each other during inspiration and expiration. There is continual movement of fluid into and out of the pleural space. This flux of fluid depends on the oncotic and hydrostatic pressures within the parietal and visceral pleura as well as the pressure within the pleural space itself. Hydrostatic pressure in the parietal pleura is similar to systemic circulation (30 cm H_2O), whereas that of the visceral pleura is similar to the pulmonary circulation (10 cm H_2O). Accordingly, most fluid in the pleural space is filtered from the higher pressure vascular structures in the parietal pleura. Because the pressure within the pleural space itself is more subatmospheric at the apex than at the base, most of the fluid filters in from the less dependent upper lung zones. Fluid is drained out primarily through lymphatics in the parietal pleura. The fluid enters through lymphatic stomas on the surface of the parietal pleura, which are located beneath the mesothelial monolayer. The normal turnover of fluid within the pleural space is 10 to 20 mL/day, with only 0.2 to 1 mL remaining in the pleural space.[6] Excess fluid can accumulate in the pleural space because of decreased removal (owing to obstruction of pleural lymphatics) or increased production (owing to an increase in hydrostatic pressure, a decrease in oncotic pressure, decreased pressure in the pleural space, or increased pleural membrane permeability; Table 99-3).

An increase in hydrostatic pressure or decrease in oncotic pressure will result in a low-protein collection of pleural fluid characterized as *transudates* (Table 99-4). For example, heart failure (Chapter 58) can produce transudates by increasing hydrostatic pressure in the pulmonary venous system, atelectasis (Chapter 90) can promote transudates by making pleural pressure more subatmospheric, and occasionally, oncotic pressure may be sufficiently reduced to cause transudates (e.g., with hypoalbuminemia). Changes in pleural membrane permeability can produce high-protein effusions, which are characterized as *exudates* and can be seen in malignancy or inflammatory states such as pneumonia, tuberculosis, or rheumatoid arthritis. Tumors can disrupt the integrity of the mesothelial layer or the integrity of the capillary epithelium, thereby resulting in exudative effusions, or they may block lymphatic drainage either through interference with stomal openings into the pleural space or obstruction of lymphatic channels.

CLINICAL MANIFESTATIONS

Patients with pleural effusions may be asymptomatic or may experience dyspnea. When the parietal pleura is actively inflamed, pain can be present, and it is generally unilateral, is sharp, and worsens with inspiration. At times, effusions may be sufficiently large to contribute to respiratory failure. Physical findings include dullness to percussion in the area of the effusion, along with diminished breath sounds and absent tactile fremitus.[7]

DIAGNOSIS

Chest radiography (Chapter 84) is often the first imaging method used to detect an effusion (Fig. 99-3). The volume of fluid in the pleural space needs to exceed 250 mL to be visualized on the chest radiograph. When an effusion is present, there is blunting of the costophrenic angle on the posteroanterior chest radiograph (see Fig. 99-3), and a meniscus can be seen posteriorly on the lateral chest radiograph (see Fig. 99-3B). Fluid may also collect in either the minor or major fissures. Occasionally, pleural fluid collections in the major or minor fissures may appear as a pulmonary mass and are referred to as *pseudotumors*. Apparent elevation or changes in the contour of the diaphragm on a posteroanterior chest radiograph may signify a subpulmonic effusion, named because it does not distort the general shape of the diaphragm but will be evident on the lateral film. A lateral decubitus chest radiograph can be obtained to determine whether fluid is free flowing or loculated. Chest CT provides much better characterization of pleural and parenchymal abnormalities by better defining loculated effusions, distinguishing between atelectasis and effusion, and distinguishing loculated effusion from lung abscess or other parenchymal processes (Fig. 99-4). The edge of a parenchymal process usually touches the chest wall and forms an acute angle, whereas the edge of an empyema is usually an obtuse angle.

To determine the etiology of the effusion, a sample of fluid can be removed from the pleural space by thoracentesis, preferably using ultrasound or CT guidance to minimize procedural complications (Fig 99-5). The tests needed to make a diagnosis require a relatively small amount of fluid (30 to 50 mL). Larger volumes of fluid can be removed (1 to 1.5 L) in an attempt to alleviate symptoms. Removing volumes greater than 1.5 L may result in re-expansion

TABLE 99-3 MECHANISMS PROMOTING PLEURAL FLUID ACCUMULATION

MICROVASCULAR CIRCULATION
Increased hydrostatic pressure (heart failure)
Decreased oncotic pressure (severe hypoalbuminemia)
Increased permeability (pneumonia)
PLEURAL SPACE
Decreased pressure (lung collapse)
LYMPHATICS
Impaired lymphatic drainage (malignant effusion)
DIAPHRAGM
Movement of fluid from the peritoneal space (hepatic hydrothorax)

TABLE 99-4 CONDITIONS THAT CAUSE PLEURAL EFFUSION

TRANSUDATES
Heart failure
Nephrotic syndrome
Hepatic hydrothorax
Superior vena cava syndrome
Peritoneal dialysis
Atelectasis
Urinothorax
EXUDATES
Parapneumonic effusions
Simple
Complicated
Empyema
Other infections
Tuberculosis
Fungal
Parasites
Nocardia
Esophageal rupture
Malignancy
Carcinoma
Lymphoma
Mesothelioma
Metastatic disease
INFLAMMATORY DISORDERS
Connective tissue disease
Rheumatoid arthritis
Systemic lupus erythematosus
Churg-Strauss syndrome
Wegener granulomatosis
Familial Mediterranean fever
Abdominal disease
Subdiaphragmatic abscess (hepatic, splenic)
Pancreatitis, pancreatic pseudocyst
Postoperative
Iatrogenic
Drug induced
Misplacement of enteral feeding tube
Esophageal endoscopic interventions
Miscellaneous
Asbestos exposure
Atelectasis
Cholesterol effusion
Chylothorax
Dressler syndrome
Meigs syndrome
Pulmonary embolus
Radiation
Sarcoidosis
Trapped lung
Uremia
Yellow nail syndrome

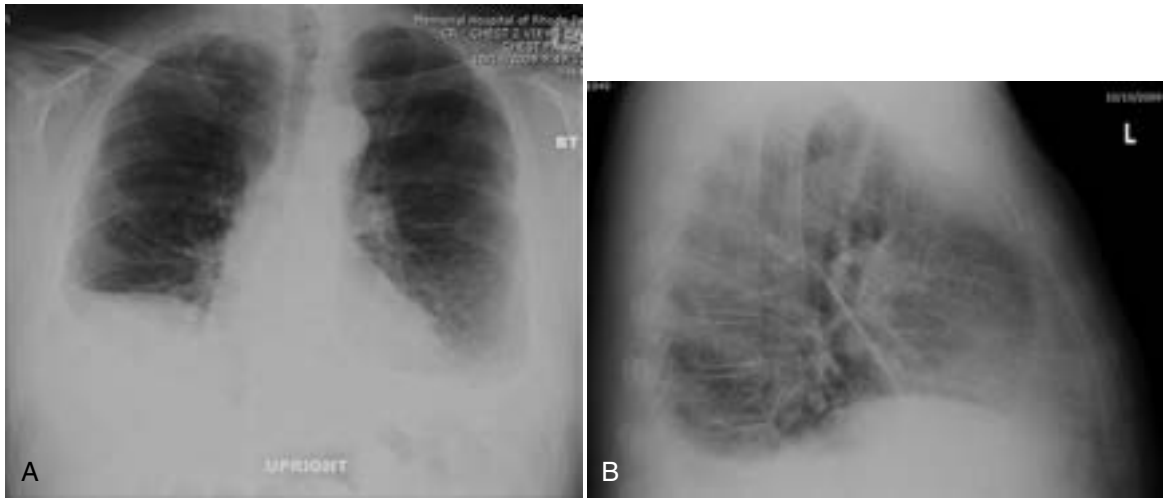

FIGURE 99-3. Patient with bilateral pleural effusions as seen on the posteroanterior radiograph of the chest (A) and lateral radiograph of the chest (B).

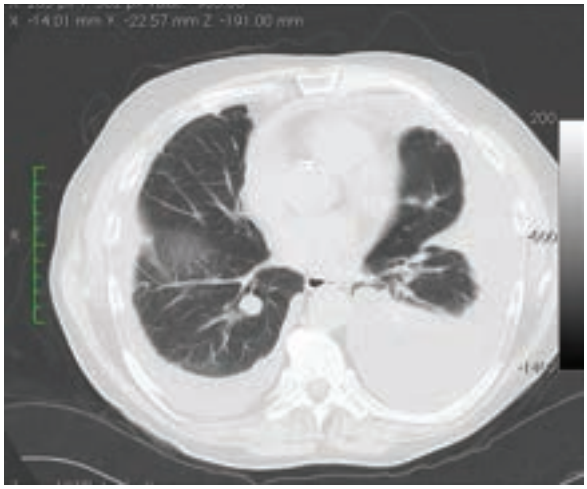

FIGURE 99-4. Chest computed tomography showing pleural effusion in the same patient as in Figure 99-3.

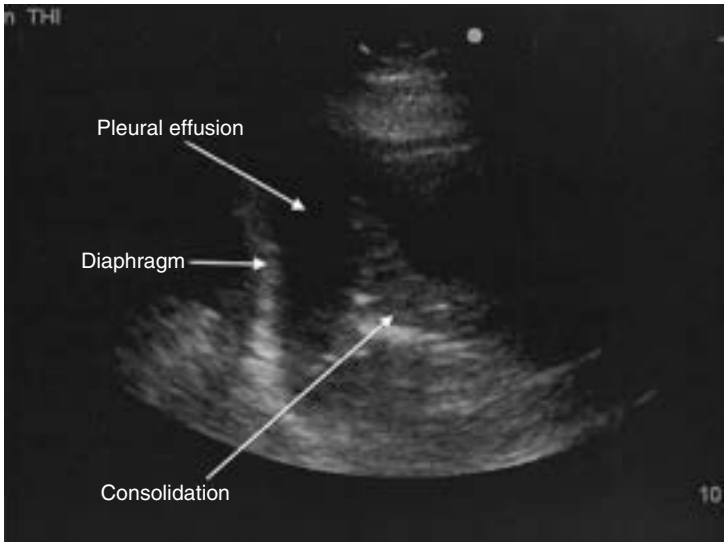

FIGURE 99-5. A right-sided pleural effusion as seen on ultrasound. The ultrasound can be performed at the bedside to aid in needle placement for thoracentesis.

pulmonary edema. Most thoracenteses can be performed at the bedside, using ultrasound guidance to enhance the procedure's safety. Relative contraindications to a diagnostic thoracentesis include a bleeding diathesis, a very small volume of pleural fluid, and a low benefit-to-risk ratio.

After fluid is obtained, a definitive diagnosis may be achieved, and the fluid can be classified as either a transudate or exudate (Table 99-5). To differentiate an exudate from a transudate, the pleural fluid needs to be analyzed for protein and lactate dehydrogenase (LDH). Simultaneous serum values of protein and LDH also need to be obtained. A pleural fluid exudate is characterized by a pleural fluid–to–serum protein ratio greater than 0.5, a pleural fluid–to–serum LDH ratio greater than 0.6, and a pleural fluid LDH greater than two thirds the normal serum value for LDH. An exudate can also be defined if the pleural fluid cholesterol is higher than 45 mg/dL.

TABLE 99-5 PLEURAL FLUID CHARACTERISTICS OF EXUDATES

LIGHT'S CRITERIA	
Protein	>0.5 pleural fluid/serum value
LDH	>0.6 pleural fluid/serum value
LDH	>⅔ upper limit of normal serum value

LDH = lactate dehydrogenase.

Transudates

Effusions that accumulate owing to changes in osmotic and hydrostatic forces usually form transudates. Transudative effusions are most commonly due to heart failure, in which the effusions are often bilateral or, if unilateral, preferentially involve the right hemithorax. Effusions caused by heart failure are usually related to elevated left and right heart pressures (Chapter 58), although right heart failure alone (such as seen in advanced pulmonary arterial hypertension) may rarely cause an effusion. Transudates may also be seen in cirrhosis (Chapter 153), nephrotic syndrome (Chapter 121), myxedema (Chapter 226), pulmonary embolism (Chapter 98), superior vena caval obstruction, and peritoneal dialysis (Chapter 131). With cirrhosis, ascites may cross from the peritoneum into the pleural space through small defects in the diaphragm (hepatic hydrothorax) (see Table 99-4). Unusual causes of transudates include peritoneal dialysis and atelectasis. Although malignancy typically causes an exudate, it can occasionally produce a transudate. Urinothorax, which is a rare cause of transudate, results from obstruction of the urinary system.

Exudates

Exudative effusions, which occur because of an alteration in vascular permeability and/or pleural fluid resorption, can be seen in inflammatory states, infection, or neoplasm. An effusion is characterized as an exudate if it meets one of the following criteria: pleural fluid–to–serum protein ratio higher than 0.5, pleural fluid–to–serum LDH ratio higher than 0.6, or pleural fluid LDH concentration higher than two thirds the normal serum value. When all three criteria are met, the sensitivity, specificity, and positive predictive value exceed 98% for defining an exudative effusion. Cholesterol levels also may be increased in exudates (>45 mg/dL). The pleural fluid analysis helps distinguish among the causes of pleural exudates (Table 99-6).

Parapneumonic Effusions

Parapneumonic effusions, which are the most common type of exudative pleural effusion, occur in up to 40% of patients with pneumonia. They

TABLE 99-6 CORRELATION OF THE CHARACTERISTICS OF PLEURAL EXUDATES WITH SPECIFIC DISEASE

TEST	DISEASE
pH < 7.2	Empyema, malignancy, esophageal rupture; rheumatoid, lupus, and tuberculous pleuritis
Glucose (<60 mg/dL)	Infection, rheumatoid pleurisy, tuberculous and lupus effusions, esophageal rupture
Amylase (>200 μg/dL)	Pancreatic disease, esophageal rupture, malignancy, ruptured ectopic pregnancy
RF, ANA, LE cells	Collagen vascular disease
↓ Complement	SLE, RA
RBCs (>5000/μL)	Trauma, malignancy, pulmonary embolus
Chylous effusion (triglycerides > 110 mg/dL)	Tuberculosis, disruption of thoracic duct (trauma, malignancy)
Cytology or biopsy (+)	Malignancy
ADA (>50 μg/L)	Tuberculosis

ADA = adenosine deaminase; ANA = antinuclear antibody; RA = rheumatoid arthritis; RBC = red blood cell; RF, rheumatoid factor; SLE = systemic lupus erythematosus.

typically occur in patients with bacterial pneumonia (Chapter 97) and can be classified as uncomplicated or complicated. With uncomplicated parapneumonic effusion, the pH is generally greater than 7.3, the glucose content more than 60 mg/dL, and the pleural fluid LDH less than 1000 IU/L. Elevated pleural fluid levels of proinflammatory markers, such as interleukin-1β (IL-1β) and IL-1 receptor antagonist, show promise in distinguishing complicated from uncomplicated effusions. Uncomplicated parapneumonic effusions are usually free flowing, do not require drainage, and will respond to the same antibiotic therapy as the pneumonia itself. Uncomplicated parapneumonic effusions can, however, rapidly transition to complicated effusions, sometimes within 24 hours. Complicated effusions, which are characterized by a pH of less than 7.2 and which often have a glucose content of less than 60 mg/dL, generally will not respond to antibiotic therapy alone but require drainage to prevent formation of an empyema, cutaneous fistulas, bronchopleural fistulas, or a thick pleural peal (fibrothorax).

Empyema

An empyema is present when frank pus is aspirated from the pleural space or when the Gram stain of the fluid is positive for bacteria or bacteria are cultured from the fluid. Pneumonia due to *Streptococcus pneumoniae* (Chapter 289) or *Staphylococcus aureus* (Chapter 288) infection can cause empyema. Individuals who aspirate are at high risk for empyema caused by anaerobic organisms, and patients with tuberculosis (Chapter 324) can develop a tuberculous empyema. Methicillin-resistant staphylococci, *Klebsiella* species pneumonia, and *Pseudomonas* species infections may cause an empyema that is difficult to treat. Uncommon infectious causes of effusions include *Actinomyces* species (Chapter 329), *Nocardia* species (Chapter 330), amebiasis (Chapter 352), *Echinococcus* species (Chapter 354), and paragonimiasis (Chapter 356).

Individuals with empyema often complain of pleuritic chest pain and have refractory fevers several days or more into the course of their pneumonia, but immunocompromised patients may develop empyema sooner and more rapidly.

Tuberculous Effusions

Tuberculosis (Chapter 324) can cause pleural effusion in up to 30% of patients who reside in locations endemic for tuberculosis. The pleural effusion typically is not due to direct mycobacterial infection but rather to increased vascular permeability of the pleural membrane because of a hypersensitivity reaction to mycobacterial proteins. The pleural fluid is generally lymphocyte predominant and culture negative for acid-fast bacilli. Adenosine deaminase levels higher than 50 U/L may be helpful in identifying tuberculous pleural effusions. A tuberculous empyema, which is distinct from a tuberculous effusion, is characterized by direct extension of the infection from thoracic lymph nodes or hematogenous spread of tuberculosis into the pleural space.

Malignancy

Malignant effusions are the second most common cause of exudative pleural effusions and may be due to seeding of the parietal or visceral pleura with malignant cells that change vascular permeability and/or impede resorption. Tumor cells also may amplify the production of mediators that promote leakage of fluid into the pleural space.[8] Rarely, a malignant effusion may be transudative. Lung cancer (Chapter 191) is the most frequent cause of malignant pleural effusion, and other malignancies that can involve the pleural space include breast cancer (Chapter 198), ovarian cancer (Chapter 199), gastric cancer (Chapter 192), and lymphoma (Chapters 185 and 186). When malignancy involves the pleural space, the prognosis is poor. However, the finding of a pleural effusion in an individual with underlying malignancy does not necessarily imply that there is a metastatic malignant process involving the pleural space. Benign effusions in patients with underlying malignancy may be due to atelectasis, postobstructive pneumonia, hypoalbuminemia, pulmonary emboli (Chapter 98), lymphatic obstruction, and complications from radiation (Chapter 20) or chemotherapy. For this reason, it is important to obtain a sample of pleural fluid in these individuals. The diagnosis of malignant pleural effusion is established by demonstrating malignant cells in the pleural fluid. Approximately 60% of malignant pleural effusions can be diagnosed with one thoracentesis, and the yield increases to 80% with repeat thoracenteses. If needed, a biopsy of the pleura may be useful in identifying the malignancy. Pleural biopsies are optimally obtained either by medical or surgical thoracoscopy (Chapter 101) rather than in a blind fashion (e.g., using a Cope or Abrams needle).

Systemic Inflammatory Disorders

Effusions may be seen in as many as 15% of patients with rheumatoid arthritis (Chapter 264), with a male preponderance to the development of effusions. Effusions typically appear within 5 years after the onset of disease but occasionally occur before the onset of joint disease. Rheumatoid factor in the pleural fluid is often greater than 1 : 320, and pleural fluid glucose is low (<60 mg/dL, or the pleural fluid–to–serum glucose ratio is <0.5). Other causes of low pleural fluid glucose include complicated parapneumonic effusions or empyema, malignant effusions, tuberculous pleurisy, lupus pleuritis, and esophageal rupture. Pleural effusions can be seen in 15 to 50% of patients with systemic lupus erythematosus (SLE). Lupus erythematosus cells, low levels of complement (C3 and C4), and pleural fluid antinuclear antibody titer of more than 1 : 160 can be seen in effusions due to SLE. Wegener granulomatosis, Sjögren syndrome, and sarcoidosis are less common causes of pleural effusions.

Pancreatitis

Patients with pancreatitis or pancreatic pseudocysts (Chapter 144) may develop exudative pleural effusions that often involve the left hemithorax. A pleural fluid amylase concentration that is greater than the upper limit of normal for serum amylase is consistent with acute or chronic pancreatitis as a cause of the effusion. Extremely high amylase levels have been reported in effusions due to pancreaticopleural fistulas. Amylase also may be seen in the pleural fluid with an esophageal rupture or malignancy. Pancreatic disease is associated with pancreatic isoenzyme amylase, whereas malignancy and esophageal rupture are characterized by a predominance of salivary amylase isoenzymes.

Chylothorax

A chylothorax has a milky-white appearance and is characterized by high levels of triglycerides (>110 mg/dL) and chylomicrons. A chylothorax is caused by leakage of lymph from the thoracic duct into the pleural space, most commonly related to mediastinal malignancy but also occurring after trauma to the thoracic duct. Major complications from a chylothorax are malnutrition and immunologic compromise when fat, protein, and lymphocytes are depleted by repeated thoracentesis or chest tube drainage. Chylous effusions must be distinguished from pseudochylous effusions, which have a white appearance but are devoid of chylomicrons, and are indicative of a chronic long-standing effusion.

Hemothorax

Blood in the thorax is easily recognized during a thoracentesis. A hemothorax has a pleural fluid hematocrit that is at least half that of the circulating hematocrit. By contrast, a bloody pleural effusion will appear red but have a lower hematocrit. A bloody effusion often suggests a malignant process. Other causes include trauma, pulmonary infarction (Chapter 98), tuberculosis (Chapter 324), collagen vascular disease (Chapters 264 and 266), and hematologic disorders. Generally, blood removed from the pleural space does not clot, whereas blood due to the trauma of the thoracentesis itself will clot

when collected. Because blood in the pleural space does not clot, it can be removed by lymphatics if the volume is small. Larger hemothoraces may require chest tube drainage.

Asbestos Exposure

Pleural effusion may occur after exposure to asbestos. The effusion is often small, unilateral, and serosanguineous, with fewer than 6000 cells/mL. The effusion tends to resolve within a year, resulting in pleural plaques that may calcify over time. Malignant mesothelioma should be excluded in these individuals.

Other Causes of Pleural Exudates

Meigs syndrome is the triad of ovarian tumor (Chapter 199), ascites, and pleural effusion. The effusion, which is usually large and on the right side, is formed when fluid moves from the abdomen to the pleural space through diaphragmatic defects. Meigs syndrome usually occurs in postmenopausal women and resolves following removal of the tumor.

Dressler syndrome may occur from 3 to 30 days after open heart surgery or myocardial infarction. The patient usually experiences pleuritic pain and has a small to moderate left-sided effusion. Treatment is as for the accompanying pericardial effusion (Chapter 77).

Uremia (Chapter 130), which can cause a polyserositis with effusion, must be distinguished from the transudate that commonly is seen in the nephrotic syndrome (Chapter 121). Subdiaphragmatic processes, such as hepatic or splenic abscesses (Chapter 151), may cause effusions. Trapped lung occurs when a lobe or segment is unable to re-expand owing to a restrictive visceral pleural peel or an endobronchial mass. Increased negative intrapleural pressures associated with trapped lung promote the formation of an effusion. A number of drugs, including amiodarone, bleomycin, dantrolene, hydralazine, isoniazid, methotrexate, methysergide, mitomycin, procainamide, and procarbazine, can cause pleural effusions. Treatment consists of discontinuing the offending agent, although treatment with oral corticosteroids may be needed (Chapter 254).

TREATMENT AND PROGNOSIS

Rx

Empyemas and complicated parapneumonic effusions require drainage by tube thoracostomy in concert with appropriate antibiotic therapy. When an empyema is present, it must be drained immediately using an indwelling chest tube. The combination of intrapleural t-PA (10 mg) and DNase (5 mg) therapy can improve chest tube drainage in patients with pleural infections.[■■] If chest tube drainage is unsuccessful in resolving the empyema, video-assisted thoracic surgery (Chapter 101) is preferred, with intrapleural streptokinase reserved for patients who are poor candidates for video-assisted thoracic surgery or are in situations in which it is not available. Occasionally, empyema requires thoracotomy and decortication.

Treatment options for malignant pleural effusion include observation, chemical pleurodesis with talc or tetracycline derivatives, and treatment of the underlying malignancy. For patients with recurrent malignant effusions, the tunneled placement of an indwelling pleural catheter can permit intermittent drainage to relieve symptoms.[9,10] In a randomized trial of patients with malignant pleural effusion and no previous pleurodesis, indwelling pleural catheters and talc pleurodesis were equivalent for relieving patient-reported dyspnea; catheter-based treatment shortened hospital length of stay but also produced more adverse effects.[■] A complete response occurs in perhaps 50% of patients. A low pleural fluid pH (<7.2) tends to impart a poor response to chemical pleurodesis. Malignant effusions with a tendency to respond to generalized chemotherapy include those related to breast cancer (Chapter 198) and small cell carcinoma of the lung (Chapter 191). Malignant effusions related to lymphomas and obstruction of lymphatic drainage of the pleural space may also respond to treatment of the underlying disease.

Treatment of inflammatory effusions centers on the use of anti-inflammatory agents and corticosteroids (Chapters 264 and 266). Treatment of chylous effusions may involve chest tube drainage, although fat malnutrition may ensue. Attempts to decrease chyle formation can be accomplished by intravenous hyperalimentation, decreased oral fat intake, and the intake of medium- and light-chain fatty acids, which are absorbed directly into the portal circulation. Ligation of the thoracic duct should be considered for traumatic chylous effusions. Thoracic duct embolization has been performed successfully for nontraumatic chylous effusion.[11]

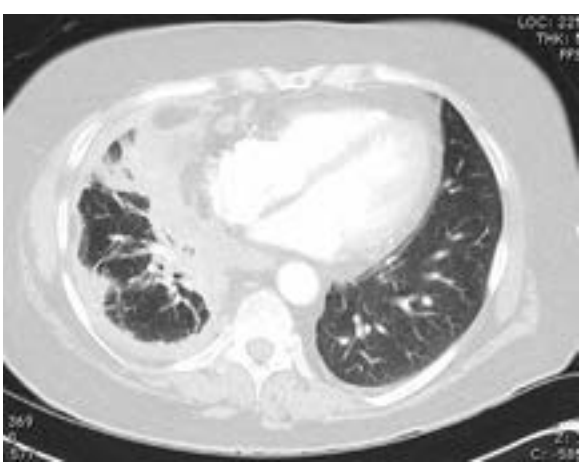

FIGURE 99-6. Computed tomographic angiogram of the chest in a patient with mesothelioma. There is diffuse thickening of the pleura and pericardium on the right with a rind-like appearance.

Mesothelioma

EPIDEMIOLOGY AND PATHOBIOLOGY

Malignant mesotheliomas (Chapter 191) are neoplasms arising from the serosal membrane of body cavities. Eighty percent of mesotheliomas originate in the pleural space, and most others arise from the peritoneum. Individuals are usually older than 55 years and often have a history of asbestos exposure in the distant past (frequently, 30 to 40 years ago). Smoking is not a risk factor for developing mesothelioma, but smoking in concert with asbestos exposure increases the risk for lung cancer. Approximately 3000 cases of mesothelioma occur in the United States each year. The annual incidence is decreasing in the United States owing to better control of occupational exposure, but it may still be increasing in other countries where there are fewer regulations.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients with mesothelioma often complain of dyspnea, weight loss, and pain. Malignant mesothelioma may present as an extremely large mass or pleural effusion occupying the entire hemithorax at times. Chest CT may demonstrate either localized or circumferential pleural thickening associated with various amounts of calcified pleural plaque (Fig. 99-6). Elevated levels of hyaluronic acid in the pleural fluid may be seen with mesotheliomas, but pleural fluid cytology frequently is insufficient for diagnosis. The most efficient way of obtaining a diagnosis is by CT-guided core biopsy or thoracoscopy. Special stains and electron microscopy of biopsy tissue may help to make the difficult distinction between metastatic adenocarcinoma and mesothelioma. The use of biomarkers such as fibulin-3 in plasma and pleural effusions may aid in detecting mesothelioma at an earlier stage and in distinguishing it from other malignancies involving the pleura.

A fraction of mesotheliomas are benign. Benign mesotheliomas are usually large and often pedunculated at the time of diagnosis.

TREATMENT AND PROGNOSIS

Rx

Unfortunately, no particular therapy, surgical or chemotherapy, has met with great success in malignant mesothelioma, and median survival is only 8 to 12 months after diagnosis.[■] Pleurodesis with talc or pleurectomy is usually performed for palliation and control of symptoms related to any pleural effusion. In highly selected patients with localized disease and no comorbid illnesses, surgical resection or radical extrapleural pneumonectomy may be attempted. Radiation therapy can provide symptom palliation and has been used following attempts at curative surgery. Treatment of benign mesothelioma involves surgical resection.

Pneumothorax

Pneumothorax refers to the accumulation of air in the pleural space (see Fig. 84-15). Normally, the pressure within the pleural space is slightly subatmospheric. However, when more than a very small amount of air accumulates within the pleural space, pressure within it becomes positive, and there is

compression of underlying lung. Typically, the visceral pleura separates from the parietal pleura, and air can be seen between the visceral pleural lining and the rib cage.

EPIDEMIOLOGY AND PATHOBIOLOGY

Pneumothorax is often associated with blunt or penetrating trauma. With penetrating trauma, air may leak into the pleural space through the injured chest wall or into the pleural space from the injured lung. Patients with underlying lung disease undergoing mechanical ventilation may acutely develop a pneumothorax when high local pressures disrupt lung tissue, thereby leading to a leak (Chapter 105).

Pneumothorax also may occur spontaneously or be secondary to underlying lung disease. Typically, spontaneous pneumothorax occurs in tall, young, thin men, presumably as a result of rupture of preexisting apical blebs. Diseases that are associated with pneumothorax include emphysema (Chapter 88), cystic fibrosis (Chapter 89), granulomatous inflammation, necrotizing pneumonia, pulmonary fibrosis (Chapter 92), eosinophilic granulomatous disease, sarcoidosis (Chapter 95), and endometriosis (Chapter 236). Catamenial pneumothorax occurs in patients who have subpleural and diaphragmatic endometriosis (Chapter 236); rupture of the endometrial nodules at the time of menstruation causes pneumothorax.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Symptoms typically include acute shortness of breath and sharp chest pain. Physical examination is characterized by tachycardia, decreased breath sounds, decreased tactile fremitus, a pleural friction rub, subcutaneous emphysema, hyper-resonance to percussion, and a tracheal shift toward the uninvolved hemithorax.

Diagnosis can be made by obtaining an upright chest radiograph, and rapid assessment with point-of-care ultrasound is increasingly being used as well. Air within the pleural space appears as an area of lucency on the chest radiograph (Fig. 99-7). With a small pneumothorax, the lucency is best appreciated at the lung apex when the patient is upright. An end-expiratory radiograph is particularly helpful in diagnosing a small pneumothorax. During expiration, the density of the lung will increase because of a reduction in volume, thereby highlighting the difference between lung parenchymal and pleural gas. When a portable chest radiograph is obtained with the patient in the supine position, such as a patient in an intensive care unit, the lucent area may be most noticeable over the lower rib cage (superior sulcus sign).

A *tension pneumothorax* is defined as a pneumothorax associated with a mediastinal shift and hemodynamic compromise, usually because high intrathoracic pressures compress the vena cava and atrium. This physiology implies an ongoing leak of air into the pleural space without opportunity for the air to escape.

TREATMENT AND PROGNOSIS

Rx

If the pneumothorax is small and the patient is not in distress, tube thoracostomy is not needed, and observation alone may be sufficient.[12] However, if a patient is symptomatic, the pneumothorax occupies more than 50% of the hemithorax, or a tension pneumothorax develops, management requires insertion of a thoracostomy tube, suction, and water-seal drainage. If there is a continuing leak despite tube thoracostomy, a bronchopleural fistula may be suspected. In this instance, chemical pleurodesis or surgical correction, usually by video-assisted thoracoscopic surgery, may be necessary (Chapter 101). In a randomized trial of patients with primary spontaneous pneumothorax, simple aspiration and drainage followed by 300 mg of minocycline pleurodesis was a safe and effective treatment, with a 30% 1-year recurrence rate, compared with a 50% 1-year recurrence rate for simple aspiration and drainage only.

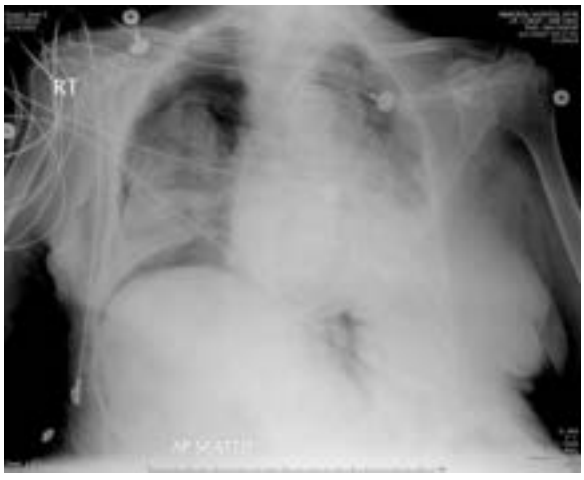

FIGURE 99-7. A portable anteroposterior chest radiograph demonstrating a right-sided pneumothorax. Note that the right lung has collapsed to less than half the size of the right hemithorax. In addition, there is a pneumoperitoneum best seen as a collection of air under the right hemidiaphragm.

MEDIASTINUM

The mediastinum, which is the central part of the thoracic cavity, lies between the right and left lungs. It contains the heart and aorta, esophagus, trachea, lymph nodes, thymus, and great vessels. It is bordered by the two pleural cavities laterally, the diaphragm inferiorly, and the thoracic inlet superiorly.

Mediastinal Masses

PATHOBIOLOGY

For clinical purposes, the mediastinum has been divided into three compartments: anterior, middle, and posterior (E-Fig. 99-E3). The anterior compartment contains the thymus, substernal extensions of the thyroid, blood vessels, pericardium, and lymph nodes. The middle compartment contains the heart, great vessels, trachea, main bronchi, lymph nodes, and the phrenic and vagal nerves. The posterior compartment contains the vertebrae, descending aorta, esophagus, thoracic duct, azygous and hemizygous veins, lower portion of the vagus, sympathetic chain, and lymph nodes.

CLINICAL MANIFESTATIONS

Mediastinal masses usually are not accompanied by symptoms, and most masses are incidentally found on either a chest radiograph or a chest CT scan. If present, however, symptoms include chest pain, cough, hoarseness, stridor, dysphasia, and dyspnea. One third of patients with a mediastinal thymoma have symptoms or weakness owing to myasthenia gravis (Chapter 422), and individuals with mediastinal lymphoma (Chapters 185 and 186) may have systemic symptoms such as fever, night sweats, and weight loss. Occasionally, a mass may compress the superior vena cava, causing partial obstruction and resulting in facial edema and dilated neck and chest veins (Fig. 99-8).

DIAGNOSIS

When a mass is identified by chest radiography or chest CT, further evaluation is mandatory. If a benign process is suspected, follow-up CT may be indicated. If a malignant process is suspected, radiologic evaluation may include angiography, positron emission tomography, or MRI. Biopsies can be obtained either with mediastinoscopy or mediastinotomy. Less invasive approaches include ultrasound-guided transbronchial needle aspiration biopsy, and direct sampling by transthoracic CT-guided needle aspiration may be useful in evaluating anterior or posterior mediastinal masses. The evaluation and differential diagnosis of mediastinal masses are guided by the compartment in which they arise (Table 99-7). The *anterior mediastinal compartment* includes lesions such as thymomas, germ cell tumors (teratomas), lymphomas, and intrathoracic thyroid tissue.[13] Thymomas make up about

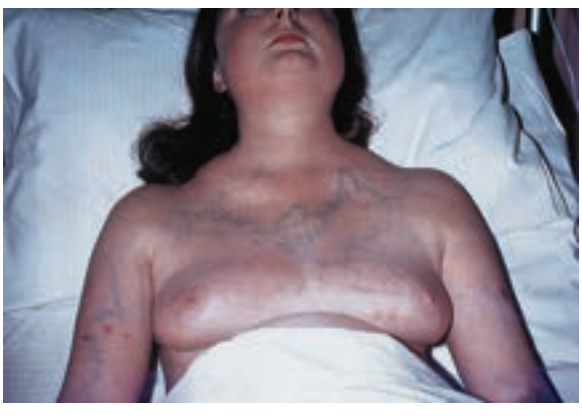

FIGURE 99-8. Superior vena cava obstruction in bronchial carcinoma. Note the swelling of the face and neck and the development of collateral circulation in the veins of the chest wall.

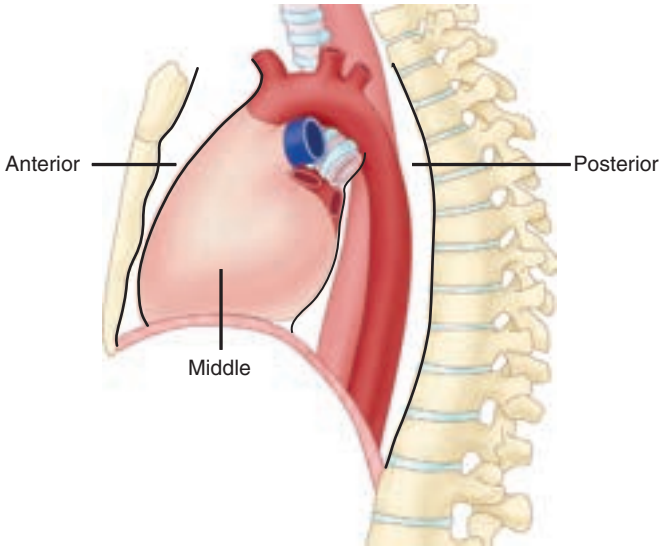

E-FIGURE 99-3. Depiction of the mediastinal compartments. The anterior compartment is bounded posteriorly by the pericardium, ascending aorta, and brachiocephalic vessels and anteriorly by the sternum. The middle compartment extends from the posterior limits of the anterior compartment to the posterior pericardial line. The posterior compartment extends from the pericardial line to the dorsal chest wall.

TABLE 99-7 CAUSES OF MEDIASTINAL MASSES

ANTERIOR	MIDDLE	POSTERIOR
Teratoma	Pericardial cyst	Neurogenic tumor
Thymoma	Lymph node hyperplasia	Esophageal tumor
Thyroid tumor	Bronchogenic tumor	Bronchogenic tumor
Goiter	Bronchogenic cyst	Bronchogenic cyst
Aneurysm	Aneurysm	Aneurysm
Lymphoma	Lymphoma	Lymphoma
Parathyroid tumor		Meningocele
Lipoma		Enteric cyst
Morgagni diaphragm hernia		Esophageal diverticula
		Bochdalek diaphragm hernia

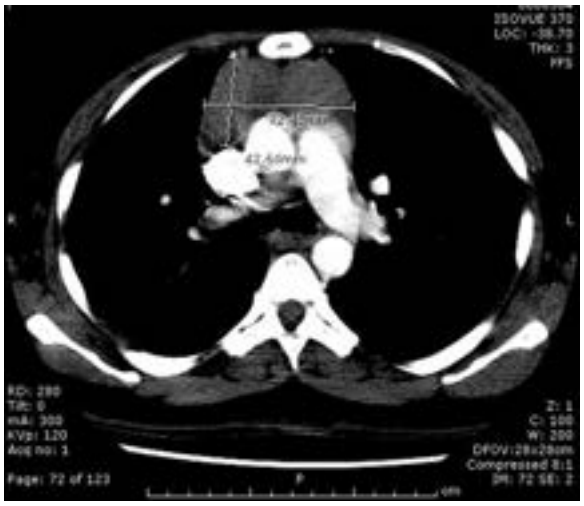

FIGURE 99-9. Chest computed tomography of a patient with an anterior mediastinal mass that proved to be a substernal goiter.

20% of mediastinal neoplasms in adults, in whom they are the most common anterior mediastinal primary neoplasm. Patients with systemic lymphoma often have involvement of the mediastinum, but only 5 to 10% of patients with lymphoma present with primary mediastinal lesions. When lung cancer (Chapter 191) presents with mediastinal adenopathy, it is at an advanced stage. Teratomas, which account for 10% of mediastinal tumors, may contain squamous cells, hair follicles, sweat glands, cartilage, and linear calcifications; about one third are malignant. Intrathoracic goiters (Chapter 226) (Fig. 99-9) may compress the trachea and cause stridor, cough, dyspnea, and, occasionally, superior vena cava obstruction. Anterior masses in the right cardiophrenic angle, which are rare and may be associated with pericardial defects or obesity, may be due to herniation of liver or intestinal contents through the foramina of Morgagni.

In the posterior mediastinum, neurogenic tumors are the most common lesions. Many of these tumors are benign and originate in the nerve sheath or sympathetic ganglion cells (ganglioneuroma). Posterior mediastinal masses also include cysts, meningocele, lymphoma (Chapter 185 and 186), aortic aneurysm (Chapter 78), and esophageal disorders (Chapter 138) such as diverticula and neoplasm. Herniation of abdominal contents into the thorax can result in posterior mediastinal masses. Herniation of abdominal contents through the foramina of Bochdalek results in a mass in the posterolateral area of the diaphragm, usually on the left side; it is the most common congenital hernia and may contain spleen or kidney. Herniation of the stomach through the esophageal hiatus (Chapter 138), which is the most common type of diaphragmatic herniation, produces a mass posterior to the heart, often with an air-fluid level.

Benign cysts can occur in the anterior, middle, or posterior compartments. They can arise in the pericardium, bronchi, thymus, thoracic duct, esophagus, and stomach and can produce compressive symptoms. Pericardial cysts, which are often located in the cardiophrenic angle, contain clear liquid. Bronchogenic cysts occur in the middle or posterior compartments and are filled with liquid and lined with respiratory epithelium; they often develop around the paratracheal area or carina and do not communicate with the tracheal bronchial tree.

TREATMENT

Rx

The treatment of a mediastinal mass depends on the underlying pathology. Some lesions, such as thymomas, teratomas, cysts, neurogenic tumors, and hernias, require surgical resection. Others, such as lymphoma, are treated with radiation or chemotherapy. Some can be carefully monitored over time.

Mediastinitis

Acute mediastinitis is most commonly caused by bacterial infection. Mediastinal infections are most commonly seen as complications after cardiothoracic surgical procedures, such as sternotomy, or procedures involving the esophagus or tracheobronchial tree. Rupture of the esophagus or trachea from trauma or tissue necrosis can cause mediastinitis. Treatment of acute mediastinitis requires antibiotics, pleural drainage, and evacuation of necrotic tissue. Chronic mediastinitis (fibrosing mediastinitis) is a progressive illness that may be idiopathic or can be caused by granulomatous infection, fungus, neoplasm, radiotherapy, or drugs (such as methysergide). Patients with chronic mediastinitis remain asymptomatic until vascular or neurologic structures are affected. When respiratory structures are involved, tracheobronchial narrowing is the most common presentation. The diagnosis and treatment often require surgical exploration, and further treatment is often not necessary unless it is due to tuberculosis (Chapter 324) or a fungal infection (Chapter 331).

Pneumomediastinum

Pneumomediastinum occurs when air infiltrates the mediastinal structures after a rupture of the esophagus, trachea, or lung dissects into the mediastinum. Loss of esophageal or tracheal integrity often results from trauma, whereas leaks from alveoli may result from trauma, occur spontaneously, or can be a complication of mechanical ventilation (Chapter 105). Pneumomediastinum rarely is seen as a complication of an asthma exacerbation (Chapter 87), violent coughing, or emesis. The diagnosis can be made by seeing thin columns of hyperlucency between mediastinal structures on chest radiography or CT scan. Pneumomediastinum may present as a sore throat, neck pain, or shortness of breath. Often, the mediastinal air dissects into the subcutaneous tissues of the neck and chest wall, where it results in subcutaneous emphysema. A characteristic crepitus is palpable when subcutaneous emphysema is present. A rare tension pneumomediastinum may compress the right ventricle. Spontaneous pneumomediastinum generally resolves without treatment. When more severe collections of subcutaneous air occur, surgical decompression is often warranted.

Grade A References

A1. Cejudo P, Lopez-Marquez I, Lopez-Campos JL, et al. Exercise training in patients with chronic respiratory failure due to kyphoscoliosis: a randomized controlled trial. *Respir Care*. 2014;59:375-382.

A2. Weinstein SL, Dolan LA, Wright JG, et al. Effects of bracing in adolescents with idiopathic scoliosis. *N Engl J Med*. 2013;369:1512-1521.

A3. Slobogean GP, MacPherson CA, Sun T, et al. Surgical fixation vs nonoperative management of flail chest: a meta-analysis. *J Am Coll Surg*. 2013;216:302-311.

A4. Kjeken I, Bo I, Ronningen A, et al. A three-week multidisciplinary in-patient rehabilitation programme had positive long-term effects in patients with ankylosing spondylitis: randomized controlled trial. *J Rehabil Med*. 2013;45:260-267.

A5. Rahman NM, Maskell NA, West A, et al. Intrapleural use of tissue plasminogen activator and DNase in pleural infection. *N Engl J Med*. 2011;365:518-526.

A6. Nie W, Liu Y, Ye J, et al. Efficacy of intrapleural instillation of fibrinolytics for treating pleural empyema and parapneumonic effusion: a meta-analysis of randomized control trials. *Clin Respir J*. 2014;8:281-291.

A7. Davies HE, Mishra EK, Kahan BC, et al. Effect of an indwelling pleural catheter vs chest tube and talc pleurodesis for relieving dyspnea in patients with malignant pleural effusion: the TIME2 randomized controlled trial. *JAMA*. 2012;307:2383-2389.

A8. Muers MF, Stephens RJ, Fisher P, et al. Active symptom control with or without chemotherapy in the treatment of patients with malignant pleural mesothelioma (MS01): a multicentre randomised trial. *Lancet*. 2008;371:1685-1694.

A9. Chen JS, Chan WK, Tsai KT, et al. Simple aspiration and drainage and intrapleural minocycline pleurodesis versus simple aspiration and drainage for the initial treatment of primary spontaneous pneumothorax: an open-label, parallel-group, prospective, randomised, controlled trial. *Lancet*. 2013;381:1277-1282.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. McCool FD, Tzelepis GE. Dysfunction of the diaphragm. *N Engl J Med*. 2012;366:932-942.
2. Hresko MT. Clinical practice: idiopathic scoliosis in adolescents. *N Engl J Med*. 2013;368: 834-841.
3. Erken HY, Burc H, Saka G, et al. Disagreements in surgical planning still exist between spinal surgeons in adolescent idiopathic scoliosis: a multisurgeon assessment. *Eur Spine J*. 2014;23: 1258-1262.
4. Vana PG, Neubauer DC, Luchette FA. Contemporary management of flail chest. *Am Surg*. 2014;80:527-535.
5. Dehghan N, de Mestral C, McKee MD, et al. Flail chest injuries: a review of outcomes and treatment practices from the National Trauma Data Bank. *J Trauma Acute Care Surg*. 2014;76:462-468.
6. Porcel JM, Light RW. Pleural effusions. *Dis Mon*. 2013;59:29-57.
7. Wilcox ME, Chong CA, Stanbrook MB, et al. Does this patient have an exudative pleural effusion? The Rational Clinical Examination systematic review. *JAMA*. 2014;311:2422-2431.
8. Stathopoulos GT, Kalomenidis I. Malignant pleural effusion: tumor-host interactions unleashed. *Am J Respir Crit Care Med*. 2012;186:487-492.
9. Van Meter MEM, McKee KY, Kohlwes RJ. Efficacy and safety of tunneled pleural catheters in adults with malignant pleural effusions: a systematic review. *J Gen Intern Med*. 2011;26:70-76.
10. Myers R, Michaud G. Tunneled pleural catheters: an update for 2013. *Clin Chest Med*. 2013;34: 73-80.
11. Nadolski GJ, Itkin M. Thoracic duct embolization for nontraumatic chylous effusion: experience in 34 patients. *Chest*. 2013;143:158-163.
12. Huang Y, Huang H, Li Q, et al. Approach of the treatment for pneumothorax. *J Thorac Dis*. 2014;6:S416-S420.
13. Shahrzad M, Le TS, Silva M, et al. Anterior mediastinal masses. *Am J Roentgenol*. 2014;203:W128-W138.

REVIEW QUESTIONS

1. A patient complains of orthopnea and dyspnea when bending. What is the best test to evaluate the possibility of bilateral diaphragm paralysis (BDP)?

A. Obtain a chest radiograph
B. Measure vital capacity
C. Measure upright and supine vital capacity
D. Measure transdiaphragmatic pressure
E. Perform a sniff test with fluoroscopic imaging

Answer: D The diagnosis of BDP can be confirmed by measurements of transdiaphragmatic pressure (Pdi). There is no change in Pdi during inspiration with bilateral diaphragm paralysis. Vital capacity is reduced with BDP and is further reduced by 30 to 50% when the patient is in the supine position; however, these findings are not specific for BDP. Chest radiographs and sniff tests are not useful when diagnosing BDP. Unilateral diaphragm paralysis may be present when there is an elevated hemidiaphragm on a chest radiograph. The diagnosis of unilateral paralysis can be confirmed by performing a "sniff test" using fluoroscopy or ultrasound: paradoxical (cephalad) movement of the hemidiaphragm dome occurs during a sniff maneuver. An alternate diagnostic test for BDP or unilateral paralysis is performing two-dimensional ultrasound of the diaphragm in the zone of apposition. The diaphragm does not thicken with inspiration when the diaphragm is paralyzed.

2. Which disease of the chest wall causes the greatest restrictive impairment?

A. Morbid obesity
B. Ankylosing spondylitis
C. Kyphoscoliosis
D. Pectus excavatum

Answer: C Kyphoscoliosis produces the most severe restrictive impairments. Total lung capacity and vital capacity may be reduced to as low as 30% of predicted values. The restriction is more severe as the degree of spinal angulation increases and when there is respiratory muscle weakness. Morbid obesity may be accompanied by mild to moderate restriction when the patient's body mass index is greater than 40 kg/m^2. Ankylosing spondylitis and pectus excavatum result in little or no restriction.

3. A patient is admitted with pneumonia and a large pleural effusion. Which feature is most consistent with the presence of a complicated parapneumonic effusion?

A. Free-flowing pleural fluid on computed tomography of the chest
B. Pleural fluid–to–serum protein ratio > 0.5
C. Pleural fluid pH < 7.2
D. Pleural fluid glucose > 60 mg/dL
E. Pleural fluid–to–serum lactate dehydrogenase (LDH) ratio higher than 0.6,

Answer: C Complicated effusions are characterized by a pH of less than 7.2 and often have a low glucose content (<60 mg/dL). These effusions often require drainage to prevent formation of an empyema, cutaneous fistulas, bronchopleural fistulas, or fibrothorax. A pleural fluid–to–serum protein ratio of more than 0.5, a pleural fluid–to–serum LDH ratio of greater than 0.6, or a pleural fluid LDH concentration higher than two thirds the normal serum value indicates the presence of an exudate. Complicated parapneumonic effusions are exudates, but exudative effusions can be observed in inflammatory states or neoplasms and are not specific for infection. Parapneumonic effusions, which are the most common type of exudative pleural effusion, can occur in up to 40% of patients with pneumonia. An empyema is present when frank pus is aspirated from the pleural space, when Gram stain of the fluid is positive for bacteria, or when bacteria are cultured from the fluid.

4. A 26-year-old nonsmoking man presents with fever and chills. A chest radiograph demonstrates no infiltrate, but a widened mediastinum is noted. Computed tomography shows a middle mediastinal mass. The most likely diagnosis is which of the following?

A. Thymoma
B. Lymphoma
C. Lung cancer
D. Schwannoma

Answer: B Classifying the location of mediastinal masses as anterior, middle, or posterior mediastinum aids in the differential diagnosis. A middle mediastinal mass accompanied by systemic symptoms such as fever, night sweats, and weight loss is most consistent with lymphoma. Lung cancer with metastasis to mediastinal nodes is possible, but the patient's age and lack of smoking history make this diagnosis unlikely. Thymomas, germ cell tumors (teratomas), and thyroid tissue usually involve the anterior comportment. Neurogenic tumors, such as schwannomas, are typically seen in the posterior compartment. Mediastinal masses may be accompanied by symptoms including chest pain, cough, hoarseness, stridor, dysphasia, and dyspnea, but most often patients with mediastinal masses are asymptomatic.

5. The following statements are true regarding pneumothorax except which one?

A. Pleural pressure becomes more subatmospheric, thereby resulting in compression of the underlying lung.
B. A tension pneumothorax produces a mediastinal shift and hemodynamic compromise.
C. Diagnosis can be made by obtaining an upright chest radiograph or a point-of-care ultrasound.
D. Tube thoracostomy and suction followed by water-seal drainage are not always needed.
E. Pneumothorax may occur spontaneously in individuals without a history of underlying lung disease or trauma.

Answer: A Normally, the pressure within the pleural space is slightly subatmospheric. However, when more than a very small amount of air accumulates within the pleural space, pressure within the pleural space becomes positive, and underlying lung is compressed. Pneumothorax is often associated with blunt or penetrating trauma, but it may occur spontaneously in some patients (typically tall, young, thin men) without underlying lung disease. Air within the pleural space appears as an area of lucency on the chest radiograph or lack of the "sliding lung sign" on ultrasound. A tension pneumothorax produces mediastinal shift and requires immediate tube thoracostomy. If the pneumothorax is small and the patient is not in distress, tube thoracostomy is not needed, and observation alone may be sufficient.

100 OBSTRUCTIVE SLEEP APNEA

ROBERT C. BASNER

DEFINITION

Obstructive sleep apnea is a chronic condition of cyclic obstruction of the upper airway during sleep, characteristically combined with associated symptoms or signs of disturbed sleep (Chapter 405), the most common being excessive daytime sleepiness and loud snoring. A frequency of at least five obstructive events (apneas, hypopneas, and/or respiratory effort–related arousals; see Pathogenesis) per hour of sleep is a minimal criterion for diagnosing obstructive sleep apnea in adults, although a higher frequency of obstructive events is more consistently correlated with an increased risk for cardiovascular and neurocognitive disorders. For continuous positive airway pressure (CPAP) treatment to be covered by Medicare in the United States, patients must be documented to have at least 15 such events per hour, or to have between 5 and 14 events per hour and also have at least one of the following: documented symptoms of excessive daytime sleepiness, impaired cognition, mood disorders, or insomnia; or documented hypertension, ischemic heart disease, or history of stroke.

EPIDEMIOLOGY

Obstructive sleep apnea is underrecognized by clinicians and underreported by patients, and most adults with moderate to severe disease remain undiagnosed. In adults, obstructive sleep apnea is characteristically found in overweight persons, although other anatomic (primarily upper airway and craniofacial) and ventilatory abnormalities may also predispose to the disorder. The presence of at least five obstructive events per hour of sleep has been found in 9 to 28% of adults without specific risk factors for, or symptoms of, obstructive sleep apnea, whereas the prevalence of such obstructive events and associated excessive daytime sleepiness is closer to 3 to 7% for adult men and 2 to 5% for adult women. Although overall prevalence appears to be much greater in men than in women, postmenopausal and obese women are at increased risk. Populations with a particularly high prevalence of obstructive sleep apnea include people older than 60 years of age and patients with systemic hypertension, particularly poorly controlled hypertension (Chapter 67); prior strokes (Chapter 407); heart failure (Chapter 58); atrial fibrillation (AF; Chapter 64), particularly recurrent AF following electrical cardioversion; prior acute myocardial infarction (Chapter 73); obesity-hypoventilation syndrome (Chapters 86 and 220); metabolic syndrome (Chapter 229); idiopathic pulmonary fibrosis (Chapter 92); and medically refractory epilepsy (Chapter 403). African Americans, particularly those younger than 25 and older than 65 years, and adult Asians have a higher incidence and/or greater severity of obstructive sleep apnea than whites, whereas Hispanic adults have a higher prevalence of snoring compared with whites. The presence of obstructive sleep apnea in a given patient more than doubles the chance of family members having the disorder compared with controls. Mean annual medical costs for patients with untreated obstructive sleep apnea are almost double those of otherwise normal people, and the costs increase in proportion to its severity.

PATHOBIOLOGY

Genetics

Despite the familial aggregation of obstructive sleep apnea, no specific genes or genetic loci have been identified to date. Most identified candidate genes share linkages to pathobiologic correlates, including obesity, craniofacial dysmorphisms, leptin, serotonin, ventilatory responsiveness, and carbonic anhydrase isoenzymes.

Pathogenesis

Obstructive sleep apnea involves complete or partial closure of the collapsible segments of the pharynx, including the velopharynx, oropharynx, and hypopharynx. *Obstructive apnea* is defined by absent airflow for at least 10 seconds associated with continued ventilatory effort (Fig. 100-1). *Obstructive hypopnea*, which is partial airway obstruction, is recognized functionally by a discrete decrease in, rather than cessation of, airflow for at least 10 seconds, associated with either a pathologic decrease in the oxygen saturation of hemoglobin (Sao_2) or an abrupt arousal from sleep. Typically the patient will display progressive ventilatory effort, flattening of the transduced nasal pressure tracing, and/or crescendo snoring (E-Fig. 100-1). A respiratory effort–related arousal occurs when such limitations of airflow are not severe enough to define hypopnea, but the progressive ventilatory effort is nevertheless associated with abrupt arousal from sleep.

Both excess weight and increasing weight are closely linked to the development and worsening of obstructive sleep apnea in adults. Conversely, weight loss, both surgical and nonsurgical, improves it (see Treatment). Anatomic and physiologic explanations of how excess weight contributes to obstructive sleep apnea include restriction of chest wall movement, with resultant mechanical and reflex upper airway narrowing; increased compliance and narrowing of the upper airway; ventilatory instability; and impaired ability to compensate for increased upper airway resistance in sleep.

The final common pathway in each obstructive event is attainment of a critical pharyngeal closing pressure. When narrowed or collapsed, the pharynx is more difficult to expand during the next inspiratory effort, thereby resulting in the characteristic generation of progressively more forceful inspiratory efforts against the obstructed upper airway and increasingly negative intrathoracic pressure excursions. Subsequent reopening of the airway is usually associated with arousal. After several hyperpneic breaths, re-transition into sleep generally occurs. The cycling of blood gases and the recurrent awake-to-asleep transitions interfere with the respiratory controllers' need to find a set point during sleep. Hypopnea or outright apnea followed by hyperpnea is the hallmark of obstructive sleep apnea.

The pathogenic importance of respiratory periodicity in obstructive sleep apnea varies by sleep stage (Chapter 405). The transition into "light" stages of sleep, termed *stages N1* and *N2 non–rapid eye movement* (NREM) *sleep*, is characterized by a tendency for arousal and sleep-wake cycling. These stages are most likely to demonstrate the periodicity of obstructive events characteristic of obstructive sleep apnea. In contrast, "deep" NREM sleep, termed *stage N3* or *slow-wave sleep*, is characteristically a time of relatively regular central nervous system output, with a decreased tendency for arousal, a

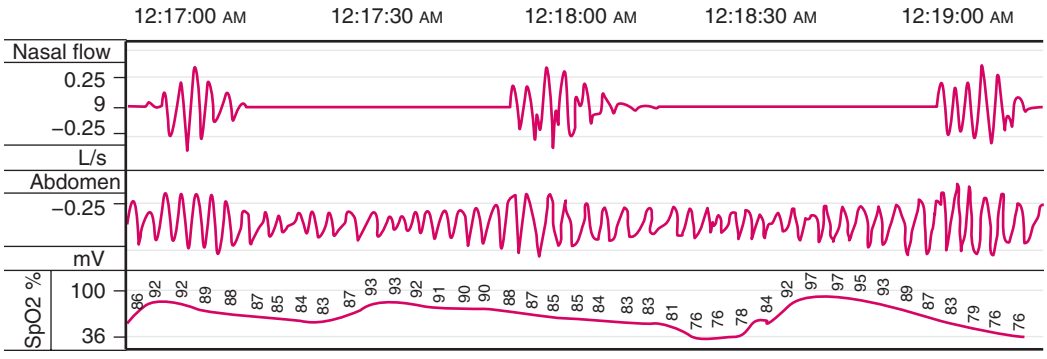

FIGURE 100-1. Polysomnographic tracing of a patient with obstructive sleep apnea during 2 minutes of non–rapid eye movement sleep. Displayed are airflow in the upper airway ("nasal flow"), recorded with a nasal pressure transducer; respiratory effort ("abdomen"), recorded by inductance plethysmography; and oxygen saturation of hemoglobin (Spo_2), recorded with pulse oximetry.

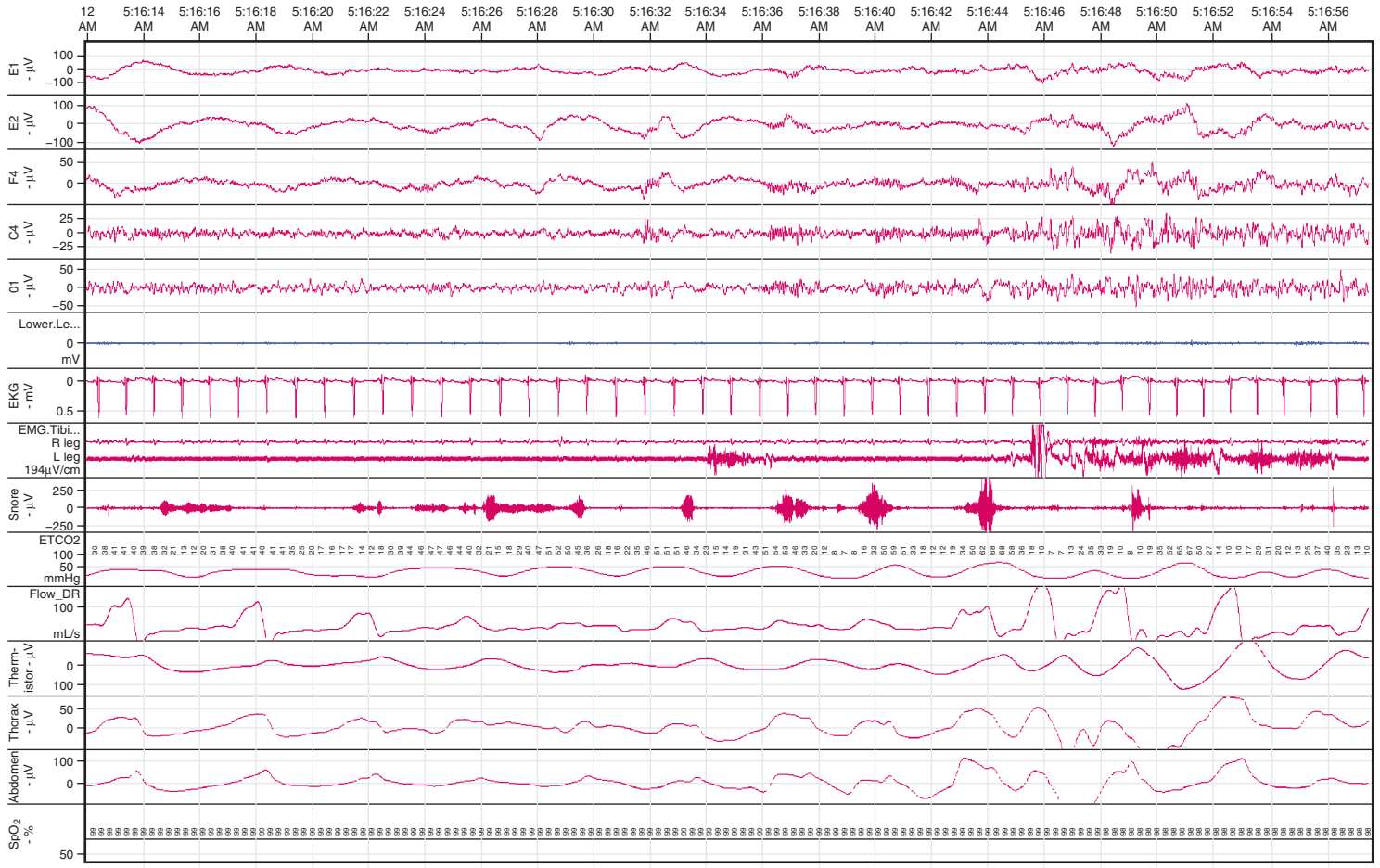

E-FIGURE 100-1. A 45-second epoch of non–rapid eye movement sleep, demonstrating an obstructive hypopnea. There is attenuation of the nasal and oronasal airflow in both the nasal pressure transducer (Flow DR) and oronasal thermistor (Thermistor), accompanied by inspiratory flattening/flow limitation of the airflow (as seen in the nasal pressure transducer), increasing thoracic and abdominal respiratory effort (inspiration upward deflection), and crescendo snoring (Snore). Note the consistently increasing end-tidal P_{CO_2} (ETCO2) throughout the hypopnea. The event is terminated with an abrupt electroencephalographic arousal (F4, C4, C3) and abrupt legs movement seen in the anterior tibialis electromyogram (EMG Tib).

regularization of breathing, and a relative paucity of obstructive events, even with generally increased upper airway resistance when compared with N1 and N2 sleep. During rapid eye movement (REM) sleep, erratic neural drive and descending neural inhibition of accessory ventilatory and upper airway muscles may lead to severe alveolar hypoventilation and apnea. Severe hypoxemia, caused by a combination of obstructive events and ventilation-perfusion mismatch, is characteristic of REM sleep in patients with obstructive sleep apnea.

Pathophysiology

With each obstructive event, the combination of progressive asphyxia, increasingly negative intrathoracic pressure, and sudden autonomic and behavioral arousal leads to acute cardiac and cerebrovascular perturbations, including increased afterload of both the left and right ventricles, decreased left ventricular compliance, increased pulmonary artery pressure, decreased coronary artery blood flow, and increased myocardial oxygen demand (Chapter 53). The abrupt arousal at the termination of the majority of obstructive events is associated with sympathetic discharge, leading to peripheral vasoconstriction and an abrupt increase in the heart rate and in systolic and diastolic blood pressure, even as cardiac output continues to fall when ventilation resumes with the airway reopened. Accordingly, systemic blood pressure characteristically fluctuates; it is relatively low during apnea and acutely elevated at termination of the obstructive event (Chapter 67). Electrocardiographic abnormalities include sinus bradycardia during obstructive events and acceleration at arousal; in REM sleep, which is a time of increased vagal tone, sinoatrial and atrioventricular block may be seen. Ventricular and supraventricular ectopy and dysrhythmia can occur with obstructive sleep apnea. At the termination of obstructive apnea, cerebral blood flow and oxygenation are decreased. The recurrent abrupt, transient arousal from sleep at the termination of each obstructive event is associated with fragmented sleep and a decreased ability to consolidate restorative sleep.

CLINICAL MANIFESTATIONS

Symptoms and Signs

The cardinal manifestations of obstructive sleep apnea include loud, chronic snoring; excessive daytime somnolence; and apneas witnessed by third parties. The snoring reflects vibratory noise from partially occluded pharyngeal soft tissue, usually with the mouth open, and it typically occurs in a crescendo pattern, with a burst of louder noise at resolution of the event. Loud snoring without obstructive sleep apnea may progress to obstructive sleep apnea, particularly if the patient gains weight.

The most useful observation for identifying patients with obstructive sleep apnea is nocturnal choking or gasping (a three-fold increase in risk). Snoring is very common in patients with obstructive sleep apnea but is not nearly as predictive, and normal-weight people with mild snoring are unlikely to have moderate or severe obstructive sleep apnea.[1]

In the hospital setting, obstructive sleep apnea may be detected because of apnea or refractory decreases in Sao_2 during surgical or endoscopic procedures that require sedation or anesthesia. In other cases, the diagnosis may be suspected when hospital staff note nocturnal heart block or dysrhythmias during sleep.

Excessive Daytime Somnolence

Patients with excessive daytime somnolence (Chapter 405) fall asleep unexpectedly; this is typically microsleep rather than sustained sleep episodes. Excessive daytime somnolence may be quantified by laboratory tests that monitor the propensity to fall asleep during the day or by questionnaires or subjective scales that assess sleepiness or decrements in quality of life. Resolution of obstructive sleep apnea does not necessarily resolve excessive daytime somnolence, thereby suggesting the possibility of sustained neurologic perturbation from chronic intermittent hypoxemia and highlighting the association of obstructive sleep apnea with metabolic, neurocognitive, respiratory, and cardiovascular perturbations. Further, many patients with otherwise clinically significant obstructive sleep apnea do not complain of excessive daytime somnolence, nor are patients typically aware of their degree of sleep fragmentation. Mood disorders, including depression and irritability, as well as perturbations in visual memory and working memory appear to be related to the severity of sleep fragmentation and hypoxemia.

Obstructed Breathing

The bed partners of obstructive sleep apnea patients often describe breathing cessation rather than obstructed breathing; close questioning usually elicits the obstructive nature of the breathing. Patients may be aware of their own snoring or complain of choking or dyspnea, particularly in relation to an inability to sleep supine. Morning dry mouth is a common symptom, as is morning headache.

Insomnia and Parasomnia

Insomnia, which is the subjective complaint of difficulty falling asleep or staying asleep (Chapter 405), is associated with the consistent sleep interruption characteristic of obstructive sleep apnea. Transient arousals during N3 sleep may result in confusional parasomnias, such as sleep walking and sleep talking. Arousals and increased work of breathing may result in restless sleep and night sweats. Nocturia, possibly mediated by atrial natriuretic receptors, may resolve with treatment.

Upper Airway Abnormalities

Nasal congestion, rhinitis, chronic sinusitis, and nasopharyngeal anatomic abnormalities are often associated with obstructive sleep apnea, as are craniofacial abnormalities such as micrognathia and retrognathia. Large tonsils, redundant soft palate tissue, and a large tongue may all be associated with a "crowded" oropharynx, but the precise role of these upper airway abnormalities in the pathogenesis of the disorder is unclear.

DIAGNOSIS

The spectrum of sleep-related breathing disorders other than obstructive sleep apnea includes hypoventilation and gas exchange disorders that may worsen with sleep, including nocturnal asthma (Chapter 87), chronic obstructive pulmonary disease (Chapter 88), neuromuscular and chest wall disorders (Chapter 99), and obesity-hypoventilation syndrome (Chapter 220), as well as other disorders in which central apnea (Chapter 405) is prominent, such as idiopathic central apnea, Cheyne-Stokes breathing, and central alveolar hypoventilation (Chapter 86). Patients with obstructive sleep apnea alone characteristically do not hypoventilate while awake, unlike patients with other disorders of hypoventilation. In patients with suspected obstructive sleep apnea, clinicians must also consider other disorders that are associated with hypersomnia, such as narcolepsy, insufficient sleep, poor sleep hygiene, and periodic limb movement disorder, as well as circadian rhythm disorders, such as shift work sleep disorder (Chapter 405). The use of standardized questionnaires such as the STOP-Bang questionnaire (Table 405-5) and the Epworth Sleepiness Scale (Table 405-3) can select patients for definitive testing.

Polysomnography

The definitive diagnostic study is polysomnography (see Fig. 100-1),[2] which generally involves all-night monitoring in a sleep laboratory by electroencephalography, electro-oculography (primarily to determine rapid eye movements characteristic of REM sleep), electrocardiography, leg and chin electromyography, and measures of respiratory effort, airflow, Sao_2, and alveolar or arterial carbon dioxide (usually end-tidal or transcutaneous CO_2). Audiovisual recordings can identify crescendo snoring and thoracoabdominal paradoxical breathing efforts to help differentiate obstructive from nonobstructive hypopnea.

The severity of obstructive sleep apnea is usually described by the apnea-hypopnea index, which is the number of obstructive apneas plus hypopneas per hour of sleep (Fig. 100-2). The term *respiratory disturbance index* includes obstructive apneas, hypopneas, and respiratory effort–related arousals. Oxygen desaturation more directly reflects the chronic intermittent hypoxia that has been increasingly linked to adverse health effects and outcomes.

It is possible to diagnose obstructive sleep apnea by means of unattended home studies that include cardiorespiratory monitoring and/or use of an autotitrating positive airway pressure (PAP) machine, which continuously self-adjusts the level of positive pressure delivered to the airway on a breath-to-breath basis in response to upper airway impedance changes. Such autotitration allows the immediate diagnosis and treatment of a patient suspected of having severe obstructive sleep apnea, with equivalent outcomes at lower cost compared with studies performed at sleep centers.

TREATMENT

Rx

The goal of treatment is to decrease sleep fragmentation and repetitive asphyxia, the resultant cardiovascular, metabolic, and cerebrovascular stress, and the increased work of breathing associated with obstructive sleep apnea.

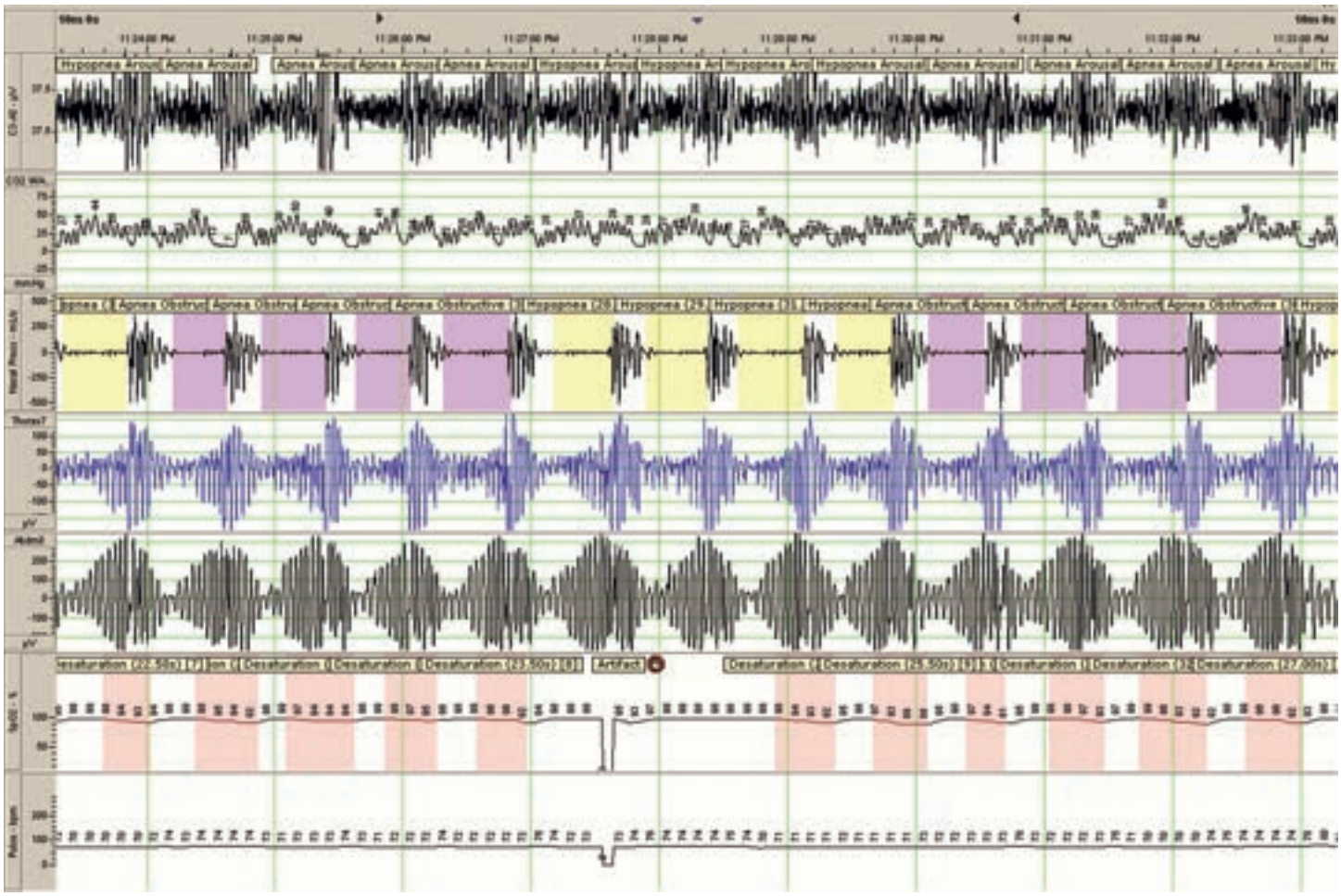

FIGURE 100-2. Continuous 10-minute polysomnographic tracing in a patient with severe obstructive sleep apnea. The 14 repetitive obstructive apneas constitute an apnea-hypopnea index of 84/hour. Each obstructive event is associated with a cyclic waxing-waning respiratory effort (seen in the thoracic and abdominal effort tracings), absent airflow (in the nasal pressure transducer), electroencephalographic arousal (C3-A2, top tracing), oxygen desaturation, and increased end-tidal Pco_2.

Mechanical Therapy

Positive Airway Pressure

CPAP, which is the current first-line therapy for obstructive sleep apnea,[3,4] consistently lowers the apnea-hypopnea index while also decreasing daytime sleepiness, oxygen desaturation, diurnal and nocturnal blood pressure[5], and pulmonary artery pressures. CPAP also improves sleep efficiency, quality of life, and executive mental function. CPAP is superior to oral appliances or no treatment at an acceptable cost, especially in patients with an apnea-hypopnea index greater than 30/hour. For patients with lower apnea-hypopnea index levels, CPAP is generally recommended for patients with prominent oxygen desaturation, daytime sleepiness, or concurrent respiratory, cardiovascular, or cerebrovascular disease. In that setting, CPAP improves cognitive function, improves left ventricular function in patients with heart failure, and decreases the risk for motor vehicle accidents.[5]

The major equipment required to administer PAP includes an interface with appropriate headgear, anchoring straps, and hosing and a compact airflow generator (Fig. 100-3). With CPAP, the fixed level of positive pressure delivered to the upper airway acts as a physiologic splint throughout the respiratory cycle, allowing the patient to achieve normal ventilation as well as more continuous and deeper sleep. CPAP does not supply ventilation above this splinting; the positive end-expiratory pressure may, however, result in improved oxygenation.

Prescribing PAP

PAP is typically prescribed after therapeutic titration in the sleep laboratory, beginning with CPAP of 2 to 4 cm H_2O and increasing by 1- to 2-cm H_2O increments to the minimum level that eliminates obstructive events in all sleep stages. Such a level should reduce other evidence of increased upper airway resistance and increased work of breathing (e.g., snoring, use of accessory inspiratory muscles, thoracoabdominal paradoxical respiration), while improving sleep continuity. Although CPAP is typically prescribed in the 8- to 12-cm H_2O range, it is not uncommon for patients with severe obstructive sleep apnea to need pressures up to 20 cm H_2O. However, as pressures increase above 12 to 14 cm H_2O, the likelihood of air leak and discomfort rises. When laboratory titration is not available, starting CPAP at a level of 10 cm H_2O is reasonable. Adding a feature that decreases expiratory pressure ("pressure relief") does not improve outcomes or adherence, or reduce side effects.

The first night of CPAP is often associated with a "rebound" of slow-wave NREM sleep and REM sleep, along with amelioration of acute fluctuations in heart rate and blood pressure. Daytime sleepiness and vigilance consistently improve with chronic use of CPAP; diurnal and nocturnal blood pressure and catecholamine levels decrease, and left ventricular ejection fraction and diastolic function improve. However, obstructive sleep apnea returns when CPAP is removed, so CPAP must be used nightly throughout sleep. The term *complex sleep apnea* refers to the commonly noted emergence of central apneas during initial CPAP titration; such an effect is generally not long-term, and the significance of this phenomenon remains to be elucidated.

Bilevel PAP

In some cases, a specific ventilatory mode of pressure delivery—generally bilevel PAP, wherein inspiratory pressure is set higher than expiratory pressure—may be helpful after the level of expiratory positive pressure necessary to prevent closure of the airway during expiration is established. Bilevel PAP may be delivered with a backup rate, similar to assist/control mode ventilation. Bilevel PAP should be considered in those with hypoventilation syndromes that overlap obstructive sleep apnea, such as obese patients who continue to have significant hypoventilation or ventilation-perfusion mismatch–related hypoxemia despite CPAP. For most obstructive sleep apnea patients beginning treatment, however, no high-quality data show improved outcomes or cost-effectiveness of bilevel PAP compared with CPAP, even when bilevel PAP is used as secondary therapy when CPAP is unsuccessful.

Autotitrating PAP

CPAP and bilevel PAP can be delivered so that the pressure automatically adjusts to the patient's breathing, based on either breath-to-breath detection of changes in airway resistance or the pattern of breathing. Short-term outcomes, including apnea-hypopnea index, symptomatic sleepiness, and quality of life, are similar, regardless of whether CPAP is guided by titration at a sleep

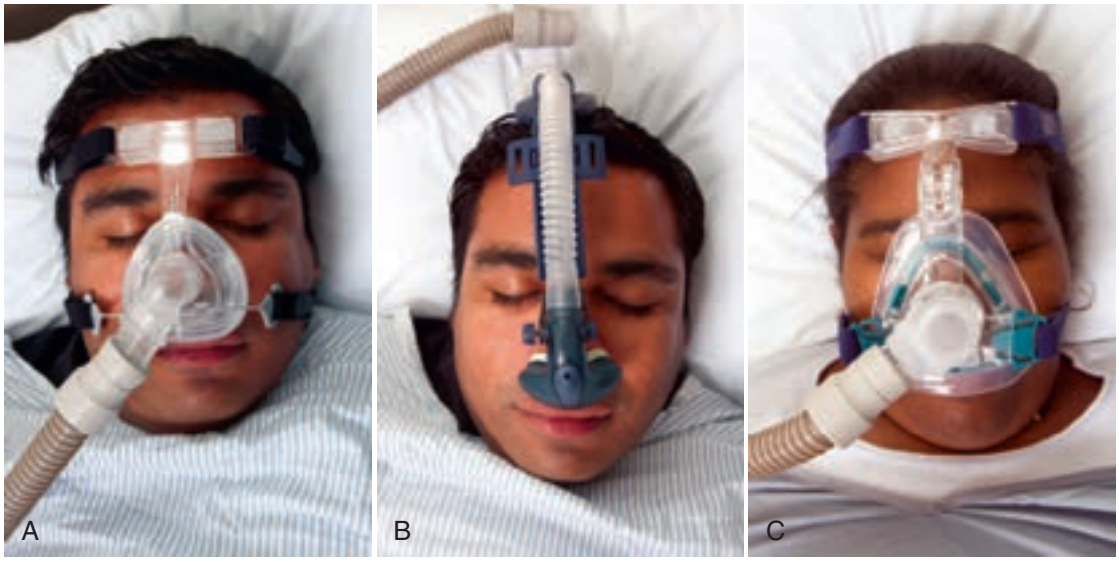

FIGURE 100-3. Positive airway pressure–patient interfaces. A, Subject wearing a nasal mask and headgear with positive airway pressure being delivered. Positive pressure is delivered only to the nasal airway; opening of the mouth may cause air to leak and decrease the efficacy of the positive-pressure regimen. **B,** Subject wearing nasal pillows. Positive pressure is delivered directly into the nares rather than covering the nose or mouth; only minimal headgear is necessary to anchor the interface in place. **C,** Subject wearing a full face mask. Positive pressure is delivered to both the nose and the mouth based on the subject's own breath-to-breath partitioning of nasal and oral airflow.

center or use of an autotitrating machine at home to determine a fixed pressure.[■] However, autotitrating PAP does not improve adherence or outcomes compared with fixed-pressure CPAP.

PAP-Patient Interface

The choice of interface is important for achieving optimal efficacy and adherence. Aside from many different shapes, sizes, and consistencies of nasal masks (see Fig. 100-3A), flexible nasal "pillows" or "prongs" (see Fig. 100-3B) can fit directly into the nares, avoiding the discomfort of a mask over the nose and pain and pressure at the bridge of the nose and over the upper teeth. Although nasal PAP is commonly used, nasal delivery may be ineffective because of nasal congestion, nasopharyngeal anatomic abnormalities, or inability to keep the mouth closed, allowing air to leak from the mouth so that ineffective pressure is delivered to the collapsible airway. Soft "hybrid" interfaces, which have a mouthpiece and nasal prongs to permit both mouth and nose breathing, may be tried in such settings. Full face masks (see Fig. 100-3C) may also improve CPAP efficacy in such situations, but gastrointestinal bloating secondary to air swallowing is a common side effect, and the risks for vomiting and aspiration are a concern. Regardless of the interface used, the patient must be vigilant to maintain the interface throughout sleep, particularly after changes in position. Cold or heated humidification may be added directly to the circuit to improve comfort and adherence, but contamination of the humidifier with pathogenic organisms must be avoided.

Adherence and Outcome

PAP adherence is generally similar to that for treatments of other chronic medical disorders, although most studies suggest that CPAP is not used optimally for maximizing restorative sleep. Adherence may be improved by systematic cognitive behavioral and educational strategies, as well as strategies that address patient-specific issues.[■] Treatment with PAP can be accurately monitored by systems that track the amount of time that physiologically successful PAP levels are delivered.

Oxygen

Although supplemental oxygen alone during sleep may ameliorate hypoxemia and improve sleep quality in obstructive sleep apnea, such treatment alone may lengthen apneas, cause paradoxical worsening of Sa_{O_2}, and fail to improve sleep fragmentation. Therefore, oxygen alone should not be used as first-line treatment for suspected or proven obstructive sleep apnea without nocturnal monitoring.

When sleep Sa_{O_2} remains low despite otherwise optimal levels of CPAP or bilevel PAP, oxygen can be added directly to the mask, to an adapter near the mask, or beneath the mask by nasal cannulae.[6,■] Higher flow rates of oxygen may be necessary as PAP levels increase, particularly if bilevel PAP is used.

Other Ways to Relieve or Bypass Obstruction

Oral appliances, specifically in the form of mandibular advancement devices fitted by an expert, improve symptoms in patients with obstructive sleep apnea, particularly with predominantly supine sleep apnea.[■] However, such devices are generally less effective than CPAP in improving apnea-hypopnea index and oxygen desaturation during sleep. They are currently recommended specifically for patients with mild or moderate obstructive sleep apnea who do not benefit from CPAP or who prefer such an appliance rather than CPAP. Other recently developed alternatives include expiratory positive airway pressure, hypoglossal nerve stimulation, and oral pressure therapy, but few data are available on their efficacy.

General Measures

Sleep Positioning and Nasal Treatments

Sleep positioning may benefit many patients who have obstructive sleep apnea predominantly in the supine position, although such positioning does not appear to be as effective as CPAP in decreasing the apnea-hypopnea index. Treatment of chronic nasal congestion and inflammation with nasal steroids, saline washes, and systemic antihistamine-decongestant regimens sometimes may ameliorate obstructive sleep apnea. Otolaryngologic consultation can be useful to identify treatable nasopharyngeal disorders, including the rare nasopharyngeal neoplasm.

Weight Loss

Weight loss is a primary goal in an overweight patient with obstructive sleep apnea. It not only affects the severity of the breathing disorder during sleep but also may contribute to regression of the associated metabolic and cardiovascular perturbations.[■] Non–morbidly obese patients who achieve nonsurgical weight loss also have significantly greater improvements in obstructive sleep apnea than do controls without such weight loss. In severely obese patients with marked obstructive sleep apnea, bariatric surgery significantly reduces weight compared with conventional medical therapy, but the incremental weight loss does not necessarily result in incremental improvement in obstructive sleep apnea itself.[7]

Treatment of Underlying Conditions

It is important to diagnose and treat underlying conditions such as diabetes (Chapter 229) and heart failure (Chapter 59) in patients with obstructive sleep apnea, but such treatment does not necessarily optimally treat the obstructive sleep apnea. For example, in patients with obstructive sleep apnea and heart failure, cardiac atrial overdrive pacing provides small improvements in the apnea-hypopnea index and sleep-related oxygen desaturation, but it is not as effective as CPAP.

Medical Therapy

There is no acceptably efficacious pharmacologic therapy for obstructive sleep apnea.[■] Respiratory stimulants, including medroxyprogesterone and acetazolamide, have not proved effective in patients with normal Pa_{CO_2} levels. Selective serotonin reuptake inhibitors, including fluoxetine and paroxetine, have been associated with a decreased apnea index during NREM but not REM sleep in a small number of patients. Tricyclic antidepressants have shown some utility in predominantly REM sleep–associated obstructive sleep apnea by decreasing REM volume, although side effects have hindered the use of such agents. Hormone replacement therapy may ameliorate the breathing disorder in postmenopausal women.

Surgical Procedures

Tracheostomy, which bypasses the site of upper airway obstruction, decreases morbidity and mortality and improves blood gas abnormalities in obstructive sleep apnea. However, tracheostomy makes speech difficult and is reserved for only the most severe cases and for patients with concomitant hypoventilation syndromes that do not respond to noninvasive forms of PAP.

Procedures to reduce uvular or palatal tissue, which were once widely recommended, have not shown a consistent benefit in high-quality studies.[8] As a result, surgical uvulopalatopharyngoplasty, radio frequency volumetric tissue reduction of the palate or tongue (or both), and laser-assisted uvuloplasty are not recommended as first-line therapy to treat symptomatic patients. A newer experimental approach is an implantable upper airway stimulator, which has shown promising preliminary results but has not been tested in randomized trials.[9]

PROGNOSIS

Population studies show an increased risk for all-cause and cerebrovascular and coronary mortality in patients with untreated severe obstructive sleep apnea and in patients with apnea-hypopnea indices of 30 or more per hour, independent of other major risk factors.[10] When patients with severe obstructive sleep apnea are prescribed CPAP, 5-year survival rates are significantly higher in those with good CPAP adherence (>6 hours/day) than in those with poor adherence. Like other patients with sleep disorders (Chapter 405), patients with obstructive sleep apnea and moderate to severe sleepiness are at greater risk for morbidity and mortality from motor vehicle accidents compared with drivers without obstructive sleep apnea.

Grade A References

A1. Rosen CL, Auckley D, Benca R, et al. A multisite randomized trial of portable sleep studies and positive airway pressure autotitration versus laboratory-based polysomnography for the diagnosis and treatment of obstructive sleep apnea: the HomePAP study. *Sleep*. 2012;35:757-767.

A2. Iftikhar IH, Valentine CW, Bittencourt LR, et al. Effects of continuous positive airway pressure on blood pressure in patients with resistant hypertension and obstructive sleep apnea: a meta-analysis. *J Hypertens*. 2014;32:2341-2350.

A3. Berry RB, Hill G, Thompson L, et al. Portable monitoring and autotitration versus polysomnography for the diagnosis and treatment of sleep apnea. *Sleep*. 2008;31:1423-1431.

A4. Wozniak DR, Lasserson TJ, Smith I. Educational, supportive and behavioural interventions to improve usage of continuous positive airway pressure machines in adults with obstructive sleep apnoea. *Cochrane Database Syst Rev*. 2014;1:CD007736.

A5. Gottlieb DJ, Punjabi NM, Mehra R, et al. CPAP versus oxygen in obstructive sleep apnea. *N Engl J Med*. 2014;370:2276-2285.

A6. Phillips CL, Grunstein RR, Darendeliler MA, et al. Health outcomes of continuous positive airway pressure versus oral appliance treatment for obstructive sleep apnea: a randomized controlled trial. *Am J Respir Crit Care Med*. 2013;187:879-887.

A7. Chirinos JA, Gurubhagavatula I, Teff K, et al. CPAP, weight loss, or both for obstructive sleep apnea. *N Engl J Med*. 2014;370:2265-2275.

A8. Mason M, Welsh EJ, Smith I. Drug therapy for obstructive sleep apnoea in adults. *Cochrane Database Syst Rev*. 2013;5:CD003002.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Myers KA, Mrkobrada M, Simel DL. Does this patient have obstructive sleep apnea? The Rational Clinical Examination systematic review. *JAMA*. 2013;310:731-741.
2. Qaseem A, Dallas P, Owens DK, et al. Diagnosis of obstructive sleep apnea in adults: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. 2014;161:210-220.
3. Qaseem A, Holty JE, Owens DK, et al. Management of obstructive sleep apnea in adults: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. 2013;159:471-483.
4. Jordan AS, McSharry DG, Malhotra A. Adult obstructive sleep apnoea. *Lancet*. 2014;383: 736-747.
5. Strohl KP, Brown DB, Collop N, et al. An official American Thoracic Society Clinical Practice Guideline: sleep apnea, sleepiness, and driving risk in noncommercial drivers. An update of a 1994 Statement. *Am J Respir Crit Care Med*. 2013;187:1259-1266.
6. Mehta V, Vasu TS, Phillips B, et al. Obstructive sleep apnea and oxygen therapy: a systematic review of the literature and meta-analysis. *J Clin Sleep Med*. 2013;9:271-279.
7. Ravesloot MJ, Hilgevoord AA, van Wagensveld BA, et al. Assessment of the effect of bariatric surgery on obstructive sleep apnea at two postoperative intervals. *Obes Surg*. 2014;24:22-31.
8. Aurora RN, Casey KR, Kristo D, et al. Practice parameters for the surgical modifications of the upper airway for obstructive sleep apnea in adults. *Sleep*. 2010;33:1408-1413.
9. Strollo PJ Jr, Soose RJ, Maurer JT, et al. Upper-airway stimulation for obstructive sleep apnea. *N Engl J Med*. 2014;370:139-149.
10. Kendzerska T, Mollayeva T, Gershon AS, et al. Untreated obstructive sleep apnea and the risk for serious long-term adverse outcomes: a systematic review. *Sleep Med Rev*. 2014;18:49-59.

REVIEW QUESTIONS

1. A hypertensive and obese 55-year-old man, now 2 days after a hip fracture reduction, has nocturnal ventricular tachycardia and complains of an inability to maintain consolidated sleep. The nursing staff reports that he has very loud and crescendo snoring when he sleeps. What is the best immediate recommendation in this setting?

A. Extended-release nonbenzodiazepine sedative-hypnotic at bedtime
B. Increased doses of opioid analgesics to improve pain control
C. Nocturnal oxygen supplementation via Venturi mask, titrated to achieve an oxygen saturation of hemoglobin (Spo_2) of more than 92%
D. Intracardiac electrophysiology study (EPS)
E. Bedside diagnostic polysomnography with positive airway pressure titration

Answer: E The most likely diagnosis is obstructive sleep apnea, which has been exacerbated by sedative and opioid medications. A trial of continuous positive airway pressure (CPAP) is indicated.

2. Which of the following patients is *not* likely to have a higher prevalence of obstructive sleep apnea than the general adult male population?

A. An 85-year-old man with a normal body mass index (BMI) and borderline systemic hypertension
B. A 50-year-old man with a normal BMI and stable heart failure
C. A 60-year-old morbidly obese woman with no known lung disease but with hypoventilation when awake
D. A 40-year-old man with cystic fibrosis and a normal BMI
E. A 55-year-old man 1 week following acute myocardial infarction

Answer: D Cystic fibrosis is not known to be directly associated with a higher prevalence of obstructive sleep apnea.

3. Current recommendations for reducing the risk associated with driving owing to sleepiness in patients with obstructive sleep apnea includes which of the following?

A. Increased nocturnal sleep time
B. CPAP
C. Armodafinil
D. Strategic napping
E. Fluoxetine

Answer: B CPAP is currently the only recommended treatment for reducing driving risk in patients with obstructive sleep apnea (Strohl KP, Brown DB, Collop N et al. An official American Thoracic Society Clinical Practice Guideline: sleep apnea, sleepiness, and driving risk in noncommercial drivers. An update of a 1994 Statement. *Am J Respir Crit Care Med.* 2013;187: 1259-1266).

101

INTERVENTIONAL AND SURGICAL APPROACHES TO LUNG DISEASE

DAVID J. FELLER-KOPMAN AND MALCOLM M. DECAMP

Interventional pulmonology and minimally invasive thoracic surgery have drastically changed the approach to the diagnosis and staging of lung cancer, the management of central airway obstruction, and the treatment of patients with pleural disease.

BRONCHOSCOPY

Flexible bronchoscopy and endobronchial ultrasound have revolutionized the evaluation of airway and parenchymal diseases, the approach to hilar and mediastinal adenopathy, and the staging of lung cancer. New technologies offer bronchoscopic treatment for patients with severe asthma and emphysema.

Endobronchial Ultrasound

Bronchoscopy is a standard component of the evaluation and staging of patients with thoracic tumors. Convex-probe endobronchial ultrasound uses a curvilinear ultrasound probe, which is incorporated into the tip of the bronchoscope, to visualize structures outside of the airway and sample hilar and mediastinal lymph nodes under direct, real-time imaging. This approach has largely replaced mediastinoscopy in the staging of patient with lung cancer[1,2] and in the diagnosis of sarcoidosis (Chapter 95) and certain lymphomas. Adding endobronchial ultrasound to surgical staging improves the sensitivity for finding nodal metastases and can reduce the unnecessary thoracotomy rate from 18 to 7%.[1] It is important to note, however, that the sensitivity of endobronchial ultrasound is not 100%, so nondiagnostic samples must be confirmed as true negatives with either surgical staging or clinical and radiologic follow-up.

In radial-probe endobronchial ultrasound, an ultrasound probe is passed through the working channel of the bronchoscope to visualize parenchymal lesions as well as to discriminate tumor invasion compared with compression of the central airways. When used with smaller bronchoscopes and advanced navigational technologies, such as electromagnetic navigation or virtual bronchoscopic navigation, this technique can improve the diagnostic yield when sampling pulmonary nodules.

Bronchial Thermoplasty

Bronchial thermoplasty uses radio frequency energy to ablate airway smooth muscle in patients whose severe asthma (Chapter 87) remains symptomatic despite medical therapy. The patient must undergo three bronchoscopic procedures to treat all of the visible airways. The right lower lobe is treated during the first procedure, the left lower lobe in the second, and both upper lobes in the third (the right middle lobe is not treated). Although asthma may be exacerbated in the immediate peritreatment period, this approach can significantly improve asthma-related quality of life, days lost from work or school, and emergency department visits in carefully selected patients for at least five years.[1]

Bronchoscopy for Central Airway Obstruction

A variety of malignant and nonmalignant diseases can obstruct the central airways (Table 101-1). Malignant causes include bronchogenic carcinoma (Chapter 191) or metastatic malignancy to the airways, as well as extrinsic compression from adenopathy. Nonmalignant causes include granulation tissue arising after an intubation or tracheostomy, adenopathy from sarcoidosis (Chapter 95), inflammatory conditions such as relapsing polychondritis (Chapter 275), amyloidosis (Chapter 188), and infectious causes such as tuberculosis (Chapter 324) and respiratory papillomatosis.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

A high index of suspicion is essential because significant airway obstruction can be present before the development of symptoms or the characteristic abnormalities on the inspiratory and expiratory flow-volume curves (Chapter 85). Patients with central airway obstruction often develop exertional dyspnea when the tracheal lumen is less than 8 mm in diameter (normal is about 18 to 20 mm) but do not develop stridor, which usually is a sign of impending respiratory failure and the need for urgent intervention, until the tracheal diameter is less than 5 mm.

Hemoptysis (Chapter 83) may be present in patients with malignancies but is also seen in infectious and inflammatory conditions. Some patients may present with a postobstructive pneumonia that responds poorly to antibiotic treatment. Most patients, however, present with nonspecific symptoms including dyspnea and cough.

Patients who have risk factors for chronic airway obstruction and present with symptoms consistent with obstruction unresponsive to conventional therapy should undergo airway imaging by computed tomography (CT) (Fig. 101-1) and flexible bronchoscopy (Fig. 101-2). Because loss of airway patency can be lethal, evaluation of these patients should be performed by individuals experienced in the management of critical airway disorders.

TABLE 101-1 CAUSES OF CENTRAL AIRWAY OBSTRUCTION

NONMALIGNANT	MALIGNANT
Congenital vascular sling	Primary airway tumors (e.g., bronchogenic, mucoepidermoid, adenoid cystic carcinoid)
Cartilage	Tumors metastatic to the airway (e.g., bronchogenic, renal cell, breast, melanoma, thyroid, colon, esophageal)
Relapsing polychondritis	External compression
Tracheobronchomalacia	Lymphadenopathy from any malignancy
Lymphadenopathy	Mediastinal tumors (e.g., thyroid, thymus, germ cell, lymphoma)
Infectious (e.g., histoplasmosis, tuberculosis)	
Sarcoidosis	
Granulation tissue associated with artificial airways, airway stenosis, aspirated foreign bodies, surgical anastomosis	
Inflammatory lesions (e.g., granulomatosis with polyangiitis, amyloidosis, papillomatosis)	
External compression (e.g., goiter)	
Internal	
Secretions	
Blood clot	

Adapted from Feller-Kopman D, Mehta AC, Wahidi MM. Therapeutic bronchoscopy. In: Broaddus VC, Mason RJ, Ernst JD, et al, eds. *Murray & Nadel's Textbook of Respiratory Medicine.* 6th ed. Philadelphia, PA: Saunders; in press (2016).

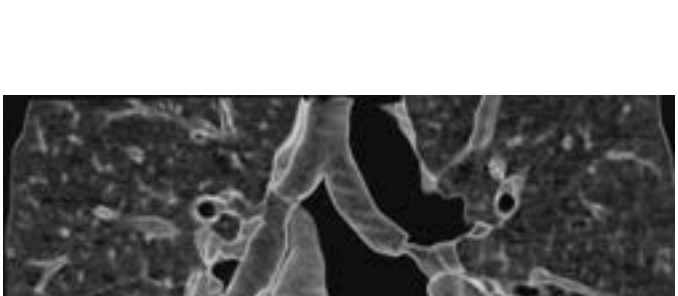

FIGURE 101-1. Three-dimensional reconstruction of a computed tomographic scan in a 20-year-old woman after bilateral lung transplantation. The image clearly shows the high-grade stenosis at the level of the left mainstem anastomosis. The right-sided airways are normal.

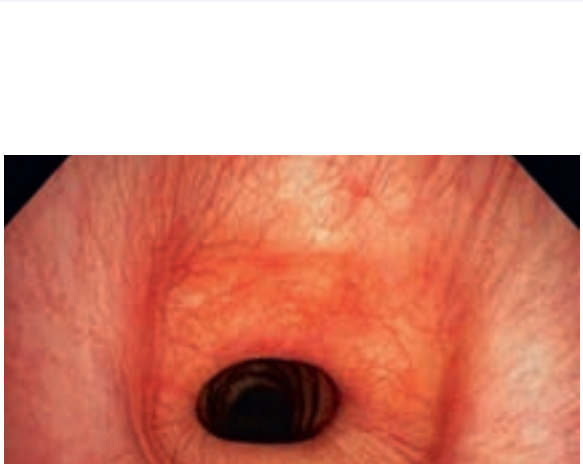

FIGURE 101-2. Bronchoscopic view of the proximal trachea in a 54-year-old man with shortness of breath and stridor. Significant circumferential narrowing is noted in the subglottic space, consistent with gastric reflux–induced subglottic stenosis.

TREATMENT

Relief of airway obstruction requires a multidisciplinary approach by expert physicians. Techniques for achieving a patent airway include coring out the tissue with the barrel of the rigid bronchoscope; laser treatment to vaporize the tissue; electrocautery or argon plasma coagulation to carbonize the tissue, so it will later slough; cryotherapy; and mechanical débridement with forceps or a microdebrider.[3] In the absence of definitive randomized trials, the choice of modality is often left to the expertise and resources of the bronchoscopist.

In contrast to endobronchial disease, airway stenting is the primary treatment for airway obstruction caused by extrinsic compression. Because airway stents can be associated with long-term complications, including stent fractures, the formation of granulation tissue, infection, and migration, they are primarily a palliative approach for malignant diseases and should be used sparingly in patients with nonmalignant airway obstruction.

PROGNOSIS

Successful endoscopic treatment can significantly improve the quality of life of patients with malignant airway obstruction and improve their survival rates to be the equivalent of similar stage patients without airway obstruction. Patients with nonmalignant disease similarly experience significant improvements in physiology (pulmonary function), exercise capacity, and quality of life.

Surgical Approaches

Open Approaches

Thoracotomy has been the standard approach to evaluate the contents of the pleural space, lung parenchyma, pulmonary hilum, and ipsilateral mediastinum and diaphragm. Selective lung ventilation, achieved by use of a double-lumen endotracheal tube or mainstem bronchial blocker, allows one lung to be collapsed and visualized. Given the precision and accuracy of contemporary preprocedure CT and magnetic resonance imaging (MRI), many procedures can be performed using smaller targeted incisions, with a muscle-sparing technique that allows specific access to regional pathology. Many lung processes can also be approached through a median sternotomy, which affords access to both lungs, although access to the left lower lobe through this approach can be challenging.

Video-Assisted Thoracoscopic Surgery

Thoracoscopy, or video-assisted thoracoscopic surgery (VATS), requires two or three incisions, termed *ports*, to place instruments in any intercostal space. A common configuration is to place one port for the video thoracoscope and two ports for endoscopic instrumentation. With the VATS approach, recovery is shorter than after an open thoracotomy because little if any muscle is divided and no mechanical rib spreading retractors are used.

The size of the surgical incisions depends on the goals of the procedure and the anatomic findings at the time of exploration. Unexpected pleural symphysis or incomplete lobar fissures may require extension of the incision to facilitate visualization. In patients undergoing anatomic resection, such as segmentectomy or lobectomy, at least one of the port incisions is extended to 4 to 8 cm in length to permit extraction of the resected lung from the hemithorax.

SURGERY FOR BENIGN LUNG DISEASE

A variety of benign lung diseases present as focal parenchymal lesions or diffuse processes that require a tissue biopsy for diagnosis. Thoracoscopy has essentially replaced a limited thoracotomy and wedge resection for this purpose. VATS provides a more complete view of the ipsilateral hemithorax, including the visceral, parietal, and mediastinal pleura, as well as access to all lung lobes. In addition, subpleural nodules that are too small to be visualized by preoperative radiography can be identified so that representative biopsy samples can be obtained.

Spontaneous Pneumothorax

Although most spontaneous pneumothoraces (Chapter 99) are uncomplicated, up to 20% of patients with pneumothoraces experience complications such as tension pneumothorax, persistent air leak despite tube drainage, or recurrent pneumothoraces either ipsilaterally or contralaterally. Patients in whom a second pneumothorax develops have a 70 to 80% chance of a third recurrence within 2 years. The current surgical approach to the treatment of

recurrent pneumothoraces is VATS resection of the subpleural blebs responsible for the pneumothorax, usually combined with mechanical abrasion of the parietal pleura or chemical pleurodesis (e.g., with 300 mg of minocycline or with insufflated talc) to induce an inflammatory reaction that will cause the visceral and parietal pleural surfaces to fuse, thereby preventing subsequent recurrence.

Giant Bullae

Most patients with chronic obstructive pulmonary disease (COPD) have diffuse parenchymal disease, but a small number of patients with COPD have dominant or giant bullae that may occupy 50% or more of the volume of the hemithorax and that compress relatively preserved lung parenchyma. The indications for bullectomy include progressive symptoms with demonstrated disability, obstructive spirometry (Chapter 85), and a single or dominant bullous lesion with radiographic demonstration of compression of the surrounding preserved lung parenchyma. Either excision or plication can remove the bullous lesion.

Malignant Lung Disease

Solitary Pulmonary Nodules

Most small pulmonary nodules (see Fig. 191-2) present in the periphery of the lung beyond the reach of diagnostic bronchoscopy. For such lesions, VATS excisional biopsy (see Video 101-1) of a small pulmonary nodule leads to a definitive diagnosis in almost all cases and is generally preferred over transthoracic needle biopsy.[4] Furthermore, thoracoscopy can allow concurrent nodal staging should a primary malignancy (Chapter 191) be confirmed. In the absence of regional adenopathy, patients with primary malignant lesions can undergo definitive resection at the same time.

Primary Lung Cancer

In patients with node-negative primary lung cancer and adequate pulmonary reserve, lobectomy or pneumonectomy is indicated to obtain optimal survival and to decrease the risk for local recurrence. Similar oncologic outcomes can be achieved by performing an open thoracotomy or by lobectomy through VATS or robotic assistance, and VATS lobectomy generally is associated with fewer complications, less pain, a shorter hospitalization, and a speedier recovery.[5] Achieving a complete lobectomy is important because sublobar resection (segmentectomy or wedge) for stage I lung cancers is associated with a two- to three-fold higher incidence of locoregional recurrence.

Metastatic Cancer

The lung is a frequent site of metastatic recurrence. Common histologies include colorectal cancer (Chapter 193), renal cell carcinoma (Chapter 197), sarcoma (Chapter 202), melanoma (Chapter 203), breast cancer (Chapter 198), and head and neck cancer (Chapter 190). VATS is often the diagnostic procedure of choice to locate and excise nodules that are too small for reliable percutaneous biopsy.

The role of pulmonary metastectomy as therapy for advanced disease remains controversial. Five-year survival rates of 20 to 30% have been reported for selected patients, especially if the disease-free interval from original diagnosis to lung metastasis is greater than 3 years. These cases often require resection of bilateral lung nodules, which in turn mandate either a median sternotomy, clamshell incision, staged thoracotomies, or bilateral VATS.

SURGERY FOR ADVANCED LUNG DISEASES: LUNG VOLUME REDUCTION SURGERY

Emphysema (Chapter 88) is the most common chronic progressive disabling lung disease treated by pulmonologists and thoracic surgeons. In eligible patients (Table 101-2), lung volume reduction surgery confers durable symptomatic, physiologic, and survival benefits compared with medical therapy (Fig. 101-3) for patients who have severe emphysema but who do not have a forced expiratory volume in 1 second (FEV_1) measure of less than 20% of predicted with either a homogeneous distribution of emphysema on CT or a diffusing capacity of less than 20% of predicted.[6,7] However, subgroup analyses suggest that the benefit is mainly in patients with upper lobe predominant disease.

For eligible patients, most programs require a 6- to 10-week preoperative pulmonary rehabilitation followed by a cardiopulmonary exercise test to assess the risks and benefits of surgery. Patients with upper lobe predominant emphysema and a low preoperative exercise capacity have a nearly 50% lower risk for death after lung volume reduction surgery compared with continued medical therapy. High-risk patients with severe airflow obstruction (FEV_1 < 20%) should be assessed for lung transplantation evaluation unless their disease is localized to the upper lobes and their gas exchange as defined by diffusing capacity is preserved (Table 101-3).

In experienced centers, bilateral stapled resection approaches yield nearly twice the physiologic benefit of unilateral lung volume reduction surgery without adversely affecting operative morbidity or mortality. Bilateral lung volume reduction surgery using the VATS approach also may reduce intensive care unit and hospital length of stay and increase the likelihood of living independently 60 days after surgery compared with median sternotomy.

TABLE 101-2 INCLUSION AND EXCLUSION CRITERIA FOR LUNG VOLUME REDUCTION SURGERY

INCLUSION CRITERIA
Radiographic evidence of emphysema, especially involving upper lobes
Hyperinflation evidenced by TLC > 100% predicted and RV > 150% predicted
FEV_1 > 20 and < 45% predicted (after bronchodilator)
DLCO > 20% predicted
Severe dyspnea
Restricted activities of daily living
Decreased quality of life
Abstinence from tobacco
EXCLUSION CRITERIA
Active smoking
Bronchiectasis
Pulmonary nodule requiring evaluation
Excessive daily sputum production
Previous thoracotomy
Obvious pleural disease
Active or inducible coronary ischemia
Pulmonary hypertension
Depressed LVEF (<45%)
Obesity (BMI > 32)
Unable or unwilling to participate in pulmonary rehabilitation
Systemic steroids, ≥20 mg prednisone/day

BMI = body mass index; DLCO = diffusion capacity for carbon monoxide; FEV_1 = first second forced expiratory volume; LVEF = left ventricular ejection fraction; RV = residual capacity; TLC = total lung capacity.
Adapted from DeCamp MM Jr, McKenna RJ Jr, Deschamps CC, et al. Lung volume reduction surgery: technique, operative mortality and morbidity. *Proc Am Thorac Soc.* 2008;5:442-446; and DeCamp MM Jr, Lipson D, Krasna M, et al. The evaluation and preparation of the patient for lung volume reduction surgery. *Proc Am Thorac Soc.* 2008;5:427-431.

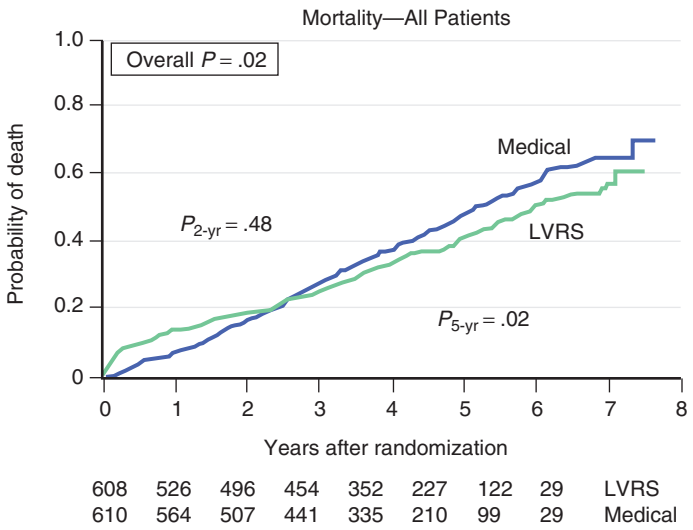

FIGURE 101-3. Long-term mortality of all patients treated with lung volume reduction surgery (LVRS) versus maximal medical therapy in the National Emphysema Treatment Trial. Note the statistically significant (P = .02) reduction in relative risk for death (RR = 0.85) in the surgical cohort. (Adapted from Naunheim KS, Wood DE, Mohsenifar Z, et al, for the National Emphysema Treatment Trial Research Group. Long-term follow-up of patients receiving lung-volume-reduction surgery versus medical therapy for severe emphysema in the National Emphysema Treatment Trial. *Ann Thorac Surg.* 2006;82:431-443.)

TABLE 101-3 DECISION GUIDE FOR SELECTION OF LUNG VOLUME REDUCTION SURGERY VERSUS TRANSPLANTATION FOR SEVERE CHRONIC OBSTRUCTIVE PULMONARY DISEASE

FACTORS FAVORING LVRS	FACTORS FAVORING TRANSPLANTATION
Age > 65 yr	$FEV_1 \leq 20\%$ predicted
Upper lobe predominant disease	$D_{LCO} \leq 20\%$ predicted
Chronic medical conditions	Homogeneous or lower lobe distribution of disease
Hepatitis B and/or C	TLC < 100% predicted
HIV infection	RV < 150% predicted
Renal insufficiency	$PaCO_2 > 60$ mm Hg
Cirrhosis	$PaO_2 < 45$ mm Hg
Neuropathy	6 MWD < 140 m or
Poorly controlled diabetes	<3 min unloaded pedaling cycle ergometer
Osteoporosis	Pulmonary hypertension
Severe GERD	Bronchiectasis
Poor esophageal motility	Recurrent pulmonary infections
Malignancy	
Unable to maintain long-term follow-up	
Psychiatric issues limiting compliance	
Insufficient social support	

6 MWD = 6-minute walk distance; D_{LCO} = diffusion capacity of carbon monoxide; FEV_1 = first second forced expired volume; GERD = gastroesophageal reflux disease; HIV = human immunodeficiency virus; LVRS = lung volume reduction surgery; RV = residual volume; TLC = total lung capacity.
Adapted from Patel N, DeCamp M, Criner GJ. Lung transplantation and lung volume reduction surgery versus transplantation in chronic obstructive pulmonary disease. *Proc Am Thorac Soc.* 2008;5:447-453.

ENDOSCOPIC MANAGEMENT OF EMPHYSEMA

The only therapies that have a proven survival advantage in the management of patients with severe emphysema are the use of oxygen and surgical lung volume reduction. Endoscopic approaches to lung volume reduction may offer a less invasive way to achieve some of the benefits of surgical lung volume reduction.[8] As with surgical lung volume reduction, the goals of these technologies are to reduce the magnitude of overdistended and poorly perfused lung tissue, thereby increasing elastic recoil, diminishing dynamic hyperinflation, and redistributing airflow to better perfused areas of the lung. Bronchoscopic options include one-way valves, coils, foam, and steam. Although some of these devices are approved in Europe, none is currently approved in the United States because randomized trials have not shown significant benefits.

LUNG TRANSPLANTATION

About 180 worldwide lung transplant centers perform more than 3700 transplantations per year.[9] Lung transplantation is now an accepted therapy for all forms of advanced lung disease. The use of ex vivo perfusion to resuscitate lungs considered unsuitable for transplant by traditional procurement criteria appears to be expanding the availability of donor organs.[10]

The most common indications for transplantation (Table 101-4) are diseases or conditions that share the following features: they produce extreme disability in affected patients, they are unresponsive to medical therapy, and they are responsible for limited life expectancy in affected patients. With the exception of a small number of cases of sarcoidosis and lymphangioleiomyomatosis, the original lung disease does not usually recur after lung transplantation.

Types of Procedures

Currently, four types of lung transplantation procedures are performed. *Single-lung transplantation,* which is typically performed through a posterolateral thoracotomy incision, requires three anastomoses: the mainstem bronchus, pulmonary artery, and pulmonary veins and left atrium. The contralateral lung is not removed, so single-lung transplantation is not performed in patients with bilaterally infected lungs (e.g., patients with cystic fibrosis or bronchiectasis) (see Table 101-3).

Bilateral lung transplantation is performed in a sequential fashion that is functionally equivalent to two single-lung transplantations completed during a single operation, most commonly through a transverse sternotomy ("clamshell") incision. It requires six anastomoses: both mainstem bronchi, both pulmonary arteries, and both sets of pulmonary veins. It is the procedure of choice for patients with bilaterally infected lungs and is also performed in certain patients with emphysema, primary pulmonary hypertension, and other diseases (see Table 101-3). Bilateral transplantation is preferred for nearly all indications because a double-lung recipient can expect a half-life of 6.9 years compared with 4.6 years for a single-lung recipient. As a result, about 75% of the world's reported lung transplants are now bilateral.

Heart-lung transplantation is now performed in only about 75 cases per year. It is an en bloc procedure with right atrial, aortic, and distal tracheal anastomoses. It is performed in patients with advanced lung disease and coexistent irreparable cardiac disease, usually associated with fixed pulmonary hypertension, and in those with Eisenmenger syndrome (Chapter 69).

Living donor lobar transplantation involves the removal of a lower lobe from each of two living donors. One is implanted into each hemithorax of the recipient in a manner similar to bilateral lung transplantation.[11]

TABLE 101-4 INDICATIONS AND CONTRAINDICATIONS FOR LUNG TRANSPLANTATION

SINGLE-LUNG TRANSPLANT	PATIENTS (%)	DOUBLE-LUNG TRANSPLANT	PATIENTS (%)
INDICATIONS			
COPD	44	CF, bronchiectasis	30
Pulmonary fibrosis, sarcoid	40	Emphysema	27
α_1-Antitrypsin deficiency	65	α_1-Antitrypsin deficiency	6
PPH, Eisenmenger	1.4	PPH, Eisenmenger	6.2
CF, bronchiectasis	2.4	Pulmonary fibrosis, sarcoid	17
Retransplantation	3	Retransplantation	2
Other*	4	Other*	6
ABSOLUTE CONTRAINDICATIONS			
Untreatable advanced extrapulmonary organ dysfunction (e.g., heart, liver, kidney)			
CAD not amenable to PCI or bypass			
Poor LV function (could consider heart-lung transplantation)			
Malignancy within 2 years (excludes cutaneous squamous or basal cell carcinoma)			
5-year disease-free interval preferred			
Noncurable extrapulmonary infection			
Infection with human immunodeficiency virus			
Hepatitis B antigen positivity			
Hepatitis C with histologic evidence of active liver disease			
Active substance abuse (including cigarettes)			
Severe musculoskeletal disease affecting the thorax			
Documented noncompliance			
Untreatable psychiatric condition that impairs compliance			
Absence of consistent and reliable social support			
RELATIVE CONTRAINDICATIONS			
Physiologic age > 65 yr			
Poor nutritional status (<70% ideal body weight)			
Severe obesity (BMI > 30 kg/m)			
Symptomatic osteoporosis			
Colonization with highly virulent and/or highly resistant fungi, mycobacteria, or bacteria			
Requirement for invasive ventilation and/or circulatory support			
Uncontrolled chronic medical conditions (e.g., diabetes, hypertension, GERD)			
Severely limited functional status with poor rehabilitation potential			
Psychosocial problems likely to affect the outcome adversely			
High-dose (>20 mg of prednisone daily) corticosteroid use			

*Other includes lymphangioleiomatosis, non-retransplantation-related obliterative bronchiolitis, and miscellaneous indications.
BMI = body mass index; CAD = coronary artery disease; CF = cystic fibrosis; COPD = chronic obstructive pulmonary disease; GERD = gastroesophageal reflux disease; LV = left ventricle; PCI = percutaneous coronary intervention; PPH = primary pulmonary hypertension.
Adapted from Yusen RD, Christie JH, Edwards LB, et al. Twenty-sixth official adult lung and heart-lung transplant report—2013. *J Heart Lung Transplant.* 2013;32:965-978; and Orens JB, Estenne M, Arcasoy S, et al. International guidelines for the selection of lung transplant candidates: 2006 update—a consensus report from the Pulmonary Scientific Council of the International Society for Heart and Lung Transplantation. *J Heart Lung Transplant.* 2006;25:745-755.

Evaluation of Potential Transplant Recipients

The ideal candidate for lung transplantation has lung disease unresponsive to medical therapy but is in otherwise good health. Patients who experience

critical illness as a result of lung disease often have poor nutritional status, coexistent major organ dysfunction, refractory infection, or other contraindications to transplantation. The specific evidence-based recommendations for referral for transplant evaluation vary with the underlying disease.

In the United States, the lung allocation system is based on expected disease-specific and patient-specific survival during the waiting period and after engraftment, thereby reflecting net transplant benefit. Early evaluations of the system, which was introduced in 2005, indicate shorter waiting times, an increase in the total number of transplantations performed, a decreased waitlist mortality, and an unchanged overall survival after transplantation.

Post-transplantation Issues

Most of the medical issues that patients and physicians face after lung transplantation are the consequence of the transplantation and post-transplantation medication rather than the underlying disease for which the transplantation was performed. Examples include immunosuppression, infections and their prophylaxis, acute allograft rejection, chronic allograft rejection, and nonpulmonary complications of transplantation.

Immunosuppression

The standard chemotherapeutic regimen for immunosuppression after lung transplantation consists of a calcineurin inhibitor such as cyclosporine or tacrolimus, azathioprine or mycophenolate mofetil, and corticosteroids. More than 50% of centers add an antilymphocyte antibody preparation in the first days after transplantation, and this practice has led to a small but statistically significant improvement in long-term survival.

Infections and Prophylaxis after Lung Transplantation

Lung transplant recipients are at high risk for bacterial, viral, fungal, and protozoal infections; infections are the leading causes of death during the early post-transplantation period. In the first 3 months after transplantation, bacterial infections are responsible for most deaths. In approximately one third of patients, pneumonia is diagnosed in the first weeks after transplantation, with gram-negative organisms as the cause in 75% of cases. Colonization and recurrent infections, usually with *Pseudomonas* species, often develop in patients with chronic rejection.

Among potential viral pathogens, *cytomegalovirus* (CMV; Chapter 376) is the most important in lung transplant recipients. Seronegative patients who receive an allograft from a seropositive donor are at particularly high risk for the development of a clinically significant CMV infection. Seronegative patients who have a seronegative donor are at low risk for infection if they are treated with seronegative blood products. Epstein-Barr virus (EBV) has been associated with the development of post-transplantation lymphoproliferative disorder.

Aspergillus species are the most common cause of invasive fungal infection (Chapter 339). Colonized patients and those deemed at risk may receive prophylactic inhaled amphotericin B.

Because of the nature of the immunosuppressive chemotherapeutic regimen used, patients are at high risk for infection by the protozoan *Pneumocystis jirovecii* (Chapter 341). The use of trimethoprim-sulfamethoxazole prophylaxis (typically 1 double-strength tablet three times weekly indefinitely) has virtually eliminated *Pneumocystis* pneumonia.

Acute Rejection

Histologically, the initial manifestation of acute rejection is a lymphocyte-predominant inflammatory response, usually centered on blood vessels, airways, or both. By convention, acute rejection is graded histologically from 0 (normal) to 4 (severe), with subclasses defined by the presence or absence of airway inflammation.

The risk for acute allograft rejection is highest in the early months after transplantation and declines with time. Multiple episodes of acute rejection are the major risk factor for the subsequent development of chronic rejection.

Clinically, patients may have fever, cough, and exertional dyspnea. Evaluation may demonstrate rales or rhonchi on chest examination, a decline in pulmonary function by spirometry, leukocytosis, opacities on chest radiography, and exertional desaturation. The clinical manifestation is often indistinguishable from infectious pneumonia, and the clinical impression is accurate in only 50% of cases.

Treatment of acute rejection most often consists of high-dose corticosteroids (typically, 1 g/day of methylprednisolone administered intravenously for 3 days).

Chronic Rejection

PATHOBIOLOGY

The bronchiolitis obliterans syndrome is thought to be a manifestation of chronic rejection. Risk factors for development of the syndrome include the number of acute rejection episodes and, in some series, previous symptomatic CMV infection. Pathologically, "early" lesions demonstrate inflammation and disruption of the epithelium of small airways, followed by growth of granulation tissue into the airway lumen and subsequent complete or partial obstruction. The granulation tissue then organizes in a stereotypical pattern with resultant fibrosis that obliterates the lumen of the airway.

CLINICAL MANIFESTATIONS

Clinically, bronchiolitis obliterans is accompanied by nonspecific symptoms.[12] Progressive exertional breathlessness typically develops, and pulmonary function testing usually demonstrates evidence of progressive airflow obstruction (Chapter 85). Bronchiolitis obliterans is classified according to the FEV_1: 0 (no significant abnormality) if FEV_1 is greater than 80% of baseline; 1 (mild) if FEV_1 is 65 to 80% of baseline; 2 (moderate) if FEV_1 is 50 to 65% of baseline; and 3 (severe) if FEV_1 is 50% or less of baseline. In early stages, chest radiography is notable only for hyperinflation, but it may show bronchiectasis as the syndrome progresses. Later stages of bronchiolitis obliterans may include a syndrome of bronchiectasis with chronic productive cough and airway colonization with *Pseudomonas* species.

DIAGNOSIS

The diagnosis of bronchiolitis obliterans is made on both clinical and pathologic grounds. Transbronchial biopsy has a low yield for demonstrating histologic evidence of bronchiolitis obliterans, but when such evidence is seen, it is diagnostic. In patients with a compatible clinical syndrome, exclusion of anastomotic stenosis and occult pulmonary infection is sufficient to establish the diagnosis.

TREATMENT Rx

A variety of therapies have been tried for chronic rejection, including pulse corticosteroids, antilymphocyte antibodies, total lymphoid irradiation, photopheresis, and nebulized cyclosporine, but none has been clearly established as effective. Most patients with bronchiolitis obliterans experience a progressive decline in pulmonary function despite immunosuppression.

PROGNOSIS

Bronchiolitis obliterans is the leading cause of late mortality after lung transplantation. Half of lung transplant recipients surviving to 5 years will have either biopsy-proven bronchiolitis obliterans or the clinical diagnosis of bronchiolitis obliterans syndrome.

Nonpulmonary Medical Complications of Lung Transplantation

Most of the nonpulmonary medical complications that arise in patients after lung transplantation are the result of immunosuppressive therapy. One or more of these complications develop in virtually all lung transplant recipients.

Osteoporosis (Chapter 243) is common because of the long-term use of corticosteroids and cyclosporine. Bone density should be monitored periodically, and pharmacologic therapy should be instituted if excessive bone loss is identified.

Chronic renal insufficiency (Chapter 130) is common and is the result of therapy with the calcineurin inhibitors cyclosporine or tacrolimus, both of which affect afferent vascular tone in the kidneys and result in an average 50% drop in the glomerular filtration rate in the first 12 months after lung transplantation. Systemic arterial hypertension is also common and is caused by corticosteroids and cyclosporine. Calcium-channel blockers, which are often used to treat hypertension, raise serum cyclosporine levels; appropriate monitoring and dose adjustment are needed when starting such therapy. Both corticosteroids and tacrolimus contribute to the development of diabetes mellitus and hyperlipidemia.

Solid organ transplantation is associated with an increased incidence of malignancy, thought to be due to pharmacologic immunosuppression and

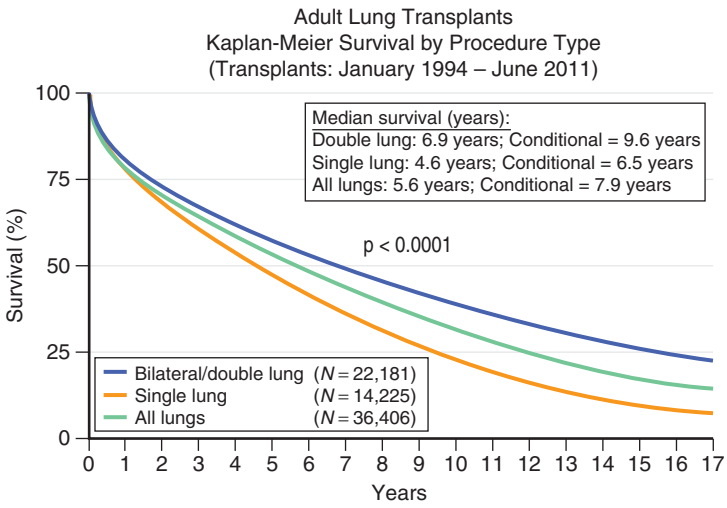

FIGURE 101-4. Kaplan-Meier survival estimates for all adult lung transplantations reported to the International Registry for Heart and Lung Transplantation from 1994 to 2011. Note the highly statistically significant survival advantage conferred by double lung grafts. Because the decline in survival is greatest during the first year after transplantation, the conditional survival (i.e., when 50% of the recipients who survive to at least 1 year have died) provides a more realistic expectation of survival time for recipients who survive the early post-transplant period. (Adapted from Yusen RD, Christie JH, Edwards LB, et al. Twenty-sixth official adult lung and heart-lung transplant report—2013. *J Heart Lung Transplant.* 2013;32:965-978.)

alteration in immune surveillance. Patients are at increased risk for lymphoproliferative malignancies and other types of cancers. Post-transplantation lymphoproliferative disorders occur in about 4% of patients after organ transplantation; most are associated with EBV. These syndromes can be polyclonal or monoclonal. Reduction in immunosuppression is sometimes therapeutic in those with polyclonal disease. The prognosis in patients with monoclonal disease is poor, with little response to modification of immunosuppression or antineoplastic chemotherapy. Patients are also at increased risk for skin, bladder, lung, cervical, and hepatobiliary malignancy after solid organ transplantation.

Outcomes after Lung Transplantation

Currently, the annual mortality rate following lung transplantation is 8 to 10% per year, largely owing to bronchiolitis obliterans syndrome. The median survival after lung transplantation is about 5.5 years (Fig. 101-4).

Grade A References

A1. Annema JT, van Meerbeeck JP, Rintoul RC, et al. Mediastinoscopy vs endosonography for mediastinal nodal staging of lung cancer: a randomized trial. *JAMA*. 2010;304:2245-2252.

A2. Wechsler ME, Laviolette M, Rubin AS, et al. Bronchial thermoplasty: Long-term safety and effectiveness in patients with severe persistent asthma. *J Allergy Clin Immunol*. 2013;132:1295-1302.

A3. Chen JS, Chan WK, Tsai KT, et al. Simple aspiration and drainage and intrapleural minocycline pleurodesis versus simple aspiration and drainage for the initial treatment of primary spontaneous pneumothorax: an open-label, parallel-group, prospective, randomised, controlled trial. *Lancet*. 2013;381:1277-1282.

A4. Naunheim KS, Wood DE, Mohsenifar Z, et al. Long-term follow-up of patients receiving lung-volume-reduction surgery versus medical therapy for severe emphysema by the National Emphysema Treatment Trial Research Group. *Ann Thorac Surg*. 2006;82:431-443.

A5. Fishman A, Martinez F, Naunheim K, et al. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med*. 2003;348:2059-2073.

A6. Sciurba FC, Ernst A, Herth FJF, et al. A randomized study of endobronchial valves for advanced emphysema. *N Engl J Med*. 2010;363:1233-1244.

A7. Shah PL, Slebos DJ, Cardoso PF, et al. Bronchoscopic lung-volume reduction with Exhale airway stents for emphysema (EASE trial): randomised, sham-controlled, multicentre trial. *Lancet*. 2011; 378:997-1005.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Detterbeck FC, Lewis SZ, Diekemper R, et al. Executive summary: diagnosis and management of lung cancer, 3rd ed. American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2013;143:7s-37s.
2. Silvestri GA, Gonzalez AV, Jantz MA, et al. Methods for staging non-small cell lung cancer: diagnosis and management of lung cancer, 3rd ed. American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2013;143:e211S-250S.
3. Bacon JL, Patterson CM, Madden BP. Indications and interventional options for non-resectable tracheal stenosis. *J Thorac Dis*. 2014;6:258-270.
4. Ost DE, Gould MK. Decision making in patients with pulmonary nodules. *Am J Respir Crit Care Med*. 2012;185:363-372.
5. Paul S, Altorki NK, Sheng S, et al. Thoracoscopic lobectomy is associated with lower morbidity than open lobectomy: a propensity-matched analysis from the STS database. *J Thorac Cardiovasc Surg*. 2010;139:366-378.
6. Berger RL, DeCamp MM, Criner GJ, et al. Lung volume reduction therapies for advanced emphysema: an update. *Chest*. 2010;138:407-417.
7. Agzarian J, Miller JD, Kosa SD, et al. Long-term survival analysis of the Canadian Lung Volume Reduction Surgery trial. *Ann Thorac Surg*. 2013;96:1217-1222.
8. Cohen E. Bronchoscopic treatment of end-stage chronic obstructive pulmonary disease. *Curr Opin Anaesthesiol*. 2014;27:36-43.
9. Yusen RD, Edwards LB, Kucheryavaya AY, et al. The registry of the International Society for Heart and Lung Transplantation: thirty-first adult lung and heart-lung transplant report—2014; focus theme: retransplantation. *J Heart Lung Transplant*. 2014;33:1009-1024.
10. Cypel M, Yeung JC, Liu M, et al. Normothermic ex vivo lung perfusion in clinical lung transplantation. *N Engl J Med*. 2011;364:1431-1440.
11. Date H, Sato M, Aoyama A, et al. Living-donor lobar lung transplantation provides similar survival to cadaveric lung transplantation even for very ill patients. *Eur J Cardiothorac Surg*. 2014;[Epub ahead of print].
12. Barker AF, Bergeron A, Rom WN, et al. Obliterative bronchiolitis. *N Engl J Med*. 2014;370:1820-1828.

REVIEW QUESTIONS

1. Which of the following statements is correct concerning endobronchial ultrasound (EBUS)?

A. The use of EBUS obviates the need for surgical lymph node sampling.
B. EBUS can provide adequate material for the analysis of molecular markers in patients with non–small cell lung cancer.
C. Radial-probe EBUS allows for real-time sampling of hilar lymph nodes.
D. The combined use of RP-EBUS and navigation bronchoscopy has a yield equivalent to transthoracic needle biopsy for the diagnosis of a 2-cm pleural-based nodule.
E. All of the above

Answer: B EBUS can obtain adequate material for the evaluation of tumor molecular markers in more than 94% of cases. Although EBUS has very high sensitivity and specificity, specimens showing lymphocytes but no tumor may need to be confirmed as true negatives by surgical lymph node dissection. Radial-probe EBUS is primarily used to identify parenchymal nodules. Transthoracic needle biopsy is associated with a higher diagnostic yield for peripheral nodules but is more likely to be complicated by the development of pneumothorax compared with radio-probe EBUS and navigation bronchoscopy.

2. Which of the following statements is true concerning patients with severe emphysema?

A. Bronchoscopic lung volume reduction can reduce emergency department visits for exacerbations of chronic bronchitis.
B. The use of oxygen and surgical lung volume reduction has been shown to decrease mortality.
C. Surgical lung volume reduction is better than medical therapy for patients with homogenous emphysema.
D. Surgical lung volume reduction improves mortality in the subset of patients with high exercise tolerance.
E. All of the above

Answer: B The use of oxygen and surgical lung volume reduction improves mortality in appropriately selected patients with advanced emphysema: patients with upper lobe–predominant disease, an FEV_1 of less than 20%, and low exercise capacity. For patients with good exercise tolerance, surgical lung volume reduction increases mortality compared with standard medical therapy. There are no data suggesting that bronchoscopic lung volume reduction reduces emergency room visits for exacerbations of chronic obstructive pulmonary disease.

3. Which of the following statements is correct concerning lung transplantation?

A. The outcomes of unilateral lung transplantation are better than for bilateral lung transplantation.
B. Lung transplantation improves survival in patients with advanced emphysema.
C. Infection with *Aspergillus* species is a common cause of mortality at 5 years.
D. The bronchiolitis obliterans syndrome is the leading cause of late mortality.
E. All of the above

Answer: D Bilateral lung transplantation has improved outcomes compared with unilateral transplantation. Transplantation confers no overall survival advantage in patients with advanced emphysema, and obliterative bronchiolitis is the leading cause of mortality at 5 years.

4. Which of the following are true concerning central airway obstruction?

A. The treatment of malignant central airway obstruction does not affect mortality.
B. Laser photoresection is better than argon plasma coagulation to treat malignant central airway obstruction.
C. Airway stenting is the procedure of choice for extrinsic airway obstruction.
D. Airway stents are well tolerated and associated with minimal complications.
E. All of the above

Answer: D Airway stents are the only endoscopic treatment that can palliate extrinsic airway obstruction. They are, however, associated with significant complications. After successful treatment of malignant airway obstruction, patients have mortality rates similar to those of patients with malignancy but no central airway obstruction.

102 APPROACH TO THE PATIENT IN A CRITICAL CARE SETTING

DEBORAH J. COOK

THE INTENSIVIST-LED MULTIDISCIPLINARY TEAM

Patients with critical illness in the intensive care unit (ICU) usually require advanced life support, such as mechanical ventilation, vasopressors, inotropic agents, or renal replacement therapy. Morbidity associated with critical illness includes complications of both acute and chronic diseases, nosocomial and iatrogenic consequences, and impaired quality of life among survivors. Critically ill patients are at a higher risk of death than any other hospital population. Accordingly, the goals of critical care are to reduce morbidity and mortality, to maintain organ function, and to restore health. Unlike many other specialties, critical care medicine is not limited to a particular population, disease, diagnosis, or organ system.

Staffing of ICUs with critical care physicians, often referred to as *intensivists*, who provide mandatory consultations or principal ongoing care is associated with a significantly reduced ICU and hospital mortality and reduced ICU and hospital lengths of stay. The addition of nighttime intensivist staffing appears to reduce mortality by 38% in ICUs with low-intensity daytime staffing but not in centers with high-intensity daytime staffing, such as academic ICUs. These findings emphasize the value of the on-site availability of trained physicians who are dedicated to appropriate triaging, diagnosis, monitoring, treatment, and palliation of critically ill patients.

Daily rounds by an ICU physician who leads the coordinated work of nurses, pharmacists, respiratory therapists, physiotherapists, dietitians, chaplains, and other physicians appear to improve outcomes. Observational studies suggest that a standardized, goal-oriented approach to care delivered by multidisciplinary clinicians, with explicitly defined roles and best practices checklists, can help improve the quality of ICU rounds.[1] The critical care process can be optimized by interprofessional leadership, communication, and a positive organizational culture.

FLUID RESUSCITATION

Intravenous fluids to maintain or to restore intravascular volume are an important component of ICU therapy. Both crystalloid and colloid solutions are in widespread use. Crystalloids are readily available and inexpensive, whereas colloids generally require less volume to achieve a specific physiologic goal.

Fluid replacement with either normal saline or 4% albumin results in similar rates of death, organ failure, and other clinical outcomes, but crystalloids may lower mortality for patients with traumatic brain injury (Chapter 399). Fluid management with hydroxyethyl starch increases the need for renal replacement therapy and increases mortality compared with crystalloid infusions. On the basis of these data, either crystalloid- or albumin-based colloid fluid resuscitation is recommended for most critically ill patients, crystalloids are recommended for head-injured patients, and starches are not recommended.

SEDATION, ANALGESIA, AND SPONTANEOUS BREATHING TRIALS

Endotracheal intubation, central venous catheterization, postoperative pain management, and other ICU procedures require that most patients receive sedation, analgesia, or both. Sedatives and analgesics are used to ensure ongoing tolerance to mechanical ventilation, particularly in patients with shock or severe acute respiratory distress syndrome (ARDS). As long as pain and anxiety are well treated, bolus injections are preferred to continuous infusions because of emerging concerns about drug-induced delirium and delayed weaning from the ventilator. If patients are receiving drug infusions, daily interruption of sedatives and analgesics, by protocols that provide an opportunity for the patient to be observed safely in a less sedated state, are associated with a shorter duration of mechanical ventilation and ICU length of stay than continuous infusions. A second key component of managing sedation and analgesia is to use a drug titration protocol and nurse-led sedation scale; in these situations, daily interruption of sedation infusions may confer no additional benefit.

Discontinuation of ventilation is affected by sedation and analgesic infusions, and vice versa. A daily sedation vacation followed by a spontaneous breathing test increases the days of breathing without assistance and shortens ICU stay and hospital stay compared with usual sedation management plus a daily spontaneous breathing test. In the year after enrollment, patients who were treated with a "wake up and breathe" protocol, which linked daily sedation vacation periods with daily spontaneous breathing trials, had a 32% better survival rate. On the basis of these data, a nurse-implemented sedation and analgesic management scale with daily drug interruption and daily spontaneous breathing trials are recommended for mechanically ventilated critically ill patients.

LONG-TERM OUTCOMES FOR SURVIVORS

Biomarkers of inflammation, residual organ dysfunction, and functional disabilities persist in most ICU survivors even after transfer out of the ICU. Treatments administered in the ICU also have serious sequelae. For example, neuromuscular blockers and corticosteroids may contribute to critical illness polyneuropathy. These problems have particularly serious adverse consequences for elderly critically ill patients who are deconditioned before hospitalization.

In addition, anxiety, post-traumatic stress, and major mood disorders are common among patients and their caregivers during recovery. Therefore, although ICU discharge and hospital discharge are milestones in a patient's trajectory, sequelae of critical illness have rarely resolved completely when patients are on the regular hospital unit. For example, residual muscle weakness is common,[2] even 5 years after ICU discharge.

The legacy of critical care and the resulting residual functional impairment increase postdischarge morbidity and costs, thereby encouraging rehabilitation interventions to improve long-term outcomes. In a randomized trial of patients who received mechanical ventilation for 72 hours or less, the addition of graduated, individualized, early physical therapy and occupational therapy during daily sedation vacation periods improved functional capacity at hospital discharge, reduced the duration of delirium, and reduced the number of ventilator days during the 28-day follow-up. Discontinuation of physiotherapy as a result of patient instability, usually patient-ventilator asynchrony, occurred in only 4% of all sessions. This trial highlights how the recovery of critically ill patients potentially can be improved by coordinated multidisciplinary care.

APPLYING EVIDENCE TO PREVENT COMPLICATIONS OF CRITICAL ILLNESS

Considerable evidence of effective preventive and therapeutic ICU interventions has emerged in randomized trials during the past decade. For example, evidence-based initial management of a patient with urosepsis and ARDS includes low tidal volume ventilation, avoidance of early high-frequency oscillation, high positive end-expiratory pressure, inotrope or vasopressor infusion, low-dose corticosteroids, early enteral small bowel nutrition, avoidance of antioxidants, head of bed elevation, oral antisepsis with chlorhexidine, stress ulcer prophylaxis,[3] thromboprophylaxis with low-molecular-weight heparin, and insulin therapy aimed at avoiding marked hyperglycemia but not achieving normoglycemia (Chapters 104 and 105). In mechanically ventilated adults, chest radiographs on demand provide clinical outcomes equivalent to those of routine radiographs, despite about one-third fewer radiographs. Later during the stabilization and recovery phase of critical illness, evidence-based management includes targeted protocol-driven sedation, daily interruption of sedation infusions, daily spontaneous breathing trials, and early mobilization.

Potential barriers to applying evidence in fast-paced ICUs include a perceived lack of responsibility, unclear decisional authority, and errors of omission. Passive dissemination of information, whether written or verbal, is generally ineffective in modifying physicians' behavior. More effective strategies to encourage the implementation of evidence-based recommendations are interactive education, audit and feedback, reminders (written or computerized), involvement of local opinion leaders, and multifaceted approaches. In the high-acuity ICU setting, preprinted physician orders may help guide (but not dictate) management (Table 102-1). For example, a statewide intervention coached local safety teams to lead multidisciplinary education about central venous catheter management strategies known to decrease infection risk, including a procedural checklist that incorporated handwashing, full barrier precautions for catheter insertion, chlorhexidine skin cleansing, avoidance of the femoral site, and removal of unnecessary catheters. This multimethod approach, which included periodic site-specific feedback,

TABLE 102-1 ICU ADMISSION ORDERS: EXAMPLE FOR A PATIENT WITH UROSEPSIS AND ARDS

MANAGEMENT STRATEGY	ORDERS	REEVALUATE
ACUTE PHASE		
Mechanical ventilation	Target TV 5-7 mL/kg of ideal body weight, PC 16 cm, rate 12, F_{IO_2} 0.7, PEEP 16 cm, plateau pressure <35 cm	PRN
Maintenance fluid	Lactated Ringer 75 mL/hr IV	PRN
Norepinephrine	Titrate to mean arterial pressure >65 mm Hg	PRN
Corticosteroids	Hydrocortisone 50 mg IV q6h while vasopressor dependent	Daily
Sedation	Midazolam 2-8 mg/hr IV, bolus 2-4 mg PRN	PRN
Analgesia	Morphine 1-4 mg IV PRN	PRN
Antibiotics	Ampicillin 2 g IV q6h	Daily
Head of bed	45-degree elevation from horizontal	PRN
Oral antisepsis	Chlorhexidine 15 mL q6h	Daily
Small bowel enteral nutrition	10 mL/hr of a commercial balanced feed containing about 1 kcal/mL; increase by 20 mL q4h to 70 mL/hr	Daily
Stress ulcer prophylaxis	Pantoprazole 40 mg IV daily	Daily
Thromboprophylaxis	Dalteparin 5000 U SC daily	Daily
Intensive insulin therapy if glucose >180 mg/dL	50 U insulin in 50 mL NS; start at 0.5 U/hr, repeat glucose q1h for 4 hr, and reassess; target 110-150 mg/dL	Daily
Glucometer calibration	Calibrate glucose from glucometer and central laboratory every morning	Daily
Tests	Glucose q4h when stable, ABG with each ventilator change, other tests as per ICU team	PRN
Monitoring	Arterial catheter for systolic blood pressure, central venous catheter for central venous pressure and mixed venous oxygen saturation, ECG, oximetry, ABGs, sedation scale, Foley catheter, others as per ICU monitoring protocols	PRN
STABILIZATION AND RECOVERY PHASES		
Sedation vacation	Daily interruption of sedation from 0700 h until 0900 h; restart at half prior infusion rate at 0900 h if necessary; aim to discontinue infusion as soon as possible	Daily
Spontaneous breathing trials	Spontaneous breathing trial when weaning readiness criteria met	Daily
Early mobility	Titrated physiotherapy and occupational therapy when able	Daily

ABG = arterial blood gas; ARDS = acute respiratory distress syndrome; ECG = electrocardiogram; F_{IO_2} = fraction of inspired oxygen; ICU = intensive care unit; IV = intravenous; NS = normal saline; PC = pressure control; PEEP = positive end-expiratory pressure; PRN = as needed; SC = subcutaneous; TV = tidal volume.

decreased catheter-related blood stream infections from 7.7 per 1000 catheter-days at baseline to 1.4 at 18 months' follow-up.[4] In a provincial cluster randomized trial addressing six evidence-based critical care practices in community ICUs, a multimethod approach including video conferencing, education, provision of algorithms, audit, and feedback resulted in a three-fold increased adoption of the six management strategies.

PREDICTIONS, PREFERENCES, AND PALLIATIVE CARE

The prognosis of many critically ill patients improves once they are in the ICU. For others, treatment responsiveness is delayed or not realized, organ dysfunction evolves but does not resolve, and complications arise. Despite best efforts of the multidisciplinary ICU team, critical illness proves fatal to between 5 and 40% of adults. Approximately 2% of ICU patients discharged to the ward are readmitted within 48 hours and about 4% within 120 hours.[5] When a therapeutic trial of critical care is started, and particularly when it is failing, it is crucial to discuss prognosis openly with families (Chapter 3). Among medical ICU patients older than 80 years at one tertiary care university hospital, ICU mortality was 46%, hospital mortality was 55%, and mortality among hospital survivors was 53% at 2 years.[6] About 15% of patients who are admitted to an ICU have clinical courses that probably should generate discussion about palliative care.[7] Families bring key information about the patient's prior function and preferences.

In the shared decision-making model dominant in many settings today, these exchanges often result in plans to withhold or to withdraw basic or advanced life support.[8] Mechanical ventilation is the most frequent life support administered to and withdrawn from critically ill patients. Ventilator withdrawal very often precedes death in the ICU. Patients undergoing ventilator withdrawal or who die while mechanically ventilated have a shorter ICU stay than patients successfully weaned from the ventilator. When life support modalities are withdrawn because their further use would be futile,[9] each can be discontinued or weaned, with attendant considerations and cautions (Table 102-2). Withdrawal may be guided by the severity of the illness and other physiologic characteristics, but it is more heavily influenced by the contemporary life support model that is attentive to a patient's values and the physician's predictions about future quality of life. This complexity underscores the need for ICU teams to be expert communicators, sensitive in eliciting patients' preferences, timely in relieving suffering, and compassionate in

TABLE 102-2 CONSIDERATIONS AND CAUTIONS IN THE WITHDRAWAL OF LIFE SUPPORT

ISSUE	RISKS	OTHER CONSIDERATIONS
Weaning from inotropes or vasopressors	No risk of physical distress	May prolong the dying process, particularly if patient requires low doses and this is the only life support withdrawn
Discontinuation of inotropes or vasopressors	No risk of physical distress	Death may not occur quickly if the patient requires low doses, particularly if mechanical ventilation is ongoing Death may occur quickly if the patient requires high doses, with or without withdrawal of mechanical ventilation
Weaning from mechanical ventilation	Low risk of dyspnea	May prolong the dying process, particularly if the patient requires low pressure settings or low oxygen levels and this is the only life support withdrawn
Discontinuation of mechanical ventilation	Risk of dyspnea	Death may not occur quickly if the patient requires low pressure settings or low oxygen levels Death may occur quickly if the patient requires high pressure settings or high oxygen levels Preemptive sedation is typically needed to blunt air hunger due to rapid changes in mechanical ventilation
Extubation	Risk of dyspnea Risk of stridor (steroids) Risk of airway obstruction (jaw thrust) Risk of noisy breathing (glycopyrrolate)	Avoids discomfort and suctioning of endotracheal tube Can facilitate oral communication Informing families about possible physical signs after extubation can prepare and reassure them Allows for the most natural appearance Not advised if the patient has hemoptysis
Discontinuation of renal replacement therapy	Low risk of physical distress	Death may take several days if this is the only advanced life support withdrawn

Reprinted with permission from Cook D, Rocker G. Dying with dignity in the intensive care unit. *N Engl J Med.* 2014;370:2506-2514. Copyright © 2014 Massachusetts Medical Society.

providing dignity to the dying while administering culturally competent, family-centered end-of-life care. A death with dignity in the ICU infers that whereas some treatments may be foregone, care can be enhanced as death ensues. Fundamental to maintaining dignity is the need to understand a patient's unique perspectives on what gives life meaning in a setting replete with depersonalizing devices. The goal is caring for patients in a manner consistent with their values at a time of incomparable vulnerability, when they cannot speak for themselves.[10]

Grade A References

A1. Kerlin MP, Small DS, Cooney E, et al. A randomized trial of nighttime physician staffing in an intensive care unit. *N Engl J Med*. 2013;368:2201-2209.

A2. Finfer S, Bellomo R, Boyce N, et al. A comparison of albumin and saline for fluid resuscitation in the intensive care unit. *N Engl J Med*. 2004;350:2247-2256.

A3. Rochwerg B, Alhazzani W, Sindi A, et al. Fluid resuscitation in sepsis: a systematic review and network meta-analysis. *Ann Intern Med*. 2014;161:347-355.

A4. Mehta S, Burry L, Cook D, et al. Daily sedation interruption in mechanically ventilated critically ill patients cared for with a sedation protocol: a randomized controlled trial. *JAMA*. 2012;308:1985-1992.

A5. Girard T, Kress JP, Fuchs BD, et al. Efficacy and safety of a paired sedation and ventilator weaning protocol for mechanically ventilated patients in intensive care (Awakening and Breathing Controlled trial): a randomised controlled trial. *Lancet*. 2008;371:126-134.

A6. Schweickert WD, Pohlman MC, Pohlman AS, et al. Early physical and occupational therapy in mechanically ventilated, critically ill patients: a randomised controlled trial. *Lancet*. 2009;373:1874-1882.

A7. Burns KE, Adhikari NK, Slutsky AS, et al. Pressure and volume limited ventilation for the ventilatory management of patients with acute lung injury: a systematic review and meta-analysis. *PLoS ONE*. 2011;6:e14623.

A8. Ferguson ND, Cook DJ, Guyatt GH, et al. High-frequency oscillation in early acute respiratory distress syndrome. *N Engl J Med*. 2013;368:795-805.

A9. Briel M, Meade M, Zhou Q, et al. Higher versus lower positive end-expiratory pressure in patients with acute lung injury and acute respiratory distress syndrome: systematic review and individual patient data meta-analysis. *JAMA*. 2010;303:865-873.

A10. Heyland D, Muscedere J, Wischmeyer PE, et al. A randomized trial of glutamine and antioxidants in critically ill patients. *N Engl J Med*. 2013;368:1489-1497.

A11. Cook D, Meade M, Guyatt G, et al. Dalteparin versus unfractionated heparin in critically ill patients. *N Engl J Med*. 2011;364:1305-1314.

A12. Finfer S, Chittock DR, Su SY, et al. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med*. 2009;360:1283-1297.

A13. Hejblum G, Chalumeau-Lemoine L, Ioos V, et al. Comparison of routine and on-demand prescription of chest radiographs in mechanically ventilated adults: a multicentre, cluster-randomized, two-period crossover study. *Lancet*. 2009;374:1687-1693.

A14. Scales DC, Dainty K, Hales B, et al. A multifaceted intervention for quality improvement in a network of intensive care units. *JAMA*. 2011;305:363-372.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Lane D, Ferri M, Lemaire J, et al. A systematic review of evidence-informed practices for patient care rounds in the ICU. *Crit Care Med*. 2013;41:2015-2029.
2. Kress JP, Hall JB. ICU-acquired weakness and recovery from critical illness. *N Engl J Med*. 2014;370:1626-1635.
3. Dellinger RP, Levy MM, Rhodes A, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Med*. 2013;39:165-228.
4. Pronovost PJ, Goeschel CA, Colantuoni E, et al. Sustaining reductions in catheter related bloodstream infections in Michigan intensive care units: observational study. *BMJ*. 2010;340:c309.
5. Brown SE, Ratcliffe SJ, Kahn JM, et al. The epidemiology of intensive care unit readmissions in the United States. *Am J Respir Crit Care Med*. 2012;185:955-964.
6. Roch A, Wiramus S, Pauly V, et al. Long-term outcome in medical patients aged 80 or over following admission to an intensive care unit. *Crit Care*. 2011;15:R36.
7. Hua MS, Li G, Blinderman CD, et al. Estimates of the need for palliative care consultation across United States intensive care units using a trigger-based model. *Am J Respir Crit Care Med*. 2014;189:428-436.
8. Curtis JR, Vincent JL. Ethics and end-of-life care for adults in the intensive care unit. *Lancet*. 2010;376:1347-1353.
9. Huynh TN, Kleerup EC, Wiley JF, et al. The frequency and cost of treatment perceived to be futile in critical care. *JAMA Intern Med*. 2013;173:1887-1894.
10. Cook D, Rocker G. Dying with dignity in the intensive care unit. *N Engl J Med*. 2014;370:2506-2514.

REVIEW QUESTIONS

1. Optimal models for the care of critically ill patients include all except which of the following?

A. Critical care delivered by a dedicated intensivist or obligatory intensivist consultation on all intensive care unit (ICU) patients
B. Nighttime in-house intensivist coverage
C. Intensivist-led multidisciplinary rounds
D. A multidisciplinary team with explicitly defined roles
E. Use of a "best practices" checklist or equivalent tool on rounds

Answer: B Strategies considered to optimize the process of critical care include a closed ICU model with intensivists delivering the care or consulting on each patient, intensivist-led multidisciplinary rounds in which clinicians have explicitly defined roles, and rounds that incorporate a best practices tool. Randomized trial evidence does not show an impact of 24/7 in-house intensivist coverage in fully staffed academic ICUs.

2. Compared with crystalloids for fluid maintenance or fluid resuscitation, which statement is true regarding starch solutions?

A. Starch decreases the risk of renal failure.
B. Starch increases the probability of needing renal replacement therapy.
C. Starch increases the probability of needing renal replacement therapy, but only in patients with severe sepsis.
D. Starch increases the risk of death in patients with severe sepsis.
E. Both B and D.

Answer: E Starches do not decrease the risk of renal failure. Large randomized trials show that in both general ICU patients and in patients with severe sepsis, starches increase the probability of needing renal replacement therapy; furthermore, in patients with severe sepsis, resuscitation with starches increases the risk of death.

3. For mechanically ventilated patients receiving infusions of sedation or analgesia, which of the following statements is true?

A. Concerns prevail that oversedation may induce delirium, prolonged weakness, and delayed liberation from the ventilator.
B. Nurse-led targeted sedation protocols represent an effective method to minimize the chance of oversedation while maintaining the patient's comfort.
C. Compared with continuing infusions, daily interruption of sedation infusions decreases the duration of ventilation and ICU stay.
D. Daily interruption of sedation infusions appears to have no additional impact on the duration of mechanical ventilation or safety of the patient if a nurse-led targeted sedation protocol is in place.
E. All of the above.

Answer: E Concern is emerging about the adverse effects of prolonged sedation and analgesia infusions. Targeted sedation by a nurse-led protocol represents an optimal approach to minimize unnecessary sedation. Daily interruption of sedation infusions is also effective but may not further improve outcomes if individualized, targeted sedation levels are already managed with a nurse-led protocol.

4. Which of the following statements is not true regarding the legacy of critical illness?

A. Disability among survivors of critical illness is pervasive, affecting many domains of quality of life.
B. Disability among ICU survivors increases the overall length of hospital stay.
C. Disability among family caregivers of ICU survivors is rare.
D. Some aspects of disability among ICU survivors may be mitigated by earlier rehabilitation in the ICU.
E. Combined early physiotherapy, sedation interruption, and spontaneous breathing tests used in combination have been shown to improve survival in critically ill patients.

Answer: C Depression, anxiety, and post-traumatic stress disorder are common in family caregivers of ICU survivors, and ICU survivors themselves can have pervasive, long-lasting multidimensional disabilities. Some disabilities are targets for early physical rehabilitation; others may benefit from counseling and other supports. Coordinated multidisciplinary approaches to help patients along the trajectory to recovery include early physiotherapy, daily sedation interruptions, and spontaneous breathing tests.

5. Which of the following statements about end-of-life care in the ICU is true?

A. Palliative care is exclusively patient centered.
B. Optimal end-of-life care is focused primarily on complete, timely management of symptoms in the final days.
C. Decisions to withdraw life support are primarily determined by acute and chronic physiologic trends during critical illness.
D. Dignified end-of-life care involves caring for patients in a manner consistent with their values when they cannot speak for themselves.
E. In contemporary ICUs, death is more commonly preceded by withdrawal of inotropes and vasopressors than by withdrawal of mechanical ventilation.

Answer: D Palliative care in the ICU ensures the patient's dignity, is holistic, is both patient and family centered, and addresses broad dimensions of health. Mechanical ventilation is the most common life support administered and withdrawn in the ICU. Decisions to withdraw mechanical ventilation are based less on pathophysiology than on the patient's values, which are typically expressed by the family, and physicians' predictions about future quality of life.

103

RESPIRATORY MONITORING IN CRITICAL CARE

JAMES K. STOLLER AND NICHOLAS S. HILL

Monitoring of the respiratory system involves a broad array of assessment techniques ranging from low-technology approaches like a careful physical examination to sophisticated technologies to monitor oxygenation and ventilation.

PHYSICAL EXAMINATION

The physical examination can provide important information about the patient's ventilation and oxygenation. Ventilation can be assessed by recording the respiratory rate (normally 12 to 20 breaths/minute in adults) as well as by closely inspecting the pattern of chest wall movement during inspiration and by noting the use of accessory inspiratory muscles (e.g., the scalene, trapezius, and sternocleidomastoid muscles). Hypopnea (shallow or slow breathing) or a slowed respiratory rate (bradypnea) can indicate decreased ventilation. Shallow breathing may relate to muscle weakness (Chapter 421) or increased lung stiffness, which is commonly accompanied by a compensatory increase in the ratio of the respiratory rate to maintain ventilation. Bradypnea may relate to a suppressed respiratory drive (e.g., excessive use of narcotics, slowing the respiratory rate). Conversely, sustained tachypnea (e.g., >35 breaths/minute in an adult) can indicate ongoing increased work of breathing, impending respiratory failure, and the need for mechanical assistance, such as noninvasive ventilation or intubation and mechanical ventilation, depending on the etiology of the respiratory failure.

Contraction of the sternocleidomastoid muscles or scalene muscles, often with a seated, bent posture, is called the tripod sign (E-Fig. 103-1). This response indicates inadequate diaphragmatic function, most commonly in the setting of emphysema with associated diaphragmatic flattening, which causes a mechanical disadvantage of diaphragmatic contraction. In this circumstance, patients may demonstrate Hoover sign, which is inspiratory retraction of the rib cage at the level of the zone of apposition, where the diaphragm inserts on the chest wall.

The physical examination of the nail beds and lips may also reveal cyanosis, which suggests hypoxemia. Cyanosis occurs when saturation falls, but it requires the presence of 5 g of desaturated hemoglobin. As such, polycythemic patients may show cyanosis with relatively high oxyhemoglobin saturation values, whereas patients with profound anemia may not demonstrate cyanosis even in the face of low values of oxyhemoglobin saturation.

SYSTEMIC ARTERIAL BLOOD GAS ANALYSIS

Sampling of arterial blood, either through a percutaneous arterial puncture or by withdrawal of blood from an indwelling arterial catheter, provides important information about the patient's oxygenation and ventilation status as well as the acuity of and compensation for derangements. The partial pressure of carbon dioxide (Pa_{CO_2}) reflects ventilation, the elimination of carbon dioxide. In many but not all cases, Pa_{CO_2} is close to the mixed alveolar Pa_{CO_2}. The Pa_{CO_2} in the arterial blood is closely related to the ratio of metabolic carbon dioxide production to alveolar ventilation:

$$Pa_{CO_2} = (K)(CO_2 \text{ production rate})/(\text{alveolar ventilation } [V_A]) \quad (1)$$

The partial pressure of oxygen (Pa_{O_2}) reflects the level of oxygenation. Normal levels of oxygenation are defined by the alveolar-arterial oxygen gradient, $P(A\text{-}a)_{O_2}$, which is calculated as

$$P(A\text{-}a)_{O_2} = F_{IO_2}(P_B - P_{H_2O} \text{ at standard pressure and body temperature}) - (Pa_{O_2} + Pa_{CO_2}/\text{respiratory quotient}) \quad (2)$$

where the respiratory quotient equals the number of moles of carbon dioxide produced for each mole of oxygen consumed (generally ~0.8 under normal metabolic conditions at rest but variable with dietary intake and metabolic rate). The normal value of the alveolar-arterial oxygen gradient varies with age and position and can be approximated by the simple equation

$$P(A\text{-}a)_{O_2} = (\text{age}/4) + 4 \quad (3)$$

Normal age-related values of Pa_{O_2} in the sitting position can be determined by the equation

$$Pa_{O_2} \text{ sitting} = 104.2 - (0.27 \times \text{age in years}) \quad (4)$$

Normal values of PaO_2 are generally in the range of 70 to 95 mm Hg, depending on the patient's age.

The Pa_{CO_2} helps assess the adequacy of the patient's ventilation. At sea level, normal values of Pa_{CO_2} range from 35 to 45 mm Hg. Values of Pa_{CO_2} below 35 mm Hg indicate hyperventilation, either as a primary respiratory event (e.g., with anxiety) or in response to another insult (e.g., hypoxemia, sepsis, liver disease). Similarly, values of Pa_{CO_2} exceeding 45 mm Hg indicate hypoventilation, hypercapnia, and respiratory acidosis, which may result either from suppression of the ventilatory drive (Chapter 86) (e.g., excess narcotics; Chapter 34) or from respiratory insufficiency (e.g., respiratory muscle weakness; Chapter 421).

Assessment of the patient's bicarbonate level (HCO_3^-) helps define the chronicity of changes in the patient's Pa_{CO_2}, where the value of bicarbonate is defined by the Henderson-Hasselbalch equation:

$$pH = 6.1 + \log_{10}[HCO_3^-]/0.003\, Pa_{CO_2} \quad (5)$$

Acute increases in Pa_{CO_2} drive the normal kidney to retain bicarbonate (Chapter 118), whereas acute decreases in Pa_{CO_2}, as in hyperventilation from anxiety or liver disease, would be expected to cause the normal kidney to waste bicarbonate to preserve the body's pH (normally 7.35 to 7.45).

The clinician can also assess whether the patient's ventilatory response to metabolic acidosis is appropriate or inadequate by the Winter equation, which predicts the expected Pa_{CO_2} in the face of a decreased bicarbonate from a

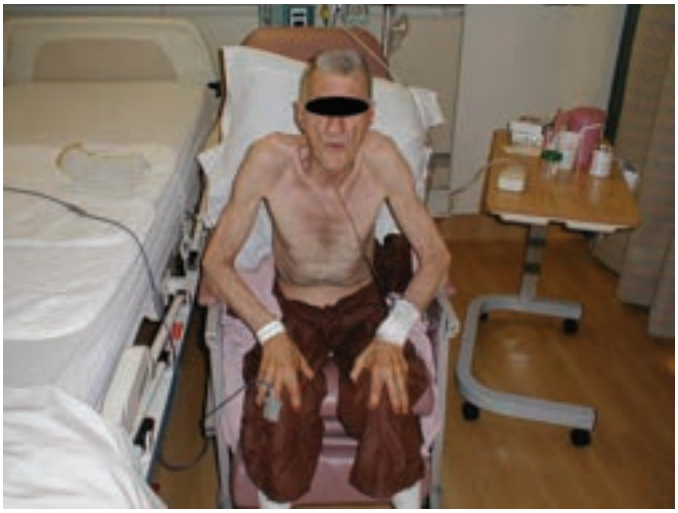

E-FIGURE 103-1. This patient with chronic obstructive pulmonary disease (COPD) demonstrates the posture referred to as the tripod sign. The patient is sitting forward with his hands on his knees to provide a mechanical advantage to the accessory muscles of respiration, such as the sternocleidomastoid and trapezius. Diaphragmatic flattening accompanying COPD lessens the diaphragm's ability to generate pressure, thereby resulting in increased dependence on the accessory muscles of respiration and leading to this posture.

metabolic acidosis (Equation 6). Specifically, a measured $Paco_2$ above the expected value indicates an inadequate ventilatory response, whereas a value of $Paco_2$ that falls within the expected range indicates an expected, appropriate ventilatory response to the metabolic derangement (i.e., the acidosis).

$$Paco_2 = (1.5[HCO_3^-] + 8) \pm 2 \quad (6)$$

When the patient is hypercapnic and hypoxemic, a useful step is to calculate the ambient air $P(A\text{-}a)o_2$ and to determine whether it is normal or increased for the patient's age. Of the six mechanisms of hypoxemia, only two (hypoventilation and breathing decreased ambient oxygen, as at altitude or from a hypoxic gas mixture) are associated with a preserved $P(A\text{-}a)o_2$ (Table 103-1). Under clinical circumstances at sea level, hypoxemia in the face of a normal $P(A\text{-}a)o_2$ indicates that the patient's hypoxemia is caused by hypoventilation and should prompt the clinician to consider the various causes of suppressed respiratory drive (Chapter 86) or respiratory insufficiency that interferes with a normal ventilatory response (e.g., respiratory muscle weakness; Chapter 421).

PULSE OXIMETRY

Pulse oximetry is a noninvasive method to assess arterial blood oxygenation.[1] The percentage of hemoglobin that is oxygenated is measured by passing light of two different wavelengths (660 nm [for deoxyhemoglobin] and 940 nm [for oxyhemoglobin]) through a blood-carrying tissue (e.g., finger, earlobe, forehead), identifying the pulsatile component (which contains arterial blood and background tissue elements), and subtracting the nonpulsatile component to isolate the arterial component. The device can estimate the percentage of oxygenated hemoglobin over the range of 100% to about 75%. Most clinicians regard the output of pulse oximeters to be inaccurate for percentage saturation values of less than 70%, although the probability of a low saturation should not be discounted (Fig. 103-1). Pulse oximetry measurements may help identify significant drops in Pao_2 below 60 to 65 mm Hg but are relatively insensitive to changes in Pao_2 from 90 to 65 mm Hg. The true value of pulse oximetry for decision-making in the emergency department setting remains uncertain.

TABLE 103-1 PHYSIOLOGIC MECHANISMS OF HYPOXEMIA AND ACCOMPANYING VALUES OF THE ALVEOLAR-ARTERIAL OXYGEN GRADIENT ON BREATHING OF ROOM AIR

MECHANISM/ PHYSIOLOGIC PROCESS	EXAMPLE	ALVEOLAR-ARTERIAL OXYGEN GRADIENT ON ROOM AIR
Ventilation-perfusion mismatch	Pneumonia	Increased
Diffusion impairment	Interstitial lung disease	Increased
Anatomic right-to-left shunt	Pulmonary arteriovenous malformation	Increased
Hypoventilation	Neuromuscular weakness	Normal
Breathing decreased ambient oxygen (from either hypobaric conditions [e.g., altitude] or breathing a gas mixture with decreased inspired oxygen fraction)	Altitude exposure	Normal
Diffusion-perfusion impairment	Hepatopulmonary syndrome	Increased

CARBON DIOXIDE MONITORING: CAPNOMETRY AND TRANSCUTANEOUS CARBON DIOXIDE MEASUREMENT

The fraction of carbon dioxide in exhaled air can be measured in real time by infrared capnometry.[2] Partial pressures can then be calculated on the basis of knowledge of atmospheric pressure. The expiratory capnogram (Fig. 103-1) represents a continuous plot of exhaled Pco_2 versus time or exhaled volume and reflects the sequential appearance of gas from various compartments (e.g., the endotracheal tube, central airways, and finally the alveoli, where the Pco_2 is in equilibrium with end-capillary blood). The shape of the capnogram provides clues to the presence of chronic obstructive pulmonary disease, in that emptying of areas of lung with increased dead space (see later) can cause the capnogram to have a rising contour (Fig. 103-1A), whereas the attainment of a so-called alveolar plateau on the normal capnogram (Fig. 103-1B) indicates that alveolar gas is composed of a mix with a relatively small contribution from areas of increased dead space. The value of $Peco_2$ measured at the end of expiration on the capnometer (i.e., the highest value recorded) represents the end-tidal $Petco_2$. Notably, the value of $Petco_2$ is always below the $Paco_2$ because there is a normal component of dead space ventilation (V_D/V_T) related to the anatomic dead space of the conducting airways (i.e., the trachea and airways to the level of gas-exchanging alveolar ducts and alveoli). The numerical difference between the $Paco_2$ and the mixed exhaled carbon dioxide tension ($Peco_2$, defined as the partial pressure of carbon dioxide that would be measured in a balloon in which the entire exhaled volume is gathered) is related to the magnitude of dead space ventilation (i.e., areas of the lung that are ventilated without accompanying blood flow, normally ~0.3 to 0.4) as defined by the Bohr equation:

$$V_D/V_T = (Paco_2 - Peco_2)/Paco_2 \quad (7)$$

The difference between $Paco_2$ and $Peco_2$ may be as low as several millimeters of mercury, but changing conditions of ventilation-perfusion matching (e.g., with pulmonary embolism [Chapter 98], atelectasis [Chapter 90]) may change the gradient over time. Measurement of the $Petco_2$ can be clinically useful to assess trends, to help detect esophageal intubation, to detect disconnection from the ventilator, and to detect perfusion during cardiopulmonary resuscitation, but it is not a reliable surrogate for $Paco_2$. Furthermore, measurement of the dead space fraction has prognostic value in patients with early acute respiratory distress syndrome (Chapter 104), in whom rising dead space is linearly related to increased mortality risk.

Measurement of transcutaneous Pco_2 by heated probes applied to the skin represents an alternative noninvasive method for estimating Pco_2. This

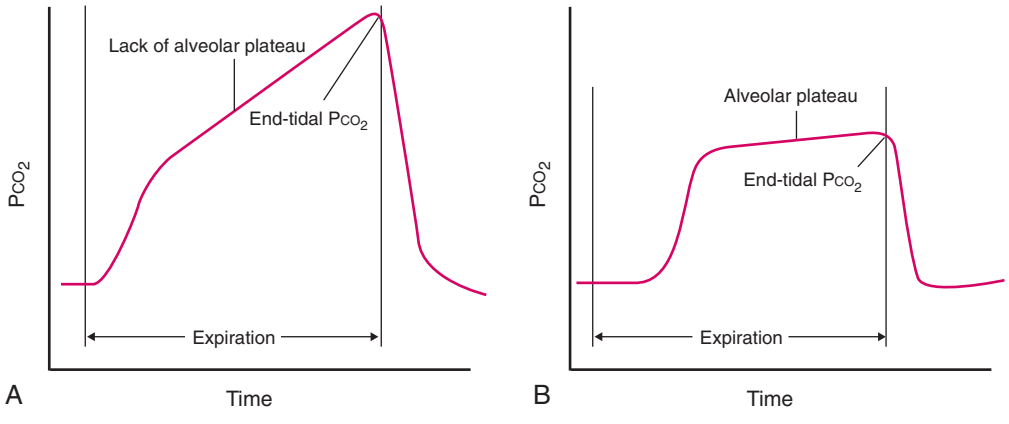

FIGURE 103-1. Abnormal and normal end-tidal capnograms. A, Illustration of a capnogram from a patient with chronic obstructive pulmonary disease in which the end-tidal Pco_2 rises throughout expiration as carbon dioxide excretion varies from different parts of the lung. **B,** Illustration of a normal capnogram in which the end-tidal Pco_2 reaches a plateau with more uniform carbon dioxide excretion. The end-tidal Pco_2 is the highest point of the alveolar plateau.

approach is less widely used clinically, at least in adults, because of technical requirements, such as site rotation for the probes and repetitive calibration, and its generally lower accuracy in estimating P_{CO_2}.

ARTERIAL OXYGEN CONTENT AND SYSTEMIC OXYGEN DELIVERY

Arterial (Ca_{O_2}) and venous oxygen content (Cv_{O_2}) are used to calculate cardiac output by the Fick equation (Equation 8), which is an alternative to determining cardiac output by the thermodilution method with a flow-directed pulmonary artery (Swan-Ganz) catheter (Chapter 57). The Fick equation is

$$\text{Oxygen consumption (mL O}_2\text{/min)} = \text{cardiac output} \times (Ca_{O_2} - Cv_{O_2}) \quad (8)$$

where oxygen content has the units of milliliters of oxygen per 100 mL of blood and is calculated as

$$\text{Oxygen content} = 1.34\,(\text{hemoglobin})\,(\%\ \text{saturation}) + 0.0031(Pa_{O_2}) \quad (9)$$

Under normal conditions (with, for example, an arterial percentage saturation of 95% and a hemoglobin level of 15 g/100 mL and an oxygen consumption of 250 mL/minute), arterial oxygen content is about 20 mL/100 mL, and because mixed venous oxygen saturation is about 75%, central venous oxygen content is about 15 mL/100 mL, making the normal arteriovenous oxygen content difference with a normal cardiac output about 5 mL/100 mL.

Systemic oxygen transport defines the amount of oxygen delivered to the tissues and multiplies the arterial oxygen content by the cardiac output:

$$\text{Systemic oxygen transport (mL/min)} = \text{cardiac output} \times Ca_{O_2} \quad (10)$$

where the normal value is about 1000 mL/minute.

MEASURING VENTILATION: MINUTE VENTILATION AND ALVEOLAR VENTILATION

Minute ventilation (V_E), which is the amount of gas exhaled from the airway per minute, is the product of the respiratory rate times the exhaled tidal volume, measured at body temperature and standardized to barometric pressure at sea level, saturated with water vapor (BTPS). The BTPS is a standard condition under which many measurements for most pulmonary function equipment and mechanical ventilators are made. These devices use an airflow meter to measure exhaled airflow and integrate the signal to derive tidal volume. An alternative way to measure tidal volume in an intensive care setting is respiratory impedance plethysmography, which uses calibrated magnetic coils in belts strapped around the chest and abdomen to monitor respiratory frequency and changes in thoracic volume.

Alveolar ventilation is the rate of gas delivery in liters per minute to gas-exchanging areas of the lung (i.e., the alveoli and alveolar ducts). The portion of minute ventilation that fails to undergo gas exchange is dead space ventilation (V_D) and is determined by Equation 7. Minute, alveolar (V_A), and dead space ventilation are related as follows:

$$V_E = V_A + V_D \quad (11)$$

It follows that conditions such as acute lung injury and acute respiratory distress syndrome (ARDS; Chapter 104) that are associated with very high dead space ratios require high V_E to achieve a sufficient V_A. Conversely, conditions that cause neuromuscular weakness (Chapter 421) are associated with small tidal volumes and have a high V_D/V_T ratio because the anatomic dead space is fixed and constitutes a higher fraction of the diminished tidal volume.

MEASURING CARBON DIOXIDE PRODUCTION

Measurement of carbon dioxide production is sometimes referred to as indirect calorimetry because it provides an index of metabolic rate and permits estimation of calorie requirements. Metabolic "carts" that simultaneously measure not only carbon dioxide production but also oxygen consumption and respiratory quotient are commonly used clinically to estimate metabolic needs to prescribe nutritional repletion (Chapter 216). The normal baseline carbon dioxide production is in the range of 200 mL/minute but is subject to wide variation because of hypermetabolic states commonly encountered in critically ill patients, such as sepsis and the systemic inflammatory response syndrome.

The respiratory quotient also gives insight into the composition of feedings because carbohydrates yield a respiratory quotient of 1, whereas fatty acids yield a ratio of 0.8 and amino acids a ratio of 0.7. Thus, balanced nutrition should yield a respiratory quotient of approximately 0.85. A respiratory quotient of 1 in combination with a high carbon dioxide production suggests that the dietary proportion of carbohydrates is excessive.

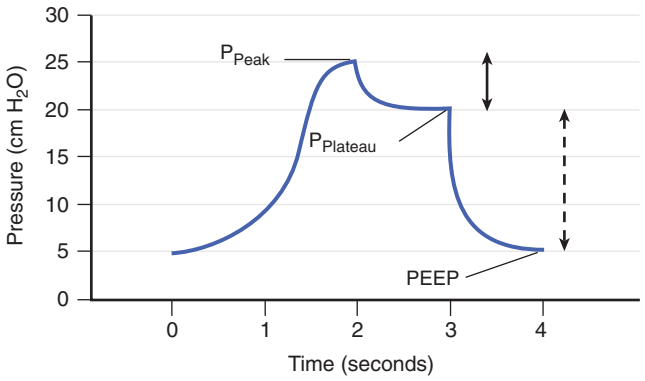

FIGURE 103-2. Illustration of inspiratory hold maneuver to determine plateau pressure ($P_{plateau}$). Airway pressure during volume-targeted mechanical ventilation rises as the tidal volume is delivered and reaches a peak. An inspiratory hold is initiated at peak pressure that prevents exhalation, so pressure falls to a "plateau" of about 20 cm H_2O. The drop in pressure reflects the pressure needed to overcome airway resistance. After slightly more than 1 second, the inspiratory hold is released, and airway pressure falls to positive end-expiratory pressure (PEEP). The difference between $P_{plateau}$ and PEEP is used to calculate static compliance by dividing the difference into the tidal volume.

MEASURING RESPIRATORY COMPLIANCE

Respiratory compliance is the change in respiratory system volume induced by a change in applied pressure (i.e., inspiratory pressure) and is the mathematical inverse of elastance. Compliance diminishes in conditions like lung injury and ARDS (Chapter 104) or pulmonary fibrosis (Chapter 92), in which diffuse inflammation and scarring alter lung structure and contribute to increased lung "stiffness." Static respiratory compliance is measured in patients receiving volume-limited mechanical ventilation by imposing a brief inspiratory hold at end inspiration. Assuming the patient has no spontaneous breathing effort, the airway pressure measured when airflow ceases is referred to as the plateau pressure ($P_{plateau}$). The difference between this pressure and the positive end-expiratory pressure (PEEP) is taken as the driving pressure required to deliver the tidal volume (Fig. 103-2). Static respiratory system compliance (C_{RS}) is then calculated as

$$C_{RS} = \Delta V(\text{exhaled tidal volume})/\Delta P\,(P_{plateau} - \text{PEEP}) \quad (12)$$

This compliance not only reflects the status of the lung but also includes contributions of the chest wall and abdomen. Thus, patients with chest wall deformities or morbid obesity have lower values of respiratory compliance even in the absence of lung abnormalities (Chapter 99). The normal respiratory compliance is in the range of 50 to 70 mL/cm H_2O, and patients with ARDS usually have values of C_{RS} of less than 30 cm H_2O. If respiratory compliance is below 20 to 25 cm H_2O, weaning from mechanical ventilation (Chapter 105) is difficult or impossible because of the high work of breathing requirements (see later).

MEASURING RESPIRATORY DRIVE

The respiratory center, located in the pons and medulla, regulates respiratory drive. Hypercapnia is a strong stimulus to ventilation (Chapter 86). This response may be blunted by chronic carbon dioxide retention or by drugs like narcotics. Hypoxemia is a weaker ventilatory stimulus that is potentiated by hypercapnia and blunted by hypocapnia.

Thus, respiratory drive can be assessed as the response to carbon dioxide in the blood in the hypercapnic ventilatory response. In one technique to measure respiratory drive, the patient rebreathes his or her exhaled air while minute ventilation and P_{ETCO_2} are monitored; a graph relating P_{ETCO_2} with minute ventilation is used to measure respiratory drive. However, this technique is impractical in an intensive care unit (ICU) setting. Another technique is to measure the negative swing in airway pressure during the first 100 msec of inspiration (P_{100}). This technique avoids the problem of diminished ventilatory response due to airway obstruction, but it is still subject to blunting by some drugs and still underestimates drive in patients with respiratory muscle weakness, a common problem in the ICU. In patients who are failing to be weaned from mechanical ventilation, a practical way to assess the integrity of respiratory drive is to determine whether the respiratory rate increases, usually into the range of 30 to 40 breaths/minute, as Pa_{CO_2} rises after the patient is removed from ventilatory support.

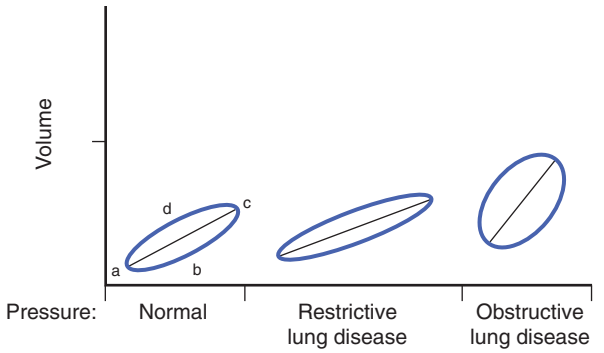

FIGURE 103-3. Pressure-volume curves illustrating components of work in a normal subject and in patients with restrictive or obstructive disease. The line between *a* and *c* represents elastic work as the lung expands, but this work is a net zero because static forces return the lung to its neutral position. The restrictive curve is flatter than normal because the lung is stiffer and volume changes less for a given unit change in pressure. The obstructive curve has a greater slope because (e.g., in emphysema) the lung is more compliant and starts inhalation from a higher volume. The *abc* curve represents resistive work during inspiration, and *cda* represents resistive work during exhalation. Resistive work during exhalation is greater in patients with obstructive lung disease.

MEASURING RESPIRATORY MUSCLE STRENGTH

Respiratory muscle weakness has long been recognized as a contributor to respiratory failure and failure to be weaned from mechanical ventilation in the ICU (Chapter 105). This recognition has intensified in recent years with the increased awareness of ICU-acquired weakness after critical illness. However, measurement of respiratory muscle strength remains challenging because of the need to differentiate between actual weakness and reduced muscle performance due to inability to cooperate or to exert a full inspiratory effort.

The most commonly used measures of respiratory muscle strength are the maximal inspiratory and expiratory pressures (PImax or MIP and PEmax or MEP). These values are obtained by measuring the pressure change with a manometer when the patient inhales with maximal force from residual volume and exhales with maximal force from total lung capacity. Normal MIP is usually more negative than −75 cm H_2O, and normal MEP is usually more positive than 125 cm H_2O. When the value for MIP is less negative than −20 or −30 cm H_2O, weaning from mechanical ventilation may be difficult, and values less positive than 60 cm H_2O suggest cough insufficiency.[3] However, these values have poor predictive value in mechanically ventilated patients because many of these patients are unable to cooperate. This problem may be addressed by attaching a one-way valve to the end of an endotracheal tube that permits exhalation but not inhalation and then measuring the inspiratory pressure efforts for 20 to 25 seconds.

MEASURING WORK OF BREATHING

Work of breathing is the product of pressure and volume for each breath (Fig. 103-3). The components include work needed to overcome elastic recoil of the lung and to displace the chest wall and abdomen as well as work needed to overcome airway resistance and lung viscosity and work needed to overcome inertia. With restrictive lung diseases, the inspiratory work of breathing is increased because of the decreased lung elasticity. With obstructive diseases, the work of breathing is increased because of increased airway resistance.

In clinical settings, a more practical way to assess the inspiratory work of breathing is to calculate the pressure-time product (in cm H_2O-seconds). The pressure-time product can be calculated by the decrease in airway pressure during inspiration, esophageal pressure (measured with an esophageal balloon manometer), or transdiaphragmatic pressure (measured with esophageal and gastric balloon manometers) as an index of diaphragmatic work. The work can be calculated as work of breathing per breath or as work of breathing per minute by multiplying the work per breath by the respiratory frequency. Commercially available devices using esophageal manometry automatically calculate the inspiratory work of breathing, which may be of some value in assessing the likelihood of weaning from mechanical ventilation. If the drop in inspiratory pressure necessary to achieve an adequate tidal volume is too large, the calculated work of breathing will be high, and the likelihood of successful weaning will be reduced.

Grade A **Grade A Reference**

A1. Schuh S, Freedman S, Coates A, et al. Effect of oximetry on hospitalization in bronchiolitis: a randomized clinical trial. *JAMA*. 2014;312:712-718.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Pretto JJ, Roebuck T, Beckert L, et al. Clinical use of pulse oximetry: official guidelines from the Thoracic Society of Australia and New Zealand. *Respirology*. 2014;19:38-46.
2. Ortega R, Connor C, Kim S, et al. Monitoring ventilation with capnography. *N Engl J Med*. 2012;367:e27.
3. McConville JF, Kress JP. Weaning patients from the ventilator. *N Engl J Med*. 2013;368:1068-1069.

REVIEW QUESTIONS

1. $Paco_2$ (arterial carbon dioxide tension) is determined by

A. CO_2 production rate and minute ventilation
B. Carbonic anhydrase level and alveolar ventilation
C. CO_2 production rate and alveolar ventilation
D. Diffusing capacity and CO_2 production rate
E. Minute ventilation and dead space (V_D/V_T)

Answer: C The $Paco_2$ is determined by the rate of CO_2 production and the alveolar ventilation.

2. A patient's ambient air arterial blood gas shows a Pao_2 of 60 torr and a $Paco_2$ of 60 torr (at sea level). What is this patient's alveolar-arterial oxygen gradient?

A. 4 torr
B. 14 torr
C. 24 torr
D. 34 torr
E. 37 torr

Answer: B The alveolar-arterial oxygen gradient at sea level (barometric pressure = 760 mm Hg) is calculated as $149 - ([1.25]\ Paco_2 + Pao_2)$. The alveolar-arterial oxygen gradients with these arterial blood gas values = 14 torr (or mm Hg).

3. Variables that affect the ratio of dead space to tidal volume (V_D/V_T) include

A. End-tidal CO_2 and $Paco_2$
B. Exhaled CO_2 tension (P_{ECO_2}) and Pao_2
C. Pao_2 and CO_2 production rate
D. Exhaled CO_2 tension (P_{ECO_2}) and $Paco_2$
E. Exhaled CO_2 tension (P_{ECO_2}) and end-tidal CO_2 tension

Answer: D The V_D/V_T is calculated by the Bohr equation, which is $(Paco_2 - P_{ECO_2})/Paco_2$.

104

ACUTE RESPIRATORY FAILURE

MICHAEL A. MATTHAY AND ARTHUR S. SLUTSKY

DEFINITION

Acute respiratory failure occurs when dysfunction of the respiratory system results in abnormal gas exchange that is potentially life-threatening. Each element of this definition is important to understand. The term *acute* implies a relatively sudden onset (from hours to days) and a substantial change from the patient's baseline condition. *Dysfunction* indicates that the abnormal gas exchange may be caused by abnormalities in any element of the respiratory system (e.g., a central nervous system abnormality affecting the regulation of breathing or a musculoskeletal thoracic abnormality affecting ventilation [Chapter 83]) in addition to abnormalities of the lung itself. The term *respiration* refers, in a broad sense, to the delivery of oxygen (O_2) to metabolically active tissues for energy use and the removal of carbon dioxide (CO_2) from these tissues (Table 104-1). Respiratory failure is a failure of the process of delivery of O_2 to the tissues or removal of CO_2 from the tissues. Abnormalities in the periphery (e.g., cyanide poisoning, circulatory shock, pathologic distribution of organ blood flow in sepsis) can lead to tissue hypoxia; although these conditions represent forms of respiratory failure in the broadest terms, this chapter focuses on respiratory failure resulting from dysfunction of the lungs, chest wall, and control of respiration.

PATHOBIOLOGY

Abnormal gas exchange is the physiologic hallmark of acute respiratory failure, which can be classified in several ways (Table 104-2). Although gas exchange can be abnormal for either oxygenation or CO_2 removal, significant hypoxemia is nearly always present when patients with acute respiratory failure breathe ambient air. If CO_2 is retained at a potentially life-threatening level under these conditions, it must be accompanied by significant hypoxemia (see later). The *life-threatening* aspect of the condition places the degree of abnormal gas exchange in a clinical context and calls for urgent treatment.

The diagnosis of acute respiratory failure requires a significant change in arterial blood gases from baseline. Many patients with chronic respiratory problems can function with blood gas tensions that would be alarming in a physiologically normal individual. Over time, patients with so-called chronic respiratory failure or chronic respiratory insufficiency develop mechanisms to compensate for inadequate gas exchange. Conversely, this chronic condition makes patients vulnerable to insults that could be easily tolerated by a previously healthy individual.

In acute respiratory failure, the O_2 content in the blood (available for tissue use) is reduced to a level at which the possibility of end-organ dysfunction increases markedly. The value of the partial pressure of O_2 in the arterial blood (Pao_2) that demarcates this vulnerable zone is often considered to be the point of the oxyhemoglobin dissociation relationship at which any further decrease in the Pao_2 results in sharp decreases in the amount of hemoglobin saturated with O_2 (Sao_2) and in the arterial blood O_2 content (Cao_2). Thus, acute respiratory failure is often defined in practice as occurring when the Pao_2 is less than about 55 mm Hg (Fig. 104-1). The oxyhemoglobin dissociation curve of venous blood, which is the partial pressure at which O_2 is being unloaded to the tissues, is a critical determinant of how much O_2 is available for the cells and their mitochondria. Other than under conditions of an extremely hypoxic environment (e.g., in utero or on the summit of Mt. Everest), the enhanced ability to unload O_2 at the tissue level more than compensates for small decreases in the amount of O_2 picked up in the lungs when the oxyhemoglobin dissociation curve is shifted rightward. With a leftward shift in the curve, O_2 is bound more tightly to hemoglobin, so less O_2 is available for tissue delivery.

These clinical considerations imply that any definition of acute respiratory failure based on an absolute level of Pao_2 is arbitrary. A healthy, young, conditioned individual climbing at high altitude may have a Pao_2 of less than

TABLE 104-1 ABBREVIATIONS COMMONLY USED IN ACUTE RESPIRATORY FUNCTION

ABG	Arterial blood gas or arterial blood gas analysis
ALI	Acute lung injury
ARDS	Acute respiratory distress syndrome
ARF	Acute respiratory failure
cm H_2O	Centimeters of water
Cao_2	Content of oxygen in arterial blood
Cco_2	Content of oxygen in end-capillary blood
CO_2	Carbon dioxide
COPD	Chronic obstructive pulmonary disease
CPAP	Continuous positive airway pressure (used when positive pressure during exhalation is applied with spontaneous ventilation)
Cvo_2	Content of oxygen in mixed venous blood
Fio_2	Fraction of inspired oxygen
g/dL	Grams per deciliter
HbO_2	Saturation of hemoglobin by oxygen
L/min	Liters per minute
mL/kg	Milliliters per kilogram
mL/min	Milliliters per minute
mm Hg	Millimeters of mercury
NIPPV	Noninvasive positive-pressure ventilation
O_2	Oxygen
$P(A\text{-}a)o_2$	Difference of partial pressure of oxygen between mean alveolar gas and arterial blood (alveolar-to-arterial oxygen difference)
$Paco_2$	Partial pressure of carbon dioxide in alveolar gas
$Paco_2$	Partial pressure of carbon dioxide in arterial blood
PAo_2	Partial pressure of oxygen in alveolar gas
Pao_2	Partial pressure of oxygen in arterial blood
Pao_2/Fio_2	Ratio of partial pressure of oxygen in arterial blood to fraction of inspired oxygen
PBW	Predicted body weight
Pc_{CO_2}	Partial pressure of carbon dioxide in end-capillary blood
P_{CO_2}	Partial pressure of carbon dioxide
Pc_{O_2}	Partial pressure of oxygen in end-capillary blood
PEEP	Positive end-expiratory pressure (used when positive pressure during exhalation is applied with mechanical ventilation)
P/F	Pao_2/Fio_2 ratio
Pio_2	Partial pressure of oxygen in inspired gas
Po_2	Partial pressure of oxygen
$Pvco_2$	Partial pressure of carbon dioxide in mixed venous blood
Pvo_2	Partial pressure of oxygen in mixed venous blood
$\dot{Q}$	Blood flow or perfusion
RR	Respiratory rate
Sao_2	Percentage of saturation of hemoglobin by oxygen in arterial blood
$\dot{V}$	Ventilation
$\dot{V}/\dot{Q}$	Ventilation-perfusion ratio
VT	Tidal volume

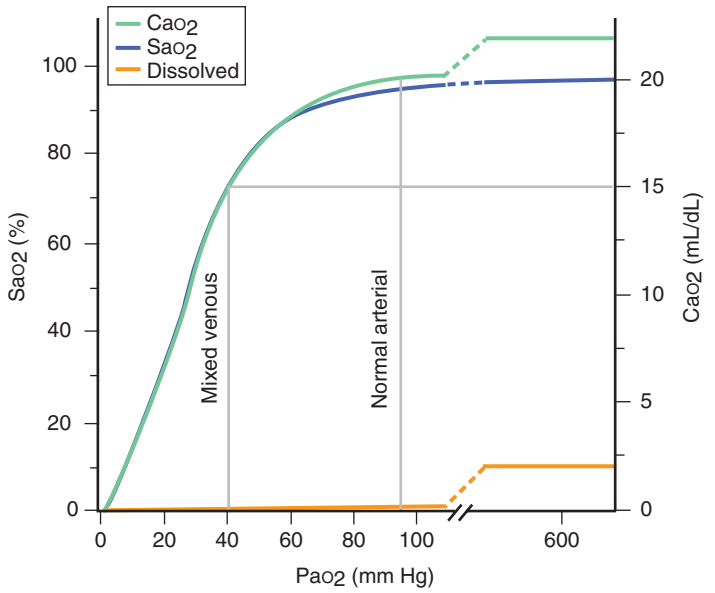

FIGURE 104-1. Oxyhemoglobin association-dissociation curve. The axis for oxygen saturation in the arterial blood (Sao_2) is on the left, and the axis for arterial content of oxygen (Cao_2) is on the right. Cao_2 is the sum of the oxygen dissolved in plasma (denoted as "Dissolved" in the figure) plus the oxygen bound to hemoglobin. With a normal hemoglobin, most of the oxygen is carried in combination with hemoglobin, with only a relatively small amount of oxygen dissolved in plasma. When the value of the arterial partial pressure of oxygen (Pao_2) is on the "flat" portion of the curve ($Pao_2 \geq 60$ to 65 mm Hg, normal partial pressure of arterial carbon dioxide [$Paco_2$], and normal pH), raising the Pao_2 further has relatively little effect on total oxygen content. Increases in temperature, Pco_2, hydrogen ion concentration, or 2,3-diphosphoglycerate cause a rightward shift in the oxyhemoglobin association-dissociation curve.

50 mm Hg because of the reduction in inspired O_2 pressure.[1] This individual is not in acute respiratory failure, even though the Pao_2 may be in the low 40s. A patient who has chronic obstructive pulmonary disease (COPD) and whose usual range of Pao_2 is 50 to 55 mm Hg would not be considered to be in acute respiratory failure if the Pao_2 was 50 mm Hg. However, if a patient's usual Pao_2 is 80 mm Hg, a sudden drop to a Pao_2 of 50 mm Hg could be associated with a substantial risk for a further life-threatening reduction in oxygenation; this patient should be considered to have acute respiratory failure.

Traditionally, the level of arterial CO_2 partial pressure ($Paco_2$) that defines acute respiratory failure has been 50 mm Hg or greater, if it is accompanied by arterial acidosis with a pH of less than about 7.30. The $Paco_2$ is linked to pH in this definition because of the general belief that acidosis is what leads to tissue dysfunction and symptoms. Patients with severe COPD may have chronic CO_2 retention, but renal compensation for the respiratory acidosis protects them against abnormalities related to the elevation in CO_2. A further acute rise in $Paco_2$ can precipitate symptoms and other organ dysfunction; however, even severe respiratory acidosis (pH 7.1) seems to be better tolerated than metabolic acidosis of the same pH in most previously healthy individuals if arterial and tissue oxygenation is adequate.

Pathophysiology

Five mechanisms can lead to a reduction in Pao_2: (1) decreased inspired partial pressure of O_2 (Pio_2) (e.g., at high altitude or when breathing a reduced percentage O_2 mixture); (2) hypoventilation; (3) ventilation-perfusion ($\dot{V}/\dot{Q}$) mismatch; (4) shunting of blood from the pulmonary to systemic circulation, bypassing the alveoli anatomically or functionally; and (5) any barrier for diffusion of O_2 from the alveoli into the capillary blood. In essence, a shunt is an extreme $\dot{V}/\dot{Q}$ mismatch in which blood perfuses alveoli with *no* ventilation; it is differentiated clinically from other $\dot{V}/\dot{Q}$ mismatching by the response to breathing of supplemental O_2 (see later).

For clinical purposes, diffusion abnormalities are not usually important causes of hypoxemia at sea level because there is sufficient time for adequate diffusion of O_2 during the transit of a red blood cell through the pulmonary capillary bed, even in the presence of severe lung disease. When diffusion abnormalities are present and contribute to hypoxemia, $\dot{V}/\dot{Q}$ mismatch nearly always coexists with the shunting, and this mismatch is an important cause of hypoxemia. Except at high altitude or when a subject is breathing a gas mixture low in O_2, hypoventilation, $\dot{V}/\dot{Q}$ mismatch, and shunting are the dominant causes of hypoxemia.

If only hypoventilation is present, the resulting hypoxemia is associated with a normal difference between the calculated alveolar and the measured arterial oxygenation levels [$P(A\text{-}a)o_2$]. In this setting, an elevated $Paco_2$ suggests disease processes that affect nonpulmonary respiratory function (e.g., central respiratory depression resulting from drug overdose, neuromuscular diseases such as Guillain-Barré syndrome, or chest wall disease such as flail chest; Chapter 86). In contrast, $\dot{V}/\dot{Q}$ mismatch and shunting are associated with an elevated $P(A\text{-}a)o_2$, which may or may not coexist with hypoventilation. The normal value for $P(A\text{-}a)o_2$ varies as a function of the fraction of inspired O_2 (Fio_2), increasing as Fio_2 increases.

TABLE 104-2 SYSTEMS TO CLASSIFY ACUTE RESPIRATORY FAILURE

HYPOXIC VERSUS HYPERCAPNIC-HYPOXEMIC

Causes of Hypoxemic Acute Respiratory Failure

Acute lung injury/ARDS
Pneumonia
Pulmonary thromboembolism
Acute lobar atelectasis
Cardiogenic pulmonary edema
Lung contusion
Acute collagen vascular disease (Goodpasture syndrome, systemic lupus erythematosus)

Causes of Hypercapnic-Hypoxemic Acute Respiratory Failure

Pulmonary disease
  COPD
  Asthma: advanced, acute, severe asthma
Drugs causing respiratory depression
Neuromuscular
  Guillain-Barré syndrome
  Acute myasthenia gravis
Spinal cord tumors
Metabolic derangements causing weakness (including hypophosphatemia, hypomagnesemia)
Musculoskeletal
  Kyphoscoliosis
  Ankylosing spondylitis
Obesity hypoventilation syndrome (often with additional acute, superimposed abnormality as cause of acute respiratory failure)

ETIOLOGIC MECHANISMS OF HYPOXEMIA

Normal $P(A\text{-}a)O_2$*

↓P_{IO_2}
  High altitude; inadvertent administration of low F_{IO_2} gas mixture
  Hypoventilation
  See causes of hypercapnic-hypoxic acute respiratory failure above

Increased $P(A\text{-}a)O_2$*

Ventilation-perfusion ($\dot{V}/\dot{Q}$) mismatch
  Airway disease
  Vascular disease, including pulmonary thromboembolism
Shunt
  Acute lung injury/ARDS
  Pneumonia
  Parenchymal lung disease
  Cardiogenic pulmonary edema
  Pulmonary infarction
Diffusion limitation†

ACUTE RESPIRATORY FAILURE WITH AND WITHOUT CHRONIC LUNG DISEASE

With Chronic Lung Disease

COPD
Asthma
Parenchymal lung diseases
Restrictive lung/chest wall diseases

Without Chronic Lung Disease‡

Acute lung injury/ARDS
Pneumonia
Pulmonary thromboembolism

ACUTE RESPIRATORY FAILURE BY ORGAN SYSTEM INVOLVED

Respiratory (Lungs and Thorax)

Airway/airflow obstruction
  COPD
  Asthma
Pulmonary parenchyma
  Pneumonia
  ARDS
  Acute flare of chronic collagen vascular disease (e.g., Goodpasture syndrome, systemic lupus erythematosus)

Central Nervous System

Respiratory depression
  Increased sedatives, tranquilizers with respiratory effect; opiates; alcohol
Brain stem and spinal cord involvement
  Tumors, trauma, vascular accidents

Neuromuscular

Guillain-Barré syndrome
Myasthenia gravis

Cardiovascular

Cardiogenic pulmonary edema
Pulmonary thromboembolism

Renal/Endocrine

Volume overload
Metabolic abnormalities

*Calculated by the alveolar-air equation; see text for description.
†See text for discussion.
‡These can also be superimposed on chronic disease.
ARDS = acute respiratory distress syndrome; COPD = chronic obstructive pulmonary disease; F_{IO_2} = fraction of inspired oxygen; $P(A\text{-}a)O_2$ = alveolar-to-arterial oxygen difference; P_{IO_2} = partial pressure of inspired oxygen; $\dot{V}/\dot{Q}$ = ventilation-perfusion ratio.

When $\dot{V}/\dot{Q}$ mismatch or shunting is the cause of hypoxemia, some alveolar regions have increased levels of P_{CO_2} and associated reduced levels of P_{O_2}; the blood in the vessels perfusing these alveoli reflects these abnormal gas tensions. The resulting increased arterial P_{CO_2} (Pa_{CO_2}) usually can be reversed by increasing overall ventilation, but this increased ventilation usually does not correct the decreased arterial P_{O_2} (Pa_{O_2}).

$\dot{V}/\dot{Q}$ mismatch is distinguished from shunting by assessing the Pa_{O_2} response to enhanced O_2 administration. Hypoxemia caused by $\dot{V}/\dot{Q}$ mismatch can be corrected to a nearly complete O_2 saturation of the hemoglobin in most patients by a relatively small increase in F_{IO_2}, such as from 0.24 to 0.28 by face mask or 1 to 2 L/minute O_2 by nasal prongs, in patients with acute exacerbations of COPD. If the airways to poorly ventilated alveoli remain open and the enriched O_2 mixture is administered for an adequate length of time (ranging from a few minutes to about 20 minutes, depending on the degree of $\dot{V}/\dot{Q}$ inequality), the increased P_{IO_2} is reflected by an increased Pa_{O_2} and an increased Pa_{O_2}. When a shunt is present (no ventilation but continued perfusion), a relatively small increase in the F_{IO_2} has little or no effect on the Pa_{O_2}, and even large increases in F_{IO_2} up to 1.0 result in only modest increases in Pa_{O_2} (Fig. 104-2).

CLINICAL MANIFESTATIONS

The hallmark of acute respiratory failure is the inability to maintain adequate oxygenation or the inability to maintain an appropriate Pa_{CO_2}. Patients are typically dyspneic and tachypneic, unless progressive respiratory failure causes fatigue—sometimes leading to respiratory arrest—or a drug overdose or neuromuscular condition prevents an appropriate respiratory response to hypoxemia or hypercapnic acidosis. Neurologic function may deteriorate, and myocardial ischemia or even infarction may be precipitated by hypoxemia. In addition, each cause has its own specific manifestations (see later).

DIAGNOSIS

As part of the diagnosis of acute respiratory failure, the physician has three objectives: (1) to confirm the clinical suspicion that acute respiratory failure is present, (2) to classify the type of acute respiratory failure (e.g., hypoxemia caused by hypoventilation vs. hypoxemia caused by $\dot{V}/\dot{Q}$ mismatch or shunting), and (3) to determine the specific cause (e.g., the acute respiratory distress syndrome [ARDS]) secondary to pulmonary or nonpulmonary sepsis or decompensated COPD because of acute bronchitis. Defining the type of acute respiratory failure and determining the specific cause are prerequisites to optimal management.

The initial approach to diagnosis consists of considering information from four sources: (1) clinical history and physical examination; (2) physiologic abnormalities, particularly arterial blood gas derangements, which help establish the mechanisms of hypoxemia; (3) chest radiographic findings; and (4) other tests aimed at elucidating specific causes. In many cases, the clinical picture from the history is so clear that the presumptive type of acute respiratory failure (and sometimes the cause) is obvious, so treatment can be started

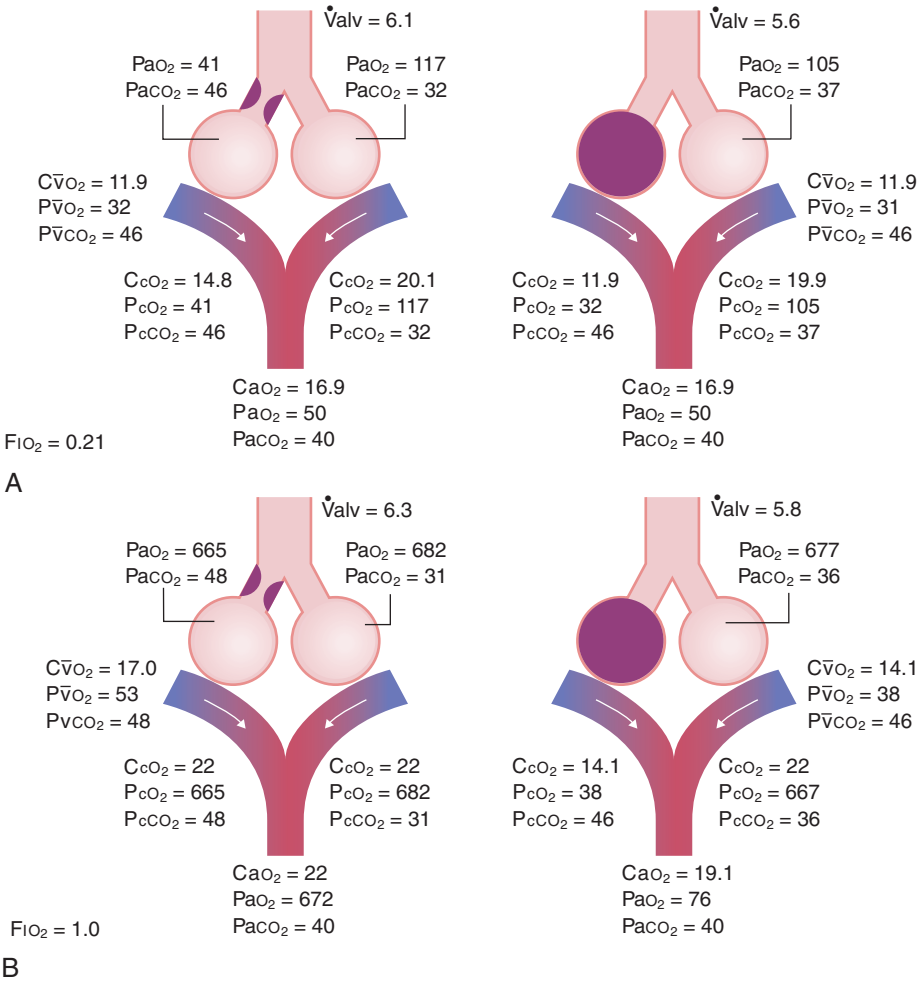

FIGURE 104-2. Arterial oxygenation. Comparison of the effect on arterial oxygenation of increasing the fraction of inspired oxygen (F_{IO_2}) from breathing of ambient air (F_{IO_2} = 0.21) (**A**) and breathing of 100% oxygen (F_{IO_2} = 1.0) (**B**) with a low ventilation-perfusion ratio ($\dot{V}/\dot{Q}$) (*left*) and a shunt (*right*), using a two-compartment lung model. Shunting and decreased $\dot{V}/\dot{Q}$ can lead to identical arterial blood gases (partial pressure of oxygen in arterial blood [P_{aO_2}] = 50 mm Hg; partial pressure of carbon dioxide in arterial blood [P_{aCO_2}] = 40 mm Hg). The response to supplemental oxygen administration is markedly different. Hypoxemia is only partially corrected by breathing of 100% oxygen when a shunt is present because arterial oxygenation represents an average of the end-capillary oxygen content (C_{cO_2}) from various parts of the lung, not an average of the partial pressures of oxygen (partial pressure of carbon dioxide in the end-capillary blood [P_{cCO_2}]). When the C_{cO_2} values are mixed, the P_{aO_2} is determined from the resultant content of oxygen in the arterial blood (C_{aO_2}) by the oxyhemoglobin association-dissociation relationship (see Fig. 104-1). With low $\dot{V}/\dot{Q}$ (as is often the case in patients with chronic obstructive pulmonary disease), an increase in F_{IO_2} increases the alveolar partial pressure of oxygen (P_{O_2}) of the low $\dot{V}/\dot{Q}$ unit and leads to a marked increase in arterial P_{O_2}. The values in this figure were generated from modeling to result in the same P_{aCO_2} (40 mm Hg) for all four situations shown; this is the reason for slight changes in alveolar ventilation ($\dot{V}$alv) for some of the conditions. Several assumptions are made: no diffusion limitation is present; oxygen consumption = 300 mL/minute, and CO_2 production = 240 mL/minute; cardiac output = 6.0 L/minute; the low $\dot{V}/\dot{Q}$ regions in the left panels represent 60% of the cardiac output perfusing alveoli with a $\dot{V}/\dot{Q}$ 25% of normal; and the shunts in the right panels represent a 37% shunt (i.e., 37% of the cardiac output is perfusing alveoli with no ventilation).

while confirmatory laboratory studies are ordered. In other cases, a clinician may be asked to see a patient because of an abnormal chest radiograph or abnormal arterial blood gases ordered by someone else and may elicit the pertinent history based on these clues. When the degree of hypoxemia is life-threatening, therapeutic decisions must be made quickly, even if data are limited. The clinician must obtain updated information continually and should view most therapeutic decisions as therapeutic trials, with careful monitoring to assess desired benefits and possible detrimental effects.

Clinical Evaluation

The presentation often reflects one of three clinical scenarios: (1) the effects of hypoxemia or respiratory acidosis, (2) the effects of primary (e.g., pneumonia) or secondary (e.g., heart failure) diseases affecting the lungs, and (3) the nonpulmonary effects of the underlying disease process. The clinical effects of hypoxemia and respiratory acidosis are manifested mainly in the central nervous system (e.g., irritability, agitation, changes in personality, depressed level of consciousness, coma) and the cardiovascular system (e.g., arrhythmias, hypotension, hypertension) (Table 104-3). In patients with underlying COPD (Chapter 88) with a gradual onset of acute respiratory failure, central nervous system abnormalities may be the major presenting findings. Cyanosis, which requires at least 5 g/dL of unsaturated hemoglobin to be detectable, may not be seen before serious tissue hypoxia develops, especially in patients with underlying anemia.

TABLE 104-3 CLINICAL MANIFESTATIONS OF HYPOXEMIA AND HYPERCAPNIA

HYPOXEMIA	HYPERCAPNIA
Tachycardia	Somnolence
Tachypnea	Lethargy
Anxiety	Restlessness
Diaphoresis	Tremor
Altered mental status	Slurred speech
Confusion	Headache
Cyanosis	Asterixis
Hypertension	Papilledema
Hypotension	Coma
Bradycardia	Diaphoresis
Seizures	
Coma	
Lactic acidosis*	

*Usually requires additional reduction in oxygen delivery because of inadequate cardiac output, severe anemia, or redistribution of blood flow.

Pulmonary symptoms and signs often reflect the respiratory disease causing the acute respiratory failure. Examples include cough and sputum with pneumonia (Chapter 97) or chest pain from pulmonary thromboembolism with infarction (Chapter 98). Dyspnea and respiratory distress are nonspecific reflections of the respiratory system's difficulty in meeting the increased demands from pulmonary and nonpulmonary diseases.

Physical findings may be associated with a particular pathologic lung process, such as pneumonia (Chapter 97), which often results in bronchial breathing and crackles on auscultation, or the crackles (rales) of cardiogenic pulmonary edema (Chapter 58). Abnormal findings may be minimal or absent in patients with ARDS or pulmonary thromboembolism (Chapter 98).

In some patients, the clinical picture is dominated by the underlying disease process, particularly with diseases that cause ARDS, such as sepsis (Chapter 108), severe pneumonia (Chapter 97), aspiration of gastric contents (Chapter 94), and trauma. In these conditions, the physical examination findings are often nonspecific, with no obvious clues except, for example, fever with sepsis or pneumonia and hypotension with septic shock.

Assessment of Physiologic Abnormalities

The clinical suspicion of acute respiratory failure must be addressed by arterial blood gas analysis to answer several questions.

Is hypoxemia present? The answer is based largely on the value of the Pa_{O_2} or Sa_{O_2}. The degree of the hypoxemia not only confirms the diagnosis of acute respiratory failure but also helps define its severity.

Is hypoventilation present? If the Pa_{CO_2} is elevated, alveolar hypoventilation is present.

Does the degree of hypoventilation fully explain the hypoxemia? If the $P(A\text{-}a)o_2$ is normal, hypoventilation fully explains the presence and degree of hypoxemia. When this is the case, the most likely causes of acute respiratory failure are central nervous system abnormalities or a chest wall abnormality. If the $P(A\text{-}a)o_2$ is increased but hypoventilation does not fully explain the hypoxemia, another condition must be present; common diagnoses include COPD (Chapter 88), severe asthma (Chapter 87), pneumonia (Chapter 97), and early stages of ARDS.

If hypoxemia exists without hypoventilation, an elevated $P(A\text{-}a)o_2$ should be confirmed, and the response to breathing of an enhanced O_2 mixture would answer this question: *Is the increase in $P(A\text{-}a)O_2$ the result of a $\dot{V}/\dot{Q}$ abnormality or of shunting?* If hypoxemia is primarily the result of a $\dot{V}/\dot{Q}$ abnormality, the likely cause is an airway disease, either COPD or acute severe asthma, or a vascular disease, such as pulmonary thromboembolism. If shunting is the major explanation for the hypoxemia, processes that fill the air spaces (e.g., cardiogenic pulmonary edema, noncardiogenic pulmonary edema or ARDS, or purulent pulmonary secretions in acute pneumonia) or, less commonly, an intracardiac or anatomic intrapulmonary shunt is the likely cause. Conditions that fill air spaces should be confirmed by abnormal findings on a chest radiograph; if the radiograph is normal, the possibility of intracardiac shunt (Chapter 69) or thromboembolism (Chapter 98) should be evaluated.

Chest Radiography

The chest radiograph in acute respiratory failure is likely to show one of three patterns (Fig. 104-3): (1) normal (or relatively normal), (2) localized alveolar filling opacities, or (3) diffuse alveolar filling opacities. Diffuse interstitial opacities are also possible, but diseases that cause this pattern usually have a more gradual onset and are associated with chronic respiratory failure. If the chest radiograph is normal (i.e., it is clear or relatively clear), airway diseases, such as COPD and asthma, or pulmonary vascular diseases, such as thromboembolism, are more likely. If a localized alveolar filling abnormality is present, pneumonia is the major consideration, but pulmonary embolism and infarction should also be considered. When diffuse (bilateral) alveolar filling abnormalities are present, cardiogenic pulmonary edema and ARDS (e.g., as seen after sepsis, trauma, pneumonia, or aspiration of gastric contents) are the major considerations. The combination of the chest radiograph and the arterial blood gas interpretation can be helpful. The finding of a significant shunt may suggest ARDS in a patient in whom this diagnosis was not clinically obvious; the chest radiograph should help confirm that possibility.

Other Evaluations

All patients with acute respiratory failure should have a complete blood count including a platelet count, routine blood chemistry tests, prothrombin time, and urinalysis to screen for possible underlying causes and comorbid conditions. Other blood tests should be guided by the clinical picture. Examples include a serum amylase level if pancreatitis is a possible cause of ARDS and thyroid indices if severe hypothyroidism is a possible cause of hypoventilation. Blood cultures are recommended when an infectious cause such as sepsis is suspected.

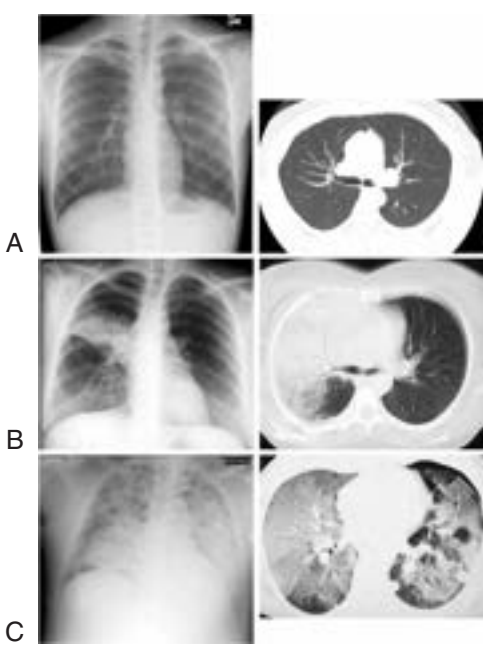

FIGURE 104-3. Chest radiographs (*left*) and computed tomography scans (*right*) of the three most common findings in diseases causing acute respiratory failure. A, Relatively clear chest, consistent with an acute exacerbation of airway disease (e.g., asthma, chronic obstructive pulmonary disease) or a central nervous system or neuromuscular disease as the cause of acute respiratory failure. **B,** Localized alveolar filling opacity, most commonly seen with acute pneumonia. **C,** Diffuse bilateral alveolar filling opacities consistent with acute lung injury and acute respiratory distress syndrome. The computed tomography scan in **C** shows a small left pneumothorax and cavities or cysts that are not apparent on the anteroposterior chest radiograph.

Any abnormal fluid collections, especially pleural effusion (Chapter 99), should be aspirated for diagnostic purposes. Sputum Gram stain and culture are indicated when pneumonia is suspected.

Other specific tests should be directed by the history, physical examinations, arterial blood gas levels, and chest radiograph. An abdominal computed tomography (CT) scan may be indicated to search for the source of infection in a patient with sepsis and ARDS. A chest CT scan may help define pulmonary disease if the chest radiograph is not definitive. CT arteriography of the pulmonary circulation may diagnose pulmonary thromboembolism (Chapter 98). A head CT scan may be indicated if a stroke involving the respiratory center is suspected. Routine blood chemistry studies can detect diabetic ketoacidosis or renal failure as contributing causes.

TREATMENT

Rx

General Measures

The management of acute respiratory failure depends on its cause, its clinical manifestations, and the patient's underlying status. Certain goals apply to all patients: improvement of the hypoxemia to eliminate or markedly reduce the acute threat to life; improvement of the acidosis if it is considered life-threatening; maintenance of cardiac output or improvement if cardiac output is compromised; treatment of the underlying disease process; and avoidance of predictable complications.

The precise methods for improving hypoxemia depend on the cause of the acute respiratory failure. However, an increase in the inspired O_2 concentration is a cornerstone of treatment for nearly all patients, even though it may not produce a marked increase in Pa_{O_2} in patients whose underlying pathophysiologic process involves a significant amount of lung with low ventilation-perfusion ratios or true shunting.

The level of acidosis that requires treatment other than for the underlying disease process is a matter of debate. Although normalization of the arterial pH was suggested in the past, respiratory acidosis is apparently well tolerated in many patients with severe ARDS, so a patient with a pH of 7.15 or higher may not require bicarbonate therapy. If the acidemia coexists with clinical complications, such as cardiac arrhythmias or a decreased level of consciousness, that have no other obvious cause, treatments to increase pH should be

considered. The therapeutic goal is alleviation or reduction of the accompanying complications by improving the level of acidosis; normalization of the pH usually is not indicated (Chapter 118).

The maintenance of cardiac output is crucial for O_2 delivery in acute respiratory failure, especially because mechanical ventilation and positive end-expiratory pressure (PEEP) may compromise cardiac output. Placement of a pulmonary artery catheter allows measurement of cardiac output and filling pressures, but most patients who have these catheters do no better than similar patients managed without them.[1] Nevertheless, selective use of diagnostic pulmonary artery catheterization can help determine the cause of the pulmonary edema (cardiogenic vs. noncardiogenic) and the physiologic basis for shock (sepsis, hypovolemia, or decreased cardiac output from impaired cardiac function) in selected patients in whom either is not clear.[2]

Many therapeutic interventions that improve short-term physiologic variables may worsen long-term, clinically important outcomes. For example, transfusing all patients to maintain a hemoglobin greater than 10 g/dL *increases* mortality in critically ill patients who have not had an acute myocardial infarction and do not have unstable angina, even though the O_2-carrying capacity of the blood is acutely increased. Use of a relatively large tidal volume (e.g., 12 mL/kg predicted body weight, which is equivalent to approximately 10 to 10.5 mL/kg measured body weight in patients who are somewhat overweight) *increases* mortality in patients with ARDS compared with a lower tidal volume (6 mg/kg predicted body weight), even though it raises Pao_2 more in the short term than does a lower tidal volume. Conservative use of fluids when vasopressors are no longer required to support the systemic blood pressure improves lung function and shortens the duration of mechanical ventilation and intensive care.[1]

Improvements in oxygenation, acid-base status, and cardiac output are of no more than temporary benefit unless the underlying disease processes are diagnosed and treated properly. In patients with ARDS, sepsis may worsen injury to the lung and other organs despite optimal supportive care. Similarly, if the precipitating cause of acute respiratory failure in a patient with COPD is not identified and treated, supportive care is likely to be futile. Complications may arise from the physiologic effects of the gas exchange abnormality, from the disease processes causing the acute respiratory failure, from being critically ill and its associated incursions on homeostasis (e.g., sleep deprivation), or from iatrogenic complications of therapy.

Mechanical Therapy to Improve Oxygenation

A Pao_2 greater than 60 mm Hg is usually adequate to produce an Sao_2 in the low to middle 90s. The Pao_2 can be increased by the administration of supplemental O_2, by pharmacologic manipulations, by continuous positive airway pressure (CPAP), by mechanical ventilation with or without maneuvers such as PEEP, and by the prone position. PEEP, pharmacologic manipulations, and positioning are used primarily in patients with ARDS (see later).

The initial choice of the concentration and amount of supplemental O_2 is based on the severity of the hypoxemia, the clinical diagnosis, the likely mechanism causing the hypoxemia, and the O_2 delivery systems available. For the tracheal FIo_2 to be the same as the delivered FIo_2, the O_2 delivery system must deliver a flow that matches the patient's peak inspiratory flow rate with gas of a known FIo_2. High-flow O_2 blenders can achieve this goal by delivering gas at 80 L/minute or more to a nonintubated patient. These systems require a large flow of O_2 (from a wall unit or tank), however, and are not universally available. Other systems for nonintubated patients (including nasal prongs, simple face masks, and non-rebreather and partial rebreather masks) use a simple regulator that mixes room air with O_2 from a wall unit or tank, with resulting flows that are frequently unable to match the patient's peak inspiratory flow rate. The patient entrains more air from the environment, and the resulting tracheal FIo_2 or partial pressure of oxygen in inspired gas (Pao_2) is unknown. The amount of air entrained depends on the patient's inspiratory pattern and minute ventilation. Although the resulting FIo_2 is unknown, these systems are satisfactory if the delivery is constant and if they result in adequate arterial O_2 saturation, as monitored by arterial blood gases or oximetry. Nasal prongs can deliver a tracheal FIo_2 of approximately 0.50, and non-rebreather masks can deliver 50 to 100% O_2; in both cases, this depends on the inspiratory pattern and flow rate. If only hypoventilation or $\dot{V}/\dot{Q}$ mismatch is present, only a small increment in FIo_2 (e.g., an FIo_2 of 0.24 or 0.28 delivered by a Venturi principle face mask or by mechanical ventilation; or 1 to 2 L/minute O_2 delivered by nasal prongs) is likely to be required. By comparison, if marked shunting or many lung units with low but not zero $\dot{V}/\dot{Q}$ are the cause of hypoxemia, a considerably higher FIo_2 (e.g., >0.7) may be required, and even this high FIo_2 may not reverse the hypoxemia. A common practice when a significant shunt is suspected is to give an FIo_2 of 1.0, then adjust the FIo_2 downward as guided by the resulting Pao_2 or Sao_2.

The O_2 concentration that is toxic to the lungs in critically ill patients is not known, but prior injury may provide tolerance to O_2 toxicity, whereas other conditioning agents, such as bleomycin, may enhance oxidative injury. An FIo_2 of 0.7 or higher is generally considered injurious to the normal human lung. Because it is unknown what lower concentration is safe, however, patients should be given the lowest FIo_2 that provides an adequate Sao_2 (≥90%). If an FIo_2 equal to or greater than 0.5 to 0.7 is required for adequate oxygenation, other measures, especially PEEP or CPAP, should be considered. Even a lower FIo_2 of about 0.5 may be associated with impaired ciliary action in the airways and impaired bacterial killing by alveolar macrophages, but the clinical importance of these effects is not known.

A low concentration of supplemental O_2 can be administered by nasal prongs or nasal cannula, which most patients find comfortable and allows them to cough, speak, eat, and drink while receiving O_2. When the nasal passages are open, the PIo_2 does not depend too much on whether the patient breathes through the nose or the mouth because O_2 is entrained from the posterior nasal pharynx during a breath taken through the mouth. The level of O_2 can be adjusted by the flow rate to the nasal prongs. In patients with COPD, flows as low as 0.5 to 2 L/minute are usually adequate unless an intrapulmonary shunt is contributing to the hypoxemia, as usually occurs in acute pneumonia. At flows greater than approximately 6 L/minute, only a small further augmentation in the PIo_2 can be achieved. Because gas flow through the nose has a drying and irritating effect, a face mask should be considered at high flow rates. O_2 face masks using the Venturi principle allow the regulation of FIo_2 and can be particularly useful when COPD is suspected, and it is important to avoid the CO_2 retention that can be associated with the unregulated administration of O_2. A higher FIo_2 of 0.5 to nearly 1.0 can be administered through a non-rebreathing face mask with an O_2 reservoir. If an FIo_2 equal to or greater than 0.70 is required for more than several hours, particularly in an unstable patient, endotracheal intubation should be considered so O_2 can be administered by a closed system with reliable maintenance of the patient's Sao_2. Indications for placement of an artificial airway in a patient with acute respiratory failure are to protect the airway against aspiration of gastric contents, to deliver an increased FIo_2, to facilitate prolonged mechanical ventilation, and possibly to aid in the control of respiratory secretions (Chapter 105).

Ventilatory maneuvers that may increase arterial oxygenation include mechanical ventilation itself and the administration of PEEP or CPAP, all of which allow ventilation of areas of the lung that were previously poorly ventilated or unventilated. Although large tidal volumes with mechanical ventilation may open areas of atelectasis and may improve oxygenation initially, these higher tidal volumes can cause lung injury, particularly if the lung is already injured (Chapter 105).[2]

CPAP refers to the maintenance of positive pressure during the respiratory cycle while breathing spontaneously. PEEP refers to the maintenance of positive pressure throughout the expiratory cycle when it is applied together with mechanical ventilation (Chapter 105). CPAP and PEEP can result in recruitment of microatelectatic regions of the lung that are perfused but were not previously ventilated, thus contributing substantially to hypoxemia. CPAP and PEEP have the theoretical advantage of keeping some of these regions open during exhalation, thus preventing cyclic closure and reopening of lung units, which may result in alveolar wall stress and injury.

Supportive Measures

Every patient with acute respiratory failure is at risk for deep venous thrombosis, pulmonary thromboembolism, and gastric stress ulceration. Prophylactic anticoagulation is recommended in patients who are not at high risk for bleeding complications; sequential leg compression therapy may be preferred for high-risk patients (Chapter 81). Nutrition is important to maintain strength needed for weaning. In patients with ARDS, limited enteral feeding for up to 6 days is as good as full enteral feeding in terms of ventilator-free days, 60-day mortality, and infectious complications, and limited feedings induce less gastrointestinal intolerance.[1]

The best means of preventing gastric stress ulceration is not known, but current evidence indicates that the use of an H_2-receptor blocker is superior to the gastric administration of sucralfate on the basis of a large randomized, controlled trial that found a higher incidence of significant bleeding in patients receiving sucralfate than in those receiving ranitidine. Evidence also indicates that proton pump inhibitors may be useful in the acute care setting (Chapter 217).

Current evidence supports maintaining the head of the bed at a 45-degree angle to reduce aspiration in critically ill patients. Attempts should be made to ensure a normal day-night sleep pattern, including minimizing activity and reducing direct lighting at night. The patient should change position frequently, including sitting in a chair and walking short distances if possible, even while receiving mechanical ventilatory support. Mobilization can enhance the removal of secretions, help maintain musculoskeletal function, reduce the risk of deep venous thrombosis, and provide psychological benefits.

SPECIFIC ACUTE RESPIRATORY FAILURE SYNDROMES

Chronic Obstructive Pulmonary Disease

EPIDEMIOLOGY AND PATHOBIOLOGY

The epidemiology and pathobiology of COPD are discussed in Chapter 88.

CLINICAL MANIFESTATIONS

When patients with COPD develop acute respiratory failure, they commonly have a history of increasing dyspnea and sputum production. Acute respiratory failure may be manifested in more cryptic ways, however, such as changes in mental status, arrhythmias, or other cardiovascular abnormalities. Acute respiratory failure must be considered whenever patients with COPD have significant nonspecific clinical changes.

DIAGNOSIS

The diagnosis can be confirmed or excluded by arterial blood gas analysis. The pH is helpful in assessing whether the hypoventilation is partly or exclusively acute. The pH declines by approximately 0.08 for each rise of 10 mm Hg in the $Paco_2$ in acute respiratory acidosis without renal compensation. By comparison, in chronic respiratory acidosis with normal renal compensation, the pH drops only about 0.03 for each rise of 10 mm Hg in the $Paco_2$.

TREATMENT

Rx

General Care

As soon as acute respiratory failure is confirmed in a patient with COPD, attention must focus on detecting potential precipitating events (Table 104-4), including decreased ventilatory drive, commonly because of oversedation; decreased muscle strength or function, often related to electrolyte abnormalities, including hypophosphatemia and hypomagnesemia; decreased chest wall elasticity, possibly related to rib fracture, pleural effusion, ileus, or ascites; atelectasis, pneumonia, or pulmonary edema; increased airway resistance, caused by bronchospasm or increased secretions; or increased metabolic O_2 requirements, such as may occur with systemic infection. Many of these abnormalities can impair the cough mechanism, diminish the clearance of airway secretions, and precipitate acute respiratory failure.

Infection

The most common specific precipitating event is airway infection, especially acute bronchitis. The role played by viral agents, *Mycoplasma pneumoniae*, chronic contaminants of the lower airway such as *Haemophilus influenzae* and *Streptococcus pneumoniae*, and other acute pathogens is difficult to determine on a clinical or even microbiologic basis. Acute exacerbations of COPD commonly result from new infections rather than from reemergence of an infection by preexisting colonization. Antibiotics modestly shorten the duration of the exacerbation, with no significant increase in toxicity, compared with placebo; the impact of antibiotics on the subsequent emergence of resistant organisms is not known. It is standard practice to use antibiotics to treat a patient with COPD who has an exacerbation severe enough to cause acute respiratory failure and who has evidence consistent with acute tracheobronchitis (Chapters 88 and 96). Pneumonia may account for 20% of cases of acute respiratory failure in patients with COPD. Compared with the physiologically normal population, patients with COPD who have community-acquired pneumonia are more likely to have gram-negative enteric bacteria or *Legionella* infections and are more likely to have antibiotic-resistant organisms.

Other Precipitating Causes

Other common precipitating causes of acute respiratory failure include heart failure and worsening of the underlying COPD, often related to noncompliance with medications. Less common and often difficult to diagnose in this setting is pulmonary thromboembolism.

Site of Care

Many patients with COPD and acute respiratory failure can be managed on a general medical hospital unit rather than in an intensive care unit if the precipitating cause of acute respiratory failure has been diagnosed and is potentially responsive to appropriate therapy, provided blood gas abnormalities respond to O_2 therapy, the patient can cooperate with the treatment, and appropriate nursing and respiratory care is available (Chapter 88). An unstable patient who requires closer observation and monitoring should be admitted to an intensive care unit.

Mechanical Therapy

The decision to institute mechanical ventilation in patients with COPD and acute respiratory failure must be made on clinical grounds and is not dictated by any particular arterial blood gas values. In general, if the patient is alert and is able to cooperate with treatment, mechanical ventilation often is not necessary. If ventilatory support is required (Chapter 105), the decision is whether to use noninvasive positive-pressure ventilation therapy (without endotracheal intubation) or endotracheal intubation with positive-pressure ventilation. A number of studies have demonstrated that noninvasive positive-pressure ventilation is preferred for patients with COPD and can decrease mortality if it is applied in appropriate patients with no factors that are likely to lead to complications.

TABLE 104-4 KEY PRINCIPLES IN THE MANAGEMENT OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE PATIENTS WITH ACUTE RESPIRATORY FAILURE

1. Monitor and treat life-threatening hypoxemia (these measures should be performed virtually simultaneously).
   a. Assess the patient clinically, and measure oxygenation by arterial blood gases and/or oximetry.
      (1) If the patient is hypoxemic, initiate supplemental oxygen therapy with nasal prongs (low flows [0.5-2. L/min] are usually sufficient) or by Venturi face mask (24 or 28% oxygen delivered).
      (2) If the patient needs ventilatory support, consider noninvasive ventilation.
      (3) Determine whether the patient needs to be intubated; this is almost always a clinical decision. Immediate action is required if the patient is comatose or severely obtunded.
   b. A reasonable goal in most patients is Pao_2 of 55-60 mm Hg or Sao_2 of 88-90%.
   c. After changes in Fio_2, check blood gases and check regularly for signs of carbon dioxide retention.
2. Start to correct life-threatening acidosis.
   a. The most effective approach is to correct the underlying cause of acute respiratory failure (e.g., bronchospasm, infection, heart failure).
   b. Consider ventilatory support, based largely on clinical considerations.
   c. With severe acidosis, the use of bicarbonate can be considered, but it is often ineffective, and there is little evidence of a clinical benefit.
3. If ventilatory support is required, consider noninvasive mechanical ventilation.
   a. The patient must have intact upper airway reflexes and be alert, cooperative, and hemodynamically stable.
   b. Careful monitoring is required; if the patient does not tolerate the mask, becomes hemodynamically unstable, or has a deteriorating mental status, consider intubation.
4. Treat airway obstruction and the underlying disease process that triggered the episode of acute respiratory failure.
   a. Treat airway obstruction with pharmacologic agents: systemic corticosteroids and bronchodilators (ipratropium and/or β-adrenergic agents).
   b. Improve secretion clearance: encourage the patient to cough, administer chest physical therapy if cough is impaired and a trial appears effective.
   c. Treat the underlying disease process (e.g., antibiotics, diuretics).
5. Prevent complications of the disease process and minimize iatrogenic complications.
   a. Pulmonary thromboembolism prophylaxis: use subcutaneous heparin if no contraindications exist.
   b. Gastrointestinal complications: administer prophylaxis for gastrointestinal bleeding.
   c. Hemodynamics: if the patient is ventilated, monitor and minimize auto-PEEP.
      (1) Treat the underlying obstruction.
      (2) Minimize minute ventilation; use controlled hypoventilation.
      (3) Use small tidal volumes; increase the inspiratory flow rate to decrease the inspiratory time and lengthen the expiratory time.
   d. Cardiac arrhythmias: maintain oxygenation and normalize electrolytes.

Fio_2 = fraction of inspired oxygen; Pao_2 = partial pressure of oxygen in arterial blood; PEEP = positive end-expiratory pressure; Sao_2 = oxygen saturation.

PROGNOSIS

Acute respiratory failure in patients with severe COPD is associated with an in-hospital mortality of 6 to 20%. The severity of the underlying disease and the severity of the acute precipitating illness are important determinants of hospital survival. Hospital mortality is higher if the respiratory failure is associated with a pH lower than 7.25 and if the patient requires invasive mechanical ventilation. However, the pH, the $Paco_2$, and other clinical characteristics are not reliable in predicting a particular patient's chances of survival.

Acute Lung Injury/Acute Respiratory Distress Syndrome

DEFINITION

ARDS is the abrupt onset of diffuse lung injury characterized by severe hypoxemia (shunting) and generalized pulmonary infiltrates on the chest

radiograph in the absence of left-sided cardiac failure.[3,4] The term *acute lung injury* has been used to include "traditional" ARDS as well as less severe forms of lung injury. Acute lung injury and its more severe manifestation, ARDS, are diagnosed by bilateral pulmonary infiltrates compatible with pulmonary edema in the absence of clinical heart failure (usually determined by the lack of elevated left atrial pressures). A recent Berlin consensus conference recommended that the term *acute lung injury* should no longer be used, with ARDS diagnosed on the basis of a Pao_2/Fio_2 of less than 300 mm Hg. With this new definition, the severity of ARDS is defined on the basis of the Pao_2 divided by the Fio_2 (Pao_2/Fio_2, also called the P/F ratio) as mild ARDS ($200 < P/F \leq 300$ mm Hg), moderate ARDS ($100 < P/F \leq 200$ mm Hg), or severe ARDS ($P/F \leq 100$ mm Hg).[5]

EPIDEMIOLOGY

ARDS is a clinical syndrome triggered by some other cause (Table 104-5), with an annual incidence of about 80 cases per 100,000 adult population. This underlying precipitating factor may affect and injure the lungs directly, such as in diffuse pneumonia or aspiration of gastric contents, or it may affect the lungs indirectly, as in severe sepsis from a nonpulmonary or a pulmonary source (Chapter 108) or severe nonthoracic trauma associated with shock (Chapter 111).[6] Severe sepsis is the most common precipitating cause of ARDS worldwide. The organisms vary widely, ranging from gram-negative and gram-positive bacteria and viruses (e.g., H1N1 influenza in 2009) to leptospiral infections or malaria. It can be difficult to determine whether pneumonia is diffuse, with endobronchial spread involving most of the lungs, or whether localized pneumonia has precipitated a sepsis syndrome, with secondary injury to other parts of the lung.

PATHOBIOLOGY

Pathology

Despite the variety of underlying disease processes leading to ARDS, the response to these insults in the lung is monotonously characteristic, with similar clinical findings, physiologic changes, and morphologic abnormalities. The pathologic abnormalities in ARDS are nonspecific and are described as *diffuse alveolar damage* by pathologists. The initial process is inflammatory, with neutrophils usually predominating in the alveolar fluid. Hyaline membranes are present in some but not all patients,[7] similar to those seen in premature infants with infant respiratory distress syndrome, presumably related to the presence of large-molecular-weight proteins that have leaked into the alveolar space. Alveolar flooding leads to impairment of surfactant, which is abnormal in quantity and quality. The result is microatelectasis, which may be associated with impaired immune function. Cytokines and other inflammatory mediators are usually markedly elevated, although with different patterns over time in the bronchoalveolar lavage fluid and the systemic blood. The resolution of ARDS depends in part on restoration of a functional alveolar epithelial barrier, capable of removing alveolar edema fluid by sodium-dependent vectorial fluid transport.[8] Lung repair is also disturbed; early evidence of profibrotic processes includes the appearance of breakdown products of procollagen in the bronchoalveolar lavage fluid, followed by fibrosis in some patients. Lung function improves over time in survivors of ARDS; however, the fibrosis is often reversible.

TABLE 104-5 DISORDERS ASSOCIATED WITH THE ACUTE RESPIRATORY DISTRESS SYNDROME

COMMON
Sepsis (gram-positive or gram-negative bacterial, viral, fungal, or parasitic infection)
Diffuse pneumonia (bacterial, viral, or fungal)
Aspiration of gastric contents
Trauma (usually severe)
LESS COMMON
Near-drowning (fresh or salt water)
Drug overdose
Acetylsalicylic acid
Heroin and other narcotic drugs
Massive blood transfusion (likely a marker of severe trauma, but also seen with severe gastrointestinal bleeding, especially in patients with severe liver disease)
Leukoagglutination reactions
Inhalation of smoke or corrosive gases (usually requires high concentrations)
Pancreatitis
Fat embolism
UNCOMMON
Miliary tuberculosis
Paraquat poisoning
Central nervous system injury or anoxia (neurogenic pulmonary edema)
Cardiopulmonary bypass

Pathophysiology

The physiologic abnormalities are dominated by severe hypoxemia with shunting, decreased lung compliance, decreased functional residual capacity, increased pulmonary dead space, and increased work of breathing. Initially, the $Paco_2$ is low or normal, usually associated with increased minute ventilation. The initial abnormalities in oxygenation are thought to be related to alveolar flooding and collapse. As the disease progresses, especially in patients who require ventilatory support, fibroproliferation may develop; the lungs (including alveoli, blood vessels, and small airways) remodel and scar, with a loss of microvasculature. In some patients, these changes may lead to pulmonary hypertension and increased pulmonary dead space; marked elevations in minute ventilation are required to achieve a normal $Paco_2$, even as oxygenation abnormalities are improving.

CLINICAL MANIFESTATIONS

In most cases of ARDS, the onset either coincides with or occurs within 72 hours of the onset of the underlying disease process; the mean time from onset of the underlying cause to onset of acute lung injury is 12 to 24 hours. The presenting picture is dominated by respiratory distress and the accompanying laboratory findings of severe hypoxemia and generalized infiltrates or opacities on the chest radiograph. Alternatively, it may be dominated by manifestations of the underlying disease process, such as severe sepsis with hypotension and other manifestations of systemic infection.

DIAGNOSIS

The key to diagnosis is to distinguish ARDS from cardiogenic pulmonary edema (Table 104-6). No specific biochemical test exists to define ARDS. Certain blood or bronchoalveolar lavage (Chapter 85) abnormalities are frequent but are not sufficiently specific to be useful clinically.

TREATMENT

Rx

Treatment of ARDS consists predominantly of respiratory support and treatment of the underlying disease (Fig. 104-4). Sepsis, which is a common predisposing condition for the development of ARDS, must be treated aggressively (Chapter 108).

Current recommendations for lung-protective mechanical ventilation by endotracheal intubation (Table 104-7) emphasize lower tidal volumes based on the patient's predicted body weight (Chapter 105). This approach also includes achieving a plateau airway pressure less than 30 cm H_2O. PEEP is a mainstay in the ventilatory strategy for ARDS; although the method for determining the optimal level of PEEP has not been established, higher PEEP levels may have some benefit for patients with moderate to severe ARDS. PEEP may allow a lower Fio_2 to provide adequate oxygenation, thereby reducing the risk of O_2 toxicity. It also may prevent the cyclic collapse and reopening of lung units, a process that is thought to be a major cause of ventilator-induced lung injury, even when adequate oxygenation can be obtained at relatively low levels of Fio_2.[9] On the basis of one clinical trial, the early use of cisatracurium besylate (15 mg rapid infusion followed by 37.5 mg/hour for 48 hours), a neuromuscular blocker, can reduce ARDS mortality rates by about 25% in patients with moderately severe ARDS with a P/F below 150 mm Hg. In patients with severe ARDS who do not respond to standard therapy but otherwise have a reasonable life expectancy and do not have multiorgan failure, extracorporeal membrane oxygenation is an acceptable albeit not fully proven rescue therapy.[10] Data also indicate that prone positioning reduces mortality in patients with moderate to severe ARDS (initial P/F < 150 mm Hg).

PROGNOSIS

Case-fatality rates are 30 to 50% and are highly dependent on disease severity and the underlying predisposing condition. Based on the degree of hypoxemia (mild, 200 mm Hg $< Pao_2/Fio_2 <$ 300 mm Hg; moderate, 100 mm Hg $< Pao_2/Fio_2 <$ 200 mm Hg; and severe, $Pao_2/Fio_2 <$ 100 mm Hg), inpatient mortality rates are about 27%, 32%, and 45%, respectively, not including patients with severe underlying conditions, such as end-stage cancer.[5] Memory, verbal fluency, and executive function are impaired in about 13%, 16%, and 49% of long-term survivors after ARDS.[11,12] Lower Pao_2 during hospitalization is associated with cognitive and psychiatric impairment.

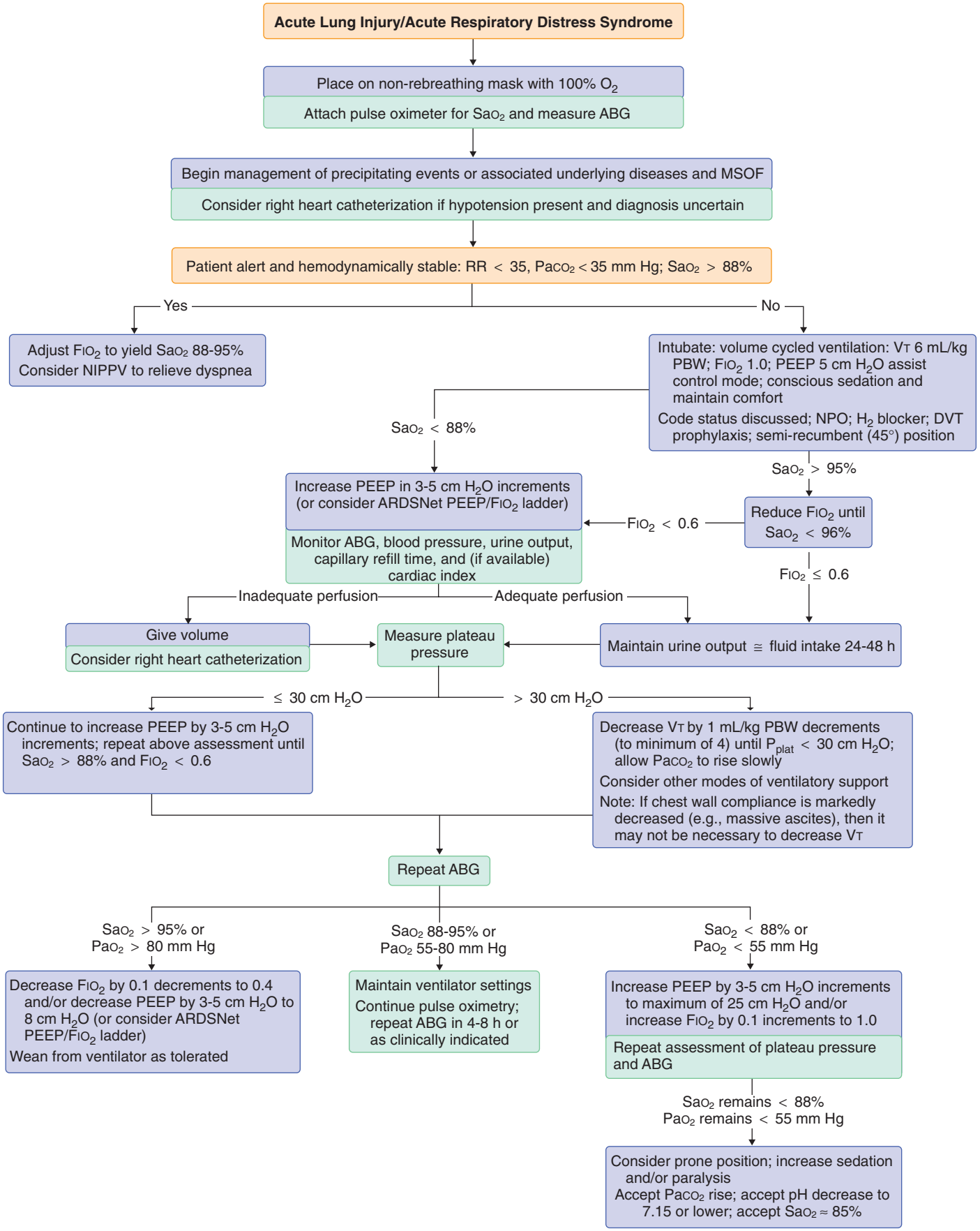

FIGURE 104-4. Algorithm for the initial management of acute respiratory distress syndrome. ABG = arterial blood gas analysis; CO_2 = carbon dioxide; DVT = deep venous thrombosis; FIO_2 = inspired oxygen concentration; MSOF = multisystem organ failure; NIPPV = noninvasive intermittent positive-pressure ventilation; O_2 = oxygen; $Paco_2$ = arterial partial pressure of carbon dioxide; Pao_2 = arterial partial pressure of oxygen; PBW = predicted body weight; PEEP = positive end-expiratory pressure; P_{plat} = plateau pressure; RR = respiratory rate; Sao_2 = arterial oxygen saturation; VT = tidal volume.

TABLE 104-6 FEATURES ASSOCIATED WITH NONCARDIOGENIC AND CARDIOGENIC PULMONARY EDEMA*

NONCARDIOGENIC EDEMA (ARDS)	CARDIOGENIC EDEMA/VOLUME OVERLOAD
PRIOR HISTORY	
No history of heart disease	Prior history of heart disease
Appropriate fluid balance (difficult to assess after resuscitation from shock or trauma)	Hypertension, chest pain, new-onset palpitations; positive fluid balance
PHYSICAL EXAMINATION	
Flat neck veins	Elevated neck veins
Hyperdynamic pulses	Left ventricular enlargement, lift, heave, dyskinesis
Physiologic gallop	S_3 and S_4; murmurs
Absence of edema	Edema: flank, presacral, legs
ELECTROCARDIOGRAM	
Sinus tachycardia, nonspecific ST-T wave changes	Evidence of prior or ongoing ischemia, supraventricular tachycardia, left ventricular hypertrophy
CHEST RADIOGRAPH	
Normal heart size	Cardiomegaly
Peripheral distribution of infiltrates	Central or basilar infiltrates; peribronchial and vascular congestion
Air bronchograms common (80%)	Septal lines (Kerley lines), air bronchograms (25%), pleural effusion
HEMODYNAMIC MEASUREMENTS	
Pulmonary artery wedge pressure <15 mm Hg, cardiac index >3.5 L/min/m^2	Pulmonary capillary wedge pressure >18 mm Hg, cardiac index <3.5 L/min/m^2 with ischemia, may be >3.5 L/min/m^2 with volume overload

*These features are neither highly sensitive nor specific. Although the findings are more commonly associated with the type of pulmonary edema as listed, they do not have high positive or negative predictive value.
ARDS = acute respiratory distress syndrome.

TABLE 104-7 ARDS NETWORK VENTILATORY MANAGEMENT PROTOCOL FOR TIDAL VOLUME AND PLATEAU AIRWAY PRESSURE

Calculate PBW:
- Male PBW: 50 + 2.3 (height in inches – 60) or 50 + 0.91 (height in centimeters – 152.4)
- Female PBW: 45.5 + 2.3 (height in inches – 60) or 45.5 + 0.91 (height in centimeters – 152.4)

Select assist control mode
Set initial V_T at 8 mL/kg PBW
Reduce V_T by 1 mL/kg at intervals < 2 hr until V_T = 6 mL/kg PBW
Set initial RR to approximate baseline minute ventilation (maximum RR = 35/min)
Set inspiratory flow rate higher than patient's demand (usually > 80 L/min)
Adjust V_T and RR further to achieve P_{plat} and pH goals
- If P_{plat} > 30 cm H_2O: decrease V_T by 1 mL/kg PBW (minimum = 4 mL/kg PBW)
- If pH ≤ 7.30, increase RR (maximum = 35)
- If pH < 7.15, increase RR to 35; consider sodium bicarbonate administration or increase V_T

ARDS = acute respiratory distress syndrome; PBW = predicted body weight; P_{plat} = plateau pressure (airway pressure at the end of delivery of a tidal volume breath during a condition of no airflow); RR = respiratory rate; V_T = tidal volume.
See the ARDSNet website (http://www.ardsnet.org) for further details about the protocol, including the approach for setting positive end-expiratory pressure and fraction of inspired oxygen.

Acute Respiratory Failure without Lung Disease

Acute respiratory failure without pulmonary abnormalities (see Table 104-2) develops in patients with depressed ventilatory drive secondary to central nervous system dysfunction and in patients with severe neuromuscular disease. The prototypical patient with suppressed ventilatory drive has taken an overdose of a sedative or tranquilizing medication (Chapter 110). The prototypical patient with neuromuscular disease has Guillain-Barré syndrome (Chapter 420). The treatment for both types of patients is supportive. In the case of a patient with a sedative overdose, the threshold for intubation with mechanical ventilatory support should be low because this temporary condition is quickly reversible when the responsible drug is eliminated. Such a patient may require intubation for airway protection against aspiration of gastric contents.

Patients with Guillain-Barré syndrome or other forms of progressive neuromuscular disease should be monitored with serial measurements of vital capacity. In general, when the vital capacity decreases to less than 10 to 15 mL/kg body weight, intubation and mechanical ventilatory support should be considered without regard to the patient's $PaCO_2$.

Grade A References

A1. Wheeler AP, Bernard GR, Thompson BT, et al. Pulmonary-artery versus central venous catheter to guide treatment of acute lung injury. *N Engl J Med*. 2006;354:2213-2224.
A2. Wiedemann HP, Wheeler AP, Bernard GR, et al. Comparison of two fluid-management strategies in acute lung injury. *N Engl J Med*. 2006;354:2564-2575.
A3. Rice TW, Wheeler AP, Thompson BT, et al. Initial trophic vs full enteral feeding in patients with acute lung injury: the EDEN randomized trial. *JAMA*. 2012;307:795-803.
A4. Williams JW, Cox CE, Hargett CW, et al. Noninvasive Positive-Pressure Ventilation (NPPV) for Acute Respiratory Failure. AHRQ Comparative Effectiveness Reviews, No. 68. Report No. 12-EHC089-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2012.
A5. Putensen C, Theuerkauf N, Zinserling J, et al. Meta-analysis: ventilation strategies and outcomes of the acute respiratory distress syndrome and acute lung injury. *Ann Intern Med*. 2009;151:566-576.
A6. Briel M, Meade M, Mercat A, et al. Higher vs lower positive end-expiratory pressure in patients with acute lung injury and acute respiratory distress syndrome: systematic review and meta-analysis. *JAMA*. 2010;303:865-873.
A7. Meade MO, Cook DJ, Guyatt GH, et al. Ventilation strategy using low tidal volumes, recruitment maneuvers, and high positive end-expiratory pressure for acute lung injury and acute respiratory distress syndrome: a randomized controlled trial. *JAMA*. 2008;299:637-645.
A8. Mercat A, Richard JC, Vielle B, et al. Positive end-expiratory pressure setting in adults with acute lung injury and acute respiratory distress syndrome: a randomized controlled trial. *JAMA*. 2008;299:646-655.
A9. Papazian L, Forel JM, Gacouin A, et al. Neuromuscular blockers in early acute respiratory distress syndrome. *N Engl J Med*. 2010;363:1107-1116.
A10. Peek GJ, Mugford M, Tiruvoipati R, et al. Efficacy and economic assessment of conventional ventilatory support versus extracorporeal membrane oxygenation for severe adult respiratory failure (CESAR): a multicentre randomised controlled trial. *Lancet*. 2009;374:1351-1363.
A11. Hu SL, He HL, Pan C, et al. The effect of prone positioning on mortality in patients with acute respiratory distress syndrome: a meta-analysis of randomized controlled trials. *Crit Care*. 2014;18:R109.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Grocott MP, Martin DS, Levett DZ, et al. Arterial blood gases and oxygen content in climbers on Mount Everest. *N Engl J Med*. 2009;360:140-149.
2. Wilson JG, Matthay MA. Mechanical ventilation in acute hypoxemic respiratory failure: a review of new strategies for the practicing hospitalist. *J Hosp Med*. 2014;9:469-475.
3. Del Sorbo L, Slutsky AS. Acute respiratory distress syndrome and multiple organ failure. *Curr Opin Crit Care*. 2011;17:1-6.
4. Matthay MA, Ware L, Zimmerman GA. The acute respiratory distress syndrome. *J Clin Invest*. 2012;122:2731-2740.
5. Force ADT, Ranieri VM, Rubenfeld GD, et al. Acute respiratory distress syndrome: the Berlin Definition. *JAMA*. 2012;307:2526-2533.
6. Villar J, Sulemanji D, Kacmarek RM. The acute respiratory distress syndrome: incidence and mortality, has it changed? *Curr Opin Crit Care*. 2014;20:3-9.
7. Thille AW, Esteban A, Fernandez-Segoviano P, et al. Comparison of the Berlin definition for acute respiratory distress syndrome with autopsy. *Am J Respir Crit Care Med*. 2013;187:761-767.
8. Matthay MA. Resolution of pulmonary edema. Thirty years of progress. *Am J Respir Crit Care Med*. 2014;189:1301-1308.
9. Slutsky AS, Ranieri VM. Ventilator-induced lung injury. *N Engl J Med*. 2013;369:2126-2136.
10. Ventetuolo CE, Muratore CS. Extracorporeal life support in critically ill adults. *Am J Respir Crit Care Med*. 2014;190:497-508.
11. Herridge MS, Tansey CM, Matte A, et al. Functional disability 5 years after acute respiratory distress syndrome. *N Engl J Med*. 2011;364:1293-1304.
12. Mikkelsen ME, Christie JD, Lanken PN, et al. The adult respiratory distress syndrome cognitive outcomes study: long-term neuropsychological function in survivors of acute lung injury. *Am J Respir Crit Care Med*. 2012;185:1307-1315.

REVIEW QUESTIONS

1. Which of the following mechanisms can lead to a reduction in the arterial oxygen tension (Pao_2)?

A. Ventilation-perfusion mismatch
B. Intrapulmonary right-to-left shunt
C. Decreased inspired partial pressure of oxygen
D. Alveolar hypoventilation
E. All of the above

Answer: E It is important for clinicians to understand the physiologic mechanisms of arterial hypoxemia. Once the physiologic basis of hypoxemia is known, the differential diagnosis as to specific cause (i.e., the type of disease at hand leading to this physiologic abnormality) can be made, and a specific treatment often can be initiated.

2. If a patient has the clinical presentation of bilateral pneumonia with no evidence of cardiac disease and has a chest radiograph that shows bilateral pulmonary infiltrates, which of the following Pao_2/Fio_2 ratios (where Pao_2 is in mm Hg) is consistent with the diagnosis of acute respiratory distress syndrome (ARDS)?

A. $Pao_2 = 80$; $Fio_2 = 0.25$
B. $Pao_2 = 200$; $Fio_2 = 0.3$
C. $Pao_2 = 200$; $Fio_2 = 0.8$
D. $Pao_2 = 190$; $Fio_2 = 0.3$
E. $Pao_2 = 490$; $Fio_2 = 0.5$

Answer: C It is the only one with a Pao_2/Fio_2 ratio of less than 300 mm Hg. It is important to be able to calculate the Pao_2/Fio_2 ratio to make the diagnosis of ARDS in a patient who has bilateral pulmonary infiltrates consistent with pulmonary edema but no evidence of cardiogenic pulmonary edema as the primary cause.

3. Which of the following treatments has been shown to reduce mortality in patients with ARDS on the basis of multicenter randomized trials.

A. High-frequency ventilation
B. Lung-protective ventilation with a tidal volume of 6 mL/kg predicted body weight and a plateau airway pressure of less than 30 cm H_2O
C. Intermittent mandatory ventilation with weaning to pressure support ventilation
D. Pressure-control ventilation
E. Liquid ventilation with perfluorocarbons

Answer: B on the basis of a National Heart, Lung, and Blood Institute–sponsored ARDS Network clinical trial in 2000. All intensivists need to know that lung-protective ventilation with this strategy markedly reduces mortality in patients with ARDS.

4. Why do most patients with an exacerbation of chronic obstructive pulmonary disease (COPD) require relatively low-flow nasal oxygen to achieve adequate oxygenation?

A. Because intrapulmonary right-to-left shunt is an important cause of the hypoxemia
B. Because there are many lung units with low ventilation-perfusion ratios, and these units are the primary cause of arterial hypoxemia in these patients
C. Because diffusion impairment is an important cause of hypoxemia and COPD
D. Because the inspired oxygen tension is low in these patients
E. Because such patients have decreased cardiac output and therefore low perfusion

Answer: B because ventilation-perfusion abnormalities predominate. In most patients with an exacerbation of COPD, most gas exchange takes place in units with low ventilation to perfusion values. In most patients, the hypoxemia can be corrected, at least in part, with low-flow oxygen. Clinicians should understand the physiologic basis of treating an exacerbation of COPD, which normally requires low-flow oxygen unless another process, such as pneumonia, pulmonary edema, or atelectasis, could be causing an intrapulmonary shunt.

5. Which of the following clinical conditions have been associated with the development of ARDS?

A. Nonpulmonary sepsis
B. Pneumonia
C. Drug overdose
D. Massive blood transfusions
E. All of the above

Answer: E Clinicians need to know clinical disorders that may be associated with ARDS.

105

MECHANICAL VENTILATION

ARTHUR S. SLUTSKY

Mechanical ventilation is a life-sustaining therapy in which a ventilator provides partial or full support for patients with respiratory failure (Chapter 104). In setting the ventilator, the clinician can use a variety of modes of ventilation and can also alter the inspired oxygen tension, the pressure at the airway opening at the end of a breath, and other facets of the volume or pressure time pattern imposed on the patient.

The main goals of ventilatory support are to maintain adequate gas exchange, to rest the respiratory muscles, and to decrease the oxygen cost of breathing. Modern ventilation strategies focus on minimizing its iatrogenic consequences, such as iatrogenic hyperinflation (from endogenously derived positive pressure at the end of a breath, i.e., auto-PEEP) and ventilator-induced lung injury. In some patients, the physician should be willing to accept arterial blood gases that are not in the normal range to avoid these complications by using lower levels of minute ventilation or relatively smaller tidal volumes.

TYPES OF MECHANICAL VENTILATORS

Negative-Pressure Ventilators

Delivery of gas to the lungs requires a hydrostatic pressure gradient between the airway opening and the alveoli. During spontaneous breathing, this pressure gradient is generated by developing negative pleural pressure due to respiratory muscle contraction. Some ventilators operate by generating negative pressure around the chest wall (e.g., cuirass) or around the entire body below the neck (e.g., iron lung). The cuirass has the major advantage of minimizing detrimental hemodynamic consequences, but it is difficult to apply because the device must have an adequate seal to the body so that the negative pressure is not dissipated to the room—a task that is not always easy to accomplish in a way that is comfortable for the patient. The iron lung makes nursing care difficult because it encircles the patient's entire body. Although iron lungs were widely used during the polio epidemic of the mid-1950s, they are rarely used today.

Positive-Pressure Ventilators

The most widely used approach to mechanical ventilation is to deliver gas to the lung with positive-pressure ventilation (PPV) applied through an endotracheal tube, a tracheostomy, or a tight-fitting mask. The approach with a mask is considered noninvasive ventilation (NIV) and is considered separately.

The most basic mode of PPV is controlled ventilation, in which a preset tidal volume at a predetermined rate is delivered, regardless of the patient's requirements or efforts. This form of ventilation is usually used in patients who cannot initiate spontaneous breaths (e.g., heavily sedated or paralyzed patients) or in those who need full ventilatory support because of extremely severe pulmonary or cardiovascular disease (e.g., severe shock). This ventilator mode may be beneficial when it is used for relatively short periods (~48 hours) in patients with the acute respiratory distress syndrome (ARDS) early in their clinical course, when they may be treated with neuromuscular blocking agents (see later). However, a paralyzed patient without any ability to make breathing efforts is at risk of asphyxia in the event of an inadvertent disconnection from the ventilator. If the patient is not making any respiratory efforts, controlled ventilation can rapidly lead to respiratory muscle atrophy. For these reasons, clinicians usually try to limit the time that a patient is paralyzed and receiving controlled ventilation. Assisted ventilation is the term used when the patient's spontaneous ventilatory efforts trigger the ventilator to deliver breaths, rather than having the breaths delivered by the ventilator at a fixed rate without regard to the patient's efforts.

Mechanical ventilation can be applied by either volume-controlled or pressure-controlled modes. In volume-controlled ventilation, the desired tidal volume and respiratory rate are set by the user, and the airway pressure is the dependent variable. The airway pressure profile depends on the mechanical properties of the patient's respiratory system and on the ventilator's flow settings. In pressure-controlled ventilation, the pressure imposed at the airway opening along with the respiratory rate is set by the user, and the tidal volume becomes the dependent variable.

Positive End-Expiratory Pressure

A key characteristic that can be combined with most ventilatory modes is the level of the end-expiratory pressure. Positive end-expiratory pressure (PEEP) is used in patients with diffuse pulmonary diseases (e.g., pulmonary edema or ARDS) to recruit collapsed alveolar regions and to maintain them in a recruited state, to reopen collapsed airways, to redistribute fluid in the lung, to increase functional residual capacity, and to redistribute ventilation to dependent regions. All these changes can improve the matching of ventilation to perfusion, thereby leading to improved oxygenation and allowing the fractional inspiratory concentration of oxygen (F_{IO_2}) to be reduced. PEEP does not usually improve alveolar ventilation and, in fact, may increase dead space by overdistending alveoli, with a concomitant decrease in alveolar capillary blood flow in certain regions of the lung. PEEP can also be administered to spontaneously breathing subjects by a technique termed continuous positive airway pressure (CPAP). In patients with exacerbations of chronic obstructive pulmonary disease (COPD), PEEP and CPAP can overcome some of the mechanical consequences of auto-PEEP (see later) to minimize the work of breathing, provided the magnitude of the PEEP is low enough that it does not cause additional hyperinflation.

Volume-Controlled Ventilation

Volume-controlled ventilation (or volume-limited ventilation) refers to mechanical ventilation in which the tidal volume is preset. The major advantage is that the delivered tidal volume is maintained even if lung mechanics change, thereby ensuring a more constant partial pressure of arterial carbon dioxide (P_{ACO_2}). The potential disadvantage is that if lung mechanics deteriorate, higher pressures may be required to achieve the tidal volume goal, and regions of overinflation may result in regional lung injury. Although controlled ventilation as described earlier can be either volume limited (preset tidal volume) or pressure limited (preset peak airway pressure), clinicians usually use the term *controlled mechanical ventilation* to refer to volume-limited ventilation with a set ventilatory rate. In volume-controlled ventilation, an upper limit to applied airway pressure is commonly used for safety reasons.

The most common form of volume-controlled ventilation is one in which the patient assists the ventilator, thus triggering at least some of the breaths. The term *assisted mechanical ventilation* can refer to either volume-limited ventilation or pressure-limited ventilation when the patient triggers some or all of the breaths, but in either case the ventilator should be set to deliver breaths if apnea occurs. This mode is also referred to as assist/control (A/C).

Intermittent Mandatory Ventilation

Intermittent mandatory ventilation (IMV) refers to a mode in which the patient is allowed to breathe spontaneously through an endotracheal tube or tracheostomy but also receives some preset (and thus mandatory) volume-limited breaths from the ventilator. In current ventilators, the mandatory breaths are triggered by the patient and are synchronized (synchronized IMV); however, if the patient ceases spontaneous ventilatory efforts, breaths at the rate set on the ventilator will still be delivered. Synchronized IMV is a form of partial ventilatory support because some breaths are spontaneous, in contrast to full ventilatory support, in which all breaths are delivered by the ventilator. This mode allows the patient to do a variable amount of the respiratory work but with the security of a set minimal backup rate should spontaneous ventilatory efforts stop.

Pressure-Controlled Ventilation

Pressure-controlled ventilation is a type of ventilation in which the ventilator delivers pressure-limited breaths to the patient; delivered volume becomes a dependent variable. The initiation of each breath may be triggered by the patient (assisted breaths) or may be initiated by the ventilator (controlled breaths). In the assist mode, a backup control rate protects any patients who cease to make inspiratory efforts on their own. The delivered tidal volume depends on the preset pressure, the ventilatory rate, the inspiratory-to-expiratory ratio, and the patient's respiratory mechanics (resistance, compliance, and auto-PEEP). At a fixed preset pressure and inspiratory-to-expiratory ratio, tidal volume decreases as respiratory frequency increases. In patients with COPD, the tidal volume at low frequencies is relatively high but decreases substantially as the respiratory rate is increased; whereas in patients with stiff respiratory systems (e.g., ARDS), the tidal volume does not change much with respiratory frequency because the lung fills with gas quickly.

Pressure-Support Ventilation

Pressure-support ventilation is a pressure-limited, patient-triggered ventilatory mode. Once the patient triggers the ventilator by creating either a small negative pressure or a low inspiratory flow at the airway opening, the ventilator switches to inspiratory mode and provides the airflow needed to maintain a preset level of pressure. In contrast to pressure-controlled ventilation, inspiration terminates when the inspiratory airflow decreases to a threshold level (the specific algorithm varies from ventilator to ventilator). This mode provides flexibility for the patient with respect to tidal volume, inspiratory flow, and ratio of time allowed for inspiration compared with expiration. Tidal volume depends on patient-related factors (effort), respiratory system mechanics, and level of pressure set for support. During pressure-support ventilation, the size of each breath is determined partially by the patient's muscle effort and partially by the ventilator. This mode can compensate for the added work of breathing imposed by the resistance of the endotracheal tube. Pressure-support ventilation has been used to wean patients from ventilatory support because it provides a simple way to reduce the magnitude of mechanical support while the patient assumes a larger fraction of the ventilatory work.

High-Frequency Ventilation

High-frequency ventilation refers to modes that have the common feature of providing ventilation at frequencies that are substantially greater than those

used during normal breathing. During high-frequency ventilation, tidal volumes may be less than the dead space, so adequate gas transport takes place by various convective and diffusive mechanisms. Interest in these modes of ventilation has waned because it appears to be no better or even worse than conventional ventilation for adults with ARDS.

Proportional Assist Ventilation and Neurally Adjusted Ventilatory Assist

One of the difficulties in providing assisted ventilation is ensuring that there is adequate synchrony between the patient's respiratory drive and the delivery of the ventilator's breaths. This issue is a particular problem for patients with severe obstructive airways disease, especially if they have significant auto-PEEP. Two newer modes of ventilation, proportional assist ventilation (PAV) and neurally adjusted ventilatory assist (NAVA), have been developed and implemented on some ventilators in part to address this concern. Both these modes deliver ventilation in proportion to the instantaneous effort of the patient, but the underlying principles are different. Although both of these modes improve patient-ventilator synchrony, data are insufficient to know whether either will improve clinically important outcomes.

PAV is based on the mathematical relationships between airway pressure and airflow; these state that the pressure applied by the respiratory muscles is used to overcome the elastic losses (i.e., compliance) and the resistive losses of the respiratory system. With PAV, the pressure that is applied during inspiration varies on the basis of the patient's inspiratory effort and respiratory system mechanics. This form of ventilation is not in widespread use, in part because of its complexity and the need to estimate the patient's compliance and resistance on a regular basis. This latter issue has been addressed in new versions of the technique in which measurements of compliance and resistance are automatically measured repeatedly.

NAVA makes use of the electrical activity of the diaphragm (E_{di}) as measured by an array of electrodes attached to a nasogastric tube inserted into the esophagus. Pressure is then delivered by the ventilator in direct proportion to the (virtually) instantaneous E_{di}. Because the initiation and delivery of the breath by the ventilator are not dependent on measurement of pressures in the lung, patient-ventilator synchrony is improved in patients with auto-PEEP. Once the array of electrodes has been inserted, the mode is relatively easy to use; the only parameter to set is the proportionality factor linking the E_{di} and the pressure delivered by the ventilator.

Noninvasive Positive-Pressure Ventilation

PPV can be provided through a mask rather than through an endotracheal tube. This method, which has been termed noninvasive because the patient is not intubated, is conceptually simple but requires appropriate implementation and monitoring for its successful application.[1] Of particular importance are patient selection and appropriate training of hospital personnel. Patients must be alert, cooperative, and hemodynamically stable. Patients must also have intact upper airway reflexes to prevent aspiration of material from the upper airway into the lung, and they must not have any facial trauma that would preclude the use of a mask. Once patients are started on NIV, they should be carefully monitored, and NIV should be discontinued if the patient's clinical condition deteriorates, if the patient develops cardiovascular instability, or if it appears that the patient is likely to aspirate. NIV can also be delivered through a "helmet" that avoids some of the problems associated with the use of face masks.

NIV has potential advantages compared with invasive ventilation. It is relatively easy to apply and can be used for short intervals because it can be started and stopped easily. The major advantages are that it avoids the complications associated with intubation, it is usually more comfortable for the patient, and it reduces the need for sedation. Patients receiving NIV are able to communicate verbally with medical staff and family members, are probably able to sleep better, and are able to eat if they are sufficiently stable to remove the mask for short periods.

However, NIV has several disadvantages. Implementation of NIV takes more time from caregivers at the bedside initially, and the time course of correction of blood gases is slower than usually occurs in patients who are intubated and ventilated. Gastric distention is an unusual occurrence; medical staff should be aware of this complication and should watch for signs of abdominal distention. Data strongly support the use of NIV for patients with COPD (see later), and it is preferred in cardiogenic pulmonary edema, but whether it provides better outcomes in other forms of respiratory failure is uncertain.

COMPLICATIONS OF MECHANICAL VENTILATION

Intubation

Endotracheal intubation can be used to secure a patient's airway, to act as a conduit to deliver gas from the ventilator to the patient, to prevent aspiration, and to help with pulmonary toilet when secretions are increased. However, intubation can be associated with complications including the risk of aspiration during insertion of the endotracheal tube, difficulty in swallowing and communicating, disruption of normal host defense mechanisms, and upper airway trauma. Pressure from the cuff of the tube that provides a pneumatic seal between the tube and trachea can lead to regions of tracheal ischemia and may eventually cause tracheal stenosis.

The endotracheal tube increases airway resistance because its diameter is smaller than the airway into which it is inserted. The magnitude of the increase depends on the length, diameter, and shape of the tube as well as on the buildup of secretions and mucus that narrow the tube's diameter. Furthermore, the upper airway is normally an effective means of heating and humidifying inspiratory gases. This natural system is bypassed by an endotracheal tube; inadequately humidified inspiratory gases can reduce mucociliary clearance and can lead to inspissation of tracheal secretions.

Intubation affects a number of factors that increase the likelihood of nosocomial pneumonia (Chapters 97 and 282). Normally, cough involves an increase in airway pressure as respiratory muscles are contracted against a closed glottis. When the glottis opens, expiratory flow sharply increases, resulting in dynamic compression of major airways. The presence of an endotracheal tube limits the buildup of airway pressure and alters the dynamics of expiratory flow, thereby greatly impairing the efficacy of the patient's cough. A cuffed endotracheal tube helps prevent gross aspiration, but pharyngeal secretions that pool at the top of the cuff often seep into the lungs. Endotracheal tubes also can often become colonized with the microorganisms that cause ventilator-associated pneumonia (Chapter 97). Silver-coated tubes can reduce this risk but are considerably more expensive than conventional endotracheal tubes and are unlikely to be used routinely for initial intubation. Endotracheal tubes with a port that allows suctioning of secretions above the cuff may also reduce the incidence of ventilator-associated pneumonia, although results of studies have been mixed.

In addition, endotracheal intubation is often not well tolerated in awake patients, and there is always the danger that the tube will inadvertently be dislodged—a complication that can have tragic consequences. For these reasons and to improve oral care and feeding, a tracheostomy can be performed. However, tracheostomy is associated with its own set of complications, and performing a tracheostomy in the first week is no better than waiting until the patient has been ventilated for about 10 days, in part because clinicians are not accurate in predicting which patients will require prolonged ventilatory support. Early tracheostomy should be avoided in patients in whom uncertainty exists as to how long invasive ventilation will be needed.

Hemodynamic Compromise

The major mechanical determinants of cardiovascular hemodynamics during mechanical ventilation are intrathoracic pressure, changes in lung volume, and the patient's circulatory volume status. An increase in lung volume can cause a beneficial decrease in pulmonary vascular resistance, if lung units that had been closed are opened as a result of mechanical ventilation, or it can lead to a detrimental increase in pulmonary vascular resistance related to overdistention of the lung with concomitant compression and lengthening of alveolar vessels.

PPV can affect cardiovascular hemodynamics through its effect on pleural pressure, an effect that is directly related to changes in lung volume and not necessarily directly reflected in measurements of airway pressure; the relation between alveolar pressure and lung volume depends on respiratory system mechanics. For example, in a patient with stiff lungs (e.g., ARDS), a given increase in airway pressure will lead to much less of an increase in lung volume than in a patient with COPD, so the increase in pleural pressure will be much less in the patient with ARDS. As a result, patients with ARDS tolerate relatively high PEEP levels, whereas similarly high levels in patients with normal lungs (e.g., in a drug overdose) or in patients with COPD would markedly reduce cardiac output. At very high lung volumes, a direct effect of the pressure of the lung on the heart can increase pericardial pressure and can thereby decrease cardiac filling.

Auto-PEEP and Dynamic Hyperinflation

A key factor that affects cardiovascular hemodynamics and other physiologic variables during mechanical ventilation is the development of auto-PEEP, which is defined as the difference between alveolar pressure and airway pressure at end expiration. Auto-PEEP is associated with dynamic hyperinflation, which is an increase in the end-expiratory lung volume above the value that would be obtained if there was complete exhalation to the static functional residual capacity. This phenomenon occurs whenever there is insufficient time for a complete exhalation to occur; the respiratory system is thus prevented from reaching its static end-expiratory volume. The major determinants of auto-PEEP and hence dynamic hyperinflation are increased expiratory airway resistance, high minute ventilation, increased respiratory system compliance, and decreased expiratory time.

Auto-PEEP may not be detected by routine measurements of pressure at the airway opening because most of the pressure drop occurs across the airways. Moreover, measurements of auto-PEEP are difficult to make in spontaneously breathing patients. When patients are not making spontaneous breathing efforts, auto-PEEP can be assessed as the difference in pressure between the set PEEP and the pressure obtained when the airway opening is occluded at the end of expiration (Fig. 105-1). It can also be assessed by the change in plateau pressure after a prolonged pause during volume cycle ventilation. If it is considered safe for the patient, a rapid estimate of the effect of auto-PEEP on cardiovascular hemodynamics can be obtained by transiently disconnecting the ventilator and allowing the auto-PEEP to approach zero during a long expiration. If the auto-PEEP is less than 5 cm H_2O, it is unlikely to cause clinically important changes in the measured intravascular pressures.

If auto-PEEP is not considered in the interpretation of respiratory mechanics, measurements of respiratory system compliance will be falsely low. Dynamic hyperinflation can be measured as the volume of gas that is released when the expiratory time of a given breath is lengthened by 20 to 30 seconds. The techniques for measuring auto-PEEP are based on the assumptions that no respiratory efforts are made and that the alveoli communicate with the airway opening, thereby allowing equilibration of pressures or exhalation of trapped gas. However, this assumption is not necessarily correct in patients with severe airways obstruction (e.g., status asthmaticus) because some airways may be completely closed.

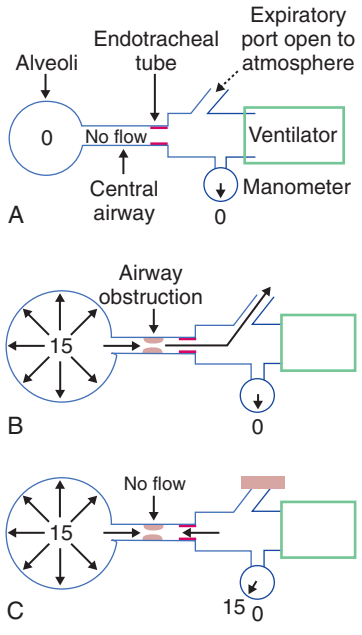

FIGURE 105-1. The relationships among alveolar, central airway, and ventilator circuit pressure at the end of exhalation under the following conditions: A, Normal conditions (no auto–positive end-expiratory pressure [auto-PEEP]). **B,** Severe dynamic airway obstruction with the expiratory port open. **C,** Severe dynamic airway obstruction with the expiratory port occluded at the end of exhalation. The auto-PEEP level is identified by creating an end-expiratory hold, thereby allowing the alveolar, central airway, and ventilator circuit pressures to equilibrate because there is no flow in the circuit. During equilibration, the level of auto-PEEP can be read on the manometer in the ventilator circuit. (Modified from Pepe PE, Marini JJ. Occult positive end-expiratory pressure in mechanically ventilated patients with airflow obstruction: the auto-PEEP effect. *Am Rev Respir Dis.* 1982;126:166-170.)

Auto-PEEP should be suspected whenever flow at end expiration is detectable or when a patient fails to trigger the ventilator consistently with inspiratory efforts. This failure to trigger the ventilator occurs because the patient must generate sufficient pressure to overcome the level of auto-PEEP before a negative deflection of pressure or generation of inspiratory flow (either of which may be used by the ventilator to detect the onset of inspiration) is sensed at the airway opening.

Auto-PEEP and the attendant dynamic hyperinflation have numerous detrimental consequences. In a patient who is not breathing spontaneously, dynamic hyperinflation increases pleural pressure and right atrial pressure, thereby leading to a decrease in the driving pressure for venous return, with a concomitant decrease in cardiac output. This effect can be magnified in patients with airway obstruction immediately after intubation and initiation of mechanical ventilation because compensatory mechanisms to enhance venous return are impaired by pharmacologic agents that are often used to prepare the patient for endotracheal tube insertion and that also reduce venous and arterial tone. In such patients, auto-PEEP can also lead to gross misinterpretation of vascular pressures. For example, the absolute value of capillary wedge pressure will be directly affected by the increase in intrathoracic pressure during auto-PEEP. The clinician may interpret this high (absolute) capillary wedge pressure as indicating adequate ventricular filling when, in fact, transmural capillary wedge pressure is low because intrathoracic pressure is also high. This misinterpretation, coupled with the decreased cardiac output related to the high intrathoracic pressure, may suggest the diagnosis of cardiogenic shock rather than the correct diagnosis of auto-PEEP.

In a spontaneously breathing patient, dynamic hyperinflation can markedly increase the oxygen cost of breathing for two reasons. First, because the respiratory system is stiffer at higher lung volumes, more energy is required to complete each ventilatory cycle. Second, to initiate flow into the lung, the patient must generate a pressure in the alveolar zone that is lower than atmospheric pressure. However, if dynamic hyperinflation is present, the patient first has to generate an inspiratory effort sufficient to overcome the (positive) end-expiratory alveolar pressure before he or she begins to lower alveolar pressure to less than atmospheric pressure to initiate airflow. The increase in lung volume associated with dynamic hyperinflation also has an impact on the effectiveness of the ventilatory muscles; at high lung volumes, the diaphragm is relatively flat, so it is at a mechanical disadvantage in producing changes in pleural pressure.

Auto-PEEP is more likely to occur in patients with airway obstruction. Avoidance of high levels of auto-PEEP by approaches such as controlled hypoventilation—during which minute ventilation is minimized, with the attendant hypercapnia—is a fundamental approach to consider in ventilating patients who have severe airway obstruction. Treatment of the detrimental hemodynamic consequences of auto-PEEP include infusion of fluids and, most important, decreasing the level of auto-PEEP, which can usually be accomplished by increasing expiratory time, decreasing airway resistance (e.g., bronchodilators, when appropriate), or decreasing minute ventilation. The last approach is usually the most effective ventilatory maneuver, but it results in an increase in the $Paco_2$.

Ventilator-Induced Lung Injury

Mechanical ventilation itself can lead to numerous types of lung injury[2,3] (Fig. 105-2) in addition to oxygen toxicity[4] when high levels of inspired oxygen concentrations are administered. Barotrauma refers to pulmonary air leaks, such as pneumothorax and pneumomediastinum. However, a much more subtle injury—diffuse alveolar damage presenting as pulmonary edema—can also occur. For both types of injury, the critical factor is the degree of overdistention of the lung, best assessed by the transpulmonary pressure (P_{tp}), the airway opening minus pleural pressure (P_{pl}). The esophageal pressure, measured with an esophageal balloon, estimates P_{pl}, although this measurement is not routinely performed in clinical practice.

The usual pressures measured during mechanical ventilation are airway pressures referenced to atmospheric pressure. The peak inspiratory pressure (PIP) is easy to measure, but its interpretation is not always simple. PIP represents the sum of the pressure needed to overcome the resistance to flow plus the pressure required to inflate the lungs. Thus, a high PIP does not necessarily indicate an increased propensity to overdistending the lung with subsequent ventilator-induced lung injury. For example, for a given inspiratory flow, use of a smaller endotracheal tube will increase PIP, but the danger of pulmonary overdistention is no greater than would be present if the patient was ventilated with a larger-bore tube and a lower PIP. The plateau pressure (P_{plat}) is the airway pressure at the end of an end-inspiratory pause (usually

>0.5 second) and is relatively easy to measure at the bedside if the patient is passive (e.g., receiving a paralytic agent). Depending on P_{pl}, it has some relationship with the development of overdistention. Although P_{pl} can vary greatly and no single value of P_{plat} can be defined as "dangerous" from a lung injury perspective, a reasonable maximal value of P_{plat} in patients with ARDS is 30 cm H_2O.

Certain caveats should be noted in interpreting P_{plat} and PIP, related to associated changes in P_{pl}. If the patient is breathing spontaneously, P_{pl} will be negative, and overdistention may occur even with a P_{plat} much lower than 30 cm H_2O. Conversely, in a patient who is either paralyzed or not making ventilatory efforts and who has a stiff chest wall (e.g., due to ascites, obesity, pregnancy), as airway pressure increases, most of the pressure drop will be dissipated across the chest wall, thus leading to values of P_{pl} that are positive. In this setting, a high P_{plat} may not be indicative of a high P_{tp} and hence may not indicate increased lung distention. Thus, the physician caring for a patient receiving mechanical ventilatory support must interpret the measured airway pressures within the clinical context. Measurement of P_{pl}, as noted earlier, may help resolve these difficulties.

During mechanical ventilation, some areas of the lung may undergo cyclic recruitment and de-recruitment. This process, which is of particular importance in patients who have ARDS, has been termed atelectrauma and can cause significant lung injury. The precise mechanisms of injury are not entirely clear but are thought to result from shear stress due to opening and closing of lung units, regional hypoxia in atelectatic lung units, and effects on surfactant. Prevention of this type of injury provides part of the rationale for the use of PEEP to maintain recruitment of lung units during tidal ventilation (Video 105-1).

Finally, evidence suggests that mechanical ventilation strategies that promote overdistention and atelectrauma can lead to an inflammatory response in the lung, a mechanism of injury termed biotrauma, with the release of proinflammatory cytokines and chemokines. To the extent that these mediators can translocate from the lung into the systemic circulation, they could potentially lead to dysfunction of other organs (Fig. 105-3). This concept suggests that optimal ventilatory strategies are important not only to maintain lung function but also to prevent the development of multiple-organ dysfunction (Chapter 104), a condition that is reasonably frequent in very sick, ventilated patients. This hypothesis may explain the decreased mortality recently observed with a strategy designed to avoid overdistention in a large randomized trial of mechanical ventilation in patients with ARDS.

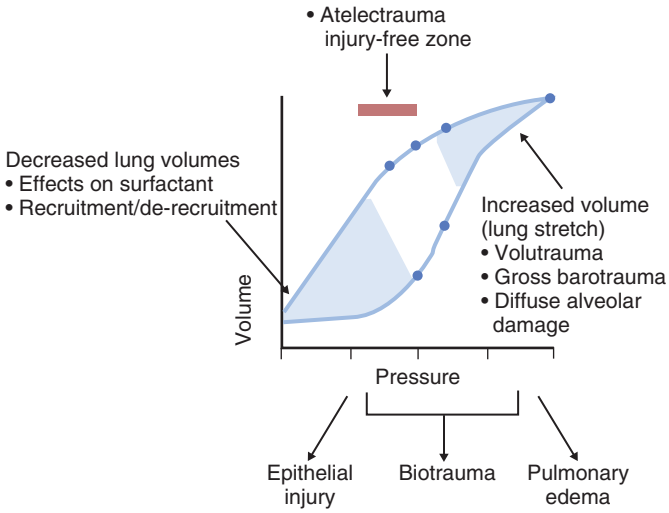

FIGURE 105-2. Schematic representation of the pressure-volume curve of a lung with diffuse alveolar edema. Mechanical ventilation can induce or worsen lung injury by numerous mechanisms when ventilation occurs at high lung volumes or when ventilation occurs at low lung volumes. Lung-protective strategies during ventilation of patients with acute respiratory distress syndrome should try to keep the ventilatory pattern in the *injury-free zone*. Data in patients confirm the benefit of ensuring that overdistention does not occur.

SPECIFIC COMMON TREATMENT SCENARIOS

Initiation of Mechanical Ventilation

The initiation of mechanical ventilation involves several steps in clinical decision making (Table 105-1). Despite the utility of such guidelines, each patient must be evaluated for specific factors that could modify the recommendation or mandate an alternative.

Acute Respiratory Distress Syndrome

Patients with ARDS (Chapter 104) have noncardiogenic pulmonary edema, with a reduced functional residual capacity and a mortality rate that commonly exceeds 25%. Although therapy may be available for the underlying disease process that led to the development of ARDS (e.g., antibiotics for a predisposing pneumonia), no effective therapy directly reverses diffuse alveolar damage. These patients require mechanical ventilation as supportive therapy to improve oxygenation and to decrease the oxygen cost of breathing until their lungs recover from the primary insult that led to the alveolar damage. The major goal in treating these patients is to provide adequate gas exchange while ensuring that damaged lungs are not further injured by whatever ventilatory strategy is required to provide sufficient oxygenation. The balanced approaches to minimize lung injury are termed lung-protective ventilation or lung-protective strategies.

Lung-Protective Ventilation Strategies

The lungs of a patient with ARDS are stiff and are characterized on computed tomographic scans by patchy, heterogeneous infiltrates that consist of airless atelectatic or consolidated regions. Many patients have a dependent zone that

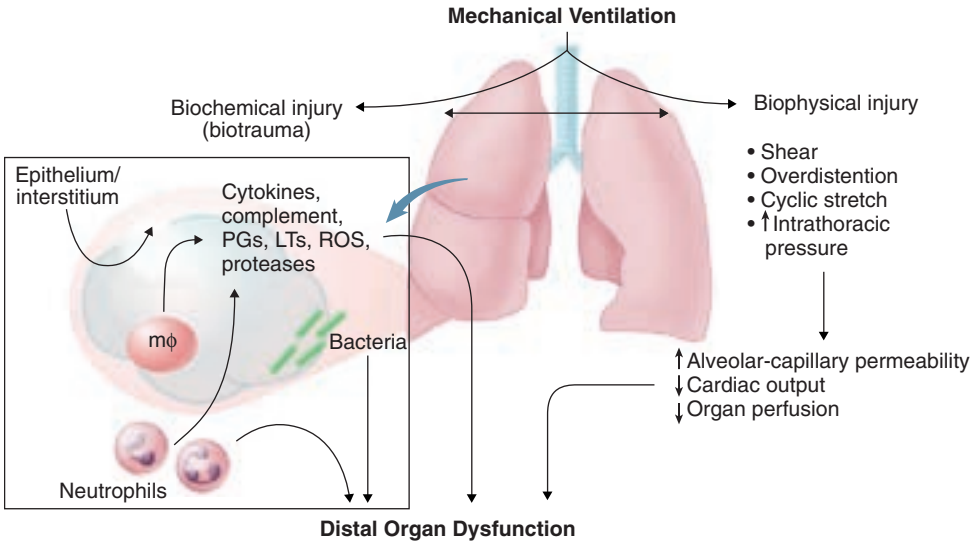

FIGURE 105-3. Mechanisms by which mechanical ventilation may lead to distal organ dysfunction. LTs = leukotrienes; mϕ = macrophages; PGs = prostaglandins; ROS = reactive oxygen species. (Modified from Slutsky AS, Tremblay LN. Multiple system organ failure: is mechanical ventilation a contributing factor? *Am J Respir Crit Care Med*. 1998;157:1721-1725.)

TABLE 105-1 STEPS AND GUIDELINES FOR INITIATION OF MECHANICAL VENTILATION*

1. Ventilatory mode
   Unintubated patients
   - NIV for patients with COPD and acute hypercapnic respiratory failure if alert, cooperative, and hemodynamically stable
   - NIV not routinely recommended for acute hypoxemic respiratory failure

   Intubated patients
   - Assist/control with volume-limited ventilation as initial mode
   - Consider specific indications for PCV or HFOV (see text) in acute lung injury
   - SIMV: consider if some respiratory effort, dyssynchrony
   - PSV: consider if patient's effort good, ventilatory needs moderate to low, and patient more comfortable during PSV trial
2. Oxygenation
   - If infiltrates on chest radiograph, then
     F_{IO_2}: begin with 0.8-1.0, reduce according to Spo_2
     PEEP: begin with 5 cm H_2O, increase according to Pao_2 or Spo_2, F_{IO_2} requirements, and hemodynamic effects; consider PEEP/F_{IO_2} "ladder" (see Fig. 105-4); goal of Spo_2 >90%, $F_{IO_2} \leq 0.6$
   - No infiltrates on chest radiograph (COPD, asthma, PTE)
     F_{IO_2}: start at 0.4 and adjust according to Spo_2 (consider starting higher if pulmonary embolism is strongly suspected)
3. Ventilation
   - Tidal volume: begin with 8 mL/kg PBW (see Fig. 105-4 for formulas); decrease to 6 mL/kg PBW over a few hours if acute lung injury present (see Fig. 105-4)
   - Rate: begin with 10-20 breaths/min (10-15 if not acidotic; 15-20 if acidotic); adjust for pH; goal pH > 7.3 with maximal rate of 35; may accept lower goal if minute ventilation high
4. Secondary modifications
   - Triggering: in spontaneous modes, adjustment of sensitivity levels to minimize effort
   - Inspiratory flow rate of 40-80 L/min; higher if tachypneic with respiratory distress or if auto-PEEP present, lower if high pressure in ventilator circuit leads to a high-pressure alarm
   - Assessment of auto-PEEP, especially in patients with increased airways obstruction (e.g., asthma, COPD)
   - I/E ratio: 1 : 2, either set or as function of flow rate; higher (1 : 3 or more) if auto-PEEP present
   - Flow pattern: decelerating ramp reduces peak pressure
5. Monitoring
   - Clinical: blood pressure, ECG, observation of ventilatory pattern including assessment of dyssynchrony, effort or work by the patient; assessment of airflow throughout expiratory cycle
   - Ventilator: tidal volume, minute ventilation, airway pressures (including auto-PEEP), total compliance
   - Arterial blood gases, pulse oximetry

*Decisions within this algorithm will be influenced by the specific conditions of the individual patient.
COPD = chronic obstructive pulmonary disease; ECG = electrocardiogram; F_{IO_2} = fraction of inspired oxygen; HFOV = high-frequency oscillatory ventilation; I/E ratio = inspiratory-to-expiratory ratio; NIV = noninvasive ventilation; Pao_2 = partial pressure of oxygen in arterial blood; PBW = predicted body weight; PCV = pressure-controlled ventilation; PEEP = positive end-expiratory pressure; PSV = pressure-support ventilation; PTE = pulmonary thromboembolism; SIMV = synchronized intermittent mandatory ventilation; Spo_2 = arterial oxygen saturation by pulse oximetry.

is consolidated, atelectatic, or fluid filled; a nondependent zone that looks relatively normal; and a middle zone that has some collapsed regions that can be recruited to resemble the nondependent regions if sufficiently increased levels of airway pressure are transiently used (these approaches are called recruitment maneuvers). Arterial oxygen saturation can often be increased by high tidal volumes but at the expense of regional overdistention of those lung units that were not affected by the disease process itself—a treatment strategy that can lead, over time, to worse lung injury and poorer clinical outcomes.

The injury caused by mechanical ventilation can be reduced by ventilatory strategies that avoid or minimize regional lung overdistention: limiting inspiratory pressure to some "safe" level or using smaller tidal volumes to limit end-inspiratory stretch, or both. However, in some patients, this lower "dose" of ventilation results in higher levels of $Paco_2$ (so-called permissive hypercapnia) and a lower pH. Higher tidal volumes (12 mL/kg predicted body weight) yielded more normal blood gases, but lower tidal volumes (6 mL/kg predicted body weight) decreased mortality by 22% (from an absolute value of 40 to 31%) in a large clinical trial (Fig. 105-4).

Data also suggest that limiting tidal volumes in ventilated patients who are intubated for reasons other than ARDS prevents injury later in the course of their intensive care unit stay.[5] A lung-protective strategy with limitation of tidal volume should be considered in ventilated patients who are at high risk for development of acute lung injury or ARDS.

Positive End-Expiratory Pressure

PEEP traditionally has been used to improve oxygenation while at the same time allowing reduction in F_{IO_2} to relatively nontoxic levels. Within the context of the current paradigm of trying to minimize iatrogenic complications of mechanical ventilation, PEEP is a therapy that can potentially minimize the injury caused by ventilation at low lung volumes by recruiting lung units and keeping them open. The critical issues are how to assess the level of PEEP in an individual patient and how to determine whether the procedures to recruit the lung units and keep them open are less harmful than allowing the lung units to remain de-recruited. One experimental option is chest computed tomography to assess whether areas of the lung are recruited, but this technique is not practical for routine assessment.

Data are inconclusive regarding the benefits of higher (≈13 cm H_2O) compared with lower (≈8 cm H_2O) PEEP levels, and PEEP levels often are individualized on the basis of a PEEP/F_{IO_2} table (Fig. 105-4). Higher PEEP levels appear to be associated with decreased mortality in ARDS patients with Pao_2/F_{IO_2} of less than 200 mm Hg but not in patients with higher Pao_2/F_{IO_2} ratios. PEEP guided by esophageal pressure measured by an intraesophageal balloon[6] can significantly increase Po_2 levels and respiratory compliance compared with treatment guided by a standard protocol.

Adjunctive Approaches for Ventilating ARDS Patients

The neuromuscular blocking agent cisatracurium (15 mg intravenous bolus followed by 37.5 mg/hour infusion) can decrease mortality in ARDS patients with Pao_2/F_{IO_2} ratios below 150 mm Hg when it is given for 48 hours in patients with early ARDS. The putative mechanism is a decrease in ventilator-induced lung injury.

The use of prone position in patients with ARDS can improve oxygenation compared with the supine position by permitting a more even distribution of pleural pressure, thereby reducing ventilator-induced lung injury and decreasing F_{IO_2}. Use of the prone position has decreased mortality by an absolute 9% in patients who have Pao_2/F_{IO_2} below 100 mm Hg and by an absolute 16% in patients with PaO_2/FIO_2 below 150 mm Hg. A critical factor in the use of the prone position is proper training of medical personnel in how to place patients safely in the prone position.

Obstructive Airways Diseases

The major physiologic abnormality in patients with obstructive airways diseases (e.g., COPD, asthma) is an increase in airway resistance leading to expiratory airflow limitation; patients may also have a concomitant increase in minute ventilation. These factors may lead to dynamic hyperinflation, which is associated with numerous complications (described earlier), including respiratory muscle compromise, increased oxygen cost of breathing, and hemodynamic compromise. Thus, the main goals in the ventilatory support of patients with obstructive airway diseases are to minimize auto-PEEP, to rest the respiratory muscles, to maintain adequate gas exchange, and to decrease the oxygen cost of breathing while simultaneously minimizing the iatrogenic complications of mechanical ventilation. These strategies allow time for the diagnosis and treatment of the primary cause of the exacerbation (Chapters 87 and 88).

Noninvasive Ventilation

For patients who have acute respiratory failure resulting from an exacerbation of COPD and who require ventilatory support, the preferred approach is NIV if the patient is hemodynamically stable, alert, and cooperative and does not need to be intubated to protect the airway.[7] It is important to choose a comfortable mask and to reassure the patient because some patients find the mask difficult to tolerate. This strategy may be applied with several ventilation modes, including pressure support and bilevel positive airway pressure. The ventilation settings are adjusted to improve gas exchange and to ensure the patient's comfort. Despite this approach, some patients with COPD require intubation and ventilation because of cardiac or respiratory arrest, agitation, increased sputum, worsening respiratory failure, or other concomitant severe disorders.

Intubation and Ventilation

The key goal after intubation is to minimize the detrimental effects of dynamic hyperinflation. The most effective way to do this is to decrease the minute

Ventilatory Strategy for Patients with ARDS*

Goal 1: Low Vt/P_{plat}

Initiation:

Calculate PBW
—Male:
50 + 2. 3 (height [inches] – 60)
—Female:
45.5 + 2.3 (height [inches] – 60)

Initiate volume assist control
—start with 8 mL/kg, and ↓ to 6 mL/kg over a few hours

↓

Keep P_{plat} (based on 0.5-sec pause) < 35 cm H_2O

If P_{plat} > 30 cm H_2O, ↓ Vt by 1 mL/kg to 5 or 4 mL/kg

If P_{plat} < 25 AND Vt < 6 mL/kg, ↑Vt by 1 mL/kg until P_{plat} > 25 cm H_2O OR Vt = 6 mL/kg

If patient severely distressed and/or breath stacking, consider ↑Vt to 7 or 8 mL/kg, as long as $P_{plat} \leq 30$ cm H_2O†

Goal 2: Adequate Oxygenation

Specific goal:

Pa_{O_2} 55-80 mm Hg
or
Sp_{O_2} 88-95%

Use only F_{IO_2} /PEEP combinations shown below to achieve this target

- if oxygenation is low, choose F_{IO_2} /PEEP combination (from F_{IO_2} /PEEP table) to the right
- if oxygenation is high, choose F_{IO_2} /PEEP combination to the left

Goal 3: Arterial pH

Goal:

pH: 7.30–7.45

Acidosis algorithm

If pH 7.15–7.30
- ↑ set rate until pH > 7.30 or Pa_{CO_2} < 25 mm Hg (max RR = 35)
- if RR = 35 & pH < 7.30 $NaHCO_3$ may be given

If pH < 7.15
- ↑ set RR to 35
- if set RR = 35 & pH < 7.15, Vt may be ↑ in 1 mL/kg steps until pH > 7.15

(P_{plat} target may be exceeded)

Alkalosis algorithm

If pH > 7.45
- ↓ set RR until patient RR > set RR

(minimum set RR = 6/min)

F_{IO_2}/PEEP Table

F_{IO_2}	0.3	0.4	0.4	0.5	0.5	0.6	0.7	0.7	0.7	0.8	0.9	0.9	0.9	1.0
PEEP	5	5	8	8	10	10	10	12	14	14	14	16	18	18–24

*Based on ARDS Network Algorithm
†If compliance of the chest wall is markedly decreased (e.g., massive ascites), it may be reasonable or necessary (if the patient is very hypoxemic) to allow a P_{plat} >30 cm H_2O.

FIGURE 105-4. Ventilatory strategy for patients with the acute respiratory distress syndrome (ARDS). Several caveats should be considered in using the low tidal volume strategy. (1) Tidal volume (Vt) is based on predicted body weight (PBW), not actual body weight; PBW tends to be about 20% lower than actual body weight. (2) The protocol mandates decreases in the Vt lower than 6 mL/kg of PBW if the plateau pressure (P_{plat}) is greater than 30 cm H_2O and allows small increases in Vt if the patient is severely distressed or if there is breath stacking, as long as P_{plat} remains at 30 cm H_2O or lower. (3) Because arterial carbon dioxide (CO_2) levels will rise, pH will fall; acidosis is treated with increasingly aggressive strategies dependent on the arterial pH. (4) The protocol has no specific provisions for the patient with a stiff chest wall, which in this context refers to the rib cage and abdomen; in such patients, it seems reasonable to allow P_{plat} to increase to more than 30 cm H_2O, even though it is not mandated by the protocol; in such cases, the limit on P_{plat} may be modified on the basis of analysis of abdominal pressure, which can be estimated by measuring bladder pressure. RR = respiratory rate; Sp_{O_2} = oxygen saturation based on pulse oximeter.

ventilation, even if this means an increase in Pa_{CO_2}—a strategy known as permissive hypercapnia or controlled hypoventilation. Judicious use of sedation may decrease carbon dioxide production and improve patient-ventilator synchrony, although the avoidance of sedation can reduce the duration of ventilation and hospitalization.[■] In a randomized study, no difference was found between dexmedetomidine and midazolam in time at targeted sedation level, but dexmedetomidine resulted in less time on mechanical ventilation, less delirium, and less hypertension and with less tachycardia but more bradycardia.[■] Care must be taken in the use of paralytic agents, especially when patients with asthma are also receiving corticosteroids, because they may lead to prolonged muscle weakness and resulting difficulty in extubation and post–intensive care unit recovery.

Increasing expiratory time by use of a higher peak inspiratory flow may be somewhat helpful, but it is not nearly as effective as decreasing minute ventilation. What level of Pa_{CO_2} (and pH) should be tolerated is not known with certainty, but maintaining pH higher than approximately 7.20 is a reasonable target if the patient is not having side effects (e.g., arrhythmias, increasing right-sided heart failure), although much lower values have been reported in clinical studies.

In patients with COPD who are spontaneously breathing, the addition of external (set) PEEP at a level that is just less than what is necessary to overcome the auto-PEEP fully may not increase P_{plat} and may decrease the inspiratory effort that the patient needs to generate to initiate inspiratory airflow. This strategy does not appear to be as effective in patients with status asthmaticus, in whom it may cause an increase in P_{plat}. Measurements of auto-PEEP by airway occlusion may be inaccurate in some patients with status asthmaticus, probably because of gas trapping at the end of expiration with closed-off lung regions that do not communicate with the central airways.

DISCONTINUATION OF MECHANICAL VENTILATION

To minimize the iatrogenic consequences of intubation and mechanical ventilation, discontinuation of ventilatory support and extubation should occur as expeditiously as possible. However, if discontinuation is attempted too early, patients may deteriorate and require urgent reintubation.

From the moment that mechanical ventilation is instituted, it is important that the clinician start planning for eventual discontinuation of ventilatory support. A key aspect of this approach is serial evaluation, with aggressive treatment of the factors contributing to the patient's ventilatory dependence, including respiratory systems factors (e.g., respiratory muscles), cardiovascular factors (e.g., myocardial ischemia), neurologic factors (e.g., respiratory muscle weakness), and metabolic factors (e.g., increased oxygen consumption).

Two major types of weaning strategies have been used historically: a ventilatory mode thought to hasten the weaning process; and daily monitoring of the patient for criteria suggesting the likelihood of successful weaning, with a trial of spontaneous breathing for patients deemed likely to succeed. Studies of ventilatory modes of weaning have included trials in which patients are allowed to breathe spontaneously from a fresh gas supply delivered to the endotracheal tube (a so-called T-tube), trials of IMV, and trials of pressure-support ventilation. With all approaches, the level of support is gradually decreased until extubation is tolerated by the patient. These methods have been compared in randomized controlled trials, with mixed results, although weaning with IMV appeared less favorable in most trials. Likewise, use of ventilatory criteria to predict weaning success has been disappointing, mainly because some patients who fail to meet the criteria will be successfully weaned if they are given the opportunity to breathe spontaneously. An easily measured variable is the so-called rapid shallow breathing index, in which the respiratory rate is divided by tidal volume (in liters), with a value of less than 105 suggesting the ability to be weaned; however, false-negative and false-positive test results commonly occur.

More recently, the approach to weaning has been based on the concept that a patient is ready to be removed from ventilatory support when the underlying disease process that led to the intubation has resolved or improved substantially. Rather than applying rigorous ventilatory criteria, the only requirements are that the patient be clinically stable (i.e., has shown improvement in the underlying process), be hemodynamically stable, and have oxygen requirements that can be met by face mask once the patient is extubated.[8] If the patient meets these general criteria, a spontaneous breathing trial is recommended (Fig. 105-5); if the patient passes the trial, the patient

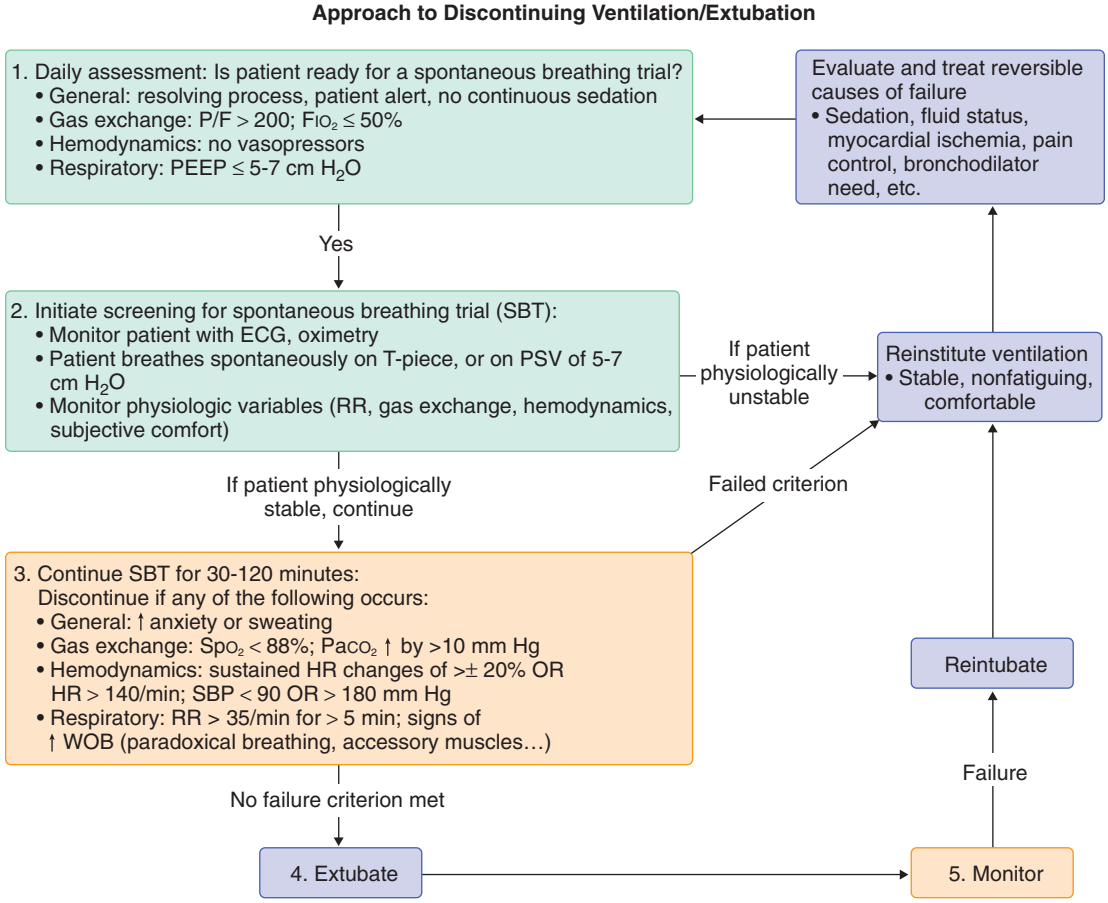

FIGURE 105-5. Algorithm for assessing whether a patient is ready to be liberated from mechanical ventilation and extubated. ECG = electrocardiogram; HR = heart rate; P/F = PaO_2/FIO_2 ratio; PSV = pressure support ventilation; RR = respiratory rate; SBP = systolic blood pressure; SpO_2 = oxygen saturation based on pulse oximeter; WOB = work of breathing.

can be extubated. A corollary is that gradual weaning is not necessary; instead, patients should be assessed on a daily basis regarding their suitability for removal from ventilatory support, and if they are not ready, a comfortable, nonfatiguing form of mechanical ventilation should be used between the assessments. Assisted modes of ventilation are preferred between the spontaneous breathing trials. After extubation, evidence suggests that noninvasive ventilation may be beneficial in hypercapnic or high-risk patients.[9]

An important recommendation in relation to weaning or discontinuation of mechanical ventilation relates to evidence that intensive care units should develop weaning or discontinuation protocols that can be implemented by health care professionals other than physicians. Three large randomized trials demonstrated that protocols implemented by health care professionals other than physicians improved care and were associated with substantial savings in costs compared with standard management approaches, even though the specifics of the protocols were different. A strategy that paired spontaneous awakening, based on the interruption of sedatives, with spontaneous breathing trials improved extubation rates, reduced intensive care length of stay, and decreased mortality by 32%.

A major issue to assess before extubation is the patency of the patient's airway and whether the patient will be able to clear secretions after extubation. Assessment of the likely patency of the upper airway can be achieved by use of the *cuff-leak volume*, which is the difference between the inspiratory and expiratory tidal volume when the cuff of the endotracheal tube is deflated. If this volume is more than 110 mL, it is usually an indication that major upper airway obstruction will not occur after extubation. Although this test is not required before extubation, a low cuff-leak volume warrants added precautions, such as the availability of equipment and personnel for managing a difficult intubation, when the patient is extubated. In patients who have been ventilated for more than 36 hours, methylprednisolone (20 mg intravenously) started 12 hours before a planned extubation and repeated every 4 hours until tube removal substantially reduces postextubation laryngeal edema and reduces the need for reintubation by 50%.

Despite the use of all these techniques, approximately 5 to 25% of patients will have to be reintubated and have mechanical ventilation reinstituted. Once a patient is reintubated, it is again necessary to reevaluate the respiratory and nonrespiratory reasons for the failure.

The choice of the specific weaning protocol should be left to the individual institution and should be individualized to the specific group of patients considered. In instituting such protocols, several key issues should be recognized. First, protocols are guides that should not replace clinical judgment. If a clinician does not follow some aspect of the protocol, there should be a mechanism in place for keeping track of what recommendations were not accepted, with an explanation of the rationale; these data should be collated and used to reassess the protocol. Second, protocols should be viewed as dynamic structures that are open to change and should be reevaluated on a regular basis. Third, implementation of a protocol requires adequate resources, and an institution must make a commitment not only to develop protocols but also to implement and regularly assess them.

Grade A References

A1. Papazian L, Forel JM, Gacouin A, et al. Neuromuscular blockers in early acute respiratory distress syndrome. *N Engl J Med*. 2010;363:1107-1116.

A2. Ferguson ND, Cook DJ, Guyatt GH, et al. High-frequency oscillation in early acute respiratory distress syndrome. *N Engl J Med*. 2013;368:795-805.

A3. Young D, Lamb S, Shah S, et al. High-frequency oscillation for acute respiratory distress syndrome. *N Engl J Med*. 2013;368:806-813.

A4. Williams JW, Cox CE, Hargett CW, et al. Noninvasive Positive-Pressure Ventilation (NPPV) for Acute Respiratory Failure. AHRQ Comparative Effectiveness Reviews, No. 68. Report No. 12-EHC089-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2012.

A5. Gray A, Goodacre S, Newby DE, et al. Noninvasive ventilation in acute cardiogenic pulmonary edema. *N Engl J Med*. 2008;359:142-151.

A6. Terragni PP, Antonelli M, Fumagalli R, et al. Early vs late tracheotomy for prevention of pneumonia in mechanically ventilated adult ICU patients: a randomized controlled trial. *JAMA*. 2010;303:1483-1489.

A7. Young D, Harrison DA, Cuthbertson BH, et al. Effect of early vs late tracheostomy placement on survival in patients receiving mechanical ventilation: the TracMan randomized trial. *JAMA*. 2013;309:2121-2129.

A8. The Acute Respiratory Distress Syndrome Network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med*. 2000;342:1301-1308.

A9. Briel M, Meade M, Mercat A, et al. Higher vs lower positive end-expiratory pressure in patients with acute lung injury and acute respiratory distress syndrome: systematic review and meta-analysis. *JAMA*. 2010;303:865-873.
A10. Talmor D, Sarge T, Malhotra A, et al. Mechanical ventilation guided by esophageal pressure in acute lung injury. *N Engl J Med*. 2008;359:2095-2104.
A11. Alhazzani W, Alshahrani M, Jaeschke R, et al. Neuromuscular blocking agents in acute respiratory distress syndrome: a systematic review and meta-analysis of randomized controlled trials. *Crit Care*. 2013;17:R43.
A12. Hu SL, He HL, Pan C, et al. The effect of prone positioning on mortality in patients with acute respiratory distress syndrome: a meta-analysis of randomized controlled trials. *Crit Care*. 2014;18:R109.
A13. Guerin C, Reignier J, Richard JC, et al. Prone positioning in severe acute respiratory distress syndrome. *N Engl J Med*. 2013;368:2159-2168.
A14. Strom T, Martinussen T, Toft P. A protocol of no sedation for critically ill patients receiving mechanical ventilation: a randomised trial. *Lancet*. 2010;375:475-480.
A15. Riker RR, Shehabi Y, Bokesch PM, et al. Dexmedetomidine vs midazolam for sedation of critically ill patients: a randomized trial. *JAMA*. 2009;301:489-499.
A16. Girard TD, Kress JP, Fuchs BD, et al. Efficacy and safety of a paired sedation and ventilator weaning protocol for mechanically ventilated patients in intensive care (Awakening and Breathing Controlled trial): a randomised controlled trial. *Lancet*. 2008;371:126-134.
A17. Francois B, Bellissant E, Gissot V, et al. 12-h pretreatment with methylprednisolone versus placebo for prevention of postextubation laryngeal oedema: a randomised double-blind trial. *Lancet*. 2007;369:1083-1089.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bello G, De Pascale G, Antonelli M. Noninvasive ventilation: practical advice. *Curr Opin Crit Care*. 2013;19:1-8.
2. Slutsky AS, Ranieri VM. Ventilator-induced lung injury. *N Engl J Med*. 2013;369:2126-2136.
3. Saddy F, Sutherasan Y, Rocco PR, et al. Ventilator-associated lung injury during assisted mechanical ventilation. *Semin Respir Crit Care Med*. 2014;35:409-417.
4. Kallet RH, Matthay MA. Hyperoxic acute lung injury. *Respir Care*. 2013;58:123-141.
5. Wilson JG, Matthay MA. Mechanical ventilation in acute hypoxemic respiratory failure: a review of new strategies for the practicing hospitalist. *J Hosp Med*. 2014;9:469-475.
6. Akoumianaki E, Maggiore SM, Valenza F, et al. The application of esophageal pressure measurement in patients with respiratory failure. *Am J Respir Crit Care Med*. 2014;189:520-531.
7. Ramsay M, Hart N. Current opinions on non-invasive ventilation as a treatment for chronic obstructive pulmonary disease. *Curr Opin Pulm Med*. 2013;19:626-630.
8. Macintyre NR. Evidence-based assessments in the ventilator discontinuation process. *Respir Care*. 2012;57:1611-1618.
9. Thille AW, Richard JC, Brochard L. The decision to extubate in the intensive care unit. *Am J Respir Crit Care Med*. 2013;187:1294-1302.

REVIEW QUESTIONS

1. Which of the following is not usually a goal when positive end-expiratory pressure (PEEP) is used in patients with the acute respiratory distress syndrome (ARDS)?

A. Decrease the negative physiologic consequences of auto-PEEP
B. Recruit alveolar regions
C. Decrease ventilator-induced lung injury
D. Increase oxygenation
E. Increase functional residual capacity

Answer: A Although external PEEP may be useful in some patients with auto-PEEP, especially when it is caused by airway obstruction, this problem usually is not an issue in patients with ARDS.

2. Which of the following is not an exclusion criterion for use of noninvasive ventilation with COPD?

A. Facial trauma
B. Apnea
C. Hypercapnia
D. Loss of consciousness
E. Hemodynamic instability

Answer: C Hypercapnia is not a contraindication to the use of noninvasive ventilation in patients with COPD. In fact, COPD is one of the indications for its use. Noninvasive ventilation should not be used in most patients with severe facial trauma, because of the problems of fitting the mask, or in patients who are unconscious or apneic, because a patient has to be able to breathe spontaneously to trigger the ventilator. In addition, patients with hemodynamic compromise commonly need full respiratory support to minimize the oxygen cost of breathing.

3. Which of the following is not first-line therapy for a patient who has ARDS and a P/F ratio below 100 mm Hg?

A. Lung-protective mechanical ventilation
B. Use of the prone position
C. Neuromuscular blocking agents
D. High-frequency ventilation
E. Use of PEEP

Answer: D Two recent studies have shown that high-frequency ventilation does not have a role in the treatment of adult patients with ARDS early in their course.

4. Which of the following is an indication *not* to try a spontaneous breathing trial?

A. A Pao_2 = 88 mm Hg while breathing an Fio_2 = 1.0
B. PEEP = 5 mm Hg
C. Patient is not receiving vasopressors
D. Patient is alert
E. Patient is receiving minimal sedation

Answer: A In general, it is best not to attempt a spontaneous breathing trial if the patient has a P/F ratio below 200 mm Hg.

5. Which of the following statements about lung-protective strategy is incorrect?

A. Ventilator-induced lung injury can be associated with release of a number of mediators that may have systemic consequences.
B. Regional lung distention is an important factor in causing ventilator-induced lung injury.
C. Use of a lung-protective strategy is important even in patients who do not have ARDS.
D. An increased plateau pressure (>~30 cm H_2O) always indicates that the patient is at increased risk of ventilator-induced lung injury.
E. The use of a lung-protective strategy in an ARDS patient has been shown to decrease mortality.

Answer: D The key variable in terms of overdistention for ventilator-induced lung injury is overdistention of regional lung units. A patient may have an increased plateau pressure because of conditions that affect chest wall mechanics (e.g., ascites), and if so, the high plateau pressure may not indicate that the lung is being overstretched.

106

APPROACH TO THE PATIENT WITH SHOCK

EMANUEL P. RIVERS

DEFINITION

The key feature of shock is tissue hypoperfusion, not a specific level of systemic arterial blood pressure. The clinical picture may be cryptic or obvious.

EPIDEMIOLOGY

More than 1.2 million patients present in shock or develop shock in U.S. hospitals each year, at an annual cost of more than $100 billion. Shock can be categorized as hypovolemic, cardiogenic (Chapter 107), extracardiac/obstructive, distributive, or dissociative.[1]

PATHOBIOLOGY

The delivery and utilization of oxygen are essential for cellular viability, and the failure to deliver or to use oxygen is central to the concept of shock and its pathogenesis (Fig. 106-1). Systemic oxygen delivery (i.e., the amount of oxygen delivered to tissues by the arterial blood) depends on the concentration of hemoglobin in the blood, the fractional saturation of the hemoglobin with oxygen (Sao_2), the amount of oxygen dissolved in the blood (Pao_2), and cardiac output. Cardiac output is a product of stroke volume and heart rate. Stroke volume is determined by ventricular preload and afterload as well as by contractility of the right or left side of the heart. Systemic vascular resistance (SVR), the force resisting cardiac contraction, can be calculated by Equation 1:

$$SVR = (MAP - CVP)*80/CO \quad (1)$$

$$MAP = \text{diastolic blood pressure} + (\text{systolic} - \text{diastolic blood pressure})/3 \quad (2)$$

MAP denotes the mean systemic arterial blood pressure, CVP denotes central venous pressure, and CO denotes cardiac output. SVR is determined primarily by the degree of vasomotor tone in the precapillary smooth muscle sphincters.

The systemic circulation is normally autoregulated, so that when systemic arterial pressure increases, vessel diameter decreases to maintain flow at a steady level. The clinical significance of these relationships is apparent when a patient presents with a decrease in cardiac output, but a compensatory increase in SVR maintains a nearly normal MAP. Despite the nearly normal blood pressure, however, the patient is in "cryptic shock" because of tissue hypoperfusion. Compensatory mechanisms are organ specific. Blood flow to organs such as the heart and brain is carefully regulated and maintained over a wide range of blood pressures. In other organs, however, such as the intestine or liver, autoregulation is not as tightly maintained.

Systemic oxygen consumption, which is the amount of oxygen consumed by the body per minute, is calculated as the systemic oxygen delivery multiplied by the systemic oxygen extraction ratio. Oxygen demand is the amount of oxygen required by the tissues to avoid anaerobic metabolism. Normally, systemic oxygen delivery is sufficient so that systemic oxygen consumption is not altered by or dependent on changes in delivery. However, if systemic oxygen delivery drops below a critical value, a compensatory increase in the oxygen extraction ratio maintains systemic oxygen consumption at adequate levels to meet systemic oxygen demands. When this compensatory response in the oxygen extraction ratio is inadequate to meet systemic oxygen demands, a switch occurs from aerobic metabolism to the less efficient anaerobic metabolism. The result is depletion of adenosine triphosphate (ATP) and intracellular energy reserves. Intracellular acidosis occurs, and anaerobic glycolysis leads to the production of lactate. Below this critical value of systemic oxygen delivery, systemic oxygen consumption is dependent on systemic oxygen delivery, a relationship termed *physiologic oxygen supply dependency.* This critical value of systemic oxygen delivery varies substantially because comorbid or preexisting conditions affect the rate of systemic oxygen utilization.

A pathologic systemic oxygen delivery dependency exists in patients with sepsis, trauma, and acute respiratory distress syndrome (ARDS) and after resuscitation from prolonged cardiac arrest. These patients have systemic oxygen delivery in the normal or elevated range but an impairment of oxygen utilization. This condition of cytopathic tissue hypoxia is a result of maldistribution of blood flow or a defect in the utilization of substrate at the microcirculatory or subcellular level. This pathologic supply dependency is accompanied by very high mixed venous oxygen saturation levels or venous hyperoxia as well as by elevated lactate levels. This process is believed to be an important mechanism of cellular damage in various forms of shock.

Compensatory Responses

Minor decreases in arterial blood pressure and systemic oxygen delivery activate the baroreceptor reflex through stretch receptors or sensing mechanisms located in the carotid sinus, splanchnic vasculature, aortic arch, right atrium, and juxtaglomerular apparatus of the kidney as well as through chemoreceptors sensitive to concentrations of carbon dioxide or oxygen located in the central nervous system, mostly in the medulla. These compensatory responses mediated by activation of the sympathetic nervous system include the following: release of cortisol, aldosterone, and epinephrine; activation of the renin-angiotensin system; release of arginine vasopressin from the posterior pituitary; augmentation of myocardial contractility and heart rate; constriction of arterial and venous capacitance vessels, particularly in the splanchnic bed, thereby augmenting venous return; redistribution of blood flow away from skeletal muscle beds and the splanchnic viscera; and creation of a local tissue environment to enhance the unloading of oxygen to tissues and to improve its extraction because of acidosis, pyrexia, and increased red blood cell 2,3-diphosphoglycerate.

Noncompensatory Responses

Noncompensatory responses develop when physiologic adjustments are exaggerated or lead to pathologic results. Vasodilatory shock results from many sources, including unregulated nitric oxide synthesis, inadequate ATP synthesis in vascular smooth muscle cells, activation of the enzyme poly(ADP-ribose) polymerase 1, lipid mediators, and opening of ATP-sensitive potassium channels in vascular smooth muscle cells. This multifaceted insult leads to interstitial fluid and cellular edema, which impairs oxygen diffusion from capillary to cell, causing a failure of energy-dependent ion transport, the production of lactate, and the inability to maintain normal transmembrane gradients of potassium, chloride, and calcium. Cells lose their ability to use available oxygen as a result of mitochondrial dysfunction, abnormal carbohydrate metabolism, and failure of many energy-dependent enzyme reactions.

Acidosis commonly accompanies shock. When a molecule of ATP is hydrolyzed to adenosine diphosphate (ADP) and inorganic phosphate, the reaction also generates a proton. The net yield of protons is positive when ATP is hydrolyzed in the cell and then regenerated only by the anaerobic

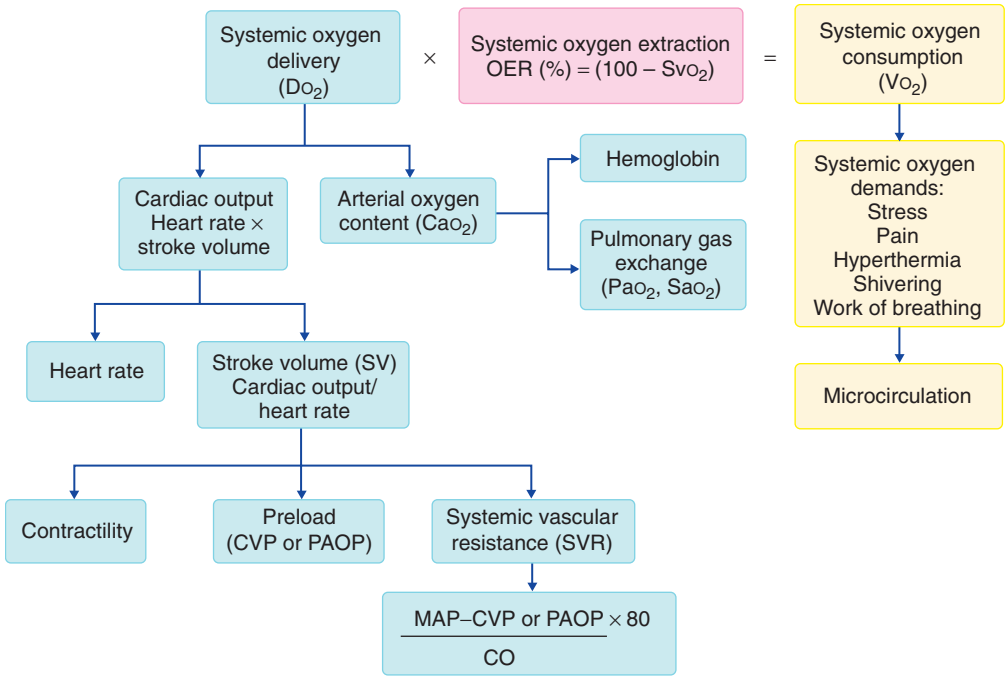

FIGURE 106-1. The hemodynamic, oxygen transport, and oxygen utilization components of shock management. Systemic oxygen delivery (Do_2) is affected by cardiac output (CO) and arterial oxygen content. The cardiac, pulmonary, and blood determinants of Do_2 are shown. CVP = central venous pressure; MAP = mean arterial pressure; OER = oxygen extraction ratio; PAOP = pulmonary artery occlusion pressure.

breakdown of glucose. Thus, during anaerobic glycolysis, the use of ATP to power cellular processes, coupled with the anaerobic production of ATP by substrate-level phosphorylation reactions, results in the development of acidosis.

Cells in organs such as the kidneys, liver, and brain can convert lactate into glucose through gluconeogenesis or oxidize lactate to pyruvate and then, ultimately, to carbon dioxide and water. Lactate levels are a reflection of tissue hypoxia, clearance, and alternative sources of production. When the splanchnic circulation is compromised in shock, hepatic lactate clearance is impaired, contributing to the buildup of lactate levels in the circulation. In sepsis, however, the rate of glycolysis increases even in the absence of tissue hypoxia. This phenomenon, which has been termed accelerated aerobic glycolysis, may reflect a change in the ratio of the active to the inactive form of pyruvate dehydrogenase, which is the rate-limiting step for the entry of substrate into the mitochondrial tricarboxylic acid cycle.

When systemic oxygen delivery continues to fail to meet systemic oxygen demands, the oxygen debt accumulates. Three stages of shock can ensue. The first stage, which is called early, reversible, or compensated shock, is characterized by compensatory responses to minimize tissue injury. This stage of shock can be self-limited, with full recovery and minimal residual morbidity, if the cause is recognized and treated early. If substantial oxygen debt persists without timely repayment or resolution, inflammation and cellular and microvascular injury define the second stage of shock, which is associated with a prolonged recovery and is typically complicated by organ failure, such as acute lung injury and acute kidney injury. The third stage is late, irreversible, or decompensated shock. In this situation, the oxygen debt is large, and repayment is slow or nonexistent. When shock reaches this point, cellular and tissue injury is extensive and largely irreversible. Progression to multisystem organ failure or death is inevitable, regardless of therapy.

CLINICAL MANIFESTATIONS

The five general types of shock are cardiogenic, distributive, hypovolemic, obstructive, and dissociative. The distinction among these five shock syndromes can be made by combining the history, clinical picture, and hemodynamic measurements. Cardiogenic shock (Chapter 107) and shock syndromes related to sepsis (Chapter 108) are covered in detail elsewhere.

The clinical manifestation of shock is variable and depends on the initiating cause and the response of multiple organs. Shock typically is manifested as absolute or relative systemic arterial hypotension and evidence of end-organ dysfunction (Table 106-1). Even a one-time hypotensive episode on hospital admission is associated with increased in-hospital mortality. The extremities are cool and pale if shock is associated with peripheral vasoconstriction, which is typical of hypovolemic, cardiogenic, and obstructive shock, but they are typically warm and pink with the peripheral vasodilation of distributive shock and dissociative shock (cyanide poisoning). Skin mottling is a physical finding that correlates with hemodynamic compromise, organ failure, and mortality.[2]

The most frequent neurologic finding in shock is alteration in the level of consciousness, ranging from confusion to coma. Many of the clinically apparent manifestations of cardiac involvement in shock result from sympathoadrenal stimulation, with tachycardia being the most sensitive indicator that shock is present. Acute lung injury, which can be immediate or delayed, results in impaired gas exchange; the work of breathing is increased, and respiratory muscle fatigue and ventilatory failure require mechanical ventilation. Hypovolemia with or without acute tubular necrosis results in oliguria, although polyuria may be seen in early shock.

Typical clinical manifestations of gut involvement during shock include abdominal pain, ileus, erosive gastritis, pancreatitis, acalculous cholecystitis, and submucosal hemorrhage. If the integrity of the gut barrier is compromised, bacteria and their toxins are translocated into the blood stream. The most common manifestation of liver involvement in shock is a mild increase in serum levels of aminotransferases and lactate dehydrogenase. With severe hypoperfusion, shock liver may be manifested with massive aminotransferase elevations and extensive hepatocellular damage.

Thrombocytopenia may result from dilution during volume repletion or from immunologic platelet destruction, which is especially common during septic shock. Activation of the coagulation cascade can lead to disseminated intravascular coagulation (Chapter 175), which results in thrombocytopenia, decreased fibrinogen, elevated fibrin split products, and microangiopathic hemolytic anemia. The finding of nucleated red blood cells on a peripheral blood smear is associated with increased in-hospital mortality.[3]

Hypovolemic Shock

Hemorrhagic shock, whether from internal or external bleeding, is the most common cause of hypovolemic shock (Table 106-2). Nonhemorrhagic hypovolemic shock can be caused by severe dehydration due to massive urinary or gastrointestinal fluid losses. Such losses are common in conditions such as diabetic ketoacidosis (Chapter 229) and diarrhea from some infectious diseases, such as cholera (Chapter 302). Massive insensible losses of water or perspiration can precipitate shock in patients with major burn injuries (Chapter 111) or heatstroke (Chapter 109). Sequestration of fluid in the extravascular compartment, commonly referred to as third spacing, can cause shock in patients as a result of surgery, bowel obstruction, hepatic failure (Chapter 154), systemic inflammation, acute pancreatitis (Chapter 144), or

TABLE 106-1 PHYSICAL EXAMINATION AND SELECTED LABORATORY SIGNS IN SHOCK

Central nervous system	Acute delirium, restlessness, disorientation, confusion, and coma, which may be secondary to decreased cerebral perfusion pressure (mean arterial pressure minus intracranial pressure). Patients with chronic hypertension or increased intracranial pressure may be symptomatic at normal blood pressures. Cheyne-Stokes respirations may be seen with severe decompensated heart failure. Blindness can be a presenting complaint or complication.
Temperature	Hyperthermia results in excess tissue respiration and greater systemic oxygen delivery requirements. Hypothermia can occur when decreased systemic oxygen delivery or impaired cellular respiration decreases heat generation.
Skin	Cool distal extremities (combined low serum bicarbonate and high arterial lactate levels) aid in identifying patients with hypoperfusion. Pallor, cyanosis, sweating, and decreased capillary refill and pale, dusky, clammy or mottled extremities indicate systemic hypoperfusion. Dry mucous membranes and decreased skin turgor indicate low vascular volume. Low toe temperature correlates with the severity of shock.
General cardiovascular	Neck vein distention (e.g., heart failure, pulmonary embolus, pericardial tamponade) or flattening (e.g., hypovolemia), tachycardia, and arrhythmias Decreased coronary perfusion pressures can lead to ischemia, decreased ventricular compliance, and increased left ventricular diastolic pressure. A "mill wheel" heart murmur may be heard with an air embolus.
Heart rate	Usually elevated. However, paradoxical bradycardia can be seen in patients with preexisting cardiac disease and severe hemorrhage. Heart rate variability is associated with poor outcomes.
Systolic blood pressure	May actually increase slightly when cardiac contractility increases in early shock and then fall as shock advances A single episode of undifferentiated hypotension with a systolic blood pressure <80 mm Hg carries an in-hospital mortality of 18%.
Diastolic blood pressure	Correlates with arteriolar vasoconstriction and may rise early in shock and then fall when cardiovascular compensation fails
Pulse pressure	Defined as systolic minus diastolic pressure and related to stroke volume and the rigidity of the aorta Increases early in shock and decreases before systolic pressure decreases
Pulsus paradoxus	An exaggerated change in systolic blood pressure with respiration (systolic blood pressure declines >10 mm Hg with inspiration) seen in asthma, cardiac tamponade, and air embolus
Mean arterial blood pressure	Diastolic blood pressure + [pulse pressure/3]
Shock index	Heart rate/systolic blood pressure. Normal = 0.5 to 0.7. A persistent elevation of the shock index (>1.0) indicates impaired left ventricular function (as a result of blood loss or cardiac depression) and is associated with increased mortality.
Respiratory	Tachypnea, increased minute ventilation, increased dead space, bronchospasm, hypocapnia with progression to respiratory failure, acute lung injury, and adult respiratory distress syndrome
Abdomen	Low-flow states may result in abdominal pain, ileus, gastrointestinal bleeding, pancreatitis, acalculous cholecystitis, mesenteric ischemia, and shock liver.
Renal	Because the kidney receives 20% of cardiac output, low cardiac output reduces the glomerular filtration rate and redistributes renal blood flow from the renal cortex toward the renal medulla, thereby leading to oliguria. Paradoxical polyuria in early sepsis may be confused with adequate hydration.
Metabolic	Respiratory alkalosis is the first acid-base abnormality, but metabolic acidosis occurs as shock progresses. Hyperglycemia, hypoglycemia, and hyperkalemia may develop.

TABLE 106-2 CLASSIFICATION OF HEMORRHAGIC SHOCK*

	CLASS I	CLASS II	CLASS III	CLASS IV
Blood loss (mL)	Up to 750	750-1500	1500-2000	>2000
% Volume	Up to 15	15-30	30-40	>40
Pulse rate (per minute)	<100	>100	>120	>140
Blood pressure	Normal	Normal	Decreased	Decreased
Pulse pressure	Normal or increased	Decreased	Decreased	Decreased
Respiratory rate (per minute)	14-20	20-30	30-40	>35
Urine output (mL/hr)	>30	20-30	5-15	Negligible
Mental status	Slightly anxious	Mildly anxious	Anxious, confused	Confused, lethargic
Fluid replacement	Crystalloid	Crystalloid	Crystalloid and blood	Crystalloid and blood

*Estimates based on a 70-kg patient. From Committee on Trauma of the American College of Surgeons. *Advanced Trauma Life Support for Doctors*. Chicago: American College of Surgeons; 1997:108.

thermal injuries (Chapter 111).[4] Regardless of whether hypovolemic shock is due to hemorrhage or fluid losses, the rate of loss is a critical component of the presentation. If volume is lost at a slow rate, compensatory mechanisms are usually effective, and any given amount of volume depletion is often better tolerated than if the same volume were lost acutely. In addition, underlying diseases, especially those that limit cardiac reserve, can substantially influence the clinical severity of a hypovolemic insult. As the importance of the microcirculation continues to be elucidated, it may become a target of future management strategies.

Distributive Shock

The most important and prevalent cause of distributive shock is septic shock (Chapter 108), but anaphylaxis (Chapter 253), drug overdose (Chapter 34), neurogenic insults, and addisonian crisis (Chapter 227) can also produce vasodilatory shock. Sepsis can be a combination of hypovolemia, vasodilation, myocardial suppression, and impaired tissue oxygen use (dissociative shock). In approximately 10 to 15% of septic shock patients, myocardial dysfunction results in a low cardiac output form of shock. Early interventions (Chapter 108) can improve outcomes substantially.

Cardiogenic Shock

Cardiogenic shock (Chapter 107) is defined by a decrease in systemic oxygen delivery caused by an acute or chronic deterioration of cardiac function due to myocardial, valvular, structural, toxic, or infectious causes. The clinical picture of cardiogenic shock is variable, depending on which structural component of the ventricle is impaired.

Extracardiac Obstructive Shock

This form of shock results from acute obstruction to flow in the circulation. Examples include impaired diastolic filling of the right ventricle (e.g., superior vena cava syndrome; Chapter 179), obstruction of right ventricular output (e.g., massive pulmonary embolism; Chapter 98), and an air embolus from cardiopulmonary bypass or central line placement (Chapter 98). Systemic arterial hypertension (Chapter 67) severe enough to impair left ventricular function or acute pericardial tamponade or constrictive pericarditis (Chapter 77) can also produce an obstructive shock pattern.

Dissociative Shock

Dissociative shock results from microvascular abnormalities, with maldistribution or shunting of blood flow, or cytopathic tissue hypoxia. Dissociative shock includes disorders that inhibit oxygen utilization, such as cyanide poisoning (Chapter 110), sodium nitroprusside use, and sepsis.

Mixed Shock States

Shock may arise from multiple causes. For example, a patient with pneumonia and a history of ischemic cardiomyopathy may present in a hypodynamic

rather than in a hyperdynamic state when combined with sepsis. Thus, a mixture of hypovolemic, distributive, cardiogenic, obstructive, and dissociative shock can potentially be seen in the same patient.

DIAGNOSIS

A key element in the approach to shock is a problem-directed history and physical examination. Some patients present with few symptoms other than generalized weakness, lethargy, or altered mental status. A discussion with the patient and family members should specifically address symptoms that suggest volume depletion, including bleeding, vomiting, diarrhea, excessive urination, insensible losses due to fever, and orthostatic lightheadedness. The history should also inquire about prior or current evidence of cardiovascular disease, especially episodes of chest pain (Chapter 51) or symptoms of heart failure (Chapter 58). Prior neurologic diseases can render patients more susceptible to complications from hypovolemia. Medication use, both prescribed and nonprescribed, must be ascertained. Some medications cause volume depletion (e.g., diuretics), whereas others depress myocardial contractility (e.g., β-blockers, calcium-channel blockers). The possibility of an anaphylactic reaction to a new medication or cardiovascular depression due to drug toxicity should be considered. A recent or remote history of steroid use may suggest adrenal insufficiency (Chapters 35 and 227).

The physical examination can provide critical information to aid in the diagnosis (see Table 106-1). Traditionally, shock is defined by a systolic blood pressure less than 90 mm Hg or 40 mm Hg less than the baseline systolic blood pressure if the patient has a history of hypertension. Ultrasound can be incorporated into a formal protocol to evaluate cardiac function, cardiac chamber filling, and certain aspects of the peripheral vasculature (Table 106-3).[5]

Acidosis

A common theme in shock is that tissue hypoxia leads to acidosis (Chapter 118), which develops as a consequence of anaerobic metabolism and generally parallels the severity of shock. Laboratory manifestations may include a base deficit, low arterial and venous pH levels, and an elevated serum lactate level. Base deficit is the absolute decrease in the serum concentration of bicarbonate (normal minus the patient's bicarbonate). A mild base deficit is −2 to −5, moderate is −6 to −14, and severe is −15 mmol/L or greater. When patients are resuscitated with large volumes of normal saline, the large fluid load can cause a dilutional acidosis, and the large chloride load can induce metabolic acidosis even in the absence of tissue hypoxia and anaerobic metabolism. Base deficit can also be caused by cocaine, alcohol (Chapter 33), and diabetic ketoacidosis (Chapter 229). Despite its limitations, base deficit provides the clinician with a quick indicator to assess the severity of tissue hypoperfusion and the adequacy of resuscitation in relieving anaerobic metabolism and oxygen debt.

The low blood pH of metabolic acidosis can result from different acids. Acidosis caused by lactate and unidentified anions produces an anion gap. The blood lactate concentration rises with increased anaerobic metabolism, as is seen in shock but also in diabetic ketoacidosis (Chapter 229), total parenteral nutrition (Chapter 217), seizures (Chapter 403), thiamine deficiency (Chapter 218), treatment of HIV infection with protease inhibitors (Chapter 388), and administration of metformin, salicylate, isoniazid, propofol, and cyanide (Chapter 110). A lactate concentration greater than 4 mmol/L is unusual in normal and non–critically ill hospitalized patients and warrants concern. A lactate concentration greater than 4 mmol/L is associated with an in-hospital mortality exceeding 25%, regardless of the cause, and failure to decrease lactate levels during the first 6 hours of shock is associated with an increased inflammatory response, the development of organ failure, and mortality. The arterial-venous difference in carbon dioxide content is inversely related to cardiac output. Whether the samples are taken from the pulmonary artery or the central venous circulation, the relationship and clinical interpretation are the same.

Urine Output

The kidneys normally receive 20% of the systemic oxygen delivery, and because of this large amount of blood flow per gram of tissue, they are highly sensitive to changes in renal blood flow. Urine output is a valuable indicator of renal perfusion and vital organ blood flow. Although a significant drop in urine output indicates reduced renal blood flow, an adequate urine output does not always indicate successful resuscitation. In fact, paradoxical polyuria may be present. Other factors that may affect urine output include the use of mannitol or diuretics. Preexisting conditions, such as renal failure, may also limit the ability of this measure to reflect the adequacy of resuscitation.

TREATMENT Rx

The goal of initial management is to restore global and microvascular perfusion to levels that sustain aerobic cellular respiration. Multiple randomized trials have shown significant and consistent reductions in mortality when shock is reversed aggressively before organ failure develops. Once this initial management is accomplished, the definitive diagnosis leads to more specific therapy based on the cause of shock.

Markers of shock serve not only as diagnostic tools for risk stratification but also as targets or end points for the early restoration of adequate tissue perfusion. Clinical monitoring of tissue oxygenation and organ function commonly involves measuring traditional end points of resuscitation, such as heart rate, blood pressure, mentation, urine output, and skin perfusion. Many clinicians continue to use these parameters as indicators that systemic oxygenation imbalances have been corrected. However, there is increasing evidence that clinical parameters may be poor indicators of the ongoing tissue hypoxia and microcirculatory dysfunction that are associated with increased mortality.

Initial Management

The initial management of shock requires immediate diagnostic and therapeutic interventions, including attention to airway, breathing, and circulation and definitive diagnosis and treatment (Chapter 63). The first step to optimize systemic oxygen delivery is to provide supplemental oxygen to increase arterial oxygen content. If any doubt exists about the patency of the airway or the adequacy of ventilation, endotracheal intubation should be performed and mechanical ventilation initiated (Chapter 105). Mechanical ventilation helps provide adequate oxygenation and carbon dioxide elimination and decreases oxygen utilization by the respiratory muscles, which may be responsible for up to 40% of systemic oxygen consumption and lactate production.

Although endotracheal intubation and mechanical ventilation may be critical for patients in shock, the sudden increase in airway pressure can lead to a series of deleterious hemodynamic complications, especially in patients who are hypovolemic or vasodilated or have compromised cardiac function. In such patients, the resulting decreased venous return, increased pulmonary

TABLE 106-3 RAPID ULTRASOUND IN SHOCK (RUSH) PROTOCOL*

RUSH EVALUATION	HYPOVOLEMIC SHOCK	CARDIOGENIC SHOCK	OBSTRUCTIVE SHOCK	DISTRIBUTIVE SHOCK
Heart	Hypercontractile LV Small LV chamber size	Hypocontractile or dilated LV	Hypercontractile LV Pericardial effusion Cardiac tamponade RV strain Cardiac thrombus	Hypercontractile LV in early sepsis, hypocontractile LV in late sepsis
Fluid status	Flat IVC Flat jugular veins Peritoneal fluid (fluid loss) Pleural fluid (fluid loss)	Distended IVC Distended jugular veins Pulmonary edema Pleural or peritoneal fluid	Distended IVC Distended jugular veins Pneumothorax	Normal or small IVC in early sepsis Peritoneal or pleural fluid
Extracardiac circulatory system	Abdominal aneurysm Aortic dissection	Normal	DVT	Normal

*Modified from Perera P, Mailhot T, Riley D, et al. The RUSH exam: Rapid Ultrasound in SHock in the evaluation of the critically ill. *Emerg Med Clin North Am.* 2010;28:29-56.
DVT = deep venous thrombosis; IVC = inferior vena cava; LV = left ventricle; RV = right ventricle.

TABLE 106-4 VASOPRESSOR AGENTS

AGENT	DOSE RANGE	PERIPHERAL VASCULATURE		CARDIAC EFFECTS			TYPICAL USE
		Vasoconstriction	*Vasodilation*	*Heart Rate*	*Contractility*	*Dysrhythmias*	
Dopamine	1-4 μg/kg/min	0	1+	1+	1+	1+	"Renal dose" does not improve renal function; may be used with bradycardia and hypotension
	5-10 μg/kg/min	1-2+	1+	2+	2+	2+	
	11-20 μg/kg/min	2-3+	1+	2+	2+	3+	Vasopressor range
Vasopressin	0.04-0.1 units/min	3-4+	0	0	0	1+	Septic shock, post–cardiopulmonary bypass shock state, no outcome benefit in sepsis
Phenylephrine	20-200 μg/min	4+	0	0	0	1+	Vasodilatory shock; best for supraventricular tachycardia
Norepinephrine	1-20 μg/min	4+	0	2+	2+	2+	First-line vasopressor for septic shock, vasodilatory shock
Epinephrine	1-20 μg/min	4+	0	4+	4+	4+	Refractory shock, shock with bradycardia, anaphylactic shock
Dobutamine	1-20 μg/kg/min	1+	2+	1-2+	3+	3+	Cardiogenic shock, septic shock
Milrinone	37.5-75 μg/kg bolus followed by 0.375-0.75 μg/min	0	2+	1+	3+	2+	Cardiogenic shock, right-sided heart failure, dilates pulmonary artery; caution in renal failure

vascular resistance, and decreased ventricular compliance may lead to hypotension and cardiovascular collapse.[6,7] Early sedation and muscle relaxation with mechanical ventilation have been shown to have outcome benefit, especially with acute lung injury. However, early use of sedatives, anxiolytics, or induction agents during and after intubation can decrease catecholamine levels, peripheral vascular resistance, and cortisol levels and may result in hypotension. If possible, preparations should be made to monitor physiologic variables, to ensure adequate fluid administration, and to provide rapid access to vasopressors should systemic arterial pressure fall to dangerously low levels.

On initial presentation, it is good practice to place one or two large-bore (≥16 gauge) peripheral intravenous catheters and to administer a crystalloid solution (normal saline or lactated Ringer solution). If MAP is less than 60 to 65 mm Hg, systolic blood pressure is less than 90 mm Hg, or evidence of tissue hypoperfusion is present, an intravenous fluid challenge (20 to 40 mL/kg crystalloid or colloid) should be given rapidly. A bolus of 500 mL every 30 minutes titrated to MAP or measurement of preload is recommended. In an 80-kg person, the average intravascular volume is 5 L. In shock states, such as septic shock, in which intravascular hypovolemia is a predominant feature, 5 to 6 L of fluid during the first 6 hours is considered an average volume resuscitation. If hemorrhage is the likely cause of shock, blood should be used to replace volume. Fluids should not be withheld, even in patients with end-stage renal disease.

Central venous access and arterial blood pressure monitoring should be established to administer vasopressors and to monitor hemodynamics and venous and arterial blood gases, respectively. Electrocardiographic monitoring and continuous measurement of oxygen saturation by pulse oximetry are useful adjuncts. Because the Trendelenburg position may impair gas exchange and promote aspiration, an alternative is to raise the patient's legs above the level of the heart with the patient in the supine position.

If the patient remains hypotensive, vasopressors such as norepinephrine, dopamine, or phenylephrine (Table 106-4) should be administered to restore adequate systemic arterial pressure while the diagnostic evaluation is ongoing. Treatment with vasopressors should not be postponed while trying to achieve euvolemia by using fluid boluses because patients with cerebrovascular and coronary artery disease may be intolerant of the hypotensive interval. However, vasopressors may also mask hypovolemia when they increase blood pressure. If the volume status remains undefined or the hemodynamic condition requires repeated fluid challenges or vasopressor treatment, a central venous catheter should be placed to determine central venous oxygen saturation, ventricular filling pressures, and intravascular volume status while echocardiography is performed. On the basis of these data, patients can usually be classified and managed according to their hemodynamic and oxygen transport patterns (Figs. 106-2 and 106-3).

Fluid Replacement

Rapid and appropriate restoration of vascular volume decreases the need for vasopressor therapy, adrenal replacement therapy, and invasive monitoring; in addition, it modulates the inflammation that arises when a patient progresses to severe shock. The goal of fluid resuscitation is not merely to achieve a predetermined volume or pressure but rather to titrate fluid to optimize systemic oxygen delivery and to meet tissue oxygen demands.[8]

To assess the adequacy of cardiac preload during the resuscitation of a patient with shock, decisions based on monitoring of CVP lead to the same outcomes as those based on measuring equivalence to wedged pulmonary artery occlusion pressure. However, the CVP does not correlate well with left ventricular end-diastolic volume; although a low CVP indicates hypovolemia, a "normal" CVP does not exclude inadequate preload as a cause of shock. A fluid challenge in a volume-responsive patient increases cardiac output by about 20% for each change of 2 cm H_2O in CVP; by comparison, cardiac output does not change if the CVP is raised in a patient with adequate left ventricular volume.

When intrathoracic pressure increases during the application of positive airway pressure in a mechanically ventilated patient, venous return decreases, and as a consequence, left ventricular stroke volume also decreases. The variation in pulse pressure or stroke volume during a positive-pressure breath can also predict the responsiveness of cardiac output to changes in preload. Pulse pressure variation is defined by Equation 3:

$$100 \times (\text{PPmax} - \text{PPmin})/[(\text{PPmax} + \text{PPmin})/2] \quad (3)$$

PPmax and PPmin are the maximal and minimal pulse pressures, respectively, in a respiratory cycle; these measurements must be made when the patient is not making any respiratory efforts on his or her own. Pulse pressure variation values of 13 to 15% suggest that hypovolemia is present and that the cardiac index will increase by at least 15% after the rapid infusion of 500 mL of crystalloid. Pulse pressure variation is a reasonable predictor of volume status and response to fluids, although atrial arrhythmias can interfere with the usefulness of this technique.

Types of Fluids

The two most commonly used crystalloid solutions are 0.9% sodium chloride solution (normal saline) and lactated Ringer solution (Table 106-5). Although these two solutions have been regarded as essentially interchangeable, accumulating data suggest that large volumes of normal saline, but not of lactated Ringer solution, promote the development of hyperchloremic metabolic acidosis and coagulopathy. Hypertonic saline is not recommended for routine use in trauma patients.

Colloids are higher-molecular-weight solutions that increase plasma oncotic pressure; they are classified as natural (albumin) or artificial (starches, hetastarch, pentastarch, dextrans, and gelatins). Colloids are dissolved in either normal saline or a balanced salt solution. Colloids stay in the intravascular space significantly longer than crystalloids do, with an intravascular half-life of 16 hours for albumin versus 30 to 60 minutes for normal saline and lactated Ringer solution. When they are titrated to the same volume status, colloids and crystalloids restore tissue perfusion to the same magnitude, but a two to four times greater volume of crystalloids is required to achieve the same end point. The outcomes of patients with hypovolemic shock are equivalent for crystalloids and albumin (see Table 106-5). By comparison, hetastarch increases renal failure and mortality in intensive care unit (ICU) patients with shock.

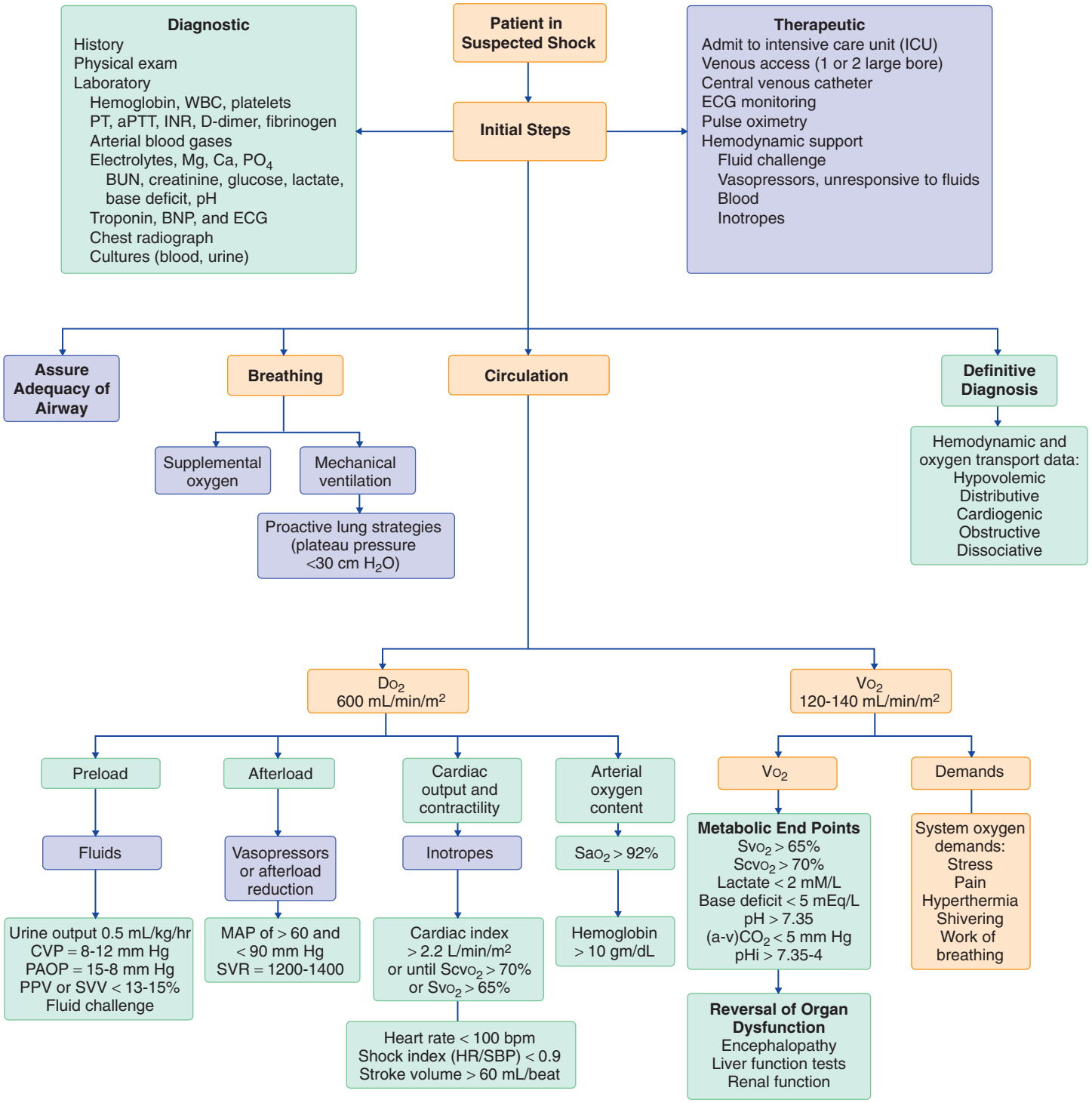

FIGURE 106-2. General hemodynamic management. aPTT = activated partial thromboplastin time; BNP = brain natriuretic peptide; bpm = beats per minute; BUN, blood urea nitrogen; CVP = central venous pressure; Do_2 = (systemic) oxygen delivery; ECG = electrocardiogram; HR = heart rate; INR = international normalized ratio; MAP = mean arterial pressure; PAOP = pulmonary artery occlusion pressure; pHi = intestinal mucosal pH; PPV = pulse pressure variation; PT = prothrombin time; SBP = systolic blood pressure; SVR = systemic vascular resistance; SVV = stroke volume variation; Vo_2 = (systemic) oxygen consumption; WBC = white blood cell.

Fluid Management Strategies

After initial aggressive fluid resuscitation within 6 hours of presentation, controversy exists regarding fluid management strategies in the next 2 days or so, especially in patients with ARDS. Beginning an average of about 48 hours after admission to the ICU and 24 hours after the establishment of ARDS, conservative fluid therapy to maintain euvolemia provides significantly better lung and central nervous system function as well as a decreased need for sedation, mechanical ventilation, and ICU care compared with more aggressive fluid therapy.[8] However, patients who receive conservative volume replacement transiently may need more vasopressor support. One of the negative attributes of an open-ended and oftentimes aggressive fluid resuscitation strategy is that patients may develop an intra-abdominal compartment syndrome in which elevated abdominal pressures impair gas exchange, decrease renal perfusion, decrease visceral organ perfusion, impair venous return, and thereby decrease cardiac output and systemic oxygen delivery.[9]

Hemoglobin

In hemorrhagic shock, the rapid administration of packed red blood cells and, if indicated, platelets and thawed fresh-frozen plasma can be life-saving. Whenever possible, fully crossmatched packed red blood cells are preferable, but type-specific blood can often be given safely when immediate therapy is warranted (Chapter 177). In dire emergencies, type O Rh-negative blood can be administered to women of childbearing potential, and type O Rh-negative or Rh-positive blood can be given to men or postmenopausal women.

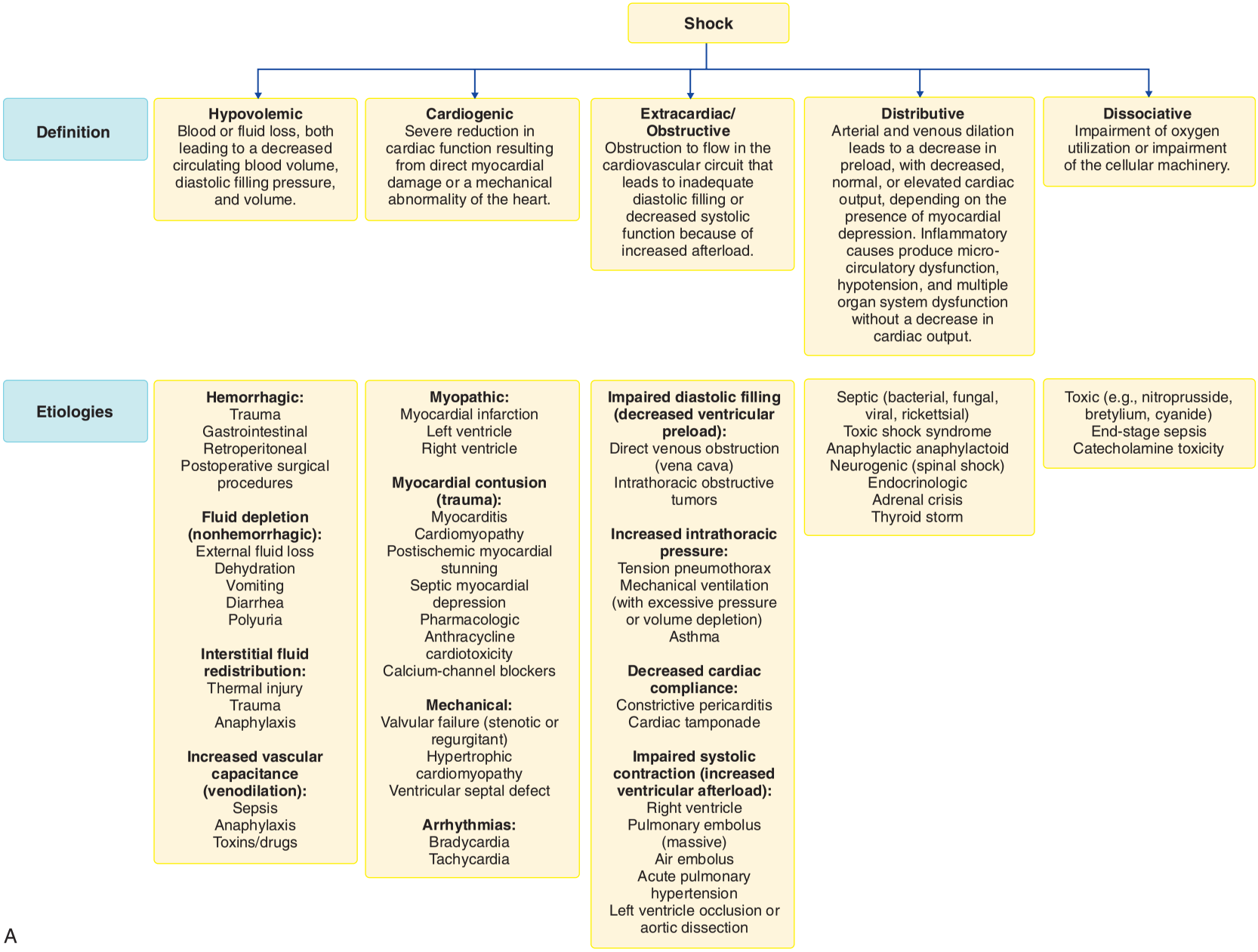

A

FIGURE 106-3. Definitions, etiologies, and therapies of various shock states. CI = cardiac index; CO = cardiac output; CT = computed tomography; CVP = central venous pressure; Do_2 = systemic oxygen delivery; ECG = electrocardiogram; LV = left ventricular; MAP = mean arterial pressure; MRI = magnetic resonance imaging; PA = pulmonary artery; PAOP = pulmonary artery occlusion pressure; RA = right atrial; RV = right ventricular; SVR = systemic vascular resistance; US = ultrasonography; VSD = ventricular septal defect.

Hemodynamic Characterization	CVP/RA ↓ PAOP ↓ $CO/CI/Do_2$ ↓ SVR ↑ $Svo_2/Scvo_2$ ↓ Comments: Preexisting heart disease may alter the filling pressures	CVP/RA ↑ PAOP ↓ early, ↑ later $CO/CI/Do_2$ ↓↓ SVR ↑ $Svo_2/Scvo_2$ ↓ Comments: If there is a VSD, there may be $Scvo_2/Svo_2$ or a step up from the right atrium to right ventricle to the pulmonary artery. Large V wave is seen on PA catheter with mitral regurgitation.	CVP/RA ↑↑ PAOP ↓ $CO/CI/Do_2$ ↑↓ SVR ↑ $Svo_2/Scvo_2$ ↓ Comments: Equalization of intracardiac diastolic pressure is consistent with pericardial tamponade. A rapid x and blunted y descent on CVP wave form indicates the inability of the heart to fill.	CVP/RA ↓ early, normal later PAOP ↓ early, normal later $CO/CI/Do_2$ ↓ early, ↑ later SVR ↑ early, ↓ later $Svo_2/Scvo_2$ ↓ early, ↑ later Comments: In the early stage, filling pressures are low, leading to low to normal cardiac output and variable $Scvo_2/Svo_2$. If myocardial depression accompanies, then a similar profile of cardiogenic shock is seen. After resuscitation a hyperdynamic circulation is seen.	CVP/RA ↓ early, normal to ↑ later PAOP ↓ early, normal to ↑ later $CO/CI/Do_2$ ↓ early, ↑ later SVR ↑ early, ↓ later $Svo_2/Scvo_2$ ↑ early, ↑↑ later Comments: Filling pressures can be variable, depending on stage of the disease. Cardiac output is generally increased. SVR is decreased and $Scvo_2/Svo_2$ and lactate are increased.
Therapy	Rapid replacement of blood, colloid or crystalloid Identify source of blood/fluid loss Replace deficient coagulation factors Consider factor VIIa Endoscopy/colonoscopy Angiography CT/MRI/US or other	LV infarction: intra-aortic balloon pump (IABP) coronary angiography Recanalization: thrombolytic therapy, angioplasty, coronary bypass surgery RV infarction: fluids and inotropes with PA catheter Mechanical abnormality: echocardiography, cardiac catheterization Corrective surgery: mechanical assist devices, transplantation	Pericardial tamponade: pericardiocentesis, surgical drainage (if needed) Pneumothorax: chest tube Pulmonary embolism: heparin ventilation-perfusion lung scan pulmonary angiography consider thrombolytic therapy, embolectomy Severe hypertension: afterload reduction	Identify site of infection Antimicrobial agents Early goal-directed therapy fluids vasopressors inotropic agents Goals: $Sco_2 > 70\%$ CVP > 8 mm Hg MAP > 65 mm Hg Improving organ function Decreasing lactate levels Consider activated protein C Requiring vasopressors, consider corticosteroids	Toxin antidote or remove agent, hyperbaric oxygen therapy Recombinant activated protein C Nitroglycerin (speculative) Prostacyclin (speculative)

B

FIGURE 106-3, cont'd.

TABLE 106-5 FLUID THERAPY

Normal saline	Normal saline is a slightly hyperosmolar solution containing 154 mEq/L of both sodium and chloride. Because of the relatively high chloride concentration and low pH, normal saline carries a risk of inducing hyperchloremic metabolic acidosis when it is given in large amounts.
Lactated Ringer solution (LR)	Lactate is metabolized to carbon dioxide (CO_2) and water by the liver, leading to the release of CO_2 in the lungs and excretion of water by the kidneys. LR is preferred to normal saline and buffers acidemia. Because LR contains a very small amount of potassium, there is a small risk of inducing hyperkalemia in patients with renal insufficiency or renal failure. LR may be incompletely metabolized in severe hepatic failure.
Albumin	Albumin is a protein derived from human plasma and is available in varying concentrations from 4 to 25%. A study comparing fluid resuscitation with albumin versus saline found similar 28-day mortalities and secondary outcomes in each arm. However, a post hoc subset analysis of patients with sepsis and acute lung injury resuscitated with albumin showed a trend toward a decrease in mortality. There was a significant increase in mortality in trauma patients, particularly those with head injury.
Hydroxyethyl starch (HES)	HES, which is a synthetic colloid derived from hydrolyzed amylopectin, causes renal impairment at recommended doses and impaired long-term survival at high doses. HES can also cause coagulopathy and bleeding complications from reduced factor VIII and von Willebrand factor levels as well as impaired platelet function. HES increases the risk of acute renal failure and reduces the probability of survival in patients with sepsis.
Dextrans	Dextrans are artificial colloids synthesized by *Leuconostoc mesenteroides* bacteria grown in sucrose media. Dextrans are used more frequently to lower blood viscosity than for rapid plasma expansion. They can cause renal dysfunction as well as anaphylactoid reactions.
Gelatins	Gelatins are produced from bovine collagen. Because they have a small molecular weight, they are not very effective at expanding plasma volume, but they cost less than other options. They have been reported to cause renal impairment as well as allergic reactions ranging from pruritus to anaphylaxis. Gelatins are not currently available in North America.

The appropriate hemoglobin level in shock remains controversial, but a transfusion threshold value of ≥7 g/dL is as good as a transfusion threshold of 9 g/dL in septic shock.[A] A hemoglobin value of 7 to 10 g/dL is appropriate when the patient is in the acute but stable phase of upper gastrointestinal bleeding.[A]

Vasopressor Therapy

To optimize end-organ perfusion, the second phase of intervention after adequate fluid therapy is to maintain perfusion pressure. A specific MAP goal has not been established for all shock states, but a MAP of at least 60 to 65 mm Hg is a reasonable target.

The most common vasopressors are agonists at various adrenergic receptors. Receptors include peripheral α-adrenergic receptors that lead to vasoconstriction; cardiac β_1 receptors with both chronotropic and inotropic effects; β_2 receptors located in the circulation and airways that mediate vasodilation and bronchodilation; and dopaminergic receptors located throughout the cardiovascular, mesenteric, and renal circulations. On the basis of these mechanisms, therapy can be tailored to a specific circumstance. For example, a patient with severe tachycardia would be best served by an agent with more α-selective activity and less β activity to avoid tachycardia and increased myocardial oxygen consumption (see Table 106-4).

Norepinephrine, which is a vasoconstrictor and an inotrope, provides better splanchnic oxygen utilization compared with dopamine. It is generally considered the first-line vasopressor for treating persistent hypotension in septic patients despite adequate resuscitation, and it may be superior to dopamine in treating cardiogenic shock. For example, a randomized trial comparing dopamine with norepinephrine in patients with shock showed no significant difference in mortality for hypovolemic and septic shock but a significant benefit of norepinephrine in cardiogenic shock.[A] In addition, there was a significant two-fold increase in arrhythmic events with dopamine compared with norepinephrine (24.1 vs. 12.4%).

Dopamine's effects result from transduction at dopaminergic receptors in the renal, mesenteric, coronary, and systemic circulations. The positive chronotropic and inotropic effects of dopamine can lead to tachycardia and tachyarrhythmias; this effect frequently limits its dosing because the increased myocardial oxygen requirements promote the development of myocardial ischemia, especially in the presence of coronary artery disease.

Phenylephrine is a synthetic catecholamine that is a selective α-adrenergic agonist and is ideal in patients with tachycardia. However, the resulting increase in myocardial oxygen consumption, decrease in splanchnic blood flow, and decrease in cardiac output can be detrimental for patients with septic shock.

Epinephrine, which is a potent α-, β_1-, and β_2-adrenergic agonist, increases peripheral arteriolar tone as well as cardiac contractility. It is the first-line agent for the treatment of anaphylactic shock and is used to support myocardial contractility after cardiac surgery. Epinephrine increases the white blood cell count and the blood lactate concentration because of accelerated aerobic glycogenolysis or maldistribution of blood flow.

Vasopressin deficiency accompanies vasodilatory shock, and administration of low doses of vasopressin (0.03–0.04 units/minute) increases arterial blood pressure in septic patients with intractable hypotension. In patients with septic shock, the addition of low-dose vasopressin to norepinephrine does not have any overall benefit,[A] but it may benefit patients who have less severe forms of shock and who also receive glucocorticoids.[10]

Adrenal Dysfunction

Beyond their metabolic functions, glucocorticoids are required to maintain responsiveness to vasopressors, intravascular volume, vascular permeability, and myocardial contractility. If the hypothalamic-pituitary-adrenocortical axis is depressed in shock, clinical findings can include unexplained fever, hypoglycemia, hyponatremia, hyperkalemia, metabolic acidosis, hypotension refractory to fluid resuscitation, and eosinophilia. Cortisol levels and the results of cosyntropin stimulation testing may not be clinically helpful.[11] If adrenal insufficiency is strongly suspected, or if patients have refractory hypotension despite vasopressors and hemodynamic optimization, stress doses of intravenous hydrocortisone (e.g., 50 mg every 6 hours) are recommended.[A]

Mechanical Support

Mechanical hemodynamic support with an intra-aortic balloon may be indicated in cardiogenic shock, but it does not improve survival in patients with acute myocardial infarction.[A] Extracorporeal membrane oxygenation is another temporary option for cardiopulmonary support in patients with ARDS until more definitive long-term interventions can be performed.[12]

PROGNOSIS

Clinical characteristics associated with a poor outcome include the severity of shock; its temporal duration, underlying cause, and reversibility; and preexisting vital organ dysfunction. Decreased systemic oxygen consumption, persistently elevated lactate levels, size of the base deficit, and severity of the anion gap are associated with increased organ failure and are prognostic in trauma, in septic shock, and after cardiac arrest. Regional measurements of pH are highly predictive of outcome; for example, if the gastric mucosal pH remains below 7.3 for 24 hours, the hospital mortality rate is about 50%. Although many of these poor prognostic signs are suggestive of microcirculatory failure, no targeted therapies yet exist to reverse this disorder. The mortality for an undiagnosed patient who is sent to a general medical ward and develops shock is three times higher than for a patient who is admitted directly to the ICU.

Grade A References

A1. Papazian L, Forel JM, Gacouin A, et al. Neuromuscular blockers in early acute respiratory distress syndrome. *N Engl J Med*. 2010;363:1107-1116.

A2. Wheeler AP, Bernard GR, Thompson BT, et al. Pulmonary-artery versus central venous catheter to guide treatment of acute lung injury. *N Engl J Med*. 2006;354:2213-2224.

A3. Wang JW, Li JP, Song YL, et al. Hypertonic saline in the traumatic hypovolemic shock: meta-analysis. *J Surg Res*. 2014;191:448-454.

A4. Rochwerg B, Alhazzani W, Sindi A, et al. Fluid resuscitation in sepsis: a systematic review and network meta-analysis. *Ann Intern Med*. 2014;161:347-355.

A5. Annane D, Siami S, Jaber S, et al. Effects of fluid resuscitation with colloids vs crystalloids on mortality in critically ill patients presenting with hypovolemic shock: the CRISTAL randomized trial. *JAMA*. 2013;310:1809-1817.

A6. Zarychanski R, Abou-Setta AM, Turgeon AF, et al. Association of hydroxyethyl starch administration with mortality and acute kidney injury in critically ill patients requiring volume resuscitation: a systematic review and meta-analysis. *JAMA*. 2013;309:678-688.
A7. Wiedemann HP, Wheeler AP, Bernard GR, et al. Comparison of two fluid-management strategies in acute lung injury. *N Engl J Med*. 2006;354:2564-2575.
A8. Holst LB, Haase N, Wetterslev J, et al. Lower versus higher hemoglobin threshold for transfusion in septic shock. *N Engl J Med*. 2014;371:1381-1391.
A9. Villanueva C, Colomo A, Bosch A, et al. Transfusion strategies for acute upper gastrointestinal bleeding. *N Engl J Med*. 2013;368:11-21.
A10. De Backer D, Biston P, Devriendt J, et al. Comparison of dopamine and norepinephrine in the treatment of shock. *N Engl J Med*. 2010;362:779-789.
A11. Russell JA, Walley KR, Singer J, et al. Vasopressin versus norepinephrine infusion in patients with septic shock. *N Engl J Med*. 2008;358:877-887.
A12. Sprung CL, Annane D, Keh D, et al. Hydrocortisone therapy for patients with septic shock. *N Engl J Med*. 2008;358:111-124.
A13. Thiele H, Zeymer U, Neumann FJ, et al. Intra-aortic balloon counterpulsation in acute myocardial infarction complicated by cardiogenic shock (IABP-SHOCK II): final 12 month results of a randomised, open-label trial. *Lancet*. 2013;382:1638-1645.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Vincent JL, De Backer D. Circulatory shock. *N Engl J Med*. 2013;369:1726-1734.
2. Ait-Oufella H, Lemoinne S, Boelle PY, et al. Mottling score predicts survival in septic shock. *Intensive Care Med*. 2011;37:801-807.
3. Desai S, Jones SL, Turner KL, et al. Nucleated red blood cells are associated with a higher mortality rate in patients with surgical sepsis. *Surg Infect (Larchmt)*. 2012;13:360-365.
4. Druey KM, Greipp PR. Narrative review: the systemic capillary leak syndrome. *Ann Intern Med*. 2010;153:90-98.
5. Kanji HD, McCallum J, Sirounis D, et al. Limited echocardiography-guided therapy in subacute shock is associated with change in management and improved outcomes. *J Crit Care*. 2014;29:700-705.
6. Funk DJ, Jacobsohn E, Kumar A. The role of venous return in critical illness and shock—part I: physiology. *Crit Care Med*. 2013;41:255-262.
7. Funk DJ, Jacobsohn E, Kumar A. Role of the venous return in critical illness and shock: part II—shock and mechanical ventilation. *Crit Care Med*. 2013;41:573-579.
8. Myburgh JA, Mythen MG. Resuscitation fluids. *N Engl J Med*. 2013;369:1243-1251.
9. Kirkpatrick AW, Roberts DJ, De Waele J, et al. Intra-abdominal hypertension and the abdominal compartment syndrome: updated consensus definitions and clinical practice guidelines from the World Society of the Abdominal Compartment Syndrome. *Intensive Care Med*. 2013;39:1190-1206.
10. Russell JA, Walley KR, Gordon AC, et al. Interaction of vasopressin infusion, corticosteroid treatment, and mortality of septic shock. *Crit Care Med*. 2009;37:811-818.
11. Boonen E, Vervenne H, Meersseman P, et al. Reduced cortisol metabolism during critical illness. *N Engl J Med*. 2013;368:1477-1488.
12. Abrams D, Combes A, Brodie D. Extracorporeal membrane oxygenation in cardiopulmonary disease in adults. *J Am Coll Cardiol*. 2014;63:2769-2778.

REVIEW QUESTIONS

1. Hydroxyethyl starch fluid resuscitation is associated with

A. Less renal failure
B. Better outcomes
C. Higher mortality
D. Lowers costs of fluid replacement
E. Lowered plasma oncotic pressure

Answer: C Hydroxyethyl starch has increasingly been shown to increase morbidity and mortality in shock.

2. Which vasoactive agent is the least likely to increase heart rate?

A. Dopamine
B. Epinephrine
C. Phenylephrine
D. Dobutamine
E. Nitroprusside

Answer: C Phenylephrine is a pure α agonist and does not have β-adrenergic activity that precipitates tachycardia.

3. A decreased central venous or mixed venous oxygen saturation is seen in all except which of the following?

A. Cardiogenic shock
B. Severe cyanide poisoning
C. Hemorrhagic shock
D. During cardiopulmonary resuscitation
E. Saddle pulmonary embolus

Answer: B Cyanide poisoning is the correct answer. This type of poisoning uncouples oxidative metabolism, so less oxygen is extracted from blood, thereby leading to an increased mixed venous oxygen content.

4. Obstructive shock can result from each of the following except which one?

A. Pericardial tamponade
B. Tension pneumothorax
C. Pulmonary embolus
D. Increased arterial vascular resistance
E. Ventricular septal defect

Answer: E A ventricular septal defect does not obstruct the flow of blood, whereas all other choices represent physiologic conditions in which obstruction can occur.

5. Shock can be accompanied by

A. Hypovolemia
B. Myocardial suppression
C. Vasodilation
D. Normal or increased lactate level
E. All of the above

Answer: E All of the noted items are associated with shock.

TABLE 107-1 DIAGNOSIS OF CARDIOGENIC SHOCK

CLINICAL SIGNS
Hypotension
Oliguria
Clouded sensorium
Cool and mottled extremities
HEMODYNAMIC CRITERIA
Systolic blood pressure < 90 mm Hg or > 30 mm Hg decrease from baseline for > 30 minutes
Cardiac index < 2.2 L/min/m^2
Pulmonary capillary wedge pressure > 18 mm Hg
OTHER
Documented myocardial dysfunction
Exclusion of hypovolemia, hypoxia, and acidosis

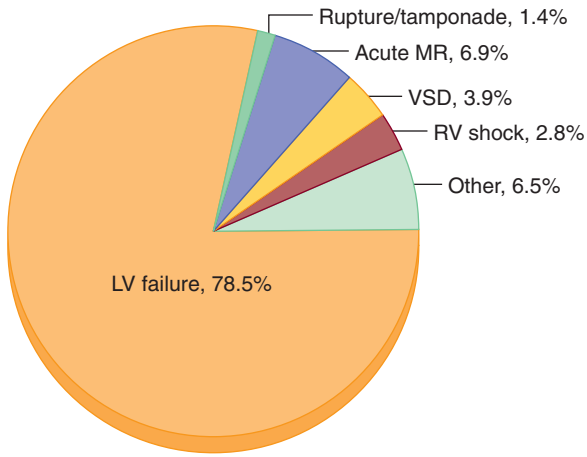

FIGURE 107-1. Causes of cardiogenic shock in patients with myocardial infarction in the SHOCK trial registry. LV = left ventricular; MR = mitral regurgitation; RV = right ventricular; VSD = ventricular septal defect. (Modified from Hochman JS, Boland J, Sleeper LA, et al. Current spectrum of cardiogenic shock and effect of early revascularization on mortality. Results of an International Registry. SHOCK Registry Investigators. *Circulation*. 1995;91:873-881).

107 CARDIOGENIC SHOCK

STEVEN M. HOLLENBERG

DEFINITION

Cardiogenic shock is the syndrome that ensues when the heart is unable to deliver enough blood to maintain adequate tissue perfusion.[1] The hemodynamic picture includes sustained systemic hypotension, pulmonary capillary wedge pressure (PCWP) greater than 18 mm Hg, and cardiac index less than 2.2 L/minute/m^2 (Table 107-1). Although systolic blood pressure less than 90 mm Hg is a commonly accepted threshold for shock, a decrease of 30 mm Hg from baseline is also used. The diagnosis of cardiogenic shock is often made on clinical grounds—hypotension combined with signs of poor tissue perfusion, including oliguria, clouded sensorium, and cool extremities, all in the setting of myocardial dysfunction. To make the diagnosis, it is important to document myocardial dysfunction and to exclude or to correct factors such as hypovolemia, hypoxemia, and acidosis.

EPIDEMIOLOGY

The predominant cause of cardiogenic shock (Fig. 107-1) is left ventricular failure secondary to acute myocardial infarction (MI)—an extensive first acute MI, the cumulative loss of myocardial function in a patient with previous MI or cardiomyopathy, or a mechanical complication of MI (Chapter 73). However, any cause of severe left ventricular (LV) or right ventricular (RV) dysfunction can lead to cardiogenic shock, including end-stage cardiomyopathy (Chapter 60), prolonged cardiopulmonary bypass, valvular disease (Chapter 75), myocardial contusion (Chapter 111), sepsis with unusually profound myocardial depression (Chapter 108), and fulminant myocarditis (Chapter 60) (Table 107-2). Stress-induced (takotsubo) cardiomyopathy (Chapter 60), a syndrome of acute apical LV dysfunction that occurs after emotional distress, may also be manifested with cardiogenic shock. Acute valvular regurgitation, most often caused by endocarditis (Chapter 76) or chordal rupture (Chapter 75), can lead to shock, as can physiologic stress in the setting of severe valvular stenosis. Cardiac tamponade (Chapter 77) and massive pulmonary embolism (Chapter 98) with acute RV failure can cause shock without pulmonary edema.

The incidence and mortality associated with cardiogenic shock appear to be declining.[2] In the past 30 years, the incidence has fallen from about 8% to 6% of MIs, largely because of the benefit of early perfusion strategies (Chapter 73). In parallel, mortality from cardiogenic shock has decreased from 70 to 80% to 50% or less, suggesting that increasingly effective early treatment and more widespread adoption of early revascularization have improved the outcomes of patients in whom shock has already developed.

Risk factors for the development of cardiogenic shock in MI parallel those for LV dysfunction and the severity of coronary artery disease (CAD). Characteristics of patients include older age, anterior MI, diabetes, hypertension, multivessel CAD, previous MI, and peripheral vascular and cerebrovascular disease. Clinical risk factors include decreased ejection fractions, larger infarctions, and lack of compensatory hyperkinesis in myocardial territories remote from the infarction. Clinical harbingers of impending shock include the degree of hypotension and tachycardia at hospital presentation. The factors that predict mortality reflect the severity of the acute insult as well as comorbid conditions.

Coronary angiography most often demonstrates multivessel CAD. About 30% of patients have a left main coronary artery occlusion, about 60% have three-vessel coronary disease, and only about 20% have single-vessel disease. Multivessel CAD helps explain the failure to develop compensatory hyperkinesis in remote myocardial segments because of either previous infarction or high-grade coronary stenoses.

Only one fourth of patients who develop cardiogenic shock are in shock when they initially present to the hospital; in the others, shock usually evolves during several hours, suggesting that early treatment may prevent shock. Comparison of the clinical characteristics of patients with early and late shock shows similar demographic, historical, clinical, and hemodynamic characteristics, but shock tends to develop earlier in patients with single-vessel CAD than in those with triple-vessel disease. This finding suggests that early shock in the setting of acute MI may be more amenable to revascularization of the culprit vessel by thrombolysis or angioplasty (Chapter 73), whereas shock developing later may require more complete revascularization with multivessel percutaneous coronary intervention (PCI) or coronary artery bypass graft (CABG) surgery (Chapter 74).

TABLE 107-2 CAUSES OF CARDIOGENIC SHOCK

ACUTE MYOCARDIAL INFARCTION
Pump failure
Large infarction
Smaller infarction with preexisting left ventricular dysfunction
Infarct extension
Reinfarction
Infarct expansion
Mechanical complications
Acute mitral regurgitation due to papillary muscle rupture
Ventricular septal defect
Free wall rupture
Pericardial tamponade
Right ventricular infarction
CARDIOMYOPATHY
Myocarditis
Peripartum cardiomyopathy
End-stage low-output heart failure
Hypertrophic cardiomyopathy with outflow tract obstruction
Stress cardiomyopathy
VALVULAR HEART DISEASE
Acute mitral regurgitation (chordal rupture)
Acute aortic regurgitation
Aortic or mitral stenosis with tachyarrhythmia or other comorbid condition causing decompensation
Prosthetic valve dysfunction
TACHYARRHYTHMIA
OTHER CONDITIONS
Prolonged cardiopulmonary bypass
Septic shock with severe myocardial depression
Penetrating or blunt cardiac trauma
Orthotopic transplant rejection
Massive pulmonary embolism
Pericardial tamponade

TABLE 107-3 CLINICAL SIGNS OF VOLUME STATUS AND PERFUSION

SIGNS AND SYMPTOMS OF CONGESTION
Orthopnea, paroxysmal nocturnal dyspnea
Jugular venous distention
Abdominojugular reflux
Rales
Hepatomegaly
Edema
Right upper quadrant tenderness
POSSIBLE EVIDENCE OF LOW PERFUSION
Narrow pulse pressure
Obtundation
Cool extremities
Cachexia, muscle loss
Decreased exercise tolerance
Renal/hepatic dysfunction
Hypotension with angiotensin-converting enzyme inhibition

PATHOBIOLOGY

Cardiogenic shock is characterized by a downward cascade in which myocardial dysfunction reduces stroke volume, cardiac output, and blood pressure; these changes compromise myocardial perfusion, exacerbate ischemia, and further depress myocardial function, cardiac output, and systemic perfusion. Concomitant diastolic dysfunction increases left atrial pressure, which leads to pulmonary congestion and hypoxemia that can exacerbate myocardial ischemia and impair ventricular performance.

Compensatory mechanisms include sympathetic stimulation, which increases heart rate and contractility, and renal fluid retention, which increases preload. Increases in heart rate and contractility raise output but also increase myocardial oxygen demand. Another compensatory mechanism, vasoconstriction to maintain blood pressure, increases myocardial afterload, further impairing cardiac performance and increasing myocardial oxygen demand. In the face of inadequate perfusion, this increased demand can worsen ischemia and perpetuate a vicious circle that, if unbroken, may culminate in death. Interruption of this circle of myocardial dysfunction and ischemia is the basis for therapeutic regimens for cardiogenic shock.

In cardiogenic shock, LV dysfunction is not always severe. In one large study, the mean LV ejection fraction was 30%, indicating that mechanisms other than primary pump failure were operative. Furthermore, systemic vascular resistance is not always elevated, suggesting that compensatory vasoconstriction is not universal. Inflammatory responses may contribute to the vasodilation and myocardial dysfunction in cardiogenic shock.

Patients in cardiogenic shock may have areas of nonfunctional but viable myocardium due to stunning or hibernation. Myocardial stunning represents postischemic dysfunction that persists despite restoration of normal blood flow. Hibernating myocardial segments have persistently impaired function at rest because of severely reduced coronary blood flow. Although hibernation is conceptually different from stunning, the two conditions may not differ much clinically. Repetitive episodes of myocardial stunning can occur in areas of viable myocardium subtended by a critical coronary stenosis. Such episodes can recapitulate the hibernation phenotype, blurring the distinction between myocardial stunning and hibernation. Regardless of the degree of overlap, their therapeutic implications differ in cardiogenic shock. The contractile function of hibernating myocardium improves with revascularization, whereas stunned myocardium retains inotropic reserve and can respond to inotropic stimulation. In addition, the severity of the antecedent ischemic insult determines the intensity of stunning, providing a rationale for reestablishing the patency of occluded coronary arteries in patients with cardiogenic shock. Finally, the notion that some myocardial tissue may recover function emphasizes the importance of measures to support the patient hemodynamically and to minimize myocardial necrosis in patients with shock.

CLINICAL MANIFESTATIONS

The physical examination should be geared toward evaluating congestion and systemic perfusion to characterize the patient's hemodynamic profile (Table 107-3). An assessment of whether the patient is "wet" or "dry" and "cold" or "warm" is integral to management. Signs of congestion (Chapter 58) include jugular venous distention (see Fig. 51-3) and pulmonary rales and may include peripheral edema and ascites. Whether the patient is cold or warm is an indication of systemic perfusion.

The majority of the cardiogenic shock patients present wet and cold. Patients with shock are usually ashen or cyanotic, and they have cool skin and mottled extremities. Cerebral hypoperfusion may cloud the sensorium. Pulses, which are rapid and faint, may be irregular in the presence of arrhythmias. Jugular venous distention and pulmonary rales are usually present, although their absence does not exclude the diagnosis. A precordial heave resulting from LV dyskinesis may be palpable. The heart sounds may be distant, and third and fourth heart sounds are usually present. A systolic murmur of mitral regurgitation or a ventricular septal defect may be heard, but either complication can occur without an audible murmur (Chapter 73).

DIAGNOSIS

After recognizing the clinical manifestations of apparent cardiogenic shock, the clinician must confirm its presence and assess its cause while simultaneously initiating supportive therapy before irreversible damage to vital organs ensues. The clinician must balance overzealous pursuit of an etiologic diagnosis before achieving stabilization with overzealous empirical treatment without establishing the underlying pathophysiologic process.

An electrocardiogram (ECG) should be performed immediately. In cardiogenic shock caused by acute MI, the ECG most commonly shows ST elevation, but ST depression or nonspecific changes are found in 25% of cases. If RV infarction is suspected, ST elevation in modified right-sided leads may be diagnostic (Chapter 73). The ECG may also provide information on previous MIs and rhythm abnormalities. A relatively normal ECG or one showing only diffuse, nonspecific changes in a patient with clinical cardiogenic shock should suggest myocarditis (Chapter 60), especially if the patient has arrhythmias. In end-stage heart failure, the ECG may show Q waves or bundle branch block, indicative of extensive disease.

Other initial diagnostic tests include a chest radiograph, complete blood count, and measurement of arterial blood gases, electrolytes, and cardiac biomarkers. A high-quality chest film can assess signs of pulmonary edema and is helpful when signs suggest an alternative diagnosis, such as a widened mediastinum indicative of aortic dissection (Chapter 78).

Echocardiography

Echocardiography should be performed as early as possible, preferably with color flow Doppler, to provide an expeditious assessment of cardiac chamber size, LV and RV function, valvular structure and motion, atrial size, and the pericardium (Chapter 55). Echocardiography can also assess or diagnose overall and regional systolic function, diastolic function, papillary muscle rupture, acute ventricular septal defect, free wall rupture, degree of mitral regurgitation, presence of RV infarction, cardiac tamponade, and valvular stenosis.

Right-Sided Heart Catheterization

If the history, physical examination, chest radiograph, and echocardiogram demonstrate systemic hypoperfusion, low cardiac output, and elevation of venous pressures, right-sided heart catheterization may not be necessary for diagnosis. However, therapy with vasopressors and inotropic agents is best optimized with hemodynamic measurements. Right-sided heart catheterization can exclude other causes of shock, such as volume depletion and sepsis. A step-up in oxygen saturation between the right atrium and pulmonary artery can indicate a ventricular septal defect (Chapter 69), and large *v* waves in the PCWP waveform can reflect acute severe mitral regurgitation. RV infarction should be suspected when the PCWP is normal but right-sided filling pressures are notably elevated.

Right-sided heart catheterization is most useful, however, to optimize therapy in unstable patients. In such patients, clinical estimates of filling pressures can be unreliable, and changes in myocardial performance or therapeutic interventions can change cardiac output and filling pressures precipitously. Although patients with a low cardiac index (<2.2 L/minute/m^2) and a PCWP greater than 18 mm Hg meet the definition of cardiogenic shock, optimal filling pressures may be even higher in individual patients with LV diastolic dysfunction.

TREATMENT

Initial Management

Initial stabilization of the patient with suspected cardiogenic shock consists of venous access, supplemental oxygen, and continuous ECG monitoring (Fig. 107-2). Many patients require endotracheal intubation and mechanical ventilation (Chapter 105), if only to reduce the work of breathing and to facilitate sedation and stabilization before cardiac catheterization. Electrolyte abnormalities should be corrected. Morphine (1 to 2 mg every 5 minutes) relieves pain and anxiety, reduces excessive sympathetic activity, and decreases oxygen demand, preload, and afterload. Atrial bradyarrhythmias or tachyarrhythmias (Chapter 64) or ventricular tachyarrhythmias can reduce cardiac output and should be corrected promptly with antiarrhythmic drugs (see Table 64-6), cardioversion, or pacing (Chapter 66).

If the cause is likely to be an acute MI, aspirin and heparin should be given immediately (Chapter 73). Some therapies routinely used in acute MI (e.g., nitrates, β-blockers, angiotensin-converting enzyme inhibitors) have the potential to exacerbate hypotension in cardiogenic shock and have recently been associated with poorer outcomes in hypotensive patients.[3] As a result, they should be avoided in patients with a tenuous hemodynamic status until they stabilize.

An initial assessment of fluid status and systemic perfusion should be performed. Patients are commonly diaphoretic, and relative hypovolemia may be present. Ischemia produces diastolic dysfunction, so high filling pressures may be necessary to maintain stroke volume in some patients. Some patients may benefit from judicious fluid replacement with predetermined rapid bolus infusions of 100 to 200 mL of normal saline titrated to clinical end points. Patients who do not respond rapidly to initial treatment should be considered for invasive hemodynamic monitoring to identify the filling pressure at which cardiac output is maximized. Maintenance of adequate preload is particularly important in patients with RV infarction.

After initial stabilization and restoration of adequate blood pressure, tissue perfusion should be assessed. If tissue perfusion is adequate but significant pulmonary congestion remains, low-dose diuretics may be used, with care taken not to remove too much fluid. If tissue perfusion remains inadequate, inotropic support or mechanical support should be initiated.

Vasopressors and Inotropes

Maintenance of adequate blood pressure is essential to break the vicious circle of progressive hypotension and further myocardial ischemia. When arterial pressure remains inadequate, therapy with vasopressor agents, titrated not only to blood pressure but also to clinical indices of perfusion and mixed venous oxygen saturation, may be required.[3] Norepinephrine and dopamine are considered first-line drugs for hypotension in this situation. Dopamine acts as both an inotrope (particularly at 3 to 10 μg/kg/minute) and a vasopressor (10 to 20 μg/kg/minute). Norepinephrine (0.02 to 1.0 μg/kg/minute) acts primarily as a vasoconstrictor, has a mild inotropic effect, and increases coronary flow. In a randomized trial of patients with shock, there was no significant difference overall in 28-day mortality between those receiving dopamine and those receiving norepinephrine, but norepinephrine reduced mortality in a prespecified subgroup of patients with cardiogenic shock.[■] Vasopressor infusions need to be titrated carefully in patients with cardiogenic shock to maximize coronary perfusion pressure with the least possible increase in myocardial oxygen demand. Invasive hemodynamic monitoring with an arterial line and temporary right-sided heart catheterization are advisable during the initial titration of vasoactive agents.[4]

If tissue perfusion remains inadequate despite norepinephrine, inotropic therapy should be initiated. Dobutamine (2.5 to 20 μg/kg/minute), a selective β_1-adrenergic receptor agonist, can improve myocardial contractility and increase cardiac output, and it is the initial agent of choice in patients with a low-output syndrome and systolic blood pressures greater than 90 mm Hg. Dobutamine may exacerbate hypotension in some patients because of its vasodilatory effects, and it can precipitate tachyarrhythmias. Milrinone (0.125 to 0.75 μg/kg/minute), a phosphodiesterase inhibitor, has fewer chronotropic and arrhythmogenic effects than catecholamines, but it has a long half-life and can cause hypotension; in patients whose clinical status is tenuous, it is usually reserved for situations in which other agents have proved ineffective. Levosimendan (0.05 to 0.2 μg/kg/minute) is a calcium sensitizer that has both inotropic and vasodilatory properties and does not increase myocardial oxygen consumption. Levosimendan may be more effective than dobutamine in treating low-output heart failure,[■] but it also may cause hypotension, and it must be used with caution in patients with cardiogenic shock. A randomized trial of nitric oxide inhibition did not show benefit.[■]

Intra-aortic Balloon Counterpulsation

An intra-aortic balloon pump (IABP) reduces systolic afterload and augments diastolic perfusion pressure without increasing oxygen demand. Despite a convincing hemodynamic rationale for its use, however, a recent randomized trial failed to show improvement in 30-day[■] or 1-year mortality[■] with IABP insertion in patients who have cardiogenic shock and who undergo early revascularization, perhaps because IABPs do not produce a significant improvement in blood flow distal to a critical coronary stenosis. Although this finding has dampened enthusiasm for their routine use, these trial results may not be applicable to all patients. Use of an IABP still may be a reasonable stabilizing measure in appropriately selected patients, such as supporting patients through a critical period of shock until definitive therapy is undertaken.

Reperfusion

Supportive therapy may improve blood pressure and cardiac output in cardiogenic shock, but it does not interrupt the vicious circle of myocardial dysfunction and ischemia. Rapid restoration of myocardial blood flow is the cornerstone of therapy for patients with cardiogenic shock due to MI (Chapter 73). Reperfusion therapy (see Fig. 73-3) restores patency of the infarcted artery and decreases the likelihood of progression to cardiogenic shock. After cardiogenic shock has already developed, however, fibrinolytic therapy is less effective at achieving and maintaining reperfusion, probably because of a combination of hemodynamic, mechanical, and metabolic factors that prevent the achievement and maintenance of patency in the infarct-related artery.

Prompt revascularization is the only intervention that consistently reduces mortality rates in patients with cardiogenic shock. In a randomized trial of patients with LV failure complicating ST elevation MI, cardiac catheterization with PCI or CABG within 48 hours of presentation reduced all-cause mortality marginally at 30 days (47% in the revascularization group vs. 56% in the medical therapy group; $P = 0.11$) and significantly at 6 months, 1 year,[■] and 6 years[■] compared with optimal medical management, including IABP, in 86% of patients. Subgroup analyses also revealed benefits in patients younger than 75 years, those with prior MI, and those randomized less than 6 hours from the onset of infarction. Another similarly designed trial, which was terminated early because of difficulties in patient recruitment, also showed a trend toward reduced 30-day and 1-year mortality. Together, these

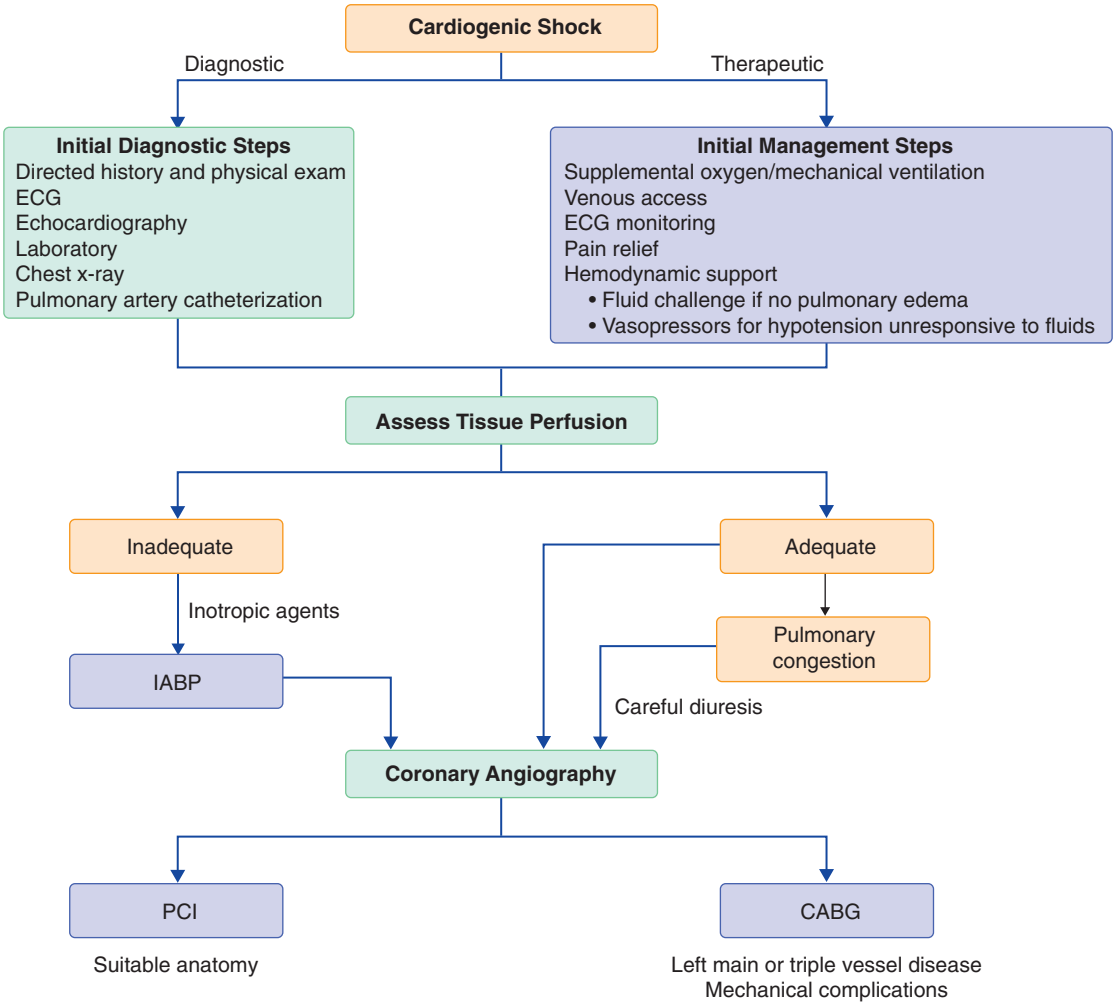

FIGURE 107-2. Approach to the diagnosis and treatment of cardiogenic shock caused by myocardial infarction. Right ventricular infarction and mechanical complications are discussed in the text. CABG = coronary artery bypass graft; ECG = electrocardiogram; IABP = intra-aortic balloon pump; PCI = percutaneous coronary intervention. (Modified from Hochman JS, Boland J, Sleeper LA, et al. Current spectrum of cardiogenic shock and effect of early revascularization on mortality. Results of an International Registry. SHOCK Registry Investigators. *Circulation.* 1995;91:873-881.)

trials suggest that about 13 patients will be saved at 1 year for each 100 patients treated.

On the basis of these results, emergent coronary revascularization, most often with PCI and stents, is the standard of care for cardiogenic shock due to pump failure in acute MI. Outcomes are best when PCI is performed within 6 hours after the onset of symptoms, but survival benefits are still demonstrable up to 48 hours after the onset of MI and 18 hours after the onset of shock. Elderly patients who are suitable for aggressive therapy also appear to benefit.

CABG surgery is more likely to provide complete revascularization and achieves long-term survival rates comparable to those of PCI, often despite worse coronary anatomy and a higher prevalence of diabetes. In practice, however, emergency CABG is performed less than 10% of the time.

When cardiogenic shock results from mechanical complications of MI (Chapter 73), surgery is recommended when feasible. For acute mitral regurgitation due to papillary muscle rupture, supportive therapy with an IABP and vasoactive agents is a temporizing measure; definitive therapy requires expeditious surgical valve repair or replacement (Chapter 75). Although mortality is 20 to 40%, survival and ventricular function are improved compared with medical therapy.

Timely surgery is also critical in patients whose cardiogenic shock is caused by ventricular septal or free wall rupture. Because perforations are exposed to shear forces, the rupture site can expand abruptly. Repair can be technically difficult because of the need to suture in areas of necrosis. Surgical mortality is 20 to 50% and is especially high for serpiginous inferoposterior ruptures, which are typically less well circumscribed than anteroapical ruptures. RV function is an important determinant of outcome in this setting. Timing of surgery has been controversial, but guidelines now recommend that operative repair be undertaken early, within 48 hours of the rupture. Placement of a septal occluding device may be helpful in selected patients.

Circulatory Support

Mechanical support with a left ventricular assist device (LVAD; Chapter 73) can interrupt the downward spiral of myocardial dysfunction, hypoperfusion, and ischemia in cardiogenic shock, allowing time for stunned or hibernating myocardium to recover. In cardiogenic shock after acute MI, percutaneous LVADs can be placed in the catheterization laboratory. These devices provide short-term support and are usually intended as a bridge to recovery or, occasionally, as a bridge to transplantation. Percutaneous LVADs provide better hemodynamics compared with IABPs, with higher cardiac indices and mean arterial pressures as well as lower filling pressures, but they have not been shown to improve mortality at 30 days. Extracorporeal support has been used in selected patients.[5]

Management of Special Conditions

At the end stage of a dilated or restrictive cardiomyopathy (Chapter 60), low cardiac output can result in cardiogenic shock. A search for reversible precipitating causes should be undertaken. Some patients will respond to inotropic therapy and will have a brief period of relative improvement. Appropriate candidates should be referred for evaluation for possible cardiac transplantation (Chapter 82). In selected patients, a surgically placed LVAD, such as a continuous-flow device, can provide 45% 2-year survival free of device surgery or disabling stroke.[6] LVADs can be used either as a bridge to transplantation or as destination therapy. A discussion about end-of-life care is also warranted.

Acute myocarditis (Chapter 60) can have a fulminant course leading to shock in 10 to 15% of cases. Patients with acute myocarditis are typically younger than those with cardiogenic shock due to MI, and they present more commonly with dyspnea rather than chest pain. Echocardiography usually shows global LV dysfunction. Supportive therapy is indicated; some patients may require circulatory support and even consideration of cardiac transplantation. Immunosuppressive therapy does not improve outcome in this setting.

Patients with hypertrophic cardiomyopathy (Chapter 60) may present with severe outflow tract obstruction and shock. Recognition of this condition is important because diuretic and inotropic therapy may worsen the obstruction. Careful volume resuscitation and use of a pure α-agonist, such as phenylephrine (0.1 to 0.3 mg/kg/minute), can increase afterload and cavity size. β-Blockers (esmolol, 0.05 to 0.2 mg/kg/minute; metoprolol, 2.5 to 5 mg IV every 2 to 5 minutes, up to 15 mg) or calcium blockers with negative inotropic properties (e.g., diltiazem, 5 mg IV every 2 minutes, up to 20 mg) can also be helpful.

Some patients with stress (takotsubo) cardiomyopathy (Chapter 60) may have LV dysfunction severe enough to produce shock. Because the presentation is similar to that of acute MI, with chest pain and ECG changes, the diagnosis is usually made in the catheterization laboratory, when significant coronary obstruction is excluded and the characteristic apical hypokinesis or dyskinesis is documented. Treatment is supportive and may include an IABP. Most patients have recovery of LV function within days to weeks, and the long-term prognosis is excellent.

Acute valvular regurgitation (Chapter 75) is manifested with pulmonary edema and decreased forward cardiac output. The regurgitant murmur may be soft or inaudible, and the diagnosis is best made by echocardiography. Acute ischemic mitral regurgitation is usually associated with rupture of the posterior papillary muscle, which has a single blood supply. Other causes include spontaneous chordal rupture, infective endocarditis (Chapter 76), rheumatic fever (Chapter 290), and trauma (Chapter 111). Immediate management includes afterload reduction (Chapter 59) and an IABP as temporizing measures. Inotropic or vasopressor therapy may also be needed to support cardiac output and blood pressure. Definitive therapy, however, consists of surgical valve repair or replacement (Chapter 75).

Acute aortic regurgitation most commonly results from infective endocarditis (Chapter 76) with leaflet destruction, but it may also be due to traumatic injury (Chapter 111) or acute aortic dissection (Chapter 78). The pulse pressure is usually narrow, indicating decreased forward stroke volume, and the bounding pulsations seen with chronic aortic regurgitation are usually absent. Temporizing measures include afterload reduction, with vasopressor and inotropic support as needed. IABP is contraindicated, and excessive slowing of the heart rate may worsen hemodynamics by prolonging diastole. Definitive therapy is surgical.

PROGNOSIS

Cardiogenic shock is still the most common cause of death in acute MI. Survival rates are improving, however, coincident with the increasing use of reperfusion therapy in appropriately selected patients. Hemodynamics predict short-term but not long-term mortality. Among patients undergoing revascularization, age, time to revascularization, and restoration of coronary blood flow independently predict survival, but the benefits of revascularization are seen at every level of risk, with an average 1-year survival of 50 to 55%. Encouragingly, the survival benefit of early revascularization is maintained at 6-year follow-up, with 5-year survival approaching 45%. The quality of life in survivors is usually excellent, with 83% either asymptomatic or having only mildly symptomatic heart failure. For patients with end-stage nonischemic myocardial disease, the prognosis is very poor in the absence of heart transplantation.

Grade A References

A1. De Backer D, Biston P, Devriendt J, et al. Comparison of dopamine and norepinephrine in the treatment of shock. *N Engl J Med*. 2010;362:779-789.

A2. Unverzagt S, Wachsmuth L, Hirsch K, et al. Inotropic agents and vasodilator strategies for acute myocardial infarction complicated by cardiogenic shock or low cardiac output syndrome. *Cochrane Database Syst Rev*. 2014;1:CD009669.

A3. Alexander JH, Reynolds HR, Stebbins AL, et al. Effect of tilarginine acetate in patients with acute myocardial infarction and cardiogenic shock: the TRIUMPH randomized controlled trial. *JAMA*. 2007;297:1657-1666.

A4. Thiele H, Zeymer U, Neumann FJ, et al. Intraaortic balloon support for myocardial infarction with cardiogenic shock. *N Engl J Med*. 2012;367:1287-1296.

A5. Thiele H, Zeymer U, Neumann FJ, et al. Intra-aortic balloon counterpulsation in acute myocardial infarction complicated by cardiogenic shock (IABP-SHOCK II): final 12 month results of a randomised, open-label trial. *Lancet*. 2013;382:1638-1645.

A6. Hochman JS, Sleeper LA, White HD, et al. One-year survival following early revascularization for cardiogenic shock. *JAMA*. 2001;285:190-192.

A7. Hochman JS, Sleeper LA, Webb JG, et al. Early revascularization and long-term survival in cardiogenic shock complicating acute myocardial infarction. *JAMA*. 2006;295:2511-2515.

A8. Jeger RV, Urban P, Harkness SM, et al. Early revascularization is beneficial across all ages and a wide spectrum of cardiogenic shock severity: a pooled analysis of trials. *Acute Card Care*. 2011;13:14-20.

A9. Cheng JM, den Uil CA, Hoeks SE, et al. Percutaneous left ventricular assist devices vs. intra-aortic balloon pump counterpulsation for treatment of cardiogenic shock: a meta-analysis of controlled trials. *Eur Heart J*. 2009;30:2102-2108.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Cooper HA, Panza JA. Cardiogenic shock. *Cardiol Clin*. 2013;31:567-580.
2. Kolte D, Khera S, Aronow WS, et al. Trends in incidence, management, and outcomes of cardiogenic shock complicating ST-elevation myocardial infarction in the United States. *J Am Heart Assoc*. 2014;3:e000590.
3. van Diepen S, Reynolds HR, Stebbins A, et al. Incidence and outcomes associated with early heart failure pharmacotherapy in patients with ongoing cardiogenic shock. *Crit Care Med*. 2014;42:281-288.
4. Hollenberg SM. Hemodynamic monitoring. *Chest*. 2013;143:1480-1488.
5. Cheng R, Hachamovitch R, Kittleson M, et al. Complications of extracorporeal membrane oxygenation for treatment of cardiogenic shock and cardiac arrest: a meta-analysis of 1,866 adult patients. *Ann Thorac Surg*. 2014;97:610-616.
6. Teuteberg JJ, Chou JC. Mechanical circulatory devices in acute heart failure. *Crit Care Clin*. 2014;30:585-606.

REVIEW QUESTIONS

1. Which of the following is the most common cause of cardiogenic shock?

A. Pump failure in acute myocardial infarction
B. Myocarditis
C. Mechanical complications of acute myocardial infarction
D. Right ventricular infarction
E. Valvular heart disease

Answer: A The predominant cause of cardiogenic shock is left ventricular failure secondary to acute myocardial infarction. The other listed answers can lead to cardiogenic shock but are less common.

2. Coronary angiography in patients with cardiogenic shock in the setting of acute myocardial infarction most commonly shows which of the following?

A. Left main coronary artery disease
B. Single-vessel left anterior descending coronary artery disease
C. Multivessel disease
D. Extensive collateralization
E. Two-vessel disease involving the right coronary artery

Answer: C In cardiogenic shock resulting from acute myocardial infarction, coronary angiography most often demonstrates multivessel disease. About 30% of patients have a left main coronary artery occlusion, about 60% have three-vessel coronary disease, and only about 20% have single-vessel disease.

3. Which of the following is true concerning the use of vasopressor agents in cardiogenic shock?

A. Norepinephrine is more arrhythmogenic than dopamine.
B. Norepinephrine is associated with decreased 28-day mortality compared with dopamine in patients with cardiogenic shock.
C. Dopamine is preferable because it improves renal function.
D. Dobutamine is more effective at raising blood pressure.
E. Blood pressure should be monitored noninvasively to avoid vascular complications.

Answer: B In a randomized trial of patients with shock, norepinephrine reduced 28-day mortality in a prespecified subgroup of patients with cardiogenic shock compared with dopamine. Dopamine was more arrhythmogenic and does not improve renal function in patients with shock, although some patients have increased urine output. Dobutamine has vasodilatory effects and can worsen hypotension. Noninvasive blood pressure monitoring can be unreliable in patients with shock; arterial line placement is recommended.

4. Complete the following statement correctly: Percutaneous left ventricular assist devices for cardiogenic shock

A. Improve hemodynamics compared with intra-aortic balloon pumping.
B. Improve 30-day mortality compared with intra-aortic balloon pumping.
C. Should be reserved for patients eligible for cardiac transplantation.
D. Are associated with decreased vascular complication rates compared with intra-aortic balloon pumping.
E. Provide support independent of right ventricular function.

Answer: A Percutaneous left ventricular assist devices (LVADs) provide short-term hemodynamic support after cardiogenic shock. By allowing time for left ventricular recovery, they are usually intended as a bridge to definitive therapy, such as transplantation. Percutaneous LVADs provide better hemodynamics compared with an intra-aortic balloon pump, with higher cardiac indices and mean arterial pressures as well as lower filling pressures; however, they have not been shown to improve mortality at 30 days. Known complications of percutaneous LVADs include limb ischemia and bleeding. The available percutaneous devices support the left ventricle and thus require adequate right ventricular function.

5. Which of the following is true concerning prognosis after cardiogenic shock in the setting of myocardial infarction?

A. Hemodynamics predict long-term mortality among patients undergoing revascularization.
B. The early survival benefits of revascularization are lost by 5-year follow-up.
C. Benefits of revascularization are seen only in younger patients.
D. Prognosis is worse than in patients with end-stage nonischemic myocardial disease.
E. Quality of life in survivors is usually excellent.

Answer: E The quality of life in survivors of cardiogenic shock complicating acute myocardial infarction is usually excellent, with 83% of patients either asymptomatic or having only mildly symptomatic heart failure. Hemodynamics predict short-term but not long-term mortality. The survival benefit of early revascularization is maintained at 6-year follow-up, with 5-year survival approaching 45%. Among patients undergoing revascularization, age and time to revascularization predict survival, but the benefits of revascularization are seen at every level of risk, with an average 1-year survival of 50 to 55%. For patients with end-stage nonischemic myocardial disease, the prognosis is very poor in the absence of heart transplantation.

108 SHOCK SYNDROMES RELATED TO SEPSIS

JAMES A. RUSSELL

DEFINITION

Sepsis is defined by presence of at least two of the four signs of the systemic inflammatory response syndrome (SIRS): (1) fever (>38° C) or hypothermia (<36° C); (2) tachycardia (>90 beats/minute); (3) tachypnea (>20 breaths/minute), hypocapnia (partial pressure of carbon dioxide <32 mm Hg), or the need for mechanical ventilatory assistance; and (4) leukocytosis (>12,000 cells/μL), leukopenia (<4000 cells/μL), or a left shift (>0% immature band cells) in the circulating white blood cell differential and suspected or proven infection. *Bacteremia* is defined as the growth of bacteria in blood cultures, but infection does not have to be proved to diagnose sepsis at the onset. *Severe sepsis* is sepsis in addition to dysfunction of one or more organ systems (e.g., hypoxemia, oliguria, lactic acidosis, thrombocytopenia, decreased Glasgow Coma Scale score). *Septic shock* is defined as severe sepsis in addition to hypotension (systolic blood pressure <90 mm Hg or a >40 mm Hg decrease from baseline) despite adequate fluid resuscitation.[1]

EPIDEMIOLOGY

Approximately 750,000 cases of severe sepsis or septic shock occur every year in the United States. Sepsis causes as many deaths as acute myocardial infarction, and septic shock and its complications are the most common causes of death in noncoronary intensive care units (ICUs). The medical care costs associated with sepsis are approximately $16.7 billion a year in the United States alone. The frequency of septic shock is increasing as physicians perform more aggressive surgery, as more resistant organisms are present in the environment, and as the prevalence of immune compromise resulting from disease and immunosuppressive drugs increases. Studies suggest that African Americans have a higher incidence of severe sepsis than whites do (6.0 vs. 3.6 per 1000 population) and a higher mortality in ICUs (32.1 vs. 29.3%; $P < .0001$), even after adjustment for poverty levels. The mechanisms of this apparent difference in risk for and mortality from sepsis are not known.

Gram-positive or gram-negative bacteria, fungi, and, very rarely, protozoa or rickettsiae can cause septic shock. Increasingly common causes of septic shock are gram-positive bacteria, especially methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant enterococci, penicillin-resistant *Streptococcus pneumoniae*, and resistant gram-negative bacilli.

The common infections causing septic shock are pneumonia, peritonitis, pyelonephritis, abscess (especially intra-abdominal), primary bacteremia, cholangitis, cellulitis, necrotizing fasciitis, and meningitis. Nosocomial pneumonia is the most common cause of death from nosocomial infection.

PATHOBIOLOGY

At onset, septic shock activates inflammation, leading to enhanced coagulation, activated platelets, increased neutrophils and mononuclear cells, and

diminished fibrinolysis. After several days, a compensatory anti-inflammatory response with immunosuppression may contribute to death. Several pathways amplify one another: inflammation triggers coagulation, and coagulation triggers inflammation, resulting in a positive feedback loop that is proinflammatory and procoagulant. Tissue hypoxia in septic shock also amplifies inflammation and coagulation. Many mediators that are critical for the homeostatic control of infection may be injurious to the host (e.g., tumor necrosis factor-α [TNF-α]), so therapies that fully neutralize such mediators are largely ineffective.

Widespread endothelial injury is an important feature of septic shock; an injured endothelium is more permeable, so the flux of protein-rich edema fluid into tissues such as the lung increases. Injured endothelial cells release nitric oxide, a potent vasodilator that is a key mediator of septic shock. Septic shock also injures epithelial cells of the lung and intestine. Intestinal epithelial injury increases intestinal permeability; this leads to epithelial translocation of intestinal bacteria and endotoxin, which further augments the inflammatory phenotype of septic shock.

Early Infection, the Innate Immune Response, Inflammation, and the Endothelium

Host defense is organized into innate and adaptive immune responses. The innate immune system responds by using pattern recognition receptors (e.g., toll-like receptors [TLRs]) to pathogen-associated molecular patterns, which are extremely well conserved molecules of microorganisms. Surface molecules of gram-positive and gram-negative bacteria (peptidoglycan and lipopolysaccharide, respectively) bind to TLR-2 and TLR-4, respectively (E-Fig. 108-1). TLR-2 and TLR-4 binding initiates an intracellular signaling cascade that culminates in nuclear transport of the transcription factor nuclear factor κB (NF-κB), which triggers transcription of cytokines such as TNF-α and interleukin (IL)–6. Cytokines upregulate adhesion molecules of neutrophils and endothelial cells, and neutrophil activation leads to bacterial killing. However, cytokines also directly injure host endothelial cells, as do activated neutrophils, monocytes, and platelets. Inhibition of early cytokine mediators of sepsis, such as TNF-α and IL-1β, has not proved successful, probably because TNF-α and IL-1β peak and then decline quickly, before these antagonist therapies can be applied clinically.

After the early cytokine inflammatory response, immune cells, including macrophages and neutrophils, release later mediators, such as high-mobility group box 1 (HMGB-1). HMGB-1 activates neutrophils, monocytes, and endothelium. Unlike TNF-α antagonists, inhibitors of HMGB-1 decrease mortality even when they are given 24 hours after the induction of experimental peritonitis.

Another adverse effect of sepsis is widespread endothelial injury that leads to increased endothelial permeability with loss of protein and fluids to the interstitial space. This endothelial permeability is a final common pathway for widespread tissue injury. Cytokines and other inflammatory mediators induce intercellular endothelial cell gaps by disrupting intercellular junctions, by changing cytoskeletal structure, or by direct damage to endothelial cells. Several pathways of altered endothelial permeability have been implicated in sepsis, including protease-activated receptor 1 (PAR-1) and disruption of the intercellular VE-cadherin, β-catenin, and p120-catenin complex. PAR-1 binding by activated protein C and low-dose thrombin is cytoprotective, whereas PAR-1 stimulation by high-dose thrombin increases endothelial permeability. Binding of Slit to Robo4 maintains the integrity of the intercellular VE-cadherin, β-catenin, and p120-catenin complexes and thus maintains healthy endothelial permeability.

Adaptive Immunity Adds Specificity and Amplifies the Immune Response

Microorganisms stimulate specific humoral and cell-mediated adaptive immune responses that amplify innate immunity. B cells release immunoglobulins that bind to microorganisms and thereby facilitate delivery of microorganisms to natural killer cells and neutrophils. In sepsis, type 1 helper T (T_H1) cells generally secrete proinflammatory cytokines (TNF-α, IL-1β), and type 2 helper T (T_H2) cells secrete anti-inflammatory cytokines (IL-4, IL-10).

Coagulation Response to Infection

Septic shock activates the coagulation system (E-Fig. 108-2) and ultimately converts fibrinogen to fibrin, which is bound to platelets to form microvascular thrombi. Microvascular thrombi further amplify endothelial injury by the release of mediators and by tissue hypoxia because of obstruction to blood flow.

Normally, natural anticoagulants (protein C, protein S, antithrombin, and tissue factor pathway inhibitor) dampen coagulation, enhance fibrinolysis, and remove microthrombi. Thrombin-α binds to thrombomodulin, which activates protein C when protein C is bound to the endothelial protein C receptor (EPCR). Activated protein C dampens the procoagulant phenotype because it inactivates factors Va and VIIIa and inhibits the synthesis of plasminogen activator inhibitor 1 (PAI-1). Activated protein C also decreases apoptosis, leukocyte activation and adhesion, and production of cytokines.

Septic shock decreases the levels of the natural anticoagulants protein C, protein S, antithrombin, and tissue factor pathway inhibitor. Furthermore, lipopolysaccharide and TNF-α decrease thrombomodulin and EPCR, thereby limiting the activation of protein C. Lipopolysaccharide and TNF-α also increase levels of PAI-1, inhibiting fibrinolysis.

Tissue Hypoxia in Septic Shock

Tissue hypoxia independently activates inflammation (by activation of NF-κB and cytokines, synthesis of nitric oxide, and activation of HMGB-1), induces coagulation (through tissue factor and PAI-1), and activates neutrophils, monocytes, and platelets. Hypoxia induces hypoxia-inducible factor-1α (HIF-1α), which upregulates erythropoietin, and vascular endothelial growth factor (VEGF). Erythropoietin is protective to brain and other tissues. VEGF inhibits fibrinolysis and increases inducible nitric oxide synthase, which augments nitric oxide–induced vasodilation. Nitric oxide has a further injurious effect: excessive nitric oxide inhibits the beneficial actions of HIF-1α (e.g., upregulating synthesis of erythropoietin) during hypoxia.

Late Septic Shock, Immunosuppression, and Apoptosis of Immune and Epithelial Cells

After about 1 week of septic shock, death can result from immunosuppression, which is suggested by anergy, lymphopenia, hypothermia, and nosocomial infection (E-Fig. 108-3). Multiple organ dysfunction may be an anti-inflammatory phenotype because of the apoptosis of immune, epithelial, and endothelial cells. Activated $CD4^+$ T cells evolve into either a T_H1 proinflammatory (TNF-α, IL-1β) or a T_H2 anti-inflammatory (IL-4, IL-10) phenotype. Sepsis leads to migration from a T_H1 to a T_H2 phenotype; for example, persistent elevation of IL-10 is associated with an increased risk of death. Immunosuppression also develops because of apoptosis of lymphocytes. Proinflammatory cytokines, activated B and T cells, and glucocorticoids induce lymphocyte apoptosis, whereas TNF-α and endotoxin induce apoptosis of lung and intestinal epithelial cells. The fact that glucocorticoids also stimulate apoptosis could be the biologic explanation for the observation that patients with septic shock who are treated with hydrocortisone have more superinfections than do patients treated with placebo.

Death from infectious disease appears to be highly heritable. Sepsis is a prime example of a polygenic disease related to the interaction of multiple genes and an environmental insult (infection). Single-nucleotide polymorphisms of cytokines (TNF-α, IL-6, IL-10), coagulation factors (protein C, fibrinogen-β), the catecholamine pathway (β-adrenergic receptor), and innate immunity genes (CD14, TLR-1, TLR-2) have been variably associated with an increased risk of death from sepsis.

Cardiovascular Dysfunction

Inadequate tissue perfusion and tissue hypoxia are the cardinal features of all types of shock. Early in septic shock, most patients have sinus tachycardia and, by definition, decreased blood pressure (<90 mm Hg systolic, a decrease of ≥40 mm Hg from baseline systolic pressure, or mean arterial pressure <65 mm Hg; Table 108-1). Septic shock is the classic form of distributive shock (Chapter 106), characterized by increased pulse pressure (bounding pulses), decreased systemic vascular resistance (warm, flushed skin), and functional hypovolemia (low jugular venous pressure). Distributive shock means that the distribution of systemic blood flow is abnormal, such that areas of both low flow (and low venous oxygen saturation) and high flow (and increased venous oxygen saturation) are present. Nevertheless, about one third of patients with septic shock initially present with findings more typical of hypovolemic shock (low central venous pressure and low central venous oxygen saturation) because the clinical features depend on the stage and severity of septic shock as well as on the degree of fluid resuscitation that has occurred. After fluid resuscitation, patients typically develop the characteristic clinical and hemodynamic features of classic distributive shock.

Ventricular preload is commonly decreased in early septic shock, for several reasons.[2] First, patients may be volume depleted because of decreased fluid intake and because of increased fluid losses as a result of fever, vomiting, and

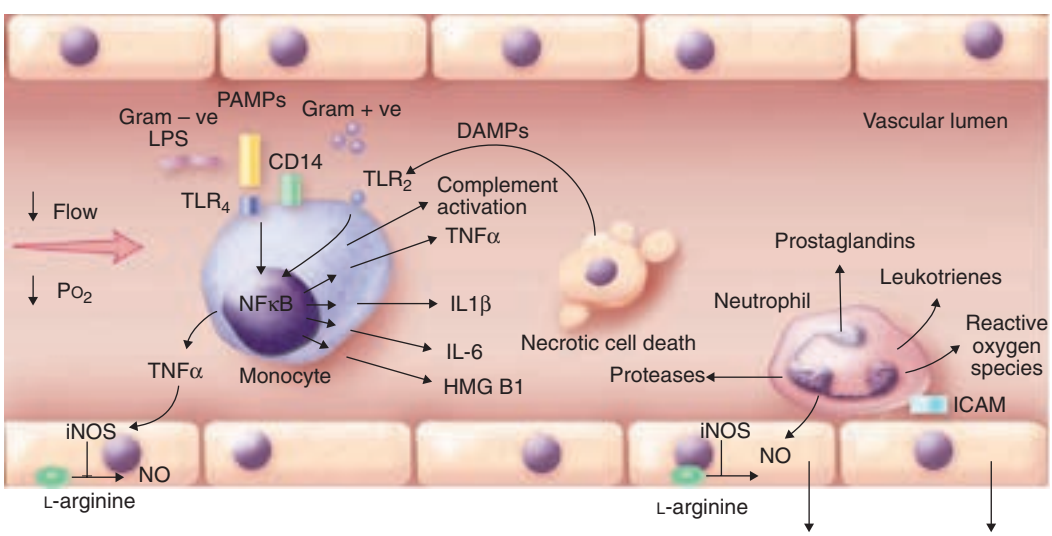

E-FIGURE 108-1. Inflammatory responses to sepsis. Gram-positive and gram-negative bacteria, viruses, and fungi have unique cell wall molecules called pathogen-associated molecular patterns (PAMPs) that bind to pattern recognition receptors (called toll-like receptors [TLRs]) on the surface of immune cells as well as C-type lectin receptors, retinoic acid inducible gene 1–like receptors, and nucleotide-binding oligomerization domain–like receptors. Nucleotide-binding oligomerization domain–like receptors alter protein complexes in the inflammasome. The lipopolysaccharide (LPS) of gram-negative bacilli binds to LPS-binding protein–CD14 complex. The peptidoglycan of gram-positive bacteria and the LPS of gram-negative bacteria bind to TLR-2 and TLR-4, respectively. TLR-2 and TLR-4 binding activates intracellular signal transduction pathways that lead to the activation of the cytosolic transcription factor nuclear factor κB (NF-κB). Activated NF-κB moves from the cytoplasm to the nucleus, binds to transcription start sites, and increases the transcription of proinflammatory cytokines such as tumor necrosis factor-α (TNF-α), interleukin-1β (IL-1β), and interleukin-6 (IL-6). Increased necrotic cell death releases damage-associated molecular patterns (DAMPs or alarmins) that bind to TLRs and thus feedback to further amplify the proinflammatory response. TNF-α, IL-1β, and IL-6 are proinflammatory cytokines that activate the adaptive immune response but also cause both direct and indirect host injury (e.g., by complement activation). Sepsis increases the activity of inducible nitric oxide synthase (iNOS), which increases the synthesis of nitric oxide (NO), a potent vasodilator. Cytokines activate endothelial cells by upregulating adhesion receptors such as intercellular adhesion molecule (ICAM), and they injure endothelial cells by the activation and binding of neutrophils, monocytes, macrophages, and platelets to endothelial cells. These effector cells release mediators such as proteases, reactive oxygen species, prostaglandins, and ICAM leukotrienes. Cytokines also activate the coagulation cascade.

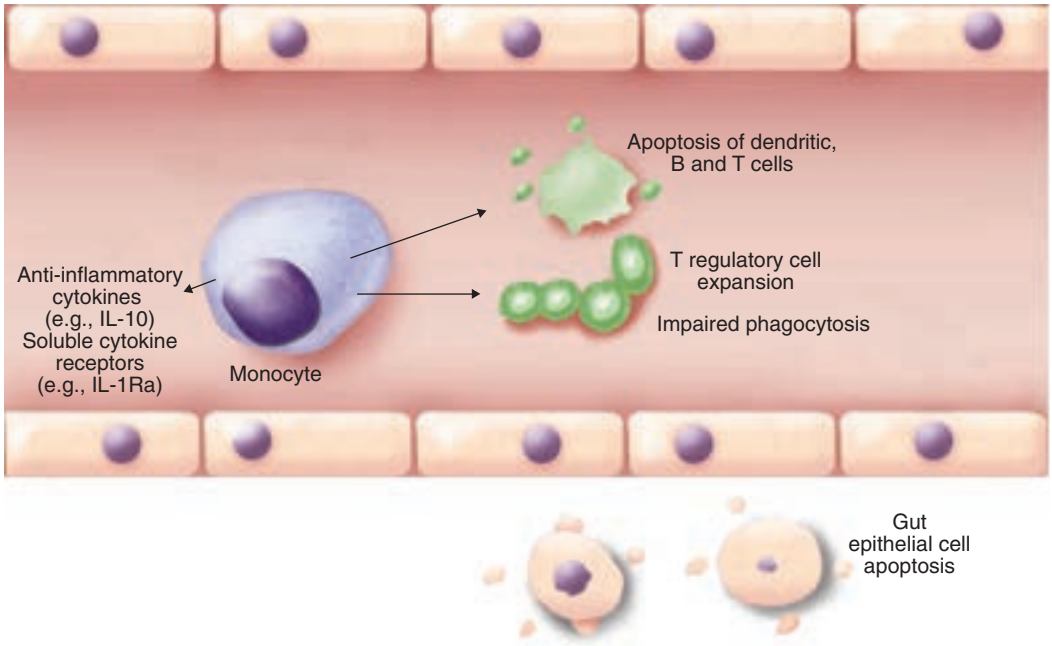

E-FIGURE 108-2. Procoagulant response in sepsis. Sepsis initiates coagulation by activating the endothelium to increase tissue factor. Protease-activated receptors (PARs), especially PAR-1, link the inflammatory and coagulation responses to sepsis. Activation of factors Va and VIIIa leads to the formation of thrombin-α, which converts fibrinogen to fibrin. Fibrin binds to platelets that adhere to endothelial cells, forming microvascular thrombi. Microvascular thrombi amplify injury by the release of mediators and by microvascular obstruction, which causes distal ischemia and tissue hypoxia. Normally, natural anticoagulants—protein C (PC), protein S (PS), antithrombin, and tissue factor pathway inhibitor (TFPI)—dampen coagulation, enhance fibrinolysis, and remove microthrombi. Thrombin-α binds to thrombomodulin on endothelial cells and thus activates the binding of PC to endothelial PC receptor (EPCR). PC forms a complex with its cofactor PS. PC binding to EPCR increases the activation of PC to activated PC (APC). APC proteolytically inactivates factors Va and VIIIa and decreases the synthesis of plasminogen activator inhibitor 1 (PAI-1). Sepsis decreases levels of PC, PS, antithrombin, and TFPI. Lipopolysaccharide and tumor necrosis factor-α (TNF-α) decrease thrombomodulin and EPCR, thus decreasing the activation of PC. PAR-1 binding by activated protein C and low-dose thrombin is cytoprotective, whereas PAR-1 stimulation by high-dose thrombin increases endothelial permeability. Lipopolysaccharide and TNF-α also inhibit PAI-1, so fibrinolysis is inhibited. HIFα = hypoxia-inducible factor-α; NO = nitric oxide; tPA = tissue plasminogen activator; VEGF = vascular endothelial growth factor.

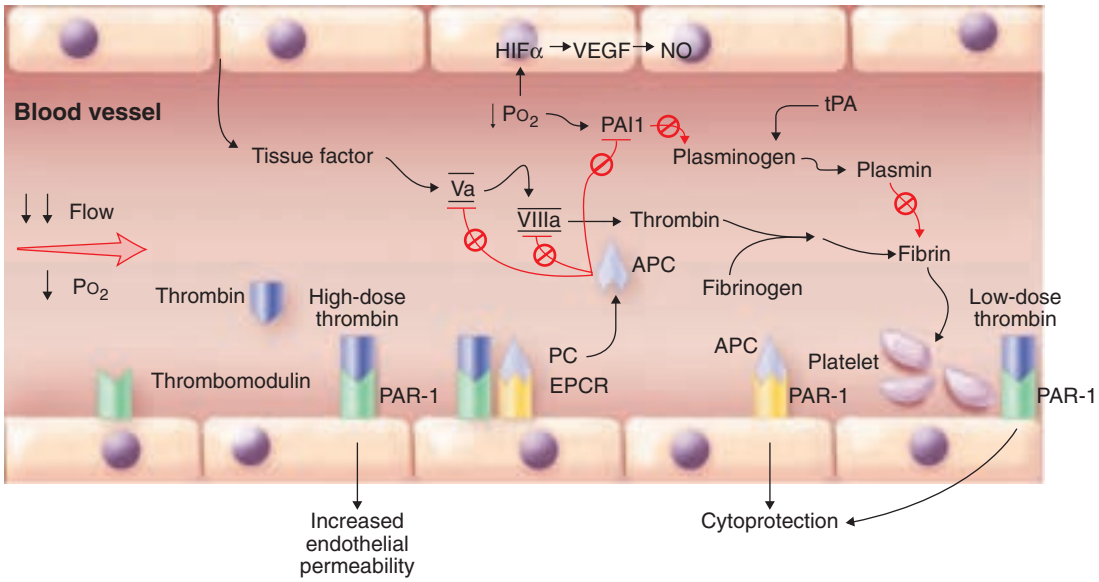

E-FIGURE 108-3. Anti-inflammatory, immunosuppressive, and proapoptotic responses in sepsis. In parallel with the proinflammatory response, the host generates anti-inflammatory, immunosuppressive, and proapoptotic responses. Activated neutrophils increase the transcription of anti-inflammatory cytokines such as interleukin-10 (IL-10), and IL-10 then inactivates macrophages and has other anti-inflammatory effects. Apoptosis of dendritic cells, T and B cells, and gut epithelial cells leads to sometimes profound immunosuppression that increases the risk of secondary nosocomial infections. There is also increased production of T-regulatory cells that then impair phagocytosis of organisms. Initially increased pituitary adrenocorticotropic hormone (ACTH)–induced adrenal release of cortisol inhibits the proinflammatory response by regulating transcription of many cytokines and other mediators of inflammation.

TABLE 108-1 HEMODYNAMIC VARIABLES, ABBREVIATIONS, AND NORMAL VALUES

Arterial pressure: systolic pressure (SAP) (>100 mm Hg), diastolic pressure, pulse pressure, mean arterial pressure (MAP) (>65 mm Hg)
Central venous pressure (CVP): normal, 6-12 mm Hg
Pulmonary artery pressure (PAP): normal, 25/15 mm Hg
Pulmonary vascular resistance (PVR): normal, 150-250 dynes/sec/cm

$$\left(\equiv \frac{PAP - PAOP}{CO} \times 80\right)$$

Pulmonary artery occlusion pressure (PAOP) or pulmonary artery wedge pressure (PAWP): normal, 8-15 mm Hg
Systemic vascular resistance (SVR): normal, 900-1400 dynes/sec/cm

$$\left(\equiv \frac{MAP - CVP}{CO} \times 80\right)$$

Cardiac output (CO): normal, 5 L/min
Left ventricular stroke work index (LVSWI): normal, (60-100 grams × meters/beats) = (SV × [MAP – PAWP] × 0.0136)
Oxygen delivery (Do_2): normal, 1 L/min (= CO × [Hg × 1.38 × Sao_2] + [0.003 × Po_2])
Oxygen consumption (Vo_2): normal, 250 mL/min (= CO × Hg × 1.38 × [Sao_2 – Svo_2] + [0.003 × (Pao_2 – Pvo_2)])
Oxygen extraction ratio: normal, 0.23-0.32 (= Vo_2/Do_2)
Hemodynamic variables are often normalized to account for different body mass by dividing by body surface area (BSA)
- Pulmonary vascular resistance index (PVRI): normal (= PVR/BSA)
- Systemic vascular resistance index (SVRI): normal (= SVR/BSA)
- Cardiac index (CI): normal, 2.5-4.2 L/min/m^2 (= CO/BSA)
- Left ventricular stroke work index (LVSWI): normal (= LVSW/BSA)
- Oxygen delivery index (Do_2I): normal, 460-650 mL/min/m^2 (= Do_2/BSA)
- Oxygen consumption index (Vo_2I): normal, 95-170 mL/min/m^2 (= Vo_2/BSA)

diarrhea if gastrointestinal disease is present. Second, fluid loss from the intravascular to the interstitial space (capillary leak) is caused by mediators that induce widespread endothelial injury, which increases capillary permeability. Increased capillary permeability leads to loss of protein-rich edema fluid into the interstitial space. In the lung, increased permeability is a key component of acute lung injury. A third reason that ventricular preload is decreased in septic shock is venodilation induced by mediators such as nitric oxide. Venodilation increases venous capacitance, leading to relative volume depletion, which compounds the absolute volume depletion. Ventricular afterload is decreased because of excessive release of potent vasodilators such as nitric oxide, prostaglandin I_2, adenosine diphosphate, and other vasodilators.

In addition to abnormal vasodilation, patients have concurrent microvascular vasoconstriction. Microvascular vasoconstriction may not be apparent clinically or hemodynamically, but it can lead to tissue hypoxia, detected by increased arterial lactate concentrations. Microvascular vasoconstriction is caused by increased norepinephrine, thromboxanes, and other local vasoconstrictors. Microvascular vasoconstriction causes focal hypoxia, which is exacerbated by microvascular obstruction by platelets and leukocytes.

The abnormal mismatch of oxygen delivery to oxygen demand can disturb the global relationship of oxygen delivery to oxygen consumption. Normally, oxygen consumption is independent of oxygen delivery over a wide range. When oxygen delivery decreases to less than the critical oxygen delivery level, oxygen consumption decreases and leads to a state in which oxygen consumption depends on oxygen delivery. At levels lower than the critical oxygen delivery level, arterial lactate increases as a result of tissue hypoxia. The clinical implication is that oxygen delivery should be increased (e.g., by increasing cardiac output by volume resuscitation, infusion of dobutamine, or transfusion of erythrocytes) to more than the critical level.

Cardiovascular function is further compromised in septic shock because of decreased ventricular contractility.[3] Decreased ventricular contractility may be difficult to detect clinically and may be diagnosed only by hemodynamic or echocardiographic assessment. Numerous circulating mediators of sepsis, including endotoxin, cytokines (e.g., IL-6, TNF-α), and nitric oxide (locally released into the coronary circulation), decrease contractility. Endotoxin signals through TLRs to upregulate the expression of proteins such as S110A8 and S100A9 to cause a receptor for advanced glycation end products (RAGE)–dependent decrease in calcium flux, which decreases the ejection fraction. Coronary ischemia resulting from microvascular obstruction by leukocytes and oxygen free radicals, which are released by neutrophils adherent to the coronary capillary endothelium, is another mechanism of decreased contractility.

Early in septic shock, patients who survive have increased left ventricular end-diastolic volume, which likely allows them to maintain cardiac output despite decreased contractility. In contrast, nonsurvivors do not have increased left ventricular end-diastolic volume, so their cardiac output is compromised. In some patients with septic shock, concurrent acute lung injury and secondary pulmonary hypertension increase right ventricular afterload, with a secondary shift of the interventricular septum from right to left. This septal shift decreases left ventricular end-diastolic volume and can also limit cardiac output.

CLINICAL MANIFESTATIONS

Cardiovascular dysfunction in septic shock is characterized by decreased preload (because of decreased intake, fluid losses, third spacing resulting from increased permeability, and venodilation), decreased afterload, and often decreased ventricular contractility. Decreased ventricular volume is detected clinically by low jugular venous pressure and hemodynamically by decreased central venous pressure. Left ventricular resistance, or afterload, is also commonly decreased and is detected clinically by warm, flushed skin and hemodynamically by decreased systemic vascular resistance.

DIAGNOSIS

Even as the diagnostic evaluation is beginning, the initial assessment of a critically ill patient must focus immediately on the airway (need for intubation), breathing (respiratory rate, respiratory distress, pulse oximetry), circulation (heart rate, blood pressure, jugular venous pressure, skin perfusion), and rapid initiation of resuscitation (Fig. 108-1). Vital signs and the leukocyte count quickly establish whether the patient has SIRS (two of the four criteria). Arterial blood gases and lactate levels are useful complementary tests. A secondary survey is designed to determine the likely source of infection and the status of organ function. Pneumonia (Chapter 97) is suggested by cough, sputum, and respiratory distress; empyema (Chapter 99) is suggested by pleuritic chest pain. Signs of peritonitis, an abdominal mass, and right upper quadrant tenderness suggest abdominal sepsis. Pyelonephritis (Chapter 284) is likely in patients with dysuria and costovertebral angle tenderness. Integumentary assessment for erythema (cellulitis), line site erythema (line sepsis), tenderness (necrotizing fasciitis), crepitus (anaerobic myonecrosis), and petechiae and purpura (meningococcemia) can be illuminating. Headache, stiff neck, and signs of meningismus raise the suspicion of meningitis (Chapter 412). Focal neurologic signs suggest brain abscess (Chapter 413).

Laboratory investigations that are helpful to identify the source of infection include appropriate cultures and Gram stains (blood, sputum, urine, fluids, and cerebrospinal fluid). Blood cultures are positive in 40 to 60% of patients who have septic shock. The chest radiograph aids in the diagnosis of pneumonia, empyema, and acute lung injury. Abdominal ultrasound and computed tomography are indicated if abdominal sepsis is suspected.

Hemodynamic assessment of the patient includes diagnostic central venous or pulmonary artery catheterization. In early septic shock, central venous pressure is usually low and increases in response to volume resuscitation. Central venous oxygen saturation, cardiac output, and ventricular filling pressures may be determined continuously. Pulmonary artery pressure is usually normal but may be increased because septic shock can cause pulmonary hypertension. Pulmonary artery occlusion (or wedge) pressure is usually low before resuscitation, but it may be normal or increased if the patient has underlying preexisting heart disease (e.g., heart failure or coronary artery disease with prior myocardial infarction) or if left ventricular contractility is decreased by sepsis. Cardiac output may be low or normal before fluid resuscitation and typically increases to higher than normal after fluid resuscitation. If fluid resuscitation increases central venous pressure and pulmonary artery occlusion pressure but cardiac output does not increase, left ventricular dysfunction is presumably present.

Echocardiographic features of decreased ventricular contractility include decreased right and left ventricular ejection fractions and increased end-diastolic and end-systolic volumes. Early in septic shock, the left ventricular ejection fraction is decreased, and it remains low in nonsurvivors. In survivors, the left ventricular ejection fraction usually returns to normal during 5 to 10 days. Bedside echocardiography can also be used to assess intravascular volume status, which can be diagnosed on the basis of collapse of the inferior vena cava, and valvular dysfunction.

Renal, hepatic, and coagulation function tests are helpful to evaluate organ function. After determination of the source of sepsis, it is crucial to address

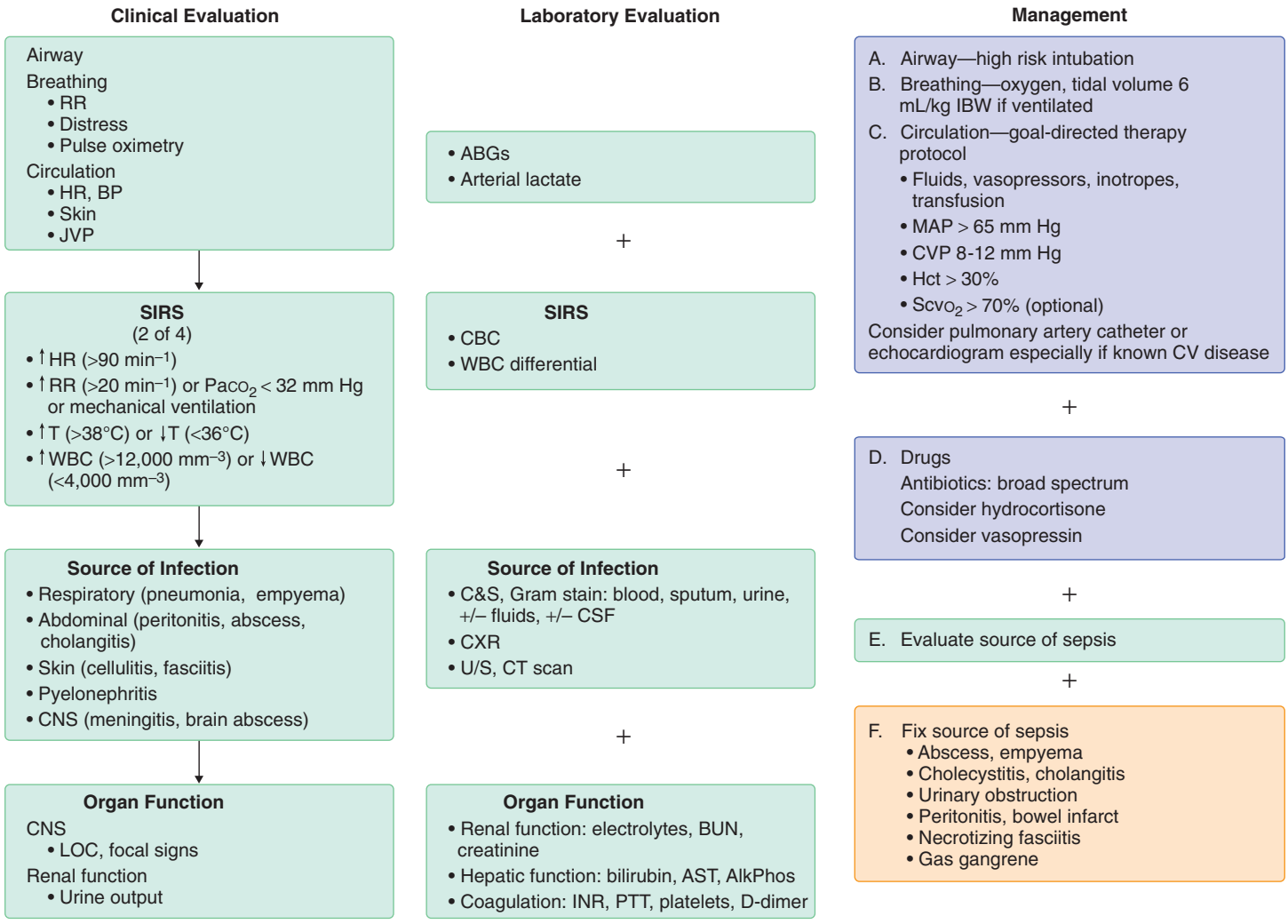

FIGURE 108-1. Algorithm for the clinical and laboratory evaluation and management of septic shock. ABGs = arterial blood gases; AlkPhos = alkaline phosphatase; AST = aspartate aminotransferase; BP = blood pressure; BUN = blood urea nitrogen; C&S = culture and sensitivity; CBC = complete blood count; CNS = central nervous system; CSF = cerebrospinal fluid; CT = computed tomography; CV = cardiovascular; CVP = central venous pressure; CXR = chest radiograph; Hct = hematocrit; HR = heart rate; IBW = ideal body weight; INR = international normalized ratio; JVP = jugular venous pressure; LOC = level of consciousness; MAP = mean arterial pressure; $Paco_2$ = partial pressure of carbon dioxide; PTT = partial thromboplastin time; RR = respiratory rate; $Scvo_2$ = central venous oxygen saturation; SIRS = systemic inflammatory response syndrome; T = temperature; U/S = ultrasound; WBC = white blood cell count.

that source by draining abscesses and empyemas; radiologically or surgically correcting urinary tract obstruction; and surgically managing peritonitis, bowel infarction, cholecystitis, cholangitis, necrotizing fasciitis, and gas gangrene.

Differential Diagnosis

The major differential diagnoses of classic septic shock are other nonseptic causes of SIRS, such as acute pancreatitis (Chapter 144), acute respiratory distress syndrome (Chapter 104), aspiration pneumonitis (Chapter 94), multiple trauma (Chapter 111), and recent major surgery without infection (Chapter 433). Other causes of distributive shock are anaphylactic shock (suggested by angioedema and hives; Chapter 440), spinal shock (recent trauma and paraplegia; Chapter 399), acute adrenal insufficiency ("tanned skin," hyperkalemia, metabolic alkalosis; Chapter 227), and acute or acute-on-chronic hepatic failure (jaundice, ascites, encephalopathy; Chapter 153).

The differential diagnosis of septic shock must include the other causes of shock: hypovolemic, cardiogenic, and obstructive shock (Chapters 106 and 107). Patients with hypovolemic shock (from internal or external fluid losses, hemorrhage) present with a suggestive history and signs of hypovolemia (low jugular venous pressure) and skin hypoperfusion (cool, clammy, cyanotic extremities). Cardiogenic shock (resulting from myocardial infarction or acute-on-chronic congestive heart failure or occurring after cardiovascular surgery) is suggested by the history, signs of increased filling pressure (increased jugular venous pressure, crackles, S_3, pulmonary edema, cardiomegaly), and skin hypoperfusion (Chapter 107). Some patients who have acute myocardial infarction and cardiogenic shock have features of SIRS without infection. Obstructive shock (from pulmonary thromboembolism, cardiac tamponade, pneumothorax) is manifested similarly to cardiogenic shock.

PREVENTION

Measures to prevent sepsis include handwashing, elevation of the head of the bed, scrupulous sterile techniques for the insertion of catheters, and possibly the use of antibiotic-impregnated catheters. New catheter insertion sites for catheter changes, isolation of patients who have resistant organisms, and isolation of significantly immunocompromised patients may also prevent infection.

Preventing the progression from sepsis to septic shock requires early diagnosis and aggressive resuscitation.[4] Early fluid resuscitation, lung-protective ventilation, and antibiotics are critical therapies in early septic shock (Table 108-2).

TREATMENT

Respiratory Therapy

All patients in septic shock require oxygen initially, and many require mechanical ventilation. Mechanical ventilation is required in most patients who have septic shock because acute lung injury is the most common complication. Lung-protective ventilation (mechanical ventilation that minimizes lung injury by using a relatively low tidal volume, such as >6 mL/kg of

TABLE 108-2 POTENTIAL ANTIBIOTIC REGIMENS FOR PATIENTS WITH SEPTIC SHOCK*

SOURCE OF SEPSIS	INITIAL ANTIBIOTIC REGIMEN	ALTERNATIVE ANTIBIOTIC REGIMEN
Community-acquired pneumonia	Third-generation cephalosporin (cefotaxime 2 g IV q6h; ceftriaxone 2 g IV q12h; ceftizoxime 2 g IV q8h) *plus* Fluoroquinolone (e.g., ciprofloxacin 400 mg IV q12h, levofloxacin 750 mg IV q24h, moxifloxacin 400 mg IV q24h) *or* Macrolide (azithromycin 500 mg IV q24h)	Piperacillin-tazobactam 3.375 g IV q6h *plus* Fluoroquinolone *or* Macrolide
Hospital-acquired pneumonia	Imipenem 0.5 g IV q6h *or* Meropenem 1 g IV q8h	Fluoroquinolone (ciprofloxacin 400 mg IV q12h) *plus* Vancomycin 1.5 g IV q12h *or* Piperacillin-tazobactam 3.375 g IV q6h *plus* Tobramycin 1.5 mg/kg q8h *plus* Vancomycin
Abdominal (mixed aerobic/anaerobic)	Piperacillin-tazobactam 3.375 g IV q6h *or* Imipenem 0.5 g IV q6h (or meropenem 1 g IV q8h)	Ampicillin 2 g IV q4h *plus* Metronidazole 500 mg IV q8h *plus* Fluoroquinolone (ciprofloxacin 400 mg IV q12h)
Urinary tract	Fluoroquinolone (ciprofloxacin 400 mg IV q12h)	Ampicillin 2 g IV q4h *plus* Gentamicin 1.5 mg/kg IV q8h *or* Third-generation cephalosporin (cefotaxime 2 g IV q6h, ceftriaxone 2 g IV q12h, or ceftizoxime 2 g IV q8h)
Necrotizing fasciitis	Imipenem 0.5 g IV q6h	Penicillin G (if confirmed group A streptococci)
Primary bacteremia (normal host)	Piperacillin-tazobactam 3.375 g IV q6h *plus* Vancomycin 1.5 g IV q12h	Imipenem 0.5 g IV q6h *plus* Vancomycin 1.5 g IV q12h
Primary bacteremia (intravenous drug user)	Vancomycin 1.5 g IV q12h *plus* Fluoroquinolone (ciprofloxacin 400 mg IV q12h)	Piperacillin-tazobactam 3.375 g IV q6h *plus* Vancomycin 1.5 g IV q12h
Febrile neutropenia	Cefepime 2 g IV q8h *plus* Vancomycin 1.5 g IV q12h	Piperacillin-tazobactam 3.375 g IV q6h *plus* Gentamicin 1.5 mg/kg q8h *or* Imipenem 0.5 g IV q6h *plus* Gentamicin 1.5 mg/kg q8h
Bacterial meningitis	Ceftriaxone 2 g IV q12h *plus* Ampicillin 3 g IV q6h *plus* Vancomycin 1.5 g IV q12h *plus* Dexamethasone 0.15 mg/kg IV q6h for 2-4 days	Gram-positive cocci: vancomycin *plus* ceftriaxone 2 g IV q12h Gram-negative diplococci: cefotaxime 2 g IV q4-6h Gram-positive bacilli: ampicillin 3 g IV q6h *plus* gentamicin Gram-negative bacilli: ceftazidime 2 g IV q8h *plus* gentamicin 1.5 mg/kg IV q8h All above *plus* dexamethasone
Cellulitis	Ciprofloxacin 400 mg IV q12h *plus* Clindamycin 900 mg IV q8h	Imipenem 0.5 g IV q6h

*Most antibiotic doses must be adjusted if there is hepatic or renal dysfunction. Some antibiotics require adjustment based on levels (e.g., gentamicin). In selecting a drug, carefully consider the patient's history of antibiotic (especially penicillin) allergy.

predicted body weight) decreases mortality from acute lung injury and acute respiratory distress syndrome (Chapter 105).

Patients who require ventilation need adequate but not excessive sedation, which can worsen hemodynamic instability, prolong ventilation, and increase the risk for development of nosocomial pneumonia. Sedation should be titrated by objective assessment. Daily interruption of sedation decreases the duration of mechanical ventilation and intensive care. Weaning from mechanical ventilation is often associated with fluid overload from prior fluid resuscitation and from the reduction in intrathoracic pressure. Patients whose weaning is guided by brain natriuretic peptide levels are weaned more quickly and have more ventilator-free days because they generally receive more aggressive diuretic therapy, without a concomitant increased need for vasopressors, an increased risk of renal dysfunction, or more electrolyte abnormalities.

Circulatory Therapy

Early therapy is the cornerstone of emergency management, but such therapy need not achieve specific central hemodynamic targets or require the placement of a central venous catheter or the administration of inotropic agents or blood transfusions. Standard therapies should have the goal of increasing tissue oxygen delivery by increasing profoundly low blood pressure, increasing inadequate blood flow, increasing low arterial oxygen saturation, and increasing mixed venous oxygen saturation. Although oxygen delivery is higher in survivors than in nonsurvivors, it is not clear that a specific oxygen delivery target is more beneficial than clinical end points. Several trials have shown that supernormal global oxygen delivery does not decrease mortality rates in sepsis and septic shock.

A mean arterial blood pressure goal of 65 to 70 mm Hg is as good as a goal of 80 to 85 mm Hg. However, a higher mean arterial pressure target (80 to 85 mm Hg) may decrease the risk of renal injury and the need for renal replacement therapy in patients with preexisting hypertension. In patients who have acute lung injury, no difference in outcomes is seen with management using a pulmonary artery catheter versus a central venous catheter. Patients whose acute lung injury is managed after 24 to 48 hours with a conservative fluid strategy (compared with a liberal fluid strategy) have significantly improved lung function and shorter duration of ventilation and ICU stay.

Fluids should be used to maintain central venous pressure at 8 to 12 mm Hg; at present, no convincing data indicate that albumin is better than normal saline solution.[5] In patients with severe sepsis, large randomized trials confirm that modified lactated Ringer solution or albumin is preferred to 10% hetastarch (a colloid) because of lower rates of acute kidney injury, less need for renal replacement therapy, and fewer deaths. As a result, hetastarch should not be used in septic shock. If central venous oxygen saturation is less than 70%, packed red cell transfusions should be used to maintain a hematocrit greater than 30%.

Vasopressors (e.g., norepinephrine, 1 to 50 µg/minute; epinephrine, 1 to 30 µg/minute) should be added if the mean arterial pressure is less than 65 mm Hg. Dobutamine (2.5 to 20 µg/kg/minute) is required if central venous pressure, mean arterial pressure, and hematocrit are optimized but the central venous oxygen saturation remains less than 70%. In a randomized trial of patients with septic shock, the combination of norepinephrine plus dobutamine resulted in a mortality similar to that with epinephrine alone, with no differences in organ dysfunction, time to resolution of shock, or adverse events. In another randomized trial, norepinephrine was slightly but not significantly better than dopamine for reducing mortality when used as the first-line vasopressor for patients with septic shock; however, norepinephrine was associated with a lower rate of arrhythmias, especially atrial fibrillation. These accumulated data suggest that norepinephrine may be preferable to dopamine as the first vasopressor in septic shock.

Clinicians can use epinephrine alone, norepinephrine alone, or norepinephrine plus dobutamine in patients with low cardiac output. As a strategic approach to persistent hypotension despite adequate fluid resuscitation, a

vasopressor such as norepinephrine (1 to 50 µg/minute) can be added first. If the cardiac index is low or if the mixed venous oxygen saturation is low (>70%) despite an adequate central venous pressure, an inotropic agent such as dobutamine should be added, initially at approximately 2 to 5 µg/kg/minute and increasing until the mixed venous oxygen saturation is adequate. In some patients in septic shock, the cardiac index is inadequate, as reflected by a low mixed venous oxygen saturation despite a high central venous pressure (>12 mm Hg) or pulmonary artery wedge pressure (>18 mm Hg) because of underlying cardiovascular dysfunction or because of acute left ventricular dysfunction resulting from sepsis. In such patients, earlier use of an inotropic agent such as dobutamine should be considered to increase left ventricular contractility. The overall goal is to achieve an adequate mean arterial pressure (>65 mm Hg), central venous pressure, and mixed venous oxygen saturation while other indices of adequate perfusion are monitored, such as hourly urine output (>0.5 mL/kg/hour), arterial lactate levels (<2 mmol/L), mental status, and skin perfusion. To assess the adequacy of early resuscitation of severe sepsis and septic shock and to guide ongoing therapy, a central venous oxygen saturation goal greater than 70% or a lactate clearance of at least 10% is an equally good measure.

Fever, which is common in septic shock, may have some beneficial effects for resisting infection but also increases oxygen demand. Reducing fever can decrease the need for vasopressors and possibly the risk of septic encephalopathy. In one large trial, external cooling for 48 hours to maintain core body temperature between 36.5° C and 37° C was safe and decreased the need for vasopressors as well as 14-day mortality rates (from 34% to 19%). Cooling to reduce fever to normothermia is especially promising for febrile septic shock patients who are receiving high doses of vasopressors, require inotropic agents, or have marked tachycardia.

Transfusion of Erythrocytes and Erythropoietin

Anemia is common in septic shock, but the optimal hemoglobin level for resuscitation has been controversial. Recently, however, a randomized trial of transfusion in critically ill patients with septic shock found that a transfusion threshold of a hemoglobin level of 7 g/dL was equivalent to a threshold of 9 g/dL in terms of ischemic events, the need for life support, or death, while reducing transfusions from a median of 4 units to a median of 1 unit.

Drugs

Antibiotics

The infected site and infecting organisms of septic shock are often not known initially, and clinicians usually need to decide on an empirical antibiotic regimen before knowing culture results in patients with septic shock. After appropriate culture specimens are obtained, intravenous broad-spectrum antibiotics should be administered on an emergency basis (within 1 hour) while considering host factors such as immune and allergic status (Chapters 280 and 281; see Fig. 108-1). Emergency, empirical antibiotic therapy (Table 108-2) should be guided by the greater frequency of gram-positive bacteria, the possibility of resistant organisms, and the local bacteriologic features. In one large trial, meropenem alone was equivalent to the combination of meropenem and moxifloxacin in terms of mortality rates and organ dysfunction. Adding empirical fluconazole treatment does not result in better outcomes in critically ill, non-neutropenic patients.

Because outcomes in patients with septic shock are worse if the organisms causing the sepsis are not sensitive to the initial antibiotic regimen, a central question is the relative value of using a single antibiotic compared with multiple antibiotics. Broad-spectrum antibiotics should be used for as short a time as possible. If a causative organism is identified (<20% of septic patients have negative cultures), the antibiotic regimen should be quickly narrowed with 3 to 5 days to decrease the emergence of resistant organisms. The duration of antibiotics should be guided by the cause of septic shock, but patients generally require 10 to 14 days of therapy.

Corticosteroids

High-dose corticosteroids do not improve outcomes in the full spectrum of patients with sepsis or acute respiratory distress syndrome, but the evidence supporting the use of low-dose corticosteroids in septic shock is controversial. In one trial, corticosteroids (hydrocortisone, 50 mg intravenously every 6 hours, plus fludrocortisone, 50-µg tablet/day per nasogastric tube or orally for 7 days) increased survival in septic patients whose serum cortisol response after stimulation with an intravenous infusion of 250 µg of corticotropin was 9 µg/dL or less. In a subsequent randomized trial of hydrocortisone (50 mg every 6 hours intravenously for 5 days) versus placebo, however, mortality was not improved, regardless of the patient's response to corticotropin stimulation. Adding fludrocortisone to hydrocortisone does not appear to be better than hydrocortisone alone, and aggressive insulin therapy to address the hyperglycemia that often accompanies corticosteroid treatment is no better than usual glucose control. Hydrocortisone treatment is often associated with a shorter duration of septic shock but can increase the risk of superinfections. Current sepsis guidelines suggest that corticosteroids should be considered only in patients who are poorly responsive to vasopressors.

Corticosteroids administered before antibiotics also decrease the neurologic sequelae of bacterial, especially pneumococcal, meningitis (Chapter 412). Enthusiasm for corticosteroid therapy must be tempered by the risk of complications such as superinfection, neuromyopathy, hyperglycemia, immune suppression, and impaired wound healing.

Recombinant Human Activated Protein C

In septic shock, a deficiency of activated protein C is nearly universal, and low levels of activated protein C are associated with increased mortality. However, recombinant human activated protein C infusion does not reduce mortality in patients who receive guideline-driven therapy for septic shock, and the drug has been removed from the worldwide market.

Vasopressin

Although vasopressin deficiency and downregulation of vasopressin receptors are common findings in septic shock, a low-dose vasopressin infusion added to norepinephrine is not significantly better than a norepinephrine infusion alone in septic shock. Whether vasopressin infusion may be beneficial in patients with less severe shock is uncertain.

Hyperglycemia and Intensive Insulin Therapy

Hyperglycemia and insulin resistance are common in septic shock, but intensive insulin therapy to control hyperglycemia is not beneficial in patients in medical ICUs because of significantly higher rates of hypoglycemia and perhaps increased mortality. Therefore, intensive insulin therapy cannot be recommended for patients with septic shock.

Renal Dysfunction and Dialysis

Acute renal failure is an important complication of septic shock because of its associated morbidity, mortality, and resource use (Fig. 108-2). In critically ill patients who have acute kidney injury, hemodialysis six times per week is no better than conventional hemodialysis, and intensive renal replacement therapy is no better than standard therapy overall or in patients who have sepsis.

Low-dose dopamine (2 to 4 µg/kg/minute) does not decrease the need for renal support, does not improve outcomes, and is not recommended. Lactic acidosis is a common complication of septic shock, but administration of sodium bicarbonate in the setting of lactic acidosis does not improve hemodynamics or the response to vasopressors.

A small randomized controlled trial evaluated the use of polymyxin B hemoperfusion in patients with abdominal sepsis to reduce blood endotoxin levels. Polymyxin B hemoperfusion increased blood pressure, decreased vasopressor requirements, improved organ dysfunction, and reduced mortality by one third. However, this intervention requires further evaluation before it can be recommended.

Other Therapies

Deep venous thrombosis prophylaxis with low-dose heparin is recommended for patients who do not have active bleeding, coagulopathy, or a contraindication to heparin (see Fig. 108-2). Low-molecular-weight heparin does not decrease risk of deep venous thrombosis compared with unfractionated heparin, but it may decrease the risk of pulmonary emboli in the critically ill. Stress ulcer prophylaxis with H_2-receptor antagonists decreases the risk of gastrointestinal hemorrhage. Proton pump inhibitors may also be effective, but they have not been fully evaluated in septic shock.

Enteral nutrition is generally safer and more effective than total parenteral nutrition, but total parenteral nutrition is sometimes required in patients with abdominal sepsis, surgery, or trauma.[6] Initial trophic feeding, which provides about 25% of normal calorie requirements, appears to be as good as full enteral feeding and is therefore recommended after stabilization of patients in septic shock. The use of sedation, neuromuscular blocking agents, and corticosteroids should be minimized because they can exacerbate septic encephalopathy and the polyneuropathy or myopathy of sepsis. Neutropenic patients may benefit from granulocyte colony-stimulating factor (Chapter 167). The risk of nosocomial infection is decreased by narrow-spectrum antibiotics, early weaning from ventilation, and periodic removal and replacement of catheters (Chapter 282).

B. Breathing—Oxygen, with a tidal volume 6 mL/kg IBW if ventilated. Wean according to ARDSNet protocol (Chapters 105 and 106)

C. Circulation
- Fluids, vasopressors, inotropes, transfusion; goals include:
  - MAP > 65 mm Hg
  - CVP 8-12 mm Hg
  - Hg 70-90 g/L
  - $Scvo_2$ > 70% (optional)

Consider pulmonary artery catheter or echocardiogram especially if known cardiovascular disease; goals include:
- Wedge pressure 8-15 mm Hg
- Cardiac index: normal or increased

D. Drugs:
- Antibiotics: Narrow spectrum to cause of infection
- Hydrocortisone (if evidence of relative adrenal insufficiency [see text]): hydrocortisone 50 mg intravenously every 6 hours and fludrocortisone 50-µg tab orally or per NG tube daily for 7 days

Other Organ Support
- Renal function: Continuous renal replacement
- DVT prophylaxis: Low-dose heparin 5000 IU subcutaneously every 12 hours
- Stress ulcer prophylaxis: H_2-receptor antagonist (e.g., ranitidine 50 mg intravenously every 8 hours)
- Nutrition: Enteral preferred
- Sedation: Intermittent with daily awakening

FIGURE 108-2. Ongoing critical care support and management in septic shock. CVP = central venous pressure; DVT = deep venous thrombosis; Hg = hemoglobin; IBW = ideal body weight; MAP = mean arterial pressure; NG = nasogastric; $Scvo_2$ = central venous oxygen saturation.

PROGNOSIS

The 28-day mortality of septic shock has decreased during the past 20 years from about 50% to about 25 to 35%,[7] especially in academic centers,[8] probably because of the earlier initiation of appropriate therapies at appropriate doses for limited periods. Early deaths (in the first 72 hours) are usually the result of refractory, progressive shock despite escalating life support. Later deaths from septic shock (after day 3) are usually secondary to multiple organ dysfunction. The number of dysfunctional organs and the progression or lack of improvement of organ dysfunction are indicators of increased risk of death. Other factors that portend a poor prognosis are increased age, underlying medical conditions, more severe illness, increased arterial lactate concentrations, and the need for high-dose vasopressors. Furthermore, a delay in achieving adequate resuscitation is associated with increased mortality.

As the number of survivors of septic shock has increased, so have the numbers with significant long-term sequelae, including cognitive dysfunction,[9] depression, and post-traumatic stress disorder. Survivors of septic shock who also had acute lung injury (Chapter 104) can have weakness, fatigue, and dyspnea on exertion after hospital discharge due to pulmonary dysfunction, neuromuscular dysfunction, or other persistent organ dysfunction. Patients who have an episode of acute kidney injury during septic shock have a significantly decreased long-term survival than patients without it.[10] Overall, the survival and quality of life after hospital discharge after septic shock remain poorer than expected for at least the next 10 years.[11,12]

Grade A References

A1. Acute Respiratory Distress Syndrome Network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med*. 2000;342:1301-1308.

A2. Mekontso Dessap A, Roche-Campo F, Kouatchet A, et al. Natriuretic peptide–driven fluid management during ventilator weaning: a randomized controlled trial. *Am J Respir Crit Care Med*. 2012;186:1256-1263.

A3. Yealy DM, Kellum JA, Huang DT, et al. A randomized trial of protocol-based care for early septic shock. *N Engl J Med*. 2014;370:1683-1693.

A4. Peake SL, Delaney A, Bailey M, et al. Goal-directed resuscitation for patients with early septic shock. *N Engl J Med*. 2014;371:1496-1506.

A5. Asfar P, Meziani F, Hamel JF, et al. High versus low blood-pressure target in patients with septic shock. *N Engl J Med*. 2014;370:1583-1593.

A6. Caironi P, Tognoni G, Masson S, et al. Albumin replacement in patients with severe sepsis or septic shock. *N Engl J Med*. 2014;370:1412-1421.

A7. Rochwerg B, Alhazzani W, Sindi A, et al. Fluid resuscitation in sepsis: a systematic review and network meta-analysis. *Ann Intern Med*. 2014;161:347-355.

A8. Annane D, Vignon P, Renault A, et al. Norepinephrine plus dobutamine versus epinephrine alone for management of septic shock: a randomized trial. *Lancet*. 2007;370:676-684.

A9. De Backer D, Biston P, Devriendt J, et al. Comparison of dopamine and norepinephrine in the treatment of septic shock. *N Engl J Med*. 2010;362:779-789.

A10. Jones AE, Shapiro NI, Trzeziak S, et al. Lactate clearance vs. central venous oxygen saturation as goals of early sepsis therapy: a randomized clinical trial. *JAMA*. 2010;303:739-746.

A11. Schortgen F, Clabault K, Katsahian S, et al. Fever control using external cooling in septic shock: a randomized controlled trial. *Am J Respir Crit Care Med*. 2012;185:1088-1095.

A12. Holst LB, Haase N, Wetterslev J, et al. Lower versus higher hemoglobin threshold for transfusion in septic shock. *N Engl J Med*. 2014;371:1381-1391.

A13. Brunkhorst FM, Oppert M, Marx G, et al. Effect of empirical treatment with moxifloxacin and meropenem vs meropenem on sepsis-related organ dysfunction in patients with severe sepsis: a randomized trial. *JAMA*. 2012;307:2390-2399.

A14. Sprung CL, Annane D, Keh D, et al. Hydrocortisone therapy for patients with septic shock. *N Engl J Med*. 2008;358:111-124.

A15. Annane D, Cariou A, Maxime V, et al. Corticosteroid treatment and intensive insulin therapy for septic shock in adults: a randomized controlled trial. *JAMA*. 2010;303:341-348.

A16. Ranieri VM, Thompson BT, Barie PS, et al. Drotrecogin alfa (activated) in adults with septic shock. *N Engl J Med*. 2012;366:2055-2064.

A17. Russell JA, Walley KR, Singer J, et al. Vasopressin versus norepinephrine infusion in patients with septic shock. *N Engl J Med*. 2008;358:877-887.

A18. Finfer S, Chittock DR, Su SY, et al. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med*. 2009;360:1283-1297.

A19. Palevsky PM, Zhang JH, O'Connor TZ, et al. Intensity of renal support in critically ill patients with acute kidney injury. *N Engl J Med*. 2008;359:7-20.

A20. Cook D, Meade M, Guyatt G, et al. Dalteparin versus unfractionated heparin in critically ill patients. *N Engl J Med*. 2011;364:1305-1314.

A21. Harvey SE, Parrott F, Harrison DA, et al. Trial of the route of early nutritional support in critically ill adults. *N Engl J Med*. 2014;371:1673-1684.

A22. Rice TW, Wheeler AP, Thompson BT, et al. Initial trophic vs full enteral feeding in patients with acute lung injury: the EDEN randomized trial. *JAMA*. 2012;307:795-803.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Angus DC, van der Poll T. Severe sepsis and septic shock. *N Engl J Med*. 2013;369:840-851.
2. Funk DJ, Jacobsohn E, Kumar A. The role of venous return in critical illness and shock—part I: physiology. *Crit Care Med*. 2013;41:255-262.
3. Vincent JL, De Backer D. Circulatory shock. *N Engl J Med*. 2013;369:1726-1734.
4. Dellinger RP, Levy MM, Rhodes A, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med*. 2013;41:580-637.
5. Myburgh JA, Mythen MG. Resuscitation fluids. *N Engl J Med*. 2013;369:1243-1251.
6. Desai SV, McClave SA, Rice TW. Nutrition in the ICU: an evidence-based approach. *Chest*. 2014;145:1148-1157.
7. Kaukonen KM, Bailey M, Suzuki S, et al. Mortality related to severe sepsis and septic shock among critically ill patients in Australia and New Zealand, 2000-2012. *JAMA*. 2014;311:1308-1316.
8. Walkey AJ, Wiener RS. Hospital case volume and outcomes among patients hospitalized with severe sepsis. *Am J Respir Crit Care Med*. 2014;189:548-555.
9. Pandharipande PP, Girard TD, Jackson JC, et al. Long-term cognitive impairment after critical illness. *N Engl J Med*. 2013;369:1306-1316.
10. Linder A, Fjell C, Levin A, et al. Small acute increases in serum creatinine are associated with decreased long-term survival in the critically ill. *Am J Respir Crit Care Med*. 2014;189:1075-1081.
11. Iwashyna TJ, Ely EW, Smith DM, et al. Long-term cognitive impairment and functional disability among survivors of severe sepsis. *JAMA*. 2010;304:1787-1794.
12. Linder A, Guh D, Boyd J, et al. Long term (10 year) mortality of younger previously healthy severe sepsis patients is worse than nonseptic patients and general population. *Crit Care Med*. 2014;42:2211-2218.

REVIEW QUESTIONS

1. Which of the following statements about dopamine in septic shock is correct?

A. Low-dose dopamine is effective in decreasing the risk of acute kidney injury.
B. Dopamine infusion was associated with increased cardiovascular adverse effects compared with norepinephrine in a large mortality-powered randomized controlled trial.
C. Dopamine increases mean arterial pressure excessively compared with norepinephrine.
D. Dopamine alters dopaminergic receptor neurologic functions and alters mental status in septic shock.
E. Randomized controlled trial evidence shows that dopamine infusion shortens the duration of vasopressor support compared with norepinephrine.

Answer: B Dopamine infusion is associated with increased cardiovascular adverse effects compared with norepinephrine (De Backer D, Biston P, Devriendt J, et al. Comparison of dopamine and norepinephrine in the treatment of septic shock. *N Engl J Med*. 2010;362:779-789; Patel GP, Grahe JS, Sperry M, et al. Efficacy and safety of dopamine versus norepinephrine in the management of septic shock. *Shock*. 2010;33:375-380.). Low-dose dopamine does not decrease the risk of acute kidney injury (Bellomo R, Chapman M, Finfer S, et al. Australian and New Zealand Intensive Care Society Clinical Trials Group. Low dose dopamine in patients with early renal dysfunction: a placebo controlled randomised trial. *Lancet*. 2000;356:2139-2143.) and is not recommended in the surviving sepsis guidelines (Dellinger RP, Levy MM, Rhodes A, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med*. 2013;41:580-637.) for this reason. There is no convincing evidence that dopamine increases mean arterial pressure excessively compared with norepinephrine; both are given as continuous infusions that can be effectively titrated to a target mean arterial pressure. Although there are dopaminergic neurons, there is no evidence that dopamine infusion alters dopaminergic receptor neurologic functions or alters mental status in septic shock. There is no evidence that dopamine shortens the duration of vasopressor support compared with norepinephrine (De Backer D, Biston P, Devriendt J, et al. Comparison of dopamine and norepinephrine in the treatment of septic shock. *N Engl J Med*. 2010;362:779-789.).

2. Septic shock has not been shown to be associated with which of the following complications during post-hospitalization recovery?

A. Cognitive dysfunction
B. Decreased long-term survival
C. Neuromuscular dysfunction
D. Impaired quality of life
E. Recurrent seizure disorder

Answer: E Cognitive function, long-term survival, neuromuscular function, and quality of life are all reduced among survivors of severe sepsis. However, survivors of septic shock do not have increased risk of recurrent seizures. (Pandharipande PP, Girard TD, Jackson JC, et al. Long-term cognitive impairment after critical illness. *N Engl J Med*. 2013;369:1306-1316. Herridge MS, Tansey CM, Matte A, et al. Functional disability 5 years after acute respiratory distress syndrome. *N Engl J Med*. 2011;364:1293-1304. Iwashyna TJ, Ely EW, Smith DM, et al. Long-term cognitive impairment and functional disability among survivors of severe sepsis. *JAMA*. 2010;304:1787-1794.)

109

DISORDERS DUE TO HEAT AND COLD

MICHAEL N. SAWKA AND FRANCIS G. O'CONNOR

TEMPERATURE REGULATION

Body temperature is regulated through two parallel processes that modify body heat balance: behavioral (clothing, shelter, physical activity) and physiologic (skin blood flow, sweating, shivering). Both peripheral (skin) and central (core) thermal receptors provide afferent input to a central nervous system integrator (hypothalamic thermoregulatory center), and any deviation between the controlled variable (body temperature) and a theoretical reference variable ("set point" temperature) results in a heat loss or conservation response (E-Fig. 109-1).

Humans normally regulate body (core) temperature at about 37° C (98.6° F), and fluctuations within the narrow range of 35° C to 41° C (95° F to 105.8° F) can be tolerated by healthy acclimatized persons; core temperatures outside this range can induce morbidity and mortality. There is no single core temperature because it varies at different deep body sites and during rest and physical exercise. Arterial blood temperature, which provides the best invasive measurement of core temperature, is slightly lower than brain temperature. The most accurate noninvasive index of core temperature is esophageal temperature, followed in order of preference by rectal, gastrointestinal tract (telemetry pill), and oral temperature. Ear (tympanic and auditory meatus) or scanned temporal artery temperature should not be relied on for clinical judgment. Rectal temperatures are most commonly recommended because they are easy to measure and are not biased by environmental conditions.

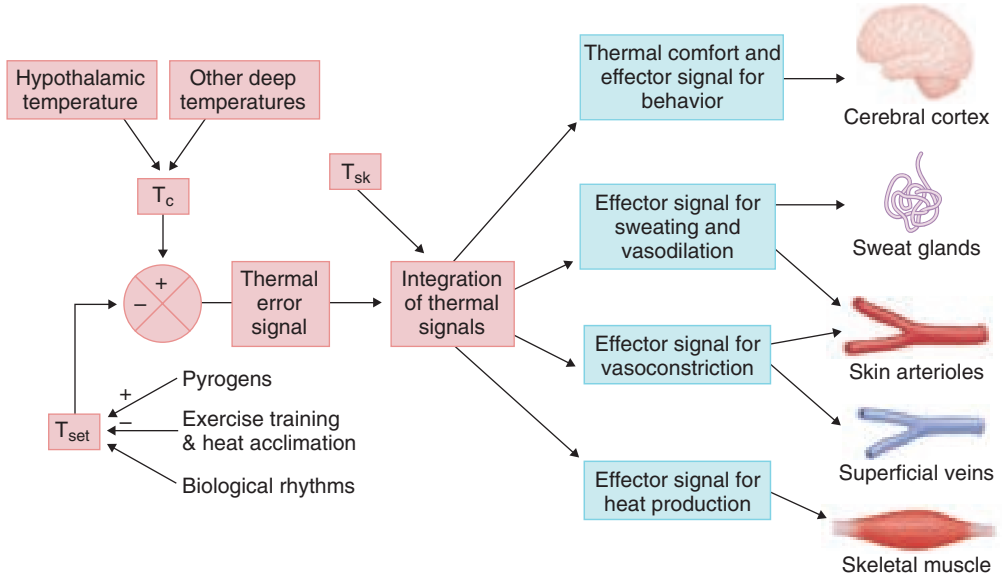

E-FIGURE 109-1. Schematic diagram of human thermoregulatory control system. T_c = core temperature; T_{set} = set point temperature; T_{sk} = skin temperature. (From Sawka MN, Leon LR, Montain SJ, et al. Integrated physiological mechanisms of exercise performance, adaptation, and maladaptation to heat stress. *Compr Physiol.* 2011;1:1883-1928.)

HEAT ILLNESS

DEFINITION

Minor heat-related illnesses include miliaria rubra, heat syncope, and heat cramps. Serious heat illness represents a continuum from heat exhaustion to heat injury and heatstroke.

EPIDEMIOLOGY

Heat illness accounts for considerable morbidity and mortality in the world today. Serious heat illness is associated with a variety of individual factors, health conditions, drugs, and environmental factors (Table 109-1). Exertional heat illness is among the leading causes of death in young athletes,[1] and its incidence appears to be increasing in the United States. Classic heat illness caused by high environmental temperatures remains a problem especially in homebound elderly persons without air conditioners.[2] Anticholinergic and sympathomimetic poisoning (Chapter 110) can induce hyperthermia. Malignant hyperthermia (Chapter 432) is a rare disorder occurring in genetically predisposed individuals; rapid and massive skeletal muscle contraction from exposure to certain anesthetic agents, most commonly halothane and succinylcholine, can trigger core temperature elevations well above 43° C (110° F). Neuroleptic malignant syndrome (Chapter 434) is an idiosyncratic hyperthermic reaction caused by skeletal muscle rigidity from treatment with neuroleptic medications (e.g., antipsychotics, antidepressants, antiemetics). Both malignant hyperthermia and neuroleptic malignant syndrome are potentially fatal without prompt recognition and early intervention.

TABLE 109-1 FACTORS PREDISPOSING TO SERIOUS HEAT ILLNESS

INDIVIDUAL FACTORS
Lack of acclimatization
Low physical fitness
Excessive body weight
Dehydration
Advanced age
Young age
HEALTH CONDITIONS
Inflammation and fever
Viral infection
Cardiovascular disease
Diabetes mellitus
Gastroenteritis
Rash, sunburn, and previous burns to large areas of skin
Seizures
Thyroid storm
Neuroleptic malignant syndrome
Malignant hyperthermia
Sickle cell trait
Cystic fibrosis
Spinal cord injury
DRUGS
Anticholinergic properties (atropine)
Antiepileptic (topiramate)
Antihistamines
Glutethimide (Doriden)
Phenothiazines
Tricyclic antidepressants
Amphetamines, cocaine, Ecstasy
Ergogenic stimulants (e.g., ephedrine, ephedra)
Lithium
Diuretics
β-Blockers
Ethanol
ENVIRONMENTAL FACTORS
High temperature
High humidity
Little air motion
Lack of shade
Heat wave
Physical exercise
Heavy clothing
Air pollution (nitrogen dioxide)

Heat illness can also occur in low-risk individuals who have taken appropriate precautions relative to situations to which they have been exposed in the past. Historically, such unexpected cases were attributed to dehydration (which impairs thermoregulation and increases hyperthermia and cardiovascular strain), but it is now suspected that a previous heat exposure or a concurrent event (e.g., sickness or injury) might make these individuals more susceptible to serious heat illness. One theory is that previous heat injury or illness primes the acute phase response and augments the hyperthermia of exercise, inducing unexpected serious heat illness. Another theory is that previous infection produces proinflammatory cytokines that deactivate the cells' ability to protect against heat shock.

PATHOBIOLOGY

Body temperature can increase from a number of mechanisms: exposure to environmental heat (impeded heat dissipation); physical exercise (increased heat production); fever from systemic illness (elevated set point with subsequent activation of shivering); and medications (neuroleptic malignant syndrome and malignant hyperthermia). In addition, febrile persons have accentuated elevations in core temperature when they are exposed to high ambient temperature, physical exercise, or both. Environmental temperature and humidity, medications, and exercise heat stress in turn challenge the cardiovascular system to provide high blood flow to the skin, where blood pools in warm, compliant vessels such as those found in the extremities. When blood flow is diverted to the skin, reduced perfusion of the intestines and other viscera can result in ischemia, endotoxemia, and oxidative stress (E-Fig. 109-2). In addition, excessively high tissue temperatures (heat shock: >41° C [105.8° F]) can produce direct tissue injury; the magnitude and duration of the heat shock influence whether cells respond by adaptation (acquired thermal tolerance), injury, or death (apoptotic or necrotic). Heat shock, ischemia, and systemic inflammatory responses can result in cellular dysfunction, disseminated intravascular coagulation, and multiorgan dysfunction syndrome. In addition, reduced cerebral blood flow, combined with abnormal local metabolism and coagulopathy, can lead to dysfunction of the central nervous system.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Minor heat illness is common and can be recognized by its clinical features. Miliaria rubra (heat rash) results from the occlusion of eccrine sweat gland ducts and can be complicated by secondary staphylococcal infection. Heat syncope (fainting) is caused by temporary circulatory insufficiency as a result of blood pooling in the peripheral veins, especially the cutaneous and lower extremity veins. Skeletal muscle cramps most commonly occur during and after intense exercise and are probably related to dehydration, loss of sodium or potassium, and neurogenic fatigue rather than to overheating itself.

Serious heat illness includes heat exhaustion, which can progress to heat injury, which then can progress to heatstroke. In many patients, the degree of severity of heat illness often is not initially clear. Patients who exhibit symptoms (e.g., dizziness, unsteady gait, ataxia, headache, confusion, weakness, fatigue, nausea, vomiting, diarrhea) should have an immediate assessment of their mental status, core (rectal) temperature, and other vital signs. Until it is proved otherwise, heatstroke should be the initial working diagnosis in anyone who is a heat casualty and has an altered mental status.

Heat exhaustion is defined as a syndrome of hyperthermia (temperature at time of event usually ≤40° C or 104° F) and debilitation that occur during or immediately after exertion in the heat, accompanied by no more than minor central nervous system dysfunction (headache, dizziness), which resolves rapidly with intervention. It is primarily a cardiovascular event (insufficient cardiac output) frequently accompanied by sweaty hot skin, dehydration, and collapse.

Heat injury is a moderate to severe illness characterized by evidence of damage to organs (e.g., liver, renal, gut) and tissues (e.g., rhabdomyolysis) without sufficient neurologic symptoms to be diagnosed as heat stroke. It is usually associated with body temperatures above 40° C (104° F).

Heatstroke is a severe illness characterized by profound mental status changes with high body temperatures, usually but not always higher than 40° C (104° F). However, patients with a core temperature higher than 40° C do not universally have a heat injury or heatstroke, and core temperatures this high can be seen transiently after stressful exercise in the heat. To establish the diagnosis of heatstroke, the entire clinical picture, including mental status and laboratory results, must be considered. Heatstroke is often categorized as classic or exertional; classic heatstroke is observed primarily in otherwise sick and compromised individuals, and exertional heatstroke is observed

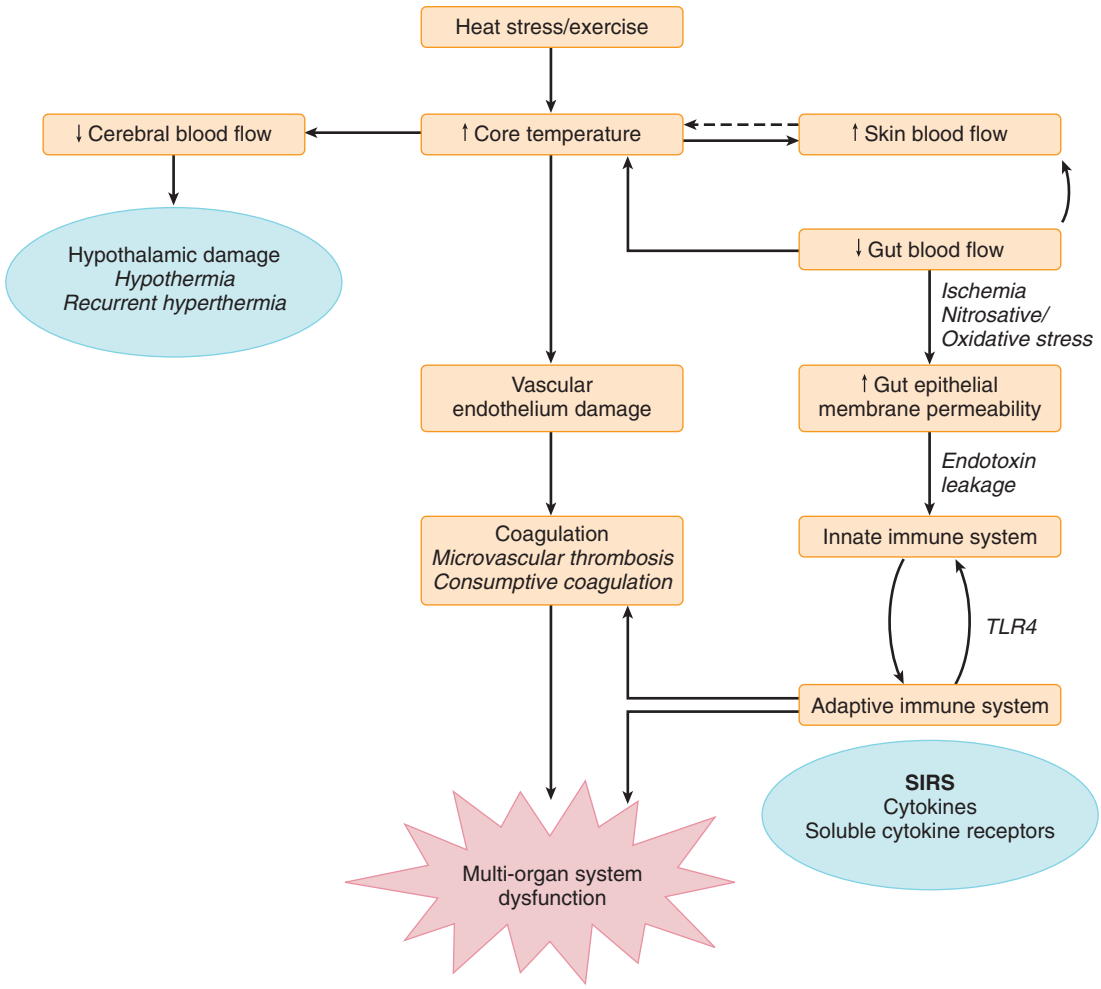

E-FIGURE 109-2. Summary of the pathophysiologic process of heat stroke that culminates in multiorgan system dysfunction and death. SIRS = systemic inflammatory response syndrome; TLR4 = toll-like receptor 4. (From Leon LR, Helwig BG. Heat stroke: role of the systemic inflammatory response. *J Appl Physiol (1985).* 2010;109:1980-1988.)

TABLE 109-2 COMPARISON OF CLASSIC AND EXERTIONAL HEATSTROKE

PATIENT CHARACTERISTICS	CLASSIC	EXERTIONAL
Age	Young children or elderly	15-55 years
Health	Chronic illness	Usually healthy
Fever	Unusual	Common
Prevailing weather	Frequent in heat waves	Variable
Activity	Sedentary	Strenuous exercise
Drug use	Diuretics, antidepressants, anticholinergics, phenothiazines	Ergogenic stimulants or cocaine
Sweating	Often absent	Common
Acid-base disturbances	Respiratory alkalosis	Lactic acidosis
Acute renal failure	Uncommon	Common (≈15%)
Rhabdomyolysis	Uncommon	Common (≈25%)
CK	Mildly elevated	Markedly elevated (500-1000 U/L)
ALT, AST	Mildly elevated	Markedly elevated
Hyperkalemia	Uncommon	Common
Hypocalcemia	Uncommon	Common
DIC	Mild	Marked
Hypoglycemia	Uncommon	Common

ALT = alanine aminotransferase; AST = aspartate aminotransferase; CK = creatine kinase; DIC = disseminated intravascular coagulation.

primarily in apparently healthy and physically fit individuals during or after vigorous exercise (Table 109-2). In heatstroke, neuropsychiatric impairments (e.g., marked confusion, disorientation, combativeness, and seizures) develop early and universally[3] but are readily reversible with early cooling. In addition, heatstroke can be complicated by liver damage, rhabdomyolysis, disseminated intravascular coagulation, water and electrolyte imbalance, and renal failure. In fulminant heat stroke, patients have the full spectrum of abnormalities associated with the systemic inflammatory response syndrome (Chapter 108).

PREVENTION AND TREATMENT

Rx

Heat illness can be prevented by heat acclimatization and acquired thermal tolerance, maintenance of adequate hydration, and avoidance of overwhelming heat exposure. Adequate fluid intake is critical, and oral rehydration solutions may be preferable to other forms of hydration for patients with heat exhaustion.

Management of serious heat illness, which should begin in the field setting, includes cooling,[4] rehydration, and monitoring (Table 109-3). The first priority should be immediately to initiate whole body cooling and to continue cooling until the core temperature falls below 38.8° C (102° F). Body cooling lowers tissue temperatures, thereby facilitating conduction and convection from the core to the shell, and reduces cardiovascular stress by causing arterial and venous constriction that redirects blood back to the heart. Immersion or soaking of the skin in cool or ice water with skin massage is the most effective method, but other effective methods include soaking of the skin followed by accelerated evaporation with fans or the use of ice sheets and ice packs. These noninvasive treatments can be supplemented with the infusion of chilled (≈5° C) normal saline. Cooling can induce shivering, which is usually not sufficient to increase body temperature, so shivering need not be treated.

In the hospital, the priority for patient care remains urgent cooling. Patients who are unconscious are at risk of poor airway control and may require endotracheal intubation to prevent aspiration. Fluid and electrolyte deficits should be corrected; restoration of plasma volume with isotonic fluids (e.g., normal saline) sufficient to sustain adequate perfusion, as judged by carefully monitored urine output, is also a priority. Rapid overcorrection of serum electrolytes (e.g., sodium) should be avoided. If rhabdomyolysis (Chapter 113) and myoglobinuria are present, maintaining urine flow helps minimize renal injury.

For exercise-induced and environmental heat illness, no pharmacologic interventions have been proved to augment cooling. For patients with malignant hyperthermia, however, dantrolene should be administered as a loading bolus of 2.5 mg/kg intravenously, with subsequent bolus doses of 1 mg/kg intravenously until the signs have abated.[5]

Patients should be carefully monitored to detect possible metabolic abnormalities (e.g., hyperkalemia), renal or hepatic failure, disseminated intravascular coagulation, cardiac arrhythmias, and acute respiratory failure. Medications to be avoided include antipyretics and sedatives with hepatic toxicities. Lorazepam (1 to 2 mg administered intravenously during a 2- to 5-minute period, repeated if necessary) is a safe sedative because of its low hepatotoxicity and rapid metabolism and may be indicated in patients who are combative or exhibit seizure activity.

TABLE 109-3 MANAGEMENT OF HEAT ILLNESS

HEAT EXHAUSTION
Rest and shade
Loosen and remove clothing
Supine position and elevate legs
Actively cool skin
Fluids by mouth
Monitor core temperature
Monitor mental status
HYPERTHERMIA
Protect the airway
Insert at least two large-bore intravenous lines
Monitor core temperature; options include rectal, pulmonary artery, esophageal probe
Actively cool the skin until core temperature reaches <39° C (<102.2° F)
Ice baths or cool water (≈22° C [71.6° F]) immersion
Wetting with water (avoid alcohol rubs)
Continuous fanning
Exposure to cool environment
Axillary or perineal ice packs and ice sheets
Infusion of room-temperature saline
Gastric or colonic iced saline lavage
Peritoneal lavage with cool saline
Monitor electrocardiogram for arrhythmia
Obtain serial diagnostic studies*

*Electrocardiogram, chest radiograph, complete blood count with differential, platelet count, urinalysis, aminotransferases, alkaline phosphatase, bilirubin, creatine kinase, blood urea nitrogen, creatinine, phosphate, calcium, glucose, electrolytes, uric acid, prothrombin time and partial thromboplastin time, fibrin split products, fibrinogen, arterial blood gases, toxicology screen.

PROGNOSIS

A single episode of heat exhaustion does not imply a predisposition to heat illness, and most patients recover within several hours after cooling and rehydration. For patients who present to a hospital with heatstroke, however, mortality rates can range from 21 to 63%. Mortality in both classic and exertional heat stroke correlates directly with the magnitude and duration of temperature elevation, the delay in time to initiation of cooling, and the number of organ systems affected.

Patients who have suffered heat injury or heatstroke should not be reexposed to heat until recovery is complete, which can be many weeks or months, and about 10% of heatstroke patients remain intolerant of heat. The long-term consequences of heatstroke likely include sustained organ damage, which presumably explains why such patients have a higher long-term mortality from cardiovascular, liver, and digestive diseases.

COLD INJURY

DEFINITION

Cold injuries are classified as hypothermia and peripheral cold injuries. Hypothermia is whole body cooling, whereas peripheral cold injuries are localized to the extremities and exposed skin. Hypothermia is further divided into three categories: mild (≈33° C to ≈35° C), moderate (≈27° C to ≈32° C), and profound (<27° C). Peripheral cold injuries can be divided into nonfreezing (chilblain, trench foot) and freezing (frostbite). Both hypothermia and peripheral cold injuries often occur simultaneously, and treatment priority should be given to rewarming in moderate and profound hypothermia.

EPIDEMIOLOGY

A variety of individual factors, health conditions, medications, and environmental factors are associated with a predisposition to cold injury

TABLE 109-4 FACTORS PREDISPOSING TO COLD INJURY

INDIVIDUAL FACTORS
Inadequate clothing and shelter
Lean and low body fat
Low physical fitness
Advanced age
Young age
Black race (men and women)
HEALTH CONDITIONS
Burns
Diabetes mellitus
Hypoglycemia
Neurologic lesions
Dementia
Hypoadrenalism, hypopituitarism, hypothyroidism
Prior frostbite or trench foot
Raynaud phenomenon
Sickle cell trait
Trauma
Spinal cord injury
DRUGS
Alcohol
Anesthetics
Antidepressants
Antithyroid agents
Sedatives and narcotics
ENVIRONMENTAL FACTORS
Cold temperatures
High air motion
Rain and immersion
Skin contact with metal and fuels
Repeated cold exposure
Physical fatigue
Immobility
High-altitude and low-oxygen-tension environments

(Table 109-4). In trauma patients (Chapter 111), hypothermia is associated with increased morbidity and mortality.

PATHOBIOLOGY

Cold exposure elicits peripheral vasoconstriction to reduce heat transfer between the body's core and shell (skin, subcutaneous fat). If sufficiently cold, the underlying tissues (e.g., muscle) constrict to thicken the isolative shell while reducing the body's core area. This vasoconstrictor response defends core temperature but at the expense of declining peripheral tissue temperatures, which contribute to peripheral cold injuries. Hypothermia depresses enzymatic activity, interferes with physiologic functions (e.g., clotting, respiration, cardiac conduction and rhythm), impairs the expression of cytokines, and can induce cellular injury and death.

The pathophysiologic mechanism of frostbite includes four overlapping pathologic phases: prefreeze, freeze-thaw, vascular stasis, and late ischemic.[6] The prefreeze phase consists of tissue cooling with accompanying vasoconstriction and ischemia but does not involve actual ice crystal formation. In the freeze-thaw phase, intracellular ice crystals form, thereby causing protein and lipid derangement, cellular electrolyte shifts, cellular dehydration, cell membrane lysis, and subsequent cell death. In the vascular stasis phase, vessels may fluctuate between constriction and dilation; blood may leak from vessels or coagulate within them. The late ischemic phase results from progressive tissue ischemia and infarction due to a cascade of inflammatory cytokines and prostaglandins, intermittent vasoconstriction with continued thrombus formation, and secondary reperfusion injury.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Hypothermia is a core temperature below 35° C (95° F), and clinical manifestations are related to the core temperature achieved (Table 109-5). The classic J wave on the electrocardiogram (Fig. 109-1) appears at a core temperature below about 33.8° C (93° F).

Chilblain (Chapter 80) appears as localized inflammatory lesions of the skin, most often involving the dorsal surface of fingers but also involving the ears, face, and exposed shins. Trench foot is caused by prolonged cold, wet exposure (e.g., wet socks or gloves), which can cause skin breakdown and nerve damage. Trench foot is often accompanied by infection and increased sensitivity to pain.

TABLE 109-5 HYPOTHERMIA: STAGES AND ASSOCIATED CLINICAL MANIFESTATIONS

STAGE	CORE TEMPERATURE °F	CORE TEMPERATURE °C	CLINICAL MANIFESTATIONS
Normothermia	98.6	37.0	
Mild hypothermia	95.0	35.0	Cold diuresis, maximal shivering
	93.0	33.8	Ataxia, poor judgment, J wave
	91.0	32.7	Amnesia, blood pressure difficult to measure
Moderate hypothermia	89.0	31.6	Stupor, pupils dilated
	87.0	30.5	Shivering ceases
	85.0	30.0	Cardiac arrhythmias, insulin inactive
	82.0	27.8	Unconsciousness, ventricular fibrillation likely
	80.0	26.6	No muscle reflexes
Profound hypothermia	78.0	25.5	Acid-base disturbances, no response to pain
	75.0	23.8	Pulmonary edema, hypotension
	73.0	22.7	No corneal reflexes
	66.0	18.8	Heart standstill
	62.0	16.6	Isoelectric electrocardiogram
	57.6	14.2	Lowest infant survival from accidental hypothermia
	48.2	9.0	Lowest adult survival from accidental hypothermia

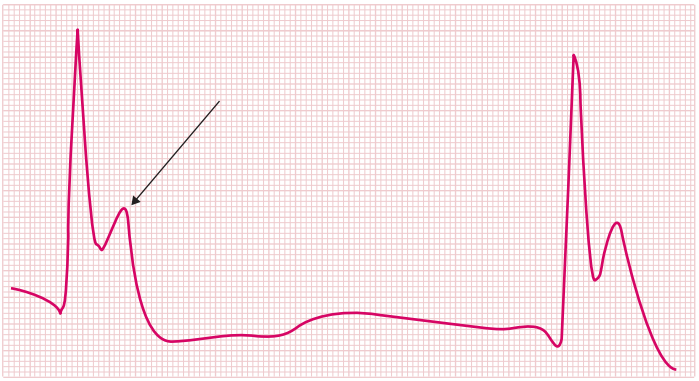

FIGURE 109-1. J (Osborne) wave.

Frostbite, which is actual freezing of tissues, has traditionally been categorized as first degree (superficial, "frostnip"), second degree (full skin), third degree (subcutaneous tissue), and fourth degree (extensive tissue and bone). For patients with frostbite, early surgical consultation is advised.

PREVENTION AND TREATMENT

Humans demonstrate minimal cold acclimatization, so prevention depends primarily on avoiding cold exposure and having adequate protection and calorie intake to support metabolism. Management of hypothermia depends on the core temperature (Table 109-6). Patients' wet clothing should be removed, and they should be provided with dry insulation. Shivering is an effective physiologic rewarming mechanism and should not be pharmacologically suppressed.

Moderately and profoundly hypothermic patients require active rewarming. Rewarming of the hypothermic patient includes both passive (insulation of the patient to prevent further heat loss) and active core (e.g., warmed saline and humidified oxygen) and external (e.g., warmed water bottles, electric blankets) techniques. Rewarming at a rate of 0.5° C to 1.0° C (0.9° F to 1.8° F) per hour is acceptable in most cases, except that aggressive rewarming is warranted in patients with significant trauma (because coagulation is hindered by hypothermia) or cardiac arrest.

Complications commonly associated with rewarming of the hypothermic individual include both afterdrop (reduction of core temperature by cold blood returning to the circulation from the periphery) and aftershock (hypotension caused by peripheral vasodilation). Another potential complication of

TABLE 109-6 TREATMENT OF HYPOTHERMIA

STAGE	MANAGEMENT	BODY REWARMING
Mild hypothermia	Monitor vital signs Warm intravenous saline Oxygen Monitor electrocardiogram for arrhythmia	Insulate Shivering Warm bath Active warming blanket
Moderate hypothermia	Diagnostic studies* Intensive care Anticipate infection and multiorgan dysfunction	Prevent extra heat loss by supplementing with airway rewarming Colonic irrigation Peritoneal dialysis
Profound hypothermia	Diagnostic studies*	Central rewarming

*See Table 109-3. Also lactate dehydrogenase, serum lactate, cortisol, thyroid-stimulating hormone, T_3, and T_4.

rewarming is ventricular fibrillation, which is more difficult to treat in the presence of moderate or profound hypothermia. If ventricular tachycardia or ventricular fibrillation develops, defibrillation should be attempted (Chapter 63). If ventricular tachycardia or ventricular fibrillation persists after a single shock, additional defibrillation attempts should be made, concurrent with rewarming but without waiting for the patient to warm to a particular target body temperature. Standard resuscitation approaches generally should be followed (Chapter 63).

Body cooling induces cold diuresis, so plasma volume must be reestablished to support adequate perfusion. Patients should receive an intravenous infusion of 250 to 1000 mL of heated (40° C to 42° C [104° F to 108° F]) 5% dextrose in normal saline. Lactated Ringer solution should be avoided because the liver cannot metabolize lactate efficiently during hypothermia. Patients should be monitored for disturbances in potassium and glucose. If hypoglycemia, alcohol, or opiate intoxication is contributing to hypothermia, intravenous glucose (50 to 100 mL of 50% dextrose), thiamine[7] (100 mg), or naloxone (1 to 2 mg), respectively, may be indicated.

Frostbitten tissues should be protected from friction or trauma but should not be thawed until there is confidence in the ability to maintain warmth because refreezing causes more injury. Gentle rewarming in a water bath (38° C to 43° C [100° F to 108° F]) is recommended. Ibuprofen should be started in the field at a dose of 12 mg/kg per day (divided twice daily) to inhibit harmful prostaglandins and increased up to a maximum dose of 2400 mg/day (divided four times daily) if the patient is experiencing pain. Common practice for blister care is selective draining of clear blisters while leaving hemorrhagic blisters intact. Case reports suggest that intravenous or intra-arterial thrombolysis within 24 hours of injury at experienced centers may mitigate the morbidity of frostbite injury.[8]

PROGNOSIS

Although noninvasive imaging with technetium pyrophosphate or magnetic resonance imaging can often predict the likelihood of tissue viability, it may take weeks to determine the precise demarcation of tissue that will require amputation. As morbidity may result from premature or unnecessary surgical intervention, consultation with a surgeon with experience evaluating and treating frostbite should be obtained to assess the need for and the timing of any amputations.

Hypothermic Syndromes

Exercise-induced bronchospasm (Chapter 87) can be triggered by exercise in cold air, particularly in patients with asthma. Livedo reticularis is patchy mottling of the limbs with cold exposure. Cryoglobulinemia (Chapter 187) occurs when immunoglobulins (IgM, IgG) reversibly precipitate after being cooled and contribute to impaired capillary blood flow in hypothermic tissues. Cold urticaria (Chapters 252 and 440) is the development of localized and general erythema and wheals in skin exposed to cold. Paroxysmal hypothermia is periodic lowering of the thermoregulatory set point and is often associated with hypothalamic abnormalities. Raynaud phenomenon (see Fig. 80-7) is intense vasoconstriction with sensitivity to pain in limbs exposed to cold.

Trauma Hypothermia

In trauma patients (Chapter 111), unintended hypothermia (<34° C [93° F]) is associated with increased morbidity and mortality due to impaired coagulation, peripheral vasoconstriction, respiratory depression, and increased risk for cardiac arrhythmias. Shivering aggravates perfusion problems by requiring blood flow to support increased metabolism in contracting muscles. Trauma patients become hypothermic because of heat loss from exposed cavities, environmental exposure, infusion of cool fluids, and ischemia, which depletes cell energy stores. Body temperature should be measured, and appropriate actions should be taken to restore normothermia during the early treatment of trauma patients.

THERAPEUTIC HYPOTHERMIA AND HYPERTHERMIA

Therapeutic Hypothermia

Therapeutic hypothermia theoretically can provide benefits by suppressing metabolism, free radical production, lipid peroxidation, inflammatory products, excitatory amino acid release, and calcium release. Among children with neonatal encephalopathy, whole body hypothermia reduces death and the combined end point of death or an IQ score of less than 70 at 6 to 7 years of age by 15%.[A2] It is being tested for spinal cord injury but has not consistently shown a benefit in children with traumatic brain injury.

For post–cardiac arrest adults, therapeutic hypothermia improves survival and neurologic outcomes.[A3] It should be initiated as soon as possible by skin cooling (e.g., cooling packs to the axilla, groin, head, and neck while simultaneously treating the patient with a cooling blanket, water mattress, or fan), endovascular cooling, or both. Endovascular cooling can be achieved either by the infusion of cool fluids (e.g., 30 mL/kg crystalloids at 4° C [40° F]) or by indwelling heat transfer devices. Cooling by peritoneal and pleural lavage is not generally used. Endovascular cooling provides a more rapid and better-controlled cooling than skin cooling. Thermoregulatory responses (shivering and peripheral vasoconstriction) will resist induced hypothermia and elevate blood pressure and should therefore be pharmacologically blunted. Low-dose meperidine (e.g., 12.5 mg intravenously) can blunt shivering without excessive toxicity. By comparison, mild pre-hospital hypothermia to a temperature of 34° C does not improve survival or neurologic outcomes,[A4] and a target temperature of 33° C is no better than a target temperature of 36° C among unconscious adults after out-of-hospital cardiac arrest.[A5]

Therapeutic Hyperthermia

Hyperthermia treatment (whole body or regional) is an experimental technique used as an adjunct to chemotherapy or radiation therapy in patients with advanced cancer. Hyperthermia (40° C to 43° C [104° F to 109° F]) alone can damage or kill cancer cells, but more important, hyperthermia might potentiate the effectiveness of chemotherapy and radiation by softening the tumor tissue, thus reducing its interstitial pressure.

Externally applied radiant heat, microwaves, or extracorporeal circulation usually induces whole body hyperthermia. Target temperatures are usually achieved during 1 to 2 hours and then maintained for approximately 1 hour, followed by a 1-hour cooling phase. Patients are usually sedated while core and skin temperatures are monitored. In regional hyperthermia, the part of the body where the tumor is located is heated while being perfused or bathed by a warmed solution containing anticancer drugs. The potential benefit of regional hyperthermia as an adjunct treatment of advanced prostate cancers is currently under investigation.[9]

Grade A References

A1. Ishikawa T, Tamura H, Ishiguro H, et al. Effect of oral rehydration solution on fatigue during outdoor work in a hot environment: a randomized crossover study. *J Occup Health*. 2010;52:209-215.

A2. Shankaran S, Pappas A, McDonald SA, et al. Childhood outcomes after hypothermia for neonatal encephalopathy. *N Engl J Med*. 2012;366:2085-2092.

A3. Arrich J, Holzer M, Havel C, et al. Hypothermia for neuroprotection in adults after cardiopulmonary resuscitation. *Cochrane Database Syst Rev*. 2012;9:CD004128.

A4. Kim F, Nichol G, Maynard C, et al. Effect of prehospital induction of mild hypothermia on survival and neurological status among adults with cardiac arrest: a randomized clinical trial. *JAMA*. 2014;311:45-52.

A5. Nielsen N, Wetterslev J, Cronberg T, et al. Targeted temperature management at 33°C versus 36°C after cardiac arrest. *N Engl J Med*. 2013;369:2197-2206.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Boden BP, Breit I, Beachler JA, et al. Fatalities in high school and college football players. *Am J Sports Med*. 2013;41:1108-1116.
2. Heat illness and deaths—New York City, 2000-2011. *MMWR Morb Mortal Wkly Rep*. 2013;62:617-621.
3. Gomez CR. Disorders of body temperature. *Handb Clin Neurol*. 2014;120:947-957.
4. DeGroot DW, Gallimore RP, Thompson SM, et al. Extremity cooling for heat stress mitigation in military and occupational settings. *J Thermal Biol*. 2013;38:305-310.
5. Malignant Hyperthermia Association of the United States. http://www.mhaus.org/; Accessed January 31, 2015.
6. McIntosh SE, Hamonko M, Freer L, et al. Wilderness Medical Society practice guidelines for the prevention and treatment of frostbite. *Wilderness Environ Med*. 2011;22:156-166.
7. Handford C, Buxton P, Russell K, et al. Frostbite: a practical approach to hospital management. *Extrem Physiol Med*. 2014;3:7.
8. Grieve AW, Davis P, Dhillon S, et al. A clinical review of the management of frostbite. *J R Army Med Corps*. 2011;157:73-78.
9. Hurwitz MD, Hansen JL, Prokopios-Davos S, et al. Hyperthermia combined with radiation for the treatment of locally advanced prostate cancer: long-term results from Dana-Farber Cancer Institute study 94-153. *Cancer*. 2011;117:510-516.

REVIEW QUESTIONS

1. Frostbite is a limb-threatening injury that requires early intervention for limb preservation. Which of the following interventions is not currently routinely recommended for the management of deep frostbite?

A. Update tetanus status as required
B. Prophylactic coverage with systemic antibiotics
C. Early and rapid rewarming of frozen tissue
D. Administration of low-dose nonsteroidal anti-inflammatory drugs until wound healing is complete
E. None of the above

Answer: B Frostbite is a limb-threatening injury that requires urgent intervention. Recently developed guidelines no longer include routine prophylactic antibiotics. Antibiotic therapy, however, is strongly recommended at the earliest signs of infection. Rewarming remains the cornerstone of therapy for frostbite. Updating tetanus status and using nonsteroidal anti-inflammatory drugs to mitigate inflammation also are recommended.

2. The management of hypothermia requires attention to detail, including careful patient handling and judicious hydration, to avoid triggering of potential fatal arrhythmias. If a patient with hypothermia develops a life-threatening arrhythmia, which treatment approach best describes current guidance on defibrillation?

A. Ventricular fibrillation in the hypothermic patient is unresponsive to defibrillation, so the patient should be treated with chest compressions until normal core body temperature is achieved.
B. Defibrillation equipment is unreliable in a hypothermic patient.
C. Antiarrhythmic medications should be used aggressively until normal body temperature is achieved, and if there is no response, continue cardiopulmonary resuscitation until return of a normal core temperature.
D. For the hypothermic patient with a temperature lower than 30° C, a single defibrillation attempt is warranted. If this attempt fails, proceed with basic life support guidelines with defibrillation attempts as required while pursuing active rewarming.

Answer: D The success rate of defibrillation and of life-saving medications in a cardiac arrest scenario is mitigated in the presence of significant hypothermia. If ventricular tachycardia (VT) or ventricular fibrillation (VF) is present, defibrillation should be attempted. If VT or VF persists after a single shock, the value of subsequent defibrillations until a target temperature is achieved is uncertain. It may be reasonable to perform further defibrillation attempts according to the standard basic life support algorithm concurrent with rewarming strategies. (Vanden Hoek TL, Morrison LJ, Shuster M, et al. Part 12: cardiac arrest in special situations: 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circulation*. 2010;122:S829-S861.)

3. When exertional heat stroke is suspected, what should be the first treatment priority?

A. Measuring rectal/core temperature
B. Starting an intravenous line for fluid resuscitation and drug delivery
C. Determining mental status
D. Removing the patient's clothes
E. Rapid whole body cooling with immersion in cooled or ice water

Answer: E Exertional heat stroke is a life-threatening emergency. Although assessing mental status, obtaining intravenous access for fluid administration, and determining the core temperature with a rectal thermometer are all important interventions, the cornerstone of therapy that should be an immediate focus is rapid cooling, which is best facilitated by rapid whole body cooling with immersion in cooled or ice water. (Casa DJ, Almquist J, Anderson SA, et al. The inter-association task force for preventing sudden death in secondary school athletics programs: best-practices recommendations. *J Athl Train*. 2013;48:546-553.)

110 ACUTE POISONING

LEWIS S. NELSON AND MARSHA D. FORD

EPIDEMIOLOGY

Each year, more than 4 million poisoning cases, suspected or verified, and 300,000 related hospital admissions occur in the United States. Poisoning-related deaths total more than 30,000 per year and are increasing. Poisonings are now the leading cause of injury-related death in the United States, where approximately 90% of these deaths involve drugs, predominantly prescription opioid analgesics.[1] Worldwide, however, pesticides and insecticides[2] are also common causes, as are intentional ingestions of toxic plants. The incidence of recurrent, purposeful self-poisoning is 12 to 18%, with most events occurring within 3 months of the original attempt. Poison centers provide support and collect data on more than 2 million exposures annually.[3] These facts emphasize the need for regulatory measures to improve the safety of prescription drugs, especially opioid analgesics, including appropriate prescribing, and aggressive treatment of poisoned patients, including early psychiatric intervention for suicidal behavior, to reduce fatalities and repeated attempts (Chapter 397).

DIAGNOSIS

Despite the vast array of toxins to which a patient may be exposed, the clinical manifestations of poisoning are fairly limited. In most cases, it is less important to predict exactly which toxin is responsible for the acute poisoning than it is to create a differential diagnosis based on a careful history and physical examination as well as basic laboratory assays. The recognition of the specific toxic syndrome, or *toxidrome*, guides the clinician toward the likely diagnosis based on reasonably solid evidence. On this basis, treatment, including initial stabilization, critical care, decontamination, and even the empirical administration of antidotes, can be guided by an understanding of the pharmacology and physiology of the patient's toxidrome. More advanced care, such as methods to enhance the elimination of specific toxicants, commonly requires serial examinations, additional history, and subsequent laboratory testing. Even then, however, the clinical picture may be clouded by exposures to multiple toxicants and an ill-defined time course since the initial exposure.

History

Details elicited about toxic exposures should include the involved drugs and other toxicants, their estimated or known amounts, the time and routes of exposure, the patient's symptoms and signs, and any treatment already administered. Intoxication may result from acute, chronic, or acute-on-chronic exposure. A *toxicant* is defined as a chemical capable of harming a biologic organism; this definition encompasses toxins, which are derived from living organisms, as well as medications, drugs of abuse, dietary supplements, and industrial and other chemicals. Determination of chronicity is important because signs and symptoms of chronic intoxication (Chapter 22) can differ from those of acute and acute-on-chronic intoxication. For example, a history of acute multisystem organ failure narrows the toxicant possibilities to a few gases, chemicals, and drugs. A listing of available medications (e.g., medications of the patient, spouse, relatives, or friends), use of nonprescription medications and herbal or dietary supplements or ethnic remedies, and occupational and avocational activities should be obtained. Occupational and avocational histories should include present and all past jobs and hobbies, with a focus on chemicals, metals, and gases. Known medical conditions may suggest classes of medications available to the patient. The patient's history, which may be incomplete if the patient is confused or suicidal, should be correlated with the clinical manifestations and course. Further history from relatives and friends and findings from the scene as reported by the transporting emergency medical services personnel may be relevant.

Physical Examination

The physical examination should focus on vital signs; the eye, ear, nose, and throat examination; and the neurologic, cardiopulmonary, gastrointestinal, and dermatologic systems. Findings can suggest certain toxidromes, which are clusters of signs and symptoms typical of poisoning. Among the dozens of toxidromes that assist in patient assessment and guide management, those due to adrenergic, anticholinergic, cholinomimetic, opioid, and sedative-hypnotic agents are most relevant to the emergency management of poisoned patients (Table 110-1). Patients may have some or all of these signs and symptoms; an incomplete clinical picture does not exclude a particular toxidrome, but it can still assist the clinician in identifying the correct category of toxicant involved.

Vital Signs

Tachycardia, which can occur with numerous toxicants, is most prominent in patients with anticholinergic and sympathomimetic toxidromes. However, because tachycardia also may occur with anxiety and other nontoxicologic conditions, it is not generally a helpful finding. The differential diagnosis for toxicant-induced *bradycardia* is more limited and includes β-adrenergic receptor antagonists, L-type calcium-channel antagonists (diltiazem or verapamil), cardioactive steroids, α-adrenergic receptor agonists (e.g., phenylephrine, whose

TABLE 110-1 TOXIDROMES AND ASSOCIATED DRUGS AND TOXICANTS

TOXIDROME	SYNDROME FEATURES		COMMON DRUGS AND TOXICANTS
	Vital Signs	*End Organ*	
Adrenergic	Hypertension, hyperthermia, tachycardia, tachypnea	Agitation, arrhythmias, diaphoresis, mydriasis, seizures	Amphetamines, caffeine, cathinone derivatives, cocaine, ephedrine, pseudoephedrine, *Ephedra* sp, phenylephrine,* theophylline
Anticholinergic	Hyperthermia, tachycardia, blood pressure generally normal	Agitation, delirium, decreased or absent bowel sounds, dry flushed skin and mucous membranes, mydriasis or blurred vision, seizures, urinary retention	First-generation H_1-receptor antagonists (e.g., diphenhydramine), belladonna alkaloids (e.g., scopolamine, atropine) from plants (e.g., *Datura* sp—deadly nightshade, henbane), benztropine, cyclic antidepressants, dicyclomine, muscle relaxants (e.g., orphenadrine, cyclobenzaprine), trihexyphenidyl
Cholinomimetic	Tachycardia or bradycardia†	Agitation, delirium, coma; bronchorrhea, bronchospasm; diaphoresis; fasciculations; lacrimation; miosis; urination; diarrhea, vomiting; seizures	Carbamate cholinesterase inhibitors (e.g., physostigmine, neostigmine, edrophonium), organophosphorus compounds including pesticides and nerve agents (e.g., somin, sarin) *Inocybe* or *Clitocybe* mushroom sp
Opioid, opiate	Bradycardia, bradypnea or apnea, hypotension (rare), hypothermia	CNS depression; hypotonia; miosis	Codeine, fentanyl, ultrapotent fentanyls, heroin, opioids (e.g., hydrocodone, oxycodone, meperidine, morphine), central α_2-agonists (e.g., clonidine, imidazolines)
Sedative-hypnotic	Generally near normal with benzodiazepines, but bradypnea or apnea, mild hypotension, and mild hypothermia can occur	Ataxia, CNS depression, hyporeflexia, slurred speech, stupor, or coma	Barbiturates, benzodiazepines, bromides, chloral hydrate, ethanol, ethchlorvynol, etomidate, glutethimide, meprobamate, methaqualone, methyprylon, propofol, zolpidem

*Reflex bradycardia can occur as a result of a pure α-adrenergic agonist effect.
†Tachycardia can occur early as a result of a preganglionic nicotinic effect; as toxicity progresses, postganglionic muscarinic effects predominate, and bradycardia develops.
CNS = central nervous system.

effects are mediated by baroreceptor reflexes), γ-hydroxybutyric acid, opioids, sedative-hypnotics, baclofen, central α_2-agonists like clonidine, organophosphorus and carbamate pesticides, muscarine-containing mushrooms (*Clitocybe, Inocybe* sp), plant- and animal-derived toxins (e.g., aconitine, andromedotoxin, ciguatoxin, and veratridine, all of which open sodium channels in the myocardium), therapeutic cholinesterase inhibitors (e.g., physostigmine), and some antiarrhythmic drugs (e.g., procainamide, flecainide, and other class IA, IC, and III drugs, such as amiodarone and sotalol). Bradycardia also is a preterminal sign for many consequential toxicants, such as cyclic antidepressants and cyanide.

Many toxicants cause *hypotension* (Chapter 8). The primary mechanisms are decreased peripheral vascular resistance, decreased myocardial contractility, hypovolemia secondary to gastrointestinal or dermal loss of intravascular volume, and, occasionally, arrhythmias. Common causes of *hypertension* (Chapter 67), generally due to vasoconstriction with or without enhanced inotropy, include amphetamines, cocaine, ephedrine and similar agents, ergots, phencyclidine, nicotine, phenylephrine, thyroid hormones, yohimbine, and chronic lead toxicity. Blood pressure can rise early in poisoning with central α_2-adrenergic agonists and monoamine oxidase inhibitors, but subsequent hypotension should be expected and is more concerning.

Hyperthermia (Chapter 109) occurs with toxicants that cause agitation or excessive psychomotor activity (e.g., cocaine, phencyclidine, monoamine oxidase inhibitors, strychnine), uncouple oxidative phosphorylation (e.g., salicylates, 2,4-dinitrophenol), increase the metabolic rate (thyroid hormones), impair sweating (e.g., first-generation antihistamines, anticholinergics, cocaine, phenothiazines, zonisamide), cause vasoconstriction (e.g., amphetamines, ephedrine), or impair vasodilation and alter perception of heat (cocaine). Other toxicant-induced states associated with hyperthermia include malignant hyperthermia, neuroleptic malignant syndrome, serotonin syndrome, metal fume fever, and aspiration. Toxicant-induced *hypothermia* is typically due to sedative-hypnotics, opioids, barbiturates, ethanol, phenothiazines, or hypoglycemic agents (such as insulin, sulfonylureas, and meglitinides) or rarely to unripe ackee fruit. Oxygen saturation (see Fig. 158-2) as measured by *pulse oximetry* decreases with true hypoxemia (see Fig. 104-1) or methemoglobinemia (Chapter 158) but remains normal or may be increased in patients with carbon monoxide poisoning (Chapter 94).

Eyes, Ears, Nose, and Throat

Toxicant-induced bilateral miosis (Fig. 110-1) has a limited differential diagnosis that includes central α_2-agonists such as clonidine, guanfacine, and the imidazolines; olanzapine; opioids; organophosphorus compounds or carbamates; therapeutic cholinesterase inhibitors (e.g., physostigmine); topical miotic ophthalmic drugs (e.g., pilocarpine, carbachol); and, variably, phencyclidine, phenothiazines, ethanol, and some sedative-hypnotics (Chapter 424). Pontine hemorrhage is a major nontoxicologic diagnosis to consider in a comatose patient with miotic pupils (Chapter 408). Mydriasis is a nonspecific finding. A unilateral dilated pupil may be due to topical ocular application of sympathomimetics (e.g., phenylephrine), antihistamines, or anticholinergic agents (e.g., inhaled anticholinergics such as ipratropium or tiotropium, dust or sap from *Datura* sp) and can be caused by a postauricular scopolamine patch. Failure of topical 4% pilocarpine ophthalmic drops to constrict the pupil supports the diagnosis of pupillary dilation from a topical mydriatic agent. Visual disturbances, including partial or total blindness as a result of systemic toxicity, have been reported with anticholinergic agents, carbon monoxide, digitalis, ethambutol, methanol, methyl bromide, quinine, and agents that are associated with pseudotumor cerebri, including antimicrobials (e.g., ampicillin, metronidazole, nalidixic acid, nitrofurantoin, sulfa drugs, tetracycline), glucocorticosteroids, lead, lithium, oral contraceptives, phenothiazines, phenytoin, and vitamin A. Nonarteritic anterior ischemic optic neuropathy has developed after the use of sildenafil and other related drugs (Chapter 234); the causal relationship is unknown.

Acute hearing loss (Chapter 428) can occur as a toxic effect of aminoglycosides, bromates, chloroquine, cisplatin, carboplatin, high-dose loop diuretics, nitrogen mustard, quinine, opioids, salicylates, vinblastine, and vincristine. Nasal septal erosions and perforations may be due to chronic exposure to intranasal cocaine (Chapter 34) or inhalation of fumes from chromium and nickel (Chapters 93 and 94).

Neurologic Signs

Many toxicants affect the central nervous system (CNS) and can produce agitated delirium, depression, or seizures (Table 110-2). Distinguishing features of various toxicants may assist in making the correct diagnosis. Patients who are withdrawing from opioids because of abstinence (as opposed to use of naloxone) are alert and oriented, whereas patients withdrawing from alcohol, barbiturates, benzodiazepines, and other sedative-hypnotics can be disoriented. Initial CNS depression can also develop with large ingestions of acetaminophen or ibuprofen. Isoniazid and theophylline are noted for producing seizures refractory to the usual doses of benzodiazepines and barbiturates. Pyridoxine treats isoniazid-induced seizures by increasing CNS γ-aminobutyric acid, which is depleted by isoniazid. Phenytoin is relatively ineffective for the majority of toxicant-induced seizures, perhaps because of the absence of a discrete seizure focus in most patients. Plant or mushroom ingestion can also produce CNS depression (e.g., *Rhododendron* sp, *Solanum* [bittersweet], *Taxus* [yew], *Sophora* [mescal bean]), CNS stimulation (e.g., *Catha edulis* [khat], *Strychnos nux-vomica* [contains strychnine], *Cicuta* sp [water hemlock], *Ephedra* [Mormon tea]), atropine-like effects (e.g., *Atropa belladonna* [deadly nightshade], *Datura* sp [jimsonweed]), and cholinomimetic effects (e.g., *Nicotiana* genus [tobacco], *Conium maculatum* [poison

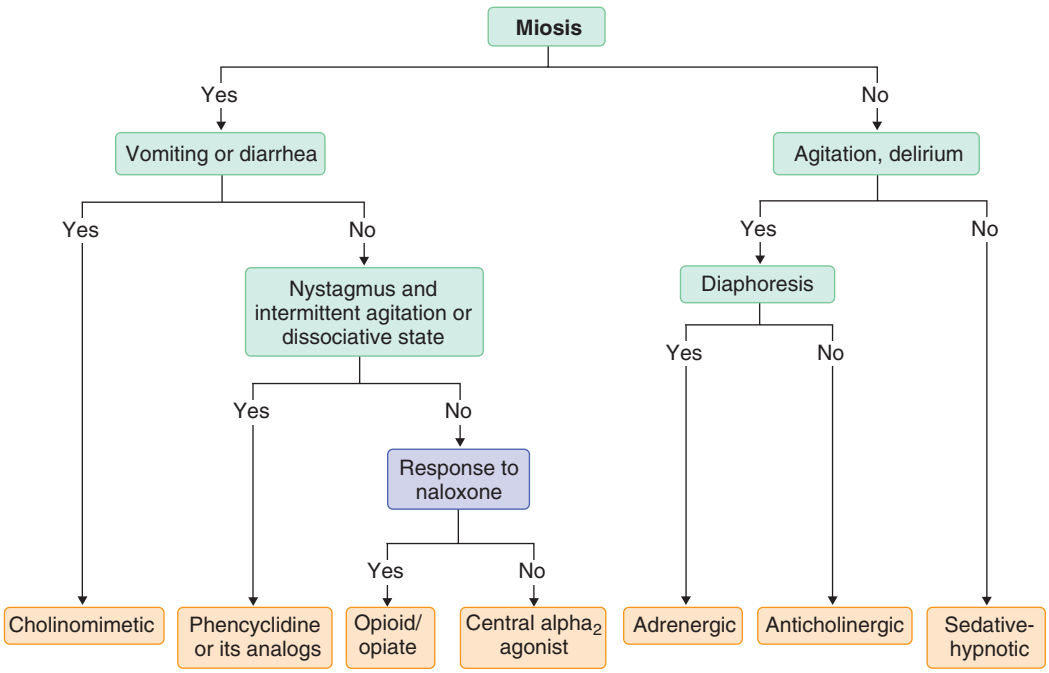

FIGURE 110-1. Diagnostic algorithm using the size of the pupils.

TABLE 110-2 CENTRAL NERVOUS SYSTEM EFFECTS OF TOXICANTS

TOXICANT CATEGORIES AND AGENTS	CNS EFFECTS		
	Agitated Delirium	*Decreased Level of Consciousness*	*Seizures*
CATEGORIES			
Adrenergic agonists	•		•
Anticholinergic agents	•	•	•
Anticonvulsants		•	• (paradoxical with some agents)
Antipsychotic drugs	•	•	•
β-Adrenergic receptor antagonists		•	•
Hallucinogenic agents	•		•
Monoamine oxidase inhibitors	•	•	•
Opioids	• (meperidine, tramadol)	•	• (meperidine, tramadol)
Sedative-hypnotics		•	• (rare)
Serotonin agonists	•	•	•
AGENTS			
Amphetamines, cocaine	•	•	•
Antidepressants	•	•	• (can be delayed with bupropion)
Antihistamines (first generation, e.g., diphenhydramine)	•	•	•
Barbiturates		•	
Benzodiazepines		•	
Cytochrome oxidase inhibitors (e.g., carbon monoxide, cyanide, hydrogen sulfide, azides)		•	•
Ephedra alkaloids and similar agents	•		•
Ethylene glycol, methanol	•	•	•
γ-Hydroxybutyrate and precursors	•	•	• (rare)
Lithium	•	•	•
Organophosphorus compounds (e.g., diazinon, malathion) and carbamates (e.g., carbaryl)		•	•
Salicylates	•	•	•
SSRIs/SRIs	•	•	• (uncommon)
Withdrawal from alcohol, barbiturates, benzodiazepines, other sedative-hypnotics	•	•	•
Withdrawal from opioids			• (reported only in neonates)

CNS = central nervous system; SRIs = serotonin re-uptake inhibitors; SSRIs = selective serotonin re-uptake inhibitors.

hemlock], *Inocybe* and *Clitocybe* mushrooms). Seizures can also occur with many of these plants and with mushrooms that contain gyromitrins (e.g., *Gyromitra* sp) and muscimol (e.g., *Amanita muscaria*, *Amanita pantherina*).

Distal axonopathy, a primary degeneration of peripheral nervous system axons with secondary degeneration of the myelin sheath, is the predominant type of toxicant-induced peripheral neuropathy. Common causative agents include acrylamide monomer, allyl chloride, arsenic (inorganic), capsaicin, carbon disulfide, chloramphenicol, cisplatin, colchicine, cyanate, dapsone, dideoxycytidine, dideoxyinosine, disulfiram, ethambutol, ethanol, ethylene oxide, gold salts, hexachlorophene, *n*-hexane, hydralazine, interferon, isoniazid, lead, linezolid, mercury, methyl bromide, methyl *n*-butyl ketone, metronidazole, nitrofurantoin, nitrous oxide, some organophosphorus compounds, phenol, platinum, podophyllotoxin, polychlorinated biphenyls, pyridoxine, tacrolimus, taxoids, thalidomide, thallium, vidarabine, vinca alkaloids, and vinyl chloride. Amiodarone, arsenic, and trichloroethylene can produce a demyelinating neuropathy, whereas pyridoxine can produce a sensory neuronopathy.

Neuronal transmission can be altered by aminoglycosides; the venom of *Latrodectus* species (widow spiders), scorpions (only the bark scorpion, *Centruroides sculpturatus*, in the United States), and crotaline (e.g., rattlesnakes) and elapid snakes; brevetoxin (shellfish) and ciguatoxin (various fish); neuromuscular blocking drugs; nicotine and related alkaloids; paralytic shellfish toxins; saxitoxin (shellfish); organophosphorus compounds and carbamates; tetrodotoxin (puffer fish [fugu], blue-ringed octopus, salamanders, newts, and others); and veratridine (e.g., false hellebore). Cranial nerves can be affected by carbon disulfide, ciguatoxin, domoic acid (shellfish), elapid venom, ethylene glycol metabolites, paralytic shellfish toxins, bark scorpion, saxitoxin, tetrodotoxin, thallium, and trichloroethylene. Mononeuropathies and vasculitic neuropathies are unlikely to be induced by toxicants (Chapter 112).

Cardiopulmonary Effects

The examination should focus on blood pressure, heart rate, electrocardiographic abnormalities (e.g., rhythm, conduction, depolarization, repolarization), and pulmonary findings, including pulse oximetry. Drugs and other toxicants that can cause arrhythmias or conduction abnormalities include β-adrenergic receptor antagonists, butyrophenones (e.g., haloperidol), L-type calcium-channel antagonists, cardioactive steroids (e.g., bufadienolides, found in toxic toad venom and incorporated into some illicit aphrodisiacs; cardenolides, such as digoxin, found in plants such as oleander and lily of the valley), chloral hydrate, chloroquine, cocaine, cyclic antidepressants, ethanol, halogenated hydrocarbons (e.g., halothane, trichloroethylene), magnesium, potassium, propoxyphene, thioridazine or mesoridazine, and antiarrhythmics and other agents that affect the myocardial voltage-gated sodium channels (e.g., bupivacaine, chloroquine, cocaine, cyclic antidepressants, flecainide, mexiletine, quinidine, procainamide, propafenone) and potassium channels (e.g., cisapride, citalopram, erythromycin [especially when taken concomitantly with cytochrome P-450 A inhibitors; Chapter 29], quinidine, sotalol, terfenadine). Bedside echocardiography may reveal depressed myocardial contractility as a result of agents that block the myocardial voltage-gated sodium channel, β-adrenergic receptor antagonists, calcium-channel antagonists, cyclic antidepressants, magnesium, arsenic, ciguatoxin, cyanide, ethanol, iron, scorpion venom, and tetrodotoxin.

Toxicants can produce myriad pulmonary effects, including airway irritation, parenchymal and pleural diseases, vascular diseases, and barotrauma. Immediate life-threatening toxic effects include acute lung injury and pulmonary edema, acute respiratory distress syndrome (Chapter 104), and rapidly developing pulmonary fibrosis or bronchiolitis obliterans. Typical syndromes and etiologic agents include pulmonary edema (β-adrenergic receptor antagonists, calcium-channel antagonists, antiarrhythmics, daunorubicin, doxorubicin), acute lung injury or acute respiratory distress syndrome (amphetamines,

cadmium, chlorine, cocaine, ethchlorvynol, methotrexate, opioids, heroin, paraquat, salicylates; inhalation of smoke, zinc chloride, methyl bromide, methyl chloride), and rapidly developing pulmonary fibrosis or bronchiolitis obliterans (nitrogen dioxide, paraquat).

Gastrointestinal Effects

Symptoms of nausea, vomiting, diarrhea, and abdominal pain are nonspecific and must be interpreted in the context of other findings. Agents that produce severe or life-threatening toxicity with early gastrointestinal findings include acid or large alkali ingestions; cardiac glycosides, colchicines, and other microtubular toxicants (e.g., podophyllin); iron; metals such as arsenic, mercury salts, and thallium; mushrooms containing amanitin (*Amanita phalloides*, *Amanita virosa*, *Amanita verna*, *Lepiota chlorophyllum*), gyromitrins (*Gyromitra esculenta*), orellanines (*Cortinarius orellanus*), or allenic norleucine (*Amanita smithiana*); nicotine; organophosphorus compounds; and theophylline. Severe abdominal pain and rigidity can occur with envenomation by *Latrodectus* species (widow spiders). Hepatotoxicity can occur as an adverse effect of the therapeutic use of many drugs; in the United States, acetaminophen and ethanol are the most common causes of toxicant-induced hepatotoxicity (Chapter 150). Other notable hepatotoxicants include aflatoxins (in food contaminated with *Aspergillus flavus*), arsenicals, carbon tetrachloride, copper sulfate, cyclopeptide mushrooms (e.g., *Amanita phalloides*), *Ephedra* species (e.g., ma huang), iron, methamphetamine, pennyroyal oil, pyrrolizidine alkaloids (various plant species used in herbal teas), and vitamin A in chronic excessive doses.

Dermatologic Signs

The skin, hair, nails, and mucous membranes should be examined for evidence of intravenous drug use; the presence or absence of skin and mucous membrane moisture; abnormal skin coloration, including erythema and cyanosis; alopecia; and nail abnormalities. Bullous skin lesions have been reported with chronic barbiturate use, glutethimide, carbon monoxide, meprobamate, methadone, and valproic acid. Cyanosis may reflect hypoxemia or methemoglobinemia. Commonly used agents that can cause methemoglobinemia include aniline dyes, benzocaine and other amide anesthetics, dapsone, naphthalene, nitrates, nitrites, phenazopyridine, rifampin, and sulfonamides. Skin erythema or flushing occurs with anticholinergic agents, boric acid ingestion, monosodium glutamate, niacin, scombroid toxicity as a result of the ingestion of inadequately refrigerated fish with a high histidine content (e.g., tuna, mahi-mahi, amberjack), vancomycin, and interactions between ethanol and numerous agents that produce disulfiram or disulfiram-like reactions (e.g., carbon disulfide, some cephalosporins, *Coprinus atramentarius* mushroom, disulfiram, griseofulvin, metronidazole, thiuram herbicides, trichloroethylene).

Specific Toxicants

Some common toxicants can be suspected from their characteristic manifestations (Table 110-3). These suspicions should guide specific diagnostic and therapeutic strategies that complement general decontamination and supportive treatments.

Diagnostic Tests

Drug testing should be guided by the history and physical examination, with emphasis on tests that can influence management. Rapid qualitative urine drug screening tests are readily available in most hospitals, but their clinical value is limited by the number of drugs that can be tested and the reliability of the tests themselves. Although such testing may be valuable for subsequent psychiatric evaluation, it is less reliable for urgent diagnosis and therapy. For example, a positive test result may be unrelated to the patient's condition because drug analytes may be detectable for days after drug use, depending on the drug, dose, and frequency of use. In addition, both false-positive and false-negative results occur (Table 110-4), so the screening result must be verified with a second method, such as gas chromatography–mass spectrometry. The type of drug use in a population should be considered in determining the drugs to be screened to decrease the incidence of false-positive results. A test with 99% specificity for a drug with a prevalence of 0.1% in a population would produce 10 false-positive results for

TABLE 110-3 PATHOPHYSIOLOGY, CLINICAL EFFECTS, AND MANAGEMENT OF SPECIFIC DRUGS AND TOXICANTS

DRUG OR TOXICANT	PATHOPHYSIOLOGY	CLINICAL EFFECTS	LABORATORY	SPECIFIC THERAPY
Acetaminophen[7]	NAPQI (toxic metabolite) binds hepatic and renal tubular cells; acetaminophen itself induces transient decrease in functional factor VII	Initial: nausea, vomiting, coma, lactic acidosis in severe cases Days 1-3: elevated INR, aminotransferase levels; RUQ tenderness; increased creatinine level in severe cases Days 4-14: gradual recovery or continued increase in INR and creatinine, lactic acidosis, coma, cerebral edema, death	Potentially toxic level ≥150 μg/mL 4 hours after ingestion* INR may be transiently elevated in first 24 hours because of decrease in functional factor VII; further increases indicate hepatic necrosis; elevated aminotransferase and bilirubin levels not predictive of hepatic failure Creatinine elevated in severe cases	NAC (see dosing guidelines in Table 110-6); NAC can increase INR but not aPTT†
Amphetamines	Increased release of presynaptic norepinephrine and dopamine Increased serotonin release (especially MDMA, PMA, DOB, other synthetic amphetamines)	Mild: euphoria, decreased appetite, repetitive behavior Moderate: vomiting, agitation, hypertension, tachycardia, mydriasis, bruxism, diaphoresis Severe: hypertension or hypotension, arrhythmias, hyperthermia, seizures, coma, multisystem organ failure, hyponatremia (SIADH), cerebral infarction or hemorrhage	Not helpful; many false-positives and false-negatives on screening tests (see Table 110-4)	IV crystalloids External cooling Benzodiazepines to control agitation or seizures Benzodiazepines or phentolamine for hypertension See SSRIs/SRIs for features and treatment of serotonin syndrome
β-Adrenergic receptor antagonists	Blocks catecholamines from β-adrenergic receptors α- and β-adrenergic receptor antagonism: carvedilol, labetalol Delayed rectifier potassium-channel blockade: sotalol	Bradyarrhythmias (with high doses), decreased myocardial contractility, hypotension, respiratory depression, decreased consciousness with seizures or coma (lipophilic agents, e.g., propranolol), prolonged QT interval (sotalol)	ECG No specific tests	IV glucagon, 3-5 mg over 2-minute period; if no increase in BP or HR, can repeat up to 10 mg; if effective, immediately start continuous infusion at 2-10 mg/hr; if still unstable, options include (1) regular insulin, 1 U/kg by IV bolus, followed by 1 U/kg/hr, plus dextrose to maintain euglycemia; (2) norepinephrine or dobutamine infusion titrated to desirable BP and HR Electrical pacing, IABP, or intravenous lipid emulsion therapy in refractory cases

TABLE 110-3 PATHOPHYSIOLOGY, CLINICAL EFFECTS, AND MANAGEMENT OF SPECIFIC DRUGS AND TOXICANTS—cont'd

DRUG OR TOXICANT	PATHOPHYSIOLOGY	CLINICAL EFFECTS	LABORATORY	SPECIFIC THERAPY
L-type calcium-channel antagonists	Blocks L-type voltage-sensitive calcium channels, thereby decreasing calcium entry into myocardial and vascular smooth muscle cells Decreases pancreatic insulin release and increases insulin resistance	Bradyarrhythmias (verapamil, diltiazem), hypotension, hyperglycemia Tachycardia and hypotension (dihydropyridines such as amlodipine, nifedipine)	ECG No specific tests	IV 10% calcium chloride, 10-20 mg/kg (0.1-0.2 mL/kg); can repeat once; if BP improves, continuous infusion at 0.2-0.5 mL/kg/hr (20-50 mg/kg/hr) Ionized Ca^{2+} levels should not exceed 2× normal (severe cases will be refractory to calcium therapy) Glucagon, high-dose insulin and dextrose, catecholamines, and milrinone (as for β-adrenergic antagonists) Intravenous lipid emulsion therapy for verapamil and diltiazem, unclear for dihydropyridines
Cardiac glycosides, including digoxin, bufadienolides (toxic toad venom), and cardenolides (e.g., oleander, lily of the valley, dogbane)	Inhibits Na^+,K^+-ATPase Decreased CNS sympathetic output Decreased baroreceptor sensitivity Increased vagal acetylcholine discharge	Bradyarrhythmias, including second- and third-degree AV block and asystole Ventricular ectopy, tachycardia, fibrillation Junctional tachycardia, paroxysmal atrial tachycardia with block Weakness, visual disturbances, nausea, vomiting	Serum digoxin level Serum potassium (hyperkalemia occurs in acute poisoning and is prognostic for poor outcome without Fab; hypokalemia may be present in chronic poisoning due to concomitant medications), magnesium, and creatinine levels	Correct hypokalemia and hypomagnesemia; do not give calcium Digoxin-specific antibody fragments (Fab) indicated if patient has hemodynamically significant arrhythmias, serum potassium ≥5 mEq/L if acute overdose, Mobitz II or third-degree AV block, ingestion of bufadienolide- or cardenolide-containing agents, or renal insufficiency Empirical dose Chronic: 2-5 vials Acute: 10-20 vials Calculated dose Chronic: number of vials = 2 × serum digoxin level (ng/mL) × 5.6 × weight (kg)/1000 Acute: number of vials = 2 × oral digoxin dose (mg) × 0.8 If hypokalemic (generally chronic overdose), replete serum potassium
Cyclic antidepressants	Myocardial sodium- and potassium-channel blockade Blockade of α-adrenergic and cholinergic muscarinic receptors Inhibition of norepinephrine re-uptake	Decreased level of consciousness (can develop rapidly), myoclonus, seizures, coma Anticholinergic toxidrome (see Table 110-1) Sinus tachycardia, ventricular conduction delays, ventricular arrhythmias, asystole Hypotension	Serum levels not helpful in management	Intermittent IV boluses of $NaHCO_3$ (1 mEq/kg) to narrow QRS to <100 msec; infusion of $NaHCO_3$ is less effective; maintain arterial pH at 7.5 because alkalemia and sodium ions can improve cardiovascular performance Contraindicated drugs: types IA and IC antiarrhythmic agents, physostigmine, flumazenil
Ethylene glycol, methanol (e.g., antifreeze, window cleaners, camping stove fuels)	Ethylene glycol: toxic metabolites produce cytotoxicity in CNS, kidneys, lungs, heart, liver, muscles; metabolic acidosis is due to glycolate accumulation; oxalate complexes with calcium, so hypocalcemia can develop Methanol: metabolized to formic acid, which is responsible for metabolic acidosis and inhibition of cytochrome aa_3; target organs include retina, optic nerve, CNS	Ethylene glycol: CNS depression, cerebral edema, seizures, anion gap metabolic acidosis, renal failure with acute tubular necrosis, pulmonary edema, myositis Methanol: nausea, vomiting; cerebral edema, hemorrhage, infarcts; necrosis of thalamus and putamen; anion gap metabolic acidosis; visual disturbances, papilledema, hyperemic optic disc, nonreactive pupils	Serum ethylene glycol and methanol levels; levels may be low or undetectable if significant metabolism has occurred Ethylene glycol: serum calcium, creatinine, BUN levels; examine urine for calcium oxalate crystals; false hyperlactatemia occurs with certain analyzers using L-lactate oxidase, which cross-reacts with glycolic and glyoxylic acid metabolites	For both: fomepizole[8] (inhibits alcohol dehydrogenase and blocks formation of toxic metabolites), 15 mg/kg IV loading dose, then 10 mg/kg IV for 4 doses during the next 48 hours, then 15 mg/kg for subsequent doses; interval dosing is q12h (q4h during hemodialysis, with dosing interval adjustments at start and finish); continue until ethylene glycol or methanol is no longer detectable Use of ethanol is no longer recommended Hemodialysis: initiate if level is ≥50 mg/dL or metabolic acidosis with end-organ toxicity; continue until acidosis resolves and serum level of ethylene glycol or methanol is undetectable (if available) Monitor for cerebral edema with possible herniation Ethylene glycol: IV calcium for symptomatic hypocalcemia Methanol: folinic acid, 50 mg IV q4h until methanol not detectable and acidosis cleared

TABLE 110-3 PATHOPHYSIOLOGY, CLINICAL EFFECTS, AND MANAGEMENT OF SPECIFIC DRUGS AND TOXICANTS—cont'd

DRUG OR TOXICANT	PATHOPHYSIOLOGY	CLINICAL EFFECTS	LABORATORY	SPECIFIC THERAPY
γ-Hydroxybutyrate (GHB) and its precursors (γ-butyrolactone and 1,4-butanediol)	Agonist effect on CNS GHB receptors; indirect action with opioid receptors (may increase proenkephalins); metabolized to GABA, interacts with $GABA_B$ receptors; decreases dopamine release	CNS: rapid loss of consciousness, with recovery typical within 2-4 hours; myoclonus (possible seizures) Respiratory depression; bradycardia; nausea, vomiting	No specific tests	Supportive care, including respiratory support as needed Withdrawal resembles sedative-hypnotic withdrawal and can be treated with benzodiazepines or pentobarbital
Lithium	Decreases brain inositol; alters CNS serotonin, dopamine, and norepinephrine; inhibits adenylate cyclases, including those that mediate vasopressin-induced renal concentration and thyroid function	Chronic toxicity usually more severe than acute toxicity: tremor, hyperreflexia, drowsiness, incoordination, clonus, confusion, ataxia; in severe cases, seizures, coma, death; recovery may take weeks, and CNS deficits may persist Sinus node dysfunction, QT prolongation, T wave abnormalities, U waves Nephrogenic diabetes insipidus, hypothyroidism, hyperthyroidism, hypercalcemia, pseudotumor cerebri Acute toxicity: nausea, vomiting, diarrhea, and milder neurologic findings	Peak serum levels: Normal dose, 2-3 hours; up to 5 hours for sustained-release lithium Acute overdose: peak may be delayed ≥4-12 hours	Replenish intravascular volume, maintain urinary output at 1-2 mL/kg/hr Consider gastrointestinal decontamination with oral polyethylene glycol electrolyte solution within 1-2 hours after acute overdose of sustained-release drug Hemodialysis[‡] in patients with altered mental status, ataxia, seizures, or coma or in patients with mild symptoms in the setting of acute overdose or renal insufficiency Ineffective or contraindicated therapies include oral activated charcoal, diuretics, and aminophylline
Opioids (e.g., opiate [natural]: morphine, codeine; semisynthetic: heroin, oxycodone, hydrocodone; synthetic: methadone, fentanyl)	Agonist effect at CNS μ, κ, and δ opioid receptors; result is cell hyperpolarization and decreased neurotransmitter release	CNS depression, respiratory depression, miosis (see Table 110-1) Dextromethorphan increases CNS serotonin and inhibits NMDA receptors, which causes hallucinations Seizure risk with tramadol, meperidine QTc prolongation and torsades de pointes with methadone	Rapid urine drug screens detect morphine, poorly detect semisynthetic opioids, and do not detect synthetic opioids; some interferents/irrelevants (see Table 110-4)	Ventilate and oxygenate IV naloxone, 0.04 mg initial dose and titrate every 2-3 minutes in patients with likely opioid dependence; in patients known to be naïve, start 0.4 mg; repeat up to 10 mg if no response Continuous infusion for recurrent symptoms or sustained-release opioid ingestion; give 50% of dose that produces desired effect 15 minutes after initial effect is obtained, then infuse two thirds of this dose every hour; infusion rate can be increased or decreased to maintain normal respiration and to avoid withdrawal symptoms Contraindicated therapies: naltrexone should not be used for acute opioid reversal[9]
Organophosphorus compounds and carbamates (e.g., diazinon, mevinphos, fenthion, aldicarb)	Inhibits acetylcholinesterase, resulting in excessive acetylcholine stimulation of nicotinic and muscarinic receptors in autonomic and somatic motor nervous systems and CNS	Nicotinic-mediated effects: tachycardia, mydriasis, hypertension, delirium, coma, seizures, muscle weakness, fasciculations Muscarinic-mediated effects: salivation, lacrimation, urination, vomiting, defecation, miosis, bronchorrhea, bronchospasm, bradycardia	Serum (butyrylcholinesterase) or RBC (acetylcholinesterase) activity <50% of normal (see Table 110-6) Clinical recovery occurs before serum cholinesterase levels normalize	Atropine, 1-2 mg by initial IV bolus; double the dose every 5 minutes (2 mg, 4 mg, 8 mg, 16 mg, and so on) until drying of bronchial secretions, adequate oxygenation, pulse >80 beats/min, systolic blood pressure >80 mm Hg achieved; continuous infusion at 10-20% of total stabilizing dose per hour; stop infusion when patient develops concerning signs or symptoms of anticholinergic toxidrome (see Table 110-1); restart infusion at lower rate when signs or symptoms abate Pralidoxime chloride 30 mg/kg (maximum 2 g) IV bolus during 30 minutes, then 8-10 mg/kg/hr (maximum 650 mg/hr) continuous infusion; administer as soon as possible after poisoning; continue 12-24 hours after atropine no longer required and symptoms resolve[§]

TABLE 110-3 PATHOPHYSIOLOGY, CLINICAL EFFECTS, AND MANAGEMENT OF SPECIFIC DRUGS AND TOXICANTS—cont'd

DRUG OR TOXICANT	PATHOPHYSIOLOGY	CLINICAL EFFECTS	LABORATORY	SPECIFIC THERAPY
Salicylates	Inhibit cyclooxygenase; decrease formation of prostaglandins and thromboxane A_2; stimulate CNS medullary respiratory center and chemoreceptor trigger zone; impair platelet function; disrupt carbohydrate metabolism; uncouple oxidative phosphorylation; increase vascular permeability	Acute toxicity Mild: nausea, vomiting, diaphoresis, tinnitus, decreased hearing, hyperpnea, tachypnea Moderate-severe: confusion, delirium, coma, seizures, hyperthermia, ARDS; death can occur within hours of overdose Chronic toxicity: same as acute, but may not have diaphoresis or vomiting Consider diagnosis in patients with new-onset confusion, anion gap metabolic acidosis, or acute lung injury	Serum salicylate level: toxic ≥30 mg/dL; level ≥100 mg/dL indicates life-threatening toxicity with possible sudden, rapid clinical deterioration; in chronic toxicity, levels may be minimally elevated (>30 mg/dL), and clinical evaluation is more reliable for gauging degree of toxicity Arterial blood gases: respiratory alkalosis with metabolic acidosis Anion gap metabolic acidosis Prolonged PT and PTT, ketonuria, ketonemia	Multidose activated charcoal q2-3h in acute overdose Hemodialysis with progressive symptoms, particularly neurologic, hyperthermia, renal failure, ARDS, or salicylate level >100 even with minor clinical findings
SSRIs/SRIs	Inhibit re-uptake of serotonin SRIs have additional effects (e.g., duloxetine inhibits norepinephrine re-uptake, nefazodone inhibits serotonergic 5-HT_2 receptors, trazodone inhibits peripheral α-adrenergic receptors, venlafaxine inhibits norepinephrine and dopamine re-uptake)	Vomiting, blurred vision, CNS depression, tachycardia Seizures and coma rare Torsades de pointes reported with citalopram Serotonin toxicity: clonus, agitation, tremor, diaphoresis, hyperreflexia; hyperthermia and hypertonicity in severe cases	No specific tests If serotonin toxicity suspected: electrolytes, BUN, glucose, liver enzymes, coagulation panel, blood gases, chest radiograph	Respiratory support as needed Benzodiazepines for agitation or seizures Serotonin toxicity: consider cyproheptadine, 12 mg PO initial dose, then 2 mg PO q2h (to a maximum of 32 mg/day) until symptoms resolve Critical therapies for hyperthermia, rhabdomyolysis, DIC, ARDS, renal and hepatic dysfunction, torsades de pointes

*A nomogram to evaluate the potential toxicity of levels drawn more than 4 hours after ingestion is provided in Rumack BH, Matthew H. Acetaminophen poisoning and toxicity. *Pediatrics.* 1975;55:871-876. The nomogram is valid only for levels drawn after a single acute ingestion.
†Intravenous *N*-acetylcysteine (NAC) has generally replaced oral NAC for the majority of cases, largely because of convenience, not efficacy. The full 21-hour course of intravenous NAC should be administered in most situations in which it is used. Oral NAC, which remains an acceptable alternative, can be discontinued in patients with uncomplicated disease after a loading dose plus six maintenance doses if hepatic aminotransferase levels are normal and acetaminophen is not detected; otherwise, the full regimen should be administered.
‡Continue hemodialysis until the serum lithium level is less than 1 mEq/L. Recheck the level 4 to 8 hours after dialysis, and restart hemodialysis if the level is higher than 1 mEq/L. Repeat this cycle until the serum lithium level remains lower than 1 mEq/L.
§Randomized, placebo-controlled trials of pralidoxime in acute organophosphorus poisoning have not found a significant difference in mortality rates or need for intubation.
aPTT = activated partial thromboplastin time; ARDS = acute respiratory distress syndrome; AV = atrioventricular; BP = blood pressure; BUN = blood urea nitrogen; CNS = central nervous system; DIC = disseminated intravascular coagulation; DOB = 4-bromo-2,5-dimethoxyamphetamine; ECG = electrocardiogram; GABA = γ-aminobutyric acid; HR = heart rate; IABP = intra-aortic balloon counterpulsation; INR = international normalized ratio; MDMA = 3,4-methylenedioxymethamphetamine; Na^+,K^+-ATPase = sodium, potassium adenosine triphosphatase; NAC = *N*-acetylcysteine; NAPQI = *N*-acetyl-*p*-benzoquinone imine; NMDA = *N*-methyl-D-aspartate; PMA = paramethoxyamphetamine; PT = prothrombin time; PTT = partial thromboplastin time; RBC = red blood cell; RUQ = right upper quadrant (abdomen); SIADH = syndrome of inappropriate antidiuretic secretion; SRIs = serotonin re-uptake inhibitors; SSRIs = selective serotonin re-uptake inhibitor.

TABLE 110-4 QUALITATIVE URINE DRUG SCREENS: CAUSES OF ERRONEOUS RESULTS*

DRUG/TOXICANT	INTERFERENTS/IRRELEVANTS†	COMMENTS
Amphetamines	Amantadine, bupropion, chlorpromazine, ephedrine, pseudoephedrine, desoxyephedrine, ephedra alkaloids (from *Ephedra* sp), mexiletine, phenylephrine, phenylpropanolamine, selegiline, trazodone	Many false positives; Vicks nasal inhaler (desoxyephedrine) and selegiline also cause positive GC-MS findings; chiral confirmation is required; newer immunoassays have eliminated false-positive results from desoxyephedrine
Benzodiazepines	Oxaprozin, sertraline	Most assays directed against oxazepam; poor detection of benzodiazepines without the oxazepam metabolite (e.g., alprazolam, lorazepam, triazolam)
Cocaine	Coca leaf teas (clinical false positive)	Few analytical false positives; urine is reliable for detecting true positives
Opiates, opioids	Poppy seeds (contain morphine), quinine, quinolones, rifampin	Assay directed against morphine; poorly detects semisynthetic opiates and does not detect synthetic opiates (e.g., fentanyls, meperidine, methadone, propoxyphene)
Phencyclidine	Dextromethorphan, diphenhydramine, doxylamine, ibuprofen, ketamine, tramadol, venlafaxine	Can be used, although not reliably, to identify dextromethorphan or ketamine misuse and abuse
Tetrahydrocannabinol	Dronabinol, efavirenz, proton pump inhibitors	Positive result is seldom clinically relevant; synthetic cannabinomimetics (e.g., spice, K2) are not reliably detected
Tricyclic antidepressants	Carbamazepine, cyclobenzaprine, cyproheptadine, diphenhydramine, phenothiazines, quetiapine	

*Advances in drug screening and variability in immunoassay results should be considered by the clinician when interpreting qualitative drug screening results. Consultation with the testing laboratory is advised. Positive screening results are considered presumptive and should be verified by GC-MS.
†Irrelevants are agents causing true-positive but clinically irrelevant results on laboratory screening tests; they vary, depending on the screening method.
GC-MS = gas chromatography–mass spectrometry.

every true-positive result. Clinically irrelevant true-positive findings also occur, such as when poppy seeds produce a positive opiate test result. Failure to consider these limitations of drug screens can result in misdiagnosis.

For a limited number of drugs and toxicants, levels in blood or urine are useful for diagnosis, therapy, or monitoring (Table 110-5). Threshold levels of certain toxicants indicate the need for specific therapies: acetaminophen (*N*-acetylcysteine), ethylene glycol (fomepizole and hemodialysis), iron (deferoxamine), methanol (fomepizole and hemodialysis), methemoglobin (methylene blue), and salicylates (urine alkalinization and hemodialysis). In chronic poisoning with some drugs, such as salicylates or theophylline, these therapies may be indicated at lower drug levels. In general, any end-organ toxicity that is evident or anticipated (on the basis of the toxicant, amount

TABLE 110-5 CLINICALLY IMPORTANT QUANTITATIVE DRUG LEVELS

DRUG OR TOXICANT	LEVELS	
	Therapeutic	***Toxic****
SOURCE: BLOOD OR SERUM		
Acetaminophen[†]	10-30 µg/mL	≥150 µg/mL 4 hours after ingestion[‡]
Carbamazepine	4-12 µg/mL	>15 µg/mL
Carboxyhemoglobin	Nonsmoker: 0.5-1.5% Smoker: 4-9%	>20%[§]
Cholinesterase[‖]		
Serum (butyrylcholinesterase)	3100-6500 U/L	<50% of normal value
Red blood cell (acetylcholinesterase)	26.7-49.2 U/g of hemoglobin	<50% of normal value
Digoxin (≥6 hours after oral dose for long-term therapy)	0.8-2.0 ng/mL[¶]	>2.0 ng/mL
Ethanol	None measured	>80 mg/dL[**]
Ethylene glycol	None measured	>25 mg/dL
Iron	50-175 µg/dL	>350 g/dL
Lead	<10 µg/dL	>25 g/dL
Lithium	0.6-1.2 mEq/L	>1.2 mEq/L[††]
Methanol	None measured	>25 mg/dL
Methemoglobin	1-2%	>15%
Phenobarbital	15-40 µg/mL	>40 g/mL
Phenytoin	10-20 µg/mL	>20 g/mL
Salicylates	≤30 mg/dL	>30 mg/dL
Theophylline	8-20 µg/mL	>20 g/mL
Valproic acid	50-100 µg/mL	>100 g/mL
SOURCE: URINE	***Normal***	***Toxic****
Arsenic	< 50 µg/day	>100 g/24-hr urine[††]
Mercury	< 20 µg/L	>20 g/L[††]
Thallium	< 5.0 µg/L	>200 g/L[††]

*The "toxic" level is provided for perspective. For many toxicants, simply being above this value does not imply a specific need for therapy or a necessarily poor prognosis. It does, however, suggest a need for additional evaluation, observation, or monitoring.
[†]False-positive levels of 16 to 28 µg/mL have been reported in patients with bilirubin levels greater than 17 mg/dL.
[‡]Levels drawn more than 4 hours after ingestion should be plotted on the nomogram provided by Rumack and Matthew (Rumack BH, Matthew H. Acetaminophen poisoning and toxicity. *Pediatrics.* 1975;55:871-876) to assess the potential for toxicity.
[§]Lower levels may be toxic in pregnant patients and in those with prolonged exposure to carbon monoxide.
[‖]Consult a reference laboratory for normal values; results are assay dependent.
[¶]Some patients may require levels above the therapeutic range to control symptoms.
[**]The value of 80 mg/dL for ethanol is the statutory limit for operating a motor vehicle. Toxic clinical effects are uncommon with concentrations below 200 mg/dL.
[††]Lower values may indicate toxicity if appropriate clinical findings are present.

ingested, and time required to produce toxic effects) is more important than a specific level in determining the need for treatment.

Occult acetaminophen ingestion with toxic serum levels occurs in 0.3 to 1.9% of intentional ingestions. Given that these patients may be asymptomatic until hepatotoxicity develops and that administration of an antidote can prevent this hepatotoxicity, the current recommendation is to test the serum for acetaminophen in all patients with intentional self-harm ingestions.

Other Blood Tests

Anion gap metabolic acidosis resulting from primary lactic acidosis can be caused by cyanide, hydrogen sulfide, iron, isoniazid, metformin, nucleoside reverse transcriptase inhibitors, phenformin, sodium azide, and, rarely, acetaminophen with high serum levels. Anion gap metabolic acidosis not related to lactic acidosis occurs with diethylene glycol, ethylene glycol, nonsteroidal anti-inflammatory drugs, methanol, salicylates, and toluene. In poisonings resulting from ibuprofen, methanol, propylene glycol, and salicylates, lactic acid can also be produced, but the level is insufficient to account for the anion gap. Anion gap metabolic acidosis can also develop in patients with ongoing agitation, hyperthermia, and muscle rigidity, such as in neuroleptic malignant syndrome (Chapter 434), or in some cases of rhabdomyolysis (Chapter 113) secondary to toxicants such as doxylamine, phencyclidine, strychnine, cocaine, and amphetamines. Elevated serum creatinine and blood urea nitrogen levels indicative of declining renal function may be seen with numerous toxicants. Direct toxicity occurs with acetaminophen, aminoglycosides, cadmium, Chinese weight-loss botanicals (containing *Stephania tetrandra* or *Magnolia officinalis*), chromium, *Crotalus durissus* venom, diethylene glycol, diquat, ethylene glycol, fluorinated anesthetics, gold, heroin, lithium (diabetes insipidus), mercury salts, mushrooms (*Amanita smithiana* and *Cortinarius* sp), paraquat, radiocontrast agents, solvents (e.g., carbon tetrachloride, trichloroethylene, tetrachloroethylene, toluene), and sulfonamides. Agents that decrease glomerular perfusion by reducing renal blood flow include amphotericin, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, cocaine, cyclosporine, mannitol (excessive chronic doses), methotrexate, and nonsteroidal anti-inflammatory drugs.

Imaging

A computed tomographic scan of the head can detect life-threatening cerebral edema secondary to toxicant-induced hepatic failure, ethylene glycol, and methanol. It also detects intracranial bleeding caused by anticoagulants, scorpion venom, and sympathomimetics (e.g., amphetamines, cocaine, phenylpropanolamine). An abdominal radiograph can reveal radiopaque ferrous sulfate tablets or metals such as arsenic, lead, mercury, and thallium.

Diagnostic Syndromes

Given the myriad combinations of signs, symptoms, and laboratory findings, making the correct diagnosis in a noncommunicative patient can be daunting. A thorough history from bystanders, friends, and prehospital medical personnel may yield crucial information. In addition, the diagnostic possibilities can be narrowed by findings that can narrow the differential diagnosis with modest certainty. For example, consider a patient with sudden loss of consciousness, anion gap metabolic acidosis, and bradycardia without hypoxemia. Among the possible causes of anion gap metabolic acidosis (see earlier) and sudden loss of consciousness are hydrogen sulfide, cyanide, and severe poisoning with sodium azide; however, sinus bradycardia in the absence of acute ischemic cardiac injury is typical only of cyanide poisoning.

TREATMENT

Initial Stabilization

Intubation and Respiratory Support

Appropriate airway management should be instituted to correct hypoxemia and respiratory acidosis and to protect against pulmonary aspiration (Fig. 110-2); intubation should be considered if the patient has depressed consciousness and a decreased gag reflex. Rapid-sequence intubation facilitates airway management. Anatomic difficulties should be anticipated in patients with caustic ingestions (e.g., hypopharyngeal burns that may perforate); angioedema caused by angiotensin-converting enzyme inhibitor therapy or envenomation by some rattlesnakes, such as the canebrake (*Crotalus horridus atricaudatus*) and eastern diamondback (*Crotalus adamanteus*; Chapter 112); and swelling secondary to direct tissue injury (e.g., huffing compressed hydrocarbons, smoking crack) or secondary to anaphylactoid and anaphylactic reactions. Endotracheal intubation by flexible fiberoptic nasopharyngoscopy may be indicated in these cases. Hypoxemia can occur with toxicants that produce CNS depression, such as antidepressants, barbiturates, sedative-hypnotics, and central α_2-adrenergic receptor agonists (clonidine), or agents causing peripheral neuromuscular impairment, such as nicotine, organophosphorus compounds, strychnine, tetrodotoxin (puffer fish, blue-ringed octopus), botulinum, or envenomation from elapids (coral snake), Mojave rattlesnakes, or certain coelenterates (box jellyfish; Chapter 112).

Respiratory acidosis can rapidly worsen the toxicities of cyclic antidepressants and salicylates; sedation of these patients should be accompanied by immediate airway support. Intoxicated patients may have an increased risk for pulmonary aspiration because of concomitant CNS depression, attenuated airway reflexes, full stomachs, and delayed gastric emptying.

Succinylcholine can cause prolonged paralysis in patients with organophosphorus poisoning and can exacerbate hyperkalemia from cardioactive steroids (e.g., digoxin), hydrofluoric acid, or rhabdomyolysis (Chapter 113).

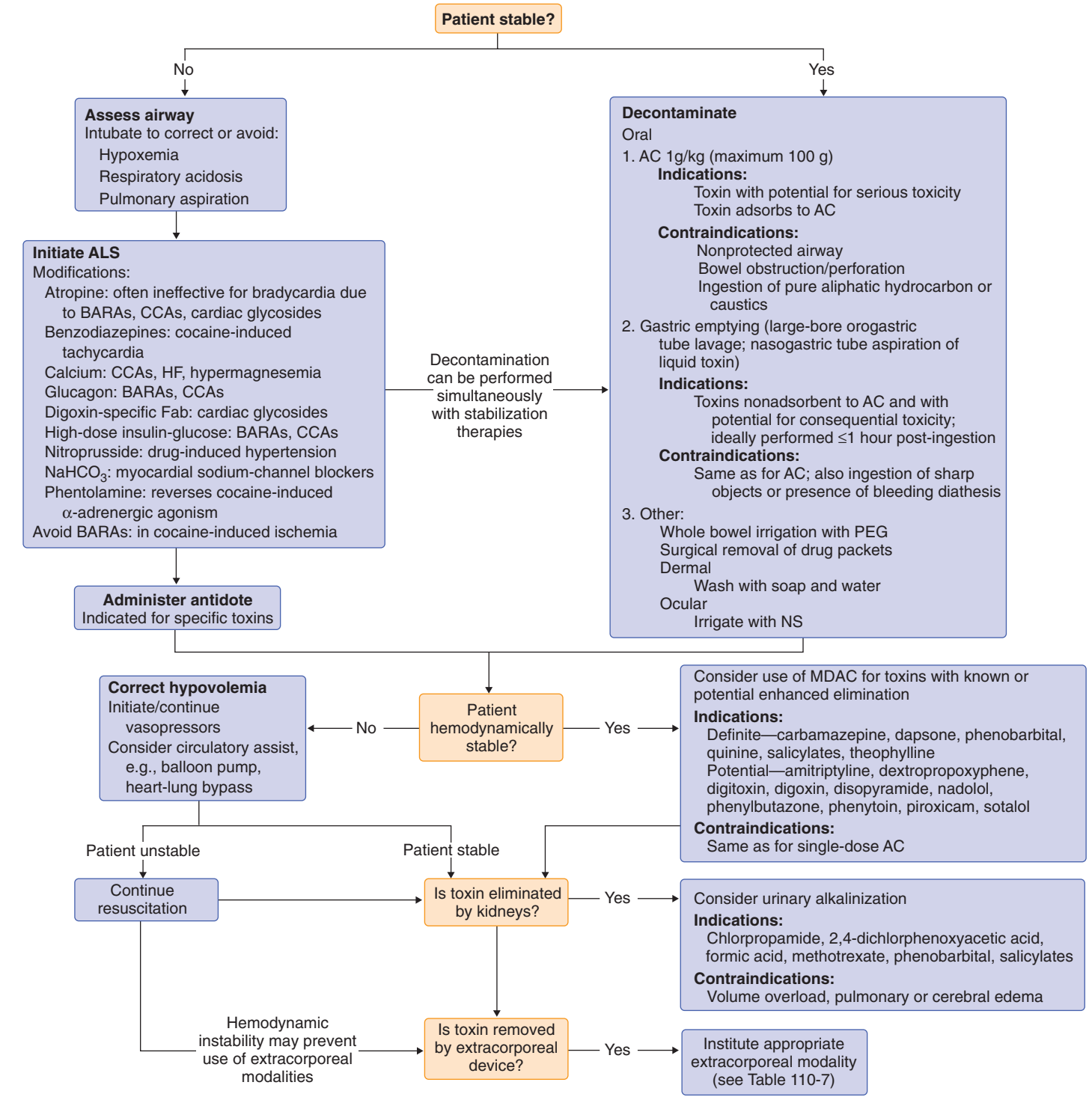

FIGURE 110-2. Algorithm for the management of acute poisoning. AC = activated charcoal; ALS = advanced life support; BARAs = β-adrenergic receptor antagonists; CCAs = L-type calcium-channel antagonists; HF = hydrofluoric acid; MDAC = multidose activated charcoal; NS = 0.9% saline solution; PEG = nonabsorbable polyethylene glycol solution.

Rhabdomyolysis has been reported with adrenergic agents, doxylamine, phencyclidine, heroin, *Tricholoma equestre* mushrooms, and envenomation by crotaline snakes, scorpions, or widow spiders (*Latrodectus* sp); short-acting nondepolarizing agents, such as vecuronium and rocuronium, are preferable in these cases.

Advanced Life Support

Standard emergency cardiovascular care algorithms (Chapter 63) must be modified for effects caused by specific poisons. Atropine often does not reverse bradycardia secondary to β-adrenergic receptor antagonists, L-type calcium-channel antagonists, or cardiac glycosides, and it may actually impair the ability to do adequate gastrointestinal decontamination. In these cases, more specific therapy with intravenous calcium (calcium-channel antagonists), high doses of glucagon (β-adrenergic receptor antagonists, calcium-channel antagonists), or digoxin-specific Fab antibody (cardiac glycosides) is indicated. High-dose insulin-glucose therapy can successfully reverse myocardial depression and conduction abnormalities in humans poisoned with β-adrenergic receptor antagonists and calcium-channel antagonists. Intravenous sodium bicarbonate may reverse cardiac conduction delays caused by antiarrhythmic drugs with sodium-channel blockade recovery rates of greater than 1 second (Vaughn-Williams classification IA and IC), cocaine, cyclic antidepressants, diphenhydramine, and quinine. β-Adrenergic receptor antagonists are contraindicated in patients with cocaine-induced myocardial syndromes because they can result in unopposed α-adrenergic–mediated vasoconstriction, but phentolamine can reverse the agonistic effects of cocaine on α-adrenergic receptors. Benzodiazepines can reverse significant sinus tachycardia from sympathomimetic agents. Calcium may also be life-saving in systemic hydrofluoric acid poisoning and severe hypermagnesemia, and it is indicated for symptomatic hypocalcemia caused by ethylene glycol toxicity. Drug-induced hypertension may be transitory; nitroprusside should be used if treatment is clinically indicated. In patients with toxicant-induced circulatory collapse refractory to maximal therapy, including vasopressors, circulatory assist devices may support the patient until sufficient toxicant is eliminated (Chapter 107).

Decontamination

Activated Charcoal

Single-dose activated charcoal without prior gastric emptying has been the preferred method of treatment for the ingestion of substances that have the potential to cause moderate to life-threatening toxicity and are known to adsorb to activated charcoal. The absence of clinical signs and symptoms does not preclude administration of activated charcoal because drug absorption and toxicity can be delayed. Activated charcoal can also be administered when the ingested toxicant cannot be identified but significant toxicity is a concern. Activated charcoal consists of pyrolysis products that have been specially cleaned to produce an internal pore structure to which substances can adsorb, thereby limiting their systemic absorption. Activated charcoal can be administered with antiemetic drugs or given through a nasogastric tube, when necessary. The oral dose is approximately 1 g/kg body weight, with a maximum single dose of 100 g. Efficacy in preventing toxicant absorption declines with time, so activated charcoal should be given as soon as possible after ingestion. However, the documented efficacy of activated charcoal for reducing toxicant blood levels has not translated into reduced mortality in reports[4] or in randomized trials.■■ The decision to administer activated charcoal should be based on a risk/benefit assessment that includes nature of the exposure, clinical effects displayed during evaluation, and abilities of the medical facility and staff. For patients likely to have a good outcome, the risk and effort associated with activated charcoal administration are not worthwhile. Its use is justified in patients who present early (1 to 2 hours) after exposures to a large amount of a concerning toxin that is likely to be adsorbed to charcoal. Activated charcoal should not be used in patients at risk for aspiration until the airway is secure to minimize aspiration; the patient's head should also be elevated unless it is contraindicated. Activated charcoal is contraindicated in patients with a perforated bowel, functional or mechanical bowel obstruction, ingestion of a pure aliphatic hydrocarbon such as gasoline or kerosene (no benefit and increased risk for aspiration), and ingestion of caustic acid and alkali (no benefit and obscures endoscopy). Certain agents, such as lithium, iron, metals, and ethanol, do not adsorb significantly to activated charcoal, but its use is not precluded if the patient has ingested other toxicants that do adsorb to activated charcoal. Pulmonary aspiration and bowel obstruction from inspissated activated charcoal are the most common complications; both occur more frequently when multidose activated charcoal is administered, but they can be avoided by withholding treatment in patients who have suboptimal bowel function or decreased fecal elimination.

Gastric Emptying

Two methods of gastric emptying, syrup of ipecac[5] and orogastric lavage through a large-bore tube, are no longer routinely used. Both are relatively ineffective therapies that potentially increase the risk for aspiration. No well-designed study has documented any benefit of gastric emptying, either by lavage or by syrup of ipecac, compared with the use of activated charcoal alone. Gastric emptying by lavage or, rarely, by syrup of ipecac may be of benefit and should be performed in patients who have ingested toxicants that do not adsorb to activated charcoal and are known to produce significant morbidity or for which aggressive decontamination may offer the best chance for survival (e.g., colchicine, sodium azide, sodium fluoroacetate). Removal of a liquid toxicant, such as ethylene glycol, may be accomplished by aspiration of gastric contents through a nasogastric tube. Contraindications to gastric emptying include those for activated charcoal, a bleeding diathesis, and the ingestion of sharp objects. Placement of an endotracheal tube before gastric lavage may be necessary to protect the airway in patients who have a decreased level of consciousness and impaired gag reflex but is not required in all cases. Major complications of gastric emptying include pulmonary aspiration, esophageal tears and perforations, and laryngospasm (with lavage).

Whole Bowel Irrigation

Whole bowel irrigation with a nonabsorbable polyethylene glycol solution has been recommended for iron and sustained-release medications, for agents not adsorbed to activated charcoal, and for body packers (smugglers who swallow packets of illicit drugs). The most common complication is vomiting, and whole bowel irrigation is contraindicated in patients with bowel perforation, obstruction, hemorrhage, or hemodynamic or respiratory instability. The initial recommended dose is 500 mL/hour given orally or by nasogastric tube, with titration to 2000 mL/hour as tolerated; treatment continues until the rectal effluent clears. Rarely, surgery may be necessary to remove packets in smugglers who have symptoms of cocaine toxicity or are obstructed; endoscopic removal of these packets should never be attempted because of the risk of packet rupture.

Antidotes

Few toxicants have specific therapies (Table 110-6). Although antidotes may be essential in treating patients exposed to certain toxicants, their use does not preclude the need for ongoing supportive care and, in some cases, extracorporeal elimination.

Enhanced Elimination

Methods to accelerate the elimination of toxicants or drugs from the body include multiple doses of activated charcoal, urinary alkalinization, and extracorporeal removal. Another method, using the oral ion exchange resins sodium polystyrene sulfonate and cholestyramine, has experimentally enhanced the elimination of lithium, digoxin, digitoxin, and organochlorines but has limited clinical usefulness.

Multiple Doses of Oral Activated Charcoal

The rationale for administering multiple doses of oral activated charcoal includes the adsorption of any toxic agent remaining in the gastrointestinal tract (e.g., sustained-release drugs or drugs that retard their absorption, such as anticholinergics); interference with the enterohepatic and enteroenteric recirculation of toxicants; and enhancement of the elimination of drugs with a long half-life, a volume of distribution less than 1 L/kg body weight, and low protein binding (termed gastrointestinal dialysis). The existing evidence shows enhanced elimination of carbamazepine, dapsone, phenobarbital, quinine, salicylates, and theophylline, but multiple doses of activated charcoal may also be effective for amitriptyline, dextropropoxyphene, digitoxin, digoxin, disopyramide, nadolol, phenylbutazone, phenytoin, piroxicam, and sotalol. Whether enhanced elimination provided by repeated doses of activated charcoal translates into decreased morbidity and mortality has not been adequately examined in large controlled clinical trials, except for yellow oleander and organophosphate ingestion, for which it has shown no benefit.■ The usual recommendations are an average dose of 12.5 g of activated charcoal (after the initial dose of 1 g/kg body weight, with a maximum single dose of 100 g) administered every 4 to 6 hours after the previous dose. The contraindications to single-dose activated charcoal also apply to multidose activated charcoal. Reported complications include pulmonary aspiration, bowel obstruction from inspissated charcoal, and fluid and electrolyte imbalance from multiple doses of a simultaneously administered cathartic.

Urinary Alkalinization

Alkalinization of the urine, which increases the renal elimination of weak acids, is used primarily to enhance the elimination of salicylates, but the elimination of chlorpropamide, 2,4-dichlorophenoxyacetic acid, formic acid, methotrexate, and phenobarbital may be increased with this method. Urinary alkalinization is accomplished by an intravenous bolus of 1 to 2 mEq of sodium bicarbonate per kilogram body weight, followed by three ampules (150 mL) of sodium bicarbonate (44 mEq/50 mL) in 850 mL of 5% dextrose in water infused at two to three times the normal maintenance fluid rate. Urinary pH should be checked hourly, and the infusion should be adjusted to maintain a urine pH of 7.5 to 8.0. Potassium should be administered simultaneously to avoid hypokalemia, which prevents urinary alkalinization because the distal tubule excretes hydrogen ion in exchange for potassium (Chapters 116 and 117). Serum pH should be monitored and kept at 7.55 or lower to avoid excessive alkalemia. Contraindications to this therapy include volume overload and cerebral or pulmonary edema. Urinary acidification is not recommended to enhance the elimination of weak bases, such as amphetamines, because of the danger of precipitating tubular myoglobin in patients with rhabdomyolysis.

Extracorporeal Removal

Extracorporeal techniques enhance the elimination of a few drugs and toxicants, especially those that exist in the blood and are otherwise poorly eliminated. Such drugs generally have single-compartment kinetics, a volume of distribution less than 1 L/kg, and endogenous clearance of less than 4 mL/minute/kg (Table 110-7). For hemodialysis, the toxicant must be water soluble, have a molecular weight less than 500, and exhibit low protein binding. For continuous renal replacement therapy, the toxicant must have a molecular weight less than the permeability limit of the filter membrane. As a group, these latter forms of elimination are slow and likely to be of limited benefit to most patients with acute poisoning.[6] Rarely, extracorporeal removal has been used for aminoglycosides, atenolol, bromide, carbamazepine, diethylene glycol, isopropanol, magnesium, metformin, methotrexate, *N*-acetylprocainamide, phenobarbital, procainamide, sotalol, and trichloroethanol (chloral hydrate).

TABLE 110-6 ANTIDOTES AND INDICATIONS FOR USE

ANTIDOTE	INDICATION FOR USE	DOSE*	TREATMENT END POINT	COMMENTS
Antivenom, *Crotalidae* (Fab)†	Crotaline snake (e.g., rattlesnakes, copperhead)	4-6 vials; repeat for persistent or worsening clinical condition; repeated doses of 2 vials at 6, 12, and 18 hours after initial antivenom dose are recommended	Halt in progression of circumferential and proximal swelling Resolving systemic effects	Better safety profile than historical equine-derived antivenom Repetitive dosing indicated for recurrent soft tissue swelling Less effective at correcting hematologic (i.e., coagulation and platelet) disorders
Antivenom, *Latrodectus* (equine)†	Black widow spider (*Latrodectus* sp)	1 vial diluted in 50-100 mL NS, infused during 1 hour; can repeat	Resolution of symptoms, vital signs normal	Dilution and slow infusion rate are **critical** to avoid anaphylactoid reaction Indications include severe pain unresponsive to opioids and severe hypertension Serum sickness can occur IV calcium is ineffective
Atropine	Carbamates Nerve agents Organophosphorus compounds	2 mg IV; double the dose every 5 minutes to achieve atropinization and hemodynamic stability; then start continuous infusion of 10-20% of total stabilizing dose per hour	Cessation of excessive oral and pulmonary secretions, >80 beats/min, systolic blood pressure >80 mm Hg	Doubling of the dose every 5 minutes (e.g., 2 mg, 4 mg, 8 mg, 16 mg) estimated to achieve atropinization within 30 minutes Stop infusion when patient develops concerning signs or symptoms of anticholinergic toxidrome (see Table 110-1); restart infusion at lower rate when signs or symptoms abate
Calcium salt†	Calcium-channel antagonists	Calcium chloride 10%, 10 mL (1 g) during 10 minutes; can be given in 1 minute if critically ill Calcium gluconate 10%, 30 mL (3 g) during 10 minutes; can be given in 1 minute if critically ill	Reversal of hypotension; may not reverse bradycardia	*All indications:* Monitor ionized calcium levels IV extravasation causes tissue necrosis, especially with calcium chloride Can administer at faster than stated rates for immediate life-threatening conditions (i.e., in 1 minute) Calcium chloride contains three times more elemental calcium than calcium gluconate does
	Hydrofluoric acid	Systemic toxicity: calcium gluconate 10%, 1-3 g (10-30 mL) per dose IV during 10-minute period; repeat as needed every 5-10 minutes	Reversal of life-threatening manifestations of hypocalcemia and hyperkalemia	Can dilute and give intra-arterially or IV with a Bier block for extremity exposures and burns
	Hyperkalemia (except cardiac glycosides)	Calcium gluconate 10%, 1 g (10 mL) per dose IV during 10-minute period; repeat as needed every 5-10 minutes	Reversal of myocardial depression and conduction delays	May precipitate ventricular arrhythmias
	Hypermagnesemia	Calcium gluconate 10%, 1-2 g (10-20 mL) per dose IV during 10-minute period; repeat as needed every 5-10 minutes	Reversal of respiratory depression, hypotension, and cardiac conduction blocks	Simultaneous therapies to increase magnesium elimination should be instituted
	Hypocalcemia (e.g., ethylene glycol)	Calcium gluconate 10%, 0.5-1.0 g (5-10 mL) per dose during 10-minute period; repeat as needed every 10 minutes	Reversal of tetany	Correct symptomatic hypocalcemia; avoid excessive administration that may increase production of calcium oxalate crystals in ethylene glycol poisoning
L-Carnitine	Valproate-induced hyperammonemia or hepatotoxicity	100 mg/kg (maximum 6 g) IV during 30 minutes, then 15 mg/kg IV during 30-minute period q4h (maximum 6 g/day)	Treat until clinical improvement occurs	Levocarnitine is active form Adjust dose for end-stage renal disease
Cyanide antidote kit Amyl nitrite Sodium nitrite Sodium thiosulfate [Hydroxocobalamin is preferred if available, see below]	Cyanide	Amyl nitrite: 0.3-mL pearls, crush and inhale during 30-second period Sodium nitrite 3%: 10 mL IV during 10-minute period Sodium thiosulfate 25%: 50 mL (12.5 g) IV during 10-minute period	Resolution of lactic acidosis and moderate to severe clinical signs and symptoms: seizures, coma, dyspnea, apnea, hypotension, bradycardia	Coordinate amyl nitrite with continued oxygenation and give only until sodium nitrite infusion is begun; nitrites may produce hypotension and excess methemoglobinemia Sodium nitrite dose must be adjusted if patient has hemoglobin <12 g/dL Sodium thiosulfate dosing can be repeated

TABLE 110-6 ANTIDOTES AND INDICATIONS FOR USE—cont'd

ANTIDOTE	INDICATION FOR USE	DOSE*	TREATMENT END POINT	COMMENTS
Deferoxamine	Iron	Initiate at 5 mg/kg/hr, titrate to 15 mg/kg/hr IV (maximum 8 g/day) Mild to moderate: administer for 6-12 hours Severe toxicity: administer 24 hours	Resolution of clinical signs and symptoms Do not use urine color, which is an unreliable marker for iron clearance Laboratory testing is unreliable while antidote is being received	Indications: symptomatic patients with lethargy, severe abdominal pain, hypovolemia, acidosis, shock; any symptomatic patient with peak serum iron level >350 g/dL Prolonged therapy can cause pulmonary toxicity
Digoxin-specific antibody fragments (Fab)	Digoxin Digitalis and related plants (e.g., oleander, lily of the valley) Other cardiac glycosides (e.g., bufadienolides [Bufo toads])	Unknown digoxin dose or serum level, or for plant or toad source: acute toxicity—10-20 vials; chronic toxicity—3-6 vials Digoxin dose known: number of vials = (mg ingested × 0.8) ÷ 0.5 Digoxin serum level known: number of vials = [serum level (ng/mL) × weight (kg)] ÷ 100 Infuse dose during 30 minutes	Resolution of hyperkalemia, symptomatic bradydysrhythmias, ventricular arrhythmias, Mobitz II or third-degree heart block	Each vial binds 0.5 mg of digoxin or digitoxin Monitor ECG and potassium levels Digoxin serum levels unreliable after antidote administered unless test is specific for free serum digoxin
Dimercaprol (BAL)	Arsenic Lead Mercury, elemental and inorganic salts	Arsenic: 3-5 mg/kg IM q4h Lead: 75 mg/m^2 (4 mg/kg) IM q4h for 5 days Inorganic mercury: 5 mg/kg IM, then 2.5 mg/kg IM q12h for 10 days or until patient is clinically improved	Arsenic: 24-hour urinary arsenic <50 µg/L Lead: encephalopathy resolved, blood lead level <100 µg/dL, and succimer therapy can be started Mercury, elemental and inorganic: 24-hour urinary mercury <20 µg/L	Formulated in peanut oil; painful IM injection and caution with allergy Maximum adult dose is 3 g/day BAL started 4 hours before initiation of concomitant $CaNa_2EDTA$ for lead encephalopathy Dosing not well established for arsenic and elemental or inorganic mercury toxicity; not used for organic mercury poisoning Adverse effects: painful injections, fever, diaphoresis, agitation, headache, salivation, nausea and vomiting, hemolysis in G6PD-deficient patients, chelation of essential metals Check essential metal levels if chelation is prolonged Succimer is replacing BAL for many indications except lead encephalopathy Treatment end points for arsenic and mercury include improving clinical condition
Edetate calcium disodium ($CaNa_2EDTA$)	Lead	1500 mg/m^2/24 hr (maximum 3 g) by continuous infusion	Treat for 5 days, followed by 2-day hiatus; repeat until encephalopathy resolved, lead level <100 µg/dL, and succimer therapy can be started	Use in patients with lead encephalopathy or lead level >100 g/dL Administer BAL 4 hours before initiating $CaNa_2EDTA$ Hydrate patient and establish good urinary output before starting therapy Avoid thrombophlebitis by diluting in NS or D_5W to a concentration ≤0.5% Substitution of Na_2EDTA can cause fatal hypocalcemia
Flumazenil	Benzodiazepines	0.1 mg/min IV to a total dose of 1 mg	Reversal of coma	Limit use to reversal of inadequate ventilation in benzodiazepine-toxic patients Acute benzodiazepine withdrawal may occur in patients dependent on benzodiazepines Increases intracranial pressure and risk for seizures in presence of underlying seizure disorder or ingestion of seizure-producing toxicants Monitor for resedation up to 2 hours after last dose
Folinic acid (tetrahydrofolic acid [leucovorin])	Methanol Methotrexate	Methanol: 50 mg IV q4h Methotrexate: 100 mg/m^2 IV during 15-30 minutes q3-6h; dosing lower when used as "rescue" in chemotherapy	Methanol: methanol undetectable, metabolic acidosis cleared Methotrexate: serum level <1 × 10^{-8} mol/L	Essential therapy for both toxicants Little concern with excessive dosing when used for methotrexate overdose Methotrexate: large overdoses may require increased dose Glucarpidase administered 2-4 hours before or after folinic acid

TABLE 110-6 ANTIDOTES AND INDICATIONS FOR USE—cont'd

ANTIDOTE	INDICATION FOR USE	DOSE*	TREATMENT END POINT	COMMENTS
Fomepizole[8]	Ethylene glycol Methanol	Dose 1: 15 mg/kg IV Doses during next 48 hours: 10 mg/kg IV All subsequent doses: 15 mg/kg IV Administer q12h, except when HD performed: HD initiation: ½ next dose if >6 hours since last dose HD ongoing: q4h End of HD (based on time of last dose): <1 hour, no dose; 1-3 hours, ½ next dose; >3 hours, next dose	For both: serum level <25 mg/dL and metabolic acidosis resolved	Start immediately if toxic alcohol suspected, without waiting for confirmatory levels Dose amount is not affected by interval timing of doses
Glucagon	β-Adrenergic receptor antagonists Calcium-channel antagonists	Bolus of 3-5 mg IV; can repeat to achieve clinical effect, then infusion of 2-10 mg/hr	Reversal of hypotension and bradycardia; taper infusion	Can precipitate vomiting; be prepared to protect airway Mild hyperglycemia occurs Maximum dosing amounts unknown; bolus doses up to 30 mg reported Duration of effect is 15 minutes; thus infusion must be started immediately
Hydroxocobalamin	Cyanide	Initial: 5 g IV during 15-minute period Second dose: 5 g IV during 15 minutes to 2 hours; maximum total dose is 10 g Follow each hydroxocobalamin dose with sodium thiosulfate 25%: 50 mL (12.5 g) IV during 10-minute period	Resolution of lactic acidosis and moderate to severe clinical signs and symptoms: seizures, coma, dyspnea, apnea, hypotension, bradycardia	Can be administered IV push if patient is in cardiac arrest Do not give hydroxocobalamin and sodium thiosulfate through the same IV line Adverse effects: red discoloration of plasma, urine, mucous membranes, skin; transient hypertension Interference with laboratory colorimetric assays: Levels increased: bilirubin; creatinine; glucose; hemoglobin; magnesium; co-oximetry total Hb, COHb%, MetHb% Levels decreased: AST, ALT, creatinine, co-oximetry O_2Hb%
Hyperbaric oxygen (HBO)	Carbon monoxide Experimental: carbon tetrachloride, cyanide, hydrogen sulfide	3.0 atm pressure for 60 minutes (25 minutes O_2, 5 minutes air, 25 minutes O_2, 5 minutes air), then 2.0 atm for 65 minutes (30 minutes O_2, 5 minutes air, 30 minutes O_2), then "surface" to 1.0 atm	One treatment Repeated treatment controversial	Carbon monoxide: treatment protocols may vary HBO indicated for loss of consciousness; seizures; cerebellar dysfunction; impaired cognition; headache, nausea/vomiting persisting after 4 hours of O_2 therapy regardless of carboxyhemoglobin level
Insulin-glucose	Calcium-channel antagonists β-Adrenergic receptor antagonists	Regular insulin, 1 U/kg bolus, followed by 0.5-1 U/kg/hr Titrate 50% dextrose IV to avoid hypoglycemia	Reversal of myocardial depression	Initiate early to reverse myocardial depression Monitor glucose and potassium; hypoglycemia can occur during and after therapy Hyperglycemia results from calcium-channel antagonist–induced insulin resistance, and initial dextrose requirements may be less than anticipated Recovery may be heralded by normalization of glucose levels, with increased dextrose required to avoid hypoglycemia
Lipid emulsion	Cardiac toxicity from local anesthetics (e.g., bupivacaine, ropivacaine) Experimental: verapamil, diltiazem, tricyclic antidepressants, bupropion, propranolol and other lipophilic toxicants	Use 20% formulation Initial bolus: 1.5 mL/kg IV during 1 minute, followed immediately by infusion of 0.25 mL/kg/min for 30-60 minutes Can repeat bolus for asystole	Return of hemodynamic stability	Use for other than bupivacaine based on animal experiments and human case reports; numerous dosing regimens have been used Use if advanced life support measures fail; continue CPR as needed during drug administration
Methylene blue	Methemoglobin-producing agents	1-2 mg/kg body weight (0.1-0.2 mL/kg of 1%) methylene blue during 5-minute period; repeat dose for persistent or recurrent symptoms or signs	Resolution of dyspnea and altered mental status	Use if patient is symptomatic (i.e., dyspneic, altered mental status) Maximum dose should not exceed 7 mg/kg (0.7 mL/kg) Contraindicated in G6PD-deficient patients; may cause hemolysis Some toxicants (e.g., dapsone) may require prolonged therapy

TABLE 110-6 ANTIDOTES AND INDICATIONS FOR USE—cont'd

ANTIDOTE	INDICATION FOR USE	DOSE*	TREATMENT END POINT	COMMENTS
N-Acetylcysteine (NAC)[7]	Acetaminophen Experimental: carbon tetrachloride, chloroform, pennyroyal oil	Oral (total 72 hours) Load: 140 mg/kg Maintenance (starting 4 hours after load): 70 mg/kg q4h × 17 doses IV (total 21 hours) Load: 150 mg/kg during 1-hour period Maintenance infusion: 12.5 mg/kg/hr during 4-hour period, then 6.25 mg/kg/hr over 16 hours as continuous infusion	At the end of therapy, repeat AST and APAP levels: if AST normal and APAP not detected, treatment complete; if AST normal and APAP detected, continue NAC; if AST elevated, continue NAC After patient has received full course of NAC therapy, if INR ≥2.0 or severe hepatotoxicity present (AST > 1000 U/L), continue NAC until INR <2.0 and aminotransferases normalize	Most effective if initiated within 8 hours after ingestion; may be started any time after ingestion and is beneficial in severe hepatotoxic states Longer duration of treatment may be required in patients with hepatotoxicity Treatment end points simplified for ease of use INR result not valid indicator if FFP recently administered
Naloxone[9]	Opioids	Opioid dependence possible: 40-50 µg (0.04-0.05 mg) IV titrated upward to reversal, while avoiding withdrawal if concerns for opioid dependence Opioid dependence not likely: 0.4 mg by any route and titrate up to 10 mg Continuous infusion: establish bolus dose required to reverse respiratory depression; begin infusing two thirds of reversal dose every hour and titrate to maintain adequate respirations; repeated bolus with half of reversal dose 15 minutes after reversal of respiratory depression	Initial: reversal of respiratory depression with resolution of hypoxia and hypercapnia Final: resolution of CNS and respiratory depression	Pre-ventilate patients with respiratory depression by bag-valve-mask or intubation before administration Use smaller doses in opioid-dependent patients Some opioids (e.g., buprenorphine) may require larger doses of naloxone Use continuous infusion for recurrent symptoms and prolonged action of some formulations (e.g., sustained-release morphine, methadone) Re-sedation can occur Do not use naltrexone to reverse acute toxicity
Octreotide	Sulfonylurea-induced hypoglycemia	50 mg SC q6h	Resolution of hypoglycemia and dextrose not required	Maintain dextrose infusion as needed Not for insulin-induced hypoglycemia
Physostigmine	Anticholinergic agents (e.g., diphenhydramine, jimsonweed [*Datura* sp], scopolamine)	1-2 mg IV during 5-minute period; can repeat in 5 minutes if no effect and cholinergic effects do not occur	Reversal of anticholinergic effects	Duration of effect is 60-90 minutes Benzodiazepine used for subsequent treatment of agitation and seizures; additional physostigmine used rarely (e.g., refractory seizures or agitation) Adverse effects include seizures, excessive oral secretions, bradyarrhythmias; contraindicated in cyclic antidepressant toxicity
Pralidoxime chloride	Organophosphorus compounds Nerve agents: sarin, VX	30 mg/kg IV bolus (maximum 2 g) during 30 minutes, followed by continuous infusion of 8-10 mg/kg/hr (maximum 650 mg/hr)	Resolution of signs and symptoms, atropine no longer required	Can give initial dose during 2-minute period for life-threatening clinical effects Administer early when diagnosis known or strongly suspected Efficacy variable, depending on the organophosphate Fat-soluble organophosphates may require prolonged treatment
Pyridoxine	Ethylene glycol Isoniazid Monomethylhydrazine (*Gyromitra esculenta* mushrooms)	100 mg IV 5 g IV, repeat for refractory seizures	One dose Resolution of seizures	Efficacy theoretical for ethylene glycol to enhance elimination of toxic metabolites Pyridoxine may be required even with benzodiazepines to stop seizures, but patient can remain comatose (isoniazid, *Gyromitra* mushrooms) Excessive dosing can cause neuropathy
Sodium bicarbonate ($NaHCO_3$)	Reversal of myocardial sodium-channel blockers (e.g., cyclic antidepressants, cocaine, sodium-channel–blocking antiarrhythmics with $\tau_{recovery}$ >1 second, piperidine phenothiazines (thioridazine, mesoridazine)	1-2 mEq $NaHCO_3$/kg by intermittent bolus; repeat as needed	Narrowing of prolonged QRS, resolution of ventricular arrhythmias, reversal of hypotension	Monitor blood pH (optimal pH approximately 7.50); avoid pH > 7.55

TABLE 110-6 ANTIDOTES AND INDICATIONS FOR USE—cont'd

ANTIDOTE	INDICATION FOR USE	DOSE*	TREATMENT END POINT	COMMENTS
Sodium bicarbonate ($NaHCO_3$) (cont'd)	Altered tissue distribution or enhanced elimination of salicylates; may be used for chlorophenoxy herbicides, formic acid, methotrexate, phenobarbital	1-2 mEq $NaHCO_3$/kg, followed by 3 ampules (150 mL) $NaHCO_3$ (44 mEq per 50 mL) in 850 mL of D_5W, infused at 2-3 times normal maintenance fluid rate	Serum salicylate <30 mg/dL and patient clinically stable	Target blood pH: 7.50-7.55 Monitor urinary pH hourly; adjust infusion to maintain urine pH of 7.5-8.0 (avoid blood pH >7.55) Monitor ABGs Maintain normokalemia
Succimer (DMSA)	Arsenic Lead Mercury, all forms	10 mg/kg/dose q8h for 5 days, followed by q12h for 14 days Drug holiday for 2 weeks; repeat if treatment end point not reached	Arsenic: 24-hour urinary arsenic <50 µg/L Lead: resolution of encephalopathy, gastrointestinal symptoms, neuropathy, nephropathy, arthralgias, myalgias, and blood lead level <70 µg/dL Mercury, elemental and inorganic: 24-hour urinary mercury <20 µg/L Mercury, organic: end point not well established	Oral chelator; adverse effects include rash, transient AST and alkaline phosphatase elevations, and gastrointestinal distress; minimal chelation of essential metals occurs Dosing for arsenic and mercury not well established Therapeutic end point for organic mercury not established; neurotoxicity not responsive to chelation therapy; suggest chelation until blood mercury level within normal value range for reference laboratory
Vitamin K	Vitamin K antagonist anticoagulants (e.g., warfarin, long-acting anticoagulant rodenticides [LAARs]) [note this is not for normalizing a supratherapeutic INR in patients prescribed Warfarin]	Subcutaneous: AquaMEPHYTON (K_1), 10-25 mg, repeat every 6-12 hours until oral vitamin K_1 started Oral: 25-50 mg q6h; larger doses may be required	INR is normal 48-72 hours after stopping vitamin K_1 therapy Can also monitor factor VII activity	Anaphylactoid reaction can occur with rapid IV administration Severe bleeding may also require FFP, prothrombin protein concentrate (off label), or factor concentrates Base decision to treat on finding of elevated INR; do not administer prophylactic vitamin K_1 Oral therapy may be required for months with LAAR poisoning because of lipophilicity of toxicant, with slow body clearance

*Dose concentrations and infusion times are not given. Drug dosages may require adjustment in patients with renal or hepatic failure.
†Administer antivenom in a monitored setting; antivenom must be reconstituted and then diluted; initially infuse at a rate of 2 to 5 mL/hr, and double the infusion rate every 5 minutes as tolerated to administer antivenom during a 1-hour period.
‡Ten percent calcium chloride solution = 100 mg/mL (27.2 mg/mL elemental calcium); 10% calcium gluconate solution = 100 mg/mL (9 mg/mL elemental calcium).
ABGs = arterial blood gases; ALT = alanine aminotransferase; APAP = acetyl-*p*-aminophenol (acetaminophen); AST = aspartate aminotransferase; BAL = British antilewisite; CNS = central nervous system; COHb% = percentage carboxyhemoglobin; CPR, cardiopulmonary resuscitation; D_5W = 5% dextrose in water; DMSA = 2,3-dimercaptosuccinic acid; ECG = electrocardiogram; FFP = fresh-frozen plasma; G6PD = glucose-6-phosphate dehydrogenase; Hb = hemoglobin; HD = hemodialysis; INR = international normalized ratio; MetHb% = percentage methemoglobinemia; NS = normal saline; O_2Hb% = percentage oxyhemoglobin; $\tau_{recovery}$ = drug blockade recovery rate.

TABLE 110-7 COMMON TOXICANTS REMOVED BY HEMODIALYSIS

TOXICANT	INDICATIONS	COMMENTS
Ethylene glycol	Serum level ≥50 mg/dL, or lower levels with concomitant metabolic acidosis and evidence of end-organ toxicity	Not routinely required in a patient with normal creatinine clearance and acid-base status who is receiving fomepizole
Lithium*	Clinical indications	Clinical indication is CNS toxicity (e.g., decreased mental status, ataxia, coma, seizures) Use dialysate containing bicarbonate to decrease Na^+/K^+ antiporter intracellular sequestration of lithium
Methanol	Serum level ≥50 mg/dL, or lower levels with concomitant metabolic acidosis and evidence of end-organ toxicity	Usually required because of slow elimination half-life in presence of fomepizole (mean, 52 hours; range, 22-87 hours), even in patients with no metabolic acidosis or evidence of end-organ toxicity
Phenobarbital	Clinical indications	Rarely necessary except when a patient is hemodynamically unstable despite aggressive support
Salicylates	Acute toxicity: serum level ≥100 mg/dL without clinical abnormality or <100 mg/dL in the presence of a clinical indication Chronic toxicity: any clinical indication	Serum protein binding decreases with increasing toxic levels, increasing the amount of free salicylate available for HD removal; clinical indications are one or more of the following: altered mental status, seizures, pulmonary edema, intractable acidosis, renal failure
Valproic acid	Severe intoxication with serum concentration >850 mg/L	Clinical indications include hepatic dysfunction; coma, especially with hyperammonemia; deteriorating clinical status despite aggressive support

*Hemodiafiltration removes lithium; the clinical benefit of this technique is unknown.
CNS = central nervous system; HD = hemodialysis.

PROGNOSIS

Almost all patients who reach the hospital alive survive with appropriate care. Inpatient mortality rates are 0.2 to 0.5%.

Grade A References

A1. Eddleston M, Juszczak E, Buckley NA, et al. Multiple-dose activated charcoal in acute self-poisoning: a randomised controlled trial. *Lancet*. 2008;371:579-587.

A2. Cooper GM, Le Couteur DG, Richardson D, et al. A randomized clinical trial of activated charcoal for the routine management of oral drug overdose. *QJM*. 2005;98:655-660.

A3. Buckley NA, Eddleston M, Li Y, et al. Oximes for acute organophosphate pesticide poisoning. *Cochrane Database Syst Rev*. 2011;2:CD005085.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Jones CM, Mack KA, Paulozzi LJ. Pharmaceutical overdose deaths, United States, 2010. *JAMA*. 2013;309:657-659.
2. Koylu R, Dundar ZD, Koylu O, et al. The experiences in a toxicology unit: a review of 623 cases. *J Clin Med Res*. 2014;6:59-65.
3. Mowry JB, Spyker DA, Cantilena LR Jr, et al. 2012 Annual report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 30th Annual Report. *Clin Toxicol (Phila)*. 2013;51:949-1229.
4. Merigian KS, Blaho KE. Single-dose oral activated charcoal in the treatment of the self-poisoned patient: a prospective, randomized, controlled trial. *Am J Ther*. 2002;9:301-308.
5. Höjer J, Troutman WG, Hoppu K, et al. Position paper update: ipecac syrup for gastrointestinal decontamination. *Clin Toxicol (Phila)*. 2013;51:134-139.
6. Garlich FM, Goldfarb DS. Have advances in extracorporeal removal techniques changed the indications for their use in poisonings? *Adv Chronic Kidney Dis*. 2011;18:172-179.
7. Hodgman MJ, Garrard AR. A review of acetaminophen poisoning. *Crit Care Clin*. 2012; 28:499-516.
8. Brent J. Fomepizole for ethylene glycol and methanol poisoning. *N Engl J Med*. 2009;360: 2216-2223.
9. Boyer EW. Management of opioid analgesic overdose. *N Engl J Med*. 2012;367:146-155.

REVIEW QUESTIONS

1. A patient presents to the emergency department with a metabolic acidosis after exposure to a toxin. For which one of the following toxic etiologies is hemodialysis indicated?

A. Acetaminophen
B. Aspirin
C. Carbon monoxide
D. Cyanide
E. Isopropyl alcohol

Answer: B Acetaminophen can cause a metabolic acidosis immediately after massive overdose or after hepatic failure develops. *N*-Acetylcysteine is the treatment, in addition to supportive care. Carbon monoxide and cyanide both inhibit oxidative phosphorylation and cause an elevated lactate level. Accepted treatments include hyperbaric oxygen therapy and hydroxocobalamin, respectively. Isopropanol does not cause a significant metabolic acidosis, and hemodialysis is rarely needed. Aspirin can uncouple oxidative phosphorylation and cause an anion gap metabolic acidosis; treatment for severe clinical effects of aspirin poisoning should include hemodialysis.

2. A middle-aged patient who is enrolled in a methadone-maintenance program presents to the emergency department with altered mental status. On physical examination, he is deeply obtunded and has a respiratory rate of 4/min with a pulse oximeter saturation reading of 85%. Which one of the following is the optimal initial approach?

A. Bag-valve-mask ventilation
B. Intubation and ventilation
C. Irritant stimulation
D. Naloxone, 2 mg IV
E. Nasal cannula oxygen

Answer: A There is no rush to reverse the opioid intoxication with naloxone in patients who are in a health care setting because only oxygenation and ventilation are impaired by such an overdose. Simply externally supporting the patient's respiratory efforts with bag-valve-mask ventilation for a short time is sufficient while assessing the patient. Delivery of oxygen alone is not sufficient as it does not address the equally important ventilatory impairment. Naloxone may be used, but at a very low initial dose, such as 0.04 mg, to avoid intubation while also not precipitating opioid withdrawal, as can occur with larger doses. Endotracheal intubation is often required in patients with acute respiratory distress syndrome or who do not respond to naloxone.

3. A young woman with a history of depression presents to the emergency department after being started on a second psychiatric medication. Her initial vital signs include a body temperature of 108° F. After external cooling, what specific pharmacotherapy can be provided?

A. Acetaminophen
B. Cyproheptadine
C. Dantrolene
D. Diazepam
E. Diphenhydramine

Answer: B Acetaminophen is ineffective for fever in which the body's set point is elevated. This patient likely has serotonin toxicity (or serotonin syndrome), which is due to excessive production of heat. Dantrolene is indicated for malignant hyperthermia and may have nonspecific effects in other hyperthermic syndromes. Diazepam is similarly nonspecific. Diphenhydramine is contraindicated because of its impairment of heat loss through sweating, given its antimuscarinic effects. Cyproheptadine is a serotonin receptor antagonist, and although it has not been adequately studied, in part given the rarity of serotonin toxicity, it is a specific therapy.

4. In a patient with an acute, intentional overdose with suicidal intent, which of the following laboratory tests is mandatory?

A. Acetaminophen level
B. β-Human chorionic gonadotropin level
C. Aspirin level
D. Carbon monoxide (carboxyhemoglobin) level
E. Urine drug screen

Answer: A A pregnancy test is indicated in women of childbearing potential but not in other patients. Blood aspirin levels and urine drug screens are often obtained, but they offer little information in acute management. Aspirin poisoning should be identified clinically. Urine drug-of-abuse screens may help in a psychiatric evaluation but are as likely to mislead the initial clinical evaluation as they are to offer any insight. Carbon monoxide is a distinctly uncommon form of suicide and should be assessed by the history or other findings. Acetaminophen overdose is extremely common because of the wide availability of this medication in various forms, and it is very difficult to detect clinically until hepatic failure develops. Given the availability of an effective antidote, early detection is critical and should be sought in all patients with suicidal intent.

5. Which of the following methods of gastrointestinal decontamination should be performed in the majority of patients with oral exposures to poisons?

A. Ipecac-induced emesis
B. No decontamination
C. Oral activated charcoal
D. Orogastric lavage
E. Whole bowel irrigation

Answer: B The benefit of gastrointestinal decontamination has been difficult to prove. For patients with high-risk exposures to significant quantities of toxin (such as calcium-channel blockers, monoamine oxidase inhibitors, colchicine), the benefits generally outweigh the risks. For the majority of exposures, it is perfectly acceptable to avoid oral decontamination, including administration of oral activated charcoal. Whole bowel irrigation is reserved for sustained-release medication (or drug mules), and orogastric lavage is generally performed only for the highest risk exposures in patients who present within several hours of ingestion. Ipecac is no longer used.

111

MEDICAL ASPECTS OF INJURIES AND BURNS

ROBERT L. SHERIDAN

Injured or burned patients are complex to manage not only because of the vast number of potential anatomic derangements but also because of the complex physiologic cascades triggered by injury. Although burns are associated with the most profound physiologic derangements, most medical issues are fairly similar across a wide range of injuries. As a result, needed interventions are often predictable regardless of the mechanism of injury.

EPIDEMIOLOGY

Trauma is an enormous public health issue. In the United States alone each year, about 2.5 million people are injured and 40,000 killed in automobile crashes, and about 78,000 are injured and 32,000 killed by gunshots. Burns and falls follow in frequency. Worldwide, injury by trauma and burns is the leading cause of death in children and young adults. In the middle-aged and elderly, injury follows only cancer and heart disease as a cause of death. Many long-term survivors of trauma and burns have high degrees of disability, which creates a particularly difficult problem given the young age of many victims.

Death after injury has a trimodal distribution.[1] At least 50% of fatalities occur within minutes of the injury as a result of massive hemorrhage or nonsurvivable brain injury. Because no medical interventions are possible in such cases, the importance of prevention is paramount. Approximately one third of deaths occur within a few hours after the injury and are usually caused by hemorrhage, anoxia, or progressive brain trauma. This interval provides an opportunity for emergent intervention. Later fatalities are usually the result of multisystem organ dysfunction or overwhelming infection in the days and weeks after the injury.

PATHOBIOLOGY

Early Local Response to Injury

The local response to blunt, penetrating, electrical, crush, thermal, blast, or other injury varies with the energy transferred. All are associated with some degree of secondary injury by progressive microvascular thrombosis, progressive edema, and secondary compromise of perfusion. The most common preventable mechanisms are related to progressive edema beneath an inelastic eschar or fascial compartments or direct injury to vascular inflow or outflow.

Early Systemic Response to Injury

The early systemic response to burns and local trauma is driven by fluid loss and release of vasoactive mediators from injured tissue. In more severe injuries, including surface burns greater than about 20% of the body surface, interstitial edema develops in unburned skin as well as in distant organs and soft tissues. These distant microvascular effects, which can compromise the function of organs that were not directly injured, explain the frequent occurrence of pulmonary and other organ dysfunctions in patients with large burns.

The capillary leak syndrome, in which the leak is typically proportional to the scope of the injury, is caused by the release of vasoactive substances from the injured and reperfused tissue. As a result, during the first 18 to 24 hours after a serious burn, both electrolytes and large colloid molecules freely diffuse into the interstitial space.

Late Local Response to Injury

At about 72 hours after the initial trauma, local wound issues are particularly important in patients who have extensive soft tissue damage, especially after burns or crush injury with extensive volumes of devitalized tissue. Even wounds that initially are generally clean can be rapidly colonized by endogenous bacteria. As these bacteria multiply in the avascular tissue during the succeeding days, proteases liquefy the eschar and necrotic tissue, which then separates and leaves a bed of granulation tissue. In healthy patients with smaller wounds and burns (<20% of the body surface), this septic process is usually tolerated. When injuries are larger, however, systemic infection results and explains the rare survival of patients who have burns in excess of 40% of body surface area or who have massive soft tissue injuries that were managed without early wound excision.

Late Hypermetabolic Systemic Response to Injury

Successfully resuscitated patients with large burns and, to a lesser extent, those with serious non-burn trauma demonstrate an initial decrease in cardiac output and metabolic rate. After successful resuscitation, a hypermetabolic response occurs, with a near doubling of cardiac output and resting energy expenditure during the next 24 to 72 hours. The magnitude of the response, which becomes greater with larger burns and more severe injuries, peaks at up to twice the normal metabolic rate in otherwise healthy patients with burn involving 60% or more of the body surface area. This hypermetabolic response is characterized by enhanced gluconeogenesis, insulin resistance, and increased protein catabolism. Although the causes of these physiologic changes are not well understood, they seem to involve the systemic release of bacterial products, the breakdown of gastrointestinal barrier dysfunction with translocation of bacteria and their byproducts into the circulation, and increases in the secretion of glucagon, cortisol, and catecholamines.

CLINICAL MANIFESTATIONS

All injuries involve the transmission of energy to viable tissue. Classic injury mechanisms include blunt, penetrating, electrical, thermal, blast, and crush (Table 111-1), but combined mechanisms are common. For example, patients crushed in building collapses frequently suffer a concomitant penetrating component, and patients who suffer high-voltage injury frequently fall from a height, such as a utility pole. In all mechanisms, edema or vascular injury can compromise perfusion and lead to secondary injury.

Blunt injury is commonly seen in motor vehicle accidents and falls. Injuries are often multisystem. Penetrating soft tissue injuries vary widely (Fig. 111-1), but the overarching concern is the possibility of occult injury to vascular, bone, or visceral structures. The degree of trauma is directly related to the energy of the injury. For example, high-velocity gunshots cause greater injury than lower velocity rounds because they create a local blast effect, called cavitation, along their trajectory.

Thermal injury (Fig. 111-2) by whatever mechanism (flame, scald, contact) is associated with a graded soft tissue injury described in degrees.[2] As thermal injuries increase beyond about 15% of the body surface, a clinically important systemic phenomenon occurs. A major manifestation of severe burns is a diffuse capillary leak that continues for 18 to 24 hours after injury and involves both burned and unburned tissue. This phenomenon can result in cardiovascular collapse and is the key physiologic derangement underlying the shock state accompanying burns.

The severity of electrical injuries (Fig. 111-3) varies with voltage, current flow, and contact quality. Low-voltage injuries are rarely associated with distant sequelae, whereas high-voltage injuries are commonly associated with compartment syndromes, cardiac complications, pigmenturia, and other trauma.

Crush injuries (Fig. 111-4) include direct soft tissue injuries as well as secondary ischemic damage due to a compartment syndrome or

TABLE 111-1 INJURY CLASSES

INJURY CLASS	CLINICAL IMPLICATIONS	COMMON ERRORS
Blunt	Graded soft tissue and bone injury, edema Occult vascular or visceral injury	Delayed consequences of edema Missed visceral injury and secondary consequences
Penetrating	Soft tissue and bone injuries vary with energy of injuring object (e.g., knife vs. fragment vs. bullet) Occult vascular or visceral injury	Missed vascular or visceral injury Secondary consequences of missed vascular injury
Thermal	Graded soft tissue injury (first- to fourth-degree burns) Capillary leak phenomenon	Inadequate appreciation of capillary leak phenomenon and consequent under-resuscitation Inaccurate wound evaluation and over-resuscitation Unanticipated local injury progression
Electrical	Range of soft tissue injury with increasing voltage and contact duration and quality	Occult cardiac and muscle injury Secondary compartment syndrome complications
Crush	Graded soft tissue, bone, and visceral injury with secondary consequences of edema and of reperfusion	Underappreciation of injury severity Missed muscle ischemia due to primary ischemia and to reperfusion edema
Blast	Graded injury severity (primary to quaternary injury patterns) Occult visceral injury	Underappreciation of secondary, tertiary, and quaternary components of injury Missed associated visceral injury
Fragmentation injury	Multiple unpredictable penetrations with associated visceral and vascular injury	Missed visceral and vascular injury Associated blast trauma

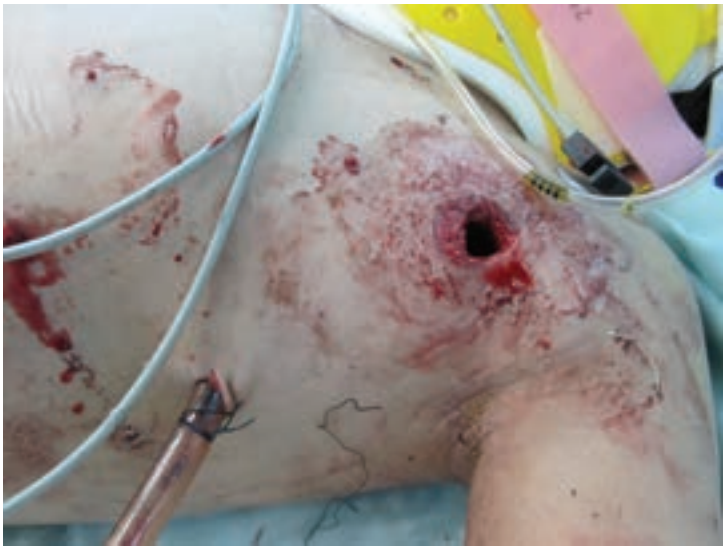

FIGURE 111-1. Penetrating injury is associated with often occult visceral and bone trauma.

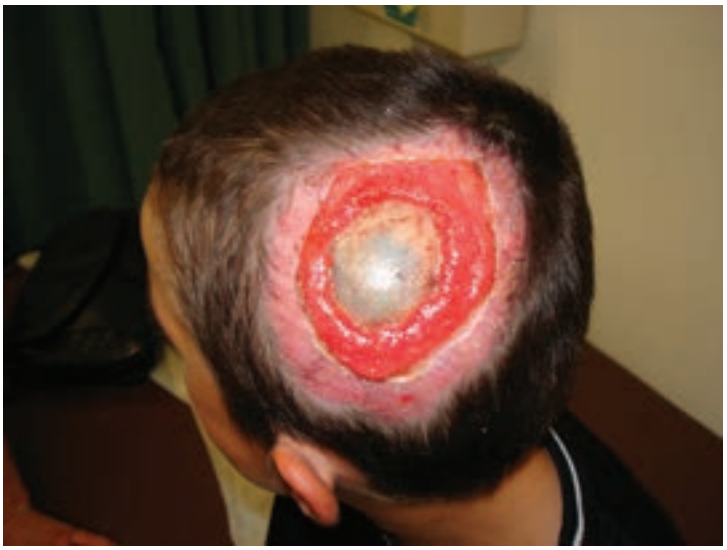

FIGURE 111-3. Electrical injury can be associated with a number of systemic sequelae, depending on current strength and pattern of flow.

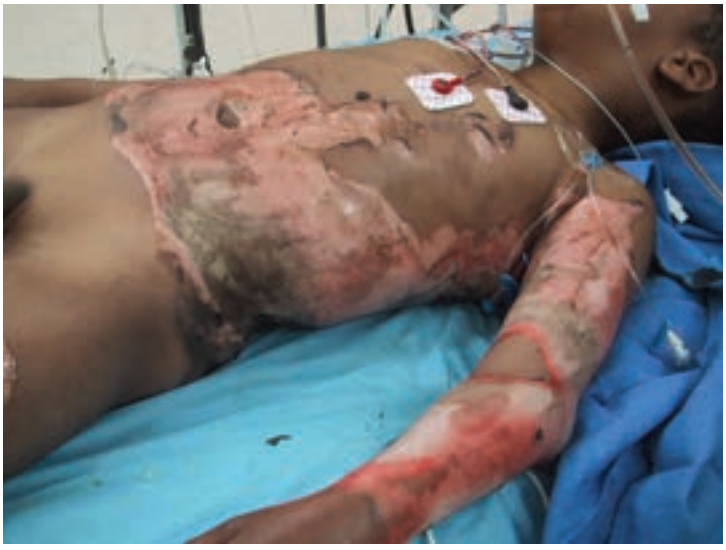

FIGURE 111-2. Thermal injury is of variable depth. Initial evaluation usually underestimates this depth.

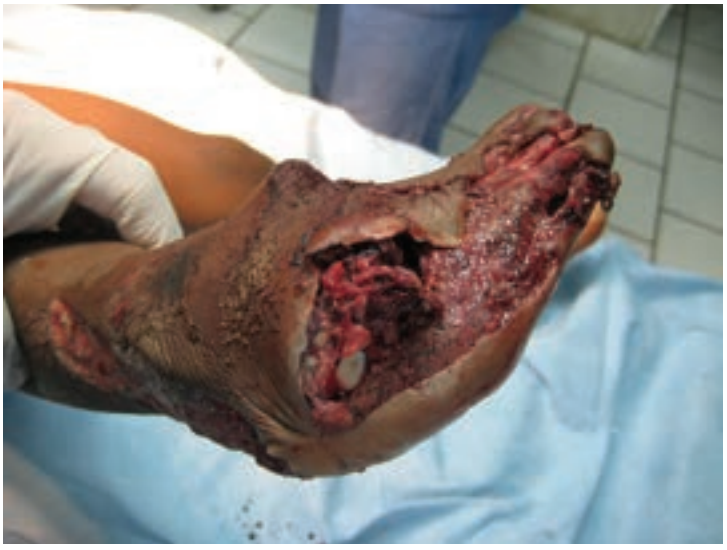

FIGURE 111-4. Crush injury is often associated with underappreciated deep muscle injury.

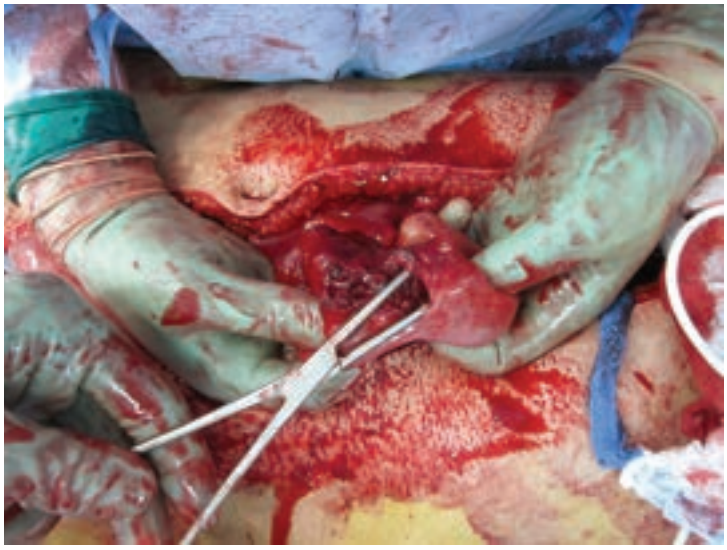

FIGURE 111-5. Blast injury is a complex of four injury subtypes. Missed injuries are common.

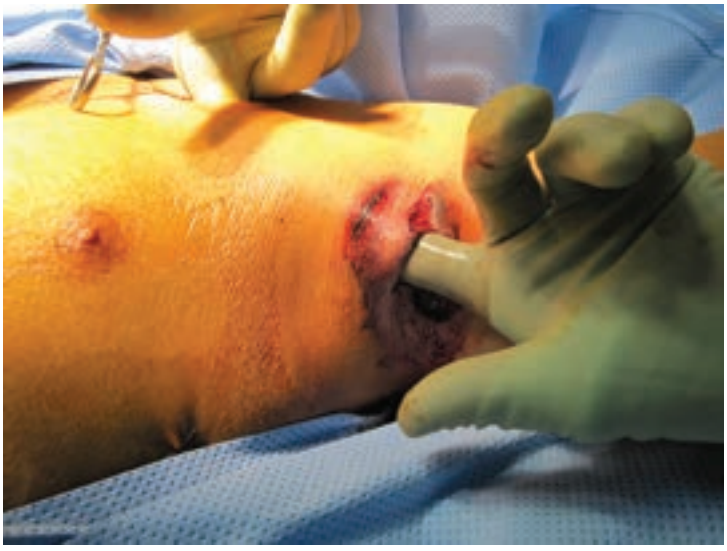

FIGURE 111-6. Fragmentation injury is a particular form of penetrating injury frequently associated with missed injuries.

ischemia-reperfusion. Associated bone, visceral, and vascular injuries are common. Major septic complications are common when necrotic soft tissues are left unexcised.

Blast injuries (Fig. 111-5) are graded complex injuries with four characteristics: primary injury to air-filled structures and the central nervous system, secondary injury from flying debris, tertiary injury from collisions with stationary objects, and quaternary injury from associated crush or other trauma. Blast injuries of all types can be subtle or have a delayed clinical presentation, and visceral injuries often are not diagnosed promptly.

Fragmentation injuries (Fig. 111-6) are penetrating injuries characterized by multiple foreign bodies of variable size and energy as well as unpredictable trajectories. As with blast injuries, the full extent of injury is often underestimated initially.

DIAGNOSIS AND TREATMENT

Rx

Complex burns and trauma are most successfully and cost-effectively managed in high-volume programs.[3] Trauma and burn center programs emphasize the comprehensive nature of injury care, including community involvement with injury prevention programs, thorough prehospital care, early resuscitation and surgery, and long-term rehabilitation and reconstruction. Most programs include imbedded specialty intensive care units and operating rooms. Multidisciplinary staff include physicians, physician assistants, nurse practitioners, nurses, physical and occupational therapists, respiratory therapists, psychiatrists, social workers, and nurse administrators. Coordination and scheduled communication are essential.

Care of individual patients with serious multisystem injury is complex and requires a longitudinal four-phase approach. Phase One, which describes the initial evaluation and resuscitation, generally is completed in the first 24 hours. Phase Two includes initial wound excision for burn patients and initial resuscitative surgery and fracture stabilization for non-burn trauma patients. This phase frequently overlaps with Phase One but is usually completed within 72 hours. Phase Three involves definitive wound closure for burns and the completion of surgery for trauma patients. The duration of this phase varies greatly with injury but is usually complete at the time of discharge. Phase Four describes the sometimes long process of reconstruction, rehabilitation, and reintegration. This phase of care commonly spans the later part of the acute hospitalization, inpatient rehabilitation, and variable amounts of time in the home-based setting.

Prehospital Care and Interhospital Transport

When prehospital and interhospital transfers are being arranged, key issues include control of the airway, secure venous access, placement of bladder and nasogastric catheters, maintenance of body temperature, fluid administration if transport time will be more than 1 hour, documentation of the events of the injury from personnel who will not be available to the receiving facility, efforts to notify family members, and clear documentation of all interventions. Hypothermia (Chapter 109) is a particular problem in burn patients because of their evaporative heat losses. Transporting vehicles and emergency department receiving areas should be heated before the patient's arrival. Initial burn dressings should be dry, clean sheets rather than wetted dressings. Cooling of a wound involving less than 15% of the body surface within minutes may help limit burn depth without causing systemic hypothermia, whereas cooling after a few minutes is generally unhelpful.

Phase One: Initial Evaluation and Resuscitation

Primary Survey

The organized approach to the initial evaluation of injured patients requires primary and subsequent patient surveys. The primary survey is an initial look for major aberrations in airway, circulatory, or neurologic status that justify emergent intervention. The airway is controlled with intubation if needed. Clinically obvious pneumothoraces (Chapter 99) should be decompressed, and hemothoraces must be diagnosed and drained. In suspicious cases, pericardial tamponade (Chapter 77) can usually be documented or excluded by handheld bedside ultrasonography. Vascular access is obtained and fluid resuscitation started. A very brief neurologic assessment (Chapter 396), including use of the Glasgow Coma Scale, is critical.

Secondary Survey

The secondary survey includes a much more detailed head-to-toe physical examination after the patient has a reliable airway and has been hemodynamically stabilized. The physical examination should assess for evidence of closed head injury, skull and facial fractures, and eye and ear injuries, with a low threshold for a head computed tomography (CT) scan. The neck must be stabilized and assessed for possible injury. Burns should be categorized as first-degree burns, involving only the epidermis; second-degree burns, involving variable amounts of dermis; third-degree burns, involving the entire dermis; and fourth-degree burns, involving fat, muscle, and bone.

A constant concern is the potentially missed injury. In the chaos of initial care, major issues can be missed if they do not have a dramatic initial presentation. Unfortunately, many injuries, such as epidural hematoma and small bowel perforations, are initially subtle and become catastrophic hours or days later. The best way to deal with this difficult reality is to have a highly organized approach to initial evaluation so that all potential injuries are considered and reasonably excluded.

Projectiles often follow an unpredictable course through tissue and bone, thereby increasing the likelihood of missed injury. Missed visceral injury is also common with fragmentation injuries, especially because surgical exploration of all potential sites of injury is often impossible. Selected exploration is guided by initial and serial examination and by imaging when it is available.

Appropriate laboratory and imaging studies should be obtained. The fundamental goal is to consider carefully the injury mechanism and to exclude all potential occult injuries to a reasonable level of confidence. High-energy injuries, such as electrical burns, motor vehicle crashes, and blasts, are notorious for generating significant injuries that are missed during the initial evaluation. Bedside ultrasonography has emerged as a routine, rapid, and repeatable method to assess for abdominal fluid, usually blood. CT scanning of the head, chest, abdomen, and pelvis is justified if the mechanism of injury is consistent with head or abdominal trauma. Rapid imaging has largely supplanted operative exploration for diagnostic purposes, except in unstable patients, in whom immediate operative exploration is performed for ongoing hemorrhage.

Chest Injuries

Life-threatening tension pneumothorax (Chapter 99), massive hemothorax, cardiac tamponade (Chapter 77), flail chest, open pneumothorax, and disruption of the thoracic aorta (Chapter 78) may be missed in the primary survey

but must be diagnosed in the secondary survey. All trauma patients should have a supine chest radiograph to examine the lung fields, the mediastinal contour, and the chest wall. Thoracic aortic injury is usually an immediately fatal complication of severe acceleration-deceleration injury, but some patients may have a contained mediastinal hematoma that requires urgent evaluation and repair.

Myocardial contusion should be suspected in patients with blunt trauma, especially if they have a sternal fracture or anterior rib fractures. Life-threatening myocardial contusion can be manifested with electrocardiographic abnormalities, ventricular arrhythmias, and even cardiogenic shock. Coronary artery injuries are uncommon, although lacerations and dissections can occur and require emergent percutaneous coronary intervention similar to a typical ST elevation myocardial infarction (Chapter 73).

Commotio cordis is sudden cardiac arrest (Chapter 63) after acute blunt chest trauma from softballs, baseballs, hockey pucks, or collisions. The trauma presumably occurs during an electrically vulnerable period between 30 and 15 msec before the T wave peak and produces ventricular fibrillation. Death is essentially universal unless the victim receives immediate resuscitation.

Abdominal Injuries

The spleen is the most frequently injured intra-abdominal organ, and splenic injury is suggested by left-sided rib fractures, left upper quadrant pain or tenderness, and pain referred to the left shoulder secondary to diaphragmatic irritation. CT scanning is the test of choice. High-grade injuries require splenectomy or, infrequently, splenorrhaphy for salvage of the damaged spleen with preservation of splenic function. Right-sided rib fractures and right upper quadrant tenderness suggest liver injury, which can also be assessed by CT. Most isolated blunt liver injuries can be managed nonoperatively. Focused abdominal sonography has a sensitivity of 95% for detection of free blood in the abdomen (Fig. 111-7) but is not usually able to identify the source. An alternative is diagnostic peritoneal aspiration or lavage. If blood is found by either approach, laparotomy is necessary to identify and to correct the source. Abdominal CT is a reliable determinant of both intraperitoneal and retroperitoneal injury in a stable patient (Fig. 111-8). CT evaluation also can identify sources of blood loss and help decide whether the problem requires operative exploration.

Head and Spine Injuries

Neurologic issues often dominate the quality of long-term outcome. The initial evaluation of the neurotrauma patient focuses on detection of reversible causes of elevated intracranial pressure or reduced perfusion (Chapter 399),[4] but routine measurement of intracranial pressures does not improve outcome. Although many outcomes are fixed from the time of injury, opportunities for early intervention should not be missed. It is particularly important to avoid secondary neurologic injury, most commonly due to reduced perfusion from hypotension and cerebral edema.

Prevention of spinal cord injury (Chapter 399) and evaluation of the spine for bone and ligamentous disruption are also important components of the early evaluation. Multidetector CT images can be obtained quickly to detect occult spinal injuries.

A multidisciplinary approach to the evaluation and management of pain and anxiety in injured patients has important short- and long-term benefits.[5] An organized approach facilitates dealing with the inevitable pain and anxiety in an organized and consistent manner and allows the unit to determine the effectiveness of new interventions.

Tertiary Survey

The tertiary survey is a planned and careful repeated physical examination, usually 1 or 2 days after admission, often accompanied by focused imaging tests. The goal again is to exclude subtle injuries that might otherwise cause significant long-term morbidity. Examples include minor fractures of the wrist or foot, small deep lacerations of the scalp, subtle eye injuries, and some abdominal injuries, particularly retroperitoneal duodenal or colonic perforations. Such injuries can become the patient's dominant source of long-term morbidity if they are missed during initial evaluation and treatment.

Key Management Issues

Fluid Resuscitation and Transfusion

The clinical goal is to administer adequate but not excessive fluid while ensuring that soft tissues are not rendered ischemic by increasing pressure beneath inelastic eschar or tense muscle compartments. The severity of hemorrhagic shock should be assessed and treated appropriately (Table 111-2).

For burn patients, several formulas based on weight, surface area, and burn size have been developed over the years. These are general guides, and physiologic monitoring is needed with individual titration of infusions to meet set resuscitation end points, such as urine output, base deficit, and vital signs (E-Table 111-1).

In resuscitation of burn patients, capillary integrity generally returns if resuscitation has been successful at 18 to 24 hours, and fluid needs markedly decrease toward approximately 1.5-fold maintenance in most patients. Overadministration of fluid should be avoided at this phase of resuscitation. With the reduction in the infusion rate of isotonic crystalloid, topical care of large wounds will have a major impact on serum electrolytes. Wounds treated with nonaqueous topical antimicrobials, such as silver sulfadiazine cream or mafenide acetate cream, promote transeschar water loss and generate a free water requirement, which is generally provided as 5% dextrose and water or free water added to enteral feedings, to avoid hypernatremia. By comparison, wounds treated with aqueous topical agents are associated with electrolyte leeching and secondary hyponatremia (Chapter 116). Serum levels of potassium, calcium, and magnesium (Chapters 117-119 and 245) should be monitored frequently and replaced as needed. In most patients, enteral feedings (Chapter 216) can be started.

Modern fluid resuscitation in trauma patients emphasizes initial "permissive hypotension" to reduce early bleeding (Chapters 104 and 106). Patients with systemic arterial pressures in the range of 60 to 80 mm Hg are resuscitated initially to modest hypotension (i.e., the range of 80 to 90 mm Hg) if they are otherwise alert and well perfused. They are then taken expeditiously to the operating room for surgical control before targeting a normotensive state. Early use of blood products such as fresh-frozen plasma and red cells rather than crystalloid can minimize the coagulopathy and improve outcomes in rapidly hemorrhaging patients.[6] Liberal use of tourniquets and compressive dressings before surgery further improves outcome by reducing blood loss. Among the various crystalloids, none seems to have any obvious advantages or disadvantages for treatment of hemorrhage. Tranexamic acid (e.g., loading dose of 1 g during 10 minutes, then infusion of 1 g during 8 hours), which inhibits thrombolysis, beginning within 8 hours of injury may reduce bleeding in selected patients who require massive transfusion.

Pigmented urine is commonly seen in the setting of high-voltage, crush, blast, or very deep thermal injury. Myoglobin and hemoglobin that are liberated from lysed muscle (Chapter 113) and red cells cause the pigmentation.

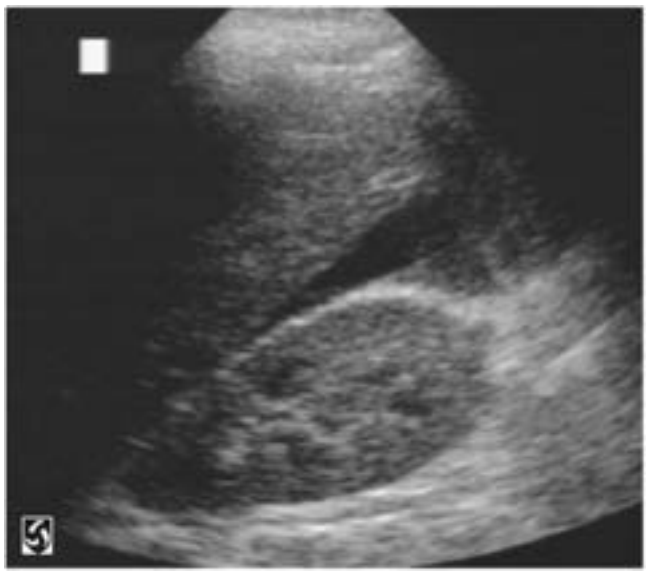

FIGURE 111-7. Positive focal abdominal sonography in trauma (FAST) study. There is a collection of fluid that shows up as a black strip between the liver and kidney. The white appearance is fat around the kidney. These findings are consistent with a fluid collection in Morison pouch. (Courtesy Robert H. Demling, MD.)

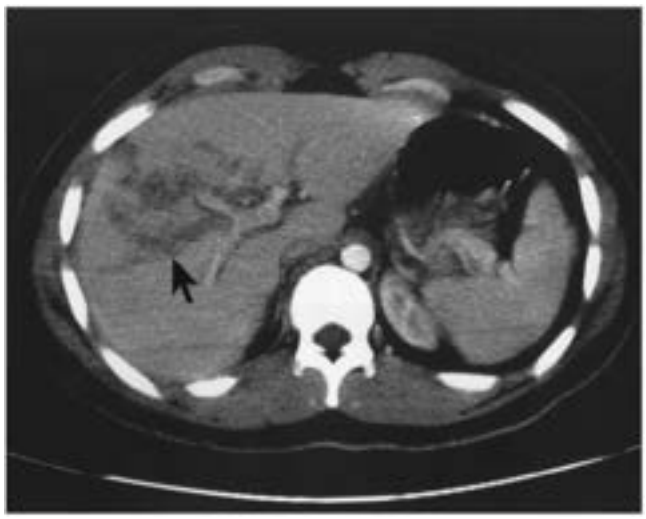

FIGURE 111-8. Abdominal injury. An abdominal computed tomography scan shows an injury to the right lobe of the liver (*arrow*). (Courtesy Robert H. Demling, MD.)

E-TABLE 111-1 BURN RESUSCITATION FORMULA

FIRST 24 HOURS

Adults and Children >20 kg

- Lactated Ringer solution: 2-4 mL/kg per percentage of total body surface area (TBSA) burned per 24 hours (first half in first 8 hours)
- Colloid: in many children, particularly those with small injuries, no colloid is advised in the first 24 hours. However, colloid, generally as 5% albumin solution, is increasingly used early in resuscitation of patients with large burn injuries. This is program specific and should ideally be discussed with the unit to which the child with a large injury will be referred.

SECOND 24 HOURS—ALL PATIENTS

- Crystalloid: to maintain urine output. If silver nitrate is used, sodium leeching will mandate continued isotonic crystalloid. If other topical is used, then free water requirement is significant. Serum sodium should be monitored closely.
- Nutritional support should begin, ideally by the enteral route.
- Colloid (5% albumin):
  - 0-30% burn: none
  - 30-50% burn: 0.3 mL/kg per percentage of TBSA burned per 24 hours
  - 50-70% burn: 0.4 mL/kg per percentage of TBSA burned per 24 hours
  - 70-100% burn: 0.5 mL/kg per percentage of TBSA burned per 24 hours

TABLE 111-2 CATEGORIZATION AND INITIAL TREATMENT OF HEMORRHAGIC SHOCK*

	CLASS I	CLASS II	CLASS III	CLASS IV
Blood loss (mL)	≤750	750-1500	1500-2000	≥2000
Blood loss (% of blood volume)	≤15	15-30	30-40	≥40
Pulse rate	<100	>100	>120	≥140
Blood pressure	Normal	Normal	Decreased	Decreased
Capillary refill test	Normal	Positive	Positive	Positive
Respiratory rate	14-20	20-30	30-40	>35
Urine output (mL/hr)	≥30	20-30	5-15	Negligible
Mental status	Slightly anxious	Mildly anxious	Anxious and confused	Confused and lethargic
Fluid replacement (3 : 1 rule)	Crystalloid	Crystalloid	Crystalloid + blood	Crystalloid + blood

*Based on a 70-kg adult. From Demling RH, Gates JD. Medical aspects of trauma and burn care. In: Goldman L, Schafer AI, eds. *Goldman's Cecil Medicine*. 24th ed. Philadelphia: Saunders-Elsevier; 2012.

To avoid renal tubular injury (Chapter 120), crystalloids should be administered to achieve a urine output of 2 mL/kg/hour (Chapter 113).

Decompression Procedures

An important component of the initial evaluation of the trauma and burn patient is to find and to rectify compartment syndromes before irreversible tissue ischemia occurs. Perfusion of the extremities can be compromised by inelastic near-circumferential burns, and the normal fascial envelopes of major muscle groups can cause high soft tissue pressure if muscles are injured by crush, thermal, or electrical injury or in association with major fractures. Even if pressures are normal in large vessels, edematous soft tissues can cause ischemia and necrosis because capillary perfusion is compromised. Extremities at risk should be dressed simply to facilitate frequent assessment for temperature, pliability, voluntary motion, pain with passive motion, detectable pulsations, and low-pressure flow by capillary refill and Doppler signals in the digital vessels and digital pulp. In assessing capillary refill, it is important to elevate the extremity, as even a mottled nonperfused extremity will demonstrate venous refill when it is dependent. Serial measurements of compartment pressures may be valuable in selected patients, with decompression recommended when measured pressures are above 30 cm H_2O. In most situations, serial clinical examination is sufficient to determine the need for escharotomy or fasciotomy, thereby avoiding the risk of bacterial seeding posed by passing pressure-monitoring catheters through contaminated wounds. Compromised extremities should be promptly decompressed by escharotomy or fasciotomy before the development of irreversible tissue necrosis (Fig. 111-9).

Abdominal Compartment Syndrome

If abdominal viscera become extremely edematous, abdominal compartment syndrome may result.[7] This syndrome, which is usually caused by edema of the bowel wall, occurs after abdominal trauma or after the gut is reperfused by resuscitative therapies. When intra-abdominal pressures exceed 25 mm Hg (34 cm H_2O), renal blood flow, inferior vena cava blood return, and diaphragmatic excursion are impaired. This syndrome typically is manifested with oliguria, hypotension, and difficult ventilation. Diagnosis is by physical examination and measurement of bladder pressure. Treatment is by laparotomy with temporary abdominal closure.

Phase Two: Initial Surgical Care

Phase Two includes initial wound excision for burn patients and initial resuscitative surgery and fracture stabilization for non-burn trauma patients. This phase frequently overlaps with Phase One but is usually completed within 72 hours. Multiple subspecialty surgical teams may need to be coordinated by a trauma surgeon who directs overall care and reconciles conflicting priorities.

Early Wound Excision

Early wound excision within the first 72 hours of injury, before heavy microbial colonization occurs, is important for patients with significant burns, crush, or other soft tissue trauma. Early identification, excision, and closure of full-thickness wounds can avoid otherwise inevitable sepsis, systemic infection, and systemic inflammatory response. Excisional débridement of nonviable soft tissues or deep burns is critically important.

Damage Control Surgery

Damage control surgery identifies the most important surgical tasks that must be done to save the patient's life and only these are performed initially, thereby allowing a more stable patient to return to the operating room later for a subsequent definitive and often time-consuming operation.[8] The prototypical example is for abdominal trauma, when bleeding and gastrointestinal contamination are addressed initially but the abdomen is left open so a warmed and more stable patient can return to the operating room in 12 to 36 hours for a definitive bowel anastomosis and abdominal closure. This concept can also be applied to the trauma care system. If multiple patients need to share limited operating room resources, truncating individual operations allows more patients to be treated urgently.

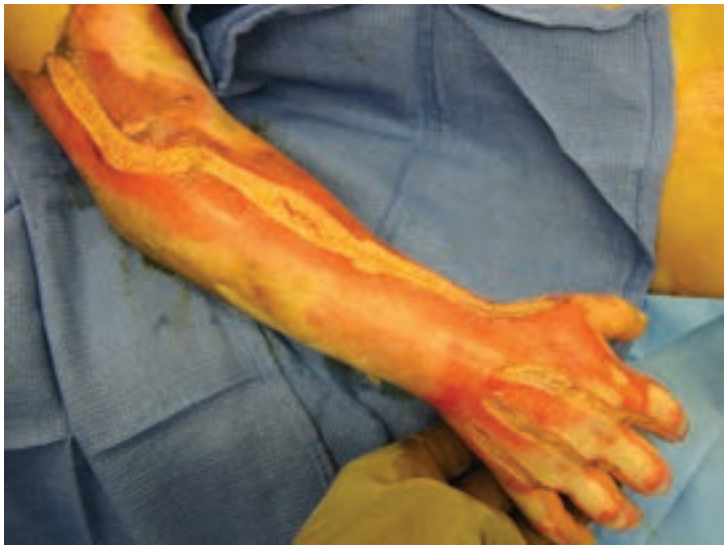

FIGURE 111-9. **Tight extremities should be promptly decompressed by escharotomy or fasciotomy before the development of irreversible tissue necrosis.**

Early Fracture Fixation

Early surgical fixation of long bone fractures benefits injured patients by reducing the complications of immobilization, enhancing rehabilitation, and likely reducing the systemic inflammatory state and proclivity to thromboembolic complications.[9]

Prophylaxis for Thromboembolic Complications and Gastrointestinal Hemorrhage

Trauma incites a hypercoagulable state, and seriously injured patients are prone to deep venous thrombosis and pulmonary embolism.[10] Pharmacologic prophylaxis, such as low-molecular-weight heparin, is recommended. Mechanical prophylaxis with automatic compression stockings is also routine in most trauma programs.

During the early hypodynamic phase with reduced splanchnic blood flow, gastrointestinal hemorrhage may occur in the seriously injured and burned patient. This complication can be sharply reduced with routine pharmacologic prophylaxis, including proton pump inhibitors or histamine-2 receptor blockers, such as with immediate-release omeprazole oral suspension (40 mg twice daily on the first day and 20 mg daily thereafter).[11]

Critical Care of the Injured or Burned

Many injured patients require transient intubation and mechanical ventilation (Chapter 105) to facilitate resuscitation, evaluation, and initial care. Some will go on to develop respiratory failure (Chapter 104) requiring protracted ventilator support.

Multiple mechanical and injury factors contribute to respiratory insufficiency in trauma patients, including chest wall trauma and pulmonary contusion. In patients with more severe injuries including flail chest, rib fixation is a useful intervention.[12]

In burn patients, inhalation injury (Chapter 94) is an important cause of respiratory failure. The diagnosis of inhalation injury is best made by history and clinical examination revealing singed nasal vibrissae and carbonaceous debris in the mouth and pharynx (E-Fig. 111-1). Chest radiographs are usually normal initially. Carbon monoxide poisoning (Chapter 94) can be seen in conjunction with inhalation injury; the standard of care is 100% oxygen at the prevailing atmospheric pressure for 6 hours, with hyperbaric oxygen reserved for patients with a carboxyhemoglobin level above 30% or neurologic changes.

Pneumonia (Chapter 97) or tracheobronchitis (Chapter 96) occurs in about 30% of burn patients with inhalation injury. The acute respiratory distress syndrome (Chapter 104) also is common in trauma patients, particularly those who go on to develop sepsis, pneumonia, and multiple organ failure. Antibiotic therapy is directed by sputum Gram stain and cultures and should not be prolonged beyond a 7- to 10-day therapeutic course. Vigorous pulmonary toilet, with directed bronchoscopy to clear secretions in selected patients, is an important component of therapy.

Post-Resuscitation Metabolic Issues

In the days that follow successful resuscitation, a hyperdynamic state predictably occurs, characterized by a high cardiac output and low peripheral resistance. Because wound contamination with bacteria and tissue hypoxia drive much of the hypermetabolic response, early excision of burned tissue and closure of burn wounds with autografting shorten hospital stays and enhance functional outcomes in patients with deep dermal and full-thickness burns. Prevention of tissue ischemia by repair of vascular injuries and decompression of edematous limbs by escharotomy or fasciotomy minimize the burden of necrotic tissue (E-Fig. 111-2). Early excision of necrotic or ischemic tissue will minimize the incidence of wound contamination and systemic sepsis and inflammation.

Several formulas have been devised to predict non-protein calorie needs (E-Table 111-2). A very rough estimate of calorie needs is a range of 25 to 35 kcal/kg/day, with the lower number applying to more stable and older patients and the upper number applying to more seriously injured and younger patients. Protein administration of 2 to 3 g per kilogram per day will adequately support the needs of most injured patients.

The route of nutritional support is ideally enteral (Chapter 216), with tube feeding beginning during resuscitation. However, some patients, particularly those with very severe and abdominal injuries or those with intervening sepsis, will not tolerate enteral feedings at goal rates and will require supplemental parenteral nutrition to ensure delivery of all needed nutrients (Chapter 217).

Safe and reliable modification of adverse components of the hypermetabolic response, particularly protein catabolism, has proved an elusive goal. Anabolic agents such as recombinant human growth hormone and anabolic steroids may help restore positive nitrogen balance, but they are not widely used because of their complications, expense, and conflicting data regarding the efficacy in most injured patients. Other attempted interventions, such as antipyretics, β-adrenergic blockade, β-adrenergic supplementation, nonsteroidal anti-inflammatory agents, recombinant growth hormone, insulin-like growth factor-I, and anabolic steroids, have been tried, but available data are inadequate to support any of these therapies as standard care.

Septic Complications and Multiple Organ Failure

Injured patients are at elevated risk for all types of infectious complications (Chapter 282). Multiple organ dysfunction and septic shock (Chapter 108) are manifestations of uncontrolled systemic inflammation for severe infection.

Phase Three: Definitive Surgical Care

Phase Three involves definitive wound closure for burns and completion surgery for trauma patients. In non-burn trauma patients, this phase may involve removal of vacuum wound dressings and definitive closure, closure of fasciotomy sites, and fixation of facial fractures. The duration of this phase varies greatly with injury, but it is usually complete at the time of acute hospital discharge. For example, in burn patients, this phase of care is defined by replacement of temporary membranes with permanent grafts and the important but time-consuming grafting of hands and face.

Phase Four: Rehabilitation and Reintegration

Phase Four describes the sometimes long process of reconstruction, rehabilitation, and reintegration.[13] This phase of care may begin in an inpatient rehabilitation setting and progress into the home environment. Depending on the specifics of the injury, the process may be a few weeks to months or years. Daily passive range of motion movements, splinting, anti-deformity positioning, and strengthening can reduce the frequency of these complications.

The environment of recovery is an important therapeutic consideration. Data suggest that specific qualities of the family have a major impact on multiple aspects of recovery, some that can potentially be modified. Community resources can be used to enhance recovery and are an important part of discharge planning.

REHABILITATION OF THE INJURED PATIENT

Rehabilitation in the Critically Ill Patient

Passive range of motion movement, splinting, and anti-deformity positioning will minimize the capsular contraction and shortening of tendon and muscle groups that otherwise occur with protracted immobilization.

Emotional recovery after severe trauma can be limited by neurologic injury and post-traumatic stress. These issues can be anticipated and their adverse effects blunted by early and continuous involvement of emotional supports for the family and patient. Ideally, psychiatric, social service, and psychological resources are available.

SPECIAL CONSIDERATIONS

Electrical Injury

Cardiopulmonary arrest can be caused by low-voltage electrical injury[14] but is more common with high-voltage electrical injury. Extensive tissue necrosis may also liberate enough potassium to cause cardiac dysfunction. Because cardiac arrhythmias may recur after resuscitation or develop 24 to 48 hours after injury, all patients who have sustained high-voltage electrical injury should undergo continuous electrocardiographic monitoring for at least 48 hours after the last documented arrhythmia.

A detailed neurologic examination must be performed on all patients with high-voltage electrical injury because dysfunction may be apparent immediately or may appear later. Recovery of function after direct electrical nerve damage is rare. Conversely, the nerves that are not injured directly commonly recover. A polyneuritic syndrome of relatively late onset can cause deficits in the function of peripheral nerves far removed from the points of contact. Delayed-onset spinal cord deficits can be manifested as quadriplegia, hemiplegia, localized nerve deficits with signs of ascending paralysis, transverse myelitis, and even an amyotrophic lateral sclerosis–like syndrome.

Among patients hospitalized with serious high-voltage electrical injuries, about 8% die and another 22% have permanent neurologic deficits despite optimal care. Most patients with low-voltage injuries, such as can be caused by electrical flash burns, also develop long-term sequelae, including neurologic (memory loss, numbness, headache, chronic pain, weakness) and musculoskeletal (pain, reduced range of motion, contracture) symptoms.

Lightning Injury

Cardiopulmonary arrest is common in patients struck by lightning.[15] Coma is also common acutely but typically resolves in a few hours. Keraunoparalysis, which is lightning-induced paralysis, is characterized by usually transient paresthesias and paralysis that develop during several days and typically involve the lower limbs. Ruptured tympanic membranes and hearing loss may also occur. With immediate cardiopulmonary resuscitation, about two thirds of lightning victims survive, and persistent neurologic deficits are relatively uncommon.

Considerations at the Extremes of Age

Injured patients at the extremes of age bring with them a number of important physiologic and psychosocial issues that directly affect their care (E-Table 111-3). Older adults do not have the depth of physiologic reserve of the young (Chapter 25). Cardiopulmonary function may be compromised, and peripheral vascular disease is common. Muscle mass and respiratory muscle strength are often reduced. Reduced renal reserve increases sensitivity to nephrotoxic drugs. The skin is relatively atrophic and consequently tolerates burning and donor harvest poorly. Nutritional needs are not well predicted by standard equations.

As a result, a major injury is often the event that changes the subsequent living condition of an elderly person.[16] Injuries are often the result of cognitive or functional changes (Chapter 27) and may be associated with syncopal episodes (Chapter 62) that need to be concurrently evaluated. Resuscitation should be carefully considered if burns are very large or trauma is severe. Data suggest that mortality is nearly 90% in patients who are older than 60 years, have burns of more than 40% of their body surface, and have concomitant inhalation injury. Patients may have advanced directives, interested families, or health care proxies who should be consulted as early as possible in the care of such patients (Chapter 3).

Older adults bring unique psychosocial issues to the trauma and burn center. They may live alone or have a spouse who cannot meet their discharge care needs for wound care, transportation, or general support. Children may live far away and be unable to support them. Discharge planning can be very involved, requiring orchestration of many community resources, and must be started early.

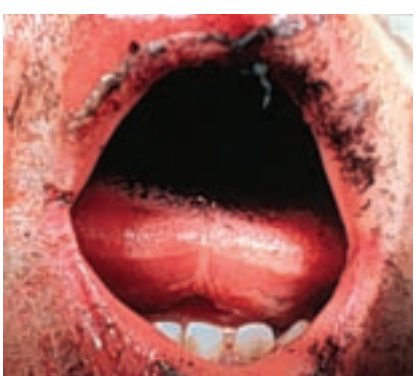

E-FIGURE 111-1. **Inhalation injury is predicted by history and physical findings, including soot in the mouth and teeth as shown here.**

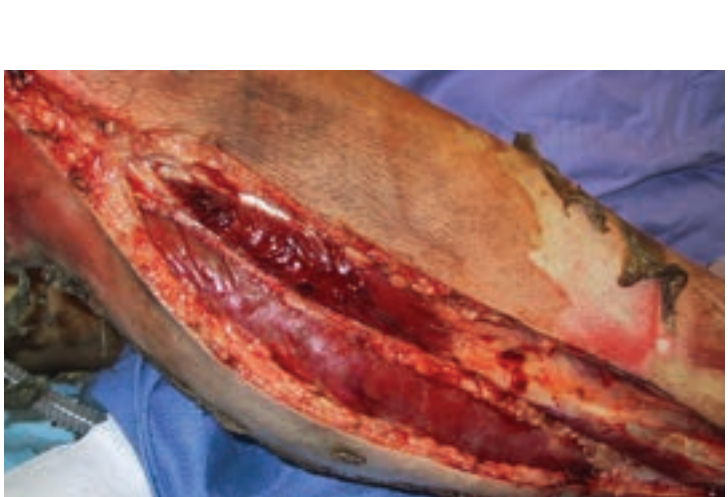

E-FIGURE 111-2. **Prevention of tissue ischemia of edematous limbs by escharotomy or fasciotomy will minimize the burden of necrotic tissue.**

E-TABLE 111-2 MODIFIED HARRIS-BENEDICT EQUATION

MALE PATIENTS
BMR = 88.362 + (13.397 × weight in kg) + (4.799 × height in cm) − (5.677 × age in years)
FEMALE PATIENTS
BMR = 447.593 + (9.247 × weight in kg) + (3.098 × height in cm) − (4.330 × age in years)

BMR = basal metabolic rate.
From Roza AM, Shizgal HM. The Harris Benedict equation reevaluated: resting energy requirements and the body cell mass. *Am J Clin Nutr.* 1984;40:168-182.

E-TABLE 111-3 PSYCHOSOCIAL AND PHYSIOLOGIC CONSIDERATIONS IN GERIATRICS

- Injury mechanisms more often involve compromised mobility or dexterity.
- Injuries may reflect an inability to safely live alone.
- Injuries occur during syncopal episodes.
- Resuscitation should be carefully considered if burns are very large, particularly in the presence of inhalation injury.
- Patients may have advanced directives, interested families, or health care proxies who should be consulted as early as possible in their care.
- Older adults often do not have the physiologic reserve of the young.
- Pulmonary function may be compromised by years of smoking.
- Occult or overt coronary artery or peripheral vascular disease may exist.
- Muscle strength, including respiratory muscles, may be reduced.
- Renal function may be reduced with resulting greater sensitivity to nephrotoxic drugs or hypotensive insults.
- Nutritional needs of the elderly are poorly predicted by standard predictive equations.
- The skin of the elderly person is thin and therefore sustains full-thickness injury more readily and tolerates repeated donor harvest less well.
- Older adults may live alone or have a spouse who cannot reasonably meet postdischarge needs.
- Discharge planning may be very involved and must be started early.

Grade A References

A1. Chesnut RM, Temkin N, Carney N, et al. A trial of intracranial-pressure monitoring in traumatic brain injury. *N Engl J Med*. 2012;367:2471-2481.

A2. Gruen RL, Brohi K, Schreiber M, et al. Haemorrhage control in severely injured patients. *Lancet*. 2012;380:1099-1108.

A3. Shakur H, Roberts I, Bautista R, et al. Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): a randomised, placebo-controlled trial. *Lancet*. 2010;376:23-32.

A4. Conrad SA, Gabrielli A, Margolis B, et al. Randomized, double-blind comparison of immediate-release omeprazole oral suspension versus intravenous cimetidine for the prevention of upper gastrointestinal bleeding in critically ill patients. *Crit Care Med*. 2005;33:760-765.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Clark DE, Qian J, Sihler KC, et al. The distribution of survival times after injury. *World J Surg*. 2012;36:1562-1570.
2. Cancio LC, Lundy JB, Sheridan RL. Evolving changes in the management of burns and environmental injuries. *Surg Clin North Am*. 2012;92:959-986.
3. Klein MB, Goverman J, Hayden DL, et al. Benchmarking outcomes in the critically injured burn patient. *Ann Surg*. 2014;259:833-841.
4. Rosenfeld JV, Maas AI, Bragge P, et al. Early management of severe traumatic brain injury. *Lancet*. 2012;380:1088-1098.
5. Barr J, Fraser GL, Puntillo K, et al. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. *Crit Care Med*. 2013;41:263-306.
6. Campion EM, Pritts TA, Dorlac WC, et al. Implementation of a military-derived damage-control resuscitation strategy in a civilian trauma center decreases acute hypoxia in massively transfused patients. *J Trauma Acute Care Surg*. 2013;75:S221-S227.
7. Carr JA. Abdominal compartment syndrome: a decade of progress. *J Am Coll Surg*. 2013;216:135-146.
8. Sharrock AE, Midwinter M. Damage control—trauma care in the first hour and beyond: a clinical review of relevant developments in the field of trauma care. *Ann R Coll Surg Engl*. 2013;95:177-183.
9. Nahm NJ, Vallier HA. Timing of definitive treatment of femoral shaft fractures in patients with multiple injuries: a systematic review of randomized and nonrandomized trials. *J Trauma Acute Care Surg*. 2012;73:1046-1063.
10. Holley AB, Petteys S, Mitchell JD, et al. Thromboprophylaxis and VTE rates in soldiers wounded in Operation Enduring Freedom and Operation Iraqi Freedom. *Chest*. 2013;144:966-973.
11. Hurt RT, Frazier TH, McClave SA, et al. Stress prophylaxis in intensive care unit patients and the role of enteral nutrition. *JPEN J Parenter Enteral Nutr*. 2012;36:721-731.
12. Slobogean GP, MacPherson CA, Sun T, et al. Surgical fixation vs nonoperative management of flail chest: a meta-analysis. *J Am Coll Surg*. 2013;216:302-311.
13. Engels PT, Beckett AN, Rubenfeld GD, et al. Physical rehabilitation of the critically ill trauma patient in the ICU. *Crit Care Med*. 2013;41:1790-1801.
14. Alemayehu H, Tarkowski A, Dehmer JJ, et al. Management of electrical and chemical burns in children. *J Surg Res*. 2014;190:210-213.
15. Sanford A, Gamelli RL. Lightning and thermal injuries. *Handb Clin Neurol*. 2014;120:981-986.
16. Staudenmayer KL, Hsia RY, Mann NC, et al. Triage of elderly trauma patients: a population-based perspective. *J Am Coll Surg*. 2013;217:569-576.

REVIEW QUESTIONS

1. A 75-year-old woman was struck by a car going 20 miles per hour while she was crossing the street. The major impact was at the level of the knees, but she was thrown up onto the hood of the car before sliding onto the street. On arrival in the emergency department, she is alert with stable vital signs and a comfortable breathing pattern. She has obvious closed bilateral lower extremity fractures. Besides progressive edema in the fractured legs leading to compartment syndrome, what is the most likely trap to be anticipated when managing this patient?

A. Delirium induced by pain medications
B. Immediate thromboembolic complications
C. Delayed presentation of visceral injury
D. Anoxia and seizures
E. Occult pneumothorax

Answer: C The primary survey will reveal the injuries to the lower extremities, but the secondary survey should include attention to possible abdominal injury. Failure to note a lacerated kidney could have disastrous consequences.

2. Three policemen in their early 30s are rushed to the emergency department after being injured by a vehicle-borne improvised explosive device. These victims are likely to suffer from all *except* which of the following?

A. Overpressure injury to air-filled structures such as lung and bowel
B. Penetrating injury from flying debris
C. Blunt injury from impact with stationary objects
D. Crush injury from falling debris
E. Globe rupture from overpressure wave

Answer: E Overpressure waves affect air-filled structures such as the lung, bowel, and auditory system. Overpressure could affect hearing, as is seen in lightning injuries.

3. In the situation described in question 2, there are only two available operating rooms. The least injured of the three patients will have to wait 2 to 4 hours for surgery. During this time, an optimal resuscitation target is

A. Systolic blood pressure of 140-160, intact sensorium, strong peripheral pulses, urine output of 4 mL/kg/hr, base deficit 0
B. Systolic blood pressure of 120-140, intact sensorium, strong peripheral pulses, urine output of 2 mL/kg/hr, base deficit 0-1
C. Systolic blood pressure of 100-120, intact sensorium, strong peripheral pulses, urine output of >1 mL/kg/hr, base deficit 0-1
D. Systolic blood pressure of 80-90, intact sensorium, palpable peripheral pulses, urine output of 0.5 mL/kg/hr, normalizing base deficit
E. Systolic blood pressure of 60-70, depressed sensorium, thready peripheral pulses, urine output of <0.5 mL/kg/hr, base deficit >5

Answer: D Recent research has shown that targeting lower systemic arterial pressure before exploratory surgery can lead to improved outcomes.

4. Typical of the hypermetabolic phase after severe burn or trauma are all of the following except which one?

A. Hyperglycemia
B. Low systemic vascular resistance
C. Modest hypotension
D. Hyperlipidemia
E. Increased cardiac output

Answer: D A massive sympathetic discharge leads to hyperglycemia, low systemic vascular resistance, increased cardiac output, and modest hypotension, but an increase in lipids is not part of this picture.

112 ENVENOMATION

GEOFFREY K. ISBISTER AND STEVEN A. SEIFERT

GENERAL CONCEPTS OF ENVENOMATION

EPIDEMIOLOGY

Terrestrial and marine envenomations result in a wide range of clinical syndromes and remain an important and unrecognized cause of morbidity and mortality, especially in the developing world. An estimated 5.5 million people are bitten by snakes each year, and snake envenomation is the most important envenomation syndrome worldwide, especially in tropical and subtropical countries. Snake bites cause an estimated 440,000 envenomation cases and 20,000 deaths each year.[1] Scorpions and spiders,[2] which also cause envenomation, are discussed in Chapter 359.

Marine envenomation is less common, except in coastal regions where hundreds of thousands and potentially millions of minor jellyfish stings occur each year. Venomous fish stings and injuries from other creatures such as sponges and sea urchins also occur, but the annual incidence is unknown. Worldwide, ciguatera is also a major clinical syndrome because of the large number of cases and morbidity in the Pacific region, where people rely on fish as a major food source. Fortunately few deaths occur. Other marine poisonings, including puffer fish poisoning and shellfish poisoning, often result in severe and potentially life-threatening effects but are rare.[3]

Venoms, Toxins, and Poisons

Envenomation includes any injury caused by a venomous creature that produces venom in a specialized gland and can deliver it to other organisms by means of fangs (e.g., snakes and spiders), stings (e.g., scorpions and jellyfish), or spines (e.g., fish). Venoms can be any combination of procoagulant, neurotoxic, myotoxic, and cytotoxic substances, not all of which affect humans because humans are not the intended victim of envenomation. Envenomation differs from poisoning, which occurs when mainly marine animals containing toxic substances (e.g., ciguatera) are ingested.

TREATMENT OF ENVENOMATION Rx

Very few first-aid techniques have been shown to be clinically effective for envenomations, so rapid transport to the hospital for definitive medical care is of utmost importance. Because envenomations such as snake bites and box jellyfish stings can result in early hypotensive collapse and cardiac arrest, the most important first aid is basic life support (Chapter 63).

Many snake venoms contain local tissue cytotoxins and techniques that concentrate venom at the bite site for prolonged periods and may result in greater local injury. Nonetheless, immobilization of the bitten extremity is recommended to minimize the venom's spread or systemic uptake, provided immobilization does not delay access to definitive therapy. With envenomations from Australasian elapids, kraits, and coral snakes, where there is likely to be only minor local tissue injury, a pressure bandage should be applied.[4] Pharmacologic agents that slow lymphatic flow (e.g., topical nifedipine or lidocaine) also may be useful as adjuncts in slowing venom uptake.[5] Other popular first-aid treatments that are often found in snake bite kits actually cause harm, including arterial or venous tourniquets, incision, suction, heat, cold, and electric shock.

Antivenom

Antivenom, which is the major treatment for envenomation, consists of a mixture of polyclonal whole or fractionated antibodies against the toxins in a specific venom. Because these antivenoms are produced in animals that are exposed to one or more venoms, they are generally specific to the venoms and their local species. Most antivenoms have never been tested in randomized controlled trials, so their true effectiveness is rarely known despite proven efficacy in preclinical studies. Antivenoms also cause a number of side effects, the most important being hypersensitivity reactions that can range from skin reactions to anaphylaxis (Chapter 253) that is not immunoglobulin E mediated and is characterized mainly by hypotension. Microbial contamination of antivenoms also can cause pyogenic reactions, which are characterized by fevers, rigors, and chills and may be associated with headache, gastrointestinal symptoms, and rarely hypotension. Anywhere from 6% to more than 30% of patients develop delayed antivenom serum sickness (Chapter 47), with influenza-like fever, arthralgia, myalgia, rash, urticaria, lymphadenopathy, headache, and gastrointestinal symptoms about 4 to 14 days after antivenom administration.[6]

The high frequency of reactions to some snake antivenoms has prompted the use of premedication, and one large randomized trial supports premedication with epinephrine (0.25 mL of 1 : 1000 solution subcutaneously) for antivenoms with a high reaction rate. By comparison, no evidence supports corticosteroids or antihistamines for premedication.

Other treatments for envenomation and poisoning focus on the specific but wide-ranging clinical effects caused by the various toxins and are not aimed at neutralizing the toxins or eliminating them. Examples include intubation and ventilation (Chapters 104 and 105) for neurotoxicity that leads to ventilatory muscle failure, hydration and renal replacement therapy in severe myotoxicity and acute kidney injury (Chapter 120), blood product replacement for severe coagulopathy with or without hemorrhage (Chapters 174 and 175), cardiac support for cardiac toxicities (Chapter 63), and pain relief (Chapter 30) in toxin-mediated pain syndromes such as scorpion stings and latrodectism and jellyfish stings.

SNAKE ENVENOMATION

Venomous snakes belong to one of five families: Viperidae, with its two subfamilies, Viperinae, or Old World vipers (e.g., saw-scaled vipers, puff adders), and Crotalinae, or pit vipers, named for the heat-sensitive organ between the eye and nostril used for hunting warm-blooded prey (e.g., rattlesnakes, copperheads); Elapidae (e.g., cobras, kraits, coral snakes); Hydrophiidae (e.g., sea snakes); Atractaspididae (e.g., asps); and Colubridae (e.g., garter snakes, corn snakes, boomslangs). Although all five families contain venomous species, the two responsible for more than 90% of venomous bites are Viperidae and Elapidae.

CLINICAL SYNDROMES

Local swelling and bruising after a bite are caused by increased vascular permeability as a result of endothelial cell damage and are mediated by hydrolases, proteases, phospholipase A_2, polypeptide toxins, metalloproteinases, and the release of endogenous autacoids such as bradykinin and histamine. Some venoms (e.g., some cobras) may cause necrotic skin lesions or extensive swelling and bruising. Other venoms (e.g., European adders and North American rattlesnakes) increase vascular permeability and result in significant regional swelling and edema. In rare cases, the loss of fluid into the tissues can cause hypovolemic shock.

One type of toxin will usually predominate in a particular snake, and the combination of toxins will result in a somewhat unique clinical syndrome for each snake. For example, bites from *Bungarus* spp (kraits) usually result in neurotoxicity, whereas *Echis* spp (saw-scaled or carpet vipers) primarily cause a coagulopathy with spontaneous bleeding.

Abnormal Coagulation

Coagulopathy (Chapters 175 and 176) is the most important snake envenomation syndrome, and venom-induced consumption coagulopathy is the most common form.[7] The syndrome results from the activation of the clotting pathway at a number of specific points due to procoagulant

toxins in snake venoms. (See also http://wikitoxin.toxicology.wikispaces.net/Venom+induced+consumption+coagulopathy+-+VICC.) Important procoagulant toxins include fibrinogenolytic toxins or thrombin-like enzymes (e.g., Malaysian pit viper, American vipers), prothrombin activators (e.g., Australian elapids), and factor X activators (e.g., Russell's viper venom), all of which result in low or undetectable fibrinogen levels.

Venom-induced consumption coagulopathy is an acquired acute clotting factor deficiency that persists until clotting factors can be resynthesized. Venom-induced consumption coagulopathy results in an abnormal international normalized ratio (INR), activated partial thromboplastin time (aPTT), and thrombin clotting time as well as elevated D-dimer levels and fibrinogen degradation products. The coagulopathy may not result in hemorrhage unless it is accompanied by associated injury or trauma. Unlike disseminated intravascular coagulation (Chapter 175), the syndrome has a rapid onset and resolution and is not associated with systemic microthrombi, end-organ failure, or poor clinical outcomes.

Some venoms can cause a thrombotic microangiopathy, with clinical manifestations (e.g., thrombocytopenia, microangiopathic hemolytic anemia, and acute renal failure) similar to what is seen in other microangiopathies, such as thrombotic thrombocytopenic purpura and the hemolytic-uremic syndrome (Chapter 172). This syndrome has been reported for Australian elapids, Russell's viper, saw-scaled viper, and desert horned viper.

Conversely, venom-induced anticoagulant coagulopathy is characterized by an elevated aPTT and a mildly abnormal INR. It has been reported in Australian black snake envenomation but does not appear to cause a clinically important coagulopathy. Thrombocytopenia occurs in many snakes and is prominent in North American Crotalinae spp because of decreased production, aggregation, and sequestration of platelets.

Neuromuscular Toxicity

Snake neurotoxins, which occur mainly with elapid and some viperid venoms, are usually classified as presynaptic neurotoxins (e.g., β-bungarotoxins, taipoxin), which injure the presynaptic neuromuscular terminal and prevent acetylcholine release, or postsynaptic toxins (e.g., α-bungarotoxins), which bind to acetylcholine receptors on the motor end plate. Presynaptic neurotoxins generally result in irreversible neurotoxicity that may take days to weeks to resolve if it is not treated early with antivenom. The clinical presentation is initially with ptosis and extraocular ophthalmoplegia, which progress to bulbar palsy, and finally respiratory and peripheral muscle paralysis.

Myotoxic envenomation syndromes are divided into two clinical types, local myotoxicity and systemic myotoxicity or rhabdomyolysis (Chapter 421). Local myotoxicity is most commonly seen when viper envenomation affects muscles near the bite site. Symptoms range from local myalgia with swelling and pain to myonecrosis. Moderate (10-fold) transient creatine kinase (CK) concentration elevations are seen for up to 48 hours. In systemic myotoxicity, widespread muscle injury results in generalized myalgia, a raised CK concentration to as high as 100,000 U/L for several days, and myoglobinuria (Chapter 113). Systemic myotoxicity is seen with sea snakes, Australian elapids, and the South American rattlesnake *Crotalus durissus*. However, acute renal failure is uncommon.

Other Effects

Some snakes (primarily vipers and Australian elapids) produce an acute hypotensive syndrome within minutes of the bite, apparently due to the release of vasodilating autacoids. *Bothrops* species inhibit the breakdown of bradykinin and angiotensinogen, which are naturally occurring, rapid-acting angiotensin-converting enzyme inhibitors. Other cardiovascular effects include vasodilation, diffuse vascular permeability, myocardial depression (Chapter 58), and atrioventricular block (Chapter 64), which may contribute to the hypotension caused by crotalines and elapids. Acute kidney injury (Chapter 120) is uncommon but can be caused by a direct renal toxin, hypotension with resulting acute tubular necrosis, or rhabdomyolysis or as part of a thrombotic microangiopathy.

Many envenomations cause nonspecific systemic symptoms, including nausea, vomiting, headache, abdominal pain, diarrhea, and diaphoresis. Although these effects are not life-threatening, they may be useful early markers of envenomation and may require symptomatic treatment.

DIAGNOSIS

Most patients present with an obvious history of a snakebite, but they may not have envenomation. The initial assessment and investigation needs to determine

- whether the patient has snake envenomation or a bite without envenomation;
- what snake or snake group is likely to have caused the bite; and
- whether antivenom is indicated and, if so, what antivenom is available and should be used.

In some patients, signs or symptoms of envenomation may be obvious. Other patients, however, initially may not have any evidence of envenomation and should be observed with serial clinical and laboratory assessment, depending on the region and type of snake potentially involved (e.g., vipers may inject large amounts of venom with delayed absorption) as well as the severity of the patient's clinical syndrome.

Laboratory Investigations

If they are available, the most important laboratory studies are coagulation studies: INR, aPTT, and platelet count. Point-of-care testing devices for measuring the INR are insensitive to the changes caused by venom-induced consumption coagulopathy and should not be relied on. If they are available, measures of fibrinogen and D-dimer can assist in the diagnosis of venom-induced consumption coagulopathy.

A complete blood count can identify thrombocytopenia and thrombotic microangiopathy; nonspecific leukocytosis and lymphopenia can occur in snake envenomation and may provide support for the diagnosis. An elevated CK level indicates systemic myotoxicity but lags behind the muscle injury by 12 to 24 hours. A serum creatinine level is important to assess possible acute kidney injury. Measurements of venom and antivenom in serum are available only in the research setting.

TREATMENT Rx

The most important first-aid measures in snake bite are to move a safe distance from the snake, to immobilize the patient as well as the bitten body part, and to organize transport to medical care. Unproven and potentially dangerous first-aid techniques that should not be used include suction, cutting the bite site, electrocution, and the application of heat, cold, or tourniquets. Pressure bandaging with immobilization (at lymphatic obstruction pressures of 55 to 70 mm Hg) is recommended for Australian snake bites and some other elapids but should not be used for the majority of other snakes, including all viperid snakes.

Patients whose envenomation is manifested as coagulopathy, myotoxicity, or neurotoxicity should be administered antivenom and admitted until the clinical effects have resolved. The specific antivenom, the indications for its use, and the dosing regimens depend on the type of snake and the geographic region. Regional and national guidelines should be consulted.

Local Tissue Injury and Local Myotoxicity

Most local injuries can be treated with supportive care, including analgesia (e.g., an anti-inflammatory such as ibuprofen, 400 to 800 mg, and opioids such as morphine, 2.5 to 10 mg intravenously, titrated to the pain), elevation, and standard wound care. Some viperid bites will cause significant local and regional swelling with increased tissue pressures, but early surgical treatment is not indicated and may be associated with worse outcomes. Late surgical débridement of necrotic tissue may be required. There does not appear to be a role for prophylactic antibiotics, which should be reserved for documented infection.

Abnormal Coagulation

The treatment of venom-induced consumption coagulopathy includes antivenom to neutralize the procoagulant toxins, supportive or interventional care for any spontaneous or traumatic hemorrhage, and blood or factor replacement. The clinical effectiveness of antivenom is supported by a number of observational studies but depends on the type of procoagulant toxin and the timing of antivenom administration. In one of the most striking observational studies, *Echis* antivenom corrected coagulopathy within 24 to 48 hours, whereas untreated patients had deranged hemostasis for 8 to 10 days.[8] Similar benefits have been reported for other antivenoms. In one randomized trial of Australian elapid bites, fresh-frozen plasma restored clotting function more rapidly but did not decrease major hemorrhage or death, and it may even worsen the coagulopathy if it is given within 6 hours of the bite.

Thrombotic microangiopathy (Chapter 172) requires close monitoring and supportive care. Antivenom may prevent thrombotic microangiopathy, but repeated doses are of no benefit. Major hemorrhage such as intracranial (Chapter 408) and gastrointestinal bleeding (Chapter 135) should be managed per standard protocols, including blood transfusion as required (Chapter 177). There is no evidence to support the use of plasmapheresis in snakebite-induced thrombotic microangiopathy.

Neuromuscular Toxicity

Antivenom is the only specific treatment for neurotoxicity, and numerous studies have demonstrated that early antivenom can prevent neurotoxicity. However, because most neurotoxicity is presynaptic, antivenom will not reverse neurotoxicity once it appears, and delayed administration may be of no benefit. For snakes that cause neurotoxicity (Chapters 420 and 422), repeated examinations should look for ptosis, extraocular muscle weakness, bulbar palsy, and respiratory abnormalities. Intubation and mechanical ventilation may be required for respiratory muscle weakness.

Systemic myotoxicity with rhabdomyolysis (Chapter 113) is irreversible but can be prevented with early antivenom administration. Serial measurement of the CK level is essential to monitor systemic myotoxicity.

Specific Snake Syndromes

The worldwide distribution of the most common snakes of medical importance is available at either www.toxinology.com or http://apps.who.int/bloodproducts/snakeantivenoms/database/snakeframeset.html.

NORTH AMERICAN SNAKES

In the United States, an estimated 8000 bites by native venomous snake species occur annually, mostly by pit vipers—rattlesnakes (*Crotalus* and *Sistrurus* spp), copperheads, and cottonmouths (*Agkistrodon* spp); fewer than 100 bites occur by coral snakes, which are the only native elapid (*Micrurus* spp and *Micruroides euyxanthus*). Fortunately, the number of fatalities, primarily caused by rattlesnakes, is fewer than 10 per year. Because the crotaline antivenom in the United States (CroFab) is effective against all North American crotaline species, because envenomations are readily clinically diagnosable, and because coral snakes are easily recognized, there is no need to capture or kill the offending snake for identification. In the prehospital setting, management priorities include avoidance of further interaction with the snake, removal of jewelry, loosening of tight-fitting clothing, loose splinting of the bitten body part, and rapid transportation to a health care facility capable of treating snakebite.

Envenomation by North American viperids invariably produces local tissue injury, with progressive swelling and pain. However, frank necrosis occurs in only about 5% of cases, and rates of hematologic (prothrombin time/INR and aPTT prolongation, thrombocytopenia, or hypofibrinogenemia), neurologic, hemodynamic, and nonspecific symptoms and signs vary. Definitive management is antivenom plus elevation of the bitten body part. Surgical intervention is not beneficial in the acute phase. Infection is uncommon, and prophylactic antibiotics are not recommended. Antibiotics may also be added if there is tissue necrosis or signs of infection. Unless there are signs of an immediate hypersensitivity reaction to the venom, there is no role for antihistamines or steroids. Local venom effects may recur within the first 24 hours. Hematologic effects may recur after several days and persist for days or weeks, and about 1% of patients may develop late bleeding complications.[9] Monitoring and management strategies are complex, and expert assistance is suggested.[10]

Clinical effects of coral snakes, primarily the Eastern coral snake (*Micrurus fulvius*), include neurotoxicity and minimal local effects. With severe envenomation, perioral paresthesias, nausea, vomiting, hypersalivation, and euphoria progress to cranial nerve paralysis (e.g., ptosis, diplopia, and dysphagia) and respiratory failure. Respiratory failure may develop within minutes of the onset of neurologic signs. Because of a shortage of antivenom, some clinicians have elected to wait for signs of envenomation before administering antivenom. If this approach is taken, patients should be given antivenom at the first signs of neurotoxicity and observed for at least 24 hours before it is concluded that no envenomation has occurred.

There are approximately 50 venomous snake bites a year in the United States by non-native (exotic) species. Antivenoms for these snakes as well as expert assistance in the management of envenomations can be obtained by contacting the appropriate poison center (1-800-222-1222).

CENTRAL AND SOUTH AMERICA

Snake envenomation in Central and South America is associated with more morbidity and mortality than in North America, despite the availability of antivenom for most venomous species. One reason is the larger variety of snakes, most importantly viperid snake species (*Crotalus*, *Bothrops*, and *Lachesis*) and the coral snakes (elapids) (see http://wikitoxin.toxicology.wikispaces.net/Latin+American+Snakebite). Various South and Central American vipers induce bite site necrosis, consumption coagulopathy, and bleeding. As in North America, coral snake envenomation causes a descending paralysis but few local effects. Polyvalent antivenoms are available in Central and South America to cover the particular vipers in different geographic regions. (See http://apps.who.int/bloodproducts/snakeantivenoms/database/snakeframeset.html.)

ASIAN SNAKES

The incidence of snake envenomation in Asia, particularly in south Asia, is potentially the highest in the world, with significant mortality and morbidity. (See http://wikitoxin.toxicology.wikispaces.net/Asian+Snakebite.) As a general rule, vipers cause coagulopathy and elapids cause neurotoxicity. Russell's vipers, saw-scaled vipers, and pit vipers (e.g., Malayan pit vipers) cause venom-induced consumption coagulopathy, except the mechanism of action of the procoagulant toxin differs (see http://wikitoxin.toxicology.wikispaces.net/Venom+induced+consumption+coagulopathy+-+VICC). Viper bites are also associated with local tissue injury and nonspecific systemic effects. Neurotoxicity, which is a major problem with krait envenomation and some cobras, results in an irreversible descending paralysis that develops during hours. Many krait bites occur at night when the snakes are active and patients are often not aware of the bite, so patients often present late with established neurotoxicity.

Treatment protocols for snake bite differ throughout Asia and depend on local resources, the availability of antivenom, and the presumed venomous snakes. Antivenom is the main treatment and is manufactured in a number of countries (http://apps.who.int/bloodproducts/snakeantivenoms/database/snakeframeset.html). Venom-induced consumption coagulopathy resulting from Asian snakes appears to respond to antivenom treatment, which may be effective days after the bite. The irreversibility of neurotoxicity means that intubation and mechanical ventilation are often required, particularly for patients presenting late when antivenom is unlikely to be effective. In krait envenomation, patients may require ventilation for days to weeks.

AFRICAN SNAKES

Snake envenomation is a major problem in Africa (see http://wikitoxin.toxicology.wikispaces.net/African+Snakebite), especially in locations with limited health care and availability of antivenom. In addition, the offending species of snakes are not as well known or categorized.

The puff adder is widely distributed and likely to cause many bites, but most of the bites cause only local swelling, blistering, and uncommonly necrosis. Systemic envenomation causes early hypotension, bradycardia, thrombocytopenia, and uncommonly coagulopathy.

Carpet vipers (*Echis*), which are the most important snakes in western and northern Africa and across to the Middle East, cause a significant coagulopathy that may persist for 7 to 14 days without antivenom. The desert vipers (*Cerastes*) of northern Africa and across into the Middle East cause local effects and rarely a mild coagulopathy.

Cobras are widely distributed elapids divided into the cytotoxic/spitting cobras (*Naja nigricolis*, *Naja mossambica*) and the neurotoxic cobras. The cytotoxic/spitting cobras can spit venom into the eye and cause venom ophthalmia. Bites by cytotoxic cobras cause nonspecific systemic symptoms and necrosis. In contrast, the neurotoxic cobras (e.g., Egyptian cobras, or *Naja haje*) cause progressive descending paralysis. Mambas are large elapids that cause neurotoxicity characterized by paresthesia, muscle fasciculations, and weakness, associated with nonspecific and autonomic effects (sweating, hypertension, tachycardia, vomiting). The boomslang (*Dispholidus typhus*) is a Colubrid that causes a severe venom-induced consumption coagulopathy that results in spontaneous and major bleeding and uncommonly thrombotic microangiopathy. The burrowing asps are reported to cause local effects and nonspecific symptoms.

Although a number of snake antivenoms are available for African snakes, they often are not readily available, and many patients do not present to a hospital. Polyvalent antivenom from South Africa covers many of the important snakes. Other antivenoms are available and details are at http://apps.who.int/bloodproducts/snakeantivenoms/database/snakeframeset.html.

EUROPEAN SNAKES

Snake envenomation is uncommon in Europe, and the only medically important species are *Vipera* spp (true vipers) that occur across most of Europe including England, almost all of continental Europe down to the Mediterranean, and also across into north Asia and northern Africa (Morocco, Algeria, and Tunisia). The most important species are *Vipera berus*, *Vipera aspis*, and

Vipera ammodytes, although the epidemiology of snake envenomation is not well defined for many countries.

The major effects of *Vipera* bites are local swelling and regional swelling, tissue injury, hematomas, nonspecific systemic symptoms (vomiting, diarrhea, abdominal pain), and hypotension. Coagulopathy and neurotoxicity are uncommon. The severity of envenomation can be graded (see table at http://wikitoxin.toxicology.wikispaces.net/European+Snakebite). Antivenoms for *Vipera* bites are available in a number of countries and are regarded as the main treatment.[11] Supportive care, including intravenous fluids for hypotension and elevation of the bitten limb, is also important.

AUSTRALASIAN SNAKES

In Australasia, the medically important venomous snakes include brown snakes (*Pseudonaja* spp), tiger snakes (*Notechis* spp), black snakes (*Pseudechis* spp), taipans (*Oxyuranus* spp), and death adders (*Acanthophis* spp). All are elapids and all look similar, except the death adder. As a result, the choice of antivenom is usually based on geography and clinical effects and only rarely on the identification of the snake.

Each snake group causes a characteristic clinical syndrome. Venom-induced consumption coagulopathy is the most common with all except black snakes and death adders. Myotoxicity occurs with black snakes and tiger snakes, and it also occurs less commonly with taipans. Neurotoxicity is the only clinical effect of death adder envenomation, is the major clinical effect in taipan envenomation, and can occur with tiger snakes. (See http://wikitoxin.toxicology.wikispaces.net/Australian+Snakebite.) Local effects are uncommon, so pressure bandaging is the current standard first-aid method.

About 90% of suspected snake bites in Australia are nonenvenomed cases or dry bites, so treatment focuses on excluding envenomation. Five monovalent antivenoms are available, depending on the clinical features and geography. A snake venom detection kit can detect venom on a wound swab from the five major snake groups and may assist in determining the type of snake. If the type of snake remains unclear on the basis of clinical effects and geography, polyvalent antivenom should be administered. The dose is one vial for all antivenoms, and redosing is not required. (See http://wikitoxin.toxicology.wikispaces.net/Australian+Snakebite.)

SNAKE HANDLERS AND EXOTIC SNAKE BITE

A considerable number of snake bites occur in snake handlers, including zoo keepers, private collectors, wild-life rescuers, and venom researchers. In the United States, for example, bites were reported from more than 90 different non-native species during a 10-year period. The case-fatality rate of exotic envenomation can be an order of magnitude higher than for native species, in part because of an immunoglobulin E–mediated anaphylactic reaction (Chapter 253) to the venom in individuals who handle snakes or who have been bitten previously. Early expert advice and contact with a poison center are essential to access information and antivenom.

MARINE ENVENOMATION

A wide variety of venomous marine creatures exist worldwide, including invertebrates and vertebrates. Fortunately, the limited number of types of marine envenomation result in only a few major envenomation syndromes, with more similarities than differences.[12] Perhaps the most important venomous marine creatures are the jellyfish, including the major box jellyfish (*Chironex fleckeri*), which is regarded as the most toxic venomous creature. Other important venomous marine creatures are the venomous fish, echinoderms (including sea urchins), sea snakes, and sponges.

Jellyfish Envenomation

EPIDEMIOLOGY

More than 100 jellyfish (Cnidaria) of medical importance exist worldwide, with a range of envenomation from very minor to potentially life-threatening. The epidemiology remains unclear, with millions of minor unreported injuries occurring in some coastal parts of the developed world and an unknown number of cases in other parts of the world. Deaths from major box jellyfish stings are rare and occur mainly in northern Australia, but such events may be underestimated in resource-poor countries.

PATHOBIOLOGY

Jellyfish are covered by millions of nematocysts (stinging cells), mainly on their tentacles. When these nematocysts are triggered by chemical or physical stimuli, such as contact with human skin, venom released from these stinging cells results in dermal injection and clinical envenomation. The severity of the envenomation depends on the amount of tentacle contact and the potency of the venom. Envenomation with the major box jellyfish, which have multiple long tentacles and highly potent venom, can be life-threatening, whereas stings with *Physalia* species tend to be minor.

CLINICAL MANIFESTATIONS

Envenomation by jellyfish can be broadly divided into two major clinical syndromes. With linear/tentacle stings, contact with one or more jellyfish tentacles results in local pain of varying severity and a raised linear erythematous or urticarial lesion that persists for a few hours to a day. Rarely these stings are associated with nonspecific systemic symptoms such as dizziness, nausea, vomiting, and malaise. With Irukandji-like stings, there is minimal local pain and evidence of the sting, but the envenomation syndrome results in systemic effects, including severe generalized back, chest, abdominal, and large muscle pain associated with nausea, vomiting, and headache within about 20 to 30 minutes after contact with the jellyfish. Sympathomimetic-like effects, including tachycardia, hypertension, anxiety, and agitation, also occur. With severe envenomation, cardiac effects, likely secondary to this catecholaminergic excess, can include electrocardiogram changes (T wave changes and ST segment depression), elevated troponin levels, myocardial depression, and even cardiogenic pulmonary edema.

Common and Important Jellyfish Groups and Syndromes

Numerous jellyfish result in specific syndromes. (See http://wikitoxin.toxicology.wikispaces.net/Jellyfish.) *Physalia* species are widespread and have different common names throughout the world, including Portuguese man-of-war, Pacific man-of-war, and blue bottle. They are the most common jellyfish stings and are responsible for hundreds of thousands of stings in America, Australia, and Europe each year, although most victims do not seek medical attention. *Physalia* stings cause typical linear erythematous reactions with moderate to severe local pain persisting for a few hours. The linear markings, which may have a characteristic ladder-like or beaded appearance, persist for several days. Localized bullous reactions and scarring can occur, but systemic effects are rare and deaths are extremely rare. *C. fleckeri*, which is a box jellyfish (Cubozoa), is reported to be the most venomous creature in the world. *C. fleckeri* has resulted in more than 70 deaths in northern Australia, the majority in young children in remote areas. Like all box jellyfish, they are cube shaped with long tentacles attached to each corner. Contact with them can result in several meters of skin involvement and severe envenomation. Fortunately, however, most stings are minor linear tentacle stings. Linear erythematous eruptions are similar to other linear tentacle stings but usually more severe, with severe pain persisting for hours and requiring opioid analgesia (Chapter 30). Local necrosis can occur along the sting line, but permanent scarring is rare. Severe envenomation is characterized by profound hypotension and early cardiovascular collapse, which can result in death in 20 to 30 minutes. Delayed hypersensitivity reactions occur in more than 50% of patients, with papular urticarial lesions developing along the stings. Numerous other jellyfish include the sea nettle (*Chrysaora quinquecirrha*), which is important in the Chesapeake Bay area of America but also occurs in Japan and the Philippines. It has effects similar to those of *Physalia*. Hair jellyfish cause stings to the eyes and cornea because the tentacles break off. *Pelagia* spp (moon jellyfish) are also common in parts of the world and cause effects similar to those of *Physalia*.

TREATMENT Rx

The treatment of the majority of jellyfish stings occurs out of the hospital. The jellyfish is rarely identified, so treatment is based on the clinical effects and the known local jellyfish. Most victims do not seek formal medical care unless they have severe, persistent pain or systemic effects that require medical intervention. First aid can be given by bystanders, surf life-savers, first-aid personnel, or ambulance services.

First aid includes washing off tentacles or any jellyfish material with sea water or carefully removing them by hand.[13] The use of fresh water may cause additional nematocyst discharge and potentially worsen the effects. For some major box jellyfish (including *C. fleckeri*) that can cause severe systemic envenomation, vinegar, which is often available on beaches where these jellyfish stings occur, should be applied immediately. Despite limited evidence, the aim of the vinegar is to prevent further nematocyst discharge and severe systemic envenomation from major jellyfish stings. Vinegar may increase the pain and is not recommended for these less severe jellyfish stings.

Box jellyfish stings have the potential to cause severe systemic envenomation within the first 30 minutes. Cardiovascular collapse or arrest should be treated with standard advanced life support (Chapter 63). Antivenom is unlikely to be effective for *C. fleckeri*, but guidelines still recommend its use in severe envenomation if it is available to be given by the intravenous route. Patients who arrive at the hospital without cardiovascular effects are highly unlikely to develop severe envenomation after arrival.

Physalia stings should be treated with hot-water immersion (45°C for 20 minutes) to treat local pain.[■] An alternative is a hot shower or a constant flow of hot water. Analgesia or local dressings are rarely required for skin reactions.

For more severe stings, such as *C. fleckeri* stings, oral or parenteral opioids (Chapter 30) may be required. Numerous topical treatments, such as bicarbonate slurry, urine, or sand, are not supported by any evidence and are not recommended. Most stings do not require any local treatment, but necrosis can occur in severe cases and should be treated similar to burns (Chapter 111) with appropriate dressings.

Irukandji syndrome or Irukandji-like stings often require in-hospital treatment for severe pain and systemic symptoms. Titrated intravenous opiates (Chapter 30) should be given in combination with an antiemetic (e.g., metoclopramide, 10 mg; see Table 132-5). Benzodiazepines may be useful for agitation and anxiety. All patients with Irukandji syndrome should have an electrocardiogram and measurement of the serum troponin level because of the risk of cardiac toxicity. Patients can be discharged if they have a normal electrocardiogram and troponin level and are pain free. However, patients with evidence of cardiac involvement should be admitted for monitoring and an echocardiogram. Acute pulmonary edema should be treated per standard protocols (Chapter 59).

Evidence for the treatment of other jellyfish stings is limited, but first aid should include washing off the tentacles with sea water. Hot water may be effective, but it has not been rigorously tested.

Venomous Fish and Stingray Injuries

EPIDEMIOLOGY

Venomous fish occur in tropical and sometimes in temperate oceans. The important venomous fish, which differ throughout the world, include catfish (Siluriformes), stonefish (Synanceiidae), bullrout, weever fish (Trachinidae), scorpion fish, and lionfish (Scorpaenidae). Venomous fish are also kept in private aquariums, and many injuries in North America are due to lionfish in captivity. Stingrays, which vary in size from smaller than a hand to more than a meter, can live in tropical and temperate oceans as well as in fresh water.

PATHOBIOLOGY

Venomous fish and stingrays have a venomous sheath-covered spine that is formed on the dorsal or ventral spine of the fish and the tail of a stingray. The venom, which is between the spine and the sheath, is injected as the spine penetrates the skin. However, the majority of the injury is simply due to penetrating trauma, particularly in the case of stingrays with their much larger spines. Although the venoms differ among types of venomous fish and stingrays, they appear to be responsible for the severe pain out of proportion to the trauma and possibly contribute to the slow recovery and risk of infection with some wounds.

The site of injury varies for different fish and depends on how they encounter humans. Stingrays most commonly cause injuries to the ankle when they are stepped on, stonefish and bullrout cause injuries to the foot when they are stepped on, and catfish cause injuries to the hands when they are being caught.

CLINICAL MANIFESTATIONS

Symptoms vary among different fish and stingrays (see http://wikitoxin.toxicology.wikispaces.net/Penetrating+Marine+Envenoming). The main effects are a puncture wound or laceration associated with pain that varies on the basis of the size of the injury and the venom injected.[14] Severe and prolonged site pain may occur with stonefish, bullrout, some marine catfish, and weever fish. In addition, local bleeding, particularly with stingrays, and surrounding swelling and edema can occur. The spines rarely remain in the wound, but small amounts of foreign material often do and represent a considerable risk of secondary marine infection. Systemic effects are rarely reported but may include nonspecific symptoms such as nausea, vomiting, and malaise.

Catfish injuries most commonly occur when fishermen remove them from fishing lines. The pain of the puncture wound varies from mild with nonvenomous spines to severe with venomous spines such as the striped catfish (*Plotosus lineatus*). Bleeding and secondary infection are rare.

Stonefish typically cause injuries to the sole of the foot. The pain is usually very severe and out of proportion to the local trauma. Significant swelling and edema of the foot may extend up the lower leg. Systemic effects have been reported but are uncommon.

Most lionfish injuries, which are the most common venomous fish injury in the United States, usually occur in private aquariums when they are being cleaned. Stings cause severe local pain that may be associated with swelling and edema. Systemic effects are rare and minor.

Stingray injuries, which cause more trauma than venom-mediated effects, generally occur when people tread on them in shallow water, although they are occasionally picked up and rarely cause thoracoabdominal trauma to divers who are swimming too close to them. Laceration, local pain, and bleeding are the major effects. The wound may become necrotic and there is a risk of secondary infection, worse with larger wounds and those with foreign debris.

TREATMENT Rx

The first aid for penetrating venomous marine injuries is washing the site and immersing it in hot water (45°C) for up to 90 minutes if this improves the pain. Unlike for jellyfish, there is limited evidence to support hot water for penetrating injuries, and the pain often recurs as soon as the hot water is removed. It is essential to test the temperature and not to immerse for more than 90 minutes so that superficial burns do not occur.

If the pain does not respond to hot water, a combination of ibuprofen (200 to 400 mg every 8 hours), acetaminophen (1 g every 4 to 6 hours), and oxycodone (5 to 10 mg every 4 hours) can be used. The pain rarely persists for more than 12 to 24 hours. Titrated intravenous opiates (Chapter 30) may be required, but local or regional local anesthetic infiltration is often more effective and can also assist with cleaning of the wound.

Careful wound management, which is key in all cases of penetrating marine injuries, includes cleaning and irrigating the wound. In some cases, radiographic images or ultrasound may be useful to evaluate or to exclude any retained foreign bodies. Most wounds do not require closure, and larger wounds will heal by secondary intention, although they may require surgical exploration and débridement. The use of prophylactic antibiotics remains unclear, with limited evidence supporting their use. Wounds that enter sterile body cavities (e.g., joints) require surgical exploration. Stingray injuries to the abdomen or chest should be managed as major thoracoabdominal trauma (Chapter 111).

Secondary infections that may uncommonly complicate penetrating marine injuries are usually *Vibrio* spp (Chapter 302) in the marine environment or *Aeromonas* spp (Chapter 359) in fresh water. Both are associated with significant morbidity and mortality if they are not treated early and aggressively. It is essential that all cases be observed closely in the first week. If there is any evidence of infection, wound swabs or aspirates for marine organisms should be sent and antibiotics (e.g., ciprofloxacin, 400 mg twice daily) should be commenced while awaiting culture and sensitivity test results. Surgical drainage may sometimes be required.

An antivenom is available for stonefish stings if the pain does not resolve with analgesia. Antivenom is given diluted by the intravenous route.

Sea Snake Envenomation

Sea snakes are found in the tropical parts of the Indian and Pacific oceans. They differ from eels because they have scales and do not have fins or gills. Sea snakes have small fangs similar to elapids, and their venom contains myotoxins and neurotoxins. Among the number of different species, the beaked sea snake (*Enhydrina schistosa*) is the most medically important.

Sea snake bites cause minimal pain, and the patient may not be aware of the bite. Systemic effects, which develop during hours, include nausea, vomiting, and headache as well as myotoxicity with myalgia of the face, neck, limbs, and trunk, sometimes with trismus. The CK level rapidly rises and usually is diagnostic. Rhabdomyolysis (Chapter 113) occurs in severe cases and may precipitate acute kidney injury. A rare sea snake bite may cause predominant neurotoxicity.

First aid is with a pressure bandage and immobilization. Polyvalent sea snake antivenom is manufactured in Australia (CSL Ltd, Melbourne). If it is available, one vial should be administered early. Intravenous fluids should be given, and renal function should be monitored.

Echinoderms

The two major echinoderms that cause human injury are sea urchins and the crown-of-thorns sea star. Sea urchins, which are worldwide in distribution, cause injuries when they are picked up or more commonly when they are stepped on. The spines, which range from chalky material to much stronger

thorns, are usually nonvenomous. Crown-of-thorns sea stars are found in the Indo-Pacific region and cause injuries similar to those of sea urchins.

Local pain is rarely severe, and the major issue is the presence of retained spines, which are often multiple, are difficult to find, and may cause ongoing pain. The treatment of sea urchin injuries is similar to that of venomous fish stings, with hot-water immersion and local wound cleaning. However, removal of the spines can be a major problem because of the potential for multiple difficult-to-remove foreign bodies. Radiography and ultrasound may assist in locating spines. Any spines close to the surface should be removed, and the patient should be observed carefully until the symptoms resolve. If patients have ongoing pain from retained spines, surgical consultation is required.

Sponges

Tedania spp (fire sponges) and *Neofibularia* spp are among a number of medically important sponges that produce toxic secretions. Stinging sponge dermatitis, which is uncommon, results from contact with the sponge.[15]

The majority of sponge stings are minor with local pain, paresthesia, itchiness, and numbness. The symptoms develop during hours and may persist for several days. Symptoms are associated with local erythema and sometimes with vesicular reactions and stiffness. An unusual manifestation of fire sponge contact is delayed pain, swelling, and erythema at the contact site 2 to 3 weeks later, followed by desquamation.

The sting site should be washed as soon as possible. Most cases can be treated with analgesia (e.g., ibuprofen, 400 mg three times daily; see Table 30-3).

MARINE POISONING

Marine poisoning is uncommon except for ciguatera, which remains a major problem in many parts of the Pacific where fish are the main source of protein in the diet. Although marine poisoning is rare, the transport of fish by air has meant that marine poisoning can occur anywhere the fish are served. Most marine poisonings, including ciguatera, shellfish poisoning, and tetrodotoxin poisoning, result in neurotoxic effects.

Ciguatera

Ciguatera is endemic to parts of the Caribbean and Indo-Pacific, but it can occur anywhere fish are transported. Ciguatera poisoning affects about 60 individuals annually in the United States.[16] It results from toxins that are accumulated in tropical reef fish. A large number of fish have been implicated, including Spanish mackerel, bass, moray eels, some cod species, coral trout, and emperors (see online table at http://wikitoxin.toxicology.wikispaces.net/Ciguatera). A major issue is that it is not possible to determine if a particular serving of fish can cause ciguatera.

The clinical manifestations of ciguatera are mainly gastrointestinal in the Caribbean and a combination of gastrointestinal and neurologic in the Indo-Pacific region. The gastrointestinal effects, which include vomiting, diarrhea, and abdominal cramping, start within hours and may persist for up to 24 hours. The major neurologic effect is a sensory polyneuropathy (Chapter 420), which is delayed and develops within 24 hours. The almost pathognomonic feature of ciguatera is cold allodynia, which is often referred to as heat reversal but is in fact an unpleasant discomfort or sensation when touching cold objects. Other neurologic features are perioral and distal paresthesia and numbness. Patients may also develop myalgia, arthralgia, and pruritus (see online table at http://wikitoxin.toxicology.wikispaces.net/Ciguatera). The clinical diagnosis is based on the history of eating fish combined with the gastrointestinal or neurologic effects.

No antidote exists for ciguatera, and treatment is symptomatic relief and supportive care. Intravenous fluids should be given for dehydration, but there is no evidence for the use of mannitol.[A5] Nonsteroidal anti-inflammatory agents (ibuprofen, 200 to 800 mg three times daily) should be given for acute symptoms. Tricyclic antidepressants, gabapentin, and calcium antagonists have been suggested for chronic symptoms, but data on their efficacy are lacking.

Tetrodotoxin (Puffer Fish) Poisoning

Tetrodotoxin poisoning results from the ingestion of a number of types of bony fish, including puffer fish and toad fish; it is also known as fugu poisoning in Japan. Tetrodotoxin is a sodium-channel blocker that interrupts nerve conduction and results in paralysis (Chapter 420).

Tetrodotoxin poisoning causes a sensorimotor neuropathy and mild gastrointestinal symptoms with mainly nausea and occasional vomiting. The rate of onset of toxicity is more rapid with more severe poisoning, which usually develops within an hour. The major clinical effects are perioral numbness and paresthesia, ataxia due to weakness, distal to proximal muscle weakness, and respiratory muscle paralysis that can be fatal. A clinical grading system is available at http://wikitoxin.toxicology.wikispaces.net/Tetrodotoxin+Poisoning.

The treatment for tetrodotoxin poisoning is supportive care because no antidote exists. Prehospital management of the airway and breathing is required in severe cases. Patients may require mechanical ventilation for 2 to 5 days.

Shellfish Poisoning

Shellfish poisoning is rare and occurs sporadically when there are algal blooms in areas where shellfish are collected. Shellfish poisoning is divided into four types, three that are neurologic and one that is gastrointestinal. Paralytic shellfish poisoning, which is the most common shellfish poisoning, is clinically indistinguishable from tetrodotoxin poisoning because it is due to other sodium-channel blockers, including saxitoxin and gonyautoxins. It has a high fatality rate because of the respiratory paralysis. Treatment is the same as for tetrodotoxin poisoning. Neurotoxic shellfish poisoning is much rarer and is a neuroexcitatory syndrome that is similar to ciguatera. Treatment is supportive.

Encephalopathic shellfish poisoning has been reported only once in North America. It results from domoic acid, and there is no specific treatment. Diarrhea from shellfish poisoning causes a severe gastroenteritis, which is rapid in onset and can be severe with fluid loss and hypovolemic shock. Treatment is supportive with aggressive fluid therapy.

Scombroid

Scombroid, which is similar to an acute hypersensitivity reaction (Chapter 47), results from ingestion of fish that contain high concentrations of histamine.[17] The high histamine concentrations are due to spoilage when the fish are stored or transported after they are caught. The most commonly implicated fish are from the Scombridae family, including kingfish, mackerel, wahoo, and tuna.

The clinical manifestations, which are due to the histamine in the fish, may resemble an acute hypersensitivity reaction. Signs and symptoms develop within a few hours and continue for about 6 hours. They include diffuse erythema, flushing, itchiness, and urticaria with nausea, vomiting, abdominal pain, and diarrhea. More severe cases can be manifested with hypotension, wheezing, and bronchospasm. The clinical diagnosis is confirmed by measurement of histamine in the ingested fish.

Treatment is supportive with intravenous fluids for dehydration and hypotension. Both H_1- and H_2-receptor antagonists (e.g., diphenhydramine, 25 to 50 mg orally or intravenously; ranitidine, 150 mg orally or 50 mg intravenously) should be given as required. Epinephrine is rarely required.

Other Marine Invertebrates

The blue-ringed octopus in Australia has saliva that contains tetrodotoxin. Bites cause effects similar to tetrodotoxin poisoning. Cone snails are a rare cause of envenomation, and their stings cause minor pain, numbness, and sometimes partial or complete paralysis. In both cases, treatment is supportive but may require mechanical ventilation.

Grade A References

A1. de Silva HA, Pathmeswaran A, Ranasinha CD, et al. Low-dose adrenaline, promethazine, and hydrocortisone in the prevention of acute adverse reactions to antivenom following snakebite: a randomised, double-blind, placebo-controlled trial. *PLoS Med*. 2011;8:e1000435.

A2. Maduwage K, Isbister GK. Current treatment for venom-induced consumption coagulopathy resulting from snakebite. *PLoS Negl Trop Dis*. 2014;8:e3220.

A3. Isbister GK, Buckley NA, Page CB, et al. A randomized controlled trial of fresh frozen plasma for treating venom-induced consumption coagulopathy in cases of Australian snakebite (ASP-18). *J Thromb Haemost*. 2013;11:1310-1318.

A4. Loten C, Stokes B, Worsley D, et al. A randomised controlled trial of hot water (45°C) immersion versus ice packs for pain relief in bluebottle stings. *Med J Aust*. 2006;184:329-333.

A5. Schnorf H, Taurarii M, Cundy T. Ciguatera fish poisoning: a double-blind randomized trial of mannitol therapy. *Neurology*. 2002;58:873-880.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Kasturiratne A, Wickremasinghe AR, de Silva N, et al. The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths. *PLoS Med*. 2008;5:e218.
2. Isbister GK, Fan HW. Spider bite. *Lancet*. 2011;378:2039-2047.
3. Isbister GK, Kiernan MC. Neurotoxic marine poisoning. *Lancet Neurol*. 2005;4:219-228.
4. Isbister GK, Brown SG, Page CB, et al. Snakebite in Australia: a practical approach to diagnosis and treatment. *Med J Aust*. 2013;199:763-768.
5. van Helden DF, Thomas PA, Dosen PJ, et al. Pharmacological approaches that slow lymphatic flow as a snakebite first aid. *PLoS Negl Trop Dis*. 2014;8:e2722.
6. Venomous snakes and antivenoms search interface. http://apps.who.int/bloodproducts/snakeantivenoms/database/snakeframeset.html. Accessed March 1, 2015.
7. Isbister GK. Procoagulant snake toxins: laboratory studies, diagnosis, and understanding snakebite coagulopathy. *Semin Thromb Hemost*. 2009;35:93-103.
8. Mion G, Larreche S, Benois A, et al. Hemostasis dynamics during coagulopathy resulting from *Echis* envenomation. *Toxicon*. 2013;76:103-109.
9. Lavonas EJ, Khatri V, Daugherty C, et al. Medically significant late bleeding after treated crotaline envenomation: a systematic review. *Ann Emerg Med*. 2014;63:71-78.
10. Lavonas EJ, Ruha AM, Banner W, et al. Unified treatment algorithm for the management of crotaline snakebite in the United States: results of an evidence-informed consensus workshop. *BMC Emerg Med*. 2011;11:2.
11. Boels D, Hamel JF, Bretaudeau Deguigne M, et al. European viper envenomings: assessment of Viperfav and other symptomatic treatments. *Clin Toxicol (Phila)*. 2012;50:189-196.
12. Schmitt C, De Haro L. Clinical marine toxicology: a European perspective for clinical toxicologists and poison centers. *Toxins (Basel)*. 2013;5:1343-1352.
13. Li L, McGee RG, Isbister G, et al. Interventions for the symptoms and signs resulting from jellyfish stings. *Cochrane Database Syst Rev*. 2013;12:CD009688.
14. Clark RF, Girard RH, Rao D, et al. Stingray envenomation: a retrospective review of clinical presentation and treatment in 119 cases. *J Emerg Med*. 2007;33:33-37.
15. Isbister GK, Hooper JN. Clinical effects of stings by sponges of the genus *Tedania* and a review of sponge stings worldwide. *Toxicon*. 2005;46:782-785.
16. Pennotti R, Scallan E, Backer L, et al. Ciguatera and scombroid fish poisoning in the United States. *Foodborne Pathog Dis*. 2013;10:1059-1066.
17. Hungerford JM. Scombroid poisoning: a review. *Toxicon*. 2010;56:231-243.

REVIEW QUESTIONS

1. Antivenom is most correctly described as which one of the following?

A. A physiologic antidote that reverses the effects of venoms
B. A mixture of antibodies that bind the toxins in a specific venom and can be associated with hypersensitivity reactions in humans
C. Whole or fractionated immunoglobulin G (IgG) antibodies that reverse all the effects of envenomation.
D. Specific toxins designed to neutralize the effects of venoms in humans
E. IgG fractions that bind to a range of receptors to prevent the action of a specific venom

Answer: B Antivenom, which consists of a mixture of polyclonal antibodies raised against the toxins in a specific venom, binds to the venom components and thereby neutralizes venom in the circulation and prevents or reverses venom effects. Antivenom can be composed of whole immunoglobulins (IgG) or fractionated IgG, either $F(ab')_2$ or Fab. The administration of antivenom can result in hypersensitivity reactions because they contain foreign proteins.

2. Which of the following is the most appropriate treatment for North American rattlesnake envenomation?

A. Early application of a pressure bandage and immobilization
B. Immediate surgical decompression if the bite site is swollen and edematous
C. Administration of antivenom for systemic and severe local effects and elevation of the bitten limb
D. Universal early administration of antivenom with prophylactic antihistamines
E. Administration of antivenom and fresh-frozen plasma to treat coagulopathy

Answer: C Definitive management of rattlesnake envenomation is antivenom plus elevation. Early surgical intervention has not been shown to be beneficial. Pressure bandaging is contraindicated in viper envenomation.

3. Which is the most correct statement about stingray injuries?

A. Stingray injuries occur most commonly when they are stepped on in shallow water.
B. Acute pulmonary edema is the major systemic effect of stingray envenomation.
C. Antibiotics are never required for the treatment of stingray injuries.
D. Most injuries occur from captive stingrays when tanks are cleaned.
E. Immersion of the sting site in hot water for 20 minutes will provide complete relief from the pain.

Answer: A Stingray injuries occur most commonly when people tread on them in shallow water. Stingray injuries cause severe local pain, bleeding, and a penetrating injury. Systemic effects are rare. Hot water may partially treat the pain, but pain usually recurs when the hot water is removed, unlike the pain from *Physalia* jellyfish stings, which usually resolves completely with hot-water immersion.

113 RHABDOMYOLYSIS

FRANCIS G. O'CONNOR AND PATRICIA A. DEUSTER

DEFINITION

Rhabdomyolysis, an acute, potentially fatal clinical syndrome, reflects the dissolution and disintegration of striated muscle, with the release of muscle cell contents into the systemic circulation.[1] Myoglobinemia and myoglobinuria are common sequelae. Skeletal muscle destruction can cause systemic effects mediated by substances released from affected muscle cells (e.g., myoglobin, calcium, potassium). Prerenal azotemia, complicated by the toxicity of free myoglobin on the renal tubules, may lead to acute kidney injury (Chapter 120), which exacerbates other metabolic abnormalities. At the extreme, arrhythmias, caused by the release of intracellular potassium and organic acids, coupled with hypocalcemia may be fatal.

EPIDEMIOLOGY

In the absence of a widely accepted laboratory definition or of clinical diagnostic criteria, the true incidence of rhabdomyolysis is unclear, but about 26,000 hospitalized cases are seen each year in the United States. The U.S. military also reports about 400 annual cases of exertional rhabdomyolysis.[2] Among patients with rhabdomyolysis, anywhere from 13 to 67% may develop acute kidney injury, accounting for 5 to 10% of all cases of acute kidney failure in the United States.

ETIOLOGY

Rhabdomyolysis is a complex and multifactorial clinical disorder, with multiple potential inherited and acquired causes (Table 113-1). In urban adults, abuse of alcohol and other drugs, muscle compression, and status epilepticus are common causes of rhabdomyolysis. In pediatric patients, the most common cause is trauma, followed by nonketotic hyperosmolar coma, viral myositis, dystonia, and malignant hyperthermia. However, exertional rhabdomyolysis from repetitive exercise is also a concern in young athletes. Importantly, children and adolescents with recurrent rhabdomyolysis are increasingly being recognized as possibly having inherited metabolic disorders.

Drugs and Intoxications

Among intoxications, which are a common cause of rhabdomyolysis, the most frequent illegal drug-associated causes are cocaine and heroin (Chapter 34), with nearly 20% of cocaine overdoses complicated by rhabdomyolysis. Other recreational drugs, such as "bath salts" (of which methylenedioxypyrovalerone is the primary ingredient) and synthetic cannabinoids (or "spice"), have been associated with rhabdomyolysis.[3] Other substances that can induce rhabdomyolysis include ethanol (Chapter 33), amphetamines, phenylalkylamine derivatives, caffeine, and statins. Statins (Chapter 206) may result in myalgias in up to 10% of patients receiving treatment, but reported rates of statin-induced rhabdomyolysis range from 0 to 2.2 cases per 1000 person-years, with cerivastatin being associated with the highest rates.[4] Dietary supplements, in particular those with combinations of stimulants, have also been associated with rhabdomyolysis and other complications.

Exertional Rhabdomyolysis

Rhabdomyolysis can also be a consequence of excessive exertion,[5] prolonged heat exposure (Chapter 109), coexisting sickle cell trait (Chapter 163), and the use of dietary supplements (e.g., ephedra). In one series, 35 of 225,000 emergency department visits to an urban tertiary care center were for exertional rhabdomyolysis. The average creatine kinase (CK) level was 40,000 U/L, but no patient developed acute kidney injury. In another series, 57% of participants in an ultramarathon had evidence of myoglobinemia, but none progressed to acute kidney injury. Exertional rhabdomyolysis has also been diagnosed in other sports settings (e.g., baseball, football, track, wrestling), but none with a higher frequency than seen with endurance events.

In the military, acute exertional rhabdomyolysis occurs in 2 to 40% of individuals undergoing basic training, usually within the first 6 days. Resolution of myoglobinuria typically occurs after 2 or 3 days, with clinical improvement within 1 week. Consistent risk factors for exertional rhabdomyolysis are low levels of physical fitness and early introduction of repetitive exercises (e.g., squats, push-ups, sit-ups). Although most cases are self-limited, with no long-term evidence of kidney or muscle injury, patients who demonstrate systemic signs, generalized clinical findings, or acute kidney injury often have an underlying metabolic myopathy. Importantly, 25% of all heatstroke cases in the military between 1980 and 2000 were associated with rhabdomyolysis; acute kidney injury developed in 33%. A retrospective review of deaths in a military basic trainee population found an increased risk for nontraumatic, exertional sudden death in African Americans with sickle cell trait; several deaths were associated with fulminant exertional rhabdomyolysis.

The extreme exertion characteristics of military service carry over to the population of correctional inmates and civil servant first responders. Unsupervised repetitive exercise in prison populations can lead to exertional rhabdomyolysis. Among New York City firefighters, 32 of 16,506 candidates (0.2%) were hospitalized for exertional rhabdomyolysis after a physical fitness test, with four requiring hemodialysis. In a group of 50 prospective police officers from Massachusetts, 13 trainees were hospitalized with exertional rhabdomyolysis and had CK levels higher than 32,000 U/L; six required dialysis, and one died 44 days later as a result of complications of heatstroke, exertional rhabdomyolysis, and kidney and hepatic failure.

PATHOBIOLOGY

Pathophysiology

The final common pathway for all cases of rhabdomyolysis is from direct or indirect injury or destruction of muscle cells, with displacement of their intracellular contents into extracellular fluid, the circulation, or both (E-Fig. 113-1). Cell function is critically dependent on the relationship

TABLE 113-1 INHERITED AND ACQUIRED CAUSES OF RHABDOMYOLYSIS

INHERITED	ACQUIRED
Glycolytic/glycogenolytic, e.g., McArdle disease (myophosphorylase deficiency)	Exertion, e.g., exercise, status epilepticus, delirium, electrical shock, status asthmaticus, cardiopulmonary resuscitation (see also Table 113-2)
Fatty acid oxidation, e.g., carnitine palmitoyltransferase II deficiency	Crush, e.g., external weight, prolonged immobility, bariatric surgery
Krebs cycle, e.g., aconitase deficiency	Ischemia, e.g., arterial occlusion, compartment syndrome, sickle cell disease, disseminated intravascular coagulation
Pentose phosphate pathway, e.g., glucose-6-phosphate dehydrogenase deficiency	Extremes of body temperature, e.g., fever, exertional heatstroke, burns, malignant hyperthermia, hypothermia, lightning
Purine nucleotide cycle, e.g., myoadenylate deaminase deficiency	Metabolic, e.g., hypokalemia, hypernatremia or hyponatremia, hypophosphatemia, pancreatitis, diabetic ketoacidosis, renal tubular acidosis, hyperthyroidism or hypothyroidism, nonketotic hyperosmolar states
Mitochondrial respiratory chain, e.g., succinate dehydrogenase deficiency	Drugs or toxins, e.g., anticholinergics, amphetamines, antihistamines, arsenic, ethanol, opiates, statins, cocaine, succinylcholine, halothane, corticosteroids, cyclosporine, itraconazole, phenothiazines, bath salts, synthetic cannabinoids
Malignant hyperthermia susceptibility, e.g., familial malignant hyperthermia (*RYR1*) mutations, myotonic dystrophy, Duchenne and Becker dystrophies	Infections, e.g., Epstein-Barr virus, human immunodeficiency virus, herpes simplex, influenza A and B, *Borrelia burgdorferi*, tetanus
Other, e.g., familial recurrent myoglobinuria	Inflammatory and autoimmune disorders, e.g., polymyositis, dermatomyositis

Modified and reproduced with permission from Warren JD, Blumbergs PC, Thompson PD. Rhabdomyolysis: a review. *Muscle Nerve.* 2002;25:332-347.

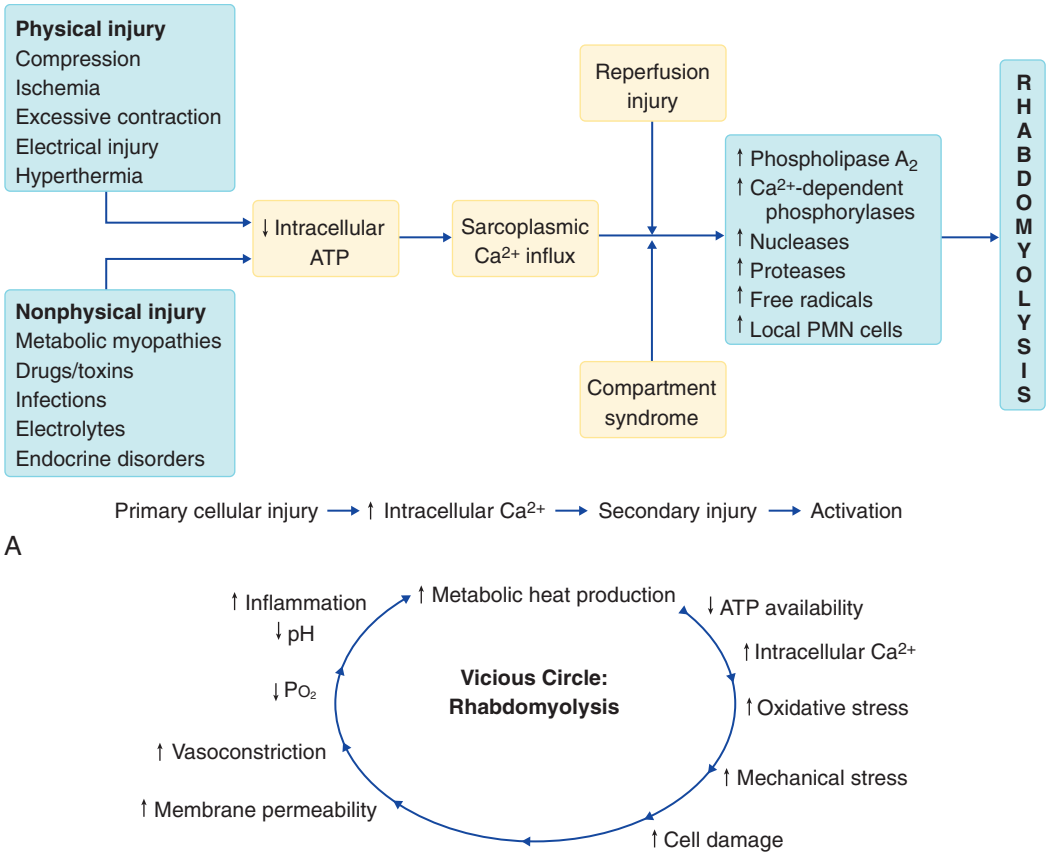

A

B

E-FIGURE 113-1. **A, Pathogenesis of rhabdomyolysis. B, Vicious circle of rhabdomyolysis. ATP = adenosine triphosphate; PMN = polymorphonuclear.**

between intracellular calcium (Ca^{2+}) and sodium (Na^{+}) concentrations. Sarcolemmal Na^{+},K^{+}-ATPase regulates extracellular Ca^{2+} concentrations by exchanging Na^{+} for Ca^{2+} across the sarcolemma. A low intracellular Na^{+} concentration creates a gradient that actively results in efflux of Ca^{2+} as it is exchanged for Na^{+} ions. This process maintains intracellular Ca^{2+} levels at several orders of magnitude lower than extracellular Ca^{2+}.

When the cell is subjected to mechanical stress, stretch-activated channels in the sarcolemma can open and cause an influx of Na^{+} and Ca^{2+}. With excessive intracellular Ca^{2+}, several pathologic processes begin: persistent contraction of myofibers, depletion of adenosine triphosphate (ATP), production of free radicals, activation of vasoactive molecules, release of proteases, and, ultimately, cell death. Cell death is followed by an invasion of neutrophils, which amplify the damage by further release of proteases and increased production of free radicals. Rather than simple necrosis, a self-sustaining, inflammatory myolytic reaction develops.

Rhabdomyolysis can be further complicated by reperfusion injury and compartment syndrome. In reperfusion injury, restoration of vascular flow after a prolonged period of ischemia results in the delivery of activated neutrophils in combination with an abundance of oxygen, which contributes to the development of highly reactive free radicals. Because most muscle groups are contained within rigid fascial compartments, rhabdomyolysis can quickly precipitate a secondary acute compartment syndrome. The swelling associated with traumatized tissue can also increase intracompartmental pressure, which can provoke additional damage by compromising both venous and arterial blood flow. Thus, compartment syndrome can also result in rhabdomyolysis.

Inherited and Acquired Rhabdomyolysis

Rhabdomyolysis can be classified as inherited or acquired (see Table 113-1). Patients who have recurrent episodes of high levels of CK in the blood triggered by low levels of stress or exertion should be evaluated for a metabolic myopathy. A number of pathways leading to the formation of ATP can be disrupted by genetic defects (e.g., inherited disorders of glycogenolysis, glycolysis, and lipid and purine metabolism). In one series of 77 patients who underwent biopsy for idiopathic myoglobinuria, 47% were found to have enzymatic defects; the most common disorders were deficiencies of carnitine palmitoyltransferase II and myophosphorylase. Recent work suggests that carnitine palmitoyltransferase, acid maltase, and lipin deficiencies are common causes of rhabdomyolysis.

Inherited Rhabdomyolysis

Malignant hyperthermia (Chapters 432 and 434) is a potentially fatal, heterogeneous, pharmacogenetic disorder triggered by volatile anesthetics in predisposed individuals. The disorder is most commonly inherited in an autosomal dominant pattern. Evidence from molecular studies indicates that 25% of patients who are susceptible to malignant hyperthermia have mutations in the ryanodine receptor (*RYR1*) gene, which encodes the protein for one of the primary Ca^{2+} release channels involved in triggering muscle contraction. When a susceptible patient is exposed to a triggering anesthetic agent, excessive release of Ca^{2+} into the myoplasm leads to a hypermetabolic state manifested by hypercapnia, tachycardia, and metabolic acidosis. Although a genetic predisposition to malignant hyperthermia may be linked to a predisposition to exertional rhabdomyolysis and exercise-induced heat injury, this association has not been proved.

Acquired Rhabdomyolysis

Drugs and Toxins

Drugs and toxins, also common causes of rhabdomyolysis, operate through a number of mechanisms, including direct membrane toxicity (e.g., herbicides), indirect metabolic derangements (e.g., anticholinergics), ischemia (e.g., cocaine), and agitation (e.g., hemlock).[6] The most commonly cited drugs precipitating rhabdomyolysis are alcohol, statins, cocaine, amphetamines, and phenothiazines, although new designer drugs (spice, bath salts) are also implicated. Alcohol can induce rhabdomyolysis through a combination of mechanisms, including immobilization, direct myotoxicity, and electrolyte abnormalities. Statins, which inhibit 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase, can be directly myotoxic and appear to trigger sustained increases in intracellular Ca^{2+}. Although the precise mechanisms underlying statin-induced myopathy are currently incompletely understood, it can be aggravated by the concomitant administration of cytochrome P-450 3A4 inhibitors (e.g., itraconazole, erythromycin, cyclosporine, danazol) and fibrates as well as by physical exercise, excessive alcohol intake, and preexisting comorbid medical conditions. Amphetamines and phenothiazines may lead to a clinical picture of rhabdomyolysis through the serotonin syndrome (Chapter 434) and the neuroleptic malignant syndrome (Chapter 418), respectively. The mechanism whereby bath salts lead to rhabdomyolysis may be direct muscle toxicity, severe hyperthermia, or electrolyte disorders.

Infections

Both viral and bacterial infections can trigger rhabdomyolysis. Either cellular invasion or generation of various toxins may precipitate infection-induced rhabdomyolysis. Influenza A and B (Chapter 364) are the most common viral causes, followed by human immunodeficiency virus (Chapter 386), coxsackievirus (Chapter 379), and Epstein-Barr virus (Chapter 377). The most common bacterial organisms that induce rhabdomyolysis are *Legionella* species (Chapter 314), followed by *Francisella tularensis* (Chapter 311) and *Streptococcus pneumoniae* (Chapter 289). Acute kidney injury develops in approximately 57% (33 to 100%) and 34% (0 to 100%) of bacterial and viral cases of rhabdomyolysis, respectively.

Trauma

Trauma is the most common cause of rhabdomyolysis. Wars, natural disasters, and traffic and occupational accidents are frequent causes of trauma-induced "crush injury syndrome" (Chapter 111). Other less frequent causes of trauma- or compression-induced rhabdomyolysis include struggling against restraints, direct blows, child abuse, torture, prolonged immobilization (e.g., anesthesia, coma, drug- or alcohol-induced stupor), and bariatric and other forms of surgery. The primary mechanism of crush syndrome and compression-induced rhabdomyolysis is reperfusion of damaged tissue after a period of ischemia.

Exertional rhabdomyolysis can result from excessive exercise in fit and unfit individuals, particularly eccentrically based activities (lengthening contractions, such as lowering a weight), but it can also be triggered by exertion in combination with thermal stress, sickle cell trait, or altitude or by the use of medications (e.g., anticholinergics) or dietary supplements (e.g., caffeine, ephedra). The spectrum of exertional rhabdomyolysis is broad and can range from a subclinical event to catastrophic collapse and death; underlying mechanisms may be either mechanical or metabolic in nature, but all are associated with elevated myoplasmic Ca^{2+} concentrations.

A number of underlying genetic polymorphisms and inherited disorders are associated with exertional rhabdomyolysis (Table 113-2).[7] Multiple mutations in the carnitine palmitoyltransferase II and myophosphorylase genes have been found, and although each mutation has been associated with exercise-induced myoglobinuria, the mutations alone may not explain the clinical episodes. Mutations and variants in the ryanodine receptor 1 (*RYR1*) gene, which are common in malignant hyperthermia (Chapter 432), have also been found in persons with exertional rhabdomyolysis. Other single-nucleotide polymorphisms associated with severe exertional rhabdomyolysis are found in the genes encoding CK muscle isoform (*CKMM* NcoI), α-actinin-3 (*ACTN3* R577X), or myosin light chain kinase (*MYLK* C37885A). However, data are insufficient to use these or other variants to predict an individual's clinical susceptibility to infection-, toxin-, exertion-, or drug-induced rhabdomyolysis.

TABLE 113-2 GENETIC MUTATIONS/VARIANTS ASSOCIATED WITH EXERTIONAL RHABDOMYOLYSIS

GENE	
Ryanodine receptor 1	*RyR1*
Myoadenylate deaminase	*AMPDA1*
Carnitine palmitoyltransferase II	*CPT2*
Myophosphorylase	*PYGM*
Phosphofructokinase	*PFKM*
Phosphorylase *b* kinase	*PHKA1*
Very long chain acyl–coenzyme A dehydrogenase	*ACAD9*
Phosphoglycerate mutase	*PGAMM*
Phosphoglycerate kinase	*PGK1*
Lactate dehydrogenase	*LDHA*
Cytochrome *c* oxidase	*COX I, II,* and *III*
Cytochrome *b* (complex III)	*CYTB*
Mitochondrial tRNA	*Mt-tRNA*
β-Sarcoglycan	*SGCB*
Mitochondrial DNA	*MT-CO2*

CLINICAL MANIFESTATIONS

The classic manifestations of rhabdomyolysis include acute myalgia and pigmenturia as a result of myoglobinuria in association with elevated serum muscle enzymes (CK in particular). Many clinical features are nonspecific, however, and the course and initial signs, symptoms, and laboratory abnormalities are clearly dependent on the underlying cause and severity of the event.

Rhabdomyolysis can be accompanied by both local and systemic features. Local features, generally noted in the area of the traumatized muscle groups, can occur within hours of the trauma and include muscle pain, tenderness, and swelling. Systemic features include tea-colored urine, chills, fever, and malaise. In extreme cases, patients complain of nausea and vomiting and demonstrate confusion, agitation, or delirium. Whenever systemic features such as chills, fever, malaise, or generalized muscle involvement are observed, an underlying metabolic myopathy should be considered.

Clinical findings may also include compartment syndrome, which can occur in muscle groups encased by fascia, especially the lower leg, forearm, and thigh muscle groups. Sensory abnormalities caused by nerve compression are an early manifestation of compartment syndrome; the loss of a pulse as a result of vascular compromise is a later finding. If compartment syndrome is not addressed within 6 to 8 hours, irreversible ischemic muscle and nerve damage may occur.

Laboratory findings are related to the degree of muscle involvement. Early findings include elevated blood levels of CK, myoglobin, potassium, urea, and phosphorus. CK levels typically peak 2 to 5 days after the initial insult; levels higher than 15,000 U/L are more likely to be associated with acute kidney injury than are lower levels.[8] Hypocalcemia, caused by the influx and deposition of Ca^{2+} in damaged muscle tissue, may accompany rhabdomyolysis. Moreover, an anion gap metabolic acidosis may develop because of the release of organic acids from damaged muscle. With resolution of rhabdomyolysis, sequestered Ca^{2+} may be released back into the circulation and cause hypercalcemia.

DIAGNOSIS

Creatine Kinase Levels

A diagnosis of rhabdomyolysis is made when there is clinical evidence of myonecrosis and release into the systemic circulation of muscle cell contents, including myoglobin, creatinine, CK, organic acids, potassium, aldolase, lactate dehydrogenase, and hydroxybutyrate dehydrogenase. The skeletal muscle subtype CK-MM is abundantly present in skeletal muscle and released as a result of muscle destruction. Serum levels exceeding 100,000 U/L are not uncommon with rhabdomyolysis. Because CK remains in the circulation longer than myoglobin and can be detected easily and efficiently, it is the most frequently used marker to diagnose rhabdomyolysis. No universally accepted clinical or laboratory definition of rhabdomyolysis currently exists, but CK elevations ranging from more than five times to more than 50 times the upper limits of normal, as well as varying requirements for serum creatinine elevation, have been proposed. Importantly, sex, ethnicity, and baseline physical activity levels all affect individual baseline CK levels. For example, African American males and young athletic men have the highest baseline CK levels, whereas non–African American women have the lowest. Thus, modifying factors such as sex, ethnicity, and physical fitness must be considered when CK is used for the diagnosis of rhabdomyolysis. In general, CK levels in excess of at least five times normal, in combination with the appropriate clinical presentation, are accepted as evidence of muscle breakdown, which may be consistent with a diagnosis of rhabdomyolysis.

Myoglobin Testing

Because myoglobinuria does not occur in the absence of rhabdomyolysis, myoglobin should be the most specific marker of rhabdomyolysis. However, testing for serum or urine myoglobin is problematic and not always consistent. Because myoglobin is normally bound to plasma globulins, only a small fraction of the myoglobin that is released into the circulation reaches the glomeruli. In the presence of severe muscle damage, blood levels of myoglobin overwhelm the binding capacity of the circulating proteins, so free myoglobin reaches the glomeruli and eventually the renal tubules. Elevations in serum myoglobin occur before the rise in serum CK, but the elimination kinetics of serum myoglobin is more rapid than that of CK, which makes the often evanescent rise in serum myoglobin a less reliable marker of muscle injury. Furthermore, the liver can quickly metabolize myoglobin.

Because most laboratories will perform urine myoglobin testing no more often than once per day, urine myoglobin is neither a timely nor accurate predictor of acute kidney injury.[9] Nevertheless, urine screening for rhabdomyolysis may be performed by dipstick. The orthotoluidine portion of urine dipsticks turns blue in the presence of hemoglobin or myoglobin, so if the urine sediment does not contain erythrocytes, one can assume, in the appropriate clinical setting, that a positive dipstick reading reflects the presence of myoglobin.

Other Laboratory Tests

Other associated laboratory findings in acute rhabdomyolysis can include hypocalcemia or hypercalcemia, hyperphosphatemia, metabolic (lactic) acidosis, thrombocytopenia, and disseminated intravascular coagulation. Muscle biopsy is not required to make a diagnosis of rhabdomyolysis, but it can be confirmatory, especially in cases of recurrent rhabdomyolysis or when the diagnosis is not clear. Histopathologic evaluation usually demonstrates muscle necrosis, loss of the cell nucleus, and muscle stria with the absence of inflammatory cells.

Differential Diagnosis

The clinical findings of acutely swollen muscles or muscle weakness (or both) with reddish brown urine are not always the result of rhabdomyolysis, and the examining clinician must be careful to scrutinize all information. The differential diagnosis includes disorders that may indirectly affect myocytes, such as Guillain-Barré syndrome and periodic paralysis. Guillain-Barré syndrome (Chapter 420) differs from rhabdomyolysis in that it is characterized as a fulminant polyneuropathy, usually after an antecedent viral infection. Periodic paralysis (Chapter 421) is frequently associated with transient electrolyte disturbances and is distinguished from rhabdomyolysis in that most cases follow periods of rest or sleep.

Myoglobinuria causes the urine to be reddish brown, but tea-colored (or cola-colored) urine does not necessarily indicate the presence of myoglobin. Other conditions associated with discoloration of urine include hemoglobinuria from hemolysis, intrinsic renal disease, porphyria, acute glomerulonephritis, "athletic pseudonephritis," and external factors such as ingestion of beets and various drugs (e.g., phenytoin, rifampin, riboflavin, or vitamin B_2).

The diagnosis of rhabdomyolysis is complete when the clinician determines the cause. This step, although it is frequently established during the history and physical examination, may require further diagnostic assessment after initiation of clinical treatment during the acute phase. Individuals with recurrent rhabdomyolysis, a positive family history of rhabdomyolysis or malignant hyperthermia, low exercise tolerance, no apparent cause, or a fulminant or explosive form of rhabdomyolysis appear to warrant further testing.

Testing may include a nonischemic forearm test, which involves isometric exercise at 70% of maximal voluntary contraction for 30 seconds under nonischemic conditions; electromyography; more in-depth blood tests for muscle enzymes (e.g., mitochondrial myopathies [Chapter 421], fatty acid transport defects [Chapter 421], glycogen storage diseases [Chapter 207], diseases associated with myoglobinuria); muscle biopsy to investigate specific metabolic myopathies and other enzyme or genetic defects; or any combination of such testing. The forearm exercise test may help identify metabolic and genetic causes of rhabdomyolysis. Patients who have had an episode of malignant hyperthermia or exertional heat illness may be candidates for a caffeine halothane contracture test, which evaluates the force produced by biopsied muscle samples after separate exposures to caffeine, halothane, and caffeine/halothane in the laboratory. Isolated, perfused muscle fibers must show an increase in tension of at least 0.2 g when exposed to 2 mM of caffeine or at least 0.7 g of tension after exposure to 3% halothane. In addition, genetic investigation for mutations of the *RYR1* receptor gene may be warranted.

PREVENTION

Approaches for preventing rhabdomyolysis induced by infections, medications, toxins, heat stress, or exercise may emerge in the future, but no definitive guidelines currently can be presented. To prevent further muscle injury, blood flow to ischemic areas must be promptly restored to minimize ischemia-reperfusion damage.

TREATMENT

Rx

Treatment of rhabdomyolysis begins with a careful history and physical examination to identify and to manage any underlying illness. Vital signs, urine output, and serial electrolyte and CK levels should be obtained as soon as possible. Patients require aggressive early management to preserve renal function (Table 113-3). Careful observation and treatment of potential early and late complications are critical, and intensive care monitoring may be required.

TABLE 113-3 STEPS IN THE PREVENTION AND TREATMENT OF RHABDOMYOLYSIS-INDUCED ACUTE KIDNEY INJURY

Check for extracellular volume status, central venous pressure, and urine output.*
Measure serum creatine kinase levels. Measurement of other muscle enzymes (myoglobin, aldolase, lactate dehydrogenase, alanine aminotransferase, aspartate aminotransferase) adds little information relevant to diagnosis or management.
Measure levels of plasma and urine creatinine, potassium, and sodium; blood urea nitrogen; total and ionized calcium, magnesium, and phosphorus; and uric acid and albumin. Evaluate acid-base status, blood cell count, and coagulation.
Perform a urine dipstick test and examine the urine sediment.
Initiate volume repletion with normal saline promptly at a rate of approximately 400 mL/hr (200-1000 mL/hr, depending on the patient size, setting, and severity) and monitor the clinical course or central venous pressure.
Target urine output of approximately 3 mL/kg body weight/hr (200-300 mL/hr).
Check serum potassium levels frequently.
Correct hypocalcemia only if symptomatic (e.g., tetany, seizures) or severe hyperkalemia occurs.
Investigate the cause of rhabdomyolysis.
Consider treatment with bicarbonate. Check urine pH: if it is <6.5, alternate each liter of normal saline with 1 L 5% dextrose or 0.45% saline plus 100 mmol bicarbonate. Avoid potassium- and lactate-containing solutions.
Consider treatment with mannitol (up to 200 g/day; cumulative dose up to 800 g). Check for plasma osmolality and calculate the plasma osmolal gap (Chapter 120). Discontinue if diuresis (>20 mL/hr) is not established.
Maintain volume repletion until myoglobinuria is cleared (as evidenced by clear urine or urine dipstick test that is negative for blood).
Consider kidney replacement therapy with resistant hyperkalemia (>6.5 mmol/L) that is symptomatic (as assessed by electrocardiography), rapidly rising serum potassium levels, oliguria (<0.5 mL of urine/kg/hr for 12 hours), anuria, volume overload, or resistant metabolic acidosis (pH < 7.1).

*In the case of crush syndrome (e.g., earthquake, building collapse, bariatric surgery), institute aggressive volume repletion promptly before evacuating the patient if creatine kinase level is >15,000 U/L.

Modified from Bosch X, Poch E, Grau JM. Rhabdomyolysis and acute kidney injury. *N Engl J Med.* 2009;361:62-72.

Hydration

Hydration is the cornerstone of preserving kidney function in patients with rhabdomyolysis, and delay of fluid administration for more than 6 hours increases the risk for acute kidney injury. Inpatient hydration is indicated for victims of collapse, trauma, or exertional heat injury as well as for patients who have moderate early symptoms, more than mild elevations in CK, or abnormal serum levels of creatinine, potassium, calcium, phosphate, or bicarbonate. In adults, the target urine output is 300 mL/hr for at least 24 hours to prevent acute kidney injury. Hydration is accomplished by the aggressive administration of isotonic intravenous fluids at a rate that results in a urine output of 200 to 300 mL/hour until CK levels begin to decline. If fluid resuscitation fails to correct intractable hyperkalemia and acidosis, renal replacement therapy should be considered (Chapter 120). By comparison, adults with mild symptoms and serum CK levels less than 3000 U/L are considered to be at low risk and may be treated as outpatients with vigorous oral hydration, limited physical activity, and careful follow-up.

Specific Therapeutic Measures

Alkalinization of the urine decreases cast formation, minimizes the toxic effects of myoglobin on the renal tubules, inhibits lipid peroxidation, and decreases the risk for hyperkalemia. However, bicarbonate therapy can cause calcium to precipitate in the soft tissues and contribute to a hyperosmolar state. Mannitol is an osmotic diuretic, volume expander, and free radical scavenger, but it should be used only after adequate kidney function is established and must be used with great caution in patients with marginal cardiac function. To date, no convincing evidence demonstrates that adding sodium bicarbonate or mannitol is superior to fluid therapy alone.[8] Sodium bicarbonate should be used only in patients with evidence of systemic acidosis, and mannitol should be used only when it is needed to maintain a urine output of 300 mL/hour.

Deposition of calcium, which occurs early in rhabdomyolysis, is directly related to the degree of muscle destruction and to the administration of calcium. Furthermore, reversal of hypocalcemia early in the patient's course may worsen ectopic calcification and exacerbate hypercalcemia during the resolution phase.[10] Accordingly, hypocalcemia should be treated only when patients develop clinical symptoms, signs of tetany, or severe hyperkalemia.

Management of Compartment Syndrome

Compartment syndrome (Chapter 111) is a well-described late complication as well as a potential cause of rhabdomyolysis. Compartment syndrome can occur as a direct consequence of muscle injury associated with increased vascular permeability, aggressive fluid resuscitation, or restoration of reperfusion. Compartment syndrome should be suspected when the muscles are tense and swollen, previously declining CK levels start to rise, or neurovascular compromise occurs. In these cases, compartment pressures should be measured; if pressures exceed 30 mm Hg, prompt fasciotomy should be considered. However, because late fasciotomy (>12 hours after the onset of symptoms) can convert a closed injury into an open wound and thereby increase the risk of uncontrollable infection, late fasciotomy is relatively contraindicated.

Management of Crush Injury

For victims of crush injury (Chapter 111), aggressive on-site hydration with intravenous normal saline is recommended. For massive damage, amputation of the extremity may be required to protect the patient's overall health. The Mangled Extremity Severity Score can identify nonsalvageable extremities on the basis of the degree of skeletal and soft tissue injury, the patient's blood pressure, the presence of a detectable pulse, and age (see http://www.mdcalc.com/mangled-extremity-severity-score-mess-score/).

Malignant Hyperthermia

Rhabdomyolysis caused by malignant hyperthermia (Chapter 432) requires rapid diagnosis and aggressive management. Anesthetics should be discontinued, and the patient should be treated with dantrolene sodium, 2.5 to 4 mg/kg intravenously, followed by about 1 mg/kg every 4 hours for up to 48 hours to avoid recrudescence.

PROGNOSIS

The most serious consequence of rhabdomyolysis is acute kidney injury, which occurs in up to 67% of all cases, regardless of cause. Predictors of the risk of needing renal replacement therapy or death in patients with rhabdomyolysis include age older than 50 years; initial serum creatinine level of 1.4 mg/dL or higher; initial serum calcium level below 7.5 mg/dL; initial serum phosphate level above 4.0 mg/dL; initial serum bicarbonate level below 19 mEq/L; and a cause other than syncope, seizures, exercise, statins, or myositis.[11] Event rates range from 0% in patients with none of these criteria to 20% or more in patients with four or more criteria. For compartment syndrome, a poor prognosis is associated with an ischemic period lasting longer than 6 hours.

The prognosis of patients with rhabdomyolysis improves markedly when treatment is started soon after the diagnosis is made. With mild episodes, the prognosis is customarily excellent, and the patient can typically resume usual activities within several weeks after CK levels have normalized. However, some patients do not return to normal and continue to experience extreme fatigue and muscle pain on exertion. These patients require additional testing (nonischemic forearm test, electromyography, muscle disease enzyme panel, muscle biopsy; Chapter 421) to determine whether an underlying metabolic myopathy exists. The results of these tests will help determine future recommendations, but the patient's tolerance and response to light and more strenuous exercise are important factors. Most authorities agree that statin-induced rhabdomyolysis is an indication for discontinuation of their use.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Zimmerman JL, Shen MC. Rhabdomyolysis. *Chest*. 2013;144:1058-1065.
2. Update: Exertional rhabdomyolysis, active component, U.S. Armed Forces 2008-2012. *MSMR*. 2013;120:21-24.
3. Ross EA, Reisfield GM, Watson MC, et al. Psychoactive "bath salts" intoxication with methylenedioxypyrovalerone. *Am J Med*. 2012;125:854-858.
4. Auer J, Sinzinger H, Franklin B, et al. Muscle- and skeletal-related side-effects of statins: tip of the iceberg? *Eur J Prev Cardiol*. 2014. [Epub ahead of print]
5. Szczepanik ME, Heled Y, Capacchione J, et al. Exertional rhabdomyolysis: identification and evaluation of the athlete at risk for recurrence. *Curr Sports Med Rep*. 2014;13:113-119.
6. Pasnoor M, Barohn RJ, Dimachkie MM. Toxic myopathies. *Neurol Clin*. 2014;32:647-670.
7. Deuster PA, Contreras-Sesvold CL, O'Connor FG, et al. Genetic polymorphisms associated with exertional rhabdomyolysis. *Eur J Appl Physiol*. 2013;113:1997-2004.
8. Scharman EJ, Troutman WG. Prevention of kidney injury following rhabdomyolysis: a systematic review. *Ann Pharmacother*. 2013;47:90-105.
9. Bosch X, Poch E, Grau JM. Rhabdomyolysis and acute kidney injury. *N Engl J Med*. 2009;361:62-72.
10. Graziani G, Calvetta A, Cucchiari D, et al. Life-threatening hypercalcemia in patients with rhabdomyolysis-induced oliguric acute renal failure. *J Nephrol*. 2011;24:128-131.
11. McMahon GM, Zeng X, Waikar SS. A risk prediction score for kidney failure or mortality in rhabdomyolysis. *JAMA Intern Med*. 2013;173:1821-1827.

REVIEW QUESTIONS

1. During the management of a patient who has multiple traumatic injuries, you note that a previously declining serum creatine kinase level now happens to rise. The patient's urine output is stable on aggressive hydration, and vital signs are normal. Which of the following represents the most likely clinical scenario accounting for the increasing creatine kinase?

A. Infection
B. Compartment syndrome
C. Medication toxicity
D. Underlying metabolic myopathy
E. Acute respiratory failure

Answer: B In the management of the trauma patient with rhabdomyolysis, serum creatine kinase levels should steadily decline after a peak has been achieved, which typically is 3 to 5 days into effective treatment with aggressive fluid hydration. When the serum creatine kinase level increases, the clinician should suspect a compartment syndrome. Excessive pain is typically a clinical clue, but the use of regional anesthetic blocks or strong analgesics may mask the symptoms.

2. Rhabdomyolysis can be a complication of crush injury in people trapped at the site of a collapsed building. Which of the following field site interventions is most prudent to preserve both life and limb?

A. Extremity hypothermia protocol
B. Hypotensive resuscitation
C. Aggressive fluid hydration
D. Low-molecular-weight dextran to increase viscosity
E. Extremity hyperthermia protocol

Answer: C Crush injury can cause both early and delayed deaths. The principal culprit for delayed morbidity and mortality is acute renal failure with associated metabolic complications. Early intravenous hydration, preferably within the first 6 hours, should aim to achieve a urine output in adults of 300 mL/hr to prevent acute renal failure.

3. Rhabdomyolysis can be a potentially life-threatening illness, with acute, subacute, and chronic complications, including compartment syndrome and acute kidney injury. Which of the following diagnostic tests is not necessary in the acute management of a significant case of rhabdomyolysis?

A. Electrocardiogram
B. Muscle biopsy
C. Metabolic panel
D. Urinalysis
E. Measurement of creatine kinase levels

Answer: B In the management of rhabdomyolysis, the immediate focus is on treating metabolic complications and preventing kidney failure. Accordingly, a metabolic panel, measurement of creatine kinase levels, electrocardiogram, and urinalysis are all valuable assessments in detecting abnormalities and assisting in ongoing management. A muscle biopsy is not indicated in the acute management of the patient with rhabdomyolysis, but it can assist in determining whether a patient with recurrent rhabdomyolysis has an underlying inherited and acquired myopathy.

4. A serum creatine kinase level is commonly obtained to aid in both diagnosis and management of the patient with rhabdomyolysis. Which of the following characteristics are important to understand the baseline creatine kinase level in an individual patient?

A. Sex
B. Race
C. Activity level
D. A and B
E. A, B, and C

Answer: E The difficulty in using serum creatine kinase levels alone to make the diagnosis of mild rhabdomyolysis partly lies in the wide variability of baseline levels. Men, African Americans, and more active people have higher baseline serum creatine kinase levels.

114 APPROACH TO THE PATIENT WITH RENAL DISEASE

DONALD W. LANDRY AND HASAN BAZARI

The prominent functions of the kidney include the excretion of nitrogenous waste; the regulated excretion of water, sodium, potassium, and acid; and the synthesis of a variety of hormones, including 1,25-dihydroxyvitamin D, erythropoietin, and renin. The kidney's elaboration of a protein-free and cell-free ultrafiltrate is uniquely responsible for the excretion of nitrogenous wastes. The approach to the patient with renal disease is largely focused on disordered ultrafiltration and not on defects in the isolated renal tubular processing of individual ions, water, or acids. A patient may, for example, present with an isolated defect in renal acid excretion (Chapter 118), but in this case the "approach to the patient" is framed for the evaluation of metabolic acidosis, an abnormality for which the kidney is only one among the many causes in a broad differential diagnosis. In contrast, acute kidney injury (Chapter 120) and chronic kidney disease (Chapter 130) refer specifically and exclusively to defects in the filtration function of the kidney. In the context of a diminished magnitude of filtration, many of the other individual functions of the kidney (e.g., hormone synthesis, electrolyte homeostasis) may fail as well.

Primary diseases of the tubules, such as acute tubular necrosis (Chapter 120) and tubulointerstitial disease (Chapter 122), also impair the rate of glomerular filtration and cause acute kidney injury and chronic kidney disease. In contrast, an impairment in ultrafiltration may, in early chronic kidney disease, be reflected solely in a decreased *quality* of glomerular filtration (e.g., the presence of albuminuria) rather than in a decreased *quantity* of filtrate with increased concentrations of nitrogenous waste. Similarly, the glomerular filtration rate (GFR) may be normal in nephrotic syndrome despite ultrafiltration defects that result in massive proteinuria. Defects in the filter can also allow passage of cells, such as red blood cells (RBCs), as is seen in the acute nephritic syndrome (Chapter 121), with or without heavy proteinuria. The paradox of glomerular hematuria without albuminuria is also possible. For example, in mild forms of immunoglobulin A (IgA) nephropathy (Chapter 121), relatively few defects in the glomerular filter will permit a detectable number of RBCs per high-power field in the urine despite a urine albumin level that still remains within normal limits (Fig. 114-1).

In this context, this chapter considers the approach to the patient with acute kidney injury, glomerular syndromes (nephrotic vs. nephritic), tubulointerstitial disease, vasculitis and vascular diseases of the kidney, papillary necrosis, and chronic kidney disease.

PATHOBIOLOGY

The approximately 2 million renal glomeruli normally filter about 180 L/day. The renal glomerulus is not simply a filter but rather a size- and charge-dependent ultrafilter that excludes not only cells but also proteins larger than 60 kD from the ultrafiltrate. Smaller proteins are variably filtered at the glomerulus and endocytosed in the proximal tubule so that the protein concentration of the urine is normally low. Kidney disease reflects a failure in the quantity or quality of the glomerular ultrafiltrate.

The normal GFR may decline in hours to days in acute kidney injury or during months to years in chronic kidney disease. An acute decline in glomerular filtration is the necessary and sufficient condition for the diagnosis of acute kidney injury, but abnormal urinary findings can assist with elucidating the etiology of the injury. Proteinuria, ranging from microscopic to nephrotic range (Chapter 121), and urinary findings, from a few cells per microscopic high-power field to gross hematuria or pyuria, may be the only evidence of the earliest stages of chronic kidney disease. As chronic kidney disease advances, the decline in the GFR progresses until dialysis or transplantation (Chapter 131) is required to forestall or to treat the syndrome of uremia.

DIAGNOSIS

Measuring Kidney Function

Although the most accurate method of evaluating kidney function is a formal measurement of GFR with iothalamate, iohexol, or similar markers, these tests are too expensive and time-consuming to be recommended for routine clinical practice. Currently, the most common methods used to estimate GFR are the serum creatinine concentration, the calculated creatinine clearance, and estimation equations based on serum creatinine.[1]

Serum creatinine is, to a first approximation, neither secreted nor reabsorbed, so the amount appearing in the urine per unit time is a measure of the amount that was filtered at the glomerulus during that period. As a result, the rate of creatinine clearance is a reasonably close estimate of the GFR. A decrement in the GFR diminishes creatinine clearance but has no immediate effect on creatinine production by muscle; as a result, the serum creatinine concentration rises. The change in serum creatinine over time indicates the tempo of the renal disease and can distinguish acute injury from chronic kidney disease. Problems with the routine use of serum creatinine alone to infer GFR stem from the differing rates of creatinine production among individuals, mainly because of variations in muscle mass. Women and the elderly can have deceptively low serum creatinine levels despite significant declines in GFR.[2] In addition, the shape of the curve relating the GFR to serum creatinine (Fig. 114-2) has an important and potentially easily overlooked clinical implication, namely, that an initial small absolute rise in creatinine usually reflects a marked fall in GFR.

Creatinine clearance can be calculated with a 24-hour urine collection to measure the creatinine concentration. The patient must be instructed to discard the first morning urine before initiating the collection and to conclude the collection by including the next morning void. The formula for calculating creatinine clearance is as follows:

$$\text{CCr} = (\text{urine Cr} \times \text{V})/(\text{plasma Cr})$$

where *CCr* is creatinine clearance, *urine Cr* is urine creatinine concentration, *V* is urine flow rate, and *plasma Cr* is plasma creatinine. The creatinine clearance overestimates GFR by about 10% owing to tubular secretion of creatinine. Calculation of creatinine clearance from a 24-hour urine collection can be cumbersome for patients and is prone to error because of inaccurate urine collection.

Because of the logistical and practical limitations of a 24-hour urine collection, several equations have been developed to estimate GFR on the basis of easily obtainable clinical data and laboratory results. To date, the most widely used equations are the Cockcroft-Gault, the Modification of Diet in Renal Disease (MDRD) Study, and the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equations (Table 114-1). Weight estimations or ideal weight estimations can make calculation and reporting of Cockcroft-Gault results problematic. The MDRD equations (both the full and abbreviated forms) use data that are readily available to laboratories, but the equations systematically underestimate GFR at higher serum creatinine values, thereby raising concern for false diagnoses of chronic kidney disease. The CKD-EPI equation appears to be more precise and accurate than the MDRD equation, especially at higher GFRs.[3] Cystatin C may provide a more accurate and prognostic measurement of GFR in patients whose creatinine levels are in the upper end of the normal range, but it does not replace estimated GFR measurements for most clinical purposes.[4,5]

Urinalysis

The normal color of the urine is derived from urochromes, which are pigments excreted in the urine. Abnormal color or appearance of the urine may be explained by many conditions (Table 114-2). The basic analysis of the urine sample involves measurements with commercially available dipsticks or microscopic analyses.

Urine Dipstick

The *specific gravity* of the urine generally is related linearly with osmolality. However, it can be raised by the presence of molecules with relatively high molecular weight, such as glucose or contrast dye. A fixed specific gravity of 1.010, so-called isosthenuria, is characteristic of chronic kidney disease (Chapter 130).

Urine pH typically is 5 as a result of daily net acid excretion. An alkaline pH often is noted after meals, when an "alkaline tide" to balance gastric acid excretion increases urine pH. A high urine pH also is seen in patients who are on a vegetarian diet. An exceptionally high urine pH is indicative of an infection with a urea-splitting organism, such as *Proteus* species (Chapter 284). An inappropriately high urine pH in the setting of systemic non–anion gap metabolic acidosis may be seen in certain forms of renal tubular acidosis (Chapter 118). In a proximal renal tubular acidosis, the urine pH is high until the tubular reabsorption threshold for bicarbonate, which is abnormally low,

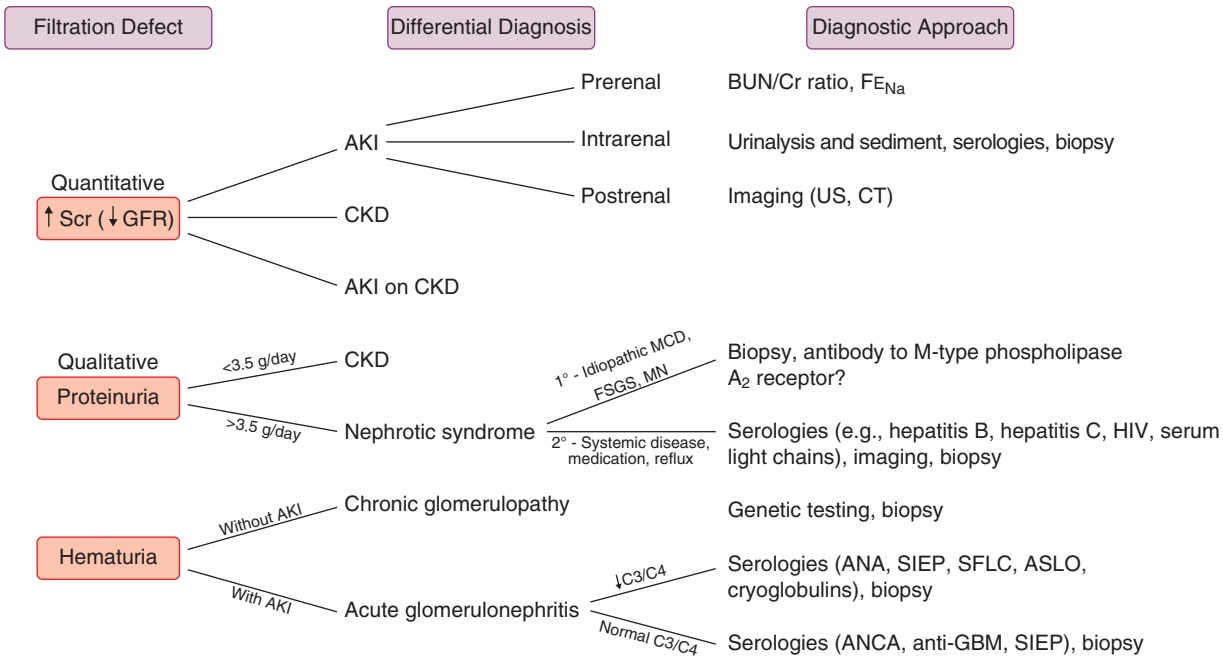

FIGURE 114-1. Overview of approach to kidney disease. Quantitative defects in filtration, manifested by elevated serum creatinine (Scr) and reduced glomerular filtration rate (GFR), should lead to a query into acute kidney injury (AKI) versus chronic kidney disease (CKD). AKI, in turn, is generally divided into prerenal, postrenal, and intrinsic causes. Qualitative defects in filtration, manifested by proteinuria or hematuria, can occur in the absence of changes in GFR and often require biopsy for diagnosis. Proteinuria of more than 3.5 g/day signals nephrotic syndrome, which may be idiopathic or secondary to systemic diseases, such as hepatitis B or C, human immunodeficiency virus (HIV) infection, or diabetes. Glomerular hematuria without AKI is consistent with a chronic glomerulopathy, such as IgA nephropathy, or familial diseases, such as thin basement membranes disease. When hematuria accompanies AKI, acute glomerulonephritis should be suspected and can diagnostically be divided into low-complement glomerulonephritides (immune complex–mediated lesions such as lupus nephritis, postinfectious glomerulonephritis, and cryoglobulinemic glomerulonephritis) and normocomplementemic glomerulonephritides (classically seen in the rapidly progressive glomerulonephropathies due to antineutrophil cytoplasmic antibody [ANCA] and anti–glomerular basement membrane [anti-GBM] antibody). ANA = antinuclear antibody; ASLO = antistreptolysin O; BUN = blood urea nitrogen; Cr = creatinine; CT = computed tomography; MCD = minimal change disease; FE_{Na} = fractional excretion of sodium; FSGS = focal segmental glomerulosclerosis; MN = membranous nephropathy; SFLC = serum free light chain; SIEP = serum immunoelectrophoresis; US, ultrasonography.

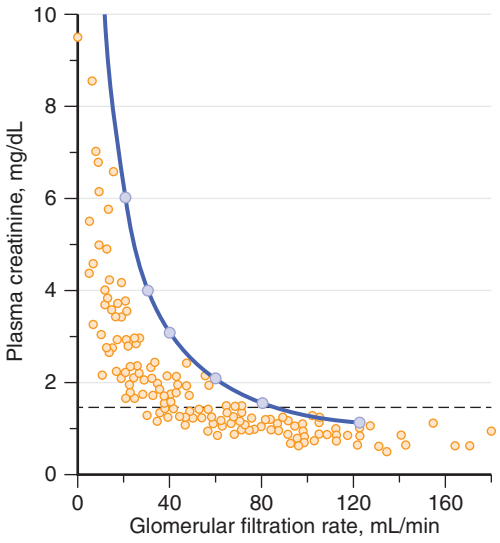

FIGURE 114-2. Relationship between plasma creatinine and glomerular filtration rate measured by inulin clearance in 171 patients (*circles*). The *continuous line* reflects the idealized relationship between these parameters if creatinine were excreted solely by glomerular filtration; the *dashed line* represents an upper limit of "normal" for the creatinine concentration of 1.4 mg/dL. (Redrawn from Shemesh O, Golbetz H, Kriss JP, et al. Limitations of creatinine as a filtration marker in glomerulopathic patients. *Kidney Int.* 1985;28:830-838.)

is reached. At this point, the urine pH decreases to 5. In distal renal tubular acidosis, the inability to create a sufficient gradient for hydrogen ions results in a urine pH that is always higher than 5.5. In type 4 renal tubular acidosis, the urine pH is often 5, and the urine net charge is often positive, thereby confirming the absence of significant amounts of ammonium in the urine; this defect is exacerbated by the accompanying hyperkalemia.

Glucose in the urine is detected by an assay using dipsticks impregnated with the enzyme glucose oxidase. Glycosuria is seen in diabetes mellitus (Chapter 229), when pregnancy causes the tubular threshold for glucose reabsorption to change, and in tubular diseases that affect the proximal convoluted tubule and cause tubular glycosuria. Evidence for pan–proximal tubular dysfunction (e.g., glycosuria, aminoaciduria, phosphaturia) indicates that Fanconi syndrome is present.

The dipstick for *protein* is a sensitive assay based on color change induced by the presence of proteins at a given pH. It is most sensitive to the presence of albumin and is much less sensitive to other proteins, such as the light chains of Bence Jones protein (Chapter 187). The presence of 1+ protein correlates with about 30 mg/dL of albuminuria, and 3+ protein correlates with more than 500 mg/dL of proteinuria. Because the dipstick is not a quantitative measurement, small amounts of proteinuria in an oliguric patient may give the false appearance of high-grade proteinuria. The excretion of abnormal quantities of albumin below the level detectable by the urine dipstick is called *microalbuminuria*. Normal albumin excretion, which is less than 30 mg/day, is best detected by radioimmunoassay or enzyme immunoassay. Microalbuminuria is the earliest clinically detectable stage of diabetic nephropathy (Chapter 124). Proteinuria of increasing severity is associated with a more rapid decline in the GFR, regardless of the GFR,[6] except in minimal change disease (Chapter 121).

The dipstick for *heme* uses the peroxidase-like activity of hemoglobin and myoglobin molecules to detect the presence of heme pigment. The reaction occurs on exposure to hemoglobin, myoglobin, or intact RBCs. The presence of myoglobin, which is found in patients with rhabdomyolysis (Chapter 113), or free hemoglobin, which is seen in patients with intravascular hemolytic anemias (Chapter 160), is suspected if the heme reaction is intensely positive and there is a paucity of cellular elements in the sediment. Persistent, isolated, asymptomatic, microscopic hematuria in adolescents and young adults is associated with a nearly 20-fold increased risk of subsequent end-stage renal disease.[7]

The dipstick detection of *leukocytes* depends on the presence of leukocyte esterase. Leukocyte esterase is usually present in infections (Chapter 284) and in inflammatory conditions.

TABLE 114-1 EQUATIONS FOR ESTIMATION OF GLOMERULAR FILTRATION RATE

COCKCROFT-GAULT	
Male	$CCr\,(mL/min) = \frac{(140 - age) \times \text{lean body wt}\,(kg)}{SCr\,(mg/dL) \times 72}$
Female	$CCr\,(mL/min) = \frac{(140 - age) \times \text{lean body wt}\,(kg) \times 0.85}{SCr\,(mg/dL) \times 72}$
MODIFICATION OF DIET IN RENAL DISEASE 1	
Black male	$GFR = 170 \times SCr^{-0.999} \times age^{-0.176} \times BUN^{-0.170} \times Albumin^{0.318} \times 1.18$
Black female	$GFR = 170 \times SCr^{-0.999} \times age^{-0.176} \times BUN^{-0.170} \times Albumin^{0.318} \times 1.18 \times 0.762$
White male	$GFR = 170 \times SCr^{-0.999} \times age^{-0.176} \times BUN^{-0.170} \times Albumin^{0.318}$
White female	$GFR = 170 \times SCr^{-0.999} \times age^{-0.176} \times BUN^{-0.170} \times Albumin^{0.318} \times 0.762$
MODIFICATION OF DIET IN RENAL DISEASE 2 (ABBREVIATED)	
Black male	$GFR = 186 \times SCr^{-1.154} \times age^{-0.203} \times 1.21$
Black female	$GFR = 186 \times SCr^{-1.154} \times age^{-0.203} \times 1.21 \times 0.742$
White male	$GFR = 186 \times SCr^{-1.154} \times age^{-0.203}$
White female	$GFR = 186 \times SCr^{-1.154} \times age^{-0.203} \times 0.742$
CHRONIC KIDNEY DISEASE EPIDEMIOLOGY COLLABORATION	
Black male, SCr ≤0.9 mg/dL	$GFR = 163 \times (SCr/0.9)^{-0.411} \times 0.993^{age}$
Black male, SCr >0.9 mg/dL	$GFR = 163 \times (SCr/0.9)^{-1.209} \times 0.993^{age}$
Black female, SCr ≤0.7 mg/dL	$GFR = 166 \times (SCr/0.7)^{-0.329} \times 0.993^{age}$
Black female, SCr >0.7 mg/dL	$GFR = 166 \times (SCr/0.7)^{-1.209} \times 0.993^{age}$
White male, SCr ≤0.9 mg/dL	$GFR = 141 \times (SCr/0.9)^{-0.411} \times 0.993^{age}$
White male, SCr >0.9 mg/dL	$GFR = 141 \times (SCr/0.9)^{-1.209} \times 0.993^{age}$
White female, SCr ≤0.7 mg/dL	$GFR = 144 \times (SCr/0.7)^{-0.329} \times 0.993^{age}$
White female, SCr >0.7 mg/dL	$GFR = 144 \times (SCr/0.7)^{-1.209} \times 0.993^{age}$

BUN = blood urea nitrogen; GFR = glomerular filtration rate; SCr = serum creatinine.

TABLE 114-2 MACROSCOPIC APPEARANCE OF URINE

APPEARANCE	CAUSE
Milky	Acid urine: urate crystals Alkaline urine: insoluble phosphates Infection: pus Spermatozoa Chyluria
Smoky pink	Hematuria (>0.54 mL blood/L urine)
Foamy	Proteinuria
Blue or green	*Pseudomonas* urinary tract infection Bilirubin Methylene blue
Pink or red	Aniline dyes in sweets Porphyrins (on standing) Blood, hemoglobin, myoglobin Drugs: phenindione, phenolphthalein Anthocyaninuria (beetroot, "beeturia")
Orange	Drugs: anthraquinones (laxatives), rifampicin Urobilinogenuria
Yellow	Mepacrine Conjugated bilirubin Phenacetin Riboflavin
Brown or black	Melanin (on standing) Myoglobin (on standing) Alkaptonuria
Green or black	Phenol Lysol
Brown	Drugs: phenazopyridine, furazolidone, L-dopa, niridazole Hemoglobin and myoglobin (on standing) Bilirubin

From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine*. 3rd ed. London: Mosby; 2003.

Urine Sediment

RBCs, white blood cells (WBCs), tubular cells, transitional cells, and squamous epithelial cells may be seen in the urine. Casts are formed in tubules, may contain cells or cellular debris, or may be acellular.

RBCs may originate from intrarenal vessels, glomeruli, tubules, or anywhere in the urogenital tract (Fig. 114-3). Dysmorphic RBCs are cells that have been deformed by transit through the glomerulus and through the medullary interstitium, as opposed to RBCs from the remainder of the genitourinary tract (Figs. 114-4 and 114-5); these cells are often lysed and less refractile than nonglomerular RBCs. Dysmorphic RBCs often fragment with poikilocytosis and with blebs, forming so-called "Mickey Mouse" RBCs. Phase contrast microscopy aids in the identification of dysmorphic RBCs. The presence of a majority of dysmorphic RBCs in a urine sediment points to a glomerular origin of the hematuria. The presence of RBC casts is often conclusive evidence for the presence of glomerulonephritis.

WBCs are seen most commonly in urinary tract infections, but they also can be seen in acute interstitial nephritis, infections with *Legionella* (Chapter 314) and *Leptospira* (Chapter 323) species, chronic infections such as tuberculosis (Chapter 324), allergic interstitial nephritis (Chapter 122), atheroembolic diseases (Chapter 125), granulomatous diseases such as sarcoidosis (Chapter 95), IgG4-related interstitial nephritis, and tubulointerstitial nephritis and uveitis syndrome. Mononuclear cells often appear with transplant rejection. Tubular cells, which are seen in many conditions involving tubulointerstitial diseases, also are seen in ischemic and nephrotoxic injury, such as with myeloma kidney (Chapter 187) or cast nephropathy. Eosinophils require special stains, with the Giemsa stain being much less sensitive than the Hansel stain (Chapter 122). Urine eosinophils classically are seen in allergic interstitial nephritis (Chapter 122), but they also are seen in atheroembolic disease (Chapter 125), prostatitis (Chapter 129), and vasculitis.

Casts, which are formed in tubules, are characterized by the arrangement of the cells in a clearly formed matrix composed of Tamm-Horsfall protein. Because casts are formed in the renal parenchyma, they may give a clue to the origin of accompanying cellular elements.

Hyaline casts are composed of Tamm-Horsfall proteins that are formed normally and are seen in increased numbers after exercise (Fig. 114-6).

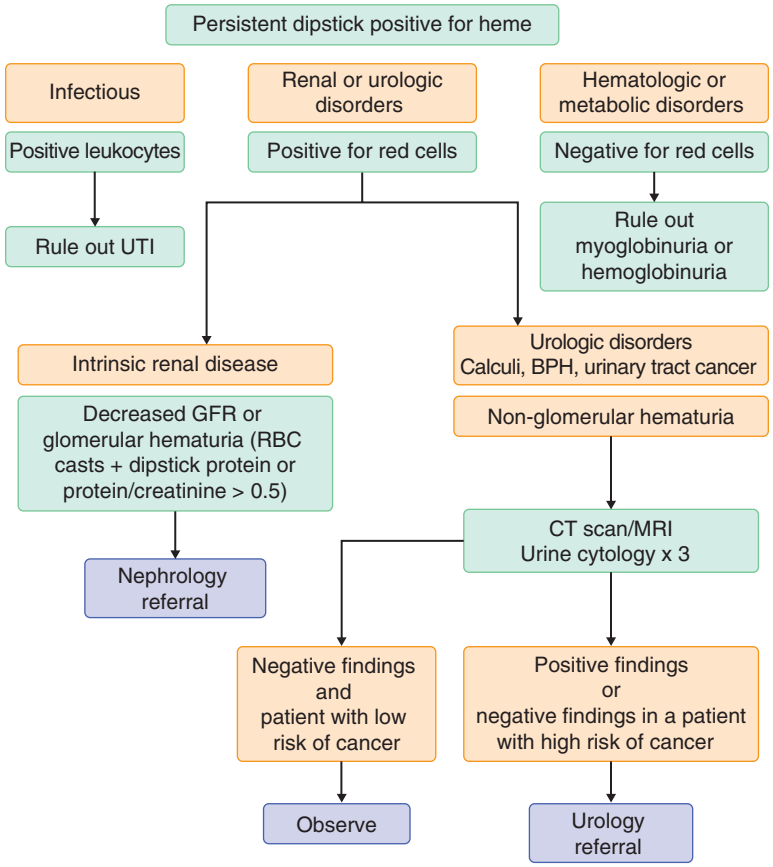

FIGURE 114-3. Algorithm for the evaluation of asymptomatic hematuria. BPH = benign prostatic hyperplasia; CT = computed tomography; GFR = glomerular filtration rate; MRI = magnetic resonance imaging; RBC = red blood cell; UTI = urinary tract infection. (Courtesy Ali Gharavi, MD. Modified from Cohen RA, Brown RS. Microscopic hematuria. *N Engl J Med.* 2003;348:2330-2338).

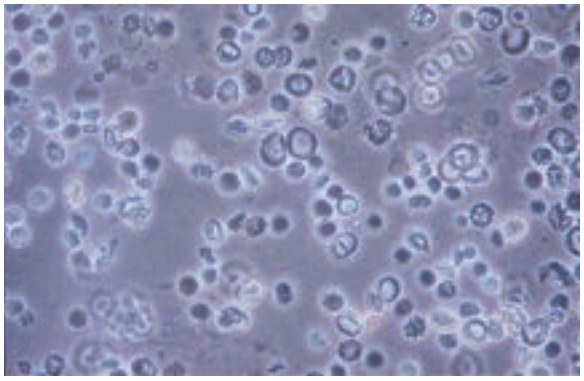

FIGURE 114-4. Dysmorphic erythrocytes. These dysmorphic erythrocytes vary in size, shape, and hemoglobin content and reflect glomerular bleeding. (From Johnson RJ, Feehally J. *Comprehensive Clinical Nephrology*. London: Mosby; 2000.)

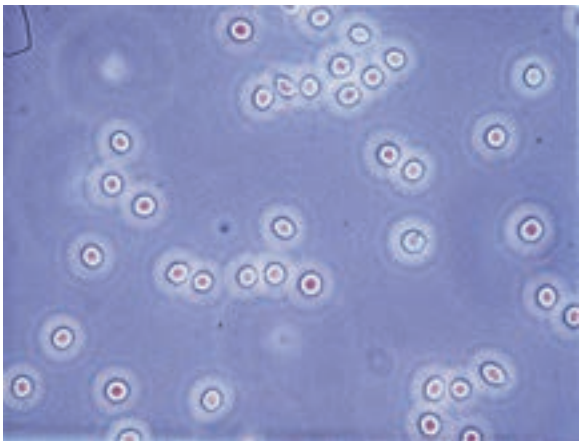

FIGURE 114-5. Isomorphic erythrocytes. These erythrocytes are similar in size, shape, and hemoglobin content. Isomorphic cells reflect nonglomerular bleeding from lesions such as calculi and papillomas or hemorrhage from cysts in polycystic renal disease. (From Johnson RJ, Feehally J. *Comprehensive Clinical Nephrology*. London: Mosby; 2000.)

Granular casts are degenerated tubular cell casts that are seen in the setting of tubular injury (Fig. 114-7). *Pigmented granular casts* are seen in rhabdomyolysis (Chapter 113) with myoglobinuria or, rarely, hemoglobinuria. *RBC casts* (Fig. 114-8) are rarely seen in allergic interstitial nephritis and diabetic nephropathy, but they are frequently seen in acute glomerulonephritis (Chapter 121). The presence of RBC casts in a patient with microscopic hematuria can narrow the focus of the evaluation to a glomerular lesion. *WBC casts* are seen commonly in pyelonephritis (Chapter 284) and in acute and chronic nonbacterial infections. They also are seen in other conditions in which WBCs are associated with parenchymal renal processes, such as allergic interstitial nephritis (Chapter 122), atheroembolic diseases (Chapter 125), and granulomatous diseases such as sarcoidosis (Chapter 95). Rarely, WBC casts can be a dominant feature of many diseases that traditionally are thought of as glomerular diseases, such as lupus nephritis (Chapter 266) and antineutrophil cytoplasmic antibody (ANCA)–associated glomerulonephritis (Chapter 270). *Tubular cell casts* are seen with any acute tubular injury and are the dominant cellular casts in ischemic acute tubular necrosis (Chapter 120). They also can be seen with nephrotoxic injury, such as with aminoglycosides and cisplatin. Some casts may contain both leukocytes and tubular cells.

Crystals can be a normal finding in the urine or serve as clues to pathophysiologic processes. Certain crystals, such as the hexagonal crystals seen with cystinuria (Chapter 128), are always abnormal (Fig. 114-9). Others, such as the octahedral calcium oxalate crystals (Fig. 114-10), may be a normal finding or may be evidence for ethylene glycol intoxication (Chapter 110). Triple phosphate crystals, which are composed of ammonium magnesium phosphate and are coffin shaped (Fig. 114-11), are seen in urinary tract infections with urea-splitting organisms (Chapter 284). Uric acid crystals,

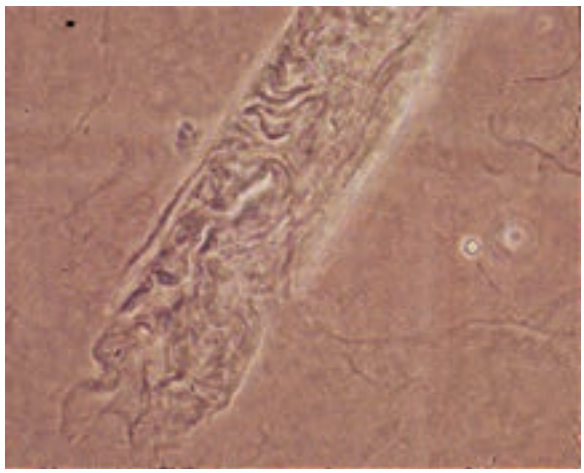

FIGURE 114-6. Hyaline cast of the type seen in small numbers in normal urine. (From Johnson RJ, Feehally J. *Comprehensive Clinical Nephrology*. London: Mosby; 2000.)

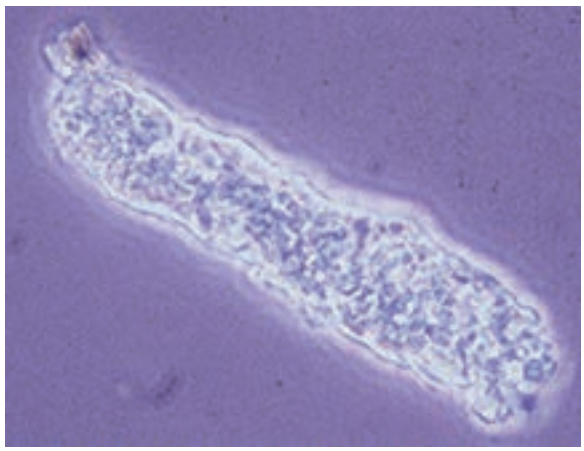

FIGURE 114-7. Number and type of granules and their density in the cast vary in different casts. The presence of erythrocytes in this cast may mean that the granules are derived partly from disrupted erythrocytes. (From Johnson RJ, Feehally J. *Comprehensive Clinical Nephrology*. London: Mosby; 2000.)

FIGURE 114-8. A cast composed entirely of erythrocytes reflects heavy hematuria and active glomerular disease. Crescentic nephritis is likely to be present if erythrocyte cast density is greater than 100/mL. (From Johnson RJ, Feehally J. *Comprehensive Clinical Nephrology*. London: Mosby; 2000.)

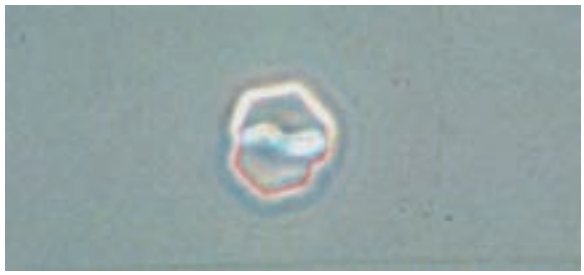

FIGURE 114-9. Typical hexagonal cystine crystal. A single crystal provides a definitive diagnosis of cystinuria. (From Johnson RJ, Feehally J. *Comprehensive Clinical Nephrology*. London: Mosby; 2000.)

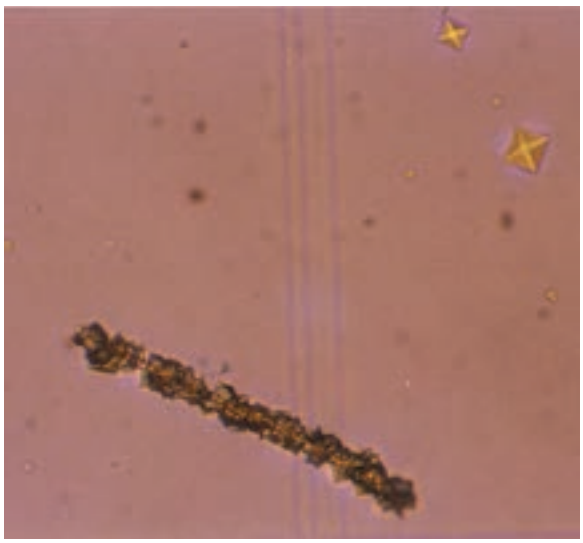

FIGURE 114-10. Oxalate crystals. A pseudocast of calcium oxalate crystals accompanied by crystals of calcium oxalate dehydrate. (From Johnson RJ, Feehally J. *Comprehensive Clinical Nephrology*. London: Mosby; 2000.)

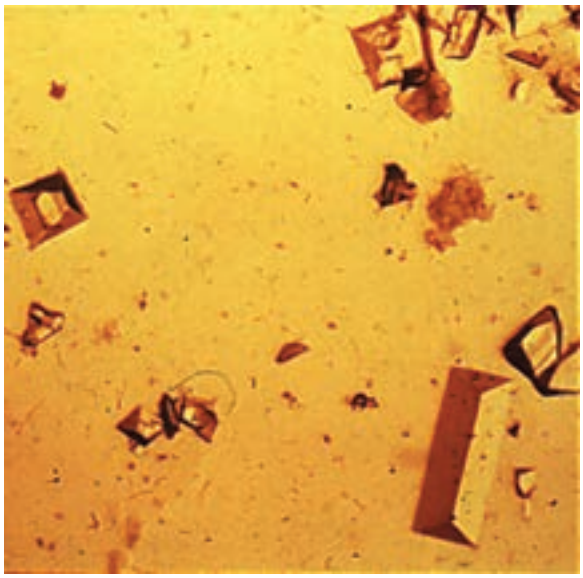

FIGURE 114-11. Coffin-lid crystals of magnesium ammonium phosphate (struvite). (From Johnson RJ, Feehally J. *Comprehensive Clinical Nephrology*. London: Mosby; 2000.)

sodium urate crystals (Fig. 114-12), and calcium phosphate amorphous crystals are common and do not usually have pathologic significance.

Other Elements

Bacteria may be seen in the urine sediment. A spun urine sediment may show rods or cocci in chains, but bacteria are identified best by Gram staining of the urine sediment. Budding yeast forms (which are highly refractile), trichomonads, and spermatozoa also may be seen in the urinary sediment.

SPECIFIC RENAL SYNDROMES

This chapter considers the approach to the patient with acute kidney injury (Chapter 120), glomerular syndromes (nephrotic vs. nephritic; Chapter 121), tubulointerstitial disease (Chapter 122), vasculitis and vascular diseases of the kidney (Chapter 125), papillary necrosis, and chronic kidney disease (Chapter 130).

Acute Kidney Injury

Acute kidney injury (Chapter 120) is a syndrome in which glomerular filtration declines during a period of hours to days. The serum creatinine level is elevated in both acute and chronic kidney disease, but an actively rising serum creatinine level confirms an acute or acute-on-chronic insult to kidney

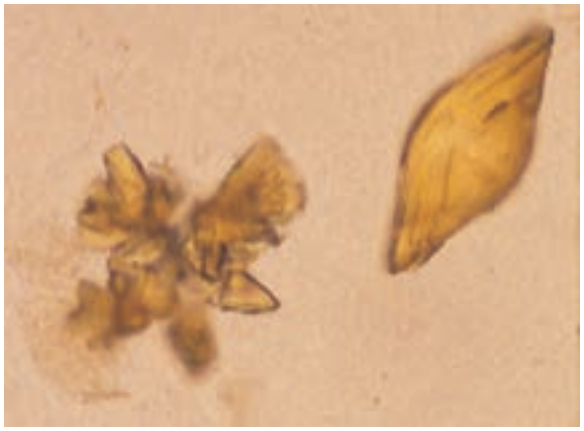

FIGURE 114-12. Urate crystals. Complex crystals suggestive of acute urate nephropathy or urate nephrolithiasis. (From Johnson RJ, Feehally J. *Comprehensive Clinical Nephrology*. London: Mosby; 2000.)

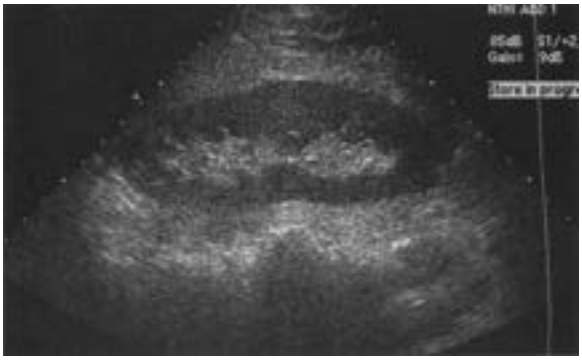

FIGURE 114-13. Normal findings on sagittal renal ultrasound. The cortex is hypoechoic compared with the echogenic fat containing the renal sinus. (From Johnson RJ, Feehally J. *Comprehensive Clinical Nephrology*. London: Mosby; 2000.)

function. As a blood filtration organ, the kidney is susceptible to an acute compromise of renal arterial perfusion (Chapter 125), such as prerenal kidney injury, or blockage in urine outflow, such as urinary obstruction due to benign prostatic hypertrophy (Chapter 129). Thus, the patient with acute renal failure is best approached by evaluation for prerenal, renal, and postrenal causes. The intrarenal causes of acute kidney injury include acute tubular necrosis (Chapter 120), acute interstitial nephritis (Chapter 122), acute glomerulonephritis (Chapter 121), and acute vasculitis and vascular disease (Chapters 121 and 125). The careful and systematic evaluation of the patient should start with a thorough history and physical examination, which should be followed by selected laboratory tests and often an imaging test, such as renal ultrasonography. Most cases of acute renal failure in the hospital have hemodynamic or toxic causes, so prerenal azotemia and acute tubular necrosis must be considered carefully and distinguished from one another.

ETIOLOGY

Prerenal Kidney Injury

Prerenal kidney injury can be caused by shock or renal hypoperfusion from a variety of conditions, including arterial underfilling secondary to edematous states (e.g., severe heart failure, decompensated cirrhosis) or, more variably, cases of nephrotic syndrome. History relevant to renal hypoperfusion states, such as a history of acute gastroenteritis, should be sought. Patients should also be asked about use of nonsteroidal anti-inflammatory drugs or blockers of the renin-angiotensin-aldosterone system (e.g., angiotension-converting enzyme inhibitors, angiotensin receptor blockers) that can exacerbate prerenal injury. Relative hypotension compared with a patient's baseline blood pressure and orthostatic changes in blood pressure and pulse indicate arterial underfilling. Relatively minor orthostatic hypotension may explain the acute decompensation of kidney function in a patient with chronic kidney disease (Chapter 130) or renal artery stenosis (Chapter 125). Lower extremity edema is common in cirrhosis (Chapter 153), heart failure (Chapter 58), and nephrotic syndrome (Chapter 121).

Acute Tubular Necrosis

Acute tubular necrosis can arise from ischemic or toxic injury to the kidneys. Prerenal azotemia can progress to acute tubular necrosis, particularly if frank hypotension occurs in the setting of infection and persists. The transition of prerenal renal failure to acute tubular necrosis may be revealed by a rise in the fractional excretion of sodium to a value greater than 1%. Alternatively, acute tubular necrosis may arise from a toxic effect, so a medication and ingestion history is critical to the evaluation of the patient.

DIAGNOSIS

Laboratory Testing

The normal concentration of blood urea nitrogen (BUN), which is a product of protein catabolism, is about 10-fold higher than the creatinine concentration. Because the BUN-to-creatinine ratio commonly rises with arterial underfilling, BUN typically is used as a marker of effective volume status. Classically, the BUN-to-creatinine ratio will be higher than 15 to 20 in prerenal azotemia but 10 or close to it in acute tubular necrosis. However, the BUN concentration (and hence its ratio to creatinine concentration) may be inappropriately high in other circumstances, such as with high protein intake, gastrointestinal bleeding, or the use of steroids or tetracyclines. The BUN concentration and its ratio to creatinine concentration may be low in patients who have a poor dietary intake of protein, malnutrition, or liver disease.

The excretion of sodium in the setting of oliguria and acute kidney injury (Chapter 120) often gives insight into the appropriateness of tubular function. The fractional excretion of sodium (FE_{Na}) is calculated as follows:

$$FE_{Na} = (\text{urine Na}/\text{plasma Na})/(\text{urine Cr}/\text{plasma Cr}) \times 100$$

where *Na* is the sodium concentration (in mmol/L) and *Cr* is the creatinine concentration (in mmol/L or mg/dL). In the setting of oliguria, FE_{Na} below 1% often denotes prerenal azotemia, whereas FE_{Na} above 1% suggests intrinsic renal damage. Although this measurement is generally useful, FE_{Na} below 1% may be seen without evidence of a prerenal component, including contrast nephropathy (Chapter 120), hepatorenal syndrome (Chapter 154), obstructive uropathy (Chapter 123), interstitial nephritis (Chapter 122), glomerulonephritis (Chapter 121), and rhabdomyolysis (Chapter 113). Conversely, a high FE_{Na} can be seen in cases in which there is a prerenal component, including diuretic use, adrenal insufficiency (Chapter 227), cerebral salt wasting, and salt-wasting nephropathy (Chapter 116). The FE_{Na} must be evaluated in the context of the clinical situation because it can be low or high in a normal patient or in a patient with chronic kidney disease. Ultimately, a patient's volume status is best at the bedside and should not be deduced solely from a measurement of electrolytes.

Imaging

Ultrasonography, which is the most commonly used renal imaging study (Fig. 114-13), provides reliable information about obstruction, kidney size, presence of masses, and renal echotexture. Ultrasonography has only a 90% sensitivity for the detection of hydronephrosis and hence is not sufficient to exclude obstruction (Chapter 123) with certainty. In addition, its inability to detect stones in the ureters and bladder limits its utility in the evaluation for kidney stones (Chapter 126). Ultrasonography can detect vascular disease, and Doppler imaging permits evaluation of the renal vessels with resistive indices. Resistive indices are crucial in ascribing renal dysfunction to the detected vascular disease (Chapter 125). A high resistive index reflects parenchymal disease with scarring and indicates that intervention on the vascular disease itself is unlikely to improve renal function.

A computed tomography (CT) scan stone protocol to assess the kidneys, ureters, and bladder is the study of choice for detecting kidney stones (Chapter 126) because of its ability to detect stones of all kinds, including uric acid stones and nonobstructing stones, as well as stones in the ureters (Fig. 114-14). Masses in the kidney can be evaluated with either contrast CT or a renal ultrasound examination. CT angiography with iodinated contrast material can assess possible renal artery stenosis (Chapter 125) with an accuracy comparable to that of magnetic resonance (MR) angiography.

Glomerular Syndromes: Nephrotic versus Nephritic

The nephrotic syndrome (Chapter 121) is characterized by the presence of proteinuria of more than 3.5 g/day/1.73 m^2, with accompanying edema, hypertension, and hyperlipidemia. Other consequences include a predisposition to infection and hypercoagulability. In general, the diseases associated

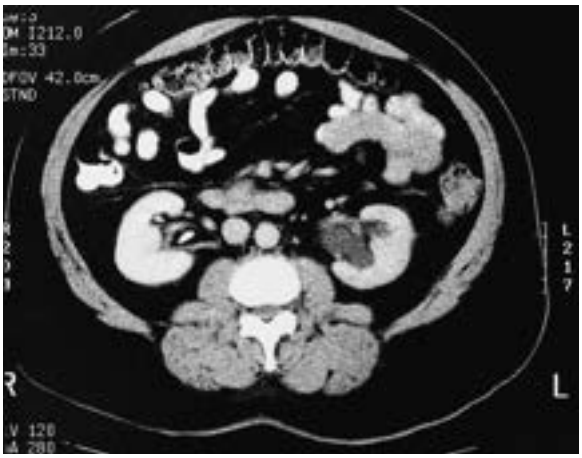

FIGURE 114-14. **Delayed excretion in the left kidney secondary to a distal calculus. Contrast-enhanced computed tomography scan shows dilated left renal pelvis. (From Johnson RJ, Feehally J. *Comprehensive Clinical Nephrology*. London: Mosby; 2000.)**

with nephrotic syndrome do not cause acute kidney injury, although acute kidney injury may be seen with minimal change disease, human immunodeficiency virus (HIV)–associated nephropathy, and bilateral renal vein thrombosis (Chapter 125). The causes of primary idiopathic nephrotic syndrome, in decreasing order of prevalence, are focal and segmental glomerulosclerosis, membranous nephropathy, minimal change disease, and membranoproliferative glomerulonephritis. Membranous nephropathy has been associated with antibodies to the M-type phospholipase A_2 receptor. Secondary causes of the nephrotic syndrome include diabetic nephropathy (Chapter 124), amyloidosis (Chapter 188), and membranous lupus nephritis (Chapters 121 and 266).

The acute nephritic syndrome is an uncommon but dramatic presentation of an acute glomerulonephritis (Chapter 121). The hallmark of the acute nephritic syndrome is the presence of dysmorphic RBCs and RBC casts, but their absence does not exclude the syndrome. The acute nephritic syndrome can be caused by any of the rapidly progressive glomerulonephropathies with ANCA-associated vasculitis (granulomatosis with polyangiitis, microscopic polyangiitis, and eosinophilic granulomatosis with polyangiitis), anti–glomerular basement membrane (anti-GBM) glomerulonephritis, and immune complex–mediated glomerulonephritis (including systemic lupus erythematosus, cryoglobulinemia, postinfectious glomerulonephritis, endocarditis, IgA nephropathy, and Henoch-Schönlein purpura). The rapid decline in renal function often warrants urgent and usually inpatient evaluation.

DIAGNOSIS

Laboratory Testing

Proteinuria (as albuminuria) of more than 3.5 g in 24 hours generally indicates glomerular disease (Chapter 121). Lesser quantities do not preclude glomerular disease, and electrophoresis gives valuable insight into the composition of the proteinuria (Chapter 187). On occasion, overflow proteinuria of a low-molecular-weight protein, such as light chains in Bence Jones proteinuria, can be higher than 3.5 g/day without any of the manifestations or implications of the nephrotic syndrome; a urine protein electrophoresis study is important in making the distinction. A comparison of the microalbumin-to-creatinine ratio with the protein-to-creatinine ratio will give an insight into the presence of Bence Jones protein because of the absence of albuminuria despite significant proteinuria. Collection must be done by discarding the first morning void and collecting all urine output for the next 24 hours, including the first morning void the next day.

The 24-hour urine collection for protein excretion is cumbersome and subject to inaccuracies. Instead, a spot urine sample for protein and creatinine can be used to estimate the amount of protein excreted. A protein-to-creatinine ratio of 3 translates to a 24-hour protein excretion of about 3 g. The ratio is most accurate when the first morning urine collection is used and may be inaccurate in patients with orthostatic proteinuria.

The evaluation of proteinuric renal dysfunction, particularly when glomerular diseases are suspected, should follow a stepwise progression from noninvasive serologic evaluation to a definitive or confirmatory diagnostic evaluation, such as a renal biopsy.[8] Sometimes an expeditious diagnosis is needed, and a biopsy may be done relatively early in the evaluation.

Serologies

An *antinuclear antibody* (ANA) titer can be useful to evaluate glomerular disease in either nephrotic or nephritic presentations. A high ANA titer (e.g., 1 : 320), especially if it is accompanied by a more specific finding such as anti–double-stranded DNA antibody or anti-Smith antibody, can be highly specific for the diagnosis of lupus nephritis (Chapter 266), which usually requires a renal biopsy. Lower titers (e.g., 1 : 80 or 1 : 40) are nonspecific.

A *rheumatoid factor* titer will usually be elevated in patients with rheumatoid arthritis (Chapter 264), but vasculitis is a relatively late and rare event. Rheumatoid factor can be detected in some forms of cryoglobulinemia (Chapter 187); for example, IgM, which is present in type II and type III cryoglobulinemia, has rheumatoid factor activity. Rheumatoid factor also can be seen as a nonspecific finding in bacterial endocarditis (Chapter 76) and systemic vasculitis (Chapter 270).

The levels of *complement* components C3 and C4 and the 50% hemolyzing dose of complement (CH_{50}) usually are measured to evaluate suspected rapidly progressive glomerulonephritis (Chapter 121). Complement levels are usually low in active systemic lupus erythematosus (Chapter 266), poststreptococcal glomerulonephritis (Chapter 121), endocarditis (Chapter 76), membranoproliferative glomerulonephritis, cryoglobulinemia (Chapter 187), shunt nephritis with infection of a ventriculoatrial shunt, and glomerulonephritis associated with visceral abscesses. A particularly depressed C4 compared with C3 should raise the suspicion of cryoglobulinemia.

Serum immunoelectrophoresis will detect elevated polyclonal IgA levels in about 50% of cases of IgA nephropathy (Chapter 121) and Henoch-Schönlein purpura (Chapter 121). Polyclonal elevation of IgG may occur in a variety of systemic diseases and is a nonspecific finding. The presence of a monoclonal protein in the serum should raise the suspicion for a monoclonal gammopathy–associated disease (Chapter 187). The differential diagnosis includes monoclonal gammopathy of uncertain significance, myeloma kidney, lymphomas (Chapter 185), amyloidosis (Chapter 188), light chain deposition disease, heavy chain deposition disease, immunotactoid glomerulonephritis, and cryoglobulinemia. The concentration of the monoclonal protein is higher when the diagnosis of multiple myeloma is made, but even small quantities of Bence Jones proteins in the serum can have clinical significance.

A *urine immunoelectrophoresis* always should be obtained concomitantly if myeloma is suspected. Because a substantial fraction of multiple myelomas can have no heavy chain excretion and small quantities of light chains may be difficult to detect by serum immunoelectrophoresis, a urine immune electrophoresis test for Bence Jones protein complements the serum immunoelectrophoresis. In light chain myeloma, patients may have Bence Jones proteinuria even in the absence of an M component in the serum immunoelectrophoresis. Bence Jones proteinuria may be present in myeloma kidney, amyloidosis, light chain deposition disease, lymphoma, or, occasionally, monoclonal gammopathy of uncertain significance. However, some patients with systemic AL (light chain) amyloidosis have a normal serum immunoelectrophoresis and no Bence Jones proteinuria (Chapter 187). More sensitive assays for serum free light chains and an assessment of the ratio of κ to λ lights chains increase the sensitivity for detection of monoclonal gammopathies.

The *antineutrophil cytoplasmic antibody* (ANCA) assay has allowed earlier and more definitive recognition of vasculitic causes of rapidly progressive glomerulonephritis (Chapter 270), especially granulomatosis with polyangiitis, microscopic polyangiitis, and eosinophilic granulomatosis with polyangiitis, when it is confirmed by enzyme-linked immunosorbent assay. The antibodies cause two different patterns of staining: perinuclear staining (p-ANCA) and cytoplasmic staining (c-ANCA). Both antigens actually have a cytoplasmic distribution, and the perinuclear staining pattern is an artifact of the fixation method. In most cases, the antigen for p-ANCA is myeloperoxidase (MPO), whereas the antigen for c-ANCA is proteinase 3 (PR3). Anti-MPO antibodies are associated with microscopic polyangiitis, idiopathic crescentic glomerulonephritis, or Churg-Strauss syndrome (eosinophilic granulomatosis with polyangiitis; Chapter 270). Anti-PR3 antibodies often correlate with the classic disease of granulomatosis with polyangiitis (formerly known as Wegener granulomatosis) (Chapter 270).

Anti–glomerular basement membrane (anti-GBM) antibodies are autoantibodies to the Goodpasture antigen (Chapter 121), which resides in a domain of the α chain of type 4 collagen. An early and accurate diagnosis of Goodpasture syndrome can be made by immunofluorescence and confirmed by Western blot analysis. Anti-GBM antibody staining also may occur in the

presence of a positive ANCA. In these cases, the theory is that exposure of the Goodpasture antigen, as a result of the glomerular injury, leads to anti-GBM antibody formation as a secondary process.

Cryoglobulins (Chapter 187) are thermolabile immunoglobulins. They are a single monoclonal type in type I cryoglobulinemia. In type II and type III cryoglobulinemia, however, the mixture of immunoglobulins includes one with rheumatoid factor activity against IgG. Type I and type II cryoglobulins are more likely to be associated with clinical disease, especially at higher titers. Type III cryoglobulinemia is often of less clinical significance. Type I cryoglobulinemia is seen with Waldenström macroglobulinemia and multiple myeloma (Chapter 187); type II, with hepatitis C infection (Chapters 148 and 149), Sjögren syndrome (Chapter 268), lymphomas (Chapters 185 and 186), and systemic lupus erythematosus (Chapter 266); and type III, with hepatitis C (Chapters 148 and 149), chronic infections, and inflammatory conditions. When cryoglobulinemia is associated with hepatitis C, the hepatitis C virus (HCV) RNA is concentrated in the cryoprecipitate; the diagnosis can be made by an RNA assay of the cryoprecipitate at 37° C.

Membranous nephropathy is associated with chronic hepatitis B infection with hepatitis B surface antigenemia (Chapter 149). Classic polyarteritis nodosa (Chapter 270) occasionally is seen with chronic hepatitis B infection, often with surface antigenemia and hepatitis B e antigenemia. M-type phospholipase A_2 receptor antibodies also have been detected as autoantibodies in idiopathic membranous nephropathy.

Hepatitis C serology is associated with a variety of renal diseases, including cryoglobulinemia, membranoproliferative glomerulonephritis, and membranous nephropathy. The evaluation may include the antibody test and an assay for HCV RNA. On occasion, the HCV RNA analysis may have to be conducted on the cryoprecipitate at 37° C.

HIV-associated nephropathy (Chapter 121) is associated with nephrotic syndrome and acute kidney injury. In the appropriate clinical setting, HIV serology and viral titers are warranted tests for both clinical syndromes.

Streptococcal infection can be confirmed as the cause of postinfectious glomerulonephritis (Chapter 121) with an anti-DNase or antistreptolysin assay. Acute and convalescent serology assays are used to confirm recent infection.

The erythrocyte sedimentation rate (ESR) is a relatively nonspecific test in the evaluation of renal disease. However, a high ESR often points to systemic vasculitis (Chapter 270), multiple myeloma (Chapter 187), or malignant disease as the underlying cause. However, the ESR often is elevated in the nephrotic syndrome (Chapter 121), including diabetic nephropathy (Chapter 124).

Renal Biopsy

No formal guidelines exist for the indications to perform a renal biopsy. Most nephrologists will perform a biopsy for adults with idiopathic nephrotic syndrome and for children with steroid-dependent or steroid-resistant nephrotic syndrome. In addition, acute kidney injury without an identifiable inciting cause is a clear indication for biopsy. Notably, patients with hospital-acquired kidney failure rarely meet this indication. Other abnormal clinical findings, such as gross or microscopic hematuria or subnephrotic proteinuria, often but not always lead to a kidney biopsy. Renal biopsy usually is performed percutaneously with real-time ultrasound or CT guidance. About 1 to 2% of patients without an underlying coagulopathy will develop bleeding that requires a transfusion. The transjugular approach can be used in patients in whom the risks for bleeding are high.

The decision to pursue a kidney biopsy should be individualized for each patient, but a renal biopsy generally is justified for most patients with two or more of the following four findings: hematuria, proteinuria above 1 g/day, renal insufficiency, or positive serologies for systemic diseases with known potential for kidney involvement (e.g., hepatitis B or C virus infection, systemic lupus erythematosus, ANCA seropositivity). The decision about whether to perform a renal biopsy in diabetic patients with suspected diabetic nephropathy should be individualized and is usually driven by the presence of atypical features or an active urine sediment.[9] In addition, in patients with renal transplants (Chapter 131) and acute or chronic renal failure, biopsy of the allograft kidney provides crucial information in guiding diagnosis and treatment.

Tubulointerstitial Diseases

Tubulointerstitial diseases (Chapter 122) vary in presentation from acute kidney injury to chronic kidney dysfunction that initially is manifested as asymptotic mild renal insufficiency (Table 114-3). The urine sediment often contains small to moderate amounts of proteinuria, usually less than 1 g/day, as well as WBCs, RBCs, tubular cells, and WBC casts. RBC casts are rare in acute interstitial nephritis and are more characteristic of glomerular disease.

TABLE 114-3 MAJOR CAUSES OF TUBULOINTERSTITIAL DISEASE

Ischemic and toxic acute tubular necrosis
Allergic interstitial nephritis
Interstitial nephritis secondary to immune complex–related collagen vascular disease, such as Sjögren disease or systemic lupus erythematosus
Granulomatous diseases: sarcoidosis, tubulointerstitial nephritis with uveitis
IgG4-related interstitial nephritis
Pigment-related tubular injury: myoglobulinuria, hemoglobinuria
Hypercalcemia with nephrocalcinosis
Tubular obstruction: drugs such as indinavir, uric acid in tumor lysis syndrome
Myeloma kidney or cast nephropathy
Infection-related interstitial nephritis: *Legionella*, *Leptospira* species
Infiltrative diseases, such as lymphoma

Vasculitis and Vascular Diseases of the Kidney

Vascular diseases of the kidney can be divided into large-vessel obstruction and medium- to small-vessel diseases (Chapter 125). Renovascular disease is a common cause of hypertension, heart failure, and renal insufficiency. About 90% of renal artery stenosis is atherosclerotic in origin, with most of the remaining caused by fibromuscular dysplasia, which is more common in women 20 to 50 years of age. Medium-sized arterial vessel diseases include polyarteritis nodosa, which is seen in patients with hepatitis B (Chapters 148 and 149), HIV infection (Chapter 121), or, rarely, hepatitis C (Chapters 148 and 149). Symptoms include abdominal pain, hypertension, and mild renal insufficiency, often with a benign sediment; diagnostic findings include microaneurysms at the bifurcation of medium-sized arteries. Other diseases involving small vessels include atheroembolic disease (Chapter 125), which is seen either spontaneously or after arteriography or surgery. This syndrome typically affects the kidneys, gastrointestinal tract, and lower extremities, but it can also involve the central nervous system when the aortic arch is affected.

The thrombotic microangiopathies include hemolytic-uremic syndrome (HUS) and thrombotic thrombocytopenic purpura (Chapter 172). Thrombocytopenic purpura is associated with an acquired inhibitor to or the congenital inherited absence of a protease that cleaves large-molecular-weight von Willebrand multimers. HUS is caused by endothelial injury. In diarrhea-positive (or typical) HUS, the endothelial injury is induced by Shiga toxin from *Escherichia coli* O157:H7 infection. In diarrhea-negative (atypical) HUS, dysregulation of the alternative complement pathway is the underlying cause of endothelial injury. The antiphospholipid antibody syndrome (Chapter 176) can cause large-vessel thrombosis and stenosis as well as a thrombotic microangiopathy with proteinuria, hypertension, and renal insufficiency. Scleroderma renal crisis, which is a manifestation of systemic sclerosis (Chapter 267), often leads to an inexorable progression to end-stage renal insufficiency if untreated.

A systemic vasculitis may be manifested in a variety of ways, including skin manifestations such as petechial rash, purpura, digital gangrene, and splinter hemorrhages. Otitis, sinusitis, epistaxis, hemoptysis, and nasal septal ulcers are common manifestations of granulomatosis with polyangiitis (Chapter 270). Pulmonary hemorrhage can be a catastrophic manifestation of Goodpasture syndrome (Chapter 121) or anti-GBM disease as well as the ANCA-associated vasculitis (Chapter 270). Abdominal pain and tenderness and gastrointestinal hemorrhage may be observed in Henoch-Schönlein purpura and classic polyarteritis nodosa (Chapter 270). Neurologic symptoms may be a manifestation of vasculitis, such as microscopic polyangiitis (Chapter 270) and cryoglobulinemia (Chapter 187).

DIAGNOSIS

Radiologic Evaluation

Magnetic resonance imaging (MRI) with MR angiography (Fig. 114-15) is highly sensitive for detecting atherosclerotic renovascular disease (Chapter 125), but it tends to overestimate the degree of stenosis. Its accuracy in detecting fibromuscular dysplasia, however, is less well validated. MRI also can be used to evaluate renal masses. MRI does not require iodinated contrast material, but gadolinium-based contrast agents for vascular studies are

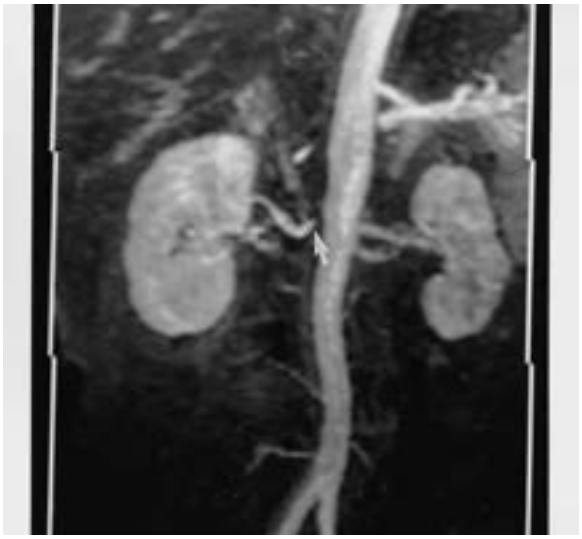

FIGURE 114-15. Magnetic resonance angiography. Coronal three-dimensional image shows right renal artery stenosis *(arrow)*. (From Johnson RJ, Feehally J. *Comprehensive Clinical Nephrology*. London: Mosby; 2000.)

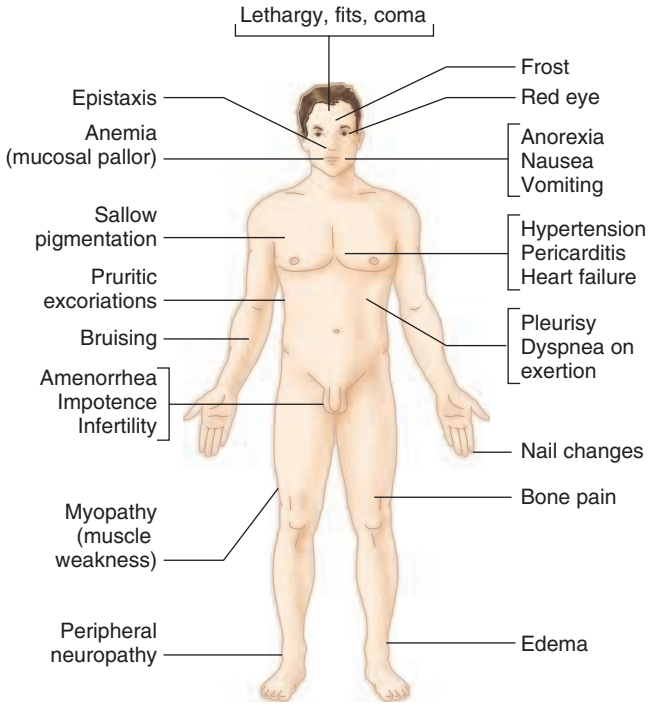

FIGURE 114-16. Common symptoms and signs of chronic renal failure. (Redrawn from Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine*. 3rd ed. London: Mosby; 2003.)

associated with the syndrome of nephrogenic systemic fibrosis in patients with advanced renal failure (Chapter 267).

Renal arteriography, which is the "gold standard" in the evaluation of renal artery stenosis (Chapter 125), also is used for the evaluation of arteriovenous malformations, polyarteritis nodosa, and other vascular lesions of the kidneys. This invasive study uses iodinated contrast material and incurs a small risk for atheroembolic disease (Chapter 125). Therapeutic angioplasty and stenting can be done at the time of angiography.

Papillary Necrosis

Acute necrosis of the renal papilla is associated with sickle cell anemia (Chapter 163), analgesic nephropathy (Chapter 122), diabetic nephropathy (Chapter 124), and obstructive pyelonephritis (Chapter 284). In sickle cell disease (Chapter 163)[10], the hypoxic and hypertonic milieu of the inner medulla promotes sickling, and chronic sickling at the vasa recta results in medullary ischemia. Massive and prolonged consumption of analgesics, particularly the combination of aspirin, caffeine, and acetaminophen, is associated with chronic interstitial nephritis and a predisposition to papillary necrosis (Chapter 122); medullary ischemia is thought to be caused by inhibition of synthesis of vasodilatory prostaglandins by aspirin, and direct toxicity is attributed to metabolites of phenacetin. Similarly, medullary perfusion is thought to be compromised in diabetic nephropathy (Chapter 124) and obstructive pyelonephritis (Chapter 123).

The clinical manifestations of papillary necrosis can include flank pain and hematuria. If the papilla is sloughed, obstruction may occur at the renal pelvis or ureter of the affected kidney, with referred pain migrating from the flank to the groin. A sloughed papilla may precipitate frank renal failure if the function of the contralateral kidney is impaired or if obstruction occurs at the level of the bladder or urethra (Chapter 123).

Classically, papillary necrosis is diagnosed on an excretory pyelogram as a calyceal defect after sloughing of a papilla, but CT with contrast enhancement is as good for advanced lesions. If the necrotic papilla is retained, however, the defect will be more subtle. Transitional cell carcinoma (Chapter 197) can occur in the setting of papillary necrosis or can mimic its appearance. Obstruction, if present, must be relieved, but treatment otherwise is limited to pain control and hydration.

Chronic Kidney Disease

Chronic kidney disease, which is defined as either kidney damage or a GFR of less than 60 mL/min/1.73 m^2 for longer than 3 months, includes five stages (Table 114-4). Kidney damage is defined as pathologic abnormalities or markers of kidney damage, including abnormalities in the composition of blood or urine or abnormalities on imaging tests. The excretion of 30 to 300 mg of albumin in a 24-hour period defines microalbuminuria. An estimated 12% of the adult U.S. population has abnormal albumin excretion in the urine, and the frequency increases with age. Kidney failure is defined as either a GFR of less than 15 mL/min/1.73 m^2 that is accompanied by signs and symptoms of uremia or a need for initiation of kidney replacement therapy for treatment of complications of decreased GFR (Fig. 114-16). End-stage renal disease includes all cases requiring treatment by dialysis or transplantation regardless of the level of GFR.

TABLE 114-4 STAGES OF CHRONIC KIDNEY DISEASE*

STAGE	DESCRIPTION	GFR (mL/min/1.73 m^2)
1	Kidney damage with normal or ↑GFR	≥90
2	Kidney damage with mild or ↓GFR	60-89
3	Moderate ↓GFR	30-59
4	Severe ↓GFR	15-29
5	Kidney failure	<15 (or dialysis)

*Chronic kidney disease is defined as either kidney damage or GFR <60 mL/min/1.73 m^2 for ≥3 months. Kidney damage is defined as pathologic abnormalities or presence of markers of damage, including abnormalities in blood or urine test results or imaging studies.
GFR = glomerular filtration rate.

Patients with chronic kidney disease warrant referral to a nephrologist. Care of these patients should focus on efforts to slow disease progression, to optimize medical management, and to make a seamless transition to renal replacement therapy (Chapter 130).[11] The care should include optimal blood pressure control, use of angiotensin-converting enzyme inhibitors and angiotensin receptor blockers if indicated, dietary counseling, careful management of calcium and phosphorus levels, control of the parathyroid hormone level, and management of anemia with the use of erythropoietin and iron supplements. Early referral for placement of access for dialysis and initiation of transplant evaluation (Chapter 131) are important components of the care of patients with chronic kidney disease.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. O'Riordan P, Stevens PE, Lamb EJ. Estimated glomerular filtration rate. *BMJ*. 2014;348:g264.
2. Hallan SI, Matsushita K, Sang Y, et al. Age and association of kidney measures with mortality and end-stage renal disease. *JAMA*. 2012;308:2349-2360.
3. Matsushita K, Mahmoodi BK, Woodward M, et al. Comparison of risk prediction using the CKD-EPI equation and the MDRD study equation for estimated glomerular filtration rate. *JAMA*. 2012;307:1941-1951.
4. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367:20-29.
5. Shlipak MG, Matsushita K, Arnlov J, et al. Cystatin C versus creatinine in determining risk based on kidney function. *N Engl J Med*. 2013;369:932-943.
6. Turin TC, James M, Ravani P, et al. Proteinuria and rate of change in kidney function in a community-based population. *J Am Soc Nephrol*. 2013;24:1661-1667.
7. Vivante A, Afek A, Frenkel-Nir Y, et al. Persistent asymptomatic isolated microscopic hematuria in Israeli adolescents and young adults and risk for end-stage renal disease. *JAMA*. 2011;306:729-736.
8. Browne OT, Bhandari S. Interpreting and investigating proteinuria. *BMJ*. 2012;344:e2339.
9. Sharma SG, Bomback AS, Radhakrishnan J, et al. The modern spectrum of renal biopsy findings in patients with diabetes. *Clin J Am Soc Nephrol*. 2013;8:1718-1724.
10. Ataga KI, Derebail VK, Archer DR. The glomerulopathy of sickle cell disease. *Am J Hematol*. 2014;89:907-914.
11. Stevens PE, Levin A. Evaluation and management of chronic kidney disease: synopsis of the Kidney Disease: Improving Global Outcomes 2012 Clinical Practice Guideline. *Ann Intern Med*. 2013;158:825-830.

REVIEW QUESTIONS

1. Idiopathic membranous nephropathy is associated with which of the following?

A. Hepatitis B infection
B. Hepatitis C infection
C. Antibodies to the M-type phospholipase A_2 receptor
D. Nonsteroidal use
E. None of the above

Answer: C Idiopathic membranous nephropathy has been identified to have antibodies to the M-type phospholipase A_2 receptor. Although secondary membranous nephropathy can be associated with hepatitis B and rarely with hepatitis C or drugs, the idiopathic disease seems to be mediated by autoimmunity.

2. The most sensitive screening test for detection of a monoclonal gammopathy is

A. Serum immunoelectrophoresis
B. Serum free light chain ratio
C. Urine for Bence Jones
D. Bone marrow biopsy
E. Urine free light chain ratio

Answer: B The serum immunoelectrophoresis is sensitive for the detection of intact immunoglobulins but less so for the detection of isolated light chains, which can occur in a substantial proportion of monoclonal gammopathies. An abnormal κ:λ ratio is the most sensitive test to detect a clonal plasma cell disorder. Urine free light chain ratios are less reliable than serum ratios. A bone marrow biopsy is used to diagnose myeloma, but small numbers of plasma cells can cause many plasma cell disorders.

3. What is the best estimate of glomerular filtration rate in patients with renal function in the upper limits of the normal range?

A. Modification of Diet in Renal Disease (MDRD) 1 equation
B. Modification of Diet in Renal Disease (MDRD) 2 equation
C. Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation
D. None of the above

Answer: C MDRD 1 and 2 equations were derived in a population of patients with chronic kidney disease. MDRD 1 equation includes race, gender, creatinine, age, blood urea nitrogen, and albumin, but the other two equations do not include blood urea nitrogen or albumin. Of these equations, CKD-EPI is the most accurate for patients with mild chronic kidney disease.

4. Which of the following statements regarding the urine dipstick is false?

A. The specific gravity can be lowered by contrast agents in the urine.
B. Red blood cells, myoglobin, or hemoglobin can be detected by the presence of heme.
C. The dipstick can miss Bence Jones proteins.
D. The urine pH can be high in the presence of infections.
E. The urine pH is usually higher than 5.5 in distal renal tubular acidosis.

Answer: A The high density of contrast material makes the urine specific gravity high rather than low. The dipstick detects all heme pigments, including myoglobin and hemoglobin, as well as red blood cells themselves. The dipstick detects albumin, but Bence Jones proteins may be missed or may cause only weak reaction in the dipstick. Hence, an immunoelectrophoresis or urine protein-to-creatinine ratio will be important; immunoelectrophoresis identifies the presence of clonal light chains, and the urine protein-to-creatinine ratio is for the actual quantification of the degree of proteinuria.

115 STRUCTURE AND FUNCTION OF THE KIDNEYS

QAIS AL-AWQATI AND JONATHAN BARASCH

The kidney regulates the ionic composition and volume of body fluids, the excretion of nitrogenous waste, the elimination of exogenous molecules (e.g., many drugs), the synthesis of a variety of hormones (e.g., erythropoietin), and the metabolism of low-molecular-weight proteins (e.g., insulin). Befitting such an array of responsibilities, the kidney receives 25% of the cardiac output. The gross anatomy of the kidney is notable for a weight of approximately 150 g and a characteristic bean shape with approximate dimensions of 11 × 6 × 2.5 cm. On bisection, a simple gross structure is evident with an outer cortex and a more central medulla that narrows to multiple papillae at the apices of so-called pyramids (Fig. 115-1).

Understanding of the kidney, however, requires an appreciation of the intricate microstructure that underlies its complex functions. Although the kidney is an organ, the *nephron* is actually the organ's definable and independent unit. The human kidney is composed of approximately 1 million essentially identical nephrons. All the functions of the kidney are performed by each individual nephron, and to a first approximation, all nephrons are independent of each other because they have their own innervation and blood supply. The nephron is made up of two functional subunits, the glomerulus and the tubules and ducts (Fig. 115-2). The glomerulus begins with the branching of the afferent arteriole, an end artery of the corresponding renal artery, to a tuft of capillaries. The glomerular capillaries invaginate an epithelium with the visceral epithelial cells adjacent to the capillary and the parietal epithelial cells outside this tuft. The space between the epithelial layers is the urinary space. The fenestrated glomerular capillary endothelium, the intervening basement membrane, and the foot processes of the visceral epithelium, so-called podocytes, make up the glomerular filtration barrier. The balance of hydrostatic and oncotic pressures drives the extrusion of a protein-free filtrate through this barrier into the urinary space. The urinary space then leads to a series of tubules and ducts: the proximal tubule, the thin limb of the loop of Henle, the thick limb of the loop of Henle, the distal convoluted tubule, the cortical collecting duct, and the medullary collecting duct. The papillary collecting duct empties through the renal papilla into the renal pelvis and then to the ureter. The glomerular capillary bed coalesces to form the efferent arteriole, a vessel that is exquisitely sensitive to angiotensin II, and then the peritubular (proximal) capillaries. This system allows efferent arteriole constriction to regulate proximal tubule reabsorption, as described later.

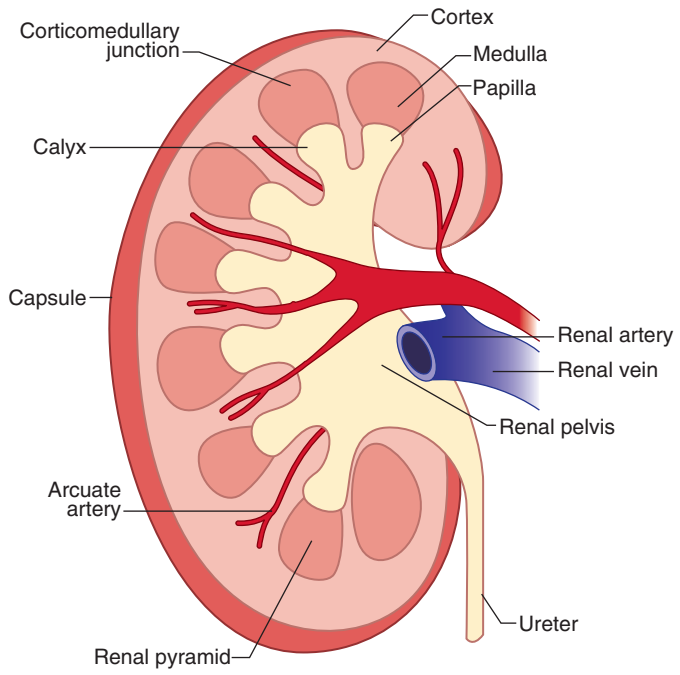

FIGURE 115-1. Sagittal section of the human kidney depicting gross anatomy and organization.

The nephron regulates homeostasis by three actions. First, in the glomerulus, nephrons produce as much as 120 mL/minute of an ultrafiltrate of blood. Second, different segments of the nephron change the composition of the filtrate by the transfer of nearly 99% of its components (e.g., glucose, NaCl, water) from the lumen to the blood. Third, additional electrolytes (e.g., NH_4^+, K^+, HCO_3^-) are secreted from the blood into the lumen.

To perform these functions, each nephron segment, with the exception of the collecting ducts, is composed of a single epithelial cell type whose luminal or apical surface (facing the urine) and basolateral surface (facing the blood) differentially express various proteins and lipids. For example, the apical membrane often has microvilli or cilia, whereas the basolateral membrane does not. Apical polarized endocytosis and exocytosis are often important in the regulation of the number of transport proteins on the apical surface. In addition, epithelia are connected to one another by tight junctions, which confer a characteristic ionic permeability on the epithelial sheet. Transepithelial transport occurs largely through the cell, but transport through the tight junction (the *paracellular pathway*) can also be important in different segments of the tubule. For instance, sodium transport begins with entry at the luminal surface down an electrochemical gradient, whereas its exit at the basolateral surface is uphill and requires adenosine triphosphate (ATP) hydrolysis. The Na^+, K^+-ATPase is located at the basolateral surface of all epithelia, and all "active" energy-consuming transport is coupled directly or indirectly to it with the exception of H^+ transport. Each segment has a distinct composition of channels, carriers, and ATPases, and each segment is regulated by different chemical and physical sensors, so the "final urine" contains the components that must be discarded to maintain constancy of body composition.

THE KIDNEY REGULATES EXTRACELLULAR FLUID VOLUME BY REGULATING ITS SODIUM CONTENT

Filtration of 180 L/day containing 24,000 mEq of sodium is followed by the reabsorption of more than 99% of the filtered sodium. Sodium reabsorption accounts for more than 90% of the oxygen consumed by the kidney. It is regulated by volume receptors that are located in the carotid artery and increase β-sympathetic output, which in turn releases renin, an aspartate protease from the granular cells of the juxtaglomerular apparatus. The renin-releasing cells are close to the afferent arterioles, where renin cleaves angiotensinogen to angiotensin I, which is then converted locally to angiotensin II. Angiotensin II binds to angiotensin receptors and constricts the efferent arteriole, thereby affecting glomerular hemodynamics. The increased hydrostatic pressure within the glomerular capillaries drives the formation of an ultrafiltrate of plasma. As filtration progresses, a protein-rich, oncotically active solution in the capillary opposes the glomerular capillary hydrostatic pressure until a pressure equilibrium is achieved before the efferent arteriole is reached. Consequently, angiotensin II may not change glomerular filtration rate (GFR) markedly, but it can increase proximal reabsorption by reducing the hydrostatic pressure and increasing the oncotic pressure in the peritubular capillaries that surround the proximal tubule in a plexus, thereby favoring reabsorption of water and solutes such as urea.

The glomerular filtrate next enters the tubular portion of the nephron, where Na^+ traverses the cell by entering the apical membrane either through a cotransporter or countertransporter or through a sodium channel, depending on the specific mechanisms of different segments. In the apical membrane of the proximal tubule, an Na^+/H^+ (NHE3) exchanger, an Na^+-coupled glucose carrier, and an Na^+-coupled amino acid and phosphate cotransporter are present. Subsequently, Na^+ is actively transported by the basolateral Na^+, K^+-ATPase into the paracellular space, thereby resulting in local hypertonicity, which causes osmosis through low-resistance tight junctions of the initial segments of the proximal tubule (Fig. 115-3).

In the thick ascending limb of Henle, Na^+ is absorbed by an Na-K-2Cl cotransporter. The driving force for this neutral carrier allows Na^+ and Cl^- to enter the cell, but K^+ is then recycled across the apical membrane, thereby resulting in depolarization of the transepithelial membrane potential. In the distal convoluted tubule, Na^+ is absorbed by a thiazide-sensitive cotransporter, which conducts Na^+ and Cl^- in a strict 1 : 1 stoichiometry. Na^+ exits as usual by the Na^+, K^+-ATPase, but there is also a basolateral Na/Ca exchanger. In this short segment, the macula densa helps control the GFR by regulating renin release through secretion of adenosine and prostaglandins.

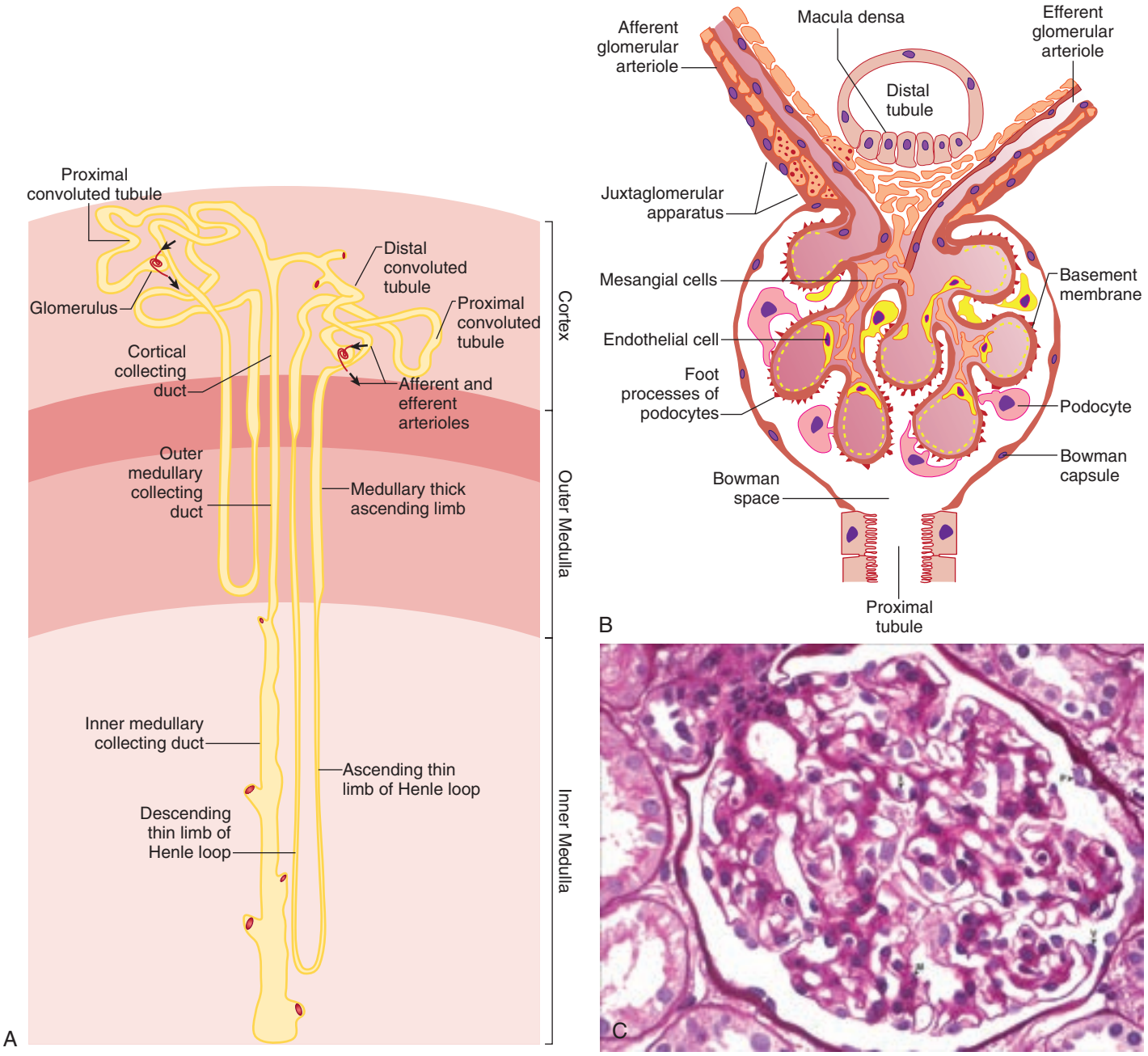

FIGURE 115-2. Structure of the nephron. A, Components of the cortical and juxtaglomerular nephrons. **B,** Anatomy of the glomerulus. **C,** Light micrograph of a human glomerulus. E = endothelial cell; M = mesangial cell; P = parietal epithelial cell; V = visceral epithelial cell. (**C** courtesy Dr. Glen Markowitz.)

In the principal cells of the collecting duct, aldosterone, derived from the zona glomerulosa of the adrenal cortex, increases reabsorption of the final 50 to 100 mEq/day of sodium remaining in the lumen by increasing the number of open sodium channels (ENac), by activating expression of α-subunits, and by increasing the activity of the basolateral Na^+, K^+-ATPase.[1] Aldosterone is the final critical regulator of sodium balance.

A number of the steps that regulate sodium reabsorption can be counteracted by atrial natriuretic peptide (ANP), which is released from the atria in response to volume overload. ANP increases filtration in the glomerulus by dilating the afferent arteriole, thereby increasing glomerular capillary pressure and lowering the oncotic pressure. In addition, ANP increases sodium excretion by inhibiting the release of renin, the production of aldosterone, and the tubular reabsorption of sodium in the terminal collecting duct.

THE KIDNEY REGULATES BODY FLUID OSMOLARITY BY REGULATING ITS WATER CONTENT

Water moves freely among all cells and compartments of the body owing to the presence of water-conducting channels called aquaporins.[2] The concentration of water (its osmolarity) is strictly regulated to prevent cells from swelling or contracting. Control of the osmolarity of the body fluids requires control of intake (through the behavioral mechanism of thirst) and its excretion, whereby the kidney can vary the osmolarity of urine. It can dilute the urine by absorbing sodium without water in the thick ascending limb, the distal tubule, and the collecting duct. Conversely, the absorbed NaCl and urea provide a hyperosmotic region in the medulla, where the limited blood flow maintains an osmotic gradient. To concentrate the urine, water is removed from the filtrate by this osmotic gradient as it passes through the cortical and medullary collecting tubules.

Antidiuretic hormone (ADH), also called *vasopressin*, is a critical component of water reclamation. The neurohypophysis releases ADH as a result of activation of TRPV1 channels, which are responsive to cell shrinkage caused by changes in osmolarity of less than 1%. In the kidney, ADH binds to the vasopressin type 2 receptor TRPV1 on collecting tubules and increases water permeability by reversibly inserting a water channel, called *aquaporin 2*, in the apical membrane by the fusion of vesicles. During states of water deprivation, the dilute urine generated by the thick ascending limb enters vasopressin-sensitive segments in the cortical collecting tubule, where the bulk of the water is absorbed into the cortex to increase the osmolarity of the urinary filtrate to 300 mOsm, which is the osmolarity of plasma. Subsequently, the urine becomes concentrated when it equilibrates with the

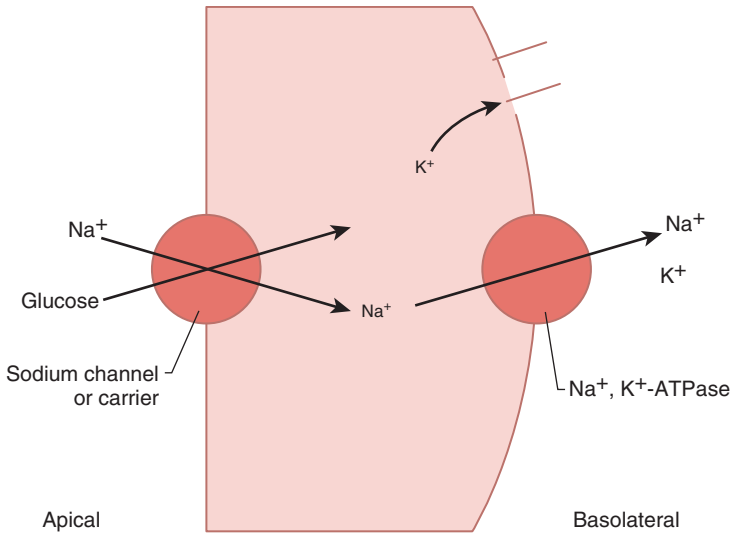

FIGURE 115-3. Sodium reabsorption in the proximal tubular cell.

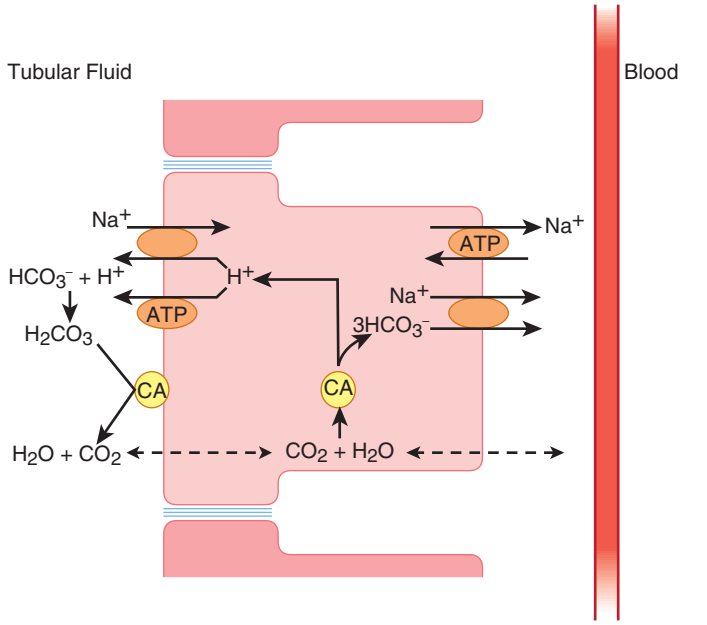

FIGURE 115-5. Ion transport in the proximal tubule. ATP = adenosine triphosphate; CA = carbonic anhydrase.

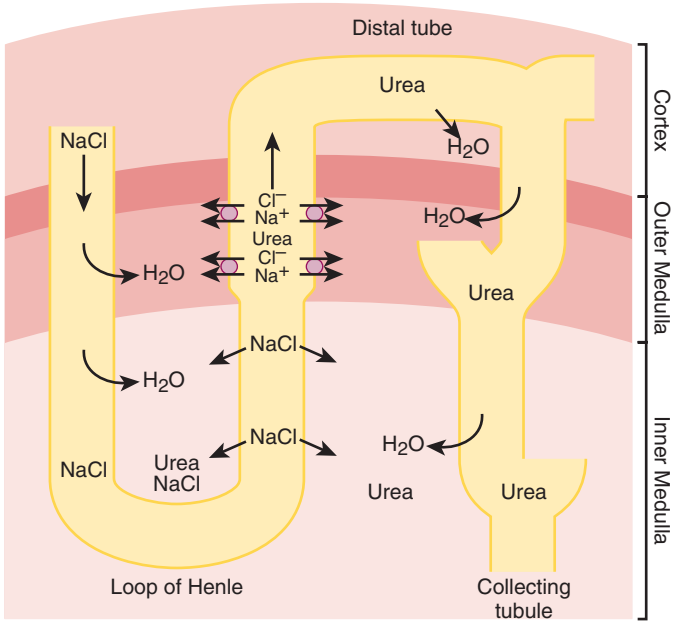

FIGURE 115-4. Regulation of water content.

osmolarity of the medulla as the urine courses down the medullary collecting duct (Fig. 115-4).

THE KIDNEY REGULATES PLASMA pH BY REGULATING HCO_3 CONTENT

The concentration of free H^+ in the intracellular and extracellular fluids is maintained at about 40 nM (pH 7.4) by the daily excretion of acid or base in amounts equal to what is generated by dietary intake and by cellular activity. The complete oxidation of carbohydrates and fats generates approximately 15 to 20 mol/day of the volatile acid CO_2, and nonvolatile acids generated by the metabolism of protein-rich diets account for 60 to 80 mEq/day. HCO_3^- is the most important buffer not only because it is consumed by metabolic acid, thereby producing CO_2 that can be exhaled, but also because it can be regenerated by the kidney. The overall relationship of this buffer system is described by the Henderson-Hasselbalch equation:

$$pH = pK_a + \log[HCO_3^-]/\alpha P_{CO_2}$$

where pK_a is 6.1 and α represents the solubility coefficient of P_{CO_2} (which is 0.03). Because all the body buffers (e.g., bone, intracellular proteins) are in equilibrium, changes in the concentration of HCO_3^- regulate the pH of body fluids.

Although one task of the kidney is to replace the HCO_3^- that is lost as a consequence of the production of acid by oxidative metabolism, it must first reabsorb the 5000 mEq/day HCO_3^- that is filtered every day. The proximal tubule reabsorbs luminal HCO_3^- by secreting H^+ through an apical Na^+/H^+ exchanger, which is directly stimulated by angiotensin II to secrete H^+ in strict 1 : 1 exchange for Na^+, thereby mediating both sodium absorption and H^+ secretion. An H^+-ATPase is also present in the microvilli and apical endocytic vesicles that fuse with the apical membrane in response to elevated blood P_{CO_2}. H^+ secretion titrates filtered HCO_3 and, catalyzed by carbonic anhydrase, converts it to CO_2 and water, thereby allowing CO_2 reabsorption. H^+ secretion leads to an excess of OH^- in the cell, where it combines with this CO_2 to produce cellular HCO_3^-, which then exits across the basolateral membrane through an Na-HCO_3^- cotransporter (Fig. 115-5).

Given that the proximal tubule regulates bicarbonate reabsorption, the collecting duct must produce "new" HCO_3 to replace what is lost during titration by nonvolatile acids. The cortical collecting tubules contain intercalated cells that mediate H^+ secretion (α-intercalated cell) or β-cell types that mediate HCO_3 secretion, whereas only α cells are present in the medulla. H^+ secretion is mediated in the α-intercalated cell by the H^+-ATPase that is delivered to the apical membrane by fusion of apical vesicles as stimulated by ambient P_{CO_2} (Fig. 115-6).

In a reaction catalyzed by carbonic anhydrase II, the excess OH^- is carboxylated by CO_2 to form HCO_3^-. HCO_3^- is subsequently transported across the basolateral surface in exchange for Cl^- by an alternately spliced form of the red cell anion exchanger (band 3). In contrast, β cells secrete HCO_3^- by an apical Cl/HCO_3 exchanger (pendrin) and a basolateral H^+-ATPase. An H^+, K^+-ATPase in the collecting tubule may also play a role in potassium absorption and perhaps in H^+ secretion.

The collecting tubule is a "tight epithelium" that can maintain electrical and concentration gradients. Secretion of H^+ reduces the pH of the filtrate; the maximal gradient is 3 pH units or 180 mV, but both the size of the pH gradient and its transepithelial membrane potential can be modified to regulate H^+ secretion. For example, Na^+ absorption by the principal cell hyperpolarizes the epithelium and can thus drive H^+ secretion. Aldosterone can stimulate not only Na^+ absorption but also independently H^+ secretion; hence, it is the major hormone that stimulates acid secretion in the collecting tubule. Finally, chronic metabolic acidosis converts the β-intercalated cells into α-intercalated cells, thereby increasing the number of acid-secreting cells and reducing the amount of HCO_3^- secretion in this segment.[3]

The secreted H^+ titrates urinary NH_3 and HPO_4^{2-}. NH_3, which is synthesized by the conversion of glutamine to α-ketoglutarate, is secreted into the proximal lumen. In the loop of Henle and distal segments, the protonated ammonium ion is transferred into the interstitium, from where it reenters the nephron as ammonia gas. Once in the lumen, ammonia gas becomes reprotonated to ammonium, thus trapping a proton. The amount of NH_3 generated from glutamine increases up to four- to five-fold in the setting of metabolic acidosis. Net acid secretion is consequently urinary NH_4^+ plus titratable weak

acids (such as HPO_4^{2-}) minus urinary HCO_3^-. Each of these components is regulated by the kidney.

THE KIDNEY REGULATES PLASMA POTASSIUM BY EXCRETION AND THE CONTROL OF EXTRACELLULAR pH

Because the membrane potential of most cells is governed by the ratio of intracellular to extracellular potassium, the plasma potassium concentration must be tightly regulated. Most dietary potassium is pumped into the cell by the Na^+, K^+-ATPase, but about 90% of ingested potassium eventually must be excreted into the urine. K^+ is filtered but is then reabsorbed by the proximal tubule and loop of Henle. K^+ is then secreted by the distal convoluted tubule and the principal cells through apical potassium inwardly rectifying (ROMK) and potassium large conductance calcium-activated (BK) channels. K^+ is secreted in response to a gradient, with cell K^+ greater than lumen K^+, which drives K^+ into the lumen. Secretion is enhanced by the increased entry of K^+ into the principal cell as a result of high extracellular concentration, by metabolic alkalosis, and by aldosterone. Aldosterone increases the synthesis of Na^+, K^+-ATPase and increases the probability that the ROMK channels will be open. Aldosterone also enhances Na^+ reabsorption by ENaC, thereby hyperpolarizing the transepithelial membrane potential, which increases the driving force for potassium secretion. High urinary flow rates deliver a large volume of fluid that is essentially K^+ free, thereby providing a concentration gradient. However, the flow rate also activates mechanosensory flagella that lead to an increase in cell calcium, which in turn activates BK channels. Potassium reabsorption is not regulated to the same extent as sodium retention, and profound potassium depletion may be required for potassium excretion by the kidneys to be eliminated completely.

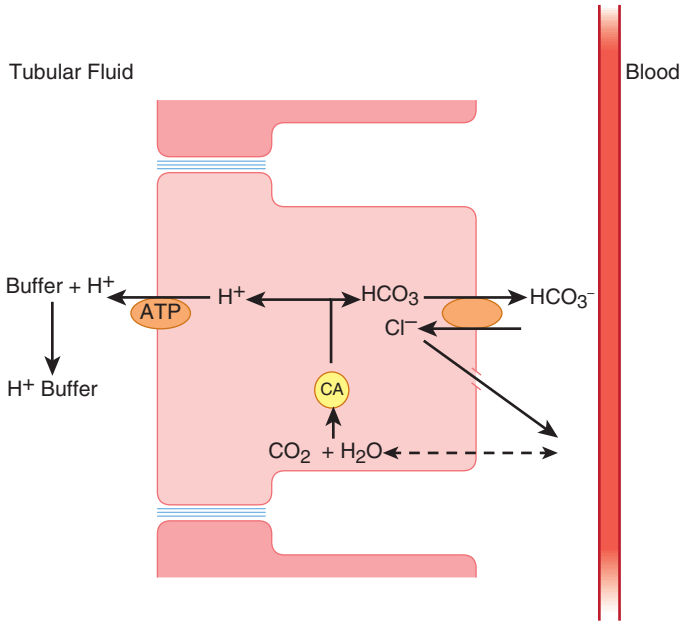

FIGURE 115-6. Mechanism of acid secretion in the collecting duct. ATP = adenosine triphosphate; CA = carbonic anhydrase.

THE KIDNEY REGULATES PLASMA PO_4 AND Ca^{2+} BY EXCRETION AND BY SYNTHESIZING VITAMIN D_3

The level of PO_4 critically regulates serum Ca^{2+} because their plasma concentrations are close to their saturation product, at which point crystallization occurs. Thus, any significant increase in PO_4 results in precipitation of $CaPO_4$.

The level of PO_4 is regulated by glomerular filtration. Initially, about 85% is reabsorbed by the proximal tubule sodium-phosphate cotransporters NaPi-2a and NaPi-2c. Parathyroid hormone and the heteromeric receptor FGFR/Klotho, which binds fibroblast growth factor 23 (FGF23), inhibit the expression of NaPi and hence increase excretion of PO_4. Conversely, the proximal tubule generates vitamin D_3 (1,25-dihydroxycholecalciferol) by capturing 25-hydroxycholecalciferol, bound to its filterable transport protein with brush border megalin (Fig. 115-7), after which 1α-hydroxylation is stimulated by low Ca^{2+} and PO_4. 1,25-Vitamin D_3 inhibits parathyroid hormone and stimulates PO_4 reabsorption in gut and kidney, thereby counteracting parathyroid hormone and FGF23-Klotho. Nonetheless, Klotho signaling blocks 1,25-vitamin D_3 synthesis, thereby suggesting that it dominates the control of PO_4.[4]

The level of serum calcium is also regulated by the kidney. Approximately 60% of serum calcium is filtered, after which it follows sodium reabsorption in the proximal tubule and in the loops of Henle, where calcium absorption is driven by the positive membrane potential generated by the Na-K-2Cl/K recycling transporter. In contrast, calcium and sodium are regulated independently in the distal convoluted tubule by parathyroid hormone and Klotho, which increase reabsorption. Phosphate and calcium metabolism is regulated by sodium reclamation, parathyroid hormone, 1,25-dihydroxycholecalciferol, and Klotho in different segments of the nephron.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

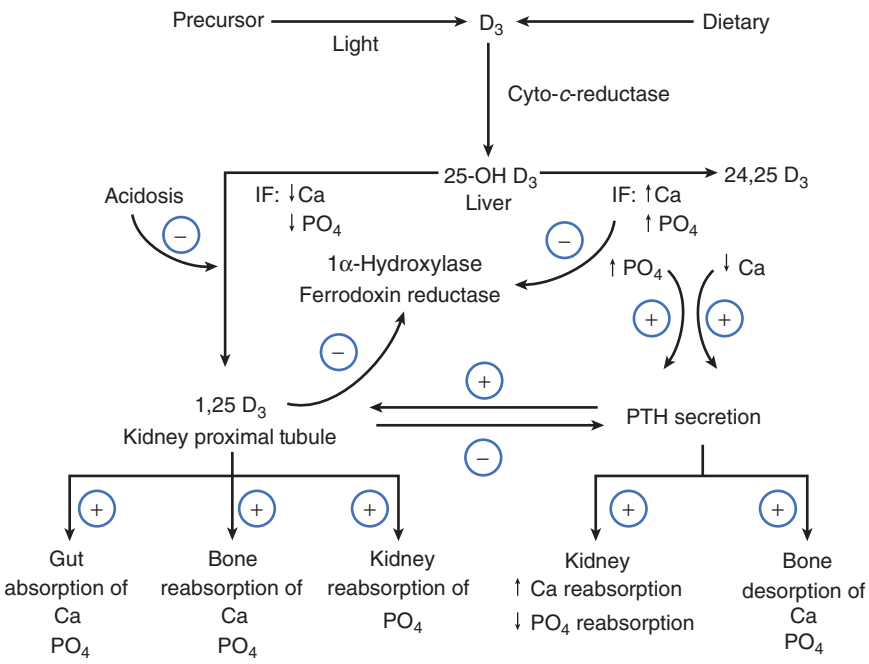

FIGURE 115-7. Calcium and phosphate metabolism. PTH = parathyroid hormone.

GENERAL REFERENCES

1. Hamm LL, Feng Z, Hering-Smith KS. Regulation of sodium transport by ENaC in the kidney. *Curr Opin Nephrol Hypertens*. 2010;19:98-105.
2. Kortenoeven ML, Fenton RA. Renal aquaporins and water balance disorders. *Biochim Biophys Acta*. 2014;1840:1533-1549.
3. Al-Awqati Q, Gao XB. Differentiation of intercalated cells in the kidney. *Physiology (Bethesda)*. 2011;26:266-272.
4. Wolf M, White KE. Coupling fibroblast growth factor 23 production and cleavage: iron deficiency, rickets, and kidney disease. *Curr Opin Nephrol Hypertens*. 2014;23:411-419.

116 DISORDERS OF SODIUM AND WATER HOMEOSTASIS

ITZCHAK SLOTKI AND KARL SKORECKI

SODIUM AND WATER HOMEOSTASIS

EPIDEMIOLOGY

Disturbances in sodium and water balance or distribution, with attendant perturbations in the volume or solute composition of body fluid compartments, are among the most frequently encountered abnormalities in clinical medicine. The principal manifestations of these disturbances are *hypovolemia*, *hypervolemia*, *dysnatremia* (*hyponatremia* or *hypernatremia*), and *polyuria*. Disturbances in body tonicity, reflected by hyponatremia or hypernatremia, are estimated to affect up to 15 to 20% of hospitalized patients, with severe disturbances (>10% deviation from normal values) affecting 1 to 2% of patients. Prevalence rates for these abnormalities in ambulatory populations are lower, and elderly individuals and patients treated with multiple medications are the most susceptible.

PATHOBIOLOGY

Approximately 60% of body mass is composed of solute-containing fluid solutions that are divided into extracellular fluid (ECF) and intracellular fluid (ICF) compartments. Water flows freely across cell membranes through specific water channels (aquaporin family of transmembrane proteins) according to the dictates of osmotic forces, thereby maintaining near equality of the solute-to-water ratio (osmolality) in the ICF and ECF. However, the composition of the major solutes differs between the ECF and ICF (Fig. 116-1A). Sodium and potassium are the major cations in the ECF and ICF, respectively. Chloride and bicarbonate are the major accompanying anions in the ECF, and negative charges on organic molecules maintain electroneutrality with potassium in the ICF. The difference in cationic solute composition between these two compartments is maintained by a pump-leak mechanism involving the activity of Na^+, K^+-adenosine triphosphatase (ATPase) operating in concert with cell membrane sodium and potassium conductance pathways. The free movement of water ensures that the sodium concentration in ECF is nearly equivalent to the potassium concentration in ICF. The magnitude of free water movement is determined by the *tonicity*, which refers to the concentration of solutes that are "effective" in eliciting a water shift between body fluid compartments. Tonicity should be distinguished from *osmolality*, which refers to the concentration of all solutes, some of which permeate freely and equilibrate across most cell membranes. Molecules in this category include urea and glucose. However, urea and glucose do contribute to the laboratory measurement of fluid osmolality. Addition or removal of effective solutes causes a sustained shift of water to restore the near equality of concentrations. The restriction of sodium to the ECF compartment by virtue of the pump-leak mechanism, together with maintenance of osmotic equilibrium between the ECF and ICF, ensures that ECF volume is determined principally by the total body fluid sodium content, which governs the partitioning of fluid between the ECF and ICF compartments. The addition or removal of water without solutes results in a proportionate reduction or increase, respectively, in both osmolality and tonicity of all body fluid compartments.

The mechanisms that govern body fluid homeostasis preserve near constancy of the volumes of the ECF and ICF compartments despite variations in dietary intake and extrarenal losses of sodium and water and adjust this balance in response to variations in the capacity of these compartments. Thus, the overriding principle of body fluid and solute homeostasis is mass balance of total body intake and output of water as well as balance of osmotically active particles. Even the slightest perturbations in these parameters activate neural and hormonal mediators for restoration of balance. This restoration of balance is achieved through adjustments in the urinary excretion of sodium and water in response to perceived changes in ECF or ICF volume. Constancy of ECF volume together with control of vascular capacitance ensures a high degree of circulatory stability.

Intracellular Water (2/3)		Extracellular Water (1/3)	
		Interstitial (2/3)	Blood (1/3)
25	Na	140	
150	K	4.5	
15	Mg	1.2	
0.01	Ca	2.4	
2	Cl	100	
6	HCO_3	25	
50	Phos	1.2	

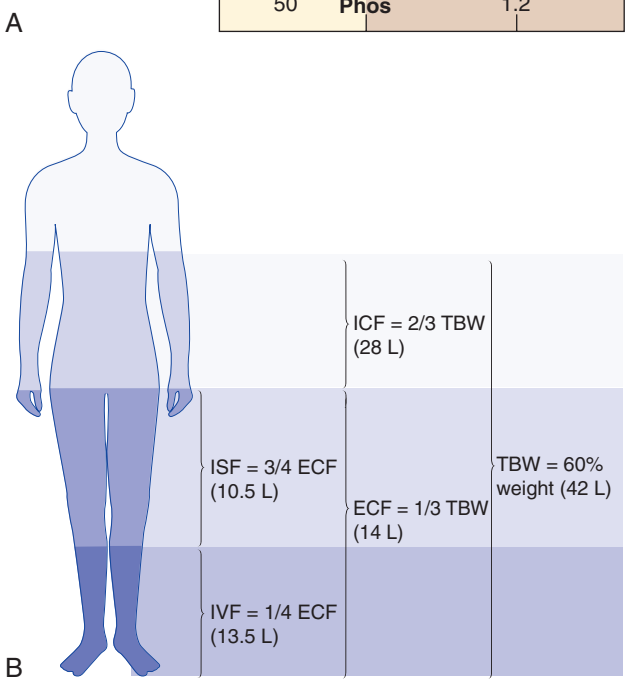

FIGURE 116-1. Composition of body fluid compartments. Schematic representation of (**A**) electrolyte composition of compartments in humans and (**B**) body fluid compartments. In **A**, electrolyte concentrations are in millimoles per liter; intracellular concentrations are typical values obtained from muscle. In **B**, shaded areas depict the approximate size of each compartment as a function of body weight. In a normally built individual, the total body water content is roughly 60% of body weight. Because adipose tissue has a low concentration of water, the relative water–to–total body weight ratio is lower in obese individuals. Relative volumes of each compartment are shown as fractions; in parentheses are shown approximate absolute volumes of the compartments (in liters) in a 70-kg adult. ECF = extracellular fluid; ICF = intracellular fluid; ISF = interstitial fluid; IVF = intravascular fluid; TBW = total body water. (**B** from Verbalis JG. Body water osmolality. In: Wilkinson B, Jamison R, eds. *Textbook of Nephrology*. London: Chapman & Hall; 1997:89-94. Reproduced with permission of Hodder Arnold.)

Sodium Balance

Sodium balance refers to the difference between intake and excretion. In the nonclinical setting, sodium intake is controlled by dietary habits. In the clinical setting, prescribed adjustments in sodium intake and the administration of sodium-containing medications or solutions cause variation to overall sodium intake. Although nonrenal loss is under some regulatory influence (e.g., aldosterone-mediated regulation of sodium concentration in stool and sweat), the fine adjustment of sodium balance in response to changes in intake is mediated by regulation of urinary sodium excretion. In the steady state, urinary excretion of sodium is closely matched to dietary salt intake. This balance depends on a series of afferent mechanisms that sense the volume of the ECF compartment relative to its capacitance and trigger effector mechanisms that modify the rate of renal sodium excretion to maintain ECF volume homeostasis. Normal functioning of these mechanisms in turn is affected by the partitioning of the ECF into two subcompartments: intravascular and extravascular (interstitial) (Fig. 116-1B). The composition and concentration of small, noncolloid electrolyte solutes in these two ECF subcompartments are nearly equivalent, but there is a higher concentration of colloid osmotic particles, mostly molecules of albumin and globulin proteins, in the intravascular compartment. Opposing transcapillary hydraulic and colloid osmotic (oncotic) pressure gradients (Starling forces) favor the net transudation of fluid from the intravascular to the interstitial compartment. At the same time, lymphatic fluid movement from interstitial sites back to the

circulation through the thoracic duct ensures that the intravascular subcompartment is replenished and maintains a nearly constant proportion of approximately 25% of the overall ECF, which corresponds to 3.5 L of plasma. The remaining approximately 75% of ECF volume (equivalent to 10.5 L in a normal 70-kg man) is contained in the interstitial space. Because intravascular volume is one of the key determinants of circulatory integrity, preservation of the constancy of ECF volume and appropriate partitioning of the ECF volume between the intravascular and interstitial subcompartments are critical for hemodynamic stability.

A hypertonic interstitial fluid compartment also exists in the skin, where it is associated with a macrophage-mediated increased vasoreactivity in precapillary arterioles, the major resistance vessel of the skin (Chapter 435). Although not directly affecting ECF volume responses, the resulting increase in peripheral resistance could contribute to higher blood pressure in salt-sensitive hypertension.[1]

Effective Arterial Blood Volume

The circulatory network is composed of central and peripheral venous compartments as well as of renal and extrarenal arterial compartments. Each compartment reflects a unique characteristic of overall circulatory function (e.g., cardiac filling, tissue perfusion, renal perfusion, and transudation of fluid into the interstitial space).

The concept of effective arterial blood volume (EABV) is crucial to understanding of the afferent mechanisms that govern the regulation of sodium homeostasis. Unlike ICF, ECF, and intravascular volume, EABV is not measurable as an anatomically defined space. Rather, EABV is best understood in functional terms as an integration of hemodynamic parameters emanating from specific sites in the arterial circuit that monitor tissue perfusion and trigger appropriate changes in urinary sodium excretion. These sites include the carotid baroreceptor and intrarenal mechanisms located at the glomerular afferent arterioles, the juxtaglomerular apparatus, and the peritubular capillaries. EABV often but not always varies directly with actual ECF volume. Low-pressure sensors (e.g., cardiac atrial transmural stretch and tension receptors) respond to the state of cardiac filling and tend to protect against overfilling of the ECF compartment, but they also have a role in renal sodium retention in states of perceived underfilling (Table 116-1). Taken together, the integrated ECF volume–sensing signals elicit an appropriate renal response for modulating sodium excretion in an effort to maintain a constant ECF volume.

The filtered load of sodium vastly exceeds net intake, so tubular reabsorption usually serves as the principal modulator of urinary sodium excretion, the preservation of sodium balance, and the maintenance of a constant ECF volume. Specific luminal membrane sodium transporters or channels at each tubular segment mediate movement of sodium from the luminal fluid into the cell in a carefully regulated manner, followed by extrusion of sodium across the basolateral surface through Na^+, K^+-ATPase and other sodium transporters. Among many others, these regulated luminal transporters include the Na/H exchanger (proximal tubule), the Na-K-2Cl cotransport pathway (loop of Henle), the NaCl cotransporter along the distal convoluted tubule, and the epithelial sodium channel along the connecting and cortical collecting tubule. At some nephron sites, sodium reabsorption is isotonic (e.g., proximal tubule), whereas at other sites, sodium reabsorption exceeds water reabsorption, so tubular fluid has a sodium concentration less than that of plasma (e.g., thick ascending limb of the loop of Henle). Sodium reabsorptive transport pathways are subject to a series of regulatory influences that sense EABV, including the renin-angiotensin-aldosterone pathway, natriuretic peptides, endothelium-derived endothelins and nitric oxide, the eicosanoid-prostaglandin system, guanylin peptides of gut origin, and urotensins. This redundancy of multiple hormonal mediators, which act together with the sympathetic nervous system, renal neural stimulation, and intrarenal physical factors (e.g., peritubular capillary Starling forces, tubule lumen sodium chloride delivery, and tubuloglomerular feedback), underscores the evolutionary importance of regulating urinary sodium excretion to maintain volume homeostasis.

TABLE 116-1 MECHANISMS FOR SENSING REGIONAL CHANGES IN BODY FLUID VOLUME

Cardiopulmonary volume sensors
Atria (neural and humoral pathways)
Ventricular and pulmonary sensing sites
Arterial volume sensors
Carotid and aortic arch baroreceptors
Renal volume sensors
Central nervous system sensors
Hepatic and gastrointestinal tract sensors

Water Balance

Water balance refers to the difference between intake (oral, enteral, or parenteral) and excretion (insensible, gastrointestinal, perspiratory, and renal). Maintaining equivalency of intake and excretion of water ensures constancy of body fluid tonicity (osmoregulation). Osmoregulation ensures that the content of effective solutes in each body fluid compartment determines the volume of that compartment. Positive water balance or negative water balance and the corresponding changes in tonicity and cell volume are sensed by osmoreceptor and thirst center cells in the hypothalamus. The osmoreceptors are situated in the supraoptic and paraventricular nuclei of the hypothalamus; the thirst center is situated in the organum vasculosum of the anterior hypothalamus. Just a 2% change in effective osmolality or tonicity elicits a change in release of the hormone arginine vasopressin (AVP) from the posterior pituitary gland and the perception of thirst (Fig. 116-2A). Endothelin-1 also is released from the posterior pituitary in response to water deprivation and increases plasma AVP levels. Stimulation of thirst depends on centrally produced angiotensin II. A reduction of more than approximately 8% of ECF volume serves as an overriding afferent signal (carried by the ninth and tenth cranial nerves) for the nonosmotically driven release of AVP and also stimulates thirst by means of angiotensin II, even when body tonicity is not elevated (Fig. 116-2B).

The urinary excretion of water depends on the delivery of isotonic sodium-containing filtrate to the thick ascending limb of the loop of Henle, reabsorption of sodium and accompanying electrolytes in the thick ascending limb of the loop of Henle and the distal tubule, and the AVP-regulated reabsorption of the appropriate volumes of solute-free water through aquaporin 2 water channels in the cells of the collecting tubule. The hydro-osmotic movement of water from the collecting tubule lumen to the hyperosmotic milieu of the renal medulla minimizes urinary excretion of the hypotonic fluid generated in the thick ascending limb of the loop of Henle and the distal tubule, thereby promoting positive water balance. In the absence of AVP, this hypotonic fluid is excreted, and negative water balance ensues. Typical levels of urine osmolality that can be achieved in the human kidney by fluctuations in AVP action range between 50 and 1200 mOsm/kg, but this range narrows at the extremes of age or in the presence of intrinsic renal disease.

PATHOPHYSIOLOGY

Disturbances in sodium balance primarily affect ECF volume, and disturbances in water balance primarily affect body fluid tonicity. A cumulative negative balance of sodium (sodium deficit) in the absence of a change in tonicity results in ECF volume contraction (hypovolemia), whereas a cumulative positive balance of sodium (sodium surfeit) results in ECF volume expansion (hypervolemia). In contrast, a cumulative positive body water balance (water surfeit) results in volume expansion of all the body fluid compartments, whereas negative water balance (water deficit) results in volume contraction of all the body fluid compartments. However, because ICF volume is double that of ECF at baseline, the more prominent expansion or contraction in absolute volume terms involves the ICF compartment. Furthermore, maintenance of a surfeit or deficit of water relative to solutes results in a uniform decrease or increase in body fluid tonicity in all fluid compartments. Because the solute composition of the plasma component of the ECF compartment is sampled, a disturbance in tonicity is most commonly detected as an abnormality in plasma sodium concentration (hyponatremia for water surfeit and hypernatremia for water deficit). Disturbances in water and sodium balance frequently occur together, and all combinations of surfeit or deficit can occur (Table 116-2). The clinical approach to a patient with a sodium or water balance disturbance (or both) can be facilitated by careful consideration of which states apply.

SODIUM BALANCE DISORDERS

Hypovolemia

DEFINITION

Hypovolemia is a reduction in the volume of the ECF compartment in relation to its capacitance. In states of absolute hypovolemia, a deficit in sodium

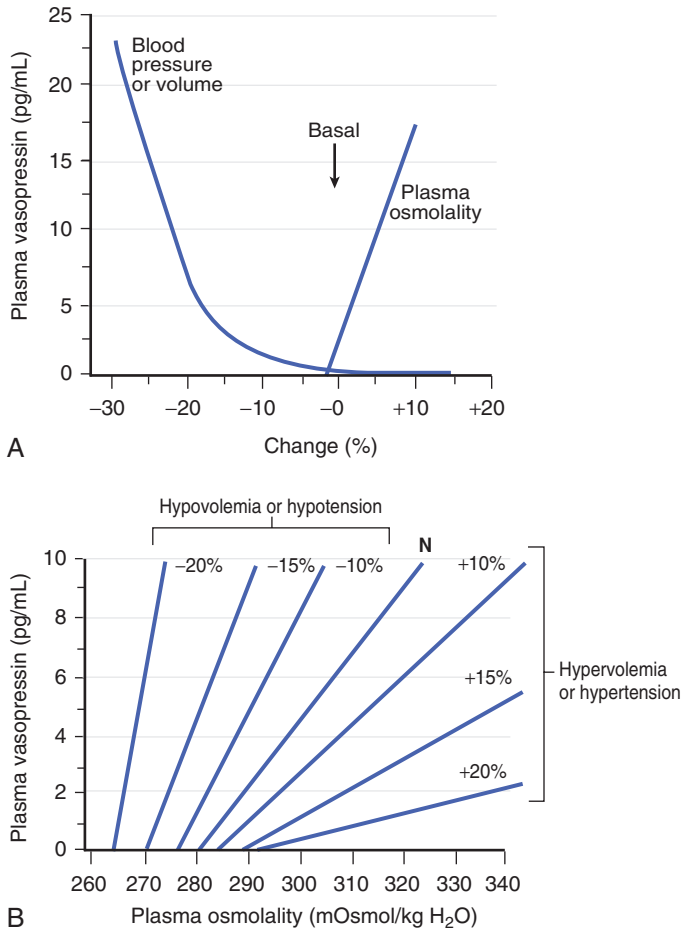

FIGURE 116-2. Relationship between arginine vasopressin (AVP) and osmolality. **A,** Comparative sensitivity of AVP secretion in response to increases in plasma osmolality versus decreases in blood volume or blood pressure in human subjects. The *arrow* indicates the low plasma AVP concentrations found at basal plasma osmolality. Note that AVP secretion is much more sensitive to small changes in blood osmolality than to changes in volume or pressure. **B,** The relationship between the osmolality of plasma and the concentration of AVP in plasma is modulated by blood volume and pressure. The line labeled N shows plasma AVP concentrations across a range of plasma osmolality in an adult with normal intravascular volume (euvolemic) and normal blood pressure (normotensive). The lines to the left of N show the relationship between plasma AVP concentration and plasma osmolality in adults whose low intravascular volume (hypovolemia) or blood pressure (hypotension) is 10%, 15%, and 20% below normal. Lines to the right of N are for volumes and blood pressure 10%, 15%, and 20% above normal. Note that hemodynamic influences do not disrupt the osmoregulation of AVP but rather raise or lower the set point and possibly also the sensitivity of AVP secretion in proportion to the magnitude of the change in blood volume or pressure.

TABLE 116-2 PATHOGENIC PROCESSES LEADING TO DISORDERS OF BODY SODIUM AND FLUID HOMEOSTASIS

CLINICAL STATE	EXTRACELLULAR FLUID VOLUME	BODY FLUID TONICITY	PATHOGENIC PROCESS
Normal	↔	↔	
Hypovolemic hypernatremia	↓	↑	Net loss of water in excess of sodium
Hypovolemic normonatremia	↓	↔	Isotonic net loss of sodium and water
Hypovolemic hyponatremia	↓	↓	Net loss of sodium in excess of water
Normovolemic hyponatremia	↔	↓	Net water gain ± sodium loss
Normovolemic hypernatremia	↔	↑	Net water loss ± sodium gain
Hypervolemic normonatremia	↑	↔	Isotonic net gain of sodium and water
Hypervolemic hyponatremia	↑	↓	Hypotonic net gain of sodium and water
Hypervolemic hypernatremia	↑	↑	Hypertonic net gain of sodium and water

↔ = unchanged.

TABLE 116-3 CAUSES OF ABSOLUTE AND RELATIVE HYPOVOLEMIA

EXTRARENAL
Absolute
Bleeding
Gastrointestinal fluid loss (diarrhea, vomiting, ileostomy or colostomy secretions)
Fluid loss from skin (burns, sweat)
Respiratory fluid loss
Relative
Third space loss
Sepsis
RENAL
Absolute
Diuretics
Inherited sodium-wasting tubulopathies
Tubulointerstitial diseases
Partial obstruction or postobstruction etiology
Endocrine disorders (e.g., hypoaldosteronism, adrenal insufficiencies)
Relative
Nephrotic syndrome

reflects past or ongoing negative sodium balance. The volume of the ECF intravascular and extravascular (interstitial) subcompartments may vary in the same or opposite directions. ICF volume is reflected by the measurement of plasma osmolality and sodium concentration and may be concomitantly disturbed (see Table 116-2).

EPIDEMIOLOGY

Causes of absolute and relative hypovolemia can be categorized into extrarenal and renal causes (Table 116-3). Gastrointestinal fluid loss and massive bleeding are the most frequent and direct causes of a reduction in the absolute volume of the intravascular subcompartment of the ECF (Chapters 106 and 135). Infectious diarrhea remains a leading cause of death from hypovolemia in many areas of the world. Another extrarenal cause of absolute hypovolemia is fluid loss from the integumentary and respiratory systems. Burns (Chapter 111), which allow the loss of large volumes of plasma and interstitial fluid, can rapidly lead to profound hypovolemia similar to what is seen with bleeding (Chapter 135). Enhanced evaporative water loss from the respiratory tract can occur with exercise, in response to heat stress (Chapter 109), in febrile states, and in patients undergoing mechanical ventilation with inadequate humidification.

In *relative hypovolemia*, ECF volume is not reduced in absolute terms but the capacitance of the ECF or intravascular compartment is expanded, thereby leading to clinical manifestations that mimic those of absolute hypovolemia. Relative hypovolemia can be classified into two categories: states of vasodilation and states of generalized edema. Vasodilation occurs in response to endogenous endothelial substances (e.g., nitric oxide) or exogenous agents (such as vasodilator drugs). Peripheral vasodilation occurs in sepsis (Chapter 108) and normal pregnancy (Chapter 239). "Third space" loss refers to states in which ECF is sequestered into compartments within the body without an evident history of fluid loss. Included in this category are gastrointestinal obstruction, sequestration of fluids in subcutaneous tissue after trauma or burns (Chapter 111), and sequestration in the retroperitoneal or peritoneal space in patients with pancreatitis (Chapter 144) or peritonitis (Chapter 142), respectively.

Renal causes of absolute hypovolemia include any situation in which tubular reabsorption of the filtered sodium load does not match the sum of this filtered load plus dietary intake, thereby leading to a renal sodium-wasting state. Most of the widely used diuretic (or, more appropriately, natriuretic) medications inhibit specific pathways for sodium reabsorption at various sites along the nephron. Loop diuretics (furosemide, bumetanide, torsemide, and ethacrynic acid) are the most potent, and their potential to induce renal sodium loss is augmented when they are combined with other classes of diuretic agents (thiazides, carbonic anhydrase inhibitors,

aldosterone antagonists, and distal epithelial sodium-channel blockers). Tubular sodium reabsorption also may be disrupted by inherited or acquired renal tubulopathies, such as various forms of proximal tubulopathy and different forms of Bartter and Gitelman syndromes (Chapter 128).

Impairment of renal tubule sodium reabsorption may also occur with nonoliguric acute kidney injury and the recovery phase after oliguric acute kidney injury or after release of urinary obstruction (Chapters 120 and 123). Interstitial renal disease (Chapter 122) may result in fluid and electrolyte abnormalities, which can include renal sodium wasting. Nonelectrolyte urinary solutes such as glucose (in severe hyperglycemia) and mannitol cause polyuria and hypovolemia. Mineralocorticoid deficiency and resistance, including adrenal insufficiency (Chapter 227), should always be considered in a patient with evidence of urinary sodium loss in the face of hypovolemia. In cerebral salt wasting (also referred to as renal salt wasting), tubular sodium reabsorption is impaired in the setting of acute head injury (Chapter 399) or intracranial hemorrhage.

PATHOBIOLOGY

Extrarenal Causes of Absolute Hypovolemia

The immediate effect of bleeding is a proportionate net loss in plasma and erythrocytes and hence an isotonic reduction in ECF volume. Compensatory hemodynamic responses include tachycardia and vasoconstriction, followed by a shift of fluid from the interstitial to the intravascular compartment because of altered transcapillary Starling hydraulic forces. Additional neural and hormonal responses result in renal sodium and water retention, with the aim of restoring intravascular volume and stabilizing the circulation.

Besides dietary intake, approximately 7 L of isotonic fluid enters the gastrointestinal tract on a daily basis, the bulk of which is reabsorbed to minimize fecal fluid loss. Because this fluid contains varying concentrations of sodium and accompanying anions, enhanced secretion or impaired reabsorption causes absolute ECF volume depletion. The composition of the fluid and electrolyte loss differs according to the cause and source of gastrointestinal fluid loss (Chapters 132 and 140).

Given its large surface area, it is not surprising that fluid losses from the integumentary system can be an important cause of hypovolemia. In the absence of medical intervention, hemoconcentration and hypoalbuminemia ensue; however, because of the isotonic composition of the lost fluid, no changes in plasma osmolality or sodium concentration are expected. Exertion in hot environments increases thermoregulatory fluid losses from the skin in the form of sweat. Fluid loss through perspiration can reach 1 L/hour or more, depending on exertion and environment. The sodium concentration in sweat varies among individuals (range, approximately 20 to 50 mmol/L). Thus, although the fluid loss is hypotonic, a significant sodium deficit and ECF volume contraction can ensue. The extent of hypovolemia and the resulting body fluid composition (plasma osmolality and sodium concentration) will depend on fluid replacement, which is determined primarily by thirst and availability.

All extrarenal causes of hypovolemia are expected to invoke a renal response, whose hallmark is sodium and fluid conservation. Obviously, this expected response is absent when the kidney itself is responsible for sodium loss, whether it is due to the effect of pharmacologic or hormonal influences or to intrinsic renal disease.

Renal Causes of Absolute Hypovolemia

When the glomerular filtration rate (GFR) and plasma sodium concentration are normal, approximately 24,000 mmol of sodium is filtered per day. Even when the GFR is markedly impaired, the quantities of sodium filtered far exceed normal dietary intake. Thus, the small quantities of sodium excreted in urine relative to the filtered load depend on the integrity of tubular sodium reabsorptive mechanisms to match urinary sodium excretion to dietary intake through volume sensing and effector mechanisms. Impairment in the integrity of one or more of these sodium reabsorptive mechanisms can result in a profound sodium deficit and absolute volume depletion.

An underappreciated but frequent clinical setting for renal sodium loss occurs after the administration of high volumes of volume-expanding, salt-containing solutions to hospitalized patients. Such patients are usually in the postoperative or post-trauma setting and may be administered many liters of saline or other sodium-containing maintenance intravenous fluids for several consecutive days, during which tubular reabsorption of sodium is downregulated. There may be a lag in the restoration of full tubular reabsorptive capacity when intravenous fluids are discontinued, and high volumes of urine rich in sodium continue to be excreted. During this lag phase, the patient may become mildly but transiently hypovolemic. This scenario can be avoided by a graduated reduction in administered sodium-containing fluids, at a pace that allows sodium reabsorptive tubular pathways to be upregulated and restored to their normal reabsorptive levels.

Diabetes insipidus represents a spectrum of conditions resulting from deficiency or tubular resistance to the action of AVP. However, because it is the tubular reabsorption of water and not of solutes that is impaired, the impact on ECF volume is generally minor in comparison to the impact on ICF and body fluid tonicity.

CLINICAL MANIFESTATIONS

The clinical manifestations of hypovolemia depend on the magnitude and rate of volume loss, the solute composition of the net fluid loss (taking into account ingested or administered fluids), and the vascular and renal responses. A detailed history will usually reveal underlying vomiting, diarrhea, bleeding, polyuria, medications, and diaphoresis.

However, the absence of symptoms does not exclude mild to moderate hypovolemia, especially if the volume loss has occurred gradually. Intravascular volume contraction of less than 5% does not usually elicit symptoms and readily escapes detection by physical examination. With greater degrees of absolute hypovolemia (corresponding to intravascular volume contraction in the range of 5 to 15%), symptoms and signs begin to appear. Patients may exhibit nonspecific symptoms related to end-organ hypoperfusion, including weakness, muscle cramps, and postural lightheadedness. Thirst may be an early manifestation but more likely reflects a concomitant hypertonic state.

A number of clinical scenarios illustrate the relationship of clinical manifestations to causes of hypovolemia. For example, a patient with an acute *gastrointestinal hemorrhage* (Chapter 135) of 0.5 L of whole blood can experience tachycardia, postural hypotension, peripheral vasoconstriction with cool extremities, lightheadedness, and oliguria, with high urine osmolality and low urine sodium concentration. Hemoglobin and albumin will probably remain constant initially, and then a drop in hemoglobin will ensue after movement of ECF from the interstitial to the intravascular compartment. The plasma solute composition (sodium and potassium concentrations, acid-base parameters) is not likely to change initially. The plasma concentration of urea may rise somewhat as a result of increased proximal tubular reabsorption of urea and the increased nitrogen load from the destruction of erythrocytes in the gastrointestinal tract. Jugular venous pressure is expected to fall, and central venous pressure (CVP) will generally be less than 5 cm H_2O in the absence of confounding factors. Urine output is expected to be low, with a high specific gravity and osmolality and a low urine sodium concentration.

In another scenario, a patient inappropriately receiving potent *loop diuretics* during a period of many days for localized peripheral edema may suffer a cumulative net ECF loss of 3 L, or about 20%; approximately one third of this lost volume will have originated from the intravascular compartment, with the remainder coming from the interstitial compartment. This degree of intravascular loss usually induces weakness, tachycardia, low jugular venous pressure, and hypotension. However, because the deficit may have accumulated over time, a degree of adaptation would attenuate the severity of these clinical manifestations and lead the clinician to underestimate the extent of hypovolemia. The plasma sodium concentration may remain unaltered because disruption of the urine-concentrating mechanism by loop diuretics attenuates the tendency to water retention, and an intact thirst mechanism prevents hypertonicity. Because loop diuretics enhance urinary potassium and ammonium excretion, hypokalemic metabolic alkalosis is expected. The ongoing effect of the loop diuretic would mitigate oliguria and lead to inappropriately high concentrations of solutes including sodium and potassium in urine. After cessation of the loop diuretic, the appropriate renal response of oliguria, high urine osmolality, and sodium concentration would be expected.

A third scenario is a patient with relative hypovolemia due to the vasodilation that typically accompanies *sepsis* (Chapter 108). No source of fluid loss would be identified in the history, but the patient would usually manifest symptoms of weakness and even prostration, accompanied by tachycardia and hypotension. The extremities might be warm, but reduced tissue perfusion could reduce the level of consciousness and cause oliguria, elevated plasma urea and creatinine levels, and lactic acidosis.

DIAGNOSIS

The most readily appreciated physical findings related to contraction of the intravascular compartment include tachycardia, orthostatic hypotension, and

reduced jugular venous pressure. Although low CVP often reliably reflects intravascular volume contraction, an elevated CVP does not necessarily exclude hypovolemia because of the possible confounding influence of cardiac or lung disease. Severe degrees of hypovolemia (corresponding to intravascular volume contraction exceeding 10 to 20%) cause hypotension (even in the supine position), peripheral cyanosis, cold extremities, and reduced levels of consciousness (extending even to coma) as a result of end-organ and cerebral hypoperfusion. Hemodynamic collapse (hypovolemic shock; Chapter 106) also can occur with more rapid volume loss, comorbid conditions, and greater degrees of hypovolemia. When the source of volume loss is purely extrarenal, oliguria is expected. Physical findings, such as reduced skin or eyeball turgor and dry mucous membranes, are not reliable indicators of hypovolemia.

Laboratory Findings

Laboratory measurements are an adjunct to clinical assessment but do not replace symptoms and physical examination findings as a primary diagnostic tool. Decreases in *hemoglobin* may indicate past or ongoing bleeding, but a normal or stable hemoglobin level does not exclude acute bleeding. Hemoconcentration is often observed when hypovolemia is not the consequence of bleeding, but comorbid disease processes that produce anemia may mitigate this rise.

The *albumin* concentration may rise with gastrointestinal, urinary, or skin losses of albumin-free fluids. Conversely, fluid losses that are accompanied by albumin loss (e.g., proteinuria, protein-losing enteropathy) may mitigate this rise or even result in hypoalbuminemia. Similarly, burns and hepatic disease are often accompanied by hypoalbuminemia as a result of the loss or third spacing of protein-rich fluid.

Serum levels of *sodium* and *other electrolytes* also can vary widely. Even subtle and subclinical degrees of hypovolemia trigger urinary water retention and result in a hypotonic, hyponatremic plasma if the patient is exposed to solutions that are more hypotonic than the fluid lost. In contrast, loss of hypotonic fluids with inadequate water ingestion or replacement results in a hypertonic plasma composition and hypernatremia.

Acid-base (Chapter 118) and *potassium* (Chapter 117) changes point to specific causes of hypovolemia. For example, hypokalemia with metabolic alkalosis frequently accompanies vomiting and some causes of diarrhea (e.g., a villous adenoma). More often, diarrheal fluid loss is associated with a non–anion gap metabolic acidosis. Loop and thiazide diuretic-induced hypovolemia is often associated with hypokalemic metabolic alkalosis, as are the inherited tubulopathies (Bartter and Gitelman syndromes) (Chapter 128), which disrupt sodium reabsorptive mechanisms at the loop of Henle and distal convoluted tubule, respectively. Severe hypovolemia with circulatory compromise and tissue hypoperfusion is often accompanied by lactic acidosis.

Increases in *plasma urea* and *creatinine* concentrations are frequently observed in hypovolemic states and reflect reduced renal plasma flow. If acute tubular injury does not supervene, the rise in plasma urea concentration is often disproportionate to the rise in plasma creatinine concentration (see Prerenal Azotemia in Chapter 120). The rise in plasma urea and creatinine concentrations is particularly common when hypovolemia is a consequence of urinary fluid losses. In such conditions, the patient is not oliguric, even though renal plasma flow and GFR are compromised. Other causes of urinary sodium loss, such as occur with adrenal insufficiency or aldosterone unresponsiveness, are accompanied by a tendency toward hyperkalemia and mild metabolic acidosis.

Urinary biochemical parameters may also help in the clinical assessment of hypovolemic states.[2] When fluid loss is extrarenal, the expected renal response of water and sodium conservation occurs, thereby resulting in oliguria with an elevated urine specific gravity (>1.020), an elevated osmolality (>400 mOsm/kg), and a sodium concentration of less than 20 mmol/L because of enhanced renal tubule reabsorptive activity. More complex indices of the appropriate renal response to hypovolemia include fractional excretion of sodium of less than 1% and fractional excretion of urea of less than 30 to 35%. Intrinsic renal injury confounds the diagnostic value of these urinary indices.

Differential Diagnosis

Relative hypovolemia secondary to arterial vasodilation mimics some of the clinical manifestations of absolute hypovolemia. With vasodilation, as seen, for example, in sepsis (Chapter 108), tachycardia and hypotension are common, but the extremities may be warm. However, tissues are actually underperfused, as reflected by reduced renal and cerebral function and lactic acidosis.

TREATMENT

Absolute Hypovolemia

The major goal in treatment of hypovolemia is to restore hemodynamic integrity and tissue perfusion. The management approach includes treatment of the underlying disease state when possible, replacement of the volume deficit, and fluid administration to maintain ECF volume in the event of continuing losses. Irrespective of specific treatments, the mainstay of therapy involves fluid administration. The important issues are the volume, rate of administration, and composition of the replacement and maintenance fluids. These factors may vary during different stages of treatment and should be adjusted according to the patient's response as determined by closely monitored clinical parameters.

The choice of oral or intravenous replacement fluids (or both) for hypovolemic states is dictated by the integrity of gastrointestinal absorptive function, by the magnitude of the volume deficit, and by the disturbances in other electrolyte and acid-base parameters. The rate of replacement is a function of the urgency of the threat to circulatory integrity and consideration of complications related to overzealous or too rapid correction.

Fluid therapy for hypovolemic states sometimes begins with a diagnostic fluid challenge. In situations in which clinical parameters do not permit a firm diagnosis of hypovolemia, the response to a fluid challenge can be informative and serve as the initial treatment step. For example, a patient with known long-standing compensated heart failure who is being maintained on a therapeutic regimen that includes diuretics may have tachycardia, reduction in blood pressure from baseline values, poor cognition, and renal dysfunction. Such a clinical scenario could have a number of different explanations, including superimposed volume depletion with inadequate left ventricular filling volume. CVP, whether measured directly or assessed by jugular venous pressure, may be misleading in the face of right ventricular dysfunction, but direct measurement of pulmonary capillary wedge pressure does not significantly improve clinical outcomes.[3] Conversely, interventional hemodynamic monitoring to guide fluid resuscitation has been shown, in randomized controlled trials, to lead to lower mortality and a reduced incidence of acute kidney injury in certain postsurgical settings.

Another example is a patient with hyponatremia in the setting of suspected volume depletion. The degree of volume depletion is often too subtle to be detected by clinical examination, and a therapeutic challenge with fluid of the appropriate composition may be the only option.

The initial volume and rate of therapeutic replacement fluid should be determined by ongoing monitoring of clinical parameters rather than by a priori estimates of volume deficit.[3] In some settings, the clinical state will dictate rapid fluid replacement, as in a patient with unambiguous hypovolemic shock and life-threatening circulatory collapse. In such cases, fluids can be administered at the most rapid rate possible, limited only by intravenous access, until blood pressure and tissue perfusion are restored. However, in most cases, much slower rates are indicated, especially in elderly patients, patients whose medical background is unclear, or those with known comorbid conditions. Replacement fluids of different compositions have disparate volumes of distribution in the body fluid compartments and therefore differ in their efficiency of restoring ECF volume. *Crystalloid solutions* with sodium as the principal cation are the mainstay in fluid replacement therapy for hypovolemic states and are indicated primarily for hypovolemic states that are caused by renal, gastrointestinal, or sweat-based sodium losses. These solutions also are useful initial agents and adjuncts to therapy for the hypovolemia of hemorrhage and burns. *Isotonic saline* is confined to the ECF compartment (except in cases of severe dysnatremia). Thus, retention of 1 L of infused isotonic saline increases plasma volume by about 300 mL, with the remaining portion distributed in the interstitial subcompartment of the ECF. In contrast, a solution of *5% dextrose in water* (D_5W) is equivalent to administering solute-free water and distributes uniformly throughout all body fluid compartments (one third of the retained volume of infusate remains in the ECF compartment and only approximately 10 to 15% in the intravascular compartment). Infusing a given volume of *half-isotonic saline (0.45% sodium chloride plus 5% glucose)* can be considered equivalent to infusing half that volume as solute-free water (distributed throughout body fluid compartments) and the other half as isotonic saline (confined to the ECF compartment). The retained solute-free volume reduces body tonicity and the plasma sodium concentration, potentially beneficial in the follow-up treatment of patients whose hypovolemia is accompanied by hypertonicity and hypernatremia but detrimental for patients with normotonic or hypotonic hypovolemia.[4]

When hypovolemia is accompanied by hypobicarbonatemia (metabolic acidosis), it may be appropriate to design a solution in which a portion of the sodium is accompanied by bicarbonate (Chapter 118). For example, it is possible to add a given quantity of hypertonic sodium bicarbonate to a solution of half-isotonic saline (in which chloride is the anion accompanying sodium) to obtain an isotonic replacement fluid appropriate for the given acid-base status of the patient. Similarly, in patients with concomitant potassium depletion (Chapter 117), especially when it is accompanied by metabolic alkalosis, addition of potassium chloride to the replacement solution may be

indicated. A number of crystalloid solutions with predetermined concentrations of potassium, lactate (converted to bicarbonate by the liver), and other electrolytes are commercially available, but it is more appropriate to begin with a sodium chloride–containing solution at a concentration appropriate to body tonicity, then to add other solutes as indicated or at a separate intravenous administration site. This approach provides maximal flexibility in tailoring individualized fluid replacement therapy to the patient's needs. Administration of chloride-restricted intravenous fluids rather than of chloride-liberal intravenous fluids may decrease the incidence of acute kidney injury in critically ill patients.[5]

Colloid-containing solutions include albumin or high-molecular-weight carbohydrate molecules (e.g., hydroxyethyl starch or dextran) at concentrations that exert a colloid osmotic pressure equal to or greater than that of plasma. Banked human plasma is also considered a colloid solution. Because large molecules such as albumin and high-molecular-weight carbohydrates do not readily cross the transcapillary barrier, they are thought to expand the intravascular compartment more rapidly and efficiently than crystalloid solutions. However, randomized trials have not shown any benefit of colloids compared with crystalloids for fluid resuscitation.■ Moreover, some large-molecular-weight carbohydrates, such as hydroxyethyl starch, appear to be nephrotoxic and probably should not be used.■ In patients with multiorgan system failure and capillary leakage, albumin is both rapidly catabolized and redistributed into the interstitial compartment, so it can aggravate interstitial edema without providing the benefit of intravascular volume repletion. Nevertheless, albumin-containing solutions may be useful in hypovolemia associated with burns (Chapter 111), when cutaneous protein losses are appreciable. Furthermore, because of the capacity for rapid intravascular volume expansion with just a small volume of replacement fluid, colloid-containing solutions are frequently used when rapid intravascular expansion is desired, such as at trauma sites outside of the hospital setting. Overall, crystalloid-containing solutions should be the mainstay of volume replacement therapy.

In theory, blood products can be used for volume replacement in hypovolemic states, and a unit of packed red blood cells remains entirely in the vascular compartment. However, erythrocytes are actually considered part of the intracellular compartment and do not contribute to organ plasma flow. The role of packed red cells in the treatment of hemorrhage is to restore the principal function of the erythrocyte in oxygen carriage and delivery, not as a means of ECF volume replacement.

In addition to replacement fluids, maintenance fluids must be provided to counteract ongoing losses. Such ongoing losses may be a continuation of the underlying disease state (e.g., continued vomiting, diarrhea, polyuric states, or severe burns). The volume, rate of administration, and composition of these replacement fluids are best determined by actual measurements of the corresponding ongoing fluid losses, with appropriate adjustments for the patient's clinical assessment parameters.

Relative Hypovolemia

The treatment approach to relative hypovolemia is more complex than for absolute hypovolemia. When relative hypovolemia is the result of peripheral vasodilation, therapy should be directed toward reversal of the underlying cause and restoration of normal vascular reactivity. Bridging to maintain circulatory integrity until the underlying cause is successfully reversed can be achieved by infusion of an *isotonic crystalloid solution* such as normal saline. In such situations, selection of volumes and rates must be done with extreme caution because there is no absolute deficit and the administered volume will have to be excreted or removed once systemic vascular resistance and vascular capacitance are restored to normal. Furthermore, it is more difficult to estimate an increase in vascular capacitance than it is to estimate an absolute volume deficit. On occasion, it is appropriate to consider the use of vasoconstrictor agents.

Hypervolemia

DEFINITION

Hypervolemia refers to expansion of ECF volume, which varies, even in normal individuals, with dietary sodium intake. Thus, an individual in steady state with low daily dietary sodium intake (e.g., 20 mmol/day, corresponding to approximately 1.2 g of table salt per day) will have correspondingly low urinary sodium excretion, equivalent to dietary intake minus extrarenal losses. A shift to much higher sodium intake (e.g., 200 mmol/day, corresponding to approximately 12 g of table salt per day) will bring the individual to a new steady state characterized by a correspondingly higher urinary sodium excretion rate. This shift is accompanied by an increase in ECF volume, which triggers the sensor and effector mechanisms for increased urinary sodium excretion (described earlier). In most individuals, this increase in ECF volume is not clinically detectable and does not have pathologic consequences. In some individuals, however, this upward shift in ECF volume increases systemic arterial blood pressure. When the sodium surfeit expands the ECF volume beyond the range necessary for the adjustment needed to restore sodium balance, a state of pathologic hypervolemia ensues.

TABLE 116-4 PRIMARY AND SECONDARY RENAL SODIUM-RETAINING STATES

PRIMARY
Oliguric renal failure
Chronic kidney disease
Glomerular disease, including nephrotic syndrome
Severe bilateral renovascular obstruction
Mineralocorticoid excess
Inherited sodium-retaining tubulopathies
SECONDARY
Cardiac failure
Cirrhosis
Idiopathic edema

EPIDEMIOLOGY

Primary and secondary renal sodium retention (Table 116-4) can lead to hypervolemia. Patients with oliguric renal failure of any cause (Chapters 120 and 130) have a limited ability to excrete both sodium and water. Urinary sodium retention can be one of the cardinal manifestations of primary glomerular diseases (Chapter 121), even when the GFR is well preserved. States of *mineralocorticoid excess* (Chapter 227) or enhanced activity are associated with a phase of sodium retention; however, because of the phenomenon of "mineralocorticoid escape," the clinical manifestation is generally that of hypertension rather than hypervolemia. Both heart failure (Chapter 58) and cirrhosis (Chapter 153) are associated with renal sodium retention.

PATHOBIOLOGY

Two pathophysiologic mechanisms can lead to sodium retention with ECF volume expansion. The first involves renal sodium retention that is primary and unrelated to the activation of afferent sensor mechanisms. This category includes primary renal diseases and endocrine disorders characterized by excess mineralocorticoid action. In the second category, EABV is reduced, and afferent sensory mechanisms activate effector responses that drive renal sodium retention. In these conditions, total ECF volume is expanded, but intravascular volume is contracted. Therefore, the volume homeostatic mechanisms of the body mimic those of hypovolemia because of the perception of reduced EABV. The degree of solute-free water retention that accompanies the sodium surfeit has a relatively small influence on the extent of hypervolemia but influences the accompanying tonicity state and determines whether the hypervolemia is hypotonic or isotonic.

When the ECF volume is expanded, the relative distribution between the intravascular and extravascular (interstitial) compartments depends on a number of factors. When cardiac and hepatic functions are normal and peripheral transcapillary Starling forces are intact, the excess ECF volume is evenly distributed between the intravascular and interstitial fluid compartments. In such cases, edema does not occur until there is a substantial surfeit of sodium, and hypertension is expected. In contrast, concomitant disruption of transcapillary Starling forces in a given microcirculatory bed would favor the accumulation of retained fluid at one or more such interstitial locations (e.g., dependent edema progressing to anasarca, ascites, pleural effusion, pulmonary congestion).

Primary Renal Sodium Retention

Patients who retain ingested or administered sodium and water loads expand their ECF volume. In patients with chronic kidney disease, the filtered load of sodium remains well above dietary intake until very late stages of severely reduced GFR; even when the GFR is decreased by as much as 90%, the daily filtered load of approximately 2400 mmol still greatly exceeds dietary intake. Nevertheless, the relationship between tubular reabsorption and filtered load may be disrupted in kidney disease.

Monogenic disorders that cause or mimic enhanced mineralocorticoid activity or are associated with enhanced activity of the distal nephron

sodium reabsorptive pathways include Liddle syndrome and pseudohypoaldosteronism type 2 (Chapters 67, 117, and 128). In these conditions and in other causes of mineralocorticoid excess, the only clue to mild hypervolemia may be hypertension, which can be severe. Mineralocorticoid excess, glucocorticoid-remediable hypertension, apparent mineralocorticoid excess, and Liddle syndrome are associated with hypokalemia, whereas pseudohypoaldosteronism type 2 (Gordon syndrome) is often accompanied by hyperkalemia.

Secondary Renal Sodium Retention

With both low-output and high-output heart failure and both systolic and diastolic dysfunction, sodium retention is typical (Chapter 58). Low cardiac output, diversion of cardiac output away from arterial intravascular volume–sensing sites, or a high cardiac output that still is not sufficient to meet tissue demands appears to be a necessary and sufficient condition for initiating renal sodium retention. In the case of cirrhosis with ascites (Chapter 153), hepatic intrasinusoidal hypertension is a sufficient and necessary condition for initiating renal sodium retention. These pathophysiologic disturbances in cardiac or hepatic function disrupt afferent signals that govern normal sodium homeostasis and trigger effector mechanisms that lead to enhanced tubular reabsorption of sodium at multiple nephron sites. At the very earliest stages of disease, sodium retention occurs independently of any measurable or detectable reduction in the volume of the intravascular compartments or any of its measurable subcompartments. At more advanced stages of disease, reduced intravascular volume serves as the overriding stimulus for renal sodium retention and thereby leads to a decompensated state of intractable ECF volume accumulation. The more advanced stages, which often are accompanied by a disproportionate degree of positive water balance and consequent hyponatremia, herald imminent compromise of the GFR. Among the many neuronal and humoral abnormalities that characterize the sodium retention associated with heart failure and cirrhosis are endothelial dysfunction, enhanced sympathetic nerve activity, activation of the renin-angiotensin-aldosterone axis, and resistance to natriuretic peptides. In cirrhosis with ascites (Chapter 153), portosystemic shunting together with translocation of intravascular volume to the splanchnic and venous circulation further compromises EABV. In addition, synthetic dysfunction with resulting hypoalbuminemia favors transudation of fluid into the interstitial compartment. At the level of intrahepatic hemodynamics, intrasinusoidal hypertension results in enhanced hepatic lymph formation. When the rate of enhanced hepatic lymph formation exceeds the capacity for return to the intravascular compartment through the thoracic duct, hepatic lymph accumulates in the form of ascites, and the intravascular compartment is further compromised.

CLINICAL MANIFESTATIONS

In addition to the clinical manifestations of the underlying disease, the clinical manifestations of *hypervolemia* depend on the amount and relative distribution of accumulated fluid in the various ECF subcompartments, including the venous and arterial components of the intravascular compartment (jugular venous distention and hypertension), the interstitial spaces of the extremities, the subcutaneous tissues of the lower back and the periorbital region (peripheral pitting edema, the predominant location of which depends on the patient's position), the peritoneal and pleural spaces (ascites and pleural effusion, respectively), and the alveolar space (pulmonary edema). When cardiac and hepatic function are normal and the transcapillary Starling forces are not disrupted, the excess volume is distributed proportionately throughout the ECF compartment. *Hypertension* may be an early manifestation, depending on cardiac function and the state of systemic vascular resistance. Jugular venous distention (see Fig. 51-3) and peripheral edema (see Fig. 51-7) may be present. Clinically detectable pitting peripheral edema usually signifies the accumulation of at least 3 L of excess interstitial volume. Because intravascular plasma volume is itself only 3 L, any state of generalized peripheral edema must signify ECF volume expansion and therefore past or ongoing renal sodium retention or both.

When cardiac function is impaired because of myocardial disease, valvular disease, or pericardial disease, pulmonary and systemic venous hypertension predominates and systemic arterial pressure may be low as a result of disproportionate accumulation of intravascular volume in the venous as opposed to the arterial circulation (Chapter 58). The presence of transudative ascites (see Fig. 146-4) signifies the substantial accumulation of excess ECF volume in the peritoneal cavity, most commonly secondary to disruption of intrahepatic hemodynamics in the setting of liver disease. Pleural effusions can also be a manifestation of hypervolemia, particularly in the setting of heart failure or advanced cirrhosis with ascites.

DIAGNOSIS

Hypervolemia usually is easily detected by findings of generalized edema, ascites, elevated jugular venous pressure, inspiratory pulmonary crackles, or evidence of the presence of pleural effusion. The prevailing systemic arterial blood pressure often provides a clue about whether the hypervolemic state is secondary to reduced EABV or instead due to primary renal sodium retention. The history and physical examination are often sufficient to yield the diagnosis of an underlying secondary cause of sodium retention, such as heart failure or cirrhosis. Adjunctive laboratory tests providing evidence of cardiac dysfunction or liver disease may be helpful. The presence of glomerular-range proteinuria with hypoalbuminemia indicates a glomerular cause of the sodium retention and hypervolemia. Elevated creatinine points to renal failure, which can be intrinsic or may occur in association with advanced stages of some of the aforementioned conditions, such as heart failure (cardiorenal syndrome) or hepatic cirrhosis (hepatorenal failure). Hypoalbuminemia is characteristic of both cirrhosis and nephrotic syndrome.

A low urine sodium concentration and low fractional excretion of sodium confirm renal sodium retention secondary to a perceived decrease in EABV in the edema states, even in the face of overall hypervolemia. More recently, elevated concentrations of brain natriuretic peptide have been used to support the diagnosis of hypervolemia, particularly in the setting of cardiac failure and renal disease.

TREATMENT Rx

The most important step in ameliorating renal sodium retention is recognition and treatment of the underlying disease. Optimization of hemodynamic parameters in heart failure (Chapter 59), improvement of liver function (Chapter 154), or remission of nephrotic syndrome (Chapter 121) improves or reverses sodium retention. Therapeutic intervention to reduce ECF volume without addressing the underlying disease is often met by complications, especially when ECF volume expansion is associated with decreased intravascular volume or EABV. Nevertheless, three treatment modalities are available to reduce ECF volume directly by inducing negative sodium balance: dietary sodium restriction, diuretics, and extracorporeal fluid removal by ultrafiltration. The modality and the desired rate of sodium removal vary with the clinical setting and depend on the relative distribution of the sodium surfeit and excess volume in the body fluid compartments. Therefore, before initiating any treatment, the clinician should identify the specific disturbances in clinical parameters that are harmful to the patient and monitor the improvement in these parameters during the course of treatment. Harmful manifestations of hypervolemia include hypertension, pulmonary congestion and edema or pleural effusions with compromised respiratory function, hepatic congestion and ascites, and degrees of peripheral edema that compromise skin integrity and predispose the patient to cellulitis. Once ECF volume reduction has removed these threats to the patient's well-being, rates of sodium removal should be slowed significantly. Thus, a patient with mild peripheral edema, small pleural effusions, minimal ascites, jugular venous distention, and normal blood pressure might be managed with sodium restriction and limited use of natriuretic medications to induce a gradual negative sodium balance during a period of many days to weeks. In contrast, a patient with limb- or life-threatening anasarca, pulmonary congestion, or hypervolemia-induced hypertension might require the continuous intravenous infusion of natriuretics or in some cases extracorporeal ultrafiltration therapy.

Sodium Restriction

In the management of chronic hypervolemia, other modalities are futile if they are not accompanied by restriction of sodium intake because renal sodium avidity results in the reaccumulation of ECF fluid as soon as the influence of diuretics has ceased. Dietary sodium restriction in the range of 50 to 100 mmol/day is often recommended and requires abstention from added salt as well as from foods rich in sodium. In acute decompensated heart failure, however, sodium restriction does not augment negative fluid balance over what can be achieved by furosemide alone.[1] Sodium substitutes can be useful, although caution needs to be exercised in patients with a tendency to hyperkalemia because some salt substitutes contain potassium. Calorie intake and nutritional parameters should be monitored to ensure that an overly draconian diet does not induce protein-energy malnutrition. In hospitalized patients, it is particularly important to ensure that the sodium content of administered intravenous fluids and sodium-containing medications is monitored and reduced to the minimum possible. The practice of infusing sodium-containing solutions on the one hand and simultaneously treating with diuretics has no sound physiologic or therapeutic basis. In one randomized trial, however, the combination of high-dose furosemide and small doses of

hypertonic saline was better than furosemide alone for treatment of patients with refractory heart failure or with ascites. Water restriction is not appropriate in hypervolemic edema states unless the plasma sodium concentration is less than 135 mmol/L or symptomatic hyponatremia supervenes.

Diuretics

Diuretics and natriuretics (Table 116-5) enhance the urinary excretion of sodium-containing fluid by inhibiting tubular reabsorption at specific nephron sites (Fig. 116-3).

Proximal Tubule Natriuretics

The cardinal example of a proximal tubule natriuretic is acetazolamide, a carbonic anhydrase inhibitor that blocks proximal reabsorption of sodium bicarbonate. Consequently, prolonged use of acetazolamide may lead to hyperchloremic acidosis, in contrast to all other natriuretics, which act at loci before the late distal nephron. Metolazone, a congener of the thiazide class of natriuretics, blocks sodium chloride absorption in the proximal tubule as well as in the early distal tubule. Because the major locus for phosphate absorption is in the proximal nephron, the phosphaturia accompanying metolazone administration considerably exceeds that observed with other thiazide-class diuretics. Proximal tubule natriuretics rarely are used as primary therapy but are used as supplements to loop natriuretics when loop natriuretics alone are insufficiently effective.

Loop Natriuretics

Loop natriuretics, such as furosemide, bumetanide, torsemide, and ethacrynic acid, induce natriuresis by inhibiting the coupled entry of Na^+, Cl^-, and K^+ across apical plasma membranes in the thick ascending limb of the loop of Henle, which is responsible for the reabsorption of approximately 25% of filtered sodium. Loop diuretics, which are the most potent diuretics, continue to be effective even in patients with relatively compromised kidney function.

Distal Tubule Natriuretics

Distal tubule natriuretics, such as hydrochlorothiazide, chlorthalidone, and metolazone, interfere primarily with sodium chloride absorption in the earliest segments of the distal convoluted tubule, where they block the sodium chloride cotransport mechanism across apical plasma membranes. Distal tubule natriuretics generally are used in the same conditions as loop natriuretics are, but not in chronic kidney disease and in disorders of calcium metabolism. Whereas loop natriuretics are calciuretic and are valuable for managing acute

TABLE 116-5 DIURETICS AND OTHER NATRIURETIC MEDICATIONS

DIURETICS IN COMMON USE	DAILY DOSE RANGE	ADVERSE REACTIONS	COMMENTS
Thiazides (oral)		Rash, neutropenia, thrombocytopenia, hyperglycemia, hyperuricemia	Usually not effective below GFR of 30-40 mL/mm (metolazone, 20-30 mL/mm)
Hydrochlorothiazide	25-100 mg		
Metolazone	2.5-5 mg		
Chlorthalidone	20-50 mg		
Loop diuretics (oral or intravenous)			Rapid onset, short duration Split doses in normal renal function; give intravenously in acute situations or if reduced gastrointestinal absorption Can use up to 500 mg furosemide (or equivalent) in severe renal insufficiency
Furosemide	20-320 mg	Ototoxicity at high doses	
Bumetanide	1-8 mg		
Torsemide	20-200 mg		
Potassium sparing		Hyperkalemia	Not very potent
Spironolactone	25-400 mg		
Triamterene	25-100 mg		
Amiloride	5-20 mg		
Eplerenone	25-50 mg		

GFR = glomerular filtration rate.

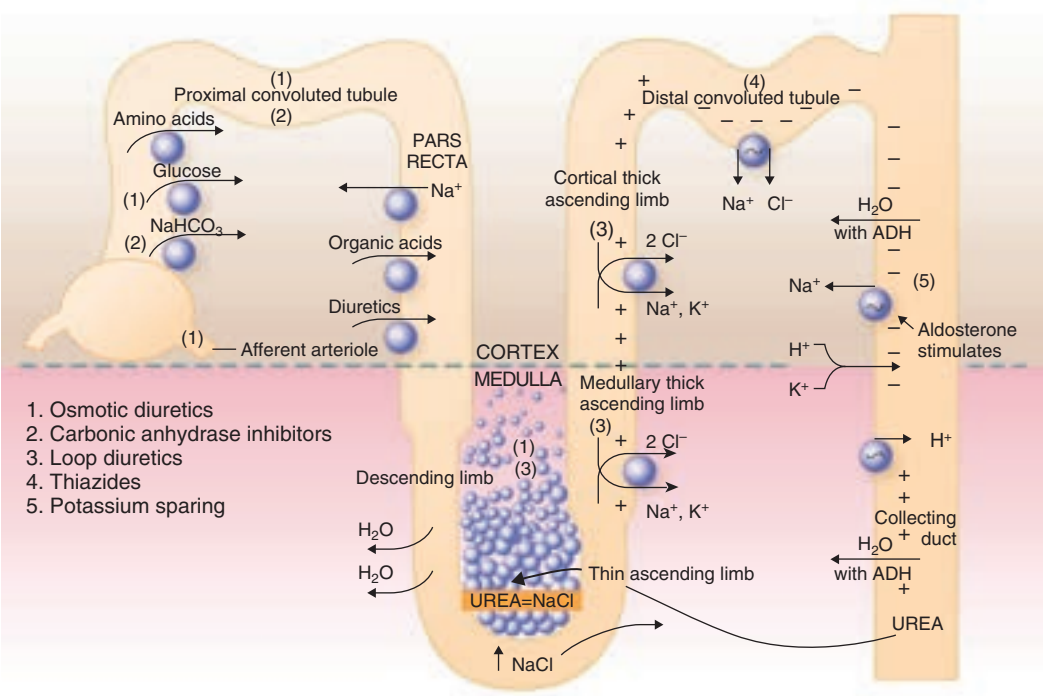

FIGURE 116-3. Major transport processes along the nephron segments and primary sites of action of diuretics. The site of action of diuretics is shown by numbers in parentheses in each nephron segment; numbers correspond to those next to the diuretics listed in the lower left section of the figure. ADH = antidiuretic hormone. (From Kokko JP. Diuretics. In: Alexander RW, Schlant RC, Fuster V, eds. *The Heart*. 9th ed. New York: McGraw-Hill; 1998.)

hypercalcemia (Chapter 245), thiazide natriuretics promote hypocalciuria and calcium retention and are useful in managing hypercalciuric states. With the exception of acetazolamide, which impairs bicarbonate absorption, the natriuretics discussed so far can cause hypokalemia and metabolic alkalosis.

Collecting Duct Natriuretics

Spironolactone and eplerenone compete with aldosterone and inhibit sodium absorption in the collecting duct while concomitantly suppressing potassium and proton secretion. Triamterene and amiloride directly block sodium uptake by collecting duct cells and concomitantly suppress potassium and proton secretion. These agents are used in combination with thiazide and loop natriuretics to offset hypokalemia. However, hyperkalemia and hyperchloremic metabolic acidosis may complicate the injudicious use of any of these agents. Spironolactone and eplerenone are useful in managing disorders characterized by secondary hyperaldosteronism (such as cirrhosis with ascites), in promoting natriuresis in hypokalemic patients, and in competitively blocking nonepithelial mineralocorticoid receptors in patients with left ventricular dysfunction (Chapter 59).

Nesiritide is the recombinant version of a naturally occurring brain natriuretic peptide with unique vasodilator and natriuretic actions. Nesiritide given alone may compromise renal function, especially when high levels are achieved after an initial bolus, so its usefulness appears to be limited; frequent monitoring of urine output and of plasma urea and creatinine concentrations is required. However, a recent randomized trial showed therapeutic benefit with the combination of an angiotensin-converting enzyme inhibitor (enalapril) and the endopeptidase inhibitor AHU-377 (a neprilysin inhibitor prodrug, which blocks the breakdown and, hence, augments the levels and activity of endogenous natriuretic peptides) for patients with heart failure.

Patients with severe degrees of renal sodium avidity can be resistant to conventionally recommended doses of individual classes of diuretic agents. In such patients, combinations of diuretic agents acting at different sites along the nephron may overcome this resistance and induce a natriuretic response.[6] The continuous intravenous infusion of furosemide, sometimes in conjunction with intermittent bolus infusions of albumin, also can overcome natriuretic resistance in some hospitalized patients. Monitoring of plasma sodium, potassium, magnesium, calcium, and phosphate concentrations is mandatory in patients treated with high or frequent doses or continuous infusions of natriuretic agents. Besides body tonicity and electrolyte disturbances, other potential adverse consequences include a reduction in GFR. Drug-specific idiosyncratic adverse responses, such as allergic cutaneous reactions, interstitial nephritis, pancreatitis, and blood dyscrasias, are much less common.

Extracorporeal Ultrafiltration

In a small subset of patients, either superimposed renal impairment or extreme resistance to natriuretic action may require the direct removal of excess ECF volume by ultrafiltration, hemodialysis, or peritoneal dialysis (Chapter 131). Chronic ambulatory peritoneal dialysis has been used for the symptomatic relief of pulmonary congestion and anasarca in some patients with chronic heart failure who are unresponsive to other therapeutic modalities and are not candidates for cardiac transplantation.

WATER BALANCE DISORDERS

Water balance disorders generally come to medical attention because of one or more of three clinical manifestations: hyponatremia,[7] hypernatremia,[8] or polyuria.

Hyponatremia

DEFINITION

Hyponatremia, which is defined as a plasma sodium concentration of less than 136 mmol/L, is the most frequently encountered electrolyte abnormality in hospitalized patients. Hyponatremia, irrespective of the underlying cause, is independently associated with higher mortality.[9]

EPIDEMIOLOGY AND PATHOBIOLOGY

Hyponatremia may be hypertonic, isotonic, or hypotonic. *Hypertonic hyponatremia* occurs when there is an accumulation in the ECF compartment of non–sodium-containing effective solutes, such as very high concentrations of glucose in diabetic patients or exogenously administered mannitol or glycerol. These non-sodium solutes lead to a shift of water from the ICF to the ECF compartments and consequent ICF shrinkage. The accumulation of a solute such as urea, which contributes to the measured plasma osmolality but is not an osmotically effective solute in terms of transcellular water shift, should not be included in the category of hypertonic hyponatremic states. *Isotonic hyponatremia* signifies the laboratory finding of hyponatremia in patients with no disturbances in body fluid tonicity and almost always reflects the interference of marked hyperlipidemia or marked hyperglobulinemia with certain laboratory techniques for the measurement of the plasma sodium concentration; these situations are termed *pseudohyponatremia* and should always be excluded before embarking on diagnostic or therapeutic measures to alter water balance or body tonicity.

True *hypotonic hyponatremia* always reflects an important underlying disorder that leads to abnormal body water retention and either past or ongoing expansion of ICF volume. Even in chronic hypotonic hyponatremic states in which cell volume has been restored to normal by osmotic adaptive mechanisms, the compensation occurs at the price of loss of intracellular solutes and compromised cell function.

Hypotonic hyponatremia can be further classified according to volume status. Heart failure (Chapter 59) and cirrhosis with ascites (Chapter 153) are examples of hypervolemic hyponatremia. In these conditions, reduced EABV stimulates the release of AVP and also may limit the delivery of glomerular ultrafiltrate to the diluting segments of the nephron, thereby leading to impaired water excretion. The hyponatremia that is seen in advanced renal failure (Chapter 130) because of impaired excretion of water may also be associated with hypervolemia.

Hypovolemic hyponatremia occurs when relatively more sodium than water is lost through the gastrointestinal tract (e.g., diarrhea) or in urine (e.g., thiazide diuretics). Decreased body tonicity can also develop even when fluid loss is isotonic or hypotonic (e.g., sweat) if the ingested or administered replacement fluid is more hypotonic than the lost fluid (e.g., ingestion of water or intravenous administration of D_5W).

Another important consideration is the potassium concentration in the lost fluid. For example, diarrheal fluid and natriuretic medication–induced polyuric urine are often rich in potassium as well as in sodium. Because water moves freely across most cell membranes, the ECF and ICF are in osmotic equilibrium. As a result, plasma tonicity is equal to the sum of the effective osmolalities of the ICF and total body water. Given that sodium is the principal determinant of ECF tonicity and potassium is the main contributor to ICF tonicity, plasma sodium is directly proportional to the sum of sodium and potassium concentrations in total body water. Therefore, even if the concentration of lost sodium alone is less than that in the ECF, combined loss of sodium plus potassium can cause hypotonicity.

CLINICAL MANIFESTATIONS

The finding of hyponatremia is often incidental on routine laboratory testing, on laboratory testing of patients with nonspecific complaints, or as part of the investigation of other clinical syndromes. The symptoms of hypotonic hyponatremia depend on its duration, severity, and rate of development. When hyponatremia develops rapidly (hours to days), acute brain swelling or cerebral edema occurs and is manifested as headache, lethargy, seizures, and a progressively decreased level of consciousness that can lead to coma and death. In addition, women between menarche and menopause are particularly susceptible to the life-threatening neurologic manifestations of acute hyponatremia, even of relatively mild degree. In contrast, when the rate of decline in plasma sodium concentration is more gradual, osmotic adaptation ensues, and even severe hyponatremia (plasma sodium concentration <120 mmol/L) may induce few or only subtle clinical manifestations.

DIAGNOSIS

After immediate assessment of the clinical urgency of the situation, the first step in the diagnostic approach to a patient with hyponatremia is to establish the presence of true hypotonic hyponatremia. Plasma electrolytes, urea, glucose, and osmolality should be checked to allow comparison of the measured with the calculated plasma osmolality according to the following equation:

$$\text{Plasma osmolality (mOsm/kg)} = 2\text{Na}^+\text{(mmol/L)} + (\text{blood urea nitrogen [mg/dL]}/2.8) + (\text{glucose [mg/dL]}/18)$$

A careful history, including a search of the medical record for previous plasma sodium values, will help to indicate the rate of decline. The history and physical examination usually provide important clues to underlying disorders, disease states, or medication exposures that can inform the diagnosis. A history of weight gain or loss also can be helpful in the assessment of recent fluid mass balance. The physical examination should focus on attempts to establish the state of ECF volume. The presence of generalized edema with jugular venous distention and ascites, especially in the setting of heart or liver disease, points clearly to hypervolemic hyponatremia. Conversely, orthostatic hypotension and tachycardia, particularly in the setting of a history of

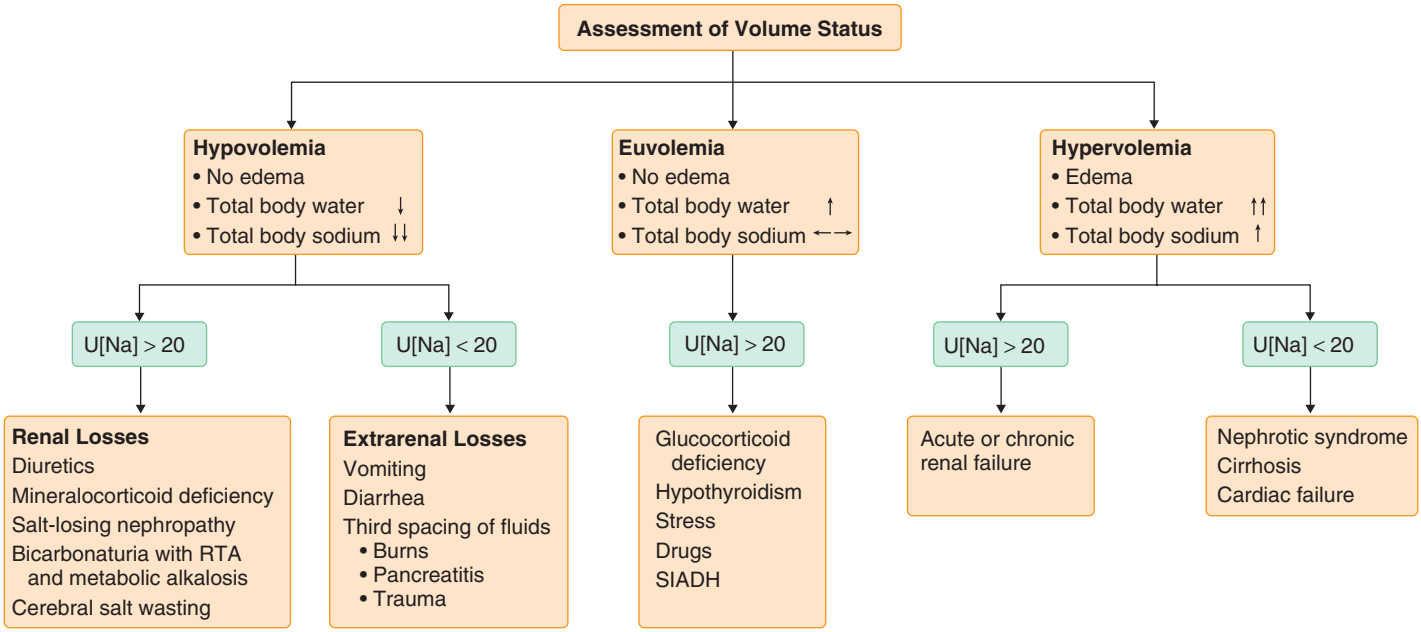

FIGURE 116-4. Diagnostic approach to hyponatremia. RTA = renal tubular acidosis; SIADH = syndrome of inappropriate antidiuretic hormone secretion. (Modified from Halterman R, Berl T. Therapy of dysnatremic disorders. In: Brady H, Wilcox C, eds. *Therapy in Nephrology and Hypertension*. Philadelphia: Saunders; 1999:256.)

natriuretic medication use or gastrointestinal fluid losses, suggest hypovolemia, but the absence of these findings does not exclude hypovolemia.

Further laboratory tests should include a repeated set of plasma electrolyte concentrations, including potassium and chloride levels, which together with determination of acid-base parameters (pH, P_{CO_2}, and bicarbonate) can point to processes not always detectable in the history, such as vomiting, diarrhea, or natriuretic medication exposure. Other laboratory tests should include liver function tests and measurement of plasma urea, creatinine, uric acid, thyroid-stimulating hormone, and cortisol concentrations and, if indicated, an adrenocorticotropic hormone stimulation test. High levels of both urea and creatinine point to intrinsic renal disease, whereas a disproportionate elevation of urea over creatinine might support hypovolemia with a tendency to prerenal azotemia (Chapter 120). In contrast, very low levels of urea and uric acid are typical of both the syndrome of inappropriate antidiuretic hormone secretion (SIADH) and the cerebral salt-wasting syndrome.

Marked elevation in the plasma glucose concentration increases both measured and calculated plasma osmolality and indicates a state of hypertonic hyponatremia that should be approached as a state of body fluid hypertonicity with cell shrinkage (see later) rather than hypotonicity. The plasma sodium concentration declines by approximately 1.6 mmol/L for each increase of 100 mg/dL (5.5 mmol/L) in plasma glucose concentration. However, this decline in plasma sodium is variable and is greater in states of progressively severe hyperglycemia. In contrast to hyperglycemia, an elevated urea concentration should not be considered as contributing to plasma or ECF tonicity, even though urea does contribute to the laboratory measurement of plasma osmolality. Thus, a hyponatremic patient with a normal or elevated laboratory measurement of plasma osmolality that can be fully attributed to an increased urea concentration should be considered as having hypotonic hyponatremia. A discrepancy in which measured plasma osmolality exceeds calculated plasma osmolality and that cannot be attributed to either glucose or urea indicates the presence of an unidentified small solute (osmolar gap), including alcohols (e.g., ethanol, methanol, ethylene glycol, and isopropyl alcohol) and the organic anions of weak acids, which raise the plasma anion gap. Because these small molecules are not effective solutes in terms of water movement, the water balance and tonicity status of the patient is determined by the plasma sodium concentration. Just as for urea, a patient with hyponatremia and normal or elevated measured plasma osmolality as a result of one of these small solutes should be approached as having a true hypotonic hyponatremia, notwithstanding the normal or elevated plasma osmolality measurement. However, the finding of such an osmolar gap should prompt a thorough investigation for poisoning, intoxication, or an organic acidosis (Chapter 118).

Once a state of true hypotonic hyponatremia has been established, determination of the cause and further diagnostic approach follow a classification into one of three categories based on assessment of the volume status of the patient (Fig. 116-4). Abnormal liver function test results can provide adjunctive support for hepatic disease and a hypervolemic hyponatremic state. The diagnosis of heart failure should be made clinically, but it can be assisted by a brain natriuretic peptide level, chest radiograph, or echocardiograph (Chapter 58). A radiograph or chest computed tomography scan may help identify intrathoracic lesions that are associated with SIADH.

Approximately 85% of hyponatremic inpatients have true hypotonic hyponatremia. Among these patients, about 25% are hypovolemic, about 25% have an edema state, about one third are normovolemic, and most of the remainder have renal failure.

In the absence of a clinically obvious edema state, a low urine sodium concentration (<20 mmol/L) or a low fractional excretion of sodium (<1%) supports the diagnosis of hypovolemic hyponatremia secondary to extrarenal losses or past renal losses that have since abated. If hypovolemia is due to ongoing renal losses, the urine sodium concentration may remain high in the face of hypovolemia, but a low fractional excretion of urea (<35%) may still be evident. High urinary concentrations of potassium point to persistent renal loss, such as from the ongoing effect of potassium-depleting natriuretic drugs. Conversely, potassium loss due to diarrhea or vomiting would lead to renal potassium retention and low urinary concentrations of potassium (<20 mmol/L). In SIADH, the urine sodium concentration often reflects sodium intake as well as mild volume expansion and is therefore most often higher than 40 mmol/L and frequently higher than 100 mmol/L. Hypotonic hyponatremia in a patient without evidence of either hypovolemia or hypervolemia, together with low plasma urea and uric acid concentrations without hypothyroidism or adrenal insufficiency, strongly suggests SIADH. If there is any doubt about the presence of hypovolemia, a carefully monitored volume challenge can be of diagnostic as well as therapeutic benefit (see later). Lack of sustained improvement after an adequate salt-containing volume challenge lends further support to the diagnosis of SIADH. Once the diagnosis of SIADH has been established in this manner, the cause should be sought, including a thorough review of medication exposure, review of the history for symptoms and signs of malignant disease, and magnetic resonance imaging or computed tomography of the brain and chest (Table 116-6).

Hypervolemic Hyponatremia

A patient with hypervolemic hyponatremia suffers from a surfeit of both sodium and water, but the surfeit of water is disproportionate to that of sodium. Once the hypervolemic state has been established, the cause of the sodium surfeit should be determined. The most frequent causes are heart failure, decompensated cirrhosis with ascites, and renal failure. The occurrence of hyponatremia in any of these conditions often signifies advanced disease, although it may be due to overzealous sodium deprivation while the patient is taking natriuretic medications. Patients who have heart failure or cirrhosis with ascites experience avid renal sodium retention with urine sodium concentrations of less than 20 mmol/L and fractional sodium excretion of less than 1% in the face of clear-cut clinical evidence of ECF volume

TABLE 116-6 CAUSES OF THE SYNDROME OF INAPPROPRIATE ANTIDIURETIC HORMONE SECRETION

MALIGNANT NEOPLASIA
Carcinoma: bronchogenic, pancreatic, duodenal, ureteral, prostatic, bladder
Lymphoma and leukemia
Thymoma and mesothelioma
CENTRAL NERVOUS SYSTEM DISORDERS
Trauma
Infection
Tumors
Porphyria
PULMONARY DISORDERS
Tuberculosis
Pneumonia
Fungal infections
Lung abscesses
Mechanical positive-pressure ventilation
DRUG INDUCED
Carbamazepine
Desmopressin
Oxytocin
Vinca alkaloids
Alkylating agents/antimetabolites
Interferons
Anticonvulsants
Antipsychotic agents
Nicotine
Cyclophosphamide
Morphine
Amitriptyline
Selective serotonin reuptake inhibitors
3-4-Methylenedioxymethamphetamine (Ecstasy)

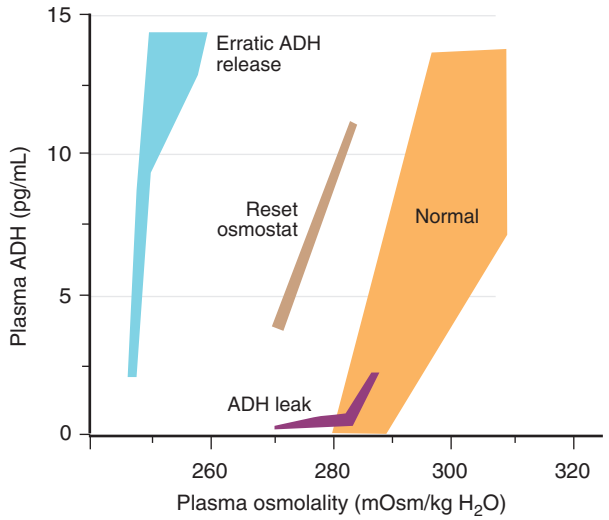

FIGURE 116-5. Patterns of serum antidiuretic hormone (ADH) abnormalities in the syndrome of inappropriate ADH secretion (SIADH). The shaded area to the right indicates the normal relationship between increases in effective extracellular osmolality and ADH levels; the normal osmotic threshold is lower than the normal serum osmolality. The three other shaded areas indicate ADH patterns in SIADH from various causes. (Modified from Zerbe R, Strope L, Robertson G. Vasopressin function in the syndrome of inappropriate diuresis. *Annu Rev Med.* 1980;31:315-327.)

expansion, usually with generalized edema. However, these urine parameters can be masked by the ongoing influence of natriuretic agents. The edema state of the nephrotic syndrome is less commonly associated with hyponatremia unless the patient has been exposed to severe salt restriction and natriuretic therapy. Water retention with hyponatremia is a feature of renal failure only in its more advanced stages (stage 4 and stage 5 chronic kidney disease; Chapter 130).

Normovolemic and Hypovolemic Hyponatremia

It is often difficult to distinguish normovolemic from hypovolemic hyponatremia because mild hypovolemia can easily escape clinical detection. The initial history and physical examination should try to establish or to exclude a cause of hypovolemia. In extrarenal hypovolemia, a low urine sodium concentration and low fractional excretion of sodium are characteristic. When the hypovolemia is due to urinary loss, the urine sodium concentration is usually elevated rather than decreased.

Hypovolemic hyponatremia always signifies past or ongoing sodium loss (often with potassium), accompanied by a degree of net water loss that does not match the electrolyte loss and hence leaves the patient hypotonic. The high levels of AVP that are associated with hypovolemic hyponatremia are usually an appropriate response to the physiologic stimulus of hypovolemia. The most common extrarenal causes of hypovolemia leading to hyponatremia are gastrointestinal fluid losses and excessive sweating. In gastrointestinal fluid losses, any concomitant nausea and vomiting may be independent triggers for the central release of AVP. In unaccompanied patients with impaired consciousness or cognitive dysfunction, clues to vomiting include the characteristic plasma and urine biochemical parameters of metabolic alkalosis (Chapter 118), often with higher than expected urinary concentrations of sodium and bicarbonate. Hyponatremia is a more common complication of diarrhea when the diarrheal fluid is secretory and rich in electrolytes. Sweating-induced hyponatremia occurs when individuals ingest high volumes of hypotonic fluid, often pure water, while losing sodium in sweat.

The renal causes of hypovolemic hyponatremia (see Fig. 116-4) include thiazide-induced hyponatremia and the cerebral salt-wasting syndrome. Thiazide-induced hyponatremia occurs when patients with impaired urinary diluting capacity excrete concentrated urine because thiazides do not affect the ability of the renal medulla to concentrate urine. Thiazide-treated patients are particularly susceptible to hyponatremia when they ingest or receive hypotonic solutions that exceed their maximal capacity to excrete electrolyte-free water in their urine. In cerebral salt wasting (also known as renal salt wasting), patients who have suffered a head injury or intracranial hemorrhage experience a state of negative sodium balance due to inappropriate renal sodium wasting. The consequent ECF volume depletion stimulates the release of AVP, and these patients are prone to the development of hyponatremia if hypotonic fluid is then ingested or administered. The combination of a high urine sodium concentration and hyponatremia makes the syndrome difficult to distinguish from hyponatremia caused by the syndrome of inappropriate secretion of AVP (see later). However, persistently low plasma uric acid levels and high fractional excretion of uric acid, even after correction of the hyponatremia, may be observed in cerebral salt wasting but not in SIADH.

In normovolemic hypotonic hyponatremia, there is neither an osmolar nor a volume stimulus to the release of AVP. Thus, concentrated urine, usually containing concentrations of sodium higher than 40 mmol/L as a result of dietary intake plus the effects of mild ECF volume expansion, indicates either inappropriate secretion or an augmented renal response to AVP. Conditions that can result in inappropriate AVP secretion or responsiveness include tumors, central nervous system lesions or disorders, intrathoracic or chest wall disease, and numerous drugs and medications (see Table 116-6). All syndromes in which the AVP level or the kidney's responsiveness is inappropriately high and not attributable to osmolar or volume stimuli are known collectively as SIADH (Chapter 225). Patterns of abnormal AVP secretion (Fig. 116-5) include an erratic release of AVP from the neurohypophysis without any apparent coordination with incoming volume or osmotic stimuli (type A pattern), a constant low-level leak of AVP from the neurohypophysis (type B pattern), a reduced threshold for osmotic release of AVP at a lower than normal plasma osmolality (type C pattern), and an abnormal renal response to circulating AVP in patients whose neurohypophysial regulation is intact (type D pattern). However, no consistent correlation between these various patterns and an underlying cause has emerged. A specific monogenic disorder involves a mutation in which the AVP V_2 receptor is constitutively active in the absence of ligand. Hypothyroidism (Chapter 226) and adrenal glucocorticoid insufficiency (Chapter 227) can be associated with hypotonic hyponatremia without clinically evident hypovolemia and with a clinical and biochemical profile that mimics SIADH; abnormal regulation of the aquaporin 2 water channel and reduced distal tubular delivery may be involved. Pregnancy, which also is associated with a reduction in both the osmotic threshold for AVP release and thirst, results in mild hyponatremia (Chapter 239). Another unusual setting for normovolemic hyponatremia is the "beer potomania" syndrome. Because the minimal urine osmolarity, even in the absence of AVP, is 30 to 50 mOsm/L, the upper limit of solute-free water excretion depends on total obligate solute excretion. A paucity of available urinary solutes sets an upper limit on the total water intake that can be

tolerated without inducing hyponatremia. For example, when patients consume large volumes of beer (rich in carbohydrates and water but poor in sodium and electrolytes), the absence of protein intake limits urea production and excretion, thereby limiting non-electrolyte urinary solutes and hence urinary water excretion. Together with the large volumes of beer ingested, the result is the unusual combination of a normovolemic hypotonic hyponatremic state with low urine osmolality.

TREATMENT

Rx

Treatment of hyponatremia varies by the urgency of the clinical situation, its cause, and underlying diagnosis (Fig. 116-6). The overall approach can be divided into the immediate treatment of symptomatic hypotonic hyponatremia and the long-term management of chronic persistent hyponatremia.

The first principle common to all causes of hypotonic hyponatremia, irrespective of the underlying cause, is that the sodium concentration and the rate of correction should be guided by the patient's age, gender, and neurologic status and any information about recent and past plasma sodium concentrations or osmolality values. Delayed correction of hyponatremia can perpetuate cerebral edema and result in irreversible neurologic damage and death, especially in women of reproductive age and in patients whose hyponatremia developed at a rapid pace that outstripped the rate of osmotic adaptation by brain cells. In contrast, overly rapid correction or correction to a sodium concentration that is above the level needed to safeguard the patient from the neurologic sequelae of cerebral edema can result in the *osmotic demyelination syndrome*.[10] This devastating and often irreversible syndrome is characterized by fluctuating levels of consciousness, pseudobulbar palsy, ataxia, dysarthria, difficulty in swallowing, and characteristic abnormalities in the region of the brain stem on magnetic resonance imaging. Osmotic demyelination syndrome can be fatal, and recovery in nonfatal cases is either slow or incomplete, often with irreversible residual neurologic sequelae.

The second principle is the importance of identifying and treating any underlying disorder. Thus, in a patient with hypervolemic hyponatremia associated with heart failure, measures to optimize cardiac function are the most appropriate and effective means of restoring a normal sodium concentration. Indeed, restoration of normal sodium concentration provides one of the most reassuring indices for successful management of this underlying disorder. Conversely, in hypovolemic hyponatremia, appropriate correction of ECF volume will lead to the resolution of hyponatremia.

Acute Hyponatremia

Current guidelines suggest that if the hyponatremia is known to be acute (<24 to 48 hours) and is accompanied by severe neurologic symptoms such as seizures or decreased level of consciousness, correction should be rapid; 4 to 6 mmol/L in 4 to 6 hours should be sufficient to reverse the most severe neurologic symptoms. The total increase in sodium concentration should not exceed 9 to 12 mmol/L (some experts recommend no more than 6 to 8 mmol/L) in the first 24 hours or surpass 18 mmol/L within 48 hours.

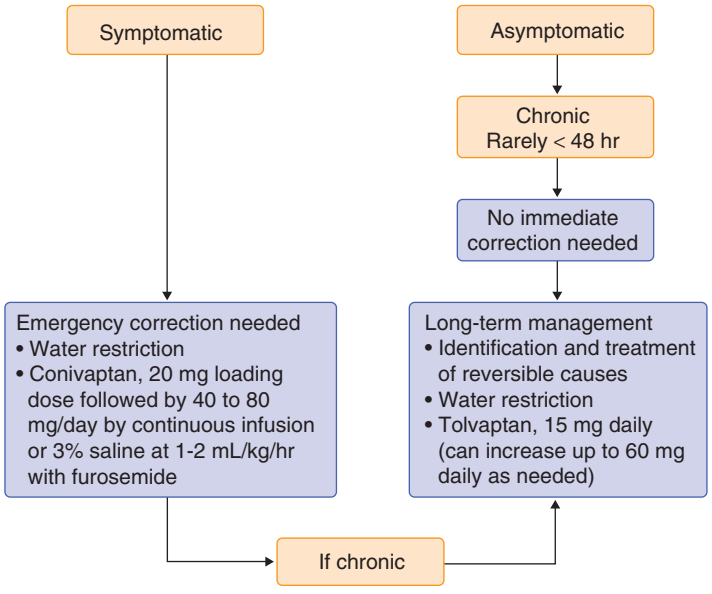

FIGURE 116-6. Treatment of severe normovolemic hyponatremia. (From Thurman J, Halterman R, Berl T. Therapy of dysnatremic disorders. In: Brady H, Wilcox C, eds. *Therapy in Nephrology and Hypertension*. 2nd ed. Philadelphia: Saunders; 2003.)

Chronic Hyponatremia

Mild degrees of hyponatremia can be tolerated for long periods but have been associated with increased risk of hip fractures.[11] Current recommendations are that only symptomatic hyponatremia or sodium concentrations below 125 to 130 mmol/L require specific additional treatment.

If the rate of decline in plasma sodium concentration has been slow, brain cells have the opportunity to undergo osmotic adaptation by extruding or eliminating intracellular solutes. It is this subgroup of patients who are most susceptible to osmotic demyelination after too rapid or overzealous correction of hyponatremia. Patients in whom there is no previous record of sodium concentration or osmolality should be considered in the same category and treated accordingly. In such cases, the targeted rate of increase in sodium concentration should not exceed 0.5 mmol/L/hour, and the total rise in sodium concentration should not exceed 8 mmol/L in any 24-hour period, even (and especially) if the initial sodium concentration is extremely low (<110 mmol/L), provided the hyponatremia is not accompanied by severe neurologic symptoms. Patients with severe degrees of chronic hyponatremia in the setting of malnutrition, alcoholism, or chronic illness are particularly susceptible to osmotic demyelination. Frequent monitoring of the plasma sodium concentration and osmolality is crucial. If the safe target rate of correction is exceeded, osmotic demyelination can be prevented by slowing the correction rate, returning to a lower plasma sodium concentration by the judicious readministration of hypotonic solutions, or administering vasopressin analogues (see later).

A target goal and rate of correction having been established, the approach varies with the underlying diagnosis. In hypervolemic hyponatremia, there is a surfeit of both sodium and water, but in tonicity terms, the excess water is disproportionate to the excess sodium. Thus, the goal of treatment is to remove both sodium and water but to replace proportionately less water than sodium. Water restriction is helpful but is often inadequate or impractical because patients are thirsty, and their adequate nutrition requires calorie intake that is accompanied by obligate water ingestion and metabolic water production. The AVP V_2-receptor antagonists tolvaptan (starting at 15 mg once daily, with a maximal dose of 60 mg once daily) ameliorates hyponatremia and improves symptoms and outcomes in patients with hypervolemic hyponatremic syndromes, including decompensated heart failure and hepatic cirrhosis.[12] Moreover, in a randomized trial in patients with acute decompensated heart failure, tolvaptan alone was found to have a diuretic effect superior to that of furosemide alone; the combination of the two drugs produced a diuresis similar to that of tolvaptan alone.[13] In another randomized placebo-controlled trial, a newer member of the vaptan group, lixivaptan, also was effective as an add-on diuretic to standard therapy in the management of acute decompensated heart failure.[14] Vaptans are also useful in normovolemic hyponatremic states, such as SIADH (see later), but should be assiduously avoided in hypovolemic hyponatremic states.

Hypovolemic Hypotonic Hyponatremia

A frequent diagnostic dilemma is the distinction between hypovolemic and normovolemic hypotonic hyponatremia. When hypovolemia is clearly evident (appropriate clinical history, orthostatic hypotension, low urine sodium concentration in the setting of extrarenal fluid losses, elevated plasma urea disproportionate to the rise in the serum creatinine level, and elevated uric acid concentrations), administration of volume repletion in the form of isotonic saline is the treatment of choice. The salutary effect of saline derives mostly from the effect of volume repletion to remove the hypovolemic stimulus to release of AVP, thereby inducing a water diuresis, with a minor contribution of the osmolar effect of the infused solute. However, great caution should be exercised in the administration of isotonic saline to these patients because the administration of small volumes of isotonic saline can sometimes induce a brisk and rapid decrease in urine osmolality and an accompanying water diuresis, with an overly rapid correction of hyponatremia. Accordingly, whenever isotonic saline or other forms of volume repletion therapy are administered to patients with known or suspected hypovolemic hyponatremia, careful hour-by-hour monitoring of urine output, urine osmolality, plasma sodium concentration, and plasma osmolality is required. A rapid drop in urine osmolality accompanied by water diuresis should prompt cessation of volume repletion and, in some cases, administration of hypotonic solutions or even analogues of AVP itself (see later) to halt or to reverse the rapid rise in sodium concentration to within the recommended guidelines so as to prevent osmotic demyelination. When hypovolemia is not clearly evident but cannot be excluded, a brisk drop in urine osmolality in response to a saline challenge confirms the suspicion of hypovolemia and simultaneously initiates therapy. In contrast, failure to induce such a response lends support to the diagnosis of normovolemic hyponatremia.

Normovolemic Hyponatremia

In patients with normovolemic hyponatremia, the appropriate therapeutic approach is to address the underlying disease. Hypothyroidism (Chapter 226) and adrenal insufficiency (Chapter 227) should be corrected with appropriate hormonal replacement therapy. Medication-induced SIADH mandates identification and cessation of the offending medication when possible. If the

underlying disease cannot be identified or reversed, treatment is aimed at removal of the water surfeit. The therapeutic outcome depends on the minimal urine osmolality that can be achieved, which in turn depends on the severity of SIADH. In many cases, when urine osmolality cannot be suppressed below certain high levels, the severity of water restriction required would not be consistent with the need for calorie intake or compatible with reasonable expectations for the patient's adherence. Therefore, maneuvers to generate a gradual net negative water balance are required. In such cases, titrated dose oral tolvaptan (beginning at 15 mg once daily and increasing to a maximum of 60 mg once daily) and reliance on an intact thirst mechanism are crucial to avoid polyuria or hypernatremia. In rare cases with an urgent need to correct the hyponatremia because of a neurologic emergency or definitive documentation that the sodium concentration has decreased acutely during a 24- to 72-hour period, intravenous conivaptan (20 mg loading dose followed by 40 to 80 mg/day by continuous infusion) can be used cautiously to avoid overly rapid correction of hyponatremia.[13] When volume status is in doubt, however, these agents are better avoided until the possibility of hypovolemic hyponatremia has been addressed. Hemodialysis, which can rapidly raise the plasma sodium concentration, should be reserved for the most extreme cases of acute life-threatening hyponatremia for which no other solution is available (Chapter 131).

Hypernatremia

DEFINITION

Hypernatremia, defined as a plasma sodium concentration higher than 144 mmol/L, always reflects a state of hypertonicity, with an increase in the ratio of the concentration of osmotically active solutes to water throughout all body fluid compartments.

EPIDEMIOLOGY AND PATHOBIOLOGY

Because sodium is an osmotically effective ECF solute, hypernatremic patients have undergone a process whereby water has moved from the ICF to the ECF compartment, accompanied by a reduction in ICF volume and cell shrinkage. Cell shrinkage in the brain is associated with intracerebral hemorrhage, which is often punctate but sometimes due to ruptured blood vessels, particularly at the brain surface and arachnoid interface. In an effort to restore their cell volume, brain cells undergo osmotic adaptation by accumulating sodium and other electrolytes and then subsequently producing non-electrolyte small solutes (osmolytes) such as inositol, taurine, glutamine, and glutamate, among others. This process partially reverses cell shrinkage, but at the price of an altered intracellular solute composition with consequent perturbations in neuronal function.

Hypernatremia is the most frequent but not the only hypertonicity state in clinical medicine. Glucose, mannitol, and glycerol can produce hypertonicity states that may not be accompanied by hypernatremia and in fact are frequently accompanied by hyponatremia (see earlier). In hypertonicity states, the measured plasma osmolality is always high, but elevated plasma osmolality is not necessarily associated with hypertonicity because a number of solutes that contribute to the measured plasma osmolality are not osmotically effective in terms of movement of water from the ICF to the ECF compartment. Thus, patients with high concentrations of urea or small alcohols (e.g., methanol, ethylene glycol, ethanol) often have elevated plasma osmolality but should not be considered to have a hypertonicity state.

Although hypernatremia can be diagnosed as an incidental laboratory abnormality, it most commonly occurs in the setting of a severe underlying disease with other accompanying disturbances in body fluid homeostasis (Table 116-7).

In *hypervolemic hypernatremia*, a disproportionate excess of sodium over water expands ECF volume, but owing to water egress from cells in an attempt to restore normal plasma tonicity, ICF volume is decreased. Hypervolemic hypernatremia usually occurs in the hospital setting because of inadvertent or overzealous administration of hypertonic saline, the administration of hypertonic sodium bicarbonate solutions during cardiopulmonary resuscitation, or dialysis against a hypertonic dialysate.

In *normovolemic hypernatremia*, a pure water deficit with no disturbance in total body sodium content or clinically perceptible decrease in ECF volume occurs owing to proportionally greater water loss from the ICF (approximately two thirds) than from the ECF compartment. Thus, for example, a 3-L pure net water deficit will consist of 2 L lost from the ICF but a clinically undetectable loss of only 1 L from the ECF compartment. Nevertheless, a 3-L or greater deficit certainly increases body fluid tonicity and the measured plasma sodium concentration. Clinical conditions in this category require a source of fluid loss that has a relatively low content of osmotically effective solutes (principally sodium and potassium and their accompanying anions), such as the various forms of diabetes insipidus or the use of AVP V_2-receptor antagonists (vaptans) without adequate monitoring. In these conditions, profuse volumes of low-osmolality urine are excreted. Even so, hypernatremia is actually uncommon as long as thirst perception and availability of water remain intact. The principal clinical manifestation is polyuria and polydipsia (see later). Insensible evaporative losses from the skin and respiratory tract also are a source of hypotonic fluid loss. Increased fluid loss can occur in febrile patients (skin and respiratory tract), patients on mechanical ventilation (respiratory tract), and patients with profuse sweating. The sweat sodium concentration decreases with increasing volumes of perspiration. These conditions also will lead to hypernatremia with body fluid hypertonicity only if the thirst mechanism or access to water is impaired.

TABLE 116-7 CAUSES OF HYPERNATREMIA CLASSIFIED BY TOTAL BODY SODIUM CONTENT

Hypervolemia	Hypertonic saline excess Hypertonic sodium bicarbonate solutions
Hypertonicity with near normovolemia	Diabetes insipidus Febrile fluid loss
Hypovolemia	Gastrointestinal loss (diarrhea, vomiting) Skin fluid loss (burn, sweat) Loop diuretics Osmotic diuresis Impaired thirst perception

Hypovolemic hypernatremia is by far the most common hypertonicity state. Patients with hypovolemic hypernatremia have lost both sodium and water, but the net loss of water is disproportionately greater than the net loss of sodium. The actual plasma sodium concentration resulting from loss of hypotonic fluid depends not only on the sodium concentration of the fluid lost but also on the concentration of other osmotically active solutes, such as potassium, and on the solute composition of concomitantly ingested or administered fluids. The extrarenal and renal causes of such fluid losses are similar to those of isotonic hypovolemia. Among gastrointestinal causes of hypovolemic hypernatremia, diarrhea is more common than vomiting, and osmotic diarrheas result in disproportionately greater loss of water than electrolytes, with a greater propensity to hypernatremia than is seen with secretory diarrheas. Among the renal sources of sodium and water loss, the two most common causes are loop natriuretic medications and osmotic diuresis. Loop natriuretic agents interfere with the countercurrent mechanism and generate large volumes of urine with an iso-osmolar composition. Because some of the solutes are non-electrolyte (urea), the impact on body tonicity may be to increase tonicity, unless there is concomitant intake or administration of hypotonic fluids. In contrast, thiazides do not interfere with the countercurrent mechanism and therefore rarely promote hypernatremia. The presence of non-electrolyte solutes in urine causes an osmotic diuresis. Such solutes can be either endogenous (e.g., urea or glucose) or exogenous (e.g., mannitol or glycerol). The presence of these solutes in tubular fluid impairs both sodium and water reabsorption, but the excretion of urine that is relatively rich in non-electrolyte solutes tends to promote body fluid hypertonicity, unless sufficient hypotonic fluids are ingested or administered concomitantly.

Failure to replace hypotonic fluid losses generally reflects impairment in thirst, disability or infirmity that prevents the patient from responding to thirst, or failure of the clinician to recognize the need for hypotonic fluid replacement. Rarely, impaired thirst in patients who are awake and alert can be caused by damage to the hypothalamic osmoreceptors that control thirst perception and response, a condition known as *primary hypodipsia*. This condition usually tends to be associated with an abnormality in the osmotic regulation of AVP secretion. However, cases have been described in which the osmotic regulation of AVP secretion has been dissociated from the osmotic regulation of thirst. Such patients suffer hypernatremia only when extrarenal fluid losses exceed their habitual water intake, as might occur in settings of thermal stress or exercise.

CLINICAL MANIFESTATIONS

The clinical features of patients with hypernatremia can be divided into those associated with the underlying disease state, those associated with a

concomitant disturbance in ECF volume, and those associated with an increase in body fluid tonicity. The main clinically relevant consequence of increased body fluid tonicity is decreased brain cell volume, with the attendant risk for intracerebral hemorrhage. Thus, the major symptoms are neurologic and include confusion, seizures, focal neurologic deficits, and a progressively decreasing level of consciousness that can progress to coma. In the absence of an underlying neurologic problem or disturbance in the thirst mechanism, the patient would be expected to complain of thirst unless the neurologic injury has disturbed consciousness.

In patients with hypernatremia of sufficient duration to enable brain cells to undergo osmotic adaptation, the risk for intracerebral hemorrhage from cell shrinkage is decreased, but a hypertonic intracellular environment with the accumulation of new intracellular solutes can perturb normal cellular function. However, few if any clinical manifestations would be observed in this situation.

DIAGNOSIS

The diagnosis of hypernatremia is made by laboratory testing of the sodium concentration, which always should be repeated to confirm its accuracy, corroborated by measurement of plasma osmolality, which is expected to be elevated in all cases. The underlying cause of the hypernatremia is usually evident from the history and physical examination. The history should include a review of recent and current medication use and questions about exercise, heat exposure, sweating, vomiting, diarrhea, urine output, recent fluid intake, and presence of thirst. Physical examination should include an assessment of ECF volume and a complete neurologic evaluation. Urine volume should be monitored, urine osmolality should be measured in several spot urine samples, and 24-hour urine osmolar excretion should be measured if polyuria is present. In the less common situation of hypervolemic hypernatremia, there is often an antecedent history of the administration of sodium-containing solutions, and the findings on physical examination are consistent with ECF volume expansion. In the absence of underlying intrinsic renal disease or diuretic action, urine osmolality should be high because of the hypertonic stimulus to AVP release, which overrides the attenuating effect of hypervolemia. In such patients, the urine sodium concentration should be elevated in response to hypervolemia.

In the more common condition of hypovolemic hypernatremia with extrarenal fluid loss, urine output should be reduced to less than 500 mL/day, and urine osmolality should be the maximum expected for age (>1000 mOsm/kg in young adulthood decreasing to >600 mOsm/kg by the seventh decade of life and beyond). Polyuria with a submaximal urine osmolality in the presence of hypernatremia suggests impaired urine-concentrating ability, such as occurs with preexisting or underlying intrinsic renal disease or exposure to diuretic agents. A spot urine osmolality measurement of less than 100 to 200 mOsm/kg or polyuria (>3 L/day) together with 24-hour urine solute excretion of less than 600 mOsm/day in the face of hypernatremia suggests diabetes insipidus. In contrast, daily solute excretion exceeding 800 to 1000 mOsm/day suggests an osmotic diuresis, which can be confirmed by measurement of glucose and urea in urine.

TREATMENT

Rx

The main components of treatment are to correct the underlying disorder and the abnormality in ECF volume, to replace the water deficit, and to provide maintenance fluids to match continuing ongoing fluid losses if they persist.

The management of serious symptomatic hypovolemic hypernatremia is challenging and often controversial. It is best to divide the therapeutic approach into two separate phases: rapid correction of the depleted ECF volume, followed by gradual replacement of the water deficit, including provision for ongoing fluid losses. When ECF volume contraction is severe, as evidenced by tissue hypoperfusion and shock, administered fluid should have a sodium concentration as close as possible to that of the patient and should distribute to the ECF, especially the intravascular compartment. Isotonic saline is generally the fluid of choice, and the volume and rate of administration should be guided by clinical parameters related to reversal of hypovolemia. After the patient's tissue perfusion has been restored, further fluid replacement should be aimed at correction of the estimated water deficit. This estimate begins with a simple calculation of the percentage deficit based on the measured sodium concentration:

$$\text{Total body water deficit} = 0.4 \times \text{premorbid weight} \times ([\text{Na}/140] - 1)$$

Total body water is used because the sodium concentration reflects tonicity in all body fluid compartments, including the ICF. Unlike the isotonic fluid replacement for ECF volume, the water replacement should be administered gradually during a period of hours to days, unless there is clear documentation that the hypernatremia has itself evolved during minutes to hours. The necessity for gradual replacement is dictated by the process of osmotic adaptation described previously, and, ideally, the rate of water replacement should match the rate at which brain intracellular solutes can be adaptively extruded or removed. More rapid rates of administration could result in brain cell swelling with attendant dangerous neurologic consequences. It is recommended that the estimated volume of the water deficit be replaced at a rate that will lead to a reduction of approximately 0.5 to 1.0 mmol/L in measured plasma sodium concentration per hour and no more than 10 to 12 mmol/L but no less than 6 mmol/L in 24 hours. In addition to the estimated water deficit, the estimated ongoing water loss during replacement should include at least 1 L per 24 hours of insensible fluid losses (greater volumes in patients who are febrile or mechanically ventilated), supplemented with any ongoing water losses (renal or gastrointestinal) resulting from continuation of the underlying disease process. Because of the need to distribute replacement of the initial water deficit, which can amount to several liters in a number of days during which ongoing water losses continue, it is not unusual for patients to require large volumes of water, sometimes reaching 5 to 10 L, for the duration of the correction period. This water deficit, together with ongoing losses, can be replaced by the dietary ingestion of tap water, if the patient's condition is suitable, or by an enteral feeding tube. If a gastrointestinal or other disease process precludes these preferred routes, a hypotonic intravenous solution such as D_5W or half-isotonic saline can be used. When D_5W is used, the glucose is either stored as glycogen or fat or metabolized into carbon dioxide and water, thus effectively providing the patient with solute-free water replacement. In the case of half-isotonic saline, for any given liter administered, only half can be considered as replacement of the water deficit, and the sodium content will either replace any remaining sodium deficit that has not been fully corrected in the first phase of treatment or be excreted if there is no impairment in urinary sodium excretion. In elderly patients with known or possible underlying cardiac, hepatic, or renal disease, caution should be exercised in the provision of excessive volumes of salt-containing solutions. In any case, the sodium concentration should be monitored at regular intervals of no less than every 4 hours to avoid too slow or too rapid correction, and ECF volume parameters should be monitored to avoid hypervolemic complications.

Special considerations apply for hypertonic states in the setting of uncontrolled diabetes with hyperglycemia (Chapter 229). The unusual cases of patients with hypervolemic hypernatremia in the hospital setting also need special attention and sometimes require continuous infusions of loop diuretics together with the administration of hypotonic solutions or, in some cases, extracorporeal means to remove both the sodium and water excess in a controlled and safe manner under careful monitoring, preferably in the intensive care unit.

The route of administration should change in accordance with the patient's response. Although an initial parenteral or nasogastric enteral route might be appropriate when the patient's neurologic status is compromised, subsequent therapy can consist of simple dietary intake of water. Once a patient is awake and alert, and if thirst mechanisms are intact, the patient will generally correct the hypertonic state by spontaneous oral fluid intake.

Polyuria

Polyuria (Table 116-8), which is defined as a urine output of more than 3 L/day, should be distinguished from urinary frequency, which can occur with frequent voiding of small volumes totaling less than this amount per day. Polyuria occurs when urine-concentrating mechanisms are not being used at any time of the day (water diuresis) or urine solute excretion is excessive (solute diuresis).

SOLUTE DIURESIS

Polyuria in association with a urine osmolality of 300 mOsm/kg or more generally indicates solute (or osmotic) diuresis. On a typical Western diet, solute excretion (mainly sodium, potassium, and urea) is 600 to 900 mOsm/day. Therefore, normal maximum urine output cannot be greater than 3 L (= 900 mOsm/300 mOsm/kg). Hence, if urine volume exceeds 3 L/day in the presence of a urine osmolality greater than 300 mOsm/kg, extra solute must be present in the urine. The composition of these excess solutes can be electrolyte or non-electrolyte. Electrolyte solute diuresis usually occurs in response to the iatrogenic administration of high volumes of electrolyte-containing solutions, which are eliminated by the kidney through normal physiologic mechanisms. Non-electrolyte solute diuresis (glucose or urea, resulting from hyperglycemia or high-protein feeding, respectively) is equivalent to osmotic diuresis in which the presence of a non-reabsorbable non-electrolyte solute in the tubular fluid prevents reabsorption of sodium and other electrolytes as well as water.

TABLE 116-8 REASONS FOR POLYURIA

WATER DIURESIS
Diabetes insipidus
Central (neurogenic)
Inherited
Acquired (e.g., tumors, trauma, hypoxia)
Nephrogenic
Hypercalcemia
Amyloidosis
Drugs (e.g., lithium, foscarnet, cidofovir, vaptans)
Sjögren syndrome
Sickle cell disease
Inherited
Polydipsia
Primary (e.g., hypothalamic)
Psychogenic
SOLUTE DIURESIS
Sodium
Excess sodium intake (oral, enteral, parenteral)
Renal sodium wasting (e.g., inherited tubulopathies, interstitial nephritis, natriuretic drugs)
Anion based (sodium is usually the associated cation)
Chloride excretion (e.g., Bartter syndrome, loop diuretic)
Bicarbonate excretion (e.g., exogenous bicarbonate, carbonic anhydrase inhibition)
Glucose/keto acids
Diabetic ketoacidosis
Hyperglycemic-hyperosmolar syndrome
Renal glycosuria
Sugar alcohols
External loading (e.g., mannitol, glycerol)
Urea
Exogenous loading (e.g., urea, protein, amino acids)
Diuretic phase of acute kidney injury
Post-obstructive diuresis
Hypercatabolic states
Hemoglobin/myoglobin driven (post-rhabdomyolysis or reabsorption of a hematoma)
Other
Radiocontrast agents

WATER DIURESIS

When polyuria is associated with a urine osmolality of less than 250 mOsm/kg, a defect in urine-concentrating ability is generally suggested. In some cases, this defect occurs in association with a more general state of intrinsic renal injury and can be part of the spectrum of interstitial injury in chronic renal disease. More specific defects in urine-concentrating ability fall into the category of diabetes insipidus (Chapter 225).

TREATMENT

Rx

Once a patient with polyuria has been classified as having a solute or water diuresis (see Table 116-8), the clinical manifestations and treatment will be those of the underlying disease, and the consequences of changes in ECF volume and tonicity are the same as discussed earlier. Depending on the nature of fluid intake and medication used at the onset of polyuria, a significant percentage of polyuric patients will have alterations in plasma sodium and ECF volume and will need attention to the underlying disease as well as correction of fluid and electrolyte abnormalities. Thus, for example, although antihyperglycemic treatment effectively corrects the solute diuresis and polyuric state of uncontrolled diabetes mellitus, initial correction of the concomitant electrolyte and ECF volume disorders takes precedence (Chapter 229).

Grade A References

A1. Prowle JR, Chua HR, Bagshaw SM, et al. Clinical review: volume of fluid resuscitation and the incidence of acute kidney injury—a systematic review. *Crit Care*. 2012;16:230.

A2. Roberts I, Blackhall K, Alderson P, et al. Human albumin solution for resuscitation and volume expansion in critically ill patients. *Cochrane Database Syst Rev*. 2011;11:CD001208.

A3. Myburgh JA, Finfer S, Bellomo R, et al. Hydroxyethyl starch or saline for fluid resuscitation in intensive care. *N Engl J Med*. 2012;367:1901-1911.

A4. Aliti GB, Rabelo ER, Clausell N, et al. Aggressive fluid and sodium restriction in acute decompensated heart failure: a randomized clinical trial. *JAMA Intern Med*. 2013;173:1058-1064.

A5. Paterna S, Fasullo S, Parrinello G, et al. Short-term effects of hypertonic saline solution in acute heart failure and long-term effects of a moderate sodium restriction in patients with compensated heart failure with New York Heart Association class III (Class C) (SMAC-HF Study). *Am J Med Sci*. 2011;342:27-37.

A6. Licata G, Tuttolomondo A, Licata A, et al. Clinical Trial: High-dose furosemide plus small-volume hypertonic saline solutions vs. repeated paracentesis as treatment of refractory ascites. *Aliment Pharmacol Ther*. 2009;30:227-235.

A7. McMurray JJ, Packer M, Desai AS, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. *N Engl J Med*. 2014;371:993-1004.

A8. Udelson JE, Bilsker M, Hauptman PJ, et al. A multicenter, randomized, double-blind, placebo-controlled study of tolvaptan monotherapy compared to furosemide and the combination of tolvaptan and furosemide in patients with heart failure and systolic dysfunction. *J Card Fail*. 2011;17:973-981.

A9. Ghali JK, Orlandi C, Abraham WT. The efficacy and safety of lixivaptan in outpatients with heart failure and volume overload: results of a multicentre, randomized, double-blind, placebo-controlled, parallel-group study. *Eur J Heart Fail*. 2012;14:642-651.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Titze J, Dahlmann A, Lerchl K, et al. Spooky sodium balance. *Kidney Int*. 2014;85:759-767.
2. Schrier RW. Diagnostic value of urinary sodium, chloride, urea, and flow. *J Am Soc Nephrol*. 2011;22:1610-1613.
3. Prowle JR, Kirwan CJ, Bellomo R. Fluid management for the prevention and attenuation of acute kidney injury. *Nat Rev Nephrol*. 2014;10:37-47.
4. Neville KA, Sandeman DJ, Rubinstein A, et al. Prevention of hyponatremia during maintenance intravenous fluid administration: a prospective randomized study of fluid type versus fluid rate. *J Pediatr*. 2010;156:313-319.
5. Yunos NM, Bellomo R, Hegarty C, et al. Association between a chloride-liberal vs chloride-restrictive intravenous fluid administration strategy and kidney injury in critically ill adults. *JAMA*. 2012;308:1566-1572.
6. Jentzer JC, DeWald TA, Hernandez AF. Combination of loop diuretics with thiazide-type diuretics in heart failure. *J Am Coll Cardiol*. 2010;56:1527-1534.
7. Adrogue HJ, Madias NE. The challenge of hyponatremia. *J Am Soc Nephrol*. 2012;23:1140-1148.
8. Lindner G, Funk GC. Hypernatremia in critically ill patients. *J Crit Care*. 2013;28:216.
9. Kovesdy CP, Lott EH, Lu JL, et al. Hyponatremia, hypernatremia, and mortality in patients with chronic kidney disease with and without congestive heart failure. *Circulation*. 2012;125:677-684.
10. Singh TD, Fugate JE, Rabinstein AA. Central pontine and extrapontine myelinolysis: a systematic review. *Eur J Neurol*. 2014;21:1443-1450.
11. Hoorn EJ, Liamis G, Zietse R, et al. Hyponatremia and bone: an emerging relationship. *Nat Rev Endocrinol*. 2012;8:33-39.
12. Lehrich RW, Ortiz-Melo DI, Patel MB, et al. Role of vaptans in the management of hyponatremia. *Am J Kidney Dis*. 2013;62:364-376.
13. Peri A. Clinical review: the use of vaptans in clinical endocrinology. *J Clin Endocrinol Metab*. 2013;98:1321-1332.

117

POTASSIUM DISORDERS

JULIAN L. SEIFTER

DEFINITION

Maintenance of a normal and narrow range of blood plasma potassium concentration, usually on the order of 3.5 to 5.0 mmol/L, is vital for health. *Hyperkalemia* refers to an increased plasma potassium concentration and *hypokalemia* to a decreased concentration. Within the human body, potassium is not equally distributed in the total body water. Approximately two thirds of body water is intracellular, and potassium is the major cation within that compartment, reaching concentrations as high as 140 mEq/L. Consequently, more than 98% of potassium resides within cells. An *excess* of total body potassium stores is less common than potassium *depletion* unless renal function is compromised.

Discordance between total body stores and the plasma concentration can cause hyperkalemia despite potassium depletion, and it can cause hypokalemia even with potassium excess. Potassium *adaptation* defines changes in regulatory mechanisms resulting from potassium excesses or deficits.

EPIDEMIOLOGY

Potassium is ubiquitous in both plant and animal dietary sources, so avoiding potassium is difficult. Nevertheless, disorders of potassium balance are common in both inpatient and outpatient settings. Diets high in potassium and low in sodium are associated with lower blood pressure and decrease the risk of cardiovascular disease, including stroke. By comparison, hypokalemia is associated with increased mortality, especially in patients with heart disease,[1] whereas renal dialysis patients with hyperkalemia have a higher cardiovascular mortality.[2] Mild decreases in plasma potassium are seen in healthy well-trained athletes and during normal pregnancy. Patients with chronic kidney disease (Chapter 130) and insulin-deficient diabetes (Chapter 229) have a tendency to development of hyperkalemia in association with high-potassium diets or treatment with medications that interfere with potassium balance. Starvation, gastrointestinal disease, commonly used diuretics, and other medications may cause hypokalemia. Some renal disorders, such as urinary tract obstruction (Chapter 123), may be associated with hyperkalemia, whereas others, such as aminoglycoside nephrotoxic injury, result in hypokalemia. Endocrine diseases, including adrenal insufficiency (Chapter 227), characteristically cause hyperkalemia, whereas hypokalemia is a common finding in patients who have adrenal aldosterone-secreting adenomas or tumors with ectopic secretion of adrenocorticotropic hormone (ACTH).

PATHOBIOLOGY

Potassium Balance

Most cells express sodium-potassium adenosine triphosphatase (Na^+, K^+-ATPase) on the cell plasma membranes and thereby use metabolic energy in the form of ATP to develop gradients of potassium and sodium. As a result, cellular potassium may exceed extracellular concentrations 35-fold. Established electrochemical gradients enable normal muscle and neural function as well as facilitate cellular nutrient uptake and transcellular solute transport in the intestine and kidney. Potassium entry into cells is balanced by extrusion by potassium channels and transporters in a regulated and coordinated fashion.

The total amount of potassium is usually on the order of 50 mmol/kg of body weight, so a 70-kg individual has a store of about 3500 mmol of potassium, mostly in skeletal muscle. By comparison, the entire extracellular fluid, which is approximately 20% of body weight, or 14 L in a 70-kg person, may have a potassium content of only 50 to 60 mmol. As a result, total body potassium is poorly reflected by the extracellular or plasma potassium concentration. Changes in the distribution between the cells and extracellular fluid can occur rapidly, within minutes, in contrast to the matching of dietary potassium intake to potassium elimination from the body, which occurs within hours. Potassium ingestion at the time of a meal may be equal to a large fraction of the total extracellular potassium. An average daily consumption on the order of 50 to 100 mmol would cause a rapid rise in extracellular potassium concentration after meals if it were not for the ability of potassium to distribute rapidly from the extracellular space to the intracellular space. The intracellular space, given its large volume and potassium content, can accommodate, or buffer, an extra load of potassium without significant changes in plasma or cellular concentration.

The Importance of Potassium

Potassium is essential for a number of critical body functions, including enzymatic reactions that regulate protein and glycogen synthesis, as well as for cell growth and division. The ability of cells to take up or to extrude potassium contributes to the regulation of cell volume during periods of osmotic stress. In excitable cells, such as cardiac myocytes, the relationship of intracellular to extracellular potassium concentration is critical in establishing the resting membrane potential. Because of the relative magnitude of cellular and extracellular concentrations, larger percentage changes tend to occur in the extracellular potassium concentration, which consequently has the greatest impact on the electrical properties across cell membranes.

Potassium has critical effects on excitable tissues, especially cardiac and skeletal muscle. A low serum potassium concentration not only hyperpolarizes most cells, thereby leading to an increase in the resting potential, but also alters potassium channels required for repolarization (Chapter 61). Thus, hypokalemia decreases or slows potassium conductance in some potassium channels.

Because of an increased potassium conductance, hyperkalemia antagonizes the normal slow depolarization of pacemaker tissue that is usually associated with a decrease in potassium conductance. Certain muscle-depolarizing anesthetic agents, such as succinylcholine, may potentiate the effects of hyperkalemia, as may gentamicin, particularly in patients with renal failure.

Potassium is as important a body fluid osmole as sodium, and losses of potassium obligate sodium to replace it in the intracellular space, thereby resulting in hyponatremia (Chapter 116). Hypokalemia contributes to the hyponatremia associated with thiazide diuretics. Iso-osmotic losses of combined potassium and sodium salts in watery diarrhea (Chapter 140) may result in isotonic extracellular volume depletion. Similarly, if potassium chloride is added to isotonic saline, a hypertonic solution results, and potassium may enter cells as sodium exits, thereby contributing to hypernatremia.

Potassium is also an important local mediator of vascular tone in muscle beds. During exercise, local interstitial fluid potassium concentrations may rise 10-fold, thereby causing local vasodilation to allow more blood supply to the exercising muscle but also resulting in sarcolemmal depolarization that creates muscle fatigue. Very little of that potassium enters the total extracellular fluid, so severe hyperkalemia does not usually occur with exercise. The trained athlete develops an adaptive increase in Na^+, K^+-ATPase to allow efficient re-uptake of potassium into muscle cells. For example, experienced marathoners know that their diet must provide adequate potassium stores needed for muscle endurance because overexertion during a state of potassium depletion can lead to rhabdomyolysis (Chapter 113).

Potassium excretion[3] follows a circadian rhythm. This rhythm appears to be a feed-forward mechanism that anticipates the highest level of urinary potassium excretion to coincide with dietary intake of potassium. The relationship likely involves potassium sensing within the gastrointestinal tract or splanchnic or hepatic circulation.[4] In healthy humans, potassium excretion is greatest during the daytime hours, and it can increase several-fold, without requiring a change in the blood potassium level. Independent of aldosterone, a relationship between increased potassium and sodium excretion may account for the beneficial association of high-potassium diets with lower blood pressures. The distribution of potassium between the extracellular and intracellular spaces also shows a circadian variation that may contribute to the diurnal variance in cardiac arrhythmias and sudden death in patients at risk.

Renal Potassium Handling

In the kidney, potassium excretion begins with filtration. Because the extracellular concentration of potassium is approximately 4 mmol/L and that of sodium is 140 mmol/L, far less potassium is filtered than sodium (about 3%). The renal proximal tubule reabsorbs potassium, primarily by the paracellular pathway, in the process of reabsorbing sodium and water. In the thick ascending limb of the loop of Henle, potassium is reabsorbed both by the apical sodium–potassium–2 chloride cotransporter (NKCC) and, like calcium and magnesium, by paracellular reabsorption of the cation. The latter mechanism is a consequence of the electropositive lumen created by the recycling of potassium from the cell to the lumen through renal outer medullary potassium channels (ROMK). Potassium that is reabsorbed in the thick ascending limb re-enters the tubular fluid when it is secreted into the thin descending limb in a process known as medullary potassium recycling. The resulting high interstitial potassium concentrations may enable potassium excretion from the medullary collecting duct by limiting potassium backleak. Luminal ammonium (NH_4^+) can substitute for potassium on the thick limb NKCC; the resulting increase in medullary interstitial fluid NH_4^+ concentrations enhances medullary collecting duct net acid excretion. In hyperkalemic states, less NH_4^+ appears in the urine because the high concentration of luminal potassium competes with NH_4^+ for reabsorption in the thick limb. As a consequence, metabolic acidosis may develop in hyperkalemic states, and lowering of elevated plasma potassium levels helps treat acidosis.

In hypokalemia, whether through proximal tubule intracellular acidosis or other mechanisms, glutaminase enzymes are increased, and more ammonia is produced. This ammonia leads to greater medullary interstitial fluid concentrations and therefore to enhanced net acid elimination. Ammonium production in hypokalemia could be considered an adaptation to allow potassium to be reabsorbed as NH_4^+ accompanies excreted anions into the urine. The increase in ammonia production may have a deleterious effect in that it may contribute to the chronic tubulointerstitial nephritis of chronic hypokalemia.

By the time the tubular fluid reaches the distal tubule and collecting duct, more than 90% of potassium has been reabsorbed. In potassium depletion, an increase in the apical membrane hydrogen-potassium ATPase (H^+, K^+-ATPase) of the collecting duct intercalated cell allows near-complete removal of potassium from the urine. However, potassium reabsorption is seldom as complete as that of sodium. It is unusual to see potassium concentrations in the urine lower than 5 to 10 mmol/L. When dietary potassium is abundant, the reabsorption of 90% of filtered potassium by the proximal and distal nephron is followed by net potassium secretion.

The secretion of potassium in the cortical collecting duct (E-Fig. 117-1), which may vary according to need by as much as 400%, is controlled by three major mechanisms: (1) development of a lumen-negative transepithelial potential difference that provides the driving force for potassium secretion into the lumen, (2) regulated apical membrane secretory potassium channels, and (3) tubular fluid flow dependency. The urinary potassium most closely reflects potassium secreted by the distal nephron.

The Aldosterone Paradox

The traditional view of regulating potassium secretion and therefore excretion has focused on the central role of aldosterone, the steroid hormone synthesized and secreted by the zona glomerulosa of the adrenal cortex.[5] Aldosterone is stimulated by angiotensin II, predominantly in the hyperreninemic states of extracellular volume depletion, and independently by potassium loading, usually from an excess of potassium in the diet. The renal effects of aldosterone show overlap of functions to conserve sodium and to eliminate potassium. In what has been called the *aldosterone paradox*, the kidney can prevent undesired losses of potassium from occurring when sodium retention is required to maintain extracellular volume (the angiotensin II stimulus), and it also can prevent undesired retention of sodium from occurring while maintaining normal potassium balance when the stimulus to

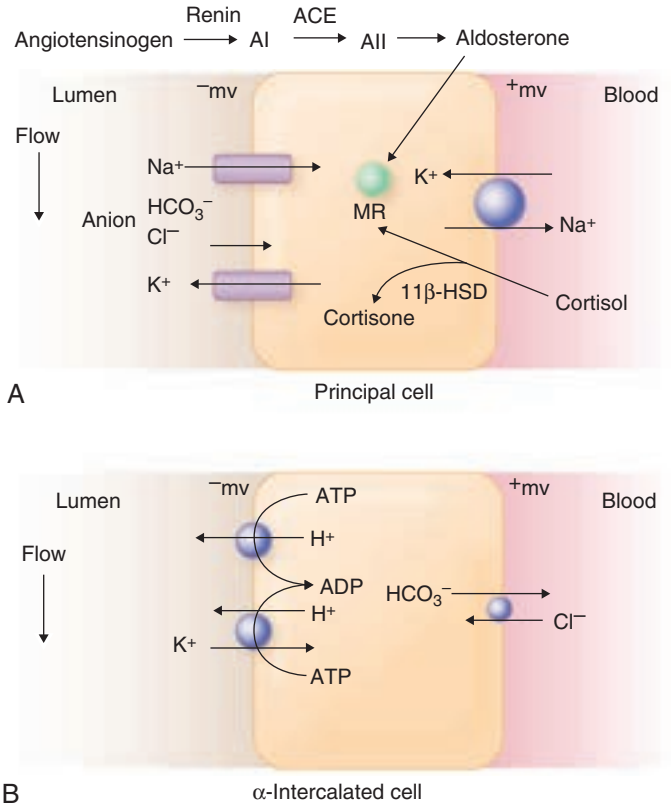

E-FIGURE 117-1. Two cell types in the cortical collecting duct. A, The principal cells mediate sodium (Na^+) reabsorption energized by the basolateral Na^+, K^+ pump. Entry from the lumen is through the epithelial Na^+ channel (ENaC), which renders the lumen negatively charged ($^-$mv). This transepithelial voltage stimulates secretion of potassium (K^+) through renal outer medullary potassium (ROMK) channels. Reabsorbable anions such as chloride (Cl^-) lessen the luminal negativity and decrease K^+ secretion. Bicarbonate (HCO_3^-) has an effect to increase K^+ secretion. High flow rates increase net K^+ secretion by activating maxi-K or big-K channels and by preventing development of high K^+ concentrations in the lumen. The effects of the renin-angiotensin-aldosterone axis are shown: increased mineralocorticoid receptor (MR) activation increases ENaC, the Na^+, K^+ pump, and K^+ channels, thereby increasing Na^+ reabsorption and K^+ secretion. Cortisol would also increase MR activity, but it is inactivated by the enzyme 11β-hydroxysteroid dehydrogenase (11β-HSD). AI = angiotensin I; AII = angiotensin II; ACE = angiotensin-converting enzyme; mv = millivolts. **B,** Intercalated cells are carbonic anhydrase–rich cells that secrete acid and reabsorb HCO_3^-. The H^+-ATPase secretes H^+ in a way favored by the negatively charged lumen in conjunction with the aldosterone-stimulated effect on Na^+ reabsorption in neighboring principal cells. The K^+, H^+-ATPase is an electroneutral pump. The K^+/H^+ exchanger, reabsorbing K^+, is important in states of K^+ depletion when urinary K^+ is decreased. Intercalated cells also have apical big-K channels, indicating a role in K^+ secretion. ADP = adenosine diphosphate; ATP = adenosine triphosphate.

aldosterone is potassium loading. The feedback mechanism coupling aldosterone with serum potassium is important in regulating the degree of potassium losses in the urine. For example, hyperaldosteronism results in potassium loss, and then the resulting hypokalemia reduces aldosterone production and subsequent potassium losses. Hyperkalemia has the opposite effect: it is an important stimulus of aldosterone synthesis and release. Thus, with volume expansion and low angiotensin II, the rise in potassium stimulates aldosterone release from the zona glomerulosa of the adrenal cortex, thereby allowing potassium to be secreted into the urine.

Mechanism of Sodium Reabsorption and Electrochemical Forces

The major regulatory site for potassium secretion resides in the aldosterone-sensitive distal nephron, which is composed of the late distal convoluted tubule as well as the principal cells of the connecting tubule and the cortical collecting duct. Aldosterone regulates sodium reabsorption in these segments: by the NaCl cotransporter in the late distal convoluted tubule and connecting tubule, and by the apical epithelial sodium channel (ENaC) in the principal cells of the connecting tubule and cortical collecting duct. Thiazide-sensitive NaCl cotransport is independent of aldosterone in the early distal convoluted tubule.

To optimize potassium secretion, sodium must be delivered in ample amounts (depending on the flow rate and sodium concentration of the tubular fluid) to result in sodium reabsorption through the apical ENaC in the principal cells. In severe prerenal states of avid sodium reabsorption, including hypovolemic states, the hepatorenal syndrome, and severe heart failure, sodium delivery from more proximal sites may become rate limiting for potassium secretion. Assuming that sodium delivery is not limiting, reabsorption of the sodium cation creates a lumen-negative transepithelial potential difference. Aldosterone affects the transepithelial potential difference in several ways. The intracellular mineralocorticoid receptor functions to increase the activity and density of ENaC and of basolateral Na^+, K^+-ATPase enzymes. Cortisol, which is normally present in higher concentrations than aldosterone, has equal affinity for the aldosterone receptor and therefore could lead to increased ENaC activity. However, the enzyme 11β-hydroxysteroid dehydrogenase type 2 is present in the principal cells and converts cortisol to inactive cortisone.

The reabsorption of sodium is dependent on the low intracellular sodium concentrations that result from the energy-requiring Na^+, K^+-ATPase on the basolateral membrane. Both increased intracellular Na^+ and extracellular K^+ also stimulate the Na^+, K^+-ATPase. Increased Na^+ entry through an apical mechanism will increase Na^+, K^+-ATPase, thereby bringing more potassium into the cells for transepithelial secretion, independent of aldosterone. However, oral potassium loads in the intestine may increase potassium secretion without prior increments in plasma potassium.

Most of the potassium that enters the principal cell from the extracellular fluid through the Na^+, K^+-ATPase is secreted into the lumen because of the lumen-negative transepithelial voltage. Some is then either recycled back to the extracellular fluid by basolateral potassium transport mechanisms or, after secretion into the cortical collecting duct lumen, reabsorbed back to the blood by neighboring intercalated cell H^+, K^+-ATPase.

Role of Potassium Secretory Channels and Tubular Flow Rate

Optimal potassium secretion requires adequate function of several types of potassium secretory channels on the luminal membranes of distal nephron cells. Two predominant potassium channels are present in the apical membranes of the collecting duct cells.

The renal outer medullary potassium channel (ROMK), which is located on the apical membrane of the principal cell in the connecting tubule and cortical collecting duct, is aldosterone sensitive, is increased by dietary potassium loads, and likely subserves constitutive K^+ secretion under basal conditions. ROMK undergoes new synthesis and increased cycling to the apical membrane under the control of aldosterone. The principal cell also mediates vasopressin-responsive water reabsorption. Factors regulating ROMK channels include antidiuretic hormone (ADH) and intracellular pH. The increase in ROMK channel activity in response to ADH, combined with an ADH effect to increase ENaC and osmotic water flow, allows the highest possible potassium concentration in the urine with the low tubular flow that accompanies antidiuresis. When ADH is suppressed, luminal potassium concentrations are lower, thereby favoring gradients for potassium secretion, but ROMK activity is reduced because of the low ADH. Cellular acidification inhibits potassium secretion through an effect on ROMK, thereby providing a renal mechanism for decreased electrogenic potassium secretion when proton secretion is necessary during metabolic and respiratory acidoses. In alkalosis, ROMK activity is increased, thereby favoring increased potassium secretion, but the result may be significant potassium loss.

Additional potassium channels, known as maxi-K or big-K channels, are prominent in the principal cells and intercalated cells of the cortical collecting duct, thereby implying that the intercalated cell also has a role in potassium secretion. The big-K channels are highly regulated by tubular fluid flow rate. One mechanism by which flow rate influences this potassium secretory channel is by the deformation by flow of the primary cilium of the principal cell, which leads to a secondary increase in cellular calcium that has both direct and indirect effects to enhance BK potassium secretion.

Excretion Mechanisms and Normal Function

Because the major regulatory site for potassium secretion resides in the aldosterone-sensitive distal nephron, the mechanisms for Na^+ and K^+ transport in each of these cell types are central to the proposed role of a network of protein kinases, including the "with no lysine" or WNK kinases in the aldosterone paradox. Note that aldosterone regulates sodium reabsorption in all segments: NaCl cotransport in the late distal convoluted tubule and connecting tubule, and ENaC in the connecting tubule and principal cells of the cortical collecting duct. ROMK is increased by aldosterone in the cortical collecting duct.

In hypovolemia, sodium reabsorption is increased by angiotensin II in the proximal tubule, the early and late distal convoluted tubule, and the connecting tubule so that very little sodium is delivered to the principal cells of the cortical collecting duct. Potassium losses are diminished because of decreased delivery of sodium, and an angiotensin II–induced switch involving WNK kinase decreases ROMK activity. Although aldosterone will still activate principal cell ENaC, the decrease in sodium delivery and decreased potassium secretion (favoring chloride reabsorption instead), will allow greater NaCl reabsorption while limiting potassium loss.

When aldosterone is present in the absence of angiotensin II, as in potassium loading, there is no earlier stimulation of sodium reabsorption, so sodium delivery to late aldosterone-sensitive distal nephron segments that contain ROMK, the NaCl cotransporter, and ENaC is increased. In the absence of angiotensin II, the WNK network of protein kinases favors increased ROMK, whereas aldosterone increases ENaC activity. The result favors potassium secretion, but overall sodium reabsorption is not increased because increased NaCl delivery from proximal sites offsets the increase in sodium reabsorption through ENaC.

Internal Potassium Balance and Associated Disorders

Because of the delay in hours before renal excretion matches dietary intake and because potassium first enters the extracellular fluid from the gastrointestinal tract, it is critical that the process of cellular buffering be effective (Table 117-1). Essentially, increases in postprandial blood potassium are minimized before potassium is eliminated from the body. A major factor in this regulation after meals is the feedback loop involving insulin and potassium. An increase in serum potassium stimulates insulin release from the β cells of the pancreatic islets. Insulin increases potassium uptake into cells, primarily muscle, independent of its effect on glucose uptake. Acutely, potassium uptake is chiefly the result of increased Na^+, K^+-ATPase activity, whereas chronically there is increased abundance in the plasma membranes in these cells.

Another important mechanism of regulating the distribution of potassium between extracellular and cellular spaces involves the sympathoadrenal

TABLE 117-1 FACTORS REGULATING INTERNAL POTASSIUM BALANCE

Circadian rhythm
Insulin
β-Adrenergic activity
Acid-base balance
Magnesium
Aldosterone
Osmolality
Thyroid hormone
Extracellular potassium
Intracellular sodium

system. β-Adrenergic activation, particularly through the β_2-receptor, increases potassium uptake into muscle and fat cells. As with insulin, potassium uptake is the consequence of increased Na^+, K^+-ATPase activity associated with increased intracellular cyclic adenosine monophosphate. The adrenergic effect is important in regulating the serum potassium concentration during exercise and is independent of the additional effect that catecholamines may have on blood glucose with the expected increases in insulin. In the trained athlete, a chronic increase in Na^+, K^+-ATPase on cell membranes may cause a transient lowering of the serum potassium concentration after exertion. Conversely, a severe stress may contribute to hypokalemia, through both direct β_2 effects and insulin action secondary to the blood glucose rise.

Phenylephrine, an α-adrenergic agonist, increases serum potassium. Importantly, epinephrine, which also has α-adrenergic effects, is associated with a transient increase in potassium release from the liver before a more prolonged period of decreased serum potassium mediated by the β_2-receptor.

Metabolic acidosis raises the potassium level more than does respiratory acidosis; both metabolic alkalosis and respiratory alkalosis lower the potassium level. Anion gap acidosis does not raise the potassium level, probably because of the movement of the organic anion (e.g., lactate) from cells into the extracellular space with an accompanying proton, whereas the ingestion of chloride salts has the most profound effect on the potassium level because chloride is restricted to the extracellular space and protons enter cells in exchange for the exit of the potassium cation. The ingestion of excessive chloride salts of arginine and lysine is associated with hyperkalemic acidosis. ε-Aminocaproic acid has also been associated with hyperkalemia and is hypothesized to exchange for cellular potassium, much like the other cationic amino acids.

Just as multiple simultaneous acid-base disturbances lead to a single blood pH level, many processes that affect the net potassium concentration can be simultaneously present. Metabolic acidosis may be associated with diarrheal or urinary losses of potassium, so that the potassium concentration is low, not high. Metabolic acidosis in diabetes can also be associated with insulin deficiency and renal failure, in which case the plasma potassium level might be elevated despite osmotically driven urinary losses of potassium and total body potassium depletion.

Other hormonal effects on potassium include thyroid and growth hormone, but patients with disorders of these hormones do not usually have significant changes in their blood potassium levels. Some patients with hyperthyroidism may have mild hypokalemia, perhaps related to increased sympathetic activity. Growth states are associated with a greater need for potassium; for example, in normal pregnancy, the maternal potassium concentration may fall as the developing fetus grows.

Hypomagnesemia frequently accompanies hypokalemia. Both magnesium and potassium are found predominantly in cells, but Na^+, K^+-ATPase requires magnesium for function. If magnesium is deficient, potassium distributes more to the extracellular fluid, thereby masking the degree of potassium deficiency. Moreover, magnesium deficiency leads to renal potassium wasting, so potassium depletion is difficult to correct until magnesium is repleted.

External Potassium Balance and Associated Disorders

Normally, no more than 10 to 20% of total potassium excretion is accomplished by the gastrointestinal tract, but colonic excretion is increased in renal failure, primarily through potassium-induced increases in epithelial Na^+, K^+-ATPase activity and aldosterone. In renal failure, the normal mechanisms to distribute potassium acquire increased importance.

Some conditions that cause the greatest losses of gastrointestinal potassium include secretory diarrheas of the colon, the result of infection or laxative abuse. Disorders of the small intestine, which may lead to large quantities of liquid stool with a low potassium concentration, engender favorable gradients for marked potassium secretion by the colon. A syndrome of watery diarrhea and hypokalemia is associated with neuroendocrine tumors (Chapter 195) that secrete vasoactive intestinal peptide. Rectosigmoid secretion of potassium may result in particularly high potassium losses, and potassium deficiency is seen in patients who have ureterosigmoidostomies. Potassium can be lost from a variety of other sources, including excess sweat or salivation, vomiting, and diarrhea.

Potassium can be depleted by vomiting or diarrhea. Urinary losses may exceed intake in renal tubular disorders, when excessive quantities of osmotic or anionic products are excreted in the urine, or in patients who take diuretics.

It is unusual for hyperkalemia to be caused by excessive potassium intake unless the patient has renal dysfunction. However, in patients who have brisk hemolysis, internal hemorrhage, or rhabdomyolysis, particularly if the hemoglobinuria or myoglobinuria also results in acute kidney injury, life-threatening hyperkalemia can quickly develop by the rapid release of cellular potassium stores.

CLINICAL MANIFESTATIONS

Hypokalemia

Clinical manifestations of potassium depletion include hypertension, decreased growth, and muscle symptoms such as weakness, cramps, fasciculations, and even paralysis. In severe cases, the diaphragm may be paralyzed, leading to respiratory failure. Cardiac arrhythmias are a critical component of low potassium states and are usually seen when the serum potassium falls below 3 mmol/L or when ischemia, hypercalcemia, or drugs such as digoxin are simultaneously present. A patient who has a chronically low potassium level (e.g., from diuretic use) may be particularly vulnerable to supraventricular and ventricular tachyarrhythmias during periods of stress, such as head trauma, or during the acute coronary syndrome, to which cardiac ischemia also contributes. The prolonged cardiac repolarization phase of hypokalemia accounts for the characteristic electrocardiographic findings of broad, flattened T waves. U waves are also indicative of this delay in repolarization (Fig. 117-1). In the intestine, hypokalemia may result in paralytic ileus, which may interfere with oral replacement. Hypokalemia may result in acute skeletal muscle weakness and even paralysis.

In addition to these systemic effects of potassium imbalance, the kidney is particularly sensitive to depletion of potassium. Structural changes in the glomeruli and tubules lead to a decreased glomerular filtration rate, increased proximal tubule ammoniagenesis, increased sodium bicarbonate reabsorption, and net acid excretion, thereby causing metabolic alkalosis. A condition of nephrogenic diabetes insipidus results when potassium depletion decreases expression of vasopressin-dependent water channels (aquaporin 2) in the collecting duct luminal plasma membranes. Hypokalemia diminishes insulin secretion and may be associated with glucose intolerance.

Hyperkalemia

In hyperkalemia, the depolarizing effect on the resting membrane potential and increased potassium channel conductance lead to the classic

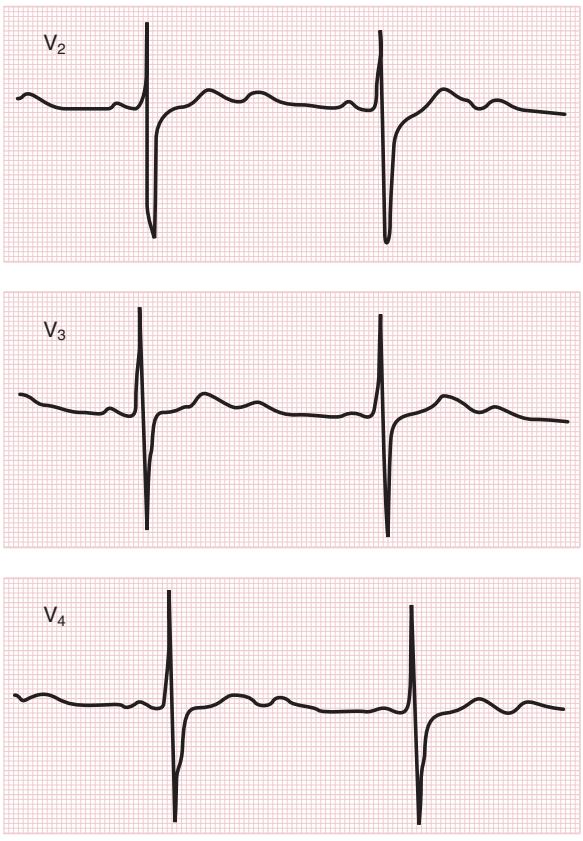

FIGURE 117-1. The electrocardiographic manifestations of hypokalemia. The serum potassium concentration was 2.2 mEq/L. The ST segment is prolonged, primarily because of a U wave following the T wave, and the T wave is flattened.

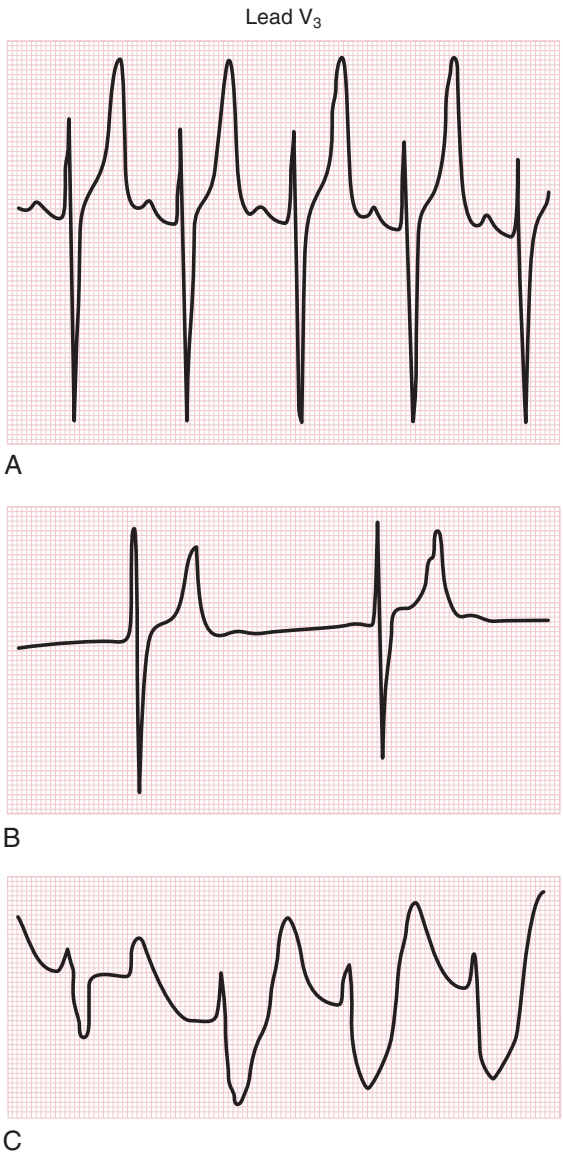

FIGURE 117-2. The effects of progressive hyperkalemia on the electrocardiogram. All of the illustrations are from lead V_3. **A,** Serum potassium concentration ($[K^+]$) = 6.8 mEq/L; note the peaked T waves together with normal sinus rhythm. **B,** Serum $[K^+]$ = 8.9 mEq/L; note the peaked T waves and absent P waves. **C,** Serum $[K^+] > 8.9$ mEq/L; note the classic sine wave with absent P waves, marked prolongation of the QRS complex, and peaked T waves.

electrocardiographic changes of hyperacute peaked T waves associated with rapid repolarization (Fig. 117-2). Hyperkalemia commonly results in sinus bradycardia. Heart block, loss of P waves on the electrocardiogram, and prolonged QRS intervals are all seen in cases of severe hyperkalemia, usually in excess of 6 mmol/L. The electrocardiogram, however, is not a sensitive indicator of severe hyperkalemia, and cardiac arrest may occur without warning. Like hypokalemia, severe hyperkalemia can cause skeletal muscle paralysis; unlike hypokalemic paralysis, it is often ascending in nature.

DIAGNOSIS

The first clue to a disorder in potassium balance usually is an abnormal serum potassium concentration obtained as part of a laboratory evaluation, not because an abnormal potassium level itself is suspected.[6] When the potassium concentration is elevated above normal, it is imperative to exclude common artifacts, known as pseudohyperkalemia. Hemolysis in the test tube is a common artifact; in cases of cold-induced hemolysis (Chapter 161), it is important to collect the blood and to allow it to clot in a warm environment. Some patients have pseudohyperkalemia resulting from high platelet counts, usually in excess of 1 million/μL, or myelogenous leukemia (Chapters 183 and 184); in such cases, potassium is released during clot formation in the test tube. Plasma potassium levels should be within the normal range. The serum potassium level may sometimes be elevated because of local ischemia related to application of the tourniquet and clenching of the fist.

A detailed medical history should focus on medications, family history, and sources of potassium excess or loss. The physical examination should pay particular attention to blood pressure, extracellular volume status, heart rate and rhythm, and muscle strength and reflexes.

Laboratory testing should include a complete blood count as well as serum levels of sodium, chloride, bicarbonate, creatinine, and blood urea nitrogen. In more serious cases, arterial blood gases and levels of creatine kinase and magnesium should be obtained. A 12-lead electrocardiogram also should be obtained.

A low urinary potassium level is expected in hypokalemia. In a hypokalemic patient, a urinary potassium concentration higher than 30 mEq/L suggests renal potassium wasting, whereas extrarenal losses are usually reflected by concentrations lower than 20 mEq/L. In a state of potassium excess, urinary potassium excretion should exceed about 35 mEq/L, unless urinary underexcretion was the cause of the hyperkalemia. A high aldosterone level causes a high urinary potassium-to-sodium ratio, whereas hypoaldosteronism may cause the opposite.

Hypokalemic Disorders

The most common cause of hypokalemia (Table 117-2) in medical practice is the use of thiazide or loop diuretics.[7] Patients may have low, normal, or high blood pressures, depending on their volume status and whether the diuretics were prescribed for hypertension or heart failure. The most common acute causes of hypokalemia are diarrhea and vomiting.

Hypokalemic Hypertensive Syndromes

If the renal principal cells develop a transepithelial electrical gradient that is more lumen negative than is needed to maintain potassium balance, urinary potassium wasting, inappropriate to the blood potassium level, occurs. Because a parallel increase in H^+ secretion will occur, it is common to see an accompanying metabolic alkalosis. If the abnormality is related to a primary increase in sodium reabsorption, hypertension or extracellular volume expansion also will develop.[8]

Evaluation of plasma renin and aldosterone levels can help distinguish among specific diagnoses (see Table 117-2).[9] Primary hyperaldosteronism is associated with low renin levels due to volume expansion. If it is corrected for plasma potassium, an aldosterone-to-renin ratio of 30:1 suggests a primary adrenal cortical tumor (aldosteronoma) or hyperplasia (Chapter 227). However, such ratios must be used with caution, and the absolute value of aldosterone is important, especially when the renin and aldosterone levels are both low. The tubular delivery of large amounts of sodium chloride in a setting of volume expansion and nonsuppressible aldosterone results in hypokalemia, which improves after sodium restriction and worsens with the administration of intravenous saline. In congenital adrenal hyperplasia (Chapter 227), such as 11β-hydroxylase deficiency, hypokalemic alkalosis is associated with excessive androgen production. In patients with renin-secreting tumors or unilateral renal artery stenosis, high renin levels stimulate angiotensin II and then aldosterone secretion, with a resulting increase in sodium reabsorption, hypertension, and hypokalemic metabolic alkalosis.

In some patients with overproduction of ACTH, as in ectopic production from lung and other malignant neoplasms (Chapter 179), cortisol may overwhelm the aldosterone receptor and result in hypertensive, hypokalemic alkalosis. The patient may not show signs of Cushing syndrome unless the syndrome is prolonged. In glucocorticoid-remediable aldosteronism, which is a familial disorder in which episodes of hypokalemia and hypertension develop, a chimeric gene duplication couples the ACTH-responsive 11β-hydroxylase promoter to the coding region of aldosterone synthase.

Glycyrrhizic acid, which is found in licorice and anisette, inhibits the renal enzyme 11β-hydroxysteroid dehydrogenase and thereby causes hypokalemia, metabolic alkalosis, and hypertension. A rare genetic disorder known as apparent mineralocorticoid excess syndrome produces the same effect due to deficiency of 11β-hydroxysteroid dehydrogenase. The syndrome produces a high ratio of cortisol to cortisone; as a result, renin and aldosterone are suppressed by the volume expansion. Hypokalemia may be precipitated by ACTH stimulation of cortisol.

Activating mutations of ENaC cause increased sodium reabsorption (Liddle syndrome; Chapter 128). The syndrome can be distinguished from primary or secondary hyperaldosteronism by a decrease in renin and aldosterone levels.

Hypokalemic Hypotensive Syndromes

In contrast to the hypokalemic hypertensive syndromes, in which an increased sodium reabsorption is a primary event, many hypokalemic alkaloses are

TABLE 117-2 CAUSES OF HYPOKALEMIA AND INCREASED POTASSIUM EXCRETION

CAUSES OF INCREASED K^+ EXCRETION AND HYPOKALEMIA	RENIN	ALDOSTERONE	EXTRACELLULAR VOLUME OR BLOOD PRESSURE	ACID-BASE STATUS
Increased ENaC: Liddle syndrome	Low	Low	High	Alkalosis
Decreased β-hydroxysteroid dehydrogenase: apparent mineralocorticoid excess, licorice	Low	Low	High	Alkalosis
Adrenal tumor or hyperplasia	Low	High	High	Alkalosis
Ectopic ACTH: Cushing syndrome	Low	Low	High	Alkalosis
Congenital adrenal hyperplasia	Low	High	High	Alkalosis
Unilateral renal artery stenosis	High	High	High	Alkalosis
Renin-secreting tumor	High	High	High	Alkalosis
Diuretics				
Thiazides	High	High	Low	Alkalosis
Furosemide	High	High	Low	Alkalosis
Acetazolamide	High	High	Variable	Acidosis
Bartter syndrome	High	High	Low	Alkalosis
Gitelman syndrome	High	High	Low	Alkalosis
Fanconi syndrome	High	High	Low	Acidosis
Distal RTA	High	High	Low	Acidosis

ACTH = adrenocorticotropic hormone; ENaC = epithelial Na^+ channel; RTA = renal tubular acidosis.

associated with extracellular volume depletion. With a physiologic response to volume depletion, the potassium losses may be a result of appropriate sodium reabsorption; signs of hypotension or extracellular volume depletion will be observed. Secretory diarrheas, whether associated with hypochloremic alkalosis or hyperchloremic acidosis, lead to extracellular volume depletion and secondary increases in renin and aldosterone; the result is both gastrointestinal and urinary potassium losses. However, because alkalosis increases potassium secretion, urinary potassium losses are most severe in gastric alkalosis or other chloride-wasting syndromes associated with alkalemia.

Diuretic use, Bartter syndrome (Chapter 128), and Gitelman syndrome (Chapter 128) are renal tubular causes of extracellular volume depletion, hypotension, and hypokalemic, hypochloremic alkalosis; sodium and chloride are lost in the urine, and secondary rises in renin and aldosterone occur. Increased urinary flow rates are important contributors to the increased potassium losses in each of these examples. Bartter syndrome affects the function of the thick ascending limb through mutations in NKCC or in potassium or chloride channels, whereas Gitelman syndrome is characterized by inactivating mutations or dysregulation of the Na^+-Cl^- cotransporter in the distal tubule. The hypokalemia in Gitelman syndrome may be caused by secondary hyperaldosteronism, bicarbonaturia, and hypomagnesemia.

Classic type 1 distal renal tubular acidosis (Chapters 118 and 128) is often associated with hypokalemia, which may improve with correction of the acidemia. In contrast, proximal renal tubular acidosis, when it is corrected with bicarbonate, often results in worsening of the hypokalemia because of greater bicarbonate wasting associated with increases in the filtered bicarbonate load.

Acetazolamide, when it is given to an alkalotic patient, is a particularly potent kaliuretic agent. Volume depletion and hyperaldosteronism contribute to the potassium losses, as does the bicarbonate wasting, which appears to have a direct effect on potassium secretion. Whenever the urine is alkaline, potassium will usually be present in significant amounts.

Tubular toxins that may be associated with severe potassium losses include aminoglycosides, cisplatinum, and ifosfamide. Amphotericin B results in significant potassium wasting accompanied by renal tubular acidosis. In many of these conditions, simultaneous use of amiloride may diminish potassium losses by as much as 50%.

Patients who present with unexplained hypokalemia and alkalosis with volume depletion should have urinary electrolytes measured to determine whether the urine chloride is low, as with vomiting or laxative abuse. If the urine contains chloride, a diuretic screen should be considered; Gitelman syndrome and Bartter syndrome are other possibilities.

Hyperkalemic Disorders

In clinical practice, acute hyperkalemia is seen most commonly with renal failure (Chapter 131), with acidosis (Chapter 118), and with acute muscle damage from rhabdomyolysis (Chapter 113) (Table 117-3). Chronic hyperkalemia is most commonly seen with medications that reduce potassium secretion and with renal tubular disorders.

Hyperkalemic Hypotensive or Normotensive Syndromes

Type 4 renal tubular acidoses (Chapter 118) that are associated with an inability to acidify the urine are caused by diseases that disrupt distal nephron function, including systemic lupus erythematosus (Chapter 266), urinary tract obstruction (Chapter 123), amyloidosis (Chapter 188), the nephropathy associated with kidney and bone marrow transplantation (Chapters 131 and 178), and sickle cell nephropathy (Chapter 125). Men who present with hyperkalemia and renal insufficiency of unknown cause should be evaluated for possible prostatic obstruction (Chapter 129). Each of these conditions may also be associated with failure to concentrate the urine (nephrogenic diabetes insipidus) or with a hyperchloremic metabolic acidosis caused by abnormalities in acid secretion.

Primary selective hypoaldosteronism or complete adrenal cortical deficiency (Chapter 227) is associated with elevated renin and low aldosterone. Secondary hypoaldosteronism may be seen in hyporenin states caused by β-blockers, renin antagonists, or nonsteroidal anti-inflammatory drugs (NSAIDs). Angiotensin-converting enzyme inhibitors and angiotensin receptor blockers increase renin and decrease aldosterone. Heparin, including low-molecular-weight and fractionated forms (Chapter 38), can lead to hyperkalemia even in small subcutaneous doses.

Disorders that affect the transepithelial potential difference and can result in hyperkalemia, acidosis, and extracellular volume depletion include inactivating mutations of ENaC (autosomal recessive pseudohypoaldosteronism type 1), which may be accompanied by high renin and aldosterone levels. The ENaC may also be inhibited by the potassium-sparing diuretics (i.e., amiloride and triamterene) and certain medications secreted by the proximal tubule organic cation transporters, such as trimethoprim and pentamidine, as well as by lithium. Inhibition of the aldosterone receptor may be the result of antagonists such as spironolactone and eplerenone.

Hyperkalemic Hypertensive Syndromes

NSAIDs can cause hyperkalemia, particularly in patients with renal disease, by decreasing sodium delivery, increasing water reabsorption, and decreasing renin and aldosterone. Hypertension with the NSAIDs is most likely caused by renal salt and water retention.

Cyclosporine or tacrolimus may produce a hyperkalemic acidosis. The mechanism may involve inhibition of cyclooxygenase 2 and therefore hyporenin-hypoaldosteronism. There may also be a decrease in apical potassium secretion in the collecting duct. Gordon syndrome is a genetic disorder (pseudohypoaldosteronism type 2) associated with hyperkalemia, volume expansion, and metabolic acidosis. It is caused by activation of the thiazide-sensitive distal convoluted tubule NaCl cotransporter related to a dysfunction of the WNK kinase regulatory role.

TABLE 117-3 CAUSES OF HYPERKALEMIA AND DECREASED POTASSIUM EXCRETION

CAUSE OF DECREASED K^+ EXCRETION AND HYPERKALEMIA	RENIN	ALDOSTERONE	EXTRACELLULAR VOLUME OR BLOOD PRESSURE	ACID-BASE STATUS
Decreased ENaC Drugs: amiloride, triamterene, trimethoprim, lithium Pseudohypoaldosteronism type 1 autosomal recessive ENaC mutation	High	High	Low or normal	Normal or acidosis
Pseudohypoaldosteronism type 1 autosomal dominant MR mutation	High	High	Low	Acidosis
MR blockade: spironolactone, eplerenone, progesterone	High	High	Low	Acidosis
Hypoaldosteronism: adrenal insufficiency	High	Low	Low	Acidosis
Hyporenin-hypoaldosteronism: NSAIDs, β-blockers, autonomic neuropathy	Low	Low	Low	Acidosis
Pseudohypoaldosteronism type 2	Low	Low	High	Acidosis

ENaC = epithelial Na^+ channel; MR = mineralocorticoid receptor; NSAIDs = nonsteroidal anti-inflammatory drugs.

TREATMENT

Rx

It may be difficult to determine the exact state of total body potassium stores from the serum potassium level because as much as 100 to 300 mmol of potassium may be lost from the body with a fall in serum potassium of only 1 mmol/L.

Hypokalemia

The goal of acute therapy for hypokalemia is to prevent or to manage potentially life-threatening arrhythmias or paralysis. Patients at greatest risk are elderly patients, patients with known liver disease or cardiac disturbances, and patients who have had an abrupt fall in serum potassium concentration to less than 2.5 mEq/L. Potassium must traverse the extracellular space before repleting intracellular stores, so it is dangerously easy to replete potassium too quickly. Oral potassium should be given, if possible. If the potassium level is greater than 3 mEq/L, an increase in dietary potassium can be considered, along with removal of the underlying cause of hypokalemia. Usual oral replacement is with potassium chloride at a dose of 40 to 100 mmol/day. The chloride salt has the advantage of treating concomitant metabolic alkalosis, but other available forms include potassium citrate (in the acidotic patient) and potassium phosphate (in patients with a phosphate deficit). Giving potassium with a non-reabsorbable anion, such as gluconate, may not replace the potassium deficit adequately. Intravenous potassium is reserved for patients who are unable to take enteral potassium and patients with symptomatic hypokalemia, paralysis, or cardiac arrhythmias. It is usually given as a solution of 20 to 40 mmol of potassium in 1 L of solution at a rate that does not exceed 10 to 20 mmol/hour. In some cases of severe hypokalemia (<2.5 mEq/L) and in symptomatic patients, higher concentrations (up to 40 mmol in 100 mL) have been used. If the potassium level is less than 3 mmol/L or if more than 10 mmol/hour is to be delivered, it may be best to treat the patient in a monitored setting to observe for cardiac complications. A central venous catheter may be required for these higher concentrations. In these unusual circumstances, it is best to consult with a renal specialist and the pharmacy.

In patients who have prerenal azotemia associated with hyperglycemia or severe metabolic alkalosis, volume expansion with sodium chloride solutions alone can result in life-threatening potassium losses, despite improvement in the extracellular volume. Potassium must be given in anticipation of such events.

In a hypokalemic patient, care must be exercised when glucose-containing solutions are given because the resulting increase in insulin may further decrease the blood potassium level. Attention to the urine output and ongoing losses is crucial. If ongoing losses of potassium are severe, it may be necessary to provide a potassium-sparing diuretic (e.g., amiloride, 5 to 10 mg orally) and to treat the cause of the ongoing losses (e.g., diarrhea). Magnesium should be measured and replaced, if necessary, in any hypokalemic patient.

In the inpatient setting, hypokalemia is a common complication of intravenous fluid administration. In patients with normal renal function, a maintenance dose of intravenous potassium can avoid hypokalemia.[10] In the outpatient setting, hypokalemia is a common side effect of diuretic therapy. The serum potassium level should be maintained within the normal range (>3.5 mEq/L), especially in high-risk patients. Addition of 40 to 100 mmol of potassium per day as the chloride salt is the usual treatment, depending on the patient's response.

Hyperkalemia

If hyperkalemia is severe, the goal is to achieve a rapid reduction in potassium concentration.[11] If hyperkalemia is associated with cardiac arrhythmias, however, the cardiac effects of the hyperkalemia require interim treatment in a monitored setting, before the serum potassium level can be expected to decline, even with aggressive therapy. Calcium gluconate, 10 mL of a 10% solution (8.9 mg calcium) during 10 to 20 minutes, is often indicated to stabilize electrical effects on cardiac excitation. Calcium chloride (3 to 4 mL of a 10% solution) is used as another alternative, but it should be administered through a central access line because extravasation of the chloride salt may result in tissue necrosis.

However, calcium does not lower the potassium concentration. Nebulized or inhaled β-agonists and intravenous insulin and glucose, either alone or the two in combination, are the best treatments. To lower the potassium level acutely, alternatives include 100 mL of 50% glucose alone or in combination with 10 units of regular insulin; the combination will, on average, provide a significantly greater reduction in the serum potassium level (0.8 mmol/L vs. 0.5 mmol/L) at 60 minutes, but about 20% of insulin-treated patients will develop hypoglycemia. Alternatively, 10 mg of regular insulin can be given with 10% glucose solution during 1 hour for a total of 30 to 50 g in the normoglycemic patient. The blood glucose level should be monitored because an abrupt increase in plasma osmolality with glucose may worsen hyperkalemia if it causes potassium to leak from cells with osmotic water flow. Albuterol by nebulizer (10 to 20 mg in 4 mL of saline during 10 minutes) can redistribute potassium acutely but should not be the sole treatment because some patients are not responsive. In some cases, intravenous β-adrenergic agonists (e.g., albuterol, 0.5 mg in 100 mL of 5% dextrose during 15 minutes) have been used to lower the serum potassium level by about 1 mmol/L within minutes to hours. Sodium bicarbonate, as an isotonic mixture calculated to correct acid-base status, should be reserved for acidemic patients who otherwise require alkalinization, while being careful to avoid hypocalcemia; complications of sodium bicarbonate infusions include hypernatremia, volume expansion, and decreased ionized calcium, potentially resulting in tetany. Diabetic patients may be potassium depleted even though they present with hyperkalemia; as they are volume repleted, they typically require potassium replacement (Chapter 229). The hyperkalemia of Gordon syndrome is highly responsive to thiazide diuretics.

During the longer term, potassium loss may be sustained by use of cation exchange resins such as sodium polystyrene sulfonate, given orally or as an enema (Chapter 131).[12] A dose of 30 to 50 g can reduce potassium levels during several hours. This resin will also provide a sodium load and bind calcium, thereby resulting in volume expansion and hypocalcemia. These resins may interfere with the absorption of lithium and thyroxine. A serious complication of sodium polystyrene sulfonate resins when they are used in combination with sorbitol is colonic ulceration and necrosis. These resins also should not be given in combination with aluminum-based antacids because the resulting concretions can obstruct the gastrointestinal tract. If the patient is volume expanded, furosemide (40 to 100 mg), chlorothiazide (500 mg), or, if the patient is also alkalotic, acetazolamide (250 to 500 mg) may enhance renal potassium clearance. If the patient is volume depleted, isotonic saline expansion may improve urine output and, with it, potassium excretion.

New medications that reduce potassium levels in high-risk patients include patiromer (a non-absorbed calcium-potassium exchange resin that works in the colon at 4.2 to 8.4 g twice daily) in patients with chronic kidney disease who are taking renin-angiotensin-aldosterone system inhibitors, or zirconium cyclosilicate (an oral crystalline agent with high binding affinity for potassium within the gastrointestinal tract at 1.25 to 10 g three times daily). However, the long-term utility and safety of these agents remains to be determined, and they are not currently FDA-approved.

Specific Clinical Syndromes

Hypokalemia

Patients with pernicious anemia who receive vitamin B_{12} to stimulate erythropoiesis may deplete extracellular potassium and suffer from hypokalemia as a cost of producing new red blood cells. Leukemias with rapid growth rates (Chapter 183) also may cause a drop in the serum potassium level, and some forms of myelogenous leukemia are associated with a high level of lysozyme, which leads to urinary potassium loss as well (Chapter 184).

Familial hypokalemic periodic paralysis is an autosomal dominant disorder usually caused by mutations in certain voltage-gated skeletal muscle sodium channels or L-type calcium channels.[13] Characteristically, periodic attacks of severe hypokalemia are precipitated by stimuli, such as the insulin response to carbohydrate ingestion or rest after exercise, that usually induce mild hypokalemia by distributing potassium into cells. In individuals with these channel mutations, the same stimuli cause progressive and severe hypokalemia, enough to result in muscle paralysis due to hyperpolarization of the sarcolemma. The cause of the syndrome is an imbalance between outwardly directed K^+ current and inwardly directed cation leaks of Na^+ and Ca^{2+}. In the familial mutations, the cation depolarizing influx, normally very small under hyperpolarizing conditions, is instead increased and, when balanced with the outward K^+ current, results in a paradoxical depolarization. In turn, the depolarization inactivates the sodium channel needed for the rapid action potential, so the muscle cell is inexcitable. In nonfamilial hypokalemic periodic paralysis, a loss in function of outwardly directed potassium channels (Kir) accounts for paradoxical depolarization, inactivation of sodium channels, and decreased excitability.

The clinical presentation of hypokalemic periodic paralysis is usually in the teenage years or early adulthood. In some familial cases, a progressive proximal myopathy (Chapter 421) may develop. Asian patients, usually males with hyperthyroidism (Chapter 226), have a decrease-in-function mutation of an outwardly directed potassium channel that causes them to develop episodic paralysis, which is precipitated by high-carbohydrate meals (insulin secretion) or by rest after exercise (when plasma potassium falls because of reuptake of potassium by the ATPase pumps). In Anderson syndrome, which is another form of hypokalemic periodic paralysis, cardiac potassium channels are also affected, and serious cardiac arrhythmias may result.

The condition is treated by a high-potassium diet as well as with β_2-blockers (e.g., propranolol, 20 to 40 mg twice daily) and carbonic anhydrase inhibitors (e.g., acetazolamide, 125 to 500 mg), which in part work by creating a hyperchloremic acidosis that offsets the urinary potassium wasting they cause.

Hyperkalemia

The abnormal distribution of potassium between cells and the extracellular space results in the hyperkalemia that is associated with acidosis, insulin-deficient states, and β_2-adrenergic blockade. Although it is not always possible, it is best to know the levels of serum glucose and potassium in an unconscious diabetic patient before infusing concentrated glucose solutions because of the risk for aggravating an already elevated potassium concentration.

Familial hyperkalemic periodic paralysis is a myopathy caused by a genetic defect in voltage-gated sodium channels in skeletal muscle. Exercise or dietary increases in plasma potassium result in mild depolarization of skeletal muscle that then unmasks the sodium channel defect, rendering the cells unexcitable. Treatment is frequent meals and acetazolamide (125 to 500 mg).

Potassium competes with digoxin-binding sites on the Na^+, K^+-ATPase, so that if hypokalemia coexists, digoxin will have an intensified effect and may lead to drug toxicity. In extreme cases of digitalis overdose, severe hyperkalemia develops as a result of generalized blockade of Na^+, K^+-ATPase.

PROGNOSIS

The prognosis of patients with hypokalemia and hyperkalemia depends on the severity and underlying illness. Most hypokalemic cases are mild (potassium concentration, 3 to 3.5 mEq/L). However, mortality of hospitalized patients with hypokalemia is increased 10-fold. Hyperkalemia is reported in 1 to 10% of hospitalized patients, of whom 10% have severe hyperkalemia (potassium concentration, >6.0 mEq/L). Hyperkalemia is associated with increased mortality (14 to 41%), and it accounts for 2 to 5% of deaths in patients with end-stage renal disease.

Grade A References

A1. Mahoney BA, Smith WA, Lo DS, et al. Emergency interventions for hyperkalaemia. *Cochrane Database Syst Rev*. 2005;2:CD003235.

A2. Chothia MY, Halperin ML, Rensburg MA, et al. Bolus administration of intravenous glucose in the treatment of hyperkalemia: a randomized controlled trial. *Nephron Physiol*. 2014;126:1-8.

A3. Weir MR, Bakris GL, Bushinsky DA, et al. Patiromer in patients with kidney disease and hyperkalemia receiving RAAS inhibitors. *N Engl J Med*. 2015;372:211-221.

A4. Packham DK, Rasmussen HS, Lavin PT, et al. Sodium zirconium cyclosilicate in hyperkalemia. *N Engl J Med*. 2015;372:222-231.

A5. Kosiborod M, Rasmussen HS, Lavin P, et al. Effect of sodium zirconium cyclosilicate on potassium lowering for 28 days among outpatients with hyperkalemia: the HARMONIZE randomized clinical trial. *JAMA*. 2014;312:2223-2233.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Yang Q, Liu T, Kuklina EV, et al. Sodium and potassium intake and mortality among US adults: prospective data from the Third National Health and Nutrition Examination Survey. *Arch Intern Med*. 2011;171:1183-1191.
2. Korgaonkar S, Tilea A, Gillespie BW, et al. Serum potassium and outcomes in CKD: insights from the RRI-CKD cohort study. *Clin J Am Soc Nephrol*. 2010;5:762-769.
3. Gumz ML, Rabinowitz L. Role of circadian rhythms in potassium homeostasis. *Semin Nephrol*. 2013;33:229-236.
4. Youn JH. Gut sensing of potassium intake and its role in potassium homeostasis. *Semin Nephrol*. 2013;33:248-256.
5. Welling PA. Regulation of renal potassium secretion: molecular mechanisms. *Semin Nephrol*. 2013; 33:215-228.
6. Medford-Davis L, Rafique Z. Derangements of potassium. *Emerg Med Clin North Am*. 2014;32: 329-347.
7. Lippi G, Favaloro EJ, Montagnana M, et al. Prevalence of hypokalaemia: the experience of a large academic hospital. *Intern Med J*. 2010;40:315-316.
8. Jain G, Ong S, Warnock DG. Genetic disorders of potassium homeostasis. *Semin Nephrol*. 2013; 33:300-309.
9. Oram RA, McDonald TJ, Vaidya B. Investigating hypokalaemia. *BMJ*. 2013;347:f5137.
10. Scotto CJ, Fridline M, Menhart CJ, et al. Preventing hypokalemia in critically ill patients. *Am J Crit Care*. 2014;23:145-149.
11. Fordjour KN, Walton T, Doran JJ. Management of hyperkalemia in hospitalized patients. *Am J Med Sci*. 2014;347:93-100.
12. Sterns RH, Rojas M, Bernstein P, et al. Ion-exchange resins for the treatment of hyperkalemia: are they safe and effective? *J Am Soc Nephrol*. 2010;21:733-735.
13. Cheng CJ, Kuo E, Huang CL. Extracellular potassium homeostasis: insights from hypokalemic periodic paralysis. *Semin Nephrol*. 2013;33:237-247.

REVIEW QUESTIONS

1. A patient with vomiting develops hypokalemia from

A. Gastrointestinal potassium losses
B. Renal excretion of filtered bicarbonate in the steady state
C. Renal potassium wasting from early bicarbonaturia and hyperaldosteronism
D. Increased proximal tubule $NaHCO_3$ reabsorption
E. Distribution of potassium to the intracellular space

Answer: C Renal potassium wasting is the cause of severe potassium depletion in vomiting and plays a more important role than gastric losses because gastric fluid K^+ is low. Increased proximal tubule Na^+/H^+ exchanger increases $NaHCO_3$ reabsorption and maintains alkalosis but does not worsen K^+ losses. The alkaline urine and hyperaldosteronism favor the loss of K^+ with HCO_3.

2. Hypokalemic periodic paralysis can be associated with

A. Severe hypokalemia due to dietary ingestion of carbohydrates and subsequent insulin release
B. Hyperthyroidism
C. A mutation of a specific calcium channel responsible for cation current into muscle cells
D. Severe cardiac arrhythmias
E. All of the above

Answer: E Mutations of sodium or calcium channels that contribute to an inward leak of cations are the cause of a paradoxical depolarization of the sarcolemma membrane precipitated by stimuli that induce mild decreases in serum potassium. Symptoms may be precipitated by hypokalemic stimuli, such as carbohydrate loading. In some patients, especially Asians, it is associated with thyrotoxicosis. In occasional cases, severe hypokalemia can cause cardiac arrhythmias.

118

ACID-BASE DISORDERS

JULIAN L. SEIFTER

DEFINITION

If arterial pH is below 7.35, acidemia is said to exist. If pH is above 7.45, alkalemia exists. However, several processes may simultaneously drive the pH upward or downward; these individual processes are known as *acidoses* or *alkaloses*. Because multiple processes may coexist, an abnormal pH is not always noted in acid-base disturbances. Because pH is related to the ratio of HCO_3^- to P_{CO_2}, the finding of an abnormal bicarbonate level alone cannot define acidosis or alkalosis.

EPIDEMIOLOGY

An acid-base disturbance should alert the clinician to the possible presence of an important underlying condition. Anion gap acidoses represent serious underlying metabolic disorders, ranging from sepsis (Chapter 108) to uremia (Chapter 130) to diabetic ketoacidosis (Chapter 229) to serious poisonings (Chapter 110). Specific renal abnormalities as well as diarrhea (Chapter 140) can cause hyperchloremic acidosis (Table 118-1). Metabolic alkaloses are commonly caused by diuretics or renal tubular abnormalities or the loss of acid from the stomach due to vomiting or nasogastric suction (Table 118-2). Respiratory acidosis and alkalosis are related to ventilation, which is increased by conditions such as sepsis (Chapter 108) and anxiety and decreased in many pulmonary conditions (Chapters 86 and 104).

PATHOBIOLOGY

One of the major requirements for cell survival, along with maintaining electrical gradients and cell volume, is the regulation of the H^+ ion concentration, or pH (defined as the negative logarithm of the hydrogen ion concentration).

TABLE 118-1 CAUSES OF HYPERCHLOREMIC ACIDOSIS

TYPE	CAUSE
Renal with hypokalemia	Proximal RTA, type 2 Distal RTA, type 1 Some anion gap acidoses with high anion clearance
Renal with hyperkalemia	Type 4 RTA; hyporenin-hypoaldosteronism
Nonrenal with hypokalemia	Diarrhea Urinary diversions: ureteroileostomy, ureterosigmoidostomy
Nonrenal with hyperkalemia	NaCl, KCl, NH_4Cl, $CaCl_2$, Arg-HCl, Lys-HCl

RTA = renal tubular acidosis.

TABLE 118-2 CAUSES OF METABOLIC ALKALOSIS

TYPE	CAUSES
Renal, hypochloremic alkalosis: chloride responsive with urine chloride concentration >20 mEq/L	Loop and distal tubule diuretics Bartter syndrome Gitelman syndrome Post-hypercapnic status
Nonrenal, hypochloremic alkalosis: chloride responsive with urine chloride concentration <20 mEq/L	Vomiting, nasogastric suction Chloridorrhea Villous adenoma
Renal, alkalosis with extracellular expansion: chloride unresponsive with urine chloride concentration >20 mEq/L	Hyperaldosteronism, primary and secondary to unilateral renal artery stenosis Liddle syndrome
Nonrenal alkalosis, chloride unresponsive	$NaHCO_3$, acetate, citrate, lactate
Other causes of metabolic alkalosis	Excessive non-reabsorbable anion excretion Hypoproteinemia

Growth, cell division, fertilization, and protein and glucose metabolism are examples of pH-sensitive processes. Acids are generated within cells during metabolism, and each cell must maintain a pH appropriate for its function. For example, cardiac contractility (Chapter 53) is reduced when cardiac myocytes are too acid. Bone and muscle develop and grow poorly in an acidic environment. Intracellular pH may be lower than extracellular pH because cells are electronegative with respect to extracellular fluid, but they are not as acidic as they would be if H^+ reached electrochemical equilibrium with the extracellular fluid, which means that all cells require energy to lose acid actively. Cells are capable of buffering an acid load, and intracellular vacuoles may use hydrogen adenosine triphosphatases (H^+-ATPases) to sequester excess acid before transport from the cell. The transport processes located on the plasma membranes can protect cells from both acid and alkaline loads. The specific mechanisms may differ from one cell type to another, but they are similar to those used by the excretory organs that finally eliminate the net acid produced by the body into the external world.[1]

At a normal arterial blood pH of 7.36 to 7.45, the hydrogen ion concentration is in the range of 40 nanoequivalents (nEq) per liter, a very small concentration in comparison to a normal plasma sodium concentration of 140 mEq/L. In severe disease states, arterial pH may fall as low as 6.8 and rise as high as 7.7. Strenuous exercise with the metabolic production of lactate may transiently but severely lower pH, even in normal healthy individuals.

The hydrogen ion concentration of body fluids is in equilibrium with each of multiple weak acids or buffers, such as proteins and phosphate, but acid-base equilibria in the body are often described and analyzed by use of the CO_2/HCO_3^- system and the relationship of the proton concentration (thus pH) to the ratio of HCO_3^- to CO_2.[2] The Henderson-Hasselbalch equation is a logarithmic expression of the relationship.

$$CO_2 + H_2O \leftrightarrow H_2CO_3 \leftrightarrow H^+ + HCO_3^-$$

$$pH = pK + \log [HCO_3]/0.03\,(P_{CO_2})$$

In this equation, p*K*, or the log of the equilibrium constant for the reaction, is 6.1; 0.03 (mM/mm Hg) is the solubility factor for CO_2 in solution. The product of $0.03 \times P_{CO_2}$ represents dissolved CO_2; the "total CO_2" in plasma is the sum of HCO_3^-, normally about 25 mM, and $0.03 \times P_{CO_2}$, normally about 1.2 mM. It is important to note that pH is a function of the *ratio* of HCO_3^- to P_{CO_2}. The HCO_3^- concentration in the numerator is regulated by the kidney, and P_{CO_2} is regulated by the lung.

Production of Carbonic Acid and the Elimination of Carbon Dioxide by the Lung

Volatile acid is the term used for the approximately 20,000 mmol/day of CO_2 produced, with an equimolar amount of water, by tissue respiration. This CO_2 is carried from tissues to the lung, where it is eliminated by alveolar ventilation. Steady-state arterial P_{CO_2} is normally 38 to 42 mm Hg.

Oxidation of carbohydrates, fat, and the carbon skeleton of amino acids results in the production of water and CO_2. To maintain a steady state, any acid (or base) produced per day must be equivalent to what is eliminated. If tissue CO_2 production exceeds CO_2 elimination by the lungs, respiratory acidosis characterized by a high P_{CO_2} will develop. If the rate of CO_2 eliminated exceeds production, respiratory alkalosis develops. The inverse relationship between alveolar ventilation (the clearance of CO_2) and P_{CO_2} is demonstrated as

$$\text{alveolar ventilation} \sim CO_2 \text{ elimination} \div P_{CO_2}$$

The circulation plays a critical role in transporting tissue CO_2 to the lungs. The process depends not only on cellular respiration but also on tissue capillary flow as well as diffusion of CO_2 into the blood and across red blood cell membranes, where it may react with hemoglobin and proteins to form carbamino compounds or combine with water for conversion to H^+ and HCO_3^- in a reaction catalyzed by carbonic anhydrase. The intracellular H^+ can combine with hemoglobin (the Bohr effect) and the HCO_3^- exchanged with plasma Cl^- through red cell anion exchangers. Most tissue CO_2 is brought to the lung as venous plasma HCO_3^-. Compared with arterial blood, venous blood has the characteristics of a respiratory acidosis. Venous pH is normally approximately 0.05 pH unit more acid than arterial pH, its P_{CO_2} is 5 to 6 mm Hg higher than that of arterial blood, and its bicarbonate concentrations are normally greater than arterial concentrations.

In disease, changes in P_{CO_2} are most often caused by changes in alveolar ventilation rather than by production of CO_2. Thus, respiratory acidosis is a consequence of decreased pulmonary ventilation because of lung, skeletal muscle, or central nervous system (CNS) disease. However, if alveolar ventilation is compromised, increased production of CO_2 will worsen CO_2 retention. Similarly, respiratory alkalosis develops because of hyperventilation rather than decreased CO_2 production. In either case, when the elimination rate of CO_2 (alveolar ventilation $\times P_{CO_2}$) again equals CO_2 production, a new steady-state P_{CO_2} will prevail, with no net retention or loss of carbonic acid.

Production of Acids and Excretion by the Kidney

Nonvolatile or fixed acid describes non-carbonic acids that are formed primarily from protein metabolism. The usual rate of formation is approximately 1 to 2 mEq of H^+ per kilogram of body weight per day. Most diets that contain animal protein have a net positive quantity of nonvolatile acids, primarily sulfates from the sulfur-containing amino acids cysteine and methionine. Other acids are produced in the form of phosphates (from phosphoproteins, phospholipids, and phosphonucleotides) and nonmetabolizable organic acids (e.g., uric acid) and chloride from salts of lysine, arginine, and histidine.

From the Henderson-Hasselbalch equation, consider the addition of metabolic acid HA (where the anion A^- could be Cl^-, lactate$^-$, HSO_4^-, or $H_2PO_4^-$) to blood that contains Na^+ and HCO_3^-.

$$HA + NaHCO_3 \rightarrow Na^+ + A^- + H^+ + HCO_3^-$$

and

$$H^+ + HCO_3^- \leftrightarrow H_2CO_3 \leftrightarrow CO_2 + H_2O$$

The CO_2 produced by this process will not raise the blood P_{CO_2} if the system is well ventilated because the contribution of metabolic acid is a small part of the daily production of CO_2. Note that the addition of protons to body fluids by these acid end products consumes bicarbonate ("lost bicarbonate"), which then must be replenished by the kidney as it eliminates the proton and A^- in the urine. The process of excreting net H^+ and A^- is equivalent to producing a "new" HCO_3^- to restore the HCO_3^- that is lost by the addition of HA, the metabolic acid. The kidney must excrete any nonvolatile acid (or alkali) load to maintain a steady-state serum HCO_3^- concentration in the 22- to 28-mEq/L range.

When the diet requires the excretion of acids, the urine pH will fall to a value as low as 5.0, and the urine will become nominally free of bicarbonate. With an alkaline load, by comparison, the kidney will reject the excess filtered HCO_3^-, and the urine pH may approach a maximal value of 8.0 to 8.5. In most humans, particularly those who eat animal protein or an "acid-ash" diet, the requirement for net acid *excretion* predominates. However, some vegetarians may have an overall "alkaline-ash" diet, for which net alkali must be excreted to match intake.

Bicarbonate and the Kidney in Acid-Base Balance

The first role of the kidney in achieving acid excretion is to reabsorb all filtered HCO_3^- (E-Fig. 118-1). At a normal glomerular filtration rate (e.g., ~180 L/day in an adult) and plasma HCO_3^- concentration of 25 mEq/L, about 4500 mEq of HCO_3^- is filtered in 1 day. Loss of even a small fraction of that amount would result in metabolic acidosis if not replaced by HCO_3^- intake. Just as reabsorption of all filtered glucose does not add glucose to the body, reabsorption of all filtered HCO_3^- will maintain the status quo but will not fulfill the need to generate new HCO_3^-. Therefore, the kidney must perform two functions: reabsorb all filtered HCO_3^- and eliminate enough additional H^+ (producing HCO_3^- in the process) to maintain balance. Without urinary buffers, the urinary pH cannot be lowered enough to excrete the amount of acid needed for this purpose.

The Proximal Tubule

About 80 to 90% of HCO_3^- reabsorption is accomplished in the proximal tubule by a proton secretory process that renders the proximal tubular fluid more acid (pH ~6.5). The brush border membranes facing the lumen of the proximal tubule cell contain transporters known as Na/H exchangers (NHE3), which carry out the greatest proportion of acidification, and vacuolar H^+-ATPases, which provide a smaller contribution. Through the normal function of basolateral membrane Na^+, K^+-ATPase, cell Na^+ is kept at low concentration so that filtered Na^+ in the lumen will be favored to enter the cell in exchange for H^+ secreted into the lumen. This H^+ rapidly combines with filtered HCO_3^- to form H_2CO_3, which then dehydrates in the lumen to form CO_2 and H_2O. This last process is greatly facilitated by the large surface

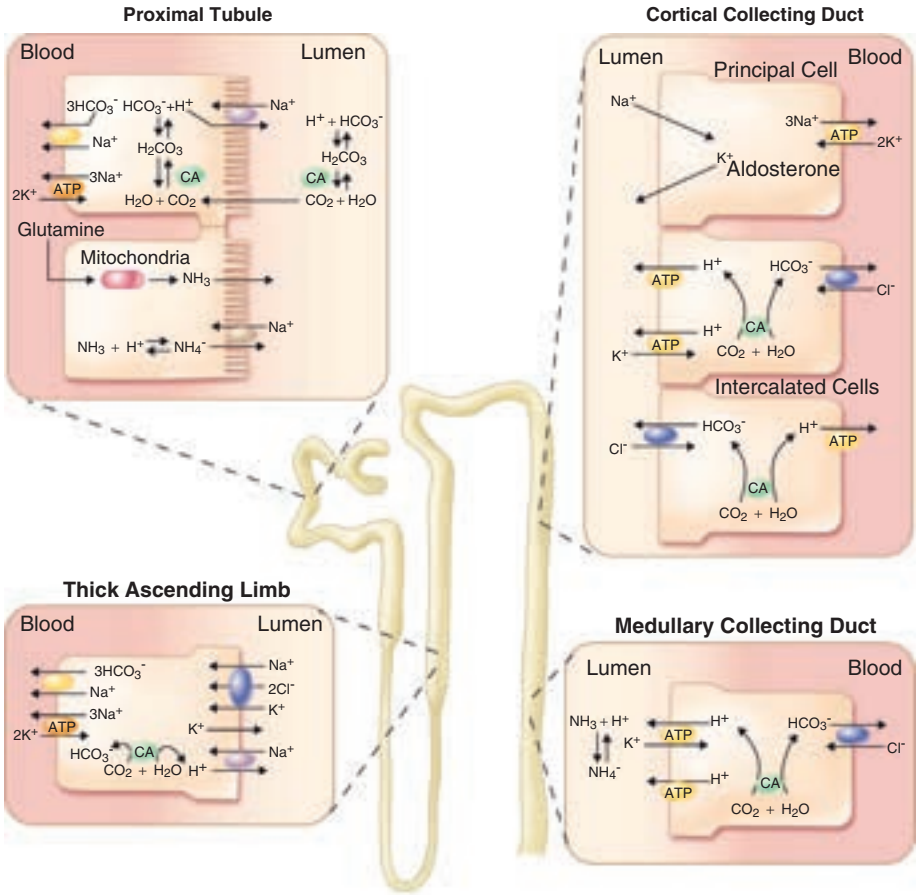

E-FIGURE 118-1. Renal acidification mechanisms. ATP = adenosine triphosphate; CA = carbonic anhydrase.

area of microvillous membranes and by luminal carbonic anhydrase (CA_{IV}). The CO_2 enters the proximal cell by diffusion through apical water channels (aquaporin 1). CO_2 within the cell re-forms HCO_3^-, a reaction catalyzed by intracellular carbonic anhydrase (CA_{II}). The cell, more alkaline by apical H^+ secretion, favors the reaction $OH^- + CO_2$ to form HCO_3^- catalyzed by carbonic anhydrase at the higher cell pH. The HCO_3^- is then transported back toward the blood by a sodium bicarbonate cotransporter (NBC), which couples 1Na and $3HCO_3^-$, thereby completing net Na^+ and HCO_3^- reabsorption. The 1 : 3 stoichiometry of NBC is necessary to provide sufficient energy to couple Na^+ and HCO_3^- in this electrogenic process, which protects the cell from an alkaline load, accomplishes HCO_3^- reabsorption, and drives uphill Na^+ reabsorption without further direct requirement for adenosine triphosphate (ATP). The entire process requires a mitochondrial source of ATP for the Na^+/K^+ pump, intact NHE3 and NBC, and two isoforms of carbonic anhydrase. In addition, there must be favorable ion gradients for luminal Na^+ entry, H^+ secretion, and basolateral HCO_3^- transport. A disturbance in any of these factors may disrupt proximal HCO_3^- reabsorption enough to cause loss of HCO_3^- in urine. It is also clear that the acidifying process in the proximal tubule provides a significant mechanism for Na^+ reabsorption, consistent with the finding that HCO_3^- reabsorption is increased by angiotensin II and the sympathetic nervous system in the defense of extracellular volume. Another 10 to 15% of filtered HCO_3^- is reabsorbed in the thick ascending limb of Henle through a mechanism similar to the mechanism in the proximal tubule, so that only small amounts of the filtered HCO_3^- are normally delivered to more distal nephron segments.

The Cortical Collecting Duct

The cortical connecting tubules and collecting ducts reabsorb less than 10% of the filtered HCO_3^-. In principal cells, Na^+ is reabsorbed from lumen to cell by the epithelial sodium channel (ENaC), driven by the inwardly directed Na^+ gradient and favorable electrical potential. With the reabsorption of Na^+, the lumen becomes electronegative, thus favoring the secretion of both K^+, through potassium channels, and H^+, through vacuolar H^+-ATPases on the luminal surface of neighboring α-intercalated cells, which are acid-secreting cells. The secreted H^+ will combine with the remaining HCO_3^- in the lumen to generate CO_2, with subsequent reabsorption of CO_2, re-formation of cellular HCO_3^- with the help of cellular carbonic anhydrase (CA_{II}), and then exchange of HCO_3^- from cell to blood for entry of Cl^- through Cl^-/HCO_3^- exchangers, similar to the erythrocytic anion exchangers (AE1) that participate in blood CO_2 transport. The distal nephron, which has a smaller requirement for bicarbonate reabsorption than the proximal tubule, lacks both brush border and luminal carbonic anhydrase. It is at this distal site that tubular fluid pH starts to fall to levels below pH 6.0.

Some collecting duct cells have reverse polarity and secrete HCO_3^- into the lumen in exchange for Cl^- entry into the cell. In these cells, the H^+-ATPase faces the blood side of the cell (β-intercalated cells). An elevated extracellular HCO_3^- concentration, as seen with an alkaline-ash diet or metabolic alkalosis, will increase HCO_3^- secretion by these cells through pendrin, which is the apical Cl^-/HCO_3^- exchanger.

The Medullary Collecting Duct

The medullary collecting duct continues to secrete protons into the luminal fluid, where the pH reaches its lowest values of close to 5.0. The mechanism is based on continued function of H^+-ATPases with an additional role of an ATP-dependent K^+/H^+ exchanger, a member of the family of K^+, H^+-ATPases found in the stomach and colon.

Once the filtered HCO_3^- is fully reabsorbed, the kidney is still required to eliminate an additional net amount of acid equivalent to that produced in metabolism. Most of this net acid excretion is in the form of ammonium (NH_4^+), which is derived from the renal synthesis of ammonia from glutamine in the proximal tubule and the titration of filtered phosphate to acid phosphate (titratable acidity).

$$NH_3 + H^+ \leftrightarrow NH_4^+ \; pK\ 9.1$$

and

$$HPO_4^{2-} + H^+ \leftrightarrow H_2PO_4^- \; pK\ 6.8$$

Urinary Buffers

As the collecting duct cells continue to secrete H^+ into urine with a diminishing luminal HCO_3^- concentration and decreasing pH, H^+ is captured by the urinary buffers. The resulting alkalinization of the cells after H^+ leaves results in the formation of cellular HCO_3^- ready for transport into blood. This process generates new HCO_3^- that is not a result of the reabsorption of filtered HCO_3^-. The amount of new HCO_3^- matches the amount of net acid eliminated and is also equal to each of the following: the amount of acid produced; the amount of body buffer consumed by that acid; and the amount of fixed acid anions, such as sulfate, phosphate, and Cl^-, that accompanied the H^+. The result is maintenance of normal acid-base equilibrium.

The ability of the kidney to lower urinary pH to values as low as 5.0 enables the buffers to capture a proton. Net acid excretion in urine is accomplished not simply by a decrease in urine pH but mostly by titration of these important urinary buffers. For example, a typical daily urine volume of 1 L at pH 5.0 contains only 10^{-5} molar hydrogen ion, or 0.01 mmol, a trivial amount compared with the amount of produced acid (~1 mmol/kg/day). In chronic kidney disease, it is the failure to produce enough ammonium that leads to a poorly buffered, although acid, urine and an inability to excrete enough net acid to stay in normal balance.

Regulation of Urinary Acid Secretion

Renal mechanisms of urinary acidification are adaptable. Transport processes such as H^+-ATPases, Na^+/H^+ exchange, and Cl^-/HCO_3^- exchange can increase or decrease their capacity to handle acid-base equivalents, depending on the challenge presented. Renal ammoniagenic mechanisms are also critically regulated to serve the acid-base needs of the individual. Metabolic acidosis and respiratory acidosis increase the capacity to reabsorb HCO_3^-, including increased expression of the transporters involved in acidifying the urine. At the same time, increased glutamine uptake into proximal cells and ammonia production enable increased acid excretion and the generation of new HCO_3^- in the distal nephron. Metabolic alkalosis and respiratory alkalosis have the opposite effects.

Ammoniagenesis, which is a key element in urinary acid excretion, provides the major acceptor for protons. Ammonia is produced predominantly in the proximal tubule cell by mitochondrial glutaminase enzymes. Production is increased by increasing the metabolic acid load in the body, respiratory acidosis, and hypokalemia. NH_4^+ can preserve potassium in the hypokalemic state by serving as a cation needed for anion excretion. Similarly, in response to metabolic acidosis, the kidney would ideally excrete chloride with ammonium and preserve the ability to retain Na^+ and K^+.

Ammonia can be secreted by nonionic diffusion into the proximal fluid, where it will pick up a proton and form ammonium (NH_4^+), or it could form ammonium within the proximal tubule cell and be secreted by Na^+/NH_4^+ exchange, a mode of operation of the Na^+/H^+ exchanger. Ammonium may be reabsorbed by the thick ascending limb of Henle on the Na-K-2Cl transporter, where it can substitute for K^+. By countercurrent multiplication, NH_3/NH_4^+ concentrates in the medullary interstitial fluid rather than remaining in the ascending limb fluid as it reaches the highly perfused renal cortex, where it would otherwise dissipate into renal venous blood. The countercurrent mechanism also allows ammonia to diffuse into the lumen of the medullary collecting duct, where it will be trapped as ammonium in the acid tubular fluid. Collecting duct cells also secrete NH_3 by way of glycoproteins that are in the family of the Rh factor, red blood cell ammonia transporters.

Regulation is accomplished at a number of levels. Hormones such as angiotensin II and catecholamines stimulate Na^+ reabsorption in the proximal tubule by increasing sodium-hydrogen exchange and $NaHCO_3$ cotransport. Aldosterone increases H^+-ATPase in the collecting duct cell and stimulates Na^+ reabsorption, thereby increasing proton secretion. Low extracellular fluid volume increases proximal HCO_3^- reabsorption, as does hypokalemia and high P_{CO_2}. Hyperkalemia (Chapter 117) may limit urinary acidification by decreasing the NH_3 entering the countercurrent multiplier in the loop of Henle and decreasing H^+ secretion by ATPases in the collecting duct as the need to secrete K^+ predominates.

The liver also contributes to urinary net acid excretion. Acidosis decreases activity of the urea cycle, which otherwise consumes NH_3 and HCO_3^- to form carbamoyl phosphate. In acidosis, NH_3 produces hepatic glutamine, which is delivered to the proximal tubule cell, where mitochondrial enzymes produce $2NH_3$ to assist in urinary net acid excretion; the associated increased α-ketoglutarate provides a substrate for renal gluconeogenesis.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Assessment of clinical acid-base disturbances usually begins with measurement of arterial blood gases (Fig. 118-1 and Table 118-3). In some situations, venous blood can be used as an alternative.

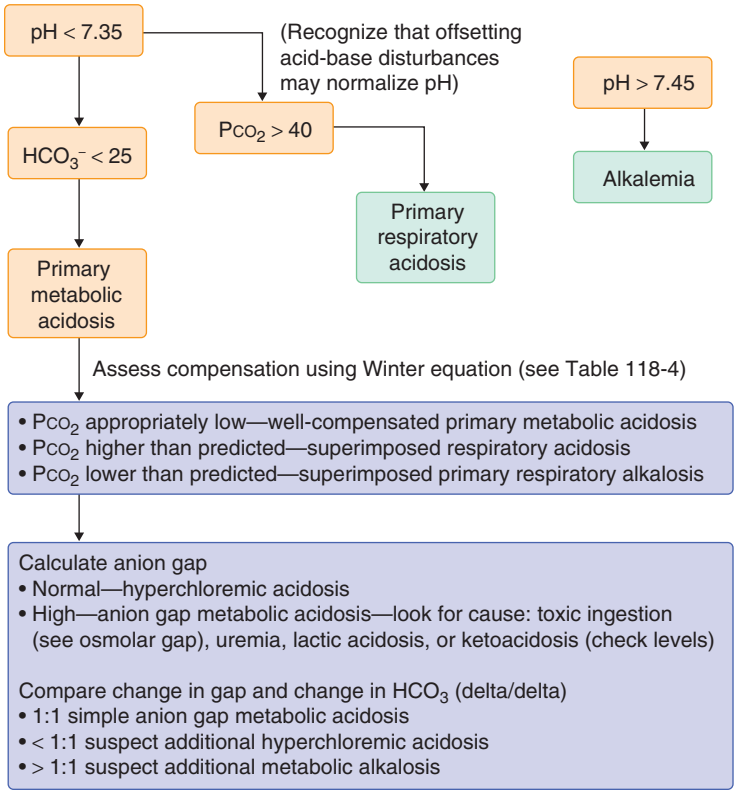

FIGURE 118-1. Evaluation of acidemia.

TABLE 118-3 LABORATORY STEPS IN IDENTIFYING ACID-BASE DISORDERS

EVALUATE pH	ACIDEMIC	ALKALEMIC
Elevated P_{CO_2}	Respiratory acidosis	Metabolic alkalosis
Elevated HCO_3	Respiratory acidosis	Metabolic alkalosis
Decreased P_{CO_2}	Metabolic acidosis	Respiratory alkalosis
Decreased HCO_3	Metabolic acidosis	Respiratory alkalosis
EVALUATE FOR EXPECTED COMPENSATION		
Meets expectation: simple disorder with compensation or could be offsetting metabolic alkalosis and acidosis		
Does not meet expectation: complex disorder, but pH indicates whether acidosis or alkalosis is dominant		
If a metabolic disorder is dominant, a P_{CO_2} greater than predicted indicates an additional respiratory acidosis. A P_{CO_2} less than predicted indicates an additional respiratory alkalosis.		
If a respiratory disorder is dominant, an HCO_3 concentration greater than predicted indicates additional metabolic alkalosis. An HCO_3 concentration less than predicted indicates an additional metabolic acidosis.		
ASSESS ANION GAP		
Elevated: metabolic acidosis is present whether acidemic or alkalemic. If alkalemic, an additional metabolic or respiratory alkalosis is present.		
If the gap is greater than the fall in HCO_3, consider an additional metabolic alkalosis or respiratory acidosis.		
If the gap is less than the fall in HCO_3, consider an additional nongap acidosis or respiratory alkalosis.		

It is useful to conceptualize acid-base disorders by a mass action shift of the variables to the right or left in the following relationship:

$$H^+ + HCO_3^- \leftrightarrow H_2CO_3 \leftrightarrow CO_2 + H_2O$$

The addition of CO_2, as in respiratory acidosis, will increase the hydrogen and bicarbonate concentrations (left shift). Removal of CO_2, as in respiratory alkalosis, will decrease CO_2, protons, and bicarbonate (shift right). The addition of protons with an anion other than HCO_3^-, as in metabolic acidosis, will lead to increased concentrations of protons and a decreased bicarbonate concentration. Removal of HCO_3^- with a cation such as Na^+, also a cause of metabolic acidosis, will increase the proton concentration and lower the HCO_3^- concentration. Metabolic alkalosis might be caused by the addition of $NaHCO_3$ with a resulting decrease in the proton concentration or by removal of H^+ with chloride, thereby leading to a decreased proton concentration and increased HCO_3^-. Because of the relationship of metabolic acidosis and alkalosis to the gain or loss of fluids and electrolytes, one could consider acid-base disorders a consequence of electrolyte imbalance.[3]

TABLE 118-4 EXPECTED DEGREES OF COMPENSATION IN ACID-BASE DISORDERS

DISORDER	EXPECTED COMPENSATION
Metabolic acidosis	Steady state in 12-36 hours Expected P_{CO_2} = 1.5 (measured HCO_3) + 8 ± 2 (Winter equation)
Metabolic alkalosis	Less predictable Expected P_{CO_2} increases 0.5 mm Hg per 1-mEq/L increase in HCO_3
Respiratory acidosis	
Acute	Expected 1-mEq/L increase in HCO_3 per 10-mm Hg rise in P_{CO_2}
Chronic, 24-36 hours	Expected 3- to 5-mEq/L increase in HCO_3 per 10-mm Hg rise in P_{CO_2}
Respiratory alkalosis	
Acute	Expected 1- to 2-mEq/L fall in HCO_3 per 10-mm Hg fall in P_{CO_2}
Chronic, after 24-36 hours	Expected 5-mEq/L fall in HCO_3 per 10-mm Hg fall in P_{CO_2}

Compensatory Changes

Few patients have an isolated acid-base disturbance. In nearly all cases, a respiratory or renal compensation (or both) occurs in response to counteract a primary acid-base process.

When they are functioning normally, the lungs may maintain a normal pH and P_{CO_2} during changes in volatile acid production. The kidneys will also maintain normal acid-base balance during changes in fixed acid production. Only excesses beyond the capacity to eliminate an acid or alkali load will lead to clinical disturbances. It follows that patients with renal or lung disease may do less well in response to metabolic and respiratory disorders.

When an acid-base disturbance develops, the initial response to modulate its severity depends on the titration of various body buffer pairs. For example, phosphate, hemoglobin, and albumin change their protonated and unprotonated concentrations. The body will then further attempt to correct the extracellular pH toward normal but usually not to normal. For metabolic disturbances caused by increased or decreased nonvolatile acid, the response is respiratory; for primary respiratory acidosis and alkalosis, the compensation is renal (Table 118-4). The direction of change in HCO_3^- and P_{CO_2} is the same when the primary disturbance is compensated; the ratio of HCO_3^- to P_{CO_2} and thus pH become more normal. These compensations tend to take time, so acid-base disturbances, particularly the respiratory conditions, are classified as acute (lasting less than 24 to 48 hours) or chronic.

Peripheral blood does not demonstrate complete compensation for most acid-base disturbances, with the occasional exception of chronic respiratory alkalosis. Full compensation for metabolic acidosis would expend large amounts of respiratory muscle energy, which could limit a prolonged response. Full compensation for metabolic alkalosis would result in excessive hypoventilation and adverse effects on oxygenation. In contrast, the CNS closely regulates its pH, with nearly full correction within 1 to 2 days. Before this compensation occurs, acute alkalemia may be associated with cerebral vasoconstriction and ischemia, whereas acidemia may result in vasodilation and cerebral edema. Rapid changes in blood P_{CO_2} affect the CNS chemosensors more quickly than do changes in HCO_3^- because of the more rapid movement of nonionic CO_2 across the blood-brain barrier. Increases in CNS CO_2 lead to acidification of the medullary center interstitial fluid and an increased ventilatory drive. Decreases in CNS CO_2 (alkalinization of the respiratory center) lead to hypoventilation. Acid-base changes are reflected in the composition of the cerebrospinal fluid (CSF).

In metabolic acidosis, peripheral chemosensors in the carotid body stimulate the CNS to increase ventilation to reduce P_{CO_2}. The fall in peripheral P_{CO_2} will lead to dissolved CO_2 leaving the CNS ahead of HCO_3^-; the

alkalinization of the medullary center interstitial fluid will then slow the hyperventilatory response until a new steady state of hypocapnia is achieved. Patients may sense dyspnea or air hunger acutely with rapid and shallow respirations. In severe cases of metabolic acidemia, the respirations are deep and gasping, typical of Kussmaul breathing.

When the bicarbonate concentration increases as a result of metabolic alkalosis, a hypoventilatory response, signaled from the peripheral chemosensors, raises P_{CO_2}. As P_{CO_2} rises, the dissolved CO_2 will enter the CSF and will acidify the medullary respiratory center. The stimulus to breathe will, in part, antagonize the peripheral signal until a steady state of hypoventilation is reached.

The acute stimulus of hypercapnia to increase net renal acid excretion disappears in chronic respiratory acidosis when, at the elevated P_{CO_2}, carbonic acid production and elimination are again equal. However, the hypochloremia, brought about by the compensatory early excretion of NH_4Cl, and elevated serum HCO_3^-, maintained by the high P_{CO_2}, persist.

In respiratory alkalosis, the primary event is a fall in P_{CO_2} because of increased alveolar ventilation. On transition from acute to chronic respiratory alkalosis, the compensatory mechanisms that initially helped maintain a more normal systemic pH are no longer required as CO_2 production and elimination become equal. Thus, the initial compensatory decrease in renal acid excretion brought about by increased loss of filtered $NaHCO_3$ ceases, but low serum HCO_3^- and high serum Cl^- concentrations are maintained.

In identifying whether an acid-base disturbance is simple (a single disturbance with its compensation) or complex (multiple primary processes simultaneously present), it is useful to compare the expected compensation for a simple process with the observed parameters of the blood gases (see Table 118-3). For example, if P_{CO_2} is lower than would be predicted in a patient with a simple, compensated metabolic acidosis, an additional respiratory alkalosis must be driving the P_{CO_2} down. If P_{CO_2} is higher than would be predicted for a low bicarbonate level in a patient with metabolic acidosis, a coexistent respiratory acidosis is present.

METABOLIC ACIDOSIS

EPIDEMIOLOGY AND PATHOBIOLOGY

In metabolic acidosis, the primary change is a fall in serum bicarbonate. The compensatory response is to increase ventilation to reduce P_{CO_2}. Worsening acidosis elicits increasing alveolar ventilation.

Primary metabolic acidosis results from an imbalance between net acid production and net acid excretion (NAE) in the form of urinary ammonium excretion and acid phosphate excretion. Consider the following relationship, where U_x represents the urinary concentration and the urinary flow rate $\dot{V}$:

$$NAE = (U_{NH_4} \times \dot{V}) + (U_{phos} \times \dot{V}) - (U_{HCO_3^-} \times \dot{V})$$

In a normal steady-state condition, the rate of excretion of net acid must be equal to the rate of production. The normal production rate depends on diet. If net acid production is normal, metabolic acidosis could occur because of a failure to reabsorb bicarbonate or a failure to elaborate enough urinary buffers, as is the case in renal failure and renal tubular acidosis. An inequality also could develop if net acid production were excessive or if large extrarenal bicarbonate losses were unable to be matched by maximal adaptive increases in net acid excretion. Endogenous sources of acid include ketoacidosis and lactic acidosis, whereas exogenous sources might be acid metabolic products of ingested ethylene glycol or methanol. On occasion, strong inorganic acids may be ingested. When net acid is retained in body fluids, the serum bicarbonate concentration falls. However, maintenance of a constant low serum HCO_3^- concentration does not guarantee that there is a new steady state in which net acid production is equal to net acid excretion because body buffers such as carbonate salts of bone may become depleted by relentless acid retention, as in chronic kidney disease and distal renal tubular acidosis.

The causes of metabolic acidosis are usually categorized according to the presence of either hyperchloremia or an elevated serum anion gap. The serum anion gap is the net charge difference when the sum of chloride and bicarbonate is subtracted from the serum sodium concentration.[4]

$$\text{Anion gap} = Na^+ - (Cl^- + HCO_3^-)$$

The normal anion gap is due to the net unmeasured anionic charge associated predominantly with albumin. When acidemia is present, albumin is in a more protonated form, which lowers the normal gap. In alkalemia, the effect of pH is to increase the gap attributed to albumin. Each 1 g/dL of albumin contributes approximately 2.5 mEq/L to the normal anion gap. The anion gap may be low with hypoalbuminemia or with an increase in unmeasured cations, such as immunoglobulin G myeloma paraproteins, calcium, lithium, or magnesium. The anion gap may be high in the presence of unmeasured anions including sulfates, bromides, iodides, and immunoglobulin A myeloma light chains. When the anion gap is increased above the normal value of approximately 10 to 12 mEq/L by a non-chloride acid anion, an anion gap metabolic acidosis exists. The accompanying proton is responsible for lowering the serum bicarbonate concentration. The degree of increase in the anion gap, sometimes referred to as the *delta anion gap*, may be estimated by the difference between the observed anion gap and a normal value of 10 to 12 mEq/L. A similar calculation for a change in serum HCO_3^- can be made by subtracting the observed HCO_3^- from the normal value of about 25 mEq/L (the *delta* HCO_3^-). Comparison of the two values (the *delta-delta*) may help identify more complicated acid-base disorders. If the increase in the anion gap is larger than the decrease in serum HCO_3^-, an additional process is raising the HCO_3^- level. The patient may have a coexisting metabolic alkalosis or be compensating for chronic respiratory acidosis. If the decrease in serum HCO_3^- is larger than the increase in the anion gap, it is a sign of another process that raises the Cl^- while lowering the HCO_3^- level, such as an additional hyperchloremic acidosis or respiratory alkalosis. In most anion gap acidoses, the increase in anion gap and the decrease in HCO_3^- is not 1 : 1 because the excretion of urinary anions with Na^+ results in a hyperchloremic component to the acidosis. Conversely, any buffering of H^+ with non-HCO_3^- buffers will decrease the drop in HCO_3^- compared with the increase in anion gap. In severe cases of anion gap acidosis, Cl^- may be displaced into cells, thereby resulting in a higher anion gap compared with the decrease in HCO_3^-—a hypochloremic anion gap acidosis.

CLINICAL MANIFESTATIONS

The effects of metabolic acidosis depend on its rapidity of onset and severity.[5] Patients often complain of fatigue and dyspnea, particularly on exertion. Nausea and vomiting are common. On examination, deep respirations, often labored with the use of accessory muscles, may be detected acutely, but hyperventilation may be less notable with long-standing metabolic acidemia. Metabolic acidemia also may be associated with vasodilation, tachycardia, and hypotension (Chapter 106). The negative inotropic effect of acidemia on the heart can exacerbate septic shock (Chapter 108). The stress of either an underlying illness or an increase in adrenergic and corticosteroid activity associated with acidemia may elevate the peripheral white blood cell count and cause hyperglycemia. Other laboratory findings include variable degrees of hyperkalemia, hyperphosphatemia, and hyperuricemia as well as hypocalcemia as a result of decreased renal synthesis of 1,25-dihydroxyvitamin D.

Anion Gap Metabolic Acidoses

A variety of abnormalities can cause anion gap acidoses. One mnemonic for the common ones is *gold mark*, for *g*lycols (ethylene and propylene), *o*xoproline, *l*-lactate, *d*-lactate, *m*ethanol, *a*spirin, *r*enal failure, and *k*etoacidosis.[6] Because some causes are life-threatening, a rapid diagnosis is required. The osmolar gap should be calculated in all cases of anion gap acidosis (Table 118-5) because unmeasured toxic, nonionic alcohols that contribute to body osmolality but not to acidity oxidize to dangerous unmeasured organic acid anions that contribute only to the anion gap. The osmolar gap is defined as the difference between the measured and the calculated serum osmolality.

TABLE 118-5 CAUSES OF INCREASED ANION AND OSMOLAR GAPS

ANION GAP METABOLIC ACIDOSIS	OSMOLAR GAP
Uremia	No
Lactic acidosis	Variable/no
D-Lactic acidosis	No
Diabetic ketoacidosis	No
Starvation ketoacidosis	No
Alcoholic ketoacidosis	If ethanol is present
Ethylene glycol	Yes
Methanol	Yes
Salicylates	No
5-Oxoprolinuria (acetaminophen)	No

The serum osmolality should be measured by a freezing point depression technique and compared with the calculated osmolality.

$$\text{Calculated osmolality} = 2(Na^+) + (\text{Glucose [mg/dL]} \div 18) + (\text{Blood urea nitrogen [mg/dL]} \div 2.8)$$

UREMIC ACIDOSIS

The metabolic acidosis of advanced chronic kidney disease (Chapter 130) may be due to tubular leakage of HCO_3^-, but it is often present when inadequate ammonia production is unable to facilitate excretion of the normal metabolic acid load. Many patients with renal failure can acidify their urine, but the lack of buffering capacity diminishes net acid excretion. Many organic and inorganic anions, such as phosphate and sulfates, are retained at glomerular filtration rates of less than 25 mL/minute and constitute an increased anion gap in association with the metabolic acidosis. The magnitude of the gap is usually less than 20 mEq/L consisting of poorly filtered sulfates and phosphates. The renal patient who is maximally producing NH_3 to stay in balance with daily acid production may be unable to accommodate any further acid production, such as a metabolic or respiratory acidosis, that would require increased ammoniagenesis. The patient with a poor glomerular filtration will retain HCO_3^-, thereby worsening both metabolic and respiratory alkalosis. The systemic acid-base disturbance in renal diseases with prominent tubular dysfunction is attributable to the kidney's inability to secrete hydrogen and to reabsorb and generate HCO_3^-. It is particularly pronounced in oliguric acute kidney injury and is exacerbated by hypercatabolic states such as infection. A significant metabolic acidosis in a patient with chronic kidney disease of unknown cause should raise the possibility of urinary tract obstruction (Chapter 123) or chronic tubulointerstitial diseases (Chapter 122), including amyloidosis (Chapter 188), myeloma (Chapter 187), autoimmune disorders, and analgesic nephropathy (Chapter 122).

It is important to treat the metabolic acidosis of chronic kidney disease.[7,8] Maintaining the serum HCO_3^- concentration above 20 to 22 mEq/L, by administering $NaHCO_3$ at a rate of 1 mEq HCO_3^-/kg/day, will slow the progression of chronic kidney disease, delay end-stage renal failure,[■] and improve nutritional status.[■]

PROGNOSIS

In population studies, a low serum bicarbonate level is associated with higher all-cause mortality. The relative risk of death is about 2.6-fold higher in patients with chronic kidney disease and about 1.7-fold higher even without it.

OVERPRODUCTION OF ENDOGENOUS ACIDS

Lactic Acidosis

EPIDEMIOLOGY AND PATHOBIOLOGY

Overproduction of lactate may occur with severe exertion, but true lactic acidosis is frequently associated with critical illness, multiorgan failure, and increased mortality. Lactate, which is the final product in the anaerobic pathway of glucose metabolism, is produced from pyruvate in a reaction catalyzed by lactate dehydrogenase:

$$NADH + \text{pyruvate} + H^+ \rightarrow \text{lactate} + NAD$$

A high reduced nicotinamide adenine dinucleotide (NADH)/NAD ratio will favor the formation of lactate. Conversion of ethanol to acetaldehyde and conversion of β-hydroxybutyrate to acetoacetate use NAD and produce NADH. Alcohol metabolism may be associated with excessive β-hydroxybutyrate and lactic acidosis.

Lactic acidosis is caused by an imbalance in the rates of lactate production and its clearance, primarily in the liver. Lactic acidosis, which increases the anion gap, is most often due to circulatory failure, hypoxia, and mitochondrial dysfunction that each increase anaerobic glycolysis and the rate of conversion of pyruvate to lactate. Sepsis (Chapter 108) is associated with an elevated lactate level because of poor clearance and impaired gluconeogenesis. Lactic acidosis can also result from seizure activity (Chapter 403) when lactate is released from muscle cells that have sustained a period of anaerobic metabolism. Other causes include thiamine deficiency (Chapter 214), hypophosphatemia (Chapter 119), isoniazid toxicity (Chapter 110), and hypoglycemic states (Chapter 230).[9] Metformin may cause lactic acidosis, particularly in elderly patients with cardiac, hepatic, or renal dysfunction.

Nucleoside antivirals (Chapter 388), including zidovudine, may cause lactic acidosis and abnormal liver function as a result of toxic mitochondrial effects. Abnormal mitochondrial function is also a feature of aspirin overdose (Chapter 37) or toxicity with hypoglycin from ingestion of the unripe ackee fruit (Jamaican vomiting sickness). The antibiotic linezolid is another cause of lactic acidosis.

Lactic acidosis can also be caused by the overproduction of lactate, which may occur with severe exertion and malignant neoplasms, particularly with a large tumor burden from lymphoma or widely metastatic cancer. Malignant cells can upregulate glycolytic activity, which may increase their uptake of glucose and decrease their dependence on mitochondrion-derived energy. These tumors can use large amounts of available glucose and inorganic phosphate, thereby leading to a syndrome of hypoglycemia, hypophosphatemia, and lactic acidosis.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

In any patient with an anion gap acidosis, the serum lactate level should be directly measured. Glucose, creatinine, and blood urea nitrogen levels also should be obtained. In cases in which a toxic ingestion is suspected (see Table 118-5), a screen for such toxins in the serum should be performed.

TREATMENT

Rx

Treatment of lactic acidosis is aimed at correction of the underlying cause. Central venous oxygen saturation should be increased, with a goal of at least 70%, by restoring tissue perfusion and ventilation, but specific therapy to increase the clearance of lactate is not of significant incremental value.[■] In a randomized trial of patients with lactic acidosis (pH of 6.9 to 7.2 and an average lactate level of 7.8 mM) in an intensive care unit, for example, sodium bicarbonate infused at a rate of 2 mEq/kg per 15 minutes did not improve hemodynamics, despite improvement in pH, but adversely lowered ionized serum calcium compared with saline.[■] Sodium bicarbonate can be considered when the arterial pH is below 7.0 or when acidemia has resulted in decreased cardiac inotropy or systemic vasodilation and shock. It is preferable to give $NaHCO_3$ as an isotonic mixture in 5% dextrose and water, rather than as a hypertonic bolus, because a hypertonic bolus carries the risk of pulmonary edema and hypernatremia. The quantity of administered sodium bicarbonate to raise arterial pH to 7.2 should be estimated by multiplying the desired minus observed bicarbonate concentration by 50% of body weight. Full correction should be avoided.

In patients with a metabolic acidosis after seizures (Chapter 403), the lactate is quickly metabolized to HCO_3^- by the liver and kidneys, and the acidosis often resolves within 60 minutes. The administration of HCO_3^- is usually unnecessary and may precipitate an overshoot metabolic alkalosis as the lactate is metabolized, which lowers the seizure threshold.

In patients with intestinal bacterial overgrowth (Chapter 140), a syndrome of disorientation, ataxia, and anion gap metabolic acidosis may develop after a carbohydrate meal because of bacterial production of D-lactate. This isomer of the mammalian L-lactate can be measured only by a specific D-lactate assay. The condition is treated with oral antibiotics and appropriate diet.

PROGNOSIS

Lactic acidosis, when severe, is associated with a high early mortality. When the pH is less than 7.2, only 17% of patients who are admitted to an intensive care unit are ultimately discharged from the hospital.

Diabetic Ketoacidosis

EPIDEMIOLOGY AND PATHOBIOLOGY

Diabetic ketoacidosis is defined as hyperglycemia with metabolic acidosis resulting from generation of the acid anions β-hydroxybutyrate (a hydroxy acid) and acetoacetate (a keto acid) in response to insulin deficiency and elevated counter-regulatory hormones such as glucagon. It is most commonly seen in cases of type 1 diabetes mellitus but can occasionally be seen in type 2 diabetes mellitus (Chapter 229). In an urban population, African Americans of low socioeconomic status were found to have more frequent episodes of diabetic ketoacidosis.[10]

The lack of insulin increases lipolysis in adipose tissue; free fatty acids are transported to the liver, where hepatic mitochondria produce ketone bodies, including acetoacetate, from acetyl coenzyme A. In the presence of high NADH/NAD ratio, the more reduced form of β-hydroxybutyrate is produced.

CLINICAL MANIFESTATIONS

Symptoms include nausea, vomiting, anorexia, polydipsia, and polyuria. Patients often exhibit Kussmaul respirations and volume depletion.

Neurologic symptoms include fatigue and lethargy with depression of the sensorium. CSF exhibits a change in acid-base status with treatment of diabetic ketoacidosis. Even without bicarbonate administration, CSF pH falls as a result of the ventilatory response to the correction of acidosis and the sudden rise in P_{CO_2}. However, no correlation between decreased CSF pH and depression of sensorium has been established. Ketoacidosis is also seen in cases of starvation, in which it is generally mild and not associated with hyperglycemia.

Keto acids in the urine may be accompanied by cations, including sodium and potassium, thereby contributing to volume depletion, potassium depletion, relative chloride retention, and a mixed anion gap and hyperchloremic acidosis. The delta HCO_3^- will exceed the delta anion gap, especially if the glomerular filtration rate and the filtered load of keto acids are high. The serum anion gap in general will be greatest when renal failure is present because the additional anions cannot be cleared from extracellular fluid.

DIAGNOSIS

The urinary dipstick nitroprusside test for ketones may underestimate the degree of ketosis because it does not detect β-hydroxybutyrate; in fact, the ketone test result may become more positive as treatment helps metabolize β-hydroxybutyrate to acetoacetate. This problem should be addressed by direct measurement of serum β-hydroxybutyrate. Diabetic patients also are more prone to lactic acidosis because an increase in NADH favors the formation of lactate from pyruvate, and pyruvate dehydrogenase is inhibited in the absence of insulin.

TREATMENT Rx

Treatment of diabetic ketoacidosis (Chapter 229) consists of volume repletion, insulin administration with dextrose if necessary to avoid hypoglycemia, and potassium replacement (Chapter 117). Bicarbonate administration should be considered only if ketoacidosis is accompanied by shock or if arterial pH is less than 7.0 or 7.1, and bolus infusion should be avoided. The administration of bicarbonate occasionally results in cerebral edema significant enough to lead to loss of consciousness and even death.

PROGNOSIS

Most patients with diabetic ketoacidosis recover. In the less than 0.5% of patients who present with coma from cerebral edema, the mortality rate ranges from 20% to as high as 90%. The cerebral edema may be exacerbated by bicarbonate administration, which is discouraged in this situation.

Salicylate Intoxication

EPIDEMIOLOGY AND PATHOBIOLOGY

Salicylate intoxication can be caused by accidental overdose, therapeutic overdose, or a suicide attempt (Chapters 37 and 110). Salicylate functions as an uncoupler of oxidative phosphorylation and consequently results in increased oxygen consumption and CO_2 production. However, the increase in alveolar ventilation resulting from stimulation of central chemoreceptors overcomes this increase in CO_2.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The most common clinical manifestation is a combined anion gap metabolic acidosis and respiratory alkalosis, although the condition also can be manifested as either one or the other only. Children are often seen with metabolic acidosis, whereas adults often have predominant respiratory alkalosis. Hypoglycemia, ketoacidosis, and lactic acidosis may result. Other manifestations of intoxication include hemorrhage, fever, nausea and vomiting, hyperventilation, diaphoresis, tinnitus, and occasionally polyuria followed by oliguria. Severe cases may lead to seizures, respiratory depression, and coma. Noncardiogenic pulmonary edema is sometimes seen in adults.

Respiratory alkalosis is the result of a direct stimulatory effect of salicylate on the medullary respiratory control center. Salicylate intoxication also increases the metabolic rate. Diagnosis is suspected by the clinical presentation and confirmed by the salicylate level (Chapters 37 and 110).

Treatment of salicylate intoxication (Chapter 110) is aimed at correction of the metabolic acidosis and removal of salicylate. Bicarbonate as a sodium salt should be administered according to an estimated calculation of the deficit if metabolic acidosis predominates. Salicylates are removed by alkaline diuresis because the less reabsorbable salicylate anion will predominate when the urine pH increases. Urinary alkalinization with acetazolamide should be used cautiously because carbonic anhydrase inhibition may impair CO_2 transport from tissue to blood and potentially worsen acidosis in the respiratory center. In severe intoxication (salicylate concentrations greater than 35 mg/dL) or when renal failure is present, dialysis may be required.

PROGNOSIS

The prognosis of salicylate toxicity is better with early diagnosis and prompt management, in which case most patients do well. Patients who ingest oil of wintergreen (methylsalicylate) may have more severe deterioration because of the high lipid solubility of the drug.

Alcoholic Ketoacidosis

EPIDEMIOLOGY AND PATHOBIOLOGY

Alcoholic ketoacidosis occurs in a patient who has been drinking very heavily without eating. The pathophysiologic mechanism is based on the overproduction of β-hydroxybutyrate and, to a lesser extent, acetoacetate because of an increased production of free fatty acids from adipose tissue. Alcohol inhibits the conversion of lactate to glucose in the liver. The oxidation of ethanol increases the ratio of NADH to NAD^+ and favors the production of β-hydroxybutyrate from acetoacetate. Damage to mitochondria by alcohol can further elevate the ratio of β-hydroxybutyrate to acetoacetate by preventing reoxidation of NADH to NAD. The oxidative metabolism of ethanol favors the reaction of dehydrogenase enzymes to form β-hydroxybutyrate and lactate (opposing glucose production).

CLINICAL MANIFESTATIONS

Alcoholic ketoacidosis usually follows binge drinking and may be associated with withdrawal symptoms (Chapters 33 and 416) and the associated hyperadrenergic state. Alcoholic ketoacidosis is associated with abdominal pain, vomiting, starvation, and volume depletion. In contrast to diabetic ketoacidosis, coma is rare. Blood glucose levels are generally low or normal, and the insulin level is frequently low, with elevated glucagon (favoring ketogenesis) and cortisol levels. Some patients have hyperglycemia because of the increased catecholamine response.

DIAGNOSIS

Patients typically have a high osmolal gap initially. Blood alcohol levels may be absent or elevated on initial evaluation. A clue to the diagnosis of toxic alcohol ingestion is the simultaneous presence of an anion gap metabolic acidosis and an osmolal gap. This osmolal gap, if secondary to ethanol, should be equal to the ethanol concentration in milligrams per deciliter divided by 4.6. If this calculation does not yield the expected gap based on the ethanol concentration, ingestion of another alcohol, such as methanol, isopropanol, or ethylene glycol, should be suspected (see Table 118-2).

If possible, ethanol, ethylene glycol, propylene glycol, and methanol levels should be measured directly; each is associated with a metabolic acidosis. In contrast, isopropanol metabolizes to acetone and causes ketosis without acidosis. An osmolar gap may be present.

TREATMENT Rx

Treatment of alcoholic metabolic acidosis consists of volume repletion with normal saline in dextrose; the administration of thiamine (50 to 100 mg intravenously) and enough glucose to treat hypoglycemia; and the correction of any hypophosphatemia (Chapter 119), hypokalemia (Chapter 117), and hypomagnesemia (Chapter 119) that may be present. The acid-base disturbance usually resolves after several hours. Both hypophosphatemia and thiamine deficiency, which may not be apparent until 12 to 24 hours after the initiation of treatment in an undernourished patient, are exacerbated by glucose administration and may contribute to an associated lactic acidosis.

PROGNOSIS

The prognosis of alcoholic ketoacidosis is usually favorable. The long-term outlook is more closely tied to other complications of continued alcohol abuse.

5-Oxoprolinuria

5-Oxoprolinuria, which is an acquired form of anion gap metabolic acidosis, is increasingly recognized in patients who have glutathione depletion associated with underlying chronic illness, malnutrition, diabetes, alcoholism, or cancer, especially in the context of a high therapeutic level or overdose of acetaminophen.[11] It has been observed without concomitant hepatic failure. Glutathione, which is a tripeptide consisting of glutamate, cysteine, and glycine, has many functions, including protection from cellular toxins, combating of oxidative stress, and shuttling of amino acids into the cytosol by the γ-glutamyl transpeptidase pathway. Hereditary forms of 5-oxoprolinuria are associated with γ-glutamyl transpeptidase cycle enzyme deficiencies (5-oxoprolinase and glutathione synthase). The usual production of 5-oxoproline is by release of the transported amino acid from the γ-glutamyl–amino acid dipeptide. The enzyme 5-oxoprolinase re-forms glutamate to re-enter the cycle as glutamylcysteine. Acetaminophen causes further depletion of glutathione due to binding of an acetaminophen metabolite, *N*-acetyl-*p*-benzoquinone imine (NAPQI), to glutathione. Treatment with *N*-acetylcysteine, in doses similar to those used for acetaminophen toxicity (Chapter 110), should be considered to decrease further glutathione depletion.

Ethylene Glycol

Ethylene glycol (Chapter 110)[12] is commonly found in antifreeze and is used as an industrial solvent. It has a sweet taste, and patients occasionally ingest it as a substitute for ethanol. Although ethylene glycol itself is not particularly damaging, its highly toxic metabolites include glyoxylate, glycolate, oxalic acid, and ketoaldehydes. Glycolic acid appears to be primarily responsible for the metabolic acidosis observed in this condition.

Intoxication is characterized by profound CNS symptoms, including diplopia, seizures and coma, severe metabolic acidosis, cardiac failure, pulmonary failure, and renal failure. Patients are often dehydrated and hypernatremic because of osmotic diuresis from the renal excretion of the alcohol.

An increased anion gap is attributable to ethylene glycol metabolites. A high osmolal gap will also be present because of the uncharged alcohol. However, an osmolal gap may not be present if all of the alcohol has been converted to the toxic anionic forms. Calcium oxalate crystals in the urine may cause intratubular obstruction and acute kidney injury. Treatment is aimed at rehydration with saline and correction of acidosis with $NaHCO_3$ based on an estimate of the bicarbonate deficit. When an osmolal gap exists, competitive inhibition of alcohol dehydrogenase should be initiated with fomepizole at a loading dose of 15 to 20 mg/kg intravenously in 100 mL normal saline during 30 minutes to 1 hour, followed by a maintenance dose of 10 mg/kg every 12 hours (Chapter 110). If ethanol is used, a solution of 10% ethanol in 5% dextrose can be given as a loading dose of 0.6 g/kg intravenously followed by a maintenance dose of 150 mg/kg per hour in alcoholic patients or 65 mg/kg per hour in nonalcoholic patients. The ethanol level should be maintained at 100 to 200 mg/dL. The goal of therapy is early recognition to prevent metabolism of the uncharged glycol to acidic products. Hemodialysis is required in severe cases. If the diagnosis is made promptly and appropriate therapy is instituted, outcomes are favorable. Renal failure may be reversible.

Methanol

Methanol, wood alcohol, is a component of shellac and windshield wiper fluid and is highly toxic to the CNS after metabolism to formaldehyde and formic acid. Some automotive fluids now contain the less toxic propylene glycol. Optic papillitis may cause blindness. Detection may be made easier if fluorescein has been added to the methanol-containing fluid.

Treatment consists of competitive inhibitors for alcohol dehydrogenase, including ethanol and fomepizole, in similar amounts as for ethylene glycol poisoning, to reduce the formation of acid anions and the anion gap while maintaining a higher level of methanol in the blood (Chapter 110). Hemodialysis may be necessary to increase elimination. Early diagnosis and treatment are associated with a favorable outcome, but visual loss may be permanent. Late presentation is associated with a poor prognosis, particularly if the amount consumed exceeds 30 mL. As with ethylene glycol, the simultaneous presence of ethanol on presentation may help slow the metabolism of methanol and improve outcome.

Isopropyl Alcohol

Toxic ingestion of isopropyl alcohol (Chapter 110), as in rubbing alcohol, does not cause an increased anion gap or ketoacidosis because the metabolite is acetone, but test results for ketones are positive, and a high osmolal gap will be present.

Propylene Glycol

On occasion, patients in the intensive care unit setting are given high doses of intravenous benzodiazepines, such as lorazepam or diazepam, that contain propylene glycol as a diluent. Other intravenous medications that also contain this diluent include phenobarbital, phenytoin, nitroglycerine, and esmolol. Propylene glycol has also been used as a less toxic substitute for methanol in windshield wiper fluid. A high osmolal gap may develop because of the propylene glycol and lead to a clinical picture of sedation, failure to be weaned from the respirator, and an increased lactate level. Propylene glycol, a 3-carbon glycol, oxidizes to lactate and pyruvate. Treatment, which consists of early recognition and withdrawal of the offending agent, usually results in a favorable prognosis.

Hyperchloremic (Normal Anion Gap) Acidosis

Hyperchloremic metabolic acidoses (see Table 118-1) can be caused by renal or nonrenal mechanisms and can be associated with an elevated, normal, or low serum potassium level.

HYPERCHLOREMIC METABOLIC ACIDOSIS OF NONRENAL ORIGIN ASSOCIATED WITH NORMAL OR INCREASED POTASSIUM LEVEL

Hyperchloremic metabolic acidoses with a normal or elevated potassium concentration can develop as a result of the addition of chloride salts such as NaCl, KCl, $CaCl_2$, NH_4Cl, arginine and lysine hydrochlorides, or HCl itself. If the quantity of Cl^- introduced exceeds the ability of the kidney to eliminate Cl^- salts in urine, hyperchloremia will develop. Electroneutrality is maintained by a decrease in the serum HCO_3^- concentration, and a hyperchloremic acidosis ensues. Renal production of NH_3 increases in an attempt to improve HCl (NH_4Cl) excretion. Hyperkalemia can occur because the acidemia favors the exit of K^+ from cells. Acidemia also inhibits K^+ secretion in the renal collecting duct.

HYPERCHLOREMIC METABOLIC ACIDOSIS OF NONRENAL ORIGIN ASSOCIATED WITH HYPOKALEMIA

Hypokalemic, hyperchloremic acidosis may result from loss of a body fluid that is low in Cl^- relative to Na^+ and K^+ compared with the ratio of Cl^- to Na^+ in extracellular fluid. For example, stool losses of Na^+, K^+, and HCO_3^- in small bowel diarrhea or organic acid anions of bacterial origin, such as butyrate, in colonic diarrhea lead to hyperchloremic acidosis (Chapter 140). Pancreatic secretions (Chapter 195) or heavy losses from ileostomy sites may lead to loss of bicarbonate-containing fluids. Secretagogues such as vasoactive intestinal peptide (VIP), which is associated with neoplasms of the pancreas or sympathetic chain (Chapter 195), cause large losses of HCO_3^- in stool, with a resulting hypokalemic, hyperchloremic metabolic acidosis. Concomitant gastric achlorhydria is part of the syndrome known as *watery diarrhea, hypokalemic, hypochlorhydric acidosis*. Urinary diversions, such as ureterosigmoidostomies and ileal loops, may increase chloride absorption in exchange for bicarbonate in the intestinal segment and lead to hyperchloremic acidosis. In the presence of urea-splitting bacteria, the net absorption of NH_4Cl can result in both hyperchloremic acidosis and hyperammonemia.

RENAL TUBULAR ACIDOSIS TYPES 1 AND 2

Proximal Renal Tubular Acidosis

Renal tubular acidosis causes the cations Na^+ and K^+ to be lost in the urine with HCO_3^- rather than with Cl^-, thereby leading to hyperchloremia. Proximal renal tubular acidosis (type 2) is characterized by a decreased threshold for bicarbonate reabsorption. HCO_3^- wasting and concomitant urinary losses of potassium occur until a lower level of serum bicarbonate reduces the filtered HCO_3^- to a level that the combined function of the dysfunctional proximal tubule and distal nephron can completely reabsorb. At that point, the urine becomes acid (pH < 5.3), and net acid production equals net acid excretion, with a steady-state low plasma HCO_3^-.

Isolated proximal renal tubular acidosis may result from mutations of specific transporters of the proximal tubule, such as the $NaHCO_3$ cotransporter, or from hereditary deficiency of carbonic anhydrase. More commonly, proximal renal tubular acidosis is associated with the Fanconi syndrome or generalized proximal tubule dysfunction. Causes (Table 118-6) include genetic diseases such as glucose-6-phosphatase deficiency (Chapter 161), cystinosis (Chapter 128), hereditary fructose intolerance (Chapter 205), and Wilson disease (Chapter 211). Multiple myeloma (Chapter 187) and Sjögren

TABLE 118-6 CAUSES OF RENAL TUBULAR ACIDOSIS*

HYPOKALEMIC DISTAL (TYPE 1) RTA
Hereditary tubule disorders
Vacuolar H^+-ATPase β-subunit gene mutations
Carbonic anhydrase type II deficiency
Cl^-/HCO_3^- exchanger (AE1) mutations
Genetic causes
Sickle cell
Fabry disease
Wilson disease
Elliptocytosis
Paroxysmal nocturnal hemoglobinuria
Medullary cystic kidneys
Autoimmune disorders
Systemic lupus erythematosus
Sjögren syndrome
Multiple myeloma and amyloidosis
Drugs: amphotericin, cisplatinum, aminoglycosides
Nephrocalcinosis and hypercalcemic disorders
Tubulointerstitial diseases
Acute tubulointerstitial nephritis
Reflux nephropathy
Analgesic nephropathy
PROXIMAL (TYPE 2) RTA
Hereditary tubule disorders
$NaHCO_3$ cotransport (NBC) mutations
Carbonic anhydrase deficiency
Generalized proximal tubular dysfunction
Hereditary Fanconi syndrome
Genetic diseases: cystinosis, glycogen storage disease (glucose-6-phosphatase deficiency), Wilson disease
Hormonal: hyperparathyroidism, vitamin D deficiency
Multiple myeloma, lysozymuria
Sjögren syndrome
Renal transplantation
Heavy metals: cobalt, mercury, lead
Drugs: ifosfamide, outdated tetracycline, tenofovir, tacrolimus, aminoglycosides
HYPERKALEMIC (TYPE 4) RTA
Renal diseases—aldosterone resistance
Diabetes mellitus
Amyloidosis
Systemic lupus erythematosus
Urinary tract obstruction
Hyporeninism
Autonomic neuropathy (diabetic)
Sickle cell anemia
Primary hypoaldosteronism
Adrenal insufficiency: Addison disease
Tubular mutations: pseudohypoaldosteronism
Drugs: potassium-sparing diuretics, amiloride, triamterene, spironolactone, nonsteroidal anti-inflammatory drugs, lithium, trimethoprim, cyclosporine, tacrolimus, renin inhibitors, angiotensin-converting enzyme inhibitors, angiotensin II receptor antagonists

*Type 3 renal tubular acidosis (RTA) is not listed separately because it is an overlap of proximal and distal dysfunction.

syndrome (Chapter 268) should be considered in an adult patient. Primary hyperparathyroidism (Chapter 245) results in proximal renal tubular acidosis and hypophosphatemia secondary to inhibition of Na^+/H^+ exchange and sodium phosphate cotransport in the proximal tubule by parathyroid hormone through cyclic adenosine monophosphate. Hyperparathyroidism is one of the few causes of metabolic acidosis with hypercalcemia. The Cl^-/phosphate ratio in plasma may be elevated. Drug toxicity with aminoglycosides, cisplatin, and ifosfamide may cause proximal tubule dysfunction. The antiretroviral drug tenofovir, a nucleotide analogue reverse transcriptase inhibitor, is a cause of the Fanconi syndrome. The syndrome also may be seen after kidney transplantation (Chapter 131).

Distal Renal Tubular Acidosis

In distal renal tubular acidosis (type 1), failure to produce ammonia leads to an inability to excrete adequate net acid, thereby leading to continuous retention of acid in the body. The degree of acidemia is often severe, with pH reaching values as low as 7.2, whereas urine pH usually exceeds 5.3.

Kindreds have been described in which mutations in genes for the distal vacuolar H^+-ATPase cause an autosomal recessive distal renal tubular acidosis with deafness. Mutations resulting in defective Cl^-/HCO_3^- exchange protein (AE1) have been linked to an autosomal dominant form of distal renal tubular acidosis.

Distal renal tubular acidosis (see Table 118-6) is also associated with autoimmune disorders, including systemic lupus erythematosus (Chapter 266) and Sjögren syndrome (Chapter 268), and genetic diseases, including sickle cell anemia (Chapter 163), Wilson disease (Chapter 211), Fabry disease (Chapter 208), cystic kidney diseases (Chapter 127), and hereditary elliptocytosis (Chapter 161). Hypercalciuria and hyperoxaluria may cause distal renal tubular acidosis; nephrocalcinosis and nephrolithiasis may be present. Increased proximal tubular citrate reabsorption, as a consequence of the chronic acidosis, also leads to hypocitraturia, which is a risk factor for calcium nephrolithiasis (Chapter 126). A chronically alkaline urine is a risk for pure $CaHPO_4$ stones (brushite). Amyloidosis (Chapter 188) may be manifested as severe acidemia and other tubular dysfunction, including nephrogenic diabetes insipidus. Chronic tubulointerstitial diseases of the kidney (Chapter 122), including reflux nephropathy (Chapter 128) and urinary obstruction (Chapter 123), may result in renal tubular acidosis with hypokalemia or hyperkalemia. Acute tubulointerstitial nephritis also may result in renal tubular acidosis. Drugs such as amphotericin B can cause hypokalemic distal renal tubular acidosis. Topiramate, used for migraines, is a carbonic anhydrase inhibitor that may cause mixed proximal and distal renal tubular acidosis.

HYPERCHLOREMIC METABOLIC ACIDOSIS OF RENAL ORIGIN ASSOCIATED WITH HYPERKALEMIA

Hyperkalemic, hyperchloremic acidosis (type 4) suggests dysfunction of the cortical collecting duct, where acidification of urine and disorders in potassium secretion may occur. Some patients with high blood potassium and hyperchloremic acidosis can lower urinary pH below 5.3, whereas others appear to have defects in both potassium balance and urinary acidification. Hyperkalemia itself may worsen metabolic acidosis by decreasing NH_3 accumulation by countercurrent multiplication in the medullary interstitium.

Causes include hyporenin-hypoaldosteronism, as seen in diabetic renal disease (Chapter 124); other tubulointerstitial diseases (Chapter 122), usually with some renal impairment; sickle cell anemia (Chapter 163); and the use of drugs such as β-blockers and nonsteroidal anti-inflammatory drugs. Low renin and aldosterone levels can also be found in cases of volume expansion with hypertension. Cyclosporine and tacrolimus may lead to decreased electrical driving forces for K^+ and H^+ secretion. Hyperkalemic acidosis with elevated renin and low aldosterone is found in adrenal insufficiency (Chapter 227), in isolated hypoaldosteronism (Chapter 227), and with the use of angiotensin-converting enzyme inhibitors, renin inhibitors, and angiotensin II receptor blockers. High renin and aldosterone levels are anticipated when the renal collecting duct cell is insensitive to aldosterone, as in urinary tract obstruction, sickle cell anemia, amyloidosis, and systemic lupus erythematosus. Inhibition of aldosterone action with spironolactone or eplerenone may cause hyperkalemic acidosis, as does ENaC inhibition by amiloride, triamterene, trimethoprim, and lithium.

Pseudohypoaldosteronism type 1 is due to autosomal recessive, inactivating mutations of the sodium channel ENaC, whereas autosomal dominant pseudohypoaldosteronism type 1 is due to mutations of the mineralocorticoid receptor. Both cause hypovolemia, metabolic acidosis, and hyperkalemia with secondary increases in renin and aldosterone. In Gordon syndrome (pseudohypoaldosteronism type 2), increases in Na^+ and Cl^- reabsorption through increased activity of the distal thiazide-sensitive NaCl transporter lead to hypertension, hyperkalemic acidosis, volume expansion, and consequently low renin and aldosterone.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The urinary anion, or charge, gap helps distinguish renal tubular acidosis[13] from extrarenal bicarbonate loss (e.g., from diarrhea). Because the normal renal response to metabolic acidosis is an increase in ammoniagenesis, the urine should contain large amounts of NH_4Cl while the kidney retains sodium and potassium; the urinary charge gap, which is $(Na^+ + K^+) - Cl^-$, should be strongly negative because of the unmeasured NH_4^+.

In renal diseases such as distal renal tubular acidosis, however, the urinary anion gap will be zero or positive because of either the failure of ammoniagenesis or the excretion of sodium plus potassium with bicarbonate. With type 2 (proximal) renal tubular acidosis, patients often have Fanconi

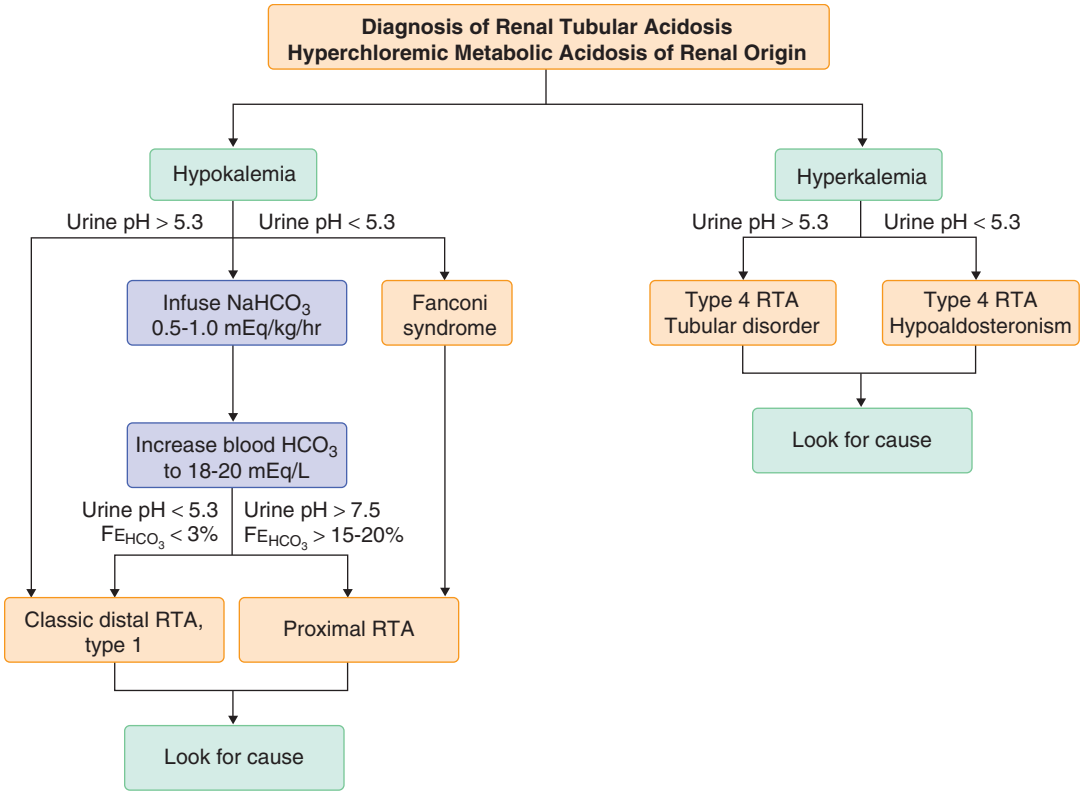

FIGURE 118-2. Diagnosis of renal tubular acidosis (RTA). FE = fractional excretion; $NaHCO_3$ = sodium bicarbonate.

syndrome with glycosuria, phosphaturia, aminoaciduria, and uricosuria. In proximal renal tubular acidosis, the steady-state urine pH is usually less than 5.3, the acidosis is not severe (i.e., HCO_3^- usually not less than 16), and acid excretion may balance acid production at this new steady state.

In contrast to proximal renal tubular acidosis, distal renal tubular acidosis (type 1) is generally a more severe metabolic disorder that may be accompanied by hypercalciuria, nephrocalcinosis, calcium phosphate kidney stones (Chapter 126), and bone disease that includes rickets in children and osteomalacia in adults. Proximal and distal renal tubular acidoses usually can be distinguished by a careful clinical evaluation (Fig. 118-2). Helpful findings include the presence of a urine pH greater than 5.3 in distal but not in proximal renal tubular acidosis during acidemia; a fractional excretion of bicarbonate as high as 10 to 15% in proximal renal tubular acidosis; and the lowering of serum potassium on correction of proximal but not of distal tubular acidosis.

In patients with an elevated serum anion gap, unmeasured anions such as keto acids and lactate, rather than NH_4^+, are present in urine, so a positive urinary anion gap does not indicate renal tubular acidosis. On occasion, however, the prompt renal excretion of organic anions with sodium and potassium may minimize the increase in the serum anion gap. In the metabolic acidosis of glue sniffers, hippurate, which is a product of toluene, is rapidly excreted, thus giving the appearance of a nongap metabolic acidosis with a positive urinary anion gap. Similarly, if keto acids are completely cleared into the urine, ketoacidosis may be manifested as a hyperchloremic acidosis rather than as an anion gap acidosis.

TREATMENT

Rx

If possible, treatment of metabolic acidosis should focus on correction of the underlying cause, such as discontinuation of an offending drug, permitting the body's homeostatic mechanisms to correct the acid-base disturbance.

Patients whose pH is less than 7.2 are typically treated with infusions of sodium bicarbonate, guided by the estimated base deficit in milliequivalents, calculated by the serum HCO_3^- concentration in milliequivalents per liter:

$$\text{Amount of } HCO_3^- = (25 - [HCO_3^-]) \times \text{wt (kg)}/2$$

In general, the correction of metabolic acidemia should be based on a calculated amount, with not more than 50% of the estimate given before recalculation. Moreover, this equation is used for deficit correction only; the ongoing losses of 1 to 2 mEq/kg per day, equivalent to the daily acid load, should be replaced in distal renal tubular acidosis with $NaHCO_3$, $KHCO_3$, or citrate salts in divided doses. Hypokalemia may accompany distal renal tubular acidosis and may improve with treatment. Citrate should be avoided as an alkalinizing salt in patients with low glomerular filtration rate.

Proximal renal tubular acidosis in children may affect growth and require large quantities of bicarbonate in excess of 1 to 2 mEq/kg per day to correct the acidosis because ingested alkali is promptly excreted in alkaline urine. In adults, treatment is often deferred because the steady-state acidosis allows a normal acid excretion rate. Hypokalemia may worsen with bicarbonate treatment of proximal tubular acidosis.

In type 4 renal tubular acidosis, treatment of hyperkalemia with a low-potassium diet, thiazide, or loop diuretics or sodium polystyrene sulfonate often improves urinary acidification without the use of bicarbonate salts.

PROGNOSIS

The prognosis of renal tubular acidosis generally depends on the presence of an underlying systemic disease, such as myeloma (Chapter 187). In children, disorders such as medullary cystic kidney disease (Chapter 127) and cystinosis (Chapter 128) usually result in renal failure by the teenage years. These patients are candidates for renal replacement therapy, including transplantation. Chronic metabolic acidosis in children, if not well treated, is associated with rickets (Chapter 244) and short stature.

METABOLIC ALKALOSIS

EPIDEMIOLOGY AND PATHOBIOLOGY

In metabolic alkalosis, the primary event is elevation of the plasma bicarbonate concentration. In response to increased systemic pH, alveolar ventilation is decreased to increase P_{CO_2} and thereby decrease pH. However, respiratory compensation is generally less effective in cases of metabolic alkalosis than in cases of metabolic acidosis. Contributing factors may include the fact that hypoventilation also decreases P_{O_2}, which is a potent stimulus for the peripheral chemoreceptors to increase alveolar ventilation when P_{O_2} falls below about 60 mm Hg. A second mechanism that may blunt respiratory compensation is intracellular acidosis in the brain in the setting of hypokalemia. In acute metabolic alkalosis, an initial paradoxical acidotic shift in CSF pH

secondary to a sudden increase in P_{CO_2}, analogous to the alkaline shift in CSF pH in acute metabolic acidosis, may activate central chemoreceptors and increase ventilatory drive despite peripheral stimulation to decrease alveolar ventilation. In chronic metabolic alkalosis, CSF pH may return to normal, so respiratory drive is controlled entirely by the peripheral chemoreceptors. The result is that the ventilatory response to metabolic alkalosis is highly varied: many patients with metabolic alkalosis maintain nearly normal P_{CO_2} levels, and the level rarely rises above 60 mm Hg.

Metabolic alkalosis requires a generation phase, in which new HCO_3^- is added to the extracellular fluid, and a maintenance phase, in which the new elevated serum HCO_3^- concentration is sustained. Without the maintenance phase, a kidney with normal filtration and tubular function has a high capacity to excrete HCO_3^-, thereby preventing alkalosis. Maintenance of a high HCO_3^- concentration usually occurs because of volume depletion, reduced glomerular filtration rate, hypokalemia, or low chloride levels.

Metabolic Alkalosis of Renal Origin Associated with Volume Depletion

Metabolic alkalosis of renal origin may be the result of excessive urinary chloride excretion, most commonly related to diuretics that inhibit reabsorption of Cl^-. The Cl^- loss results in hypochloremia, with a compensatory increase in plasma HCO_3^- to maintain electroneutrality. Extracellular volume depletion stimulates the renin-angiotensin-aldosterone pathway, and high aldosterone levels superimposed on increased distal urinary flow rates result in increased K^+ excretion and hypokalemia. The volume depletion and hypokalemia enhance proximal HCO_3^- reabsorption, thereby maintaining the alkalosis, and the prerenal fall in the glomerular flow rate limits HCO_3^- filtration.

Important but rare genetic syndromes characterized by urinary chloride wasting include Bartter syndrome and Gitelman syndrome. Bartter syndrome is an autosomal recessive salt-losing state associated with extracellular volume depletion and excessive urinary chloride loss that results in hypokalemia and hypochloremic metabolic alkalosis. Secondary increases of plasma renin and aldosterone occur, as does renal juxtaglomerular cell hyperplasia. The syndrome resembles the effects of furosemide on the thick ascending limb of Henle; gene mutations in the Na-K-2Cl cotransporter, the renal outer medullary potassium channel (ROMK), and chloride channels have been described. Because calcium reabsorption occurs in the thick ascending limb of Henle, Bartter syndrome (Chapter 128), like furosemide, causes hypercalciuria and nephrocalcinosis as well as polyuria due to decreased urinary concentrating ability.

Gitelman syndrome is an autosomal recessive cause of extracellular volume depletion, urinary chloride wasting, and hypokalemic metabolic alkalosis. It is due to inactivating mutations in the *SLC12A3* gene encoding the thiazide-sensitive NaCl cotransporter of the renal distal tubule. Urinary concentrating ability is preserved, and patients are hypocalciuric because decreased NaCl reabsorption in the distal tubule is associated with a decrease in calcium excretion. Hypomagnesemia may also be severe.

Metabolic Alkalosis of Nonrenal Origin with Extracellular Volume Depletion

Metabolic alkalosis may develop as a result of gastrointestinal Cl^- loss from vomiting, nasogastric suctioning, or secretory diarrhea. In such cases, extracellular volume is usually contracted, hypochloremia develops, and the urinary chloride level is usually less than 20 mEq/L.

In Zollinger-Ellison syndrome (Chapter 195), excessive gastrin-induced gastric acid secretion may result in an acidic stool with high chloride content. Diarrhea does not cause metabolic alkalosis unless the electrolyte relationship $[(Na^+ + K^+) - Cl^-]$ in the stool is less than plasma HCO_3^-.

Infectious gastroenteritis, congenital chloridorrhea, and villous adenomas also cause chloride losses in stool. Congenital chloridorrhea is an autosomal recessive disorder of defective intestinal apical Cl^-/HCO_3^- exchange associated with the downregulated adenoma (*DRA*) gene.

With vomiting, the initiating event is loss of HCl. This secretion of HCl into the stomach lumen by the parietal cell is coupled to the absorption of HCO_3^- in exchange for chloride at the basolateral membrane. When gastric acid is normally secreted, a mild increase in serum HCO_3^- spills into urine and causes an "alkaline tide." With vomiting, however, the net loss of HCl generates the alkalosis. Initially, this increased HCO_3^- is filtered by the glomeruli and excreted in urine accompanied by Na^+ and K^+; volume depletion begins to develop. As vomiting continues, extracellular volume depletion worsens, glomerular filtration falls, HCO_3^- filtration is limited, volume depletion increases the renin–angiotensin II–aldosterone system, proximal fluid and HCO_3^- reabsorption increase, distal Na^+ reabsorption increases under the influence of aldosterone, and greater H^+ secretion enhances HCO_3^- reabsorption. These effects reduce renal Na^+ loss but at the expense of maintaining the metabolic alkalosis. Significant K^+ losses, which occur as a result of the bicarbonaturia and hyperaldosteronism, lead to hypokalemia, which is actually due to renal, not gastrointestinal, losses as a consequence of attempts to maintain extracellular volume. The hypokalemia further increases proximal $NaHCO_3$ reabsorption, distal H^+ secretion, and K^+ reabsorption, all at the expense of further reabsorption of HCO_3^-. At the new steady state after vomiting or nasogastric suctioning ceases, the paradoxical aciduria of metabolic alkalosis develops as HCO_3^- reabsorption is complete, and the urine contains low levels of Na^+, K^+, and Cl^-. The patient may be hypovolemic, hypokalemic, and alkalemic, but because Na^+, K^+, and acid-base balance are intrinsically linked, life-threatening volume depletion, potassium depletion, and alkalemia are usually avoided.

Most nonrenal metabolic alkaloses with volume depletion are due to gastrointestinal losses. However, some patients with cystic fibrosis (Chapter 89) may develop hypochloremic alkalosis as a consequence of excessive sweat chloride content related to the *CFTR* gene mutation.

The sweat gland, like the principal cell in the kidney, contains the aldosterone-sensitive epithelial sodium channel, so Na^+ absorption from the glandular duct renders the lumen electronegative. When Cl^- absorption is decreased in cystic fibrosis (Chapter 89), the lumen becomes more negative, thereby decreasing Na^+, Cl^-, and fluid absorption and also leading to salty sweat; the proportionally large Cl^- loss generates a hypochloremic metabolic alkalosis.

Metabolic Alkalosis of Renal Origin with Volume Expansion and Hypertension

The renal conditions that cause metabolic alkalosis and volume expansion are due to a proportionately greater increase in Na^+ reabsorption above what is required to maintain a steady state of Na^+ balance, rather than primary loss of the Cl^- anion. As Na^+ is reabsorbed, electroneutrality is maintained by an increase in plasma HCO_3^-. The plasma Na^+ concentration may be increased, and Cl^- balance is normal; Cl^- appears in urine, and hypochloremia is not present. In the kidney, the loss of net acid as NH_4Cl in excess of the acid produced generates a metabolic alkalosis, in which the new bicarbonate generated is due to proton secretion by the distal nephron through H^+-ATPases. The H^+ then combines with NH_3 to form NH_4^+ in urine.

Na^+ is reabsorbed independently of Cl^- in the cortical collecting duct through the aldosterone-sensitive cells containing the ENaC. When Na^+ is reabsorbed by the principal cells of the cortical collecting duct, the tubule lumen becomes electronegative and stimulates both K^+ and H^+ secretion by the electrogenic H^+-ATPases. To the extent that HCO_3^- remains in the lumen, the secreted protons complete HCO_3^- reabsorption. Additional secreted protons combine with NH_3 and phosphates and lead to net acid excretion. Any increase in the distal H^+ secretory mechanism will produce more urinary net acid; more new HCO_3^- will be generated and returned to the now expanded extracellular fluid, and metabolic alkalosis will develop. The increased plasma HCO_3^- will be filtered, but in the absence of a stimulus to increase proximal HCO_3^- reabsorption, the HCO_3^- will flow distally to be reabsorbed by the increased H^+ secretion of the collecting duct. At first, the alkalosis is mild, but increased cortical collecting duct Na^+ reabsorption will also lead to increased K^+ secretion and hypokalemia. Hypokalemia increases the capacity for proximal HCO_3^- reabsorption, thereby opposing the effect of volume expansion, so that distal delivery of HCO_3^- decreases. The higher than normal distal H^+ secretion titrates urinary buffers, so further new HCO_3^- is formed and the alkalosis worsens. Metabolic alkaloses in the hypermineralocorticoid syndromes are sustained by hypokalemia.

Metabolic Alkalosis of Nonrenal Origin Associated with Normal or Expanded Volume

If an alkalotic patient is not hypochloremic, electroneutrality must be maintained either by depletion of an alternative anion or by an excessive concentration of a cation. An example of a metabolic alkalosis associated with depletion of a non-chloride anion is hypoproteinemic alkalosis, with hypoalbuminemia and a small anion gap. Chloride balance is normal and chloride appears in urine.

Alkalosis also may result from the addition of alkali salts of organic anions. The normal response to the ingestion of $NaHCO_3$ is rapid urinary alkalinization because of an unaltered threshold for HCO_3^- reabsorption. However, a

marked excess of HCO_3^-, as may be administered in an attempt to alkalinize a patient's urine, expands volume and causes an alkalemia, especially in the presence of volume depletion or low glomerular filtration. Milk-alkali syndrome, usually seen when patients in renal failure ingest milk or calcium antacids, is associated with hypercalcemia, alkalemia, and normal chloride concentration.

Other situations in which intake of alkali salts results in metabolic alkalosis include infusion of large quantities of sodium salts of metabolizable organic compounds, such as acetate, citrate, lactate, or bicarbonate; hyperalimentation with acetate salts; chronic peritoneal dialysis with acetate or lactate dialysate; and excessive transfusions or plasmapheresis, in which large quantities of citrate, used as an anticoagulant, are delivered.

CLINICAL MANIFESTATIONS

Mild metabolic alkalosis up to a pH of 7.50 is usually asymptomatic. When the pH exceeds 7.55, however, the alkalosis itself and the compensatory hypoventilation are frequently associated with metabolic encephalopathy. Symptoms include confusion, obtundation, delirium, and coma. The seizure threshold is lowered; tetany, paresthesias, muscle cramping, and other symptoms of low calcium are seen. In patients with hypocalcemia, these signs may be seen at pH values above 7.45. Other findings include cardiac tachyarrhythmias and hypotension. Lactate production increases as a result of the increased anaerobic glycolysis.

DIAGNOSIS

In diagnosis of the cause of metabolic alkalosis, it is important to distinguish whether the condition is chloride responsive or chloride unresponsive. Metabolic alkalosis is generally divided into two categories on the basis of its responsiveness to chloride (see Table 118-2). Chloride-responsive metabolic alkalosis is associated with extracellular fluid and chloride depletion and is seen in cases of gastric fluid loss and diuretic use. A diagnostic clue comes from the serum electrolytes. HCO_3^- is increased with a corresponding fall in serum chloride (hypochloremic alkalosis). Chloride-unresponsive metabolic alkalosis is seen in patients with extracellular fluid expansion in conditions such as primary aldosteronism and hypokalemia. Entry of hydrogen ions into cells can also lead to metabolic alkalosis in patients with hypokalemia.

Vomiting, nasogastric suction, and diarrhea are usually obvious sources of metabolic alkalosis. However, the Zollinger-Ellison syndrome (Chapter 195), villous adenomas (Chapter 193), and VIPomas (Chapter 195) may be more difficult to diagnose unless the index of suspicion is high.

Patients who present with hypokalemic metabolic alkalosis with normal or low blood pressure and have urinary chloride concentrations above 25 mEq/L may be taking diuretics such as furosemide or thiazides surreptitiously; a diuretic screen can document the presence of the drug. If the screen is negative, Bartter or Gitelman syndrome (Chapter 128) should be considered. Bartter syndrome is less common, is usually more severe, and presents in young patients. The presence of hypercalciuria favors Bartter syndrome, whereas hypocalciuria and hypomagnesemia suggest Gitelman syndrome.

Specific causes of renal alkalosis with volume expansion and hypertension can be classified according to levels of renin and aldosterone. Primary increases in renin with secondary increases in aldosterone can be seen in patients with unilateral renal artery stenosis (Chapter 125), renin-secreting tumors of the kidney (Chapter 67), and malignant hypertension (Chapter 67). Low renin and elevated aldosterone levels are characteristic of primary hyperaldosteronism from adrenal adenoma or hyperplasia (Chapter 227). A high cortisol level with volume expansion is seen in hypercortisolism and adrenocorticotropic hormone–secreting tumors (Chapter 227). Inhibition of the intracellular enzyme 11β-hydroxysteroid dehydrogenase Type 2, which normally inactivates cortisol to form cortisone in the principal cell, will also result in low renin levels, low aldosterone levels, and hypokalemic alkalosis. Both genetic mutations (the apparent mineralocorticoid excess syndrome) and an excess consumption of glycyrrhizic acid found in licorice and anisette are causes of this enzyme block. Another cause of hypertension with hypokalemic alkalosis but with low renin and aldosterone levels is Liddle syndrome (Chapter 128), in which an activating mutation in the cortical collecting duct sodium channel (ENaC) leads to increased Na^+ reabsorption.

Metabolic alkalosis may also develop without volume expansion when a non-reabsorbable anion is presented to the cortical collecting duct lumen. Nitrates, sulfates, and certain antibiotics such as nafcillin, carbenicillin, and ticarcillin may obligate K^+ and H^+ secretion as Na^+ is reabsorbed. Topical administration of silver nitrate to burn victims may result in alkalosis.

TREATMENT

Rx

In chloride-responsive patients (see Table 118-2), treatment is directed at increasing urinary excretion of bicarbonate. In patients with mild to moderate alkalosis, liberalizing salt intake and administering potassium chloride is effective in increasing renal HCO_3^- excretion. The K^+ deficit is likely to be at least 100 mEq for every decrease of 1 mEq/L in serum potassium. Unless potassium chloride is also replenished, the improvement in filtration and proximal reabsorption will result in severe potassium wasting as bicarbonaturia develops and aldosterone's effects remain. In addition, complete resolution of alkalosis will not occur until K^+ is normalized. In a patient with renal failure and vomiting, the elevation in HCO_3^- may be more severe because of poor HCO_3^- filtration. In cases of volume expansion and alkalosis, acetazolamide may be administered carefully while monitoring its potential for losing K^+. If this agent fails to work, dilute solutions of HCl (0.1N HCl) may be cautiously administered. The amount of H^+, in milliequivalents, to be given may be calculated as the product of the desired change in serum HCO_3^- concentration (mEq/L) times 0.5 of body weight in kilograms. It is likely that this calculation will overestimate the amount of acid needed for correction, so no more than one third of the amount should be given before recalculating to avoid metabolic acidosis. Full correction of HCO_3^- should not be the goal. In the absence of renal failure, intravenous acetazolamide (250 to 500 mg every 8 hours) may be effective but may greatly increase K^+ losses.

Chloride-unresponsive patients (see Table 118-2) include those with mineralocorticoid excess. In these patients, the metabolic alkalosis can be lessened by potassium replacement or by blocking Na^+ reabsorption with aldosterone antagonists such as spironolactone, starting with 25 mg orally, or amiloride, beginning with 5 mg orally. Indomethacin effectively treats Bartter syndrome (Chapter 128) by interfering with prostaglandin E_2 to allow greater NaCl reabsorption in the thick ascending limb. Gitelman and Bartter syndromes are best treated with combinations of potassium chloride, a potassium-sparing diuretic, and magnesium if needed.

RESPIRATORY ACIDOSIS

Respiratory acidosis is characterized by a primary elevation in P_{CO_2} as reflected by reduced arterial pH with variable elevation in the HCO_3^- concentration. It is most frequently caused by a decrease in alveolar ventilation due to pulmonary disease (Chapter 104), respiratory muscle fatigue, musculoskeletal abnormalities of the chest wall, or abnormalities in ventilatory control (Chapter 86).

CLINICAL MANIFESTATIONS

Clinical findings in respiratory acidosis are related to the degree and duration of the respiratory acidosis and whether hypoxemia is present. A precipitous rise in P_{CO_2} can lead to confusion, anxiety, psychosis, asterixis, seizures, and myoclonic jerks, with progressive depression of the sensorium and coma at an arterial P_{CO_2} greater than 60 mm Hg (CO_2 narcosis). Hypercapnia, which increases cerebral blood flow and volume, can lead to symptoms and signs of elevated intracranial pressure, including headaches and papilledema. Other findings in acute respiratory acidosis include signs of catecholamine release, such as skin flushing, diaphoresis, and increased cardiac contractility and output. Symptoms of chronic hypercapnia include fatigue, lethargy, and confusion, in addition to the findings seen in acute hypercapnia.

The slow time course of many of these diseases allows the kidney to compensate adequately as the disease progresses by increasing its excretion of hydrogen ion as ammonium and generating and reabsorbing bicarbonate to restore systemic pH toward normal values. This compensatory process is not maximal until 3 to 5 days after the onset of respiratory acidosis. Chronic $NaHCO_3$ retention and edema often accompany chronic respiratory acidosis.

DIAGNOSIS

Disorders that cause a respiratory acidosis include central effects of drugs, stroke, and infection; airway obstruction; primary parenchymal processes, such as chronic obstructive pulmonary disease (Chapter 88) and acute respiratory distress syndrome (Chapter 104); disorders of ventilation (Chapter 86); and neuromuscular diseases, such as myasthenia gravis (Chapter 422) and muscular dystrophies (Chapter 421). Permissive hypercapnia has been used clinically in patients with acute respiratory distress syndrome to limit pulmonary damage secondary to mechanical ventilation (Chapter 105).

TREATMENT

Rx

Treatment of both chronic and acute respiratory acidosis aims primarily to correct the underlying cause and to ensure adequate ventilation. In acute respiratory acidosis, measures to relieve severe hypoxemia and acidemia should be instituted immediately, including intubation and assisted mechanical ventilation (Chapter 105) if necessary. Patients with myxedema coma require thyroid replacement (Chapter 226).

In patients with compensated chronic respiratory acidosis, rapid and complete correction of hypercapnia can result in post-hypercapnic metabolic alkalosis. Patients who are chronically hypercapnic and hypoxemic should receive necessary oxygen even though their P_{CO_2} will rise. The rise is not necessarily due to loss of a hypoxic drive to ventilation but may instead be because of release of CO_2 from hemoglobin in the presence of oxygen or because oxygen-induced pulmonary arteriolar vasodilation increases perfusion to poorly ventilated alveoli. Patients recovering from an acute-on-chronic respiratory acidosis should be monitored carefully to correct hypokalemia, hypochloremia, and hypovolemia so that adequate renal excretion of bicarbonate can occur.

Bicarbonate therapy is not indicated for respiratory acidosis unless the pH falls below 7.0 and the patient is about to be intubated. There is a role for bicarbonate therapy in patients with renal failure (Chapter 130), in whom adequate compensatory acid excretion cannot take place.

RESPIRATORY ALKALOSIS

EPIDEMIOLOGY AND PATHOBIOLOGY

In respiratory alkalosis, a primary decrease in P_{CO_2} is reflected by increases in arterial pH and variable decreases in plasma bicarbonate concentration. The most common cause is alveolar hyperventilation, not underproduction of CO_2.

Acute hypocapnia results in an initial increase in the pH of both the CSF and the brain's intracellular environment. However, this increase is quickly offset by a decrease in bicarbonate levels. In acute respiratory alkalosis, one of the primary mechanisms of this fall in bicarbonate appears to be the generation of lactate as a result of vasoconstriction, hypoxia, and increased hemoglobin affinity for oxygen. The combination of increased oxygen demand and decreased oxygen delivery may contribute to adverse clinical outcomes in hypocapnic alkalosis.

Cerebral blood flow is significantly decreased by hypocapnia, which is a potent vasoconstrictor. As in respiratory acidosis, the CNS is immediately affected by decreases in systemic P_{CO_2} because of the blood-brain barrier's permeability to CO_2. In addition, as in respiratory acidosis, CSF and intracellular pH show an initial short-lived response that parallels the systemic increase in pH.

Renal compensation for sustained hypocapnia is complete in 36 to 72 hours. The mechanism rests primarily in the kidney's net reduction of hydrogen ion excretion, which it accomplishes largely by decreasing ammonium and titratable acid excretion. The threshold for bicarbonate excretion is also lowered, and bicarbonaturia develops. As a result, systemic bicarbonate levels decrease, and arterial pH returns toward normal values.

Acute exposure to high altitude (Chapter 94) results in hypoxia-induced hyperventilation. Compensation requires at least several days and is characterized by a gradual further increase in hyperventilation, a steadily decreasing P_{CO_2}, and a recovering P_{O_2}. The effect of the hypoxic stimulus to ventilate is initially modulated by the effects of alkalosis, both peripherally and centrally. However, as HCO_3^- falls in the CSF, inhibition of the central stimulus to ventilate decreases. Once a steady state is achieved, the drive to ventilate is determined by the effects of hypoxemia and alkalemia on the peripheral chemoreceptors.

CLINICAL MANIFESTATIONS

The clinical manifestations of respiratory alkalosis depend on the degree and duration of the condition but are primarily those of the underlying disorder. Chronic hypocapnia does not appear to be associated with any significant clinical symptoms.

Symptoms of acute hypocapnia are largely attributable to the alkalemia and include dizziness, perioral or extremity paresthesias, confusion, asterixis, hypotension, seizures, and coma. Most symptoms, which are manifested only when the P_{CO_2} falls below 25 or 30 mm Hg, can be related to decreased cerebral blood flow or reduced free calcium because alkalosis increases calcium's protein-bound fraction. Shortness of breath and chest wall pain, which frequently may be seen when patients hyperventilate because of pain or anxiety, do not appear to be related to hypocapnia.

DIAGNOSIS

Alveolar hyperventilation leading to respiratory alkalosis is seen with hypoxemia from pulmonary disease (Chapter 83), heart failure (Chapter 58), high altitudes (Chapter 94), or anemia. Mechanical ventilation (Chapter 105) is also a common cause of respiratory alkalosis.

Another common cause of respiratory alkalosis is primary stimulation of the central chemoreceptor, as seen in sepsis (Chapter 108), hepatic cirrhosis (Chapter 153), salicylate intoxication (Chapters 37 and 110), correction of metabolic acidosis, hyperthermia (Chapter 109), and pregnancy, as well as cortical hyperventilation from anxiety and pain. In these situations, central signals override peripheral chemoreceptors until the primary stimulus is removed.

Primary neurologic diseases that can stimulate alveolar hyperventilation include acute stroke, infection, trauma, and tumors. Two patterns of respiration are seen: central hyperventilation and Cheyne-Stokes respiration (Chapter 86). Central hyperventilation, which is associated with lesions at the pontine-midbrain level, is regular, but with an increased rate and tidal volume. Cheyne-Stokes breathing, which is characterized by periods of hyperventilation alternating with apnea, is seen in patients with bilateral cortical and upper pontine lesions and in patients with heart failure.

TREATMENT

Rx

Treatment of respiratory alkalosis must address the underlying cause of the disturbance. Hyperventilation syndrome is a diagnosis of exclusion, but patients who exhibit symptoms, such as tetany and syncope, and who do not have more serious causes of hyperventilation can be treated with a rebreathing mask. Hypophosphatemia can be seen in these patients, but it usually improves with treatment of the alkalosis. Patients with respiratory alkalosis associated with mountain sickness can be pretreated with acetazolamide to induce a metabolic acidosis, thereby preventing extreme elevations in pH (Chapter 94).

Grade A References

A1. Susantitaphong P, Sewaralthahab K, Balk EM, et al. Short- and long-term effects of alkali therapy in chronic kidney disease: a systematic review. *Am J Nephrol.* 2012;35:540-547.

A2. de Brito-Ashurst I, Varagunam M, Raftery MJ, et al. Bicarbonate supplementation slows progression of CKD and improves nutritional status. *J Am Soc Nephrol.* 2009;20:2075-2084.

A3. Jones AE, Shapiro NI, Trzeciak S, et al. Lactate clearance vs central venous oxygen saturation as goals of early sepsis therapy: a randomized clinical trial. *JAMA.* 2010;303:739-746.

A4. Cooper DJ, Walley KR, Wiggs BR, et al. Bicarbonate does not improve hemodynamics in critically ill patients who have lactic acidosis. A prospective, controlled clinical study. *Ann Intern Med.* 1990;112:492-498.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Brown D, Wagner CA. Molecular mechanisms of acid-base sensing by the kidney. *J Am Soc Nephrol*. 2012;23:774-780.
2. Berend K, de Vries AP, Gans RO. Physiological approach to assessment of acid-base disturbances. *N Engl J Med*. 2014;371:1434-1445.
3. Seifter JL. Integration of acid-base and electrolyte disorders. *N Engl J Med*. 2014;371:1821-1831.
4. Kraut JA, Nagami GT. The serum anion gap in the evaluation of acid-base disorders: what are its limitations and can its effectiveness be improved? *Clin J Am Soc Nephrol*. 2013;8:2018-2024.
5. Rice M, Ismail B, Pillow MT. Approach to metabolic acidosis in the emergency department. *Emerg Med Clin North Am*. 2014;32:403-420.
6. Mehta AN, Emmett JB, Emmett M. GOLD MARK: an anion gap mnemonic for the 21st century. *Lancet*. 2008;372:892.
7. Dobre M, Yang W, Chen J, et al. Association of serum bicarbonate with risk of renal and cardiovascular outcomes in CKD: a report from the Chronic Renal Insufficiency Cohort (CRIC) study. *Am J Kidney Dis*. 2013;62:670-678.
8. Yaqoob MM. Treatment of acidosis in CKD. *Clin J Am Soc Nephrol*. 2013;8:342-343.
9. Andersen LW, Mackenhauer J, Roberts JC, et al. Etiology and therapeutic approach to elevated lactate levels. *Mayo Clin Proc*. 2013;88:1127-1140.
10. Randall L, Begovic J, Hudson M, et al. Recurrent diabetic ketoacidosis in inner-city minority patients: behavioral, socioeconomic, and psychosocial factors. *Diabetes Care*. 2011;34:1891-1896.
11. Liss DB, Paden MS, Schwarz ES, et al. What is the clinical significance of 5-oxoproline (pyroglutamic acid) in high anion gap metabolic acidosis following paracetamol (acetaminophen) exposure? *Clin Toxicol (Phila)*. 2013;51:817-827.
12. Wiener SW. Toxicologic acid-base disorders. *Emerg Med Clin North Am*. 2014;32:149-165.
13. Gil-Pena H, Mejia N, Santos F. Renal tubular acidosis. *J Pediatr*. 2014;164:691-698.

REVIEW QUESTIONS

1. A patient with type 1 diabetes mellitus, on hemodialysis, underwent a deceased donor combined renal-pancreas organ transplantation. The pancreatic duct was drained through the recipient's urinary bladder. The function of the transplanted kidney and pancreas was excellent. Which of the following compensations would occur for the post-transplantation acid-base disorder?

A. Respiratory hypoventilation
B. Respiratory hyperventilation and increased renal ammoniagenesis
C. Metabolic alkalosis due to the severe urinary potassium wasting
D. Renal chloride losses to match pancreatic bicarbonate loss
E. Initial HCO_3^- losses followed by paradoxical aciduria

Answer: B The patient would lose pancreatic $NaHCO_3$ in the bladder, thereby causing a severe hyperchloremic acidosis, which would be compensated by hyperventilation and increased renal attempts to excrete the acid load by increasing ammonia production in the newly functional kidney.

2. A 36-year-old woman returns from an overseas trip with profuse watery diarrhea and is found to have a urinary Na^+ of 34 mEq/L, a urinary K^+ of 41 mEq/L, and a urinary Cl^- of 6 mEq/L. Which one of the following statements is most likely to be correct?

A. She has severe metabolic acidosis from the diarrhea.
B. She has both metabolic acidosis from diarrhea and bicarbonaturia from an unsuspected renal tubular acidosis.
C. She is acidemic from the diarrhea but must also have been vomiting.
D. She has not been vomiting but has metabolic alkalosis and alkalemia.
E. Her urinary NH_4^+ must be very high.

Answer: D She has chloride-wasting secretory diarrhea, which has caused hypochloremic metabolic alkalosis. There is bicarbonaturia from the net gain of plasma bicarbonate, and that is why her urinary Na^+ and K^+ are high. She is conserving chloride. If she was acidemic with a mixed hypochloremic alkalosis from vomiting and an acidosis from diarrhea, the net Cl^- in the blood would be high and her urinary anion gap would be less than 0 ($Na^+ + K^+ - Cl^- < 0$). There is no evidence that urinary NH_4^+ is high, and the positive urinary anion gap is normally seen in hypochloremic alkalosis. It does not imply a renal tubular acidosis. The teaching point is that not all diarrhea causes metabolic acidosis, and diarrhea may cause metabolic alkalosis if the stool has high Cl^-.

119

DISORDERS OF MAGNESIUM AND PHOSPHORUS

ALAN S. L. YU

MAGNESIUM METABOLISM

Magnesium is an important mineral component of the bony skeleton, a cofactor for many metabolic enzymes, and a regulator of ion channels and transporters in excitable tissues.

Normal Magnesium Metabolism

The majority of total body magnesium is intracellular or in bone, with only 1% in extracellular fluid. The normal serum magnesium concentration is 1.8 to 2.3 mg/dL (1.5 to 1.9 mEq/L). The average daily intake of magnesium is 300 mg, the main sources of which are green vegetables, nuts, whole grain cereals, milk, and seafood. Magnesium is absorbed mainly in the jejunum and ileum. In the kidney, 70 to 80% of serum magnesium is filtered at the glomerulus, with the majority being reabsorbed along the length of the tubule, particularly in the thick ascending limb of Henle. In states of magnesium deficiency or excess, renal tubule reabsorption is tightly regulated so that magnesium excretion is adjusted accordingly.

MAGNESIUM DEFICIENCY

PATHOBIOLOGY

Magnesium deficiency is usually detected when hypomagnesemia becomes evident. However, because magnesium is stored primarily intracellularly, substantial depletion of total body magnesium can occur before serum magnesium levels drop appreciably.

Magnesium deficiency may be due to nutritional deficiency, intestinal malabsorption, redistribution into bone, or losses via cutaneous, lower gastrointestinal, or renal routes (Table 119-1). The recommended daily allowance of magnesium is 420 mg for males and 320 mg for females. Approximately 25% of alcoholics are chronically hypomagnesemic because of a combination of poor nutritional intake and increased renal loss. Magnesium deficiency can occur, rarely, in protein-calorie malnutrition and may be associated with acute hypomagnesemia during refeeding because of rapid cellular magnesium uptake. Fat malabsorption in conditions such as celiac disease, Crohn disease, and small intestinal resection causes magnesium deficiency because free fatty acids accumulate in the intestinal lumen, where they combine with magnesium to form insoluble soaps. Proton pump inhibitors also can cause hypomagnesemia, primarily in patients concurrently using diuretics.[1] This is thought to be due to inhibition of intestinal absorption. Lower gastrointestinal tract secretions are rich in magnesium, so diarrhea of colonic origin is a common cause of hypomagnesemia. Sweat contains significant amounts of magnesium, and transient hypomagnesemia can occur after prolonged, intense exercise such as marathon runs. Magnesium is also lost from burned skin surfaces, and 40% of patients with severe burns (Chapter 111) are hypomagnesemic.

In patients with severe hyperparathyroidism (Chapter 245) and high bone turnover, continued sequestration of minerals within bone may continue for several days after parathyroidectomy and cause transient hypocalcemia, hypomagnesemia, and hypophosphatemia. Renal magnesium losses can occur in any polyuric state, including the recovery phase of acute tubular necrosis or urinary tract obstruction. Hypomagnesemia is common in diabetes mellitus (Chapter 229), in which it is thought to be due to a combination of poor intestinal absorption owing to autonomic neuropathy, osmotic diuresis, and decreased renal tubule reabsorption. Failure of sodium reabsorption in the thick ascending limb of Henle as a result of the use of loop diuretics and in the distal convoluted tubule as a result of thiazide diuretics inhibits tubular magnesium reabsorption and leads to urinary magnesium wasting. Drugs that are tubular toxins are also common causes of renal magnesium wasting. Such drugs include cisplatin, carboplatin, amphotericin B, and aminoglycosides, which are commonly associated with hypokalemia and rarely with renal tubule acidosis, as well as calcineurin inhibitors such as cyclosporine and tacrolimus, which also cause hyperkalemia. Antibodies to the epidermal growth factor receptor, such as cetuximab and panitumumab, which are used to treat metastatic colorectal cancer, downregulate a distal tubule magnesium channel and are an increasingly common cause of isolated severe hypomagnesemia.[2]

Inherited hypomagnesemia is usually caused by renal magnesium loss and can be subdivided into three main types, depending on the coexistence of other electrolyte disturbances: Bartter and Gitelman syndromes, which are associated with renal salt wasting and hypokalemic metabolic alkalosis; familial hypomagnesemia with hypercalciuria and nephrocalcinosis; and isolated hypomagnesemia, which is usually associated with hypocalcemia.[3]

TABLE 119-1 CAUSES OF MAGNESIUM DEFICIENCY

Nutritional deficiency
Alcoholism*
Malnutrition
Refeeding syndrome
Intestinal malabsorption*
Proton pump inhibitors
Lower gastrointestinal losses
Colonic diarrhea*
Intestinal fistula
Laxative abuse
Cutaneous losses
Burns*
Exercise-induced sweating
Redistribution into bone
Hungry bone syndrome
Renal losses
Polyuria (including diabetes mellitus)*
Volume expansion
Hyperaldosteronism
Bartter and Gitelman syndromes
Hypercalcemia
Loop and thiazide diuretics*
Nephrotoxins (cisplatin, amphotericin, aminoglycosides, pentamidine, cyclosporine)*
Epidermal growth factor monoclonal antibodies (cetuximab, panitumumab)*

*Common causes.

CLINICAL MANIFESTATIONS

Mild-to-moderate hypomagnesemia or magnesium deficiency is frequently asymptomatic. Manifestations of increased neuronal excitability are the most common symptoms, including paresthesias, tetany, and seizures. These may be associated with Chvostek's sign (twitching of the cheek muscles in response to tapping the facial nerve in front of the ear) or Trousseau sign (carpal spasm induced by compressing the upper arm with a tourniquet or blood pressure cuff). Cardiac disturbances also may occur and range in severity from mild electrocardiographic abnormalities (nonspecific T wave changes, U waves, prolonged QT interval, and repolarization alternans) to ventricular tachycardia, torsades de pointes, and ventricular fibrillation (Chapter 65).

Coexistent hypokalemia is very common, for two reasons: many of the causes of hypomagnesemia are also causes of potassium loss, and hypomagnesemia itself causes renal potassium wasting. Severe hypomagnesemia also impairs parathyroid hormone secretion and induces tissue resistance to its actions, thereby leading to hypocalcemia.

DIAGNOSIS

The cause of the magnesium deficiency is often obvious from the history. In difficult diagnostic cases, a random urine sample should be collected and the fractional excretion of magnesium (FE_{Mg}) determined.

$$FE_{Mg} = \frac{\text{Urine magnesium} \times \text{Serum creatinine}}{0.7 \times \text{Serum magnesium} \times \text{Urine creatinine}}$$

With extrarenal magnesium loss (usually malabsorption or laxative abuse), the FE_{Mg} is appropriately suppressed (<4%). Higher FE_{Mg} levels indicate renal magnesium wasting, often secondary to surreptitious diuretic use or one of the familial magnesium-wasting disorders.

TREATMENT

Rx

It is unclear whether mild, asymptomatic hypomagnesemia needs to be treated. Magnesium repletion is recommended in hypomagnesemic patients if they are symptomatic, have underlying cardiac or seizure disorders, exhibit concurrent severe hypocalcemia or hypokalemia, or have severe hypomagnesemia (<1.4 mg/dL). In mild cases or in the outpatient setting, oral magnesium salts such as magnesium oxide (250 to 500 mg four times daily) can be used for repletion, but these substances frequently cause diarrhea, particularly at high doses. In the inpatient setting, intravenous magnesium sulfate (1 to 2 g every 6 hours) can be used for repletion. Because the redistribution of magnesium from extracellular to intracellular compartments is relatively slow, the serum magnesium concentration may normalize before total body magnesium stores are replete. It is therefore prudent to continue intravenous magnesium for an additional 1 to 2 days after restoration of normomagnesemia. In patients with normal renal function, any excess magnesium is simply excreted renally. Adverse effects from intravenous magnesium administration are primarily due to transient hypermagnesemia and include flushing, hypotension, and flaccid paralysis. Amiloride (10 to 20 mg PO once daily) abrogates renal magnesium wasting in some patients with this problem, but the mechanism is unknown.

PROGNOSIS

Hospitalized patients with hypomagnesemia have a longer length-of-stay and higher mortality, presumably because it is a marker of more severe illness.[4] In patients with a self-limited cause of magnesium deficiency, repletion is easily accomplished. However, in patients with persistent magnesium wasting, such as in Gitelman syndrome (Chapter 128), it can be difficult to keep up with the ongoing losses with oral therapy. Fortunately, these individuals tend to adapt to their chronic hypomagnesemia and tolerate it fairly well

HYPERMAGNESEMIA

Transient hypermagnesemia can occur in patients given large doses of intravenous magnesium, for example, in the setting of preeclampsia. It has also been reported in individuals taking large doses of magnesium-containing antacids or cathartics, particularly in settings in which intestinal absorption is enhanced, such as inflammatory bowel disease and intestinal obstruction. However, the kidney has a very large capacity to excrete excess magnesium. Thus, persistent hypermagnesemia is seen almost exclusively in patients who have chronic renal insufficiency (Chapter 130) who are also taking excessive amounts of magnesium in the form of antacids, cathartics, or enemas.

CLINICAL MANIFESTATIONS

Magnesium toxicity is a serious and potentially fatal condition. Mild hypermagnesemia (serum magnesium level > 4 to 6 mg/dL) causes hypotension, nausea, vomiting, facial flushing, urinary retention, and ileus. Above serum levels of 8 to 12 mg/dL, flaccid skeletal muscle paralysis and hyporeflexia may ensue, along with bradyarrhythmias, respiratory depression, coma, and cardiac arrest. A low or even negative serum anion gap is sometimes seen.

TREATMENT Rx

Mild hypermagnesemia in a patient with good renal function usually requires no treatment because renal clearance is rapid and the normal serum half-life of magnesium is approximately 1 day. In the event of serious toxicity, the effects of magnesium can be temporarily antagonized by the administration of intravenous calcium salts (5 to 10 mL of 10% calcium chloride). Renal magnesium excretion can be enhanced by administering furosemide (20 to 40 mg every 4 hours) together with a saline infusion (0.9% NaCl at 150 mL/hour, titrated to replace urinary losses). In patients with advanced renal insufficiency, the most effective method of magnesium removal is hemodialysis.

PROGNOSIS

Severe hypermagnesemia is potentially fatal. Lesser degrees of hypermagnesemia usually respond well to treatment.

PHOSPHORUS METABOLISM

Phosphorus has many critical roles. It is a major component of bone mineral, of phospholipids in cell membranes, and of nucleic acids. It forms high-energy phosphate bonds in compounds such as adenosine triphosphate (ATP), is post-translationally bound to proteins as an intracellular signal, and acts as a major pH buffer in serum and urine.

Normal Phosphorus Metabolism

Of the total body phosphorus content, 85% is in bone, 14% is in intracellular compartments, and only 1% is in extracellular fluid. The normal concentration of phosphorus in plasma is 3 to 4.5 mg/dL (1 to 1.5 mM). Daily intake of phosphorus is 800 to 1500 mg. Phosphorus is present in many foods, including dairy products, meats, and grains, and it is absorbed in the small intestine. The kidneys excrete excess phosphorus, which is the principal mechanism by which the body regulates extracellular phosphate balance. Ninety percent of serum phosphate is filtered at the glomerulus, of which 80 to 97% is reabsorbed along the nephron, primarily in the proximal tubule. Parathyroid hormone increases renal phosphate excretion by inhibiting the sodium-phosphate cotransporter in the proximal tubule, whereas vitamin D enhances intestinal phosphate absorption.

HYPOPHOSPHATEMIA

PATHOBIOLOGY

Hypophosphatemia may be caused by decreased intake, impaired intestinal absorption, redistribution into cells or bone, and renal losses (Table 119-2). Phosphate is frequently depleted in alcoholism (Chapter 33) because of the intake of a carbohydrate-rich, phosphate-poor diet, as well as renal phosphate wasting. Divalent cation-containing antacids bind phosphate in the intestinal lumen to form insoluble salts, thereby preventing their absorption. Vitamin D deficiency also leads to decreased intestinal phosphate absorption and hence to hypophosphatemia. Respiratory but not metabolic alkalosis (Chapter 118) may cause transient hypophosphatemia. In this disorder, intracellular pH is increased, thereby stimulating glycolysis, which depletes the intracellular inorganic phosphate pool and leads to a shift of phosphate into cells.

TABLE 119-2 CAUSES OF HYPOPHOSPHATEMIA

Nutritional deficiency
Alcoholism*
Impaired intestinal absorption
Antacids
Vitamin D deficiency*
Redistribution into cells
Respiratory alkalosis*
Insulin*
Refeeding syndrome
Burns*
Redistribution into bone
Hungry bone syndrome
Tyrosine kinase inhibitors (imatinib, sorafenib)
Renal losses
Hyperparathyroidism*
Renal tubulopathy
Fanconi syndrome
Drugs (pentamidine, foscarnet, acetazolamide)
Phosphatonin excess syndrome
Oncogenic osteomalacia
Familial hypophosphatemic rickets

*Common causes.

Insulin is also a strong stimulus for shifting phosphate into cells. Patients with diabetic ketoacidosis (Chapter 229) are often hyperphosphatemic because of a shift of phosphate out of cells under insulinopenic conditions, but their total body phosphate is actually depleted as a result of urinary losses. Subsequent treatment with insulin may uncover severe hypophosphatemia. Similarly, in malnourished patients (Chapter 215), whose total body phosphate stores may be depleted, overzealous intravenous refeeding with carbohydrate-rich fluids may stimulate insulin release and cause acute hypophosphatemia. The tyrosine kinase inhibitors imatinib, sorafenib, and nilotinib, which are used in the treatment of various cancers (Chapter 184), can cause profound hypophosphatemia, which appears to be due either to inhibition of bone resorption or a partial Fanconi syndrome.[5]

Renal phosphate wasting is usually due to impaired proximal tubule phosphate reabsorption. In primary hyperparathyroidism (Chapter 245), hypercalcemia is typically associated with hypophosphatemia. Fanconi syndrome is a generalized proximal tubule disorder characterized by hypophosphatemia in association with glycosuria, aminoaciduria, hypokalemia, and type II renal tubular acidosis; it can be caused by a variety of inherited metabolic disorders, multiple myeloma (Chapter 187), heavy metal intoxication (Chapter 22), and drugs such as ifosfamide, cidofovir, and tenofovir.[6] Phosphaturia also can occur with diuretics, particularly carbonic anhydrase inhibitors, and with antimicrobial agents such as pentamidine and foscarnet.

Oncogenic osteomalacia is a paraneoplastic syndrome (Chapter 179) associated primarily with mesenchymal tumors that secrete a variety of phosphaturic factors collectively known as phosphatonins. A similar phenotype is found in X-linked and autosomal dominant hypophosphatemic rickets; these inherited disorders are characterized by an increase in a circulating phosphatonin called *fibroblast growth factor-23*.[7] Phosphatonins inhibit both renal tubular phosphate reabsorption and 1α-hydroxylation of 25-hydroxycholecalciferol, thereby leading to hypophosphatemia, rickets, or osteomalacia (Chapter 244) and inappropriately low serum levels of 1,25-dihydroxycholecalciferol.

CLINICAL MANIFESTATIONS

Clinical complications, which are usually observed only with severe hypophosphatemia (<1 mg/dL), are thought to be due to the disruption of cell membrane composition, depletion of ATP (which particularly affects high energy–consuming tissues such as skeletal and cardiac muscle), and depletion of 2,3-diphosphoglycerate in erythrocytes, with impaired tissue oxygen delivery. Manifestations of severe hypophosphatemia include encephalopathy, dilated cardiomyopathy, generalized muscle weakness that can lead to

respiratory failure, rhabdomyolysis, and hemolysis. Hypophosphatemia also impairs renal ammoniagenesis and reduces the availability of urinary buffer, thereby impairing renal acid excretion and causing metabolic acidosis. Chronic hypophosphatemia leads to resorption of bone and osteomalacia.

DIAGNOSIS

The cause of hypophosphatemia is often evident from the history and physical examination. If not, measurement of either 24-hour urinary phosphate excretion or fractional excretion of phosphate (FE_{PO4}) in a spot urine sample is often helpful.

$$FE_{PO4} = \frac{\text{Urine phosphate} \times \text{Serum creatinine}}{\text{Serum phosphate} \times \text{Urine creatinine}}$$

In the setting of hypophosphatemia, the normal response of the kidney is to reduce urinary phosphate excretion to less than 100 mg/day or to reduce FE_{PO4} to less than 5%. Higher values suggest one of the causes of renal phosphate wasting.

TREATMENT

Rx

Patients with asymptomatic mild-to-moderate hypophosphatemia, normal total body phosphorus stores, and minimal ongoing phosphorus losses (e.g., a patient with hypophosphatemia as a result of acute respiratory alkalosis) do not require treatment. Phosphate should be repleted in patients who are symptomatic, are suspected of having severely depleted intracellular phosphorus stores (malnourished or alcoholic patients), have ongoing gastrointestinal or renal losses, or have severe hypophosphatemia (<1 mg/dL).[8] Oral repletion can be accomplished with sodium or potassium phosphate salts (1 to 2 g/day) or with skimmed milk. Intravenous phosphorus repletion at a dose of 0.16 to 0.64 mmol/kg over 4 to 8 hours is recommended for severe hypophosphatemia but is contraindicated in patients with renal insufficiency or hypercalemia. Complications of phosphate therapy include hypocalcemia, metastatic calcification, hypotension, acute renal failure, and arrhythmias, as well as concomitant hypernatremia or hyperkalemia, depending on which salt is administered.

PROGNOSIS

Most patients with hypophosphatemia respond well to treatment.

HYPERPHOSPHATEMIA

PATHOBIOLOGY

Pseudohyperphosphatemia may occur in blood specimens that are hemolyzed or hyperglobulinemic, such as in multiple myeloma (Chapter 187). True hyperphosphatemia is caused by excessive phosphate intake, increased intestinal absorption, redistribution from intracellular stores, or impaired renal excretion (Table 119-3 and Fig. 119-1).[9] Overzealous phosphate repletion can obviously cause hyperphosphatemia. The phosphorus in some laxatives and enemas may be absorbed and cause hyperphosphatemia. Intoxication with vitamin D or its analogues increases intestinal absorption of both calcium and phosphorus. Conditions associated with massive cell lysis, such as rhabdomyolysis (Chapter 113) and tumor lysis syndrome (Chapter 179), cause the release of intracellular phosphate into the extracellular fluid. Patients with diabetic ketoacidosis (Chapter 229) are often hyperphosphatemic at initial evaluation because of the redistribution of phosphate out of cells in the insulin-deficient state. Decreased phosphate excretion is most commonly due to acute or chronic renal failure (Chapter 130). With a normal diet, serum phosphate levels can be maintained within the normal range until the glomerular filtration rate falls below 25 mL/minute. However, even mild degrees of renal insufficiency may predispose to hyperphosphatemia if there is a concurrent excessive intake of phosphate-containing compounds such as laxatives. Finally, because parathyroid hormone stimulates proximal tubule phosphate excretion, primary hypoparathyroidism (Chapter 245) is often associated with mild hyperphosphatemia together with hypocalcemia.

CLINICAL MANIFESTATIONS

Acute hyperphosphatemia increases the risk for precipitation of calcium phosphate and subsequent metastatic calcification in soft tissues, including the kidney, in which it can cause acute renal failure. The resultant hypocalcemia (Chapter 245) can cause tetany, hypotension, seizures, and cardiac arrhythmias. In the chronic hyperphosphatemia of chronic renal insufficiency, patients with a serum phosphate concentration greater than 6.5 mg/dL have higher mortality. Hyperphosphatemia in this setting is a risk factor for coronary and other vascular calcification, which is associated with increased mortality.

TABLE 119-3 CAUSES OF HYPERPHOSPHATEMIA

Phosphate intake
Phosphate repletion
Phosphate-containing laxatives and enemas*
Increased intestinal absorption
Vitamin D toxicity
Redistribution from intracellular stores
Rhabdomyolysis*
Tumor lysis syndrome
Diabetic ketoacidosis
Decreased renal excretion
Renal failure*
Hypoparathyroidism
Pseudohypoparathyroidism
Familial tumoral calcinosis

*Common causes.

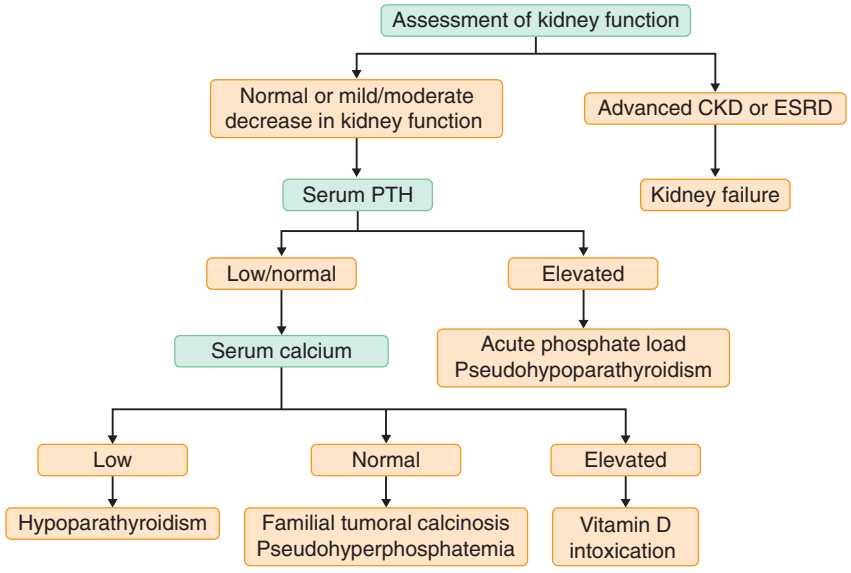

FIGURE 119-1. Diagnostic approach to chronic hyperphosphatemia, a parathyroid hormone (PTH)-based diagnostic algorithm. CKD = chronic kidney disease; ESRD = end-stage renal disease; PTH = parathyroid hormone. (Redrawn from Leaf DE, Wolf M. A physiologic-based approach to the evaluation of a patient with hyperphosphatemia. *Am J Kidney Dis.* 2013;61:330-336).

TREATMENT

Rx

Acute hyperphosphatemia in an asymptomatic patient with normal renal function often resolves spontaneously as excess phosphate is excreted. In symptomatic patients and those with impaired renal function, phosphate should be removed by extracorporeal therapy. Because of the slow rate of phosphate mobilization from intracellular stores, continuous venovenous hemodiafiltration is considerably more effective than intermittent hemodialysis. Chronic hyperphosphatemia (Chapter 130) can be managed by minimizing dietary phosphorus intake and administering oral phosphate binders such as calcium salts (e.g., calcium acetate 1334 mg with each meal), lanthanum carbonate (500 mg with each meal), or sevelamer (800 to 1600 mg with each meal) (Table 119-4).[10] Aluminum hydroxide (300 to 600 mg with meals) is also a very effective phosphate binder, but prolonged use leads to aluminum accumulation and results in encephalopathy and osteomalacia. Cinacalcet (30 to 180 mg/day), a calcimimetic used in patients with chronic renal insufficiency to treat secondary hyperparathyroidism, also reduces the serum phosphate concentration and calcium-phosphate product. However, it does not significantly reduce the risk for death or major cardiovascular events in patients undergoing dialysis.

TABLE 119-4 MEDICATIONS FOR HYPERPHOSPHATEMIA

MEDICATION	USUAL DOSE WITH EACH MEAL	COMMENTS
CALCIUM SALTS		
Calcium acetate	1334 mg	Calcium level will increase approximately 0.5 mg/dL
Calcium carbonate	500-1000 mg	Calcium level will increase approximately 0.5 mg/dL
MAGNESIUM SALTS		
Magnesium hydroxide	311-622 mg	Can cause diarrhea or hypermagnesemia
Magnesium carbonate	63-126 mg	Can cause diarrhea or hypermagnesemia
ALUMINUM SALTS		
Aluminum hydroxide	300-600 mg	Encephalopathy and osteomalacia with prolonged use
OTHERS		
Sevelamer hydrochloride	800-1600 mg	Gastrointestinal side effects
Sevelamer carbonate	800-1600 mg	Gastrointestinal side effects
Lanthanum carbonate	250-500 mg	Monitor serum bicarbonate and chloride levels as well as folic acid and vitamin D, E, and K levels

PROGNOSIS

Severe acute hyperphosphatemia can be life-threatening owing to metastatic calcification and multiorgan failure, but it generally responds well to prompt therapy. Chronic hyperphosphatemia in patients with chronic kidney failure (Chapter 130) is often fairly resistant to treatment, particularly in poorly compliant individuals, and is associated with increased long-term mortality.

Grade A Reference

A1. Chertow GM, Block GA, Correa-Rotter R, et al. Effect of cinacalcet on cardiovascular disease in patients undergoing dialysis. *N Engl J Med*. 2012;367:2482-2494.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Danziger J, William JH, Scott DJ, et al. Proton-pump inhibitor use is associated with low serum magnesium concentrations. *Kidney Int*. 2013;83:692-699.
2. Petrelli F, Borgonovo K, Cabiddu M, et al. Risk of anti-EGFR monoclonal antibody-related hypomagnesemia: systematic review and pooled analysis of randomized studies. *Expert Opin Drug Saf*. 2012;11(suppl 1):S9-S19.
3. Ferre S, Hoenderop JJ, Bindels RJ. Role of the distal convoluted tubule in renal Mg^{2+} handling: molecular lessons from inherited hypomagnesemia. *Magnes Res*. 2011;24:S101-S108.
4. Wolf F, Hilewitz A. Hypomagnesaemia in patients hospitalised in internal medicine is associated with increased mortality. *Int J Clin Pract*. 2014;68:111-116.
5. Ianotto JC, Tempescul A, Amet Y, et al. Imatinib mesylate induces massive and nonspecific aminoaciduria in CML patients. *Am J Hematol*. 2012;87:437-439.
6. Hall AM, Hendry BM, Nitsch D, et al. Tenofovir-associated kidney toxicity in HIV-infected patients: a review of the evidence. *Am J Kidney Dis*. 2011;57:773-780.
7. Gattineni J. Inherited disorders of calcium and phosphate metabolism. *Curr Opin Pediatr*. 2014;26:215-222.
8. Felsenfeld AJ, Levine BS. Approach to treatment of hypophosphatemia. *Am J Kidney Dis*. 2012;60:655-661.
9. Leaf DE, Wolf M. A physiologic-based approach to the evaluation of a patient with hyperphosphatemia. *Am J Kidney Dis*. 2013;61:330-336.
10. Bhan I. Phosphate management in chronic kidney disease. *Curr Opin Nephrol Hypertens*. 2014;23:174-179.

REVIEW QUESTIONS

1. A 55-year-old woman with ovarian cancer is being treated with carboplatin, Taxol, and pelvic irradiation. She now complains of chronic diarrhea and is found to have the following blood serum levels: magnesium 0.8 mg/dL, calcium 6.1 mg/dL, albumin 3 g/dL, and creatinine 1.3 mg/dL. Her urine magnesium is 17 mg/dL, with a urine creatinine 58 mg/dL. The most likely cause of her laboratory abnormalities is:

A. Carboplatin.
B. Diarrhea.
C. Vitamin D deficiency.
D. Radiation nephritis.
E. Primary hypoparathyroidism.

Answer: A The fractional excretion of magnesium is 68%, thereby indicating that the cause of the hypomagnesemia is severe renal magnesium wasting rather than diarrheal losses. Platinum-based chemotherapeutic agents, which are toxic to the renal tubule, are a well-recognized cause of hypomagnesemia. This toxicity, which is more common with cisplatin but also has been reported with carboplatin, can persist after discontinuation of therapy. The hypocalcemia in this case is likely secondary to the severe hypomagnesemia, which impairs parathyroid hormone secretion and also causes peripheral resistance to its actions.

2. A 46-year-old man with diabetes mellitus, chronic stage IV kidney disease, and chronic constipation has been taking magnesium citrate for 4 consecutive days. He now presents with hypotension, muscle weakness, and hyporeflexia. His serum magnesium level is 5.8 mg/dL and his serum creatinine 3.6 mg/dL. All of the following treatments would be appropriate *except:*

A. Forced diuresis with saline and furosemide.
B. Hemodialysis.
C. Calcium chloride intravenously.
D. Sodium polystyrene sulphonate.
E. Lactulose

Answer: D Magnesium ingestion in patients with renal insufficiency can cause hypermagnesemia. Hypotension, flaccid muscle paralysis, and hyporeflexia are typical manifestations. Hypermagnesemia can be treated by antagonizing the effects of magnesium with intravenous calcium salts, by increasing the excretion of magnesium with furosemide, or by hemodialysis. Alternative laxatives that are safe to use with impaired renal function include lactulose, docusate, bisacodyl, and senokot. Sodium polystyrene sulfonate is effective at binding intestinal potassium but not magnesium; if administered in this setting, it could cause hypokalemia that would exacerbate the muscle weakness.

3. An 18-year-old woman with anorexia nervosa and malnutrition was admitted for parenteral nutrition and subsequently developed hypophosphatemia with a serum phosphorus level of 0.6 mg/dL. The mechanism for hypophosphatemia is:

A. Renal phosphate wasting.
B. Fanconi syndrome.
C. Shift of phosphorus into intracellular stores.
D. Diarrheal phosphate losses.
E. Excess bone deposition.

Answer: C In malnourished patients, total body phosphate stores are depleted. Intravenous refeeding without appropriate phosphate supplementation with carbohydrate-rich fluids may stimulate insulin release and cause an acute shift of phosphate into intracellular stores.

4. A 55-year-old man with HIV infection is maintained on antiretroviral therapy with tenofovir, emtricitabine, ritonavir, and atazanavir. Routine laboratory examination revealed the following: serum sodium 133 mEq/L, potassium 2.6 mEq/L, chloride 110 mEq/L, bicarbonate 15 mEq/L, glucose 96 mg/dL, urea nitrogen 16 mg/dL, creatinine 1.8 mg/dL, phosphorus 0.7 mg/dL, and uric acid 2.6 mg/dL. Urinalysis shows 4+ glucose and trace protein. The most likely cause of his hypophosphatemia is:

A. HIV-associated nephropathy.
B. Tenofovir.
C. Acute tubular necrosis.
D. Multiple myeloma.
E. Oncogenic osteomalacia.

Answer: B This patient has hypophosphatemia, hypokalemia, a probable hyperchloremic metabolic acidosis, hypouricemia, and glycosuria without hyperglycemia. All of these findings are features of Fanconi syndrome or proximal tubulopathy. Tenofovir is an acyclic nucleotide analogue reverse-transcriptase inhibitor (as are cidofovir and adefovir) that is transported into the proximal renal tubule by organic anion transporters and causes mitochondrial toxicity. Fanconi syndrome occurs in as many as 20% of tenofovir-treated patients. Tenofovir also can cause acute tubular necrosis and acute kidney injury, but these clinical syndromes would tend to increase the serum phosphate concentration. Multiple myeloma can also cause Fanconi syndrome, but it is much less common. Oncogenic osteomalacia is a rare syndrome caused by mesenchymal tumors that secrete circulating phosphaturic factors, which cause renal phosphate wasting; however, oncogenic osteomalacia is not associated with Fanconi syndrome.

5. A 34-year-old man presents with muscle soreness and voiding dark brown urine soon after initiating a high-intensity home exercise regimen. His serum potassium is 6.5 mEq/L, creatinine 4.6 mg/dL, phosphorus 8.3 mg/dL, and creatine kinase 17,500 U/L. The most appropriate treatment for his hyperphosphatemia is:

A. Calcium acetate.
B. Calcium carbonate.
C. Insulin and glucose.
D. Sodium bicarbonate.
E. Hemodialysis.

Answer: E This patient has rhabdomyolysis, as evidenced by the very high creatine kinase value. Hyperkalemia and hyperphosphatemia resulting from release of intracellular ions is typical in this disease. Myoglobinuria is thought to be the cause of acute kidney injury. Given the degree of renal insufficiency and concomitant hyperkalemia, extracorporeal therapy with either hemodialysis or continuous venovenous hemodiafiltration would be appropriate and would effectively remove phosphate. Forced saline or alkaline diuresis is contraindicated in the setting of renal insufficiency. Calcium salts would not be very effective because they would bind only dietary phosphate in the intestinal lumen; moreover, some of the calcium would be absorbed and increase the likelihood of metastatic precipitation of calcium-phosphate crystals. Insulin would shift some phosphorus into cells, but this would not have a large or lasting effect. Sodium bicarbonate can be used in an attempt to alkalinize the urine, which is believed to decrease the renal toxicity of myoglobin, but it would not affect the phosphate level.

120 ACUTE KIDNEY INJURY

BRUCE A. MOLITORIS

DEFINITION

Acute kidney injury (AKI) is a clinical syndrome defined as a functional or structural kidney abnormality that manifests with an increase in serum creatinine (Cr) of 0.3 mg/dL or greater within 48 hours, an increase in serum Cr of 1.5 or greater times baseline within 7 days or a urine volume less than 0.5 mL/kg/hour for 6 hours (Table 120-1).[1] Diagnostically, the reduction in kidney function in AKI is staged according to the maximal rise in serum Cr or reduction in urine output with oliguria. The use of a 50% change in serum Cr over baseline should not be used in patients with a very low baseline volume.[2]

EPIDEMIOLOGY

Most episodes of AKI occur in the hospital, with an incidence of 20% among all hospitalized patients[3] and up to 50% among patients in intensive care units. AKI is the number one reason for hospital nephrology consult. By contrast, the incidence of community-acquired AKI is no more than 1%.

The various causes of AKI are divided broadly into three anatomic categories: prerenal, intrarenal or intrinsic, and postrenal (Fig. 120-1). Each of the categories represents a unique pathophysiologic process with distinctive diagnostic parameters and prognosis.

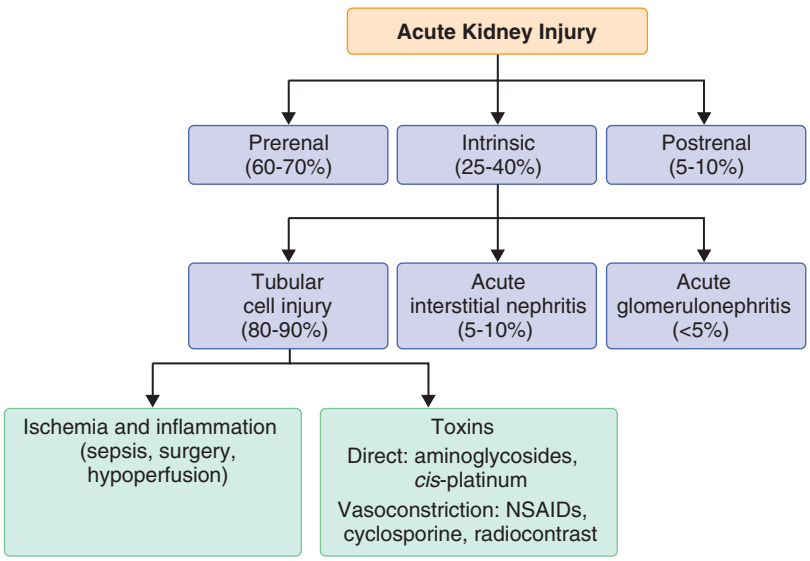

FIGURE 120-1. Main categories of acute kidney injury. NSAIDs = nonsteroidal anti-inflammatory drugs.

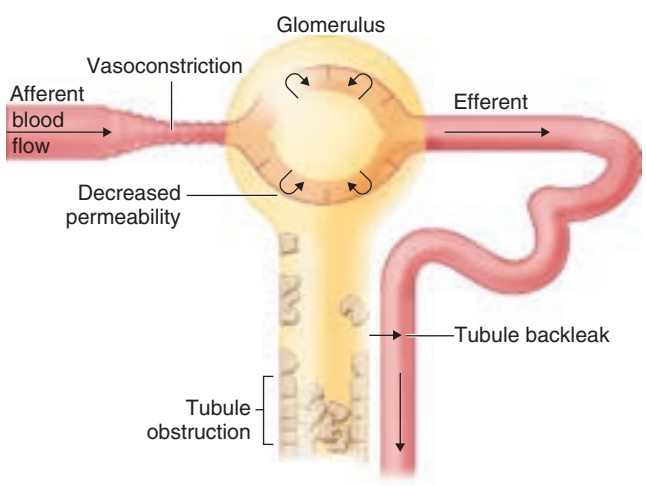

FIGURE 120-2. Mechanisms of prerenal and intrinsic acute renal injury. See text for descriptions.

TABLE 120-1 KDIGO ACUTE KIDNEY INJURY CLASSIFICATION

STAGE	SERUM CREATININE	URINE OUTPUT
1	1.5-1.9 times baseline OR ≥0.3 mg/dL (≥26.5 µmol/L) increase	<0.5 mL/kg/hr for 6-12 hr
2	2.0-2.9 times baseline	<0.5 mL/kg/hr for ≥12 hr
3	3.0 times baseline OR Increase in serum creatinine to ≥4.0 mg/dL (≥353.6 µmol/L) OR Initiation of renal replacement therapy OR In patients <18 yr, decrease in eGFR to <35 mL/min per 1.73 m^2	<0.3 mL/kg/hr for ≥24 hr OR Anuria for ≥12 hr

eGFR = estimated glomerular filtration rate; KDIGO = Kidney Disease Improving Global Outcomes

TABLE 120-2 COMMON RENAL TUBULAR TOXINS

Aminoglycosides
Radiocontrast agents
Acyclovir
Cisplatin
Sulfonamides
Methotrexate
Cyclosporine
Tacrolimus
Amphotericin B
Foscarnet
Pentamidine
Ethylene glycol
Toluene
Cocaine
HMG-CoA reductase inhibitors

HMG-CoA = 3-hydroxy-3-methylglutaryl coenzyme A.

TABLE 120-3 MEDICATIONS ASSOCIATED WITH ACUTE INTERSTITIAL NEPHRITIS

β-LACTAM ANTIBIOTICS
Penicillin
Cephalosporins
Ampicillin
Methicillin
Nafcillin
DIURETICS
Furosemide
Hydrochlorothiazide
Triamterene
OTHER ANTIBIOTICS
Sulfonamides
Vancomycin
Rifampin
Acyclovir
Indinavir
NSAIDS
Ibuprofen
Naproxen
Indomethacin

NSAIDs = nonsteroidal anti-inflammatory drugs.

Prerenal Azotemia

Prerenal azotemia, which is the most common cause of AKI, is a result of renal hypoperfusion. It accounts for approximately 60 to 70% of community-acquired and 40% of hospital-acquired cases. Hypoperfusion occurs in disease states that reduce effective intravascular volume, such as volume depletion from bleeding over-diuresis, sepsis (Chapter 108), heart failure (Chapter 58), or liver failure (Chapter 154). Additionally, medications that act directly to reduce glomerular capillary perfusion, such as angiotensin-converting enzyme (ACE) inhibitors, angiotensin-receptor blockers (ARBs), and nonsteroidal anti-inflammatory drugs (NSAIDs), also can cause prerenal AKI. The use of these agents in a patient with underlying renal hypoperfusion is to be avoided.

Intrinsic Acute Kidney Injury

Intrarenal AKI often results when untreated or untreatable severe hypoperfusion leads to cellular injury and ischemic AKI. The diverse causes of intrinsic AKI can involve any portion of the renal vasculature, nephron, or interstitium (Fig. 120-2). Ischemic and septic injury are major causes. Renal toxins, such as radiocontrast agents and aminoglycosides, also can damage tubules both directly and indirectly (Table 120-2). Fortunately, AKI does not develop in every patient exposed to these agents, but elderly patients with diabetes mellitus, hemodynamically unstable patients, and patients with a reduced effective arterial volume (heart failure, burns, cirrhosis, hypoalbuminemia) are the most susceptible to toxic renal injury. In fact, the incidence of aminoglycoside antibiotic nephrotoxicity increases from 3 to 5% to 30 to 50% in these high-risk patients.

AKI secondary to injury to the renal interstitium is termed *acute interstitial nephritis*. Commonly implicated medications for interstitial nephritis include penicillins, cephalosporins, sulfonamides, and NSAIDs (Table 120-3) (Chapter 122). Bacterial and viral infections also can be the causative agents. Interstitial nephritis is also associated with a kidney-confined or systemic autoimmune process, such as systemic lupus erythematosus (Chapter 266), Sjögren syndrome (Chapter 268), cryoglobulinemia (Chapter 187), and primary biliary cirrhosis (Chapter 153).

Postrenal Acute Kidney Injury

Postrenal AKI can occur in the setting of bilateral urinary outflow obstruction or in a patient with a solitary kidney when a single urinary outflow tract is obstructed (Chapter 123). Most commonly, this type of outflow obstruction is observed in patients with prostatic hypertrophy (Chapter 129), prostatic or cervical cancer (Chapter 199), or retroperitoneal disorders, including lymphadenopathy. A functional obstruction also can be observed in patients with a neurogenic bladder. In addition, intraluminal obstruction can be seen in patients with bilateral renal calculi (Chapter 126), papillary necrosis, blood clots, and bladder carcinoma, whereas extraluminal obstruction can develop in connection with retroperitoneal fibrosis, colon cancer, and lymphomas. Finally, intratubular crystallization of compounds such as uric acid, calcium oxalate, acyclovir, sulfonamide, and methotrexate, as well as myeloma light chains, can result in tubular obstruction.

PATHOBIOLOGY

The causes of AKI are diverse, and it can arise from a number of physiologic insults that injure the kidney and reduce the glomerular filtration rate (GFR). Decreased kidney perfusion and a reduced GFR can occur with or without cellular injury; toxic, ischemic, or obstructive injury to the nephron; inflammation and edema of the tubulointerstitium; and a primary glomerular disease process.

TABLE 120-4 CONDITIONS THAT LEAD TO ISCHEMIC ACUTE RENAL FAILURE

MECHANISM	CONDITION
Intravascular volume depletion and hypotension	Hemorrhage; gastrointestinal, renal, and dermal losses
Decreased effective intravascular volume	Heart failure, cirrhosis, hepatorenal syndrome, peritonitis
Systemic vasodilation, renal vasoconstriction	Sepsis, hepatorenal syndrome
Large-vessel renal vascular disease	Renal artery thrombosis or embolism, intraoperative arterial cross-clamping, renal artery stenosis, cholesterol emboli
Small-vessel renal vascular disease	Sepsis, vasculitis, atheroembolism, hemolytic-uremic syndrome, malignant hypertension, scleroderma, preeclampsia, sickle cell anemia, hypercalcemia, transplant rejection
Impaired renal blood flow	Cyclosporine, tacrolimus, ACEIs, ARBs, NSAIDs, radiocontrast agents

ACEIs = angiotensin-converting enzyme inhibitors; ARBs = angiotensin-receptor blockers; NSAIDs = nonsteroidal anti-inflammatory drugs.

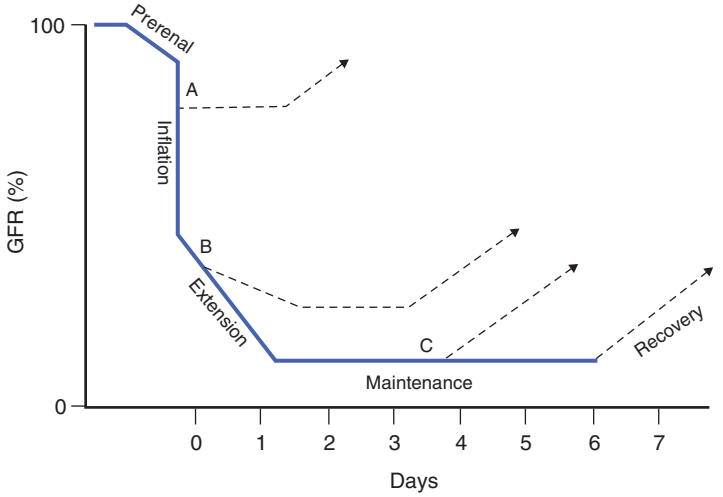

FIGURE 120-3. Phases of acute kidney injury. GFR = glomerular filtration rate. (From Sutton TA, Fisher CJ, Molitoris BA. Microvascular endothelial injury and dysfunction during ischemic acute renal failure. *Kidney Int.* 2002;62:1539-1549.)

Prerenal Acute Kidney Injury

The precipitating event for prerenal AKI is hypoperfusion of the kidney (see Fig. 120-2), which can be caused by a reduction in the total or intravascular fluid volume or disease states associated with normal or even increased total or intravascular fluid volumes but decrements in effective arterial volume, such as in sepsis, heart failure, and advanced cirrhosis. Prerenal azotemia is also divided functionally into volume responsive and nonresponsive azotemia, based on the response to hydration. For example, in severe heart failure (Chapter 58), additional intravascular volume may not improve kidney perfusion, whereas afterload reduction may improve perfusion by increasing cardiac output. Early in the course of prerenal AKI, the renal parenchyma remains intact and functional. During this initial phase, the GFR remains largely intact because kidney hypoperfusion initiates a neurohormonal cascade that results in afferent arteriolar dilation and efferent arteriolar constriction, thereby maintaining glomerular perfusion pressure. Because prerenal azotemia is often easily reversible, and mortality rates are low, early diagnosis and correction of the underlying pathophysiology are of critical importance. However, without early medical corrective intervention, prerenal azotemia progresses, ischemia worsens, and the resulting injury to tubular epithelial cells further decreases the GFR. This progression from prerenal azotemia to ischemic AKI is a continuum that depends on the severity and duration of the pathophysiologic insult.

Intrarenal Acute Kidney Injury

Intrinsic AKI is classified according to the primary histologic site of injury: tubules, interstitium, vasculature, or glomerulus. Renal tubular epithelial cell injury, commonly termed *acute tubular necrosis* (ATN), occurs more commonly in the setting of ischemia, although the renal tubules also can be damaged by specific kidney toxins. Ischemia can arise from a number of different clinical scenarios, but the common underlying pathogenesis is reduced renal blood flow (Table 120-4) with progression from prerenal azotemia to ischemic AKI in four distinct clinical and cellular phases: initiation, extension, maintenance, and recovery. Each of these phases encompasses distinct cellular events and declines in GFR as the kidneys respond to the insult and attempt to maintain and reestablish function (Fig. 120-3).[4] The initiation phase, which marks the transition from prerenal to tubular cell injury and dysfunction, is characterized by severe cellular depletion of adenosine triphosphate. Renal tubular epithelial cell injury, especially of proximal tubular cells, is a prominent feature during this phase, but injury to endothelial and vascular smooth muscle cells also has been documented. During this phase, extensive signaling between the proximal tubular cells and adjacent endothelial cells results in endothelial dysfunction and an inflammatory endothelial response.[5] Leukocytes of all types play a role in ongoing inflammation and cell injury. Dendritic cells, macrophages, neutrophils, and lymphocytes have been shown to play either a detrimental or protective role. The time course of involvement varies according to the cell type, and changes in macrophage phenotype from M1 to M2 mediate conversion from a proinflammatory form to a repair-mediating form.

During the extension phase, microvascular congestion with continued hypoxia and inflammation are most pronounced in the corticomedullary junction of the kidney, where reperfusion is limited owing to endothelial dysfunction at the capillary and postcapillary venule levels, with white blood cell adhesion and rouleaux formation.[6] The GFR is at its ebb during the maintenance phase as cells undergo repair, migration, and proliferation and as the kidney attempts to reestablish cellular and tubular integrity. Finally, during the recovery phase, the GFR begins to improve as cellular differentiation continues and normal cellular and organ function returns. Proximal tubular cells undergo cellular repair, and terminally differentiated epithelial cells re-express stem cell markers and divide to repopulate the nephron.[7] This last phase is often heralded by increasing urine output.

The S3 segment of the proximal tubule is located in the outer stripe of the medullary region of the nephron. This region is particularly susceptible to continued reduced perfusion after injury, and ongoing or worsening hypoxia results in continued cellular injury. Proximal tubular cell injury during the initiation phase of renal ischemia is first manifested as bleb formation in the apical membranes, with loss of the brush border. Proximal tubule cells also lose the polarity of the surface membrane and the integrity of their tight junctions. As the injury progresses, both live and necrotic proximal cells detach and enter the tubular lumen, where they ultimately form casts in the distal tubule. Casts contribute to a reduction in GFR by obstructing tubular urine flow, thereby preventing further filtration into that nephron. In addition, loss of the epithelial cell barrier and cell tight junctions allows back-leak of the glomerular filtrate into the interstitium, thus further compromising GFR (see Fig. 120-2).

Common agents that can cause direct tubular cell toxicity (see Table 120-2) include the aminoglycoside antibiotics, intravenous radiocontrast agents, and cisplatin. Other agents such as radiocontrast dyes, NSAIDs, and cyclosporine induce vasoconstriction and reduce kidney perfusion. Cocaine and the 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors can damage skeletal muscle and cause rhabdomyolysis (Chapter 113), thereby resulting in the release of myoglobin that is toxic to the tubular epithelium. Finally, the precipitation of some compounds or their metabolites can cause intratubular obstruction; agents in this category include acyclovir, sulfonamides, ethylene glycol (calcium oxalate metabolite; Chapter 110), methotrexate, and the light chains of multiple myeloma (Chapter 187).

Sepsis is a very common cause of intrinsic AKI. Although the causes of septic AKI are often multifactorial, ischemia owing to poor microvascular perfusion is a major factor. Interestingly, proximal tubular cells act as part of the innate immune system to detect danger-associated molecular patterns and establish pathogen-associated recognition patterns via toll-like receptors (TLRs).[8] Lipopolysaccharide is a prime example of proximal tubule TLR4-mediated uptake that results in subsequent signaling via cytokines and oxidative stress. The resulting histology in humans is patchy involvement of cells with apoptosis and minimal cellular necrosis that cannot fully explain the severity of sepsis-induced kidney dysfunction. Coagulation abnormalities of the microvasculature also play an important role in kidney dysfunction and ongoing ischemia.

In AKI caused by the interstitial injury, a mixed inflammatory infiltrate composed of T lymphocytes, monocytes, and macrophages, is seen. These inflammatory lesions can be diffuse or patchy in distribution. Occasionally, granulomas also can be observed, especially in drug hypersensitivity reactions. Acute interstitial nephritis that persists and becomes chronic is characterized by interstitial fibrosis and tubular atrophy, although foci of inflammatory cells can persist. This process can lead to chronic and even end-stage kidney disease requiring chronic dialysis.

Vascular causes of intrinsic AKI can include microvascular and macrovascular processes. Classic microvascular disorders, which include thrombotic thrombocytopenic purpura (Chapter 172), sepsis (Chapter 108), hemolytic-uremic syndrome (Chapter 172), and the HELLP syndrome (*h*emolysis, *e*levated *l*iver enzymes, and *l*ow *p*latelet count; Chapters 146 and 239), cause AKI as a result of glomerular capillary thrombosis and microvascular occlusion. Macrovascular disease such as atherosclerosis can cause AKI secondary to atheroembolization (Chapter 125), especially during or after an invasive or interventional vascular procedure in a patient with preexisting atherosclerotic disease.

A less common cause of AKI is glomerulonephritis (Chapter 121), which can be seen in systemic lupus nephritis (Chapter 266), granulomatosis with polyangiitis (Chapter 270), polyarteritis nodosa (Chapter 270), Goodpasture syndrome (Chapter 121), Henoch-Schönlein purpura (Chapter 270), and hemolytic-uremic syndrome (Chapter 172). AKI in this setting is termed *rapidly progressive glomerulonephritis* and results from direct inflammatory glomerular or vascular injury.

Postrenal Acute Kidney Injury

Postrenal AKI is caused by obstruction to luminal flow of the glomerular filtrate. This obstruction results in a relatively complex pathophysiology that begins with transmission of backpressure to Bowman space of the glomerulus. Intuitively, this backpressure would be expected to reduce the GFR. However, by dilation of the glomerular afferent arteriole, the GFR remains largely preserved. Unfortunately, such compensation is only transient, and the GFR will begin to attenuate if the obstruction is not rapidly relieved. With continued obstruction for more than 12 to 24 hours, renal blood flow and intratubular pressure decline and large unperfused and underperfused areas of the kidney cortex result in a reduction in GFR.

CLINICAL MANIFESTATIONS

AKI, even when advanced, frequently is first diagnosed by abnormalities observed in a patient's laboratory studies and not by any specific symptom or sign. The clinical manifestations associated with AKI are frequently protean, occur late in the course, and are often not apparent until kidney dysfunction has become severe. The clinical findings of AKI also depend on the stage at which it is diagnosed. Patients with AKI may report symptoms such as anorexia, fatigue, nausea and vomiting, and pruritus, as well as a decline in urine output or dark-colored urine. Furthermore, if the patient has become volume overloaded, shortness of breath and dyspnea on exertion may be noted.

A thorough physical examination with special emphasis on determination of volume status and effective arterial volume is essential. If volume overload is present, jugular venous distention, pulmonary crackles, and peripheral edema may be found (Chapter 58). Findings such as asterixis, myoclonus, or a pericardial rub may be seen in severe AKI.

DIAGNOSIS

A systematic approach that considers each of the three major categories in the pathogenesis of AKI will ensure that an accurate diagnosis and an appropriate therapeutic plan will be achieved. An appropriate diagnostic strategy is to exclude prerenal and postrenal causes first and then, if needed, begin an evaluation for possible intrinsic causes.

Laboratory analysis of blood and urine samples of patients with AKI reveals the level of dysfunction, will frequently suggest a cause, and may also direct the rapidity with which a specific therapy needs to be instituted. All patients with clinical findings of AKI should be evaluated with serum measurements of electrolytes, Cr, calcium, and phosphorus; a blood urea nitrogen level; and complete blood count with differential. In addition, urine studies, including sodium, potassium, chloride, and Cr determinations for calculation of the fractional excretion of sodium (FE_{Na}), are important. The formula for calculating FE_{Na} is as follows:

$$FE_{Na} = \frac{\text{Urine Na} \times \text{Plasma Cr}}{\text{Plasma Na} \times \text{Urine Cr}} \times 100$$

TABLE 120-5 FE_{NA} VALUES FOR THE VARIOUS CAUSES OF ACUTE KIDNEY INJURY

CAUSE OF ACUTE KIDNEY INJURY	FE_{Na}	BUN-TO-SERUM CREATININE RATIO
Prerenal	<1%	>20
Intrarenal		<10-15
Tubular necrosis	≥1%	
Interstitial nephritis	≥1%	
Glomerulonephritis (early)	<1%	
Vascular disorders (early)	<1%	
Postrenal	≥1%	>20

TABLE 120-6 COMMON URINALYSIS FINDINGS IN ACUTE KIDNEY INJURY

CAUSE OF ACUTE KIDNEY INJURY	URINALYSIS
Prerenal	Normal or hyaline casts
Intrarenal	
Tubular cell injury	Muddy-brown, granular, epithelial casts
Interstitial nephritis	Pyuria, hematuria, mild proteinuria, granular and epithelial casts, eosinophils
Glomerulonephritis	Hematuria, marked proteinuria, red blood cell casts, granular casts
Vascular disorders	Normal or hematuria, mild proteinuria
Postrenal	Normal or hematuria, granular casts, pyuria

The numerical value of FE_{Na} can be helpful in determining the potential cause of the AKI (Table 120-5). In some cases, it is better to use FE_{Cl} because urinary sodium can be elevated during systemic alkalosis when a high urinary bicarbonate level obligates the loss of sodium. Urine dipstick and microscopy (Table 120-6) should be performed on a fresh urine sample because important cellular elements that could indicate potential causes degrade rapidly with time. Finally, renal ultrasound to determine the presence or absence of outlet obstruction also should be included in the initial evaluation. Measurement of urine and serum levels of structural biomarkers, such as kidney injury molecule 1 (KIM-1) and inflammatory markers, such as neutrophil gelatinase–associated lipocalin (NGAL) and interleukin 18, may aid in the diagnosis of AKI, although the data are not yet conclusive.[9]

Prerenal Azotemia

Prerenal azotemia, which is the most common cause of renal dysfunction, often can be determined by the patient's history. Common historical features in patients with prerenal azotemia include vomiting, diarrhea, and poor oral intake. Heart failure can suggest a possible prerenal cause of reduced renal perfusion from over-diuresis or as exacerbation of the heart failure itself. Other medications that can attenuate renal perfusion, such as NSAIDs, ACE inhibitors, and ARBs, can cause prerenal azotemia. Common physical examination findings include tachycardia, systemic or orthostatic hypotension (or both), and dry mucous membranes.

Laboratory studies in patients with prerenal azotemia demonstrate elevated serum Cr and blood urea nitrogen (BUN) levels. FE_{Na} is typically less than 1%. However, in a patient taking diuretics such as furosemide, FE_{Na} may be greater than 1% even though the patient has prerenal azotemia because of diuretic-induced natriuresis. For these clinical situations, the fractional excretion of urea can be used and is calculated in similar fashion:

$$FE_{urea} = \frac{\text{Urine Urea} \times \text{Plasma Cr}}{\text{Plasma Urea} \times \text{Urine Cr}} \times 100$$

FE_{urea} less than 35% suggests prerenal AKI. Other causes of an FE_{Na} greater than 1% include the presence of a non-reabsorbable solute such as bicarbonate, glucose, or mannitol. Chronic kidney disease, ATN, and late obstructive nephropathy are also associated with FE_{Na} greater than 1%. Therefore, in these disease states, FE_{Na} cannot provide reliable diagnostic information regarding AKI unless the FE_{Na} is less than 1%. Moreover, FE_{urea} has not been validated for these clinical entities.

Another laboratory parameter to assist in diagnosing prerenal AKI is the ratio of BUN to serum Cr. Commonly, a patient with prerenal azotemia will have a ratio of BUN to serum Cr of greater than 20 : 1.

Intrarenal Acute Kidney Injury

A history of hypotension or exposure to a nephrotoxin or medication is a common finding in patients with intrarenal AKI. The nephrotoxin can be a specific tubular toxin that causes ATN or a medication that causes an allergic reaction as in acute interstitial nephritis (see Tables 120-2 and 120-3). Physical examination may reveal signs and symptoms of volume depletion or fluid overload. It is important to remember that ATN often results from a persistent severe prerenal state and the prerenal state must first be corrected to prevent ongoing or worsening ATN. Rash may accompany acute interstitial nephritis. Cholesterol embolism in patients with severe atherosclerotic disease (Chapter 125) may manifest classically as cyanotic digits and AKI; this finding is frequently seen after invasive vascular surgery or an interventional study.

Laboratory studies will demonstrate elevated serum Cr and BUN levels in intrarenal AKI. ATN and acute interstitial nephritis are frequently associated with FE_{Na} greater than 1, whereas FE_{Na} is typically less than 1 in early radiocontrast-induced AKI, sepsis, glomerulonephritis, and vascular disorders. Peripheral eosinophilia and urinary eosinophils may be present in acute interstitial nephritis, although the latter are neither sensitive nor specific for this type of AKI. Urinary eosinophils are also associated with cholesterol microembolic disease (Chapter 125). Intrarenal AKI has specific urinalysis findings that can be helpful in making diagnostic and therapeutic decisions (see Table 120-6).

Postrenal Acute Kidney Injury

A history of prostatic hypertrophy (Chapter 129), prostate cancer (Chapter 201), lymphoma (Chapter 185), cervical cancer (Chapter 199), or retroperitoneal disease often can be found in patients with postrenal AKI. Postrenal AKI should always be in the differential diagnosis of patients with severe oliguria (urine output < 450 mL/day) or anuria (urine output < 100 mL/day). However, many patients with postrenal AKI are neither oliguric nor anuric. Beyond an elevation in a patient's serum Cr and BUN levels, laboratory studies generally yield benign results. Bladder catheterization can be both diagnostic and therapeutic in postrenal AKI. However, renal ultrasound is the diagnostic test of choice, although it may be falsely negative early in postrenal AKI.

TREATMENT

The cornerstones of therapy for AKI are rapid recognition and correction of reversible causes such as hypoperfusion, avoidance of any further renal injury, and correction and maintenance of a normal electrolyte and fluid volume milieu. Preventive therapy or medical interventions performed during the initiation and extension phases of AKI provide the greatest chance for minimizing the extent of injury (see Fig. 120-3, *lines A* and *B*) and hastening renal recovery (see Fig. 120-3, *line B*); interventions provided during the maintenance phase of AKI have not proved beneficial (see Fig. 120-3, *line C*). If prerenal AKI is not addressed early in a patient's course or if the patient is seen late in the course, ATN may occur and markedly increase morbidity and mortality.[10]

Prerenal azotemia in its early stages often can be rapidly corrected by aggressive normalization of effective arterial volume, although more care must be taken during volume resuscitation in patients with a history of heart failure, cirrhosis, and sepsis. Key approaches include administering volume (e.g., normal saline) to achieve euvolemia, improving cardiac output by afterload reduction (Chapter 59), or normalizing systemic vascular resistance.

Postrenal AKI secondary to prostatic hypertrophy frequently can be corrected by placement of a bladder catheter. However, outlet obstruction from a neoplastic process will usually require urologic consultation for consideration of ureteral stenting or placement of a percutaneous nephrostomy tube.

Intrarenal AKI can be the most complex and difficult to treat. AKI caused by glomerulonephritis (Chapter 121) or vasculitis (Chapters 266 and 270) will frequently require immunosuppressive therapy. For suspected acute interstitial nephritis, the offending medication must be determined and discontinued; a 2-week tapering course of glucocorticoids, beginning with 1 mg/kg of prednisone (up to 60 mg) for 3 days, is commonly recommended despite the absence of data from randomized trials.

General supportive measures include avoiding any further nephrotoxins and paying careful attention to the patient's fluid balance by monitoring weight and daily input and output. In addition, serum electrolytes, Cr, and BUN should be monitored at least daily more frequently if the patient's renal function appears to be tenuous. Patients with AKI also should receive a diet low in sodium, potassium, and protein, which can be liberalized as the patient's renal function improves. A phosphate binder (e.g., calcium acetate [1334 mg], lanthanum carbonate [500 mg], sevelamer [300 to 1000 mg], or aluminum hydroxide [300 to 600 mg] with each meal; see Table 119-4) is also usually helpful in controlling the serum phosphate level by minimizing the absorption of dietary phosphate.

Early nephrology consultation will ensure that the patient receives optimal care. Some patients will warrant urgent hemodialysis because of marked metabolic acidosis unresponsive to sodium bicarbonate infusions; electrolyte abnormalities, such as hyperkalemia that is unresponsive to medical management; pulmonary edema not responding to diuretic therapy; and uremic symptoms of encephalopathy, seizures, and pericarditis. In the absence of acute indications, however, when to initiate dialysis in AKI remains unresolved.

For AKI, early initiation of dialysis (Cr level 7.5 mg/dL) is no better than standard dialysis beginning at a Cr level of approximately 10 mg/dL.[■] Furthermore, intensive dialytic renal support therapy is no better than standard dialytic therapy, and intermittent hemodialysis and continuous renal replacement therapy lead to similar clinical outcomes in acute renal failure. However, it is important to verify that the prescribed dialysis is received and that standardized measures are achieved. Some patients—especially patients in an increased catabolic state, trauma patients, and patients receiving glucocorticoids—may require dialysis more than three times per week to achieve adequate therapy (Chapter 131). Neither furosemide nor low-dose dopamine improve outcome, even though low-dose dopamine may temporarily improve metrics of renal physiology.

PREVENTION

Given the marked increase in morbidity and mortality associated with AKI, especially for critically ill patients, potential measures to prevent AKI are essential. The first step in prevention, however, is being aware of patients who are at highest risk for AKI because of known kidney disease or comorbid medical conditions, such as chronic kidney disease, diabetes, hypertension, nephrotic syndrome, heart failure, age, and peripheral vascular disease.

Of all the risk factors for acquiring AKI, the presence of preexisting chronic kidney disease is the most predictive. Recent data document a vicious cycle involving AKI and CKD (Fig. 120-4). Appropriate hospital surveillance measures include avoiding nephrotoxic medications (e.g., NSAIDs and aminoglycosides); minimizing diagnostic procedures that require radiocontrast material, especially in the prerenal patient; and careful monitoring of urinary output with daily determination of serum electrolyte and Cr levels after any procedures known to induce AKI. Additionally, educating the patient regarding common nonprescription nephrotoxins such as NSAIDs can reduce the risk for AKI in outpatients.

Before a potentially nephrotoxic exposure, early consultation with a nephrologist is warranted for this high-risk group to advise whether a specific medication or intervention may reduce the risk for AKI or whether an alternative medication or procedure, such as magnetic resonance imaging instead of computed tomography with intravenous radiocontrast agents, may be preferred. All potential nephrotoxins, such as NSAIDs, should be discontinued before a potentially nephrotoxic procedure and avoided after it. The patient's volume and hemodynamic status must be maximized both before and after the event.

In high-risk patients, a renal protective intervention is often instituted before exposure to the agent. Interventions that may be useful for preventing AKI associated with intravenous radiocontrast agents include hydration with intravenous sodium chloride[■■] and short-term rosuvastatin (40 mg on admission then 20 mg/day or 10 mg/day for 2 days before and 3 days after

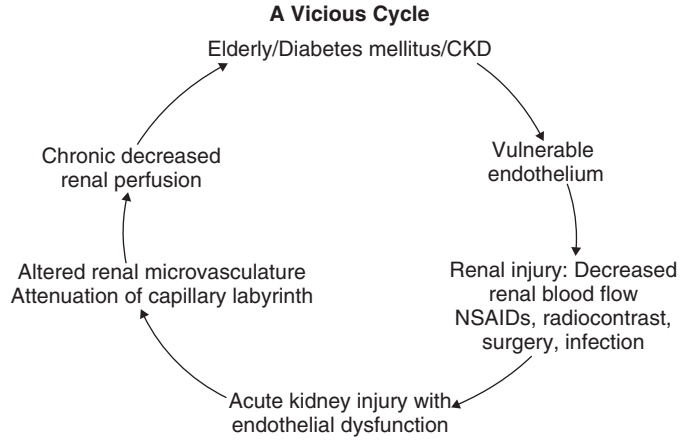

FIGURE 120-4. A vicious cycle exists between acute kidney injury (AKI) and chronic kidney disease (CKD), with CKD increasing the risk for developing AKI, and AKI accelerating progression of CKD.

the procedure).[■■] *N*-Acetylcysteine is no longer recommended for this purpose.[■]

PROGNOSIS

Typically, AKI secondary to prerenal causes, if diagnosed and treated early, has the best prognosis for renal recovery. Patients with prerenal AKI commonly return to their baseline level of renal function and have a mortality rate of less than 10%. Similarly, patients with postrenal AKI also have a good prognosis for renal recovery if the outlet obstruction is promptly diagnosed and definitively treated.

In contrast, patients with intrarenal AKI have a less predictable renal outcome, and mortality in this group varies between 30 and 80%, depending on the severity of injury. Higher mortality rates occur in older patients with hospital-acquired AKI admitted to ICUs. Furthermore, mortality in patients with AKI is incremental, and seemingly modest increases in serum Cr can result in marked increases in the mortality rate. Even a rise in serum Cr of only 0.3 mg/dL results in a significantly increased mortality risk.

The clinical course after recovery from ATN is subsequent tubular regeneration with recovery of renal function. However, this outcome is less ensured in patients with preexisting kidney disease. In addition, given the frequent systemic nature of their illness, patients with ATN, glomerulonephritis, and vasculitic causes of AKI may not fully recover to their baseline renal function.[11] Recovery is often difficult to quantify using serum creatinine because muscle wasting alone results in lower creatinine values.[12] Patients who have a severe episode of AKI requiring hemodialysis may not recover their renal function and may need hemodialysis indefinitely (Chapter 131), especially if they have a preexisting history of chronic kidney disease.[13] AKI hastens progression of chronic kidney disease to end-stage kidney disease and is often the major factor that causes such progression. It is also important that a nephrologist see most patients with hospital-acquired AKI and follow them in terms of any progression, hypertension, or other abnormalities.

Grade A References

A1. Jamale TE, Hase NK, Kulkarni M, et al. Earlier-start versus usual-start dialysis in patients with community-acquired acute kidney injury: a randomized controlled trial. *Am J Kidney Dis*. 2013;62:1116-1121.

A2. Koc F, Ozdemir K, Altunkas F, et al. Sodium bicarbonate versus isotonic saline for the prevention of contrast-induced nephropathy in patients with diabetes mellitus undergoing coronary angiography and/or intervention: a multicenter prospective randomized study. *J Investig Med*. 2013;61:872-877.

A3. Brar SS, Aharonian V, Mansukhani P, et al. Haemodynamic-guided fluid administration for the prevention of contrast-induced acute kidney injury: the POSEIDON randomised controlled trial. *Lancet*. 2014;383:1814-1823.

A4. Han Y, Zhu G, Han L, et al. Short-term rosuvastatin therapy for prevention of contrast-induced acute kidney injury in patients with diabetes and chronic kidney disease. *J Am Coll Cardiol*. 2014;63:62-70.

A5. Leoncini M, Toso A, Maioli M, et al. Early high-dose rosuvastatin for contrast-induced nephropathy prevention in acute coronary syndrome: results from the PRATO-ACS Study (Protective Effect of Rosuvastatin and Antiplatelet Therapy On contrast-induced acute kidney injury and myocardial damage in patients with Acute Coronary Syndrome). *J Am Coll Cardiol*. 2014;63:71-79.

A6. Acetylcysteine for prevention of renal outcomes in patients undergoing coronary and peripheral vascular angiography: main results from the randomized Acetylcysteine for Contrast-Induced Nephropathy Trial (ACT). *Circulation*. 2011;124:1250-1259.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO clinical practice guide for acute kidney injury. *Kidney Int Suppl*. 2012;2:1-138.
2. Zeng X, McMahon GM, Brunelli SM, et al. Incidence, outcomes, and comparisons across definitions of AKI in hospitalized individuals. *Clin J Am Soc Nephrol*. 2014;9:12-20.
3. Susantitaphong P, Cruz DN, Cerda J, et al. World incidence of AKI: a meta-analysis. *Clin J Am Soc Nephrol*. 2013;8:1482-1493.
4. Sharfuddin AA, Molitoris BA. Pathophysiology of ischemic acute kidney injury. *Nat Rev Nephrol*. 2011;7:189-200.
5. Bellomo R, Kellum JA, Ronco C. Acute kidney injury. *Lancet*. 2012;380:756-766.
6. Molitoris BA. Therapeutic translation in acute kidney injury: the epithelial/endothelial axis. *J Clin Invest*. 2014;124:2355-2363.
7. Kusaba T, Lalli M, Kramann R, et al. Differentiated kidney epithelial cells repair injured proximal tubule. *Proc Natl Acad Sci U S A*. 2014;111:1527-1532.
8. Gomez H, Ince C, De Backer D, et al. A unified theory of sepsis-induced acute kidney injury: inflammation, microcirculatory dysfunction, bioenergetics, and the tubular cell adaptation to injury. *Shock*. 2014;41:3-11.
9. Parikh CR, Coca SG, Thiessen-Philbrook H, et al. Postoperative biomarkers predict acute kidney injury and poor outcomes after adult cardiac surgery. *J Am Soc Nephrol*. 2011;22:1748-1757.
10. Prowle JR, Kirwan CJ, Bellomo R. Fluid management for the prevention and attenuation of acute kidney injury. *Nat Rev Nephrol*. 2014;10:37-47.
11. Chawla LS, Eggers PW, Star RA, et al. Acute kidney injury and chronic kidney disease as interconnected syndromes. *N Engl J Med*. 2014;371:58-66.
12. Prowle JR, Kolic I, Purdell-Lewis J, et al. Serum creatinine changes associated with critical illness and detection of persistent renal dysfunction after AKI. *Clin J Am Soc Nephrol*. 2014;9:1015-1023.
13. Stads S, Fortrie G, van Bommel J, et al. Impaired kidney function at hospital discharge and long-term renal and overall survival in patients who received CRRT. *Clin J Am Soc Nephrol*. 2013;8:1284-1291.

REVIEW QUESTIONS

1. Acute kidney injury is a common disorder in hospitalized patients, occurring in up to 20% of all adult admissions. Within a broad differential, which particular process is most likely as the cause of acute kidney injury in hospitalized patients?

A. Outflow obstruction
B. Glomerulonephritis
C. Acute interstitial nephritis
D. Prerenal azotemia
E. Ischemic acute tubular necrosis

Answer: D Up to 50% of patients diagnosed with acute kidney injury in a hospital setting have prerenal azotemia. It is important to make this diagnosis, and it usually is a rapidly reversible alteration that is responsive to fluid therapy, with low morbidity and mortality if treated rapidly and appropriately. However, prerenal azotemia also can lead to ischemic acute tubular necrosis if it persists or worsens.

2. The diagnosis of prerenal azotemia is made after an appropriate history and physical examination. What urinary and serum biomarkers can be used to help confirm the diagnosis of prerenal azotemia?

A. A ratio of BUN to creatinine greater than 20
B. A fractional excretion of sodium less than 1%
C. A bland urine analysis
D. A and B
E. All of the above

Answer: E Prerenal azotemia is an adaptation by the kidney to hypoperfusion without cellular injury. In the prerenal state, the kidney is being inadequately perfused but not yet to a level that results in low cellular ATP and cell injury. The kidney tries to increase the volume status of the patient by reabsorbing nearly all of the filtered sodium, thereby resulting in a low FE_{Na}. Urea reabsorption is also increased, thereby resulting in the high ratio of BUN to creatinine.

3. Risk factors for the development of acute kidney injury include all of the following *except*?

A. Chronic kidney disease
B. Volume depletion
C. Use of nonsteroidal anti-inflammatory agents
D. Hypertension with blood pressure of 160/100 mm Hg
E. Nephrotoxins

Answer: D Hypertension causes chronic kidney injury but does not usually cause acute kidney injury except with malignant hypertension. Of the remaining four risk factors, chronic kidney disease is the most important; the more severe the chronic kidney disease, the more likely acute kidney injury is to occur.

4. Serum creatinine is used as a marker of glomerular filtration and therefore kidney function in patients. Numerous attempts have been made to develop estimating equations based around serum creatinine and other serum markers in an attempt to help quantify glomerular filtration rate without undertaking timely and costly studies. In a patient with acute kidney injury and either a rising or falling serum creatinine, the calculation of an eGFR is inappropriate for which of the following reasons?

A. Serum creatinine is not in equilibrium and therefore it is not possible to calculate eGFR.
B. There may be an alteration in the rate of creatinine released from muscles in a hospitalized patient.
C. Other factors such as fever, glucocorticoids, or trauma, influence creatinine release.
D. With complete loss of kidney function, the serum creatinine can rise between 1 and 4 mg/dL/day depending on the patient's circumstances.
E. All of the above

Answer: E Although a stable serum creatinine can be used to estimate the glomerular filtration rate using estimating equations, the equations are reasonably accurate only when the serum creatinine level is stable. Multiple factors, including fever, glucocorticoids, or trauma, affect the muscle release of creatine, which is the precursor of serum creatinine.

5. Dialysis for acute kidney injury is a sign of moderate-to-severe injury with an enhanced morbidity and mortality for the patient. Which of the following is true regarding dialysis for acute kidney injury?

A. Dialysis should be intensive both in frequency and duration to provide the patient with the most appropriate internal milieu.
B. The major reasons to start dialysis in a patient with acute kidney injury include acidosis, hyperkalemia, and volume overload.
C. Continuous renal replacement has been shown to have a better outcome for patients than intermittent hemodialysis.
D. The serum creatinine level at the initiation of dialysis inversely correlates with outcomes in patients with acute kidney injury.
E. Starting dialysis early in patients with sepsis improves outcomes.

Answer: B The patient's volume, potassium level, and acid-base status are major indications for dialysis in AKI. Adequate dialysis is critically important, but neither increasing the intensity of dialysis nor initiating it earlier improves outcomes.

121

GLOMERULAR DISORDERS AND NEPHROTIC SYNDROMES

GERALD B. APPEL AND JAI RADHAKRISHNAN

GLOMERULAR DISORDERS

DEFINITION

Each *glomerulus,* the basic filtering unit of the kidney, consists of a tuft of anastomosing capillaries formed by the branching of the afferent arteriole. Approximately 1 million glomeruli comprise approximately 5% of the kidney weight and provide almost 2 m^2 of glomerular capillary filtering surface. The glomerular basement membrane (GBM) provides both a size- and charge-selective barrier to the passage of circulating macromolecules. Renal pathologic processes involving all glomeruli are called *diffuse* or *generalized;* if only some glomeruli are involved, the process is called *focal.* When dealing with the individual glomerulus, a process is *global* if the whole glomerular tuft is involved and *segmental* if only part of the glomerulus is involved. The modifying terms *proliferative, sclerosing,* and *necrotizing* are often used (e.g., focal and segmental glomerulosclerosis; diffuse global proliferative lupus nephritis). Extracapillary proliferation or crescent formation is caused by the accumulations of macrophages, fibroblasts, proliferating epithelial cells, and fibrin within Bowman space. In general, crescent formation in any form of glomerular damage conveys a serious prognosis. Scarring of the tissue between the tubules and glomeruli, interstitial fibrosis, is also a poor prognostic sign in every glomerular disease.

EPIDEMIOLOGY

At present, more than 10% of the U.S. population may have proteinuria or renal dysfunction, often caused by glomerular diseases. More than 500,000 Americans are in end-stage renal disease (ESRD) programs, which cost approximately $33 billion per year and account for approximately 28% of all Medicare spending, largely as a result of renal involvement by glomerular diseases. Diabetic renal damage alone affects many millions of persons and is the major cause of ESRD in the United States (Chapter 124). Glomerular diseases associated with infectious agents such as malaria (Chapter 345), schistosomiasis (Chapter 355), human immunodeficiency virus (HIV) and the hepatitis B and C viruses (Chapter 149) are major worldwide health problems. Manifestations of glomerular injury range from asymptomatic microhematuria and albuminuria to rapidly progressive oliguric renal failure. Some patients develop massive fluid retention and edema at onset of their glomerular disease, whereas others present with only the slow insidious signs and symptoms of chronic renal failure (Chapter 130).

PATHOBIOLOGY

Common mechanisms, such as breaks in the glomerular capillary wall leading to hematuria and loss of the selective barrier to particles based on size and charge associated with proteinuria, are characteristic of glomerular diseases. Nevertheless, the nature of the initiating processes varies among different glomerular diseases. In some, such as diabetes and amyloidosis, structural and biochemical alterations are clearly present in the glomerular capillary wall. In others, immune-mediated renal injury is caused by deposition of circulating immune complexes, in situ formation of immune complexes, or the localized effects of anti–glomerular basement membrane (anti-GBM) antibodies. In still other diseases, genetic or acquired defects in the glomerular podocytes are associated with proteinuria and progressive renal dysfunction.

CLINICAL MANIFESTATIONS

Findings indicative of a glomerular origin of renal disease include erythrocyte casts and dysmorphic red blood cells (RBCs) in the urinary sediment, as well as large amounts of albuminuria. Persistent urinary excretion of more than 500 to 1000 erythrocytes per milliliter (or more than 5 RBCs per high-power field on microscopy) is abnormal. Dysmorphic RBCs, which are deformed as a result of their passage through the glomerular capillary wall and tubules, as well as RBC casts, which are formed when these erythrocytes become enmeshed in a proteinaceous matrix in the lumen of the tubules, are also indicative of glomerular disease.

In a normal person, the urinary excretion of albumin is less than 30 mg/day and the total urinary excretion of protein is less than 150 mg/day. Although increases in urinary protein excretion may come from the filtration of abnormal circulating proteins (e.g., light chains in multiple myeloma) or from the deficient proximal tubular reabsorption of normal, filtered low-molecular-weight proteins (e.g., β_2-microglobulin), the most common cause of proteinuria, and specifically albuminuria, is glomerular injury. Proteinuria associated with glomerular disease may range from several hundred milligrams to more than 30 g daily. In some diseases, such as minimal change nephrotic syndrome, albumin is the predominant protein in the urine. In others, such as focal sclerosing glomerulonephritis and diabetes, the proteinuria, although still largely composed of albumin, is nonselective and contains many higher-molecular-weight proteins.

DIAGNOSIS

Some patients have asymptomatic microhematuria or proteinuria discovered by routine evaluations. Microscopic hematuria associated with deformed RBCs or RBC casts is likely to be glomerular in origin. Subnephrotic levels of proteinuria may arise from orthostatic proteinuria, exercise, hypertension, tubular disease, or glomerular damage.

In patients with asymptomatic urinary abnormalities of glomerular origin, the underlying renal lesion may represent the early phase of a progressive glomerular disease or may be due to a benign, nonprogressive glomerular lesion. In general, if patients have less than 1 g of proteinuria daily or only glomerular microhematuria but a normal glomerular filtration rate (GFR) and no evidence of systemic disease, it is not necessary to proceed to a diagnostic renal biopsy. Most of these patients need no immunosuppressive therapy. The patient can be followed closely, and biopsy need be performed only in patients with progressively increasing proteinuria or evidence of a decreasing GFR.

THE NEPHROTIC SYNDROME

DEFINITION

The nephrotic syndrome (Table 121-1) is defined by albuminuria of more than 3 to 3.5 g/day accompanied by hypoalbuminemia, edema, and hyperlipidemia. In practice, many clinicians refer to "nephrotic range" proteinuria regardless of whether patients have the other manifestations of the full syndrome, because these are a consequence of the proteinuria.

EPIDEMIOLOGY

The nephrotic syndrome may be primary and idiopathic (Table 121-2), or it may be caused by a known underlying condition, such as diabetes, amyloidosis, or systemic lupus erythematosus (Table 121-3). Although minimal change disease is the most common cause of nephrotic syndrome in children, idiopathic membranous nephropathy and focal segmental glomerular sclerosis are the most common causes in adults, with the former being most common in whites and the latter in blacks.

PATHOBIOLOGY

Hypoalbuminemia, which is largely a consequence of urinary protein loss, also may be due to proximal tubular catabolism of filtered albumin, the redistribution of albumin within the body, and reduced hepatic synthesis of albumin. As a result, the relationship among urinary protein loss, the level of the serum albumin, and other secondary consequences of heavy albuminuria is inexact.

TABLE 121-1 TYPICAL FINDINGS OF NEPHROTIC SYNDROME VERSUS NEPHRITIS

NEPHROTIC SYNDROME	NEPHRITIS
RENAL INSUFFICIENCY	
Uncommon at presentation	Common at presentation
PROTEINURIA	
Typically high (>3 g/day)	Variable
Urine RBCs	
Few	Prominent
Urine RBC CASTS	
Unlikely	Likely

RBCs = red blood cells.

TABLE 121-2 CAUSES OF IDIOPATHIC NEPHROTIC SYNDROME IN ADULTS

CAUSE	INCIDENCE (%)
Minimal change disease	5-10
Focal segmental glomerulosclerosis	20-25
Membranous nephropathy	25-30
Membranoproliferative glomerulonephritis	5
Other proliferative and sclerosing glomerulonephritides	15-30

The salt and volume retention in the nephrotic syndrome may occur through at least two different major mechanisms. The classic teaching is that hypoalbuminemia reduces the oncotic pressure of plasma, the resulting intravascular volume depletion leads to activation of the renin-angiotensin-aldosterone axis, and this activation increases the retention of renal sodium and fluid. However, primary salt retention in the distal nephron also may occur independently of the renin-angiotensin-aldosterone axis.

Venous thrombosis often occurs in patients with the nephrotic syndrome. Explanations include urinary loss of antithrombin III, protein C, and protein S, which prevent thrombosis, as well as increased synthesis of acute phase reactants that promote thrombosis.

CLINICAL MANIFESTATIONS

Patients may present with weight gain, peripheral edema, and periorbital edema. Hypertension is common, and varying degrees of hematuria may be present. A 24-hour urine sample usually shows more than 3 to 3.5 g of proteinuria. Nephrotic patients often have a hypercoagulable state and are predisposed to deep vein thrombosis (Chapter 81), pulmonary emboli (Chapter 98), and renal vein thrombosis (Chapter 125). Patients with nephrotic syndrome have increased risk for atherosclerotic complications. Most nephrotic patients have elevated levels of total and low-density lipoprotein (LDL)

TABLE 121-3 NEPHROTIC SYNDROME ASSOCIATED WITH SPECIFIC CAUSES (SECONDARY NEPHROTIC SYNDROME)

SYSTEMIC DISEASES
Diabetes mellitus Systemic lupus erythematosus and other collagen vascular diseases Amyloidosis (amyloid AL- or AA-associated) Vasculitic-immunologic disease (mixed cryoglobulinemia, granulomatous polyangiitis, microscopic polyangiitis, rapidly progressive glomerulonephritis, Henoch-Schönlein purpura, anti-GBM disease)
INFECTIONS
Bacterial (post-streptococcal, congenital and secondary syphilis, subacute bacterial endocarditis, cerebral ventriculoatrial shunt nephritis) Viral (hepatitis B, hepatitis C, HIV infection, infectious mononucleosis, cytomegalovirus infection) Parasitic (malaria, toxoplasmosis, schistosomiasis, filariasis)
MEDICATION RELATED
Gold, mercury, and the heavy metals Penicillamine Nonsteroidal anti-inflammatory drugs, including COX2 inhibitors Lithium Paramethadione, trimethadione Captopril "Street" heroin Others: Probenecid, chlorpropamide, rifampin, tolbutamide, phenindione, pamidronate
ALLERGENS, VENOMS, AND IMMUNIZATIONS
NEOPLASMS
Hodgkin lymphoma and leukemias/lymphomas (with minimal change lesion) Solid tumors (with membranous nephropathy)
HEREDITARY AND METABOLIC DISEASES
Alport syndrome Fabry disease Sickle cell disease Congenital (Finnish type) nephrotic syndrome Familial nephrotic syndrome Nail-patella syndrome Partial lipodystrophy
OTHER
Pregnancy related (includes preeclampsia) Transplant rejection Serum sickness Accelerated hypertensive nephrosclerosis Unilateral renal artery stenosis Massive obesity–sleep apnea Reflux nephropathy

COX = cyclooxygenase; GBM = glomerular basement membrane; HIV = human immunodeficiency virus.

TABLE 121-4 SERUM COMPLEMENT LEVELS IN GLOMERULAR DISEASES

DISEASES WITH A REDUCED COMPLEMENT LEVEL
Post-streptococcal glomerulonephritis
Subacute bacterial endocarditis, visceral abscess, shunt nephritis
Systemic lupus erythematosus
Cryoglobulinemia
Idiopathic membranoproliferative glomerulonephritis
DISEASES ASSOCIATED WITH A NORMAL SERUM COMPLEMENT
Minimal change nephrotic syndrome
Focal segmental glomerulosclerosis
Membranous nephropathy
Immunoglobulin A nephropathy
Henoch-Schönlein purpura
Anti–glomerular basement disease
Pauci-immune rapidly progressive glomerulonephritis (e.g., granulomatous polyangiitis and microscopic polyangiitis)

cholesterol with low or normal high-density lipoprotein (HDL) cholesterol. Lipoprotein(a) levels are elevated as well and normalize with remission of the nephrotic syndrome.

DIAGNOSIS

Initial evaluation of the nephrotic patient includes laboratory tests to define whether the patient has primary, idiopathic nephrotic syndrome (see Table 121-2) or a secondary cause related to a systemic disease, toxin, or medication (see Table 121-3).[1] Common screening tests include the fasting blood sugar and glycosylated hemoglobin tests for diabetes, an antinuclear antibody test for collagen vascular disease, and a serum complement level, which screens for many immune complex–mediated diseases (Table 121-4). In selected patients, cryoglobulins, hepatitis B and C serologies, HIV testing, antineutrophil cytoplasmic antibodies (ANCAs), anti-GMB antibodies, serum protein and immunoelectrophoresis, and other tests may be useful.

After exclusion of secondary causes, a renal biopsy is often required in the adult nephrotic patient. Biopsy results in patients with heavy proteinuria and the nephrotic syndrome are likely to provide a specific diagnosis, determine prognosis, and guide therapy.

TREATMENT Rx

Treatment varies by the specific cause.[2] Elevated lipid levels should be treated (Chapter 206). Anticoagulation is not recommended routinely but is needed if any thrombotic complications occur. It is commonly recommended in patients with additional risk factors for thrombosis (e.g., a prior idiopathic thromboembolic event, immobilization, severe heart failure, or morbid obesity), and may be considered in patients who have membranous nephropathy and are at low risk for bleeding. Edema is treated with a low-salt diet and diuretics. The combination of a loop diuretic plus albumin may be more effective than a loop diuretic alone for some patients who have refractory edema and evidence of intravascular volume depletion. In childhood-onset relapsing or steroid-dependent nephrotic syndrome, rituximab (375 mg/m^2 once weekly for 4 weeks) can provide a median relapse-free interval of about 9 months.

Idiopathic Nephrotic Syndrome

MINIMAL CHANGE DISEASE

Minimal change disease, which is the most common pattern of nephrotic syndrome in children, accounts for only approximately 5 to 10% of cases of idiopathic nephrotic syndrome in adults. A similar histologic pattern may be seen as an adverse reaction to certain medications (nonsteroidal anti-inflammatory drugs [NSAIDs], lithium) or in association with certain tumors (e.g., Hodgkin disease).

CLINICAL MANIFESTATIONS

Patients typically present with weight gain and periorbital and peripheral edema related to the proteinuria. Proteinuria in adult patients can average as much as 10 g/day, and subnephrotic levels are rare. Approximately 30% of adults are hypertensive and 30% have microscopic hematuria. However, an active urinary sediment with RBC casts is not typical of this disease. Many adult patients have mild-to-moderate azotemia, which may be related to hypoalbuminemia and intravascular volume depletion. Complement levels and serologic test results are normal.

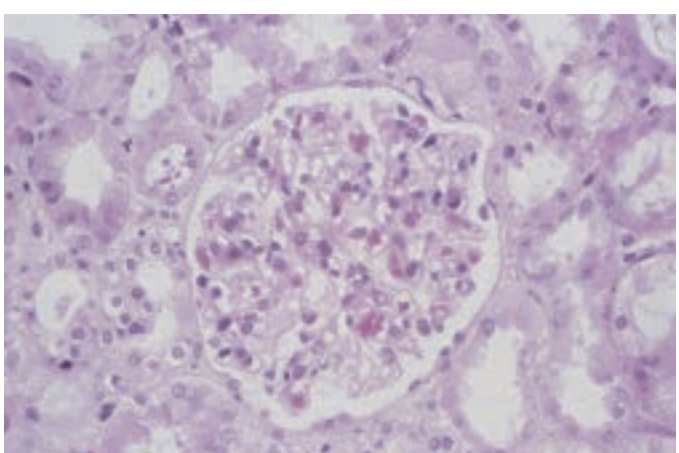

FIGURE 121-1. Unremarkable light microscopic appearance of minimal change disease glomerulopathy. Glomerular basement membranes are thin, and there is no glomerular hypercellularity.

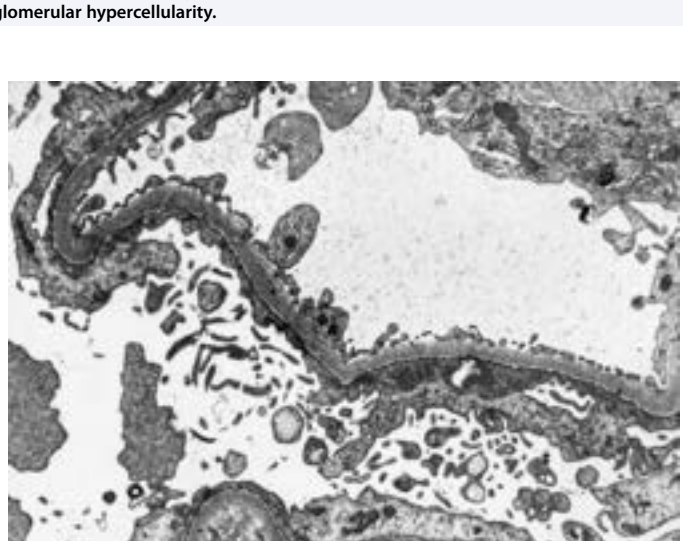

FIGURE 121-2. Minimal change disease. Electron micrograph shows widespread effacement of foot processes with microvillous transformation of the visceral epithelium. No electron-dense deposits are present (uranyl acetate, lead citrate stain; 6000×).

DIAGNOSIS

In true minimal change disease, histopathology typically reveals no glomerular abnormalities on light microscopy (Fig. 121-1). The tubules may show lipid droplet accumulation from absorbed lipoproteins (hence the older term *lipoid nephrosis*). Immunofluorescence staining and electron microscopy (Fig. 121-2) show no immune-type deposits. By electron microscopy, the GBM is normal and effacement or "fusion" of the visceral epithelial foot processes is noted along virtually the entire distribution of every capillary loop.

TREATMENT Rx

The course of minimal change nephrotic syndrome is often one of remissions followed by relapses, which are typically responsive to corticosteroids.[3] When treated with corticosteroids for 8 weeks, 85 to 95% of children experience a remission of proteinuria. In adults, the response rate is somewhat lower, with 75 to 85% of patients responding to regimens of daily prednisone (1 mg/

kg, maximum 80 mg) or alternate-day therapy (2 mg/kg, maximum 120 mg), tapered after 2 months of treatment for a total of 5 to 6 months of therapy. The time to clinical response is slower in adults, and they are not considered steroid resistant until they have failed to respond to 16 weeks of treatment. Tapering of the steroid dose after remission should be gradual over 1 to 2 months. Approximately 40% of adults relapse by 1 year. Most clinicians treat the first relapse similarly to the initial episode. Patients who relapse a third time or who become corticosteroid dependent (unable to decrease the prednisone dose without proteinuria recurring) may be treated with a 2- to 3-month course of the alkylating agent cyclophosphamide at a dose of up to 2 mg/kg/day. Up to 50% of patients will have a remission of at least 5 years, but the response rate is lower in corticosteroid-dependent patients. Other alternative treatments for patients who frequently relapse or are steroid dependent include rituximab (375 mg/m^2 weekly for 4 weeks), low-dose cyclosporine (3 to 5 mg/kg/day for 4 months), tacrolimus (0.05 to 0.1 mg/kg/day), and mycophenolate mofetil (750 to 1000 mg twice daily). All provide approximately equivalent remission and relapse rates, but cyclosporine and tacrolimus are potentially nephrotoxic, so their trough blood levels must be monitored.

PROGNOSIS

The prognosis of minimal change disease is excellent, and most patients who have a decline in kidney function actually have focal segmental glomerulosclerosis on a subsequent kidney biopsy. However, more than 50% of patients have relapses and 10 to 20% may become steroid dependent.

FOCAL SEGMENTAL GLOMERULOSCLEROSIS

Approximately 20 to 25% of adults with idiopathic nephrotic syndrome are found on biopsy to have focal segmental glomerulosclerosis (FSGS).[4] The incidence of FSGS is increasing in all races, but it is especially common in African Americans. FSGS may be either idiopathic or secondary (e.g., associated with heroin abuse, HIV infection, sickle cell disease, obesity, reflux of urine from the bladder to the kidneys, and lesions associated with single or remnant kidneys). FSGS can occur in multiple family members owing to genetic defects in components of the podocyte, which is the visceral epithelial cell. The most common defects include an autosomal recessive pattern caused by mutations in the structural protein podocin; and autosomal dominant forms, caused by mutations in the structural protein α-actinin 4, the TRPC6 glomerular slit diaphragm-associated channel, or INF2, which encodes a formin (actin-regulating protein). The predisposition of African Americans to FSGS is partly related to alleles of the gene apolipoprotein L1.

CLINICAL MANIFESTATIONS

Patients with idiopathic FSGS typically present with either asymptomatic proteinuria or edema. Although two thirds are fully nephrotic at presentation, proteinuria may vary from less than 1 to more than 30 g/day. Hypertension is found in 30 to 50% of patients, and microscopic hematuria occurs in approximately half of patients. The GFR is decreased at presentation in 20 to 30% of patients. Complement levels and other serologic test results are normal.

DIAGNOSIS

By light microscopy, only some glomeruli initially have areas of segmental scarring (Fig. 121-3). As renal function declines, repeat biopsy specimens show more glomeruli with segmental sclerosing lesions and increased numbers of globally sclerotic glomeruli. By immunofluorescence staining, immunoglobulin M (IgM) and C3 commonly are trapped in the areas of glomerular sclerosis. Electron microscopy, however, shows no immune-type deposits and only visceral epithelial cell foot process effacement. The histopathologic variants of FSGS are associated with epidemiologic, clinical, and prognostic differences. For example, the "tip lesion" variant has a relatively benign course, whereas the "collapsing variant" progresses more rapidly to renal failure.

TREATMENT Rx

For primary (idiopathic) FSGS, corticosteroids are used as initial therapy (e.g., prednisone, 1 mg/kg/day, maximum 80 mg, or 2 mg/kg every other day, maximum 120 mg, as tolerated) for a minimum of 3 to 4 months and slowly tapered over the next 3 to 6 months if remission is achieved. A complete or partial remission may be seen in up to 40 to 60% of patients, with preservation of long-term renal function. In patients who relapse after initial therapy or who are steroid resistant, cyclosporine (5 to 6 mg/kg/day) appears to be as good as the combination of pulse dexamethasone (0.9 mg/kg per dose [maximum 40 mg] daily on 2 consecutive days weekly for the first 8 weeks, every other week for weeks 8 to 26, then every 4 weeks until week 50) plus mycophenolate mofetil (25 to 36 mg/kg/day, maximum 2 g/day).■ Tacrolimus (0.05 to 0.1 mg/kg/day) is as good as cyclophosphamide.■ As a result, either cyclosporine or tacrolimus is recommended for 12 to 24 months in steroid-resistant patients who respond, with slow tapering thereafter. With these calcineurin inhibitors, careful monitoring of renal function and serum levels are necessary to avoid nephrotoxicity. Abatacept, a cytotoxic T lymphocyte–associated antigen 4-immunoglobulin fusion protein, has been used in small numbers of resistant patients.

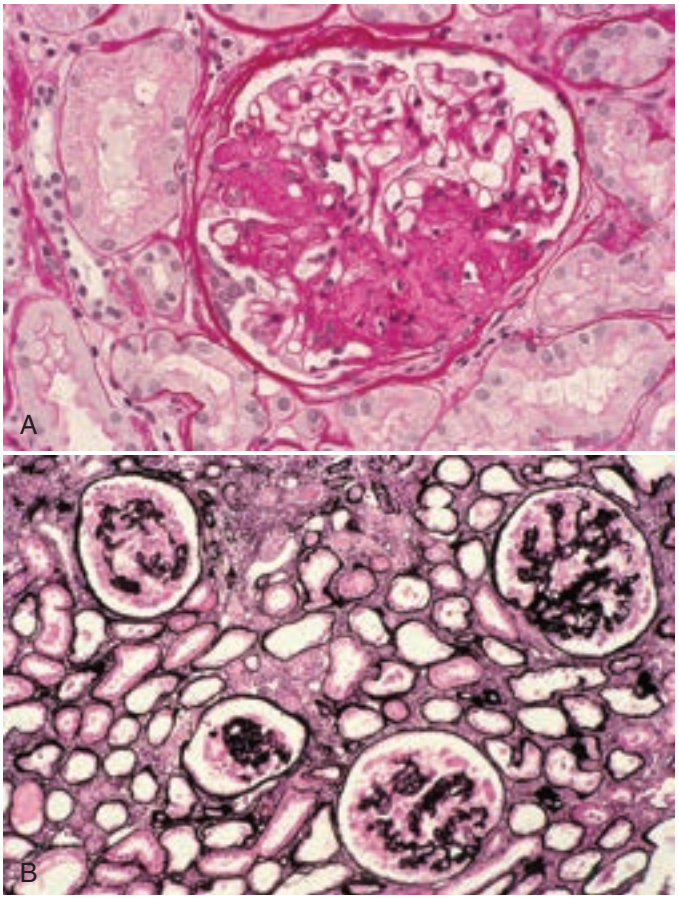

FIGURE 121-3. Focal segmental glomerulosclerosis. A, Light micrograph of classic focal segmental glomerulosclerosis. **B,** Silver stain of collapsing focal segmental glomerulosclerosis showing collapse of the glomerular tufts.

PROGNOSIS

The course of untreated FSGS is usually one of progressive proteinuria and declining GFR. Only a minority of patients experiences a spontaneous remission of proteinuria, and most untreated patients who do not remit eventually develop ESRD within 5 to 20 years. Patients with a sustained remission of their nephrotic syndrome are unlikely to progress to ESRD.

Most patients with genetic forms of FSGS are steroid resistant, have a progressive course, and do not experience recurrences of the FSGS if they receive a renal transplant. Overall, however, FSGS recurs in the transplanted kidney in up to 30% of cases, often in association with elevated levels of a circulating permeability factor. Younger patients, those with a rapid course to renal failure, and those with a prior recurrence are more likely to have allograft recurrence.

MEMBRANOUS NEPHROPATHY

Membranous nephropathy is the most common pattern of idiopathic nephrotic syndrome in white patients. It also may be associated with infections (syphilis [Chapter 319] and hepatitis B and C [Chapter 149]), systemic

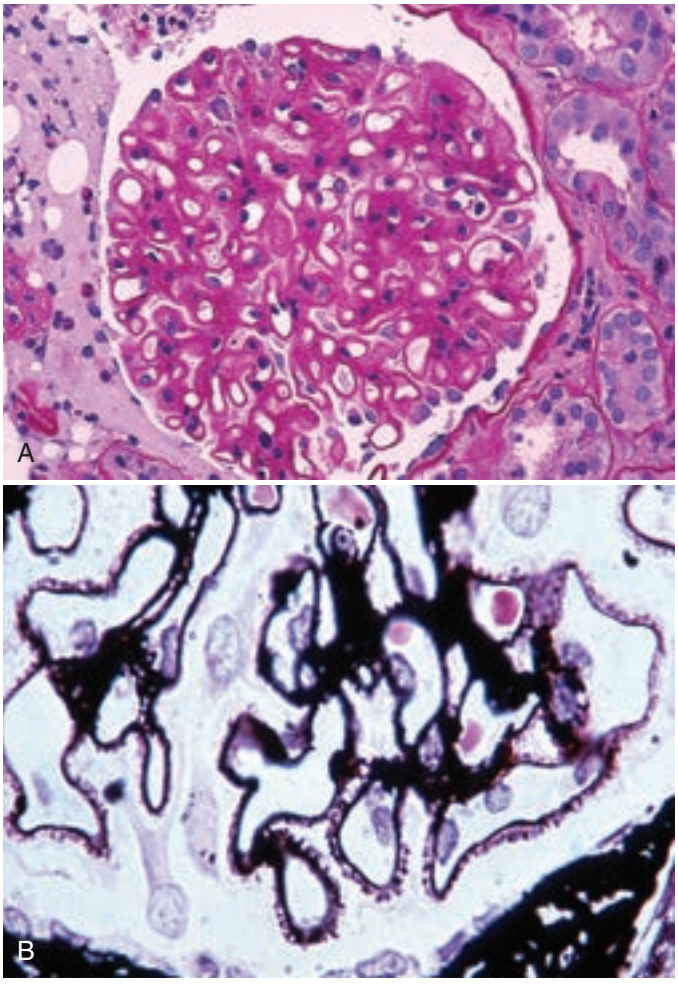

FIGURE 121-4. Membranous nephropathy. A, Light micrograph of membranous nephropathy demonstrating thickening of the glomerular capillary wall but no hypercellularity. **B,** Silver stain of idiopathic membranous nephropathy showing spike formation along the outer aspect of the glomerular basement membrane corresponding to projections of the basement membrane between the epimembranous deposits.

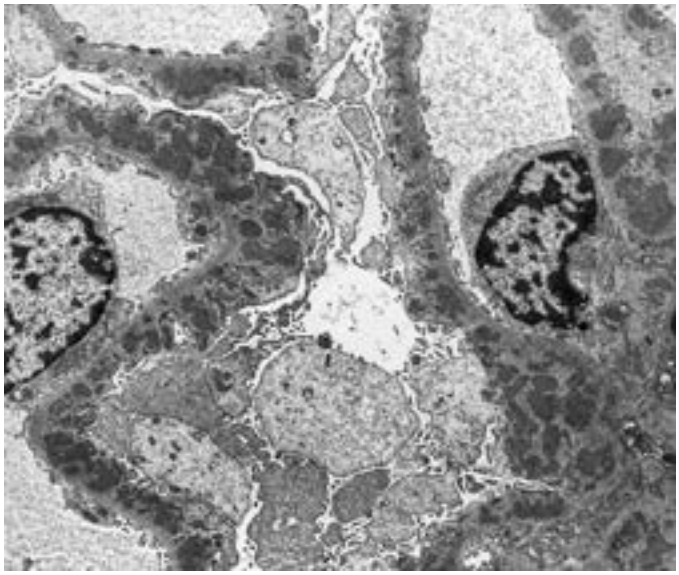

FIGURE 121-5. Membranous glomerulopathy. On ultrastructural examination, there are numerous, closely apposed epimembranous electron-dense deposits separated by basement membrane spikes (uranyl acetate, lead citrate stain; 2500×).

lupus erythematosus (SLE; Chapter 266), medications (gold salts, NSAIDs), and certain tumors (solid tumors and lymphomas). In most patients with idiopathic membranous nephropathy, the antigen in the immune deposits is the M-type phospholipase A_2 receptor, which is present in podocytes. Circulating antibodies combine with this antigen to form in-situ immune complexes on the epithelial side of the basement membrane, thereby leading to proteinuria.

CLINICAL MANIFESTATIONS

Membranous nephropathy typically manifests with proteinuria and edema. Hypertension and microhematuria are not infrequent, but renal function and GFR are usually preserved at presentation. Despite the finding of complement in the glomerular immune deposits, serum complement levels are normal. Membranous nephropathy is the most common pattern of the nephrotic syndrome to be associated with a hypercoagulable state and renal vein thrombosis (Chapter 125). The presence of sudden flank pain, deterioration of renal function, or symptoms of pulmonary disease in a patient with membranous nephropathy should prompt an investigation for renal vein thrombosis and pulmonary emboli.

DIAGNOSIS

On light microscopy, the glomerular capillary loops often appear rigid or thickened (Fig. 121-4), but there is no cellular proliferation. Immunofluorescence staining and electron microscopy show subepithelial immune dense deposits all along the glomerular capillary loops (Fig. 121-5). The presence of circulating antibodies to the M-type phospholipase A_2 receptor are highly sensitive and specific for idiopathic membranous nephropathy.

TREATMENT

Rx

Most studies using corticosteroids alone to treat membranous nephropathy have not shown significant benefit in terms of remission of the nephrotic syndrome or preservation of renal function. By comparison, the combination of corticosteroid therapy (methylprednisolone, 1 g intravenously (IV) daily for 3 days, then prednisone, 0.5 mg/kg/day orally (PO) for 27 days, in months 1, 3, and 5) plus oral cytotoxic therapy (cyclophosphamide, 2 to 2.5 mg/kg/day, for 30 days in months 2, 4, and 6) given in alternating months over 6 months results in more remissions and better preservation of renal function compared with symptomatic therapy.[5] The combination of cyclosporine (3.5 mg/kg/day adjusted to serum levels of 125 to 225 µg/L) plus prednisone (0.15 mg/kg/day up to a maximum of 15 mg/day) for 6 months also is more likely to induce remission of nephrotic syndrome compared with placebo or prednisone alone. Success with tacrolimus has been similar to that with cyclosporine. Other agents used successfully in uncontrolled or small controlled trials in membranous nephropathy include mycophenolate mofetil, adrenocorticotropic hormone, and the monoclonal anti-CD20 antibody, rituximab. In one report, two thirds of 100 consecutive patients treated with rituximab (375 mg/m^2 weekly or subsequent doses based on CD20 count recovery > 5/µL) achieved complete or partial remission, with a median time to remission of 7.1 months. Patients with membranous nephropathy are at increased risk for developing renal vein thrombosis, especially when their serum albumin levels decline. A risk benefit approach to this problem is recommended; however, no clear guidelines exist on when to start anticoagulation.[5]

PROGNOSIS

In most large series, renal survival is more than 75% at 10 years,[6] with a spontaneous remission rate of 20 to 30%. In general, older patients, males, and those with heavy persistent proteinuria are most likely to progress to renal failure and hence to benefit from therapy.

MEMBRANOPROLIFERATIVE GLOMERULONEPHRITIS

Idiopathic membranoproliferative glomerulonephritis was formerly divided into three types (type I, II, and II) based on the location of electron dense deposits. Now, however, the classification is based on immunofluorescence findings into immunoglobulin-mediated and complement-mediated membranoproliferative glomerulonephritis. Immunoglobulin-mediated membranoproliferative glomerulonephritis is associated with immune complex diseases, including systemic lupus erythematosus (Chapter 266), infections such as hepatitis C (Chapter 149), and monoclonal gammopathy (Chapter 187).[7] Complement-mediated membranoproliferative glomerulonephritis includes dense deposit disease and C3 glomerulonephritis. All these stimuli have been proposed to incite the glomerular mesangial cells to grow out along the capillary wall and split the GBM.[8]

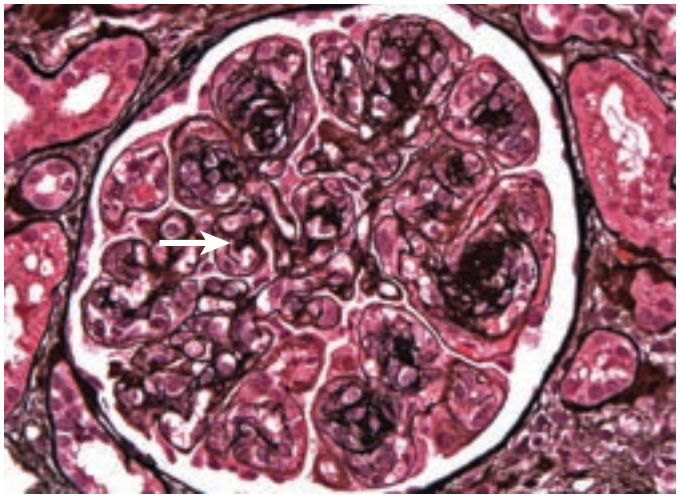

FIGURE 121-6. Membranoproliferative glomerulonephritis with lobulation of the glomerular tuft and splitting of the basement membrane as seen by silver stain.

The entity formerly called type II membranoproliferative glomerulonephritis, or dense deposit disease, and the associated C3 glomerulonephritis are rare glomerular diseases associated with uncontrolled systemic activation of the alternative complement pathway. Many patients have C3 nephritic factor, an autoantibody directed against the C3 convertase of the alternate complement pathway, whereas others have deficiencies of factor H or I or other inhibitors of the alternate complement cascade. Dense deposit disease may be associated with partial lipodystrophy. Patients with alternate complement pathway activation also may have a proliferative glomerulonephritis with electron-dense deposits dissimilar to dense deposit disease. These patients have C3 glomerulonephritis. Indeed, some of the patients originally classified as having membranoproliferative glomerulonephritis type I are found to have a pathogenesis closer to those of dense deposit disease and C3 glomerulonephritis.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Most patients with idiopathic membranoproliferative glomerulonephritis are children or young adults who present with proteinuria or the nephrotic syndrome. A low serum complement level is found intermittently in immunoglobulin-associated membranoproliferative glomerulonephritis, whereas the C3 level is usually reduced in dense deposit disease. The diagnosis of dense deposit disease requires a renal biopsy that shows complement C3 in a characteristic ribbon-like pattern around the capillary loops (Fig. 121-6).

TREATMENT AND PROGNOSIS

Rx

Attempts to treat idiopathic membranoproliferative glomerulonephritis have included corticosteroids, other immunosuppressive medications, anticoagulants, and antiplatelet agents. No therapy has been proved effective in a randomized trial in adults, but corticosteroids have had some success in children. Dense deposit disease and C3 glomerulonephritis have recently been treated with eculizumab, a blocker of the complement system, with variable efficacy. Some may also respond to immunosuppressive therapies.

ACUTE GLOMERULONEPHRITIS AND THE NEPHRITIC SYNDROME

EPIDEMIOLOGY AND PATHOBIOLOGY

Known inciting causes of acute glomerulonephritis include infectious agents, such as type 12 and type 49 "nephritogenic" strains of group A streptococci and endocarditis caused by *Staphylococcus aureus* and *Streptococcus viridans*. Acute glomerulonephritis can also be caused by the deposition of immune complexes in autoimmune diseases such as SLE (Chapter 266) and the damaging effect of circulating antibodies directed against the GBM, as in Goodpasture syndrome.

Invading neutrophils and monocytes, as well as resident glomerular cells, can damage the glomerulus through a number of mediators, including oxidants, chemoattractant agents, proteases, cytokines, and growth factors. Some factors, such as transforming growth factor-β, have been related to eventual glomerulosclerosis and chronic glomerular damage.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients with acute glomerulonephritis often present with a nephritic picture characterized by a decreased GFR, azotemia, oliguria, hypertension, and an active urinary sediment (see Table 121-1). Hypertension is common and is caused by intravascular volume expansion, although renin levels may not be appropriately suppressed for the degree of volume expansion. Patients may note dark, smoky, or cola-colored urine in association with an active urinary sediment, which is composed of erythrocytes, leukocytes, and a variety of casts, including RBC casts. Although many patients with acute glomerulonephritis have proteinuria, sometimes even in the nephrotic range, most have lesser degrees of albuminuria, especially when the GFR is markedly reduced. Regardless of the inciting cause, acute glomerulonephritis is characterized on light microscopy by hypercellularity of the glomerulus, which may be composed of infiltrating inflammatory cells, proliferation of resident glomerular cells, or both.

Immunoglobulin A Nephropathy

EPIDEMIOLOGY AND PATHOBIOLOGY

IgA nephropathy, the most frequent form of idiopathic glomerulonephritis worldwide, represents 15 to 40% of primary glomerulonephritides in parts of Europe and Asia.[9] In geographic areas where renal biopsies are performed for milder urinary findings, a higher incidence of IgA is noted. In the United States, some centers report this diagnosis in up to 20% of all primary glomerulopathies. Males outnumber females, with peak occurrence in the second to third decades of life.

In IgA nephropathy, most patients and their direct blood relatives have elevated circulating levels of IgA molecules that are galactose deficient at the hinge region. The predominant form of IgA is composed of polymeric IgA1. Patients, but not relatives, have circulating IgG and IgA autoantibodies against this galactose-deficient IgA. It is thought that a second "hit" phenomenon (e.g., infection, oxidative stress, etc.) stimulates production of these antibodies, and then the resulting immune complexes deposit in the glomeruli, thereby leading to an inflammatory reaction and consequent glomerulosclerosis and interstitial fibrosis.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

IgA nephropathy often presents either as asymptomatic microscopic hematuria with or without proteinuria (the most common presentation in adults) or as episodic gross hematuria after an upper respiratory tract infection or exercise (the most common presentation in children and young adults).[10] Approximately 20 to 50% of all patients have hypertension. Increased serum IgA levels, noted in one third to half of cases, do not correlate with the course of the disease. Serum complement levels are normal. The diagnosis of IgA nephropathy is established by finding glomerular IgA deposits as either the dominant or the codominant immunoglobulin on immunofluorescence microscopy (Fig. 121-7). Deposits of C3 and IgG also are often found. Light microscopy varies from the most common pattern of mild mesangial proliferation to severe crescentic glomerulonephritis. By electron microscopy, immune-type dense deposits are typically found in the mesangial and paramesangial areas.

TREATMENT

Rx

The pathogenesis of IgA nephropathy may involve abnormal antigenic stimulation of mucosal IgA production and subsequent immune complex deposition in the glomeruli. Efforts to treat the disease by preventing antigenic stimulation, including broad-spectrum antibiotics (e.g., doxycycline), tonsillectomy, and dietary manipulations (e.g., gluten elimination), generally have been controversial or unsuccessful. Trials using fish oils to decrease proteinuria and slow progressive disease have given conflicting results. Controlled trials support the use of angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers (ARBs) (see Table 67-7, in Chapter 67) to decrease proteinuria and protect renal function. Randomized trials suggest that a 6-month course of glucocorticoids (e.g., 6 months of oral prednisone, starting with 0.8 to 1 mg/kg/day for 2 months and then reduced by 0.2 mg/kg/day per month for the next 4 months) reduces both proteinuria and the risk for kidney failure. ACE inhibitors or ARBs with glucocorticoid

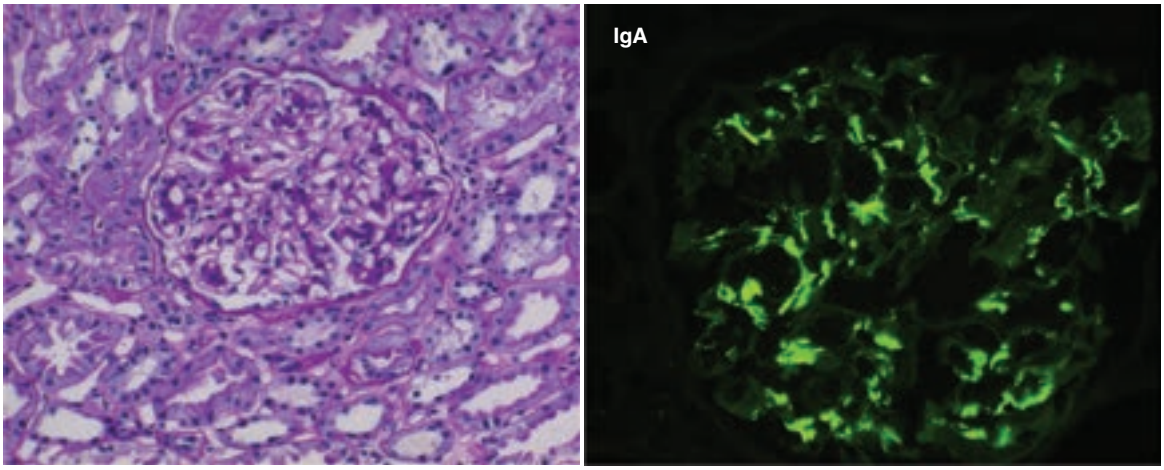

FIGURE 121-7. Immunoglobulin A (IgA) nephropathy with mesangial cell proliferation by light microscopy and IgA deposition on immunofluorescence.

therapy appear to be of incremental benefit. There are conflicting data on the benefit of immunosuppressive agents (e.g., cyclophosphamide, azathioprine, mycophenolate mofetil). For the few patients with crescentic IgA nephropathy, cytotoxic agents have been used.

PROGNOSIS

The course is variable, with some patients showing no decline in GFR over decades and others developing the nephrotic syndrome, hypertension, and renal failure. Factors predictive of a poor outcome in IgA nephropathy include hypertension, persistent proteinuria greater than 1 g/day, male gender, an elevated serum creatinine level, and the histologic features of mesangial and endothelial proliferation and sclerosis or of tubulointerstitial damage and crescent formation. Levels of galactose-deficient IgA and levels of autoantibodies against this galactose-deficient IgA correlate with renal prognosis. Renal survival rates are estimated at 80% to 85% at 10 years and 65% at 20 years. A significant percentage of transplant recipients have a morphologic recurrence in the allograft, but graft loss is uncommon.

Henoch-Schönlein Purpura

Henoch-Schönlein purpura (HSP; Chapter 270) is characterized by a small-vessel vasculitis with arthralgias, skin purpura, and abdominal symptoms, as well as a proliferative acute glomerulonephritis. HSP is predominantly a disease of childhood, although cases occur at all ages. As with IgA nephropathy, patients and their relatives have elevated levels of circulating IgA molecules that are galactose deficient. Renal biopsies are identical to those seen in IgA nephropathy. Some investigators feel IgA nephropathy and HSP are two sides of a spectrum of the same pathogenetic disease. No infectious agent or allergen has been defined as causative, and serum complement levels are normal.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The clinical manifestations of HSP (Chapter 270) include dermatologic, gastrointestinal, rheumatologic, and renal findings. Skin involvement typically starts with a macular rash that coalesces into purpuric lesions (see Fig. 270-2 in Chapter 270) on the ankles, legs, and occasionally arms and buttocks. Gastrointestinal symptoms include cramps, diarrhea, nausea, and vomiting, with melena and bloody diarrhea in the most severely involved cases. Although arthralgias of the knees, wrists, and ankles are common, true arthritis is uncommon. Symptoms of different organ system involvement may occur concurrently or separately, and recurrent episodes during the first year are not uncommon. The renal histopathology of HSP is similar to that of IgA nephropathy. Skin biopsies typically show a small-vessel leukocytoclastic angiitis with immune deposition of IgA.

TREATMENT AND PROGNOSIS

Rx

Like IgA nephropathy, HSP has no proved therapy. Episodes of rash, arthralgias, and abdominal symptoms usually resolve spontaneously. Some patients with severe abdominal findings have been treated with short courses of high doses of corticosteroids. Patients with severe glomerular involvement may benefit by modalities used to treat patients with severe IgA nephropathy, that is, ACE inhibitors and ARBs, and a 6-month course of corticosteroids. Although most patients with HSP recover fully, patients with a more severe nephritic or nephrotic presentation and more severe glomerular damage on renal biopsy have an unfavorable long-term prognosis.

Post-streptococcal Glomerulonephritis

Acute post-streptococcal glomerulonephritis may occur in either an epidemic form or as sporadic cases after infection with nephritogenic strains of group A β-hemolytic streptococci (Chapter 290). Post-streptococcal glomerulonephritis is largely a disease of childhood, but severe disease in adults is well documented. The disease is most common after episodes of pharyngitis, but it can follow streptococcal infections at any site and subclinical cases greatly outnumber clinical cases. Post-streptococcal glomerulonephritis is an acute immune complex disease characterized by the formation of antibodies against streptococci with the localization of immune complexes and complement in the kidney.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Most cases manifest with hematuria, proteinuria, hypertension, and the nephritic syndrome (see Table 121-1) 10 days to several weeks after a streptococcal infection. Throat cultures and skin cultures of suspected sites of streptococcal involvement often are no longer positive for group A β-hemolytic streptococci. A variety of antibodies (e.g., antistreptolysin O [ASLO], antihyaluronidase [AHT]) and a streptozyme panel of antibodies against streptococcal antigens (which includes ASLO, AHT, antistreptokinase, and anti-DNase) often show high titers, but a change in titer over time is more indicative of a recent streptococcal infection. More than 95% of patients with post-streptococcal glomerulonephritis secondary to pharyngitis and 85% of patients with streptococcal skin infections have positive antibody titers. The serum total hemolytic complement levels and C3 levels are decreased in more than 90% of patients during the episode of acute glomerulonephritis.

In a patient with a classic acute nephritic episode after a documented streptococcal infection, with a change in streptococcal antibody titer and a depressed serum complement level, a renal biopsy adds little to the diagnosis. In other cases, a biopsy may be necessary to confirm or refute the diagnosis. On light microscopy (Fig. 121-8), glomeruli are markedly enlarged and often fill Bowman space. Glomeruli exhibit hypercellularity with infiltration of monocytes and polymorphonuclear cells and a proliferation of the glomerular cellular elements. The capillary lumens often are compressed. Some cases demonstrate extracapillary proliferation with crescents. On immunofluorescence microscopy, there is coarse granular deposition of IgG, IgM, and complement, especially C3, along the capillary wall. Electron microscopy shows large dome-shaped, electron-dense subepithelial deposits resembling the humps of a camel at isolated intervals along the GBM.

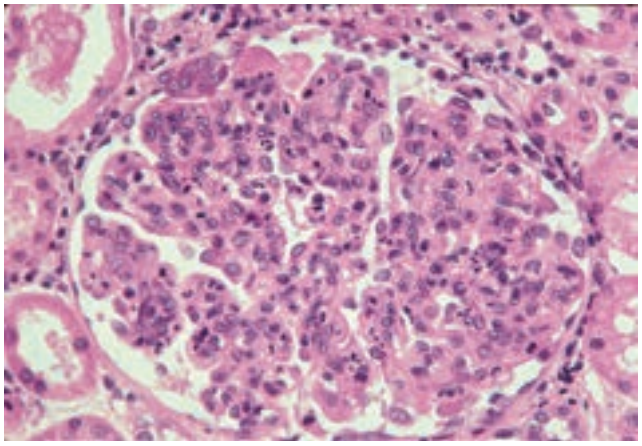

FIGURE 121-8. Post-streptococcal glomerulonephritis with hypercellular glomerulus filling Bowman space and infiltrated by polymorphonuclear and other cells.

TREATMENT AND PROGNOSIS

Rx

Therapy is symptomatic and directed at controlling the hypertension and fluid retention with antihypertensive agents (see Table 67-7 in Chapter 67) and diuretics. In most patients, this is a self-limited disease, with recovery of renal function and disappearance of hypertension in several weeks. However, the presence of underlying renal disease, especially diabetic nephropathy, is associated with a worse prognosis. Proteinuria and hematuria may resolve more slowly over months.

Glomerulonephritis with Endocarditis and Visceral Abscesses

Various glomerular lesions are found in patients with acute and chronic bacterial endocarditis (Chapter 76). Although embolic phenomena can lead to glomerular ischemia and infarcts, a common finding is an immune complex glomerulonephritis. With *S. viridans* endocarditis, both focal and diffuse proliferative glomerulonephritides are common. With the increased incidence of *S. aureus* endocarditis, 40 to 80% of patients have clinical evidence of an immune complex proliferative glomerulonephritis. Glomerulonephritis is now more common with acute than subacute bacterial endocarditis.

Patients often have hematuria and urinary RBC casts, proteinuria ranging from less than 1 g/day to nephrotic levels, and progressive renal failure. Serum total complement and C3 levels are usually reduced. Renal insufficiency may be mild and reversible with appropriate antibiotic therapy, or it may be progressive and lead to dialysis and irreversible renal failure.

A proliferative immune complex glomerulonephritis also can occur in patients with deep visceral bacterial abscesses and infections, such as empyema of the lung (Chapter 99) and osteomyelitis (Chapter 272). Immune complex forms of acute glomerulonephritis also have been noted in patients with bacterial pneumonias, including *Mycoplasma* (Chapter 317), and patients with chronically infected cerebral ventriculoatrial shunts for hydrocephalus. Many of these patients have nephrotic-range proteinuria and only mild renal dysfunction. With appropriate antibiotic therapy, most patients' glomerular lesions heal, and renal function recovers.

RAPIDLY PROGRESSIVE GLOMERULONEPHRITIS

Rapidly progressive glomerulonephritis (RPGN) includes glomerular diseases that progress to renal failure in a matter of weeks to months (Table 121-5). The renal biopsy in all RPGN shows extensive extracapillary proliferation, that is, crescent formation. Patients with primary RPGN can be divided into three patterns as defined by immunologic pathogenesis: those with anti-GBM disease (e.g., Goodpasture syndrome); those with immune complex deposition (e.g., SLE, HSP, post-streptococcal); and those without immune deposits or anti-GBM antibodies (so-called pauci-immune, usually ANCA-positive RPGN).

Anti–Glomerular Basement Membrane Disease

The disease has two peaks of occurrence: in the third decade of life predominantly in men and after 60 years of age predominantly in women. Anti-GBM disease (Table 121-6) is caused by circulating antibodies that are directed against the noncollagenous domain of the α_3 chain of type IV collagen. These antibodies damage the GBM, thereby resulting in an inflammatory response, breaks in the GBM, and the formation of a proliferative and often crescentic glomerulonephritis. If the anti-GBM antibodies cross-react with and damage the basement membrane of pulmonary capillaries, the patient develops pulmonary hemorrhage and hemoptysis, an association called *Goodpasture Syndrome*.

TABLE 121-5 CLASSIFICATION OF RAPIDLY PROGRESSIVE (CRESCENTIC) GLOMERULONEPHRITIS

PRIMARY
Anti–glomerular basement membrane antibody disease, Goodpasture syndrome (with pulmonary disease)
Immune complex mediated
Pauci-immune (usually antineutrophil cytoplasmic antibody-positive)
SECONDARY
Membranoproliferative glomerulonephritis
IgA nephropathy, Henoch-Schönlein purpura
Post-streptococcal glomerulonephritis
Systemic lupus erythematosus

TABLE 121-6 COMMON RENAL DISEASES WITH ASSOCIATED PULMONARY DISEASES

DISEASE	MARKER
Goodpasture syndrome	+Anti–glomerular basement membrane antibodies
Small vessel vasculitis (granulomatous polyangiitis and microscopic polyangiitis)	+Antineutrophil cytoplasmic antibodies
Systemic lupus erythematosus	+Anti-DNA antibodies, low complement
Nephrotic syndrome, renal vein thrombosis, pulmonary embolus	+Lung scan or +CT angiography
Pneumonia with immune complex glomerulonephritis	Low complement, circulating immune complexes
Uremic lung	Elevated creatinine level

CT = computed tomography.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients present with a nephritic picture (see Table 121-1). Renal function may deteriorate from normal to dialysis-requiring levels in a matter of days to weeks. Patients with pulmonary involvement may have life-threatening hemoptysis with dyspnea and with diffuse alveolar infiltrates on chest radiograph. The pathology of anti-GBM disease shows a proliferative glomerulonephritis, often with severe crescentic proliferation in Bowman space. There is linear deposition of immunoglobulin (usually IgG) along the GBM by immunofluorescence (Fig. 121-9), but electron microscopy does not show electron-dense deposits.

Although the treatment of this rare disease has not been studied in large controlled trials, intensive immunosuppressive therapy with cyclophosphamide (e.g., 2 mg/kg/day as tolerated) and corticosteroids (e.g., pulse methylprednisolone, 15 to 30 mg/kg to a maximum of dose of 1000 mg IV daily for three doses, followed by prednisone, 1 mg/kg/day PO to a maximum of 60 to 80 mg/day and slowly tapered after achieving clinical remission) to reduce the production of anti-GBM antibodies, combined with daily plasmapheresis to remove circulating anti-GBM antibodies, has been successful in many patients. Rapid treatment is recommended to prevent irreversible renal damage and is necessary in patients with pulmonary hemorrhage. The optimal duration of therapy is uncertain, but daily plasmapheresis should be performed, preferably until anti-GBM antibody is undetectable, and corticosteroids and cyclophosphamide should be continued until clinical remission is achieved, typically between 3 and 6 months. Patients who already require dialysis at the time of treatment generally do not regain renal function despite aggressive therapy. Relapses of the disease are rare.

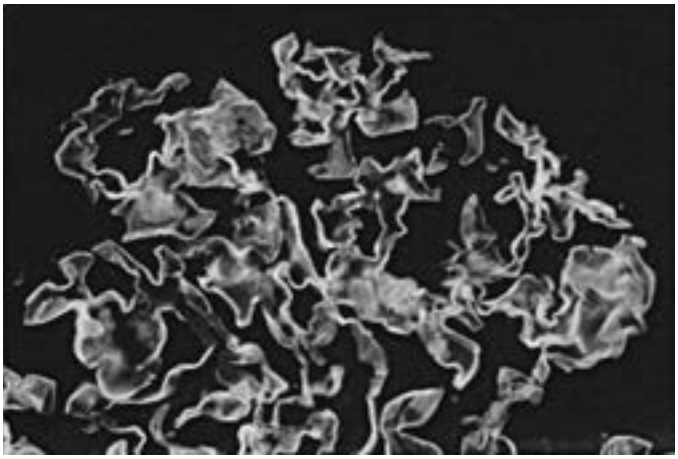

FIGURE 121-9. Anti–glomerular basement membrane (GBM) glomerulonephritis. An immunofluorescence micrograph of a portion of a glomerulus with anti-GBM glomerulonephritis shows linear staining of GBM for immunoglobulin G (IgG) (fluorescein isothiocyanate anti-IgG stain, 600×). (From Falk RJ, Jennette JC, Nachman PH. Primary glomerular disease. In: Brenner BM, ed. *Brenner and Rector's The Kidney.* 7th ed. Philadelphia: Elsevier; 2004.)

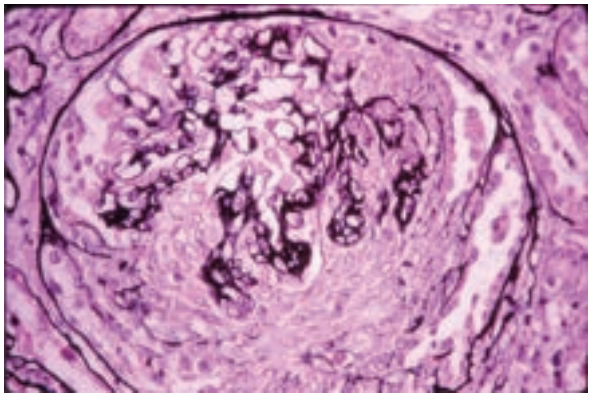

FIGURE 121-10. Crescentic glomerulonephritis typical of both anti–glomerular basement membrane disease and antineutrophil cytoplasmic antibody–positive pauci-immune glomerulonephritis.

Immune Complex Rapidly Progressive Glomerulonephritis

RPGN-associated immune complex–mediated damage to the glomeruli can be seen with idiopathic glomerulopathies, such as IgA nephropathy and idiopathic membranoproliferative glomerulonephritis, or with systemic diseases such as postinfectious glomerulonephritis and SLE. Many cases of crescentic postinfectious glomerulonephritis resolve with successful treatment of the underlying infection. The treatment of severe lupus nephritis is described later.

Pauci-immune and Vasculitis–Associated Rapidly Progressive Glomerulonephritis

Pauci-immune RPGN includes patients with and without evidence of systemic vasculitis. Most patients have circulating ANCAs that are directed against components of neutrophil primary granules. Some patients have granulomatous polyangiitis (formerly called Wegener granulomatosis) with upper and lower respiratory tract involvement by granulomatous angiitis (Chapter 270) along with the pauci-immune glomrulonephritis. Others have microscopic polyangiitis akin to what was formerly called a subgroup of polyarteritis. Finally, others have eosinophilic polyangitis, formerly called Churg-Strauss disease (Chapter 270).

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients often present with progressive renal failure and a nephritic picture (see Table 121-1). Patients with microscopic polyangiitis often have circulating perinuclear ANCA (antibodies directed against granulocyte myeloperoxidase) and a systemic clinical picture (Chapter 270) with arthritis, dermal leukocytoclastic angiitis, pulmonary disease, and constitutional and systemic signs. Patients with granulomatous polyangiitis often have circulating cytoplasmic ANCA (antibodies directed against a granulocyte serine proteinase, anti-PR3), upper and lower respiratory tract involvement by granulomatous angiitis, as well as pauci immune glomrulonephritis (Chapter 270). However, there is considerable overlap among these groups, and some patients have both ANCA and anti-GBM antibodies (Fig. 121-10). Patients with eosinophilic polyangiitis usually have a history of asthma, pulmonary infiltrates, and circulating eosinophila. If they are ANCA positive, they often have associated glomerular disease. Although there is no direct correlation between ANCA titers and disease activity, patients with high titers (especially high anti-PR3 titers) and patients with a major recent increase in titers are more likely to have flares of their disease.

TREATMENT AND PROGNOSIS

Rx

For induction therapy, combination therapy with corticosteroids (e.g., pulse methylprednisolone, 10 to 15 mg/kg/day up to a maximum of 500 to1000 mg IV daily, for 3 days, followed by prednisolone, 1 mg/kg/day PO) and cyclophosphamide (e.g., 15 mg/kg IV every 2 to 3 weeks or 1.5 to 2 mg/kg/day PO), with or without plasmapheresis, has markedly improved renal and patient survival rates in patients with granulomatous polyangiitis and microscopic polyangiitis (Chapter 270). Rituximab, a monoclonal antibody against CD20-positive B cells (375 mg/m^2 weekly for 4 weeks), is as effective and safe as a cyclophosphamide at 6 and 18 months. Methotrexate is as effective as cyclophosphamide in achieving remission, but it leads to a higher relapse rate, so it is not an initial alternative to cyclophosphamide or rituximab in most patients. In severe renal vasculitis, renal survival, but not patient survival, is improved with the addition of plasmapheresis. Maintenance regimens using rituximab (500 mg at months 6, 12, and 18), azathioprine (1.5 mg/kg/day), mycophenolate mofetil (1000 mg twice daily) or methotrexate (20 to 25 mg/week) should be administered for 12 to 18 months after achieving remission. Corticosteroids should be slowly tapered, as determined by the presence of clinical symptoms.

As in all forms of RPGN, renal function may deteriorate rapidly. In pauci-immune RPGN, high-risk patients include older patients, patients with severe pulmonary involvement, and patients with severe renal failure. Analyses have found no difference in prognosis in patients who have systemic vasculitis compared with isolated crescentic RPGN.

GLOMERULAR DISEASES ASSOCIATED WITH GENETIC DEFECTS

As described earlier, some patients with FSGS have abnormalities in genes encoding for podocyte proteins or channels. Other patients, often with a history of clinical renal disease in siblings and other relatives, have other forms of hereditary nephritis.

Alport syndrome is a hereditary form of glomerulonephritis that often manifests with asymptomatic urinary findings. In approximately 85% of cases, it is an X-linked condition with hematuria and proteinuria, often in association with high-pitched hearing loss and abnormalities of the lens of the eye (lenticonus). Most of these patients have a localized mutation in the α_5 chain of type IV collagen (COL4A5). Other families have different patterns of inheritance, more often with mutations in the α_3 and α_4 chains of type IV collagen (COL4A3, COL4A4). Although the light microscopy findings vary from mild mesangial proliferative to advanced sclerosing lesions depending on the stage of biopsy, electron microscopy typically shows areas of GBM thinning and other areas of GBM splitting with lamellations. Some patients with mutations in the collagen IV gene have microhematuria and proteinuria with only areas of extreme GBM thinning on electron microscopy (so-called thin basement membrane disease). In males, Alport syndrome often leads to progressive glomerulosclerosis and ESRD. Use of ACE-inhibitors (Table 67-7) at the onset of proteinuria may slow progression of renal failure.

Fabry disease (Chapter 208), which is caused by an X-linked recessive genetic defect of α-galactosidase, leads to the deposition of ceramide trihexose in the kidneys and other organs. It may cause progressive proteinuria and renal insufficiency in males and in some female carriers. It is associated with telangiectasias of the skin, typically in the bathing suit area, acroparesthesias, cardiac abnormalities, and eye changes. Replacement with intravenous recombinant enzyme agalsidase β is associated with clinical improvement.

Nail-patella syndrome, associated with skeletal and nail deformities, is a rare cause of the nephrotic syndrome. It is due to an autosomal dominant mutation in the LMX1B transcription factor that regulates collagen, nephrin, and podocin gene expression.

OTHER GLOMERULAR DISEASES

Systemic Lupus Erythematosus

The pattern and degree of renal involvement greatly influences the course and therapy of SLE (Chapter 266). Although the incidence of clinical renal disease in SLE varies from 15 to 75%, histologic evidence of renal involvement is found in most biopsy specimens.

The International Society of Nephrology biopsy classification of lupus nephritis can provide a guide to therapy and prognosis (see Table 121-7). In general, class I and II patients have mild lesions that require no therapy directed at the kidney. All patients with class IV lesions (diffuse proliferative lupus nephritis) on biopsy deserve some form of vigorous therapy for their nephritis. Many class III (focal proliferative lupus nephritis) patients, especially those with active necrotizing lesions and large amounts of subendothelial deposits (Fig. 121-11), also benefit from vigorous therapy. For class V (membranous lupus nephritis) patients, the optimal therapy is less clear, and recommendations vary from uniform vigorous treatment to reserving such therapy for patients with serologic activity or more severe nephrotic syndrome.[11]

Induction therapy for severe proliferative lupus nephritis (either active class III or class IV) includes one of two regimens: corticosteroids (prednisone up to 1 mg/kg PO), tapering according to the clinical response over 6 to 12 months) with either daily mycophenolate mofetil (up to 1500 mg PO twice daily) or with cyclophosphamide pulses (0.5 to 1 g/m^2 IV/month for 6 months or 500 mg every 2 weeks for 6 doses) followed by maintenance therapy to prevent flares of disease or progression to renal failure. The addition of plasamapheresis has not been shown to improve outcome. Maintenance therapy with mycophenolate mofetil (1000 mg twice daily) or azathioprine (2 mg/kg/day) is more effective and less toxic than continued intravenous cyclophosphamide therapy after the 6-month induction period.

Rituximab, an anti-CD20 monoclonal antibody, has not proved beneficial in inducing remissions when added to full doses of other immunosuppressive agents in controlled trials of patients with lupus nephritis. It may have a role in refractory or relapsing disease or as a steroid-sparing agent. Other agents under investigation include blockers of T- and B-cell costimulatory molecules, and anti-B LyS therapy.

Many patients with lupus nephritis (40 to 50%) produce autoantibodies against certain phospholipids, including anticardiolipin antibodies. Those patients who experience clotting in the glomeruli and arterioles, require anticoagulation or antiplatelet agents, or both, as well as immunosuppressive medications.

TABLE 121-7 CLASSIFICATION OF LUPUS NEPHRITIS

CLASS	CLINICAL FEATURES
I. Minimal mesangial LN	No renal findings
II. Mesangial proliferative LN	Mild clinical renal disease; minimally active urinary sediment; mild-to-moderate proteinuria (never nephrotic) but may have active serology
III. Focal proliferative LN < 50% glomeruli involved A. Active A/C. Active and chronic C. Chronic	More active sediment changes; often active serology; increased proteinuria (~25% nephrotic); hypertension may be present; some evolve into class IV pattern; active lesions require treatment, chronic do not
IV. Diffuse proliferative LN (>50% glomeruli involved); all may be with segmental or global involvement (S or G) A. Active A/C. Active and chronic C. Chronic	Most severe renal involvement with active sediment, hypertension, heavy proteinuria (frequent nephrotic syndrome), often reduced glomerular filtration rate; serology very active. Active lesions require treatment
V. Membranous LN glomerulonephritis	Significant proteinuria (often nephrotic) with less active lupus serology
VI. Advanced sclerosing LN	More than 90% glomerulosclerosis; no treatment prevents renal failure

LN = lupus nephritis.

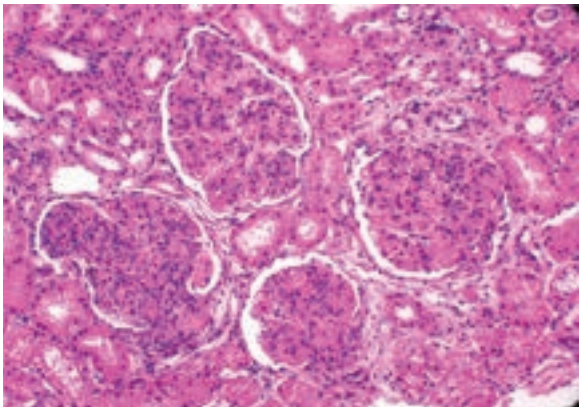

FIGURE 121-11 Diffuse proliferative lupus nephritis with involvement of all glomeruli.

Diabetes Mellitus

Diabetic nephropathy, which is the most common form of glomerular damage seen in developed countries, is discussed in detail in Chapter 124.

Amyloidosis

Renal amyloid deposits, whether due to AL, AA, or hereditary-genetic forms of amyloid, are predominantly found within the glomeruli, where they often appear as amorphous eosinophilic extracellular nodules (Chapter 188). All amyloid is due to fibril formation of proteins that have a tendency to conform into β-pleated sheets. All stain positively with Congo red and display apple-green birefringence under polarized light. By electron microscopy, amyloid appears as nonbranching rigid fibrils 8 to 10 nm in diameter. In AL amyloid, overproduction of an abnormal light chain (λ 80%, κ 20%) can be detected by immunofluorescence staining for only λ or only κ light chains. Most patients have a clonal proliferation of plasma cells that typically does not reach levels seen with symptomatic multiple myeloma. In AA amyloid, antisera to the AA protein stain the glomeruli. Some patients with renal amyloidosis have genetically abnormal forms of proteins such as transthyretin, LECT2, and lipoproteins that lead to fibrillogenesis and deposition of amyloid fibrils in the kidney.

Almost 80% of patients with AL amyloid have renal disease. Renal manifestations include albuminuria and renal insufficiency. Approximately 25% of these patients present with nephrotic syndrome, which eventually is diagnosed in up to half of patients. Extrarenal amyloid involvement may cause cardiac disease (Chapter 60) or neuropathy (Chapter 420). Diagnosis may be made from organ biopsy other than the kidney (e.g., myocardial, gingival, rectal, or fat pad biopsy). Treatment strategies for renal AL amyloidosis are similar to those for multiple myeloma and other plasma cell dyscrasias (Chapter 187).

AA amyloid is usually associated with chronic inflammatory conditions such as rheumatoid arthritis (Chapter 264), familial Mediterranean fever (Chapter 261), inflammatory bowel disease (Chapter 141), osteomyelitis (Chapter 272), and other chronic infections. Treatment is directed at the underlying inflammatory process. Specific therapy against the primary inflammatory disease (e.g., anti–tumor necrosis factor therapy in rheumatoid arthritis and colchicine in familial Mediterranean fever) can prevent fibrillogenesis in AA amyloid patients. Eprodisate, a compound that inhibits polymerization and deposition of amyloid fibrils, can slow the progression of renal disease in patients with AA amyloidosis.

Light-Chain Deposition Disease

Light-chain deposition disease, like AL amyloidosis, is a systemic disease caused by the overproduction and extracellular deposition of a monoclonal immunoglobulin light chain (Chapter 187). However, the deposits do not form β-pleated sheets, do not stain with Congo red, and are granular rather than fibrillar. Most patients with light chain deposition disease have a lymphoplasmacytic B-cell disease similar to multiple myeloma (Chapter 187).

Albuminuria is common, and the nephrotic syndrome is found in half of patients at presentation, often accompanied by hypertension and renal insufficiency. On light microscopy, most glomeruli contain eosinophilic mesangial glomerular nodules. Some biopsy samples show associated light-chain cast nephropathy with eosinophilic laminated and fracturing casts obstructing the tubules, as seen in myeloma. By immunofluorescence, a single class of immunoglobulin light-chain (κ in 80% of cases) stains in a diffuse linear pattern along the GBM, in the nodules and along the tubular basement membranes.

The treatment for most patients with light-chain deposition disease is chemotherapy similar to that for myeloma (Chapter 187).

Fibrillary Glomerulopathy–Immunotactoid Glomerulopathy

Some patients with renal disease have glomerular lesions with deposits of nonamyloid fibrillar proteins ranging in size from 12 to more than 50 nm. Patients with these lesions have been divided into two groups: those with fibrillary glomerulonephritis with fibrils of 20 nm in diameter and those with immunotactoid glomerulonephritis, a much rarer disease often associated with lymphoproliferative disorders, in which the fibrils are much larger (30 to 50 nm). Proteinuria is found in almost all patients, and hematuria, the nephrotic syndrome, and renal insufficiency eventually develop in most. There is no proved therapy for fibrillary glomerulopathy. In patients with immunotactoid glomerulopathy, a search for a treatable underlying B cell disorder is important.

Human Immunodeficiency Virus–Associated Nephropathy

Infection with human immunodeficiency virus (HIV) (Chapter 386) is associated with a number of patterns of renal disease, including acute kidney injury and a unique form of glomerulopathy now called HIV-associated nephropathy.[12] HIV-associated nephropathy is characterized by heavy proteinuria and rapid progression to renal failure. On light microscopy, biopsies show global collapse of the glomerular tufts, severe tubulointerstitial changes with interstitial inflammation, edema, microcystic dilation of tubules, and severe tubular degenerative changes. On electron microscopy, tubuloreticular inclusions can be seen in the glomerular endothelium. The use of ACE inhibitors or ARBs (see Table 67-7 in Chapter 67) and antiretroviral therapy may slow the progression to renal failure and decrease proteinuria. Corticosteroids (e.g., prednisone, 1 mg/kg for 1 month followed by a taper over several months) may be beneficial in selected patients with HIV-associated nephropathy.

Mixed Cryoglobulinemia

Cryoglobulinemia (Chapter 187) is caused by the production of circulating immunoglobulins that precipitate on cooling and resolubilize on warming. Cryoglobulinemia may be found in association with infections, collagen vascular disease, and lymphoproliferative diseases such as multiple myeloma and Waldenström macroglobulinemia. Many patients with what was originally described as glomerulonephritis resulting from essential mixed cryoglobulinemia have been found to have hepatitis C–associated renal disease. Some patients develop an acute nephritic picture with acute renal insufficiency. Most patients have proteinuria, and approximately 20% present with the nephrotic syndrome. Most patients with renal disease have a slow, indolent course characterized by proteinuria, hypertension, hematuria, and renal insufficiency. Hypocomplementemia, especially of the early components C1q to C4, is a characteristic finding in cryoglobulinemic glomerulonephritis, whether hepatitis C–related or idiopathic. Treatment of hepatitis C–associated cryoglobulinemia includes antiviral therapy (Chapter 149). When significant renal disease is present, various combinations of corticosteroids with or without rituximab or cyclophosphamide or plasmapheresis have been used.

Thrombotic Microangiopathies

A number of systemic diseases, including hemolytic-uremic syndrome (Chapters 125 and 172), thrombotic thrombocytopenic purpura (TTP; Chapter 172), and the antiphospholipid syndrome (Chapter 176), as well as microangiopathy associated with drugs such as mitomycin and cyclosporine, are characterized by microthromboses of the glomerular capillaries and small arterioles. The renal findings may be dominant or only part of a more generalized picture of microangiopathy.

Renal manifestations of the thrombotic microangiopathies may include gross or microscopic hematuria, proteinuria that is typically less than 2 g/day but may reach nephrotic levels, and renal insufficiency. Patients may have oliguric or nonoliguric acute kidney injury. The histologic findings in all of the microangiopathies resemble each other and include glomerular capillary thromboses, areas of ischemic damage, and intimal proliferation with luminal narrowing by thrombi of arterioles and small arteries. In all thrombotic microangiopathies, treatment includes correcting hypovolemia, controlling hypertension, and the use of dialytic support for those with severe renal failure. In TTP associated with an acquired or hereditary deficiency of the von Willebrand convertase ADAMTS-13 and in some other cases, plasmapheresis with fresh-frozen plasma is beneficial (Chapter 172). In some patients whose hemolytic-uremic syndrome is not associated with Shiga toxin, defects in the alternate complement system are found. These patients with atypical hemolytic-uremic syndrome may benefit from a monoclonal blocker of the fifth component of complement, eculizumab (Chapter 172). In the antiphospholipid syndrome, anticoagulation with heparin and then warfarin is useful (Chapter 176).

Grade A References

A1. Dharmaraj R, Hari P, Bagga A. Randomized cross-over trial comparing albumin and furosemide infusions in nephrotic syndrome. *Pediatr Nephrol*. 2009;24:775-782.

A2. Iijima K, Sako M, Nozu K, et al. Rituximab for childhood-onset, complicated, frequently relapsing nephrotic syndrome or steroid-dependent nephrotic syndrome: a multicentre, double-blind, randomised, placebo-controlled trial. *Lancet*. 2014;384:1273-1281.

A3. Gipson DS, Trachtman H, Kaskel FJ, et al. Clinical trial of focal segmental glomerulosclerosis in children and young adults. *Kidney Int*. 2011;80:868-878.

A4. Ren H, Shen P, Li X, et al. Tacrolimus versus cyclophosphamide in steroid-dependent or steroid-resistant focal segmental glomerulosclerosis: a randomized controlled trial. *Am J Nephrol*. 2013;37:84-90.

A5. Chen Y, Schieppati A, Cai G, et al. Immunosuppression for membranous nephropathy: a systematic review and meta-analysis of 36 clinical trials. *Clin J Am Soc Nephrol*. 2013;8:787-796.

A6. Cheng J, Zhang W, Zhang XH, et al. ACEI/ARB therapy for IgA nephropathy: a meta analysis of randomised controlled trials. *Int J Clin Pract*. 2009;63:880-888.

A7. Lv J, Xu D, Perkovic V, et al. Corticosteroid therapy in IgA nephropathy. *J Am Soc Nephrol*. 2012;23:1108-1116.

A8. Specks U, Merkel PA, Seo P, et al. Efficacy of remission-induction regimens for ANCA-associated vasculitis. *N Engl J Med*. 2013;369:417-427.

A9. Jayne DR, Gaskin G, Rasmussen N, et al. Randomized trial of plasma exchange or high-dosage methylprednisolone as adjunctive therapy for severe renal vasculitis. *J Am Soc Nephrol*. 2007;18:2180-2188.

A10. Guillevin L, Pagnoux C, Karras A, et al. Rituximab versus azathioprine for maintenance in ANCA-associated vasculitis. *N Engl J Med*. 2014;371:1771-1780.

A11. Henderson LK, Masson P, Craig JC, et al. Induction and maintenance treatment of proliferative lupus nephritis: a meta-analysis of randomized controlled trials. *Am J Kidney Dis*. 2013;61:74-87.

A12. Dooley MA, Jayne D, Ginzler EM, et al. Mycophenolate versus azathioprine as maintenance therapy for lupus nephritis. *N Engl J Med*. 2011;365:1886-1895.

A13. Dember LM, Hawkins PN, Hazenberg BP, et al. Eprodisate for the treatment of renal disease in AA amyloidosis. *N Engl J Med*. 2007;356:2349-2360.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hogan J, Mohan P, Appel GB. Diagnostic tests and treatment options in glomerular disease: 2014 update. *Am J Kidney Dis*. 2014;63:656-666.
2. Cybulsky AV, Walsh M, Knoll G, et al. Canadian Society of Nephrology Commentary on the 2012 KDIGO Clinical Practice Guideline for Glomerulonephritis: Management of Glomerulonephritis in Adults. *Am J Kidney Dis*. 2014;63:363-377.
3. Hogan J, Radhakrishnan J. The treatment of minimal change disease in adults. *J Am Soc Nephrol*. 2013;24:702-711.
4. D'Agati VD, Kaskel FJ, Falk RJ. Focal segmental glomerulosclerosis. *N Engl J Med*. 2011;365:2398-2411.
5. Lee T, Biddle AK, Lionaki S, et al. Personalized prophylactic anticoagulation decision analysis in patients with membranous nephropathy. *Kidney Int*. 2014;85:1412-1420.
6. van den Brand JA, van Dijk PR, Hofstra JM, et al. Long-term outcomes in idiopathic membranous nephropathy using a restrictive treatment strategy. *J Am Soc Nephrol*. 2014;25:150-158.
7. Zand L, Fervenza FC, Nasr SH, et al. Membranoproliferative glomerulonephritis associated with autoimmune diseases. *J Nephrol*. 2014;27:165-171.
8. Sethi S, Fervenza FC. Membranoproliferative glomerulonephritis: a new look at an old entity. *N Engl J Med*. 2012;366:1119-1131.
9. Yu HH, Chiang BL. Diagnosis and classification of IgA nephropathy. *Autoimmun Rev*. 2014;13:556-559.
10. Wyatt RJ, Julian BA. IgA nephropathy. *N Engl J Med*. 2013;368:2402-2414.
11. Hogan J, Appel GB. Update on the treatment of lupus nephritis. *Curr Opin Nephrol Hypertens*. 2013;22:224-230.
12. Hilton R. Human immunodeficiency virus infection and kidney disease. *J R Coll Physicians Edinb*. 2013;43:236-239.

REVIEW QUESTIONS

1. A 24-year-old Asian man presents to his family physician with one episode of painless gross hematuria. He is otherwise healthy with no significant family history. On physical examination, his blood pressure is normal and there are no other abnormalities. Laboratory examination shows a normal metabolic profile, hemogram, liver enzymes, and renal function tests, as well as a negative urine culture. Serologic tests including complement levels, antinuclear antibody, antineutrophil cytoplasmic antibody, and a hepatitis panel are all negative. Urine examinations show 3+ blood, 3+ protein, and many RBCs and RBC casts on microscopic examination. His renal ultrasound is normal. A repeat physical examination and laboratory tests after 1 week show 1+ heme, 5 to 10 RBCs, and trace urinary protein, which quantitates to 150 mg/day. Your recommendation at this point should be:

A. Continue observation.
B. Renal biopsy.
C. Empiric therapy with alternate day steroids.
D. Genetic testing for patient and his family.
E. Urology referral.

Answer: A The presence of gross hematuria associated with RBC and RBC casts and proteinuria point to a glomerular origin as opposed to a urologic origin (hence a urologic consult is not appropriate). In this patient with an Asian ancestry, negative serologies (particularly normal complement levels), and a self-limiting course, an IgA nephropathy is the most likely diagnosis. A renal biopsy is not appropriate at this time because he has no high-risk features for progressive disease, such as proteinuria greater than 1 g/day or impaired renal function. Likewise, empiric treatment is not warranted with these low-risk features. Although hereditary nephritis may present like this, an insidious progression to renal failure without episodes of gross hematuria is more typical, and the absence of a family history makes this entity less likely.

Wyatt RJ, Julian BA. IgA nephropathy. *N Engl J Med.* 2013;368: 2402-2414.

2. A 50-year-old man is admitted with shortness of breath, 1 week of hemoptysis, and mild swelling of his lower extremities. He also notes a 1-month history of myalgia and arthralgia, partially relieved by ibuprofen, which he takes twice daily. On examination, his blood pressure is 150/90 mm Hg, he is afebrile, and his lung examination shows bibasilar crackles. Laboratory tests reveal mild leukocytosis with a neutrophilic preponderance and hemoglobin of 11.2 g/dL. His urine shows 2+ protein, many RBCs, and a few RBC casts. His creatinine is 2.4 mg/dL. His chest radiograph is notable for bilateral diffuse infiltrates, which worsen in the next 24 hours. Serologic tests are pending. Your management at this point would include:

A. Continue observation until serologic tests are available
B. Arrange for an urgent renal biopsy
C. Choice B + intravenous steroids
D. Choice B + intravenous steroids + plasmapheresis

Answer: D This patient has the pulmonary renal syndrome with a nephritic urine sediment. The possibilities are immune complex disease, anti-GBM antibody disease (Goodpasture syndrome), and small vessel vasculitis. The hemoptysis and worsening pulmonary infiltrate require empirical therapy with high-dose steroids and plasmapheresis while awaiting renal biopsy and serologic results to confirm the diagnosis. Plasmapheresis is indicated in view of the pulmonary hemorrhage. Cyclophosphamide (or rituximab in small vessel vasculitis) is administered after confirming the diagnosis.

Bolton WK. Pulmonary renal syndrome and emergency therapy. *Contrib Nephrol.* 2010;165:166-173. Plasmapheresis therapy for diffuse alveolar hemorrhage in patients with small-vessel vasculitis.

Klemmer PJ, Chalermskulrat W, Reif MS, et al. Plasmapheresis therapy for diffuse alveolar hemorrhage in patients with small-vessel vasculitis. *Am J Kidney Dis.* 2003;42:1149-1153.

3. A 65-year-old white woman with nephrotic syndrome (urine protein 6 g/day, normal renal function, serum albumin 2.2 g/dL) recently has been diagnosed with membranous nephropathy. All of the following statements about membranous nephropathy are true *except:*

A. Further investigations should include a malignancy workup.
B. The patient should start on cyclic cyclophosphamide alternating with steroids.
C. There is a significant risk for thrombotic events, and prophylactic anticoagulation may be considered in patients at low risk for bleeding.
D. She should be screened for hepatitis B, which is known to be associated with membranous nephropathy.
E. She is at risk for progressive renal disease if proteinuria remains at the same level after 6 months.

Answer: B Patients with membranous nephropathy should be evaluated for secondary causes, including malignancies, infections (e.g., hepatitis viruses), systemic lupus, and medications. The risk for thrombotic events is significant and rises when the serum albumin level is less than 2.8 g/dL; some authorities recommend prophylactic anticoagulation in such patients if they are at low risk for bleeding. The risk for progressive renal disease is associated with persistent high-grade proteinuria. However, approximately 30% of patients undergo spontaneous remission, so observation for 3 to 6 months is recommended before considering specific immunosuppression.

Waldman M, Austin HA, 3rd. Treatment of idiopathic membranous nephropathy. *J Am Soc Nephrol.* 2012;23:1617-1630.

4. A 71-year-old man with controlled hypertension has been experiencing progressive shortness of breath and lower extremity edema for 1 month. He has just been discharged after being admitted with a diagnosis of heart failure with preserved ejection fraction. On examination, his blood pressure is 110/70 mm Hg supine but drops to 90/60 mm Hg on standing with no increase in his heart rate. His liver is enlarged 5 cm below the costal margin. His lung examination shows diminished breath sounds over the bases. Diagnostic tests done during his hospitalization show a normal hemogram, a serum albumin 2.5 g/dL, serum creatinine 1.5 mg/dL, and urine protein 8 g/day with no cells in the urine sediment. Further diagnostic evaluation shows normal serologies (including complement levels), an elevated λ light chain, and a decreased κ/λ ratio. A renal biopsy shows mesangial expansion with an amorphous material on light microscopy. All of the following statements are true about this patient's disease, *except:*

A. The mesangial deposits are Congo red positive with 10-nm fibrils on electron microscopy.
B. Treatment directed against the abnormal plasma cell clone may help stabilize his disease course.
C. Bone marrow biopsy typically will not show features of multiple myeloma.
D. There is usually an underlying inflammatory condition as the basis of this disease.
E. Fat pad biopsy may obviate the need for a renal biopsy.

Answer: D This patient with nephrotic-range proteinuria, cardiac failure with preserved ejection fraction, and autonomic neuropathy likely has AL amyloid, which on biopsy (renal, cardiac, fat pad, or gingival) will show the typical deposits that stain apple green on Congo red staining under a polarizing microscope. Treatment is similar to that for myeloma even though the bone marrow does not show evidence for myeloma. The elevated λ light-chain levels are a clue for AL amyloid; definitive diagnosis is based on staining the biopsy material for light chains, which will show the monoclonal nature of the deposit. The diagnosis of AA amyloid is made when the Congo red stain shows the typical birefringence but the deposits stain for AA protein.

Desport E, Bridoux F, Sirac C, et al. AL amyloidosis. *Orphanet J Rare Dis.* 2012;7:54.

122 TUBULOINTERSTITIAL NEPHRITIS

ERIC G. NEILSON

DEFINITION

Interstitial nephritis can be primary and begin in the tubulointerstitium or appear as a secondary event and spread from blood vessels, including the glomerular capillaries. Injury to the tubulointerstitial compartment can be the result of autoimmunity, toxic insult, infection, or exposure to drugs. In all cases, however, the inflammatory process has an immunologic component that leads to the release of tissue cytokines, which attract T lymphocytes and other monocytes, which eventually convert tubular epithelia into fibroblasts to produce fibrosis.

Tubulointerstitial nephritis can be arbitrarily divided into acute and chronic types. The acute form of interstitial nephritis often begins abruptly. When inciting events subside, so does the nephritis, and the glomerular filtration rate tends to normalize, with little residual damage except in patients with preexisting disease. Chronic interstitial nephritis is persistent and over time reduces the number of functioning nephrons by encasing and dismantling them with irreversible fibrosis. So-called toxic nephropathy is similar to this form of nephritis. Sometimes acute and chronic injury is difficult to distinguish because global destruction of the tubulointerstitium can occur within a matter of weeks.

EPIDEMIOLOGY

Acute interstitial nephritis appears unexpectedly in otherwise healthy individuals from a variety of causes. Approximately 1% of patients with hematuria and proteinuria have acute interstitial nephritis, and it is seen in 1 to 15% of

TABLE 122-1 CAUSES OF ACUTE INTERSTITIAL NEPHRITIS

DRUGS
Antibiotics
Penicillins
Rifampin, ethambutol
Sulfa
Vancomycin
Ciprofloxacin
Cephalosporins
Erythromycin
Trimethoprim-sulfamethoxazole
Acyclovir
Nonsteroidal Anti-Inflammatory Drugs
Selective and Nonselective
Cyclooxygenase-2 Inhibitors
Diuretics
Thiazides
Furosemide
Triamterene
Miscellaneous
Captopril
Ranitidine
Omeprazole
Phenobarbital
Phenytoin
Sodium valproate
Carbamazepine
Allopurinol
Interferon
Interleukin-2
All-*trans*-retinoic acid
INFECTIONS
Bacteria
Legionella
Brucella
Diphtheria
Streptococcus
Staphylococcus
Yersinia
Salmonella
Escherichia coli
Campylobacter
Viruses
Epstein-Barr virus
Cytomegalovirus
Hantavirus
Herpes simplex virus
Hepatitis B virus
Other
Mycoplasma
Rickettsia
Leptospira
Mycobacterium tuberculosis
Schistosoma mekongi
Toxoplasma
Chlamydia
AUTOIMMUNE DISEASES
Anti–tubular basement membrane disease
Tubulointerstitial nephritis and uveitis (TINU) syndrome
Kawasaki disease

TABLE 122-2 CAUSES OF CHRONIC INTERSTITIAL NEPHRITIS

HEREDITARY DISEASES
Mitochondrial mutations
METABOLIC DISTURBANCES
Hypercalcemia, nephrocalcinosis
Hyperoxaluria
Hypokalemia
Hyperuricemia
Cystinosis
Methylmalonic acidemia
DRUGS AND TOXINS
Analgesics
Cadmium
Lead
Health food botanicals, herbs
Lithium
Cyclosporine, tacrolimus
Cisplatin, methotrexate
Nitrosoureas
AUTOIMMUNE DISEASES
Renal allograft rejection
Granulomatosis with polyangiitis
Immunoglobulin G4–related tubulointerstitial nephropathy
Sjögren syndrome
Systemic lupus erythematosus, vasculitis
Tubulointerstitial nephritis and uveitis (TINU) syndrome
Sarcoidosis
HEMATOLOGIC DISTURBANCES
Multiple myeloma, light chains
Lymphoma
Sickle cell disease
INFECTIONS
Complicated pyelonephritis
Human immunodeficiency virus (HIV)
Epstein-Barr virus
Malakoplakia
Xanthogranulomatous pyelonephritis
OBSTRUCTIVE NEPHROPATHY
Tumors
Stones
Outlet obstruction
Vesicoureteral reflux
MISCELLANEOUS
Age-related vascular disease
Hypertension
Ischemia
Balkan (endemic) nephropathy
Radiation nephritis

autopsy series. In recent decades its prevalence has increased in individuals over 65 years of age, perhaps owing to their more frequent exposure to prescription drugs.[1]

Although acute interstitial nephritis is largely due to the use of pharmaceuticals,[2] other important causes include infection and idiopathic autoimmune diseases (Table 122-1). Penicillin moieties (less so nafcillin and piperacillin), cephalosporins, sulfa-like drugs, and nonsteroidal anti-inflammatory drugs (NSAIDs) top the list. NSAIDs cause both acute interstitial nephritis and chronic analgesic nephropathy. Diphtheria in children (Chapter 292), legionellosis (Chapter 314), leptospirosis (Chapter 323), histoplasmosis (Chapter 332), tuberculosis (Chapter 324), and DNA viruses such as cytomegalovirus (Chapter 376) and Epstein-Barr virus (Chapter 377) are well-recognized agents of acute interstitial nephritis. Anti–tubular basement membrane disease is a rare cause of autoimmune interstitial nephritis. Although sarcoidosis or the tubulointerstitial nephritis and uveitis (TINU) syndrome can manifest as acute interstitial nephritis on biopsy, they often quickly evolve into chronic disease.

All forms of injury to the kidney, regardless of origin, progress to end-stage renal disease through a terminal phase of chronic interstitial nephritis. In addition to glomerulonephritides (Chapter 121), cystic diseases (Chapter 127), and diabetes (Chapter 229), a wide variety of renal conditions start slowly in the tubulointerstitium and often go unrecognized until late in the course, when biopsy shows chronic interstitial nephritis.

Primary chronic interstitial nephritis can be caused by a variety of toxic, metabolic, hematologic, obstructive, and infectious processes. Ingestion of six or more tablets per day of acetaminophen, aspirin, or NSAIDs, alone or together, for at least 3 years puts patients at risk for analgesic nephropathy. A careful history of drug or toxin exposure, previous renal images, and a family history often point to a probable diagnosis (Table 122-2).

PATHOBIOLOGY

Pathophysiology

Regardless of the origin of renal inflammation, the kidneys do not fail until interstitial nephritis, fibrosis, and tubular atrophy develop. Interstitial

nephritis is the pathologic equivalent of clinical progression because it is the final common pathway to permanent tissue damage.[3] The degree of reduction in the glomerular filtration rate correlates with the degree of interstitial injury. Urinary flow is impeded by tubular obstruction. Increased vascular resistance causes progressive tubular injury and fibrosis. A net reduction in the cross-sectional area of peritubular vessels increases postglomerular resistance to the extent that the compensatory increase in glomerular hydrostatic pressure cannot fully restore filtration to normal levels. Tubuloglomerular feedback assumes increasing importance in the transition from acute to chronic glomerulonephritis when autoregulation of renal blood flow is disrupted by tubulointerstitial fibrosis. Loss of autoregulation by tubuloglomerular feedback results from the absence or insensitivity of the afferent arteriole. Perhaps more significant is the effect of interstitial pressure on the sensitivity of the feedback mechanism. Tubular atrophy may disrupt the normal renal osmotic gradient by decreasing sodium chloride transport along the proximal tubule or thick ascending loop of Henle. The result is poor abstraction of water from the filtrate, with hyposthenuria and polyuria. Such an increase in solute and water within the tubular fluid results in adaptive downregulation of the glomerular filtration process.

The antigen targets engaging the inflammasome and the immune system in interstitial nephritis are slowly unfolding. Drugs act as haptens, mimic endogenous structures in the interstitium, alter regulation of the immune system, or function in some combination of the foregoing. Bacteria, fungi, and viruses can infect the kidney and cause mononuclear cell infiltration or activate toll-like receptors on tubular epithelia, which subsequently educate the adaptive immune response to events in the interstitium. Autoimmune diseases such as anti–tubular basement membrane disease or spontaneous interstitial nephritis remain confined to the kidney, whereas systemic diseases spread to the kidney, where they cause persistent, chronic interstitial nephritis.

Although the adaptive immune response is similar to that of other tissues, T-cell activation figures prominently in interstitial nephritis. Antibodies (anti–tubular basement membrane disease) and immune complex deposition along the tubular basement membrane (systemic lupus erythematosus) are rarely seen. Antigens presented by class II major histocompatibility complex molecules on macrophages, dendritic cells, and adjacent tubular epithelia, in conjunction with associative recognition molecules, engage the CD4/CD8 T-cell repertoires. The resultant cytokine and protease activity injures tubular nephrons and basement membranes and causes fibroblasts to form locally by epithelial-mesenchymal transition and to proliferate. Transforming growth factor-β (TGF-β), fibroblast growth factor 2, and platelet-derived growth factor are particularly active in this transition. If the nephritis persists, fibrogenesis dismantles nephrons and causes tubular atrophy; in late stages, the inflammatory reaction outgrows its survival factors and lymphocytes and fibroblasts disappear by apoptosis and leave an acellular fibrotic scar.[4]

Viruses, including Epstein-Barr virus, have long been suspected of contributing to idiopathic, chronic interstitial nephritis. Malakoplakia and xanthogranulomatous pyelonephritis are probably not defects in nephrogenesis but rather are destructive responses to bacterial inflammation in the interstitium. Focal abnormalities in kidney structure can be a nidus for infections associated with perinephric, psoas, or peritoneal abscesses. Children with vesicoureteral reflux can have chronic or repeating episodes of pyelonephritis, but whether the reflux or the infection is more important to progression to renal failure is unclear. There is also no agreement on whether recurrent pyelonephritis by itself produces chronic interstitial nephritis in adults.

Pathology

Both kidneys are typically involved, except in cases of unilateral infection, obstruction, or trauma. The inflammatory reaction in acute interstitial nephritis consists mainly of T lymphocytes and monocytes, but neutrophils, plasma cells, and eosinophils can be present. The T cells are of a mixed phenotype with a distinct preference for CD4+ lymphocytes. The infiltrative process is associated with interstitial edema, which displaces tubules away from one another and causes the kidneys to swell. The tubular basement membrane may be disrupted in more severe cases, but immune deposits are rarely found by immunofluorescence.

In chronic interstitial nephritis, the kidney assumes an irregular or contracted appearance. The tubular epithelia sit on thickened or disrupted tubular basement membranes and are often effaced against dilated lumens; the tubules eventually dismantle and atrophy. *Chronic* is a relative term, because fibrotic changes can be seen within 7 to 10 days of continuing inflammation. Normal glomeruli in primary interstitial nephritis are eventually surrounded by periglomerular fibrosis and subsequently undergo segmental or global sclerosis. Chronic vascular thickening and glomerular changes are present in advanced stages of disease, so pathologic determination of the primary cause may be difficult in some biopsy samples. Progressive glomerular sclerosis also occurs with aging and must be factored in when interpreting the biopsy findings.

A third pathologic category, granuloma formation, can be seen in either acute or chronic interstitial nephritis. In acute granulomatous interstitial nephritis, granulomas are sparse and non-necrotic and giant cells are rare. The granulomas in chronic interstitial nephritis contain an abundance of giant cells, and those caused by tuberculosis may become necrotic. Drugs are a common cause of this lesion in the acute setting, and most of the drugs associated with acute interstitial nephritis have been reported to cause granuloma formation. In the absence of drug exposure, sarcoidosis (Chapter 95), Wegener granulomatosis with polyangiitis (Chapter 270), histoplasmosis (Chapter 332), or tuberculosis (Chapter 324) should be considered, depending on the context, when numerous granulomas are present. The renal granulomas seen in granulomatosis with polyangiitis are almost always accompanied by glomerular and vascular pathology.

ACUTE INTERSTITIAL NEPHRITIS

CLINICAL MANIFESTATIONS

Most patients present with an asymptomatic rise in the serum creatinine level or an abnormal urinalysis, and it is important to consider acute interstitial nephritis in any patient with an unexplained precipitous diminution in renal function.[5] Because injury is often asymptomatic, patients already may have substantial renal failure on initial presentation. Patients also may present with nonspecific symptoms such as lethargy or weakness, and many patients have fever and oliguria owing to severe acute kidney injury (Chapter 120).

Several features can distinguish acute interstitial nephritis from acute tubular necrosis (Chapter 120) or glomerulonephritis (Chapter 121) (Table 122-3). Fever and occasional flank pain over the kidneys occur in infection or with drug-induced acute interstitial nephritis. Lumbar pain, sometimes unilateral, is due to distention of the renal capsule. Allergic reactions are associated with maculopapular rash, fever, and eosinophilia, but the entire triad is seen in less than 33% of patients, and such signs are uncommon when NSAIDs cause acute interstitial nephritis. These signs and symptoms of drug reaction have been codified and referred to as the DRESS syndrome (*d*rug *r*ash, *e*osinophilia, and *s*ystemic *s*ymptoms; Chapter 440), which is associated with interstitial nephritis in up to 40% of patients with persistent exposure to selected drugs.[6]

The course of renal failure in acute interstitial nephritis takes several days to weeks and follows the kinetics of the primary immune response. However, renal failure can be precipitous, especially in patients re-exposed to a previous agent. Rarely, the course can be protracted, with the glomerular filtration rate declining over a period of months if the diagnosis is not recognized. This protracted course is more common with diuretic-induced interstitial nephritis. The onset of drug-induced nephritis ranges from days to weeks after the initiation of therapy, and a previous allergic history is rare. The classic setting for a drug reaction is a febrile patient with an infectious process and who defervesces while taking antibiotics but then develops recurrent fever several days later.

DIAGNOSIS

Urinalysis is particularly helpful. Mild-to-moderate proteinuria and hematuria are seen in most cases, and gross hematuria is observed rarely. The sediment typically shows red and white blood cells, and white blood cell casts are commonly seen. Conversely, red blood cell casts suggest a glomerular

TABLE 122-3 TYPICAL CLINICAL MANIFESTATIONS OF ACUTE INTERSTITIAL NEPHRITIS

History of drug hypersensitivity or recent infection and taking antibiotics
Sudden onset of fever lasting several days to weeks
Variable degrees of hypertension
Rise in creatinine with FE_{Na} > 1.0; no expected acute tubular necrosis or glomerulonephritis
Kidney size normal or increased
Hematuria with mild proteinuria (<1.0 g)
Presence of WBC casts and WBCs on urinalysis; rarely eosinophils

FE_{Na} = fractional excretion of sodium; WBC = white blood cell.

TABLE 122-4 WHEN TO CONSIDER A RENAL BIOPSY TO DIAGNOSE NEPHRITIS

The setting, history, or clinical findings do not support a diagnosis of acute tubular necrosis or volume depletion
The clinical setting warrants a tissue diagnosis to determine the type of lesion, the extent of involvement, or the degree of fibrosis
The patient is stable enough to undergo biopsy and receive immunosuppressive drugs
The physician believes that the choice of therapy or the length of treatment is partially determined by the type of tissue injury

diagnosis. The finding of eosinophils in the urine supports the diagnosis of allergic interstitial nephritis, but the positive predictive value is low, even with more than 5% eosinophils in the urine, and the absence of eosinophiluria does not exclude the diagnosis of acute interstitial nephritis.[7]

An elevated serum creatinine level is usually the first abnormal laboratory result in renal injury. The normal serum creatinine of 0.6 to 1.3 mg/dL varies with muscle mass, age, and gender. Early recognition of acute interstitial nephritis requires a high degree of clinical suspicion because the serum creatinine level may be only mildly elevated even after the kidneys lose half their function.

The magnitude of proteinuria in acute interstitial nephritis is nearly always less than 3 g/24 hours and is typically less than 1 g/24 hours. Nephrotic-range proteinuria is not seen unless there is a coexisting glomerular lesion, such as a concurrent minimal-change lesion, or after exposure to NSAIDs. Many patients with acute interstitial nephritis also have a fractional excretion of sodium (FE_{Na}) greater than 1, but occasionally they are oliguric.

Imaging is of little diagnostic value. The kidney in acute interstitial nephritis is usually normal or slightly increased in size on echographic or tomographic images. Increased cortical echogenicity may correlate with diffuse interstitial infiltrates on renal biopsy. Gallium scanning is not particularly useful because a variety of other renal processes can cause gallium uptake, including minimal change glomerulonephritis, cortical necrosis, and acute tubular necrosis; in addition, acute interstitial nephritis can be found on biopsy in those with a normal scan.

DIFFERENTIAL DIAGNOSIS

It is sometimes difficult to distinguish among nonoliguric acute tubular necrosis, acute interstitial nephritis, and glomerulonephritis without a biopsy. Exposure to pharmaceuticals, particularly antibiotics and NSAIDs, is responsible for most cases of acute interstitial nephritis, followed by infections and autoimmune disease. Selective tubular defects and tubular syndromes, such as proximal acquired Fanconi syndrome or distal renal tubular acidosis, can be seen in subacute or chronic interstitial nephritis but argue against acute interstitial nephritis.

Biopsy

Ultimately, the diagnosis can be established with certainty only by renal biopsy, which confirms and assesses the extent of acute interstitial inflammation. A biopsy should be performed in patients with acute renal failure who have suggestive signs or symptoms of an interstitial process and in whom prerenal azotemia and obvious acute tubular necrosis cannot be excluded on clinical grounds (Table 122-4). In primary acute interstitial nephritis, the biopsy demonstrates inflammatory cells that typically spare the glomeruli until late in the course (Fig. 122-1A). Lesions that reduce renal function are usually diffuse, but drug-induced interstitial injury is often patchy, beginning deep in the cortex before spreading.

TREATMENT

Rx

Biopsy is important to confirm acute interstitial nephritis, because chronic interstitial fibrosis rarely responds to aggressive treatment. The principal intervention for acute interstitial nephritis is to remove the inciting drug or treat the infection. Switching to different derivatives of a suspected drug is unwise. Concomitantly, or if the serum creatinine concentration does not fall after a few days, steroids (prednisone 0.75 to 1.0 mg/kg PO) can be given daily for approximately 1 week. If no further improvement occurs, cyclophosphamide (1 to 2 mg/kg/day PO) can be added for several more weeks. In patients who respond, cyclophosphamide can be steroid sparing, particularly in those with persistent sarcoidosis. It is important not to continue high-dose immunosuppression without some evidence of benefit because immunosuppressive drugs in patients with azotemia can lead to serious infection and even death. It is better to reserve these drugs for use with kidney transplantation if the primary disease does not respond.

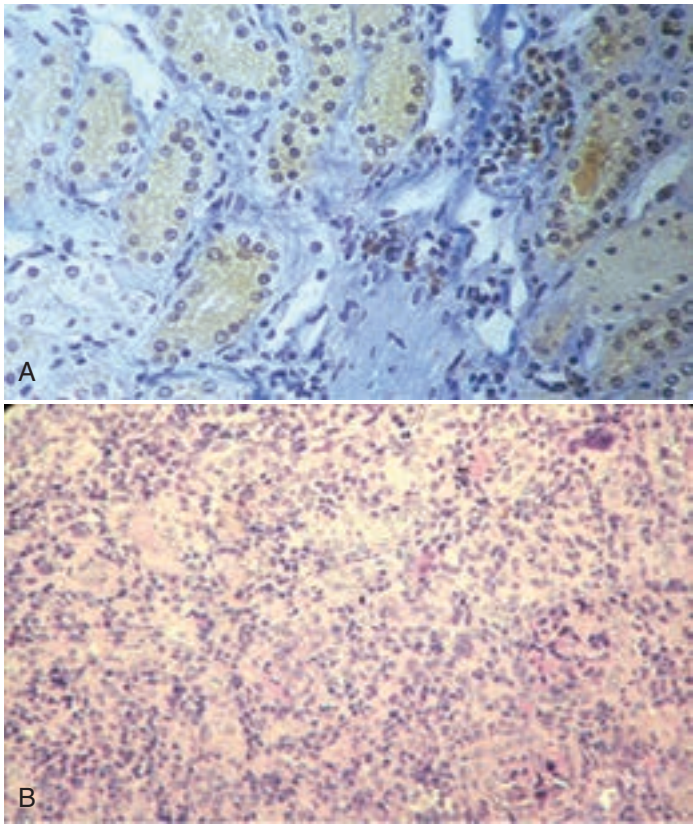

FIGURE 122-1. Tubulointerstitial nephritis on biopsy. A, Acute interstitial nephritis can be most aggressive when the interstitium is crowded with mononuclear cells and giant cells that destroy nearly all tubular nephrons (hematoxylin-eosin). **B,** Chronic interstitial nephritis is a slower process, with substantial collagen deposition (*blue color;* trichrome), tubular dropout, and fibroblasts in the interstitial spaces widened by fibrosis.

PROGNOSIS

The prognosis for acute interstitial nephritis is good if it is recognized early in the setting of minimal fibrosis. In nonrandomized, observational series, patients treated with steroids tend to do somewhat better.[8] Early removal of offending agents or prompt treatment with antibiotics or immunosuppressive drugs can be renoprotective.

CHRONIC INTERSTITIAL NEPHRITIS

CLINICAL MANIFESTATIONS

Patients with primary chronic interstitial nephritis typically have elevated levels of serum creatinine and signs and symptoms of renal failure, including hematuria, hyposthenuria, nocturia, fatigue, and nausea. Urinalysis shows a fixed specific gravity of about 1.010, occasional glycosuria, and non–nephrotic-range proteinuria (often <1 g/L), with red and white blood cells and granular casts. Pyuria and positive urine cultures for bacteria are seen occasionally, and varying degrees of metabolic acidosis and hyperphosphatemia may be present. Before the glomerular filtration rate falls below 25 to 30 mL/minute, tubular acidosis is common. Anemia is often out of proportion to the degree of renal failure, and many patients have hypertension but only minimal edema until advanced stages of renal failure. Acquired Fanconi syndrome can be seen in patients with a serum creatinine level less than 2.5 mg/dL in the setting of drug exposure, myeloma, human immunodeficiency virus (HIV) infection, lead exposure, and herbal nephropathy.

DIAGNOSIS

A careful dietary history is critical. As for any patient with evidence of renal failure, the evaluation includes laboratory tests to determine possible causes and severity. These tests include measures of renal function (serum creatinine

level and blood urea nitrogen level), as well as levels of serum electrolytes, calcium, phosphate, uric acid, and albumin. Urinalysis shows a fixed specific gravity of approximately 1.010, occasional glycosuria, proteinuria (often < 1 g/L), and red cells, white cells, and granular casts. Depending on the clinical situation, the search for specific causes may include serum and urine protein electrophoresis, blood cultures, serologic tests for autoimmune diseases (e.g., cryoglobulin level, antinuclear antibodies, anti–neutrophil cytoplasmic antibody, and anti–glomerular basement membrane antibody levels [see Table 257-2]), or viral infection, particularly after renal transplantation (Chapter 131).

Selective tubular defects and tubular syndromes, such as proximal acquired Fanconi syndrome (bicarbonaturia with a plasma carbon dioxide [CO_2] content < 20 mEq/L, aminoaciduria, phosphate wasting, uricosuria, and glycosuria) or distal renal tubular acidosis type 1 (urine pH > 5.6, plasma CO_2 content < 20 mEq/L, with low or high potassium) can be seen occasionally in subacute or chronic interstitial nephritis. Patients with Fanconi syndrome, in particular, exhibit proximal tubular epithelium alterations that variably impair transporter function in the area of injury. These tubular defects are classically and occasionally seen in light-chain myeloma, cystinosis, Lowe syndrome, TINU syndrome, biliary cirrhosis, or after exposure to selected drugs, such as tenofovir or ifosfamide. Patients with Fanconi syndrome quickly develop an alkaline urine (pH > 7.5) and increased fractional excretion of urine bicarbonate when serum bicarbonate is elevated above 20 mEq/L by intravenous infusion.

Classic images of analgesic nephropathy on tomography are quite specific (Fig. 122-2) and show a decrease in overall kidney size, with atrophic scars and an irregular cortical contour, sometimes accompanied by papillary necrosis.

On biopsy (see Table 122-4), chronic interstitial nephritis is manifest by a cellular infiltrate that is eventually replaced by tubulointerstitial fibrosis (see Fig. 122-1B). Infiltrates of lymphocytes and rare neutrophils are scattered and less abundant than in acute interstitial nephritis.

Specific Causes

Analgesics

Aspirin, acetaminophen, and NSAIDs alone or together are a source of toxic metabolites and can induce medullary ischemia and papillary necrosis, sometimes with papillary calcification. The likelihood of analgesic nephropathy from taking acetaminophen alone is much less than with the others. Uroepithelial malignancies also occur with increased frequency in this group of patients.

Aristolochic Acid Nephropathy

Aristolochic acid has been implicated as a cause of Balkan nephropathy and so-called Chinese herbal nephropathy. A growing number of people are taking vitamins and herbal preparations purchased from health food stores (Chapter 39), and some of these remedies contain botanicals that produce chronic interstitial nephritis.[9] Patients who are dieting often use these remedies and are first seen when they already have late-stage disease, which increases the risk for uroepithelial malignancies (Chapter 197).

Human Immunodeficiency Virus Tubulointerstitial Nephropathy

Predominate tubulointerstitial nephropathy accounts for about 25% of the renal lesions seen on biopsy in HIV-infected patients. Biopsies show two general forms: a tubulopathy or an interstitial nephritis. In the tubulopathy, damage is to the proximal tubule; 80% of cases are associated with drug exposure, particularly tenofovir (Chapter 360), and approximately 30% improve with time. Patients in the tubulointerstitial nephritis group have a mononuclear infiltration and more persistent viral loads, but 50 to 60% recover.[10]

Tubulointerstitial Nephritis with Uveitis Syndrome

Seen at any age but more commonly in young women and children, the TINU syndrome may be idiopathic, genetic, or in response to pharmaceutical exposure.[11] Unilateral or bilateral anterior panuveitis (Chapter 423) may precede or follow evidence of interstitial nephritis that starts acutely but may persist as chronic renal injury. If the renal function is not greatly impaired, concomitant Fanconi syndrome may suggest proximal tubular involvement. In the absence of controlled trials, treatment with varying doses of methylprednisolone followed by oral prednisone may be beneficial in some patients. Varying durations of supplemental or steroid-sparing immunosuppression with mycophenolate mofetil or cyclophosphamide also have been used.

Vascular Disease

Chronic renal ischemia from vascular injury can lead to interstitial nephritis, nephrosclerosis, and fibrosis, which are the classic renal lesions of untreated

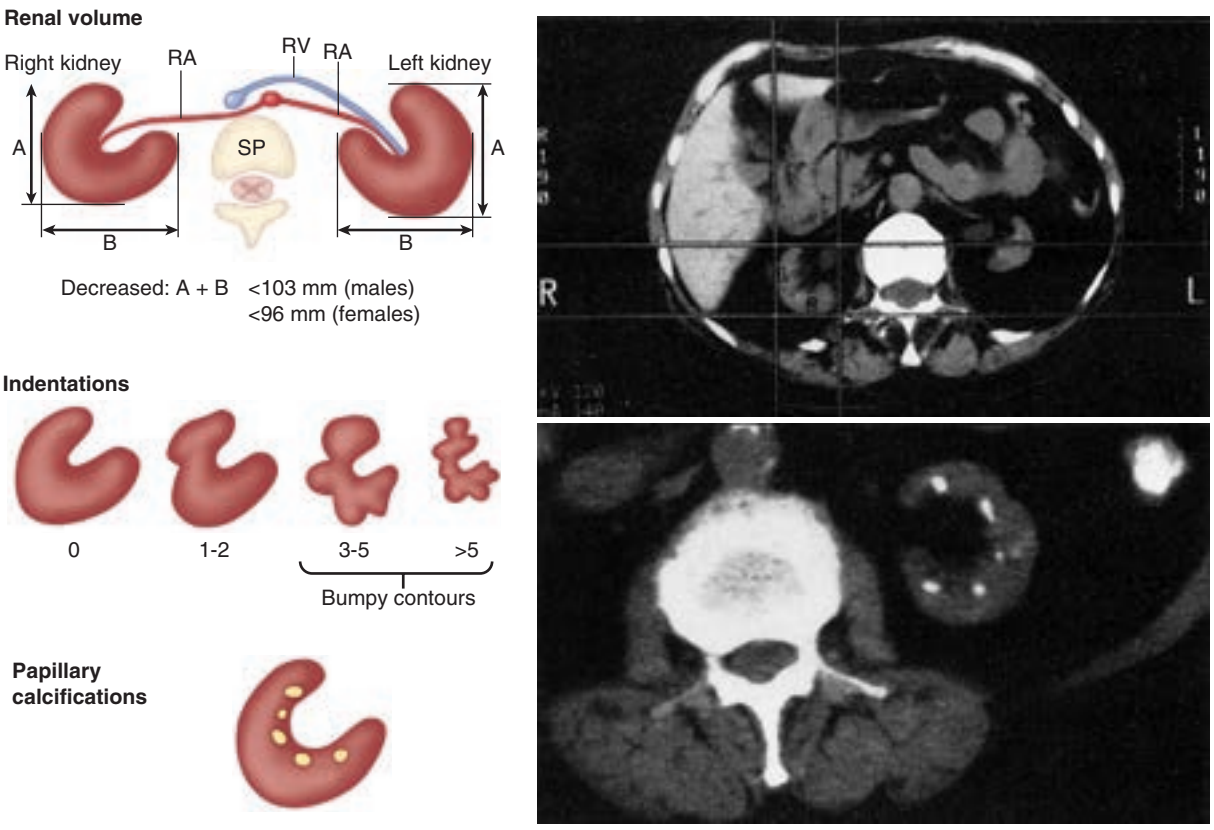

FIGURE 122-2. Renal changes in analgesic nephropathy seen by tomographic imaging. Structural changes, including reduced volume, nodularity, and calcifications, are seen on computed tomography. RA = right artery; RV = right vein; SP = spinal vertebra. (From Elseviers MM, De Schepper A, Corthouts R, et al. High diagnostic performance of CT scan for analgesic nephropathy in patients with incipient to severe renal failure. *Kidney Int.* 1995;48:1316.)

essential hypertension (Chapter 67). Similar injury is seen with aging, diabetes (Chapter 124), sickle cell disease (Chapter 163), and radiation nephritis (Chapter 20). This tubulointerstitial injury from the vascular diseases is quite different from the aggressive necrosis seen with acute vasculitis. In patients taking calcineurin inhibitors such as cyclosporine or tacrolimus, renal ischemia from vasoconstriction can cause interstitial fibrosis that is sometimes difficult to distinguish from chronic allograft rejection (Chapter 131).

Immunoglobulin G4–Related Tubulointerstitial Nephritis

Immunoglobulin G4 (IgG4)-related tubulointerstitial nephritis is a relatively new systemic syndrome that expresses in the kidney as inflammatory masses associated with plasma cells and interstitial mononuclear cell infiltrates. Patients may have lesions in other organs, such as the liver, pancreas, thyroid, and myocardium. Eosinophilia and low complement levels can be seen. Serum IgG4 levels and tissue plasma cells are increased, but it is not clear whether the IgG4 antibodies are a biomarker or causative. The renal lesions in many patients respond briskly to corticosteroid treatment.[12] IgG4 is also separately associated with membranous nephropathy (Chapter 121) and nephrotic-range proteinuria.

Obstruction

Significant urinary obstruction (Chapter 123) owing to occlusion of both ureters by bladder tumors, cervical carcinoma, ureteral valve disease, or bladder outlet obstruction is an important cause of chronic interstitial nephritis. Complete or partial urinary tract obstruction is accompanied by a decline in glomerular filtration and classic tubular abnormalities, including diminished reabsorption of solutes, impaired excretion of H^+ and K^+, and a vasopressin-resistant concentrating defect in the medulla. Obstruction is associated with a fall in the glomerular filtration rate because of reduced plasma flow and hydraulic pressure associated with the release of angiotensin II, leukotrienes, and nitric oxide, a process leading to mononuclear cell infiltration. Growth factors such as TGF-β, released by infiltrating cells, may contribute to the interstitial and glomerular fibrosis. Obstruction is more common in men than in women and is part of the routine assessment of renal failure by renal ultrasound (Chapters 120 and 123). Almost all obstructed kidneys eventually become infected if the obstruction is not relieved.

Hypercalcemia

Hypercalcemia can decrease glomerular filtration through renal vasoconstriction, a decrease in the glomerular ultrafiltration coefficient, and volume depletion as a result of a vasopressin-resistant concentrating defect associated with nephrocalcinosis and calcium deposition around the basement membranes of the distal tubules and collecting ducts. Such deposition secondarily leads to mononuclear cell infiltration and tubular death. Nephrocalcinosis also occurs in normocalcemic disorders of augmented calcium absorption through the gut (sarcoidosis [Chapter 95], vitamin D intoxication [Chapter 245]), skeletal breakdown (neoplasms or multiple myeloma [Chapter 187]), or classic distal renal tubular acidosis.

Myeloma

The chronic renal failure of multiple myeloma (Chapter 187) is caused by several mechanisms, including cast nephropathy ("myeloma kidney"), coexistent volume depletion, hypercalcemia (Chapter 245), nephrocalcinosis (Chapter 245), and uric acid nephropathy.[13] Proteinaceous casts form in dilated, atrophic distal nephron segments that are surrounded by multinucleated giant cells in interstitial infiltrates. The casts typically contain both Tamm-Horsfall protein and a pathologic light chain. Interstitial plasma cells and mononuclear infiltrates, calcifications in the interstitium, and amyloid deposits in the vessels and glomeruli are often present. Light chains are nephrotoxic by direct injury to tubular cells or through intrarenal obstruction from cast formation. In the setting of excess light chain production, the proximal tubule reabsorptive capacity is overwhelmed, leading to their urinary excretion as Bence Jones proteins. An elevated intratubular pressure partly accounts for the decline in glomerular filtration in experimental cast nephropathy.

Lead Toxicity

Epidemiologic analyses support the association between excess lead burden (Chapter 22) and chronic renal failure.[14] Blood lead levels reflect only recent, not chronic, exposure and can be normal in patients with a significant lead burden. Lead preferentially deposits in the proximal tubule, and nuclear inclusions within proximal tubular cells are characteristic of lead nephropathy.

Ingestion of moonshine liquor, with its high lead content, can be an important historical clue to the diagnosis. In adults, lead nephropathy produces chronic interstitial nephritis, fibrosis, and nephrosclerosis. Proximal tubular dysfunction may produce isolated tubule defects or a full Fanconi syndrome. Patients often have recurrent gout, and hyperuricemia and hypertension may be present. Some laboratories can measure δ-aminolevulinic acid dehydratase, which is inhibited by lead. Although chelation studies may document lead burden, this test is difficult to perform in patients with renal failure. X-ray fluorescent measurements of in vivo skeletal lead stores correlate well with ethylenediaminetetraacetic acid (EDTA) chelation tests and have the advantage of being rapid and noninvasive.

Cadmium Toxicity

Cadmium nephropathy (Chapter 22) is seen in regions with contamination from smelters that result in prolonged low-level exposure. Cadmium is bound to metallothionein, and proximal tubular cells take up these complexes. The liver and kidney are the two major organs in which cadmium accumulates. Its half-life in the body is longer than 10 years. Like blood levels of lead, blood levels of cadmium fall after acute exposure because of extensive tissue deposition. Once a threshold of renal deposition is exceeded, excess cadmium is excreted in urine. Cadmium intoxication produces irreversible proximal tubular dysfunction, hypercalciuria, nephrolithiasis, and metabolic bone disease with pain (called "ouch-ouch" disease in Japan).

Hyperuricemia

Hyperuricemia, especially in acutely treated myeloproliferative disease, can cause acute renal failure. Many patients with chronic renal failure have serum uric acid levels higher than 10 mg/dL, attributable to diminished glomerular filtration and the effects of diuretics. However, most studies have not demonstrated an independent association of hyperuricemia with chronic interstitial disease that could not otherwise be attributed to hypertension, vascular disease, calculi, or aging.

TREATMENT

Rx

Chronic interstitial nephritis tends to progress slowly. Inciting factors such as obstruction, infection, drugs, or toxins should be removed whenever possible. The treatment is similar to that of other causes of chronic renal failure. Angiotensin-converting enzyme inhibitors or angiotensin II receptor blockers (Table 67-7) are used early to slow disease progression, with a systolic blood pressure goal of 140 mm Hg (Chapters 67 and 130), except when hyperkalemia limits their use. Early treatment of acidosis with sodium bicarbonate, starting at 600 mg orally (PO) three times daily. Anemia is treated with erythropoiesis-stimulating agents (e.g., darbepoetin alfa 0.45 μg/kg weekly to keep the hemoglobin concentration between 10 and 12 g/L), hyperphosphatemia with oral phosphate binders (Table 119-4) and hyperparathyroidism with calcitriol (starting at 0.25 μg/day) can improve performance status and protect against bone loss (Chapters 130 and 131). There is no clear role for immunosuppressive drugs in the treatment of chronic interstitial nephritis, except perhaps in early sarcoidosis (Chapter 95).

For certain specific causes of chronic interstitial nephritis, specific therapeutic approaches are warranted. For analgesic nephropathy, stopping analgesic use can help reduce progression. For hypercalcemia (Chapter 245), therapy is directed toward the primary disease—reduction of the serum calcium concentration, when appropriate, and correction of acid-base disturbances.

Appropriate therapy for presumed cast nephropathy in multiple myeloma includes chemotherapy to ameliorate excess light chain production (Chapter 187); treatment of hypercalcemia (Chapter 245); alkalinization of the urine with the addition of bicarbonate to hypotonic fluids; and avoidance of radiocontrast agents, which may enhance the nephrotoxicity of light chains. Loop diuretics should be used with caution, particularly in the setting of volume depletion.

EDTA is advocated as chelation therapy for lead toxicity (Chapter 22). The goal of chelation is to normalize the EDTA mobilization test. In occasional patients, this may arrest or reverse the progression of the renal failure.

PROGNOSIS

The prognosis for chronic interstitial nephritis is highly variable and depends on the underlying condition and on comorbid conditions, including cardiovascular disease and diabetes mellitus, which become increasingly common in these patients over time.

Grade A Reference

A1. de Brito-Ashurst I, Varagunam M, Raftery MJ, et al. Bicarbonate supplementation slows progression of CKD and improves nutritional status. *J Am Soc Nephrol.* 2009;20:2075-2084.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Goicoechea M, Rivera F, Lopez-Gomez JM, et al. Increased prevalence of acute tubulointerstitial nephritis. *Nephrol Dial Transplant*. 2013;28:112-115.
2. Praga M, Sevillano A, Aunon P, et al. Changes in the aetiology, clinical presentation and management of acute interstitial nephritis, an increasingly common cause of acute kidney injury. *Nephrol Dial Transplant*. 2014; [Epub ahead of print].
3. Zeisberg M, Neilson EG. Mechanisms of tubulointerstitial fibrosis. *J Am Soc Nephrol*. 2010;21:1819-1834.
4. Tampe D, Zeisberg M. A primer on the epigenetics of kidney fibrosis. *Minerva Med*. 2012;103:267-278.
5. Raghavan R, Eknoyan G. Acute interstitial nephritis—a reappraisal and update. *Clin Nephrol*. 2014;82:149-162.
6. Criado PR, Criado RF, Avancini Jde M, et al. Drug reaction with eosinophilia and systemic symptoms (DRESS) / drug-induced hypersensitivity syndrome (DIHS): a review of current concepts. *An Bras Dermatol*. 2012;87:435-449.
7. Muriithi AK, Nasr SH, Leung N. Utility of urine eosinophils in the diagnosis of acute interstitial nephritis. *Clin J Am Soc Nephrol*. 2013;8:1857-1862.
8. Raza MN, Hadid M, Keen CE, et al. Acute tubulointerstitial nephritis, treatment with steroid and impact on renal outcomes. *Nephrology (Carlton)*. 2012;17:748-753.
9. Allard T, Wenner T, Greten HJ, et al. Mechanisms of herb-induced nephrotoxicity. *Curr Med Chem*. 2013;20:2812-2819.
10. Zaidan M, Lescure FX, Brocheriou I, et al. Tubulointerstitial nephropathies in HIV-infected patients over the past 15 years: a clinico-pathological study. *Clin J Am Soc Nephrol*. 2013;8:930-938.
11. Reddy AK, Hwang YS, Mandelcorn ED, et al. HLA-DR, DQ class II DNA typing in pediatric panuveitis and tubulointerstitial nephritis and uveitis (TINU). *Am J Ophthalmol*. 2014;157:678-686.
12. Cornell LD. IgG4-related kidney disease. *Curr Opin Nephrol Hypertens*. 2012;21:279-288.
13. Sanders PW. Light chain-mediated tubulopathies. *Contrib Nephrol*. 2011;169:262-269.
14. Sabath E, Robles-Osorio ML. Renal health and the environment: heavy metal nephrotoxicity. *Nefrologia*. 2012;32:279-286.

REVIEW QUESTIONS

1. A 35-year-old man employed in a long-term modeling contract sees you for the first time. He is worrying that he had seen blood in his urine. An ardent disciple of exercise, dieting, and supplements, he is in a monogamous relationship and taking no pharmaceuticals. On routine testing he has a serum creatinine of 2.3 mg/dL with +2 protein and blood in his urine without bacteria. Computed tomographic scan revealed normal-sized kidneys and a bladder mass. The most likely cause of his renal injury is:

A. Herbal nephropathy
B. BK virus nephropathy
C. Myeloma kidney
D. Sjögren syndrome
E. Silent pyelonephritis

Answer: A Patient was taking various unknown herbal agents freely available over the counter. The likely exposure to aristolochic acid in these supplements has led to chronic interstitial nephritis and a uroepithelial tumor in the bladder producing hematuria. BK virus nephropathy is seen only after transplantation on immunosuppression. This is the wrong age for myeloma kidney, there are no symptoms of Sjögren syndrome, and pyelonephritis is unlikely in the absence of fever and bacteria in the urine.

2. A 44-year-old female office manager for a well-known U.S. senator has suffered from self-described stomach pains for many years. In the distant past, she was periodically given short courses of tetracycline from a friend for presumed gastric ulcers with unclear relief. Her regular physician has had her on long-term proton pump inhibitors for chronic gastroesophageal reflux. Following a persistent low-grade fever, she visited her physician, and testing revealed a recent rise in her serum creatinine to 1.5 mg/dL, pyuria without bacteria, and a fever of 100.5° F. The most likely diagnosis is:

A. Omeprazole nephropathy
B. Sarcoidosis
C. Tetracycline nephropathy
D. Malakoplakia
E. Renal calculi

Answer: A Patient was taking proton pump inhibitors for a long time, which can cause acute interstitial nephritis; the reaction is idiosyncratic. Tetracyclines are an unlikely cause because she is not taking them presently. She has no symptoms of sarcoidosis. Malakoplakia is extremely rare and typically involves the lower urinary tract. Renal calculi often produce hematuria, a different pain syndrome, and raise serum creatinine usually as a result of bilateral obstruction.

3. A 67-year-old retired auto mechanic living in a rural town visits his physician with complaints of fatigue, shortness of breath, and generally not feeling well. He takes no pharmaceuticals or over-the-counter drugs. His vital signs were normal, HIV screen was negative, serum calcium was 11.9 mg/dL, serum hemoglobin was 7.9 g/dL, and serum creatinine was 2.8 mg/dL. His urine protein was 723 mg/L, and only a few white blood cells were found on urinalysis. He was admitted to the hospital when a lucency on his left femur was seen on a bone radiograph. The most likely diagnosis is:

A. Myeloma nephropathy
B. Focal glomerulosclerosis
C. Xanthogranulomatous pyelonephritis
D. Balkan nephropathy
E. Ouch-Ouch disease

Answer: A The patient presents with substantial clinical evidence of classic plasma cell dyscrasia, probably a form of myeloma. His urine contained highly abnormal levels of κ light chains with a monoclonal spike. Biopsy revealed cast nephropathy with diffuse interstitial nephritis, and his elevated calcium with lytic lesions suggests systemic bony involvement. With normal vital signs, less than a gram of protein in the urine, and a negative HIV test, focal glomerulosclerosis seems unlikely. The absence of fever or obvious pyuria eliminates xanthogranulomatous pyelonephritis. Patients with Balkan nephropathy or Ouch-Ouch disease (cadmium toxicity) do not normally present with hypercalcemia and lytic bone lesions, and there is no history of exposure to herbals or toxic metals.

4. A 19-year-old female college student appeared in a student health clinic with complaints of generalized weakness, a sense of increased respiration, and bilateral blurry vision. She was taking birth control pills but had been otherwise well. Her college physical examination was completely normal, and her childhood uneventful. Her vital signs were normal, and her retina was normal on ophthalmoscopy. Her blood work showed a serum creatinine of 1.6 mg/dL, serum HCO_3 of 19 mEq/L, serum glucose of 90 mg/dL, serum calcium of 9.1 mg/dL, and a serum phosphate of 3.2 mg/dL. Her chest radiograph was normal, but her urinalysis demonstrated pyuria and glycosuria. She was admitted urgently for a kidney biopsy and a formal ophthalmologic examination. The most likely cause is:

A. Tubulointerstitial nephritis and uveitis (TINU) syndrome
B. Sarcoidosis
C. Epstein-Barr virus nephropathy
D. Lead nephropathy
E. Light chain nephropathy

Answer: A The patient presents with substantial clinical evidence of Fanconi syndrome with an elevated serum creatinine level. Her kidney biopsy will likely show interstitial nephritis, and her eye examination will likely reveal anterior uveitis. Sarcoidosis is unlikely with a normal chest radiograph and serum calcium level. She has no findings of Epstein-Barr virus infection, no history of lead exposure, and is too young for light chain disease.

123 OBSTRUCTIVE UROPATHY

MARK L. ZEIDEL

DEFINITION

Each day an average adult produces 1.5 to 2 L of urine, which must flow from the kidneys to the end of the urethra, a process that requires proper functioning of each renal pelvis, the ureters, bladder, and urethra. *Obstructive uropathy* occurs when a structural or functional defect in the urinary tract blocks or reduces urine flow. *Obstructive nephropathy* ensues when obstructive uropathy impairs renal function. Increased hydrostatic pressure from downstream obstruction may dilate upstream elements of the urinary tract, thereby causing *hydronephrosis.* Because recovery of renal function relates inversely to the duration and severity of the obstruction, prompt recognition and treatment of obstructive uropathy is essential for preserving renal function in this condition.

EPIDEMIOLOGY

Although there are few studies of unselected populations, in autopsy series hydronephrosis occurs at a rate of 3.1% in all subjects (2.9% in females, 3.3% in males). Autopsy in children younger than 16 years reveal hydronephrosis in 2.2% of boys and 1.5% of girls; 80% of hydronephrosis occurs in children younger than 12 months. In adults, hydronephrosis occurs with equal frequency in both sexes in those younger than 20 years, but owing to pregnancy and uterine cancer, it is more common in women than in men between the ages of 20 and 60 years. In individuals older than age 60 years of age, obstructive uropathy occurs more commonly in men because of prostate disease. The annual frequency of hospitalization for obstructive uropathy in the United States is 166 per 100,000. Each year, the 2000 or so patients who begin treatment for end-stage renal disease (ESRD) because of a presumed diagnosis of obstructive nephropathy represent approximately 2% of patients with ESRD. Among these patients, 4% are younger than 20 years of age, 44% are 20 to 64 years, and the others are older than 64 years.

PATHOBIOLOGY

Rhythmic, coordinated contractions of the renal pelvis "milk" urine from the renal papilla into the proximal end of the ureter. Peristaltic contractions of the ureter coordinate with periodic openings of the ureterovesical junction to propel the urine to the bladder. As the bladder fills, stretch is detected in its muscular wall and possibly its lining epithelium, the urothelium, thereby activating relaxation reflexes that suppress contraction of the bladder wall musculature and tighten the urethral sphincter to allow the bladder to expand without large increases in intravesicular pressure (E-Fig. 123-1; Chapter 26). When filling reaches a critical level, the relaxation reflex is suppressed and the voiding reflex is initiated. Suppression of detrusor muscle contraction ends and stimulation begins while the urethral sphincter is relaxed, leading to the buildup of pressure needed for voiding.

Obstructive uropathy results from functional or mechanical defects of the entire urinary tract. Functional failures include an inability to open the ureteropelvic or ureterovesical junction, failure to open the urethrovesical junction, or failure of bladder reflexes. Partial or complete mechanical blockade of the urinary tract at any level can lead to obstruction.

Functional or mechanical obstruction can occur at any point along the urinary tract from the renal pelvis and proximal urethra to the end of the urethra (phimosis). Because diagnosis and treatment depend heavily on the location of the obstruction, disorders are classified by anatomic location and whether the obstruction is due to factors within the urinary tract (intrinsic obstruction) or factors outside the tract (extrinsic obstruction) (Table 123-1). Intrinsic obstruction may be due to intraluminal or intramural causes. Intraluminal causes include stones or sludging of material, such as sloughed papillae or clots in papillary necrosis. Intramural causes may be anatomic (e.g., tumors or strictures) or functional (e.g., uncoordinated ureteral peristalsis or failure to open the ureteropelvic or ureterovesical junction). Extrinsic

TABLE 123-1 CAUSES OF URINARY TRACT OBSTRUCTION

INTRARENAL
Uric acid nephropathy
Sulfonamide precipitates
Acyclovir, indinavir precipitates
Multiple myeloma
URETERAL
Intrinsic
Intraluminal
Nephrolithiasis
Papillary necrosis
Blood clots
Fungus balls
Intramural
Ureteropelvic junction dysfunction
Ureterovesical junction dysfunction
Ureteral valve, polyp, or tumor
Ureteral stricture
Schistosomiasis
Tuberculosis
Scarring from instrumentation
Drugs (e.g., nonsteroidal anti-inflammatory agents)
Extrinsic
Vascular system
Aneurysm: Abdominal aorta or iliac vessels
Aberrant vessels: Ureteropelvic junction
Venous: Retrocaval ureter
Gastrointestinal tract
Crohn disease
Diverticulitis
Appendiceal abscess
Colon cancer
Pancreatic tumor, abscess, or cyst
Reproductive system
Uterus: Pregnancy, prolapse, tumor, endometriosis
Ovary: Abscess, tumor, ovarian remnants
Gartner duct cyst, tubo-ovarian abscess
Retroperitoneal disease
Retroperitoneal fibrosis: Radiation, drugs, idiopathic
Inflammatory: Tuberculosis, sarcoidosis
Hematoma
Primary tumor (e.g., lymphoma, sarcoma)
Metastatic tumor (e.g., cervix, ovarian, bladder, colon)
Lymphocele
Pelvic lipomatosis
BLADDER
Neurogenic bladder
Diabetes mellitus
Spinal cord defect
Trauma
Multiple sclerosis
Stroke
Parkinson disease
Spinal anesthesia
Anticholinergics
Bladder neck dysfunction
Bladder calculus
Bladder cancer
URETHRA
Urethral stricture
Prostate hypertrophy or cancer
Obstruction from instrumentation

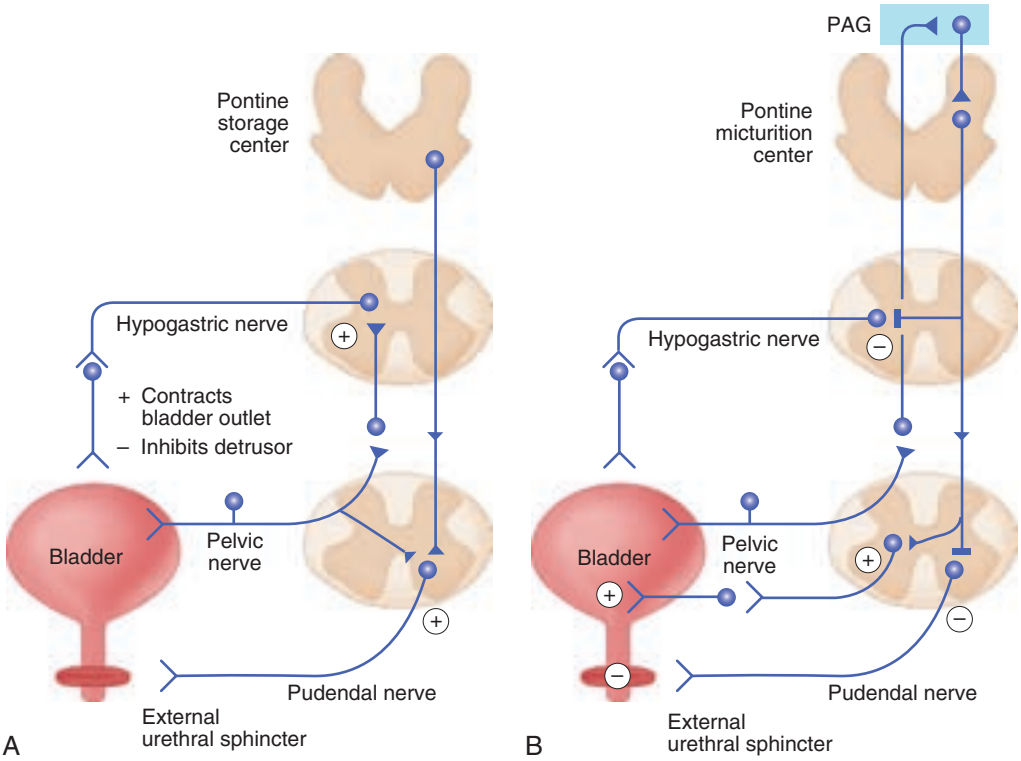

E-FIGURE 123-1. Neural circuits controlling continence and micturition. A, Reflexes mediating urine storage and continence. As the bladder fills, distention stimulates low-level firing of vesical afferents (pelvic nerve), which in turn stimulate sympathetic outflow to the bladder outlet (hypogastric nerve to contract the internal sphincter and inhibit detrusor activity) and pudendal outflow to the external urethral sphincter. These responses occur by spinal reflex pathways that promote continence. The pontine storage center in the rostral pons augments pudendal nerve firing to enhance external urethral sphincter activity. **B,** Voiding reflexes. As the bladder fills, afferents fire more intensely and activate spinobulbospinal reflex pathways passing through the pontine micturition center. These reflexes stimulate parasympathetic outflow to the bladder and urethral smooth muscle (hypogastric nerve) and inhibit sympathetic and pudendal outflow to the urethral outlet. Ascending afferent input from the spinal cord may pass through the periaqueductal gray (PAG) matter before reaching the cortex, leading to the sensation of urgency. (Modified from DeGroat WC. Integrative control of the lower urinary tract: a preclinical perspective. *Br J Pharmacol.* 2006;147[Suppl 2]:S25-S40.)

causes of obstruction are grouped according to the organ system causing the obstruction.

Pathology and Pathophysiology

Acute obstruction of urine flow out of the nephron reversibly alters renal blood flow, glomerular filtration, and tubular function. Acute unilateral obstruction may cause minimal systemic clinical disturbance because, absent other disease, the contralateral kidney compensates for the loss of function in the affected kidney. Obstructive uropathy is most often partial and of prolonged duration; this chronic obstruction leads to fibrosis and permanent damage.

Because of ease of study in animal models, the pathophysiology of acute complete obstruction is better understood than is partial obstruction. In acute complete obstruction, glomerular filtration ceases and tubular transport is markedly reduced. Immediately after the onset of complete ureteral obstruction, blockage of urine flow markedly increases tubular intraluminal pressure, which is transmitted back to the glomerulus. Initial dilation of the afferent arteriole maintains glomerular filtration. However, local production of the potent vasoconstrictors angiotensin II and thromboxane A_2 soon decreases the renal blood flow, glomerular filtration pressure, and glomerular filtration rate (GFR). Angiotensin and thromboxane also contract glomerular mesangial cells, reducing the glomerular capillary bed surface area available for filtration. At the same time, prostaglandin E_2 and I_2 levels rise and attenuate the level of vasoconstriction.

Obstruction also shuts down the ability of renal tubules (including the proximal tubule, the medullary thick ascending limb of Henle, and the cortical and medullary collecting ducts) to absorb sodium, secrete potassium and acid, and concentrate and dilute the urine. Reduced tubular transport results from the local release of mediators, such as prostaglandin E_2, that inhibit transport, the local accumulation of macrophages, and the release of inflammatory mediators, as well as mechanisms intrinsic to tubular epithelial cells. When urine flow is halted or markedly slowed, reduced delivery of solutes to tubular cells slows the rate of apical sodium entry, resulting in reduced synthesis and deployment to the plasma membrane of crucial transporter proteins, such as Na^+, K^+-ATPase, and apical sodium entry pathways, such as the epithelial sodium channel and Na/K/Cl cotransporter. As obstruction becomes more prolonged, renal fibrosis and permanent damage ensue. In addition to attenuation of salt reabsorption, reduced solute reabsorption in the thick ascending limb leads to loss of high solute concentrations in the medullary interstitium. Obstruction also markedly reduces the synthesis and membrane trafficking of aquaporins, especially aquaporin 2. The combined impact of the absence of medullary solute accumulation and reduced aquaporin activity leads to an inability to concentrate and dilute the urine.

With bilateral complete obstruction, the loss of function of both kidneys leads to the accumulation of salt, water, and uremic toxins; acidosis; and hyperkalemia. Accumulation of salt and water leads to elevated levels of salt-wasting hormones, such as atrial natriuretic peptide, kinins, and prostaglandins, and reduced levels of salt-retaining hormones, such as angiotensin II, catecholamines, and aldosterone. If the kidneys have not been severely damaged by the obstruction, these hormonal changes act synergistically with the postobstructive state of the kidneys to enhance glomerular filtration and reduce tubular salt reabsorption after the obstruction is released.

Obstructive nephropathy markedly attenuates the ability of distal nephron segments to secrete potassium and acid, so it can lead to hyperkalemia (Chapter 117) and a non–anion gap metabolic acidosis (Chapter 118) in patients with chronic partial obstruction. With acidemia, failure to acidify the urine may be revealed by a high urine pH (>5.5) and a positive urine anion gap (urine sodium and potassium higher than urine chloride), which indicates distal nephron failure to excrete ammonium in urine. In elderly patients, especially those with azotemia, chronic partial obstruction is associated with hyporeninemic hypoaldosteronism. In this condition, hyperkalemia and non–anion gap metabolic acidosis result from a combination of inadequate aldosterone production for the level of potassium and blood pH and an inadequate tubular response to aldosterone secondary to tubular dysfunction.

Chronic partial urethral obstruction, such as that caused by prostatic hypertrophy in men, can lead to dilation and remodeling of the bladder. Under normal circumstances, as the bladder fills, stretch receptors in the bladder wall and possibly in the epithelium sense the filling. Signaling via afferents to brain stem centers transmits efferent impulses to inhibit bladder wall contraction, permitting the bladder to fill with a modest increase in hydrostatic pressure (see E-Fig. 123-1). These bladder-filling reflexes also tighten the internal urethral sphincter and allow the maintenance of continence without the need for voluntary contraction of the external sphincter. However, bladder filling to volumes of 200 to 300 mL in women and 300 to 400 mL in men activates additional stretch receptors, stimulating brain stem micturition centers (see E-Fig. 123-1). Efferents from these centers augment reflex contraction of the bladder detrusor musculature, relax the internal sphincter, and alert the cortex of the need to void. As the bladder fills further, the micturition reflex becomes stronger, the urge to void becomes uncomfortable and urgent, and the bladder begins to contract against the voluntary, external sphincter, rendering it difficult to maintain continence.

With chronic urethral obstruction, micturition requires higher contractile pressure, resulting in detrusor muscle hypertrophy. The bladder empties less completely, and residual volumes increase. Initially, retained urine owing to incomplete emptying diminishes the volume capacity of the bladder between micturitions, thereby resulting in frequency and nocturia. Over time, with bladder remodeling and changes in autonomic reflexes, the transition from bladder accommodation to the micturition reflex may be delayed and occur at increasingly higher bladder-filling volumes. When the micturition reflex is suddenly activated in patients whose bladders are dilated, urgency, dribbling, and frank incontinence ensue. Some of these same features occur in women with pelvic floor disturbances that impede normal bladder function. Bladder wall remodeling and elevated pressures on voiding may increase back pressure up the ureters and result in the physiologic changes of chronic obstruction, including diminished ability to acidify and concentrate the urine, as well as reduced glomerular filtration.

The renal response to the release of obstruction depends on several factors, including whether the obstruction is unilateral or bilateral and the extent and duration of the obstruction. Release of acute unilateral obstruction leads to gradual reversal of renal vasoconstriction and rapid recovery of the glomerular filtration rate. Because tubular transport mechanisms may still be inhibited, postobstructive salt wasting, inability to secrete potassium and acid, and inability to concentrate and dilute the urine persist and lead to the production of a high quantity of isosthenuric urine (urine with a tonicity similar to that of plasma) from the affected kidney. However, the normal contralateral kidney compensates for these abnormalities in tubular transport. Release of an acute bilateral obstruction can lead to high volumes of urine output and striking salt wasting.

CLINICAL MANIFESTATIONS

The clinical appearance of obstructive uropathy depends on the extent (partial or complete), duration (acute or chronic), and location of the obstruction, as well as on whether one or both kidneys are affected (Table 123-2). Patients may be asymptomatic even with severe obstruction, especially when the obstruction has developed gradually.

TABLE 123-2 CLINICAL MANIFESTATIONS AND LABORATORY FINDINGS IN URINARY TRACT OBSTRUCTION

No symptoms (chronic hydronephrosis)
Intermittent pain (chronic hydronephrosis)
Elevated levels of blood urea nitrogen and serum creatinine with no other symptoms (chronic hydronephrosis)
Renal colic (usually caused by ureteral stones or papillary necrosis)
Changes in urinary output
Anuria or oliguria (acute renal failure)
Polyuria (incomplete or partial obstruction)
Fluctuating urinary output
Hematuria
Palpable masses
Flank (hydronephrotic kidney, usually in infants)
Suprapubic (distended bladder)
Hypertension
Volume dependent (usually caused by chronic bilateral obstruction)
Renin dependent (usually caused by acute unilateral obstruction)
Repeated urinary tract infections or infection refractory to treatment
Hyperkalemic, hyperchloremic acidosis (usually caused by defective tubular secretion of hydrogen and potassium)
Hypernatremia (seen in infants with partial obstruction and polyuria)
Polycythemia (increased renal production of erythropoietin)
Lower urinary tract symptoms: Hesitancy, urgency, incontinence, postvoid dribbling, decreased force and caliber of the urinary stream, nocturia

Symptoms

In patients with acute obstruction, bladder distention caused by the inability to relax the urethral sphincter (e.g., postoperatively) gives rise to sharp pain. By contrast, in the setting of gradual urethral obstruction, as can occur from prostatic enlargement (Chapter 129), the bladder may be able to fill to enormous volumes without causing significant pain.

Renal colic from the abrupt distention of the ureter is a common manifestation of the passage of a renal calculus (Chapter 126), with or without acute ureteral obstruction. Renal colic is a severe, stabbing pain localized to the flank (when the stone is in the upper third of the ureter) or radiating to the groin or pelvic structures (when the stone is located in the lower two thirds of the ureter).

Patients with chronic partial obstruction may have no symptoms or may have intermittent pain. Chronic partial ureteral obstruction may cause intermittent flank pain. Abdominal pain that radiates to the flank during voiding may indicate vesicoureteral reflux. Increasing the urine volume with the administration of fluid loads or diuretics may elicit pain in patients with partial obstruction by stretching the ureteral wall.

In patients with complete bilateral ureteral obstruction, complete obstruction of a solitary functioning kidney, or complete obstruction to urine flow beyond the bladder, anuric acute renal failure occurs. Patients with partial obstruction may have normal urine volumes or polyuria. In some cases, partial obstruction prevents urinary concentration, leading to polyuria, increased thirst, and sometimes hypernatremia. Although unusual, a history of oligoanuria alternating with polyuria or the sudden onset of anuria strongly suggests obstructive uropathy.

Bilateral complete obstruction or complete obstruction of a single functioning kidney may cause signs, symptoms, and laboratory evidence of acute renal failure (Chapter 120), with volume overload, hypertension, and metabolic disturbances. By contrast, unilateral obstruction with a functioning contralateral kidney usually does not lead to manifestations of renal failure because the functioning kidney may compensate in large part for the failure of filtration and tubular transport in the obstructed kidney. If the obstruction affects both kidneys, patients may have symptoms of impaired renal function, including nocturia and polyuria from failure to concentrate the urine and increased levels of potassium, phosphate, creatinine, and blood urea nitrogen (Chapter 120).

Given the complexity of bladder filling and emptying, a variety of disorders of these processes can lead to distressing lower urinary tract symptoms, such as incontinence (Chapter 26), urgency, frequency, and dysuria. These disorders, which include interstitial cystitis and urethral symptoms, are present in varying severity in 20 to 40% of adults over age 50 years.

Symptoms such as reduced force and caliber of the urine stream, urinary frequency, hesitancy, incontinence, nocturia, postvoid dribbling, and urgency often arise with urethral obstruction. Neurogenic bladder may alter micturition and result in frequency, urgency, and incontinence. Incontinence (Chapter 26) may occur because an inadequate sensation of bladder fullness or an inability to void properly leads to overfilling of the bladder and reflex emptying (overflow incontinence).

Physical Examination

In patients with acute ureteral obstruction, the physical examination may be normal or it may reveal flank tenderness. Flank tenderness can denote obstruction or pyelonephritis (Chapter 284). Renal colic may cause abdominal distention and evidence of reduced peristalsis with diminished bowel sounds. Kidneys enlarged by chronic hydronephrosis may be palpable on the abdominal examination or may cause costovertebral angle tenderness and flank rigidity. In patients with acute obstruction below the bladder, acute bladder distention may be detectable as a suprapubic mass, and the bladder may be tender. In the setting of bladder distention, the rectal examination in men may reveal prostatic enlargement, whereas the pelvic examination in women may reveal pelvic masses. Obstructive uropathy may cause hypertension owing to salt and water retention. Examination of the sensory and reflex pathways of the sacral nerves may reveal neurologic causes of urinary retention.

Laboratory Findings

Patients with obstruction may exhibit gross hematuria, especially in the setting of ureteral stones, which may abrade the urothelium and cause bleeding as they pass. Microscopic examination of the urine reveals round, regular red blood cells, which can be distinguished from the dysmorphic red blood cells that are typically seen in the hematuria of glomerular disease, in which red cells cross the glomerulus and remain in the tubules for a prolonged period. Gross hematuria from any cause may lead to clots, which themselves can cause obstruction.

A urinary tract infection in a younger man or repeated urinary tract infections in women without apparent cause suggest a structural lesion in the urinary tract (Chapter 128) and may be associated with partial or complete obstruction. Infection occurs more commonly in patients with obstruction involving the bladder or urethra, likely as a result of disruption of normal defenses against bacterial access and adherence to the bladder urothelium. The finding of unusual organisms (e.g., *Pseudomonas* or *Proteus* sp) in noninstrumented patients suggests the disruption of normal defense mechanisms and possible obstruction.

Depending on the extent and duration of obstruction, obstructive uropathy impairs renal function. Hyperkalemia and a nonanion acidosis, owing to distal renal tubular acidosis (Chapter 118), commonly develop. Chronic and more complete obstruction causes permanent renal damage and can lead to ESRD. Any patient with no previous history of kidney disease who has significant renal impairment should be evaluated for obstructive uropathy, especially if the urinary sediment is bland (Chapter 120). In addition, obstruction should be considered as a potential cause of accelerating deterioration of renal function in patients with underlying disease. In some obstructed kidneys, vasoconstriction may reduce cortical blood flow and oxygen tension, leading to increased erythropoietin production and polycythemia, which reverses with relief of the obstruction.

DIAGNOSIS

Because obstructive uropathy may be asymptomatic or may manifest in many different ways, the diagnosis may not be apparent. However, early diagnosis (Table 123-3) and prompt treatment reduce the extent of long-term renal damage.

Age, gender, and concomitant conditions often help identify the cause of obstruction. In children, congenital sources of obstruction at the ureteropelvic or ureterovesical junction are a major cause of ESRD. In adult women, complications of pregnancy or reproductive malignancies such as cervical or uterine cancer may cause obstruction resulting from compression of the ureters or ureterovesical junction. Infravesical obstruction explains less than 50% of lower urinary tract symptoms in men. In older men, prostatic hypertrophy (Chapter 129) or cancer (Chapter 201) often causes urethral obstruction.[1]

In outpatients, a history of renal colic, flank pain, or hematuria may suggest stone disease leading to ureteral obstruction. Changes in the volume or frequency of urination, including anuria, polyuria, or swings from oligoanuria to polyuria, may suggest obstruction. Bladder dysfunction is suggested by symptoms of frequency, urgency, and nocturia. In addition, a history of conditions that predispose to obstructive uropathy, such as sickle cell disease (Chapter 163), chronic ingestion of high levels of pain relievers (papillary necrosis; Chapter 125), previous stone disease (Chapter 126), or abdominal cancer (which may lead to ureteral obstruction), should raise the suspicion

TABLE 123-3 DIAGNOSTIC TESTS FOR OBSTRUCTIVE UROPATHY

UPPER URINARY TRACT OBSTRUCTION
Sonography (ultrasound)
Plain films of the abdomen (KUB)
Excretory or intravenous pyelography
Retrograde pyelography
Isotopic renography
Computed tomography
Magnetic resonance imaging
Pressure flow studies (Whitaker test)
LOWER URINARY TRACT OBSTRUCTION
Computed tomography
Magnetic resonance imaging
Cystoscopy
Voiding cystourethrography
Retrograde urethrography
Urodynamic tests
Cystometrography
Electromyography
Urethral pressure profile

From Klahr S. Obstructive uropathy. In: Jacobson HR, Striker GE, Klahr S, eds. *The Principles and Practice of Nephrology*. Toronto: BC Decker; 1991:432-441.
KUB = kidneys, ureter, bladder.

of obstruction. Finally, the presence of a single functioning kidney should raise the possibility that unilateral obstruction may be causing azotemia. In the inpatient setting, monitoring the pattern of urine output may reveal oligoanuria or polyuria.

Laboratory Studies

Initial laboratory evaluation includes a careful urinalysis and standard chemistry panel. The urine may reveal hematuria in the case of stones, bacteriuria and numerous granulocytes in the setting of obstruction or infection, or a urine pH greater than 7.5 in the case of chronic infection with urea-splitting organisms. Serum chemistries may reveal hyperkalemia, non–anion gap acidosis, and, more rarely, hypernatremia. Corresponding urine chemistry evaluation may reveal a pH higher than 5.5, lack of a negative anion gap (see earlier), and isosthenuria. The urine sediment also may reveal evidence of crystals (uric acid or calcium oxalate) that suggest stone disease. In addition, laboratory measurements should include blood urea nitrogen and creatinine levels to assess the adequacy of glomerular filtration.

DIAGNOSTIC TESTING

When obstructive uropathy is suspected, ultrasonography is the best screening modality because it is highly sensitive, safe, and inexpensive and does not expose the patient to contrast material or ionizing radiation. Because of its safety and low cost, ultrasonography is often used in patients with acute renal failure to exclude obstruction, although ultrasound performed in the absence of clinical suspicion of obstruction rarely finds a clinically significant obstructive nephropathy.[2] Ultrasound may reveal dilation of the calyces, renal pelvis, and, on occasion, proximal ureter (Fig. 123-1). It is also the preferred test to diagnose a renal calculus (see Fig. 126-2, Chapter 126). In patients with stones, measurement of the renal resistive index by bilateral color Doppler ultrasound can help predict which patients will progress to ureteral dilatation and obstruction.[3] False-positive findings (dilation in the absence of obstruction) occur in patients with congenital anomalies, during diuresis, and in many patients with ileal conduits. False-negative findings may occur because the pelvis and calyces fail to dilate despite obstruction, as may occur with retroperitoneal fibrosis or volume depletion. In such situations, the addition of color Doppler ultrasound can sometimes detect obstruction and its causes.[4] Because of the possibility of false-negative ultrasound results, computed tomography (CT) is warranted when the clinical setting strongly suggests obstruction and the ultrasound findings are negative. CT scanning may define the anatomic location of obstruction in patients found to have hydronephrosis on ultrasound, or it may identify obstruction in patients with negative ultrasound studies. In the setting of cancer or other structural lesions obstructing the ureters or invading the bladder, CT may help identify the cause of obstruction. CT, or in some centers magnetic resonance imaging (MRI), is the optimal method for diagnosing retroperitoneal fibrosis (Fig. 123-2), a rare condition that can cause partial or total bilateral ureteral obstruction.

Diffusion-weighted MRI can noninvasively detect changes in renal perfusion and in tissue density that occur during acute ureteral obstruction, and MRI is highly sensitive for detecting suspected obstructive uropathy.[5] However, the use of MRI is limited by the risk for nephrogenic systemic fibrosis in patients who are azotemic, especially if the glomerular filtration rate is below 30 mL/minute. Definitive diagnosis of the location of obstruction can be obtained by retrograde pyelography, in which contrast material is injected directly into the ureters via catheters inserted into the urethra and bladder, or by antegrade pyelography, in which the contrast material is injected into the renal pelvis via a percutaneous catheter. Retrograde pyelography is performed when obstruction has been diagnosed or is strongly suggested. This procedure precisely localizes the site of obstruction and guides the urologist in the placement of stents to clear the obstruction.

When bladder dysfunction or lesions in the bladder have been identified, retrograde cystograms may define bladder anatomy. In addition, cystometrography with urodynamic testing can define the force of detrusor function, determine whether the detrusor and sphincter act in a coordinated fashion (lack of coordination is referred to as *dyssynergy*), and define the extent to which pressure within the bladder is elevated and is causing obstruction to urine flow (Chapter 26).

DIFFERENTIAL DIAGNOSIS

Because obstructive uropathy may have subtle manifestations that mimic many other conditions, the differential diagnosis depends on the initial clinical symptoms and signs. Though suggestive of obstructive uropathy, anuria and acute renal failure (Chapter 120) may result from intrarenal diseases such as glomerulonephritis or acute tubular necrosis. Patients with polyuria, hypernatremia, and dilute urine may have nephrogenic or central diabetes insipidus (Chapter 225). Obstructive uropathy is a rare cause of nephrogenic diabetes insipidus. Patients with hyperkalemic hyperchloremic metabolic acidosis may have hyporeninemic hypoaldosteronism, which is associated with chronic mild obstruction or other tubular disorders (Chapter 122). Renal colic may resemble abdominal pain secondary to diseases of the gastrointestinal or reproductive tract, such as appendicitis (Chapter 142) or an ovarian cyst, especially when the colic is associated with nausea, vomiting, and diaphoresis.

In some patients, the cause of disturbing lower urinary tract symptoms may be elusive. Men should be evaluated for prostatic disease (Chapter 129), and both men and women should be evaluated in terms of their bladder function (Chapter 26).[6]

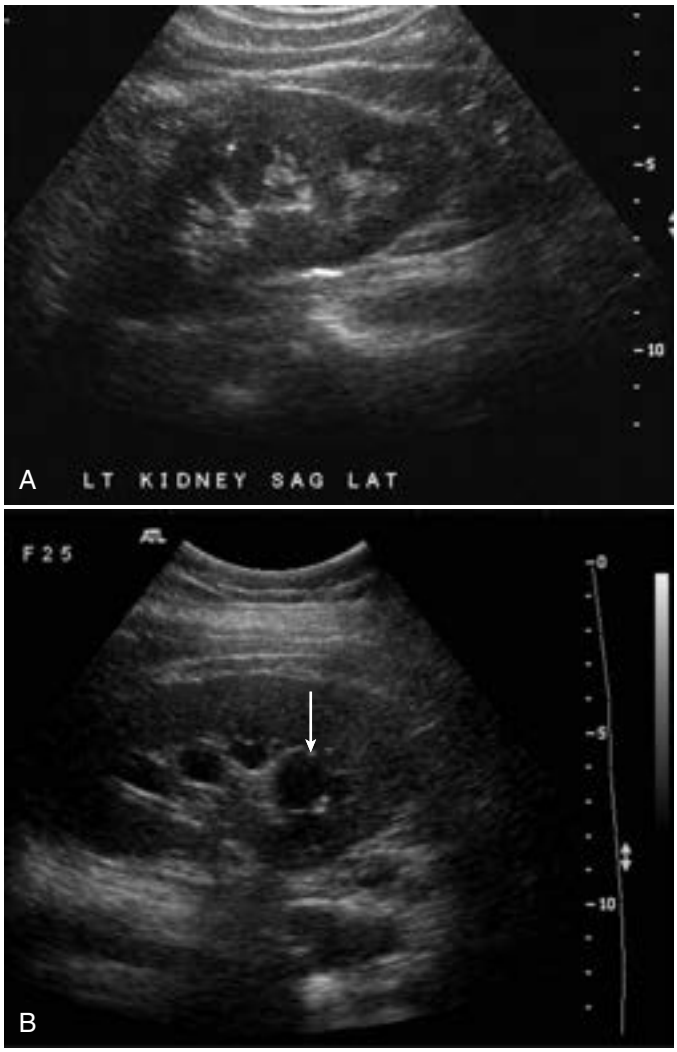

FIGURE 123-1. A, Renal ultrasound of a normal kidney. The outline of the kidney is clearly seen, and the calyces (*darker areas*) are small and somewhat indistinct. **B, Renal ultrasound of an obstructed kidney.** The *arrow* points directly at a dilated renal calyx. Other dilated calyces are seen as large, round, dark areas adjacent to the calyx indicated by the *arrow*. For orientation, the tail end of the *arrow* overlies the margin of the renal cortex. (Courtesy Jonathan Kruskal, MD, PhD.)

TREATMENT

Rx

Once obstructive nephropathy has been identified, therapy focuses on the rapid restoration of normal urine flow, treatment of any accompanying infection, and management of postobstructive complications. The degree to which renal function recovers depends on several factors, including the extent and duration of the obstruction and the extent of previous renal dysfunction.

Acute Obstruction

Complete obstruction causes acute renal failure. Because the extent and rate of recovery of renal function depend on the speed of relief, prompt

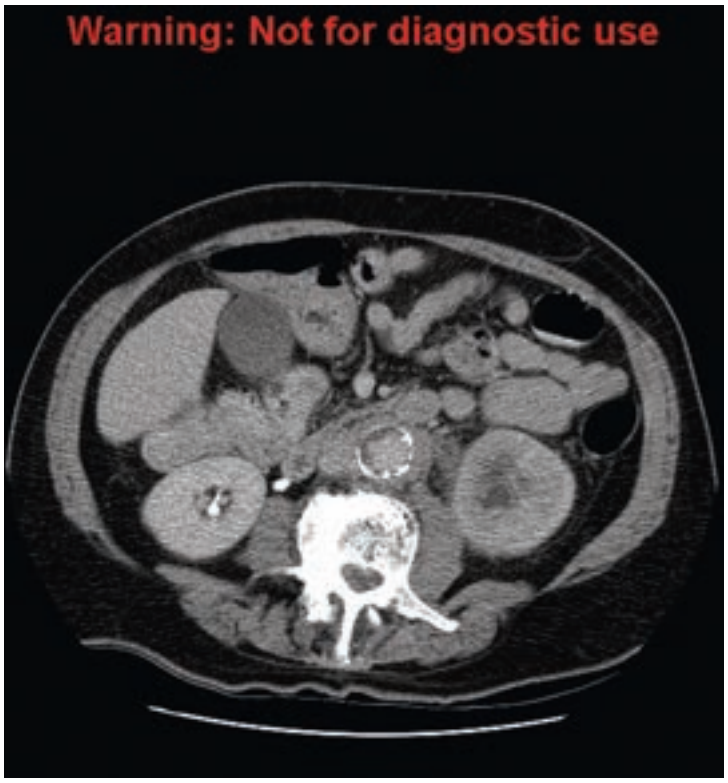

FIGURE 123-2. Contrast-enhanced axial computed tomography scan through the mid-abdomen, showing retroperitoneal fibrosis encircling the aorta and causing left-sided hydronephrosis owing to ureteral obstruction. (Courtesy of Jonathan Kruskal, MD, PhD.)

resolution of obstruction obviates the complications of uremia and the need for acute dialysis in patients with bilateral obstruction or obstruction of a single functioning kidney. In the setting of antecedent renal disease, partial obstruction may lead to permanent renal damage, so prompt relief can salvage significant renal function. In all cases of obstruction, the urine should be examined and cultured to identify and treat infections (Chapter 284). In patients with urinary sepsis and obstruction, the sepsis cannot be treated successfully until the obstruction is relieved; in such patients, it is also crucial to look for perinephric abscesses and drain them if present.

The site of the obstruction and its cause determine the therapeutic approach. Obstruction in the urethra or owing to bladder dysfunction may be relieved by placement of a urethral catheter. If catheters cannot be passed through the urethra, urgent suprapubic cystotomy is needed, followed by a more permanent approach, such as surgical diversion or ileal conduits, to prevent recurrent obstruction. If the obstruction is in the upper urinary tract, retrograde ureteral catheters with stents or nephrostomy tubes may be needed to relieve the obstruction. Retrograde catheters have the advantage that internal stents can be left in place to restore normal voiding, avoiding the need to maintain percutaneous drainage tubes.

Acute Obstruction Caused by Calculi

Calculi (Chapter 126) are the most common cause of ureteral obstruction. The cornerstones of therapy include analgesia, relief of the obstruction, and treatment of concomitant infections (Chapter 284). Stones 5 mm or less often pass without procedural intervention, but larger (7 to 15 mm) stones are more likely to cause complete obstruction and are progressively less likely to pass without intervention. If the stone is above the pelvic brim and is less than 15 mm, extracorporeal shock wave or ultrasonic lithotripsy is 90% effective, with passage of the fragments within 3 months. It is important to increase the volume of urine flow after these approaches to help the patient pass the fragments. For stones located below the pelvic brim or for larger stones, endoureteroscopy with direct removal may be performed via catheters passed through the urethra. In all patients with stone disease, it is crucial to identify the cause and initiate appropriate measures to prevent further stones (Chapter 126). The common practice of routinely placing a ureteral stent after ureteroscopy increases irritative lower urinary symptoms without any demonstrable clinical benefit.

Chronic Partial Obstruction

Although patients with chronic partial obstruction may do well for prolonged periods, the obstructive process should be relieved because it poses a long-term threat to renal function. Prompt relief is mandatory when partial obstruction progresses to frank urinary retention, the obstruction is accompanied by urinary sepsis or repeated urinary tract infections, the obstruction is causing renal damage, or the patient has symptoms such as voiding dysfunction, flank pain, or dysuria. Most often, chronic partial obstruction results from lesions in the lower urinary tract, including urethral blockage from prostate enlargement (Chapter 129). In men, benign prostatic hypertrophy, which may remain stable for long periods, usually responds to medications, but therapeutic decisions, including surgery, depend on symptoms, the presence of infection, and the risk for permanent bladder or renal dysfunction. The possibility of prostate cancer also must be considered (Chapter 201).

Idiopathic retroperitoneal fibrosis, which is an unusual cause of chronic ureteral obstruction, can progress to complete acute obstruction.[7] Treatment has not been definitively established, but prednisone (1 mg/kg/day tapered over 6 months), has been shown to be effective in up to 90% of patients for avoiding or delaying interventional treatments.

Chronic obstruction at the bladder neck or urethra can lead to bladder dilation and remodeling, with attendant persistence of dysfunction and symptoms even after relief of the obstruction. On this basis, it may be appropriate to relieve the obstruction before infection, major symptoms, or renal dysfunction occurs. Urethral strictures can be treated by dilation or urethrotomy.[8]

Postobstructive Diuresis

Though usually self-limited, postobstructive diuresis can last several days to a week and may result in clinically important depletion of sodium, potassium, and chloride. Because postobstructive diuresis is prolonged and promoted by excessive fluid replacement, administration of volume is justified only when excessive losses result in clear volume depletion. Proper replacement is guided by measurement of urine chemistries and osmolality. Because the urine is generally isosthenuric, with relatively high sodium levels as a result of residual tubular dysfunction, appropriate replacement fluid is often 0.45% saline given at a rate somewhat slower than that of urine output. By careful monitoring of vital signs, volume status, urine output, and serum and urine chemistry and osmolality, coupled with judicious fluid replacement, the diuresis can be limited and will not cause serious volume or electrolyte abnormalities. As basic research continues to unravel the signaling mechanisms that induce fibrosis rather than repair in obstructed kidneys, new therapeutic approaches may be possible after obstruction.[9]

Lower Urinary Tract Symptoms

Because lower urinary tract symptoms in the absence of demonstrable obstruction, infection (Chapter 284), or prostatic disease (Chapter 129) are so poorly understood, therapy is empirical and often ineffective.[10] Options in men usually focus on treatments for prostatic hyperplasia.[11] For women, treatments tend to focus on bladder function (Chapter 26) and menopause-related changes in the urogenital tract (Chapter 240).[12]

PROGNOSIS

The recovery of renal function depends on the duration and completeness of the obstruction. If the obstruction involves only one kidney and the other kidney has relatively normal function, the preserved kidney will compensate for the loss of function in the obstructed kidney. If obstruction is of relatively short duration and is partial, renal function will likely improve, leaving the patient with no symptoms of renal failure. However, if obstruction is bilateral, complete, or near-complete and persists for a week or more, significant permanent renal damage ensues, particularly if renal function was impaired before the onset of obstruction. Vigorous postobstructive diuresis is associated with renal recovery, but it may take weeks for renal function to recover. Approximately 20% of patients may have persistent chronic renal failure, and 5 to 10%[13] may need chronic renal replacement therapy (Chapter 131). Patients with lower urinary tract symptoms of unknown cause often have persisting symptoms despite therapies aimed at prostatic or bladder disease.[14]

Grade A Reference

A1. Vaglio A, Palmisano A, Alberici F, et al. Prednisone versus tamoxifen in patients with idiopathic retroperitoneal fibrosis: an open-label randomised controlled trial. *Lancet.* 2011;378:338-346.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. D'Silva KA, Dahm P, Wong CL. Does this man with lower urinary tract symptoms have bladder outlet obstruction?: The Rational Clinical Examination: a systematic review. *JAMA*. 2014;312:535-542.
2. Podoll A, Walther C, Finkel K. Clinical utility of gray scale renal ultrasound in acute kidney injury. *BMC Nephrol*. 2013;14:188.
3. Piazzese EM, Mazzeo GI, Galipo S, et al. The renal resistive index as a predictor of acute hydronephrosis in patients with renal colic. *J Ultrasound*. 2012;15:239-246.
4. Pepe F, Pepe P. Color Doppler ultrasound (CDU) in the diagnosis of obstructive hydronephrosis in pregnant women. *Arch Gynecol Obstet*. 2013;288:489-493.
5. Muthusami P, Bhuvaneswari V, Elangovan S, et al. The role of static magnetic resonance urography in the evaluation of obstructive uropathy. *Urology*. 2013;81:623-627.
6. Mangera A, Chapple C. Modern evaluation of lower urinary tract symptoms in 2014. *Curr Opin Urol*. 2014;24:15-20.
7. Scheel PJ Jr, Feeley N. Retroperitoneal fibrosis. *Rheum Dis Clin North Am*. 2013;39:365-381.
8. Mertens S, Zeegers AG, Wertheimer PA, et al. Efficacy and complications of urinary drainage procedures in idiopathic retroperitoneal fibrosis complicated by extrinsic ureteral obstruction. *Int J Urol*. 2014;21:283-288.
9. Ito I, Waku T, Aoki M, et al. A nonclassical vitamin D receptor pathway suppresses renal fibrosis. *J Clin Invest*. 2013;123:4579-4594.
10. Silva J, Silva CM, Cruz F. Current medical treatment of lower urinary tract symptoms/BPH: do we have a standard? *Curr Opin Urol*. 2014;24:21-28.
11. Fullhase C, Soler R, Gratzke C. New strategies in treating male lower urinary tract symptoms. *Curr Opin Urol*. 2014;24:29-35.
12. Singh S, van Herwijnen I, Phillips C. The management of lower urogenital changes in the menopause. *Menopause Int*. 2013;19:77-81.
13. Hamdi A, Hajage D, Van Glabeke E, et al. Severe post-renal acute kidney injury, post-obstructive diuresis and renal recovery. *BJU Int*. 2012;110:E1027-E1034.
14. Maserejian NN, Chen S, Chiu GR, et al. Treatment status and progression or regression of lower urinary tract symptoms in a general adult population sample. *J Urol*. 2014;191:107-113.

124

DIABETES AND THE KIDNEY

RAYMOND C. HARRIS

EPIDEMIOLOGY

In the industrialized world, diabetes mellitus is the single leading cause of end-stage renal disease (ESRD). Despite the improved care of patients with diabetes, both the incidence and prevalence of ESRD secondary to diabetes continue to rise. In the United States, more than 30% of patients undergoing either dialytic therapy or renal transplantation have ESRD as a result of diabetic nephropathy, and 40% of the new (incident) cases of ESRD are attributable to diabetes. Currently, more than 200,000 patients receive ESRD care as a result of diabetic nephropathy.

In the United States, Europe, and Japan, more than 90% of patients with diabetes have type 2 rather than insulinopenic type 1 diabetes (Chapter 229). The incidence of renal disease is equivalent, and more than 80% of the ESRD secondary to diabetes is also seen in patients with type 2 diabetes. Although it was previously supposed that ESRD secondary to type 2 diabetes was less common than with type 1 diabetes, when cohorts of patients with type 1 and type 2 diabetes are monitored for an extended period, the degree of renal involvement is similar. The demographics of ESRD secondary to type 2 diabetes mirror the prevalence of type 2 diabetes in the U.S. population, with a higher incidence in women and in African Americans, Hispanic Americans, Native Americans, and Asian Americans and a peak incidence in the fifth to seventh decade. Much of the increased mortality in type 2 diabetes is associated with the prevalence of nephropathy.[1] Given the global epidemic of obesity in developed countries, an increasing incidence of diabetic nephropathy is being widely appreciated. Smoking and elevated blood cholesterol may be predisposing factors for the development of diabetic nephropathy in type 2 patients with diabetes.

PATHOBIOLOGY

Hyperglycemia

The metabolic sequelae of hyperglycemia appears to be the most important causative factor in the development of diabetic nephropathy. Hyperglycemia leads to increased generation of reactive oxygen species; depletion of the reduced form of nicotinamide dinucleotide (phosphate); activation of the polyol pathway, which can lead to de novo synthesis of diacylglycerol and increased protein kinase C activity; alterations in the hexosamine pathway; and nonenzymatic protein glycation (advanced glycosylation end products), all of which have been implicated in the development of diabetic nephropathy, as well as other diabetic microvasculopathies. Although this can be marked by individual variations, better glucose control generally reduces the risk of nephropathy and other microvascular complications. For example, in a cohort of patients who had glucokinase mutations, which are associated with milder hyperglycemia (average hemoglobin A_{1c} [HbA_{1c}] levels, 6.9%), they had far less proteinuria, microalbuminuria, or nephropathy than patients with long-standing type 2 diabetes with HbA_{1c} levels averaging 7.8%.[2] Furthermore, randomized interventional studies clearly demonstrate that relatively better control of blood sugar decreases the development of nephropathy in type 1 diabetes, and observational studies with repeat renal biopsies show that the renal lesions of diabetic nephropathy may reverse after long-term functioning pancreas transplantation.

Hemodynamics

Patients with type 1 and, to a lesser extent, type 2 diabetes exhibit an increased glomerular filtration rate (GFR), so-called hyperfiltration, that is mediated by proportionately greater relaxation of the afferent arteriole than the efferent arteriole. This hyperfiltration leads to increased glomerular blood flow and elevated glomerular capillary pressure. With poorly controlled diabetes, patients also develop glomerular hypertrophy, with an increased glomerular capillary surface area. These intraglomerular hemodynamic and structural alterations may contribute to the development or progression (or both) of diabetic renal injury. Because angiotensin-converting enzyme (ACE) inhibitors and decreased dietary protein reduce this elevated intraglomerular capillary pressure in experimental animals, the hyperfiltration hypothesis provides one rationale for the success of these interventions in resisting the progression of diabetic nephropathy (see later discussion).

Hormones and Cytokines

Studies in experimental animals have implicated a number of cytokines, hormones, and intracellular signaling pathways in either development or progression of diabetic nephropathy, notably transforming growth factor ß, connective tissue growth factor, angiotensin II, vascular endothelial growth factor, endothelin, prostaglandins, and nitric oxide. Because these factors have also been implicated in a variety of nondiabetic kidney diseases, it is likely that they will not prove to be specific for diabetic nephropathy. However, agents that interrupt angiotensin II production and signaling have proven to be very effective in slowing the progression of diabetic nephropathy. Furthermore, agents that interrupt intracellular pathways activated by these factors or by other consequences of hyperglycemia may provide future therapeutic opportunities.

Genetics

At present, it is not possible to predict which patients will develop diabetic nephropathy. Although poor glycemic and blood pressure control undoubtedly contribute, nephropathy may or may not develop in an individual patient even after many years of hypertension and hyperglycemia. Both type 1 and type 2 diabetes cluster in families. Those with type 1 diabetes with siblings who have diabetic nephropathy have a greater than 70% lifetime risk of diabetic nephropathy developing in themselves. Patients with type 2 diabetes also appear to have a hereditary predisposition for or against the development of diabetic nephropathy.

However, diabetic nephropathy is likely to be a polygenic disease, and its development and progression are likely related to the inheritance of multiple polymorphisms with variable effect sizes.[3] For example, African Americans with the apolipoprotein-1 gene do not have an increased predisposition to develop diabetic nephropathy, but nephropathy will have an accelerated progression in such patients.[4] In addition, the lack of nephropathy in patients with glucokinase mutations may be more pronounced than can be explained just by their relatively lower HgA_{1c} levels. Studies also suggest a long-term programming or memory effect in the development of diabetic kidney disease, such that patients whose type 1 diabetes was poorly controlled in the past will develop nephropathy at an increased rate despite subsequent excellent glycemic control. These findings suggest the possible role for epigenetic programming in the development of diabetic nephropathy.[5]

CLINICAL MANIFESTATIONS

Natural History

Although a minority of patients with diabetic nephropathy have type 1 diabetes, the natural history of the disease is best exemplified in this population because the onset of diabetes is more clearly definable and typically occurs at an early enough age to permit long-term follow-up. Furthermore, patients with type 1 diabetes usually do not initially have comorbid essential hypertension, atherosclerotic cardiovascular disease, obesity, and other conditions that are often associated with type 2 diabetes and that may independently produce chronic renal injury. However, the similarity of the nephropathic progressions in type 1 and 2 diabetes is exemplified by the Pima Indians, who exhibit a strong genetic predisposition for the development of type 2 diabetes by the fourth decade of life and in whom the diabetic nephropathy progresses in a similar pattern as seen in type 1 diabetic patients.

Diabetic nephropathy progresses through four relatively distinct stages (Fig. 124-1).

Stage I

In stage I, which commences soon after the overt manifestations of diabetes, renal blood flow and GFR increase by up to 50%, and the kidneys' glomeruli and tubules hypertrophy compared with age- and weight-matched normal control subjects. Although patients with type 2 diabetes also tend to have an elevated GFR during the early course of their disease, the GFR increases are not usually as pronounced as seen with insulin-dependent diabetes mellitus. At this stage, no macroalbuminuria is detectable, but transient microalbuminuria can occasionally be measured by radioimmunoassay, enzyme-linked immunosorbent assay, or special dipsticks, especially when induced by stress, physical exertion, concurrent illness, or poor glycemic control. Hypertension is usually absent in the early stages in patients with type 1 diabetes but is present in 10% to 25% of type 2 patients with diabetes at their initial evaluations.

	Stage I	Stage II	Stage III	Stage IV
Median year of onset	0	10	15-17	18-20
Patients with diabetes (%)	100	30-35	30	30

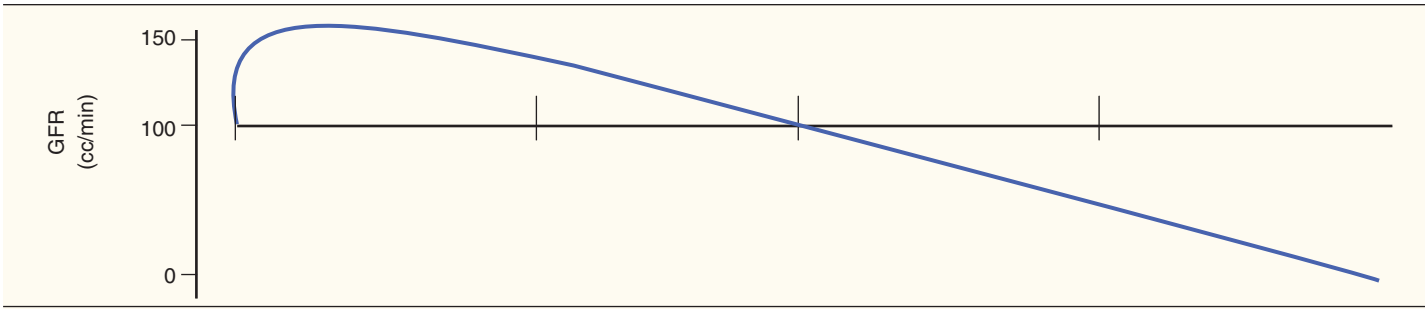

Urinary protein	Occasional and transient microalbuminuria	Fixed microalbuminuria	Proteinuria (>500 mg/24 hr) and microalbuminuria (200 mg/24 hr)	Nephrotic-range proteinuria (<3.5 g/24 hr)
Systemic manifestations	Hypertension: absent in type 1; often present in type 2		Hypertension: absent in type 1 and worsening in type 2	Manifestations of chronic renal insufficiency
Renal morphology and history	Kidney hypertrophy	Glomerular basement membrane thickening and mesangial matrix expansion	Focal glomerulosclerosis (± nodular or Kimmelstiel-Wilson lesions) Microvascular hyalinosis and tubulointerstitial fibrosis	Kidney may still be inappropriately large for level of renal insufficiency Global glomerulosclerosis and tubulointerstitial fibrosis

FIGURE 124-1. Stages of diabetic nephropathy. GFR = glomerular filtration rate.

Stage II

Approximately 30% of patients with type 1 diabetes progress to stage 2, which is characterized by fixed microalbuminuria of at least 30 mg/24 hr, after a median of about 10 years of diabetes. Although the GFR either remains elevated or is within the normal range at this stage, renal histology becomes abnormal and is manifested as glomerular and tubular basement membrane thickening and the inception of mesangial matrix expansion. Microalbuminuria is more likely in patients who have evidence of other microvascular insults, especially proliferative retinopathy. Microalbuminuria is a more specific sign of diabetic nephropathy in type 1 diabetes than in type 2 diabetes because of the high incidence of hypertension, which itself may lead to microalbuminuria, in the latter.

Stage III

The great majority of patients who are initially seen with fixed microalbuminuria progress to overt nephropathy (stage III) within 5 to 7 years. In this stage, patients have overt proteinuria (>500 mg of total protein per 24 hours) and macroalbuminuria (>200 mg/24 hr), which are detectable with a routine urinary protein dipstick. With the onset of stage III, estimated GFR (eGFR) is usually below normal levels for age and continues to decrease as the disease progresses. Blood pressure begins to rise in patients with type 1 diabetes with stage III nephropathy. In patients with type 2 diabetes, who frequently have preexistent hypertension, blood pressure commonly becomes more difficult to control.

Renal biopsy reveals diffuse or nodular (Kimmelstiel-Wilson) glomerulosclerosis. Although the Kimmelstiel-Wilson lesion is considered pathognomonic of advanced diabetic nephropathy, only approximately 25% of patients manifest this lesion. A nodular pattern of glomerulopathy mimicking Kimmelstiel-Wilson lesions may also be seen in light-chain nephropathy (Chapter 187), and historic descriptions of "diabetic nephropathy without overt hyperglycemia" based solely on light microscopic analysis actually may have represented light-chain disease. Nodular glomerular lesions can also be observed in amyloidosis (Chapter 188) and membranoproliferative glomerulonephritis type II (Chapter 121).

An additional pathognomonic feature of diabetic nephropathy is the finding of both afferent and efferent arteriolar hyalinosis, which can be distinguished from the isolated afferent arteriolar lesion of essential hypertension. In overt diabetic nephropathy, progressive tubulointerstitial fibrosis correlates most closely with the decline in renal function. The GFR begins to decline from the normal range, but the serum creatinine level may remain in the normal range.

Stage IV

Stage IV, or advanced diabetic nephropathy, is characterized by a relentless decline in renal function and progression to ESRD. Patients typically have heavy or nephrotic-range proteinuria (>3.5 g/24 hr) and systemic hypertension but have no evidence of inflammatory glomerular (red blood cell casts) or tubulointerstitial (white blood cells, white blood cell casts) lesions. The kidneys may be inappropriately large for the observed degree of renal insufficiency. However, a subset of patients with type 2 diabetes develops chronic kidney disease without nephrotic-range proteinuria. Whether this difference represents a fundamental difference in the pathophysiology of the two conditions or represents the synergistic effects of other kidney injuries, such as hypertensive renal disease, is unclear.

Other Renal Complications

Patients with diabetes also have an increased rate of other kidney and genitourinary abnormalities. Type IV (hyporeninemic, hypoaldosteronemic) metabolic acidosis (Chapter 118) with hyperkalemia is commonly encountered in patients with diabetes and mild to moderate renal insufficiency. These patients should be carefully monitored for the development of severe hyperkalemia (Chapter 117) in response to volume depletion or after the initiation of drugs that interfere with the renin-angiotensin system, such as ACE inhibitors, AT1 receptor blockers (ARBs), β-adrenergic blockers, both nonselective and selective cyclooxygenase-2 (COX-2) nonsteroidal anti-inflammatory agents, and heparin, as well as potassium-sparing diuretics.

Patients with diabetes have an increased incidence of bacterial and fungal infections of the genitourinary tract (Chapter 284). In addition to lower urinary tract infections, they have an increased risk for pyelonephritis and intrarenal and perinephric abscess formation (Chapter 284).

Unilateral or bilateral renal artery stenosis (Chapter 125) is more frequent in the type 2 diabetic population than in age-matched nondiabetic individuals and should be considered if a patient with diabetes has intractable hypertension or a rapidly rising serum creatinine level immediately after initiation of therapy with an ACE inhibitor or AT1 receptor blocker. Other causes of acute deterioration in renal function include papillary necrosis with ureteral obstruction owing to sloughing of a papilla, obstructive uropathy (Chapter 123) caused by bladder dysfunction as a result of autonomic neuropathy, and contrast media–induced acute tubular necrosis (Chapter 120). In addition, prerenal azotemia or acute tubular necrosis may develop in patients with diabetes as a result of heart failure or volume depletion owing to vomiting induced by gastroparesis (Chapter 229) or diarrhea from autonomic neuropathy.

Stage I	Tight glucose control BP control—consider use of ACEI or ARB
Stage II	Tight glucose control ACEI or ARB BP control Smoking cessation Weight reduction Exercise Annual eye examination
Stage III	ACEI or ARB BP control Restriction of dietary protein (to 0.8g/kg/day of ideal body weight) Antihyperlipidemic medications
Stage IV	Treat manifestations of nephrotic syndrome and chronic renal insufficiency Prepare for renal replacement therapy, including prevention of abnormalities in calcium and phosphorus metabolism and prevention of anemia by early use of erythropoietin

FIGURE 124-2. Treatment of diabetic nephropathy. ACEI = angiotensin-converting enzyme inhibitor; ARB = angiotensin-receptor blocker; BP = blood pressure.

DIAGNOSIS

The diagnosis of overt diabetic nephropathy is made by three main criteria: the presence of proteinuria within an appropriate time frame, the presence of retinopathy (found in 90%-95% of patients with type 1 diabetes and in 60%-65% of patients with type 2 diabetes), and the absence of other causes of nephrotic syndrome or renal insufficiency. For patients with type 1 diabetes who develop diabetic nephropathy, significant proteinuria is seen within 17 ± 6 years of the onset of diabetes. Although microalbuminuria does not always progress to overt proteinuria in these patients, it always precedes overt proteinuria in patients who do progress. The American Diabetes Association recommends screening all patients with type 1 diabetes for microalbuminuria 5 years after the diagnosis and yearly thereafter. For patients with type 2 diabetes, the recommendation is screening for microalbuminuria at the time of diagnosis and yearly thereafter.

PREVENTION AND TREATMENT

Rx

Glycemic control significantly lessens the incidence of nephropathy in patients with type 1 diabetes for at least 20 or more years[A1] but does not completely eliminate the risk (Fig. 124-2). However, tight glycemic control to reduce to a $HgBA_{1c}$ target level of 6.5% or less in patients with type 2 diabetes does not reduce the risk of nephropathy compared with standard therapy to a goal of 7% to 7.9%.[A2,A3]

Elevated blood pressure is an important risk factor in the progression of diabetic nephropathy, and it was previously believed that blood pressure goals should be lower than for the general population. However, recent studies have shown detrimental effects of low blood pressures in patients with diabetic nephropathy, and more moderate blood pressure control with systolic blood pressures of 130 to 140 mm Hg is advocated.[A4]

In latent (stage II) and overt (stage III) diabetic nephropathy, renal function declines. With declining renal function, oral hypoglycemic agents become contraindicated. Because of the increased risk of prolonged hypoglycemia, sulfonylureas are contraindicated in patients with eGFRs less than 45 mL/min. Because of the potentially increased risk of lactic acidosis with metformin therapy in patients with renal insufficiency, treatment guidelines currently recommend it be avoided in patients with a serum creatinine level greater than 1.7 mg/dl, although this strict guideline is currently debated and may be changed in the future. As GFR declines, insulin requirements may decrease owing to reduced insulin degradation and clearance by the failing kidney.

Medications that interfere with the renin–angiotensin system, either ACE inhibitors or ARBs, are the preferred agents and appear to have additional benefits beyond lowering systemic blood pressure.[A5,A6] However, combination therapy using both an ACE inhibitor and an ARB is contraindicated because of increased side effects.[A7]

Patients whose diabetic nephropathy is treated with ACE inhibitors or ARBs should have their serum potassium and creatinine levels monitored closely in the first week after the initiation of therapy because of their high prevalence of type IV renal tubular acidosis and renal artery stenosis. If blood pressure control is not achieved with these agents, diuretics and other antihypertensive agents, including cardioselective β-blockers, α-blockers, and nondihydropyridine calcium channel blockers, can be added (Chapter 67). Dihydropyridine calcium channel blockers induce selective afferent arteriolar vasodilation and may increase intraglomerular capillary pressure, so they are usually reserved for patients who need them to control blood pressure after other agents have failed.

Physicians should encourage smoking cessation (Chapter 32) and should prescribe statins for patients with hyperlipidemia (Chapter 206), since patients with diabetic nephropathy have a significantly increased risk of morbidity and mortality from cardiovascular disease.[6] Judicial restriction of dietary protein to 0.8 g/kg of ideal body weight per day is recommended by the American Diabetes Association. Although further dietary protein restriction may retard the progression of diabetic nephropathy, considerations of such restrictions must be balanced against the individual patient's nutritional requirements.

Renal Replacement Therapy

More than 80% of patients with end-stage diabetic nephropathy receive dialysis as their modality of renal replacement therapy (Chapter 131), with about five times as many undergoing hemodialysis compared with peritoneal dialysis. Because of their associated cardiovascular, cerebrovascular, and peripheral vascular disease as well as their increased risk for infection, the mortality rate of patients with diabetes who receive either type of dialysis is 1.5 to 2.0 times higher than in nondiabetic patients, corresponding to a 5-year survival rate of less than 20% in patients with diabetes undergoing maintenance dialysis. Outcomes are worse in patients whose HbA_{1c} levels are above 8.5%.[7]

In general, the management of a patient with diabetes nearing ESRD is similar to that of a nondiabetic patient (Chapter 130). Stage III patients should be under the care of a nephrologist, and planning should be initiated for the modality of dialysis. Although dialysis is generally initiated when the GFR declines to less than 10 mL/min, earlier initiation of dialysis is sometimes necessary in patient with diabetes because their volume-dependent hypertension or hyperkalemia is not otherwise manageable or when their uremia and gastroparesis lead to malnutrition or uncontrollable recurrent emesis.

Approximately 25% of renal transplants performed in the United States are in patient with diabetes, and more than 90% of these are in patient with type 1 diabetes because of their younger age and lesser degrees of macrovascular comorbidity. Long-term survival and quality of life are generally superior after transplantation compared with chronic dialysis. However, the other microvascular complications (retinopathy, neuropathy) are not improved by renal transplantation alone. Pancreas and combined kidney and pancreas transplantation can significantly improve the quality of life of patients with diabetic nephropathy by improving autonomic neuropathy, retarding or possibly correcting retinopathy, and avoiding the potential complications of insulin administration. However, all transplantation options remain limited by organ availability.

Grade A References

A1. de Boer IH, Sun W, Cleary PA, et al. Intensive diabetes therapy and glomerular filtration rate in type 1 diabetes. *N Engl J Med*. 2011;365:2366-2376.

A2. Gerstein HC, Miller ME, Byington RP, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008;358:2545-2559.

A3. Patel A, MacMahon S, Chalmers J, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2008;358:2560-2572.

A4. Cushman WC, Evans GW, Byington RP, et al. Effects of intensive blood-pressure control in type 2 diabetes mellitus. *N Engl J Med*. 2010;362:1575-1585.

A5. Lv J, Perkovic V, Foote CV, et al. Antihypertensive agents for preventing diabetic kidney disease. *Cochrane Database Syst Rev*. 2012;12:CD004136.

A6. Wu HY, Huang JW, Lin HJ, et al. Comparative effectiveness of renin-angiotensin system blockers and other antihypertensive drugs in patients with diabetes: systematic review and bayesian network meta-analysis. *BMJ*. 2013;347:f6008.

A7. Fried LF, Emanuele N, Zhang JH, et al. Combined angiotensin inhibition for the treatment of diabetic nephropathy. *N Engl J Med*. 2013;369:1892-1903.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Afkarian M, Sachs MC, Kestenbaum B, et al. Kidney disease and increased mortality risk in type 2 diabetes. *J Am Soc Nephrol*. 2013;24:302-308.
2. Steele AM, Shields BM, Wensley KJ, et al. Prevalence of vascular complications among patients with glucokinase mutations and prolonged, mild hyperglycemia. *JAMA*. 2014;311:279-286.
3. Reidy K, Kang H, Hostetter T, et al. Molecular mechanisms of diabetic kidney disease. *J Clin Invest*. 2014;124:2333-2340.
4. Parsa A, Kao WH, Xie D, et al. APOL1 risk variants, race, and progression of chronic kidney disease. *N Engl J Med*. 2013;369:2183-2196.
5. Kato M, Natarajan R. Diabetic nephropathy-emerging epigenetic mechanisms. *Nat Rev Nephrol*. 2014;10:518-530.
6. Haynes R, Staplin N, Emberson J, et al. Evaluating the contribution of the cause of kidney disease to prognosis in CKD: results from the Study of Heart and Renal Protection (SHARP). *Am J Kidney Dis*. 2014;64:40-48.
7. Hill CJ, Maxwell AP, Cardwell CR, et al. Glycated hemoglobin and risk of death in diabetic patients treated with hemodialysis: a meta-analysis. *Am J Kidney Dis*. 2014;63:84-94.

REVIEW QUESTIONS

1. A 64-year-old Hispanic man with moderately controlled type 2 diabetes, hypertension, and chronic kidney disease with nephrotic-range proteinuria and serum creatinine of 2.4 is admitted to the hospital because he has had a stroke. Which is the best single answer concerning this patient?

A. Intensive glycemic control of type 2 diabetes does not decrease the incidence of nephropathy or cardiovascular mortality
B. Patients with type 2 diabetes with chronic kidney disease always have increased albuminuria.
C. In type 2 diabetes, proteinuria correlates with an increased incidence of coronary heart disease but not with stroke.
D. In both type 1 and type 2 diabetes, hypertension usually indicates evidence of diabetic nephropathy
E. Nephrotic-range proteinuria is greater than 1.5 g/day of proteinuria.

Answer: A Intensive glycemic control of type 2 diabetes does not decrease the incidence of nephropathy and cardiovascular mortality. The ACCORD trial indicated an increased risk of overall mortality with intensive glucose lowering in this population. Recent meta-analyses have indicated a decrease in cardiovascular disease outcomes, specifically nonfatal myocardial infarctions and risk of progression of nephropathy, but could not confirm that there was any decrease in the incidence of nephropathy or cardiovascular mortality. B is false because about 20% of type 2 diabetic patients with chronic kidney disease do not have albuminuria. C is false because proteinuria correlates with an increased incidence of coronary heart disease and stroke in patients with type 2 diabetes. D is false because patients with type 2 diabetes are older and often have hypertension before the onset of diabetic kidney injury. E is false because the definition of nephrotic range proteinuria is greater than 2.5 g/day.

2. Which is the best single answer concerning structural changes to the kidney during the course of diabetic nephropathy?

A. Afferent and efferent arteriolar hyalinosis are nonspecific findings and are found in a variety of glomerular diseases.
B. More than 90% of patients with type 2 diabetes have evidence of retinopathy at the time of diagnosis of overt nephropathy.
C. Thickening of the glomerular basement membrane is a late finding in diabetic nephropathy.
D. Kimmelstiel-Wilson lesions are observed in a minority of patients diagnosed with diabetic nephropathy.
E. Kidney hypertrophy always indicates that the patient will develop overt nephropathy.

Answer: D Kimmelstiel-Wilson lesions are observed in a minority of patients diagnosed with diabetic nephropathy. A is false because the combination of afferent and efferent arteriolar hyalinosis is characteristic of diabetic nephropathy. B is false because only 60% to 65% of type 2 diabetic patients have retinopathy concomitant with onset of nephropathy. C is false because thickening of the glomerular basement membrane can be observed 2 to 5 years after the onset of diabetes and occurs before the onset of overt proteinuria or decline in renal function. E is false because kidney hypertrophy does not always indicate that the patient will develop overt nephropathy.

125 VASCULAR DISORDERS OF THE KIDNEY

THOMAS D. DUBOSE, JR., AND RENATO M. SANTOS

Vascular disorders that significantly alter renal perfusion can impact the glomerular filtration rate (GFR); tubular function; and, ultimately, kidney function. Stenosis, thrombosis, emboli, atherosclerosis, inflammation, and hypertension may involve the renal arteries, arterioles, microvasculature, and renal veins.

RENAL ARTERY STENOSIS

DEFINITION

Renal artery stenosis is the prototype for secondary hypertension. It is defined as narrowing of the renal arteries, resulting in hypoperfusion of one or both kidneys. The two most common forms are atherosclerotic disease and fibromuscular dysplasia, but other causes include vasculitis (Chapter 270), neurofibromatosis (Chapter 417), congenital bands, extrinsic compression, and radiation injury (Chapter 20).

EPIDEMIOLOGY

The prevalence of renal artery stenosis is estimated to be about 2% in the fifth decade of life rising to 20% to 25% in the ninth decade. The prevalence may be somewhat higher in African Americans and Latinos compared with whites.

An estimated 2% of all patients with systemic hypertension (Chapter 67) have renal artery stenosis,[1] although the prevalence is much higher in patients whose hypertension is resistant to multiple medications.[2] Atherosclerotic disease, which is the most common form of renal artery stenosis, accounts for approximately 90% of all lesions. As with other forms of atherosclerosis, it is more common in elderly adults and in patients with cardiovascular risk factors such as hypertension, hyperlipidemia, smoking, and diabetes (Chapter 229). It is also more common in patients with heart failure (Chapter 58), multivessel coronary disease (Chapter 71), and peripheral vascular disease (Chapters 79 and 80).

In contrast, fibromuscular dysplasia is more common in young patients, with 90% of cases occurring in women at a mean age of 52 years at diagnosis.[3] Its true prevalence is unknown, but in one study of more than 2500 individuals evaluated as possible kidney donors by renal computed tomography angiography (CTA), the prevalence was 26%, 31% of whom had previously diagnosed hypertension.[4] It is an important cause of treatable hypertension in young patients without cardiovascular risk factors and is pathologically and angiographically distinct from atherosclerotic disease. The cause remains unknown, but genetic, hormonal, and mechanical factors may predispose to this disorder.

PATHOBIOLOGY

The pathobioloy of atherosclerotic renal artery stenosis is identical to that of atherosclerotic disease in other arterial beds (Chapter 70). Fibromuscular dysplasia, by contrast, includes four distinct histopathologic types: medial fibroplasia, which is the most common type and accounts for 75% to 80% of cases; perimedial fibroplasia, with irregular thickening of the media; medial hyperplasia, with smooth muscle hyperplasia without fibrosis; and intimal fibroplasia. Unlike atherosclerotic disease, which localizes at ostial and proximal segments of the renal arteries, fibromuscular dysplasia more commonly involves the middle and distal arterial segments.

The renal hypoperfusion caused by renal artery stenosis activates the renin–angiotensin–aldosterone system (RAAS) and results in an increase in systemic blood pressure. This mechanism explains the ability of angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers (ARBs) to control the resulting hypertension in animal models and in patients with fibromuscular dysplasia. Although hypoperfusion-stimulated activation of the RAAS is essential for initiating the pressor response, this effect is transient. Over time, the pathophysiology of atherosclerotic disease transitions to pressor mechanisms independent of the RAAS, including vasoconstriction from oxidative stress, endothelial dysfunction, endothelin release, and sympathetic activation. Confounding risks—including smoking, advanced age, dyslipidemia, diabetes, and hypertension—can also contribute to vascular injury. This activation of inflammatory pathways combined with typical atherosclerotic risk factors contribute to chronic pressor mechanisms that may not respond to revascularization.

Ischemic nephropathy, which is defined as impairment of renal function beyond the decrease in perfusion typical of hemodynamically significant renal artery stenosis, is difficult to assess. Unlike cardiac or cerebral tissue, perfusion to the kidneys is primarily determined by glomerular ultrafiltration and exceeds its own metabolic needs by more than 10-fold. Therefore, a severe reduction in perfusion to the entire renal parenchyma is necessary to cause kidney injury. This relationship is supported by the observation that hemodynamically significant lesions of the fibromuscular dysplasia are rarely associated with renal dysfunction. Conversely, a decline in kidney function that is common in atherosclerotic renal artery stenosis may be caused in part by atherogenic stimuli, which can magnify oxidative stress and activate pro-inflammatory and profibrogenic pathways. Repetitive bouts of hypoperfusion in atherosclerotic disease may lead to renal tubular injury and, over time, can lead to tubulointerstitial fibrosis. Thus, a combination of repetitive perfusion insults superimposed on atherogenic risk factors, not hypoperfusion alone, likely explains renal dysfunction.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The clinical features of renal artery stenosis are related primarily to renovascular hypertension and ischemic nephropathy. Patients typically present with resistant hypertension unresponsive to high doses of multiple antihypertensive agents.

Patients with fibromuscular dysplasia often complain of headache and pulsatile tinnitus. Fibromuscular dysplasia should be suspected in young patients, especially young women, who have a recent onset of hypertension without other cardiac risk factors or a family history of hypertension. Because patients with fibromuscular dysplasia often benefit from renal artery revascularization, renal Doppler ultrasonography, which is the preferred screening test in these patients, is recommended. If the renal artery ultrasonography suggests significant renal artery stenosis, then renal angiography (Fig. 125-1) should be considered.

In patients at risk for atherosclerotic renal artery stenosis, any enthusiasm for making the diagnosis must be tempered by the lack of efficacy of renal artery revascularization for reducing blood pressure or improving outcomes

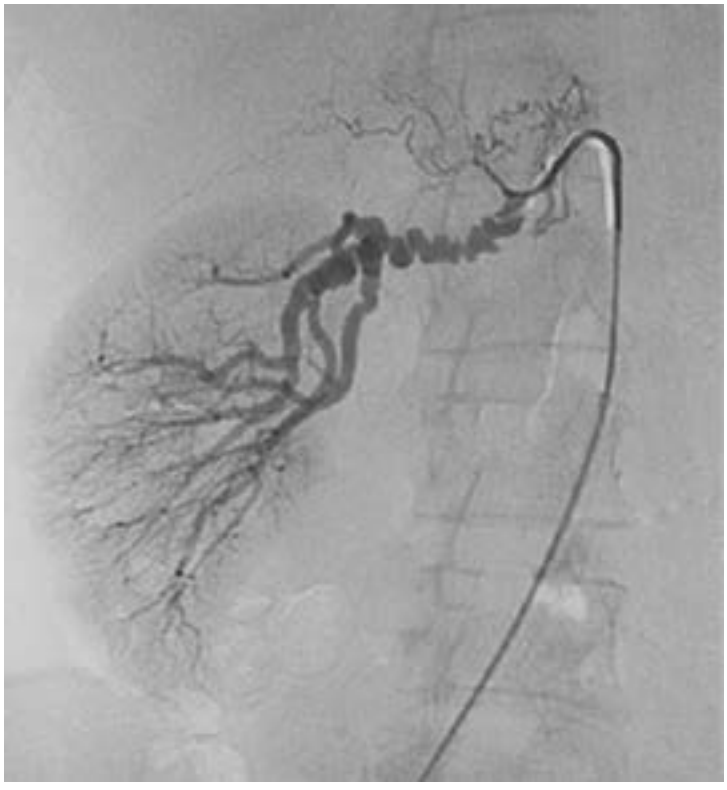

FIGURE 125-1. Selective right renal arteriogram demonstrating fibromuscular dysplasia. Typical features of medial form of fibromuscular dysplasia are illustrated by the "beads on a string" appearance.

(Chapter 67). For the small subgroup of patients whose atherosclerotic vascular artery stenosis may benefit from revascularization (see later), screening by renal Doppler or magnetic resonance angiography (MRA) is reasonable (Fig. 125-2). MRA requires gadolinium enhancement, which may precipitate a rare but severe condition called nephrogenic systemic fibrosis (Chapter 267), particularly when the linear gadolinium chelate gadodiamide is used in patients with advanced chronic kidney disease or acute kidney injury. Gadolinium is contraindicated if the estimated GFR is less than 30 mL/min. The risk of nephrogenic systemic fibrosis limits the applicability of gadolinium-enhanced MRA for patients with advanced renal dysfunction, but these patients are also less likely to benefit from renal revascularization. CTA of the renal artery may be required in some cases, but it has the disadvantages of requiring iodinated contrast as well as exposure to ionizing radiation. Digital subtraction angiography is used when revascularization is planned (Fig. 125-3). Although the complication rates are low, the usual risks of catheterization should be considered, including access site trauma, contrast reactions, contrast nephropathy (Chapter 57), and atheroembolic renal disease (Chapter 29).

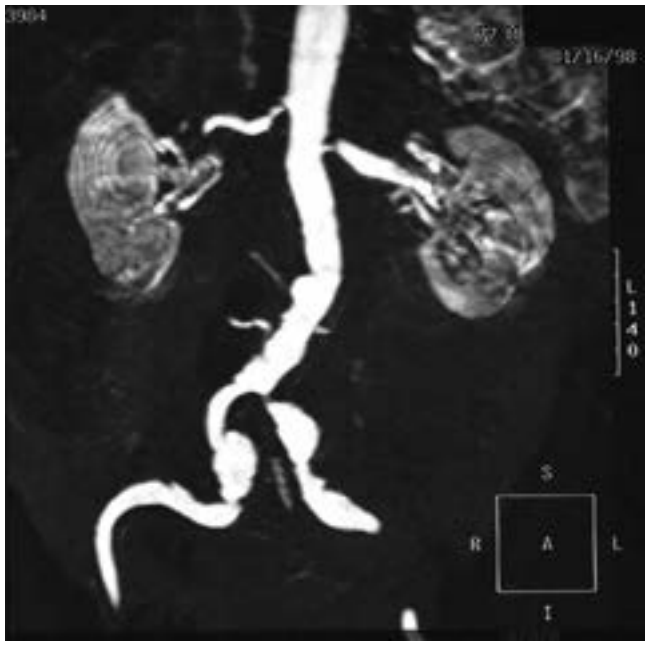

FIGURE 125-2. Magnetic resonance angiogram of the abdominal aorta showing bilateral renal artery stenosis. Significant iliac stenosis is also demonstrated.

TREATMENT

In patients with fibromuscular dysplasia, balloon angioplasty is the treatment of choice[5] because the stenosis can progress to renal artery occlusion despite adequate blood pressure control (Video 125-1). Stenting is rarely needed, and restenosis rates are generally low. Close follow-up with serial blood pressure measurements and evaluation of renal function should be performed every 3 to 4 months.

By comparison, routine renal artery revascularization is no longer recommended even for patients with severe atherosclerotic renal artery stenosis because medical therapy, especially with an ACE inhibitor or ARB, is superior to revascularization. Atherosclerotic renal artery stenosis should be viewed as a biologic marker of cardiovascular disease and cardiac morbidity. The goals of therapy are to control blood pressure, stabilize renal function, and reduce cardiovascular complications. Efforts to optimize medical therapy for secondary prevention include aspirin (81 mg/day), statins to treat dyslipidemia (Chapter 206), management of diabetes (Chapter 229), and control of blood pressure (Chapter 67). Medical therapy for blood pressure control should include ACE inhibitors or ARBs because of their proven benefit in renal protection. Although both ACE inhibitors and ARB therapy are usually well tolerated, the serum creatinine and estimated GFR must be monitored carefully during the first weeks after their initiation, particularly in elderly patients. An increase of 1.0 mg/dL or more in the serum creatinine level suggests significant bilateral renal artery stenosis, renal artery stenosis in a unilateral functional kidney, or renal artery stenosis in the kidney transplant allograft. This complication, which arises as a result of inhibition of autoregulation, is an absolute indication to discontinue ACE inhibitor or ARB therapy. Smoking cessation, weight control, and increased exercise should be universally recommended.

In selected circumstances, renal artery intervention may be considered for patients whose atherosclerotic renal artery stenosis is associated with recurrent episodes of pulmonary edema that cannot be explained by cardiac lesions, particularly if left ventricular function is preserved.[6] Revascularization may also be considered for acute reversible renal dysfunction associated with an ACE inhibitor or ARB therapy. Finally, critical bilateral renal artery stenosis or stenosis in a single functioning kidney may be an indication for intervention. Procedures should be performed at high-volume centers with capability for digital subtraction angiography and experienced operators to minimize the risk of contrast-induced nephropathy.

PROGNOSIS

When adjusted for baseline variables, atherosclerotic renal artery stenosis remains an independent predictor of cardiovascular mortality, with a 4-year adjusted mortality rate of 25% to 40%. Factors associated with higher mortality rates include an elevated baseline serum creatinine level, more severe renal artery stenosis, worsening renal function, age, advanced diabetes, other cardiovascular disease, and heart failure. Improvement of blood pressure control or renal function after revascularization is associated with improved survival even though revascularization does not affect overall survival.

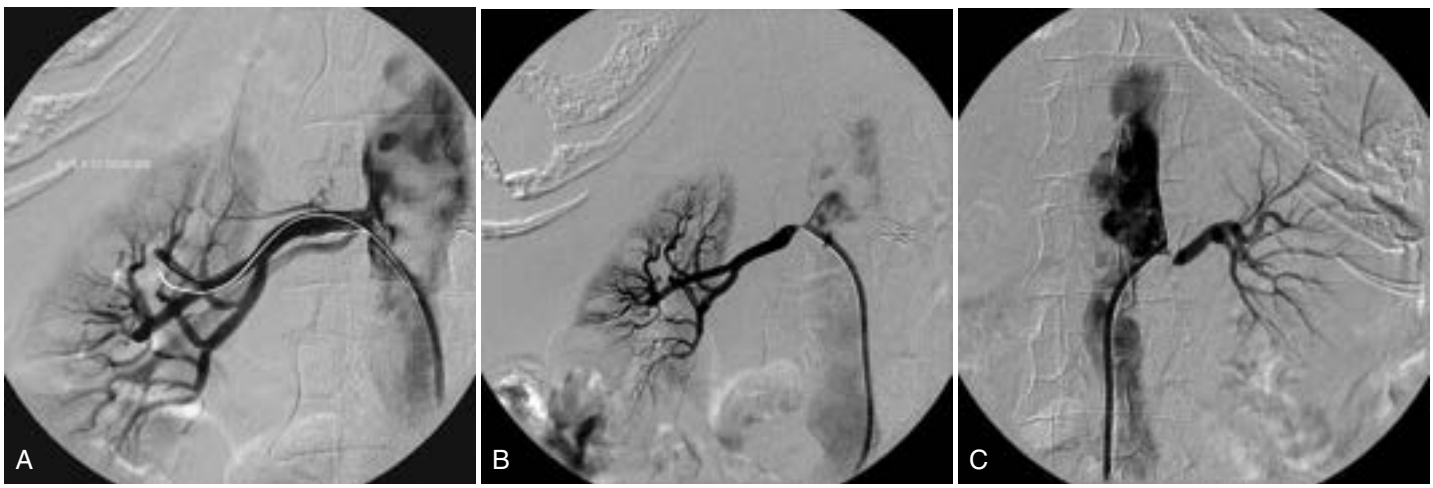

FIGURE 125-3. Renal angiograms from an elderly patient with heart failure. Cardiac catheterization revealed normal coronary arteries, but after initiation of therapy with an angiotensin-converting enzyme inhibitor and spironolactone, progressive kidney disease with hyperkalemia and poor blood pressure control ensued. Renal Doppler ultrasonography suggested bilateral renal artery stenosis, as confirmed by angiography (**A** and **B**). The patient underwent successful percutaneous revascularization in stages, leading to a return to normal left ventricular function and improved blood pressure control (**C**).

VIDEO 125-1. Renal artery stent.

For fibromuscular dysplasia, angioplasty corrects the hypertension in approximately 45% of patients.[7] Younger age, milder hypertension, and shorter duration of hypertension are associated with successful outcomes. An atrophic kidney (<8 cm), however, is unlikely to recover with revascularization.

THROMBOEMBOLIC OCCLUSION OF THE RENAL ARTERIES

EPIDEMIOLOGY AND PATHOBIOLOGY

Acute occlusion of the renal arteries and segmental branches may arise as a result of intrinsic pathology of the renal arteries, abdominal trauma, or embolization of thrombi arising in the heart or proximal aorta. Thrombosis can occur as a complication of progressive atherosclerosis, in which case it may be an important cause of progressive renal insufficiency. In other patients, thrombosis may be associated with thrombophilic states (Chapter 176), such as the antiphospholipid antibody syndrome. Thrombosis also may occur as a consequence of inflammatory disorders, including Takayasu arteritis (Chapter 78); syphilis (Chapter 319); thromboangiitis obliterans (Chapter 80); and systemic vasculitides, especially granulomatosis with polyangiitis (Chapter 270). It has also been reported after infliximab infusions. In situ thrombosis has been observed in structural lesions of the renal arteries, such as fibromuscular dysplasia or renal artery aneurysms. In patients younger than 60 years, traumatic thrombosis is the most common cause. Blunt trauma and deceleration injuries can cause acute thrombosis as a result of intimal tears, contusion against the vertebral column, or compression from a retroperitoneal hematoma. Iatrogenic causes include diagnostic angiography or arterial intervention in the renal arteries or vessels proximal to the kidneys.

Embolization, which is a more common cause of renal artery occlusion than in situ thrombosis, is generally unilateral but is bilateral in 15% to 30% of cases. Total infarction of the kidney is much less common than segmental infarction or ischemia. Approximately 90% of thromboemboli to the renal arteries originate in the heart. Atrial fibrillation with embolization of atrial thrombus is the most common cause (Chapter 64), but left ventricular thrombus (Chapter 73), valvular heart disease (Chapter 75), bacterial endocarditis (Chapter 76), nonbacterial (aseptic) endocarditis (Chapter 76), and atrial myxoma (Chapter 60) are other instigators. Noncardiac sources include aortic atheroma (Chapter 78) and mural thrombus, as well as paradoxical emboli through an atrial septal defect or patent foramen ovale (Chapter 69).

CLINICAL MANIFESTATIONS

The clinical presentation of a renal infarction can be variable, and the diagnosis is often confused with more common disorders such as renal colic or pyelonephritis. Occlusion of a primary or secondary branch of the renal artery in a patient with preexisting disease and established collateral circulation, such as in long-standing renal artery stenosis, may produce little or no infarction and minimal symptoms. Acute thrombosis and infarction may cause a sudden onset of flank pain, fever, nausea, vomiting, and, on occasion, hematuria. Pain may be localized to the abdomen, back, or even the chest, but pain is absent in more than 50% of cases. Hypertension, which occurs with infarction, is a result of the release of renin from the ischemic renal parenchyma and may be severe.

Anuria suggests bilateral involvement or occlusion of the artery to a solitary kidney. Urinalysis usually, but not always, reveals microscopic hematuria, and mild proteinuria may be present.

If infarction occurs, leukocytosis usually develops, and serum enzyme levels of aspartate aminotransferase, lactate dehydrogenase (LDH), and alkaline phosphatase may be elevated. These laboratory findings are nonspecific, but elevation of the urinary LDH level is more specific because its concentration should be normal in extrarenal disorders. Blood urea nitrogen and creatinine levels typically increase transiently with unilateral infarction, but more severe and protracted renal dysfunction may follow bilateral renal infarction or infarction of a solitary kidney.

DIAGNOSIS

The diagnosis of renal artery occlusion is most reliably established by computed tomography (CT), with and without contrast. CT is accurate, can be performed rapidly, and can identify associated traumatic injury. Findings may include filling defects in the main or segmental renal arteries, as well as the absence of enhancement of renal tissue, indicating a lack of perfusion (Fig. 125-4). The complication of contrast nephropathy is a major concern in patients with a creatinine level of 2.0 mg/dL or greater or an estimated GFR less than 60 mL/m. Alternatives to contrast administration should always be considered in patients with chronic kidney disease, acute renal failure, or diabetes mellitus and in elderly patients. MRA has a high diagnostic accuracy and may be preferable to contrast CT in elderly patients and in patients with diabetes mellitus, although it is contraindicated in patients with renal insufficiency. Radioisotope renograms, excretion urograms, and duplex ultrasound scanning are not recommended for the diagnosis of acute occlusions. Invasive angiography carries inherent risks but is occasionally required if the diagnosis remains uncertain or if percutaneous reperfusion is considered.

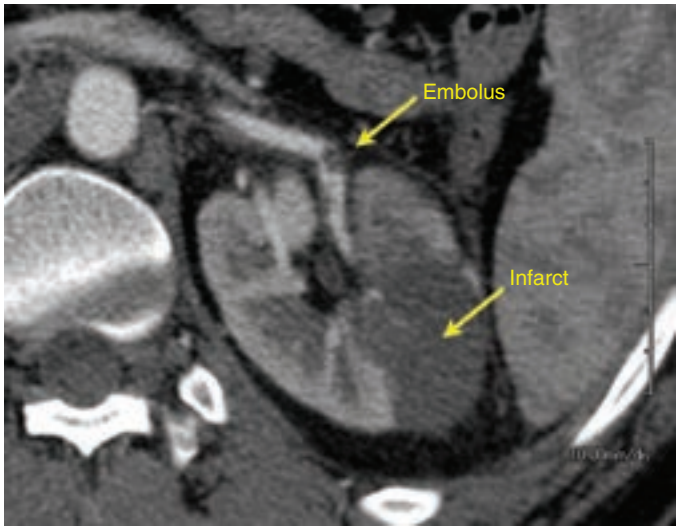

FIGURE 125-4. Computed tomography demonstrating a clot in the main renal artery and segmental renal infarction.

In patients with suspected embolic renal artery occlusion, echocardiography is indicated to search for a possible intracardiac thrombus. In nontraumatic thrombotic occlusion, evaluation for thrombophilia (Chapter 176), vasculitides (Chapter 270), or progressive atherosclerosis (Chapter 79) should be considered.

TREATMENT

Rx

The human kidney can tolerate 60 to 90 minutes of warm ischemia, although the presence of collateral circulation may permit longer ischemic times. As a result, acute renal artery thrombosis requires urgent treatment in an attempt to reopen the artery. Options for nontraumatic renal artery thrombosis include systemic anticoagulation with unfractionated heparin (see Table 81-4 in Chapter 81) or low-molecular-weight heparin (LMWH) (see Table 81-3 in Chapter 81) for about 7 to 10 days, with oral warfarin begun at about day 3 and continued for 1 year to maintain an international normalized ratio of 2.0 to 3.0. Alternatively, intraarterial thrombolytic therapy may be considered. Surgical revascularization is associated with a higher mortality rate than medical therapy without improved renal salvage rates, so it is not recommended as primary therapy; it can be considered, however, for patients with bilateral renal artery occlusion or occlusion of the renal artery of a solitary kidney. Percutaneous endovascular therapies (e.g., local thrombolysis, thrombectomy, stent placement) have also been successful in acute renal artery occlusion, but there have been no comparative studies of endovascular intervention compared with medical therapy for renal artery occlusion.

For traumatic renal artery thrombosis, surgery is the treatment of choice but usually can salvage renal function only if accomplished immediately. For iatrogenic occlusion of the renal artery as a result of angiographic manipulations or angioplasty, intraarterial stent placement may be considered.

Patients with renal artery thrombosis also require rigorous medical care with assiduous attention to control of their blood pressure. To achieve a target blood pressure between 140/90 and 110/70 mm Hg may require multiple parenteral agents, as is recommended for malignant hypertension, before switching to oral agents, preferably a combination of ACE inhibitors, ARBs, or nondihydropyridine calcium channel blockers (Chapter 67). Adequate hydration is also imperative.

PROGNOSIS

The mortality rate is high, especially in patients who require hemodialysis, and it correlates with the severity of the underlying conditions. For patients undergoing surgical revascularization for complete acute renal artery

occlusion, the mortality rate is 11% to 25%. The risk of end-stage renal disease (ESRD) is variable, and rates of 0% to greater than 50% have been reported. Hypertension, which may develop as a late sequela of treated renal artery occlusion, is preferably treated with ACE inhibitors, ARBs, or nondihydropyridine calcium channel blockers (see Table 67-7 in Chapter 67).

ARTERIOLES AND MICROVASCULATURE

Atheroembolic Disease of the Renal Arteries

EPIDEMIOLOGY AND PATHOBIOLOGY

Cholesterol crystal embolization is a potential complication of widespread atherosclerosis. The risk factors are similar to those for all atherosclerotic disease (Chapter 52), including smoking, hypertension, hyperlipidemia, diabetes, and older age. Atheroembolic disease appears to be more common in whites, but the condition may be underdiagnosed in African Americans owing to the difficulty of assessing livedo reticularis in this population.

The most common triggering events are manipulation of a thrombus or of the abdominal aorta or renal arteries during angiography or transluminal angioplasty. It is not unusual, therefore, for cholesterol crystal embolization to be mistaken for contrast-induced nephropathy. Atheroemboli can also be associated with anticoagulant or thrombolytic therapy and may be accompanied by the finding of a cyanotic toe on physical examination. Spontaneous atheroembolism after detachment of a mural plaque is uncommon. Cholesterol crystals typically do not occlude arterial flow but induce an inflammatory response and subsequent endothelial proliferation, so the clinical manifestations may occur some time after the initial insult.

CLINICAL MANIFESTATIONS

Although all organ systems can be affected (Chapter 80), the kidneys are most commonly involved followed by the spleen and gastrointestinal tract.[8] Acute renal failure, hypertension, or both typically occur, and many patients progress to chronic kidney disease or even ESRD.[9] Nonspecific complaints may include fever, myalgias, headaches, and weight loss. Evidence of cholesterol embolization may be present in the retina, muscles, or skin, where livedo reticularis (see Fig. 80-3 in Chapter 80) or cyanotic digits may be observed. Embolization can also result in cerebrovascular events, acute pancreatitis, ischemic bowel, and gangrene of the extremities (Chapter 80).

DIAGNOSIS

Although most atheroembolic events are diagnosed because of an acute clinical change, clinically silent, chronic, low-grade embolization may be overlooked because patients at risk for this complication often have other chronic illnesses associated with renal failure, hypertension, and atherosclerosis. Urinalysis may not be helpful because cholesterol crystals often are not present, but mild proteinuria, eosinophiluria, and increased cellularity are frequently observed. Transient eosinophilia is common, and an elevated erythrocyte sedimentation rate (ESR), hypocomplementemia, anemia, and leukocytosis may also be present. Up to 80% of patients may have a serum creatinine level exceeding 2 mg/dL.

Although the demonstration of cholesterol crystals in the renal microvasculature with subsequent vessel occlusion is considered a diagnostic feature, a kidney biopsy generally is not necessary for patients with typical clinical features, including livedo reticularis and violaceous mottling of the toes, eosinophilia, and an elevated ESR. When uncertainty exists, noninvasive biopsy of the skin of the lower extremity, muscle of the calf or thigh, or gastric mucosa can be diagnostic in up to 80% of patients.

TREATMENT Rx

There is no curative therapy for atheroembolic disease, so supportive care is often all that can be offered. Aggressive control of dyslipidemia with statins (Chapter 206) is recommended and may have the added benefit of stabilizing the endothelial surface. Anticoagulants are of no value and may delay the healing of ulcerating atherosclerotic lesions; if possible, anticoagulant therapy should be discontinued. Adequate hydration is important to sustain renal perfusion. Hypertension should be treated with angiotensin II antagonists and vasodilators (see Table 67-7 in Chapter 67), and volume should be controlled with diuretics, with care to avoid hypotension. Intravascular radiologic procedures and vascular surgery should be avoided if possible.

PROGNOSIS

Recent series suggest up to an 80% survival rate at 1 year. Most patients die of cardiovascular complications. With adequate blood pressure control for several months or years, renal function may recover sufficiently to avoid dialysis. Patients who develop ESRD have a significantly higher mortality rate.

RENAL VEINS

Renal Vein Thrombosis

EPIDEMIOLOGY

Unilateral or bilateral thrombosis of the major renal veins or their segments is often a subtle disorder that can develop in a variety of conditions but is especially prominent with nephrotic syndrome[10] (Chapter 121) or renal cell carcinoma (Chapter 197). Cases have also been associated with trauma, oral contraceptive use, hypovolemia, renal transplantation (Chapter 131), and thrombophilic states (Chapter 176). The reported incidence of renal vein thrombosis ranges from 5% to 62% in patients with nephrotic syndrome and is typically associated with membranous nephropathy, but it may also occur with membranoproliferative glomerulonephritis, focal glomerular sclerosis, sickle cell nephropathy, amyloidosis, diabetic nephropathy, renal vasculitis, and lupus nephritis. Spontaneous renal vein thrombosis is unusual in patients without underlying risk factors.

PATHOBIOLOGY

The precipitating factors are apparently abnormalities in coagulation or fibrinolysis. Antithrombin III and plasminogen levels may be depressed as a result of urinary excretion of antithrombin III in patients with nephrotic syndrome. Thrombocytosis, increased platelet activation, hyperfibrinogenemia, inhibition of plasminogen activation, and altered circulating levels of proteins S and C in nephrotic syndrome contribute to thromboembolic complications.

Extrinsic compression of the renal veins from retroperitoneal sources such as lymph nodes, fibrosis, abscess, aortic aneurysm, or tumor may cause renal vein thrombosis as a result of sluggish renal venous flow. Acute pancreatitis, trauma, and retroperitoneal surgery also may predispose to renal vein thrombosis. Renal cell carcinoma characteristically invades the renal vein and compromises venous flow, resulting in renal vein thrombosis. Renal vein thrombosis in the setting of severe volume depletion and impaired renal blood flow has been described in young adults.

CLINICAL MANIFESTATIONS

The manifestations of renal vein thrombosis depend on the extent and rapidity of the development of the occlusion. Patients with acute renal vein thrombosis may have nausea, vomiting, flank pain, leukocytosis, hematuria, compromised renal function and an increase in kidney size. These features may be confused with renal colic or pyelonephritis. Adult patients with nephrotic syndrome and chronic renal vein thrombosis may have more subtle findings, such as a dramatic increase in proteinuria or evidence of tubular dysfunction, including glycosuria, aminoaciduria, phosphaturia, and impaired urinary acidification. Chronic renal vein thrombosis may first present in association with a pulmonary embolus.

DIAGNOSIS

For acute renal vein thrombosis, which is typically associated with thrombophilia, contrast-enhanced CT shows an enlarged kidney, stretching of the calyces, and notching of the ureters. A venogram is rarely required but can be considered in cases of acute renal failure in which thrombectomy or thrombolysis is considered. In chronic renal vein thrombosis, an incidentally noted renal vein thrombus may be seen on imaging studies ordered for other reasons. The tumor thrombus associated with renal cell carcinoma often extends into the inferior vena cava and occasionally to the level of the right heart. Routine screening for thrombus is not recommended for patients with nephrotic syndrome, but contrast-enhanced CT is recommended in patients with suggestive clinical manifestations.

TREATMENT Rx

The most widely accepted form of therapy for both acute and chronic renal vein thrombosis is anticoagulation with LMWH (see Table 81-3 in Chapter 81) or unfractionated heparin (see Table 81-4 in Chapter 81) for about 7 to 10 days, with oral warfarin begun on about day 3 and continued for at least 1 year to

achieve an international normalized ratio of 2.0 to 3.0. In patients with ongoing risk factors, such as persistent nephrotic syndrome, or recurrent thrombosis, anticoagulation should be continued indefinitely. In patients who have an underlying renal cell carcinoma, long-term LMWH is preferred over warfarin because it is better for preventing cancer-related thrombosis (Chapter 176). Fibrinolytic therapy may be considered in patients with acute renal vein thrombosis associated with acute renal failure.

PROGNOSIS

The prognosis of patients with renal vein thrombosis depends entirely on the underlying condition causing or associated with it. Renal vein thrombosis associated with membranous glomerulonephritis and nephrotic syndrome (Chapter 121) usually resolves if the underlying condition responds to therapy or spontaneously resolves. In contrast, renal vein thrombosis associated with renal cell carcinoma (Chapter 197) has a very poor prognosis. Renal vein thrombosis associated with trauma or hypovolemia may resolve after appropriate therapy.

Grade A References

A1. Wheatley K, Ives N, Gray R, et al. Revascularization versus medical therapy for renal-artery stenosis. *N Engl J Med*. 2009;361:1953-1962.

A2. Bax L, Woittiez AJ, Kouwenberg HJ, et al. Stent placement in patients with atherosclerotic renal artery stenosis and impaired renal function: a randomized trial. *Ann Intern Med*. 2009;150:840-848.

A3. Cooper CJ, Murphy TP, Cutlip DE, et al. Stenting and medical therapy for atherosclerotic renal-artery stenosis. *N Engl J Med*. 2014;370:13-22.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Piecha G, Wiecek A, Januszewicz A. Epidemiology and optimal management in patients with renal artery stenosis. *J Nephrol*. 2012;25:872-878.
2. Benjamin MM, Fazel P, Filardo G, et al. Prevalence of and risk factors of renal artery stenosis in patients with resistant hypertension. *Am J Cardiol*. 2014;113:687-690.
3. Olin JW, Froehlich J, Gu X, et al. The United States Registry for Fibromuscular Dysplasia: results in the first 447 patients. *Circulation*. 2012;125:3182-3190.
4. McKenzie GA, Oderich GS, Kawashima A, et al. Renal artery fibromuscular dysplasia in 2,640 renal donor subjects: a CT angiography analysis. *J Vasc Interv Radiol*. 2013;24:1477-1480.
5. Chrysant SG, Chrysant GS. Treatment of hypertension in patients with renal artery stenosis due to fibromuscular dysplasia of the renal arteries. *Cardiovasc Diagn Ther*. 2014;4:36-43.
6. Ritchie J, Green D, Chrysochou C, et al. High-risk clinical presentations in atherosclerotic renovascular disease: prognosis and response to renal artery revascularization. *Am J Kidney Dis*. 2014;63:186-197.
7. Trinquart L, Mounier-Vehier C, Sapoval M, et al. Efficacy of revascularization for renal artery stenosis caused by fibromuscular dysplasia: a systematic review and meta-analysis. *Hypertension*. 2010;56:525-532.
8. Scolari F, Ravani P. Atheroembolic renal disease. *Lancet*. 2010;375:1650-1660.
9. Quinones A, Saric M. The cholesterol emboli syndrome in atherosclerosis. *Curr Atheroscler Rep*. 2013;15:315.
10. Barbano B, Gigante A, Amoroso A, et al. Thrombosis in nephrotic syndrome. *Semin Thromb Hemost*. 2013;39:469-476.

REVIEW QUESTION

1. You are asked to review a case regarding a 67-year-old man who is being evaluated for progressive chest pain. He has type 2 diabetes mellitus with chronic kidney disease (HbA_{1c} = 8.7%, serum Cr = 2.1 mg/dL). Other tests include ratio of total cholesterol to high-density lipoprotein of 7.2 and a distant history of smoking (15 pack years), but he quit 20 years ago. He is also noted to have poorly controlled hypertension and despite being treated with atenolol, hydrochlorothiazide, and amlodipine, with systolic pressures remaining above 180 mm Hg. He states that he is compliant with all medications but does not follow his diet very well. His ejection fraction by echocardiography is 40%, and stress echocardiography showed stress-induced anterior wall motion abnormalities. What should you recommend?

A. A gadolinium-enhanced renal artery magnetic resonance image after overnight hydration with 0.5% normal saline. If it is positive for renal artery stenosis, perform renal angiography at the time of the cardiac catheterization that he also needs.
B. Proceed with a cardiac catheterization after overnight hydration. Perform renal angiography only if the patient will require coronary bypass surgery. If coronary stenting is performed, delay the renal angiography for 6 weeks to reduce his acute exposure to contrast dye.
C. Start statin therapy and improve his diabetes management. When his cardiac evaluation is completed, consider initiating an ACE inhibitor if his renal function remains stable and careful follow-up of the serum creatinine concentration.
D. Order renal artery Doppler studies with resistive indices. If there is greater than 60% stenosis of one or both renal arteries and the resistive index is less than 0.8, proceed with renal artery stenting of the most severe lesion and stage the second artery if necessary.
E. None of the above

Answer: C The patient is taking multiple hypertensive medications, but his blood pressure remains poorly controlled, and he has renal dysfunction, left ventricular dysfunction, and probable coronary artery disease. Therefore, he has many risk factors for atherosclerotic renal artery stenosis. Multiple randomized trials have shown no efficacy for renal revascularization for atherosclerotic renal artery stenosis, so screening in this scenario would be of no clinical benefit. Furthermore, his estimated glomerular filtration rate is likely to be less than 30 mL/min, thereby making gadolinium contraindicated. On the other hand, treatment of his risk factors and addition of an ACE inhibitor for left ventricular dysfunction and diabetes mellitus has the most clinical benefit for reducing major adverse cardiovascular events.

126

NEPHROLITHIASIS

DAVID A. BUSHINSKY

Kidney stones are composed of crystals in a protein matrix. Most crystals contain calcium, which is generally complexed with oxalate, phosphate, or both; other stones are composed of uric acid, magnesium ammonium phosphate (struvite), or cystine, alone or in combination. Kidney stones form when the urinary saturation of their components exceeds the solubility of the solid phase.

EPIDEMIOLOGY

The annual incidence of kidney stones in industrialized nations exceeds one per 1000 persons, with a lifetime risk of about 7% in women and about 11% in men.[1] The incidence of stone formation peaks in the third and fourth decades of life, and the prevalence increases with age until approximately 70 years in men and 60 years in women. In the United States, the increase in lifetime risk of nephrolithiasis from 3% in the late 1970s to almost 9% in 2010 is far faster than can be accounted for by alterations in our genome. The increasing prevalence suggests that changes to our diets and lifestyle may account for this increased stone formation. The estimated yearly economic cost of kidney stones exceeds $5 billion in the United States.

In the United States, whites are more likely to develop renal stones than are other ethnic groups. Stones are more common in hot, dry climates, perhaps because greater fluid loss through the skin and respiration leads to more concentrated urine. Many occupations make it inconvenient to use a restroom, and these individuals often avoid fluids in an effort to avoid urination, thereby leading to excretion of a concentrated urine. Insufficient rehydration among people who engage in physical activity and have large insensible losses also leads to concentrated urine.

Obesity is correlated with the risk for kidney stone formation. Individuals weighing more than 220 lb or having a body mass index (BMI) greater than 30 are significantly more likely to form stones than are individuals who weigh less than 150 lb or have a BMI between 21 and 22.9.[2]

The types of stones vary around the world. In the United States, the majority of stones are calcium oxalate or calcium phosphate (>80%), and fewer than 10% are pure uric acid stones. By comparison, about 70% of stones in the Mediterranean and Middle East are composed of uric acid. Magnesium ammonium phosphate (struvite) stones account for 10% to 25% of stones, and cystine stones comprise 2% of stones.

PATHOBIOLOGY

Human bone is composed of calcium and phosphate, principally in the form of apatite. When humans reach their adult height and their skeletons are fully mineralized, the net amount of calcium that is absorbed by nonpregnant individuals must be excreted in the urine. Similarly, absorbed phosphate that is not needed for bone mineralization or cell growth must be excreted. Oxalate is an end product of metabolism, and it too must be excreted in the urine. The need for water conservation by terrestrial man often leads to excretion of these ions in relatively scant amounts of urine, thereby leading to increasing saturation with respect to the solid phases of calcium oxalate and calcium phosphate. Increasing saturation drives the formation of the solid crystal phase and is expressed as the ratio of calcium oxalate (or phosphate) ion activity to its solubility. At ratios of greater than 1, termed *supersaturation*, a solid phase can form, but the substances remain in solution at ratios less than 1. When urine is supersaturated, ions can bond to form the more stable, solid phase, which is termed *nucleation*. Homogeneous nucleation refers to the bonding of similar ions into crystals. The more common and thermodynamically favored heterogeneous nucleation occurs when crystals grow on dissimilar crystals or substances as cellular debris in the urine. Although humans produce inhibitors of stone formation, such as osteopontin and Tamm-Horsefall protein, supersaturation can overwhelm this inhibition, and a solid phase will form.

Calcium Stones

About 70% to 80% of kidney stones contain calcium, which is often complexed with oxalate or phosphate.[3] Calcium-containing kidney stones are most often caused by excessive excretion of calcium (hypercalciuria), oxalate (hyperoxaluria), or urate (hyperuricosuria), or an insufficient excretion of citrate (hypocitraturia)

Calcium-containing stones have a strong genetic component. Idiopathic hypercalciuria is thought to be a polygenic disorder in which a generalized dysregulation of calcium transport in the kidney, intestine, and bone leads to excessive urine calcium excretion.[4]

An initial genome-wide association study has identified sequence variants in genes encoding claudin 14, which regulates calcium reabsorption in the thick ascending limb of Henle loop.

Urinary oxalate is derived from endogenous metabolism of glyoxylate and ascorbic acid or from dietary sources, such as cocoa, nuts, tea, and certain leafy green vegetables (e.g., spinach). The three main causes of hyperoxaluria are excessive oxalate ingestion (dietary oxaluria), the excessive intestinal absorption of oxalate (enteric oxaluria) that is paradoxically observed in malabsorptive gastrointestinal (GI) disorders, and the excessive endogenous oxalate production seen in certain hepatic enzyme deficiencies (primary hyperoxaluria; Chapter 205). Additionally, ethylene glycol, a common automobile antifreeze, is metabolized to oxalate and can cause excessive urinary oxalate excretion in conjunction with severe metabolic acidosis and renal failure (Chapter 110).

Enteric oxaluria results in elevated urinary oxalate levels (60-100 mg/day). In GI malabsorptive conditions such as Crohn disease (Chapter 141), celiac sprue (Chapter 140), jejunoileal bypass (Chapter 220), and chronic pancreatitis (Chapter 144), malabsorbed fatty acids bind dietary calcium, thereby allowing oxalate to be readily absorbed in the colon. The colonic mucosa becomes more permeable to oxalate owing to exposure to bile acids.

Primary hyperoxaluria (Chapter 205) results from hepatic enzyme deficiencies that lead to substantial endogenous oxalate production and a marked elevation of urinary oxalate (80-300 mg/day).[5] Oxalate deposits in numerous organs, including the heart, bone marrow, muscle, and renal parenchyma, where they lead to renal failure, cardiomyopathy, and bone marrow suppression at an early age. In type 1 primary hyperoxaluria, the hepatic enzyme alanine glyoxylate aminotransferase is deficient because of one of several mutations in the *AGXT* gene. In the less common type 2 primary

hyperoxaluria, patients lack D-glycerate reductase and glyoxylate reductase owing to mutations in the *GRHPR* gene. Type 3 is a pediatric disease that is due to mutations in the *HOGA1* gene and does not lead to kidney failure.

Nephrolithiasis and nephrocalcinosis can also result from a variety of monogenic disorders, such as Dent disease (X-linked recessive nephrolithiasis), McCune-Albright syndrome (Chapters 231 and 248), osteogenesis imperfecta type 1 (Chapter 260), and congenital lactase deficiency (Chapter 140).

Citrate combines with calcium to form a soluble complex that reduces the availability of calcium to bind with oxalate or phosphate. The principal risk factor for hypocitraturia is a high protein intake. Men often have lower urinary citrate concentrations than women. Distal renal tubular acidosis (Chapter 118) leads to calcium phosphate stones owing to bone demineralization and an alkaline tubular pH.

Calcium oxalate kidney stones form on calcium phosphate deposits, termed Randall plaques, which are located in the renal papillae. Randall plaque formation is positively correlated with urine calcium excretion and negatively correlated with urine volume and pH. These calcium phosphate crystals, in the form of apatite, originate around the thin loop of Henle and then extend into the interstitium without eroding into the tubular lumen or damaging the tubular cells. The crystals move toward the urinary space, where they form Randall plaques, which are visible on cystoscopy but not generally visible on a routine CT scan. In the presence of urine that is supersaturated with respect to calcium oxalate, calcium oxalate crystals bind to these plaques and increase in size. If the calcium oxalate crystals then break off from the Randall plaques, the free stone can enter, migrate, irritate, and possibly obstruct the ureter, where it can cause severe pain.

Uric Acid Stones

The incidence of uric acid stones in the United States appears to be rising in parallel with the increase in obesity, which leads to renal insulin resistance and a very low urine pH.[6] Most patients with uric acid stones have a reduced urinary pH, and some have low urine volumes or elevated urinary uric acid levels. More than five times as much uric acid is soluble in a urine with a pH of 6.5 compared with a pH of 5.3. Diarrhea and diets high in animal protein can contribute to an acidic urinary pH. Uric acid stone formers have greater body weight and a higher incidence of insulin resistance and type 2 diabetes mellitus. Insulin resistance leads to impaired urinary ammonium excretion, thereby resulting in the excretion of more hydrogen ions as titratable acids than as ammonium and, therefore, a lower urinary pH.

Hyperuricosuria may be seen in patients who ingest large quantities of dietary purine, such as found in organ meats, shellfish, certain fish (e.g., anchovies, sardines, herring, and mackerel), and meat extracts such as bouillon and consommé, and protein. Hyperuricemic disorders, including gout (Chapter 273), myeloproliferative disorders, tumor lysis syndrome, and certain inborn errors of metabolism, can also contribute to an increased urinary uric acid. Medications such as salicylates and probenecid can be hyperuricosuric.

Struvite Stones

Struvite stones, sometimes called *triple phosphate stones, magnesium ammonium phosphate stones,* and *infection stones,* comprise only about 10% to 25% of all stones but constitute the majority of staghorn calculi, which are large stones that extend beyond a single renal calyx.[7] Struvite stones are far more common in women than in men, in large part owing to women's increased susceptibility to urinary tract infections (UTIs) (Chapter 284). Similarly, any patient with urinary stasis, such as patients with neurogenic bladders, indwelling urinary catheters, or spinal cord lesions, is susceptible to struvite stones.

Struvite stones form only in the presence of both ammonium ions and an alkaline urinary pH of 7 or greater, which only occur with urease-producing bacteria. *Proteus* spp. (Chapter 305) are common urease-producing bacteria, but other gram-negative and gram-positive bacteria (e.g., *Klebsiella* spp. and *Staphylococcus epidermidis*) as well as *Mycoplasma* spp. (Chapter 317) and yeast species have been implicated in urease production. By comparison, *Escherichia coli* does not produce urease.

Cystine Stones

Cystinuria (Chapter 128), which is an autosomal recessive disorder caused by mutations of the *SLC3A1* gene or the *SCLC7A9* gene, results in decreased renal tubular reabsorption and excessive urinary excretion of the dibasic amino acids cystine, ornithine, lysine, and arginine.[8] The resulting urinary excretion of cystine in the typical volume of urine exceeds its solubility of about 300 mg/L, so stones form. Although normal people excrete about 30 to 50 mg/day of cystine, heterozygotes for cystinuria excrete about 400 mg/day, and homozygotes often excrete about 600 mg/day. Thus, homozygotes must continually excrete 2 L or more of urine each day to avoid stone formation. Cystinuria is unrelated to the much more severe disorder *cystinosis* (Chapter 128), which results in extensive intracellular cystine accumulation.

CLINICAL MANIFESTATIONS

Patients with kidney stones often present with pain or hematuria (or both) and less often with UTIs (Chapter 284) or acute kidney injury (Chapter 120), either owing to bilateral obstruction or unilateral obstruction of a single functioning kidney (Chapter 123). Patients often complain of severe ureteral colic. The pain is of abrupt onset and can intensify into severe, excruciating flank pain. The pain may migrate anteriorly along the abdomen and inferiorly to the groin, testicles, or labia majora as the stone moves down the ureter toward the ureterovesical junction. The pain resolves only after the stone passes or is removed. Hematuria, even gross hematuria, is common, and patients occasionally present with painless hematuria (see Fig. 114-1 in Chapter 114), and the finding of a stone on radiographic examination does not preclude another cause. Conversely, even large calculi may be asymptomatic and be discovered during the investigation of unrelated symptoms. Obstruction caused by calculi may also be painless, and nephrolithiasis should always be considered in the differential diagnosis of unexplained acute or chronic kidney disease (Chapters 120, 123, and 130).

DIAGNOSIS

The physical examination can provide clues to the diagnosis of kidney stones but is not diagnostic. Some patients have demonstrable flank tenderness, and a rare patient with hyperuricemia will have tophi (Chapter 273). However, the physical examination is most helpful for not showing other potential causes of pain (Fig. 126-1). The suspicion of a kidney stone generally obligates radiographic evaluation. Radiographic studies can be deferred, however, in patients in whom the clinical diagnosis is clear, who have no evidence of infection, who are able to eat and drink, and who can be managed on oral analgesics.

Ultrasonography is an easy and rapid way to detect possible urinary obstruction. Ultrasonography can detect clinically significant renal calculi, although it has only a 19% sensitivity in detecting the ureteral stones that cause acute symptoms in many patients. The main advantage of ultrasonography is that it does not use radiation and is clearly the modality of choice when radiation exposure must be minimized.

A helical computed tomographic (CT) scan without radiographic contrast can detect kidney stones with both a sensitivity and a specificity exceeding 95% (Fig. 126-2). Based on the density of the stone, it also can often differentiate a calcium-containing stone from a cysteine or uric acid stone. An added benefit is that helical CT is often helpful in determining the cause of non–stone-induced abdominal pain.

However, a helical CT scan, especially when performed both with and without contrast, exposes the patient to ionizing radiation, which increases the risk of cancer. In a recent trial, initial ultrasonography was associated with lower cumulative radiation exposure than an initial helical CT, without significant differences in high-risk diagnoses with complications.

In certain situations, however, such as in patients with HIV thought to have stones induced by protease inhibitor, a helical CT scan with contrast is often required because the stones are not radiopaque and do not obstruct the ureter.

Approximately 90% of kidney stones are radiopaque and may be detected on a simple abdominal radiograph. Unfortunately, however, the stone is often obscured by stool, vertebrae, or abdominal gas, so the sensitivity of a plain abdominal radiograph is about 50%, and its specificity is only about 75%. Uric acid stones are radiolucent and cannot be detected radiographically without contrast.

Intravenous pyelography (IVP) has a sensitivity of about 75% and a specificity of more than 90% for detecting renal calculi. IVP is also useful for identifying structural abnormalities of the urinary tract such a medullary sponge kidney (Chapter 127) that predisposes to stone formation. However, an IVP often does not detect nonobstructing radiolucent stones because they do not create a filling defect. An IVP exposes the patient to more radiation than a plain radiograph but less than a CT scan. It also carries the risk of radiographic contrast material, which is greater in individuals with underlying renal compromise. With the widespread availability of ultrasonography and helical CT scans, an IVP is rarely indicated.

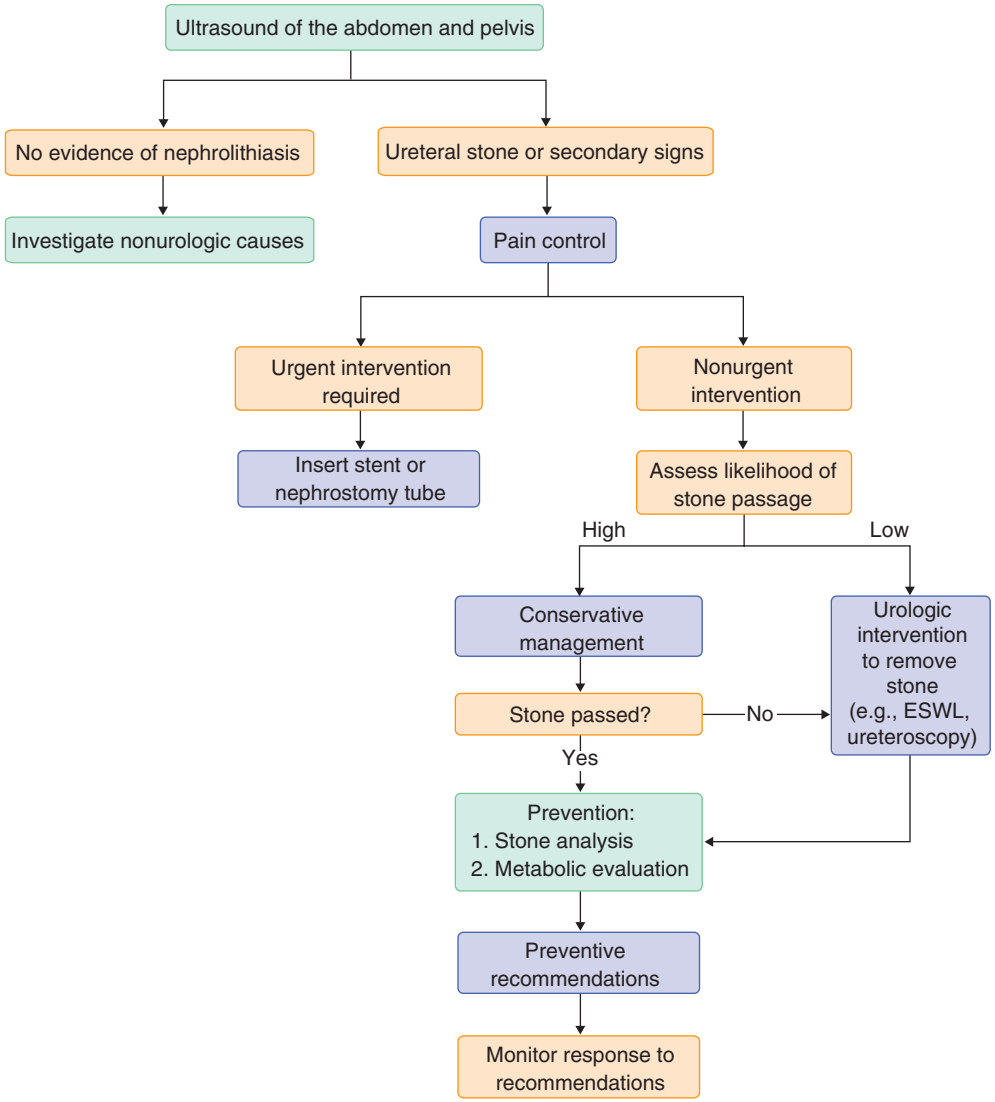

FIGURE 126-1. Algorithm for evaluation of suspected renal colic. CT = computed tomography; ESWL = extracorporeal shock-wave lithotripsy.

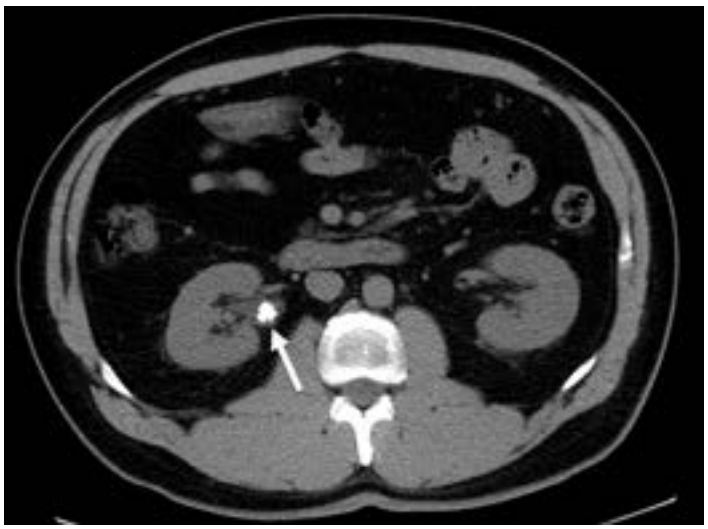

FIGURE 126-2. High-resolution helical computed tomography scan of the upper part of the abdomen demonstrating a stone in the right renal pelvis and a smaller stone in the left kidney *(arrows)*. There is no hydronephrosis. (Courtesy of Marc Brown, MD, and Lawrence H. Schwartz, MD, Department of Radiology, Columbia University Medical Center.)

TREATMENT

Medical Therapy

Because the pain of renal colic can be excruciating, pain control is critical after the definitive diagnosis has been made (see Fig. 126-1). If nausea and vomiting prevent the use of oral medication, parenteral medication is typically required. Nonsteroidal anti-inflammatory drugs (NSAIDs) are as effective as opiates for renal colic and are preferred because they have fewer side effects. An intravenous (IV) option is ketorolac (30 mg), and ibuprofen (200-400 mg/dose every 4-6 hours with a maximum daily dose of 1.2 g) is a common oral option in patients who can tolerate oral medication. Morphine (5-10 mg IV) and hydromorphone (1-2 mg IV) are options if ketorolac is insufficient to control pain. Oral oxycodone (5-15 mg every 4 to 6 hours as needed) may be added to ibuprofen for outpatient pain control. Ondansetron (2-4 mg IV) is helpful if antiemetics are required. Because NSAIDs can cause acute kidney injury, especially in patients who are dehydrated or who have chronic underlying kidney disease, adequate hydration (e.g., normal saline with dextrose at 75-150 mL/hr) is essential, but neither high-volume fluid therapy nor diuretics promote the passage of the stone.

Medical Expulsive Therapy

Kidney stones 5 mm or smaller in size have about a 70% probability of passing spontaneously; stones between 5 and 10 mm have less than a 50% chance of passing spontaneously. Medical expulsive therapy using α-adrenergic receptor blockers, such as tamsulosin (0.4 mg/day orally [PO]), terazosin, and doxazosin, or the calcium-channel blocker nifedipine (nifedipine XL 30 mg/day or twice daily) reduces spasm of the ureteral smooth muscle, allows peristalsis to move the stone more effectively, and increases spontaneous passage rates by about 50%. Medical expulsive therapy may be cautiously attempted with ureteral stones smaller than 10 mm in diameter for 4 to 6 weeks if pain

is controlled, kidney function is normal, and there is no evidence of UTI or significant obstruction. The patient must be followed closely, generally with repeat ultrasound examinations.

Initial Surgical Treatment

Stones that cause obstruction, infection, or intractable pain must be removed expeditiously. In general, stones larger than 10 mm are less likely to pass spontaneously, even with medical expulsive therapy, and are often an indicator for earlier intervention.

The approach to stone removal depends on its size, location, and composition, as well as on urinary tract anatomy. Options include extracorporeal shock-wave lithotripsy, ureteroscopic extraction, percutaneous nephrolithotomy, and only rarely open surgical extraction.

Extracorporeal shock-wave lithotripsy focuses external sound waves onto the kidney stone, thereby fragmenting the stone into smaller stones that can more easily be passed spontaneously. Kidney stones smaller than about 15 mm, proximal ureteral stones, upper and middle pole kidney stones, and stones not composed of cystine or calcium oxalate monohydrate respond best. *Ureteroscopy* involves the passage of a semirigid or flexible scope through the bladder and into the ureter. It is a mainstay of surgical stone extraction for lower pole renal calculi smaller than 1 cm,[■] for most ureteral stones and especially for distal ureteral stones.[■] Endocorporeal lithotripsy can directly fragment visualized stones. *Percutaneous nephrolithotomy* involves placement of a fiberoptic catheter with an open lumen through the flank and into the collecting system. Instruments and lasers are inserted through the lumen and used to fragment and remove the stone. This technique is quite effective in removing large (>2 cm) or staghorn calculi, as well as stones that do not fragment well with extracorporeal shock-wave lithotripsy.

Medical Evaluation

After an initial stone episode, the recurrence rate in nontreated patients is estimated to be about 50% at 5 years. As a result, every patient who forms an initial stone should undergo evaluation with the goal of preventing recurrent stones. The history should focus on uncovering the reason a patient may have formed a stone at this point in time. All stones should be analyzed to assist in defining the underlying metabolic abnormality and to guide therapy.

A careful dietary history, including an estimate of fluid intake, is essential to determine its potential contribution to stone formation. Many stone formers are erroneously instructed to eliminate all calcium from their diet, a practice that not only increases stone formation but also enhances bone demineralization. Sodium intake should be estimated because sodium excretion obligates calcium excretion, thereby potentially leading to urinary supersaturation. Excessive animal protein intake increases metabolic acid production, which induces bone demineralization and increases calciuria.

Hypercalcemic disorders—including malignancy (Chapter 245), hyperparathyroidism (Chapter 245), and sarcoidosis (Chapter 95)—often result in hypercalciuria, increased urinary supersaturation, and calcium stone formation. Malabsorptive GI disorders, such as Crohn disease (Chapter 141) and celiac disease (Chapter 140), or weight reduction surgery (Chapter 220), such as ileal resection or jejunoileal bypass, will often result in calcium oxalate stone formation owing to increased oxalate absorption and excretion and volume depletion.

Medications that can cause calcium stone formation include loop diuretics, which increase urinary calcium excretion, as well as salicylates and probenecid, which increase urinary uric acid excretion. Other medications that can themselves precipitate into stones include IV acyclovir, high-dose sulfadiazine, triamterene, and the antiretroviral agents indinavir and nelfinavir. Still other medications, such as acetazolamide and topiramate, inhibit tubular carbonic anhydrase activity, thereby leading to metabolic acidosis, bone resorption, hypercalciuria, lower urinary citrate excretion, and higher urinary pH, all of which can result in the formation of calcium phosphate stones.

The number and frequency of stones formed, the patient's age at the time of the first stone, the size of the stone, and an analysis of the composition of the stone are also important clues. Stones that develop at a young age suggest a genetic disorder, such as primary hyperoxaluria or cystinuria. Large staghorn calculi in elderly patients are consistent with struvite stones. A stone's response to intervention is also helpful: cystine stones do not fragment well with lithotripsy, and stones that recur frequently in a single kidney suggest a unilateral anatomic abnormality.

The basic evaluation of a stone patient includes measurement of serum electrolytes (sodium, potassium, chloride, and bicarbonate), creatinine, calcium, phosphorus, and uric acid, as well as a 25 hydroxyvitamin D level and the level of thyroid stimulating hormone. If the serum calcium is above the midrange of normal and the serum phosphorus is below the midrange of normal, a serum parathyroid hormone level should be obtained.

An elevated urine specific gravity suggests inadequate fluid intake. Patients with struvite stones generally have an elevated urine pH (>7.4) owing to the splitting of urea into ammonia and bicarbonate, but a low urinary pH (<5.5) raises the suspicion of uric acid stones. Hematuria may indicate irritation to the urothelial lining by a stone. Characteristic crystals (see Fig. 114-10 in Chapter 114) may be seen in stone formers and are more common than in nonstone formers. The presence of hexagonal crystals (see Fig. 114-9 in Chapter 114) mandates that cystinuria be excluded. The combination of an elevated urine pH with bacteriuria (Chapter 284) suggests struvite stones. Urease production adequate to stimulate struvite stone formation may be present despite low bacterial colony counts, so the laboratory should be asked to identify all bacteria and determine antibiotic sensitivities even with low colony counts. If no bacteria are isolated, cultures for *Ureaplasma urealyticum* should be requested. The combination of an elevated urinary pH (6.5-7.2) with a low serum potassium or serum bicarbonate level strongly suggests distal renal tubular acidosis (Chapter 118).

In patients with recurrent stones and in patients with high-risk characteristics for recurrence, a 24-hour urine collection can determine the levels of calcium, oxalate, citrate, sodium, urate, phosphorus, and creatinine, as well as supersaturation with respect to calcium oxalate, calcium phosphate, and uric acid (Table 126-1). The presence of supersaturation should lead the clinician to determine the individual urinary components that are causing the increased supersaturation, and efforts can then be made to rectify these abnormalities. Cystine should also be measured at least once in every stone former to exclude cystinuria and regularly in patients who form cysteine stones. Multivitamins should be discontinued about 5 days before the collection to prevent any antioxidant effect on the urine sample.

TABLE 126-1 OPTIMAL 24-HOUR URINE VALUES IN PATIENTS WITH NEPHROLITHIASIS*

PARAMETER	VALUE
Volume	>2-2.5 L
pH	>5.5 and <7.0 (24-hr specimen not required)
Calcium	<300 mg or <3.5-4.0 mg/kg in men; <250 mg or <3.5-4.0 mg/kg in women
Oxalate	<40 mg
Sodium	<3000 mg
Uric acid	<800 mg in men <750 mg in women
Phosphorus	<1100 mg
Citrate	>320 mg
Supersaturation with respect to calcium oxalate	<5
Supersaturation with respect to calcium phosphate	0.5-2
Supersaturation with respect to uric acid	0-1

*Urine creatinine should be measured to ensure adequacy of collection and should be >15 mg/kg in men and >10 mg/kg in women. Supersaturation is the ratio of the ion activity product and its solubility product.

PREVENTING RECURRENT STONES

General treatment to prevent recurrent stone formation (Fig. 126-3) consists of advising the patient to increase oral fluid intake to result in a urine volume greater than 2 L/day,[9] which will decrease recurrent stone formation by about 50%.[■]

Calcium Stones

Because urine sodium excretion is directly correlated with urine calcium excretion, decreasing dietary sodium intake will decrease urine calcium excretion and reduce supersaturation. Patients with calcium stones should be instructed to limit their daily sodium intake to no more than 2 g/day. As the metabolism of animal protein leads to hypercalciuria, patients should also reduce animal protein intake to 0.8 to 1.0 g/kg/day.[■]

The recommended calcium intake for a 19- to 50-year-old man or woman is 1000 mg of elemental calcium per day, and a number of studies demonstrate decreased formation of calcium stones when people consume diets adequate in calcium, presumably because appropriate calcium intake is needed for intestinal binding of dietary oxalate by dietary calcium. Dairy products are preferred over calcium supplements because data suggest that women taking supplemental vitamin D and calcium have a significant increase in stone formation.

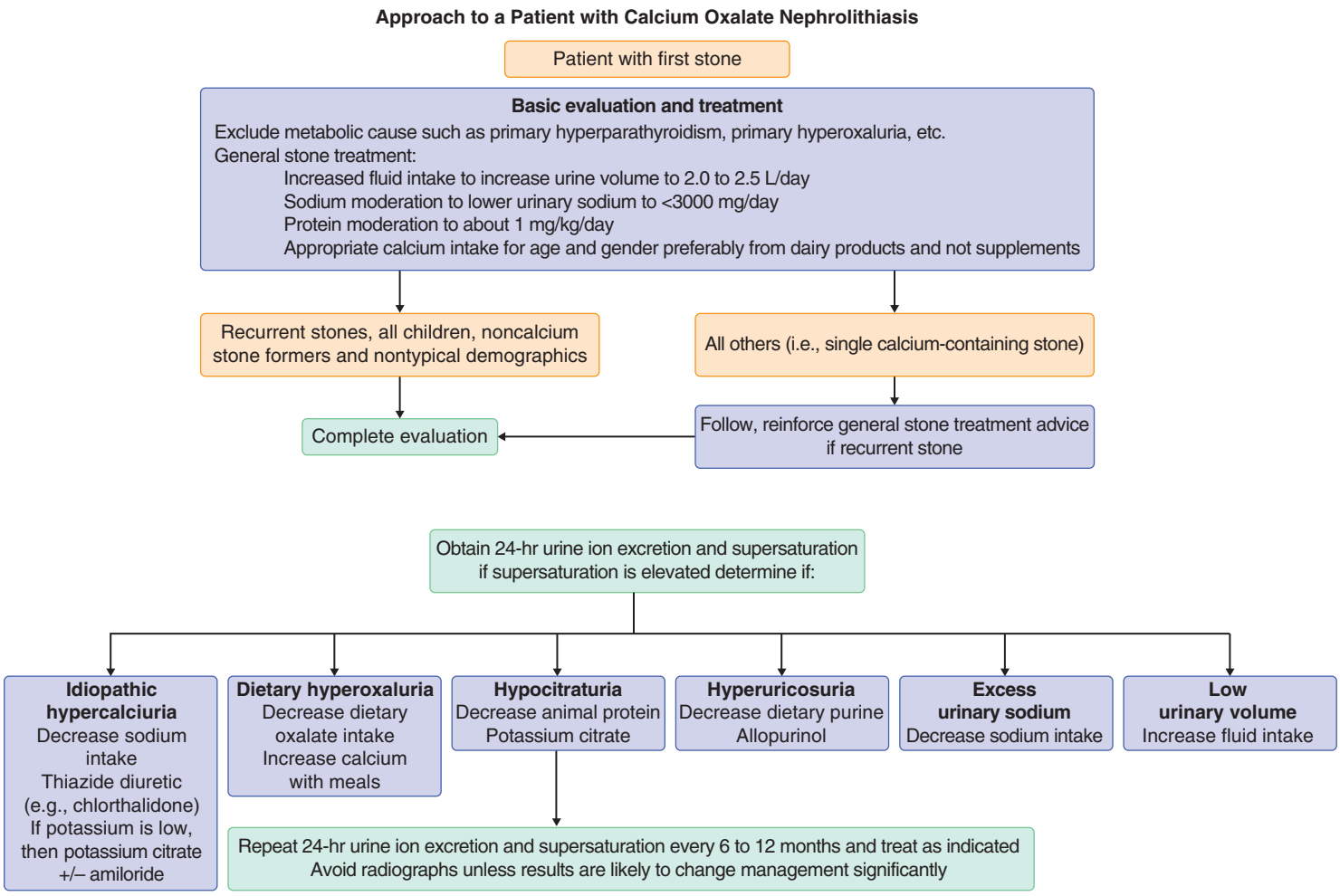

FIGURE 126-3. Approach to a patient with calcium oxalate nephrolithiasis.

Patients with persistent hypercalciuria often benefit from thiazide diuretics (e.g., chlorthalidone [12.5-25 mg/day]), which directly lower urinary calcium and reduce recurrent stone formation by about 50%.[■] Thiazides are effective only if patients restrict dietary sodium. In patients with hypercholesterolemia or hyperglycemia, indapamide (1.25-5 mg/day) can be used because it has less effect on these parameters. If patients develop hypokalemia, dietary potassium intake should be increased or a potassium supplement can be administered. Potassium citrate (10-40 meq/day) will increase urinary citrate excretion, bind urinary calcium, and further decrease recurrent stone formation. Potassium citrate is available as a wax-matrix tablet. A 24-hour urine collection can be repeated in a month or two to assess the response to therapy.

Patients with dietary hyperoxaluria should be instructed to limit or avoid foods, such as cocoa, nuts, tea, and leafy green vegetables such as spinach, that have a high oxalate content. Because dietary calcium binds dietary oxalate, patients should consume these foods with calcium-containing foods. For enteric hyperoxaluria, treatment is first directed at the underlying disorder and then at instituting therapy for cause of the steatorrhea (Chapter 140). Dietary oxalate should be restricted, and dietary calcium and oxalate should be ingested at the same meal. A diet that is low in fat; high in fruits, vegetables, and low-fat dairy products; that is rich in grains, fish, poultry, beans, seeds, and nuts; and that contains less sweets, added sugar, and red meat appears to be a reasonable alternative to a low-oxalate diet.[■] Additional fluid intake and potassium citrate are often beneficial.

In some patients with type 1 primary hyperoxaluria, pyridoxine (vitamin B_6) can increase enzyme activity and reduce oxalate production. All patients with primary hyperoxaluria should be treated with measures that reduce calcium oxalate precipitation, such as ample fluid supplementation and potassium citrate (10-40 meq/day). These patients should be seen by specialists because prompt and effective treatment can forestall kidney failure. Liver transplantation (Chapter 154) can be curative.

Calcium stones may be found in patients with hyperuricosuria.[10] These patients often excrete excess amounts of urinary uric acid but normal amounts of urinary calcium and oxalate. Compared with patients with pure uric acid stones, they generally have a higher urinary pH (≈5.5). The mechanism by which uric acid promotes calcium stone formation is unclear. Therapy generally consists of dietary purine restriction, increased fluid intake, and the addition of allopurinol (300 mg/day) if necessary.

If moderation of dietary protein is not successful in patients with hypocitraturia, oral potassium citrate is given in a wax-matrix formulation (10 to 40 mEq/day). Serum levels of potassium and bicarbonate must be closely monitored, especially in patients with chronic kidney disease.

Uric Acid Stones

Uric acid stones are radiolucent and are thus most often visualized with ultrasonography or CT. Therapy for patients with uric acid stones begins with nonspecific measures such as increasing fluid intake, a low purine diet, and lowering animal protein intake to increase urinary pH. Ideally, the urinary pH should be elevated to approximately 6.5 to 7.0, a level that not only prevents new stone formation but also can dissolve existing uric acid stones without promoting calcium phosphate deposition. Potassium citrate, again in a wax-matrix formulation (10-40 mEq/day), may be required to raise the urinary pH sufficiently. If all else fails, the carbonic anhydrase inhibitor acetazolamide (250-500 mg/day) may be initiated to raise urine pH. Because hyperuricemia usually persists, allopurinol (100-300 mg/day) is commonly indicated to lower the serum uric acid level.

Struvite Stones

Struvite stones rapidly grow to a large size and may promptly recur if they are not completely removed. As a result, struvite stone therapy requires complete surgical stone removal coupled with appropriate long-term antibiotic therapy (Chapter 284) selected on the basis of cultures of stone fragments retrieved from surgery. Antibiotics should be continued at full doses until the urine is sterile and then continued at a lower dose. Monthly surveillance cultures should be continued until the urine remains sterile for 3 consecutive months.

Antibiotics can then be discontinued with monthly surveillance urine cultures for another year.

Cystine Stones

Cystine kidney stones, which usually develop by the second or third decade of life, are radiopaque and may appear as staghorn calculi or multiple stones. The disease should be suspected in any patient with an early onset of stones, frequently recurrent nephrolithiasis, and a family history of the disease. Although the presence of the classic hexagonal cystine crystals in the urine (see Fig. 114-9 in Chapter 114) may suggest the diagnosis, anyone suspected of the disorder should have a quantitative cystine measurement on a 24-hour urine sample.

The goal of treatment is to lower the urinary cystine concentration below the limits of solubility. Patients are advised to drink sufficient quantities of water to keep all excreted cystine in solution. Patients should moderate dairy products and high-protein foods because they contain large amounts of methionine, which is a precursor of cystine. Because cystine is more soluble at a higher pH, urinary alkalinization with potassium citrate (10-40 mEq/day) can be used to maintain a urinary pH between 6.5 and 7.0. Chelating agents can reduce the free cystine concentration by forming more soluble compounds, but they should be prescribed only by a specialist because of their high risk of side effects.

PROGNOSIS

The recurrence rate of calcium oxalate nephrolithiasis is about 50% at 5 to 10 years, and the recurrence rate is higher for cystine, uric acid, and struvite stones. Patients with kidney stones have a nearly twofold increase in all-cause mortality because of their age, male gender, race, and poverty level but no increase after adjusting for these factors.[11]

Grade A References

A1. Lindqvist K, Hellstrom M, Holmberg G, et al. Immediate versus deferred radiological investigation after acute renal colic: a prospective randomized study. *Scand J Urol Nephrol*. 2006;40:119-124.

A2. Smith-Bindman R, Aubin C, Bailitz J, et al. Ultrasonography versus computed tomography for suspected nephrolithiasis. *N Engl J Med*. 2014;371:1100-1110.

A3. Holdgate A, Pollock T. Nonsteroidal anti-inflammatory drugs (NSAIDs) versus opioids for acute renal colic. *Cochrane Database Syst Rev*. 2005:CD004137.

A4. Worster AS, Bhanich Supapol W. Fluids and diuretics for acute ureteric colic. *Cochrane Database Syst Rev*. 2012;2:CD004926.

A5. Campschroer T, Zhu Y, Duijvesz D, et al. Alpha-blockers as medical expulsive therapy for ureteral stones. *Cochrane Database Syst Rev*. 2014;4:CD008509.

A6. Srisubat A, Potisat S, Lojanapiwat B, et al. Extracorporeal shock wave lithotripsy (ESWL) versus percutaneous nephrolithotomy (PCNL) or retrograde intrarenal surgery (RIRS) for kidney stones. *Cochrane Database Syst Rev*. 2014;11:CD007044.

A7. Aboumarzouk OM, Kata SG, Keeley FX, et al. Extracorporeal shock wave lithotripsy (ESWL) versus ureteroscopic management for ureteric calculi. *Cochrane Database Syst Rev*. 2012;5:CD006029.

A8. Fink HA, Wilt TJ, Eidman KE, et al. Medical management to prevent recurrent nephrolithiasis in adults: a systematic review for an American College of Physicians Clinical Guideline. *Ann Intern Med*. 2013;158:535-543.

A9. Escribano J, Balaguer A, Roque IFM, et al. Dietary interventions for preventing complications in idiopathic hypercalciuria. *Cochrane Database Syst Rev*. 2014;2:CD006022.

A10. Noori N, Honarkar E, Goldfarb DS, et al. Urinary lithogenic risk profile in recurrent stone formers with hyperoxaluria: a randomized controlled trial comparing DASH (Dietary Approaches to Stop Hypertension)-style and low-oxalate diets. *Am J Kidney Dis*. 2014;63:456-463.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Scales CD Jr, Smith AC, Hanley JM, et al. Prevalence of kidney stones in the United States. *Eur Urol*. 2012;62:160-165.
2. Rendina D, De Filippo G, D'Elia L, et al. Metabolic syndrome and nephrolithiasis: a systematic review and meta-analysis of the scientific evidence. *J Nephrol*. 2014;27:371-376.
3. Wu W, Yang B, Ou L, et al. Urinary stone analysis on 12,846 patients: a report from a single center in China. *Urolithiasis*. 2014;42:39-43.
4. Lieske JC, Turner ST, Edeh SN, et al. Heritability of urinary traits that contribute to nephrolithiasis. *Clin J Am Soc Nephrol*. 2014;9:943-950.
5. Cochat P, Rumsby G. Primary hyperoxaluria. *N Engl J Med*. 2013;369:649-658.
6. Sakhaee K. Epidemiology and clinical pathophysiology of uric acid kidney stones. *J Nephrol*. 2014;27:241-245.
7. Iqbal MW, Youssef RF, Neisius A, et al. Contemporary management of struvite stones using combined endourologic and medical treatment: predictors of unfavorable clinical outcome. *J Endourol*. 2013. [Epub ahead of print].
8. Sumorok N, Goldfarb DS. Update on cystinuria. *Curr Opin Nephrol Hypertens*. 2013;22:427-431.
9. Xu H, Zisman AL, Coe FL, et al. Kidney stones: an update on current pharmacological management and future directions. *Expert Opin Pharmacother*. 2013;14:435-447.
10. Arowojolu O, Goldfarb DS. Treatment of calcium nephrolithiasis in the patient with hyperuricosuria. *J Nephrol*. 2014;27:601-605.
11. Tang J, Mettler P, McFann K, et al. The association of prevalent kidney stone disease with mortality in US adults: the National Health and Nutrition Examination Survey III, 1988-1994. *Am J Nephrol*. 2013;37:501-506.

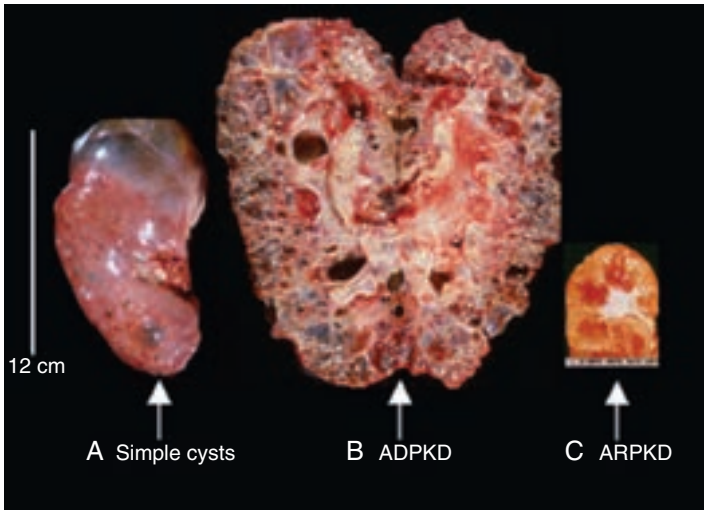

FIGURE 127-1. Gross pathology of selected cystic kidney diseases. A, Photograph of a kidney with multiple simple cysts. The cysts bulge out from the surface of a normal-sized kidney. **B,** Sagittal cross section of a kidney from an adult with autosomal dominant polycystic kidney disease (ADPKD). Multiple macroscopic cysts have resulted in an enlarged but still reniform kidney (note the evidence of prior hemorrhage within some of the cysts). **C,** Sagittal cross section of a kidney segment from a neonate with autosomal recessive polycystic kidney disease (ARPKD). The kidney is enlarged, with numerous small cysts. (Courtesy Dr. Robert Colvin, Massachusetts General Hospital.)

127

CYSTIC KIDNEY DISEASES

M. AMIN ARNAOUT

DEFINITION AND EPIDEMIOLOGY

The term *cystic kidney diseases* refers to a heterogeneous group of hereditary and acquired disorders characterized by the presence of unilateral or bilateral renal cysts. When acquired singly or in small numbers and in the absence of any other pathology, renal cysts are termed *simple cysts*, which are present in approximately 50% of individuals older than 40 years, are usually not loculated, and tend to bulge out from the renal surface (Fig. 127-1). The *polycystic kidney diseases* (PKDs), by comparison, constitute a clinically important group of genetically mediated disorders characterized by prominent, expanding, typically bilateral renal cysts. PKDs are classified as dominant, recessive, or X-linked, based on their pattern of inheritance. Autosomal dominant PKD (ADPKD), with a prevalence rate between one in 400 and one in 1000, is the most common monogenic disease in humans and accounts for about 8% to 10% of all end-stage renal disease (ESRD) in the United States. ADPKD develops in an age-dependent manner and affects mainly adults. Autosomal recessive PKD (ARPKD), by contrast, is a relatively rare childhood disorder that appears in one in every 6000 to 50,000 live births. Renal cysts are also seen in several other rare hereditary kidney diseases and in several syndromic PKDs (Table 127-1). Collectively, the hereditary PKDs generally affect both genders and all races equally and cost more than $1 billion annually to manage in the United States alone. *Acquired cystic kidney disease* refers to the multiple bilateral renal cysts that occur in 90% of patients who have been receiving renal replacement therapy for 8 years or longer and are associated with increased rates of renal cell carcinoma.

AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE

PATHOBIOLOGY

ADPKD is a systemic disorder characterized by cyst formation in multiple organs, including the kidneys, other ductal organs, and the cardiovascular system. Renal cysts originate as outpouchings of tubules and may arise from any portion of the nephron, with up to 1% of nephrons involved. The outpouchings expand and eventually separate from the parent tubules, yielding cysts (Fig. 127-2). Cyst growth is caused by proliferation of the cyst lining cells and by abnormal fluid accumulation that results when chloride-driven fluid secretion outpaces absorption.[1] Cyst expansion and fibrosis (induced by cystic epithelium-derived chemokines, cytokines, and growth factors that attract macrophages and fibroblasts) cause compression and obstruction of noncystic normal tubules, thereby resulting in upstream tubular dilation. The kidneys become massively enlarged, and kidney function progressively declines.

Genetics

Heterogeneous mutations in two genes, *PKD1* and *PKD2*, cause ADPKD. Heterogeneous mutations in *PKD1* and PKD2 account for approximately

TABLE 127-1 COMPARISON OF CLINICAL FEATURES OF CYSTIC KIDNEY DISEASES

DISEASE	INHERITANCE	FREQUENCY	GENE PRODUCT	AGE OF ONSET	CYST ORIGIN	RENOMEGALY	CAUSE OF ESRD	OTHER MANIFESTATIONS
ADPKD	AD	1:400-1000	Polycystin-1 Polycystin-2	20s and 30s	Anywhere (including Bowman capsule)	Yes	Yes	Liver cysts Cerebral aneurysms Hypertension Mitral valve prolapse Kidney stones, UTIs
ARPKD	AR	1:6000-10,000	Fibrocystin/ polyductin	First year of life	Distal nephron, CD	Yes	Yes	Hepatic fibrosis Pulmonary hypoplasia Hypertension
ACKD	No	90% of ESRD patients at 8 yr	None*	Years after onset of ESRD	Proximal and distal tubules	Rarely	No	None
Simple cysts	No	50% in those older than 40 yr	None*	Adulthood	Anywhere (usually cortical)	No	No	None
NPHP	AR	1:80,000	Nephrocystins (NPHP1–11)	Childhood or adolescence	Medullary DCT	No	Yes	Retinal degeneration; neurologic, skeletal, hepatic, cardiac malformations
MCKD	AD	Rare	MUC1 and Uromodulin	Adulthood	Medullary DCT	No	Yes	Hyperuricemia, gout
MSK	No	1:5000-20,000	None*	30s	Medullary CD	No	No	Kidney stones
Tuberous sclerosis	AD	1:10,000	Hamartin (TSC1), tuberin (TSC2)	Childhood	Loop of Henle, DCT	Rarely	Rarely	Renal cell carcinoma Tubers, seizures, angiomyolipoma, hypertension
VHL syndrome	AD	1:40,000	VHL protein	20s	Cortical nephrons	Rarely	Rarely	Retinal angioma, CNS hemangioblastoma, renal cell carcinoma, pheochromocytoma
Oral-facial-digital syndrome-1	XD	1:250,000	OFD1 protein	Childhood or adulthood	Renal glomeruli	Rarely	Yes	Malformation of the face, oral cavity, and digits; liver cysts; mental retardation
BBS	AR	1:65,000-160,000	BBS 1-14	Adulthood	Renal calyces	Rarely	Yes	Syndactyly and polydactyly, obesity, retinal dystrophy, male hypogenitalism, hypertension, mental retardation

*No known genetic susceptibility.
ACKD = acquired cystic kidney disease; AD = autosomal dominant; ADPKD = autosomal dominant polycystic kidney disease; AR = autosomal recessive; ARPKD = autosomal recessive polycystic kidney disease; BBS = Bardet-Biedl syndrome; CD = collecting duct; CNS = central nervous system; DCT = distal convoluted tubule; ESRD = end-stage renal disease; MCKD = medullary cystic kidney disease; MSK = medullary sponge kidney; NPHP = nephronophthisis; UTI = urinary tract infection; VHL = von Hippel-Lindau; XD = X-linked dominant.

85% and 15% of cases of ADPKD, respectively. In a minority of ADPKD cases, no demonstrable *PKD1* or *PKD2* mutations are found, suggesting that a third gene may be involved.

PKD1 is 54 kb long and is located on chromosome 16p13.3, adjacent to the tuberous sclerosis 2 (*TSC2*) gene. *PKD1* has 46 exons, generates a 14-kb transcript, and encodes a 4302- residue protein called polycystin-1 (PC1), the first described member of an expanding polycystin protein family. The 5′ region of human *PKD1* (to exon 33) is replicated on the same chromosome, resulting in approximately six copies of *PKD1*-like pseudogenes, which must be distinguished from the *PKD1* gene in direct mutational analysis. To date, 1923 truncating mutations have been identified throughout the gene (http://pkdb.mayo.edu/) but especially in the 3′ half. Mutations in the 5′ half of the gene are associated with more severe disease, with only 19% of patients retaining adequate renal function at 60 years of age (vs. 40% of those with 3′ mutations), and they are more likely to have intracerebral aneurysms and aneurysm rupture.

The 68-kb *PKD2* gene is located on chromosome 4q13-23, transcribes 15 exons, and generates a 5-kb transcript, which encodes a protein of 968 residues called polycystin-2 (PC2). To date, 241 heterogeneous mutations in *PKD2* have been identified. The greater number of cysts that occur at an early age in PKD1 compared with PKD2 disease is likely a result of a higher frequency of mutations in the longer *PKD1* coding region and the presence of six *PKD1*-like pseudogenes that may be involved in recombination-based gene conversion and rearrangements.

Because renal cysts develop from approximately 1% of nephrons, somatic inactivation of the normal *PKD1* or *PKD2* allele (somatic second hit model) has been proposed as the main mechanism of cyst initiation. The *PKD1* or *PKD2* haploinsufficiency state may produce wide stochastic fluctuations in level of the normal gene product and may reduce it to below a critical disease-causing threshold in the absence of a somatic second hit (haploinsufficiency model). Increasing evidence seems to support the latter model. It remains plausible that the genomic instability associated with the *PKD1* or *PKD2* haploid state may increase the likelihood of somatic second hits, which contribute to disease progression by providing cysts with a growth or survival advantage.

There are wide variations in the onset and severity of ADPKD even among affected members of the same family. This variability could arise from variable frequency and timing of the somatic inactivation of the respective normal allele. Variability in disease onset and severity can also result from bilineal inheritance of *PKD1* and *PKD2* mutant alleles or by inheritance of hypomorphic or incompletely penetrant variants of either gene. Mosaicism in a parent in whom a mutation has arisen de novo could result in one sibling having the disease while another is disease free despite sharing the identical inherited haplotype at the *PKD* locus. Digeneic inheritance of mutant *PKD1*, *PKHD1, TSC2,* or *HNF1B* alleles can result in ADPKD in infancy. Paradoxically, combining conditional inactivation of the polycystins with ablation of cilia (by inactivating, for example, the ciliopathy genes encoding heterotrimeric kinesin component Kif3a or the intraflagellar transport [protein ift20]) suppress cyst growth in developing and adult kidney and liver. Other modifier loci beyond known disease genes as well as nongenetic risk factors (e.g., smoking, caffeine and male hormones) also likely influence disease variability.

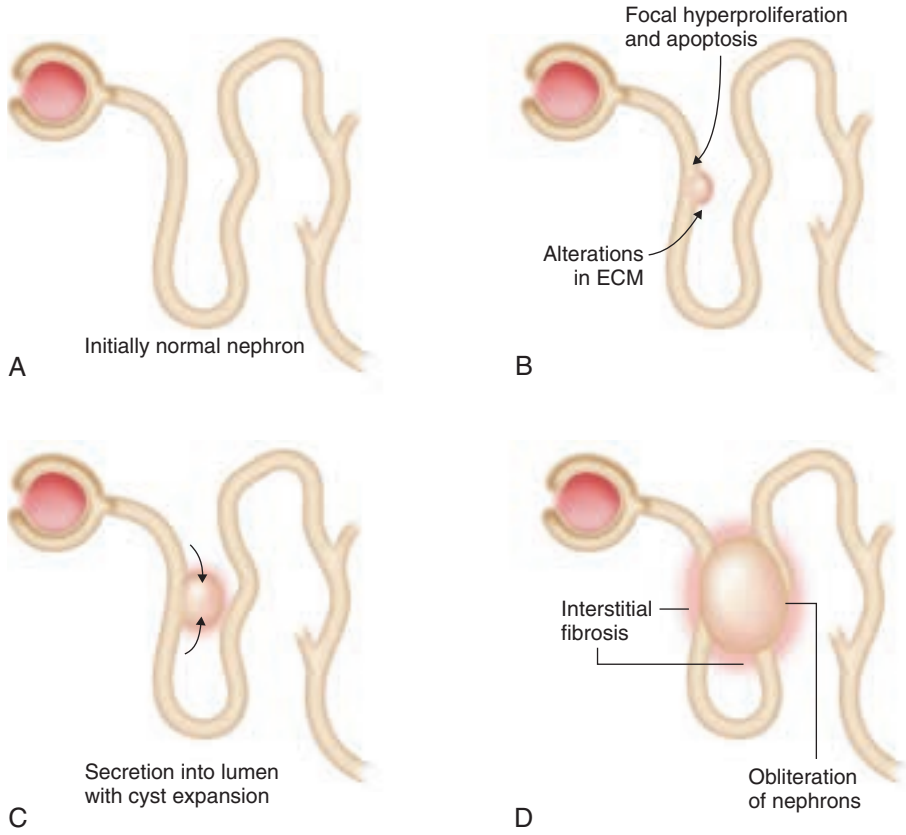

FIGURE 127-2. **A to D, Steps involved in cyst formation in autosomal dominant polycystic kidney disease. Note that this process occurs hundreds or thousands of times during the natural history of the disease. ECM = extracellular matrix.**

Gene Products

PKD1 encodes PC1, which consists of a large extracellular modular ectodomain followed by an 11-pass transmembrane segment and a short cytoplasmic carboxyl terminus (E-Fig. 127-1). The PC1 ectodomain contains multiple functional motifs, which are also present in cell adhesion receptors, thereby suggesting a role in cell–cell or cell–matrix interactions (or both). Cleavage of PC1 at a membrane proximal G protein–coupled receptor proteolytic site found within the G-protein–coupled receptor autoproteolysis inducing domain is necessary for normal function and results in N-terminal and C-terminal fragments that remain tethered. PC2, the protein encoded by *PKD2*, is a six-membrane spanner that acts as a nonselective voltage-dependent, calcium-permeable ion channel with cytoplasmic amino- and carboxy termini. Its six-transmembrane segment bears topologic and sequence similarity to the carboxyl terminal six-transmembrane segment of PC1. The coiled–coil domains within the carboxyl termini of PC1 and PC2 interact, thereby facilitating PC1 translocation to the plasma membrane and stabilizing the channel activity of PC2. The latter function is regulated by casein kinase 2 (CK2β), which binds the intracellular PLAT domain of PC1.

PC1 and PC2 are widely expressed in tissues, with some overlap consistent with their direct interaction. PC1 expression is highest in fetal tissue and progressively declines thereafter, becoming mainly limited to the collecting duct in the normal adult kidney. PC1 expression is induced in injured adult kidney. By comparison, renal expression of PC2 is maintained throughout development and predominates in the medullary thick ascending limb and distal cortical tubules in the normal adult kidney. PC1 is found at multiple cell membrane sites, including primary cilia, adherens junctions, desmosomes, and focal adhesions, as well as in intracellular vesicles. PC2 is found in primary cilia as well as the basolateral membrane, endoplasmic reticulum, centrosome, and mitotic spindles in dividing cells. Both PC1 and PC2 are also found in urinary exosome-like vesicles.

Experimental studies have suggested that developing renal tubules with high proliferative indices are more sensitive to the reduction in PC1 or PC2 levels, with disease severity radically altered depending on the time of gene inactivation. Whereas *Pkd1* or *Pkd2* inactivation in mice during proliferation of immature tubular epithelium leads to massive cysts and embryonic or neonatal death, inactivation when the renal epithelium has already differentiated into recognizable nephron segments causes mild disease. Maintenance of higher proliferative indices in distal nephron epithelium relative to that in the proximal nephron may explain the predominance of cysts in the distal nephron in adult human ADPKD.

CLINICAL MANIFESTATIONS

ADPKD has a highly variable presentation, even within families. The clinical features of *PKD2*-associated ADPKD are indistinguishable from those of *PKD1*-associated disease. However, *PKD2* disease is milder, with an older average age of onset—74 years versus about 54 years for *PKD1*-associated disease—owing to the development of fewer cysts at an early age rather than to a slower cyst growth rate.[2]

Despite an estimated 100% penetrance by age 90 years, only half the individuals with heterozygous mutations in *PKD1* or *PKD2* are ever diagnosed clinically with ADPKD. Of these patients, the majority present in the third or fourth decade of life with symptoms referable to renal cystic disease. However, ADPKD can develop at any age, including infancy, and it can have a nonrenal presentation. Renomegaly may predominate the clinical picture, with abdominal distention, discomfort, or pain; however, renomegaly can also be discovered incidentally on physical examination or after radiographic studies of the abdomen.

Nocturia, one of the earliest signs of abnormal renal function in ADPKD, reflects the early impairment in urinary concentration owing to disruption of renal architecture by the cysts. Hematuria is typical, but proteinuria is less prominent than in many other renal diseases. Cyst hemorrhage and sometimes rupture, which can occur spontaneously or after trauma, present as sharp pain and hematuria. Anemia features less prominently than in other renal diseases, probably because of the relatively well-preserved erythropoietin secretion. Kidney stones (Chapter 126) occur in 20% to 36% of patients, with uric acid stones slightly more common than calcium oxalate stones; predisposing factors include urinary stasis, defective trapping of urea in the medulla, low urine pH, hypocitraturia, hyperoxaluria, hypercalciuria, and hypomagnesuria. Recurrent cyst infection (Chapter 284), usually by common urinary tract–infecting organisms, is characterized by flank or abdominal pain, fever, rigors, leukocytosis, and occasionally sepsis.

Cardiovascular disease, which is common in ADPKD, manifests as biventricular diastolic dysfunction even in young patients with normal blood pressure and renal function, thoracic and abdominal aortic aneurysms, and

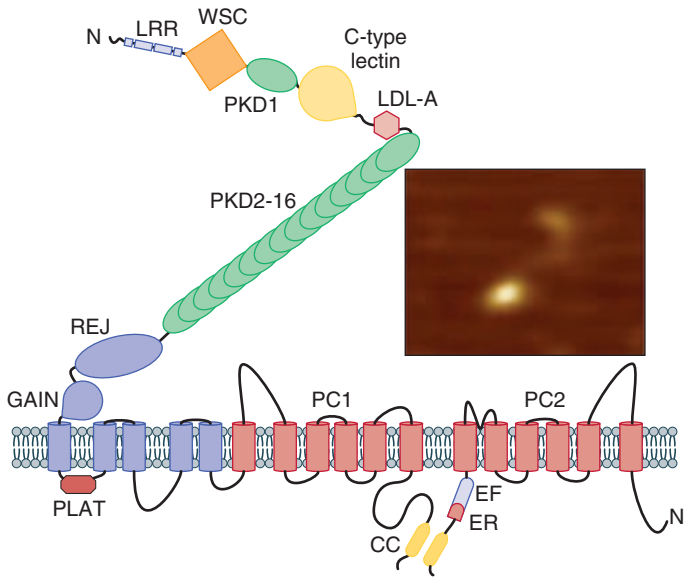

E-FIGURE 127-1. Schematic of polycystin-1 (PC1) and PC2. PC-1 is a multidomain glycoprotein with 11 putative transmembrane (TM) segments, the C-terminal six of which (colored in *orange*) bear homology to the six transmembrane segments of PC2. PC1 domains are leucine-rich region (LRR); cell wall and stress component (WSC); C-type lectin; low-density lipoprotein receptor (LDL) A-like domain (LDL-A); immunoglobulin-like domain (PKD); receptor for egg jelly domain (REJ); G-protein-couple receptor autoproteolysis inducing domain (GAIN), which contains the G protein–coupled receptor proteolytic site (GPS) needed for PC-1 activation; and PC1–lipoxygenase-α toxin domain (PLAT). *Inset:* The PC1 schematic is drawn after atomic force microscopy imaging of PC1. Full-length PC1 appeared as two unequally sized blobs. The smaller one representing the N-terminal LRR, LDL-A, the first PKD domain, and C-type lectin, and the larger one includes REJ, GAIN, all transmembrane domains, and cytoplasmic tail. The 35-nm string connecting the two blobs likely consists of the tandem PKD domains. The PC2 domain includes a calcium-binding motif (EF), an endoplasmic reticulum (ER) retention signal, and a predicted coiled-coiled (CC) domain that interacts with a CC domain in PC1.

cervicocephalic and coronary aneurysms.[3] Arterial hypertension is present in approximately 70% of cases before renal dysfunction is detected. Development of hypertension at a young age is associated with a fourfold increased risk of ESRD and increased cardiovascular morbidity. The risk for preeclampsia is also higher than in the general population.

An estimated 4% to 17% of individuals with ADPKD develop saccular cerebral aneurysms (Chapter 408),[4] a prevalence rate that is four to 10 times greater than in the general population. These aneurysms tend to segregate in families, making ADPKD one of a group of diseases characterized by autosomal dominant inheritance of cerebral aneurysms. ADPKD-associated aneurysms tend to rupture at a smaller size and in younger individuals—on average, 10 years younger than among the general population. Although usually clinically silent, intact cerebral aneurysms can present with focal neurologic symptoms and headaches. By contrast, aneurysms that rupture lead to subarachnoid hemorrhage (Chapter 408) and have dramatic presentations that include severe headaches, seizures, altered sensorium, and death. Aortic and coronary artery aneurysms are also more prevalent in patients with ADPKD, and the frequency of aortic insufficiency is increased.

Although almost never severe enough to cause end-stage liver disease, age-dependent hepatic cysts occur in 30% to 80% of patients with ADPKD and can lead to signs and symptoms of mass effect, infection, hemorrhage, and rupture.[5] Estrogen intake and multiparity in women are risk factors in developing larger and more symptomatic cysts. The cysts that occasionally form in other organs, such as the pancreas, spleen, brain, ovaries, epididymis, and prostate, are usually asymptomatic.[6] Sperm abnormalities and defective motility may occur but rarely cause male infertility. Bronchiectasis (Chapter 90) is threefold more common, inguinal hernias may be more prevalent, and colonic diverticulosis and diverticulitis are more common in ESRD patients with ADPKD.

DIAGNOSIS

Renomegaly, typically presenting in the third or forth decade of life with a positive family history and common extrarenal manifestations, such as hypertension, are useful findings in making a presumptive diagnosis of ADPKD. In these cases, renal ultrasonography is frequently diagnostic. Because only about 60% of individuals give a family history of ADPKD, ultrasound screening of asymptomatic parents or grandparents may be required to uncover diagnostically relevant silent ADPKD.

A diagnosis of ADPKD can be made in an asymptomatic individual by ultrasonography (Fig. 127-3A).[7] To account for the common age-dependent appearance of simple cysts, ADPKD is diagnosed if at least three renal cysts (distributed in one or both kidneys) are present in individuals 15 to 29 years old (sensitivity, 0.82; specificity, 1.0), 30 to 39 years old (sensitivity, 0.96; specificity, 1.0), if at least two renal cysts are present in each kidney in individuals age 40 to 59 years (sensitivity, 1.0; specificity, 0.99), or if at least four renal cysts are detected in each kidney in individuals 60 years of age or older (sensitivity, 1.0; specificity, 1.0). Fewer than two renal cysts in an individual from an ADPKD family with an unknown genotype age 40 years or older or the absence of renal cysts in such an individual age 30 to 39 years is sufficient to exclude the disease, with negative predictive values of 100% and 99.3%, respectively. A negative ultrasound result is less accurate in excluding disease in individuals younger than 30 years, so computed tomography (CT) or magnetic resonance imaging (MRI) is recommended. When a young relative is being considered as a potential kidney donor to a family member with end-stage ADPKD, contrast-enhanced, three-dimensional CT or magnetic resonance angiography (MRA) is required because these tests can detect 3-mm cysts compared with ultrasonography's ability to detect 10-mm cysts.

Cerebral aneurysms are increasingly being detected with MRI in ADPKD patients. Four-vessel cerebral angiography remains the gold standard and is often used for surgical planning.

DNA-based diagnosis of ADPKD by means of direct sequencing is commercially available and detects mutations in more than 90% of affected individuals. Next-generation exome sequencing may have a much higher sensitivity and specificity for the diagnosis.[8] Testing is recommended when imaging study results are equivocal, in a young transplant donor from an ADPKD family when imaging study results are negative, or to facilitate preimplantation genetic diagnosis. The marked allelic heterogeneity of the *PKD1* and *PKD2* mutations, as well as the paucity of phenotype–genotype correlations, contributes to the complexity of DNA-based diagnostics.

Monitoring Renal Disease Progression

Despite ongoing renal cyst expansion, the glomerular filtration rate (GFR) and the serum creatinine level are generally maintained within the normal range in ADPKD patients until the fourth to sixth decade of life. These markers are therefore insensitive for monitoring disease progression, especially in young patients. Reduced renal blood flow is the most sensitive prognostic indicator of renal progression. Estimating height-adjusted total kidney volume from measurements of kidney length, width, and thickness using MRI, CT, or ultrasonography can also reliably monitor disease progression in the early stages of ADPKD when the GFR is still in the normal range. A baseline kidney volume of 600 cm^3 or greater predicts the development of renal insufficiency with a 75% accuracy within an 8-year follow-up period. Although gadolinium-enhanced MRI provides excellent detail of kidney structures, MRI without gadolinium is recommended to monitor kidney volume, especially in ADPKD patients with a reduced GFR, because of concern about nephrogenic systemic fibrosis (Chapter 267) and further impairment of renal function associated with its use. The kidney volume in ADPKD patients progressively increases in most cases but at widely differing rates, ranging from less than 1% to more than 10% annually, and renal function declines as the kidneys enlarge.[9] Kidney volumes greater than 1,500 mL are frequently associated with a decreased GFR and gross hematuria and invariably with arterial hypertension.

PREVENTION AND TREATMENT

Rx

Management strategies are aimed at monitoring for complications of ADPKD and treating them, as well as providing counseling. Frequent monitoring and effective treatment of hypertension (Chapter 67; see Tables 67-7 and 67-8) are essential because hypertensive patients have a greater annual increase in kidney volume, as well as a higher prevalence of left ventricular

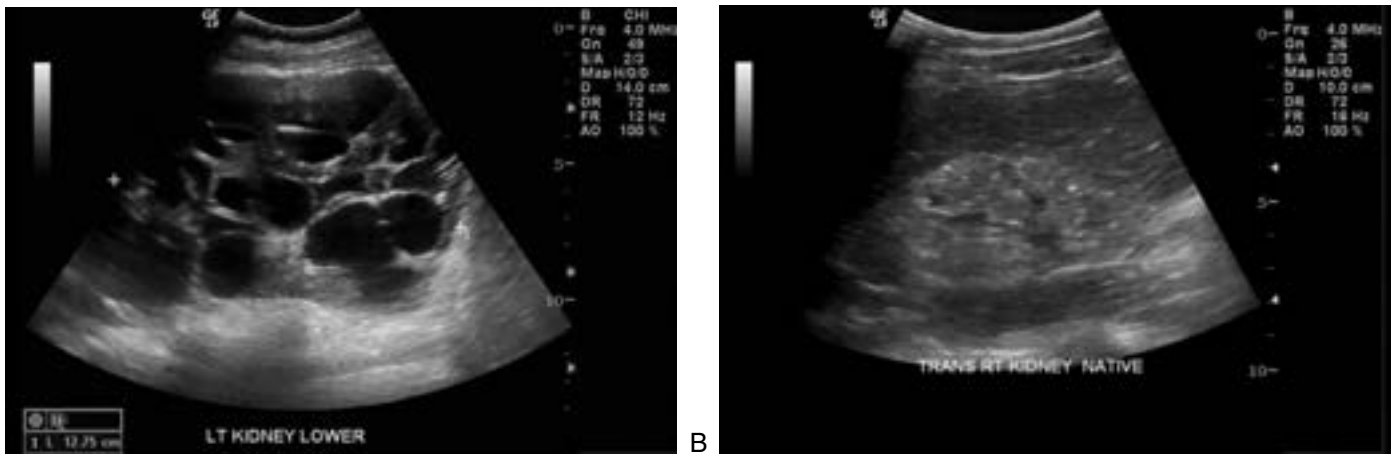

A B

FIGURE 127-3. A, Ultrasound image of a kidney from a patient with autosomal dominant polycystic kidney disease. B, Ultrasound image of a kidney from a patient with autosomal recessive polycystic kidney disease. (Courtesy Dr. Javier M. Romero and Jennifer A. McDowell, Massachusetts General Hospital.)

hypertrophy, ischemic heart disease, and stroke, compared with normotensive patients. The goals of blood pressure control are the same as for other patients with renal disease, including the attainment of a symptom-free blood pressure of 120/80 mm Hg or less. Experimental and clinical data suggest that angiotensin-converting enzyme (ACE) inhibitors may be the preferred type of medication because they promote more reversal of left ventricular hypertrophy for the same of blood pressure reduction compared with calcium channel blockers in patients with ADPKD.

Treatment of urinary tract infection (Chapter 284) and prevention of nephrolithiasis (Chapter 126) are the same as in the general population and include standard antimicrobial therapy and increased fluid intake, respectively. Renal and hepatic cyst infections are optimally treated with lipophilic antibiotics that possess cyst-penetrating capabilities, including ciprofloxacin, trimethoprim, clindamycin, and vancomycin. Blood or urine cultures and sensitivities are used to guide the choice of antibiotic therapy.

Cyst hemorrhage and rupture, with resultant pain and hematuria, are usually managed conservatively with rest and analgesics; nonsteroidal anti-inflammatory drugs (NSAIDs) are avoided owing to their antiplatelet action and potential renal toxicity. Alternatives include acetaminophen 500 mg up to every 4 hours for mild to moderate pain, nonopiates such as tramadol 50 mg up to every 4 hours for moderate to severe pain, and the addition of an oral or transdermal opioid (see Table 30-5 in Chapter 30) as needed. Patients with enlarged kidneys should be advised to avoid playing contact sports, and those with massively enlarged kidneys should refrain from wearing belts and seatbelts. Some patients with unusually painful cysts respond to cyst fluid aspiration, cyst deroofing, or ethanol-induced sclerosis.

Nephrectomy is rarely indicated before the onset of ESRD. Renal replacement therapies, including renal transplantation (Chapter 131), are at least as effective as in other causes of ESRD.

The massive cystic enlargement of the liver commonly seen by midlife in women with ADPKD makes it prudent to avoid estrogen intake and repeated pregnancies. Partial hepatectomy has been successful in improving the quality of life in patients with massive liver enlargement.

Magnetic resonance angiography screening is not routinely recommended for asymptomatic patients without a family history of cerebral aneurysm or subarachnoid hemorrhage, but it is recommended for patients with a positive personal or family history and for patients who experience a new onset of severe headache or central nervous system symptoms or signs. It may also be considered in those with high-risk occupations, such as airline pilots, and for patients with incapacitating anxiety. The risk of a patient with a positive family history developing a new aneurysm after an initially negative MRA result is 2.6% at a mean follow-up period of 9.8 years, so rescreening every 5 to 10 years in such cases may be appropriate, especially in patients with early-onset hypertension or a history of heavy smoking. Asymptomatic patients with positive family history of cerebral aneurysms and positive MRA results should be followed closely with a neurosurgeon and monitored by annual MRA screening. The decision to perform surgical clipping or endovascular embolization should take into consideration the age of the patient, the size of the aneurysm and its location, and a previous history of cerebral bleeding (Chapter 408).

Patients should be advised that their children have a 50% probability of inheriting a disease-causing germline mutation. DNA-based diagnostics are most useful in identifying the germline mutation prenatally or preimplantation. As DNA-based diagnostics become more widely available and less expensive and as promising new treatment options are developed, counseling will become an increasingly important component of prevention.

Experimental Therapies

Current therapeutic efforts are targeting increased cell proliferation of cystic epithelium and abnormal fluid secretion into cysts (E-Fig. 127-2). So far, the only effective specific therapy is oral tolvaptan (45-90 mg in the morning and 15-30 mg in the afternoon as tolerated), which slows the increase in total kidney volume and the decline in kidney function over a 3-year period compared with placebo.[■] However, tolvaptan also is associated with a fourfold increase in liver injury and could not be tolerated by nearly 25% of patients. As a result, tolvaptan has not been approved by the Food and Drug Administration for the treatment of ADPKD. The metabolically stable somatostatin analogue octreotide (at a dose of 40 mg every 28 days) can slow the expansion of renal cysts but does not slow the decline in renal function, and it also is associated with substantial side effects.[■■■] In meta-analyses, rapamycin complex 1 inhibitors, such as sirolimus and everolimus, also have not been beneficial.[■■] These data suggest that cyst growth may not be the key determinant of renal function decline.

PROGNOSIS

The rate of progression of renal disease is highest in men with poorly controlled hypertension, an early age at diagnosis, and mutations in *PKD1*. Kidney cysts develop earlier in *PKD1* than in *PKD2* disease, but the rate of expansion is similar, indicating that the basic difference in phenotype is in cyst initiation not expansion. Whereas the presence of one affected family member who developed ESRD by age 60 years is highly predictive of PKD1 disease (positive predictive value, 100%; sensitivity, 75%), the development of ESRD after age 70 years in a family member is highly predictive of PKD2 disease (positive predictive value, 95%; sensitivity, 75%). Approximately 5% of all ADPKD patients with cerebral aneurysms die from aneurysmal rupture.

AUTOSOMAL RECESSIVE POLYCYSTIC KIDNEY DISEASE

DEFINITION AND EPIDEMIOLOGY

ARPKD is a multisystem childhood disorder characterized by severe and early PKD dominated by dilation of the kidney collecting ducts, systemic hypertension, biliary ductal plate dysgenesis in neonates, and portal tract fibrosis in older children.

PATHOBIOLOGY

ARPKD has been linked to heterogeneous mutations in a single gene, *PKHD1* (polycystic kidney and hepatic disease 1). Located on chromosome 6q21, *PKHD1* has 67 exons and spans a genomic region of approximately 470 kb, the longest open reading frame of which is 12,222 base pairs long. *PKHD1* encodes a unique type I membrane protein, fibrocystin/polyductin, comprising 4074 amino acids, with a large extracellular segment and a short cytoplasmic carboxyl terminus. The precise physiologic function of fibrocystin/polyductin in collecting duct and biliary duct epithelium is not clear, but its domains are known to mediate cell motility and invasion, extracellular protein and carbohydrate binding, and catalysis of polysaccharide hydrolysis. Alternatively-spliced transcripts of *PKHD1* encode a membrane protein with variable extracellular domains, as well as forms lacking the transmembrane segment. Homozygous truncating mutations in *PKHD1* are associated with perinatal renal disease, whereas patients with homozygous missense mutations present later because some functional protein is produced.

Fibrocystin/polyductin protein encoded by *PKHD1* is predominantly expressed in the cortical and medullary collecting ducts and the thick ascending limbs of Henle. It is also found to a lesser degree in the pancreas, liver, and lungs, which are also affected in ARPKD. In common with many cystogenic proteins, fibrocystin/polyductin is found in basal bodies and primary apical cilia, suggesting that it is important in maintaining the structural integrity of cilia.

Loss of the fibrocystin/polyductin protein downregulates PC2. Fibrocystin/polyductin–PC2 interactions may regulate calcium influx mediated by PC2. It has been proposed that cystogenesis in ADPKD and ARPKD share a common mechanism, with the variability in the ARPKD phenotype traced to the degree of PC2 expression among patients. However, renal cysts can originate from any part of the nephron and are detached from the nephron proper in ADPKD, but they originate from and remain attached to the distal nephron in ARPKD, suggesting that additional signaling pathways specific for each disease and gene product likely explain these anatomic differences.

CLINICAL MANIFESTATIONS

Although ARPKD can present as renal cysts discovered radiographically either antenatally or during adulthood, it usually manifests as bilateral abdominal masses and renal insufficiency in infancy. It carries a 30% mortality rate owing to severe pulmonary hypoplasia; oligohydramnios, presumably linked to in utero renal disease, likely accounts for the pulmonary hypoplasia. Hypertension is almost universal, typically develops before renal impairment is apparent, and probably accelerates the decline in renal function. Findings related to renal tubular dysfunction may be present and include polyuria, enuresis, hyponatremia, and hyperchloremic metabolic acidosis. Cystic complications related to infection and rupture also occur, although hematuria is an infrequent finding. ESRD can take up to 20 years to develop, and in rare instances, it never occurs.

Hepatic fibrosis, secondary to dilation of the intrahepatic and extrahepatic bile ducts, manifests as recurrent ascending cholangitis (Chapter 155) and portal hypertension with splenomegaly and esophageal varices. Pancreatic fibrosis is rarely a clinical concern.

DIAGNOSIS

The demonstration by abdominal ultrasonography (see Fig. 127-3, *B*) or CT of symmetrically enlarged polycystic kidneys that retain their reniform shape (owing to uniform microcystic dilation of collecting ducts) and hepatic

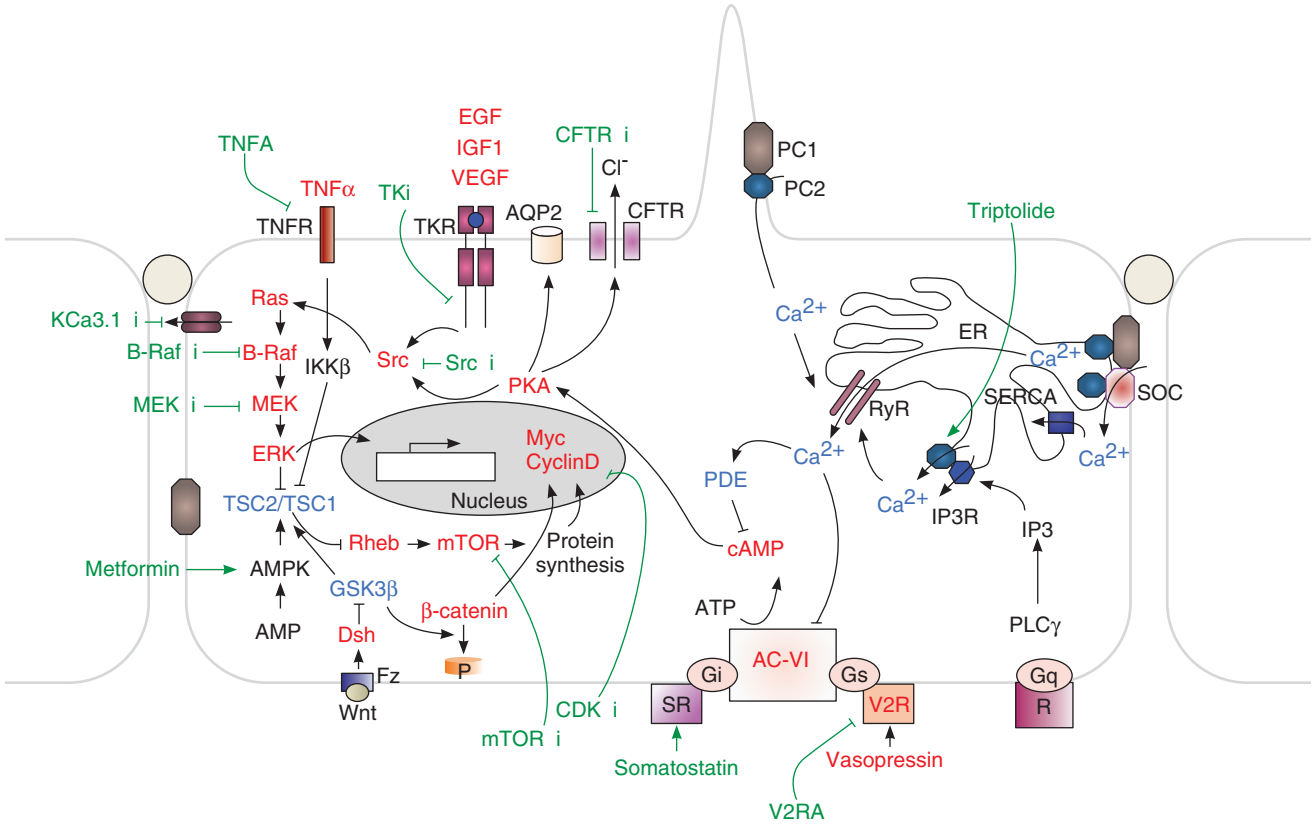

E-FIGURE 127-2. Signaling pathways that may be up- or downregulated in polycystic kidney disease and sites of current therapeutic interventions. Upregulated pathways are in *red,* downregulated pathways are in *blue,* and drug targets are in *green*. Polycystin-1 (PC1) and PC2 mediate calcium (Ca^{2+}) entry into the cell, which triggers Ca^{2+} release from the endoplasmic reticulum (ER) via the ryanodine receptor (RyR). PC2 interacts with PC1 to regulate store operated Ca^{2+} channel (SOC) activity, and it interacts with inositol 1,4,5-triphosphate (IP3) receptor (IP3R) to regulate calcium release from the ER. Reduced PC1/PC2–mediated calcium influx increases intracellular cyclic adenosine monophosphate (cAMP) levels, which stimulates, via protein kinase A (PKA), chloride, and water secretion across the luminal membrane through cystic fibrosis transmembrane conductance regulator (CFTR) and the vasopressin-sensitive aquaporin-2 (AQP2) channels, respectively, and activates mitogen-activated protein kinases/extracellular signal-regulated kinases (MAPK/ERK) signaling, which may also be activated by the mislocalized tyrosine kinase receptors (TKR). Phosphorylation of tuberin by ERK, its inadequate targeting to the plasma membrane owing to defective interaction with PC1, upregulation of tumor necrosis factor-α (TNFα), or downregulation of 5′ AMP = activated protein kinase (AMPK) signaling may dissociate the tuberin–hamartin complex, leading to the activation of Ras homolog enriched in brain guanosine triphosphate binding protein (Rheb) and mammalian target of rapamycin (mTOR). ERK and mTOR activation promote G1/S transition and cell proliferation through upregulation of cyclin D and protein translation, respectively. Upregulation of Wnt signaling also activates mTOR and the β-catenin mitogenic pathway. AC-VI = adenylate cyclase 6, the predominant AC in collecting duct principal cells; CDK = cyclin-dependent protein kinase; Dsh = dishevelled; Fz = frizzled receptor; Gq R = G-protein q–coupled receptors; i = inhibitor; KCa3.1 = Ca^{2+}-dependent K^+ channel 3.1; P = proteosome; PDE = phosphodiesterase (PDE1 in collecting duct principal cells); PKA = protein kinase A; PLCγ = phospholipase Cγ; SERCA = sarcoplasmic reticulum Ca^{2+} pump; SR = somatostatin sst2 receptor; TNFA = tumor necrosis factor antagonist; TSC = tuberous sclerosis proteins hamartin (TSC1) and tuberin (TSC2); V2R = vasopressin V2 receptor; V2RA = vasopressin V2 receptor antagonists. (Modified from Torres VE, Harris PC. Autosomal dominant polycystic kidney disease: the last 3 years. *Kidney Int.* 2009;76:149-168.)

fibrosis is sufficient to diagnose ARPKD. In contrast to ADPKD cysts, ARPKD cysts tend to retain their connections with the originating nephron. Aside from an occasional affected sibling, a family history is often not elicited. Distinguishing ARPKD from ADPKD, especially in patients presenting in childhood or adulthood, may require a liver biopsy to document otherwise undetectable hepatic fibrosis. Gene-based diagnostics in ARPKD are helpful in making a firm diagnosis, especially in patients with late-onset disease, and they are useful in preimplantation and in prenatal diagnosis.

PREVENTION AND TREATMENT

Rx

In the absence of specific therapy for ARPKD, management goals focus on early detection and on treatment of the complications of hypertension, urinary tract or cyst infection, ESRD, and portal hypertension. Kidney transplantation may be necessary in late childhood in patients with ARPKD who present with renal disease perinatally. Treatment of portal hypertension may require liver transplantation or portosystemic shunting (Chapter 154). Treatment of hypertension begins with ACE inhibitors and angiotensin receptor blockers (see Tables 67-7 and 67-8 in Chapter 67), which are generally effective. As in all children with ESRD, attention to nutrition and renal osteodystrophy is paramount (Chapter 130). With improving sensitivity of gene-based diagnostics, genetic counseling will play a more active role in prevention.

PROGNOSIS

For patients with ARPKD, the highest mortality rates occur during the first year of life. Approximately 50% to 80% of patients survive to 15 years of age.

NEPHRONOPHTHISIS

DEFINITION AND EPIDEMIOLOGY

Nephronophthisis (NPHP) is the most common genetic cause of ESRD in childhood and adolescence, accounting for 5% to 15% of cases of ESRD.[10] It is an autosomal recessive disorder caused by mutations in a number of genes that encode nephrocystins, which are expressed in the centrosome and basal body of primary cilia, with some also localizing to adherens junctions or focal adhesions in epithelia. Homozygotic and compound heterozygotic mutations in at least 11 known genes (*NPHP1* through *NPHP11*) account for about 30% of cases. Infantile, juvenile, and adolescent variants have been described based on the median age of onset of ESRD. Infantile NPHP is characterized by mutations in the *NPHP2* gene. Whereas the juvenile form caused by mutations in *NPHP1* is the most common, the adolescent form is caused by mutations in *NPHP3*. As in other genetic renal diseases, the rate of progression to ESRD is determined in part by type and severity of the genetic defect.

Nephronophthisis is characterized pathologically by renal interstitial fibrosis, tubular atrophy with basement membrane thickening and disruption, and renal cysts and diverticula that are largely restricted to the loops of Henle and distal tubules at the corticomedullary junctions. Kidney size is generally normal or reduced, except in the rare infantile variant that leads to ESRD by 3 years of age. Expression of NPHP genes at extrarenal sites accounts for the associated retinal, neural, liver, and skeletal abnormalities. Each of the NPHP genes is associated with a somewhat different phenotype, although some share common phenotypes, explained by protein–protein interactions between two or more nephrocystins. Cerebello-ocular-renal syndromes associated with NPHP can sometimes include early-onset retinitis pigmentosa, when the syndrome is caused by mutations in *NPHP5* and *NPHP6*.

CLINICAL MANIFESTATIONS

Presenting symptoms include polyuria, growth failure, and anemia. Polyuria occurs early, owing to reduced urinary concentrating ability and to salt wasting. Decreased growth rate is related to chronic dehydration, and growth failure and anemia occur with the onset of ESRD. Blood pressure is normal, and edema is absent before the onset of renal failure. Hematuria and proteinuria are absent or minimal. Patients with mutations in specific NPHP genes may also present with eye defects, oculomotor apraxia, congenital amaurosis, retinitis pigmentosa, neural anomalies, cerebellar ataxia, seizures, liver fibrosis, skeletal defects, scoliosis, cleft palate, and situs inversus.

DIAGNOSIS

The diagnosis relies on clinical suspicion of the disorder in a pediatric or adolescent patient who presents with ESRD and extrarenal manifestations such as abnormal eye movements, blindness, mental retardation, and polydactyly. Differential diagnosis includes renal dysplasia, early-onset ADPKD, urinary tract obstruction, and ARPKD. Abdominal ultrasonography, MRI (without gadolinium), electroretinography, and full neurologic and ophthalmologic evaluations should assess the patient's renal, liver, retinal, and neurologic status. Renal ultrasonography showing loss of corticomedullary differentiation, increased parenchymal echogenicity, and occasionally small medullary or corticomedullary cysts in normal-sized or moderately small kidneys is highly suggestive of juvenile NPHP in a child with severe uremia. Genetic testing may be required for a conclusive diagnosis. A renal biopsy showing evidence of chronic tubulointerstitial nephritis may be necessary to confirm the diagnosis if genetic testing is not available.

PREVENTION AND TREATMENT

Rx

Treatment is largely supportive, focusing on the progressive renal failure and the need for dialysis and transplantation. Prenatal genetic testing in families with a genetic diagnosis of NPHP is feasible. DNA-based diagnostics are most useful in identifying the germline mutation prenatally or preimplantation.

MEDULLARY CYSTIC KIDNEY DISEASE

Medullary cystic kidney disease is an autosomal dominant interstitial kidney disease that results in renal failure after the fourth decade of life. It is more rare than autosomal recessive NPHP, with which it shares a number of histologic features, including tubular basement membrane disintegration, tubular cyst formation, and tubulointerstitial inflammation and fibrosis.

Mutations in two genes, *MCKD1* and *MCKD2*, cause medullary cystic kidney disease. *MCKD1* encodes mucin 1 (MUC1), a transmembrane protein expressed in the kidney but with unknown function. Knockout studies in mice show that *Muc1* is not essential, suggesting a dominant-negative and/or gain-of-function mode of action for the mutant protein in humans. *MCKD2* encodes an 85-kD nonciliary protein, uromodulin, which is expressed on the luminal side of renal epithelium in the thick ascending limb of Henle loop and early distal convoluted tubules. Uromodulin has been associated with urate metabolism, inhibition of stone formation and renal immune response. Familial juvenile hyperuricemic nephropathy and glomerulocystic kidney disease, rare but distinct disorders, are allelic to *MCKD2*. The former is characterized by glomerular cysts, hyperuricemia-associated gouty arthritis, and early-onset ESRD, and the latter is characterized by impaired urine concentrating ability and reduced uric acid excretion.

Mutations in *MCKD1* and *MCKD2* result in a similar clinical picture except for an earlier onset of ESRD and precocious gout with mutations in *MCKD2*. Polyuria and anemia are usually not clinically present in the early stages of the renal disease. Hypertension is likely secondary to renal failure.

Corticomedullary cysts, which are present in most adult patients, cannot always be recognized on ultrasonography or CT because they tend to be very small. Except for the treatment of gout (Chapter 273), management is similar to that of patients with NPHP.

MEDULLARY SPONGE KIDNEY

Medullary sponge kidney, which is a rare disorder of unknown pathogenesis, is characterized by congenitally acquired inner medullary and papillary collecting duct dilations, hypercalciuria, and a mild defect in urinary concentration and acidification owing to tubular dysfunction.[11] Patients present with hematuria and recurrent kidney stones, usually by the second or third decade of life. Medullary sponge kidney may also be an incidental finding on an intravenous pyelogram that shows the characteristic pooling of contrast material within the cystic collecting ducts. ESRD is uncommon, and the long-term prognosis is excellent.

OTHER INHERITED CYSTIC SYNDROMES

Bardet-Biedl syndrome (BBS) is an autosomal recessive disorder characterized by vision loss; obesity; hypertension; dystrophy of the hands, kidneys, and male genitalia; delayed development of motor skills; and behavioral problems. Mutations in at least 14 BBS genes involved in maintenance and function of cilia have been identified. About 45% of cases result from mutations in *BBS1* or *BBS10*, but in about 25% of cases, the defective gene remains to be identified. Calyceal cysts and calyceal clubbing predominate the renal lesion and are best diagnosed by intravenous urography rather than ultrasonography. Renal impairment is frequent and is an important cause of death.

Oral-facial-digital syndrome is a rare neurodevelopmental ciliopathy characterized by malformations of the brain, face, oral cavity, and digits. It is inherited in an X-linked dominant pattern and is caused by defects in the *OFD1* gene, which encodes OFD1 protein expressed in the centrosome and basal body of primary cilia but has an undetermined function. Renal (primarily glomerular) cysts are found in as many as 50% of patients, all females; males carrying the mutation die in utero. ESRD has been reported in affected girls and women ranging in age from 11 to 72 years.

Autosomal dominant renal cyst formation is also seen in tuberous sclerosis and von Hippel-Lindau syndrome (Chapter 417). In tuberous sclerosis, cyst formation is commonly associated with hypertension; this disorder can resemble ADPKD and is associated with about a 5% incidence of renal cell carcinoma. In von Hippel-Lindau syndrome, cyst formation can also lead to features of ADPKD; more important, the syndrome is associated with a 25% incidence of renal cell carcinoma.

ACQUIRED CYSTIC KIDNEY DISEASE

Acquired cystic kidney disease is largely confined to the ESRD population on dialysis (Chapter 131). Cysts arise from proximal and distal tubule dilations in small end-stage kidneys regardless of cause, mode of dialysis, or presence of a functioning kidney transplant. Identifiable risk factors include duration of ESRD, older age, male gender, black race, and chronic hypokalemia.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Acquired cystic kidney disease is usually asymptomatic, but it occasionally leads to enlarged kidneys with associated abdominal discomfort and pain. Cyst hemorrhage, which is more common than cyst infection, presents with flank pain, anemia, or hematuria. The most significant complication of acquired cystic kidney disease is malignant conversion of cysts into renal cell carcinoma (Chapter 197). Carcinomas commonly present as hematuria and are two to 200 times more common in patients with acquired cystic kidney disease than in the general dialysis population.

Acquired cystic kidney disease is diagnosed by ultrasonography or CT demonstrating multiple and bilateral renal cysts in a patient with preexisting chronic renal failure or ESRD. In contrast to ADPKD and ARPKD, the kidneys are usually not enlarged, and there is no family history of PKD. Renal CT or MRI is preferable to detect cysts in small kidneys and to assess for malignant conversion.

PREVENTION AND TREATMENT

Rx

There are no strategies to prevent the appearance or delay the expansion of renal cysts in patients on hemodialysis, but cysts may stabilize or regress after successful renal transplantation. New or frank hematuria raises the concern of renal cell carcinoma (Chapter 197), which should be assessed using ultrasonography and contrast-enhanced CT. Any evidence of septa formation, solid material, or contrast enhancement within a cyst is suspicious for carcinoma and warrants consideration of nephrectomy.

PROGNOSIS

Asymptomatic acquired cystic kidney disease does not affect survival. The incidence of renal cell carcinoma in patients with acquired cystic kidney disease is approximately 0.18% per year. Although metastasis is less common at the time of diagnosis in patients with acquired cystic kidney disease than in other patients with renal cell carcinoma, the 5-year mortality rates are higher, likely related to the almost invariable coexistence of ESRD.

Grade A References

A1. Torres VE, Chapman AB, Devuyst O, et al. Tolvaptan in patients with autosomal dominant polycystic kidney disease. *N Engl J Med*. 2012;367:2407-2418.

A2. Hogan MC, Masyuk TV, Page LJ, et al. Randomized clinical trial of long-acting somatostatin for autosomal dominant polycystic kidney and liver disease. *J Am Soc Nephrol*. 2010;21:1052-1061.

A3. Caroli A, Perico N, Perna A, et al. Effect of longacting somatostatin analogue on kidney and cyst growth in autosomal dominant polycystic kidney disease (ALADIN): a randomised, placebo-controlled, multicentre trial. *Lancet*. 2013;382:1485-1495.

A4. Myint T, Rangan G, Webster A. Treatments to slow progression of autosomal dominant polycystic kidney disease: systematic review and meta analysis of randomized trials. *Nephrology (Carlton)*. 2014;19:217-226.

A5. Liu YM, Shao YQ, He Q. Sirolimus for treatment of autosomal-dominant polycystic kidney disease: a meta-analysis of randomized controlled trials. *Transplant Proc*. 2014;46:66-74.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Rowe I, Boletta A. Defective metabolism in polycystic kidney disease: potential for therapy and open questions. *Nephrol Dial Transplant*. 2014;29:1480-1486.
2. Cornec-Le Gall E, Audrezet MP, Chen JM, et al. Type of PKD1 mutation influences renal outcome in ADPKD. *J Am Soc Nephrol*. 2013;24:1006-1013.
3. Ecder T. Cardiovascular complications in autosomal dominant polycystic kidney disease. *Curr Hypertens Rev*. 2013;9:2-11.
4. Niemczyk M, Gradzik M, Niemczyk S, et al. Intracranial aneurysms in autosomal dominant polycystic kidney disease. *AJNR Am J Neuroradiol*. 2013;34:1556-1559.
5. Abu-Wasel B, Walsh C, Keough V, et al. Pathophysiology, epidemiology, classification and treatment options for polycystic liver diseases. *World J Gastroenterol*. 2013;19:5775-5786.
6. Luciano RL, Dahl NK. Extra-renal manifestations of autosomal dominant polycystic kidney disease (ADPKD): considerations for routine screening and management. *Nephrol Dial Transplant*. 2014; 29:247-254.
7. Pei Y, Hwang YH, Conklin J, et al. Imaging-based diagnosis of autosomal dominant polycystic kidney disease. *J Am Soc Nephrol*. 2014; [Epub ahead of print].
8. Tan AY, Michaeel A, Liu G, et al. Molecular diagnosis of autosomal dominant polycystic kidney disease using next-generation sequencing. *J Mol Diagn*. 2014;16:216-228.
9. Higashihara E, Nutahara K, Okegawa T, et al. Kidney volume and function in autosomal dominant polycystic kidney disease. *Clin Exp Nephrol*. 2014;18:157-165.
10. Benzing T, Schermer B. Clinical spectrum and pathogenesis of nephronophthisis. *Curr Opin Nephrol Hypertens*. 2012;21:272-278.
11. Fabris A, Anglani F, Lupo A, et al. Medullary sponge kidney: state of the art. *Nephrol Dial Transplant*. 2013;28:1111-1119.

REVIEW QUESTIONS

1. In confirming a diagnosis of autosomal dominant polycystic kidney disease in an adult patient with a positive family history, which of the following is needed?

A. Ultrasonography
B. Gene mutational analysis of the patient
C. Gene mutational analysis of the patient and at least two family members (one of whom has the disease)
D. MRI to determine total cyst volume
E. All of the above

Answer: A A positive ultrasonography result in a patient with a positive family history of autosomal dominant polycystic kidney disease is sufficient to confirm the diagnosis. Unfortunately, a negative ultrasound result does not necessarily exclude the disease in such a patient.

2. Risk factors for disease progression in autosomal dominant polycystic kidney disease include

A. male gender.
B. early onset of hypertension.
C. proteinuria.
D. height-adjusted total kidney volume greater than 600 cc/m
E. all the above.

Answer: E Male gender, early onset of hypertension, total kidney volume greater than 600 cc/m, and proteinuria are risk factors for the progression of disease.

128

HEREDITARY NEPHROPATHIES AND DEVELOPMENTAL ABNORMALITIES OF THE URINARY TRACT

LISA M. GUAY-WOODFORD

HEREDITARY NEPHROPATHIES

The proximal tubule is responsible for reclaiming most of the filtered glucose, amino acids, uric acid, phosphate, bicarbonate, and low-molecular-weight proteins. The loop of Henle and the distal nephron reabsorb approximately 30% of the filtered sodium chloride and 50% of the filtered divalent cations. The collecting duct, under the regulatory control of aldosterone, fine-tunes sodium reabsorption and secretes hydrogen and potassium ions. In the terminal collecting duct, antidiuretic hormone regulates water reabsorption and urine concentration.

Inherited renal tubular disorders are a group of conditions in which the normal renal tubular reabsorption of ions, organic solutes, and water (Chapter 116) is disrupted because of defects in single genes.[1] These defects can be categorized by the nephron segment affected (Table 128-1).

Disorders of Proximal Tubule Function

CYSTINURIA

Cystinuria is characterized by defective proximal tubular reabsorption of cystine and dibasic amino acids, resulting in increased excretion of cystine and the risk of forming cystine-containing urinary stones (Chapter 126).[2] This autosomal recessive trait has an estimated prevalence of 1 in 7000 individuals. Two cystinuria genes have been identified: *SLC7A9*, which encodes the luminal transport channel itself; and *SLC3A1*, which encodes the transporter regulatory subunit. Several large studies indicate that mutations in *SLC3A1* are more common than mutations in *SLC7A9*. Mutations in *SLC3A1* cause cystinuria type A, mutations in *SLC7A9* cause cystinuria type B, and mutations in both genes (compound heterozygotes) cause cystinuria type AB.

Although the severity of the disease is similar in all types of cystinurias, the clinical presentation can be variable, and the onset of disease may occur from infancy to the seventh decade of life. Cystine stones are radiopaque and often form the nidus for secondary calcium oxalate stones. Symptoms include renal colic, which may be associated with urinary tract obstruction or infection. Affected children can be identified by elevated urinary cystine levels, but testing must be performed after tubular transport has fully matured (at 2 years of age). Genetic testing is available, but at this point it does not offer any clinical therapeutic benefit. Conservative therapy with high urine volume and urinary alkalinization is sufficient for many patients with cystinuria. However, recurrent stone formation may cause renal damage and warrants treatment with thiol-containing agents, such as D-penicillamine (pediatric dose, 15 to 30 mg/kg/day in four divided doses; adult dose, 2 g/day in four divided doses), α-mercaptopropionylglycine (pediatric dose, 10 to 15 mg/kg/day in three divided doses; adult dose, 0.8 to 1.0 g/day in three divided doses), or captopril (pediatric dose, 12.5 mg/kg/day in two divided doses), to form soluble mixed disulfides with cystine and to maintain free urine cystine levels below 200 mg per gram of creatinine.

CYSTINOSIS

Cystinosis is the most common inherited cause of renal Fanconi syndrome; it also affects the eyes, muscles, central nervous system, lungs, and various endocrine organs. Cystinosis is an autosomal recessive disorder caused by mutations in the gene *CTNS*, which encodes cystinosin, a lysosomal cystine transporter. Defects in this transporter lead to the accumulation of intralysosomal cystine crystals and widespread cellular destruction.

TABLE 128-1 HEREDITARY NEPHROPATHIES BY NEPHRON SEGMENT

NEPHRON SEGMENT	DISORDER	INHERITANCE	OMIM	MAJOR RENAL FEATURES
PROXIMAL TUBULE				
	Renal glycosuria	AR	233100	Isolated glycosuria
	Proximal renal tubular acidosis	AR	604278	Hyperchloremic, hypokalemic, metabolic acidosis
	Carbonic anhydrase II deficiency	AR	259730	Mixed proximal and distal renal tubular acidosis
	Hartnup disease	AR	234500	Neutral aminoaciduria
	Cystinuria	AR	Type A: 220100 Type B: 604144	Urinary calculi
	Cystinosis	AR	Infantile: 219800 Late-onset: 219900 Non-nephropathic: 219750	Renal Fanconi syndrome
	Dent disease	X-linked	Dent disease 1: 300009 Dent disease 2: 300555	Nephrocalcinosis, urinary calculi; low-molecular-weight proteinuria
	Lowe syndrome	X-linked	309000	Renal Fanconi syndrome
	Hereditary fructose intolerance	AR	229600	Renal Fanconi syndrome
	Tyrosinemia, type I	AR	276700	Renal Fanconi syndrome
	Wilson disease	AR	277900	Renal Fanconi syndrome
LOOP OF HENLE				
	Bartter syndrome	AR AD	Type 1: 601678 Type 2: 241200 Type 3: 607364 Type 4: 602522 Type 5: 601199	Hypokalemic, hypochloremic metabolic alkalosis
DISTAL TUBULE				
	Gitelman syndrome	AR	263800	Hypokalemic, hypochloremic, metabolic alkalosis
	Familial hypomagnesemia with hypercalciuria	AR	248250, 248190	Severe renal magnesium and calcium wasting
	Isolated hypomagnesemia	AD	154020	Renal magnesium wasting
COLLECTING DUCT				
	Liddle syndrome	AD	177200	Low-renin hypertension
	Glucocorticoid-remediable hyperaldosteronism	AD	103900	Low-renin hypertension
	Apparent mineralocorticoid excess	AR	218030	Low-renin hypertension
	Pseudohypoaldosteronism, type 1	AR, AD	AR: 264350 AD: 177735	Hyponatremic, hypokalemic, metabolic acidosis
	Pseudohypoaldosteronism type 2 (Gordon syndrome)	AD	114300	Low-renin hypertension with hyperkalemia
	Distal renal tubular acidosis	AR, AD	AR: 602722, 605239 AD: 179800, 611590	Hyperchloremic, hypokalemic, metabolic acidosis
	Carbonic anhydrase II deficiency	AR	259730	Mixed proximal and distal renal tubular acidosis
	Nephrogenic diabetes insipidus	X-linked AR and AD	X-linked: 304800 AD and AR: 125800	Urinary concentrating defect

AD = autosomal dominant; AR = autosomal recessive; OMIM = entries in Online Mendelian Inheritance in Man, available at www.ncbi.nlm.nih.gov/omim.

Three clinical presentations have been described.[3] The most severe is infantile (classic) cystinosis, which is manifested in the first year of life with renal tubular acidosis, impaired growth, and evidence of renal Fanconi syndrome, including aminoaciduria, glucosuria, phosphaturia, and low-molecular-weight proteinuria. Progressive renal failure reaches end-stage renal disease in childhood. A less severe, late-onset (juvenile or intermediate) form causes renal dysfunction in adolescence and involves cystine deposits in the cornea. The mildest form, an ocular, non-nephropathic form, features photophobia but no renal problems.

The mainstay of cystinosis therapy is oral cysteamine (dose: 60 to 90 mg/kg/day or 1.35 to 1.90 g/m^2/day, divided every 6 hours), an aminothiol that can lower intracellular cystine content by 90%. In well-treated adolescent and young adult patients, cysteamine delays renal glomerular deterioration, enhances growth, prevents hypothyroidism, and lowers muscle cystine content. Therefore, early diagnosis and prompt, proper treatment are critical for preventing or significantly delaying the complications of cystinosis. In a randomized trial, twice-daily dosing with delayed-release cysteamine bitartrate (at approximately 70% of the patient's usual dose) was as efficacious as cysteamine for reducing white blood cell cystine levels in patients with nephropathic cystinosis,■ thereby suggesting that it is an equally effective therapy.

Disorders of Loop of Henle and Distal Tubule Function

THE BARTTER-GITELMAN DISORDERS

The Bartter-Gitelman syndromes are a group of disorders characterized by markedly reduced salt transport in the thick ascending limb of Henle (Bartter syndrome) or in the distal convoluted tubule (Gitelman syndrome).[4] Most patients with Gitelman syndrome have defects in *SLC12A3*, the gene encoding the sodium-chloride cotransporter NCCT. However, a minority of patients with the Gitelman phenotype have mutations in *CLCNKB*.

Bartter syndrome can be caused by mutations in one of four genes: *SLC12A2*, encoding the sodium-potassium-chloride cotransporter NKCC2; *KCNJ1*, encoding the ROMK1 potassium ion channel; *CLCNKB*, encoding the ClC-Kb basolateral chloride ion channel; and *BSND*, encoding barttin, a regulatory subunit required for basolateral chloride channel targeting to the membrane. These mutations cause autosomal recessive Bartter syndrome

TABLE 128-2 FEATURES OF THE INHERITED RENAL TUBULAR ACIDOSES

DISORDER	RENAL TRANSPORT DEFECT	MINIMAL URINE pH DURING ACIDOSIS	ALKALI SUPPLEMENTATION	UAG DURING ACIDOSIS
Proximal renal tubular acidosis	↓Proximal bicarbonate reabsorption	<5.5	Children: 10-15 mEq HCO_3^-/kg/day	0 or +
Carbonic anhydrase II deficiency	↓Proximal bicarbonate reabsorption and ↓distal acidification	Variable	Variable	0 or +
Distal renal tubular acidosis	↓Distal acidification	>5.5	Adults: 1-3 mEq HCO_3^-/kg/day Children: 3-6 mEq HCO_3^-/kg/day	+

HCO_3^- = bicarbonate; UAG = urinary anion gap = $[Na^+] + [K^+] - [Cl^-]$. In renal tubular acidosis, the UAG is usually 0 or positive. By comparison, the UAG is negative in metabolic acidosis associated with diarrheal illness.

types 1, 2, 3, and 4, respectively. Defects in any of these genes disrupt salt transport in the thick ascending limb, causing a furosemide-like effect (E-Fig. 128-1). In addition, severe gain-in-function mutations in *CASR*, the gene encoding the extracellular calcium ion–sensing receptor CaSR, can cause a Bartter-like phenotype (referred to as Bartter syndrome type 5) that is distinguished from the others by autosomal dominant transmission and associated hypocalcemic hypercalciuria.

Individuals with Bartter syndrome exhibit renal salt wasting, lowered blood pressure, polyuria, hypokalemic metabolic alkalosis, and hypercalciuria with a variable risk of nephrocalcinosis. In comparison, individuals with Gitelman syndrome exhibit milder renal salt wasting, normal blood pressure, hypokalemic metabolic alkalosis, hypomagnesemia, and hypocalciuria. This clinical disorder resembles the effects of long-term thiazide administration. Clinical differences between Bartter and Gitelman syndromes relate to the severity of salt wasting, whereas phenotypic differences among Bartter syndrome types 1 through 5 correlate with the specific physiologic roles played by the individual transporters or channels in the kidney and other organs.

The mainstay of treatment includes replacing salt and water losses and providing potassium supplementation to maintain serum levels greater than 3 mEq/dL. In patients with perinatal (type 1 or 2) Bartter syndrome, cyclooxygenase inhibitors (e.g., indomethacin, 2 to 4 mg/kg/day in two to four divided doses) may be beneficial. In patients with Gitelman syndrome and some patients with Bartter syndrome type 3, oral magnesium supplementation may be required to maintain serum levels above 1.2 mg/dL.

Disorders of Collecting Duct Function

LIDDLE SYNDROME (PSEUDOALDOSTERONISM)

Liddle syndrome is an autosomal dominant form of salt-sensitive hypertension (Chapter 67) caused by mutations in the α-, β-, or γ-subunits of the epithelial sodium channel, which is expressed at the apical surface of collecting duct cells and plays a critical role in maintaining salt balance and blood pressure. Both the β- and the γ-subunits regulate the channel activity of the α-subunit. Mutations in either of these regulatory subunits result in increased epithelial sodium channel activity and Liddle syndrome.

Severe hypertension typically is manifested in childhood, with features of hypokalemic metabolic alkalosis that resemble primary aldosteronism. However, renin and aldosterone secretion is suppressed in this disorder. The clinical abnormalities can be ameliorated by a low-salt diet plus a potassium-sparing diuretic (e.g., amiloride, 5 to 10 mg/day), which acts as an antagonist of the epithelial sodium channel.

DISTAL RENAL TUBULAR ACIDOSIS

Distal renal tubular acidosis (dRTA) results from failure of the collecting duct α-intercalated cells to excrete fixed acids (see Fig. 118-1). Both autosomal dominant and autosomal recessive forms of dRTA have been described. These heritable disorders include mutations in genes encoding carbonic anhydrase II, the chloride-bicarbonate exchanger AE1, and subunits of the hydrogen adenosine triphosphatase (H^+-ATPase) proton pump. Mutations in the *SLC4A1* gene encoding AE1 cause autosomal dominant dRTA and are rarely associated with recessive forms of the disease. Mutations in subunits of H^+-ATPase are the primary causes of autosomal recessive dRTA. Vacuolar H^+-ATPases (V-type ATPases) are ubiquitous, multisubunit protein complexes that mediate the ATP-dependent transport of protons. In the kidney, V-type ATPases are the major proton-secreting pumps in the distal nephron and are involved in net proton secretion (bicarbonate generation) or proton reabsorption (net bicarbonate secretion). Defects in two genes, *ATP6B1* and *ATP6N1B*, cause dRTA with or without associated sensorineural deafness.

Clinical consequences include hypokalemic, hyperchloremic metabolic acidosis; impaired growth; hypercalciuria; hypocitraturia; nephrocalcinosis; nephrolithiasis; rickets in children; and osteomalacia in adults. Classic dRTA can be distinguished from other metabolic acidoses by an inappropriately high urine pH (>5.5), diminished net acid excretion, positive urinary anion gap, and low urinary ammonium concentration (Table 128-2). Treatment with alkali supplementation (1 to 3 mEq/kg/day in adults and 3 to 6 mEq/kg/day in children) is usually effective in correcting the acidosis. In contrast to proximal RTA, urinary potassium wasting can be ameliorated with alkali therapy alone.

DEVELOPMENT OF THE KIDNEY AND URINARY TRACT

The human kidney and urogenital tract develop from three principal embryonic structures: the metanephric mesenchyme, the mesonephric (wolffian) duct, and the cloaca (Fig. 128-1). At 4 to 5 weeks of gestation, the ureteric bud originates as a diverticulum of the mesonephric duct. Reciprocal interactions between the branching ureteric bud and the metanephric mesenchyme induce kidney development, with the metanephros undergoing an epithelial transformation to form the glomeruli and the proximal and distal tubules. The ureteric bud branches give rise to the collecting ducts, the renal pelvis, the ureter, and the bladder trigone. Nephrogenesis is completed by 36 weeks of gestation.

Concurrent with the initial nephrogenic events, the urorectal fold divides the cloaca into the urogenital sinus and the future rectum. The mesonephric duct opening into the bladder becomes the vesicoureteric orifice of the trigone. Between 5 and 6 weeks of gestation, the second genital duct (müllerian duct) appears and runs in parallel with the wolffian duct. In males, the müllerian duct subsequently regresses; the wolffian duct proceeds to form the epididymis, the vas deferens, the seminal vesicle, and the ejaculatory duct. In females, the wolffian duct regresses, and the müllerian ducts fuse to form the ureterovaginal primordium, which merges with the urogenital sinus and eventually gives rise to the uterus, the oviducts, and the proximal vagina. The remnants of the allantois form the urachus, a fibrous cord that connects the bladder to the umbilicus.

Congenital abnormalities of the kidney and urinary tract are detected in about 1 in 500 fetal ultrasound examinations and account for approximately 20 to 30% of all anomalies identified in the prenatal period. Some urinary tract anomalies are asymptomatic and inconsequential, but many renal tract malformations are important causes of infant mortality as well as morbidity in older children and adults, including the progression to renal failure.[5]

ABNORMALITIES OF THE URINARY TRACT

Renal Parenchymal Malformations

Congenital defects in renal development may result in the absence of a kidney (agenesis) or abnormalities in kidney size, structure, or position. Irregularities in the renal contour may arise from the persistence of fetal lobulation or a depression in the midpole of the left kidney caused by the spleen (a "dromedary hump"). Neither irregularity impairs renal function.

RENAL AGENESIS

Renal agenesis reflects a complete failure of nephrogenesis. Unilateral agenesis can occur as an isolated abnormality or as a component of syndromic

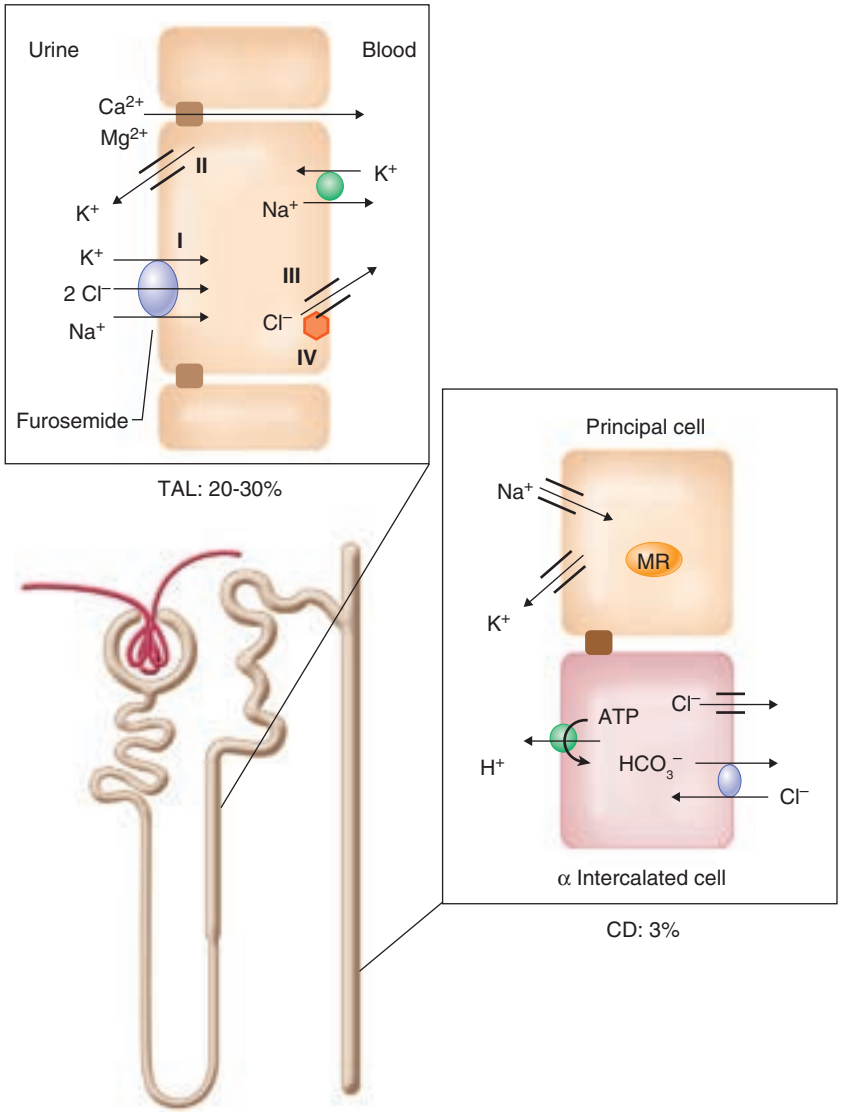

E-FIGURE 128-1. Transport in the thick ascending limb (TAL) and the pathophysiology of Bartter syndrome. Sodium (Na^+), potassium (K^+), and chloride (Cl^-) are reabsorbed across the apical membrane through the sodium-potassium-chloride cotransporter NKCC2 (defective in Bartter syndrome type 1). NKCC2 can be inhibited by furosemide. The optimal function of this transporter requires the binding of all four ions. Because of the low luminal concentration of potassium, K^+ binding becomes the rate-limiting step. Therefore, to ensure an adequate luminal supply, potassium is recycled through the ROMK1 channel (defective in Bartter syndrome type 2). These apical transport processes result in a relative excess of positive charges in the tubular lumen, providing the driving force for paracellular absorption of calcium (Ca^{2+}) and magnesium (Mg^{2+}). Na^+ exits across the basolateral membrane through the sodium-potassium adenosine triphosphatase (Na^+, K^+-ATPase) pump, whereas Cl^- exits through the chloride channels ClC-Kb (defective in Bartter syndrome type 3) and ClC-Ka. Both channels require barttin (defective in Bartter syndrome type 4) for proper membrane localization. The TAL reabsorbs approximately 25% of the filtered sodium load, whereas the collecting duct (CD) is responsible for reabsorbing about 3% of the sodium load. The profound defect in TAL sodium reabsorption leads to salt wasting, stimulation of the renin-aldosterone system, and an aldosterone-induced increase in potassium and hydrogen ion secretion in the CD, with resultant hypokalemic metabolic alkalosis. MR = mineralocorticoid receptor.

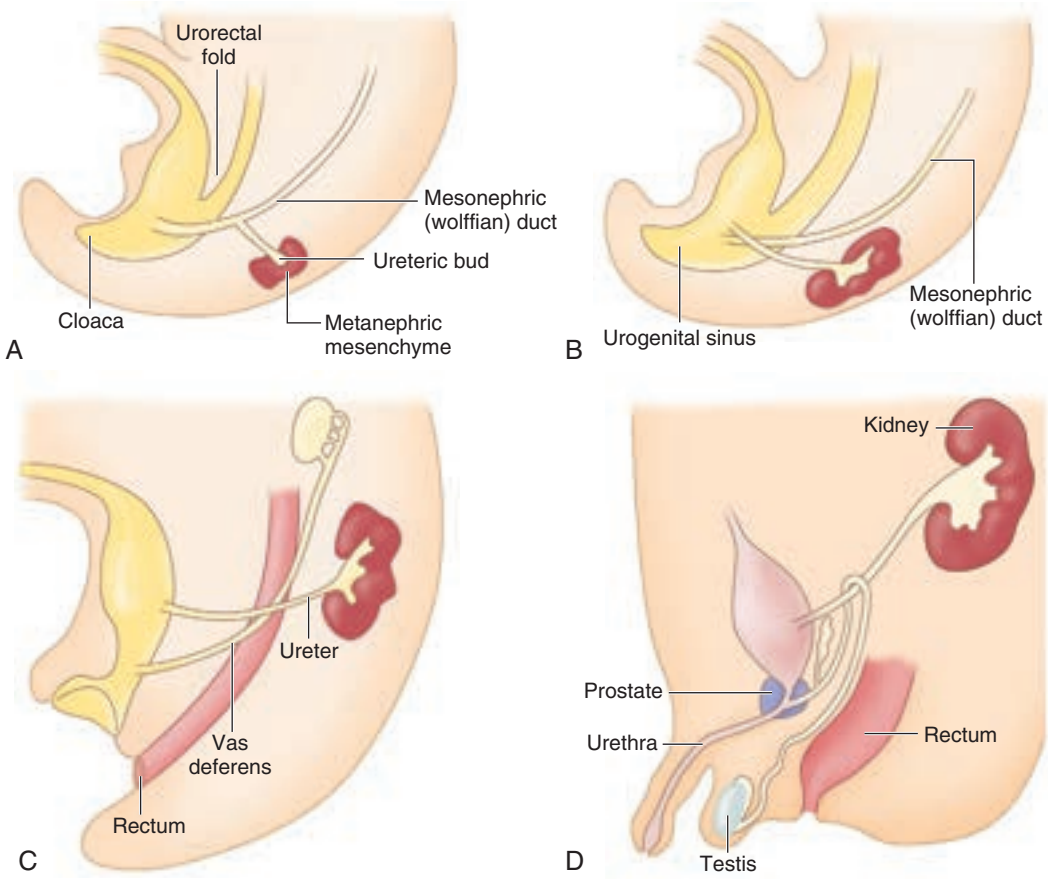

FIGURE 128-1. Key events in the development of the urinary tract. In the 4-week embryo, the ureteric bud emerges from the wolffian duct (**A**). Reciprocal interactions between the branching ureteric bud and the metanephric mesenchyme induce kidney development. Concurrently, the cloaca is divided by the urorectal fold into the urogenital sinus and the future rectum (**B**). In the 8-week male embryo, the wolffian duct begins to give rise to the epididymis, the seminal vesicles, and the caudal part of the vas deferens (**C**). By 9 weeks, axial growth of the fetal spine prompts the developing kidney to ascend from the pelvis to its final lumbar position. The external genitalia develop between 8 and 16 weeks, and testicular descent begins in month 7 of gestation (**D**).

disorders, such as Turner syndrome (Chapter 233). As an isolated entity, the complete absence of one kidney occurs in 1 in 1000 to 1500 individuals. The incidence is higher in males and occurs somewhat more frequently on the left side; in about half the patients, the ipsilateral ureter and hemitrigone are also absent. The remaining kidney is usually enlarged as a result of compensatory hypertrophy, but it may be ectopic or malrotated. Vesicoureteral reflux is observed on the contralateral side in about 30% of patients.

Renal agenesis is commonly associated with genital anomalies, suggesting that it represents a developmental field defect. In females, absence of the ipsilateral oviduct and malformation of the uterus and vagina result from maldevelopment of the müllerian duct; whereas in males, wolffian duct–derived structures, such as the vas deferens and the seminal vesicles, are often absent. Other associated anomalies can involve the cardiovascular system (30%), the musculoskeletal system (14%), and the adrenal gland (10%). Unilateral renal agenesis is found in 30% of patients with the vertebral, imperforate anus, trachea-esophageal, and renal (VATER) syndrome.

Bilateral renal agenesis has an estimated incidence of 1 in 4000 births and is associated with the Potter phenotype, which includes pulmonary hypoplasia, a characteristic facies, and deformities of the spine and limbs. At birth, these neonates have a critical degree of pulmonary hypoplasia that is incompatible with survival. The familial association of unilateral and bilateral renal agenesis, renal dysplasia, and congenital hydronephrosis occurs in hereditary renal adysplasia syndrome (Online Mendelian Inheritance in Man entry 191830), a rare autosomal dominant disorder with variable penetrance.

RENAL HYPOPLASIA

The term *renal hypoplasia* describes small kidneys with normally differentiated nephrons that are reduced in number. *Oligomeganephronia* describes a form of bilateral renal hypoplasia with a marked reduction in nephron number and associated hypertrophy of individual glomeruli and tubules. This abnormality occurs sporadically as an isolated developmental defect that must be differentiated from acquired renal atrophy and the nephronophthisis–medullary cystic disease complex. Renal function declines slowly, with progression to end-stage renal failure in the second to third decade of life.

RENAL DYSPLASIA

Renal dysplasia, which can be associated with various abnormalities of kidney size, results from abnormal metanephric differentiation that causes anomalous or incompletely differentiated renal elements. Small dysplastic kidneys are commonly referred to as *aplastic.* Large dysplastic kidneys are often cystic; the most extreme type is referred to as *multicystic dysplastic kidney.*

Unilateral dysplasia may be asymptomatic well into adult life. Small aplastic and large multicystic dysplastic kidneys are nonfunctioning and can be distinguished from renal agenesis by imaging studies. The ipsilateral ureter is typically atretic. Contralateral malformations, including obstruction and vesicoureteral reflux, are common. Unilateral multicystic kidneys involute over time and often disappear. Unilateral aplasia and multicystic dysplasia may be manifestations of the hereditary renal adysplasia syndrome. Bilateral multicystic dysplastic kidneys are incompatible with neonatal survival.

Renal and Ureteral Structural Abnormalities

RENAL MALROTATION AND ECTOPIA

Metanephric kidney development begins caudally. By 9 weeks of gestation, the kidney has ascended to its normal level (L1 to L3), and the renal pelvis has rotated 90 degrees toward the midline. Anomalies of ascent and failure of rotation are common. Mutations in the dual serine-threonine and tyrosine protein kinase gene (*DSTYK*) are found in 2.3% of patients with congenital abnormalities of the kidney or urinary tract.[6] Bilateral renal ectopia is often associated with kidney fusion. The most common fusion anomaly is the horseshoe kidney, which occurs in 1 in 500 newborns with a 2 : 1 male predominance. Renal ascent is prevented by the root of the inferior mesenteric artery (Fig. 128-2A). Crossed renal ectopia can occur with or without fusion. Supernumerary (extra) kidneys are typically ectopic and vary in location.

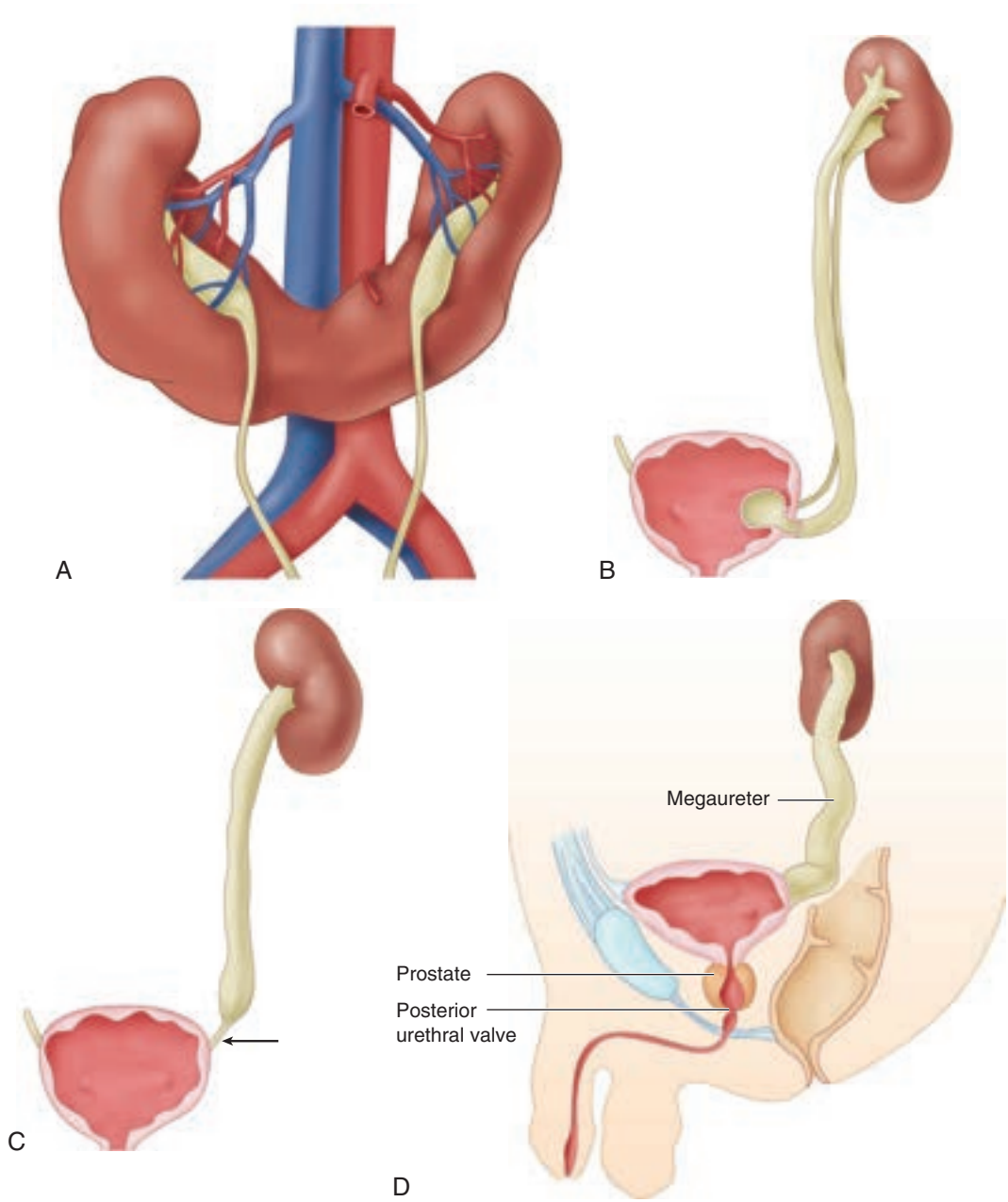

FIGURE 128-2. Developmental abnormalities of the urinary tract. A, Horseshoe kidney. **B,** Ectopic ureter associated with a ureterocele. **C,** Megaureter with the aperistaltic segment (*arrow*). **D,** Bladder outlet obstruction caused by posterior urethral valves.

Although almost one third of patients with renal ectopia remain asymptomatic, the associated malrotation of the renal pelvis increases the risk of hydronephrosis, infection, and stone formation.

PELVIURETERAL ABNORMALITIES

Obstruction of the ureteropelvic junction impedes the flow of urine from the renal pelvis into the ureter. It is one of the most frequently occurring urinary tract anomalies in infants, occurring in 1 in 500 live births. In congenital obstruction of the ureteropelvic junction, urologic anomalies in the contralateral system are common, including renal agenesis, renal dysplasia, multicystic dysplasia, ureteropelvic junction obstruction, and vesicoureteral reflux. Ureteropelvic junction obstruction may occur in adults secondary to external compression, kinking, or stenosis of the proximal ureter. Surgical intervention is indicated if there is associated renal function impairment, pyelonephritis, stones, or pain.

Hydrocalyx or *hydrocalycosis* refers to dilation of a major calyx that occurs in the context of intrinsic obstruction, as in infundibular stenosis, or in the context of extrinsic compression of the pelvis, as caused by a vessel or a parapelvic cyst. In comparison, *megacalycosis* represents a nonobstructive, dysplastic lesion seen primarily in males, in which the calyces are dilated and usually increased in number. Associated renal medullary hypoplasia causes malformation of the renal papillae.

Calyceal diverticula are cystic structures connected by a narrow channel to an adjacent minor calyx. In imaging studies, these diverticula typically fill with contrast material, which distinguishes them from renal parenchymal cysts.

Partial duplication of the renal pelvis and ureter is a common anomaly that occurs more frequently in females, is typically unilateral, and is clinically insignificant.

URETERIC ANOMALIES

Ectopic ureters usually reflect complete ureteric and renal duplication. Approximately 10% are bilateral. The ectopic ureter typically drains the dysplastic upper pole of a duplex kidney and inserts below the normal vesicoureteral junction into the lower trigone or the proximal urethra. Ectopic ureters occur much more frequently in females, and the insertion sites can include the vagina and the vulva, with resulting incontinence. An ectopic ureter is often associated with a ureterocele, a cystic dilation of the terminal ureter (Fig. 128-2B). In children, ureteroceles can be associated with urinary tract infection and obstruction of the bladder neck or even of the contralateral ureter. In adults, the clinical presentation usually involves an associated infection, ureteric stones, or both.

A megaureter, or grossly dilated ureter, has multiple potential causes, including intrinsic ureteric obstruction by a stone, bladder outflow obstruction, vesicoureteral reflux, and external compression of the distal ureter. In

contrast, primary megaureter results from a functional obstruction of the distal ureter caused by an aperistaltic segment (Fig. 128-2C).

VESICOURETERAL REFLUX

In the normal urinary tract, urinary reflux from the bladder into the ureters is prevented by a functional valve-like mechanism at the vesicoureteral junction. The competence of this valve is dependent on several critical factors, such as the intramural length of the ureter, the position of the ureteric orifice in the bladder, and the integrity of the bladder wall musculature.

Primary vesicoureteral reflux results from incompetence of the vesicoureteral junction due to the shortened length of the ureter's submucosal segment and the lateral, ectopic position of its orifice. It is estimated to occur in 1 to 2% of children. Genetic factors appear to contribute to the pathogenesis of primary vesicoureteral reflux because there is a 30- to 50-fold increased risk in immediate relatives of an index case. A deleterious heterozygous mutation (T3257I) in the gene encoding tenascin XB (*TNXB* in 6p21.3), which is expressed in the human uroepithelial lining of the ureterovesical junction and may be important for generating tensile forces, is associated with familial vesicoureteral reflux.[7] As the intramural ureter lengthens with age, primary vesicoureteral reflux tends to remit or to disappear. Vesicoureteral reflux can also be secondary to obstructive maldevelopment of the lower urinary tract in children, such as in triad syndrome and posterior urethral valves, or secondary to masses that obstruct the bladder or urethra (in adults). In both primary and secondary vesicoureteral reflux, intrarenal reflux can lead to the development of reflux nephropathy, a tubulointerstitial lesion (Chapter 122) associated with gross scarring at the renal poles. In addition, the development of a glomerular lesion consistent with focal and segmental glomerulosclerosis (Chapter 121) can cause proteinuria, hypertension, and progressive loss of renal function.

Primary vesicoureteral reflux is diagnosed in about 30 to 40% of children who have imaging studies after urinary tract infections (Chapter 284). Management of these children has been controversial with respect to both antibiotic prophylaxis and surgical correction. Nevertheless, there is general agreement that frequent urinary tract infection, higher grades of vesicoureteral reflux, and the presence of bladder and bowel dysfunction are particular risk factors for renal cortical scarring and that the rates of spontaneous resolution of reflux and of endoscopic surgical success depend on bladder and bowel dysfunction.[8,9] In a randomized trial, antibiotic prophylaxis substantially reduced recurrent urinary tract infections but did not reduce renal scarring[A1] in children with vesicoureteral reflux.

Surgical correction is the current standard of care for severe grades of vesicoureteral reflux and for recurrent symptomatic infections despite medical management, particularly for patients with secondary forms associated with maldevelopment of the lower urinary tract. However, endoscopic polyacrylate polyalcohol copolymer injection can correct grade IV and grade V vesicoureteral reflux with an 83% success rate that is comparable to surgical correction, so this alternative should be considered as a minimally invasive treatment option.[10]

Lower Urinary Tract Abnormalities

TRIAD SYNDROME (PRUNE-BELLY SYNDROME, EAGLE-BARRETT SYNDROME)

Triad syndrome, also referred to as prune-belly syndrome or Eagle-Barrett syndrome, involves a constellation of anomalies including congenital absence or deficiency of the abdominal wall musculature, gross ureteral dilation, bladder wall thickening, prostatic hypoplasia, and bilateral undescended testes (cryptorchidism). The full syndrome is expressed only in males, and surviving individuals are typically infertile. Patients with an incomplete syndrome can have anomalies of the abdominal wall musculature, bladder, and upper urinary tract; 3% of these patients are females. Although the specific molecular events have yet to be defined, defects in mesenchymal development appear to cause poor prostate and bladder differentiation, ureteral smooth muscle aplasia with consequent ureteral aperistalsis, and varying degrees of renal dysplasia. Three fourths of patients with triad syndrome have associated malformations in the cardiopulmonary system, gastrointestinal tract, and skeleton. In the immediate postnatal period, prognosis depends on the severity of extragenitourinary anomalies. Long-term outcome is based on the degree of renal dysplasia and the success of urodynamic management.

BLADDER ABNORMALITIES

Bladder exstrophy results from a midline closure defect involving the lower anterior abdominal wall, the bladder, and the external genitalia. These abnormalities have been attributed to a primary defect in the differentiation of the cloacal membrane, but the precise molecular events are unclear. In severe cases, bladder exstrophy may be associated with imperforate anus and rectal atresia. However, other congenital anomalies are rarely associated. Clinical studies indicate that there is a correlation between the success of bladder reconstruction and long-term preservation of renal function.

In adults, neuropathic or neurogenic bladder (Chapter 26) has numerous etiologic contributors, including central nervous system trauma, stroke, disorders such as Parkinson disease, spinal trauma, multiple sclerosis, and peripheral nerve damage caused by trauma or surgery. In children, myelomeningocele (spina bifida) is the most common cause of neurogenic bladder dysfunction. Other forms of myelodysplasia, such as spinal dysraphism (spina bifida occulta) and sacral agenesis, are less common causes.

POSTERIOR URETHRAL VALVES

In male infants, posterior urethral valves are the most common cause of bladder outflow obstruction, with resulting bilateral hydronephrosis and megaureters. However, among all infants with hydronephrosis, only 10% have posterior urethral valves. The urethral obstruction results from defective reabsorption of mucosal folds in the posterior urethra, just distal to the verumontanum. As a result, dilation of the proximal urethra, bladder wall hypertrophy and trabeculation, associated vesicoureteral reflux, and varying degrees of renal dysplasia are present (Fig. 128-2D). Surgical management strategies are dictated by the age of the child and the degree of associated renal insufficiency. Survival and long-term renal outcome depend on the severity of the associated renal dysplasia.

Grade A References

A1. Langman CB, Greenbaum LA, Sarwal M, et al. A randomized controlled crossover trial with delayed-release cysteamine bitartrate in nephropathic cystinosis: effectiveness on white blood cell cystine levels and comparison of safety. *Clin J Am Soc Nephrol*. 2012;7:1112-1120.

A2. Hoberman A, Greenfield SP, Mattoo TK, et al. The RIVUR Trial Investigators. Antimicrobial prophylaxis for children with vesicoureteral reflux. *N Engl J Med*. 2014;370:2367-2376.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Devuyst O, Knoers NV, Remuzzi G, et al. Rare inherited kidney diseases: challenges, opportunities, and perspectives. *Lancet*. 2014;383:1844-1859.
2. Sumorok N, Goldfarb DS. Update on cystinuria. *Curr Opin Nephrol Hypertens*. 2013;22:427-431.
3. Nesterova G, Gahl WA. Cystinosis: the evolution of a treatable disease. *Pediatr Nephrol*. 2013;28: 51-59.
4. Fremont OT, Chan JC. Understanding Bartter syndrome and Gitelman syndrome. *World J Pediatr*. 2012;8:25-30.
5. Rodriguez MM. Congenital Anomalies of the Kidney and the Urinary Tract (CAKUT). *Fetal Pediatr Pathol*. 2014;33:293-320.
6. Sanna-Cherchi S, Sampogna RV, Papeta N, et al. Mutations in DSTYK and dominant urinary tract malformations. *N Engl J Med*. 2013;369:621-629.
7. Gbadegesin RA, Brophy PD, Adeyemo A, et al. TNXB mutations can cause vesicoureteral reflux. *J Am Soc Nephrol*. 2013;24:1313-1322.
8. Tekgül S, Riedmiller H, Hoebeke P, et al. EAU guidelines on vesicoureteral reflux in children. *Eur Urol*. 2012;62:534-542.
9. Springer A, Subramaniam R. Relevance of current guidelines in the management of VUR. *Eur J Pediatr*. 2014;173:835-843.
10. De Badiola FI, Soria R, Vagni RL, et al. Results of treatment of grades IV and V vesicoureteral reflux with endoscopic injection of polyacrylate polyalcohol copolymer. *Front Pediatr*. 2013;1:32.

129 BENIGN PROSTATIC HYPERPLASIA AND PROSTATITIS

STEVEN A. KAPLAN

BENIGN PROSTATIC HYPERPLASIA

DEFINITION

The prostate gland is composed of four zones: peripheral, central, transitional, and stroma. It can also be divided by lobes: anterior, posterior, lateral, and median (Fig. 129-1). Benign prostatic hyperplasia (BPH) is a condition that can lead to lower urinary tract symptoms, which can have a significant negative impact on quality of life.[1] Complaints associated with BPH may relate to difficulty with voiding (e.g., urinary hesitancy, weak stream, straining, and prolonged voiding) or difficulty with controlling urinary storage (e.g., urinary urgency, nocturia). BPH is a histologic diagnosis defined by the proliferation of smooth muscle and epithelial cells within the prostatic transition zone.

EPIDEMIOLOGY

Age is the major risk factor for BPH. A histologic diagnosis of BPH will develop in approximately 50% of men older than 40 years. Of these men, approximately 50% will develop notable lower urinary tract symptoms,

which increase in prevalence in a linear fashion between the ages of 40 and 80 years.[2]

PATHOBIOLOGY

An enlarged prostate may cause lower urinary tract symptoms by directly obstructing the flow of urine or by increasing the muscle tone of the prostate. In addition, changes in the vascularity of the prostate or the urinary bladder can contribute to the development of symptoms. The degree of prostatic enlargement, which can contribute to and affect the severity of the symptoms, is highly variable. Enlargement typically is a combination of stromal hypertrophy and glandular hyperplasia, mostly in the central zone. In BPH, the calculated volume exceeds 30 mL.

CLINICAL MANIFESTATIONS

The presence of lower urinary tract symptoms is often indicative of bladder outlet obstruction secondary to BPH. Symptoms may significantly impair health-related quality of life and are classified as voiding (hesitancy, weak stream, straining, and prolonged voiding), storage (frequency, urgency, nocturia, urge incontinence, and voiding of small volumes), or postmicturition (postvoid dribble, incomplete emptying). Most patients who have lower urinary tract symptoms present with a combination of these symptoms. The symptoms of overactive bladder (Chapter 26) and lower urinary tract symptoms secondary to BPH often overlap.

DIAGNOSIS

Assessment should begin with a medical history and review of the patient's medications. The medical history should include any causes that may lead to bladder dysfunction, such as cerebrovascular disease, previous surgical procedures, and a history of prostatic disease. Diuretics and over-the-counter preparations, such as nasal decongestants and antihistamines, may exacerbate the patient's urinary symptoms. Furthermore, dietary factors such as water, caffeine, alcohol, and artificial sweeteners can be important contributors to the overall clinical manifestations of symptoms because they serve as direct bladder irritants and diuretics. Urinary symptoms should be assessed in a standardized fashion with validated instruments such as the International Prostate Symptom Score (IPSS) (Fig. 129-2) and observed over time.

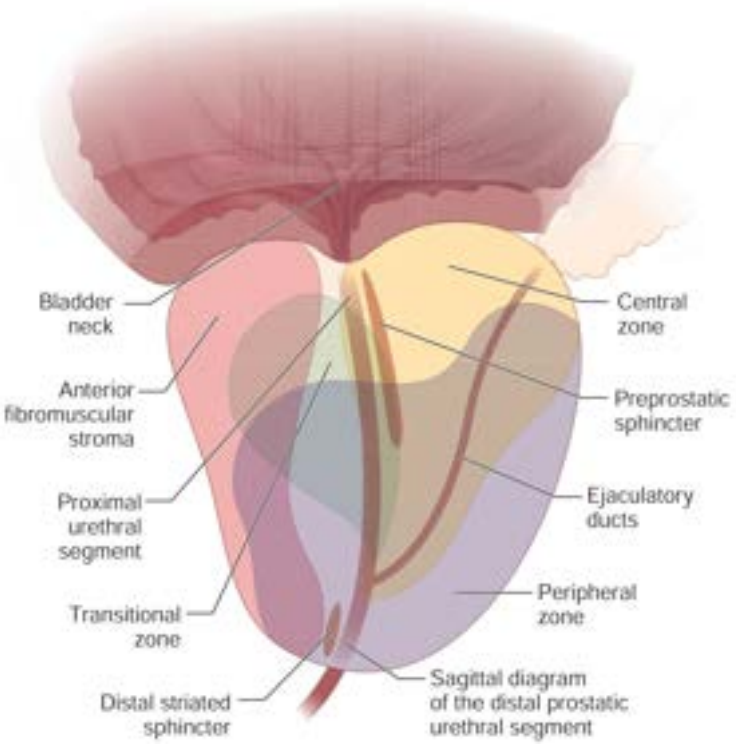

FIGURE 129-1. Anatomy of the prostate.

	Not at all	Less than 1 time in 5	Less than half the time	About half the time	More than half the time	Almost always
1. Over the past month or so, how often have you had a sensation of not emptying your bladder completely after you finished urinating?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
2. Over the past month or so, how often have you had to urinate again less than two hours after you finished urinating?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
3. Over the past month or so, how often have you found you stopped and started again several times when you urinated?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
4. Over the past month or so, how often have you found it difficult to postpone urination?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
5. Over the past month or so, how often have you had a weak urinary stream?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
6. Over the past month or so, how often have you had to push or strain to begin urination?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐

7. Over the past month, how many times did you most typically get up to urinate from the time you went to bed at night until the time you got up in the morning?

0 ☐ none	1 ☐ 1 time	2 ☐ 2 times	3 ☐ 3 times	4 ☐ 4 times	5 ☐ 5 or more times

Total IPSS Score = sum of questions 1–7 = _____

Quality of life due to urinary symptoms

If you were to spend the rest of your life with your urinary condition just the way it is now, how would you feel about that?

Delighted	Pleased	Mostly satisfied	Mixed—about equally satisfied and dissatisfied	Mostly dissatisfied	Unhappy	Terrible
0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐

FIGURE 129-2. International Prostate Symptom Score (IPSS). The seven symptom questions constitute a scale initially developed by the American Urological Association. The eighth question about quality of life is scored separately. (From Barry MJ, Fowler FJ Jr, O'Leary MP, et al. The American Urological Association symptom index for benign prostatic hyperplasia: the Measurement Committee of the American Urological Association. *J Urol.* 1992;148:1549.)

Abdominal examination should be performed to identify the presence of a palpable bladder, which could be a sign of urinary retention. The physical examination should include a prostate examination to evaluate its size and the possible presence of nodules. The digital rectal examination gives only an approximate estimate of size because only the posterior half is palpated. A focused neurologic examination is also important to assess a patient's mental status, ambulatory status, lower extremity neuromuscular function, and anal sphincter tone.

Laboratory Findings

Urinalysis should be performed to screen for hematuria and urinary tract infection (Chapter 284). A serum prostate-specific antigen (PSA) level should be measured because it can detect symptomatic prostate cancer (Chapter 201) that may require treatment even in men who would not be candidates for treatment if they were asymptomatic. However, the results must be interpreted with caution because the serum PSA level correlates with prostatic volume even in the absence of cancer.[3] Objective parameters such as maximum urinary flow by uroflowmetry and postvoid residual by ultrasound should also be measured if the diagnosis is in question.[4] Although there is not a direct correlation between lower urinary tract symptoms and objective parameters such as prostate size or flow rate, prostate volume and serum PSA levels generally predict worsening of symptoms.

Other causes of bladder dysfunction that should be considered during the assessment of men presenting with lower urinary tract symptoms include bladder cancer (Chapter 197), diabetes (Chapter 229), urethral strictures, and bladder stones. The absence of hematuria makes bladder cancer very unlikely, but patients with BPH symptoms require cystoscopy if they also have hematuria. Neurologic disorders including Parkinson disease (Chapter 409) and multiple sclerosis (Chapter 411) may also cause lower urinary tract symptoms in men.

TREATMENT

Rx

The treatment of BPH aims to improve subjective symptoms and quality of life as well as to prevent progression of the disease. Over time, treatment has evolved away from surgical therapy and largely to medical therapy.

Men who are mildly symptomatic, defined as a score of 7 or less on the IPSS questionnaire, can be observed (Fig. 129-3). Lifestyle changes that may improve symptoms include fluid restriction, timed voiding, and double voiding. Patients who have moderate symptoms, defined as 8 to 19 points on

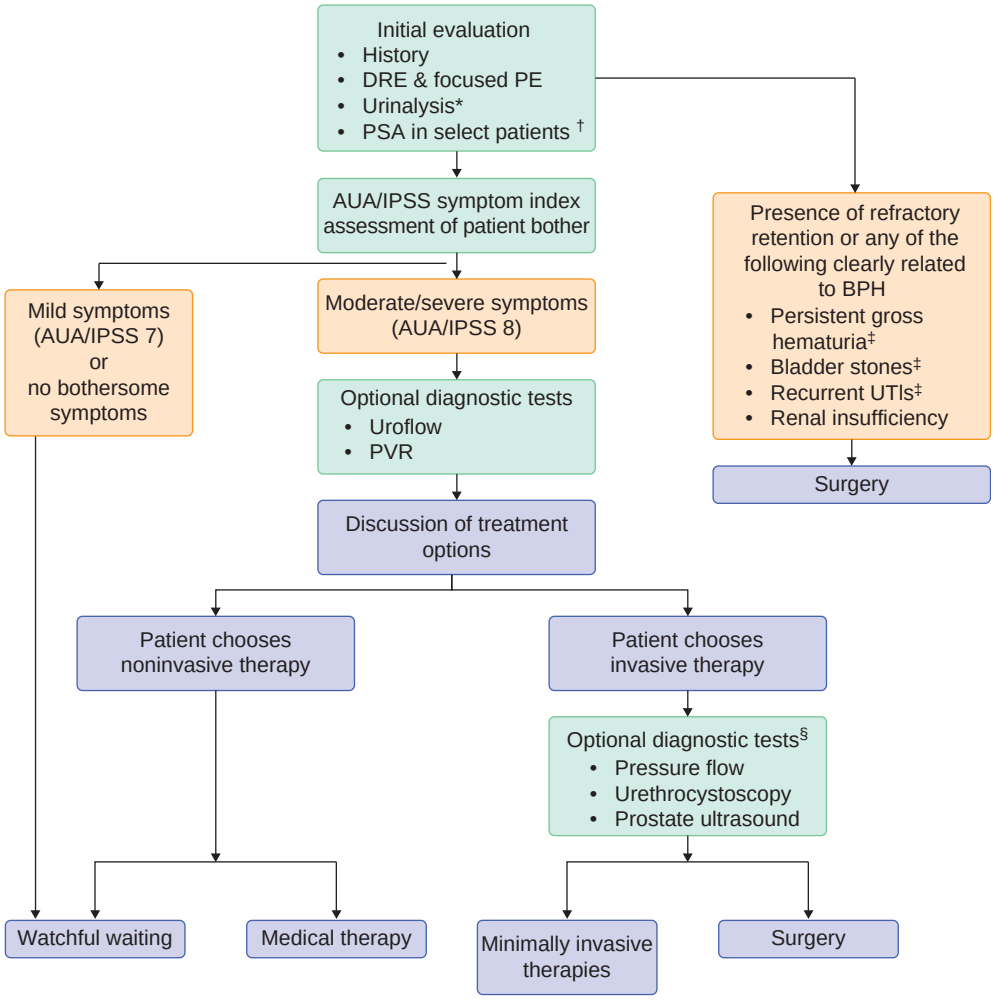

*In patients with clinically significant prostatic bleeding, a course of a 5 alpha-reductase inhibitor may be used. If bleeding persists, tissue ablative surgery is indicated.
†Patients with at least a 10-year life expectancy for whom knowledge of the presence of prostate cancer would change management or patients for whom the PSA measurement may change the management of voiding symptoms.
‡After exhausting other therapeutic options.
§Some diagnostic tests are used in predicting response to therapy. Pressure-flow studies are most useful in men prior to surgery.

AUA, American Urological Association; DRE, digital rectal exam; IPSS, International Prostate Symptom Score; PE, physical exam; PSA, prostate-specific antigen; PVR, postvoid residual urine; UTI, urinary tract infection.

FIGURE 129-3. Algorithm of management of benign prostatic hyperplasia (BPH). (©2003 American Urological Association Educations and Research, Inc.)

TABLE 129-1 MEDICATIONS AND RECOMMENDED DAILY DOSES FOR MALE LOWER URINARY TRACT SYMPTOMS

α-BLOCKERS	5α-REDUCTASE INHIBITORS	α-BLOCKER AND 5α-REDUCTASE INHIBITOR	ANTICHOLINERGICS	ANTICHOLINERGIC PATCHES	PHOSPHODIESTERASE TYPE 5 INHIBITORS
Alfuzosin 10 mg	Dutasteride 0.5 mg	Dutasteride and tamsulosin 0.5/0.4 mg	Darifenacin 7.5, 15 mg	Oxybutynin transdermal 3.9 mg	Tadalafil 2.5, 5, 10, 20 mg
Doxazosin 1-8 mg	Finasteride 5 mg		Fesoterodine 4, 8 mg		Sildenafil 20, 50, 100 mg
Tamsulosin 0.4 mg			Oxybutynin 5, 10, 15 mg		Vardenafil 10, 20 mg
Terazosin 1-10 mg			Oxybutynin XL 5, 10, 15 ER mg		
Silodosin 4, 8 mg			Tolterodine 1, 2 mg Tolterodine LA 2, 4 ER mg Trospium 20, 60 mg Solifenacin 5, 10 mg		

ER = extended-release.

the IPSS, and who are not seriously bothered by their symptoms may also be observed with periodic reassessment to monitor whether urinary retention, refractory hematuria, infections, or other complications develop.

Medications

α-Adrenergic antagonists, which decrease bladder outlet resistance by relaxing urethral smooth muscle and possibly striated sphincter tone, are considered the first-line treatment of male lower urinary tract symptoms (Table 129-1). Their side effects include orthostatic hypotension and dizziness. All medications in this class should be discontinued before cataract surgery for fear of floppy-iris syndrome.

The 5α-reductase inhibitors finasteride and dutasteride are also effective for the treatment of BPH. These agents reduce prostate size by suppressing testosterone and dihydrotestosterone production. As a result, these agents reduce the number of episodes of acute urinary retention and decrease the need for surgical treatment of BPH, and they are particularly helpful for the treatment of patients with higher prostate volumes. Side effects of these medications include erectile dysfunction, reduced libido, and decreased ejaculate volume.

Phosphodiesterase type 5 inhibitors (tadalafil, sildenafil, vardenafil) also improve lower urinary tract symptoms and IPSS scores secondary to BPH, although these agents do not similarly improve maximal flow, and their long-term efficacy is not as well studied as that of α-adrenergic antagonists. Currently, only tadalafil is approved for the treatment of BPH.

Antimuscarinic agents, which target the muscarinic cholinergic receptors in the bladder to reduce overactivity that occurs as a result of changes in detrusor function, are safe and efficacious for the treatment of BPH. Options include tolterodine, solifenacin, and fesoterodine. However, men with bladder outlet obstruction due to BPH often have overactive bladder detrusor function, and antimuscarinic agents block this hyperstimulation and could theoretically decrease detrusor contractility in the setting of bladder outlet obstruction, thereby resulting in urinary retention.

Combination drug therapy is an option when single-drug therapy is insufficient. Data suggest that the best combination is adding a 5α-reductase inhibitor to an α-adrenergic blocker. Saw palmetto, the most commonly used phytotherapeutic agent, has not improved urinary parameters during a treatment period of 12 months in randomized trials and is not recommended even as add-on therapy.

If oral medications are insufficient to control symptoms, another option is onabotulinumtoxinA (Botox). Injection of onabotulinumtoxinA into the prostate, through either a transrectal or transperineal route, improves IPSS scores and maximum flow rates. However, its role in the treatment of lower urinary tract symptoms secondary to BPH remains to be elucidated.

Surgical Treatments

A variety of minimally invasive treatment options are available for BPH.[5] Transurethral needle ablation uses interstitial radiofrequency needles to necrotize tissue. Various laser options can ablate, coagulate, resect, enucleate, or vaporize tissue. Data suggest that these techniques provide results that are similar to open transurethral resection of the prostate (TURP), but many of the reports of their use are from single-site trials, and their efficacy requires further study. Transurethral microwave therapy, during which the prostate is heated by a microwave antenna mounted on a urethral catheter, is associated with lower risks for retrograde ejaculation, treatment for strictures, hematuria, and blood transfusions but also with increased risks for dysuria, urinary retention, and need for retreatment compared with TURP.

Electrosurgically based TURP is the "gold standard" in endoscopic treatment of symptomatic BPH. During the procedure, an antenna is inserted into the urethra and microwaves are emitted. This heat energy destroys the enlarged prostate without damaging surrounding tissue. Potential complications include urinary retention, infection, incontinence, and urethral stricture. Because of improvements in medical therapy and minimally invasive options, the number of TURP procedures performed in the United States has declined, although it still remains a commonly performed urologic procedure.

Bipolar transurethral resection of the prostate is another alternative. With use of the same equipment as for standard TURP, it resects large amounts of prostatic tissue from the transitional zone and central zone but decreases complications, such as perioperative bleeding requiring transfusion. As a result, it can produce results comparable to those of regular TURP with fewer complications. Long-term studies will be needed to confirm the durability of its results in these early trends. For patients with very large prostate glands (i.e., 80 g and larger), TURP procedures may require prolonged operative times, and an open surgical approach may be necessary for adequate debulking of the obstructing prostatic tissue.

PROGNOSIS

Over time, the symptoms of BPH often increase, requiring medications or a procedure. Patients should be counseled about the likelihood of progression, the natural history of lower urinary tract symptoms related to BPH, and the treatment options that can be offered.

PROSTATITIS

EPIDEMIOLOGY AND PATHOBIOLOGY

Prostatitis is usually a clinical diagnosis based on signs and symptoms that occur as a result of inflammation of the prostate gland. The overall prevalence of prostatitis is approximately 8%, and it affects men of a wide age range. Swelling or inflammation of the prostate gland, which may be due to various causes, can have a significant impact on quality of life. The current classification system defines the various types of prostatitis on the basis of whether it is acute or chronic, associated with infection, or associated with pelvic pain (Table 129-2).

The causative organisms of acute bacterial prostatitis are usually similar to those that cause other common genitourinary infections (Chapter 284) and include *Escherichia coli* and *Enterococcus* spp. About 60% of patients with chronic bacterial prostatitis have evidence of ongoing infection based on polymerase chain reaction (PCR) testing of their expressed prostatic secretions, with *Chlamydia trachomatis*, *Ureaplasma urealyticum*, *Mycoplasma genitalium*, and *Mycoplasma hominis* the most common organisms detected.[6] The pathogenesis of chronic prostatitis/chronic pelvic pain syndrome, however, remains unclear.

TABLE 129-2 CLASSIFICATION OF PROSTATITIS

	CATEGORIES			
	I	II	III	IV
Term	Acute bacterial prostatitis	Chronic bacterial prostatitis	Chronic prostatitis/chronic pelvic pain syndrome (CP/CPPS) IIIa. Inflammatory CPPS IIIb. Noninflammatory CPPS	Asymptomatic inflammatory prostatitis
Characteristics	Acute infection of prostate gland	Recurrent or relapsing infection caused by the same organism; not acute	90% of cases of chronic prostatitis are presumed to be nonbacterial; diagnosis of exclusion Characteristic symptoms Discomfort or pain in pelvic region for more than 3 months within the past 6 months Pelvic, perineal, penile, or ejaculatory pain; irritative or obstructive voiding symptoms, sexual dysfunction No documented recurrent urinary tract infections; repeated negative cultures Classification into IIIa or IIIb determined by presence of leukocytes in semen, post–prostate massage urine, or prostatic secretion	No symptoms of prostatitis Leukocytes or inflammatory cells present in prostate tissue, semen, or expressed prostatic secretions

CLINICAL MANIFESTATIONS

Acute bacterial prostatitis (type I) is characterized by acute infection of the prostate gland. Presenting symptoms include pelvic, perineal, penile, or ejaculatory pain as well as irritative or obstructive voiding symptoms and sexual dysfunction. Patients with severe infections can present with fever and chills and can become septic (Chapter 108). Type II prostatitis is characterized by recurrent or relapsing infection caused by the same organism. These patients tend to be less sick during each episode and usually present with voiding symptoms and pain.

Type III prostatitis is characterized by pelvic discomfort or pain for more than 3 of the 6 months before evaluation. Type III patients have repeatedly negative urine cultures. Classification into type IIIa or IIIb is contingent on the presence of leukocytes in semen, post–prostate massage urine, or prostatic secretions. Patients with type IV prostatitis do not experience any symptoms of prostatitis. Leukocytes or inflammatory cells are found in prostate tissue, semen, or expressed prostatic secretions.

DIAGNOSIS

The diagnosis of prostatitis requires a careful history, physical examination, and examination of the urine. The physician must ask pertinent questions about voiding history, sexual history, symptoms, pain, neurologic disorders, and prior pelvic surgery. The National Institutes of Health Chronic Prostatitis Symptom Index (NIH-CPSI), which is a standardized tool for the evaluation and assessment of prostatitis, consists of nine parts that outline three major areas of prostatitis: pain, urinary symptoms, and quality of life (Fig. 129-4). The NIH-CPSI is not specific for making the diagnosis of prostatitis, but it is very useful for longitudinally monitoring changes in symptoms over time after a diagnosis of prostatitis has been established.[7] The physical examination should include an abdominal, external genital, perineal, and digital rectal examination. Attention should be placed on identifying pelvic wall discomfort, structural abnormalities, or prostatic pain on digital rectal examination.

Laboratory Evaluation

Urinalysis and urine culture should be performed for every patient. Historically, the four-specimen test was recommended, with specimens obtained from the initial voided bladder urine, midstream voided bladder urine, expressed prostatic secretions obtained during prostate massage, and voided bladder urine collected after prostate massage. Leukocytes in the third specimen suggest the diagnosis of prostatitis. Leukocytes without bacteria suggest inflammation consistent with nonbacterial prostatitis. However, this four-step approach is rarely used today because it has not proved to be useful as a diagnostic tool or for directing treatment. It has been replaced by either a semen culture or a midstream urine culture and by examination of a voided urine specimen after prostate massage.

Urodynamic methods offer valuable insight for patients experiencing predominantly voiding symptoms. Other conditions, such as prostatic obstruction, primary bladder neck obstruction, dysfunctional voiding, urethral obstruction, and detrusor-sphincter dyssynergia, may be defined with the help of postvoid residual, pressure-flow urodynamics, or videourodynamics. These conditions, in comparison to prostatitis, have many effective treatment options.

Measurement of a postvoid residual urine volume by ultrasound can be used to assess incomplete emptying because urinary retention can be a risk factor for recurrent infections. Low maximum urine flow rates suggest bladder outlet obstruction, decreased detrusor contractility, or both as an alternative explanation for lower urinary tract symptoms and poor flow. However, for those who fail to respond to treatment, PCR analysis of semen to evaluate for possible fastidious organisms (e.g., *C. trachomatis*, *U. urealyticum*, *Mycoplasma* species, and *Neisseria gonorrhoeae*) may be useful for the evaluation of chronic prostatitis. Cultures of urethral swabs may be used to evaluate patients with potentially undiagnosed sexually transmitted diseases (Chapter 285). Other diagnostic tests include semen analysis and culture, which are useful for patients with complaints of abnormal-smelling semen or infertility.

Cystoscopy is an adjunct to urodynamics in the evaluation of chronic prostatitis/chronic pelvic pain syndrome, especially before any surgical intervention. Cystoscopy is also performed for evaluation of hematuria or abnormal cytology findings because prostate cancer (Chapter 201) or bladder cancer (Chapter 197) can cause symptoms similar to chronic pelvic pain syndrome.

PSA testing should be ordered on the basis of prostate cancer screening (Chapter 201) but has no specific role in evaluation of prostatitis symptoms. Acute prostatitis can increase the serum level of PSA, but it usually returns to normal levels with appropriate antibiotics within 1 to 3 months. PSA can be elevated and even can wax and wane in patients with chronic prostatitis/chronic pelvic pain syndrome. Patients with chronic prostatitis have a less well defined decrease in PSA after a course of antibiotics. Anywhere from one third to two thirds of men undergoing prostate biopsy have chronic inflammation, but the correlation with symptoms of prostatitis is unclear.

Transrectal ultrasound can identify a prostatic abscess and is generally recommended for patients who have recurrent prostatitis or who do not respond to treatment. The potential utility of pelvic computed xerographic scanning and transrectal magnetic resonance imaging is unclear.

TREATMENT

Acute Bacterial Prostatitis

Oral or intravenous antibiotics are usually effective for curing acute prostatitis, and progression to chronic bacterial prostatitis is uncommon. Typical first-course antibiotics include oral fluoroquinolones (e.g., levofloxacin 500 mg once daily or ofloxacin 300 mg twice daily) and sulfonamides (e.g., trimethoprim/sulfamethoxazole 160 mg/800 mg twice daily) for 6 weeks. Patients who are unable to tolerate oral medications and patients with signs of sepsis may require broad-spectrum intravenous antibiotics (e.g., ampicillin 2 g every 6 hours plus gentamicin 1.5 mg/kg every 8 hours until afebrile) followed by 6 weeks of oral therapy as before.[8]

Chronic Bacterial Prostatitis

Antibiotic therapy is the mainstay of treatment for chronic bacterial prostatitis. Antibiotics with good lipid solubility, good enteric bacterial coverage, and a high pK_a have the best prostatic penetration. These antibiotics include quinolones, sulfa-based preparations, macrolides, tetracyclines, and aminoglycosides. The fluoroquinolones (e.g., ciprofloxacin 500 mg twice daily, levofloxacin 500 mg twice daily, or ofloxacin 300 mg twice daily) have equivalent success rates in patients with chronic bacterial prostatitis and are generally the first-line treatment.[■] In cases in which atypical bacteria, such as chlamydia, are suspected to be the cause of chronic bacterial prostatitis, better results may be achieved by macrolide antibiotics, such as azithromycin (500 mg twice daily).

NIH-Chronic Prostatitis Symptom Index (NIH-CPSI)

Pain or Discomfort

1. In the last week, have you experienced any pain or discomfort in the following areas?

	Yes	No
a. Area between rectum and testicles (perineum)	❑1	❑0
b. Testicles	❑1	❑0
c. Tip of the penis (not related to urination)	❑1	❑0
d. Below your waist, in your pubic or bladder area	❑1	❑0

2. In the last week, have you experienced:

	Yes	No
a. Pain or burning during urination?	❑1	❑0
b. Pain or discomfort during or after sexual climax (ejaculation)?	❑1	❑0

3. How often have you had pain or discomfort in any of these areas over the last week?

❑ 0 Never
❑ 1 Rarely
❑ 2 Sometimes
❑ 3 Often
❑ 4 Usually
❑ 5 Always

4. Which number best describes your AVERAGE pain or discomfort on the days that you had it, over the last week?

❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑ ❑
0 1 2 3 4 5 6 7 8 9 10
NO PAIN — PAIN AS BAD AS YOU CAN IMAGINE

Urination

5. How often have you had a sensation of not emptying your bladder completely after you finished urinating, over the last week?

❑ 0 Not at all
❑ 1 Less than 1 time in 5
❑ 2 Less than half the time
❑ 3 About half the time
❑ 4 More than half the time
❑ 5 Almost always

6. How often have you had to urinate again less than two hours after you finished urinating, over the last week?

❑ 0 Not at all
❑ 1 Less than 1 time in 5
❑ 2 Less than half the time
❑ 3 About half the time
❑ 4 More than half the time
❑ 5 Almost always

Impact of Symptoms

7. How much have your symptoms kept you from doing the kinds of things you would usually do, over the last week?

❑ 0 None
❑ 1 Only a little
❑ 2 Some
❑ 3 A lot

8. How much did you think about your symptoms, over the last week?

❑ 0 None
❑ 1 Only a little
❑ 2 Some
❑ 3 A lot

Quality of Life

9. If you were to spend the rest of your life with your symptoms just the way they have been during the last week, how would you feel about that?

❑ 0 Delighted
❑ 1 Pleased
❑ 2 Mostly satisfied
❑ 3 Mixed (about equally satisfied and dissatisfied)
❑ 4 Mostly dissatisfied
❑ 5 Unhappy
❑ 6 Terrible

Scoring the NIH-Chronic Prostatitis Symptom Index Domains

Pain: Total of items 1a, 1b, 1c,1d, 2a, 2b, 3, and 4 = _____

Urinary Symptoms: Total of items 5 and 6 = _____

Quality of Life Impact: Total of items 7, 8, and 9 = _____

FIGURE 129-4. NIH Chronic Prostatitis Symptom Index. (Modified from Litwin MS, McNaughton-Collins M, Fowler FJ Jr, et al. The National Institutes of Health chronic prostatitis symptom index: development and validation of a new outcome measure. Chronic Prostatitis Collaborative Research Network. *J Urol.* 1999;162:369-375.)

Most studies demonstrate effective treatment with 30 days of therapy, but some clinicians prescribe 6 weeks of therapy as for acute prostatitis because the recurrence rate is as high as 40% within a year. Delivery of antibiotics by intraprostatic injection or anal submucosal injection is rarely used today.

Treatments for Chronic Prostatitis/ Chronic Pelvic Pain Syndrome

The optimal regimen for the treatment of chronic prostatitis/chronic pelvic pain syndrome is not known, and the response to treatment is often disappointing.[9] Prolonged courses of antibiotics are not generally effective. One option is to perform PCR testing of expressed prostatic secretions and to use antibiotics only if the test result is positive. Current empirical therapy uses a combination of α-blockers, adrenergic antagonists, and anti-inflammatory agents (e.g., ibuprofen 400 mg three times daily or naproxen 200 mg twice daily). Neuroleptics, such as gabapentin, have been tried but appear to be no better than placebo and have side effects. The potential value of allopurinol is unclear. Many phytotherapies have been tried, but improvements have not been dramatic or consistent in various trials.

Conservative supportive therapies include warm sitz baths and special diets that avoid spicy foods, caffeine, alcohol, and other urinary irritants. Behavioral

therapies and stress reduction have also been used. Therapies that aim to improve relaxation and to reeducate pelvic floor muscle function can improve symptoms in highly stressed individuals with dysfunctional voiding. Options include biofeedback and bladder retraining. Unproven treatments include trigger point massage combined with relaxation and electromagnetic pelvic floor therapy, acupuncture, and percutaneous tibial nerve stimulation. Prostate massage can be combined with other therapies, but its efficacy has been variable.

PROGNOSIS

Most patients with type I prostatitis are effectively treated with oral or intravenous antibiotics, although some cases do not respond to treatment and progress to type II prostatitis. The natural history of type II, type III, and type IV prostatitis remains undefined.

Grade A References

A1. McVary KT, Roehrborn CG, Avins AL, et al. Update on AUA guideline on the management of benign prostatic hyperplasia. *J Urol*. 2011;185:1793-1803.

A2. Oelke M, Bachmann A, Descazeaud A, et al. EAU guidelines on the treatment and follow-up of non-neurogenic male lower urinary tract symptoms including benign prostatic obstruction. *Eur Urol*. 2013;64:118-140.

A3. Park T, Choi JY. Efficacy and safety of dutasteride for the treatment of symptomatic benign prostatic hyperplasia (BPH): a systematic review and meta-analysis. *World J Urol*. 2014;32:1093-1105.

A4. Kaplan S, McConnell J. Combination therapy with doxazosin and finasteride for benign prostatic hyperplasia in patients with lower urinary tract symptoms and a baseline total prostate volume of 25 mL or greater. *J Urol*. 2006;175:217-220.

A5. Gacci M, Corona G, Salvi M, et al. A systematic review and meta-analysis on the use of phosphodiesterase 5 inhibitors alone or in combination with alpha-blockers for lower urinary tract symptoms due to benign prostatic hyperplasia. *Eur Urol*. 2012;61:994-1003.

A6. Abrams P, Kaplan S, De Koning Gans HJ, et al. Safety and tolerability of tolterodine for the treatment of overactive bladder in men with bladder outlet obstruction. *J Urol*. 2006;175:999-1004.

A7. Fullhase C, Chapple C, Cornu JN, et al. Systematic review of combination drug therapy for non-neurogenic male lower urinary tract symptoms. *Eur Urol*. 2013;64:228-243.

A8. Barry MJ, Meleth S, Lee JY, et al. Effect of increasing doses of saw palmetto extract on lower urinary tract symptoms: a randomized trial. *JAMA*. 2011;306:1344-1351.

A9. Marberger M, Chartier-Kastler E, Egerdie B, et al. A randomized double-blind placebo-controlled phase 2 dose-ranging study of onabotulinumtoxinA in men with benign prostatic hyperplasia. *Eur Urol*. 2013;63:496-503.

A10. Lee SW, Choi JB, Lee KS, et al. Transurethral procedures for lower urinary tract symptoms resulting from benign prostatic enlargement: a quality and meta-analysis. *Int Neurourol J*. 2013;17:59-66.

A11. Hoffman RM, Monga M, Elliott SP, et al. Microwave thermotherapy for benign prostatic hyperplasia. *Cochrane Database Syst Rev*. 2012;9:CD004135.

A12. Perletti G, Marras E, Wagenlehner FM, et al. Antimicrobial therapy for chronic bacterial prostatitis. *Cochrane Database Syst Rev*. 2013;8:CD009071.

A13. Zhu Y, Wang C, Pang X, et al. Antibiotics are not beneficial in the management of category III prostatitis: a meta analysis. *Urol J*. 2014;11:1377-1385.

A14. Anothaisintawee T, Attia J, Nickel JC, et al. Management of chronic prostatitis/chronic pelvic pain syndrome: a systematic review and network meta-analysis. *JAMA*. 2011;305:78-86.

A15. Aboumarzouk OM, Nelson RL. Pregabalin for chronic prostatitis. *Cochrane Database Syst Rev*. 2012;8:CD009063.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Sarma AV, Wei JT. Clinical practice. Benign prostatic hyperplasia and lower urinary tract symptoms. *N Engl J Med*. 2012;367:248-257.
2. Abrams P, Chapple C, Khoury S, et al. Evaluation and treatment of lower urinary tract symptoms in older men. *J Urol*. 2013;189:S93-S101.
3. Park DS, Oh JJ, Hong JY, et al. Serum prostate-specific antigen as a predictor of prostate volume and lower urinary tract symptoms in a community-based cohort: a large-scale Korean screening study. *Asian J Androl*. 2013;15:249-253.
4. D'Silva KA, Dahm P, Wong CL. Does this man with lower urinary tract symptoms have bladder outlet obstruction?: The Rational Clinical Examination: a systematic review. *JAMA*. 2014;312:535-542.
5. Bhojani N, Gandaglia G, Sood A, et al. Morbidity and mortality after benign prostatic hyperplasia surgery: data from the American College of Surgeons National Surgical Quality Improvement Program. *J Endourol*. 2014;28:831-840.
6. Choi YS, Kim KS, Choi SW, et al. Microbiological etiology of bacterial prostatitis in general hospital and primary care clinic in Korea. *Prostate Int*. 2013;1:133-138.
7. Clemens JQ, Calhoun EA, Litwin MS, et al. Rescoring the NIH chronic prostatitis symptom index: nothing new. *Prostate Cancer Prostatic Dis*. 2009;12:285-287.
8. Schiller DS, Parikh A. Identification, pharmacologic considerations, and management of prostatitis. *Am J Geriatr Pharmacother*. 2011;9:37-48.
9. Cohen JM, Fagin AP, Hariton E, et al. Therapeutic intervention for chronic prostatitis/chronic pelvic pain syndrome (CP/CPPS): a systematic review and meta-analysis. *PLoS ONE*. 2012;7:e41941.

REVIEW QUESTIONS

1. A 50-year-old man presents with lower urinary tract symptoms. Digital rectal examination reveals an enlarged prostate complicated by lower urinary tract symptoms. In discussing the potential side effects of dutasteride with the patient, which of the following would *not* apply in regard to sexual function?

A. Decreased libido
B. Decreased ejaculatory volume
C. Erectile dysfunction
D. Ejaculatory pain
E. A and C

Answer: D 5α-Reductase inhibitors, such as dutasteride, were developed to block the conversion of testosterone to dihydrotestosterone, thus reducing prostate volume and thereby decreasing bladder outlet obstruction. The side effects of a 5α-reductase inhibitor include decreased libido, decreased ejaculatory volume, and erectile dysfunction.

2. A 60-year-old man has significant lower urinary tract symptoms refractory to other therapies. He does not have any major morbid conditions that would be contraindications to surgery. Which of the following is not an absolute indication for surgery?

A. The patient has developed acute urinary retention.
B. The patient has renal insufficiency secondary to benign prostatic hyperplasia.
C. The patient has recurrent gross hematuria due to benign prostatic hyperplasia.
D. Cystoscopy revealed a bladder diverticulum.
E. B and D

Answer: D Surgical intervention is an appropriate treatment alternative for patients with moderate to severe lower urinary tract symptoms and for patients who have developed acute urinary retention or other benign prostatic hyperplasia–related complications. Surgery is recommended for patients who have renal insufficiency secondary to benign prostatic hyperplasia; for patients who have recurrent urinary tract infections, bladder stones, or gross hematuria due to benign prostatic hyperplasia; and for those who have lower urinary tract symptoms refractory to other therapies. The presence of a bladder diverticulum is not an absolute indication for surgery unless it is associated with recurrent urinary tract infection or progressive bladder dysfunction.

3. A 35-year-old man presents with the following symptoms indicative of chronic prostatitis/chronic pelvic pain syndrome: frequent urination, increased urgency, and post-ejaculatory pain for the past 4 months. In this case, which of the following evaluations would not be mandatory?

A. Abdominal examination
B. Digital rectal examination
C. Semen analysis
D. Urinalysis
E. Urine culture

Answer: C The physician must ask pertinent questions about voiding history, sexual history, characterization of symptoms, and pain. A thorough medical and surgical history should focus on neurologic disorders and pelvic surgery. Physical examination must include abdominal, external genital, perineal, and digital rectal examination. Attention should be placed on identifying pelvic wall discomfort, structural abnormalities, or prostatic pain on digital rectal examination. Urinalysis and urine culture should be performed for every patient.

4. A 45-year-old man has a history of recurrent urinary tract infections. Bacterial growth on culture of expressed prostatic fluid and a post–prostate massage urine specimen confirms a diagnosis of chronic bacterial prostatitis. The expressed prostatic secretion contains more than 10 white blood cells per high-power field. Digital rectal examination demonstrates a normal benign prostate. Which antibiotic class is usually recommended for treatment of his prostatitis?

A. Fluoroquinolones
B. Macrolides
C. Trimethoprim/sulfamethoxazole
D. Doxycycline
E. None of the above

Answer: A Fluoroquinolones have been recommended as first-line therapy for chronic bacterial prostatitis. With high lipid solubility, they show the best penetration into the prostate and seminal fluid. The usual course is 4 to 6 weeks.

5. After 48 hours of therapy with appropriate intravenous broad-spectrum antibiotics for his prostatitis, a 40-year-old man continues to be febrile and highly symptomatic with chills and malaise. Which of the following is the "gold standard" imaging modality for a patient with a suspected prostatic abscess?

A. Computed tomography scan
B. Magnetic resonance imaging scan
C. Abdominal plain film and intravenous pyelography
D. Transrectal ultrasound
E. There is no gold standard imaging modality for a patient with a suspected prostatic abscess.

Answer: D Transrectal ultrasound is a useful tool in identifying a prostatic abscess and is considered the gold standard. This modality can also identify seminal vesicle or ejaculatory duct abnormalities, especially in patients with ejaculatory pain or hematospermia.

TABLE 130-1 CAUSES OF CHRONIC RENAL FAILURE

Diabetic glomerulosclerosis*
Hypertensive nephrosclerosis
Glomerular disease
Glomerulonephritis
Amyloidosis, light chain disease*
Systemic lupus erythematosus, granulomatosis with polyangiitis
Tubulointerstitial disease
Reflux nephropathy (chronic pyelonephritis)
Analgesic nephropathy
Obstructive nephropathy (stones, benign prostatic hypertrophy)
Myeloma kidney*
Vascular disease
Scleroderma*
Vasculitis*
Renovascular renal failure (ischemic nephropathy)
Atheroembolic renal disease*
Cystic disease
Autosomal dominant polycystic kidney disease
Medullary cystic kidney disease

*Systemic disease involving the kidney.

130 CHRONIC KIDNEY DISEASE

WILLIAM E. MITCH

DEFINITION

Chronic kidney disease (CKD) refers to the many clinical abnormalities that progressively worsen as kidney function declines. CKD results from a large number of systemic diseases that damage the kidney or from disorders that are intrinsic to the kidney (Table 130-1). The severity of CKD is graded by the depressed level of the glomerular filtration rate (GFR); a GFR persistently below 60 mL/min/1.73 m^2 serves to identify patients who will most likely develop clinical manifestations as a result of progressive loss of kidney function. Measurement of GFR is cumbersome, so estimates of GFR (eGFR) based on a patient's serum creatinine and demographics (Chapter 114) are used to monitor the course of CKD.

Unlike acute kidney damage (Chapter 120), which can be repaired with a resulting improvement in kidney function, the kidney damage in CKD is rarely repaired, so loss of function persists. In most CKD patients, in fact, loss of kidney function generates more kidney damage, so CKD can progressively worsen even if the disorder that initially caused it becomes inactive.

CKD includes a spectrum of clinical dysfunctions that range from abnormalities detectable only by laboratory testing to a syndrome known as uremia. *Uremia*, which literally means "urine in the blood," results from the accumulation of unexcreted ions and waste products and the metabolic abnormalities they induce. When the kidney fails to perform most of its functions, the clinical state is called *end-stage renal disease* (ESRD), and dialysis or transplantation is required to sustain life (Chapter 131). Before this stage, treatment strategies are directed at slowing the loss of kidney function, postponing the onset of ESRD, and ameliorating the symptoms of uremia.

EPIDEMIOLOGY

The most widely used method for identifying degrees of CKD is the 2009 Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation, which has been modified to include information about creatinine or cystatin C to ascertain the risks for development of complications of CKD (see Table 114-1).[1] On the basis of these equations plus demographic and medical information from the National Health and Nutrition Examination Survey, from the National Institutes of Health United States Renal Data System,[2] and from the Kidney Disease Improving Global Outcomes work group,[3] an estimated 23.1 million Americans, representing 11.5% of noninstitutionalized adults older than 20 years, have evidence of CKD (Table 130-2). The largest group of CKD patients has an eGFR below 60 mL/min/1.73 m^2 (i.e., stages 3 to 5). These individuals are at higher risk for having their GFR decline below a threshold that will precipitate progressive CKD and its complications (e.g., hypertension, anemia, hyperphosphatemia, and acidosis). The overall prevalence of adults with stage 3 CKD is increasing, probably related to the aging of the population as well as the increasing prevalence of obesity and type 2 diabetes.[4] Fortunately, data suggest that an individual's risk for development of CKD has not been rising since 2001. Nevertheless, the number of adults with ESRD is estimated to be about 615,000, mainly because increasing numbers of CKD patients are 70 years of age and older.

Besides the elderly, CKD occurs more widely in African Americans, Asian Americans, Hispanics, and Native Americans, including Hawaiians and Pacific Islanders. Two disorders account for more than 70% of all adult CKD patients in the United States: 44% have diabetes mellitus (Chapter 229), and 28% have hypertension. The frequency of CKD is also increased in patients with albuminuria and in patients with a family member with CKD. Data on the genetic epidemiology of CKD are limited, but African Americans with two copies of high-risk variants of the *APO1* gene encoding apolipoprotein-1 have about a 2-fold higher probability of suffering rapid decline in eGFR and

TABLE 130-2 PREVALENCE OF STAGES OF CHRONIC KIDNEY DISEASE AND FREQUENCY OF COMPLICATIONS

STAGE	DESCRIPTION	GFR* (mL/min/1.73 m^2)	ADULT PREVALENCE (MILLIONS)	SYMPTOMS OR SIGNS
1	Chronic kidney damage; normal or increased GFR	>90	4.6	Anemia 4% Hypertension 40% 5-year mortality 19%
2	Mild GFR loss	60-89	5.0	Anemia 4% Hypertension 40% 5-year mortality 19%
3	Moderate GFR loss	30-59	12.5	Anemia 7% Hypertension 55% 5-year mortality 24%
4	Severe GFR loss	15-29	0.8	Hyperphosphatemia 20% Anemia 29% Hypertension 77% 5-year mortality 46%
5	Kidney failure	<15 or dialysis	0.2	Hyperphosphatemia 50% Anemia 69% Hypertension >75% 3-year mortality 14%

*The formula for estimating the glomerular filtration rate (GFR) of adults with chronic kidney disease (CKD) is derived from data obtained during the National Health and Nutrition Examination Survey (NHANES 2001-2008). The 2009 Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation is

$$eGFR = 141 \times \min(SCr/\kappa, 1)^{\alpha} \times \max(SCr/\kappa, 1)^{-1.209} \times 0.993^{Age} \; [\times 1.018 \text{ if female}] [\times 1.159 \text{ if black}]$$

where SCr is serum creatinine (in mg/dL), κ is 0.7 for females and 0.9 for males, α is −0.329 for females and −0.411 for males, min is the minimum of SCr/κ or 1, and max is the maximum of SCr/κ or 1.

TABLE 130-3 FUNCTIONS OF THE KIDNEY AND IMPAIRMENT OF KIDNEY FUNCTION IN PATIENTS WITH CHRONIC KIDNEY DISEASE

KIDNEY FUNCTION	CONSEQUENCES OF DYSFUNCTION
Maintain concentration and body content of electrolytes and fluid volume	Hyponatremia, hyperkalemia, low total potassium content, hypocalcemia, hyperphosphatemia, decreased tolerance to electrolyte or mineral loading
Regulate blood pressure	Hypertension, cardiovascular disease
Endocrine mediator	Anemia (low erythropoietin), hypertension (renin system activation), bone disease (secondary hyperparathyroidism), low vitamin D activation, prolonged half-lives of peptide hormones (e.g., insulin)
Waste product excretion	Anorexia, nausea, soft tissue deposition of oxalates and phosphates, neurologic dysfunction, loss of muscle protein

adverse renal outcomes.[5] Other epidemiologic factors associated with an increasing risk of progressive CKD are cardiovascular disease, smoking, and hyperlipidemia.

PATHOBIOLOGY

The intact nephron hypothesis helps explain the importance of GFR as a measure of remaining kidney function. The nephron consists of the glomerulus, proximal tubule, loop of Henle, distal tubule, and collecting duct. Individuals are born with 0.75 million to 1.25 million nephrons per kidney, but if nephrons are lost, new ones are not regenerated. The intact nephron hypothesis is that each nephron functions as an independent unit, so the sum of the functions of all remaining nephrons determines the whole kidney's GFR, the most accurate estimate of remaining kidney function.

Physiologic and metabolic functions of the kidney include the regulation of blood pressure, several endocrine functions, and ion concentrations in the extracellular and intracellular fluids as well as the excretion of waste products (Table 130-3). Loss of these functions yields several direct and derivative consequences of CKD. For example, a limitation in the ability to excrete acid causes hyperventilation and a decrease in P_{CO_2}. In muscle, acidosis activates the ubiquitin-proteasome enzymatic process to degrade protein, causing loss of muscle mass. Bone buffers acid by releasing calcium and phosphates, a response that leads to demineralization and secondary hyperparathyroidism, both of which make bones more susceptible to fracture.

Balance and Steady-State Considerations

Metabolic balance is the state in which the intake or production of a substance equals its elimination. In response to CKD, the ability to excrete sodium falls as nephrons are lost, but the remaining nephrons respond at least partially by excreting a greater fraction of the sodium filtered by each glomerulus. Similar phenomena adjust the excretion of other ions and substances, thereby allowing the patient with CKD to reduce the accumulation of ions and to avoid adverse consequences such as hyperkalemia. The ability to achieve balance between intake and excretion, however, has a limit; if the sodium balance is positive because the intake of sodium exceeds its excretion by the kidney, hypertension and edema will develop.

A related concept is that of steady state. A patient is in steady state when the internal environment is constant and the intake and production of an ion or compound equal its output and metabolism. Although a constant weight indicates that sodium intake is equal to sodium output, this steady state does not necessarily indicate normal conditions. For example, a patient who is grossly edematous may be in the steady state because the intake of sodium equals its excretion but at the price of sodium accumulation in the extracellular fluid.

The Tradeoff Hypothesis

Another important principle, the tradeoff hypothesis, refers to the activation of pathophysiologic responses that produce adverse consequences. In response to CKD, the loss of nephrons will initially reduce salt excretion, thereby leading to sodium retention, expansion of extracellular fluid, and a rise in weight and blood pressure. The pathophysiologic response to sodium retention triggers adaptations that increase sodium excretion by raising blood pressure and suppressing the reabsorption of filtered sodium. The tradeoffs include the development of volume-dependent hypertension and the persistence of impaired reabsorption of filtered sodium by the remaining tubules. An abrupt decrease in salt intake will elicit only a sluggish increase in sodium reabsorption, and the result will be a loss of sodium and a reduction in extracellular and intravascular volume, which will impair kidney perfusion and decrease GFR.

The most extensively studied tradeoff is the adaptation that stimulates secondary hyperparathyroidism (Fig. 130-1). In CKD, the loss of nephrons impairs the kidney's ability to excrete phosphates, which accumulate and result in an increased formation of calcium-phosphate complexes. The resulting reduction in the level of ionized calcium stimulates calcium-sensing receptors in the parathyroid gland. These responses stimulate the production and secretion of parathyroid hormone (PTH). The increase in PTH is beneficial because it suppresses Na/Pi type II, the cotransporter of phosphates, thereby reducing phosphate reabsorption by the kidney and promoting the excretion of the accumulated phosphates. Besides stimulating PTH

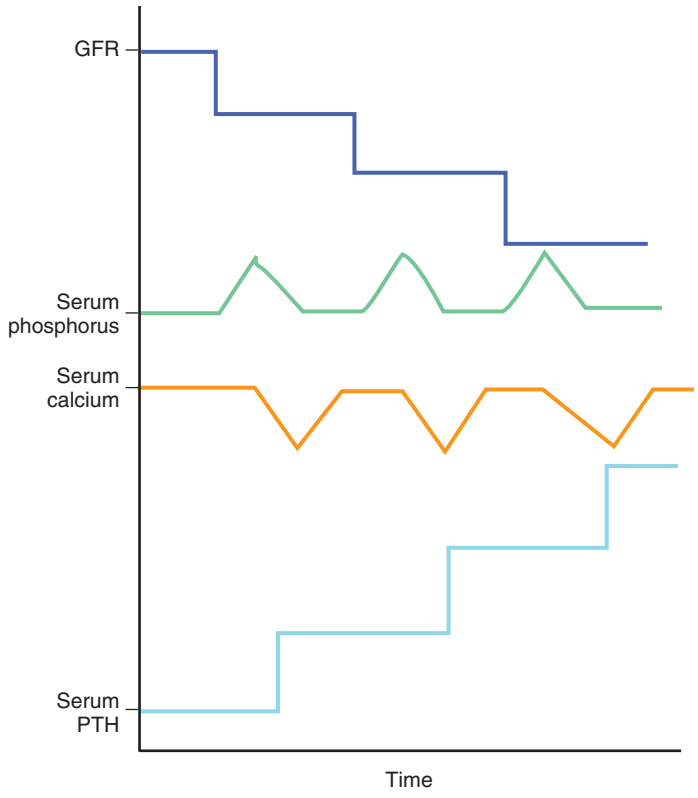

FIGURE 130-1. A decrease in glomerular filtration rate (GFR) is followed by an increase in serum phosphorus and a decrease in serum calcium. An increase in serum parathyroid hormone (PTH) returns phosphorus and calcium to normal levels, but the tradeoff is PTH-induced bone disease.

secretion, an increase in the levels of circulating phosphates will suppress the production of 1,25-dihydroxycholecalciferol (calcitriol), which is the most potent form of vitamin D. Finally, the increase in circulating phosphates stimulates bone osteoclasts to secrete fibroblast growth factor 23 (FGF23). FGF23 can interact with the cofactor Klotho in the proximal tubule, thereby suppressing the Na/Pi type II phosphate transporter. The result is that the increase in PTH and FGF23 stimulates urinary phosphate excretion. The tradeoff for eliminating phosphates is the development of renal osteodystrophy because the steady-state increase in ionized calcium and reduction in circulating phosphates can be maintained only if the circulating concentrations of PTH and FGF23 are increased to cause phosphate excretion (see Fig. 130-1). The adverse consequences of these responses include PTH-mediated stimulation of osteoclasts and an increase in mortality associated with a high FGF23 level (see later).

Hypertension

Hypertension, like anemia, is almost universal in CKD patients and is often the first clinical indication of CKD. The coincidence of CKD and high blood pressure is particularly important because hypertension contributes to the development of cardiovascular disease, which is the leading cause of morbidity and mortality in CKD patients. Hypertension in CKD patients largely reflects an expanded extracellular volume due to a salt-rich diet plus impaired capacity to excrete sodium; activation of the renin-angiotensin-aldosterone system also plays a role. In terms of a tradeoff, when sodium retention raises extracellular volume, blood pressure and sodium excretion increase, thereby contributing to a balance between sodium intake and its excretion; salt balance can be maintained only as long as blood pressure is high. Two practical implications arise from these relationships. First, treatment of hypertensive patients with vasodilating drugs alone is frequently unsuccessful because the initial reduction in blood pressure stimulates sodium retention and expansion of the extracellular volume, which raise blood pressure. Second, a salt-rich diet cancels the benefits of diuretics even in normal adults.

Another mechanism for hypertension in CKD patients is activation of the renin-angiotensin-aldosterone system and the sympathetic nervous system. When inhibitors of the renin-angiotensin-aldosterone system, such as angiotensin-converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARBs), are given to patients with CKD, they can slow the loss of GFR. Evidence for activation of the sympathetic nervous system includes higher circulating levels of norepinephrine, which can contribute to vasoconstriction, and suppressed production of nitric oxide. Inhibition of this system in CKD patients, unlike blocking of the renin-angiotensin-aldosterone system, does not, however, slow the loss of GFR.

Endocrine Disorders

CKD, even in patients with serum creatinine values as low as 2.5 mg/dL, reduces the ability of insulin to stimulate glucose uptake by muscle and other organs, an abnormality known as *insulin resistance* (Chapter 229). The resulting reduction in glucose uptake raises blood glucose levels and causes a compensatory increase in the release of insulin, which acts to maintain blood glucose levels near normal. Insulin resistance in nondiabetic CKD patients is generally associated with blood glucose values within the normal range, and blood glucose levels rarely exceed 200 mg/dL. Insulin resistance also impairs intracellular signaling and interferes with the metabolism of both glucose and protein, thereby causing loss of muscle proteins.

The metabolic acidosis (Chapter 118) of CKD contributes to the development of insulin resistance, impairs the ability of growth hormone to stimulate insulin-like growth factor-I, and depresses circulating levels of thyroxine and triiodothyronine (Chapter 226). Fortunately, most of these metabolic changes, like the acidosis-induced loss of bone density and muscle protein, are reversed simply by treating CKD patients with sodium bicarbonate or other alkalizing agents.

Another mechanism that affects endocrine status in CKD patients is the kidney's impaired ability to degrade small proteins, including several hormones. For example, diabetic patients with CKD can progressively lose the ability to degrade insulin, thereby lengthening its half-life. This response can cause hypoglycemia in patients with progressive CKD who are treated with standard doses of insulin. Second, incomplete degradation of PTH by the damaged kidney can affect the interpretation of the circulating PTH concentration. Fragments of partially degraded PTH are recognized by the PTH immunoassay, thereby leading to a misinterpretation that levels of PTH are excessively high.

In patients with advanced, stage 4 CKD, normochromic, normocytic anemia (Chapter 158) is almost universal, principally because of impaired production of erythropoietin by interstitial cells in the damaged kidney. Other factors contributing to anemia in CKD patients include a shortened half-life of erythrocytes, gastrointestinal bleeding, and deficiencies of vitamins and iron.

Renal Bone Disease

Renal bone disease, also called *renal osteodystrophy*, afflicts virtually all CKD patients to different degrees. Findings on bone biopsies of patients with renal bone disease range from features indicative of increased bone turnover (i.e., greater numbers of osteoclasts, osteoblasts, and osteocytes) to abnormalities reflecting low bone turnover (i.e., reduced numbers of osteoclasts and osteoblasts and the accumulation of demineralized matrix). PTH is a major stimulus for the development of renal osteodystrophy in CKD patients; patients with high bone turnover have increased circulating PTH levels, whereas patients with low bone turnover exhibit only a small increase in circulating PTH. A third type of disease is mixed uremic osteodystrophy, which has features of hyperparathyroidism plus defective mineralization. The pathophysiologic process of this bone disorder involves an increase in circulating phosphates, which increase circulating PTH, which in turn activates osteoclasts to reduce bone mass. Besides stimulating PTH, the increase in circulating phosphates suppresses the activation of calcitriol, and this decrease in calcitriol limits intestinal absorption of calcium and phosphates and also suppresses PTH production. The physicochemical interaction between circulating phosphates and "free" or ionized calcium lowers the level of ionized calcium. Consequently, the binding of ionized calcium to calcium receptors on parathyroid chief cells is blunted, thereby stimulating the production and secretion of PTH. Second, hyperphosphatemia can act directly on parathyroid cells to stimulate PTH production. Other factors that contribute to CKD-induced bone disease include defects in cellular signaling by the calcium receptor and changes in vitamin D metabolism. Hyperphosphatemia also can blunt the ability of ACE inhibitors or ARBs to slow the loss of GFR.

A G protein–coupled plasma membrane receptor present in chief cells of the parathyroid gland and in certain renal tubular cells responds directly to changes in calcium ions. This receptor can then interact with calcium or with cinacalcet, a small, orally available molecule that activates the calcium

receptor, thereby leading to suppression of the expression and a release of PTH from parathyroid chief cells. In contrast, hyperphosphatemia and a reduced level of circulating ionized calcium increase the production and release of PTH (see Fig. 130-1). Stimulation of PTH secretion is negated by cinacalcet even when there is hyperphosphatemia and low ionized calcium levels. Because of the complexity of its action, use of cinacalcet requires careful monitoring to avoid hypoparathyroidism and hypocalcemia.

Another factor that regulates the circulating calcium level and affects the development of renal bone disease is vitamin D. Under normal conditions, activation of vitamin D proceeds by repeated hydroxylation of the parent molecule, cholecalciferol (vitamin D_3). The initial hydroxylation occurs in the liver, where 25-hydroxyvitamin D_3 is formed. This form of vitamin D stimulates the absorption of calcium and phosphates from the intestines and is believed to change the function and metabolism of muscle and possibly other organs by mechanisms that are poorly defined. Moreover, low circulating values of 25-hydroxyvitamin D_3 in CKD patients are associated with an increase in the risk of all-cause mortality.

Calcitriol, the most active form of vitamin D, is produced when 25-hydroxyvitamin D_3 is hydroxylated by 25-hydroxycholecalciferol 1α-hydroxylase in the proximal tubules of the kidney. Activity of the 1α-hydroxylase is regulated by factors that change mineral metabolism; for example, its activity is reduced by hyperphosphatemia, which decreases the production of calcitriol and hence absorption of calcium and phosphates from the intestine. Alternatively, the β-glucuronidase Klotho is a tissue-specific, locally secreted cofactor of FGF23. It increases the excretion of phosphates by the kidney. In addition, the Klotho-FGF23 combination can bind to the FGF receptor to decrease the expression of the 1α-hydroxylase, thereby suppressing the production of calcitriol. Finally, the Klotho-FGF23 combination interacts with its receptor in the parathyroid gland and down-regulates the secretion of PTH. These complex regulatory processes emphasize how loss of kidney function disrupts the normal function and turnover of bone.

Accumulation of Uremic Toxins

When diets contain protein-rich foods, the protein is metabolized to amino acids that can be used to build body protein stores (E-Fig. 130-1). Amino acids not used for this purpose are metabolized to form urea or are converted into potentially toxic products that are accumulated in patients with CKD.[6] Because urea production is directly proportional to the amount of protein eaten, it follows that excess protein in the diet increases the production of urea and uremic toxins. Another source of uremic toxins results when bacteria in the colon metabolize amino acids into uremic toxins (e.g., tryptophan or histidine can be converted into *p*-cresol and indoxyl sulfate).

The concentration of creatinine in serum is determined by the degree of renal insufficiency and the rate of creatinine production, which is proportionate to lean body mass. For example, a serum creatinine concentration of 1.4 mg/dL in an adult with a small muscle mass signifies a much greater loss of kidney function than it does in an individual with a large muscle mass. Creatinine is formed from creatine and creatine phosphate by a nonenzymatic reaction, so creatinine production in subjects with a stable weight should be constant. The production of creatinine is affected by the diet; creatine and creatine phosphate are highly concentrated in muscle, and extensive cooking of meat converts creatine to creatinine. At least 4 months are required to reach a new steady state of the conversion of creatine and creatine phosphate to creatinine when CKD patients change the protein in their diets. Consequently, if serum creatinine rises or falls in response to a treatment affecting kidney function, conclusions that the treatment has slowed the loss of GFR must be delayed until 4 months have passed.

With CKD, the accumulation of peptides (also known as *middle molecules*) has been associated with disorders that range from the induction of anorexia to neurologic abnormalities. Alternatively, high uric acid levels, which are related to excess protein intake, can cause gout (Chapter 273) and may participate in the development of hypertension and inflammatory responses in blood vessels. Another association between protein-rich diets and increased levels of uremic toxins arises because foods rich in protein invariably raise the intake of phosphates, sodium, acid, potassium, and other ions. These ions aggravate phosphate-induced renal osteodystrophy, volume-dependent hypertension, and acidosis-stimulated loss of muscle protein.

Ideally, circulating levels of uremic toxins should be monitored, but such measurements are not practical. Fortunately, the blood urea nitrogen (BUN) provides a readily available index of the level of uremic toxins because the production of urea is directly proportional to the intake of proteins and hence the accumulation of ions and other unexcreted waste products (see E-Fig. 130-1). The amount of protein in the diet can be reliably estimated from the 24-hour excretion of urea nitrogen as long as the patient is in the steady state (i.e., BUN and weight are stable; Table 130-4). When the ratio of BUN to serum creatinine is chronically below 10 : 1, the patient is eating a protein-restricted diet. When the BUN concentration exceeds 10 times the serum creatinine concentration, three possibilities should be considered. First, the patient may have gastrointestinal bleeding or may be suffering from a severely catabolic condition (e.g., trauma or high-dose glucocorticoid administration) that results in catabolism of endogenous proteins to amino acids and hence to urea. Second, the patient may be eating excessive amounts of protein, which yields more urea than the impaired kidney can excrete. Finally, extracellular volume depletion or severe liver or heart disease can stimulate active reabsorption of sodium and fluid by the proximal tubule and thereby increase the passive reabsorption of urea and raise the BUN level. The corollary is that a decrease in urea production is associated with a decrease in the load of uremic toxins. Consequently, the goal of dietary manipulation in CKD is to minimize production of urea while ensuring an adequate intake to maintain body protein stores. To accomplish this goal, the dietary content of protein should be monitored to limit the excessive production of urea (see later).

TABLE 130-4 ESTIMATION OF DIETARY PROTEIN FROM 24-HOUR UREA NITROGEN EXCRETION

ASSUMPTIONS
The patient is in the steady state, and neither the serum urea nitrogen concentration nor body weight is changing; there is no edema.
The patient is in nitrogen balance, so that nitrogen intake equals nitrogen excretion.
Protein is 16% nitrogen.
The nonurea nitrogen excretion (the nitrogen in urinary creatinine, uric acid, and peptides plus feces) is 0.031 g nitrogen/kg/day.
CASE 1
A 50-year-old patient with a stable weight of 70 kg is prescribed a diet containing 0.8 g protein/kg/day. His 24-hour urea nitrogen excretion is 6.8 g nitrogen/day. How much protein is he eating?
His diet should contain 70 kg × 0.8 g protein/kg, or 56 g protein. His intake of nitrogen from this diet is approximately 9 g (56 g protein × 0.16 = 8.96 g nitrogen). His nitrogen excretion is 6.8 g urea nitrogen + 2.17 g nonurea nitrogen/day (70 × 0.031 g nonurea nitrogen/kg/day). The total nitrogen excretion is 8.97 g, so the patient is compliant with the prescribed diet.
CASE 2
A 40-year-old woman weighing 60 kg is confident that she is eating a diet containing 0.6 g protein/kg/day. Her 24-hour urea nitrogen excretion is 10 g nitrogen/kg/day. Does she require additional investigation?
Her diet should contain 60 kg × 0.6 g protein/kg, or 36 g protein. Therefore, her intake of nitrogen is approximately 5.8 g (36 g protein × 0.16 = 5.76 g nitrogen). Her nitrogen excretion is 10 g urea nitrogen + 1.86 g nonurea nitrogen (60 kg × 0.031 g nonurea nitrogen/kg/day). Her total nitrogen excretion is 11.9 g/day, far in excess of the amount of protein she believes she is eating. Consequently, the patient requires investigation for gastrointestinal bleeding.

Progression of Chronic Kidney Disease

Persistence of diseases affecting the kidney (e.g., diabetes or inflammatory conditions such as systemic lupus erythematosus) is not the only factor that determines the rapidity of the loss of kidney function. Even when diseases that initially damage the kidney are no longer active, kidney function continues to decline, perhaps because of systemic hypertension, hemodynamic injury to the kidney, proteinuria, and the accumulation of nephrotoxins.

There is no agreement that treatment of hypertension aggressively will change the course of progressive loss of kidney function. Clinical observations suggest that hypertension adversely affects the kidney; malignant hypertension damages endothelial cells of both the afferent arteriole and the glomerulus and may even cause thrombosis in these vessels. Chronic hypertension is frequently associated with ischemic injury to glomeruli and can result in glomerulosclerosis. Furthermore, the degree of hypertension in CKD patients is directly correlated with the rate of loss of kidney function.

Experimentally, angiotensin II–related, progressive glomerular damage (Table 130-5) arises because it preferentially constricts the glomerular efferent arteriole to a greater extent than the afferent arteriole. This imbalance raises intracapillary pressure and tends to increase glomerular filtration, thereby leading to the hyperfiltration mechanism, but the tradeoff for the

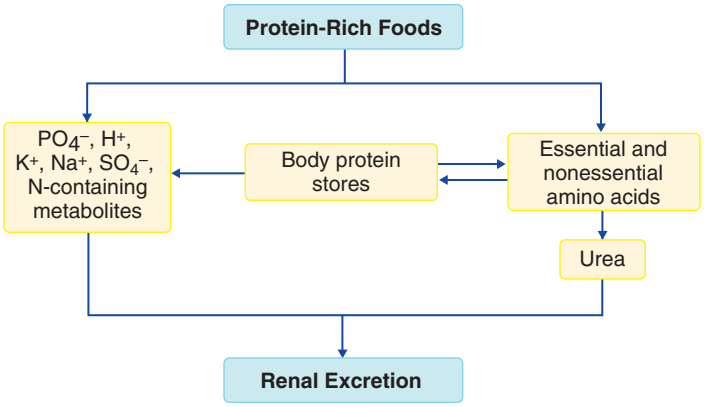

E-FIGURE 130-1. The breakdown of dietary protein enlarges the pool of ions plus the essential and nonessential amino acids, which are converted into urea plus waste products or used to synthesize body protein. Protein breakdown also increases the production of urea and other nitrogenous waste products, which, like inorganic ions, accumulate because they are not excreted by the kidney.

TABLE 130-5 ANGIOTENSIN II RESPONSES IN CHRONIC KIDNEY DISEASE*

Hemodynamic responses
Systemic hypertension
Vasoconstriction
Salt retention (aldosterone)
Intraglomerular hypertension
Efferent arteriolar vasoconstriction
Nonhemodynamic responses in the kidney
Macrophage infiltration and inflammation
Interstitial matrix accumulation
Increased transforming growth factor-β
Increased plasminogen activator inhibitor type 1
Increased aldosterone

*Includes the proposed actions of angiotensin II that can contribute to the development of cardiovascular disease and progressive loss of kidney function.

TABLE 130-6 COMPLICATIONS OF CHRONIC KIDNEY DISEASE

AFFECTED SYSTEM	CAUSE OR MECHANISM	CLINICAL SYNDROME
Systemic symptoms	Anemia, inflammation	Fatigue, lassitude
Skin	Hyperparathyroidism, calcium-phosphate deposition	Rash, pruritus, metastatic calcification
Cardiovascular disease	Hypertension, anemia, hyperhomocysteinemia, vascular calcification	Atherosclerosis, heart failure, stroke
Serositis	Unknown	Pericardial or pleural pain and fluid, peritoneal fluid
Gastrointestinal	Unknown	Anorexia, nausea, vomiting, diarrhea, gastrointestinal tract bleeding
Immune system	Leukocyte dysfunction, depressed cellular immunity	Infections
Endocrine	Hypothalamic-pituitary axis dysfunction	Amenorrhea, menorrhagia, impotence, oligospermia, hyperprolactinemia
Neurologic	Unknown	Neuromuscular excitability, cognitive dysfunction progressing to coma, peripheral neuropathy (restless leg syndrome or sensory deficits)

increase in GFR is damage to glomerular capillaries. Because angiotensin II is the mediator of preferential efferent arteriolar constriction, ACE inhibitors and ARBs are used to prevent both hyperfiltration and damage to the kidney.

ACE inhibitors and ARBs not only lower blood pressure but also block growth factor properties of angiotensin II and its ability to activate transforming growth factor-β, plasminogen activator inhibitor type 1, and other cytokines that can contribute to the development of interstitial damage to the kidney (see Table 130-5). Aldosterone may also contribute to the development of interstitial damage and collagen deposition in the kidney. ACE inhibitors or ARBs also generally reduce albuminuria, and a decrease in albuminuria is highly associated with a slowing of progressive CKD, perhaps because albumin or molecules bound to albumin can directly injure kidney cells.

CLINICAL MANIFESTATIONS

Unfortunately, progressive loss of kidney function produces no clinically distinct signs or symptoms. Findings that should raise the possibility of CKD include hypertension, urinary abnormalities such as hematuria or repeated urinary tract infections, and edema.

As the GFR declines, clinical abnormalities become more frequent, but symptoms are mostly nonspecific even with stage 4 CKD (see Table 130-2). Some patients complain only of exercise intolerance, fatigue, or anorexia. If these symptoms are present, serum creatinine and BUN levels should be measured, and the urine should be examined for albuminuria. For this assessment, a "spot" urine can suffice if the urine albumin and creatinine concentrations are measured; an albumin-to-creatinine ratio of more than 30 mg/g in urine is associated with more adverse outcomes. As CKD progresses, patients frequently develop anemia (Chapter 158), metabolic acidosis (Chapter 118), hyperkalemia (Chapter 117), hyperphosphatemia (Chapter 119), hypocalcemia (Chapter 245), and hypoalbuminemia (Chapter 215), each of which may be associated with symptoms (Table 130-6).

Specific syndromes can be associated with proteinuria and CKD. For example, severe albumin losses (>3 g/day) plus edema and hypercholesterolemia define the nephrotic syndrome (Chapter 121), which can lead to the loss of the relatively small (59 kD) vitamin D–binding protein that is attached to 25-hydroxyvitamin D_3, thereby aggravating renal osteodystrophy. Advanced proteinuria also can be associated with losses of clotting factors IX, XI, and XII, causing coagulation defects (Chapter 174). Conversely, urinary losses of antithrombin III can result in thrombosis (Chapter 171), especially when inflammation increases levels of acute phase reactant proteins, including fibrinogen.

Some patients with renal bone disease complain of vague, ill-defined pain in the lower back, hips, knees, and other locations. In advanced renal osteodystrophy, pain can be so severe that it decreases exercise tolerance, and the resulting immobilization can increase the risk of fractures even with minimal trauma.

Another clinical syndrome related to hyperphosphatemia and hence to the development of renal bone disease is vascular calcification, which causes vascular stiffness, an increase in systolic blood pressure, and the development of left ventricular hypertrophy. A more disabling manifestation is calcification in the tunica media of blood vessels (i.e., Mönckeberg sclerosis) as well as calcifications that impair the function of multiple organs, including the lungs, myocardium, and skin. Calcification of the skin and cutaneous vessels defines the syndrome of calciphylaxis.

DIAGNOSIS

If CKD is suspected, emphasis should be placed on eliciting a history of hypertension, urinary abnormalities, and treatment with drugs that might affect kidney function (e.g., ACE inhibitors, ARBs, and nonsteroidal anti-inflammatory drugs; see Chapter 120). The family history should focus on family members with kidney diseases, kidney stones, surgery involving the urinary tract, diabetes, and hypertension. The physical examination should include lying and standing blood pressure measurements in both arms and a search for findings associated with CKD, such as skin abnormalities, persistent itching, a palpable polycystic kidney (Chapter 127), evidence of lost lean body mass, peripheral edema, and neurologic abnormalities.

Staging

The severity of CKD is divided into five stages based on persistent reductions in estimated GFR (see Table 130-2). Two assessments of impaired kidney function are required: estimated GFR (eGFR) and the degree of albuminuria. The eGFR is calculated from the serum creatinine concentration or the serum level of the protease inhibitor cystatin C plus age, body weight, gender, and race (see Table 130-2).

Shortcomings of the creatinine-based method of assessing kidney function are that it is influenced by both kidney function and creatinine production, the latter of which is directly proportional to lean body mass and is also influenced by the amount of well-cooked meat in the diet. Another shortcoming is that the serum creatinine concentration can remain in the nominally normal range until as much as 50% of kidney function is lost. The assessment of eGFR from cystatin C is limited because the level of this protein is influenced by inflammation, cigarette smoke, and excess acid. In addition, the measurement of cystatin C concentration is not as standardized as the measurement of serum creatinine concentration. Regardless, a low eGFR estimated by the cystatin C method or by a combination of the creatinine and cystatin C methods is closely associated with the development of complications of CKD.[7]

The progression of CKD, which can be estimated by the decline in the reciprocal of the creatinine level (1/serum creatinine), is linear in most patients (Fig. 130-2), and deviations from this linearity suggest a change in the course of CKD. Other markers are not as accurate for estimating changes in kidney function; the BUN concentration, for example, is determined not only by the remaining kidney function but also by the amount of protein in the diet.

A careful microscopic examination of the urine is critical for diagnosis of CKD or changes in progression. The presence of erythrocytes and

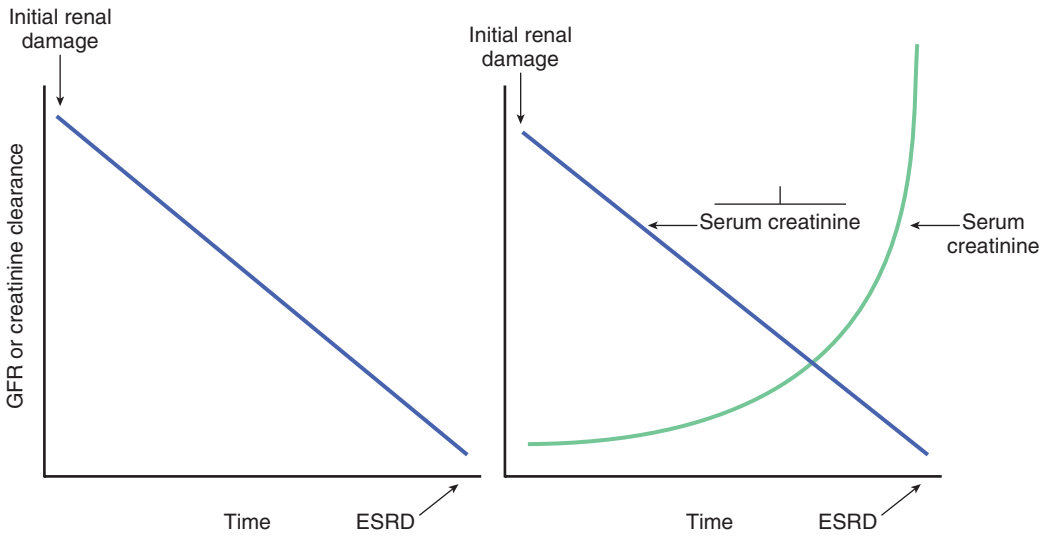

FIGURE 130-2. **The loss of kidney function in chronic kidney disease is constant from initial damage to end-stage renal disease (ESRD, *left panel*).** The constant loss of glomerular filtration rate (GFR) or creatinine clearance is estimated most easily by plotting the reciprocal of serum creatinine against time. (Reprinted with permission from *Annual Review of Medicine* 35. ©1984 by Annual Reviews, www.annualreviews.org.)

erythrocyte casts in the urine sediment is consistent with glomerulonephritis (Chapter 121), the presence of fine granular casts plus protein suggests diabetic kidney disease (Chapter 124), urine samples containing leukocytes plus fine and coarse granular casts suggest interstitial nephritis (Chapter 122), and eosinophils in the urine suggest a drug reaction with interstitial damage (Chapter 122).

Microalbuminuria is defined as 30 to 300 mg albumin/24 hours in a urine specimen, and the abnormality should be present in at least two specimens. A urinary albumin-to-creatinine ratio higher than 30 mg/g in at least two specimens is also associated with high risks of morbidity and mortality in patients with CKD. However, it is uncertain whether albuminuria precedes and potentially causes the loss of eGFR or whether it is a marker of kidney damage.[8] Albuminuria (>300 mg/24 hours) is defined as a persistent excretion rate exceeding that of microalbuminuria.

The urea nitrogen content of the 24-hour urine collection can be used to calculate the amount of dietary protein, which is composed of 16% nitrogen (see E-Fig. 130-1 and Table 130-4). When the amount of urea nitrogen excreted in the steady state (i.e., stable values of weight and BUN) is added to the amount of nonurea nitrogen excreted daily (i.e., 0.031 g nitrogen per kilogram of ideal body weight per day, which includes the nitrogen present in proteinuria of 5 g), the total daily intake of nitrogen can be calculated. If the calculated total of nitrogen excreted exceeds the nitrogen contained in a prescribed diet, the physician should suspect dietary nonadherence, gastrointestinal bleeding, or a catabolic illness.

Other Laboratory Tests

Blood chemistries that evaluate the consequences of CKD include concentrations of sodium, potassium, chloride, bicarbonate, calcium, and phosphorus as well as uric acid levels. The blood glucose level and hemoglobin A_{1c} level should be monitored in diabetics. In patients with suspected glomerulonephritis (Chapter 121), testing commonly includes measurement of antinuclear antibodies, double-stranded DNA antibodies, serum complement levels, antineutrophil cytoplasmic antibody levels, and serologies for hepatitis viruses. The hematocrit or hemoglobin level should be monitored because anemia can develop even with mild renal dysfunction and tends to worsen as CKD progresses because of reduced erythropoietin production and iron deficiency. Iron deficiency can be recognized if the serum iron level is low, the serum ferritin concentration is less than 200 ng/mL, and the transferrin saturation is less than 20% (Chapter 159). Such findings should raise the possibility of gastrointestinal bleeding. To detect CKD-induced bone disease, serum levels of PTH, alkaline phosphatase, calcium, and phosphorus should be obtained.

Imaging

The initial evaluation should include an ultrasound examination of the kidney and bladder to ensure that there is no obstruction of urine flow (Chapter 123) or evidence of polycystic kidney disease (Chapter 127). Enlarged kidneys suggest that CKD may be caused by diabetes (Chapter 124), HIV-associated nephropathy (Chapter 121), infiltrative diseases (e.g., amyloidosis; Chapters 121 and 188), or polycystic kidney disease. Small kidneys, especially with a shrunken kidney cortex, suggest chronic glomerular (Chapter 121) or interstitial (Chapter 122) diseases. If the size of the two kidneys differs substantially, stenosis of the renal artery (Chapter 125) of the smaller kidney should be considered, especially in hypertensive patients.

TREATMENT

Rx

Cardiovascular disorders are common in patients with CKD, in part because hypertension is almost universal. The 2014 Eighth Joint National Committee (JNC 8) for the management of high blood pressure in adults concludes that results from randomized, controlled trials of blood pressure control support lowering of blood pressure to 140/90 mm Hg,[9] although some suggest a lower target of 130/80 mm Hg. Lowering of blood pressure to 140/90 mm Hg with ACE inhibitors or ARBs in combination with a diuretic should be included in the initial treatment plan to slow the loss of kidney function and to reduce cardiovascular events (see Table 67-7). Combining an ACE inhibitor and an ARB provides no additional benefit in terms of protecting the kidney and is associated with more frequent adverse events, so this combination is not recommended. In patients with type 2 diabetes, a systolic blood pressure goal of 135 to 140 mm Hg is preferable to a blood pressure below 120 mm Hg, and the goal for those with hypertensive renal diseases is about 140/85 mm Hg because lowering of blood pressure to 130/80 mm Hg does not further reduce the loss of GFR.

If the serum creatinine concentration increases shortly after ACE inhibitor or ARB therapy is started, it should not trigger an automatic discontinuation of the drug but rather a search for other mechanisms that cause kidney damage (e.g., an increase in blood pressure, urinary infection, use of drugs that adversely affect kidney function, or an exacerbation of the underlying renal disease) and an evaluation to be sure that the patient does not have bilateral renal artery stenosis (Chapter 125). Hyperkalemia also occurs with ACE inhibitor or ARB treatment because a reduced angiotensin II level will suppress aldosterone release. Other causes of an increase in serum potassium concentration should be sought and corrected (e.g., stop treatment with nonsteroidal anti-inflammatory drugs or potassium-sparing diuretics, the development of metabolic acidosis, or increased intake of potassium-rich foods). The abnormality will be corrected in some hyperkalemic patients by reducing dietary potassium-rich foods and adding a loop diuretic (e.g., 40 mg furosemide when the serum creatinine level is below 2 to 3 mg/dL, with higher doses for patients with more advanced CKD). If ACE inhibitor or ARB therapy produces persistent hyperkalemia or an increase in serum creatinine concentration, the dose should be reduced by 50% and amlodipine should be instituted. If the increase in serum potassium concentration persists, the potassium-binding resin Kayexalate can be added to control hyperkalemia (Chapter 117).

For blood pressure goals to be achieved, patients almost always must restrict dietary salt to 2 g sodium/day, which is the equivalent to 86 mEq sodium/day in the urine. Dietary salt restriction to 60 to 80 mEq/day alone

(approximately 1.5 to 2 g of sodium) can reduce blood pressure by 10/4 mm Hg[10] and is routinely recommended. This goal is achievable if patients are taught to avoid salt-rich foods. If blood pressure remains above 140/90 mm Hg, a loop diuretic such as furosemide should be added because it does not compromise renal blood flow (see Table 67-11). A critical strategy is to monitor body weight; an increase in weight and edema signifies salt retention. Conversely, a rapid loss of weight and edema would be the first clue that the diuretic dose should be reduced.

Despite the ability of calcium-channel blockers to combat hypertension, they are not as effective as ACE inhibitors or ARBs in reducing albuminuria, and they can cause peripheral edema. The loop diuretics are preferred for patients with more advanced CKD because they maintain renal blood flow, have few adverse effects, and, unlike thiazide diuretics, remain effective even at GFRs below 25 mL/minute. As renal insufficiency advances, higher daily doses of loop diuretics (e.g., 80 to 160 mg furosemide orally) may be required to reduce extracellular volume. The dose-response relationship of loop diuretics is sigmoidal, so once a dose that reduces edema and decreases body weight is identified, it should not be changed or given in divided doses because its effectiveness will be sharply reduced.

Stage 1 and Stage 2 Chronic Kidney Disease

In patients with stage 1 or stage 2 CKD, uremic symptoms are unusual because kidney function is sufficient to control the levels of potential uremic toxins. Therapy therefore emphasizes reducing blood pressure to 140/90 mm Hg plus intensive treatment of the underlying disease (e.g., treating infections, normalizing the blood glucose concentration in diabetic patients). Physicians often monitor changes in albuminuria and the rate of loss of GFR (Fig. 130-3), but there is no proven benefit of doing so.[11]

Stage 3 and Stage 4 Chronic Kidney Disease

Patients with stage 3 CKD should be referred to a nephrologist to maximize preventive measures and to search for remediable disorders. When stage 4 CKD is reached, a nephrologist should instruct patients about the advantages of therapies such as hemodialysis, peritoneal dialysis, and transplantation (Chapter 131).

In stage 3 and stage 4 CKD, the doses of many drugs that are excreted by the kidney must be reduced to prevent overdosing (Chapter 29). Treatable complications of CKD, including hypertension, secondary hyperparathyroidism, acidosis, and uremic symptoms, must be addressed. Radiologic tests with nephrotoxic contrast dye should be avoided if possible.

The development of many CKD complications (see Table 130-6) requires modification of the diet. Current levels of the estimated intake of protein in the diets of CKD patients substantially exceed recommended amounts.[12] These unrestricted high-protein diets also increase the intake of salt, acid precursors, and phosphates and can lead to the development of metabolic acidosis, hyperkalemia, hyperphosphatemia, edema, hypertension, and uremic symptoms. Recommended diets for CKD patients and especially for patients with complications of CKD should contain 0.8 g protein per kilogram of ideal body weight per day, not the patient's actual weight, which may be a function of edema or obesity. This amount will maintain body protein stores and reduce the likelihood for development of other complications. Calorie intake should be reduced to 30 kcal or fewer calories per kilogram of ideal body weight per day for largely sedentary patients; more calories are needed by those who exercise vigorously. If CKD-induced uremic symptoms persist, dietary protein can be restricted to a minimum of 0.6 g protein/kg/day, with a calorie intake of 30 kcal/kg/day. With both diets, successful implementation requires the advice, guidance, and close monitoring of a dietitian to take

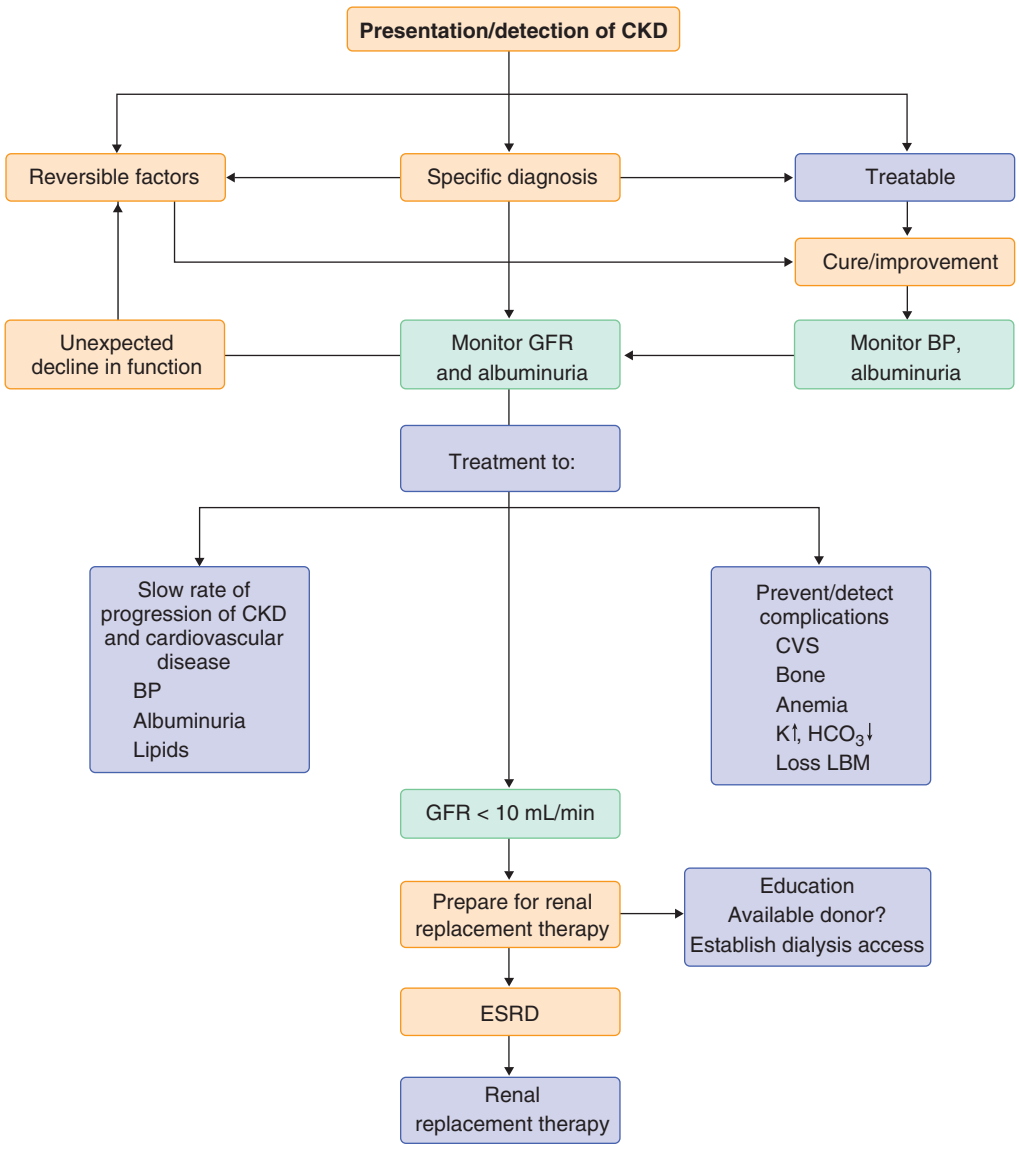

FIGURE 130-3. Management of patients in the various stages of chronic kidney disease (CKD). BP = blood pressure; CVS = cardiovascular system; ESRD = end-stage renal disease; GFR = glomerular filtration rate; LBM = lean body mass.

advantage of the patient's food preferences. Dietary compliance with protein and salt restriction can be monitored by measuring the 24-hour excretion of urea nitrogen and sodium (see Table 130-4), and the adequacy of protein stores should be assessed regularly by measuring body weight and serum protein levels. Although low-protein diets may not slow the loss of kidney function, they reduce uremic symptoms and can delay the need for dialysis without compromising the maintenance of protein stores.[A5] Most patients on a protein-restricted diet should be given a daily supplement of water-soluble vitamins (Chapter 218); fat-soluble vitamins should be prescribed for documented deficiencies.

Renal Bone Disease

Successful treatment of renal bone disease depends on correction of the principal disorder, which is the accumulation of phosphates. Even in patients with serum phosphorus levels below 5.5 mg/dL, abnormalities in calcium and phosphate metabolism can be detected by measuring the serum level of the intact PTH (i.e., the fraction that does not include PTH fragments) or by measuring the serum phosphorus level after a normal meal. A high PTH level means that kidney damage has limited the capacity to excrete phosphates despite the presence of two counterbalancing hormones, PTH and FGF23, both of which increase renal phosphate excretion. For such patients, the dietary content of phosphates must be reduced to less than 800 mg/day. Because the dietary content of phosphates is proportional to dietary protein and because phosphates are present in so many foods, CKD patients require training from a skilled dietitian or nutritionist to interpret labels on prepared foods and to change their diet successfully. The need to modify dietary phosphates and salt is emphasized because excesses of either nutrient can block the beneficial effects of ACE inhibitors or ARBs on slowing the loss of GFR. Fortunately, carefully planned diets are nutritionally sound even when the protein content is markedly restricted, provided there is adequate monitoring.

If dietary restriction proves insufficient to maintain serum phosphorus concentration of 5.5 mg/dL or less, "phosphate binders" (see Table 119-4) can lower phosphate levels.[13] Their efficacy is based on the daily secretion into the intestines of approximately 13 liters of fluid, which has a phosphate content similar to that in blood. Oral administration of binders leads to elimination of phosphates in intestinal fluids, thereby lowering phosphate levels. Second, phosphate binders are composed of a cation (e.g., calcium, lanthanum, or aluminum) that is loosely complexed with an anion (e.g., carbonate, lactate, or hydroxide), so negatively charged phosphates in intestinal fluids can bind to the cation and be eliminated in the stool. Sevelamer hydrochloride is somewhat different because it is a cationic resin that binds phosphates, thereby promoting their intestinal elimination; the chloride anion is buffered.[14]

Phosphate binders should be used in patients with serum phosphorus levels higher than 5.5 mg/dL but lower than 6.5 mg/dL. Traditionally, calcium-based binders have been used for such patients because they are relatively inexpensive and can be effective (e.g., calcium carbonate, initially one or two 500-mg tablets with each meal, or calcium acetate, initially one or two 667-mg tablets with each meal). In general, patients with CKD and hyperphosphatemia should not be given calcium supplements because of the increased risk of soft tissue calcification. The exception is the emergency use of intravenous calcium to treat hyperkalemia (Chapter 117) or neuromuscular irritability (Chapter 245). Even in these conditions, Trousseau sign should be demonstrated to confirm that the ionized calcium level is low. The admonition to avoid calcium supplements for patients with hyperphosphatemia is supported by the reports that these patients have an increased risk of vascular calcification, which may contribute to the 22% higher all-cause mortality rate compared with patients treated with sevelamer or lanthanum carbonate.[A6,A7] Therefore, for patients with serum phosphate levels below 6.5 mg/dL, sevelamer hydrochloride (400- to 800-mg tablets, initially 1.2 g/day in divided doses) is preferred to calcium-based binders. Long-term experience with lanthanum carbonate is limited, but it also may cause fewer cardiovascular problems compared with calcium-based binders. For patients with prolonged levels of serum phosphorus above 8 mg/dL, aluminum hydroxide binders are occasionally administered because they can rapidly lower the serum phosphorus level, but these compounds are generally avoided because they are associated with aluminum accumulation, deposition in bone, and potential neurotoxicity. For patients with a serum phosphorus level above 8 mg/dL, sevelamer hydrochloride (800-mg tablets, initially 1.6 g/day in divided doses) is preferred because it avoids the adverse responses to calcium-based binders.

Calcitriol is often administered to patients with CKD because it can suppress the development of hyperparathyroidism. However, it also increases the intestinal absorption of both calcium and phosphates, so it should not be given to patients who have elevated serum phosphorus levels. For CKD patients with normal serum phosphorus levels, however, calcitriol or paricalcitol can reduce proteinuria by 16% and therefore might improve the course of CKD.

A low 25-hydroxyvitamin D_3 level is frequent in CKD patients and is associated with increased mortality. Patients with insufficient levels of calcitriol or 25-hydroxyvitamin D_3 can be treated with cholecalciferol 1000 units/day. Careful monitoring is needed to avoid hypercalcemia or urinary calcium values above 250 mg/day because this level increases the risk for development of kidney stones.

Calciphylaxis, which results from deposition of calcium phosphate crystals and the secondary inflammation in blood vessels and soft tissues, is an unusual complication in patients with unrelenting hyperphosphatemia and high PTH levels. Calciphylaxis is painful, and treatment options are directed at aggressively restricting dietary phosphates and reducing serum phosphorus with phosphate binders. If the disorder persists despite correction of the serum phosphorus level, a trial of cinacalcet (initial dose of 30 mg and increasing the dose as needed to reduce circulating PTH) is warranted. However, successful treatment generally requires parathyroidectomy.

Anemia

Because of the decreased absorption of oral iron in patients with advanced CKD, ferumoxytol (two 500-mg intravenous injections separated by 3 to 8 days) provides a better response than oral iron supplements and has a low risk of side effects.[A9] In predialysis patients, treatment with erythropoietin (e.g., weekly injections of darbepoetin alfa 0.45 µg/kg) corrects anemia and improves quality of life. The hemoglobin concentration should not be raised above 12 g/dL to avoid an increased risk of stroke.[A10] Consequently, erythropoietin-stimulating agents should be used to maintain the hemoglobin concentration between 10 and 12 g/dL.

Acidosis

Treating metabolic acidosis of CKD with $NaHCO_3$ (initially with two tablets of 650 mg each two or three times daily) to raise the serum bicarbonate concentration above 22 mM can slow the loss of kidney function and improve the metabolism of muscle and bone.[A1,15]

Atherosclerosis

Cardiovascular disease is the most common cause of mortality in CKD. Contributing factors include hypertension, diabetes, increased low-density-lipoprotein cholesterol levels, and vascular calcification. Statins benefit patients with stage 2 or early stage 3 CKD and possibly those with the nephrotic syndrome.[A8] The benefits of antiplatelet therapy among persons with CKD are uncertain and are potentially outweighed by bleeding hazards.[A3]

PROGNOSIS

Because the rate of loss of kidney function varies widely, even among patients who have the same type of kidney disease, it is critical to monitor the course of declining kidney function in each CKD patient by plotting the estimated GFR or 1/serum creatinine versus time (see Fig. 130-2). If the serum creatinine level remains unchanged after 4 months, treatment has probably slowed the progression of CKD. Monitoring of albuminuria provides additional prognostic information because persistent microalbuminuria, and especially albuminuria, is associated with an increased risk of cardiovascular disease and a more rapid loss of kidney function.

Although population data are of limited value for the individual patient with CKD, epidemiologic studies indicate that one third of patients with stage 4 CKD (see Table 130-2) will progress to ESRD within 3 years. Dialysis or transplantation is generally required when the serum creatinine concentration reaches 10 mg/dL. However, patients should be informed about treatment options well before this stage of CKD because the frequency of complications rises sharply when dialysis is initiated on an emergency basis.

Grade A References

A1. de Brito-Ashurst I, Varagunam M, Raftery MJ, et al. Bicarbonate supplementation slows progression of CKD and improves nutritional status. *J Am Soc Nephrol*. 2009;20:2075-2084.

A2. Moe SM, Zidehsarai MP, Chambers MA, et al. Vegetarian compared with meat dietary protein source and phosphorus homeostasis in chronic kidney disease. *Clin J Am Soc Nephrol*. 2011;6:257-264.

A3. Jamerson K, Weber MA, Bakris GL, et al. Benazepril plus amlodipine or hydrochlorothiazide for hypertension in high-risk patients. *N Engl J Med*. 2008;359:2417-2428.

A4. Appel LJ, Wright JT Jr, Greene T, et al. Intensive blood-pressure control in hypertensive chronic kidney disease. *N Engl J Med*. 2010;363:918-929.

A5. McMahon EJ, Bauer JD, Hawley CM, et al. A randomized trial of dietary sodium restriction in CKD. *J Am Soc Nephrol*. 2013;24:2096-2103.

A6. Fouque D, Laville M. Low protein diets for chronic kidney disease in non diabetic adults. *Cochrane Database Syst Rev*. 2009;3:CD001892.

A7. Block GA, Wheeler DC, Persky MS, et al. Effects of phosphate binders in moderate CKD. *J Am Soc Nephrol*. 2012;23:1407-1415.

A8. Jamal SA, Vandermeer B, Raggi P, et al. Effect of calcium-based versus non-calcium-based phosphate binders on mortality in patients with chronic kidney disease: an updated systematic review and meta-analysis. *Lancet*. 2013;382:1268-1277.

A9. Lu M, Cohen MH, Rieves D, et al. FDA report: ferumoxytol for intravenous iron therapy in adult patients with chronic kidney disease. *Am J Hematol*. 2010;85:315-319.

A10. Pfeffer MA, Burdmann EA, Chen CY, et al. A trial of darbepoetin alfa in type 2 diabetes and chronic kidney disease. *N Engl J Med*. 2009;361:2019-2032.

A11. Tonelli M, Wanner C. Lipid management in chronic kidney disease: synopsis of the Kidney Disease: Improving Global Outcomes 2013 clinical practice guideline. *Ann Intern Med*. 2014;160:182-189.
A12. Palmer SC, Di Micco L, Razavian M, et al. Effects of antiplatelet therapy on mortality and cardiovascular and bleeding outcomes in persons with chronic kidney disease: a systematic review and meta-analysis. *Ann Intern Med*. 2012;156:445-459.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Shlipak MG, Matsushita K, Arnlov J, et al. Cystatin C versus creatinine in determining risk based on kidney function. *N Engl J Med*. 2013;369:932-943.
2. United States Renal Data System. *2014 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States*. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 2014.
3. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl*. 2013;3:1-150.
4. Grams ME, Juraschek SP, Selvin E, et al. Trends in the prevalence of reduced GFR in the United States: a comparison of creatinine- and cystatin C–based estimates. *Am J Kidney Dis*. 2013;62:253-260.
5. Parsa A, Kao WH, Xie D, et al. APOL1 risk variants, race, and progression of chronic kidney disease. *N Engl J Med*. 2013;369:2183-2196.
6. Dobre M, Meyer TW, Hostetter TH. Searching for uremic toxins. *Clin J Am Soc Nephrol*. 2013;8:322-327.
7. Stevens PE, Levin A. Evaluation and management of chronic kidney disease: synopsis of the Kidney Disease: Improving Global Outcomes 2012 clinical practice guideline. *Ann Intern Med*. 2013;158:825-830.
8. Krolewski AS, Niewczas MA, Skupien J, et al. Early progressive renal decline precedes the onset of microalbuminuria and its progression to macroalbuminuria. *Diabetes Care*. 2014;37:226-234.
9. James PA, Oparil S, Carter BL, et al. 2014 Evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014;311:507-520.
10. Sacks FM, Campos H. Dietary therapy in hypertension. *N Engl J Med*. 2010;362:2102-2112.
11. Qaseem A, Hopkins RH Jr, Sweet DE, et al. Screening, monitoring, and treatment of stage 1 to 3 chronic kidney disease: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. 2013;159:835-847.
12. Moore LW, Byham-Gray LD, Scott Parrott J, et al. The mean dietary protein intake at different stages of chronic kidney disease is higher than current guidelines. *Kidney Int*. 2013;83:724-732.
13. Bhan I. Phosphate management in chronic kidney disease. *Curr Opin Nephrol Hypertens*. 2014;23:174-179.
14. Carrero JJ, Cozzolino M. Nutritional therapy, phosphate control and renal protection. *Nephron Clin Pract*. 2014;126:1-7.
15. Abramowitz MK, Melamed ML, Bauer C, et al. Effects of oral sodium bicarbonate in patients with CKD. *Clin J Am Soc Nephrol*. 2013;8:714-720.

REVIEW QUESTIONS

1. A 65-year-old African American patient with stage 4 chronic kidney disease (CKD; estimated glomerular filtration rate, 20 mL/minute) is examined in your office. Among laboratory abnormalities uncovered, the subject's serum phosphorus level is 5.6 mg/dL. Regarding the increase in serum phosphorus, which of the following options is correct?

A. An increase in serum phosphorus has not been associated with adverse outcomes in nondialysis patients with CKD and it can be ignored.
B. Use of phosphate binders in nondialysis CKD patients has been shown to correct the high serum phosphorus level and should be encouraged.
C. The patient should meet with a nutritionist for counseling in methods that control phosphate intake. The dietitian's help is needed because phosphates have been added to many foods, and patients must learn how to avoid foods rich in phosphates and how to use phosphate binders.
D. The parathyroid hormone and fibroblast growth factor 23 (FGF23) levels should be measured as they could be subnormal.

Answer: C An increase in serum phosphorus is a strong predictor of mortality and adverse outcomes in CKD patients. Because phosphates are added to prepared foods, it is difficult to reduce dietary sources of phosphates. However, clinical trials of reducing serum phosphorus by giving phosphate binders to predialysis CKD patients have not demonstrated improved mortality.[1] Both FGF23 and parathyroid hormone increase urinary losses of phosphates, so they rise when serum phosphorus levels rise.

1. Chue CD, Townend JN, Moody WE, et al. Cardiovascular effects of sevelamer in stage 3 CKD. *J Am Soc Nephrol.* 2013;24:842-852.

2. Which one of the following statements about the interplay of diet and treatment for CKD patients is true?

A. High-salt diets interfere with angiotensin-converting enzyme (ACE) inhibitor therapy, thereby compromising blood pressure control and the ability of ACE inhibition to slow the progression of CKD.
B. A low-salt, high-potassium diet can lower systolic blood pressure in hypertensive, otherwise normal individuals.
C. High-salt diets interfere with diuretic therapy and should be avoided.
D. Low-protein diets can reduce uremic symptoms in patients with advanced CKD.
E. All of the above

Answer: E References 2 and 3 document that inadequate control of the diet counteracts the benefits of ACE inhibitor therapy. Reference 4 has details about the low-salt, high-potassium DASH diet. References 5 and 6 provide information that excess dietary salt overcomes the benefits of diuretics and how CKD-induced metabolic problems can be ameliorated by modifying the diet.

2. Zoccali C, Ruggenenti P, Perna A, et al. Phosphate may promote CKD progression and attenuate renoprotective effect of ACE inhibition. *J Am Soc Nephrol.* 2011;22:1923-1930.

3. Vegter S, Perna A, Postma MJ, et al. Sodium intake, ACE inhibition, and progression to ESRD. *J Am Soc Nephrol.* 2012;23:165-173.

4. Appel LJ, Moore TJ, Obarzanek E, et al. A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. *N Engl J Med.* 1997;336:1117-1124.

5. Kelly RA, Wilcox CS, Mitch WE, et al. Response of the kidney to furosemide. II. Effect of captopril on sodium balance. *Kidney Int* 1983;24:233-239.

6. Mitch WE, Remuzzi G. Diets for patients with chronic kidney disease, still worth prescribing. *J Am Soc Nephrol.* 2004;15:234-237.

3. A 30-year-old man who has become an expert in computer-based games has been told he has kidney disease and needs dialysis. He comes to you for a second opinion. His physical examination reveals 3+ edema and a blood pressure of 200/98 mm Hg plus. Abnormal blood chemistry values include the following: serum creatinine, 2 mg/dL; blood urea nitrogen (BUN), 50 mg/dL; serum potassium, 5 mEq/L; serum bicarbonate, 16 mM/μL; and serum phosphorus, 6.2 mg/dL. Which of the following initial evaluations will be most helpful in the treatment of the patient?

A. A discussion of the role of dialysis in avoiding complications of end-stage renal disease
B. A careful family history to determine if there are inherited kidney diseases and potential family members who might donate a kidney
C. A dietary history plus collection of a 24-hour urine specimen to measure the content of urea nitrogen, creatinine, albumin, and sodium
D. Administration of hydralazine and amlodipine to reduce his blood pressure
E. Measurements of calcitriol and 25-hydroxyvitamin D_3 levels. If the values are low, supplement him with vitamin D. To compensate for protein losses, patients with the nephrotic syndrome should gain 2.5 to 3.5 g/kg of nonedematous body weight daily.

Answer: C The patient has complications of CKD, but the degree of CKD is within stage 3 (estimated glomerular filtration rate, 45 mL/min/1.73 m^2). The 24-hour urine sample can guide dietary changes to reduce complications of CKD. His urinary albumin-to-creatinine ratio will provide more information about the severity of CKD, and his excretion of urea nitrogen will be used to restrict dietary protein to lower his BUN and to correct the retention of phosphates and acid. Sodium excretion will be used to design a salt-restricted diet because the patient has hypertension and edema. A and B: the degree of CKD is too low to consider dialysis or transplantation at this time. D: the preferred antihypertensive drug is an ACE inhibitor or angiotensin receptor blocker, and it will be difficult to correct his blood pressure without diuresis and dietary salt restriction. E: because he remains inside to play computer gains, he may have low vitamin D levels. However, he has an elevated serum phosphorus level, which is a contraindication to vitamin D administration.

131 TREATMENT OF IRREVERSIBLE RENAL FAILURE

DAVID COHEN AND ANTHONY MICHAEL VALERI

Chronic kidney disease (Chapter 130) tends to progress over time owing to progressive nephron dropout, glomerular capillary hypertension, and glomerular hyperfiltration, whether it is caused by primary glomerular injury or by tubulointerstitial or vascular injury. Irreversible and advanced end-stage renal disease (ESRD) requires renal replacement therapy, which can be broadly categorized as hemodialysis, peritoneal dialysis, and renal transplantation.[1] In the United States, about 640,000 people receive some form of renal replacement therapy each year.

Renal replacement therapy must be initiated when fluid and electrolyte derangements, particularly hyperkalemia and acidosis, can no longer be adequately controlled with dietary modifications and medications (Chapter 130) or if uremic symptoms, such as anorexia, nausea, vomiting, gastritis, pericarditis, or encephalopathy, develop (Table 131-1). Renal replacement therapy typically is required when the estimated glomerular filtration rate (eGFR) is below 10 mL/minute, although it may be needed at an eGFR of 10 to 15 mL/minute when comorbid conditions, particularly heart failure, make medical management even more challenging and difficult. However, preventive earlier initiation of renal replacement therapy at an eGFR of 10 to 14 mL/minute is no better than later initiation at an eGFR of 5 to 7 mL/minute or when uremic symptoms require it.

Hemodialysis relies on diffusion across a semipermeable artificial membrane, whereas peritoneal dialysis brings the blood and the dialysate solution in contact across a natural biologic membrane. The diffusion of solutes along their respective concentration gradients across a semipermeable membrane removes nitrogenous waste products and corrects imbalances of potassium, calcium, magnesium, phosphorus, and acid. In addition, plasma water filters across the membrane and, by convection, drags solutes across the membrane in approximately the same concentration as in the plasma water. The electrolyte concentrations in the dialysate solution are not necessarily physiologic but are intentionally varied with respect to their potassium, calcium, magnesium, and bicarbonate concentrations to favor correction of the plasma toward a more normal physiologic state. For example, a typical dialysate solution might use a potassium concentration of 2 mEq/L and a bicarbonate concentration of 35 mEq/L to produce concentration gradients that favor correction of hyperkalemia and uremic metabolic acidosis. Convection across the dialysis membrane is driven by either a hydrostatic pressure gradient applied across the membrane (hemodialysis) or an oncotic pressure gradient using high-dextrose concentrations or a large, poorly absorbed carbohydrate polymer in the dialysate solution (peritoneal dialysis) for the removal and filtration of excess salt and water from the body.[2]

TABLE 131-1 INDICATIONS FOR DIALYSIS IN CHRONIC KIDNEY DISEASE

Uremic encephalopathy or neuropathy
Pericarditis or pleuritis
Bleeding attributable to uremia
Fluid overload refractory to diuretics
Hypertension poorly responsive to medication
Persistent hyperkalemia, metabolic acidosis, hypercalcemia, hypocalcemia, or hyperphosphatemia refractory to medical therapy
Malnutrition or weight loss
Persistent nausea and vomiting

From Tolkoff-Rubin N. Treatment of irreversible renal failure. In: Goldman L, Schafer AI, eds. *Goldman's Cecil Medicine*. 24th ed. Philadelphia: Elsevier Saunders; 2012.

DIALYSIS

Hemodialysis

Conventional hemodialysis is typically provided at an outpatient dialysis unit where patients are treated three times per week for 3 or 4 hours per session. The measures used for determining the adequacy of treatment are based on urea clearance (as a surrogate marker molecule for the generation and clearance of small-molecular-weight nitrogenous waste products, <500 daltons). The simplest measure is the urea reduction ratio, which is the percentage fall in the blood urea nitrogen level with each dialysis session, with the goal being a 65% or greater fall in blood urea nitrogen with each dialysis session on a thrice-weekly dialysis schedule. More precise measurements can fine-tune the dialysis in an individual patient.[3] A more intensive protocol for increased urea clearance on a thrice-weekly schedule does not improve survival of patients, perhaps because it increases the likelihood of hypotension during dialysis.[4] More frequent hemodialysis, whether as in-center hemodialysis or nocturnal home hemodialysis six times per week, improves outcomes such as kidney-specific measures of quality of life, blood pressure, regression of left ventricular hypertrophy, and serum phosphorus levels, but it has not reduced mortality.

Patients on hemodialysis are exposed to a large volume of dialysate solution (typically 36 to 48 L/hour during dialysis) and must be protected against even small quantities of impurities, such as trace minerals, bacteria, and bacterial endotoxins, in the dialysis solution. To this end, dialysate solutions are prepared from concentrates diluted with water prepared by reverse osmosis systems or deionizer tanks to remove undesired trace cations and anions and then filtered through small micron-sized pore filters to remove bacteria and their byproducts. Exposure to low levels of bacteria and bacterial byproducts may contribute to the chronic inflammation seen in some patients on hemodialysis.

In critically ill patients who are hemodynamically unstable, one alternative is continuous venovenous hemofiltration, which requires central venous access (double-lumen catheter) and blood flows between 150 and 200 mL/minute. Plasma water under pressure passes across one side of a highly permeable membrane, thereby allowing both water and solutes up to about 60 kD to be removed. In contrast to hemodialysis, urea, creatinine, and phosphate are cleared at similar rates (convective clearance) during hemofiltration. The filtrate is discarded, and the fluid lost is partially replaced with a solution containing the major crystalloid components of the plasma at physiologic levels. However, there is no evidence from randomized trials that continuous venovenous hemofiltration offers a survival advantage compared with intermittent hemodialysis in patients with acute renal failure.

New modes of hemodialysis, used predominantly outside the United States, have sought to take advantage of convection to increase the clearance of middle molecules by using high-flux dialysis membranes and ultrafiltration rates of more than 20 L of fluid per dialysis session. In a large randomized trial, this approach, termed high-efficiency postdilution online hemofiltration, reduced all-cause mortality by 30%.

The most common complications during hemodialysis are vascular access problems, hypotension, muscle cramps, nausea, vomiting, headache, and chest pain. Excessive fluid removal is the most frequent cause of hypotension, but persistent hypotension may be caused by sepsis (Chapter 108), myocardial ischemia (Chapter 72), pericardial tamponade (Chapter 77), arrhythmias (Chapters 62-65), and active bleeding. A rare complication is an air embolus (Chapter 98), which is manifested with agitation, cough, dyspnea, and chest pain. The patient should be given 100% oxygen and be positioned with the left side down in an attempt to trap air in the right ventricle.

Peritoneal Dialysis

Peritoneal dialysis, which is an alternative mode of renal replacement therapy, is usually performed at home by the patient or the patient's family. A peritoneal dialysis catheter is implanted through a surgically created tunnel in the abdominal wall and inserted into the peritoneal cavity. The catheter's tip in the pelvis is used to instill a dialysis solution into the peritoneal cavity and then to drain the solution from it. Peritoneal dialysis uses

the peritoneal membrane, which consists of the visceral peritoneum, the interstitial tissues, and the mesenteric capillaries, as the filter across which the diffusion of solutes and convection of plasma water occur. Different forms include continuous ambulatory peritoneal dialysis, which is typically performed with 1.5- to 3-L exchanges of peritoneal dialysis solution instilled in the abdomen through a Tenckhoff catheter four times per day; automated peritoneal dialysis, which is performed with a cycler machine at night cycling 1.5 to 3 L of peritoneal dialysis fluid into and out of the abdomen four or five times overnight; and continuous cyclic peritoneal dialysis, a hybrid of continuous ambulatory peritoneal dialysis and automated peritoneal dialysis that uses cycler therapy at night combined with one or two manual exchanges during the day. *Kt/V* urea is the total volume clearance of urea (*Kt*) normalized to the urea space, *V*, which is approximately equal to total body water.

Clearance of small solutes is a key predictor of survival in patients undergoing peritoneal dialysis, and residual renal function also plays a critical role. Current guidelines advocate a minimal target *Kt/V* urea of at least 1.7 per week.[1] Every effort should be made to maintain residual renal function as long as possible by avoiding nephrotoxins such as nonsteroidal anti-inflammatory drugs, iodinated contrast agents, and aminoglycosides. As residual renal function diminishes over time, the peritoneal dialysis prescription needs to be adjusted accordingly.

MEDICAL ISSUES

Renal replacement therapy is concerned not only with adequate dialytic clearance of nitrogenous waste products, the restoration of acid-base and electrolyte balance, and the control of salt and water balance but also with preserving adequate access sites for dialysis, nutrition, and the management of anemia, bone diseases, and cardiovascular risk.

Access Issues

The optimal dialysis access for hemodialysis is a native vein arteriovenous fistula, which can be created by the surgical anastomosis of the radial artery to the cephalic vein or of the brachial artery to either the brachiocephalic or the basilic vein. If a patient's veins are inadequate, a synthetic graft can be placed between the radial, brachial, or axillary artery and the brachiocephalic, basilic, or axillary vein as an alternative. Most commonly, these grafts are made from a synthetic polymer, expanded polytetrafluoroethylene. Less frequently, urgent temporary access can be obtained with a dual-lumen central venous dialysis catheter that is inserted into the internal jugular (preferable) or subclavian vein (right preferred to left). Dialysis access requires periodic Doppler monitoring to detect stenoses that develop as a result of intimal hyperplasia from high turbulent flow and that can lead to reduced or even inadequate blood flow and recirculation within the access. These stenoses can be treated by percutaneous transluminal angioplasty, although stenting is sometimes required to maintain adequate luminal patency.

Nutrition

Adequate nutrition is important for optimizing outcome with renal replacement therapy, and rates of hospitalization are reduced when patients ingest at least 1 g/kg/day of protein. Another marker of nutritional status is the serum albumin level, which generally serves as a good predictor of outcomes with renal replacement therapy.

Anemia

Anemia in ESRD is due to the reduced production of erythropoietin by the diseased kidneys, a shortened half-life of red blood cells, and the potential loss of red blood cells in the extracorporeal dialysis circuit and the gastrointestinal tract (related to intermittent anticoagulation for the hemodialysis procedure). Erythropoietic stimulating agents should aim to avoid transfusion[5] by keeping the hemoglobin level generally above 9 g/dL but not higher than 11 g/dL[1] because higher levels are associated with more cardiovascular complications. Patients who have relative or absolute iron deficiency, as evidenced by a transferrin saturation below 20% or a serum ferritin level below 200, are often relatively resistant to erythropoietic stimulating agents and may benefit from the concurrent administration of a 1-g course (during 8 to 10 dialysis sessions) of intravenous iron (iron sucrose, ferric gluconate, or ferumoxytol).[1]

Metabolic Bone Disease

Metabolic bone disease and phosphorus balance are common in ESRD. Clinical manifestations of renal osteodystrophy (Chapter 130) can range from adynamic bone disease to osteomalacia (Chapter 244) to secondary hyperparathyroidism (Chapter 245) and osteitis fibrosa cystica. The goals of therapy are to achieve a serum phosphorus level within or close to the normal physiologic range, typically about 2.5 to 5.5 g/dL; a corrected calcium ($0.8 \times [4 - \text{the serum albumin level}] + \text{the serum calcium level}$])–phosphorus product less than 55; and an intact parathyroid hormone level within two to nine times the upper limit of normal.[6] Management includes the restriction of dietary phosphate to 750 to 1000 mg/day of elemental phosphorus; the use of phosphate binders (calcium carbonate or acetate, sevelamer carbonate, lanthanum carbonate, or potentially other cationic agents, magnesium or ferric ion) with meals to bind phosphate in the intestines and to reduce its absorption (see Table 119-4); the use of activated forms of vitamin D (e.g., calcitriol [0.25 to 5.0 μg], paricalcitol [1 to 15 μg], or doxercalciferol [1 to 7 μg] intravenously at each hemodialysis session or 50% of that dose orally daily) to stimulate the parathyroid cells to suppress parathyroid hormone secretion; and cinacalcet (30 to 180 mg daily) to significantly reduce secondary hyperparathyroidism by nearly 70%,[1] even though it does not provide a survival benefit over activated vitamin D therapy alone.[1]

Cardiovascular Disease

Cardiovascular disease is the leading cause of morbidity and mortality in patients on dialysis. The mainstay of blood pressure control in renal replacement therapy is to achieve the lowest tolerated postdialysis weight to facilitate blood pressure control and to enhance the body's sensitivity to antihypertensive agents. The general approach to antihypertensive medication therapy is similar to the approach for other patients, with the exception that diuretics are not helpful (see Tables 67-7 and 67-11).

Dialysis-dependent chronic kidney disease is an indicator for the same type of aspirin therapy and β-adrenergic blockade used in survivors of a myocardial infarction (Chapters 72 and 73). Statin therapy (Chapter 206) is often recommended, but statins are not of proven benefit for reducing major cardiovascular events or overall mortality in dialysis patients despite clinically relevant reductions in low-density lipoprotein cholesterol levels.[1] The addition of spironolactone (25 mg daily) can reduce the risk of a cardiovascular or cerebrovascular event by about 50%.[1] Control of diabetes (Chapter 229) is important, but mortality in hemodialysis patients is not increased until the hemoglobin A_{1c} level is 8.5% or higher.[7]

Amyloidosis

Dialysis-associated amyloidosis (Chapter 188), which is typically seen in patients who have been on dialysis for more than 10 to 12 years, is caused by the deposition of end-products of β_2-microglobulin catabolism as amyloid fibrils. The clinical syndrome includes carpal tunnel syndrome (Chapter 420), arthropathy, and autonomic neuropathy (Chapter 418).

Infection

Infection is the second leading cause of morbidity and mortality in dialysis patients. In hemodialysis, vascular access sites are the primary sources of bacteria and account for about 75% of all cases of bacteremia. In a patient with possible catheter-related bacteremia, blood culture specimens should be obtained both from the catheter and from a peripheral vein. Most vascular access infections are caused by staphylococcal organisms (Chapter 288) and should be treated empirically with 1 g of vancomycin, continued every 3 days for at least 2 weeks while blood levels are monitored, unless another organism is cultured. In patients with systemic sepsis (Chapter 108) with hemodynamic instability, the line should be pulled promptly and be reinserted only after blood cultures are negative for at least 48 hours and the patient has defervesced. Interim dialysis can use an alternative access site, such as a temporary nontunneled catheter. A prolonged course of antibiotic therapy (4 to 8 weeks) is recommended for bacteremia or fungemia that persists after the catheter is removed or in patients with evidence of endocarditis (Chapter 76), septic arthritis (Chapter 272), osteomyelitis (Chapter 272), epidural abscess (Chapter 413), or other metastatic infection.

With peritoneal dialysis, infection is the most frequent reason that the catheter must be removed and therapy must be discontinued. Infection is suspected by erythema, tenderness, induration, or purulent or bloody drainage. Peritonitis must be suspected in patients with abdominal pain and cloudy dialysate; the physical examination usually shows abdominal tenderness, often with rebound. The peritoneal fluid white blood cell count is typically above 100/μL with a predominance of neutrophils, although lymphocytes may predominate with fungal or mycobacterial infections. Polymicrobial infections should raise the possibility of a perforated diverticulum (Chapter

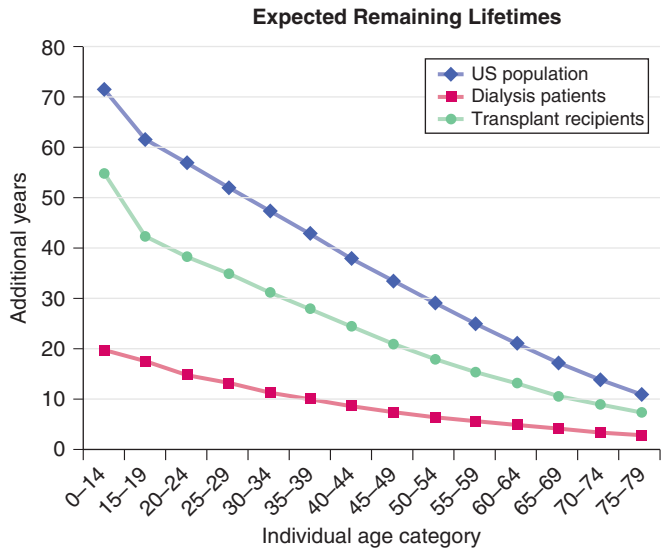

FIGURE 131-1. Life expectancy in end-stage renal disease compared with the general population. (Source: Organ Procurement and Transplantation Network. *Concepts for Kidney Allocation*. http://optn.transplant.hrsa.gov/SharedContentDocuments/KidneyConceptDocument.PDF. Accessed February 2, 2015.)

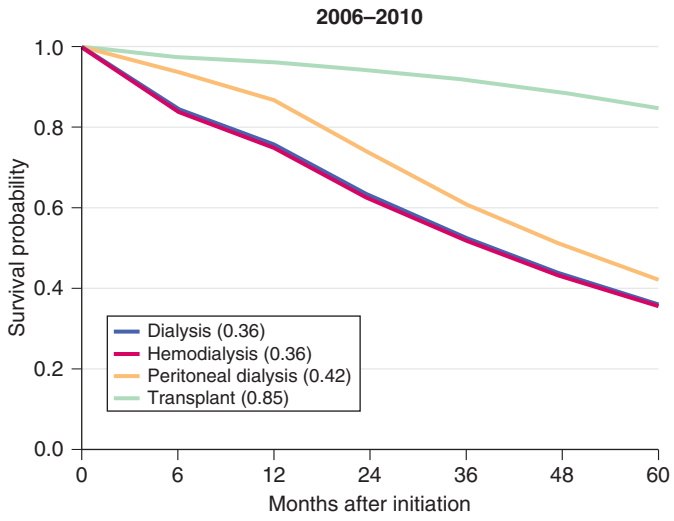

FIGURE 131-2. Adjusted 5-year survival, by modality. (Source: The U.S. Renal Data System. *USRDS 2013 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States*. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 2013.)

142), ruptured appendix (Chapter 142), ischemic bowel (Chapter 143), incarcerated hernia (Chapter 142), pancreatitis (Chapter 144), or gynecologic conditions (Chapters 199, 235, and 236) and should prompt an emergent abdominal computed tomography scan. *Staphylococcus epidermidis* (Chapter 288) is the most common cause of peritonitis, but *Pseudomonas* species (Chapter 306) account for 5 to 8% of cases. Recommended initial empirical treatment is intraperitoneal, unless there is evidence of bacteremia or hematogenous spread of the infection. Options include vancomycin (e.g., 1 g every 5 to 7 days, as guided by serum levels, for at least 2 weeks) or a cephalosporin (e.g., cefazolin, 15 mg/kg in one exchange per day) together with intravenous or intraperitoneal administration of a third-generation cephalosporin with antipseudomonal activity (e.g., ceftazidime, 1 to 1.5 g in one exchange per day) or gentamicin (0.6 mg/kg in one exchange per day or 80 mg intravenously) for at least 2 weeks.

Staphylococcus aureus (Chapter 288) is the most common organism for exit site and tunnel infections. Treatment options include empirical oral penicillinase-resistant penicillins (e.g., dicloxacillin, 250 to 500 mg twice daily for 14 days), fluoroquinolones (e.g., ciprofloxacin, 250 to 500 mg twice daily for 14 days), trimethoprim-sulfamethoxazole (e.g., 40/800 mg for 14 days), and cephalosporins (e.g., cephalexin, 500 mg twice daily for 14 days). Vancomycin should be avoided as first-line therapy except for methicillin-resistant *S. aureus.* Mupirocin nasal ointment applied to the exit site twice daily for 5 days every 4 weeks significantly reduces the incidence of *S. aureus* exit site infections.

RENAL TRANSPLANTATION

The success of renal transplantation has increased dramatically during the past several decades owing to improved surgical technique, improved medical care, and more effective and safe immunosuppressive medications. Although transplant recipients do not have a normal life expectancy (Fig. 131-1), current 5-year survival rates are nearly twice as high as for similar patients who remain on dialysis or on the transplant waiting list (Fig. 131-2). Several long-held axioms have been proved false, most notably the need for long-term steroid therapy, the impossibility of transplantation in the face of a positive crossmatch, the inability to transplant ABO-incompatible donor-recipient pairs, and the inability to create long-term tolerance. Despite this remarkable progress, the number of patients on the waiting list far exceeds the number of available kidneys, and the long-term survival rates for kidneys and patients remain disappointing.

EPIDEMIOLOGY AND DEMOGRAPHICS

Between 17,000 and 18,000 total renal transplants, including combined kidney-pancreas, kidney-liver, and kidney-heart, are performed annually in the United States. The most common disease leading to renal transplantation is diabetic nephrosclerosis (Chapter 124), followed by hypertensive nephrosclerosis (Chapter 125) and other forms of glomerulonephritis (Chapter 121). Kidney transplant recipients vary in age from infants to older than 80 years, with the majority being between 35 and 64 years of age. About 65% of the kidneys transplanted in the United States are from deceased donors (E-Fig. 131-1), with the other 35% from living donors, who include both living genetically related family members and genetically unrelated donors, most often spouses or friends. Unfortunately, the number of kidney donors, both living and deceased, has remained relatively constant, whereas the number of patients awaiting transplant continues to increase. The number of patients on the active waitlist is currently three times larger than the supply of kidneys (E-Fig. 131-2), and as a result, the median waiting time for a patient listed for kidney transplantation is currently about 4.3 years. About 10% of newly listed candidates die within 3 years of listing without having received a transplant.[1]

Renal Transplant Success Rates

For deceased donor kidney transplants, the overall 1-year success rate is approximately 91%, and the 10-year success rate is just below 50%. For recipients of live donor kidney transplants, the overall 1-year success rate is more than 96%, and the 10-year success rate is more than 60% (E-Fig. 131-3).

Kidney Donation

Potential *living kidney donors* undergo a thorough evaluation to ensure that age-adjusted renal function is normal, that the surgical risk of donor nephrectomy is acceptably low, and that there are no medical conditions that would increase the risk of future renal disease in the donor. In addition, all donors are screened for transmissible infections or malignant neoplasms. Potential donors must be capable of understanding the risks and benefits of live kidney donation, cannot be coerced into donating, and cannot condition the donation on the receipt of money or other valuable goods. Federal law prohibits the buying and selling of organs. Recent data suggest that live kidney donors have a slightly elevated relative risk[8] for development of ESRD compared with healthy nondonors during the 10 years after donation, but this risk remains less than 1%.

Potential recipients who have a willing and medically and psychosocially suitable live donor but who either have preformed antidonor HLA antibodies or are blood group incompatible with their donor can receive successful transplants by protocols that involve intravenous immune globulin and plasmapheresis to reduce the level of antidonor HLA antibodies or isoagglutinins. Despite remarkable short-term success, the high incidence of antibody-mediated rejection has made long-term success rates somewhat disappointing, at about 60 to 70% at 5 years. An increasingly popular alternative is kidney-paired donation, in which two live donor-recipient pairs are found and each donor is immunologically incompatible with his/her intended recipient but is compatible with and donates to the other recipient. These paired donations now represent nearly 10% of all live donor kidney transplants in the United States. In some situations, even more complicated chains may involve three or more donor-recipient pairs.

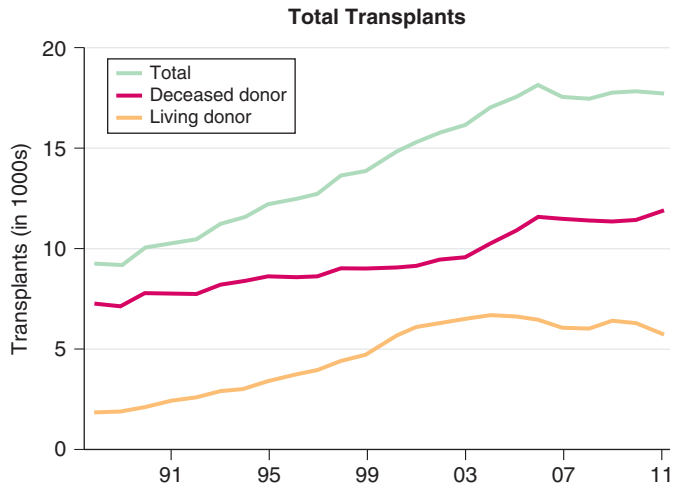

E-FIGURE 131-1. Volume of renal transplants in the United States. (From The U.S. Renal Data System. *USRDS 2013 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States*. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 2013.)

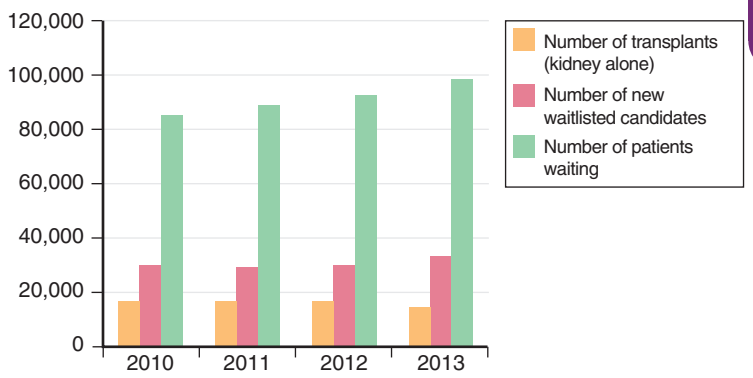

E-FIGURE 131-2. Supply of kidneys versus demand. (Source: Organ Procurement and Transplantation Network and the Scientific Registry of Transplant Recipients, 2012 Annual Data Report. Based on data from Matas AJ, Smith JM, Skeans MA, et al. OPTN/SRTR 2012 Annual Data Report: kidney. *Am J Transplant.* 2014;14(Suppl 1):11-44.)

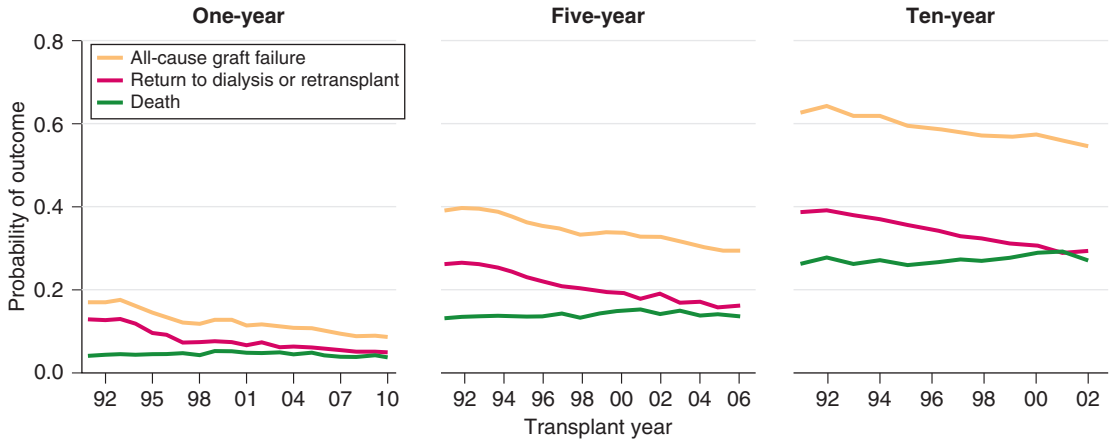

A. Deceased donor transplants

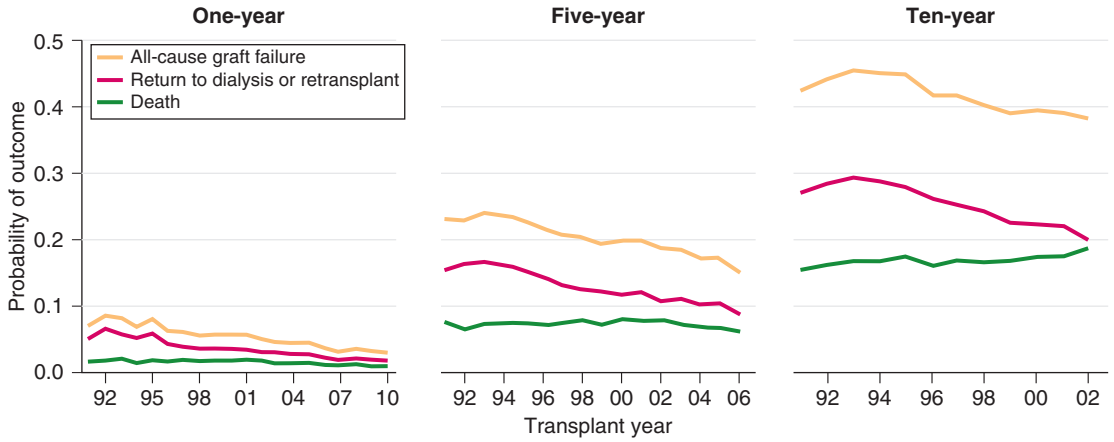

B. Live donor transplants

E-FIGURE 131-3. A, Deceased donor transplants. B, Live donor transplants. (From The U.S. Renal Data System. *USRDS 2013 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States*. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 2013.)

The majority of *deceased donors* are brain dead, as a result of head trauma (Chapter 399), cerebrovascular catastrophe (Chapter 408), or anoxia (Chapter 63), but continue to maintain a normal circulation. The Uniform Anatomical Gift Act of 1968 and subsequent revisions equates brain death with death (Chapter 2) and permits organ donation after a declaration of brain death, provided there is consent from the family of the deceased. Events surrounding brain death and organ procurement may result in acute kidney injury at the time of procurement, and renal biopsies are frequently performed before implantation to assess kidney quality and potential longevity.

The shortage of available donor organs continues to plague the field of clinical transplantation. Attempts to increase the number of donor organs have included the use of kidneys of lower quality, previously termed "expanded criteria" deceased donors. The former categories of Standard Criteria Donor and Expanded Criteria Donor have been replaced by a continuous grading system of deceased donor kidneys—the Kidney Donor Profile Index (KDPI). Multiple donor characteristics are used to calculate the KDPI, including the donor's age, weight, serum creatinine, history of hypertension, history of diabetes, and cause of death. Lower KDRI values—higher quality kidneys—are associated with a lower inherent estimated risk of kidney failure (longer anticipated function after transplantation), whereas transplantation of kidneys with higher KDRI values will result in shorter estimated allograft survival times. This prognosis allows for better informed decision-making on the part of both patients and physicians. Developing algorithms to facilitate the optimal use of sub-optimal kidneys—those with higher KDRI—will continue to be a challenge. Protocols have also been developed to allow organ recovery from deceased donors after cessation of cardiac function (Donation after Cardiac Death or DCD donors). These kidneys function as well as those recovered by traditional donation after brain death.

Who Is a Candidate for Renal Transplantation?

All patients with advanced chronic kidney disease must be informed about the option of transplantation. Candidates must undergo a comprehensive medical and psychosocial evaluation to determine their suitability. Because cardiovascular disease is common in ESRD patients, cardiovascular testing is often undertaken in patients older than 50 years and in patients with a history of diabetes (see Table 431-5).[9] Up to 30% of patients never complete their evaluation or are determined not to be suitable candidates for transplantation.

Allocation

The national system for the allocation of deceased donor organs in the U.S. is divided into 58 donor service areas, each with an organ procurement organization responsible for procuring and allocating deceased donor organs. The distribution of these organs is first local (within the donor service area), then regional, and finally national if no local compatible recipient exists. The current allocation system (in effect as of December 2014) largely prioritizes waitlisted candidates based on waiting times—whoever has waited the longest—as determined by the earlier of the date of initiating dialysis or the date of being placed on the waitlist, which is permitted once the eGFR is <20 mL/min, is first in line for the next available compatible kidney within their blood group. In addition, each waitlisted patient will receive an Estimated Post-Transplant Survival (EPTS) score based on based on four factors: the candidate's time on dialysis, current diagnosis of diabetes, prior solid organ transplants, and age. Patients with the best EPTS score are prioritized to receive donor organs with longer anticipated function (lower KDPI), beginning the practice of longevity matching in kidney allocation. Other special consideration is given to children who are waitlisted prior to age 18 years, to patients with high levels of anti-HLA antibodies (for whom compatible donors are difficult to find), and to donor-recipient pairs with high degrees of HLA matching.[10]

Immunosuppression

Notwithstanding several reports of successful medication-free long-term renal allograft survival in small numbers of carefully selected patients undergoing complicated conditioning regimens, long-term immunosuppression continues to remain obligatory in virtually all patients to achieve long-term renal allograft survival. The principal drivers of the immune response to an organ allograft and the principal obstacle to organ acceptance are the HLA antigens (Chapter 49). HLA-identical siblings enjoy the best long-term survival rates, and increasing degrees of HLA antigen *mis*match continue to be associated with inferior long-term outcomes (Chapter 49).

Immunosuppressive treatment can be divided into three phases: induction, maintenance, and antirejection (Table 131-2). Induction immunosuppression consists of antilymphocyte antibodies given within the first week immediately after transplantation. Maintenance immunosuppression refers to those medications given daily as long as the allograft is functioning. Antirejection therapy is given to treat acute rejection episodes when they occur.

For about two thirds of patients, the standard maintenance therapy for the long-term prevention of renal allograft rejection is a three-drug regimen: tacrolimus, mycophenolate, and prednisone. In one third of patients, steroids are withdrawn within the first week after transplantation, but patients are maintained with a dual-therapy regimen of mycophenolate and tacrolimus. Outcomes appear to be similar with either regimen.

Induction therapy consists of either polyclonal or monoclonal antilymphocyte antibodies administered in the first week after transplantation. Approximately 80% of kidney transplant recipients are in centers that administer induction therapy, the majority with rabbit-derived polyclonal antilymphocyte serum (usually 600 mg/kg total dose) and the others with either basiliximab (20 mg, day 0 and day 4) or alemtuzumab (30-mg single dose).

TABLE 131-2 IMMUNOSUPPRESSIVE MEDICATIONS USED IN RENAL TRANSPLANTATION

CLASS	DRUGS	HOW USED	MECHANISM OF ACTION	MAJOR ADVERSE EFFECTS
Steroids	Prednisone, methylprednisolone	Induction, maintenance, antirejection	Multiple sites, anti-inflammatory; inhibit production of multiple cytokines	Poor wound healing, weight gain, diabetes, acne, hypertension, cushingoid appearance
Calcineurin inhibitors	Tacrolimus and cyclosporine	Maintenance	Inhibit calcineurin, prevent interleukin-2 production	Nephrotoxicity, hypertension, diabetes mellitus
Purine synthesis inhibitors	Azathioprine Mycophenolic acid	Maintenance Maintenance	Metabolized to 6-mercaptopurine Inhibits purine synthesis by the inhibition of inosine monophosphate dehydrogenase, specific for the purine de novo pathway on which lymphocytes depend	Anemia, leukopenia, thrombocytopenia Anemia, leukopenia, thrombocytopenia; teratogenic
mTOR inhibitors	Sirolimus, everolimus	Maintenance	Inhibit mTOR (mammalian target of rapamycin), block lymphocyte response to growth factors	Poor wound healing, proteinuria, hyperlipidemia
Costimulation blockade	Belatacept	Maintenance	Human recombinant fusion protein combining the extracellular portion of cytotoxic T lymphocyte–associated antigen-4 (CTLA-4) with the Fc fragment of human IgG; this binds to the CTLA-4 receptor, preventing engagement of CTLA-4 on the T cell, preventing costimulation	Central nervous system post-transplantation lymphoproliferative disease
T cell–depleting antibodies	Thymoglobulin Alemtuzumab	Induction, antirejection Induction	Rabbit-derived polyclonal antilymphocyte serum (multiple targets) Anti-CD52	Infusion reaction, leukopenia, thrombocytopenia
Anti-IL2R monoclonal antibodies	Basiliximab	Induction	Block interleukin-2 receptor	Rare infusion reaction

The goals of induction therapy include a reduction in the number of acute rejection episodes in the first several weeks or months after transplantation, a long-term reduction in the overall number of acute rejection episodes, and a resulting improvement in long-term success rates.

Rejection and Its Treatment

Despite the availability of more effective and safer immunosuppressive medications, rejection remains an ever-present threat, and a significant proportion of allograft failures are ultimately caused by immune-mediated injury. Classically, there are three types of rejection: hyperacute, acute, and chronic.

Hyperacute rejection occurs when a kidney is transplanted into a recipient who is presensitized to the donor, that is, has preformed, circulating antibodies reactive against antigens expressed on the donor kidney. Antidonor antibodies causing hyperacute rejection may target donor HLA antigens, blood group (ABO) antigens, or endothelial cell antigens. The result is immediate endothelial injury and irreversible thrombosis of the transplant. Because of the prescreening of all donor-recipient pairs for the presence of such antidonor antibodies, hyperacute rejection occurs in less than 1% of all renal transplants.

Acute rejection is characterized clinically by an increase in the serum creatinine level during a period of days to weeks, most often in the absence of any symptoms. On pathologic examination, acute rejection is most frequently "cellular" and characterized by T-lymphocyte infiltration of the tubules, of the interstitium, and sometimes into the vascular structures in more severe rejection. Acute rejection now occurs in only 10 to 15% of transplant patients during the first post-transplantation year and in only about another 10% in the second year. The majority of acute rejections occur in the first 3 to 6 months after transplantation and are clinically mild, treatable, and reversible. Other causes of allograft dysfunction (e.g., hemodynamic factors such as hypotension and volume depletion, urinary tract obstruction, drug-induced nephrotoxicity, and BK polyomavirus nephropathy) must be excluded. Percutaneous biopsy of the allograft is required to establish the diagnosis with certainty, to determine the type and severity of the rejection, and thereby to guide therapeutic decisions. Acute cellular rejection should be treated immediately, either with high-dose intravenous corticosteroid therapy for milder forms or with polyclonal antilymphocyte sera for steroid-resistant or more severe acute rejection episodes. Such treatment is usually effective.

About 10 to 20% of patients have acute antibody-mediated rejection injury, usually as evidenced by the deposition of C4d, which is a metabolite of a complement component C4, in the peritubular capillaries, accompanied by inflammatory cells in the peritubular capillaries. This antibody-mediated rejection may occur[11] alone or in combination with cellular rejection. Early antibody-mediated rejection is much more common in patients with a prior history of exposure to alloantigen through transfusion, pregnancy, or prior transplantation. The development of sensitive reagents for detection of circulating antidonor antibodies has greatly facilitated the diagnosis of antibody-mediated rejection. Treatment for antibody-mediated rejection is more problematic and usually involves plasmapheresis to remove antidonor antibodies and the administration of various regimens of intravenous immune globulin. Although this treatment is often effective in reversing acute antibody-mediated rejection, the successful elimination of antidonor antibody and the prevention of alloantibody resynthesis are much more difficult, and chronic antibody-mediated rejection frequently follows an episode of acute antibody-mediated rejection. Anti–B cell and anti–plasma cell drugs have also been tried, but none are currently approved by the Food and Drug Administration for this use. High-dose corticosteroids and polyclonal antilymphocyte serum are often administered as well.

Although *chronic rejection* can occur in the absence of any documented episodes of acute rejection and after many years of stable allograft function, early acute rejection is a major risk factor for later chronic rejection and allograft failure. Chronic rejection is characterized by a slow decline in allograft function, generally during a period of months to years, and frequently accompanied by proteinuria. The development of donor-specific antibodies appears to be a strong risk factor for subsequent allograft failure, and antidonor antibodies are detectable in a high percentage of cases of chronic allograft dysfunction. Biopsy characteristically shows evidence of T cell–induced injury, antibody-induced injury with deposits of C4d, or a combination of the two. In many patients with long-term deterioration of allograft function, however, the only findings are interstitial fibrosis and tubular atrophy without clear-cut evidence as to the cause. Unfortunately, no treatment has been proved to be effective for chronic rejection.

Other Complications of Renal Transplantation

Immunosuppression in renal transplant recipients is associated with a number of adverse effects, the most important of which are infection, malignant neoplasia, and nephrotoxicity.

Infection

Infections occurring in the first few weeks after transplantation are generally nosocomial (Chapter 282), donor derived, present in the recipient at the time of transplantation, or a result of complications of the surgical procedure, such as wound, catheter-associated, or urinary tract infections (Chapter 284) (Fig. 131-3).[12] Next is a period of 5 to 6 months during which opportunistic infection or reactivation of latent infection, such as cytomegalovirus (CMV) infection (Chapter 370) or Epstein-Barr virus (EBV) infection (Chapter 370), is most likely to occur. Later in the course, when the doses of immunosuppressive medication are decreased, community-acquired infections are most common. This timeline is partly a function of infection prophylaxis, which is given to all renal transplant recipients to minimize infectious complications during periods of highest risk, including perioperative antibiotics to prevent wound infection and urinary tract infection (Chapter 284), trimethoprim-sulfamethoxazole to prevent *Pneumocystis jiroveci* pneumonia (Chapter 341), oral nystatin or clotrimazole to prevent oral candidiasis (Chapter 338), and oral valganciclovir (see Table 360-4) to prevent CMV disease (Chapter 376). Urinary tract infections (Chapter 284) are the most frequent bacterial infections that lead to hospitalizations in kidney transplant recipients, and the most common organism is *E. coli*.

Up to 30% of patients develop active infection with the BK polyomavirus,[13] which is tropic for the urinary epithelia. In 5 to 10% of patients, this infection leads to tubulitis and interstitial nephritis, which may result in significant allograft injury. BK polyoma nephritis is often difficult to distinguish from acute rejection without the use of a special SV40 stain that identifies the large T antigen of BK virus. Approximately 80 to 90% of the general population has serologic evidence of exposure to BK polyomavirus, so infection in the renal transplant recipient may be due to reactivation, superinfection, or primary infection. Interestingly, BK nephropathy is very rare in equally immunosuppressed recipients of other transplants. The most important noninvasive indicator of possible BK viral nephropathy is BK viremia, which can be documented in about 10 to 20% of renal transplant recipients. It is presently estimated that 5 to 10% of all renal transplant recipients develop BK virus–induced nephritis and allograft dysfunction, which leads to allograft failure in about half of the patients with nephritis. Monitoring for BK viremia or viruria is strongly recommended. BK viremia and BK viral nephropathy are treated initially with a reduction in immunosuppression, but effective antiviral therapy has yet to be developed.

CMV infection (Chapter 370) continues to be a major cause of morbidity in renal transplant recipients. Like BK polyomavirus, CMV is widely distributed in the general population, with seroprevalence rates ranging from 40 to 90%, depending on the population, so infections may be due to reactivation, superinfection, or primary infection. CMV disease may be limited to fever and malaise, but it frequently causes significant morbidity and even mortality after infecting the gastrointestinal tract, the lungs, or the liver. Most transplant centers use oral valganciclovir (see Table 360-4) for CMV prophylaxis for 7 months or more after transplantation for all patients at risk of CMV disease, especially seronegative recipients who receive transplants from seropositive donors. Another alternative is frequent post-transplantation monitoring, with treatment reserved only for patients who develop viremia. Regardless of the initial strategy, approximately 10% of patients have later viral activation and require additional treatment. Antiviral therapy (Chapters 360 and 370) is usually highly effective.

EBV infection (Chapter 370) also may represent a primary infection, reactivation, or superinfection. EBV may cause a mononucleosis-like syndrome, but the more serious concern in renal transplant patients is a form of lymphoma termed post-transplantation lymphoproliferative disease (Chapters 370 and 185).

Malignant Neoplasia

Renal transplant patients develop some malignant neoplasms at higher rates than in the general population. These include post-transplantation lymphoproliferative disorder related to EBV, Kaposi sarcoma (Chapter 392) caused by human herpesvirus 8 (Chapter 374), and cervical (Chapter 199) and anal (Chapter 145) carcinomas caused by human papillomavirus (Chapter 373). Post-transplantation lymphoproliferative disease most commonly involves

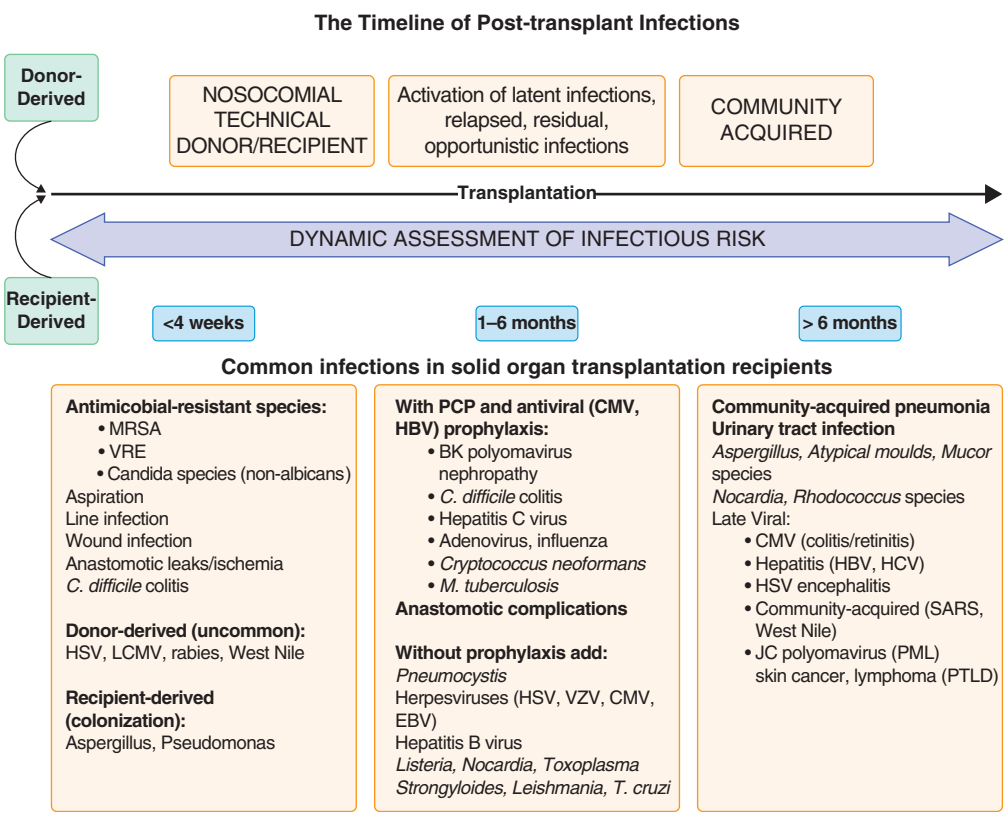

FIGURE 131-3. Timeline of post-transplantation infections. CMV = cytomegalovirus; EBV = Epstein-Barr virus; LCMV = lymphocytic choriomeningitis virus; MRSA = methicillin-resistant *Staphylococcus aureus*; PCP = *Pneumocystis* pneumonia; HBV = hepatitis B virus; HCV = hepatitis C virus; HSV = herpes simplex virus; PML = progressive multifocal leukoencephalopathy; PTLD = post-transplantation lymphoproliferative disease; SARS = severe acute respiratory syndrome; VRE = vancomycin-resistant enterococci; VZV = varicella-zoster virus. (From Fishman JA. Introduction: infection in solid organ transplant recipients. *Am J Transplant.* 2009;9(Suppl 4):S3-6.)

the renal allograft, but it may involve any organ, including the gastrointestinal tract, lung, and central nervous system.[14] Treatment usually involves a reduction in immunosuppression along with chemotherapy (Chapter 185). Long-term patient survival rates approach 75%.

The relative risks of many common solid carcinomas are also elevated about 2-fold, but with the exception of nonmelanoma skin cancers, the absolute risks remain low.[15] For nonmelanoma skin cancers (Chapter 203), the lifetime risk may be as high as 70%. Annual skin screening and the attentive use of ultraviolet-blocking skin creams are strongly recommended.[16] Otherwise, transplant recipients should follow the same standard guidelines as the general population for screening for breast (Chapter 198), colon (Chapter 193), prostate (Chapter 201), and lung (Chapter 191) cancer. Whether immunosuppression should be reduced in patients who develop malignant neoplasms other than post-transplantation lymphoproliferative disease is unknown.

Calcineurin Inhibitor–Induced Nephrotoxicity

Calcineurin inhibitors (cyclosporine and tacrolimus)—arguably the most effective antirejection medications used in kidney transplantation—ironically cause a chronic, usually mild and stable, impairment of GFR in virtually all patients who receive them. Superimposed, acute reversible cyclosporine or tacrolimus nephrotoxicity may occur in patients who experience sudden elevations in drug blood levels. These acute episodes usually resolve when the dose of cyclosporine or tacrolimus is reduced. Of greater concern is long-term, irreversible calcineurin inhibitor–induced nephrotoxicity, which can cause allograft failure. This diagnosis is often difficult to make because there are few if any absolutely specific histologic biopsy findings. As a result, careful monitoring of blood calcineurin inhibitor levels is essential to avoid suprathrapeutic levels (risking nephrotoxicity and infection) or subtherapeutic levels (risking rejection).

New-Onset Diabetes

By 36 months after transplantation, between 30 and 40% of patients develop glucose intolerance or frank diabetes, which may be caused by corticosteroids, tacrolimus, and, to a lesser extent, cyclosporine. Treatment is as for diabetes in the general population (Chapter 229), with the understanding that doses of insulin and oral agents may need to be adjusted for renal function. Patients who develop diabetes have poorer outcomes, including higher rates of cardiovascular events, graft failure, and death.[17]

Management of the Patient after Renal Transplantation

In the initial few months after transplantation, patients are generally seen weekly. The interval then gradually expands to once monthly for the remainder of the first year and every 2 to 6 months thereafter, depending on the patient. Routine testing includes measurement of blood pressure, renal function, calcineurin inhibitor blood levels, lipid levels, glucose levels, BK polyomavirus assays, and proteinuria. Polymerase chain reaction analysis for CMV or EBV is indicated in selected patients. Monitoring for metabolic bone disease is also recommended, by parathyroid hormone and vitamin D levels as well as by bone densitometry. Influenza and pneumococcal vaccinations are strongly recommended for renal transplant recipients, but live vaccines are contraindicated (Chapter 18).

PROGNOSIS

Approximately 6% of long-term transplants fail each year. Allograft failure with return to dialysis or retransplantation accounts for about 50% of long-term allograft failures. Progressive loss of function in the long-term renal allograft is most commonly the result of chronic rejection but may also be due to nonimmunologic causes, such as calcineurin inhibitor–induced nephrotoxicity, hypertension, diabetes, or recurrent glomerular disease (Table 131-3). A standardized evaluation (Table 131-4) can help establish the cause.

The importance of immunologic (or alloantigen-dependent) factors is illustrated by the fact that living donor transplants between HLA-identical siblings have a 70 to 80% 10-year graft survival, compared with 50 to 60% for parental or other less well matched living donor kidneys and 40 to 50% overall for cadaveric kidney recipients. Deceased donor kidneys with zero and one HLA-A, B, or DR mismatchs have a higher success rate (close to 65% at 10 years) than do deceased donor kidneys with two or more antigen mismatches, which have 15 to 30% lower absolute survival rates.

TABLE 131-3 COMMON CAUSES OF RENAL ALLOGRAFT DYSFUNCTION

Volume depletion: Nausea, vomiting, diarrhea, poor fluid intake, hemorrhage
Medication induced: Diuretics, antihypertensive medications, nonsteroidal anti-inflammatory drugs, calcineurin inhibitors, angiotensin-converting enzyme inhibitors
Urinary tract obstruction: Bladder outlet obstruction, ureteral obstruction
Transplant renal artery stenosis
Infection: Bacteremia, urinary tract infection
Rejection: Cell mediated, antibody mediated

TABLE 131-4 EVALUATION OF RENAL ALLOGRAFT DYSFUNCTION

History: Fluid intake and loss, medication changes, fever
Physical examination: Blood pressure, pulse rate, temperature, weight, edema, allograft tenderness or swelling
Laboratory tests: Serum creatinine level, BK virus detection by polymerase chain reaction analysis of urine or blood, calcineurin inhibitor levels, antidonor antibody testing, urine protein analysis, urine and blood cultures
Doppler ultrasound to assess for possible urinary tract obstruction or transplant renal artery stenosis
Renal biopsy

Factors unrelated to antiallograft immunity are also associated with reduced long-term success rates. These characteristics include hypertension, a kidney from an older (>50 years) donor, hyperlipidemia, diabetes mellitus, calcineurin inhibitor nephrotoxicity, recurrent glomerular disease, and a maladaptive response to hemodynamic injury, which is caused by elevated glomerular flow and pressure. Chronic calcineurin inhibitor toxicity leads to ESRD in at least 5 to 10% of recipients of nonrenal transplants and is likely to have at least a comparable long-term impact in renal transplant patients. Virtually all primary glomerular diseases (Chapter 121), including focal segmental glomerulosclerosis, membranous nephropathy, membranoproliferative glomerulonephritis, and immunoglobulin A nephropathy, can recur after transplantation. Recurrent glomerulonephritis is a common cause of proteinuria and nephrotic syndrome after transplantation, and recurrent glomerulonephritis may be responsible for up to 10% of all graft failures.[18] Renal biopsy is required to establish the diagnosis in a long-term renal allograft with deteriorating function or proteinuria and to distinguish immune-related from non–immune-related causes. For long-term allograft survival rates to improve significantly, the discovery of biomarkers of actual or impending allograft injury is essential.[19]

In the other 50% of failures, a patient dies with a functioning allograft. The most common cause of death in these patients is cardiovascular, due to the high prevalence of diabetes, hypertension, and hyperlipidemia as well as the high burden of preexisting cardiovascular disease in transplant candidates. Although there are no specific guidelines for the prevention, monitoring, or management of cardiovascular disease after renal transplantation, standard risk factor reduction guidelines are generally considered applicable, including the use of statins (Chapter 206)[20] and the use of antihypertensive medications to meet blood pressure targets (Chapter 67).[21] The other leading causes of long-term death in renal transplant recipients are infection, sepsis, and malignant disease.

Grade A References

A1. Cooper BA, Branley P, Bulfone L, et al. A randomized, controlled trial of early versus late initiation of dialysis. *N Engl J Med*. 2010;363:609-619.
A2. Eknoyan G, Beck GJ, Cheung AK, et al. Effect of dialysis dose and membrane flux in maintenance hemodialysis. *N Engl J Med*. 2002;347:2010-2019.
A3. Chertow GM, Levin NW, Beck GJ, et al. In-center hemodialysis six times per week versus three times per week. *N Engl J Med*. 2010;363:2287-2300.
A4. Rocco MV, Lockridge RS Jr, Beck GJ, et al. The effects of frequent nocturnal home hemodialysis: the Frequent Hemodialysis Network Nocturnal Trial. *Kidney Int*. 2011;80:1080-1091.
A5. Maduell F, Moreso F, Pons M, et al. High-efficiency postdilution online hemodiafiltration reduces all-cause mortality in hemodialysis patients. *J Am Soc Nephrol*. 2013;24:487-497.
A6. Paniagua R, Amato D, Vonesh E, et al. Effects of increased peritoneal clearances on mortality rates in peritoneal dialysis: ADEMEX, a prospective, randomized, controlled trial. *J Am Soc Nephrol*. 2002;13:1307-1320.
A7. Drueke TB, Locatelli F, Clyne N, et al. Normalization of hemoglobin level in patients with chronic kidney disease and anemia. *N Engl J Med*. 2006;355:2071-2084.
A8. Singh AK, Szczech L, Tang KL, et al. Correction of anemia with epoetin alfa in chronic kidney disease. *N Engl J Med*. 2006;355:2085-2098.
A9. Pfeffer MA, Burdmann EA, Chen CY, et al. A trial of darbepoetin alfa in type 2 diabetes and chronic kidney disease. *N Engl J Med*. 2009;361:2019-2032.
A10. Coyne DW, Kapoian T, Suki W, et al. Ferric gluconate is highly efficacious in anemic hemodialysis patients with high serum ferritin and low transferrin saturation: results of the Dialysis Patients' Response to IV Iron with Elevated Ferritin (DRIVE) Study. *J Am Soc Nephrol*. 2007;18:975-984.
A11. Parfrey PS, Chertow GM, Block GA, et al. The clinical course of treated hyperparathyroidism among patients receiving hemodialysis and the effect of cinacalcet: the EVOLVE trial. *J Clin Endocrinol Metab*. 2013;98:4834-4844.
A12. Chertow GM, Block GA, Correa-Rotter R, et al. Effect of cinacalcet on cardiovascular disease in patients undergoing dialysis. *N Engl J Med*. 2012;367:2482-2494.
A13. Palmer SC, Navaneethan SD, Craig JC, et al. HMG CoA reductase inhibitors (statins) for dialysis patients. *Cochrane Database Syst Rev*. 2013;9:CD004289.
A14. Matsumoto Y, Mori Y, Kageyama S, et al. Spironolactone reduces cardiovascular and cerebrovascular morbidity and mortality in hemodialysis patients. *J Am Coll Cardiol*. 2014;63:528-536.
A15. Woodle ES, First MR, Pirsch J, et al. A prospective, randomized, double-blind, placebo-controlled multicenter trial comparing early (7 day) corticosteroid cessation versus long-term, low-dose corticosteroid therapy. *Ann Surg*. 2008;248:564-577.
A16. Palmer SC, Navaneethan SD, Craig JC, et al. HMG CoA reductase inhibitors (statins) for kidney transplant recipients. *Cochrane Database Syst Rev*. 2014;1:CD005019.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. *United States Renal Data System (USRDS) 2014 Annual Data Report*. http://www.usrds.org/adr.aspx. Accessed November 20, 2014.
2. Himmelfarb J, Ikizler TA. Hemodialysis. *N Engl J Med*. 2010;363:1833-1845.
3. Daugirdas JT, Leypoldt JK, Akonur A, et al. Improved equation for estimating single-pool Kt/V at higher dialysis frequencies. *Nephrol Dial Transplant*. 2013;28:2156-2160.
4. McCausland FR, Brunelli SM, Waikar SS. Dialysis dose and intradialytic hypotension: results from the HEMO study. *Am J Nephrol*. 2013;38:388-396.
5. Gaweda AE, Aronoff GR, Jacobs AA, et al. Individualized anemia management reduces hemoglobin variability in hemodialysis patients. *J Am Soc Nephrol*. 2014;25:159-166.
6. *Kidney Disease Improving Global Outcomes (KDIGO)*. www.kdigo.org. Accessed November 20, 2014.
7. Hill CJ, Maxwell AP, Cardwell CR, et al. Glycated hemoglobin and risk of death in diabetic patients treated with hemodialysis: a meta-analysis. *Am J Kidney Dis*. 2014;63:84-94.
8. Muzaale AD, Massie AB, Wang MC, et al. Risk of end-stage renal disease following live kidney donation. *JAMA*. 2014;311:579-586.
9. Lentine KL, Costa SP, Weir MR, et al. Cardiac disease evaluation and management among kidney and liver transplantation candidates: a scientific statement from the American Heart Association and the American College of Cardiology Foundation: endorsed by the American Society of Transplant Surgeons, American Society of Transplantation, and National Kidney Foundation. *Circulation*. 2012;126:617-663.
10. Organ Procurement and Transplantation Network. *Kidney Allocation System*. http://optn.transplant.hrsa.gov/learn/professional-education/kidney-allocation-system/. Accessed November 20, 2014.
11. Mengel M, Sis B, Haas M, et al. Banff 2011 Meeting report: new concepts in antibody-mediated rejection. *Am J Transplant*. 2012;12:563-570.
12. The American Society of Transplantation Infectious Diseases Guidelines 3rd Edition. *Am J Transplant*. 2013;13(special issue):1-371.
13. Masutani K. Current problems in screening, diagnosis and treatment of polyomavirus BK nephropathy. *Nephrology (Carlton)*. 2014;19(suppl 3):11-16.
14. Green M, Michaels MG. Epstein-Barr virus infection and posttransplant lymphoproliferative disorder. *Am J Transplant*. 2013;13(suppl 3):41-54.
15. Sampaio MS, Cho YW, Qazi Y, et al. Posttransplant malignancies in solid organ adult recipients: an analysis of the U.S. National Transplant Database. *Transplantation*. 2012;94:990-998.
16. Harwood CA, Mesher D, McGregor JM, et al. A surveillance model for skin cancer in organ transplant recipients: a 22-year prospective study in an ethnically diverse population. *Am J Transplant*. 2013;13:119-129.
17. Sarno G, Muscogiuri G, De Rosa P. New-onset diabetes after kidney transplantation: prevalence, risk factors, and management. *Transplantation*. 2012;93:1189-1195.
18. Ponticelli C, Glassock RJ. Posttransplant recurrence of primary glomerulonephritis. *Clin J Am Soc Nephrol*. 2010;5:2363-2372.
19. Suthanthiran M, Schwartz JE, Ding R, et al. Urinary-cell mRNA profile and acute cellular rejection in kidney allografts. *N Engl J Med*. 2013;369:20-31.
20. Wanner C, Tonelli M. KDIGO Clinical Practice Guideline for Lipid Management in CKD: summary of recommendation statements and clinical approach to the patient. *Kidney Int*. 2014;85:1303-1309.
21. James PA, Oparil S, Carter BL, et al. 2014 Evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014;311:507-520.

132 APPROACH TO THE PATIENT WITH GASTROINTESTINAL DISEASE

KENNETH R. MCQUAID

The luminal gastrointestinal (GI) tract (esophagus, stomach, duodenum, small and large intestine, and anus) and pancreas are responsible for digestion, for the absorption of nutrients and fluids, and for the temporary storage and excretion of undigested waste. The GI tract has an epithelial lining with an enormous surface area that provides nutrient absorption and serves as a barrier to microorganisms. In addition, the GI tract has a large innate and adaptive immune system that interfaces with luminal food antigens, host proteins, commensal and pathogenic bacteria, and parasites and must decide which antigens to tolerate and which require immune activation.[1] The GI tract also contains an extensive enteric endocrine system that regulates food intake, weight control, and glucose homeostasis as well as secretions from the stomach, intestine, and pancreas. Finally, it has an enteric nervous system that is integrated with the autonomic and central nervous systems to control gastric emptying, intestinal motility, and defecation.[2]

Numerous diseases within and outside the GI tract may alter normal function by causing structural damage (erosion, ulceration, perforation, stenosis, or obstruction), bleeding, inflammation, abnormal absorption or secretion of nutrients and electrolytes, or abnormal motility. Despite its anatomic and physiologic complexity, the GI system has only a limited repertoire of symptoms and signs to express conditions that may be either serious or clinically insignificant: abdominal pain, heartburn, regurgitation, dysphagia, odynophagia, dyspepsia, nausea and vomiting, gas and bloating, weight loss, diarrhea, constipation, overt or occult GI tract bleeding, and incontinence.

GENERAL APPROACH TO PATIENTS WITH GASTROINTESTINAL SIGNS AND SYMPTOMS

An appropriate history and physical evaluation usually can narrow the differential diagnosis of GI complaints. A specific diagnosis can almost always be established thereafter by the judicious use of laboratory, endoscopic, or imaging studies (Table 132-1).

Clinical History

The clinician should elicit the nature of the complaint, including its acuity, severity, location, radiation, duration, pattern (steady vs. colicky; abrupt vs. gradual onset), and relationship to food, meals, and bowel movements. Symptoms that arise from the GI tract are almost always improved or worsened by eating or by bowel movements. For symptoms of recent onset, it is important to elicit recent dietary intake, a medication history, potential exposure to enteric infections or sexually transmitted diseases (Chapter 285), and recent travel. It is also useful to establish whether there are signs or symptoms that suggest a systemic illness, including fever, weight loss, arthralgias, fatigue, weakness, and rash.

Most nonsurgical GI diseases are manifested with mild to moderate symptoms that develop gradually and do not require immediate attention. Acute symptoms that require urgent assessment are severe abdominal pain and overt GI bleeding (Chapter 135) that is manifested by hematemesis, melena, or large-volume hematochezia. Severe or dramatic abdominal pain that develops acutely during minutes to hours requires urgent evaluation to determine whether surgical intervention is required. Severe vomiting or diarrhea with signs of dehydration also warrants urgent attention.

Mild to moderate chronic or intermittent symptoms that have been present for a long period can be evaluated in a deliberate fashion. A substantial proportion of chronic GI complaints has no obvious organic or biochemical basis and ultimately is classified as *functional* GI disorders (Chapter 137). Complaints that have been ongoing for years rarely are attributable to readily remedied structural disorders. In patients with chronic GI symptoms, it is important to elicit and to address the current reason for seeking evaluation, which may include concern for underlying serious illness (especially cancer), a change in the character or severity of symptoms, life stressors, and depression. Asking the patient what he or she thinks or fears may provide insights into the proportion of the complaint attributable to these amplifying issues, regardless of whether the problem is functional or structural in origin.

A dietary history (Chapter 214) should be obtained. For acute symptoms of nausea, vomiting, diarrhea, or abdominal pain, intake during the previous 24 to 48 hours should be reviewed for clues to a food-borne illness, including possible exposure to a contaminated food or water source and similar symptoms in other people (Chapter 283). For chronic or intermittent complaints, a recall of meals and types of foods eaten during the previous 1 to 2 days provides insight into eating habits and the amounts and types of fruits and vegetables, whole grains, fiber, protein, fat, and dairy products ingested. A relationship between specific foods and symptoms may be found. For example, dairy products (lactose intolerance), whole grains, legumes or cruciferous vegetables, or fatty meals (malabsorption) may cause pain, flatulence, or diarrhea, and a low-fiber diet may cause chronic constipation. Recent and long-term changes in body weight should be elicited. Involuntary loss of more than 5% of body weight during the prior 12 months is worrisome for serious disease and significant malnutrition (Chapter 215).

The number and consistency of bowel movements should be elicited, and any change in bowel habits must be explored. Signs of acute GI bleeding (melena or hematochezia) or inflammatory colitis (blood, mucus, or pus) should be elicited. Improvement in symptoms after passage of flatus or a bowel movement suggests a disorder of the colon or anorectum.

Past Medical History

The past medical history should be reviewed for conditions that may cause acute or chronic GI symptoms, including endocrine disorders such as diabetes (Chapter 229) and thyroid dysfunction (Chapter 226), cardiovascular diseases such as heart failure (Chapter 58) and peripheral vascular disease (Chapter 79), chronic liver disease and portal hypertension (Chapter 153), neurologic conditions such as Parkinson disease (Chapter 409) and neuromuscular disorders (Chapter 396), and rheumatologic and collagen vascular disorders (Chapter 256). In addition to their impact on GI tract function, the severity of these conditions must be considered in weighing the risks of diagnostic studies, especially endoscopy. Patients with symptomatic or advanced respiratory insufficiency (Chapter 104), sleep apnea (Chapter 100), valvular heart disease (Chapter 75), coronary artery disease (Chapter 51), heart failure (Chapter 58), cirrhosis (Chapter 153), cerebrovascular disease (Chapter 406), neuromuscular disease (Chapter 422), or dementia (Chapter 402) have an increased risk of sedation-related complications during endoscopy.

A list of prescription and nonprescription medications, vitamins, minerals, and other nutritional supplements should be obtained, paying particular attention to any that were recently initiated or changed. Herbal supplements (Chapter 39) are commonly used but are seldom reported without direct questioning. Medications are potential causes of odynophagia, dyspepsia, nausea or vomiting, abdominal pain, diarrhea, and constipation. The use of antiplatelet agents, including aspirin and anticoagulants, should be determined. The risks of stopping versus continuing these medications must be weighed in patients who have acute or chronic GI bleeding or in whom a therapeutic procedure is to be performed.[3]

Social History

The patient's personal relationships, employment history, quality of life, alcohol intake (Chapter 33), and smoking (Chapter 32) history should be determined. It can be very informative to observe both verbal and nonverbal interactions between the patient and a partner or caregiver during an interview. Alcohol may cause heartburn, dyspepsia, nausea, diarrhea, or chronic liver disease. Many patients are reluctant to disclose the full extent of their alcohol intake on direct questioning; therefore, in addition to asking how often they imbibe (days/week and drinks/day), it may be revealing to inquire about their preferred beverage and how it is purchased (location, volume, and frequency). Cigarette smoking is associated with an increased risk of heartburn, peptic ulcer disease, Crohn disease, and GI malignant neoplasms.

Clinicians should inquire about the degree to which GI symptoms are disrupting a patient's life. GI illness may affect dietary and bowel habits, sleep, and sense of vitality. Concerns about dietary intolerances, inability to eat, inability to have comfortable bowel movements, uncontrolled diarrhea or gas, fecal urgency, or fecal incontinence may affect a patient's social life, personal and sexual relationships, employment, and sense of optimism.

The social history should also be reviewed for recent stressors that may precipitate or exacerbate GI symptoms, including marital or interpersonal discord, personal or family illness, bereavement, financial pressures, job loss, and change in employment. To elicit such information, it may be helpful to

tell patients that stress worsens many conditions and to inquire whether they believe stress may be contributing to their problem.

For elderly, disabled, or marginally housed patients, it is important to elicit how they obtain and prepare their meals and how they access toilet facilities. For patients undergoing GI procedures, it is important to determine whether they have mental, physical, or social barriers that would make it difficult to comply with preprocedure instructions (including bowel preparation) and whether they have an able-bodied adult who can accompany them to the procedure and observe them at home, if necessary, afterward.

Family History

The family history should be reviewed for GI disorders with a heritable component, especially celiac disease (Chapter 140), inflammatory bowel diseases (Chapter 141), and neoplasms of the GI, gynecologic, and genitourinary tracts.

Physical Examination

Nonabdominal Examination

The nonabdominal examination should assess nutritional status (Chapter 214) and any signs of systemic conditions that may cause GI symptoms or that must be considered in weighing the risks and benefits of further testing, especially endoscopy. Vital signs should be obtained in all patients. Low-grade fever (<100.5° F) is common in inflammatory conditions, including gastroenteritis, inflammatory bowel disease, appendicitis, cholecystitis, and diverticulitis. High fever (>102° F) suggests sepsis, pelvic or intra-abdominal infections (e.g., cholangitis, pelvic inflammatory disease, pyelonephritis), or peritonitis. Hemodynamic instability (hypotension or tachycardia) suggests intravascular depletion due to poor oral intake, acute GI or intra-abdominal bleeding, severe diarrhea, or peritonitis. A body mass index of less than 18 suggests malnourishment.

A general survey should be performed to assess for signs of weight loss (fat and muscle wasting), malnutrition (dry or thin skin, hair loss, edema, anasarca), and vitamin deficiencies (pellagra, scurvy). Skin lesions may provide clues to systemic conditions such as liver disease (jaundice, spider telangiectasias, palmar erythema), inflammatory bowel disease (erythema nodosum, pyoderma gangrenosum), celiac disease (dermatitis herpetiformis), vasculitis, and rare GI malignant neoplasms, polyposis syndromes, and pancreatic endocrine tumors (Chapters 192, 193, and 195). An oral examination looks for mucocutaneous candidiasis (which may reflect immunosuppression), ulcerations (which may reflect inflammatory bowel disease, vasculitis, viral infection, or vitamin deficiencies), and glossitis or angular cheilitis (which may reflect vitamin deficiencies). With the exception of supraclavicular lymph nodes, peripheral lymph nodes are uninvolved with GI diseases but should be examined when systemic infection or advanced malignant disease is suspected (Chapter 168). Examination of the lungs and cardiovascular system should focus on evidence of conditions that might increase the risk of moderate sedation in the event that endoscopy is required (respiratory insufficiency, heart failure) and of conditions that increase the risk of intestinal ischemia (atrial fibrillation, valvular heart disease, peripheral vascular disease) (Chapter 143). The extremities should be evaluated for edema and peripheral pulses. Finally, a brief neurologic assessment should be performed to screen for intracranial mass lesions or other neurologic disorders that may be manifested with GI symptoms.

Abdominal Examination

The abdominal examination begins with a visual inspection of the abdomen and inguinal region for scars (due to prior surgeries or trauma), asymmetry (suggesting a mass or organomegaly), distention (due to obesity, ascites, or intestinal ileus or obstruction), prominent periumbilical veins (suggesting portal hypertension), or hernias (umbilical, ventral, inguinal). The examination proceeds with auscultation followed by percussion, and it ends with light and deep palpation.

In patients *without* abdominal pain, auscultation of bowel sounds to assess intestinal motility has limited usefulness and may be omitted. Percussion may be performed before or in conjunction with light and deep palpation. Initial cursory light percussion across the upper, mid, and lower abdomen is useful to denote areas of dullness and tympany as well as to elicit unanticipated areas of pain or tenderness before palpation. More extensive percussion provides limited but useful information about the size of the liver and spleen, gastric or intestinal distention, bladder distention, and ascites (Chapters 146 and 153). Gentle, light palpation promotes abdominal relaxation and allows the detection of muscle resistance (guarding), abdominal tenderness, and superficial masses of the abdominal wall or abdomen. Deeper palpation of the abdominal organs (liver, spleen, kidneys, aorta) and abdominal cavity may detect enlargement or abnormal masses. Superficial or deep masses should be assessed for size, location, mobility, content (solid, liquid, or air), and the presence or absence of tenderness. The consistency of a patient's response to palpation with and without distraction is particularly useful in those with suspected chronic functional abdominal discomfort. Superficial masses include hernias, lymph nodes, subcutaneous abscesses, lipomas, and hematomas. Neoplasms (liver, gallbladder, pancreas, stomach, intestine, kidney), abscesses (appendicitis, diverticulitis, Crohn disease), or aortic aneurysms may represent deep abdominal masses.

Examination of the right upper quadrant should assess the liver size, contour, texture, and tenderness. Liver size is crudely estimated by percussion of the upper and lower borders of liver dullness in the midclavicular line. Liver contour and tenderness are best assessed during held inspiration by deep palpation along the costal margin. Examination of the left upper quadrant is useful to detect splenomegaly (Chapter 168), although a normal-sized or even an enlarged spleen often cannot be detected. Percussion in the left upper quadrant near the tenth rib (posterior to the midaxillary line) may detect splenic dullness that is distinct from gastric or colonic tympany. The tip of an enlarged spleen may be palpated during inspiration if the examiner supports the left costal margin with the left hand while palpating below the costal margin with the right hand. Ascites should be suspected in a patient with a protuberant abdomen and bulging flanks. To screen for ascites, percussion of the flanks should be performed to assess the level of dullness. If the level of flank dullness appears to be increased, the most sensitive test for ascites is to check for "shifting" dullness when the patient rolls from the supine to the lateral position.

Digital Rectal and Pelvic Examinations

The digital rectal examination is intrusive and uncomfortable and should be performed only when necessary, such as in patients with perianal or rectal symptoms, incontinence, difficult defecation, suspected inflammatory bowel disease, and acute abdominal pain. The digital examination, with or without fecal occult blood testing, is not a useful screening test for colorectal cancer (Chapter 193). However, in patients with acute or chronic GI bleeding (Chapter 135), it is a rapid means of assessing the stool for color and occult blood. The perianal area should be visually inspected for rashes, soilage (suggesting incontinence or fistula), fistulas, fissures, skin tags, external hemorrhoids, and prolapsed internal hemorrhoids (Chapter 145). After gentle digital insertion, the anal canal should be assessed for resting tone and voluntary squeeze. The distal rectal vault should be swept circumferentially to palpate for mass lesions, tenderness, or fluctuance.

Laboratory Studies

Blood Tests

Blood tests routinely obtained in the evaluation of patients with GI symptoms include a complete blood count, liver tests (Chapter 147), serum chemistries, and, in selected cases, pancreatic enzymes and markers of inflammation. GI causes of anemia include acute or chronic GI blood loss, inflammatory bowel disease, nutrient malabsorption (folate, iron, or vitamin B_{12}), and chronic liver disease. Microcytosis suggests iron deficiency due to chronic GI blood loss or malabsorption. Macrocytosis may be attributable to folate or B_{12} malabsorption, medications (e.g., immunomodulators used for inflammatory bowel disease), or chronic liver disease. An elevated platelet count suggests chronic inflammation (e.g., inflammatory bowel disease) or GI blood loss with compensatory marrow production. A low platelet count may be attributable to portal hypertension with splenic sequestration. Low serum albumin may be caused by chronic GI disorders that result in weight loss, nutrient malabsorption, chronic inflammation, loss of protein across abnormal GI mucosa (i.e., protein-losing enteropathy), or decreased hepatic synthesis (e.g., chronic liver disease). Abnormal liver test results may be due to acute or chronic liver diseases, disorders of the pancreas or biliary tract, and medications (Chapter 147). Serum amylase and lipase are obtained to screen for pancreatitis (Chapter 144) in patients with acute abdominal pain. Increased levels of inflammatory markers, such as an elevated erythrocyte sedimentation rate and C-reactive protein, are nonspecific but useful in the management of patients with inflammatory bowel disease (Chapter 141).

Serum ferritin reflects total body iron and may be decreased in patients with chronic GI blood loss or intestinal malabsorption (e.g., celiac disease). Deficiencies in the fat-soluble vitamins (A, D, E, K) (Chapter 140) may reflect disorders of malabsorption that result in steatorrhea. The serum

TABLE 132-1 APPROACH TO COMMON GASTROINTESTINAL SYMPTOMS AND SIGNS

	ABDOMINAL PAIN	GI BLEEDING	DIARRHEA	STEATORRHEA	CONSTIPATION
History (ascertain the following)	Duration: acute vs. chronic Onset: sudden vs. 1-2 hr vs. gradual Character: visceral (vague or dull, steady or cramping, diffuse) or parietal (severe, well localized, worse with movement) Location: upper, middle, or lower; radiation Associated symptoms: vomiting, hematemesis, diarrhea, hematochezia, melena, constipation, obstipation, jaundice Previous episodes Other diseases	Acute vs. chronic (duration); intermittent vs. continuous; quantity; hematemesis, melena, or hematochezia; associated pain and location; symptoms of anemia (e.g., dyspnea, chest pain, lightheadedness); medication use (especially aspirin, NSAIDs, anticoagulants); previous episodes; risk factors for chronic liver disease (alcohol, hepatitis)	Acute (<2 wk) vs. chronic (duration); fever, weight loss, or abdominal pain; stool character: number per 24 hr, watery or bloody, large vs. small volume, change in volume with eating, greasy; dietary history (especially lactose); history of IBD, pancreatic disease, intestinal surgery, DM; recent change in medications or antibiotic use; community outbreak or similar symptoms in family members; potential exposure to contaminated food; elderly immunosuppressed host; risk of HIV or sexually transmitted disease	Duration; weight loss; stool number, consistency (greasy), presence of blood; abdominal pain; flatulence; history of excessive alcohol, chronic liver disease, pancreatitis, intestinal dysmotility, surgery, DM; history of easy bruising, night blindness, bone pain, osteoporosis, dermatitis herpetiformis	Acute vs. chronic (duration); age; number of stools per week; difficulty with defecation (straining, incomplete evacuation, digital manipulation); bloating or discomfort; blood on stools, weight loss; dietary fiber and fluid intake; chronic illness (DM, neuromuscular, endocrine); abdominal surgeries; medications, impaired mobility
Physical findings (evaluate for the following)	Fever, HR, BP Appearance: calm, restless, motionless Inspect: skin, distention, hernias Bowel sounds: present, absent, roaring Percussion and palpation: organomegaly, mass (abscess), focal tenderness, guarding Peritoneal signs: sharp pain with cough, shaking, percussion, light palpation	HR, BP, orthostatic findings; abdominal pain present or absent; signs of chronic liver disease and portal hypertension (which may indicate varices); jaundice, spider angiomas, hepatosplenomegaly; ascites; examine nasogastric aspirate for blood ("coffee grounds" vs. bright red); examine stool for blood (Hemoccult) and color (melena, maroon, or bright red)	HR, BP, orthostatic findings; fever; wasting; presence of abdominal tenderness or mass; perianal disease; extraintestinal symptoms of IBD (e.g., oral ulcers, arthritis, erythema nodosum)	Wasting, presence or absence of abdominal mass or tenderness; rash (vitamin deficiencies) or excessive bruising (vitamin K deficiency); jaundice or signs of chronic liver disease	Assess mobility and chronic medical conditions; abdominal distention; palpable stool within bowel in left lower quadrant; rectal examination—impacted stool, anal fissure, rectal prolapse, pelvic floor descent with straining, rectocele
Laboratory tests	CBC, BUN, Cr, glucose, amylase, lipase, liver tests (ALT, AST, bilirubin, alkaline phosphatase), albumin, INR, U/A, urine β-HCG	CBC, BUN, Cr, liver tests, INR, type and cross	CBC, BUN, Cr, glucose, electrolytes, liver tests, albumin, C-reactive protein Selected cases: consider serum chromogranin A, VIP, calcitonin, gastrin, glucagon, urinary 5-HIAA; stool for culture, ova, parasites; consider fecal weight, fat, electrolytes, laxative screen	CBC, glucose, liver tests, albumin, electrolytes, celiac disease antibodies (anti-tTG or antiendomysial); assessment of vitamin and mineral absorption (A, D) and INR (K is fat soluble), folate, iron, calcium, phosphate, B_{12}; stool for fecal elastase and qualitative fat; stool for quantitative fecal fat; hydrogen breath test for bacterial overgrowth	CBC, Chem-7, calcium, magnesium, phosphate, TFTs; selected patients with severe constipation may undergo colonic transit studies or anal manometry with balloon expulsion
Endoscopy	EGD, colonoscopy	EGD, colonoscopy, enteroscopy, wireless capsule study	Colonoscopy (including ileal inspection) with biopsies; EGD with duodenal biopsies; wireless capsule study	EGD with duodenal biopsies	Colonoscopy if recent change in bowel habits
Imaging	CT scan or ultrasound; angiography; small bowel enterography	Tagged RBC scan, angiography	Small bowel enterography: CT, MRI, or barium (Crohn disease); somatostatin scintigraphy	CT of the abdomen (pancreatic calcifications; biliary dilation)	Usually not necessary; MRI or defecography

ALT = alanine transaminase; AST = aspartate transaminase; BP = blood pressure; BUN = blood urea nitrogen; CBC = complete blood count; CNS = central nervous system; COPD = chronic obstructive pulmonary disease; Cr = creatinine; CT = computed tomography; DM = diabetes mellitus; EGD = esophagogastroduodenoscopy; ESR = erythrocyte sedimentation rate; GERD = gastroesophageal reflux disease; GI = gastrointestinal; HbA_{1C} = hemoglobin A_{1C}; HCG = human chorionic gonadotropin; 5-HIAA = 5-hydroxyindoleacetic acid; HIV = human immunodeficiency virus; HR = heart rate; IBD = inflammatory bowel disease; INR = international normalized ratio; MRI = magnetic resonance imaging; NSAIDs = nonsteroidal anti-inflammatory drugs; RBC = red blood cell; TFTs = thyroid function tests; tTG = tissue transglutaminase; U/A = urinalysis; VIP = vasoactive intestinal polypeptide.

Modified from Proctor DD. Approach to the patient with gastrointestinal disease. In: Goldman L, Ausiello D, eds. *Cecil Textbook of Medicine*. 23rd ed. Philadelphia: Saunders-Elsevier; 2008.

NAUSEA AND VOMITING	DYSPHAGIA	ODYNOPHAGIA	HEARTBURN AND REGURGITATION	ANOREXIA	WEIGHT LOSS
Nausea with or without emesis; acute vs. chronic (duration); intermittent vs. constant; presence or absence of severe abdominal pain, comorbid illnesses, especially peptic ulcer, endocrine (DM), cardiac, psychiatric; medications; history of excessive alcohol	Oropharyngeal vs. esophageal dysphagia; solids vs. liquids; acute vs. chronic (duration); intermittent vs. progressive; GERD symptoms present or absent; weight loss; history of food impactions, allergies, atopic conditions, skin changes, cold hands (Raynaud phenomenon)	Duration of pain with swallowing; underlying immunosuppression (e.g., HIV infection, DM); caustic ingestion; use of medications that cause topical injury (especially NSAIDs, KCl, bisphosphonates, iron, antibiotics, zidovudine)	Duration of symptoms; location; relation to meals or specific foods; nocturnal symptoms; dysphagia or chest pain; extraesophageal manifestations: cough, hoarseness, asthma	Acute vs. chronic (duration); association with different foods; psychiatric disease (e.g., depression, dementia); chronic or undiagnosed medical conditions (e.g., DM, thyroid or adrenal disease, COPD, advanced heart failure, renal insufficiency, malignant disease, HIV infection); medication use	Acute vs. chronic (duration); age; total amount (>5% is significant); intentional vs. unintentional; appetite increased or decreased; rapid vs. gradual; change in physical activity; documented vs. undocumented; fever or sweats; anorexia, nausea, vomiting; diarrhea, steatorrhea, blood in stool; abdominal pain; history or symptoms of chronic medical, neurologic, or psychiatric illness; medications; alcohol and substance abuse
Acute with severe abdominal pain: evaluate for GI obstruction, pancreatitis, mesenteric ischemia, biliary colic, appendicitis, or other conditions causing peritonitis Acute without abdominal pain: evaluate for pregnancy, medications, food poisoning, infectious gastroenteritis, hepatitis, CNS disease, postoperative ileus Chronic: evaluate for medications, chronic gastric outlet obstruction (due to ulcer disease or malignant disease), impaired GI motility (gastroparesis), other chronic medical conditions, intracranial disorders, psychiatric disease (bulimia)	Usually normal; examine oropharynx and neck for lymphadenopathy and masses; evaluate the skin for sclerodermatous changes	Usually normal; evaluate oropharynx for thrush, herpetic lesions, caustic injury; general examination for signs of underlying immunosuppression	Usually normal, unless extraesophageal manifestations are present	Wasting; fever; signs of bulimia (e.g., loss of tooth enamel, knuckle ulcerations and calluses); abdominal masses; enlarged lymph nodes	Wasting; malnutrition; poor dentition or poorly fitting dentures; thyromegaly; COPD or heart failure; abdominal masses; enlarged lymph nodes; pelvic masses in women; diabetic neuropathy; signs of depression, dementia, or bulimia
β-HCG, CBC, serum electrolytes, BUN, Cr, glucose, HbA_{1c}, liver tests, albumin, TFTs, cortisol	CBC; eosinophilia or elevated IgE in some patients with eosinophilic esophagitis	CBC, HIV test, fasting glucose	Usually normal	CBC, Chem-7, liver tests, albumin, HIV test, TFTs	CBC, Chem-7, HbA_{1c}, TFTs, liver tests, C-reactive protein or ESR, calcium, phosphate, albumin, HIV test, morning cortisol
EGD to exclude gastric outlet obstruction	EGD with biopsies or dilation, esophageal motility study	EGD with biopsies	EGD (to detect erosive esophagitis or Barrett esophagus); ambulatory pH/impedance probe	Directed at detecting underlying disease, e.g., if a GI cause is suspected, EGD or colonoscopy with biopsies may be helpful	Directed at detecting underlying disease, e.g., if a GI cause is suspected, EGD or colonoscopy with biopsies may be helpful
CT of the abdomen; if chronic, also consider head CT, gastric emptying study, small bowel enterography	Esophagogram (barium swallow) will show stricture, Schatzki ring, mass	Usually not necessary	Usually not necessary	Directed at detecting underlying disease, e.g., if a GI cause is suspected, abdominal CT may be helpful	Directed at detecting underlying disease, e.g., chest or abdominal CT may be helpful

international normalized ratio may be elevated in patients with cholestasis because of malabsorption of vitamin K or in patients with chronic liver disease because of decreased hepatic synthetic function. Serum B_{12} may be decreased in patients with autoimmune gastritis (pernicious anemia), gastric bypass surgery, or malabsorption because of small bowel bacterial overgrowth or disease of the terminal ileum (e.g., Crohn disease).

Specialized laboratory tests that may be useful for the diagnosis of specific diseases include stool *Helicobacter pylori* antigen in patients with duodenal ulcer disease or dyspepsia, antibodies to tissue transglutaminase IgA in celiac disease, antibodies to microbial antigens or autoimmune markers in inflammatory bowel disease (anti–*Saccharomyces cerevisiae*, perinuclear antineutrophil cytoplasmic antibody), and CA19-9 in pancreaticobiliary malignant disease. Because of their limited sensitivity and specificity, these tests are not useful for screening but may be helpful in circumscribed situations in which the results may shift the diagnostic probability.

Stool Examination

Fecal occult blood testing is useful to evaluate iron deficiency anemia and acute or chronic GI blood loss. In patients with acute diarrhea, assessment of fecal leukocytes or culture of common pathogens is routine, and in selected patients, testing for parasites (*Giardia, Entamoeba histolytica*), *Clostridium difficile*, *Escherichia coli* O157:H7, or other specific organisms may be warranted. To distinguish among the causes of chronic diarrhea (Chapter 140), stool samples may be sent for assessment of electrolytes, leukocytes, and fecal fat.

Endoscopy and Radiology

Endoscopy (Chapter 134) and radiographic studies (Chapter 133) play a major role in the evaluation and management of many GI disorders. Esophageal manometry and esophageal pH and impedance monitoring can be useful for the evaluation of heartburn, reflux, and other esophageal symptoms (Chapter 138). Anorectal manometry may be useful in some patients with fecal incontinence and defecatory dysfunction (Chapter 145). Breath tests are commonly used to diagnose *H. pylori* infection (a urease breath test; Chapter 139), lactose intolerance, and small bowel bacterial overgrowth (a hydrogen breath test with lactulose or glucose; Chapter 140).

The diagnosis of a functional GI disorder is made after organic disorders have been excluded by clinical evaluation and limited, directed diagnostic testing. "Overtesting" should be avoided. Thereafter, the emphasis should switch from finding a "cause" of the symptoms to implementing successful coping and adaptive behaviors.

ABDOMINAL PAIN

Abdominal pain, which is a frequent complaint among outpatients in the office setting and emergency department, may be benign and self-limited or the presenting symptom of severe, life-threatening disease. Chronic abdominal pain that has been present for months or years in the absence of other organic illness is almost always functional in origin and does not require urgent evaluation. By contrast, most patients with severe acute abdominal pain require a thorough but emergent evaluation, which may quickly reveal an acute surgical illness (Chapter 142).

PATHOBIOLOGY

Stimulation of hollow abdominal viscera is mediated by splanchnic afferent fibers within the muscle wall, visceral peritoneum, and mesentery that are sensitive to distention and contraction. Visceral afferent nerves are loosely organized, innervate several organs, and enter the spinal cord at several levels. Thus, visceral pain is vague or dull in character and diffuse; patients attempting to localize the pain often move their entire hand over the upper, middle, or lower abdomen. Most visceral pain is steady, but cramping, intermittent pain or "colic" results from peristaltic contractions caused by partial or complete obstruction of the small intestine, ureter, or uterine tubes. In contrast to visceral innervation, a dense network of nerve fibers that follow a spinal T6 to L1 somatic distribution innervates the parietal peritoneum. Pain fibers of the parietal peritoneum are stimulated by stretch or distention of the abdominal cavity or retroperitoneum; direct irritation from infection, pus, or secretions (e.g., caused by a ruptured viscus); or inflammation caused by contact between the parietal peritoneum and an adjacent inflamed organ (e.g., appendicitis). Parietal pain is sharp, well characterized, and localized by the patient to a precise location on the abdomen, often by pointing with one finger.

The GI viscera (liver, biliary system, pancreas, and GI tract) arise during embryologic development from midline structures that have bilateral innervation. Thus, GI visceral pain is typically localized to the abdominal midline.

Acute Abdominal Pain

CLINICAL MANIFESTATIONS

History

The history should determine the time course, character, and location and radiation pattern of the pain (Table 132-2). Severe abdominal pain that begins suddenly during seconds to minutes indicates a catastrophic event, such as esophageal rupture, perforated peptic ulcer or viscus, ruptured ectopic pregnancy, ruptured aortic aneurysm, acute mesenteric ischemia, or myocardial infarction. Pain that progresses within 1 to 2 hours is consistent with a rapidly progressive inflammatory disorder (e.g., cholecystitis, appendicitis, pancreatitis), acute obstruction of a viscus (small intestinal obstruction, ureteral colic), or organ ischemia caused by a strangulated blood supply (volvulus, strangulated hernia, ovarian torsion). Pain that is less severe and develops during several hours is more commonly caused by a medical rather than a surgical condition, including upper GI disorders (dyspepsia), intestinal disorders (gastroenteritis, inflammatory bowel disease), liver disorders (hepatitis, abscess), urinary disorders (cystitis, pyelonephritis), and gynecologic infections; however, the slow evolution of surgical disorders such as cholecystitis (Chapter 155), appendicitis or diverticulitis (Chapter 142), and intra-abdominal abscesses must not be overlooked.

The character of the pain provides important information about whether the symptoms are due to visceral stimulation or parietal stimulation (peritonitis). Patients with peritonitis may report severe localized pain or irritation with activities or maneuvers that stretch or move the parietal peritoneum, such as walking, moving in bed, and coughing; as a result, they tend to lie quietly to avoid painful stimulation. By contrast, patients with visceral pain may move or walk restlessly or attempt a bowel movement in an effort to relieve their symptoms.

The location of pain in the upper, middle, or lower abdomen is a crude but important indicator of the diagnosis (Fig. 132-1). Visceral pain arising from the foregut (esophagus, stomach, proximal duodenum, bile duct, gallbladder, pancreas) most often is manifested in the epigastrium. Pain derived from the midgut (small intestine, appendix, ascending colon, proximal transverse colon) occurs in a periumbilical location. Pain derived from the hindgut (distal transverse colon, left colon, rectum) localizes to the lower midline between the umbilicus and symphysis pubis. Paired intra-abdominal organs such as the kidneys, ureters, ovaries, and fallopian tubes have unilateral innervation that localizes pain to the side of the involved organ. As some surgical conditions progress, the character and location of the pain shift from a visceral to a parietal pain pattern. Thus, early cholecystitis (Chapter 155) may be manifested with vague midline epigastric pain that progresses to sharp right upper quadrant pain as localized peritoneal irritation develops. Likewise, appendicitis (Chapter 142) commonly begins with vague, diffuse periumbilical pain that evolves to sharp, well-localized right lower quadrant pain as peritonitis ensues.

Anorexia, vomiting, diarrhea, distention, and constipation are commonly seen with abdominal pain caused by both medical and surgical disorders. Although nonspecific, the *absence* of any of these symptoms is evidence against an emergent surgical or medical disorder because severe illness usually leads to reflex stimulation or inhibition of gastric and intestinal peristalsis. Vomiting is common in medical and surgical disorders involving the upper GI tract, including acute gastroenteritis, pancreatitis, gastric and small intestinal obstruction, and biliary tract disease. Pain that precedes the onset of vomiting is typical of surgical conditions, whereas the reverse is true of medical conditions (e.g., food poisoning, gastroenteritis). Abdominal pain with prominent diarrhea is most commonly caused by a medical condition (e.g., gastroenteritis, inflammatory bowel disease). Although constipation alone is a nonspecific complaint, the absence of stool passage and flatus is consistent with complete bowel obstruction or paralytic ileus.

Jaundice accompanying acute abdominal pain virtually always indicates a hepatobiliary disorder (Chapter 147), including obstruction of the biliary duct (choledocholithiasis, pancreatic carcinoma, cholangiocarcinoma), complications of acute cholecystitis, acute hepatitis (viral, ischemic), and hepatic malignant neoplasms. The possibility of cholangitis should be considered and excluded in all patients with acute abdominal pain and jaundice, especially if the patient has fever, chills, hypotension, altered mental status, or leukocytosis. Hematemesis with upper abdominal pain suggests a Mallory-Weiss tear, alcoholic gastritis, or peptic ulcer disease. Hematochezia with abdominal pain is most commonly caused by medical conditions such as infectious

TABLE 132-2 TYPICAL MANIFESTATIONS OF KEY CAUSES OF ACUTE AND CHRONIC ABDOMINAL PAIN

CONDITION	LOCATION	QUALITY	ONSET	AGGRAVATING OR RELIEVING FACTORS	ASSOCIATED SYMPTOMS OR SIGNS	DIAGNOSTIC STUDIES
Peptic ulcer disease (Chapter 139)	Epigastric, occasionally RUQ, rarely LUQ	Dyspepsia: mild to moderate aching discomfort, pain, burning, gnawing, postprandial fullness	Days	Variable relief with antacids; may be relieved by, worsened by, or unrelated to meals	Recurrent; associated factors (e.g., *Helicobacter pylori,* aspirin, NSAIDs)	Anemia, upper endoscopy, *H. pylori* testing
Acute pancreatitis (Chapter 144)	Epigastric, radiates to midback (occasionally RUQ or LUQ)	Diffuse, steady, stabbing, penetrating	1-2 hr	Aggravated by food; better when lying still and with narcotics	Severe nausea and vomiting; reduced or absent bowel sounds; associated factors (e.g., alcohol, gallstones)	Elevated amylase and lipase, CT
Acute cholecystitis (Chapter 155)	Epigastric, then moves to RUQ; may radiate to right scapula	Gradual, steady increase, moderate to severe	Hours	May follow a fatty meal; better with narcotics and surgery	Nausea, some vomiting, fever	Elevated WBC count, US or CT
Acute appendicitis (Chapter 142)	Periumbilical, then moves to RLQ	Vague initially; gradual, steady increase to intense, localized, pain	Hours	Unprovoked; better with narcotics and surgery	Anorexia, nausea, obstipation; occasional vomiting, fever late	Elevated WBC count, US or CT
Diverticulitis (Chapter 142)	LLQ or suprapubic	Moderate to severe, steady or cramping, sharp or aching, localized	Hours to days	Unprovoked; better with narcotics and antibiotics or surgery	Anorexia, nausea, distention, constipation or loose stools; partial relief with passage of flatus or BM; fever late	Elevated WBC count, CT
Ruptured viscus and peritonitis (Chapter 142)	Diffuse	Intense	Minutes to hours	Worse with cough or movement; better when lying still or with narcotics or surgery	Fever, anorexia, nausea, vomiting; lack of bowel sounds; tenderness with percussion, light touch, rebound; guarding and rigidity (late); loath to move	Elevated WBC count, CT
Intestinal ischemia (Chapter 143)	Small intestine—periumbilical; proximal (right) colon—periumbilical or RLQ; distal colon—LLQ	Severe, stabbing pain out of proportion to physical findings	Minutes	Chronic ischemia—occurs after eating; acute ischemia—usually unprovoked; better with narcotics, thrombus dissolution, stenting, surgical resection	Nausea, bloody diarrhea; associated factors (e.g., hypotension, cardiac arrhythmias)	Elevated WBC count, CT or MR with angiography, or colonoscopy (colonic ischemia)
Strangulated hernia (Chapter 142)	Localized	Sharp, localized, intense; crampy or steady	Minutes to hours	Previous hernia history; unprovoked; better with narcotics and decompression, including surgery	Anorexia, nausea, vomiting, no stool or flatus passage if obstruction; bowel sounds variable—hyperactive early if obstruction present, but absent bowel sounds late, especially with peritonitis	Elevated WBC count, CT, US
Small or large bowel obstruction (Chapter 142)	Small intestine—periumbilical; proximal (right) colon—periumbilical or right abdomen; distal (left) colon—LLQ	Early—diffuse, colicky, crampy; late—steady and better localized	Hours to days	Aggravated by food; better with narcotics, NGT decompression, or surgery	Distention, anorexia, nausea, vomiting; no stool or flatus passage; small intestine—increased hyperperistaltic (rushes) bowel sounds (early) or quiet abdomen (late); large intestine—bowel sounds variable; associated factors (e.g., hernia, previous surgery)	CT
Abdominal abscess (Chapter 142)	Located over the abscess, usually LLQ or RLQ	Insidious, intense, constant	Days	May be aggravated by movement; better with abscess drainage	Fever, anorexia, nausea, abdominal mass	Elevated WBC count, CT
Acute hepatitis (Chapter 148)	RUQ	Dull or intense; localized	Days	Worse with deep inspiration	Jaundice, anorexia, nausea; liver enlarged and tender to palpation; associated factors (e.g., alcohol, infection)	Abnormal liver test results
GERD (Chapter 138)	Substernal or epigastric	Burning, gnawing	Days to years	Provoked by large or fatty meals or recumbency; relief with antacids	Recurrent; may have regurgitation, dysphagia, or extraesophageal manifestations (e.g., asthma, chronic cough, laryngitis)	Upper endoscopy (usually normal), ambulatory pH/impedance probe

TABLE 132-2 TYPICAL MANIFESTATIONS OF KEY CAUSES OF ACUTE AND CHRONIC ABDOMINAL PAIN—cont'd

CONDITION	LOCATION	QUALITY	ONSET	AGGRAVATING OR RELIEVING FACTORS	ASSOCIATED SYMPTOMS OR SIGNS	DIAGNOSTIC STUDIES
Nonulcer (functional) dyspepsia (Chapter 137)	Epigastric	Mild to moderate discomfort, pain, burning, gnawing, postprandial fullness	Years	May be worsened by meals; cannot be reliably distinguished from ulcer disease by history alone	Other symptoms of functional disorders (IBS, fibromyalgia, pelvic pain)	Normal EGD
IBS (Chapter 137)	Variable; usually lower abdomen	Vague, crampy, sense of urgency	Years	Pain may be precipitated by dietary factors or stress; associated with change in bowel characteristics (e.g., frequency, form, difficulty with passage); relieved with stool passage	Bloating and abdominal distention	Normal sigmoidoscopy, colonoscopy, and CT, but these are usually not necessary for diagnosis
Chronic pancreatitis (Chapter 144)	Epigastric or periumbilical, radiates to midback	Intense, localized	Days to years	Aggravated by food; better with narcotics	Anorexia, nausea, vomiting; associated factors (e.g., alcohol); DM (with advanced disease)	Amylase and lipase may be normal; CT may show calcifications, dilated pancreatic duct, pseudocyst; increased fecal fat and decreased fecal elastase if pancreatic insufficiency
Inflammatory or infectious enterocolitis (Chapters 142 and 283)	Small intestine—periumbilical; large intestine—right or left side of the abdomen over the colon; rectum—tenesmus	Crampy	Hours to days	Better with stool passage and treatment of underlying cause	Nausea, vomiting, bloody diarrhea; associated factors (e.g., infectious—food transmission; IBD—prolonged duration, family history)	Stool studies for culture, colonoscopy with biopsies
Malignant disease (Chapter 193)	Variable, depending on cancer location	Variable; intense and crampy if bowel obstruction; steady and vague if local invasion	Days	Better with narcotics and cancer therapy	Primary vs. metastatic disease	CT and biopsies, PET
Pneumonia/pleurisy (Chapters 97 and 99)	Upper abdomen: epigastric, RUQ, or LUQ	Localized; worse with deep breathing	Hours to days	Painful breathing; better with antibiotics	Cough, fever, dyspnea	CXR
Angina and myocardial infarction (Chapters 71-73)	Retrosternal or epigastric	Pressure, squeezing, heaviness, or intense	Minutes	Worse with exertion; relief with nitroglycerin	Dyspnea, diaphoresis	ECG, cardiac enzymes, stress testing
Genitourinary disorders (Chapters 126, 284, and 285)	Bladder—suprapubic; renal colic—abrupt, excruciating LLQ or RLQ pain radiating to the groin; prostate—dull, suprapubic; kidney—CVA	Constant or colicky; stone passage—restless, cannot find a comfortable position	Minutes to days	Better with antibiotics and pain medications (pyelonephritis or nephrolithiasis)	Hematuria, dysuria, prostate tenderness, fever	Urinalysis, urine culture, CT for stone disease
Ovarian cysts or torsion (Chapters 199 and 235)	LLQ or RLQ	Constant, intense	Minutes	Better with NSAIDs or surgery (torsion)	Nausea, vomiting; may be recurrent	US
Ruptured ectopic pregnancy (Chapter 239)	LLQ or RLQ	Constant, intense, stabbing	Minutes	Better with surgery	Rebound and guarding present, abnormal menses or amenorrhea	Acute anemia, elevated β-HCG, US
Musculoskeletal disorders	Specific muscle groups	Aching	Days	Better with heat or NSAIDs; aggravated by movement	History of muscle injury or exertion	Normal laboratory results
Herpes zoster (Chapter 375)	Dermatomal distribution	Burning, itching, neuropathic, constant	Days	Aggravated by touching the dermatome; better with pain or antiviral medications	Recurrent; rash may or may not be present	Skin culture or biopsy

TABLE 132-2 TYPICAL MANIFESTATIONS OF KEY CAUSES OF ACUTE AND CHRONIC ABDOMINAL PAIN—cont'd

CONDITION	LOCATION	QUALITY	ONSET	AGGRAVATING OR RELIEVING FACTORS	ASSOCIATED SYMPTOMS OR SIGNS	DIAGNOSTIC STUDIES
Metabolic disorders (e.g., DM; Chapter 229)	Epigastric or generalized	Intense, constant	Hours to days	Worse with poor metabolic control (e.g., poor glucose control)	Recurrent; nausea, vomiting, diabetic neuropathy	Specific metabolic parameters abnormal (e.g., elevated glucose in DM)
Abdominal epilepsy (Chapter 403)	Epigastric or umbilical	Constant	Hours to days	Unprovoked; better with antiseizure therapy	Recurrent; may have associated seizure disorder	EEG
Dissecting or leaking abdominal aortic aneurysm (Chapter 78)	Over the aneurysm, radiates to the back or groin	Severe, searing, constant	Minutes to hours to days	History of HTN or CAD	Shock, pulsatile mass; bruit *not* usually present	Acute anemia, CT, angiography

BM = bowel movement; CAD = coronary artery disease; CT = computed tomography; CVA = costovertebral angle; CXR = chest radiograph; DM = diabetes mellitus; ECG = electrocardiography; EEG = electroencephalography; EGD = esophagogastroduodenoscopy; GERD = gastroesophageal reflux disease; HCG = human chorionic gonadotropin; HTN = hypertension; IBD = irritable bowel disease; IBS = irritable bowel syndrome; LLQ = left lower quadrant; LUQ = left upper quadrant; MR = magnetic resonance; NGT = nasogastric tube; NSAIDs = nonsteroidal anti-inflammatory drugs; PET = positron emission tomography; RLQ = right lower quadrant; RUQ = right upper quadrant; US = ultrasonography; WBC = white blood cell.
Modified from Proctor DD. Approach to the patient with gastrointestinal disease. In: Goldman L, Ausiello D, eds. *Cecil Textbook of Medicine*. 23rd ed. Philadelphia: Saunders-Elsevier; 2008.

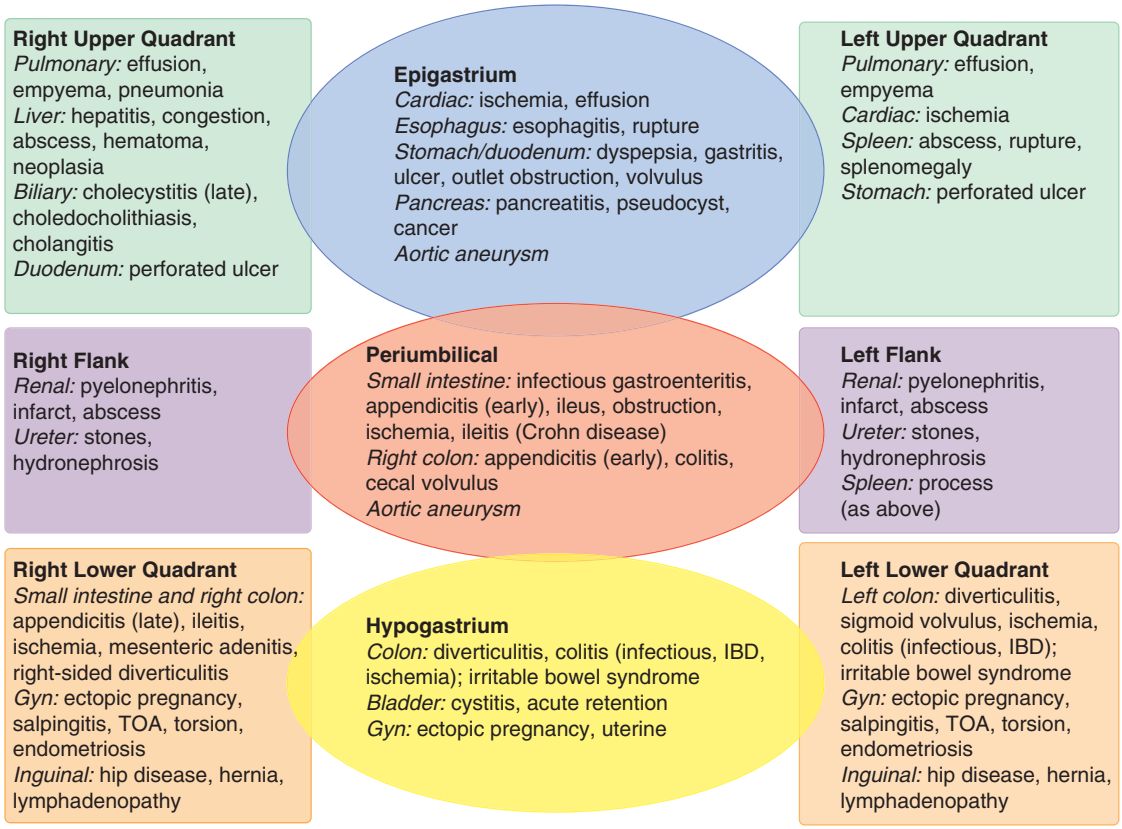

FIGURE 132-1. Differential diagnosis of abdominal pain by its initial location. IBD = inflammatory bowel disease; TOA = tubo-ovarian abscess.

gastroenteritis or inflammatory bowel disease, but it also may be caused by ischemic colitis or mesenteric ischemia. Gross hematuria may be due to cystitis (Chapter 284) or a ureteral stone (Chapter 126). Abdominal pain with weight loss may be due to inflammatory bowel disease, chronic mesenteric ischemia, or advanced GI malignant neoplasms. In women, a missed menstrual period, adnexal pain, spotting, or cramping may suggest pregnancy, ectopic pregnancy, or spontaneous abortion. Acute pain between cycles may be caused by ovarian follicles or ruptured corpus luteum cysts. Pelvic pain with fever, chills, or cervical discharge suggests pelvic inflammatory disease.

The past medical history and review of systems can provide clues about systemic and extra-abdominal conditions that may be manifested with abdominal pain. Acute coronary syndromes (Chapter 72), heart failure (Chapter 58), pneumonia (Chapter 97), or empyema may cause dyspepsia, epigastric or right or left upper quadrant pain, nausea, and vomiting. Metabolic conditions such as uremia (Chapter 130), diabetes with hyperglycemia or ketoacidosis (Chapter 229), hypercalcemia (Chapter 245), or acute adrenocortical insufficiency (Chapter 227) may cause pain, nausea, vomiting, and diarrhea. Acute intermittent porphyria (Chapter 210) and familial Mediterranean fever (Chapter 275) may cause recurrent episodes of severe pain and peritonitis that

may be misdiagnosed, leading to unnecessary surgeries. Other causes of acute abdominal pain include narcotic withdrawal (Chapter 34), insect or reptile bites (Chapter 112), and lead or arsenic poisoning (Chapter 22).

Physical Examination

The physical examination must identify life-threatening illnesses that require urgent surgical evaluation. Nevertheless, the examination must be orderly, careful, and complete. If the examiner immediately palpates the site of maximal pain, the patient is unlikely to relax and to cooperate for the remainder of the examination.

First, the patient should be observed and the abdomen inspected. Most patients remain calm, cooperative, and freely capable of moving during the examination. Patients who are writhing or restless may have pain due to visceral distention (e.g., renal colic, intestinal obstruction), whereas patients who lie motionless may have peritonitis. Gentle shaking of the bed or having the patient cough may elicit sharp, well-localized pain in patients with parietal but not with visceral pain. Auscultation should be performed before percussion or palpation so that intestinal activity is undisturbed. An abdomen that is quiet except for infrequent squeaks or tinkles suggests peritonitis or ileus. Loud peristaltic rushes that occur in synchrony with abdominal pain suggest small bowel obstruction. Light percussion across the upper, middle, and lower abdomen can determine any site of focal tenderness suggestive of peritonitis. Light palpation should be performed with one or two fingers (not the whole hand), beginning away from where the patient localizes the pain and gradually moving to the site of pain. Thereafter, gentle, deeper palpation of the entire abdomen is performed gradually, including the region of tenderness. An attempt should be made to palpate for an abdominal aortic aneurysm (Chapter 78). Examination also should include the inguinal and femoral canals, umbilicus, and surgical scars for evidence of incarcerating hernias. The presence of focal tenderness indicates parietal peritoneal irritation. Voluntary or involuntary tightening of the muscle wall ("guarding") may occur during palpation. With gentle, steady compression of the abdomen with one hand, voluntary guarding usually subsides, allowing the examination to proceed. Persistent involuntary guarding indicates peritonitis with reflex muscle wall contraction. Testing for "rebound tenderness" in patients with suspected peritonitis is not recommended because it causes significant pain and is usually not necessary to establish the diagnosis. When the presentation strongly suggests a nonserious GI disorder but the patient has significant tenderness with palpation, it is useful to use the stethoscope ostensibly to listen for bowel sounds but actually to reproduce the pressure of palpation. A significant discrepancy in the tenderness elicited by the stethoscope and by digital palpation may be seen in patients who are anxious, have functional complaints, or are seeking secondary gain. A digital rectal examination should be performed in most patients with acute abdominal pain to evaluate for tenderness or fluctuance that suggests a perirectal abscess and to assess the stool for signs of overt or occult blood. Women with lower abdominal pain should have a pelvic examination by a skilled examiner to evaluate for gynecologic disease. Some specific and dramatic findings point to particular diagnoses (Table 132-3).

TABLE 132-3 PHYSICAL SIGNS IN PATIENTS WITH ACUTE ABDOMINAL PAIN

SIGN	DESCRIPTION	DIAGNOSIS
Murphy sign	Cessation of inspiration during right upper quadrant examination	Acute cholecystitis
McBurney sign	Tenderness located midway between anterior superior iliac spine and umbilicus	Acute appendicitis
Cullen sign	Periumbilical bluish discoloration	Retroperitoneal hemorrhage Pancreatic hemorrhage Ruptured abdominal aortic aneurysm
Grey Turner sign	Bluish discoloration of flanks	Retroperitoneal hemorrhage Pancreatic hemorrhage Ruptured abdominal aortic aneurysm
Kehr sign	Severe left shoulder pain	Splenic rupture Ectopic pregnancy rupture
Obturator sign	Pain with flexed right hip rotation	Appendicitis
Psoas sign	Pain with straight leg raising against resistance (right side)	Appendicitis

Special Populations

Increased diligence is required in the evaluation of patients in whom abdominal signs and symptoms may be minimal until the disease process is far advanced. Such patients include the elderly (Chapter 25) and patients who have dementia (Chapter 402), psychiatric disturbances (Chapter 397), or spinal cord injuries. An admitting diagnosis of "altered mental status," "failure to thrive," "obstipation," or "fever of unknown origin" may stem from serious intra-abdominal conditions. Disorders that may be overlooked in the elderly include bowel perforation, bowel obstruction, cholecystitis, diverticulitis, volvulus, mesenteric ischemia, and abdominal aortic aneurysm. In patients with chronic liver disease, the presence of ascites may mask the signs and symptoms of serious surgical conditions such as cholecystitis, appendicitis, and diverticulitis. Even in the presence of perforation, signs of peritonitis may be lacking because the ascites fluid separates the visceral peritoneum and parietal peritoneum. Likewise, immunocompromised populations, who are at risk for infectious, drug-related, and iatrogenic complications, may manifest few physical findings or laboratory abnormalities. Owing to the limitations of the clinical evaluation in these vulnerable populations, there should be a low threshold for the use of abdominal imaging.

Abdominal Pain Developing in the Hospital

When pain develops as a new problem in a hospitalized patient, it is usually caused by a limited number of conditions. Postprocedural complications may cause perforation, infection, or bleeding (intraperitoneal, retroperitoneal, or within solid organs). Shunting of splanchnic blood flow in severely ill medical or surgical patients may cause stress gastritis, nonocclusive mesenteric ischemia, or acalculous cholecystitis. Adynamic ileus or acute colonic pseudoobstruction is common in critically ill or postoperative patients and is manifested as diffuse abdominal pain and distention. *Clostridium difficile* (Chapter 296) colitis is a common cause of pain, diarrhea, and distention, especially in patients receiving antibiotics. Constipation (Chapter 136), which is a common problem in hospitalized patients, may go unnoticed until pain and distention develop. Finally, many medications can cause dyspepsia and abdominal pain.

DIAGNOSIS

Patients with acute abdominal pain should have a complete blood count with differential; leukocytosis is present in most acute surgical conditions (Fig. 132-2). A pregnancy test is required in women of childbearing age. Serum levels of electrolytes, glucose, blood urea nitrogen, and creatinine assess hydration, acid-base status, and renal function. Liver chemistries and pancreatic enzymes should be obtained in most patients, but especially in those with upper abdominal pain, jaundice, or vomiting. An elevation in aspartate or alanine aminotransferase levels may reflect choledocholithiasis with acute biliary obstruction (Chapter 155), acute gallstone pancreatitis (Chapter 144), or a hepatocellular process (Chapter 148). Painful jaundice with a significant rise in the alkaline phosphatase level usually reflects cholestasis caused by extrahepatic biliary obstruction (Chapter 155). Amylase and lipase levels are elevated in most patients with acute pancreatitis, but minor amylase elevations also occur with a perforated viscus or mesenteric ischemia (Chapter 143). Urinalysis may demonstrate pyuria, hematuria, or bacteriuria due to ureteral calculi (Chapter 126) or urinary tract infection (Chapter 284).

Imaging

Ultrasound is preferred in suspected pregnancy and to evaluate for other acute gynecologic disorders, such as tubo-ovarian abscess, ruptured corpus luteum cyst, or ovarian torsion; it is also preferred for the initial evaluation of suspected acute cholecystitis (Chapter 155) and ureteral stones with hydronephrosis (Chapter 123) and for the bedside evaluation of unstable patients. In most other settings, abdominal computed tomography (CT) with oral and intravenous administration of contrast material (when possible) is preferred and can provide a definitive diagnosis in up to 90% of patients with acute severe abdominal pain (Chapter 133). Abdominal CT may be falsely negative early in the course of acute pancreatitis, mesenteric ischemia, cholecystitis, appendicitis, and diverticulitis, especially if it is performed without contrast enhancement.

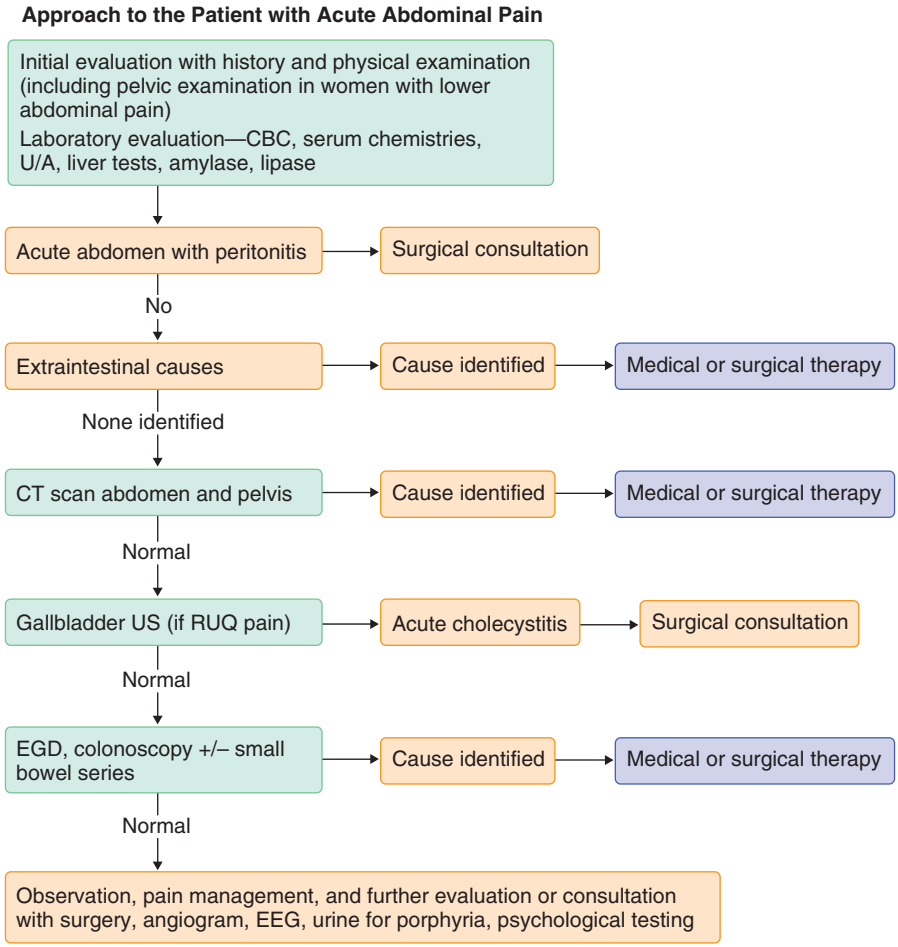

FIGURE 132-2. Approach to the patient with acute abdominal pain. CBC = complete blood count; CT = computed tomography; EEG = electroencephalography; EGD = esophagogastroduodenoscopy; RUQ = right upper quadrant; U/A = urinalysis; US = ultrasonography.

TREATMENT

Rx

Once the diagnosis is clear, treatment of the underlying condition is initiated. In patients with nonspecific acute abdominal pain and no clear diagnosis, early laparoscopy is useful for diagnosis, but outcomes such as complication rates, readmission rates, and length of hospitalization are no better than with a strategy of active observation.

Chronic Abdominal Pain

Chronic or recurrent abdominal pain that has been present for months to years may be caused by structural (organic) disease, but the majority of patients have a functional disorder such as irritable bowel syndrome (Chapter 137). Common organic causes of chronic abdominal pain include medications with GI side effects, peptic ulcer disease (Chapter 139), inflammatory bowel disease (Chapter 141), chronic pancreatitis (Chapter 144), biliary tract disease (Chapter 155), GI cancers (Chapters 192 and 193), and endometriosis (Chapter 236). The clinician should attempt to distinguish patients with symptoms or signs of organic disease, in whom further diagnostic investigation is warranted, from those with probable functional disease (Fig. 132-3). Although functional disorders occur in all age groups, the symptoms usually begin before the age of 40 years. "Alarm" features that suggest a structural disorder and are inconsistent with a functional disorder are fever, severe pain, significant weight loss, jaundice, progressive dysphagia, recurrent vomiting, nocturnal pain or diarrhea, and stools that are bloody or positive for fecal occult blood. Laboratory study findings should be normal with functional disorders; therefore, an unrevealing evaluation for anemia, leukocytosis, and levels of iron, albumin, C-reactive protein, and vitamins A, D, or B_{12} argues against structural or organic disease.

In patients younger than 50 years with a suspected functional disorder and no alarm features (e.g., family history of colon cancer or inflammatory bowel disease or abnormalities on screening blood tests), further testing should be minimized, and the emphasis should be shifted to managing symptoms, coping, and making lifestyle changes (Chapter 137). In patients who may have organic disease, testing often includes a combination of upper GI endoscopy, colonoscopy, and ultrasound or CT imaging.

GAS AND BLOATING

Belching

Belching (eructation), which is the involuntary or voluntary release of gas from the esophagus or stomach, commonly occurs during or after a meal. Virtually all belching is caused by swallowed air, which may be increased by eating quickly, drinking carbonated beverages, chewing gum, and smoking. Gas also may be produced within the stomach by antacids, especially sodium bicarbonate, which rapidly neutralize gastric acid and release carbon dioxide. Belching seldom reflects serious GI dysfunction but may be increased in patients with gastroesophageal reflux (Chapter 138), functional dyspepsia (Chapter 137), or gastroparesis (Chapter 136). Chronic, excessive, repetitive belching is a functional disorder caused by transient ingestion of air into the esophagus (caused by subconscious diaphragmatic contraction and upper esophageal sphincter relaxation) and its subsequent expulsion; it is treated with behavioral modification.[4]

Flatus

Flatus or "gas" is a normal byproduct of digestion. Otherwise healthy adults pass flatus 10 to 20 times daily and excrete up to 1500 mL. Thus, it is difficult to distinguish patients with abnormal or excessive gas production from those with only a heightened awareness of or sensitivity to normal production. Increased flatulence with diarrhea may be symptomatic of

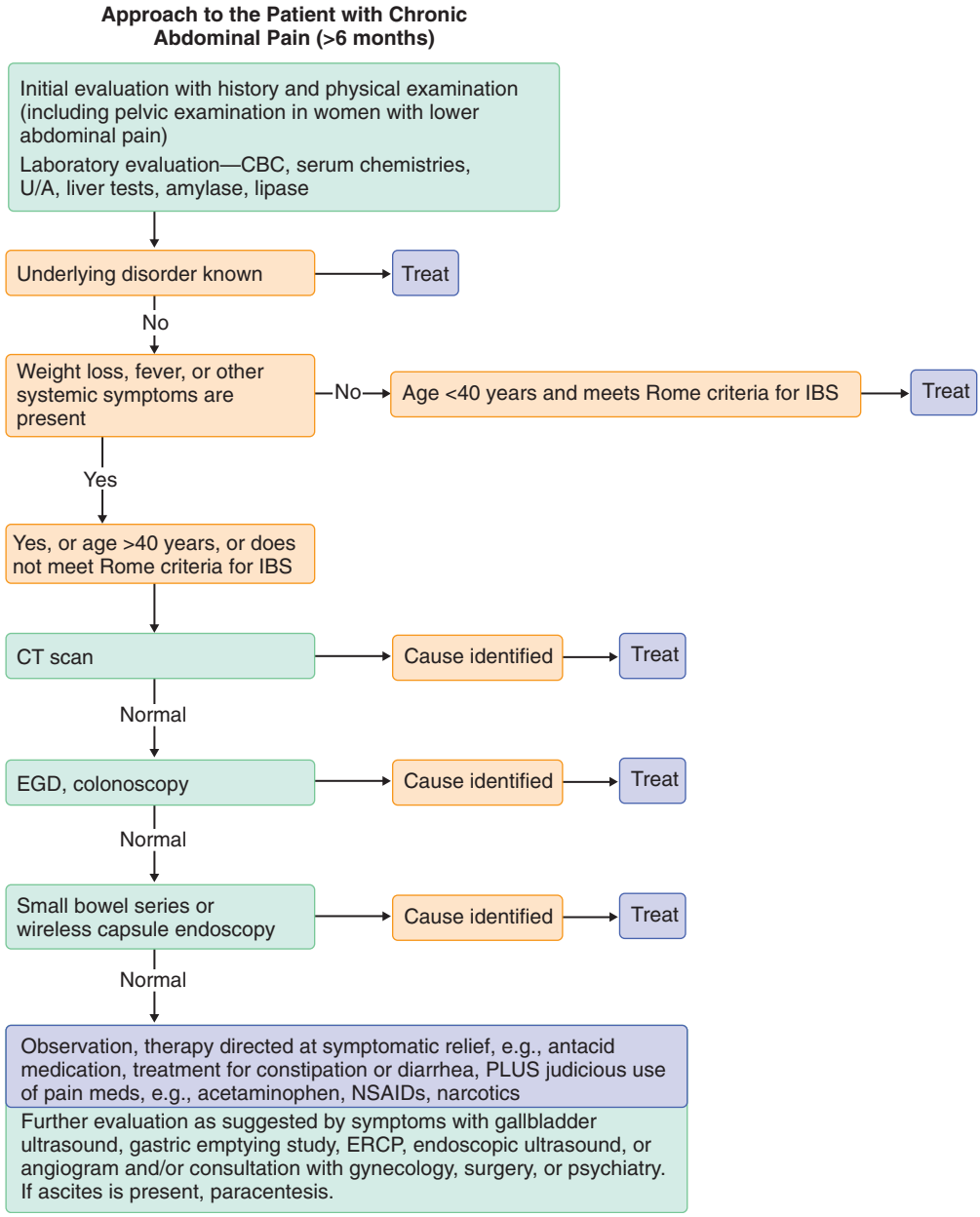

FIGURE 132-3. Approach to the patient with chronic abdominal pain. CBC = complete blood count; CT = computed tomography; EGD = esophagogastroduodenoscopy; ERCP = endoscopic retrograde cholangiopancreatography; IBS = irritable bowel syndrome; NSAIDs = nonsteroidal anti-inflammatory drugs; U/A = urinalysis.

disorders of malabsorption, including celiac disease (Chapter 140), pancreatic insufficiency (Chapter 144), and small intestinal bacterial overgrowth (Chapter 140).

In normal adults, flatus is derived from two sources: swallowed air and colonic bacterial fermentation of FODMAPs (fermentable oligosaccharides, disaccharides, and monosaccharides and polypols), which are short-chain carbohydrates that may be incompletely absorbed in the small intestine and result in the colonic production of carbon dioxide or methane. FODMAPs include lactose (dairy products), fructose, fructans, polypols, and galacto-oligosaccharides. Fructose is present in fruit, especially apples and pears, and is a major component of corn syrups that are widely used as sweeteners. Polyols include sorbitol, which is a natural sugar in stone fruit (peaches, apricots, plums, prunes) and a common added sweetener in sugar-free candies, as well as trehalose, which is present in mushrooms. Fructans and oligosaccharides are plentiful in cruciferous vegetables (cabbage, broccoli, cauliflower, Brussels sprouts, turnips, rutabagas), garlic, onions, legumes (beans, soy, lentils, peas), pasta, and whole grains.[5]

TREATMENT

Rx

Patients with long-standing flatulence in the absence of other symptoms or signs of GI disease can be treated conservatively. Avoidance of carbonated beverages, chewing gum, sorbitol- and fructose-containing sweeteners, and gas-producing vegetables improves symptoms in most patients.[■■] Lactase deficiency may be confirmed by a lactose breath test. Underlying GI illness is suggested by the recent onset of flatulence with other symptoms of organic disease, including weight loss, abdominal pain, diarrhea, distention, and abnormal laboratory studies (Chapter 140). A positive fecal fat analysis confirms malabsorption and merits further investigation (see Table 140-6). Suspected small bowel bacterial overgrowth may be confirmed by carbohydrate breath tests or treated empirically with antibiotics.

Bloating and Distention

Bloating and distention are common complaints among patients with functional GI disorders (Chapter 137). As chronic, isolated symptoms, they are almost never caused by serious structural disease. Functional bloating may be caused by heightened sensitivity to minor increases in intestinal gas or impaired transit of gas, even though the total volume of intestinal gas is within normal limits. The acute onset of distention in conjunction with alarm symptoms such as cramping pain, weight loss, nausea, vomiting, obstipation, or diarrhea warrants further evaluation for disorders that cause intestinal obstruction (Chapter 142) or malabsorption (Chapter 140). Rifaximin (550 mg three times daily for 2 weeks) is effective for functional bloating, pain, and loose or watery stools[6]; but dietary and behavioral changes and reassurance may also be useful.

INVOLUNTARY WEIGHT LOSS

The unintentional loss of more than 5% of baseline weight within a 12-month period is frequently due to a serious underlying medical or psychiatric illness. Weight loss is seldom the sole presenting sign of medical disorders, but it is often revealed during the clinical evaluation of other complaints. Chronic weight loss in the elderly is commonly caused by depression, dementia, difficulty with chewing or swallowing, malignant disease, medications, alcoholism, or physical and social limitations to procuring, preparing, and eating meals (Table 132-4) (Chapter 24). Gradual, mild weight loss occurs in some elderly patients because of the loss of lean body mass. In young patients, weight loss is more commonly caused by eating disorders (Chapter 219), endocrine disorders (Chapters 226 and 227), or chronic GI conditions such as inflammatory bowel disease (Chapter 141) or celiac disease (Chapter 140). In chronic medical conditions, involuntary weight loss is usually caused by a combination of decreased appetite (anorexia) and varying degrees of cachexia; examples include advanced malignant disease, chronic infections (HIV, tuberculosis), heart failure, chronic kidney or liver disease, end-stage lung disease, and adrenal insufficiency. Weight loss that occurs in the presence of normal or increased appetite suggests increased metabolism and energy expenditure caused by endocrine disorders, such as poorly controlled diabetes (Chapter 229) or hyperthyroidism (Chapter 226), or GI disorders that result in food malabsorption (Chapter 140). Chronic GI disorders that cause progressive narrowing or obstruction of the esophagus (cancer, achalasia), stomach (cancer, peptic ulcer disease with gastric outlet obstruction), small intestine (Crohn disease), or arterial circulation (chronic mesenteric ischemia) may cause weight loss as a result of dysphagia, vomiting, or postprandial pain that limits the ability to ingest sufficient calories.

DIAGNOSIS

The cause of weight loss (see Table 132-4) is usually evident from the history, physical examination, and routine laboratory studies, including complete blood count, electrolytes, liver chemistries, thyroid-stimulating hormone, urinalysis, and, when appropriate, HIV serology (Fig. 132-4). A chest radiograph should be obtained in patients who smoke, have any respiratory symptoms, or are older than 40 years. Signs of dehydration or severe malnutrition may require an assessment for nutritional deficiencies (Chapter 214) and nutritional support (Chapters 216 and 217).

Other symptoms and signs necessitate consideration of additional diagnostic testing. Weight loss with increased appetite merits an assessment of thyroid function (Chapter 226), glucose intolerance (Chapter 229), and malabsorption (Chapter 140). Suspected malabsorption may be confirmed by a positive fecal fat analysis. GI symptoms suggesting obstruction or occult GI malignant disease can be evaluated with upper GI endoscopy, upper GI radiographic series, colonoscopy, or abdominal CT. Psychiatric evaluation may be warranted in patients with signs of depression, early dementia, or eating disorders. In up to 25% of patients, no cause of weight loss is found.

NAUSEA AND VOMITING

Nausea is an unpleasant feeling of the impending need to vomit. *Vomiting* is the forceful oral expulsion of gastric contents as a result of retrograde contraction of the duodenum and antrum with compression of the thoracoabdominal musculature. Nausea and vomiting may be caused by a number of GI and non-GI disorders, but they are best categorized according to chronicity and the presence of abdominal pain. The acute onset of vomiting *with* severe abdominal pain suggests a serious illness potentially requiring surgical intervention, including GI obstruction (Chapter 142), mesenteric ischemia (Chapter 143), pancreatitis (Chapter 144), biliary colic (Chapter 155), and conditions causing peritonitis (Chapter 142), such as appendicitis or a perforated viscus. Acute vomiting *without* abdominal pain is most commonly caused by medications (including chemotherapy), motion sickness (Chapter 428), food poisoning (Chapter 283), infectious gastroenteritis (Chapter 283), hepatitis (Chapters 148 and 149), upper GI bleeding, postoperative ileus, or acute central nervous system disease. Chronic or recurrent nausea and vomiting *with* abdominal pain are commonly caused by GI disorders that result in the partial or intermittent obstruction of the stomach or small intestine. Chronic nausea and vomiting *without* abdominal pain may be due to disorders that impair gastric emptying or small intestine motility and non-GI causes, including medications, pregnancy, intracerebral disorders, cardiac disease, endocrine disease, labyrinth disorders, psychiatric disease (including bulimia), and functional disorders. Vomiting of undigested food eaten hours earlier suggests gastric obstruction or gastroparesis. Abdominal distention or feculent emesis suggests obstruction of the small intestine.

DIAGNOSIS

Most cases of acute vomiting without abdominal pain are self-limited and require no evaluation (Fig. 132-5). Medication-related symptoms and pregnancy should be excluded. With severe vomiting, serum electrolyte values should be obtained. Hyperglycemia may cause acute gastroparesis. Increased liver chemistries or pancreatic enzymes suggest hepatobiliary or pancreatic disease. In patients with acute abdominal pain and vomiting, abdominal plain radiography or CT is performed to look for evidence of GI obstruction, a perforated viscus, or pancreaticobiliary disease. In patients with chronic vomiting of uncertain cause, the goal is to distinguish structural GI disorders, GI motility disorders, and non-GI disorders. Esophagogastroduodenoscopy, enterography, abdominal cross-sectional imaging, GI motility studies, and head CT or magnetic resonance imaging may be indicated.

TREATMENT

Rx

The approach to the medical treatment of nausea and vomiting depends on the cause (Table 132-5). Patients who are receiving moderately emetogenic chemotherapy are frequently managed with a 5-HT_3 receptor antagonist and dexamethasone; aprepitant is added for highly emetogenic regimens. For patients with mildly emetogenic regimens or vomiting from other causes, treatment with single or combinations of anticholinergic agents, dopamine receptor antagonists, or 5-HT_3 receptor antagonists usually provides symptomatic relief.[7]

OTHER GASTROINTESTINAL COMPLAINTS

Heartburn, esophageal regurgitation, dysphagia, odynophagia, and noncardiac chest pain suggest esophageal disease (Chapter 138). *Dyspepsia*, which refers to bothersome, intermittent, mild to moderate upper abdominal or epigastric symptoms (burning, pain, early satiation, postprandial fullness) can be caused by peptic ulcer disease (Chapter 139) or esophageal disease (Chapter 138), or it can be functional in origin (Chapter 137). An orderly diagnostic approach (Fig. 132-6) can help distinguish among the various causes, avoid unnecessary testing, and minimize symptoms.[8]

Diarrhea, which is defined pathophysiologically as an increase in stool weight to more than 200 g/day, can be caused by malabsorption of osmotically active substances or by increased intestinal secretion of electrolytes and water. In clinical practice, however, stool weight is seldom quantified, and the term *diarrhea* refers to an increase in stool liquidity or frequency (more than three bowel movements per day). Acute and chronic diarrhea should be distinguished because the evaluation and treatment are different[9] (Chapter 140).

Constipation (Chapter 136), which is the most common digestive symptom, occurs in 15% of the population. Constipation may refer to fewer than three bowel movements per week; hard or lumpy stools; or difficulty during defecation, characterized by straining, a sensation of obstruction or incomplete evacuation, or the need to engage in manual manipulations to promote evacuation. Constipation may be caused by systemic conditions that slow colonic transit, including neuromuscular disease, endocrine disorders, and electrolyte abnormalities, or by lesions that obstruct the passage of stool through the distal colon or anorectum, such as neoplasms, strictures, prolapse, and aganglionosis (Hirschsprung disease). Most patients, however, do not have an apparent cause and are deemed to have functional constipation.[10]

TABLE 132-4 CAUSES OF INVOLUNTARY WEIGHT LOSS

CONDITION	QUALITY	DURATION	AGGRAVATING OR RELIEVING FACTORS	ASSOCIATED SYMPTOMS OR SIGNS	DIAGNOSTIC STUDIES
WEIGHT LOSS SECONDARY TO GASTROINTESTINAL CAUSES					
GI, pancreatic, or hepatobiliary malignant disease (Chapters 192-196)	Progressive, fast	Months	Better with cancer therapy (e.g., surgery, XRT, chemotherapy)	Dysphagia (esophageal); anorexia, nausea, vomiting (gastric, small or large bowel obstruction); visible or occult blood in stool; altered bowel habits; jaundice or hepatomegaly (biliary obstruction, hepatic tumor, metastatic disease); iron deficiency anemia	CBC, FOBT, ferritin, CEA, CA19-9, AFP, EGD, colonoscopy, abdominal CT, PET
Malabsorption (Chapter 140) (poor absorption of nutrients due to pancreatic insufficiency, small intestinal mucosal disorders, or bacterial overgrowth)	Progressive, slow	Months to years	Diarrhea or steatorrhea, excessive flatulence; worse with eating and resolves with NPO status	Usually associated with increased appetite; may have anemia (iron, B_{12}, folate), osteoporosis, or osteomalacia (vitamin D, calcium, phosphorus); easy bruising (vitamin K), night blindness (vitamin A)	72-hr stool for fecal fat; fecal elastase; vitamins A and D and INR; calcium, ferritin, B_{12}, albumin; celiac disease antibodies (e.g., anti-tTG, antiendomysial antibodies); EGD with small bowel biopsy; breath test for bacterial overgrowth
Inflammatory bowel disease (especially Crohn disease) (Chapter 141)	Progressive, slow	Months	Eating causes pain, cramps, increased diarrhea and urgency; improved by low-residue diet or NPO status	Bloody stools, abdominal cramps and pain, perianal disease, extraintestinal manifestations (e.g., oral ulcers, uveitis, erythema nodosum, arthralgias)	CBC, albumin, ESR, CRP, colonoscopy with biopsies, CT or MR enterography, wireless capsule study
GI motility disorders (Chapter 136)	Intermittent, slow	Years	Worse with eating	Nausea, vomiting, distention, diarrhea, or constipation may be present	EGD and colonoscopy, gastric emptying study, CT or MR enterography, surgical full-thickness intestinal biopsies
Cirrhosis (Chapter 153)	Muscle wasting with edema, so weight may increase	Months to years	Worse with salt or fluid intake	Ascites, peripheral edema	Liver biopsy
Chronic intestinal ischemia (Chapter 143)	Progressive	Months to years	Worse with eating	Afraid to eat; postprandial abdominal pain, nausea; associated atherosclerotic disease	CT or MR angiography
WEIGHT LOSS SECONDARY TO NONGASTROINTESTINAL CAUSES					
Poor or inadequate calorie intake due to social factors (Chapter 215)	Intermittent or progressive, acute (hospitalized) or chronic	Days to months to years	Common in elderly, teenagers; exacerbated by poor dentition or poorly fitting dentures	Will eat if food is made available	Review dietary log and how food is obtained and prepared
Medications	Intermittent or progressive	Months	Worse with medication; resolves with discontinuation of offending drug	Anorexia, nausea, vomiting	Review drug profile
Non-GI malignant disease	Progressive	Months	Better with cancer therapy (e.g., surgery, XRT, chemotherapy)	Anorexia, nausea, vomiting; pain; metastatic disease	Calcium, cortisol; CT for underlying disease, PET
Endocrine disorders: DM, hyperthyroidism, adrenal insufficiency (Chapters 226, 227, and 229)	DM—appetite increased or decreased, early satiety; hyperthyroidism—increased appetite	Months to years	Worse with disease chronicity	DM: gastroparesis, neuropathy, retinopathy, nephropathy Adrenal insufficiency: nausea, vomiting, diarrhea, abdominal pain	Serum glucose, TFT, cortisol
Chronic infections, including HIV and TB (Chapters 324 and 390)	Progressive, fast	Months	Better with directed therapy, megestrol acetate (Megace)	Nausea, anorexia, other infections	HIV test, PPD, cultures, biopsies if necessary
Systemic inflammatory disorders	Progressive, moderate	Months to years	Better with directed therapy, megestrol acetate (Megace)	Arthritis, rash, vasculitis	ANA, RF, ESR, CRP
Chronic renal failure (Chapter 130)	Progressive, slow; edema may increase weight	Months to years	Better with dialysis, megestrol acetate (Megace)	Nausea, anorexia, weight gain	BUN, Cr, 24-hr creatinine clearance
Advanced COPD or heart failure (Chapters 58 and 88)	Progressive, slow	Months to years	Better with oxygen and specific treatment	Fatigue, dyspnea, edema, wasting	Pulmonary function testing or two-dimensional echocardiography
Psychiatric illness: depression, manic-depressive illness (Chapter 397)	Progressive, slow	Months to years		Depression common in elderly; flat affect; manic phase associated with hyperactivity and decreased intake	Psychological testing

TABLE 132-4 CAUSES OF INVOLUNTARY WEIGHT LOSS—cont'd

CONDITION	QUALITY	DURATION	AGGRAVATING OR RELIEVING FACTORS	ASSOCIATED SYMPTOMS OR SIGNS	DIAGNOSTIC STUDIES
Psychogenic eating disorders—anorexia nervosa, bulimia (Chapter 219)	Intermittent or progressive	Months to years	Worse with stressors	Refusal to eat, loss of tooth enamel, calluses and healing ulcerations of hand	Psychiatric testing
Substance abuse (alcohol, opiates, CNS stimulants)	Intermittent or progressive	Months	Resolves with discontinuation	Anorexia, nausea, vomiting	Careful interview; patients may deny or minimize

AFP = α-fetoprotein; ANA = antinuclear antibody; BUN = blood urea nitrogen; CBC = complete blood count; CEA = carcinoembryonic antigen; CNS = central nervous system; COPD = chronic obstructive pulmonary disease; Cr = creatinine; CRP = C-reactive protein; CT = computed tomography; DM = diabetes mellitus; EGD = esophagogastroduodenoscopy; ESR = erythrocyte sedimentation rate; FOBT = fecal occult blood test; GI = gastrointestinal; HIV = human immunodeficiency virus; INR = international normalized ratio; MR = magnetic resonance; NPO = nothing orally; PET = positron emission tomography; PPD = purified protein derivative; RF = rheumatoid factor; TB = tuberculosis; TFT = thyroid function test; tTG = tissue transglutaminase; XRT = x-ray therapy.
Modified from Proctor DD. Approach to the patient with gastrointestinal disease. In: Goldman L, Ausiello D, eds. *Cecil Textbook of Medicine*. 23rd ed. Philadelphia: Saunders-Elsevier; 2008.

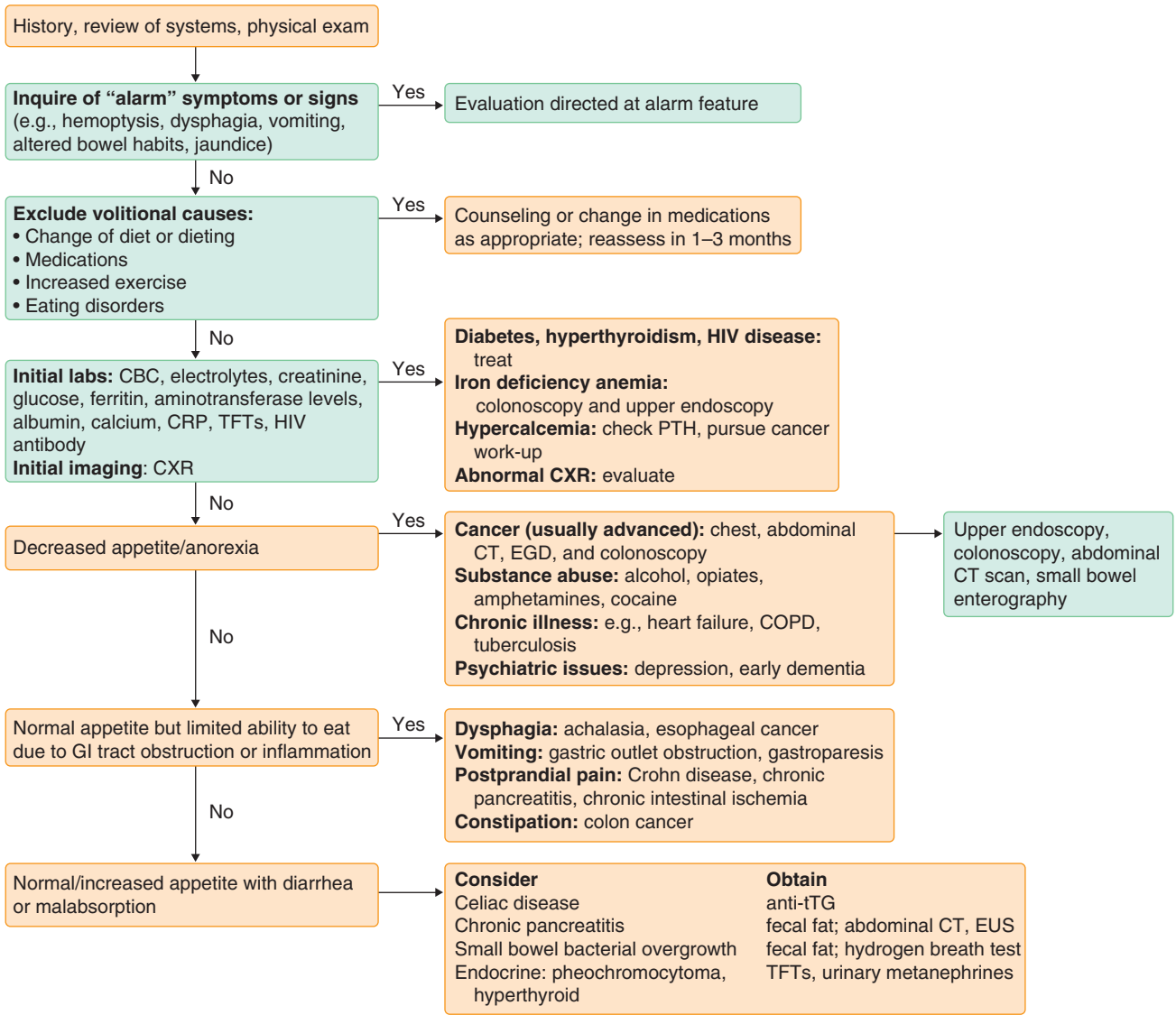

FIGURE 132-4. Approach to the patient with unintentional weight loss of more than 5%. CBC = complete blood count; COPD = chronic obstructive pulmonary disease; CRP = C-reactive protein; CT = computed tomography; CXR = chest radiograph; EGD = esophagogastroduodenoscopy; EUS = endoscopic ultrasound; GI = gastrointestinal; HIV = human immunodeficiency virus; PTH = parathyroid hormone; TFTs = thyroid function tests; tTG = tissue transglutaminase; U/A = urinalysis.

GI bleeding (Chapter 135) may be acute and clinically apparent (overt) or chronic, slow, and clinically inapparent (occult). The location of acute GI bleeding is described as either upper or lower, according to whether the source is proximal or distal to the ligament of Treitz (distal duodenum). Upper GI bleeding, which is three times more common than lower GI bleeding, is manifested by bloody emesis (hematemesis), coffee ground emesis, and, in most cases, black stools (melena). Common causes of significant bleeding are peptic ulcer disease, esophageal varices, Mallory-Weiss tears, erosive gastritis or esophagitis, and vascular ectasias.[11] Major lower GI bleeding is manifested by large-volume maroon or bright red bloody stools (hematochezia). Although 80 to 90% of patients with hematochezia have a lower source of bleeding, massive upper GI bleeding also may cause hematochezia.

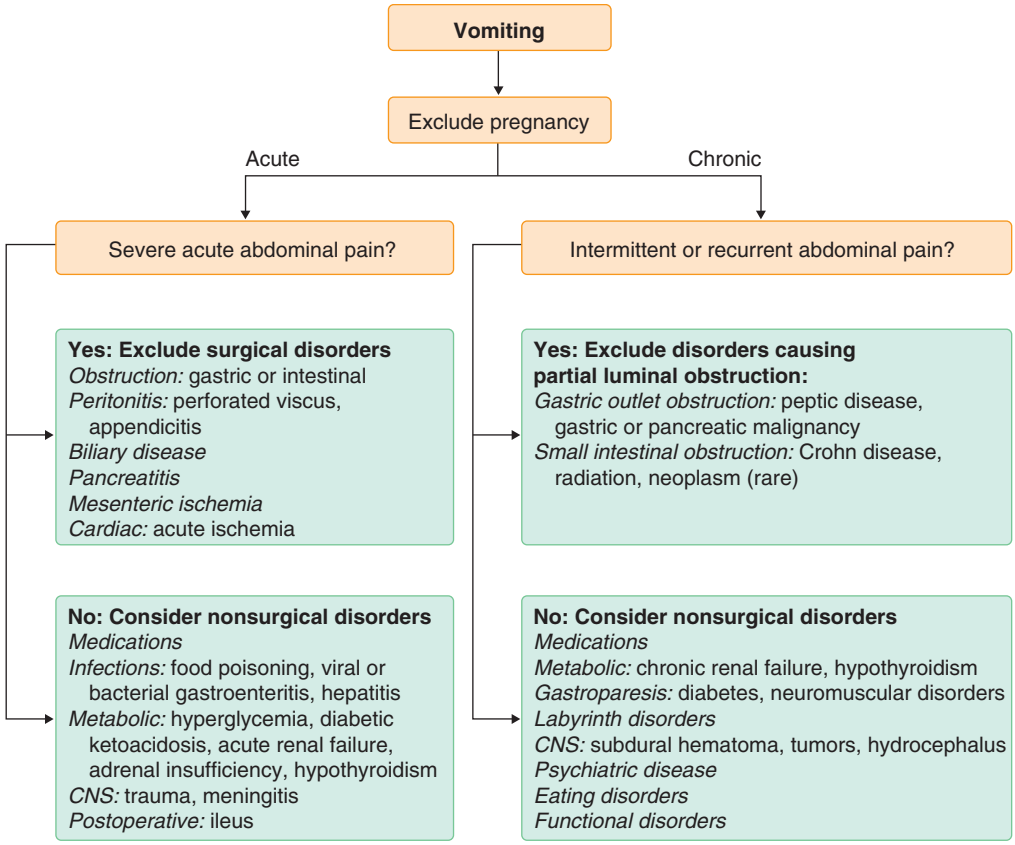

FIGURE 132-5. **Approach to the patient with vomiting.** CNS = central nervous system.

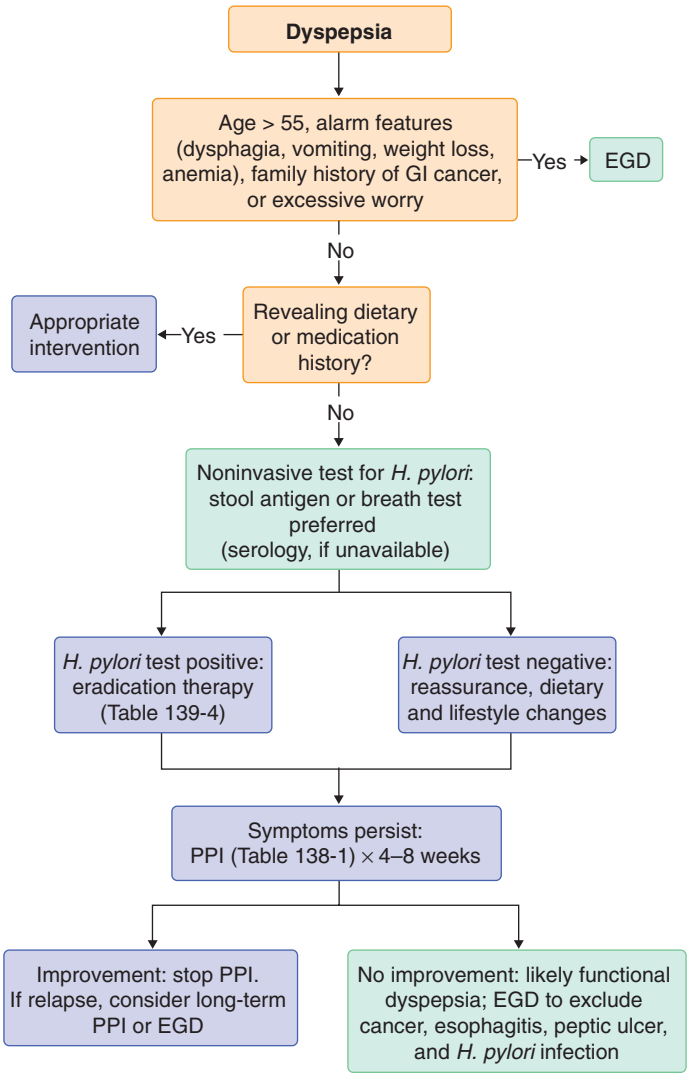

FIGURE 132-6. **Approach to the patient with dyspepsia.** EGD = esophagogastroduodenoscopy; GI = gastrointestinal; PPI = proton pump inhibitor.

TABLE 132-5 MEDICAL TREATMENT OF NAUSEA AND VOMITING

DRUG	USUAL INDICATIONS	USUAL DOSE (RANGE)	ROUTE	COMMENTS
ANTICHOLINERGIC-ANTIHISTAMINE AGENTS				Side effects: sedation, dizziness, delirium, blurred vision, glaucoma, bronchospasm, tachycardia, urinary retention Avoid concomitant alcohol or CNS depressants; use with caution in elderly patients
Scopolamine patch	MS	1.5 mg/72 hr	Patch	
Dimenhydrinate	MS	50 mg (50-100 mg) q4-6h	PO, IM, IV	Maximum 400 mg/24 hr
Cyclizine	MS, GIDz	50 mg q8h	PO, IM	Maximum 200 mg/24 hr
Meclizine	MS, V	25-50 mg q24h	PO	
Diphenhydramine	GIDz	25-50 mg q6h 50-100 mg q6h	PO, IV IM	
Promethazine	GIDz, PONV, MS	25 mg (12.5-25 mg) q6-12h 25 mg (12.5-50 mg) q4-6h	PO, PR IV, IM	Phenothiazine derivative, but lacks significant antidopaminergic effects Avoid perivascular extravasation or subcutaneous injection (severe tissue necrosis)
Trimethobenzamide	GIDz, PONV	200 mg q6-8h	IM	
DOPAMINE RECEPTOR ANTAGONISTS				Side effects: neuromuscular (extrapyramidal) symptoms—agitation, restlessness, involuntary movements, dystonia, torticollis, laryngospasm, Parkinson-like features
Prochlorperazine	GIDz, PONV, CTX	5-10 mg q6-8h 25 mg q12h	PO, IV, IM PR	Maximum dose 20-40 mg/24 hr; avoid subcutaneous injection (irritation)
Metoclopramide	GIDz CTX	10 mg (10-20 mg) q6-8h 1-2 mg/kg before and 2 hr after CTX	PO, IV, IM IV	Modest efficacy at these doses High doses infrequently used owing to availability of safer, more effective CTX regimens; use with diphenhydramine to reduce adverse side effects
Droperidol	PONV	2.5 mg (1.25-5 mg) preinduction and q4-6h as needed	IV, IM	May cause QTc prolongation and torsades de pointes; use is restricted to patients who fail to respond to other agents
CORTICOSTEROIDS				
Dexamethasone	PONV CTX	4-8 mg once preinduction 8-20 mg on day 1; 8 mg on days 2-4	PO, IV PO, IV	Most beneficial when used with other agents (e.g., 5-HT$_3$ RA, neurokinin-1 RA)
BENZODIAZEPINES				Used to reduce anxiety and anticipatory vomiting
Lorazepam	CTX	1-2 mg q4-6h	PO, IV	
CANNABINOIDS				May stimulate appetite; adverse side effects (sedation, dizziness, dysphoria, dry mouth) limit use
Dronabinol	GIDz, CTX	5-10 mg q6-8h	PO	
Nabilone	GIDz, CTX	1-2 mg q12h	PO	
5-HT$_3$ RECEPTOR ANTAGONISTS	PONV, CTX			PONV prevention: give IV immediately before anesthesia induction Prevention of CTX-induced vomiting: give 30 min (IV) to 1 hr (PO) before chemotherapy
Ondansetron	PONV CTX, RadTx	4 mg once 4-8 mg 8 mg once, 8 mg twice daily	IV PO PO	
Granisetron	CTX, RadTx	1 mg twice daily 1 mg once	PO IV	
Dolasetron	CTX, PONV	100 mg once daily	PO only	
Palonosetron	CTX PONV	0.25 mg once 0.5 mg 0.075 mg	IV PO, 1-3 days IV	
NEUROKININ-1 RECEPTOR ANTAGONISTS	Highly emetogenic CTX			Used exclusively in combination with a 5-HT$_3$ RA or dexamethasone
Aprepitant		125 mg on day 1 80 mg on days 2-3	PO	
Fosaprepitant		150 mg on day 1	IV	Aprepitant 80 mg PO on days 2-3
ANTIEMETIC REGIMENS FOR CHEMOTHERAPY				
Mildly emetogenic CTX	Option 1 Option 2	Dexamethasone 8 mg Dopamine receptor antagonist	IV or PO	One dose only One dose only
Moderately emetogenic CTX		Day 1: 5-HT$_3$ RA plus dexamethasone 8 mg	IV or PO	Days 2-3: continue oral 5-HT$_3$ RA and dexamethasone 8 mg to reduce delayed emesis
Highly emetogenic CTX		Day 1: 5-HT$_3$ RA plus dexamethasone 12 mg plus neurokinin-1 RA	IV or PO	Give aprepitant 80 mg PO days 2-3 and dexamethasone 8 mg PO days 2-4 to reduce delayed emesis

CNS = central nervous system; CTX = chemotherapy; GIDz = gastrointestinal disorders associated with nausea and vomiting; 5-HT$_3$ = serotonin or 5-hydroxytryptamine$_3$; MS = motion sickness; PONV = postoperative nausea and vomiting; PR = per rectum; RA = receptor antagonist; RadTx = radiation therapy–induced nausea and vomiting; V = vertigo.
Modified from Proctor DD. Approach to the patient with gastrointestinal disease. In: Goldman L, Ausiello D, eds. *Cecil Textbook of Medicine.* 23rd ed. Philadelphia: Saunders-Elsevier; 2008.

Approximately 95% of major lower GI bleeding arises from the colon and 5% from the small intestine. Lower GI bleeding is increased in patients older than 50 years, in whom diverticulosis accounts for 60% of cases; the remainder are due to ischemia, neoplasms, ulcers, vascular ectasias, or hemorrhoids. In patients younger than 50 years, bleeding is more commonly attributable to inflammatory bowel disease, hemorrhoids, or infectious colitis.

Occult GI bleeding refers to GI blood loss that is small in volume and not apparent to the patient but is detectable by tests for fecal occult blood. Chronic occult bleeding may result in iron deficiency anemia. Both upper endoscopy and colonoscopy should be performed to look for a source of occult bleeding, most commonly gastroesophageal or colonic neoplasia, erosive esophagitis or gastritis, ulcer disease, or vascular ectasia. In patients with recurrent iron deficiency and occult blood loss in whom no source is found on upper and lower endoscopy, video capsule endoscopy or enteroscopy is performed to look for a small bowel source (vascular ectasia, ulcer, or neoplasm).

Fecal incontinence (Chapter 145) is dependent on a number of factors, including a solid or semisolid stool, a compliant and distensible rectal reservoir, the ability to sense rectal fullness, an intact internal anal sphincter (an involuntary muscle innervated by the enteric nervous system), an intact external anal sphincter and puborectalis (voluntary muscles innervated by the pudendal nerve), and the mental and physical ability to reach a toilet facility when needed.[12] Minor incontinence, which occurs in 10% of people older than 70 years, is characterized by the inability to control flatus or by the seepage of fecal matter that results in soiling of the perianal area and undergarments. It tends to be intermittent, occurring after bowel movements; when coughing, lifting, or passing flatus; or when stools are loose. Major incontinence is characterized by the partial or complete inability to reliably control bowel movements, resulting in gross, involuntary loss of feces and the need to wear a diaper. It occurs in less than 1% of the population and is virtually always caused by a central nervous system disorder that results in diminished awareness of bowel needs, neuropathy, or damage to the anal sphincters.

Grade A References

A1. Kim K, Kim YH, Kim SY, et al. Low-dose abdominal CT for evaluating suspected appendicitis. *N Engl J Med*. 2012;366:1596-1605.

A2. Begtrup LM, Engsbro AL, Kjeldsen J, et al. A positive diagnostic strategy is noninferior to a strategy of exclusion for patients with irritable bowel syndrome. *Clin Gastroenterol Hepatol*. 2013;11:956-962.

A3. Biesiekierski JR, Peters SL, Newnham ED, et al. No effects of gluten in patients with self-reported non-celiac gluten sensitivity after dietary reduction of fermentable, poorly absorbed, short-chain carbohydrates. *Gastroenterology*. 2013;145:320-328.

A4. Halmos EP, Power VA, Shepherd SJ, et al. A diet low in FODMAPs reduces symptoms of irritable bowel syndrome. *Gastroenterology*. 2014;146:67-75.

A5. dos Santos LV, Souza FH, Brunetto AT, et al. Neurokinin-1 receptor antagonists for chemotherapy-induced nausea and vomiting: a systematic review. *J Natl Cancer Inst*. 2012;104:1280-1292.

A6. Mazzoleni LE, Sander GB, Francesconi CF, et al. *Helicobacter pylori* eradication in functional dyspepsia: HEROES trial. *Arch Intern Med*. 2011;171:1929-1936.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. *Nat Rev Genet*. 2012;13:260-270.
2. Farré R, Tack J. Food and symptom generation in functional gastrointestinal disorders: physiological aspects. *Am J Gastroenterol*. 2013;108:698-706.
3. Baron TH, Kamath PS, McBane RD. Management of antithrombotic therapy in patients undergoing invasive procedures. *N Engl J Med*. 2013;368:2113-2124.
4. Bredenoord AJ. Management of belching, hiccups, and aerophagia. *Clin Gastroenterol Hepatol*. 2013;11:6-12.
5. Gibson PR, Shepherd SJ. Food choice as a key management strategy for functional gastrointestinal symptoms. *Am J Gastroenterol*. 2012;107:657-666.
6. Menees SB, Maneerattannaporn M, Kim HM, et al. The efficacy and safety of rifaximin for the irritable bowel syndrome: a systematic review and meta-analysis. *Am J Gastroenterol*. 2012;107:28-35.
7. Furyk JS, Meek R, McKenzie S. Drug treatment of adults with nausea and vomiting in primary care. *BMJ*. 2014;349:g4714.
8. Yang JC, Lu CW, Lin CJ. Treatment of infection: current status and future concepts. *World J Gastroenterol*. 2014;20:5283-5293.
9. DuPont HL. Acute infectious diarrhea in immunocompetent adults. *N Engl J Med*. 2014;370:1532-1540.
10. Bharucha AE, Dorn SD, Lembo A, et al. American Gastroenterological Association medical position statement on constipation. *Gastroenterology*. 2013;144:211-217.
11. Greenspoon J, Barkun A, Bardou M, et al. Management of patients with nonvariceal upper gastrointestinal bleeding. *Clin Gastroenterol Hepatol*. 2012;10:234-239.
12. Shah BJ, Chokhavatia S, Rose S. Fecal incontinence in the elderly: FAQ. *Am J Gastroenterol*. 2012;107:1635-1646.

REVIEW QUESTIONS

1. A 60-year-old man reports the recent onset of mild to moderate epigastric pain during the last few weeks. The pain is unrelated to meals and does not radiate. He has had some decrease in appetite but no nausea, vomiting, weight loss, or dysphagia. He denies ingestion of any aspirin or nonsteroidal anti-inflammatory drugs. He tried taking an over-the-counter H_2-receptor antagonist (ranitidine) without symptomatic improvement. Physical examination findings are normal except for mild tenderness in the epigastrium. A stool specimen is negative for occult blood. Which of the following is the *most* appropriate initial diagnostic test?

A. Upper gastrointestinal endoscopy
B. *H. pylori* serology
C. Abdominal computed tomography scan
D. Ambulatory pH/impedance test
E. An upper gastrointestinal series of radiographs, with a barium swallow

Answer: A The patient has dyspepsia (i.e., mild to moderate epigastric pain or discomfort), which is of recent onset but is not associated with any "alarm features." Current guidelines recommend upper gastrointestinal endoscopy in patients with recent onset of dyspepsia who have alarm features or who are older than 55 years because of an increased likelihood of significant medical disorders, including peptic ulcer disease, gastroesophageal reflux disease, and malignant disease. Gastric biopsy specimens should be obtained to exclude *H. pylori* infection, which may cause dyspepsia and peptic ulcer disease. In patients without alarm features who are younger than 55 years, testing for *H. pylori* should be performed with a stool antigen assay or a urea breath test, and eradication therapy should be given if the test result is positive. Patients who do not have *H. pylori* infection or whose symptoms persist after *H. pylori* eradication may be treated empirically with a proton pump inhibitor for 4 to 8 weeks. Patients with persistent symptoms should undergo upper gastrointestinal endoscopy. (Talley NJ, Vakil NB, Moayyedi P. American Gastroenterological Association Technical Review on the evaluation of dyspepsia. *Gastroenterology.* 2005;129:1756-1780.)

2. A 65-year-old man reports progressive unintentional weight loss of 25 pounds (10% of body weight) during the past 12 months. He reports no dysphagia, nausea, vomiting, or jaundice. His appetite remains good. He reports loose stools and excessive, foul-smelling flatus. He has a history of diabetes mellitus, for which he takes metformin and insulin. He has a prior history of heavy ethanol use but has been abstinent for several years. A physical examination is unrevealing, including a nontender abdomen and no organomegaly or abnormal masses. Initial blood work including a compete blood count, electrolytes, and liver chemistries is normal. Which of the following is the *most likely* cause of the patient's weight loss?

A. Metformin
B. Chronic pancreatitis
C. Colon cancer
D. Diabetic gastroparesis
E. Pancreatic cancer

Answer: B The patient's unintentional weight loss of more than 5% of his former body weight is significant and warrants evaluation (see Fig. 132-4). The presence of loose stools and increased flatus is suggestive of malabsorption. Given his history of heavy alcohol use, chronic pancreatitis resulting in malabsorption and diabetes mellitus is the most likely cause of weight loss. Metformin can cause a number of gastrointestinal symptoms, including dyspepsia and diarrhea, but it does not cause malabsorption. Although the patient could have an underlying malignant neoplasm, nonmetastatic colon cancer usually does not cause significant weight loss, and malabsorption is not a typical finding with pancreatic cancer. Although gastroparesis can cause weight loss, the patient has no early satiety or vomiting to suggest this diagnosis. (Forsmark CE. Management of chronic pancreatitis. *Gastroenterology.* 2013;144:1282-1291.)

3. A 73-year-old woman reports the onset of a constant aching, nonradiating left lower quadrant pain 2 days ago. She has continued to eat without nausea or vomiting, but her appetite is poor. She has not had a bowel movement for 2 days but has continued to pass flatus without relief of her discomfort. She has had no fever or chills. Her past medical history is noncontributory. She has chronic constipation but has noted no recent changes in her bowel habits. On physical examination, she is alert and moves without obvious pain. She has a normal blood pressure and pulse, with a temperature of 38° C. Bowel sounds are present but diminished. She has no tenderness to cough or percussion. On light and deep palpation, tenderness is noted in the left lower quadrant with some guarding. A mass is not palpable on abdominal or pelvic examination. Digital rectal examination is normal. Her white blood cell count is 12,000/μL. Which is the *most likely* cause of abdominal pain?

A. Acute appendicitis
B. Diverticulitis
C. Colon cancer
D. Ovarian torsion
E. Pancreatic cancer

Answer: B The patient presents with 2 days of left lower quadrant pain without signs or peritoneal irritation (see Table 132-2 and Fig. 132-1). Although appendicitis must be considered, it uncommonly is manifested with pain in the left lower quadrant. An obstructing cancer in the distal colon can cause acute large bowel obstruction, but the absence of a change in her bowel habits and the presence of acute, localized pain make this diagnosis unlikely. Ovarian torsion can present at any age, usually with a more sudden onset of intense pain. The absence of a mass on pelvic examination makes this diagnosis unlikely. Pancreatic cancer pain usually has a more insidious onset with gradual progression. Diverticulitis usually is manifested with left lower quadrant pain because diverticulosis is more common in the left colon. The pain, which may be intermittent or constant, commonly is associated with obstipation or diarrhea. Diverticulitis may be associated with localized inflammation and microperforation, with formation of an inflammatory mass and abscess, or with free perforation and peritonitis. Abdominal CT with oral and intravenous administration of contrast material has a high sensitivity and specificity (>90%) for the diagnosis of diverticulitis and helps exclude other cause of acute abdominal pain. (Biondo BS, Lopez Barao J, Millan M, et al. Current status of the treatment of acute colonic diverticulitis: a systematic review. *Colorectal Dis.* 2012;14:e1-e11.)

4. A 35-year-old woman reports chronic bloating and "gas" for several years. On closer questioning, she reports one to three stools per day that vary in consistency from well-formed to loose, without mucus or visible blood. She also reports a large amount of flatus. She denies nausea, heartburn, belching, vomiting, or abdominal pain. Her appetite is excellent, and her weight is stable. She has no significant prior medical history and takes no medications. She is a vegetarian and avoids all dairy products. Physical examination findings are normal. Laboratory data including complete blood count, albumin, C-reactive protein, and anti–tissue transglutaminase antibody are normal. A qualitative analysis of a stool specimen is negative for fecal fat. The *most likely* cause of this woman's symptoms is

A. Irritable bowel syndrome
B. Crohn ileitis
C. Celiac disease
D. Carbohydrate maldigestion
E. Lactose intolerance

Answer: D Excessive flatus production is caused by either swallowed air or intestinal (usually colonic) bacterial fermentation of undigested carbohydrates. A large variety of short-chain carbohydrates, known as FODMAPs, are present in varying amounts in fruits, vegetables, and grains and can lead to increased colonic gas production in otherwise healthy people. A careful dietary history of foods containing these FODMAPs should be performed.[1] Patients with disorders causing food malabsorption, such as celiac disease, chronic pancreatitis, and Crohn disease, may also have increased flatus due to malabsorption of fats and carbohydrates. The negative anti–tissue transglutaminase antibody is strong evidence against celiac disease. The lack of abdominal pain and normal laboratory data (including inflammatory markers) makes Crohn disease unlikely. Although irritable bowel syndrome may be associated with complaints of bloating, the hallmark of this entity is abdominal pain, which the patient denies. She avoids all dairy products, so lactose intolerance is an unlikely cause. (Gibson PR, Shepherd SJ. Food choice as a key management strategy for functional gastrointestinal symptoms. *Am J Gastroenterol.* 2012;107:657-666.)

5. A 35-year-old woman seeks evaluation for abdominal pain that has been present for several years. The pain, which is diffuse and relatively constant, is unrelated to menses, meals, or bowel movements. She has no associated nausea, vomiting, weight loss, diarrhea, or constipation. She reports chronic fatigue, headaches, and intermittent myalgias. Her menses are irregular. She has a history of depression but currently is not taking medications. She has seen several other providers and reports having undergone upper gastrointestinal endoscopy, colonoscopy, ambulatory pH study, abdominal ultrasound, and abdominal computed tomography scan, all of which have been unrevealing. On physical examination, she has diffuse abdominal tenderness without guarding or signs of peritoneal irritation. She has no organomegaly or abnormal masses. Laboratory data are normal, including complete blood count, electrolytes, liver chemistries, C-reactive protein, and thyroid-stimulating hormone. The *most likely* cause of this patient's abdominal pain is

A. Crohn disease
B. Acute porphyria
C. Functional pain syndrome
D. Endometriosis
E. Gallstones

Answer: C Chronic abdominal pain that is present for months to years can be caused by structural (organic) disorders, but such causes are usually apparent after careful medical evaluation. Common causes of chronic pain include gastroesophageal reflux, peptic ulcer disease, inflammatory bowel disease, chronic pancreatitis, biliary tract disease, gastrointestinal malignant neoplasms, and endometriosis. With these entities, the pain usually is localized and intermittent. Crohn disease usually is associated with altered bowel habits, anemia, and elevated inflammatory markers; it is almost always demonstrated by endoscopy or abdominal cross-sectional imaging. Acute porphyria may cause acute abdominal pain and neuropsychiatric features, but the attacks are intermittent. Endometriosis usually causes pelvic pain that is intermittent and worsened before menses. This patient's pain strongly suggests a functional origin: it is constant and diffuse, it is unrelated to physiologic functions (i.e., eating or bowel movements), and it is associated with other chronic somatic complaints (fatigue, myalgias, depression). The patient has already had an extensive but unrevealing evaluation for organic disorders. Further studies are extremely unlikely to be of benefit. The physician should endeavor to shift her focus from finding the cause of the pain to coping with her chronic illness. A therapeutic patient-provider relationship—which includes a nonjudgmental attitude, reassurance, empathy, education, and an assessment of psychosocial contributors, fears, and cultural beliefs—is essential to helping her adjust and improve. (Clouse RE, Mayer EA, Aziz Q, et al. Functional abdominal pain syndrome. *Gastroenterology.* 2006;130:1492-1497.)

133

DIAGNOSTIC IMAGING PROCEDURES IN GASTROENTEROLOGY

DAVID H. KIM AND PERRY J. PICKHARDT

A wide range of diagnostic imaging modalities are available for evaluation of diseases of the gastrointestinal (GI) tract and the hepatopancreaticobiliary system. Once the workhorses of GI radiology, conventional radiography and fluoroscopy are still relevant but have largely given way to more advanced cross-sectional imaging studies, such as ultrasonography, computed tomography (CT), and magnetic resonance imaging (MRI). Many of the visceral vascular evaluations undertaken by conventional angiography have been replaced by these noninvasive modalities as well. These cross-sectional technologies have become the preferred methods of evaluation, allowing more precise and accurate diagnoses. In addition, cross-sectional techniques can be used to guide a wide variety of interventional procedures. With the emergence of molecular imaging, there has been renewed interest in nuclear medicine, most notably positron emission tomography (PET) and the combination modalities of PET/CT and PET/MR.

CONVENTIONAL RADIOGRAPHY

Conventional radiographs, often referred to as plain films, remain useful for a limited number of abdominal indications but are generally much less sensitive and specific for disease compared with techniques such as CT. Advantages of radiography include its wide availability, low cost, and portability, allowing the acquisition of images in acute clinical situations. Supine and upright frontal abdominal radiographs can assess rapidly for bowel obstruction or perforation in the setting of an acute abdomen (Fig. 133-1). Serial abdominal radiographs remain a practical approach for observing patients

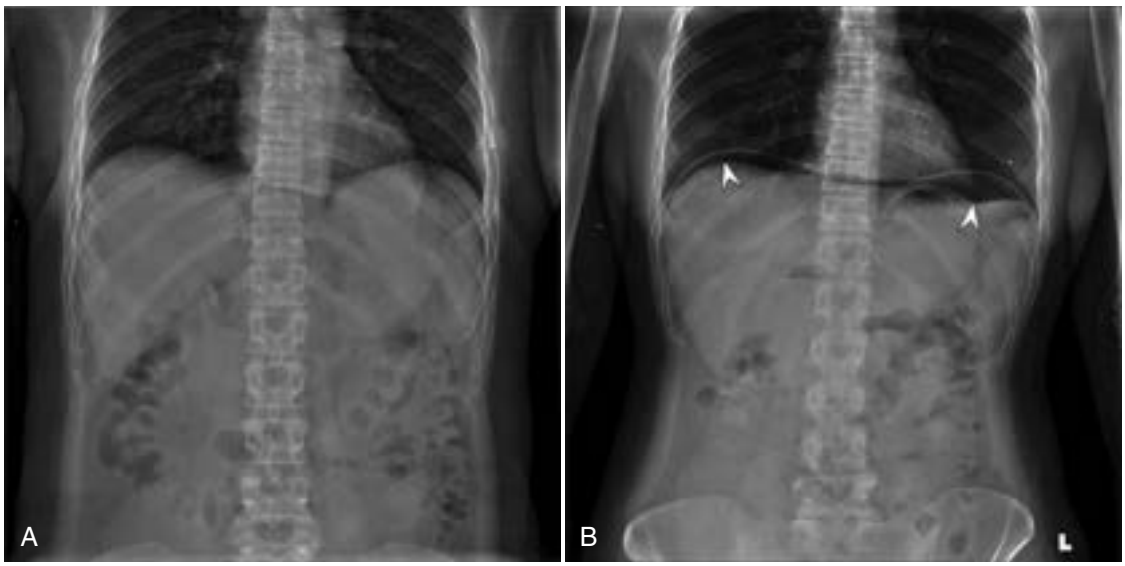

FIGURE 133-1. Pneumoperitoneum from bowel perforation on conventional radiography. A 36-year-old renal transplant patient presents with abdominal pain after colonoscopy. **A,** Supine abdominal radiograph is grossly normal with scattered nondistended bowel. **B,** Upright radiograph centered over the diaphragm reveals a lucent area below the diaphragm (*arrowheads*) consistent with free intraperitoneal air. This view is required for the detection of pneumoperitoneum because the supine examination does not show free air unless it is present in large amounts.

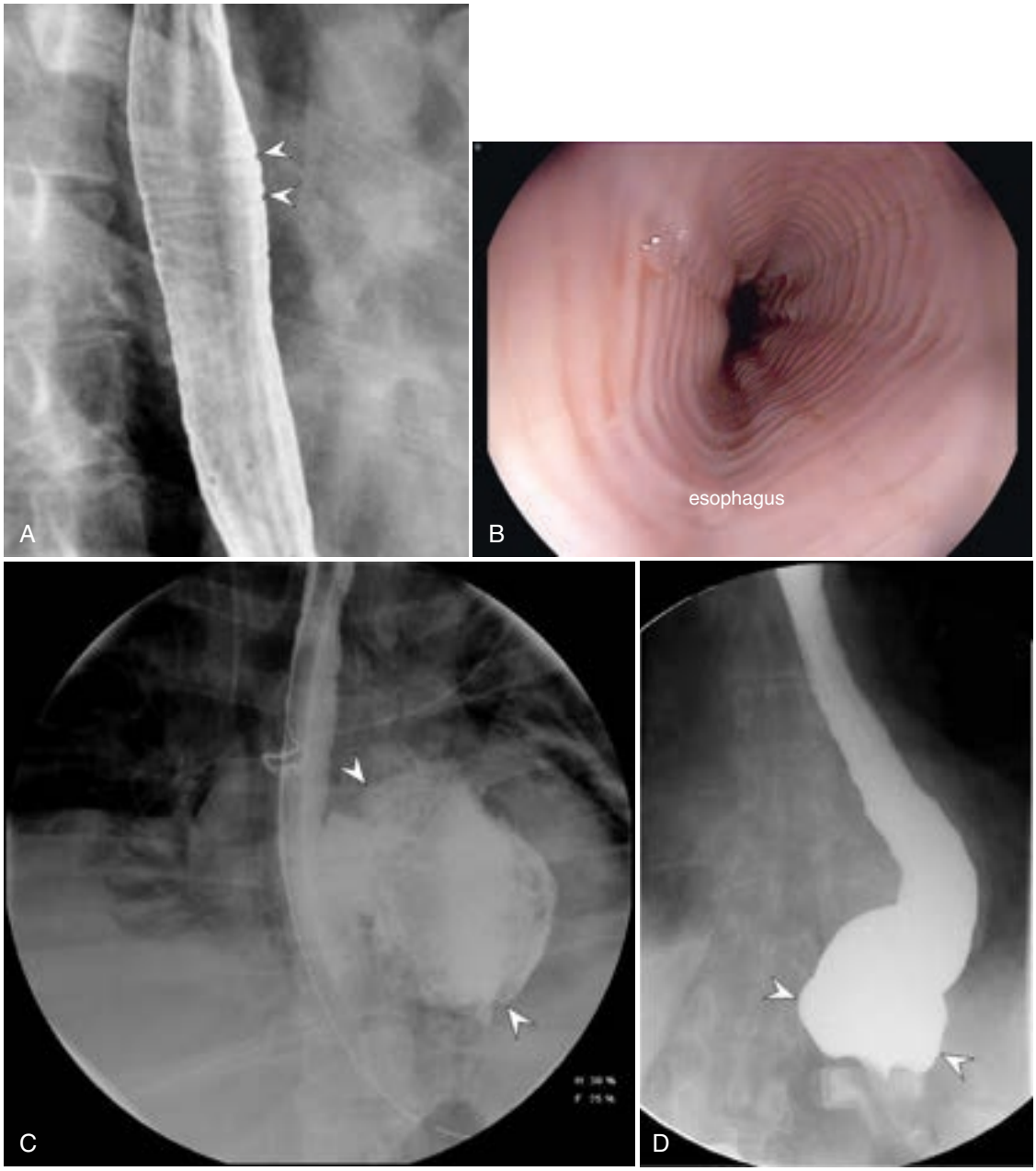

FIGURE 133-2. Evolving nature of contrast fluoroscopy. A, Double-contrast image of the esophagus shows fine mucosal constrictions or rings (*arrowheads*) in a young adult man with dysphagia suggestive of eosinophilic esophagitis. This imaging approach is now used less commonly. **B,** Digital photograph from esophagogastroduodenoscopy confirms the diagnosis (numerous eosinophils at biopsy). **C,** Single-column esophagography is a more modern radiographic technique to confirm a leak (*arrowheads*) of water-soluble contrast material from the esophagus after trauma or to demonstrate obstruction and dilation (**D,** *arrowheads*) of the esophagus and proximal stomach upstream of an overly tight gastric band placed for weight control.

with an abnormal bowel gas pattern suggestive of either evolving small bowel obstruction or adynamic ileus.[1] Conventional radiographs can demonstrate abnormal abdominal calcifications and radiopaque foreign bodies. In each of these cases, however, cross-sectional modalities, such as CT, have increased sensitivity and provide better delineation of disease processes. CT is often undertaken when the findings on initial plain film evaluation are normal or to provide better information when the findings on conventional radiography are abnormal.

FLUOROSCOPIC PROCEDURES

The role of contrast fluoroscopy has changed dramatically during the past two decades.[2] Double-contrast radiographic images (Fig. 133-2) to depict mucosal details of the esophagus, stomach, small bowel, and colon have largely been supplanted by endoscopy (Chapter 134) and advanced radiologic techniques. The diagnoses of an erosion, ulcer, polyp, or mass are now largely in the domain of endoscopy, supplemented by the newer cross-sectional modalities, such as CT or MR enterography for the small bowel (replacing the double-contrast small bowel enteroclysis) and CT colonography (replacing the double-contrast barium enema).

Single-column contrast images to depict overall anatomic structure and for problem-solving now constitute the bulk of fluoroscopic studies. In single-column examinations, the luminal structure of concern is distended by contrast material only, either thin barium or a water-soluble iodinated contrast agent, to delineate the gross structure and without trying to determine fine mucosal detail. In the esophagus and stomach, single-column luminal examinations are commonly used to assess for postoperative anastomotic breakdown after esophagectomy, gastrectomy, or bariatric surgery (see Fig 133-2). In the small bowel, single-contrast examinations (i.e., small bowel series/follow-through) can be used preoperatively as anatomic "road maps" to delineate fistulas or to guide bowel resection in Crohn disease. For the colorectum, the single-contrast barium enema remains an important diagnostic tool in such settings as suspected colonic obstruction, postoperative leak or fistula, and ileocolic intussusception in children (Chapter 142). Many institutions continue to perform fluoroscopic defecography to help delineate functional abnormalities in patients with evacuation disorders (Chapter 145), although dynamic MR cine series have replaced defecography at some institutions.

In addition to these single-column examinations, videofluoroscopy remains a mainstay for evaluation of swallowing problems. In this examination, the

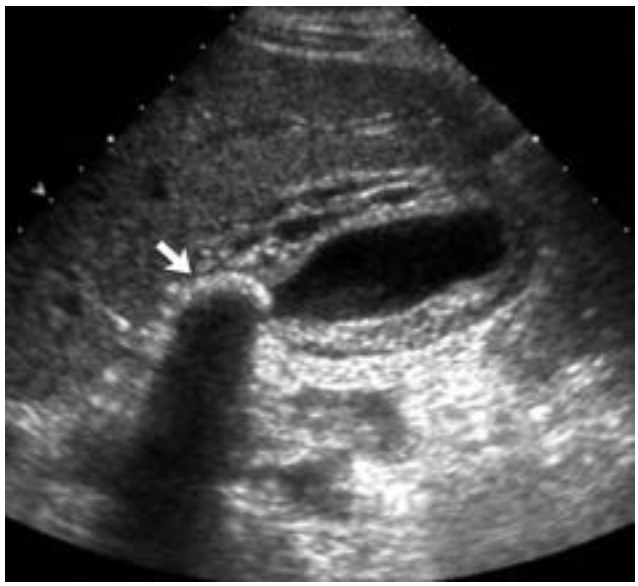

FIGURE 133-3. Acute cholecystitis on ultrasound examination. Image from right upper quadrant sonography shows diffuse gallbladder wall thickening and a shadowing impacted gallstone (*arrow*). A sonographic Murphy sign was present. These findings are diagnostic for acute calculous cholecystitis.

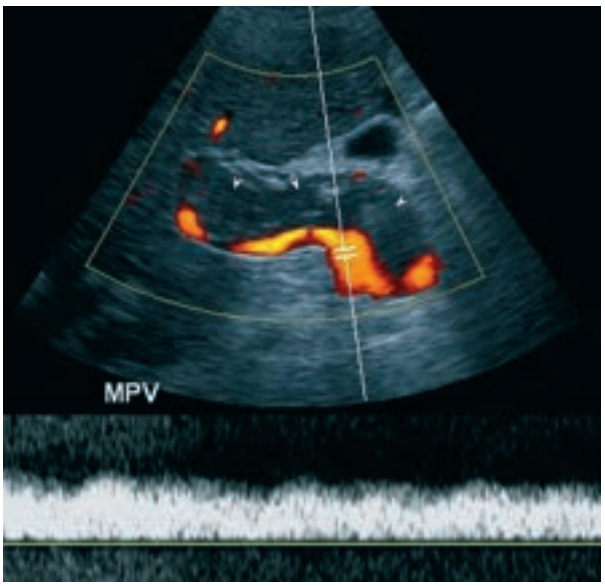

FIGURE 133-4. Portal vein thrombosis on ultrasound examination. Ultrasound gray-scale image with both power color Doppler and spectral Doppler interrogation shows a tubular hypoechoic structure (*arrowheads*) consistent with nonocclusive thrombus filling the majority of the main portal vein (MPV). Flow patency is seen in the deep peripheral aspect of the vessel.

patient swallows varying consistencies of barium, ranging from thin liquids to solids, typically in the form of a barium cookie. This test is an excellent way to assess the swallowing mechanism dynamically and to exclude aspiration.

ULTRASONOGRAPHY

The introduction of harmonic and compound imaging, advances in high-resolution transducers, and improvements in color Doppler evaluation have all combined to enhance the diagnostic capabilities of portable ultrasound. In general, ultrasound is useful for imaging solid organs and fluid-filled structures, but it is unable to penetrate gas-filled structures. For example, overlying bowel gas often precludes a complete sonographic evaluation of the pancreas. Ultrasound is a relatively versatile imaging technique in that it can be performed by many different routes, including transabdominal, endoscopic (as part of esophagogastroduodenoscopy), transrectal, intravascular, and endovaginal approaches. In addition, it is excellent for many image-guided interventions because of its real-time capabilities.

Ultrasonography is the most frequently used initial modality to evaluate the liver and biliary system. Suspected acute cholecystitis (Chapter 155) in the setting of right upper quadrant pain is a common indication for right upper quadrant sonography; classic findings include cholelithiasis, gallbladder wall thickening, and a sonographic Murphy sign (reproducible pain when the transducer is pressed over the gallbladder) (Fig. 133-3). The sensitivity for detection of gallstones with ultrasonography exceeds 95%. Acalculous cholelithiasis can be a more challenging diagnosis because the findings overlap with nonspecific gallbladder wall thickening in critically ill patients. Ultrasound is typically the first imaging test obtained in patients with new-onset jaundice or cholestatic laboratory findings because it offers a rapid, noninvasive evaluation of the biliary tree to differentiate obstruction from other causes.[3] If biliary ductal dilation is present, the level and cause of the obstruction can sometimes be demonstrated on ultrasound; common causes include choledocholithiasis and pancreatic head masses. In most cases of biliary obstruction, additional imaging tests will be necessary, consisting of CT, MR cholangiopancreatography (MRCP), endoscopic retrograde cholangiopancreatography, or percutaneous transhepatic cholangiography, depending on the specific circumstances.

Although ultrasonography is typically less sensitive and specific than CT or MRI for the detection or characterization of focal liver lesions (Chapters 151 and 196), it is useful for distinguishing cystic lesions from solid lesions. Although they are not approved for use in the United States, intravenous contrast agents for ultrasound have been studied fairly extensively in other countries and appear to offer similar advantages seen with CT and MRI contrast agents.

In diffuse liver disease, ultrasound is an alternative to CT or MR in screening of patients with viral hepatitis for possible cirrhosis and hepatocellular carcinoma (Chapters 153 and 196). Although it is less sensitive for hepatocellular carcinoma, ultrasound holds the advantages of being relatively inexpensive, convenient, and not requiring ionizing radiation. Sonographic findings in cirrhosis include a heterogeneously coarsened parenchymal echotexture, nodular surface contour, predominantly right-sided volume loss, and evidence of portal hypertension, including ascites, splenomegaly, and portosystemic collaterals. Ultrasound elastography uses either an acoustic pulse or mechanical vibration to create a shear wave that can assess the stiffness of the liver as a nonspecific measurement of the degree of liver fibrosis.[4] Studies have shown good agreement between this technique and histology, particularly at the extremes (none and severe fibrosis). Recently, the FibroScan or transient elastrography has been used for the longitudinal evaluation of liver fibrosis, but it is not yet ready for integration with routine clinical practice.

Ultrasonography can be used to detect hepatic steatosis (fatty liver; Chapter 152) when the parenchyma demonstrates increased echogenicity and decreased penetration of the sound beam. The findings of steatosis can be focal, multifocal, or diffuse; ultimately, MRI is more specific and can confirm the diagnosis.

Color and power Doppler evaluation allows the noninvasive sonographic assessment of vascular patency. Doppler evaluation of the liver is commonly performed in patients with end-stage liver disease (Chapter 154) to evaluate the portal system and to search for portosystemic collaterals. Abnormal portal vein findings include hepatofugal flow and thrombosis (Fig. 133-4). Doppler ultrasound is also used for the evaluation of transjugular intrahepatic portosystemic shunts (TIPS), both before and after stent placement. In orthotopic liver transplant recipients, Doppler evaluation is frequently performed to assess the hepatic vasculature, with particular attention to the hepatic arterial supply.

COMPUTED TOMOGRAPHY

CT has revolutionized the imaging of abdominal disease, providing a rapid, reproducible, and comprehensive evaluation. The introduction of single-detector helical or spiral CT, followed by multidetector scanners, has resulted in improved resolution and faster acquisition of true volumetric data. High-resolution scans of the entire abdomen and pelvis can now be easily acquired in a single short breath-hold. With the automated, high-rate injection of intravenous contrast materials and advanced processing, specialized CT examinations are replacing many traditional modalities.

A major challenge of CT scanning, however, is to minimize radiation doses as CT use increases, especially in patients who may undergo these because of the possibility of multiple scans for nonmalignant indications throughout their lifetime.[5] Newer image reconstruction techniques hold the promise of markedly decreasing the dose of radiation while maintaining the fidelity of the CT images.

The clinical indications for abdominal CT are broad. One common use is the diagnostic evaluation of a nontraumatic acute abdomen (Chapter 142). Common inflammatory conditions such as appendicitis and diverticulitis

are diagnosed by CT. Other common indications include evaluation for intra-abdominal abscess, pancreatitis, and small bowel obstruction. In cases of relatively high-grade bowel obstruction, CT can often localize the transition point, elucidate the underlying cause, and evaluate for vascular compromise. In the setting of an acute abdomen due to blunt trauma (Chapter 111), CT has become invaluable for the prompt detection of significant abdominal injury.

In the nonacute setting, multiphase CT with intravenous administration of contrast material can characterize lesions and often results in a noninvasive diagnosis, particularly in combination with the clinical history (Fig. 133-5). Primary abdominal malignant neoplasms, such as hepatocellular carcinoma and pancreatic cancer, are often first detected on CT. CT has become the modality of choice for abdominal staging for metastatic disease, including hematogenous, lymphatic, peritoneal, and local spread, and in assessing the response to various therapies.

A rapidly growing indication of CT is to replace traditional fluoroscopy in the diagnostic evaluation of bowel complaints because of its better sensitivity and specificity. CT enterography protocols combine neutral (i.e., water density) oral contrast agents with dynamic, high-resolution imaging that provides detailed multiplanar evaluation of the small bowel. Indications include evaluation of occult GI bleeding due to small bowel masses and low-grade chronic obstruction. Dedicated CT enterography and capsule endoscopy yield a complementary and comprehensive evaluation of the small bowel. Although it is accurate and useful for the diagnosis of Crohn disease,[6] CT enterography has been largely supplanted by its MR counterpart for follow-up assessments so that radiation exposure can be limited. For the large bowel, CT colonography, also referred to as virtual colonoscopy, combines two- and three-dimensional evaluation of the prepared and distended colon for the detection of colorectal polyps and masses (Fig. 133-6; Video 133-1). CT colonography, which has replaced the double-contrast barium enema at many institutions, can be used to screen for colorectal cancer in average-risk

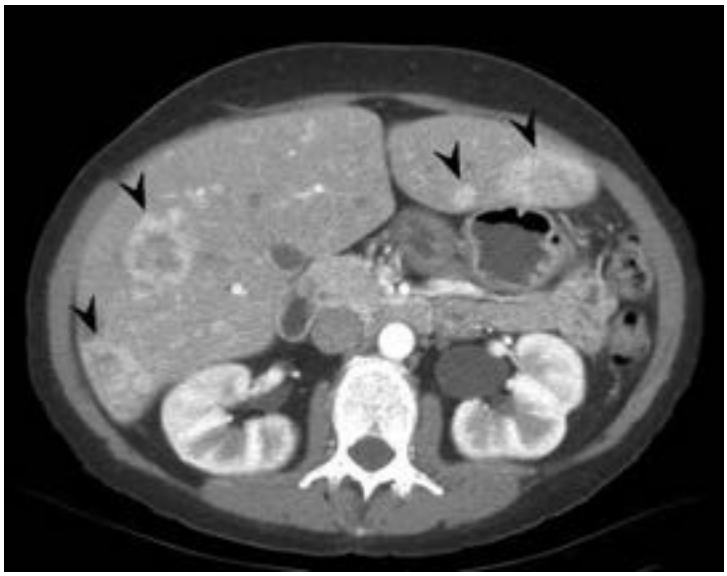

FIGURE 133-5. Multiple hypervascular liver lesions on CT. Dynamic contrast-enhanced CT image obtained during the arterial phase shows multiple hypervascular liver lesions (*arrowheads*), which proved to be hepatic adenomas in a patient with von Gierke disease.

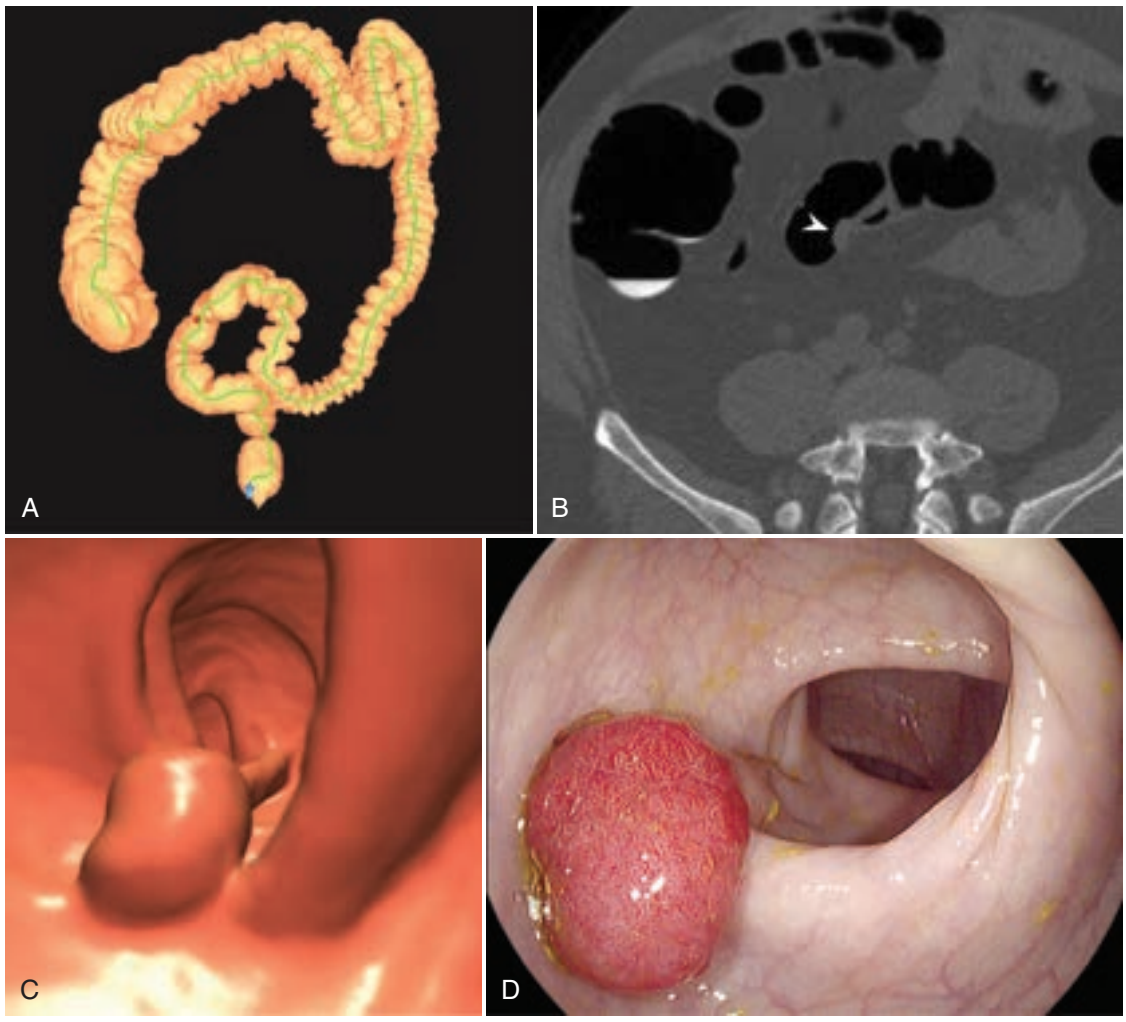

FIGURE 133-6. Pedunculated tubulovillous adenoma on screening virtual colonoscopy (CT colonography). A, Colonic color map allows precise documentation of location of this sigmoid polyp (*red dot*). **B,** Transverse two-dimensional image confirms that the lesion is composed of soft tissue (*arrowhead*). **C,** Three-dimensional endoluminal view nicely matches the pedunculated appearance at colonoscopy (**D**).

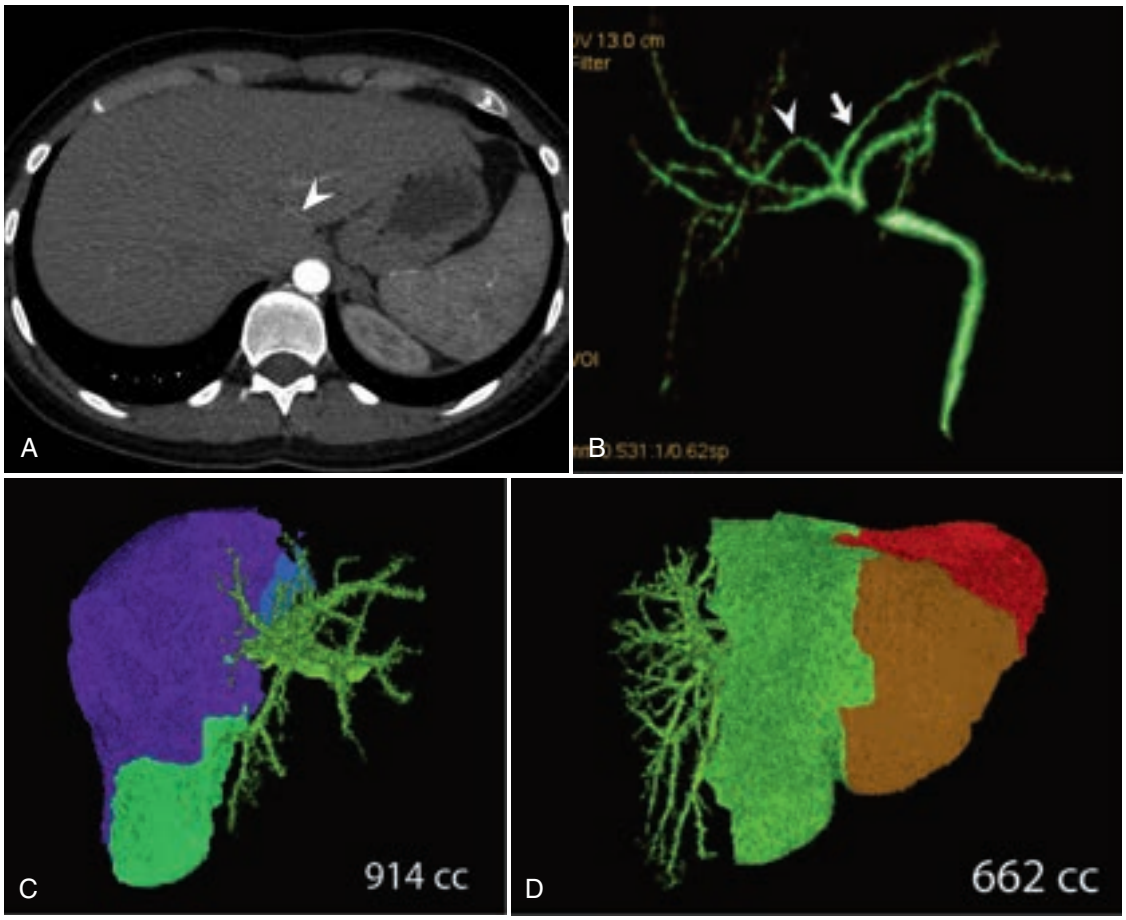

FIGURE 133-7. Preoperative anatomic evaluation of potential living related liver donors. Specialized CT protocols with vascular and cholangiographic contrast can allow complete evaluation. **A,** Transverse CT image shows a small-caliber accessory left hepatic artery (*arrowhead*). Other levels (not pictured) show normally branching right and left intrahepatic arteries at the porta hepatis. **B,** Three-dimensional cholangiographic reconstruction shows an anatomic variant, in which a posterior segment hepatic duct (*arrowhead*) does not join the anterior segment but rather drains anomalously into a left hepatic duct (*arrow*), thereby potentially influencing the choice of the transplant donor. **C** and **D,** Segmented liver volumes allow prediction of hepatic adequacy for both the donor and the recipient.

individuals, with reported sensitivity of 90% for detection of polyps 10 mm and larger. Optical colonoscopy is still required for polypectomy.

The technologic advances in CT drive an ever-expanding number of applications in abdominal imaging. Visceral CT angiography is largely replacing conventional diagnostic angiography. For example, in many institutions, evaluation of the vascular anatomy before hepatic transplantation (Chapter 154) is now undertaken by CT rather than by catheter angiography. Similarly, CT cholangiography can accurately map the biliary system of a living related liver transplant donor, and CT volumetric analysis can help assess whether enough liver will remain after such donation (Fig. 133-7).[7]

MAGNETIC RESONANCE IMAGING

The advantages of MRI over CT for abdominal evaluation include superior soft tissue contrast resolution and lack of ionizing radiation. Drawbacks include decreased spatial resolution, longer examination times, increased expense, decreased availability, and inability to scan some patients due to claustrophobia or implanted devices such as cardiac pacemakers. Imaging artifacts can also make MRI interpretation more difficult and less uniform across different readers. At many institutions, MRI primarily is used to assess a specific known condition. For example, MR is commonly used to characterize a liver lesion seen on another modality or to determine the local staging of a rectal cancer rather than ordered to evaluate nonspecific abdominal pain or to detect possible abdominal abscesses in a patient with fever of unknown origin.

Contrast-enhanced MRI offers a dynamic evaluation comparable to CT for the solid abdominal organs. Its sensitivity and specificity for detection of focal lesions are better than CT when image quality is good. For example, many institutions screen cirrhotic populations for hepatocellular cancer by MR rather than by CT because of its high accuracy and lack of ionizing radiation. MR also is a good examination in assessing response to local ablative therapies (Fig 133-8). Unfortunately, patient factors (e.g., the inability to hold a breath adequately or large amounts of ascites in cirrhotic patients) can often limit the quality of the MR image, thereby decreasing accuracy.

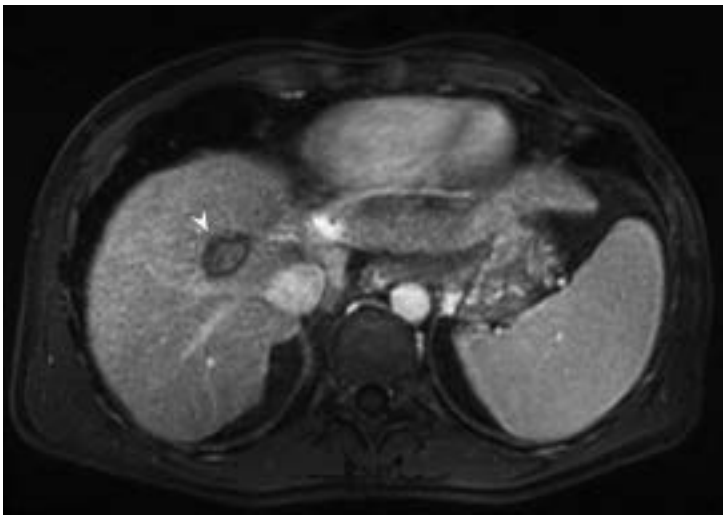

FIGURE 133-8. Surveillance after microwave ablation of hepatocellular carcinoma. Contrast-enhanced fat-suppressed gradient echo MRI shows a bland postablation site (*arrowhead*) with central carbonaceous "char" and no evidence of enhancing recurrence in the periphery.

Intravenous gadolinium-based agents with hepatocyte-specific uptake increase MRI's diagnostic capabilities in evaluating focal hepatic lesions. For example, hypervascular lesions, which retain these agents, can be diagnosed as benign focal nodular hyperplasia without the need for biopsy. Because these agents are excreted into the biliary system, the evaluation of possible biliary disease is also improved. The high accuracy of MRI in diagnosis of hepatic steatosis (Chapter 152) can sometimes prevent unnecessary biopsy, particularly in cases of focal fatty infiltration that simulates metastatic disease.

MRI is also sensitive for detection of iron overload within the liver and other organs related to primary hemochromatosis (Chapter 212) and secondary hemosiderosis (most often due to multiple transfusions). Similar to CT, MRI can provide quality arterial and venous angiographic imaging, such that conventional angiography is generally reserved for therapeutic interventions.

Patients with decreased renal function should not receive gadolinium because of their risk for development of the rare nephrogenic systemic fibrosis (Chapter 267), which is characterized by involvement of the skin, eyes, joints, and internal organs. Whether newer formulations of MR gadolinium-based contrast agents may decrease this risk is under investigation.

Three specialized MR examinations have significantly changed practice patterns in recent years. MRCP, a heavily T2-weighted imaging technique for the noninvasive diagnostic evaluation of the biliary and pancreatic ductal systems, relies not on the administration of contrast material but on the presence of static fluid.[8] MRCP can be a useful screening tool to select appropriate candidates for more invasive therapeutic procedures, such as endoscopic retrograde cholangiopancreatography and percutaneous transhepatic cholangiography (Fig. 133-9). MRCP is useful for diagnosis of biliary and pancreatic ductal obstruction, choledocholithiasis, primary sclerosing cholangitis, and cystic conditions such as Caroli disease. T1-weighted MR cholangiography with intravenous administration of contrast agents that undergo biliary excretion can be useful in evaluating for bile leaks, analogous to hepatobiliary scintigraphy.

MR enterography has become the preferred approach for assessment of disease activity in patients with Crohn disease (Chapter 141). Given the lack of ionizing radiation, it is particularly advantageous in young patients who require multiple examinations over a lifetime. Oral contrast agents such as polyethylene glycol and nonabsorbable low-concentration barium preparations are given to distend the small bowel; spasmolytics are typically administered to decrease bowel peristalsis. Similar to CT, fast breath-held imaging with intravenous administration of contrast material allows evaluation of mucosal and wall enhancement or thickening, suggesting active disease. Unlike CT, MRI can assess intrinsic signal characteristics on T2-weighted images to improve specificity and to distinguish active inflammation from chronic fibrostenotic disease (Fig. 133-10).

The use of neoadjuvant rather than adjuvant chemoradiation for stage II to stage IV rectal cancer has made MR staging an important preoperative test[9] before surgical resection. MR is preferred to endoscopic rectal ultrasound for determination of the relationship of the cancer to the mesorectal fascia as well as for detection of lateral pelvic lymph nodes, which cannot be seen by ultrasound because of its limited field of view (Fig. 133-11).

INTERVENTIONAL PROCEDURES

Ultrasound, CT, fluoroscopy, and even MR techniques have been used for guidance in performing a wide variety of abdominal interventional procedures. Percutaneous image-guided biopsy, whether by fine-needle aspiration or core biopsy, is a relatively safe procedure that is commonly performed for tissue diagnosis[10] and has drastically reduced the need for open surgical biopsy. Other common nonvascular procedures that use image guidance include abscess drainage, biliary interventions, gastrostomy, and tumor ablation. In the case of peridiverticular and periappendiceal abscesses, CT-guided drainage can often simplify the ultimate operative approach and turn high-risk emergent surgery into a safer elective procedure. Biliary interventions include transhepatic access of an obstructed system for stenting or external drainage and cholecystostomy tube placement. Percutaneous CT- or ultrasound-guided tumor ablation is a rapidly evolving technique that is particularly useful in poor operative candidates or in conjunction with surgical resection of other lesions.[11] A variety of ablation methods have been employed, including radio frequency, alcohol, microwave, and cryoablation.

Diagnostic conventional angiography has been replaced largely by noninvasive CT and MR techniques, but direct catheter angiography remains an important procedure for directing various therapies. Vascular interventions include angioplasty, stenting, embolization, and thrombolysis. TIPS placement (Chapter 153) is a commonly performed angiographic procedure in patients with portal hypertension complicated by variceal bleeding or

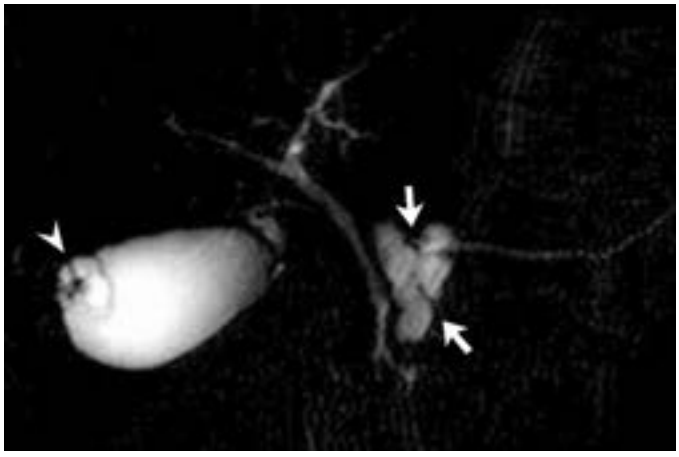

FIGURE 133-9. Pancreatic intraductal papillary mucinous neoplasm on MRCP. Heavily T2-weighted MR image shows a lobulated cystic lesion in the pancreatic head region (*arrows*) that represents a side branch intraductal papillary mucinous neoplasm. Note the mild focal irregularity of the gallbladder (*arrowhead*), consistent with the fundal form of adenomyomatosis. The intrahepatic and extrahepatic biliary ducts are normal.

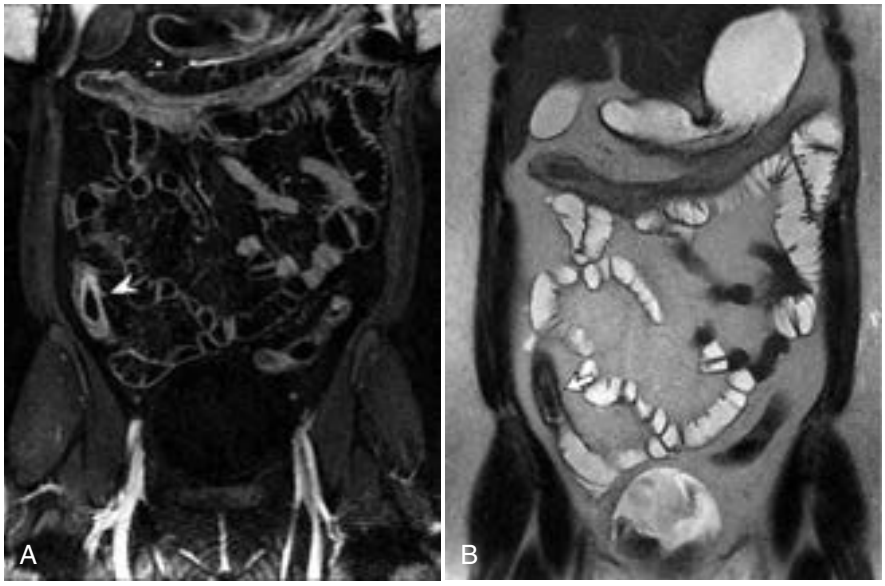

FIGURE 133-10. Active Crohn disease on MR enterography. A, Coronal three-dimensional volume-acquired breath-hold, T1-weighted, gradient-echo image with dynamic gadolinium administration and fat saturation shows wall thickening and enhancement (*arrowhead*) of an abnormal segment of terminal ileum. **B,** Coronal two-dimensional single-shot, fast spin-echo, T2-weighted image shows increased signal in this area (*arrow*), signifying edema and active disease.

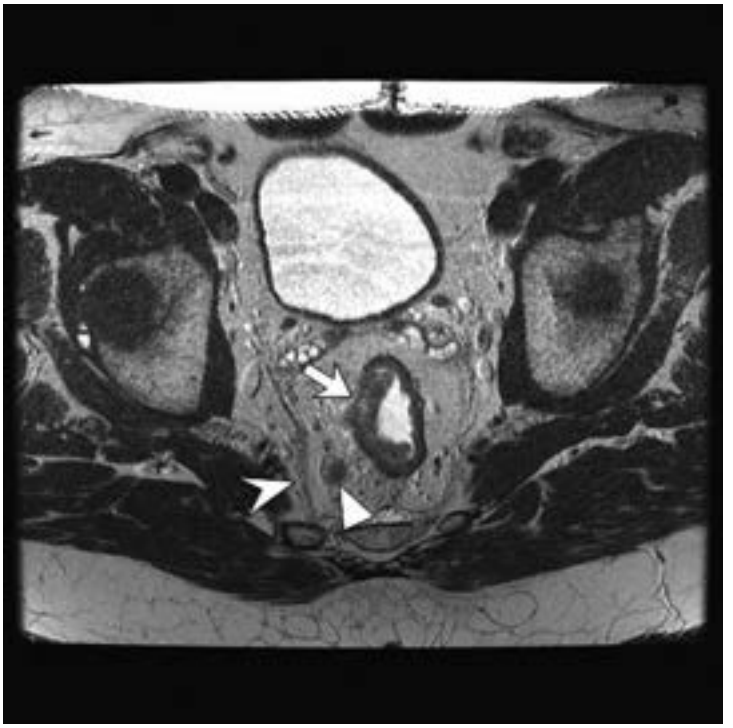

FIGURE 133-11. Preoperative rectal cancer staging by MR. Fast spin-echo T2 image shows the clear margin between the mesorectal fascia (*arrowhead*) and primary cancer (*arrow*) as well as a metastatic lymph node (*triangle*).

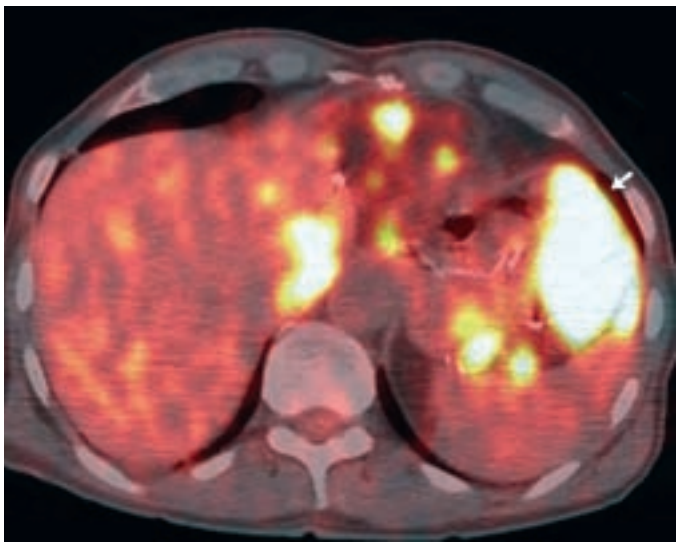

FIGURE 133-12. Metastatic gastrointestinal stromal tumor (GIST) on fused PET/CT. Transverse fused PET/CT image shows a dominant hypermetabolic mass (*arrow*) representing a gastric GIST. Multiple smaller peritoneal and hepatic hypermetabolic foci are consistent with metastatic deposits. Note the utility of combining the functional information from PET with the anatomic localization provided by CT.

intractable ascites. Placement of a TIPS stent creates a low-pressure communication between the portal and hepatic venous systems. Chemoembolization can provide palliation for those with advanced hepatic malignant disease, whether primary or metastatic.

NUCLEAR MEDICINE (RADIONUCLIDE SCINTIGRAPHY)

Owing to the emergence of PET/CT for oncologic evaluation, nuclear medicine is more relevant now than ever before in abdominal imaging. PET/CT is a powerful diagnostic tool that combines functional and anatomic imaging. PET is useful for both initial staging and evaluating the response to therapy for a wide range of primary malignant tumors, especially when it is combined with CT (Fig. 133-12). Currently, clinical PET imaging most often uses ^{18}F-fluorodeoxyglucose, but other positron-emitting agents may be used for specific purposes. PET/MR also is a newer combined modality.[12]

Several other nuclear medicine studies are used to evaluate GI and hepatobiliary diseases. Injection of red blood cells labeled with technetium Tc 99m provides a useful test for GI bleeding. Advantages of performing this as the initial diagnostic imaging study include its noninvasive nature, high sensitivity for active bleeding, and ability to rescan the patient hours later without the need for repeated injection. Disadvantages include relatively poor anatomic localization and lack of therapeutic ability. The use of tagged red blood cell scintigraphy for the diagnosis of hepatic cavernous hemangioma has decreased significantly owing to advances in CT and MRI. Hepatobiliary scintigraphy remains a useful tool in equivocal cases of cholecystitis, particularly acalculous disease, and it can confirm suspected biliary leaks. Scintigraphic imaging with ^{111}In-octreotide is valuable for the diagnosis, staging, and follow-up of GI neuroendocrine tumors, such as carcinoid and pancreatic islet cell tumors.

Grade A Reference

A1. Kim K, Kim YH, Kim SY, et al. Low-dose abdominal CT for evaluating suspected appendicitis. *N Engl J Med*. 2012;366:1596-1605.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Taylor MR, Lalani N. Adult small bowel obstruction. *Acad Emerg Med*. 2013;20:528-544.
2. Yeh BM, Carucci LR, Fidler JL, et al. Luminal imaging in the 21st century. *AJR Am J Roentgenol*. 2011;197:28-29.
3. Lalani T, Couto CA, Rosen MP, et al. ACR appropriateness criteria jaundice. *J Am Coll Radiol*. 2013;10:402-409.
4. Zeng J, Huang ZP, Zheng J, et al. Diagnostic accuracy of two-dimensional shear wave elastography for the non-invasive staging of hepatic fibrosis in chronic hepatitis B: a cohort study with internal validation. *Eur Radiol*. 2014;24:2572-2581.
5. Costello JE, Cecava ND, Tucker JE, et al. CT radiation dose: current controversies and dose reduction strategies. *AJR Am J Roentgenol*. 2013;201:1283-1290.
6. Fidler JL, Fletcher JG, Bruining DH, et al. Current status of CT, magnetic resonance, and barium in inflammatory bowel disease. *Semin Roentgenol*. 2013;48:234-244.
7. Hahn LD, Emre SH, Israel GM. Radiographic features of potential donor livers that precluded donation. *AJR Am J Roentgenol*. 2014;202:W343-W348.
8. Tirkes T, Sandrasegaran K, Sanyal R, et al. Secretin-enhanced MR cholangiopancreatography: spectrum of findings. *Radiographics*. 2013;33:1889-1906.
9. Kaur H, Choi H, You YN, et al. MR imaging for preoperative evaluation of primary rectal cancer: practical considerations. *Radiographics*. 2012;32:389-409.
10. Howlett DC, Drinkwater KJ, Lawrence D, et al. Findings of the UK national audit evaluating image-guided or image-assisted liver biopsy. Part II. Minor and major complications and procedure-related mortality. *Radiology*. 2013;266:226-235.
11. Ahmed M, Brace CL, Lee FT Jr, et al. Principles of and advances in percutaneous ablation. *Radiology*. 2011;258:351-369.
12. Torigian DA, Zaidi H, Kwee TC, et al. PET/MR imaging: technical aspects and potential clinical applications. *Radiology*. 2013;267:26-44.

134 GASTROINTESTINAL ENDOSCOPY

PANKAJ JAY PASRICHA

IMPORTANCE AND USE OF ENDOSCOPY

Technologic advances in radiologic and endoscopic imaging have transformed medicine in the past few decades. With its remarkable accessibility, the gastrointestinal tract, perhaps more than any other organ system, has benefited particularly from the endoscopic approach. The major advantages of endoscopy over contrast radiography for evaluation of diseases of the alimentary tract include direct visualization, resulting in a more accurate and sensitive evaluation of mucosal lesions; the ability to obtain biopsy specimens from superficial lesions; and the ability to perform therapeutic interventions. These advantages make endoscopy the procedure of choice in most cases in which mucosal lesions or growths are suspected. Conversely, computed tomography (CT) or, occasionally, contrast radiography may be indicated when extrinsic or intrinsic distortions of anatomy are suspected, such as volvulus, intussusceptions, subtle strictures, or complicated postsurgical changes (Chapter 133).

On the therapeutic front, the endoscopic approach (transorally or transrectally) is increasingly being used to replace traditional and more invasive forms of surgery, a trend that is expected to gain strength in the years to come. In this context, the flexible endoscope is rapidly becoming an essential tool for both gastroenterologists and surgeons alike.

INSTRUMENTS AND PROCEDURES

Endoscopic procedures and their therapeutic applications are described in Table 134-1.

Luminal Endoscopy: Conventional and Wireless

The modern gastrointestinal endoscope is a "videoscope," with a charged couple device chip at its tip. The scope itself is tethered to a light source and video processer, and the image is displayed on one or more monitors. The endoscopic shaft not only carries the optical elements for imaging but also contains channels that enable various functions, such as air insufflation, water irrigation, suction, and passage of diagnostic and therapeutic devices.

Endoscopy no longer requires tethering to a light source. The capsule endoscope, a disposable plastic capsule that is approximately the size of a large vitamin pill, contains a chip camera, batteries, and a radio transmitter

TABLE 134-1 ENDOSCOPIC PROCEDURES AND GENERAL APPLICATIONS

ENDOSCOPIC PROCEDURE	THERAPEUTIC APPLICATIONS
LUMINAL ENDOSCOPY	
Common procedures Esophagogastroduodenoscopy Colonoscopy Flexible sigmoidoscopy	Hemostasis Luminal restoration (dilation, ablation, stenting) Lesion removal (e.g., polypectomy, mucosal ablation) Provision of access (percutaneous endoscopic gastrostomy and jejunostomy)
Less common procedures Enteroscopy Capsule endoscopy	Barrier strengthening (antireflux procedures)
PANCREATOBILIARY IMAGING	
Endoscopic retrograde cholangiopancreatography	Lesion (stone) removal Luminal restoration (dilation, stenting) Provision of access (sphincterotomy) Drainage (bile, pancreatic pseudocyst)
TRANSLUMINAL IMAGING	
Endoscopic ultrasonography	Analgesic block Delivery of therapeutic agents (experimental)

TABLE 134-2 COMPLICATIONS OF ENDOSCOPY

ENDOSCOPIC COMPLICATION	INCIDENCE (%)	SPECIFIC PROPHYLAXIS
GENERAL COMPLICATIONS		
Complications related primarily to sedation (cardiovascular and respiratory depression, aspiration)	0.6-0.7	Airway protection with massive upper gastrointestinal bleeding Preprocedure medical evaluation, intraprocedure and postprocedure monitoring Anesthesiology consultation for high-risk patients
Perforation	0.1-0.3 (upper endoscopy) 0.14-0.25 (colonoscopy)	None (except careful technique)
Bleeding	0.3 (upper endoscopy) 0.7-2.5 (polypectomy)	Carefully balance risk and benefits Discontinue or reduce anticoagulant use before high-risk procedures
Bacteremia and infectious complications (endocarditis, bacterial ascites)	<0.1	Antibiotics for patients at risk for endocarditis (patients with artificial valves, pulmonary-systemic shunts, previous history of endocarditis), with synthetic vascular grafts, and with bacterial ascites (cirrhotics)
Death	0.6 (upper endoscopy) 0.2 (colonoscopy)	
COMPLICATIONS ASSOCIATED WITH SPECIALIZED PROCEDURES		
Pancreatitis (ERCP)	3-20	Rectal indomethacin
Cholangitis (ERCP)	0.1-2	Preprocedure antibiotics
Wound infections (PEG)	3-4	Preprocedure antibiotics

ERCP = endoscopic retrograde cholangiopancreatography; PEG = percutaneous endoscopic gastrostomy.

that wirelessly sends images to a device that the patient wears as a belt. At the end of the procedure, the information is downloaded to a computer, and the capsule itself passes out harmlessly in the stool. The capsule endoscope has allowed routine evaluation of the entire length of the small bowel, a feat that was challenging and oftentimes impossible with conventional instruments. Variations of the capsule endoscope have been developed for esophageal and colonic imaging, but the utility of these capsules in routine clinical practice has not been established.

Ancillary Organ Imaging: Endoscopic Retrograde Cholangiography and Pancreatography

Endoscopic retrograde cholangiopancreatography (ERCP) uses a side-viewing endoscope that accesses the second part of the duodenum, where a small catheter is then introduced into the bile or pancreatic duct to inject radiographic contrast medium under fluoroscopic monitoring. Successful cannulation and imaging can be achieved in up to 95% of cases. In some instances, a fine-caliber endoscope can also be introduced into the duct of interest (cholangioscopy or pancreatoscopy) for direct visualization of intraductal disease.

Mural and Transmural Imaging: Endoscopic Ultrasonography

An ultrasonic transducer in the tip of a flexible endoscope or a stand-alone ultrasound probe inserted through the channel of a regular endoscope can image lesions within the wall of the gut as well as adjacent lymph nodes, vascular structures, and neighboring organs such as the pancreas. Endoscopic ultrasonography (EUS) can guide fine-needle aspiration of suspicious lesions more accurately than abdominal ultrasonography or CT.

COMPLICATIONS AND PRE-ENDOSCOPIC PREPARATION

Diagnostic endoscopy is a remarkably safe and well-tolerated procedure.[1] It can be performed under conscious sedation with a combination of benzodiazepines and narcotics or increasingly in the United States with propofol. Although propofol provides faster and deeper sedation with rapid recovery (Chapter 432), it adds significantly to the cost of the procedure, principally because of the need for monitored anesthesia care. Recent Food and Drug Administration approval of a computer-assisted semiautomated delivery and monitoring system for propofol may address some of these concerns.

Although rare, potential complications (Table 134-2) must be carefully explained to the patient as part of the informed consent process.[2,3] In general, routine blood tests, radiographs, or electrocardiograms are not necessary before endoscopy unless a careful history and physical examination suggest possible hematologic, cardiovascular, or pulmonary and airway problems. Women of childbearing age should be questioned about the possibility of pregnancy and tested if there is any doubt.

Diagnostic endoscopies, including those with mucosal biopsies, are considered low enough risk that they do not warrant discontinuation of anticoagulant medication. Similarly, patients undergoing screening colonoscopy can continue aspirin or other nonsteroidal anti-inflammatory agents. In high-risk elective procedures, the decision to withhold anticoagulants and antiplatelet agents should be individualized. Depending on the underlying thromboembolic risk, patients may require "bridging" therapy with agents in the heparin family (Chapters 38 and 431). For patients with acute bleeding, reversal therapy or platelet replacement may be considered (Chapters 171 through 175).

ERCP is associated with the highest risk of serious complications, with about 5% of cases developing pancreatitis. Most experts will place a short-term pancreatic stent in high-risk cases as a reasonable preventive measure. Prophylactic rectal indomethacin (two 50-mg suppositories immediately after the completion of the procedure) reduces the incidence of post-ERCP pancreatitis by almost 50%[■] and is becoming the standard of care.

SPECIFIC INDICATIONS

Most indications for gastrointestinal endoscopy are based on the presenting symptoms of the patient (e.g., dysphagia, bleeding, diarrhea). In other instances, endoscopy is required to evaluate specific lesions found by other diagnostic imaging, such as a gastric ulcer or colon polyp discovered by barium radiography. Finally, screening endoscopy is often performed in asymptomatic individuals on the basis of their risk for commonly occurring and preventable conditions, such as colon cancer (see later).

Implicit in the decision to perform endoscopy is the assumption that it will have a bearing on future management strategy. In evaluating gastrointestinal symptoms, several questions need to be addressed by the referring physician and the endoscopist. Which patients need endoscopy? When should the endoscopy be done? What is the endoscopist looking for? What endoscopic therapy, if any, should be planned?

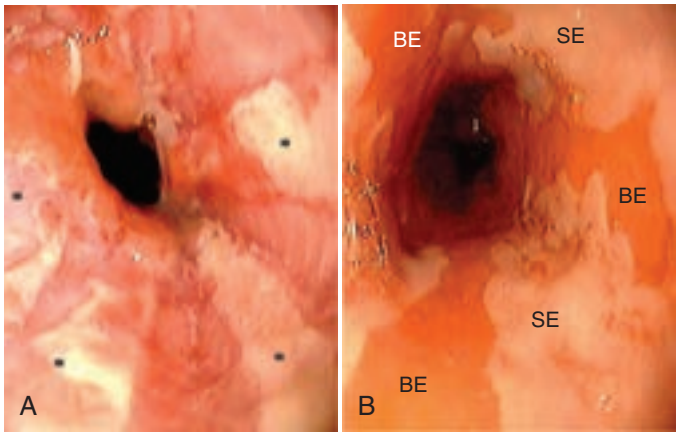

FIGURE 134-1. Severe reflux esophagitis (*left*) with mucosal erythema and linear ulcers with yellow exudates (*asterisks*). It is thought that such changes eventually lead to Barrett esophagus (*right*), in which the normal white squamous epithelium (SE) is replaced by red columnar epithelium (BE). These pictures are from different patients.

Gastroesophageal Reflux and Heartburn (Chapters 138 and 139)

Gastroesophageal reflux disease (GERD) is an extremely common condition in the general population. The fact that its cardinal symptom, heartburn, is relatively specific for this condition justifies an empirical approach to treatment by a combination of lifestyle modifications and over-the-counter or even prescription drugs. Endoscopy is not therefore necessary to make the diagnosis of GERD. Indeed, normal findings on endoscopy do not exclude the diagnosis of GERD because the overall sensitivity of endoscopy in GERD is only about 70%. If necessary, further evaluation with ambulatory esophageal manometry and pH monitoring may be indicated to establish the diagnosis. However, there are several circumstances in which endoscopy should be considered for patients with reflux, including patients with associated warning symptoms ("red flags"), such as dysphagia, odynophagia, regurgitation, weight loss, gastrointestinal bleeding, or frequent vomiting (Fig. 134-1). These symptoms imply either the development of a GERD-related complication (erosive esophagitis, stricture, or adenocarcinoma) or another disorder masquerading as GERD (esophageal cancer or a gastric-duodenal lesion such as cancer or peptic ulcer). Another group of patients who are candidates for endoscopy are those with severe, persistent, or frequently recurrent symptoms that suggest significant esophagitis and hence a risk for complications, such as stricture or Barrett esophagus, which is intestinal metaplasia of the esophageal epithelium.

If a significant length of Barrett esophagus, especially more than 3 cm, is discovered (see Fig. 134-1), most experts recommend some form of periodic surveillance endoscopy because of the increased risk for the development of adenocarcinoma. Control of reflux by either pharmacologic or surgical means does not generally lead to regression of established Barrett esophagus (Chapter 138). For patients whose high-grade dysplasia associated with Barrett esophagus poses a serious risk for future cancer, endoscopic ablation or resection may provide a potentially curative alternative to surgical esophagectomy. Ablation can be achieved by a variety of modalities, including radio frequency, cryotherapy, electrical cautery, argon plasma coagulation, and photodynamic therapy. An alternative to ablation is endoscopic mucosal resection, which is en bloc resection of the mucosa to allow a complete pathologic analysis and to minimize the risk for regrowth of the abnormal mucosal lining.

Several endoscopic techniques are potential alternatives to surgical fundoplication for patients whose reflux is not satisfactorily managed by medical therapy.[4] These include methods that provide thermal energy to the lower esophageal sphincter and others that serve to tighten the gastroesophageal "valve" area. Other procedures that are in clinical trials include electrical stimulation by implantable electrodes.

Heartburn in immunocompromised patients often indicates an esophageal infection with an opportunistic organism, such as *Candida albicans*, cytomegalovirus, or herpesvirus. Because most patients with the acquired immunodeficiency syndrome (AIDS) and esophagitis have candidiasis, an empirical course of antifungal therapy may be justified. Patients who do not respond to this approach, however, should almost always have endoscopy and biopsy so that more specific therapy can be instituted.

Dysphagia (Chapter 138)

Dysphagia can often be categorized as oropharyngeal on the basis of the clinical features of nasal regurgitation, laryngeal aspiration, or difficulty in moving the bolus out of the mouth. These symptoms are usually associated with a lesion in the central or peripheral nervous system. Although endoscopy is often performed in these patients, videofluoroesophagography (modified barium swallow or cine-esophagogram) is the procedure of choice because it allows a frame-by-frame evaluation of the rapid sequence of events involved in transfer of the bolus from the mouth to the esophagus. Common causes of dysphagia in the esophageal body include malignant as well as benign processes (peptic strictures secondary to reflux, Schatzki ring) and motility disturbances. Endoscopic examination is considered mandatory in all patients with esophageal dysphagia. However, contrast esophagography is helpful; it can provide guidance for endoscopy that is anticipated to be difficult (e.g., a patient with a complex stricture or diverticulum), suggest a disturbance in motility, and occasionally detect subtle stenoses that are not appreciated on endoscopy (the scope diameter is typically ≤10 mm, whereas some symptomatic strictures can be considerably wider).

Endoscopic treatment options are available for many causes of esophageal dysphagia. Tumors can be ablated by thermal means (cautery or laser) or stented with prosthetic devices. Metallic expandable stents have become the palliative procedure of choice for most patients with symptomatic esophageal cancer. Benign lesions of the esophagus, such as strictures or rings, can also be dilated endoscopically, usually with excellent results. Finally, some motility disturbances, such as achalasia, may be approached endoscopically with the use of large balloon dilators for the lower esophageal sphincter or, in the case of high-risk patients, local injection of botulinum toxin.

Dyspepsia (Chapter 137)

Dyspepsia, which is chronic or recurring pain or discomfort centered in the upper abdomen, is a common condition that can be caused by a variety of disorders, including peptic ulcer, reflux esophagitis, gallstones, gastric dysmotility, and, rarely, gastric or esophageal cancer. However, up to 60% of patients with chronic (>3 months) dyspepsia belong to the so-called functional category in which there is no definite structural or biochemical explanation for the symptoms. Although *Helicobacter pylori* gastritis is found frequently in these patients, there is no definite evidence to prove a cause-and-effect relationship between these two findings. If a diagnostic test is to be performed, endoscopy, sometimes with biopsies to detect *H. pylori*, is clearly the procedure of choice (see Fig. 139-2), with accuracy of about 90% compared with about 65% for double-contrast radiography. Because dyspepsia is a recurrent condition and because patients who do not respond to empirical therapy eventually almost always undergo endoscopy, many gastroenterologists opt for early endoscopy, if only for the reassurance that a normal examination provides.

Upper Gastrointestinal Bleeding (Chapter 135)

Acid peptic disease (including ulcers, erosions, and gastritis), variceal bleeding, and Mallory-Weiss tears account for most cases of upper gastrointestinal bleeding. Other less common but important lesions are angiomas, gastric vascular ectasia ("watermelon" stomach), and the uncommon Dieulafoy lesion (a superficial artery that erodes through the gut mucosa). Finally, upper gastrointestinal cancers are occasionally associated with significant bleeding. Endoscopy is mandatory in all patients with upper gastrointestinal bleeding, with the rare exception being the terminally ill patient in whom the outcome is unlikely to be affected. Endoscopy is able to detect and to localize the site of the bleeding in 95% of cases and is clearly superior to contrast radiography (with an accuracy of only 75 to 80%). The endoscopic appearance of bleeding lesions can also help predict the risk of rebleeding, thus facilitating the triage and treatment process. Bleeding can be effectively controlled during the initial endoscopic examination itself in the majority of cases. The risk of recurrent bleeding is diminished, thereby resulting in a shorter duration of hospital stay as well as a reduction in the need for surgery.

In general, endoscopy should be performed only after adequate stabilization of hemodynamic and respiratory parameters. The role of gastric lavage before endoscopy is controversial; some endoscopists prefer that it be done, occasionally even with use of a large-bore tube, whereas others avoid such preparation because of the fear of producing artifact. The timing of subsequent endoscopy depends on two factors: the severity of the hemorrhage and the risk status of the patient. Patients with active, persistent, or severe bleeding (>3 units of blood) require urgent endoscopy. In these patients,

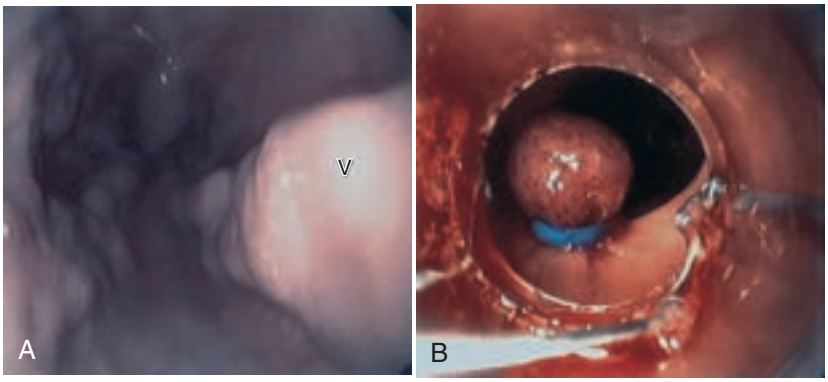

FIGURE 134-2. Endoscopic view of esophageal varices (*left*) in the wall of the esophagus (V). *Right*, Image of a varix that has been endoscopically ligated with a band.

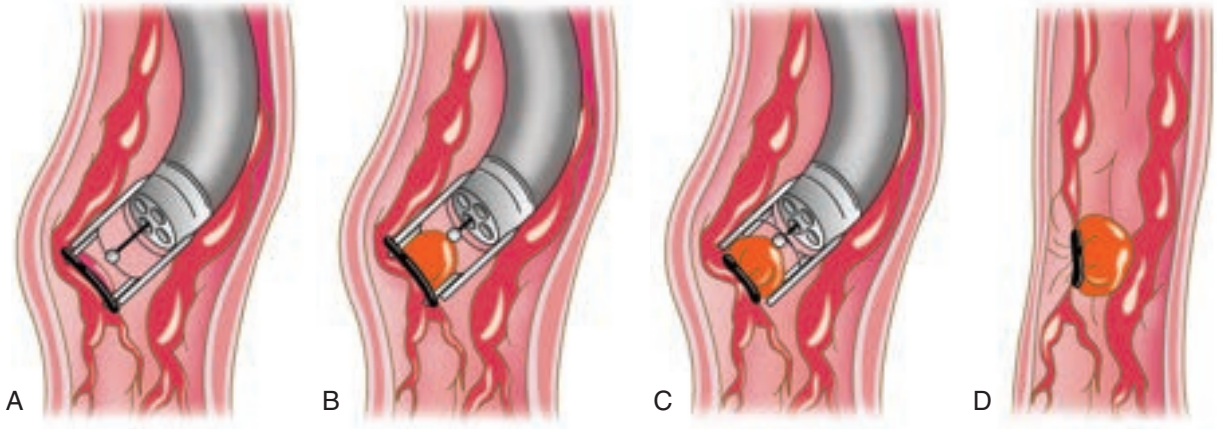

FIGURE 134-3. Endoscopic variceal ligation technique. A, The endoscope, with attached ligating device, is brought into contact with a varix just above the gastroesophageal junction. **B,** Suction is applied, drawing the varix-containing mucosa into the dead space created at the end of the endoscope by the ligating device. **C,** The tripwire is pulled, releasing the band around the aspirated tissue. **D,** Completed ligation.

endoscopy is best performed in the intensive care unit because of the risk for aspiration and the occasional need for emergent intubation to provide respiratory protection and ventilation. Patients with slower or inactive bleeding may be evaluated by endoscopy in a "semielective" manner (usually within 12 to 20 hours), but a case can be made to perform endoscopy early even in these stable patients (perhaps in the emergency department itself) to allow more confident triage and efficient resource management.[5]

Nonvariceal bleeding vessels can be treated by a variety of means, including injections of various substances (epinephrine, saline, sclerosants), thermal coagulation (laser or electrocautery), and mechanical means (clipping). In the United States, the most popular approach to a bleeding peptic ulcer lesion is a combination of injection with dilute epinephrine and electrocoagulation. Initial hemostasis can be achieved in more than 90% cases; rebleeding, which may occur in up to 20% of cases, responds about half of the time to a second endoscopic procedure. Patients who continue to bleed (typically patients with large ulcers in the posterior wall of the duodenal bulb) are usually managed by interventional angiography (with embolization of the bleeding vessel) or surgically.

Variceal bleeding is also effectively managed endoscopically, with a success rate similar to that with bleeding ulcers (Fig. 134-2). Hemostasis with band ligation (Fig. 134-3) has replaced the older methods of sclerotherapy because of fewer side effects. Even if initial endoscopic hemostasis is successful, long-term prevention of rebleeding requires a program of ongoing endoscopic sessions until variceal obliteration is complete. Patients who do not respond to endoscopic treatment are considered candidates for a transjugular intrahepatic portosystemic shunt. In patients whose large esophageal varices have never bled, β-blockers are considered first-line treatment, but endoscopic band ligation may be useful in selected patients.

Acute Lower Gastrointestinal Bleeding

The most common cause of acute lower gastrointestinal bleeding is angiodysplasia, followed by diverticulosis, neoplasms, and colitis. In about 10% of patients presenting with hematochezia, a small bowel lesion may be responsible. In contrast to upper gastrointestinal bleeding, there is no single best test for acute lower gastrointestinal bleeding (Fig. 134-4). In young patients (<40 years old) with minor bleeding, features that are highly suggestive of anorectal origin (e.g., blood on the surface of the stool or on the wipe) may warrant only flexible sigmoidoscopy. Conversely, patients presenting with hemodynamic compromise may need upper endoscopy first to exclude a lesion in the upper gastrointestinal tract (typically postpyloric) bleeding so briskly that it presents as hematochezia. Colonoscopy has been traditionally recommended after bleeding has slowed or stopped and the patient has been given an adequate bowel purge. However, a disadvantage of delaying endoscopy is that when a pathologic lesion such as an arteriovenous malformation (see Fig. 134-4) or diverticulum is found, it may be impossible to implicate it confidently as the site of bleeding (complementary information by radiography or scintigraphy becomes particularly important in this situation). Some

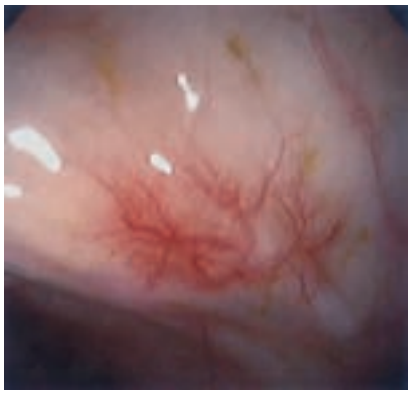

FIGURE 134-4. Mucosal telangiectasia (arteriovenous malformation) in the colon. The patient presented with hematochezia. The lesion was subsequently cauterized endoscopically.

experts therefore recommend urgent diagnostic endoscopy with little or no preparation for acute lower gastrointestinal hemorrhages and have reported significant diagnostic as well as therapeutic success rates. However, such recommendations have not been universally accepted and remain logistically difficult to implement in most hospital settings. If an acute bleeding site cannot be identified by upper and lower gastrointestinal endoscopy, capsule endoscopy is better than angiography to find the source of bleeding.

Occult Gastrointestinal Bleeding or Iron Deficiency Anemia

Normal fecal blood loss is usually less than 2 to 3 mL/day. Most standard fecal occult blood tests detect blood loss of only 10 mL/day or more. Therefore, even if this test result is negative, patients with iron deficiency anemia and no other obvious source of blood loss should always undergo aggressive gastrointestinal evaluation, which uncovers a gastrointestinal lesion in the majority of cases. Although most lesions that cause overt gastrointestinal bleeding can also cause occult blood loss, occult bleeding should almost never be ascribed to diverticulosis or hemorrhoids. Endoscopy is always preferable to radiographic studies for evaluation of occult blood loss or iron deficiency anemia because of its ability to detect flat lesions, particularly vascular malformations, which may be found in 6% or more of patients. If the findings on both upper and lower endoscopy are normal, the next test is capsule endoscopy, which may be helpful to detect small bowel lesions, such as erosions, tumors, or angiomas. Although it is relatively contraindicated in patients with suspected narrowing or strictures of the small bowel, capsule endoscopy has become the diagnostic procedure of choice in patients with obscure gastrointestinal bleeding (with normal findings on upper and lower endoscopies) and when mucosal lesions of the small bowel are suspected. Findings on capsule endoscopy may prompt the consideration of enteroscopy (with specialized balloon-assisted or spirally advancing endoscopes), which can theoretically access the entire small bowel and permit biopsy or therapy of suspected lesions.

Colorectal Neoplasms (Chapter 193)

Colonoscopy is the most accurate test for detection of mass lesions of the large bowel or colon that are suspected on clinical or radiologic grounds. However, the greatest impact of endoscopy on colorectal neoplasia may be in the area of screening and prevention. The adenoma to carcinoma sequence of progression in colorectal cancer provides a unique opportunity for prophylaxis. Thus, if screening programs can identify patients with polyps and if these polyps are removed, cancer can largely be prevented. Various techniques are available for safe and effective polypectomy, depending on the size, presence of a stalk, and location (Fig. 134-5). Colonoscopy is currently recommended for screening of patients at average risk, that is, anyone older than 50 years. Adenomatous polyps are removed, and patients are entered into a surveillance program with follow-up colonoscopies at intervals that depend on the nature and number of the initial lesions. Patients who do not have any polyps generally do not require follow-up colonoscopies more than once every 10 years. More aggressive screening strategies are required for patients considered at high risk for colorectal cancer, including patients with well-defined hereditary syndromes as well as those with a history of colorectal cancer in a first-degree relative. In addition, patients with ulcerative colitis (Chapter 141) with long-standing (>8 years) disease affecting the entire colon have an increased risk for development of colon cancer, about 0.5 to 3% after 20 years.

CT colonography (Chapter 133), which involves the digital construction of an endoluminal view of the colon on the basis of data from abdominal CT, has generally replaced barium enema as an alternative method for colon cancer screening. It has not, however, generally replaced colonoscopy, in part because it misses smaller polyps and in part because any abnormalities found on CT colonoscopy require colonoscopic follow-up. Its major utility is for screening patients who have an incomplete colonoscopy. Imaging of the colon by capsule endoscopes is currently not adequate and not recommended.

Chronic Diarrhea (Chapter 140)

Endoscopy may be a valuable aid in the evaluation of patients with persistent diarrhea. The timing of the endoscopy in these patients often depends on the clinical features of the illness. Patients with bloody diarrhea should have lower endoscopy as part of their initial evaluation to determine if inflammatory bowel disease is present (Chapter 141). In most patients with chronic diarrhea, endoscopy is often done when initial routine testing does not yield a specific diagnosis. Both upper and lower endoscopies may be used, depending on the clinical presentation. Thus, the patient thought to have a malabsorptive process may require upper endoscopy with jejunal or duodenal biopsies to look for celiac sprue or rare lesions such as lymphoma or Whipple disease because endoscopic biopsy has largely replaced blind intestinal biopsies for these conditions. Conversely, patients thought to have a secretory cause of diarrhea require a colonoscopy with biopsies to look for overt inflammatory bowel disease or more subtle variants such as microscopic or lymphocytic colitis, in which cases the diagnosis requires careful examination of the biopsy specimens.

The endoscopic approach to diarrhea in immunocompromised patients, such as those with HIV infection, is guided by the degree of immunosuppression and the need to find treatable infections. When results of routine stool tests are negative, patients with CD4 counts less than 100/mm^3 should undergo endoscopic evaluation to detect pathogens, such as cytomegalovirus, *Mycobacterium avium* complex, and microsporidiosis. Small-volume stools with tenesmus suggest proctocolitis, for which sigmoidoscopy (rather than a full colonoscopy) with biopsies is usually adequate. In patients with upper gastrointestinal symptoms (large-volume diarrhea, bloating, and dyspepsia), upper endoscopy with small bowel biopsy may be attempted first.

Miscellaneous Indications

The upper endoscope has provided a relatively quick and noninvasive means for removal of accidentally or deliberately ingested foreign bodies. Timing is critical for removal, however, because objects are usually beyond endoscopic retrieval when they reach the small bowel. Any foreign object that is causing symptoms should be removed, as should potentially dangerous devices such as batteries and sharp objects. In general, objects larger than 2.5 cm in width or 13 cm in length are unlikely to leave the stomach and so should also be removed. On occasion, patients with food impacted in the esophagus require endoscopic removal (Fig. 134-6). This condition almost always indicates an underlying functional or structural problem (Chapter 138) and should prompt a thorough diagnostic evaluation after the acute problem has been addressed.

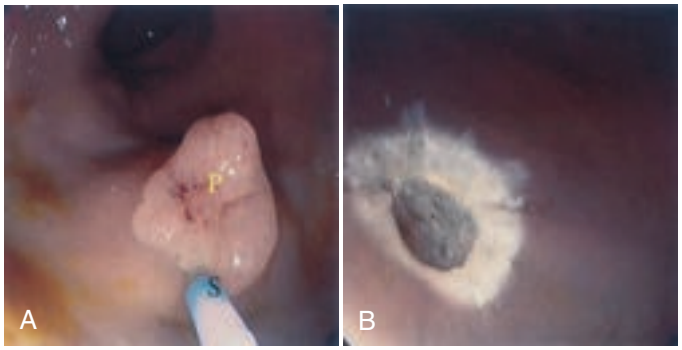

FIGURE 134-5. Endoscopic polypectomy. *Left*, A snare (S) has been passed through the endoscope and positioned around the polyp (P). *Right*, Subsequently, cautery was applied and the polyp guillotined, leaving behind a clean mucosal defect.

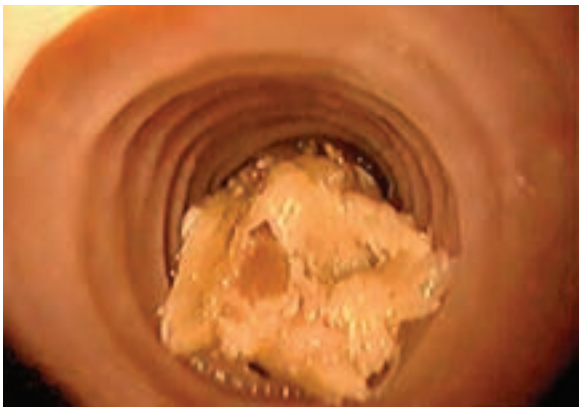

FIGURE 134-6. Impacted food bolus in a young male patient who was found to have a ringed esophagus on endoscopy. This presentation is characteristic and may be either congenital or acquired secondary to reflux-induced or eosinophilic esophagitis.

Because of the relatively poor correlation between oropharyngeal lesions and more distal visceral injury, upper endoscopy is usually recommended urgently in patients with corrosive ingestion (Chapter 110). Endoscopy allows patients to be divided into high- or low-risk groups for complications, with institution of appropriate monitoring and therapy.

Malignant obstruction of the gastrointestinal lumen including the esophagus (Fig. 134-7), pylorus or duodenum, and colon can now be safely and effectively palliated endoscopically by expandable metal stents, thereby avoiding the need for surgery. Colonoscopy is also useful in patients with pseudo-obstructive (nonobstructive) colonic dilation or Ogilvie syndrome; such patients are at risk for colonic rupture at diameters of more than 9 to 12 cm, and colonoscopic decompression is often required, sometimes on an emergent basis.

A major advance in enteral feeding has been the introduction of percutaneous endoscopic gastrostomy (PEG), a relatively quick, simple, and safe endoscopic procedure that has virtually eliminated surgical placement of gastric tubes. The most common indication for these procedures is the need for sustained nutrition in patients with neurologic impairment of swallowing or with head and neck cancers. Patients with a short life expectancy are not suitable candidates for PEG and can be managed by nasoenteral tubes. Further, despite its intuitive appeal, there is little or no evidence that PEG feeding alters clinical or nutritional outcomes or significantly improves quality of life. A variation of PEG is percutaneous endoscopic jejunostomy (PEJ), in which a long tube is passed through the gastric tube, past the pylorus, and into the jejunum. PEJ does not prevent aspiration, but it is effective in patients who have significant impairment of gastric emptying. Retrograde tube migration with PEJ is common, however, and may require frequent replacement.

Therapeutic endoscopy is increasingly an option for patients who have leaks, perforations, or even fistulas.[6] However, its precise role in these situations remains to be clarified.

PANCREATOBILIARY ENDOSCOPY (IMAGING)

Suspected Biliary Disease (Chapter 155)

The diagnostic approach to patients with cholestasis begins with an attempt to differentiate obstructive from hepatocellular causes. The most common causes of obstructive jaundice are common bile duct stones and tumors of the pancreatic and bile ducts. Less invasive conventional imaging with ultrasonography, CT, or magnetic resonance imaging demonstrates dilated bile ducts and mass lesions but is not sensitive or specific for the detection or delineation of pathologic changes in the distal common bile duct and pancreas, two regions where the majority of obstructing lesions are found. Furthermore, some biliary diseases, such as sclerosing cholangitis, do not result in dilated ducts but have a characteristic appearance on cholangiography. Finally, the ability to use devices such as cytology brushes and biopsy forceps during cholangiography provides an additional aid in the diagnosis of biliary lesions. Both percutaneous and endoscopic cholangiographic techniques are associated with a high rate of success in experienced hands, but the endoscopic approach allows visualization of the ampullary region and the performance of sphincterotomy and also avoids the small risk of a biliary leak associated with puncture of the liver capsule.

In the last few years, magnetic resonance cholangiopancreatography (MRCP), a digital reconstruction technique based on an abdominal MR imaging scan, has become popular as an imaging modality for the pancreatobiliary system, with excellent sensitivity and specificity. Because of its relative safety, this procedure should be routinely used for screening of patients with a low likelihood of disease because it avoids the risk of pancreatitis associated with ERCP. In patients with a higher probability of having a definite lesion, however, ERCP is still the procedure of choice because of its therapeutic options.

Of the approximately 600,000 patients undergoing cholecystectomy in the United States, 5 to 10% may present with bile duct stones before or after surgery. Endoscopic stone removal is successful in 90% or more of these cases and usually requires a sphincterotomy (Fig. 134-8). The sphincter of Oddi is a band of muscle that encircles the distal common bile duct and pancreatic duct in the region of the ampulla of Vater; cutting of this muscle, or sphincterotomy, is one of the mainstays of endoscopic biliary treatment and is accomplished with a special tool called a papillotome or sphincterotome. This procedure is often sufficient for the treatment of small stones in the bile ducts, but larger stones may require additional procedures, such as mechanical, electrohydraulic, or laser lithotripsy, which can be performed endoscopically. In addition to stone disease, sphincterotomy can be curative for patients with papillary stenosis. In other patients with suspected spasm of the biliary sphincter (termed sphincter of Oddi dysfunction), sphincter pressures can be measured by manometry, although the role of sphincterotomy in the treatment of these patients is much more controversial and associated with some of the highest risks for pancreatitis. Finally, by enlarging the access to the bile duct, sphincterotomy facilitates the passage of stents and other devices into the bile duct.

Endoscopic placement of indwelling metal stents for malignant biliary obstruction is superior to both radiologic and surgical techniques.

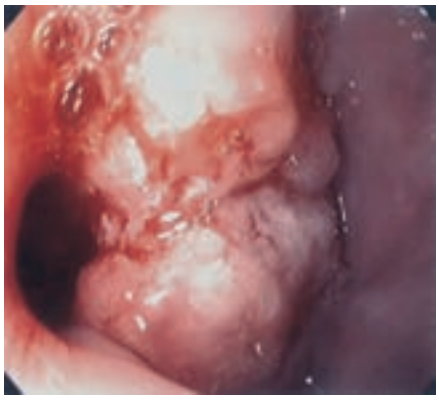

FIGURE 134-7. Large malignant mass at the gastroesophageal junction as seen endoscopically.

Pancreatic Neoplasms

EUS is probably the single best test for diagnosis of pancreatic tumors (Chapter 195), particularly the small endocrine varieties, with sensitivities approaching 95% (Fig. 134-9). It is also the procedure of choice for imaging of submucosal and other mural lesions of the gastrointestinal tract (overall accuracy of 65 to 70%) as well as for staging of a variety of gastrointestinal tumors (overall accuracy of 90% or more), especially esophageal and pancreatic cancer. EUS-directed celiac plexus neurolysis appears to be effective for

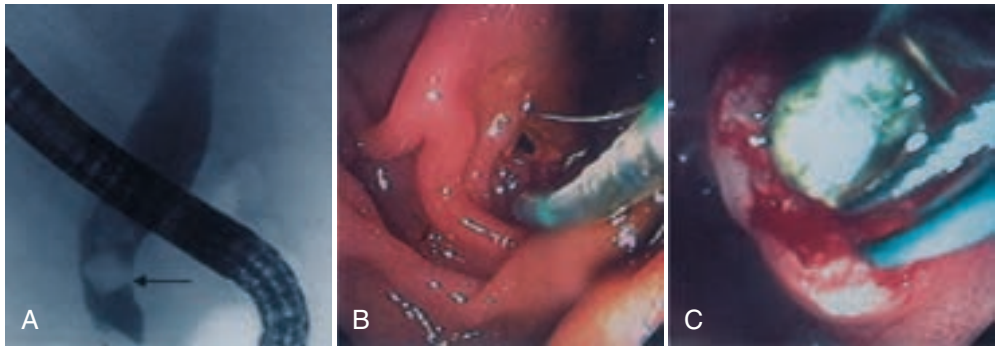

FIGURE 134-8. Biliary sphincterotomy and stone removal from the bile duct. *Left*, Endoscopic retrograde cholangiographic image showing stones (*arrow*) in the distal common bile duct. *Center*, Endoscopic image of a sphincterotome in the bile duct with the wire cutting the roof of the ampulla (sphincter). *Right*, A stone is being removed from the bile duct by an endoscopically passed basket.

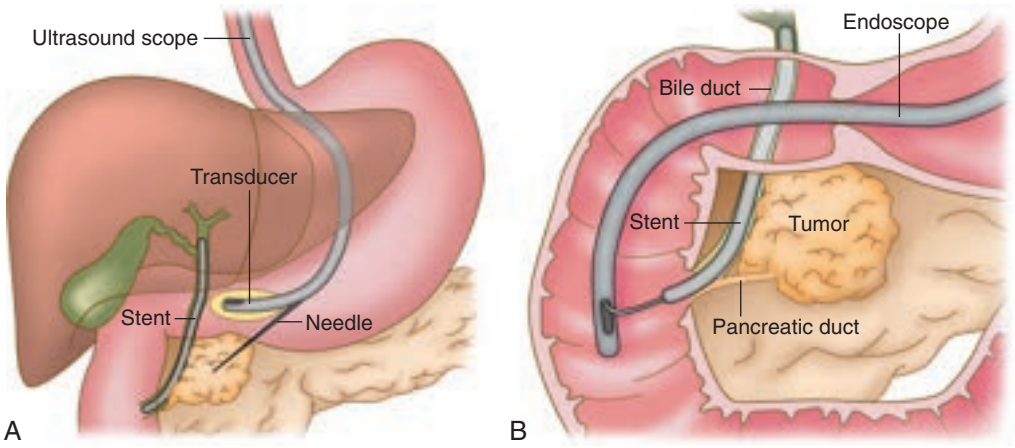

FIGURE 134-9. Biopsy of a pancreatic mass guided by endoscopic ultrasonography (A) and the placement of a stent into a malignant bile duct stricture with endoscopic retrograde cholangiopancreatography (B). (From Brugge WR, Van Dam J: Pancreatic and biliary endoscopy. *N Engl J Med.* 1999;341:1808-1816. Copyright ©1999 Massachusetts Medical Society. All rights reserved.)

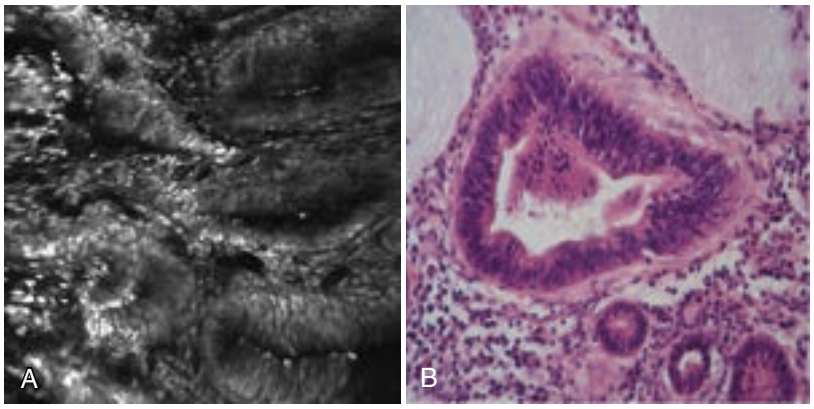

FIGURE 134-10. Confocal microscopy (*left*) showing high-grade intraepithelial neoplasia of a colorectal polyp during endoscopy. Acriflavine was used as a contrast agent (0.02%), highlighting the cellular and nuclei architecture. The histologic picture on the right was taken from the same polyp. (Courtesy Dr. Ralph Kiesselich, University of Mainz.)

the treatment of pain in patients with pancreatic cancer, although it does not appear to work as well in patients with chronic pancreatitis.

Nonmalignant Pancreatic Disease (Chapter 144)

ERCP has replaced MRCP as the test of choice for patients with acute or recurrent pancreatitis without any obvious risk factors on history or routine laboratory evaluation. Imaging of the pancreatic duct by either method may delineate anatomic abnormalities that may be responsible for the pancreatitis, such as congenital variants (pancreas divisum, annular pancreas) or intraductal tumors. In patients with chronic pancreatitis, which is most often due to excessive alcohol intake, pancreatography can confirm the diagnosis, provide useful information about the severity of the disease, and identify ductal lesions that may be amenable to therapy by either endoscopic or surgical means. Adjunctive diagnostic measures include the collection and analysis of bile or pancreatic juice in the duodenum. Bile duct crystals (so-called microlithiasis) can result in pancreatitis in some patients even in the absence of macroscopic stones. In more subtle cases, collection and analysis of the electrolyte content of pancreatic juice after stimulation with secretin may be useful in establishing exocrine impairment and hence in confirming chronic pancreatic injury.

ERCP also has a role in some patients with acute pancreatitis (Chapter 144) that is caused by obstructing biliary stones. Patients presenting with severe biliary pancreatitis may benefit from urgent ERCP early in their course, with the intention of detecting and removing stones from the common bile duct. Similarly, patients who have smoldering acute pancreatitis that does not appear to be improving satisfactorily with conservative treatment may require ERCP for identification and treatment of any obstructing lesions in the pancreatic or distal biliary duct.

Therapeutic endoscopy for chronic pancreatic disease is still evolving. Relief of ductal obstruction (e.g., by endoscopic removal of pancreatic stones or dilation of strictures) can provide short to intermediate pain relief in some patients with chronic pancreatitis, although it is not as effective as surgery in the long term. Endoscopic pseudocyst drainage by a variety of techniques is now technically feasible, with results that appear to be comparable to those of surgical or radiologic techniques. Patients with ductal disruptions (e.g., those with pancreatic ascites) can often be treated successfully with endoscopic stent placement. Pancreatic papillotomy may also be useful for some patients with recurrent pancreatitis, such as when pancreas divisum is thought to play a role.

EVOLVING TECHNIQUES AND FUTURE DIRECTIONS

Significant technologic and procedural innovations in both diagnostic and therapeutic endoscopy are being tested. Many of these are so-called optical biopsy techniques that include confocal microscopy (Fig. 134-10), optical coherence tomography, and a variety of different forms of spectroscopy. These techniques have the ability to provide microscopic images of cells at the surface as well as within deeper layers, thereby providing virtual real-time histology.[7] Furthermore, with use of targeted probes, it is possible to image function as well as form (E-Fig. 134-1), thereby adding another dimension to diagnosis. Innovations in endoscopic therapy include natural orifice transluminal endoscopic surgery, by which the endoscopist or surgeon introduces an endoscope through a natural orifice (mouth, vagina, or anal canal), traverses the wall of the viscus, and accesses the peritoneal cavity to perform diagnostic and therapeutic procedures. Although this approach remains controversial, it has paved the way for much more aggressive forms of intraluminal endoscopic therapy. Thus, removal and resection of large tumors is possible with endoscopic approaches, sometimes requiring the closure of real or potential leaks with adjunctive measures, such as clipping or suturing. Another example is the so-called POEM procedure (*per*oral *e*ndoscopic *m*yotomy) for achalasia,[8] in which esophageal myotomy is performed by a submucosal tunneling technique from within the esophagus. Yet another example will be the ability to obtain full-thickness biopsy specimens of the intestinal tract to allow systematic analysis of changes in nerves and muscle

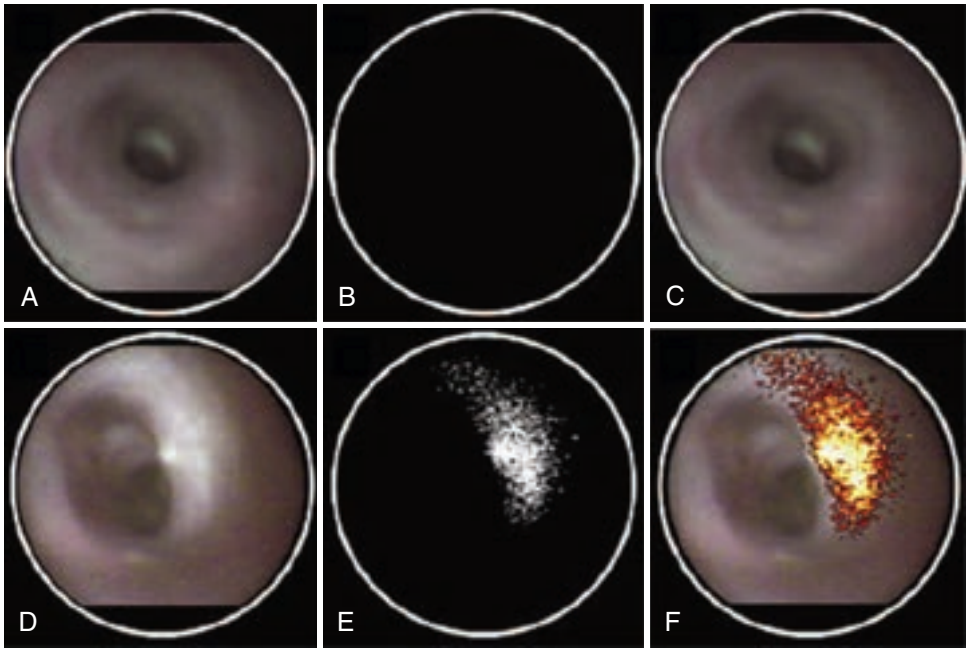

E-FIGURE 134-1. Functional endoscopic imaging. Near-infrared endoscopic imaging of protease activity in a small animal model. Increase in near-infrared fluorescence intensity reveals protease activity from a colonic adenocarcinoma (*bottom row*) in comparison to normal colonic mucosa (*top row*) in an orthotopically implanted mouse model. **A** and **D**, In vivo white light endoscopic images of normal (**A**) and cancerous (**D**) murine colonic mucosa. **B** and **E**, Near-infrared fluorescence images after intravenous injection of ProSense 680, a protease-activated probe sensitive to cathepsin B, showing increased intensity at the site of the tumor but not in normal mucosa. **C** and **F**, False color overlay of white light with fluorescence images showing integration of structural and functional data. (From Alencar H, Funovics MA, Figueiredo J, et al. Colonic adenocarcinomas: near-infrared microcatheter imaging of smart probes for early detection—study in mice. *Radiology*. 2007;244:236.)

in motility disorders. Endoscopic methods also may become viable and less morbid alternatives to more traditional forms of bariatric surgery for obesity and diabetes.

Grade A References

A1. Elmunzer BJ, Scheiman JM, Lehman GA, et al. A randomized trial of rectal indomethacin to prevent post-ERCP pancreatitis. *N Engl J Med*. 2012;366:1414-1422.

A2. Phoa KN, van Vilsteren FG, Weusten BL, et al. Radiofrequency ablation vs endoscopic surveillance for patients with Barrett esophagus and low-grade dysplasia: a randomized clinical trial. *JAMA*. 2014;311:1209-1217.

A3. Leung WK, Ho SS, Suen BY, et al. Capsule endoscopy or angiography in patients with acute overt obscure gastrointestinal bleeding: a prospective randomized study with long-term follow-up. *Am J Gastroenterol*. 2012;107:1370-1376.

A4. Atkin W, Dadswell E, Wooldrage K, et al. Computed tomographic colonography versus colonoscopy for investigation of patients with symptoms suggestive of colorectal cancer (SIGGAR): a multicentre randomised trial. *Lancet*. 2013;381:1194-1202.

A5. Halligan S, Wooldrage K, Dadswell E, et al. Computed tomographic colonography versus barium enema for diagnosis of colorectal cancer or large polyps in symptomatic patients (SIGGAR): a multicentre randomised trial. *Lancet*. 2013;381:1185-1193.

A6. Cahen DL, Gouma DJ, Nio Y, et al. Endoscopic versus surgical drainage of the pancreatic duct in chronic pancreatitis. *N Engl J Med*. 2007;356:676-684.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Gorospe EC, Oxentenko AS. Preprocedural considerations in gastrointestinal endoscopy. *Mayo Clin Proc*. 2013;88:1010-1016.
2. Ben-Menachem T, Decker GA, Early DS, et al. Adverse events of upper GI endoscopy. *Gastrointest Endosc*. 2012;76:707-718.
3. Day LW, Kwon A, Inadomi JM, et al. Adverse events in older patients undergoing colonoscopy: a systematic review and meta-analysis. *Gastrointest Endosc*. 2011;74:885-896.
4. Kaindlstorfer A, Koch OO, Antoniou SA, et al. A randomized trial on endoscopic full-thickness gastroplication versus laparoscopic antireflux surgery in GERD patients without hiatal hernias. *Surg Laparosc Endosc Percutan Tech*. 2013;23:212-222.
5. Jairath V, Kahan BC, Logan RF, et al. Outcomes following acute nonvariceal upper gastrointestinal bleeding in relation to time to endoscopy: results from a nationwide study. *Endoscopy*. 2012;44:723-730.
6. Raju GS. Endoscopic clip closure of gastrointestinal perforations, fistulae, and leaks. *Dig Endosc*. 2014;26(suppl 1):95-104.
7. Liu J, Dlugosz A, Neumann H. Beyond white light endoscopy: the role of optical biopsy in inflammatory bowel disease. *World J Gastroenterol*. 2013;19:7544-7551.
8. Bredenoord AJ, Rosch T, Fockens P. Peroral endoscopic myotomy for achalasia. *Neurogastroenterol Motil*. 2014;26:3-12.

REVIEW QUESTIONS

1. A 22-year-old woman with recurrent right upper quadrant pain is admitted for another attack. She undergoes an ultrasound and computed tomography (CT) examination of the abdomen, which does not reveal any significant abnormalities. Routine chemistries reveal mild (<2×) elevations in transaminases. She is thought to have sphincter of Oddi dysfunction and is scheduled for sphincter manometry followed by possible endoscopic retrograde cholangiopancreatography (ERCP) and sphincterotomy. What is the most important measure to prevent pancreatitis after this procedure?

A. Octreotide given subcutaneously before the procedure
B. Rectal indomethacin right after the procedure
C. Intravenous antibiotics 6 hours before the procedure
D. Performance of the procedure under general anesthesia
E. Pancreatic enzymes begun before and continued after the procedure

Answer: B A single indomethacin suppository administered to patients immediately after ERCP significantly reduces the risk for pancreatitis compared with placebo in high-risk patients. Other measures have not been shown to be of value in this setting.

2. A 55-year-old man undergoes colonoscopy for colorectal cancer screening. The colonoscopy does not reveal any pathologic lesions. On recovery, as you discuss these results with the patient, he points out that his mother's cousin recently died of colorectal cancer. On hearing this, which of the following is most appropriate?

A. Reassure the patient and schedule the next colonoscopy at 10 years.
B. Express concern about the family history and bring the patient back for another colonoscopy within a year.
C. Schedule the next colonoscopy at 3 years.
D. Schedule the next colonoscopy at 5 years.
E. Schedule CT colonography (or virtual colonoscopy) within the next 3 months in case you missed something.

Answer: A Only a history of colorectal cancer or adenoma in a first-degree relative is deemed to put a patient at higher risk than the general population. Therefore, a standard interval of 10 years after a normal screening colonoscopy is recommended.

3. A 35-year-old man presents with recent-onset dysphagia. He admits to occasional heartburn but no chest pain. He does not smoke or drink, and there is no family history of relevance. You perform an upper endoscopy, which reveals a smooth large bulge in the lateral wall of the midesophagus with normal-appearing overlying mucosa. Endoscopic biopsy findings are normal. Your next step is to

A. Reassure the patient and bring him back for another endoscopy within 6 months or a year to monitor the progress of the lesion
B. Refer the patient for surgery
C. Schedule the patient for a CT scan of the chest
D. Schedule the patient for an endoscopic ultrasound
E. Dismiss the finding as a normal variant and treat the patient with a course of antireflux medications.

Answer: D These findings suggest a submucosal lesion such as a leiomyoma. The best way to determine its origin is by endoscopic ultrasound, which can also detect any enlarged nodes that may be targeted for cytologic aspiration.

4. A 67-year-old man is being worked up for iron deficiency anemia. Occult blood is detected in the stool, and he undergoes an upper endoscopy and colonoscopy, both of which have normal findings, except for diverticulosis of the colon. Which of the following would you recommend?

A. Trial of iron infusion therapy and monitor blood counts during the next year
B. Empirical trial of estrogen therapy for suspected (occult) telangiectasia of the intestine
C. Proceed to enteroscopy
D. Schedule a capsule endoscopy
E. Refer to a surgeon for exploration of the small bowel by laparoscopy for occult tumor

Answer: D Capsule endoscopy is the procedure of choice to examine the mucosa of the small bowel because of its accuracy and noninvasive nature. If a lesion requires further intervention (biopsy, removal), deep enteroscopy may be undertaken, depending on its nature and location within the small bowel.

5. An 82-year-old man admitted to the orthopedic service after back surgery complains of progressively distended abdomen and obstipation. A plain radiograph of his abdomen reveals marked dilation of the colon, with a cecal diameter of 9 cm. Despite correction of hypokalemia and a trial of neostigmine, there is no improvement. On examination, his abdomen is visibly distended but not tender, and there are no peritoneal signs. The next step for this patient is

A. Trial of intravenous erythromycin
B. Continue with nasogastric suction and intravenous rehydration for another 48 hours
C. Colonoscopy with attempted decompression
D. Immediate surgery and exploration for mechanical obstruction
E. Urgent cecostomy

Answer: C This patient has acute colonic pseudo-obstruction or Ogilvie syndrome. Neostigmine is a first medical measure if there are no cardiac contraindications. Because of the large size of his cecum, he is at risk for perforation if decompression is not successful. Decompression is best attempted with colonoscopy; if this procedure fails, a cecostomy is an appropriate further step.

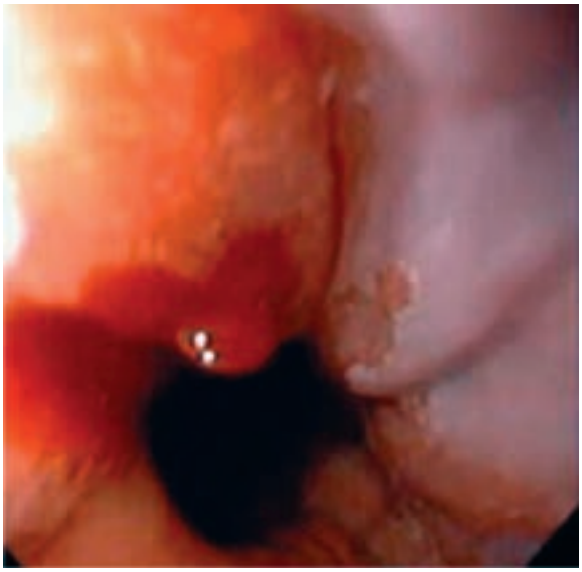

FIGURE 135-1. Bleeding esophageal varix at the gastroesophageal junction. (Courtesy Pankaj Jay Pasricha, MD.)

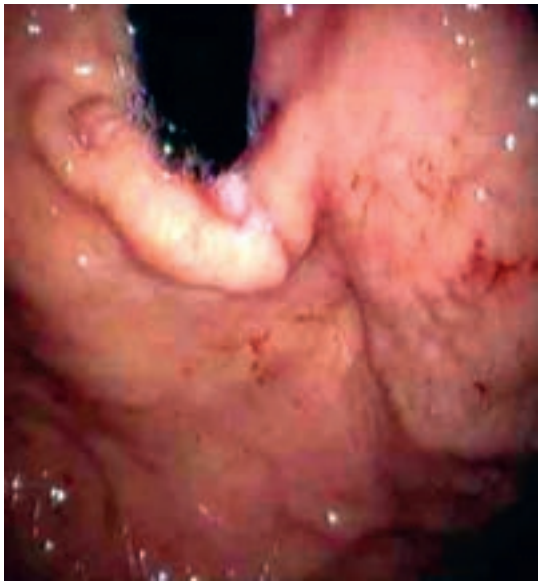

FIGURE 135-2. Retroflexed endoscopic image of a Mallory-Weiss tear at the gastroesophageal junction.

135 GASTROINTESTINAL HEMORRHAGE

THOMAS O. KOVACS AND DENNIS M. JENSEN

BACKGROUND

Gastrointestinal (GI) hemorrhage can be manifested clinically as overt bleeding from the upper GI tract (esophagus, stomach, and duodenum), lower GI tract (colon), or obscure locations (usually in the small intestine). Alternatively, it can occur as occult bleeding detected by iron deficiency anemia (Chapter 159) or by a positive result of fecal occult blood testing (Chapter 193).

GI hemorrhage is a common worldwide clinical problem and continues to be associated with significant morbidity and mortality. The annual hospitalization rate for upper GI bleeding is estimated to be 30 to 100 patients per 100,000, or about 400,000 hospitalizations per year for acute nonvariceal upper GI bleeding in the United States. A report using a national inpatient database showed that hospitalizations for upper GI hemorrhage decreased by more than 20% in the decade of 2001 to 2009. Lower GI bleeding occurs less frequently, with an incidence of 6 to 20 per 100,000, and also has decreased during the past decade. The incidence of lower GI bleeding increases substantially with age (200 per 100,000 by the age of 80 years), and lower GI hemorrhage may occur more frequently than upper GI bleeding in the elderly. Overall, for patients hospitalized for GI bleeding, 40% occurred in the upper GI tract, 25% in the lower GI tract, and 35% in an undefined location.[1] Mortality rates from upper GI hemorrhage are high, varying from 3.5 to 7% in the United States.

UPPER GASTROINTESTINAL BLEEDING

CLINICAL MANIFESTATIONS

Upper GI bleeding occurs proximal to the ligament of Treitz. Patients with upper GI bleeding usually present with hematemesis (vomiting blood or coffee-ground material) or melena (black, tarry stool). In large series, about 50% of patients have hematemesis and melena, about 30% have hematemesis alone, and about 20% have only melena. On occasion, however, hematochezia (passage per rectum of red blood or clots) may be the only manifestation of a bleeding ulcer, and about 15% of all patients who present with hematochezia have an upper GI source.[2] Peptic ulcer disease is the most common cause of acute upper GI hemorrhage, accounting for about 40% of cases.[3] Other common causes are esophageal and gastric varices (Fig. 135-1) and erosive esophagitis (see Fig. 134-1). Variceal bleeding, which occurs in the setting of portal hypertension, is discussed in Chapter 153. Other conditions, such as Mallory-Weiss tears (Fig. 135-2; Chapter 138), angiodysplasia, watermelon stomach, tumors, and Dieulafoy lesion, occur less frequently than peptic ulcer (Table 135-1). The mortality from nonulcer bleeding is comparable to that from ulcer hemorrhage in high-risk patients,[4] so all causes of upper GI hemorrhage contribute to the morbidity and cost of care associated with it.

DIAGNOSIS

Initial assessment includes a medical history, vital signs, physical examination (including digital rectal examination), and nasogastric lavage in an attempt to localize the source of melena or hematochezia to the upper GI tract. Patients should be asked questions that can help determine the diagnostic possibilities for the bleeding source. For example, peptic ulcer bleeding (Chapter 139) should be suspected in patients taking daily aspirin or nonsteroidal anti-inflammatory drugs (NSAIDs). For patients with known or suspected liver disease, bleeding related to portal hypertension (such as varices or portal hypertensive gastropathy; Chapter 153) should be strongly considered. Heavy alcohol intake or vomiting should suggest a Mallory-Weiss tear (Chapter 138). A feeding tube or a chronic nasogastric tube and a history of

TABLE 135-1 CAUSES OF SEVERE UPPER GASTROINTESTINAL (GI) BLEEDING IN ONE LARGE CENTER

DIAGNOSIS	%
Peptic ulcer (gastric or duodenal)	38
Gastric or esophageal varices	16
Erosive esophagitis	13
Upper GI tumors	7
Upper GI angiomas*	6
Mallory-Weiss tear	4
Gastric or duodenal erosions	4
Dieulafoy lesion	2
Other†	2
No upper GI cause found‡	8

*Upper GI angiomas include single or multiple angiectasia, watermelon stomach, and Osler-Weber-Rendu telangiectasia.
†Other lesions were surgical anastomoses, Cameron ulcers, aortoenteric fistulas, and hemobilia.
‡No cause found in esophagus, stomach, or duodenum, but 2% had mouth, nose, or pharyngeal bleeding sites.From Kovacs TO, Jensen DM. Endoscopic therapy for severe ulcer bleeding. *Gastrointest Endosc Clin N Am.* 2011;21:681-696.

TABLE 135-2 ENDOSCOPIC STIGMATA OF RECENT ULCER HEMORRHAGE

ENDOSCOPIC APPEARANCE	FREQUENCY (%)	RISK OF REBLEEDING (%)	REBLEEDING RISK AFTER ENDOSCOPIC HEMOSTASIS (%)*
Active arterial bleeding	12	80-90	15-30
Nonbleeding visible vessel	22	40-50	15-30
Adherent clot	10	30-35	0-5
Oozing without other stigmata	14	10-20	0-5
Flat spot	10	5-10	—†
Clean ulcer base	32	3	—†

*Reduction in bleeding risk is with the administration of a proton pump inhibitor, after successful endoscopic hemostasis.
†Endoscopic hemostasis is not recommended for these stigmata.
From Kovacs TO, Jensen DM. Endoscopic therapy for severe ulcer bleeding. *Gastrointest Endosc Clin N Am.* 2011;21:681-696.

gastroesophageal reflux disease raise the suspicion for severe erosive esophagitis (Chapter 138).

The physician should check the vital signs with attention to signs of hypovolemia, such as hypotension, tachycardia, and orthostasis. The patient's skin should be examined for petechiae (see Fig. 436-5), purpura (see Fig. 436-11), spider angiomas, and palmar erythema (see Fig. 146-2), and the abdomen should be assessed for ascites (see Fig. 146-4), hepatomegaly, or splenomegaly, which may indicate portal hypertension. Tenderness or a mass may indicate an intra-abdominal tumor.

Nasogastric or orogastric tube placement to aspirate gastric contents can potentially be useful to localize bleeding to the upper GI tract and to determine the amount of red blood, coffee-ground material, or nonbloody fluid present. There is little use in testing of nasogastric tube aspirates for blood because the trauma of tube placement may cause bleeding and false-positive results. Patients who have witnessed coffee-ground emesis or fresh bloody emesis do not require a nasogastric tube for diagnostic purposes but may need one to help clear the gastric blood for better endoscopic visualization and to minimize the risk of aspiration.

Peripheral blood should be sent for standard hematology, chemistry, liver, and coagulation studies as well as for typing and crossmatching for packed red blood cells (Chapter 177). Hemoglobin concentration and hematocrit may not accurately reflect blood loss because equilibration with extravascular fluid requires 24 to 72 hours. A low platelet count suggests chronic liver disease, dilution, drug reaction, or a hematologic disorder. In upper GI bleeding, the blood urea nitrogen level typically increases to a greater extent than the creatinine level owing to increased intestinal absorption of urea after the breakdown of blood proteins. However, this phenomenon can be misleading in the setting of renal insufficiency or rapid transit of blood. An elevated international normalized ratio can be observed in chronic liver disease and in patients who are taking warfarin.

TREATMENT

Acute Management

Resuscitation efforts should be initiated simultaneously with assessment in the emergency department and continue during the hospitalization. Large-bore (14- or 16-gauge) intravenous catheters are recommended, with normal saline infused as fast as necessary to maintain hemodynamic stability.[5,6] A restrictive transfusion strategy (red blood cell transfusions given only at hemoglobin level <7 g/dL, with a post-transfusion target of 7 to 9 g/dL) is better than a liberal strategy for reducing 45-day mortality and further bleeding in patients with acute upper GI hemorrhage.■ However, these results cannot be generalized to patients with upper GI bleeding who are hypotensive due to severe hemorrhage or have associated cardiovascular disease, in whom a hemoglobin target of 9 to 10 g/dL is recommended. General guidelines are also to use blood products as needed to maintain the platelet count above 50,000/μL and the international normalized ratio below 2. To prevent aspiration, which can cause considerable morbidity and mortality, endotracheal intubation should be considered in patients with active hematemesis or altered mental status.

Endoscopic Evaluation and Therapy

Endoscopy can identify the site of bleeding and provide therapeutic hemostasis in most patients.[7,8] Patients with evidence of active bleeding (red blood by nasogastric lavage or hypotension) should undergo emergency endoscopy as soon as possible after medical resuscitation.■ An intravenous prokinetic agent (either erythromycin, 250 mg, or metoclopramide, 10 mg) 30 to 60 minutes before endoscopy may help move blood out of the stomach and into the small intestine, improving endoscopic visualization.■

In addition to localization of the bleeding source, endoscopic evaluation can provide prognostic information and stratify the risk of rebleeding on the basis of the presence or absence of stigmata of recent hemorrhage (Table 135-2). In addition to endoscopic stigmata, other clinical and laboratory factors that predict a poor prognosis include older age, bleeding onset in the hospital, medical comorbidities, shock, coagulopathy, fresh blood in the nasogastric lavage, and the need for multiple blood transfusions. Clinical scoring systems, such as the Rockall score (Table 135-3), use clinical information and endoscopic findings to predict clinical outcomes. Endoscopic Doppler ultrasound may also help in the risk stratification of patients with ulcer hemorrhage. Persistence of a positive Doppler signal after endoscopic treatment correlates with rebleeding.

The goal of endoscopic therapy is to stop acute bleeding and to reduce the risk of recurrent bleeding. Most endoscopic therapies were designed for peptic ulcer hemorrhage, but they can be used for other causes of nonvariceal upper GI bleeding, in which underlying arteries are the source of bleeding. Available treatments include injection (epinephrine or sclerosants), thermal coagulation (with multipolar/bipolar or heater probe), and mechanical compression (hemostatic clips). Injection therapy is effective, safe, and inexpensive, but it is sometimes inadequate for definitive hemostasis when it is used alone. When epinephrine injection is combined with either thermal coagulation or hemoclips, hemostasis is achieved in more than 95% of patients with active bleeding, and rebleeding rates are decreased by more than 50%.■ Endoscopic therapy is reserved for lesions that have high-risk stigmata for rebleeding (e.g., active arterial bleeding, nonbleeding visible vessel, or adherent clot) or moderate-risk stigmata, such as oozing blood without associated clot or nonbleeding visible vessel. It is not warranted for low-risk stigmata (clean-based ulcer or flat pigmented spot), which have the lowest risk of rebleeding (Fig. 135-3). Some causes of upper GI bleeding, such as gastric or duodenal erosions and neoplasms, generally are not amenable to endoscopic treatment and require appropriate medical or surgical therapy.

Medical Therapy

The most common cause of upper GI bleeding is peptic ulcer disease, and the most common cause of peptic ulcer disease is NSAID use. Aspirin and other NSAIDs should be discontinued in patients with bleeding peptic ulcers, unless there is a contraindication to do so (e.g., secondary prophylaxis for stroke). Patients with documented *Helicobacter pylori* infection should be treated with combination antibiotics and proton pump inhibitors (Chapter 139).[9] After documentation of *H. pylori* eradication, maintenance antisecretory treatment is not needed unless the patient also requires NSAIDs or anticoagulation therapy.

Proton pump inhibitors (see Table 138-1) are the mainstay of medical therapy for hemostasis and healing of peptic lesions. Acid suppression can promote platelet aggregation and clot formation as well as reduce the risk of rebleeding. High-dose intravenous proton pump inhibitors (bolus followed by

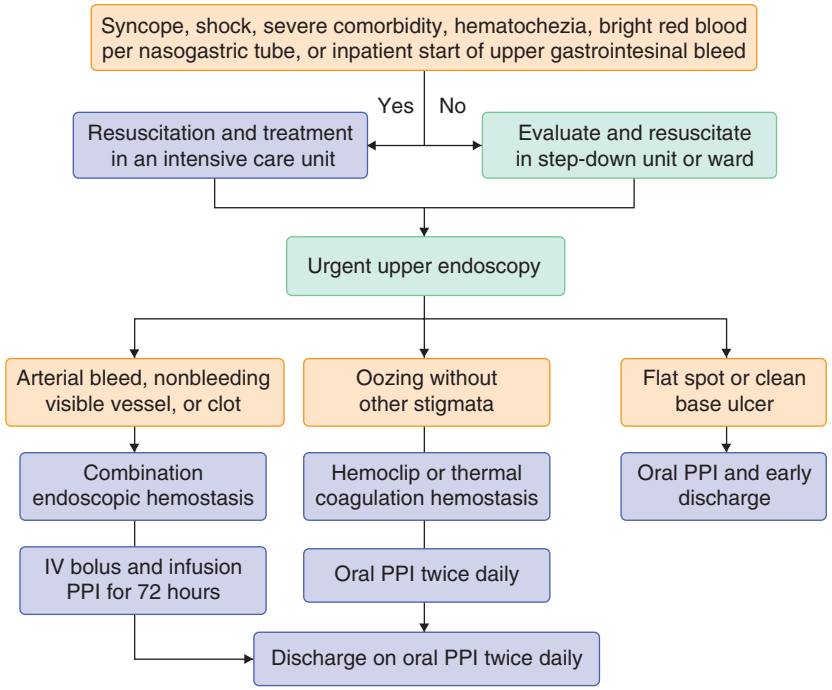

FIGURE 135-3. Management algorithm for nonvariceal upper gastrointestinal hemorrhage. IV = intravenous; PPI = proton pump inhibitor.

TABLE 135-3 COMPLETE ROCKALL SCORING SYSTEM FOR UPPER GASTROINTESTINAL BLEEDING

VARIABLE	0 POINTS	1 POINT	2 POINTS	3 POINTS
Age (years)	<60	60-79	>80	—
Shock				
Pulse rate (beats/min)	<100	>100	—	—
Systolic blood pressure	>100	>100	<100	—
Comorbidity	None	—	Ischemic heart disease, cardiac failure, other major illness	Renal failure, hepatic failure, metastatic cancer
Endoscopic stigmata of recent hemorrhage	No stigmata or dark spot in ulcer base	—	Blood in upper gastrointestinal tract, adherent clot, visible vessel, active bleeding	—
Diagnosis	Mallory-Weiss tear or no lesion seen	All other diagnoses	Malignant lesions	—

PRE-ENDOSCOPY SCORE	MORTALITY (%)	POST-ENDOSCOPY SCORE	REBLEED RATE (%)	MORTALITY (%)
0	0.2	0	4.9	0
1	2.4	1	3.4	0
2	5.6	2	5.3	0.2
3	11	3	11.2	2.9
4	24.6	4	14.1	5.3
5	39.6	5	24.1	10.8
6	48.9	6	32.9	17.3
7	50	7	43.8	27
—	—	8+	41.8	41.1

From Rockall TA, Logan RFA, Devlin HB, Northfield TC. Risk assessment after acute upper gastrointestinal haemorrhage. *Gut.* 1996; 38:316-321.

continuous infusion [e.g., pantoprazole, 80 mg, followed by 8 mg/hour for 72 hours]) after successful endoscopic hemostasis reduce rebleeding rates and mortality in patients with high-risk stigmata of recent hemorrhage.[9] For patients with high-risk stigmata, either this same regimen or intermittent high-dose proton pump inhibitor therapy appears to be equally effective.[9] Patients with low-risk or no stigmata can be treated with an oral proton pump inhibitor (e.g., esomeprazole or pantoprazole, 40 mg twice daily) and considered for early discharge. Pre-endoscopic proton pump inhibitor therapy does not alter clinical outcomes,[9] but it may downstage the severity of the lesion and decrease the need for endoscopic intervention. Therefore, in patients with suspected ulcer bleeding, proton pump inhibitors given as an intravenous bolus may be considered while waiting for the endoscopy, but this therapy cannot substitute for endoscopy and should not delay it.

The optimal management approach for patients who develop peptic ulcer–related GI bleeding while receiving antiplatelet or anticoagulant therapy is controversial. In a randomized trial of patients who had a bleeding peptic ulcer while receiving low-dose aspirin therapy followed by successful endoscopic hemostasis and high-dose intravenous proton pump inhibitor therapy, the incidence of recurrent ulcer bleeding at 30 days was twice as high (10% vs. 5%) in patients who continued low-dose aspirin compared with those who did not. However, the patients who continued low-dose aspirin had a significantly lower mortality rate compared with the placebo group because of lower rates of cardiovascular and cerebrovascular complications.[9] Current recommendations suggest that patients with upper GI hemorrhage who need secondary cardiovascular prophylaxis should resume low-dose aspirin treatment as soon as the cardiovascular risks outweigh the gastrointestinal risks, usually within 7 days, while continuing proton pump inhibitors (e.g., esomeprazole or pantoprazole, 40 mg daily).[10] Patients with idiopathic (non–*H. pylori*, non-NSAID) peptic ulcers should receive long-term proton pump inhibitor therapy.

LOWER GASTROINTESTINAL BLEEDING

DEFINITION AND EPIDEMIOLOGY

Lower GI hemorrhage generally refers to bleeding from the colon and anorectum. Severe lower GI bleeding occurs with an annual incidence of 20 per 100,000 population, much less frequently than upper GI bleeding. Patients are usually older and present with painless hematochezia, typically without orthostasis.[11] If orthostasis is present, a brisk upper GI bleed, which occurs in 15 to 20% of cases of severe hematochezia, should be considered. The most common cause of severe colorectal hemorrhage is from diverticulosis (Chapter 142). Other frequent causes include internal hemorrhoids (Chapter 145), ischemic colitis (Chapter 143), rectal ulcers (Chapter 145), and delayed bleeding from post-polypectomy ulcers at a median of 8 days (range, 5 hours to 17 days) after the procedure (Table 135-4; Chapter 134). In most patients, the bleeding stops and does not recur. The overall mortality rate from lower GI bleeding is 2 to 4%.

DIAGNOSIS

Patients with hematochezia should have a careful history, physical examination, and laboratory evaluation, analogous to the evaluation for upper GI bleeding. The history may help in the differential diagnosis. A history of diverticulosis (see Figs. 142-5 and 142-6) may raise the suspicion for a diverticular bleed (Chapter 142).[12] Abdominal cramping followed by bloody diarrhea suggests ischemic colitis (see Fig. 143-2). A recent polypectomy makes a post-polypectomy bleed more likely.

TABLE 135-4 MOST COMMON COLONIC SOURCES OF SEVERE HEMATOCHEZIA IN A SERIES OF 486 CASES (EXPRESSED AS PERCENTAGE OF COLONIC SOURCES)*

Diverticulosis	32%
Internal hemorrhoids	13%
Ischemic colitis	12%
Rectal ulcers	8%
Colonic angiodysplasia, angiectasia, angiomas, or radiation telangiectasias	7%
Ulcerative colitis, Crohn, other colitis	6%
Post-polypectomy ulcer	5%
Other lower gastrointestinal tract sources	6%

*Less common causes include surgical anastomotic ulcers or sutures, nonsteroidal anti-inflammatory agent colopathy, metastases, portal hypertensive colopathy, lymphoma, and endometriosis.
Modified from Jensen DM. Management of patients with severe hematochezia—with all current evidence available. *Am J Gastroenterol.* 2005;100:2403-2406.

After medical resuscitation, most patients will need to undergo colonoscopy, although a flexible sigmoidoscopy or anoscopy may be an alternative in cases in which it is highly likely that the bleeding source is from the anorectum or distal colon (Fig. 135-4). Colonoscopy is critical for identifying a luminal source of bleeding as well as for potential hemostasis of amenable lesions. Urgent colonoscopy within 12 hours of admission improves diagnostic yield but has not been proved to reduce the rate of rebleeding. To cleanse the colon adequately and to visualize the colonic mucosa, a bowel purge with 6 L or more of a polyethylene glycol solution should be administered before the procedure. Colonoscopic treatment (Fig. 135-4) of focal bleeding sources with stigmata of hemorrhage uses the same methods (epinephrine and either thermocoagulation or hemoclips) as for upper GI bleeding.[13,14] In addition, coagulation of angiectasias and bleeding biopsy sites is feasible during colonoscopy.

If colonoscopy does not reveal the source of bleeding, an upper endoscopic evaluation should be performed. If both are negative, capsule endoscopy (Chapter 134), if it is locally available, may be the preferred way to look for a small intestinal source of bleeding. Alternatively, if bleeding persists or is too rapid for a colonoscopy to be performed, a tagged red blood cell nuclear scan or angiography may be performed to localize the bleeding.[15] The red blood cell scan can help localize bleeding if it occurs at a rate of at least 0.1 mL/minute. Angiography has the advantage that it can treat the bleeding source with embolization of the bleeding vessel. However, it requires a faster bleeding rate (at least 0.5 mL/minute). Neither red blood cell scans nor angiography can identify nonbleeding stigmata, and usually neither yields an etiologic diagnosis.

Surgical management rarely is needed for hemostasis of lower GI bleeding because most bleeding is either self-limited or easily managed with medical or endoscopic therapy. The main indications for surgery are malignant lesions (Chapters 145 and 193), diffusely bleeding lesions that fail to respond to medical therapy (such as ischemia), and recurrent diverticular hemorrhage (Chapter 142). If the bleeding source can be localized preoperatively to a particular area of the colon, a segmental colonic resection can be performed rather than a subtotal colectomy.

OBSCURE AND OCCULT GASTROINTESTINAL BLEEDING

DEFINITION AND EPIDEMIOLOGY

Obscure GI bleeding is persistent or recurrent bleeding, despite negative initial GI evaluation, including upper endoscopy, colonoscopy, and radiologic evaluation of the small intestine, such as small bowel follow-through. Obscure GI bleeding can be classified as either overt (melena, maroon stool, or hematochezia) or occult bleeding (positive result of fecal occult blood testing, usually in the setting of iron deficiency anemia).

In most large series of hospitalized patients, 5% of overt GI bleeding cases are considered to be obscure, and 75% of these patients have bleeding from the small intestine that is beyond the reach of an upper endoscope or

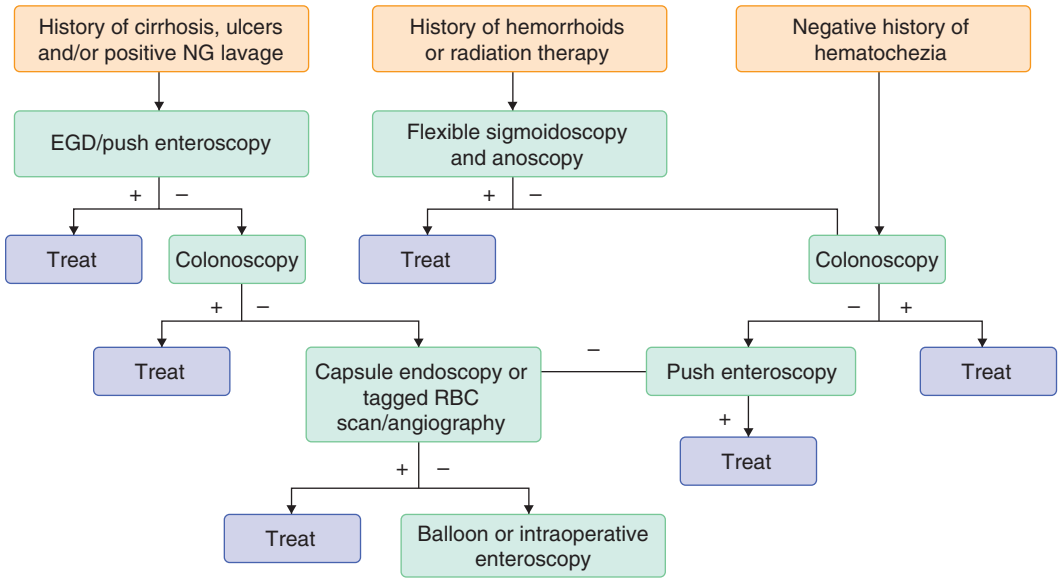

FIGURE 135-4. Management algorithm for severe hematochezia. EGD = esophagogastroduodenoscopy; NG = nasogastric; RBC = red blood cell.

colonoscope. Angiectasias (see Fig. 134-4) are the most common source of small intestinal bleeding, followed by ulcers and tumors (Table 135-5). Other causes for obscure GI bleeding include lesions that are within reach of standard endoscopes but that were not recognized as the bleeding site (e.g., a large hiatal hernia with lesions known as Cameron ulcers on endoscopy, or internal hemorrhoids on colonoscopy) and intermittently bleeding lesions such as a Dieulafoy lesion (Fig. 135-5), which is an aberrant submucosal vessel without an ulcer.

Iron deficiency (Chapter 159) has a prevalence of 2 to 5% among adult men and postmenopausal women. It can occur from overt or occult blood loss (e.g., GI tract lesions, menorrhagia), iron malabsorption (celiac disease, atrophic gastritis), and red blood cell destruction (hemolysis). It should be suspected in patients with low mean corpuscular volume, low ferritin level, or low transferrin saturation.

TABLE 135-5 SMALL INTESTINAL LESIONS FOUND IN 488 PATIENTS DURING DOUBLE-BALLOON ENTEROSCOPY FOR OBSCURE GASTROINTESTINAL BLEEDING

LESION	FREQUENCY (RANGE)
None	40% (0-57)
Angiectasias	31% (6-55)
Ulcerations	13% (2-35)
Malignancy	8% (3-26)
Other	6% (2-22)

From Rajir GS, Gerson L, Das A, et al. American Gastroenterological Association (AGA) Institute technical review on obscure gastrointestinal bleeding. *Gastroenterology.* 2007;133:1697-1717.

DIAGNOSIS

The approaches to overt and occult obscure GI bleeding are similar (Fig. 135-6). In the case of recurrent overt bleeding, upper endoscopy and colonoscopy should be repeated, with the type of bleeding dictating which endoscopic procedure to do first. Colonoscopy with anoscopy should be done first if there is hematochezia. If there is melena, a push small bowel enteroscopy

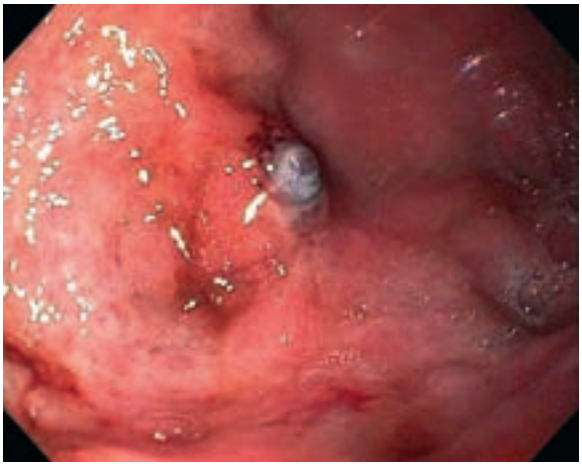

FIGURE 135-5. Dieulafoy lesion.

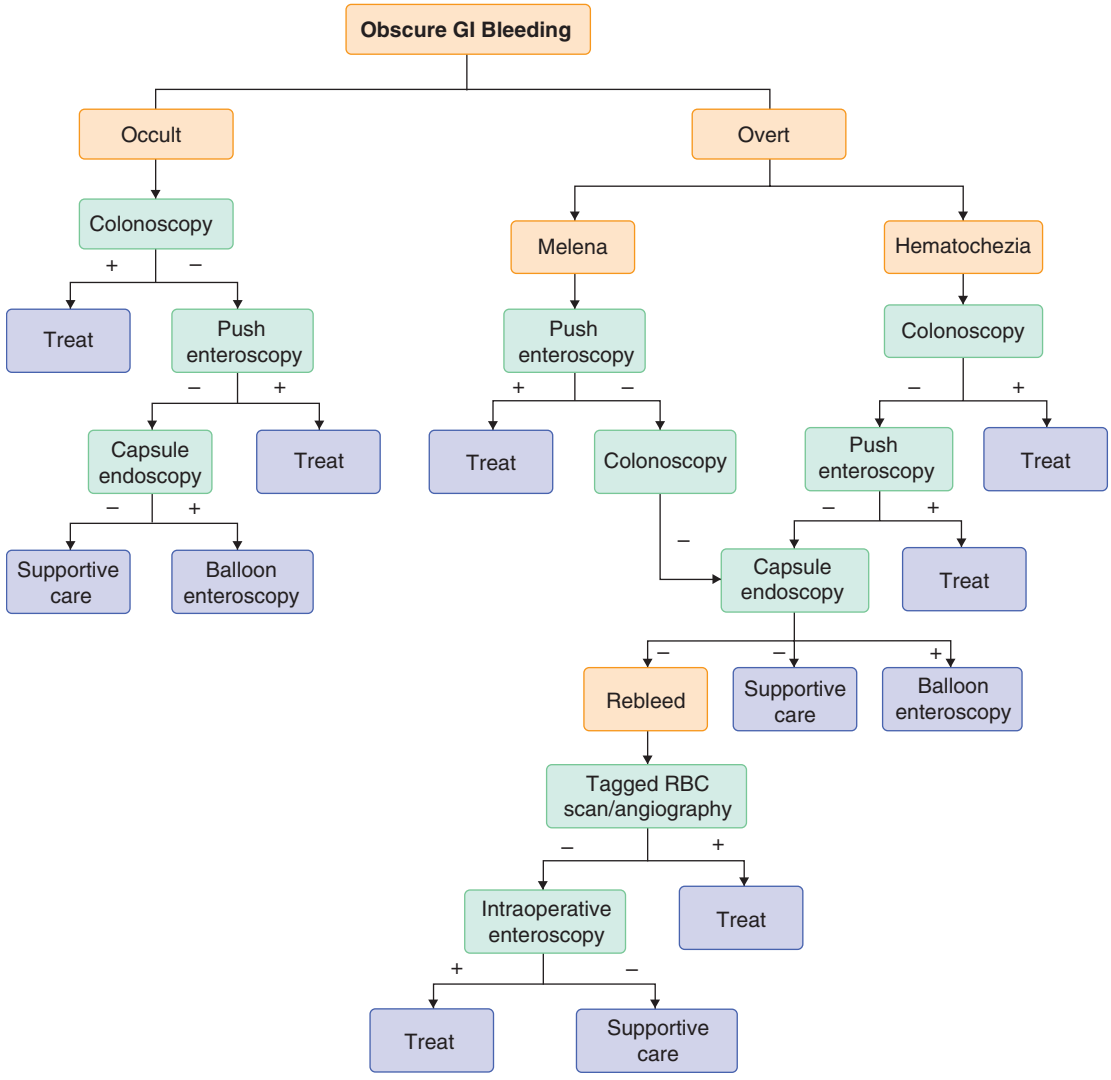

FIGURE 135-6. Management algorithm for obscure gastrointestinal bleeding. GI = gastrointestinal; RBC = red blood cell.

should be performed. If the first procedure is unremarkable, evaluation should be undertaken from the opposite end. If the second test result is negative, capsule endoscopy should be performed (Chapter 134).[16] If bleeding from the small intestine is seen on capsule endoscopy, further attempts to diagnose and to treat the bleeding should be performed with either deep enteroscopy (using a balloon overtube to slide much farther along into the small intestine) or intraoperative enteroscopy. All of these long enteroscopes facilitate diagnosis and hemostasis. If the capsule endoscopy is negative and rapid rebleeding recurs, a tagged red blood cell scan or angiography may be used to localize the bleeding site and assist with subsequent intraoperative enteroscopy.

For occult GI bleeding, colonoscopy should be performed first because fecal occult blood testing was designed to screen for colorectal cancer (Chapter 193). Upper endoscopy and push enteroscopy should follow if the colonoscopy is unremarkable. Afterward, the algorithm is the same as for overt bleeding; however, if the capsule endoscopy is negative, efforts should be focused toward providing supportive care rather than further evaluation.

Iron Deficiency Anemia

GI evaluation of iron deficiency anemia (Chapter 159) is indicated in adult men, regardless of age, and postmenopausal women. Women who have not reached menopause may warrant a GI evaluation after obvious or potential causes of iron deficiency and blood loss, such as chronic menorrhagia, have been excluded. Colonoscopy should be performed first, followed by upper endoscopy and push enteroscopy if the colonoscopy is negative. Duodenal biopsy specimens should be taken to look for evidence of celiac disease (Chapter 140). If all three of these endoscopic procedures are unrevealing, capsule endoscopy should be performed. If the capsule study is negative, investigation into non-GI causes of iron deficiency (Chapter 159) may be pursued.

Grade A References

A1. Villanueva C, Colomo A, Bosch A, et al. Transfusion strategies for acute upper gastrointestinal bleeding. *N Engl J Med*. 2013;368:11-21.

A2. Barkun AN, Bardou M, Kuipers EJ, et al. International consensus recommendations on the management of patients with nonvariceal upper gastrointestinal bleeding. *Ann Intern Med*. 2010;152:101-113.

A3. Barkun AN, Bardou M, Martel M, et al. Prokinetics in acute upper GI bleeding: a meta-analysis. *Gastrointest Endosc*. 2010;72:1138-1145.

A4. Vergara M, Bennett C, Calvet X, et al. Epinephrine injection versus epinephrine injection and a second endoscopic method in high-risk bleeding ulcers. *Cochrane Database Syst Rev*. 2014;10:CD005584.

A5. Sung JJ, Barkun A, Kuipers EJ, et al. Intravenous esomeprazole for prevention of recurrent peptic ulcer bleeding: a randomized trial. *Ann Intern Med*. 2009;150:455-464.

A6. Sachar H, Vaidya K, Laine L. Intermittent vs continuous proton pump inhibitor therapy for high-risk bleeding ulcers: a systematic review and meta-analysis. *JAMA Intern Med*. 2014;174:1755-1762.

A7. Sreedharan A, Martin J, Leontiadis GI, et al. Proton pump inhibitor treatment initiated prior to endoscopic diagnosis in upper gastrointestinal bleeding. *Cochrane Database Syst Rev*. 2010;7:CD005415.

A8. Sung JJ, Lau JY, Ching JY, et al. Continuation of low-dose aspirin therapy in peptic ulcer bleeding: a randomized trial. *Ann Intern Med*. 2010;152:1-9.

A9. Laine L, Shah A. Randomized trial of urgent vs. elective colonoscopy in patients hospitalized with lower GI bleeding. *Am J Gastroenterol*. 2010;105:2636-2641.

A10. Leung WK, Ho SS, Suen BY, et al. Capsule endoscopy or angiography in patients with acute overt obscure gastrointestinal bleeding: a prospective randomized study with long-term follow-up. *Am J Gastroenterol*. 2012;107:1370-1376.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Laine L, Yang H, Chang SC, et al. Trends for incidence of hospitalization and death due to GI complications in the United States from 2001 to 2009. *Am J Gastroenterol*. 2012;107:1190-1195.
2. Kim BS, Li BT, Engel A, et al. Diagnosis of gastrointestinal bleeding: A practical guide for clinicians. *World J Gastrointest Pathophysiol*. 2014;5:467-478.
3. Kovacs TO, Jensen DM. Endoscopic therapy for severe ulcer bleeding. *Gastrointest Endosc Clin N Am*. 2011;21:681-696.
4. Marmo R, Del Piano M, Rotondano G, et al. Mortality from nonulcer bleeding is similar to that of ulcer bleeding in high-risk patients with nonvariceal hemorrhage: a prospective database study in Italy. *Gastrointest Endosc*. 2012;75:263-272.
5. Laine L, Jensen DM. Management of patients with ulcer bleeding. *Am J Gastroenterol*. 2012;107: 345-360.
6. Feinman M, Haut ER. Upper gastrointestinal bleeding. *Surg Clin North Am*. 2014;94:43-53.
7. Hwang JH, Fisher DA, Ben-Menachem T, et al. The role of endoscopy in the management of acute non-variceal upper GI bleeding. *Gastrointest Endosc*. 2012;75:1132-1138.
8. Lu Y, Loffroy R, Lau JY, et al. Multidisciplinary management strategies for acute non-variceal upper gastrointestinal bleeding. *Br J Surg*. 2014;101:e34-e50.
9. McColl KE. Clinical practice. *Helicobacter pylori* infection. *N Engl J Med*. 2010;362:1597-1604.
10. Lau JYW, Barkun A, Fan D, et al. Challenges in the management of acute peptic ulcer bleeding. *Lancet*. 2013;381:2033-2043.
11. Cavallaro LG, Monica F, Germana B, et al. Time trends and outcome of gastrointestinal bleeding in the Veneto region: a retrospective population based study from 2001 to 2010. *Dig Liver Dis*. 2014;46:313-317.
12. Jensen DM. The ins and outs of diverticular bleeding. *Gastrointest Endosc*. 2012;75:388-391.
13. Strate LL, Naumann CR. The role of colonoscopy and radiological procedures in the management of acute lower intestinal bleeding. *Clin Gastroenterol Hepatol*. 2010;8:333-343.
14. Kaltenbach T, Watson R, Shah J, et al. Colonoscopy with clipping is useful in the diagnosis and treatment of diverticular bleeding. *Clin Gastroenterol Hepatol*. 2012;10:131-137.
15. Lepileur L, Dray X, Antonietti M, et al. Factors associated with diagnosis of obscure gastrointestinal bleeding by video capsule enteroscopy. *Clin Gastroenterol Hepatol*. 2012;10:1376-1380.
16. Eliakim R. Video capsule endoscopy of the small bowel. *Curr Opin Gastroenterol*. 2013;29: 133-139.

REVIEW QUESTIONS

1. A 50-year-old man without any prior medical problems began taking ibuprofen 800 mg three times daily for lower back pain after a work-related injury. He subsequently developed nausea followed by hematemesis and melena. He now presents to the emergency department for further evaluation. On the basis of this presentation and epidemiologic studies, what is the most likely cause of the suspected upper gastrointestinal (GI) hemorrhage?

A. Peptic ulcer
B. Mallory-Weiss tear
C. Esophagitis
D. Esophageal varices
E. Dieulafoy lesion

Answer: A This patient began taking a nonsteroidal anti-inflammatory drug (NSAID) and subsequently presented with melena and hematemesis indicative of upper GI bleeding. Both epidemiologic data and the clinical history support peptic ulcer disease as the most likely cause of the bleeding. A Mallory-Weiss tear is possible on the basis of the clinical presentation but is less likely to be associated with NSAID use and is not the likely cause of his bleeding. (Laine L, Jensen DM. Management of patients with ulcer bleeding. *Am J Gastroenterol.* 2012;107:345-360.)

2. A 70-year-old man who had a mechanical aortic valve implanted for aortic stenosis presents to the emergency department with painless hematochezia. He is receiving warfarin and takes aspirin 81 mg daily. He is volume resuscitated with 6 units of packed red blood cells and stabilized. An urgent colonoscopy demonstrates no fresh blood in the terminal ileum or colon. There are a few, scattered diverticula. What would you do next?

A. Upper endoscopy
B. Tagged red blood cell scan
C. Nothing
D. Angiography
E. Surgery

Answer: A This patient presents with hematochezia that is suggestive of a lower bleeding source. However, up to 15 to 18% of cases of significant hematochezia have an upper GI tract cause, such as peptic ulcer or esophageal varices. For this reason, the patient should be further evaluated with an upper GI endoscopy. As a matter of routine practice, high-volume hematochezia (as indicated by hemodynamic compromise, high volume requirements for volume resuscitation, and a large fall in the hematocrit, particularly when it occurs in conjunction with risk factors for peptic ulceration or portal hypertension) should prompt an urgent upper GI endoscopy as the first intervention after resuscitation, followed by a rapid bowel purge and colonoscopy. (Kovacs TOG, Jensen DM. Upper or small bowel hemorrhage that presents as hematochezia. *Tech Gastrointest Endosc.* 2001;3:206-215; Laine L, Jensen DM. Management of patients with ulcer bleeding. *Am J Gastroenterol.* 2012;107:345-360.)

3. A 80-year-old woman presents with melena, hematemesis, and syncope. Examination reveals hypotension and tachycardia. What is the first step in management?

A. Emergent endoscopy
B. Nasogastric lavage
C. Intravenous proton pump inhibitor
D. Tagged red blood cell scan
E. Intravenous access and intravascular volume repletion

Answer: E The first step in any active GI hemorrhage is volume resuscitation. In this scenario, the patient has lost at least 50% of her intravascular volume, so aggressive resuscitation with both blood and crystalloid products is required. Emergent endoscopy (A), intravenous proton pump inhibitor (C), and nasogastric lavage (B) may be necessary, but only after adequate repletion of the intravascular volume. In general, few patients die of complete exsanguination, particularly in the setting of peptic ulcer bleeding, whereas the majority die as a consequence of the physiologic impact of sudden, dramatic volume shifts, with a disproportionate mortality in the elderly, who may suffer a stroke, myocardial infarction, and acute renal failure if they are not volume resuscitated promptly. Volume resuscitation is initiated with saline, and it is inappropriate to delay while waiting for blood transfusion. (Laine L, Jensen DM. Management of patients with ulcer bleeding. *Am J Gastroenterol.* 2012;107:345-360.)

4. A 70-year-old man with a history of coronary artery disease is admitted to the hospital for melena and dizziness. He takes aspirin 81 mg daily for secondary prophylaxis of cardiovascular disease without concurrent acid suppression. Endoscopy reveals an antral ulcer with a nonbleeding visible vessel that is treated with endoscopic hemostasis. Biopsy shows no evidence of *H. pylori*. He is started on a proton pump inhibitor (PPI). What would be the best course of action?

A. Continue PPI daily and do not restart aspirin.
B. Continue PPI daily and restart aspirin in 4 to 6 weeks.
C. Continue PPI daily and restart aspirin within 1 week.
D. Stop PPI in 4 to 6 weeks and restart aspirin on discharge.
E. Stop PPI and restart aspirin in 4 to 6 weeks.

Answer: C In patients who need aspirin for cardioprotection, the antiplatelet agent should be restarted as soon as cardiovascular risks outweigh GI risks. In a study of patients randomized to restart aspirin versus placebo soon after endoscopic treatment of bleeding ulcer disease, those taking aspirin had more bleeding but significantly lower mortality and cardiovascular events. Therefore, aspirin usually should be restarted within the first week after endoscopic therapy while continuing PPI therapy as well. (Sung JJ, Lau JY, Ching JY, et al. Continuation of low-dose aspirin therapy in peptic ulcer bleeding: a randomized trial. *Ann Intern Med.* 2010;152:1-9.)

5. A 70-year-old woman presents to the emergency department with dizziness and five episodes of bright red blood per rectum in the last 24 hours. Nasogastric tube lavage yields bilious fluid without blood. What is the most common cause of severe hematochezia?

A. Diverticulosis
B. Colonic angiodysplasia
C. Internal hemorrhoids
D. Ulcerative colitis
E. Ischemic colitis

Answer: A Diverticulosis, which is the most common cause of severe hematochezia, accounts for about one third of the colonic sources of severe hematochezia. Internal hemorrhoids and ischemic colitis are the next most common causes. Diverticular hemorrhage may be self-limited, but in severe cases, either colonoscopy or angiography can find and treat the bleeding diverticulum. Several studies have shown at least some benefit of early colonoscopy compared with radiologic testing to determine the location of bleeding and to decrease overall cost. (Jensen DM. The ins and outs of diverticular bleeding. *Gastrointest Endosc.* 2012;75:388-391.)

136

DISORDERS OF GASTROINTESTINAL MOTILITY

MICHAEL CAMILLERI

DEFINITION

Motility disorders result from impaired control of the neuromuscular apparatus of the gastrointestinal tract. Associated symptoms include recurrent or chronic nausea, vomiting, bloating, abdominal discomfort, and constipation or diarrhea in the absence of intestinal obstruction.

PATHOBIOLOGY

Normal Physiology

Neuroenteric Control

Motor function of the gastrointestinal tract depends on the contraction of smooth muscle cells and their integration and modulation by enteric and extrinsic nerves as well as the interstitial cells of Cajal. Extrinsic neural control of gastrointestinal motor function comprises the cranial and sacral parasympathetic outflow (excitatory to nonsphincteric muscle) and the thoracolumbar sympathetic supply (excitatory to sphincters, inhibitory to nonsphincteric muscle). The cranial outflow is predominantly through the vagus nerve, which innervates the gastrointestinal tract from the stomach to the right colon. Parasympathetic innervation of the colon is provided by the vagal fibers coursing along the ileocolonic branches of the superior mesenteric artery and the S2 to S4 parasympathetic supply to the distal colon. Sympathetic fibers to the stomach and small bowel arise from T5 to T10 levels of the intermediolateral column of the spinal cord. Sympathetic innervation of the colon arises from T11 to L3 levels of the spinal cord. The prevertebral ganglia play an important role in the integration of afferent impulses between the gut and the central nervous system and in the reflex control of abdominal viscera.

The enteric nervous system is an independent nervous system comprising approximately 100 million neurons organized into ganglionated plexuses. The larger myenteric or Auerbach plexus is situated between the longitudinal and circular muscle layers of the muscularis externa; this plexus contains neurons responsible for gastrointestinal motility. The submucosal or Meissner plexus controls absorption, secretion, and mucosal blood flow. The enteric nervous system also plays an important role in visceral afferent function.

The interstitial cells of Cajal are spontaneously active pacemaker cells that coordinate muscle contraction and sense distortion. They form a non-neural pacemaker system predominantly at the interface of the circular and longitudinal muscle layers of the intestine as well as within the muscle layers themselves and function as intermediaries between the neurogenic enteric nervous system and the myogenic control system. Electrical control activity spreads through interneurons in the contiguous segments of the gut through neurochemical activation by transmitters that may be excitatory (e.g., acetylcholine, substance P) or inhibitory (e.g., nitric oxide, somatostatin).

Gastric and Small Bowel Motility

The motor functions of the stomach and small intestine are characterized by distinct patterns of motor activity in the fasting and postprandial periods (Fig. 136-1). The fasting or interdigestive period is characterized by a cyclic motor phenomenon, the interdigestive migrating motor complex. In healthy individuals, one cycle of this complex is completed every 60 to 90 minutes. The complex has three phases: a period of quiescence (phase I), a period of intermittent pressure activity (phase II), and an activity front (phase III) during which the stomach and small intestine contract at highest frequencies (3 per minute in the stomach, 12 per minute in the duodenum, 8 per minute in the ileum). Another characteristic motor pattern in the distal small intestine is the giant migrating complex, or power contraction, which empties residue from the ileum into the colon in bolus transfers.

In response to food ingestion, the proximal stomach accommodates food by a vagally mediated reduction in tone, thereby facilitating the ingestion of food without an increase in pressure. Liquids empty from the stomach in an exponential manner, and the rate of emptying varies with calorie content and viscosity. The half-emptying time for non-nutrient liquids in healthy individuals is usually less than 20 minutes. Solids are retained selectively in the stomach, where they undergo acid and peptic digestion as well as "churning" or trituration by high liquid shearing forces in the antrum. Digestible food particles are emptied after their size is reduced by trituration to less than 2 mm. Gastric emptying of solids is characterized by an initial lag period followed by a linear postlag emptying phase. Secretion of hormones that mediate the motor and digestive process (e.g., gastrin for acid secretion; cholecystokinin for gallbladder contraction and bile and pancreatic secretion; and insulin, glucagon, and incretins such as glucagon-like peptide 1 for glucose regulation) is integrated with the arrival of food or chyme at different levels of the gut to ensure optimal digestion.

The small intestine transports solids and liquids at approximately the same rate. As a result of the lag phase for the transport of solids from the stomach, liquids typically arrive in the colon before solids. Chyme moves from ileum to colon intermittently in boluses propelled by contractions.

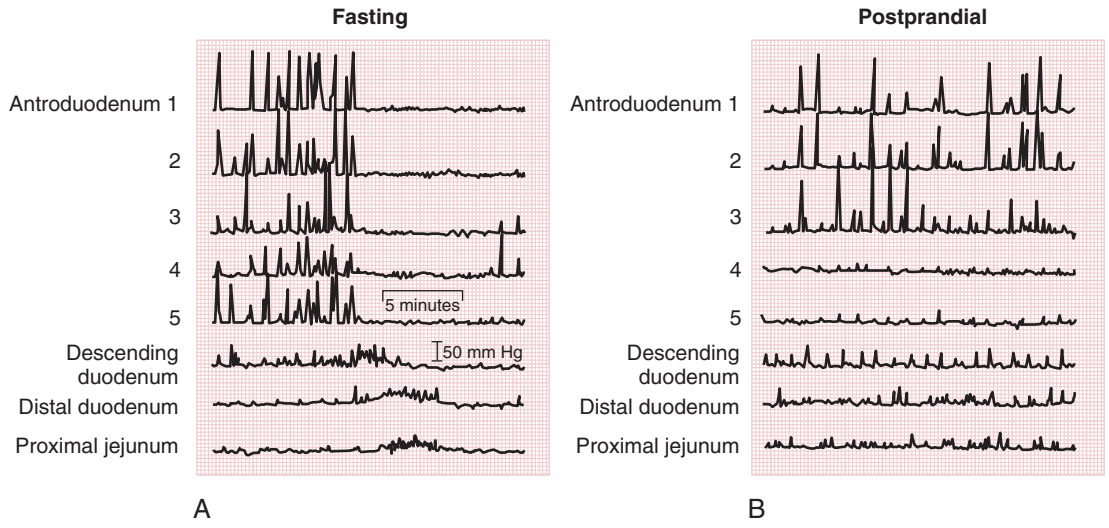

FIGURE 136-1. Fasting and postprandial gastroduodenal manometric recordings in a healthy volunteer. A 535-kcal meal is ingested during the study. Note the cyclic interdigestive migrating motor complex (**A**) and the sustained, high-amplitude but irregular pressure activity after a meal (**B**). (From Coulie B, Camilleri M. Intestinal pseudo-obstruction. *Annu Rev Med*. 1999;50:37-55.)

In the postprandial period, the interdigestive migrating motor complex is replaced by an irregular pattern of variable amplitude and frequency. This pattern, which enables mixing, digestion, and absorption, is observed in the gastrointestinal regions in contact with food. The maximum frequency of contractions is lower than during phase III of the interdigestive motor complex, and the duration of this postprandial contractile activity is proportional to the number of calories consumed (about 1 hour for each 200 kcal ingested). Segments of the small intestine that are not in contact with food continue with interdigestive motor patterns.

Vomiting is characterized by a stereotypic sequence of motor events, including contractions of the stomach, abdominal muscles, and diaphragm. This sequence is followed immediately in the proximal small bowel by a propagated, rhythmic contractile response similar to the migrating motor complex.

Colonic Motility

The normal colon displays short-duration (phasic) contractions and a background contractility or tone. Nonpropagated phasic contractions have a role in segmenting the colon into haustra, which compartmentalize the colon and facilitate mixing, retention of residue, and formation of solid stool. High-amplitude propagated contractions, which are characterized by an amplitude greater than 75 mm Hg, propagation over a distance of at least 15 cm, and a propagation velocity of 0.15 to 2.2 cm/second, contribute to the mass movements in the colon. In health, these contractions occur on average five or six times per day, most often postprandially and between 6 AM and 2 PM.

Colonic transit is a discontinuous process, slow most of the time and rapid at other times. Residue may be retained for prolonged periods in the right colon, and a mass movement may deliver the contents to the sigmoid colon in seconds. Feeding stimulates movement of colonic content (referred to as the gastrocolonic response). In health, the average mouth-to-cecum transit time is about 6 hours, and transit times through the right colon, left colon, and sigmoid colon are about 12 hours each. As the ingestion of dietary fiber increases, mean colonic transit time decreases, stool frequency increases, and stool consistency is softer. Decreased calorie intake slows colonic transit, whereas a meal (typically >500 kcal, especially a fat-rich meal) stimulates colonic motor function and propulsion of colonic content. Outlet obstruction in patients with pelvic floor dysfunction or voluntary suppression of defecation often is associated with slow colonic transit and decreased motor response to feeding.

Fluid reabsorption influences gastrointestinal transit.[1] Approximately 9 L of fluid enters the gut from oral intake and endogenous secretions. The small intestine delivers about 1.5 L of fluid to the colon, where most is reabsorbed, leaving a maximum of 200 mL of water excreted in normal stool. Up to 3 L of fluid can be reabsorbed by the colon in a 24-hour period, unless the rate of ileocolonic flow or colonic motility overwhelms the colon's capacity or reabsorptive ability.

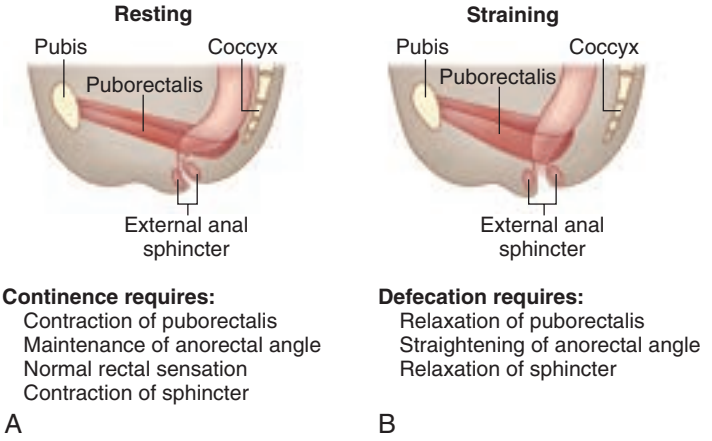

FIGURE 136-2. Pelvic floor and anorectal functions during continence and defecation. Sagittal view through the pelvis in the resting (**A**) and straining (**B**) postures. Coordinated functions of pelvic floor (puborectalis) and anal sphincter are essential for continence and defecation.

Defecation and Continence

Normal defecation requires a series of coordinated actions of the colon, rectum, pelvic floor, and anal sphincter muscles (Fig. 136-2). Filling of the rectum by a volume of 10 mL may be sensed, although the rectum can accommodate 300 mL before a sense of fullness and urge to defecate develop. Distention of the rectum results in the relaxation of the internal anal sphincter (rectoanal inhibitory reflex) and simultaneous contraction of the external anal sphincter to maintain continence. The anal transition zone can sense the difference between solid or liquid stool and gas.

DISEASES OF SLOW TRANSIT THROUGH THE STOMACH AND SMALL BOWEL

PATHOBIOLOGY

Gastrointestinal motility disturbances (Table 136-1) result from disorders of the extrinsic nervous system, enteric nervous system, interstitial cells of Cajal (or intestinal pacemakers), or smooth muscle. Neuropathic patterns are characterized by normal amplitude but incoordinated contractions, whereas myopathies are characterized by low-amplitude contractions (average of less than 40 mm Hg in the antrum and less than 10 mm Hg in the small bowel). Combined disorders occur in systemic sclerosis (Chapter 267), amyloidosis (Chapter 188), and mitochondrial cytopathy (Chapter 421), which can be manifested initially with neuropathic patterns and later display myopathic characteristics with disease progression.

TABLE 136-1 CLASSIFICATION OF GASTROPARESIS AND PSEUDO-OBSTRUCTION

TYPE	NEUROPATHIC	MYOPATHIC
Infiltrative	Progressive systemic sclerosis Amyloidosis	Progressive systemic sclerosis Amyloidosis Systemic lupus erythematosus Ehlers-Danlos syndrome Dermatomyositis
Familial	Familial visceral neuropathies	Familial visceral myopathies Metabolic myopathies
Idiopathic	Sporadic hollow visceral myopathy	Idiopathic intestinal pseudo-obstruction
Neurologic	Porphyria Heavy metal poisoning Brain stem tumor Parkinson disease Multiple sclerosis Spinal cord transection	Myotonia Other dystrophies
Infectious	Chagas disease Cytomegalovirus Norwalk virus Epstein-Barr virus	
Drug induced	Tricyclic antidepressants Narcotic agents Anticholinergic agents Antihypertensives Dopaminergic agents Vincristine Laxatives	
Paraneoplastic	Small cell lung cancer Carcinoid syndrome	
Postsurgical	Postvagotomy with or without pyloroplasty or gastric resection	
Endocrine	Diabetes mellitus Hypothyroidism or hyperthyroidism Hypoparathyroidism	

Genetic defects that result in congenital dysmotilities include abnormalities of *RET*, the gene that encodes the tyrosine kinase receptor, and abnormalities in the endothelin B system. Neural crest cells migrate from the vagal and sacral crest to the developing gut and, over time, colonize the entire developing alimentary canal and its appendages. Endothelin B serves to retard maturation of migrating neural crest cells, thus facilitating colonization of the entire gut with nerve cells. Other abnormalities resulting in congenital dysmotility involve other transcription factors, such as Sox10, which enhances maturation of neural precursors, and Kit, a marker for the interstitial cells of Cajal. Defects of *RET*, endothelin B, and Sox10 are associated with the phenotypic picture recognized in Hirschsprung disease, whereas *KIT* defects have been associated with idiopathic hypertrophic pyloric stenosis and congenital megacolon. c-*KIT* mutations are associated with gastrointestinal stromal tumors (Chapter 192).

Extrinsic Neuropathic Disorders

Extrinsic neuropathic processes include vagotomy, trauma, Parkinson disease (Chapter 409), diabetes (Chapter 229), amyloidosis (Chapter 188), and a paraneoplastic syndrome usually associated with small cell carcinoma of the lung (Chapter 191). Another common "neuropathic" problem in clinical practice results from the effect of medications, such as α_2-adrenergic agonists, glucagon-like peptide 1 analogues, opiates, and anticholinergics, on neural control.

Damage to the autonomic nerves by trauma, infection, neuropathy, and neurodegeneration may lead to motor, secretory, and sensory disturbances, most frequently resulting in constipation. Patients with spinal cord injury (Chapter 399) above the level of the sacral segments have delayed proximal and distal colonic transit attributable to parasympathetic denervation. In these patients, fasting colonic motility and tone are normal, but the response to feeding generally is reduced or absent. Spinal cord lesions involving the sacral segments and damage to the efferent nerves from these segments disrupt the neural integration of rectosigmoid expulsion and anal sphincter control. In patients with these injuries, there is loss of contractile activity in the left colon and decreased rectal tone and sensitivity, which may lead to colorectal dilation and fecal impaction. Parkinson disease (Chapter 409) and multiple sclerosis (Chapter 411) frequently are associated with constipation.

Enteric and Intrinsic Neuropathic Disorders

Disorders of the enteric nervous system or interstitial cells of Cajal are usually the result of an infectious, degenerative, immune, or inflammatory process. Virus-induced gastroparesis (e.g., rotavirus, Norwalk virus [Chapter 380], cytomegalovirus [Chapter 376], or Epstein-Barr virus [Chapter 377]) is associated with infiltration of the myenteric plexus with inflammatory cells. In idiopathic chronic intestinal pseudo-obstruction, in which there is no disturbance of the extrinsic neural control, degeneration of the interstitial cells of Cajal, inflammation, or herpesvirus infection may contribute.

Smooth Muscle Disorders

Disturbances of smooth muscle may result in significant disorders of gastric emptying and of transit through the small bowel and colon. These disturbances include, in decreasing order of prevalence, systemic sclerosis (Chapter 267), amyloidosis (Chapter 188), dermatomyositis (Chapter 269), myotonic dystrophy (Chapter 421), and metabolic muscle disorders (Chapter 421). Motility disturbances may be the result of metabolic disorders, such as hypothyroidism (Chapter 226) and hyperparathyroidism (Chapter 245), but these patients more commonly present with constipation. Scleroderma may result in focal or general dilation, wide-mouthed diverticula, and delayed transit in the stomach, small bowel, and colon. The amplitude of contractions is reduced, and bacterial overgrowth may result in steatorrhea or pneumatosis intestinalis. Mitochondrial neurogastrointestinal encephalomyopathy, or familial visceral myopathy type II, is an autosomal recessive condition that may be manifested with hepatic failure in neonates, seizures or diarrhea in infants, and hepatic failure or chronic intestinal pseudo-obstruction in adults.

Gastroparesis and Pseudo-Obstruction

CLINICAL MANIFESTATIONS

The clinical features of gastroparesis[2] and chronic intestinal pseudo-obstruction[3] are similar and include nausea, vomiting, early satiety, abdominal discomfort, distention, bloating, and anorexia. Severe cases, which occur mostly in patients with disorders of smooth muscle, may be accompanied by considerable dilation as well as by weight loss, with depletion of mineral and vitamin stores. Diarrhea and constipation indicate that the motility disorder extends beyond the stomach. Vomiting may result in aspiration pneumonia (Chapter 94) or Mallory-Weiss esophageal tears (Chapters 135 and 138), and patients with a generalized motility disorder may have abnormal swallowing or delayed colonic transit.

A careful family and medication history is essential. Review of systems may reveal an underlying collagen vascular disease (e.g., scleroderma) or disturbances of extrinsic autonomic neural control, including orthostatic dizziness, difficulties with erection or ejaculation, recurrent urinary tract infections, dry mouth, dry eyes, dry vagina, difficulties with visual accommodation in bright lights, and absence of sweating.

On physical examination, a succussion splash indicates stasis, typically in the stomach. The hands and mouth may reveal signs of Raynaud phenomenon or scleroderma (Chapter 267). Testing of pupillary responses (to light and accommodation), external ocular movements, blood pressure in the lying and standing positions, and general features of a peripheral neuropathy can identify patients with an associated neurologic disturbance (e.g., diabetic neuropathy) or with the oculogastrointestinal dystrophy that typically is found with mitochondrial cytopathies (see under the Smooth Muscle Disorders section in this chapter).

The differential diagnosis includes mechanical obstruction, functional gastrointestinal disorders, anorexia nervosa, and the rumination syndrome. The rumination syndrome is a relatively common, underdiagnosed condition that is manifested with early (0 to 30 minutes) postprandial, effortless regurgitation of undigested food after virtually every meal.

DIAGNOSIS

A motility disorder of the stomach or small bowel should be suspected whenever large volumes are aspirated from the stomach, particularly after an overnight fast, or when undigested solid food or large volumes of liquids are observed during esophagogastroduodenoscopy. The clinician should assess the acuity of symptoms and the patient's state of hydration and nutrition. The

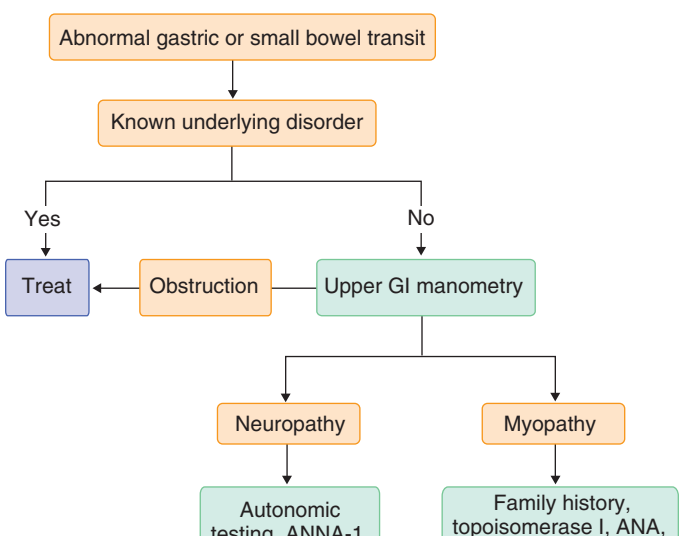

FIGURE 136-3. Flow diagram outlines steps in diagnosis of idiopathic gastroparesis and intestinal pseudo-obstruction. ANA = antinuclear antibody; ANNA-1 = type 1 antineuronal nuclear antibody; CK = creatine kinase; GI = gastrointestinal.

goals of the evaluation are to determine what regions of the digestive tract are malfunctioning and whether the symptoms are due to a neuropathy or a myopathy (Fig. 136-3). Key steps include the following:

1. *Suspect and exclude mechanical obstruction.* In symptomatic patients with pseudo-obstruction, plain radiographs of the abdomen typically show dilated loops of small bowel with associated air-fluid levels. Mechanical obstruction should be excluded by upper gastrointestinal endoscopy and small bowel imaging studies, including a small bowel follow-through, computed tomographic enterography, and magnetic resonance enteroclysis. Capsule endoscopy should be avoided because of the potential risk of retention of the capsule. Barium studies may suggest the presence of a motor disorder, particularly if there is gross dilation, dilution of barium, or retained solid food within the stomach. These studies rarely identify the cause, however, except for systemic sclerosis and mitochondrial cytopathy, which are characterized by megaduodenum, multiple small bowel diverticula, and pneumatosis intestinalis.
2. *Assess gastric and small bowel motility.*[4] After mechanical obstruction and alternative diagnoses such as Crohn disease (Chapter 141) have been excluded, a transit profile of the stomach or small bowel should be performed. The preferred test is a gastric emptying study, in which ingestion of a radiolabeled solid meal is followed by scintigraphy at 0, 1, 2, 3, 4, and 6 hours. An alternative is measurement of gastric emptying by a stable isotope breath test. If the cause of the motility disturbance is obvious, such as gastroparesis in a patient with long-standing diabetes mellitus, it is usually unnecessary to pursue further diagnostic testing. If the cause is unclear, gastroduodenal manometry by use of a multilumen tube with sensors in the distal stomach and proximal small intestine can differentiate a neuropathic process (normal-amplitude contractions but abnormal patterns of contractility) from a myopathic process (low-amplitude contractions in the affected segments). An alternative that can assess contraction amplitude is the wireless motility capsule.
3. *Identify the pathogenesis* (see Table 136-1). In patients with neuropathic causes of uncertain origin, tests should assess autonomic dysfunction (Chapter 421), measure type 1 antineuronal nuclear autoantibodies and other autoantibodies associated with paraneoplastic syndromes, and consider the possibility of a brain stem lesion. In patients with a myopathic disorder of unclear cause, the evaluation should consider amyloidosis (immunoglobulin electrophoresis, fat aspirate, or rectal biopsy; Chapter 188), systemic sclerosis (topoisomerase I; Chapter 267), and thyroid disease (Chapter 226). In appropriate settings, porphyria (Chapter 210) and Chagas disease (Chapter 347) may need to be excluded. In refractory cases, referral to a specialized center may result in genetic testing or full-thickness biopsy of the small intestine (Fig. 136-4) to identify metabolic muscle disorders and mitochondrial cytopathies.

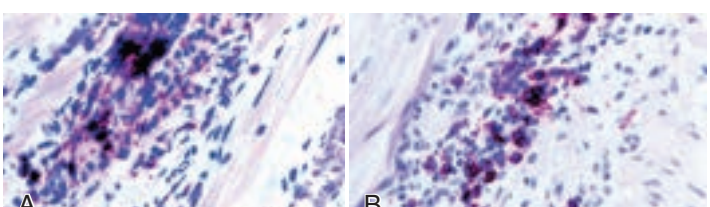

FIGURE 136-4. Micrographs showing both types of T lymphocytes, CD4 (A) and CD8 (B), detectable within the myenteric plexus of the small intestine (proximal ileum) of a 20-year-old man with chronic intestinal pseudo-obstruction. Note the intense CD4 and CD8 immunoreactivities that represent the predominant component of the immune infiltrate observed in cases of lymphocytic ganglionitis. Alkaline phosphatase anti–alkaline phosphatase immunohistochemical technique (×120) in **A** and **B**. (From De Giorgio R, Camilleri M. Human enteric neuropathies: morphology and molecular pathology. *Neurogastroenterol Motil.* 2004;16:515-531.)

Diabetes mellitus (Chapter 229) is associated with gastroparesis, pylorospasm, intestinal pseudo-obstruction, diarrhea, constipation, and fecal incontinence.[5] All of these manifestations may be caused by poor glycemic control, autonomic dysfunction, or changes in the structure and function of the interstitial cells of Cajal and enteric nervous system. Patients may become poorly nourished with vitamin deficiencies. The prevalence of constipation is 22% among diabetic patients with neuropathy but only 9.2% in diabetic patients without neuropathy, a rate that is not significantly different from that of nondiabetic controls.

4. *Identify complications of the motility disorder, including bacterial overgrowth, dehydration, and malnutrition.* In patients presenting with diarrhea, it is important to assess nutritional status and to exclude bacterial overgrowth by culture of small bowel aspirates or glucose-hydrogen breath test (Chapter 140). Bacterial overgrowth is relatively uncommon in neuropathic disorders but is found more often in myopathic conditions, such as scleroderma, that are associated more often with dilation or low-amplitude contractions. An empirical trial of antibiotics (see later) often is used instead of formal testing.

TREATMENT

Rx

Rehydration, electrolyte repletion, and nutritional supplementation are particularly important during acute exacerbations of gastroparesis and chronic intestinal pseudo-obstruction. Initial nutritional measures include low-fiber supplements with the addition of iron, folate, calcium, and vitamins D, K, and B_{12} at the usually recommended daily levels. In patients with more severe symptoms, enteral or parenteral supplementation may be required. If it is anticipated that enteral supplementation may be needed for more than 3 months, a jejunostomy feeding tube is recommended. These tubes may be placed with the aid of endoscopy. Gastrostomy tubes should be avoided in patients with gastroparesis except for venting purposes. The rumination syndrome is treated with behavioral approaches such as diaphragmatic breathing in the early postprandial period.

Medical Therapy

Medications increasingly are being used to treat neuromuscular motility disorders, but there is little evidence of effectiveness in myopathic disturbances except for the rare case of dystrophia myotonica affecting the stomach and for small bowel systemic sclerosis. Small randomized trials demonstrate symptomatic benefit of metoclopramide and domperidone over placebo in patients with gastroparesis. Metoclopramide is a dopamine antagonist with prokinetic and antiemetic properties. Antiemetic effects are due in part to its anti–5-hydroxytryptamine type 3 (HT_3) antagonist actions. The recommended duration of treatment with metoclopramide is 3 months; longer-term use may result in tremor and parkinsonian-like symptoms. It is available in tablet or elixir form or as a parenteral preparation and, typically, is taken orally 30 minutes before meals and at bedtime. Usual doses are 5 to 10 mg four times daily, but patients may experience side effects (changes in affect, anxiety) at relatively low doses (even 30 to 40 mg/day). The U.S. Food and Drug Administration recommends its use only in patients who do not respond to other treatments and for periods of less than 3 months.

Domperidone (10 to 20 mg, three times per day before meals) is another dopamine antagonist that is approved in some countries but not in the United States, where it may be available through an investigational drug application to the U.S. Food and Drug Administration and local Institutional Review Board. Its efficacy appears similar to metoclopramide, with a lower incidence of somnolence and much lower incidence of involuntary movements.

Domperidone-induced cardiac arrhythmias may occur in patients who have specific genetic polymorphisms in the cytochrome P-450 3A4 gene, which encodes the enzyme that metabolizes domperidone. Clinically, it is important to modify dose if the patient has renal or liver failure or ingests medications that inhibit cytochrome P-450 3A4 enzyme (Chapter 29). Genetic screening is not recommended in any guidelines.

Erythromycin, a macrolide antibiotic that stimulates motilin receptors at higher doses (250 to 500 mg) and cholinergic mechanisms at lower doses (40 to 80 mg), results in the dumping of solids from the stomach. It accelerates gastric emptying in gastroparesis, increases the amplitude of antral contractions, and improves antroduodenal coordination. Erythromycin is most effective when it is used intravenously (3 mg/kg every 8 hours by slow infusion) during acute exacerbations of gastroparesis. For oral erythromycin, tolerance and gastrointestinal side effects often prevent use for longer than 1 month, but sometimes liquid erythromycin can be tolerated at 40 to 80 mg three times daily before meals. Because of the limited availability of medications approved for long-term treatment, older medications such as pyridostigmine (30 to 60 mg every 6 hours) are sometimes used.

Octreotide (50 µg subcutaneously at bedtime), a cyclized analogue of somatostatin, induces small intestinal activity that mimics phase III of the interdigestive migrating motor complex. It retards gastric emptying, decreases postprandial gastric motility, and inhibits small bowel transit. Octreotide appears to be useful in the treatment of dumping syndromes associated with accelerated transit. Octreotide may be used before sleep at night to induce migrating motor complex activity, to sweep residue toward the colon, and to avoid bacterial overgrowth. If it is required during the daytime, octreotide often is combined with an oral prokinetic to "normalize" the gastric emptying rate.

Antiemetics, including diphenhydramine (25 mg orally up to two times per day for up to 3 months), promethazine (orally or by suppository, 25 mg up to two times per day), and metoclopramide (5 to 10 mg orally up to three times per day for up to 3 months), can treat nausea and vomiting in patients with gastroparesis and intestinal pseudo-obstruction. The more expensive serotonin 5-HT_3 antagonists (e.g., ondansetron) or NK_1 antagonists (e.g., aprepitant) have not proved to be of greater benefit than these less expensive alternatives in these patients. Nortriptyline is not effective in patients with idiopathic gastroparesis.[6]

Antibiotic therapy is indicated in patients with documented, symptomatic bacterial overgrowth. Although formal clinical trials have not been conducted, it is common practice to use different antibiotics for 7 to 10 days each month, in an attempt to avoid bacterial resistance. Common antibiotics include doxycycline, 100 mg twice daily; metronidazole, 500 mg three times daily; ciprofloxacin, 500 mg twice daily; double-strength trimethoprim-sulfamethoxazole, two tablets twice daily; and rifaximin, 275 mg twice daily. Use of antibiotics in patients with diarrhea and fat malabsorption secondary to bacterial overgrowth results in significant symptomatic relief.

Surgical Therapy

Surgical decompression is rarely necessary in patients with chronic pseudo-obstruction. Venting enterostomy (jejunostomy) is effective, however, in relieving abdominal distention and bloating and in reducing the frequency with which nasogastric intubations and hospitalizations are required for acute exacerbations compared with the period before vent placement. Access to the small intestine by enterostomy also provides a means to deliver nutrients and should be considered in patients with intermittent symptoms. Surgical treatment should be considered whenever the motility disorder is localized to a resectable portion of the gut: completion gastrectomy for patients with post–gastric surgical stasis syndrome, and colectomy with ileorectostomy for intractable constipation associated with chronic colonic pseudo-obstruction.

Gastric electrical stimulation, a treatment approved for humanitarian use, may improve symptoms in some patients with severe gastroparesis, but data on its efficacy are inconclusive, and its use is restricted to a few centers in the United States. Small bowel transplantation currently is limited to patients with intestinal failure who have reversible liver disease induced by total parenteral nutrition or who have life-threatening or recurrent catheter-related sepsis. Isolated transplantation of the small intestine is associated with higher graft and patient survival and fewer complications related to rejection and infection.

DISEASES OF RAPID TRANSIT THROUGH STOMACH AND SMALL BOWEL

Dumping Syndrome and Accelerated Gastric Emptying

Dumping syndrome and accelerated gastric emptying typically follow truncal vagotomy and gastric drainage procedures (Chapter 139) or fundoplication for gastric esophageal reflux disease (Chapter 138). With the widespread use of highly selective vagotomy and the advent of effective antacid secretory therapy, these problems are becoming rare. A high calorie (usually carbohydrate) content of the liquid phase of the meal evokes a rapid insulin response with secondary hypoglycemia. These patients may also have impaired antral contractility and gastric stasis of solids, which paradoxically may result in a clinical picture of gastroparesis (for solids) and dumping (for liquids).

The management of dumping syndrome and accelerated gastric emptying emphasizes dietary maneuvers, such as avoidance of high-nutrient liquid drinks and possibly addition of guar gum or pectin to retard gastric emptying of liquids. Rarely, pharmacologic treatment with octreotide, 25 to 100 µg subcutaneously before meals, is needed to retard intestinal transit and to inhibit the hormonal responses that lead to hypoglycemia.

Rapid Transit Dysmotility of the Small Bowel

Rapid transit of material through the small bowel may occur in the setting of the irritable bowel syndrome–diarrhea predominant subtype (Chapter 137), postvagotomy diarrhea (Chapter 140), short bowel syndrome (Chapter 140), diabetic diarrhea (Chapter 140), and carcinoid diarrhea (Chapter 232). With the exception of irritable bowel syndrome, these conditions may cause severe diarrhea and result in significant losses of fluid and electrolytes. Ileal resection or disease and idiopathic bile acid malabsorption may represent an inability of the distal ileum to reabsorb bile acids because of rapid transit and reduced contact time with the ileal mucosa; this condition may induce colonic secretion and secondary diarrhea. Accelerated transit may be confirmed by scintigraphic studies.

Treatment goals are to restore hydration and nutrition and to slow small bowel transit. Dietary interventions include avoiding hyperosmolar drinks and replacing them with iso-osmolar or hypo-osmolar oral rehydration solutions. The fat content in the diet should be reduced to approximately 50 g/day to avoid delivery of unabsorbed fat to the colon. All electrolyte and nutritional deficiencies of calcium, magnesium, potassium, and water-soluble and fat-soluble vitamins should be corrected. In patients with less than 1 m of residual small bowel, it may be impossible to maintain fluid and electrolyte homeostasis without parenteral support. In patients with a longer residual segment, oral nutrition, pharmacotherapy, and supplements are almost always effective.

The opioid agent loperamide (4 mg 30 minutes before meals and at bedtime for a total dose of 16 mg/day) suppresses the motor response to feeding and improves symptoms but may be ineffective or cause side effects (e.g., hypotension). Bile acid binding, such as with cholestyramine (4 g three times daily) or colesevelam (1.875 g twice daily), is indicated for patients with suspected or proven bile acid malabsorption. Verapamil (40 mg twice daily), clonidine (0.1 mg twice daily), or a 5-HT_3 antagonist (e.g., ondansetron, 4 to 8 mg three times daily) is used rarely in addition to loperamide. Octreotide (50 µg subcutaneously three times daily before meals) may be used in patients for whom the oral agents are ineffective or poorly tolerated. 5-HT_3 antagonists (e.g., alosetron, 0.5 to 1 mg orally up to two times per day) may be efficacious in the treatment of carcinoid diarrhea and diarrhea-predominant irritable bowel syndrome, but it should be reserved for patients with severe, unresponsive diarrhea, and the dose should be titrated to avoid constipation.

COLONIC MOTILITY DISORDERS

Constipation

EPIDEMIOLOGY

Constipation is a common clinical problem, reported by about 20% of the population, and 40% of Americans report needing to strain excessively to pass their bowel movements.[7]

PATHOBIOLOGY

In functional constipation, transit is normal, and there is no evacuation disorder. These patients may have pain in association with constipation, and there is overlap with constipation-predominant irritable bowel syndrome (Chapter 137). In patients with acquired slow-transit constipation, unassociated with colonic dilation, the number of interstitial cells of Cajal in the different layers of the colon is reduced compared with controls.

Idiopathic megarectum and megacolon can be either congenital or acquired; an enteric nervous system defect is suspected. In megacolon, the dilated segment shows normal phasic contractility but decreased colonic tone, with smooth muscle hypertrophy and fibrosis of the muscularis mucosa, circular muscle, and longitudinal muscle layers.

Acquired defects in the enteric nervous system may result in constipation in Chagas disease (Chapter 347), which is caused by infection with

TABLE 136-2 CLINICAL CLUES SUGGESTIVE OF AN EVACUATION DISORDER

HISTORY
Prolonged straining to expel stool
Taking up unusual postures on the toilet to facilitate stool expulsion
Support of perineum or digitation of rectum or vagina to facilitate rectal emptying
Inability to expel enema fluid
Constipation after subtotal colectomy for constipation
RECTAL EXAMINATION (WITH PATIENT IN LEFT LATERAL POSITION)
Inspection
Anus "pulled" forward during attempts to simulate strain during defecation
Anal verge descends <1 cm or >4 cm during attempts to simulate strain during defecation
Perineum balloons down during straining, and rectal mucosa prolapses through anus
Palpation
High anal sphincter tone at rest precludes easy entry of examining finger (in absence of painful perianal condition, e.g., anal fissure)
Anal sphincter pressure during voluntary squeeze is minimally higher than tone at rest
Perineum descends <1 cm or >4 cm during attempts to simulate strain during defecation
Puborectalis muscle palpable through posterior rectal wall is tender
Palpable mucosal prolapse during straining
"Defect" in anterior wall of the rectum, suggestive of rectocele
ANORECTAL MANOMETRY AND BALLOON EXPULSION (WITH PATIENT IN LEFT LATERAL POSITION)
Average anal sphincter resting tone >80 cm H_2O or squeeze pressure >240 cm H_2O
Failure of balloon expulsion despite addition of 200 g weight

Trypanosoma cruzi and results in the destruction of myenteric neurons. Acquired aganglionosis also has been reported with circulating antineuronal antibodies, with or without associated neoplasm.

DIAGNOSIS

It is essential to distinguish an evacuation disorder, also called functional outlet obstruction (Table 136-2), from constipation resulting from slow transit or other causes. In a tertiary center, about 25% of 1411 patients presenting with constipation had impaired evacuation, and the remainder had constipation associated with either normal transit (also called functional constipation) or delayed colonic transit (also called slow-transit constipation). Characterization of constipated patients (Fig. 136-5) relies on the measurement of transit with radiopaque markers or scintigraphy.

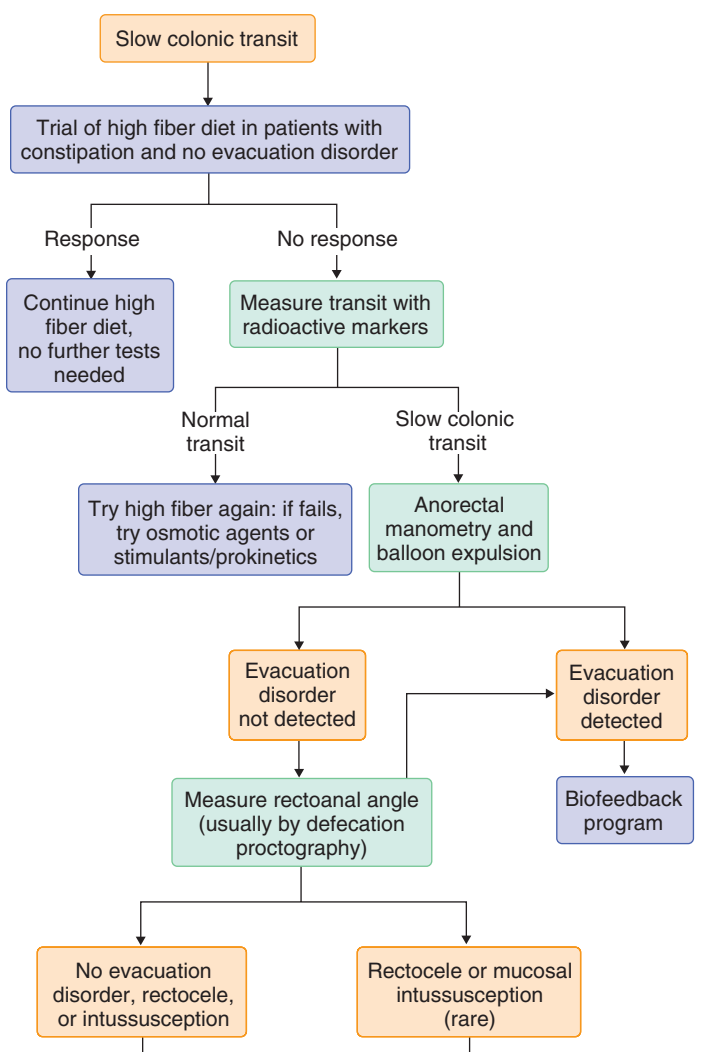

FIGURE 136-5. Algorithm in the management of constipation.

TREATMENT

The average daily fiber intake is around 12 g/day. In patients with normal-transit constipation, 12 to 30 g/day is effective in relief of constipation. In patients with slow-transit constipation, drug-induced constipation, or evacuation disorders, however, supplementation of 30 g of fiber per day does not result in any improvement in constipation. A second step is to add an osmotic laxative, such as a magnesium salt or polyethylene glycol solution, to enhance the retention of fluid within the lumen by osmotic forces, to increase the fluidity, and to ease aboral transport of colonic content. Polyethylene glycol solutions (such as GoLYTELY, NuLYTELY, MiraLAX, OCL solution) are used frequently as a first-line therapy.[■] If these measures do not suffice, a prokinetic or stimulant agent, such as bisacodyl (5 to 10 mg every 1 to 2 days) may be added.

Newer medications that accelerate colonic transit include prucalopride (2 mg daily), which is beneficial in chronic constipation[■■]; lubiprostone (24 µg twice daily), a chloride channel activator; and linaclotide (145 µg or 290 µg daily), a guanylate cyclase C agonist that induces chloride and fluid secretion.[■] These drugs[■] as well as naloxegol (a peripheral µ-opioid antagonist at 12.5 or 25 mg daily)[■] and methylnaltrexone (0.15 mg/kg subcutaneously every other day for 2 weeks) are effective in patients with opiate-induced constipation as a complication of advanced illness.[■]

When these approaches do not work, the patient should be reassessed to exclude an evacuation disorder. For evacuation disorders, a biofeedback treatment program with muscle relaxation of the anal sphincters and pelvic floor results in a 70% or greater cure rate for the constipation. The response to this treatment program is influenced by comorbidity, such as the coexistence of eating disorders or a psychological or psychiatric diagnosis.

Surgical Therapy

In patients whose constipation is not associated with an evacuation disorder and does not respond to aggressive medical therapies (including combinations described earlier), subtotal colectomy with ileorectostomy is effective in relieving constipation. Laparoscopic colectomy with ileorectostomy achieves the same success rate with less morbidity compared with open colectomy with ileorectostomy.

Hirschsprung Disease

DEFINITION

Hirschsprung disease occurs in 1 in 5000 live births. It is characterized by a localized segment of narrowing of the distal colon as a result of failure of local development of intrinsic nerves in the myenteric plexus.[8]

PATHOBIOLOGY

A relative deficiency of *KIT*-positive interstitial cells of Cajal has been reported in Hirschsprung disease and chronic intestinal pseudo-obstruction. The majority of familial and sporadic Hirschsprung disease is associated with *RET* oncogene mutations, but mutations in *EDNRB*, *GDNF*, and *EDN3* have been reported as well. Hirschsprung disease is well characterized histologically by the absence of ganglion cells in the myenteric and submucosal plexus and the presence of hypertrophied nerve trunks in the space normally occupied by the ganglion cells. In the muscle layers of the colon involved with Hirschsprung disease, deficiency of nerve growth factor receptors also

contributes to the dysfunction. The narrowing and failure of relaxation in the aganglionic segment are thought to be due to the lack of neurons containing nitric oxide synthase.

CLINICAL MANIFESTATIONS

Hirschsprung disease is usually diagnosed at birth because of failure to pass meconium or the presence of megacolon, although it may be identified in childhood as a result of fecal retention, constipation, or abdominal distention. Onset of symptoms or diagnosis after the age of 10 years is rare.

DIAGNOSIS

Diagnosis is based on the typical focal narrowing of the colon, the absence of the rectoanal inhibitory reflex (relaxation of anal sphincter pressure at rest during distention of a balloon in the rectum depends on natural preservation and maturation of intrinsic nerves in the distal bowel), and a deep rectal biopsy specimen showing absence of submucosal neurons with hypertrophied nerve trunks.

TREATMENT

Rx

Treatment involves excision of the affected bowel segment or a pull-through procedure by which normal bowel is anastomosed to the cuff of the rectum, just above the anal sphincters.

Grade A References

A1. Dipalma JA, Cleveland MV, McGowan J, et al. A randomized multicenter, placebo-controlled trial of polyethylene glycol laxative for chronic treatment of chronic constipation. *Am J Gastroenterol*. 2007;102:1436-1441.

A2. Camilleri M, Kerstens R, Rykx A, et al. A placebo-controlled trial of prucalopride for severe chronic constipation. *N Engl J Med*. 2008;358:2344-2354.

A3. Shin A, Camilleri M, Kolar G, et al. Systematic review with meta-analysis: highly selective 5-HT_4 agonists (prucalopride, velusetrag or naronapride) in chronic constipation. *Aliment Pharmacol Ther*. 2014;39:239-253.

A4. Lembo AJ, Schneier HA, Shiff SJ, et al. Two randomized trials of linaclotide for chronic constipation. *N Engl J Med*. 2011;365:527-536.

A5. Ford AC, Brenner DM, Schoenfeld PS. Efficacy of pharmacological therapies for the treatment of opioid-induced constipation: systematic review and meta-analysis. *Am J Gastroenterol*. 2013;108:1566-1574.

A6. Chey WD, Webster L, Sostek M, et al. Naloxegol for opioid-induced constipation in patients with noncancer pain. *N Engl J Med*. 2014;370:2387-2396.

A7. Thomas J, Karver S, Cooney GA, et al. Methylnaltrexone for opioid-induced constipation in advanced illness. *N Engl J Med*. 2008;358:2332-2343.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bharucha AE, Rao SS. An update on anorectal disorders for gastroenterologists. *Gastroenterology*. 2014;146:37-45.
2. Camilleri M, Parkman HP, Shafi MA, et al. Clinical guideline: management of gastroparesis. *Am J Gastroenterol*. 2013;108:18-37.
3. Gabbard SL, Lacy BE. Chronic intestinal pseudo-obstruction. *Nutr Clin Pract*. 2013;28:307-316.
4. Bredenoord AJ, Smout AJ. Advances in motility testing—current and novel approaches. *Nat Rev Gastroenterol Hepatol*. 2013;10:463-472.
5. Shin AS, Camilleri M. Diagnostic assessment of diabetic gastroparesis. *Diabetes*. 2013;62:2667-2673.
6. Parkman HP, Van Natta ML, Abell TL, et al. Effect of nortriptyline on symptoms of idiopathic gastroparesis: the NORIG randomized clinical trial. *JAMA*. 2013;310:2640-2649.
7. Bharucha AE, Pemberton JH, Locke GR 3rd. American Gastroenterological Association technical review on constipation. *Gastroenterology*. 2013;144:218-238.
8. Burkardt DD, Graham JM Jr, Short SS, et al. Advances in Hirschsprung disease genetics and treatment strategies: an update for the primary care pediatrician. *Clin Pediatr (Phila)*. 2014;53:71-81.

137

FUNCTIONAL GASTROINTESTINAL DISORDERS: IRRITABLE BOWEL SYNDROME, DYSPEPSIA, CHEST PAIN OF PRESUMED ESOPHAGEAL ORIGIN, AND HEARTBURN

EMERAN A. MAYER

DEFINITIONS

Irritable bowel syndrome (IBS), functional dyspepsia, functional chest pain of presumed esophageal origin, and functional heartburn are characterized by chronic, recurrent symptoms of pain and discomfort referred to the lower abdomen, the epigastrium and upper abdomen, and the retrosternum, respectively. They belong to the family of functional gastrointestinal (GI) disorders that comprise a wide spectrum of chronic GI disorders common in both the adult and pediatric populations. In the absence of disease-specific biomarkers, each syndrome is classified by symptoms and the absence of other conditions that can account for the symptoms. Despite the benign prognoses, functional GI diseases can affect health-related quality of life at least as much as organic diseases. Because effective pharmacologic therapies are limited, disease management may include cognitive behavioral approaches and alternative medicine approaches (Chapter 39).

PATHOBIOLOGY

The pathophysiology of functional GI diseases remains incompletely understood, but these diseases are characterized by alterations in bidirectional interactions between the brain and the gut (brain-gut axis; E-Fig. 137-1), with variable contributions of peripheral (gut microbiota, mucosal immune activation, motility, bile acids) and central factors (enhanced perception of visceral signals by the central nervous system). In addition, altered signaling from the nervous system to the GI tract through the autonomic nervous system can modulate esophagogastrointestinal function. Each diagnostic category of functional GI disease is defined by symptomatic criteria that include different subsets of patients who exhibit different patterns of the brain-gut axis dysregulation and that result in varying abnormalities in GI motility, secretion, immune function, or visceral sensitivity. Despite this heterogeneity, however, functional GI diseases all share certain features, including a greater prevalence in women, enhanced sensitivity to stress, enhanced perception of visceral signals, the frequent coexistence of psychiatric and chronic pain disorders, and the response to centrally targeted pharmacologic and nonpharmacologic therapies.

Enhanced Perception of Visceral Pain

About 30 to 70% of patients with functional GI diseases have an altered perception of visceral afferent stimuli ("visceral hypersensitivity"), in which normally innocuous stimuli, such as physiologic contractions, distentions by food or gas, or chemical stimulation of the intestine (bile salts, microbial signaling molecules), stomach, or esophagus (hydrochloric acid, bile acids) lead to the sensation of pain or discomfort. The stimulus may be spontaneous peristaltic activity or result from distention by luminal contents, such as ingested food, liquids, gas, or feces. Visceral hypersensitivity may be associated with aberrant referral of visceral sensations to a particular body area, and this referral is often atypical in location and larger compared with most individuals. Many patients also have enhanced perception of somatic pain owing to alterations in sensory processing and modulation systems.

Altered Stress Responsiveness

Abnormal autonomic and neuroendocrine responses to psychosocial stressors are a key feature of functional GI disease and may play an important role in both its cause and its exacerbation. For example, stressful events are more likely to lead to abdominal pain and a change in stool pattern in patients with IBS compared with healthy controls, and stress has been correlated with bowel symptoms and physician visit. Patients with functional GI disease report more lifetime stressful events than healthy controls, including a higher frequency and severity of early adverse life events.

IRRITABLE BOWEL SYNDROME

DEFINITION

IBS is defined as chronic, recurring abdominal pain or discomfort that is associated with defecation or a change in bowel habits; the diagnosis also requires the absence of detectable organic disease that may explain the symptoms[1] (Table 137-1). Given the high prevalence of IBS in the general population, comorbid organic GI syndromes may coexist with IBS, including ulcerative colitis (Chapter 141), microscopic and collagenous colitis (Chapter 140), and celiac disease (Chapter 140). As a result of such comorbidity, some colitis patients report symptom severity disproportionate to the degree of mucosal inflammation and may continue to have GI symptoms after their inflammatory bowel disease or other types of colitis have been successfully treated. For example, a small subset of ulcerative colitis patients exhibit IBS-like symptoms and have demonstrable changes in colonic motility when they are in clinical remission.

IBS is further subdivided into IBS with diarrhea, IBS with constipation, or IBS with mixed bowel habits depending on the predominant bowel habit. Another subset is postinfectious IBS, which includes patients who develop

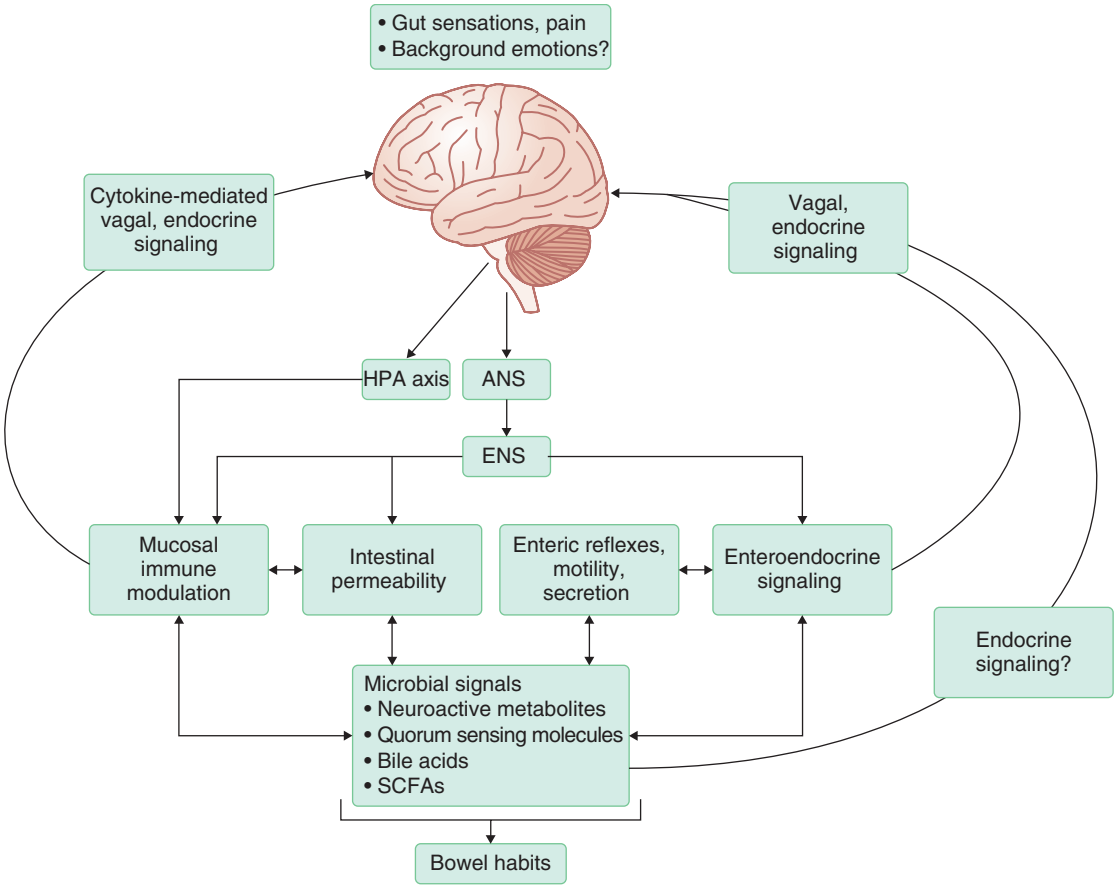

E-FIGURE 137-1. Brain-gut microbiome axis. Sensory and immune-related signals from the gastrointestinal (GI) tract, including signals from the gut microbiome, are transmitted by viscerosensory (vagal and spinal) afferents, as well as by various blood-borne (endocrine) signaling molecules to the brain (*left* and *right ascending arrows*). In the healthy person, most of these signals are not consciously perceived but nevertheless provide input to various regulatory feedback loops from the brain to the gut by the autonomic nervous system (ANS) and the hypothalamic-pituitary-adrenal (HPA) axis. This feedback reaches specific target cells in the gut (immune, epithelial, enteroendocrine, smooth muscle, neuronal cells) to modulate various gut functions (including mucosal immune activation, intestinal permeability, peristalsis, secretion, and neurotransmitter release). Many of these gut functions have a bidirectional relationship with the gut microbiome, which itself is in bidirectional communication with the gut epithelium through multiple signaling mechanisms. Although the integration of these interactions is the major determinant of bowel habits, the ascending signals to the brain also play a role in the generation of abdominal pain and discomfort as well as in background emotions. ENS = enteric nervous system; SCFAs = short-chain fatty acids. (From Mayer EA, Savidge T, Shulman RJ. Brain-gut microbiome interactions and functional bowel disorders. *Gastroenterology.* 2014;146:1500-1512.)

TABLE 137-1 ROME III DIAGNOSTIC CRITERIA FOR IRRITABLE BOWEL SYNDROME

1. Recurrent abdominal pain or discomfort at least 3 days per month in the last 3 months (but with symptom onset for at least 6 months) associated with two or more of the following:
   - Improvement with defecation
   - Onset associated with a change in frequency of stool
   - Onset associated with a change in form (appearance) of stool

   Symptoms that cumulatively support the diagnosis of irritable bowel syndrome
   - Abnormal stool frequency: ≤3 bowel movements per week or >3 bowel movements per day
   - Abnormal stool form: lumpy/hard stool or loose/watery stool
   - Defecation straining
   - Urgency
   - Feeling of incomplete bowel movement
   - Passing mucus
   - Bloating or feeling of abdominal distention
2. Absence of alarm symptoms:
   - Weight loss
   - Bloody stool
   - Anemia
   - Family history of inflammatory bowel disease
   - Colon cancer
   - Celiac disease

persistent IBS-like symptoms despite the resolution of an episode or episodes of bacterial gastroenteritis. When patients do not fulfill criteria for these subtypes, their disease is classified as unspecified IBS.

EPIDEMIOLOGY

IBS is a common disorder, with a worldwide prevalence that ranges between 5 and 15%. As is the case with many related functional pain disorders, it is more common in women, with a female-to-male ratio of 2 : 1 to 2.5 : 1. It is generally assumed that IBS most commonly presents between the ages of 30 and 50 years, but IBS is also common in the pediatric population, in which such symptoms have traditionally been referred to as recurrent abdominal pain.

The socioeconomic burden of IBS is substantial, with patients taking three times as many days of sick leave compared with individuals without IBS; about 8% of patients retire early because of their symptoms. IBS is estimated to account for about 12% of primary care visits and 19% of GI specialty visits. In the United States, IBS is estimated to account for $1.6 billion in direct costs and $19.2 billion in indirect costs annually.

PATHOBIOLOGY

Although IBS is the most common and best studied functional GI disease, there is no general agreement on its pathophysiology. Both peripheral and central alterations are implicated in the brain-gut axis that best explains the clinical manifestations of IBS.[2]

Altered Gastrointestinal Motility and Secretion

GI motility is quantitatively but not qualitatively different in 25 to 75% of IBS patients compared with healthy controls, but these measured differences are not sufficiently reliable to be used as diagnostic markers. Colonic transit is accelerated in about 45% of IBS patients with diarrhea, and high-amplitude propagating contractions occur more frequently in such patients than in controls. These high-amplitude propagating contractions correlate with crampy abdominal pain and may be the mechanism underlying urgency, diarrhea, and associated fecal incontinence in this patient subgroup. By contrast, slowed colonic transit is observed in about 25% of IBS patients with constipation. Exaggerated or prolonged colonic motility responses to food intake (gastrocolonic response) may be present in approximately 30% of patients who report an exacerbation of abdominal pain after food intake. Likely contributors to alterations in bowel habits include abnormalities in intestinal water and electrolyte secretion and absorption as well as altered synthesis and secretion of bile acids.

Acute stress-induced activation of contractions and secretions of the hindgut is mediated by sacral parasympathetic pathways, and IBS patients have increased activation of these pathways in response to severe laboratory stressors. Tonic upregulation of sympathetic and sacral parasympathetic activity may result in neuroplastic changes in peripheral target mechanisms within the gut, including the enteric nervous system.

Evidence supports increased intestinal permeability predominantly in IBS patients with diarrhea. Metabolites from gut microbiota, GI infections, and chronic psychological stress can decrease epithelial barrier function in susceptible individuals.

In response to luminal signals such as bile acids, moving intestinal contents, and microbial products, serotonin is released from enterochromaffin cells on the basolateral side of the intestinal epithelium, where it stimulates vagal afferent nerves and enteric neurons involved in secretion and motility. An increased release of serotonin from these cells following a meal may contribute to increased motility and secretion in IBS patients with diarrhea.

Data indicate decreased diversity in small bowel microbiota of patients with IBS, with an increased abundance of gram-negative organisms and a decrease of *Bifidobacterium* and *Lactobacillus*, as well as increased ratios of Firmicutes to Bacteroidetes.[3,4] Metabolites produced by microbiota, including short-chain fatty acids and neurotransmitter substances, have been implicated in mediating the effects of dysbiosis on the gut. Small bowel bacterial overgrowth may contribute to IBS symptoms in some patients, especially those with predominant bloating-type symptoms, in whom treatment with nonabsorbable antibiotics may reduce symptoms.

Altered synthesis and secretion of bile acids may result in increased fluid secretion and motility patterns. Certain foods that may trigger symptoms include milk, wheat products, and fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAP).

Genetics

Familial aggregation has been demonstrated in several studies of IBS patients, and the genetic heritability of functional GI diseases has been estimated to be between 22 and 57%. Candidate gene association studies suggest possible association of IBS with polymorphisms in genes related to signaling systems within the brain-gut axis, including bile acid physiology, serotonin, noradrenaline, corticotropin-releasing factor, and voltage-gated sodium channels.[5] Epigenetic factors, including the influence of certain types of stress, have been identified as an important influence on the risk for developing IBS.

CLINICAL MANIFESTATIONS

Symptoms

The location of abdominal symptoms in IBS is highly variable, but pain is most typically referred to the lower abdomen. Pain and discomfort, which occur mainly while the patient is awake, are frequently aggravated by emotion or stress, poor sleep, and intake of food, but these aggravating factors cannot be elicited in all patients. By definition, abdominal symptoms are relieved by defecation, but this relief may be temporary. IBS symptoms vary widely, from mild to very severe. Most patients seen by primary care physicians have mild symptoms, whereas most patients seen by gastroenterologists have moderate to severe symptoms.

Psychological symptoms and psychiatric diagnoses (Chapter 397), such as anxiety disorders (e.g., generalized anxiety disorder, panic disorder, and posttraumatic stress syndrome), depression, somatization, hypochondriasis, and phobias, are more common in patients with IBS, even in mildly symptomatic patients. The prevalence of coexisting psychiatric disorders can be as high as 40% to more than 90% in patients seen in tertiary referral centers but is lower in patients seen in primary care practices. Even in the absence of demonstrable psychiatric comorbidity, psychosocial stressors play an important role in exacerbating IBS symptoms in patients seen in all settings.

Patients with IBS have a three-fold or higher prevalence of fibromyalgia (Chapter 274) and migraine headaches (Chapter 398) compared with patients without IBS. IBS also frequently coexists with chronic fatigue syndrome, and 60% or more of patients with chronic fatigue syndrome have IBS symptoms. Interstitial cystitis, chronic prostatitis (Chapter 129), chronic pelvic pain, and temporomandibular disorders are also common in IBS patients.

These associations of IBS with comorbid conditions and other symptoms are more common in female patients, but all patients with IBS on average see primary care physicians for non-GI complaints three times more frequently than do healthy subjects. IBS patients also often complain of extraintestinal symptoms such as dyspareunia, fatigue, loss of energy, impotence, urinary frequency, backache, and dysmenorrhea.

DIAGNOSIS

The Rome III criteria (see Table 137-1) can establish the diagnosis of IBS without additional extensive testing. Alternative diagnoses should be considered if symptoms awaken patients from sleep, are unrelated to defecation or

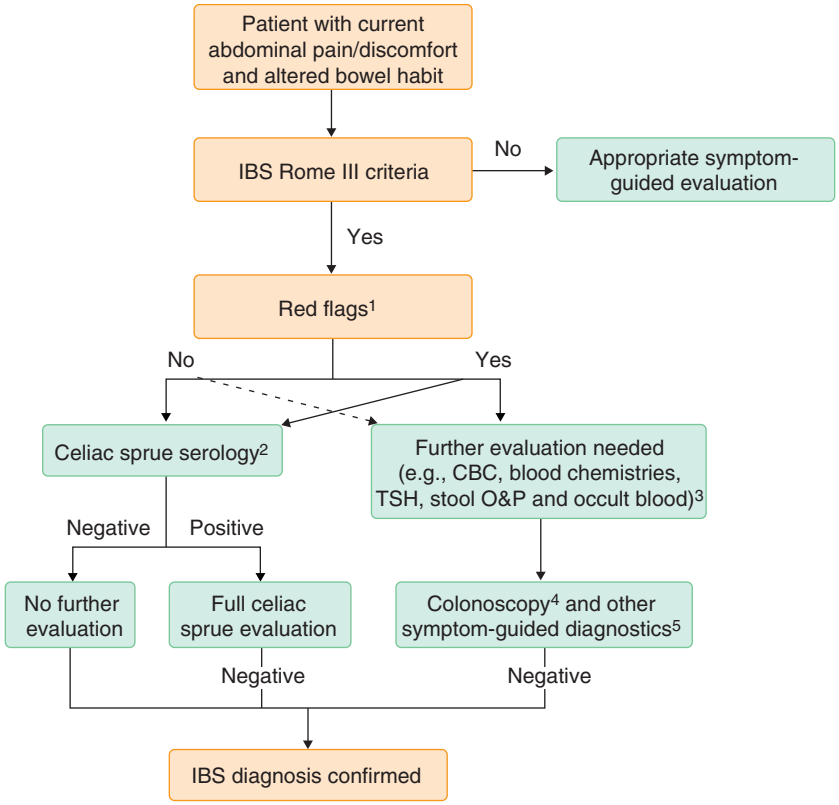

FIGURE 137-1. Diagnostic algorithm for irritable bowel syndrome (IBS). (1) Red flags include rectal bleeding, anemia, weight loss, fever, family history of colon cancer, first symptom onset after age 50 years, or major symptom change. (2) Testing for celiac sprue may be useful in patients meeting Rome criteria, in particular in irritable bowel syndrome with diarrhea, if there are red flags, and in populations in which the background prevalence of celiac sprue is high. (3) In the absence of red flags, basic complete blood count (CBC), serum biochemistry, stool testing for occult blood and ova and parasites (O&P), and thyroid-stimulating hormone (TSH) levels are only indicated in cases of a supportive clinical history. (4) Colonoscopy is only recommended in the patient with positive red flags. However, according to colon cancer screening guidelines, routine colonoscopy should be performed in IBS patients at age 50 years or older regardless of IBS symptoms. (5) If there has been a major qualitative change in the pattern of chronic symptoms, a new comorbid condition producing these symptoms should be suspected, and a more comprehensive diagnostic approach is warranted. (From Mayer EA. Clinical perspectives: irritable bowel syndrome. *N Engl J Med*. 2008;358:1692-1699.)

food intake, are constant, or are not relieved by any physiologic intervention. Weight loss is uncommon in IBS except in patients with depression (Chapter 397) or eating disorders (Chapter 219), and its presence mandates investigation of an underlying organic cause. Other alarm features raising concern for serious alternative diagnosis include bloody stools, anemia, and a family history of inflammatory bowel disease, colon cancer, or celiac disease.

Careful review of medications and dietary supplements may reveal an etiologic agent because many prescribed medications can cause constipation and some diet supplements can lead to the inadvertent ingestion of laxatives. Specific dietary agents rarely cause IBS symptoms, but lactose malabsorption can be assessed by history or by a short trial of a lactose-free diet (Chapter 140), and the diagnosis of celiac disease (Chapter 140) should be excluded. A brief psychosocial assessment is useful in identifying risk factors for chronic pain, such as traumatic early life events or somatization, as well as potential symptom triggers or exacerbating factors, such as anxiety, depression, and significant life stress.

The physical examination in IBS is usually normal but may reveal abdominal tenderness, especially in the left lower quadrant, or a tender, palpable sigmoid colon. In patients with constipation-predominant IBS and associated defecatory dysfunction, rectal examination may reveal paradoxical contraction of the puborectalis muscle or decreased descent of the pelvic floor when simulating a bowel movement.

Diagnostic Testing

In an otherwise healthy person who is younger than 50 years and who fulfills the Rome III criteria, further diagnostic testing should be minimal and guided by the individual presentation (Fig. 137-1). It remains controversial when to order a complete blood count, sedimentation rate, or thyroid studies, but most agree that it is not cost-effective to obtain these tests in all patients with presumed IBS. Serologic screening for celiac disease (Chapter 140) should be pursued in patients with diarrhea-predominant or mixed bowel habits, in patients with a family history of celiac disease, and in patients who have identified gluten-containing foods as a trigger of their symptoms. Colonic biopsies for collagenous or microscopic colitis (Chapter 140) should be considered in the setting of severe persistent diarrhea. The onset of symptoms after age 50 years and the presence of rectal bleeding or unexpected weight loss are indications for colonoscopy to assess for colorectal cancer (Chapter 193). Lower abdominal discomfort occurring with menses or with associated weight loss should trigger gynecologic evaluation. If the duration of diarrhea symptoms is short, it may be useful to assess for giardiasis (Chapter 351) or *Campylobacter* species infection (Chapter 303), but routine stool testing of all patients with diarrhea is not indicated.

Differential Diagnosis

Differential diagnosis for IBS with diarrhea predominance or mixed bowel habit includes inflammatory bowel diseases (e.g., Crohn disease, ulcerative colitis, collagenous or microscopic colitis) (Chapters 140 and 141), celiac disease (Chapter 140), infection, small bowel bacterial overgrowth (Chapter 140), malabsorptive syndromes (including lactose intolerance) (Chapter 140), bile salt malabsorption (Chapter 140), and pancreatic insufficiency (Chapter 144) or malignancy (including neuroendocrine tumors [Chapter 195] and colorectal adenocarcinoma [Chapter 193]). The differential diagnosis for constipation-predominant IBS includes colonic inertia, GI manifestations of Parkinson disease, pseudo-obstruction, paradoxical pelvic floor contraction, or pelvic outlet obstruction from structural causes such as rectal prolapse, rectocele, or short-segment Hirschsprung disease (Chapter 136).

TREATMENT

General Principles

Symptomatic treatment attempts to normalize bowel habits and decrease abdominal pain, in part by providing the patient with a plausible biologic explanation for their symptoms as well as reassurance that the symptoms are real and the prognosis is benign. Dietary intervention can be helpful, and

TABLE 137-2 MEDICATIONS USED IN THE TREATMENT OF IRRITABLE BOWEL SYNDROME*

SYMPTOMS AND MEDICATION	INITIAL DOSE (mg/day)†	TARGET DOSE (mg/day)†	COMMON OR SERIOUS SIDE EFFECTS	DEGREE OF EVIDENCE		FDA APPROVED	
				of the Symptom	*of IBS*	*for the Symptom*	*for IBS*
CONSTIPATION							
Laxatives‡ and Secretory Stimulators							
Polyethylene glycol 3350 (MiraLAX)	17,000	70,000	Diarrhea, bloating, cramping	+++	–		
Lactulose (Kristalose)	10,000-20,000	20,000-40,000	Diarrhea, bloating, cramping	+++	–		
Lubiprostone (Amitiza)		24, twice a day	Nausea, diarrhea, headache, abdominal pain and discomfort	+++	–	Yes§	No
Linaclotide (Linzess)		145,000 (IBS-C) 290,000 (Constip)	Diarrhea, abdominal pain, flatulence	+++	+++	Yes	Yes
DIARRHEA							
Loperamide (Imodium)	2	2-8	Constipation	+++	–	Yes	No
Alosetron (Lotronex)			Constipation, ischemic colitis (rare)	–	+++	No	Yes¶
BLOATING							
Antibiotics							
Rifaximin		400, three times a day	Abdominal pain, diarrhea, bad taste	–	+	No	No
Probiotics**							
Bifidobacterium infantis 35624		1 capsule per day	None	+	+	No	No
Activia *Bifidus regularis*		1 pod, twice a day	None	+		N/A	N/A
PAIN							
Tricyclic Antidepressants††			Dry mouth, dizziness, weight gain				
Amitriptyline (Elavil)	10, at bedtime	10-75, at bedtime		++	+	No	No
Desipramine (Norpramin)	10, at bedtime	10-75, at bedtime		++	+	No	No
Selective Serotonin Reuptake Inhibitors‡‡			Sexual dysfunction, headache, nausea, sedation, insomnia, sweating, withdrawal symptoms				
Paroxetine (Paxil CR)		10-60		–	+	No	No
Citalopram (Lexapro)		5-20		+	+	No	No
Fluoxetine (Prozac)		20-40	Somnolence, dizziness, headaches, insomnia	+	–	No	No
Nonselective Reuptake Inhibitors							
Duloxetine (Cymbalta)	30	60	Nausea, dry mouth, headache, dizziness	++	-	Yes	No

*This list is not exhaustive but includes major medications for which there is evidence from well-designed clinical trials of effectiveness for global irritable bowel syndrome (IBS) symptoms or for individual symptoms (e.g., constipation, diarrhea, or abdominal pain and discomfort). In the column about evidence, + denotes some evidence from at least one controlled trial; ++, moderate evidence from several controlled trials or from meta-analysis of such trials; +++, strong evidence from well-designed, controlled clinical trials; and –, no evidence. FDA = U.S. Food and Drug Administration.
†Dosages are in milligrams per day unless otherwise noted.
‡A wide range of osmotic and irritant laxatives, including fiber products, are available over the counter.
§Lubiprostone is FDA approved for the treatment of chronic constipation.
Linaclotide is FDA approved for the treatment of chronic constipation (145 mcg/day) and IBS with constipation (290 mcg/day).
¶Lotronex use is restricted for women with severe diarrhea-predominant IBS, unresponsive to other medications, because of side effects.
**Many probiotics are available over the counter and are not listed. Align is a probiotic for which a beneficial effect for IBS symptoms has been shown in a high-quality, randomized, controlled trial.
††A wide range of tricyclic antidepressants with various side effects and side-effect profiles are available. Two commonly prescribed tricyclic antidepressants are listed.
‡‡Many selective serotonin reuptake inhibitors are available. Only those that have been evaluated in IBS trials are listed.

specific medications can be targeted to the management of individual symptoms, such as constipation, diarrhea, and abdominal pain (Table 137-2).

Physician-Patient Relationship and Patient Education

The doctor-patient relationship is of utmost importance in the treatment of IBS and other functional GI diseases. Many patients have been told previously that their symptoms are "all in their head" and have had their concerns dismissed. The physician must listen to and determine the patient's understanding of the illness and related concerns because patients frequently seek validation from the physician that their symptoms are real. Patients with functional GI disease typically visit a physician because of a flare of their usual symptoms, and it is important to identify such triggering factors, in particular various psychosocial stressors or, less commonly, a gastroenteric infection.

A thorough explanation of the symptoms and relationship to functional GI disease should be provided to the patient, including the natural history and benign prognosis of the disorder. Patients who are not properly informed tend to have more health care visits, whereas symptoms generally are reduced when diagnostic and prognostic information is explained. The physician must then set realistic short-term and long-term goals, including methods of adapting to symptoms that are not amenable to treatment. Patients often benefit from being involved in their treatment, and the maintenance of a symptom diary is one such example. Apart from providing the patient with a sense of empowerment, the symptom diary can help identify erroneous beliefs as well as lifestyle or dietary factors that may exacerbate symptoms; lifestyle modifications based on this information may provide symptomatic relief.

Dietary Modification

Some patients with functional GI disease, especially patients with IBS and functional dyspepsia, often complain that certain types of food exacerbate their symptoms. Others complain that any type of food, even a sip of water, may trigger symptoms. Perceived food sensitivity may be related to a variety

of factors, including conditioned fear responses related to anticipatory anxiety, food intake in general, volume of the meal, or sensitivities to certain food items. For example, a diet low in fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAP) can effectively reduce symptoms of IBS in some patients. Some IBS patients report an exacerbation of symptoms with high-fat food, milk products, gas-producing food, gluten-containing food, alcohol, and caffeine. It is important to inform the patient that there is no universal diet that is successful in every IBS patient. However, a therapeutic trial of a lactose-free diet or a gluten-free diet[6] may benefit some patients, even in the absence of obvious lactose intolerance or documentable celiac disease. If, however, such dietary changes are not effective in consistently reducing symptoms, patients should resume consumption of the eliminated foods.

Pharmacologic Therapy

Medications for Bowel Habit Abnormalities and Bloating

Few medications have been proved to benefit patients with IBS.[7] In IBS with constipation, osmotic laxatives, such as polyethylene glycol or lactulose, or secretory stimulators, such as lubiprostone, or linaclotide can be useful. Dietary fiber supplements often increase symptoms of gas, bloating, and flatulence and should not be used as first-line therapy.

In the IBS patient with diarrhea, loperamide is generally effective in reducing urgency and uncontrollable bowel movements. The 5-HT_3 receptor antagonist alosetron (0.5 to 1 mg orally twice per day) is clinically effective but should be considered only in patients who have severe diarrhea and who have failed all other therapies because of its infrequent but potentially serious side effects.

Manipulation of the intestinal flora with antibiotics (rifaximin 550 mg three times daily for 2 weeks) may be effective in some patients, particularly patients with abdominal bloating symptoms. It should not be used as a first-line drug, however, because its long-term effects on the gut microbiome are unknown. Clinical evidence supports the use of probiotics for nonpainful symptoms, but the response rate is low. Herbal laxatives such as aloe, rhubarb root products, and peppermint oil may be of use in some patients.

Antidepressants and Psychological Therapies

Low doses of tricyclic antidepressants (e.g., nortriptyline, amitriptyline, or imipramine, 10 to 50 mg at bedtime) are frequently used to treat IBS based on data that about one patient in four may benefit. Doses should be started as low as 5 mg once daily and gradually advanced to a maximum of 50 mg once daily if tolerated. The therapeutic effect should be expected within days to 2 weeks of initiating therapy; if no beneficial effect is observed at 50 mg once daily or if significant side effects are experienced, the drug should be discontinued. Amitriptyline may be most useful in patients with prominent abdominal pain and diarrhea symptoms and in patients with sleep problems, but other tricyclics should be tried if side effects are experienced.

To maximize compliance and reduce side effects, the patient should be informed about the rationale of the treatment choice (i.e., the goal is the treatment of pain and not psychiatric symptoms) and informed about the much lower risk for side effects at the low dose range compared with full psychiatric doses.

If the treatment goal is aimed at comorbid depression or anxiety disorders, a selective serotonin reuptake inhibitor (SSRI) should be tried. In such patients, either a combination of a low-dose tricyclic antidepressant with a full therapeutic dose of an SSRI (see Table 137-2) or the use of a nonselective uptake inhibitor (SNRI) such as duloxetine, milnacipran, or venlafaxine should be considered.

Cognitive-behavioral therapy, which involves relaxation, change in beliefs, and self-management, can reduce gastrointestinal symptoms in at least 50% of patients. Novel delivery forms of cognitive-behavioral therapy, such as minimal contact therapy or Internet-assisted therapy, are cost-effective approaches where available. Several controlled trials support the effectiveness of hypnotherapy.

PROGNOSIS

The natural history of IBS is periods of exacerbation followed by periods of remission, but about 50% of IBS patients become asymptomatic. Patients with coexisting psychiatric disorders are less likely to have their IBS symptoms resolve.

FUNCTIONAL DYSPEPSIA

As with other functional GI diseases, the diagnosis of functional dyspepsia is based on specific symptoms (Tables 137-3 and 137-4). Functional dyspepsia is thought to originate from the upper GI tract, but no detectable organic disease can explain the symptoms. Symptoms may include epigastric pain, epigastric burning, postprandial fullness, and early satiation. Abdominal bloating and nausea also may be experienced, but they are less specific and are not considered cardinal symptoms of functional dyspepsia. Symptoms overlap with atypical manifestations of gastroesophageal reflux disease (GERD; Chapter 138), and previous studies on functional dyspepsia may have inadvertently included patients with atypical GERD symptoms.

TABLE 137-3 ROME III DIAGNOSTIC CRITERIA FOR FUNCTIONAL DYSPEPSIA*

1. One or more of the following:
   - Bothersome postprandial fullness
   - Early satiation
   - Epigastric pain
   - Epigastric burning

and

2. No evidence of structural disease (including at upper endoscopy) that is likely to explain the symptoms

*The criteria must be fulfilled for the last 3 months with symptom onset at least 6 months before diagnosis.

TABLE 137-4 ROME III DIAGNOSTIC CRITERIA FOR SUBGROUPS OF PATIENTS WITH FUNCTIONAL DYSPEPSIA*

POSTPRANDIAL DISTRESS SYNDROME

One or both of the following:
1. Bothersome postprandial fullness, occurring after ordinary sized meals, at least several times per week
2. Early satiation that prevents finishing a regular meal, at least several times per week

Supportive criteria

1. Upper abdominal bloating or postprandial nausea or excessive belching can be present.
2. Epigastric pain syndrome may coexist.

EPIGASTRIC PAIN SYNDROME

One or more of the following:
1. Pain or burning localized to the epigastrium of at least moderate severity at least once per week
2. Intermittent pain
3. Pain not generalized or localized to other abdominal or chest regions
4. Pain not relieved by defecation or passage of flatus
5. Pain not fulfilling criteria for gallbladder and sphincter of Oddi disorders

Supportive criteria

1. The pain may be of a burning quality but without a retrosternal component.
2. The pain is commonly induced or relieved by ingestion of a meal but may occur while fasting.
3. Postprandial distress syndrome may coexist.

*The criteria must be fulfilled for the last 3 months with symptom onset at least 6 months before diagnosis.

In the current Rome III criteria, symptoms are divided into the meal-related postprandial distress syndrome and the meal-unrelated epigastric pain syndrome (see Table 137-4). The clinical utility of these subgroups is controversial because these two entities often overlap in the same patient. Moreover, the concept of distinct pathophysiologies correlating with distinct symptom patterns has not been confirmed.[8]

EPIDEMIOLOGY

Functional dyspepsia is a common disorder, with an estimated prevalence of 3 to 10% based on Rome III criteria and of up to 40% when less restrictive criteria are used. The socioeconomic burden of functional dyspepsia is substantial; patients with functional dyspepsia take three times as much sick leave as patients with duodenal ulcers. In the United Kingdom, an estimated 2 to 5% of primary care visits and more than 10% of primary care drug expenditures are related to functional dyspepsia. Approximately one of two individuals with functional dyspepsia seeks health care for symptoms at some time in their life. In the United States, the diagnosis, treatment, and work absenteeism related to functional dyspepsia have been estimated to cost $18.4 billion per year.

Pain or discomfort in the upper abdomen may be assumed to be related to the upper GI tract, but on detailed questioning, the "dyspepsia" may be related to bowel disturbances. One third of patients with functional dyspepsia have concurrent symptoms of IBS, and approximately 40% of IBS patients

also report symptoms of functional dyspepsia. Moreover, transitions between the two syndromes in the same patient over time are common. In a 1-year follow-up of patients with IBS or dyspepsia, 22% of IBS patients reported a change in their symptom profile to that of functional dyspepsia, and 16% of patients with functional dyspepsia reported a change to an IBS symptom profile. This transition of patients between different diagnostic categories puts into question the concept that these symptom-based entities are really distinct pathophysiologic syndromes. Compared with patients with non-life-threatening organic GI disease, patients with functional dyspepsia have higher anxiety but not depression or neuroticism scores.

PATHOBIOLOGY

The pathobiology of functional dyspepsia is not fully understood, but both central and peripheral mechanisms have been proposed. Although enhanced perception of gastric stimuli may be a key central mechanism, the roles of gastric acid, acute and chronic gastric mucosal infections, and gastroduodenal dysmotility remain to be determined. Attempts to classify subtypes of functional dyspepsia based on pathophysiologic abnormalities or predominant symptoms have been unsuccessful so far.

As with IBS patients, visceral hypersensitivity is common in functional dyspepsia, and 34 to 65% of patients with functional dyspepsia report pain and discomfort at lower volumes of gastric distention than healthy control subjects or patients with dyspepsia from organic causes. Chemical sensitivity to capsaicin also has been reported. Central sensory augmentation of painful and nonpainful stimuli also likely contribute to the observed gastric hypersensitivity. For example, the presence of lipids in the duodenum increases the sensitivity of patients with functional dyspepsia to gastric balloon distention compared with controls. This abnormal modulation of gastric perception thresholds to distention by lipid in a distant site supports the concept of a centrally mediated mechanism and is consistent with the clinical observation that fatty foods worsen symptoms in such patients. Furthermore, about 50% of patients with functional dyspepsia experience altered viscerosomatic referral patterns in response to gastric balloon distention. Some patients are hypersensitive to either intraduodenal or intra-antral acid or suffer from atypical manifestations of GERD.

In a small subset of patients, chronic *Helicobacter pylori* infection (Chapter 139) may be related to functional dyspepsia. About 20% of patients with functional dyspepsia develop their symptoms after an acute episode of presumed viral or bacterial gastroenteritis or *Giardia* species infection (Chapter 351).

Alterations in gastroduodenal motility, such as delayed gastric emptying, are found in some patients with functional dyspepsia, but the low concordance between symptoms and altered motility argues against a pathophysiologic link. Anxiety, depression, somatization, and stress contribute to the severity of symptoms and their impact on quality of life.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Dyspepsia can be suspected to be functional based on a clinical history consistent with the Rome III criteria (see Tables 137-3 and 137-4) and the absence of alarm features (see Table 137-1). The presence of anxiety, in particular symptom-related anxiety and comorbid IBS, increases the likelihood of functional dyspepsia. Nonsteroidal anti-inflammatory medications, alcohol, and certain foods can trigger dyspeptic symptoms. A psychosocial history may reveal underlying stressors that contribute to symptoms.

The physical examination is generally normal, although epigastric tenderness may be present. In contrast to gastroparesis, a succussion splash indicative of delayed gastric emptying is typically absent. Although confirmation of the functional dyspepsia diagnosis requires a normal upper endoscopic examination, invasive testing in the absence of alarm features should be considered only in a minority of symptomatic patients, such as patients older than 50 years with new-onset or changing symptoms or a poor response to initial therapy. Evaluation for *H. pylori* (Chapter 139) may be performed by stool antigen, urea breath test, or gastric biopsy.

Differential Diagnosis

Common organic causes of dyspepsia include gastroesophageal reflux disease (Chapter 138) and peptic ulcer disease (Chapter 139). Mild to moderately delayed gastric emptying is present in about 30% of patients with functional dyspepsia but is characteristic and more pronounced in patients with diabetic or idiopathic gastroparesis (Chapter 136). Delayed vomiting of undigested food is characteristic of these forms of gastroparesis, but not of dyspepsia. Gastric and esophageal cancers (Chapter 192) may also present with symptoms of dyspepsia but are much less common. Pancreaticobiliary disorders (Chapters 144 and 155) (including sphincter of Oddi dysfunction, chronic pancreatitis, or pancreatic cancer) also occasionally mimic dyspepsia.

TREATMENT

H. pylori infection, if present, should be eradicated[9] (Chapter 139), and then symptoms should be reassessed. About 10 to 15% of patients with functional dyspepsia respond to acid suppression therapy, and options include histamine (H_2)-receptor antagonists (e.g., famotidine, 20 mg twice per day) and proton pump inhibitors (e.g., omeprazole, 20 mg per day).[10] Young patients who respond well to a trial of proton pump inhibitor (see Table 138-1) therapy or *H. pylori* eradication (see Table 139-4) do not require further investigation unless alarm features are identified. If symptoms persist, some patients respond to treatment with a low-dose tricyclic antidepressant (e.g., amitriptyline, desipramine, or imipramine 10 to 50 mg every night at bedtime). In patients with comorbid depression or anxiety, the combination of low-dose tricyclics with a full dose of an SSRI or the use of an SNRI (see Table 137-2) should be considered.

The benefits of cognitive-behavioral therapy and hypnosis for functional dyspepsia have not been studied in high-quality trials, but such therapies may be reasonable in patients who do not respond to pharmacotherapy. Prokinetic drugs such as domperidone and acotiamide are of uncertain benefit and are not approved by the U.S. Food and Drug Administration for this purpose.

PROGNOSIS

As with all functional disorders, functional dyspepsia has a benign prognosis with a high rate of spontaneous remissions, although symptoms can be persistent.

FUNCTIONAL CHEST PAIN OF PRESUMED ESOPHAGEAL ORIGIN AND FUNCTIONAL HEARTBURN

DEFINITION

Functional chest pain of presumed esophageal origin (Table 137-5) is a chronic, unexplained midline chest pain that is thought to be of esophageal origin. To make the diagnosis, cardiac causes, gastroesophageal reflux, and well-defined motility disorders (achalasia, scleroderma) must be excluded. Functional heartburn is defined as a burning retrosternal discomfort of pain that persists for at least 3 months in the absence of GERD or an esophageal motility disorder.[11]

EPIDEMIOLOGY

Functional chest pain is quite common, with prevalence rates as high as 25%, evenly divided between men and women. Its prevalence appears to decline with advancing age. Most patients who present to a physician with acute chest pain (see Table 51-2) have a noncardiac cause, and many are cases of functional chest pain. Risk factors for developing functional chest pain are not well defined but include younger age, adversity in childhood, and a history of other functional gastrointestinal conditions. Once symptoms occur, more than 50% of patients have persisting symptoms for longer than 6 months. Even after the initial diagnosis, however, many patients with chronic pain undergo repeated and unnecessary diagnostic cardiac evaluations. Patients with functional chest pain may have comorbid anxiety or panic attacks, but formal referral for psychological treatment is infrequent. When compared with patients who have known cardiac chest pain, patients with functional chest pain report

TABLE 137-5 ROME III DIAGNOSTIC CRITERIA FOR FUNCTIONAL CHEST PAIN OF PRESUMED ESOPHAGEAL ORIGIN*

Must include all of the following:

- Midline chest pain or discomfort that is not of burning quality
- Absence of evidence that gastroesophageal reflux is the cause of symptoms
- Absence of histopathology-based esophageal motility disorders

*The criteria must be fulfilled for the last 3 months with symptom onset at least 6 months before diagnosis.

greater impairment of health-related quality of life in the domains of mental health and vitality. These findings suggest a link between psychological disturbance and functional chest pain, at least in a subgroup of patients.

Functional heartburn has been reported in about 20% of patients evaluated in a tertiary referral population for heartburn that is refractory to proton pump inhibitor therapy. However, its prevalence in the general population is unknown.

PATHOBIOLOGY

The pathophysiology of both functional chest pain and functional heartburn are incompletely understood, although visceral hypersensitivity, altered esophageal motility, and psychological factors have all been implicated. A substantial proportion of patients show enhanced sensitivity to esophageal distention by a balloon. These same patients will often also have increased perceptual responses to intraesophageal acid infusion. The origin of such findings is not clear, although alterations in central processing of sensory signals from the esophagus (central sensory augmentation) have been implicated. In addition, altered esophageal motility can be observed in a subset of patients with functional chest pain, with or without visceral hypersensitivity. Increased contraction of the esophageal muscle has long been considered a source of chest pain, although reproducible studies to prove this hypothesis are lacking.

CLINICAL MANIFESTATIONS

Patients with functional chest pain may complain of pain that is typical for myocardial ischemia or of pain with a variety of characteristics that would be considered atypical for ischemia (Chapters 51 and 71). The location of pain is typically substernal, but radiation to the arm and neck can be described. The pain typically is not precipitated by exertion and may persist for hours. Nitroglycerin may sometimes be helpful acutely in patients with coexisting esophageal spasm. Patients may describe the discomfort using a variety of adjectives, and those descriptions may be indistinguishable from angina.

Patients with functional heartburn complain of typical heartburn symptoms that usually are triggered by food intake and psychosocial stress but are refractory to acid suppressive therapy with proton pump inhibitors. By definition, they also have negative endoscopic and acid monitoring evaluations of the esophagus.

DIAGNOSIS

Patients who present with chest pain suspicious for angina should undergo prompt cardiac evaluation (Chapters 51 and 71). In patients with an initial negative cardiac evaluation, causes of functional chest pain can be categorized based on historical features (Table 137-6), and repeated cardiac evaluation for recurrent pain is of low yield. Clinical features, such as association with certain foods or pain location, can help differentiate well between acid-induced and non-acid-related causes. Patients with prominent chest wall pain and tenderness on palpation or with changes in pain with movement usually have musculoskeletal rather than esophageal pain. The remaining patients may have an esophageal source of pain (Chapter 138) and should be divided into patients with and without alarm symptoms, such as weight loss, progressive dysphagia, or anemia. Upper endoscopy is useful in patients with alarm features but is of lower yield in patients without them.

Many patients with functional chest pain or heartburn will have atypical GERD and should be given a therapeutic trial of a proton pump inhibitor (e.g., omeprazole, 20 mg daily).[12] Patients who respond to a 1- to 2-week trial of daily proton pump inhibitor likely have acid-related symptoms and should be treated for GERD. In patients whose response is equivocal, a longer proton pump inhibitor trial of 4 to 8 weeks, endoscopy, or 24-hour pH testing can be considered.

For patients who do not respond to a proton pump inhibitor, a disorder of esophageal motility or visceral hypersensitivity may be present. Esophageal motility disorders such as high-amplitude contractions ("nutcracker esophagus") or diffuse esophageal spasm can be identified by esophageal manometry, but the low concordance between symptoms and manometric findings suggests that such testing should be performed only in highly selected patients based on the advice of a gastroenterologist.

TABLE 137-6 DIFFERENTIAL DIAGNOSIS OF FUNCTIONAL CHEST PAIN

ORGANIC ESOPHAGEAL CAUSES
Gastroesophageal reflux disease
Achalasia
Virus- or pill-induced esophagitis
NONGASTROINTESTINAL CAUSES
Cardiac chest pain
Chest wall pain
Pulmonary disease
Panic attack

TREATMENT

Rx

For both functional chest pain and functional heartburn, low-dose tricyclic antidepressants (e.g., amitriptyline, imipramine, or desipramine, 10 to 50 mg per day) may be useful, particularly in patients suspected to have visceral hypersensitivity.[13] Comorbid psychological symptoms should be addressed with either pharmacologic (addition of an SSRI to the tricyclic antidepression or trial with an SNRI [see Table 137-2]) or cognitive behavioral approaches. Even in the absence of a specific psychiatric diagnosis, relaxation therapy or cognitive behavioral therapy may be of benefit.

In the patient with functional chest pain, fears of cardiac disease should be addressed explicitly, and the significance of a negative cardiac evaluation must be reinforced. If the functional chest pain is responsive to a proton pump inhibitor trial, patients should be continued on such therapy or treated with alternative acid-suppressing medications. The optimal duration of treatment is unclear, but it is reasonable to attempt withdrawal of medication and observe patients who have remained asymptomatic for several months.

PROGNOSIS

The natural course of functional chest pain is not well understood, likely because most patients undergo a cardiac evaluation but then may have no further evaluation of their symptoms. However, both functional chest pain and functional heartburn have a benign prognosis, although symptoms may persist and continue to diminish quality of life.

Grade A References

A1. Halmos EP, Power VA, Shepherd SJ, et al. A diet low in FODMAPs reduces symptoms of irritable bowel syndrome. *Gastroenterology*. 2014;146:67-75.

A2. Atluri DK, Chandar AK, Bharucha AE, et al. Effect of linaclotide in irritable bowel syndrome with constipation (IBS-C): a systematic review and meta-analysis. *Neurogastroenterol Motil*. 2014;26:499-509.

A3. Menees SB, Maneerattannaporn M, Kim HM, et al. The efficacy and safety of rifaximin for the irritable bowel syndrome: a systematic review and meta-analysis. *Am J Gastroenterol*. 2012;107:28-35.

A4. Moayyedi P, Ford AC, Talley NJ, et al. The efficacy of probiotics in the treatment of irritable bowel syndrome: a systematic review. *Gut*. 2010;59:325-332.

A5. Pajak R, Lackner J, Kamboj SK. A systematic review of minimal-contact psychological treatments for symptom management in irritable bowel syndrome. *J Psychosom Res*. 2013;75:103-112.

A6. Wang WH, Huang JQ, Zheng GF, et al. Effects of proton-pump inhibitors on functional dyspepsia: a meta-analysis of randomized placebo-controlled trials. *Clin Gastroenterol Hepatol*. 2007;5:178-185.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Barbara G, Cremon C, Stanghellini V. Inflammatory bowel disease and irritable bowel syndrome: similarities and differences. *Curr Opin Gastroenterol*. 2014;30:352-358.
2. Camilleri M. Peripheral mechanisms in irritable bowel syndrome. *N Engl J Med*. 2012;367:1626-1635.
3. Mayer EA, Savidge T, Shulman RJ. Brain-gut microbiome interactions and functional bowel disorders. *Gastroenterology*. 2014;146:1500-1512.
4. Simren M, Barbara G, Flint HJ, et al. Intestinal microbiota in functional bowel disorders: a Rome foundation report. *Gut*. 2013;62:159-176.
5. Beyder A, Mazzone A, Strege PR, et al. Loss-of-function of the voltage-gated sodium channel NaV1.5 (channelopathies) in patients with irritable bowel syndrome. *Gastroenterology*. 2014;146:1659-1668.
6. Volta U, De Giorgio R. New understanding of gluten sensitivity. *Nat Rev Gastroenterol Hepatol*. 2012;9:295-299.
7. Barboza JL, Talley NJ, Moshiree B. Current and emerging pharmacotherapeutic options for irritable bowel syndrome. *Drugs*. 2014;74:1849-1870.
8. Tack J, Talley NJ. Functional dyspepsia: symptoms, definitions and validity of the Rome III criteria. *Nat Rev Gastroenterol Hepatol*. 2013;10:134-141.
9. Suzuki H, Moayyedi P. Helicobacter pylori infection in functional dyspepsia. *Nat Rev Gastroenterol Hepatol*. 2013;10:168-174.
10. Lacy BE, Talley NJ, Locke GR 3rd, et al. Review article: current treatment options and management of functional dyspepsia. *Aliment Pharmacol Ther*. 2012;36:3-15.
11. Kumar AR, Katz PO. Functional esophageal disorders: a review of diagnosis and management. *Expert Rev Gastroenterol Hepatol*. 2013;7:453-461.
12. Coss-Adame E, Erdogan A, Rao SS. Treatment of esophageal (noncardiac) chest pain: a review. *Clin Gastroenterol Hepatol*. 2014;12:1224-1245.
13. Zerbib F, Bruley des Varannes S, Simon M, et al. Functional heartburn: definition and management strategies. *Curr Gastroenterol Rep*. 2012;14:181-188.

138

DISEASES OF THE ESOPHAGUS

GARY W. FALK AND DAVID A. KATZKA

NORMAL ANATOMY AND PHYSIOLOGY

The esophagus, which averages about 27 cm in length, is a hollow muscular tube consisting of the mucosa, submucosa, and muscularis layers, with the

notable absence of a serosal layer. The mucosa is a stratified squamous nonkeratinized epithelium that transitions to a columnar epithelium at the gastroesophageal junction. The muscular layer of the esophagus is composed of striated muscle in the upper one third and smooth muscle in the lower two thirds. These muscular components are arranged as an inner circular and an outer longitudinal layer. Located between the circular and longitudinal muscle layers is Auerbach (myenteric) plexus, whereas Meissner plexus is located within the submucosa and innervates the muscularis mucosae. The esophagus is bound by an upper esophageal sphincter proximally and the lower esophageal sphincter distally. The upper esophageal sphincter contains functional contributions from the inferior pharyngeal constrictor proximally and the cricopharyngeus distally. By contrast, the lower esophageal sphincter is anatomically and histologically indistinguishable from the lower esophagus. Blood supply for the cervical portion is derived from branches of the inferior thyroid artery. The intrathoracic segment of the esophagus receives its blood supply from bronchial arteries and direct branches from the aorta, and the left gastric and the inferior phrenic arteries supply the abdominal portion of the esophagus. Venous drainage follows the arterial supply in the cervical and abdominal portions, whereas the thoracic esophagus drains into the azygous and hemiazygos system. Likewise, the lymphatic drainage of the esophagus is segmental, with the cervical portion draining into deep cervical lymph nodes, the thoracic portion into the superior and posterior mediastinal lymph nodes, and the abdominal portion into the gastric and celiac lymph nodes.

The motor functions of the esophagus are to transport a food bolus from the oropharynx into the stomach and then to keep food from returning to the esophagus after it has entered the stomach. The upper esophageal sphincter and the proximal third of the esophagus compose the first portion of the esophagus. The upper esophageal sphincter is approximately 2 to 4 cm in length. The recurrent laryngeal nerve inferiorly and a pharyngeal plexus superiorly supply the upper esophageal sphincter, which is approximately 2 to 4 cm in length. The muscle layers close the esophageal lumen and shorten the esophagus to facilitate forward transport through the proximal esophagus. After food traverses the proximal esophagus, it moves into the distal two thirds of the esophagus, where peristalsis is achieved by sequential muscular contraction mediated through an interplay of inhibitory and excitatory neurotransmitters. Peristalsis may be primary, that is, initiated by a swallow, or secondary, that is, stimulated by refluxed gastric contents. Although local mechanisms control most esophageal motor function, vagal input is important in the distal esophagus, where smooth muscle myopathies and autonomic neuropathies can cause dysfunction. The distal esophagus is separated from the stomach by the lower esophageal sphincter, which is 4 to 5 cm in length and is functionally distinct because it maintains a tonic high-pressure zone. This sphincter relaxes nearly completely upon swallowing to allow the passage of food and then regains its tone to provide a barrier against reflux. It also relaxes transiently during normal functions, such as belching and vomiting. The vagus nerve, acetylcholine, and nitric oxide influence tone, but the lower esophageal sphincter tone is influenced by the crural diaphragm as the esophagus traverses the diaphragmatic hiatus.

Esophageal Functional Testing

Options for esophageal functional testing include barium esophagography, high-resolution esophageal manometry, and esophageal impedance testing. Barium esophagography (Chapter 133) reveals both anatomic and physiologic information about luminal lesions, such as malignancies, ulceration, diverticula, hiatal hernia, and strictures; intramural lesions, such as leiomyomas; and extrinsic lesions, such as occur from vascular (aorta, right atrium, subclavian artery) impingement or solid lesions (pulmonary malignancy, adenopathy) that compress the esophagus. Radiography is also an excellent tool for studying motility patterns, such as peristalsis with either liquid or solid contrast material, while precisely visualizing how the esophagus handles a bolus rather than by implying function from pressure or impedance changes.

High-resolution esophageal manometry measures pressure changes generated by esophageal wall contraction and changes in tone using multiple sensors that simultaneously measure pressure from the pharynx to the lower esophageal sphincter. Esophageal impedance testing detects the movement of an intraluminal bolus by measuring conductance between catheter-based electrodes based on the substance that is in contact with each electrode. Air, which is a poor conductor of electric current, will yield high impedance, whereas swallowed or refluxed liquids, which are excellent conductors of electricity, will generate a low impedance signal. From these measurements, the direction and velocity of the transport of air and bolus can help assess peristaltic function and the reflux of acid and nonacid gastric contents.

Symptoms of Esophageal Disease

The most common symptom of esophageal disease is heartburn, which is defined as a sensation of substernal burning. Chest pain without typical heartburn may occur in a variety of esophageal disorders, including gastroesophageal reflux and motor disorders such as in achalasia. However, esophageal pain and even heartburn can be indistinguishable from cardiac angina (Chapter 51), so care must be taken when a patient at risk for coronary artery disease complains of heartburn for the first time.

Dysphagia, or difficulty swallowing, is another cardinal symptom of esophageal disease. Dysphagia with only solid food tends to occur with structural lesions, which cause esophageal constriction, whereas dysphagia with both liquids and solids occurs more often with motility disorders. Patients with oropharyngeal dysphagia will commonly complain of a feeling of food "sticking" in the throat or the inability to propel the bolus from the mouth to the pharynx; they may also complain of the need for multiple swallowing motions to clear the bolus. Since the cranial nerves that generally control the initial phases of swallowing are responsible for other functions as well, symptoms that may be associated with oropharyngeal dysphagia include drooling, dysarthria (due to tongue dysfunction), nasal regurgitation (due to failure to seal off the nasal passage), or coughing and aspiration (due to failure to elevate and cover the laryngeal vestibule). Dysphagia that results from abnormalities in the body of the esophagus may be referred to the chest or the neck, so the location of pain does not predict the location of the disease. Dysphagia may also lead to a variety of behavioral accommodations, including maneuvers such as slow eating, food aversion, avoidance of hard solid food, and drinking of large amounts of liquids with solid meals.

Regurgitation, which is another typical esophageal symptom, may be described as the feeling of food coming up into the chest or, more dramatically, into the mouth. When regurgitation occurs early in the meal, it suggests a proximal lesion. Regurgitation later in the meal suggests a motility abnormality such as achalasia.

Food impaction is an extreme esophageal symptom. When impaction occurs in the oropharynx, patients may develop a "steakhouse" syndrome, in which an impacted food bolus leads to tracheal impaction or compression. With more distal esophageal lesions, impaction may occur any time during the meal, almost always from a mechanical cause. Patients experience the sudden onset of chest pain and the sensation of food sticking, typically after solids such as meats, raw vegetables, and sticky rice. With complete impaction, patients who cannot handle secretions because of the obstructing bolus are at risk for aspiration, esophageal rupture, and perforation.

GASTROESOPHAGEAL REFLUX DISEASE

DEFINITION

Gastroesophageal reflux disease (GERD) develops when the reflux of stomach contents into the esophagus causes troublesome symptoms or complications.

EPIDEMIOLOGY

It is estimated that GERD, defined as at least weekly heartburn or acid regurgitation, has a prevalence ranging from 10 to 20% in the Western world and less than 5% in Asia. The prevalence also tends to be higher in North America than Europe and higher in northern Europe than in southern Europe. Risk factors for developing GERD include obesity, particularly central obesity, and possibly increasing age. A genetic component may also play a role because GERD is more common in patients with a positive family history and in monozygotic twins than in dizygotic twins.

PATHOBIOLOGY

The esophagus is protected from the harmful effects of refluxed gastric contents by the antireflux barrier at the gastroesophageal junction, by esophageal clearance mechanisms, and by epithelial defensive factors. The antireflux barrier consists of the lower esophageal sphincter, crural diaphragm, phrenoesophageal ligament, and angle of His, which causes an oblique entrance of the esophagus into the stomach. The attachment of the lower esophageal sphincter to the crural diaphragm results in increased pressure during inspiration and when intra-abdominal pressure increases. Disruption of normal defense mechanisms leads to pathologic amounts of reflux.[1]

Reflux of gastric contents from the stomach into the esophagus occurs in healthy individuals, but refluxed gastric contents are normally cleared in a two-step process: volume clearance by peristaltic function and neutralization of small amounts of residual acid by weakly alkaline swallowed saliva. In

normal healthy individuals, physiologic reflux occurs primarily when the lower esophageal sphincter transiently relaxes in the absence of a swallow because of a vagally mediated reflex that is stimulated by gastric distention. In GERD patients, transient relaxation of the lower esophageal sphincter or a low resting lower esophageal sphincter pressure can result in regurgitation, especially when intra-abdominal pressure is increased.

A hiatal hernia, which results in axial and vertical spatial separation between the augmenting effects of the crural diaphragm and the lower esophageal sphincter, predisposes to reflux events by widening the opening of the gastroesophageal junction and decreasing the pressure of the lower esophageal sphincter. The result is an increased exposure of the esophagus to acid and gastric contents, with increased reflux events during transient physiologic relaxation of the lower esophageal sphincter and/or increased gastric pressure. Hernias also act as a reservoir for gastric contents when normal esophageal clearance mechanisms result in trapping of fluids in the hernia sac. These contents can reflux into the esophagus when the lower esophageal sphincter relaxes during subsequent swallowing.

Normal individuals also have an unbuffered acid pocket in the gastric cardia, which escapes the buffering effects of a meal in the postprandial period. This region is a source of postprandial reflux and may explain the chronic inflammation often seen in the cardia and distal esophagus. In reflux patients, the acid pocket is more common and longer in length than in normal individuals. Displacement of the acid pocket into a hiatal hernia also appears to increase acidic reflux in patients with GERD.

Increased intra-abdominal fat associated with obesity increases intragastric pressure, which increases the gastroesophageal pressure gradient and the frequency of transient lower esophageal sphincter relaxation, thereby predisposing gastric contents to migrate into the esophagus. In addition, obesity enhances the spatial separation of the crural diaphragm and the lower esophageal sphincter, thereby predisposing obese individuals to a hiatal hernia. The metabolic syndrome (Chapter 229) that is associated with obesity may also have an independent effect in promoting esophageal injury in GERD.

The normal defense mechanisms based on peristalsis and saliva can also be impaired. Peristaltic dysfunction is associated with an increasing severity of esophagitis, and ineffective peristaltic clearance may occur when the amplitude of esophageal contractions is less than 20 mm Hg. Saliva production may be impaired by a variety of mechanisms, such as smoking and Sjögren syndrome (Chapter 268).

The esophageal mucosa contains several lines of defense. A pre-epithelial barrier constitutes a small unstirred water layer combined with bicarbonate from swallowed saliva and from the secretions of submucosal glands. A second epithelial defense is composed of cell membranes and tight intercellular junctions, cellular and intercellular buffers, and cell membrane ion transporters. The postepithelial line of defense is composed of the blood supply to the esophagus. Acid and acidified pepsin in the refluxate are the key factors that damage the intercellular junctions, increase intracellular permeability, and dilate intercellular spaces. If sufficient quantities of refluxate diffuse into the intercellular spaces, cellular damage may occur. Signs and symptoms of GERD occur when defective epithelium comes into contact with refluxed acid, pepsin, or other noxious gastric contents. In addition to the direct noxious effects of refluxed acid, pepsin, and bile, refluxed gastric juice stimulates esophageal epithelial cells to secrete chemokines that attract inflammatory cells into the esophagus, thereby damaging the esophageal mucosa.

CLINICAL MANIFESTATIONS

The classic symptoms of GERD are heartburn and acid regurgitation; atypical symptoms include chest pain, dysphagia, and odynophagia. Extraesophageal manifestations of reflux disease can include cough (Chapter 83), laryngitis (Chapter 429), asthma (Chapter 87), and dental erosions, but these symptoms are more easily attributable to GERD if accompanied by classic signs and symptoms of reflux disease. Other proposed associations that are not clearly established include pharyngitis, sinusitis, otitis media, and idiopathic pulmonary fibrosis (Fig. 138-1).

When excessive gastric contents overwhelm the mucosal protective factors in the esophagus, esophagitis may be manifest as erosions or ulceration of the esophagus and may also lead to fibrosis with stricturing, columnar metaplasia (Barrett esophagus) or esophageal adenocarcinoma (Chapter 192). However, approximately two thirds of individuals with reflux symptoms have no evidence of esophageal damage by endoscopy.

DIAGNOSIS

When GERD is associated with typical signs and symptoms, such as heartburn or acid regurgitation, that are responsive to antisecretory therapy, no diagnostic evaluation is warranted.[2,3] Diagnostic endoscopy is warranted in individuals who fail to respond to 4 to 8 weeks of therapy or have alarm symptoms or signs such as dysphagia, weight loss (Fig. 132-4), anemia, gastrointestinal bleeding, or persistent heartburn (Fig. 138-2).[4] Endoscopy permits the detection of erosive esophagitis and complications such as a peptic stricture (Fig. 138-3) and Barrett esophagus (Fig. 138-4); mucosal biopsy, which is crucial in these settings, also excludes conditions that can mimic GERD, such as eosinophilic esophagitis. However, most patients have no mucosal damage seen on endoscopy, regardless of whether they are on or off antisecretory therapy.

Esophageal manometry is useful to exclude achalasia in patients with suggestive symptoms. Esophageal reflux testing may be performed using 24-hour transnasal pH monitoring, 48-hour devices attached to the esophageal lumen, or 24-hour combined impedance and pH monitoring. Testing of a patient who is not receiving antisecretory therapy can document abnormal esophageal acid exposure and establish the relationship between symptoms and reflux events. Testing of a patient who is on therapy is best accomplished by combined impedance pH monitoring, which can establish the relationship, if any, between symptoms and reflux events, thereby permitting exclusion of GERD as the cause of persistent symptoms. Barium radiography has no role in the diagnostic evaluation of patients with reflux disease.

TREATMENT

Rx

Although avoidance of foods or beverages that may provoke symptoms, such as alcohol, coffee, spicy foods, and late meals, makes physiologic sense, data from clinical trials to support these maneuvers are lacking. Similarly, elevation of the head of the bed for patients with nocturnal regurgitation or

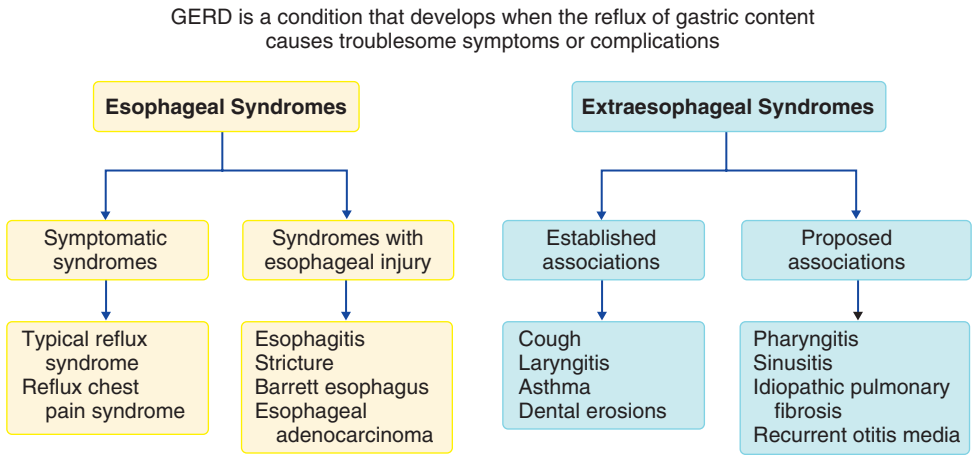

FIGURE 138-1. Montreal classification of gastroesophageal reflux disease (GERD). (From Vakil N, van Zanten S, Kahrilas P, et al. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *Am J Gastroenterol.* 2006;101:1900-1920.)

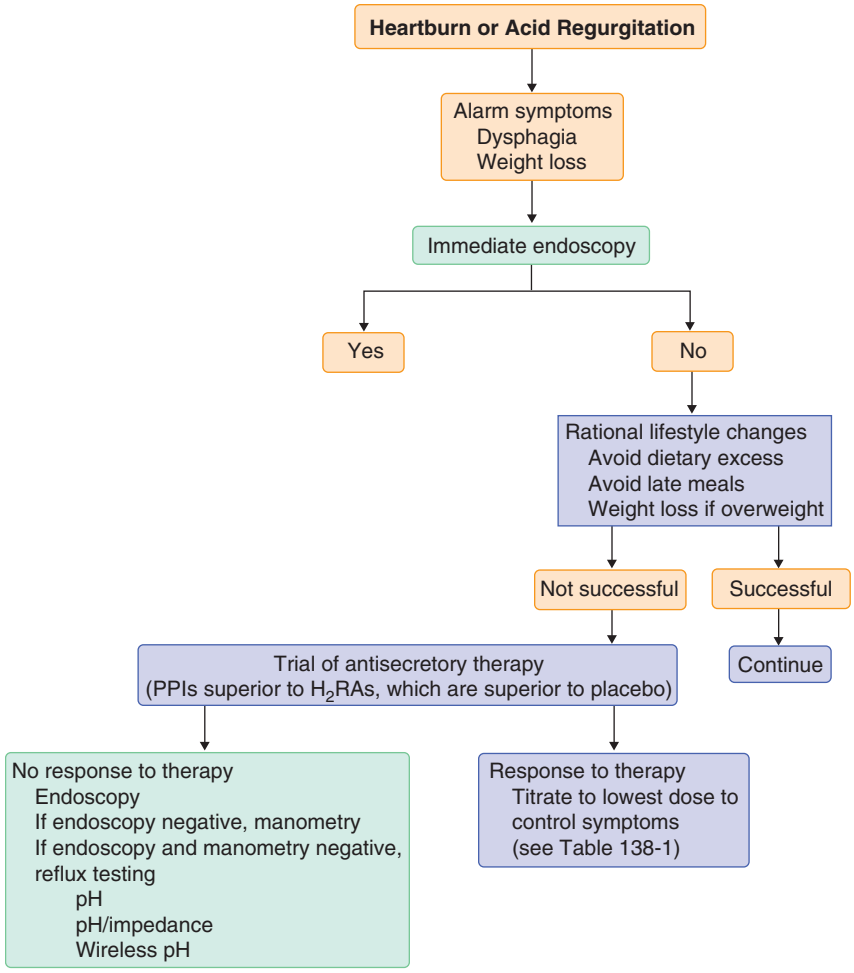

FIGURE 138-2. Algorithm for the management of heartburn or regurgitation symptoms. Surgery is indicated only for patients who are intolerant of antisecretory therapy or who have ongoing symptoms, especially regurgitation if reflux is well documented. H_2RA = histamine-2 receptor antagonist; PPI = proton pump inhibitor. (Based on Kahrilas PJ, Shaheen NJ, Vaezi MF. American Gastroenterological Association Institute technical review on the management of gastroesophageal reflux disease. *Gastroenterology.* 2008;135:1392-1413.)

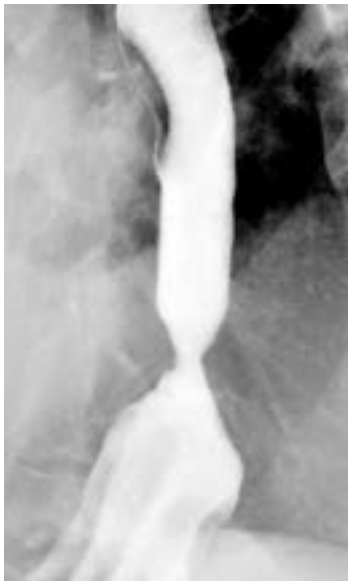

FIGURE 138-3. Barium radiograph of a peptic stricture. (Courtesy Marc Levine, MD.)

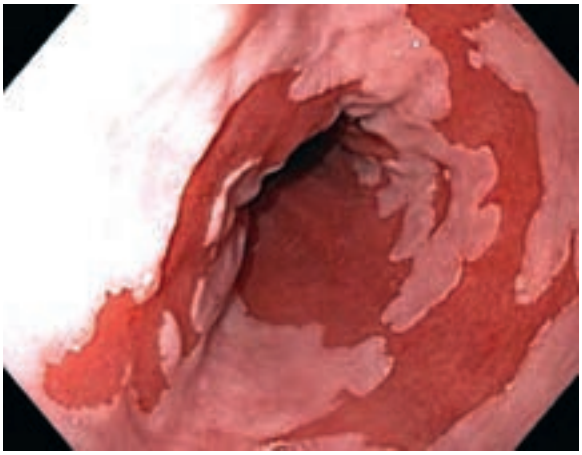

FIGURE 138-4. Endoscopic appearance of Barrett esophagus. Note the white-appearing normal squamous mucosa displaced above the true end of the esophagus. The intervening mucosa appears salmon-pink.

heartburn also is logical. Given the association of obesity and GERD symptoms, weight loss should be part of any treatment program for obese patients.

Inhibition of gastric acid secretion (Table 138-1) is the cornerstone of the acute treatment of GERD, and proton pump inhibitors are superior to histamine (H_2)-receptor antagonists for both the healing of esophagitis and the control of symptoms. However, the healing of esophagitis is more predictable than improvement in heartburn symptoms, even with proton pump inhibitors. There are no major differences in treatment efficacy among the various proton pump inhibitors, and once-daily dosing is adequate in most patients.

Given the chronicity of reflux symptoms, long-term maintenance therapy with proton pump inhibitors is typically required and is mandatory for patients with erosive esophagitis. Dosing should be titrated to the lowest dose necessary to control symptoms. Data to support the use of proton pump inhibitors in the management of extraesophageal GERD syndromes are weak, although selected patients may benefit. The safety profile of proton pump inhibitors is

TABLE 138-1 DRUG THERAPY FOR ESOPHAGEAL DISORDERS

AGENT	DOSE
ANTACIDS: LIQUID (TO BUFFER ACID AND INCREASE LESP)	
For example, Mylanta II/Maalox TC (acid-neutralizing capacity, 25 mEq/5 mL)*	15 mL qid 1 hr after meals and at bedtime or as needed
GAVISCON (TO DECREASE REFLUX VIA A VISCOUS MECHANICAL BARRIER AND BUFFER ACID)	
$Al(OH)_3$, $NaHCO_3$, Mg trisilicate, alginic acid	2-4 tablets qid and at bedtime or as needed
H_2-RECEPTOR ANTAGONISTS (TO DECREASE ACID SECRETION)	
Cimetidine	400 mg bid or 200 mg qid
Ranitidine	150 mg bid-qid or 10 mL qid; maintenance dose, 150 mg bid, 10 mL bid
Famotidine	20-40 mg bid or 2.5-5 mL bid
Nizatidine	150 mg bid
PROTON PUMP INHIBITORS (TO DECREASE ACID SECRETION AND GASTRIC VOLUME)†	
Omeprazole	20 mg/day; maintenance dose, 20 mg/day
Lansoprazole	15-30 mg/day; maintenance dose, 15 mg/day
Pantoprazole	40 mg/day; maintenance dose, 40 mg/day
Rabeprazole	20 mg/day; maintenance dose, 10-20 mg/day
Esomeprazole	20-40 mg/day; maintenance dose, 20 mg/day
Dexlansoprazole	30-60 mg/day; maintenance dose, 30 mg/day

*Patients with reflux are not generally hypersecretors of gastric acid, so the therapeutic doses of antacids are based on their capacity to buffer (normal) basal acid secretion rates of approximately 1 to 7 mEq/hr (mean, 2 mEq/hr) and peak meal-stimulated acid secretion rates of about 10 to 60 mEq/hr (mean, 30 mEq/hr).
†High-dose therapy is a twice-daily administration of the usually daily dose.
LESP = lower esophageal sphincter pressure.

excellent, but short-term adverse events such as headaches and diarrhea may occur. Long-term proton pump inhibitors use is associated with vitamin B_{12} deficiency[5] and may be associated with an increased risk for *Clostridium difficile* infection, community-acquired pneumonia, and hip fracture.

Although proton pump inhibitors are superior to H_2-receptor antagonists for long-term maintenance therapy as well as for short-term relief, H_2-receptor antagonists are superior to placebo and are useful in patients who cannot tolerate proton pump inhibitors. No high-quality data exist to support the common practice of using metoclopramide as either monotherapy or an adjunct to acid suppression therapy; furthermore, its significant adverse effects argue against the use of this drug in any GERD patients.

Antireflux surgery is an option for patients who have documented esophagitis, who are intolerant of proton pump inhibitors or unresponsive to them, or who regurgitate large volumes. Laparoscopic antireflux surgery is equivalent to continued proton pump inhibitor therapy for the healing of esophagitis and for the treatment of chronic GERD in patients who initially respond to proton pump inhibitors.[■■] However, surgery has a number of serious complications that may affect quality of life, including dysphagia, vagal nerve injury, gas bloat syndrome, and diarrhea. Endoscopic approaches to GERD are being studied but are not currently part of routine care.

PROGNOSIS

Patients with GERD generally do well with conservative antireflux measures and proton pump inhibitor therapy. When surgery is required, the outcome is usually excellent.

Peptic Strictures

Esophageal strictures are a well-recognized complication of GERD, especially in older patients with long-standing reflux symptoms, but population-based studies suggest that the incidence of new and recurrent strictures is declining. Peptic strictures are thought to be a consequence of severe inflammation, which leads to fibrosis, scarring, esophageal shortening, and loss of compliance of the lumen.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Symptoms are typically dysphagia to solids with or without antecedent symptoms of heartburn or acid regurgitation. Strictures may be diagnosed by barium radiography or with upper endoscopy, but barium esophagrams (see Fig. 138-3) have a higher sensitivity for detecting subtle lesions, especially if performed with a solid challenge such a barium-impregnated pill. However, peptic strictures must be distinguished from a wide variety of other causes of luminal narrowing, including pills, prior nasogastric tube intubation, neoplasia, infection, radiation, surgical anastomosis, some systemic diseases, caustic substances, and extrinsic compression. As a result, endoscopic biopsy and cytology are critical for distinguishing benign from malignant causes of strictures.

TREATMENT

Rx

Endoscopic dilation, which remains the cornerstone of therapy, should be done gradually to achieve a luminal diameter that is sufficiently large to relieve symptoms—typically a diameter of 13 mm or greater. After dilation is accomplished, patients should receive chronic proton pump inhibitor therapy (see Table 138-1). For recalcitrant peptic strictures, injection of triamcinolone into the stricture is superior to sham injection in patients who receive balloon dilation and postprocedure proton pump inhibitors.[■]

PROGNOSIS

Endoscopic dilation will usually alleviate symptoms, but repetitive dilation is required in a significant minority of patients.

Barrett Esophagus

Barrett esophagus, which is an acquired condition that results from severe esophageal mucosal injury, is a metaplastic change in the lining of the distal tubular esophagus, where the normal squamous epithelium is replaced by a columnar epithelium.[6] Barrett esophagus would be of little importance if not for its well-recognized association with adenocarcinoma of the esophagus (Chapter 192). However, the risk for cancer in an individual patient with Barrett esophagus is low.[7]

EPIDEMIOLOGY

It is estimated that Barrett esophagus is found in approximately 5 to 15% of patients who undergo endoscopy for symptoms of GERD. Population-based studies suggest that the prevalence of Barrett esophagus is approximately 1.3 to 1.6%, but about 45% of affected patients do not have reflux symptoms. Barrett esophagus is predominantly a disease of middle-aged white men, but about 25% of patients are women or are younger than 50 years. The prevalence of Barrett esophagus increases until a plateau is reached between the seventh and ninth decades. Risk factors include frequent and long-standing reflux episodes, smoking, male gender, older age, and central male pattern obesity.

PATHOBIOLOGY

Barrett esophagus results from severe esophageal mucosal injury. Patients who develop Barrett esophagus typically have more esophageal acid and bile exposure, the former based on 24-hour pH monitoring, and almost always have a hiatal hernia, which is typically longer and associated with larger defects than in patients without Barrett. However, why some patients with GERD develop Barrett esophagus whereas others do not remains unclear, as does the cell of origin of columnar metaplasia. Candidates include dedifferentiation of squamous epithelium into columnar epithelium or stimulation of stem cells from either the basal layer of the esophageal epithelium, the esophageal submucosal glands, residual embryonal stem cells, or the bone marrow. The transcription factor CDX2, which can be induced by both acid and bile salts, appears to play a role in promoting the development of the columnar epithelium in the distal esophagus and gastroesophageal junction. A small subset of patients may have an inherited predisposition to Barrett esophagus, although the genetics of the disease are unknown.

CLINICAL MANIFESTATIONS

The development of reflux symptoms at an earlier age, an increased duration of reflux symptoms, an increased severity of nocturnal reflux symptoms, and prior complications of GERD such as esophagitis, ulceration, stricture, and

bleeding may raise the likelihood of Barrett's esophagus. Nevertheless, patients with Barrett's esophagus are difficult to distinguish clinically from patients whose GERD is uncomplicated by a columnar-lined esophagus. Patients with Barrett's esophagus may paradoxically have impaired sensitivity to esophageal acid perfusion compared with patients with uncomplicated GERD.

DIAGNOSIS

Endoscopically, Barrett's esophagus is characterized by displacement of the squamocolumnar junction so that it is now proximal to the gastroesophageal junction, which is defined by the proximal margin of gastric folds (see Fig. 138-4). The diagnosis of Barrett's esophagus is established if the squamocolumnar junction is displaced proximal to the gastroesophageal junction and if intestinal metaplasia, which is characterized in part by acid mucin-containing goblet cells, is detected by biopsy.

The precise junction of the stomach and the esophagus may be difficult to determine endoscopically owing to the presence of a hiatal hernia, inflammation, and the dynamic nature of the gastroesophageal junction. If the squamocolumnar junction is above the level of the esophagogastric junction, as defined by the proximal margin of the gastric folds, biopsy specimens should be obtained for confirmation of columnar metaplasia.

Intestinal or columnar metaplasia may be seen in the cardia of normal individuals as well as in persons with chronic reflux disease, and the prevalence of intestinal metaplasia at a normal-appearing gastroesophageal junction varies from 5 to 36%. Dysplasia and an increased risk for carcinoma have been reported in patients who have intestinal metaplasia of the gastroesophageal junction or cardia, but the magnitude of that risk appears to be less than that of Barrett's esophagus.

TREATMENT

Rx

Patients who are diagnosed with Barrett esophagus require surveillance endoscopy at regular intervals (Fig. 138-5). They characteristically worry about cancer risk, which they may overestimate, face higher life insurance premiums, and may receive conflicting information on how best to treat their condition.

Proton pump inhibitors (see Table 138-1), which are the cornerstone of medical therapy for Barrett esophagus, consistently relieve symptoms and heal esophagitis. However, proton pump inhibitors, even at high doses, provide no more than modest regression of Barrett histology, perhaps because alleviation of reflux symptoms is not necessarily equivalent to normalization of esophageal acid exposure. In fact, abnormal acid exposure persists in approximately 25% of Barrett esophagus patients despite twice-daily proton pump inhibitors. The importance of complete control of esophageal acid exposure in patients with Barrett esophagus remains unknown, although data suggest that effective therapy can protect against the development of cancer.[8] Antireflux surgery effectively alleviates GERD symptoms, and the indications for surgery are the same as those for patients with GERD without Barrett esophagus. Surgery should not, however, be viewed as a better cancer prevention tool.

Current practice guidelines, based on observational data, recommend endoscopic surveillance of patients with documented Barrett esophagus in an attempt to detect dysplasia and cancer at an early and potentially curable stage.[9] Before entering into a surveillance program, patients should be advised about risks and benefits, including the limitations of surveillance endoscopy as well as the importance of adhering to appropriate surveillance intervals. Other considerations include age, likelihood of survival over the next 5 years, and ability to tolerate either endoscopic or surgical interventions for early esophageal adenocarcinoma.

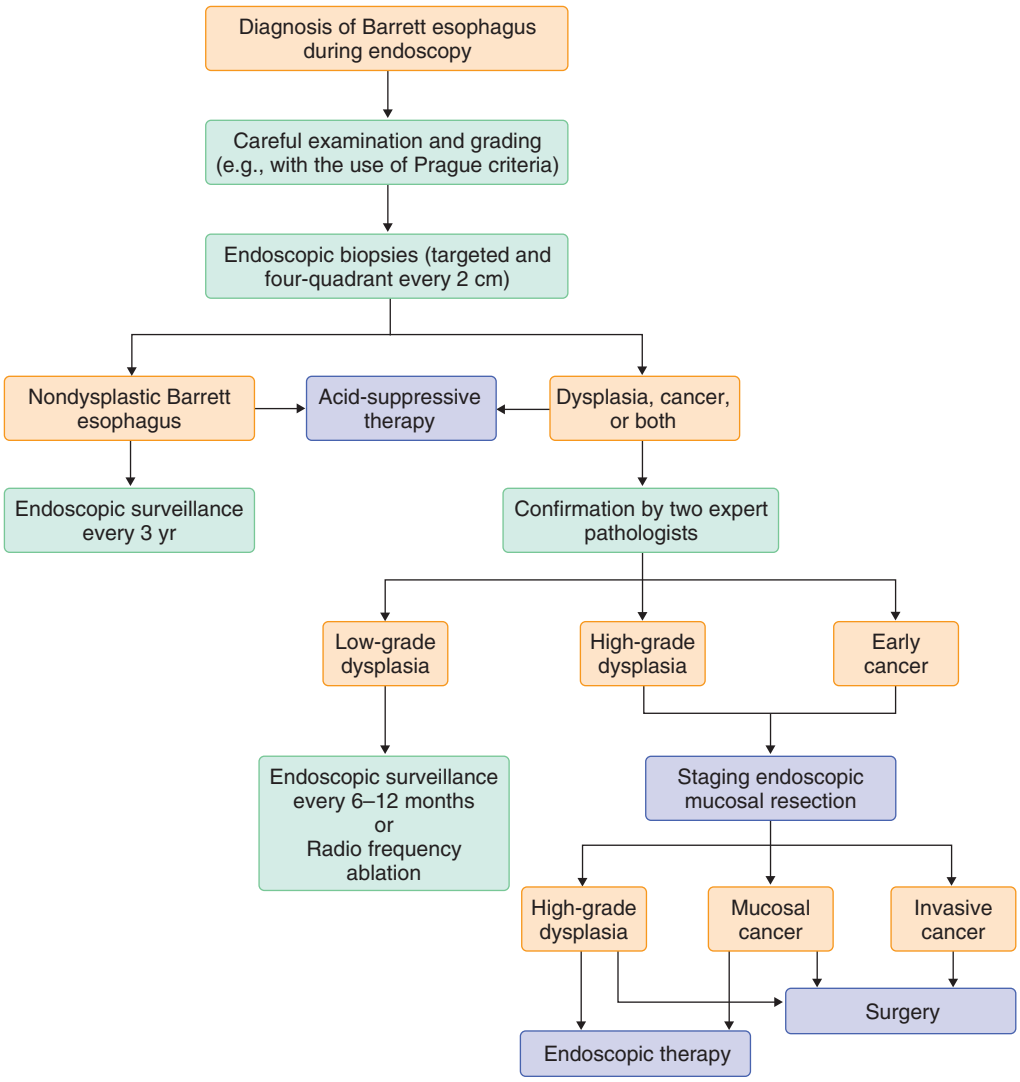

FIGURE 138-5. Proposed treatment algorithm for patients with Barrett esophagus. (From Sharma P. Barrett's esophagus. *N Engl J Med.* 2009;361:2548-2556, Fig. 3).

Systematic four-quadrant biopsies should be obtained at 2-cm intervals along the entire length of the Barrett segment after inflammation related to GERD is controlled with antisecretory therapy. Mucosal abnormalities, especially in the setting of high-grade dysplasia, should be resected endoscopically. Surveillance intervals, determined by the presence and grade of dysplasia, are based on a limited understanding of the biology of esophageal adenocarcinoma. Surveillance for patients without dysplasia should occur every 3 to 5 years after an initial negative examination. If low-grade dysplasia is found, the diagnosis should first be confirmed by an expert gastrointestinal pathologist because of the marked interobserver variability in the interpretation of these biopsies. If confirmed, aggressive proton pump inhibitor therapy (see Table 138-1) is recommended to decrease inflammation and regeneration, which may make pathologic interpretation difficult. A repeat endoscopy should then be performed within 6 months of the initial diagnosis. If low-grade dysplasia is confirmed, options include continued surveillance at 6- to 12-month intervals or radiofrequency ablation, but ablation significantly reduces the risk for progression to high-grade dysplasia or to adenocarcinoma over the next 3 years.

If high-grade dysplasia is found, an experienced independent gastrointestinal pathologist should confirm the diagnosis. For patients with confirmed high-grade dysplasia, endoscopic ablation therapy with radio frequency ablation, endoscopic mucosal resection, or a combination is now recommended instead of surgery or continued surveillance. For high-grade dysplasia patients with any visible abnormality, endoscopic mucosal resection is recommended for optimal diagnosis and staging.[10] Surgery with esophagectomy is reserved only for patients who fail to respond to endoscopic ablation therapy.

PROGNOSIS

The risk that a patient with Barrett esophagus will develop esophageal adenocarcinoma is now estimated to be approximately 0.1 to 0.3% annually,[11] and thus, most patients with Barrett esophagus will never develop esophageal adenocarcinoma but will die of other causes. Observational data suggest that aspirin, nonsteroidal anti-inflammatory drugs, and statins are associated with a reduced risk for neoplastic progression, but no clinical trial data yet support their routine use.

ESOPHAGITIS

Eosinophilic Esophagitis

Eosinophilic esophagitis is probably caused by an aberrant immune or antigenic response to food and aeroallergens that trigger chronic inflammation in the esophageal mucosa.[12] The disease is most common in children and adolescents, but adults are commonly affected as well. Patients often have a personal and family history of other allergic disorders. A genetic predisposition is suggested by abnormal gene profiles in almost 50% of children with this disorder.[13]

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Children present with dyspeptic symptoms, whereas adults present with solid food dysphagia, food impaction, or chest pain. Booerhaave syndrome (see later) may also occur with this disease. Diagnosis is made by classic findings on endoscopy, such as linear furrowing, white exudates, and multiple rings (Fig. 138-6), accompanied by biopsies demonstrating eosinophilic infiltration (Table 138-2) in the absence of gastroesophageal reflux disease.

TREATMENT AND PROGNOSIS

Rx

Treatment options for allergic eosinophilic esophagitis include food elimination diets, 2 months of topical steroids (e.g., swallowed fluticasone, 440 to 880 µg twice daily, or budesonide suspension, 1 mg twice daily, for 15 days, followed by 0.25 mg twice daily), or occasionally a 1-month course of systemic steroids (e.g., prednisone, starting at 40 mg and then tapering). Patients also should undergo a formal allergy evaluation (Chapter 249) to determine whether any food trigger can be identified and avoided. Another option is an empirical six-food elimination diet that eliminates cow's milk protein (casein), soy, wheat, egg, peanut/tree nuts, and seafood. An elemental diet may be used for severe disease, particularly in children. Most patients do well after treatment, although the optimal approach to chronic therapy remains uncertain.

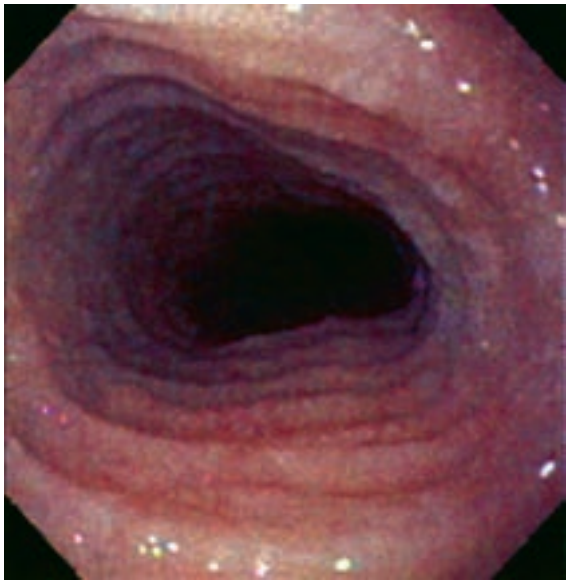

FIGURE 138-6. Endoscopic corrugated (ringed) appearance of esophagus in eosinophilic esophagitis.

TABLE 138-2 GUIDELINES FOR THE DIAGNOSIS OF EOSINOPHILIC ESOPHAGITIS

- Clinical symptoms of esophageal dysfunction
- At least 15 eosinophils in 1 high-power field
- Lack of responsiveness to high-dose proton pump inhibition (see Table 138-1) or normal pH monitoring of the distal esophagus

Pill-Induced Esophagitis

Pills (Table 138-3) can induce esophageal injury by producing a caustic acid solution (e.g., ascorbic acid and ferrous sulfate), producing a caustic alkaline solution (e.g., alendronate, button batteries), placing a hyperosmolar solution in contact with the esophageal mucosa (e.g., potassium chloride), or causing direct drug toxicity to the esophageal mucosa (e.g., tetracycline). Because prolonged contact is an essential part of the injury, predisposing factors for pill-induced injury include anatomic barriers, such as a stricture, a prominent aortic arch that compresses the esophagus, or improper ingestion of the pill because of inadequate fluid or improper positioning (i.e., lying down directly after taking the pill). The most common medications are tetracycline and its derivatives, but other commonly implicated medications include nonsteroidal anti-inflammatory drugs, bisphosphonates, ferrous sulfate, quinidine, and potassium chloride.

Patients usually complain of the acute onset of severe odynophagia. Radiographic or endoscopic findings may range from discrete ulceration to diffuse esophagitis. Treatment is generally supportive with discontinuation of the medication until the injury resolves. Although acid suppression is commonly recommended, there is no proof that this approach is beneficial. Patients should be given careful instructions to avoid lying down immediately after ingesting medication and to drink adequate fluids to prevent injury. Rarely, pill-induced injury may lead to strictures and even fistulas.

Caustic Injury

Potentially devastating caustic esophageal injuries may be caused by highly alkaline solutions, such as sodium hydroxide, or highly acidic solutions, such as sulfuric acid. The most common products that contain these substances are drain cleaners and industrial strength cleaners, but other corrosive substances include hair relaxers, oven and toilet bowl cleaners, and button batteries. Patients require emergent endoscopy to determine the degree of injury, which helps predict long-term prognosis. No clear evidence supports routine steroids or antibiotics. Many patients will have lifelong disease marked by chronic strictures that require frequent dilation and even esophageal reconstruction. In patients with a severe initial injury, the risk for esophageal cancer is significantly increased.

TABLE 138-3 MEDICATIONS COMMONLY ASSOCIATED WITH ESOPHAGITIS OR ESOPHAGEAL INJURY

ANTIBIOTICS
Tetracycline
Doxycycline
Clindamycin
Penicillin
Rifampin
ANTIVIRAL AGENTS
Zalcitabine
Zidovudine
Nelfinavir
BISPHOSPHONATES
Alendronate
Etidronate
Pamidronate
CHEMOTHERAPEUTIC AGENTS
Dactinomycin
Bleomycin
Cytarabine
Daunorubicin
5-Fluorouracil
Methotrexate
Vincristine
NONSTEROIDAL ANTI-INFLAMMATORY DRUGS
Aspirin
Naproxen
Ibuprofen
OTHER MEDICATIONS
Quinidine
Potassium chloride
Ferrous sulfate
Ascorbic acid
Multivitamins
Theophylline

ESOPHAGEAL MOTOR DISORDERS

Oropharyngeal Dysfunction

The cricopharyngeus and inferior pharyngeal constrictor are composed of striated muscle and innervated by upper motor neurons, the brain stem, and the cerebral cortex. Both primary myopathic and neuropathic disorders can result in dysfunction. The most common neurologic cause of oropharyngeal dysphagia is a cerebrovascular accident (Chapter 407). Other neuropathic disorders that may also affect function include myasthenia gravis (Chapter 422), brain stem tumors (Chapter 189), amyotrophic lateral sclerosis (Chapter 419), Parkinson disease (Chapter 409), Alzheimer disease (Chapter 402), postpolio syndrome (Chapter 379), Guillain-Barré syndrome (Chapter 420), and botulism (Chapter 296). Myogenic disorders that cause dysfunction include paraneoplastic antibody-mediated syndromes (Chapter 179), thyroid disease (Chapter 226), primary myopathies (Chapter 421) such as dermatomyositis and inclusion body myositis, and drugs that cause myopathy such as statins and amiodarone.

Patients with oropharyngeal abnormalities experience dysphagia, often accompanied by postprandial coughing, hoarseness, and aspiration pneumonia. Treatment focuses on the underlying myopathic or neurologic cause and swallowing therapy, but prognosis is often poor owing to limited treatment options.

Achalasia

Achalasia, which is the prototypic esophageal motility disorder, is characterized by insufficient relaxation of the lower esophageal sphincter accompanied by loss of esophageal peristalsis.[14]

EPIDEMIOLOGY AND PATHOBIOLOGY

Achalasia can occur in patients of almost any age, from infants to nonagenarians, but it most commonly presents between 30 and 60 years of age. The prevalence is 10 per 100,000 in the United States, with all races affected and an equal distribution in men and women. The pathophysiology of achalasia most likely reflects an antibody-mediated autoimmune myenteric plexopathy in the lower esophageal sphincter and a generalized neuropathy in the esophageal body. The triggering event is unclear, but a viral cause is suggested. Injury to the lower esophageal sphincter neurons leads to a relative selective deficiency of nitric oxide. With this loss of the main functional inhibitory neurotransmitter, the sphincter loses its ability to relax. The neurochemical process that leads to aperistalsis is unclear.

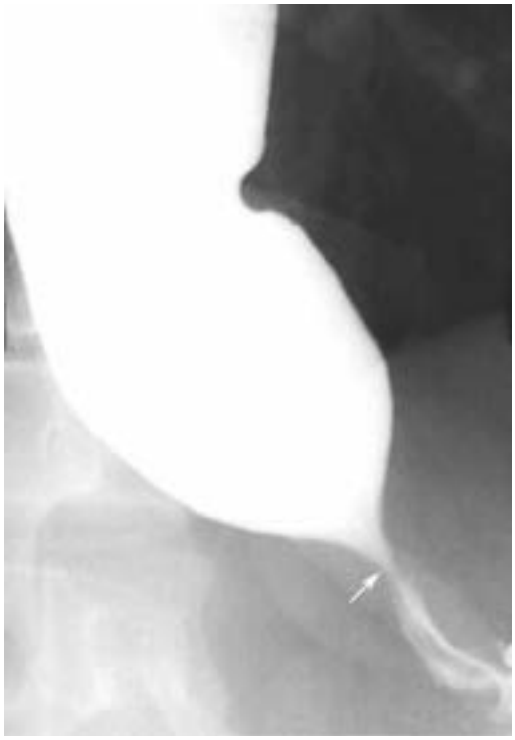

FIGURE 138-7. Esophagogram of a patient with idiopathic achalasia. Note the dilated esophagus with an air-fluid level and distal tapering providing a "bird's beak" deformity in the area of the lower esophageal sphincter. (Courtesy Marc Levine, MD.)

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The cardinal symptoms of achalasia are dysphagia to both liquids and solids, regurgitation, and chest pain. Some patients may have more subtle symptoms, including heartburn, presumably caused by esophageal stasis of acidic food content, weight loss, and aspiration pneumonia; in these settings, diagnosis is often delayed.

The diagnosis of achalasia relies on esophageal manometry and barium radiography. The classic radiographic appearance is esophageal dilation, stasis of contrast material, and a "bird's beak" appearance to the lower esophageal sphincter (Fig. 138-7).[15] Manometry demonstrates high residual pressures of the lower esophageal sphincter and either simultaneous contractions or complete absence of peristaltic contractions. Whether classification systems based on manometric findings can predict response is being studied. Endoscopy is recommended to exclude secondary causes of achalasia, such as gastroesophageal junctional cancer. Occasionally, patients present with a massively dilated esophagus and marked food retention. Radiographically, these patients develop a "sigmoid" esophagus with a radiographic picture similar to a sigmoid colon.

Achalasia may represent a paraneoplastic presentation of some malignancies, particularly small cell lung cancer (Chapter 191); in these patients, the tumor produces an antineuronal antibody (anti-Hu) that mediates the autoimmune lower esophageal sphincter plexopathy and produces a syndrome identical to primary achalasia. Some tumors, such as proximal gastric cancer, may also metastasize to or directly extend into the lower esophageal sphincter and produce an achalasia-like picture, possibly owing to extrinsic compression or tumor infiltration.

TREATMENT

Rx

Treatment options to decrease the functional obstruction at the level of the lower esophageal sphincter include the intrasphincteric injection of botulinum toxin, pneumatic dilation, and surgical myotomy. Injection of botulinum

toxin during endoscopy reduces symptoms and improves esophageal emptying in up to 90% of patients. Because symptoms typically recur within 6 to 24 months, this treatment is best for patients who are not candidates for more definitive therapies or to confirm the diagnosis of achalasia when clinical, radiographic, and manometric criteria are not conclusive.

In pneumatic dilation, a 30- to 40-mm pneumatic balloon is placed fluoroscopically to straddle the lower esophageal sphincter; the balloon is then inflated to tear the muscle fibers of the lower esophageal sphincter. In general, one pneumatic dilation will achieve 5 years of symptomatic remission in 70% of patients, and three dilations will succeed in 90% of patients. The downside of this procedure is the risk for perforation, which occurs in up to 2% of patients, even in experienced hands.

The third approach is a Heller myotomy, which is now typically performed laparoscopically. This long myotomy starts at least 2 cm below the lower esophageal sphincter and extends for about 6 cm upward past the sphincter; a loose fundoplication is performed to prevent gastroesophageal reflux. The 5-year success rate approaches 90%. Randomized controlled trials suggest that either pneumatic dilation or Heller myotomy will provide comparable clinical results over 2 years.

A new endoscopic approach known as peroral endoscopic myotomy (referred to as POEM) is an alternative to laparoscopic myotomy.[16] Early results are encouraging, but this approach is not yet standard clinical practice.

In patients with massive dilation, esophageal dysfunction may warrant total esophagectomy because of life-threatening symptoms such as continued weight loss, recurrent aspiration pneumonia, or tracheal compression. In patients with underlying paraneoplastic malignancy, treatment of the tumor may be helpful, and botulinum toxin has been tried with anecdotal success.

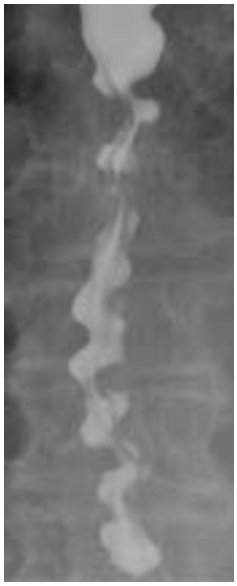

FIGURE 138-8. Barium esophagogram showing a "corkscrew" esophagus in a patient with diffuse esophageal spasm. The patient had dysphagia, chest pain, and normal endoscopic findings. (Image courtesy of Marc Levine, MD.)

PROGNOSIS

No treatment approach is curative, and all are palliative. Recurrent symptoms may be related to an incomplete myotomy, herniation or unwrapping of the fundoplication, esophageal strictures, Barrett esophagus, or just the natural history of the disease. Achalasia patients may also be predisposed to squamous cell carcinoma of the esophagus, although current evidence does not support cancer screening.

Diffuse Esophageal Spasm

Diffuse esophageal spasm is found in fewer than 5% of patients who undergo manometry for symptoms of chest pain, dysphagia, or both. The pathophysiology of diffuse esophageal spasm is not well understood, but a deficiency of nitric oxide in the esophageal body may lead to a loss of control of esophageal peristalsis, high pressures, and rapid velocity contractions.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Classically, patients have symptoms of intermittent chest pain (Chapter 137), dysphagia, or both. On endoscopic ultrasound, patients may demonstrate thickening of the circular and longitudinal muscle layers. Diffuse esophageal spasm is defined manometrically by premature rapid contractions in at least 20% of all swallows, accompanied by normal peristalsis. Radiographically, diffuse esophageal spasm is characterized by a "corkscrew" esophagus with multiple simultaneous contractions that obliterate the lumen (Fig. 138-8).

TREATMENT AND PROGNOSIS

Rx

Treatment of these patients is challenging, and clinical trials have not demonstrated efficacy of any therapy. Empirical therapy may be tried with agents that relax smooth muscle or augment the nitric oxide content, such as hyoscyamine (0.125 mg sublingually), calcium-channel antagonists (nifedipine, 10 mg sublingually), nitroglycerin (0.3 mg sublingually), and sildenafil (50 mg orally). Injection of botulinum toxin into the esophageal body has had equivocal results. Antidepressants, particularly low-dose tricyclics (e.g., imipramine, 10 to 50 mg before bedtime), have had some success. For patients with dysphagia, sphincter options include botulinum toxin injection and pneumatic dilation, but surgery is not indicated unless there is evidence of achalasia. In general, these disorders can be difficult to manage but are not life-threatening.

Other Motility Disorders

Jackhammer esophagus and hypertensive peristalsis are recently described manometric diagnoses that formerly were called nutcracker esophagus. Both are characterized by high-amplitude peristaltic contractions and are of uncertain clinical significance. Antisecretory therapy with proton pump inhibitors (see Table 138-1) is often warranted.

A hypertensive lower esophageal sphincter is of unknown significance, but isolated incomplete relaxation of the lower esophageal sphincter may represent an achalasia variant and should be treated as such in patients with dysphagia. Patients with manometric findings of low-amplitude contractions, failed contractions, or contractions with large breaks have weak peristalsis, which is commonly seen in patients with severe underlying GERD.

STRUCTURAL ABNORMALITIES

Cricopharyngeal Bars

A cricopharyngeal bar, which is caused by a prominent impression of the cricopharyngeus, reflects an inability to relax the upper esophageal sphincter maximally during the flow of barium, with a sustained decrease in compliance. The decrease in cross-sectional area of the upper esophageal sphincter is likely caused by fibrosis. This condition may be asymptomatic or accompanied by symptoms of oropharyngeal dysphagia to solids. Treatment of symptomatic patients is usually surgical, and the response is generally excellent.

Esophageal Diverticula

Esophageal diverticula are encountered in fewer than 1% of upper gastrointestinal radiographic studies and account for less than 5% of cases of dysphagia. Esophageal diverticula may occur in one of three locations: above the upper esophageal sphincter, in the mid-esophagus, and just above the lower esophageal sphincter.

Zenker diverticulum is a pouch that protrudes posteriorly above the upper esophageal sphincter (Fig. 138-9). This protrusion occurs through a triangular region known as Killian triangle, which is bordered above by fibers of the inferior constrictor muscle and below by the cricopharyngeal muscle. It is thought to be caused by increased hypopharyngeal pressure that results from decreased compliance and impaired opening of the upper esophageal sphincter. Small diverticula may be asymptomatic, but increasing size is associated with globus, dysphagia to solids or liquids, regurgitation of undigested food, halitosis, and aspiration. Although no treatment is needed for asymptomatic patients, open or endoscopic surgery is indicated when symptoms occur. Prognosis after surgery is excellent.

Mid-esophageal diverticula, which are focal outpouchings of the middle of the esophagus, are thought to be related to an underlying abnormality of esophageal motility. Past tuberculosis disease may result in such diverticula. Symptoms typically include dysphagia with or without regurgitation, and diagnosis is usually made with barium contrast radiography. Surgical diverticulectomy with myotomy is reserved for patients with symptoms.

Epiphrenic diverticula are herniations of mucosa and submucosa through the muscular layers of the distal 10 cm of the esophagus. They are most

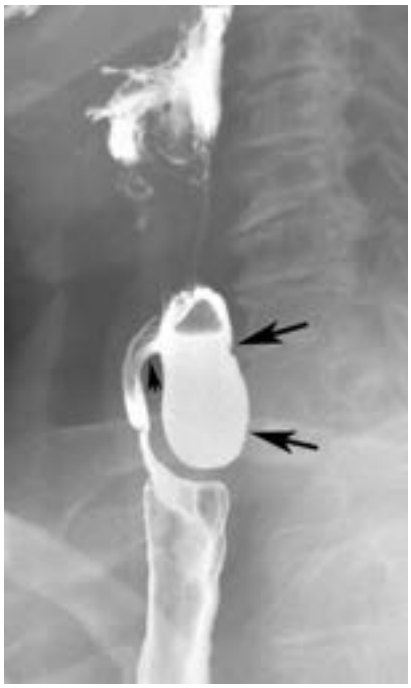

FIGURE 138-9. Barium radiograph of Zenker diverticulum *(arrow)* and cricopharyngeus *(arrowhead)*. (Courtesy Marc Levine, MD.)

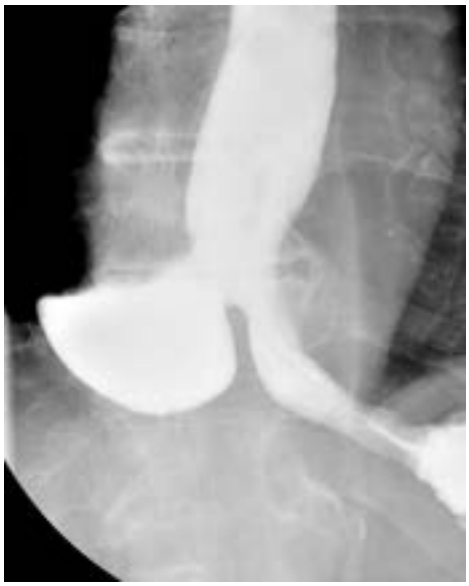

FIGURE 138-10. Barium radiograph of an esophageal epiphrenic diverticulum. (Courtesy Marc Levine, MD.)

FIGURE 138-11. Barium radiograph of Schatzki ring. (Courtesy Marc Levine, MD.)

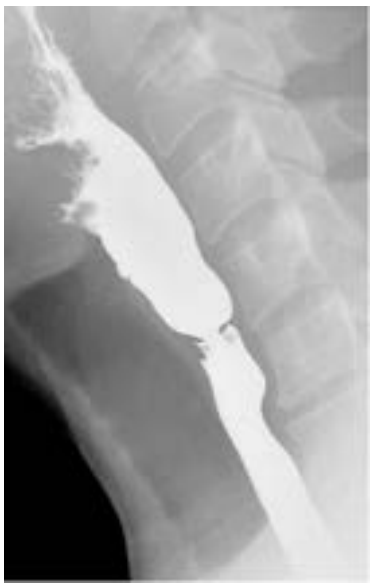

FIGURE 138-12. Barium radiograph of an esophageal web. (Courtesy of Marc Levine, MD.)

commonly caused by functional lower esophageal sphincter obstruction, owing to underlying motility abnormalities such as achalasia or diffuse esophageal spasm, but may also be associated with mechanical obstruction, owing to a leiomyoma, prior surgery, stenosis, stricture tumor, or web. Epiphrenic diverticula may be asymptomatic or cause dysphagia, regurgitation, odynophagia, chest pain, heartburn, or aspiration. Diagnosis is typically made by barium radiography (Fig. 138-10), but endoscopy is recommended to exclude a structural cause of obstruction, and esophageal manometry is recommended to evaluate for underlying motility abnormalities. No therapy is warranted in asymptomatic patients, but symptomatic patients generally do well with surgical diverticulectomy, repair of the defect in the esophageal wall, and relief of the underlying obstruction, usually with a myotomy and partial fundoplication.

Rings and Webs

Esophageal rings are concentric areas of narrowing, usually in the distal esophagus. *Schatzki ring* (Fig. 138-11) is a thin, fixed, circumferential membrane-like narrowing at the gastroesophageal junction, typically at the proximal border of a hiatal hernia. The cause may be congenital, secondary to a pleat of redundant mucosa, or related to gastroesophageal reflux. Symptoms typically occur if the ring results in a lumen that is 13 mm or less in diameter. Classic symptoms are intermittent dysphagia to solids or impaction of solid food. Diagnosis is best accomplished with barium radiography, especially for rings larger than 13 mm in diameter, which may be missed at the time of endoscopy. Treatment involves use of large-caliber dilators at least 18 mm in diameter. Long-term therapy with proton pump inhibitors (see Table 138-1) may prevent relapses. Most patients relapse after a single dilation.

Muscular rings, which are located several centimeters above the squamocolumnar junction, are composed of mucosa, submucosa, and muscle. Barium radiography and endoscopy reveal a focal constriction of variable diameter. Optimal therapy is unclear. Dilation leads to only partial or temporary relief. Other treatment modalities include injection of botulinum toxin and anticholinergic agents.

Esophageal webs are thin, eccentric, membranous areas of narrowing that may be found anywhere in the esophagus but most commonly are in the proximal region (Fig. 138-12). The pathogenesis of webs is unknown, but

they are associated with a number of systemic diseases, including bullous skin disorders, chronic graft-versus-host disease, and iron deficiency anemia.

Hiatal Hernia

A hiatal hernia involves herniation of elements of the abdominal cavity through the diaphragmatic hiatus. A sliding or type I hernia, in which the gastroesophageal junction is displaced above the diaphragmatic hiatus, is the most common type. Type I hiatal hernias typically are not associated with symptoms or are associated with heartburn or acid regurgitation. Treatment is the same as for GERD.

In a type II or true paraesophageal hernia, which is uncommon, the gastroesophageal junction is in its normal location, but the fundus and parts of the greater curvature of the stomach herniate into the mediastinum alongside the esophagus. With type III or mixed paraesophageal hernia, the gastroesophageal junction and a large part of the stomach herniate into the mediastinum. Both types of paraesophageal hernias present with symptoms of postprandial distress, such as epigastric pain, chest pain, substernal fullness, shortness of breath, nausea, or vomiting. Iron deficiency anemia may be seen with large hiatal hernias in which at least one third of the stomach is in the chest; linear gastric erosions at the top of gastric folds at the level of the diaphragm have been implicated as a cause of chronic blood loss. Asymptomatic paraesophageal hernias do not require surgery. Symptomatic paraesophageal hernias warrant surgical therapy because of the risk for strangulation, bleeding, perforation, or obstruction.

THE ESOPHAGUS IN SYSTEMIC DISEASES

Scleroderma

Up to 90% of patients with scleroderma (Chapter 267) have esophageal involvement, thereby making it the most common gastrointestinal abnormality in the disease. The classic manometric and radiologic findings are aperistalsis of the distal two thirds of the esophagus and hypotensive or patulous lower esophageal sphincter on manometry or radiography, respectively. These findings result initially because of neuropathy and later because of the myopathy. The main clinical manifestation is severe gastroesophageal reflux. Delayed gastric emptying may also occur. Marked esophageal stasis may also lead to *Candida* esophagitis. A patulous (on radiography) or hypotensive (on manometry) lower esophageal sphincter strongly supports the diagnosis. Proton pump inhibitor therapy (see Table 138-1) is the cornerstone of treatment. Lifestyle changes, including small frequent meals and avoiding nighttime meals, may reduce the gastric symptoms. Antireflux surgery can exacerbate dysphagia by creating a functional high-pressure zone in the distal esophagus and is thus contraindicated. Unfortunately, many patients suffer from chronic GERD and are at risk for developing Barrett esophagus, recurrent strictures, and adenocarcinoma.

Amyloidosis

Amyloidosis (Chapter 188) may lead to smooth muscle and autonomic nervous system dysfunction that involves the esophagus in a pattern similar to scleroderma. Patients will have both dysphagia and severe reflux because of esophageal aperistalsis and a hypotensive lower esophageal sphincter. Dysphagia may result not only from the motility changes but also from diffuse esophageal rigidity and loss of compliance owing to amyloid infiltration of the esophageal wall. Rarely, an achalasia-like pattern may develop. No treatment improves the esophageal manifestations of amyloidosis, other than treatment for the underlying disease and high-dose proton pump inhibitors given twice daily.

Other Systemic Diseases

Dermatomyositis (Chapter 269) principally involves the striated muscle of the oropharynx and proximal esophagus, but sometimes the distal esophagus may lose normal peristaltic function. Symptoms can be oropharyngeal or esophageal depending on the site of greatest muscular involvement. Dysphagia may be an early presenting symptom of dermatomyositis, and the esophagus may be involved in 10 to 50% of patients. Treatment is directed at the generalized myositis because there are no specific treatments for the esophagus. Swallowing therapy may sometimes be helpful. Return of swallowing commonly lags behind recovery of other striated muscle symptoms.

Esophageal involvement in *systemic lupus erythematosus* (Chapter 266) is not as prominent as in scleroderma, dermatomyositis, or mixed connective tissue disease. Dysphagia occurs in fewer than 15% of patients and may be caused by decreased salivation from secondary Sjögren syndrome or reduced esophageal peristalsis. Other causes of esophageal symptoms in patients with systemic lupus erythematosus include GERD, esophageal ulcers, esophageal infection, and medication-induced esophagitis. *Behçet disease* (Chapter 270) causes esophageal ulcers in less than 15% of affected patients. Ulcers are typically located in the middle third of the esophagus and are often associated with ulcers in the stomach, ileum, or colon. Rare esophageal lesions include strictures, varices, and fistulas connecting with the trachea. These complications can be severely debilitating.

Esophageal involvement in *Crohn disease* (Chapter 141) is uncommon but has been described in up to 1 to 2% of patients. Occasionally, isolated esophageal disease may occur. Ulceration is the most common manifestation, but strictures, fissures, esophagobronchial fistulas, mediastinal abscesses, and aphthoid lesions have been described. Patients typically complain of dysphagia and odynophagia. Treatment and prognosis are as for the underlying Crohn disease.

In the *Ehlers-Danlos syndrome* (Chapter 260), hiatal hernias are common, and structural esophageal defects such as giant epiphrenic diverticula, megaesophagus, and spontaneous esophageal rupture may also be seen.

Skin Disorders That Involve the Esophagus

Lichen planus (Chapter 438) involves the esophagus. Patients generally present in the fourth to seventh decade of life. Histologically, the lesions in up to 25 to 50% of patients show a characteristic lymphohistiocytic inflammatory infiltrate and apoptotic basal keratinocytes known as Civatte bodies. Lesions classically occur on the buccal mucosa and the tongue, but the disease may also involve other areas of the oral cavity and the esophagus, conjunctiva, nose, larynx, stomach, bladder, and anus. Most patients are asymptomatic or have only minor symptoms, but patients with severe esophageal lichen planus often develop strictures and may present with dysphagia, odynophagia, and weight loss. Strictures are typically proximal but may be variable in length and location, sometimes involving most of the esophageal body. Patients may present with esophageal lichen planus in the absence of extraesophageal disease. Endoscopic findings, which can be subtle and nonspecific, include peeling mucosa, hyperemic focal abnormalities, and submucosal plaque or papules. High-dose systemic steroids, starting with at least 40 mg prednisone and then tapering over 1 to 2 months, are often successful, but relapse is common when steroids are tapered. Topical steroids have also been tried. Dilation and intralesional steroids can alleviate the symptoms associated with strictures, but symptoms frequently recur in less than a year and necessitate repeat dilations. The esophageal lesions of lichen planus may rarely have malignant potential.

Pemphigus vulgaris (Chapter 439) patients who experience acute flares may have upper gastrointestinal symptoms (dysphagia, odynophagia, or retrosternal burning) in 80% of cases and biopsy-proven esophageal pemphigus lesions in nearly 50% of cases. Pemphigus vulgaris may rarely be isolated to the esophagus. Paraneoplastic pemphigus (Chapter 439), most commonly reported with lymphoreticular disease, also may involve the esophagus. Like lesions on the skin, esophageal lesions may be flaccid blisters or erosions, but they may also appear as red longitudinal lines along the entire esophagus. Like cutaneous and oral pemphigus, esophageal pemphigus is generally treated with corticosteroids.

In *mucous membrane pemphigoid* (cicatricial pemphigoid, Chapter 439), the esophagus is the most common site of gastrointestinal involvement, but esophageal disease occurs in less than 15% of patients. The esophageal disease can present as many as 10 years after disease onset and may be the only presentation of the disease. Patients complain of dysphagia and odynophagia. Imaging typically shows erosions, strictures, bullae, or webs. Treatment generally is as for the skin disease.

In *dystrophic epidermolysis bullosa* (Chapter 439), approximately 70 to 95% of patients develop esophageal stenosis or strictures. Strictures are especially common in children with recessive dystrophic epidermolysis bullosa, whereas fewer than 10% of patients with dominant dystrophic epidermolysis bullosa develop strictures by age 50 years.

ESOPHAGEAL INFECTIONS

Herpes Simplex Virus

Herpes simplex virus (HSV; Chapter 374) esophagitis, which is usually caused by HSV-1, is a well-recognized disease in immunocompromised patients but may also occur in immunocompetent hosts. HSV typically presents as severe odynophagia. Endoscopy and radiology characteristically demonstrate multiple ulcers and friable mucosa, predominantly involving the

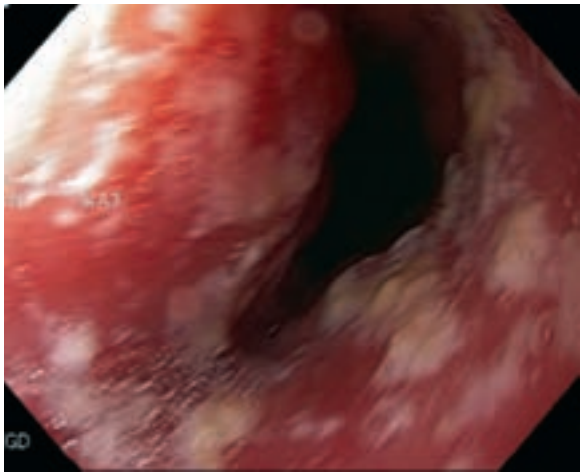

FIGURE 138-13. Candidal esophagitis.

distal esophagus. Mucosal biopsies demonstrate typical intranuclear inclusion bodies. Depending on the severity of the esophageal disease and the underlying health of the patient, treatment options may include intravenous acyclovir, 5 mg/kg every 8 hours for 7 days, or oral acyclovir, 800 mg five times daily for 7 days, combined with symptomatic treatment.

Candidiasis

Esophageal candidiasis is common in patients with HIV infection or with impaired cellular immunity owing to hematologic malignancies, immunosuppressive therapy, or diabetes mellitus. Esophageal candidiasis is also occasionally seen in immunocompetent patients with marked esophageal stasis, such as patients with advanced achalasia or scleroderma. Esophageal candidiasis characteristically presents with symptoms of odynophagia, dysphagia, and chest pain. Endoscopy demonstrates scattered or coalescent yellow-white mucosal plaques (Fig. 138-13). Given the high prevalence of esophageal candidiasis in HIV patients, some recommend treating symptomatic HIV patients empirically and reserving endoscopy for refractory symptoms. Mild oropharyngeal candidiasis may be treated with topical clotrimazole (10 mg troche five times daily for 7 to 14 days) or nystatin (600,000 units four times daily for 7 to 10 days), whereas oral fluconazole (100 mg daily for 7 to 14 days) is needed for moderate to severe oropharyngeal and esophageal candidiasis. Intravenous fluconazole or amphotericin B deoxycholate (Chapter 331) are appropriate options for severe esophageal candidiasis or for patients who cannot tolerate oral therapy.

Cytomegalovirus

Cytomegalovirus (CMV; Chapter 370) infection of the esophagus is exclusively found in immunocompromised patients. Patients infected with the HIV virus with low CD4 counts are most commonly affected, but CMV also occurs in transplant recipients on immunosuppressive therapy and in immunosuppressed patients with malignancy. Patients typically have severe odynophagia with evidence of radiographic or endoscopic esophageal ulcers. Treatment is usually with intravenous ganciclovir (5 mg/kg once or twice daily for 10 to 14 days) or foscarnet (90 mg/kg every 8 to 12 hours until healing occurs).

Bacterial Esophagitis

Bacterial infection of the esophagus is uncommon but can occur in immunosuppressed patients, typically those with neutropenia and malignancy. Patients present with chest pain or odynophagia, or both, and endoscopy reveals extensive erosions, usually in the distal esophagus. Biopsy and Gram stain reveal acute and chronic inflammation and bacteria, commonly gram-positive organisms, especially Viridans-group streptococci, *Staphylococcus aureus*, *Staphylococcus epidermidis,* and *Bacillus* species. Treatment is with the appropriate antibiotics administered parenterally.

Human Papillomavirus

Esophageal infections with human papillomavirus (HPV; Chapter 373) are typically asymptomatic. HPV lesions are most frequently found in the mid to distal esophagus as erythematous macules, white plaque, nodules, or exuberant frond-like lesions. The diagnosis is made by histologic demonstration of koilocytosis (an atypical nucleus surrounded by a ring), giant cells, or immunohistochemical stains. Treatment is usually not necessary, although large lesions may require endoscopic removal. Other treatments, such as interferon, bleomycin, and etoposide, have yielded varying results. HPV infection is a risk factor for esophageal squamous cell carcinoma (Chapter 192), but the value of endoscopic surveillance is not known.

MISCELLANEOUS ESOPHAGEAL CONDITIONS

Esophageal Emergencies

BOERHAAVE SYNDROME

Boerhaave syndrome, or spontaneous rupture of the esophagus, is a transmural full-thickness tear of the esophageal wall. A sudden rise in intraesophageal pressure during forceful vomiting is the cause in most cases. The tear is most commonly in the lower third of the esophagus, 2 to 3 cm proximal to the gastroesophageal junction.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The classic presentation is vomiting, lower thoracic pain, and subcutaneous emphysema. Other findings may include pleural effusions, especially left sided, tachypnea, abdominal rigidity, fever, and hypotension. Because the condition is rare and classic antecedent vomiting is not always reported, rupture is often recognized only after the development of mediastinitis.

The chest radiograph may demonstrate mediastinal widening, a unilateral pleural effusion, hydropneumothorax, and pneumomediastinum. The esophagram, typically with a water-soluble agent, reveals extravasation, although false-negative results may be encountered. Computed tomography (CT) scanning with an oral contrast agent, which is perhaps the best diagnostic option, typically demonstrates air in the mediastinum.

TREATMENT AND PROGNOSIS

Rx

Successful therapy depends on early recognition and the underlying condition of the patient. The classic approach is operative repair in conjunction with broad-spectrum antibiotics and nutritional support, especially if the diagnosis is made within the first 24 hours. Aggressive conservative therapy with percutaneous drains, broad-spectrum parenteral antibiotics, and nutrition has also been successful. In selected individuals, endoscopic insertion of self-expanding covered metallic stents is an option. The mortality rate can be as high as 100% without treatment.

MALLORY-WEISS TEAR

A Mallory-Weiss tear (see Fig. 135-2) is a mucosal tear, often at the gastroesophageal junction, usually caused by severe vomiting or retching. The syndrome is more common in alcoholic patients. In about 85% of patients, it is associated with acute upper gastrointestinal bleeding, which requires transfusion in about 70% of patients and urgent intervention in about 10% of cases. Mortality is about 5%, similar to the mortality in severe bleeding ulcers.[17] Most tears heal spontaneously within about 48 hours.

IATROGENIC PERFORATION

Esophageal perforation may occur as a result of a variety of iatrogenic causes, including endoscopy, dilation, endosonography, or surgery as well as with foreign body ingestion. Typical symptoms include chest pain with or without abdominal pain, subcutaneous emphysema, and fever. Diagnosis may be made immediately at the time of endoscopy or radiography. Chest radiographs may show pneumomediastinum, subcutaneous emphysema, hydrothorax, or hydropneumothorax. A Gastrografin swallow may show a leak, and chest CT may show mediastinal air. Acute perforations are life-threatening emergencies that warrant immediate closure before contamination of the mediastinum. For acute perforation caused by endoscopy, endoscopic clips or stents are an option. Conservative therapy also is an option for well-contained perforations with minimal mediastinal, pleural, or peritoneal contamination and no obvious signs of sepsis. Prognosis is generally good if the perforation is recognized and treated early, preferably within 12 to 24 hours.

FOREIGN BODIES

Impaction of food or foreign bodies is a common gastrointestinal emergency. Meat impaction is most commonly seen in adults, but other foreign objects

that also may be ingested accidentally or by design include batteries, coins, and bones. Many patients have an underlying esophageal abnormality, such as Schatzki ring, stricture, eosinophilic esophagitis, tumor, or achalasia. The clinical presentation includes dysphagia, odynophagia, foreign body sensation, chest pain, excessive salivation, and difficulty handling secretions. Immediate endoscopic extraction with concomitant airway protection is recommended because prolonged impaction may result in penetration into the esophageal wall followed by perforation and mediastinitis.[18]

ESOPHAGEAL VARICES

Esophageal varices are described in Chapters 135 and 153.

ESOPHAGEAL FISTULA

Tracheoesophageal fistulas may arise from a variety of different causes, including trauma, infectious esophagitis, necrosis after prolonged endotracheal intubation, esophageal cancer, and radiation therapy. Patients are often critically ill and may not tolerate surgery. Metallic stents are a treatment option.

Atrial-esophageal fistulas are a rare complication after radiofrequency ablation (Chapter 66) for atrial fibrillation. The clinical presentation is nonspecific and includes features such as dysphagia, fever, leukocytosis, bacteremia, massive intestinal bleeding, and septic shock. Diagnosis requires a high index of suspicion and is aided by a chest CT scan. Surgery is required; the prognosis is excellent with early recognition and treatment but can be dire if treatment is delayed.

Congenital Abnormalities

ESOPHAGEAL ATRESIA AND TRACHEOESOPHAGEAL FISTULA

Children with congenital esophageal anomalies commonly survive into adulthood after successful surgery. As adults, these patients typically have gastroesophageal reflux that warrants proton pump inhibitor therapy. Strictures at prior surgical anastomoses are common, and patients may require periodic endoscopic dilation.

HETEROTOPIC GASTRIC MUCOSA (INLET PATCH)

Inlet patches, which are areas of gastric columnar epithelium, are present in up to 4.5% of adults and are usually found incidentally at the time of routine endoscopy. Typically, a red columnar patch varying in size from a few millimeters to a few centimeters is seen just below the cricopharyngeus. Inlet patches can be associated with hoarseness, sore throat, and globus sensation. Strictures, ulcers, and rare malignant transformation have also been described. Treatment options include proton pump inhibitors (see Table 138-1) and endoscopic ablation.

OTHER CONGENITAL ESOPHAGEAL DISORDERS

Duplication cysts, which may be located anywhere in the esophagus, may be in continuity with or separate from the esophageal lumen. Symptoms, which may present initially in adulthood, include dysphagia owing to luminal compression, chest pain, and regurgitation. Treatment is surgical.

Dysphagia lusoria is caused by an aberrant right subclavian artery that originates from the right aortic arch and causes partial compression of the esophagus as it crosses over to the left side. Patients may complain of dysphagia. This condition is most commonly detected incidentally during barium radiography, where it is visualized as a crossing diagonal impression at the junction of the proximal and middle thirds of the esophagus. Surgery is rarely indicated because of the complexity of the operation and the difficulty in establishing a clear relationship between the radiographic findings and symptoms.

Grade A References

A1. Sigterman KE, van Pinxteren B, Bonis PA, et al. Short-term treatment with proton pump inhibitors, H2-receptor antagonists and prokinetics for gastro-oesophageal reflux disease-like symptoms and endoscopy negative reflux disease. *Cochrane Database Syst Rev*. 2013;5:CD002095.

A2. Lundell L, Miettinen P, Myrvold HE, et al., for the Nordic GEORD Study Group. Seven-year follow-up of a randomized clinical trial comparing proton-pump inhibition with surgical therapy for reflux oesophagitis. *Br J Surg*. 2007;94:198-203.

A3. Galmiche JP, Hatlebakk J, Attwood S, et al. Laparoscopic antireflux surgery vs esomeprazole treatment for chronic GERD: the LOTUS randomized clinical trial. *JAMA*. 2011;305:1969-1977.

A4. Ramage JI Jr, Rumalla A, Baron TH, et al. A prospective, randomized, double-blind, placebo-controlled trial of endoscopic steroid injection therapy for recalcitrant esophageal peptic strictures. *Am J Gastroenterol*. 2005;100:2419-2425.

A5. Phoa KN, van Vilsteren FG, Weusten BL, et al. Radiofrequency ablation vs endoscopic surveillance for patients with Barrett esophagus and low-grade dysplasia: a randomized clinical trial. *JAMA*. 2014;311:1209-1217.

A6. Shaheen NJ, Sharma P, Overholt BF, et al. Radiofrequency ablation in Barrett's esophagus with dysplasia. *N Engl J Med*. 2009;360:2353-2355.

A7. van Vilsteren FG, Pouw RE, Seewald S, et al. Stepwise radical endoscopic resection versus radiofrequency ablation for Barrett's oesophagus with high-grade dysplasia or early cancer: a multicentre randomised trial. *Gut*. 2011;60:765-773.

A8. Arias A, Gonzalez-Cervera J, Tenias JM, et al. Efficacy of dietary interventions for inducing histologic remission in patients with eosinophilic esophagitis: a systematic review and meta-analysis. *Gastroenterology*. 2014;146:1639-1648.

A9. Straumann A, Conus S, Degen L, et al. Budesonide is effective in adolescent and adult patients with active eosinophilic esophagitis. *Gastroenterology*. 2010;139:1526-1537.

A10. Boeckxstaens GE, Annese V, des Varannes SB, et al. Pneumatic dilation versus laparoscopic Heller's myotomy for idiopathic achalasia. *N Engl J Med*. 2011;364:1807-1816.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bredenoord AJ, Pandolfino JE, Smout AJ. Gastro-oesophageal reflux disease. *Lancet*. 2013;381:1933-1942.
2. Kahrilas PJ, Shaheen NJ, Vaezi MF, et al. American Gastroenterological Association Medical Position Statement on the management of gastroesophageal reflux disease. *Gastroenterology*. 2008;135:1383-1391.
3. Katz PO, Gerson LB, Vela MF. Guidelines for the diagnosis and management of gastroesophageal reflux disease. *Am J Gastroenterol*. 2013;108:308-328.
4. Shaheen NJ, Weinberg DS, Denberg TD, et al. Upper endoscopy for gastroesophageal reflux disease: best practice advice from the clinical guidelines committee of the American College of Physicians. *Ann Intern Med*. 2012;157:808-816.
5. Lam JR, Schneider JL, Zhao W, et al. Proton pump inhibitor and histamine 2 receptor antagonist use and vitamin B12 deficiency. *JAMA*. 2013;310:2435-2442.
6. Spechler SJ, Souza RF. Barrett's esophagus. *N Engl J Med*. 2014;371:836-845.
7. Hvid-Jensen F, Pedersen L, Drewes AM, et al. Incidence of adenocarcinoma among patients with Barrett's esophagus. *N Engl J Med*. 2011;365:1375-1383.
8. Kastelein F, Spaander MC, Steyerberg EW, et al. Proton pump inhibitors reduce the risk of neoplastic progression in patients with Barrett's esophagus. *Clin Gastroenterol Hepatol*. 2013;11:382-388.
9. American Gastroenterological Association, Spechler SJ, Sharma P, et al. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. *Gastroenterology*. 2011;140:1084-1091.
10. Bennett C, Vakil N, Bergman J, et al. Consensus statements for management of Barrett's dysplasia and early-stage esophageal adenocarcinoma, based on a Delphi process. *Gastroenterology*. 2012;143:336-346.
11. Verbeek RE, Leenders M, Ten Kate FJ, et al. Surveillance of Barrett's esophagus and mortality from esophageal adenocarcinoma: a population-based cohort study. *Am J Gastroenterol*. 2014;109:1215-1222.
12. Dellon ES, Gonsalves N, Hirano I, et al. ACG clinical guideline: Evidence based approach to the diagnosis and management of esophageal eosinophilia and eosinophilic esophagitis (EoE). *Am J Gastroenterol*. 2013;108:679-692.
13. Kottyan LC, Davis BP, Sherrill JD, et al. Genome-wide association analysis of eosinophilic esophagitis provides insight into the tissue specificity of this allergic disease. *Nat Genet*. 2014;46:895-900.
14. Vaezi MF, Pandolfino JE, Vela MF. ACG clinical guideline: diagnosis and management of achalasia. *Am J Gastroenterol*. 2013;108:1238-1249.
15. Boeckxstaens GE, Zaninotto G, Richter JE. Achalasia. *Lancet*. 2014;383:83-93.
16. Pescarus R, Shlomovitz E, Swanstrom LL. Per-oral endoscopic myotomy (POEM) for esophageal achalasia. *Curr Gastroenterol Rep*. 2014;16:369.
17. Ljubicic N, Budimir I, Pavic T, et al. Mortality in high-risk patients with bleeding Mallory-Weiss syndrome is similar to that of peptic ulcer bleeding: results of a prospective database study. *Scand J Gastroenterol*. 2014;49:458-464.
18. ASGE Standards of Practice Committee, Ikenberry SO, Jue TL, et al. Management of ingested foreign bodies and food impactions. *Gastrointest Endosc*. 2011;73:1085-1091.

139 ACID PEPTIC DISEASE

ERNST J. KUIPERS AND MARTIN J. BLASER

Acid peptic diseases can involve the esophagus (Chapter 138), the stomach, and the duodenum. Dyspeptic symptoms also can occur in patients who have no endoscopic abnormalities, in whom it is termed *nonulcer dyspepsia* (Chapter 137).

DEFINITIONS

Gastric and duodenal ulcers usually occur in an area of inflamed mucosa. This inflammation, termed *gastritis, duodenitis, or bulbitis*, can sometimes be recognized during endoscopy by signs of edema and erythema of the mucosa, but microscopic evaluation of endoscopic biopsy specimens is required for a definitive diagnosis of mucosal inflammation.

Gastritis is categorized by endoscopic and histologic criteria, with granulocytes predominating in active gastritis and mononuclear cells in chronic gastritis. Gastritis is further classified by the segment of the involved stomach: antral-predominant gastritis, corpus-predominant gastritis, or pangastritis. Finally, the absence or presence of premalignant stages of damage to the mucosa as a result of long-standing inflammation defines the categories of nonatrophic and atrophic gastritis, respectively. Endoscopic findings usually are not specific, unless the gastric mucosa has either a typical miniature cobblestone appearance, termed *nodular gastritis* (a finding particularly in children colonized by *Helicobacter pylori*), or grossly enlarged folds without evidence of cancer, termed *hypertrophic gastritis*.

A *peptic ulcer* is a mucosal defect at least 0.5 cm in diameter that penetrates the muscularis mucosae. Smaller mucosal defects are called *erosions* (Fig. 139-1). Gastric ulcers are subdivided into proximal ulcers, located in the body of the stomach, and distal ulcers, located in the antrum and angulus of the stomach. Gastric ulcers are located mainly along the lesser curvature, in particular at the transitional zone of corpus- to antral-type mucosa. This transitional zone is often in the area of the angulus but may shift proximally. Duodenal ulcers usually are located on the anterior or posterior wall of the duodenal bulb (Fig. 139-2), or occasionally at both sites ("kissing" ulcers). Lesions distal to the duodenal bulb are termed *postbulbar ulcers.* Patients who previously underwent a distal gastric resection (Billroth I or II procedure) can develop ulceration at the gastroduodenal anastomosis (anastomotic ulcer). However, ulcers occurring after Billroth II resection are located predominantly in the jejunal mucosa at the junction between the afferent and efferent loops. Anastomotic ulcers also occur after gastric bypass surgery (Chapter 220), for instance, bariatric Roux-en-Y gastric bypass. Other peptic ulcers can occur at sites of metaplastic or heterotopic gastric mucosa, for example, in Meckel diverticulum, the rectum, or Barrett esophagus. Patients with a large hiatal hernia can develop gastric ulceration, known as *Cameron ulcers*, at the level of the herniation. *Dieulafoy ulcers*, which are small mucosal defects over an intramural arteriole, can lead to severe bleeding. Although these lesions can occur throughout the gastrointestinal tract, two thirds occur in the stomach.

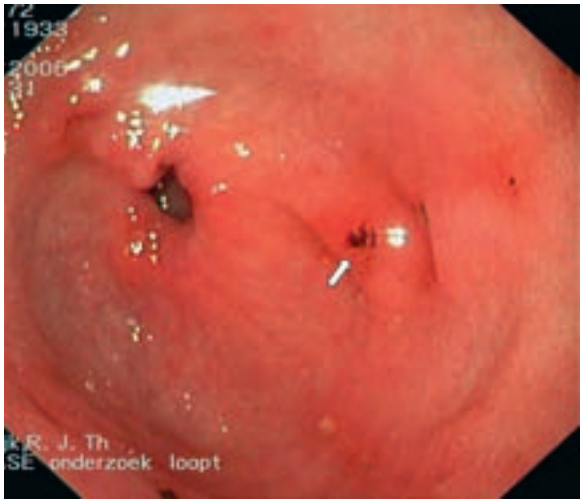

FIGURE 139-1. Endoscopic view of uncomplicated erosive gastritis. The erosion appears as a small, superficial mucosal break with a black base *(arrow)*.

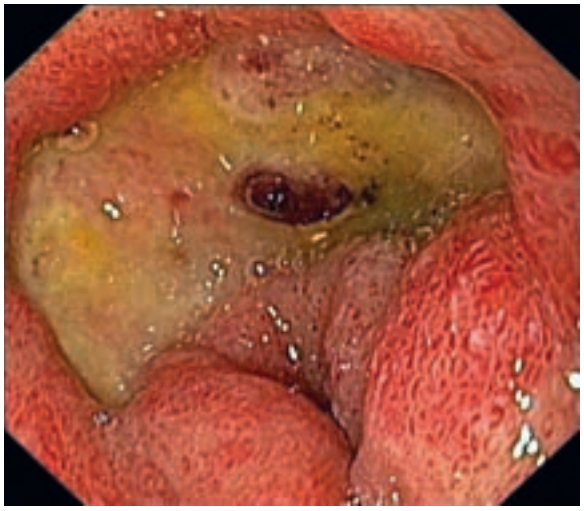

FIGURE 139-2. Endoscopic view of an ulcer at the anterior wall of the duodenal bulb. The ulcer has a clean base, with a visible vessel appearing as a dark red protruding spot close to the lower ulcer rim. The surrounding mucosa is inflamed and swollen.

EPIDEMIOLOGY

The worldwide prevalence of gastritis reflects the prevalence of *H. pylori*. Colonization with this bacterium is virtually always associated with chronic active gastritis, which persists as long as an individual remains colonized and only slowly disappears 6 to 24 months after the eradication of *H. pylori*. Colonization with *H. pylori* usually occurs in the first decade of life and then remains lifelong. In developing countries, high colonization rates result in a high prevalence of *H. pylori* gastritis (often ≥80%) in all age groups, including children. In Western countries, the colonization pressure in children has decreased markedly in recent decades, thereby leading to a birth-cohort phenomenon for the prevalence of *H. pylori* gastritis—currently less than 20% in young adults but 40 to 60% in elderly persons. Some recent studies suggest that this decrease in *H. pylori* prevalence in children has slowed or stopped,[1] owing to factors such as the increasing use of daycare facilities.[2] Nevertheless, the decline of *H. pylori* in developed country populations has been inexorable.

Although peptic ulcer disease is strongly related to *H. pylori* gastritis and duodenitis, the epidemiology of ulcer disease has shown secular variation even when *H. pylori* was ubiquitous. The incidence of peptic ulcer disease rose steeply in Western countries in the late 19th and early 20th centuries and has decreased over the past 40 years; nevertheless, peptic ulceration remains a common disorder. The decline in incidence, associated with a decrease in hospital admissions and surgery for ulcer disease, is believed mostly to reflect the decreasing prevalence of gastric colonization with *H. pylori*. The declining incidence of ulcer disease is also the result of the widespread application of eradication therapy, which strongly reduced recurrent ulcers in *H. pylori*–positive patients. Other factors may include the widespread use of acid suppressive medications. Nevertheless, hospital admissions for complications of ulcers and mortality from ulcer disease have shown a far less marked decline in both the United States and other countries because the reduction in *H. pylori*–associated ulcers in younger persons has been counterbalanced by an increase in ulcers related to nonsteroidal anti-inflammatory drugs (NSAIDs) in older persons.

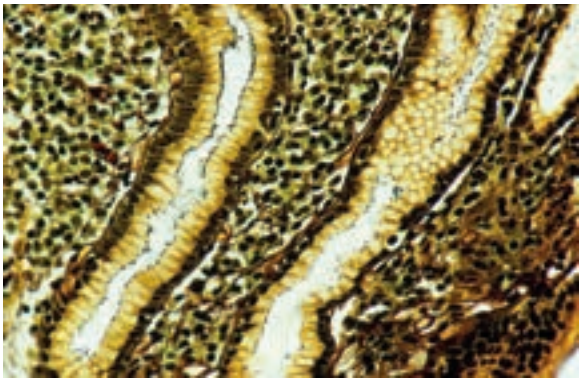

FIGURE 139-3. Gastric mucosa colonized with *Helicobacter pylori* appearing as curved bacilli on the mucosal surface.

In Western countries, duodenal ulcers occur more frequently than gastric ulcers. The predominant age at which duodenal ulcers occur is between 20 and 50 years, whereas gastric ulcers most commonly occur in patients older than 40 years. The incidence of gastroduodenal ulcer disease is approximately 1 to 2 per 1000 inhabitants per year. Although the prevalence of *H. pylori* colonization is nearly identical in males and females, two thirds of patients with ulcers are male. The risk for recurrent disease after initial healing is high; more than 50% of patients have a recurrent ulcer within 12 months of healing in the absence of treatment. Maintenance acid suppressive therapy reduces this recurrence rate, but only therapeutic measures that remove the underlying cause of the ulcer can prevent most ulcer recurrences.

PATHOBIOLOGY

Helicobacter pylori

Most peptic ulcers are associated with colonization with *H. pylori* (Fig. 139-3). Initial clinical studies of the association between *H. pylori* and ulcer disease reported that approximately 85% of patients with gastric ulcer disease and 95% of patients with duodenal ulcer disease were colonized by *H. pylori*. Most persons who are *H. pylori* positive do not have any specific complaints, nor do they develop ulcer disease. The estimated risk for the development of ulcer disease during persistent *H. pylori* colonization is 5 to 15%—that is, three- to eight-fold higher than the risk in patients who are *H. pylori* negative. The risk for the development of an ulcer in the presence of *H. pylori* is determined by a combination of host- and bacteria-related factors. Host factors include immune response, smoking, and stress. A recent genome-wide association study in two independent European cohorts and a subsequent meta-analysis identified a relationship between specific genetic variations in the toll-like receptor (TLR)-1 gene and the prevalence of *H. pylori*.[3] This relationship may in part explain the variation in risk for *H. pylori* colonization and thus the risk for *H. pylori*-associated disease, including peptic ulcer. Bacterial factors that increase the risk for ulcer include a high production of the *VacA* product, which reflects the presence of the s1m1 genotype; a high level of cytokine induction, owing to the presence of genes in the *cag* pathogenicity island; and enhanced adherence, resulting from bacterial *babA* expression.

Nonsteroidal Anti-inflammatory Drugs and Aspirin

The other common cause of gastroduodenal ulcer disease is the use of NSAIDs. At least 2 to 4% of the population in many countries use acetylsalicylic acid, acetic acid derivatives (diclofenac, indomethacin, sulindac), or propionic acid derivatives (ibuprofen, ketoprofen, naproxen) on a daily basis. The risk for ulcer disease is dose and duration dependent. Within 14 days after the start of such treatment, about 5% of patients develop mucosal breaks, that is, erosions and ulcers. In patients who continue therapy for 4 weeks or longer, this proportion increases to 10%, but many are clinically silent. The concomitant presence of *H. pylori* infection increases the incidence of NSAID-related ulcers.

TABLE 139-1 DIFFERENTIAL DIAGNOSIS OF PEPTIC ULCER DISEASE

ORIGIN	CONDITION	FREQUENCY*	DIAGNOSTIC TEST	FINDINGS
Microbes	*Helicobacter pylori*	Very common	*H. pylori* tests Histology	Bacteria, enzymes, antigens, antibodies Gastritis
	Helicobacter heilmannii	Rare	Histology	Spiral bacteria, gastritis
	Treponema pallidum	Very rare	Serology	Antibodies
	Mycobacterial infection	Very rare	Histology, immune response testing, chest radiograph	Acid-fast bacteria, granuloma, immune response, pulmonary infiltrate
	Cytomegalovirus, herpes simplex virus type 1; Epstein-Barr virus	Rare	Histology, serology	Virus inclusions, antibodies
Drug use	NSAIDs, aspirin	Very common	History, urine test	NSAID use
	Bisphosphonates	Rare	History	Bisphosphonate use
	Corticosteroids	Rare	History	Corticosteroid use, comorbidity
	Amphetamines, cocaine	Rare	History, drug testing	Drug use
	Anticoagulants, coagulopathy	Rare	Endoscopy	Ulcer after intramural bleed
Malignancy	Gastric cancer	Common	Histology	Malignancy
	Duodenal cancer	Rare	Histology, CT	Malignancy
	Pancreatic cancer	Common	Histology, CT	Malignancy
	Mucosa-associated lymphoid tissue lymphoma	Rare	Histology	Malignancy
	Metastatic cancer	Rare	Histology	Malignancy
Gastritis syndromes	Eosinophilic gastritis	Rare	Histology	Eosinophilic infiltration
	Lymphocytic gastritis	Rare	Histology, celiac disease screening	Lymphocytic infiltration, villous atrophy
Hyperacidicity syndromes	Zollinger-Ellison syndrome	Rare	Serum gastrin, secretin test	Extreme hypergastrinemia, positive secretin test
	Antral G-cell hyperfunction	Very rare	Serum gastrin, secretin test	Moderate hypergastrinemia, negative secretin test
	Retained gastric antrum	Very rare	Medical history, gastrin	Billroth II resection, hypergastrinemia
	Systemic mastocytosis	Very rare	Histology of affected sites	Mast cell infiltration
	Chronic myelogenous leukemia	Very rare	Leukemia evaluation	Leukemia
Ischemia	Mesenteric vascular occlusion	Common	Angiography	Vascular disease
	Polycythemia vera	Rare	Blood counts	Polycythemia
Specific ulcer types	Cameron ulcer	Common	Endoscopy	Ulcer in large hiatal hernia
	Marginal ulcer	Common	Endoscopy	Ulcer at anastomosis
	Dieulafoy ulcer	Common	Endoscopy	Singular bleeding focus with minimal mucosal disruption
Systemic inflammation	Crohn disease	Common	Histology, ileocolonoscopy	Inflammation, granulomas
	Vasculitides	Rare	Histology, systemic evaluation	Vasculitis, signs of systemic disease
	Gastric amyloidosis	Very rare	Histology	Amyloid deposition
Other conditions	Stress ulcer	Fairly common in patients in intensive care units	Endoscopy	—
	Radiation therapy, chemotherapy	Rare	Endoscopy, history	—

*Frequency as a cause of gastroduodenal ulcer disease.
CT = computed tomography; NSAID = nonsteroidal anti-inflammatory drug.

The risk for developing an ulcer during NSAID use is higher in patients who are older than 60 years; patients with a previous ulcer; patients who use corticosteroids, selective serotonin reuptake inhibitors, or aldosterone antagonists[4]; and patients with major comorbid diseases. In patients who use anticoagulants, such as warfarin and new oral anticoagulants (thrombin and factor Xa inhibitors), or who have severe comorbid disease, an NSAID-induced ulcer is more likely to lead to life-threatening gastroduodenal hemorrhage.

On the basis of their activities, NSAIDs are divided into cyclooxygenase 1 (COX1) and COX2 inhibitors (Chapter 37). The COX1 enzyme is involved in the production of prostaglandins, which play a role in normal cell regulation. The COX2 enzyme, which is also involved in the production of prostaglandins, is induced by inflammatory responses. Most NSAIDs have a nonselective COX inhibitory effect; selective COX2 inhibitors are associated with fewer gastroduodenal ulcers, but their use is limited by adverse coronary effects (Chapter 37). Because of the strong association between NSAIDs and ulcer disease and the risk for recurrence of ulcers with their continued use, patients with ulcers must be thoroughly assessed for any use of NSAIDs.

Helicobacter pylori–Negative, Non-NSAID Ulcer Disease

In most series, *H. pylori* and NSAID use account for 80 to 95% of cases of ulcer disease. The remaining cases are often referred to as *idiopathic* or *H. pylori–negative, non-NSAID* acid peptic disease. The proportion of ulcer disease that is idiopathic is increasing throughout the world as the prevalence of *H. pylori* decreases. Further, it is likely that some ulcers in *H. pylori*–positive patients were not caused by *H. pylori.* Consistent with this notion is the fact that some *H. pylori*–positive patients develop recurrent ulcers after successful bacterial eradication, so presumably their ulcer disease was idiopathic. It is not known whether the increase in idiopathic ulcer disease is simply proportional to the decrease in *H. pylori*–associated ulcers or whether it reflects a true increase in the incidence of idiopathic ulcers.

In patients with idiopathic ulcer disease, specific clues to the underlying cause are often provided by the medical history, including comorbidity and drug use; the endoscopic appearance of the ulcer; and the histologic features of the ulcer's margins and surroundings. In most cases, these initial data can be used to direct further diagnostic studies (Table 139-1).

Malignant Ulcer Disease

Gastroduodenal ulcers can result from underlying malignancies. In the stomach, such tumors are related to gastric adenocarcinoma and, rarely, to mucosa-associated lymphoid tissue (MALT) lymphomas (Chapter 192). Malignant ulcers in the duodenum may result from primary duodenal carcinomas or from penetrating pancreatic cancers. Duodenal cancers have an association with polyposis syndromes, especially familial adenomatous polyposis and, to a much lesser extent, MYH-associated polyposis and Peutz-Jeghers syndrome (Chapter 193). In both the stomach and the duodenum, ulcer disease also may be caused by metastatic tumors, including cancers of the breast, colon, thyroid, or kidney, or by melanoma, disseminated lymphoma, or Kaposi sarcoma. Malignant ulcers are characteristically irregular in shape with heaped borders, but they also may be flat or depressed lesions.

Current high-resolution and magnification endoscopes allow visualization of the altered mucosal structure surrounding an ulcer, including changes in the microvascular pattern. For a definite diagnosis of malignancy, multiple biopsy specimens are needed, usually from the ulcer margins.

Systemic Inflammatory Disorders

A few gastroduodenal ulcers are caused by systemic inflammatory diseases, in particular, Crohn disease (Chapter 141). Patients with Crohn disease affecting the proximal gastrointestinal tract often have multiple ulcers characterized by irregular longitudinal shapes. Ulcers in the duodenum occur on top of Kerckring folds. Patients with gastroduodenal ulcers from Crohn disease do not invariably have evidence of disease elsewhere in the digestive tract, nor do blood tests always suggest an active inflammatory bowel disorder. The demonstration of ulcerative inflammation elsewhere in the digestive tract, in particular in the terminal ileum and colon, strongly supports the diagnosis of Crohn disease, as do noncaseating granulomas on biopsy specimens. However, the absence of granulomas does not exclude Crohn disease, and these lesions are not specific for Crohn disease; they also are associated with *H. pylori* gastritis and other conditions, particularly sarcoidosis (Chapter 95). Sarcoidosis can also lead to gastroduodenal ulcer disease.

Other inflammatory disorders that can cause gastritis or gastroduodenal ulcers include various forms of vasculitis affecting the mesenteric system, in particular Behçet disease (Chapter 270), Henoch-Schönlein purpura (Chapter 270), Takayasu arteritis (Chapters 78 and 270), polyarteritis nodosa (Chapter 270), systemic lupus erythematosus (Chapter 266), Churg-Strauss syndrome (Chapter 270), and granulomatosis with polyangiitis (Chapter 270). Lymphocytic gastroduodenitis, which is strongly associated with celiac disease (Chapter 140), may lead to duodenal ulceration and subsequent stenotic web formation. Ulcer disease also may occur in patients with polycythemia vera (Chapter 166), possibly in relation to reduced mucosal blood flow. Vasculitis underlying ulcer disease should be considered in patients with chronic or recurrent ulceration in whom other causes have been excluded. Lymphocytic phlebitis, which is a rare vasculitic inflammatory disorder that affects the mesenteric veins, may cause gastric ulcers. Systemic amyloidosis (Chapter 188) affecting the stomach wall may lead to gastric ulcers. Rare cases of duodenal ulceration have been described in the presence of annular pancreas or congenital bands obstructing the descending duodenum.

Hypergastrinemic Syndromes

Peptic ulcers can result from chronic gastric hyperacidity related to hypergastrinemia. The most important hypergastrinemic disorder is Zollinger-Ellison syndrome (Chapter 195), a condition of marked hyperacidity leading to severe peptic ulcer disease caused by a gastrin-producing endocrine tumor. These patients usually have multiple bulbar and postbulbar duodenal ulcers that are resistant to conventional acid suppressive therapy. The diagnosis can be confirmed by the presence of a high fasting serum gastrin level (often but not always ≥10-fold increased and >1000 pg/mL). Similar gastrin levels are sometimes seen in patients treated for chronic ulcer disease with high-dose proton pump inhibitors. For clarification, secretin testing can be performed: in patients with Zollinger-Ellison syndrome, the injection of secretin (1 U/kg) increases serum gastrin levels by more than 50%, or 120 pg/mL or greater in those with fasting gastrin levels less than 10-fold above normal. Imaging techniques, such as computed tomography (CT), magnetic resonance imaging, isotope scanning, endoscopic ultrasonography, videocapsule endoscopy, and balloon-assisted enteroscopy, may be used to detect the primary tumor, which is often located in either the pancreas or the proximal small bowel. In some patients, Zollinger-Ellison syndrome occurs as part of the multiple endocrine neoplasia syndrome (Chapter 231), particularly in association with hyperparathyroidism. Other hypergastrinemic hyperacidity syndromes are the retained gastric antrum syndrome (see later) and antral G-cell hyperfunction. In the latter, fasting serum gastrin levels are only modestly increased and do not rise after the injection of secretin, but they respond in an exaggerated way to meals, thereby leading to hyperacidity. When the condition occurs in an *H. pylori*–positive patient, bacterial eradication therapy may be curative. However, some patients with G-cell hyperfunction are *H. pylori* negative.

Ischemia

Stenosis or occlusion of the celiac trunk or the superior mesenteric artery (Chapter 143) also can lead to ulceration in the mucosa of the proximal digestive tract (Fig. 139-4). These ulcers typically occur in elderly patients who have known severe atherosclerosis or risk factors for it, but they can also occur in younger subjects with mesenteric obstruction due to other causes. Ischemic ulcers tend to heal slowly and to recur. Pallor of the mucosa, consistent with decreased mucosal blood flow, may be noted at endoscopy. Upper mesenteric ischemia is often associated with upper abdominal pain, which can be elicited by a meal or by physical activity. These symptoms may cause patients to decrease their food intake, leading to weight loss before their clinical presentation. The prevalence of upper mesenteric ischemia with secondary ulcer disease is unknown, in part owing to its variable presentation, often with a history of gradual symptoms; the lack of standardized and reliable diagnostic tests; and clinicians' unfamiliarity with the condition. Diagnostic evaluation includes a duplex ultrasound scan for vascular flow and conventional or CT angiography of the affected arteries. Validated functional tests for gastroduodenal mucosal perfusion are not widely available, but a technique for directly measuring mucosal oxygen saturation during endoscopy is investigational.

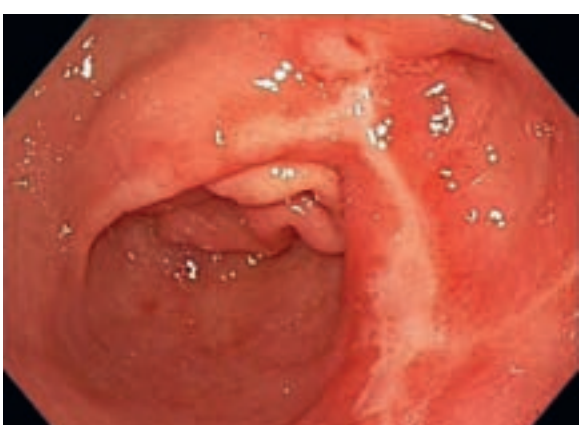

FIGURE 139-4. Endoscopic view of an irregular gastric ulcer at the posterior wall and smaller curvature in a patient with chronic mesenteric ischemia due to subtotal stenosis of the celiac trunk.

Stress Ulcers

Patients with severe medical conditions, such as major trauma, sepsis, extensive burns, head injury, or multiorgan failure, can develop stress ulcers in the stomach or duodenum. Major risk factors for stress ulceration in severely ill patients include mechanical ventilation, coagulopathy, and hypotension, but factors such as hepatic and renal failure and the use of ulcerogenic medications may contribute. Stress ulcers occur independently of *H. pylori* colonization. Ulcers associated with head injury are known as Cushing ulcers, and ulcers associated with extensive burns are known as Curling ulcers. Stress ulcers were once common in patients in intensive care units, but improvements in overall management, including respiratory and hemodynamic care, acid inhibition, and emphasis on adequate feeding, have reduced the incidence of these ulcers, which currently affect 1 to 2% of these patients. Stress ulcers may be asymptomatic, but they can also cause complications, especially bleeding.

Other Factors

Cameron Ulcer

Patients with large hiatal hernias (Chapter 138) may present with proximal gastric ulcers, termed *Cameron ulcers*, at the level of the hiatus, where the stomach is compressed. These ulcers are usually asymptomatic but may cause occult or overt bleeding. During upper gastrointestinal endoscopy, patients with large hiatal hernias and iron deficiency anemia should be carefully examined in normal and retroverted positions for the presence of Cameron ulcers.

Anastomotic or Marginal Ulceration

Patients who have undergone partial gastrectomy sometimes develop recurrent ulcers, often located at the anastomosis or within the jejunum immediately opposite the anastomosis. Ischemia and chronic inflammation in particular resulting from biliary reflux may cause such ulcers. If biopsies exclude cancer, treatment includes acid suppression and *H. pylori* eradication, if needed. Anastomotic ulcers after bariatric gastric surgery appear to be related to local ischemia. They have no clear correlation with *H. pylori*, and

preoperative *H. pylori* eradication does not appear to reduce the incidence of such ulcers. Peptic ulcer disease can be associated also with the retained gastric antrum syndrome when the antrum is not completely excised from the detached duodenum during partial gastrectomy surgery. Because it then lacks exposure to acid and is thus not physiologically downregulated, it continues to secrete gastrin despite normal or even high acid levels. Marginal ulcers can also occur after bariatric Roux-en-Y gastric bypass surgery.

Other Microbial Organisms

Colonization with *Helicobacter heilmannii* (formerly known as *Gastrospirillum hominis*), which is probably a zoonotic organism, is associated with mild gastritis and sometimes with transient ulcer disease. Ulcer disease also may be infectious, resulting from secondary syphilis (Chapter 319), mycobacterial infection (Chapter 324), infection with herpes simplex virus type 1 (Chapter 374), varicella-zoster virus (Chapter 375), cytomegalovirus infection (Chapter 376), or Epstein-Barr virus (Chapter 377).

Alcohol

Short-term heavy alcohol use or long-term moderate to heavy alcohol use can lead to signs of acute and chronic gastritis. No evidence indicates that this type of gastritis is associated with a significant risk for peptic ulceration, although alcohol use increases the risk for bleeding in patients with peptic ulcer disease.

Hyperhistaminic Syndromes

Similar to the hypergastrinemic syndromes, persistent elevation of histamine can lead to hyperacidity as a result of the chronic stimulation of parietal cells. Elevated histamine levels are observed in two rare syndromes. *Systemic mastocytosis* (Chapter 255) is characterized by a proliferation of mast cells in the bone marrow, skin, liver, spleen, and gastrointestinal tract, often associated with both spontaneous and trigger-induced (e.g., alcohol) release of histamine and other vasoactive substances. Patients with systemic mastocytosis often have gastrointestinal symptoms, including pain, diarrhea, and blood loss. Ulceration results from chronic gastric acid hypersecretion. Clues to the diagnosis include symptoms of pruritus, urticaria, or rash. The bone marrow and affected organ mast cell infiltrates carry a specific *C-kit* mutation and express CD2 and CD25. Histamine hypersecretion leading to peptic ulcer disease also can occur in *chronic myelogenous leukemia* (Chapter 184) with basophilia.

Other Drugs

Oral bisphosphonates, used widely for osteoporosis (Chapter 243), may induce gastric erosions and ulcerations in an estimated 3 to 10% of treated patients. The risk for ulcer disease may be synergistically increased by NSAID use but is probably independent of *H. pylori* colonization. Although corticosteroid treatment can be complicated by peptic ulcer disease, the relative risk is only slightly increased, except in patients with serious comorbid diseases, using long-term or high-dose therapy, or with prior ulcers. Other patients who use corticosteroids are not at serious risk for ulcer disease and therefore do not require measures to prevent ulcers. Similarly, the use of aldosterone antagonists is also associated with an increased risk for peptic ulcer and ulcer bleeding, likely related to impaired mucosal healing.

Persons who use amphetamines and crack cocaine (Chapter 34) frequently develop ulcer disease, often with perforation, possibly as a result of vascular insufficiency. Chemotherapy, particularly when given selectively as a high-dose intra-arterial infusion in the celiac system, can be complicated by ulcer disease. Patients on anticoagulant therapy and those with other coagulopathies may rarely develop intramural hematoma of the gastrointestinal tract. Depending on the location, these hematomas may cause obstruction, but they can also leave large ulcers when they rupture into the lumen. Radiation therapy of the upper abdomen is sometimes complicated by chronic ischemic ulceration, especially as a late complication.

CLINICAL MANIFESTATIONS

The clinical manifestations of acid peptic disease (Table 139-2) do not always predict the various morphologic presentations found at endoscopy. Indeed, a silent ulcer may be recognized only when it presents abruptly with a complication, most commonly hemorrhage or perforation, or it may be discovered incidentally when a diagnostic test is performed for other reasons. Nevertheless, the typical presentation of acid peptic disease is recurrent episodes of pain. The pain is almost invariably located in the epigastrium and may radiate to the back or, less commonly, to the thorax or other regions of the abdomen (see Table 139-2). Some patients describe the pain as burning or piercing, whereas others describe it as an uncomfortable feeling of emptiness of the stomach, referred to as *painful hunger.* Indeed, the pain may improve with the ingestion of food, only to return in the postprandial period. The timing of the pain in relation to meals and to the soothing effects of food is nonspecific, however, and may also occur in patients with functional dyspepsia without ulcer. Nocturnal epigastric pain that awakens a patient several hours after a late meal is more likely to represent ulcer pain.

TABLE 139-2 KEY SYMPTOMS AND SIGNS OF PEPTIC ULCER

UNCOMPLICATED ULCER
No symptoms ("silent ulcer" in up to 40% of cases)
Epigastric pain
Pain may radiate to the back, thorax, other parts of abdomen (cephalad most likely, caudad least likely)
Pain may be nocturnal (most specific), "painful hunger" relieved by food, or continuous (least specific)
Nausea
Vomiting
Heartburn (mimics or associated with gastroesophageal reflex)
COMPLICATED ULCER
Severe abdominal pain
Shock
Abdominal board-like rigidity (and rebound and other signs of peritoneal irritation)
Free intraperitoneal air
Hemorrhage
Hematemesis and/or melena
Hemodynamic changes, anemia
Previous history of ulcer symptoms (80%)
Gastric outlet obstruction
Satiation, inability to ingest food, eructation
Nausea, vomiting (and related disturbances)
Weight loss

Aside from the pain during symptomatic episodes, patients may complain of retrosternal burning (heartburn) or acidic regurgitation into the throat, symptoms that reflect associated gastroesophageal reflux (Chapter 138), which is aggravated by hyperacidity or delayed gastric emptying. Nausea and vomiting may also occur but are nonspecific. The presence of significant diarrhea should raise the possibility of Zollinger-Ellison syndrome (Chapter 195), but diarrhea also may result from the intensive use of magnesium-containing antacids. In untreated patients, symptoms tend to be intermittent, with flares of daily pain lasting 2 to 8 weeks, separated by prolonged asymptomatic intervals. During periods of remission, patients may feel well and may be able to eat even heavy or spicy meals without apparent discomfort.

Physical Examination

The physical examination is usually unrevealing. If significant bleeding has occurred (Chapter 135), the patient may present with pallor and may be hypovolemic (Chapter 106). It is always useful to inquire about the characteristics of the stool because ulcer-related bleeding may manifest not only obviously in the form of hematemesis but also insidiously as melena (black feces). In the case of massive ulcer bleeding with the rapid bowel passage of blood, patients may also present with red rectal blood loss. When a patient has acute perforation, severe epigastric and abdominal pain usually develops, and the patient appears distressed. Characteristically, intense contracture of the abdominal muscles is apparent on palpation, together with rebound tenderness and other signs of peritoneal irritation. With large amounts of intra-abdominal air, percussion may reveal hypertympany over the liver.

DIAGNOSIS

In a patient who presents with symptoms consistent with ulcer disease, the diagnostic evaluation should proceed along two different but complementary paths: confirmation of the anatomic abnormality and investigation of its cause (Table 139-3). In most patients, it is advisable to follow both diagnostic paths simultaneously, but sometimes it is reasonable to skip the anatomic verification as a cost-saving strategy and proceed to management based on probable cause.

Anatomic Diagnosis

Upper gastrointestinal endoscopy (Chapter 134) is the primary investigative tool in patients suspected of having acid peptic disease. This technique can

TABLE 139-3 DIAGNOSTIC PATHS AND TOOLS IN ULCER DISEASE

PATH 1: MORPHOLOGIC DIAGNOSIS
Gastroduodenoscopy Barium contrast (inferior alternative) Endoscopic ultrasound (selected cases only) Computed tomography (useful in selected cases)
PATH 2: ETIOLOGIC DIAGNOSIS
***Helicobacter pylori* Testing**
Histologic examination of gastric mucosa with appropriate stains Stool antigen test Carbon-13 urea breath test Serum antibodies
Ulcer Associated with Nonsteroidal Anti-Inflammatory Drug Use
History of drug ingestion Decreased platelet adherence Molecular identification of drugs, pro-drugs, metabolites (complex, expensive)
Acid Hypersecretory Syndromes
Serum gastrin elevation Gastrin provocative tests (intravenous secretin, meal) Gastric analysis (acid titration)

detect erosive gastritis (see Fig. 139-1) or an ulcer in the gastric wall or duodenal bulb (see Fig. 139-2). Because of the high prevalence and the spontaneous improvement of dyspeptic symptoms, endoscopy generally should not be performed immediately; rather, its use should be restricted to patients with persistent or recurrent symptoms. However, immediate endoscopy is indicated in patients with alarm symptoms such as weight loss, dysphagia, anorexia, considerable vomiting, anemia, or signs of occult or overt bleeding.

Endoscopy, which is the gold standard for diagnosis, is both highly sensitive and highly specific for the detection of ulcer disease. The most common locations for a peptic ulcer are the stomach and duodenal bulb, but peptic ulcers sometimes occur in the esophagus, the small bowel, and a Meckel diverticulum lined with heterotopic gastric mucosa. Endoscopic ultrasound may detect an unsuspected submucosal component or enlarged lymph nodes, such as may occur in gastric neoplasia, especially lymphoma and linitis plastica (Chapter 192). Ulcers in the dorsal wall of the duodenal bulb, especially at the transition from the bulb to the postbulbar descending portion of the duodenum, are most difficult to visualize, and they sometimes require a side-viewing endoscope, particularly when endoscopic treatment is needed. Other regions where gastroduodenal ulcers can be easily missed are the cardia and the gastric angulus. Dieulafoy lesions may be difficult to diagnose because of their small mucosal defects and intermittent bleeding. Some patients require more than one endoscopy, preferably during acute bleeding, to localize the lesion. Endoscopy is also useful for ascertaining the presence of concomitant disorders, including esophagitis and duodenitis, or complications, such as bleeding or a visible vessel (see Fig. 139-2); obtaining biopsy specimens, such as for histologic examination and to assess for *H. pylori* (see Fig. 139-3); and performing therapeutic interventions.

In rare cases, such as stenosis that blocks the advancing endoscope, conventional barium contrast radiographs (Chapter 133) are indicated. Additional investigations by endosonography or CT are needed when underlying malignant disease is suspected. The endoscopist should obtain biopsy samples from all gastric ulcers, especially those with a suspicious appearance, to exclude potential malignant disease. Because duodenal ulcers are less likely to be malignant, biopsies are usually not required unless malignancy is specifically suspected.

Etiologic Diagnosis

Diagnosis must focus on establishing the cause of the ulcer. The first step is to determine whether *H. pylori* or NSAID use is present because these are the major risk factors for peptic ulcers and can be contributing factors in ulcers with other precipitating causes.

Testing for Helicobacter pylori

In populations with high *H. pylori* prevalence, nearly all patients with peptic ulcer disease are positive for *H. pylori*, so diagnostic testing has little value except when antimicrobial susceptibility testing is needed. The prevalence of *H. pylori* remains high in immigrants from developing countries, where most people become *H. pylori* positive in youth. In Western countries, approximately 50% of individuals who are older than 65 years are colonized with *H. pylori*, but the prevalence is less than 20% in those younger than 30 years. In these younger persons, the proportion of patients with ulcers who are *H. pylori* negative is higher than in older patients, making diagnostic testing for *H. pylori*, followed by targeted therapy in those who are positive for the bacterium, more attractive than empirical therapy.

The presence of *H. pylori* can be ascertained by four possible approaches. *Histologic examination* of gastric mucosal biopsies (either routine hematoxylin and eosin stain or a specific stain, such as Warthin-Starry stain), which is the standard procedure when diagnostic endoscopy is initially performed, is sensitive and specific for *H. pylori.* However, the accuracy of this technique may be affected by sampling error, improper orientation of the specimen, and recent therapy with proton pump inhibitors or antibiotics.

A second option is *serology*, which is a relatively simple, inexpensive test that has reduced predictive value in areas where the prevalence of *H. pylori* is low. Serology is not helpful to verify whether *H. pylori* has been eradicated with antibiotics because it may take more than 6 months or even 1 to 2 years for *H. pylori* antibodies to decrease to undetectable levels. None of the diagnostic tests that do not involve endoscopy can determine whether an ulcer is present.

A third option is a *stool H. pylori antigen test*, which is similar in accuracy to standardized serologic testing. Finally, the *carbon-13* or *carbon -14 urea breath test*, which relies on the detection of *H. pylori* urease activity, is a noninvasive and relatively simple test, but it is more expensive than stool or blood testing. Although the test usually becomes negative as soon as *H. pylori* treatment is begun, a minimum interval of 6 to 8 weeks after therapy has ended is recommended to reduce false-negative results.

Nonsteroidal Anti-inflammatory Drugs

NSAIDs are usually established as the putative cause of an ulcer based on information obtained from the patient. Assessment of NSAID use in an individual patient presenting with ulcer disease can be difficult both because NSAID use is common and often intermittent and because many different NSAIDs are widely available over the counter in most countries. NSAID use is usually evaluated by detailed medical history, focusing not only on current and recent drug use, and especially over-the-counter treatments but also on symptoms of pain, including musculoskeletal complaints. Further information from family members, family practitioners, and pharmacists is sometimes helpful. If surreptitious use of NSAIDs is suspected, direct serum and urine testing for aspirin and NSAID derivatives is feasible, or aspirin use can be assessed indirectly by a platelet adherence assay; however, these tests are not commonly used in clinical practice.

Hypersecretory Syndromes

Hypersecretory syndromes not related to *H. pylori* or NSAIDs are rare causes of ulcer disease and are diagnosed by special tests (see Table 139-3). Zollinger-Ellison syndrome should be strongly considered in patients with multiple ulcers, particularly in atypical locations such as distal to the duodenal bulb, and when diarrhea is present, because these finding are uncommon in *H. pylori*–related peptic ulcer disease. Hypergastrinemic syndromes (e.g., Zollinger-Ellison syndrome, antral G-cell hyperplasia) are best diagnosed by a determination of serum gastrin levels, both basal and after stimulation with intravenous secretin (gastrinoma detection) or a test meal (antral G-cell hyperplasia detection). When detected early, gastrinomas may be resectable (Chapter 195).

Gastric analysis, which is performed by placing a nasogastric tube to aspirate gastric juice and to quantify gastric acid output (both basal and after stimulation with subcutaneous pentagastrin), is indicated in only two rare circumstances: patients with elevated serum gastrin levels suggestive of Zollinger-Ellison syndrome or antral G-cell hyperplasia, but with equivocal responses to standard gastric provocative tests, and patients with indirect signs of gastric hypersecretion (e.g., enlarged folds and abundant clear fluid at endoscopy), normal gastrin levels, and negative provocative gastrin tests but who may still be hypersecretors, such as patients with recurrent ulcer disease despite a prior vagotomy with or without antrectomy. A basal acid output greater than 15 mEq/hour or greater than 5 mEq/hour in a postoperative patient is considered a positive test result.

The diagnosis of Zollinger-Ellison syndrome is best confirmed by gastric analysis showing a basal acid output greater than 15 mEq/hour in

conjunction with a fasting serum gastrin level exceeding 1000 pg/mL in the presence of gastric pH less than 2. To skip the cumbersome gastric analysis, a gastric pH determination showing a fasting pH of 2 or less is adequate. For serum gastrin levels in the range 100 to 1000 pg/mL and intragastric pH greater than 2, an increase in the serum gastrin to more than 200 pg/mL after a secretin stimulation test is suggestive of the diagnosis. An elevated serum gastrin level alone is not sufficient to diagnose Zollinger-Ellison syndrome because serum gastrin levels tend to increase over time with atrophic gastritis and also can be markedly increased in patients receiving proton pump inhibitor therapy.

Other Causes

In patients in whom a gastroduodenal ulcer cannot be ascribed to colonization with *H. pylori*, use of NSAIDs, or a hypersecretory syndrome, the establishment of a definite etiologic diagnosis may require a more thorough evaluation, starting with a medical history that focuses on the use of other ulcerogenic agents and the presence of symptoms that could suggest an underlying systemic disease. The next step is to evaluate biopsy samples from ulcer borders and from the antrum, corpus, and duodenum. The ulcer specimens may reveal overt or suspicious signs of malignancy, in particular adenocarcinoma (Chapter 192) or lymphoma. In these cases, further diagnostic evaluation should include staging of the malignancy. Alternatively, the biopsies may provide evidence of other infectious conditions, specific types of gastritis, celiac disease, ischemia, amyloidosis, or a systemic inflammatory condition. These data can be combined with clues provided by the endoscopic evaluation, including the character and location of the ulcer, signs of ischemia, and signs of inflammation at other locations. Further evaluation should focus on the presence of systemic disorders and may include a chest radiograph, angiography, ileocolonoscopy, and abdominal CT.

Differential Diagnosis

The differential diagnosis of ulcer-like symptoms includes many disorders of the upper abdominal organs, including malignant diseases of the stomach (Chapter 192), duodenum (Chapter 193), pancreas (Chapter 194), or bile ducts (Chapter 196). The differential diagnosis of upper abdominal symptoms also includes liver and gallstone disease (Chapter 155), pancreatitis (Chapter 144), and motility disorders (Chapter 136). In many patients with upper abdominal dyspeptic complaints, no underlying cause can be identified. In this "nonulcer" or functional dyspepsia group, complaints characteristic of gastroesophageal reflux, ulcer symptoms, or dysmotility symptoms may be prominent. Some patients (generally 5%) benefit from eradication of *H. pylori*, with a slow decrease of dyspeptic complaints over 12 to 24 months, but functional dyspepsia is not a proven or widely accepted indication for treatment of *H. pylori.* If such treatment is considered, both the patient and the physician should be prepared for persistent symptoms despite eradication of *H. pylori* and potentially for the emergence of antimicrobial-resistant *H. pylori*, which may result in a spiral of multiple courses of therapy in an attempt to remove resistant organisms.

Diagnostic Scenarios: Acute or Initial Clinical Presentation

Younger (≤45 years old) patients without alarm symptoms or signs such as anemia, rapid weight loss, or other evidence of serious disease do not necessarily require endoscopy, and evidence indicates that malignant gastric disease is unlikely. When a physician is treating a patient who lives in an area of the world with a relatively high prevalence of *H. pylori* infection (>10% of the population is positive), a test-and-treat approach can begin with an *H. pylori* stool antigen determination, urea breath test, or *H. pylori* serologic examination. However, although commonly practiced, the test-and-treat approach has never been documented to improve outcomes, except in the setting of endoscopically confirmed peptic ulcer disease. If the *H. pylori* test is positive, the patient can be treated with the appropriate *H. pylori* eradication drugs (see later) and observed for 4 to 6 weeks. If a patient with dyspeptic symptoms is taking an NSAID, either orally or parenterally, the first therapeutic approach is to discontinue these drugs and to prescribe a proton pump inhibitor for 4 to 6 weeks. In patients who test negative for *H. pylori* and who are not taking NSAIDs or who do not improve after these drugs are discontinued, endoscopy is indicated to determine whether an ulcer is present.

Conversely, gastroenterologists more commonly proceed directly to endoscopy. If no abnormalities are apparent or the endoscopic study shows only "gastritis" without an overt ulcer, a biopsy should be obtained to ascertain by histologic examination or urease testing whether *H. pylori* is present. If *H. pylori* is found by endoscopic biopsy, eradication treatment may be pursued. However, the efficacy of *H. pylori* eradication for the relief of functional dyspeptic symptoms is only about 5 to 13% greater than with placebo,[■■] and eradication therapy increases the risk for reflux-related symptoms (Chapter 138), especially in Asian patients.[5]

If endoscopic examination reveals an ulcer, its location determines the subsequent approach. An ulcer in the duodenal bulb has only a remote chance of representing a malignant lesion and need not routinely be examined by biopsy. By contrast, biopsy is mandatory for a gastric ulcerative lesion identified at endoscopy because malignant gastric disease may present with similar clinical manifestations and may resemble benign ulcer disease morphologically. Even if histologic assessment does not identify a malignant process, repeat endoscopy is recommended about 1 month after therapy to verify complete healing and for biopsy of the scar.

TREATMENT Rx

Helicobacter pylori–Associated Ulcers

H. pylori–associated ulcers often heal spontaneously, but acid suppressive therapy accelerates healing and ameliorates symptoms. Four weeks of acid suppressive therapy heals 70 to 80% of ulcers, and this number increases to more than 90% after 8 weeks of therapy. If *H. pylori* colonization persists, however, ulcers recur in 50 to 90% of patients within 12 to 24 months; this rate can be reduced to 20 to 30% with maintenance acid suppression and to less than 5% with *H. pylori* eradication.[■] Eradication treatment is therefore mandatory (Table 139-4). The success of eradication treatment strongly depends on therapy adherence and antimicrobial resistance. Resistance of *H. pylori* to metronidazole varies between 10 and 80% throughout the world. Clarithromycin resistance is increasing and ranges between 10 and 30% in many regions owing to the widespread use of macrolides to treat upper respiratory infections. Resistance to amoxicillin and tetracycline is rare and is not usually relevant in clinical practice. Because of this worldwide increase in prevalence of antimicrobial resistance, *H. pylori* eradication regimens continue to evolve.

The conventional regimen is "triple" therapy for 7 to 14 days.[6] Triple therapy combines a proton pump inhibitor with two antibiotics, usually combinations of amoxicillin, a nitroimidazole, and clarithromycin, the latter sometimes being replaced by levofloxacin.[■] Triple treatment for 10 to 14 days has about a 4 to 8% advantage over 7-day therapy,[■] which explains why 7-day therapy has become obsolete. Furthermore, double-dose proton pump inhibitor therapy (dose equivalent to omeprazole 40 mg twice daily) also increases eradication rates by approximately 10%. For patients in whom such therapy fails, a 10-day course of quadruple therapy is advised. This second-line regimen eradicates *H. pylori* in an additional 80% of patients.[7]

Because of the increased prevalence of antimicrobial resistance, triple therapies are increasingly being replaced by initial quadruple therapies.[8,9] Bismuth-based quadruple therapy consists of a proton pump inhibitor, a bismuth compound, and two antibiotics, usually tetracycline and a nitroimidazole (see Table 139-4). This regimen leads to eradication in 80 to 95% of patients. Non-bismuth-based quadruple therapies consist of a proton pump inhibitor plus three antibiotics, usually given for 10 days but often extended to 14 days (see Table 139-4). The three forms of non-bismuth-based quadruple therapy (sequential, hybrid, and concomitant therapy) differ in their antibiotic dosing schedule. Sequential therapy consists of a proton pump inhibitor with amoxicillin, usually given for 5 days, followed by a proton pump inhibitor with clarithromycin and a nitroimidazole for another 5 days. Ten-day sequential therapy is superior to 7- to 10-day triple therapy and equally effective as 14-day triple therapy.[■] Hybrid therapy starts with the same 5-day combination of a proton pump inhibitor with amoxicillin and then continues the amoxicillin for the next 5 days together with clarithromycin and metronidazole. Concomitant therapy gives these same drugs all together for 10 to 14 days. Other combinations of antibiotics are occasionally used (see Table 139-4). Concomitant therapy is the most effective, in particular when given for 14 days with a proton pump inhibitor in a dose of 40 mg twice daily.[■]

Continuation of acid suppressive therapy after antibiotic treatment is needed only when symptoms persist, or in cases of complicated ulcer disease until eradication of *H. pylori* has been confirmed. Ascertainment of therapeutic efficacy must be delayed at least 1 month after the end of treatment to prevent false-negative results related to the organism's temporary suppression but not eradication. When repeat endoscopy is needed (e.g., a gastric ulcer requires repeated histologic examination to exclude underlying malignancy), repeat screening for *H. pylori* can be performed using the gastric biopsy specimens for histologic examination, culture, or urease testing. If no clinical indication exists for repeat endoscopy, *H. pylori* status can be determined by a carbon-13 urea breath test, stool *H. pylori* antigen, or repeated serology. Serologic determination is based on a more than 40 to 50% decrease in immunoglobulin G antibody levels in the first 6 months after treatment compared with pretreatment levels in that patient; ideally, the (frozen) pretreatment and

TABLE 139-4 OVERVIEW OF ANTIBIOTICS USED FOR *HELICOBACTER PYLORI* ERADICATION

DRUG CLASS	DRUG	TRIPLE THERAPY* DOSE	BISMUTH-BASED QUADRUPLE THERAPY† DOSE	NON-BISMUTH-BASED QUADRUPLE THERAPY‡ DOSE
Acid suppression	Proton pump inhibitor	20-40 mg bid§	20-40 mg bid§	20-40 mg bid§
Standard antimicrobials	Bismuth compound‖		2 tablets bid	
	Amoxicillin	1 g bid	1 g bid	1 g bid
	Metronidazole¶	500 mg bid	500 mg tid	500 mg bid
	Clarithromycin	500 mg bid		500 mg bid
	Tetracycline		500 mg qid	
Salvage antimicrobials	Levofloxacin	500 mg bid	500 mg bid	500 mg bid
	Rifabutin	150 mg bid		
	Furazolidone	100 mg bid	100 mg bid	
	Doxycycline		100 mg bid	100 mg bid
	Nitazoxanide		500 mg bid	500 mg bid

*Triple therapy consists of a proton pump inhibitor or bismuth compound, together with two of the listed antibiotics, usually given for 7 to 14 days.
†Bismuth-based quadruple therapy consists of a proton pump inhibitor plus the combination of a bismuth compound and two antibiotics given for 7 to 14 days.
‡Non-bismuth-based quadruple therapy consists of a proton pump inhibitor, plus three antibiotics usually given for 10 days and sometimes extended to 14 days. The three forms of non-bismuth-based quadruple therapy differ in their antibiotic dosing schedules: (1) *sequential therapy* gives amoxicillin for the first half of the course, and then metronidazole and clarithromycin for the second half; (2) *hybrid therapy* starts with amoxicillin for the first half, and then continues the second half with amoxicillin, clarithromycin, and metronidazole; (3) *concomitant therapy* combines all three antibiotics throughout the 10- to 14-day therapy. Other combinations of antibiotics are occasionally used.
§Proton pump inhibitor dose equivalent to omeprazole 20 to 40 mg bid. (See Table 138-1 for doses of other proton pump inhibitors).
‖Bismuth subsalicylate or subcitrate.
¶An alternative is tinidazole 500 mg bid.

post-treatment specimens should be examined simultaneously by the same laboratory using the same assay.

After successful *H. pylori* eradication, the risk for recurrent infection in most populations is small. In a minority of patients, ulcers recur either owing to reinfection or in the presence of another ulcerogenic factor, particularly NSAID use.

Disease Related to Nonsteroidal Anti-inflammatory Drug Use

In patients who are diagnosed with acid peptic disease while they are taking NSAIDs or aspirin, the first step is to stop such therapy. Acid suppression with a proton pump inhibitor (in doses similar to those used for *H. pylori*) leads to healing of 85% of NSAID-induced gastric ulcers and more than 90% of duodenal ulcers within 8 weeks of therapy, whereas acid suppression with a histamine-2 (H_2)-receptor blocker, equivalent to ranitidine 300 mg twice daily (see Table 138-1), heals approximately 70% of ulcers within 8 weeks. The mucosal protective drug misoprostol (200 mg four times daily) has a similar effect to the H_2 blockers. Treatment must be continued for at least 8 weeks, and maintenance therapy is needed in patients who continue to take NSAIDs. Gastric ulcers, larger lesions, and recurrent lesions heal more slowly.

Ulcer occurrence during NSAID therapy suggests a causal relationship, but patients should also be tested for *H. pylori*. In patients who are *H. pylori* positive, eradication therapy should be considered because there are no clear clinical parameters distinguishing between these etiologic factors. In patients who continue to take NSAIDs, maintenance therapy with a proton pump inhibitor (see Table 138-1) is superior to *H. pylori* eradication for the prevention of recurrent ulcer,[■] except in patients who use aspirin, for whom *H. pylori* eradication alone may be curative.

Idiopathic Ulcer Disease

Patients with idiopathic ulcer disease despite a thorough assessment for underlying causes are treated primarily with an acid suppressant, usually a proton pump inhibitor, because they are at considerable risk for recurrent ulcer disease. After the underlying cause is identified and adequately treated, acid suppressive therapy can be withdrawn if there are no additional risk factors for ulcer disease, such as NSAID therapy or *H. pylori* infection. If the cause of the idiopathic ulceration is not clarified and there is doubt about the adequacy of the diagnostic testing for *H. pylori*, empirical eradication therapy can be considered, especially when there is histologic evidence of chronic active gastritis without further explanation. If related to unidentified *H. pylori*, gastritis should slowly disappear after successful eradication therapy.

PREVENTION

Primary Prevention

A test-and-treat strategy for *H. pylori* colonization is sometimes considered for patients with dyspeptic symptoms, but there is no specific way to prevent *H. pylori*–associated ulcer disease. By contrast, primary prevention of NSAID-associated ulcer disease is widely advocated for patients at high risk because of a prior ulcer, severe concomitant disease, use of warfarin or high-dose corticosteroids, or older age (>65 years).[10] H_2 blockers (at a dose equivalent to ranitidine 300 mg twice daily; Table 138-1) partially prevent duodenal ulcer disease during NSAID therapy but have no effect on preventing gastric ulcers unless a higher dose (equivalent to famotidine 40 mg twice daily) is given. Proton pump inhibitors (at a dose equivalent to omeprazole 20 mg once daily; Table 138-1) and misoprostol (in doses varying between 400 and 800 mg/day) partially protect against both gastric and duodenal ulcers during NSAID use. Misoprostol and proton pump inhibitors are equally effective,[■] but adherence with therapy is lower with misoprostol owing to its side effects. Patients should be advised of the importance of adherence because less than 80% adherence to gastroprotection is associated with a more than two-fold increased risk for ulcer disease compared with those who are fully adherent. During low-dose aspirin therapy, primary prevention of ulcers is advocated for the same risk groups, using a proton pump inhibitor[■] or an H_2-receptor antagonist.[■]

Secondary Prevention

Secondary prevention of *H. pylori*–associated ulcer disease is mandatory and consists of successful bacterial eradication. Testing to ascertain *H. pylori* status after eradication therapy is indicated in patients with prior complicated ulcer disease or with persistent or recurrent symptoms after therapy, as well as in patients who fail to complete the therapeutic course.

Secondary prevention of NSAID-related ulcer disease is preferentially achieved by the withdrawal of NSAIDs. In patients who must continue taking NSAIDs, a change to a selective COX2 inhibitor in combination with a proton pump inhibitor at a dose equivalent to esomeprazole 20 mg twice daily is advocated, especially for patients with complicated ulcer disease.[11] This combination is associated with a lower risk for secondary peptic ulcer complications than treatment with a COX2 inhibitor alone.

Secondary prevention of recurrent ulcers in patients who use aspirin may depend on *H. pylori* status. In *H. pylori*–positive patients, *H. pylori* eradication is as effective as a proton pump inhibitor for the prevention of recurrent ulcers. In *H. pylori*–negative patients, additional acid suppressive therapy at a dose equivalent to esomeprazole 20 mg twice daily should be prescribed.[■] Secondary prevention of idiopathic ulcer disease consists primarily of maintenance therapy with a proton pump inhibitor and treatment of the underlying condition. When there is doubt about the accuracy of the diagnostic assessment for *H. pylori*, an empirical course of eradication treatment can be considered.

Complications

Hemorrhage

Hemorrhage (Chapter 135), which is the most common complication of peptic ulcer disease, occurs in about one in six patients with ulcers over the course of their ulcer activity. Ulcers caused by NSAIDs account for a larger proportion of these hemorrhages. Peptic ulcer is thus the most common cause of nonvariceal upper gastrointestinal bleeding, accounting for 40 to

TABLE 139-5 PARAMETERS OF THE ROCKALL, BLATCHFORD AND AIMS65 SCORING SYSTEMS FOR UPPER GASTROINTESTINAL BLEEDING*

SCORING CATEGORIES	SCORING PARAMETERS	ROCKALL SYSTEM† Parameter	ROCKALL SYSTEM† Score	BLATCHFORD SYSTEM‡ Parameter	BLATCHFORD SYSTEM‡ Score	AIMS65§ Parameter	AIMS65§ Score
Age	Age (yr)	60-79	1	N/A	—	>65	1
		≥80	2				
Clinical status	Systolic blood pressure	<100	2	100-109	1	≤90	1
				90-99	2		
				<90	3		
	Pulse	>100	1	>100	1	N/A	—
	Melena	N/A	—	Present	1	N/A	—
	Syncope	N/A	—	Present	1	N/A	—
	Altered mental status	N/A	—	N/A	—	Present	1
	Comorbidities	Any major comorbidity	2	Hepatic disease	2	N/A	—
		Renal or liver failure, or disseminated malignancy	3	Cardiac failure	2		
Laboratory parameters	Blood urea, mmol/L	N/A	—	6.5-7.9	2	N/A	—
				8.0-9.9	3		
				10-24.9	4		
				>25.0	6		
	Hemoglobin, g/L	N/A	—	Men; 120-130	1	N/A	—
				Women; 100-120	1		
				Men; 100-120	3		
				Men and women; <100	6		
	INR	N/A	—	N/A	—	>1.5	1
	Albumin	N/A	—	N/A	—	<3.0 g/dL	1
Endoscopy	Endoscopic diagnosis	No focus, or Mallory-Weiss tear	0	N/A	—	N/A	—
		Upper GI malignancy	2				
		All other diagnoses	1				
	Endoscopic SRH	None/dark spot only	0	N/A	—	N/A	—
		Blood/clot/vessel	2				

*The Blatchford (Stanley AJ, Ashley D, Dalton HR, et al. Outpatient management of patients with low-risk upper-gastrointestinal haemorrhage: multicentre validation and prospective evaluation. *Lancet.* 2009;373:42-47) and AIMS65 (Saltzman JR, Tabak YP, Hyett BH, et al. A simple risk score accurately predicts in-hospital mortality, length of stay, and cost in acute upper GI bleeding. *Gastrointest Endosc.* 2011;74:1215-1224) systems are pre-endoscopy scoring systems; the Rockall system (Rockall TA, Logan RF, Devlin HB, et al. Risk assessment after acute upper gastrointestinal haemorrhage. *Gut.* 1996;38:316-321) has a pre- and post-endoscopy component. The Blatchford system is a tool to select patients for early discharge and later endoscopy during office hours. The Rockall score can help to assess the risk for rebleeding and mortality. The AIMS65 predicts mortality.
†Scores before and after endoscopy. Low risk defined as scores of ≤2, high risk defined as scores ≥6.
‡Low risk defined as score of 0.
§Low risk defined as score of ≤1, high risk as score ≥2.
GI = gastrointestinal; INR = international normalized ratio; N/A = not applicable; PUD = peptic ulcer disease; SRH = stigmata of recent haemorrhage.

60% of cases in most populations. Bleeding is associated with a 5 to 15% risk for rebleeding and up to a 10% risk for mortality. Hemorrhage may occur along a continuum from a serious acute event associated with hemodynamic shock and high mortality to slow or intermittent blood loss leading to chronic anemia. Approximately 80% of patients with bleeding ulcers describe a prior history of symptomatic disease, and about 20 to 30% have suffered a previous hemorrhage. Assessment of the magnitude of bleeding is of paramount importance in determining the need for transfusion and subsequent management (Table 139-5). Initial hematocrit levels may be misleading and are likely to fall because of hemodilution. Rapid bleeding is usually apparent on the basis of clinical signs (pallor, systolic blood pressure ≤100 mm Hg, pulse ≥100/minute); immediate fluid resuscitation and transfusions are indicated to prevent circulatory collapse. Mortality is particularly related to complications of the bleed, such as aspiration, and exacerbation of underlying disease, such as pulmonary, cardiovascular, renal, and hepatic disease. On rare occasions, patients may actually bleed to death, especially when a larger artery is affected, such as when an ulcer in the posterior wall of the duodenal bulb perforates the gastroduodenal artery. The stopping of such bleeding is therefore a medical emergency.[12]

Initial treatment aims at hemodynamic stabilization. Endoscopy is the mainstay of therapy and should be performed emergently or within 24 hours of presentation in high-risk cases, especially patients who are hemodynamically unstable, require transfusion, or have more severe comorbidities. Endoscopy can be performed to ascertain the origin of the bleeding and, if necessary, provide therapy to stop the bleed and reduce the risk for rebleeding.

The appearance of the ulcer determines the need for endoscopic treatment and the risk for rebleeding (Table 139-6) and mortality. "Clean base" ulcers and those with flat and pigmented spots carry a low risk for rebleeding and do not require endoscopic treatment. In contrast, ulcers with active bleeding or stigmata of recent bleeding, in particular a visible vessel or adherent clot, require treatment to stop the bleed and reduce the otherwise high risk for recurrence.

TABLE 139-6 ENDOSCOPY RESULTS IN PATIENTS WITH BLEEDING ULCERS

ENDOSCOPY RESULT	ULCER CHARACTERISTICS*	RISK FOR RECURRENT BLEEDING (%)
Active bleeding	Arterial bleeding	80-90
	Oozing bleeding	10-30
Stigmata of recent bleeding	Nonbleeding visible vessel	50-60
	Adherent clot	25-35
	Flat pigmented spot	0-8
No signs of bleeding	Clean ulcer base	0-12

*The ulcer characteristics determine the risk for recurrent bleeding during follow-up.

Endoscopic therapy may lead to a three-fold reduction in episodes of recurrent bleeding and in the need for surgical intervention, as well as a 40% reduction in mortality.[A] Treatment modalities include injection therapy with epinephrine or a sclerosant, thermocoagulation, and mechanical pressure by means of clips. Thermocoagulation can be performed by direct contact, such as with a heater probe, or by a noncontact method, such as argon plasma coagulation. The efficacy of these methods is generally comparable, except that epinephrine injection alone is inferior to the other modalities but can be useful when combined with any of the others.[A] If an adherent clot is found, an attempt should be made to remove it by water flushing or

snaring to allow an assessment of the underlying ulcer base and the treatment of any underlying visible vessels to reduce the risk for rebleeding.

The pre-endoscopy administration of intravenous proton pump inhibitor therapy (at a dose equivalent to a bolus of 80 mg esomeprazole, followed by a continuous infusion of 8 mg/hour until endoscopy) reduces bleeding and the need for emergent endoscopic treatment, but it has no effect on the need for transfusion or the occurrence of rebleeding or death. For patients with active bleeding or stigmata of recent bleeding, endoscopic treatment should be followed by an intravenous proton pump inhibitor given as a bolus at a dose equivalent to 80 mg esomeprazole over 30 minutes, followed by a continuous infusion at a dose equivalent to esomeprazole 8 mg/hour for 72 hours, to reduce rebleeding and the need for further intervention. Other therapies, including tranexamic acid, vasopressin, somatostatin, and octreotide, should be considered experimental (Chapter 135). Second-look endoscopy is not routinely indicated but can be considered in very high-risk cases, particularly when there is doubt about the adequacy of initial visualization or treatment.[13]

About 70 to 80% of rebleeds occur within the first 3 days and generally should be managed by repeat endoscopy. If endoscopy fails to stop the bleed or prevent further rebleeding, surgery and interventional radiology are equivalent options. Surgery includes stitching of the ulcer and occlusion of the feeding artery, usually the gastroduodenal artery. Interventional radiology uses angiography to insert a coil in the culprit vessel at the site of the bleed. Single-center observational experience suggests that each of these methods is equally effective in the hands of experienced clinicians, and the choice depends on local availability and expertise.

The risk for a fatal outcome of an upper gastrointestinal hemorrhage can be estimated based on five clinical and endoscopic parameters (see Table 139-5). In several studies, mortality in patients with a bleeding peptic ulcer was less than 2% among those with a score of 2 points or less, 10% in those with 3 to 5 points, and up to 46% in those with 6 points or more. Management of patients who recover after a peptic ulcer hemorrhage is similar to the treatment of patients with uncomplicated ulcers. Eradication of *H. pylori* provides excellent protection against both recurrence and rebleeding of *H. pylori*–related ulcers. NSAID-induced ulcers are preferentially managed by the withdrawal of NSAIDs or, if this is not feasible, by the combination of a COX2 inhibitor and a proton pump inhibitor at a dose equivalent to esomeprazole 20 mg twice daily. In patients with a history of ulcer bleeding and concomitant cardiovascular disease requiring antiplatelet therapy, the combination of low-dose aspirin and a proton pump inhibitor at a dose equivalent to esomeprazole 20 mg twice daily is associated with a lower risk for complicated ulcer than is clopidogrel monotherapy. If a patient who requires antiplatelet therapy presents with a bleeding ulcer, antiplatelet therapy should be continued or restarted as soon as possible if the risk for a cardiovascular event outweighs the risk for recurrent bleeding.

Perforation

Perforation may manifest as an acute event, whereby gastric contents spill into the peritoneal cavity, or more insidiously as the ulcer slowly penetrates into surrounding tissues. Acute free perforation typically causes abrupt and severe abdominal pain associated with abdominal muscular spasm that produces board-like rigidity of the abdomen and other manifestations of peritoneal irritation. Secondary hemodynamic shock is common. The clinical diagnosis can be confirmed in approximately 80% of patients by a plain chest radiograph with the patient standing (Fig. 139-5); a CT scan can be obtained if doubt persists. Leukocytosis and elevated C-reactive protein levels develop rapidly, and mild hyperamylasemia may occur. Treatment begins by correcting hemodynamic, fluid, and electrolyte imbalances. Nasogastric suction is helpful, and prophylactic antibiotics (e.g., amoxicillin-clavulanic acid 1 g every 8 hours intravenously) are usually administered. Unless a specific contraindication exists, emergency surgery is usually indicated, although more conservative approaches are sometimes appropriate. Given the success in achieving the long-term cure of ulcer disease through the eradication of *H. pylori* and the withdrawal of NSAIDs, suturing of the perforated ulcer may be adequate, permitting the patient to avoid a more radical vagotomy with or without gastric resection.

Intractability

Intractability is a term strictly applied to an ulcer that persists even after intensive and prolonged proton pump inhibitor therapy. Symptoms may or may not be present. These rare cases may result from poor compliance with recommended treatment, surreptitious use of ulcerogenic drugs, or other diseases (e.g., Crohn disease, ischemia, infection with bacteria other than *H. pylori*, viral infection). If these issues are recognized and these diagnoses are pursued, further complications and interventions such as surgical vagotomy and pyloroplasty can almost always be avoided.

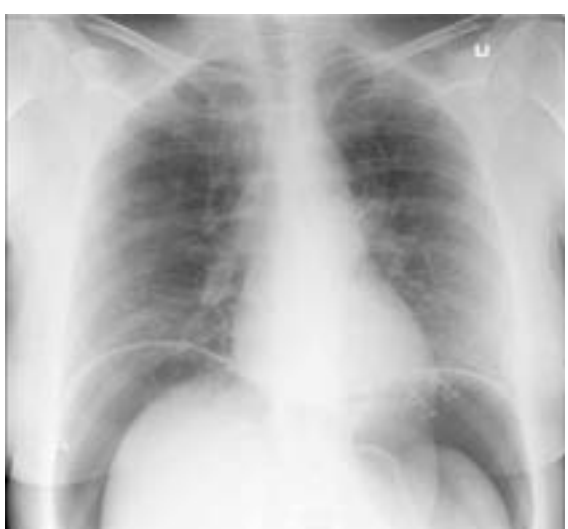

FIGURE 139-5. Plain chest radiograph in an upright patient with a perforated ulcer. The radiograph shows free air under the diaphragm.

Acid peptic disease related to alcohol or bisphosphonates should be addressed by discontinuing the precipitating agent. Treatment of Zollinger-Ellison syndrome requires high-dose proton pump inhibitors and/or surgery (Chapter 195). Those rare ulcers caused by Crohn disease (Chapter 141), vasculitis (Chapter 270), sarcoidosis (Chapter 95), polycythemia vera (Chapter 166), amyloidosis (Chapter 188), and other rare disorders should be addressed by treating the underlying condition. Stress ulcers and Cameron ulcers are treated by potent acid suppressive therapy (e.g., omeprazole 20 mg twice daily).

Stenosis

Gastric outlet obstruction is now a rare complication of ulcer disease because of the early detection and treatment of most ulcers. Most patients who develop clinically relevant gastric outlet obstruction have had an ulcer in the duodenal bulb and/or pyloric channel. Edema and inflammation play an important role, and occasionally a patient with active disease presents with symptoms of outlet obstruction as manifested by nausea, vomiting, and gastric stasis without a tight, chronic stenosis. Management therefore involves three key steps. The first is nasogastric tube aspiration and gastric lavage to clear the stomach of retained debris, followed by early endoscopy. This step facilitates an accurate diagnosis. Nasogastric suction may need to be maintained for several days if vomiting resumes when the tube is clamped. The second step consists of intense antisecretory therapy using intravenous proton pump inhibitors in a dose equivalent to a bolus of 80 mg esomeprazole over 30 minutes, followed by a continuous infusion of 8 mg/hour. Finally, the cause of the ulcer needs to be addressed, usually by eradicating *H. pylori* and withdrawing NSAIDs. If the initial treatment resolves the clinical situation and the patient can resume eating, it is often not necessary to undertake further treatment of the outlet stenosis; however, tight, fibrous scarring may require endoscopic balloon dilation or surgery.

PROGNOSIS

Most peptic ulcers heal spontaneously within weeks to months. However, if the underlying condition is not adequately treated, a large proportion of ulcers recur. Both initial and recurrent ulcers can give rise to complications. The four major complications are intractability, perforation, hemorrhage, and stenosis. Each distinct situation requires specific management approaches. Patients with complicated ulcer disease are at particular risk for recurrent complications and need careful assessment for secondary prevention. In patients who take long-term acid suppressive therapy with either H_2 antagonists or proton pump inhibitors, the risk for vitamin B_{12} deficiency (Chapters 164 and 218) is increased about two-fold.[14]

Grade A References

A1. Moayyedi P, Soo S, Deeks J, et al. Eradication of Helicobacter pylori for non-ulcer dyspepsia. *Cochrane Database Syst Rev*. 2006;2:CD002096.

A2. Mazzoleni LE, Sander GB, Francesconi CF, et al. Helicobacter pylori eradication in functional dyspepsia: HEROES trial. *Arch Intern Med*. 2011;171:1929-1936.

A3. Ford AC, Delaney BC, Forman D, et al. Eradication therapy for peptic ulcer disease in Helicobacter pylori positive patients. *Cochrane Database Syst Rev*. 2006;2:CD003840.

A4. Peedikayil MC, Alsohaibani FI, Alkhenizan AH. Levofloxacin-based first-line therapy versus standard first-line therapy for Helicobacter pylori eradication: meta-analysis of randomized controlled trials. *PLoS ONE*. 2014;9:e85620.

A5. Yuan Y, Ford AC, Khan KJ, et al. Optimum duration of regimens for Helicobacter pylori eradication. *Cochrane Database Syst Rev*. 2013;12:CD008337.

A6. Gatta L, Vakil N, Vaira D, et al. Global eradication rates for Helicobacter pylori infection: systematic review and meta-analysis of sequential therapy. *BMJ*. 2013;347:f4587.

A7. Molina-Infante J, Romano M, Fernandez-Bermejo M, et al. Optimized nonbismuth quadruple therapies cure most patients with Helicobacter pylori infection in populations with high rates of antibiotic resistance. *Gastroenterology*. 2013;145:121-128.

A8. Chan FK, Chung SC, Suen BY, et al. Preventing recurrent upper gastrointestinal bleeding in patients with Helicobacter pylori infection who are taking low-dose aspirin or naproxen. *N Engl J Med*. 2001;344:967-973.

A9. Graham DY, Agrawal NM, Campbell DR, et al. Ulcer prevention in long-term users of nonsteroidal anti-inflammatory drugs: results of a double-blind, randomized, multicenter, active- and placebo-controlled study of misoprostol vs lansoprazole. *Arch Intern Med*. 2002;162:169-175.

A10. Taha AS, McCloskey C, Prasad R, et al. Famotidine for the prevention of peptic ulcers and oesophagitis in patients taking low-dose aspirin (FAMOUS): a phase III, randomised, double-blind, placebo-controlled trial. *Lancet*. 2009;374:119-125.

A11. Chan FK, Wong VW, Suen BY, et al. Combination of a cyclo-oxygenase-2 inhibitor and a proton-pump inhibitor for prevention of recurrent ulcer bleeding in patients at very high risk: a double-blind, randomised trial. *Lancet*. 2007;369:1621-1626.

A12. Barkun AN, Martel M, Toubouti Y, et al. Endoscopic hemostasis in peptic ulcer bleeding for patients with high-risk lesions: a series of meta-analyses. *Gastrointest Endosc*. 2009;69:786-799.

A13. Laine L, McQuaid KR. Endoscopic therapy for bleeding ulcers: an evidence-based approach based on meta-analyses of randomized controlled trials. *Clin Gastroenterol Hepatol*. 2009;7:33-47.

A14. Kahi CJ, Jensen DM, Sung JJ, et al. Endoscopic therapy versus medical therapy for bleeding peptic ulcer with adherent clot: a meta-analysis. *Gastroenterology*. 2005;129:855-862.

A15. Sreedharan A, Martin J, Leontiadis GI, et al. Proton pump inhibitor treatment initiated prior to endoscopic diagnosis in upper gastrointestinal bleeding. *Cochrane Database Syst Rev*. 2010;7:CD005415.

A16. Sung JJ, Barkun A, Kuipers EJ, et al. Intravenous esomeprazole for prevention of recurrent peptic ulcer bleeding: a randomized trial. *Ann Intern Med*. 2009;150:455-464.

A17. Chan FK, Ching JY, Hung LC, et al. Clopidogrel versus aspirin and esomeprazole to prevent recurrent ulcer bleeding. *N Engl J Med*. 2005;352:238-244.

A18. Sung JJ, Lau JY, Ching JY, et al. Continuation of low-dose aspirin therapy in peptic ulcer bleeding: a randomized trial. *Ann Intern Med*. 2010;152:1-9.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. den Hoed CM, Vila AJ, Holster IL, et al. Helicobacter pylori and the birth cohort effect: evidence for stabilized colonization rates in childhood. *Helicobacter*. 2011;16:405-409.
2. Bastos J, Carreira H, La Vecchia C, et al. Childcare attendance and Helicobacter pylori infection: systematic review and meta-analysis. *Eur J Cancer Prev*. 2013;22:311-319.
3. Mayerle J, den Hoed CM, Schurmann C, et al. Identification of genetic loci associated with Helicobacter pylori serologic status. *JAMA*. 2013;309:1912-1920.
4. Masclee GM, Valkhoff VE, Coloma PM, et al. Risk of upper gastrointestinal bleeding from different drug combinations. *Gastroenterology*. 2014;147:784-792.
5. Xie T, Cui X, Zheng H, et al. Meta-analysis: eradication of Helicobacter pylori infection is associated with the development of endoscopic gastroesophageal reflux disease. *Eur J Gastroenterol Hepatol*. 2013;25:1195-1205.
6. Malfertheiner P, Megraud F, O'Morain CA, et al. Management of Helicobacter pylori infection: the Maastricht IV/ Florence Consensus Report. *Gut*. 2012;61:646-664.
7. Liu KS, Hung IF, Seto WK, et al. Ten day sequential versus 10 day modified bismuth quadruple therapy as empirical firstline and secondline treatment for Helicobacter pylori in Chinese patients: an open label, randomised, crossover trial. *Gut*. 2014;63:1410-1415.
8. Federico A, Gravina AG, Miranda A, et al. Eradication of Helicobacter pylori infection: which regimen first? *World J Gastroenterol*. 2014;20:665-672.
9. Malfertheiner P. Helicobacter pylori infection: management from a European perspective. *Dig Dis*. 2014;32:275-280.
10. Lanza FL, Chan FK, Quigley EM, et al. Guidelines for prevention of NSAID-related ulcer complications. *Am J Gastroenterol*. 2009;104:728-738.
11. Barkun AN, Bardou M, Kuipers EJ, et al. International consensus recommendations on the management of patients with nonvariceal upper gastrointestinal bleeding. *Ann Intern Med*. 2010;152:101-113.
12. Lau JY, Barkun A, Fan DM, et al. Challenges in the management of acute peptic ulcer bleeding. *Lancet*. 2013;381:2033-2043.
13. El Ouali S, Barkun AN, Wyse J, et al. Is routine second-look endoscopy effective after endoscopic hemostasis in acute peptic ulcer bleeding? A meta-analysis. *Gastrointest Endosc*. 2012;76:283-292.
14. Lam JR, Schneider JL, Zhao W, et al. Proton pump inhibitor and histamine 2 receptor antagonist use and vitamin B12 deficiency. *JAMA*. 2013;310:2435-2442.

REVIEW QUESTIONS

1. Which of the following is a likely cause of duodenal ulceration in patients who are *Helicobacter pylori* negative?

A. Bordetella pertussis infection
B. Zollinger-Ellison syndrome
C. Pheochromocytoma
D. Mercury poisoning
E. None of the above

Answer: B Zollinger-Ellison syndrome is caused by a gastrin-producing endocrine tumor, which is usually in the pancreas or small bowel and leads to marked hyperacidity. This syndrome usually gives rise to severe peptic ulcer disease with multiple, concomitant duodenal ulcers that are resistant to conventional acid suppressive therapy. The other conditions listed are not known to be associated with an increased risk for peptic ulcer.

2. An 80-year-old woman with degenerative joint disease involving her fingers presents with a gastric ulcer with smooth edges. She denies being prescribed nonsteroidal anti-inflammatory drugs (NSAIDs), and biopsies and serology for *H. pylori* are negative. Which of the following is the most likely cause of her ulcer?

A. Gastric rheumatoid nodules
B. Systemic lupus erythematosus
C. Over-the-counter medications containing NSAIDs
D. Ankylosing spondylitis
E. Surreptitious laxative abuse

Answer: C Over-the-counter NSAID use is very common, in particularly in elderly patients. Patients often do not recognize that they use NSAIDs even when specifically asked. Targeted questioning and checking of home medication is therefore indicated. In some series, surreptitious use of NSAIDs or aspirin was found as the underlying cause in 30 to 60% of patients with previously unexplained peptic ulcer.

3. Which of the following is the major factor contributing to current changes in the incidence of peptic ulcer disease in developed countries?

A. The rise in *H. pylori*–associated gastric ulcers
B. The fall in NSAID-associated duodenal ulcers
C. The rise in NSAID-related gastric ulcers
D. The rise in NSAID-related duodenal ulcers
E. The rise in intensive care unit (ICU)-related stress ulcers

Answer: C *H. pylori*–related ulcers have considerably decreased in developed countries owing to a decrease in the declining prevalence of the bacterium in recent generations and the use of eradication therapy. ICU-related stress ulcers have become less common owing to improvements in overall management, including respiratory and hemodynamic care, acid inhibition, and emphasis on adequate feeding. NSAID use has become more common and has given rise to an increase of NSAID-related ulcer disease.

4. Which of the following is correct about peptic ulcer disease?

A. Risk is greatest in females.
B. The highest relative risk is in persons younger than 20 years.
C. Treatment of autoimmune diseases with biologic agents (e.g., anti–tumor necrosis factor [TNF]-α agents) has increased risk.
D. Most deaths are due to gastric cancer.
E. Smoking increases risk.

Answer: E Peptic ulcer is more common in men and in elderly people. Anti-TNF-α treatment is not causally related to peptic ulcer disease, and most patients with peptic ulcers do not develop gastric cancer. In *H. pylori*–positive subjects, however, smoking increases the risk for peptic ulceration. This increased relative risk has been reported to be up to 12-fold.

5. Which of the following involves an increased risk for peptic ulcer disease?

A. Scleroderma
B. Pernicious anemia
C. Celiac disease
D. Systemic mastocytosis
E. Cryptococcosis

Answer: D In vitamin B_{12} deficiency with pernicious anemia, gastric atrophy is caused by an autoimmune gastritis. However, this inflammation is not associated with peptic ulcer disease. Autoimmune inflammation of the small bowel due to celiac disease also does not cause ulceration. Various infectious conditions can give rise to peptic ulcer, but cryptococcosis is not one of them. Systemic vasculitides can be associated with peptic ulcers, but they are not a common presenting manifestation of scleroderma. Systemic mastocytosis is, however, associated with a clearly increased risk for peptic ulceration.

6. The effect of *H. pylori* eradication therapy always needs to be assessed in patients with which of the following?

A. A bleeding peptic ulcer
B. Reflux esophagitis
C. Nonulcer dyspepsia
D. Uncomplicated peptic ulcer
E. Chronic active gastritis

Answer: A Patients with complicated peptic ulcer disease, such as peptic ulcer bleeding, are at considerably increased risk for recurrent ulcer complications. Patients with previous peptic ulcer bleeding in the presence of *H. pylori* therefore always need to be assessed to document the success of eradication therapy. If treatment has failed, they need repeat eradication treatment.

7. Which of the following statements is *false* in patients with peptic ulcer bleeding?

A. Endoscopic treatment is the mainstay of therapy.
B. Preemptive proton pump inhibitor (PPI) therapy reduces the need for endoscopic treatment.
C. High-dose continuous intravenous PPI treatment reduces the risk for rebleeding.
D. Patients with a low Blatchford score usually do not require intervention.
E. Recurrent bleeding requires surgery.

Answer: E In patients with peptic ulcer bleeding, endoscopic treatment is the mainstay to stop ongoing bleeding and reduce the risk for rebleeding. Preemptive PPI therapy reduces the need for endoscopic treatment, but it has no effect on rebleeding and mortality. High-dose continuous intravenous PPI treatment reduces the risk for rebleeding, the need for blood transfusion, and mortality in patients with a high risk for rebleeding. The Blatchford scale can be used to predict the need for intervention, such as endoscopic treatment and blood transfusion; a low score is associated with a low chance that an intervention is needed. Patients with recurrent bleeding despite endoscopic and PPI treatment can often first be retreated endoscopically. If this retreatment fails, both angiographic embolization of the feeding vessel and surgery are alternative rescue treatments. Recurrent bleeding is thus not a routine indication for surgery.

140

APPROACH TO THE PATIENT WITH DIARRHEA AND MALABSORPTION

CAROL E. SEMRAD

DEFINITIONS

Normal stool frequency ranges from three times per week to three times per day. As a symptom, diarrhea can be described as a decrease in stool consistency (increased fluidity), stools that cause urgency or abdominal discomfort, or an increase in the frequency of stool. Consistency is defined as the ratio of fecal water to the water-holding capacity of fecal insoluble solids, which are composed of bacterial mass and dietary fiber. Because it is difficult to measure stool consistency and because stool is predominantly (60 to 85%) water, stool weight becomes a reasonable surrogate of consistency.

As a sign, diarrhea is defined by the weight or volume of stool measured over a 24- to 72-hour period. Daily stool weights of children and adults are less than 200 g, and greater stool weights are an objective definition of diarrhea; however, this definition misses 20% of diarrheal symptoms in patients who have loose stools that are less than this daily weight.

Acute diarrheas persist for less than 2 to 3 weeks or, rarely, 6 to 8 weeks. The most common cause of acute diarrhea is infection. Chronic diarrheal conditions persist for at least 4 weeks and, more typically, 6 to 8 weeks or longer. The four mechanisms of diarrhea are osmotic, secretory, exudative, and altered motility. Because many diarrheal diseases are due to more than one of these mechanisms, it is clinically useful to categorize diarrhea as malabsorptive (fatty), watery, and inflammatory.

EPIDEMIOLOGY

Diarrhea is the second leading cause of mortality worldwide and is particularly problematic for elderly people and for children younger than 5 years of age in developing nations. Infectious diarrheal conditions cause approximately 750,000 worldwide childhood deaths annually, despite the improved use of oral rehydration solutions, zinc, and vitamin A supplements. Rotavirus infection (Chapter 380) is the most common cause of fatal childhood diarrhea, but the introduction of the oral monovalent rotavirus vaccine has significantly decreased its mortality rate in both developing and developed nations.

In the United States, norovirus has surpassed rotavirus as the leading cause of gastroenteritis requiring medical care.[1] Approximately 48 million Americans suffer from food-borne illness each year, including around 130,000 annual hospitalizations and 3000 deaths, most in elderly people. The major pathogens that cause diarrhea result in an estimated $14 to 16 billion in annual health care costs and days lost from work.

PATHOBIOLOGY

Fluid and Electrolyte Transport

Whether a hypotonic meal, such as a steak and water, or a hypertonic meal, such as milk and a doughnut, is consumed, the volume of the meal is augmented by gastric, pancreatic, biliary, and duodenal secretions. The permeable duodenum then renders the meal approximately isotonic with an electrolyte content similar to that of plasma by the time it reaches the proximal jejunum. As the intestinal slurry moves toward the colon, the Na^+ concentration in the luminal fluid remains constant, but Cl^- is reduced to 60 to 70 mmol/L, and bicarbonate (HCO_3^-) is increased to a similar concentration as the result of Cl^- and HCO_3^- transport mechanisms in the enterocyte and HCO_3^- secretion in the ileum (E-Fig. 140-1A and B). In the colon, K^{2+} is secreted, and the Na^+ transport mechanism of the colonocyte, together with the low epithelial permeability, extracts Na^+ and fluid from the stool. As a result, the Na^+ content of stool decreases to 30 to 40 mmol/L; K^{2+} increases from 5 to 10 mmol/L in the small bowel to 75 to 90 mmol/L; and poorly absorbed divalent cations, such as Mg^{2++} and Ca^{2++}, are concentrated in stool to values of 5 to 100 mmol/L. The anion concentrations in the colon change drastically because bacterial degradation of carbohydrate (i.e., unabsorbed starches, sugars, and fiber) creates short-chain fatty acids that attain concentrations of 80 to 180 mmol/L; at colonic pH, organic anions, such as acetate, propionate, and butyrate, are present. In the setting of carbohydrate malabsorption, the generation of high concentrations of these short-chain fatty acids may decrease stool pH to 4 or lower. The osmolality of the stool is approximately that of plasma (280 to 300 mOsm/kg H_2O) when it is passed.

At the cellular level, Na^{++} transport by the epithelium from lumen to blood (by Na^{++}-coupled sugar and amino acid transport in the small intestine, by Na^{++}/H^{++} exchange proteins in the small intestine and proximal colon, and by aldosterone-regulated $^+Na^{++}$ channels in the distal colon) creates a favorable osmotic gradient for absorption (see E-Fig. 140-1A and C). Chloride transport by the epithelium from blood to lumen (by cystic fibrosis transmembrane conductance regulator [CFTR] and the calcium-activated chloride channel in the small intestine and colon) creates an osmotic gradient for secretion (see E-Fig. 140-1B). Normally, the intestine is in a net absorptive state, regulated by extrinsic adrenergic nerves and proabsorptive neuropeptides and hormones (see E-Fig. 140-1D). Stimulation of secretion by neurotransmitters, hormones, and inflammatory mediators (Table 140-1) can offset this balance. A heterozygous missense mutation (c.2519G→T) in GUCY2C on chromosome 12 causes familial diarrhea by increasing GC-C signaling.

Diarrhea is due primarily to alterations of intestinal fluid and electrolyte transport and less to smooth muscle function. Each 24 hours, 8 to 10 L of fluid enters the duodenum. The diet supplies 2 L of this fluid; the remainder comes from salivary, gastric, hepatic, pancreatic, and intestinal secretions. The small intestine normally absorbs 8 to 9 L (80%) of this fluid and presents 1.5 L to the colon for absorption. Of the remaining fluid, the colon absorbs all but approximately 100 mL. Diarrhea can result from increased secretion by the small intestine or the colon if the maximal daily absorptive capacity of the colon (4 L) is exceeded. Alternatively, if the colon is diseased so that

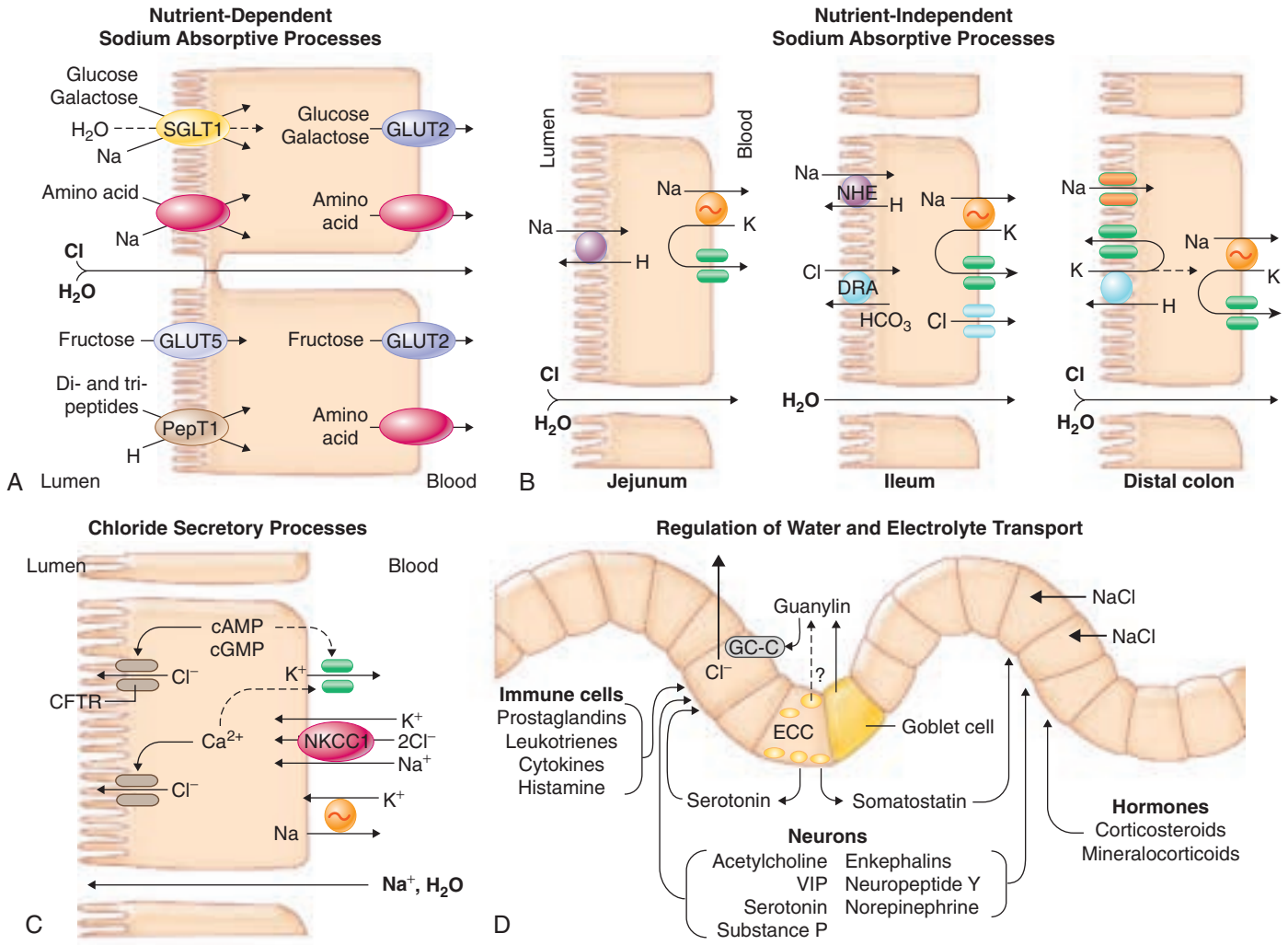

E-FIGURE 140-1. **Mechanisms of intestinal transport of water and electrolytes. A,** Sodium also is absorbed by nutrient-independent transport processes in the small intestine and colon. The Na^{++}/H^{++} (NHE) and Cl^-/HCO_3^- (DRA) exchangers are inhibited by agents that elevate intracellular cyclic adenosine monophosphate (*cAMP*), cyclic guanosine monophosphate (*cGMP*), or calcium. **B,** Chloride secretion by intestinal crypt cells. Chloride can be secreted actively throughout the small intestine and colon. Intracellular mediators of secretion (cAMP, cGMP, Ca^{2++}) open apical Cl^- channels (cystic fibrosis transmembrane conductance regulator [*CFTR*], calcium-activated chloride channel [TMEM16]) and basolateral K^{++} channels. Chloride moves from crypt cells into the intestinal lumen, favoring movement of Cl^- from the blood into cells by the $Na^{++}/K^{++}/2Cl^-$ co-transporter (*NKCC1*). Bicarbonate (HCO_3^-) also may be secreted via the CFTR channel. **C,** Intestinal sodium absorption. Sodium is actively absorbed in villus cells of the small intestine and surface cells of the colon. The sodium-potassium adenosine triphosphatase (Na^{++},K^{++}-ATPase) present on the cell basolateral membrane maintains a low intracellular Na^{+++} concentration and an electronegative cell interior favoring Na^{+++} movement across the apical membrane from lumen into cell. In the small intestine, glucose and galactose are taken up with sodium and water at the apical membrane by the sodium-glucose ligand transporter (*SGLT1*). Several different sodium-dependent amino acid carriers, some with overlapping substrate specificities, transport cationic, anionic, and neutral amino acids into villus cells. Dipeptides and tripeptides are transported by a hydrogen-coupled oligopeptide carrier, *PepT1*, that is driven by luminal hydrogen ions generated by the epithelial Na^+/H^{++} exchanger. Fructose is taken up by the facilitative glucose transporter (*GLUT5*). **D,** Regulation of intestinal water and electrolyte transport. Normally, the intestine is in a net absorptive state under the control of extrinsic adrenergic nerves from the sympathetic nervous system. Guanylin, the natural ligand for the *Escherichia coli* stable-toxin receptor (membrane-bound guanylyl cyclase [*GC-C*]), may be important in regulating local chloride secretion. The normal tone of the intestine is modified by the enteric nervous system, endocrine and inflammatory cells in the intestinal mucosa, and circulating hormones. The enteric nervous system releases a variety of neurotransmitters, some that stimulate chloride secretion (e.g., vasoactive intestinal peptide [*VIP*], acetylcholine) and others that promote sodium absorption (e.g., enkephalins, neuropeptide Y). Hormones produced locally from enterochromaffin cells (*ECC*) in the intestinal epithelium and inflammatory mediators released from immune cells directly affect enterocytes and nearby nerves. Circulating hormones (e.g., aldosterone, glucocorticoids) enhance sodium absorption in the intestine. Glucocorticoids also inhibit release of arachidonic acid and production of prostaglandin by inflammatory cells. DRA = down-regulated in adenoma gene; NHE = sodium-hydrogen exchanger.

it cannot absorb even the 1.5 L normally presented to it by the small intestine, diarrhea results.

Watery diarrheas may be due to osmotic, secretory, or inflammatory mechanisms. With ingestion of a poorly absorbed (e.g., Mg^{2++}) or unabsorbable (polyethylene glycol, lactulose or, in lactase-deficient individuals, lactose) solute, the osmotic force of the solute pulls water and secondarily sodium and chloride ions into the intestinal lumen. A considerable proportion of the osmolality of stool results from the nonabsorbed solute. This gap between stool osmolality and the sum of the electrolytes in the stool causes osmotic diarrhea.

TABLE 140-1 STIMULI OF INTESTINAL SECRETION

AGENT	INTRACELLULAR MEDIATOR	RELATED DIARRHEAL ILLNESS
Cholera, *E. coli* heat labile toxin, *Salmonella, Yersinia*	cAMP	Travelers, endemic
E. coli heat stable toxin	cGMP	
Rotatoxin (NSP4)	?	Viral gastroenteritis
Serotonin, PAF	Ca	Inflammatory, allergic
PG, leukotrienes	cAMP, Ca	Invasive enteric bacteria* inflammatory bowel diseases
PG	cAMP	Villous adenoma
Histamine	Ca	Intestinal allergies, mastocytosis, scombroid poisoning
VIP	cAMP	VIPoma, ganglioneuromas
5-HT, substance P, bradykinin	Ca	Malignant carcinoid
Calcitonin	?	Medullary carcinoma thyroid
Acetylcholine	Ca	Insecticides, nerve gas poisoning, cholinergic drugs
Ricinoleic acid	cAMP, Ca	Laxative abuse†
Caffeine	cAMP	Coffee, sodas, tea

5-HT = 5-hydroxytryptamine; Ca = calcium; cAMP = cyclic adenosine monophosphate; cGMP = cyclic guanosine monophosphate; PAF = platelet-activating factor; PG = prostaglandin; VIP = vasoactive intestinal peptide.
*_Shigella_ sp, _Clostridium difficile_, enteroinvasive, _E. coli, Vibrio parahaemolyticus, Clostridium perfringens_.
†Also phenolphthalein, anthraquinone, bisacodyl, dioctyl sodium sulosuccinate, and senna.

Active chloride secretion or inhibited sodium absorption, which also creates an osmotic gradient favorable for the movement of fluids from blood to lumen, explains the pathophysiology of the secretory diarrheas. Agents that increase enterocyte cyclic adenosine monophosphate (cAMP) (e.g., cholera toxin, prostaglandins), cyclic guanosine monophosphate (cGMP) (e.g., *Escherichia coli* stable toxin), or intracellular ionized calcium (Ca^{2++}) (e.g., acetylcholine) (see Table 140-1) inhibit non-nutrient Na^{++} absorption and stimulate Cl^- secretion (see Table 140-1 and E-Fig. 140-1B and D).

Inflammatory diarrheas, which may be watery or bloody, are characterized by enterocyte damage, villus atrophy, and crypt hyperplasia. The damaged enterocyte membrane of the small intestine has decreased disaccharidase and peptide hydrolase activity, reduced or absent Na^{++}-coupled sugar or amino acid transport mechanisms, and reduced or absent sodium chloride absorptive transporters. Conversely, the hyperplastic crypt cells maintain their ability to secrete Cl^- (and perhaps HCO_3^-). If the inflammation is severe, immune-mediated vascular damage or ulceration allows blood, pus, and protein to leak (exudate) from capillaries and lymphatics and contribute to the diarrhea. Activation of lymphocytes, phagocytes, and fibroblasts releases various inflammatory mediators that induce intestinal chloride secretion (see E-Fig. 140-1D). Interleukin-1 (IL-1) and tumor necrosis factor, which also are released into the blood, cause fever, anorexia, and malaise.

ACUTE DIARRHEA

CLINICAL MANIFESTATIONS

Approximately 80% of acute diarrheas are due to infections with viruses, bacteria, and parasites. The remainder is due to medications that have an osmotic force, stimulate intestinal fluid secretion, damage the intestinal epithelium, or contain poorly or nonabsorbable sugars (e.g., sorbitol), or less commonly fecal impaction, pelvic inflammation (e.g., acute appendicitis [Chapter 142]), or intestinal ischemia (Chapter 143).

Food-Borne and Water-Borne Infectious Diarrhea

Most infectious diarrheas are acquired through fecal-oral transmission from water, food, or person-to-person contact (Table 140-2). Patients with infectious diarrhea often complain of nausea, vomiting, and abdominal cramps that are associated with watery, malabsorptive, or bloody diarrhea and fever (dysentery) (Chapters 302 through 312, 336, 337, 350 to 352, 356, 357, 379, and 380). As documented using polymerase chain reaction methods of diagnosis, most outbreaks of nonbacterial acute gastroenteritis in the United States and other countries are caused by noroviruses (Norwalk agent; Chapter 380). Rotavirus (Chapter 380) predominantly causes diarrhea in infants, usually in the winter months, but also may cause nonseasonal acute diarrhea in adults, particularly in elderly people. Mechanisms for diarrhea

TABLE 140-2 EPIDEMIOLOGY OF ACUTE INFECTIOUS DIARRHEA AND INFECTIOUS FOOD-BORNE ILLNESS

VEHICLE	CLASSIC PATHOGENS
Water (including foods washed in such water)	*Vibrio cholerae*, norovirus (Norwalk agent), *Giardia, Cryptosporidium*
Food	
Poultry	*Salmonella, Campylobacter, Shigella* sp
Beef, unpasteurized fruit juice	Enterohemorrhagic *Escherichia coli*
Pork	Tapeworm
Seafood and shellfish (including raw sushi and gefilte fish)	*V. cholerae, Vibrio parahaemolyticus,* and *Vibrio vulnificus; Salmonella* and *Shigella* sp; hepatitis A and B viruses; tapeworm; anisakiasis
Cheese, milk	*Listeria* sp
Eggs	*Salmonella* sp
Mayonnaise-containing foods and cream pies	Staphylococcal and clostridial food poisonings
Fried rice	*Bacillus cereus*
Fresh berries	*Cyclospora* sp
Canned vegetables or fruits	*Clostridium* sp
Sprouts	Enterohemorrhagic *E. coli, Salmonella* sp
Animal-to-person (pets and livestock) contact	*Salmonella, Campylobacter, Cryptosporidium,* enterohemorrhagic *E. coli*, and *Giardia* sp
Person-to-person (including sexual) contact	All enteric bacteria, viruses, and parasites
Daycare center	*Shigella, Campylobacter, Cryptosporidium,* and *Giardia* sp; viruses; *Clostridium difficile*
Hospitalization, antibiotics, or chemotherapy	*C. difficile*
Swimming pool	*Giardia* and *Cryptosporidium* sp
Foreign travel	*E. coli* of various types; *Salmonella, Shigella, Campylobacter, Giardia,* and *Cryptosporidium* sp; *Entamoeba histolytica*

Modified from Powell DW. Approach to the patient with diarrhea. In: Yamada T, Alpers DH, Owyang C, et al, eds. *Textbook of Gastroenterology,* 3rd ed. Philadelphia: Lippincott-Raven; 1999.

include decreased fluid absorption due to destruction of villus enterocytes and stimulation of fluid secretion by NSP4 rotatoxin and viral activation of the enteric nervous system. Ebola virus (*Filoviridae*) infects endothelial cells, macrophages, and dendritic cells. The mechanism for the massive watery diarrhea is not known (Chapter 381).

Food-borne bacterial diseases in the United States are primarily due to *Salmonella* (Chapter 308), *Campylobacter jejuni* (Chapter 303), and *E. coli* O157:H7 (Chapter 304), and less commonly *Shigella* (Chapter 309). The incidence of *Vibrio* infection is increasing owing to the consumption of raw shellfish. Outbreaks of *E. coli* O157:H7 have been associated with petting zoos, uncooked ground beef, and green leafy vegetables. These bacteria most often invade the distal small bowel and colon, where they multiply intracellularly and damage the epithelium. Diarrhea is due to the stimulation of intestinal secretion by inflammatory mediators, decreased absorption across the damaged epithelium, and exudation of protein into the lumen. *Shigella* species and enterohemorrhagic *E. coli* produce a similar toxin, the "Shiga toxin," which is cytotoxic to intestinal epithelial cells and causes inflammation, cell damage, and diarrhea with blood and pus.

Outbreaks of *Cryptosporidium* (Chapter 350) have been reported in water parks. This parasite causes diarrhea by adhering and fusing to the epithelial cell membrane in the small bowel, thereby causing cell damage. Organisms that are specific for seafood include *Vibrio parahaemolyticus* (Chapter 302), which causes either watery or bloody diarrhea, and *Vibrio vulnificus,* which causes watery diarrhea and, especially in patients with liver disease, a fatal septicemia. Ingestion of meat contaminated by anthrax (Chapter 294) causes fever, diffuse abdominal pain, and bloody stool or vomitus. Anthrax invades the intestinal mucosa; the organism, or anthrax toxin, causes inflammation, ulceration, and necrosis.

In addition to enteric infections, certain systemic infections (e.g., viral hepatitis [Chapter 148], listeriosis [Chapter 293], legionellosis [Chapter 314]), mycoplasma, and emerging infections (e.g., Hanta virus [Chapter 381], severe acute respiratory syndrome [SARS, Chapter 366], avian influenza [Chapter 364]) may cause or manifest with substantial diarrhea.

Environmental and Food Poisonings

Food poisoning refers to the accumulation of toxin in food owing to the growth of toxin-producing organisms, most commonly *Staphylococcus aureus* (Chapter 288), *Bacillus cereus, Clostridium perfringens* (Chapter 296), and *Clostridium botulinum* (Chapter 296). Diarrhea is usually of rapid onset, as early as 4 hours after ingestion, and is often associated with vomiting. Natural toxins also are responsible for mushroom (*Amanita*) poisoning (Chapter 110), which can also cause acute liver and kidney failure.

Environmental poisonings may be caused by heavy metals (arsenic from rat poison, gold, lead, mercury) that impair cell energy production. Arsenic (Chapter 22) also induces cardiovascular collapse at high doses. Insecticide (organophosphates and carbamates) poisoning occurs most commonly in field workers or from the ingestion of contaminated herbs or teas (Chapter 110); diarrhea, excessive saliva, and pulmonary secretions are caused by acetylcholine-stimulated chloride secretion in intestine and other epithelia. Patients often have associated vomiting and abdominal cramps.

Seafood is a common source of food poisoning, particularly fin fish and bivalve shellfish. Most of these toxins cause varying combinations of gastrointestinal (nausea, vomiting, diarrhea) and neurologic symptoms (tingling and burning around the mouth, facial flushing, sweating, headache, palpitations, and dizziness) within hours of seafood ingestion (Chapter 112). Similar symptoms are reported in patients with scombroid poisoning, which is caused by ingestion of decaying flesh of blood fish (tuna, mahi-mahi, marlin, or mackerel) that release large amounts of histamine (Chapter 112).

Marine dinoflagellates (algae) produce toxins that can cause paralytic shellfish poisoning, diarrhetic shellfish poisoning, and ciguatera (Chapter 112). Sporadic outbreaks of diarrhetic shellfish poisoning "red tides" occur when bivalve mollusks ingest dinoflagellates that produce saxitoxins (voltage-sensitive sodium-channel blocker) and okadaic acid (a lipid-soluble toxin that inhibits serine and threonine protein phosphatases 1 and 2A). Ingestion of contaminated mollusks by humans results in diarrhea and neurologic symptoms. Saxitoxins cause predominantly neurologic symptoms (paralytic, neurotoxic, or amnestic shellfish poisonings) and okadaic acid gastrointestinal symptoms (diarrhetic shellfish poisoning).

Food-chain passage of another dinoflagellate species (*Gambierdiscus toxicus*) to fin fish (mackerel, amberjack, snapper, grouper, or barracuda) results in the accumulation of ciguatoxin (Chapter 112) that causes a seafood poisoning called ciguatera. Ciguatoxin activates voltage-sensitive sodium channels and causes neurologic and gastrointestinal symptoms. Fish from the Albemarle-Pamlico estuary (eastern United States) ingest toxic dinoflagellates that cause *Pfiesteria piscicida* poisoning. The dinoflagellate toxins cause nausea, vomiting, abdominal pain, diarrhea, and neurologic symptoms such as fatigue, myalgias, pruritus, circumoral paresthesias, reversal of hot and cold sensation, psychiatric abnormalities, and memory loss. The neurologic symptoms may persist for months to years. Puffer fish poisoning by tetrodotoxin, a voltage-sensitive sodium-channel blocker produced by the fish, causes neurologic symptoms, respiratory paralysis, and death.

Traveler's Diarrhea

North American travelers to developing countries and travelers on airplanes and cruise ships are at high risk for acute infectious diarrhea. Common causes of traveler's diarrhea (Chapter 286) include enterotoxic or enteroaggregative *E. coli, Shigella,* giardiasis, and norovirus.[2,3] *E. coli* heat-stable toxin binds to guanylate cyclase in the enterocyte brush-border membrane, where it results in elevation of intracellular cGMP. *E. coli* heat-labile toxin, similar to cholera toxin, binds to the monosialoganglioside GM_1 in the brush-border membrane, thereby resulting in the activation of adenylate cyclase and the elevation of intracellular cAMP. cAMP and cGMP stimulate intestinal chloride secretion and inhibit the nutrient-independent absorption of sodium and chloride. Sodium-glucose absorption is not affected, hence the basis for oral rehydration therapy. Cholera toxin permanently binds to adenylate cyclase until the natural turnover of the intestinal epithelium in 5 to 7 days, thereby resulting in persistent secretion and severe diarrhea. DNA sequencing to detect genome sequences of *Vibrio* cholera isolates, suggest human introduction as the cause of recent outbreaks. Of the annual average of six cases of cholera reported in the United States, most are travel associated.

Antibiotic-Associated Diarrheas

Antibiotics are a common cause of hospital-acquired diarrheas that occur in approximately 20% of patients receiving broad-spectrum antibiotics; approximately 30% of these diarrheas are due to *Clostridium difficile* (Chapter 296). Strains that produce increased levels of toxins A and B and a binary toxin have emerged. These strains are associated with an increase in the incidence and severity of *C. difficile* infections, including fulminant *C. difficile* colitis that can lead to colectomy or even death. The A and B toxins produced by *C. difficile* can cause diarrhea. In animal models, IL-8, substance P, and leukotriene B_4 were found to mediate toxin A–stimulated intestinal fluid secretion. *C. difficile* can cause severe diarrhea, pseudomembranous colitis, or toxic megacolon.

Nosocomial Hospital Diarrhea

Diarrhea is the most common nosocomial illness among hospitalized patients and residents in long-term care facilities. Common causes include antibiotic-associated diarrhea, *C. difficile* infection, medications, fecal impaction, tube feeding, and underlying illness. Magnesium-containing laxatives, antacids, and lactulose cause osmotic diarrheas. Bisacodyl laxatives cause secretory diarrhea. Liquid formulations of medications cause diarrhea (elixir diarrhea) because of the high content of sorbitol or other nonabsorbable sugars (e.g., mannitol) used to sweeten the elixir; patients prescribed liquid medications through feeding tubes may receive more than 20 g of sorbitol daily. An important but poorly understood cause of diarrhea is enteral (tube) feeding (Chapter 216), particularly in critically ill patients, who often develop diarrhea. Dysmotility, increased intestinal permeability, and low sodium content in enteral formulas may be contributing factors.

Patients in mental health institutions and nursing homes have a high incidence of nosocomial infectious diarrhea (e.g., *C. difficile* and less commonly *Shigella, Salmonella,* hemorrhagic *E. coli, Giardia, Entamoeba histolytica*). Infectious diarrhea, 50% or more of which is caused by *C. difficile,* is also common in acute-care hospitals. Severe *C. difficile* infection has also been reported among peripartum women. If outside foods are not brought to hospitalized patients, the likelihood of a nosocomial infection caused by *Salmonella* or *Shigella* is extremely rare. Immunosuppressed patients are also susceptible to nosocomial viral infections (rotavirus, norovirus, adenovirus, and coxsackievirus).

Cancer Treatment and Medication-Related Diarrhea

Abdominal or whole body radiation virtually always causes an increased frequency of bowel movements that are often watery. Cancer chemotherapy with amsacrine, azacitidine, cytarabine, dactinomycin, daunorubicin, doxorubicin, floxuridine, 5-fluorouracil, 6-mercaptopurine, methotrexate,

plicamycin, IL-2, and resveratrol may cause mild to moderate diarrhea. Irinotecan (CPT-11) or oxaliplatin and the combination of 5-fluorouracil plus leucovorin are frequent causes of severe watery diarrhea.

Olmesartan, an angiotensin II receptor antagonist (ARB), causes severe diarrhea as a result of a sprue-like enteropathy. Angiotensin-converting enzyme (ACE) inhibitors may cause abdominal pain and diarrhea resulting from visceral angioedema. Colchicine, neomycin, methotrexate, and *para*-aminosalicylic acid damage the enterocyte membrane. Cholestyramine, colestipol, and colesevelam bind bile salts and can result in malabsorptive diarrhea, especially in patients who have had an ileal resection. Gold therapy causes intestinal inflammation and diarrhea.

Daycare Diarrhea

More than 7 million children in the United States attend daycare, where diarrhea is extremely common, and secondary infection of family members occurs in 10 to 20% of cases. Most outbreaks of diarrhea are due to rotavirus or norovirus; less common causes are *Shigella* (Chapter 309), *Giardia* (Chapter 351), and *Cryptosporidium* (Chapter 350).

Runner's Diarrhea

Diarrhea occurs in 10 to 25% of individuals who exercise vigorously, especially women marathon runners and triathletes. Some athletes have associated abdominal cramps, urgency, nausea, or vomiting. The pathophysiology of runner's diarrhea is unknown. Release of intestinal secretogogues, especially prostaglandins, hormones, or ischemia, may be involved.

Diagnosis

Acute watery diarrhea may be due to infections, food toxins, or medications, or the acute diarrhea may signal the onset of a chronic disease (Fig. 140-1; see Tables 140-1 and 140-2) (Chapters 302 through 312, 336, 337, 351, 352, 356, 357, 379, and 380). The diagnostic approach in patients with fever and watery or bloody diarrhea should focus on stool cultures for *Campylobacter, Salmonella,* and *Shigella* sp. Routine stool culture is not indicated when diarrhea occurs after 3 to 5 days of hospitalization, except in patients with neutropenia, human immunodeficiency virus (HIV) infection, or signs of enteric infection. In patients with a history of recent antibiotic use, hospitalization, or peripartum, stools for *C. difficile* toxin should be obtained. Organisms that cause diarrhea but are not routinely tested by clinical microbiology laboratories include *Yersinia, Plesiomonas,* enterohemorrhagic *E. coli* serotype O157:H7, *Aeromonas, Cyclospora,* microsporidia, and noncholera *Vibrio.* Parasites such as *Giardia, Cryptosporidium,* and *Strongyloides* can be difficult to detect in stool but may be diagnosed by stool antigen testing or intestinal biopsy. Despite all testing techniques available, 20 to 40% of acute infectious diarrheas remain undiagnosed.

TREATMENT

Rx

Goals for the treatment of diarrhea include fluid replacement, antidiarrheal agents, nutritional support, and antimicrobial therapy when indicated.[4] Because death in patients with acute diarrhea is caused by dehydration, the first task is to assess the degree of dehydration and to replace fluid and electrolyte deficits.

Fluid Replacement

Severely dehydrated patients should be treated with intravenous Ringer lactate or saline solution, with additional potassium and bicarbonate as needed. Oral rehydration solutions, which are used extensively to replace diarrheal fluid and electrolyte losses, are effective because they contain sodium, sugars, and, often, amino acids that use nutrient-dependent sodium uptake transporters. In alert patients with mild to moderate dehydration, oral rehydration solution is equally effective as intravenous hydration in repairing fluid and electrolyte losses. Oral rehydration solutions can be given to infants and children in volumes of 50 to 100 mL/kg over 4 to 6 hours; adults may need to drink 1000 mL/hour. Reduced-osmolarity solutions (Na^{++} 75 mmol/L, osmolarity 245 mmol/L versus Na^{++} 90 mmol/L, osmolarity 311 mmol/L in standard solutions) are better tolerated and effective in noncholera diarrhea but may cause hyponatremia in patients with high-volume diarrhea, particularly children.[5] Glucose-based solutions, although effective in rehydrating the patient, may worsen the diarrhea. In contrast to glucose-based solutions, polymeric rice-based solutions decrease diarrhea in cholera victims; rice is digested to many glucose monomers that aid in the absorption of intestinal secretions. These solutions may not decrease stool output in acute diarrhea, but they will effectively rehydrate the patient despite continued diarrhea. After rehydration has been accomplished, oral rehydration solutions are given at rates equaling stool loss plus insensible losses until the diarrhea ceases.

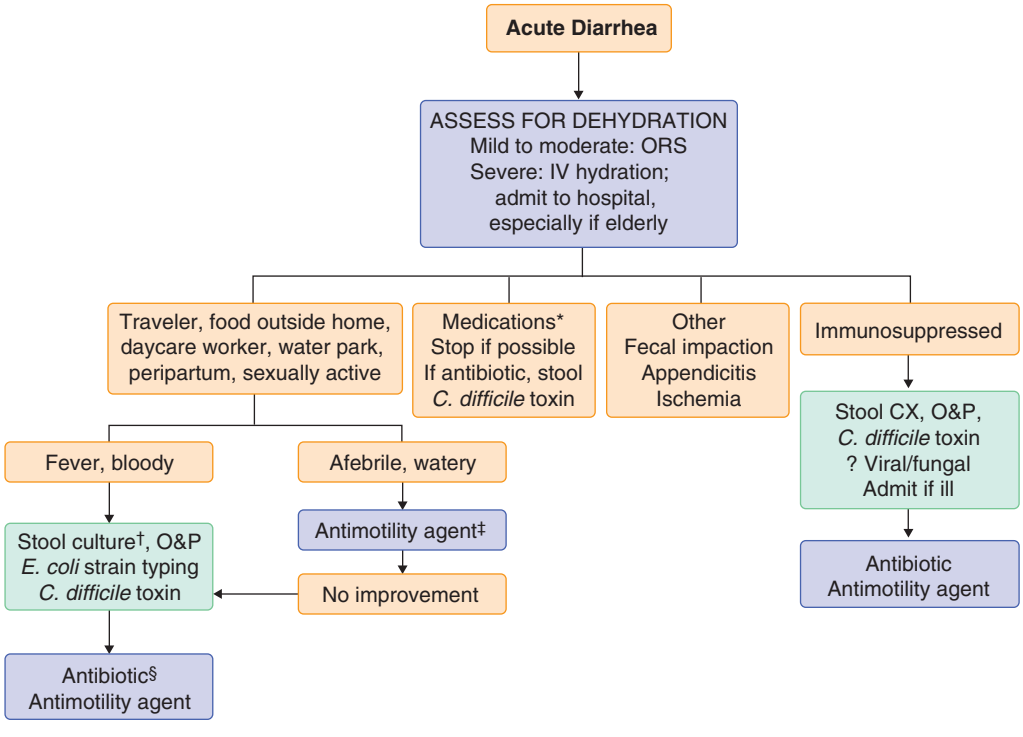

FIGURE 140-1. Approach to the diagnosis of acute diarrhea. *More than 700 medications cause diarrhea, including furosemide, caffeine, protease inhibitors, thyroid preparations, metformin, mycophenolate mofetil, sirolimus, cholinergic drugs, colchicine, theophylline, selective serotonin reuptake inhibitors, proton pump inhibitors, histamine-2 blockers, 5-ASA derivatives, angiotensin-converting enzyme inhibitors, bisacodyl, senna, aloe, anthraquinones, and magnesium- or phosphorus-containing medications. †Specifically request culture for *Yersinia, Plesiomonas,* enterohemorrhagic *Escherichia coli* serotype O157:H7, and *Aeromonas* if suspected. ‡If high suspicion for *Clostridium difficile* or invasive bacterial infection, wait for stool culture and toxin studies before starting. Racecadotril has antisecretory effects without paralyzing intestinal motility and can be used if available. §Not recommended for patients with bloody diarrhea due to *E. coli* O157:H7. CX = culture; IV therapy = intravenous rehydration; O&P = ova and parasites; ORS = oral rehydration solution.

Reducing Diarrhea

Bismuth subsalicylate (Pepto-Bismol, 525 mg orally [PO] every 30 minutes to 1 hour for five doses, may repeat on day 2) is safe and efficacious in bacterial infectious diarrheas. Opiates and anticholinergic drugs are not recommended for invasive bacterial infectious diarrheas because these drugs paralyze intestinal motility and predispose to increased colonization, invasion, and prolonged excretion of infectious organisms. The opiate loperamide is safe in acute or traveler's diarrhea, provided that it is not given to patients with dysentery (high fever, with blood or pus in the stool), and especially when administered concomitantly with effective antibiotics. A combination of loperamide (2 mg PO four times daily) plus simethicone (125 mg PO four times daily) may reduce the abdominal cramps and duration of traveler's diarrhea. Racecadotril (100 mg PO three times daily in adults, 1.5 mg/kg of body weight PO three times daily in children), an intestinal enkephalinase inhibitor that is antisecretory but does not paralyze intestinal motility, is effective in the treatment of acute diarrhea in children and adults. The diarrhea associated with enteral nutrition (Chapter 216) often can be managed with pectin (4 g/kg body weight daily) or, if there are no contraindications, with loperamide (2 mg PO four times daily for 3 to 7 days, maximal dose 16 mg daily), and diarrhea is not a reason to stop tube feeding unless stool volumes exceed 1 L/day.

Anxiolytics (e.g., diazepam 2 mg PO two to four times daily) and antiemetics (e.g., promethazine 12.5 to 25 mg PO once or twice daily) that decrease sensory perception may make symptoms more tolerable and are safe. Some foods or food-derived substances (green bananas, pectins [amylase-resistant starch], zinc) lessen the amount or duration of diarrhea in children. Unabsorbed amylase-resistant starches are metabolized in the colon to short-chain fatty acids that enhance fluid absorption. Zinc supplementation (20 mg of elemental zinc PO once daily) is effective in preventing recurrences of diarrhea in malnourished children; copper deficiency is a potential complication of prolonged zinc therapy.

Probiotics may be of benefit in children with acute diarrhea, predominantly that caused by rotavirus infection. *Lactobacillus* GG (10^{10} colony-forming units [CFU]/250 mL/day until diarrhea stops) added to an oral rehydration solution decreases the duration of diarrhea.

Antibiotics

While the clinician is awaiting stool culture results to guide specific therapy (Chapter 287), the fluoroquinolones (e.g., ciprofloxacin, 500 mg PO twice daily for 1 to 3 days, or levofloxacin, 500 mg PO daily for 1 to 3 days) are the treatment of choice when antibiotics are indicated (see Fig. 140-1). Trimethoprim-sulfamethoxazole (1 double-strength tablet PO twice daily for 5 days or 2 single-strength tablets PO twice a day for 5 days) is second-line therapy. If the symptom complex suggests *Campylobacter* infection, azithromycin (500 mg/day PO for 3 days) should be added. Regardless of the cause of infectious diarrhea, patients should be treated with antibiotics if they are immunosuppressed; have valvular, vascular, or orthopedic prostheses; have congenital hemolytic anemias (especially if salmonellosis is involved); or are extremely young or old.

Certain infectious diarrheas should be treated with antibiotics, including those associated with shigellosis (Chapter 309), cholera (Chapter 302), pseudomembranous enterocolitis (Chapter 296), parasitic infestations (Chapters 350 to 352 and 357), and sexually transmitted diseases (Chapter 285). Treatment of *E. coli* serotype O157:H7 infection is not recommended at present because current antibiotics do not appear to be helpful and the incidence of complications (hemolytic-uremic syndrome) may be greater after antibiotic therapy. Antibiotics are not effective for viral diarrhea or cryptosporidiosis.

For traveler's diarrhea, ciprofloxacin (500 mg PO twice daily for 3 days) is an effective treatment. The nonabsorbable antibiotic rifaximin (200 mg taken PO three times daily or 400 mg twice daily for 3 days) is safe and effective for treatment of traveler's diarrhea in Mexico, but it may not be effective against *Campylobacter* and *Shigella* infections.

Fluoroquinolone-resistant and trimethoprim-sulfamethoxazole–resistant strains of *Shigella, E. coli, Salmonella, Campylobacter,* and *C. difficile* have emerged. Azithromycin (500 mg PO on day 1 and 250 mg/day PO for 4 days) may be an effective alternative treatment for resistant strains of *Shigella* and *Campylobacter* and for traveler's diarrhea acquired in Mexico.

If *C. difficile* is suspected on an epidemiologic basis, metronidazole (250 mg PO four times daily or 500 mg PO three times daily for 10 days) or oral vancomycin (125 to 250 mg PO four times daily for 10 days) should be prescribed.[6] In patients with recurrent *C. difficile* infection that is associated with low serum antibody titers to toxin A, immunotherapy with monoclonal antibodies against toxin A and B may decrease recurrence rates. Fecal bacteriotherapy is more effective than vancomycin for the treatment of recurrences. Non–*C. difficile* antibiotic-induced diarrhea is generally mild and self-limited, and it usually clears spontaneously or in response to cholestyramine therapy (4 g PO four times daily for 2 weeks).

Treatment for chemotherapy-induced and radiation-induced mild to moderate diarrhea includes loperamide (2 mg PO four times daily) and nonsteroidal anti-inflammatory drugs (NSAIDs) (e.g., naproxen, 250 to 500 mg PO twice daily). Octreotide may be an effective treatment in those with severe diarrhea in doses up to 700 µg/day subcutaneously (SC).

PREVENTION

Rotavirus vaccination (Chapter 380) reduces the risk for infection and death and generally results in milder symptoms among those infected. Travelers to high-risk countries (Central America and parts of Latin America, Africa, Asia, the Middle East) should avoid ingestion of tap water and ice and of raw meat, raw seafood, and raw vegetables. An oral cholera vaccine against recombinant toxin B subunit and killed whole-cell (rBS-WC) is effective in preventing infection from the O1 El Tor strain and partially effective against enterotoxigenic *E. coli* strains. Cholera vaccination is recommended for relief workers and health professionals who work in endemic countries and for individuals who are immunocompromised or have chronic illnesses or hypochlorhydria. Rifaximin (200 to 600 mg/day PO for 2 weeks) and fluoroquinolones (e.g. norfloxacin 400/day PO for 2 weeks) are effective for reducing the risk for traveler's diarrhea in Mexico, and the combination of rifaximin plus loperamide is better than either one alone. Bismuth subsalicylate (525 mg PO four times daily for up to 3 weeks) is also effective. Loperamide and NSAIDs are taken prophylactically by many runners who are susceptible to runner's diarrhea, but it is not clear whether they are effective. Oral probiotics (e.g., lactobacilli and bifidobacteria) are not effective for preventing antibiotic-associated acute diarrhea.

CHRONIC DIARRHEA

An estimated 5% of the U.S. population suffers from chronic diarrhea, and approximately 40% of these individuals are older than 60 years of age.[7] The causes of chronic diarrhea include persistent infectious or inflammatory diarrheas, malabsorptive syndromes, and watery diarrheas (Table 140-3).

CLINICAL MANIFESTATIONS

Patients with malabsorption (Table 140-4) can present with a variety of gastrointestinal or extraintestinal manifestations (Table 140-5). Significant malabsorption of fat and carbohydrate usually causes chronic diarrhea, abdominal cramps, gas, bloating, and weight loss. Steatorrhea (fat in the stool) manifests as oily, foul-smelling stools that are difficult to flush down the toilet. Stools may be large and bulky (e.g., pancreatic insufficiency) or watery (e.g.,

TABLE 140-3 CAUSES OF CHRONIC DIARRHEA

- Persistent infectious diarrheas (see Table 140-2)
  - Brainerd diarrhea
- Malabsorptive syndromes (see Tables 140-4 and 140-6)
  - Common causes
    - Gastric bypass surgery
    - Dumping syndrome
    - Chronic pancreatitis
    - Intestinal bacterial overgrowth
    - Lactase deficiency
    - Celiac disease
    - Tropical sprue
    - *Giardia lamblia* infection
    - HIV/AIDS-related
    - Crohn disease (Chapter 141)
    - Radiation enteritis (Chapters 20 and 142)
- Watery diarrhea
  - Osmotic diarrhea
    - Magnesium, sodium phosphate, sulfate
    - Sorbitol, fructose
    - Glucose-galactose malabsorption
    - Disaccharidase deficiencies
  - Rapid intestinal transit
- Functional watery diarrhea (irritable bowel syndrome; Chapter 137)
  - Hormone-secreting tumors (VIPoma, carcinoid, gastrinoma, medullary thyroid cancer)
  - Systemic mastocytosis
  - Villous adenoma
  - Diabetes
  - Alcohol
  - Factitious
- Idiopathic
- Inflammatory diarrheas
  - Inflammatory bowel disease (Chapter 141)
  - Eosinophilic gastroenteritis
  - Microscopic colitis (collagenous or lymphocytic)
  - Food allergy

AIDS = acquired immunodeficiency syndrome; HIV, human immunodeficiency virus; VIP = vasoactive intestinal peptide.

TABLE 140-4 CAUSES OF MALABSORPTION

MECHANISM OF MALABSORPTION	CONDITIONS
Impaired mixing	Partial/total gastrectomy Gastric bypass surgery
Impaired lipolysis	Chronic pancreatitis Pancreatic cancer Congenital pancreatic insufficiency Congenital colipase deficiency Gastrinoma
Impaired micelle formation	Severe chronic liver disease Cholestatic liver disease Bacterial overgrowth Crohn disease Ileal resection Gastrinoma
Impaired mucosal absorption	Lactase deficiency Congenital enterokinase deficiency Abetalipoproteinemia Giardiasis Celiac disease Tropical sprue Agammaglobulinemia Amyloidosis AIDS-related (infections, enteropathy) Radiation enteritis Graft-versus-host disease Whipple disease Eosinophilic gastroenteritis Megaloblastic gut Collagenous sprue Refractory celiac disease Lymphoma Bacterial overgrowth Autoimmune enteritis Short-bowel syndrome
Impaired nutrient delivery	Congenital lymphangiectasia Lymphoma Tuberculosis Constrictive pericarditis Severe congestive heart failure
Unknown	Hypoparathyroidism Adrenal insufficiency Hyperthyroidism Carcinoid syndrome

AIDS = acquired immunodeficiency syndrome.

TABLE 140-5 CLINICAL CONSEQUENCES OF MALABSORPTION OF NUTRIENTS, WATER, AND ELECTROLYTES

NUTRIENT MALABSORBED	CLINICAL MANIFESTATION
Protein	Wasting, edema
Carbohydrate and fat	Diarrhea, abdominal cramps and bloating, weight loss and growth retardation
Fluid and electrolytes	Diarrhea, dehydration
Iron	Anemia, cheilosis, angular stomatitis
Calcium and vitamin D	Bone pain, fractures, tetany
Magnesium	Paresthesias, tetany
Vitamin B_{12} and folate	Anemia, glossitis, cheilosis, paresthesias, ataxia (vitamin B_{12} only)
Vitamin E	Paresthesias, ataxia, retinopathy
Vitamin A	Night blindness, xerophthalmia, hyperkeratosis, diarrhea
Vitamin K	Ecchymoses
Riboflavin	Angular stomatitis, cheilosis
Zinc	Dermatitis, hypogeusia, diarrhea
Selenium	Cardiomyopathy
Essential fatty acids	Dermatitis
Copper	Anemia, mental deterioration, neuropathy

bacterial overgrowth, mucosal diseases). Individuals with malabsorption also can present with manifestations of vitamin and mineral deficiencies. Dyspnea can be caused by anemia from iron, copper, folate, or vitamin B_{12} deficiency. Manifestations of calcium, magnesium, or vitamin D malabsorption include paresthesias and tetany resulting from hypocalcemia or hypomagnesemia and bone pain from osteomalacia or osteoporosis-related fractures. Paresthesias and ataxia are manifestations of cobalamin and vitamin E deficiency. Alternatively, neuropathy may be due to malnutrition or copper deficiency. Dermatitis herpetiformis is a blistering, burning, itchy rash on the extensor surfaces and buttocks that is associated with celiac disease.

Inflammatory diarrheas may manifest with fever and abdominal pain or with edema to suggest chronic protein loss. Patients may have multiple, low-volume, bloody stools with tenesmus to suggest proctitis or have severe diarrhea as a result of graft-versus-host disease (GVHD) or celiac disease. Systemic manifestations of inflammatory bowel disease include polymigratory arthritis, sacroiliitis, erythema nodosum, pyoderma gangrenosum, leukocytoclastic angiitis, uveitis, and oral aphthous ulcers.

DIAGNOSIS OF CHRONIC DIARRHEA

A detailed history and physical examination lead to a diagnosis in 25 to 50% of patients with chronic diarrheas (see Table 140-1 and Fig. 140-2). The addition of stool culture and examination for ova and parasites, determination of stool fat, and flexible sigmoidoscopy or colonoscopy with biopsy raises the diagnostic rate to approximately 75%. The remaining 25% of patients with chronic diarrhea may need extensive testing and perhaps hospitalization to make a diagnosis.

A history of 10 to 20 daily bowel movements that do not respond to fasting suggests secretory diarrhea (Fig. 140-3). A history of peptic ulcer should suggest gastrinoma (Chapter 195) or systemic mastocytosis (Chapter 255). Physical examination is helpful only if the thyromegaly of medullary carcinoma (Chapter 246), the cutaneous flushing of the neuroendocrine tumors and systemic mastocytosis, the dermatographism of systemic mastocytosis, or the migratory necrolytic erythema of glucagonoma (Chapter 195) is evident. Autonomic dysfunction (e.g., postural hypotension, impotence, gustatory sweating) is almost invariably present in diabetic diarrhea.

Evaluation for malabsorption begins with a careful elicitation of bowel habits and a description of the stool, weight loss, travel, food or milk tolerance, underlying gastrointestinal, pancreatic, or liver diseases, abdominal surgery, radiation or chemotherapy treatments, family history, and drug and alcohol use.

Blood Tests

Blood measurements (see Fig. 140-2) of iron, folate, vitamin B_{12}, vitamin D, or prothrombin time (vitamin K) help evaluate malabsorption. Specific antibody tests should be sent when celiac disease is suspected (see later). Peripheral blood findings of leukocytosis, eosinophilia, elevated erythrocyte sedimentation rate, hypoalbuminemia, or low total serum protein suggests an inflammatory diarrhea, whose hallmark is the presence of blood, either gross or occult, and leukocytes in the stool. There are no bedside screening tests to establish the diagnosis in watery diarrheas.

Imaging

Malabsorption may be suggested if a flat plate radiograph of the abdomen shows pancreatic calcification (Chapter 133). Some diseases (e.g., previous gastric surgery, gastrocolic fistulas, blind loops from previous intestinal anastomoses, small intestine strictures, multiple jejunal diverticula, abnormal intestinal motility that could lead to bacterial overgrowth) may be shown by computed tomography (CT) or magnetic resonance imaging (MRI) of the abdomen after administration of oral contrast agents or by a traditional upper gastrointestinal radiographic series with small intestine follow-through. A small bowel barium study may show thickening of the intestinal folds (e.g., amyloidosis, lymphoma or Whipple disease), uniform or patchy abnormalities (e.g., lymphoma or lymphangiectasia), or flocculation of barium and ilealization of jejunum to suggest celiac disease. Routine contrast radiographs of the gastrointestinal tract usually are not helpful in the diagnosis of watery diarrheas, unless they show extensive small bowel resection, the presence of a tumor (carcinoid or villous adenoma), or a bowel filled with fluid (endocrine tumor). Abdominal contrast imaging, particularly CT or MR enterography that uses oral neutral contrast to enhance the bowel wall, may show diagnostic evidence of advanced inflammatory bowel disease or changes suggestive of eosinophilic gastroenteritis or radiation enterocolitis. Somatostatin receptor scintigraphy with indium-111–labeled octreotide can be useful in localizing gastrinomas, pancreatic endocrine tumors, and carcinoid tumors.

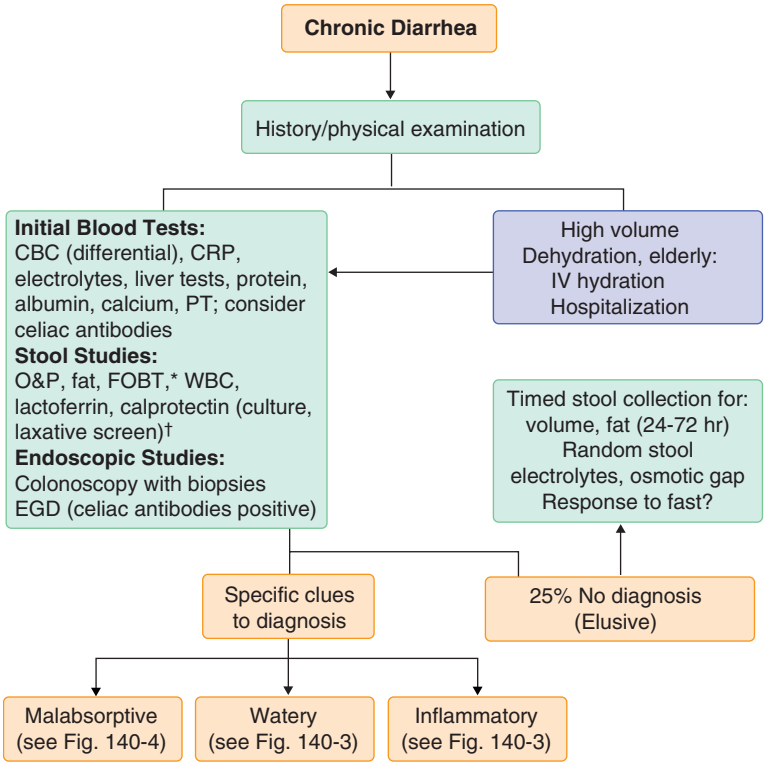

FIGURE 140-2. Initial approach to chronic diarrhea. *Fecal occult blood testing (FOBT) is a sensitive test for underlying bowel inflammation. †Perform stool culture in those who are immunosuppressed; perform laxative screen if laxative abuse is suspected. CBC = complete blood count; CRP = C-reactive protein; EGD = esophagogastroduodenoscopy; IV = intravenous; O&P = ova and parasites; PT = prothrombin time; WBC = white blood cells.

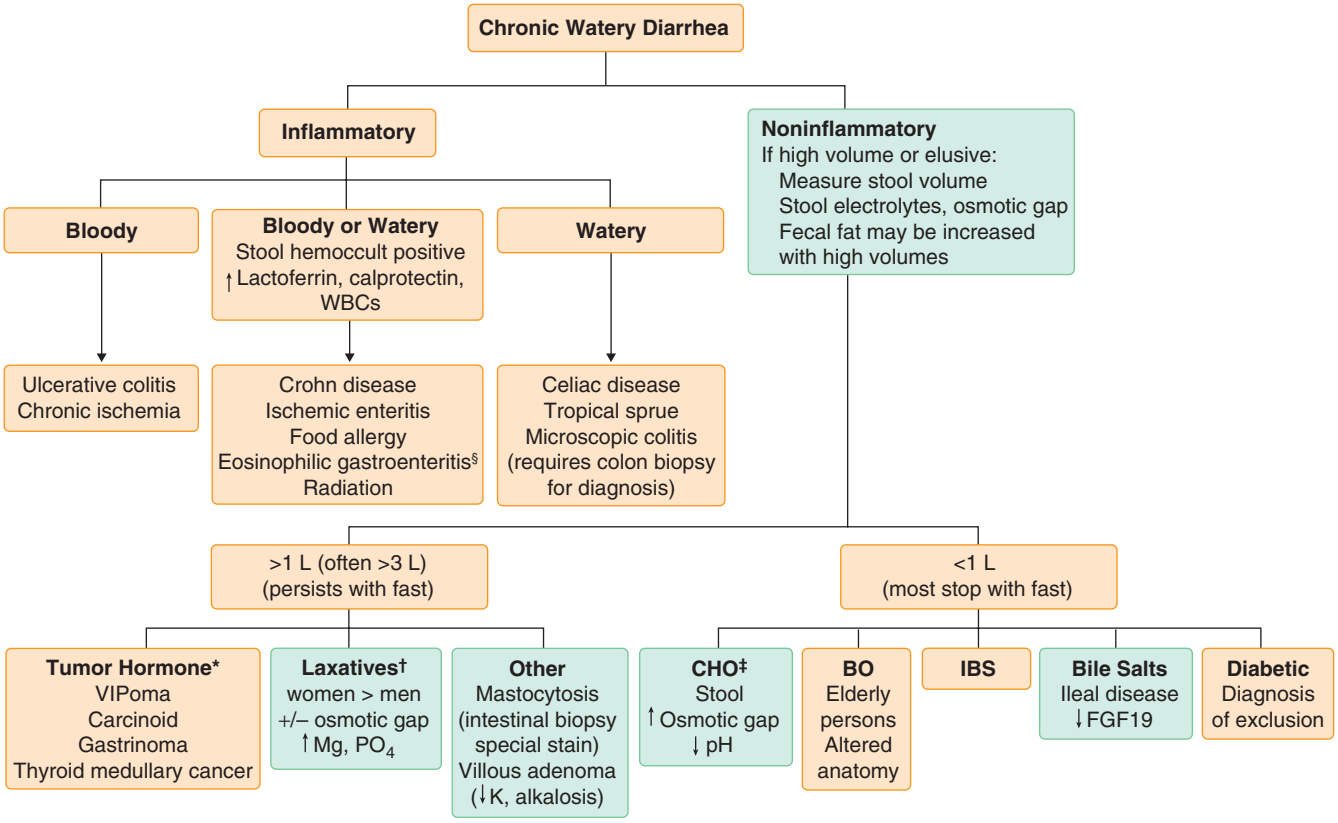

FIGURE 140-3. Approach to the evaluation of watery diarrheas. Many diarrheas have more than one mechanism (i.e., osmotic, secretory, inflammatory). Other causes include medications, postsurgical (vagotomy, Nissan wrap, cholecystectomy), hyperthyroidism, and alcohol. *VIPoma: >3 L output daily "pancreatic cholera," elevated VIP level. Carcinoid: elevated urine 5-hydroxyindole acetic acid, positive OctreoScan. Gastrinoma (Zollinger-Ellison syndrome): elevated gastrin level, positive secretin stimulation test, diarrhea due to high volume of acid secretion. Thyroid medullary cancer: elevated calcitonin level. †May be high or low volume depending on dose ingested, may respond to fast. ‡Carbohydrate malabsorption (CHO) may be due to lactase deficiency, dietary fructose, sorbitol in diabetic candies or liquid medications. §Full-thickness biopsy may be needed for diagnosis. BO = bacterial overgrowth; FGF = fibroblast growth factor; IBS = irritable bowel syndrome; WBCs = white blood cells.

Endoscopy and Biopsy

Upper endoscopy with distal duodenal biopsy should be undertaken if serologic tests for celiac disease are positive or diagnostic clues suggest small bowel mucosal malabsorption (Chapter 134). Small bowel biopsy is virtually always abnormal when the tTG immunoglobulin A (IgA) antibody level is very high (more than five-fold the normal range), and antiendomysial antibody (EMA) is positive. A biopsy may be avoided in this setting if gastrointestinal symptoms and the HLA risk alleles for celiac disease are present. Some patients may have patchy mucosal disease and require enteroscopy with jejunal biopsies for diagnosis. Wireless video capsule endoscopy (Chapter 134) and balloon-assisted enteroscopy are increasingly used to diagnose diseases that reside deep in the small bowel. Patients with severe watery or elusive diarrhea should have a colonoscopy to assess for villous adenomas, microscopic colitis, mastocytosis, or early inflammatory bowel disease. Colonoscopy also may show brown pigmentation suggestive of melanosis coli due to chronic use of anthracene laxatives. Terminal ileal biopsy may indicate infectious or inflammatory bowel disease.

Other Laboratory Tests

Malabsorption

If chronic diarrhea is the presenting symptom, a stool examination for ova and parasites and a stool antigen-capture enzyme-linked immunosorbent assay (ELISA) test for *Giardia* should be obtained. A stool test for fat on a high-fat diet (70 to 100 g/day) is the best available screening test for malabsorption (Table 140-6). If the fecal fat test result is negative, selective carbohydrate malabsorption or other causes of watery diarrhea should be considered. If the fecal fat test result is positive, further testing should be based on clinical suspicion for particular diseases. If pancreatic insufficiency is suspected, imaging studies of the pancreas should be performed. If bacterial overgrowth is suspected, culture of an intestinal aspirate or a breath test should be obtained. Small bowel contrast imaging is useful in detecting structural abnormalities that predispose to bacterial overgrowth (Table 140-7). If proximal mucosal damage is suspected, multiple small intestinal biopsy specimens should be obtained. If there are no clues, CT or MR enterography may help to detect middle and distal small bowel mucosal diseases. Some

TABLE 140-6 TESTS FOR THE EVALUATION OF MALABSORPTION*

TEST	COMMENTS
GENERAL TESTS OF ABSORPTION	
Quantitative stool fat test	Gold standard test of fat malabsorption, with which all other tests are compared. Requires ingestion of a high-fat diet (100 g) for 2 days before and during the collection. Stool is collected for 3 days. Normally, <7 g/24 hr is excreted on a high-fat diet. Borderline abnormalities of 8-14 g/24 hr may be seen in secretory or osmotic diarrheas that are not caused by malabsorption. There are false-negative findings if fat intake is inadequate. False-positive results can occur if the nonabsorbable fat olestra is ingested or mineral oil laxatives or rectal suppositories (e.g., cocoa butter) are given to the patient before stool collection.
Qualitative stool fat test	Sudan stain of a stool sample for fat. Many fat droplets per medium-power field (40×) constitute a positive test result. The nuclear magnetic resonance method determines the percentage of fat in the stool (normal, <20%). The test depends on an adequate fat intake (100 g/day). There is high sensitivity (90%) and specificity (90%) with fat malabsorption of >10 g/24 hr. Sensitivity drops with stool fat in the range of 6-10 g/24 hr.
Hydrogen breath test	Most useful in the diagnosis of lactase deficiency. An oral dose of lactose (1 g/kg body weight) is administered after measurement of basal breath H_2 levels. The sole source of H_2 in the mammal is bacterial fermentation; unabsorbed lactose makes its way to colonic bacteria, resulting in excess breath H_2. A *late peak* (within 3-6 hr) of > 20 ppm of exhaled H_2 after lactose ingestion suggests lactose malabsorption. Absorption of other carbohydrates (e.g., sucrose, glucose, fructose) also can be tested.
SPECIFIC TESTS FOR MALABSORPTION	
Tests for Pancreatic Function	
Secretin stimulation test	The gold standard test of pancreatic function. Requires duodenal test intubation with a double-lumen tube and collection of pancreatic juice in response to intravenous secretin. Allows measurement of bicarbonate (HCO_3^-) and pancreatic enzymes. A sensitive test of pancreatic function, but labor intensive and invasive.
Fecal elastase-1 test	Stool test for pancreatic function. Equal sensitivity to the secretin stimulation test for the diagnosis of moderate-to-severe pancreatic insufficiency. More specific than the fecal chymotrypsin test. Unreliable with mild insufficiency. False-positive results occur with increased stool volume and intestinal mucosal diseases.
Tests for Bacterial Overgrowth	
Quantitative culture of small intestinal aspirate	Gold standard test for bacterial overgrowth. Greater than 10^5 colony-forming units (CFU)/mL in the jejunum suggests bacterial overgrowth. Requires special anaerobic sample collection, rapid anaerobic and aerobic plating, and care to avoid oropharyngeal contamination. False-negative results occur with focal jejunal diverticula and when overgrowth is distal to the site aspirated.
Hydrogen breath test	The 50-g glucose breath test has a sensitivity of 90% for growth of 10^5 colonic-type bacteria in the small intestine. If bacterial overgrowth is present, increased H_2 is excreted in the breath. A hydrogen level (within 2 hr) of > 20 ppm suggests bacterial overgrowth. False-negative results occur with non–hydrogen-producing organisms. Concomitant measurement of breath methane improves test sensitivity.
Tests for Mucosal Disease	
Small bowel biopsy	Obtained for a specific diagnosis when there is a high index of suspicion for small intestinal disease. Several biopsy specimens (4-5) must be obtained to maximize the diagnostic yield. Distal duodenal biopsy specimens are usually adequate for diagnosis, but occasionally enteroscopy with jejunal biopsy specimens is necessary. Small intestinal biopsy provides a specific diagnosis in some diseases (e.g., intestinal infection, Whipple disease, abetalipoproteinemia, agammaglobulinemia, lymphangiectasia, lymphoma, amyloidosis). In other conditions, such as celiac disease and tropical sprue, the biopsy specimens show characteristic findings, but the diagnosis is made on improvement after treatment.
Tests of Ileal Function	
Schilling test	A test of vitamin B_{12} absorption (see Table 164-4 in Chapter 164).
^{75}SeHCAT test	This is a test of bile acid absorption. Seven days after ingestion of radiolabeled synthetic selenium–homocholic acid conjugated with taurine (^{75}SeHCAT), whole body retention is measured by a gamma-counting device. The result is expressed as a fraction of baseline ingestion. Retention values of <10% are abnormal and indicate bile acid malabsorption with a sensitivity of 80-90% and specificity of 70-100%. The radiation dose is equivalent to that of a plain chest x-ray. Liver disease and bacterial overgrowth may give false results. Not approved for use in the United States.

*Not all these tests are readily available. A strong suspicion for any disease may warrant foregoing an extensive work-up and obtaining the test with highest diagnostic yield. In some cases, empirical treatment, such as removing lactose from the diet of an otherwise healthy individual with lactose intolerance, is warranted without any testing.

TABLE 140-7 ABNORMALITIES CONDUCIVE TO BACTERIAL OVERGROWTH

STRUCTURAL
Afferent loop syndrome after gastrojejunostomy
Ileocecal valve resection
End-to-side intestinal anastomoses
Duodenal and jejunal diverticula
Strictures (Crohn disease, radiation enteritis)
Adhesions (postsurgical)
Gastrojejunocolic fistulas
MOTOR
Scleroderma
Diabetes mellitus
Idiopathic pseudo-obstruction
HYPOCHLORHYDRIA
Atrophic gastritis
Proton pump inhibitors
Acquired immunodeficiency syndrome
Acid-reducing surgery for peptic ulcer disease
MISCELLANEOUS
Immunodeficiency states
Pancreatitis
Cirrhosis
Chronic renal failure

individuals with celiac disease present with selective nutrient deficiencies without diarrhea. In these cases, tTG antibody tests and intestinal biopsy should be performed. In patients hospitalized for severe diarrhea or malnutrition, a more streamlined evaluation usually includes a stool for culture, ova and parasites, and fat; an abdominal imaging study; and a biopsy of the small intestine and colon.

Watery Diarrhea

Breath tests to measure the respiratory excretion of H_2 and methane after administration of carbohydrates can assess carbohydrate malabsorption or bacterial overgrowth (see Table 140-6).

The diagnosis of endocrine tumors, such as carcinoids, gastrinoma, VIPoma, medullary carcinoma of the thyroid, glucagonoma, somatostatinoma, and systemic mastocytosis, is made by showing elevated blood levels of serotonin, chromogranin A, or urinary 5-hydroxyindoleacetic acid and serum levels for gastrin, vasoactive intestinal peptide, calcitonin, glucagon, somatostatin, histamine, or prostaglandins (Chapter 195). Somatostatin receptor scintigraphy has proved to be sensitive and useful in the diagnosis and evaluation of Zollinger-Ellison syndrome (Chapter 195).

Inflammatory Diarrhea

Stool occult blood, white blood cells, or lactoferrin and calprotectin (components of leukocytes) are helpful tests for bowel inflammation. Video capsule endoscopy (Chapter 134) of the small bowel may detect ulcerations deep in the small bowel not reachable by standard upper or lower endoscopy and not detected with conventional barium contrast radiography. However, the risk for capsule retention in the small bowel is high in patients with Crohn disease or NSAID use, particularly when there is a history of obstructive symptoms. The most sensitive test for protein-losing enteropathy is measurement of intestinal protein loss by 24-hour stool excretion or clearance of α_1-antitrypsin.

Stool Examination in Elusive Diarrhea

An important adjunct to diagnosing the cause of diarrhea is to examine the stool. The greasy, bulky stool of steatorrhea and the bloody stool of gut inflammation are distinctive. Stool collections (see Table 140-6) can be analyzed for weight, volume, fat, electrolytes ($^+Na^{++}$, K^{++}, Cl^-), osmolality, pH, and a laxative screen (SO_4^{2-}, PO_4^{2-}, Mg^{2+}). Stool or urine can be analyzed for emetine (a component of ipecac), bisacodyl, castor oil, or anthraquinone.

Carbohydrate malabsorption lowers stool pH because of colonic fermentation of carbohydrate to short-chain fatty acids. Stool pH less than 5.3 usually means pure carbohydrate malabsorption, whereas in the generalized malabsorptive diseases, stool pH is greater than 5.6 and usually greater than 6.0.

The normal stool osmotic gap, which is the difference between stool osmolality (or 290 mOsm) and twice the stool Na^{++} and K^{++} concentrations, is 50 to 125. In secretory diarrheas, the colon's capacity for adjusting electrolyte concentrations is overwhelmed, the stool osmotic gap is less than 50, and stool electrolytes more nearly resemble plasma electrolytes ($^+Na^{++}$ concentrations are usually > 90 mmol/L, K^+ concentrations usually < 10 mmol/L), except for higher HCO_3^- concentrations (usually > 50 mmol/L). In osmotic diarrhea, the presence of uncharged solute or unmeasured cation in the colonic lumen draws in water, depresses stool Na^{++} (usually < 60 mmol/L) and K^+ concentrations, and results in a stool osmotic gap greater than 125. Stools with Na^{++} concentrations between 60 and 90 mmol/L and calculated osmotic gaps between 50 and 100 can result from either secretory or malabsorptive abnormalities. Patients with Mg^{2++}-induced diarrhea may be diagnosed by fecal Mg^{2+} values of more than 50 mmol/L. Sodium anion–induced diarrheas (Na_2SO_4, Na_2PO_4) mimic secretory diarrhea because the stool Na^{++} content is high (>90 mmol/L) and there is no osmotic gap; this diarrhea may be diagnosed by determining stool Cl^- concentration because these anions displace stool Cl^-, resulting in a depressed stool Cl^- value (usually < 20 mmol/L). A low stool osmolality suggests contamination of stool with urine or water in the case of factitious diarrhea.

SPECIFIC CAUSES OF CHRONIC DIARRHEA

Prolonged, Persistent Infectious Diarrheas

Prolonged infectious diarrheas (>2 weeks) may be due to persistent or recurrent infections. These diarrheas occur most commonly in children exposed to unsafe drinking water in developing countries, patients who have acquired immunodeficiency syndrome (AIDS) or are immunosuppressed for other reasons, and recent travelers. The most common causes in children in developing countries are enteropathogenic and enteroadherent *E. coli* infections (Chapter 304). Other common organisms include *Giardia* (Chapter 351), *Cryptosporidium* (Chapter 350), *Entamoeba* (Chapter 352), *Isospora* (Chapter 390), and microsporidia (Chapter 350). Recurrent or prolonged infectious diarrhea may lead to severe malnutrition and death (mortality rate, 50%). Treatment includes nutrition support with supplemental vitamin A (200,000 IU twice yearly) and zinc (20 mg elemental daily for 14 days). Severe disease may require total parenteral nutrition.

In patients with AIDS, protracted diarrhea may be caused by treatable agents such as *E. histolytica, Giardia,* or *Strongyloides* or by organisms such as *Cryptosporidium, Isospora belli*, and microsporidia that are difficult to treat or untreatable. The most effective treatment is retroviral therapy to improve the immune system (Chapter 388).

Up to 10% of travelers returning from developing countries have infectious diarrhea that persists for longer than 3 to 4 weeks. Stool should be examined for culture and for ova and parasites; in patients with a recent history of antibiotic use, stool also should be sent for *C. difficile* toxin. Any specific organisms that are identified should be treated. If treatment with trimethoprim-sulfamethoxazole or a fluoroquinolone has been unsuccessful, tetracycline (250 mg PO four times daily for 7 to 10 days) or metronidazole (250 mg PO three times daily for 7 to 10 days) can be tried. After documented infectious diarrhea, 25% of patients experience pain, bloating, urgency, a sense of incomplete evacuation, and loose stools for 6 months or longer; some of these patients have celiac disease, so screening (see later) is warranted in this setting. When no other cause is found, these patients are deemed to have postinfectious irritable bowel syndrome (Chapter 137).

Sporadic outbreaks of severe, prolonged diarrhea, often greater than 1 year in duration, occasionally have been reported. This form of prolonged diarrhea is called *Brainerd diarrhea*. The organism has yet to be identified. The diarrhea is difficult to treat; cholestyramine (4 g PO three times daily) may be helpful.

Malabsorptive Syndromes

Malabsorption is caused by many different diseases, drugs (e.g., the lipase inhibitor orlistat; Chapter 220), and nutritional products (the nonabsorbable fat olestra) that impair intraluminal digestion, mucosal absorption, or delivery of the nutrient to the systemic circulation (E-Fig. 140-2; see Table 140-4). Dietary fat is the nutrient most difficult to absorb. Fatty stools (steatorrhea) are the hallmark of malabsorption; a stool test for fat is the best screening test. Malabsorption does not always cause diarrhea. Clinical signs of vitamin or mineral deficiencies may occur in the absence of diarrhea. A careful history is crucial in guiding further testing to confirm the suspicion of malabsorption and to make a specific diagnosis (Fig. 140-4). The goals of treatment are to correct or treat the underlying disease and replenish losses of water, electrolytes, and nutrients.

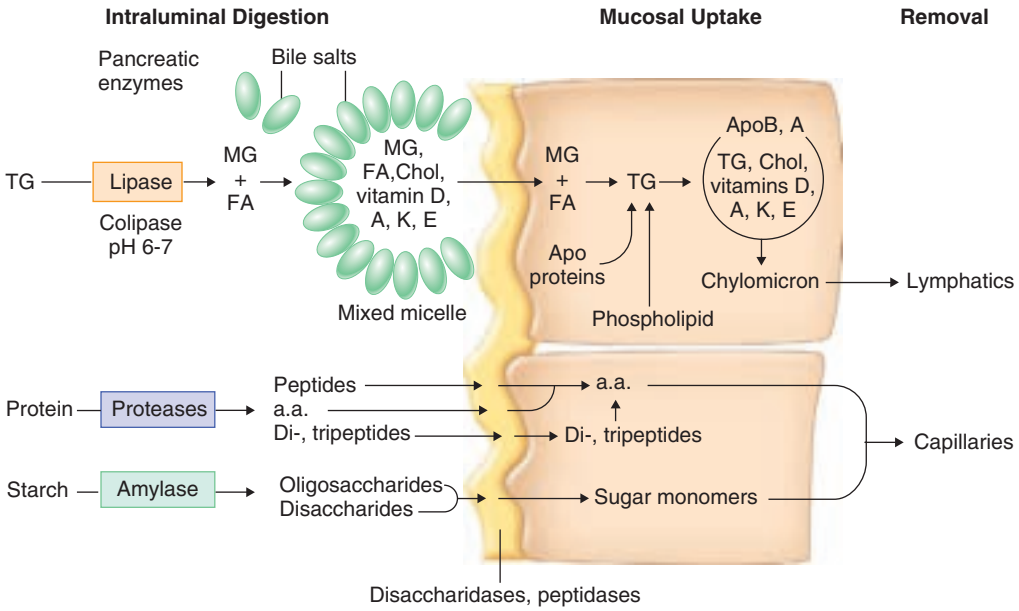

E-FIGURE 140-2. **Phases of intestinal digestion and absorption of dietary fat, protein, and carbohydrate. a.a. = Amino acids; ApoB, A = apolipoproteins B and A; Chol = cholesterol; FA = fatty acids; MG = monoglycerides; TG = triglycerides.**

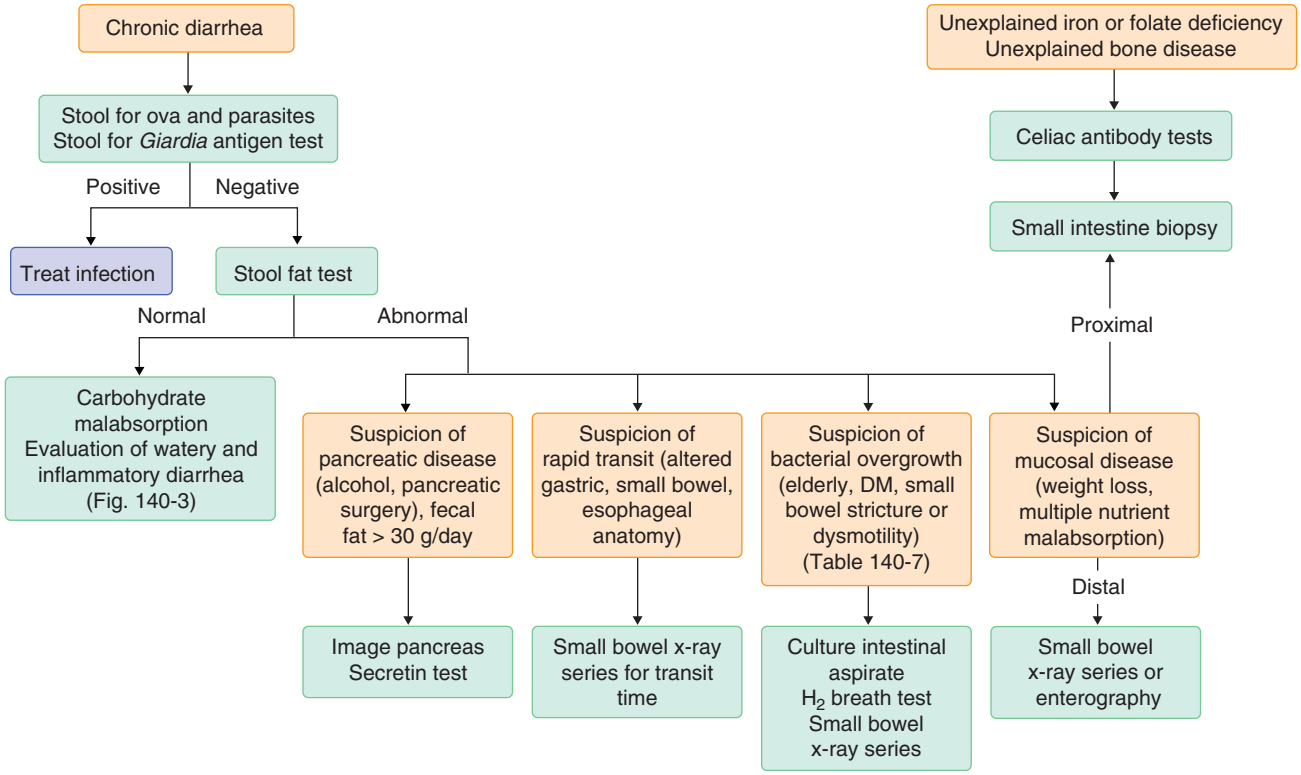

FIGURE 140-4. Approach to the diagnosis of malabsorption. DM = diabetes mellitus.

Conditions That Impair Intraluminal Digestion

Most digestion and absorption of nutrients occur in the small intestine (see E-Fig. 140-2). Carbohydrates and most dietary proteins are water soluble and readily digested by pancreatic enzymes. Pancreatic proteases (trypsinogen, chymotrypsinogen, procarboxypeptidases) are secreted from acinar cells in inactive forms. The cleavage of trypsinogen to trypsin by the duodenal brush-border peptidase enteropeptidase (enterokinase) allows trypsin to cleave the remaining trypsinogen and other proteases to their active form.

Most dietary lipids (long-chain triglycerides, cholesterol, and fat-soluble vitamins) are water insoluble and must undergo lipolysis and incorporation into mixed micelles before they can be absorbed across the intestinal mucosa. Pancreatic lipase, in the presence of its cofactor, colipase, cleaves long-chain triglycerides into fatty acids and monoglycerides. The products of lipolysis interact with bile salts and phospholipids to form mixed micelles, which also incorporate cholesterol and fat-soluble vitamins (D, A, K, and E) in their hydrophobic centers. Bicarbonate secreted from pancreatic duct cells is physiologically important to neutralize gastric acid because pancreatic enzyme activity and bile salt micelle formation are optimum at a luminal pH of 6 to 8.

IMPAIRED MIXING

Surgical alterations, such as partial gastrectomy with gastrojejunostomy (Billroth II anastomosis) or gastrointestinal bypass surgeries for obesity, result in the release of biliary and pancreatic secretions into the intestine at a site remote from the site of entry of gastric contents. This imbalance can result in impaired lipolysis and impaired micelle formation, with subsequent fat malabsorption. Bypass of the duodenum also impairs absorption of iron, folate, and calcium. Rapid transit through the jejunum contributes to the malabsorption of nutrients. Individuals with these conditions also have surgical anastomoses that predispose to bacterial overgrowth.

DUMPING SYNDROME

After esophageal (distal esophagectomy, myomectomy for achalasia), gastric (Nissan wrap, hiatal hernia repair, gastrojejunostomy), and bariatric (Roux-en-Y and duodenal switch gastric bypass) surgeries, the unregulated delivery of concentrated sugars and food into the duodenum and jejunum results in altered insulin regulation, maldigestion, osmotic movement of fluid into the intestinal lumen, and rapid transit such that intestinal contact time is insufficient for absorption of nutrients.

Treatment is with a diet that is low in concentrated sugars divided into six small meals. Administration of pectin (15 g with each meal) may slow gastric emptying. In patients who are refractory to dietary measures, a short-acting somatostatin analogue (e.g., octreotide, 25 to 200 μg SC three times daily) or the better tolerated intramuscular preparation (10 to 20 mg monthly) improves dumping symptoms. In patients with predominant reactive hypoglycemia 1 to 3 hours after a meal (late dumping), an α-glycosidase hydrolase inhibitor (e.g., acarbose, 50 to 100 mg PO three times daily) that blocks carbohydrate absorption in the small bowel may be beneficial. Continuous tube feeding is also effective.

IMPAIRED LIPOLYSIS

A deficiency in pancreatic lipase may be caused by the congenital absence of pancreatic lipase or by destruction of the pancreatic gland as a result of alcohol-related pancreatitis, cystic fibrosis, or pancreatic cancer. Pancreatic lipase also can be denatured by excess secretion of gastric acid (e.g., Zollinger-Ellison syndrome; Chapter 195). In such cases, lipase denaturation can be offset by treatment with a high-dose proton pump inhibitor (e.g., omeprazole 60 mg/day PO) to block acid secretion.

CHRONIC PANCREATITIS

Chronic pancreatitis (Chapter 144) is the most common cause of pancreatic insufficiency and impaired lipolysis. In the United States, chronic pancreatitis most commonly results from alcohol abuse; in contrast, tropical (nutritional) pancreatitis is most common worldwide. Malabsorption of fat does not occur until more than 90% of the pancreas is destroyed.

CLINICAL MANIFESTATIONS

Individuals with pancreatic causes of malabsorption typically present with bulky, fat-laden stools (usually > 30 g of fat daily), abdominal pain, and diabetes, although some present with diabetes in the absence of gastrointestinal symptoms. Stools usually are not watery because undigested triglycerides form large emulsion droplets with little osmotic force and, in contrast to fatty acids, do not stimulate water and electrolyte secretion in the colon. Deficiency of fat-soluble vitamins is seen only rarely, presumably because gastric and residual pancreatic lipase generates enough fatty acids for some micelle formation. In severe disease, subclinical protein malabsorption, manifested by the presence of undigested meat fibers in the stool, and subclinical carbohydrate malabsorption, manifested by gas-filled, floating stools, can occur. Weight loss, when it occurs, is most often caused by decreased oral intake to

avoid abdominal pain or diarrhea and less commonly by malabsorption. Pancreatic enzyme replacement and analgesics are the mainstays of treatment for chronic pancreatitis (Table 144-5 in Chapter 144).

In the dumping syndrome, patients may present with severe diarrhea, malabsorption, abdominal cramping, gas, and weight loss. Some have associated sweatiness, dizziness, and altered cognition because of postprandial hypoglycemia.

DIAGNOSIS

Between 30 and 40% of individuals with alcohol-related chronic pancreatitis have calcifications on abdominal radiographs. A qualitative or quantitative test for fecal fat is positive in individuals whose pancreas is more than 90% destroyed. Noninvasive tests of pancreatic function are not sensitive enough to detect mild to moderate insufficiency, so the secretin stimulation test is preferred (see Table 140-6) if it can be obtained. A modified oral glucose tolerance test that shows late (120 to 180 minutes) hypoglycemia and an early (30 minutes) rise in hematocrit with an increased pulse rate suggests the dumping syndrome in patients with consistent symptoms. A small-bowel barium study to assess transit time may be helpful in the diagnosis.

IMPAIRED MICELLE FORMATION

Bile salt concentrations in the intestinal lumen can fall to less than the critical concentration (2 to 3 mmol/L) needed for micelle formation because of decreased bile salt synthesis (severe liver disease), decreased bile salt delivery (cholestasis), or removal of luminal bile salts (bacterial overgrowth, terminal ileal disease or resection, cholestyramine therapy, acid hypersecretion). Fat malabsorption resulting from impaired micelle formation is generally not as severe as malabsorption resulting from pancreatic lipase deficiency, presumably because fatty acids and monoglycerides can form lamellar structures, which to a certain extent can be absorbed. Malabsorption of fat-soluble vitamins (D, A, K, and E) may be marked, however, because micelle formation is required for their absorption.

Decreased Bile Salt Synthesis and Delivery

Malabsorption can occur in individuals with cholestatic liver disease or bile duct obstruction. The clinical consequences of malabsorption are seen most often in women with primary biliary cirrhosis because of the prolonged nature of the illness. Although these individuals can present with steatorrhea, osteoporosis or, less commonly, osteomalacia is the most common presentation. The cause of bone disease in these patients is poorly understood and often is not related to vitamin D deficiency. Bone disease is treated with calcium supplements (and vitamin D if a deficiency is documented), weight-bearing exercise, and a bisphosphonate (e.g., alendronate, 10 mg/day PO or 70 mg PO once weekly).

Intestinal Bacterial Overgrowth

In health, only small numbers of lactobacilli, enterococci, gram-positive aerobes, or facultative anaerobes can be cultured from the upper small bowel lumen. Motility and acid are the most important factors in keeping the number of bacteria in the upper small bowel low. Any condition that produces local stasis or recirculation of colonic luminal contents allows development of a predominantly "colonic" flora (coliforms and anaerobes, such as *Bacteroides* and *Clostridium*) in the small intestine (see Table 140-7). Anaerobic bacteria cause impaired micelle formation by releasing cholylamidases, which deconjugate bile salts. The unconjugated bile salts, with their higher pK_a, are more likely to be in the protonated form at the normal upper small intestinal pH of 6 to 7 and can be absorbed passively. As a result, the concentration of bile salts decreases in the intestinal lumen and can fall to less than the critical micellar concentration, causing malabsorption of fats and fat-soluble vitamins. Vitamin B_{12} deficiency and carbohydrate malabsorption also can occur with generalized bacterial overgrowth. Anaerobic bacteria ingest vitamin B_{12} and release proteases that degrade brush-border disaccharidases. Although anaerobic bacteria use vitamin B_{12}, they synthesize folate. Individuals with bacterial overgrowth usually have low serum vitamin B_{12} levels but normal or high folate levels; this helps distinguish bacterial overgrowth from tropical sprue, in which vitamin B_{12} and folate levels are usually low because of decreased mucosal uptake.

CLINICAL MANIFESTATIONS

Individuals with bacterial overgrowth can present with diarrhea, abdominal cramps, gas and bloating, weight loss, and signs and symptoms of vitamin B_{12} and fat-soluble vitamin deficiency. Watery diarrhea occurs because of the osmotic load of unabsorbed carbohydrates and stimulation of colonic secretion by unabsorbed fatty acids.

DIAGNOSIS

The diagnosis of bacterial overgrowth should be considered in elderly people and in individuals with predisposing underlying disorders (see Table 140-7).[8] Bacterial overgrowth may be associated with the irritable bowel syndrome (Chapter 137). The identification of greater than 10^5 CFU/mL in a culture of small intestinal aspirate is the gold standard in diagnosis but is not readily available. The noninvasive tests with a sensitivity and specificity comparable to those of intestinal culture are the glucose hydrogen and methane breath test; in individuals with low vitamin B_{12} levels, a Schilling test before and after antibiotic therapy can be diagnostic if available (Chapter 164).

TREATMENT Rx

The goals of treatment are to correct the structural or motility defect, if possible; to eradicate offending bacteria; and to provide nutritional support. Acid-reducing agents should be stopped, if possible. Treatment with antibiotics should be based on culture results whenever possible; otherwise, empirical treatment is given. Rifaximin (400 mg PO three times daily) is effective,[■] but less so in individuals with an excluded (blind) intestinal loop. Tetracycline (250 to 500 mg PO four times daily) or a broad-spectrum antibiotic against aerobes and enteric anaerobes (ciprofloxacin, 500 mg PO twice daily; amoxicillin-clavulanic acid, 250 to 500 mg PO three times daily; cephalexin, 250 mg PO four times daily with metronidazole, 250 mg PO three times daily) should be given for 14 days. Prokinetic agents such as metoclopramide (10 mg PO four times daily) or erythromycin (250 to 500 mg PO four times daily) can be tried to treat small bowel motility disorders, but often they are not efficacious. Octreotide (50 µg SC every day) may improve motility and reduce bacterial overgrowth in individuals with scleroderma. If the structural abnormality or motility disturbance cannot be corrected, the patient is at risk for malnutrition and deficiencies of vitamin B_{12} and fat-soluble vitamins. Cyclic treatment (1 to 3 weeks of every 4 to 6 weeks) with rotating antibiotics may be required in these patients to prevent recurrent bouts of bacterial overgrowth. If supplemental calories are needed, medium-chain triglycerides should be given because they do not depend on micelle formation for their absorption. Monthly treatment with vitamin B_{12} should be considered, along with supplemental vitamins D, A, K, and E and calcium.

ILEAL DISEASE

Disease of the terminal ileum is most commonly due to Crohn disease (Chapter 141), which also may lead to ileal resection, but it also can be caused by radiation enteritis, tropical sprue, tuberculosis, *Yersinia* infection, or idiopathic bile salt malabsorption. These diseases cause bile salt wasting in the colon.

The clinical consequences of bile salt malabsorption are related directly to the length of the diseased or resected terminal ileum. In an adult, if less than 100 cm of ileum is diseased or resected, watery diarrhea results because of stimulation of colonic fluid secretion by unabsorbed bile salts. Bile acid diarrhea responds to cholestyramine (2 to 4 g taken at breakfast, lunch, and dinner).[9] If more than 100 cm of ileum is diseased or resected, bile salt losses (>3 g/day) in the colon exceed the capacity for increased bile salt synthesis in the liver, the bile salt pool shrinks, and micelle formation is impaired. As a result, steatorrhea ensues, and fatty acid–induced intestinal secretion synergizes with the bile acid–induced secretion to cause diarrhea. Treatment is with a low-fat diet, vitamin B_{12} (300 to 1000 µg SC once every month or 2 mg/day PO), dietary supplements of calcium (500 mg PO two or three times daily, monitor 24-hour urine calcium for adequacy of dose), and a multiple vitamin and mineral supplement. An antimotility agent should be given for diarrhea. Bile salt binders may worsen diarrhea. Screening for fat-soluble vitamin deficiencies (vitamins A and E, 25-OH vitamin D, and prothrombin time) and bone disease (bone densitometry, serum calcium, intact parathyroid hormone, 24-hour urine for calcium) should be done.

Three long-term complications of chronic bile salt wasting and fat malabsorption are renal stones, bone disease (osteoporosis and osteomalacia), and gallstones. Oxalate renal stones occur as a consequence of excess free oxalate absorption in the colon. Free oxalate is generated when unabsorbed fatty acids bind luminal calcium, which is then unavailable for binding oxalate. Renal oxalate stones sometimes can be avoided with a low-fat, low-oxalate diet and calcium supplements. Bone disease is caused by impaired micelle formation with a resulting decrease in absorption of vitamin D; year-round sun exposure

reduces this complication. Vitamin D (50,000 U PO one to three times per week) and calcium supplements (500 mg PO two to three times daily day) should be given to susceptible individuals, but vitamin D levels and serum and urinary calcium must be monitored for response to treatment because excess vitamin D can be toxic. The mechanism of gallstone formation in these individuals is unclear; pigmented gallstones are most common.

Conditions That Impair Mucosal Absorption

PATHOBIOLOGY

Nutrients are absorbed along the entire length of the small intestine, with the exception of iron and folate, which are absorbed in the duodenum and proximal jejunum, and bile salts and cobalamin, which are absorbed in the distal ileum. The efficiency of nutrient uptake at the mucosa is influenced by the number of villus absorptive cells, the presence of functional hydrolases and specific nutrient transport proteins on the brush-border membrane, and transit time. Transit time determines the contact time of luminal contents with the brush-border membrane and influences the efficiency of nutrient uptake across the mucosa.

Mucosal malabsorption can be caused by specific (usually congenital) brush-border enzyme or nutrient transporter deficiencies or by generalized diseases that damage the small intestinal mucosa or result in surgical resection or bypass of the small intestine. The nutrients malabsorbed in these general malabsorptive diseases depend on the site of intestinal injury (proximal, distal, or diffuse) and the severity of damage. The main mechanism of malabsorption in these conditions is a decrease in surface area available for absorption. Some conditions (infection, celiac disease, tropical sprue, food allergies, and GVHD) are characterized by intestinal inflammation and villus flattening; others are characterized by ulceration (ulcerative jejunitis, NSAIDs, Crohn disease), infiltration (amyloidosis), or ischemia (radiation enteritis, mesenteric ischemia).

Long-chain fatty acids are transported across the microvillus membrane of villus epithelial cells by the fatty acid transport protein FATP4. The bile salts from mixed micelles remain in the intestinal lumen and are absorbed in the distal ileum by sodium-dependent co-transport. Oligosaccharides and larger oligopeptides (products of pancreatic enzyme digestion), sucrose, and lactose are hydrolyzed further by enzymes present in the brush-border membrane of villus epithelial cells before they are absorbed. Although only sugar monomers (glucose, galactose, fructose) can be taken up at the apical epithelial cell membrane, dipeptides and tripeptides are readily taken into the cell.

Water-soluble vitamins are readily absorbed throughout the small intestine. Fat-soluble vitamins, minerals, and cobalamin are more difficult to absorb because of the requirement for micelle formation (vitamins D, A, K, and E), a divalent charge (magnesium, calcium, iron), or selected sites of uptake in the intestine (iron, cobalamin). Calcium is absorbed best in the proximal small intestine by a vitamin D–dependent calcium channel (TRPV6). Magnesium is absorbed in the small intestine by members of the transient receptor potential family (TRPM6 and TRPM7). Mutations in TRPM6 have been identified in the rare disorder hereditary hypomagnesemia. Ferrous iron is transported into intestinal epithelial cells by a proton-coupled metal-ion transporter (Nramp2) that has specificity for Fe^{2++} and other divalent cations (Zn^{2++}, Mn^{2++}, Co^{2++}, Cd^{2++}, Cu^{2++}, Ni^{2++}, and Pb^{2++}). The absorption of calcium and nonheme iron is enhanced by solubilization with hydrochloric acid. Intraluminal compounds such as oxalate, phytate, and long-chain fatty acids bind to calcium and magnesium, decreasing their absorption. Individuals with severe mucosal disease or short-bowel syndrome with high fecal fluid outputs lose magnesium and zinc from endogenous secretions.

Folates (Chapters 164 and 218) are both taken in the diet and produced by bacteria in the colon. Dietary folates are absorbed in the proximal small intestine through a reduced folate carrier (RFC1). Deficiency can be caused by poor intake or malabsorption secondary to intestinal disease or drugs. The cobalamins (Chapters 164 and 218) are abundant in foods containing animal proteins (e.g., meat, seafood, eggs, milk). Cobalamin (vitamin B_{12}) deficiency in industrialized countries is rarely due to poor dietary intake but rather reflects the inability to absorb cobalamin. This inability may be caused by a lack of intrinsic factor, consumption of cobalamin by overgrowth of anaerobic bacteria in the small bowel lumen, ileal disease or resection, or defective transcobalamin II. Large amounts of cobalamin are present in the liver (2 to 5 mg), and cobalamin is reabsorbed from bile through the enterohepatic circulation, thereby limiting daily losses to less than 1 μg. It usually takes 10 to 12 years for cobalamin deficiency to develop after it is eliminated from the diet, but deficiency can occur more rapidly (2 to 5 years) with malabsorptive syndromes. If lack of gastric acid causes food-cobalamin malabsorption, treatment with oral cyanocobalamin supplementation (Chapter 164) is curative.

LACTASE DEFICIENCY

EPIDEMIOLOGY

Acquired lactase deficiency is the most common cause of selective carbohydrate malabsorption. Most individuals, except those of northern European descent, begin to lose lactase activity by the age of 2 years. The prevalence of lactase deficiency is highest (85 to 100%) in persons of Asian, African, and Native-American descent.

PATHOBIOLOGY

The persistence or nonpersistence of lactase activity is associated with a single nucleotide polymorphism C/T_{-13910} that is found upstream of the lactase gene on chromosome 2q21-22. Hypolactasia is associated with the C/C_{-13910} genotype in diverse ethnic groups. The mechanism by which this variant downregulates the lactase gene is not known, but functional studies suggest genotype-dependent alterations in levels of messenger RNA.

Clinical Manifestations

Adults with lactase deficiency typically complain of gas, bloating, and diarrhea after the ingestion of milk or dairy products but do not lose weight. Unabsorbed lactose is osmotically active, drawing water followed by ions into the intestinal lumen. On reaching the colon, bacteria metabolize lactose to short-chain fatty acids, carbon dioxide, and hydrogen gas. Short-chain fatty acids are transported with sodium into colonic epithelial cells, facilitating the reabsorption of fluid in the colon. If the colonic capacity for the reabsorption of short-chain fatty acids is exceeded, an osmotic diarrhea results (see later discussion of carbohydrate malabsorption in watery diarrheas).

Diagnosis

The diagnosis of acquired lactase deficiency can be made by empirical treatment with a lactose-free diet, which results in resolution of symptoms; by the hydrogen breath test after oral administration of lactose; or by genetic testing.[10] Many intestinal diseases cause secondary reversible lactase deficiency, including viral gastroenteritis, celiac disease, giardiasis, and bacterial overgrowth.

CONGENITAL ENTEROPEPTIDASE (ENTEROKINASE) DEFICIENCY

Enteropeptidase is a brush-border protease that cleaves trypsinogen to trypsin, triggering the cascade of pancreatic protease activation in the intestinal lumen. The rare congenital deficiency of enteropeptidase results in inability to activate all pancreatic proteases and leads to severe protein malabsorption. It manifests in infancy as diarrhea, growth restriction, and hypoproteinemic edema.

ABETALIPOPROTEINEMIA

Formation and exocytosis of chylomicrons at the basolateral membrane of intestinal epithelial cells are necessary for the delivery of lipids to the systemic circulation. One of the proteins required for assembly and secretion of chylomicrons is the microsomal triglyceride transfer protein, which is mutated in individuals with abetalipoproteinemia. Children with this disorder have fat malabsorption and the consequences of vitamin E deficiency (retinopathy and spinocerebellar degeneration). Biochemical tests show low plasma levels of apoprotein B, triglyceride, and cholesterol. Membrane lipid abnormalities result in red blood cell acanthosis (burr cells). Intestinal biopsy is diagnostic; the tissue is characterized by engorgement of epithelial cells with lipid droplets. Calories are provided by treatment with a low-fat diet containing medium-chain triglycerides. Poor absorption of long-chain fatty acids sometimes can result in essential fatty acid deficiency. High doses of fat-soluble vitamins, especially vitamin E, often are needed. Mutations in the apolipoprotein B gene (hypobetalipoproteinemia) and intracellular retention of chylomicrons (Anderson disease) cause a similar although less severe clinical syndrome with rare fat malabsorption.

CELIAC DISEASE

DEFINITION AND EPIDEMIOLOGY

Celiac disease is an inflammatory condition of the small intestine precipitated by the ingestion of wheat, rye, and barley in individuals with certain genetic predispositions.[11] Screening studies for the antiendomysial (EMA) and

anti–tissue transglutaminase (anti-tTG) antibodies that are associated with celiac disease suggest a prevalence in white populations of approximately 1%. High-risk groups for celiac disease include first-degree relatives and individuals with type 1 diabetes mellitus, autoimmune thyroid disease, primary biliary cirrhosis, Turner syndrome, or Down syndrome. Approximately 20% of patients diagnosed with irritable bowel syndrome or with microscopic (lymphocytic) colitis have celiac disease.

PATHOBIOLOGY

Environmental and genetic factors are important in the development of celiac disease. Approximately 15% of first-degree relatives of affected individuals are found to have celiac disease. Predisposition has been mapped to the human leukocyte antigen (HLA)-D region on chromosome 6. More than 90% of northern Europeans with celiac disease have the DQ2 heterodimer encoded by alleles DQA1*0501 and DQB1*0201, compared with 20 to 30% of controls. A smaller celiac group carries HLA DQ8. Many non-HLA alleles identified in genome-wide association studies account for a small portion of genetic risk. Most such genes are involved in adaptive and innate immune responses. Overlap variants have been identified in diabetes, rheumatoid arthritis, and Crohn disease.

The alcohol-soluble protein fraction of wheat gluten, the gliadins, and similar prolamins in rye and barley trigger intestinal inflammation in susceptible individuals. Oat grains, which have prolamins rich in glutamine but not proline, are rarely toxic. Gliadins and similar prolamins with high proline content are relatively resistant to digestion by human proteases. Many gliadin and prolamin peptides have been identified that stimulate HLA-DQ2 and DQ8 restricted intestinal T-cell clones from individuals with celiac disease. In blood lymphocyte comprehensive screen assays, three immunodominant prolamin peptides have been identified from wheat, barley, and rye in DQ2 celiac individuals; dominant peptides differ in those with DQ8. The DQ2 protein expressed on antigen-presenting cells has positively charged binding pockets; tTG (the autoantigen recognized by EMA) may enhance intestinal inflammation by deamidation of select glutamine residues in gliadin and similar prolamins to negatively charged glutamic acid. In the deamidated form, most gliadin peptides have a higher binding affinity for DQ2 and are more potent stimulants of gluten-sensitized T cells. Villous atrophy may be caused by inflammation that is triggered by γ-interferon released from DQ2- or DQ8-restricted CD4 T cells in the lamina propria. Alternatively, intraepithelial lymphocytes may directly kill intestinal epithelial cells under the influence of IL-15 released from stressed enterocytes.

CLINICAL MANIFESTATIONS

Celiac disease usually manifests early in life, at approximately 2 years of age (after wheat has been introduced into the diet), or later in the second to fourth decades of life, but it can occur at any age.[12] It may first manifest clinically after abdominal surgery or an episode of infectious diarrhea.

Breast-feeding and the time of introduction of wheat in the diet may lessen the risk or delay the onset of celiac disease in infants at risk. Adults with celiac disease in the United States often present with anemia or osteoporosis without diarrhea or other gastrointestinal symptoms. These individuals most likely have proximal disease that impairs iron, folate, and calcium absorption but an adequate surface area in the remaining intestine for absorption of other nutrients. Other extraintestinal manifestations of celiac disease include rash (dermatitis herpetiformis), neurologic disorders (peripheral neuropathy, ataxia, epilepsy), psychiatric disorders (depression, paranoia), reproductive disorders (infertility, spontaneous abortion), short stature, dental enamel hypoplasia, pancreatitis, chronic hepatitis, or cardiomyopathy.

Individuals with significant mucosal involvement present with watery diarrhea, weight loss or growth retardation, and the clinical manifestations of vitamin and mineral deficiencies. Cobalamin deficiency is more common (10% of patients) than previously thought and usually corrects itself on a gluten-free diet. Symptomatic individuals require supplementation of vitamin B_{12}. Diarrhea is caused by many mechanisms, including a decreased surface area for water and electrolyte absorption, the osmotic effect of unabsorbed luminal nutrients, an increased surface area for chloride secretion (crypt hyperplasia), and the stimulation of intestinal fluid secretion by inflammatory mediators and unabsorbed fatty acids. Some individuals have impaired pancreatic enzyme secretion caused by decreased mucosal cholecystokinin release or bacterial overgrowth that may contribute to diarrhea. Individuals with nonceliac gluten sensitivity have wheat-related intestinal and extraintestinal symptoms similar to those of celiac disease but lack intestinal inflammation or celiac serologic markers. Fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAPs) or gluten may be the offending agent.

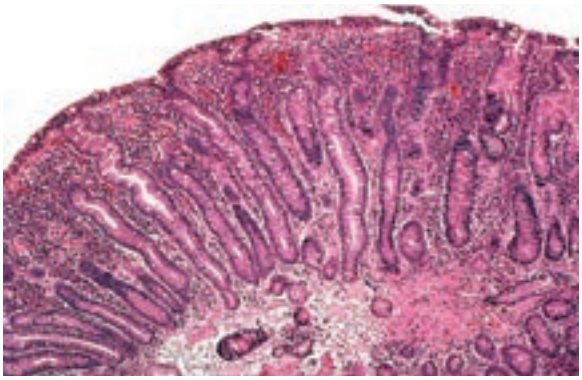

FIGURE 140-5. Intestinal biopsy appearance of flattened villi, hyperplastic crypts, and increased intraepithelial lymphocytes. (Courtesy John Hart, MD.)

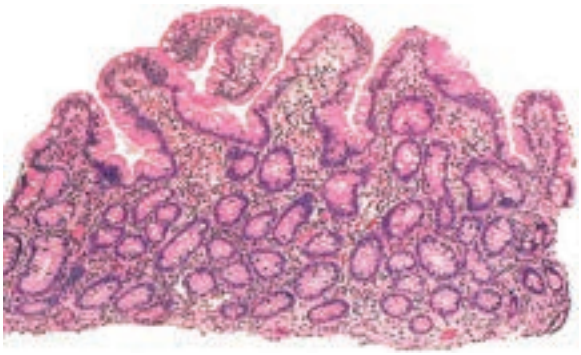

FIGURE 140-6. Regeneration of villi after initiation of a gluten-free diet. (Courtesy of John Hart, MD.)

DIAGNOSIS

Anti-tTG IgA antibody testing, when obtained with a serum IgA level, is a cost-effective strategy for screening high-risk groups; very high titers of the anti-tTG IgA and EMA antibodies are virtually diagnostic of celiac disease. EMA immunoglobulin A (IgA) antibodies, detected by indirect immunofluorescence, are highly sensitive (90%) and specific (95 to 99%) for active celiac disease in skilled laboratory testing. An enzyme-linked immunosorbent assay (ELISA) test to detect antibodies against tTG has equal sensitivity to the EMA test but is less specific. The anti-deamidated gliadin (a biotinylated synthetic γ-gliadin peptide with glutamic acid substituted for glutamine) IgA and IgG antibody immunofluorometric assay has a sensitivity and specificity that approaches that of anti-tTG IgA antibody.[13]

Patients with mild disease may have negative antibody studies. Anti-tTG, gliadin peptide, and EMA IgA antibodies tests are negative in individuals with selective IgA deficiency (present in up to 2.6% of individuals with celiac disease). In these patients, anti-tTG or gliadin peptide IgG antibodies may be helpful in diagnosis and monitoring. In equivocal cases (negative serologic findings and equivocal biopsy result or positive serologic findings and normal biopsy result), HLA genotyping is useful to exclude the diagnosis of celiac disease in persons who lack the DQ2 or DQ8 gene.

The diagnosis of celiac disease is confirmed by characteristic abnormalities seen on a small intestinal biopsy sample and improving when a gluten-free diet is instituted (Figs. 140-5 and 140-6). Biopsy is still required for diagnosis in most adults, who commonly present with atypical symptoms or are asymptomatic first-degree relatives detected by screening. In children, biopsy is not required if the patient has a greater than 5-fold increase in anti-tTG, a positive EMA serologic test, a DQ2 or DQ8 genotype, and typical gastrointestinal symptoms.[14] Mucosal flattening may be observed endoscopically as scalloped or reduced duodenal folds. Characteristic features found on intestinal biopsy include blunted or absent villi, crypt hyperplasia, increased intraepithelial lymphocytes, and infiltration of the lamina propria with plasma cells and lymphocytes. In some individuals, the only abnormal biopsy finding is increased intraepithelial lymphocytes. A hypoplastic mucosa indicates irreversible (end-stage) intestinal disease.

TABLE 140-8 VITAMIN AND MINERAL DOSES USED IN THE TREATMENT OF MALABSORPTION

VITAMIN	ORAL DOSE	PARENTERAL DOSE
Vitamin A*	Water-soluble A, 25,000 U/day†	
Vitamin E	Water-soluble E, 400-800 U/day†	
Vitamin D‡	25,000-50,000 U/day	
Vitamin K	5 mg/day	
Folic acid	1 mg/day	
Calcium§	1500-2000 mg elemental calcium/day Calcium citrate, 500 mg calcium/tablet† Calcium carbonate, 500 mg calcium/tablet†	
Magnesium	Liquid magnesium gluconate† 1-3 tbsp (12-36 mEq magnesium) in 1-2 L of ORS or sports drink sipped throughout the day Magnesium chloride hexahydrate† 100-600 mg elemental magnesium/day	2 mL of a 50% solution (8 mEq) both buttocks IM
Zinc	Zinc gluconate† 20-50 mg elemental zinc/day‖	
Iron	150-300 mg elemental iron/day Polysaccharide-iron complex† Iron sulfate or gluconate	Iron sucrose¶ Sodium ferric gluconate complex¶ Iron dextran (as calculated for anemia) (IV or IM¶; Chapter 159)
B-complex vitamins	1 megadose tablet/day	
Vitamin B_{12}	2 mg/day	1 mg IM or SC/mo**
Copper	copper sulfate 2-3 mg/day	1-2 mg IV/day

IM, intramuscularly; IV, intravenously; ORS = oral rehydration solution; SC = subcutaneously.
*Monitor serum vitamin A level to avoid toxicity, especially in patients with hypertriglyceridemia.
†Form best absorbed or with least side effects.
‡Monitor serum calcium and 25-OH vitamin D levels to avoid toxicity.
§Monitor 24-hr urine calcium to assess adequacy of dose.
‖If intestinal output is high, additional zinc should be given. Monitor for copper deficiency with high doses.
¶Parenteral therapy should be given in a supervised outpatient setting because of the risk for fatal reactions. Decreased risk for fatal reactions when compared with iron dextran.
**For vitamin B_{12} deficiency, 1 mg IM or SC twice per week for 4 wk, then once per month.

TREATMENT

Rx

Treatment consists of a lifelong gluten-free diet,[15] and even asymptomatic EMA-positive patients benefit. Wheat, rye, and barley grains should be excluded from the diet. Rice and corn grains are tolerated. Oats (if not contaminated by wheat grain) are tolerated by most. Early referral to a reputable celiac support group or website is often helpful in maintaining dietary compliance. Owing to secondary lactase deficiency, a lactose-free diet should be recommended until symptoms improve. Bone densitometry should be performed on all individuals with celiac disease because up to 70% have osteopenia or osteoporosis. Patients with diarrhea and weight loss should be screened for vitamin and mineral deficiencies. Documented deficiencies of vitamins and minerals should be replenished (Table 140-8), and women of childbearing age should take folic acid supplements. Bone mass often improves on a gluten-free diet alone. Patients with vitamin D or calcium deficiency should receive supplements (Chapter 218), with the dose monitored by 25-OH vitamin D levels and a 24-hour urine test for calcium.

PROGNOSIS

Of patients with celiac disease treated with a gluten-free diet, 90% experience symptomatic improvement within 2 weeks. The most common cause of a poor dietary response is continued ingestion of gluten. Other possibilities include a missed intestinal infection (see later), an alternative diagnosis (e.g., ARB use particularly in elderly patients with a sprue-like enteropathy but negative celiac serologic findings, agammaglobulinemia [diagnosed by hypogammaglobulinemia and lack of plasma cells on small bowel biopsy], autoimmune enteritis [diagnosed by a positive antienterocyte antibody and crypt apoptosis or loss of goblet cells on small bowel biopsy]), bacterial overgrowth, pancreatic insufficiency, microscopic colitis, or other food allergies (cow's milk, soy protein). Up to 40% of patients with celiac disease with symptomatic improvement have incomplete histologic recovery on a gluten-free diet; in such patients, a stricter diet may further improve symptoms and histology. In a small percentage of patients, symptoms and enteropathy do not improve despite a strict gluten-free diet. In such patients, repeat intestinal biopsy is indicated. Some patients will have collagen deposition beneath the surface epithelium (collagenous sprue) or a polyclonal population of intraepithelial lymphocytes (refractory celiac disease type I), Others will have ulcerative jejunitis or a monoclonal population of intraepithelial T cells with an aberrant phenotype or clonal T-cell receptor-γ gene rearrangements (refractory celiac disease type II), which are predictive of enteropathy-associated T-cell lymphoma (Chapter 185) that portends a poor prognosis. Video capsule endoscopy and device-assisted enteroscopy may be helpful in establishing these diagnoses. Patients with collagenous sprue, autoimmune enteritis, or refractory celiac disease type I often respond to prednisone (20 to 40 mg/day PO) or budesonide (9 mg PO daily).

Patients with celiac disease have a higher likelihood of having other autoimmune conditions, such as type 1 diabetes, thyroiditis, rheumatoid arthritis, inflammatory bowel disease, systemic lupus erythematosus, Sjögren syndrome, primary biliary cirrhosis, autoimmune hepatitis, vitiligo, and pancreatitis. Interestingly, approximately one third of patients with idiopathic sporadic ataxia have transglutaminase 6 antibodies, consistent with gluten-induced disease.

Individuals with celiac disease are at increased risk for B-cell lymphoma (Chapter 185),[16] gastrointestinal tract carcinomas (esophageal, small bowel, and colonic adenocarcinomas), and increased mortality; a strict gluten-free diet for life may lessen these risks. Intestinal T-cell lymphoma is rare and should be suspected in individuals who have abdominal pain, recurrence of symptoms after initial response to a gluten-free diet, or refractory celiac disease.

TROPICAL SPRUE

Tropical sprue is an inflammatory disease of the small intestine associated with the overgrowth of predominantly coliform bacteria. It occurs in residents or travelers to the tropics, especially India, Southeast Asia, Puerto Rico, and parts of the Caribbean. With the expansion of tourism and the global economy, this may be an under-recognized cause of enteropathy or mistaken for celiac disease.[17] Individuals classically present with diarrhea and megaloblastic anemia secondary to vitamin B_{12} and folate deficiency, but some have anemia only. Intestinal biopsy characteristically shows subtotal and patchy villous atrophy in the proximal and distal small intestine, which may be caused by the effect of bacterial toxins on gut structure or by the secondary effects of vitamin B_{12} deficiency on the gut (megaloblastic gut). Diagnosis is based on history, documentation of vitamin B_{12} or folate deficiency, and the presence of an abnormal small intestinal biopsy report. Treatment is a prolonged course of tetracycline (250 mg PO four times daily) or doxycycline (100 mg PO two times daily), folic acid (5 mg/day PO), and, with coexistent deficiency, vitamin B_{12} injections (1000 µg weekly) until symptoms resolve. Relapses or reinfection occurs in 20%, mainly in natives of the tropics.

GIARDIA LAMBLIA

Giardia lamblia (Chapter 351) infection, the most common protozoal infection in the United States, can cause malabsorption in individuals infected with many trophozoites, especially the immunocompromised or IgA-deficient hosts. Malabsorption occurs when many organisms cover the epithelium and cause mucosal inflammation, which results in villous flattening and a decrease in absorptive surface area. Stool for ova and parasites at this stage of infection is often negative because of the attachment of organisms in the proximal small intestine. Diagnosis can be made by a stool antigen-capture ELISA test but may require duodenal aspiration and biopsies.

HUMAN IMMUNODEFICIENCY VIRUS

Diarrhea, malabsorption, and wasting are common in individuals with AIDS but are seen less frequently with improved antiretroviral therapy (Chapter 390). In patients who are receiving highly active antiretroviral therapy, diarrhea is more likely to be due to protease inhibitors than to enteric infection.

Malabsorption is usually due to infection with cryptosporidia, *Mycobacterium avium-intracellulare* complex, *I. belli,* or microsporidia. An organism can

be identified by stool examination or intestinal biopsy approximately 50% of the time. *AIDS enteropathy* (a term used if no organism is identified) also can cause malabsorption. Mechanisms of malabsorption and diarrhea include villous atrophy, increased intestinal permeability, rapid small bowel transit (in patients with protozoal infection), and ultrastructural damage of enterocytes (in AIDS enteropathy). Among individuals with AIDS and diarrhea, results of fecal fat absorption are frequently abnormal. Serum albumin, vitamin B_{12}, and zinc levels are often low. Vitamin B_{12} deficiency is caused mainly by ileal disease, but low intrinsic factor and decreased transcobalamin II may be contributing factors. Management of malabsorption should focus on restoring the immune system by treating the underlying HIV infection with antiviral therapy. If possible, the offending organism should be treated with antibiotics. If the organism cannot be eradicated, chronic diarrhea and malabsorption result; treatment in these cases consists of antimotility agents and a lactose-free, low-fat diet. Pancreatic enzyme replacement therapy can be tried in HIV-infected individuals who are taking highly active antiretroviral therapy or nucleoside analogues and who have fat malabsorption of obscure origin. If supplemental calories are needed, liquid oral supplements that are predigested and high in medium-chain triglycerides (semi-elemental) are tolerated best. Vitamin and mineral deficiencies should be screened for and treated.

WHIPPLE DISEASE

Whipple disease (Chapters 142 and 275), a rare cause of malabsorption, manifests with gastrointestinal complaints in association with systemic symptoms, such as fever, joint pain, or neurologic manifestations.[18] Approximately one third of patients have cardiac involvement, most commonly culture-negative endocarditis. Occasionally, individuals present with ocular or neurologic disease without gastrointestinal symptoms. Men are affected more commonly than women, particularly white men. The organism responsible for causing Whipple disease is a gram-positive actinomycete, *Tropheryma whippelii.* The epidemiology and pathogenesis of Whipple disease are poorly understood. The prevalence of the disease is higher in farmers than in other workers, which suggests that the organism lives in the soil. Using polymerase chain reaction, *T. whippelii* has been detected in sewage and in duodenal biopsy specimens, gastric juice, saliva, and stool of individuals without clinical disease. Whether the latter represents a carrier state or the presence of nonpathogenic organisms is not known. Immunologic defects, IL-16, and an association with the HLA-B27 gene may be disease factors. Small intestinal biopsy shows villous blunting and infiltration of the lamina propria with large macrophages that stain positive with the periodic acid–Schiff method and are filled with the organism. It is important to distinguish these macrophages from macrophages infected with *M. avium-intracellulare* complex, which stain positive on acid-fast staining and are found in individuals with AIDS. Treatment is with a prolonged course of broad-spectrum antibiotics (e.g., ceftriaxone, 2 g/day intravenously (IV) or meropenem 1 g IV three times daily; then trimethoprim 160 mg and sulfamethoxazole 800 mg PO two times daily for 1 year or trimethoprim 160 mg and sulfamethoxazole 800 mg PO two times daily for 1 year). Relapses occur, but initial treatment with parenteral ceftriaxone or meropenem appears to be associated with a low relapse rate.

GRAFT-VERSUS-HOST DISEASE

Diarrhea occurs frequently after allogeneic bone marrow or stem cell transplantation (Chapter 178). Immediately after transplantation, diarrhea is caused by the toxic effects of cytoreductive therapy on the intestinal epithelium. From 20 to 100 days after transplantation, diarrhea is usually due to GVHD or infection. Patients with GVHD present clinically with a skin rash, hepatic cholestasis, buccal mucositis, anorexia, nausea, vomiting, abdominal cramps, and diarrhea. The diagnosis of GVHD in the gastrointestinal tract can be made on biopsy of the stomach, small intestine, or colon. In mild cases, the mucosa appears normal on inspection at endoscopy, but apoptosis of gastric gland or crypt cells can be found on biopsy. In severe cases, denudation of the intestinal epithelium results in diarrhea and malabsorption and often requires parenteral nutritional support. Octreotide (50 to 250 µg SC three times daily) may be helpful in controlling voluminous diarrhea. Treatment of GVHD is with steroids and antithymocyte globulin combined with parenteral nutritional support until intestinal function returns.

SHORT-BOWEL SYNDROME

Malabsorption caused by small bowel resection or surgical bypass is called the short-bowel syndrome. The most common causes in the United States are massive resection of the jejunum, owing to strangulated bowel, volvulus, or ischemia (mesenteric or after intra-abdominal surgery), and jejunal exclusion, owing to gastric bypass surgery. Short-bowel syndrome resulting from Crohn disease and radiation enteritis now is less common because of improved medical and radiation therapies. The severity of malabsorption depends on the site and extent of resection; the capacity for hyperplasia,[19] dilation, and elongation; and the function of the residual bowel. Mechanisms of malabsorption after small bowel resection include a decreased absorptive surface area, decreased luminal bile salt concentration, rapid transit, and bacterial overgrowth. Limited jejunal resection usually is tolerated best because bile salt and vitamin B_{12} absorption remain normal. Ileal resection is less well tolerated because of the consequences of bile salt wasting and the limited capacity of the jejunum to undergo adaptive hyperplasia. Adaptive hyperplasia in residual small bowel after resection depends on nutrients, endogenous secretions (pancreatic and biliary juice), local factors (trefoil peptides, prostaglandins, polyamines), growth hormone, and growth factors (epidermal growth factor [EGF], insulin-like growth factor-1 [IGF1], transforming growth factor-α [TGFα], interleukin 11 [IL11]). The glucagon-like peptide 2 (GLP2) produced in L cells in the terminal ileum and colon is a potent stimulant of adaptive hyperplasia in the jejunum in response to a meal, Using intestinal stem cell technology, epithelial organoids have been successfully grown in culture systems.

When less than 100 cm of jejunum remains, the colon takes on an important role in caloric salvage and fluid reabsorption. Malabsorbed carbohydrates are digested by colonic bacteria to short-chain fatty acids, which are absorbed in the colon.

TREATMENT

Rx

Parenteral nutrition may be avoided by a diet rich in complex carbohydrates, oral rehydration solutions, and acid-reducing and antimotility agents. In comparison, individuals with fewer than 100 cm of jejunum and no colon have high jejunostomy outputs and often require intravenous fluids or parenteral nutrition to survive. These individuals waste sodium, chloride, bicarbonate, magnesium, zinc, and water in their ostomy effluent. Dietary modifications should include a high-salt, nutrient-rich diet given in small meals. An oral rehydration solution with a sodium concentration greater than 90 mmol/L is absorbed best. Oral vitamin and mineral doses higher than the usual U.S. recommended daily allowances are required (Table 140-8). Vitamin B_{12} should be given parenterally (500 to 1000 µg SC every month). Magnesium deficiencies are often difficult to replenish with oral magnesium because of its osmotic effect in the intestinal lumen. A liquid magnesium preparation added to an oral rehydration solution and sipped throughout the day may minimize magnesium-induced fluid losses. Potent antimotility agents, such as tincture of opium (0.5 to 1 mL PO four times daily) or liquid morphine 20 mg/mL (1 mL PO four times daily), often are needed to slow transit and maximize contact time for nutrient absorption. High-volume jejunostomy outputs can be lessened by inhibiting endogenous secretions with a proton pump inhibitor (e.g., omeprazole, 40 mg PO one or two times daily, or lansoprazole, 30 mg PO one or two times daily) and, in severe cases, octreotide (100 to 250 µg SC three times daily; if effective, convert to an equivalent long-acting monthly dosage). The benefit of octreotide may be offset by its potential to inhibit intestinal adaptation and impair pancreatic enzyme secretion with doses greater than 300 µg/day.

In the most severe cases, supplemental calories must be provided by nocturnal tube feeding or parenteral nutrition. Treatment with growth hormone (0.1 mg/kg/day SC) with or without glutamine (30 g/day PO) for 4 weeks may reduce parenteral nutrition requirements in patients who have had massive intestinal resections. Teduglutide (0.05 mg/kg/day SC), a glucagon-like peptide-2 analogue that stimulates adaptive hyperplasia in remnant intestine after resection, reduces parenteral nutrition requirements.[20] Small bowel transplantation should be considered for individuals who require parenteral nutrition to survive and then develop progressive liver disease or venous access problems.[21]

PROGNOSIS

Long-term complications include bone disease, renal stones (oxalate stones if the colon is present, urate stones with a jejunostomy), gallstones, bacterial overgrowth, fat-soluble vitamin deficiencies, essential fatty acid deficiency, and D-lactic acidosis.

Conditions That Impair Nutrient Delivery to the Systemic Circulation

Insoluble lipids (present in chylomicrons) are exocytosed across the basolateral membrane of epithelial cells into the intestinal lymphatics. From there, they enter the mesenteric lymphatics and the general circulation through the thoracic duct. Sugar monomers, amino acids, and medium-chain fatty acids are transported across the basolateral membrane of intestinal epithelial cells into capillaries and into the portal circulation. Sugar monomers are

transported across the basolateral membrane by the facilitative glucose transporter isoform (GLUT2) and amino acids by facilitative amino acid carriers (see E-Fig. 140-1C).

IMPAIRED LYMPHATIC DRAINAGE

Diseases that cause intestinal lymphatic obstruction, such as primary congenital lymphangiectasia (malunion of intestinal lymphatics), and diseases that result in secondary lymphangiectasia (lymphoma, tuberculosis, Kaposi sarcoma, retroperitoneal fibrosis, constrictive pericarditis, severe heart failure) result in fat malabsorption. The increased pressure in the intestinal lymphatics leads to leakage and sometimes rupture of lymph into the intestinal lumen, with the loss of lipids, γ-globulins, albumin, and lymphocytes. The diagnosis of lymphangiectasia can be made by intestinal biopsy, but the specific cause may be more difficult to identify. Individuals with lymphangiectasia malabsorb fat and fat-soluble vitamins and have protein loss into the intestinal lumen. The most common presentation is hypoproteinemic edema. Nutritional management includes a low-fat diet and supplementation with medium-chain triglycerides, which are absorbed directly into the portal circulation. Fat-soluble vitamins should be given if deficiencies develop.

WATERY DIARRHEA

Watery diarrhea may be due to osmotic, secretory, inflammatory, or often combined mechanisms (see Fig. 140-3).

Ingestion of Nonabsorbable or Poorly Absorbable Solutes

MAGNESIUM AND SODIUM PHOSPHATE AND SULFATE DIARRHEAS

Magnesium, phosphate, and sulfate are poorly absorbed minerals. Individuals who ingest significant amounts of magnesium-based antacids or high-potency multimineral and multivitamin supplements or those who surreptitiously ingest magnesium-containing laxatives or nonabsorbable anion laxatives, such as Na_2PO_4 (neutral phosphate) or Na_2SO_4 (Glauber or Carlsbad salt) may develop osmotically induced, watery diarrhea that may be high volume.

SORBITOL AND FRUCTOSE DIARRHEA

Dietetic food, chewing gum, candies, and medication elixirs that are sweetened with sorbitol, which is an unabsorbable carbohydrate, can cause diarrhea. Excessive consumption of pears, prunes, peaches, and apple juice, which also contain sorbitol and fructose, a poorly absorbable sugar, can result in diarrhea.[22] Most soft drinks are now sweetened with fructose-containing corn syrup and may be a cause of diarrhea when ingested in high concentrations.

Glucose-Galactose Malabsorption and Disaccharidase Deficiencies

Primary and secondary lactase deficiency is the most common cause of disaccharidase deficiency (see discussion of malabsorption). Congenital lactase deficiency causes diarrhea at birth with the first breast-feed. Congenital sucrose-isomaltose deficiency manifests in infancy when table sugar is introduced into the diet. Glucose-galactose malabsorption is due to mutations in the *SGLT1* gene and causes diarrhea at birth. The mechanism of diarrhea in these disorders is osmotic. Stools are acidic owing to conversion of unabsorbed sugars to short-chain fatty acids in the colon. Treatment is the substitution of fructose for other sugars in the diet. Patients who develop gas, bloating, or diarrhea after the ingestion of mushrooms may have a deficiency in the disaccharidase trehalase.

Rapid Intestinal Transit

A small amount of carbohydrate in the diet is unabsorbed by the normal small intestine. Diets that are high in carbohydrate and low in fat may allow rapid gastric emptying and rapid small intestinal motility, thereby leading to carbohydrate malabsorption and osmotic diarrhea. Rapid transit time also occurs in thyrotoxicosis (Chapter 226). Because of the production of H_2 and carbon dioxide gas by colonic bacteria, abdominal gas and cramping may be the predominant symptoms.

Bile Acid Malabsorption

Ileal malabsorption of bile salts results in the stimulation of colonic fluid secretion and watery diarrhea. Three types of bile acid malabsorption induce diarrhea. Type 1 results when severe disease (e.g., Crohn disease), resection, or bypass of the distal ileum allows bile salts to escape absorption (see earlier). Type 2 may be congenital, rarely, owing to a defect in the apical sodium bile acid transporter, or more commonly may be idiopathic. The idiopathic type has been associated with decreased levels of FGF19, an intestinal fibroblast growth factor that normally downregulates bile salt synthesis in the liver and increased levels of 7-α-hydroxy-4-cholesten-3-one (C4) (a marker of bile acid synthesis) in blood. The result is increased bile salt production that overwhelms reabsorption in the ileum. Type 3 is caused by various conditions, including prior cholecystectomy, celiac disease, pancreatic insufficiency, microscopic colitis, bacterial overgrowth, gastric surgery, or vagotomy. Postulated mechanisms include a bile salt storage problem, increased production, decreased recycling, or saturation of absorption.

TREATMENT Rx

Diarrhea due to types 1 and 2 often responds to cholestyramine (2 to 4 g PO two to four times daily) or the more potent and better tolerated bile salt binder, colesevelam (625-mg tablet PO two to six times daily). Fat-soluble vitamin deficiency is a potential risk with chronic use of bile salt binders. Although many patients with type 3 respond to cholestyramine or colesevelam, some do not. In these patients, motility-altering drugs such as opiates (e.g., loperamide, 2 to 4 mg PO two to four times daily) and anticholinergics (e.g., hyoscyamine sulfate, 0.125 to 0.250 mg PO two to four times daily) may be of benefit.

Functional Watery Diarrhea (Irritable Bowel Syndrome)

See Chapter 137.

TRUE SECRETORY DIARRHEAS

Endocrine diseases that can cause secretory diarrheas (see Fig. 140-3) include carcinoid tumors (Chapter 232), gastrinomas (Chapter 195), VIPomas of the pancreas (Chapter 195), and medullary carcinoma of the thyroid (Chapter 246). Diarrhea is also seen in 60 to 80% of patients with systemic mastocytosis (Chapter 255). Diarrhea resulting from gastrinoma is distinct in that it is caused by high volumes of hydrochloric acid secretion that overwhelm the reabsorptive capacity of the colon and by maldigestion of fat owing to pH inactivation of pancreatic lipase and precipitation of bile salts.

Villous Adenomas

Large (4 to 18 cm) villous adenomas (Chapter 193), particularly in the rectum or occasionally the sigmoid colon, may cause secretory diarrhea of 500 to 3000 mL/24 hours characterized by hypokalemia, chloride-rich stool, and metabolic alkalosis. Increased numbers of goblet cells and increased prostaglandin E_2 are responsible for the diarrhea. Chloride wasting in the stool and metabolic alkalosis are also found in congenital chloridorrhea, which is caused by a defect in the intestinal Cl^-/HCO_3^- transporter. The metabolic alkalosis distinguishes these two diarrheas from most other diarrheas that cause metabolic acidosis. A villous adenoma is usually diagnosed by colonoscopy. The prostaglandin antagonist indomethacin (25 to 100 mg/day PO) reduces the diarrhea in some patients; resection is curative.

Diabetes Mellitus–Related Diarrhea

Constipation is more common than diarrhea in patients with diabetes. High-volume, watery diarrhea, often with nocturnal incontinence, occurs in 20% of patients with poorly controlled type 1 diabetes. These patients usually have concomitant neuropathy, nephropathy, and retinopathy. The diarrhea may be due to several causes, including celiac disease, anal incontinence, bacterial overgrowth related to dysmotility, medications (metformin, acarbose), and autonomic neuropathy. If no specific cause is found, clonidine (initial dose 0.1 mg PO twice daily and titrated slowly to a maximal dose of 0.5 to 0.6 mg PO twice daily) may be helpful. Patients with neuropathy frequently have impaired anal sphincter function, and high-dose loperamide (4 mg PO four times daily) may improve the incontinence.

Alcoholic Diarrhea

Diarrhea related to alcohol ingestion (Chapter 33) may be due to rapid intestinal transit, decreased bile and pancreatic secretion, nutritional deficiencies such as folate or vitamin B_{12}, or alcohol-related enteric neuropathy. Diarrhea may be acute with binge drinking, or it may be chronic and watery and persist for days or weeks. The diarrhea slowly resolves with abstinence from alcohol, proper nutrition, and the repletion of vitamin deficiencies.

Factitious Diarrhea

Approximately 30% of patients referred to tertiary centers have chronic diarrhea from laxative abuse. The diarrhea is usually severe and watery, often with nocturnal symptoms. Some patients may have abdominal pain, weight loss, nausea, vomiting, hypokalemic myopathy, and acidosis. Stool volumes range

from 300 to 3000 mL per day depending on the dose of laxative ingested. In the United States, bisacodyl is the most common cause. Other culprits include anthraquinone (senna, cascara, aloe, rhubarb) or osmotic laxatives (neutral phosphate, Epsom salts, and magnesium citrate). Some patients abuse other agents that cause diarrhea, such as the diuretics furosemide and ethacrynic acid.

More than 90% of laxative abusers are women who have underlying eating disorders such as anorexia nervosa or bulimia (Chapter 219) or middle-aged women who have complicated medical histories and who often work in health care. In patients with unexplained diarrhea, laxative screening of stool and urine (see later) should be performed to exclude this syndrome before an extensive medical evaluation is performed for other causes of chronic diarrhea.

Chronic Idiopathic Secretory Diarrhea

In a small subset of patients with secretory diarrhea, no cause is found despite an extensive evaluation. These cases are labeled as chronic idiopathic secretory diarrhea. In most patients, the diarrhea resolves within 6 to 24 months, which suggests a possible postinfectious or Brainerd diarrhea. If no diagnosis is found after thorough testing and a search for surreptitious laxative abuse, a therapeutic trial with bile salt–binding drugs (e.g., cholestyramine, 4 g PO before meals three times daily, or the more potent colesevelam, 625-mg tablet two to six times daily) or opiates (e.g., loperamide, 2 mg PO four times daily, maximal dose 16 mg daily) is warranted.

INFLAMMATORY DIARRHEAS

Diarrhea resulting from inflammation is characterized by watery or bloody stools, fecal leukocytes, and loss of protein in the stool (see Fig. 140-3).

Inflammatory Bowel Disease

See Chapter 141.

Eosinophilic Gastroenteritis

Eosinophilic gastroenteritis is an increasingly recognized condition of unknown etiology characterized by infiltration of eosinophils in the mucosa, muscle, or serosal layers of the gastrointestinal tract.[23] Approximately 50% of patients have atopic histories. Infestation with nematodes (Chapter 357) must be excluded before this diagnosis is made. Diarrhea occurs in 30 to 60% of patients with mucosal disease. Patients with involvement of the muscle layer often present with abdominal pain, nausea, and vomiting indicative of gastric outlet or intestinal obstruction. Peripheral eosinophilia is present in most patients. The disease may involve the entire gastrointestinal tract from esophagus to anus, or it may be isolated to a segment. With diffuse involvement, patients may have steatorrhea, protein-losing enteropathy, and blood loss.

Microscopic (Collagenous and Lymphocytic) Colitis

These two conditions, collectively known as *microscopic colitis,* may or may not be the same disease or variants of the same disease.[24] Lymphocytic colitis is equally prevalent in men and women, whereas collagenous colitis occurs 10 times more often in middle-aged or elderly women. These conditions may be associated with autoimmune disease or with NSAID use. There is an increased prevalence (15%) of microscopic colitis among individuals with celiac disease. These diseases may be categorized as either inflammatory or secretory diarrheas. An epidemiologic relationship to medications such as NSAIDs, H_2-receptor blockers, proton pump inhibitors, selective serotonin reuptake inhibitors, and smoking has been reported, and increased luminal prostaglandin levels may cause the diarrhea. Enteric infections, food hypersensitivity, or intraluminal bile has been proposed as a trigger for prostaglandin release from lymphocytes.

Antidiarrheal agents such as loperamide (2 mg PO four times daily) are the mainstay of therapy, and the disease usually has a benign and self-limiting course.[25] Budesonide (9 mg PO daily) is the most effective therapy.[Rx] In patients who do not tolerate or respond to it, alternatives include bismuth subsalicylate therapy (8 chewable 262-mg tablets PO daily) and 5-aminosalicylates (e.g., mesalamine, 400 to 800 mg PO three times daily). Patients with refractory disease may require corticosteroids (e.g., prednisone, 40 mg/day PO), a trial of azathioprine or anti–TNF-α antibodies, or, as a last resort, fecal stream diversion surgery.

Food Allergy

Food allergies or sensitivities, especially to cow's milk and soy protein, are a well-established cause of enterocolitis in children, with an estimated frequency of 5%. Symptoms of abdominal cramps, diarrhea, and sometimes vomiting occur shortly after ingestion of the allergen (Chapter 253). The role of food allergy in causing diarrhea in adults is less clear owing to the lack of a reliable diagnostic test. Allergy testing correlates poorly with intestinal allergy. The most common food allergens are milk, soy, eggs, seafood, nuts, and wheat. Sequential elimination diets can be diagnostic and therapeutic.

RADIATION ENTERITIS

Patients who receive pelvic radiation for malignancies of the female urogenital tract or the male prostate may develop chronic radiation enterocolitis 6 to 24 months after total doses of radiation greater than 40 to 60 Gy (Chapters 20 and 142), but symptoms can develop as late as 20 years after treatment. Early abnormalities include an increase in inflammatory mediators, an increase in cholinergic stimulation of intestinal tissue, and endothelial cell apoptosis that precedes epithelial cell apoptosis. The last finding suggests that vascular injury is the primary event. Diarrhea may be caused by bile acid malabsorption if the ileum is damaged, by bacterial overgrowth if radiation causes small intestinal strictures or bypass, or by radiation-induced chronic inflammation of the small intestine and colon. Rapid transit also may contribute to malabsorption and diarrhea.

TREATMENT [Rx]

Treatment is often unsatisfactory. Anti-inflammatory drugs (sulfasalazine, corticosteroids) and antibiotics have been tried with little success. Cholestyramine (4 g PO three times daily) and NSAIDs (e.g., naproxen, 250 to 500 mg PO twice daily) may help, as may opiates (loperamide, 2 mg PO four times daily, or loperamide-*N*-oxide, 3 mg PO two times daily).

PROTEIN-LOSING GASTROENTEROPATHY

Severe protein loss through the gastrointestinal tract can be caused by mucosal diseases such as lymphangiectasia, lymphatic obstruction, bacterial or parasitic infection, gastritis (Chapter 139), gastric cancer, collagenous colitis, inflammatory bowel disease (Chapter 141), celiac disease, sarcoidosis (Chapter 95), lymphoma (Chapter 185), tuberculosis (Chapter 324), Ménétrier disease (Chapter 192), eosinophilic gastroenteritis, and food allergies. A variety of extraintestinal diseases, including systemic lupus erythematosus (Chapter 266), heart failure (Chapter 58), and constrictive pericarditis (Chapter 77), also can be causative. Patients with systemic lupus erythematosus (Chapter 266) may present with protein-losing enteropathy as the only manifestation of their disease. Treatment focuses on the underlying disease.

MISCELLANEOUS DISEASES

Although acute mesenteric arterial or venous thrombosis manifests as an acute bloody diarrhea, chronic mesenteric vascular ischemia (Chapter 143) may manifest as watery diarrhea. Gastrointestinal tuberculosis (Chapter 324) and histoplasmosis (Chapter 332) manifest as diarrhea that may be either bloody or watery, as do certain immunologic diseases, such as Behçet syndrome or Churg-Strauss syndrome. All of these diseases may be misdiagnosed as inflammatory bowel disease (Chapter 141). Neutropenic enterocolitis, an ileocolitis that occurs in patients with neutropenia and leukemia, sometimes is caused by *C. difficile* infection.

TREATMENT OF CHRONIC DIARRHEA [Rx]

Antidiarrheal Therapy

Antidiarrheal agents are of two types: those used for mild to moderate diarrheas and those used for severe secretory diarrheas. A major shortcoming of opiates, the most commonly prescribed antidiarrheal agents, is that they have no antisecretory effect. Rather, they act by decreasing intestinal motility, thereby allowing longer contact time with the mucosa for improved fluid absorption. The exception is racecadotril, an enkephalinase inhibitor, that blocks intestinal fluid secretion without affecting motility.

Bulk-forming agents (psyllium, 7 g in 8 oz water PO up to five times daily] and methylcellulose [3 to 6 tablets twice daily with 300 mL of water]) act by binding water and increasing the consistency of stool. Pectin has been shown to have proabsorptive activity. These agents may be useful in patients with fecal incontinence. Bismuth subsalicylates (524 mg PO every hour up to eight

doses daily) have mild antisecretory and antimotility effects and are effective and safe in mild diarrheas.

The opiates may be symptomatically useful in mild to moderate diarrheas. Paregoric, deodorized tincture of opium, codeine, and diphenoxylate with atropine largely have been supplanted by loperamide. Loperamide does not pass the blood-brain barrier and has a high first-pass metabolism in the liver; it has a high therapeutic-to-toxic ratio and is essentially devoid of addiction potential. It is safe in adults, even in total doses of 24 mg/day. The usual dose is 2 to 4 mg two to four times daily. Opiates may be harmful in patients with severe diarrheas because large volumes of fluid may pool in the intestinal lumen (third space), and stool output is no longer a reliable gauge for replacing fluid losses. The antimotility effects are a problem in infectious diarrheas because stasis may enhance bacterial invasion and delay clearance of microorganisms from the bowel. Opiates and anticholinergics also are dangerous in severe inflammatory bowel disease or severe *C. difficile* infection, where they may precipitate megacolon.

Antidiarrhea agents that are used for the treatment of severe secretory and inflammatory diarrheas generally have profiles with more serious side effects. The somatostatin analogue octreotide (initial dose, 100 to 600 µg SC in two to four divided doses daily; maximal dose, 1500 µg daily) lessens diarrhea in the carcinoid syndrome and in neuroendocrine tumors because it inhibits hormone secretion by the tumor. It is also effective in the treatment of dumping syndrome and chemotherapy-related diarrheas. Long-acting subcutaneous octreotide preparations (20 to 30 mg intramuscularly intragluteally every month) are now available for once-a-month dosing. Octreotide can suppress pancreatic enzyme secretion and make diarrhea worse; it also may be of only limited usefulness in short-bowel syndrome and AIDS diarrhea. Agents such as phenothiazine and calcium-channel blockers have mild antisecretory effects, but side effects limit their use. Clonidine (initial dose, 0.1 mg PO twice daily, titrated slowly to a maximal dose of 0.5 to 0.6 mg twice daily) is most useful in opiate withdrawal diarrhea and is sometimes useful in diabetic diarrhea; postural hypotension may limit its use, particularly in patients with diabetes. Alosetron (0.5 mg PO twice daily for 4 weeks, maximal dose 1 mg PO twice daily) may be justified for severe diarrhea-predominant irritable bowel syndrome; associations with ischemic colitis and severe constipation have limited its use. Indomethacin (250 to 500 mg PO twice daily), a cyclooxygenase blocker that inhibits prostaglandin production, is useful in the treatment of diarrheas caused by acute radiation, AIDS, or villous adenomas of the rectum or colon; occasionally, it may be useful in neuroendocrine tumors and food allergy. For eosinophilic gastroenteritis, corticosteroids (prednisone, 20 to 40 mg/day PO for 7 to 10 days) are the mainstay of therapy, but disodium cromoglycate (200 mg PO four times daily) also may be useful; food elimination diets are not usually effective. Treatment of inflammatory bowel disease is described in Chapter 141.

Grade A References

A1. Lowy I, Molrine DC, Leav BA, et al. Treatment with monoclonal antibodies against *Clostridium difficile* toxins. *N Engl J Med*. 2010;362:197-205.

A2. van Nood E, Vrieze A, Nieuwdorp M, et al. Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *N Engl J Med*. 2013;368:407-415.

A3. Richardson V, Hernandez-Pichardo J, Quintanar-Solares M, et al. Effect of rotavirus vaccination on death from childhood diarrhea in Mexico. *N Engl J Med*. 2010;362:299-305.

A4. Sur D, Lopez AL, Kanungo S, et al. Efficacy and safety of a modified killed-whole-cell oral cholera vaccine in India: an interim analysis of a cluster-randomised, double-blind, placebo-controlled trial. *Lancet*. 2009;374:1694-1702.

A5. Hu Y, Ren J, Zhan M, et al. Efficacy of rifaximin in prevention of travelers' diarrhea: a meta-analysis of randomized, double-blind, placebo-controlled trials. *J Travel Med*. 2012;19:352-356.

A6. Alajbegovic S, Sanders JW, Atherly DE, et al. Effectiveness of rifaximin and fluoroquinolones in preventing travelers' diarrhea (TD): a systematic review and meta-analysis. *Syst Rev*. 2012;1:39.

A7. DuPont HL, Jiang ZD, Okhuysen PC, et al. A randomized, double-blind, placebo-controlled trial of rifaximin to prevent travelers' diarrhea. *Ann Intern Med*. 2005;142:805-812.

A8. Allen SJ, Wareham K, Wang D, et al. Lactobacilli and bifidobacteria in the prevention of antibiotic-associated diarrhoea and *Clostridium difficile* diarrhoea in older inpatients (PLACIDE): a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet*. 2013;382:1249-1257.

A9. Lauritano EC, Gabrielli M, Scarpellini E, et al. Antibiotic therapy in small intestinal bacterial overgrowth: rifaximin versus metronidazole. *Eur Rev Med Pharmacol Sci*. 2009;13:111-116.

A10. Kurppa K, Paavola A, Collin P, et al. Benefits of a gluten-free diet for asymptomatic patients with serologic markers of celiac disease. *Gastroenterology*. 2014;147:610-617.

A11. Feurle GE, Junga NS, Marth T. Efficacy of ceftriaxone or meropenem as initial therapies in Whipple's disease. *Gastroenterology*. 2010;138:478-486.

A12. Miehlke S, Madisch A, Kupcinskas L, et al. Budesonide is more effective than mesalamine or placebo in short-term treatment of collagenous colitis. *Gastroenterology*. 2014;146:1222-1230.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Payne DC, Vinje J, Szilagyi PG, et al. Norovirus and medically attended gastroenteritis in U.S. children. *N Engl J Med*. 2013;368:1121-1130.
2. Zboromyrska Y, Hurtado JC, Salvador P, et al. Aetiology of traveller's diarrhoea: evaluation of a multiplex PCR tool to detect different enteropathogens. *Clin Microbiol Infect*. 2014; 20:O753-O759.
3. Zaidi D, Wine E. An update on travelers' diarrhea. *Curr Opin Gastroenterol*. 2015;31:7-13.
4. DuPont HL. Acute infectious diarrhea in immunocompetent adults. *N Engl J Med*. 2014;370: 1532-1540.
5. Binder HJ, Brown I, Ramakrishna BS, et al. Oral rehydration therapy in the second decade of the twenty-first century. *Curr Gastroenterol Rep*. 2014;16:376-383.
6. Iv EC, Iii EC, Johnson DA. Clinical update for the diagnosis and treatment of infection. *World J Gastrointest Pharmacol Ther*. 2014;5:1-26.
7. Schiller LR, Pardi DS, Spiller R, et al. Gastro 2013 APDW/WCOG Shanghai working party report: chronic diarrhea: definition, classification, diagnosis. *J Gastroenterol Hepatol*. 2014;29:6-25.
8. Grace E, Shaw C, Whelan K, et al. Review article: small intestinal bacterial overgrowth: prevalence, clinical features, current and developing diagnostic tests, and treatment. *Aliment Pharmacol Ther*. 2013;38:674-688.
9. Wilcox C, Turner J, Green J. Systematic review: the management of chronic diarrhoea due to bile acid malabsorption. *Aliment Pharmacol Ther*. 2014;39:923-939.
10. Perets TT, Shporn E, Aizic S, et al. A diagnostic approach to patients with suspected lactose malabsorption. *Dig Dis Sci*. 2014;59:1012-1016.
11. Bai JC, Fried M, Corazza GR, et al. World Gastroenterology Organisation global guidelines on celiac disease. *J Clin Gastroenterol*. 2013;47:121-126.
12. Fasano A, Catassi C. Clinical practice: celiac disease. *N Engl J Med*. 2012;367:2419-2426.
13. van der Windt DA, Jellema P, Mulder CJ, et al. Diagnostic testing for celiac disease among patients with abdominal symptoms: a systematic review. *JAMA*. 2010;303:1738-1746.
14. Lundin KE, Sollid LM. Advances in coeliac disease. *Curr Opin Gastroenterol*. 2014;30:154-162.
15. Rubio-Tapia A, Hill ID, Kelly CP, et al. ACG clinical guidelines: diagnosis and management of celiac disease. *Am J Gastroenterol*. 2013;108:656-676.
16. Lebwohl B, Granath F, Ekbom A, et al. Mucosal healing and risk for lymphoproliferative malignancy in celiac disease: a population-based cohort study. *Ann Intern Med*. 2013;159:169-175.
17. Brown IS, Bettington A, Bettington M, et al. Tropical sprue: revisiting an underrecognized disease. *Am J Surg Pathol*. 2014;38:666-672.
18. Puechal X. Whipple's disease. *Ann Rheum Dis*. 2013;72:797-803.
19. Tappenden KA. Pathophysiology of short bowel syndrome: considerations of resected and residual anatomy. *JPEN J Parenter Enteral Nutr*. 2014;38:14s-22s.
20. Jeppesen PB. Pharmacologic options for intestinal rehabilitation in patients with short bowel syndrome. *JPEN J Parenter Enteral Nutr*. 2014;38:45s-52s.
21. DiBaise JK. Short bowel syndrome and small bowel transplantation. *Curr Opin Gastroenterol*. 2014;30:128-133.
22. Fedewa A, Rao SS. Dietary fructose intolerance, fructan intolerance and FODMAPs. *Curr Gastroenterol Rep*. 2014;16:370.
23. Lucendo AJ. Eosinophilic diseases of the gastrointestinal tract. *Scand J Gastroenterol*. 2010;45: 1013-1021.
24. Munch A, Langner C. Microscopic colitis: clinical and pathologic perspectives. *Clin Gastroenterol Hepatol*. 2014;13:228-236.
25. O'Toole A, Coss A, Holleran G, et al. Microscopic colitis: clinical characteristics, treatment and outcomes in an Irish population. *Int J Colorectal Dis*. 2014;29:799-803.

141

INFLAMMATORY BOWEL DISEASE

GARY R. LICHTENSTEIN

DEFINITION

Inflammatory bowel disease refers to two chronic idiopathic inflammatory disorders, ulcerative colitis and Crohn disease. Characteristic clinical, endoscopic, and histologic features are critical for the diagnosis of these disorders, but no single individual finding is absolutely diagnostic for one disease or the other. Ulceration from Crohn disease may be transmural and may occur anywhere in the gastrointestinal tract, most commonly in the distal ileum and proximal colon. The hallmark of ulcerative colitis is continuous ulceration starting in the rectum and limited to the colon. Approximately 10% of patients with inflammatory bowel disease have indeterminate colitis, a term used when Crohn colitis cannot be distinguished from ulcerative colitis.

EPIDEMIOLOGY

Inflammatory bowel disease occurs worldwide, but the highest incidence is found in North America, the United Kingdom, and northern Europe. Data suggest an increasing incidence and prevalence over time and in different regions around the world, although ulcerative colitis remains slightly more prevalent than Crohn disease.[1] The incidence of ulcerative colitis in North America is estimated to be 19.3 per 100,000 person years and 24.3 per 100,000 person years in Europe, with a prevalence of approximately 250 per 100,000 persons in North America and 500 per 100,000 persons in Europe. The incidence of Crohn disease in North America is estimated to be 20.2 per 100,000 person years and 12.7 per 100,000 person years in Europe, with a prevalence of approximately 320 per 100,000 in North America and in Europe.

Crohn disease and ulcerative colitis may occur at any age, but both have their peak incidence in the second to fourth decade, with a second peak in the seventh decade. The female-to-male ratio for both ulcerative colitis and Crohn disease suggests no gender preference.

Crohn disease and ulcerative colitis are polygenic disorders, for which family history is a risk factor. Crohn disease and ulcerative colitis occur in all ethnic and socioeconomic groups, but their incidence is highest in white Caucasians and Jewish people of Eastern European (Ashkenazi) descent. In North America and the United Kingdom, however, the incidence of Crohn disease in African Americans and African Caribbeans appears to be approaching that of whites. Studies of migrants from underdeveloped countries in South Asia to the United Kingdom suggest an increased prevalence of inflammatory bowel disease in subsequent generations, presumably as a result of environmental influences.

Cigarette smoking is associated with a worse prognosis in patients with Crohn disease but an improved course in ulcerative colitis. Nonsteroidal anti-inflammatory drugs (NSAIDs) appear to be associated with new onset of inflammatory bowel disease and with exacerbations of disease. Appendectomy has been suggested as protective against the development of ulcerative colitis. Diet does not clearly affect the course of inflammatory bowel disease.

PATHOBIOLOGY

Although the trigger for inflammatory bowel disease is not known, three major pathways likely activate the disease: a genetic predisposition, immune dysregulation, and an environmental antigen. A possible explanation is that the inability of the innate immune system to clear microbial antigens, combined with increased intestinal epithelial permeability to antigens, eventually leads to an overactive adaptive immune response.

Genetics

Of patients with inflammatory bowel disease, 5 to 20% have another family member with inflammatory bowel disease. First-degree relatives have a 10- to 15-fold increased risk for developing inflammatory bowel disease. The concordance rate of developing Crohn disease in identical twins, siblings, and first-degree relatives is 50%, 0 to 3%, and 5 to 10%, respectively. Ulcerative colitis follows similar genetic patterns but with slightly lower risk rates. Twenty percent of patients with a positive family history of inflammatory

bowel disease will have discordant disease type: one family member with Crohn disease and another with ulcerative colitis.

More than 163 gene susceptibility loci have been linked to inflammatory bowel disease, with at least 30 specific for Crohn disease and more than 20 specific for ulcerative colitis. Some of these genes also may correlate with the severity of disease. The first gene discovered to be associated with Crohn disease was *NOD2/CARD15,* which is located on chromosome 16 (16q12) and expressed in intestinal epithelial Paneth cells, macrophages, and dendritic cells. This gene is involved in the expression of an intracellular receptor that senses muramyl dipeptide, a peptidoglycan component of gram-positive bacteria. Activation of *NOD2* leads to activation of nuclear factor κ-B (NF-κB), which mediates transcription of numerous proinflammatory cytokines. A mutation in the leucine-rich domain of the NOD2 protein, which interacts with bacterial lipopolysaccharide, leads to failure in activation of NF-κB and is associated with the development of Crohn disease.

The *ATG16L1* gene on chromosome 2 and the *IRGM* gene on chromosome 5 also have been associated with increased susceptibility to Crohn disease. Both are members of a family of genes involved in autophagy, an autonomous process that involves the maintenance of cellular homeostasis and organelle turnover, as well as the processing of intracellular pathogens, the subsequent presentation of antigens, and the regulation of cell signaling. Toll-like receptor-4 gene polymorphisms are associated with both Crohn disease and ulcerative colitis. Polymorphisms of the interleukin-23 (IL-23) receptor gene are associated with ulcerative colitis and a varied risk for Crohn disease. Human leukocyte antigen (HLA) class II polymorphisms, especially in HLA-DR molecules, may confer increased risk for ulcerative colitis and possibly Crohn as well. The *OCTN1* gene, located on chromosome 5q31, and the *DLG5* gene, located on chromosome 10, have been found to be associated with Crohn disease. *DLG5,* which encodes a scaffolding protein that is important for maintaining epithelial integrity in various organs, may interact with the *NOD2/CARD15* gene to increase susceptibility to Crohn disease. *OCTN1* encodes for an ion channel and also increases the risk for Crohn disease; mutations in this gene may disrupt ion channels through altered function of cation transporters and cell-to-cell signaling in the intestinal epithelium.

Inflammatory bowel disease also has been associated with Turner syndrome (Chapter 233), glycogen storage type Ib (Chapter 207), and the Hermansky-Pudlak syndrome (triad of albinism, platelet aggregation defect, and accumulation of ceroid-like pigment in tissue; and Chapter 173). Inflammatory bowel disease is associated with various diseases that have known genetic predisposition, including ankylosing spondylitis (Chapter 265), psoriasis (Chapter 438), atopy (Chapter 249), eczema (Chapter 438), celiac sprue (Chapter 140), cystic fibrosis (Chapter 89), primary sclerosing cholangitis (Chapter 155), multiple sclerosis (Chapter 411), autoimmune thyroid disease (Chapter 226), autoimmune hemolytic anemia (Chapter 160), primary biliary cirrhosis (Chapter 155), myasthenia gravis (Chapter 422), and Cogan syndrome (Chapter 270).

PATHOPHYSIOLOGY

Microbes likely play a part in the development of inflammatory bowel disease. In several animal models of colitis, colitis does not develop in a sterile environment but can be induced after the introduction of commensal bacteria. Diverting the fecal stream away from active mucosal inflammation, such as in an ileostomy, also helps alleviate inflammation in Crohn disease. Crohn disease and ulcerative colitis preferentially occur in the terminal ileum and colon, which contain the highest concentration of bacteria, on the order of approximately 10^{12} organisms per gram of luminal contents. Antibiotics, particularly antibiotics with broad-spectrum anaerobic coverage, are helpful in the treatment of Crohn disease. More recently, several genetic polymorphisms associated with sensing the intestinal microbial environment and triggering an immune response have been linked to inflammatory bowel disease.

Both Crohn disease and ulcerative colitis are products of a dysregulated innate immune system that triggers T cells and a humoral response. T_H17 cells, which are activated in Crohn disease and ulcerative colitis, are stimulated by IL-23, which is produced by antigen-presenting cells.

Pathology

Crohn Disease

As a result of a dysregulated immune system, patients with Crohn disease develop aphthous ulcers, which are superficial mucosal ulcers. As the disease progresses, the ulceration becomes deeper, transmural, and discrete; it may form a serpiginous pattern and may occur anywhere from the esophagus to the anus in a noncontinuous pattern. The most common location for ulceration is the ileocecal region. In some patients, chronic disease leads to the formation of fibrotic strictures, and approximately 30% of patients may develop fistulas.

In early Crohn disease, the histopathologic findings are characterized by an acute inflammatory infiltrate in the lamina propria, with cryptitis, and crypt abscesses. Later in the disease process, the crypt architecture becomes distorted, with a lymphocytic infiltrate and a resulting branching and shortening of the crypts. Noncaseating granulomas, which are present in up to 15% of endoscopic biopsy specimens and as many as 70% of surgical specimens, are not unique to Crohn disease but help confirm the diagnosis when other classic features are present.

Surgical specimens also may show transmural intestinal wall inflammation and fat creeping on the serosal surface.

Ulcerative Colitis

In mild ulcerative colitis, the mucosa is granular, hyperemic, and edematous in appearance. As the disease becomes more severe, the mucosa ulcerates, and the ulcers may extend into the lamina propria. Ulcerative colitis starts in the rectum and may extend proximally in a continuous pattern, but it affects only the colon. Pseudopolyps may form owing to epithelial regeneration after recurrent acute attacks. With chronic disease, the colonic mucosa may lose the normal fold pattern, the colon may shorten, and the colon may appear narrowed.

In early ulcerative colitis, the histopathologic findings are characterized by epithelial necrosis, an acute inflammatory infiltrate in the lamina propria, cryptitis, and crypt abscesses. In chronic disease, a predominant lymphocytic infiltrate and distortion of crypt architecture are seen.

CLINICAL MANIFESTATIONS

Symptoms of inflammatory bowel disease are varied and may be a consequence of the location of the disease, the duration of disease, and any anatomic complications of the disease, such as strictures and fistulas in Crohn disease (Table 141-1).

Symptoms

Crohn Disease

The terminal ileum is affected in about 70% of patients with Crohn disease. Primary ileal disease occurs in 30% of patients, whereas ileocolonic disease occurs in 40%. Symptoms may include abdominal pain, typically in the right lower quadrant, diarrhea, hematochezia, and fatigue. With more severe disease, fever and weight loss may be present. Some patients may present with obstructive symptoms, such as abdominal pain, abdominal distention, and nausea.

Only approximately 5% of patients develop Crohn disease in the upper gastrointestinal tract, and esophageal Crohn disease occurs in less than 2% of patients. Subjects with upper gastrointestinal Crohn disease may present with dysphagia, odynophagia, chest pain, or heartburn. Gastroduodenal disease occurs in 0.5 to 4% of patients and commonly occurs along with distal disease. Symptoms may include upper abdominal pain. Isolated jejunal disease is rare; if the jejunum is involved, there is also distal small bowel

TABLE 141-1 CLINICAL CHARACTERISTICS OF CROHN DISEASE AND ULCERATIVE COLITIS

CHARACTERISTICS	CROHN DISEASE	ULCERATIVE COLITIS
Peak age of onset (years of age)	15-30, 2nd peak in the 7th decade	20-40, 2nd smaller peak beyond the 7th decade
Sex distribution (F/M)	1.2/1	1/1
Potential sites of gastrointestinal involvement	Esophagus to anus	colon
Skipped areas of involvement	+	–
Transmural inflammation	+	–
Type of ulceration	Usually discrete	Continuous
Fistula	+	–
Stricture	–	–
Perianal disease (fissure, skin tags)	+	–

involvement. Up to 30% of patients have perianal disease (Chapter 145) that may include the development of fistulas, abscesses, fissures, and skin tags. Symptoms of perianal disease include pain and discharge. Fever may be present if there is an abscess.

Fistulas, which are internal tracts that can occur anywhere in the gastrointestinal tract and connect to various sites, occur in 20 to 40% of Crohn patients. Penetrating Crohn disease also may cause intra-abdominal and perianal abscesses owing to a fistula with a blind end or intestinal perforation. External fistulas, which present with symptoms of fluid discharge from the cutaneous opening, can be enterocutaneous or perianal. Internal fistulas can be enteroenteric, rectovaginal, or enterocolonic. Patients may present with persistent abdominal pain and fever with an abscess in this location.

Ulcerative Colitis

As with Crohn disease, symptoms and signs of ulcerative colitis depend on the extent and severity of disease. At the time of diagnosis, 14 to 37% of patients have pancolitis, 36 to 41% have disease extending beyond the rectum, and 44 to 49% have proctosigmoiditis. Symptoms include hematochezia, diarrhea, tenesmus, production of excessive mucus, urgency to defecate, and abdominal pain. In the setting of proctitis or proctosigmoiditis, patients may have constipation with difficulty defecating. With more extensive and severe colonic involvement, patients also may have weight loss and fever. They also may have nausea and vomiting because of abdominal pain, fatigue because of anemia, and peripheral edema because of hypoalbuminemia.

Physical Examination

Signs on physical examination are representative of the type of disease as well as its location and severity. Oral ulcers may be present in Crohn disease. The location of abdominal tenderness usually reflects the location of intestinal involvement. In Crohn disease, abdominal tenderness is classically in the right lower quadrant and may include fullness or a mass depending on the severity of inflammation. Peritoneal signs may occur when penetrating Crohn disease causes intestinal perforation. Rectal examination may reveal skin tags, hemorrhoids, fissure, and fistulae.

Extraintestinal Manifestations

Arthropathy, the most common extraintestinal manifestation (Table 141-2), affects up to 10 to 20% of subjects.[2] Peripheral arthralgias, arthritis, ankylosing spondylitis (Chapter 265), and sacroiliitis may exacerbate with gastrointestinal symptoms. Dermatologic disorders, such as erythema nodosum (10 to 15%; see Fig. 440-24) and pyoderma gangrenosum (1 to 2%; Chapter 261), develop in up to 15% of patients. Eye disorders, especially uveitis and episcleritis (Chapter 423), may occur in 5 to 15%. Patients with inflammatory bowel disease also have up to a 10% risk for renal calculi, especially calcium oxalate stones (Chapter 126), in the setting of fat malabsorption with Crohn disease in the small bowel. Uric acid stones can occur in the setting of severe volume depletion. Patients with inflammatory bowel disease, especially patients with ulcerative colitis, are at increased risk for primary sclerosing cholangitis—2 to 7.5% of patients develop this disorder, and 70-80% of patients with this disorder have inflammatory bowel disease (Chapter 155).

TABLE 141-2 EXTRAINTESTINAL COMPLICATIONS OF INFLAMMATORY BOWEL DISEASE

COMPLICATIONS	CROHN DISEASE	ULCERATIVE COLITIS
Ocular disorders (uveitis, episcleritis)	+	+
Arthropathy	+	+
Oral ulcers	+	–
Skin disorders (pyoderma gangrenosum, erythema nodosum)	+	+
Nephrolithiasis	+	+
Primary sclerosing cholangitis	+	+
Bone disorders (osteoporosis, osteomalacia)	+	–
Thromboembolic disease	+	+
B_{12} deficiency	+	–

Extraintestinal Complications

Patients with inflammatory bowel disease are susceptible to extraintestinal complications from the disease itself or medications used to treat disease. These complications include osteoporosis, osteomalacia, arthritic complications, thromboembolic events, pulmonary disease, and renal, dermatologic, and neurologic complications. Osteoporosis occurs in approximately 15% of patients, and steroid therapy (Chapter 35) is the major risk factor; avascular necrosis of the hip and septic arthritis are unusual complications of steroids or other immunosuppressive therapies. Cheilitis may be a result of iron deficiency anemia (Chapter 159). Patients with inflammatory bowel disease are at an increased risk for thromboembolic disease, especially in the setting of active intestinal disease, even when compared with other autoimmune diseases such as rheumatoid arthritis and celiac disease. Secondary amyloidosis with renal involvement can be a consequence of chronic inflammation. Asthma is the most common pulmonary disorder observed in association with Crohn disease. Patients also are at risk for multiple sclerosis (Chapter 411) and for peripheral neuropathy (Chapter 420) from vitamin B_{12} deficiency, which may occur as a result of poor absorption owing to active small bowel disease or surgical resection.

DIAGNOSIS

When diarrhea (Chapter 140) is the predominant symptom, the initial evaluation should include a thorough medical history, testing for infectious colitis (Chapter 140), and screening for endocrine-metabolic disorders such as hyperthyroidism (Chapter 226) and hypocalcemia (Chapter 245). Infections with organisms such as *Shigella* (Chapter 309), *Amoeba* (Chapter 352), *Giardia* (Chapter 351), *Escherichia coli* O157:H7 (Chapter 304), and *Campylobacter* (Chapter 303) can be accompanied by bloody diarrhea, abdominal cramps, and an endoscopic mucosal appearance identical to that of ulcerative colitis. Stool studies are needed to diagnose or exclude these infections. If hematochezia and abdominal pain are the predominant symptoms, the differential diagnosis is broad (Table 141-3).

TABLE 141-3 DIFFERENTIAL DIAGNOSIS OF ILEITIS AND COLITIS

INFECTIONS
BACTERIAL
Aeromonas
Campylobacter jejuni
Chlamydia (proctitis)
Clostridium difficile
Mycobacterium tuberculosis
Salmonella
Shigella
Enterohemorrhagic *Escherichia coli*
Yersinia
VIRAL
Cytomegalovirus
Herpes simplex virus (proctitis)
Human immunodeficiency virus
FUNGAL
Histoplasma capsulatum
PARASITIC
Entamoeba histolytica
Helminths
VASCULAR
COLLAGEN VASCULAR DISEASE
Behçet disease
Churg-Strauss syndrome
Henoch-Schönlein purpura
Systemic lupus erythematosus
Polyarteritis nodosa
ISCHEMIA

MEDICATIONS/TOXINS
NONSTEROIDAL ANTI-INFLAMMATORY DRUGS
PANCREATIC ENZYME SUPPLEMENTS—FIBROSING COLOPATHY
PHOSPHOSODA BOWEL PREPARATIONS
RADIATION
INFLAMMATORY
APPENDICITIS
DIVERTICULAR DISEASE
EOSINOPHILIC GASTROENTERITIS
NONGRANULOMATOUS ULCERATIVE JEJUNOILEITIS (CELIAC DISEASE)
NEOPLASIA
CARCINOID
CARCINOMA PRIMARY OR METASTATIC
LYMPHOMA
MYCOSIS FUNGOIDES
MALIGNANT HISTIOCYTOSIS
MISCELLANEOUS
AMYLOIDOSIS
SARCOIDOSIS
ENDOMETRIOSIS
TUBO-OVARIAN ABSCESSES

Modified from Aberra FN, Lichtenstein GR. Crohn disease. In Talley NJ, Kane SV, Wallace MD, eds. *Practical Gastroenterology and Hepatology: Small and Large Intestine*. Wiley-Blackwell, 2010:225-235.

Diagnostic Evaluation

Endoscopic Evaluation

In a patient with symptoms suggestive of inflammatory bowel disease and no evidence for an infection to explain the symptoms, endoscopic evaluation is essential. Colonoscopy is the initial endoscopic test for patients who present with lower gastrointestinal symptoms such as diarrhea and hematochezia, except in the presence of acute severe peritoneal symptoms. Colonoscopy to the terminal ileum is important if there is a potential diagnosis of inflammatory bowel disease. Small bowel imaging (such as small bowel follow-through or computed tomography [CT] enterography) also may be needed to determine whether there is small bowel disease or to determine the distribution of disease. Capsule endoscopy is useful if all other endoscopic and radiologic testing is nondiagnostic, but Crohn disease of the small bowel is still suspected. Findings on capsule endoscopy should be followed by endoscopy to obtain biopsies. Capsule endoscopy should not be performed if Crohn disease is complicated by a known small bowel stricture.

Crohn Disease

Early endoscopic findings in Crohn disease include superficial small mucosal ulcers, also called aphthous ulcers. As the severity of Crohn disease progresses, the ulcerations become deeper and may become round, linear, or serpiginous. Intersecting longitudinal and transverse ulcers cause a cobblestone mucosal appearance, with "stone" areas representing normal mucosa (Fig. 141-1). Areas of ulceration, which are typically interspersed with normal "skip" areas, may occur anywhere from the esophagus to anus but are most common in the ileocecal region. Isolated colonic disease occurs in 25% of patients, and 60% will have rectal involvement, thereby making it at times difficult to differentiate from ulcerative colitis.

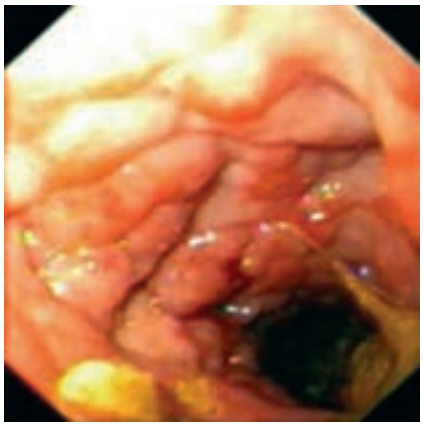

FIGURE 141-1. Endoscopic appearance of Crohn disease with cobblestoning.

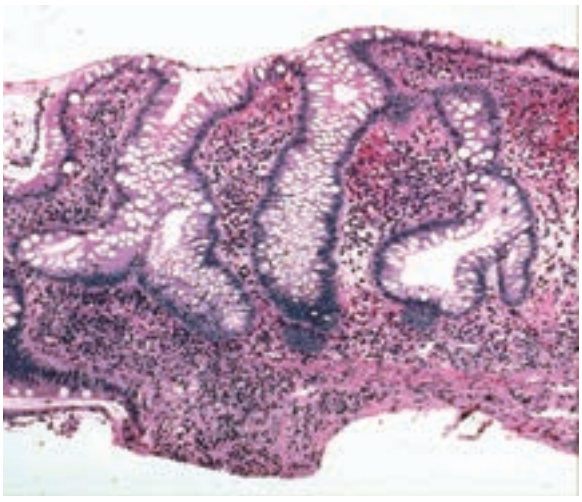

FIGURE 141-2. In ulcerative colitis, histopathology from colonic biopsies reveals features of crypt distortion and lymphocytic infiltration in the mucosa. (Modified from AGA Institute GastroSlides 2010.)

The diagnosis of inflammatory bowel disease is contingent upon accurate histopathologic results, so biopsy of the affected area(s) is key. Findings of an inflammatory infiltrate in the lamina propria and distortion of the crypt architecture support the diagnosis (Fig. 141-2). The diagnosis of Crohn disease may be made by histopathologic examination alone if noncaseating granulomas are seen, but granulomas are rarely found on endoscopic biopsies. The diagnosis of Crohn disease is usually based on a combination of information gleaned from histopathologic findings, colonoscopy, and small bowel imaging. A skip pattern of ulceration, ulceration in the small bowel or upper gastrointestinal tract, or the presence of fistulas support the diagnosis of Crohn disease. Colonic and small bowel ulceration occur in several other disorders, including infections that may not be detected by routine stool studies (such as enterohemorrhagic *E. coli*), vascular disorders, immune-related enterocolitis, neoplasia, diverticulitis, radiation, and medications such as NSAIDs (Table 141-3).

Ulcerative Colitis

The diagnosis of ulcerative colitis is based on endoscopic findings and histopathology. Early in the disease process, patients develop diffuse mucosal erythema with loss of the normal mucosal vascular pattern. In mild disease, the mucosa may have a granular and edematous appearance. As the disease becomes more severe, the mucosa becomes more friable, bleeds easily when the mucosa is touched, and may eventually ulcerate (Fig. 141-3). Endoscopic findings, which start in the rectum and may extend proximally in a continuous pattern, affect only the colon. The term "backwash ileitis" describes a spillover effect from ulcerative colitis and should not be construed as actual involvement of the terminal ileum by ulcerative colitis. Pseudopolyps may form owing to epithelial regeneration after recurrent attacks in patients with long-standing disease. With chronic disease, the colonic mucosa may lose its normal fold pattern, and the colon may shorten and appear narrowed. A new endoscopic index based on the observed vascular pattern, bleeding, and ulceration shows promise for grading the severity of ulcerative colitis and assessing its response to treatment.[3]

Features such as crypt distortion, continuous mucosal inflammation starting from the rectum, absence of granulomas, and absence of small bowel disease are consistent with ulcerative colitis. Early in the disease process, chronic inflammatory findings, such as crypt distortion, may not be present, and the diagnosis may be more difficult to confirm.

Radiology

Radiologic imaging is vital and almost always should be obtained when inflammatory bowel disease, particularly Crohn disease, is suspected. Barium studies such as an upper gastrointestinal series, small bowel follow-through, and barium enema are usually necessary to diagnose fistulas and strictures in Crohn disease. If Crohn disease is suspected by colonoscopic examination, a small bowel follow-through is generally obtained to assess the extent, severity, and type of disease (strictures and fistulas) in the small intestine. CT enterography and magnetic resonance imaging (MRI) enterography are alternatives to a small bowel follow-through. CT enterography may be preferred for the detection of abdominal abscesses, whereas MRI may be preferred for the detection of perineal abscesses and strictures.

Laboratory Findings

Anemia may result from chronic disease, blood loss, or nutritional deficiencies of iron, folate, or vitamin B_{12}. A modestly elevated leukocyte count is

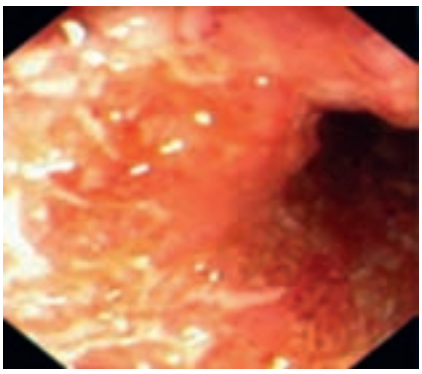

FIGURE 141-3. Endoscopic appearance of ulcerative colitis.

indicative of active disease, but a marked elevation suggests an abscess or another suppurative complication. The erythrocyte sedimentation rate and C-reactive protein are nonspecific serum inflammatory markers that are commonly used to monitor the activity of disease. Hypoalbuminemia is an indication of malnutrition and is common with active disease. Ileal disease or resection of more than 100 cm of distal ileum results in a diminished serum vitamin B_{12} level because of malabsorption.

Serologic Markers

Serologic markers are supportive but may not be used independently to diagnose inflammatory bowel disease. Anti–*Saccharomyces cerevisiae* antibodies (ASCA), which are antibodies to yeast, are present in 40 to 70% of patients with Crohn disease and in less than 15% of patients with ulcerative colitis. The combination of elevated ASCA immunoglobulin A (IgA) and IgG titers is highly specific for Crohn disease, ranging from 89 to 100%. Perinuclear antineutrophil cytoplasmic antibodies (pANCA) are present in 20% of Crohn patients, primarily in colon-predominant disease, and in 55% of patients with ulcerative colitis. ASCA-positive and pANCA-negative disease are associated with a sensitivity of 55% and specificity of 93% for Crohn disease. The antimicrobial antibodies anti-I2 (Crohn disease–related protein from *Pseudomonas fluorescens*), anti-Cbir1 (flagellin-like antigen), and anti-OmpC (*E. coli* outer membrane porin C) are also associated with Crohn disease.

TREATMENT

Rx

The aim of medical therapy is to reduce inflammation and subsequently induce and maintain clinical remission. Medications used to treat inflammatory bowel disease include the categories of 5-aminosalicylate (5-ASA), antibiotics, corticosteroids, immunomodulators, and biologics (infliximab, adalimumab, certolizumab pegol, golimumab, vedolizumab, and natalizumab (Table 141-4).[4] The specific medical therapy selected is based on the location, extent (nonpenetrating and nonstricturing, stricturing, and penetrating and fistulizing disease), and severity of disease (Fig. 141-4; see Fig. 141-3). Supportive medical therapy, such as antidiarrheal and antispasmodic medications, also may be used.

Categories of Medical Therapy

5-Aminosalicylate

5-ASA, which acts as a topical anti-inflammatory within the lumen of the intestine, is used to treat mild to moderate ulcerative colitis and as maintenance therapy for patients in remission. Sulfasalazine is the combination of a sulfapyridine with 5-ASA; 5-ASA is responsible for the anti-inflammatory property of this drug, whereas sulfapyridine is the carrier that allows 5-ASA to be delivered into the colon. Other oral formulations of 5-ASA allow it to be delivered to the intestine by different mechanisms. Mesalamine is released in the intestine based on a pH delivery model, whereas sulfasalazine, olsalazine, and balsalazide are released in the intestine by bacterial cleavage of a covalent bond between 5-ASA and a prodrug. For rectal and sigmoid disease, 5-ASA suppository and enema preparations are also effective for induction and maintenance of remission in patients with ulcerative colitis. Adverse events associated with 5-ASAs are rare and may include nausea, dyspepsia, hair loss, headache, worsening diarrhea, and hypersensitivity reactions.

Corticosteroids

Corticosteroids are primarily used to treat flares of ulcerative colitis and Crohn disease. Oral formulations may be used for mild to moderate disease, whereas systemic corticosteroids are used for moderate to severe disease.

TABLE 141-4 MEDICAL THERAPIES FOR INFLAMMATORY BOWEL DISEASE

DRUG	DOSE	RELEASE SITE
5-AMINOSALICYLATES		
Sulfasalazine (Azulfidine)	2-6 g/day	Colon
Mesalamine (Asacol, Lialda, Apriso)	2.4-4.8 g/day	Distal ileum, colon
Olsalazine (Dipentum)	1-3 g/day	Colon
Balsalazide (Colazal)	6.25 g/day	Colon
Mesalamine (Pentasa)	2-4 g/day	Duodenum, jejunum, ileum, colon
Mesalamine (Rowasa), enema, suppository	4 g/day (enema) 1 g/day (suppository)	Rectum/sigmoid Rectum
Mesalamine (Canasa), suppository	1 g/day (suppository)	Rectum
CORTICOSTEROIDS		
Budesonide (Entocort EC)	Induction: 9 mg PO daily Maintenance: 6 mg PO daily	Small intestine
Budesonide (MMX, UCERIS)	Induction 9 mg PO daily	Colon
Prednisone	0.25-0.75 mg/kg PO daily	Systemic
Methylprednisolone	40-60 mg IV daily	Systemic
IMMUNOMODULATORS		
6-Mercaptopurine	1.5 mg/kg/day	Systemic
Azathioprine	2.5 mg/kg/day	Systemic
Methotrexate	Induction: 25 mg SC weekly × 4 mo. Maintenance: 15-25 mg SC weekly	Systemic
Cyclosporine	2-4 mg/kg/day IV	Systemic
BIOLOGICS		
Infliximab	Induction: 5 mg/kg IV weeks 0, 2, 6 Maintenance: 5-10 mg/kg IV every 8 weeks	Systemic
Adalimumab	Induction: 160 mg SC week 0, 80 mg week 2 Maintenance: 40 mg SC every other week	Systemic
Golimumab	Induction: 200 mg SC week 0 and 100 mg week 2 Maintenance: 100 mg SC every 4 weeks	Systemic
Certolizumab pegol	Induction: 400 mg SC weeks 0, 2, 4 Maintenance: 400 mg SC every 4 weeks	Systemic
Natalizumab	300 mg IV every 4 weeks	Systemic
Vedolizumab	Induction: 300 mg IV at 0, 2, and 6 weeks Maintenance: 300 mg IV every 8 weeks	Systemic

IV = intravenously; PO = orally; SC = subcutaneously.

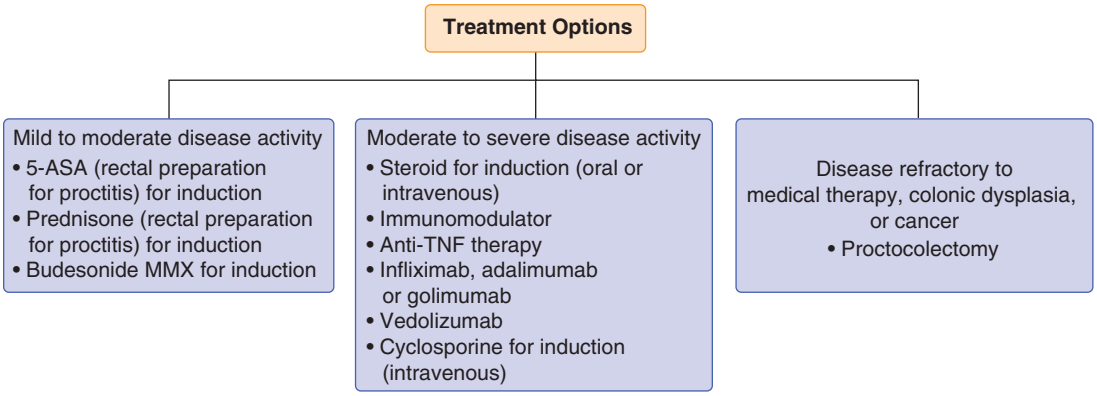

FIGURE 141-4. Ulcerative colitis treatment algorithm.

Enteric-coated budesonide, a pH-dependent ileal release formulation, is an oral corticosteroid with high topical activity and low systemic bioavailability (10%). Enteric-coated budesonide is indicated for treatment of active mild to moderate ileocecal Crohn disease. Budesonide MMX is a budesonide formulation that is released in the colon and is available for treatment of mild to moderately active ulcerative colitis. Oral corticosteroids such as prednisone and methylprednisolone are used for moderate to severe disease, starting at doses ranging from 40 to 60 mg/day. Intravenous methylprednisolone is used for severe disease, with dosing ranging from 40 to 60 mg/day. Maintenance with systemic corticosteroids is not recommended because of their substantial side effects (Chapter 35).

Immunomodulatory therapy

In patients who remain symptomatic despite 5-ASA therapy or who have moderate to severe Crohn disease or ulcerative colitis, the thiopurine analogues (6-mercaptopurine and azathioprine) may be used. Methotrexate also may be used for moderate to severe Crohn disease. Azathioprine, the prodrug of 6-mercaptopurine, typically is prescribed at a dose of 2 to 3 mg/kg/day; the equivalent dose of 6-mercaptopurine is 1.5 mg/kg/day. A disadvantage of the thiopurine analogues is the slow clinical response that may not be evident for as long as 12 weeks. Their side effects include allergic reactions, pancreatitis, myelosuppression, nausea, infections, hepatotoxicity, and malignancy, especially lymphoma.[5] The white blood cell count and liver chemistries must be monitored routinely. Methotrexate, which is a folic acid antagonist, is given as 25 mg intramuscularly (IM) or subcutaneously (SC) once per week for 16 weeks for active Crohn disease and, 15 mg to 25 mg IM or SC once per week for maintaining remission.

Antibiotics

The exact mechanism for the beneficial effect of broad-spectrum antibiotics in the treatment of inflammatory bowel disease is not known. Potential mechanisms include eliminating bacterial overgrowth, eradicating a bacterially mediated antigenic trigger, and potential immunosuppressive properties (e.g., metronidazole). The primary role of antibiotics is in Crohn disease, where metronidazole (10 to 20 mg/kg/day for 4 to 8 weeks), ciprofloxacin (500 mg orally (PO) twice daily for 4 to 8 weeks), or both are primary inductive therapies for perianal fistulae and fissures. Metronidazole also may be a helpful adjunctive treatment for colonic Crohn disease and to prevent postoperative recurrence in Crohn disease as well. In addition, a novel enteric form of rifaximin may be of benefit for mild to moderate Crohn disease.

Biologicals

Anti–Tumor Necrosis Factor-α Agents

Monoclonal antibody therapy directed against tumor necrosis factor-α (anti–TNF-α) include infliximab, which is a chimeric mouse-human IgG1 monoclonal antibody that is approved to treat moderate to severe Crohn disease, fistulizing Crohn disease, and moderate to severe ulcerative colitis that has failed to respond to conventional therapy. Adalimumab (Humira) and certolizumab pegol (Cimzia) have been approved to treat moderate to severe Crohn disease that has failed to respond to conventional therapy, and adalimumab (Humira) and golimumab (Simponi) have been approved to treat moderate to severe ulcerative colitis that has failed to respond to conventional therapy. Adalimumab and golimumab are fully human IgG1 antibodies that are self-administered subcutaneously. Certolizumab pegol, which is a chimeric pegylated Fab fragment to TNF-α, also is administered subcutaneously. Before anti-TNF therapy is considered, risk versus benefit needs to be assessed in each individual patient, given the potential risk for infection and malignancy.[6]

Antiadhesion Molecules

Natalizumab, a humanized IgG4 monoclonal antibody, binds to the α_4 subunit of $\alpha_4\beta_1$ and $\alpha_4\beta_7$ integrins expressed on all leukocytes except neutrophils. Natalizumab inhibits the interactions between α_4 integrins on the surface of leukocytes and adhesion molecules on vascular endothelial cells in the gastrointestinal tract, thereby preventing adhesion and recruitment of leukocytes. Natalizumab is approved for the treatment of moderate to severe Crohn disease that is refractory to other therapies, but there are strict guidelines for prescribing natalizumab because of its associated risk for progressive multifocal leukoencephalopathy (Chapter 370).

Another small adhesion molecule, vedolizumab, is approved for patients with moderate to severe ulcerative colitis and adult patients with moderate to severe Crohn disease when one or more standard therapies (corticosteroids, immunomodulators, or TNF blocker medications) have not provided an adequate response. Because this agent is gut-selective and is not associated with impairment of central nervous system immunosurveillance, the risk for progressive multifocal leukoencephalopathy appears to be very low in this molecule.

Crohn Disease Medical Therapy

Mild to Moderate Crohn Disease

Sulfasalazine (3 to 6 g/day), is superior to placebo for treating active ileocolonic and colonic Crohn disease, with response rates ranging from 45 to 55% for mild to moderate disease, but is not clearly effective for small bowel disease alone (Fig. 141-5).[7] Mesalamine may provide a modest benefit compared with placebo for mild to moderate disease. However, the 5-ASAs are not effective for maintaining remission in Crohn disease. In a phase 2 randomized trial of patients with moderately active Crohn disease (800 mg of extended intestinal release rifaximin twice daily for 12 weeks) induced remission in 63% of patients compared with 43% of controls, with few adverse events. For mild to moderate Crohn disease involving the distal small intestine or proximal colon budesonide (9 mg/day) provides approximately a 70% response rate after 8 weeks and is significantly more effective than mesalamine (4 g/day) for distal ileal and right colonic disease. As a maintenance agent at 3 or 6 mg, the effects of budesonide wane and disappear within 1 year.

Upper gastrointestinal Crohn disease (jejunal, duodenal, gastric, and esophageal) is uncommon, and few clinical trials are available to assess therapies for this location. Because local therapies such as 5-ASAs and budesonide are not released in these locations, systemic immunosuppressants (azathioprine, mercaptopurine, infliximab, adalimumab, and certolizumab pegol) are the mainstays of therapy.

Moderate to Severe Crohn Disease

Patients with moderate to severe disease are initially treated with systemic corticosteroids, but corticosteroids should not be used as maintenance therapy. Options to induce a remission or maintain a steroid-induced remission include 6-mercaptopurine, azathioprine, methotrexate, infliximab, adalimumab, and certolizumab. Infliximab (5 mg/kg at 0, 2, and 6 weeks and then every 8 weeks) alone or infliximab plus azathioprine (2.5 mg/kg/day) is more effective than azathioprine alone, and initial combined therapy (corticosteroids, daily azathioprine, and infliximab) is preferable to reserving it only for patients who do not respond to corticosteroids plus azathioprine or who have aggressive disease.[8] Infliximab therapy also can decrease the need for hospitalization and surgery.

Of patients with Crohn disease who are treated for at least 1 year with infliximab and an antimetabolite agent, approximately 50% will experience a relapse within 1 year after discontinuation of infliximab. Patients who do not respond to conventional therapy, including an anti-TNF agent, may be considered for natalizumab.

For severe Crohn disease, patients should be hospitalized, given nothing by mouth, rehydrated with intravenous fluids, and administered parenteral corticosteroids. Patients who respond to parenteral corticosteroids should be switched to high-dose oral corticosteroids (prednisone, 40 to 60 mg/day), with the dose of prednisone gradually reduced. Patients who have severe Crohn disease and who do not respond to parenteral corticosteroids within a week should be considered for either infliximab or surgery.[9] A course of total parenteral nutrition (Chapter 217) may be useful as adjunctive therapy.

Fistulizing Crohn Disease

Fistulas (Chapter 145) occur in one third of patients with Crohn disease, and perianal fistulas represent the most common location. Asymptomatic internal fistulas rarely require therapy. A concomitant abscess, which may occur in the

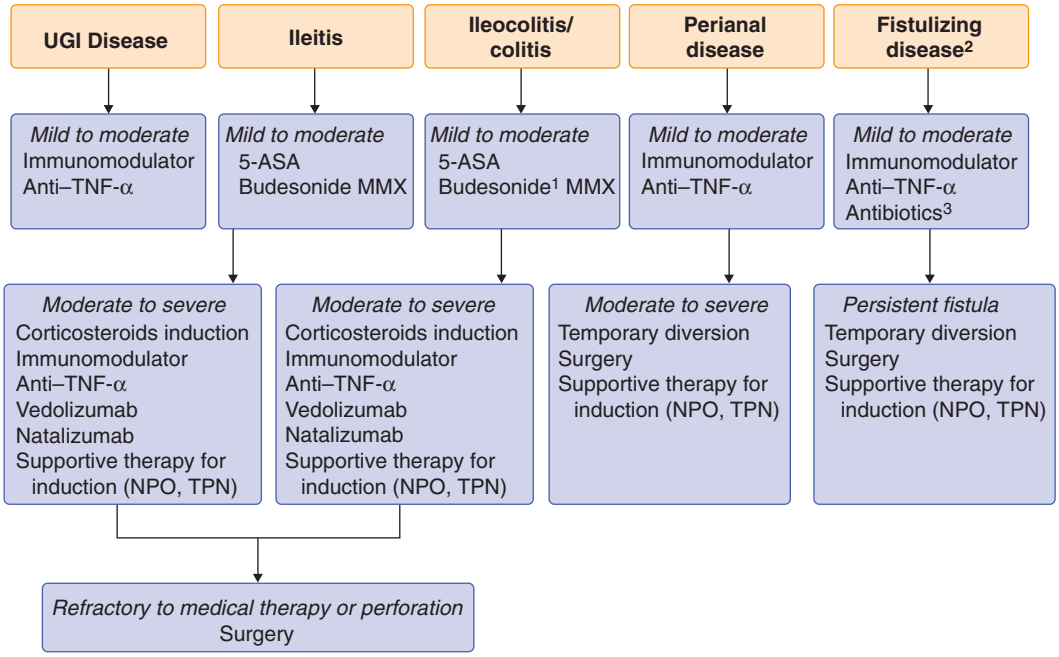

[1]Proximal colon disease involvement.
[2]Abscess should be excluded before initiating medical therapy.
[3]Perianal location.

FIGURE 141-5. Crohn disease treatment algorithm. NPO = nothing by mouth; TNF-α = tumor necrosis factor-α; TPN = total parenteral nutrition; UGI = upper gastrointestinal.

setting of a fistula, must be excluded before initiating immunosuppressive therapy. Surgery may be required. Medical treatment depends on the location and associated complications.

High-output enterocutaneous fistulas in the setting of proximal small bowel involvement can lead to outputs of more than 500 mL/day and can cause severe volume depletion. Initial management requires volume repletion. In the postoperative setting, a fistulous opening is usually in the area of a wound, and it is imperative to protect the healing skin from infection caused by the drainage from either an ostomy bag or a catheter used for a high-output fistula. High-output fistulas will rarely close spontaneously and typically will require surgical closure. Low-output fistulas may be treated initially with azathioprine (or 6-mercaptopurine), methotrexate, or anti–TNF-α therapy (infliximab, adalimumab, or certolizumab).

Perianal fistulas are classified into simple and complex (Chapter 145). A simple fistula is located below the dentate line (i.e., most of the anal sphincter) and has one track. A complex fistula passes through the intersphincteric (high location), transsphincteric, or suprasphincteric region and may have multiple tracks. Simple fistulas respond well to medical therapy, initially with metronidazole (10 to 20 mg/kg/day PO for 4 to 8 weeks) and ciprofloxacin (500 mg PO twice daily for 4 to 8 weeks) for the fistula and treatment of concurrent mucosal disease. Treatment with immunomodulators or anti–TNF-α agents is also beneficial. Patients with fistulas without rectal mucosal Crohn disease may respond well to fistulotomy, whereas patients with mucosal involvement may benefit from seton placement rather than fistulotomy. Complex fistulas usually require a combination of surgical and medical therapy. In the setting of intractable disease, colonic or ileal diversion may allow for rectal and perianal healing; in severe cases, proctocolectomy may be necessary.

For Crohn disease–related rectovaginal fistulas, medical therapy with antimetabolite therapy or anti–TNF-α agents is usually considered before surgery. Surgical therapy such as fistulotomy and mucosal flap may be considered.

Enterovesicular or colovesicular fistulas may be treated with antimetabolite therapy or anti–TNF-α agents, or both, but recurrent urinary tract infection is an indication for surgery. Surgery usually involves resection of involved bowel and closure of the bladder defect.

Asymptomatic internal fistulas such as enteroenteric fistulas, do not require surgical intervention, but treatment with an immunomodulator may be considered. Internal fistulas, such as cologastric and coloduodenal, may cause substantial symptoms because of bypass of part of the intestine. If medical management fails or if an abscess forms, surgery is recommended.

Medical Management of Ulcerative Colitis

The anatomic distribution of ulcerative colitis guides therapy. Options include suppositories, retention enemas, topical foam, oral therapy, and parenteral therapy. Suppositories are effective to treat proctitis in the distal 20 cm of the colon. Topical foam and enemas are effective for distal and left-sided colitis. Oral therapy and parenteral therapy are effective for all locations of disease.

Proctitis

For active ulcerative proctitis, topical 5-ASA (enema and suppository) in combination with oral treatment is superior to oral treatment alone.[10] Topical 5-ASA is superior to topical corticosteroids for treatment of active ulcerative proctitis, rectal 5-ASA therapy produces a faster response when given with oral 5-ASA. Corticosteroid enemas, suppositories, or foam also can be used if 5-ASA fails. 5-ASA or corticosteroid retention enemas can be used for active disease up to the splenic flexure (i.e., the rectum, sigmoid colon, and descending colon). Another approach to proctitis or distal colitis is an oral aminosalicylate, although a response may not be evident for 3 to 4 weeks. Additionally, once-a-day extended-release budesonide (budesonide MMX) is effective for mild to moderately active ulcerative proctosigmoiditis, with fewer steroid-related side effects than conventional corticosteroids.

Extensive Colitis

In patients with ulcerative colitis of mild to moderate activity and extension of disease proximal to the splenic flexure, the initial drug of choice is an oral 5-ASA; efficacy increases with increasing doses. Even with more extensive disease, supplementation of oral 5-ASA with 5-ASA enemas or suppositories may help reduce the symptoms of urgency that result from rectal involvement, and budesonide MMX provides incremental benefit. In patients with more than five or six bowel movements per day, in patients in whom a more rapid response is desired, or in patients who have not responded to 3 to 4 weeks of 5-ASA, the treatment of choice is oral prednisone. Patients with severe diarrhea, systemic symptoms, or significant amounts of blood in their stool should be started on 40 mg/day; most patients respond to oral corticosteroids within a few days. After the symptoms are controlled, prednisone can be tapered gradually by 5 mg every 1 to 2 weeks. Patients who respond to oral prednisone and can be fully withdrawn from it should be maintained on 5-ASA.

If patients with severe ulcerative colitis do not begin to respond to corticosteroids at the equivalent dose of methylprednisolone 60 mg IV within 5 days or do not completely respond within 7 to 10 days, options include colectomy, infliximab, or cyclosporine. In a phase 2 trial, tofacitinib (an oral inhibitor of Janus kinases at doses ranging from 0.5 mg to 15 mg twice daily for 8 weeks) improved symptoms in patients with moderate to severe active ulcerative colitis. For patients whose disease flares whenever the corticosteroids are withdrawn or their corticosteroid dose is lowered, the continuation of high-dose corticosteroid therapy is the most common management error. In patients whose disease flares when their steroid dose is reduced, a trial of an immunomodulator (azathioprine or 6-mercaptopurine), infliximab, adalimumab, golimumab, or vedolizumab should be attempted. If the patient requires a substantial dose (>15 mg/day of prednisone) for more than 6 months, a trial of an immunomodulator, infliximab, adalimumab, golimumab, or vedolizumab should be considered for maintenance of remission, and attempts should be made to reduce the steroid dose.

The most common indication for hospitalization in patients with ulcerative colitis is intractable diarrhea, although blood loss is also common. Patients with severely active ulcerative colitis should be evaluated for toxic megacolon

by abdominal radiography or CT. Antidiarrheal medications and anticholinergic medications are contraindicated in patients with severe ulcerative colitis because of the risk for precipitating toxic megacolon. The mainstays of therapy for severe ulcerative colitis are rehydration with intravenous fluids and intravenous corticosteroids (hydrocortisone, 300 mg/day; prednisolone, 60 to 80 mg/day; or methylprednisolone, 40 to 60 mg/day). Total parenteral nutrition (Chapter 217) may be necessary in patients with malnutrition. Patients with peritoneal signs or signs of systemic infection should be treated with parenteral antibiotics (Chapter 142). Patients who do not improve in 7 to 10 days should be considered for either colectomy, a trial of intravenous cyclosporine, or a trial of infliximab.

Maintenance Therapy

Aminosalicylates reduce recurrent disease in patients with ulcerative colitis, and essentially all patients should receive maintenance therapy with original or newer 5-ASA preparations. Corticosteroids are not effective as maintenance therapy and should not be used in this way. Azathioprine,[■] 6-mercaptopurine, infliximab,[■] adalimumab (160 mg at week 0, 80 mg at week 2, and then 40 mg every other week), and golimumab[■] are effective for maintenance therapy in patients whose ulcerative colitis is not controlled by 5-ASA.

Surgical Therapy

Crohn Disease

Surgical resection does not cure Crohn disease and recurrences are likely after resection, so the approach should be conservative in terms of the amount of bowel resected. Nevertheless, nearly 50% of patients with Crohn disease undergo surgery within 10 years of their diagnosis.[11] Failure of medical management is a common cause for resection in patients with Crohn disease, but complications (e.g., obstruction, fistula, and abscess) are often the indications for resection. For Crohn disease of the small bowel, the most common surgical procedure is segmental resection for obstruction or fistula; the incidence of a recurrence severe enough to require repeat surgery after ileal or ileocolic resection is approximately 25% after 10 years and 35% after 15 years. For patients with extensive colonic disease that includes the rectum, the procedure of choice is total proctocolectomy with a Brooke (end) ileostomy. Total colectomy with ileal pouch anal anastomosis is not appropriate in Crohn colitis because recurrence of Crohn disease in the ileal segment of the new pouch would require a repeat operation and loss of a long segment of ileum.

Ulcerative Colitis

For ulcerative colitis, colectomy is a curative procedure. Approximately 40% of patients with extensive ulcerative colitis eventually undergo colectomy, usually because their disease has not responded adequately to medical therapy. Emergency colectomy may be required in patients with toxic megacolon or a severe fulminant attack without toxic megacolon. The standard operation for ulcerative colitis is proctocolectomy and a Brooke ileostomy. The most popular alternative operation is total proctocolectomy with an ileal pouch anal anastomosis. In this procedure, a pouch is constructed from the terminal 30 cm of ileum and the distal end of the pouch is pulled through the anal canal. Ileoanal anastomosis is sometimes complicated by inflammation in the ileal pouch (termed *pouchitis*), which can be treated with antibiotics (typically, metronidazole, 500 mg three times daily or 20 mg/kg/day, or ciprofloxacin, 500 mg twice daily for 2 weeks). The decision for or against colectomy and among types of surgery is influenced by the patient's age, social circumstances, and duration of disease, and this decision requires expert consultation. When other indications are equivocal, the risk for malignancy (see later) may be an indication for colectomy.

Complications

Crohn Disease

Abscesses

Abscesses, which are common complications in Crohn disease, result from extension of a mucosal fissure or ulcer through the intestinal wall and into extraintestinal tissue. Leakage of intestinal contents through a fissure into the peritoneal cavity results in an abscess. Abscesses occur in 15 to 20% of patients with Crohn disease, especially in the terminal ileum. The typical clinical manifestation of an intra-abdominal abscess is fever, abdominal pain, abdominal tenderness, and leukocytosis. A CT scan is the preferred modality to diagnose intra-abdominal abscess. Broad-spectrum antibiotic therapy, including anaerobic coverage, is indicated. Percutaneous drainage of abscesses in patients with Crohn disease may improve the clinical picture but does not provide adequate therapy because of persistent communication between the abscess cavity and the intestinal lumen. Resection of the involved intestine is usually required for definitive therapy.

Obstruction

Obstruction is a common complication of Crohn disease, particularly in the small intestine, and is a leading indication for surgery. In Crohn disease, small bowel obstruction may be caused by mucosal thickening from acute inflammation, by muscular hyperplasia and scarring as a result of previous inflammation, or by adhesions. Obstruction also may occur because of impaction of a bolus of fibrous food in a stable, long-standing stricture. Cramping abdominal pain and diarrhea, which worsen after meals and resolve with fasting, suggest obstruction. Strictures may be evaluated by CT enterography, MRI enterography, oral contrast studies, barium enema, or colonoscopy, depending on the anatomic location.[12] Corticosteroids (e.g., methylprednisolone, 40 to 60 mg/day IV, or hydrocortisone, 200 to 300 mg/day IV for 5 to 14 days) are useful if acute inflammation is an important component of the obstructive process, but not if the obstruction is caused by fibrosis. A common error in the management of Crohn disease is inappropriate treatment with long courses of corticosteroids in patients who have obstructive symptoms from fixed anatomic lesions. If the obstruction does not resolve with nasogastric suction and corticosteroids, surgery is necessary.

Perianal Disease

Perianal disease is a potentially disabling complication of Crohn disease. Ulcerations in the anal canal may coalesce and result in fistula formation (Chapter 145). The fistulous openings are most commonly found in the perianal skin but can occur in the groin, vulva, or scrotum. Fistulas are accompanied by drainage of serous or purulent material. If the fistula does not drain freely, there is local accumulation of pus (perianal abscess) with redness, pain, and induration. The pain of a perianal abscess is exacerbated by local pressure that may result from defecation, sitting, or walking. The typical physical manifestation of an abscess is redness with tenderness on digital examination; fluctuance also may be present. Adequate evaluation of perianal disease generally requires proctoscopic examination under anesthesia. Cross-sectional CT or MRI can define the presence and extent of perianal abscesses. The goals of therapy for perianal disease are relief of local symptoms and preservation of the sphincter. Limited disease can be approached with sitz baths and metronidazole, but most cases also require adequate external drainage. Azathioprine, infliximab, adalimumab, or certolizumab pegol may be useful in healing perianal disease, but the disease may reactivate when the drug is stopped. Persistent severe perianal Crohn disease can result in destruction of the anal sphincter and subsequent fecal incontinence.

Ulcerative Colitis

One of the most significant complications of ulcerative colitis is toxic megacolon, which is dilation of the colon to a diameter greater than 6 cm associated with worsening of the patient's clinical condition and the development of fever, tachycardia, and leukocytosis. Physical examination may reveal postural hypotension, abdominal tenderness over the distribution of the colon, and absent or hypoactive bowel sounds. Agents that reduce gastrointestinal motility, such as antispasmodics and antidiarrheal agents, are likely to initiate or exacerbate toxic megacolon. Medical therapy is designed to reduce the likelihood of perforation and return the colon to normal motor activity as rapidly as possible. The patient is given nothing by mouth, and nasogastric suction is begun. Intravenous fluids should be administered to replete water and electrolyte abnormalities, broad-spectrum antibiotics (e.g., ampicillin-sulbactam, given as 1 g ampicillin plus 0.5 g sulbactam, 1.5 to 3 g every 6 hours for 5 to 14 days; levofloxacin, 500 mg/day IV, plus metronidazole, 500 mg IV or PO twice daily; cefazolin, 500 mg IV three times daily, plus metronidazole, 500 mg IV or PO twice daily; or trimethoprim-sulfamethoxazole, 8 to 10 mg/kg/day IV or PO in two to four divided doses, plus metronidazole, 500 mg IV or PO twice daily for approximately 7 days or until symptomatic improvement) are given in anticipation of possible peritonitis as a result of perforation, and parenteral corticosteroids are administered at a dose equivalent to more than 40 to 60 mg of prednisone per day. Signs of improvement include a decrease in abdominal girth and the return of bowel sounds. Deterioration is marked by the development of rebound tenderness, increasing abdominal girth, and cardiovascular collapse. If the patient does not begin to show signs of clinical improvement during the first 24 to 48 hours of medical therapy, the risk for perforation increases markedly, and surgical intervention colectomy is indicated.

Follow-Up

Colon Cancer, Dysplasia, and Colonoscopic Surveillance

The risk for colorectal cancer is increased beginning after 8 years of disease and continues to increase in subsequent years.[13] The incidence of colorectal adenocarcinoma is 60% higher in persons with inflammatory bowel disease than in the general population and has been stable over time. Patients with extensive ulcerative colitis have a markedly increased risk for colon cancer, patients with left-sided disease have an intermediate risk, and patients with long-standing ulcerative colitis are at risk for colorectal cancer even if their symptoms have been relatively mild or even quiescent for 10 to 15 years.

Colon cancers are commonly submucosal and may be missed at colonoscopy. Colon cancer in patients with ulcerative colitis is most commonly associated with dysplastic changes in the mucosa, often at multiple sites in the colon. Although recent data have demonstrated that most dysplasia is visible, not all dysplasia can be identified by visual inspection, so microscopic examination of biopsy specimens is required.

Current practice guidelines recommend colonoscopy with random biopsies in patients with long-standing ulcerative colitis beginning 8 years after the

onset of disease and repeated every 1 to 2 years. If the specimens show dysplasia, colectomy is recommended.

The risk for colon cancer in patients with Crohn colitis is similar to the risk in patients with a similar extent of ulcerative colitis. Surveillance colonoscopy is also recommended in patients with Crohn colitis.

Pregnancy

Fertility in women with inflammatory bowel disease usually is normal or only minimally impaired, and the incidence of prematurity, stillbirth, and developmental defects in the offspring of women with inflammatory bowel disease, except fetal complications may be somewhat more likely when the mother's disease is clinically active, regardless of drug therapy. Previous proctocolectomy or the presence of an ileostomy is not an impediment to successful completion of a pregnancy, but women who have had ileal pouch anal anastomosis surgery with a total proctocolectomy have markedly reduced fertility.

If a woman's disease is inactive at the time of conception, it is likely that it will remain inactive during the course of the pregnancy. Ulcerative colitis that is active at the time of conception tends to worsen. In patients with active Crohn disease at the time of conception, the degree of activity remains the same in two thirds of women; of the other one third, some improve clinically and others deteriorate.

Sulfasalazine does not harm the fetus, but pregnant women have an increased requirement for folic acid and sulfasalazine interferes with folate absorption by competitively inhibiting the jejunal enzyme folate conjugase. Therefore, women who are taking sulfasalazine and who are pregnant or considering pregnancy should receive folate supplementation (1 mg twice daily) to ensure that the fetus receives adequate amounts for normal development. The use of corticosteroids by pregnant women with inflammatory bowel disease is associated with an increased rate of premature rupture of the membranes and a higher rate of cleft lip. In general, it appears that the risk to the pregnancy of treatment with sulfasalazine or corticosteroids is less than the risk in allowing disease activity to go untreated.

Most of the data on the teratogenicity of azathioprine and 6-mercaptopurine in pregnancy are derived from the transplant literature and involve higher doses than are commonly used for inflammatory bowel disease. Reported fetal effects in the transplant population include congenital malformations, immunosuppression, prematurity, and growth retardation. The risks of these medications in the inflammatory bowel disease population are not completely known, given that only small a number of such patients have been formally studied.

Although the IgGI Fc antibody components of infliximab, adalimumab, and golimumab cross the placenta, these agents are considered to be safe during pregnancy. Certolizumab's Fc component is IgG4 and does not cross the placenta to the same extent. However, the risks of these various agents are not completely known, because only a relatively small number of patients have been formally studied.

PROGNOSIS

Ulcerative Colitis

Recurrent flares and remissions characterize typical ulcerative colitis. A rapidly progressive initial attack results in serious complications in approximately 10% of patients. Complete recovery after a single attack may occur in another 10% of patients. Some patients may actually have had an acute undetected infection rather than true ulcerative colitis. The probability that a patient with clinically inactive disease will remain in remission the following year is 80 to 90%. By comparison, patients with clinically active disease have a 70% probability of relapse during the following year.

Patients who present with ulcerative proctitis have the best overall prognosis, and only approximately 5% of patients with proctitis will require colectomy over a lifetime. Severe complications are very uncommon, but the disease will spread more proximally in the colon in up to 50% of patients. Ulcerative colitis–related mortality has decreased substantially since the introduction of corticosteroids, and recent studies suggest that long-term survival rates for patients with ulcerative colitis are similar to those of the general population.

Crohn Disease

The manifestations of Crohn disease wax and wane. A patient with clinically active Crohn disease has a 70 to 80% chance of having active disease in the subsequent year, whereas 80% of patients in remission will remain so over the following year. Over the course of a 4-year period, approximately 25% of patients will have persistently active disease after diagnosis, 25% will remain in remission, and 50% will have a fluctuating course with years of remission and years with clinically active disease. Approximately 75 to 80% of all patients with luminal and fistulizing Crohn disease will require surgical intervention for their disease, with approximately 50% of them requiring surgery within 6 months of diagnosis. The rate for a second surgery for luminal Crohn disease ranges from 25 to 38% within 5 years, and 40 to 70% will need reoperation by 15 years.

Patients with Crohn disease have an increased mortality rate, approximately 1.3 to 1.5 times higher than the general population, unrelated to whether they have small intestine or large intestine involvement, or both. This excess mortality, which is most notable in the first few years after diagnosis, is most commonly related to complications of Crohn disease (e.g., colorectal cancer, shock, volume depletion, protein-calorie malnutrition, and anemia). Whether aggressive use of immunomodulators and biologic therapy will alter the natural course of disease is unknown.

Grade A References

A1. Feagan BG, Macdonald JK. Oral 5-aminosalicylic acid for maintenance of remission in ulcerative colitis. *Cochrane Database Syst Rev*. 2012;10:CD000544.

A2. Marshall JK, Thabane M, Steinhart AH, et al. Rectal 5-aminosalicylic acid for maintenance of remission in ulcerative colitis. *Cochrane Database Syst Rev*. 2012;11:CD004118.

A3. Marshall JK, Thabane M, Steinhart AH, et al. Rectal 5-aminosalicylic acid for induction of remission in ulcerative colitis. *Cochrane Database Syst Rev*. 2010;1:CD004115.

A4. Chande N, Tsoulis DJ, MacDonald JK. Azathioprine or 6-mercaptopurine for induction of remission in Crohn's disease. *Cochrane Database Syst Rev*. 2013;4:CD000545.

A5. McDonald JW, Wang Y, Tsoulis DJ, et al. Methotrexate for induction of remission in refractory Crohn's disease. *Cochrane Database Syst Rev*. 2014;8:CD003459.

A6. Khan KJ, Ullman TA, Ford AC, et al. Antibiotic therapy in inflammatory bowel disease: a systematic review and meta-analysis. *Am J Gastroenterol*. 2011;106:661-673.

A7. Ford AC, Bernstein CN, Khan KJ, et al. Glucocorticosteroid therapy in inflammatory bowel disease: systematic review and meta-analysis. *Am J Gastroenterol*. 2011;106:590-599.

A8. Colombel JF, Sandborn WJ, Reinisch W, et al. Infliximab, azathioprine, or combination therapy for Crohn's disease. *N Engl J Med*. 2010;362:1383-1395.

A9. Costa J, Magro F, Caldeira D, et al. Infliximab reduces hospitalizations and surgery interventions in patients with inflammatory bowel disease: a systematic review and meta-analysis. *Inflamm Bowel Dis*. 2013;19:2098-2110.

A10. Travis SP, Danese S, Kupcinskas L, et al. Once-daily budesonide MMX in active, mild-to-moderate ulcerative colitis: results from the randomised CORE II study. *Gut*. 2014;63:433-441.

A11. Danese S, Fiorino G, Peyrin-Biroulet L, et al. Biological agents for moderately to severely active ulcerative colitis: a systematic review and network meta-analysis. *Ann Intern Med*. 2014;160:704-711.

A12. Timmer A, McDonald JW, Tsoulis DJ, et al. Azathioprine and 6-mercaptopurine for maintenance of remission in ulcerative colitis. *Cochrane Database Syst Rev*. 2012;9:CD000478.

A13. Reinisch W, Sandborn WJ, Rutgeerts P, et al. Long-term infliximab maintenance therapy for ulcerative colitis: the ACT-1 and -2 extension studies. *Inflamm Bowel Dis*. 2012;18:201-211.

A14. Sandborn WJ, Feagan BG, Marano C, et al. Subcutaneous golimumab maintains clinical response in patients with moderate-to-severe ulcerative colitis. *Gastroenterology*. 2014;146:96-109.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*. 2012;142:46-54.
2. Zippi M, Corrado C, Pica R, et al. Extraintestinal manifestations in a large series of Italian inflammatory bowel disease patients. *World J Gastroenterol.* 2014;20:17463-17467.
3. Travis SP, Schnell D, Krzeski P, et al. Reliability and initial validation of the ulcerative colitis endoscopic index of severity. *Gastroenterology*. 2013;145:987-995.
4. Bryant RV, Brain O, Travis SP. Conventional drug therapy for inflammatory bowel disease. *Scand J Gastroenterol*. 2015;50:90-112.
5. Khan N, Abbas AM, Lichtenstein GR, et al. Risk of lymphoma in patients with ulcerative colitis treated with thiopurines: a nationwide retrospective cohort study. *Gastroenterology*. 2013;145:1007-1015.
6. Lichtenstein GR, Feagan BG, Cohen RD, et al. Serious infection and mortality in patients with Crohn's disease: more than 5 years of follow-up in the TREAT™ registry. *Am J Gastroenterol*. 2012;107:1409-1422.
7. Kalla R, Ventham NT, Satsangi J, et al. Crohn's disease. *BMJ*. 2014;349:g6670.
8. Lichtenstein GR. Comprehensive review: antitumor necrosis factor agents in inflammatory bowel disease and factors implicated in treatment response. *Therap Adv Gastroenterol.* 2013;6:269-293.
9. Cheifetz AS. Management of active Crohn's disease. *JAMA*. 2013;309:2150-2158.
10. Harris MS, Lichtenstein GR. Review article: delivery and efficacy of topical 5-aminosalicylic acid (mesalazine) therapy in the treatment of ulcerative colitis. *Aliment Pharmacol Ther*. 2011;33:996-1009.
11. Frolkis AD, Dykeman J, Negrón ME, et al. Risk of surgery for inflammatory bowel diseases has decreased over time: a systematic review and meta-analysis of population-based studies. *Gastroenterology*. 2013;145:996-1006.
12. Rieder F, Zimmermann EM, Remzi FH, et al. Crohn's disease complicated by strictures: a systematic review. *Gut*. 2013;62:1072-1084.
13. Nieminen U, Farkkila M. Malignancies in inflammatory bowel disease. *Scand J Gastroenterol*. 2015;50:81-89.

REVIEW QUESTIONS

1. Which one of the following complications is a known consequence of long-standing ulcerative colitis?

A. Fecaluria (as a manifestation of fistula formation to the bladder from adjacent loops of bowel)
B. Perianal abscess formation
C. Colonic stricture formation with subsequent small bowel obstruction
D. Development of dysplasia and carcinoma
E. Development of left-sided hydoureter

Answer: D The transmural inflammatory process of Crohn disease (but not ulcerative colitis) predisposes to the formation of fistulas, and the presence of fistulae signifies that the transmural inflammation has penetrated into adjacent organs, tissue, or skin. Ulcerative colitis is limited to the colon and, unlike Crohn disease, does not cause perianal abscesses. In Crohn disease, inflammation can lead to ileal or colonic strictures, and marked inflammation can cause local reactions that have a mass effect and lead to obstruction of the right ureter, where the ileum and the ureter cross anatomically. Patients with ulcerative colitis do not develop strictures or hydroureter. Dysplasia and carcinoma may develop in patients with longstanding Crohn disease but are much more typical of ulcerative colitis. This risk provides the rationale for endoscopic surveillance of patients with long-standing inflammatory bowel disease.

2. Which one of the following features is most typical for ulcerative colitis?

A. Deep serpiginous ulcers
B. Aphthous ulcers
C. Pseudopolyps
D. Fistulas
E. Colonic stricturing

Answer: C Pseudopolyps are typically found in patients with ulcerative colitis, not in patients with Crohn disease. Typical radiologic and endoscopic features of Crohn disease include deep linear ulcers and superficial or aphthous ulcers overlying lymphoid follicles. The presence of fistulas should signal the presence of Crohn disease. Patients with Crohn disease may have strictures, which are not classically seen in ulcerative colitis. When strictures are observed in patients with ulcerative colitis, coexisting colorectal cancer should be suspected.

3. Which one of the following statements regarding mesalamine (5-ASA) is true?

A. Oral mesalamine is effective for the induction and for the maintenance of remission in patients who have Crohn disease with a colonic disease distribution.
B. Topical mesalamine for ulcerative proctitis (suppositories) and for left-sided ulcerative colitis (enemas) is effective for the induction and maintenance of remission.
C. Mesalamine is first-line therapy for patients with moderate to severe ulcerative colitis.
D. Mesalamine is efficacious as primary therapy for perianal fistulizing Crohn disease when active disease is present in the rectum.
E. A combination of topical and oral mesalamine therapy in patients with left-sided Crohn disease is more effective than topical or oral mesalamine alone.

Answer: B Oral mesalamine is effective for inducing and maintaining remission in mild to moderate ulcerative colitis. Mesalamine suppositories treat the lower 10 to 20 cm of the colon but do not reach the left colon or splenic flexure; they are extremely effective for inducing and maintaining remission in ulcerative proctitis. In contrast, mesalamine liquid enemas will reach the left colon and as high as the splenic flexure; they are extremely effective for inducing and maintaining remission in left-sided ulcerative colitis. Both oral and topical therapies are effective for treating patients with active ulcerative colitis, and the combination of topical and oral mesalamine therapy is more effective than either mesalamine therapy alone in patients with left-sided ulcerative colitis. Mesalamine is no more effective than placebo for inducing or maintaining remission in Crohn disease patients.

142

INFLAMMATORY AND ANATOMIC DISEASES OF THE INTESTINE, PERITONEUM, MESENTERY, AND OMENTUM

JOHN F. KUEMMERLE

CONGENITAL STRUCTURAL ABNORMALITIES

Meckel Diverticulum

A Meckel diverticulum, which is the most common congenital anomaly of the gastrointestinal (GI) tract, is present in 2% to 3% of the population and is more common in men.[1] Meckel diverticulum occurs when the omphalomesenteric or vitelline duct connecting the fetal yolk sac to the primordial gut fails to close during development. Located on the antimesenteric border, the Meckel diverticulum is commonly found within about 100 cm of the ileocecal valve and typically is 1 to 10 cm in size. Heterotopic tissue is found in about 50% of Meckel diverticula, most commonly gastric or pancreatic tissue.[2] The presence of heterotopic tissue correlates with the development of symptomatic complications, with a lifetime risk of about 6%.

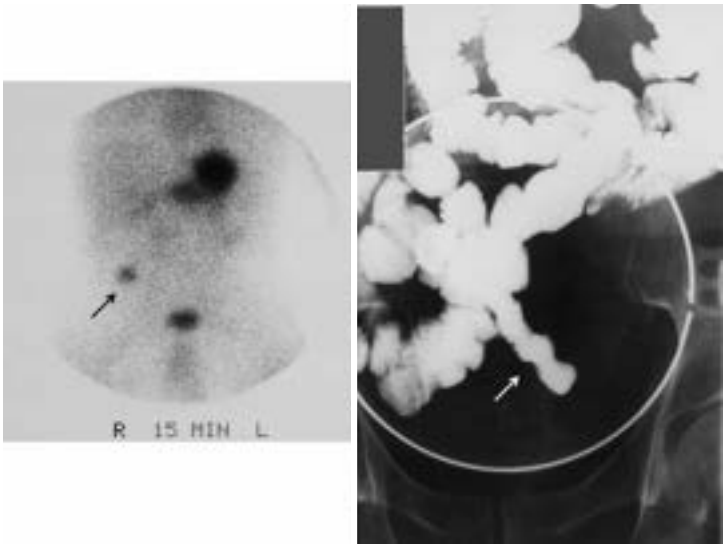

FIGURE 142-1. Meckel diverticulum. Nuclear medicine imaging with ^{99m}Tn pertechnetate scan shows tracer uptake in a Meckel diverticulum (*arrow* in left panel) and in the stomach and bladder. Barium radiography in the same patient also shows the Meckel diverticulum (*arrow* in right panel).

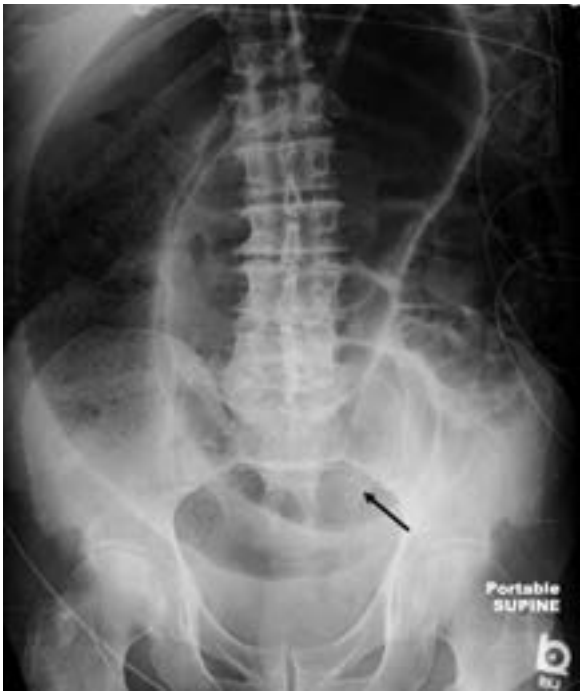

FIGURE 142-2. Sigmoid volvulus. Plain abdominal radiograph shows the presence of a sigmoid volvulus *(arrow)*.

CLINICAL MANIFESTATIONS

The complications from a Meckel diverticulum include bleeding, obstruction, diverticulitis, and perforation. Bleeding can occur when the production of acid from heterotopic gastric mucosa causes ileal ulcerations. Obstruction can result from volvulus around the diverticulum, intussusception of the diverticulum into the intestine, or herniation of the diverticulum and adjacent intestine. Inguinal, femoral, and umbilical hernias all can occur. Repeated and chronic inflammation at the neck of the diverticulum and nearby ileum also can lead to intestinal fibrosis and bowel obstruction. The most common complications are intestinal bleeding in children and obstruction in adults.

DIAGNOSIS

The diagnosis of a Meckel diverticulum can be challenging. Radionuclide imaging with sodium pertechnetate (^{99m}Tn) can be used in cases of bleeding because both normal and heterotopic gastric mucosa take up the tracer (Fig. 142-1). This test has high sensitivity and specificity in children but higher rates of false-positive and false-negative tests in adults. Crohn disease (Chapter 141) and other ileal inflammatory diseases can yield false-positive results. Radiographic imaging with a barium small bowel follow-through typically is not helpful because the diverticulum does not fill with barium contrast. Angiography can visualize the vestigial vitelline artery that arises from the superior mesenteric artery or a superior mesenteric artery branch that directly feeds the diverticulum or adjacent ileum. Both small bowel capsule endoscopy (Chapter 134) and double-balloon enteroscopy can identify a Meckel diverticulum during the evaluation of obscure bleeding.

TREATMENT AND PROGNOSIS

Rx

The management of bleeding, obstruction, or perforation that occurs in association with a Meckel diverticulum is open or laparoscopic surgical resection of the diverticulum and possibly of adjacent ulcerated and bleeding ileum.[3] The surgical resection of a Meckel diverticulum incidentally identified at the time of surgery for another condition is controversial given the low lifetime risk of complications but can be considered in young men, patients with large diverticula, or patients with suspected heterotopic tissue. After identification and treatment, the prognosis is excellent because the Meckel diverticulum is removed, and the risk of complications is eliminated.

Intestinal Atresia and Stenosis, Malrotation, Gastroschisis, and Omphalocele

The congenital disorders of intestinal atresia and stenosis, malrotation, gastroschisis, and omphalocele disorders usually present early in infancy and childhood, but sometimes the diagnosis of stenosis or malrotation may be made in adulthood. With the exception of pyloric stenosis, significant long-term morbidity and mortality are associated with these anomalies even after surgical correction. Patients with malrotation or gastroschisis typically present with complications of their surgery, including adhesions, bowel obstruction, or abdominal wall hernias. Patients with intestinal atresia and resulting short gut syndrome also have significant morbidity and mortality related to intestinal failure.

ACQUIRED STRUCTURAL DISORDERS

Volvulus

Intestinal volvulus, which is pathologic twisting of the intestine around the mesentery, can result in obstruction of the proximal bowel. Mesenteric involvement may lead to vascular compromise, bowel necrosis with resulting perforation, and peritonitis. The most susceptible regions for volvulus are the sigmoid colon, cecum, and occasionally the transverse colon with an estimated annual incidence of two to six cases per 100,000.[4] Elderly persons, especially individuals who are institutionalized, are at the highest risk. Small bowel volvulus is uncommonly observed in U.S. adults but can result from preexisting anomalies such as malrotation or congenital bands of Ladd.

CLINICAL MANIFESTATIONS

Volvulus can present with symptoms and signs of acute bowel obstruction (Chapter 132), including pain that may be out of proportion to physical findings. Nausea and vomiting are usually present. The presentation can also be more insidious or intermittent with constipation, laxative use, and a previously recognized dilated colon. Physical findings include abdominal distention, tympanic percussion, rebound, guarding, and rigidity. Escalating pain and tenderness can indicate colonic ischemia and perforation.

DIAGNOSIS

The diagnosis of colonic volvulus can be made using abdominal radiographs, which demonstrate a distended colon, loss of haustrations, and a typical "bent inner tube" sign with the apex in the right upper quadrant of the abdomen (Fig. 142-2). In cases of a cecal volvulus, the dilated cecum is observed in the epigastrium or in the left upper quadrant. A water-soluble contrast enhanced radiograph can identify the point of obstruction due to volvulus.

TREATMENT AND PROGNOSIS

Rx

Patients with colonic volvulus should have nothing per oral cavity (NPO) with nasogastric (NG) tube decompression and receive appropriate fluid volume resuscitation. In the absence of complete obstruction or signs of ischemia or perforation, patients with a sigmoid volvulus can undergo emergent colonoscopy and attempted reduction, which is successful in up to 75% of cases.[5] Surgical intervention is indicated for volvulus involving the cecum, transverse colon, or small intestine, as well as after colonoscopic reduction of a sigmoid volvulus because of the risk of recurrence. The mortality rate is about 9% for sigmoid volvulus, 7% for cecal volvulus, and about 17% for combined sigmoid and cecal volvulus or transverse colon volvulus.[6]

Intussusception

Intestinal intussusception occurs when a segment of bowel invaginates into the adjacent distal intestine and results in bowel obstruction and ischemia. Intussusception usually involves just the small intestine, but it can also present as small intestinal intussusception into the colon. Although intussusception is a common cause of small bowel obstruction in pediatric patients, especially after rotavirus vaccination (Chapter 380), it is rare in adults and accounts for only about 5% of small bowel obstruction. The cause of intussusception is infrequently identified in children, but a cause can be identified in about 90% of adult cases.[7] Typical precipitating causes in adults include inflammatory bowel disease (Chapter 141), postoperative adhesions, Meckel diverticula, feeding tubes, and small intestinal polyps and tumors (including leiomyomas, neurofibromas, and lymphomas).

CLINICAL MANIFESTATIONS

Most patients present with symptoms of partial bowel obstruction, including pain, nausea, and vomiting, and some patients have diarrhea with occult or overt bleeding. The clinical picture can be confusing when the patient has intermittent symptoms from a spontaneously resolved event. A mass may be palpable on examination. Passage of "currant jelly" stools is characteristic of intussusception, especially in children.

DIAGNOSIS

The diagnosis of intussusception is usually made using computed tomography (CT), which reveals a characteristic alternating high- and low-attenuation target-like or sausage-shaped lesion that represents the invaginated intestinal segments. However, because of the confusing presentation of intussusception in adults, a combination of plain radiographs, upper GI series, and barium enema frequently is required for an adequate evaluation.

TREATMENT AND PROGNOSIS

Colonic intussusceptions are treated surgically in adults because of the high likelihood that a colonic malignancy is the causative lesion.[8] For small intestinal intussusception, pneumatic reduction is successful in children and has been tried in adults in whom no other significant causative lesion is present. However, adults with small intestinal intussusception frequently have an underlying pathological cause, so their treatment consists primarily of surgical intervention and bowel resection, which not only resolves the obstruction but also provides a diagnosis of the causative lesion. If the predisposing cause can be diagnosed and corrected, the prognosis is good, and recurrence rates are low. If the underlying cause is not fully correctable such as with neurofibromatosis (Chapter 417) or adhesions, intussusceptions may recur.

Hernias

Anatomically, hernias comprise a herniated viscus, the hernial sac (internal wall of the hernia lined by peritoneum), and the hernial ring. Whereas an external hernia occurs when the viscus lies outside the abdomen, an internal hernia occurs when the viscus lies in an abnormal location within the abdominal cavity. Secondary hernias can occur at previous sites of incision or injury. Incisional, inguinal, and umbilical hernias comprise 90% of all hernias.[9] Hernias are common and occur in about 5% of the population within their lifetimes. For inguinal hernias alone, the lifetime cumulative incidence is estimated to be 43% in men and 6% in women.[10] By comparison, femoral, umbilical, and incisional hernias occur twice as often in women.

CLINICAL MANIFESTATIONS, DIAGNOSIS, AND TREATMENT OF SPECIFIC HERNIAS

Rx

Epigastric hernias occur at sites of congenital weakness in the midline between the xiphoid and umbilicus along the linea alba. Small epigastric hernias may be asymptomatic or difficult to identify. Larger epigastric hernias can present as nodules, sometimes with tenderness. Multiple hernias may be present. They can be repaired surgically if symptomatic or if complications are present.

Umbilical hernias occur in association with obesity, in multiparous women, and in patients with ascites. They present as a protuberant mass palpable at the umbilicus. Incarceration of small bowel or omentum is common, and strangulation occurs in about one third of umbilical hernias. Umbilical hernias can be repaired surgically if they are symptomatic or associated with complications, but ascites should be controlled for the hernia repair to succeed.

Groin hernias present with bulging in this region, particularly with Valsalva maneuvers. Whereas direct inguinal hernias occur at the site of weakness at the base of the Hesselbach triangle, indirect inguinal hernias occur lateral to the Hesselbach triangle. Both direct and indirect inguinal hernias are above the inguinal ligament. Femoral hernias occur below the inguinal ligament in the femoral canal. Pain is typically mild, but more severe pain or colicky abdominal pain suggests incarceration or strangulation. Palpation can reveal the presence of a groin hernia that increases in size with standing or increased intra-abdominal pressure, such as a Valsalva maneuver, but palpation may be difficult in obese patients. In unclear cases, CT imaging can be helpful. The differential diagnosis of an inguinal bulge also includes adenopathy, lipoma or other tumors, testicular torsion of an undescended testicle, and abscess. Femoral hernias should be repaired when first diagnosed because of their risk for strangulation, but watchful waiting is an acceptable option for men with minimally symptomatic inguinal hernias.[■] The treatment of symptomatic groin hernias is surgical, now usually by open mesh-based techniques or laparoscopic repair, which appear to provide equivalent results.[■] Because the presence of strangulation can reliably be made only at surgery, more severe symptoms warrant early surgical intervention.

Pelvic hernias occur through a weakened pelvic floor and are sixfold more common in women, especially with advancing age. The most common form is an obturator hernia, but less common forms include a sciatic hernia through the sciatic foramen or perineal hernias through the pelvic floor musculature. Most obturator hernias present with acute bowel obstruction. A tender mass may be palpable near the obturator canal on rectal or vaginal examination. Inner thigh pain on internal rotation of the hip may be present in 50% of patients. Diagnosis can be aided using CT imaging. The treatment is surgical.

Incisional hernias can develop after 1% to 4% of laparotomy incisions. Incisional hernias may cause chronic abdominal discomfort, especially with maneuvers that increase intra-abdominal pressure. Repair is usually performed with prosthetic mesh.

More rare hernias include lumbar hernias (which are more common in men and after surgery for trauma), Spigelian hernias occurring through the linea semilunaris in elderly patients, and internal hernias that occur when an intraperitoneal organ protrudes into a separate compartment within the abdomen. Some hernias occur in surgically created defects or because of congenital defects (e.g., paraduodenal, pericecal, or foramen of Winslow). Up to 15% of internal hernias occur through mesenteric or omental defects. Most patients present with intermittent symptoms of pain and bowel obstruction or strangulation. Radiologic studies can be of variable assistance. The differential diagnosis should include volvulus, adhesions, and tumors. Surgery is needed to reduce the herniated viscus and close any defect.

INFLAMMATION OF THE INTESTINE AND COLON

Appendicitis

Appendicitis is the most common intra-abdominal pathology that requires emergency surgery. The lifetime prevalence of appendicitis is 8.7% in men and 6.9% in woman. Lifetime rates of appendectomy are higher, 12% in men and 23% in women, because the diagnosis may be difficult to confirm noninvasively and because of the practice of operating on patients in whom the condition is highly suspected.

About one third of patients have luminal obstruction of the vermiform appendix, most commonly caused by an appendicolith but also occasionally by lymphoid hyperplasia or tumors, including carcinoid tumors (Chapter 232). Gangrenous appendicitis is almost always associated with luminal obstruction.

TABLE 142-1 DIFFERENTIAL DIAGNOSIS OF ACUTE APPENDICITIS

SURGICAL CAUSES
Intestinal obstruction
Intussusception
Acute cholecystitis
Mesenteric adenitis (especially from adenoviral infection; Chapter 365)
Meckel diverticulitis
Right-sided colonic diverticulitis
UROLOGIC CAUSES
Right nephrolithiasis
Right pyelonephritis
GYNECOLOGIC CAUSES
Ectopic or tubal pregnancy
Ruptured or torsed ovarian cyst
Right-sided salpingitis or tubo-ovarian abscess
MEDICAL CAUSES
Yersinia (Chapter 312) or *Campylobacter* (Chapter 303) enterocolitis
Crohn ileitis
Pneumonia
Diabetic ketoacidosis
Herpetic neuralgia (especially right 10th and 11th nerves)
Porphyria
Tuberculous colitis

TABLE 142-2 SCORING SYSTEM FOR ACUTE APPENDICITIS*

	VARIABLE	VALUE
Symptoms	Migration of pain to the right iliac fossa	1
	Anorexia	1
	Nausea or vomiting	1
Signs	Tenderness in right lower quadrant	2
	Rebound of pain	1
	Elevation of temperature (≥37.3°C)	1
Laboratory	Leukocytosis (WBC count >10,000/μL)	2
	Shift to the left (>75% neutrophils)	1
Total score		10

*An aggregate score of 5 or 6 is compatible with the diagnosis of acute appendicitis. A score of 7 or 8 indicates a probable appendicitis, and a score of 9 or 10 indicates a very probable acute appendicitis.
WBC = white blood cell.
Adapted from Alvarado, A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med.* 1986;15:557-564.

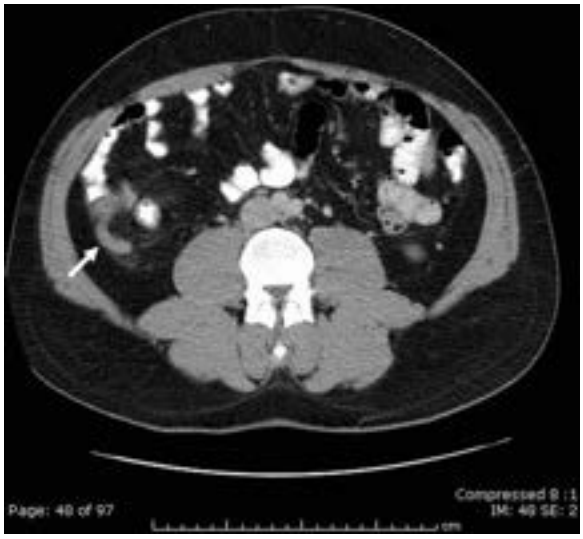

FIGURE 142-3. Appendicitis. A computed tomography scan shows an inflamed appendix with a diameter greater than 1 cm *(arrow)* consistent with acute, uncomplicated appendicitis. (Courtesy of Charlene Prather, MD.)

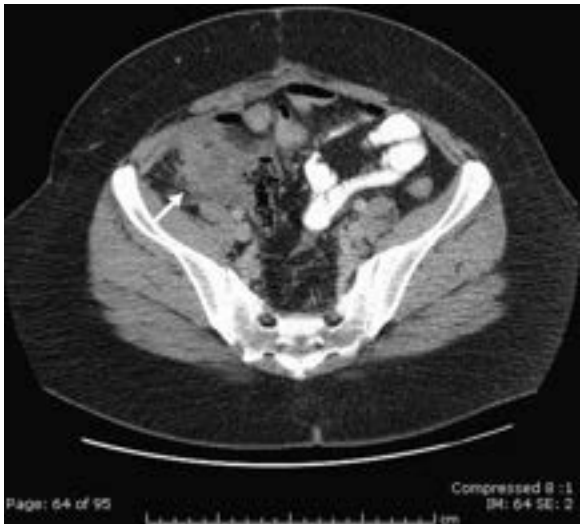

FIGURE 142-4. Appendicitis. A computed tomography scan shows appendicitis complicated by perforation with abscess formation *(arrow)*. (Courtesy of Charlene Prather, MD.)

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The differential diagnosis of appendicitis is extensive (Table 142-1).[11] Current guidelines for diagnosis include characteristic history and physical findings of abdominal pain; localized tenderness; or other signs of acute appendicitis, including increased right lower quadrant pain with cough, pain with flexion and internal rotation of the hip, pain with passive extension of the right hip, and increased right lower quadrant pain during palpation of the left lower quadrant. Laboratory evidence of inflammation include leukocytosis, greater than10,000/μL but usually less than 18,000/μL unless perforation has occurred, with left shift, and elevated markers (Table 142-2) such as the C-reactive protein or procalcitonin level. However, none of these tests are accurate enough to make or exclude the diagnosis of appendicitis.[12] The preferred diagnostic test is multidetector CT (Fig. 142-3), which has a sensitivity and specificity of at least 94% and perhaps higher, and can also detect perforation (Fig. 142-4). Ultrasonography is less sensitive and specific, 83% and 93%, respectively, but is useful when CT is contraindicated, such as in pregnant women or a suspected ectopic pregnancy.

TREATMENT AND PROGNOSIS

When acute appendicitis is suspected, emergent surgical consultation and appendectomy are indicated. Laparoscopic appendectomy is increasingly used in preference to open appendectomy because of lower rates of postoperative complications and a more rapid return to normal eating and activity. Preoperative antibiotics (e.g., cefotetan, 2 g intravenously, or cefoxitin, 2 g intravenously followed by three postoperative doses or ticarcillin–clavulanic acid) reduce infectious complications in otherwise uncomplicated appendicitis. Because perforation of the appendix increases the risk of mortality from 0.0002% to 3% and increases the morbidity rate from 3% to 47%, the traditional approach is that a negative laparotomy is an acceptable trade-off to missing true appendicitis. A nonoperative approach using antibiotics to treat uncomplicated appendicitis can reduce routine surgical morbidity but at the expense of about a 2% risk of rupture and 1% risk of gangrenous appendicitis. For a perforated appendix with abscess formation, immediate appendectomy yields similar results to a strategy of percutaneous ultrasound- or CT-guided drainage, intravenous (IV) antibiotics, and laparoscopic appendectomy about 10 weeks later. In the setting of perforation, once-daily dosing with ceftriaxone and metronidazole for 7 to 10 days is as good as triple-dose therapy.

Complications develop in more than 15% of patients, with an overall mortality rate of about 3% in patients with perforated appendicitis. Complications are uncommon in surgically treated nonperforated appendicitis. Patients who have appendectomies for suspected but not confirmed appendicitis have a prognosis that depends on whether they had an underlying disease, such as Crohn (Chapter 141) or carcinoid (Chapter 232).

Diverticulitis of the Colon

Colonic diverticula are technically pseudodiverticula. They form when the colonic mucosa and submucosa herniate through the muscularis propria of

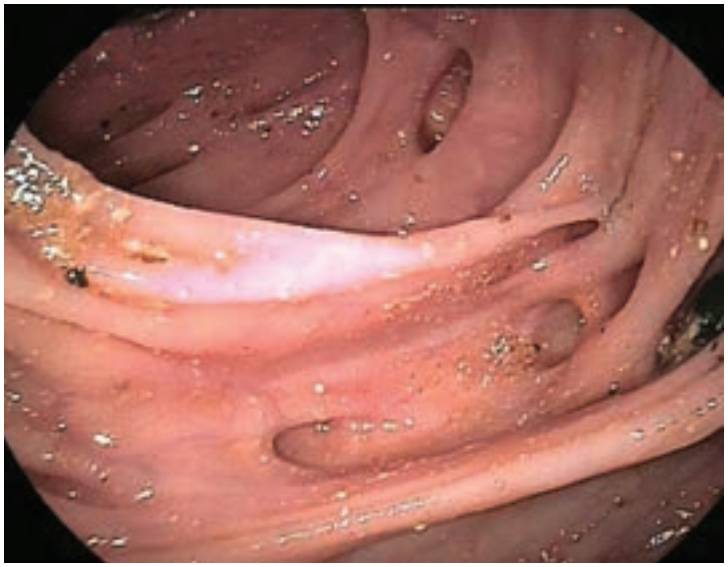

FIGURE 142-5. Sigmoid diverticulosis. The colonoscopic appearance of sigmoid diverticulosis coli.

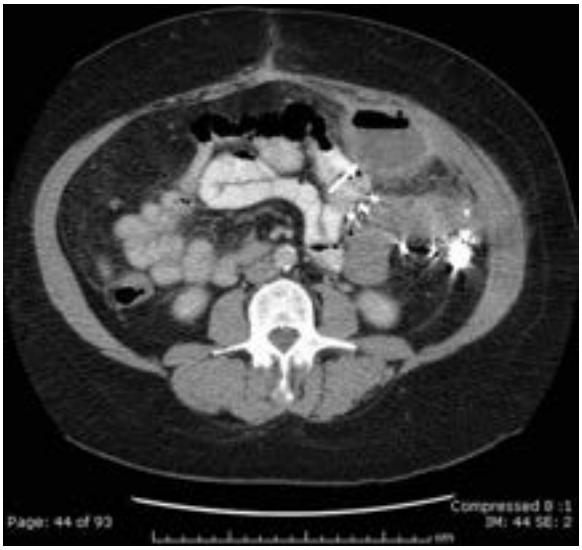

FIGURE 142-6. Diverticulitis. A computed tomography scan shows acute diverticulitis with perforation. An abscess *(arrow)* is seen presenting as an air-filled collection. (Courtesy of Charlene Prather, MD.)

the colon. A spectrum of problems can result from diverticulosis, including diverticulitis, which is an infected diverticulum, or diverticular bleeding, which is manifested as acute lower GI bleeding (Chapter 135). Diverticular disease (Fig. 142-5 and E-Fig. 142-E1) affects about 10% of middle-aged adults and increases in prevalence up to about 80% in elderly adults.

Colonic diverticula form at the site where the nutrient artery, the vasa recta, penetrates the muscularis propria. Diverticulosis in Western populations is most common in the left colon and has been thought to be associated with the low-fiber content of the typical Western diet, although recent epidemiologic studies cast doubt on this hypothesis. Diverticulosis can occur anywhere in the colon, however, and it is more commonly observed in the right colon in Asian populations.

The lifetime risk of diverticulitis is up to 25%.[13] Diverticulitis is thought to occur when impacted material in the diverticulum compresses the blood supply, thereby resulting in a microperforation. Diverticulitis can be complicated further by free perforation, abscess, or fistula formation.

CLINICAL MANIFESTATIONS

The majority of patients with colonic diverticulosis are asymptomatic. Patients with diverticulitis commonly present with localized pain, fever, and anorexia. The pain may radiate to the back, flank, or suprapubic region. Nausea and vomiting, constipation or diarrhea, or urinary symptoms may be variably present. Physical examination typically reveals left lower quadrant tenderness, sometimes with localized guarding or a palpable mass. Rebound tenderness or peritoneal signs should suggest the presence of free perforation. Visible diverticular bleeding is rare in the setting of acute diverticulitis. Leukocytosis is present. When the acutely inflamed diverticulum is adjacent to the bladder, sterile pyuria may be found.

DIAGNOSIS

The diagnosis of acute diverticulitis can be confirmed in the appropriate setting by leukocytosis and an ultrasound examination or a CT scan showing diverticulosis coli with localized inflammation of the colonic wall and pericolic fat at the site of acute diverticulitis. A CT scan also can demonstrate free perforation, abscess, or fistula formation (Fig. 142-6). Because of the increased risk of perforation, invasive testing such as barium enema or colonoscopy is contraindicated when a diagnosis of acute diverticulitis is being entertained. The differential diagnosis of acute diverticulitis includes inflammatory bowel disease (Chapter 141), gastroenteritis, appendicitis, and colon cancer (Chapter 193) with perforation.

TREATMENT

Rx

Uncomplicated acute diverticulitis can be treated with antibiotics. A 7- to 10-day course of oral antibiotics (e.g., ciprofloxacin 750 mg twice daily and metronidazole 500 mg four times daily), perhaps after a single IV dose of antibiotics, is as effective and safe as hospitalization for IV antibiotics. Patients who can tolerate clear liquids can be treated in the outpatient setting with gradual advancement of their diet. Patients who are unable to tolerate eating should be admitted to the hospital for IV fluids and antibiotics (e.g., levofloxacin 750 mg daily and metronidazole 500 mg every 6 hours or piperacillin–tazobactam 3.375 g every 6 hours).

Although the risks associated with one episode of uncomplicated diverticulitis are low, with a mortality rate that is less than 1%, complicated diverticulitis, defined as diverticulitis with abscess, fistula formation, free perforation, or obstruction, is associated with increased inpatient morbidity in up to 25% of cases and has a mortality rate as high as 5%. Furthermore, these risks increase with a second episode of complicated diverticulitis. Elective segmental colectomy typically has been recommended to patients after 2 to 3 episodes of complicated diverticulitis and in young patients even after a first episode. However, recent evidence suggests that the risk of complicated diverticulitis after recovery from uncomplicated diverticulitis is only about 5% and is lower not higher after subsequent episodes of uncomplicated diverticulitis. As a result, prophylactic surgery is not indicated in patients whose diverticulitis is uncomplicated and can be medically treated.[14] Patients with diverticulitis with abscess formation require percutaneous CT-guided drainage and subsequent surgery, usually laparoscopic, typically after 6 weeks. Acute diverticulitis can be complicated by colitis or late stricture formation. About 40% of patients with complicated diverticulitis have significant morbidity, and their mortality rate is about 6%, but it is only about 2% in patients without perforation. There is no convincing evidence that a high-fiber diet prevents recurrent diverticular disease[15] or that nuts or any particular foods should be favored or avoided.

Other Intestinal Inflammatory Conditions

SMALL INTESTINAL ULCERS

Most primary idiopathic ulcers are found in the mid- to distal ileum, where they can be solitary or multiple. A careful history is necessary to exclude other precipitating causes of small intestinal ulcers, including drug exposure and other systemic diseases (Table 142-3). Pathologically, these ulcers can be differentiated from Crohn disease or chronic ulcerative jejunoileitis by the absence of granulomas. Barium contrast studies including enteroclysis can make the diagnosis, and CT or magnetic resonance enterography can also be helpful. In the absence of bowel obstruction, wireless capsule endoscopy is often the preferred test. Therapy with anti-inflammatory or immunosuppressive medications has not proven helpful. Therapy is typically directed to complications, including perforation and obstruction, with segmental surgical resection. However, the risk of ulcer recurrence is high.

Drug-induced ulcerations are common and can result from nonsteroidal anti-inflammatory drugs (NSAIDs), potassium chloride preparations, vasoactive medications, antimetabolites, and cocaine. Wireless capsule endoscopy can make the diagnosis (Fig. 142-7). NSAID-induced injury is similar to Crohn disease, with transmural injury and the risk of stricture formation. Treatment is aimed at avoiding the offending agent, if possible. Unlike for

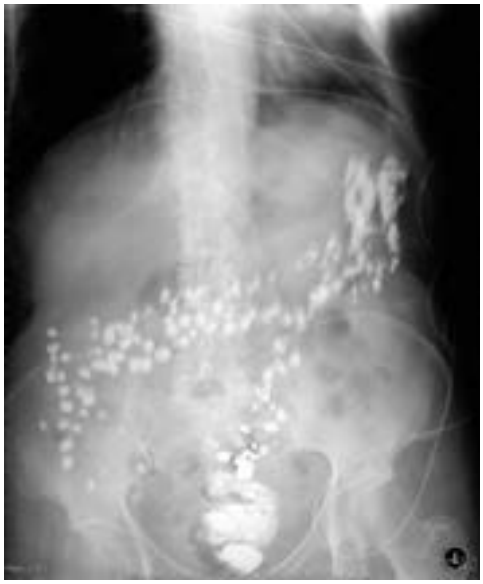

E-FIGURE 142-1. Colonic diverticula. An abdominal flat plate shows residual barium in diverticula scattered throughout the colon. (Courtesy of Charlene Prather, MD.)

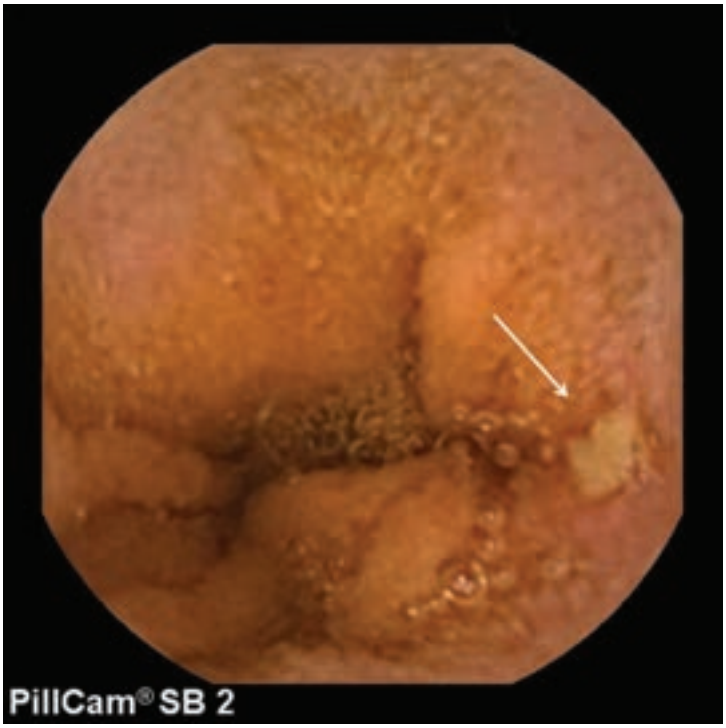

FIGURE 142-7. Nonsteroidal anti-inflammatory drug (NSAID)—induced enteropathy. The wireless capsule endoscopy appearance of a jejunal ulceration *(arrow)* caused by NSAID use.

TABLE 142-3 CAUSES OF SMALL INTESTINAL ULCERS

CATEGORY	CAUSES
Acidic	Meckel diverticulum, Zollinger-Ellison syndrome
Drug-induced	Potassium chloride, NSAID, antimetabolite
Idiopathic	Primary ulcer, Behcet disease
Infectious	Tuberculosis, typhoid, *Yersinia* infection, *Strongyloides* superinfection
Inflammatory	Crohn disease, SLE, chronic jejunoileitis
Metabolic	Uremia
Neoplastic	Malignant histiocytosis, lymphoma, adenocarcinoma
Radiation	Radiation enteritis
Vascular	Mesenteric vascular insufficiency, vasculitis, arteritis

NSAID = nonsteroidal anti-inflammatory drug; SLE = systemic lupus erythematosus.

gastric ulceration, data are conflicting as to whether the concomitant use of a proton pump inhibitor mitigates or exacerbates the condition.[16]

A variety of systemic conditions also can manifest with small intestinal ulcerations. Patients with Crohn disease (Chapter 141) can cause ulcers of any portion of the GI tract. Systemic lupus erythematosus (Chapter 266), rheumatoid arthritis (Chapter 264), scleroderma (Chapter 267), polyarteritis nodosa (Chapter 270), and Henoch-Schönlein purpura (Chapter 270) can present with small intestinal ulcerations that are thought to be secondary to microthrombosis and vasculitis. Mesenteric vasculitis presents as nausea, vomiting, fever, and GI bleeding.

Behçet disease (Chapter 270) is a systemic process that causes intestinal ulcers, typically in the ileocecal region, in fewer than 1% of cases. Although symptoms are similar to those of Crohn disease, pathologically Behçet disease-related ulcers are deep and do not have surrounding inflammation or granulomas. The optimal treatment of this disease has yet to be delineated. Patients frequently are treated with immunomodulators (Chapter 270), and surgery is reserved for ulcer-related complications.

Sarcoidosis (Chapter 95) uncommonly involves the intestine. Its presentation is similar to Crohn disease, with small and large intestinal ulceration and noncaseating granulomas. Patients commonly have nausea, vomiting, diarrhea, abdominal pain, and protein-losing enteropathy. Treatment is targeted to controlling the systemic disease.

Mycobacterium tuberculosis (Chapter 324) typically involves the distal ileum and cecum. A waxing and waning course can mimic the symptoms and location commonly seen in Crohn disease. The diagnosis is suspected based on a history of exposure, particularly in endemic regions, and confirmed by colonoscopy with biopsy, stains, and culture. Treatment is as for disseminated tuberculosis (Chapter 324).

Histoplasma capsulatum (Chapter 332) presents with ulcerations and polypoid masses that mimic tumors, typically in an immunocompromised patient. Patients present with diarrhea, bleeding, and obstruction. The granulomas seen on biopsy must be differentiated from Crohn disease, tuberculosis, and sarcoidosis. Treatment involves managing the systemic infection.

Neutropenic enterocolitis, or typhlitis (E-Fig. 142-E2), is an inflammation of the intestine or colon, usually during the neutropenic phase 10 to 14 days after high-dose induction chemotherapy.[17] It involves right lower quadrant abdominal pain, distention, and diarrhea, sometimes with bleeding. These findings in of themselves are nonspecific and are similar to *Clostridium difficile*–associated colitis (Chapter 296), ischemic colitis (Chapter 143), or pseudo-obstruction (Chapter 136). Diagnosis is based on typical CT scan findings of thickened bowel wall; bowel distention, especially of the cecum; and associated inflammatory changes. Treatment is conservative, with bowel rest and decompression when bowel dilation is present, and broad-spectrum antibiotics. Leukocyte-stimulating agents are often used to reverse the neutropenia (Chapter 167). Treatment of bleeding is supportive with transfusions and correction of any coagulopathy. Surgery is indicated in the setting of intractable bleeding or perforation. The symptoms resolve rapidly with resolution of neutropenia.

VISCERAL ANGIOEDEMA

Visceral angioedema can be idiopathic, or it can be a complication of hereditary and acquired C1 esterase inhibitor deficiency (Chapter 252), hypocomplementemia, drugs (especially angiotensin-converting enzyme inhibitors), or foods. Patients with GI angioedema commonly present with abdominal pain and distention, nausea, vomiting, and diarrhea. Some patients also have evidence of mucous membrane swelling, hives, wheezing, or dyspnea. On CT scan, thickened, fluid-filled loops of small bowel and ascites may also be seen. Symptoms commonly occur episodically, persist for 1 to 3 days and may recur periodically. The diagnosis is established by discontinuing the implicated drug or food, low serum levels of C4, low C1 esterase quantitative levels, or reduced C1 esterase functional activity. C1 esterase inhibitor deficiency can be treated successfully with C1 inhibitor therapy (Chapter 252).

RADIATION ENTERITIS

Radiation injury (Chapters 20 and 140) can occur in the small or large intestine as a result of therapy for gynecologic, urologic, rectal, or retroperitoneal tumors. Acute injury, which can occur during the course of treatment, is associated with nausea, diarrhea, and abdominal or rectal discomfort. The risk is related to radiation dose. Chronic injury typically presents 6 months to 2 years after treatment and presents with progressive bowel obstruction owing to continued inflammation and fibrosis.[18] Patients with radiation proctitis present with rectal pain, bleeding, and occasionally diarrhea.

Acute radiation injury is usually self-limited to the period of treatment, but up to 5% of patients can progress to chronic radiation injury to the small intestine, and up to 15% of patients experience chronic radiation proctitis. Radiation enteritis and stricture formation can be seen on barium radiographs or CT scans. The characteristic neovascularization pattern of telangiectasias from radiation proctitis is seen on endoscopy.

One third to half of patients develop bleeding, which can be minor and intermittent or more substantial. Whereas patients who do not require transfusion have a 70% chance of remission, patients with chronic radiation enteritis who need transfusion have a low rate of remission (20%) and have subsequent high morbidity and even mortality rates. Less invasive treatment options include sucralfate enemas. Several small studies suggest a benefit from endoscopic argon plasma coagulation.

An algorithmic approach emphasizing specific approaches, as outlined earlier, to each specific symptom, can improve outcome compared with routine clinical care. Surgery, which can be required in up to one third of patients with strictures or bleeding, is associated with a high complication rate and should be avoided if possible.[19]

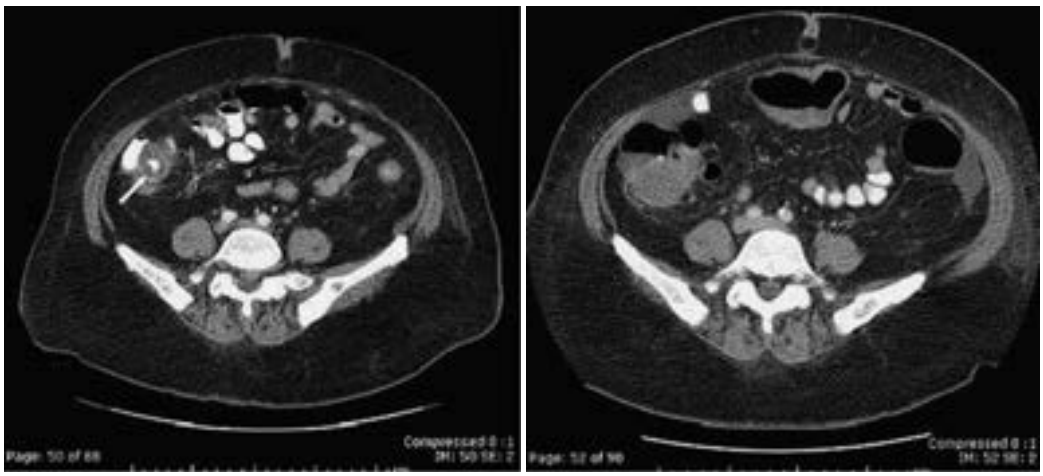

E-FIGURE 142-2. Computed tomography scans showing neutropenic colitis. The *left* scan shows a thickened cecal wall *(arrow)*. Eight days later, a follow-up scan shows resolution (*right* scan). (Courtesy of Charlene Prather, MD.)

PERITONEAL DISORDERS

Peritonitis

Peritonitis, which is a local or generalized inflammation that involves the visceral and parietal peritoneum, can occur as a primary or secondary process. Primary peritonitis in adults is the spontaneous infection of ascites in a patient with cirrhosis (Chapter 153) in the absence of an overt intra-abdominal source. The use of acid-suppressive therapy appears to increase the risk of developing spontaneous bacterial peritonitis threefold in hospitalized patients with cirrhosis.

Secondary peritonitis develops when disease or injury to the intestine results in bacterial contamination from a perforated viscus (e.g., peptic ulcer disease, appendicitis, diverticulitis, penetrating trauma, or iatrogenic), from iatrogenic causes (e.g., peritoneal dialysis [Chapter 131] or surgical contamination), from granulomatous disease (e.g., tuberculosis or fungal infections), or from chemical or aseptic exposures (e.g., bile, urine, or radiographic barium).[20]

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Abdominal pain is the hallmark of peritonitis. It can be sudden in onset in the setting of a perforated viscus or more insidious in nature in the setting of granulomatous or chemical causes. Patients with peritonitis typically lie supine with flexed knees and exhibit shallow breathing. Physical examination reveals a distended abdomen with tenderness to palpation, localized or generalized guarding, and rigidity. Associated symptoms include fever, nausea, vomiting, and leukocytosis with a left shift. Some patients with bacterial peritonitis rapidly develop septic shock (Chapter 108).

Granulomatous peritonitis has a more insidious presentation, and 70% of patients have symptoms for 4 months before diagnosis. Systemic symptoms include fever, malaise, anorexia, and weight loss. The abdomen is diffusely tender. Leukocytosis can be absent.

Patients with tuberculous peritonitis often have a positive skin test result or infiltrates on their chest radiograph. The peritoneal fluid typically has a high protein level (<3 g/dL), a low glucose (<30 mg/dL), and elevated leukocytes (<250 cells/μL); fluid stains and cultures are unhelpful, but polymerase chain reaction–based tests can be diagnostic. The laparoscopic appearance of tuberculous peritonitis is characteristic, with fibrous masses from the parietal peritoneum and granulomas.

Plain abdominal radiographs may show evidence of paralytic ileus, and free air under the diaphragm on upright views confirms the presence of a perforated viscus. CT scan is more sensitive, 70% to 100%, than plain radiographs for detecting free air and may also demonstrate the underlying cause (Fig. 142-8).

In young patients, appendicitis and a perforated duodenal ulcer (Chapter 139) are common causes. In older patients, perforated diverticula and cancer (Chapter 193) are more common. In young women, tubal pregnancy and a ruptured tubo-ovarian abscess must be considered (Chapters 285 and 299).

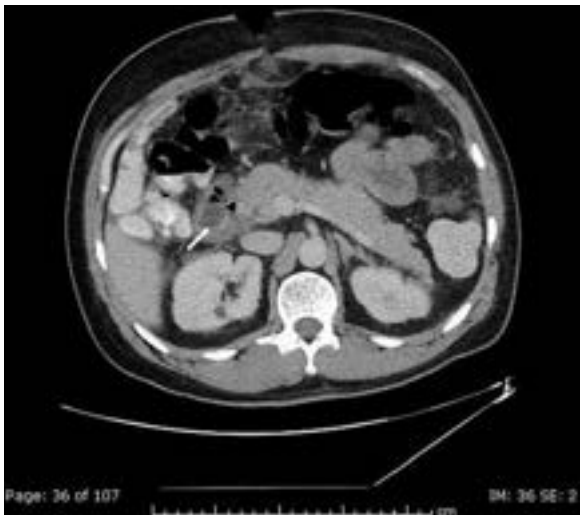

FIGURE 142-8. Peritonitis. A computed tomography scan in a patient with peritonitis showing a thickened duodenal wall from a perforated duodenal ulcer found at the time of surgical exploration. (Courtesy of Charlene Prather, MD.)

Secondary peritonitis is usually caused by a mixed flora of bacteria, including *Escherichia coli, Streptococcus faecalis, Pseudomonas aeruginosa, Klebsiella mirabilis, Bacteroides fragilis, Clostridium* spp., and anaerobic streptococci.

TREATMENT AND PROGNOSIS

Rx

Treatment of acute suppurative peritonitis relies on prompt adequate resuscitation with IV fluids and broad-spectrum IV antibiotics. Treatment regimens include: piperacillin–tazobactam, 3.375 g every 6 hours; ampicillin–sulbactam, 3.0 g every 6 hours; ciprofloxacin, 400 mg every 12 hours, and metronidazole, 1 g every 12 hours; levofloxacin, 750 mg every 24 hours; cefepime, 2 g every 12 hours, and metronidazole, 1 g every 12 hours; or imipenem–cilastatin sodium, 500 mg every 6 hours. Early diagnosis and surgical intervention for acute peritonitis from perforated viscus is critical. The mainstay of treatment for tuberculous peritonitis is at least 6 months of a multidrug regimen (Chapter 324).

Chemical aseptic peritonitis can be complicated by secondary bacterial infection. Treatment is similar to that of acute suppurative peritonitis, with intervention to control the source of peritoneal contamination.

The outcome of peritonitis depends on its cause as well as the rapidity of treatment. The mortality rate can be as low as 15% in patients who have correctable causes, such as a perforated appendix, and who do not develop multiorgan failure before treatment but as high as 50% for postoperative infective peritonitis.

PERITONITIS AS A COMPLICATION OF CHRONIC AMBULATORY PERITONEAL DIALYSIS

The most common complication of chronic ambulatory peritoneal dialysis is infectious peritonitis from bacterial contamination (Chapter 131), which commonly results from poor technique.[21] In contrast to polymicrobial acute suppurative peritonitis, peritonitis complicating chronic ambulatory peritoneal dialysis is typically monomicrobial with gram-positive cocci in the majority of cases and gram-negative species in the remainder of cases. Fungal peritonitis is rare. Symptoms can be milder than in other forms of peritonitis, with patients often presenting with diffuse abdominal pain, low-grade fever, and leukocytosis. The exchange fluid is characteristically turbid and may be the only sign of infection. The diagnosis is based on these physical findings, turbid dialysate with less than 100 leukocytes/μL, and a positive dialysate culture that determines the microbe involved. Treatment is with intraperitoneal antibiotics. The dialysis catheter should be removed in the setting of an inadequate response to therapy, fungal or tuberculous peritonitis or concomitant skin infection.

ADHESIONS

Peritoneal adhesions, which are the most commonly observed cause of bowel obstruction,[22] can occur at any time after a laparotomy. Patients who have had intra-abdominal infection, ischemia, and peritonitis are at increased risk. Patients present with complete or incomplete bowel obstruction, which is usually manifested as colicky abdominal pain, nausea, and vomiting (including feculent vomiting), abdominal distention, and an absence of flatus or stooling. Bowel obstruction can be diagnosed on plain radiographs or CT scan from the presence of dilated bowel and air-fluid levels, with decompressed bowel distal to the site of obstruction (Fig. 142-9). Treatment of bowel obstructions is with NG tube decompression and fluid resuscitation. A nonoperative approach can be attempted in cases of partial bowel obstruction or in patients who have had numerous prior laparotomies, which make surgical exploration more complicated. However, urgent laparotomy for lysis of adhesions must be performed before bowel ischemia develops. In patients with chronic abdominal pain, surgical exploration for the intent of lysing adhesions without clear evidence of obstructing adhesions should be avoided. Patients who develop postoperative adhesions can progress to bowel obstruction and develop recurrent adhesions despite surgical lysis of the adhesions. Data suggest that oxidized regenerated cellulose and hyaluronate carboxymethylcellulose adhesion barriers can safely reduce clinically relevant consequences of adhesions.

PERITONEAL CARCINOMATOSIS AND MALIGNANT ASCITES

Peritoneal carcinomatosis results when malignancy spreads throughout the peritoneal cavity and eventually encases the viscera. The cause can be primary tumors of the peritoneal cavity (e.g., mesothelioma and sarcoma), dissemination of an intra-abdominal malignancy (e.g., gastric, colon, pancreatic,

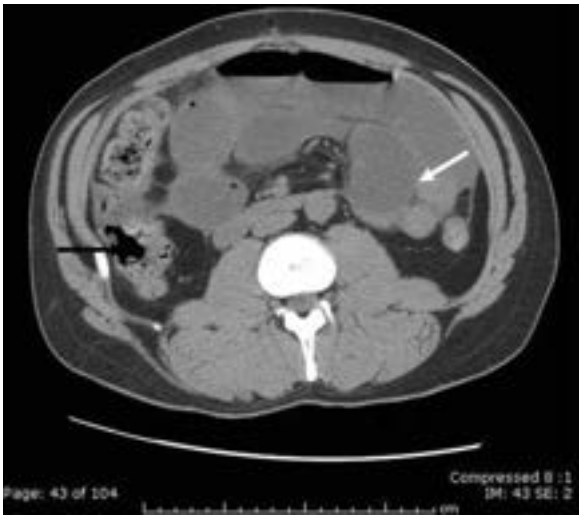

FIGURE 142-9. Small bowel obstruction. A computed tomography scan shows dilated, fluid-filled small bowel *(white arrow)* and a nondilated colon *(black arrow)* in a patient with a small bowel obstructions from adhesions in the midileum found at the time of surgical exploration. (Courtesy of Charlene Prather, MD.)

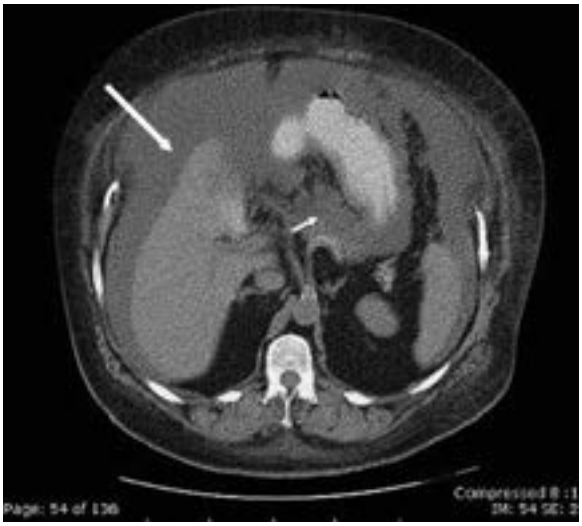

FIGURE 142-10. Malignant ascites. A computed tomography scan shows ascites *(large arrow)* in a patient with gastric cancer *(small arrow)*. (Courtesy of Charlene Prather, MD.)

ovarian, or neuroendocrine tumors), lymphomas, metastatic spread from extra-abdominal malignancy (e.g., breast, lung, or melanoma), and pseudomyxoma peritonei (a rare condition with gelatinous peritoneal implants from mucinous neoplasms of the appendix or ovary).

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The presentation of peritoneal carcinomatosis is nonspecific, with complaints of abdominal pain, anorexia, nausea, vomiting, malaise, and weight loss. Ascites is the most common physical finding. Malignant ascites is rare, accounting for less than 10% of cases of ascites.

Although ultrasound examination may document ascites, CT is preferred to identify carcinomatosis and the origin of the underlying cause (Fig. 142-10). Diagnostic paracentesis can be performed to obtain cell count and differential, culture, and cytologic examination. A serum to ascites albumen ratio of less than 1.1 mg/dL suggests malignant ascites, but the differential diagnosis also includes pancreatic ascites (Chapter 144), nephrotic syndrome (Chapter 121), and peritoneal tuberculosis. Laparoscopy can disclose the typical tumor implant stubbing the peritoneum. Biopsy and appropriate stains can verify the presence of mesothelioma (Chapters 99 and 191) by detecting hyaluronic acid on Alcian stain in a specimen that has a negative carcinoembryonic antigen stain and no mucin on a periodic acid-Schiff stain. By contrast, carcinomas have no hyaluronic acid but show mucin and have a positive carcinoembryonic antigen stain.

TABLE 142-4 MESENTERIC DISEASES

PRIMARY INFLAMMATORY DISEASE
Panniculitis
Retractile mesenteritis
CYSTS
Developmental
Traumatic
Neoplastic
Infectious
TUMORS
Benign tumors
Lipoma
Leiomyoma
Hemangioma
Malignant tumors
Liposarcoma
Leiomyosarcoma
Rhabdomyosarcoma
Metastatic tumor
Mesenteric fibromatosis

TREATMENT AND PROGNOSIS

Rx

Treatment is typically palliative owing to the late presentation of disease. Malignant ascites responds poorly to diuretic therapies, and repeated therapeutic paracentesis may be needed. Recent data suggest some benefit from cytoreductive surgery and hyperthermic intraperitoneal chemotherapy.[23]

DISEASES OF THE MESENTERY AND OMENTUM

The differential diagnosis of mesenteric and omental disorders include a variety of rare disorders, including inflammation, cysts, and tumors that can be benign or malignant (Table 142-4).

Mesenteric panniculitis presents in middle or later age with a slight male predilection. Symptoms typically include nonspecific abdominal pain, weight loss, nausea, vomiting, and low-grade fever, but patients occasionally can present with an acute abdomen. A palpable mass is felt in the majority of patients. The diagnosis can be confirmed by CT scan. Lesions should be biopsied to exclude malignancy but generally are not amenable to complete resection. Significant improvement in the symptoms of this generally self-limited process has been reported with progesterone, corticosteroids, azathioprine, or cyclophosphamide, but evidence to recommend their general use is lacking.

Mesenteric and omental cysts and solid tumors are rare disorders. Patients present with a constellation of nonspecific symptoms. They can be identified with CT scan. Surgical resection will provide the definitive diagnosis and determine whether a benign or malignant condition is present.

Mesenteric fibromatosis (desmoid tumors) is a rare noninflammatory condition that may be associated with familial adenomatosis coli (Chapter 193) and Gardner syndrome (Chapter 193). Fibromatoses are locally aggressive tumors that may present as stable or rapidly growing intraabdominal masses. They have a high rate of recurrence after incomplete surgical resection, in patients with multicentric disease, or if precipitated by surgical trauma itself.

Grade A References

A1. Fitzgibbons RJ Jr, Giobbie-Hurder A, Gibbs JO, et al. Watchful waiting vs repair of inguinal hernia in minimally symptomatic men: a randomized clinical trial. *JAMA*. 2006;295:285-292.

A2. Karthikesalingam A, Markar SR, Holt PJ, et al. Meta-analysis of randomized controlled trials comparing laparoscopic with open mesh repair of recurrent inguinal hernia. *Br J Surg*. 2010;97:4-11.

A3. Kim K, Kim YH, Kim SY, et al. Low-dose abdominal CT for evaluating suspected appendicitis. *N Engl J Med*. 2012;366:1596-1605.

A4. Ohtani H, Tamamori Y, Arimoto Y, et al. Meta-analysis of the results of randomized controlled trials that compared laparoscopic and open surgery for acute appendicitis. *J Gastrointest Surg*. 2012;16:1929-1939.

A5. Varadhan KK, Neal KR, Lobo DN. Safety and efficacy of antibiotics compared with appendicectomy for treatment of uncomplicated acute appendicitis: meta-analysis of randomised controlled trials. *BMJ*. 2012;344:e2156.

A6. St Peter SD, Aguayo P, Fraser JD, et al. Initial laparoscopic appendectomy versus initial nonoperative management and interval appendectomy for perforated appendicitis with abscess: a prospective, randomized trial. *J Pediatr Surg*. 2010;45:236-240.
A7. St Peter SD, Tsao K, Spilde TL, et al. Single daily dosing ceftriaxone and metronidazole vs standard triple antibiotic regimen for perforated appendicitis in children: a prospective randomized trial. *J Pediatr Surg*. 2008;43:981-985.
A8. Biondo S, Golda T, Kreisler E, et al. Outpatient versus hospitalization management for uncomplicated diverticulitis: a prospective, multicenter randomized clinical trial (DIVER Trial). *Ann Surg*. 2014;259:38-44.
A9. Andreyev HJ, Benton BE, Lalji A, et al. Algorithm-based management of patients with gastrointestinal symptoms in patients after pelvic radiation treatment (ORBIT): a randomised controlled trial. *Lancet*. 2013;382:2084-2092.
A10. ten Broek RP, Stommel MW, Strik C, et al. Benefits and harms of adhesion barriers for abdominal surgery: a systematic review and meta-analysis. *Lancet*. 2014;383:48-59.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Park JJ, Wolff BG, Tollefson MK, et al. Meckel diverticulum: the Mayo Clinic experience with 1476 patients (1950-2002). *Ann Surg*. 2005;241:529-533.
2. Uppal K, Tubbs RS, Matusz P, et al. Meckel's diverticulum: a review. *Clin Anat*. 2011;24:416-422.
3. Ruscher KA, Fisher JN, Hughes CD, et al. National trends in the surgical management of Meckel's diverticulum. *J Pediatr Surg*. 2011;46:893-896.
4. Martin MJ, Steele SR. Twists and turns: a practical approach to volvulus and intussusception. *Scand J Surg*. 2010;99:93-102.
5. Tan KK, Chong CS, Sim R. Management of acute sigmoid volvulus: an institution's experience over 9 years. *World J Surg*. 2010;34:1943-1948.
6. Halabi WJ, Jafari MD, Kang CY, et al. Colonic volvulus in the United States: trends, outcomes, and predictors of mortality. *Ann Surg*. 2014;259:293-301.
7. Potts J, Al Samaraee A, El-Hakeem A. Small bowel intussusception in adults. *Ann R Coll Surg Engl*. 2014;96:11-14.
8. Varban OA, Ardestani A, Azagury DE, et al. Contemporary management of adult intussusception: who needs a resection? *World J Surg*. 2013;37:1872-1877.
9. Jensen KK, Kjaer M, Jorgensen LN. Abdominal muscle function and incisional hernia: a systematic review. *Hernia*. 2014;18:481-486.
10. Zendejas B, Ramirez T, Jones T, et al. Incidence of inguinal hernia repairs in Olmsted County, MN: a population-based study. *Ann Surg*. 2013;257:520-526.
11. Rao KS, Prabhakar J, Sudhakar W. Progress in the diagnosis of appendicitis. *Ann Surg*. 2015; 261:e88-e89.
12. Yu CW, Juan LI, Wu MH, et al. Systematic review and meta-analysis of the diagnostic accuracy of procalcitonin, C-reactive protein and white blood cell count for suspected acute appendicitis. *Br J Surg*. 2013;100:322-329.
13. Shahedi K, Fuller G, Bolus R, et al. Long-term risk of acute diverticulitis among patients with incidental diverticulosis found during colonoscopy. *Clin Gastroenterol Hepatol*. 2013;11: 1609-1613.
14. Morris AM, Regenbogen SE, Hardiman KM, et al. Sigmoid diverticulitis: a systematic review. *JAMA*. 2014;311:287-297.
15. Unlu C, Daniels L, Vrouenraets BC, et al. A systematic review of high-fibre dietary therapy in diverticular disease. *Int J Colorectal Dis*. 2012;27:419-427.
16. Chan FK. Proton pump inhibitors and nonsteroidal anti-inflammatory drug-related lower gastrointestinal adverse events. *Clin Gastroenterol Hepatol*. 2014;12:904-906.
17. Shafi MA, Bresalier RS. The gastrointestinal complications of oncologic therapy. *Gastroenterol Clin North Am*. 2010;39:629-647.
18. Theis VS, Sripadam R, Ramani V, et al. Chronic radiation enteritis. *Clin Oncol (R Coll Radiol)*. 2010;22:70-83.
19. Shadad AK, Sullivan FJ, Martin JD, et al. Gastrointestinal radiation injury: prevention and treatment. *World J Gastroenterol*. 2013;19:199-208.
20. Weledji EP, Ngowe MN. The challenge of intra-abdominal sepsis. *Int J Surg*. 2013;11:290-295.
21. Segal JH, Messana JM. Prevention of peritonitis in peritoneal dialysis. *Semin Dial*. 2013;26: 494-502.
22. ten Broek RP, Issa Y, van Santbrink EJ, et al. Burden of adhesions in abdominal and pelvic surgery: systematic review and met-analysis. *BMJ*. 2013;347:f5588.
23. Brucher BL, Piso P, Verwaal V, et al. Peritoneal carcinomatosis: cytoreductive surgery and HIPEC—overview and basics. *Cancer Invest*. 2012;30:209-224.

REVIEW QUESTIONS

1. An 18-year-old college freshman reports to the Student Health Center. He complains of a 12-hour history of abdominal pain and fever. He complains of anorexia, nausea, and vomiting. He denies diarrhea and blood per rectum. His past history is unremarkable. His family history is notable for heart disease, hypertension, and diverticulosis coli. On physical examination, he is febrile to 38° C. His abdomen has normal bowel sounds but is tender in the right lower quadrant with rebound tenderness. On laboratory examination, his WBC count is elevated at 11,000/μL with a left shift. His urinalysis reveals 1+ protein. The most likely diagnosis in this patient is

A. acute colonic diverticulitis.
B. ulcerative colitis.
C. acute appendicitis.
D. Behçet disease.
E. nephrolithiasis and renal colic.

Answer: C This patient has an acute appendicitis. He has typical presenting signs and symptoms. Acute diverticulitis typically presents at an older age, and it most commonly localizes to the left side of the abdomen. Although ulcerative colitis presents at this age, its symptoms are typically bloody diarrhea and tenesmus. This patient does not have the multisystem presentation that is common for Behçet disease, in which small bowel involvement is rare. His normal urinalysis result makes a diagnosis of nephrolithiasis and renal colic unlikely.

2. A 64-year-old woman reports to the emergency department with a complaint of 2 days of left lower quadrant pain, fever, and diarrhea, as well as nausea and vomiting. She denies recent travel and antibiotic use. She admits to many years of constipation and uses laxatives intermittently. Her past history is notable for hypertension and glaucoma. Her medications include hydrochlorothiazide 25 mg/day and eye drops for glaucoma. Her family history is positive for hypertension, diabetes mellitus, and coronary artery disease. Her physical examination is notable for temperature of 37.6° C, and a heart rate of 105 beats/min. Heart examination shows a regular tachycardia and a grade 2/6 midsystolic murmur. Her abdomen is soft with tenderness in the left lower quadrant. There is no distention or involuntary rebound. On digital rectal examination, there is no stool in the rectal vault, and you palpate internal hemorrhoids. The most like diagnosis is

A. Meckel diverticulum.
B. acute colonic diverticulitis.
C. Crohn colitis.
D. colonic intussusception.
E. acute intestinal obstruction.

Answer: B This patient has the typical presenting signs and symptoms of sigmoid colon diverticulitis. Complications from Meckel diverticulum, such as right lower quadrant pain from ulceration and bleeding, are not present. Crohn colitis can occur in this age group, but the presentation usually is more insidious with diarrhea and abdominal pain. The patient does not have the symptoms of bowel obstruction or the currant jelly stools that are typical of colonic intussusception.

3. An 82-year-old man is referred to the emergency department from a nursing home. He complains of abdominal distension and progressive constipation. He has not passed flatus or stool for 2 days. His history is notable for dementia and a prior stroke. He has used laxatives for constipation for many years. His other medications include clopidogrel and aspirin. He is allergic to intravenous contrast dye. On examination, you find he is afebrile with a blood pressure of 155/93 mm Hg and a tachycardia of 110 beats/min. His abdomen is distended and tympanic to percussion, with a diffuse tenderness. Bowel sounds are decreased. A small amount of stool in the rectal vault is brown but Hemoccult positive. Laboratory tests reveal a lactate of 2 and a normal WBC and Hgb levels. The next step in his management is

A. exploratory laparotomy.
B. abdominal radiographs.
C. bleeding scan.
D. mesenteric angiography.
E. barium upper GI series.

Answer: B This patient has a sigmoid volvulus. A volvulus is commonly seen in institutionalized patients and in patients with a history of chronic constipation and laxative use. The presence of a volvulus or intestinal obstruction from other causes can be investigated cost effectively using plain radiographs as the initial imaging modality. There is no massive gastrointestinal (GI) bleeding to warrant a bleeding scan, angiography, or upper GI series.

4. A 55-year-old woman presents to your office complaining of vague abdominal pain and fatigue. Her past history is notable for osteoarthritis. She is postmenopausal and on hormone replacement therapy. Her medications also include ibuprofen 600 mg tid. She has noted intermittent black tarry stools. Physical examination is unremarkable. Laboratory examination reveals Hgb 8.7 g/dL with an MCV 79. You obtain and esophagogastroduodenoscopy (EGD) and a colonoscopy for her anemia and abdominal complaints. EGD is normal, and she has mild sigmoid diverticulosis coli. The most likely cause of her anemia is

A. diverticular bleeding.
B. gastric ulcer bleeding.
C. Meckel diverticulum.
D. Crohn disease.
E. NSAID-induced enteropathy.

Answer: E This patient has nonsteroidal anti-inflammatory drug (NSAID)–induced enteropathy. The episodes of melena are suggestive of ulceration and bleeding from the upper gastrointestinal (GI) tract, but peptic ulcer disease was not seen on EGD. She regularly uses NSAIDs, which can be a cause of secondary small intestinal ulcerations. Meckel diverticulum and bleeding are more rare than NSAID-induced enteropathy. Although sigmoid diverticuli were identified on colonoscopy, diverticular bleeding presents as massive lower GI bleeding, not melena. Crohn disease can present with anemia and bleeding, but melena is rare.

5. A 63-year-old man presents to your office with a complaint of 5 months of a poor appetite, a 35-lb weight loss, and low grade fevers up to 37.4° C. Over the past month, he has noted increasing abdominal girth and tenderness. He has been otherwise healthy and takes no medications. His family history is notable for stroke. He denies tobacco use. He drinks two glasses of wine each night. He retired at age 55 years and has traveled extensively throughout Southeast Asia. On physical examination, you note that his abdomen is markedly distended and diffusely tender. Bowel sounds are normal, but a fluid wave is present. Findings of the laboratory examination, including a comprehensive metabolic profile and complete blood count, are normal. A diagnostic paracentesis reveals protein, 3.5 g/dL; neutrophils, 800/μL; glucose, 20 mg/dL; and amylase, 50 IU/L. No bacteria are seen on stain, and fluid cultures show no growth. The most likely diagnosis for this man's ascites is

A. alcoholic cirrhosis.
B. pancreatic ascites.
C. nephrotic syndrome.
D. metastatic lung cancer.
E. tuberculous peritonitis.

Answer: E This patient has tuberculous peritonitis, which characteristically presents insidiously in patients who have been exposed to tuberculosis, as this patient probably was exposed during extensive travel in endemic regions like Southeast Asia. The ascites is typically high in protein and neutrophils and low in glucose. The patient's alcohol use and the characteristics of his ascites make alcoholic cirrhosis an unlikely cause. The ascitic fluid amylase level is normal, so pancreatic ascites is unlikely. Ascitic fluid in adult patients with the nephrotic syndrome is low in protein and is associated with hypoalbuminemia and heart failure. The characteristics of his ascites are not typical of metastatic carcinomatosis.

TABLE 143-1 CONDITIONS PREDISPOSING TO SECONDARY MESENTERIC ISCHEMIA

Adhesions
Herniation
Volvulus
Intussusception
Mesenteric fibrosis
Retroperitoneal fibrosis
Carcinoid syndrome
Malignancy (peritoneal, mesenteric, colonic)
Neurofibromatosis
Amyloidosis
Trauma

143

VASCULAR DISEASES OF THE GASTROINTESTINAL TRACT

STEPHEN CRANE HAUSER

INTESTINAL ISCHEMIA

Intestinal ischemia can occur as a result of a variety of conditions that decrease intestinal blood flow. Both diminished arterial blood flow to the gut and compromised venous circulation from the intestine can cause intestinal or mesenteric ischemia. Several conditions, such as adhesions and malignancy (Chapter 193), may predispose to mesenteric ischemia by secondarily diminishing blood flow through extrinsic compression of otherwise normal intestinal arteries or veins (Table 143-1). These disorders and esophageal varices (Chapters 135 and 153) are discussed elsewhere.

EPIDEMIOLOGY

Intestinal ischemia is responsible for about one per 1000 hospital admissions. When considering the diagnosis of intestinal ischemia, it is important to distinguish *primary* (occlusive or non-occlusive) from *secondary* (extrinsic to the blood vessel) mesenteric ischemia, *acute* manifestations from *chronic, arterial* versus *venous,* and *small bowel* versus *colonic* ischemia. Risk factors for intestinal ischemia include older age (all of the disorders discussed) and conditions that predispose to arterial embolism (e.g., cardiac arrhythmias, cardioversion, heart failure, cardiomegaly, dyskinesia, valvular heart disease, recent myocardial infarction, cardiac catheterization, intracardiac thrombus, atheromatous cholesterol embolism), occlusion of arteries (atherosclerosis, fibromuscular dysplasia, abdominal aortic aneurysm, trauma, vasculitis), low-flow states (sepsis, dialysis, reduced cardiac output, vasoconstrictive drugs), and pathologic thromboses (largely venous; hypercoagulable and hyperviscosity states, portal hypertension, trauma, malignancy, inflammation).

PATHOBIOLOGY

Arterial or venous disease of the esophagus, stomach, duodenum, and rectum is very unusual for anatomic reasons. The esophagus receives its main blood supply segmentally through multiple small vessels from the aorta, right intercostal artery, bronchial arteries, inferior thyroid artery, left gastric artery, short gastric artery, and left phrenic artery. Likewise, the stomach, duodenum, and rectum have numerous arterial inputs with rich collateralization. Patients who have undergone extensive surgical resection of the esophagus, stomach, or duodenum are at increased risk for ischemia. Vasculitic disorders, which can involve small or large arteries or veins, may affect the esophagus, stomach, duodenum, or rectum.

The arterial supply of blood to the small and large intestine is from the *celiac artery, superior mesenteric artery* (SMA), and *inferior mesenteric artery* (IMA). Collateral vessels, which vary from person to person, may include the meandering mesenteric artery or arc of Riolan at the base of the mesentery (connecting the SMA and IMA), the marginal artery of Drummond along the mesenteric border (connecting the SMA and IMA), the pancreaticoduodenal arcade (connecting the celiac artery and SMA), the arc of Barkow (connecting the celiac artery and SMA), and the arc of Buhler (connecting the celiac artery and SMA). These collaterals can rapidly enlarge in response to localized mesenteric ischemia. During states of low arterial flow, such as in patients with low systemic arterial blood pressure, "watershed" areas such as the splenic flexure, which is the farthest away from arterial flow, are more likely to be involved. By contrast, when a major arterial vessel such as the IMA is suddenly occluded, the splenic flexure is less likely to be involved because of collaterals from the SMA circulation.

Intestinal blood flow, which accounts for approximately 10% of the cardiac output, increases to as much as 25% of the cardiac output after eating a meal. Blood flow to the intestine is regulated by the sympathetic nervous system and a variety of systemic (angiotensin II, vasopressin) and local (prostaglandins, leukotrienes) humoral factors.

Mesenteric ischemia can occur as a result of decreased *arterial* blood flow, which can be *occlusive* (arterial embolus, arterial thrombus, and vasculitis) or *nonocclusive* (low-flow states). *Venous* obstruction (thrombosis, vasculitis) can also result in mesenteric ischemia.

Whatever the cause of mesenteric ischemia, the gut is able to adapt to as much as a 75% reduction in normal blood flow for as long as 12 hours. Increased flow through available and newly opened collateral vessels and increased oxygen extraction help compensate. However, with a more prolonged and more severe reduction in blood flow, generalized mesenteric arterial vasoconstriction often develops and can become irreversible, even with correction of the original underlying condition (i.e., relief of focal arterial obstruction or resolution of a low-flow state). Hypoxia and reperfusion injury by oxygen radicals, reduced endothelial synthesis of nitric oxide, and an enhanced cellular inflammatory response cause microvascular and end-organ damage. Initially, the end-organ damage is primarily mucosal, but damage can rapidly progress to transmural necrosis (gangrene). Some ischemic segments of bowel will heal with fibrosis (strictures).

CLINICAL MANIFESTATIONS

Symptoms of small intestinal ischemia at initial evaluation may be acute (sudden, lasting hours),[1] subacute (days), or chronic (intermittent, occurring over a period of weeks to months).[2] With acute and many subacute manifestations, abdominal pain is often the cardinal symptom. Usually the pain is severe, persistent and periumbilical or poorly localized. Initially, the pain is typically more severe than the findings on abdominal palpation (i.e., pain out of proportion to tenderness). With or without pain, other initial features may include fever, altered mental status, abdominal distention, difficulty eating, nausea, vomiting, and diarrhea. With small bowel ischemia, overt gastrointestinal (GI) bleeding (Chapter 135) is a late and ominous finding that often suggests small bowel infarction.

Findings on physical examination can include hypotension, tachycardia, abdominal distention, initially increased and later decreased bowel sounds, and nonspecific diffuse abdominal tenderness, often mild at first. Over time, peritoneal signs with localized to generalized abdominal tenderness, rebound, and rigidity may become manifest. Occult GI bleeding can be an early finding.

DIAGNOSIS

As the diagnostic evaluation commences, appropriate attention must be directed concurrently to emergent therapy, including fluid resuscitation, antibiotics, and invasive procedures (Fig. 143-1).

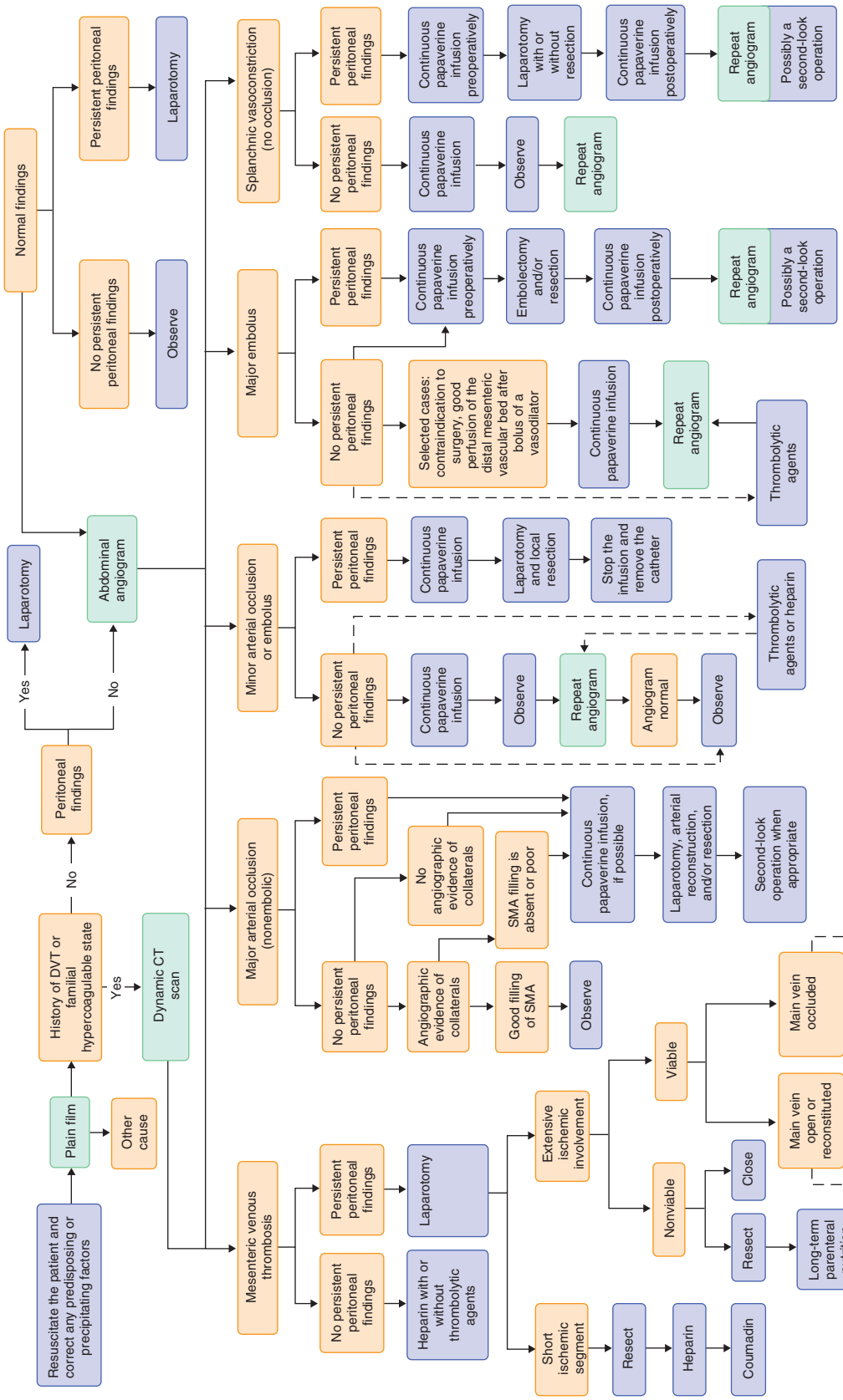

FIGURE 143-1. Algorithm for managing patients with suspected acute mesenteric ischemia: diagnosis and management. *Solid lines* indicate an accepted management plan; *dashed lines* indicate an alternative management plan. CT = computed tomography; DVT = deep vein thrombosis; SMA = superior mesenteric artery. (From American Gastroenterological Association Medical Position Statement: guidelines on intestinal ischemia. *Gastroenterology*. 2000;118:951-953 [corrected algorithm in *Gastroenterology*. 2000;119:281].)

Initial Diagnostic Evaluation

The initial laboratory findings in patients with an acute onset of small bowel ischemia can be entirely normal. Nonspecific abnormalities such as leukocytosis with a predominance of neutrophils and hemoconcentration may be observed. Elevated serum levels of amylase, lactate, aminotransferases, lactate dehydrogenase, creatine kinase, and phosphate often portend more advanced (necrotic) small bowel ischemia, but these findings lack sensitivity as well as specificity.

Noninvasive Imaging

The presence or absence of radiographic features suggestive of ischemia in patients with acute-onset mesenteric ischemia varies and depends on the duration and extent of ischemia. Plain abdominal radiographs are useful in helping exclude secondary causes of mesenteric ischemia, as well as other causes of acute abdominal pain, nausea, vomiting, or distention, such as obstruction and perforation. Radiographic findings such as "thumbprinting" (caused by submucosal hemorrhage), an ileus pattern, or formless loops of small bowel, or with more advanced disease, pneumatosis intestinalis or portal venous gas (often a sign of transmural necrosis or gangrene) occasionally may be observed. Contrast-enhanced abdominal-pelvic computed tomography (CT) is also helpful to exclude alternative diagnoses. CT may demonstrate entirely normal findings in acute mesenteric ischemia, or findings such as segmental bowel wall thickening, submucosal hemorrhage, mesenteric stranding, mesenteric venous thrombosis, pneumatosis, and portal venous gas may be present (Fig. 143-2). Multidetector CT angiography (CTA) has a greater sensitivity and specificity (each up to 95%) than traditional CT (≈65%) and is the preferred imaging study to diagnose acute small bowel mesenteric ischemia. Magnetic resonance angiography (MRA) is less sensitive for more peripheral emboli. Subacute manifestations of bowel ischemia may be due to a wide variety of causes, including mesenteric venous thrombosis, which is best diagnosed by CT scan. However, the time needed to obtain a CT scan should not delay resuscitation or arteriography in very ill patients with suspected acute-onset ischemia.

TREATMENT

Rx

Acutely ill patients require prompt, definitive diagnosis and treatment,[3] which often requires selective mesenteric angiography (Fig. 143-3). Options for arterial reconstruction in appropriately selected patients include open surgery or endovascular treatment.[4] If transmural intestinal necrosis (gangrene) is suspected from peritoneal signs, pneumatosis, or portal venous gas on imaging procedures, emergency laparotomy is indicated. The presence of predisposing conditions (e.g., arrhythmias, systemic hypotension) and their extraintestinal manifestations (e.g., heart failure, sepsis, respiratory insufficiency, acute renal failure, anemia) dictate the initial therapy,, which includes volume replacement, optimization of cardiac output, management of respiratory function, avoidance of splanchnic vasoconstrictors such as digoxin, and administration of broad-spectrum antibiotics (e.g., meropenem, imipenem–cilastatin, metronidazole and a third-generation cephalosporin, ciprofloxacin and metronidazole, or piperacillin) until symptoms resolve to cover aerobic gram-negative and anaerobic organisms and to prevent sepsis secondary to translocation of bacteria across ischemic gut mucosa.

PROGNOSIS

Acute primary arterial mesenteric ischemia involving the small bowel is an urgent condition, which, if unidentified or untreated, can result in death within hours. Mortality rates may be as high as 70% but are much lower with early diagnosis and prompt therapy. Overall, patients with colonic ischemia have a much better prognosis than do those with small bowel ischemia. Patients with mesenteric venous thrombosis also have a much better prognosis than do those with acute primary arterial mesenteric ischemia of the small intestine.

Specific Ischemic Bowel Syndromes

SUPERIOR MESENTERIC EMBOLISM

Embolization to the intestine through the SMA *(SMA embolus)* accounts for 5% of peripheral emboli and nearly 50% of cases of primary noncolonic mesenteric ischemia. Emboli originate most commonly from the heart, with

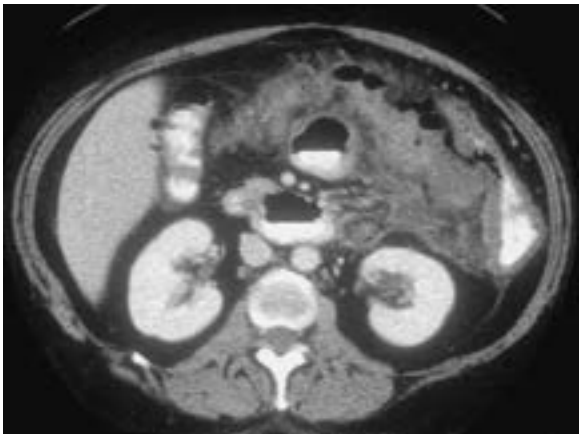

FIGURE 143-2. Computed tomography of the abdomen in a patient with ischemic colitis as a result of superior mesenteric vein thrombosis. A segmental area of the transverse colon demonstrates a thick wall, as well as considerable fluid and soft tissue stranding in the adjacent mesentery. (From Johnson CL, Schmit GD, eds. *Mayo Clinic Gastrointestinal Imaging Review*. Boca Raton, FL: Mayo Clinic Scientific Press, Taylor and Francis Group; 2005. By permission of the Mayo Foundation for Medical Education and Research. All rights reserved).

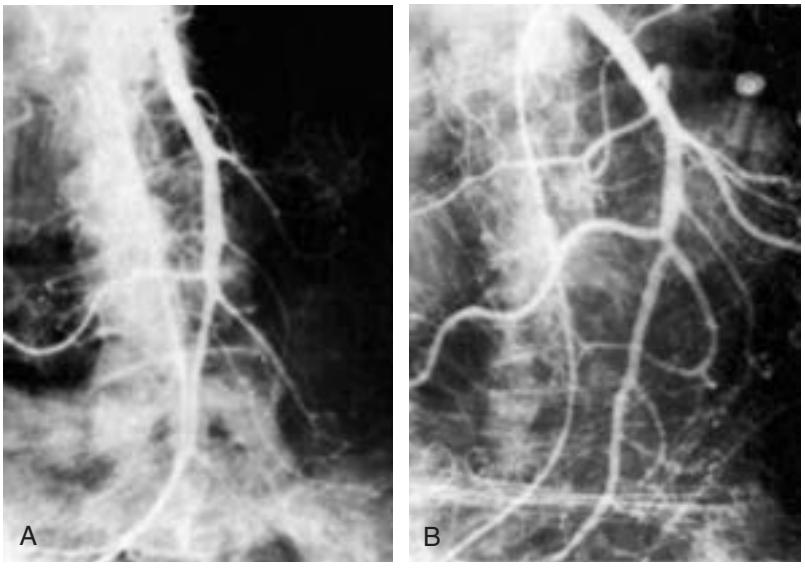

FIGURE 143-3. Selected films from superior mesenteric angiography. A, Diffuse vasoconstriction characteristic of nonocclusive mesenteric ischemia. **B,** Intra-arterial infusion of papaverine (30-60 mg/hr) resulted in vasodilation.

TABLE 143-2 CONDITIONS ASSOCIATED WITH EMBOLIZATION TO THE GASTROINTESTINAL TRACT

Cardiac arrhythmias
Valvular heart disease
Heart failure
Myocardial infarction
Intracardiac thrombus
Cardiac catheterization
Cardioversion
Atherosclerosis of the aorta

an aortic origin being less common (Table 143-2), and tend to obstruct beyond the origin of the SMA.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients who are evaluated early in the course of their illness may have entirely normal CT scans, or the CT findings may be consistent with mesenteric ischemia without features that would suggest alternative diagnoses (e.g., perforation, obstruction). Multidetector CTA is much more likely than standard CT to diagnose embolic disease reliably in the mesenteric arterial vasculature. Selective mesenteric angiography offers the possibility of therapy as well as diagnosis.

TREATMENT Rx

Select patients with acute onset of a partial or small SMA branch occlusion may be candidates for thrombolytic therapy (e.g., streptokinase, urokinase, tissue plasminogen activator) infused through an arterial catheter directly into the vicinity of the embolus; this therapy can lyse the embolus and resolve symptoms such as abdominal pain. Because segmental arterial embolic occlusion of a small portion of the SMA vascular bed results in widespread splanchnic visceral arterial vasoconstriction, which may persist even after the original inciting event (i.e., an embolus) is rectified, infusion of a vasodilator such as papaverine (often given as a 60-mg bolus followed by a continuous infusion of 30 to 60 mg/hr for 12-48 hours) through an arterial catheter reverses this reflex vasoconstriction and improves the outcome, including mortality rates. The same scenario occurs with other arterial occlusive lesions (SMA thrombi), arterial nonocclusive disease (nonocclusive mesenteric ischemia), and disorders associated with mesenteric venous occlusion.

Patients evaluated in the course of their acute embolic illness with peritoneal signs require laparotomy, with or without resection and with or without embolectomy, which is usually performed during surgical exploration. A "second-look" operation 24 hours after embolectomy to make sure that all necrotic tissue has been resected may be necessary.

Any patient in whom SMA embolization is diagnosed requires preoperative systemic anticoagulation (e.g., intravenous heparin) to prevent propagation of clot around the embolus and to guard against further embolization to the intestine or other organs (i.e., brain, coronary arteries, kidneys, extremities). Anticoagulation is usually discontinued before surgery and is often resumed 24 to 48 hours postoperatively, depending on the operative findings. Mortality can be as high as 70%.

SUPERIOR MESENTERIC THROMBOSIS

Thrombosis of the SMA *(SMA thrombus)* accounts for nearly 15% of cases of primary noncolonic mesenteric ischemia. Risk factors include older age, atherosclerosis (e.g., hypertension, diabetes mellitus, hyperlipidemia, smoking history), low-flow states, hypercoagulable states, and less often vasculitis and aortic or mesenteric aneurysms.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Nearly one-third of these patients have a history of symptomatic chronic mesenteric ischemia (see later) antedating their acute manifestation of SMA thrombosis. Proximal mesenteric arterial occlusions are well recognized by multidetector CTA, MRA, and Doppler ultrasonography, but similar abnormalities are common in asymptomatic elderly persons. Similar to acute SMA embolism, the diagnosis is confirmed by selective mesenteric angiography, with intra-arterial infusion of a vasodilator used to reverse reflex-generalized vasoconstriction.

TREATMENT Rx

Although thrombolytic therapy has been helpful in a limited number of case reports, surgical thrombectomy or bypass grafting, with or without bowel resection, is the most common therapeutic approach. Because many thrombi occur near the SMA origin, angioplasty may be therapeutic in very select cases, but the risk for reocclusion is high. Similar to acute SMA embolism, anticoagulation (intravenous heparin) is important preoperatively and at some point postoperatively in the acute state, as is administration of broad-spectrum antibiotics (see Small Intestinal Ischemia).

ACUTE NONOCCLUSIVE, NONCOLONIC PRIMARY ARTERIAL ISCHEMIA

Nonocclusive mesenteric ischemia, which accounts for about 20% of primary noncolonic mesenteric ischemia cases, is caused by low arterial blood flow to the intestine. Risk factors include advanced age, decreased systolic blood pressure (e.g., cardiac arrhythmia, heart failure, myocardial infarction, shock, sepsis, burns, pancreatitis, hemorrhage, multiple organ failure, dialysis, perioperative states), vasospasm (e.g., digoxin, vasopressin, amphetamines, cocaine), and atherosclerotic disease.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The clinical findings are generally indistinguishable from those of embolic or thrombotic vascular disease except that symptoms may be less acute. As a result, patients initially may be seen without acute abdominal pain but rather with more nonspecific symptoms such as distention, nausea, emesis, diarrhea, fever, altered mental status, and borderline or low systolic blood pressure. Selective mesenteric angiography establishes the diagnosis (lack of embolus or thrombus, alternating areas of vessel spasm and dilation, vascular pruning and spasm).

TREATMENT Rx

The best specific treatment is prolonged intra-arterial instillation of a vasodilator (e.g., papaverine, often given as a 60-mg bolus followed by a continuous infusion of 30-60 mg/hr) to reverse the vasospasm. Avoidance of vasospastic medications and optimization of cardiac output, blood volume, and blood pressure are crucial. Anticoagulation is generally not necessary, but broad-spectrum antibiotics (similar to those recommended earlier) should be administered to cover aerobic gram-negative and anaerobic organisms. Although many of these patients have serious conditions that predispose them to low-flow states and their ultimate prognosis depends on the outcomes of these serious conditions, the diagnosis and treatment of acute nonocclusive mesenteric ischemia with therapeutic angiography can be life saving.

MESENTERIC VENOUS THROMBOSIS

Occlusive disease of the mesenteric venous circulation *(mesenteric venous thrombosis)* usually involves the superior mesenteric vein (SMV) and may be accompanied by symptoms that are acute (hours to days) or subacute (weeks to months) in onset.[5]

EPIDEMIOLOGY AND PATHOBIOLOGY

Thrombosis of the SMV accounts for about 10% of cases of primary noncolonic mesenteric ischemia. Colonic involvement with ischemic colitis is much less common. In contrast to arterial occlusive disease, risk factors for and causes of SMV thrombosis are more numerous and diverse. Individuals with a personal or family history of a hypercoagulable state or deep vein thrombosis are at increased risk for SMV thrombosis. Hypercoagulable states, hyperviscosity syndromes, portal hypertension, intra-abdominal infections (e.g., pyelophlebitis, diverticulitis, appendicitis) or inflammation (e.g., Crohn disease, pancreatitis), malignancy, vasculitis, and trauma may all cause thrombosis of the SMV (Table 143-3).

CLINICAL MANIFESTATIONS

Symptoms in acute-onset cases are similar to those observed in acute occlusive and nonocclusive arterial mesenteric ischemia—abdominal pain, anorexia, nausea, vomiting, abdominal fullness, diarrhea, and constipation—but tend to persist over a longer time period. Some patients may have bacteremia, especially infection with *Bacteroides* spp. GI hemorrhage, if present, is

TABLE 143-3 RISK FACTORS FOR MESENTERIC VENOUS THROMBOSIS

HYPERCOAGULABLE AND HYPERVISCOSITY STATES
Protein S deficiency
Protein C deficiency
Antithrombin III deficiency
Factor V Leiden mutation
Hyperfibrinogenemia
Antiphospholipid syndrome
Primary myeloproliferative neoplasm
Sickle cell disease
Estrogen or progesterone
INTRA-ABDOMINAL INFECTIONS AND INFLAMMATION
Appendicitis
Diverticulitis
Crohn disease
Abscess
Pancreatitis
Cholecystitis
Pyelophlebitis
Neonatal omphalitis
PORTAL HYPERTENSION
Variceal sclerotherapy
MALIGNANCY
TRAUMA
VASCULITIS

often indicative of infarction. However, many patients with SMV thrombosis experience more vague symptomatic abdominal pain, nausea, distention, or diarrhea over a period of weeks to months (subacute).

DIAGNOSIS

Abdominal-pelvic contrast-enhanced CT, which is the preferred diagnostic test, usually (>90% sensitivity) demonstrates SMV thrombosis with or without portal vein or splenic vein thrombosis. By definition, chronic mesenteric venous thrombosis is asymptomatic and usually detected as an incidental CT finding in patients with portal hypertension, pancreatitis (acute or chronic), or malignancy. The presence of abundant collateral vessels suggests chronic or sometimes subacute mesenteric venous obstruction.

Small bowel radiography may demonstrate segmental bowel wall thickening and separation of bowel loops. Selective mesenteric angiography is not generally necessary.

TREATMENT Rx

Therapy for acute-onset cases may include laparotomy with or without bowel resection when infarction is suspected, fluid resuscitation, broad-spectrum antibiotics (similar to those recommended earlier to cover aerobic gram-negative and anaerobic organisms), avoidance of vasoconstrictors, and anticoagulation (e.g., intravenous heparin) in the absence of GI bleeding. Selected patients may be candidates for thrombolytic therapy (e.g., streptokinase, urokinase, tissue plasminogen activator) followed by anticoagulation. Underlying conditions such as hypercoagulable states, portal hypertension, intra-abdominal infections, intra-abdominal inflammation, and malignancy require concomitant diagnosis and therapy. The indications for anticoagulation in the chronic setting are uncertain, and it is generally avoided in patients who have portal hypertension but do not have symptoms related to their mesenteric venous thrombosis.

CHRONIC MESENTERIC ISCHEMIA

Chronic atherosclerotic stenosis of the visceral arteries is the cause of most cases of chronic mesenteric ischemia, sometimes called *intestinal angina.*

EPIDEMIOLOGY AND PATHOBIOLOGY

Risk factors for chronic mesenteric ischemia are principally older age and the same risk factors for atherosclerosis. Some patients may develop chronic mesenteric ischemia after their malignancy is treated with radiotherapy, chemotherapy, or both. Rarely, vasculitis or an aortic aneurysm can be manifested as chronic mesenteric ischemia. Atherosclerotic stenoses usually involve the origins of two or all of the three major visceral arteries supplying the intestine. However, many age-matched patients also harbor atherosclerotic lesions and do not have symptoms of chronic mesenteric ischemia.

CLINICAL MANIFESTATIONS

Patients typically complain of episodic ischemic abdominal pain. The pain is usually upper or midabdominal, typically begins 15 to 30 minutes after a meal, lasts 1 to 3 hours, and progresses in severity over time, as well as occurs after smaller meals and more frequently after meals. Patients may lose weight as a result of fear of eating (sitophobia). Nausea, vomiting, bloating, diarrhea, and constipation can also occur. Malabsorption with steatorrhea; otherwise unexplained gastroduodenal ulcerations; and small bowel biopsy findings of villous atrophy, nonspecific surface cell flattening, and chronic inflammation may be seen in some patients. More than half of patients have a bruit on abdominal examination. In some patients with episodic symptoms, acute thrombotic mesenteric ischemia develops suddenly.

DIAGNOSIS AND TREATMENT Rx

Atherosclerotic lesions usually can be identified by Doppler ultrasonography because they are proximal in these vessels and demonstrate increased flow velocity through areas of marked stenosis. Multidetector CTA and MRA are also useful to screen for arterial stenoses consistent with chronic mesenteric ischemia in symptomatic patients, but neither technique is adequately sensitive to exclude the diagnosis of chronic mesenteric ischemia when the pretest probability is high. Thus, selective mesenteric angiography is important to ensure that the anatomic findings are consistent with the diagnosis. Patients must be evaluated thoroughly to exclude other causes of abdominal pain (i.e., gastric cancer, gastroparesis, gastric volvulus, partial small bowel obstruction, small bowel bacterial overgrowth states, pancreatic cancer, biliary disease, paraesophageal hernias). For symptomatic patients with appropriate findings on angiography and no other causes of symptoms, surgical reconstruction provides better long-term outcomes than angioplasty and stenting in patients who are not at too high risk for surgery.[6]

ISCHEMIC COLITIS

EPIDEMIOLOGY AND PATHOBIOLOGY

Ischemic colitis, which is the single most common cause of mesenteric ischemia, accounts for nearly 50% of all cases and for almost one in 2000 hospital admissions. Many cases are acute and self-limited and occur in persons older than 60 years without any apparent cause; these cases are probably attributable to transient nonocclusive hypoperfusion involving a segment of the colon. It is controversial whether subtle hypercoagulable states contribute to the pathogenesis of idiopathic cases. Atherosclerotic or thrombotic occlusion of the IMA or its branches and low-flow states are recognizable causes of ischemic colitis. Less common causes include hypercoagulable states (especially in younger persons); iatrogenic ligation of the IMA (e.g., with aortic surgery); embolism; vasculitis; and any cause of colonic obstruction, including malignancy, stricture, and fecalith, that can produce localized compression of the vasculature with an upstream segment of ischemia. Other unusual associations include long-distance running (dehydration, mechanical trauma to the vasculature, generally involving the cecum), pit viper bite, scuba diving, and intra-abdominal infections or inflammatory disease. A variety of medications, illicit drugs, and chemicals also can result in a chemical picture identical or similar to ischemic colitis (Table 143-4), sometimes probably secondary to vasoconstriction that can affect other parts of the GI tract, liver, and other organ systems (e.g., cocaine, amphetamines, pseudoephedrine), sometimes caused by constipation (e.g., alosetron), sometimes caused by a hypercoagulable effect (e.g., estrogens), and sometimes as a result of a chemical effect (e.g., sodium polystyrene with sorbitol enemas).

CLINICAL MANIFESTATIONS

The clinical presentation of ischemic colitis is acute and, in most patients, includes abdominal pain (mostly left lower quadrant), often with urgency, diarrhea, and passage of bright red blood per rectum.[7] Anorexia, nausea, vomiting, abdominal distention, and passage of maroon material per rectum may also occur. Although the blood loss is not usually enough to require transfusion, some patients may be orthostatic because of loss of blood and fluid. Fever, tachycardia, abdominal tenderness over the affected portion of the colon, and distention may be found on physical examination.

TABLE 143-4 MEDICATIONS AND DRUGS ASSOCIATED WITH ISCHEMIC COLITIS

Digitalis
Vasopressin
Pseudoephedrine
Amphetamines
Cocaine
Ergot
Sumatriptan
Gold
Danazol
Estrogens
Progestins
Alosetron
Psychotropics
Nonsteroidal anti-inflammatory drugs
Various enemas
Tegaserod
Interferon/ribavirin

DIAGNOSIS

Laboratory findings range from normal to nonspecific findings such as leukocytosis and hemoconcentration to those observed in persons with bowel necrosis (see earlier). Evaluation of patients younger than 50 years should include tests for thrombophilic disorders (Chapter 176).

Plain radiographs of the abdomen may reveal "thumbprinting" or may be normal. Similar to small bowel ischemia, CT scanning can be useful to help exclude other disorders, especially in more symptomatic and ill patients; the findings may be consistent with segmental colonic edema and inflammation, with or without adjacent pericolonic inflammatory stranding. These radiographic features are consistent with ischemic colitis in an appropriate clinical setting but are nonspecific and may be seen in patients with other disorders such as acute diverticulitis (Chapter 142), infectious colitis (Chapter 140), and inflammatory bowel disease (Chapter 141). The diagnosis is best made by colonoscopy, which should provide endoscopic and histologic findings consistent with acute ischemic colitis: segmental patchy ulceration, edema, erythema, single stripe sign, and submucosal bluish purple hemorrhagic nodules (Fig. 143-4).

Typically, visceral angiography is not required because most patients with ischemic colitis have self-limited involvement of the left colon or distal transverse colon–splenic flexure with sparing of the rectum, and findings on urgent angiography in these patients are usually normal. However, about 10% of patients with acute ischemic colitis have predominantly right-sided involvement of the cecum, ascending colon, hepatic flexure, and proximal transverse colon. Because the arterial supply to the right colon is through the ileocolic branch of the SMA, there may be concomitant distal ileal ischemia, often owing to low-flow states (especially hemodialysis patients) or embolization. These patients have more pain, less bleeding, and a much worse outcome and are at risk for small bowel necrosis.

Differential Diagnosis

Gastrointestinal infections, such as with *Escherichia coli* O157:H7, *Clostridium difficile, Klebsiella oxytoca*, and cytomegalovirus (CMV), can mimic ischemic colitis clinically and even histologically. Acute-onset inflammatory bowel disease involving the colon can also be difficult to distinguish from ischemic colitis. However, patients with subacute or chronic pain, diarrhea, obstructive symptoms, weight loss, or bleeding may be thought to have complicated diverticular disease, Crohn disease, or malignancy with stricture, and chronic ischemic stricture of the colon may not be correctly diagnosed until after surgery is performed.

Stool culture can exclude infection, especially with *E. coli* O157:H7, *C. difficile,* and parasites. In immunocompromised patients, colonic biopsy can be performed to diagnose CMV infection.

TREATMENT

Rx

Patients with right-sided ischemic colitis require multidetector CTA or visceral angiography not only for diagnosis but also for intra-arterial administration of vasodilators (e.g., papaverine as a 60-mg intravenous bolus followed by an infusion of 30 to 60 mg/hr). Some patients may require urgent surgery.

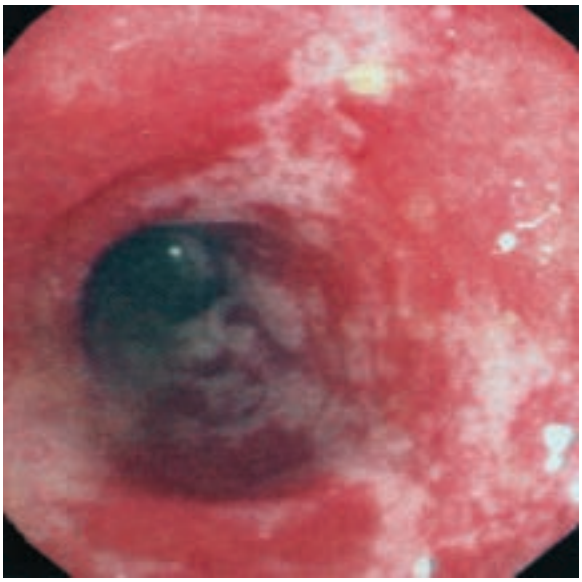

FIGURE 143-4. Endoscopy of the splenic flexure of the colon in a patient with ischemic colitis. Note the shallow, irregular, exudative ulceration with interspersed erythema. (From Emory TS, Carpenter HA, Gostout CJ, et al, eds. *Atlas of Gastrointestinal Endoscopy and Endoscopic Biopsies*. Washington, DC: Armed Forces Institute of Pathology, American Registry of Pathology; 2000.)

The clinical course in patients with right-sided ischemic colitis may be subacute, with a mortality rate as high as 50% or greater.

By contrast, left-sided acute ischemic colitis, which accounts for most cases, tends to resolve within hours to a few days with supportive therapy, including volume replacement, correction of any low-flow state; broad-spectrum antibiotics (similar to those recommended earlier for patients with small bowel ischemia); avoidance of vasoconstrictive medications; and rarely, blood transfusion; surgery is required only in patients with signs and symptoms of transmural necrosis, perforation, or massive bleeding. Occasional patients with acute-onset left-sided ischemic colitis have persistent or recurrent symptoms of pain, diarrhea, bleeding, sepsis, or stricture formation that develop over a period of weeks to months and may require segmental surgical resection.

PROGNOSIS

As many as 10% to 20% of patients may require urgent surgery. Nonocclusive ischemic colitis, acute renal failure, extent of bowel ischemia, serum lactate, and duration of catecholamine therapy predict survival.

VASCULITIS

Many vasculitic syndromes can involve the GI tract. Usually, but not always, other organ systems are also involved.

Large and Medium Vessel Vasculitis

Takayasu arteritis (Chapters 78 and 270) and *giant cell arteritis* (Chapter 271), which affect large to medium-sized muscular arteries, rarely involve the GI tract. Takayasu arteritis has been associated rarely with inflammatory bowel disease.

Medium to Small Vessel Vasculitis

Polyarteritis nodosa (Chapter 270) is characterized by segmental microaneurysms typically involving small- and medium-sized arteries. The small bowel is involved more commonly than the large bowel. Many patients will have abdominal pain, fever, hypertension, and multiple organ involvement. GI bleeding or perforation will develop in some patients. The gallbladder, spleen, pancreas, and liver may also be involved. About one-third of patients with polyarteritis nodosa are infected with hepatitis B virus.

Both *granulomatosis* with *polyangiitis* (Chapter 270) and *Churg-Strauss syndrome* (Chapter 270) affect small- and medium-sized arteries. Although GI involvement is not common in Wegener granulomatosis with granulomatous inflammation, in up to one third of patients with Churg-Strauss syndrome, abdominal pain or GI bleeding may develop as a result of ischemia. Mesenteric venous involvement can also occur with Churg-Strauss syndrome.

Thromboangiitis obliterans (Buerger disease; Chapter 80) involves small- and medium-sized arteries and can cause multiple distal occlusions of the mesenteric arterial circulation. Patients with *Behçet disease* (Chapter 270) often have lymphocytic inflammation of small- and medium-sized arteries, as well as veins. Similar to Crohn disease, the ileocecal region is frequently involved with ulceration. Abdominal pain, diarrhea, GI bleeding, and perforation may occur.

Degos disease, also known as malignant atrophic papulosis, is a rare condition characterized by vasculitis of small- and medium-sized arteries. Multiple organ systems can be affected, and GI perforation owing to mesenteric ischemia represents a major cause of mortality.

Small Vessel Vasculitis

Small vessel involvement with immunoglobulin A immune complex deposition in blood vessel walls is typical in *Henoch-Schönlein purpura* (Chapter 270). These patients usually have palpable purpura, arthritis, nephritis, and abdominal pain, as well as GI bleeding. *Hypersensitivity vasculitis* (Chapter 270), which affects small arterioles, venules, and capillaries, is related to a variety of drugs, infections, and chemicals; on occasion there may be GI involvement. *Cryoglobulinemia* (Chapter 187) with immune complex involvement of small blood vessels can sometimes involve the GI tract. These patients are often infected with hepatitis C virus.

HEMORRHAGIC VASCULAR DISORDERS

Angiodysplasia

DEFINITION AND EPIDEMIOLOGY

Angiodysplasia or vascular ectasia is a thin-walled, dilated, punctate red vascular structure in the mucosa or submucosa of the bowel; it typically involves adjacent venules, capillaries, and arterioles.[8] Angiodysplasia is found in the colon, especially the right colon, in up to 1% of persons and is found also in the stomach and small bowel but rarely in the esophagus. Angiodysplastic lesions may be single or multiple, and they increase in frequency with age. Some data suggest associations with chronic renal failure (Chapter 130), von Willebrand disease (Chapter 173), and aortic stenosis (Chapter 75); whether correction of these associated disorders diminishes future GI hemorrhage (Chapter 135) from angiodysplasia is uncertain.

CLINICAL MANIFESTATIONS

Clinically, these lesions can produce painless bleeding, which may be occult and manifest only by guaiac-positive stools or iron deficiency anemia, or the bleeding may be overt, with hematochezia, maroon stools, melena, and hematemesis.

DIAGNOSIS

Endoscopic procedures most often make the diagnosis of bleeding secondary to angiodysplasia (Fig. 143-5). In some patients, endoscopic procedures may need to be repeated, especially in volume-depleted patients and after the administration of narcotics. Small bowel angiodysplasia, beyond the reach of both a colonoscope from below and an extended-length endoscope from above, may be the cause of major bleeding (Chapter 135) and require video capsule endoscopy (Chapter 134) or small bowel balloon-assisted enteroscopy for diagnosis and treatment.

TREATMENT Rx

Electrocoagulation laser therapy or argon plasma coagulation can be accomplished during endoscopy. When very active bleeding makes urgent colonoscopy technically difficult, visceral angiography can be diagnostic and permit embolization of bleeding lesions or intra-arterial infusion of a vasoconstrictor. Rarely, bowel resection is required.

PROGNOSIS

More than 90% of GI angiodysplasias never bleed. When angiodsyplasias are found incidentally in patients who have no history of past or concurrent bleeding, they typically should not be treated.

Dieulafoy Lesion

Dieulafoy lesion is an unusually large submucosal artery typically found in the proximal portion of the stomach within 6 cm of the gastroesophageal junction. Similar lesions may also occur in the rectum; colon; small bowel; and far less often, the esophagus. Dieulafoy lesion is manifested clinically as sudden, massive bleeding, which may be recurrent.

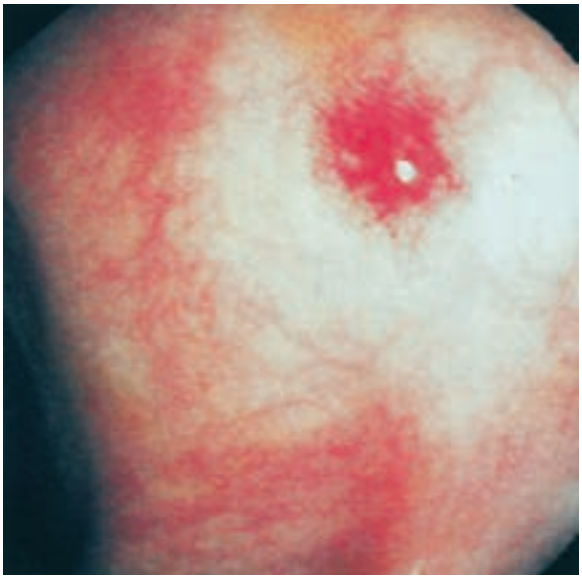

FIGURE 143-5. Endoscopy of the sigmoid colon in a patient with angiectasia. The lesion is discrete and contains a tight, radiating cluster of ectatic mucosal vessels. (From Emory TS, Carpenter HA, Gostout CJ, et al, eds. *Atlas of Gastrointestinal Endoscopy and Endoscopic Biopsies*. Washington, DC: Armed Forces Institute of Pathology, American Registry of Pathology; 2000.)

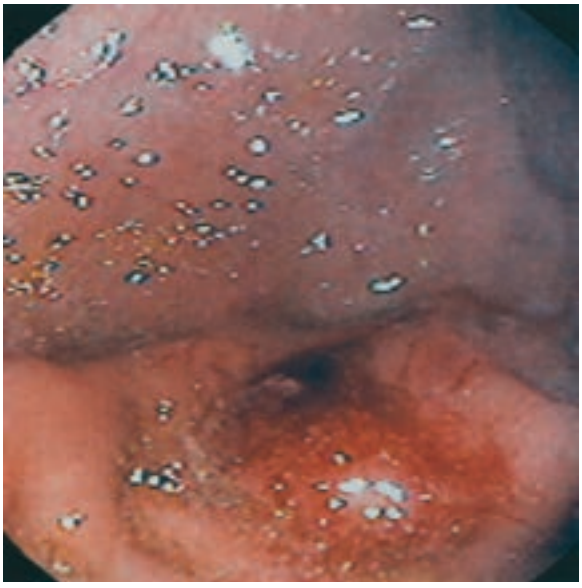

FIGURE 143-6. Endoscopy of the stomach in a patient with Dieulafoy lesion. Note the visible vessel manifested as a pale protuberance surrounded by a clot with adjacent normal-appearing mucosa. (From Emory TS, Carpenter HA, Gostout CJ, et al, eds. *Atlas of Gastrointestinal Endoscopy and Endoscopic Biopsies*. Washington, DC: Armed Forces Institute of Pathology, American Registry of Pathology; 2000.)

DIAGNOSIS AND TREATMENT Rx

Urgent endoscopy is required to identify what is usually a very small vascular protuberance (Fig. 143-6) but can rapidly become inapparent when the acute bleeding stops. Ulceration is not seen, and repeat endoscopic procedures during active bleeding may be required to make the diagnosis. Sometimes the diagnosis requires angiography during a bleeding episode. Endoscopic injection and electrocoagulation therapy are generally effective, but endoscopic band therapy and hemoclips may also be used, and surgery is sometimes required.

PROGNOSIS

Endoscopic therapy is successful long term in nearly 90% of patients.

Other Ectasias

Telangiectases are similar to angiodysplasias but occur in all the layers of the bowel wall, are usually congenital and often occur in other organ systems. *Hereditary hemorrhagic telangiectasia* (Osler-Weber-Rendu disease; Chapter 173) is an autosomal dominant disorder with telangiectases involving the lips; mucous membranes, especially in the mouth and nose; GI tract, especially the stomach and small bowel; liver; lung; retina; and central nervous system. Patients with *Turner syndrome* (Chapter 235) or *scleroderma* (Chapter 267), and the *CREST syndrome* (Chapter 267) (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, telangiectasia) may also have GI tract telangiectases.

Vascular ectasias involving venules and capillaries can also be seen in the small bowel *(congestive enteropathy)*; in the colon *(congestive colopathy)*; and more commonly, in the stomach *(congestive gastropathy)* in patients with portal hypertension (Chapter 153). In contrast to angiodysplasias, these lesions tend to be more diffuse, appear as multiple, fine punctate red spots or as a mosaic pattern similar to the gastritis of *Helicobacter pylori* and are more often found in the proximal than the distal part of the stomach. Therapies that decrease portal hypertension can reduce or eliminate these lesions and bleeding from them.

Gastric antral vascular ectasia (GAVE), or watermelon stomach, also involves venules and capillaries with thrombosis as well as ectasia. Erythematous streaks similar to the stripes on a watermelon are typically seen in the antrum radiating toward the pylorus. The gastric cardia may be involved as well. Patients usually have occult bleeding and less often melena. GAVE is associated with connective tissue diseases (e.g., systemic lupus erythematosus [Chapter 266], mixed connective tissue disease [Chapter 267], scleroderma [Chapter 267]), pernicious anemia (Chapter 164), and portal hypertension (Chapter 153). However, unlike congestive gastropathy, treatment of portal hypertension does not eliminate GAVE or bleeding from it. Argon plasma coagulation is the usual therapy if iron replacement alone is not effective. Antrectomy is rarely needed.

Neoplastic Vascular Lesions

Hemangiomas are uncommon, usually benign vascular tumors that can be found throughout the GI tract, often in the rectum or colon. They may be single or multiple, bluish purple, and sessile or polypoid. In some persons, these lesions are multiple and associated with skin lesions, such as the *blue rubber bleb nevus syndrome* with purple-blue cutaneous hemangiomas or the *Klippel-Trenaunay* syndrome with port-wine-colored cutaneous hemangiomas, hemihypertrophy, and varicose veins. Rare vascular malignant neoplasms of the GI tract include *angiosarcomas* and *hemangioendotheliomas.*

Miscellaneous Vascular Disorders

Aortoenteric fistulas, which most commonly occur after surgery for an aortic aneurysm (Chapter 78), may be related to infection of the graft and can result in torrential GI bleeding. Many of these fistulas communicate with the duodenum. Evaluation of bleeding in persons who have previously undergone abdominal aortic surgery should include urgent extended-length upper endoscopy to document a fistula or diagnose another definitive source of the bleeding. Angiography and radiographic tests (CT, magnetic resonance imaging [MRI]) are helpful only if the findings are abnormal (i.e., there is evidence of a fistula) because of their poor sensitivity for the presence of aortoenteric fistulas. If no clear alternative source for the bleeding can readily be found, explorative surgery is indicated. *Atrioesophageal fistulas* can occur as a consequence of thermal damage within the heart or esophagus, such as after radiofrequency ablation procedures for atrial fibrillation or for Barrett esophagus.

Celiac artery compression syndrome (median arcuate ligament syndrome) is a very rare pseudo-ischemic syndrome. Patients are often young and healthy and have postprandial upper abdominal pain, most likely caused by extrinsic compression of the celiac axis by the median arcuate ligament of the diaphragm. Sitophobia can result in considerable weight loss, and there may be a loud systolic bruit in the epigastric region on physical examination. Visceral angiography supports the diagnosis, but bruits and celiac axis compression may occur without symptoms. Surgical therapy is indicated after other possible causes of the patient's symptoms have been excluded.

Patients with *Ehlers-Danlos syndrome type IV* (Chapter 260) can develop small bowel ischemia and perforation as well as arterial rupture. Similar vascular catastrophes with GI or intraperitoneal hemorrhage can occur in patients with *pseudoxanthoma elasticum type I* (Chapter 260) or with *visceral artery aneurysms* (secondary to atherosclerosis, fibrodysplasia, portal hypertension, pregnancy, pancreatitis, vasculitis, or trauma).

HEPATIC AND SPLENIC VASCULAR DISEASE

Budd-Chiari Syndrome

Budd-Chiari syndrome can occur as a result of any process that interferes with the normal flow of blood out of the liver, including constrictive pericarditis (Chapter 77) and veno-occlusive disease (Chapter 150). Hepatic vein thrombosis, which is the main cause of Budd-Chiari syndrome, may involve one, two, or all three of the major hepatic veins, with or without partial or complete occlusion of the inferior vena cava. Often, Budd-Chiari syndrome is caused by a hypercoagulable state (Chapter 176), such as a chronic myeloproliferative disorder (e.g., polycythemia vera [Chapter 166], essential thrombocythemia [Chapter 166], myeloid metaplasia [Chapter 166]), paroxysmal nocturnal hemoglobinuria, or other hypercoagulable conditions (Chapter 176), such as factor V (Leiden) gene mutation, antiphospholipid antibody syndrome, protein C deficiency, protein S deficiency, or antithrombin III deficiency. A high percentage of these patients harbor *JAK2* V617F mutations. Nearly 50% of patients with Budd-Chiari syndrome have more than one risk factor. Malignancies (direct compression or invasion of hepatic veins, hypercoagulable state), infections (liver abscess), pregnancy, inflammatory disorders (e.g., Behçet syndrome, inflammatory bowel disease, connective tissue disease, sarcoidosis), and membranous obstruction (webs) of the inferior vena cava are also associated with Budd-Chiari syndrome.

CLINICAL MANIFESTATIONS

The syndrome is usually subacute or chronic, occurs over a period of weeks to months, and is characterized by the insidious onset of upper abdominal pain, hepatomegaly, and ascites. Fulminant and acute presentations, including encephalopathy, jaundice, ascites, and liver failure, are rare.

DIAGNOSIS

Liver function testing usually reveals normal or mild to moderate nonspecific elevations in serum aspartate and alanine aminotransferase levels. *JAK2* mutation analysis should be part of the initial evaluation. Doppler ultrasonography of the liver is the initial diagnostic test of choice, but the absence of hepatic venous flow or venous thrombosis (or both) is also readily apparent with contrast-enhanced CT scanning or MRA. Hepatic venography can confirm the diagnosis (Fig. 143-7) and, with imaging of the inferior vena cava, as well as selective venous pressure measurements, can help guide therapy.

TREATMENT

Rx

Therapy includes diagnosis and treatment of underlying conditions, anticoagulation (intravenous heparin; see Table 81-4 in Chapter 81) to prevent the propagation of thrombi, and treatment of ascites (e.g., diuretics; Chapter 153). To decompress the congested liver, most patients require interventional radiologic procedures, such as angioplasty, stenting, or transjugular intrahepatic portosystemic shunts, to restore hepatic venous flow.[9,10] Surgical procedures such as surgical shunts to drain the portal or mesenteric venous system into the inferior vena cava can also decompress the liver. Liver transplantation (Chapter 154) should be considered for patients with fulminant liver failure, cirrhosis, or both. Most patients with Budd-Chiari syndrome require lifelong warfarin anticoagulation (Chapter 38) even after liver transplantation. Selective JAK2 inhibitors (Chapter 166) are being tested in preclinical and clinical studies.

PROGNOSIS

Overall, the 5-year survival rate of patients with the Budd-Chiari syndrome is more than 80%. Indices such as levels of serum albumin and bilirubin, the international normalized ratio, ascites, and encephalopathy, and the Child-Pugh score (see Table 153-2 in Chapter 153) can be useful in determining the prognosis. About 50% of patients will have at least one episode of major bleeding, half of which are related to invasive therapy.

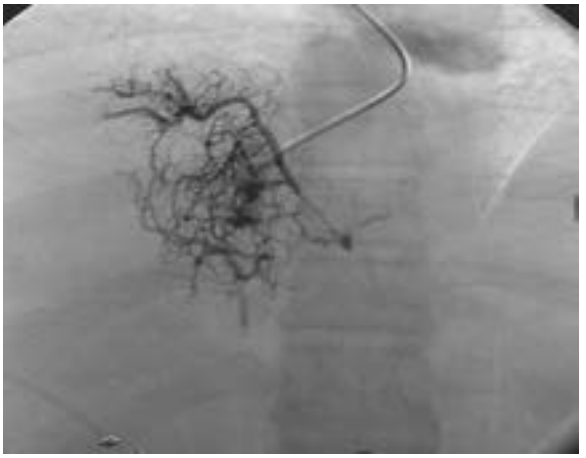

FIGURE 143-7. Budd-Chiari syndrome. A hepatic vein contrast study depicts the "spider web" pattern of venovenous collaterals attempting to bypass a thrombosed hepatic vein. (Courtesy of Patrick Kamath.)

Portal Vein Thrombosis

In adults, cirrhosis, hypercoagulable states, intra-abdominal malignancy, inflammatory disorders (e.g., pancreatitis, Crohn disease), and medical procedures (e.g., splenectomy, cholecystectomy, gastrectomy, liver transplantation, transjugular intrahepatic portosystemic shunt) are most often the cause of acute portal vein thrombosis. Similar to Budd-Chiari syndrome, a substantial number of these patients harbor *JAK2* V617F mutations.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Clinical manifestations include portal hypertension with variceal bleeding and ascites. Abdominal pain and diarrhea may indicate extension of the thrombus into the SMV with intestinal ischemia. The diagnosis of acute portal vein thrombosis is confirmed by Doppler ultrasound, multidetector CTA, or MRA. CT imaging may reveal multiple small liver abscesses.

TREATMENT

Rx

Anticoagulation (see Table 81-4 in Chapter 81) therapy for at least 3 months is recommended for patients with acute portal vein thrombosis and should be continued long term in those persons with permanent thrombotic risk factors not otherwise correctable, as well as in patients with extension of thrombus into the mesenteric veins.[11]

High fever, chills, a tender liver, and sepsis suggest pylephlebitis, which usually requires treatment with parenteral antibiotics such as piperacillin–tazobactam, ticarcillin–clavulanate, carbapenem, or a third-generation cephalosporin plus metronidazole for at least 6 weeks. Blood cultures can help to guide the choice and course of antibiotics, which should be administered intravenously for at least 2 weeks.

Treatment is less clear in cirrhotic patients with acute or chronic portal vein thrombosis. Endoscopy for the diagnosis and treatment of varices (Chapter 134), with or without pharmacologic treatment of the portal hypertension (e.g., β-blockade with propranolol; Chapter 153), is often beneficial, and surgical shunts are rarely necessary. Antibiotics, such as those recommended previously for pylephlebitis, should be administered in patients with any sign of infection.

Chronic portal vein thrombosis is defined as an obstructed portal vein replaced by collateral veins. Doppler ultrasonography, multidetector CTA, or MRA will confirm the diagnosis. Patients may be asymptomatic but often have hypersplenism and portal hypertension (e.g., subclinical encephalopathy, rare ascites). Some patients may develop jaundice owing to biliary cholangiopathy and will require endoscopic placement of biliary stents. Endoscopic screening and treatment of varices (Chapter 134), with or without pharmacologic treatment of portal hypertension (Chapter 153), is recommended. After treatment of the varices, long-term anticoagulation therapy (Chapter 81) should be considered in noncirrhotic patients whose permanent thrombotic risk factors are not otherwise correctable unless there is a contraindication.

PROGNOSIS

Mortality rates after treatment of acute portal vein thrombosis are less than 10%, and the prognosis for patients with chronic portal vein thrombosis over 5 years is similar.

Splenic Vein Thrombosis

Splenic vein thrombosis is usually secondary to malignancy (e.g., pancreatic cancer), pancreatitis, or trauma. In many of these patients, isolated gastric varices develop and are difficult to treat by therapeutic endoscopy. Liver function and portal pressure are normal. Most patients with splenic vein thrombosis have splenomegaly (Chapter 168). Doppler ultrasonography, MRI, and CT assist in making the diagnosis. Patients with symptomatic isolated splenic vein thrombosis (e.g., gastric variceal bleeding, hypersplenism) are best treated by splenectomy.

Hepatic and Splenic Arterial Disease

Hepatic arterial disease may be nonocclusive or occlusive. Nonocclusive disease, termed *ischemic hepatitis,* occurs when arterial blood flow to the liver is insufficient, usually because of cardiogenic hypotension, volume depletion, or sepsis. Typically, serum aminotransferase rises acutely to levels greater than 1000 U/L. With restoration of adequate hepatic arterial blood flow, serum aminotransferase levels eventually fall back to their baseline by about 40% to 60% per day. Hepatic artery thrombosis is extremely rare except in post–liver transplantation (Chapter 154) patients, in whom it may be manifested as mild abnormalities in liver function test results, bile duct injury (e.g., biliary stricture, cholangitis, liver abscess), or liver failure. Doppler ultrasonography and angiography confirm the diagnosis, and these patients often require biliary stents, drainage of abscesses, surgical reconstruction of the hepatic artery, or retransplantation of the liver.

The splenic artery or hepatic artery may be predisposed to the development of aneurysmal dilation, usually secondary to atherosclerosis, trauma, portal hypertension, pancreatitis, pregnancy, infection, or vasculitis. Common clinical manifestations include abdominal pain and intra-abdominal hemorrhage. Hemobilia may occur with hepatic artery aneurysms. Angiography is usually required to make the diagnosis. Symptomatic as well as sizable (variably defined, usually 1 cm or greater for a hepatic aneurysm and 2 cm or greater for a splenic artery aneurysm) aneurysms require surgery. Splenic artery aneurysms discovered during pregnancy are more likely to bleed and should be treated, usually by interventional radiology.

Fistulas from the hepatic artery to the portal vein can occur as a result of trauma, malignancy, or the inherited disorder hereditary hemorrhagic telangiectasia. The resultant portal hypertension may cause abdominal pain, ascites, and GI bleeding, and involvement of the hepatic artery may result in biliary strictures and hepatobiliary infection. Radiographic embolization of these fistulas, surgery, or liver transplantation may be required.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Sise MJ. Acute mesenteric ischemia. *Surg Clin North Am*. 2014;94:165-181.
2. Bobadilla JL. Mesenteric ischemia. *Surg Clin North Am*. 2013;93:925-940.
3. Acosta S, Bjorck M. Modern treatment of acute mesenteric ischaemia. *Br J Surg*. 2014;101: e100-e108.
4. Beaulieu RJ, Arnaoutakis KD, Abularrage CJ, et al. Comparison of open and endovascular treatment of acute mesenteric ischemia. *J Vasc Surg*. 2014;59:159-164.
5. Singal AK, Kamath PS, Tefferi A. Mesenteric venous thrombosis. *Mayo Clinic Proc*. 2013;88: 285-294.
6. Davenport DL, Shivazad A, Endean ED. Short-term outcomes for open revascularization of chronic mesenteric ischemia. *Ann Vasc Surg*. 2012;26:447-453.
7. Tadros M, Majumder S, Birk JW. A review of ischemic colitis: is our clinical recognition and management adequate? *Expert Rev Gastroenterol Hepatol*. 2013;7:605-613.
8. Sami SS, Al-Araji SA, Ragunath K. Review article: gastrointestinal angiodysplasia—pathogenesis, diagnosis and management. *Aliment Pharmacol Ther*. 2014;39:15-34.
9. Seijo S, Plessier A, Hoekstra J, et al. Good long-term outcome of Budd-Chiari syndrome with a step-wise management. *Hepatology*. 2013;57:1962-1968.
10. Tripathi D, Macnicholas R, Kothari C, et al. Good clinical outcomes following transjugular intrahepatic portosystemic stent-shunts in Budd-Chiari syndrome. *Aliment Pharmacol Ther*. 2014;39: 864-872.
11. Plessier A, Rautou PE, Valla DC. Management of hepatic vascular diseases. *J Hepatol*. 2012;56(suppl 1):S25-S38.

144

PANCREATITIS

CHRIS E. FORSMARK

ACUTE PANCREATITIS

DEFINITION

Acute pancreatitis, which is a discrete episode of cellular injury and inflammation in the pancreas, is triggered by the release of activated digestive enzymes into the pancreas and peripancreatic tissues. Acute pancreatitis usually presents with symptoms of abdominal pain, nausea, and vomiting

accompanied by an elevation in serum levels of amylase, lipase, or both and by radiographic evidence of pancreatic inflammation, edema, or necrosis. Although pancreatic morphology and function may recover after such an episode, complete recovery is unlikely if the initial damage is substantial, particularly if the original episode is associated with significant pancreatic necrosis. With repeated episodes, there can be a shift from acute inflammation, necrosis, and apoptosis to a milieu of chronic inflammation, the activation of pancreatic stellate cells, continued tissue destruction, and ultimately the fibrosis characteristic of chronic pancreatitis. About 25% of patients with acute pancreatitis will have recurrence, and about 10% will develop chronic pancreatitis.[1]

EPIDEMIOLOGY

Acute pancreatitis, which is the most common cause of hospitalization for a gastrointestinal condition in the United States,[2] accounts for approximately 275,000 hospitalizations yearly. This rate of hospital admissions has doubled over the past 2 decades. The incidence of acute pancreatitis ranges from 13 to 45 per 100,000 population. The total cost of caring for these patients is substantial, with estimates of $6 billion annually. The incidence of acute pancreatitis is increasing in the United States and in many other countries, perhaps because of heightened clinical suspicion for the diagnosis, the more widespread use of serum-based and radiologic testing, and an increasing incidence of gallstones in the midst of the obesity epidemic.

The risk of acute pancreatitis increases fourfold between ages 25 and 75 years. The risk is two- to threefold higher among the black population in the United States compared with whites. Increased abdominal adiposity but not body mass index increases the risk of acute pancreatitis approximately twofold.

PATHOBIOLOGY

The mechanisms that lead to acute pancreatitis include exposures to potential disease triggers and genetic polymorphisms that predispose to acute pancreatitis. Acute pancreatitis is characterized by premature activation of pancreatic digestive enzymes within the pancreas. In many models of pancreatitis, abnormal calcium signaling and the activation of specific protein kinases lead to the generation of inflammatory mediators, the activation of enzymes within the acinar cell, misdirected exocytosis, and ultimately the cellular injury and death that characterize acute pancreatitis.[3] The activation of trypsinogen to trypsin may be a critical initial step, with trypsin having the capacity to activate other proteases within the gland. These activated enzymes produce cellular injury and death. Necrosis may involve not only the pancreas but also surrounding fat and structures, leading to fluid extravasation into the retroperitoneal spaces ("third-space" losses). Although some degree of microscopic necrosis may be present in most cases of acute pancreatitis, more substantial necrosis (visible on an enhanced contrast computed tomography [CT] scan) is termed *acute necrotizing pancreatitis* and is distinguished from the milder *acute interstitial pancreatitis*, in which necrosis is not visible on a CT scan.

In addition to local damage within and around the pancreas, acute pancreatitis may be associated with distant organ system failure. The release of inflammatory cytokines and activated digestive enzymes into the systemic circulation can produce a systemic inflammatory response syndrome (SIRS; Chapters 106 and 108) and associated organ system failure. The most common manifestations of this process in severe acute pancreatitis include hypotension, renal failure, and respiratory failure. Gallstones and alcohol account for about 70% to 80% of all cases of acute pancreatitis, and the cause remains unknown in about 10% of cases (Table 144-1).

TABLE 144-1 CAUSES OF ACUTE PANCREATITIS

ETIOLOGY	EXAMPLES	COMMENTS
Gallstones	Gallstones Microlithiasis	Best detected by EUS
Drugs and toxins	Ethyl and methyl alcohol Tobacco Azathioprine, 6-mercaptopurine, pentamidine, didanosine, sulfonamides, thiazides, aminosalicylates, valproic acid, and others Scorpion venom Insecticides	 Usually idiosyncratic Caused by hyperstimulation of pancreatic secretion
Metabolic	Hypertriglyceridemia Hypercalcemia	Usually requires triglyceride level >1000 mg/dL
Trauma	Post-ERCP Blunt or penetrating trauma Postoperative	Risk varies with indication and may be reduced by rectal NSAIDs and pancreatic duct stents
Obstruction of the pancreatic duct	Benign pancreatic duct stricture Benign ampullary stricture (e.g., celiac disease, diverticulum) Ampullary adenoma or adenocarcinoma Pancreatic ductal adenocarcinoma Intraductal papillary mucinous neoplasm Pancreas divisum Sphincter of Oddi dysfunction	 Controversial
Infections	Cytomegalovirus, mumps, rubella, Coxsackie B *Candida*, histoplasmosis Ascaris	
Genetics	*PRSS1* mutations *CFTR* mutation *SPINK1* mutation Others (chymotrypsin C, calcium sensing receptor, claudin-2, others)	Mutation sufficient to cause disease Mutations or polymorphisms predispose to pancreatitis
Autoimmune pancreatitis	Type 1 Type 2	Elevations in serum levels of IgG4 may be seen in Type 1
Idiopathic pancreatitis		

ERCP = endoscopic retrograde cholangiopancreatography; EUS = endoscopic ultrasonography; NSAID = nonsteroidal anti-inflammatory drug.

Gallstones and Obstruction

Passage of a gallstone through the ampulla of Vater, with transient obstruction of the pancreatic duct, is the initiating event for gallstone pancreatitis. Only about 5% of all patients with gallstones develop pancreatitis, and patients with smaller gallstones (≤5 mm), which can pass the cystic duct and reach the ampulla, are at highest risk. Microlithiasis describes very tiny gallstones that are not easily visible on standard transabdominal ultrasonography but may cause gallstone pancreatitis.

In addition to gallstones and microlithiasis, pancreatic duct obstruction owing to pancreatic ductal adenocarcinoma (Chapter 194), ampullary adenoma or carcinoma, or intraductal papillary mucinous neoplasm can cause acute pancreatitis. Benign strictures of the ampulla may cause acute pancreatitis, owing to duodenal diseases such as celiac disease and periampullary diverticula. Sphincter of Oddi dysfunction, defined by high pressures of the pancreatic sphincter, and pancreas divisum, in which the larger dorsal pancreas drains through the smaller minor papilla, are controversial causes of acute pancreatitis because patients with them often have coexistent underlying genetic mutations that contribute to this disease.

Alcohol

Long-standing use of substantial alcohol (Chapter 33), usually more than 5 years of intake averaging more than 5 to 8 drinks daily, is required before pancreatitis develops. Even then, the absolute risk of pancreatitis is only 2% to 5%, thereby emphasizing important cofactors such as a high-fat diet, genetic variability, and smoking. Interestingly, binge drinking does not appear to be a risk factor for pancreatitis, and many patients develop their first episode of pancreatitis several days after stopping drinking. The mechanism by which alcohol causes pancreatic injury is uncertain but likely involves a mixture of direct toxicity, oxidative stress, and alterations in pancreatic enzyme secretion.

Drugs, Toxins, and Metabolic Factors

Serum triglyceride levels greater than 500 mg/dL and usually greater than 1000 mg/dL can cause acute pancreatitis (Chapter 206). The mechanism is

not known. Pancreatitis also can be precipitated by the administration of estrogens, which can exacerbate underlying hypertriglycideremia. Hypercalcemia is an exceedingly rare cause of acute pancreatitis. Drug-induced pancreatitis is rare and is generally an idiosyncratic event. Implicated drugs include 6-mercaptopurine and azathioprine (up to a 4% attack rate), didanosine, pentamidine, valproic acid, furosemide, sulfonamides, and aminosalicylates. Toxins that may cause acute pancreatitis include methyl alcohol (Chapter 110), organophosphate insecticides, and venom from certain scorpions (Chapter 359).

Trauma

Iatrogenic trauma to the pancreas and pancreatic duct during performance of an endoscopic retrograde cholangiopancreatography (ERCP; Chapter 134) is a common cause of pancreatitis. The risk of acute pancreatitis ranges from less than 5% for patients with simple common bile duct stones or malignancy to as high as 20% for patients with suspected sphincter of Oddi dysfunction.[4] Penetrating trauma and blunt trauma, ranging from a contusion to the gland to a severe crush injury and even transection of the gland, can also cause acute pancreatitis. Ischemic injury to the gland can occur after surgical procedures, especially in patients who undergo cardiopulmonary bypass.

Infections

Ascaris lumbricoides (Chapter 357) may cause pancreatitis by obstructing the pancreatic duct as the worms migrate through the ampulla. A number of viruses may infect the pancreatic acinar cells directly, including cytomegalovirus (Chapter 376), Coxsackie B virus (Chapter 379), Echovirus (Chapter 379), and mumps virus (Chapter 369). Fungal infections of the pancreas are exceedingly rare but may be seen in the setting of immunosuppression.

Genetic and Autoimmune Causes

Mutations in several genes predispose to the development of acute and chronic pancreatitis.[5] Most patients with these mutations will not develop pancreatitis, but those who do often develop relapsing acute pancreatitis and ultimately chronic pancreatitis.

Gain-of-function mutations in the cationic trypsinogen gene (*PRSS1*) predispose to hereditary pancreatitis with such a high penetrance that nearly all affected individuals will ultimately develop chronic pancreatitis (see later) and have a more than 35-fold lifetime risk of developing pancreatic ductal adenocarcinoma (Chapter 194) by age 70 years. Mutations in the cystic fibrosis conductance regulator (*CFTR*; Chapter 89), serine protease inhibitor Kazal type 1 (*SPINK1*), chymotrypsin C (*CTC*), calcium-sensing receptor gene, and claudin-2 genes predispose to both relapsing acute and chronic pancreatitis. With the exception of *PRSS1*, these mutations are best viewed as cofactors that interact with other risk factors and disease modifiers to cause pancreatitis.

Two forms of autoimmune pancreatitis, classically presenting as chronic pancreatitis (see later), have been identified.[6] Type 1 is a systemic disease that affects the salivary glands, retroperitoneum, biliary ducts, kidneys, and other organs and rarely presents as acute pancreatitis. Type 2 only affects the pancreas and may occasionally present as unexplained acute pancreatitis.

CLINICAL MANIFESTATIONS

Abdominal pain, nausea, and vomiting are the hallmark symptoms of acute pancreatitis. The abdominal pain is usually in the epigastric region and often radiates to the back. The pain is steady, reaches its maximum intensity over 30 to 60 minutes, and persists for days. These characteristic symptoms may be masked in patients who present with delirium, multiple organ system failure, or coma.

The physical examination usually reveals tachycardia, and more severe cases often present with or develop hypotension, tachypnea, and fever. Confusion, delirium, and even coma may occur. The abdomen is often distended with diminished bowel sounds. Tenderness to palpation of the abdomen, which may be epigastric or more diffuse, is typical, and rebound and guarding are observed in more severe cases. Dullness to percussion in the lower lung fields may be noted owing to a pleural effusion. Rare physical findings include ecchymoses of the flank (Grey Turner sign) or umbilicus (Cullen sign), which occur when fluid and blood track into these spaces from the retroperitoneum. Jaundice may be present if there is biliary obstruction by a stone.

The presence of dyspnea, tachypnea, oxygen desaturation, hypotension, or tachycardia portends a worse prognosis. Severe acute pancreatitis is defined by the presence of organ system failure (usually cardiovascular, renal, or pulmonary) or by the presence of pancreatic complications such as pancreatic and peripancreatic necrosis (Table 144-2).[7]

TABLE 144-2 COMPLICATIONS OF ACUTE PANCREATITIS

COMPLICATION	EXAMPLES
Systemic complications	Hypotension and shock Adult respiratory distress syndrome Acute renal failure Disseminated intravascular coagulation Hypocalcemia Hypertriglyceridemia Hyperglycemia Encephalopathy and coma
Gastrointestinal bleeding	Stress ulceration Pseudoaneurysm
Local complications	Acute peripancreatic fluid collection Pseudocyst Pancreatic necrosis (infected or sterile) Acute necrotic collection Walled-off pancreatic necrosis Duodenal or biliary obstruction

DIAGNOSIS

The diagnosis of acute pancreatitis requires the presence of two of three primary features: abdominal pain, elevations in serum amylase or lipase levels, and imaging studies consistent with acute pancreatitis. Accurate diagnosis also requires that other conditions that can mimic acute pancreatitis, such as intestinal infarction or small bowel obstruction, be excluded. It is equally important to define the most likely cause and the severity of pancreatitis because both strongly influence management.

Laboratory Tests

Amylase and Lipase

Nearly all patients with acute pancreatitis have an elevation in serum levels of amylase or lipase within a few hours after the onset of symptoms. Elevation more than three times the upper limit of normal is the recommended cutoff for diagnosing acute pancreatitis. Lipase remains elevated longer than amylase. Amylase and lipase levels may be normal in rare patients with acute pancreatitis, particularly if the measurement is delayed for several days after the onset of symptoms. In addition, marked hypertriglyceridemia can interfere with the accurate measurement of amylase and lipase. Both enzymes are cleared by the kidney, and renal failure can raise the level of these enzymes up to five times the upper limit of normal in the absence of pancreatitis. Both amylase and lipase can be elevated in a variety of other conditions, some of which may mimic acute pancreatitis. These include intestinal ischemia and infarction (Chapter 143), bowel obstruction (Chapter 142), cholecystitis (Chapter 155), and choledocholithiasis (Chapter 155). In addition, amylase may be elevated from ectopic pregnancy, acute salpingitis, and a variety of extraabdominal conditions such as parotitis (Chapter 369), lung cancer (Chapter 191), head trauma (Chapter 399), and others. In some patients, only amylase or lipase levels may be elevated. Given its improved specificity, equal cost, and equal sensitivity, lipase is preferred over amylase as a single diagnostic test.[8] Frequent serial measurements of amylase or lipase levels in patients with acute pancreatitis are not helpful in clinical decision making.

Other Laboratory Tests

In addition to amylase and lipase levels, all patients should have blood testing for renal function, liver chemistries, electrolyte concentrations, and levels of calcium and triglycerides. In severe pancreatitis, leukocytosis, hemoconcentration, and azotemia may be seen. Hyperglycemia, hypocalcemia, and mild hypertriglyceridemia can develop in more severe cases. Liver chemistries may be elevated in patients with gallstone pancreatitis. Elevations in alanine aminotransferase levels more than three times the upper limit of normal are most suggestive of gallstones as the cause of pancreatitis, although any significant elevation in liver chemistries should raise the possibility of gallstones (Chapter 155).

Imaging Studies

Imaging studies are used not only in establishing the diagnosis but also in determining the etiology and prognosis. In most patients, ultrasonography,

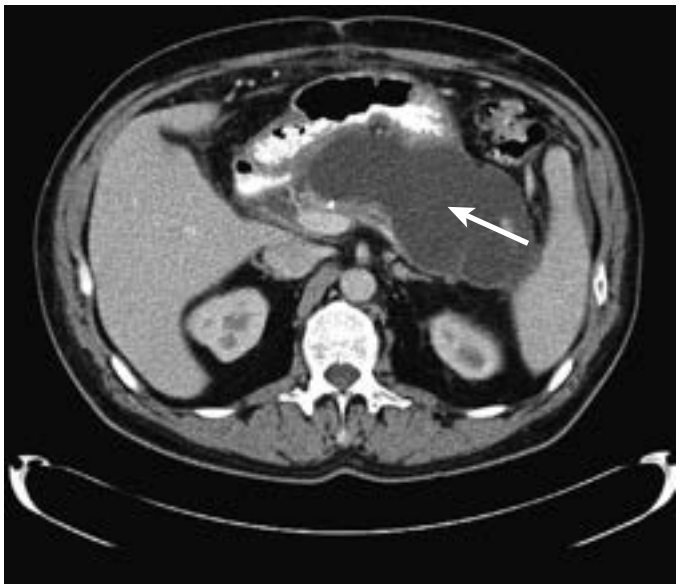

FIGURE 144-1. A computed tomography scan demonstrating a large area of pancreas that does not enhance with intravenous contrast *(arrow)*, consistent with pancreatic necrosis.

CT, or magnetic resonance imaging (MRI) are used in a complementary fashion.[9]

Abdominal ultrasonography can confirm the presence of acute pancreatitis by documenting pancreatic enlargement, edema, or associated peripancreatic fluid collections. Visualization of the pancreas may be inadequate owing to body habitus or overlying intestinal gas. Ultrasonography is accurate in identifying a ductal gallstone as the definitive cause of acute pancreatitis. Alternatively, gallstones in the gallbladder or a dilated common bile duct strongly suggests gallstones as the cause of acute pancreatitis.

Computed tomography is more accurate than ultrasonography for confirming the diagnosis of acute pancreatitis and for documenting the presence of pancreatic necrosis and peripancreatic fluid collections. CT is also particularly helpful in excluding some of the intraabdominal conditions that can mimic acute pancreatitis. However, CT is less accurate than ultrasonography in identifying gallstones. On contrast-enhanced CT, the pancreatic parenchyma that opacifies with intravenous contrast is considered still viable, but the parenchyma that does not opacify is necrotic (Fig. 144-1). The amount of pancreatic necrosis has some prognostic importance, but the degree of necrosis cannot be accurately identified on CT until 3 or more days after the onset of the disease. CT scans are not routinely required in patients with acute pancreatitis but should be performed in patients with a first attack, with severe disease, with disease that is slow to improve, or when the diagnosis is not clear.[10]

Magnetic resonance imaging is equivalent to CT in its ability to document the presence of acute pancreatitis, identify the presence of necrosis, and document nonpancreatic diseases that could mimic acute pancreatitis. In addition, *magnetic resonance cholangiopancreatography (MRCP)* is much better than CT in identifying the presence of gallstones and in assessing abnormalities of the pancreatic duct, such as pancreas divisum or a disrupted pancreatic duct. MRI is more difficult than CT to perform in critically ill patients.

Endoscopic procedures, including ERCP and endoscopic ultrasonography, are important in both diagnosis and therapy of acute pancreatitis. Endoscopic ultrasonography, which is primarily used to establish the cause when the initial evaluation is unrevealing, is particularly accurate in identifying underlying malignancy, premalignant lesions such as ampullary adenoma, and small gallstones or microlithiasis. ERCP is never used as a diagnostic test but may be used to evaluate rare causes of pancreatitis, such as pancreas divisum or sphincter of Oddi dysfunction, in patients with unexplained relapsing pancreatitis.

Determining Etiology

To identify the cause of acute pancreatitis, the history should focus on alcohol and tobacco use, previous biliary colic, drug history, family history, and recent trauma. Alcohol use may need to be corroborated with family members. All patients should undergo transabdominal ultrasonography, with CT or MR considered for patients who have a first attack or a severe attack, who fail to rapidly improve, or who do not have a clear diagnosis. Gallstones should be suspected if stones are seen on ultrasonography, CT, or MRI or if liver chemistries are abnormal, particularly if liver chemistries improve or normalize over a few days. If these initial studies are unrevealing, endoscopic ultrasonography is usually performed to assess for small gallstones, microlithiasis, or underlying malignancy, particularly in patients older than age 40 years. More specialized investigations such as ERCP, sphincter of Oddi manometry, or genetic testing are usually reserved for patients seen in referral centers after multiple attacks of pancreatitis.

On initial evaluation, about 25% of patients do not have an identified cause, but surreptitious alcohol use and microlithiasis are probably the most common underlying causes in these patients. After a detailed evaluation, approximately 10% of patients are ultimately diagnosed with idiopathic pancreatitis.

Determining and Predicting Severity

Severe pancreatitis is defined as organ system failure that persists for more than 48 hours or by local pancreatic and peripancreatic complications such as necrosis, acute fluid collections, or pseudocysts. Moderately severe pancreatitis is characterized by transient organ failure for less than 48 hours, by a local complication, or by a systemic complication owing to worsening of an underlying comorbid disease. Mild acute pancreatitis implies the absence of these features.

Organ failure can be single or multiple, early or late in onset, and progressive and persistent or transient. Patients may exhibit altered mental status, hypoxia, tachypnea, massive third space fluid loss, and intravascular volume depletion. In severe acute pancreatitis, renal failure, pulmonary failure, and circulatory failure most commonly occur as part of the SIRS response. Multiple organ system failure, particularly if it persists beyond 48 hours after admission, is associated with prolonged hospitalization, intensive care unit (ICU) admission, need for surgery, and death.

Local pancreatic and peripancreatic complications help define the severity of acute pancreatitis. The degree of pancreatic necrosis, which is defined on contrast-enhanced CT as areas of pancreas that do not enhance with intravenous contrast infusion, correlates with a worse outcome, particularly if infection develops in the devitalized necrotic tissue. Fluid collections may also accumulate around the pancreas in various retroperitoneal and peritoneal spaces. Much of this inflammatory fluid usually resolves, but some may form into a more circumscribed *pseudocyst* over several weeks. However, some fluid collections seen on contrast-enhanced CT may initially be termed *pseudocysts* when in reality they contain solid necrotic material as well as fluid and actually represent *walled-off pancreatic necrosis* that will require a different therapeutic approach than simple pseudocysts.

TREATMENT

General Supportive Care

The majority of patients will recover within several days, but it is usually not possible to identify these patients accurately at the time of admission. Patients initially should not be given any oral food or fluids. In patients with more severe pancreatitis, admission to an ICU is appropriate.[11] Pain control usually requires parenteral narcotics (e.g., hydromorphone 1-2 mg every 4-6 hours initially or via patient controlled analgesia). Antiemetic agents (Table 132-5) are often required. Early and aggressive hydration (e.g., ≥250 cc/hr or even more) in the first 12 to 24 hours may be necessary to normalize the blood urea nitrogen (BUN), hematocrit, and vital signs and to generate adequate urine output.[12] Lactated Ringer solution may be preferred over normal saline. Care must be taken to ensure patients receive sufficient volume but not enough to cause fluid overload or the development of an abdominal compartment syndrome.

Patients can begin to be fed, beginning with a low-fat solid diet, when bowel sounds have returned and nausea has resolved, without necessarily waiting until all abdominal pain has resolved. Early nasoenteric feeding is no better than an oral diet started 72 hours after admission, with enteral feeding reserved for those who cannot tolerate oral feeding. Enteral nutrition with an elemental or semi-elemental formula is associated with fewer complications and less cost compared with total parenteral nutrition.

Treatment of Complications

Most patients who develop acute gallstone pancreatitis have already passed the offending gallstone into the duodenum, but those with a

persistent or multiple stones are at higher risk of developing cholangitis and possibly more severe pancreatitis. Early ERCP is recommended in patients with gallstone pancreatitis and concomitant cholangitis (fever, jaundice, right upper quadrant pain) and in patients who have strong evidence of a persistent bile duct stone at 48 hours after admission based on a visible persistent stone on an imaging study, jaundice, a persistently dilated bile duct, or worsening liver chemistries. By comparison, early ERCP is not recommended for patients with severe pancreatitis, as evidenced by early and progressive organ system failure, but without cholangitis or suspicion of a persistent bile duct stone. When unsure, endoscopic ultrasonography or MRCP can help identify persistent bile duct stones before consideration of ERCP.

The systemic complications that develop in patients with severe acute pancreatitis are similar to those commonly encountered in other ICU patients, as well as specific metabolic issues that occur in the setting of severe pancreatitis. Hyperglycemia, which develops particularly if parenteral nutrition is used, contributes to higher rates of infections. Hypocalcemia is common in severe pancreatitis, but ionized calcium levels are usually normal and treatment is not needed in the absence of signs of hypocalcemia (Chvostek's sign or Trousseau sign; Chapter 245). Hypertriglyceridemia is usually mild, but even levels greater than 1000 mg/dL usually drop promptly when the patients do not eat. However, occasional patients with sustained severe hypertriglyceridemia may require plasmapheresis.

Acute peripancreatic fluid collections are common in acute pancreatitis, and most fluid collections will resolve spontaneously. Some, however, will mature into an encapsulated, fluid-filled pseudocyst outside of the confines of the pancreas. A pseudocyst also does not require therapy unless it causes abdominal pain or obstruction of a hollow viscus or it is associated with infection or bleeding; in these situations, endoscopic therapy is preferred. Arterial bleeding from a pseudoaneurysm caused by a pseudocyst may be massive and require an emergent CT scan for diagnosis followed by embolization.

In addition, patients with necrotizing pancreatitis may develop infected pancreatic necrosis. Infection of preexisting necrosis typically occurs 2 to 3 weeks into the illness and is heralded by fever, leukocytosis, and worsening abdominal pain. The responsible organisms are usually gram-negative rods and other intestinal flora, but *Staphylococcus aureus* is an important agent as well. If infected necrosis is suspected, a contrast-enhanced CT scan should be obtained to identify the extent of necrosis and assess for indirect evidence of infected necrosis (i.e., gas in the necrotic collection). A CT-directed fine-needle aspiration of the necrotic area for culture and Gram stain can allow antibiotic therapy to be tailored; otherwise, broad-spectrum empiric antibiotics should counter possible infective agents (Table 108-2). Prophylactic antibiotics to prevent infection in patients with preexisting sterile pancreatic necrosis is not recommended, although many patients with severe or necrotizing pancreatitis may ultimately receive antibiotics for treatment of various hospital-acquired infections. Ideally, conservative therapy is continued for at least 4 weeks to allow the infected necrotic material to demarcate, begin to liquefy, and become encapsulated so it can be more easily drained. Percutaneous, endoscopic, or minimally invasive surgical draining procedures are as effective and safer than early open surgical debridement, which is reserved for very rare patients with progressive clinical deterioration.

Any hospital-acquired infection (Chapter 282) dramatically worsens prognosis. Common infections include urinary tract infections (Chapter 284), pulmonary infections (Chapter 97), line infections, and *Clostridium difficile* (Chapter 296).

Prevention

The use of a rectal nonsteroidal anti-inflammatory drug suppository (e.g., indomethacin 100 mg or diclofenac 100 mg) placed either just before or just after ERCP reduces the risk of post-ERCP pancreatitis by about 50%. Placement of a temporary pancreatic duct stent provides equivalent protection. By comparison, preventing recurrent acute pancreatitis is more challenging. Abstinence from alcohol (Chapter 33) and tobacco (Chapter 32), which can be achieved in many patients, can reduce recurrent attacks and should be strongly encouraged. Cholecystectomy (Chapter 155) prevents subsequent attacks of gallstone pancreatitis and should be undertaken within a few weeks of discharge, at the latest. In patients who are not surgical candidates, endoscopic sphincterotomy provides reasonable protection from subsequent attacks. Control of serum lipids (Chapter 206) prevents subsequent attacks of hyperlipidemic pancreatitis. Therapy of lesions that obstruct the pancreatic duct such as strictures, ampullary adenomas, and possibly sphincter of Oddi dysfunction and pancreas divisum may also prevent recurrences.

PROGNOSIS

The case-fatality rate for acute pancreatitis has decreased over time and now averages approximately 1% to 2%. More than 80% of all patients with acute pancreatitis recover promptly and are discharged within a few days. In patients with severe acute pancreatitis, however, the mortality rate is between 10% and 20%. The mortality rate may even approach 30% in patients with more severe and numerous comorbid conditions and in patients who develop pancreatic necrosis, particularly infected necrosis, or organ system failure.

A number of scoring systems and other methods have been developed in an attempt to help guide clinicians predict prognosis, but none has been documented to be superior to experienced clinical judgment. The Ranson criteria, which are of historical interest only, have been replaced by APACHE (Acute Physiology and Chronic Health Evaluation) II and by more simplified systems using multiple-factor scoring. Practice guidelines suggest a cutoff of greater than 8 APACHE II points as the definition of severe disease, but this cutoff has a high false positive rate. An elevated BUN or hematocrit that does not return to normal with fluid therapy is associated with increased mortality rates. A C-reactive protein level greater than 150 mg/L at 48 hours is as accurate as many multifactorial scoring systems at predicting poor outcome. The BISAP score (BUN >25 mg/dL, impaired mental status, SIRS, age >60 years, and pleural effusion) has a possible score of 0 to 5, depending on the number of criteria present. Mortality ranges from less than 1% for a BISAP score of 0 or 1 up to 27% for a BISAP score of 5. For patients with alcoholic pancreatitis, the risk of progression to chronic pancreatitis is about 14% in patients who stop drinking and smoking after the first episode of acute pancreatitis but greater than 40% in those who do not change these behaviors.

CHRONIC PANCREATITIS

DEFINITION

Chronic pancreatitis, which is a syndrome with multiple predisposing risk factors, culminates in a final common pathway of irreversible and permanent pancreatic damage characterized by chronic inflammation, destruction of normal cellular (acinar) structures, and fibrosis. Chronic pancreatitis usually evolves after episodes of acute pancreatitis, some of which may have been subclinical, but the transition between acute and chronic pancreatitis may be difficult to identify.

EPIDEMIOLOGY

The prevalence of symptomatic chronic pancreatitis in Western countries is about 50 per 100,000 population, with an estimated incidence of five to 12 cases per 100,000. In the United States, chronic pancreatitis accounts for about 125,000 outpatient visits and 25,000 hospitalizations yearly. Interestingly, the prevalence of histologic evidence of chronic pancreatitis in autopsy studies approaches 5%. Many people apparently develop chronic damage to the pancreas as a consequence of normal aging, other diseases, or exposure to toxins (e.g., social consumption of alcohol) but do not develop any symptoms or signs of chronic pancreatitis during life.

PATHOBIOLOGY

Multiple episodes of acute inflammation, whether clinical or subclinical, eventually change the inflammatory milieu of the pancreas, with a shift to chronic inflammation, cellular loss, and the activation of pancreatic stellate cells with production of fibrosis. This process becomes self-sustaining and produces the characteristic histologic features in which a chronic fibrosis gradually replaces the acute inflammation.

The pathophysiology of pain, the most common symptom of chronic pancreatitis, is complex, involving both local pancreatic nociception as well as central nervous system responses. Chronic pancreatitis associated pain produces visceral, spinal cord, and central hyperalgesia, and the pain may become self-perpetuating even if therapy on the pancreas is successful.

Alcohol and Tobacco

Alcohol causes about 40% of all cases of chronic pancreatitis in the United States and other developed countries.[13] As with acute pancreatitis, which clinically or occasionally subclinically will precede chronic pancreatitis, substantial and prolonged ingestion of alcohol is usually required, on the order of 5 to 8 drinks daily over more than 5 years. The risk of chronic pancreatitis is only 2% to 5% in patients who consume this much alcohol, pointing to important cofactors such as host genetics and cigarette smoking. There is also evidence that tobacco alone can cause chronic pancreatitis, and smoking alone may be responsible for up to 25% of cases. The combination of alcohol and tobacco is synergistic in causing chronic pancreatitis.

Genetics

Hereditary pancreatitis is an autosomal dominant disease characterized by early onset of acute and chronic pancreatitis, the development of exocrine and endocrine pancreatic insufficiency, and a very high risk of pancreatic

ductal adenocarcinoma (Chapter 194). Mutations in the trypsinogen (*PRSS1*) gene appear to cause a gain in function in which the mutant trypsinogen, once activated to trypsin, is difficult to inactivate. This trypsin, if present in an amount that overwhelms normal protective mechanisms, can activate other pancreatic enzymes and lead to pancreatic damage and eventually to chronic pancreatitis. One of the protective mechanisms is a trypsin inhibitor called SPINK1. Loss of function mutations in *SPINK1* mutations may predispose to chronic pancreatitis, but unlike PRSS1 mutations, are not sufficient alone to cause chronic pancreatitis. Major mutations in the cystic fibrosis conductance regulator (CFTR) lead to cystic fibrosis (Chapter 89), which may be associated with chronic pancreatitis and pancreatic atrophy. Some mutations in CFTR predispose to chronic pancreatitis without causing the sinopulmonary features of cystic fibrosis. Combined mutations of SPINK1 and CFTR may place patients at particularly high risk for chronic pancreatitis. Other mutations and polymorphisms associated with chronic pancreatitis include chymotrypsin C and the calcium-sensing receptor gene. Polymorphisms of claudin 2, an X-linked gene, work synergistically with alcohol and may partially explain the increased risk of alcoholic chronic pancreatitis in men.

Other Causes

Autoimmune pancreatitis most often presents as a mass-like lesion with obstructive jaundice, mimicking cancer. It may also present as chronic pancreatitis and rarely as acute pancreatitis. Type 1 autoimmune pancreatitis, which usually occurs in the fifth or sixth decade of life, is characterized by focal or diffuse swelling of the pancreas, elevations in serum IgG4, and involvement of other organs. Biliary strictures, salivary gland inflammation, retroperitoneal fibrosis, and renal lesions are commonly seen. Histology shows infiltration of these organs by chronic inflammatory cells and especially by plasma cells bearing IgG4 on their surfaces. The target of the autoimmune process is not known. Type 2 autoimmune pancreatitis is limited to the pancreas and occurs in a broader age group, including children.

Tropical pancreatitis is seen primarily in southern India. Characteristic features include childhood onset, exocrine insufficiency, diffuse pancreatic calcifications, and inevitable diabetes. There is a strong genetic component (SPINK1 and others), but cofactors such as malnutrition and dietary toxins have been suggested. In southern India, this disease is becoming rarer and is being replaced by alcohol and tobacco as the most common cause of chronic pancreatitis.

Recurrent or severe acute pancreatitis, particularly a severe acute attack that causes substantial pancreatic necrosis, can destroy enough of the gland to produce exocrine and endocrine insufficiency. In addition, diseases that cause repeated attacks of pancreatitis can lead to chronic pancreatitis. One example is *hypertriglyceridemia*, which causes acute pancreatitis but commonly leads to chronic pancreatitis.

CLINICAL MANIFESTATIONS

The most common symptom of chronic pancreatitis is pain. The pain may be episodic or constant and is generally felt in the epigastrium with radiation to the back. If pain is episodic, the patient may be labeled as having acute pancreatitis or an acute flare of chronic pancreatitis. When pain is severe, nausea and vomiting may occur. Pain may worsen, improve, or remain stable over time. Pain is the symptom that is most responsible for medical care and the symptom that most detracts from quality of life. A small percentage of patients do not have pain and instead present with exocrine (steatorrhea, weight loss) or endocrine (diabetes) pancreatic insufficiency.

Most patients present initially with an episode of acute pancreatitis but then develop evidence of chronic pancreatitis; others have obvious chronic pancreatitis at their first presentation. The disease tends to be progressive over time even if the original cause (e.g., alcohol) is removed.

DIAGNOSIS

The diagnosis may be suspected based on the clinical features but must be confirmed by tests that identify either structural damage to the pancreas or derangements in pancreatic function (Table 144-3). Chronic pancreatitis is a slowly progressive disease, and visible damage to the gland (e.g., on a CT scan) and functional failure (e.g., steatorrhea or diabetes) may not be apparent for years. All diagnostic tests are most accurate when the disease is far advanced, and all are far less accurate in the early stages of disease. Early diagnosis, when pain may be severe but imaging study results are normal or equivocal, is difficult.

No clear cause is found in a significant number of patients with chronic pancreatitis. In modern studies from referral centers, almost half of women and about 25% of men are labeled as having *idiopathic chronic pancreatitis*. Some have underlying genetic mutations that put them at particular risk, but gene testing may not be feasible or possible. Even if genetic testing is performed, many commercially available screens (e.g., for CFTR) only test a small percentage of all known mutations, and management will not necessarily be affected. Genetic testing for *PRSS1* is recommended if the family history is suggestive of an autosomal dominant disorder. Others may be surreptitiously using alcohol or may be smokers.

Tests of Pancreatic Structure

Plain abdominal radiographs may demonstrate diffuse or focal pancreatic calcification in patients with advanced chronic pancreatitis. Although specific for chronic pancreatitis, these findings are quite insensitive.

Abdominal ultrasonography is of limited accuracy owing to its inability to visualize the entire pancreas. A dilated pancreatic duct, pancreatic calcifications, gland atrophy, or changes in echotexture are seen in about 60% of patients.

Computed tomography is the most widely used diagnostic test for chronic pancreatitis. High-quality images can be obtained of the pancreas and pancreatic duct. Characteristic findings include a dilated pancreatic duct, ductal or parenchymal calcifications, and atrophy (Fig. 144-2). These structural changes take years to develop, so CT is not as accurate in early or less advanced chronic pancreatitis. Similar to CT, MRI allows detailed images of the pancreas, and the addition of MRCP allows even better assessment of pancreatic duct morphology. At some centers, secretin is administered at the time of MRCP to allow better visualization of the pancreatic duct.

Endoscopic retrograde cholangiopancreatography provides the most detailed images of the pancreatic duct. Changes in the duct include dilation,

TABLE 144-3 DIAGNOSTIC TESTS FOR CHRONIC PANCREATITIS

STRUCTURAL	FUNCTIONAL
Biopsy	Hormonal (secretin) test
Endoscopic ultrasonography	Using an oroduodenal tube
Endoscopic retrograde cholangiopancreatography	Using an endoscope
Magnetic resonance imaging with magnetic resonance cholangiopancreatography	Fecal elastase
Computed tomography	Serum trypsin
Ultrasonography	Fecal fat
Plain radiography	Blood glucose

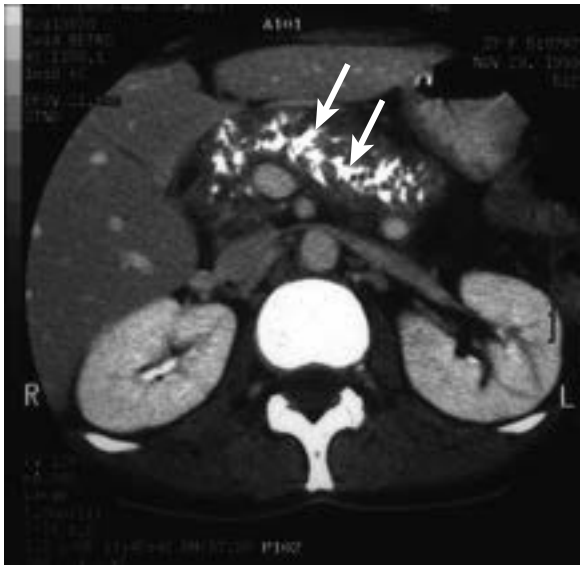

FIGURE 144-2. A computed tomography scan demonstrating diffuse pancreatic calcification in a patient with long-standing chronic pancreatitis *(arrows)*.

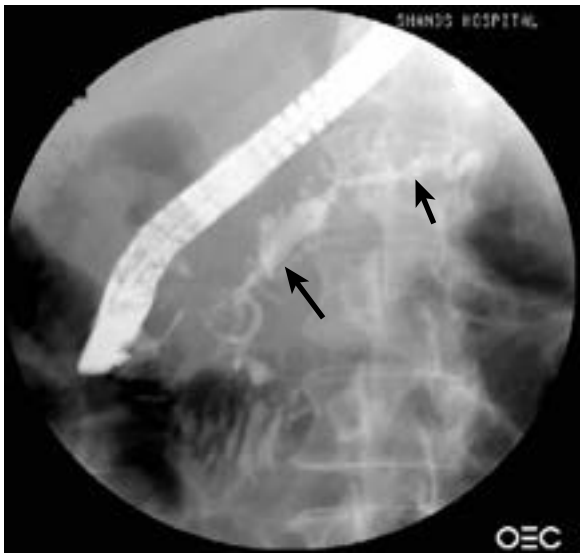

FIGURE 144-3. Endoscopic retrograde cholangiopancreatography demonstrating a very irregular pancreatic duct with areas of dilation and structuring in a patient with chronic pancreatitis *(arrows)*.

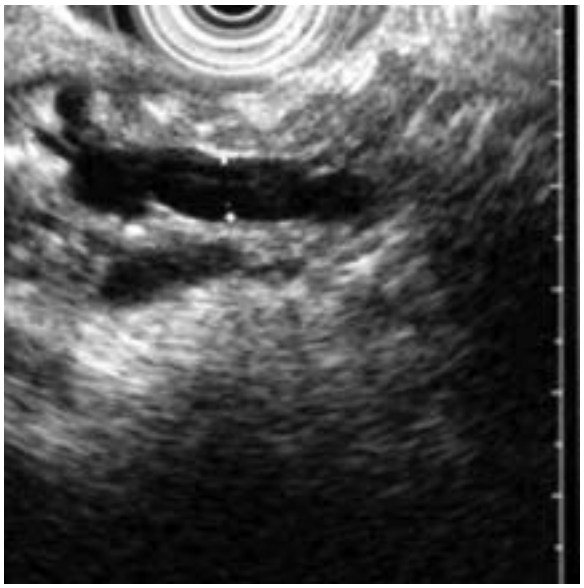

FIGURE 144-4. Endoscopic ultrasonography in a patient with chronic pancreatitis, demonstrating a dilated pancreatic duct (*marks* on margin of main duct).

irregularity, ductal stones, and strictures (Fig. 144-3). These findings are not completely specific for chronic pancreatitis and can be seen in other situations, including pancreatic cancer, after a pancreatic duct stent, and in very elderly individuals. Because of its risk, ERCP should be undertaken only when therapy involving the pancreatic duct is appropriate. Endoscopic ultrasonography allows very detailed images of pancreatic parenchyma and duct (Fig. 144-4) without the risk of ERCP. Normal endoscopic ultrasound results exclude chronic pancreatitis, but very abnormal endoscopic ultrasound results are highly consistent with chronic pancreatitis. However, many endoscopic ultrasound studies show intermediate findings, which are not specific for chronic pancreatitis.

Tests of Pancreatic Function

Serum trypsinogen is abnormally low in patients with far advanced chronic pancreatitis. Levels below 20 ng/mL are seen in patients with chronic pancreatitis that is sufficient to cause functional failure (e.g., steatorrhea). Serum levels of amylase and lipase are of little diagnostic utility for chronic pancreatitis. Serum glucose is elevated in those with endocrine insufficiency.

Quantification of fat in stool during a 72-hour collection while on a high-fat diet can be used to document steatorrhea but is rarely performed. Qualitative analysis of fat with Sudan staining of a stool specimen has poor sensitivity and specificity. Fecal levels of pancreatic elastase are diminished in patients with advanced chronic pancreatitis and steatorrhea. Fecal elastase below 100 mcg/g stool is consistent with advanced chronic pancreatitis. The test can be performed while patients are taking pancreatic enzyme therapy.

One pancreatic function test involves passing an oroduodenal tube and administering a supraphysiologic dose of secretin. Pancreatic secretions are collected over the course of 1 hour and analyzed for their bicarbonate concentration. A normal study is defined by a peak bicarbonate concentration of greater than 80 mEq/L. This test result becomes abnormal earlier in the disease process than any other test but is not widely available. An alternative, using endoscopy instead of a tube, is slightly less sensitive.

Diagnostic Approach

As the disease advances, typically over years, the structural and functional damage accumulate to the point that essentially all diagnostic test results are positive. In most patients, the diagnosis can be or will have been established by routine tests such as CT or MRI. Endoscopic ultrasonography and ERCP are rarely needed for diagnostic purposes in patients with long-standing chronic pancreatitis. The diagnostic challenge lies with patients who present with a severe pain syndrome suggestive of chronic pancreatitis but who have normal CT or MRI results. In these patients, endoscopic ultrasonography is the best choice unless the patient can have access to a secretin-based pancreatic function test. ERCP should not be used for purely diagnostic purposes because of the risk of complications, especially post-ERCP pancreatitis.

TREATMENT Rx

Abdominal Pain

Pseudocysts, obstruction of a surrounding hollow organ (e.g., duodenum or bile duct), and superimposed carcinoma cause chronic pain. A good-quality CT or MRI is usually sufficient to exclude these possibilities and to help choose appropriate therapy. Patients who have a dilated (generally >5 mm) pancreatic duct are candidates for endoscopic and surgical decompression therapy to relieve pain. Patients without ductal dilation are generally not appropriate for endoscopic and surgical therapy and must rely instead on medical therapy (Table 144-4).

Medical therapy starts with vigorous and structured attempts to assist patients in stopping alcohol and tobacco, if applicable. Most patients require analgesics. It is appropriate to start with the less potent agents first (e.g., tramadol, 50 mg four times daily), although many patients require more potent agents (Table 30-4) and may benefit from an adjunctive agent (e.g., gabapentin, pregabalin, selective serotonin-reuptake inhibitors, or tricyclic antidepressants; see Table 30-3 in Chapter 30) to potentiate the narcotic effect. Antioxidants (mixtures of selenium, vitamins E and C, β-carotene, and methionine) have been studied in two large randomized trials, with mixed results. Pancreatic enzyme therapy (see later) may have some beneficial effect on pain.[14]

Endoscopic retrograde cholangiopancreatography can be used to dilate ductal strictures and place stents. Ductal stones, if they are not too large and are not impacted, may also be removed. Lithotripsy of larger stones is usually required to reduce the stone to manageable fragments. This approach is technically successful in more than 80% of carefully selected patients, with pain relief in 70% to 80% of patients. Unfortunately, only a subset of patients with chronic pancreatitis has ductal anatomy that is amenable to this type of therapy.

Endoscopic ultrasound-guided celiac plexus block, which uses a local anesthetic and a steroid, or neurolysis, which uses absolute alcohol, can reduce the pain of chronic pancreatitis for weeks to months. However, the durability of those approaches has not been demonstrated, so they should be viewed as temporizing measures at best.

Surgery to decompress the pancreatic duct can provide more effective and durable long-term outcomes than endoscopic therapy for chronic pancreatitis. The most commonly performed procedure involves a longitudinal incision of the pancreatic duct from the body of the pancreas to as close to the duodenum as possible, and this "filleted" duct is overlaid with a defunctionalized Roux limb. At the time of surgery, ductal strictures can be incised and ductal stones can be removed. The procedure is relatively simple in those with a dilated pancreatic duct (>5 mm) and preserves maximal pancreatic parenchyma. Pain relief in the short term is good (>80%), with about 50% obtaining long-term relief of pain. Alternative surgical procedures for pain include partial pancreatic resection, typically the head of the gland. More ambitious procedures, including pancreaticoduodenectomy and total pancreatectomy, usually coupled with autotransplantation of harvested islet cells, are performed as a last resort at a small number of specialized centers.

TABLE 144-4 TREATMENT FOR PAIN ASSOCIATED WITH CHRONIC PANCREATITIS

TREATMENT	EXAMPLES
Medical therapy	Alcohol and tobacco cessation Analgesics and adjunctive agents Antioxidants Non–enteric-coated enzymes
Neurolysis	Celiac plexus block or neurolysis EUS guided CT guided
Endoscopic therapy	Stent Stone removal, lithotripsy
Surgical therapy	Pancreaticojejunostomy (modified Puestow operation) Partial pancreatic resection (Whipple operation, duodenum preserving pancreatic head resection, others) Total pancreatectomy with islet cell autotransplantation

CT = computed tomography; EUS = endoscopic ultrasonography.

TABLE 144-5 ENZYME THERAPY FOR EXOCRINE PANCREATIC INSUFFICIENCY*

PRODUCT	AVAILABLE STRENGTHS	COMMENTS
	USP lipase units/capsule or tablet	
Zenpep	3000; 5000; 10,000; 15,000; 20,000; 25,000	Enteric-coated capsule
Creon	3000; 6000; 12,000; 24,000; 36,000	Enteric-coated capsule
Pancreaze	4200; 10,500; 16,800; 21,000	Enteric-coated capsule
Ultresa	13,800; 20,700; 23,000	Enteric-coated capsule
Pertzye	8000; 16,000	Enteric-coated capsule with bicarbonate
Viokace	10,440; 20,880	Non–enteric-coated tablet

*For the treatment of pain, non–enteric-coated preparations are used. For exocrine insufficiency, cotreatment with acid-reducing medications is necessary when using non–enteric-coated preparations.

Exocrine Insufficiency

Steatorrhea and maldigestion do not occur until approximately 90% of pancreatic enzyme secretion is lost usually after at least 5 to 10 years of chronic pancreatitis. Patients may note weight loss and oily stools but often do not complain of diarrhea. Patients with chronic pancreatitis and exocrine insufficiency maldigest fat, protein, and carbohydrates, but fat maldigestion is most severe. In addition to weight loss, malabsorption of fat-soluble vitamins, particularly vitamin D, is common. A formal 72-stool fat analysis, which is the most accurate method to document steatorrhea and to gauge effectiveness of therapy, is rarely done. Instead, the clinical features and a fecal elastase less than 100 mcg/g stool, coupled with an appropriate response to enzyme replacement therapy, is the best substitute for 72-hour fecal fat testing.

Pancreatic enzymes (Table 144-5) include both enteric-coated (capsules) and non-enteric-coated (tablets) preparations. Non–enteric-coated preparations are the agents of choice if the goal is to treat pain. They can also be used to treat exocrine insufficiency, although the enteric-coated preparations are used more frequently for this indication. No generic products are currently available. The goal of enzyme therapy, which is to administer at least 10% of normal pancreatic output with each meal, translates to approximately 90,000 USP units of lipase with each meal. Because most patients are still producing some digestive enzymes and have a compensatory increase in gastric lipase, it may not be necessary to prescribe the full dosage of 90,000 USP units with each meal. An initial starting dosage of 50,000 to 70,000 units of lipase per meal, with subsequent assessment of the clinical response, is reasonable.

If non–enteric-coated preparations are used, then cotreatment with an H2-blocker or proton pump inhibitor (Table 138-1) is required to prevent acid denaturation of enzymes, a critical point of emphasis. Enzymes should be administered during and immediately after the meal. Supplementation with vitamin D and calcium is appropriate because osteoporosis and osteopenia are very common. Supplementation with other fat- and water-soluble vitamins may also be needed.

Successful enzyme replacement therapy is generally defined as weight gain, absence of visible oil in the stool, and normalization of fat-soluble vitamin levels. Failure of enzyme therapy is most often caused by an inadequate dose. Increasing the dose up to the full 90,000 USP units with meals and encouraging compliance is appropriate as a first step. In patients using a non–enteric-coated preparation, the dose of the H2-blocker or proton pump inhibitor can be increased to reduce the acid destruction of enzymes. Some patients may not respond because a second disease, such as small intestinal bacterial overgrowth (Chapter 140), is contributing to the malabsorption.

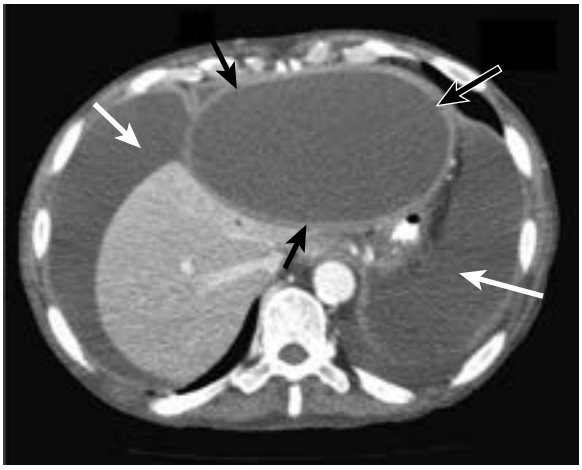

FIGURE 144-5. On computed tomography, a large pseudocyst is seen *(black arrows)*. In addition, ascites surrounding the liver *(white arrow)* is caused by a leak from the pseudocyst (pancreatic ascites).

Endocrine Insufficiency

Diabetes mellitus (Chapter 229) is a very late complication of chronic pancreatitis. Some patients will develop type 2 diabetes, some develop type 3C diabetes in which there is a loss of both insulin and glucagon secretion.[15] In such patients, overly aggressive therapy may lead to hypoglycemia, which cannot be reversed by the usual natural glucagon surge. Treatment-induced hypoglycemia can be fatal in these patients, especially if they are also malnourished. As a result, treatment should avoid exceedingly tight glucose control.

Complications

Pseudocysts, when they are discovered in patients with chronic pancreatitis, are generally mature and have a visible capsule surrounding them. As in acute pancreatitis, pseudocysts in chronic pancreatitis do not require therapy if they are not producing symptoms and are not rapidly enlarging. By comparison, symptomatic pseudocysts require drainage by endoscopic, percutaneous, or surgical procedure.

Pseudocysts may leak into the peritoneal compartment (pancreatic ascites) or track into the chest (pancreatic pleural effusion). Patients usually present with abdominal distention or dyspnea, respectively, rather than abdominal pain. Amylase level in the fluid is usually greater than 4000 U/L. Endoscopic therapy with stent placement across the connection between pseudocyst and pancreatic duct is highly effective in this situation (Fig. 144-5).

Cystic neoplasms require resection. Features that suggest a cystic neoplasm include a cyst with a thick wall or nodules in the wall, a cyst with multiple internal septations, or a cyst occurring in a patient who does not have a history of pancreatitis.

Chronic pancreatitis is also a strong risk factor for pancreatic ductal adenocarcinoma (Chapter 194), with a lifetime risk of about 4% to 5%. The risk is much higher in patients with hereditary pancreatitis and in patients who smoke. Equally important, it may be very difficult to distinguish cancer from benign disease, particularly in those with autoimmune pancreatitis.

PREVENTION

There is not currently any reliable method to prevent chronic pancreatitis, although patients who have fewer episodes of acute pancreatitis are less likely to develop chronic pancreatitis. Patients at risk for chronic pancreatitis and patients with recurrent episodes of acute pancreatitis should avoid alcohol and tobacco. Patients with autoimmune pancreatitis should be treated with steroids (see earlier) to reduce the risk of progression.

PROGNOSIS

The prognosis of chronic pancreatitis is heavily influenced by its cause, as well as by concurrent smoking and ongoing alcohol use. With prolonged follow-up of 10 to 20 years, the majority of patients will develop exocrine or endocrine

insufficiency. The survival rate of patients with chronic pancreatitis is lower than in age-matched control participants. Death is usually not attributable to pancreatitis itself but rather to malignancy, postoperative complications, and complications of tobacco and alcohol.[16] Overall, the 10-year survival rate approximates 70% and the 20-year survival rate is 45%. Patients who are older, smoke, or have alcohol as the cause are at highest risk of mortality.

Grade A References

A1. Wu BU, Hwang JQ, Gardner TH, et al. Lactated Ringer's solution reduces systemic inflammation compared with saline in patients with acute pancreatitis. *Clin Gastroenterol Hepatol*. 2011;9:710-717.

A2. Larino-Noia J, Lindkvist B, Iglesias-Garcia J, et al. Early and/or immediately full caloric diet versus standard refeeding in mild acute pancreatitis: a randomized open-label trial. *Pancreatology*. 2014;14:167-173.

A3. Bakker OJ, van Brunschot S, van Santvoort HC, et al. Early versus on-demand nasoenteric tube feeding in acute pancreatitis. *N Engl J Med*. 2014;371:1983-1993.

A4. Petrov MS, van Santvoort HC, Besselink MG, et al. Early endoscopic retrograde cholangiopancreatography versus conservative management in acute biliary pancreatitis without cholangitis: a meta-analysis of randomized trials. *Ann Surg*. 2008;247:250-257.

A5. Varadarajulu S, Bang JY, Sutton BS, et al. Equal efficacy of endoscopic and surgical cystogastrostomy for pancreatic pseudocyst drainage in a randomized trial. *Gastroenterology*. 2013;145:583-590.

A6. Bakker OJ, van Santvoort HC, van Brunschot S, et al. Endoscopic transgastric vs surgical necrosectomy for infected necrotizing pancreatitis: a randomized trial. *JAMA*. 2012;307:1053-1061.

A7. Sethi S, Sethi N, Wadhwa V, et al. A meta-analysis on the role of rectal diclofenac and indomethacin in the prevention of post-endoscopic retrograde cholangiopancreatography pancreatitis. *Pancreas*. 2014;43:190-197.

A8. Nordback I, Pelli H, Lappalainen-Lehto R, et al. The recurrence of acute alcohol-associated pancreatitis can be reduced: a randomized controlled trial. *Gastroenterology*. 2009;136:848-855.

A9. Olesen SS, Bouwense SA, Wilder-Smith OH, et al. Pregabalin reduces pain in patients with chronic pancreatitis in a randomized, controlled trial. *Gastroenterology*. 2011;141:536-543.

A10. Dite P, Ruzicka M, Zboril V, et al. A prospective, randomized trial comparing endoscopic and surgical therapy for chronic pancreatitis. *Endoscopy*. 2003;35:553-558.

A11. Cahen DL, Gouma DJ, Laramee P, et al. Long-term outcomes of endoscopic vs surgical drainage of the pancreatic duct in patients with chronic pancreatitis. *Gastroenterology*. 2011;141:1690-1695.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Yadav D, Lowenfels AB. The epidemiology of pancreatitis and pancreatic cancer. *Gastroenterology*. 2013;144:1252-1261.
2. Peery AF, Dellon ES, Lund J, et al. Burden of gastrointestinal disease in the United States: 2012 update. *Gastroenterology*. 2012;143:1179-1187.
3. Lerch MM, Gorelick FS. Models of acute and chronic pancreatitis. *Gastroenterology*. 2013;144:1180-1193.
4. Balmadrid B, Kozarek R. Prevention and management of adverse events of endoscopic retrograde cholangiopancreatography. *Gastrointest Endosc Clin N Am*. 2013;23:385-403.
5. Whitcomb DC. Genetic risk factors for pancreatic disorders. *Gastroenterology*. 2013;144:1292-1302.
6. Kamisawa T, Chari ST, Lerch MM, et al. Recent advances in autoimmune pancreatitis: type 1 and Type 2. *Gut*. 2013;62:1373-1380.
7. Banks PA, Bollen TL, Dervenis C, et al. Classification of acute pancreatitis-2012: revision of the Atlanta classification and definitions by international consensus. *Gut*. 2013;62:102-111.
8. Tenner S, Baillie J, DeWitt J, et al. American College of Gastroenterology guideline: management of acute pancreatitis. *Am J Gastroenterol*. 2013;108:1400-1415.
9. Turkvatan A, Erden A, Turkoglu MA, et al. Imaging of acute pancreatitis and its complications. Part 1: acute pancreatitis. *Diagn Interv Imaging*. 2015;96:151-160.
10. Rogers P, Adlan T, Page G. Non-invasive imaging in pancreatitis. *BMJ*. 2014;349:g5223.
11. Wu BU, Banks PA. Clinical management of patients with acute pancreatitis. *Gastroenterology*. 2013;144:1272-1281.
12. Johnson CD, Besselink MG, Carter R. Acute pancreatitis. *BMJ*. 2014;349:g4859.
13. Cote GA, Yadav D, Slivka A, et al. Alcohol and smoking as risk factors in an epidemiology study of chronic pancreatitis. *Clin Gastroenterol Hepatol*. 2011;9:266-273.
14. Forsmark CE. Management of chronic pancreatitis. *Gastroenterology*. 2013;144:1282-1291.
15. Cui Y, Andersen DK. Pancreaticogeneic diabetes: special considerations for management. *Pancreatology*. 2011;11:279-294.
16. Bang UC, Benfield T, Hyldstrup L, et al. Mortality, cancer, and comorbidities associated with chronic pancreatitis: a Danish nationwide matched-cohort study. *Gastroenterology*. 2014;146:989-994.

REVIEW QUESTIONS

1. A 52-year-old man is admitted with presumed acute alcoholic pancreatitis. On admission, his blood pressure is 111/60 mm Hg, pulse is 110 beats/min, respiration rate is 18 breaths/min, and temperature is 37.8° C. His initial physical examination is notable for a tender abdomen without rebound but is otherwise normal. Initial laboratory results include a WBC of 14,000/μL, Hgb 14.1, normal electrolytes, BUN of 26, creatinine of 1.2, and lipase of 720. Liver tests, calcium, and the triglyceride level are normal. Ultrasonography in the emergency department notes a normal gallbladder and bile duct. A computed tomography scan obtained in the emergency department notes some peripancreatic fluid and haziness of the peripancreatic fat. He is admitted to an intermediate care unit and treated with vigorous fluid resuscitation, antiemetics, and analgesics. On day 2, his blood pressure and pulse have normalized, but he has developed a low-grade fever of 38.1° C. He continues to have significant pain and some nausea. A CT scan obtained on day 3 now reveals that 40% of the pancreas is necrotic; there is no gas in the necrotic area. Fever continues with a maximum temperature of 38.2° C, the WBC is still 14,000/μL with a left shift. What would you recommend now?

A. Initiate imipenem.
B. Perform a fine-needle aspiration (FNA) and culture of the necrotic collection.
C. Place a percutaneous drain in necrotic collection.
D. Obtain a surgical consultation.
E. Continue the current conservative therapy.

Answer: E The patient has moderately severe acute pancreatitis with necrosis. Prophylactic antibiotics are not recommended. Infection of the necrosis usually occurs after 1 to 2 weeks of disease, and the clinical features and imaging are not suggestive of infection; as a result, FNA is not needed. The necrosis is not yet walled off or liquefied, so placement of a drain in the collection would be harmful. There is no indication for surgery in this patient. Conservative therapy, which could include placement of a nasoduodenal tube to initiate nutrition, should be continued.

2. A 42-year-old woman is seen in the emergency department with abdominal pain, nausea, and vomiting over the past 12 hours. On physical examination, she is in obvious pain. She is tachycardic, but her vital signs are otherwise normal. Her general physical examination results are normal; her abdomen is mildly distended and tender to palpation without rebound. Laboratory results include amylase of 4500, AST of 220, ALT of 170, alkaline phosphatase of 200, total bilirubin of 1.8, WBC of 12,000/μL, and Hgb of 12; the remaining blood test results are normal. Abdominal ultrasonography reveals several small stones in the gallbladder; the common bile duct is 7 mm. A CT scan reveals interstitial pancreatitis. She is admitted and treated with intravenous fluid, antiemetics, and analgesics. On the following day, her AST is 140, ALT is 160, total bilirubin is 1.6, and WBC is 11,000/μL. She remains afebrile. What would you recommend now?

A. Initiate antibiotics
B. Urgent endoscopic retrograde cholangiopancreatography
C. Repeat ultrasonography now
D. Endoscopic ultrasonography now
E. Continue current conservative therapy

Answer: E This patient has gallstone pancreatitis, which is resolving. There are no features to suggest concomitant cholangitis, and liver chemistries are improving. Urgent ERCP is not required. Prophylactic antibiotics are likewise not indicated. There is no reason to repeat the ultrasonography, and the liver chemistries are improving. An EUS is not needed because there is no clinical concern for microlithiasis. Continued conservative management is appropriate.

3. You see a 58-year-old man who was recently discharged after a 2-week admission for acute pancreatitis. The cause of the pancreatitis is not known, his liver chemistries and triglycerides were normal, and his abdominal ultrasonography revealed no gallstones. He does not drink but does smoke; he was on no medications known to cause pancreatitis. He is now feeling well. A CT scan during the hospitalization revealed enlargement of the head of the pancreas with peripancreatic fluid and stranding but no necrosis. What would you recommend now?

A. Serum IgG4 level
B. Endoscopic ultrasonography
C. Genetic testing
D. ERCP
E. No further testing unless a second attack occurs

Answer: B This patient with an unexplained episode of pancreatitis and at an age older than 40 years must be evaluated for the possibility of malignancy underlying his unexplained pancreatitis. Endoscopic ultrasonography (EUS) provides the most effective method to search for underlying malignancy and to exclude microlithiasis. Genetic testing might be considered if his EUS results are negative. The patient might have autoimmune pancreatitis, but this is a diagnosis of exclusion.

4. A 28-year-old woman is evaluated as an outpatient for chronic abdominal pain. This is her first visit with you, and she reports that she has chronic pancreatitis. The pain is continuous, epigastric, without radiation, and has been present for 2 years. She has not lost weight, does not smoke or drink, and has no family history of pancreatic disease. She is currently treated with oxycodone four times daily, with little relief of pain. You obtain previous medical records, which include several emergency department visits; on one of these visits, her amylase was elevated to 156 (normal <140). Results of several CT scans obtained during these visits were normal. Her physical examination result is normal. Results of laboratory tests, including amylase, lipase, liver chemistries, and triglycerides, are normal. What would you recommend now?

A. ERCP
B. EUS
C. Celiac plexus block
D. Initiate gabapentin
E. Fecal elastase

Answer: B This patient has a chronic pain syndrome but does not have sufficient evidence of chronic pancreatitis. ERCP is not indicated as a diagnostic procedure. Celiac plexus block would only be considered if a diagnosis of chronic pancreatitis was confirmed and even then only in rare circumstances. Gabapentin as an adjunctive agent is reasonable but again only after an accurate diagnosis is made. Fecal elastase would be normal in this patient; even if she has chronic pancreatitis, it is not far enough advanced for fecal elastase to be abnormal. EUS provides the best method to assess this patient for chronic pancreatitis.

5. A 52-year-old man with an 8-year history of chronic pancreatitis attributable to alcohol and tobacco is evaluated for weight loss. He has chronic pain that is managed with tramadol. He has also been treated with pancreatic enzymes and is currently taking 20,000 units of an enteric-coated lipase preparation with each meal. He notes a 20-lb weight loss over the past 6 months. He reports a normal appetite but does have some loose stools. His physical examination is notable for evidence of weight loss. Laboratory testing includes a normal CBC, amylase, and lipase. A CT scan reveals a dilated pancreatic duct with diffuse pancreatic calcifications. Fecal elastase is 85 mcg/g of stool. What would you recommend now?

A. Empiric treatment for small intestinal bacterial overgrowth
B. Add a proton pump inhibitor
C. Increase pancreatic enzyme dosage
D. Pancreatic duct stent
E. Refer for pancreatic surgery

Answer: C This patient has exocrine insufficiency and is on an inadequate dose of enzymes (which should be at least 50,000 units of lipase with each meal and up to 90,000 units). Small intestinal bacteria overgrowth might be considered if he fails to respond to an appropriate dose of enzymes. Acid suppression is not needed with enteric-coated preparations. Placement of a pancreatic duct stent or surgery might be considered for intractable pain, but this patient has pain manageable with simple medical measures.

145 DISEASES OF THE RECTUM AND ANUS

ROBERT D. MADOFF

ANATOMY

The Rectum

The rectum and anal canal make up the final portion of the hindgut. Several definitions exist to delineate the boundaries of each. In general, the rectum begins at the level of the sacral promontory, where the taeniae coli splay to form a continuous longitudinal muscle layer and extends 12 to 18 cm distally. The peritoneum covers the upper two thirds of the rectum anteriorly and is more limited laterally. The rectum and its mesentery are surrounded by endopelvic fascia, and this anatomic package contains the relevant lymphovascular structures that should be removed intact in rectal cancer surgery. The rectum has two or three curves within its lumen created by submucosal folds called the valves of Houston. The second valve is often used as a rough guideline for the intraperitoneal cavity anteriorly.

Blood supply to the rectum originates from the inferior mesenteric artery and internal iliac arteries. The inferior mesenteric artery terminates as the superior rectal (hemorrhoidal) artery, which supplies the rectum and the upper third of the anal canal. Additionally, internal iliac arteries give off the middle rectal (hemorrhoidal) arteries and the inferior rectal (hemorrhoidal) arteries (inferior via the internal pudendal artery) to supply the distal rectum and anal canal. The majority of the rectum drains into the superior hemorrhoidal venous plexus and then to the inferior mesenteric vein and portal and hepatic system. By contrast, the caudal rectum and the anal canal drain into the systemic venous circulation through the inferior and middle rectal veins into the internal iliac veins and inferior vena cava. As a general rule, the lymphatic drainage of the rectum follows the arterial supply via the inferior mesenteric and internal iliac lymph nodes. Innervation to the rectum involves both the sympathetic and parasympathetic plexus. Whereas the sympathetic nerves arise from the first three lumbar segments of the spinal cord, the parasympathetic nerve supply originates from the caudal three sacral nerve roots.

The Anal Canal

The anal canal, which begins at the level of the *levator ani* muscle and extends to the anal verge opening, is about 2.5 to 5 cm in length and is surrounded by the internal and external anal sphincter muscles. The anorectal junction, which can be easily appreciated on digital rectal examination, is the point where the rectum angulates posteriorly from the axis of the anal canal. The internal anal sphincter, which is responsible for about 70% of the resting anal tone, is an extension of the inner circular smooth muscle layer of the rectum. The external sphincter muscle comprises skeletal muscle and is under voluntary control.

The dentate (pectinate) line lies about 2 cm proximal to the anal verge (opening). Hemorrhoids are classified as proximal or distal to the dentate line.

The mucosal lining changes histologically along the course of the anal canal. Superiorly, the anal canal consists of columnar epithelium that mirrors the rectum. Approximately 1 to 2 cm above the level of the dentate line is a transitional zone, where columnar, cuboidal, transitional, and squamous epithelia cells are found. This admixture of cells constitutes the derivation of the term "basaloid" because it relates to the anal canal cancers and marks the proximal boundary for anal cancer surveillance. Distal to this line, the squamous epithelium extends to the anal verge and perianal skin, eventually adding glandular and hair follicles that resemble skin elsewhere on the body. Infected anal glands are a frequent cause of perianal abscess and fistula. Lymphatic drainage below the dentate line drains to the inguinal nodes. The external anal sphincter is innervated by fibers from S4 and the internal pudendal nerve. Somatic sensation of the anal canal comes from the inferior rectal nerve via the pudendal nerve. This somatic sensation ceases 1 to 2 cm above the dentate line, which explains why hemorrhoid ligation can be performed without anesthesia.

SPECIFIC ANORECTAL CONDITIONS

Hemorrhoids

EPIDEMIOLOGY AND PATHOBIOLOGY

An estimated 10 million or more individuals in the United States experience symptoms related to hemorrhoids each year, and these individuals generate more than 1 million annual office visits. The overall prevalence of symptomatic hemorrhoids is estimated to be approximately 5%, but many anorectal complaints attributed to "hemorrhoids" are caused by other conditions, so the true incidence and prevalence of hemorrhoids are unknown.[1]

Despite the commonly held notion that hemorrhoids are always abnormal, they actually are normal structures, identifiable even in fetuses. Hemorrhoids are vascular cushions that consist of connective tissue, smooth muscle, and both arterioles and veins. Functionally, they may aid with overall continence by serving as a malleable gasket to optimize the seal of the anal canal.

Hemorrhoids are *not* rectal varices, which are a distinct entity caused by portal hypertension. *Hemorrhoidal disease* is a more appropriate term to describe the pathologic state that generates symptoms. The arteriolar component explains why hemorrhoidal bleeding is typically bright red in color and can be copious in quantity. Causative factors that can provoke hemorrhoidal symptoms include constipation, diarrhea, older age, pregnancy, and prolonged straining at stool. Repeated stretching of the anal canal also may damage the supporting tissue and result in downward displacement of the vascular cushions. Although none of these precipitating factors has been established rigorously, each can cause either increased abdominal pressure or obstruction of venous return, thereby leading to engorgement and enlargement of the vascular cushions.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The most common symptom of internal hemorrhoids is bright red bleeding. This bleeding is typically painless and is most often seen on the toilet tissue or in the toilet bowel. The quantity of blood is variable, but some patients complain of blood dripping or squirting into the toilet bowl. Passage of dark blood or blood mixed in the stool suggests a more proximal source. Internal

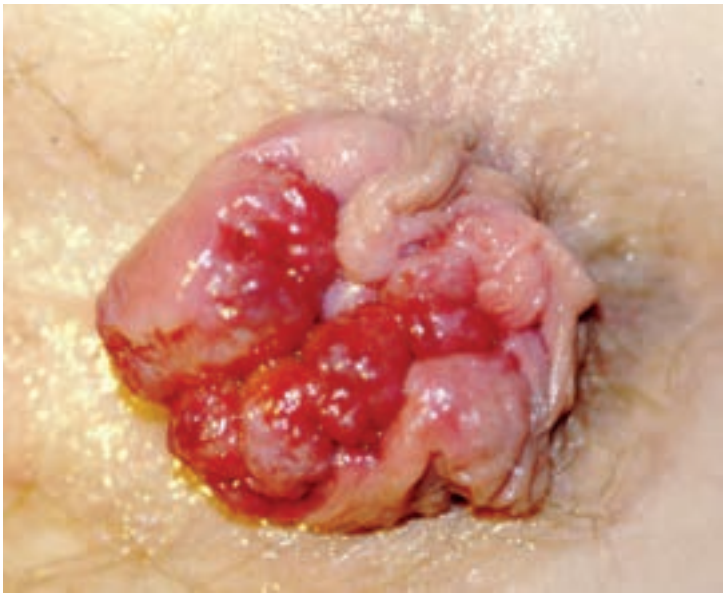

FIGURE 145-1. Grade 4 nonreducible internal and external hemorrhoids.

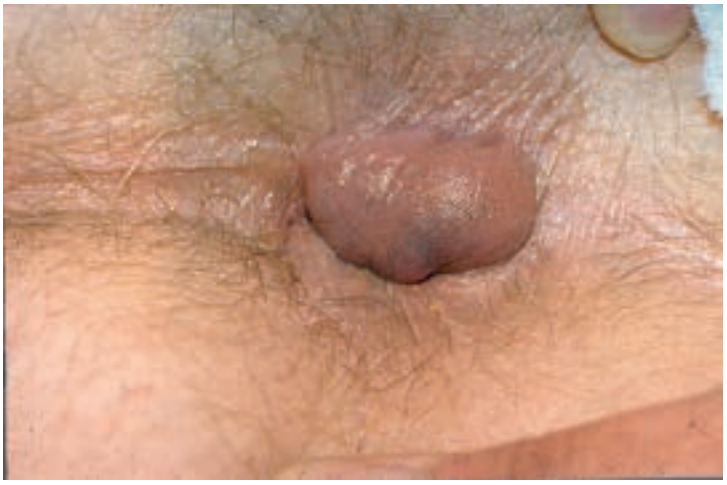

FIGURE 145-2. Thrombosed external hemorrhoid.

hemorrhoids are classified based on the symptoms they cause. As internal hemorrhoids enlarge, they become associated with redundant rectal mucosa that protrudes from the anus with defecation. Grade 1 hemorrhoids bleed but do not prolapse. Early protrusion reduces spontaneously (grade 2); more advanced protrusion requires digital reduction (grade 3) and, at its most advanced, becomes irreducible (grade 4) (Fig. 145-1). Internal hemorrhoids are insensate and are not itchy themselves, but they can cause itching owing to associated perianal soiling or mucus deposition caused by mucosal prolapse. Additionally, patients may complain of mucus discharge, extra tissue at the verge (i.e., mucosal prolapse), or a sensation of incomplete evacuation.

Individuals with anorectal symptoms frequently present complaining of "hemorrhoids" or are even referred from other physicians with a diagnosis of hemorrhoids. Although hemorrhoids may be present, they are often not the source of the underlying complaint. Therefore, it is always incorrect—and sometimes dangerous—for the physician to apply this diagnosis without completing an adequate evaluation. Fortunately, the initial evaluation is simple, and the correct diagnosis is often suspected based on history alone and confirmed by a limited visual and endoscopic examination. Internal hemorrhoids are best visualized with a slotted anoscope. Rectal bleeding should never be attributed to hemorrhoids alone, even if hemorrhoids are visible; at a minimum, flexible sigmoidoscopy is required to exclude more proximal pathology. For more concerning symptoms or high-risk patients (e.g., personal or family history of colorectal cancer, persistent bleeding despite therapy, unscreened individuals older than age 50 years), a full evaluation of the large intestine should be performed by colonoscopy.

External hemorrhoids are usually asymptomatic and should be differentiated from perianal skin tags, which occasionally cause difficulties with hygiene. External hemorrhoids become symptomatic when they thrombose to cause acute-onset pain and swelling (Fig. 145-2). Thrombosed external hemorrhoids are diagnosed by simple inspection. Occasional patients who present with extensive, circumferential thrombosis require urgent surgical consultation.

TREATMENT

Rx

Grade 1 hemorrhoids often respond to dietary manipulation alone, including increased dietary fiber, addition of a fiber supplement, and increased water intake (Table 145-1). The goal, which may not be readily achievable, is approximately 25 to 30 g of fiber and 8 glasses of water daily. More advanced hemorrhoids require specific therapy, which is almost always office based. The most popular and simplest technique is rubber band ligation, whereby a tiny rubber band (internal diameter ≈1 mm) is placed around a quantity of redundant rectal mucosa and prolapsing hemorrhoid well above the dentate line. The banded tissue sloughs in 7 to 14 days, an event sometimes heralded by rectal bleeding. Banding can be repeated at 3- to 4-week intervals until bleeding and protrusion are controlled. Bands placed too close to the dentate line cause immediate severe pain and must be removed promptly. Patients who are at high risk of bleeding because of intrinsic coagulopathies or treatment with anticoagulant agents should not undergo rubber band ligation owing to the increased risk of postprocedure bleeding. Alternative therapies for moderate internal hemorrhoids include injection sclerotherapy and infrared coagulation, both of which are office-based procedures.

Operative hemorrhoidectomy is needed in only a minority of patients with advanced disease. Indications for hemorrhoidectomy include irreducible prolapse, a substantial external component, and failure of more conservative therapies. The most common approach is an excisional hemorrhoidectomy, which is generally performed on an outpatient basis. A more recent approach is the stapled hemorrhoidopexy, which entails resecting a ring of rectal mucosa proximal to the internal hemorrhoids using a circular stapling device. This technique is associated with less postoperative pain and a shorter period of disability, but its drawbacks include a higher recurrence rate than conventional surgery as well as a small but worrisome risk of significant complications, such as chronic pain, rectovaginal fistula, and staple line bleeding. Another alternative approach is Doppler-guided hemorrhoidal artery ligation, after which recurrence rates are 5% to 15%.

Thrombosed external hemorrhoids are usually treated by excision in the office under local anesthesia. However, because the pain associated with thrombosis generally abates within 7 to 10 days, patients who present with resolving symptoms are often best managed conservatively with standard doses of over-the-counter analgesics, Sitz baths, and 25 g/day of fiber supplementation.

TABLE 145-1 INTERNAL HEMORRHOIDS: GRADING AND MANAGEMENT

GRADE	SYMPTOMS AND SIGNS	MANAGEMENT
1	Bleeding No prolapse	Dietary modifications* Rubber band ligation Infrared coagulation Injection sclerotherapy
2	Prolapse with spontaneous reduction Bleeding, seepage	Rubber band ligation Infrared coagulation Dietary modifications Injection sclerotherapy Doppler hemorrhoidal artery ligation
3	Prolapse requiring digital reduction Bleeding, seepage	Surgical hemorrhoidectomy Surgical hemorrhoidopexy Rubber band ligation Dietary modifications Doppler hemorrhoidal artery ligation
4	Prolapsed, cannot be reduced Strangulated	Surgical hemorrhoidectomy Urgent hemorrhoidectomy

*Dietary modifications include increasing the consumption of fiber, bran, or psyllium and water. Dietary modifications are always appropriate for the management of hemorrhoids and to prevent recurrence after banding or surgery (or both).

Perianal Abscess

EPIDEMIOLOGY AND PATHOBIOLOGY

Several superficial and deep spaces around the rectum and anal canal normally contain loose areolar tissue but serve as potential sites for perianal infections. Perianal abscess is a common condition, but its incidence is not well documented. Approximately 80% of perianal abscesses are caused by infection of the anal glands that track toward the skin, but other causes include simple skin infections, trauma, inflammatory bowel disease, anorectal surgery, malignancy, and immunosuppression.[2]

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients most commonly present with complaints of perianal pain and swelling. In most cases, a local area of erythema, tenderness, and fluctuance can be appreciated on physical examination. However, these findings are often absent in an intersphincteric abscess, which is a small abscess in the plane between the internal and external sphincter muscles, as well as in supralevator and deep ischiorectal abscesses. These abscesses should be suspected based on a history of increasing pain and fever, as well as the physical finding of focal perianal tenderness.

As with all acutely painful anal conditions, digital examination should generally be avoided. Likewise, office instrumentation with an anoscope or proctoscope is contraindicated because these examinations cause substantial pain and yield little if any diagnostic information. When the cause of the acute pain cannot be determined in the office, prompt examination under anesthesia should be performed. Adjunctive computed tomographic (CT) scanning, magnetic resonance imaging (MRI), or endorectal ultrasound can provide valuable information in occult, recurrent or complex disease.

TREATMENT Rx

Treatment for perianal abscess is prompt incision and drainage,■ which usually can be performed in the office. Large or deep (e.g., postanal or horseshoe) abscesses are best evaluated in the operating room with proper sedation. Antibiotics do not adequately penetrate abscess cavities, and extension of a local infection can lead to sepsis and complex long-term problems. Therefore, antibiotic therapy is inadequate and should never be given in an attempt to avoid or delay incision and drainage. Antibiotics generally are not indicated after incision and drainage, but exceptions include immunocompromised patients (those with poorly controlled HIV, transplant recipients, patients undergoing chemotherapy, patients with diabetes), patients with extensive cellulitis or severe systemic symptoms, and patients at high risk for endovascular infection (e.g., patients with cardiac shunts or prosthetic valves). In patients who fail to improve after incision and drainage, prompt surgical reevaluation is usually warranted because the problem is likely to be a residual abscess that requires further drainage.[3]

Anal Fistula

EPIDEMIOLOGY

An anal fistula, which represents the chronic form of a perianal abscess, usually manifests as one or more chronic tracts from the anal canal to the perianal skin. The incidence of anal fistula is about 8.6 per 100,000. Fistulae are two to three times more common in men than in women.

PATHOBIOLOGY

After a perianal abscess is drained, there is about a 30% to 50% chance that the internal opening—the site at the dentate line where the infected gland originated—will remain patent, thereby leaving a source for recurrent infection. Multiple or atypical anal fistulae should raise always the diagnostic suspicion of Crohn disease, which is isolated to the perianal area in approximately 10% of cases.

Anal fistulae are characterized by their relationship to the sphincter complex. The simplest and most common (≈70%) fistula is intersphincteric—located in the plane between the internal and external sphincter muscles. Transsphincteric fistulas (about 20%-25%), which traverse both the internal and external sphincter muscles, are classified either as low fistulae, which traverse only the distal external sphincter, or high fistulae, which traverse the more proximal portions of the external sphincter. Suprasphincteric fistulae originate at the dentate line and loop over the entire sphincter complex. Extrasphincteric fistulae have internal openings remote from the dentate line; most originate from a pelvic abscess caused by a ruptured appendix (Chapter 142), diverticulitis (Chapter 142), or Crohn disease (Chapter 141). A horseshoe fistula is one with external openings on both sides of the midsagittal plane; these most commonly have a single internal opening in the posterior midline.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Anal fistulae sometimes present as recurrent abscesses in the same location as the original one or as persistent purulent drainage from an abscess site that has failed to heal completely. Patients may often think the area has healed for several weeks or longer before experiencing the same feeling of a "boil" forming in the identical area, spontaneously rupturing, and relieving their symptoms.

The diagnosis of an anal fistula is established by history and by visualization of an external opening in the perianal skin. A fibrous fistula track can sometimes be palpated along the course of the fistula from the skin toward the anal canal. An internal opening is occasionally visible on anoscopy, but it is not necessary to identify one to make a presumptive diagnosis.

TREATMENT Rx

Treatment of anal fistulae is surgical. Most fistulae are cured by being laid open to eliminate the original source of infection at the internal opening. The fistula track heals by secondary intention. However, this approach divides sphincter muscle and puts the patient at risk for impaired fecal continence in proportion to the quantity of muscle involved. In general, intersphincteric and low transsphincteric fistulae can be safely laid open if the patient has normal baseline continence and no underlying predisposing factors for diarrhea (e.g., colitis) or recurrent fistulas (e.g., Crohn disease). Because the anterior sphincter mechanism is relatively short and subject to injury after vaginal delivery, fistulotomy for anterior fistulae in women must be undertaken only after careful consideration.

When a high fistula is identified, the first step is often placement of a seton, a suture, or other material (now commonly a Silastic vessel loop) that is passed though the fistula tract, out the anus, and secured to itself. The seton guarantees that the external fistula opening will not heal over, so a recurrent abscess is much less likely to supervene. After being left in for several weeks, the tract has often scarred around the seton and become fibrotic; treatment options to eliminate the internal opening include endorectal advancement flap repair and ligation of the intersphincteric fistula tract. Because Crohn disease-associated fistulae tend to be multiple and recurrent, fistulotomy is avoided, except for the most superficial fistulae. In general, patients with Crohn disease are best served by placement of long-term draining setons and medical therapy for their underlying disease (Fig. 145-3).

Anal Fissure

EPIDEMIOLOGY

Anal fissures can occur at any age but most frequently affect young adults. Men and women are equally affected.

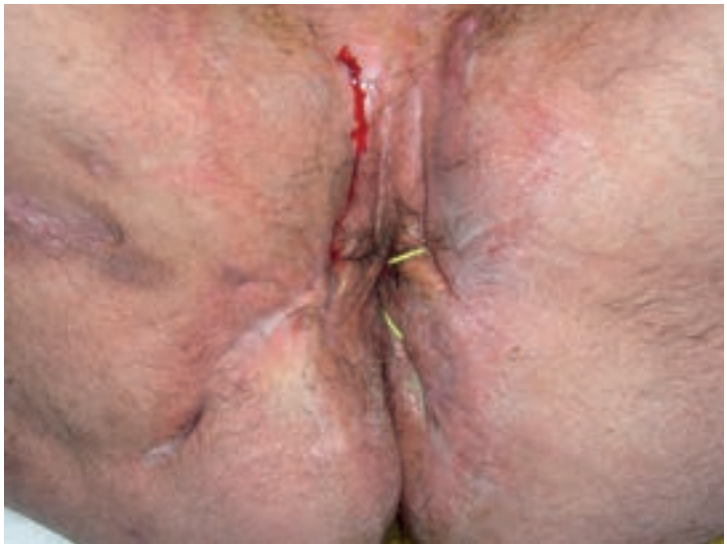

FIGURE 145-3. Crohn disease fistulae. Silastic setons in place with several additional draining tracts and evidence of prior surgery.

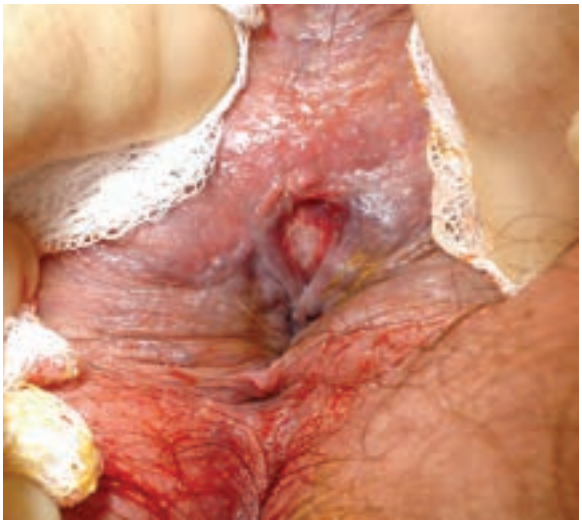

FIGURE 145-4. Anal fissure. This longitudinal tear occurs just inside the anal margin.

PATHOBIOLOGY

An anal fissure is a longitudinal tear in the anoderm that occurs just inside the anal margin (Fig. 145-4). The underlying pathophysiology of anal fissures is hypertonia of the internal anal sphincter. Typical anal fissures occur at the midline; most commonly, they occur posteriorly, but about 15% are found anteriorly or both anteriorly and posteriorly. "Off-the-midline" fissures may represent a routine fissure, but they generally require examination under anesthesia with culture, biopsy, and pathological evaluation to exclude causes such as anal cancer, Crohn disease, syphilis, HIV, leukemia, or tuberculosis.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients with anal fissures generally present with pain after a bout of constipation or a period of excessive diarrhea. After subsequent bowel movements, patients describe severe anal pain that may persist for hours or even continue until exacerbation by the next bowel movement. There is sometimes an association with minor bright red bleeding, most commonly seen in small quantities that streak the stool or the toilet tissue. Whereas an acute anal fissure appears as a superficial split in the perianal skin and anoderm, chronic anal fissures, defined by their presence for at least 6 to 8 weeks, are associated with a "sentinel" skin tag (so called because its presence should suggest the presence of an underlying fissure) and a hypertrophied anal papilla located just proximal to the fissure at the dentate line. Well-established fissures may have fibrotic margins and visible internal anal sphincter fibers at their base.

Most fissures can be readily observed on physical examination by applying opposing traction to the buttocks. In general, after a classic anal fissure is identified, no further examination is performed at that time. Digital and endoscopic examination typically should be delayed until the patient has healed to avoid causing pain. However, patients should be advised that they will require subsequent flexible sigmoidoscopy to exclude proximal pathology.

TREATMENT Rx

All therapies are directed at the underlying hypertonia of the internal anal sphincter. Approximately 40% of fissures heal with fiber supplementation and increased fluid intake alone, including the great majority of acute fissures. Patients are also advised to take warm Sitz baths for symptomatic relief, especially after bowel movements.

Two pharmacologic approaches can augment diet and Sitz baths: topical sphincter relaxants and botulinum toxin injection. Topical nitroglycerine ointment (0.2%-0.8%) or diltiazem gel (2%) applied to the anal orifice two to four times daily reduces sphincter tone and often leads to healing. Both drugs have similar efficacy, but nitroglycerine has the significant disadvantage of causing headaches, lightheadedness, or syncope owing to systemic vasodilation in approximately 25% of patients. Unfortunately, these nonsurgical approaches are only marginally better than placebo for healing chronic fissures, and recurrent fissures occur in approximately 50% of patients.

For patients who fail medical therapy or who are simply too miserable subjectively to pursue it, lateral internal sphincterotomy is an appropriate and generally safe approach that is easily performed under monitored local anesthesia in an outpatient setting. Recovery is rapid, and fissure healing is expected in more than 90% of cases. Sphincterotomy as first-line therapy provides higher healing rates, fewer relapses, and fewer side effects than topical nitroglycerine, with fewer symptoms, greater satisfaction, and no difference in continence at long-term follow-up. However, a small percentage of patients who undergo sphincterotomy develop minor seepage or actual incontinence, so sphincterotomy should be avoided in individuals who have underlying impaired continence, known sphincter injuries, or diarrheal disorders. When such patients have refractory fissures, a trial of botulinum toxin injection is often a good alternative.

Pruritus Ani

EPIDEMIOLOGY

The reported incidence of pruritus ani ranges from 1% to 5% in the general population, although is likely much higher. Men are more commonly affected than women (4 : 1), and this condition is most common in the fourth through the sixth decades of life. Primary or idiopathic pruritus is responsible for 50% to 90% of all cases of pruritus ani.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Pruritus ani, or perianal itching, is a very common symptom but is not a disease. The most common cause is likely inadequate perianal hygiene, often exacerbated by scratching, which leads to excoriation of the skin and further inflammation. Secondary causes include dermatologic, infectious, systemic, and lower gastrointestinal disease, as well as local irritants. Some of these sources include prolapsing hemorrhoids, anal fistulas, anal incontinence, as well as specific dermatologic conditions such as contact dermatitis, psoriasis, lichen sclerosis, squamous intraepithelial neoplasia (Bowen disease), and perianal Paget disease (intraepidermal adenocarcinoma).

The diagnosis is typically based on a history of intractable itching despite trials of several over-the-counter medications and physical examination to exclude an obvious inciting source.[4] The severity of the condition can be classified as stage 0—normal skin; stage 1—red and inflamed skin; stage 2—lichenified skin; or stage 3—lichenified skin as well as coarse ridges and often ulcerations.

TREATMENT Rx

A therapeutic trial of symptomatic empiric management is effective in more than 90% of patients. Options include improved hygiene, avoidance of potential contact allergens (e.g., soaps and over the counter topical treatments), and use of either talc to absorb excess moisture or a barrier cream such as zinc oxide. Mechanical scratching will perpetuate the cycle of inflammation and must be strictly avoided. Some patients benefit from the avoidance of certain foods, including caffeinated beverages, alcohol, milk, chocolate, and tomatoes. Mild topical steroids such as 1% hydrocortisone are sometimes helpful, but stronger steroid preparations are frequently counterproductive and should be avoided when they are not indicated for a specific dermatologic diagnosis. The placement of a small fluff of absorbent cotton at the anal verge can help to wick away moisture and collect any drainage before it can be in prolonged contact with the skin. Other adjuncts include drying the perianal skin with a hairdryer after bathing and replacing toilet paper with nondetergent, nondeodorant, nonalcohol hypoallergenic wipes. Skin biopsies and dermatologic consultation should be obtained when a primary dermatologic condition is suspected or when the perianal irritation fails to heal with conservative therapy.

Fecal Incontinence

EPIDEMIOLOGY AND PATHOBIOLOGY

Involuntary loss of stool or flatus is a relatively common problem that, when severe, can be socially isolating and debilitating. About 2% of individuals in the United States report incontinence symptoms, but the prevalence is substantially higher in patients visiting primary care providers and gastroenterologists.[5] Almost half of U.S. nursing home patients suffer from fecal incontinence. Factors such as embarrassment, denial, and the degree of symptoms hinder a determination of true prevalence. Patients presenting with fecal incontinence often have concomitant urinary incontinence

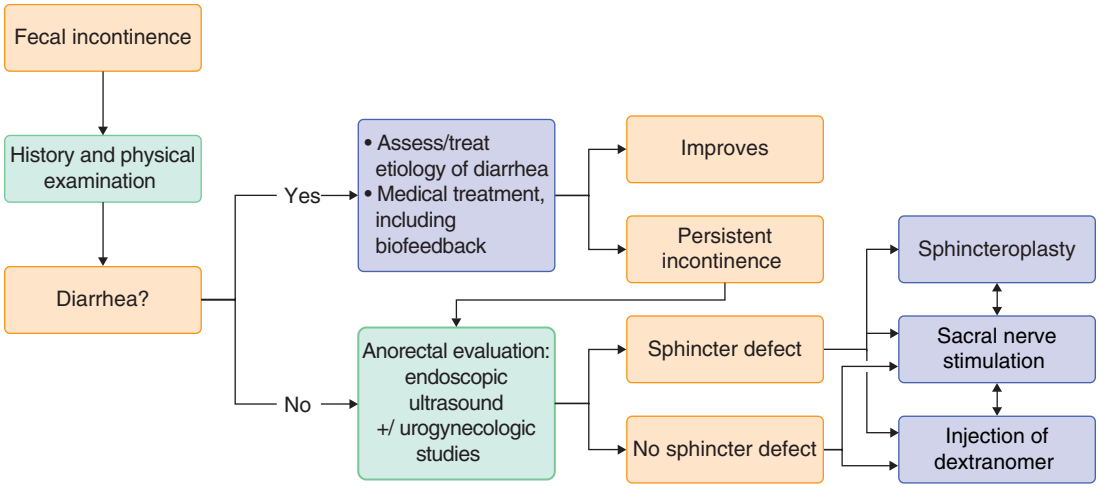

FIGURE 145-5. Algorithm for fecal incontinence. (Adapted from Madoff RD, Parker SC, Varma MG, Lowry AC. Fecal incontinence in adults. *Lancet*. 2004;364:621-632)

(Chapter 26), and up to 25% of women with fecal incontinence have at least one associated pelvic floor disorder.

Normal continence involves the coordinated interaction between multiple different neuronal pathways and the pelvic and perineal structures. In addition, several factors, including bowel motility, stool consistency, evacuation efficiency, mental status, and sphincter integrity, all play roles in normal regulation. Sphincter disruption related to vaginal delivery is a common cause that often affects young women; many incontinent women who present later in life have an underlying sphincter injury for which they can no longer compensate. Other causes of incontinence include fecal impaction, surgical or traumatic sphincter injury, rectal prolapse, neurologic disorders (e.g., diabetic neuropathy, stroke, multiple sclerosis, brain or spinal cord injury), chronic diarrheal states, dementia, impaired mobility, and poor access to toilet facilities.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Because incontinence is a symptom and not a disease, the diagnosis is based on history alone. Specific evaluation of the incontinent patient can include anal manometry, anal ultrasonography, pudendal nerve testing, and defecography. Anal endoscopic ultrasonography, which is generally the most helpful test, accurately depicts sphincter anatomy and reliably identifies sphincter defects. Some patients with fecal incontinence may have concomitant disorders such as impaired evacuation, pelvic prolapse, or urinary incontinence, urogynecologic evaluation should be performed when appropriate.

TREATMENT

Rx

Mild incontinence is often treated successfully with dietary management, addition of a fiber supplement, and use of an antimotility agent such as loperamide 2 to 4 mg up to four times daily. Biofeedback is sometimes successful.

For individuals with sphincter disruption, surgical repair usually leads to substantial improvement (Fig. 145-5). Another alternative is sacral nerve stimulation, which is highly effective for fecal incontinence, with about 40% of patients regaining continence and another 45% noting improvement at 36-months. A less invasive option that stimulates the posterior tibial nerve has shown encouraging early results.

For patients without a sphincter defect, sacral nerve stimulation is one option, but another option is transanal submucosal injection of dextranomer in stabilized hyaluronic acid.[6] These injections add bulk to the perianal tissues, thereby allowing the perianal tissues to coapt together. After two treatments of four injections at each, fecal incontinence decreases by about 50%, but many patients require repeat injections.[7] This treatment should not be used in patients who are immunosuppressed or who have had prior radiation therapy, rectal prolapse, or inflammatory bowel disease.

For patients with severe refractory incontinence, creation of a colostomy should be strongly considered. Initial reluctance notwithstanding, most patients regain control of their bowel function and report a substantially improved quality of life.

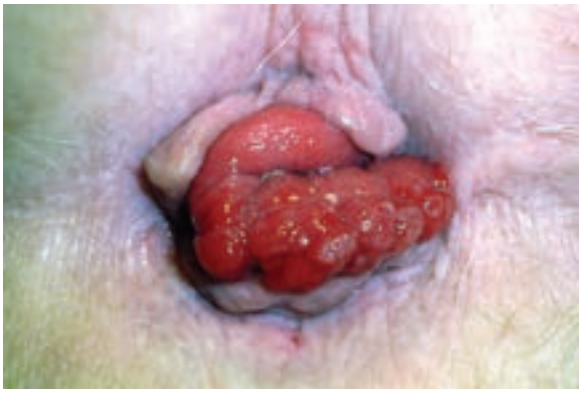

FIGURE 145-6. Rectal mucosal prolapse. The main clinical manifestation of rectal prolapse is the protruding rectal mass.

Rectal Prolapse

EPIDEMIOLOGY AND PATHOBIOLOGY

Rectal prolapse is a full-thickness protrusion of the rectum beyond the anal sphincter. It occurs in about 1% of adults older than age 65 years, and about 90% of cases are in women. Prolapse is caused by an internal rectal intussusception that eventually becomes more severe and protrudes externally. Risk factors include multiparity, a history of pelvic surgery, higher body mass index, chronic diarrhea or constipation, connective tissue disorders, and neurologic diseases. Uncorrected prolapse frequently leads to fecal incontinence by mechanically stretching the sphincter complex and causing a stretch injury to the pudendal nerves.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The main clinical manifestation of rectal prolapse is the protruding rectal mass (Fig. 145-6).[8] The protrusion most commonly occurs with bowel movements, but with time, it may occur with coughing or sneezing, and eventually it can occur spontaneously. Approximately 75% of patients have at least minor complaints of fecal incontinence, and complaints of "constipation," which are often caused by unsuccessful attempts to evacuate the intussuscepting rectum, occur in 15% to 65%. Other associated symptoms include chronic mucus discharge, pelvic discomfort, and minor bleeding. Patients can rarely present with an incarcerated or strangulated prolapse that mandates urgent intervention.

The diagnosis of rectal prolapse is confirmed on physical examination. Full-thickness prolapse, which is characterized by concentric mucosal folds, must be differentiated from circumferential mucosal prolapse, which is characterized by radial folds. The prolapse is often best demonstrated by having the patient strain on a commode. Ancillary studies do not routinely alter management, but colonoscopy may help exclude other pathology. Defecography is most helpful to diagnose internal rectal intussusception and associated pelvic floor abnormalities, such as rectocele and enterocele.

TREATMENT

Rx

Nonoperative measures will not correct rectal prolapse, so conservative measures should be considered only in patients who have very minor and minimally symptomatic prolapse or who are poor surgical candidates. Surgical techniques use one or both of the basic principles of rectal prolapse repair: rectal fixation to the sacrum and resection or plication of redundant bowel. Both transabdominal (laparoscopic or open, with or without mesh) and transperineal approaches provide good outcomes, but the abdominal approach may be associated with lower long-term recurrence rates.

Human Papillomavirus

EPIDEMIOLOGY AND PATHOBIOLOGY

Human papillomavirus (HPV), which is the most common sexually transmitted infection, is a cause of anal dysplasia and cancer. The approximately 40 HPV subtypes that can cause anogenital infections are divided into low-risk types (e.g., types 6 and 11) that cause anal warts (condyloma acuminata) and high-risk types (e.g., types 16, 18, and 33) that can cause anal dysplasia and cancer. HPV infection is associated with cervical, vulvar, vaginal, and penile cancers, and some patients have HPV-related dysplasia or cancer in multiple sites. Approximately 90% of anal cancers are attributable to HPV infection.

The incidence of anal cancer has been steadily increasing, from 0.6 per 100,000 in 1973 to 1.0 per 100,000 in 2001. Over this same period of time, the female-to-male ratio decreased from 1.6 to 1 to 1.2 to 1. In the United States alone, an estimated 5000 or more individuals will develop anal cancer annually, and more than 700 will die from it. These epidemiologic trends have been attributed to a particularly rapid increase in anal cancers among men who have sex with men, especially men infected with HIV. Anal HPV, including the high-risk serotypes, is highly prevalent in at-risk populations (sex workers, intravenous drug users, transplant recipients, men who have sex with men, and HIV-positive men and women). Furthermore, anal dysplasia is common in at-risk populations, as evidenced by the 40% to 60% prevalence rate of high-grade anal dysplasia among HIV-positive men who have sex with men seen in specialty clinics in New York and San Francisco. Other associated risk factors for both dysplasia and anal cancer include a history of other HPV-related genital dysplasia or malignancy (cervical, vaginal or vulvar), prior sexually acquired diseases, anoreceptive intercourse, multiple sexual partners, and cigarette smoking.

ANAL WARTS

CLINICAL MANIFESTATION AND DIAGNOSIS

Anal warts can occur in the perianal skin and within the anal canal. The lesions are raised, epithelialized (when external), and narrow based. They can appear as scattered individual warts or as a confluent mass (Fig. 145-7).

TREATMENT

Rx

External warts can be treated with topical podophyllin or imiquimod, but extensive warts usually require surgical excision or fulguration. Untreated warts can rarely progress to form Bushke-Löwenstein tumors—giant, locally invasive condyloma acuminata that frequently contain in situ or invasive cancer. Anal warts, especially in high-risk patients, have a high risk of recurrence and often require multiple treatments for eradication.

ANAL CANCER

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The most common symptoms of anal cancer are bleeding, pain, and a palpable mass. The cancer may be seen externally as an ulcerated mass or may be palpated within the anal canal. Anal dysplastic lesions may appear as anal warts or as flat, pigmented lesions, or they may be invisible to the naked eye. They are best visualized using anal microscopy after topical application of 3% to 5% acetic acid (see later). Biopsy is necessary to make the diagnosis.

Thorough palpation of the inguinal lymph nodes is important to detect the presence of any clinically relevant adenopathy, but proper staging for anal canal malignancies requires a chest, abdominal, and pelvic CT scan. The size of the primary lesion should be measured, sometimes complemented by lower endoscopic ultrasonography or MRI to assess the size and depth of the lesion, as well as whether the sphincter is involved.[9]

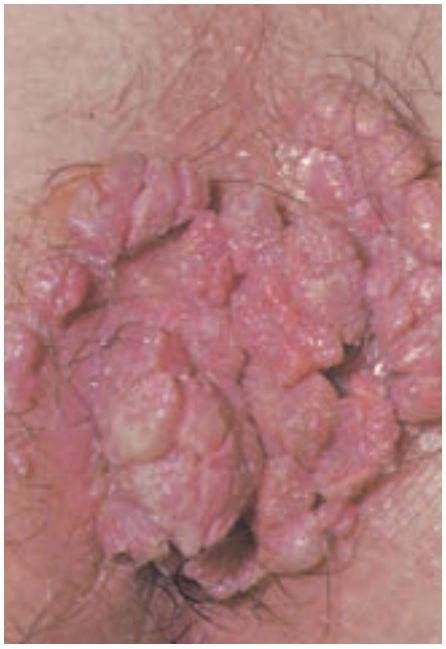

FIGURE 145-7. Anal condyloma. These can occur as individual warts or as a confluent mass.

The terminology of anal cancer is complex, but the great majority of tumors (including epidermoid, cloacogenic, and basaloid carcinomas) are variants of squamous cell carcinoma. The terminology of preinvasive squamous anal lesions is even more confusing because several terms exist for histologically identical pathology. The terms *anal dysplasia, anal intraepithelial neoplasia* (AIN), and *squamous intraepithelial lesion* (SIL) are used interchangeably. SILs are divided into low-grade and high-grade groups; AIN is similarly divided into AIN 1, 2, and 3, with AIN 2 and 3 being classified as high-grade lesions. Squamous cell carcinoma in situ corresponds to high-grade SIL and AIN 3; these terms are preferable to *Bowen disease*, which has historically been applied to this lesion. Other significant but less common anal cancers include adenocarcinoma, melanoma, and Paget disease.

Squamous cell carcinoma of the anus is divided into two groups based on tumor location: cancers of the anal margin (extending from the anal orifice for a distance up to 5 cm) and cancers of the anal canal. Tumors visible externally, but extending into the anal canal, are considered anal canal lesions.

TREATMENT

Rx

Management of patients with anal dysplasia is controversial. Many authorities advocate a screening and treatment approach based on that used for cervical cancer, another HPV-associated disease. High-risk individuals are screened with anal Papanicolaou smears, and high-definition anal microscopy (analogous to colposcopy of the cervix) is performed when abnormal cytology is detected. Using this technique, dysplastic lesions can be identified and focally ablated or treated with topical imiquimod or 5-fluorouracil. Reported complete response rates vary significantly (≈30%-80%), and side effects (pain, irritation, and ulceration) may require withdrawal of therapy.[10] More extensive dysplasia requires microscopy-directed targeted ablation in the operating room.

Chemoradiotherapy is now standard first-line treatment for squamous cell cancer of the anal canal. Current radiation protocols most frequently use 45-Gy external beam radiation therapy in 25 fractions, with a boost to the primary tumor and involved inguinal nodes to a total dose of 54 to 59 Gy. Standard chemotherapy uses 5-fluorouracil (1000 mg/m^2 per 24 hours continuous infusion for 96 hours, starting on days 1 and 29) in combination with mitomycin C (most commonly 10 mg/m^2 intravenous bolus on days 1 and 29). Abdominoperineal resection with permanent colostomy is reserved for tumors that fail to respond to chemoradiation and those that recur. Similarly, groin dissection is performed only when involved inguinal nodes fail chemoradiation.

Early squamous cell cancers of the anal margin can be locally excised if a satisfactory margin can be obtained without injuring the anal sphincter and if

there is no evidence of nodal spread. More advanced anal margin tumors are treated with chemoradiotherapy as described for anal canal tumors. Combined chemoradiation therapy is the primary treatment for most squamous cell carcinomas of the anal canal.

PREVENTION

In a randomized trial, use of an HPV vaccine reduced anal intraepithelial neoplasm by about 50% in men who have sex with men.[A9] Vaccination is warranted in all such individuals.

Other Sexually Transmitted Anorectal Diseases

A number of sexually transmitted diseases (Chapter 285) of the anorectum occur most frequently in individuals who practice anoreceptive intercourse. Common causative agents include *Treponema pallidum* (Chapter 319), *Neisseria gonorrhoeae* (Chapter 299), *Chlamydia trachomatis* (Chapter 318), herpes simplex (Chapter 374), and HIV (Chapter 384). Other sexually transmitted pathogens are *Shigella* (Chapter 309), *Campylobacter jejuni* (Chapter 303), *Haemophilus ducreyi* (Chapter 301), *Calymmatobacterium granulomatis* (Chapter 316), *Entamoeba histolytica* (Chapter 352), *Giardia lamblia* (Chapter 351), and *Isospora belli* (Chapter 353).

The widely variable presentations range from asymptomatic to anal pain, pruritus, discharge, fever, cramps, and bloody diarrhea. Clinical suspicion followed by appropriate and specific testing is necessary to make the correct diagnosis, and clinicians should consider the possibility of simultaneous infections. Treatment addresses the specific infection.

Grade A References

A1. Shanmugam V, Thaha MA, Rabindranath KS, et al. Rubber band ligation versus excisional haemorrhoidectomy for haemorrhoids. *Cochrane Database Syst Rev*. 2005;3:CD005034.

A2. Elmer SE, Nygren JO, Lenander CE. A randomized trial of transanal hemorrhoidal dearterialization with anopexy compared with open hemorrhoidectomy in the treatment of hemorrhoids. *Dis Colon Rectum*. 2013;56:484-490.

A3. Malik AI, Nelson RL, Tou S. Incision and drainage of perianal abscess with or without treatment of anal fistula. *Cochrane Database Syst Rev*. 2010;7:CD006827.

A4. Nelson RL, Thomas K, Morgan J, et al. Non surgical therapy for anal fissure. *Cochrane Database Syst Rev*. 2012;2:CD003431.

A5. Nelson RL, Chattopadhyay A, Brooks W, et al. Operative procedures for fissure in ano. *Cochrane Database Syst Rev*. 2011;11:CD002199.

A6. Brown CJ, Dubreuil D, Santoro L, et al. Lateral internal sphincterotomy is superior to topical nitroglycerin for healing chronic anal fissure and does not compromise long-term fecal continence: six-year follow-up of a multicenter, randomized, controlled trial. *Dis Colon Rectum*. 2007;50:442-448.

A7. Norton C, Cody JD. Biofeedback and/or sphincter exercises for the treatment of faecal incontinence in adults. *Cochrane Database Syst Rev*. 2012;7:CD002111.

A8. Ratto C, Litta F, Parello A, et al. Sacral nerve stimulation in faecal incontinence associated with an anal sphincter lesion: a systematic review. *Colorectal Dis*. 2012;14:e297-e304.

A9. Palefsky JM, Giuliano AR, Goldstone S, et al. HPV vaccine against anal HPV infection and anal intraepithelial neoplasia. *N Engl J Med*. 2011;365:1576-1585.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Jacobs D. Clinical practice. Hemorrhoids. *N Engl J Med*. 2014;371:944-951.
2. Sneider EB, Maykel JA. Anal abscess and fistula. *Gastroenterol Clin North Am*. 2013;42: 773-784.
3. Steele SR, Feingold D, Kumar R, et al. Practice parameters for the management of perianal abscess and fistula-in-ano. ASCRS Standards Committee. *Dis Colon Rectum*. 2011;54:1465-1474.
4. Nasseri YY, Osborne MC. Pruritus ani: diagnosis and treatment. *Gastroenterol Clin North Am*. 2013;42:801-813.
5. Bharucha AE, Rao SS. An update on anorectal disorders for gastroenterologists. *Gastroenterology*. 2014;146:37-45.
6. Roy AL, Gourcerol G, Menard JF, et al. Predictive factors for successful sacral nerve stimulation in the treatment of fecal incontinence: lessons from a comprehensive treatment assessment. *Dis Colon Rectum*. 2014;57:772-780.
7. Maslekar S, Smith K, Harji D, et al. Injectable collagen for the treatment of fecal incontinence: long-term results. *Dis Colon Rectum*. 2013;56:354-359.
8. Melton GB, Kwaan MR. Rectal prolapse. *Surg Clin North Am*. 2013;93:187-198.
9. Steele SR, Varma MG, Melton GB, et al. Practice parameters for anal squamous neoplasms. *Dis Colon Rectum*. 2012;55:735-749.
10. Goldstone SE, Johnstone AA, Moshier EL. Long-term outcome of ablation of anal high-grade squamous intraepithelial lesions: recurrence and incidence of cancer. *Dis Colon Rectum*. 2014;57: 316-323.

REVIEW QUESTIONS

1. A 34-year-old man presents with perianal pain, fever, and leukocytosis. Physical examination demonstrates no obvious source. Which of the following is the most appropriate next step?

A. Repeat examination in 24 hours
B. 10-day course of oral antibiotics
C. Digital rectal examination
D. Office-based needle aspiration
E. Examination under anesthesia

Answer: E Supralevator, intersphincteric, and deeper abscesses may present with symptoms but no obvious source on physical examination. Although adjunctive radiological testing can be considered, the best approach to this potentially serious and urgent problem is examination under anesthesia. Antibiotics alone will not resolve an abscess that needs to be drained. Digital rectal examination and blind needle aspiration in the office should be avoided.

2. A 55-year-old otherwise healthy woman complains of a mass at her anal verge with straining. Which of the following is the most characteristic for differentiating rectal prolapse from other anorectal disorders?

A. Presence of radial folds
B. Concomitant bleeding
C. Presence of concentric mucosal folds
D. A patulous anal sphincter
E. Associated incontinence

Answer: C The diagnosis of rectal prolapse is made clinically, with the major finding of concentric mucosal folds—the so-called "bulls eye." The presence of radial folds is seen in mucosal and hemorrhoidal prolapse. Although bleeding, incontinence, and a patulous anus may be seen in rectal prolapse, each may be also present in many other anorectal disorders and none are specific for rectal prolapse.

146 APPROACH TO THE PATIENT WITH LIVER DISEASE

PAUL MARTIN

The liver serves multiple key functions, including metabolism of the products of ingested food, production of amino acids to form proteins, detoxification of ingested drugs, conversion of nitrogenous substances from the gut into urea, formation of clotting factors, metabolism of bilirubin, processing of lipids absorbed from the intestine, and excretion of its products as bile. The liver also stores glycogen, which is a source of glucose, and helps contain infections by removing bacteria from the blood stream. These diverse functions reflect the activities of hepatocytes, bile duct cells called cholangiocytes, Kupffer cells, endothelial cells, and portal fibroblasts.

The liver has a dual blood supply: 70% delivered by the portal vein, which drains the intestine, and the remainder by the hepatic artery. After arrival in the liver, nutrient-rich portal blood passes along the hepatic sinusoids in close contact with lining hepatocytes before draining into the hepatic vein. The hepatocytes detoxify, metabolize, and synthesize the products of digestion. Bilirubin, which is produced by breakdown of red cells and other hemoproteins by reticuloendothelial cells predominantly in the liver and spleen, is transported to the hepatocytes, bound to albumin, and solubilized by them for biliary excretion.

Liver disease causes loss of hepatocellular activity, with diminished detoxification, excretory, and synthetic functions. Hepatocyte dysfunction results in impaired production of clotting factors, albumin, and other proteins, as well as reduced endogenous formation of lipids. Hepatocyte injury from a variety of causes, including viruses, alcohol, autoimmune disorders, and drug hepatotoxicity, is accompanied by leakage of cellular enzymes into the systemic circulation (Chapter 147). Coagulopathy, decreased serum albumin, and hyperbilirubinemia are observed in more profound hepatocellular injury. Portal hypertension occurs because of disruption of the low-pressure intrahepatic blood flow from the portal to the systemic venous circulation owing to hepatic fibrosis. Consequences of portal hypertension include accumulation of abdominal ascites and the development of portal-systemic venous collaterals with portal-systemic shunting, thereby resulting in the formation of varices and hepatic encephalopathy. Vascular disorders, including portal vein thrombosis, can result in portal hypertension in the absence of parenchymal liver disease.

The diversity of the liver's functions, its complicated blood supply, and its intimate relationship with the biliary tree contribute to the divergent manifestations of liver diseases. The initial complaint often reflects whether the cause is diffuse, such as acute viral hepatitis with widespread hepatocyte injury that manifests as malaise or fatigue, or whether the cause is discrete, such as when biliary obstruction from a gallstone in the common bile duct manifests with severe abdominal pain. Patients may present with multiple complaints, such as nausea and anorexia owing to hepatocellular disease accompanied by right upper quadrant discomfort due to stretching of the hepatic capsule by parenchymal cell edema and inflammation. Patients with more advanced liver disease, such as decompensated cirrhosis (Chapter 153), may have marked hepatocellular dysfunction with jaundice and coagulopathy in addition to portal hypertension with ascites and bleeding esophageal varices. Many patients who present with hepatic symptoms or signs may have extrahepatic disease; for example, a tender, enlarged liver may be caused by a systemic disorder, such as heart failure with hepatic congestion, rather than a primary hepatic disorder. In patients with cirrhosis, the initial presentation of previously unrecognized liver disease may be a major complication such as variceal hemorrhage, which in turn can precipitate hepatic encephalopathy and other features of frank hepatic decompensation.

HISTORY

Patients with liver disorders come to medical attention for a variety of reasons, ranging from the incidental discovery of abnormal liver chemistries to decompensated cirrhosis. Many complaints related to liver disease, such as fatigue, are nonspecific; unless liver disease is considered in the differential diagnosis, recognition of the hepatic origin of these complaints may be delayed.

In clinical practice, a frequent manifestation of asymptomatic liver disease is discovery of abnormal liver biochemistries during a life insurance application, annual physical examination, or attempt to donate blood.[1] It is important to inquire about occasions when liver biochemistries may have been obtained, to determine whether hepatic dysfunction is long-standing or more recent. In a patient with hepatic dysfunction, inquiry should be made about the presence of malaise, anorexia, fatigue, and weight change. Jaundice (Chapter 147) is a dramatic manifestation of possible liver disease. A patient may first notice lighter-colored stools or dark urine rather than scleral icterus. The absence of these latter changes suggests that unconjugated hyperbilirubinemia is due to hemolysis rather than intrinsic liver disease. Not infrequently, a patient may be unaware of jaundice until it is noted by others.

Abdominal pain (Chapter 132) related to liver disease can have a variety of causes. Symptomatic gallstones (Chapter 155) can manifest with the abrupt onset of severe epigastric or right upper quadrant discomfort, often after a large meal and frequently associated with nausea and vomiting. The pain is often steady rather than colicky and can radiate widely, including to the chest and back. A patient may not be able to achieve a position that lessens the pain, which may last several hours. More persistent pain, particularly if associated with weight loss and jaundice, raises concern about malignant bile duct obstruction. Pain is also common in parenchymal liver disease in the absence of biliary tract disease. Many patients with chronic hepatocellular disorders, such as chronic hepatitis C (Chapter 149) or nonalcoholic fatty liver disease (Chapter 152), complain of vague right upper quadrant discomfort that has no particular relieving or aggravating factors. Abdominal pain, which can be severe, is also frequent in acute viral hepatitis (Chapter 148), as well as with the hepatic congestion that results from back pressure in heart failure or hepatic vein occlusion, as in Budd-Chiari syndrome (Chapter 143).

Fatigue, anorexia, and malaise can be present in both acute and chronic liver disease. In acute liver disorders such as acute viral hepatitis (Chapter 148), drug-induced liver disease (Chapter 150), or an acute manifestation of autoimmune hepatitis (Chapter 149), patients may report profound fatigue, nausea, and malaise with decreased appetite and substantial associated weight loss. Distaste for cigarettes is said to be characteristic of acute viral hepatitis. Fatigue is also prominent in chronic liver disease such as chronic hepatitis C (Chapter 149). Pruritus is a prominent feature of cholestatic disorders, such as primary biliary cirrhosis, sclerosing cholangitis, or cholestatic drug reactions, particularly when patients are frankly icteric; however, pruritus also occurs in chronic parenchymal liver disease, most notably chronic hepatitis C, and in acute viral hepatitis. Easy bruisability in those with liver disease reflects coagulopathy and thrombocytopenia.

Fever in a patient with hepatic dysfunction is experienced in the prodrome of acute hepatitis A, as well as in alcoholic hepatitis and drug-induced liver disease. In a patient with suspected biliary obstruction, fever suggests complicating bacterial cholangitis or acute cholecystitis. Ascites is most frequently a manifestation of cirrhosis and portal hypertension in a patient with liver disease. Patients note increasing abdominal girth, which may be preceded by ankle edema. Weight gain owing to fluid retention may be masked by concomitant loss of muscle mass. The onset of ascites in the absence of a history of liver disease suggests a vascular event, such as hepatic vein occlusion (Chapter 143), or a nonhepatic cause of ascites, such as nephrotic syndrome or heart failure. Accumulation of ascites in a patient with liver disease may be subtle, with a slowly increasing waist circumference, or it may be more rapid, such as in a cirrhotic patient who receives fluid resuscitation after gastrointestinal bleeding. Although ascites in a patient with liver disease implies the presence of cirrhosis, ascites also can complicate severe acute liver disease, including alcoholic hepatitis and viral hepatitis, in which it suggests a poor prognosis.

Hepatic encephalopathy (Chapter 153), which is a neuropsychiatric disorder in patients with liver disease, can range from subtle cognitive impairment to deep coma. Early symptoms include a disturbed sleep pattern with nocturnal insomnia and daytime somnolence. More advanced encephalopathy can result in impairment of memory, confusion, and difficulty completing routine tasks. However, new-onset confusion or coma in a patient with liver disease should not be presumed to reflect hepatic encephalopathy unless other explanations, such as sedative overdose or subdural hematoma, have been excluded. Important precipitants of hepatic encephalopathy in a cirrhotic patient include gastrointestinal bleeding, bacterial infection (e.g., spontaneous bacterial peritonitis), electrolyte imbalance, and renal insufficiency,

all of which need to be excluded during the initial clinical evaluation. In a patient with acute liver failure, coma as a consequence of cerebral edema may be impossible to distinguish from advanced hepatic encephalopathy unless the increased intracranial pressure results in papilledema.

Gastrointestinal hemorrhage resulting from bleeding varices is usually profuse and often abrupt in onset. It classically manifests with hematemesis or melena (Chapter 135), and coexisting postural hypotension and presyncope can reflect profound blood loss. The increased protein load in the gut can precipitate hepatic encephalopathy. Nonvariceal causes of gastrointestinal bleeding in a patient with liver disease include portal gastropathy (Chapter 135).

RISK FACTORS FOR LIVER DISEASE

An important aspect of the history is identification of possible risk factors for liver disease. The history should include directed questioning about alcohol consumption, including frequency and pattern (Chapter 33). The age of initial alcohol use and whether consumption has increased with age should be ascertained. Family members also should be asked about their perception of the patient's alcohol use and whether it has resulted in difficulties in personal relationships or work performance. Other clues to alcohol abuse are a history of convictions for driving under the influence of alcohol, motor vehicle accidents, and physical symptoms of alcohol dependence (Chapter 33). More circumspect questioning may be required to elicit a history of recreational drug use, especially given societal disapproval of this activity. Not infrequently, a patient with suspected viral hepatitis admits to smoking marijuana or snorting cocaine but does not acknowledge intravenous drug use. With the increasing frequency of nonalcoholic fatty liver disease as a cause of hepatic dysfunction (Chapter 152), comorbid conditions such as diabetes mellitus, hyperlipidemia, or weight gain should be noted.[2]

Medication use, whether prescription or over the counter, must be assessed because drug-induced liver disease is an important cause of apparently cryptogenic hepatic dysfunction and is not limited to therapeutic drugs (Chapter 150). Increasingly, herbal and "natural" products (Chapter 39) are ingested for a variety of maladies, and patients may fail to disclose their use because they do not perceive these agents to have side effects or may sense that the physician does not endorse their use.[3] As with alcohol, it is important to quantify the amount of medication ingested and over what period. The social history should include details about recent travel and contact with individuals with viral hepatitis through intimate, household, or occupational contact. It is also important to ask about vigorous physical activity that can result in elevated aminotransferase levels of nonhepatic origin.

ASSESSING DURATION OF LIVER DISEASE

The differential diagnosis in a patient with liver disease is determined to a large extent by manifesting symptoms, such jaundice or ascites. In many patients, however, the timing of more subtle findings such as elevated aminotransferases is difficult to determine. Prior blood test results should be retrieved to determine whether hepatic dysfunction is long-standing or more recent. Hepatic dysfunction of less than 6 months' duration is regarded as acute and is frequently self-limited, whereas abnormalities that persist for more than 6 months are unlikely to resolve spontaneously. If the patient has had a cholecystectomy, it is important to determine the indication. Incidental gallstones are sometimes assumed to be the cause of abnormal liver chemistries in a patient with parenchymal liver disease and can lead to an unnecessary operation. Thrombocytopenia owing to portal hypertension in a patient with unrecognized cirrhosis may have been investigated in the past without a firm conclusion being reached.

REVIEW OF OTHER ORGAN SYSTEMS

While focusing on liver-related symptoms, it is important not to overlook other diagnostic clues, associated disorders, and complications. Sicca symptoms, including dry eyes and mouth, are common in primary biliary cirrhosis (Chapter 155); florid features of scleroderma and CREST syndrome (Chapter 267) are other associations. Dyspnea in a patient with hepatic dysfunction may reflect cardiac failure with hepatic congestion (Chapter 58). Other explanations include hepatopulmonary syndrome (Chapter 153), with the characteristic complaint of platypnea–dyspnea (often with chest tightness) that is worse in the upright position owing to aggravation of the ventilation-perfusion mismatch because of intrapulmonary shunting. A hydrothorax in decompensated cirrhosis can cause dyspnea, as can emphysema in patients with liver disease caused by α_1-antitrypsin deficiency. A history of premature menopause is common in middle-aged women with cirrhosis, as is decreased libido and sexual potency in cirrhotic men. Arthralgias are often reported in viral hepatitis, and hemochromatosis (Chapter 212) may manifest with involvement of the proximal interphalangeal joints or chondrocalcinosis of the knees; increased skin pigmentation and diabetes mellitus are other features of this disorder. Accelerated osteopenia occurs in many liver diseases, including primary cirrhosis, primary sclerosing cholangitis, and alcoholic cirrhosis; osteopenia may be aggravated by corticosteroid use in autoimmune chronic active hepatitis. Alcoholic peripheral neuropathy (Chapter 416) can manifest with pain and paresthesia. Tremor and inattentiveness in a younger patient with hepatic dysfunction suggests Wilson disease (Chapter 211). Diarrhea and rectal bleeding in a patient with cholestatic liver disease suggests associated inflammatory bowel disease (Chapter 141).

FAMILY HISTORY

The family history should inquire not only about relatives with liver disease but also associated extrahepatic conditions. Hereditary hepatic conditions (see later), such as Wilson disease and hemochromatosis, may occur in several members of a sibship. In α_1-antitrypsin deficiency, some family members may experience predominantly emphysema rather than cirrhosis. Similarly, renal failure in a family member of a patient with hepatic cysts suggests adult polycystic disease (Chapter 127). A family history of inflammatory bowel disease may be a clue to primary sclerosing cholangitis in a patient with cholestatic liver chemistries. The history in patients with suspected alcoholic liver disease may reveal other family members with alcoholism.

PHYSICAL EXAMINATION

General Condition

In a patient with suspected liver disease, it is crucial to resist the temptation to palpate the abdomen immediately, thus potentially ignoring other important diagnostic clues. Apart from seeking icterus, the initial observation should note whether muscle wasting, cutaneous stigmata of liver disease, abdominal distention, and peripheral edema are present. The vital signs may reflect the hyperdynamic circulation characteristic of cirrhosis, with a resting tachycardia, wide pulse pressure, and low blood pressure resulting from peripheral vasodilation. Fetor hepaticus, which is described as a musty smell, may be detected when a cirrhotic patient exhales and must be distinguished from more frequent halitosis caused by poor dental hygiene.

Mucocutaneous Findings

Icterus (Fig. 146-1) is best confirmed by examination of the sclera or, if necessary, under the tongue, where elastin tissue retains bilirubin. Grayish skin discoloration in hemochromatosis may be most evident in the skin folds in the groin or axilla. Acanthosis nigricans can be observed in nonalcoholic fatty liver disease (Chapter 152). A Kayser-Fleischer ring (see Fig. 211-2), caused by the deposition of copper in Descemet membrane, is a brownish circle around the periphery of the iris and may require slit lamp examination to detect; it always should be sought in a patient with suspected Wilson disease (Chapter 211). Poor dentition is characteristic in alcoholic or drug-abusing individuals, and excessive dental caries may result from decreased saliva production in sicca syndrome. Parotid gland swelling is occasionally observed in alcoholic patients. Central cyanosis and clubbing are found in the hepatopulmonary syndrome. Temporal muscle wasting and "paper money" facial skin,

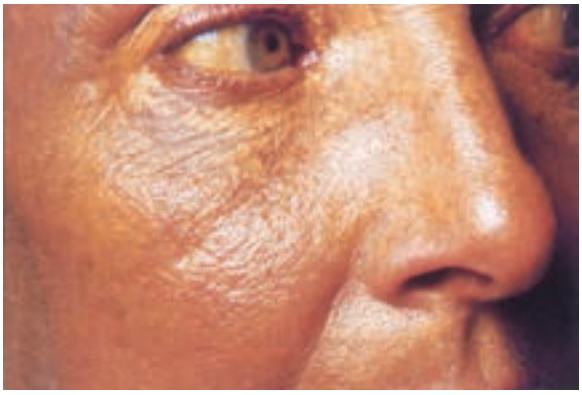

FIGURE 146-1. Scleral icterus.

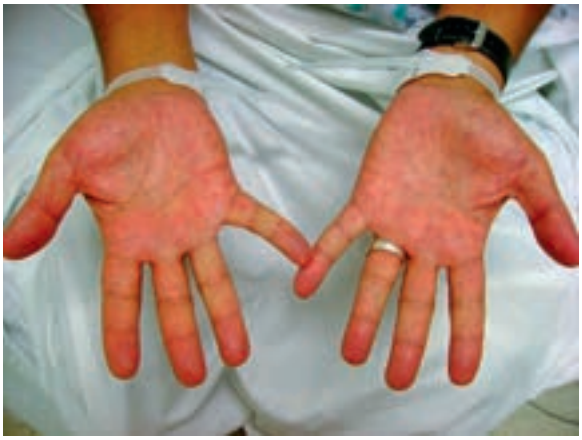

FIGURE 146-2. Palmar erythema.

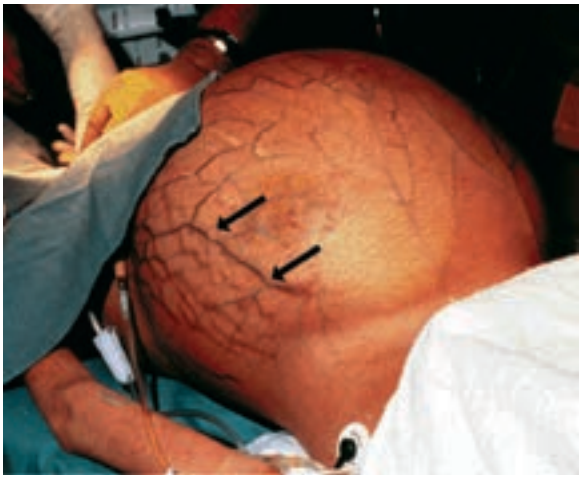

FIGURE 146-3. Caput medusae. Photograph shows caput medusae accentuated by a large amount of ascites in a patient being prepared for liver transplantation. An extensive plexus of veins is seen emanating from the umbilical region and radiating across the anterior abdominal wall. Note the large vein coursing inferiorly along the right flank (*arrows*). This is the superficial epigastric vein, which drains into the external iliac vein. (From Henseler KP, Pozniak MA, Le FT, et al. Three-dimensional CT angiography of spontaneous portosystemic shunts. *Radiographics.* 2001;21:691-704.)

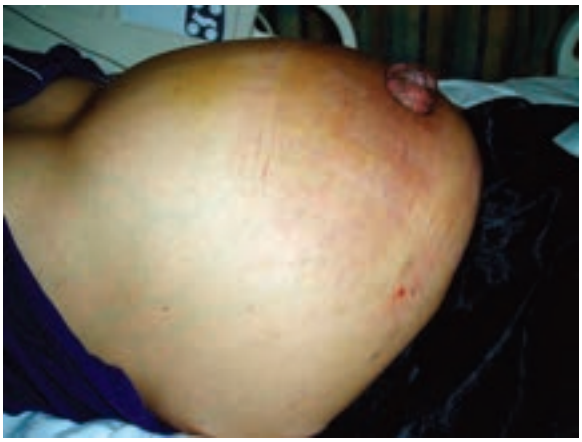

FIGURE 146-4. Ascites.

owing to atrophy with telangiectasia, are signs of advanced liver disease. Xanthelasma from lipid deposits may be observed on the eyelids and skin around the orbits in patients with cholestatic liver disease. Spider nevi on the face and thorax are not pathognomonic of liver disease, especially in women, but they are suggestive if more than a few are present. Palmar erythema (Fig. 146-2) may be normal in women but suggests liver disease in men. Dupuytren contracture (see Fig. 152-1), the retraction of the palmar fascia with subsequent contracture of the palms and fingers can be a sign of alcoholic liver disease, although it is also described in patients with epilepsy or diabetes mellitus, as well as in individuals who have work-related contractures. Petechiae and ecchymoses reflect impaired production of clotting factors and hypersplenism in advanced liver disease. Patches of white discoloration on the nails may be present in advanced liver disease. Scratch marks from pruritus may be observed on the trunk and extremities of patients with cholestatic liver disease. Sparing of the center of the back can lead to a less pigmented butterfly-shaped area because patients cannot reach that area with their fingernails.

Examination of the Abdomen

Abdominal distention with ascites and dilation of collateral veins owing to portal hypertension represent florid signs of advanced liver disease. Caput medusae (Fig. 146-3) in the periumbilical area implies recanalization of the umbilical vein with portal hypertension.

Abdominal percussion may confirm the presence of ascites (Fig. 146-4). Shifting dullness results from movement of ascites to the most dependent portion of the abdomen. The subject should be examined in the supine position, with percussion of the abdomen from the midline toward the right or left flank. A change from a tympanic sound to a dull sound signifies a change from air to fluid, and the location of that change identifies the surface of the fluid pool. Next, the examiner should percuss below the point at which dullness is elicited and ask the subject to turn toward the examiner. With the subject on his or her side, the examiner percusses again at the same point where tympany converted to dullness. If that spot is now tympanic, shifting dullness has been detected as a result of movement of the air-fluid boundary; this finding supports the presence of ascites. This maneuver should be performed sequentially on each side for confirmation. A fluid wave can be felt by placing the medial border of one hand on the abdomen and tapping the right or left lateral abdominal walls; the resulting wave is felt by the first hand. Scrotal edema and abdominal wall hernias are often present in patients with long-standing ascites. Abdominal tenderness in a patient with ascites suggests peritonitis (e.g., spontaneous bacterial peritonitis or the result of a perforated viscus). However, it is important to note that abdominal tenderness is frequently absent in spontaneous bacterial peritonitis.

The liver is dull to percussion. Percussion of the right upper quadrant can determine the liver span, normally 6 to 12 cm in the midclavicular line. The liver span may be diminished in a patient with cirrhosis, whereas hepatomegaly (Fig. 146-5) is detected in hepatic congestion resulting from heart failure, nonalcoholic fatty liver disease, and cholestatic forms of cirrhosis. The liver is best examined with the patient in the supine position, arms parallel to the side of the body, and knees bent to relax the abdominal muscles. Palpation should begin in the right lower quadrant of the abdomen and move upward toward the rib cage so that the liver edge is felt on the way up. A normal liver edge is smooth and sometimes slightly tender when palpated. In general, a liver edge that is felt up to 2 cm below the right costal margin is considered normal, but a normal-sized liver can be displaced downward by other abnormalities, such as emphysema. In thin subjects, the liver edge may be felt on deep inspiration, even if it is normal in size.

The liver can feel hard and irregular, as in cirrhosis, or slightly tender, enlarged, and smooth, as in acute viral hepatitis, alcoholic hepatitis, or hepatic congestion owing to congestive heart failure. The liver can extend across the midline, and the left lobe can be felt in the epigastrium. When the location of the edge of the liver is unclear, the scratch test may be helpful. The bell of the stethoscope is placed on the right upper quadrant over the rib cage while scratching the surface of the abdominal wall from the mid-abdomen toward the liver; the sound of the scratch is amplified in an area under which the liver lies.[4] In the presence of ascites, the liver edge may be detected by exerting quick pressure with the fingertips below the rib cage.

A palpable gallbladder suggests obstruction of the biliary system, whereas tenderness elicited by palpation during inspiration (the Murphy sign) suggests acute cholecystitis. Marked hepatic tenderness with hepatomegaly is observed in patients with a hepatic abscess (Chapter 151).

Splenomegaly (Chapter 168) is suggested by dullness to percussion between the 9th and 11th ribs in the left midaxillary line. A palpable spleen tip implies portal hypertension in a patient with chronic liver disease, although an enlarged spleen also can be detected in acute viral hepatitis and infiltrative disorders that involve both the liver and the spleen (see Table 168-7). Rectal examination is obligatory if gastrointestinal bleeding is suspected because of melena, anemia, or unexplained hepatic encephalopathy.

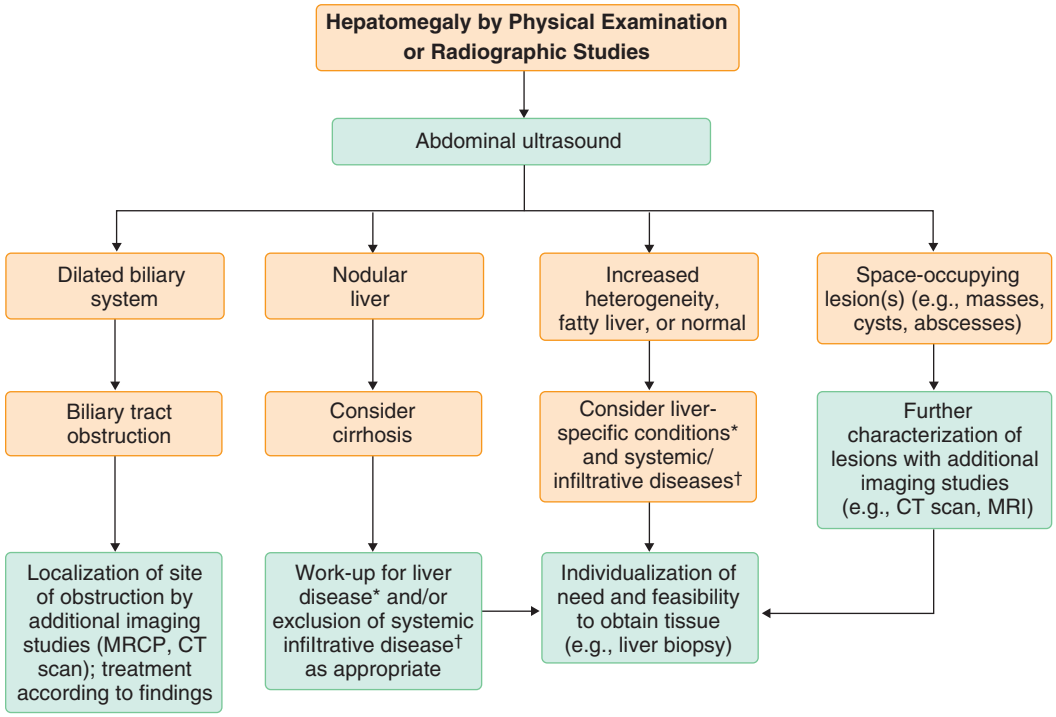

FIGURE 146-5. Diagnostic approach to hepatomegaly. *Conditions to be excluded include viral hepatitis; alcohol- and drug-induced liver disease; steatohepatitis; autoimmune liver diseases; and metabolic disorders, including hemochromatosis, Wilson disease, and α_1-antitrypsin deficiency. †Systemic and infiltrative diseases include amyloidosis, lymphoma, sarcoidosis, and infectious processes such as disseminated tuberculosis and fungemia. CT = computed tomography; MRCP = magnetic resonance cholangiopancreatography; MRI = magnetic resonance imaging.

TABLE 146-1 APPROACH TO COMMON HEPATIC COMPLAINTS

PRESENTATION	COMMON SYMPTOMS	COMMON PHYSICAL SIGNS	DIAGNOSTIC STUDIES	COMMON DIAGNOSES
Ascites	Abdominal distention and pain, ankle edema	Flank dullness Shifting dullness Fluid wave	Ultrasound with Doppler Diagnostic paracentesis Urinalysis	Cirrhosis Budd-Chiari syndrome Heart failure Nephrotic syndrome
Hepatic encephalopathy	Sleep disorientation, confusion, coma	Asterixis Altered mentation Fetor hepaticus	Serum ammonia Blood cultures Stool Hemoccult Serum creatinine and electrolytes	Decompensated cirrhosis Acute liver failure Other metabolic encephalopathies (renal, respiratory)
Hepatic mass	None or abdominal pain	Hepatic bruit or rub	α-Fetoprotein Ultrasound CT scan MRI Biopsy	Benign lesions: Hemangioma, adenoma, focal nodular hyperplasia Malignant lesions: Hepatocellular carcinoma, cholangiocarcinoma, metastases
Abdominal pain	Nausea, vomiting, fever	Right upper quadrant tenderness Palpable gallbladder Murphy sign	Ultrasound HIDA scan Paracentesis for ascites if present	Biliary colic Acute cholecystitis Hepatic congestion Hepatic metastases

CT = computed tomography; HIDA = hepatobiliary iminodiacetic acid; MRI = magnetic resonance imaging.

Complete Physical Examination

The presence of rales and elevation of the jugular venous pressure suggest heart failure or pericardial disease as a cause of hepatic congestion. Loss of secondary sexual characteristics in liver disease is reflected by a loss of axillary and pubic hair, as well as feminization of body habitus in a male patient. Testicular atrophy also may be present. Peripheral edema is common in decompensated cirrhosis and may occur before ascites is obvious.

Neuropsychiatric alterations can include subtle changes in personality or more overt hepatic encephalopathy. Constructional apraxia (e.g., inability to draw a five-pointed star or to write legibly) in a fully conscious patient is a typical finding in hepatic encephalopathy. Asterixis, which is characterized by a series of extensor and flexor wrist movements,[5] can be elicited by having the patient extend the arms, with dorsiflexion of the wrists, while separating the fingers for at least 15 seconds. Tremors (Chapter 410) are nonspecific but are also common in advanced cirrhosis.

DIAGNOSTIC STUDIES (TABLE 146-1)

Laboratory Studies

The initial evaluation of liver disease involves a battery of blood tests, which can assess hepatic necroinflammation (serum aminotransferases), cholestatic biliary tract dysfunction (alkaline phosphatase, γ-glutamyl transpeptidase), excretory function (bilirubin), and synthetic function (coagulation factors, albumin) (Chapter 147). In hepatocellular dysfunction caused by viral hepatitis (Chapter 148), aminotransferase levels are elevated, with the serum alanine aminotransferase (ALT) level higher than the aspartate aminotransferase (AST) level. In alcoholic hepatitis (Chapter 152), AST elevation exceeds that of ALT. In patients with biliary obstruction or cholestatic liver diseases (Chapter 155), such as primary biliary cirrhosis or a cholestatic drug reaction, bilirubin and generally alkaline phosphatase levels are elevated. Impairment of synthetic function leads to a decreased serum albumin level

over days to weeks. The prothrombin time, expressed as the international normalized ratio (INR), is a more sensitive test of hepatic synthetic function and becomes prolonged within several hours after a major hepatic insult. Portal hypertension owing to advanced hepatic fibrosis or cirrhosis results in a decreased platelet count resulting from hypersplenism. Although these general patterns of liver dysfunction are helpful in initial evaluation of liver disease, they are nonspecific. For example, a patient with a predominantly hepatocellular process commonly has elevated bilirubin and alkaline phosphatase levels and a low serum albumin level if the injury is severe. Nevertheless, these blood tests should identify the predominant pattern of abnormality and direct further diagnostic evaluation with serologic studies and abdominal imaging. A variety of approaches have attempted to incorporate readily available tests, such as the ratio of AST to platelet count, or serum markers of fibrosis, such as type IV collagen, to assess hepatic fibrosis, but none of these alternatives can substitute for the accuracy of liver biopsy.

Abdominal Imaging

Plain abdominal radiographs generally do not have a major role in the evaluation of suspected liver disease. An exception is a patient with severe abdominal pain, in whom it is important to exclude a perforated viscus with free air under the diaphragm.

Ultrasonography should be the initial investigation in patients with obstructive jaundice. It can confirm dilated bile ducts in patients with biliary obstruction and often can identify the cause, such as a pancreatic mass or a gallstone lodged in the common bile duct. Ultrasonography also can determine whether the hepatic parenchyma is diffusely abnormal, such as in acute viral hepatitis; it can identify bright hepatic echo texture in nonalcoholic fatty liver disease or a coarsened echo texture in cirrhosis. In addition to confirming the presence of ascites, ultrasonography can identify other signs of portal hypertension, such as splenomegaly or intra-abdominal varices. A Doppler flow study can evaluate blood flow through the portal and hepatic vessels. An ultrasound study can identify hepatic masses and distinguish a cystic mass from a solid lesion (see later). Computed tomography (CT) and magnetic resonance imaging (MRI) add greater detail to the assessment of the hepatic vasculature, hepatic masses, and hepatic vascular structures. MRI can obtain a detailed cholangiogram, thereby avoiding a more invasive endoscopic retrograde cholangiopancreatography in many patients with suspected bile duct obstruction or disease (Chapters 133 and 134). Liver elastrography assesses hepatic fibrosis by measuring liver stiffness by propagation of waves in liver tissue. Generally liver stiffness correlates with fibrosis. *Fibroscan,* which uses high-frequency shear waves, can predict the amount of fibrosis in most patients, but it is less accurate in patients with a high body mass index or extensive steatosis.[6]

Liver Biopsy

Liver biopsy remains the definitive test to assess the severity of hepatic inflammation and fibrosis extent in diffuse hepatocellular disease. In addition, it can confirm certain diagnoses suspected by noninvasive testing, such as autoimmune hepatitis, or it can suggest other diagnoses, such as drug hepatotoxicity. However, because of its potential complications, such as intra-abdominal bleeding, biopsy is recommended only when less invasive testing does not yield a definitive diagnosis or prognosis or when additional information, such as quantitative determination of hepatic copper in Wilson disease or hepatic iron in hemochromatosis, is necessary for definitive diagnosis (see later).

Transjugular pressure measurements are indicated in patients with atypical presentations of portal hypertension (e.g., when it is unclear whether liver disease is the cause) or to titrate medications to reduce portal pressure. A catheter is advanced under fluoroscopy into the hepatic vein, and the free hepatic venous pressure is measured. The catheter is then advanced further until it becomes "wedged" in a small hepatic vein venule. A small balloon occludes the venule, and the wedged hepatic vein pressure, which reflects hepatic sinusoidal pressure, is obtained. The portal pressure gradient, which is derived by subtracting the free pressure measurement from the wedged pressure measurement, is normally less than 5 mm Hg. Varices form at a gradient greater than 10 mm Hg, whereas ascites and variceal hemorrhage occur only when the gradient greater than 12 mm Hg. A transjugular approach also increases the safety of liver biopsy in the presence of ascites, coagulopathy, or thrombocytopenia when standard percutaneous liver biopsy is hazardous. Endoscopy is indicated to screen for varices in any patient suspected of having cirrhosis to determine the need for prophylaxis against hemorrhage.

APPROACH TO A HEPATIC MASS

A hepatic mass may be discovered under a number of different circumstances and is often an incidental finding on abdominal imaging performed for another indication.[7,8] The initial evaluation is directed as to whether the mass is cystic or solid (Fig. 146-6).

Hepatic Cysts

Most hepatic cysts are benign and incidental. Simple hepatic cysts are usually solitary and asymptomatic, although larger cysts may cause abdominal discomfort. Ultrasonography, which is usually diagnostic, shows an anechoic fluid-filled space, an imperceptible wall, and posterior acoustic enhancement (Fig. 146-7A and B). Worrisome features, including the presence of symptoms or increasing size, require exclusion of a cystadenoma. On ultrasound a cystadenoma is hypoechoic with irregular walls and septations. Hepatic resection is indicated because malignant transformation can occur.

Multiple hepatic cysts may suggest the presence of autosomal dominant adult polycystic disease.[9] Polycystic liver disease can occur in conjunction with polycystic kidney disease (Chapter 127) or without it. Isolated polycystic liver disease is associated with mutations in the *PRCKSCH* gene, which encodes the glucosidase 2 subunit β enzyme, and in Sec63, which is part of a large protein complex involved in protein translocation in the endoplasmic reticulum. Adult polycystic liver disease typically is symptomatic, although some patients note dull right upper quadrant pain, fullness, sense of a mass, and increasing abdominal girth. Rupture of a cyst, hemorrhage into a cyst, infection of a cyst, or torsion of a cyst may cause severe pain. Physical examination may include hepatomegaly, cachexia due to weight loss, and ascites. As cysts increase in number and size with age, they may become palpable. On ultrasound, multiple fluid-filled cysts are present without internal echoes unless bleeding or infection has occurred. Interventions indicated for symptoms include aspiration or resection. Polycystic liver disease usually does not result in clinical hepatic impairment, and prognosis depends on the severity of any concurrent polycystic renal disease. In some cases, however, large, symptomatic, or bleeding cysts may raise the consideration of liver transplantation, often with combined kidney transplantation (Chapter 131) for associated renal cysts and failure.

Caroli disease, which is a rare congenital abnormality with cystic dilatation of the intrahepatic biliary tree, can be associated with cholangitis and biliary stones. Many patients ultimately develop cholangiocarcinoma (Chapter 155). Magnetic resonance cholangiopancreatography demonstrates intrahepatic dilatations with normal ducts in between and a normal common bile duct. Endoscopic treatment is indicated for cholangitis, although resection or liver transplantation may ultimately be required.

Solid Hepatic Mass

Most solid hepatic masses are asymptomatic, but some patients may have vague right upper quadrant discomfort. The abrupt onset of more severe pain suggests a complication, such as hemorrhage or rupture. The approach to a solid hepatic mass (Fig. 146-8) is influenced by the presence or absence of underlying chronic liver disease, such as chronic viral hepatitis, as well whether an extraheptic malignancy with possible hepatic metastases is a concern. If chronic liver disease is present, a solid mass must be assumed to represent primary hepatocellular carcinoma (Chapter 196) until proved otherwise. By comparison, a solid lesion in the absence of underlying liver disease is more likely to be benign and incidental. Benign hepatic adenomas are commonly discovered in young to middle-aged women and are not associated with hepatic dysfunction. With modern contrast studies, it is usually possible to make a confident radiologic diagnosis about the nature of a solid hepatic mass. Important examples include the initial peripheral filling and retention of contrast dye by a hemangioma (see Fig. 146-8A, B, and C), the rapid arterial filling of a hepatocellular carcinoma with subsequent "washout" of contrast during the portal venous phase (see Fig. 196-5), or the central scar typical of focal nodular hyperplasia (see Fig. 146-8D), although biopsy is indicated if it is not possible to exclude malignancy.

Hepatic hemangiomas are the most common benign solid hepatic mass. On ultrasound, the lesion is hyperechoic. It has a typical hypervascular appearance on contrast CT or MRI. Generally, no intervention is required unless there is a clear association with abdominal symptoms, increasing size, or a complication such as rupture.

Hepatic adenomas (Chapter 196) are benign epithelial hepatic tumors associated with use of oral contraceptives with a higher estrogen content. On ultrasound, adenomas are hyperechoic, reflective of their fat content,

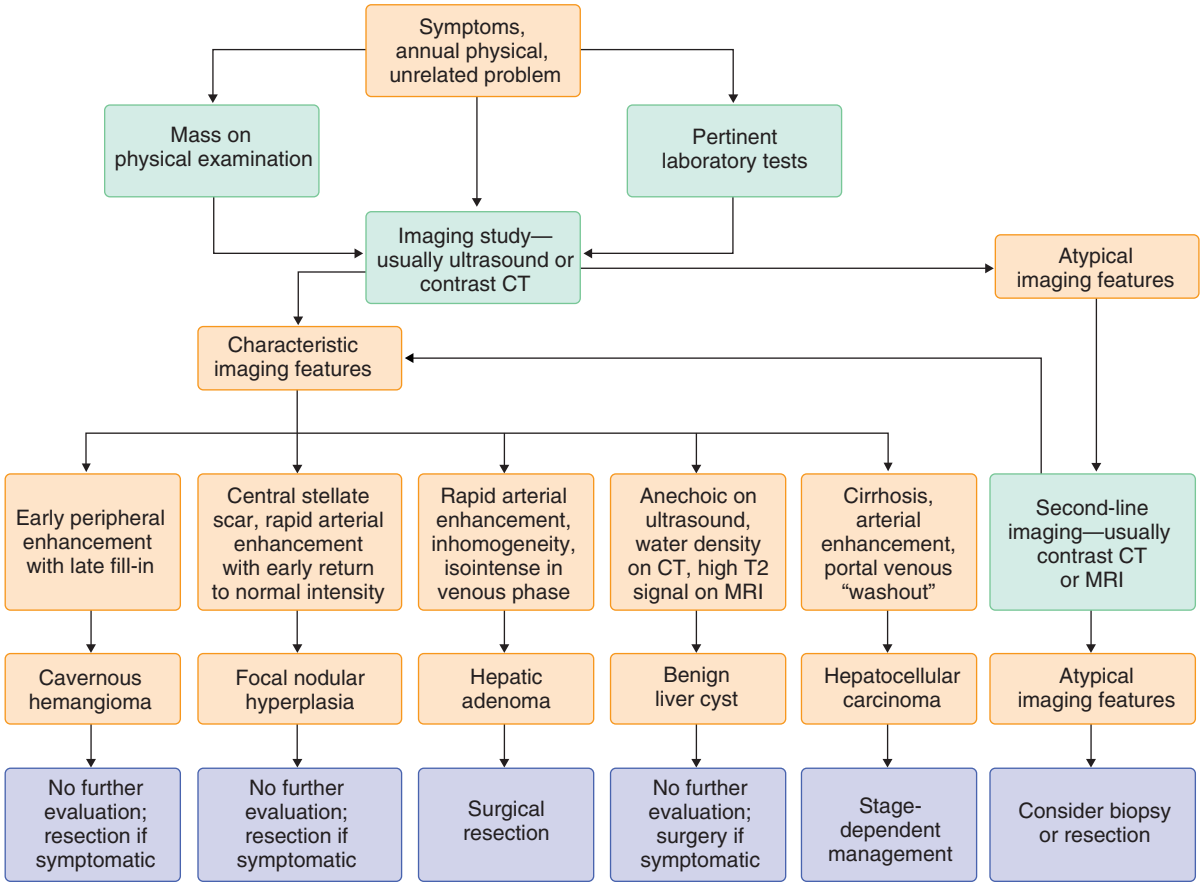

FIGURE 146-6. Approach to evaluating the patient with a mass lesion in the liver. Flow chart showing an algorithm for evaluating and managing common liver mass lesions. CT = computed tomography; MRI = magnetic resonance imaging. (From Roberts LR. Liver and biliary tract tumors. In: Goldman L, Schafer AI, eds. *Goldman's Cecil Medicine*, 24th ed. Philadelphia: Saunders; 2012.)

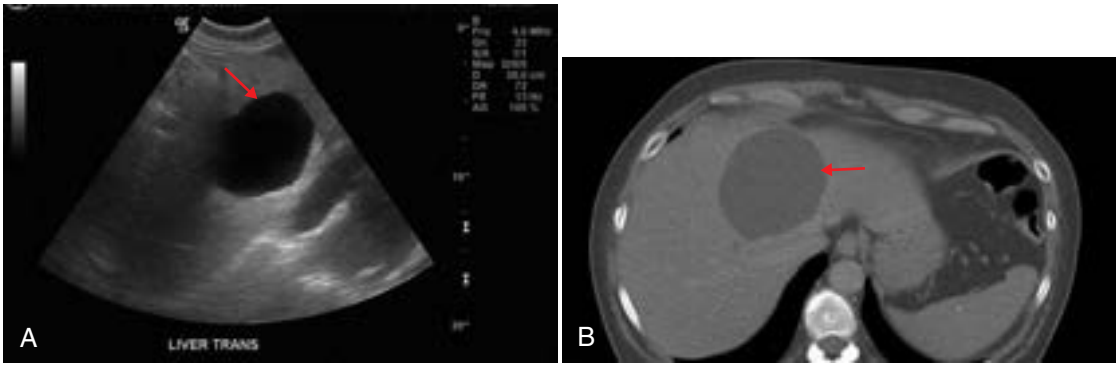

FIGURE 146-7. Simple hepatic cyst. A, Ultrasound demonstrates clear contents and an imperceptible wall (*arrow*). **B,** On a computed tomographic scan, a simple hepatic cyst is characterized by the lack of septation and an imperceptible wall (*arrow*). (Courtesy Dr. B. Madrazo.)

but become anechoic if hemorrhage has occurred. Contrast CT and MRI show arterial enhancement and varying signal intensity reflecting steatosis and hemorrhage if present. Biopsy may be needed for diagnosis. Potential complications of adenomas, especially if greater than 5 cm, include spontaneous hemorrhage and rupture, as well as rare malignant transformation. Discontinuation of oral contraceptives may lead to a decrease in size. Prophylactic hepatic resection is indicated for adenomas greater than 5 cm.

Focal nodular hyperplasia, which is a benign lesion characterized by a central scar and is thought to represent a hyperplastic response due to a vascular malformation. Identification of the scar by ultrasound, CT, or MRI is diagnostic (see Fig. 146-8D). Observation in the absence of symptoms is advised.

Nodular regenerative hyperplasia is characterized by multiple regenerative 1- to 3-mm nodules clustered around portal triads. It is associated with a variety of systemic disorders, predominantly autoimmune, including rheumatoid arthritis, systemic lupus erythematosus, and polymyalgia rheumatica, neoplastic such as myeloproliferative syndromes, and medications including thiopurines. Symptoms, if present, reflect associated portal hypertension. Diagnosis is by biopsy, which shows hypertrophied hepatocytes clustered around the center of the nodules with peripheral atrophy. Management is directed to the associated portal hypertension (Chapter 153).

Hepatocellular carcinoma (Chapter 196) is a frequent complication of cirrhosis of any cause as well as chronic hepatitis B infection even in the absence of cirrhosis. Patients usually will have evidence of cirrhosis with hepatic dysfunction. Because surveillance for hepatocellular carcinoma is now recommended in individuals at risk, this tumor is increasingly identified when asymptomatic and potentially curable.[10] On ultrasound, hepatocellular carcinoma appears solid. Owing to its predominantly arterial blood supply compared with that of normal hepatic parenchyma, intravenous contrast is

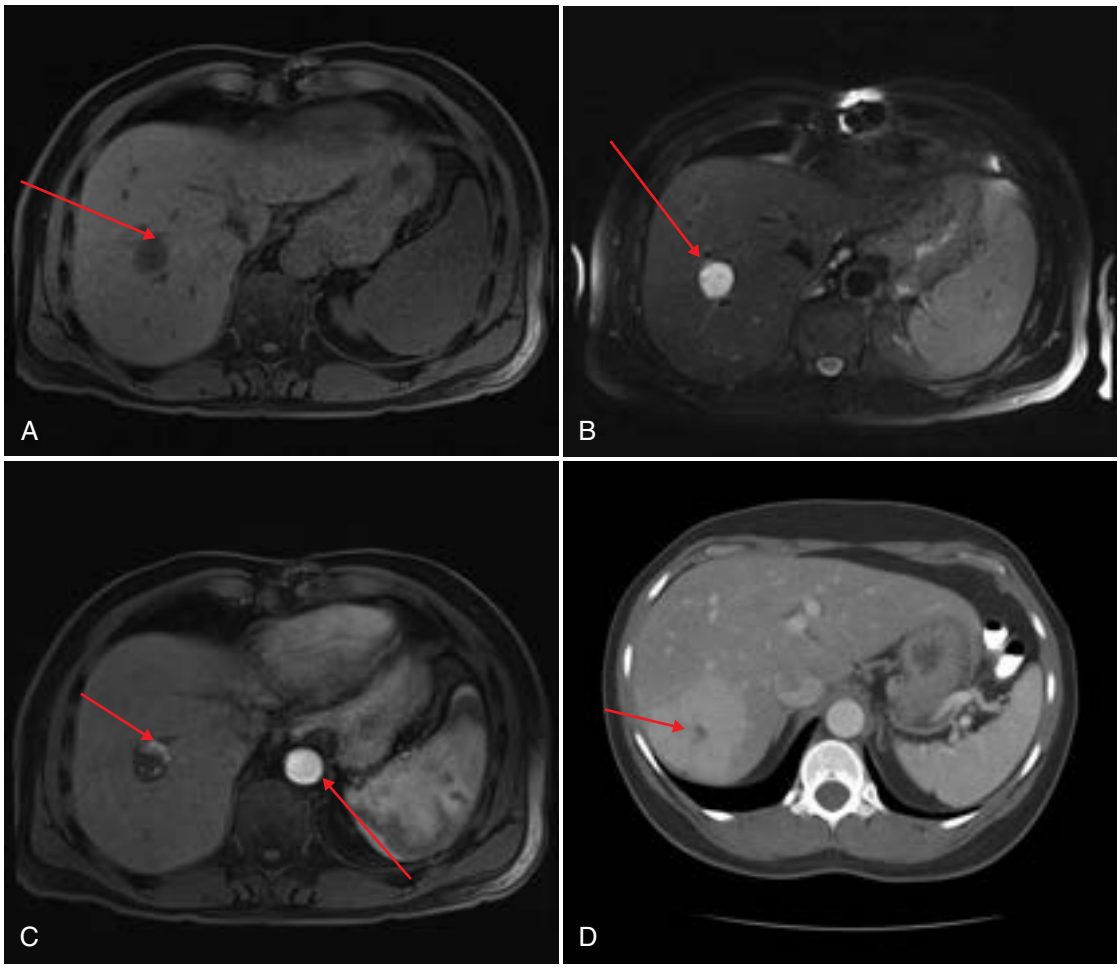

FIGURE 146-8. Solid hepatic mass. A, Hemangioma seen as a solid mass (*arrow*) on the T1 phase of a magnetic resonance (MR) imaging scan. **B,** Hemangioma on MR T2 image is characterized by a high signal density (*arrow*). **C,** A hemangioma fills slowly (*arrow*) from the periphery after intravenous contrast is administered (*arrow*). **D,** Focal nodular hyperplasia is characterized by a central scar (*arrow*) in the late venous phase of a contrast computed tomographic scan. (Courtesy Dr. B. Madrazo.)

rapidly taken up by a hepatic cellular carcinoma and then "washes out" on CT scan or MR imaging (see Fig. 196-5). Serum α-fetoprotein is elevated in approximately 60% of cases. Therapeutic options depend on the tumor's size and severity of associated liver disease.

APPROACH TO THE PATIENT WITH INHERITED LIVER DISEASE

The major inherited liver diseases in adults are hemochromatosis, Wilson disease, cystic fibrosis, and α_1-antitrypsin deficiency. These disorders may be suspected by their extrahepatic manifestations or may be detected during evaluation of unexplained hepatic dysfunction.

Hereditary hemochromatosis (Chapter 212), which is the most common inherited liver disorder, is most frequently due to mutations in the *HFE* gene (located on chromosome 6p22.2), typically at C282Y and H63D. Approximately 1 in 250 individuals of Northern European descent are homozygous for C282Y, the major *HFE* mutation, although only a minority develop phenotypic manifestations.[11] Liver blood test abnormalities, if present, are nonspecific. Definitive diagnosis is confirmed by detection of C282Y homozygote or compound heterozygote (C282Y/H63D). MRI or CT scan can confirm excessive hepatic iron. Liver biopsy is unnecessary for diagnosis but is advised for a serum ferritin greater than 1000 μg/L because the patient is likely to be cirrhotic. Iron accumulates in hepatocytes but not Kupffer cells in a periportal to pericentral gradient (Fig. 146-9). By comparison, secondary iron overload accumulates in Kupffer cells in conditions such as ineffective erythropoiesis. Quantitative iron measurement in liver tissue, calculated as the μmol of iron/g dry weight liver/age, can confirm the diagnosis. Treatment of hemochromatosis is by phlebotomy (Chapter 212).

α_1-Antitrypsin deficiency, which results from inherited mutations in this serine protease inhibitor, prevents its export from hepatocytes, where its accumulation causes liver inflammation, fibrosis, and cirrhosis. Deficiency of circulating α_1-antitrypsin results in panlobular emphysema (Chapter 88) even in nonsmokers. The disease is most frequent in persons of European ancestry. Inheritance is autosomal recessive with codominant inheritance because each allele provides 50% of circulating α_1-antitrypsin. A large number of alleles have been identified in the proteinase inhibitor (Pi) gene,

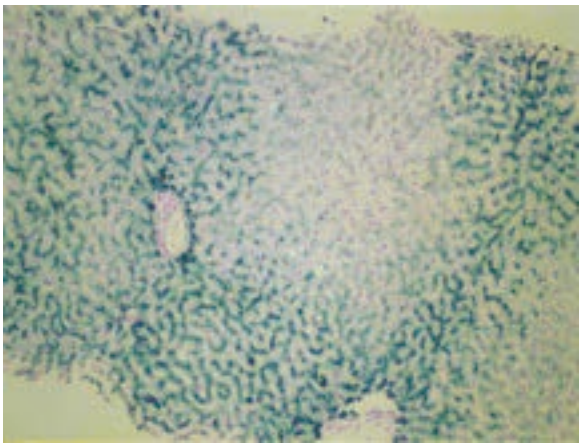

FIGURE 146-9. Liver biopsy in hemochromatosis (Perls Prussian blue stain for iron). Iron deposition is found in a periportal distribution, predominantly in parenchymal cells (hepatocytes). (From Bacon BR. Inherited and metabolic disorders of the liver. In: Goldman L, Schafer AI, eds. *Goldman's Cecil Medicine,* 24th ed. Philadelphia: Saunders; 2012.

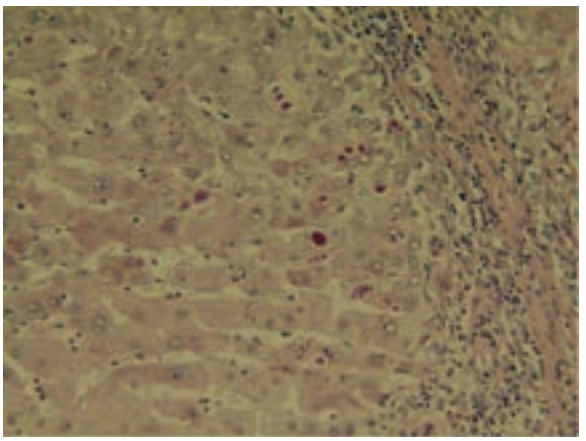

FIGURE 146-10. Liver biopsy in α_1-antitrypsin deficiency. On a periodic acid–Schiff, diastase-resistant stain, α_1-antitrypsin globules, with a characteristic magenta color, are found at the periphery of the lobule. (From Bacon BR. Inherited and metabolic disorders of the liver. In: Goldman L, Schafer AI, eds. *Goldman's Cecil Medicine*, 24th ed. Philadelphia: Saunders; 2012.)

with MM associated with normal circulating α_1-antitrypsin levels and the most common abnormal alleles, S and Z, associated with reduced levels. PiZZ homozygotes, who have the most severe α_1-antitrypsin deficiency, have only 15% of normal circulating levels and are prone to liver and lung disease. Liver disease has a bimodal distribution, with neonatal hepatitis and cholestatic jaundice manifesting in early life and cirrhosis in adulthood. Only approximately one third of adults with the PiZZ phenotype develop cirrhosis, thereby implying that other cofactors are involved in disease progression.[12] Other phenotypes with α_1-antitrypsin deficiency (e.g., PiMZ or PiSZ) increase the risk for liver disease, whereas null variants have lung but not liver disease. The condition may be suspected by a low α_1-1 globulin level on a routine protein electrophoresis, and the diagnosis can be confirmed by demonstration of a low serum α_1-antitrypsin level followed by Pi testing to determine the phenotype. Liver function tests are generally nonspecific, although the albumin level can be low and the INR prolonged if cirrhosis is present. Liver biopsy, which shows the extent of fibrosis, is characterized by periodic acid–Schiff (PAS)-positive, diastase-resistant globules in the periphery of the lobule (Fig. 146-10). No specific therapy is available for the liver disease, although weekly intravenous α_1-antitrypsin protein concentrate infusions, which are approved by the U.S. Food and Drug Administration, might prevent progression of lung disease by restoring serum and alveolar α_1-antitrypsin levels.

Wilson disease (Chapter 211) is excessive copper accumulation in liver and extrahepatic tissues owing to an inherited disorder in copper transport from hepatocytes to bile.[13] Hepatic manifestations include chronic hepatitis and cirrhosis, most typically appearing by the fourth decade of life, although they can manifest later. A characteristic finding in a fulminant manifestation of Wilson disease is hemolysis with an abnormally low alkaline phosphatase level, jaundice, and coagulopathy. More typically, patients have elevated aminotransferase levels without hemolysis at presentation. Diagnosis usually is based on an elevated 24-hour urinary copper level and low serum ceruloplasmin levels, but in unclear cases the quantification of increased hepatic copper content by biopsy is definitive. Copper chelation prevents disease progression.

Cystic fibrosis (Chapter 89) is usually dominated by pulmonary complications, but liver disease is an important cause of morbidity and mortality in adults.[14] In patients with elevated alkaline phosphatase and bilirubin levels, biliary cirrhosis (Chapter 155) can develop owing to abnormal bile viscosity. Ursodeoxycholic acid may improve biochemical test results but has no clear benefit on outcome. Liver transplantation (Chapter 154) may be necessary for decompensated cirrhosis.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Lee TH, Kim WR, Poterucha JJ. Evaluation of elevated liver enzymes. *Clin Liver Dis*. 2012;16:183-198.
2. Younossi ZM, Stepanova M, Afendy M, et al. Changes in the prevalence of the most common causes of chronic liver diseases in the United States from 1988 to 2008. *Clin Gastroenterol Hepatol*. 2011;9:524-530.
3. Bunchorntavakul C, Reddy KR. Review article: herbal and dietary supplement hepatotoxicity. *Aliment Pharmacol Ther*. 2013;37:3-17.
4. Gupta K, Dhawan A, Abel C, et al. A re-evaluation of the scratch test for locating the liver edge. *BMC Gastroenterol*. 2013;13:35.
5. Pal G, Lin MM, Laureno R. Asterixis: a study of 103 patients. *Metab Brain Dis*. 2014;29:813-824.
6. Cassinotto C, Lapuyade B, Mouries A, et al. Noninvasive assessment of liver fibrosis with impulse elastography: comparison of supersonic shear imaging with ARFI and Fibroscan. *J Hepatol*. 2014;61:550-557.
7. Wills M, Harvey CJ, Kuzmich S, et al. Characterizing benign liver lesions and trauma with contrast-enhanced ultrasound. *Br J Hosp Med (Lond)*. 2014;75:91-95.
8. Wills M, Harvey CJ, Kuzmich S, et al. Characterizing malignant liver lesions with contrast-enhanced ultrasound. *Br J Hosp Med (Lond)*. 2014;75:151-154.
9. Lantinga MA, Gevers TJ, Drenth JP. Evaluation of hepatic cystic lesions. *World J Gastroenterol*. 2013;19:3543-3554.
10. Hennedige T, Venkatesh SK. Imaging of hepatocellular carcinoma: diagnosis, staging and treatment monitoring. *Cancer Imaging*. 2013;12:530-547.
11. Wood MJ, Powell LW, Dixon JL, et al. Clinical cofactors and hepatic fibrosis in hereditary hemochromatosis: the role of diabetes mellitus. *Hepatology*. 2012;56:904-911.
12. Clark VC, Dhanasekaran R, Brantly M, et al. Liver test results do not identify liver disease in adults with alpha(1)-antitrypsin deficiency. *Clin Gastroenterol Hepatol*. 2012;10:1278-1283.
13. Kanwar P, Kowdley KV. Metal storage disorders: Wilson disease and hemochromatosis. *Med Clin North Am*. 2014;98:87-102.
14. Rowland M, Gallager C, Gallagher CG, et al. Outcome in patients with cystic fibrosis liver disease. *J Cyst Fibros*. 2015;14:120-126.

REVIEW QUESTIONS

1. Fever is a frequent feature of which of the following?

A. Acute alcoholic hepatitis
B. Acute hepatitis C
C. Autoimmune hepatitis
D. Chronic hepatitis B
E. Chronic hepatitis C

Answer: A With the exception of the prodrome of acute hepatitis A, fever is not prominent in viral hepatitis nor in autoimmune hepatitis. It is a feature of acute alcoholic hepatitis even in the absence of infection.

2. Which of following suggests biliary obstruction as the cause of scleral icterus?

A. Dark urine and dark stool
B. Pale urine and pale stool
C. Dark urine and pale stool
D. Dark urine and alternating pale and dark stool
E. Pale urine and dark stool

Answer: C With jaundice from biliary obstruction, bilirubin is predominantly conjugated, water-soluble, and excreted in the urine, which is therefore dark. Because of biliary obstruction, bile salts cannot reach the intestine, so stools are pale. By comparison, the excess bilirubin from hemolysis is unconjugated and not water soluble, so it is not excreted in the urine.

3. Which of the following presentations suggests variceal hemorrhage?

A. Unexplained iron deficiency anemia
B. Dysphagia
C. Abdominal bloating
D. Abdominal pain
E. Hematemesis

Correct: E Variceal hemorrhage is typically profuse, painless, and not associated with esophageal symptoms such as dysphagia. Because bleeding is overt, unexplained iron deficiency is not usually due to bleeding varices.

4. Which of the following disorders has an autosomal dominant inheritance?

A. Wilson disease
B. Hemochromatosis
C. α_1-Antitrypsin deficiency
D. Adult polycystic disease
E. Hepatic adenoma

Correct: D Adult polycystic disease has an autosomal dominant inheritance, whereas the other conditions are sporadic, such as adenoma, or autosomal recessive, such as Wilson disease.

147

APPROACH TO THE PATIENT WITH JAUNDICE OR ABNORMAL LIVER TESTS

PAUL D. BERK AND KEVIN M. KORENBLAT

JAUNDICE AND HYPERBILIRUBINEMIA

DEFINITION

Jaundice, from the French *jaune* ("yellow"), is the yellow-orange discoloration of the skin, conjunctivae, and mucous membranes that results from an elevated plasma bilirubin level. Mild hyperbilirubinemia may be clinically undetectable, but jaundice becomes evident at plasma bilirubin greater than 3 to 4 mg/dL, depending on the patient's normal skin pigmentation, the conditions of observation, and the bilirubin fraction that is elevated. Hyperbilirubinemia may result from hepatocellular dysfunction (pure hyperbilirubinemia) or from either increased bilirubin production or inherited or acquired defects in specific aspects of hepatic bilirubin disposition.

PATHOBIOLOGY

Bilirubin Metabolism

Bilirubin Production

Bilirubin is the degradation product of the heme moiety of hemoproteins, a class of proteins involved in the transport or metabolism of oxygen (Fig. 147-1). Normal adults produce approximately 4 mg of bilirubin per kilogram of body weight per day. Between 70 and 90% of bilirubin is derived from the hemoglobin of erythrocytes, which are sequestered and destroyed by the mononuclear phagocytic cells of the reticuloendothelial system, principally in the spleen, liver, and bone marrow. The remainder results primarily from the turnover of nonhemoglobin hemoproteins such as myoglobin, the P-450 cytochromes, catalase, and peroxidase, principally in the liver; a minor fraction reflects ineffective erythropoiesis, which is the premature destruction of newly formed erythrocytes within the bone marrow. Although bilirubin production occurs principally in the liver, bilirubin has proved to have antioxidant properties, and recent studies suggest that a limited, regulated production of bilirubin from heme may occur in many cell types and contribute to regulating the intracellular antioxidant environment.[1]

The two-step conversion of heme to bilirubin begins with the opening of the heme molecule at its α bridge carbon by the microsomal enzyme *heme oxygenase,* a process that results in the formation of equimolar quantities of carbon monoxide and the green tetrapyrrole biliverdin. This nontoxic, water-soluble pigment is the main excretory product of heme in birds, reptiles, and amphibians. However, biliverdin cannot cross the placenta. Accordingly, its reduction to bilirubin in mammals by a second enzyme, *biliverdin reductase,* allows its transplacental removal from the fetus into the maternal circulation. The unconjugated bilirubin produced in the periphery is transported to the liver in plasma. Because of its insolubility in aqueous media, it is kept in solution by tight but reversible binding to albumin. A number of compounds, including sulfonamides, furosemide, and radiographic contrast agents, can competitively displace bilirubin from its binding sites on albumin, a phenomenon that is of little clinical significance except in neonates, in whom the resulting increased concentration of unbound bilirubin raises the risk for kernicterus.

Disposition of Bilirubin by the Liver

Excretion of bilirubin from the body is a major function of the liver (see Fig. 147-1), where the specialized microanatomy enhances the extraction of tightly protein-bound compounds from the circulation. Hepatic translocation of bilirubin from blood to bile involves four distinct steps: (1) uptake of unconjugated bilirubin, principally by an incompletely characterized facilitated transport process and to a lesser extent by diffusion; (2) intracellular binding, mainly to various cytosolic proteins of the glutathione-*S*-transferase family; (3) conversion of unconjugated bilirubin to bilirubin monoglucuronides and diglucuronides by a specific uridine 5′-diphosphoglucuronosyltransferase (UDP-glucuronosyltransferase) isoform designated UGT1A1, encoded by the *UGT1* gene complex; and (4) transfer of bilirubin monoglucuronides and diglucuronides into bile by a canalicular membrane

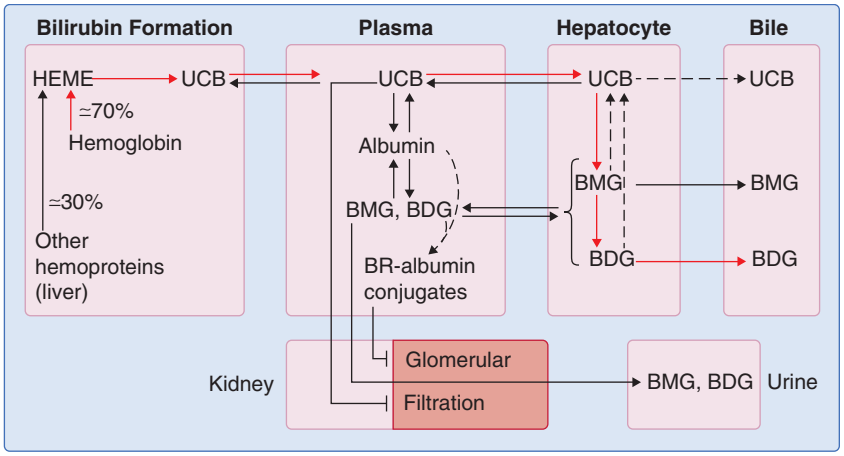

FIGURE 147-1. Overview of bilirubin metabolism. Unconjugated bilirubin (UCB) formed from the breakdown of heme from hemoglobin and other hemoproteins is transported in plasma reversibly bound to albumin and is converted in the liver to bilirubin monoglucuronide (BMG) and diglucuronide (BDG), the latter being the predominant form secreted in bile. BMG and BDG together normally account for less than 5% of normal serum bilirubin. In patients with hepatobiliary disease, BMG and BDG accumulate in plasma and appear in urine. Bilirubin glucuronides in plasma also react nonenzymatically with albumin and possibly other serum proteins to form protein conjugates, which do not appear in urine and have a plasma half-life similar to that of albumin. BR = bilirubin.

adenosine triphosphate (ATP)-dependent transporter designated multidrug resistance–associated protein 2 (MRP2) or canalicular multispecific organic anion transporter (cMOAT). MRP2/cMOAT is encoded by the gene *ABCC2*, a member of ATP-binding cassette (ABC) transporter superfamily of genes. ABC genes transport molecules across intracellular and extracellular membranes. The superfamily is divided into seven subfamilies. MRP2/cMOAT belongs to the MRP family, whose substrates include drug conjugates and unmodified anticancer drugs, which are pumped out of cells.

Conjugation of unconjugated bilirubin to bilirubin monoglucuronides and diglucuronides is a critical process that greatly increases the aqueous solubility of bilirubin, thereby enhancing its elimination from the body while simultaneously reducing its ability to diffuse across biologic membranes, including the blood-brain barrier. In newborn infants, a decreased capacity to conjugate bilirubin leads to unconjugated hyperbilirubinemia (physiologic jaundice of the newborn). If severe, this hyperbilirubinemia may lead to irreversible central nervous system toxicity. Phototherapy by exposure to blue light converts bilirubin to water-soluble photoisomers that are readily excreted in bile, thereby protecting the central nervous system from bilirubin toxicity. Gilbert syndrome and Crigler-Najjar syndrome types 1 and 2, which result from genetic defects in bilirubin conjugation, are characterized by unconjugated hyperbilirubinemia; by contrast, Dubin-Johnson syndrome, which results from inheritable defects in MRP2/cMOAT (see later), is characterized by conjugated or mixed hyperbilirubinemia. The recent discovery of the causative mutations in Rotor syndrome, which is also characterized by conjugated or mixed hyperbilirubinemia, has led to the finding of an additional step in bilirubin transport, characterized by the export of intracellular bilirubin conjugates across the sinusoidal membrane and their subsequent reuptake by transporters in downstream hepatocytes. The process, which is also involved in the disposition of drug metabolites, is believed to prevent the local saturation of upstream hepatocytes with bilirubin and drug conjugates.[2]

Enterohepatic Circulation and Excretion of Bilirubin

Normal human bile contains an average of less than 5% unconjugated bilirubin, 7% bilirubin monoconjugates, and 90% bilirubin diconjugates. Following canalicular secretion, conjugated bilirubin passes down the gastrointestinal tract without reabsorption by either the gallbladder or intestinal mucosa. Although some bilirubin reaches the feces, most is converted to urobilinogen and to related compounds by bacteria within the ileum and colon, where the urobilinogen is reabsorbed, returns to the liver through the portal circulation, and is re-excreted into bile in a process of enterohepatic recirculation. Any urobilinogen not taken up by the liver reaches the systemic circulation, from which it is cleared by the kidneys. Normal urine urobilinogen excretion is 4 mg/day or less. With hemolysis, which increases the load of bilirubin entering the gut and therefore the amount of urobilinogen formed and reabsorbed, or with liver disease, which decreases its hepatic extraction, plasma urobilinogen levels rise and more urobilinogen is excreted in the urine. Severe cholestasis, bile duct obstruction, or antibiotics that reduce or eliminate the bacterial conversion of bilirubin to urobilinogen markedly decrease the formation and urinary excretion of urobilinogen.

Unconjugated bilirubin ordinarily does not reach the gut except in neonates or, by ill-defined alternative pathways, in the presence of severe unconjugated hyperbilirubinemia (e.g., Crigler-Najjar type 1). In these circumstances, unconjugated bilirubin is reabsorbed from the gut, thereby amplifying the hyperbilirubinemia.

Measurement of Bilirubin in Plasma

The total plasma bilirubin concentration in normal adults is less than 1 to 1.5 mg/dL, depending on the measurement method. Modern analytic techniques show that normal plasma contains principally unconjugated bilirubin, with only a trace of conjugated bilirubin. Clinical laboratories typically quantify plasma bilirubin by a reaction in which bilirubin is cleaved by a diazo reagent, such as diazotized sulfanilic acid, to azodipyrroles that are readily quantitated spectrophotometrically. Bilirubin conjugates react rapidly ("prompt" or "direct"-reacting bilirubin). Unconjugated bilirubin reacts slowly because the site of attack by the diazo reagent is protected by internal hydrogen bonding. Accordingly, accurate measurement of the total plasma bilirubin concentration requires addition of an accelerator, such as ethanol or urea, to disrupt this internal hydrogen bonding and to ensure complete reaction of any unconjugated bilirubin.

The "indirect"-reacting bilirubin is calculated by subtracting the direct-reacting bilirubin from the total. Although physicians traditionally equate the direct-reacting fraction of bilirubin in plasma with conjugated bilirubin and the indirect fraction with unconjugated bilirubin, this approach is, at best, a rough approximation. The unqualified interpretation of direct and indirect fractions as reflecting conjugated and unconjugated bilirubin, respectively, may lead to diagnostic errors, particularly in the diagnosis of hereditary hyperbilirubinemias. In common practice, 10 to 20% of the bilirubin in normal plasma gives a prompt (direct) diazo reaction even though more than 95% of total bilirubin in normal plasma is unconjugated. Thus, at virtually any total bilirubin concentration, a direct-reacting fraction of less than 15% of the total bilirubin can be considered as essentially all unconjugated. When the direct-reacting fraction is greater than 15%, a simple dipstick test for bilirubinuria may clarify the situation. Unconjugated bilirubin is not excreted in urine regardless of the height of its plasma concentration because its binding to albumin is too tight for effective glomerular filtration and it is not secreted by the tubules. The canalicular transport mechanism for excretion of bilirubin conjugates is especially sensitive to injury. Accordingly, in parenchymal liver disease or mechanical bile duct obstruction, bilirubin conjugates within the hepatocyte or biliary tract may reflux into the blood stream, resulting in a mixed or, less often, a purely conjugated hyperbilirubinemia. Conjugated bilirubin, which is normally loosely bound to albumin, is readily filtered at the glomerulus; even modest degrees of conjugated hyperbilirubinemia result in bilirubinuria, which is *always* a pathologic finding. With prolonged conjugated hyperbilirubinemia, some of the conjugated bilirubin binds *covalently* to albumin and produces what is designated the δ-bilirubin fraction. Although δ-bilirubin gives a direct diazo reaction, it is not filterable by the glomerulus and does not appear in the urine; it disappears slowly from the plasma, with the 14- to 21-day half-life of the albumin to which it is

bound. δ-Bilirubin can account for the slow rate at which conjugated (direct) hyperbilirubinemia sometimes resolves as hepatitis improves or biliary obstruction is relieved. Although δ-bilirubin is not easily measured, its presence can be inferred when an elevated direct-reacting bilirubin persists after bilirubinuria resolves.

Transcutaneous bilirubinometry is an alternative method for measuring bilirubin in infants with neonatal jaundice. The method is based on the measurement of the light reflected from the percutaneous transmission of visible light and is conceptually analogous to pulse oximetry.[3]

Bilirubin Kinetics

The plasma unconjugated bilirubin concentration ([UCB]) is determined by a balance between the bilirubin production rate (BRP) and hepatic bilirubin clearance (C_{BR}) according to the relationship:

$$[UCB] \approx BRP/C_{BR}$$

C_{BR} is analogous to the creatinine clearance test of kidney function; it is a measure of the rate at which bilirubin is extracted from plasma, and it is a true quantitative test of liver function. Whereas BRP and C_{BR} are not easily quantified clinically, investigative measurements have yielded useful pathophysiologic insights. This equation indicates that [UCB] is directly proportional with increases in BRP and inversely proportional with CBR, thereby providing a basis for classifying unconjugated hyperbilirubinemias according to their pathogenesis.

Increased Bilirubin Production

An increased production of bilirubin and a resulting unconjugated hyperbilirubinemia can be caused by hemolysis, an accelerated destruction of transfused erythrocytes, resorption of hematomas, or ineffective erythropoiesis owing to lead poisoning, megaloblastic anemias related to deficiency of either folic acid or vitamin B_{12}, sideroblastic anemia, congenital erythropoietic porphyria, or myeloproliferative or myelodysplastic diseases. In these settings, other liver tests are typically normal and the hyperbilirubinemia is modest, rarely exceeding 4 mg/dL; higher values imply concomitant hepatic dysfunction. However, after brisk blood transfusion or resorption of massive hematomas caused by trauma, the increased bilirubin load may be transiently sufficient to lead to frank jaundice. The causes of hemolysis are numerous (Chapters 160 to 163). Besides specific blood disorders, mild hemolysis accompanies many acquired diseases. In the setting of systemic disease, which may include a degree of hepatic dysfunction, hemolysis may produce a component of conjugated hyperbilirubinemia in addition to an elevated unconjugated bilirubin concentration. Prolonged hemolysis may lead to the formation of bilirubin gallstones, which may cause biliary tract disease (Chapter 155).

Decreased Hepatic Bilirubin Clearance

Decreased Bilirubin Uptake

Kinetic studies suggest that hepatocellular bilirubin uptake involves both facilitated and diffusive components. Several drugs (e.g., rifampin, novobiocin, and various cholecystographic contrast agents) competitively inhibit the hepatocellular uptake of bilirubin, thereby suggesting the existence of a component of facilitated uptake. Decreased hepatic bilirubin uptake is also believed to contribute to the unconjugated hyperbilirubinemia of Gilbert syndrome, although the principal molecular basis for that syndrome is a reduction in bilirubin conjugation. The identity of the bilirubin transporter remains controversial.

Impaired Bilirubin Conjugation

The most frequent cause of decreased bilirubin clearance is a decrease in bilirubin conjugating activity. Bilirubin conjugation with glucuronic acid is catalyzed principally by a specific UDP-glucuronosyltransferase, which is designated UGT1A1 and encoded by the *UGT1* gene complex. The *UGT1A1* gene is assembled by alternative splicing of a bilirubin-specific variant of exon 1, designated exon A_1, with four common exons (exons 2 to 5) that encode the shared carboxyl terminal end of all *UGT1*-encoded proteins. Its promoter region normally contains an $A(TA)_6TAA$ TATA box–like construct.

APPROACH TO THE PATIENT WITH HYPERBILIRUBINEMIA

Hyperbilirubinemia and jaundice (see Fig. 146-1) may result from isolated disorders of bilirubin metabolism, liver disease, or obstruction of the biliary tract. Jaundice represents the most visible sign of hepatobiliary disease of many causes (Table 147-1).

Genetic Disorders of Bilirubin Conjugation

The hereditary hyperbilirubinemias (Table 147-2) are a group of five syndromes in which hyperbilirubinemia occurs as an isolated biochemical abnormality, without evidence of either hepatocellular necrosis or cholestasis.[4]

CRIGLER-NAJJAR AND GILBERT SYNDROMES

Crigler-Najjar syndrome types 1 and 2 and Gilbert syndrome are hereditary forms of unconjugated hyperbilirubinemia that have been known for more than two decades to result principally from mutations in *UGT1A1*. In Crigler-Najjar type 1, essentially no functional enzyme activity is present, whereas patients with Crigler-Najjar type 2 have up to 10% of normal and patients with Gilbert syndrome have 10 to 33% of normal activity, leading to bilirubin concentrations of 18 to 45, 6 to 25, and 1.5 to 4 mg/dL, respectively (see Table 147-2). Because total UGT1A1 enzymatic activity must be reduced to less than 50% of normal to produce unconjugated hyperbilirubinemia, phenotypic expression of mutations in this enzyme requires either homozygosity or double heterozygosity, and each of these disorders is inherited as an autosomal recessive trait. Patients with Crigler-Najjar types 1 and 2 are either homozygotes or double heterozygotes for structural mutations within the coding region. In Western countries, patients with Gilbert syndrome are typically homozygous for an $A(TA)_7TAA$ promoter mutation; this polymorphism is designated *UGT1A1*28*. Structural mutations in exon 1 of UGT1A1 causing modest reductions in UGT1A1 enzymatic activity have been reported in some Japanese patients with Gilbert syndrome. To date, 130 different mutations in *UGT1A1* associated with hereditary hyperbilirubinemia have been identified, including 59 linked to Crigler-Najjar type 1 and 48 to Crigler-Najjar type 2.[5] Recent studies have identified mutations of multiple additional regions of *UGT1A* and also of the solute carrier protein SCLO1B1, both of which participate to varying degrees in the defective glucuronidation capacity of these three syndromes.[6] The resulting complex genotypes may pose a risk not only for hyperbilirubinemia but also for susceptibility to glucuronidation-associated drug toxicity and, perhaps, for subtle diverse benefits as well.[7]

Crigler-Najjar Syndrome Type 1

Crigler-Najjar type 1 is characterized by striking unconjugated hyperbilirubinemia that appears in the neonatal period, persists for life, and is unresponsive to phenobarbital. The majority of patients (type 1A) exhibit defects in the glucuronide conjugation of a spectrum of substrates in addition to bilirubin as the result of mutations in one of the common exons (2 to 5) of the *UGT1* complex. In a smaller subset (type 1B), a mutation in the bilirubin-specific exon A1 limits the defect to bilirubin conjugation. Fifty-nine structurally diverse *UGT1A1* mutations can cause Crigler-Najjar type 1; their common feature is they all encode proteins with absent or, at most, traces of enzymatic activity. Before the availability of phototherapy, most patients with Crigler-Najjar type 1 died of bilirubin encephalopathy (kernicterus) in infancy or early childhood.[8] Optimal treatment for a neurologically intact patient includes (1) approximately 12 hours/day of phototherapy from birth throughout childhood, perhaps supplemented by exchange transfusion in the neonatal period; (2) use of tin-protoporphyrin to blunt transient episodes of increased hyperbilirubinemia; and (3) early liver transplantation, before the onset of brain damage.[9] Plasmapheresis has been used to temporarily control abrupt increases in bilirubin[10] and transplantation with isolated allogeneic hepatocytes has been evaluated as an experimental therapeutic approach.

Crigler-Najjar Syndrome Type 2

Bilirubin concentrations are typically lower in Crigler-Najjar type 2, and plasma bilirubin levels can be reduced to 3 to 5 mg/dL by phenobarbital. At least 48 different mutations of *UGT1A1* have been associated with Crigler-Najjar type 2; all encode a bilirubin-UDP-glucuronosyltransferase with markedly reduced but detectable enzymatic activity. Although uncommon in Crigler-Najjar type 2, kernicterus has occurred at all ages, typically associated with factors that temporarily raise the plasma bilirubin concentration above baseline (e.g., fasting, intercurrent illness). For this reason, phenobarbital therapy is often recommended; a single bedtime dose usually maintains clinically safe plasma bilirubin concentrations.

Gilbert Syndrome

Gilbert syndrome is the most common form of the hereditary hyperbilirubinemias, with a genotypic prevalence of approximately 12% and a

TABLE 147-1 DIFFERENTIAL DIAGNOSIS OF HYPERBILIRUBINEMIA AND JAUNDICE

ISOLATED DISORDERS OF BILIRUBIN METABOLISM
Unconjugated Hyperbilirubinemia
Increased bilirubin production
Examples: Hemolysis, ineffective erythropoiesis, blood transfusion, resorption of hematomas
Decreased hepatocellular uptake
Examples: Drugs (e.g., rifampin)
Decreased conjugation
Examples: Gilbert and Crigler-Najjar syndromes, physiologic jaundice of the newborn, breast milk jaundice, HIV protease inhibitors
Conjugated or Mixed Hyperbilirubinemia
Decreased canalicular transport: Dubin-Johnson syndrome
Decreased reuptake of bilirubin conjugates: Rotor syndrome
LIVER DISEASE
Acute or Chronic Hepatocellular Dysfunction
Acute or subacute hepatocellular injury
Examples: Viral hepatitis A, B, C, E; hepatotoxins (e.g., ethanol, acetaminophen, mushroom *[Amanita phalloides]* poisoning); drugs (e.g., isoniazid, α-methyldopa), metabolic diseases (e.g., Wilson disease, Reye syndrome); pregnancy-related (e.g., acute fatty liver of pregnancy, HELLP); hepatic ischemia (e.g., hypotension, postoperative, hepatic artery thrombosis)
Chronic hepatocellular disease
Examples: Hepatitis B, B + D, C, and E; hepatotoxins (e.g., vinyl chloride, vitamin A), nonalcoholic and alcoholic fatty liver disease; autoimmune hepatitis; metabolic disease (Wilson disease, hemochromatosis, α_1-antitrypsin deficiency)
Hepatic Disorders with Prominent Cholestasis
Familial cholestatic disorders
Single gene disorders
Examples: Benign recurrent intrahepatic cholestasis types 1-3; progressive familial intrahepatic cholestasis types 1-3
Familial cholestatic disorders of unknown pathogenesis
Examples: Aagenaes syndrome, Navajo neurohepatopathy, North American Indian cholestasis
Diffuse infiltrative disorders
Examples: Granulomatous diseases (e.g., mycobacterial and fungal infections, sarcoidosis, lymphoma, drugs), amyloidosis, infiltrative malignancies
Inflammation of intrahepatic bile ductules and/or portal tracts
Examples: Primary biliary cirrhosis, liver allograft rejection, graft-versus-host disease, drugs (e.g., chlorpromazine, erythromycin)
Miscellaneous conditions
Examples: Uncommon presentations of viral or alcoholic hepatitis, intrahepatic cholestasis of pregnancy, contraceptive jaundice, estrogens, anabolic steroids, postoperative cholestasis, cholestasis of sepsis, total parenteral nutrition, bacterial infections, drugs
OBSTRUCTION OF THE BILE DUCTS
Choledocholithiasis
Examples: Cholesterol gallstones, pigment gallstones
Diseases of the Bile Ducts
Inflammation, infection
Examples: Primary sclerosing cholangitis, AIDS cholangiopathy, hepatic arterial chemotherapy, postsurgical strictures
Neoplasms (e.g., cholangiocarcinoma)
Extrinsic Compression of the Biliary Tree
Neoplasms
Examples: Pancreatic carcinoma, metastatic lymphadenopathy, hepatoma
Pancreatitis with or without pseudocyst formation
Vascular enlargement (e.g., aneurysm, cavernous transformation of portal vein)

AIDS = acquired immunodeficiency syndrome; HELLP = hemolysis, elevated liver enzymes, and low platelets; HIV = human immunodeficiency syndrome.

phenotypic prevalence of approximately 7% in whites. Its high prevalence may explain the frequency of mild unconjugated hyperbilirubinemia in liver transplant recipients. Plasma bilirubin concentrations are most often less than 3 mg/dL, although both higher and lower values are frequent, with increases of two-fold to three-fold commonly occurring with fasting and intercurrent illness. The phenotypic distinction between mild Gilbert syndrome and a normal state is often blurred. Phenobarbital's ability to induce hepatic enzyme activity normalizes both the bilirubin concentration and C_{BR}. Oxidative drug metabolism and the disposition of many, but not all, xenobiotics that are metabolized by glucuronidation appear to be normal in Gilbert syndrome. A critical exception is the antitumor agent irinotecan (CPT-11), whose active metabolite (SN-38) is glucuronidated specifically by UGT1A1. In patients with Gilbert syndrome, CPT-11 can cause intractable diarrhea, myelosuppression, and other serious toxicities. Significant adverse events have not been described when individuals with Gilbert syndrome are prescribed many other agents that are metabolized by glucuronidation. Although Gilbert syndrome has no association with disease, occasional reports have demonstrated an association of the *UGT1A1*28* allele with decreased risk for cardiovascular and specific neoplastic diseases, an association that has been attributed to the vasodilatory and antioxidant effects of bilirubin and heme oxygenase.

UNCONJUGATED HYPERBILIRUBIN IN THE NEWBORN PERIOD

Most neonates develop mild unconjugated hyperbilirubinemia between days 2 and 5 after birth because of hepatic immaturity and low UGT1A1 levels. Peak bilirubin levels are typically less than 5 to 10 mg/dL, and levels return to normal within 2 weeks as mechanisms fostering bilirubin disposition mature. Prematurity, with hemolysis, is associated with higher bilirubin levels that may require phototherapy. The progestational steroid 3α,20β-pregnanediol and certain fatty acids that are found in breast milk (but not serum) of some mothers inhibit bilirubin conjugation and can cause excessive neonatal hyperbilirubinemia *(breast milk jaundice)*. By comparison, a UGT1A1 inhibitor, which is found in maternal serum, causes *transient familial neonatal hyperbilirubinemia* (Lucey-Driscoll syndrome).

CONJUGATED OR MIXED HYPERBILIRUBINEMIA

Two phenotypically similar but mechanistically distinct inherited disorders, Dubin-Johnson syndrome and Rotor syndrome, are characterized by conjugated or mixed hyperbilirubinemia with normal values for other standard liver tests (see Table 147-2). Dubin-Johnson syndrome results from any of several mutations in the gene encoding the ATP-dependent canalicular organic anion transporter MPR2/cMOAT (see Fig. 147-1). Individuals with Rotor syndrome exhibit the simultaneous disruption of two genes that code for the organic anion transporting polypeptides OATP1B1 and OATP1B3, which mediate the reuptake by downstream hepatocytes of conjugated bilirubin secreted into the sinusoid via the sinusoidal efflux transporter ABCC3 in more upstream hepatocytes.[11] Despite the conjugated hyperbilirubinemia, patients with Rotor syndrome are not cholestatic and can be distinguished noninvasively both from normal subjects and persons with Dubin-Johnson syndrome by their two-fold to five-fold increase in total copropophryin

TABLE 147-2 PRINCIPAL FEATURES OF THE HEREDITARY DISORDERS OF BILIRUBIN METABOLISM

FEATURE	CRIGLER-NAJJAR SYNDROME *Type I*	CRIGLER-NAJJAR SYNDROME *Type II*	GILBERT SYNDROME	DUBIN-JOHNSON SYNDROME	ROTOR SYNDROME
Incidence	Very rare	Uncommon	Up to 12% of population	Uncommon	Rare
Total serum bilirubin (mg/dL)	18-45 (usually >20), unconjugated	6-25 (usually ≤20), unconjugated	Typically ≤4 in absence of fasting or overt hemolysis; mostly unconjugated	Typically 2-5, less often ≤25; approximately 60% direct reacting	Usually 3-7, occasionally ≤20; approximately 60% direct reacting
Defect(s) in bilirubin metabolism*	Bilirubin UGT1A1 conjugating activity markedly reduced: trace to absent	Bilirubin UGT1A1 conjugating activity reduced: ≤10% of normal.	Complex haplotype Bilirubin UGT1A1 conjugating activity typically reduced to 10-33% of normal; reduced bilirubin uptake in some cases; mild hemolysis in up to 50% of patients	Impaired canalicular secretion of conjugated bilirubin owing to MRP2/cMOAT mutation	Impaired hepatic secretion or storage of conjugated bilirubin owing to impaired reuptake of bilirubin conjugates resulting from simultaneous OATP1B1 and OATP1B3 mutations.
Routine liver tests	Normal	Normal	Normal	Normal	Normal
Serum bile acids	Normal	Normal	Normal	Usually normal	Normal
Plasma sulfobromophthalein removal (% retention of 5 mg/kg dose at 45 min)†	Normal	Normal	Usually normal (<5%); mild 45-min (<15%) retention in some patients	Slow initial decline in plasma concentration (retention ≤20% at 45 min) with secondary rise at 90-120 min	Very slow initial decline in plasma concentration (45-min retention = 30-45%) without secondary rise
Oral cholecystography	Normal	Normal	Normal	Faint or non-visualization of gallbladder	Usually normal
Pharmacologic responses/special features	No response to phenobarbital	Phenobarbital reduces bilirubin by ≤75%	Phenobarbital reduces bilirubin, often to normal	Increased bilirubin concentration with estrogens; diagnostic urine coproporphyrin isomer pattern (total is normal, with isomer I increased to ≥80% of total)	Characteristic urine coproporphyrin excretion pattern (total is increased ≤2.5-fold in ~65% of cases but isomer I always <80% of total)
Major clinical features	Kernicterus in infancy if untreated; may occur later despite therapy	Rare late-onset kernicterus with fasting	None	Occasional hepatosplenomegaly	None
Hepatic morphology/histology	Normal	Normal	Normal; occasionally increased lipofucin pigment	Liver grossly black; coarse, dark centrilobular pigment	Normal
Bile bilirubin fractions‡	>90% unconjugated	Largest fraction (mean 57%) monoconjugates	Mainly diconjugates but monoconjugates are increased (mean 23%)	Mixed conjugates, reported increase in diconjugates	Increased conjugates
Inheritance (all autosomal)	Recessive	Recessive	Promoter mutation is recessive; missense mutation often dominant	Recessive; rare kindred appears dominant	Recessive
Diagnosis	Clinical and laboratory findings, lack of response to phenobarbital	Clinical and laboratory findings, response to phenobarbital	Clinical and laboratory findings; promoter genotyping; liver biopsy rarely necessary	Clinical and laboratory findings; liver biopsy unnecessary if coproporphyrin studies available; BSP disappearance	Clinical and laboratory findings; urine coproporphyrin analysis; BSP disappearance
Treatment	Phototherapy or tin protoporphyrin as short-term therapy; liver transplantation definitive	Consider phenobarbital if baseline bilirubin ≥8 mg/dL	None necessary	Avoid estrogens; no other therapy necessary	No treatment necessary

*UGT1A1 = bilirubin specific isoform of the UGT1 family of uridine diphosphate glucuronosyltransferases.
†Sulfobromophthalein (BSP) studies: Previously used to help distinguish Dubin-Johnson and Rotor syndromes if coproporphyrin isomer studies not available. However, BSP is no longer approved for clinical use in the United States.
‡Bilirubin in normal bile: <5% unconjugated bilirubin, with an average of 7% bilirubin monoconjugates and 90% bilirubin diconjugates.

excretion into the urine (see Table 147-2). Both syndromes carry a benign prognosis without specific therapy.

LIVER AND BILIARY TRACT DISEASE

Jaundice is a common sign of generalized hepatobiliary dysfunction, both acute and chronic. Icteric hepatobiliary disease is readily distinguished from the isolated disorders of bilirubin metabolism because the increase in plasma bilirubin concentration occurs in association with other markers of hepatobiliary disease (Fig. 147-2). Liver diseases can be categorized as those in which the primary injury results from inflammation and hepatocellular necrosis, inhibition of bile flow (cholestasis), or a combination of the two. The cholestatic disorders can be further subdivided into those resulting from mechanical obstruction of the bile duct flow and those from intrahepatic cholestasis, from a multitude of conditions that include several familial cholestatic syndromes; infiltrative disorders (Chapters 149 to 153), particularly those involving the intrahepatic biliary tree; certain other inflammatory or neoplastic conditions; and drug reactions (see Table 147-1 and Chapters 149 to 153).

Familial Cholestasis Syndromes

Bile secretion, which is essential both for the elimination of metabolic wastes and for the solubilization and subsequent absorption of specific nutrients, is a complex, energy-dependent process in which three ABC transporters (ATP8B1, ABCB11, and ABCB4) play critical roles. ATP8B1, also known as

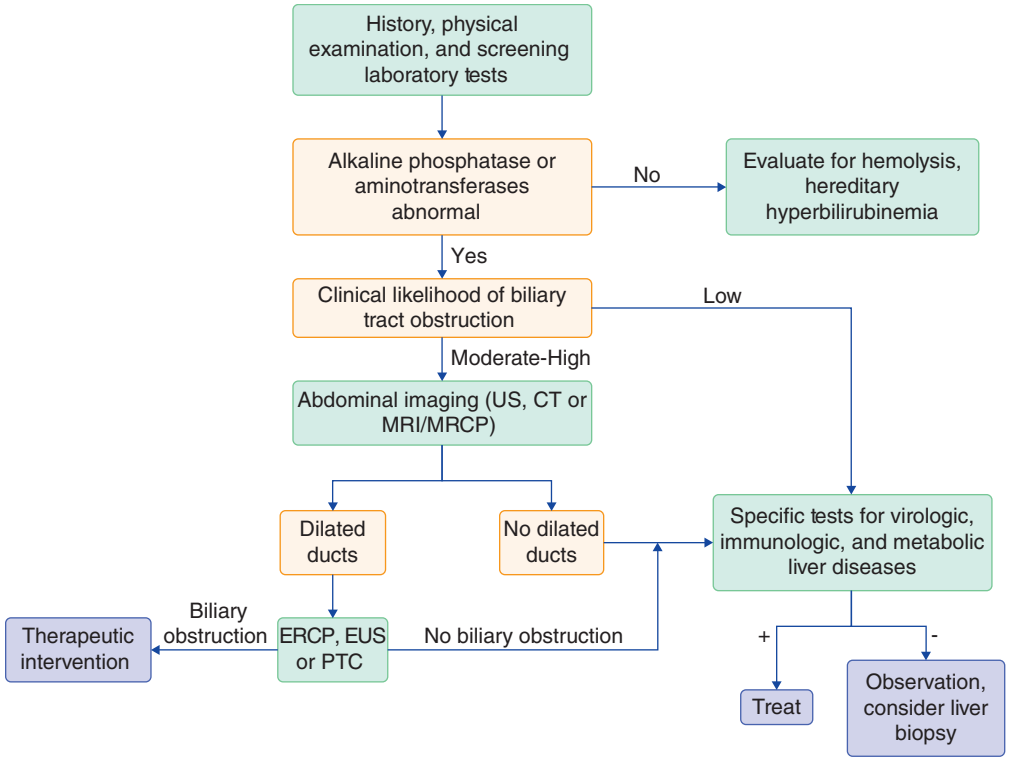

FIGURE 147-2. Diagnostic algorithm for the evaluation of hyperbilirubinemia and other liver test abnormalities and signs and symptoms suggestive of liver disease. CT = computed tomography; ERCP = endoscopic retrograde cholangiopancreatography; EUS = endoscopic ultrasound; MRCP = magnetic resonance cholangiopancreatography; MRI = magnetic resonance imaging; PTC = percutaneous transhepatic cholangiogram; US = ultrasound. (Modified from Lidofsky SD, Scharschmidt BF. Jaundice. In: Feldman M, Scharschmidt BF, Sleisenger MH, eds. *Gastrointestinal and Liver Disease.* 6th ed. Philadelphia: WB Saunders; 1998:227.)

familial intrahepatic cholestasis 1 (FIC1), is a phosphatidylserine flippase that translocates phosphatidylserine from the outer (canalicular) to the inner (cytoplasmic) leaflet of the canalicular plasma membrane; ABCB11, also known as the bile salt export pump (BSEP) or the sister of P glycoprotein (SPGP), translocates bile salts from the interior of the hepatocyte across the canalicular membrane into the bile. ABCB4 (multidrug resistance-associated protein 3 [MDR3]) is a phosphatidylcholine (lecithin) floppase translocating phosphatidylcholine outward from the cytoplasmic to the canalicular leaflet of the canalicular membrane.[12] A major function of ATP8B1 and ABCB4 is to maintain an appropriate physicochemical state of the canalicular membrane by regulating the balance of appropriate membrane phospholipids. Three different forms of severe pediatric cholestatic liver disease (designated progressive familial intrahepatic cholestasis types 1, 2, and 3) result from severe, homozygous mutations in these transporters. Less severe, nonprogressive cholestatic liver diseases (designated benign recurrent intrahepatic cholestasis [BRIC] types 1, 2, and 3) result from less severe mutations in these same genes.

Progressive familial intrahepatic cholestasis describes three phenotypically related syndromes of cholestasis during infancy and end-stage liver disease during childhood. All three disorders are inherited in an autosomal recessive pattern. In types 1 and 2, γ-glutamyl transpeptidase (GGT) levels are low despite elevations in alkaline phosphatase; by comparison, elevations in both GGT and alkaline phosphatase occur in type 3. In contrast to the selective bilirubin transport defect in Dubin-Johnson syndrome, the conjugated hyperbilirubinemia in these syndromes is caused by generalized bile secretory failure. Type 1 is the result of mutations in the *FIC1* gene that is also the cause of BRIC type 1. Types 2 and 3 are the result of mutations in genes *ABCB11* and *ABCB4,* respectively. *ABCB11* encodes a bile salt export pump, and *ABCB4* encodes the multidrug resistance protein 3 (MDR3), a protein responsible for the translocation of phosphatidylcholine.

Benign recurrent intrahepatic cholestasis is a rare, autosomal recessive disorder characterized by recurrent attacks of malaise, pruritus, and jaundice beginning in childhood or adulthood and varying in duration from weeks to months. Intervals between attacks may vary from months to years. This benign disorder does not progress to chronic liver disease or cirrhosis, and there is complete resolution between episodes. Treatment during the cholestatic episodes is symptomatic. Type 1 results from a mutation in the gene *familial intrahepatic cholestasis 1 (FIC1),* which is more severely mutated in progressive FIC1. It encodes the translocase protein ATP8B1 that transports aminophospholipids from the outer to the inner leaflet of various cell membranes. Phenotypically similar benign recurrent intrahepatic cholestasis type 2 and type 3 variants result from mutations in *ABCB11* and *ABCB4,* respectively.

Acquired Conjugation Defects

A modest reduction in bilirubin conjugating capacity occurs in advanced hepatitis or cirrhosis (Chapters 149 and 153). However, in these settings, conjugation is better preserved than other aspects of bilirubin disposition, such as canalicular excretion. Pharmacologic and metabolic perturbations may lead to acquired reductions in bilirubin conjugation. Unconjugated hyperbilirubinemia related to selective inhibition of UGT1A1 occurs with several human immunodeficiency virus (HIV) protease inhibitors (e.g., indinavir, atazanavir). Various other drugs (e.g., pregnanediol and the antibiotics novobiocin, chloramphenicol, gentamycin) also may cause unconjugated hyperbilirubinemia by inhibiting UGT1A1. In all settings in which UGT1A1 inhibitors cause unconjugated hyperbilirubinemia, the degree of hyperbilirubinemia is greater in patients with underlying Gilbert syndrome.

Jaundice in Pregnancy

Jaundice in pregnancy (Chapter 239) includes any liver disease that occurs during pregnancy. Conditions unique to pregnancy include a generally modest and self-limited elevation of the aminotransferase and bilirubin levels during the first trimester, often in patients with hyperemesis gravidarum. Intrahepatic cholestasis of pregnancy, which occurs during the second and third trimesters, is associated with intrauterine demise and resolves spontaneously after delivery. Mutations in genes encoding biliary transporters, especially *ABCB4,* have been reported in some but not all cases. Acute fatty liver or the HELLP (*h*emolysis, *e*levated *l*iver enzymes, and *l*ow *p*latelets) syndrome occurs in association with preeclampsia in the third trimester (Chapters 172 and 239). Acute fatty liver may resemble fulminant hepatic failure, with early delivery a prerequisite to maternal recovery.

Postoperative Jaundice

This multifactorial syndrome can be caused by increased bilirubin production (e.g., breakdown of transfused erythrocytes, resorption of hematomas), decreased hepatic bilirubin clearance (e.g., bacteremia, endotoxemia, parenteral nutrition, perioperative hypoxia), or both. Hyperbilirubinemia, which is the main biochemical feature, is often accompanied by a several-fold increase in alkaline phosphatase, GGT, or both levels. Aminotransferases are, at most, minimally elevated, and synthetic function is typically normal. The differential diagnosis includes biliary obstruction (Chapter 155) or hepatocellular injury related to shock, anesthetic injury (Chapter 150), or viral hepatitis (Chapter 148). Postoperative jaundice per se is not a threat to the patient, and it usually resolves in parallel with the patient's overall condition.

DIAGNOSIS

Accurate diagnosis and the distinction between acute and chronic disease often depend on appropriate selection and interpretation of a spectrum of laboratory and imaging studies. Tests used in initial evaluation of liver disease fall into two categories: tests that indicate injury, such as release of intracellular enzymes, and tests that measure, or at least reflect, actual function. Tests that reflect injury do not measure liver function and should not be called liver function tests. Liver tests must be chosen with care and interpreted within the overall clinical context. In specific situations, serial determinations are often helpful to assess the course of disease or the effects of therapy

Serum Enzyme Tests

The levels of hepatic enzymes found in plasma are a measure of hepatocyte turnover or injury. Enzymes released during normal hepatocyte turnover are believed to be the basis for normal circulating levels. Cell injury and cell death activate phospholipases that create holes in the plasma membrane, thereby increasing the release of intracellular contents.

Aminotransferases

The aminotransferases (formerly called transaminases) catalyze transfer of the α-amino group of aspartate (aspartate aminotransferase [AST]) or alanine (alanine aminotransferase [ALT]) to the α-keto group of α-ketoglutarate, with pyridoxal phosphate (vitamin B_6) as a cofactor. Laboratory methods that assay aminotransferase activity require supplementation with vitamin B_6 to avoid falsely decreased activity in subjects who are vitamin B_6 deficient.

Normal serum levels, which are established locally from samples obtained from normal populations, may vary appreciably in different populations but are typically 40 IU/L or less (see Appendix). Values can exceed 1000 IU/L in acute hepatocyte injury, for example, from viral infection (Chapter 148) or toxins (Chapter 150). ALT is a purely cytosolic enzyme. Distinct isoforms of AST are present in cytosol and mitochondria, and AST is also found on the plasma membrane. Expression of the mitochondrial isoform and its physiologic export from the hepatocyte are upregulated by ethanol. Circulating levels of AST and ALT are elevated in most hepatic diseases, and the degree of aminotransferase activity found in plasma roughly reflects the current activity of the disease process. There are, however, critical exceptions. In even the most severe cases of alcoholic hepatitis, aminotransferase levels greater than 200 to 300 IU/L are uncommon (Chapter 152). In nonalcoholic fatty liver disease, normal values for ALT may be found in an appreciable fraction of patients with active nonalcoholic steatohepatitis, fibrosis, or even cirrhosis. By contrast, aminotransferase activities of 1000 IU/L or greater are often present in even mild acute viral hepatitis (Chapter 148) or shortly after acute biliary obstruction, for example, during passage of a gallstone (Chapter 155). Conversely, aminotransferase levels may decline during the course of massive hepatic necrosis because liver injury is so extensive that little enzyme activity remains (Chapter 154). In rare circumstances, antibodies to AST result in an antibody-enzyme complex called a macroenzyme that has a delayed clearance from the circulation.

Aminotransferase levels are useful in several distinct ways. First, they provide a relatively specific screening test for hepatobiliary disease. Although AST levels may be increased with disease of other organs (notably myocardial and skeletal muscle), values 10 times the upper limit of normal or greater almost invariably indicate hepatobiliary pathology. Moreover, in the total clinical context, the source of increased aminotransferase activity is usually obvious. Aminotransferase levels are also used to monitor the activity of an acute or chronic parenchymal liver disease and its response to therapy. However, levels in a given patient may correlate poorly with the severity of the disease as assessed by liver biopsy, particularly in chronic hepatitis C (Chapter 149) and nonalcoholic fatty liver (Chapter 152).[13] Aminotransferases are also often normal despite advanced cirrhosis (Chapter 153), in which they are of limited prognostic value. Finally, aminotransferase levels may provide diagnostic clues. AST levels 15 or more times normal are unusual in chronic bile duct obstruction without cholangitis, and AST levels 6 or more times normal are uncommon in alcoholic liver disease in the absence of other causes. In most liver diseases, the ratio of AST to ALT is usually 1 or less; however, ratios are typically 2 or higher in alcoholic fatty liver and alcoholic hepatitis (Chapter 152), reflecting increased synthesis and secretion of mitochondrial AST into plasma and selective loss of ALT activity because of the pyridoxine deficiency commonly seen in alcoholism. An elevated AST/ALT ratio also occurs in fulminant hepatitis related to Wilson disease (Chapters 146 and 211).

Alkaline Phosphatase

Alkaline phosphatases are widely distributed enzymes (e.g., liver, bile ducts, intestine, bone, kidney, placenta, and leukocytes) that catalyze the release of orthophosphate from ester substrates at an alkaline pH. The normal activity level in adult serum is highly dependent on the measurement method, age, and sex. Two widely used current methods have upper limits of normal in adults of 85 and 110 IU/L (see Appendix). Higher levels are normal in children and in pregnancy. Results must always be compared with the appropriate normal range. In bone, alkaline phosphatase participates in the deposition of hydroxyapatite in osteoid. In other sites, including the liver, its phosphatase activity may facilitate movement of molecules across cell membranes. Serum alkaline phosphatase activity principally reflects the contribution of hepatic and bone isoforms; the intestinal form may account for 20 to 60% of the total after a fatty meal. There is a substantial placental contribution to the alkaline phosphatase level late in pregnancy; the Regan isozyme, a variant that appears identical to the placental form, is associated with hepatocellular cancer (Chapter 196), lung cancer (Chapter 191), and other tumors.

Elevations in the serum alkaline phosphatase activity in cholestatic hepatobiliary disease result from two distinct mechanisms: increased synthesis and secretion of the enzyme and solubilization from the apical (canalicular) surface of hepatocytes and the luminal surface of biliary epithelial cells by the increased local concentrations of bile acids that occur with cholestasis. Serum alkaline phosphatase activity also may be increased in bone disorders (e.g., Paget disease [Chapter 247], osteomalacia [Chapter 244], bone metastases [Chapter 202]), during rapid bone growth in children, in the later stages of pregnancy, with chronic renal failure (Chapter 130), and, occasionally, in the presence of malignancy not involving bones or liver. The source is often obvious, but when it is not, fractionation techniques can distinguish hepatobiliary alkaline phosphatase from other forms. A simpler alternative is to measure serum levels of GGT or 5′-nucleotidase (5′-NT), which tend to parallel levels of alkaline phosphatase in hepatobiliary disease but are usually not increased in bone disease. With a serum half-life of approximately 1 week, serum alkaline phosphatase levels may remain elevated for days to weeks after resolution of biliary obstruction. This delay may be especially misleading when it is accompanied by prolonged direct-reacting hyperbilirubinemia owing to delayed clearance of δ-bilirubin.

Modest increases in serum alkaline phosphatase activity (<3 times normal) occur in many hepatic parenchymal disorders, including hepatitis and cirrhosis. In the absence of bone disease, larger increases (3 to 10 times normal) usually indicate obstruction of bile flow. Although the highest levels usually reflect obstruction of the common bile duct, elevations also can occur with obstruction of intrahepatic bile ducts from infiltrative conditions that arise from granulomatous processes (sarcoidosis [Chapter 95], *Mycobacterium avium-intracellulare* infection [Chapter 325], mycobacteria tuberculosis [Chapter 324]), malignancy (lymphoma, cholangiocarcinoma, or metastatic cancer), or amyloidosis (Chapter 188).

Other Hepatic Enzymes

5′-NT is a plasma membrane enzyme that cleaves orthophosphate from the 5′ position on the pentose sugar of adenosine or inosine phosphate. Leucine aminopeptidase (LAP) is a ubiquitous cellular peptidase. The serum levels of both usually increase in cholestasis. Accordingly, their major use is to confirm whether an elevated serum alkaline phosphatase is of hepatobiliary origin. Both enzymes may be increased in the latter stages of a healthy pregnancy.

GGT is present in many tissues. Its serum activity increases in hepatobiliary disease but also after myocardial infarction; in neuromuscular diseases, pancreatic disease (even in the absence of biliary obstruction), pulmonary disease, and diabetes; and during the ingestion of ethanol and other inducers of microsomal enzymes. Nevertheless, because serum GGT levels are usually normal in bone disease, the enzyme may be helpful in confirming the hepatic origin of alkaline phosphatase. Measurement of GGT has been proposed as a sensitive screening test for hepatobiliary disease and for monitoring abstinence from ethanol. Because of its low specificity, many persons who test positive have no identifiable liver disease on further study. GGT offers no clear advantage over LAP or 5′-NT for identifying the source of increased serum alkaline phosphatase activity except in pregnancy. Serum GGT levels may be normal despite elevated hepatobiliary alkaline phosphatase levels in certain rare disorders, including benign recurrent intrahepatic cholestasis and progressive familial intrahepatic cholestasis types 1 and 2 (see earlier and Chapter 155).

Lactate dehydrogenase levels are often elevated in hepatic ischemia and other conditions that result in hepatic necrosis; otherwise the enzyme is too ubiquitous in other body tissues to be diagnostically useful.

Tests Based on Clearance of Metabolites and Drugs

A major function of the liver is to remove various metabolites and toxins from the blood (Chapter 150). In liver disease, clearance of such molecules may be impaired because of loss of parenchymal cells, diminished bile secretion, biliary obstruction, decreased cellular uptake or metabolism, or reduced or heterogeneous hepatic blood flow. When a metabolite is produced at a relatively constant rate (e.g., bilirubin), its serum level can be a sensitive indicator of liver function. The removal rate from plasma of certain exogenous drugs and dyes can be similarly interpreted. However, plasma removal rates of sulfobromophthalein (BSP) and indocyanine green, once widely used as tests of liver function, have essentially been abandoned.

Bilirubin

The differential diagnosis of hyperbilirubinemia (see earlier) includes generalized liver disease, inherited disorders of bilirubin metabolism (e.g., Gilbert, Crigler-Najjar, Dubin-Johnson, and Rotor syndromes) and nonhepatic conditions (e.g., hemolysis). Higher bilirubin levels correlate with a poorer prognosis in most forms of chronic liver disease.

Ammonia

Ammonia, a byproduct of amino acid metabolism, is removed from blood by the liver, converted to urea in the Krebs-Henseleit cycle, and excreted by the kidneys (Chapter 115). In the setting of portosystemic shunting or severe hepatic dysfunction (e.g., fulminant hepatic failure), ammonia levels rise. Measurements of blood ammonia are principally used to confirm a diagnosis of hepatic encephalopathy or response to treatment of the encephalopathy. However, serum ammonia levels do not strongly correlate with the severity of hepatic encephalopathy (Chapter 153). Correlations may be somewhat better if the measurement is made rapidly on an iced arterial blood sample. Elevated ammonia levels also occur when ammonia production is increased by intestinal flora (e.g., after a high-protein meal or gastrointestinal bleeding), by the kidney (in response to metabolic alkalosis or hypokalemia), or in rare genetic diseases that affect the pathway of urea synthesis (Chapter 205).

Drug Clearance

The rate of hepatic clearance of compounds such as lidocaine and aminopyrine from the circulation can be measured chemically or with radiolabeled tracers. Although such tests can quantify hepatic function, they are rarely used in clinical practice.

Tests Reflecting Hepatic Synthetic Function

Coagulation Tests

See also Chapters 38 and 171.

Prothrombin Time

The prothrombin time (PT) reflects the plasma concentrations of both extrinsic and common pathway factors, that is, factors VII, X, and V, prothrombin, and fibrinogen. A prolonged PT most often results from vitamin K deficiency, liver disease, or both. Vitamin K, a fat-soluble vitamin, is found in many foods and is also synthesized by gut bacteria (Chapter 174). Vitamin K deficiency can be caused by poor dietary intake and malabsorptive states, including the fat malabsorption that results from cholestasis, and it also occurs with antibiotic suppression of gut flora, particularly in patients who receive inadequate vitamin K replacement.

The half-lives of clotting factors are typically less than 1 day. Factor VII, which has the shortest half-life, is usually the earliest and most severely depressed during periods of defective hepatic synthesis. Because the PT is dependent on the level of factor VII, it responds rapidly with changes in hepatic synthetic function; it is useful for following the course of acute liver diseases, in which a significant or growing prolongation of the PT may indicate a poor prognosis (Chapter 148). An abnormal PT that is due solely to vitamin K deficiency usually becomes normal within 24 to 48 hours after parenteral repletion. However, if decreased synthesis of clotting factors reflects hepatocyte dysfunction, there may be little or no response to vitamin K.

Although the prolongation of PT with liver disease is generally accepted as indicative of defective clotting factor synthesis, the synthesis of both procoagulant and anticoagulant proteins is disturbed in liver disease.[14] The physiologic consequences of this rebalancing of coagulation may preserve the coagulation response. Prolongation of the PT may also reflect disseminated intravascular coagulation (Chapter 175), which should always be considered in the context of both acute liver failure and end-stage chronic liver disease.

Partial Thromboplastin Time

This test reflects both the intrinsic and common pathway factors, that is, all of the classical clotting factors except factor VII, and is, therefore, complementary to the PT. It is especially useful in detecting circulating anticoagulants (Chapter 175) but adds little to the PT in evaluating hepatic synthetic function.

Albumin

Albumin is produced solely by the liver. Its plasma concentration reflects a balance between its synthetic rate of approximately 100 to 200 mg/kg/day and its plasma half-life of approximately 21 days. The synthetic rate is affected by the patient's nutritional state, thyroid and glucocorticoid hormone levels, plasma colloid osmotic pressure, exposure to hepatotoxins (e.g., alcohol), and presence of systemic disorders, liver disease, or both. Many conditions increase albumin losses and shorten its plasma half-life, including nephrotic syndrome (Chapter 121), protein-losing enteropathy (Chapter 140), severe burns (Chapter 111), exfoliative dermatitis, and major gastrointestinal bleeding (Chapter 135). In cirrhosis with ascites (Chapters 153 and 154), hypoalbuminemia indicates diminished synthesis or redistribution into ascitic fluid. Thus, a reduced serum albumin concentration can be considered an indicator of decreased hepatic synthetic function only when these factors are not involved.

Hematologic Tests in Liver Disease

In moderate to severe acute liver diseases, varying degrees of cytopenias can be seen with all three cell lineages. The most common finding is thrombocytopenia from hypersplenism, which can be a surrogate marker of portal hypertension. Anemia may reflect low-grade hemolysis or marrow depression. Bone marrow suppression may be caused by ethanol or drugs, and aplastic anemia is an uncommon but well-recognized complication of acute viral hepatitis (Chapters 148 and 165). Zieve syndrome (hemolytic anemia and hypertriglyceridemia) is a rare but well-characterized complication of severe alcoholic liver disease (Chapters 152 and 153). Modest leukopenia, often with atypical lymphocytes, also may be present. Chronic liver disease, especially if cholestatic, may be accompanied by target cells in the peripheral blood smear. Target cells are erythrocytes with an expanded cell membrane that reflects abnormalities in serum lipids. Spur cells (acanthocytes), most often found in advanced alcoholic cirrhosis, reflect a still greater increase in membrane cholesterol.

Tests for Specific Liver Diseases

Patients who present with a picture of acute or chronic parenchymal liver disease are most likely to fall into one of three categories: viral or toxic hepatitis, including alcoholic liver disease; autoimmune liver disease; or an inherited or acquired metabolic disorder (Chapter 146). Specific tests for viral antigens, nucleic acids, and antibodies are available for the conventional hepatitis viruses, including A, B, C, D (delta), and E, chronic forms of which are being increasingly seen in immunosuppressed patients (Chapters 148 and 149), as well as Epstein-Barr virus (Chapter 377), cytomegalovirus (Chapter 376), and herpesviruses (Chapters 374 and 375), which are well-established but less common causes of liver disease. The major autoimmune diseases of the liver include primary biliary cirrhosis (Chapter 155), autoimmune hepatitis (Chapter 149), primary sclerosing cholangitis (Chapter 155), and various overlap syndromes. The starting point for establishing a specific diagnosis within this category is the search for specific autoantibodies in serum, including antimitochondrial antibodies against epitopes of the pyruvate dehydrogenase complex, which are virtually diagnostic of primary biliary cirrhosis (Chapter 155), and antinuclear, anti–smooth muscle, and anti–liver/kidney microsomal antibodies, which suggest a diagnosis of one of the subtypes of autoimmune hepatitis (Chapter 149). The most prevalent of the hereditary metabolic disorders affecting the liver include hemochromatosis, α_1-antitrypsin deficiency, and Wilson disease (Chapter 211). Nonalcoholic fatty liver disease is the most frequent acquired metabolic liver disease. The disorder is not associated with any known serologic markers, although nonspecific elevations in antinuclear antibody may be present (Chapter 152).

Liver Biopsy

Liver biopsy can be of great help in the diagnosis of diffuse or localized parenchymal diseases, including chronic hepatitis, cirrhosis, and primary or metastatic malignancy in the liver. The value of liver biopsy in acute hepatitis or acute cholestatic jaundice may be primarily prognostic because the histologic changes in these settings may be nonspecific. However, drug-induced liver injury due to certain specific agents (Chapter 150) may display diagnostic features. Liver biopsy for assessment of diffuse disease can be performed percutaneously after localization of the liver by physical examination or ultrasonographic visualization. When specific lesions, such as tumors, must be sampled, the biopsy can be guided by ultrasonographic or radiographic imaging or performed under direct visualization during laparoscopy. Relative or absolute contraindications include coagulopathy, high-grade biliary obstruction, biliary sepsis, ascites, and right pleural disease. Although liver biopsy remains the standard for assessment of hepatic histology in diffuse disease (Chapter 146), the procedure's invasiveness and concern for sampling error have generated interest in noninvasive measures of hepatic fibrosis. These noninvasive studies fall into two categories. One category uses magnetic resonance imaging or ultrasound to measure liver stiffness as a surrogate for fibrosis. The second category comprises panels of biomarkers accessible by blood testing to predict the severity of necroinflammation and fibrosis. These serologic panels typically include standard laboratory measures of hepatic injury (GGT, total bilirubin) and other serum markers (e.g. haptoglobin, hyaluronic acid, apolipoprotein A1). However, the ability of currently available noninvasive markers to assess the extent of hepatic fibrosis reliably across the clinically relevant histologic spectrum, especially in an individual patient, remains to be established.[15]

APPROACH TO THE PATIENT WITH JAUNDICE OR ABNORMAL LIVER TESTS

History, Physical Examination, and Initial Laboratory Studies

Patients with liver disease may present with jaundice or with other signs or symptoms, or the disease may be detected in the asymptomatic patient by the finding of abnormal liver tests during a routine evaluation. Regardless of how the patient comes to medical attention, the diagnostic approach (see Fig. 147-2) begins with a careful history and physical examination (Chapter 146) and screening laboratory studies (complete blood cell count, measurement of plasma bilirubin concentration, assay of ALT, AST, and alkaline phosphatase levels, and PT) to formulate an initial differential diagnosis.[16] The ability to distinguish expeditiously between liver disease and extrahepatic biliary tract obstruction is the major goal of the initial evaluation, in part because the latter may call for prompt surgical intervention. Appropriate selection of second-level laboratory tests and imaging studies leads to a definitive diagnosis in most patients. Care in selecting tests, particularly imaging studies, can both maximize the likelihood of making a correct diagnosis and protect the patient from unnecessary discomfort, risk, and expense.

If the patient is asymptomatic and hepatic tests other than bilirubin are normal, hemolysis or an isolated disorder of bilirubin metabolism should be considered. If signs, symptoms, or laboratory abnormalities indicate hepatobiliary disease, certain patterns of findings help to distinguish intrinsic liver disease from biliary obstruction (Table 147-3). Pain in the right upper quadrant accompanied by a predominant increase in serum alkaline phosphatase activity suggests biliary obstruction (Chapter 155), as does a history of biliary surgery, right upper quadrant scars, or an abdominal mass. Fever and rigors, indicative of cholangitis, strengthen this conclusion. The incidences of gallstone disease and malignant neoplasm increase with age, although risk factors such as obesity or recent extensive diet-induced weight loss increase the risk for gallstones. Other risk factors (e.g., hepatitis exposure, transfusions, intravenous drug use, alcohol use, certain medications, obesity, and family history of genetic diseases) and a predominant elevation in serum aminotransferase levels favor a diagnosis of parenchymal liver disease. Physical evidence of cirrhosis (e.g., spider angiomas, gynecomastia, ascites, splenomegaly) supports the diagnosis of chronic parenchymal disease.

Despite the general validity of these patterns, many exceptions exist. In particular, parenchymal disorders with prominent cholestasis may mimic biliary obstruction. Both alkaline phosphatase and GGT are usually elevated in patients with cholestasis; the combination of an elevated alkaline phosphatase and normal GGT suggests that the alkaline phosphatase is from bone. Conversely, an isolated elevation of GGT may result from certain drugs (e.g., diphenylhydantoin) or alcohol consumption even in the absence of liver disease. Because of the risk for life-threatening infection in the setting of unrelieved biliary tract obstruction, this possibility must always be considered and excluded if an alternative diagnosis is not definitely established.

TABLE 147-3 OBSTRUCTIVE JAUNDICE VERSUS CHOLESTATIC LIVER DISEASE

FEATURE	SUGGESTS OBSTRUCTIVE JAUNDICE	SUGGESTS PARENCHYMAL LIVER DISEASE
History	Abdominal pain Fever, rigors Prior biliary surgery Acholic stools	Anorexia, malaise, myalgias, suggestive of viral prodrome Known infectious exposure Use of injection drugs or intranasal cocaine Exposure to known hepatotoxin Family history of jaundice
Physical examination	High fever Abdominal tenderness Palpable abdominal mass Abdominal scar	Ascites Other stigmata of liver disease (e.g., prominent abdominal veins, gynecomastia, spider angiomata, asterixis, encephalopathy, Kayser-Fleischer rings)
Laboratory studies	Predominant elevation of serum bilirubin and alkaline phosphatase Prothrombin time that is normal or normalizes with vitamin K administration Elevated serum amylase	Predominant elevation of serum aminotransferases Prolonged prothrombin time that does not correct with vitamin K administration Blood tests indicative of specific liver disease

Imaging Studies

If extrahepatic obstruction is suspected, its site and nature can now be determined in virtually all patients (see Fig. 147-2). A reasonable initial step is the use of a noninvasive imaging study (Chapter 133) such as ultrasonography or magnetic resonance cholangiopancreatography to determine whether the intrahepatic, extrahepatic biliary system, or both are dilated, implying mechanical obstruction. Because of its lesser expense, portability, and convenience, ultrasound is often the procedure of choice, especially if gallstones are suspected. Magnetic resonance cholangiopancreatography may provide more precise resolution, including stricturing of intrahepatic ducts characteristic of primary sclerosing cholangitis. However, each of these techniques can fail to identify dilated ducts, particularly in patients with cirrhosis. Conversely, a modest degree of ductal dilatation is common in a patient with a previous cholecystectomy and does not necessarily signify current obstruction. If dilated ducts are found, the biliary tree should be examined by endoscopic retrograde cholangiopancreatography (ERCP) or percutaneous transhepatic cholangiography (PTC) (Chapter 134). ERCP involves positioning an endoscope in the duodenum, inserting a catheter through the ampulla of Vater, and injecting contrast medium into the distal common bile duct, pancreatic duct, or both. PTC involves percutaneous passage of a needle through the hepatic parenchyma into a peripheral bile duct, followed by injection of contrast medium into the biliary tree through the peripheral duct. The choice of procedure is based on the suspected site of obstruction (proximal vs. distal); the presence of coagulopathy; a history of abdominal surgery that might complicate PTC or ERCP; the likely need for a therapeutic procedure (e.g., stent placement or endoscopic sphincterotomy); and the skills of available staff. Endoscopic ultrasound (EUS) is a complementary approach that permits internal ultrasonographic analysis of the pancreas, extrahepatic bile ducts, and regional lymph nodes and blood vessels. EUS combined with fine needle aspiration permits tissue sampling of abnormalities in areas such as the bile ducts and pancreas that typically have been difficult to sample percutaneously.

Selection of Imaging Tests

Liver ultrasound is an ideal screening test to evaluate the liver architecture, assess for surface nodularity and parenchymal mass lesions, and exclude biliary obstruction. Ultrasound is relatively inexpensive in comparison to other imaging modalities, is widely available, and avoids ionizing radiation. The identification of mass lesions on ultrasound will commonly prompt further cross-sectional imaging with either computed tomography or magnetic resonance imaging. If there is evidence of biliary obstruction on imaging or, if obstruction is still considered likely despite imaging findings, direct

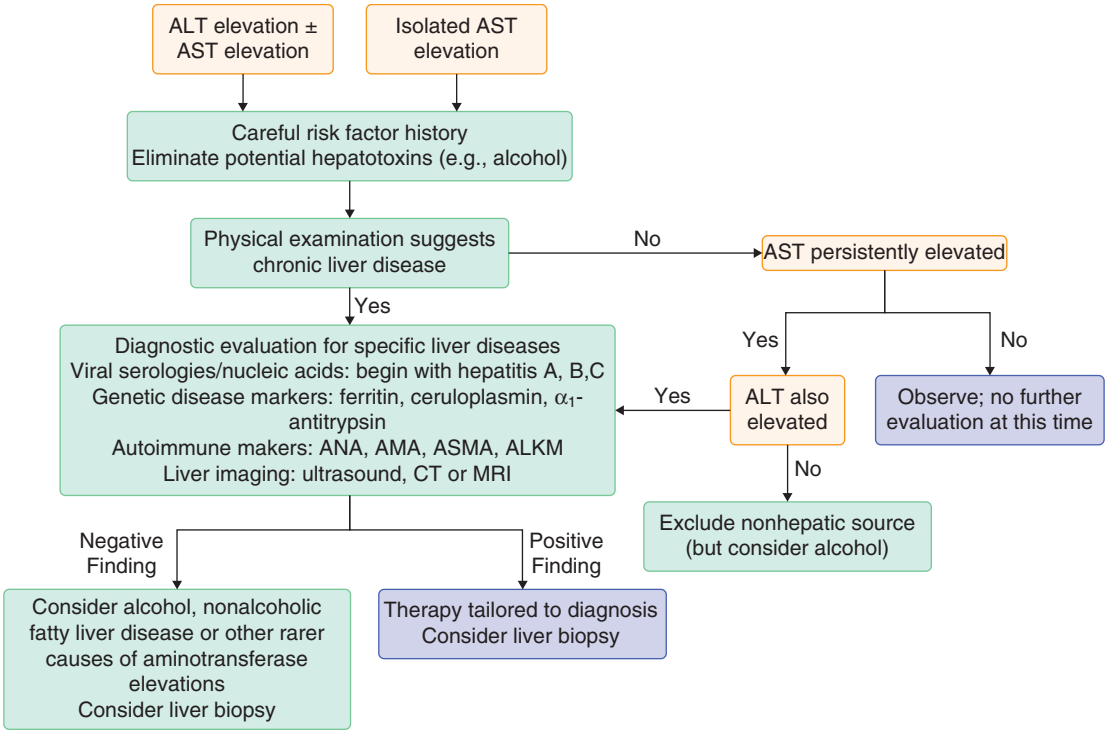

FIGURE 147-3. **Approach to the evaluation of isolated elevated levels of serum alanine aminotransferase (*ALT*), aspartate aminotransferase (*AST*), or both in the asymptomatic patient.** ANA = antinuclear antibody; AMA = antimitochondrial antibody; ASMA = anti–smooth muscle antibody; ALKM = anti–liver/kidney microsomal antibody; CT = computed tomography; MRI = magnetic resonance imaging.

cholangiography by ERCP or PTC, which offer therapeutic as well as diagnostic capabilities, may be an appropriate choice. If obstruction is considered possible but not highly likely, noninvasive imaging with MRCP is a reasonable study. Individual radiology suites have different levels of expertise for these procedures, and the local radiology staff may be quite helpful in recommending the best procedure for a given patient.

The apparently healthy patient with an isolated abnormality of the aminotransferase or alkaline phosphatase levels requires careful evaluation to identify any underlying disease while avoiding unneeded testing. Often, no significant disease is found despite extensive evaluation. Common causes of abnormal enzyme tests include alcohol consumption, hepatitis C infection, nonalcoholic fatty liver disease, bone disease, and muscle injury.

Asymptomatic Aminotransferase Elevation

Epidemiologic data suggest that up to 25% of asymptomatic adult Americans have a mild to moderate elevation of aminotransferase levels. The incidental discovery of such abnormalities is currently the most frequent means by which liver disease is first recognized. Whereas up to one third of such patients have no elevation on subsequent testing, many others prove to have steatohepatitis (Chapter 152) or chronic hepatitis C (Chapter 149). Further evaluation is generally indicated only in patients with persistent abnormalities (Fig. 147-3). Initial screening should include a careful history of exposure to hepatotoxins (alcohol, prescription drugs, over-the-counter medications, herbs, chemicals, and occupational exposures). If the abnormal test was an AST, a hepatic origin for the enzyme elevation should be confirmed with an ALT determination. If the ALT is normal, a muscle source is likely or the elevation may reflect the presence of a macroenzyme. If the ALT level is abnormal, the patient should be screened serologically for hepatitis B, C, and (at least in immunosuppressed patients) E; markers of autoimmune liver disease; and serologic markers of inherited metabolic disorders (Chapter 146). AST abnormalities caused by alcohol-induced injury should become normal with several weeks of abstinence. If the abnormalities persist for 6 to 12 months without an apparent cause, liver biopsy should be considered.

Asymptomatic Alkaline Phosphatase Elevation

Many patients with isolated elevation of the alkaline phosphatase level have nonhepatic causes, such as pregnancy or bone disease. The origin of an elevated alkaline phosphatase should be confirmed with a fasting sample because intestinal alkaline phosphatase may be elevated after a meal (Fig. 147-4).

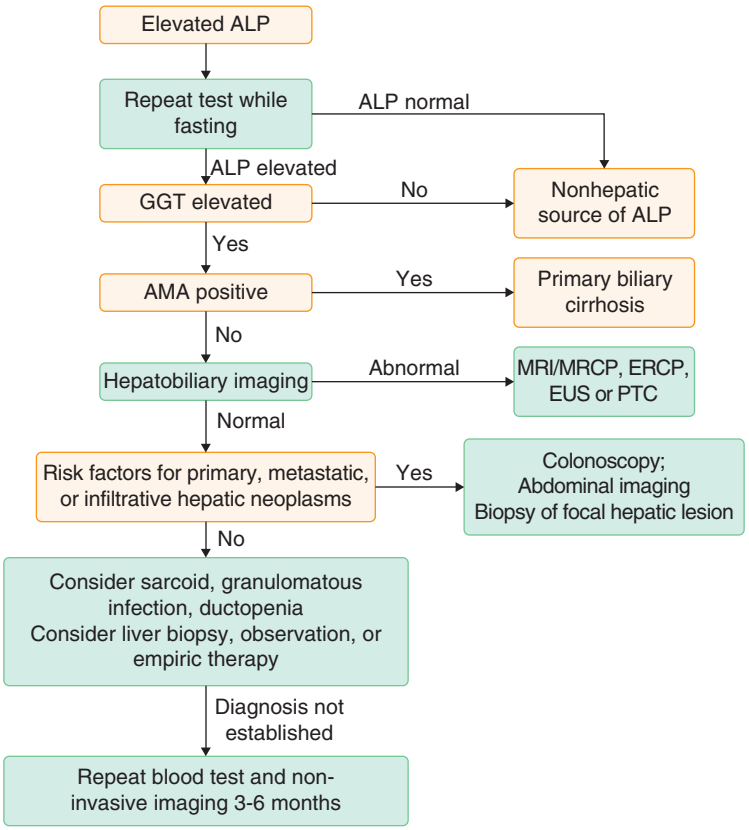

FIGURE 147-4. **Approach to the asymptomatic patient with isolated elevated levels of serum alkaline phosphatase (*ALP*).** In cases with a high index of suspicion of biliary tract disease (e.g., sclerosing cholangitis), cholangiography may be warranted even in the face of normal ultrasound imaging. AMA = antimitochondrial antibody; CT = computed tomography; ERCP = endoscopic retrograde cholangiopancreatography; EUS = endoscopic ultrasound; GGT = γ-glutamyl transpeptidase; MRCP = magnetic resonance cholangiopancreatography; MRI = magnetic resonance imaging; PTC = percutaneous transhepatic cholangiogram; US = ultrasonography.

A hepatic source is highly likely if the serum GGT is also abnormal. Serologic studies should include an antimitochondrial antibody test; a positive result suggests primary biliary cirrhosis (Chapter 155). A careful history identifies patients at risk for intrahepatic cholestasis related to drugs or toxins. Essentially all other patients with persistently abnormal alkaline phosphatase should receive a hepatobiliary sonogram or other noninvasive imaging test. Demonstration of dilated intrahepatic or extrahepatic bile ducts should prompt direct visualization of the biliary tract by ERCP or PTC (Chapters 134 and 155). Evidence of an intrahepatic mass should prompt thorough evaluation for possible malignancy (Chapters 155 and 196). Because colon cancer often metastasizes to liver, colonoscopy may be useful in appropriate cases (Chapter 193). Infiltrative diseases, including amyloidosis and granulomatous hepatitis (Chapter 151), should be considered. In the absence of evidence of biliary obstruction or a cause identifiable by noninvasive means, liver biopsy should be strongly considered to complete the evaluation of cholestatic liver test abnormalities.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Wegiel B, Nemeth Z, Correa-Costa M, et al. Heme oxygenase-1: a metabolic nike. *Antioxid Redox Signal*. 2014;20:1709-1722.
2. Dhumeaux D, Erlinger S. Hereditary conjugated hyperbilirubinaemia: 37 years later. *J Hepatol*. 2013;58:388-390.
3. O'Connor MC, Lease MA, Whalen BL. How to use: transcutaneous bilirubinometry. *Arch Dis Child Educ Pract Ed*. 2013;98:154-159.
4. Erlinger S, Arias IM, Dhumeaux D. Inherited disorders of bilirubin transport and conjugation: new insights into molecular mechanisms and consequences. *Gastroenterology*. 2014;146:1625-1638.
5. Canu G, Minucci A, Zuppi C, et al. Gilbert and Crigler Najjar syndromes: an update of the UDP-glucuronosyltransferase 1A1 (*UGT1A1*) gene mutation database. *Blood Cells Mol Dis*. 2013;50:273-280.
6. Skierka JM, Kotzer KE, Lagerstedt SA, et al. *UGT1A1* genetic analysis as a diagnostic aid for individuals with unconjugated hyperbilirubinemia. *J Pediatr*. 2013;162:1146-1152.
7. Strassburg CP. Gilbert-Meulengracht's syndrome and pharmacogenetics: is jaundice just the tip of the iceberg? *Drug Metab Rev*. 2010;42:168-181.
8. Watchko JF, Tiribelli C. Bilirubin-induced neurologic damage: mechanisms and management approaches. *N Engl J Med*. 2013;369:2021-2030.
9. Tu ZH, Shang DS, Jiang JC, et al. Liver transplantation in Crigler-Najjar syndrome type I disease. *Hepatobiliary Pancreat Dis Int*. 2012;11:545-548.
10. Sellier AL, Labrune P, Kwon T, et al. Successful plasmapheresis for acute and severe unconjugated hyperbilirubinemia in a child with Crigler Najjar type I syndrome. *JIMD Rep*. 2012;2:33-36.
11. van de Steeg E, Stranecky V, Hartmannova H, et al. Complete OATP1B1 and OATP1B3 deficiency causes human Rotor syndrome by interrupting conjugated bilirubin reuptake into the liver. *J Clin Invest*. 2012;122:519-528.
12. Chan J, Vandeberg JL. Hepatobiliary transport in health and disease. *Clin Lipidol*. 2012;7:189-202.
13. Kleiner DE, Berk PD, Hsu JY, et al. Hepatic pathology among patients without known liver disease undergoing bariatric surgery: observations and a perspective from the Longitudinal Assessment of Bariatric Surgery (LABS) study. *Semin Liver Dis*. 2014;34:98-107.
14. Tripodi A, Mannucci PM. The coagulopathy of chronic liver disease. *N Engl J Med*. 2011;365:147-156.
15. Crespo G, Fernandez-Varo G, Marino Z, et al. ARFI, FibroScan, ELF, and their combinations in the assessment of liver fibrosis: a prospective study. *J Hepatol*. 2012;57:281-287.
16. Woreta TA, Alqahtani SA. Evaluation of abnormal liver tests. *Med Clin North Am*. 2014;98:1-16.

REVIEW QUESTIONS

1. A 51-year-old man presents with 3 months of painless jaundice, pale stools, and dark urine, as well as spiking fevers. His Hgb is 10.2 g/dL; WBC 6500/μL with 64% PMNs; and platelets 145,000. Total bilirubin is 28 mg/dL; direct-reacting bilirubin 21 mg/dL; alkaline phosphatase 336, AST 94, ALT 116, albumin 3.1, PT INR 1.8. Urine urobilinogen is undetectable, and urine bilirubin is ++++. Abdominal ultrasound shows proximal dilatation of the intrahepatic bile ducts, and magnetic resonance cholangiopancreatography reveals a constricting mass in the common bile duct several centimeters distal to the junction of the right and left hepatic ducts. After being started on intravenous broad-spectrum antibiotics, the patient is taken to surgery, where the mass is resected and the common bile duct is anastomosed over a stent. Two weeks later his total bilirubin is 14 mg/dL, direct-reacting bilirubin 10 mg/dL, alkaline phosphatase 225, and urine urobilinogen still undetectable. Urine bilirubin is also undetectable. Because of the persistent direct-reacting hyperbilirubinemia and elevated alkaline phosphatase, the surgeon is considering taking the patient back to the operating room and re-exploring his biliary tract. Which of the following test results argues *against* that decision?

A. Total bilirubin
B. Direct-reacting bilirubin
C. Alkaline phosphatase
D. Urine urobilinogen
E. Urine bilirubin

Answer: E After successful relief of chronic bile duct obstruction, both the total bilirubin and the alkaline phosphatase may take several weeks to normalize. The combination of elevated direct-reacting bilirubin in plasma and an absence of bilirubin from the urine indicates that the direct-reacting fraction in plasma now consists of δ-bilirubin, a covalent complex of conjugated bilirubin with albumin, that is not cleared by the kidney. Although the absence of urobilinogen from the urine at the time of admission reflected complete bile duct obstruction, its persistence after 2 weeks of broad-spectrum antibiotic therapy no longer reflects biliary obstruction, but rather elimination by the antibiotics of the bacteria necessary to convert bilirubin in the gut to urobilinogen. In this case, use of a dipstick that detects bilirubinuria could save the patient from an unnecessary second surgical procedure.

2. A 24-year-old nurse presented to the employee health department of her hospital complaining of arthralgias, weakness, and fatigue for 2 weeks. She had previously been in good health and had a normal physical examination, CBC, and metabolic panel, including serum bilirubin determination, at her annual physical examination 2 months earlier. Her family history included relatives with rheumatoid arthritis and lupus erythematosus. On physical examination she was pale, with faintly icteric sclerae and a palpable spleen tip. Her Hgb was 7.6, Hct 23%, reticulocytes 14%, plasma total bilirubin 2.1 mg/dL, direct-reacting bilirubin 0.2 mg/dL, and indirect bilirubin 1.9 mg/dL. A lupus erythematosus preparation was negative, but because a direct Coombs test was strongly positive, the diagnosis of autoimmune hemolytic anemia was made. She consented to participate in a research study and was found to have a normal hepatic bilirubin clearance (C_{BR}) of 0.53 mL/min/kg (normal range: 0.65 ± 0.18 [SE]), a plasma bilirubin turnover of 14.4 mg/kg/day (normal: 3.9 ± 0.7), a circulating red blood cell mass of 15.3 mL/kg (normal: 25-35), and a calculated mean red blood cell lifespan of 16 days (normal in females: 94 ± 10 days). She was begun on prednisone, 60 mg/day. One week later her Hgb was 7.9 mg/dL, Hct 24%, reticulocytes 12.3%, and red blood cell mass 16.0 mL/kg. Her hepatic bilirubin clearance (C_{BR}) was 0.56 mL/min/kg, her plasma unconjugated bilirubin concentration was 0.5 mg/dL, and her calculated mean red blood cell lifespan was 62 days. There was a disagreement among the consultants on her case. Based on her persistent anemia and reticulocytosis, a rheumatologist argued that the current therapy was ineffective and urged increasing her prednisone dose to 80 mg/day. The hematologist said that she was essentially cured and urged continuation of the current treatment with follow-up in 2 weeks. When the patient returned in 2 weeks, her data fully confirmed the hematologist's prediction. Her Hgb was 12.0 g/dL, Hct 36%, reticulocytes 2.8%, and red blood cell mass 24.0 mL/kg. Plasma unconjugated bilirubin was 0.6 mg/dL. Which of the following test results led the hematologist to that correct conclusion?

A. The Hgb
B. The Hct
C. The reticulocyte count
D. The C_{BR}
E. The plasma unconjugated bilirubin concentration

Answer: E Hepatic bilirubin clearance (C_{BR}), plasma bilirubin turnover (BRT), and plasma unconjugated bilirubin concentration (BR) are related by the formula: $BR \approx BRT / C_{BR}$. In the present case, the plasma unconjugated bilirubin concentration fell by 76% during the first week of prednisone therapy, whereas C_{BR} was virtually unchanged. These data meant that the fall in plasma bilirubin concentration must have resulted from a decrease in plasma bilirubin turnover, which, in fact, also declined by 75%. Because the major source of bilirubin is the dying red blood cell, the conclusion is that RBC death must have declined by approximately the same 75%, precisely accounting for the nearly four-fold increase in calculated red blood cell lifespan. Thus, the fall in bilirubin concentration accurately reflected the beneficial effects of steroids on red blood cell survival in this patient with autoimmune hemolysis. By contrast, the continued reticulocytosis during the first week of prednisone treatment reflected the bone marrow's attempt to heal the ongoing anemia, not to the hemolytic process per se.

3. A 36-year-old man with a history of injection drug use is discovered to have hepatitis C and HIV coinfection during a medical screening examination. He has no signs or symptoms of chronic liver disease or portal hypertension, nor is there a history of AIDS-defining illnesses. Laboratory studies reveal albumin 4.2 mg/dL, AST 76 IU/mL, ALT 113 IU/mL, total bilirubin 0.6 mg/dL, and CD4 cells 450 cells/μL. Both HCV RNA and HIV RNA are detectable by PCR assays. Antiretroviral therapy is initiated with a combination of medications that include atazanavir. Evaluation following the start of antiretroviral therapy discloses mild scleral icterus. Laboratory results obtained following the start of treatment revealed albumin 4.0 mg/dL, AST 81 IU/mL, ALT 145 IU/mL, total bilirubin 2.4 mg/dL, direct bilirubin 0.3 mg/dL and haptoglobin 150 mg/dL. Which of the following is the most appropriate next step with regard to the jaundice?

A. Observation
B. Liver ultrasonography
C. Discontinuation of atazanavir
D. Liver biopsy
E. Magnetic resonance cholangiopacreatography

Answer: A Rates of HIV-HCV coinfection vary widely in populations but are highest in individuals with a history of injection drug use. The development of a mild indirect hyperbilirubinemia in the absence of evidence of hemolysis is strong evidence of Gilbert syndrome. Certain medications, including the antiretroviral protease inhibitor atazanavir, are inhibitors of UGT1A1 and also result in a mild, indirect hyperbilirubinemia similar to that seen in Gilbert syndrome. Atazanavir-associated hyperbilirubinemia may be more pronounced in individuals with Gilbert syndrome or with baseline hyperbilirubinemia from underlying liver disease. Although liver injury can occur with antiretroviral medications, there is no such factor in this case. Instead, the pattern is consistent with an acquired indirect hyperbilirubinemia related to atazanavir. No further diagnostic evaluation is warranted.

4. A 22-year-old woman is referred to your care for the evaluation of jaundice. She gives a history of episodic jaundice that is most pronounced around times of illness. On physical examination, she is well developed and healthy appearing. Scleral icterus is noted. The abdomen is soft, and there is neither shifting dullness nor organomegaly. Laboratory studies reveal hemoglobin 13 g/dL, platelets 225,000, alkaline phosphatase 95 IU/mL, AST 13 IU/mL, ALT 20 IU/mL, albumin 4.1 mg/dL, total bilirubin 2.4 mg/dL, direct bilirubin 1.5 mg/dL. Liver ultrasonography reveals cholelithiasis without gallbladder wall thickening. Which of the following is the next most appropriate diagnostic step?

A. Cholecystectomy with intraoperative cholangiography
B. Liver biopsy
C. Mutation analysis of the ABC transporter *ATP8B1*
D. Laboratory testing for antimitochondrial antibodies
E. Measurement of urinary coproporphyrin isomers

Answer: E Heritable disorders of bilirubin metabolism should be considered in the setting of a mixed or predominately direct hyperbilirubinemia in the absence of evidence of cholestasis. Dubin-Johnson syndrome is an autosomal recessive disorder arising from mutations in the ATP-dependent canalicular organic anion transporter MPR2/cMOAT. The diagnosis of Dubin-Johnson syndrome can be established by the measurement of urinary coproporphyrin isomers; the coproporphyrin isomer 1 is generally greater than 80% of total coproporphyrin concentration. Although liver biopsy is not required nor necessary for the diagnosis, the liver in Dubin-Johnson is described as containing a dark pigment comprising polymerized epinephrine metabolites.

5. A 34-year-old Hispanic woman presents for the preoperative evaluation for cholecystectomy. For the past 3 months, she has experienced episodic, colicky, right upper quadrant pain. On examination, she is obese (BMI 35 kg/m^2). She has no signs or symptoms of liver disease or portal hypertension. Laboratory evaluation reveals: ALT 44 IU/mL, AST 42 IU/mL, total bilirubin, alkaline phosphatase, and γ-glutamyl transpeptidase (GGT) are within the laboratory's reference range. Tests for chronic viral hepatitis, common genetic causes of chronic liver disease, and autoimmune liver disease are all negative. Her ferritin level is 400 ng/mL, and her transferrin saturation 27%. Liver ultrasonography reveals increased echogenicity of the liver and cholelithiasis without thickening of the gallbladder wall. In addition to cholecystectomy, which of the following options should be recommended?

A. *Hfe* genotype
B. Serologic studies for celiac disease
C. Nucleic acid testing for viral hepatitis
D. Intraoperative liver biopsy during cholecystectomy
E. Treatment with vitamin E (d-α) 800 IU daily.

Answer: D Liver diseases are commonly associated with aminotransferase elevations 2 or greater times the upper limits of the reference range. In some individuals, however, ALT levels at or within the upper limits of the reference range may be observed with nonalcoholic fatty liver disease (NAFLD). Thus, in patients at risk for NAFLD, it is imperative that aminotransferases be interpreted in the context of the patient's history, physical examination, and other laboratory findings. The presence of obesity and modest elevations in the ferritin level (without an increase in transferrin saturation) are commonly seen in subjects with NAFLD. Although the ferritin level is elevated, the transferrin saturation is not, so hereditary hemochromatosis is unlikely. Intraoperative liver biopsy is a safe and effective tool for diagnosing NAFLD and assessing the extent of fibrosis.

148 ACUTE VIRAL HEPATITIS

JEAN-MICHEL PAWLOTSKY

Infection with a hepatotropic virus causes an acute episode of liver inflammation, referred to as *acute hepatitis,* which can lead to either spontaneous clearance of the infectious agent or its persistence, which in turn leads to chronic infection for a subset of these viruses. Five hepatitis viruses are responsible for the vast majority of acute hepatitis cases (Table 148-1): hepatitis A virus (HAV); hepatitis B virus (HBV); hepatitis C virus (HCV); hepatitis D, or delta, virus (HDV), which is a defective viroid using the hepatitis B surface antigen (HBsAg) as its envelope; and hepatitis E virus (HEV).[1] Other viruses may cause acute inflammatory liver disease, including members of the Herpesviridae family such as human cytomegalovirus, Epstein-Barr virus, or herpes simplex virus. It is unclear to what extent other viruses, such as parvovirus B19 or human herpesvirus 6, can also cause acute hepatitis. Patients who present with an acute viral hepatitis syndrome but negative virologic tests are referred to as having non-A-to-E hepatitis, perhaps attributable to hepatotropic viruses that have yet to be identified. The worldwide incidence of acute viral hepatitis is decreasing because of global improvement in hygiene and the development and use of efficient vaccines against HAV and HBV, and perhaps in the future, HEV.

TABLE 148-1 VIRUSES RESPONSIBLE FOR ACUTE VIRAL HEPATITIS AND LIKELIHOOD OF CHRONIC EVOLUTION

VIRUS	EVOLUTION TO CHRONIC VIRAL HEPATITIS
Hepatitis A	Never
Hepatitis B	>90% (perinatal acquisition) to < 1% (adult infection)
Hepatitis C	50-80%
Hepatitis D or delta	2% (coinfection) to 90% (superinfection)
Hepatitis E	Occasionally in immunosuppressed patients
Other viruses Human cytomegalovirus Epstein-Barr Herpes simplex Human herpesvirus 6 Parvovirus B_{19}	May establish chronic infection, not associated with chronic hepatitis

GENERAL FEATURES OF ACUTE VIRAL HEPATITIS

PATHOBIOLOGY

Acute viral hepatitis is characterized by acute necroinflammation of the liver. Because none of the hepatotropic viruses is cytopathic, liver injury is mediated by a strong cytotoxic T cell–mediated reaction against infected hepatocytes that express viral antigens at their surface. Proinflammatory cytokines, natural killer cells, and antibody-dependent cellular cytotoxicity also appear to play a role in liver necroinflammation. Successful immune elimination may lead to viral clearance, which may or may not be associated with lifelong immunity, depending on the infecting agent. The immune reaction is sometimes so potent that the patient develops subfulminant or even fulminant hepatitis that requires liver transplantation (Chapter 154). In some patients—the proportion varies, according to the virus responsible for acute hepatitis—the immune response fails and chronic infection is established (Chapter 149).

CLINICAL MANIFESTATIONS

After infection, there is an incubation period of a few days to a few weeks, depending on the causative agent (Fig. 148-1). This incubation period is generally characterized by nonspecific symptoms, including fatigue, nausea, loss of appetite, flulike symptoms, and/or right upper quadrant pain (Table 148-2). The incubation period is often characterized by leukopenia and relative lymphocytosis. Immune-mediated symptoms, including rash, hives, arthralgias, angioneurotic edema, and fever, are observed in 10 to 20% of patients during the preicteric phase.

During the acute stage of the disease, symptoms may vary widely, from asymptomatic to subicteric, icteric or severe, and fulminant. The icteric form, which is not frequent, is characterized by fatigue, anorexia, nausea, dysgeusia, jaundice, dark urine, light-colored stool, and weight loss. Physical examination reveals jaundice and hepatic tenderness. Hepatomegaly and splenomegaly may be present. On laboratory testing, acute viral hepatitis is characterized by elevated total and direct serum bilirubin levels and aminotransferase levels that are often greater than 10 times the upper limit of normal. Cholestatic acute hepatitis is associated frequently with prolonged and fluctuating jaundice and pruritus. After 1 to 3 weeks, on average, both clinical and laboratory signs progressively improve and return to normal. Some patients, however, may experience a relapse before definitive resolution.

Signs of hepatic failure (Chapter 154), including changes in personality, aggressive behavior, sleeping disorders, and hepatic encephalopathy

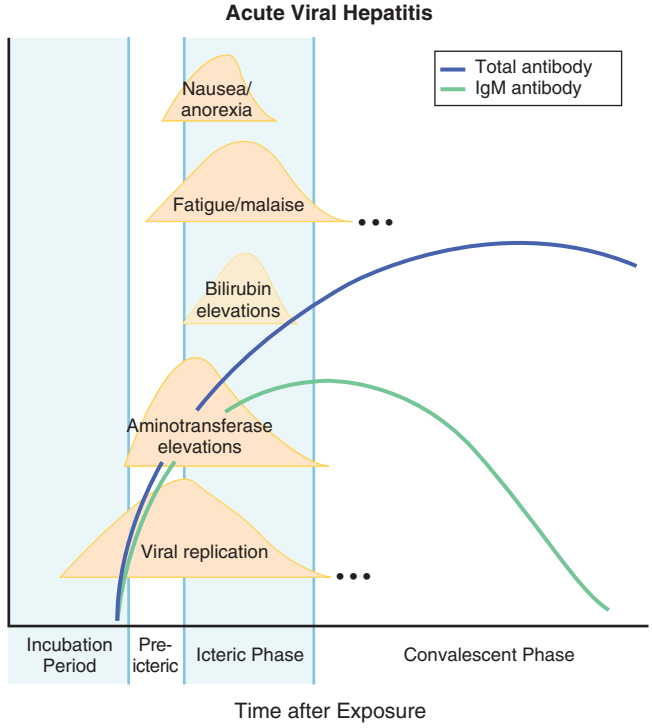

FIGURE 148-1. Typical course of acute viral hepatitis. IgM = immunoglobulin M.

TABLE 148-2 CLINICAL MANIFESTATIONS OF VIRAL HEPATITIS

PHASES OF INFECTION	DURATION*	MANIFESTATIONS*
Incubation	2-20 wk	Virus detectable in blood Aminotransferase and bilirubin levels normal No antibody detectable
Preicteric	3-10 days	Nonspecific symptoms: fatigue, anorexia, nausea, vague right upper quadrant pain Viral titers peak Aminotransferase levels begin to rise Serum sickness–like reaction (≈10-20% of cases) with rash, hives, arthralgias, fever
Icteric	1-3 wk	Jaundice appears; dark urine and light stools seen Nonspecific symptoms worsen Weight loss, dysgeusia, pruritus may occur Hepatosplenomegaly may develop Aminotransferase levels typically >10 times normal Antibodies appear Viral titers decline Rare extrahepatic manifestations (aseptic meningitis, encephalitis, seizures, ascending flaccid paralysis, nephrotic syndrome, seronegative arthritis)
Recovery	Up to 6 mo	Symptoms resolve gradually Antibody levels rise Aminotransferase and bilirubin levels normalize
Chronic	After 6 mo	See Chapter 149

*Varies by virus.

TABLE 148-3 LABORATORY TESTING FOR SUSPECTED ACUTE VIRAL HEPATITIS

GENERAL EVALUATION
Alanine aminotransferase Aspartate aminotransferase Alkaline phosphatase International normalized ratio (INR)
FIRST-LINE DIAGNOSTIC TESTS
Anti-HAV IgM HBsAg, anti-HBc IgM Anti-HCV antibodies HCV RNA
SECOND-LINE DIAGNOSTIC TESTS
Anti-HAV IgM present: none HBsAg present: HBeAg, anti-HBe antibodies, HBV DNA, HDV antigen, anti-HDV antibodies HCV RNA present with or without anti-HCV antibodies: none No virologic marker: See Fig. 147-3

HAV = hepatitis A virus; HBc = hepatitis B core; HBeAg = hepatitis B e antigen; HBsAg = hepatitis B surface antigen; HBV = hepatitis B virus; HCV = hepatitis C virus; HDV = hepatitis D virus; IgM = immunoglobulin M.

characterize fulminant forms of acute viral hepatitis. Coma can supervene rapidly, and widespread hemorrhage may develop.

DIAGNOSIS

The diagnosis of acute hepatitis is suspected based on elevated serum aminotransferase levels, which are generally more than 10 times the upper limit of normal (Table 148-3). Total and direct bilirubin levels are elevated if the acute hepatitis is subicteric or icteric. Alkaline phosphatase levels may be elevated in cases of cholestatic hepatitis.

Serologic and eventually molecular testing identifies the causal agent. Liver biopsy or a noninvasive assessment of liver inflammation and fibrosis are generally not required. All cases of acute hepatitis should be reported to the local, state, or national health department as soon as possible after diagnosis.

TREATMENT Rx

In addition to virus-specific therapy for HCV and hopefully in the future for HBV, patients with acute viral hepatitis should avoid alcohol consumption and acetaminophen. Sexual contact should be avoided if the partner is not protected. Patients with subfulminant or fulminant hepatitis should be evaluated for possible liver transplantation and supported in an intensive care unit setting (Chapter 154).

PROGNOSIS

Prognosis depends on the degree of prolongation of the prothrombin time, as well as the degree of elevation of the bilirubin and lactate levels. A factor V level less than 40% or any signs of encephalopathy are indications for hospitalization. Death is extremely rare and occurs only in fulminant cases. Other signs of poor prognosis are persistently worsening jaundice, ascites, and an acute decrease in the size of the liver. Serum aminotransferase levels and viral genome levels have no prognostic value.

HEPATITIS A

DEFINITION

The Pathogen

HAV is a member of the Picornaviridae family, genus *Hepatovirus*. The hepatitis A viral particle is a 27-nm nonenveloped icosahedral nucleocapsid that expresses the hepatitis A antigen and contains a positive-stranded RNA genome approximately 7.5 kb long. At least four different HAV genotypes have been described in humans (genotypes I, II, III, and VII), with genotype I predominating worldwide. Other genotypes have been isolated in nonhuman primates. It is currently unclear to what extent different genotypes are associated with distinct clinical courses of infection.

EPIDEMIOLOGY

HAV infection has a worldwide distribution, and infections can be sporadic or occur in epidemic outbreaks. The incidence of acute cases and the seroprevalence vary according to the hygiene, sanitation, housing, and socioeconomic standards of the region, with seroprevalences as low as approximately 13% in Sweden but up to 100% in many developing countries. In developing countries, infection generally occurs at a young age and most of the population has been exposed and is protected after age 10 years. In developed countries, however, infection can occur at any age, and the prevalence of exposed, immune subjects slowly increases with age. In the United States, according to the Centers for Disease Control and Prevention, the incidence of acute hepatitis A declined from 12.0 cases per 100,000 individuals in 1995 to 0.5 cases per 100,000 individuals by 2012.

HAV is generally transmitted via the oral-fecal route, most often directly from person to person or through the ingestion of fecally contaminated food or water. Transmission by blood transfusion has been reported, and isolated cases of apparent perinatal transmission have been described. High-risk groups for acute hepatitis A include travelers to developing countries, children in daycare centers and their parents, men who have sex with men, injection drug users, hemophiliacs who receive plasma products, and persons in institutions.

PATHOBIOLOGY

The genome serves as a messenger RNA and contains a single open reading frame that encodes both structural and nonstructural viral proteins. After attachment to a specific receptor at the surface of hepatocytes, the virus penetrates into cells and is uncoated. Subsequent events occurring exclusively in the cytoplasm include translation of the single open reading frame into a polyprotein that is later processed to generate the mature viral proteins; replication in a membrane-bound replication complex that generates new viral genomes, which are subsequently used for viral protein production and viral particle assembly; and packaging of newly formed genomes into new particles that are exported out of the cells. The virus is secreted into bile and, to a lesser extent, serum.

CLINICAL MANIFESTATIONS

Typically, the incubation period is 15 to 45 days (see Table 148-2). In most cases, acute infection takes a mild and often unrecognized course. The incidence of symptomatic, icteric cases increases with the age at infection. Acute hepatitis A in adults may require hospitalization in up to 13% of cases[2]; prolonged courses of 6 to 9 months have been reported in 10% of adult patients with a diagnosis of acute hepatitis A. Hepatitis A is the most common cause of relapsing cholestatic hepatitis.

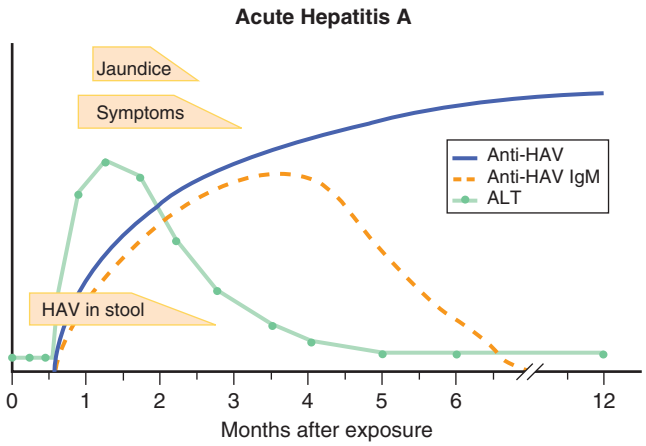

FIGURE 148-2. Serologic course of acute hepatitis A. ALT = alanine aminotransferase; HAV = hepatitis A virus; IgM = immunoglobulin M.

DIAGNOSIS

The diagnosis of acute hepatitis A is based on the detection of anti-HAV immunoglobulin M (IgM) in serum by enzyme immunoassay. IgM serum levels peak during the second month of infection (Fig. 148-2). HAV RNA can be transiently detected in stool and other body fluids by polymerase chain reaction (PCR) 3 to 10 days before the onset of illness and for 1 to 2 weeks thereafter; however, HAV RNA testing is generally not necessary. When the infection resolves, anti-HAV IgM disappears after 4 to 12 months, but anti-HAV IgG persists for life and confers definitive and durable protection against infection.

TREATMENT

Rx

Because HAV infection is self-limited, no specific antiviral treatment is required. In severe cases, patients may need to be hospitalized. If liver function is deteriorating, patients may need to be assessed for liver transplantation, which is the only therapeutic option for acute liver failure (Chapter 154).

PREVENTION

HAV vaccines (Chapter 18) consist of inactivated hepatitis A antigen purified from cell culture. Two doses of the vaccine are recommended at a 6- to 18-month interval. All vaccines are highly immunogenic, and virtually all healthy persons who are vaccinated develop protective anti-HAV antibodies. Patients with chronic liver disease also respond to vaccination but may display lower anti-HAV titers. An accelerated vaccine schedule, with vaccination on days 0, 7, and 21, is also effective and may be recommended for those planning to travel to endemic areas. HAV vaccines are well tolerated, and no serious adverse events have been linked with their administration; they can be safely administered with other vaccines or immunoglobulins without compromising the development of protective antibodies. A combination HAV and HBV vaccine is also available. Seroconversion rates are lower in patients with human immunodeficiency virus (HIV) infection and in other immunocompromised individuals.

HAV vaccination is recommended for nonimmune individuals who plan to travel to endemic countries, medical health professionals, men who have sex with men, persons who are in contact with hepatitis A patients, and individuals with chronic liver diseases. A childhood vaccination program leads to a significant decline in HAV infection, justifying its use as part of control efforts in endemic countries. Serologic testing for anti-HAV IgG can be performed before vaccination in adults born in endemic countries and in individuals older than 50 years born in industrialized areas; persons with detectable IgG are protected and should not be vaccinated. For post-exposure prophylaxis, both HAV vaccination and immunoglobulin are effective. Immunoglobulin confers a slightly lower rate of protection than vaccine,■ and HAV vaccine and immunoglobulin should be used together in this setting.

Long-term follow-up studies after complete HAV vaccination show that anti-HAV titers sharply decline during the first year after vaccination but remain detectable in almost all individuals for at least 10 years. Protective anti-HAV antibody titers persist for at least 27 years after the successful vaccination of children and young adults.

PROGNOSIS

Acute hepatitis A infection generally resolves without complications in 3 to 4 weeks and never evolves to chronic infection. Prolonged elevation of serum aminotransferase levels has been reported. Relapses a few weeks after the acute case also have been observed. Prolonged courses may occur in children and immunosuppressed individuals.

Cholestatic hepatitis A is unusual and has a good prognosis, with full recovery within a few weeks. Fulminant hepatitis A is rare, occurring in less than 0.1% of cases, but its incidence and mortality increase with the patient's age at acquisition. In the United States, 4% of all cases of fulminant hepatitis are caused by HAV infection. Overall, mortality of acute hepatitis A is 1.8% in patients older than 50 years. In patients with chronic hepatitis B, superinfection with HAV is associated with a 6- to 23-fold higher morbidity and mortality.

ACUTE HEPATITIS B

DEFINITION

The Pathogen

HBV is a member of the Hepadnaviridae family, genus *Hepadnavirus*. The infectious virion, the Dane particle, is 42 to 47 nm in diameter. It possesses an envelope and a capsid or core that contains the partially double-stranded, circular DNA genome. The HBV genome is the smallest known human virus genome, with approximately 3000 nucleotides.

EPIDEMIOLOGY

Two billion individuals worldwide have been in contact with HBV, and more than 350 million individuals have chronic infection. In the United States, approximately 0.5% of the population is chronically infected. HBV virions are produced and circulate in very high amounts in HBV-infected individuals, who are highly contagious. Four principal routes of transmission are responsible for acute HBV infections: (1) sexual transmission, which is the principal route in industrialized areas; (2) perinatal mother-to-infant transmission, which is associated with a very high (>90%) rate of chronic infection and is the principal cause of HBV transmission in Asia; (3) horizontal transmission through nonsexual interindividual contact, which is frequent at a young age in Africa and is associated with evolution to chronicity in approximately 15% of cases; and (4) percutaneous transmission by blood and blood products, unsafe medical or surgical materials, or injection drug use.

In industrialized countries, groups at high risk for HBV infection include individuals born in areas where HBV is endemic, including immigrants and adopted children; individuals who were not vaccinated as infants and whose parents were born in regions where HBV is endemic; household and sexual contacts of HBsAg-positive individuals; persons who have ever injected drugs; persons with multiple sexual partners or a history of sexually transmitted disease; men who have sex with men; inmates of correctional facilities; patients infected with HCV or HIV; patients undergoing renal dialysis; recipients of blood or blood products before 1987; and health care workers.

PATHOBIOLOGY

The HBV genome contains at least four overlapping open reading frames that encode a number of structural and nonstructural viral proteins. The pre-S/S gene encodes the three surface proteins—small (S), middle (M), and large (L)—that express HBsAg. The pre-C/C gene encodes the core protein that expresses the hepatitis B core (HBc) antigen and the hepatitis B e (HBe) protein, a nonstructural protein that plays a role in immune tolerance to HBV replication. The P gene encodes the HBV polymerase, whose two motifs—a reverse transcriptase motif and an RNAse H motif—code for two enzymes involved in HBV replication. Finally, the X gene encodes the X protein, which is a transactivator involved in HBV replication that bears oncogenic properties. The blood of infected patients contains not only infectious viruses but also a large excess of empty, noninfectious HBV envelopes.

The complex HBV life cycle involves multiple steps: fixation to an as yet unidentified receptor complex at the surface of hepatocytes; internalization; fusion and release of the nucleocapsid containing the HBV DNA genome and the associated HBV polymerase molecule in the cell cytoplasm; transport into the nucleus, where decapsidation occurs and the DNA genome molecule is released; transformation of the viral genome by the viral polymerase into a covalently closed circular DNA, which is the episomal form

responsible for persistence of the HBV genome in the nucleus of infected hepatocytes; generation of messenger RNAs and viral protein synthesis; generation of a pregenomic RNA, which serves as a template for reverse transcription that generates the long DNA strand; degradation of the pregenomic RNA by the RNAse H activity of the viral polymerase; DNA-dependent DNA polymerase activity of the reverse transcriptase motif, which synthesizes the short complementary DNA strand in newly formed nucleocapsids; and, finally, budding into the endoplasmic reticulum, maturation, and export of newly formed virions.

Nine HBV genotypes (A through I), which differ by approximately 8% of their genomic nucleotide sequence, have different geographic distributions and may be associated with different clinical outcomes. Genotype A predominates in Northern and Western Europe, whereas genotype D is the most frequent genotype in the Mediterranean area and in Eastern Europe. In non-Asian populations in the United States, genotype A predominates in men who have sex with men, whereas genotype D is the most frequent in intravenous drug users. In Asia and in Asian immigrants living in industrialized countries, genotypes B and C predominate. Genotype C has been associated with a higher incidence of severe liver disease and hepatocellular carcinoma compared with genotype B in Asia, perhaps because this genotype spread earlier than the others.

CLINICAL MANIFESTATIONS

Typically, the incubation period is 30 to 150 days. Jaundice has been reported in up to one third of adult patients with acute hepatitis B, but most cases are unrecognized.[3] Among symptomatic patients (see Table 148-2), the manifestations are similar to those of other causes of acute viral hepatitis.

DIAGNOSIS

Four markers should be sought for the diagnosis of acute hepatitis B: HBsAg, total anti-HBc antibodies, anti-HBc IgM, and anti-HBs antibodies (Table 148-4). Acute hepatitis B is characterized by the simultaneous presence of both HBsAg and anti-HBc IgM (Fig. 148-3). Total anti-HBc antibodies are also present, whereas anti-HBs antibodies are not. During the convalescence phase, patients lose HBsAg before the appearance of anti-HBs antibodies; they have isolated anti-HBc antibodies, and the diagnosis is based on the presence of anti-HBc IgM. Recovery is characterized by the appearance of anti-HBs antibodies. The presence of both total anti-HBc and anti-HBs antibodies characterizes recovery from acute hepatitis B. Anti-HBc IgG remains at high levels for the patient's lifetime, whereas anti-HBs antibody titers may fluctuate and sometimes become undetectable after several years. Quantitative HBsAg assessment may be useful during the course of acute hepatitis B because if the HBsAg level does not rapidly decrease, the patient is at risk for chronic evolution. Patients who remain positive for HBV DNA or HBeAg 6 weeks after the onset of symptoms are likely to develop chronic infection.

TABLE 148-4 SEROLOGIC PROFILES OBSERVED IN DIFFERENT PHASES OF ACUTE, SELF-RESOLVING HEPATITIS B

PHASE OF INFECTION	HBsAg	ANTI-HBc IgM	TOTAL ANTI-HBc ANTIBODIES	ANTI-HBs ANTIBODIES
Incubation	+	+/–	+/–	–
Acute hepatitis	+	+	+	–
Convalescence	–	+	+	–
Recovery	–	–	+	+

HBc = hepatitis B core; HBs = hepatitis B surface; HBsAg = hepatitis B surface antigen; IgM = immunoglobulin M.

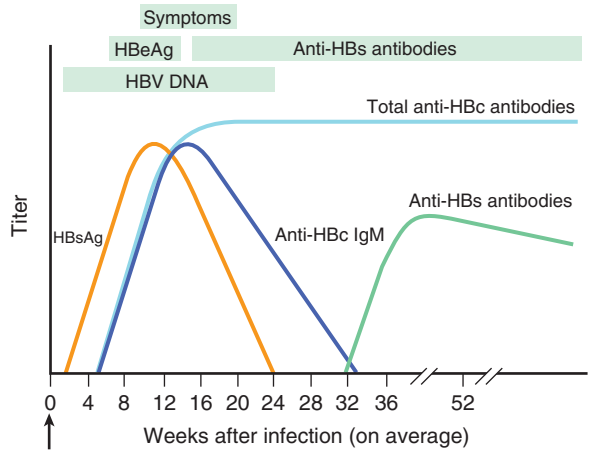

FIGURE 148-3. Kinetics of hepatitis B virus (HBV) markers during acute self-resolving hepatitis B. The *arrow* indicates infection. HBc = hepatitis B core; HBeAg = hepatitis B e antigen; HBs = hepatitis B surface; HBsAg = hepatitis B surface antigen; IgM = immunoglobulin M.

TREATMENT

Rx

Acute HBV infection usually does not require antiviral therapy, and most patients spontaneously clear the infection. Antiviral therapy with lamivudine can decrease HBV DNA levels more rapidly but does not result in better clinical or biochemical improvement and may be associated with lower levels of protective anti-HBs at 1 year.[A] In a small randomized trial of patients with severe but nonfulminant acute hepatitis B, lamivudine (100 mg daily) led to more rapid viral clearance but no difference in clinical outcomes.[A] Nevertheless, nonrandomized data suggest that early antiviral therapy is safe and may reduce the need for liver transplantation in patients with fulminant hepatitis B. Although most of the experience has been with lamivudine (100 mg daily), more potent drugs with no risk for HBV resistance selection, such as tenofovir 300 mg daily or entecavir 0.5 mg daily, are recommended in this setting.[4]

PREVENTION

Because not everyone has been vaccinated, individuals who are aware that they are infected with HBV should take steps to avoid transmitting the infection to others. This is accomplished by ensuring that their sexual contacts and household members are vaccinated; using barrier protection during sexual intercourse; not sharing instruments such as toothbrushes, razors, and combs; cleaning blood spills with detergent or bleach; and not donating blood, organs, or sperm.

Prevention of HBV infection is based on vaccination. Universal infant vaccination programs have been initiated in many countries. High-risk individuals (e.g., health care workers, dialysis patients, family members and sexual partners of HBV carriers, pregnant women, and men who have sex with men) should be screened for HBV infection by HBsAg and anti-HBs antibody testing, and seronegative persons should be vaccinated.[5]

Vaccination consists of the administration of recombinant HBsAg in three injections at 0, 1, and 6 months in adults. A lower dose is given at the same time points in newborns, children, and adolescents. Adults on dialysis require four injections at months 0, 1, 2, and 6. HBV vaccination elicits a potent neutralizing response, characterized by the presence of anti-HBs antibodies at high titers. A titer greater than 10 U/L is considered protective. The seroconversion rate is higher than 90% in healthy individuals. HBV vaccines are well tolerated. Injection site reactions within 1 to 3 days, as well as mild general reactions, are common and transient. Postvaccination testing for anti-HBs antibodies to document seroconversion is not routinely recommended. However, persons who remain at risk for HBV infection, such as infants of HBsAg-positive mothers, health care workers, dialysis patients, and sexual partners of HBV carriers, should be tested to determine their response to vaccination.

Of vaccinated persons, 3 to 10% respond poorly or do not respond, especially smokers, obese patients, and elderly individuals. Nonresponders should receive another full course of vaccination, often with an increased dose. Other options include intradermal application and the coadministration of adjuvants and cytokines. One third to two thirds of vaccinated individuals lose their anti-HBs antibodies after 10 to 15 years. It is unclear whether these persons are still protected. Thus, persons at risk should receive booster immunization if anti-HBs antibodies have been lost.

When nonimmune persons or vaccinated individuals with an anti-HBs titer below 10 IU/L have contact with HBV-contaminated materials (e.g., needles) or have sexual intercourse with an HBV-infected person, active-passive immunization (i.e., infusion of hepatitis B immunoglobulin) plus active immunization (vaccination) is recommended within 48 hours after exposure. Vaccination alone is sufficient in persons with anti-HBs antibody titers between 10 and 100 IU/L, and no action is required if the anti-HBs titer is above 100 IU/L.

Infants born to HBsAg-positive mothers must receive hepatitis B immunoglobulin and be vaccinated within 12 hours of birth; this regimen reduces the rate of vertical HBV transmission from 95% to less than 5%. Mothers with

HBV DNA levels above 5×10^7 IU/mL should also be treated during pregnancy with a potent nucleoside/nucleotide analogue with no teratogenic risk; the safest and most potent drug in this setting is tenofovir. Cesarean section is not needed if active-passive immunization is to be performed. Mothers of vaccinated infants can breast-feed, unless oral antiviral medications are present in the breast milk.

PROGNOSIS

Fulminant hepatitis is more frequent in acute HBV infection than in other types of acute viral hepatitis, with an incidence of approximately 0.1%. Factors associated with adverse outcomes of acute hepatitis B include advanced age, female sex, and perhaps some strains of virus. Whether infection with a precore mutant strain is associated with more severe or fulminant disease is still debated.

Among patients infected at birth, the rate of spontaneous recovery after an acute HBV infection is less than 5%, whereas adult infections spontaneously resolve in 95 to 99% of cases. Spontaneous resolution confers lifelong immunity, which is usually characterized by the presence of anti-HBs antibodies. Anti-HBs antibodies may become undetectable several years after resolution, but patients rapidly produce protective antibodies if they are re-exposed to HBV.

ACUTE HEPATITIS C

DEFINITION

The Pathogen

HCV is a member of the Flaviviridae family, genus *Hepacivirus*. The genome is a single-stranded, positive, linear RNA molecule whose 5′ end contains an internal ribosome entry site involved in polyprotein translation; the genome also includes one single open reading frame and a short 3′ noncoding region involved in replication. The genome is contained in a protein capsid or core, which itself is surrounded by a lipid bilayer envelope into which two inserted viral glycoproteins mediate attachment of the viral particle to receptor molecules at the surface of target cells.

EPIDEMIOLOGY

HCV is present in all continents, and an estimated 170 million individuals are chronically infected. In industrialized countries, the incidence of HCV infection has declined considerably owing to blood screening and measures to prevent viral infections in intravenous drug users. However, approximately 17,000 new cases of acute hepatitis C still occur annually in the United States according to the Centers for Disease Control and Prevention. In France, approximately 2500 new infections occur each year. HCV incidence and prevalence are higher in developing areas of the world, where the main route of HCV infection is unsafe medical or surgical procedures; only approximately 50% of blood products are screened for anti-HCV antibodies in these countries, and approximately 40% of all injections are given with reused equipment. Egypt's estimated 9% prevalence is the highest worldwide, owing to unsafe injection campaigns for the treatment of schistosomiasis. Although the incidence of acute hepatitis C has declined in Egypt during the past 15 years, up to 10% of cases of acute hepatitis are still caused by HCV.

HCV is transmitted almost exclusively by infected blood. Preventive screening with highly sensitive enzyme immunoassays and, more recently, nucleic acid testing has virtually eliminated the risk for post-transfusional HCV infection (theoretical risk: 1 in 2 million donations in the United States, 1 in 8 million donations in France). As a result, the principal route for HCV transmission in industrialized countries is now intravenous drug use, which is responsible for 60 to 80% of new cases. The incidence of HCV infection in this high-risk group is as high as 39 per 100 person years. In this context, imprisonment is an important risk factor for acquiring HCV infection in industrialized countries.

Nosocomial transmission through the use of improperly decontaminated materials or the contaminated hands or gloves of health care workers is responsible for a substantial number of new infections worldwide. HCV also can be transmitted by tattooing, piercing, or acupuncture if standard precautions are not implemented. Although HCV can be acquired after accidental needlestick exposure, the risk for infection is low (<1%), and health care workers have only a slightly higher prevalence of HCV than the general population. HCV can be transmitted to household members who share instruments such as scissors, razors, and combs. Sexual transmission is unusual, but outbreaks of acute hepatitis have been reported in HIV-positive communities of men who have sex with men. The risk for mother-to-infant transmission of HCV is less than 5% and is generally related to exposure to the mother's blood in the perinatal period. Cesarean sections are not recommended, and breast-feeding is not contraindicated. The risk for perinatal transmission is higher when the mother is coinfected with HIV. Other factors possibly associated with high transmission rates include the level of HCV viremia and maternal intravenous drug abuse. In 10 to 30% of cases, no apparent risk factor for HCV infection is identifiable, suggesting other potential sources of community-acquired hepatitis C.

PATHOBIOLOGY

Entry of HCV into cells is followed by fusion. Decapsidation of viral nucleocapsids liberates free genomic RNAs into the cell cytoplasm, where they serve, together with newly synthesized RNAs, as messenger RNAs for synthesis of the HCV polyprotein. The post-translational processing of the HCV polyprotein results in the generation of at least 10 proteins, including 3 structural proteins (the core protein and the two envelope glycoproteins) and 7 nonstructural proteins. HCV replication takes place in the replication complex that associates viral proteins, cellular components, and nascent RNA strands. It is catalyzed by the RNA-dependent RNA polymerase. Viral particle formation is initiated by the interaction of the core protein with genomic RNA. Newly produced virus particles leave the host cell via constitutive secretory pathways.

Phylogenetic analyses of HCV strains isolated in various regions of the world have identified seven main HCV genotypes, designated 1 through 7. These HCV types comprise a large number of subtypes, identified by lowercase letters (1a, 1b, etc.). The genotypes' nucleotide sequences differ by 31 to 34%, and their amino acid sequences differ by approximately 30%; in contrast, the subtypes' nucleotide sequences differ by between 20 and 23%, with marked differences in particular genomic regions. The high prevalence and diversity of HCV genotype 3 and 6 strains in Asia and of genotype 1, 2, 4 and 5 strains in Africa suggest that these types and subtypes emerged and diversified in these regions. In industrialized countries, a small number of HCV genotypes, including 1a, 1b, 2a, 2b, 2c, 3a, and 4a, have been introduced and spread rapidly among exposed populations. Subtypes 1a and 1b predominate all over the world. The most common genotypes in the United States are 1a and 1b (~75%), 2a and 2b (~15%), and 3a (~7%). Genotype 3a is more prevalent in Western Europe, where it accounts for up to 35% of cases, especially among intravenous drug users. Genotype 4 is highly prevalent in the Middle East and Africa. Its incidence and prevalence are increasing in intravenous drug users in industrialized countries. Genotype 5 is rare outside South Africa, and genotype 6 is rare outside Southeast Asia. Infections with different genotypes do not differ in terms of the clinical manifestations, progression, or disease severity (although this is debated), but the HCV genotype is an important determinant of the response to interferon-α–based therapies.

CLINICAL MANIFESTATIONS

HCV RNA becomes detectable in the serum 3 to 7 days after exposure. HCV RNA levels rise rapidly during the first weeks, followed by serum aminotransferase levels 2 to 8 weeks after exposure. Anti-HCV antibodies arise late in the course of acute hepatitis C and may not be present at the onset of symptoms and serum aminotransferase elevation.

After an incubation period that ranges from 15 to 120 days, acute hepatitis C usually remains asymptomatic and is undiagnosed.[6] Nonspecific symptoms such as fatigue, low-grade fever, myalgias, nausea, vomiting, or itching may be present. Jaundice occurs in only 20 to 30% of patients, usually 2 to 12 weeks after infection. Serum aminotransferase levels commonly exceed 10 times the upper limit of normal in the acute stage, even in the absence of symptoms. Fulminant hepatitis C has been reported but appears to be exceptional in the absence of another chronic underlying liver disease.

DIAGNOSIS

When acute hepatitis C is suspected, patients should be tested for both anti-HCV antibodies by enzyme immunoassay and HCV RNA with a sensitive molecular biology technique (i.e., an HCV RNA assay with a lower limit of detection of ≤ 50 IU/mL). Four marker profiles can be observed, based on the presence or absence of either marker (Table 148-5). The presence of HCV RNA in the absence of anti-HCV antibodies is strongly indicative of acute HCV infection, which will be confirmed by seroconversion (i.e., the appearance of anti-HCV antibodies) a few days to weeks later. Acutely infected patients can have both HCV RNA and anti-HCV antibodies at the time of diagnosis; in this case, it is difficult to distinguish acute hepatitis C from an acute exacerbation of chronic hepatitis C or acute hepatitis of another cause in a patient with chronic hepatitis C.

TABLE 148-5 PATTERNS OF HEPATITIS C VIRUS (HCV) MARKERS AND THEIR SIGNIFICANCE DURING ACUTE HEPATITIS C

ANTI-HCV ANTIBODIES	HCV RNA	DIAGNOSIS
–	–	Not acute hepatitis C
–	+	Acute hepatitis C
+	–	Probably not acute hepatitis C (retest in a few weeks)
+	+	Difficult to differentiate acute from chronic hepatitis C

HCV = hepatitis C virus.

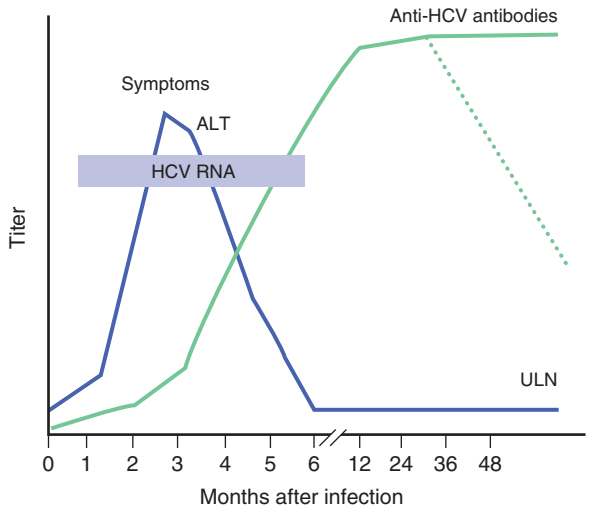

FIGURE 148-4. Kinetics of hepatitis C virus (HCV) markers during acute self-resolving hepatitis C. ALT = alanine aminotransferase; ULN = upper limit of normal.

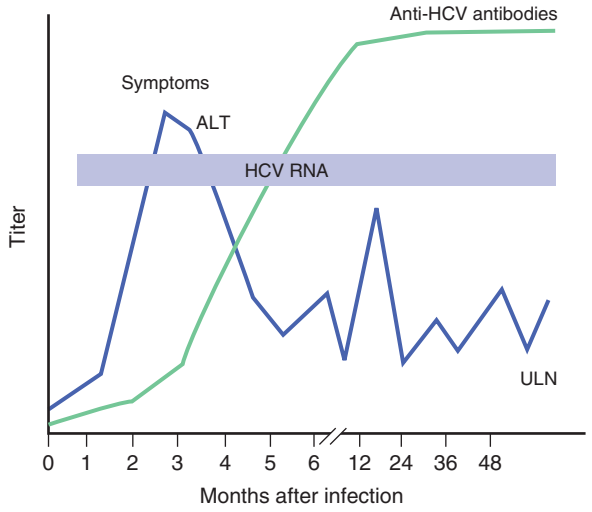

FIGURE 148-5. Kinetics of hepatitis C virus (HCV) markers during acute hepatitis C that evolves toward chronic infection. ALT = alanine aminotransferase; ULN = upper limit of normal.

Acute hepatitis C is very unlikely if both anti-HCV antibodies and HCV RNA are absent or if anti-HCV antibodies are present without HCV RNA (Fig. 148-4). Patients in the latter group should be retested in a few weeks, however, because HCV RNA may be temporarily undetectable owing to transient, partial control of viral replication before the infection becomes chronic (Fig. 148-5). Apart from such cases, the presence of anti-HCV antibodies in the absence of HCV RNA is generally seen in patients who have recovered from a past HCV infection. Nevertheless, this pattern cannot be differentiated from a false-positive enzyme immunoassay result, the exact prevalence of which is unknown.

TREATMENT

Rx

Treatment of acute hepatitis C should be considered not only because it prevents chronic HCV infection, which can lead to serious clinical sequelae, but also because HCV viremia, which is associated with a risk for transmitting HCV to other persons, may have social, legal, and economic consequences, especially for infected health care workers. Classically, treatment of acute hepatitis C is based on the use of pegylated interferon-α2a or interferon-α2b at doses of 180 µg/week or 1.5 µg/kg/week, respectively. Treatment is usually administered for 24 weeks, but 12 weeks is probably sufficient in most patients, especially patients with baseline parameters that predict a rapid viral clearance, such as symptomatic disease, high alanine aminotransferase levels, and low HCV RNA levels. Ribavirin should be added in patients who have a delayed or slow HCV RNA decrease with treatment, genotype 1 infection, low or normal baseline alanine aminotransferase levels, or a combination of these. With this treatment, 70% to more than 90% of patients have sustained viral clearance. Although delaying treatment can avoid the side effects of treatment in the 20 to 50% of patients who will spontaneously clear the virus, immediate treatment provides higher rates of viral clearance, mostly because asymptomatic patients tend to be less adherent with the close following needed to monitor patients for the institution of appropriately timed delayed therapy.

If treatment fails to clear the virus, patients can be re-treated with a combination of pegylated interferon-α2a (180 µg/week) or interferon-α2b (1.5 µg/kg/week) and ribavirin (0.8 g/day) for 48 weeks. No randomized data are available for new anti-HCV drugs. However, given its efficacy in patients with chronic hepatitis C, the oral combination of sofosbuvir 400 mg daily and ribavirin (1 or 1.2 g/day according to body weight less or more than 75 kg) for 24 weeks can be used to treat acute hepatitis C. More data will be generated soon with other antiviral drug combinations without ribavirin.

After accidental needlestick exposure, neither immunoglobulin nor preemptive therapy is recommended. Patients should be monitored with HCV RNA and aminotransferase level testing at baseline, week 2, and 4 and 6 months after exposure. Patients who have documented infection should be treated as described earlier.

PREVENTION

Prevention of HCV transmission is based on standard precautions such as screening of blood and blood products, application of standard medical and surgical hygiene procedures, and safe use of syringes and materials for drug preparation in drug users. Needle exchange programs and education regarding the risks of drug use (including intranasal cocaine) and the risk for transmission from shared injection equipment are important.

No prophylactic vaccine is or will soon be available against HCV.

PROGNOSIS

Approximately 50 to 80% of patients are unable to clear HCV spontaneously and develop chronic infection.[7] Spontaneous recovery is more frequent if infection is acquired at birth (~50%) and if acute hepatitis is symptomatic. It is unclear whether the genotype influences recovery rates. Other factors associated with better rates of spontaneous recovery include female sex, early decline of HCV RNA levels, and high aminotransferase or bilirubin levels. Patients who spontaneously recover may retain detectable anti-HCV antibodies for years to decades, but they are not protected against HCV reinfection.

ACUTE HEPATITIS D OR DELTA

DEFINITION

The Pathogen

HDV, which is a satellite of HBV, can be transmitted only to patients who are acutely or chronically infected with HBV. Its single, negative-strand, circular RNA genome of approximately 1700 nucleotides folds in native conditions into a nearly complementary rodlike structure that contains a ribozyme. The HDV genome encodes one single structural protein, the hepatitis D (HD) protein, which expresses HDAg. The 36-nm infectious HDV virion comprises the HD protein and the genome, both enclosed within an HBsAg coat derived from empty HBV envelopes.

EPIDEMIOLOGY

Five percent of chronic HBV carriers, or 15 million to 20 million individuals worldwide, are also infected with HDV. The prevalence of HDV infection in

HBV-infected patients varies according to the geographic area because it is transmitted primarily through parenteral exposure. As a result, its prevalence is relatively higher in HBsAg-positive intravenous drug users in Western countries, where approximately 8 to 12% of HBsAg-positive patients are infected with HDV. By comparison, the prevalence of HDV has decreased substantially in southern Europe, probably owing to universal HBV vaccination programs, improvement in hygiene and living conditions, and implementation of standard precautions to prevent HIV infection. The incidence of HDV is increasing in Russia, Eastern Europe, Japan, and India.

PATHOBIOLOGY

HDV uses host RNA polymerase II for its replication, following the rolling circle model. Within cells, HDV RNA is associated with multiple copies of the HD protein to form a ribonucleoprotein complex. This complex is exported by the HBV envelope, which contains the three HBV envelope proteins, into the Golgi apparatus before being secreted.

HDV has at least eight genotypes, which differ from one another by at least 15% of their nucleotide sequences. Genotype I is the most prevalent HDV worldwide. Additional genotypes have recently been identified in Africa.

CLINICAL MANIFESTATIONS

HDV can be acquired at the same time as HBV (coinfection) or by a chronic HBsAg carrier (superinfection).[8] Coinfection is characterized by one or two episodes of acute hepatitis, depending on the respective amounts of HBV and HDV present in the inoculum; acute hepatitis can range from mild to fulminant. In contrast, when chronic HBV carriers are superinfected by HDV, acute hepatitis D is generally severe, often fulminant, and generally becomes chronic.

DIAGNOSIS

Three markers of HDV infection are total anti-HD antibodies, anti-HD IgM, and HDV RNA; the latter can be detected and quantified by real-time PCR. All HBsAg-positive patients should be tested.

In patients with an HBV-HDV coinfection, HDV RNA is only transiently present and is often missed. Anti-HBc IgM indicates concomitant acute HBV infection.

In HDV superinfection of a chronic HBsAg carrier, no anti-HBc IgM is present. HDV RNA is found in serum or plasma before and during the acute episode, whereas both total anti-HD antibodies and anti-HD IgM are present during the acute phase. Both serologic markers remain at high levels if the disease becomes chronic.

TREATMENT Rx

No treatments are of proven benefit for acute hepatitis D.[9]

PREVENTION

The most effective means of preventing HDV infection is HBV vaccination, because individuals who are protected against HBV cannot be infected with HDV. In chronic HBsAg carriers, standard hygiene and behavioral precautions should be practiced to avoid superinfection with HDV.

PROGNOSIS

In patients who are acutely coinfected with HBV and HDV, only approximately 2% become chronic HDV carriers. In contrast, when HDV infection is acquired by chronic HBV carriers, approximately 90% also become chronic carriers of HDV.

ACUTE HEPATITIS E

DEFINITION

The Pathogen

HEV is a member of the genus *Hepevirus* in the Hepeviridae family. HEV is a small, nonenveloped virus. Its genome is a positive sense, single-stranded RNA molecule. Five HEV genotypes have been described: genotypes 1 and 2 appear to be strictly human, whereas genotypes 3 and 4 appear to be of swine origin but can also infect humans.

PATHOBIOLOGY AND EPIDEMIOLOGY

HEV transmission is principally oral-fecal. HEV is endemic in most developing areas of the world, where acute infections are sporadic or occur during large epidemics related to the contamination of drinking water. Genotype 1 has been found principally in Asia and North Africa, whereas genotype 2 has been isolated in cases from Central America and Western Africa. No animal reservoir is known for these genotypes, and transmission appears to be linked to the contamination of food or drinking water. In industrialized countries, HEV genotypes 1 and 2 are not present. HEV genotypes 3 and 4 are endemic in swine, and zoonotic transmission appears to be the main route of transmission in Europe, the United States, and Asia. Diagnosed cases of acute hepatitis E have increased constantly in Western Europe and North America in recent years, and high HEV seroprevalences have been described in special at-risk populations, such as butchers or farmers. Transmission may be favored by the consumption of uncooked meat and direct contact with infected animals. It is now thought to be the most common cause of acute viral hepatitis worldwide.[10]

CLINICAL MANIFESTATIONS

The incubation period is 3 to 8 weeks. HEV infection causes only mild, nonspecific symptoms in the majority of cases, especially if the infection is acquired early in life. Immunocompetent individuals clear the virus spontaneously.[11] The peak of viremia occurs early during infection, whereas the peak of aminotransferase activity is reached approximately 6 weeks after infection. Severe disease is more frequent in pregnant women and in patients with underlying chronic liver disease, who may rarely progress to fulminant hepatic failure. In one European report, 5% of patients with Guillain-Barré syndrome (Chapter 420) had a preceding acute hepatitis E infection.[12]

DIAGNOSIS

Patients with otherwise unexplained acute hepatitis should be tested for hepatitis E. Diagnosis of acute hepatitis E is based on the detection of anti-HEV IgM antibodies, which appear within 6 weeks and persist for 3 to 12 months. Unfortunately, current assays lack sensitivity and specificity, and none are yet approved by the U.S. Food and Drug Administration. HEV RNA also can be detected in feces, serum, or plasma, where its presence is transient. Anti-HEV IgG generally persists for life after acute infection.

TREATMENT Rx

Treatment of acute hepatitis E is not recommended because the vast majority of patients recover spontaneously.[13] Severe and fulminant cases should be referred to specialized units, where ribavirin monotherapy can be successful.

PREVENTION

Improved public hygiene is the best defense against hepatitis E in developing countries. Travelers to areas of the world where HEV is endemic, particularly pregnant women, should be cautioned about drinking the water and eating uncooked food. An efficient prophylactic vaccine based on recombinant HEV proteins was 100% effective in a phase 3 trial and a hepatitis E vaccine is approved in China. However, the duration of protection is unclear, so the vaccine's efficacy in preventing the spread of HEV has yet to be determined.

PROGNOSIS

Acute cases can be severe in elderly patients, and fulminant cases are frequent among pregnant women who are infected during large-scale waterborne epidemics. Overall case fatality rates range from zero to 10%, and less than 1% of fatal cases of acute hepatitis in the United States are attributed to hepatitis E. HEV genotypes 3 and 4 are less virulent than genotypes 1 and 2. Sporadic cases in industrialized areas are generally benign. HEV does not commonly evolve into chronic infection, but immunosuppressed patients and HIV-positive individuals can become chronic carriers.

OTHER TYPES OF ACUTE VIRAL HEPATITIS

Infection by members of the Herpesviridae family, such as cytomegalovirus (Chapter 376), Epstein-Barr virus (Chapter 377), and herpes simplex virus

(Chapter 374), should be considered in the differential diagnosis of unclear episodes of elevated liver enzymes in the absence of markers of acute hepatitis virus infection, especially in immunocompromised individuals. For instance, cytomegalovirus infection can be associated with graft loss after liver transplantation (Chapter 154). Parvovirus B19 (Chapter 371) may persist in the liver and worsen liver disease in patients with chronic hepatitis B and perhaps in patients with chronic HCV infection. Human herpesvirus 6 variant A has been linked with syncytial giant-cell hepatitis.

Non-A-to-E Hepatitis

Rare patients develop acute hepatitis of presumed viral cause but have no markers of known hepatitis viruses. Some of these cases may be due to variants of known hepatitis viruses, in particular HBV, that are not detected by the usual serologic and molecular methods. However, the existence of other, unknown hepatotropic viruses cannot be excluded.

Grade A References

A1. Victor JC, Monto AS, Surdina TY, et al. Hepatitis A vaccine versus immune globulin for postexposure prophylaxis. *N Engl J Med*. 2007;357:1685-1694.

A2. Kumar M, Satapathy S, Monga R, et al. A randomized controlled trial of lamivudine to treat acute hepatitis B. *Hepatology*. 2007;45:97-101.

A3. Wiegand J, Wedemeyer H, Franke A, et al. Treatment of severe, nonfulminant acute hepatitis B with lamivudine vs placebo: a prospective randomized double-blinded multicentre trial. *J Viral Hepat*. 2014;21:744-750.

A4. Jeckel E, Cornberg M, Wedemeyer H, et al. Treatment of acute hepatitis C with interferon alfa-2b. *N Engl J Med*. 2001;345:1452-1457.

A5. Santantonio T, Fasano M, Sagnelli E, et al. Acute hepatitis C: a 24 week-course of peg-interferon alpha-2b versus a 12 week-course of peg-interferon alpha-2b alone or with ribavirin. *Hepatology*. 2014;59:2101-2109.

A6. Deterding K, Gruner N, Buggisch P, et al. Delayed versus immediate treatment for patients with acute hepatitis C: a randomised controlled non-inferiority trial. *Lancet Infect Dis*. 2013;13:497-506.

A7. Zhu FC, Zhang J, Zhang XF, et al. Efficacy and safety of a recombinant hepatitis E vaccine in healthy adults: a large-scale, randomised, double-blind placebo-controlled, phase 3 trial. *Lancet*. 2010;376:895-902.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Klevens RM, Liu S, Roberts H, et al. Estimating acute viral hepatitis infections from nationally reported cases. *Am J Public Health*. 2014;104:482-487.
2. Tekin R, Yolbas I, Dal T, et al. Evaluation of adults with acute viral hepatitis A and review of the literature. *Clin Ter*. 2013;164:537-541.
3. Squadrito G, Spinella R, Raimondo G. The clinical significance of occult HBV infection. *Ann Gastroenterol*. 2014;27:15-19.
4. Tillmann HL, Zachou K, Dalekos GN. Management of severe acute to fulminant hepatitis B: to treat or not to treat or when to treat? *Liver Int*. 2012;32:544-553.
5. Schillie S, Murphy TV, Sawyer M, et al. CDC guidance for evaluating health-care personnel for hepatitis B virus protection and for administering postexposure management. *MMWR Recomm Rep*. 2013;62:1-19.
6. Sagnelli E, Santantonio T, Coppola N, et al. Acute hepatitis C: clinical and laboratory diagnosis, course of the disease, treatment. *Infection*. 2014;42:601-610.
7. Bunchorntavakul C, Jones LM, Kikuchi M, et al. Distinct features in natural history and outcomes of acute hepatitis C. *J Clin Gastroenterol*. 2014; [Epub ahead of print].
8. Alvarado-Mora MV, Locarnini S, Rizzetto M, et al. An update on HDV: virology, pathogenesis and treatment. *Antivir Ther*. 2013;18:541-548.
9. Heidrich B, Manns MP, Wedemeyer H. Treatment options for hepatitis delta virus infection. *Curr Infect Dis Rep*. 2013;15:31-38.
10. Kamar N, Bendall R, Legrand-Abravanel F, et al. Hepatitis E. *Lancet*. 2012;379:2477-2488.
11. Hoofnagle JH, Nelson KE, Purcell RH. Hepatitis E. *N Engl J Med*. 2012;367:1237-1244.
12. van den Berg B, van der Eijk AA, Pas SD, et al. Guillain-Barré syndrome associated with preceding hepatitis E virus infection. *Neurology*. 2014;82:491-497.
13. Kamar N, Dalton HR, Abravanel F, et al. Hepatitis E virus infection. *Clin Microbiol Rev*. 2014;27:116-138.

149

CHRONIC VIRAL AND AUTOIMMUNE HEPATITIS

JEAN-MICHEL PAWLOTSKY

DEFINITION

Chronic hepatitis is defined by chronic necroinflammation of the liver and may be due to various causes, including hepatotropic viruses, autoimmunity, alcohol (Chapter 152), and metabolic disorders (Chapter 146). Chronic infection by hepatitis viruses is by far the main cause of chronic hepatitis worldwide, with more than 500 million individuals chronically infected with hepatitis B virus (HBV) or hepatitis C virus (HCV). Chronic viral hepatitis B and C are the leading cause of cirrhosis (Chapter 153) and hepatocellular carcinoma (Chapter 196) worldwide and account for more than 1 million deaths per year. Chronic HBV infection can be associated with infection by hepatitis D virus (HDV). Hepatitis A virus does not cause chronic hepatitis. Hepatitis E virus (HEV) generally does not cause chronic hepatitis, except in immunosuppressed patients.

CLINICAL MANIFESTATIONS

The clinical symptoms of chronic viral and autoimmune hepatitis are typically nonspecific, and many patients have no symptoms. Fatigue, sleep disorders, and right upper quadrant pain may be present. Often the diagnosis is made when liver test abnormalities are identified by blood testing during a routine health evaluation or assessment for an unrelated problem or at the time of voluntary blood donation. More advanced symptoms include poor appetite, nausea, weight loss, muscle weakness, itching, dark urine, and jaundice. Patients can progress to full-blown cirrhosis (Chapter 153), with its typical clinical manifestations. If cirrhosis is present, weakness, weight loss, abdominal swelling, edema, bruisability, gastrointestinal bleeding, and hepatic encephalopathy with mental confusion may arise. Other findings may include spider angiomas, palmar erythema (see Fig. 146-2), ascites (see Fig. 146-4), edema, and skin excoriations.

TABLE 149-1 DIAGNOSIS OF CHRONIC HEPATITIS

DIAGNOSIS	SCREENING TESTS	CONFIRMATORY TESTS
Chronic hepatitis B	HBsAg	HBV DNA, HBeAg
Chronic hepatitis C	Anti-HCV antibodies	HCV RNA
Chronic hepatitis D	Anti-HDV antibodies	HDV RNA
Autoimmune hepatitis	ANA (anti-LKM1)	Exclusion of other causes and patterns of clinical disease
Drug-induced liver disease	History	Rechallenge, if necessary, if considered safe
Wilson disease	Ceruloplasmin	Urine copper concentration
Cryptogenic hepatitis	Exclusion of other causes	

ANA = antinuclear antibody; anti-LKM1 = anti–liver-kidney microsomal 1 antibody; HBeAg = hepatitis B e antigen; HBsAg = hepatitis B surface antigen; HBV = hepatitis B virus; HCV = hepatitis C virus; HDV = hepatitis D virus.

DIAGNOSIS

Levels of alanine aminotransferase (ALT) and aspartate aminotransferase (AST) are usually two to five times the upper limit of normal. The ALT level is generally higher than the AST level, but both can be normal in mild or inactive disease or 10 to 25 times the upper limit of normal during acute exacerbations. Biologic tests can establish the specific diagnosis (Table 149-1).

Alkaline phosphatase and γ-glutamyl transpeptidase levels are usually minimally elevated unless cirrhosis is present. Serum bilirubin and albumin levels and the prothrombin time are normal unless the disease is severe or advanced. Serum immunoglobulin levels are mildly elevated or normal in chronic viral hepatitis but may be very elevated in autoimmune hepatitis. Results that suggest the presence of advanced fibrosis are a platelet count below 160,000, AST levels higher than ALT levels, elevation in serum bilirubin, decrease in serum albumin, prolongation of the prothrombin time, elevation in α-fetoprotein levels, and presence of rheumatoid factor or high globulin levels.

Liver ultrasound can determine the texture and size of the liver and spleen, exclude hepatic masses, and assess the gallbladder, intrahepatic bile ducts, and portal venous flow. Computed tomography and magnetic resonance imaging of the liver are helpful if a mass or other abnormality is found by ultrasound. Hepatic transient elastography or acoustic radiation force impulse imaging can assess liver stiffness as a surrogate marker of fibrosis.

Liver biopsy is usually critical for diagnosis and to determine the severity of disease. Hepatocellular necrosis is typically eosinophilic degeneration or ballooning degeneration throughout the parenchyma, greater in the periportal area, spotty, or piecemeal. Fibrosis also typically begins in the periportal regions and can link adjacent portal areas or portal and central areas (bridging fibrosis), distort the hepatic architecture, and lead to cirrhosis and portal hypertension. The histologic grade of chronic hepatitis can be determined by combining scores for periportal necrosis and inflammation, lobular necrosis and inflammation, and portal inflammation. Recently, ultrasound-based methods or serologic markers have proved accurate for the assessment of mild disease and cirrhosis, but they are less accurate for identifying moderate to severe inflammation, except in patients with chronic hepatitis C. Patients with suspected chronic viral or autoimmune hepatitis should be evaluated carefully for fatty liver, alcohol-induced (Chapter 152) or drug-induced (Chapter 150) liver disease, and metabolic liver diseases (Chapter 146), each of which can coexist with hepatitis. Liver biopsy can exclude other diagnoses that mimic chronic hepatitis, including fatty liver, alcoholic liver disease, steatohepatitis (Chapter 152), drug-induced liver disease (Chapter 150), sclerosing cholangitis (Chapter 155), iron overload (Chapter 212), and veno-occlusive disease (Chapter 143).

TREATMENT AND PROGNOSIS

Rx

Chronic HBV infection is not curable, but it usually can be controlled by appropriate antiviral drugs. HCV infection is curable, and more than 80% of patients who have access to new therapies are cured. Autoimmune hepatitis responds to immunosuppression with corticosteroids and azathioprine.

CHRONIC HEPATITIS B

EPIDEMIOLOGY

More than 240 million individuals, or approximately 4% of the world's population, are chronic HBV carriers. Two billion individuals, or one human being in three, have been in contact with the virus. In North America, Western and Northern Europe, and Australia, less than 2% of the population is chronic hepatitis B surface antigen (HBsAg) carriers. In the United States, the prevalence is approximately 0.4%; that is, approximately 1.25 million Americans are infected. In Eastern Europe, South America, the Mediterranean basin, and the Indian subcontinent, the prevalence of chronic HBsAg carriage is between 2 and 8%. Highly endemic areas, with a rate of chronic HBsAg carriage exceeding 8%, include China, Southeast Asia, sub-Saharan Africa, and the native populations of the Far North of America.

HBV is the main cause of primary liver cancer (hepatocellular carcinoma, Chapter 196) worldwide, with approximately 350,000 new cases attributable to HBV every year. Hepatocellular carcinoma is more likely in the presence of underlying cirrhosis, but HBV has oncogenic properties of its own, and hepatocellular carcinoma can occur in noncirrhotic HBV patients.

PATHOBIOLOGY

HBV is not a cytopathic virus. Rather, liver injury in chronic hepatitis B is a consequence of the local immune response at the immune elimination phase. In particular, liver injury is related to cytotoxic T cells that recognize and kill infected hepatocytes that express HBV antigens at their surface and to the local production of cytokines. Chronic inflammation triggers fibrogenesis through the activation of hepatic stellate cells. The hepatitis B X protein may also directly activate fibrogenesis. As a result, many patients with chronic hepatitis B have progressive fibrosis, which may evolve into cirrhosis.

The rate of chronicity after an acute HBV infection is more than 95% among patients infected at birth. This risk diminishes as the age at acquisition increases, and it is less than 5% in those infected as adults. Chronic HBV infection is defined by HBsAg carriage that persists for more than 6 months after the acute episode. Chronic HBsAg carriage typically evolves through three phases: immune tolerant, immune elimination, and inactive.

The immune tolerant phase is generally short if the infection occurred during adulthood, but it persists for years to decades in patients infected at birth or during early childhood. At the immune tolerant stage, the immune response of the host "tolerates" HBV infection and does not cause liver inflammation or hepatocyte destruction. The immune tolerant phase is characterized by the presence of hepatitis B e antigen (HBeAg), very high levels of HBV DNA in blood, normal serum or plasma aminotransferase levels, and no or minimal inflammatory activity on liver biopsy.

The immune elimination phase is characterized by an active immune response that causes necroinflammatory lesions and triggers hepatic fibrogenesis and progressive fibrosis. ALT and AST levels are increased, but HBV DNA levels are lower than during the immune tolerant phase and frequently fluctuate. The immune elimination phase has a variable duration, ranging from a few weeks to several decades. HBeAg, when present, defines HBeAg-positive chronic hepatitis B; HBeAg can be cleared and replaced by anti-HBe antibodies, defined as HBe seroconversion; or HBeAg can be absent while anti-HBe antibodies are present and define HBeAg-negative chronic hepatitis B. Patients with HBeAg-positive chronic hepatitis B are infected with a wild-type virus and are able to secrete the HBe protein. Patients with HBeAg-negative chronic hepatitis B are infected with so-called precore mutant viruses, which cannot produce the HBe protein because they have a stop codon in the pre-C gene, and/or with core promoter mutant viruses, which produce considerably lower amounts of HBe protein. Because of the prevalence of the HBV genotype D in Euro-Mediterranean and African countries, HBeAg-negative/anti-HBe-positive chronic hepatitis B is seven to nine times more frequent than HBeAg-positive disease in those locations.

The inactive HBsAg carriage phase is the result of successful immune elimination leading to HBe seroconversion. ALT and AST levels are normal, HBV DNA is undetectable or at very low levels, and patients without preexisting cirrhosis have normal liver histology.

CLINICAL MANIFESTATIONS

Chronic hepatitis B is usually asymptomatic. The most common symptom is fatigue, but sleep disorders, difficulty concentrating, and upper right quadrant pain are often observed. Chronic hepatitis B is characterized biologically by elevated aminotransferase levels, and ALT levels can fluctuate substantially during the immune elimination phase. Moderate cholestasis, with mildly elevated alkaline phosphatase and γ-glutamyl transpeptidase levels, also can be present, especially in patients with cirrhosis.

HBeAg-negative chronic hepatitis B is generally more severe than the HBeAg-positive variety. The incidence of spontaneous HBe seroconversion among HBeAg-positive patients is 8 to 12% per year when they are in the immune elimination phase; HBe seroconversion often follows a transient ALT flare. Some of these patients evolve toward inactive HBsAg carriage, whereas others switch into an HBeAg-negative form of chronic hepatitis B, with elevated ALT levels and an HBV DNA level greater than 2000 IU/mL.

The annual incidence of cirrhosis varies from 2 to 10% in patients with chronic HBV infection, with a cumulative incidence of approximately 20% at 5 years. The risk for cirrhosis is two- to four-fold higher in HBeAg-negative patients compared with HBeAg-positive ones, probably because they are older and have more severe disease at the time of diagnosis. The annual incidence of hepatocellular carcinoma in patients with chronic hepatitis B varies from 1% in patients without cirrhosis to 2 to 8% in cirrhotic patients, with the higher rates occurring in older patients. Patients with cirrhosis (Chapter 153), hepatocellular carcinoma (Chapter 196), or both have the typical signs associated with these conditions. Rarely, chronic HBV infection is associated with extrahepatic manifestations, including glomerulonephritis (Chapter 121), most often in children, and polyarteritis nodosa (Chapter 270), mostly in adults.

DIAGNOSIS

Serologic markers used to diagnose chronic hepatitis B (Table 149-2) include HBsAg, anti-HBs antibodies, total anti–hepatitis B core (HBc) antibodies and anti-HBc immunoglobulin M (IgM), HBeAg, and anti-HBe antibodies. Molecular markers include HBV DNA and HBV resistance substitutions; real-time polymerase chain reaction (PCR)–based assays are the best way to detect and quantify HBV DNA.

Chronic HBV infection is defined by the persistence of HBsAg in the serum for more than 6 months after the acute episode. The majority of subjects with isolated anti-HBc antibodies are not viremic. However, some individuals who test positive for anti-HBc antibodies, but not for HBsAg or anti-HBs antibodies, may be viremic; in these cases, the virus's amino acid substitutions in the HBsAg sequence make HBsAg undetectable with current enzyme immunoassays. Other individuals may have such low-level HBV replication in their livers that HBV DNA is not detectable in blood ("occult" hepatitis B). Screening is recommended only in high-risk individuals.[1]

Serum or plasma ALT and HBV DNA levels are important markers of severity and prognosis. For both HBV and HCV, the assessment of severity, including the grade of necroinflammation and the stage of fibrosis, is based on the liver biopsy (Fig. 149-1). Noninvasive assessment using serologic markers, transient elastography, or acoustic radiation force impulse imaging can discriminate cirrhosis from mild hepatitis and fibrosis; although they are not accurate enough for intermediate stages, these methods will likely replace liver biopsy in the pretreatment assessment of the severity of chronic hepatitis B in many patients in the future.

PREVENTION AND TREATMENT

Rx

Patients with chronic hepatitis B should be vaccinated against hepatitis A virus, abstain from alcohol, and avoid immunosuppressive therapies unless absolutely necessary. HBV-infected patients who require corticosteroids, rituximab, or other chemotherapy for other conditions should receive entecavir 0.5 mg/day or tenofovir 300 mg/day during therapy and for 12 months after its cessation as prophylaxis against reactivation of hepatitis B.

The goals of therapy are to suppress HBV replication, reduce the histologic activity of chronic hepatitis, and lessen the risk for cirrhosis and hepatocellular carcinoma. HBV infection cannot be completely eradicated because of the

TABLE 149-2 VIROLOGIC MARKER PROFILES IN PATIENTS WITH CHRONIC HEPATITIS B VIRUS INFECTION

	HBsAg	ANTI-HBs Ab	HBeAg	ANTI-HBe Ab	*Anti-HBc Ab* IgM	*Anti-HBc Ab* TOTAL	HBV DNA
Chronic hepatitis							
HBeAg-positive	+	–	+	–	–*	+	$>2 \times 10^4$ IU/mL
HBeAg-negative	+	–	–	+	–*	+	$>2 \times 10^3$ IU/mL
Inactive carrier	+	–	–	+	–	+	$<2 \times 10^3$ IU/mL
Reactivation	+	–	+/–	+/–	+/–	+	$>2 \times 10^3$ IU/mL

*Anti-HBc IgM can be detected at low titers.
Ab = antibodies; Ag = antigen; HBc = hepatitis B core; HBe = hepatitis B e; HBs = hepatitis B surface; HBV = hepatitis B virus; IgM = immunoglobulin M.

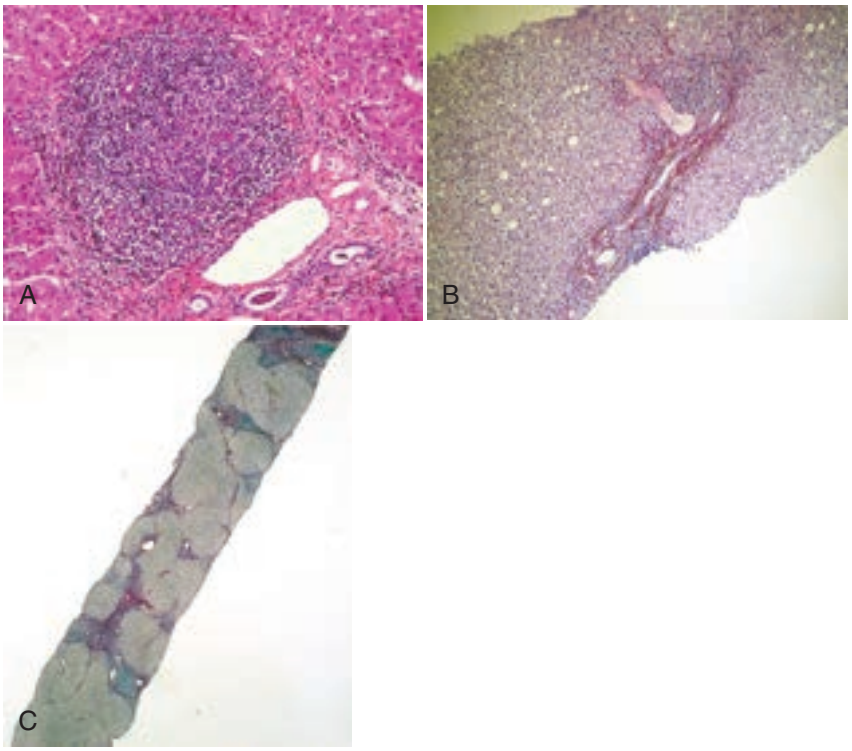

FIGURE 149-1. Liver biopsies in patients with chronic hepatitis C. A, Lymphoid nodule with germinal center; minimal interface hepatitis (hematein-eosin, magnification 200×). **B,** Mild fibrosis, Metavir score F1 (picrosirius-hemalun, magnification 100×). **C,** Extensive fibrosis, Metavir score F3 (picrosirius-hemalun, magnification 20×). (Courtesy Prof. Elie-Serge Zafrani, Department of Pathology, Henri Mondor Hospital, Créteil, France.)

persistence of covalently closed circular DNA in the nucleus of infected hepatocytes. As a result, therapy attempts to reduce HBV DNA levels as much as possible—ideally, below the limit of detection of real-time PCR assays (10 to 15 IU/mL)—to ensure a degree of viral suppression that will lead to biochemical remission, histologic improvement, and prevention of complications.

Two different types of drugs for the treatment of chronic hepatitis B are pegylated interferon-α (IFN-α) and nucleoside/nucleotide analogues. Pegylated IFN-α2a, administered subcutaneously at a dose of 180 μg/week for 48 weeks, improves various markers of HBV infection in both HBeAg-positive (Table 149-3) and HBeAg-negative (Table 149-4) patients. Pegylated IFN-α2b (1.5 μg/kg) is very similar to IFN-α2a and is used by many liver experts, although it is not currently approved for HBV. The most frequent side effects of IFN-α are flulike symptoms after the injections, fatigue, anorexia, weight loss, and alopecia. The most concerning side effects are neutropenia, thrombocytopenia, anxiety, irritability, depression, and suicidal ideation.

Nucleoside analogues (lamivudine, telbivudine, and entecavir) require triple phosphorylation to be active, whereas nucleotide analogues (adefovir and tenofovir) need only two phosphorylations to be active. These drugs are given orally (PO) once daily at the following dosages: 100 mg for lamivudine, 600 mg for telbivudine, 0.5 mg for entecavir, 10 mg for adefovir (administered as the pro-drug adefovir dipivoxil), and 300 mg for tenofovir (administered as the pro-drug tenofovir disoproxil fumarate). All have short-term benefits on various markers of HBV infection in HBeAg-positive (see Table 149-3) and HBeAg-negative (see Table 149-4) patients. Entecavir and tenofovir are two of the most potent inhibitors of HBV replication, and they are the least likely to select for resistant HBV variants. These drugs are generally well tolerated. However, adefovir is nephrotoxic at doses higher than those used for HBV therapy; renal impairment and decreases in mineral bone density are rarely seen with tenofovir; myopathy is a rare complication of telbivudine, and peripheral neuropathy has been observed when telbivudine is combined with pegylated IFN-α.

Patients should be considered for treatment when their HBV DNA levels are greater than 2000 IU/mL and/or their serum ALT levels are abnormal if the liver biopsy shows moderate to severe active necroinflammation, fibrosis, or both. Indications for treatment must also take into account the patient's age and health status and the availability of antiviral agents in individual countries. Patients in the immune tolerant phase and those with mild hepatitis on liver biopsy (or noninvasive markers) should not be treated, but follow-up ALT levels and HBV DNA assays are mandatory. Patients with compensated cirrhosis and detectable HBV DNA may be considered for treatment even if their ALT levels are normal, their HBV DNA levels are below 2000 IU/mL, or both. Patients with decompensated cirrhosis require urgent antiviral treatment.

Two treatment strategies can be considered: a 48-week course of pegylated IFN-α or long-term oral treatment with nucleoside/nucleotide analogues. Whether their combined use improves the rate of sustained virologic response after treatment is currently under study. Pegylated IFN-α can provide a sustained virologic response, defined as a sustained HBe seroconversion (clearance of HBeAg, which is replaced by anti-HBe antibodies) and an HBV DNA level that remains below 2000 IU/mL after a 48-week course of treatment; an ALT flare can be observed at the time of HBeAg loss in patients in whom treatment is successful. Pegylated IFN-α treatment should be reserved for patients with the best chance of a sustained virologic response off

TABLE 149-3 SHORT-TERM (1-YEAR) RESPONSES TO APPROVED ANTIVIRAL THERAPIES AMONG TREATMENT-NAÏVE PATIENTS WITH HBeAg-POSITIVE CHRONIC HEPATITIS B

	PLACEBO/CONTROL GROUPS	PEGYLATED IFN-α2A (48 WK)	LAMIVUDINE (48-52 WK)	TELBIVUDINE (52 WK)	ENTECAVIR (48 WK)	ADEFOVIR (48 WK)	TENOFOVIR (48 WK)
Loss of serum HBV DNA (%)*	0-17	25	40-44	60	67	21	76
HBeAg loss (%)	6-12	30/34†	17-32	26	22	24	NA
HBe seroconversion (%)	4-6	27/32‡	16-21	22	21	12	21
HBsAg loss (%)	0-1	3	1	0	2	0	3
ALT normalization (%)	7-24	39	41-75	77	68	48	68
Histologic improvement (%)	NA	38‡	49-56	65	72	53	74
Durability of response (%)§	NA	NA	50-80	≈80	69	≈90	NA

Modified from Lok AS, McMahon BJ. American Association for the Study of Liver Diseases practice guidelines: chronic hepatitis B—update 2009, http://www.aasld.org.
*HBV DNA levels were assessed with various molecular assays, with different lower limits of detection.
†Responses at week 48/week 72 (24 wk after stopping treatment).
‡Post-treatment biopsies at week 72 (24 wk after stopping treatment).
§No or short duration of consolidation treatment for lamivudine and entecavir; most patients had consolidation treatment for adefovir and telbivudine.
Ag = antigen; ALT = alanine aminotransferase; HBe = hepatitis B e; HBs = hepatitis B surface; HBV = hepatitis B virus; IFN = interferon; NA = not available.

TABLE 149-4 SHORT-TERM (1-YEAR) RESPONSES TO APPROVED ANTIVIRAL THERAPIES AMONG TREATMENT-NAÏVE PATIENTS WITH HBeAg-NEGATIVE CHRONIC HEPATITIS B

	PLACEBO/CONTROL GROUPS	PEGYLATED IFN-α2A (48 WK)	LAMIVUDINE (48-52 WK)	TELBIVUDINE (52 WK)	ENTECAVIR (48 WK)	ADEFOVIR (48 WK)	TENOFOVIR (48 WK)
Loss of serum HBV DNA (%)*	0-20	63	60-73	88	90	51	93
ALT normalization (%)	10-29	38	60-79	74	78	72	76
Histologic improvement (%)	33	48	60-66	67	70	64	72
Durability of response (%)	NA	≈20	≈10	NA	3	≈5	NA

*HBV DNA levels were assessed with various molecular assays, with different lower limits of detection.
ALT = alanine aminotransferase; HBeAg = hepatitis B e antigen; HBV = hepatitis B virus; IFN = interferon; NA = not available.
Modified from Lok AS, McMahon BJ. American Association for the Study of Liver Diseases practice guidelines: chronic hepatitis B—update 2009, http://www.aasld.org.

treatment—HBeAg-positive patients with high baseline ALT levels (more than three times the upper limit of normal) and HBV DNA levels below 2.10^6 IU/mL. Pegylated IFN-α therapy is contraindicated in patients with advanced cirrhosis and in immunosuppressed patients. Patients infected with HBV genotypes A and B generally respond better to IFN-α therapy than do patients infected with genotypes C and D, but the predictive value of the HBV genotype for an individual patient is weak. Patients who fail to achieve a sustained virologic response after a single course of pegylated IFN-α are candidates for nucleoside/nucleotide analogue therapy.

Long-term treatment with nucleoside/nucleotide analogues is indicated in the majority of patients with chronic hepatitis B. Tenofovir or entecavir, which are the most potent drugs with an optimal resistance profile, are recommended as first-line monotherapies. HBV DNA should be suppressed to undetectable levels (<10 to 15 IU/mL) with a sensitive real-time PCR-based assay. If the HBV DNA level is reduced but still detectable in a compliant patient, the other agent may be added; however, the long-term safety of combined tenofovir and entecavir is unknown. When HBeAg-positive patients seroconvert to negative or HBeAg-negative patients lose their HBsAg, treatment should be continued for an additional 6 to 12 months at least. In all other cases, treatment should be continued for life and adherence is particularly important.

Virologic breakthroughs—defined by a subsequent increase of the HBV DNA level of 1 log or more above the nadir level—in adherent patients are due to HBV resistance to the administered antiviral drug or drugs. HBV DNA levels most often increase back to baseline levels, and the virologic breakthrough is generally followed a few weeks later by a biochemical breakthrough in which a previously normal ALT level rises above normal. The cumulative rates of resistance in newly treated patients are 70% at 5 years for lamivudine, 17% at 2 years for telbivudine, 1.2% at 6 years for entecavir, 29% at 5 years for adefovir, and 0% at 6 years for tenofovir. These figures illustrate the high genetic barrier against resistance of entecavir and tenofovir. In a patient who develops resistance to any of the available nucleoside/nucleotide analogues, adding a second drug without cross-resistance or switching to tenofovir are the only efficient strategies, although the long-term safety of some of these combinations is not known. In patients with lamivudine, telbivudine, or entecavir resistance, treatment should be switched to tenofovir or tenofovir should be added. Patients who have resistance to adefovir should be switched to tenofovir and eventually also given lamivudine, telbivudine, or entecavir. In patients with tenofovir resistance (never reported thus far), any of these three drugs should be added. The combination of tenofovir and emtricitabine, a nucleoside analogue similar to lamivudine, in a single tablet (approved for HIV but not for HBV therapy) is also a valid option in cases of resistance to any of these drugs.

In patients with cirrhosis, nucleoside/nucleotide analogue therapy is the only option because pegylated IFN-α is contraindicated if cirrhosis is decompensated. Because resistance in this population can be life-threatening, some experts recommend de novo treatment with two potent drugs without cross-resistance. In those with decompensated disease, efficient antiviral treatment most often stabilizes the patient's condition and can also delay or obviate the need for liver transplantation. If transplantation is needed, post-transplant administration of anti-HBV immunoglobulins in combination with a potent nucleoside/nucleotide analogue prevents recurrent HBV in the vast majority of cases. With the advent of potent drugs with a high barrier to resistance, such as tenofovir or entecavir, the utility of immune globulins is debated.

PROGNOSIS

Every year, approximately 0.5% of inactive HBsAg carriers spontaneously lose HBsAg, and most of them acquire anti-HBs antibodies. Reactivations are possible in inactive HBV carriers, especially if they become immunosuppressed, such as when they are treated for other conditions with corticosteroids, rituximab, or other chemotherapeutic agents. HBV reactivations often evolve into a subfulminant or fulminant form.

The risk of cirrhosis (Chapter 153) in chronic hepatitis B is 2 to 10% per year, and it is significantly associated with a higher HBV DNA level, older age, alcohol consumption, coinfection with other hepatotropic viruses, and coinfection with human immunodeficiency virus (HIV). The cumulative incidence of liver decompensation is about 15 to 20% at 5 years in patients with compensated cirrhosis. The complications of cirrhosis, including

hepatocellular carcinoma, are among the main causes of mortality in HBV-infected patients, and the annual incidence of death is about 3 to 4%.

The likelihood of developing hepatocellular carcinoma (Chapter 196) is approximately 1% per year in patients without cirrhosis and 2 to 8% per year in those with cirrhosis, and it is significantly associated with a higher HBV DNA level, male sex, old age, reversion from anti-HBe–positive to HBeAg-positive, and coinfection with other hepatotropic viruses. HBV carriers at risk for hepatocellular carcinoma should be screened every 6 to 12 months, with an ultrasound examination and an α-fetoprotein level.

CHRONIC HEPATITIS C

EPIDEMIOLOGY

HCV, which is present on all continents, is estimated to cause chronic infection in approximately 185 million individuals, or approximately 3% of the world's population. The prevalence of chronic HCV infection, which varies geographically, is estimated to be about 1.3% in the United States (affecting 2.7 million individuals),[2] 3.5% in Asia, 1.9% in the Americas overall, 5.2% in Africa, 1.7% in Europe, and 1.8% in Oceania. The highest prevalence is in Egypt (9% overall, but up to 40% in certain rural areas), where infection was initially spread by intramuscular injections for schistosomiasis during treatment campaigns several decades ago. By 2007, hepatitis C virus superseded HIV as a cause of death in the United States.

PATHOBIOLOGY

Acute HCV infection evolves into chronic infection in 50 to 80% of cases. Even patients who spontaneously recover and maintain detectable anti-HCV antibodies are not protected against reinfection. Persistence of infection is related to a qualitatively and quantitatively altered CD4+ T-helper cell and cytotoxic T lymphocyte response that fails to eradicate infection. The plasticity of the viral genomes is responsible for the coexistence of closely related but genetically different viral populations in equilibrium in the patient's replicative environment. This genetic diversity allows continuously generated variant viral populations to be selected by timely changes in the replicative environment.

Chronic HCV infection is responsible for necroinflammatory lesions of varying severity, sometimes associated with steatosis, which is the accumulation of triglycerides in hepatocytes. HCV is not a cytopathic virus. Liver injury in chronic hepatitis C is related to the action of immune effectors that recognize and kill infected hepatocytes that express HCV antigens at their surface. Chronic inflammation triggers fibrogenesis through the activation of hepatic stellate cells. Fibrosis progresses at nonlinear rates that are generally faster in older patients, in males, and in the presence of chronic alcohol intake, viral coinfections, or immunosuppression. The severity of chronic hepatitis is independent of the HCV RNA level and of the HCV genotype. This chronic inflammation and progression of fibrosis predispose patients to cirrhosis (Chapter 153) and hepatocellular carcinoma (Chapter 196).

CLINICAL MANIFESTATIONS

Acute hepatitis C (Chapter 148) is most often asymptomatic and therefore undiagnosed. The most common symptom associated with chronic HCV infection is fatigue, but it may remain inapparent for years. ALT levels are usually moderately elevated and fluctuate, but they can remain normal for weeks to months despite active hepatitis on liver biopsy. Moderate cholestasis can be present in patients with cirrhosis. Patients with cirrhosis (Chapter 153), hepatocellular carcinoma (Chapter 196), or both have the typical signs associated with these conditions.

HCV is the main cause of type II and type III mixed cryoglobulinemia (Chapter 187). Low levels of circulating cryoglobulins, which contain HCV RNA, anti-HCV antibodies, rheumatoid factor, and low levels of complement, can be found in 50 to 70% of cases, whereas elevated rheumatoid factor (Chapter 264) is found in 70% of cases. Fewer than 1% of HCV-infected patients develop symptoms of cryoglobulinemic vasculitis (Chapter 270), including fatigue, myalgias, arthralgias, rash (purpura, hives, leukocytoclastic vasculitis), neuropathy, and membranoproliferative glomerulonephritis. Cryoglobulinemia can be severe and lead to end-stage renal disease or severe neuropathies, and long-term cryoglobulinemia has been linked to non–Hodgkin B-cell lymphomas (Chapter 185).

Low titers of antinuclear and anti–smooth muscle antibodies can be found in HCV-infected patients, but they do not have any clinical significance. HCV has been reported to trigger the symptoms of porphyria cutanea tarda (Chapter 210), and an association with lichen planus (Chapter 438) has been suggested.

DIAGNOSIS

Chronic HCV infection is defined by the persistence of HCV RNA for more than 6 months. In patients with clinical, biologic, or both signs of chronic liver disease, chronic hepatitis C is diagnosed by the simultaneous presence of anti-HCV antibodies and HCV RNA. Detectable HCV replication in the absence of anti-HCV antibodies is observed almost exclusively in patients who are profoundly immunosuppressed, on hemodialysis, or agammaglobulinemic. In the United States, screening for HCV is recommended in high-risk individuals, and one-time screening is recommended in all persons born between 1945 and 1965 (Chapter 15).[3] The level of HCV replication does not correlate with the severity of liver disease or with the risk for progression to cirrhosis or hepatocellular carcinoma.

The HCV genotype, which has important therapeutic implications, should be determined. Anti-HCV IgM, which is found in approximately 50% of patients with chronic hepatitis, is of no significance. Laboratory testing often reveals high levels of monoclonal rheumatoid factor and cryoglobulins.

TREATMENT

Chronic HCV infection is curable. The goal of therapy is to achieve a sustained virologic response, defined by undetectable HCV RNA at 12 to 24 weeks after the end of therapy using a sensitive HCV RNA assay with a lower limit of detection of 10 to 20 IU/mL.[4]

The decision to treat chronic hepatitis C depends on a precise assessment of the severity of liver disease, the presence of absolute or relative contraindications to therapy, and the patient's willingness to be treated. Treatment typically requires a liver biopsy, but serologic markers of liver fibrosis, fibrogenesis, or both and transient elastography or acoustic radiation force impulse imaging have been validated in large series of patients with chronic hepatitis C. In patients with no indication for therapy or with contraindications to therapy, repeated assessments of aminotransferase levels are recommended on a yearly basis. Assessment of liver inflammation and fibrosis by liver biopsy or noninvasive serologic or ultrasound-based testing is indicated for patients with persistently or intermittently elevated aminotransferase levels.

Treatment strategies for chronic hepatitis C are evolving rapidly. Options vary in different countries based on the availability of newer medications.[5,6]

Historic Standard Therapy

For more than 15 years, the standard treatment for chronic hepatitis C has been the combination of ribavirin (0.8 to 1.4 g/day PO) with either pegylated IFN-α2a (180 μg subcutaneously [SC] once weekly) or pegylated IFN-α2b (1.5 μg/kg SC once weekly). This therapy yields sustained virologic response rates of the order of 50% in patients infected with HCV genotype 1 and 75-80% in those infected with HCV genotypes 2 or 3. Two inhibitors of the HCV protease, telaprevir (750 mg every 8 hours or 1250 mg every 12 hours PO) and boceprevir (800 mg every 8 hours PO), have been approved for use in combination with pegylated IFN-α and ribavirin exclusively in patients infected with HCV genotype 1, in whom this regimen yields sustained virologic response rates of the order of 65 to 75%. These regimens remain the only available treatments in many areas of the world where new therapies are not yet available.

The most common side effects of IFN-α are influenza-like symptoms (which can be prevented by acetaminophen), neutropenia, thrombocytopenia, irritability, difficulty concentrating, memory disturbances, thyroiditis, hair loss, sleep disorders, and weight loss. The principal side effect of ribavirin is hemolytic anemia. As a result of these side effects, dose modification is frequently required during therapy. Ribavirin should be decreased in 200-mg increments in patients with severe anemia. For IFN-induced side effects, the pegylated IFN-α dose should be decreased stepwise, from 180 to 135 to 90 μg/week for pegylated IFN-α2a, and from 1.5 to 1 to 0.5 μg/kg/week for pegylated IFN-α2b. In more than 50% of cases, telaprevir administration is associated with rash, which can be severe in approximately 5% of cases and can rarely manifest as *d*rug *r*eaction with *e*osinophilia and *s*ystemic *s*ymptoms (DRESS) syndrome or Stevens-Johnson syndrome. Boceprevir induces dysgeusia. Both telaprevir and boceprevir aggravate ribavirin-induced anemia.

The main contraindications to therapy with pegylated IFN-α and ribavirin are decompensated liver disease, renal failure, severe immunosuppression, solid organ transplantation, cytopenias, severe psychiatric disease, and active substance abuse. Ribavirin is also contraindicated in patients with anemia, significant coronary or cerebrovascular disease, or renal insufficiency. Because ribavirin is teratogenic, it is essential that adequate contraception be practiced during therapy and for at least 6 months thereafter in both men and women.

Newly Available Regimens

New treatment regimens (Chapter 360), which provide a sustained virologic response for 80 to 95% of patients, have changed the approach to hepatitis C in countries in which they are available.[7] Simeprevir (150 mg/day PO), which is an inhibitor of the HCV protease, has antiviral activity against genotypes 1 and 4 and a low barrier to resistance. Sofosbuvir (400 mg/day PO), which is a

nucleotide analogue inhibitor of the HCV RNA-dependent RNA polymerase, is active against all HCV genotypes with a high barrier to resistance. Both drugs are well tolerated. Adverse reactions in patients receiving simeprevir are rash (including photosensitivity), pruritus, and nausea, as well as mild, transient hyperbilirubinemia not accompanied by changes in other liver parameters. The only side effects reported with sofosbuvir are headache and fatigue.

In patients for whom therapy is deemed appropriate, treatment is guided by the patient's HCV genotype and eligibility for pegylated IFN-α administration. Patients infected with HCV genotype 1 should be treated with one of five regimens: daily sofosbuvir plus weight-based ribavirin plus weekly pegylated IFN-α for 12 weeks; daily sofosbuvir plus daily simeprevir with or without weight-based ribavirin for 12 weeks; daily simeprevir plus weight-based ribavirin plus weekly pegylated IFN-α for 12 weeks followed by 12 weeks of pegylated IFN-α and ribavirin in treatment-naïve patients and for 36 weeks in treatment-experienced patients; the combination of sofosbuvir with the NS5A inhibitor ledipasvir in one single pill with or without weight-based ribavirin for 12 weeks; or the combination of ritonavir-boosted paritaprevir, ombitasvir, and dasabuvir with or without weight-based ribavirin for 12 or 24 weeks. The combination of simeprevir, pegylated IFN-α, and ribavirin should not be used in patients who are infected with HCV subtype 1a with a detectable Q80K substitution in the HCV protease sequence before therapy and who should benefit from the other options.

Patients infected with HCV genotype 2 must be treated with daily sofosbuvir plus ribavirin for 12 weeks (eventually combined with pegylated IFN-α if the patient has failed prior therapy). Patients infected with HCV genotype 3 should be treated with either of two regimens: daily sofosbuvir plus weight-based ribavirin for 24 weeks or daily sofosbuvir plus weight-based ribavirin plus pegylated IFN-α for 12 weeks.

Patients infected with HCV genotype 4 should be treated with one of three regimens: daily sofosbuvir plus weight-based ribavirin plus weekly pegylated IFN-α for 12 weeks; daily sofosbuvir plus daily simeprevir with or without weight-based ribavirin for 12 weeks; or daily simeprevir plus weight-based ribavirin plus weekly pegylated IFN-α for 12 weeks followed by 12 weeks of pegylated IFN-α and ribavirin in treatment-naïve patients and 36 weeks in treatment-experienced patients. Patients infected with HCV genotypes 5 or 6 should be treated with daily sofosbuvir plus weight-based ribavirin plus weekly pegylated IFN-α for 12 weeks.[6]

Emerging Regimens

Numerous new anti-HCV drugs are in clinical development. They include protease inhibitors, nucleoside/nucleotide analogues, and non-nucleoside inhibitors of the HCV RNA-dependent RNA polymerase, inhibitors of the nonstructural 5A protein of HCV, and cyclophilin A inhibitors. Oral regimens with sustained virologic response rates greater than 90% are rapidly replacing IFN-based therapies.[6-8] Treatment regimens likely to become available soon include the combination of sofosbuvir with the NS5A inhibitor daclatasvir; and the triple combination of the protease inhibitor paritaprevir boosted by ritonavir, the NS5A inhibitor ombitasvir, and the non-nucleoside inhibitor of the HCV polymerase dasabuvir, with or without ribavirin.

End-Stage Disease

In patients with end-stage liver disease, liver transplantation (Chapter 154) is the only option. However, in the absence of treatment, the graft becomes infected in 100% of patients who are viremic at the time of transplantation. Treatment with daily sofosbuvir and ribavirin until liver transplantation (up to 48 weeks) efficiently prevents graft infection when HCV RNA has been undetectable for at least 30 days before transplantation.

PROGNOSIS

Spontaneous HCV clearance in patients with chronic hepatitis C is exceptional. The HCV RNA level has no prognostic value in chronic hepatitis C. An estimated 20% of patients with chronic hepatitis C develop cirrhosis (Chapter 153) after an average of 20 years of progression in the absence of therapy. Cirrhosis remains compensated for many years in the vast majority of patients, but decompensation occurs at an annual rate of 2 to 5% in cirrhotic patients. After a first decompensation, the mortality rate related to portal hypertension, hepatocellular insufficiency, and hepatocellular carcinoma is 10% per year, with a 50% survival rate at 5 years. The risk for death increases with advancing age, male gender, and the severity of cirrhosis.

Hepatocellular carcinoma (Chapter 196) is rare in patients with chronic hepatitis C without cirrhosis. In patients with cirrhosis, the incidence of hepatocellular carcinoma is 2 to 4% per year, most often in patients with compensated cirrhosis. HCV has become the most common cause of hepatocellular carcinoma in most industrialized countries. However, hepatocellular carcinoma is less likely to develop in patients who have a sustained virologic response to treatment,[9] so this risk will hopefully decline now that better treatments are available.

Long-term follow-up shows that HCV does not recur in greater than 99% of patients who achieve a sustained virologic response, even in those who are immunosuppressed or receive chemotherapy. However, the liver disease may continue to evolve even after the infection has been eradicated. In addition, patients with chronic hepatitis C should abstain from alcohol and, unless there are other contraindications, should be vaccinated for hepatitis A and B (Chapter 18).

CHRONIC HEPATITIS D

EPIDEMIOLOGY

HDV infection occurs only in HBsAg carriers. Only approximately 2% of patients acutely coinfected with HDV and HBV develop chronic hepatitis D. In chronic HBV carriers superinfected by HDV, however, 90% of patients become chronic HDV carriers.

CLINICAL MANIFESTATIONS

Chronic hepatitis D is generally severe, with more than 80% of patients developing cirrhosis. Compared with patients who have chronic hepatitis B alone, patients with chronic infection with both HBV and HDV are three times as likely to develop hepatocellular carcinoma and twice as likely to die.

DIAGNOSIS

Markers of HDV infection should be sought at least once in every chronic HBsAg carrier. Both total anti-HD antibodies and anti-HD IgM remain at high levels in chronic HDV infection, and HDV RNA is present. Although all chronic HDV carriers also are chronic HBsAg carriers, chronic HDV carriers generally have low or undetectable HBV DNA levels because HDV inhibits HBV replication.

TREATMENT

Rx

High doses (9 million units three times/week for 1 year) of standard, nonpegylated IFN-α result in sustained normalization of ALT levels 24 weeks after the end of therapy in approximately 50% of cases, sometimes for as long as 20 years. Some patients clear HDV RNA and, eventually, HBsAg.

Pegylated IFN-α2b, 1.5 μg/kg once weekly for 12 months, provides a sustained virologic response in 20 to 40% of cases. Although there is no clear consensus, most experts now recommend 1 year of pegylated IFN-α as first-line treatment of chronic HDV infection.

PREVENTION

Chronic HDV infection is best prevented by preventing primary HBV infection, because individuals who are protected against HBV cannot be infected with HDV. In chronic HBsAg carriers, standard hygiene and behavioral precautions should be practiced to avoid superinfection with HDV. Once acute HDV infection occurs, no secondary prevention strategy is successful.

CHRONIC HEPATITIS E

HEV infection is usually an acute self-limiting disease, but in developed countries it causes chronic infection with rapidly progressive cirrhosis in organ transplant recipients, patients with hematologic malignancies that require chemotherapy, and immunosuppressed HIV-infected individuals, either from a latent virus reactivated by immunosuppression or from a virus transmitted at the time of transplantation. Almost all cases of chronic infection have been in immunosuppressed patients.[10,11]

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Chronic hepatitis E occurs after transplantation. Diagnosis is based on the detection of anti-HEV IgM antibodies. HEV RNA also can be detected in blood or feces, where its presence is transient. Solid organ transplant recipients with chronic hepatitis E harbor repeatedly positive anti-HEV antibodies and HEV RNA in blood.

TREATMENT AND PREVENTION

Rx

There is no validated treatment of chronic HEV infection. Ribavirin (600 mg/day),[12] pegylated IFN, or a combination of the two has been used in small case series. Reducing the level of immunosuppression can result in viral clearance in approximately one third of patients. Whether candidates for transplantation or immunosuppressive therapies should be vaccinated to prevent chronic HEV infection remains to be determined.

AUTOIMMUNE HEPATITIS

Autoimmune hepatitis is a chronic inflammatory liver disorder characterized by the presence of autoantibodies in serum, high levels of serum immunoglobulins, and a frequent association with other autoimmune diseases.[13]

EPIDEMIOLOGY AND PATHOBIOLOGY

Autoimmune hepatitis typically manifests between the ages of 15 and 25 years or between the ages of 45 and 60 years, and it is more common in women. The incidence rate is approximately 1.7 per 100,000 population per year.[14] Along with primary biliary cirrhosis (Chapter 155) and primary sclerosing cholangitis (Chapter 155), autoimmune hepatitis is one of the three major autoimmune liver diseases.

Autoimmune hepatitis is believed to be caused by autoimmune reactions against normal hepatocytes in genetically predisposed persons or persons exposed to unidentified triggers of an autoimmune process against liver antigens. Associations are seen with the human leukocyte antigen (HLA) class I B8 and class II DR3 and DR52a loci. In Asians, autoimmune hepatitis is associated with HLA DR4.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Autoimmune hepatitis tends to be more severe at its onset than chronic hepatitis B or C, and it progresses to end-stage liver disease if not treated with immunosuppression. Although it is occasionally detected by elevated serum aminotransferase levels on a routine health evaluation, most patients present with fatigue and jaundice. Elevations of bilirubin or alkaline phosphatase indicate more severe or advanced disease. Patients typically have marked elevations in serum γ-globulin, specifically immunoglobulin G, as well as autoantibodies directed at non–organ-specific cellular constituents.

Type 1 (classic) autoimmune hepatitis is characterized by the presence of titers of 1 : 80 or higher (>1 : 20 in children) of antinuclear, anti–smooth muscle, antiactin, and anti-asialoglycoprotein receptor antibodies. Type 2 autoimmune hepatitis is characterized by similar elevations of anti–liver-kidney microsomal 1 antibodies and anti–liver cytosol 1 antibodies, without antinuclear or anti–smooth muscle antibodies. Liver biopsy shows features that are typical of all chronic types of hepatitis, except plasma cell infiltrates.

TREATMENT

Rx

The clinical symptoms and liver test abnormalities of autoimmune hepatitis generally respond promptly to prednisone, usually at a dose of 20 to 30 mg/day, with a decrease in serum aminotransferase levels to the normal or near-normal range within 1 to 3 months; higher doses may be required in patients with more severe disease. Lack of a biochemical or clinical response should lead to reevaluation of the diagnosis. Azathioprine 50 to 100 mg can be combined with prednisone or added later to reduce long-term steroid side effects. Maintenance doses, which are typically required indefinitely, are often 5 to 10 mg/day of prednisone combined with 50 to 150 mg/day of azathioprine. Sometimes patients can be maintained on azathioprine (2 mg/kg/day) alone. After 3 years or more of remission, therapy can be carefully withdrawn, but severe and even fatal flares can occur weeks to months later.

PROGNOSIS

The prognosis is generally related to the histologic stage of the disease. Patients who initially respond to therapy may do well for many years. Patients who progress to end-stage liver disease require liver transplantation (Chapter 154).[15]

CRYPTOGENIC CHRONIC LIVER DISEASE

Cryptogenic chronic liver disease refers to chronic hepatitis or cirrhosis of unknown cause after excluding hepatitis B, C, D, and E; autoimmune hepatitis; steatohepatitis (Chapter 152); alcoholic liver disease (Chapter 152); drug-induced hepatitis (Chapter 150); and inherited and metabolic liver diseases (Chapter 146). Tests to exclude these conditions include serum levels of α_1-antitrypsin, iron, and ceruloplasmin and, if necessary, urine and liver copper concentrations. In its later stages, nonalcoholic steatohepatitis may be associated with little or no fat.

Grade A References

A1. Loomba R, Rowley A, Wesley R, et al. Systematic review: the effect of preventive lamivudine on hepatitis B reactivation during chemotherapy. *Ann Intern Med*. 2008;148:519-528.
A2. Lau GK, Piravisuth T, Luo KX, et al. Peginterferon alfa-2a, lamivudine and the combination for HBeAg-positive chronic hepatitis B. *N Engl J Med*. 2005;352:2682-2695.
A3. Chang TT, Gish RG, de Man R, et al. A comparison of entecavir and lamivudine for HBeAg-positive chronic hepatitis B. *N Engl J Med*. 2006;354:1001-1010.
A4. Marcellin P, Gane E, Buti M, et al. Regression of cirrhosis during treatment with tenofovir disoproxil fumarate for chronic hepatitis B: a 5-year open-label follow-up study. *Lancet*. 2013;381:468-475.
A5. Lawitz E, Mangia A, Wyles D, et al. Sofosbuvir for previously untreated chronic hepatitis C infection. *N Engl J Med*. 2013;368:1878-1887.
A6. Jacobson IM, Gordon SC, Kowdley KV, et al. Sofosbuvir for hepatitis C genotype 2 or 3 in patients without treatment options. *N Engl J Med*. 2013;368:1867-1877.
A7. Kowdley KV, Gordon SC, Reddy KR, et al. Ledipasvir and sofosbuvir for 8 or 12 weeks for chronic HCV without cirrhosis. *N Engl J Med*. 2014;370:1879-1888.
A8. Zeuzem S, Dusheiko GM, Salupere R, et al. Sofosbuvir and ribavirin in HCV genotypes 2 and 3. *N Engl J Med*. 2014;370:1993-2001.
A9. Sulkowski MS, Gardiner DF, Rodriguez-Torres M, et al. Daclatasvir plus sofosbuvir for previously treated or untreated chronic HCV infection. *N Engl J Med*. 2014;370:211-221.
A10. Kowdley KV, Lawitz E, Poordad F, et al. Phase 2b trial of interferon-free therapy for hepatitis C virus genotype 1. *N Engl J Med*. 2014;370:222-232.
A11. Afdhal N, Zeuzem S, Kwo P, et al. Ledipasvir and sofosbuvir for untreated HCV genotype 1 infection. *N Engl J Med*. 2014;370:1889-1898.
A12. Afdhal N, Reddy KR, Nelson DR, et al. Ledipasvir and sofosbuvir for previously treated HCV genotype 1 infection. *N Engl J Med*. 2014;370:1483-1493.
A13. Feld JJ, Kowdley KV, Coakley E, et al. Treatment of HCV with ABT-450/r-ombitasvir and dasabuvir with ribavirin. *N Engl J Med*. 2014;370:1594-1603.
A14. Zeuzem S, Jacobson IM, Baykal T, et al. Retreatment of HCV with ABT-450/r-ombitasvir and dasabuvir with ribavirin. *N Engl J Med*. 2014;370:1604-1614.
A15. Poordad F, Hezode C, Trinh R, et al. ABT-450/r-ombitasvir and dasabuvir with ribavirin for hepatitis C with cirrhosis. *N Engl J Med*. 2014;370:1973-1982.
A16. Ferenci P, Bernstein D, Lalezari J, et al. ABT-450/r-ombitasvir and dasabuvir with or without ribavirin for HCV. *N Engl J Med*. 2014;370:1983-1992.
A17. Wedemeyer H, Yurdaydin C, Dalekos GN, et al. Peginterferon plus adefovir versus either drug alone for hepatitis delta. *N Engl J Med*. 2011;364:322-331.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. LeFevre ML. Screening for hepatitis B virus infection in nonpregnant adolescents and adults: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;161:58-66.
2. Denniston MM, Jiles RB, Drobeniuc J, et al. Chronic Hepatitis C Virus Infection in the United States, National Health and Nutrition Examination Survey 2003 to 2010. *Ann Intern Med*. 2014;160:293-300.
3. Moyer VA. Screening for hepatitis C virus infection in adults: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2013;159:349-357.
4. Pawlotsky JM. New hepatitis C virus (HCV) drugs and the hope for cure: concepts in anti-HCV drug development. *Semin Liver Dis*. 2014;34:22-29.
5. European Association for the Study of the Liver. EASL recommendations on treatment of hepatitis C 2014. *J Hepatol*. 2014;61:373-395.
6. Kohli A, Shaffer A, Sherman A, et al. Treatment of hepatitis C: a systematic review. *JAMA*. 2014;312:631-640.
7. American Association for the Study of Liver Diseases, Infectious Diseases Society of American. Recommendations for testing, managing, and treating hepatitis C. http://www.hcvguidelines.org/full-report-view. Accessed March 5, 2015.
8. Pawlotsky JM. New hepatitis C therapies: the toolbox, strategies, and challenges. *Gastroenterology*. 2014;146:1176-1192.
9. Morgan RL, Baack B, Smith BD, et al. Eradication of hepatitis C virus infection and the development of hepatocellular carcinoma: a meta-analysis of observational studies. *Ann Intern Med*. 2013;158:329-337.
10. Pavri TM, Herbst DA, Reddy KR. Chronic hepatitis E virus infection: challenges in diagnosis and recognition in the United States. *J Clin Gastroenterol*. 2015;49:86-88.
11. Hoofnagle JH, Nelson KE, Purcell RH. Hepatitis E. *N Engl J Med*. 2012;367:1237-1244.
12. Kamar N, Izopet J, Tripon S, et al. Ribavirin for chronic hepatitis E virus infection in transplant recipients. *N Engl J Med*. 2014;370:1111-1120.
13. Heneghan MA, Yeoman AD, Verma S, et al. Autoimmune hepatitis. *Lancet*. 2013;382:1433-1444.
14. Grønbæk L, Vilstrup H, Jepsen P. Autoimmune hepatitis in Denmark: incidence, prevalence, prognosis, and causes of death—a nationwide registry-based cohort study. *J Hepatol*. 2014;60:612-617.
15. Carbone M, Neuberger JM. Autoimmune liver disease, autoimmunity and liver transplantation. *J Hepatol*. 2014;60:210-223.

REVIEW QUESTIONS

1. Which of the following viruses can cause chronic hepatitis?

A. Only hepatitis A virus
B. Hepatitis B and C viruses
C. Hepatitis B, C, and D viruses
D. Hepatitis B, C, D, and E viruses
E. Hepatitis A, B, C, D, and E viruses

Answer: D Hepatitis A virus never causes chronic infection and related liver disease. HBV infection becomes chronic in 95% of infants at birth and in less than 1% of adults. HCV infection becomes chronic in approximately 80% of cases. HDV infection rarely becomes chronic in cases of coinfection but almost systematically becomes chronic in case of superinfection in a chronic HBV carrier. HEV infection can become chronic in immunosuppressed patients.

2. Which of the following viruses are the main causes of hepatocellular carcinoma (primary liver cancer) worldwide?

A. Only hepatitis A virus
B. Hepatitis A and B viruses
C. Hepatitis B and C viruses
D. Hepatitis A and C viruses
E. Hepatitis C and E viruses

Answer: C Because of the prevalence of their chronic infection, their carcinogenic properties, and their ability to induce cirrhosis after many years of infection, both HBV and HCV currently are the main causes of hepatocellular carcinoma, with a predominance of HBV in developing regions and HCV in industrialized regions.

3. Which of the following sentences about hepatitis E virus (HEV) is true?

A. HEV infections are always self-limiting.
B. Chronic HEV infection occurs exclusively in immunosuppressed patients.
C. HEV infection cannot be diagnosed biologically.
D. The treatment of chronic HEV infection is based on antiviral molecules active only against HEV.
E. HEV RNA is never detected in blood.

Answer: B HEV infection is generally self-limiting but may become chronic in immunosuppressed patients. It is diagnosed by the presence of anti-HEV antibodies, in particular anti-HEV IgM at the acute stage, although HEV RNA can be detected in feces and blood. Treatment is based on the use of pegylated IFN-α and/or ribavirin, nonspecific antiviral molecules that are also used in the treatment of chronic hepatitis B or C.

150

TOXIN- AND DRUG-INDUCED LIVER DISEASE

WILLIAM M. LEE

DEFINITION

Toxin-induced and drug-induced hepatotoxicity, defined as any degree of liver injury caused by a drug or a toxic substance, is a frequent cause of acute liver injury and accounts for more than 50% of all cases of acute liver failure with hepatic encephalopathy in the United States. Hepatotoxicity has been described with many drugs, although the number of cases is low, given the number of prescriptions written.[1]

EPIDEMIOLOGY

Few data are available on the epidemiology of toxin-induced and drug-induced liver disease. The precise number of drug-induced liver injuries in the United States is unknown, but European data on adverse drug reactions indicate approximately 20 cases of drug-induced liver disease per 1 million people per year.[2] In developing parts of the world, drug-induced liver disease is much less common and is related to fewer drugs. It is estimated that less than 10% of actual cases are reported, so the true incidence of toxin-induced and drug-induced liver disease may be difficult to determine except in relatively closed populations.

PATHOBIOLOGY

The liver is central to the metabolism of exogenous substances. Most drugs and xenobiotics cross the intestinal brush border because they are lipophilic.

Biotransformation is the process by which lipophilic therapeutic agents are rendered more hydrophilic by the liver, resulting in drug excretion in urine or bile. In most instances, biotransformation changes a nonpolar to a polar compound through several steps. Foremost are oxidative pathways (e.g., hydroxylation) mediated by the cytochromes (CYPs) P-450 (Chapter 29). The next step is typically esterification to form sulfates and glucuronides, a process that results in the addition of highly polar groups to the hydroxyl group. These two enzymatic steps are referred to as *phase I* (CYP oxidation) and *phase II* (esterification). Other important metabolic pathways involve glutathione-*S*-transferase, acetylating enzymes and alcohol dehydrogenase, but the principal metabolic pathways for most pharmacologic agents involve CYPs and subsequent esterification.

Pathogenesis

The exact details of the pathogenesis of liver injury are unclear for most drugs.[3] A single drug may cause toxic effects in several ways. One overarching approach suggests that high-energy unstable metabolites of the parent drug, the result of CYP activation, bind to cell proteins or DNA and disrupt cell function. Perhaps the best example is acetaminophen. Although used universally for non-narcotic pain relief, acetaminophen, when taken in large quantities, causes profound centrilobular necrosis. The metabolic pathway of acetaminophen involves phase I and phase II reactions, glutathione detoxification, and the formation of reactive intermediates (E-Fig. 150-1). The presence of alcohol, which competes for CYP P-450 2E1, not only inhibits the formation of *N*-aminoparaquinoneamine (NAPQI) but also induces the enzyme so that its half-life is slowed and more enzyme is present. After the cessation of alcohol ingestion, NAPQI formation is enhanced by the presence of the induced enzyme and the lack of competition from alcohol. Toxicity is a dynamic process and may be most pronounced in the 24 hours after the cessation of alcohol. Glucuronidation and sulfation occur as the initial detoxifying step because the parent compound contains a hydroxyl group. Glucuronidation and sulfation capacity greatly exceeds daily needs, so even patients with very advanced liver disease continue to have adequate glucuronidation capacity, which explains why no obvious enhancement of toxicity is observed when cirrhotic patients take acetaminophen.

TABLE 150-1 FACTORS THAT MAY INFLUENCE THE METABOLIC FATE OF DRUGS

AGE
The elderly seem to be affected more often; adults are more susceptible than children to some drugs (acetaminophen, halothane, isoniazid) and less susceptible to others (aspirin, valproic acid)
ALCOHOL: ACUTE AND CHRONIC INGESTION
Induction of CYP2E1 affects drugs metabolized by this pathway, including acetaminophen and isoniazid
GENDER
Females are affected more often, but the reason is unknown
PREGNANCY
Effects of drugs in pregnancy have been poorly studied
PREEXISTING LIVER DISEASE
Hepatic disease may *protect* against idiosyncratic reactions and may *enhance* the toxicity of dose-dependent hepatotoxins (e.g., acetaminophen)
RENAL DISEASE
Slowed disappearance of the parent compound yields higher concentrations and affects P-450 (e.g., enhancement of tetracycline toxicity in renal disease)
CERTAIN FOODS
Grapefruit has an unknown substance that interferes with the metabolism of some drugs
CONCOMITANT DRUGS
Drug-drug interactions are common causes of adverse effects (e.g., valproate and chlorpromazine together lead to enhanced cholestasis)
GENETIC FACTORS
Enzyme polymorphisms (e.g., enhanced phenytoin liver disease in patients with defective epoxide hydrolase activity), HLA phenotypes (e.g., nitrofurantoin susceptibility)

Genetics

Enzyme Polymorphism

Although acetaminophen is a dose-related toxin, the rarity of idiosyncratic drug toxicity (1 in 10,000 patients) suggests the importance of environmental and host factors (Table 150-1). Genetically variant CYP isoenzymes may partially explain the observed individual variation in response to drugs. An example is debrisoquine, an antihypertensive drug marketed in Europe that is hydroxylated by CYP2D6, an isoform that is totally absent in 5% of normal individuals. Lack of CYP2D6 greatly prolongs the half-life of the parent compound in affected individuals. Another example is the phenomenon of fast versus slow acetylation, which affects different ethnic groups and has been implicated in the differential metabolism of isoniazid. Most of the known genetic variants that occur relatively frequently, however, cannot explain the formation of a toxic intermediate in only a rare individual.

Most drugs are small organic compounds, unlikely to evoke an immune response. Although some toxic drug reactions are associated with an obvious allergic response, most are not. Nevertheless, immune mechanisms not associated with systemic allergic immunoglobulin E (IgE) reactions or skin hypersensitivity might be involved. Studies suggest that the products of CYP P-450 metabolism, the highly reactive intermediates formed within the microsomes, covalently bind to the enzyme itself to form a drug-hapten adduct that disables the enzyme and injures the cell. Haptenization then evokes an immune response directed against the newly formed antigen or neoantigen. P-450s have been shown to traffic to the plasma membrane, thereby allowing the drug–P-450 adduct to become the target of a subsequent cytolytic attack. It is unclear whether the targets are these adducts or the smaller peptides processed and presented by the major histocompatibility complex class I and class II schemes. The association among neoantigens, autoantibodies, and hepatotoxic drugs implicates an immunologic mechanism, as does latency, which is the delay between the first ingestion and evidence of toxicity.[4]

Regardless of whether an individual drug causes significant cell necrosis, the drug–P-450 adducts can evoke an immune response. Any subsequent drug–P-450 adduct present on the hepatocyte surface would evoke a further response. Responses may be antibody mediated or occur as a result of direct cytolytic attack by primed T cells. Specific genetically determined components of the immune response may be important. A specific human leukocyte antigen (HLA) haplotype has been associated with hepatitis induced by amoxicillin-clavulanate and other polymorphisms encoding for increased susceptibility have been identified. However, it is unlikely that a single polymorphism will be found for most hepatocellular reactions, even when the phenotype of injury is well-characterized, such as for isoniazid. For every patient with a severe injury caused by drugs, there are often many more individuals with asymptomatic aminotransferase elevations that subsided despite continuing the drug—sometimes referred to as an *adaptive response.*

Other Mechanisms

In drug-induced cholestasis, disruption of specific transport channels in hepatocytes or cholangiocytes may be the key event. Estrogen or androgenic steroids may cause multiple canalicular membrane transport changes that affect, among others, the canalicular bile salt pump. For a few drugs, a specific uncoupling of mitochondrial respiration may lead to microvesicular steatosis and lactic acidosis.

Hepatotoxic Agents

Although there are a few dose-related toxins, most drugs involved in liver disease cause idiosyncratic, unpredictable toxicity.[5] The use of herbal and dietary supplements and the toxicity resulting from them appears to be increasing in the United States. Most common are body building supplements containing androgenic steroids that lead to severe cholestasis; weight loss products may cause hepatocellular injury of varying severity, including occasional fatalities.[6]

Intrinsic (Dose-Dependent) Agents

Acetaminophen (see later) and a few other agents seem to have a clear dose-response effect, although idiosyncrasy usually plays a role as well (Table 150-2). Some toxins, such as α-amanitine produced by *Amanita* mushrooms, cause dose-related injury. *Amanita* poisoning may occur after ingestion of the mushrooms *Amanita phalloides* (death cap) or *Amanita virosa* (deadly agaric). The dose-dependent toxic effect on the liver is attributed to amatoxin, an ingredient of the mushrooms that enhances the toxic effect by its enterohepatic recirculation characteristics; the toxic effect is exerted on each cycle of the recirculation through the liver.

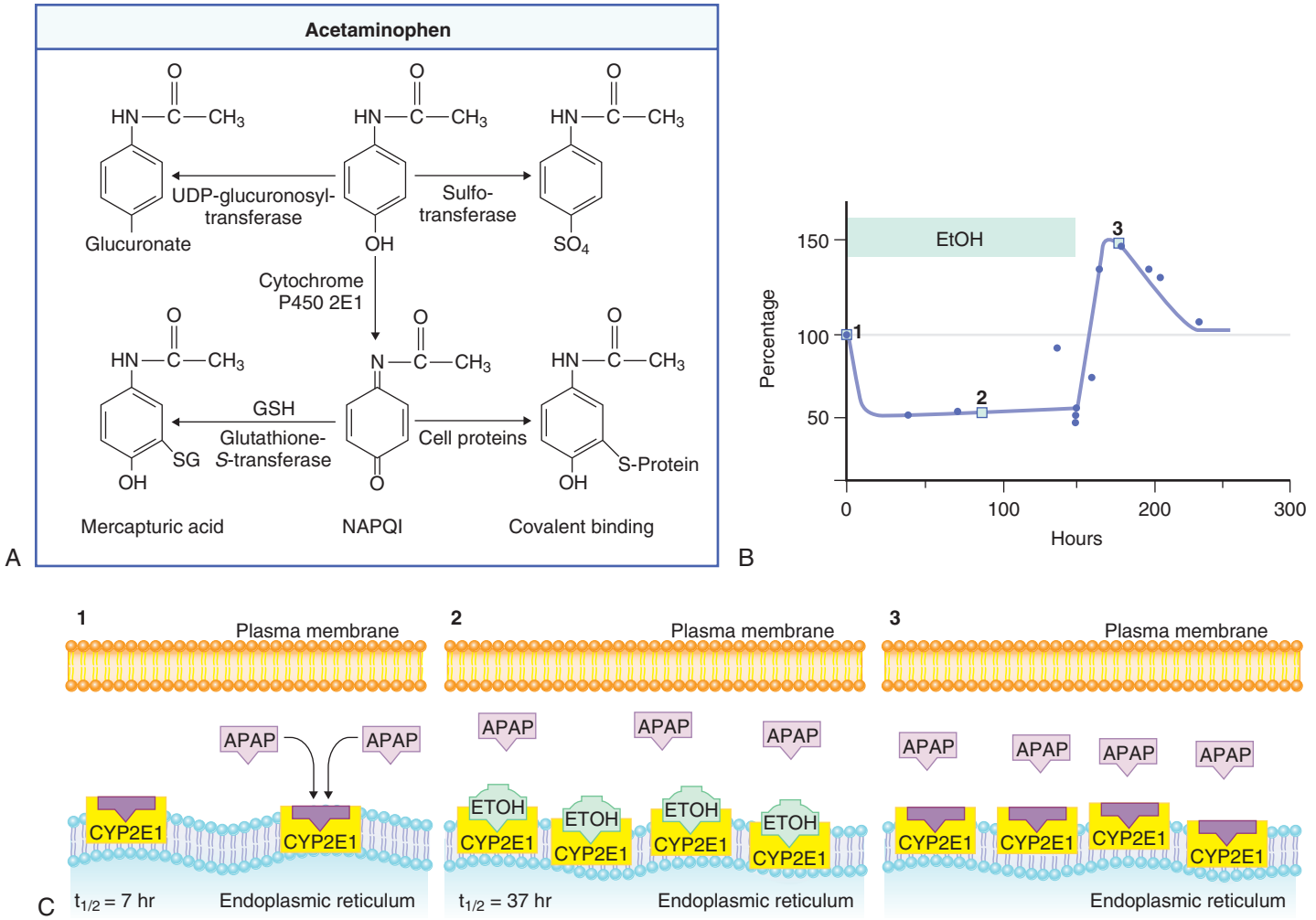

E-FIGURE 150-1. Acetaminophen metabolism. Although acetaminophen (*APAP*) is metabolized largely by sulfation or glucuronidation under normal conditions, exceeding the recommended dose increases the proportion metabolized by cytochrome P-450 2E1, which leads to the highly reactive intermediate *N*-aminoparaquinoneamine (*NAPQI*) (**A**). This compound leads to liver injury unless scavenging by glutathione yields the inert, water-soluble mercapturic acid. The presence of alcohol (*EtOH*), which competes for cytochrome P-450 2E1, not only inhibits NAPQI formation but also induces the enzyme so that its half-life is slowed and more enzyme is present (**B** and **C**). After cessation of alcohol, NAPQI formation is enhanced by the presence of the induced enzyme and the lack of competition from alcohol. Toxicity is a dynamic process and may be most pronounced in the 24 hours after cessation of alcohol. (Modified from Thummel KE, Slattery JT, Ro H, et al. Ethanol exposure and acetaminophen hepatotoxicity: inhibition and induction of hepatotoxic metabolite formation. *Clin Pharmacol Ther.* 2000;67:5991-5999.)

TABLE 150-2 DRUGS AND TOXINS IN WHICH A DOSE-RESPONSE EFFECT IS OBSERVED

DRUG OR TOXIN	RESPONSE
Acetaminophen	Total dose, single vs. multiple time points
Amiodarone	Total dose over time
Bromfenac	Toxicity occurs only after extended use
Cocaine	Dose-related vascular collapse
Cyclophosphamide	Dose related, worse with previous ALT elevations
Cyclosporine	Cholestasis with toxic blood levels, CYP3A phenotype
Methotrexate	Aminotransferase, fibrosis; single dose/total dose
Niacin	Large doses cause vascular collapse
Oral contraceptives	Prolonged use causes hepatic adenomas
Tetracycline	Total dose, renal dysfunction
Toxins (yellow phosphorus, carbon tetrachloride, *Amanita* toxin, bacterial toxins)	Total dose

ALT = alanine aminotransferase.

TABLE 150-3 TYPES OF TOXIN AND DRUG REACTIONS

REACTION TYPE	IMPLICATED DRUGS OR TOXINS
Autoimmune (attack on cell surface markers)	Lovastatin, methyldopa, nitrofurantoin
Cholestatic (attack on bile ducts)	Anabolic steroids, carbamazepine, chlorpromazine, estrogen, erythromycin
Fibrosis (activation of stellate cells leads to fibrosis)	Methotrexate, vitamin A excess
Granulomatous (macrophage stimulation)	Allopurinol, diltiazem, nitrofurantoin, quinidine, sulfa drugs
Hepatocellular (damage to smooth endoplasmic reticulum and immune cell surface)	Acetaminophen, *Amanita* poisoning, diclofenac, isoniazid, lovastatin, nefazodone, trazodone, venlafaxine
Immunoallergic (cytotoxic cell attack on surface determinants)	Halothane, phenytoin, sulfamethoxazole
Mixed (see earlier)	Amoxicillin-clavulanate, carbamazepine, cyclosporine, herbs, methimazole
Oncogenic (hepatic adenoma formation)	Oral contraceptives, androgenic agents
Steatohepatitis (mitochondrial dysfunction: β-oxidation and respiratory chain)	Amiodarone, perhexiline maleate, tamoxifen
Vascular collapse (ischemic damage)	Cocaine, ecstasy, nicotinic acid
Veno-occlusive disease (endotheliitis of sinusoidal endothelial cells)	Busulfan, cytoxan

Idiosyncratic Reactions

Drug reactions occur in 1 in 1000 to 1 in 200,000 patients. Characteristics of these idiosyncratic reactions include their infrequent occurrence, varying time intervals between the initial exposure and the reaction, and varying severity of reactions in affected individuals. There are also similarities such as "class effects" (similar drugs exhibit similar features; Table 150-3), a consistent pattern for each drug, and the fact that rechallenge with a responsible agent usually leads to a more severe reaction with a shorter latency than seen after the initial exposure.

Antibiotics, anticonvulsants, and nonsteroidal anti-inflammatory drugs are associated more frequently with drug-induced liver disease, whereas hormones, antihypertensive drugs, digoxin, and antiarrhythmic drugs are implicated rarely. In some cases, idiosyncratic reactions are so infrequent that a drug continues to be used if its effectiveness or uniqueness makes the risk acceptable. An example is isoniazid, which is among the few drugs implicated in drug-induced liver injury in developing countries. In individuals receiving isoniazid as a single agent for tuberculosis prophylaxis, increased aminotransferase levels may develop in 15 to 20%, but severe hepatic necrosis develops in only 0.1 to 1% (Chapter 324)—a rate that is high in comparison to idiosyncratic drug reactions, yet low enough for isoniazid, because of its effectiveness, to remain a key drug. Drugs that require relatively high dosing, more than 100 mg/day, and that have high lipophilicity are associated with most liver injury, not those at lower dose levels.[7]

CLINICAL MANIFESTATIONS

Patients may have few or nonspecific complaints despite very elevated aminotransferase levels. Clinical features include nausea, fatigue, occasional right upper quadrant pain, and nonspecific symptoms similar to those seen in other forms of acute hepatitis (Chapter 148). Fever or pharyngitis (typically seen in phenytoin reactions) may be present. No specific physical findings to raise suspicion of drug toxicity are noted, except possibly a rash. Any patient in whom jaundice develops is at risk for having a severe or fatal outcome, and patients who continue taking the drug despite jaundice are at highest risk.

DIAGNOSIS

Abnormal aminotransferase levels with the use of a new drug should raise the suspicion for a drug-induced reaction and prompt immediate discontinuation of the drug rather than awaiting diagnostic tests to confirm or exclude the diagnosis. Immediate discontinuation of medication at the first sign of liver disease can prevent most fatal liver injuries.

Evaluation of a patient with a suspected drug reaction is directed toward establishing the timeline for all drugs or herbs the patient may have taken. Responsible drugs have usually been started between 5 and 90 days before the onset of symptoms. Evidence of viral hepatitis (Chapters 148), gallstones (Chapter 155), alcoholic liver disease (Chapter 152), pregnancy (Chapter 239), severe right heart failure (Chapter 58), or a period of hypotension (Chapter 106) points to these specific causes. Less commonly, cytomegalovirus (Chapter 376), Epstein-Barr virus (Chapter 377), or herpesviruses (Chapters 374 and 375) can cause hepatic injury, primarily in immunosuppressed individuals. If all these causes can be excluded, the temporal relationship fits, and the patient begins to improve after withdrawal of the drug, the diagnosis is more secure. Liver biopsy is of limited value because the histologic picture in most cases of drug-induced liver injury is no different from that of viral hepatitis (Chapter 148). Nevertheless, an occasional liver biopsy specimen in an enigmatic case might reveal eosinophils or granulomas, consistent with a drug reaction.

Types of Drug Reactions

Although most liver injury involves direct hepatocyte necrosis or apoptosis (hepatocellular injury), some drugs injure primarily the bile ducts or canaliculi and cause cholestasis without significant damage to hepatocytes. Other drugs affect sinusoidal cells or present a particular pattern of liver injury affecting multiple cell types (mixed type). Another approach to drug reactions emphasizes the histologic changes involved and the cell type (see Table 150-3 and E-Fig. 150-2).

Hepatocellular Reactions

Hepatocellular reactions are the most common type of drug-induced liver disease and account for 90% of cases (Table 150-4). They are characterized by a pattern of serum liver test results that reflect hepatocellular injury. Usually, improvement is quick after discontinuation of the drug (1 to 2 months), and fulminant, acute liver failure with hepatic encephalopathy develops in only a few patients.

Histologic findings include necrosis and cellular infiltration. The necrosis may be zonal (e.g., induced by acetaminophen or carbon tetrachloride) or diffuse (e.g., induced by halothane), and the inflammatory response consists of lymphocytes or eosinophils. Massive necrosis may cause acute liver failure and death.

Acetaminophen toxicity is the most common form of acute liver failure in the United States and is the best understood example of direct hepatocyte toxicity. The incidence of acetaminophen poisoning varies widely throughout the world, but it is becoming more frequent and widespread. Liver injury occurs predictably after an intentional suicidal overdose (Chapter 110); it also occurs when acetaminophen is used in excessive doses or sometimes even in therapeutic doses for pain relief. Enhanced toxicity occurs when patients are fasting or are chronic alcohol users because of enzyme induction and depletion of glutathione by alcohol and fasting; by comparison, acute alcohol intake may protect against acetaminophen toxicity during the period of alcohol ingestion. Thereafter, a rebound increase in available CYP2E1 results in increased toxicity in the 12 hours after ingestion because of enzyme

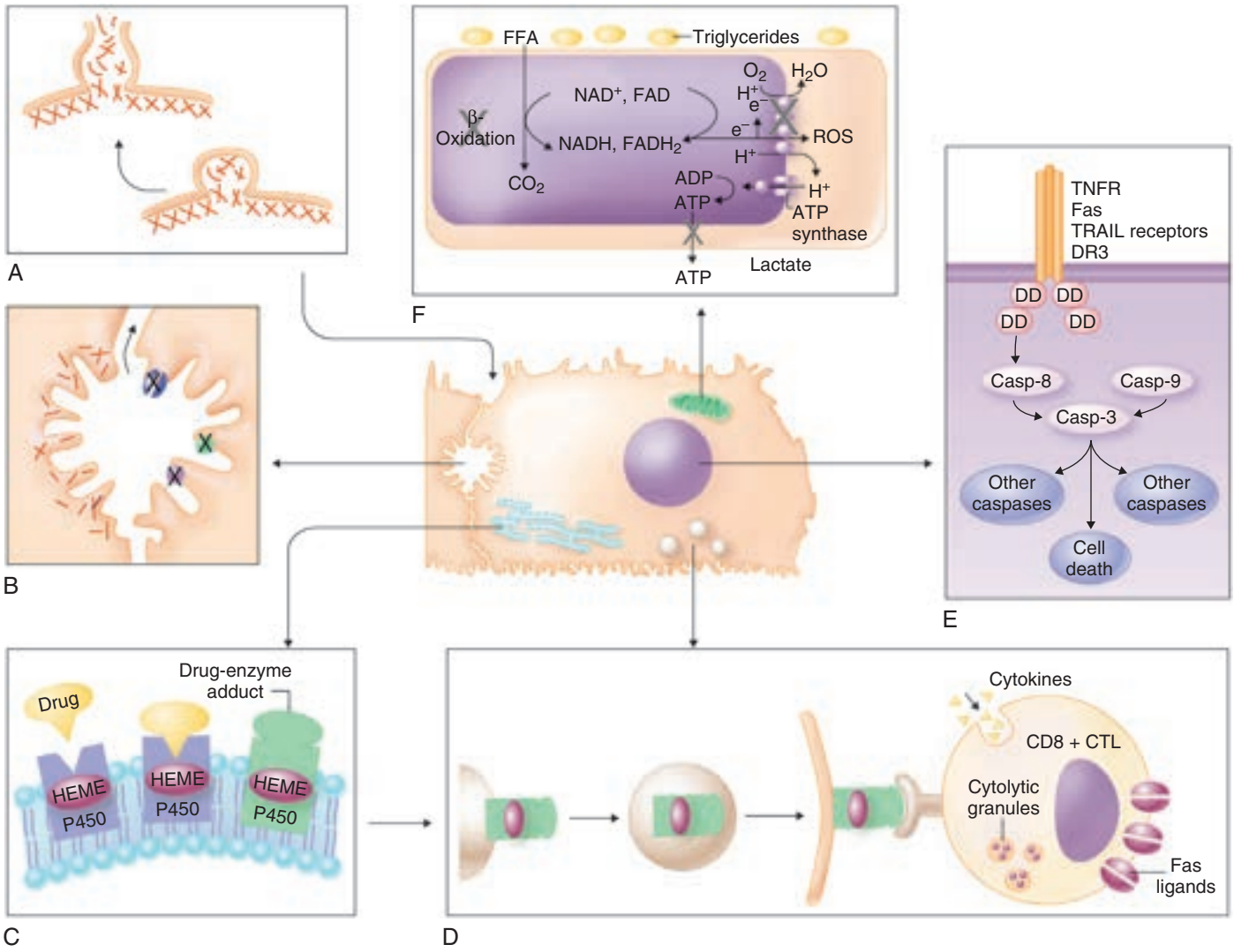

E-FIGURE 150-2. Mechanisms of liver injury. Each form of liver injury targets specific organelles, although multiple organelles may be affected. The hepatocyte in the center may be affected in at least six ways. **A,** Disruption of intracellular calcium leads to actin fibril disassembly at the hepatocyte surface, which results in blebbing of the cell membrane and subsequent rupture and cell lysis. **B,** In cholestatic diseases, disruption of actin filaments may occur with the loss of villous processes. Interference with ion pumps limits the excretion of bilirubin and other organic compounds. **C,** Most hepatocellular reactions involve the cytochrome P-450 system. The high-energy reaction involved may lead to binding of drug to enzyme and create a new adduct. **D,** Enzyme-drug adducts may traffic to the cell surface and serve as target immunogens for cytolytic attack by T cells. **E,** Activation of apoptotic pathways results in cell death. **F,** Inhibition of β-oxidation or respiration in mitochondria results in microvesicular fat accumulation and lactic acidosis, a pattern characteristic of a variety of agents, including nucleoside analogues, tetracycline, and aspirin. ATP = adenosine triphosphate; FAD = flavin adenine dinucleotide; FFA = free fatty acid; NAD = nicotinamide adenine dinucleotide; ROS = reactive oxygen species; TNFR = tumor necrosis factor receptor. (**A,** Modified from Farrell GC. *Drug-Induced Liver Disease*. Edinburgh: Churchill Livingstone; 1994: 44; **B,** Modified from Trauner M, Meier PJ, Boyer J. Mechanisms of disease: molecular pathogenesis of cholestasis. *N Engl J Med*. 1998;339:1217-1227; **C,** Modified from Watkins Zimmerman HJ. *Hepatotoxicity*, 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2000; **D,** From Robin M-A, Le Roy M, Descatoire V, Pessayre D. Plasma membrane cytochromes P450 as neoantigens and autoimmune targets in drug-induced hepatitis. *J Hepatol.* 1997;26(Suppl 1):23-30; **E,** From Reed JC. Apoptosis-regulating proteins as targets for drug discovery. *Trends Mol Med.* 2001;7:314-319; and **F,** Modified from Pessayre D, personal communication.)

TABLE 150-4 SCORING SYSTEM TO ASSESS CAUSALITY OF HEPATOCELLULAR REACTIONS

FACTOR	SCORE*
TEMPORAL RELATIONSHIP OF START OF DRUG TO START OF ILLNESS	
Initial treatment: 5-90 days; subsequent treatment course: 1-15 days	+2
Initial treatment: <5 or >90 days; subsequent treatment course: >15 days	+1
From cessation of drug: ≤15 days†	+1
COURSE	
ALT decreases ≥ 50% from peak within 8 days	+3
ALT decreases ≥ 50% from peak within 30 days	+2
If the drug is continued, inconclusive	0
RISK FACTORS	
Alcohol‡	+1
No alcohol‡	0
Age ≥ 55 yr	+1
Age < 55 yr	0
CONCOMITANT DRUG	
Concomitant drug with suggestive time of onset	−1
Concomitant drug known to be a hepatotoxin with suggestive time of onset	−2
Concomitant drug with further evidence of involvement (rechallenge)	−3
NUMBER OF NONDRUG CAUSES	
Hepatitis A, B, or C; biliary obstruction; alcoholism (AST ≥ 2 × ALT); recent hypotension; and CMV, EBV, and HSV infection all excluded	+2
4-5 causes excluded	+1
<4 causes excluded	−2
Nondrug cause highly probable	−3
PREVIOUS INFORMATION ON HEPATOTOXICITY OF DRUG IN QUESTION	
Package insert mentions	+2
Published case reports but not on package label	+1
Reaction unknown	0
RECHALLENGE	
Positive (ALT doubles with drug alone)§	+2
Compatible (ALT doubles, compounding features)§	+1
Negative (increase in ALT but ≤ 2 × ULN)§	−2
Not done	0

Modified from Danan G, Benichou C. Causality assessment of adverse reactions to drugs. I. A novel method based on the conclusions of international consensus meetings: application to drug-induced liver injuries. *J Clin Epidemiol.* 1993;46:1323-1330; and Benichou C, Danan G, Flahault A. Causality assessment of adverse reactions to drugs. II. An original model for validation of drug causality assessment methods: case reports with positive rechallenge. *J Clin Epidemiol.* 1993;46:1331-1336.
*Causality is highly probable (score >8), probable (score 6-8), possible (score 3-5), unlikely (score 1-2), or excluded (score ≤0).
†For cholestatic reactions, ≤ 30 days.
‡For cholestatic reactions, alcohol, or pregnancy.
§For cholestatic reactions, substitute alkaline phosphatase (or total bilirubin) for ALT.
ALT = alanine aminotransferase; AST = aspartate aminotransferase; CMV = cytomegalovirus; EBV = Epstein-Barr virus; HSV = herpes simplex virus; ULN = upper limit of normal.

induction (see E-Fig. 150-1). Patients with an unintentional acetaminophen overdose may fare worse than suicidal patients because the former seek treatment later in their course, even though suicidal patients take larger doses. The better outcome after an acute overdose may be explained by earlier medical attention and the use of *N*-acetylcysteine, an effective antidote. Nevertheless, one fifth of suicide attempts using acetaminophen are associated with severe liver injury and the potential for a fatal outcome.

The extremely elevated aminotransferase values (often > 6000 IU/L, and sometimes as high as 30,000 IU/L) observed in suicidal and unintentional acetaminophen ingestion help distinguish these cases from viral hepatitis or other drug injury. This signature of hyperacute injury (high aminotransferase levels, low bilirubin levels) are almost pathognomonic of acetaminophen injury, which can go unrecognized if a careful history is not elicited.[8] The availability of an antidote makes this diagnosis especially important. The antidote *N*-acetylcysteine (Chapter 110) may be given by nasogastric tube on admission and for the next 72 hours to provide glutathione substrate. The standard treatment is intravenous *N*-acetylcysteine beginning at a dose of 140 mg/kg in 300 mL of 5% dextrose given over 1 hour, followed by a dose of 70 mg/kg in 5% dextrose given over a 1-hour period every 4 hours for 48 hours. A loading dose of 140 mg/kg PO can be given, followed by 70 mg/kg every 4 hours for 17 doses (72 hours). Expected survival rates are greater than 80%, although liver transplantation is occasionally required.

Cholestatic Reactions

Cholestatic reactions have been described for many drugs. Cholestasis is best defined as failure of bile to reach the duodenum, and common symptoms are jaundice and pruritus. *Pure cholestasis,* with no signs of hepatocellular necrosis, is seen almost exclusively in patients taking oral contraceptives, anabolic steroids, or sex hormone antagonists such as tamoxifen. Acute *cholestatic hepatitis* is characterized histologically by cholestasis (dilated canaliculi, brown granules in the cytoplasm of hepatocytes), some degree of liver cell necrosis and bile duct injury, and inflammatory infiltration by polymorphonuclear leukocytes. Drugs that cause this type of reaction include carbamazepine, trimethoprim-sulfamethoxazole, captopril, and body building dietary supplements containing androgenic compounds

Generally, drug-induced cholestasis takes longer to resolve than drug-induced hepatotoxicity. In some cases, segments of the intrahepatic biliary tree may be destroyed progressively, the so-called vanishing bile duct syndrome that occurs after a protracted course (>6 months) of drug-induced cholestasis. The result is a state of chronic cholestasis that resembles primary biliary cirrhosis (Chapter 155). Approximately 30 drugs have been implicated in the vanishing bile duct syndrome, including levofloxacin and, occasionally, other antibiotics. A sclerosing cholangitis–like syndrome with jaundice caused by intrahepatic and extrahepatic strictures in the bile ducts is sometimes observed in patients receiving intra-arterial floxuridine chemotherapy for hepatic metastases of colorectal cancer.

Immunoallergic Reactions

Drugs also may be associated with definite allergic reactions. A combined toxic-immunologic mechanism is involved in liver injury caused by halothane, a fluorinated hydrocarbon anesthetic that causes severe, often fatal liver injury after multiple exposures (Chapter 432). Other fluorinated hydrocarbons, including isoflurane and desflurane, occasionally result in the same response. Although halothane has never been withdrawn, its use has been limited by the advent of safer agents. Hypersensitivity reactions, such as fever, eosinophilia, and rash, are common. Halothane may induce fever, eosinophilia, and antimitochondrial antibodies. Direct cytotoxicity and immune-mediated toxicity are observed, consistent with the clinical observation that severe halothane toxicity occurs with repeated exposure. Although evidence of injury can usually be identified within 1 week of the first exposure, the interval to toxicity is shortened and the damage is more severe with each successive exposure, as befits an immune reaction.

Phenytoin (Chapter 403) induces the simultaneous onset of fever, rash, lymphadenopathy, and eosinophilia. The mechanisms responsible for the combined allergic and hepatotoxic reaction are unknown, but the slow resolution of the illness suggests that the allergen remains on the surface of the hepatocyte for weeks or months. A concurrent mononucleosis-like picture is frequently confused with a viral illness or streptococcal pharyngitis. If phenytoin is not discontinued promptly despite signs of hepatitis, a severe Stevens-Johnson drug eruption (Chapters 439 and 440) and prolonged fever may result. As with any therapeutic agent, rapid recognition of the presence of a toxic drug reaction and immediate discontinuation of the compound are key to limiting hepatic damage. Systemic features of an allergic reaction may not be obvious, even when eosinophilia or granulomas are present on liver biopsy.

Steatohepatitis

Fatty liver disease (Chapter 152) related to the metabolic syndrome is increasingly evident in the United States and elsewhere. Differentiating this underlying condition from de novo fatty liver caused by a drug reaction can be difficult. In addition, certain agents such as statins may be associated with aminotransferase elevations independent of fatty liver disease. As a general rule, statins very rarely cause significant liver injury and should not be withheld from patients with hypercholesterolemia, even when fatty liver is present (Chapter 206). Steatosis in the liver (Chapter 152) can be present in a

microvesicular or macrovesicular pattern. Macrovesicular steatosis, the most common form, is characterized histologically by a single vacuole of fat filling up the hepatocyte and displacing the nucleus to the cell's periphery. Macrovesicular steatosis is typically caused by alcohol, diabetes, or obesity. Sometimes drugs such as corticosteroids or methotrexate may cause these hepatic changes. Amiodarone (Chapters 64 and 65) has been associated with a picture resembling alcoholic hepatitis, occasionally with progression to cirrhosis. The pathophysiology involves accumulation of phospholipids in the liver, eyes, thyroid, and skin. Treatment is primarily withdrawal of the drug and observation, although the half-life of amiodarone is prolonged. Tamoxifen, which has been used in long-term regimens for the prevention of recurrent breast cancer (Chapter 198), has also been associated with steatohepatitis evolving to cirrhosis.

In microvesicular steatosis, hepatocytes contain numerous small fat vesicles that do not displace the nucleus. Valproic acid, an anticonvulsant (Chapter 403), causes hepatotoxicity, either as a result of microvesicular fat deposition, resembling Reye syndrome, or in a more chronic, indolent fashion associated with macrovesicular fat accumulation. Toxicity is more severe and frequent in children. These lesions are associated with disruption of mitochondrial DNA, resulting in anaerobic metabolism that leads to lactic acidosis in the most severe cases. Macrovesicular and microvesicular lesions may be observed concomitantly in some patients, and microvesicular lesions are more often associated with a poor prognosis. Hepatocellular necrosis also may be present. Acute fatty liver of pregnancy (Chapters 147 and 239) and Reye syndrome are two examples of severe liver diseases caused by microvesicular steatosis.

Drugs involved in microvesicular steatosis include valproate, tetracycline, and fialuridine. Aspirin use in children has been associated with Reye syndrome, but the incidence of Reye syndrome has decreased dramatically since warnings were issued concerning aspirin use in children.

Effects of Sex Steroids

Anabolic steroids, such as methyltestosterone, may cause cholestasis. Androgens or estrogens may cause peliosis hepatis and benign or malignant tumors. Oral contraceptives (Chapter 238) may cause cholestasis, hepatic adenomas, or Budd-Chiari syndrome (hepatic vein thrombosis). Antiandrogens used to treat prostate cancer (Chapter 201), such as flutamide and nilutamide, and antipituitary drugs, such as cyproterone acetate, also have been associated with severe hepatocellular injury.

Other Drug Reactions

Other less severe drug reactions involving the liver include granulomatous reactions, fibrosis, ischemic injury, and chronic autoimmune liver injury (see Table 150-3). The type of reaction observed can be helpful in determining the probable agent because most drugs have a specific injury profile.

A pattern of veno-occlusive disease with obliteration of small intrahepatic veins, sinusoidal congestion, and necrosis is observed frequently in bone marrow transplant patients (Chapter 178) who receive chemotherapy with cyclophosphamide (Cytoxan) or busulfan. Symptoms, including rapidly accumulating ascites, painful hepatomegaly, and jaundice, develop soon after the chemotherapeutic regimen has begun. Oxiplatin can cause portal venular injury that leads to nodular regenerative hyperplasia and life-long portal hypertension. Rarely, herbal medicines (Chapter 39) such as pyrrolizidine alkaloids (*Crotalaria* and *Senecio* found in Jamaican bush tea) may cause veno-occlusive disease.

Toxins are associated with direct injury to hepatocytes in a dose-dependent fashion. Organic solvents such as carbon tetrachloride and trichloroethylene (Chapter 110) cause centrilobular injury. Yellow phosphorus, found in firecrackers and rat poisons, is a rare cause of liver injury from either accidental or intentional exposure. Symptoms of poisoning are similar to those of any other type of hepatitis.

Mushroom poisoning (Chapter 110), which follows the ingestion of *A. phalloides* and related species, typically occurs in amateur mushroom fanciers in a dose-related fashion. The associated muscarinic effects, including severe diarrhea, vomiting, and profuse sweating, predominate in the first hours after ingestion. Hepatic failure follows if antidotes (see later) are not given. The overall prognosis for spontaneous recovery is poor; liver transplantation may be life-saving.

Differential Diagnosis

The differential diagnosis of toxin-induced and drug-induced liver injury includes almost the entire spectrum of liver diseases. Some cases previously ascribed to drugs may now be linked to previously unsuspected hepatitis E (Chapters 148 and 149).[9] Because the clinical picture of drug-induced liver injury ranges from pure hepatocellular to pure cholestatic variants, a high index of suspicion must be maintained, even when toxin-induced or drug-induced liver injury is not obvious initially.

For dose-dependent hepatotoxins, the diagnosis may be easier to establish than for idiosyncratic drug reactions. Serum levels of acetaminophen, a thorough history, and characteristic biochemical abnormalities (high aminotransferase levels) usually reveal an acetaminophen overdose, whereas a diagnosis of *Amanita* poisoning depends on the history, symptoms of gastroenteritis (muscarinic reaction), and positive mushroom identification.

For idiosyncratic drug reactions, the diagnosis is sometimes more difficult to establish. A standardized reporting form called the Roussel-Uclaf Causality Assessment Method (RUCAM) (see Table 150-4), developed by an international panel, provides a worthwhile scoring system. These guidelines outline the steps an experienced clinician might use to assess the likelihood of a drug reaction. Causality assessment factors typically include the temporal relationship, course after cessation of the drug, risk factors, concomitant drugs, a search for nondrug causes (viral hepatitis), previous information concerning the drug, and response to rechallenge, which is typically not required. Recently, the Drug-Induced Liver Injury Network has used an expert opinion system that is helpful in defining phenotypes but is not readily applicable for day-to-day use.

TREATMENT

Rx

Prompt discontinuation of a suspected drug is mandatory. Available antidotes should be used for acetaminophen (*N*-acetylcysteine) and *Amanita* poisoning (penicillin 300,000 to 1 million U/kg/day intravenously (IV) and thioctic acid 5 to 100 mg every 6 hours IV have been recommended, but there are no controlled trials). General supportive therapy ranges from intravenous fluid replacement to intensive monitoring and treatment of patients with hepatic encephalopathy secondary to acute liver failure (Chapter 153). *N*-Acetylcysteine is the standard antidote for acetaminophen overdose, but it may improve outcomes in some cases of acute liver failure not associated with acetaminophen, such as severe alcoholic hepatitis.[10] Almost 10% of patients die or require liver transplantation within six months after serious drug-induced liver injury and about another 20% have persistent evidence of liver injury.[11] Liver transplantation (Chapter 154) is performed in more than 50% of patients with idiosyncratic drug-induced acute liver failure because the survival rate in this setting without transplantation is less than 20%.

FUTURE DIRECTIONS

Research in pharmacogenomics may allow the patient's own genetic information to guide individualized drug therapy and monitoring of idiosyncratic drug reactions. The genetic information would probably concentrate initially on enzymes with variant alleles associated with poor metabolism, such as CYP1A2 or CYP2C19 for isoniazid, CYP2C9 for piroxicam, or CYP2D6 for nortriptyline. Better postmarketing surveillance of all drugs to identify those with previously unappreciated hepatotoxicity should be a high priority.

PREVENTION

It is reasonable to consider a drug reaction whenever an episode of apparent hepatitis is unexplained, particularly if a new agent has been introduced in the previous 3 months. It is prudent to defer embracing new drugs during their first year of introduction, particularly if they show no unique advantages over accepted formulations. Physicians must strive to instill in their patients a healthy level of alertness with regard to drug-induced liver injury, particularly for agents with known hepatotoxicity. Monitoring of aminotransferase levels on a monthly basis is suggested for known hepatotoxins such as isoniazid or diclofenac, but it is unlikely to be cost-effective when adverse reactions occur less frequently, such as in only 1 in 50,000 patients. Because many drug reactions develop within days, monitoring provides no guarantee. Most fatal drug reactions could have been prevented if the offending agent were withdrawn immediately, at the first sign of illness.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hoofnagle JH, Serrano J, Knoben JE, et al. LiverTox: a website on drug-induced liver injury. *Hepatology*. 2013;57:873-874.
2. Bjornsson ES, Bergmann OM, Bjornsson HK, et al. Incidence, presentation, and outcomes in patients with drug-induced liver injury in the general population of Iceland. *Gastroenterology*. 2013;144:1419-1425.
3. Urban TJ, Shen Y, Stolz A, et al. Limited contribution of common genetic variants to risk for liver injury due to a variety of drugs. *Pharmacogenet Genomics*. 2012;22:784-795.
4. Uetrecht J, Naisbitt DJ. Idiosyncratic adverse drug reactions: current concepts. *Pharmacol Rev*. 2013;65:779-808.
5. Leise MD, Poterucha JJ, Talwalkar JA. Drug-induced liver injury. *Mayo Clin Proc*. 2014;89:95-106.
6. Navarro VJ, Barnhart H, Bonkovsky HL, et al. Liver injury from herbals and dietary supplements in the U.S. Drug-Induced Liver Injury Network. *Hepatology*. 2014;60:1399-1408.
7. Chen M, Borlak J, Tong W. High lipophilicity and high daily dose of oral medications are associated with significant risk for drug-induced liver injury. *Hepatology*. 2013;58:388-396.
8. Khandelwal N, James LP, Sanders C, et al. Unrecognized acetaminophen toxicity as a cause of indeterminate acute liver failure. *Hepatology*. 2011;53:567-576.
9. Davern TJ, Chalasani N, Fontana RJ, et al. Acute hepatitis E infection accounts for some cases of suspected drug-induced liver injury. *Gastroenterology*. 2011;141:1665-1672.
10. Bass S, Zook N. Intravenous acetylcysteine for indications other than acetaminophen overdose. *Am J Health Syst Pharm*. 2013;70:1496-1501.
11. Fontana RJ, Hayashi PH, Gu J, et al. Idiosyncratic drug-induced liver injury is associated with substantial morbidity and mortality within 6 months from onset. *Gastroenterology*. 2014;147:96-108.

REVIEW QUESTIONS

1. A 24-year-old woman is brought to the emergency department by her family because of altered mental status. She is rousable but drowsy, with normal vital signs except for tachycardia. Initial testing shows AST 10,200 IU/L, ALT 8,100 IU/L, INR 3.2, T bilirubin 3.7 mg/dL. The most plausible explanation for this scenario is:

A. Alprazolam (Xanax) overdose.
B. Alcohol binge drinking.
C. Aspirin overdose.
D. Acetaminophen overdose.
E. Methyl alcohol ingestion.

Answer: D Although drowsiness would follow alprazolam and possibly an alcoholic binge, only acetaminophen is associated with very high aminotransferases. Approximately 75% of acetaminophen overdoses occur in women.

2. The proper course of treatment for this patient would be:

A. Gastric lavage.
B. Normal saline at 250 mL/hour.
C. Syrup of ipecac.
D. Narcan by injection.
E. Intravenous *N*-acetylcysteine.

Answer: E *N*-Acetylcysteine is a sulfhydryl donor that repletes glutathione and is effective in preventing liver injury when given early in acetaminophen overdose. Even after severe liver damage has occurred, it is still likely to be of some benefit.

3. A college student who is trying out for the wrestling team is seen for new-onset itching and jaundice. The most likely cause would be:

A. Hepatitis A.
B. Hepatitis D.
C. Anabolic steroid use.
D. Hepatitis E.
E. Hepatitis C.

Answer: C Body-building supplements contain androgenic compounds highly associated with cholestatic liver disease, which is usually self-limited but very slow to resolve.

4. A 49-year-old truck driver with a body mass index of 42 is evaluated for life insurance. His cholesterol level is 285 mg/dL. He complains of right upper quadrant pain, and his abdominal ultrasound reveals a fatty liver. His AST is 42 IU/L, ALT is 34 IU/L. He should be considered for which of the following treatments?

A. Phen-phen for weight reduction
B. Herbal weight loss product
C. Lovastatin 40 mg/day
D. Daily aspirin 81 mg
E. Support stockings

Answer: C Lovastatin will deal with his hypercholesterolemia. He also needs exercise and dietary programs to help lose weight.

5. A 42-year-old woman is found incidentally to have multiple masses in her liver, but she is totally asymptomatic. Biopsy of one of the masses discloses an hepatic adenoma. Which of her medications should be discontinued at this time?

A. Oral contraceptive
B. Metformin
C. Lisinopril
D. Hydrochlorthiazide
E. Alprazolam

Answer: A Oral contraceptives, particularly after prolonged use, can be associated with hepatic adenomas. These adenomas regress but must be watched closely over time.

151 BACTERIAL, PARASITIC, FUNGAL, AND GRANULOMATOUS LIVER DISEASES

K. RAJENDER REDDY

INFECTIONS OF THE LIVER

Infections of the liver can be due to a variety of pathogens, including bacteria, fungi, amebae, protozoa, helminths, spirochetes, and rickettsiae. The manifestations of these infections are protean; some are generic to all infections, whereas others are specific to particular infections. The epidemiology can vary and depend on the geographic region of the world. In endemic areas, *Entamoeba histolytica* is a key consideration in the differential diagnosis of a liver abscess (Table 151-1).

Bacterial Infections

PYOGENIC LIVER ABSCESS

DEFINITION

Pyogenic liver abscess is a focal collection of purulent bacterial material and necroinflammatory debris. It can be solitary or multiple and can be caused by one or more aerobic and anaerobic bacteria (Fig. 151-1).

EPIDEMIOLOGY

Pyogenic liver abscess has an estimated global incidence of approximately 1.1 to 2.3 per 100,000 person-years, whereas in the United States, the incidence is approximately 3.6 per 100,000 and has been rising.[1] Biliary obstruction, caused by either a malignant or benign disease, accounts for 50 to 60% of pyogenic liver abscesses, whereas portal pyemia, due to appendicitis or other intra-abdominal infections, accounts for about 20% of cases. Recently, a number of studies from Asia, where *Klebsiella pneumoniae* is the primary etiologic source of pyogenic liver abscess, have proposed a correlation with underlying colorectal neoplasms,[2] some of which were evident at the time of diagnosis but also with a 5- to 8-fold higher risk of newly diagnosed colorectal cancers in the 3 years after diagnosis.[3] However, it is yet to be determined whether these findings can be extrapolated to other regions of the world.

PATHOBIOLOGY

Bacteria can enter the liver through the portal system from infections in areas drained by the mesenteric system into the portal system, such as appendicitis (Chapter 142). Other mechanisms for pyogenic liver abscess include bacterial cholangitis due to benign or malignant obstruction and infection of the liver from a systemic bacteremia, such as an infection of the oral cavity. Pyogenic liver abscess can also be caused by blunt or penetrating trauma, including such unusual causes as the ingestion of a toothpick or fish bone that can cause an intestinal perforation, fistula to the liver, and subsequent abscess formation. Liver abscesses can occur in a transplanted graft owing to vascular compromise caused by hepatic artery thrombosis and ischemic bile duct strictures.

Multiple organisms can cause pyogenic liver abscess. The most common organism, *Klebsiella pneumoniae*,[4] often is associated with biliary tract

TABLE 151-1 FEATURES OF BACTERIAL AND AMEBIC ABSCESSES

	DEMOGRAPHICS	RISK FACTORS	SYMPTOMS	LABORATORY FINDINGS	RADIOGRAPHIC FEATURES	DIAGNOSIS	TREATMENT
Bacterial liver abscess	50-70 years old Male = female	Recent bacterial infection, biliary obstruction, diabetes mellitus	Fevers, chills, malaise, anorexia, diarrhea, cough, pleuritic chest pain, RUQ pain	Leukocytosis, anemia, elevated alkaline phosphatase and bilirubin, low albumin, positive blood cultures (50%)	Multifocal (50%), usually right lobe, irregular margins	Aspirate (70-80% positive)	Percutaneous drainage and antibiotics
Amebic liver abscess	18-50 years old Male > female	Alcohol intake, HLA-DR3, oral and anal sex, contaminated enema apparatus, travel to or living in an endemic area	Fever, RUQ pain, hepatic tenderness, anorexia, weight loss, uncommon to have colitis	Leukocytosis, no eosinophilia, mild anemia, elevated alkaline phosphatase, elevated ESR, positive serology	Single abscess (80%), usually right lobe, wall enhancement seen on CT scan with IV contrast	Aspirate (trophozoites rarely seen) can rule out superimposed bacterial infection, positive serology and risk factors	Metronidazole and iodoquinol

RUQ = right upper quadrant; ESR = erythrocyte sedimentation rate; CT = computed tomography; IV = intravenous.

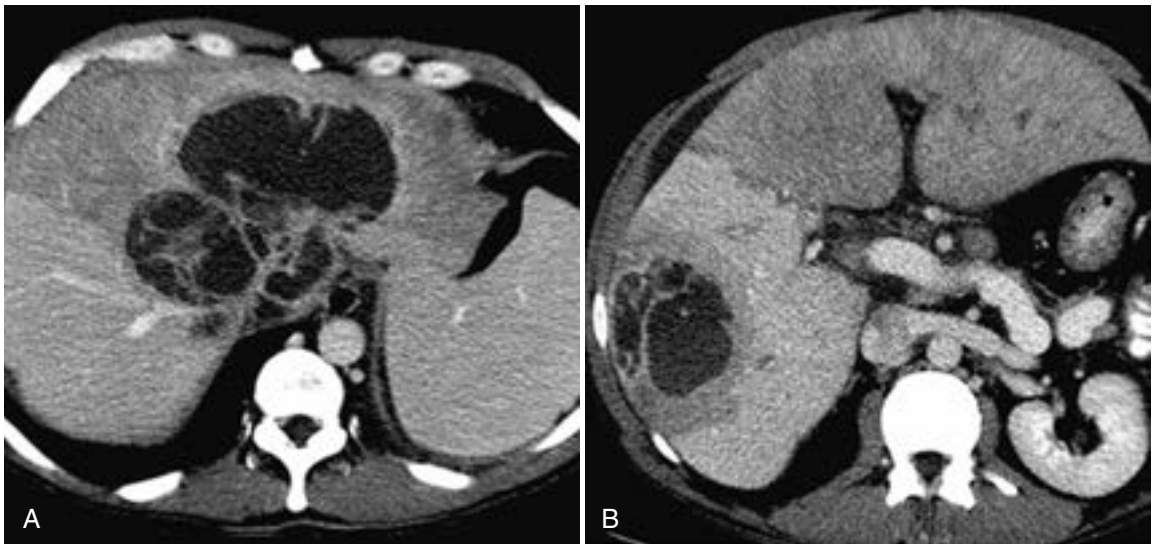

FIGURE 151-1. A and B, Computed tomographic scans of pyogenic liver abscess lesions. (Courtesy Dr. Chalermrat Bunchorntavakul, Bangkok, Thailand.)

disease. Other aerobes include *Escherichia coli*, group D streptococci, β-hemolytic streptococci, and *Staphylococcus aureus*. Anaerobic infection is often seen with colonic disease. Less common causes of liver abscesses include *Actinomyces*, *Nocardia asteroides*, *Yersinia pseudotuberculosis* and *Yersinia enterocolitica*, *Listeria monocytogenes*, *Campylobacter jejuni*, *Legionella pneumophila*, *Mycobacterium tuberculosis*, *Salmonella typhi* or *Salmonella paratyphi*, *Candida albicans*, and *Bartonella henselae*. Most often, the organism recovered from an abscess cavity is single, but multiple organisms can be isolated in as many as a third of patients. Bacteria may not be isolated from the abscess because of prior antibiotic therapy or the failure to perform proper anaerobic cultures.

CLINICAL MANIFESTATIONS

The signs and symptoms associated with a liver abscess typically include fever, right upper quadrant abdominal pain, chills, nausea, vomiting, weight loss, and jaundice. The presentation can be acute or indolent. An associated bacteremia is seen in approximately 50% of the patients, but frank sepsis is rare. About 10% of patients develop metastatic infection to other sites.[5]

DIAGNOSIS

Appropriate diagnosis requires a high degree of clinical suspicion, and diagnosis is sometimes delayed. Usually, however, the diagnosis is made promptly with the wide availability of the various radiologic modalities. The two common types of liver abscesses are pyogenic and amebic abscesses, and it is important to make a distinction because the prognosis and management differ. Amebic abscesses can become secondarily infected with other bacteria.

The diagnosis of an abscess is based on a constellation of clinical, bacteriologic, and radiologic features. Ultrasonography and computed tomography (CT) scanning are the most common radiologic modalities to diagnose an abscess cavity reliably, either as single or multiple lesions (see Fig. 151-1). CT is the preferred diagnostic indicator of an abscess, with sensitivity of more than 90%. Any concurrent biliary obstruction also can be diagnosed by these imaging modalities. Approximately 50% of patients have multiple abscesses. On CT, the lesion is seen as a fluid collection with irregular borders and wall edema. One drawback to this imaging modality is that no specific features differentiate a pyogenic abscess from other infectious causes (i.e., amebic or fungal). Another drawback is that a very early stage abscess may not be well formed and may have characteristics more suggestive of a solid mass. It is critical to distinguish an abscess from a tumor (Chapter 196) or a simple cyst, a process that can be complicated on noninvasive imaging by various events, such as bleeding into a cyst, calcification, necrosis, or bleeding into a tumor. A finding of smaller (<2 cm) peripheral lesions that surround a central abscess (the cluster sign) can help exclude a hepatic neoplasm. Calcification suggests bleeding into a tumor, but it also can be seen in the wall of an echinococcal cyst. Nonspecific radiologic features, including elevation of the right hemidiaphragm, right lower lung lobe atelectasis, and right pleural effusion, are seen in up to 30% of patients with pyogenic liver abscesses. Magnetic resonance imaging and tagged white blood cell scans add little to the diagnosis of a liver abscess.

Microbial cultures are essential. In addition to blood cultures, the abscess cavity should be aspirated percutaneously, with either ultrasound or CT guidance, and the aspirate should be cultured for aerobic and anaerobic organisms as well as for amebae if there is any suspicion of an amebic abscess. Blood cultures are positive in only approximately 50% of cases. In approximately 15 to 20% of cases, multiple organisms are identified in the abscess cavity, but 20 to 50% of cases may have negative cultures despite appropriate culture techniques.[6]

TREATMENT

Rx

Immediate broad-spectrum antibiotic coverage and the prompt identification and treatment of the source of the infection are essential for successful outcomes. Any underlying biliary source must be resolved, and any biliary obstruction must be relieved.

As soon as the causative bacterium is identified, the antibiotic regimen can be tailored appropriately. Recommendations should be guided by the culture and by prevailing bacterial resistance patterns. Monotherapy with a β-lactam or β-lactamase inhibitor such as ampicillin-sulbactam (3 g IV every 6 hours) or piperacillin-tazobactam (4.5 g IV every 6 hours) or ticarcillin-clavulanate (3.1 g IV every 4 hours) can be used, or a third-generation cephalosporin such as ceftriaxone (1 g IV daily) *plus* metronidazole (500 mg IV every 8 hours) can be considered. Other regimens for consideration include a fluoroquinolone (such as ciprofloxacin 400 mg IV every 12 hours or levofloxacin 500 or 750 mg IV daily) *plus* metronidazole (500 mg IV every 8 hours) and monotherapy with a carbapenem such as imipenem-cilastatin (500 mg IV every 6 hours), meropenem (1 g IV every 8 hours), or ertapenem (1 g IV daily).

The duration of therapy is partly based on the response to therapy. An average of 4 to 6 weeks of antibiotic therapy is reasonable, and the last 2 to 4 weeks of the antibiotic regimen can be administered orally. Abscesses that are difficult to drain or slow to demonstrate radiographic resolution require longer courses of therapy. Importantly, radiologic abnormalities resolve more slowly than clinical and biochemical features do, so the latter should be used as an indicator for tailoring of the therapeutic regimen.

Antibiotics alone can sometimes successfully resolve small, multiple pyogenic abscesses, but most patients will require drainage of the abscesses. The standard approach is placement of a drain for about 7 days. Alternatively, needle aspiration, repeated as needed when the abscess is large enough to be drained percutaneously, provides equivalent results.[■■] Surgical drainage seldom is the first option except in patients who also have a surgically correctable precipitating lesion, such as appendicitis or biliary obstruction.[7] More often, biliary obstruction is treated with endoscopic retrograde cholangiopancreatography or transhepatic cholangiography with accompanying biliary drainage. For pyogenic abscesses in transplanted livers, management and proper biliary drainage may be difficult to achieve because of the diffuse nature of the biliary strictures.

PROGNOSIS

Abscesses smaller than10 cm can take up to 16 weeks to resolve, whereas abscesses larger than10 cm may take, on average, an additional 6 weeks to resolve. Mortality from pyogenic liver abscess is associated with older age and with comorbidities such as cirrhosis, diabetes, chronic renal failure, and malignant disease. Jaundice is an ominous sign. In developed countries, the mortality rate ranges from 2 to 12%.[8]

NON-ABSCESS HEPATIC BACTERIAL INFECTIONS

Bacterial infections of the liver can also cause more diffuse infections without frank abscess formation. Implicated organisms include *Listeria monocytogenes* (Chapter 293), *Yersinia enterocolitica* (Chapter 312), *Salmonella typhi* and *Salmonella paratyphi* (Chapter 308), *Legionella* (Chapter 314), *Ehrlichia* (Chapter 327), and gonococci (Chapter 299). There are no specific features associated with these infections, and these organisms do not necessarily cause abscesses. Patients with chronic liver disease are especially at risk for *Listeria* infection. Patients with active enteric *Yersinia* infection can have secondary liver involvement with or without liver abscesses. Disseminated gonococcal infections (Chapter 299) can cause a perihepatitis (Fitz-Hugh–Curtis syndrome), which can be manifested with right upper quadrant pain and tenderness.

Systemic bacteremia can cause a variety of hepatobiliary abnormalities, which may range from elevated aminotransferase and alkaline phosphatase levels (Chapter 147) to the cholestasis of sepsis with the development of jaundice. Several organisms can disrupt normal liver function after entering the blood stream, the most common of which are *E. coli*, *Klebsiella* spp, *Streptococcus pneumoniae*, and *Staphylococcus aureus*. Typically, in patients suffering from jaundice, sources of bacteremia include pneumonia, urinary tract infection, and soft tissue infection, although organisms may originate from a range of other sites. The hepatic biochemical abnormalities associated with bacteremia may be related to factors such as hemodynamic instability and liver hypoperfusion as well as to the infection. Associated renal failure, a blood transfusion–derived bilirubin load, and drugs may complicate the picture. The course of cholestasis may be prolonged for several days to a few weeks, but it typically resolves with resolution of the systemic infection. There is no specific treatment for the cholestasis related to a systemic infection, but it is important to rule out drug-induced cholestasis that may evolve as a consequence of one or more of the antibiotics.

Fungal Diseases of the Liver

Except for hepatosplenic candidiasis, clinically significant fungal diseases of the liver are unusual. Typically, other fungal infections will also be manifested as liver granulomas but will not show the high, swinging pyrexia that is characteristic of hepatosplenic candidiasis. Tissue cultures are required to confirm a diagnosis.

HEPATOSPLENIC CANDIDIASIS

EPIDEMIOLOGY AND PATHOBIOLOGY

Hepatosplenic candidiasis typically is caused by *Candida albicans*, but other species, including *C. tropicalis*, *C. parapsilosis*, *C. glabrata*, and *C. krusei*, have occasionally been reported. It occurs as part of disseminated candidiasis

(Chapter 338), almost exclusively in patients with acute leukemia but rarely in patients with lymphoma, aplastic anemia, and sarcoma. With the current widespread use of prophylactic antifungal agents early in the course of disease in patients with acute leukemia, hepatosplenic candidiasis develops in only 1 to 2% of patients, more commonly with acute lymphoblastic leukemia than with acute myeloid leukemia. Hepatosplenic candidiasis presumably results from translocation of *Candida* species from the gastrointestinal tract into the blood stream as a result of prolonged neutropenia and a breach in mucosal integrity.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Hepatosplenic candidiasis is manifested with persistent and high spiking fevers in a patient who was previously neutropenic and has now returned to a normal neutrophil count. Right upper quadrant pain, nausea, vomiting, and anorexia may accompany fever.

Patients typically have elevated levels of alkaline phosphatase and, less frequently, of aminotransferases, bilirubin, and leukocytes. CT, which is the imaging modality of choice, classically shows multiple lucencies representing microabscesses in the liver, spleen, and kidneys. If the CT scan is nondiagnostic but clinical suspicion remains high, magnetic resonance imaging should be performed. A definitive diagnosis is usually made on a liver biopsy specimen that shows multiple granulomas and may show yeast and hyphal forms with special stains. However, biopsy often will not show evidence of infection, especially in cases in which antifungal therapy has been used. Although biopsy is the only means to establish a definitive diagnosis, it often is not required because the clinical, laboratory, and radiographic manifestations of the disease are almost always sufficient to establish a specific diagnosis.

TREATMENT AND PROGNOSIS

Rx

The mainstay of therapy is antifungal therapy. In clinically stable patients, oral fluconazole (400 mg, orally daily) can be used. In acutely ill patients, a lipid formulation of amphotericin B (3 to 5 mg/kg IV daily) is recommended. If the lipid formulation of amphotericin B is not used, other options are caspofungin (loading dose of 70 mg, then 50 mg IV daily), anidulafungin (loading dose of 200 mg, then 100 mg IV daily), and micafungin (100 mg IV daily). After 1 to 2 weeks, oral fluconazole at 400 mg daily should be started. Treatment should continue until the lesions resolve on follow-up CT scans, typically within 6 months of treatment. Patients receiving chemotherapy or stem cell transplants are at a higher risk for development of hepatosplenic candidiasis; if these treatments are indicated in patients who have a history of hepatosplenic candidiasis, fluconazole (400 mg orally, daily) should be used prophylactically to prevent relapse.

PROGNOSIS

With prolonged treatment with antifungal agents, there has been good success at treating hepatosplenic candidiasis.

OTHER FUNGAL DISEASES

Several additional fungal infections affect the liver, although rarely. *Coccidioides immitis* (Chapter 333) is often asymptomatic but may lead to fungal hepatitis characterized by increased alkaline phosphatase and the development of hepatic granulomas. Hepatic *Cryptococcus neoformans* (Chapter 336) infection is rare but has a higher prevalence in patients with AIDS, in whom it typically causes hepatomegaly. In the non–HIV-infected patient, disseminated cryptococcosis less frequently can result in focal granulomatous hepatitis, which may clinically mimic viral hepatitis, or can be manifested as obstructive jaundice secondary to sclerosing cholangitis. *Histoplasma capsulatum* (Chapter 332) infects individuals who inhale the fungus, but most cases are subclinical. In symptomatic hepatic histoplasmosis, two thirds of patients will present with hepatomegaly, with some showing splenic enlargement as well. On histologic examination, histoplasmosis can cause multiple granulomas diffusely distributed throughout the liver, although a more common finding is portal lymphohistiocytic infiltrate. *Paracoccidioides brasiliensis* (Chapter 335) most commonly infects adult men, and autopsy series have shown hepatic involvement in up to 50% of patients who die of this infection. Some individuals may present with hepatomegaly or jaundice, although jaundice is found in less than 6% of patients. Aminotransferase levels are often elevated in the early stages of the disease; changes in alkaline phosphatase or bilirubin levels tend to occur in the later stages. Biopsy may reveal lesions ranging from small granulomas to diffuse infiltration of yeast forms and fibrosis, often with bile duct involvement. In all of these examples, fungal cultures are necessary to establish a definitive diagnosis.

Parasitic, Protozoal, and Helminthic Infections of the Liver

AMEBIC LIVER ABSCESS

EPIDEMIOLOGY

Entamoeba histolytica (Chapter 352) is found throughout the world where the barriers between human feces and food and water are inadequate. After malaria, it is the second leading cause of death from parasitic diseases worldwide, accounting for an estimated 40,000 to 100,000 deaths annually. In the United States, most cases of amebiasis arise in immigrants from endemic areas and people living in states that border Mexico. Travelers to endemic areas are also at risk; ingestion of amebic cysts and colonization of the gastrointestinal tract can occur years before the development of a liver abscess. Amebic liver abscesses mainly affect men between the ages of 18 and 50 years but are also more common in postmenopausal women, thereby suggesting a hormonal protective effect. Other risk factors include alcohol intake, HLA-DR3, oral and anal sex, and contaminated enema apparatuses.[9]

PATHOBIOLOGY

E. histolytica has a simple life cycle consisting of the cyst (infectious form) and trophozoite (the motile stage associated with disease); it infects only humans and some nonhuman primates. Cysts are ingested and mature into trophozoites in the intestinal lumen. The development of amebic colitis is not essential for liver abscess formation. *E. histolytica* trophozoites penetrate through the mucosa and submucosal tissues and enter the portal circulation. *E. histolytica* blocks intrahepatic portal venules. When the trophozoites reach the liver, they create their unique abscesses, which are well-circumscribed regions of dead hepatocytes, liquefied cells, and cellular debris that are surrounded by a rim of connective tissue, a few inflammatory cells, and amebic trophozoites. The adjacent liver parenchyma is unaffected. Given the small numbers of amebae relative to the size of the abscess, it is suggested that *E. histolytica* can cause hepatocyte death without direct contact.

CLINICAL MANIFESTATIONS

Patients can present with amebic liver abscesses months to years after traveling to an endemic area, so a detailed travel history is essential for the diagnosis.[10] The disease should be suspected in patients with an appropriate travel history, fever, right upper quadrant pain, and substantial hepatic tenderness.[11] Jaundice is extremely uncommon. Symptoms are usually acute (<10 days in duration) but can be chronic, with anorexia and weight loss. Patients with acute disease tend to have multifocal disease, whereas patients with a more indolent course tend to have a solitary lesion. Laboratory data tend to demonstrate a leukocytosis without eosinophilia, mild anemia, elevated alkaline phosphatase level, and high erythrocyte sedimentation rate.

DIAGNOSIS

Although some individuals with amebic liver abscesses have concurrent amebic colitis, the majority of patients have no bowel symptoms; hence, the results of stool microscopy for *E. histolytica* trophozoites and cysts are usually negative. The diagnosis relies on identification of space-occupying lesions in the liver and a positive amebic serology. Both ultrasound and CT scan are sensitive (Fig. 151-2), but neither provides absolute specificity for amebic liver abscesses. Serologic testing of the blood is highly sensitive (>94%) and specific (>95%). False-negative test results can be obtained within the first 7 to 10 days of infection, but results on repeated testing will usually be positive. Polymerase chain reaction testing of the abscess aspirate has proved valuable for making the diagnosis in returning travelers. Aspiration of the lesion may be necessary to exclude a primary or secondary bacterial infection.

TREATMENT AND PROGNOSIS

Rx

Metronidazole (500 to 750 mg orally three times daily or a loading dose of 15 mg/kg followed by 7.5 mg/kg every 6 hours intravenously) will usually provide evidence of clinical improvement within 72 to 96 hours but should be continued for 5 to 10 days. Nitroimidazole tinidazole at 2 g daily for 5 days is also effective. Drainage is not necessary, except in patients who have abscesses larger than 10 cm or do not respond within 5 days.[12] The abscess usually will shrink by about 50% within a week, but the mean time to complete radiologic resolution is 3 to 9 months. Repeated imaging in a clinically improving patient may lead to unwarranted concern and unnecessary treatment.

Treatment must also address the removal of all of the cysts from the intestinal lumen in patients with evidence of intraluminal infection (Chapter 352). A recommended treatment is diiodohydroxyquinoline (650 mg orally three

times a day for 20 days) to prevent continued colonization and possible recurrence of the liver abscess. With prompt diagnosis and adequate medical treatment, the mortality rate from amebic abscess is 1 to 3%.

OTHER PROTOZOAN LIVER DISEASES

In addition to *Entamoeba histolytica*, other protozoan diseases that affect the liver include *Cryptosporidium*, *Toxoplasma gondii*, *Leishmania*, *Plasmodium*, and *Babesia microti* (Table 151-2).

HELMINTH INFECTIONS

Echinococcosis and Hydatid Cyst Disease

EPIDEMIOLOGY AND PATHOBIOLOGY

Human cystic echinococcosis is a zoonosis caused by the larval cestode *Echinococcus granulosus*. It is often referred to as hydatid cyst disease because of the watery cysts that characterize the infection.

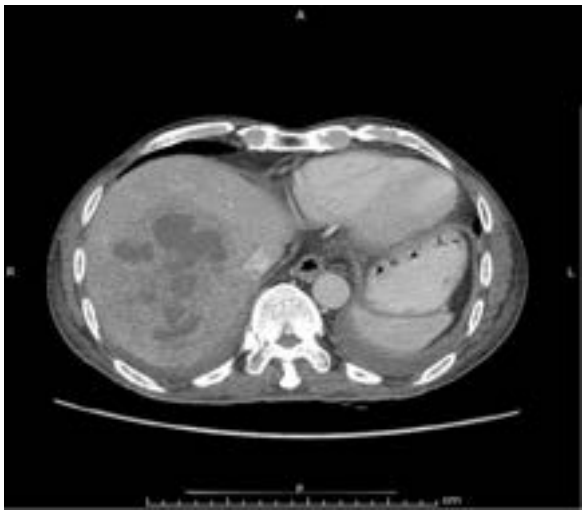

FIGURE 151-2. Computed tomographic scan of an amebic liver abscess. (Courtesy Dr. Chalermrat Bunchorntavakul, Bangkok, Thailand.)

The disease remains endemic in sheep-raising areas of the world, including Africa, the Mediterranean region of Europe, the Middle East, Asia, South America, Australia, and New Zealand. Dogs are the definitive hosts for *E. granulosus*, and sheep are the major intermediate hosts, although yaks, goats, and camels are other relevant intermediate hosts. Humans are only accidental hosts when they ingest food or water that is fecally contaminated with eggs. Human contact with sheepdogs that are in frequent contact with livestock is a major risk for infection.

The disease cycle begins when an adult tapeworm infects the intestinal tract of the definitive host (dogs usually). The adult tapeworm then produces eggs, which are expelled in the host's feces. Intermediate hosts become infected through the ingestion of parasitic eggs in fecally contaminated food. Inside the intermediate host, the eggs hatch and release tiny hooked embryos (called oncospheres), which travel in the blood stream and eventually lodge in the liver, lungs, or kidneys, where they develop into hydatid cysts. Inside these cysts grow thousands of tapeworm larvae, the next stage in the life cycle of the parasite.

CLINICAL MANIFESTATIONS

The initial infection is asymptomatic, but an enlarging hydatid liver cyst can cause abdominal pain, nausea, hepatomegaly, or a palpable mass.[13] Patients may describe symptoms of mild upper right quadrant pain, urticarial rash, and episodes of pruritus. If the cyst ruptures, serious complications can develop. Cysts that perforate into the peritoneum can lead to the development of extrahepatic cysts and may induce an allergic reaction leading to an increase of eosinophils in the blood, pruritic urticaria, and systemic anaphylaxis. In most cases, however, cysts rupture into bile ducts, which can result in cholestatic jaundice, cholangitis, or biliary pain.

DIAGNOSIS

On ultrasound or CT scans (Fig. 151-3), the hydatid cysts are often large with a flaky appearance that is referred to as hydatid sand. CT imaging also may show multiple daughter cysts or a fluid density cyst with peripheral focal areas of calcification. Fluid is of variable density, depending on the amount of proteinaceous debris.

Hydatid cysts of the liver can be diagnosed by a serologic assay, the Weinberg reaction, but it can be falsely negative in up to 38% of cases. An enzyme-linked immunosorbent assay may be more sensitive. Eosinophilia is not a feature unless the cyst ruptures; in fact, there are usually no changes in blood chemistries.

TABLE 151-2 PARASITIC INFECTIONS INVOLVING THE LIVER

	CHARACTERISTICS	ENDEMIC AREAS	RISK FACTORS	MAJOR HEPATIC MANIFESTATIONS
MAJOR PROTOZOA				
Entamoeba histolytica	Ingested cysts develop into invasive trophozoites that colonize the colon and occasionally spread to the liver by the portal blood	Mexico, regions of Central and South America, India, and regions of Africa	Male gender, alcohol intake, HLA-DR3, oral and anal sex, and contaminated enema apparatuses	Amebic liver abscesses develop as a tissue response to trophozoite invasion with acute and chronic manifestations (see text)
OTHER PROTOZOA				
Cryptosporidium sp and microsporidia	Ingested cysts develop into trophozoites in intestinal mucosa	Worldwide distribution	AIDS	Biliary tract infection with obstruction and cholangitis
Toxoplasma gondii	Ingestion of oocysts in contaminated soil or water or in infected meat; systemic spread of tachyzoites in the circulation	Worldwide distribution	Consumption of undercooked meat, contact with soil, and travel outside the United States, Europe, or Canada	Immunocompetent: asymptomatic or hepatomegaly and mild LFT elevations Immunocompromised: occasional overt hepatitis
Leishmania sp	Sand fly bite transmits promastigotes; proliferation in the reticuloendothelial system	Worldwide distribution	Children younger than 10 years and immunocompromised adults Contact with sand flies	Hepatosplenomegaly months to years after infection
Plasmodium sp	Mosquito (*Anopheles*) bite transmits sporozoites	Multiple regions throughout the world	Exposure to anopheline mosquito bites	Proliferation in hepatocytes causes hepatomegaly, enzyme elevations, and jaundice
Babesia microti	Tick bite transmits the agent, which parasitizes erythrocytes	Europe	Asplenia is a risk for fatal hepatic failure, especially bovine babesiosis	Mild liver enzyme elevations

TREATMENT AND PROGNOSIS

Rx

Drug therapy includes albendazole 400 mg twice a day for three- to six-month cycles, but treatment with albendazole alone is not effective, so drainage is essential to the effective treatment of hydatid cysts. In a randomized trial, percutaneous drainage consisting of puncture, aspiration, injection, and reaspiration of scolicidal solutions resulted in a rate of cyst disappearance similar to that of open surgical drainage, but with fewer side effects, provided patients received preprocedure and postprocedure albendazole therapy.[■] Chlorhexidine, hydrogen peroxide, 80% alcohol, and 0.5% cetrimide are preferred scolicidal agents rather than hypertonic saline or formalin. If the cyst communicates with the biliary tree, however, injection of scolicidal agents carries an almost universal risk of secondary sclerosing cholangitis, and so it is contraindicated. In such cases, the cyst must be treated surgically, by either cystectomy or hepatic resection.

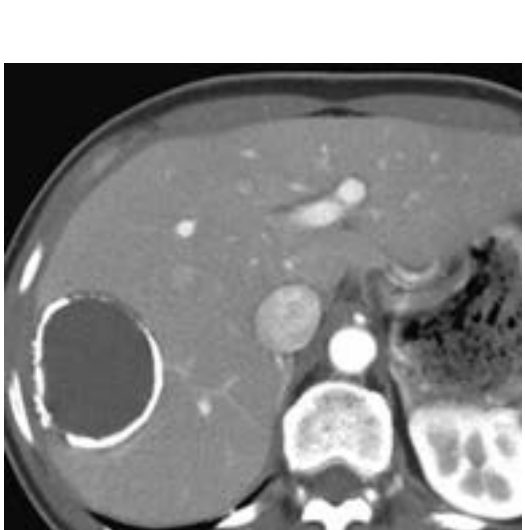

FIGURE 151-3. Computed tomographic scan of a hepatic echinococcal cyst.

The prognosis is generally good, with complete cure expected after successful percutaneous or surgical treatment. However, spillage occurs in 2 to 25% of cases, depending on the location of the cyst and the surgeon's experience; the operative mortality rate varies from 0.5 to 4% for the same reasons.

Schistosomiasis

EPIDEMIOLOGY AND PATHOBIOLOGY

Schistosomiasis is an infection of trematodes. *Schistosoma* (Chapter 355) causes periportal fibrosis and liver cirrhosis by deposition of eggs in the small portal venules. *S. mansoni* and *S. japonicum* lead to liver disease. Infection with *S. mansoni* is found in parts of South America, Africa, and the Middle East. Infection with *S. japonicum* is found in the Far East, mostly China and the Philippines. Although primary infection does not occur in the United States, 5% of the world's population (200 million people) may be infected, thereby making it a major international health concern and highly prevalent in immigrants.

Humans become infected after contact with water that contains the infective stage (cercaria) of schistosomes. After penetration of the skin, the larvae migrate to the lungs and then to the venules of the mesentery, urinary bladder, or ureters. They release eggs in the venules of the mesentery, and the eggs enter the liver through the portal vein, where they become lodged in the terminal branches of the portal venules. The lodged eggs cause a granulomatous inflammation, and the lesions heal by periportal fibrosis. *S. japonicum* is more virulent than *S. mansoni* because its infections produce 10 times more eggs.

CLINICAL MANIFESTATIONS

Initial infection is manifested as itching that is caused by skin penetration by larvae.[14] Several weeks later, patients may complain of fever, diarrhea, chills, headaches, or hives. At this time, patients will have eosinophilia. During the next 5 to 15 years, periportal liver fibrosis develops and leads to presinusoidal portal hypertension, splenomegaly (Chapter 168), and gastroesophageal varices (Chapter 138). With *S. japonicum*, however, the progression can be much more rapid, with little interval between the acute and chronic disease. Hepatic function is generally well preserved, and patients usually present with hematemesis from ruptured gastroesophageal varices.

SYMPTOMS AND SIGNS	LABORATORY FINDINGS	RADIOGRAPHIC FEATURES	DIAGNOSIS	TREATMENT
Fever, RUQ pain, and substantial hepatic tenderness	Leukocytosis without eosinophilia, mild anemia, elevated serum AP, and high ESR	US, CT, and MRI can detect abscess but cannot always differentiate amebic from pyogenic. On CT or MRI, amebic abscess sometimes appears "cold" with bright rim.	Imaging, serology, stool antigen test (microscopic evaluation of stool has a poor yield)	Metronidazole, 500-750 mg PO tid × 5-10 days, or tinidazole, 2 g daily × 3 days Iodoquinol, 650 mg tid × 20 days, also needed to eradicate intestinal colonization
See Chapter 350	See Chapter 350	See Chapter 350	See Chapter 350	See Chapter 350
See Chapter 349	See Chapter 349	See Chapter 349	See Chapter 349	See Chapter 349
See Chapter 348	See Chapter 348	See Chapter 348	See Chapter 348	See Chapter 348
See Chapter 345	See Chapter 345	See Chapter 345	See Chapter 345	See Chapter 345
See Chapter 353	See Chapter 353	See Chapter 353	See Chapter 353	See Chapter 353

TABLE 151-2 PARASITIC INFECTIONS INVOLVING THE LIVER—cont'd

	CHARACTERISTICS	ENDEMIC AREAS	RISK FACTORS	MAJOR HEPATIC MANIFESTATIONS
MAJOR HELMINTHS				
Schistosoma sp	Cercaria in fresh water penetrate the skin, travel by the circulation to portal vein radicals	*S. mansoni* found in South America, Africa, and Middle East *S. japonicum* found in Far East (mostly China and Philippines)	Contact with fresh water containing cercaria of schistosomes	Progressive presinusoidal blood flow obstruction, periportal fibrosis, portal hypertension, varices, ascites, splenomegaly
Echinococcus granulosus	Eggs of small (3-7 mm) tapeworms in stool of canid hosts; ingested eggs produce larval oncospheres that migrate to the liver and form cysts in sheep, humans, and other intermediate hosts	Worldwide distribution, found especially in sheep-raising areas (Africa, the Mediterranean region of Europe, the Middle East, Asia, South America, Australia, and New Zealand)	Ingestion of food or water fecally contaminated with eggs and human contact with sheepdogs	Initial infection asymptomatic Liver cysts increase in diameter by 1-5 cm yearly and cause variable abdominal pain, hepatomegaly, and variable eosinophilia Occasional cyst rupture, secondary bacterial infection
Echinococcus multilocularis	Eggs of small tapeworms in stool of foxes; ingested eggs produce oncospheres in the liver of rodents, humans, and other intermediate hosts	Endemic in Northern Hemisphere	Human exposure increasing with growing fox populations	Metacestodes colonize the liver as a tumor-like mass of small vesicles
Fasciola sp	Leaf-shaped flukes up to 13 × 30 mm derived from ingested cysts; the fluke excysts in the duodenum, migrates directly across the bowel wall into the peritoneal cavity, and burrows directly into the liver (or occasionally out to the skin)	Worldwide distribution	Consumption of freshwater or aquatic plants contaminated by colonized livestock	Adult flukes live in the common and hepatic bile ducts, causing obstruction that leads to thickening of the ducts, dilation, and fibrosis of the proximal biliary tree
Opisthorchis sp and *Clonorchis sinensis*	Flukes of 8-25 mm derived from ingested cysts; the fluke excysts in the duodenum and migrates into the bile ducts	*Opisthorchis* sp: Southeast Asia, central and eastern Europe (particularly Siberia) *C. sinensis*: China, Japan, Vietnam, Korea	Consumption of raw, pickled, dried, smoked, or salted freshwater fish or crayfish originating from East Asia or, in the case of *Opisthorchis felineus*, Russia and eastern Europe	Acute: typically asymptomatic Chronic: abdominal pain, fever, anorexia, tender hepatomegaly, sometimes eosinophilia Late sequelae: intermittent biliary obstruction, cholelithiasis, cholecystitis, cholangitis, secondary bacterial abscesses, cholangiocarcinoma
Toxocara sp	Nematode infection disseminates to cause visceral larva migrans after ingestion of soil contaminated with dog or cat feces	Highest prevalence in southeastern United States	Consumption of food contaminated with soil containing eggs; distributed throughout the United States	Often an asymptomatic cause of eosinophilia (exclude *Trichinella*, *Strongyloides*, filaria, hookworm, schistosomiasis) Hepatomegaly is common, but nonhepatic manifestations dominate the clinical picture
OTHER HELMINTHS				
Ascaris lumbricoides	Ingested eggs develop into larvae that migrate to the lungs and are coughed and swallowed; develop into roundworms 15-30 mm long in the small intestine	Global distribution with higher prevalence in Africa, South America, India, and the Far East; 20% of the world's population is colonized	Consumption of fecally contaminated food or water, particularly in young children	Colonization is typically asymptomatic with eosinophilia Biliary migration of worms can cause symptomatic biliary obstruction, cholangitis, cholecystitis, and secondary bacterial liver abscess
Capillaria hepatica	Ingested eggs develop into larvae in the intestinal mucosa; larvae migrate to the liver by portal blood flow and develop into short-lived roundworms	Human infection is rare	Consumption of food contaminated with rodent feces	Fever, eosinophilia, and hepatomegaly; subsequent foci of liver fibrosis, granulomas, and calcification in involved areas
Strongyloides stercoralis	Ingested eggs develop into 1.5- to 2.5-mm nematodes that invade the hepatic vasculature, lymphatics, and biliary tract	Tropical and subtropical areas including southeast United States and southern and eastern Europe	Consumption of food contaminated with soil containing eggs, infection with HTLV-1, and immunocompromised individuals	Hepatic disease in the setting of immunosuppression: jaundice, abdominal pain; eosinophilia is uncommon

RUQ = right upper quadrant; AP = alkaline phosphatase; ESR = erythrocyte sedimentation rate; US = ultrasonography; CT = computed tomography; MRI = magnetic resonance imaging; ELISA = enzyme-linked immunosorbent assay; ERCP = endoscopic retrograde cholangiopancreatography; HTLV = human T-cell lymphotrophic virus.
Modified from Neuschwander-Tetri BA. Bacterial, parasitic, fungal, and granulomatous liver disease. In: Goldman L, Ausiello D, eds. *Cecil Textbook of Medicine.* 23rd ed. Philadelphia: WB Saunders; 2007.

SYMPTOMS AND SIGNS	LABORATORY FINDINGS	RADIOGRAPHIC FEATURES	DIAGNOSIS	TREATMENT
Initial infection presents as itching; later presentations include fever, diarrhea, chills, headaches, or hives Hematemesis from ruptured gastroesophageal varices	Eosinophilia and splenomegaly Seroconversion occurs within 4-6 weeks	Extensive calcification with typical "turtle-back" appearance along portal tracts	Rectal biopsy, liver biopsy, microscopic examination of stool	Praziquantel (single dose 40 mg/kg) with a second dose 6-12 weeks later if necessary
Enlarged hydatid cyst can cause abdominal pain, nausea, hepatomegaly, or a palpable mass Mild RUQ pain, urticaria, and episodes of pruritus	Positive Weinberg reaction (false-negative reaction in 38% of cases) or ELISA analysis Eosinophilia seen in ruptured cysts	On US or CT, cysts appear flaky, sometimes showing daughter cysts or peripheral focal calcification; fluid is of variable density	Imaging, Weinberg reaction, or ELISA	Albendazole (400 mg) 2× daily for 3-6 monthly cycles with 10- to 14-day intervals Drainage of the cyst is essential Surgical resection if the cyst communicates with the biliary tree
See Chapter 354	See Chapter 354	See Chapter 354	See Chapter 354	See Chapter 354
Acute: fever, abdominal pain, eosinophilia Chronic: symptomatic biliary obstruction, variable eosinophilia	Serologic tests include hemagglutination, complement fixation, ELISA, and counterimmunoelectrophoresis Anemia, leukocytosis, eosinophilia, elevated AP, and hypergammaglobulinemia are often seen	CT is most useful: liver shows hypodense nodules or tortuous tracks; thickening of the liver capsule, subcapsular hematoma, and parenchymal calcification may also be seen	Serology, stool examination	Triclabendazole, 10 mg/kg once or twice
Chronic infection presents as dyspepsia, abdominal pain, diarrhea, nausea, vomiting, anorexia, weight loss, fevers, hepatomegaly, and urticaria Acute presentation includes serum sickness, intrahepatic pigment stones, and facial edema Other rare complications are cholangitis, pancreatitis, and obstructive jaundice	Anemia, leukocytosis, eosinophilia, elevated AP, and hypergammaglobulinemia	US may detect flukes in biliary tree CT shows small hypodense nodules	Serology and stool examination	Praziquantel, 25 mg/kg q8h × 3 doses
See Chapter 358	See Chapter 358	See Chapter 358	See Chapter 358	See Chapter 358
Sensitized patients during pulmonary phase present with asthma-like symptoms, hemoptysis, chest pain, and cyanosis Urticaria and other allergic reactions sometimes seen Patients in intestinal phase show cognitive and nutritional impairment with abdominal pain, hepatomegaly, cholangitis, and obstructive jaundice	Leukocytosis with eosinophilia Hyperbilirubinemia occasionally seen	Movement of the worms within the biliary tree can sometimes be observed A "bull's-eye" appearance can be seen on cross-sectional imaging	Stool examination, imaging, and ERCP	Albendazole, 400 mg once
Persistent fever, eosinophilia, and hepatomegaly most common Splenomegaly, anorexia, nausea, vomiting, night sweats, and altered bowel habits also seen	Anemia, eosinophilia, moderately elevated liver enzymes, increased ESR, and hypergammaglobulinemia	US shows nonspecific hyperechoic areas in the portal spaces	Stool and liver biopsy	Mebendazole, 200 mg bid × 20 days
Recurrent urticaria, abdominal pain, diarrhea, and cough; mild jaundice and hepatomegaly in the absence of splenomegaly	Eosinophilia and hypoalbuminemia Patients may have elevated liver enzymes	Imaging studies not used in diagnosis	Serology and stool examination	Ivermectin, 200 µg/kg/day × 2 days, or albendazole, 400 mg/day × 7 days

DIAGNOSIS

Schistosomal eggs typically have lateral or terminal spines and are easy to detect on microscopic examination of feces or on a rectal biopsy specimen. Seroconversion occurs within 4 to 8 weeks of infection but cannot distinguish active infection from a history of exposure. Newer, more sensitive and specific molecular and immunodiagnostic techniques (including polymerase chain reaction–enzyme-linked immunosorbent assay systems) for detection of schistosome DNA in feces or serum and plasma have the potential to diagnose schistosomiasis in all stages of clinical disease, not only when the adult worms are producing eggs.[15] Imaging may show extensive calcification with the typical "turtle-back" appearance along the portal tracts reflecting the clustered, calcified eggs along the portal triads. Liver biopsy may demonstrate an egg of *S. mansoni* along with intense granulomatous change in the portal tract (Fig. 151-4).

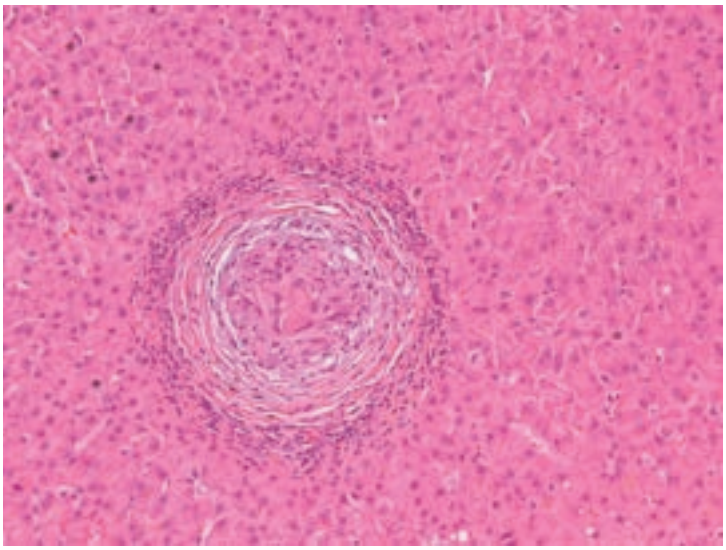

FIGURE 151-4. Liver biopsy specimen demonstrating an egg of *Schistosoma mansoni* and intense granulomatous change in the portal tract. (Courtesy Alberto Q. Farias, MD, São Paulo, Brazil.)

TREATMENT AND PROGNOSIS

Praziquantel (a single or two divided doses of 40 mg/kg for *S. mansoni* infections, and two divided doses of 60 mg/kg for *S. japonicum*), which is the preferred treatment, is effective in 70 to 100% of cases, but a second dose can be given 6 to 12 weeks later, particularly in patients with eosinophilia, high antibody titers, or persistent symptoms. Treatment of portal hypertension (Chapter 153) may be necessary. The mortality is 0.05% with heavy *S. mansoni* infection and 1.8% for severe *S. japonicum* infection. Bleeding from esophageal varices is the most serious complication. Chronic infection can lead to hepatocellular carcinoma.

GRANULOMATOUS DISEASES OF THE LIVER

Granulomas, which are found in up to 15% of liver biopsy specimens, can be an incidental finding or may represent a wide array of liver diseases (Table 151-3). A granuloma is an accumulation of epithelioid cells, including transformed macrophages, mononuclear cells, and other inflammatory cells. Granulomas may have associated necrosis (caseating), as in tuberculosis (Chapter 324), or may be non-necrotizing (noncaseating), as in sarcoidosis (Chapter 95). In addition, fibrin-ring granulomas are characterized by a vacuole encircled by a ring of fibrinoid necrosis and surrounded by lymphocytes and histiocytes, as seen in Q fever (Chapter 327).

An incidental granuloma requires minimal further evaluation. Tuberculosis should be excluded (Chapter 324). Sarcoidosis should also be considered as it is the most common cause of granulomas in the United States. An antimitochondrial antibody should be ordered to rule out primary biliary cirrhosis. The use of potentially causative drugs should be stopped (see Table 151-3).

Sarcoidosis

Sarcoidosis (Chapter 95), which is the most frequently identified cause of hepatic granuloma in the United States, affects all racial and ethnic groups and occurs at all ages. In patients with sarcoidosis, the liver is the third most commonly involved organ, after the lymph nodes and lungs; about 50 to 80% of patients with sarcoidosis will have granulomas in their livers. The classic granuloma is found mainly in the portal triads, with a cluster of large epithelioid cells and often with multinucleated giant cells.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Hepatic involvement of sarcoidosis is often subclinical, and only a minority of patients will present with pruritus, fever, abdominal pain, hepatomegaly,

TABLE 151-3 CAUSES OF GRANULOMATOUS LIVER DISEASE

DIAGNOSIS	SPECIAL AND UNIQUE FEATURES
Sarcoidosis	Evidence of pulmonary sarcoidosis, granulomas found on biopsy of other organs, elevated ACE level
Bacterial infections	
Mycobacterium tuberculosis	Caseating granulomas on biopsy, positive PPD response or interferon-gamma release assay, active pulmonary tuberculosis
Other mycobacteria (*M. avium-intracellulare, M. leprae, M. mucogenicum, M. bovis*)	HIV, exposure history
Other bacteria (brucellosis, listeriosis, melioidosis, tularemia, yersiniosis, bartonellosis, Q fever, syphilis, psittacosis)	Fever, exposure history
Viral infections (cytomegalovirus, Epstein-Barr virus, hepatitis C, hepatitis B, hepatitis A)	Serology for acute or recent exposure
Fungal infections (histoplasmosis, coccidioidomycosis, blastomycosis, nocardiosis, candidiasis)	Fever, immunocompromised
Parasitic infections (schistosomiasis, *Ascaris lumbricoides*, toxoplasmosis, visceral leishmaniasis)	Travel to endemic regions, positive serologic testing
Primary biliary cirrhosis	Female sex, positive AMA, elevated IgM
Malignant diseases (Hodgkin disease, non-Hodgkin lymphoma, renal cell carcinoma)	Evidence of malignant disease in kidney or bone marrow
Drug reactions (allopurinol, chlorpropamide, phenylbutazone, sulfonamides, carbamazepine, glyburide, quinidine, quinine, diltiazem, hydralazine, rosiglitazone, phenytoin, methyldopa, procainamide, amoxicillin–clavulanic acid, mebendazole, mesalamine, acetaminophen, pyrazinamide, halothane, isoniazid, norfloxacin)	Exposure history
Toxins (beryllium, copper sulfate, Thorotrast)	Previous exposure history
Miscellaneous (talc, Crohn disease, granulomatosis with polyangiitis, post–jejunoileal bypass, mineral oil lipogranulomas, hepatic allograft rejection, chronic granulomatous disease)	History of IV drug use, history of liver transplantation, diarrhea

ACE = angiotensin-converting enzyme; PPD = purified protein derivative; HIV = human immunodeficiency virus; AMA = antimitochondrial antibody; IV = intravenous.

cholestatic jaundice, or portal hypertension. In some patients, the granulomatous injury and destruction of the interlobular bile ducts eventually cause ductopenia and a histologic picture similar to primary biliary cirrhosis (Chapter 155). In other patients, damage to the large bile ducts may lead to a syndrome that mimics primary sclerosing cholangitis (Chapter 155). Other patients may present with focal liver lesions suggestive of malignant neoplasm.

Patients typically have markedly elevated serum alkaline phosphatase levels. Angiotensin-converting enzyme levels also are characteristically elevated but may not be helpful in differentiating sarcoidosis from other chronic liver diseases, such as primary biliary cirrhosis, in which it also may be elevated. The presence of noncaseating granulomas in a patient with clinically suspected sarcoidosis generally establishes the diagnosis. However, granulomas do not correlate with liver function test results or duration of disease.

TREATMENT AND PROGNOSIS

Rx

In general, hepatic sarcoidosis does not need to be treated except in patients with other indications for treatment (Chapter 95). Corticosteroids improve liver function test results but do not alleviate portal hypertension, and serial biopsies often show little improvement. Sarcoidosis that leads to portal hypertension and fibrosis may ultimately require liver transplantation (Chapter 154), although sarcoidosis may recur in the new organ.

Other Granulomatous Liver Diseases

In addition to sarcoidosis, another major cause of granulomatous liver disease is tuberculosis (Chapter 324). Military tuberculosis, caused by *Mycobacterium tuberculosis*, commonly results in hepatic granulomas, and patients present with hepatomegaly, abnormal liver function test results, or both. A biopsy revealing caseating granulomas along with a positive PPD (purified protein derivative) response and active pulmonary tuberculosis can help in establishing a diagnosis. Other mycobacteria are also known to cause granulomatous liver disease, including *M. avium-intracellulare*, *M. genavense*, and *M. scrofulaceum* (Chapter 325).

Other causes of granulomatous liver disease include zoonotic infections such as cat-scratch disease (Chapter 315), Q fever (Chapter 327), and brucellosis (Chapter 310). Cat-scratch disease, which primarily affects children, is caused by *Bartonella henselae*, with cats serving as the main reservoir for the organism. Patients typically present with lymphadenopathy associated with persistent pyrexia of unknown origin, abdominal pain, and weight loss. *B. henselae* can be identified by a Warthin-Starry stain, or an imaging study may demonstrate scattered defects in the liver. Q fever is caused by the intracellular gram-negative rickettsial organism *Coxiella burnetii*. Most infections are asymptomatic but may have self-limited influenza-like symptoms, pneumonia, and hepatitis. A liver biopsy specimen may demonstrate fibrin-ring granulomas (doughnut shaped) in the background of nonspecific reactive hepatitis and steatosis. Brucellosis, which is not common in the United States, is caused by at least four species of *Brucella*: *B. abortus*, *B. suis*, *B. melitensis*, and *B. canis*. The disease is manifested as recurrent high fevers, drenching sweats, malaise, arthralgia, fatigue, abdominal pain, anorexia, and headaches. Patients often have hepatomegaly and elevated serum levels of aminotransferases and alkaline phosphatase. The granulomas formed by this infection are typically smaller than those caused by sarcoidosis or tuberculosis, and a definitive diagnosis can be established through serologic testing.

It is important to keep in mind that hepatic granulomas are formed naturally as a consequence of the immune response, so they may be seen in infections with hepatic involvement. Many of the diseases noted can be manifested as granulomatous liver disease (see Table 151-3): listeriosis, yersiniosis, candidiasis, histoplasmosis, coccidioidomycosis, and schistosomiasis. Viral infections (e.g., cytomegalovirus, hepatitis A, hepatitis B, and hepatitis C), primary biliary cirrhosis, malignant diseases, and certain drug reactions have been etiologically linked to granulomatous liver disease. Table 151-3 categorically lists the major causes of hepatic granulomas and the means by which a differential diagnosis can be established.

Grade A References

A1. Singh O, Gupta S, Moses S, et al. Comparative study of catheter drainage and needle aspiration in management of large liver abscesses. *Indian J Gastroenterol*. 2009;28:88-92.

A2. Yu SC, Ho SS, Lau WY, et al. Treatment of pyogenic liver abscess: prospective randomized comparison of catheter drainage and needle aspiration. *Hepatology*. 2004;39:932-938.

A3. Chavez-Tapia NC, Hernandez-Calleros J, Tellez-Avila FI, et al. Image-guided percutaneous procedure plus metronidazole versus metronidazole alone for uncomplicated amoebic liver abscess. *Cochrane Database Syst Rev*. 2009;1:CD004886.

A4. Nasseri-Moghaddam S, Abrishami A, Taefi A, et al. Percutaneous needle aspiration, injection, and re-aspiration with or without benzimidazole coverage for uncomplicated hepatic hydatid cysts. *Cochrane Database Syst Rev*. 2011;1:CD003623.

A5. Danso-Appiah A, Olliaro PL, Donegan S, et al. Drugs for treating *Schistosoma mansoni* infection. *Cochrane Database Syst Rev*. 2013;2:CD000528.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Meddings L, Myers RP, Hubbard J, et al. A population-based study of pyogenic liver abscesses in the United States: incidence, mortality, and temporal trends. *Am J Gastroenterol*. 2010;105:117-124.
2. Qu K, Liu C, Wang ZX, et al. Pyogenic liver abscesses associated with nonmetastatic colorectal cancers: an increasing problem in Eastern Asia. *World J Gastroenterol*. 2012;18:2948-2955.
3. Jeong SW, Jang JY, Lee TH, et al. Cryptogenic pyogenic liver abscess as the herald of colon cancer. *J Gastroenterol Hepatol*. 2012;27:248-255.
4. Lo JZ, Leow JJ, Ng PL, et al. Predictors of therapy failure in a series of 741 adult pyogenic liver abscesses. *J Hepatobiliary Pancreat Sci*. 2015;22:156-165.
5. Yoon JH, Kim YJ, Jun YH, et al. Liver abscess due to *Klebsiella pneumoniae*: risk factors for metastatic infection. *Scand J Infect Dis*. 2014;46:21-26.
6. Ali AH, Smalligan RD, Ahmed M, et al. Pyogenic liver abscess and the emergence of *Klebsiella* as an etiology: a retrospective study. *Int J Gen Med*. 2013;7:37-42.
7. O'Farrell N, Collins CG, McEntee GP. Pyogenic liver abscesses: diminished role for operative treatment. *Surgeon*. 2010;8:192-196.
8. Chen YC, Lin CH, Chang SN, et al. Epidemiology and clinical outcome of pyogenic liver abscess: an analysis from the National Health Insurance Research Database of Taiwan, 2000-2011. *J Microbiol Immunol Infect*. 2014;[Epub ahead of print].
9. Choudhuri G, Rangan M. Amebic infection in humans. *Indian J Gastroenterol*. 2012;31:153-162.
10. Wuerz T, Kane JB, Boggild AK, et al. A review of amoebic liver abscess for clinicians in a nonendemic setting. *Can J Gastroenterol*. 2012;26:729-733.
11. Hoque MI, Uddin MS, Sarker AR, et al. Common presentation of amebic liver abscess—a study in a tertiary care hospital in Bangladesh. *Mymensingh Med J*. 2014;23:724-729.
12. Bammigatti C, Ramasubramanian NS, Kadhiravan T, et al. Percutaneous needle aspiration in uncomplicated amebic liver abscess: a randomized trial. *Trop Doct*. 2013;43:19-22.
13. McManus DP, Gray DJ, Zhang W, et al. Diagnosis, treatment, and management of echinococcosis. *BMJ*. 2012;344:e3866.
14. Gryseels B, Polman K, Clerinx J, et al. Human schistosomiasis. *Lancet*. 2006;368:1106-1118.
15. Cavalcanti MG, Silva LF, Peralta RH, et al. Schistosomiasis in areas of low endemicity: a new era in diagnosis. *Trends Parasitol*. 2013;29:75-82.

152 ALCOHOLIC AND NONALCOHOLIC STEATOHEPATITIS

NAGA P. CHALASANI

Alcoholic liver disease and nonalcoholic fatty liver disease (NAFLD), which represent two of the most common forms of liver disease, can lead to cirrhosis, liver failure, and death. Although these two conditions have different risk factors and natural histories, in both conditions (Table 152-1) the hepatocytes are characterized by macrovesicular steatosis, which is the accumulation of triglycerides as one large cytoplasmic globule that displaces the nucleus. In microvesicular steatosis, cytoplasmic accumulation of fat occurs as multiple small globules with a central nucleus.

ALCOHOLIC LIVER DISEASE

DEFINITION

Excessive alcohol consumption (Chapter 33) causes alcoholic liver disease and can significantly worsen other liver disorders, such as viral hepatitis (Chapter 149) and hemochromatosis (Chapter 212). Although most individuals who consume alcohol do not consume it excessively and do not develop any physical or social consequences, some alcoholics consume sufficient alcohol and, presumably because of other predisposing factors, develop alcoholic liver disease. Alcoholic liver disease is a spectrum of chronic liver diseases ranging from alcoholic fatty liver to alcoholic hepatitis and cirrhosis.

Alcoholic fatty liver disease will develop in nearly 90% of individuals who consume alcohol heavily (on average, >6 drinks per day), but only some individuals develop the more severe conditions of alcoholic hepatitis and alcoholic cirrhosis. Genetic predisposition is likely to play a role in the pathogenesis of acute alcoholic hepatitis and alcoholic cirrhosis, but these genetic factors have not been well defined.[1] Nearly 50% of the patients with alcoholic

TABLE 152-1 COMMON CAUSES OF MACROVESICULAR AND MICROVESICULAR STEATOSIS

MACROVESICULAR STEATOSIS	MICROVESICULAR STEATOSIS
Obesity, type 2 diabetes, metabolic syndrome, and dyslipidemia (nonalcoholic fatty liver disease)	Reye syndrome
Excessive alcohol consumption	Medications (valproate, antiretroviral medicines, intravenous tetracycline)
Hepatitis C (genotype 3)	Heat stroke
Wilson disease	Acute fatty liver of pregnancy
Lipodystrophy starvation	HELLP syndrome
Jejunoileal bypass	Inborn errors of metabolism (lecithin–cholesterol acyltransferase deficiency, cholesterol ester storage disease, Wolman disease)
Parenteral nutrition	
Medications (amiodarone, methotrexate, tamoxifen, corticosteroids, antipsychotics)	

HELLP = *h*emolysis, *e*levated *l*iver enzymes, and *l*ow *p*latelets.

hepatitis have preexisting cirrhosis (Chapter 153), and individuals who do not yet have cirrhosis are at high risk for its development, especially if they continue to consume alcohol.

EPIDEMIOLOGY

The true prevalence of alcoholic liver disease is not known, but nearly 1% of North American adults are believed to have alcoholic liver disease. Even this figure is considered an underestimation because milder forms of alcoholic liver disease are asymptomatic and often unrecognized. It has been estimated that alcoholic liver disease accounts for 40% of deaths from cirrhosis and 28% of all deaths from liver disease. It is the second most common indication for liver transplantation in the United States once abstinence from alcohol has been established.

PATHOBIOLOGY

The mechanisms underlying alcoholic liver injury can be broadly categorized into those caused by the effects of alcohol directly on hepatocytes and those caused by the effects mediated by Kupffer cells.[2] The hepatocyte mechanisms include the altered redox state induced by alcohol and aldehyde dehydrogenase reactions; the oxidative stress and lipid peroxidation caused by the induction of CYP2E1 enzymes and the mitochondrial electron transfer system; and the effects of alcohol on the nuclear transcription factors AMP kinase and SREBP-1c, protein adduct formation, and altered methionine and folate metabolism with resulting endoplasmic reticulum stress. Chronic alcohol consumption increases gut permeability, and the resulting portal endotoxemia activates Kupffer cells. Activated Kupffer cells release a number of proinflammatory mediators. These include tumor necrosis factor-α (TNF-α); transforming growth factor-β1; interleukins 1, 6, 8, and 10; and platelet-derived growth factor. TNF-α has plethora of biologic effects and causes hepatocyte apoptosis, whereas transforming growth factor-β1 and platelet-derived growth factor play important roles in stellate cell activation, collagen production, and hepatic fibrosis.

Among the known risk factors for development of alcoholic liver disease (Table 152-2), the amount of alcohol consumed is the single most important. For unclear reasons, only 30 to 35% of individuals with heavy and long-term drinking develop alcoholic hepatitis, and less than 20% develop cirrhosis. Women are at higher risk; for example, the risk of alcoholic cirrhosis increases after 10 years of alcohol consumption at quantities of more than 60 to 80 g/day in men, whereas in women, it can develop at quantities of only more than 20 g/day. Moreover, the peak incidence of alcoholic liver disease in women is approximately a decade earlier than in men. The type of alcoholic beverage consumed may not be as critical, but "spirits" and beer may be more hepatotoxic than wine. African American and Hispanic ethnic groups may be predisposed to more significant alcoholic liver injury. Both obesity and protein-calorie malnutrition, in which micronutrients and antioxidant capacity are diminished, also are important predispositions.

Polymorphisms in genes associated with alcohol metabolism (alcohol and aldehyde dehydrogenases and cytochrome P-450 enzymes) and dysregulated cytokine production (e.g., TNF-α) may also influence genetic susceptibility. In patients with other forms of chronic liver disease (e.g., viral hepatitis B or C), concomitant alcohol consumption significantly aggravates liver injury.

CLINICAL MANIFESTATIONS

Patients with alcoholic liver disease may have signs and symptoms from underlying alcoholism as well as those caused by liver disease. Stigmata of chronic alcoholism include palmar erythema (see Fig. 146-2), spider nevi, bilateral gynecomastia, testicular atrophy, bilateral parotid enlargement, and Dupuytren contractures (Fig. 152-1). The clinical features of liver disease will depend on the stage of alcoholic liver disease, that is, whether a patient has alcoholic fatty liver or more advanced liver disease, such as alcoholic hepatitis and cirrhosis.

Patients with alcoholic fatty liver disease are generally asymptomatic, but some patients may have anorexia, fatigue, right upper quadrant discomfort, and tender hepatomegaly. These patients may also have biochemical evidence of alcoholism and alcoholic liver disease with macrocytosis as well as elevated levels of aspartate aminotransferase (AST) and γ-glutamyl transpeptidase. Patients with alcoholic fatty liver typically do not have jaundice, ascites, or splenomegaly.

Patients with alcoholic hepatitis may have a more dramatic presentation with severe malaise, fatigue, anorexia, fever, evidence of protein-calorie malnutrition, and features of decompensated liver disease, including jaundice, coagulopathy, ascites, and encephalopathy. However, these classic features of acute alcoholic hepatitis are not universally present. Physical examination invariably shows at least some features of chronic alcoholism, and jaundice (see Fig. 146-1), ascites (see Fig. 146-4), and splenomegaly are common. The laboratory examination findings are typically abnormal. Common hematologic abnormalities include leukocytosis with neutrophil predominance, macrocytic anemia (Chapter 164), thrombocytopenia (Chapter 172), and prolonged prothrombin time. Liver biochemistries (Chapter 147) are abnormal with an elevated AST and ratio of AST to alanine aminotransferase (ALT), alkaline phosphatase, γ-glutamyl transpeptidase, and total bilirubin but decreased levels of serum albumin. The AST rarely exceeds 300 IU/L. Serum electrolyte abnormalities including hypokalemia (Chapter 117), hypomagnesemia (Chapter 119), hypocalcemia (Chapter 245), and hypophosphatemia (Chapter 119) are frequent. Patients with alcoholic cirrhosis have the same clinical features that are common to other types of cirrhosis (Chapter 153) but also with striking features of underlying chronic alcoholism.

DIAGNOSIS

The diagnosis of alcoholic liver disease strongly depends on the history of excessive alcohol consumption and the presence of liver disease. Although laboratory abnormalities are not specific for alcoholic liver disease, they can be suggestive in the context of excessive alcohol consumption. An AST/ALT ratio of more than 2 is typical in alcoholic liver disease, and ALT values greater than 150 to 200 IU/L are very rare in alcoholic liver disease. Serology testing for coexisting chronic viral hepatitis (Chapter 149) is critical. Diagnostic dilemmas arise when a patient denies excessive alcohol consumption in the face of clinical features that are suggestive of alcoholic liver disease. Interviewing family members about specific alcohol consumption may be helpful in the accurate ascertainment of alcohol consumption. Elevated blood levels of carbohydrate-deficient transferrin, which is a form of transferrin with fewer than the four sialic acid chains present in normal transferrin, can identify recent heavy alcohol consumption. Hepatic imaging by ultrasound, computed tomography, or magnetic resonance imaging will show changes consistent with hepatic steatosis or more advanced forms of liver disease, such as cirrhosis and portal hypertension. Imaging is also important to exclude other forms of liver disease, including malignant disease and

TABLE 152-2 RISK FACTORS FOR ALCOHOLIC LIVER DISEASE AND NONALCOHOLIC FATTY LIVER DISEASE (NAFLD)

ALCOHOLIC LIVER DISEASE	NAFLD
MAJOR	**MAJOR**
Amount and duration of alcohol consumption	Obesity
Female gender	Type 2 diabetes
Genetic factors	Dyslipidemia
Protein-calorie malnutrition	Metabolic syndrome
MINOR	**MINOR**
Type of beverage	Polycystic ovary syndrome
Binge drinking	Hypothyroidism
Obesity	Obstructive sleep apnea
African American and Hispanic ethnicity	Hypopituitarism
	Hypogonadism

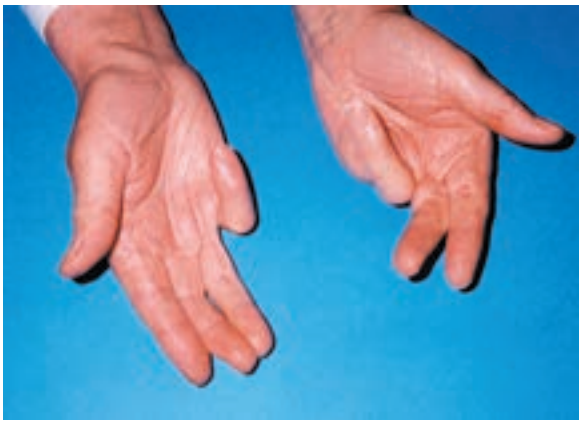

FIGURE 152-1. Dupuytren contracture. (From Gudmundsson KG, Jonsson T, Arngrimsson R. Guillaume Dupuytren and finger contractures. *Lancet*. 2003;362:165-168.)

biliary obstruction. Imaging findings specific for alcoholic liver disease include an enlarged caudate lobe, greater visualization of the right posterior hepatic notch, and focal fat sparing or geographic fat distribution.

Because specific treatment for alcoholic hepatitis may be harmful in patients with other liver diseases, it is important to exclude other predominant or coexisting liver diseases, including chronic viral hepatitis (Chapter 149) and drug-induced liver injury, especially from acetaminophen (Chapter 150), by history, blood tests, and biopsy if needed. Hyperferritinemia generally reflects an acute phase reactant, rather than an iron overload disorder, so it usually will return to normal when the acute liver injury resolves.

Liver biopsy is the key to precisely characterizing the nature of alcoholic liver disease and determining whether a patient has fatty liver or more advanced alcoholic hepatitis. Histologic features of alcoholic fatty liver include macrovesicular steatosis that is predominantly centrilobular (zone 3) in nature. In alcoholic hepatitis, the biopsy is more striking and reveals macrovesicular steatosis, lobular neutrophilic infiltration, Mallory hyaline, balloon degeneration of the hepatocytes, and perivenular fibrosis. In general, patients with alcoholic hepatitis also have histologic evidence of chronic liver injury in the form of more advanced fibrosis (periportal or bridging fibrosis or cirrhosis).

TREATMENT

Rx

Total abstinence, which is the most important treatment measure, is mandatory for the improvement of the clinical and histologic features of alcoholic liver disease. Its benefits are unequivocal, even in patients with severe decompensation. However, long-term abstinence is difficult to achieve, so a multidisciplinary approach with counseling and medications that promote abstinence should be considered. Disulfiram is not commonly used because of its poor tolerability and hepatotoxicity. Opioid antagonists, such as naltrexone (50 mg/day for up to 6 months or even longer), nalmefene (20 mg/day as maintenance), and acamprosate (333-mg tablets, 2 tablets three times each day for 1 year), can help promote abstinence when they are used as part of a multidisciplinary approach (Chapter 33).

Alcoholic fatty liver disease requires no specific treatment other than abstinence. Patients with alcoholic hepatitis, however, have increased short- and long-term mortality and should be considered for therapeutic interventions in addition to mandatory abstinence.[3] If a patient's liver biopsy findings are consistent with alcoholic hepatitis and there is no evidence of other inflammatory liver diseases, such as hepatitis C (Chapter 149), prednisolone (40 mg/day for 4 weeks) should be given to carefully selected patients with severe alcoholic hepatitis who have a score higher than 32 on Maddrey's discriminant function (4.6 × [patient's prothrombin time − control prothrombin time] + total bilirubin level) and encephalopathy but do not have gastrointestinal bleeding or systemic infection. Studies have suggested that a Model for End-Stage Liver Disease (MELD) score (see Table 153-2) higher than 21 can substitute for Maddrey's score to guide the use of prednisolone. Adding intravenous *N*-acetylcysteine (on day 1 at a dose of 150, 50, and 100 mg per kilogram of body weight in 250, 500, and 1000 mL of 5% glucose solution during a period of 30 minutes, 4 hours, and 16 hours, respectively; and on days 2 through 5, 100 mg per kilogram per day in 1000 mL of 5% glucose solution) to prednisolone appears to be superior to prednisolone alone in terms of a significantly lower mortality at 1 month (8% vs. 24%) and a somewhat but not significantly lower mortality at 6 months (22% vs. 34%; $P = .06$). Although randomized trials also have shown a benefit of pentoxifylline (400 mg three times daily for 28 days) for severe alcoholic hepatitis, prednisolone is more effective than pentoxifylline and the combination of pentoxifylline with prednisolone is no better than prednisolone alone.

The anti–TNF-α agents infliximab and etanercept are not effective and have significant side effects. All patients with alcoholic hepatitis and alcoholic cirrhosis should be assessed and treated for protein-calorie malnutrition and micronutrient deficiency. Hospitalized patients with severe decompensation should be considered for enteral nutrition (Chapter 216).

Complications such as ascites, spontaneous bacterial peritonitis, encephalopathy, variceal bleeding, hepatorenal syndrome, osteoporosis, and hepatopulmonary syndrome may occur in patients with decompensated alcoholic cirrhosis and must be managed carefully (Chapter 153). Liver transplantation is a reasonable option in patients with decompensated alcoholic cirrhosis, and observational data suggest that early liver transplantation can improve survival in patients with severe, medically refractory alcoholic hepatitis.[4] In general, however, 6 months of abstinence and strong social support are required for eligibility (Chapter 154). Patients with alcoholic cirrhosis are at risk for hepatocellular carcinoma (Chapter 196) and should be screened with semiannual liver imaging and serum α-fetoprotein levels. They are also at risk for extrahepatic malignant disease, notably head and neck, lung, and esophageal cancer. For Dupuytren contracture, injection of collagenase clostridium histolyticum can reduce the severity and improve the range of motion significantly.

PROGNOSIS

Alcoholic fatty liver is generally reversible with total abstinence for a few months. Alcoholic hepatitis carries high mortality, with nearly 40% of patients dying within 6 months after its presentation. Predictors of poor outcome include the severity of disease as quantified by approaches such as the MELD score[5] (Chapters 153 and 154) and the severity of fibrosis and the neutrophilic infiltration on liver biopsy.[6]

NONALCOHOLIC FATTY LIVER DISEASE

DEFINITIONS

On histologic examination, NAFLD resembles alcoholic liver disease, but it occurs in individuals without significant alcohol consumption. Average alcohol consumption of more than two drinks per day in men and more than one drink per day in women generally is not consistent with a diagnosis of NAFLD. In addition, the definition of NAFLD excludes patients with a history of exposure to steatogenic medications, such as amiodarone, methotrexate, and tamoxifen.

NAFLD encompasses a spectrum of abnormal liver histologic features, ranging from simple steatosis to nonalcoholic steatohepatitis (NASH) and cirrhosis. In simple steatosis, liver histology reveals macrovesicular steatosis without ballooning degeneration of hepatocytes or liver fibrosis. NASH, which is a more advanced form of NAFLD, is histologically characterized by macrovesicular steatosis, ballooning degeneration of the hepatocytes, and sinusoidal fibrosis.

EPIDEMIOLOGY

NAFLD is one of the most common causes of elevated liver enzymes and chronic liver disease in the Western world. Its incidence in adults and children is rising rapidly because of the ongoing epidemics of obesity (Chapter 220), type 2 diabetes mellitus (Chapter 229), and metabolic syndrome. Its prevalence is high in certain populations of patients; for example, nearly 80% of type 2 diabetic patients and 90% of morbidly obese individuals have imaging evidence of NAFLD. Nearly one third of adults in westernized countries are estimated to have NAFLD, and about 5% of adults may have NASH.[7] These percentages compare reasonably well with other data suggesting that the prevalence of cirrhosis from NAFLD is about 2%. Hispanics and whites are at higher risk for NAFLD, whereas its prevalence is intriguingly low in African Americans.

PATHOBIOLOGY

The major risk factors for NAFLD include obesity, type 2 diabetes mellitus (Chapter 229), metabolic syndrome, and dyslipidemia (see Table 152-2). Other comorbidities associated with NAFLD include polycystic ovary syndrome (Chapter 235), hypothyroidism (Chapter 226), hypopituitarism (Chapter 224), hypogonadism (Chapter 234), and sleep apnea (Chapter 100).

Two fundamental defects in NAFLD are insulin resistance/hyperinsulinemia and excessive levels of nonesterified fatty acids within the hepatocytes. An excessive influx of nonesterified fatty acids into the hepatocytes results in macrovesicular steatosis, which is predominantly centrilobular in location.

In addition, patients with NAFLD have increased de novo intrahepatic lipogenesis. Although patients with NAFLD robustly esterify free fatty acids into neutral triglycerides, free fatty acids within the hepatocytes are considered the primary mediators of cell injury (lipotoxicity). In the background of hepatic steatosis, factors that promote cell injury, inflammation, and fibrosis include oxidative stress, endoplasmic reticulum stress, apoptosis, adipocytokines, and stellate cell activation. The sources of oxidative stress include mitochondria and microsomes. Adipocytokines that play an important role in the pathogenesis of NAFLD include adiponectin and TNF-α. It is unclear why some patients with NAFLD exhibit NASH, whereas other patients with a comparable risk factor profile have only simple steatosis. There is a consistent and significant relationship of *PNPLA3* genetic polymorphisms with the severity of steatosis and other histologic features of NAFLD. However, the genetic factors that play a role in NASH and NAFLD have not been fully elucidated.

CLINICAL MANIFESTATIONS

NAFLD is often asymptomatic but may rarely cause fatigue and right upper quadrant pain. Physical examination may reveal hepatomegaly, palmar erythema (see Fig. 146-2), and spider nevi. If liver disease is advanced,

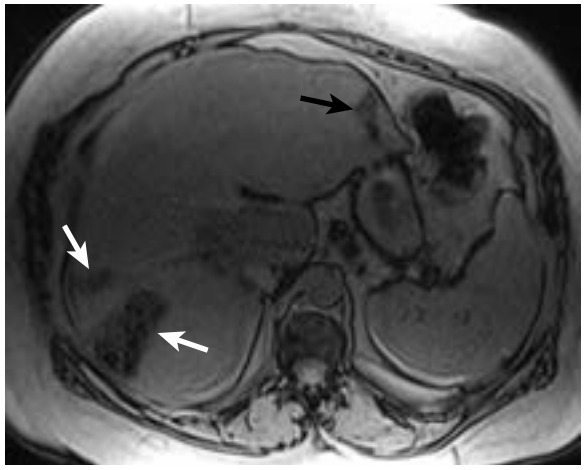

FIGURE 152-2. Magnetic resonance image of chronic liver disease due to nonalcoholic fatty liver disease. Out-of-phase T1-weighted images show focal fat infiltration of varying shapes in right lobe (*white arrows*) and left lobe (*black arrow*). (Courtesy Professor Kumar Sandrasegaran, Indiana University School of Medicine, Indianapolis, Ind.)

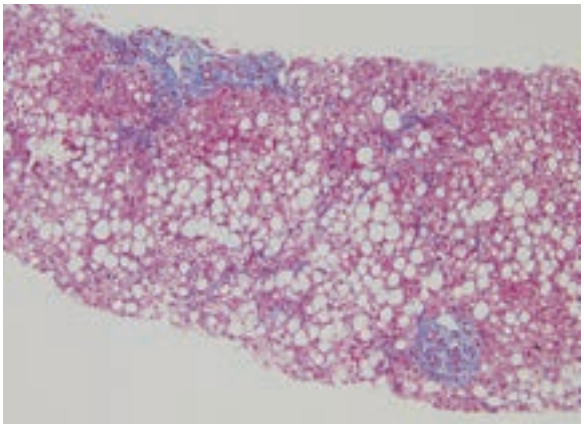

FIGURE 152-3. Liver biopsy specimen showing nonalcoholic steatohepatitis, with increased fat and with early cirrhosis. (Courtesy the NASH Clinical Research Network.)

the features of liver failure, such as ascites, encephalopathy, and abdominal collateral vessels, are present. Simple steatosis is benign with a minimal risk of cirrhosis, whereas NASH is progressive and can lead to cirrhosis (Chapter 153) and liver failure (Chapter 154). In up to 20% of patients with NASH, liver histologic features will worsen and cirrhosis will develop during a 10- to 15-year period. Severe obesity, advancing age, and diabetes are believed to be risk factors for disease progression. Disease progression during the early phase can be identified only with a repeated liver biopsy, but in later stages, the signs and symptoms of portal hypertension (e.g., abdominal collateral vessels and low platelet count) indicate the development of cirrhosis. Patients with NASH-induced cirrhosis are at risk for development of hepatocellular carcinoma (Chapter 196). Patients with NAFLD have several metabolic risks that predispose them to atherosclerosis, and coronary artery disease is the single most common cause of death in patients with NAFLD.

DIAGNOSIS

NAFLD is generally suspected when aminotransferase levels are asymptomatically elevated in an individual with metabolic risk factors (obesity, diabetes, or the metabolic syndrome) or when liver imaging (ultrasound, computed tomography, or magnetic resonance imaging) obtained for another reason shows fatty infiltration (Fig. 152-2). The diagnosis of NAFLD requires that there is no history of previous or ongoing significant alcohol consumption, no exposure to steatogenic medications, and no evidence of other causes of liver disease, such as viral hepatitis B or C.[8] Elevated levels of aminotransferases, although common, are not required for the diagnosis of NAFLD. In contrast to alcoholic liver disease, ALT levels are higher than AST levels, but they rarely exceed 250 IU/L. In general, AST and ALT levels do not have diagnostic or prognostic significance. Mild hyperferritinemia is common and should not be confused with hereditary hemochromatosis (Chapter 212). Similarly, low-grade autoantibody (antinuclear antibody, anti–smooth muscle antibody) positivity is not uncommon and should not be confused with autoimmune liver disease (Chapter 149). Because steatosis is common in patients with Wilson disease (Chapter 211), serum ceruloplasmin levels should be obtained as part of the diagnostic evaluation.

Fatty liver on ultrasonography has a positive predictive value of only 77% and a negative predictive value of only 67% compared with liver biopsy. Abdominal magnetic resonance imaging is more accurate, but its high cost limits its usefulness in routine practice. Because neither of these imaging tests can differentiate simple steatosis from NASH or identify cirrhosis until hepatic fibrosis has caused nodular liver or overt portal hypertension, liver biopsy is required to establish the presence of NASH or cirrhosis. Common indications for a percutaneous liver biopsy in patients with NAFLD include persistently high aminotransferase levels, inability to exclude a competing or a coexisting cause (e.g., iron overload or autoimmune liver disease), and clinical suspicion of NASH and advanced fibrosis. In patients with NASH, liver histology shows steatosis, inflammation, ballooning, and fibrosis (Fig. 152-3). However, liver histology alone cannot reliably distinguish NASH from alcoholic hepatitis because these two entities have remarkably similar microscopic features.

TREATMENT

Rx

Lifestyle modification with dietary restriction and regular exercise is the first choice of treatment for NAFLD. Weight reduction and increased physical activity consistently reduce liver fat, improve glucose control and insulin sensitivity, and may improve histopathologic features.[9] It is generally recommended that patients with NAFLD lose 10% of their body weight in a gradual fashion, but this goal is difficult to achieve. If resources are available, a multidisciplinary approach with behavioral therapy, dietary advice, and monitoring by a professional nutritionist and an exercise expert is more successful than a prescriptive approach.[10]

Statins (e.g., atorvastatin 20 mg daily) with or without vitamins C and E can improve liver test results and reduce subsequent NAFLD. In one large trial, 800 IU of vitamin E administered daily for 2 years significantly improved liver histologic findings in adults, but other trials did not show such convincing benefits in children. Thiazolidinedione insulin sensitizers (pioglitazone and rosiglitazone) improve steatosis, inflammation, and ballooning, and perhaps fibrosis, but with the side effect of an average gain of nearly 10 pounds. Unfortunately, the weight gain that is common with thiazolidinediones may offset the histologic benefits that they offer. Treatment with pentoxifylline (400 mg three times daily) can also improve liver enzymes and liver histologic findings in individuals with NASH.

Patients with NAFLD often have dyslipidemia that puts them at excessive risk for coronary artery disease; their dyslipidemia (Chapter 206) should be treated aggressively with statins and other lipid-lowering agents, which can be safely administered to patients with NAFLD and NASH. In morbidly obese individuals with NASH and other significant metabolic comorbidities, foregut bariatric surgery can lead to significant improvement in hepatic histologic features, but the physician must exclude the presence of portal hypertension before offering this type of surgery.[11] Carefully selected patients with decompensated cirrhosis due to NASH can be treated with liver transplantation (Chapter 154), but recurrence during the post-transplantation period is common.

PREVENTION

The measures to prevent NAFLD include maintaining optimal body weight, exercising regularly, and treating any associated metabolic comorbidities such as diabetes and dyslipidemia. Avoidance of saturated fat, high fructose intake, and alcohol consumption may reduce the development of NAFLD.

PROGNOSIS

Steatosis alone is generally benign, whereas steatohepatitis is often progressive. One third of NAFLD patients may have remission within 7 years, mostly depending on modest weight reduction. Otherwise, however, data are sparse regarding the risk and risk factors for the progression of NAFLD and NASH to cirrhosis and liver failure.

Because NAFLD often coexists with one or more components of the metabolic syndrome, its presence reflects guarded long-term overall prognosis. The long-term complications in patients with simple steatosis generally result from cardiovascular disease and atherosclerosis, not from liver failure.

Patients with NASH also are at risk for liver failure and liver cancer (Chapter 196), in addition to their significantly increased risk of cardiovascular disease. For example, among patients with NASH on their initial biopsy,

one third or more develop progressive fibrosis during a mean follow-up interval of about 5 years. Older age, diabetes, and ballooning and fibrosis on liver biopsy are important predictors of progression.

Grade A References

A1. Mathurin P, O'Grady J, Carithers RL, et al. Corticosteroids improve short-term survival in patients with severe alcoholic hepatitis: meta-analysis of individual patient data. *Gut*. 2011;60:255-260.

A2. Nguyen-Khac E, Thevenot T, Piquet MA, et al. Glucocorticoids plus *N*-acetylcysteine in severe alcoholic hepatitis. *N Engl J Med*. 2011;365:1781-1789.

A3. Parker R, Armstrong MJ, Corbett C, et al. Systematic review: pentoxifylline for the treatment of severe alcoholic hepatitis. *Aliment Pharmacol Ther*. 2013;37:845-854.

A4. Park SH, Kim DJ, Kim YS, et al. Pentoxifylline vs. corticosteroid to treat severe alcoholic hepatitis: a randomised, non-inferiority, open trial. *J Hepatol*. 2014;61:792-798.

A5. Mathurin P, Louvet A, Duhamel A, et al. Prednisolone with vs without pentoxifylline and survival of patients with severe alcoholic hepatitis: a randomized clinical trial. *JAMA*. 2013;310:1033-1041.

A6. Hurst LC, Badalamente MA, Hentz VR, et al. Injectable collagenase clostridium histolyticum for Dupuytren's contracture. *N Engl J Med*. 2009;361:968-979.

A7. Promrat K, Kleiner DE, Niemeier HM, et al. Randomized controlled trial testing the effects of weight loss on nonalcoholic steatohepatitis (NASH). *Hepatology*. 2010;51:121-129.

A8. Athyros VG, Tziomalos K, Gossios TD, et al. Safety and efficacy of long-term statin treatment for cardiovascular events in patients with coronary heart disease and abnormal liver tests in the Greek Atorvastatin and Coronary Heart Disease Evaluation (GREACE) study: a post-hoc analysis. *Lancet*. 2010;376:1916-1922.

A9. Foster T, Budoff MJ, Saab S, et al. Atorvastatin and antioxidants for the treatment of nonalcoholic fatty liver disease: the St. Francis Heart Study randomized clinical trial. *Am J Gastroenterol*. 2011;106:71-77.

A10. Sanyal AJ, Chalasani N, Kowdley KV, et al. Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. *N Engl J Med*. 2010;362:1675-1685.

A11. Lavine JE, Schwimmer JB, Van Natta ML, et al. Effect of vitamin E or metformin for treatment of nonalcoholic fatty liver disease in children and adolescents: the TONIC randomized controlled trial. *JAMA*. 2011;305:1659-1668.

A12. Boettcher E, Csako G, Pucino F, et al. Meta-analysis: pioglitazone improves liver histology and fibrosis in patients with non-alcoholic steatohepatitis. *Aliment Pharmacol Ther*. 2012;35:66-75.

A13. Zein CO, Yerian LM, Gogate P, et al. Pentoxifylline improves nonalcoholic steatohepatitis: a randomized placebo-controlled trial. *Hepatology*. 2011;54:1610-1619.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Stickel F, Hampe J. Genetic determinants of alcoholic liver disease. *Gut*. 2012;61:150-159.
2. Orman ES, Odena G, Bataller R. Alcoholic liver disease: pathogenesis, management, and novel targets for therapy. *J Gastroenterol Hepatol*. 2013;28(suppl 1):77-84.
3. Singal AK, Kamath PS, Gores GJ, et al. Alcoholic hepatitis: current challenges and future directions. *Clin Gastroenterol Hepatol*. 2014;12:555-564.
4. Mathurin P, Moreno C, Samuel D, et al. Early liver transplantation for severe alcoholic hepatitis. *N Engl J Med*. 2011;365:1790-1800.
5. Cuthbert JA, Arslanlar S, Yepuri J, et al. Predicting short-term mortality and long-term survival for hospitalized US patients with alcoholic hepatitis. *Dig Dis Sci*. 2014;59:1594-1602.
6. Altamirano J, Miquel R, Katoonizadeh A, et al. A histologic scoring system for prognosis of patients with alcoholic hepatitis. *Gastroenterology*. 2014;146:1231-1239.
7. Hyysalo J, Mannisto VT, Zhou Y, et al. A population-based study on the prevalence of NASH using scores validated against liver histology. *J Hepatol*. 2014;60:839-846.
8. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease. Practice Guideline by the American Association for the Study of Liver Disease, American College of Gastroenterology, and American Gastroenterology Association. *Gastroenterology*. 2012; 142:1592-1609.
9. Thoma C, Day CP, Trenell MI. Lifestyle interventions for the treatment of non-alcoholic fatty liver disease in adults: a systematic review. *J Hepatol*. 2012;56:255-266.
10. Corrado RL, Torres DM, Harrison SA. Review of treatment options for nonalcoholic fatty liver disease. *Med Clin North Am*. 2014;98:55-72.
11. Hafeez S, Ahmed MH. Bariatric surgery as potential treatment for nonalcoholic fatty liver disease: a future treatment by choice or by chance? *J Obes*. 2013;2013:839275.

REVIEW QUESTIONS

1. A 17-year-old adolescent boy presents with right upper quadrant discomfort of 3 to 4 months in duration. This pain is constant and is dull in nature. He consumes no alcohol. He has a long-standing history of type 1 diabetes, for which he takes subcutaneous insulin twice daily. Physical examination reveals a relatively thin individual with tender hepatomegaly. His blood test results were positive for elevated alanine aminotransferase (82 U/L; normal, <45 U/L), blood glucose (181 mg/dL; normal, <125 mg/dL), and hemoglobin A_{1c} (9%; normal, <7%). A liver ultrasound examination revealed increased echogenicity. Which is the likely diagnosis for this patient's abdominal discomfort, elevated alanine aminotransferase level, and tender hepatomegaly?

A. Nonalcoholic fatty liver disease
B. Glycogen hepatopathy
C. Wilson disease
D. Hemochromatosis
E. Glycogen storage disorder

Answer: B Tender hepatomegaly, elevated alanine aminotransferase level, and echogenic liver by ultrasound evaluation are suggestive of some form of infiltrative liver disorder. In type 1 diabetics, especially those with poorly controlled blood glucose levels, there is risk of glycogen hepatopathy, which may be manifested with abdominal discomfort, hepatomegaly, and elevated liver test results. Glycogen accumulation also enhances echogenicity on ultrasound examination and may mimic fatty liver. One must keep in mind that type 1 diabetes is rarely if ever associated with nonalcoholic fatty liver disease. It is believed that hyperinsulinemia is needed for development of nonalcoholic fatty liver disease.

2. A 46-year-old woman presents with 3 weeks' duration of fatigue, anorexia, low-grade fever, and jaundice. She is a heavy drinker and has been consuming more than 7 drinks of vodka each night for the previous 5 years. She is otherwise healthy and is not taking any medications. Physical examination reveals an emaciated woman with deep scleral icterus, peripheral muscle wasting, tender hepatomegaly, and pedal edema. Laboratory test results show macrocytic anemia, mild leukocytosis, elevated aminotransferase levels, and direct bilirubinemia. Which of the following treatments is not a suitable treatment option for this individual?

A. Alcohol abstinence
B. Prednisolone
C. *N*-Acetylcysteine
D. Liver transplantation
E. Enteral nutrition

Answer: D On the basis of the clinical picture, the most likely diagnosis for this individual's condition is acute alcoholic hepatitis. Liver transplantation is not a suitable treatment option for acute alcoholic hepatitis in individuals who are actively consuming alcohol. Most transplant centers require up to 6 months of total abstinence and some form of behavioral therapy for underlying alcoholism. In eligible patients, 28 days of oral prednisolone improves short-term survival in individuals with acute alcoholic hepatitis. Similarly, some previous clinical investigations have shown promising results for *N*-acetylcysteine and enteral nutrition.

3. A 37-year-old morbidly obese man is noted to have an echogenic liver incidentally when he had an abdominal ultrasound examination performed for his hematuria. He has no right upper quadrant pain. He rarely consumes alcohol and takes no medications. His physical examination is essentially unremarkable except for morbid obesity with a body mass index of 42 kg/m^2. Blood test results show mildly elevated aminotransferase levels but otherwise are within normal range. Your likely diagnosis is nonalcoholic fatty liver disease (NAFLD), and you recommend weight loss. Which of the following best represents his liver condition?

A. NAFLD is a rare condition, and its incidence has remained relatively steady during the last decade.
B. Type 2 diabetes, obesity, dyslipidemia, and metabolic syndrome are the major risk factors for NAFLD.
C. Hyperthyroidism is a common cause of NAFLD.
D. Sleep disturbance is rare in NAFLD.
E. This condition can be ignored because it is rarely of any clinical significance.

Answer: B Hypothyroidism, hypopituitarism, hypogonadism, polycystic ovary syndrome, and obstructive sleep disorders are emerging risk factors for NAFLD. Although hyperthyroidism is associated with liver test abnormalities, it is not known to be associated with fatty liver disease. In fact, thyroid hormone analogues may improve hepatic steatosis, although they may carry significant adverse events and are not a recommended treatment. Clearly, the incidence of NAFLD is increasing because of the increasing prevalence of obesity and related comorbid conditions, and it is now one of the most common causes of cirrhosis, liver failure, and liver cancer. Therefore, it cannot be ignored as clinically insignificant.

4. A 46-year-old woman presents with chronically elevated liver enzymes and mild hepatomegaly. She has type 2 diabetes, which is diet controlled, and has dyslipidemia, which is managed with a low-dose statin. She denies significant alcohol consumption. Her liver biopsy showed 30% hepatic steatosis with ballooned hepatocytes and mild sinusoidal fibrosis. Your clinical and histologic diagnosis is nonalcoholic steatohepatitis. Which of the following treatments is not a suitable option for her?

A. Weight loss
B. Pioglitazone
C. Vitamin E
D. Ursodeoxycholic acid
E. Pentoxifylline

Answer: D Weight loss is the first-line treatment option for both NAFLD and NASH. Loss of 3 to 5% of body weight will improve hepatic steatosis, whereas a loss of 7 to 10% is needed to see an improvement in steatohepatitis. Calorie restriction and physical exercise are advised to achieve the desired weight loss in individuals with NAFLD and NASH. When such measures are not achieved, therapy with agents such as pioglitazone, vitamin E, and pentoxifylline can be considered. In clinical studies, these compounds have shown therapeutic efficacy in various subgroups of patients. Ursodeoxycholic acid, a medication used to treat primary biliary cirrhosis, has been tested in individuals with NASH but has not proved to be effective. Metformin is not effective for improving liver histology findings in NASH and therefore should not be used for this purpose.

153

CIRRHOSIS AND ITS SEQUELAE

GUADALUPE GARCIA-TSAO

DEFINITION

Cirrhosis, which can be the final stage of any chronic liver disease, is a diffuse process characterized by fibrosis and conversion of normal architecture to structurally abnormal nodules (Fig. 153-1). These "regenerative" nodules lack normal lobular organization and are surrounded by fibrous tissue. The process involves the whole liver and generally is considered irreversible. Although cirrhosis is histologically an "all-or-nothing" diagnosis, clinically it can be classified by its status as compensated or decompensated. Decompensated cirrhosis is defined by the presence of ascites, variceal bleeding, encephalopathy, or jaundice, which are complications that result from the main consequences of cirrhosis: portal hypertension and liver insufficiency.

EPIDEMIOLOGY

Because many patients with cirrhosis are asymptomatic until decompensation occurs, it is difficult to assess the real prevalence and incidence of cirrhosis in the general population. The prevalence of chronic liver disease or cirrhosis worldwide is estimated to be 100 (range, 25 to 400) per 100,000 subjects, but it varies widely by country and by geographic region.

Cirrhosis is an important cause of morbidity and mortality worldwide and in the United States. According to the World Health Organization, about 800,000 people die of cirrhosis annually. In the United States, cirrhosis accounts for about 32,000 deaths each year, or a death rate of 10.3 per 100,000, thereby making it the 12th leading cause of death overall. Importantly, chronic liver disease and cirrhosis are the sixth leading cause of death in the United States in individuals between 25 and 44 years of age and fifth in individuals between 45 and 64 years of age. As chronic liver disease affects people in their most productive years of life, it has a significant impact on the economy as a result of premature death, illness, and disability.

Any chronic liver disease can lead to cirrhosis (Table 153-1). Chronic viral hepatitis C and alcoholic liver disease are the most common causes of cirrhosis, followed by nonalcoholic fatty liver disease (in particular, nonalcoholic steatohepatitis) and chronic hepatitis B (Chapters 149 and 152). However, other causes of cirrhosis include cholestatic and autoimmune liver diseases, such as primary biliary cirrhosis, primary sclerosing cholangitis (Chapter 155), and autoimmune hepatitis (Chapter 149), and metabolic diseases, such as hemochromatosis, Wilson disease, and α_1-antitrypsin deficiency (Chapter 146). When all potential causes have been investigated and excluded, cirrhosis is considered "cryptogenic." Many cases of cryptogenic cirrhosis are now thought to be due to nonalcoholic fatty liver disease (Chapter 152).

Although the entity termed primary biliary cirrhosis assumes the presence of cirrhosis, this term is actually misleading. Primary biliary cirrhosis (Chapter 155) is an immune-mediated cholestatic chronic liver disease that is characterized by progressive destruction of intrahepatic bile ducts and progresses over time from an initial stage in which fibrosis is minimal (stage 1) to a final stage in which there is well-established cirrhosis (stage 4).

PATHOBIOLOGY

Liver Fibrosis and Cirrhosis

The key pathogenic feature underlying liver fibrosis and cirrhosis is activation of hepatic stellate cells. Hepatic stellate cells, which are known as *Ito cells* or *perisinusoidal cells*, are located in the space of Disse between hepatocytes and sinusoidal endothelial cells. Normally, hepatic stellate cells are quiescent and serve as the main storage site for retinoids (vitamin A). In response to injury, hepatic stellate cells become activated, as a result of which they lose their vitamin A deposits, proliferate, develop a prominent rough endoplasmic reticulum, and secrete extracellular matrix (collagen types I and III, sulfated proteoglycans, and glycoproteins). In addition, they become contractile hepatic myofibroblasts.[1]

Unlike other capillaries, normal hepatic sinusoids lack a basement membrane. The sinusoidal endothelial cells themselves contain large fenestrae (100 to 200 nm in diameter) that permit the passage of large molecules with molecular masses up to 250,000 daltons. Collagen deposition in the space of Disse, as occurs in cirrhosis, leads to defenestration of the sinusoidal endothelial cells ("capillarization" of the sinusoids), thereby altering exchange between plasma and hepatocytes and resulting in a decreased sinusoidal diameter that is further exacerbated by the contraction of stellate cells.

Complications of Cirrhosis

The two main consequences of cirrhosis are portal hypertension, with the accompanying hyperdynamic circulatory state, and liver insufficiency (Fig. 153-2). The development of varices and ascites is a direct consequence of portal hypertension and the hyperdynamic circulatory state, whereas jaundice occurs as a result of an inability of the liver to excrete bilirubin (i.e., liver insufficiency). Encephalopathy is the result of both portal hypertension and liver insufficiency. Ascites, in turn, can become complicated by infection, which is called *spontaneous bacterial peritonitis*, and by functional renal failure, which is called *hepatorenal syndrome.*

Portal Hypertension and the Hyperdynamic Circulatory State

In cirrhosis, portal hypertension results from both an increase in resistance to portal flow and an increase in portal venous inflow. The initial mechanism is increased sinusoidal vascular resistance (Fig. 153-3) secondary to (1) deposition of fibrous tissue and subsequent compression by regenerative nodules (fixed component), which could theoretically be amenable to antifibrotic agents and could be ameliorated by resolving the underlying etiologic process, and (2) active vasoconstriction (functional component), which is amenable to the action of vasodilators such as nitroprusside and is caused by a deficiency in intrahepatic nitric oxide (NO) as well as enhanced activity of vasoconstrictors.

Early in the portal hypertensive process, the spleen grows and sequesters platelets and other formed blood cells, thereby leading to hypersplenism. In addition, vessels that normally drain into the portal system, such as the coronary vein, reverse their flow and shunt blood away from the portal system to the systemic circulation. These portosystemic collaterals are insufficient to decompress the portal venous system and offer additional resistance to portal

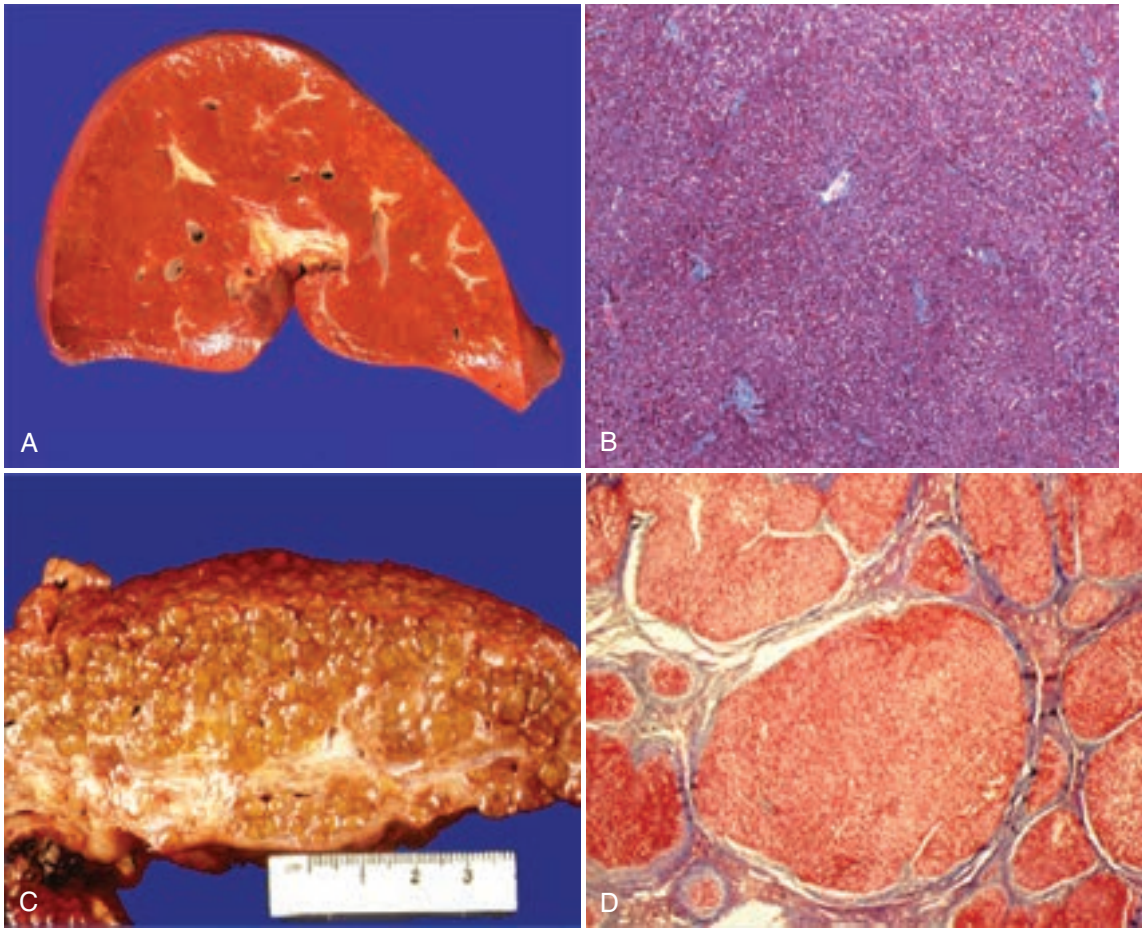

FIGURE 153-1. Gross and microscopic images of a normal and cirrhotic liver. A, Gross image of a normal liver with a smooth surface and homogeneous texture. **B,** On microscopic examination, liver sinusoids are organized, and vascular structures are normally distributed. **C,** Gross image of a cirrhotic liver. The liver has an orange-tawny color with an irregular surface and a nodular texture. **D,** On microscopic examination, the architecture is disorganized, and there are regenerative nodules surrounded by fibrous tissue.

TABLE 153-1 CAUSES OF CIRRHOSIS

MAIN FACTORS CAUSING CIRRHOSIS
Chronic hepatitis C
Alcoholic liver disease
Nonalcoholic fatty liver disease
Chronic hepatitis B
OTHER CAUSES OF CIRRHOSIS (<2% OF ALL CASES)
Cholestatic and autoimmune liver diseases
Primary biliary cirrhosis
Primary sclerosing cholangitis
Autoimmune hepatitis
Intrahepatic or extrahepatic biliary obstruction
Mechanical obstruction
Biliary atresia
Cystic fibrosis
Metabolic disorders
Hemochromatosis
Wilson disease
α_1-Antitrypsin deficiency
Glycogen storage diseases
Abetalipoproteinemia
Porphyria
Hepatic venous outflow obstruction
Budd-Chiari syndrome
Veno-occlusive disease
Right-sided heart failure
Drugs and toxins
Intestinal bypass
Indian childhood cirrhosis

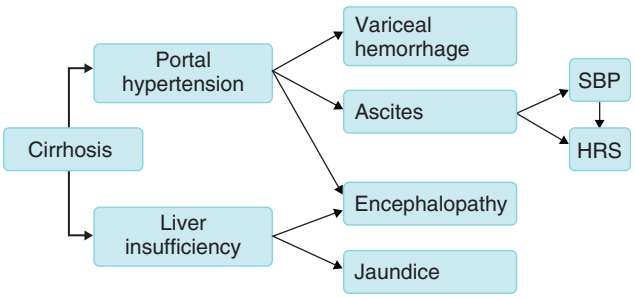

FIGURE 153-2. Complications of cirrhosis result from portal hypertension or liver insufficiency. Varices and variceal hemorrhage are a direct consequence of portal hypertension. Ascites results from sinusoidal portal hypertension and can be complicated by infection (spontaneous bacterial peritonitis [SBP]) or renal dysfunction (hepatorenal syndrome [HRS]). Hepatic encephalopathy results from portosystemic shunting (i.e., portal hypertension) and liver insufficiency. Jaundice results solely from liver insufficiency.

flow. As collaterals develop, an increase in portal blood inflow, which results from splanchnic vasodilation, maintains the portal hypertensive state. Splanchnic arteriolar vasodilation is, in turn, secondary to increased production of NO. Thus, the paradox in portal hypertension is that a deficiency of NO in the intrahepatic vasculature leads to vasoconstriction and increased resistance, whereas overproduction of NO in the extrahepatic circulation leads to vasodilation and increased portal flow.

In addition to splanchnic vasodilation, there is systemic vasodilation, which by causing a decreased *effective* arterial blood volume leads to activation of neurohumoral systems (renin-angiotensin-aldosterone system),

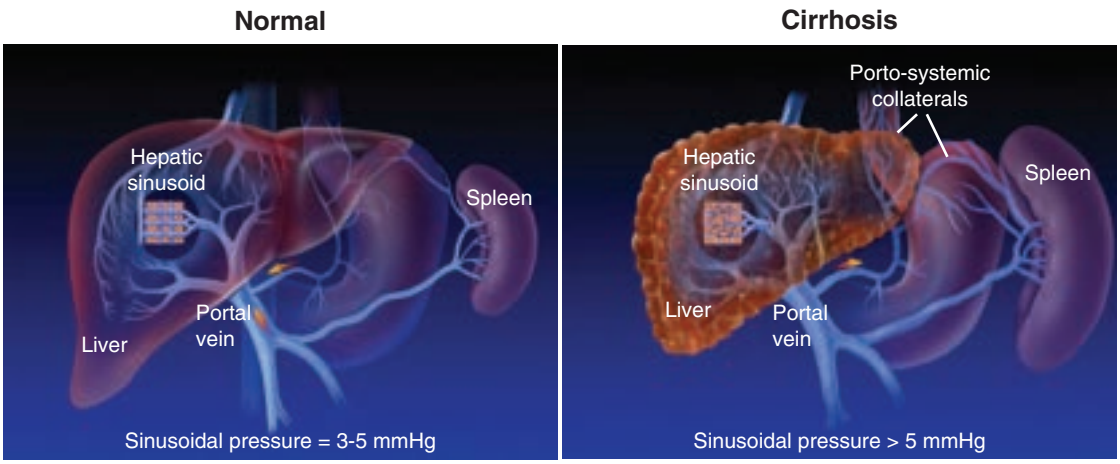

FIGURE 153-3. Hepatic sinusoidal pressure is increased in cirrhosis. In the normal liver (*left*), the intrahepatic vasculature is compliant, and the hepatic venous pressure gradient, which is a measure of sinusoidal pressure, is 5 mm Hg or lower. In the cirrhotic liver (*right*), the sinusoidal architecture is distorted by regenerative nodules and fibrosis that lead to increased intrahepatic resistance, portal hypertension, splenomegaly, and portosystemic collaterals; the hepatic venous pressure gradient is above 5 mm Hg. Complications of cirrhosis develop when the gradient increases above 10 to 12 mm Hg.

retention of sodium, expansion of plasma volume, and development of a hyperdynamic circulatory state. This hyperdynamic circulatory state maintains portal hypertension, thereby leading to the formation and growth of varices, and plays an important role in the development of all other complications of cirrhosis.

Varices and Variceal Hemorrhage

The complication of cirrhosis that results most directly from portal hypertension is the development of portal-systemic collaterals, the most relevant of which are those that form through dilation of the coronary and gastric veins and constitute gastroesophageal varices. The initial formation of esophageal collaterals depends on a threshold portal pressure, clinically established by a hepatic venous pressure gradient of 10 to 12 mm Hg, below which varices do not develop.

Development of a hyperdynamic circulatory state leads to further dilation and growth of varices and eventually to their rupture and variceal hemorrhage, one of the most dreaded complications of portal hypertension. Tension in a varix determines variceal rupture and is directly proportional to variceal diameter and intravariceal pressure and inversely proportional to variceal wall thickness.

Ascites and Hepatorenal Syndrome

Ascites, which is the accumulation of intraperitoneal fluid, in cirrhosis is secondary to sinusoidal hypertension and retention of sodium. Cirrhosis leads to sinusoidal hypertension by blocking hepatic venous outflow both anatomically by fibrosis and regenerative nodules and functionally by increased postsinusoidal vascular tone. Similar to the formation of esophageal varices, a threshold hepatic venous pressure gradient of 12 mm Hg is needed for the formation of ascites. In addition, retention of sodium replenishes the intravascular volume and allows the continuous formation of ascites. Retention of sodium results from vasodilation that is mostly due to an increase in NO production because NO inhibition in experimental animals increases urinary sodium excretion, lowers plasma aldosterone levels, and reduces ascites. With progression of cirrhosis and portal hypertension, vasodilation is more pronounced, thereby leading to further activation of the renin-angiotensin-aldosterone and sympathetic nervous systems and resulting in further sodium retention (refractory ascites), water retention (hyponatremia), and renal vasoconstriction (hepatorenal syndrome).

Spontaneous Bacterial Peritonitis

Spontaneous bacterial peritonitis, an infection of ascitic fluid, occurs in the absence of perforation of a hollow viscus or an intra-abdominal inflammatory focus, such as an abscess, acute pancreatitis, or cholecystitis. Bacterial translocation, or the migration of bacteria from the intestinal lumen to mesenteric lymph nodes and other extraintestinal sites, is the main mechanism implicated in spontaneous bacterial peritonitis. Impaired local and systemic immune defenses are a major element in promoting bacterial translocation and, together with shunting of blood away from the hepatic Kupffer cells through portosystemic collaterals, allow a transient bacteremia to become more prolonged, thereby colonizing ascitic fluid. Spontaneous bacterial peritonitis occurs in patients with reduced ascites defense mechanisms, such as a low complement level in ascitic fluid. Another factor that promotes bacterial translocation in cirrhosis is bacterial overgrowth attributed to a decrease in small bowel motility and intestinal transit time. Infections, particularly from gram-negative bacteria, can precipitate renal dysfunction through worsening of the hyperdynamic circulatory state.

Encephalopathy

Hepatic encephalopathy is brain dysfunction caused by liver insufficiency, portosystemic shunting, or both.[2,3] Ammonia, a toxin normally removed by the liver, plays a key role in its pathogenesis. In cirrhosis, ammonia accumulates in the systemic circulation because of shunting of blood through portosystemic collaterals and decreased liver metabolism (i.e., liver insufficiency). The presence of large amounts of ammonia in the brain damages supporting brain cells or astrocytes and leads to structural changes characteristic of hepatic encephalopathy (Alzheimer type II astrocytosis). Ammonia results in upregulation of astrocytic peripheral-type benzodiazepine receptors, the most potent stimulants of neurosteroid production. Neurosteroids are the major modulators of γ-aminobutyric acid, which results in cortical depression and hepatic encephalopathy. Other toxins, such as manganese, also accumulate in the brain, particularly the globus pallidus, where they lead to impaired motor function. Other yet-to-be-elucidated toxins may also be involved in the pathogenesis of encephalopathy.

Jaundice

Jaundice (Chapter 147) in cirrhosis is a reflection of the inability of the liver to excrete bilirubin and is therefore the result of liver insufficiency. However, in cholestatic diseases leading to cirrhosis (e.g., primary biliary cirrhosis, primary sclerosing cholangitis, vanishing bile duct syndrome), jaundice is more likely due to biliary damage than to liver insufficiency. Other indicators of liver insufficiency, such as the presence of encephalopathy or prolongation of the international normalized ratio, help determine the most likely contributor to hyperbilirubinemia (Chapter 147).

Cardiopulmonary Complications

The hyperdynamic circulatory state eventually results in high-output heart failure with decreased peripheral utilization of oxygen, a complication that has been referred to as *cirrhotic cardiomyopathy.* Vasodilation at the level of the pulmonary circulation leads to arterial hypoxemia, the hallmark of hepatopulmonary syndrome. Normal pulmonary capillaries are 8 μm in diameter, and red blood cells (slightly less than 8 μm) pass through them one cell at a time, thereby facilitating oxygenation. In *hepatopulmonary syndrome,* the pulmonary capillaries are dilated up to 500 μm, so passage of red cells through the pulmonary capillaries may be many cells thick. As a result, a large number of red cells are not oxygenated, which causes the equivalent of a right-to-left shunt.

Conversely, *portopulmonary hypertension* occurs when the pulmonary bed is exposed to vasoconstrictive substances that may be produced in the splanchnic circulation and bypass metabolism by the liver; the initial result

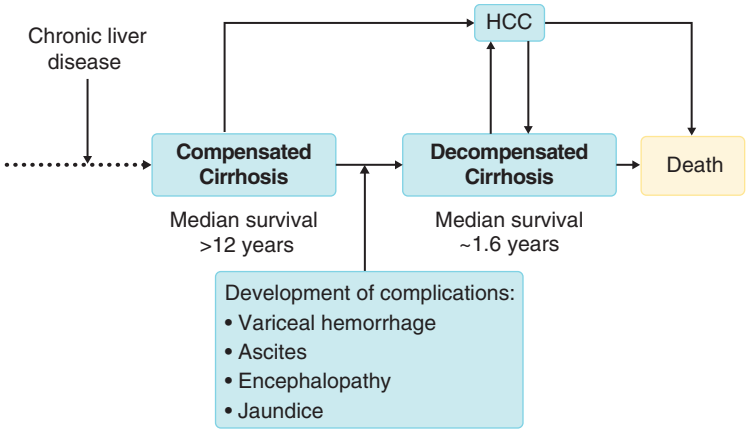

FIGURE 153-4. Natural history of cirrhosis. Any chronic liver disease will lead to cirrhosis. Initially, cirrhosis will be compensated (median survival, >12 years), but once complications (ascites, variceal hemorrhage, encephalopathy, jaundice) develop, it becomes decompensated (median survival, 1.6 years). Hepatocellular carcinoma (HCC) can develop at any stage and precipitate decompensation and death.

is reversible pulmonary hypertension. However, because these factors result in endothelial proliferation, vasoconstriction, in situ thrombosis, and obliteration of vessels, irreversible pulmonary hypertension ensues.

CLINICAL MANIFESTATIONS

The clinical manifestations of cirrhosis range widely, depending on the stage of cirrhosis, from an asymptomatic patient with no signs of chronic liver disease to a patient who is confused and jaundiced with severe muscle wasting and ascites. The natural history of cirrhosis is characterized by an initial phase, termed *compensated* cirrhosis, followed by a rapidly progressive phase marked by the development of complications of portal hypertension or liver dysfunction (or both), termed *decompensated* cirrhosis (Fig. 153-4).[4] In the compensated phase, liver synthetic function is mostly normal, and portal pressure, although increased, is below the threshold level required for the development of varices or ascites. As the disease progresses, portal pressure increases and liver function worsens, thereby resulting in the development of ascites, portal hypertensive gastrointestinal bleeding, encephalopathy, and jaundice. The development of any of these clinically detectable complications marks the transition from a compensated to a decompensated phase. Progression to death may be accelerated by the development of other complications, such as recurrent gastrointestinal bleeding, renal impairment (refractory ascites, hepatorenal syndrome), hepatopulmonary syndrome, and sepsis (spontaneous bacterial peritonitis). The development of hepatocellular carcinoma (Chapter 196) may accelerate the course of the disease at any stage (see Fig. 153-4). Transition from a compensated to a decompensated stage occurs at a rate of approximately 5 to 7% per year. The median time to decompensation, or the time at which half the patients with compensated cirrhosis will become decompensated, is about 6 years.

Compensated Cirrhosis

In this stage, cirrhosis is mostly asymptomatic and is diagnosed either during the evaluation of chronic liver disease or fortuitously during routine physical examination, biochemical testing, imaging for other reasons, endoscopy showing gastroesophageal varices, or abdominal surgery in which a nodular liver is detected. Nonspecific fatigue, decreased libido, or sleep disturbances may be the only complaints. About 40% of patients with compensated cirrhosis have esophageal varices. Nonbleeding gastroesophageal varices are asymptomatic, and their presence (without bleeding) does not denote decompensation.

Decompensated Cirrhosis

At this stage, there are signs of decompensation: ascites, variceal hemorrhage, jaundice, hepatic encephalopathy, or any combination of these findings. Ascites, which is the most frequent sign of decompensation, is present in 80% of patients with decompensated cirrhosis.

Variceal Hemorrhage

Gastroesophageal varices are present in approximately 50% of patients with newly diagnosed cirrhosis. The prevalence of varices correlates with the severity of liver disease and ranges from 40% in Child A cirrhotic patients (Table 153-2) to 85% in Child C cirrhotic patients.

TABLE 153-2 THE TWO MOST COMMONLY USED SCORING SYSTEMS IN CIRRHOSIS

1. CHILD-TURCOTTE-PUGH (CTP) SCORE (RANGE, 5-15)

	POINTS ASCRIBED		
Parameters	**1**	**2**	**3**
Ascites	None	Grade 1-2 (or easy to treat)	Grade 3-4 (or refractory)
Hepatic encephalopathy	None	Grade 1-2 (or induced by a precipitant)	Grade 3-4 (or spontaneous)
Bilirubin (mg/dL)	<2	2-3	>3
Albumin (g/dL)	>3.5	2.8-3.5	<2.8
Prothrombin time (seconds > control) or INR	<4 <1.7	4-6 1.7-2.3	>6 >2.3

CTP classification: Child A: score of 5-6; Child B: score of 7-9; Child C: score of 10-15

2. MODEL OF END-STAGE LIVER DISEASE (MELD) SCORE (RANGE, 6-40)

$[0.957 \times \text{LN (creatinine in mg/dL)} + 0.378 \times \text{LN (bilirubin in mg/dL)} + 1.12 \times \text{LN (INR)} + 0.643] \times 10$

INR = international normalized ratio; LN = natural logarithm.

Both the development of varices and the growth of small varices occur at a rate of 7 to 8% per year. The incidence of a first variceal hemorrhage in patients with small varices is about 5% per year, whereas medium and large varices bleed at a rate of approximately 15% per year. Large varices, severe liver disease, and red wale markings on varices are independent predictors of variceal hemorrhage. Bleeding from gastroesophageal varices can be manifested as overt hematemesis, melena, or both (Chapter 135).

Ascites and Hyponatremia

Ascites is the most common cause of decompensation in cirrhosis and occurs at a rate of 7 to 10% per year. The most frequent symptoms associated with ascites are increased abdominal girth, which is often described by the patient as tightness of the belt or garments around the waist, and recent weight gain. When it is present in small to moderate amounts, ascites can be identified on examination by bulging flanks, flank dullness, and shifting dullness (Chapter 146).

Hyponatremia, which is defined as a serum sodium concentration below 130 mEq/L (Chapter 116), is present in about 25% of patients with cirrhosis and ascites. However, patients with cirrhosis do not usually have significant neurologic manifestations or hyponatremia, presumably because it typically develops gradually. Nevertheless, hyponatremia is a marker of the severity of cirrhosis and is associated with poorer quality of life and the development of hepatic encephalopathy.[5]

Hepatorenal syndrome is a type of prerenal kidney injury that occurs in patients with cirrhosis and ascites.[6] It represents the extreme of the spectrum of abnormalities that lead to cirrhotic ascites and is the result of maximal peripheral vasodilation as well as maximal activation of hormones that cause the retention of sodium and water and intense vasoconstriction of the renal arteries. Hepatorenal syndrome is divided into two types based on clinical characteristics and prognosis. Type 1 hepatorenal syndrome is rapidly progressive *acute* kidney injury in which the rise in serum creatinine concentration occurs within a 2-week period. Type 2 hepatorenal syndrome is more slowly progressive and associated with ascites that is refractory to diuretics. Patients with hepatorenal syndrome usually have tense ascites that responds poorly to diuretics, but no specific symptoms or signs typify this entity.

Spontaneous Bacterial Peritonitis

About one third of cirrhotic patients are admitted for bacterial infection or acquire a bacterial infection during hospitalization, the most common being spontaneous bacterial peritonitis. The two most important predictors of the development of bacterial infection are the severity of liver disease and admission for gastrointestinal hemorrhage. The most frequent clinical manifestations of spontaneous bacterial peritonitis are fever, jaundice, and abdominal pain. On physical examination, there is typically abdominal tenderness, with or without rebound tenderness, or ileus (or both). However, up to one third of patients with spontaneous bacterial peritonitis initially may have no abdominal symptoms or symptoms of infection[7] and may present instead with encephalopathy, acute kidney injury, or evidence of shock.

Hepatic Encephalopathy

Hepatic encephalopathy, which is the neuropsychiatric manifestation of cirrhosis, occurs at a rate of approximately 2 to 3% per year. Hepatic encephalopathy associated with cirrhosis is of gradual onset and rarely fatal. It is manifested as a wide spectrum of neurologic and psychiatric abnormalities ranging from subclinical alterations to coma. Clinically, it is characterized by alterations in consciousness and behavior ranging from inversion of the sleep-wake pattern and forgetfulness (grade 1); to confusion, bizarre behavior, and disorientation (grade 2); to lethargy and profound disorientation (grade 3); to coma (grade 4). On physical examination, early stages may demonstrate only a distal tremor, but the hallmark of overt hepatic encephalopathy is the presence of asterixis (Chapter 154). In addition, patients with hepatic encephalopathy may have sweet-smelling breath, a characteristic termed *fetor hepaticus*.

Pulmonary Complications

Hepatopulmonary syndrome is associated with exertional dyspnea, which can lead to extreme debilitation. Clubbing of the fingers, cyanosis, and vascular spiders may be seen on physical examination. Hepatopulmonary syndrome is present in approximately 5 to 10% of patients awaiting liver transplantation.

Portopulmonary hypertension is manifested as exertional dyspnea, syncope, and chest pain. On examination, an accentuated second sound and right ventricular heave are prominent (Chapter 68).

DIAGNOSIS

The diagnosis of cirrhosis should be considered in any patient with chronic liver disease.[8] In asymptomatic patients with *compensated cirrhosis*, typical signs of cirrhosis may not be present, and the physical examination and laboratory test findings may be entirely normal. The diagnosis may often require histologic confirmation by liver biopsy, which is the "gold standard" for the diagnosis of cirrhosis. However, liver biopsy is an invasive procedure subject to sampling error, and the presence of cirrhosis can often be confirmed noninvasively by a combination of serum biomarkers, imaging techniques, and measurements of liver stiffness.

Physical Examination

On physical examination, stigmata of cirrhosis consist of muscle atrophy, mainly involving the bitemporal muscle regions and the thenar and hypothenar eminences; spider angiomas, mostly on the trunk, face, and upper limbs; and palmar erythema involving the thenar and the hypothenar eminences and the tips of the fingers. Although muscle atrophy is a marker of liver insufficiency, spider angiomas and palmar erythema are markers of vasodilation and a hyperdynamic circulation. Men may have hair loss on the chest and abdomen, gynecomastia, and testicular atrophy. Petechiae and ecchymoses may be present as a result of thrombocytopenia or a prolonged prothrombin time. Dupuytren contracture, which is a thickening of the palmar fascia, occurs mostly in alcoholic cirrhosis. A pathognomonic feature of cirrhosis is the finding on abdominal examination of a small right liver lobe, with a span of less than 7 cm on percussion, and a palpable left lobe that is nodular with increased consistency. Splenomegaly may also be present and is indicative of portal hypertension. Collateral circulation on the abdominal wall (caput medusae) may also develop as a consequence of portal hypertension. Absence of any of these physical findings does not exclude cirrhosis.

Laboratory Tests

Laboratory test results suggestive of cirrhosis include even subtle abnormalities in serum levels of albumin or bilirubin or elevation of the international normalized ratio. The most sensitive and specific laboratory finding suggestive of cirrhosis in the setting of chronic liver disease is a low platelet count (<150,000/μL), which occurs as a result of portal hypertension and hypersplenism. Other serum markers that are often abnormal include levels of aspartate aminotransferase, γ-glutamyl transpeptidase, hyaluronic acid, α_2-macroglobulin, haptoglobin, tissue metalloproteinase inhibitor 1, and apolipoprotein A. Combinations of these tests have been used to predict the presence of cirrhosis, but they are not as accurate as imaging studies.[9]

Imaging Studies

Confirmatory imaging tests include computed tomography, ultrasound, and magnetic resonance imaging. Findings consistent with cirrhosis include a nodular contour of the liver, a small liver with or without hypertrophy of the left or caudate lobe, splenomegaly, and, in particular, identification of intra-abdominal collateral vessels indicative of portal hypertension (Fig. 153-5). With increasing fibrosis, the liver becomes stiff, and this stiffness can be measured by ultrasound (transient elastography, acoustic radiation force impulse imaging) or magnetic resonance imaging. Measurement of liver stiffness, a new noninvasive technique, appears to be useful in the diagnosis of cirrhosis and in excluding its presence.[10] These tests are becoming more widely available, and typical findings on any of these imaging studies, together with a compatible clinical picture, are indicative of the presence of cirrhosis. A liver biopsy then would not be required unless the degree of inflammation or other features require investigation.

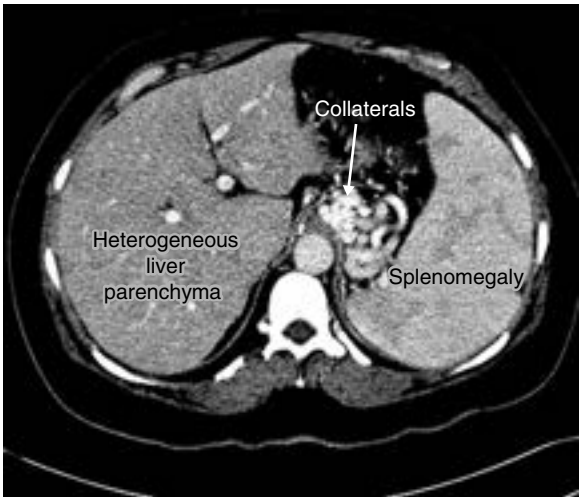

FIGURE 153-5. **Computed tomography in a patient with compensated cirrhosis. The liver parenchyma is heterogeneous, there is splenomegaly, and, importantly, there are portosystemic collaterals.**

In *decompensated cirrhosis*, detection of ascites, variceal bleeding, or encephalopathy in the setting of chronic liver disease essentially establishes the diagnosis of cirrhosis, so a liver biopsy is not necessary to establish the diagnosis. Patients with decompensated cirrhosis often exhibit malnutrition, more severe muscle wasting, more numerous vascular spiders, and hypotension and tachycardia as a result of the hyperdynamic circulatory state.

Portal Pressure Measurements

Direct measurements of portal pressure involve catheterization of the portal vein, are cumbersome, and may be associated with complications. Hepatic vein catheterization with measurement of wedged and free pressure is the simplest, safest, most reproducible, and most widely used method to indirectly measure portal pressure. Portal pressure measurements are expressed as the hepatic venous pressure gradient: the gradient between wedged hepatic venous pressure, which is a measure of sinusoidal pressure, and free hepatic or inferior vena cava pressure, which is used as an internal zero reference point. In a patient with clinical evidence of portal hypertension (e.g., varices), the hepatic venous pressure gradient is useful in the differential diagnosis of the cause of portal hypertension: it will be normal (3 to 5 mm Hg) in prehepatic causes of portal hypertension, such as portal vein thrombosis (Chapter 143), and in intrahepatic but presinusoidal causes, such as schistosomiasis (Chapter 355); but it will be abnormal (≥6 mm Hg) in sinusoidal causes of portal hypertension, such as cirrhosis, and in postsinusoidal causes, such as veno-occlusive disease. In patients with viral or alcoholic cirrhosis, a hepatic venous pressure gradient of 10 mm Hg or greater ("clinically significant" portal hypertension) predicts the development of complications of portal hypertension, and its reduction with pharmacologic therapy predicts a favorable outcome in patients with cirrhosis. In patients with variceal hemorrhage, a hepatic venous pressure gradient of more than 20 mm Hg predicts recurrent variceal hemorrhage and may portend death.

Complications of Cirrhosis

Varices and Variceal Hemorrhage

Upper gastrointestinal endoscopy (Chapter 134) remains the main method for diagnosis of varices and variceal hemorrhage. Varices are classified as small (straight, minimally elevated veins above the esophageal mucosal surface), medium (tortuous veins occupying less than one third of the esophageal

lumen), or large (occupying more than one third of the esophageal lumen). The diagnosis of variceal hemorrhage is made when diagnostic esophagogastroduodenoscopy shows one of the following: active bleeding from a varix, a "white nipple" overlying a varix, clots overlying a varix, or varices with no other potential source of bleeding.

Ascites

The most common cause of ascites is cirrhosis, which accounts for 80% of cases. Peritoneal malignant disease (e.g., peritoneal metastases from gastrointestinal tumors or ovarian cancer), heart failure (Chapter 58), and peritoneal tuberculosis (Chapter 324) together account for another 15% of cases. The initial, most cost-effective, and least invasive method to confirm the presence of ascites is abdominal ultrasonography.

Diagnostic paracentesis is a safe procedure that should be performed in every patient with new-onset ascites, even in those with coagulopathy. Ultrasound guidance should be used in patients in whom percussion cannot locate the ascites or in whom a first paracentesis attempt does not yield fluid. The fluid in a patient with new-onset ascites should always be evaluated for albumin (with simultaneous estimation of serum albumin), total protein, and polymorphonuclear (PMN) blood cell count, and bacteriologic cultures and cytology should be performed. The PMN cell count and bacteriologic culture are useful to exclude infection (either spontaneous or secondary bacterial peritonitis), and cytologic evaluation is needed if peritoneal carcinomatosis is suspected. Depending on the clinical setting, additional tests can be performed on the fluid: glucose and lactate dehydrogenase levels (if secondary bacterial peritonitis is suspected), smear and culture for acid-fast bacilli (if peritoneal tuberculosis is suspected), and amylase level (if pancreatic ascites is suspected).

The serum-ascites albumin gradient and ascites protein levels are useful in the differential diagnosis of ascites (Table 153-3). The serum-ascites albumin gradient correlates with sinusoidal pressure and will therefore be elevated (>1.1 g/dL) in patients in whom the source of ascites is the hepatic sinusoid (e.g., cirrhosis or cardiac ascites). Protein levels in ascitic fluid are an indirect marker of the integrity of the hepatic sinusoids: normal sinusoids are permeable structures that "leak" protein, whereas sinusoids in cirrhosis are "capillarized" and do not leak as much protein. The three main causes of ascites—cirrhosis, peritoneal malignant disease or tuberculosis, and heart failure—can easily be distinguished by combining the results of both the serum-ascites albumin gradient and ascites total protein content. Cirrhotic ascites typically has a high serum-ascites albumin gradient and low protein, cardiac ascites has a high serum-ascites albumin gradient and high protein, and ascites secondary to peritoneal malignant disease typically has a low serum-ascites albumin gradient and high protein. A high serum B-type natriuretic peptide has a high diagnostic accuracy in the diagnosis of ascites due to heart failure.

Hepatorenal Syndrome

Hepatorenal syndrome, which is a diagnosis of exclusion, should be diagnosed only after diuretics have been discontinued, any condition that leads to worsening of the hemodynamic status of the cirrhotic patient has been excluded or treated, and intravascular volume has been expanded with albumin. Ascites unresponsive to diuretics is universal, and dilutional hyponatremia is almost always present. The differential diagnosis includes conditions that worsen vasodilation, such as sepsis, the use of vasodilators, and large-volume paracentesis not accompanied by albumin infusion; conditions that decrease effective arterial blood volume, such as gastrointestinal hemorrhage, overdiuresis, or diarrhea (often induced by overdoses of lactulose); conditions that induce renal vasoconstriction, such as nonsteroidal anti-inflammatory drugs; and nephrotoxic insults, such as from aminoglycosides.

TABLE 153-3 USING THE SERUM-ASCITES ALBUMIN GRADIENT AND THE ASCITES TOTAL PROTEIN LEVEL TO DIAGNOSE THE CAUSE OF ASCITES

CONDITION	SERUM-ASCITES ALBUMIN GRADIENT*	ASCITES TOTAL PROTEIN LEVEL†
Cirrhosis	High	Low
Malignant ascites	Low	High
Cardiac ascites	High	High

*High is more than 1.1 g/dL; low is less than 1.1 g/dL.
†High is more than 2.5 g/dL; low is less than 2.5 g/dL.

Spontaneous Bacterial Peritonitis

A high index of suspicion and early diagnosis are key in the management of spontaneous bacterial peritonitis. Diagnostic paracentesis should be performed in any patient with symptoms or signs of spontaneous bacterial peritonitis, including unexplained encephalopathy and renal dysfunction. Because spontaneous bacterial peritonitis is often asymptomatic and frequently community acquired, diagnostic paracentesis should be performed when any cirrhotic patient is admitted to the hospital, regardless of the cause for admission.

The diagnosis of spontaneous bacterial peritonitis is established by an ascitic fluid PMN count greater than 250/μL. Bacteria can be isolated from ascitic fluid in only 40 to 50% of cases, even with sensitive methods such as inoculation directly into a blood culture bottle. Spontaneous bacterial peritonitis is mostly a monobacterial infection, usually with gram-negative enteric organisms. However, the widespread use of antibiotic prophylaxis in cirrhosis has led to an increased prevalence of infections with multidrug-resistant organisms. Anaerobes and fungi very rarely cause spontaneous bacterial peritonitis; their presence, as well as a polymicrobial infection, should raise suspicion of secondary bacterial peritonitis.

Hepatic Encephalopathy

The diagnosis of overt hepatic encephalopathy is mainly clinical and based on a history and physical examination that shows alterations in consciousness and behavior as well as the presence of asterixis. Ammonia levels are unreliable, and there is poor correlation between the grade of hepatic encephalopathy and ammonia blood levels. High levels (>150 μmol/L) are, however, indicative of hepatic encephalopathy and can be useful in the evaluation of a patient with neurocognitive disturbances of unknown origin. Psychometric tests and an electroencephalogram, which typically shows generalized slow waves and the presence of triphasic waves, are commonly used in research but are not generally used for clinical diagnosis. Minimal or subclinical hepatic encephalopathy, which is present in up to 80% of patients with cirrhosis, is diagnosed solely on the basis of abnormal results of psychometric and neuropsychological tests of attention (e.g., number connection test, digit symbol test) and psychomotor function (e.g., grooved pegboard). Screening of cirrhotic patients for minimal hepatic encephalopathy is not widely recommended because diagnostic tests are not standardized and the benefits of treatment are uncertain.

Hepatopulmonary Syndrome and Portopulmonary Hypertension

The diagnostic criteria for hepatopulmonary syndrome are arterial hypoxemia with a Pao_2 of less than 80 mm Hg or an alveolar arterial oxygen gradient of greater than 15 mm Hg, along with evidence of pulmonary vascular shunting on contrast echocardiography (Chapter 55) or a ^{99m}Tc-labeled macroaggregated albumin scan demonstrating abnormal shunting of radioactivity to the brain.[11] Portopulmonary hypertension is diagnosed by the presence of mean pulmonary arterial pressure higher than 25 mm Hg on right-sided heart catheterization, provided pulmonary capillary wedge pressure is less than 15 mm Hg.

TREATMENT

Rx

Treatment of cirrhosis should ideally be aimed at interruption or reversal of fibrosis. Although antifibrotic drugs have not been shown to reverse fibrosis consistently or to improve outcomes in cirrhotic patients, eradication of the hepatitis C or the hepatitis B virus has been associated with reversal of fibrosis. Treatment of compensated cirrhosis is currently directed at preventing the development of decompensation by (1) treating the underlying liver disease (e.g., antiviral therapy for hepatitis C or B) to reduce fibrosis and to prevent decompensation; (2) avoiding factors that could worsen liver disease, such as alcohol, hepatotoxic drugs, and superimposed viral infections; and (3) screening for varices (to prevent variceal hemorrhage) and for hepatocellular carcinoma (to treat it at an early stage) (Fig. 153-6). Treatment of decompensated cirrhosis focuses on specific decompensating events and the option of liver transplantation. Increasingly, data reveal that different therapies for the same complication may be applicable to patients with different risk profiles, mainly based on severity of the disease (see Table 153-2).

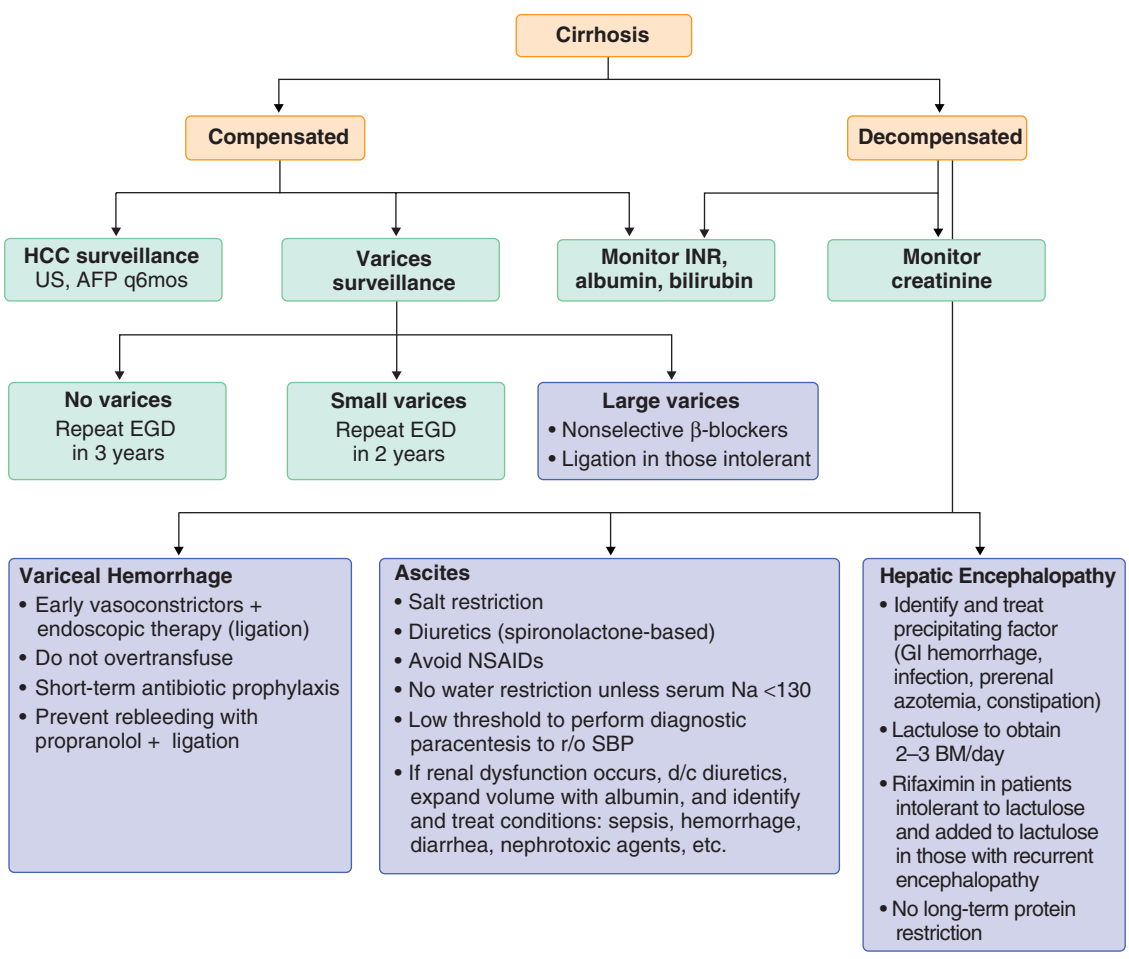

FIGURE 153-6. Summary of the management of compensated and decompensated cirrhosis. AFP = α-fetoprotein; BM = bowel movement; d/c = discontinue; EGD = esophagogastroduodenoscopy; GI = gastrointestinal; HCC = hepatocellular carcinoma; INR = international normalized ratio; Na = sodium; NSAIDs = nonsteroidal anti-inflammatory drugs; r/o = rule out; SBP = spontaneous bacterial peritonitis; US = ultrasound.

Varices and Variceal Bleeding

Reducing portal pressure decreases the risk for the development of varices and variceal hemorrhage as well as the risk for ascites and death. Nonselective β-adrenergic blockers (propranolol, nadolol) reduce portal pressure by producing splanchnic vasoconstriction and decreasing portal venous inflow. In patients with cirrhosis and medium or large varices that have never bled, nonselective β-blockers significantly reduce the risk for first variceal hemorrhage. Treatment options have included propranolol (initiated at a dose of 20 mg orally twice a day) and nadolol (initiated at a dose of 20 mg orally every day), with the dose titrated to produce a resting heart rate of about 50 to 55 beats per minute.[12] β-Adrenergic blockers also reduce portal pressure and lower the risk for development of ascites. Endoscopic variceal ligation (see Fig. 134-3), a therapy that aims to obliterate varices by placing rubber rings on variceal columns, is at least as useful as traditional nonselective β-blockers to prevent a first variceal hemorrhage.[■] Ligation has no effect on portal pressure and can lead to hemorrhage from ligation-induced ulcers. Carvedilol (a nonselective β-blocker with vasodilating properties at a dose of 12.5 mg/day) has been shown to be superior to ligation[■] and to lower portal pressure in patients who do not respond to propranolol[■]; however, its use in patients with ascites may not be advisable because of its vasodilating effect. A rational approach is to start therapy with propranolol or nadolol and to use ligation in patients who cannot tolerate or have contraindications to β-blockers. In the compensated patient, carvedilol could be used in patients who cannot tolerate propranolol or nadolol.

In patients with no varices, nonselective β-blockers do not prevent the development of varices and are associated with more side effects. In patients with small varices, data are insufficient to recommend therapy with nonselective β-blockers. Endoscopy should be repeated every 2 to 3 years in patients with no varices, every 1 to 2 years in patients with small varices, and sooner in patients with decompensated disease so that effective therapy can be instituted before the varices grow in size and bleed.[13]

Patients with cirrhosis and variceal hemorrhage require resuscitation in an intensive care unit. However, overtransfusion should be avoided because it can precipitate rebleeding.[■] Hemoglobin values should be maintained at about 8 g/dL. Prophylactic antibiotics should be used in this setting not only to prevent bacterial infections but also to decrease rebleeding and death.[■] The recommended antibiotic is oral norfloxacin at a dose of 400 mg twice daily for 5 to 7 days, although intravenous ceftriaxone at a dose of 1 g/day for 5 to 7 days is preferable in patients with advanced liver disease (malnutrition, ascites, encephalopathy, and jaundice) or in those already receiving norfloxacin prophylaxis.[■]

The most effective specific therapy for the control of active variceal hemorrhage is the combination of a vasoconstrictor with endoscopic therapy. Safe vasoconstrictors include terlipressin, somatostatin, and the somatostatin analogues octreotide and vapreotide[■]; they can be initiated at admission to the hospital and continued for 2 to 5 days. The vasoconstrictor currently available in the United States is octreotide, which is used as a 50-μg intravenous bolus followed by an infusion at 50 μg/hour. The transjugular intrahepatic portosystemic shunt (TIPS) is generally recommended for patients who fail to respond to standard therapy (Fig. 153-7). In patients at high risk of failure, Child C (score 10-13) patients and Child B patients who have actively bleeding varices at endoscopy, preemptive TIPS placement (24 to 48 hours after admission) is associated with a reduced failure rate and a significant improvement in survival.[■] Therefore, preemptive "early" TIPS should be considered in this high-risk subpopulation of patients with variceal hemorrhage. Even though patients with cirrhosis and coagulopathy (prolonged international normalized ratio or decreased platelet count) have a higher risk of not responding to hemostasis for variceal hemorrhage or bleeding from procedures (ligation, paracentesis, TIPS, surgery), neither recombinant factor VII nor eltrombopag has proved beneficial in randomized trials, and neither is recommended.

After control of hemorrhage, the 1-year recurrence of hemorrhage without treatment is high at about 60%. Therefore, therapy to prevent rebleeding should be instituted before the patient is discharged. The recommended therapy is combination of nonselective β-blockers (propranolol or nadolol), with or without isosorbide mononitrate, and endoscopic variceal ligation.[■] The dose of β-blockers should be the maximal dose tolerated, and endoscopic variceal ligation should be repeated every 2 to 4 weeks until the varices are obliterated. Patients who had TIPS placed during the episode of acute variceal hemorrhage do not require β-blockers or ligation but do require periodic Doppler examination of the shunt to assess its patency.

Shunt therapy, either surgical or through radiologic placement of a TIPS, should be used in patients whose variceal bleeding has persisted

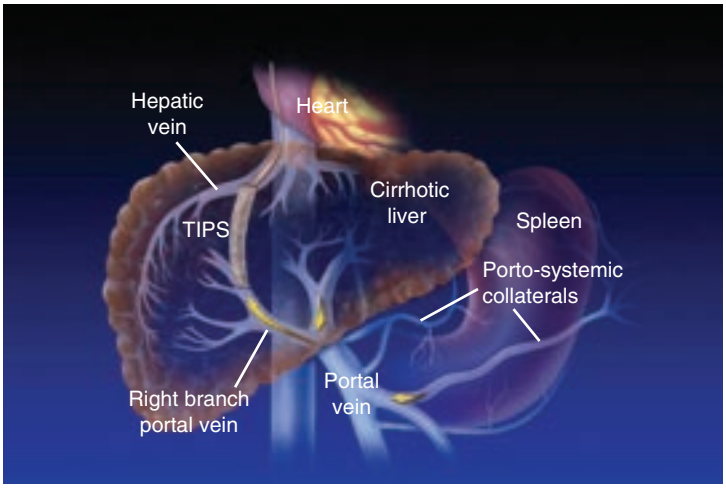

FIGURE 153-7. Transjugular intrahepatic portosystemic shunt (TIPS). This shunt is performed by interventional radiologists and consists of an expandable metal stent (most often coated with polytetrafluoroethylene) that connects a branch of the portal vein (high-pressure vein) to a branch of the hepatic vein (low-pressure vein). The shunt decompresses the portal system and portosystemic collaterals, and therefore it is used in the treatment of selected patients with cirrhosis and variceal hemorrhage. The shunt also decompresses the hepatic sinusoids and therefore is also used in the treatment of refractory ascites.

or recurred despite combined pharmacologic and endoscopic therapy. Both types of shunts are equally effective,[NEW] and the choice will depend on local expertise. Although the uncovered TIPS frequently occludes, newer polytetrafluoroethylene-covered stents are associated with lower occlusion rates and lower rates of hepatic encephalopathy.

Ascites

Salt restriction and diuretics constitute the mainstay of management of ascites. Dietary sodium intake should be restricted to 2 g/day. A more restrictive diet is not recommended and may compromise nutritional status.

Spironolactone, which is more effective than loop diuretics, should be started at a dose of 100 mg/day (once a day in the morning). The dose should be adjusted every 3 to 4 days to a maximal effective dose of 400 mg/day. Furosemide, at an escalated dose from 40 to 160 mg/day, can be started concurrently with spironolactone if ascites is tense or added subsequently if weight loss is inadequate or if hyperkalemia develops with spironolactone alone. The goal is weight loss of 1 kg in the first week and 2 kg/week subsequently. However, diuretics should be reduced if the rate of weight loss is more than 0.5 kg/day in patients without peripheral edema or more than 1 kg/day in patients with peripheral edema. Side effects of diuretic therapy include hypovolemic hyponatremia, hyperkalemia, renal dysfunction, encephalopathy, and, with spironolactone, painful gynecomastia.

In the 10 to 20% of patients with ascites who are refractory to diuretics, large-volume paracentesis, aimed at removal of all or most of the fluid, plus albumin at a dose of 6 to 8 g intravenously per liter of ascites removed, particularly when more than 5 L is removed at once, is a reasonable approach. The frequency of large-volume paracentesis is dictated by the rapidity at which the ascites reaccumulates. TIPS with uncovered stents is more effective than large-volume paracentesis plus albumin in preventing recurrent ascites but is associated with a higher rate of encephalopathy.[NEW] In patients requiring frequent large-volume paracentesis (more than twice per month), a polytetrafluoroethylene-covered TIPS should be considered. A peritoneovenous shunt, with use of a subcutaneously placed silicone tube that transfers ascites from the peritoneal cavity to the systemic circulation, can be used in patients who are not candidates for TIPS or liver transplantation. Automated flow pumps to move ascitic fluid to the bladder are under investigation.

Hyponatremia

Fluid restriction to 1.5 liters daily is recommended for severe hyponatremia (serum sodium concentration <130 mEq/L), but adherence with this recommendation is poor. Although V2-receptor antagonists such as tolvaptan are effective at increasing free water excretion and raising the serum sodium level in hyponatremic cirrhotic patients,[NEW] they have no overall effect on survival and are not approved for this use because of their hepatotoxicity. Because hyponatremia is a marker of neurologic dysfunction and increased mortality, short-term use of tolvaptan (15 mg/day, increased to 30 to 60 mg/day if needed) may be indicated only as a bridge to liver transplantation in patients in whom transplantation is imminent (Chapter 154).

Hepatorenal Syndrome

Because hepatorenal syndrome is a functional kidney injury that results from hemodynamic abnormalities secondary to end-stage liver disease and severe portal hypertension and is associated with a high mortality despite specific therapy, documented hepatorenal syndrome in the absence of prerenal azotemia or acute tubular necrosis is an indication for transplantation (Chapter 154). Specific therapies for hepatorenal syndrome that have been used to "bridge" a patient to transplantation include vasoconstrictors (terlipressin, norepinephrine, octreotide plus midodrine) plus albumin, TIPS, and extracorporeal albumin dialysis, which is an experimental hemofiltration dialysis method that uses an albumin dialysate. The largest experience is with the use of terlipressin, which at a dose 0.5 to 2.0 mg intravenously every 4 to 6 hours leads to a higher reversal rate of hepatorenal syndrome compared with placebo.[NEW] Because terlipressin is not yet available in the United States, the combination most used is octreotide (100 to 200 µg subcutaneously three times a day) plus midodrine (7.5 to 12.5 mg orally three times a day), with the dose adjusted to obtain an increase of at least 15 mm Hg in mean arterial pressure. Improvements may become clinically noticeable at day 7.

Spontaneous Bacterial Peritonitis

Empirical antibiotic therapy with an intravenous third-generation cephalosporin (e.g., cefotaxime, 2 g intravenously every 12 hours, or ceftriaxone, 2 g intravenously every 24 hours) should be initiated as soon as the diagnosis is established and before culture results are available; the minimal duration of therapy should be 5 days. Response to recommended empirical antibiotics is significantly lower in patients with health care–associated infections (i.e., infections that occur in patients who have been in a health care facility within the prior 3 months) and particularly low in nosocomial infections (i.e., infections that occur >48 hours after hospitalization)[14] because more infections are caused by multidrug-resistant organisms. In these settings, broader spectrum antibiotics (e.g., vancomycin-tazobactam, imipenem, ertapenem [Chapter 108]) should be given initially and then modified after culture results and antibiotic sensitivities are obtained. Aminoglycosides should be avoided because of the high incidence of renal toxicity in cirrhotic patients.

Repeated diagnostic paracentesis should be performed 2 days after antibiotics are started, by which time the number of PMN neutrophils in ascitic fluid should have decreased by more than 25% from baseline. Lack of response should prompt further investigations to exclude secondary peritonitis. The renal dysfunction associated with spontaneous bacterial peritonitis can be prevented by the intravenous administration of albumin, particularly in patients who have any evidence of renal dysfunction (blood urea nitrogen >30 mg/dL or creatinine >1 mg/dL, or both) or serum bilirubin concentration higher than 4 mg/dL at the time of diagnosis. Albumin has been used at a dose of 1.5 g per kilogram of body weight at diagnosis, repeated on the third day at an intravenous dose of 1 g per kilogram of body weight. However, this dosing is empirical and should not exceed 100 g per dose.

The administration of nonabsorbable (or poorly absorbable) antibiotics can prevent the development of spontaneous bacterial peritonitis and other infections in cirrhosis by selectively eliminating gram-negative organisms in the gut. However, the widespread use of prophylactic norfloxacin is associated with a higher rate of infections by antibiotic-resistant organisms. Long-term antibiotic prophylaxis with oral norfloxacin at a dose of 400 mg/day may be justified only in two groups: patients who have recovered from a previous episode of spontaneous bacterial peritonitis and patients who have an ascites protein level of less than 1 g/L with advanced liver and circulatory dysfunction as evidenced by the presence of jaundice, hyponatremia, or renal dysfunction.

Hepatic Encephalopathy

Treatment of overt hepatic encephalopathy starts by exclusion of alternative causes of altered mental status. Once the diagnosis of hepatic encephalopathy has been made, treatment involves identifying and treating the precipitating factor and reducing the ammonia level.[15] Precipitating factors include infections, overdiuresis, gastrointestinal bleeding, high oral protein load, and constipation. Narcotics and sedatives contribute to hepatic encephalopathy by directly depressing brain function. TIPS is a common precipitant of hepatic encephalopathy, and shunt reduction or occlusion may be required. Among agents aimed at decreasing ammonia production in the gut, lactulose (15 to 30 mL orally twice daily adjusted to obtain two or three soft bowel movements per day) has been the first choice for the treatment of episodic overt encephalopathy. Polyethylene glycol 3350-electrolyte solution (4 L orally or by nasogastric tube over 4 hours), however, may lead to a more rapid clinical response.[NEW] Other agents include orally administered nonabsorbable antibiotics, such as rifaximin (550 mg two times per day), neomycin (500 mg to 1 g three times per day), and metronidazole (250 mg two to four times per day). Drugs that may increase ammonia fixation in the liver, such as L-ornithine-L-aspartate, benzoate, and glycerol phenylbutyrate, are being studied.

Once an episode of overt encephalopathy has resolved, secondary prophylaxis with lactulose is recommended. If a precipitating factor has been identified and is well controlled or when liver function or nutritional status has improved, prophylactic therapy may be discontinued. In patients with recurrent encephalopathy, rifaximin together with lactulose is useful in preventing further recurrence.[NEW] Switching dietary protein from an animal source to a vegetable source may be beneficial in recurrent or persistent

encephalopathy, but protein restriction is not necessary and should not be used long term.[16]

Pulmonary Complications

Hepatopulmonary syndrome rarely resolves spontaneously, and medical therapy is disappointing. TIPS is not generally recommended. The only viable treatment is liver transplantation (Chapter 154).

By comparison, liver transplantation is indicated only in a subset of patients with portopulmonary hypertension. In fact, a mean pulmonary arterial pressure higher than 45 mm Hg is an absolute contraindication to liver transplantation. The use of vasodilators should be considered in these patients (Chapter 68).

Surgical Therapy

Liver Transplantation

Orthotopic liver transplantation (Chapter 154), which is the definitive therapy for cirrhosis, is indicated when the risk for dying of liver disease is greater than the risk for dying of transplantation, as determined by a Child-Pugh score of 7 or higher (see Table 153-2) or a Model for End-Stage Liver Disease (MELD) score of 15 or higher. MELD (see Table 153-2), which is a mathematical model that estimates the risk for 3-month mortality, is used to determine the priority for liver transplantation (see E-Table 154-1). The number of available deceased donor organs is lower than the number of patients awaiting liver transplantation; as a result, 15 to 20% of patients awaiting liver transplantation in the United States die before an organ becomes available.

PRIMARY PREVENTION

Treatment of the underlying liver disease, before the development of cirrhosis, is a primary prevention strategy. Because the major causes of cirrhosis are related to lifestyle choices such as injection drug use (Chapter 34), alcohol consumption (Chapter 33), obesity, and unprotected sex, primary prevention programs that focus on encouraging alcohol abstinence, reducing high-risk behavior for hepatitis virus infection, weight reduction, and vaccination for hepatitis B are even better prevention strategies.

PROGNOSIS

The outcome of cirrhosis depends on the patient's stage. Patients with compensated cirrhosis die of liver disease only after transition to a decompensated stage. The 10-year survival rate of patients who remain in a compensated stage is approximately 90%, whereas their likelihood of decompensation is 50% at 10 years. Inception cohort studies of patients with compensated cirrhosis show a median survival of all patients, including those in whom decompensation develops over time, of about 10 years, whereas the median survival after decompensation is about 2 years. Survival is even lower in patients with refractory ascites, hyponatremia, or recurrent variceal hemorrhage and is lowest in patients who are hospitalized with an acute decompensating event, in whom 28-day mortality is about 30% and correlates with the number of organ failures present.[17] Hepatocellular carcinoma develops at a fairly constant rate of 3% per year and is associated with a worse outcome at whatever stage it develops.

Predictors of survival are different in compensated and decompensated patients. Parameters of portal hypertension (varices, splenomegaly, platelet count, γ-globulin) assume greater importance in compensated patients, whereas renal dysfunction, hemorrhage, and hepatocellular carcinoma are important predictive factors in patients with decompensated cirrhosis. In clinical practice, the Child-Pugh score is applicable to all cirrhotic patients, and the MELD score is used in decompensated patients to determine priority for liver transplantation.

Grade A References

A1. Marcellin P, Gane E, Buti M, et al. Regression of cirrhosis during treatment with tenofovir disoproxil fumarate for chronic hepatitis B: a 5-year open-label follow-up study. *Lancet*. 2013;381:468-475.

A2. Gluud LL, Krag A. Banding ligation versus beta-blockers for primary prevention in oesophageal varices in adults. *Cochrane Database Syst Rev*. 2012;8:CD004544.

A3. Tripathi D, Ferguson JW, Kochar N, et al. Randomized controlled trial of carvedilol versus variceal band ligation for the prevention of the first variceal bleed. *Hepatology*. 2009;50:825-833.

A4. Sinagra E, Perricone G, D'Amico M, et al. Systematic review with meta-analysis: the haemodynamic effects of carvedilol compared with propranolol for portal hypertension in cirrhosis. *Aliment Pharmacol Ther*. 2014;39:557-568.

A5. Villanueva C, Colomo A, Bosch A, et al. Transfusion strategies for acute upper gastrointestinal bleeding. *N Engl J Med*. 2013;368:11-21.

A6. Chavez-Tapia NC, Barrientos-Gutierrez T, Tellez-Avila F, et al. Meta-analysis: antibiotic prophylaxis for cirrhotic patients with upper gastrointestinal bleeding—an updated Cochrane review. *Aliment Pharmacol Ther*. 2011;34:509-518.

A7. Fernandez J, Ruiz del Arbol L, Gomez C, et al. Norfloxacin vs ceftriaxone in the prophylaxis of infections in patients with advanced cirrhosis and hemorrhage. *Gastroenterology*. 2006;131:1049-1056.

A8. Wells M, Chande N, Adams P, et al. Meta-analysis: vasoactive medications for the management of acute variceal bleeds. *Aliment Pharmacol Ther*. 2012;35:1267-1278.

A9. Garcia-Pagan JC, Caca K, Bureau C, et al. Early use of TIPS in patients with cirrhosis and variceal bleeding. *N Engl J Med*. 2010;362:2370-2379.

A10. Thiele M, Krag A, Rohde U, et al. Meta-analysis: banding ligation and medical interventions for the prevention of rebleeding from oesophageal varices. *Aliment Pharmacol Ther*. 2012;35:1155-1165.

A11. Henderson JM, Boyer TD, Kutner MH, et al. Distal splenorenal shunt versus transjugular intrahepatic portal systematic shunt for variceal bleeding: a randomized trial. *Gastroenterology*. 2006;130:1643-1651.

A12. Chen RP, Zhu Ge XJ, Huang ZM, et al. Prophylactic use of transjugular intrahepatic portosystemic shunt aids in the treatment of refractory ascites: metaregression and trial sequential meta-analysis. *J Clin Gastroenterol*. 2014;48:290-299.

A13. Dahl E, Gluud LL, Kimer N, et al. Meta-analysis: the safety and efficacy of vaptans (tolvaptan, satavaptan and lixivaptan) in cirrhosis with ascites or hyponatraemia. *Aliment Pharmacol Ther*. 2012;36:619-626.

A14. Gluud LL, Christensen K, Christensen E, et al. Systematic review of randomized trials on vasoconstrictor drugs for hepatorenal syndrome. *Hepatology*. 2010;51:576-584.

A15. Rahimi RS, Singal AG, Cuthbert JA, et al. Lactulose vs polyethylene glycol 3350-electrolyte solution for treatment of overt hepatic encephalopathy: the HELP randomized clinical trial. *JAMA Intern Med*. 2014;174:1727-1733.

A16. Bass NM, Mullen KD, Sanyal A, et al. Rifaximin treatment in hepatic encephalopathy. *N Engl J Med*. 2010;362:1071-1081.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Novo E, Cannito S, Paternostro C, et al. Cellular and molecular mechanisms in liver fibrogenesis. *Arch Biochem Biophys*. 2014;548:20-37.
2. Weissenborn K. Portosystemic encephalopathy. *Handb Clin Neurol*. 2014;120:661-674.
3. Sturgeon JP, Shawcross DL. Recent insights into the pathogenesis of hepatic encephalopathy and treatments. *Expert Rev Gastroenterol Hepatol*. 2014;8:83-100.
4. Tsochatzis EA, Bosch J, Burroughs AK. Liver cirrhosis. *Lancet*. 2014;383:1749-1761.
5. Yu C, Sharma N, Saab S. Hyponatremia: clinical associations, prognosis, and treatment in cirrhosis. *Exp Clin Transplant*. 2013;11:3-11.
6. Adebayo D, Morabito V, Davenport A, et al. Renal dysfunction in cirrhosis is not just a vasomotor nephropathy. *Kidney Int*. 2015;87:509-515.
7. Chinnock B, Hendey GW, Minnigan H, et al. Clinical impression and ascites appearance do not rule out bacterial peritonitis. *J Emerg Med*. 2013;44:903-909.
8. Udell JA, Wang CS, Tinmouth J, et al. Does this patient with liver disease have cirrhosis? *JAMA*. 2012;307:832-842.
9. Castera L. Noninvasive methods to assess liver disease in patients with hepatitis B or C. *Gastroenterology*. 2012;142:1293-1302.
10. Asrani SK, Talwalkar JA, Kamath PS, et al. Role of magnetic resonance elastography in compensated and decompensated liver disease. *J Hepatol*. 2014;60:934-939.
11. Machicao VI, Balakrishnan M, Fallon MB. Pulmonary complications in chronic liver disease. *Hepatology*. 2014;59:1627-1637.
12. Garcia-Tsao G, Bosch J. Management of varices and variceal hemorrhage in cirrhosis. *N Engl J Med*. 2010;362:823-832.
13. Bloom S, Kemp W, Lubel J. Portal hypertension—pathophysiology, diagnosis and management. *Intern Med J*. 2015;45:16-26.
14. Fernandez J, Acevedo J, Castro M, et al. Prevalence and risk factors of infections by multiresistant bacteria in cirrhosis: a prospective study. *Hepatology*. 2012;55:1551-1561.
15. Leise MD, Poterucha JJ, Kamath PS, et al. Management of hepatic encephalopathy in the hospital. *Mayo Clin Proc*. 2014;89:241-253.
16. Vilstrup H, Amodio P, Bajaj J, et al. Hepatic encephalopathy in chronic liver disease: 2014 practice guideline by the European Association for the Study of the Liver and the American Association for the Study of Liver Diseases. *J Hepatol*. 2014;61:642-659.
17. Moreau R, Jalan R, Gines P, et al. Acute-on-chronic liver failure is a distinct syndrome that develops in patients with acute decompensation of cirrhosis. *Gastroenterology*. 2013;144:1426-1437.

REVIEW QUESTIONS

1. A 55-year-old man with chronic hepatitis C complains of fatigue. On physical examination, he has vascular spiders, palmar erythema, and a palpable left lobe of his liver. He has no ascites or asterixis. Laboratory analysis demonstrates aspartate aminotransferase 100, alanine aminotransferase 67, alkaline phosphatase 145, and platelet count of 120,000. Abdominal computed tomography (CT) scan shows a nodular liver contour, portosystemic collaterals, but no masses. Which of the following would you recommend as the next step?

A. Liver biopsy
B. Upper endoscopy
C. Start nonselective β-blockers
D. Start spironolactone
E. Magnetic resonance imaging of the liver

Answer: B Although the "gold standard" in the diagnosis of cirrhosis is liver biopsy, this procedure is not necessary when clinical, laboratory, and imaging results are sufficient to establish the diagnosis. In the setting of chronic liver disease, a palpable left lobe of the liver is almost diagnostic of cirrhosis. The most sensitive and specific laboratory finding suggestive of cirrhosis in the setting of chronic liver disease is a low platelet count (<150,000/μL) due to portal hypertension and hypersplenism.[1] Confirmatory imaging tests include ultrasound, CT, and magnetic resonance imaging. Findings consistent with cirrhosis include a nodular liver surface, a small liver with or without left or caudate lobe hypertrophy, splenomegaly, and especially the identification of intra-abdominal collaterals, which are indicative of portal hypertension. A liver biopsy is not required if these findings are present. Any patient with newly diagnosed cirrhosis should undergo upper endoscopy to investigate the presence and size of gastroesophageal varices.[2] β-Blockers should be initiated in patients with medium or large varices or in patients with high-risk small varices (i.e., those present in a Child class C patient or with red wale marks). β-Blockers are not useful in patients without varices and are associated with significant adverse events.[3] Studies have not shown diuretics to be useful in patients with compensated cirrhosis. In the absence of masses on CT scan or ultrasound, performing a magnetic resonance imaging study is unnecessary.

1. Poynard T, Bedossa P. Age and platelet count: a simple index for predicting the presence of histological lesions in patients with antibodies to hepatitis C virus. METAVIR and CLINIVIR Cooperative Study Groups. *J Viral Hepat.* 1997;4:199-208.

2. Garcia-Tsao G, Sanyal AJ, Grace ND, et al. Prevention and management of gastroesophageal varices and variceal hemorrhage in cirrhosis. *Hepatology.* 2007;46:922-938.

3. Groszmann RJ, Garcia-Tsao G, Bosch J, et al. Beta-blockers to prevent gastroesophageal varices in patients with cirrhosis. *N Engl J Med.* 2005;353:2254-2261.

2. A 69-year-old obese, diabetic man with biopsy-proven compensated cirrhosis and without a prior history of variceal hemorrhage has large esophageal varices on a screening endoscopy study. Which of the following medications would you recommend at this point?

A. No specific medication
B. Metoprolol
C. Nadolol
D. Spironolactone
E. Pantoprazole

Answer: C Nonselective β-blockers are the treatment of choice for the primary prevention of variceal bleeding.[1] In patients with large varices, the 2-year rate of first variceal hemorrhage is significantly lower in patients receiving nonselective β-blockers (15%) compared with those receiving no specific medication. β_1-Specific blockers (atenolol, metoprolol) are not as effective as nonselective (β_1 and β_2) β-blockers (nadolol or propranolol) or a nonselective β-blocker with vasodilating properties (carvedilol)[2,3] because the most important portal pressure–reducing effect results from β_2-blockade. Spironolactone is a diuretic that may reduce portal pressure but has not been shown to prevent first variceal hemorrhage. Although there appears to be a benefit in blocking gastric acid with proton pump inhibitors for the reduction of post-ligation ulcers and bleeding, there is no benefit in preventing a first variceal bleed.

1. D'Amico G, Pagliaro L, Bosch J. Pharmacological treatment of portal hypertension: an evidence-based approach. *Semin Liver Dis.* 1999;19:475-505.

2. Hillon P, Lebrec D, Munoz C, et al. Comparison of the effects of a cardioselective and a nonselective beta-blocker on portal hypertension in patients with cirrhosis. *Hepatology.* 1982;2:528-531.

3. Mills PR, Rae AP, Farah DA, et al. Comparison of three adrenoreceptor blocking agents in patients with cirrhosis and portal hypertension. *Gut.* 1984;25:73-78.

3. A 64-year-old obese man presents for evaluation of new ascites and peripheral edema. He is a heavy smoker and typically drinks two to four alcoholic beverages daily, but he has no history of liver disease. Physical examination reveals decreased breath sounds, moderate ascites, and peripheral edema but no other abnormalities. Laboratory tests reveal a total bilirubin 1.2 mg/dL, aspartate aminotransferase 37 IU/mL, alanine aminotransferase 35 IU/mL, albumin 3.7 g/dL, and international normalized ratio of 1.4. Paracentesis discloses clear yellow fluid with the following characteristics: total protein 3.5 g/dL, albumin 1.7 g/dL, white blood cells 640/μL (12% neutrophils). Which one of the following is the most appropriate action?

A. Begin intravenous cefotaxime and await ascites culture results
B. Perform a liver biopsy
C. Obtain serum B-type natriuretic peptide and perform cardiac echocardiography
D. Perform an abdominal CT scan to investigate peritoneal malignant disease
E. Begin therapy for tuberculosis

Answer: C This patient without documented chronic liver disease has ascites with a high serum-ascites albumin gradient (>1.1 g/dL), indicating ascites that is secondary to hepatic sinusoidal hypertension.[1] The fluid also has a high protein content (>2.5 g/dL) indicative of normal "leaky" sinusoids, so the problem is unlikely to be in the liver. A liver biopsy would not be the most indicated test. In this clinical setting, the most likely possibility is right-sided heart failure, which should be evaluated, initially with an echocardiogram. A high serum B-type natriuretic peptide has a high diagnostic accuracy in the diagnosis of ascites due to heart failure.[2] Although the ascites total white blood cell count is somewhat elevated, the fluid contains only 77 neutrophils (640 × 0.12), and therefore it is not consistent with bacterial peritonitis, which is based on more than 250 neutrophils/μL.[3] In peritoneal causes of ascites (carcinomatosis, tuberculosis), the ascites protein is high and the serum-to-ascites albumin gradient is low (<1/1 g/dL) because fluid formation does not result from sinusoidal hypertension.

1. Runyon BA, Montano AA, Akriviadis EA, et al. The serum-ascites albumin gradient is superior to the exudate-transudate concept in the differential diagnosis of ascites. *Ann Intern Med.* 1992;117:215-220.

2. Farias AQ, Silvestre OM, Garcia-Tsao G, et al. Serum B-type natriuretic peptide in the initial workup of patients with new onset ascites: a diagnostic accuracy study. *Hepatology.* 2014;59:1043-1051.

3. Rimola A, Garcia-Tsao G, Navasa M, et al. Diagnosis, treatment and prophylaxis of spontaneous bacterial peritonitis: a consensus document. International Ascites Club. *J Hepatol.* 2000;32:142-153.

4. A 53-year-old woman with cirrhosis secondary to primary biliary cirrhosis presents with 2 weeks of weight gain and an increase in abdominal girth. On physical examination, she has gained 8 pounds, and she has a modestly distended abdomen with flank dullness and shifting dullness. She does not have asterixis. Her laboratory tests include creatinine 0.8, blood urea nitrogen 24, Na 136, K 3.5. Analysis of her ascitic fluid is compatible with cirrhotic ascites, without any evidence of spontaneous bacterial peritonitis. An ultrasound examination shows no hepatic masses. Appropriate management at this point should be

A. Large-volume paracentesis
B. Salt and water restriction alone
C. Transjugular intrahepatic portosystemic shunt (TIPS)
D. Spironolactone
E. Intravenous furosemide

Answer: D The patient has the new onset of uncomplicated, nontense ascites. The mainstay of therapy is sodium restriction and diuretics. Water restriction is not indicated unless her serum sodium concentration is below 125 mEq/L.[1] The most effective diuretic is spironolactone, which should be initiated alone at a dose of 100 mg once daily in the morning, particularly in patients with modest ascites. Furosemide is not as effective as spironolactone[2] because sodium that is not reabsorbed in the loop of Henle is taken up at the distal and collecting tubules as a result of the hyperaldosteronism that is present in most patients with cirrhosis and ascites. Therefore, furosemide should not be used as the sole agent in the treatment of cirrhotic ascites and should not be used intravenously. Large-volume paracentesis is indicated to decrease discomfort in patients with tense ascites and is first-line therapy for patients with refractory ascites. TIPS placement is second-line therapy for patients with refractory ascites.[1]

1. Runyon BA. Introduction to the revised American Association for the Study of Liver Diseases Practice Guideline management of adult patients with ascites due to cirrhosis 2012. *Hepatology.* 2013;57:1651-1653.

2. Perez-Ayuso RM, Arroyo V, Planas R, et al. Randomized comparative study of efficacy of furosemide versus spironolactone in nonazotemic cirrhosis with ascites. Relationship between the diuretic response and the activity of the renin-aldosterone system. *Gastroenterology.* 1983;84:961-968.

5. A 48-year-old man with chronic hepatitis C virus infection, cirrhosis, and ascites presents with 2 days of confusion. He denies fever, chills, hematemesis, or melena. On physical examination, his temperature is 35.5° C, blood pressure 90/60, heart rate 100 beats per minute, respiratory rate 30/minute. He is confused and agitated, and asterixis is present. His abdomen has tense ascites but is nontender, and his stool is negative for blood. His laboratory tests include white blood cell 8.6, creatinine 2.1 (previously 0.8), and bilirubin 7.7 (previously 2.3). He is making only small amounts of urine, but his urinalysis is normal. Which of the following is the most appropriate test?

A. Diagnostic paracentesis and blood cultures
B. Large-volume paracentesis
C. Upper endoscopy
D. Brain CT
E. Renal ultrasound

Answer: A The presence of unexplained encephalopathy or deterioration in renal function in a patient with ascites should always raise the suspicion of spontaneous bacterial peritonitis (SBP), so therefore a diagnostic paracentesis should always be performed in this setting.[1] The ascites culture is negative in about 40% of patients with SBP, so ascites and blood cultures should be performed whenever SBP is suspected. In this patient, blood cultures are particularly important because he presents with evidence of systemic inflammatory response system (hypothermia, tachycardia, and tachypnea) and jaundice and therefore may be septic. Large-volume paracentesis in this setting could theoretically lead to further deterioration in renal function and should be avoided. There is no evidence of gastrointestinal bleeding, so endoscopy is not indicated. Brain scan is not an initial test in a cirrhotic patient with encephalopathy unless there is a history of trauma or other causes of encephalopathy are suspected. A renal ultrasound study is a low-yield test because obstruction is an unlikely cause of his decompensation.

1. Rimola A, Garcia-Tsao G, Navasa M, et al. Diagnosis, treatment and prophylaxis of spontaneous bacterial peritonitis: a consensus document. *J Hepatol.* 2000;32:142-153.

154

HEPATIC FAILURE AND LIVER TRANSPLANTATION

GREGORY T. EVERSON

In the United States, 130 programs perform about 6000 transplants per year, and about 17,000 patients are on waiting lists because recipients needing liver transplantation exceed the donor liver supply. The mortality rate while waiting on a list is 116 deaths per 1000 patient-years.

Since 1982, patient survival after liver transplantation has steadily increased by 20 to 30%, whether it is measured at 3 months, 1 year, 5 years, or 10 years, largely because of improvements within the first year after transplantation (Fig. 154-1). The positive shift in survival during the first 3 months after transplantation is related to improvements in surgical techniques and immediate postoperative care. By comparison, lack of further improvement beyond 3 months is because of long-term complications, such as recurrent hepatitis C (Chapter 149), recurrent autoimmune disease (Chapters 149 and 155), chronic allograft rejection, renal dysfunction (Chapter 130), hypertension (Chapter 67), and diabetes mellitus (Chapter 229). Further improvement in long-term survival after liver transplantation will require better management of these chronic complications.

GENERAL SELECTION CRITERIA

By far, the most common indication for liver transplantation is noncholestatic cirrhosis (Table 154-1). Livers may be donated from deceased donors (94%) or living donors (6%). Regardless of the type of transplantation, three fundamental questions must be addressed at the time of evaluation.[1]

1. Is liver transplantation indicated? The patient should have liver failure, complications of liver disease, or a metabolic condition that is best treated by liver transplantation and for which there are no alternative treatments.
2. Are there contraindications to transplantation (Table 154-2)? Comorbid conditions that could severely compromise graft or patient outcome must be identified.
3. Will the patient be able to tolerate and comply with immunosuppression and post-transplantation management? Inability to comply with the rigors of post-transplantation care and management could lead to graft loss and recipient death.

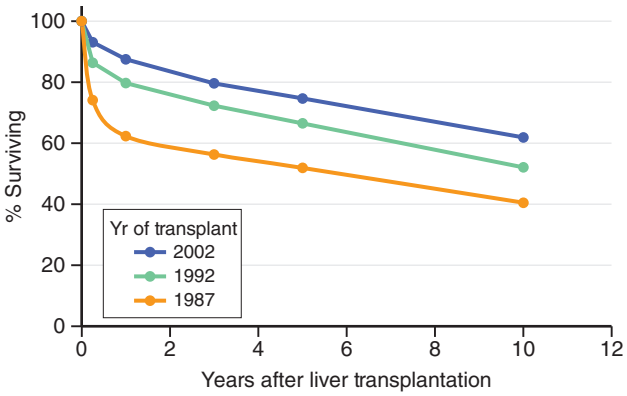

FIGURE 154-1. Survival of patients for three different years of transplantation, 1987, 1992, and 2002, by years after transplantation. The curves diverge early within the first year and then are nearly parallel, thereby suggesting that most of the improvement in survival is attributed to improvements in surgical techniques, intensive care unit management, and early post-transplantation care. Annual mortality rates between years 1 and 10 were 2.4% for the 1987 cohort, 3.1% for the 1992 cohort, and 2.8% for the 2002 cohort. (From Organ Procurement and Transplantation Network and Scientific Registry of Transplant Recipients. *OPTN/SRTR 2011 Annual Data Report*. Rockville, MD: Department of Health and Human Services, Health Resources and Services Administration, Healthcare Systems Bureau, Division of Transplantation; 2012).

TABLE 154-1 INDICATIONS FOR LIVER TRANSPLANTATION

CATEGORY	PERCENTAGE OF LISTED PATIENTS (TOTAL N = 15,748)	
Cirrhosis, non-cholestatic		72%
Hepatitis C	(27%)	
Alcoholic	(18%)	
Hepatitis C plus alcohol	(5%)	
Nonalcoholic steatohepatitis	(9%)	
Cryptogenic	(5%)	
Autoimmune	(4%)	
Hepatitis B	(3%)	
Other	(1%)	
Cholestatic liver disease		9%
Primary sclerosing cholangitis	(5%)	
Primary biliary cirrhosis	(3%)	
Other	(1%)	
Malignant neoplasms (primarily hepatocellular carcinoma)		7%
Acute liver failure		2%
Metabolic diseases		2%
Biliary atresia		1%
Other (e.g., Budd-Chiari syndrome) or unknown		8%

Based on Organ Procurement and Transplantation Network data as of May 30, 2014. http://optn.transplant.hrsa.gov/latestData/rptData.asp.

TABLE 154-2 CONTRAINDICATIONS TO LIVER TRANSPLANTATION

- Active substance abuse and inability to comply with rehabilitation
- Inadequate social support, extreme psychosocial dysfunction, active psychosis, or other underlying psychosocial disease that makes it impossible for the patient to comply with peri-transplantation care and postoperative management
- Unstable, active cardiopulmonary disease
  - Symptomatic ischemic coronary disease not amenable to revascularization
  - Severe pulmonary hypertension (mean PAP >45 mm Hg despite pharmacologic interventions)
- Active, incurable extrahepatic malignant disease
  - Metastatic nonhepatic malignant disease
  - Hepatoma with macrovascular invasion, extrahepatic metastases, or exceeding Milan or UCSF criteria and not able to be downstaged
  - Cholangiocarcinoma with percutaneous transperitoneal biopsy (likely seeding from the biopsy), tumor diameter >3 cm, or extrahepatic spread
- Active, uncontrolled, and untreatable sepsis or other serious infectious disease
- Active HIV infection with AIDS-defining illness, high-titer HIV RNA, very low CD4 count, or unresponsive to HAART
- Anatomic anomaly or extensive vascular thromboses precluding hepatic transplantation

PAP = pulmonary artery pressure; UCSF = University of California, San Francisco; HIV = human immunodeficiency virus; AIDS = acquired immunodeficiency syndrome; HAART = highly active antiretroviral therapy.

Medical Assessment

Certain components of the medical evaluation apply to all potential recipients. Viral serologies characterize the status of ongoing infection, prior exposure, or vaccination related to the hepatitis viruses (Chapters 148 and 149), Epstein-Barr virus (Chapter 377), cytomegalovirus (Chapter 376), and human immunodeficiency virus (HIV; Chapter 388). Colonoscopy is recommended for all patients older than 50 years to exclude colon cancer or polyps (Chapter 193) and in all patients with primary sclerosing cholangitis (Chapter 155) to detect inflammatory bowel disease (Chapter 141). Cardiopulmonary assessment (Chapters 51, 71, and 85) is performed in patients older than 50 years and in any patient with risk factors for coronary artery disease, such as hypertension, hypercholesterolemia, a significant family history, cigarette smoking, or diabetes. Cardiac catheterization (Chapter 57) is performed in patients with positive cardiac stress test results to confirm and to delineate the extent of coronary disease. Ultrasonography, computed tomography, and magnetic resonance imaging are useful to examine the biliary tract (Chapter 146), to exclude hepatocellular cancer (Chapter 196) or cholangiocarcinoma (Chapter 196), and to determine the patency of major vascular structures, such as the hepatic artery, celiac axis, splenic artery, portal vein, and superior mesenteric vein (Chapter 143).

The primary principle for consideration of liver transplantation is that a patient's predicted survival with transplantation must exceed expected survival without transplantation. For deceased donor liver transplantation, this principle is fulfilled by a Model for End-Stage Liver Disease (MELD) score of approximately 12 to 15 (see Table 153-2). In addition, certain complications of cirrhosis (Chapter 153), such as ascites, spontaneous bacterial peritonitis, encephalopathy, uncontrolled variceal hemorrhage, nutritional wasting, failure to thrive, and development of hepatoma (Chapter 196), shorten survival. Immediate evaluation and listing for transplantation are indicated for any patient with a MELD score of 15 or more or any of these complications.

Other complications that are not considered life-threatening may also be indications for consideration of liver transplantation. Examples include metabolic bone disease resulting in osteopenia and bone fractures (Chapters 243 and 244), inadequate nutrition, muscle wasting, severe fatigue, poor concentrating ability, and intractable pruritus.

Because the supply of donor livers is limited, transplant centers must select patients who have the greatest chance for a successful outcome. Elderly, obese, and deconditioned patients as well as patients with underlying vascular disease or long-standing diabetes mellitus are poor candidates for liver transplantation.

Surgical Assessment

The three phases of surgery are native liver dissection, the anhepatic phase, and revascularization of the graft. Native liver dissection, which is characterized by meticulous dissection and prompt control of bleeding vessels, can be complicated by morbid obesity, significant portal hypertension, portal thromboses, prior portosystemic shunt surgery, or prior abdominal surgery. The length of this phase is usually 1 to 2 hours, with blood losses ranging from 0 to 5 units of blood.

During the anhepatic phase, which usually lasts 1.5 to 3 hours and is associated with blood loss ranging from 0 to 5 units of blood, the vascular supply of the liver is completely interrupted, and the native liver is excised. Toward the end of the anhepatic phase, vessels are anastomosed, with the donor hepatic veins typically grafted to the recipient vena cava with only one caval anastomosis. Recipients with severe coagulopathy, cachexia, and nutritional deficiencies may be prone to excessive bleeding and metabolic complications.

When the venous clamps are removed, patients are at risk for primary fibrinolysis or consumptive coagulopathy. The arterial anastomosis is typically performed after unclamping to shorten the warm ischemia period.

Psychosocial Assessment

A patient referred for liver transplantation may have a history of alcohol, drug, or substance abuse. Psychosocial evaluation is important to help determine

the risk of recidivism after transplantation. Today, nearly all transplant centers require a minimum 6-month period of documented abstinence and enrollment in education or rehabilitation programs. Social workers and psychiatrists may require the patient to undergo alcohol or drug treatment, counseling, and unannounced random screens for alcohol and drugs. Other key aspects of the psychosocial assessment include (1) evaluating the patient for any underlying psychiatric illness (Chapter 397), its severity, and how it should be managed; (2) defining and establishing social and psychological support for the potential recipient, including family, friends, and significant others; and (3) in living donor liver transplantation, evaluating the potential living donors.

Assigning Priority: Model for End-Stage Liver Disease

The prioritization of U.S. liver transplantation candidates and the allocation of deceased donor livers are based on the MELD score (see Table 153-2), which ranges from 6 (best prognosis) to 40 (worst prognosis). The 3-month mortality rate increases from 10% with a MELD score of 20 to 60% at a MELD score of 35, and essentially to 100% with a MELD score above 40. A patient's survival is not improved by deceased donor liver transplantation if the MELD score is below about 15. As a result, deceased donor livers must be made available more broadly if no local patient has a MELD score of 15 or higher and must be made available to local or regional patients with scores of 35 or higher. Patients with acute liver failure or post-transplantation patients with graft failure due to hepatic artery thrombosis have a MELD score of 40 and top priority status. This allocation system (E-Table 154-1) has improved transplantation rates and has not altered waiting list or post-transplantation mortality.

In patients with cirrhosis (Chapter 153), hyponatremia (Chapter 116) is associated with hepatorenal syndrome, ascites, and death from liver disease. Hyponatremia is an independent predictor of mortality up to a MELD score of 30. As a result, the MELD score is now adjusted for serum sodium concentration in allocating deceased donor livers.

The most common exception to liver allocation by MELD score is for hepatocellular carcinoma (Chapter 196) with a tumor diameter of more than 2 cm. Other exceptions can include patients with the hepatorenal syndrome (Chapter 153), the portopulmonary syndrome (Chapter 153), cholangiocarcinoma (Chapter 155), cystic fibrosis (Chapters 89 and 146), familial amyloid polyneuropathy (Chapter 188), and primary hyperoxaluria (Chapter 205) (E-Table 154-2).

DISEASE-SPECIFIC INDICATIONS FOR TRANSPLANTATION

Acute Liver Failure

Acute liver failure is defined as hepatic injury of fewer than 26 weeks in duration, an international normalized ratio of 1.5 or greater, and altered mental status in the absence of chronic liver disease except for Wilson disease (Chapter 211), vertically acquired hepatitis B, or autoimmune hepatitis (Chapter 149).[2] The main causes of acute liver failure in the United States are acetaminophen toxicity (Chapter 110), drug-induced liver injury (Chapter 150), viral hepatitis (Chapters 148 and 149), autoimmune hepatitis (Chapter 149), Wilson disease, mushroom poisoning (Chapters 110 and 150), acute hepatic ischemia (Chapter 143), Budd-Chiari syndrome, acute fatty liver of pregnancy (sometimes associated with the HELLP syndrome—hemolysis, elevated liver enzymes, low platelets; Chapters 160 and 239), and malignant infiltration of the liver (Chapter 196).

TREATMENT

Rx

Patients with acute liver failure criteria should be transferred to the intensive care unit for enhanced monitoring, and a liver transplant center should be contacted for potential transfer of the patient. Specific treatments include *N*-acetylcysteine for acetaminophen hepatotoxicity (Chapters 110 and 150), *N*-acetylcysteine plus either silibinin or penicillin G for mushroom poisoning (Chapters 110 and 150), *N*-acetylcysteine and removal of the offending drug for drug-induced liver injury (Chapter 150), nucleos(t)ide analogues for hepatitis B (Chapters 148 and 149), acyclovir for herpes or varicella-zoster hepatitis (Chapters 374 and 375), dialysis and copper chelation for Wilson disease (Chapter 211), corticosteroids for autoimmune hepatitis (Chapter 149), delivery of the infant for acute fatty liver of pregnancy, and transjugular intrahepatic portosystemic shunting for the Budd-Chiari syndrome (Chapter 143). In addition, all causes of acute liver failure can cause a range of systemic complications.

A key issue in the general management of patients with acute liver failure is monitoring and treatment of encephalopathy. Progressive encephalopathy in acute liver failure is characteristically associated with cerebral edema and an elevated intracranial pressure (ICP). Elevated ICP is rare in patients with grade I (mild confusion) or grade II (agitated state) encephalopathy but occurs in 25 to 35% of patients with grade III encephalopathy (stuporous) and 65 to 75% of patients with grade IV (coma, but responsive to deep pain) encephalopathy. Irreversible brain injury is likely when the ICP is above 50 mm Hg and the cerebral perfusion pressure (mean arterial pressure minus ICP) is below 40 mm Hg for more than 2 hours. Liver transplantation may be contraindicated under these circumstances.

Because ammonia may play a role in the pathogenesis of cerebral edema, lactulose (20 g every 4 to 6 hours), rifaximin (550 mg twice daily), or both are recommended, especially for lower grades of encephalopathy. Patients with grade III/IV encephalopathy require intubation, mechanical ventilation, elevation of the head of the bed, sedation (e.g., propofol, initial dose of 0.005 mg/kg/minute IV, maintenance dose of 0.005 to 0.05 mg/kg/minute IV, which can be increased in increments of 0.005 mg/kg/minute every 5 minutes), and paralysis (cisatracurium, initial dose of 0.1 to 0.2 mg/kg IV, maintenance dose of 1 to 3 µg/kg/minute IV). Seizures (Chapter 403) should be treated with phenytoin (initial dose of 15 mg/kg IV, maintenance dose of 3 mg/kg every 12 hours IV), propofol (1 mg/kg IV loading dose, 3 to 7 mg/kg/hour IV maintenance), pentobarbital (13 mg/kg IV loading dose, 2 to 3 mg/kg/hour IV maintenance), or midazolam (0.2 mg/kg IV loading dose, 0.1 to 0.25 mg/kg/hour IV maintenance). Levetiracetam (500 [up to 1500] mg every 12 hours IV) may be used if renal function is preserved. Placement of an intracranial transducer for monitoring of ICP is desirable, but severe underlying coagulopathy may prohibit its placement. Elevated ICP may require treatment with mannitol (0.5 to 1.0 mg/kg every 4 to 6 hours IV), hypertonic saline (boluses every 2 to 4 hours of either 30 mL of 23.4% saline or 2 mL/kg of 7.5% saline), hyperventilation (target pH 7.45), pentobarbital (dose described before), and hypothermia (32° C to 34° C). The goal for these interventions is to maintain ICP below 20 mm Hg. Corticosteroids do not improve cerebral edema or lower ICP in acute liver failure,[3] and their use is not recommended.

Patients with acute liver failure are at increased risk for bacterial and fungal infections. Common infections include line sepsis, pneumonia (Chapter 97), and urinary tract infections (Chapter 284). Prolonged ventilation, dialysis, invasive procedures, and use of multiple antibiotics increase the risk for fungal infection (Chapter 282). Surveillance cultures of blood, urine, and sputum and periodic chest radiography are required every 2 days. Prophylactic antibiotic or antifungal therapy is not currently recommended.

In addition to an elevated international normalized ratio, many patients develop thrombocytopenia because of consumptive coagulopathy (Chapter 175). However, prophylactic use of transfusion of clotting factors or platelets should be restricted to the treatment of hemorrhage or in preparation for invasive procedures.

Patients with acute liver failure have a low systemic vascular resistance and a low mean arterial pressure. In patients with elevated ICP, the decrease in mean arterial pressure may further compromise cerebral perfusion pressure and blood flow. A goal of vasopressor therapy is to maintain cerebral perfusion pressure from 60 to 80 mm Hg. Treatment of hypotension to maintain a mean arterial pressure above 75 mm Hg may include dextrose-containing crystalloid, albumin, norepinephrine, and vasopressin (Chapter 106). Careful monitoring of volume status is critical, and care should be taken to avoid volume overload, which may worsen cerebral edema.

Acute renal failure (Chapter 120) is common in patients with acute liver failure, particularly when it is caused by acetaminophen, mushroom poisoning, or Wilson disease. Renal replacement therapy, when required, should be administered as continuous venovenous dialysis (Chapter 131). Intermittent modes of hemodialysis should be avoided because of their adverse hemodynamic effects.

In acute liver failure, hypoglycemia is a common complication that should be avoided and managed by continuous glucose infusions. Supplementation with phosphate, magnesium, and potassium (Chapter 117) may be required. Enteral feedings (Chapter 216) should be initiated early, and parenteral nutrition (Chapter 217) should be started if enteral feedings are contraindicated.

Selecting for Liver Transplantation

Clinical variables associated with reduced likelihood for spontaneous recovery include severe encephalopathy, advanced age, certain causes (hepatitis B, drug-induced hepatitis, sporadic non-A/non-B/non-C hepatitis, Wilson disease), long duration of jaundice, massive necrosis on liver biopsy, and markedly diminished liver volume on radiologic imaging. Transplant-free survival in patients with acute liver failure is approximately 50% for acetaminophen toxicity, hepatitis A, shock liver, and acute fatty liver of pregnancy/HELLP but less than 25% for all other causes.[4] Indications for listing for transplantation are based on the severity of disease but vary somewhat among countries (Table 154-3).[5]

E-TABLE 154-1 THE SEQUENCE FOR ALLOCATION OF ADULT DONOR LIVERS FOR DECEASED DONOR LIVER TRANSPLANTATION IN THE U.S.

COMBINED LOCAL AND REGIONAL
1. Status 1A candidates in descending point order
2. Status 1B candidates in descending order
LOCAL AND REGIONAL
3. Candidates with MELD score 35 or higher in descending order of MELD score, with local candidates ranked above regional candidates at each level of MELD score
LOCAL
4. Candidates with MELD score 29 to 34 in descending order of MELD score
NATIONAL
5. Liver-intestine candidates in descending order of status and mortality risk scores
LOCAL
6. Candidates with MELD score 15 to 28 in descending order of MELD score
REGIONAL
7. Candidates with MELD score 15 to 34 in descending order of MELD score
NATIONAL
8. Status 1A candidates in descending order of MELD score
9. Status 1B candidates in descending order of MELD score
10. Candidates with MELD score 15 or higher in descending order of MELD score
LOCAL
11. Candidates with MELD score 15 or less in descending order of MELD score
REGIONAL
12. Candidates with MELD score 15 or less in descending order of MELD score
NATIONAL
13. Candidates with MELD score 15 or less in descending order of MELD score

MELD = Model for End-Stage Liver Disease (see Table 153-2).

E-TABLE 154-2 LIVER CANDIDATES WITH EXCEPTIONAL CASES NOT REQUIRING REVIEW BY THE REGIONAL REVIEW BOARD IN THE U.S.

Hepatopulmonary syndrome. The patient must have evidence of intrapulmonary shunt and Pao_2 below 60 mm Hg on room air. These patients are listed with MELD 22 with 10% increase every 3 months if the Pao_2 remains under 60 mm Hg.

Cholangiocarcinoma. The liver center requesting the MELD exception must have a written United Network for Organ Sharing–approved protocol that includes neoadjuvant chemoradiation and pretransplantation operative staging to exclude metastatic disease. MELD 22 with 10% increase every 3 months as long as tumor is controlled by the treatment protocol and there is no evidence of metastases.

Cystic fibrosis. The patient must have FEV_1 below 40% of normal range. MELD 22 with 10% increase every 3 months.

Familial amyloid polyneuropathy. The patient must lack significant cardiac involvement with echocardiogram showing ejection fraction greater than 40% and have ambulatory status, appropriate *TTR* gene mutation, and biopsy-proven amyloid. MELD 22 with 10% increase every 3 months.

Primary hyperoxaluria (type I). The patient must have renal failure due to oxalate stone disease and alanine:glyoxylate aminotransferase deficiency proven by liver biopsy (sample analysis with measurement of alanine:glyoxylate aminotransferase enzyme activity and/or genetic analysis of the alanine:glyoxylate aminotransferase gene). MELD 28 with 10% increase every 3 months.

Portopulmonary syndrome. The patient must have pulmonary hypertension diagnosed by appropriate measurements of mean pulmonary artery pressure (MPAP), pulmonary vascular resistance (PVR), and transpulmonary gradient. With treatment, the MPAP should be less than 35 mm Hg and PVR should be less than 400 dynes/sec/cm^{-5}. MELD 22 with 10% increase every 3 months.

MELD = Model for End-Stage Liver Disease.

TABLE 154-3 CRITERIA FOR SELECTION OF PATIENTS WITH ACUTE LIVER FAILURE FOR LIVER TRANSPLANTATION

KING'S COLLEGE, LONDON, UNITED KINGDOM
Acetaminophen toxicity
Acidosis (pH <7.3), or
INR >6.5 plus creatinine >3.4 mg/dL
Other causes of acute liver failure
INR >6.5, or
Any three of the following:
Age <10 years or >40 years
Non-A, non-B hepatitis or drug-induced disease
Duration of jaundice before encephalopathy >7 days
INR >3.5
Bilirubin >17.5 mg/dL
HÔPITAL PAUL-BROUSSE, VILLEJUIF, FRANCE
For all causes of acute liver failure:
Hepatic encephalopathy, and
Factor V <20% of normal in patient younger than 30 years or
Factor V <30% of normal in patient 30 years of age or older

INR = international normalized ratio.

Transplantation for acute liver failure accounts for less than 5% of all liver transplants. About 45% of patients with acute liver failure are listed for transplantation; about 10% die on the waiting list, about 5% improve without transplantation, and about 30% undergo liver transplantation. Overall 1- and 5-year survival rates after liver transplantation are 79% and 71%, respectively, but post-transplantation survival is better for patients with acute acetaminophen overdose and worse for patients with severe encephalopathy before transplantation. Living donor and auxiliary liver transplantation may be considered, but their use is controversial.

Chronic Liver Failure

Cirrhosis (Chapter 153) due to the hepatitis C virus (Chapter 149) is the most common indication for liver transplantation in the United States. Hepatitis C universally recurs in the liver allograft of recipients who are viremic at the time of liver transplantation and is associated with early graft loss and death of the patient. Pretransplantation treatment (Chapter 149) with an interferon-free, pangenotypic regimen of sofosbuvir (400 mg daily) plus weight-based ribavirin (1.0 to 1.2 g/day) for 48 weeks eliminates detectable hepatitis C virus RNA in nearly all patients within 4 weeks, and patients with undetectable hepatitis C virus RNA for 30 days or more before transplantation have only about a 4% risk of becoming viremic after transplantation.[6] A similar regimen given for 24 weeks after transplantation can achieve a sustained virologic response in about 70% of recipients with hepatitis C virus genotype I infection.[7] The combination of ledipasvir, sofosbuvir, and ribavirin yields sustained virologic response rates of approximately 90% in patients with decompensated cirrhosis and in transplant recipients with recurrent hepatitis C and advanced fibrosis or cirrhosis.

Alcoholic cirrhosis (Chapter 152) is a major albeit controversial indication for liver transplantation because many patients who totally abstain from alcohol may recover to the point that transplantation is no longer indicated.[8] Candidates for transplantation must typically adhere to at least 6 months of rehabilitation and document abstinence from alcohol by testing of urine or blood. Risk factors for noncompliance and return to alcohol use include polysubstance abuse, poor social support, joblessness, and underlying psychiatric illness. However, properly selected and compliant patients have excellent post-transplantation outcomes, even though 8 to 33% return to some degree of drinking. The 1-, 3-, and 5-year patient survival rates range from 81 to 92%, 78 to 86%, and 73 to 86%, respectively.

Transplantation for alcoholic hepatitis is even more controversial. One nonrandomized study from seven French liver transplant units suggested that carefully selected patients with severe alcoholic hepatitis could substantially benefit from liver transplantation, with a 2-year survival of 77% with transplantation compared with 23% without it.[9] It is not known whether this experience can be duplicated in other centers or countries, and alcoholic hepatitis currently is considered a relative contraindication to liver transplantation in most U.S. centers.

Patients with hepatocellular carcinoma (Chapter 196) are candidates for liver transplantation unless they have large tumors, multicentric tumors, macrovascular invasion, or extrahepatic spread. The Milan criteria recommend transplantation for a single hepatocellular carcinoma less than 5 cm in diameter or up to three hepatocellular carcinomas with no single lesion more than 3 cm in diameter. The post-transplantation survival of patients satisfying these Milan criteria is similar to that of patients transplanted for other indications. The University of California, San Francisco (UCSF) criteria recommend transplantation for a single tumor less than 6.5 cm in diameter or up to five tumors with total cumulative diameter of less than 8.5 cm. Patients who are "downstaged" by transarterial chemoembolization or other locoregional treatments from UCSF to Milan criteria have a post-transplantation outcome similar to that of patients who initially met Milan criteria.

The rare fibrolamellar variant of hepatocellular carcinoma constitutes only 0.85% of all cases of primary liver cancer. Compared with patients with typical hepatocellular carcinoma, patients with fibrolamellar carcinoma are younger (mean age, 39 years vs. 65 years) and more likely to be female (52% vs. 26%) and white (85% vs. 57%). Fibrolamellar carcinoma is characterized by broad bands of fibrosis, no association with cirrhosis, slow growth, and easier resection. Although post-transplantation recurrence rates of fibrolamellar carcinoma are similar to those for standard hepatocellular carcinoma, 5-year survival is higher in patients with fibrolamellar carcinoma.

Cholangiocarcinoma (Chapter 196), in the absence of underlying liver or biliary disease, typically is manifested in elderly patients with significant comorbidities and is not an indication for liver transplantation. By contrast, cholangiocarcinoma arising in younger patients with underlying biliary disease, such as primary sclerosing cholangitis (Chapter 155), may be considered for transplantation if staging is negative for vascular, lymphatic, or neural invasion. The 2- and 5-year post-transplantation survivals of patients with cholangiocarcinoma limited to the perihilar region of the liver and treated with adjuvant chemoradiation are 78% and 65%, respectively.[10] Factors associated with early recurrence and diminished survival are tumor diameter of more than 3 cm, performance of a percutaneous transperitoneal biopsy of the tumor, metastatic disease at time of surgery, and history of a prior malignant neoplasm.

Rare Primary Hepatic Malignant Neoplasms

Other primary hepatic malignant neoplasms—epithelioid hemangioendotheliomas, hemangiosarcomas, and hepatoblastomas—represent less than 10% of all tumors undergoing transplantation. Epithelioid hemangioendothelioma is generally thought to be indolent in nature, and reported post-transplantation survivals have been favorable. By contrast, hemangiosarcoma universally recurs, so transplantation is contraindicated.

In patients with liver metastases from primary endocrine tumors, such as carcinoid tumor (Chapter 232), gastrinoma, insulinoma, glucagonoma, and VIPoma (Chapters 195 and 230), liver transplantation is best restricted to those with metastases that are confined to the liver and are unresponsive to chemotherapy, local ablative therapies, surgical resection, and hormonal therapy.[11] Even in highly selected patients, tumor recurrence is the rule, with a 5-year post-transplantation patient survival of 52% but a disease-free 5-year survival of only 30%.

With the advent of effective antiviral regimens, the post-transplantation outcomes for patients transplanted for hepatitis B (Chapter 149) are now among the best for any indication,[12] with a 3-year survival of 87%. In the United States, potential recipients are treated before transplantation with either tenofovir or entecavir monotherapy, which is continued indefinitely in the post-transplantation period.

Nonalcoholic steatohepatitis (Chapter 152) has become an increasing indication for liver transplantation.[13] The likelihood of 1-, 3-, and 5-year survival is similar to that of other recipients because their higher risk for post-transplantation cardiovascular death is offset by a lower risk of graft failure.

For primary biliary cirrhosis (Chapter 153), the mean survival for patients with a serum bilirubin level above 10 mg/dL is only 1.4 years. By comparison, actuarial survival for patients undergoing liver transplantation is 83%, 78%, and 67% at 1 year, 5 years, and 10 years, respectively.[14] Although primary biliary cirrhosis may recur in the allograft, recurrence rarely leads to graft failure, retransplantation, or the patient's death.

Among patients with primary sclerosing cholangitis (Chapter 155), the 1-, 2-, 5-, and 10-year actuarial survival rates after liver transplantation are 94%, 92%, 86%, and 70%, respectively. Primary sclerosing cholangitis recurs after transplantation at a rate of approximately 4% per year, and recurrent primary sclerosing cholangitis occasionally may progress and require retransplantation.

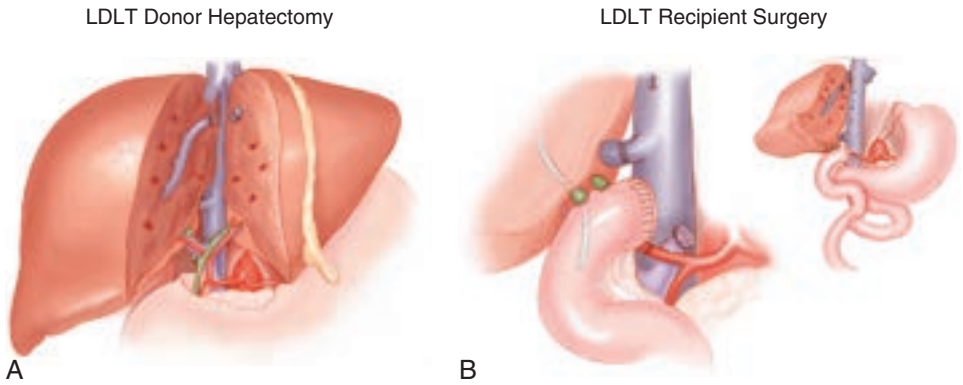

FIGURE 154-2. Living related donor hepatectomy (A) and orthotopic implantation of the right lobe graft (B). LDLT = living donor liver transplantation.

Up to 30% of patients with primary sclerosing cholangitis will develop cholangiocarcinoma, which diffusely infiltrates bile ducts, liver parenchyma, neural elements, lymphatic vessels, and surrounding tissues. Cholangiocarcinoma complicating primary sclerosing cholangitis, either before or after transplantation, reduces survival,[15] although the outcome is excellent if the cholangiocarcinoma is detected during surveillance imaging, treated with adjuvant chemoradiation, and then removed with the explant during transplantation.

In patients with autoimmune hepatitis (Chapter 149), 1-year patient survival rates after liver transplantation range from 83 to 92%, with an estimated 10-year survival of 75%. Recurrence is observed in 17% of patients after 4.6 ± 1 years, and an occasional patient with recurrent autoimmune hepatitis will require retransplantation.

For hemochromatosis (Chapter 212), liver transplantation now yields 1-, 3-, and 5-year survival rates similar to the survivals of all other transplant recipients. Patients with hemochromatosis who are considered for liver transplantation should undergo screening for hepatocellular carcinoma and also have a complete cardiologic evaluation. It is uncertain whether pretransplantation phlebotomy improves survival.

Patients with either the ZZ (n = 50) or SZ (n = 23) α_1-antitrypsin deficiency (Chapter 146) have excellent outcomes after liver transplantation.[16] Pulmonary evaluation is critical because one third of adults transplanted for α_1-antitrypsin deficiency will have significant underlying obstructive respiratory disease, which typically stabilizes and then improves after liver transplantation but sometimes may progress despite it.

In patients with polycystic liver disease, liver transplantation is rarely indicated but is highly successful in relieving symptoms of abdominal fullness, early satiety, and related complaints. The post-transplantation outcome is better for patients with isolated polycystic liver disease compared with polycystic liver in the setting of polycystic kidney disease (Chapter 127),[17] thereby highlighting the higher morbidity and mortality associated with cystic renal disease.

In patients with primary hyperoxaluria,[18] which is inherited as an autosomal dominant trait, liver transplantation is required to remove the source for the overproduction of oxalate and to halt the progression of renal disease. For patients with familial amyloid polyneuropathy, which is an inherited and fatal systemic amyloidosis that is caused by a point mutation in the transthyretin gene, liver transplantation halts the production of the amyloidogenic variant transthyretin, halts the progression of the disease, and significantly improves survival.[19]

An increasing number of HIV-infected patients are being considered for liver transplantation, primarily because of coinfection with hepatitis C virus.[20] Selection criteria include the absence of AIDS or AIDS-defining illness, adequate CD4 counts, low HIV levels, and lack of resistance to highly active antiretroviral therapies. The overall patient survival at 3 years after transplantation is 60% for HIV/hepatitis C virus coinfection, compared with 79% for hepatitis C virus monoinfection.

RISK FACTORS FOR POOR OUTCOME AFTER LIVER TRANSPLANTATION

Patients older than 60 years have about a 10% lower survival because of infection, cardiac complications, neurologic disease, and malignant disease. Survival is also lower in patients with ischemic heart disease, diabetes, persistent smoking, or chronic obstructive pulmonary disease, although patients who undergo liver transplantation for α_1-antitrypsin deficiency may show improvement in their pulmonary function test results.

LIVING DONOR LIVER TRANSPLANTATION IN ADULTS

Living donor liver transplantation is an option to expedite transplantation for adult patients with end-stage liver disease. Living donor liver transplantation already is the most common form of liver transplantation in Asia, where cultural barriers have limited liver donation from deceased donors, but it represents only 5% of all adult liver transplants in the United States. In adults, either the left or right lobe may be used, but most U.S. experts prefer the right hepatic lobe. In this procedure, the entire native liver of the recipient is removed and replaced with the right lobe of the liver from a living donor (Fig. 154-2).

The major clinical advantage of living donor liver transplantation is a reduction in recipient mortality, largely related to a reduction in pretransplantation mortality among patients who otherwise would languish on a waiting list. By comparison, both short-term and long-term survival after transplantation is nearly identical with living donor liver transplantation and deceased donor liver transplantation.[21]

The major disadvantage of living donor liver transplantation is donor safety. Most donors undergo successful right hepatectomy uneventfully, but significant complications, such as bile leakage and infection, occur in 10 to 20% of patients, and the death rate among living donors is 0.2%.[22] In the majority of donors, the quality of life returns to baseline by 6 months, although mild abdominal complaints and pain are common. More than 90% of donor-recipient pairs report that they have returned to their original predonation relationships at 1 year.

RECIPIENT OUTCOMES AFTER LIVER TRANSPLANTATION

Recipients of either deceased donor liver transplantation or living donor liver transplantation may encounter a number of complications that occur at different times after transplantation (Fig. 154-3).[23] Graft injury within the first 3 days is most often due to either primary nonfunction or hepatic artery thrombosis. Less common causes of graft dysfunction during this period include hyperacute rejection, portal vein thrombosis, and obstruction of the inferior vena cava. Between 3 and 14 days, graft dysfunction is most commonly related to acute cellular rejection, recurrent hepatitis C infection, hepatic artery thrombosis, or biliary leak or cholangitis. Infrequent causes of dysfunction during this period are portal vein thrombosis, drug hepatotoxicity, and functional cholestasis. From 14 days to 3 months, the most common causes of graft dysfunction include allograft rejection, recurrent hepatitis C infection, biliary complications, cytomegalovirus hepatitis, and drug hepatotoxicity. Vascular thromboses rarely are manifested after the first 3 months, and hepatitis B recurrence (in untreated patients) is typically delayed beyond 1 month.

The most severe form of early graft dysfunction is primary nonfunction, which is characterized by acute liver failure (encephalopathy, ascites, coagulopathy, unstable hemodynamics), elevated liver enzymes, and development of multiorgan failure (renal failure and pulmonary complications). Primary nonfunction is encountered in only 1% of recipients after living donor liver transplantation, because the graft is implanted quickly after it has been removed from the donor, but in 7 to 8.5% of recipients after deceased donor liver transplantation. Anywhere from 20 to 45% of recipients experience

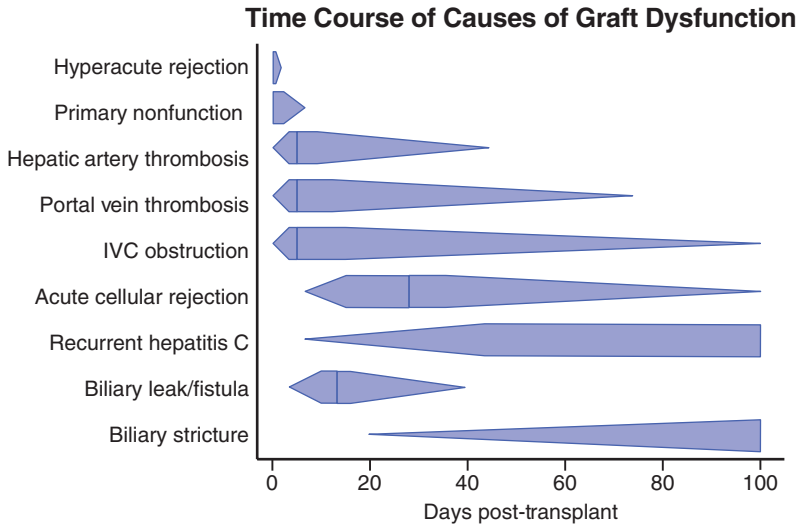

FIGURE 154-3. The time course of graft dysfunction varies by etiology. Hyperacute rejection is rare (ABO incompatibility) but occurs within the first few hours to days. Primary nonfunction, vascular thrombosis, and biliary leaks are early events, and acute rejection episodes begin after the first 7 days. Recurrent hepatitis C may begin early but typically is manifested after the first 2 or 3 weeks after transplantation. Biliary strictures evolve more slowly and tend to be later events. IVC = inferior vena cava.

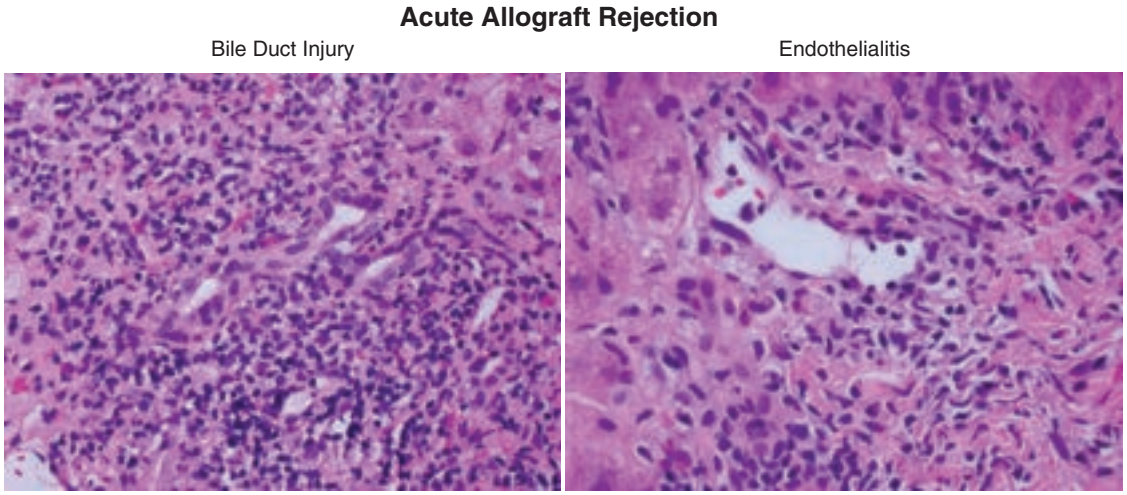

FIGURE 154-4. Liver histology of acute allograft rejection (hematoxylin and eosin stain). The immune inflammatory response is centered around the portal triad, and the typical features include a mixed cellular infiltrate (eosinophils, neutrophils, plasma cells, lymphocytes), lymphocytic cholangitis, and endothelialitis.

postoperative bleeding, including 5 to 15% who will need reoperation for bleeding. Hemorrhage is second only to infection as a cause of death.

Hepatic artery thrombosis, which complicates 3 to 10% of adult liver transplants, is one of the leading causes of graft failure in the immediate postoperative period. It is manifested as fever, bacteremia from a biliary source, and sudden elevations in liver enzymes. The preferred initial investigation of the hepatic artery is Doppler ultrasonography, followed by angiography if positive. In general, emergent revascularization is required for graft survival.

Portal Vein Thrombosis

Portal venous thrombosis, which is less common than hepatic arterial thrombosis, is manifested as hepatic failure and portal hypertension. Portal hypertension may be treated by radiologic thrombectomy, lytic therapy, stenting of underlying portal vein stenosis, or decompressive shunt surgery, but patients with significant hepatic dysfunction should be considered for revascularization or retransplantation.

Hepatic venous outflow obstruction may be manifested clinically by hepatic dysfunction, coagulopathy, and jaundice or even by hepatomegaly and ascites. Chronic outflow obstruction with preservation of graft function may be managed conservatively with diuretic therapy, radiologic evaluation of venous anastomoses, and angioplasty or stenting of outflow stenosis, but acute outflow obstruction can be a graft- and life-threatening condition that may require emergent revision of the outflow anastomosis.

Biliary tract complications, leaks, and strictures occur after approximately 5 to 30% of all liver transplants, usually within the first 3 months. Biliary scintigraphy can be diagnostic for large bile leaks, but cholangiography may be required for small, contained leaks. Anastomotic biliary strictures are focal and localized to the choledochocholedochostomy or choledochoenterostomy, whereas ischemic strictures are multiple and diffuse. Patients may present with jaundice, cholangitis, or asymptomatic elevations in liver test results. Choledocholithiasis may complicate strictures. Cholangiography is the "gold standard" for the diagnosis of biliary strictures as well as for their management with dilation and stenting. Surgical revision is reserved for patients in whom these procedures are unsuccessful, but diffuse ischemia strictures may necessitate retransplantation.

Rejection, which occurs in about 15 to 30% of liver transplants, usually is initially seen as an acute rejection within the first 30 days. Any acute rejection occurring beyond 30 days should raise the suspicion of subtherapeutic levels of immunosuppressive medications, drug interactions affecting immunosuppressant levels, or noncompliance of the patient with the medical regimen. Most patients are asymptomatic, although fever, malaise, abdominal pain, and worsening of portal hypertension with associated clinical manifestations can occur. Liver biopsy is essential for diagnosis. The classic findings are a mixed inflammatory infiltrate of lymphocytes, plasma cells, eosinophils, and neutrophils and nonsuppurative destructive cholangitis and endothelialitis (Fig. 154-4). Treatment consists of pulse doses of methylprednisolone (1 g IV daily for 1 to 3 days) and a taper of oral prednisone. Patients who fail to respond to corticosteroid treatment may be rescued by anti–T-cell therapy, such as thymoglobulin. In some cases in which the acute rejection may be mediated by B cells, plasmapheresis and anti–B-cell treatment (e.g., rituximab, 100 to 375 mg/m^2 weekly for two or three doses) may be required. Graft loss due to acute rejection is rare.

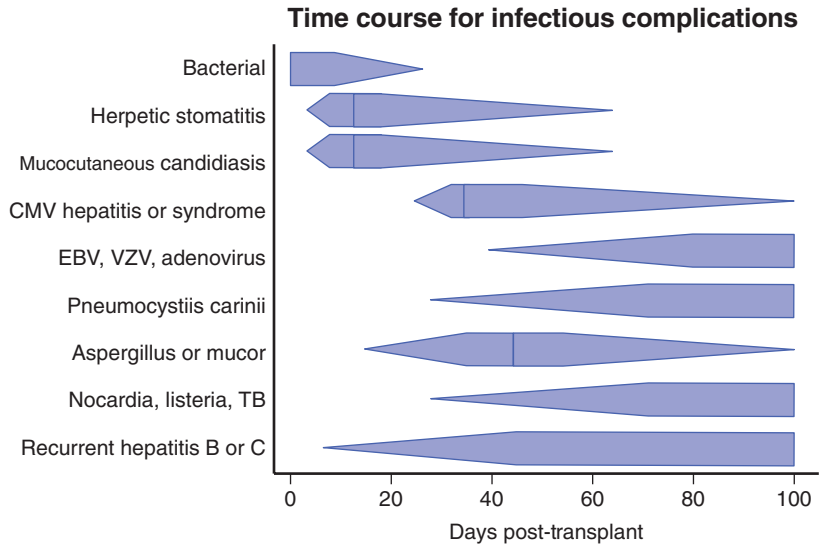

FIGURE 154-5. The time course of infectious complications varies by type of infection. Bacterial and mucocutaneous viral or candidal infections occur early, and most of the others occur after the first 2 or 3 weeks after transplantation. CMV = cytomegalovirus; EBV = Epstein-Barr virus; VZV = varicella-zoster virus; TB = tuberculosis.

Chronic allograft rejection occurs in 2 to 3% of liver transplants. Patients may be asymptomatic early in the course or present with jaundice or pruritus later in the course. Laboratory test results reveal elevated cholestatic liver enzymes, such as alkaline phosphatase or γ-glutamyltransferase, and bilirubin. Liver biopsy reveals paucity of inflammation, portal fibrosis, and loss of intralobular and interlobular bile ducts. Chronic rejection is poorly responsive to immunosuppressive therapy and often progresses to graft loss, need for retransplantation, and ultimately death.

Infections

Most infectious complications of liver transplantation occur within the first 3 months of transplantation (Fig. 154-5).[24] Bacterial and mucocutaneous herpetic (Chapter 374) or candidal (Chapter 338) infections dominate the immediate post-transplantation period. Other viral infections (such as cytomegalovirus [Chapter 376], Epstein-Barr virus [Chapter 377], varicella-zoster [Chapter 375], and adenovirus [Chapter 365]), *Pneumocystis jiroveci* infection (Chapter 341), or fungal infections tend to occur or to activate after the first 30 days. Hepatitis C replication begins immediately after transplantation in patients who are viremic at the time of transplantation, and recurrent hepatitis can occur as early as 7 days but more typically occurs after 30 days. Current strategies with nucleos(t)ides and hepatitis B immune globulin (Chapters 148 and 149) have essentially eliminated recurrent hepatitis B.

Approximately 80 to 90% of adults undergoing liver transplantation have serologic evidence of prior exposure to cytomegalovirus. Recipients who have positive cytomegalovirus serology or who receive a liver from a cytomegalovirus-positive donor are at risk for either cytomegalovirus syndrome (influenza-like illness) or cytomegalovirus hepatitis. Cytomegalovirus hepatitis requires liver biopsy for diagnosis (E-Fig. 154-1). The greatest risk is in a cytomegalovirus-negative recipient who receives a liver from a cytomegalovirus-positive donor (~60% risk for cytomegalovirus disease). Prophylaxis with oral valganciclovir (900 mg orally daily) during the first 90 days after transplantation effectively reduces risk for cytomegalovirus disease.

Immunosuppressive Medications

Immunosuppressive medications are managed by experts at the patient's liver transplant center. The primary care provider, however, should be aware of the typical maintenance immunosuppressive drugs, how they are monitored, and some of the common toxicities (Table 154-4). The backbone of most maintenance immunosuppressive regimens is an inhibitor of either calcineurin (cyclosporine and tacrolimus) or mTOR (sirolimus and everolimus). These immunosuppressants are metabolized through the hepatic enzyme CYP3A4, which is a pathway for the metabolism of about 60% of commonly prescribed medications. Coadministration of potent inhibitors of CYP3A4 (e.g., ritonavir, erythromycin, telaprevir, and boceprevir) with cyclosporine, tacrolimus, sirolimus, or everolimus will increase the plasma concentrations of the latter

TABLE 154-4 IMMUNOSUPPRESSIVE MEDICATIONS: MONITORING AND ADVERSE SIDE EFFECTS

TYPE	DOSE	MONITORING	ADVERSE EFFECTS
Cyclosporine[a]	100-200 mg bid	Blood level[b]	Nephrotoxicity[c] Neurotoxicity[d] Hypertension
Tacrolimus[e]	1-2 mg bid	Blood level[b]	Nephrotoxicity Neurotoxicity Diabetes mellitus[f]
Prednisone	5-20 mg qd	Clinical	Hypertension Diabetes mellitus[f] Neurotoxicity Fluid retention
Azathioprine	50-200 mg qd	CBC	Neutropenia Thrombocytopenia Anemia
Mycophenolate[g]	500-1500 mg bid	CBC	Neutropenia Dose-related increase in risk of HSV, CMV[h] Gastrointestinal symptoms
Sirolimus[i]	1-3 mg qd	Blood level	Neutropenia Thrombocytopenia Hyperlipidemia Vascular thrombosis
Everolimus[i]	1-2 mg bid	Blood level	Neutropenia Thrombocytopenia Hyperlipidemia

[a]Several different forms of cyclosporine are available. Additional adverse effects of long-term use of cyclosporine include hirsutism, gingival hyperplasia, and dyslipidemia.
[b]Dosages of cyclosporine and tacrolimus are adjusted primarily from trough plasma levels. There are different methods for measuring cyclosporine and tacrolimus levels. The therapeutic range will differ by the method used and whether whole blood or plasma is assayed.
[c]Nephrotoxicity is related to doses and plasma concentrations of cyclosporine or tacrolimus.
[d]Neurotoxicity includes paresthesias, neuropathy, and seizures. Neurotoxicity may be more common with tacrolimus than with cyclosporine.
[e]Several different forms of tacrolimus are available. Additional adverse effects of long-term use of tacrolimus include hirsutism, gingival hyperplasia, and dyslipidemia.
[f]Diabetes mellitus incidence is reduced by steroid withdrawal. Monotherapy with tacrolimus is more often associated with diabetes than is monotherapy with cyclosporine.
[g]Mycophenolate mofetil and mycophenolic acid are inhibitors of inosine 5′-monophosphate dehydrogenase and used as steroid-sparing or calcineurin inhibitor–sparing agents.
[h]Risk of CMV infection under mycophenolate immunosuppression is dose related, with greatest risk at doses of 3 g/day or more.
[i]Sirolimus and everolimus are inhibitors of mTOR, the mammalian target of rapamycin.
CBC = complete blood count; HSV = herpes simplex virus; CMV = cytomegalovirus.

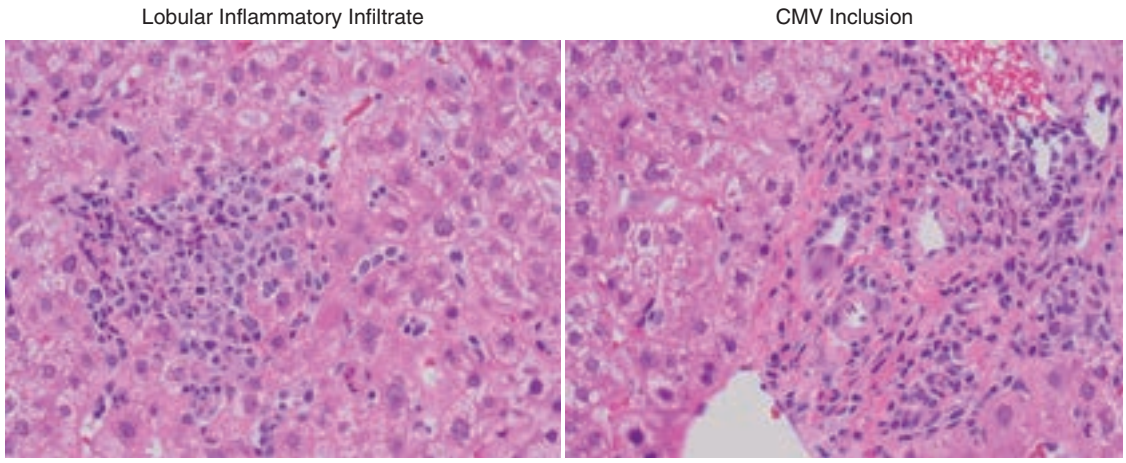

E-FIGURE 154-1. Liver histology of cytomegalovirus (CMV) hepatitis (hematoxylin and eosin stain). The characteristic lobular inflammatory infiltrate (*left*) and magenta-colored CMV inclusion body in a biliary epithelial cell (*right*) are shown. (Courtesy Kalpana M. Devaraj, MD, MHS, Department of Surgical Pathology, Gastrointestinal, Liver, and Heart Transplant Pathology, University of Colorado, Denver.)

drugs and increase the risk for toxicity. Potent enhancers of CYP3A4 (e.g., rifampin, phenobarbital, St. John's wort, and ketoconazole) will decrease plasma concentrations and increase the risk for rejection.

General Management Issues

The liver recipient requires ongoing monitoring and management to promote overall health and wellness. Special attention must be focused on cardiovascular risk factors (hypertension, diabetes, obesity, dyslipidemia, cigarette smoking), monitoring of renal function, adjustment of immunosuppressant medications, attention to drug-drug interactions, monitoring of bone health with bone densitometry every 2 years, and screening or surveillance for cancer. Patients receiving immunosuppressive medications are at long-term risk for infection, and fever should be carefully evaluated for viral, bacterial, atypical bacterial (e.g., tuberculosis), and fungal infections.

Epstein-Barr virus–associated post-transplantation lymphoproliferative disease is an uncommon but serious complication, with an incidence of 0.9 to 2.9% (Chapters 185 and 377). Fever and lymphadenopathy are the usual presenting findings. Initial treatment is reduction of immunosuppressive drugs, but other therapies including rituximab or chemotherapy may be required (Chapter 185).

Grade A Reference

A1. Teperman LW, Poordad F, Bzowej N, et al. Randomized trial of emtricitabine/tenofovir disoproxil fumarate after hepatitis B immunoglobulin withdrawal after liver transplantation. *Liver Transpl.* 2013;19:594-601.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Martin P, DiMartini A, Feng S, et al. Evaluation for liver transplantation in adults: 2013 practice guideline by the American Association for the Study of Liver Diseases and the American Society of Transplantation. *Hepatology*. 2014;59:1144-1165.
2. Bernal W, Wendon J. Acute liver failure. *N Engl J Med*. 2013;369:2525-2534.
3. Karkhanis J, Verna EC, Chang MS, et al. Steroid use in acute liver failure. *Hepatology*. 2014; 59:612-621.
4. Rutherford A, King LY, Hynan LS, et al. Development of an accurate index for predicting outcomes of patients with acute liver failure. *Gastroenterology*. 2012;143:1237-1243.
5. O'Grady J. Timing and benefit of liver transplantation in acute liver failure. *J Hepatol*. 2014;60: 663-670.
6. Curry MP, Forns X, Chung RT, et al. Sofosbuvir and ribavirin prevent recurrence of HCV infection after liver transplantation: an open-label study. *Gastroenterology*. 2015;148:100-107.
7. Charlton M, Gane E, Manns MP, et al. Sofosbuvir and ribavirin for treatment of compensated recurrent hepatitis C virus infection after liver transplantation. *Gastroenterology*. 2015;148:108-117.
8. Singal AK, Chaha KS, Rasheed K, et al. Liver transplantation in alcoholic liver disease current status and controversies. *World J Gastroenterol*. 2013;19:5953-5963.
9. Mathurin P, Moreno C, Samuel D, et al. Early liver transplantation for severe alcoholic hepatitis. *N Engl J Med*. 2011;365:1790-1800.
10. Darwish Murad S, Kim WR, Harnois DM, et al. Efficacy of neoadjuvant chemoradiation, followed by liver transplantation, for perihilar cholangiocarcinoma at 12 US centers. *Gastroenterology*. 2012;143:88-98.
11. Le Treut YP, Gregoire E, Klempnauer J, et al. Liver transplantation for neuroendocrine tumors in Europe—results and trends in patient selection: a 213-case European liver transplant registry study. *Ann Surg*. 2013;257:807-815.
12. McCaughan GW. Current management of HBV pre and post liver transplant. *Curr Hepatitis Rep*. 2013;12:119-123.
13. Wang X, Li J, Riaz DR, et al. Outcomes of liver transplantation for nonalcoholic steatohepatitis: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2014;12:394-402.
14. Carbone M, Neuberger JM. Autoimmune liver disease, autoimmunity and liver transplantation. *J Hepatol*. 2014;60:210-223.
15. Rizvi S, Gores GJ. Pathogenesis, diagnosis, and management of cholangiocarcinoma. *Gastroenterology*. 2013;145:1215-1229.
16. Carey EJ, Iyer VN, Nelson DR, et al. Outcomes for recipients of liver transplantation for alpha-1-antitrypsin deficiency–related cirrhosis. *Liver Transpl*. 2013;19:1370-1376.
17. Gevers TJ, Drenth JP. Diagnosis and management of polycystic liver disease. *Nat Rev Gastroenterol Hepatol*. 2013;10:101-108.
18. Cochat P, Rumsby G. Primary hyperoxaluria. *N Engl J Med*. 2013;369:649-658.
19. Barreiros AP, Galle PR, Otto G. Familial amyloid polyneuropathy. *Dig Dis*. 2013;31:170-174.
20. Fox AN, Vagefi PA, Stock PG. Liver transplantation in HIV patients. *Semin Liver Dis*. 2012;32:177-185.
21. Olthoff KM, Abecassis MM, Emond JC, et al. Outcomes of adult living donor liver transplantation: comparison of the Adult-to-adult Living Donor Liver Transplantation Cohort Study and the national experience. *Liver Transpl*. 2011;17:789-797.
22. Abecassis MM, Fisher RA, Olthoff KM, et al. Complications of living donor hepatic lobectomy—a comprehensive report. *Am J Transplant*. 2012;12:1208-1217.
23. Lucey MR, Terrault N, Ojo L, et al. Long-term management of the successful adult liver transplant: 2012 practice guideline by the American Association for the Study of Liver Diseases and the American Society of Transplantation. *Liver Transpl*. 2013;19:3-26.
24. Fagiuoli S, Colli A, Bruno R, et al. Management of infections pre- and post-liver transplantation: report of an AISF consensus conference. *J Hepatol*. 2014;60:1075-1089.

REVIEW QUESTIONS

1. A 50-year-old male liver transplant recipient comes to your office with community-acquired pneumonia, for which he had been prescribed erythromycin 7 days ago by an emergency department physician. His maintenance immunosuppression is tacrolimus and mycophenolate mofetil. He is now confused, and his laboratory tests reveal Na 135, K 6.2, HCO_3 15, Cl 108, blood urea nitrogen 45, creatinine 4.1. The results of his complete blood count and the rest of the chemistry profile are normal. A computed tomography scan of the brain is unremarkable. The most likely cause of this patient's clinical presentation is

A. Viral meningitis
B. Urinary tract infection
C. Tacrolimus toxicity
D. Mycophenolate mofetil toxicity
E. Erythromycin toxicity

Answer: C A key to appropriate management of the liver transplant recipient is an understanding of potential drug-drug interactions. Calcineurin and mTOR inhibitors are cornerstones of maintenance immunosuppression in most liver recipients. These drugs are processed by P-glycoprotein and metabolized by CYP3A4, both of which are key drug disposition pathways located in intestinal epithelial cells, hepatocytes, and renal tubular cells. Erythromycin inhibits CYP3A4, thereby blocking the metabolism of tacrolimus and leading to high tacrolimus levels and tacrolimus renal and neurologic toxicity. Physicians must check for any drug-drug interactions before prescribing medications to liver transplant recipients who are taking immunosuppressive medications.

2. A 62-year-old woman who received a liver transplant 5 years ago comes for a routine physical examination. She has a body mass index of 32 kg/m^2, blood pressure (BP) 165/105, total cholesterol 285 mg/dL (low-density lipoprotein 205 mg/dL), triglycerides 305 mg/dL, and fasting blood glucose of 165 mg/dL. Her current immunosuppression regimen is cyclosporine and prednisone. You notify her transplant center, and they switch her immunosuppression to low-dose tacrolimus and mycophenolate mofetil. The expected outcomes of the change in immunosuppression are

A. Weight loss, lower BP, improved lipids, improved insulin sensitivity
B. Weight gain, lower BP, improved lipids, improved insulin sensitivity
C. Weight loss, higher BP, improved lipids, improved insulin sensitivity
D. Weight loss, lower BP, higher lipids, improved insulin sensitivity
E. Weight loss, lower BP, improved lipids, worsened insulin resistance

Answer: A Each immunosuppressive medication has a characteristic pattern of side effects. Most are manageable by dose reduction or changing the immunosuppressive prescription. In this case, both cyclosporine and prednisone contributed to the constellation of obesity, hypertension, dyslipidemia, and glucose intolerance. Mycophenolate mofetil does not have these side effects, and low-dose tacrolimus is less likely than cyclosporine to cause obesity, dyslipidemia, or glucose intolerance.

3. A 47-year-old woman underwent liver transplantation for autoimmune hepatitis 5 years ago. After transplantation, she experienced steroid-refractory rejection, which was treated with antilymphocyte globulin, and cytomegalovirus hepatitis, which was treated with intravenous ganciclovir. She now presents with fever, leukopenia, anemia, and lymphadenopathy. Viral, bacterial, and fungal cultures are negative. Epstein-Barr virus serology indicates past exposure, but Epstein-Barr virus DNA is negative. Lymph node biopsy is likely to demonstrate

A. Mycobacteria
B. Epstein-Barr Virus
C. Cytomegalovirus
D. Fungi
E. CD20-positive lymphocytes

Answer: E Risk factors for post-transplantation lymphoproliferative disease (PTLD) include primary Epstein-Barr virus infection, cytomegalovirus disease, cytomegalovirus donor-recipient mismatch, and augmented immunosuppression, especially antilymphocyte globulin. The association of PTLD and Epstein-Barr virus is variable in adult liver transplant recipients; later onset PTLD is less likely to be associated with Epstein-Barr virus activation or infection. Fever, cytopenias, and lymphadenopathy are the key clinical features. Initial treatment is a reduction in immunosuppression. Monoclonal anti-CD20 antibodies and chemoradiation treatments may be required. Antiviral therapy is ineffective.

155

DISEASES OF THE GALLBLADDER AND BILE DUCTS

EVAN L. FOGEL AND STUART SHERMAN

GALLBLADDER

Gallstones

EPIDEMIOLOGY

Gallstone disease is one of the most common and costly digestive diseases, with an estimated annual direct cost of $15 billion in the United States. Newly diagnosed gallstone disease occurs in more than one million people annually in the United States, and more than 750,000 cholecystectomies are now performed annually. The prevalence of gallstones is about 10 to 15% in American and European adults, with women affected about twice as often as men. Cholesterol gallstones are uncommon in individuals younger than 20 years, but a sharp increase is noted with each decade up to approximately the age of 70 years, particularly in women. About 20% of women and 10% of men have gallstones by 60 years of age. Together, approximately 12% of Americans or 36 million men and women harbor gallstones.

In the United States, the prevalence of stones is highest in Mexican American women (26%), followed by white women (17%) and African American women (14%). The prevalence of gallstones is extremely high in Native Americans, especially in women. In Chileans and Bolivians of Indian ancestry, gallstones are also common, and gallstone-associated cancer is the most common gastrointestinal cancer in these countries.

Environmental factors and genetic predisposition are likely to play an interactive role for gallstone formation. Pregnancy may contribute to the predominance of cholesterol stones in younger women as it is associated with progesterone-induced impaired gallbladder emptying and estrogen-mediated increased cholesterol saturation of bile. The prevalence of gallbladder stones in nulliparous women is approximately one tenth of that noted in multiparous women (1.3% vs. 13%). Exogenous estrogen administration in the form of hormone replacement therapy and oral contraceptives is also associated with stone formation. Other medications, including somatostatin analogues, ceftriaxone, and clofibrate, have been associated also with an increased incidence of gallbladder stones.

Obesity is a major risk factor for development of cholesterol stones. Obese individuals have an increase in biliary cholesterol secretion relative to bile acid and lecithin secretion, thereby resulting in bile supersaturation. However, rapid weight loss is also associated with an increased risk of gallstones. Diminished ileal absorption of bile acids, due to surgical resection or bypass or to active inflammation (e.g., Crohn disease), may also lead to an increased likelihood of gallstones.

PATHOBIOLOGY

Gallstone Formation

Gallstones represent a failure to maintain certain biliary solutes, primarily cholesterol and calcium salts, in a solubilized state. Gallstones are classified by their cholesterol content as either cholesterol or pigment stones. Pigment stones are further classified as either black or brown. Most cholesterol stones contain calcium salts in their core, and pure cholesterol gallstones are uncommon (10%). In most American populations, 70 to 80% of gallstones are cholesterol, and black pigment stones account for most of the remaining 20 to 30%.

In normal bile, cholesterol is soluble in the form of mixed micelles with optimal concentration of bile salts and phospholipids. With disproportionate concentrations, bile becomes supersaturated, and the excess cholesterol precipitates as monohydrate crystals. These crystals become embedded in gallbladder mucin gel with bilirubinate to form biliary sludge, which may eventually aggregate into a gallbladder stone.

Black pigment stones make up a small proportion of gallstones. These stones consist of polymerized calcium bilirubinate, precipitated as a result of exceeding the solubility of calcium and unconjugated bilirubin. In addition to increasing age, the formation of black pigment stones is more common in individuals who have conditions that create an excessive amount of unconjugated bilirubin (e.g., chronic hemolysis in hemoglobinopathies, cirrhosis, ineffective erythropoiesis), who are being fed by total parenteral nutrition (Chapter 217), or who have ileal diseases. Black pigment stones are typically tarry, are usually not associated with infected bile, and are located almost exclusively in the gallbladder.

By contrast, brown pigment stones are coarse in texture and are primarily formed in the bile duct as a result of bacterial infection that releases β-glucuronidase to hydrolyze glucuronic acid from bilirubin. Phospholipid hydrolysis also increases, thereby leading to precipitation of calcium, bilirubin, and free fatty acids and also resulting in the formation of brown pigment stones. Brown pigment stones account for 30 to 90% of gallstones in Asian populations, may also occur throughout the entire biliary tree, and are frequently associated with pyogenic cholangiohepatitis.

Gallbladder contractility is impaired in some patients with gallstones. Although gallbladder dysfunction can be the consequence of gallstone disease or of excessive cholesterol infiltration into the gallbladder's smooth muscle, evidence suggests that gallbladder stasis itself can lead to gallbladder stone formation. A normal gallbladder ejects 10 to 20% of its contents into the duodenum in response to enteric nervous stimulation. The presence of postprandial intestinal fat further increases gallbladder contractility, which is mediated by the enteric nervous system and cholecystokinin. Gallbladder stasis is frequently evident in patients with risk factors for forming gallstones, including obesity, pregnancy, rapid weight loss, and prolonged fasting (Table 155-1). Furthermore, gallbladder dysmotility is an independent risk factor for recurrent gallstones in patients who have been treated with extracorporeal shock wave lithotripsy.

Acute Calculous Cholecystitis

The most common complication of gallstone disease is acute cholecystitis, which occurs in 15 to 20% of symptomatic patients. Acute cholecystitis results when a stone becomes lodged at the gallbladder–cystic duct junction, where it impairs gallbladder outflow and drainage. The extent of inflammation and the progression of acute cholecystitis are related to the duration and degree of obstruction. In the most severe cases, this process can lead to ischemia and necrosis of the gallbladder wall. More often, the stone spontaneously dislodges, and the inflammation gradually resolves. Acute cholecystitis is primarily an inflammatory rather than an infectious process, but about 50%

TABLE 155-1 RISK FACTORS ASSOCIATED WITH GALLSTONE FORMATION

NONMODIFIABLE FACTORS	MODIFIABLE FACTORS
Increasing age	Pregnancy and parity
Female gender	Obesity
Ethnicity	Low-fiber, high-calorie diet
Genetics, family history	Prolonged fasting
	Medications: clofibrate, estrogens, octreotide
	Low-level physical activity
	Rapid weight loss
	Hypertriglyceridemia, low high-density lipoprotein
	Metabolic syndrome
	Gallbladder stasis
	Terminal ileal disease or resection
	Total parenteral nutrition, fasting state

of patients with acute cholecystitis have secondary bacteriobilia, most commonly with *Escherichia coli*.

CLINICAL MANIFESTATIONS

The clinical spectrum of cholelithiasis ranges from the asymptomatic state to fatal complications. Among patients with asymptomatic gallstones, approximate annual risks are 1% for biliary pain, 0.3% for acute cholecystitis, 0.2% for symptomatic choledocholithiasis, and 0.04 to 1.5% for gallstone pancreatitis. However, these low individual percentages represent a huge population-wide number of symptomatic patients, given the frequency of gallstones. Overall, about 1 to 2% of asymptomatic individuals with gallstones develop serious symptoms or complications each year related to their gallstones.

The majority of gallstones are asymptomatic and are discovered on imaging studies performed for other indications. In such individuals, the gallbladder fills and empties normally, and the gallstones remain in the gallbladder and do not obstruct the cystic duct. Over time, however, asymptomatic gallstones can become symptomatic and be manifested as biliary colic due to impaction of a gallstone at the neck of the gallbladder or cystic duct. Although the pain is commonly termed biliary colic, the majority of patients actually note constant pain due to obstruction of the cystic duct and a progressive increase in gallbladder wall tension, rather than the paroxysmal pain of typical colic. The pain usually is located in the right upper quadrant or epigastrium and frequently radiates to the back and right scapula. Although biliary colic classically occurs after fatty meals, an association with meals is present in only 50% of patients, and the pain often develops more than 1 hour after eating. The duration of pain is typically 1 to 5 hours, but it may persist up to 24 hours. Pain persisting beyond 24 hours suggests that acute inflammation or cholecystitis is present. Episodes of biliary colic are usually less frequent than one episode per week. Other symptoms, such as nausea and vomiting, accompany each episode in 60 to 70% of cases. Bloating and belching are also present in 50% of patients. Fever and jaundice occur much less frequently with simple biliary colic. Although some patients with gallstones have continuous pain, predominantly in the back or the left upper quadrant, rather than episodic pain, alternative causes should be considered in such patients.

Acute Calculous Cholecystitis

Patients with acute cholecystitis typically present with right upper quadrant pain similar to biliary colic. In acute cholecystitis, however, the pain is usually unremitting, may last several days, and is often associated with nausea, emesis, anorexia, and fever. On physical examination, patients usually have a low-grade fever, with localized right upper quadrant tenderness and guarding. The presence of Murphy sign, an inspiratory arrest during deep palpation of the right upper quadrant, is the classic physical finding of acute cholecystitis. A palpable right upper quadrant mass is appreciated in one third of patients and usually represents omentum that has migrated to the area around the gallbladder in response to the inflammation. Mild jaundice (bilirubin level < 6 mg/dL) may be present. Significant jaundice is rare with acute cholecystitis but when present suggests the presence of common bile duct stones, cholangitis, or obstruction of the common hepatic duct by severe pericholecystic inflammation because of the impaction of a large stone in Hartmann pouch, which mechanically obstructs the bile duct. High fever suggests ascending cholangitis, often with bacterial infection (Fig. 155-1). Acute cholecystitis may coexist with choledocholithiasis or its complications of acute cholangitis and gallstone pancreatitis.

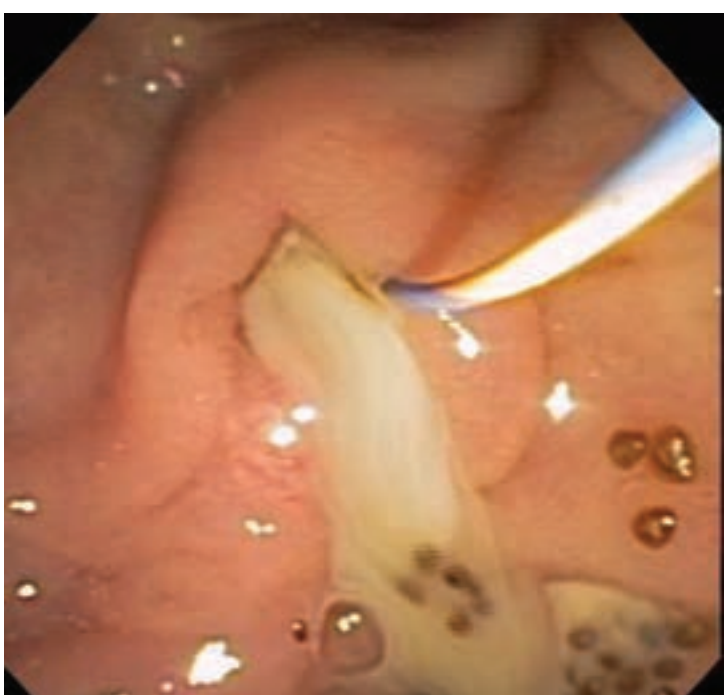

FIGURE 155-1. Endoscopic image of pus exiting the biliary orifice in this patient who presented with ascending cholangitis secondary to choledocholithiasis (note several small stones in duodenum).

FIGURE 155-2. Ultrasound showing a gallstone. (From Afdhal N. Diseases of the gallbladder and bile ducts. In: Goldman L, Schafer A, eds. *Goldman's Cecil Medicine.* 24th ed. Philadelphia: Elsevier Saunders; 2012:1017.)

DIAGNOSIS

Transabdominal ultrasound is the radiologic procedure of choice to identify gallstones (Fig. 155-2).[1] Because the ultrasound waves cannot penetrate the stones, acoustic shadowing is seen posterior to the stones, thereby facilitating diagnosis. Free-floating gallbladder stones will also move to a dependent position when the patient is repositioned during scanning. If both of these features are present, the positive predictive value of ultrasound approaches 100%. Nonshadowing echoes alone, however, may be caused by gallbladder polyps. Gallstones may be missed because of a lack of contrasting bile around the stones, as may occur with an impacted cystic duct stone or when the

gallbladder is filled with stones. Small gallstones may not cast an acoustic shadow. An ileus with increased abdominal gas, as can occur with acute cholecystitis or pancreatitis (Chapter 144), may limit visualization of the gallbladder. Overall, the false-negative rate of ultrasound for detection of gallstones is less than 5% but may increase to 15% with acute cholecystitis. Ultrasound may also demonstrate dilation of the intrahepatic and extrahepatic bile ducts. Dilated ducts may signify obstruction due to stones in the common bile duct, distal strictures, or malignant obstruction (Chapters 194 and 196).

Ultrasound has a sensitivity and specificity of 85% and 95%, respectively, for diagnosis of acute cholecystitis. In addition to the presence of gallstones, findings suggestive of acute cholecystitis include thickening of the gallbladder wall (>4 mm) and the presence of pericholecystic fluid. Focal tenderness directly over the gallbladder (sonographic Murphy sign) also is suggestive of acute cholecystitis.

Cholescintigraphy provides a noninvasive, anatomic, and functional evaluation of the liver, gallbladder, bile duct, and duodenum, although it has generally been superseded by ultrasound for this purpose, except in certain situations, such as identification of a suspected bile leak after cholecystectomy or a functional gallbladder disorder (see later). In this procedure, technetium Tc 99m–labeled iminodiacetic acid derivatives are injected intravenously, taken up by the liver, and excreted into the bile. These hepatobiliary iminodiacetic acid (HIDA) scans provide functional information about the liver's ability to excrete radiolabeled substances into a nonobstructed biliary tree. The tracer should be taken up by the liver, gallbladder, common bile duct, and duodenum within 1 hour. Nonvisualization of the gallbladder 1 hour after the injection of the radioisotope with filling of the bile duct and duodenum indicates an obstructed cystic duct and in the acute clinical setting is highly sensitive (95%) and specific (95%) for acute cholecystitis, although false-positive results are often seen in the setting of gallbladder stasis (e.g., critically ill patients, total parenteral nutrition). Slow uptake of the tracer by the liver suggests hepatic parenchymal disease. Filling of the gallbladder and bile duct with delayed or absent filling of the intestine may suggest an obstruction at the level of the major papilla.

Abdominal computed tomography (CT) is less sensitive than ultrasound for diagnosis of gallstones and is primarily indicated for the diagnosis of complications of gallstone disease, such as acute cholecystitis, choledocholithiasis, pancreatitis, and gallbladder cancer. By comparison, *plain abdominal radiographs* are of little value in the evaluation of gallbladder disease because only 15% of gallstones contain sufficient calcium to appear radiopaque. Plain films can be useful, however, for diagnosis of other causes of acute abdominal pain (e.g., perforated viscus, bowel obstruction). Rarely, abdominal films may show a calcified gallbladder wall (Fig. 155-3) in chronic cholecystitis or findings such as pneumobilia or gallstone-associated ileus in acute cholecystitis.

A CT scan also is frequently performed to evaluate an acutely ill patient with abdominal pain. The CT scan can detect gallstones, gallbladder wall thickening, pericholecystic fluid and edema, and air in the gallbladder or gallbladder wall (emphysematous cholecystitis), but it is generally less sensitive than ultrasonography for finding these conditions.

Magnetic resonance imaging is highly sensitive for diagnosis of both gallstones and common duct stones, but stones smaller than 3 mm may be missed. Endoscopic ultrasound provides excellent imaging of the gallbladder and biliary tree but is rarely the primary imaging modality for detection of a gallbladder stone. Endoscopic collection of bile for crystal analysis may serve as a surrogate for microlithiasis not seen on transabdominal ultrasound.

With cholecystitis, laboratory evaluation can show a mild leukocytosis, with a white blood cell count of 12,000 to 15,000 cells/μL. However, many patients have a normal white blood cell count. Leukocytosis greater than 20,000 cells/μL should suggest further complications of cholecystitis, such as gangrene, perforation, or cholangitis. Mild elevations in serum bilirubin, alkaline phosphatase, aminotransferase, and amylase levels may also be seen with acute cholecystitis.

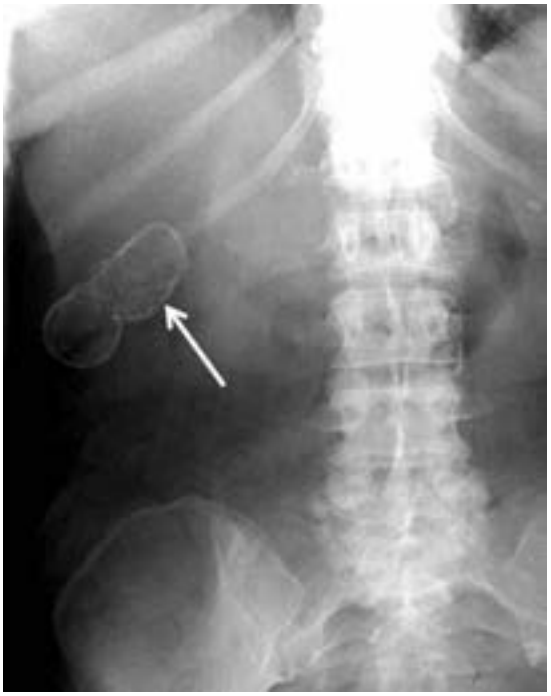

FIGURE 155-3. Plain abdominal radiograph illustrating a porcelain gallbladder. Note the calcified gallbladder wall.

TREATMENT

Silent Gallstones

The longer that stones remain silent, the less likely they are to cause symptoms. Furthermore, almost all patients will develop symptomatic disease before developing one of the serious complications of gallstones. Therefore, prophylactic cholecystectomy is not generally indicated in patients with asymptomatic gallstones.

In select groups of patients, however, prophylactic cholecystectomy should be considered, even when gallbladder stones are absent. A porcelain gallbladder with a calcified gallbladder wall is associated with a 5% or higher risk of malignant transformation, high enough to justify cholecystectomy. Patients with a long common channel between the bile and pancreatic ducts (i.e., anomalous pancreaticobiliary duct junction; see later) also are at significant risk for gallbladder cancer and should undergo prophylactic cholecystectomy. Acute cholecystitis is a potentially life-threatening condition in immunosuppressed patients, so prophylactic cholecystectomy is generally recommended before or at the time of major organ transplantation. Data also support cholecystectomy in patients undergoing bariatric weight loss surgery even in the absence of gallstones because of their nearly 30% risk for developing gallstones and requiring cholecystectomy during rapid weight loss in the first year after surgery. Prophylactic cholecystectomy adds minimal morbidity and mortality risks to most bariatric operations and is clearly indicated in patients with gallstones. Some patients with silent gallbladder stones may also benefit from prophylactic cholecystectomy. In patients with sickle cell disease (Chapter 163), for example, cholecystitis can precipitate a crisis with substantial operative risks. Large gallstones (>3 cm) are more frequently associated with acute cholecystitis and gallbladder carcinoma, so prophylactic cholecystectomy may also be indicated in these patients.

Symptomatic Gallstones

The operative management of gallstones has been the standard of care for more than a century. Early surgery within 24 hours is generally preferred for patients with biliary colic,[■] and day surgery is as safe and effective as an overnight stay.[■] More than 90% of these cholecystectomies are performed laparoscopically, with about 3% of elective procedures converted to an open procedure in the operating room. Contraindications to laparoscopic surgery include significant bleeding and Child's class C cirrhosis (Chapter 153). Some patients with severe chronic obstructive pulmonary disease or heart failure may not tolerate the pneumoperitoneum required for laparoscopic surgery, and the prior upper abdominal surgery may increase the difficulty of or preclude laparoscopic cholecystectomy. Serious complications of laparoscopic cholecystectomy are rare, with a reported incidence of 0.6 to 1.5% for any bile duct leaks and 0.3 to 0.6% for a major bile duct injury. Although these risks are higher than for open surgery, the overall mortality rate (<0.3%) is lower for laparoscopic surgery, and the postoperative recovery is much easier.

Nonsurgical options for the treatment of gallstone disease are rarely used today because of their limited efficacy and the widespread application of laparoscopic cholecystectomy. Oral dissolution therapy (e.g., ursodeoxycholic acid, 15 mg/kg/day) may be considered in symptomatic but not in asymptomatic patients who have cholesterol gallstones in a functioning gallbladder, but it completely dissolves stones in only 40% of patients, and stones recur in up to 50% of patients within 5 years after therapy is stopped. Lifelong therapy, therefore, may be necessary. The direct infusion of organic solvents (methyl *tert*-butyl ether, continuously infused and aspirated manually four to six times

per minute, for an average of 5 hours/day, for 1 to 3 days) into the gallbladder also is efficacious only for cholesterol gallstones, and the recurrence rate is similar to that of oral dissolution therapy. Extracorporeal shock wave lithotripsy can be considered for a single stone of any type (i.e., cholesterol or calcium bilirubinate) 0.5 to 2 cm in diameter, but only a small percentage of symptomatic patients fit these criteria. As a result, this therapy is limited to a very select group of patients.

Acute Calculous Cholecystitis

After medical stabilization of the patient with intravenous fluids as needed, broad-spectrum antibiotics (e.g., piperacillin-tazobactam, 3.375 g every 6 hours; or ceftriaxone, 1 to 2 g once daily, plus metronidazole, 500 mg every 6 hours; or levofloxacin, 500 mg once daily, plus metronidazole), and parenteral analgesics as necessary (see Table 30-5), the treatment of choice for acute cholecystitis is cholecystectomy. When antibiotic therapy is initiated, the duration is tailored to clinical improvement. Laparoscopic cholecystectomy significantly reduces morbidity, length of hospital stay, and time to return to work, favoring patients compared with open surgery. However, the conversion rate to an open procedure is up to 25% compared with about 3% for elective laparoscopic surgery.

Prospective randomized trials have shown that early laparoscopic cholecystectomy (within 3 days of symptom onset) can be accomplished with a morbidity and mortality rate similar to that of delayed cholecystectomy,■ with no differences in the conversion rate to open cholecystectomy. Length of hospital stay, and therefore costs, are significantly reduced in patients undergoing early surgery. Because about 20% of patients fail to stabilize with initial medical therapy and require operation during the initial admission or before the end of the planned cooling-off period, the current recommendation is to proceed with early laparoscopic cholecystectomy for acute cholecystitis unless there are contraindications to it. If the bilirubin level is <4 mg/dL and the patient has no evidence of cholangitis, common duct exploration is probably not required.■

In high-risk patients whose medical conditions preclude cholecystectomy, a percutaneous cholecystostomy can allow prompt gallbladder drainage. If such drainage and appropriate antibiotics do not lead to clear improvement within 24 hours, however, laparotomy is indicated because failure to improve after percutaneous drainage is usually caused by gangrene of the gallbladder or perforation. If cholecystostomy is successful and the acute episode resolves, the patient can electively undergo either cholecystectomy or percutaneous stone extraction and removal of the cholecystostomy tube. Alternatively, an endoscopic retrograde cholangiopancreatography (ERCP)–guided stent can be placed to drain the gallbladder if the cystic duct is patent.

PREVENTION

Moderate physical activity and dietary management (high fiber intake, avoidance of saturated fatty acids) may lower the risk of gallstone disease. Daily administration of cholecystokinin (3.5 μg) in patients receiving prolonged total parenteral nutrition may prevent formation of gallbladder sludge. Oral ursodeoxycholic acid (15 mg/kg/day) has been clearly demonstrated to be beneficial in prevention of gallstone disease during rapid weight loss■ and in patients who need long-term somatostatin therapy. For secondary prevention (i.e., in patients with gallstones already present), there are insufficient data to support the use of any medical therapy.

Complications of Acute Calculous Cholecystitis

Several complications of acute cholecystitis include empyema of the gallbladder, emphysematous cholecystitis, and gangrene leading to gallbladder perforation. Each of these complications can be associated with significant morbidity and mortality and therefore requires prompt surgical intervention. In 1 to 2% of patients with acute cholecystitis, the gallbladder will perforate into an adjacent hollow viscus, thereby creating a cholecystenteric fistula; the duodenum (79%) and the hepatic flexure of the colon (17%) are the most common sites. The episode of acute cholecystitis generally resolves as the gallbladder spontaneously decompresses after the fistula forms. If a large gallstone passes from the gallbladder into the small intestine, a mechanical bowel obstruction, termed gallstone ileus, may result. Gallstone ileus occurs in 10 to 15% of patients with a cholecystenteric fistula. Patients with gallstone ileus present with signs and symptoms of intestinal obstruction—nausea, vomiting, and abdominal pain. Abdominal films will demonstrate small bowel distention and air-fluid levels and may give additional clues to the source of the obstruction (pneumobilia or a calcified gallstone distant from the gallbladder). The initial management of gallstone ileus includes relieving the obstruction. Most frequently, this goal can be achieved by removing the gallstone through an enterotomy, although endoscopic retrieval of the offending stone can be done, depending on where the stone is located.

Acute Acalculous Cholecystitis

Acute acalculous cholecystitis, which accounts for 5 to 10% of all cases of acute cholecystitis, usually occurs in critically ill patients after trauma, burns, long-term parenteral nutrition, and major nonbiliary operations (e.g., abdominal aneurysm repair, cardiopulmonary bypass). The cause of acute acalculous cholecystitis remains unclear, although gallbladder stasis with increased bacterial colonization and ischemia have been implicated.

The symptoms and signs of acute acalculous cholecystitis are similar to those of acute calculous cholecystitis, with right upper quadrant pain and tenderness, fever, and leukocytosis. The disease often has a more fulminant course than acute calculous cholecystitis and more frequently progresses to gangrene, empyema, or perforation. Except for the absence of gallstones, ultrasound and CT findings are similar to those of calculous cholecystitis, including gallbladder wall thickening and pericholecystic fluid. On cholescintigraphy, the gallbladder does not fill; however, the false-positive rate (absent gallbladder filling without acute acalculous cholecystitis) may be as high as 40%.

Emergency cholecystectomy is recommended if the diagnosis is established or even if clinical suspicion is high because the risk of gangrene, perforation, or empyema exceeds 50%. Cholecystectomy rather than cholecystostomy is usually required, but percutaneous cholecystostomy or endoscopic gallbladder stenting is recommended in patients unable to undergo surgery. The mortality rate for acute acalculous cholecystitis can be as high as 40%, mostly because of the concomitant illnesses in patients who develop this disease.

Functional Gallbladder Disorder

Some patients present with typical symptoms of biliary colic but do not have any evidence of gallstones on ultrasound examination. If further investigations such as liver chemistries, amylase and lipase levels, CT scan, and even upper gastrointestinal endoscopy are unremarkable, the diagnosis of a functional gallbladder disorder should be considered.[2] The pathobiology is poorly understood, but one possibility is that the obesity epidemic has increased the population-wide accumulation of fat in the gallbladder wall—cholecystosteatosis—thereby decreasing the gallbladder's ability to empty. Some patients may have intermittent gallbladder outlet obstruction due to cystic duct spasm, poor coordination between the contraction of the gallbladder and the sphincter of Oddi, or dysmotility of the gallbladder. In a cholecystokinin-stimulated ^{99m}Tc-HIDA scan, cholecystokinin is infused intravenously after the gallbladder has filled with the ^{99m}Tc-labeled radionuclide, and a gallbladder ejection fraction is calculated 20 minutes later. An ejection fraction of less than 35% at 20 minutes is considered abnormal, and most of these patients have histopathologic evidence of chronic cholecystitis, although a low gallbladder ejection fraction is not specific for a functional gallbladder disorder (Table 155-2). The efficacy of laparoscopic cholecystectomy is controversial in this setting, but the Society of American Gastrointestinal and Endoscopic Surgeons recommends it. The percentage of patients undergoing cholecystectomy for functional gallbladder disorder in the United States during the past 15 years has increased from less than 5% to more than 20% of patients having the gallbladder removed.

Tumors of the Gallbladder

Benign

Cholesterol polyps are not true neoplasms but rather are cholesterol-filled projections of gallbladder mucosa that protrude into the lumen. These polyps account for approximately 50% of all polypoid gallbladder lesions, are usually

TABLE 155-2 SETTINGS IN WHICH A LOW GALLBLADDER EJECTION FRACTION MAY BE IDENTIFIED

Functional gallbladder disorder (chronic acalculous cholecystitis)
Cystic duct obstruction
Sphincter of Oddi dysfunction
Asymptomatic, healthy individuals
Diabetes
Pregnancy
Cirrhosis
Obesity
Celiac disease
Medications (narcotic analgesics, calcium-channel antagonists, oral contraceptive agents, benzodiazepines, histamine H_2-receptor antagonists)

smaller than 1 cm, are typically found incidentally on imaging studies as nonmobile filling defects, are usually asymptomatic unless associated with gallstones, and do not have malignant potential.

Adenomyosis consists of a hypertrophic gallbladder muscle layer with mucosal diverticula called Rokitansky-Aschoff sinuses. This condition may affect the gallbladder locally, particularly in the fundus, where it appears as a hemispheric lesion with a central dimple; segmentally, as an annular stricture; or diffusely, when it involves the entire gallbladder wall. The cause is not entirely clear but may be secondary to a gallbladder motility disorder. Isolated adenomyosis can cause biliary-type symptoms and can progress to gallbladder cancer, so prophylactic cholecystectomy is recommended.

Gallbladder adenomas are benign epithelial tumors with malignant potential. Adenomas usually are manifested as solitary, nonmobile filling defects on ultrasound. Polyps smaller than 0.5 cm, regardless of total polyp number, can be observed with serial imaging studies every 3 to 6 months. Larger polyps, however, may harbor a carcinoma in situ, and cholecystectomy is recommended for polyps larger than 1 cm and for any patients with biliary symptoms.

Malignant

Gallbladder cancer is the most common biliary tract malignant neoplasm and fifth most common gastrointestinal cancer overall, with approximately 7000 new cases diagnosed annually in the United States (2.5 cases/100,000 population). The usual age at onset is the sixth or seventh decade, with a female-to-male ratio of 3:1. Gallbladder cancer is more common in Native Americans, Mexicans, Alaskans, and American Hispanics as well as in residents of Israel, Chile, and northern Japan.

Perhaps because of chronic inflammation, gallbladder cancer is strongly associated with gallstones, which are identified in more than 90% of patients with gallbladder carcinoma. Conversely, only 1% of patients with gallstones develop gallbladder carcinoma. Choledochal cysts are associated with an increased risk of malignant neoplasms throughout the biliary tree, including the gallbladder, perhaps because of increased stasis, chronic inflammation, and infection. In patients with choledochal cysts, excision of extrahepatic cysts is recommended to avoid biliary tract cancer. By comparison, biliary sphincterotomy alone may be adequate for a type III extrahepatic cyst.

About 90% of gallbladder cancers are adenocarcinomas (90% scirrhous, 5% papillary, 5% colloid). The remaining tumors are anaplastic or squamous cell cancers. Gallbladder cancers spread by local extension and direct invasion of adjacent structures, including the common hepatic duct, liver, duodenum, and colon. Lymphatic drainage is to adjacent lymph nodes, and disseminated disease is to the liver and the peritoneal surface.

Most patients (80%) present with abdominal pain of less than 1 month in duration, which may be difficult to distinguish from symptoms of biliary colic or acute cholecystitis. Nausea and vomiting (50%) and weight loss (40%) are often present, and jaundice (30%) is a poor prognostic sign that typically signifies porta hepatis involvement with tumor. Up to 20% of gallbladder cancers are found at cholecystectomy performed for gallstones, whereas incidental cancers are found at 1% of cholecystectomies.

Diagnosis may be difficult preoperatively because laboratory test results may be normal or nonspecific even when advanced disease is manifested by hypoalbuminemia and anemia. There is no reliable tumor marker. Liver test results are abnormal when the tumor or periportal lymphadenopathy is associated with biliary obstruction. Ultrasonography has a sensitivity of 75 to 80% for detection of gallbladder cancer, with findings ranging from a complex luminal mass to gallbladder wall thickening, polypoid mass, or gallstones. CT or magnetic resonance cross-sectional imaging can assess the extent of disease, including regional and distant metastases. Endoscopic ultrasound may aid in determining the extent of local invasion and nodal involvement, but it rarely is necessary in the preoperative evaluation. ERCP is indicated only in patients who have clinical evidence of biliary obstruction and are being considered for stent placement to palliate their jaundice.

PREVENTION AND TREATMENT

Rx

Patients with gallstones larger than 3 cm have a 10-fold increased risk for development of gallbladder cancer, so prophylactic cholecystectomy should be considered even in an asymptomatic patient. A porcelain gallbladder, with diffuse calcification of the gallbladder wall, is an indication for cholecystectomy in the asymptomatic patient because of the increased risk of cancer. Cholecystectomy also is indicated for any gallbladder polyp larger than 1 cm, in patients with choledochal cysts other than type III, in anomalous pancreaticobiliary union, and in adenomyosis of the gallbladder.

The resectability rates for gallbladder cancer range from 15 to 30%. When the tumor does not extend beyond the muscle layer of the gallbladder wall (T_1), simple cholecystectomy alone may be curative, with a 5-year survival rate approaching 100%. Tumors that extend through the gallbladder wall (stage II/III) require a more extensive resection (cholecystectomy, partial hepatectomy, lymph node dissection). Stage II tumors (no invasion beyond the gallbladder serosa) may have up to a 60 to 80% 5-year survival rate, whereas stage III tumors have a 25% 5-year survival rate. Median survival with unresectable (stage IV) disease is only 2 to 3 months. Radiation therapy has not been shown to be effective, whereas chemotherapy regimens in unresectable disease (typically similar to regimens for pancreatic cancer [Chapter 194]) are associated with response rates of approximately 20%. Because of the late presentation of this disease and spread of tumor at diagnosis, the overall 5-year survival is less than 10%.

BILE DUCTS

Bile Duct Stones

PATHOBIOLOGY

Bile duct stones, or choledocholithiasis, can be classified as either primary or secondary. Primary duct stones develop de novo within the bile ducts, whereas secondary stones develop in the gallbladder and subsequently pass into the bile duct. In the Western world, more than 85% of all bile duct stones are secondary. Primary duct stones typically occur in conditions associated with bile stasis (e.g., benign biliary strictures, sclerosing cholangitis, choledochal cysts, periampullary diverticula), which promotes bacterial overgrowth with subsequent bilirubin deconjugation and the breakdown of biliary lipids, thereby resulting in the formation of brown pigment stones.

CLINICAL MANIFESTATIONS

Bile duct stones are discovered incidentally in 5 to 12% of patients during the evaluation of gallbladder stones and suspected cholecystitis. It is difficult to determine whether the existing bile duct stones are asymptomatic in patients who present with biliary pain alone because pain can originate from either the gallbladder stones or bile duct stones. More than 50% of patients with retained bile duct stones experience recurrent symptoms during a follow-up period of 6 months to 13 years, and 25% of cases develop serious complications.

Common clinical symptoms and signs of bile duct stones include epigastric or right upper quadrant pain, fever, and jaundice (referred to as Charcot triad). Pain can be mild or severe, and severe episodes must be differentiated from other potentially life-threatening events. Occasional patients may present with painless jaundice and weight loss mimicking pancreaticobiliary malignant disease (Chapter 194).

DIAGNOSIS

Patients with cholangitis, with or without associated pancreatitis (Chapter 144), typically have elevated serum aminotransferase levels. The serum bilirubin level usually is less than 15 mg/dL with choledocholithiasis because most bile duct stones cause intermittent, incomplete biliary obstruction. In unusual cases, the serum aminotransferase levels can be profoundly elevated (up to 2000 IU/L), mimicking acute viral hepatitis.

Although ultrasound is the most common initial test for patients with suspected gallbladder stones, it has a low sensitivity rate (25 to 60%) for detection of bile duct stones, in part because the bile duct may not be dilated in acute obstruction. CT scan may demonstrate calcified bile duct stones (Fig. 155-4), but its sensitivity for this purpose is generally little better. CT is useful, however, for identifying other potential causes of biliary obstruction (e.g., mass lesion) and local complications, such as a liver abscess (Chapter 151). Magnetic resonance cholangiopancreatography (MRCP) and endoscopic ultrasound can detect bile duct stones with an accuracy comparable to that of ERCP (Fig. 155-5). Because of potential procedure-related risks, ERCP is now reserved for patients with confirmed or a high suspicion of biliary disease who are likely to require therapeutic intervention.

TREATMENT

Rx

Given the potential serious complications of bile duct stones (i.e., cholangitis, pancreatitis), specific therapy is generally required regardless of symptoms.

About 85 to 90% of bile duct stones can be removed at ERCP by standard balloon dilation and basket extraction after biliary endoscopic sphincterotomy, with a complication rate, including pancreatitis, bleeding, cholangitis, cholecystitis, and perforation, of less than 10%. Endoscopic biliary orifice dilation without sphincterotomy may reduce some acute complications but increases the risk of pancreatitis and may lead to more subsequent procedures.

The 10 to 15% of bile duct stones that cannot be removed by standard ERCP are generally larger than 1.5 cm, impacted, or located above a stricture. Alternative therapies include the use of large-diameter dilation balloons (12 to 18 mm) and fragmentation by mechanical or electrohydraulic lithotripsy. Whenever stones cannot be completely removed endoscopically, biliary stents should be placed to ensure adequate biliary drainage and to prevent recurrent symptoms while awaiting further therapy. Long-term biliary stenting can also be used in patients with severe comorbid medical conditions that preclude surgery or repeated endoscopic interventions.

Ideally, patients with concomitant gallbladder and bile duct stones would be best treated with a single laparoscopic cholecystectomy and bile duct exploration, which is preferable to ERCP followed by cholecystectomy.[■] However, only a minority of surgeons can successfully perform laparoscopic bile duct exploration, so open common bile duct exploration is generally performed if endoscopic and laparoscopic approaches are unsuccessful.

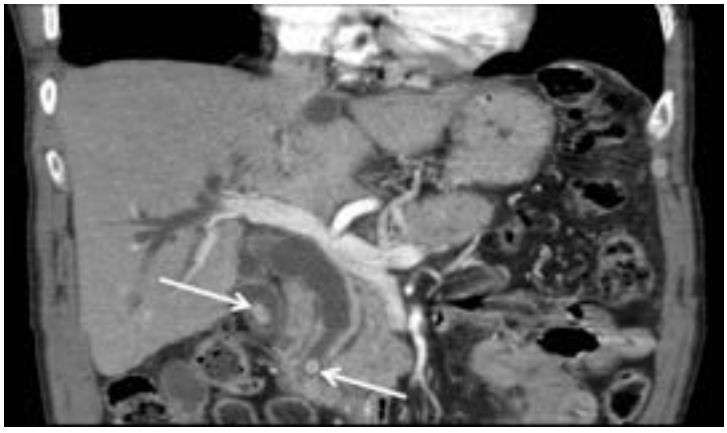

FIGURE 155-4. **Computed tomography scan demonstrating calcified gallstones and a distal bile duct stone.**

Complications of Bile Duct Stones

Cholangitis is a potentially life-threatening disease that results from bacterial infection of obstructed bile. Systemic toxicity occurs when intraductal pressure is sufficiently elevated to cause reflux of bacteria or endotoxin into the blood. About 80 to 90% of acute cholangitis is caused by choledocholithiasis, with the remaining cases caused by a benign biliary stricture (e.g., primary sclerosing cholangitis, chronic pancreatitis, postoperative bile duct injury, or narrowing at an anastomosis) or by malignant biliary obstruction, typically after previous endoscopic instrumentation and stent placement. In certain parts of the world, parasitic biliary obstruction (e.g., *Ascaris*; Chapter 357) may be manifested with cholangitis. The most common bacteria are gram-negative bacilli and *Streptococcus* spp, but *Enterococcus* spp are frequently seen in patients with occluded biliary stents. Prompt antibiotic therapy (e.g., intravenous ceftriaxone, 1 to 2 g once daily; ampicillin-sulbactam, 1.5 to 3 g every 6 hours; piperacillin-tazobactam, 3.375 g every 6 hours; ciprofloxacin, 400 mg twice daily; or levofloxacin, 500 mg orally once daily) is critical and usually can permit conservative management with endoscopic biliary decompression within 24 to 48 hours.[■] However, urgent decompression is indicated if improvement is not seen within a few hours. The advantages of ERCP are that it can delineate the cause of obstruction, obtain bile for culture, and rapidly decompress the biliary tree definitively by removing the stone or temporarily by placing a stent without removing the stone. Routine stenting is not indicated after successful stone removal,[■] unless the adequacy of biliary drainage is uncertain.

Acute gallstone pancreatitis accounts for up to 50% of cases of acute pancreatitis in Western countries (Chapter 144). Most patients quickly respond to conservative therapy, but some develop severe pancreatitis. Although early ERCP with biliary sphincterotomy and stone removal (Fig. 155-6) would appear

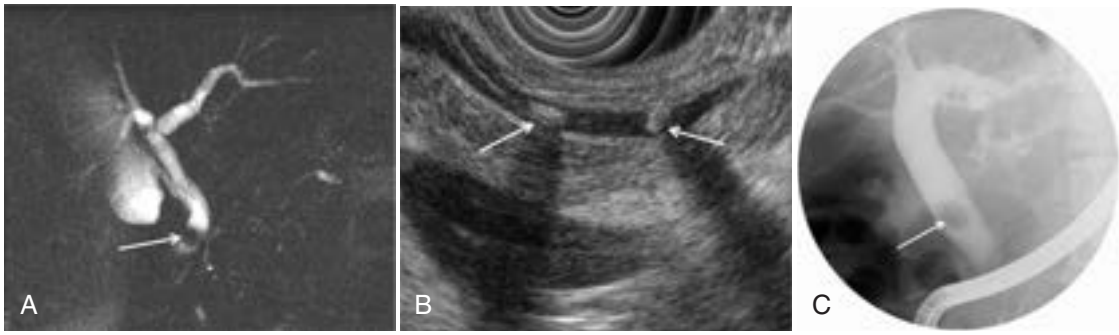

FIGURE 155-5. **Common bile duct stones seen at magnetic resonance cholangiopancreatography (A), endoscopic ultrasound (B), and endoscopic retrograde cholangiopancreatography (C).**

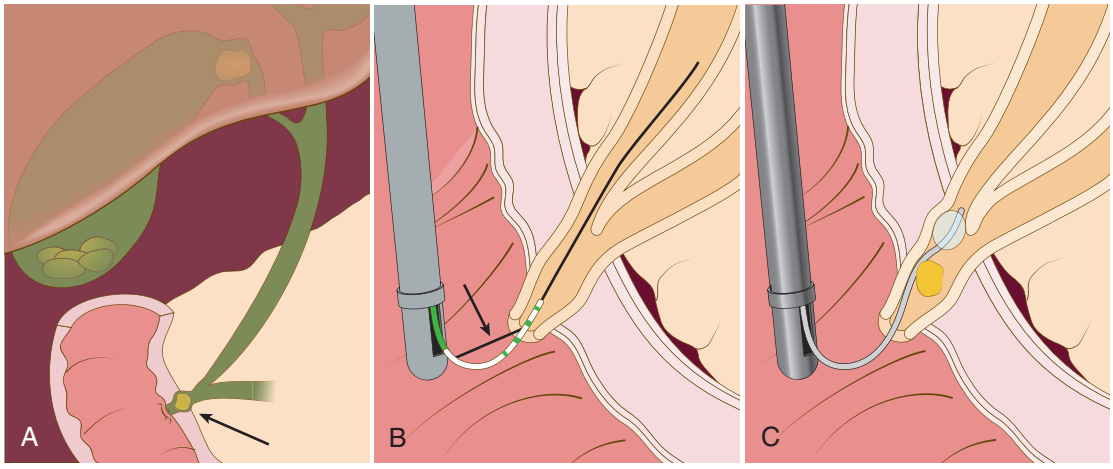

FIGURE 155-6. **A, A stone that is impacted in the distal bile duct is causing biliary pancreatitis, with gallbladder stones also seen. B, Performance of biliary sphincterotomy. C, Subsequent stone removal by balloon sweep.**

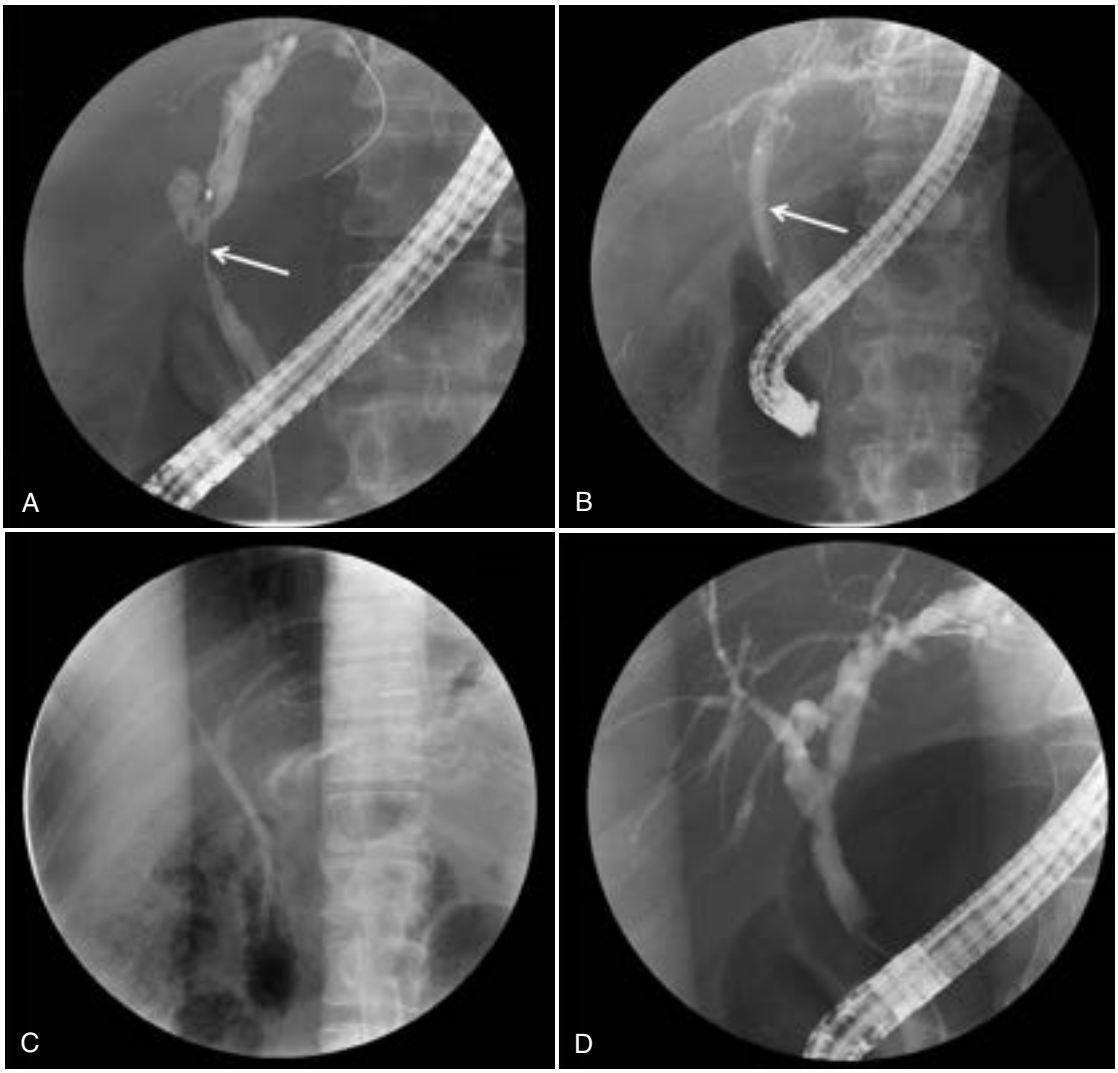

FIGURE 155-7. **Endoscopic retrograde cholangiopancreatography images obtained from a patient who presented with painless jaundice 8 months after cholecystectomy. A, Benign common hepatic duct stricture. B, Balloon dilation of the stricture. C, Multiple stents placed. D, Resolution of stricture after a 1-year stenting interval.**

to be an attractive therapeutic option, early ERCP does not reduce mortality or complications except in patients with biliary obstruction or cholangitis.[3] After recovery from an episode of biliary pancreatitis, laparoscopic cholecystectomy with intraoperative cholangiography is recommended to prevent further episodes, preferably during the same hospital admission. If a common bile duct stone is found at intraoperative cholangiography, laparoscopic or open common bile duct exploration and stone removal can be accomplished with high success rates in experienced hands, or postoperative ERCP can remove any retained stones.

PREVENTION

Up to 25% of patients may have recurrent bile duct stones, with or without a gallbladder, but it remains uncertain what proportion of these recurrent stones are in fact overlooked residual stones from a prior event. A dilated extrahepatic bile duct (≥13 mm) and periampullary diverticula are risk factors for recurrent stones, perhaps by increasing biliary stasis. Identification and treatment of biliary strictures, papillary stenosis, and gallstones in patients with gallbladder in situ are essential for preventing recurrent stones. Unfortunately, no preventive therapy has been proved effective, although ursodeoxycholic acid (15 mg/kg/day) appears to reduce the risk of gallstones during weight loss.

Benign Biliary Strictures

Postoperative extrahepatic bile duct strictures occur after 0.25 to 1% of cholecystectomies. Most such lesions are manifested as abnormal liver test results, obstructive jaundice, and cholangitis within 2 to 3 months postoperatively, although the presentation can be delayed. The cholangiogram commonly shows a short, smooth narrowing near the cystic duct stump with proximal duct dilation (Fig. 155-7). Strictures typically must be redilated, and stents are exchanged at 3- to 4-month intervals for 8 to 12 months until the stricture is nearly as open as the downstream bile duct. About 80% of patients will have a good result,[4] although some patients will ultimately require a bilioenteric bypass. Strictures more than 2 cm in length, strictures with clips placed securely across the duct, or strictures associated with resected segments of duct require surgical intervention. Biliary strictures complicating liver transplantation (Chapter 154) are usually treated similarly with good results.

Intrapancreatic common bile duct strictures, which may occur in 3 to 46% of patients with chronic pancreatitis, can lead to secondary biliary cirrhosis or recurrent cholangitis. With the complication of cholangitis or jaundice, intervention is clearly indicated, typically with ERCP and stent placement. In the absence of cholangitis or jaundice, either surgical repair or endoscopic biliary decompression with multiple plastic stents (Fig. 155-8) is generally recommended when the alkaline phosphatase level is consistently more than twice the upper limit of normal during a 6-month period of observation.

ORIENTAL CHOLANGIOHEPATITIS

Recurrent cholangitis with hepatolithiasis has a prevalence of more than 10% in parts of East Asia, especially in Taiwan, owing to infection with *Ascaris lumbricoides* (Chapter 357) and *Clonorchis sinensis* (Chapter 356). This condition results in local strictures and dilation of the intrahepatic biliary tree. Biliary stasis and subsequent bacterial infection cause brown stones to form. Most patients have recurrent cholangitis, but cholangiocarcinoma can also

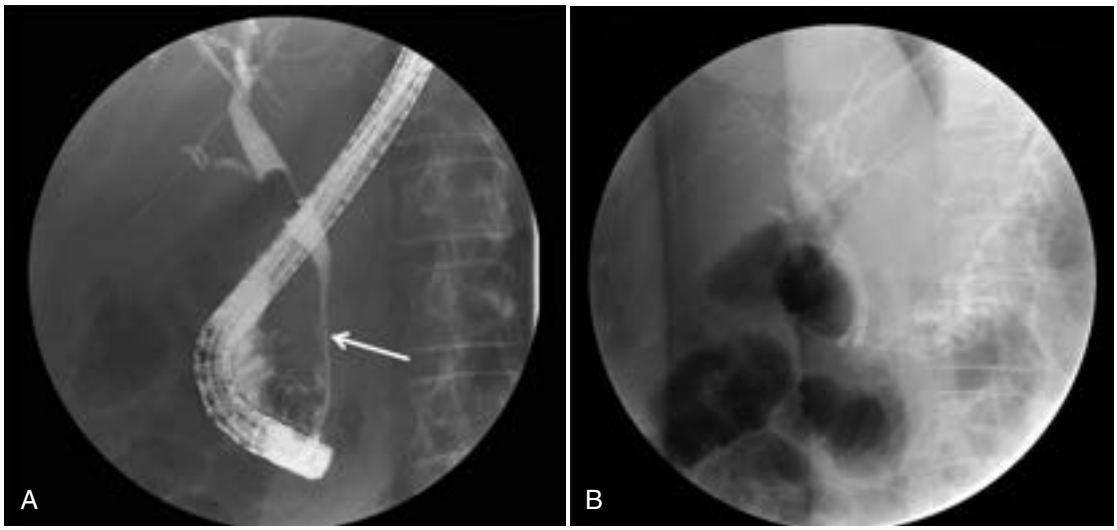

FIGURE 155-8. Endoscopic retrograde cholangiopancreatography images obtained from a patient with a history of alcohol abuse who presented with pruritus and was found to have markedly elevated alkaline phosphatase. A, A smooth distal bile duct stricture is seen *(arrow)*. **B,** Four stents have been placed through the stricture.

ensue. Ultrasound or CT can establish the diagnosis. Treatment includes intravenous fluids and antibiotics. Endoscopic stone removal is usually the preferred treatment, but localized surgical resection targeted to the cultured organisms may be necessary.

PRIMARY SCLEROSING CHOLANGITIS

DEFINITION

Primary sclerosing cholangitis is a chronic cholestatic disease characterized by fibrosing inflammation of segments of the intrahepatic and extrahepatic bile ducts.[5] It results in progressive narrowing of the duct lumen and ultimately may be manifested with recurrent episodes of ascending cholangitis or, alternatively, may progress to secondary biliary cirrhosis and its associated complications. Cholangiocarcinoma (Chapter 196) is a dreaded complication with a reported incidence of 25 to 40% at autopsy or liver transplantation.

EPIDEMIOLOGY

The true prevalence of primary sclerosing cholangitis is unknown, but current estimates are 0.2 to 8.5 per 100,000 in the U.S. population. Its prevalence is much higher in populations in which inflammatory bowel disease (Chapter 141) is more common. Affected men outnumber women in a 2 : 1 ratio, with the mean age at diagnosis of 32 to 40 years. However, primary sclerosing cholangitis has been reported in infants, children, and the elderly.

PATHOBIOLOGY

The cause of primary sclerosing cholangitis and the mechanisms responsible for its progression are unknown.[6] However, autoimmune and genetic causes are supported by its frequent association with inflammatory bowel disease and the increased prevalence of the HLA B8, DR3 haplotype. About two thirds of patients with primary sclerosing cholangitis have ulcerative colitis or Crohn colitis, and it is rarely associated with Crohn disease that is limited to the small bowel. However, only 1 to 13% of patients with colitis are diagnosed with primary sclerosing cholangitis during their lifetime. First-degree relatives of patients with primary sclerosing cholangitis have a 9- to 39-fold increased risk for development of the disease.

On pathologic examination, involved segments of bile ducts in strictured areas show diffuse thickening with a mononuclear inflammatory cell infiltrate. The most characteristic biopsy features are bile duct proliferation, periductal fibrosis, periductal inflammation, and loss of bile ducts. Obliterative cholangitis with a chronic inflammatory cell infiltrate and periductular "onion ring" fibrosis is strongly associated with primary sclerosing cholangitis but is infrequently observed in biopsy specimens. Given the patchy nature of the disease and possible lack of significant intrahepatic involvement, however, the histologic appearance of primary sclerosing cholangitis is variable and may resemble extrahepatic biliary obstruction, chronic active hepatitis, or, rarely, primary biliary cirrhosis.

CLINICAL MANIFESTATIONS

Most patients with primary sclerosing cholangitis are asymptomatic at presentation and are identified after investigation of an elevated alkaline phosphatase level (Chapter 147). Overall, about 90% of patients have an elevated alkaline phosphatase level, with or without mildly elevated serum aminotransferase levels. Fatigue, anorexia, malaise, and weight loss are common but may erroneously be attributed to a patient's inflammatory bowel disease. Patients may exhibit signs or symptoms of cholestatic liver disease, including pruritus, upper abdominal pain, and fever. The serum bilirubin concentration is elevated in only about 40% of patients at presentation. Some patients have anemia, hypoalbuminemia, or hypergammaglobulinemia, and a prolonged international normalized ratio suggests biliary obstruction or synthetic dysfunction. Nearly 90% of patients will have a positive perinuclear antineutrophilic cytoplasmic antibody, but this antibody is also nonspecific and may be found in both ulcerative colitis (Chapter 141) and autoimmune hepatitis (Chapter 149). Antinuclear or anti–smooth muscle antibodies are found in 25% of patients but are not specific, and a positive antimitochondrial antibody suggests primary biliary cirrhosis as the diagnosis.

DIAGNOSIS

Cholangiography is necessary to establish the diagnosis of primary sclerosing cholangitis (Fig. 155-9).[7] MRCP generally is the test of choice,[8] but ERCP may be indicated if MRCP is inconclusive, particularly when the disease is confined to small intrahepatic ducts. The role of biopsy remains uncertain owing to the segmental nature of the disease and the overlap of the histologic features with other disease states. Diffuse multifocal strictures are usually short, with intervening normal or dilated segments that give a beaded appearance. Other frequent findings on cholangiography include pseudodiverticula, mural irregularities, and biliary stones and sludge. Secondary causes of sclerosing cholangitis include obstruction (postoperative, autoimmune cholangiopathy, choledocholithiasis, and recurrent pyogenic cholangitis), ischemic (hepatic artery instillation of the chemotherapeutic agent 5-fluorouracil, radiation, and paroxysmal nocturnal hemoglobinuria), and neoplastic (cholangiocarcinoma, hepatocellular carcinoma, lymphoma, and metastasis).

TREATMENT

Rx

The chronic cholestasis of primary sclerosing cholangitis can be treated with cholestyramine (4 to 8 g/day), ursodeoxycholic acid (15 mg/kg/day), rifampicin (300 to 600 mg/day), or phenobarbital (30 to 120 mg/day) with modest success. Fat-soluble vitamin deficiencies (Chapter 218) must be corrected. The prevalence of osteoporosis (Chapter 243) in primary sclerosing cholangitis is between 4 and 10%, so bone densitometry should be performed at diagnosis and every 2 to 3 years thereafter. Supplementation with oral vitamin D and calcium seems prudent, even in the absence of symptomatic deficiency. Treatment with bisphosphonates (Chapter 243) is reserved for

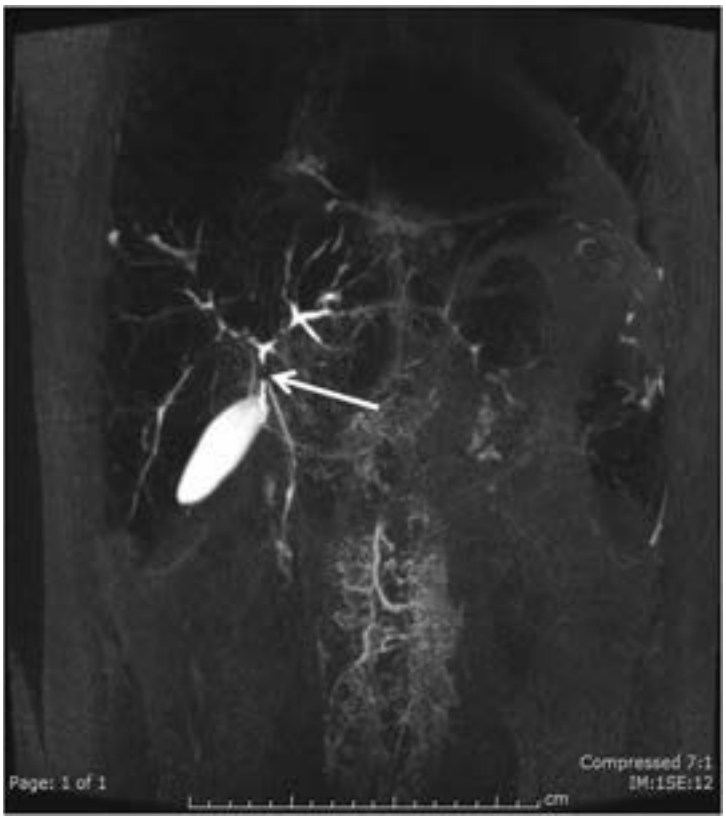

FIGURE 155-9. Magnetic resonance cholangiopancreatography demonstrating the typical cholangiographic features of primary sclerosing cholangitis. Note the narrowed segment of the common bile duct *(arrow)* as well as the diffuse strictures and dilated segments of several intrahepatic bile ducts, giving the classic "beaded" appearance.

patients with confirmed osteoporosis. Unfortunately, no medical treatment slows disease progression; ursodeoxycholic acid, D-penicillamine, corticosteroids, cyclosporine, methotrexate, and colchicine have all been shown to be ineffective for improving survival or delaying the time to liver transplantation. Whether repeated endoscopic treatment to maintain bile duct patency can improve outcomes is unknown.

Liver transplantation (Chapter 154) is the only potentially curative therapy. The 1-year and 5-year survival rates typically are in the 90% and 80% range, respectively. Primary sclerosing cholangitis may recur in the transplanted organ in 15 to 20% of patients.

PROGNOSIS

The natural history of primary sclerosing cholangitis is variable and incompletely understood. Asymptomatic patients have a much better prognosis than symptomatic patients, with 10-year actuarial survival rates of 80% and 50%, respectively. In symptomatic patients, the median time of survival until death or liver transplantation is 9 years, compared with 12 to 18 years for all patients with primary sclerosing cholangitis, regardless of symptoms. An elevated serum bilirubin level and hepatomegaly appear to correlate with a poor prognosis, whereas cholangiographic appearance, the presence or absence of inflammatory bowel disease, and the patient's age do not. Cholangiocarcinoma (Chapter 196) is a dreaded complication of primary sclerosing cholangitis, and the risk appears to be greatest in patients with long-standing ulcerative colitis and cirrhosis.

Choledochal Cysts and Anomalous Pancreaticobiliary Duct Junction

Choledochal cysts are uncommon anomalies of the biliary tree that are manifested as cystic dilation of the intrahepatic or extrahepatic ducts (or both). The incidence is 1 in 100,000 to 150,000 births in Western populations and 1 in 1000 in Asian populations. There is a 3 : 1 to 4 : 1 female-to-male preponderance. These cysts (E-Fig. 155-1) usually involve only the extrahepatic biliary tree, but they can present as extrapancreatic bile duct diverticula, involve only the intraduodenal part of the common bile duct, or present as multiple intrahepatic and extrahepatic cysts (Table 155-3). An anomalous pancreaticobiliary duct junction (E-Fig. 155-2) is frequently associated with choledochal cysts but may be found in isolation, especially in Asian populations.

TABLE 155-3 CLASSIFICATION OF CHOLEDOCHAL CYSTS

TYPES	DESCRIPTION	PROPORTION OF CHOLEDOCHAL CYSTS
I	Segmental or diffuse fusiform dilation of the bile duct	50-80%
II	Choledochal diverticulum	2%
III	Dilation of the intraduodenal portion of the bile duct	1.4-5%
IVa	Multiple intrahepatic and extrahepatic cysts	15-35%
IVb	Multiple extrahepatic cysts	
V (Caroli disease)	Single or multiple dilations of the intrahepatic ducts	20%

Patients with choledochoceles commonly have biliary colic, cholangitis, jaundice, or unexplained pancreatitis. Bile reflux may also result in acute pancreatitis or predispose to biliary cancers.

Cholangiography, preferably by ERCP, is the diagnostic "gold standard," although MRCP can also delineate the anatomy noninvasively. Because of the increased risk for development of biliary tract cancers, cyst resection (including cholecystectomy) is the generally recommended treatment, although it does not eliminate the risk entirely.[9]

Biliary Fistula

A biliary fistula represents an injury to the bile duct, most commonly seen as a complication of cholecystectomy, common bile duct exploration, or inadvertent operative injury of the bile duct or as a consequence of a local infection. Rarely, biliary fistulas result from long-standing untreated biliary tract disease. With more widespread use of laparoscopic cholecystectomy, the incidence of bile duct injury, including biliary fistula, has increased.

Postoperative bile duct leaks are usually manifested within a week after surgery, with patients presenting with abdominal pain (90%), tenderness (80%), fever (75%), nausea and vomiting (50%), and jaundice (40%). Clinically detectable ascites is rare. Biochemical testing is usually nonspecific, with variable elevations in serum liver test values and the white blood cell count.

Patients with suspected biliary fistulas often undergo abdominal ultrasonography or CT to look for evidence of a biloma as well as a hepatobiliary scan to diagnose the leak. However, ERCP is the most sensitive test to detect a biliary fistula. Treatment options for biliary leaks include percutaneously or endoscopically placed biliary drains or stents and surgical drainage and repair of the leak.

Vanishing Bile Duct Syndromes

The vanishing bile duct syndrome is characterized by a paucity of intrahepatic bile ducts, an elevated alkaline phosphatase level, and cholestasis. Causes include primary biliary cirrhosis, primary sclerosing cholangitis, autoimmune hepatitis (Chapter 149), graft-versus-host disease, chronic liver transplant rejection (Chapter 154), ischemia, intrahepatic chemotherapy, drug toxicity (e.g., ampicillin, amoxicillin, flucloxacillin, erythromycin, tetracycline, doxycycline, cotrimoxazole), human immunodeficiency virus (HIV) infection (Chapter 390), sarcoidosis (Chapter 95), and histiocytosis. Ursodeoxycholic acid (15 mg/kg) can increase bile flow, but the condition inexorably progresses to biliary cirrhosis, which ultimately requires liver transplantation.

PRIMARY BILIARY CIRRHOSIS

Primary biliary cirrhosis is an obliterative autoimmune cholangiopathy that involves the small and medium-sized bile ducts and that slowly progresses during a decade or so. As the ducts are obliterated, patients develop cholestasis, fibrosis, and, ultimately, liver failure.[10]

EPIDEMIOLOGY

About 95% of patients with primary biliary cirrhosis are women, and the peak age at onset is between 20 and 60 years. The incidence of the disease may be increasing. In the United States, the estimated annual incidence is about 4.5 per 100,000 per year for women and 0.7 for men. Because of the limited life

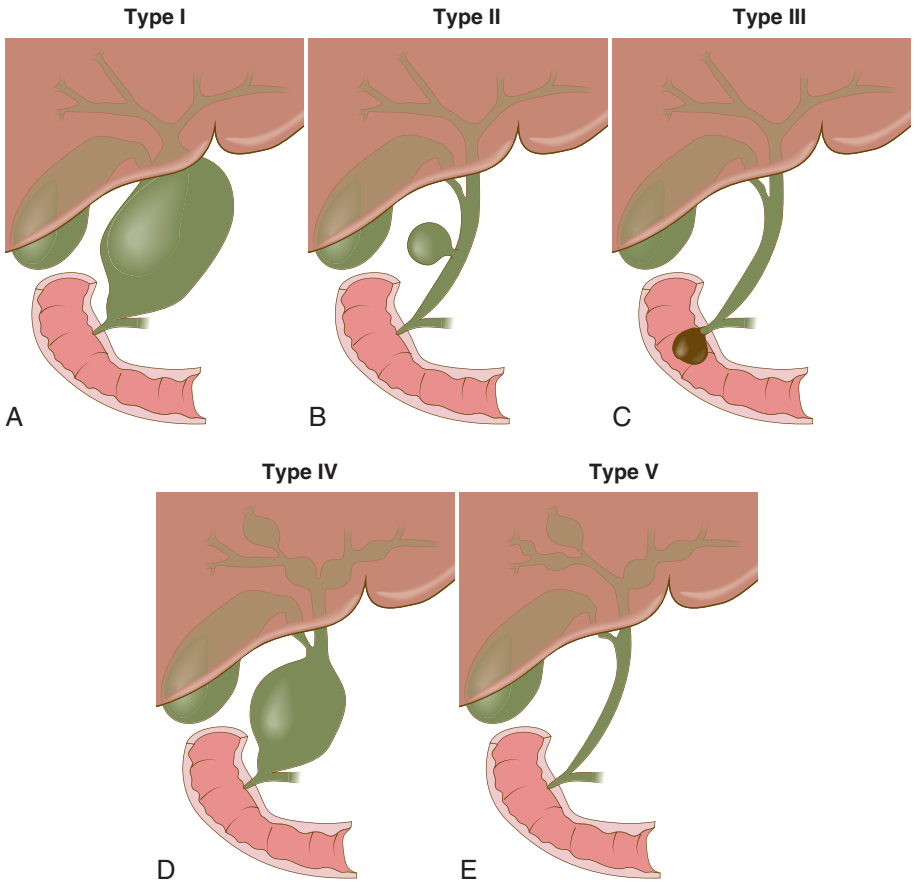

E-FIGURE 155-1. **Schematic diagram of choledochal cysts.**

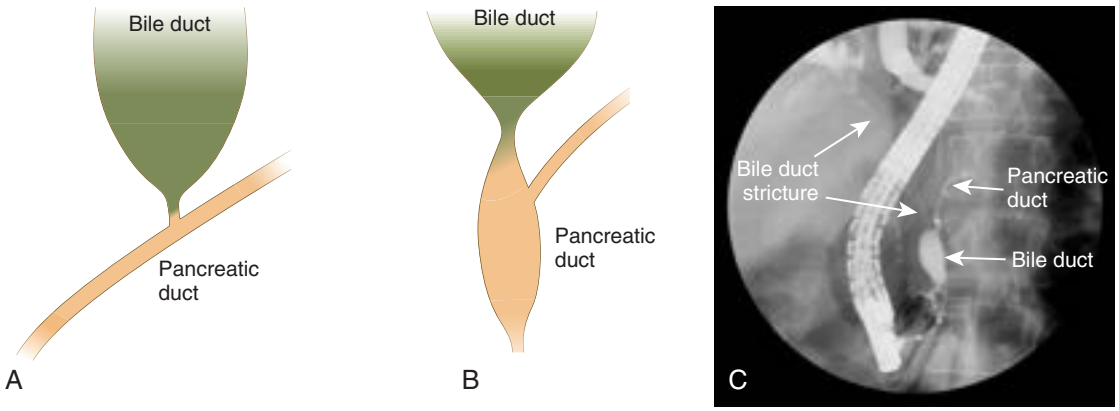

E-FIGURE 155-2. **A,** Schematic diagram of anomalous pancreaticobiliary duct junction, with the terminal bile duct draining directly into the pancreatic duct (B-P type). **B,** Schematic diagram of P-B type of anomalous pancreaticobiliary duct junction. **C,** ERCP image illustrating the P-B type of anomalous pancreaticobiliary duct junction, with a bile duct stricture due to gallbladder cancer.

expectancy of affected patients, the age- and gender-adjusted prevalence of primary biliary cirrhosis is about 65 per 100,000 in women and about 12 per 100,000 in men.

PATHOBIOLOGY

Although the mechanism of progressive destruction of the small interlobular ducts is unknown, primary biliary cirrhosis is considered to be an autoimmune disorder. Genome-wide studies show an association with HLA, interleukin-12A, and interleukin-12RB2 variants, suggesting that interleukin-12 signaling might be important. The disease progresses slowly and can eventually lead to biliary cirrhosis, portal hypertension, and liver failure. The classic histologic finding is noncaseating granulomas and paucity of bile ducts in the portal tracts.

CLINICAL MANIFESTATIONS

The most common symptoms are fatigue (50%), which can be debilitating and is unrelated to the degree of underlying liver disease, and pruritus (30%), but about 50% of patients are asymptomatic at the time of diagnosis. Many patients are initially seen by dermatologists for pruritus, which may be first noticed in pregnancy but persists after delivery.

Autoimmune syndromes associated with primary biliary cirrhosis include autoimmune thyroid dysfunction (Chapter 226), Sjögren syndrome (Chapter 268), Raynaud phenomenon (Chapter 267), and celiac disease (Chapter 140). Vitamin D malabsorption can also lead to metabolic bone disease (Chapters 243 and 244).

DIAGNOSIS

The first clue to primary biliary cirrhosis is an elevated serum alkaline phosphatase level, which should be confirmed by an elevated γ-glutamyl transpeptidase level (Chapter 147). The antimitochondrial antibody level has a sensitivity and specificity of more than 95% when the titer is higher than 1 : 40, and it may be positive even before there is any clinical or biochemical evidence of the disease. By comparison, the bilirubin level often is not elevated until later in the course of the disease, with most of the elevation typically due to an elevation in conjugated bilirubin. Total immunoglobulins are generally normal, but IgM levels can be elevated.

An ultrasound examination of the biliary tree is critical to confirm the absence of extrahepatic disease. A liver biopsy is occasionally needed to confirm the diagnosis, particularly in antimitochondrial antibody–negative patients, and to stage the disease.

TREATMENT

Rx

Ursodeoxycholic acid therapy (12 to 15 mg/kg) improves serum bilirubin, alkaline phosphatase, and cholesterol levels and has a variable effect on pruritus. Unfortunately, it does not relieve fatigue, reduce mortality, or delay the need for liver transplantation. Bezafibrate therapy also generally improves liver chemistry test results but not pruritus, liver-related mortality, or overall mortality.

There is no definite benefit from steroids, colchicine, azathioprine, or methotrexate. Bisphosphonates (Chapter 243) are commonly prescribed for the accompanying metabolic bone disease, but their benefit is uncertain. Liver transplantation (Chapter 154) is indicated for refractory disease. The posttransplantation prognosis is excellent, with 2-year and 5-year survival rates of 80% and 70%, respectively. However, studies suggest an 8 to 40% recurrence in the transplanted liver.

PROGNOSIS

Up to two thirds of asymptomatic patients become symptomatic within 2 to 4 years. Significant bridging fibrosis or cirrhosis on biopsy carries a worse prognosis. Prognosis is also influenced by the serum bilirubin and albumin levels, the international normalized ratio, older age, and the presence of peripheral edema. Liver failure develops in about 25% of patients within 10 years after diagnosis, and median survival after diagnosis is 12 to 15 years.

Bile Duct Tumors

BENIGN

Benign bile duct tumors are exceedingly rare compared with malignant tumors and are much less common than benign gallbladder tumors. They can be divided into three histologic types (papillomas, adenomas, cystadenomas), may be solitary or multiple, and often are found incidentally during the evaluation of bile duct dilation or intraductal filling defects. Patients may be asymptomatic or have symptoms of biliary obstruction. Treatment typically consists of surgical bile duct resection with hepaticojejunostomy reconstruction. Even benign tumors tend to recur after excision, and some undergo malignant change.

CHOLANGIOCARCINOMA

Cholangiocarcinoma is discussed in Chapter 196.

Ampullary Tumors

Benign ampullary lesions seen on endoscopic or radiologic studies include heterotopic gastric mucosa, a lipoma, or an impacted common bile duct stone. Primary tumors of the ampulla of Vater can be premalignant or malignant, but the overwhelming majority (>95%) are either adenomas or adenocarcinomas. The prevalence of ampullary adenomas has been estimated to be 0.04 to 0.12% in autopsy series, but the prevalence is higher in patients with hereditary polyposis syndromes (Chapter 193), in which ampullary adenomas occur in up to 80% of individuals and progress to malignancy in 4%.

Malignant ampullary lesions are most commonly adenocarcinomas, although metastatic breast cancer, renal cell cancers, and melanomas also have been identified. Carcinoid tumors (Chapter 232) and other neuroendocrine tumors are rare. Ampullary adenomas probably follow an adenoma to carcinoma sequence similar to colorectal adenocarcinoma (Chapter 193), with a 25 to 85% risk of transformation to carcinoma.

Patients with ampullary lesions may present with biliary colic, obstructive jaundice, pancreatitis, or nonspecific upper abdominal pain, with or without fluctuating serum liver test results, malaise, and anorexia. However, ampullary lesions are often found incidentally on cross-sectional imaging or during upper endoscopy performed for a different indication.[11]

For ampullary adenomas, options include observation with surveillance biopsies and attempts to resect the lesion completely through endoscopy or surgery. Surveillance of an ampullary adenoma in the setting of familial adenomatous polyposis is reasonable if the lesion is small (<1 cm) and asymptomatic, but resection is preferred if advanced histology (e.g., villous features, dysplasia) is identified. Surgical resection has been the standard for ampullary tumors. Treatment modalities include pancreatoduodenectomy, which has a high rate of morbid and even fatal complications, and transduodenal excision, which is associated with high recurrence rates. The 5-year survival rates after pancreaticoduodenectomy range from 64 to 80% for patients with node-negative disease and from 17 to 50% for node-positive disease. Limited data exist regarding adjuvant therapy, the benefits of which are uncertain. A common practice is to treat these patients in a manner similar to patients with resected pancreatic adenocarcinomas (Chapter 194). Patients who present with unresectable disease tend to receive combination gemcitabine (1000 mg/m^2) plus cisplatin (25 mg/m^2) chemotherapy, each administered on days 1 and 8, every 3 weeks for eight cycles. Palliative biliary stenting can be performed in individuals who have a short life expectancy.

In patients with small, localized, and clearly benign ampullary adenomas, endoscopic resection represents an alternative to surgical therapy in appropriately selected patients (E-Fig. 155-3). Whether endoscopic resection is effective for larger or higher risk adenomas is uncertain. Furthermore, recurrence rates after endoscopic papillectomy approach 20%, thereby emphasizing the need for careful follow-up endoscopic surveillance.

Sphincter of Oddi Dysfunction

Sphincter of Oddi dysfunction is a benign, noncalculous obstruction to flow of bile or pancreatic juice through the pancreaticobiliary junction. It may be manifested clinically by pain, pancreatitis (Chapter 144), abnormal liver test results, or abnormal pancreatic enzymes. Post-cholecystectomy pain resembling the patient's preoperative biliary colic occurs in at least 10 to 20% of patients.

Evaluation of patients with suspected sphincter of Oddi dysfunction includes standard serum liver chemistries, serum amylase and lipase levels, and an abdominal ultrasound examination or CT scan. The specimens for serum enzyme studies should be drawn during bouts of pain, if possible. Mild elevations (less than two times the upper limits of normal) are frequent in sphincter of Oddi dysfunction, whereas greater abnormalities are more suggestive of stones, tumors, and parenchymal liver disease. The findings on CT and abdominal ultrasound studies are usually normal, but abnormal liver or pancreatic enyzmes or a dilated bile duct or pancreatic duct may occasionally be found. ERCP and sphincter of Oddi manometry may be considered in patients who have objective evidence of pancreatic or biliary disease

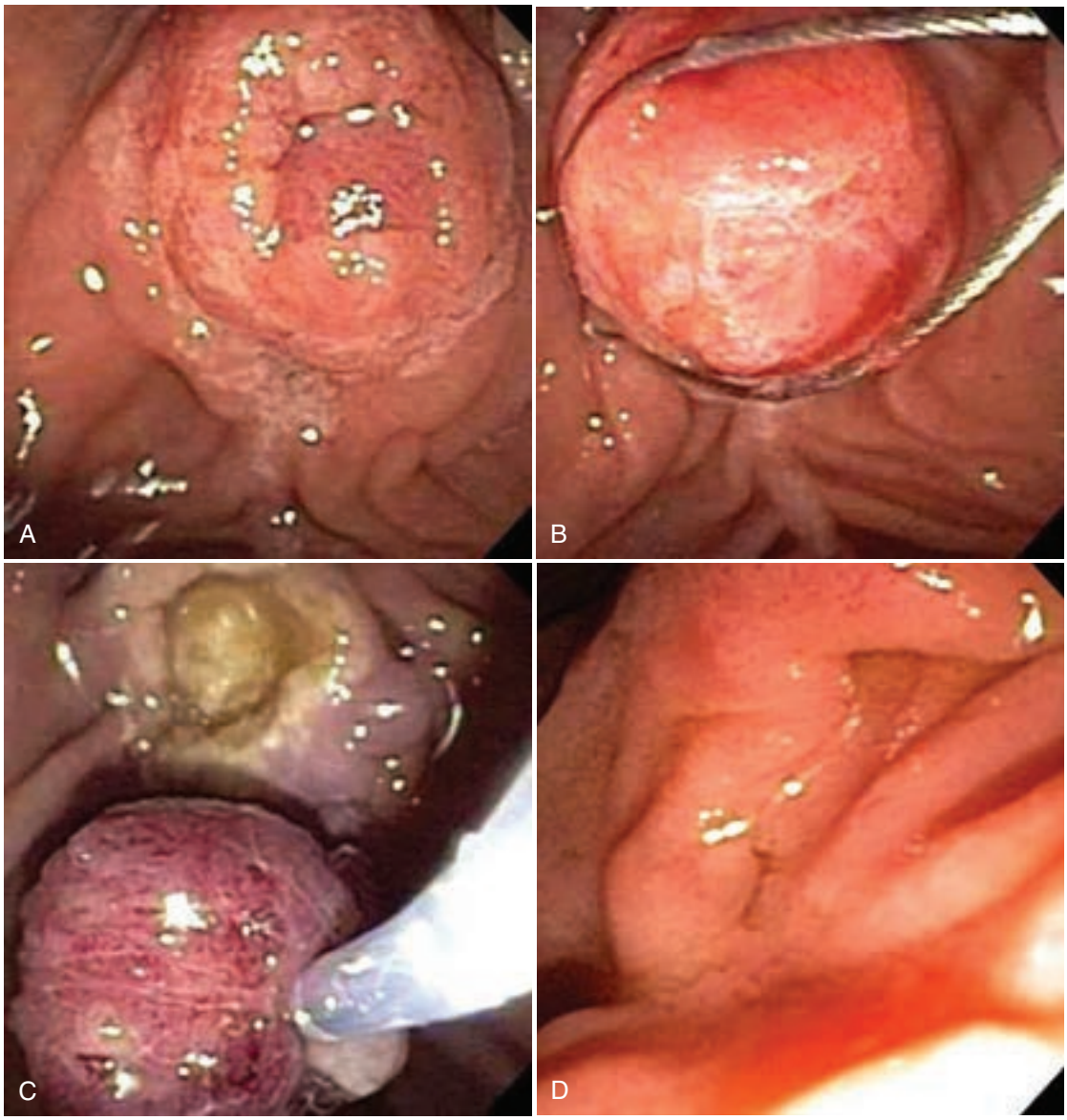

E-FIGURE 155-3. A 58-year-old man was found to have a prominent major papilla during upper endoscopy performed for evaluation of dysphagia. Duodenoscopy confirmed these findings (**A**), and biopsies revealed this to be a tubulovillous adenoma. This was grasped with a polypectomy snare (**B**), removed in one piece by electrocautery, and retrieved for pathologic analysis (**C**). A temporary pancreatic duct stent was placed (not shown), and rectal indomethacin was given to minimize the risk of postprocedure pancreatitis. No residual adenoma was seen on surveillance endoscopy 1 year later (**D**).

(abnormal liver or pancreatic enzymes or a dilated bile or pancreatic duct) or clinically significant or disabling symptoms and in whom definitive sphincter ablation is planned if abnormal sphincter function is found. In patients without objective evidence of pancreatic or biliary disease, ERCP and manometry are no longer recommended.

Medical therapy with nonspecific antispasmodics (e.g., dicyclomine, 10 to 20 mg every 6 hours; hyoscyamine, 0.375 mg every 12 hours) or smooth muscle relaxants (e.g., nifedipine, 60 mg daily) for a 1-month trial should be considered in patients with pancreaticobiliary-type pain (with or without abnormal liver enzyme, amylase, or lipase levels) or a dilated bile or pancreatic duct. If patients do not respond satisfactorily, ERCP and endoscopic sphincterotomy can improve pain in 55 to 95% of patients with abnormal findings on laboratory testing or abdominal imaging. Alternatively, only about 25% of patients without objective evidence of pancreatic or biliary disease improve after sphincterotomy.

Grade A References

A1. Gurusamy KS, Koti R, Fusai G, et al. Early versus delayed laparoscopic cholecystectomy for uncomplicated biliary colic. *Cochrane Database Syst Rev*. 2013;6:CD007196.

A2. Vaughan J, Gurusamy KS, Davidson BR. Day-surgery versus overnight stay surgery for laparoscopic cholecystectomy. *Cochrane Database Syst Rev*. 2013;7:CD006798.

A3. Gurusamy KS, Davidson C, Gluud C, et al. Early versus delayed laparoscopic cholecystectomy for people with acute cholecystitis. *Cochrane Database Syst Rev*. 2013;6:CD005440.

A4. Iranmanesh P, Frossard JL, Mugnier-Konrad B, et al. Initial cholecystectomy vs sequential common duct endoscopic assessment and subsequent cholecystectomy for suspected gallstone migration: a randomized clinical trial. *JAMA*. 2014;312:137-144.

A5. Stokes CS, Gluud LL, Casper M, et al. Ursodeoxycholic acid and diets higher in fat prevent gallbladder stones during weight loss: a meta-analysis of randomized controlled trials. *Clin Gastroenterol Hepatol*. 2014;12:1090-1100.

A6. Ding G, Cai W, Qin M. Single-stage vs. two-stage management for concomitant gallstones and common bile duct stones: a prospective randomized trial with long-term follow-up. *J Gastrointest Surg*. 2014;18:947-951.

A7. Teoh AY, Cheung FK, Hu B, et al. Randomized trial of endoscopic sphincterotomy with balloon dilation versus endoscopic sphincterotomy alone for removal of bile duct stones. *Gastroenterology*. 2013;144:341-345.

A8. Zhang RL, Zhao H, Dai YM, et al. Endoscopic nasobiliary drainage with sphincterotomy in acute obstructive cholangitis: a prospective randomized controlled trial. *J Dig Dis*. 2014;15:78-84.

A9. Gurusamy KS, Nagendran M, Davidson BR. Early versus delayed laparoscopic cholecystectomy for acute gallstone pancreatitis. *Cochrane Database Syst Rev*. 2013;9:CD010326.

A10. Lindor KD, Kowdley KV, Luketic VA, et al. High-dose ursodeoxycholic acid for the treatment of primary sclerosing cholangitis. *Hepatology*. 2009;50:808-814.

A11. Rudic JS, Poropat G, Krstic MN, et al. Ursodeoxycholic acid for primary biliary cirrhosis. *Cochrane Database Syst Rev*. 2012;12:CD000551.

A12. Rudic JS, Poropat G, Krstic MN, et al. Bezafibrate for primary biliary cirrhosis. *Cochrane Database Syst Rev*. 2012;1:CD009145.

A13. Rudic JS, Giljaca V, Krstic MN, et al. Bisphosphonates for osteoporosis in primary biliary cirrhosis. *Cochrane Database Syst Rev*. 2011;12:CD009144.

A14. Valle J, Wasan H, Palmer DH, et al. Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. *N Eng J Med*. 2010;362:1273-1281.

A15. Cotton PB, Durkalski V, Romagnuolo J, et al. Effect of endoscopic sphincterotomy for suspected sphincter of Oddi dysfunction on pain-related disability following cholecystectomy: the EPISOD randomized clinical trial. *JAMA*. 2014;311:2101-2109.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Yarmish GM, Smith MP, Rosen MP, et al. ACR Appropriateness criteria right upper quadrant pain. *J Am Coll Radiol*. 2014;11:316-322.
2. Hansel SL, DiBaise JK. Functional gallbladder disorder: gallbladder dyskinesia. *Gastroenterol Clin North Am*. 2010;39:369-379.
3. Fogel EL, Sherman S. ERCP for gallstone pancreatitis. *N Engl J Med*. 2014;370:150-157.
4. Garcia-Cano J. Endoscopic management of benign biliary strictures. *Curr Gastroenterol Rep*. 2013;15:336.
5. Hirschfield GM, Karlsen TH, Lindor KD, et al. Primary sclerosing cholangitis. *Lancet*. 2013; 382:1587-1599.
6. Eaton JE, Talwalkar JA, Lazaridis KN, et al. Pathogenesis of primary sclerosing cholangitis and advances in diagnosis and management. *Gastroenterology*. 2013;145:521-536.
7. Chapman R, Fevery J, Kalloo A, et al. Diagnosis and management of primary sclerosing cholangitis. *Hepatology*. 2010;51:660-678.
8. Dave M, Elmunzer BJ, Dwamena BA, et al. Primary sclerosing cholangitis: meta-analysis of diagnostic performance of MR cholangiopancreatography. *Radiology*. 2010;256:387-396.
9. Ohashi T, Wakai T, Kubota M, et al. Risk of subsequent biliary malignancy in patients undergoing cyst excision for congenital choledochal cysts. *J Gastroenterol Hepatol*. 2013;28:243-247.
10. Yimam KK, Bowlus CL. Diagnosis and classification of primary sclerosing cholangitis. *Autoimmun Rev*. 2014;13:445-450.
11. El Hajj II, Cote GA. Endoscopic diagnosis and management of ampullary lesions. *Gastrointest Endosc Clin N Am*. 2013;23:95-109.

REVIEW QUESTIONS

1. A 55-year-old woman is found to have several large calcifications in the right upper quadrant on plain radiographs of her spine, performed as part of her routine chiropractic examination. She denies abdominal pain and has normal serum liver test results. She has diabetes mellitus but no other significant comorbidities. Subsequent abdominal ultrasound reveals a 3.8-cm gallbladder stone, three smaller stones each less than 10 mm in diameter, a normal gallbladder wall, and a nondilated bile duct. The *most appropriate* next step in the management of this patient is

A. Magnetic resonance imaging/magnetic resonance cholangiopancreatography (MRI/MRCP) for further evaluation of her biliary tree
B. Oral dissolution therapy with ursodeoxycholic acid
C. Surgical consultation for elective laparoscopic cholecystectomy
D. Extracorporeal shock wave lithotripsy
E. Continue a conservative watch-and-wait approach, with surveillance ultrasound in 12 months

Answer: C As discussed for asymptomatic gallstones, gallbladder stones larger than 3 cm are associated with an increased incidence of acute cholecystitis, and there is a 10-fold increased risk of gallbladder cancer. Prophylactic cholecystectomy is recommended in this scenario. MRI/MRCP may be considered, but it is unlikely to add additional information in this patient because significant bile duct disease is less likely in the setting of a nondilated bile duct on ultrasound, with normal liver test results. The nonoperative methods listed (ursodeoxycholic acid, extracorporeal shock wave lithotripsy) are not options in patients with gallbladder stones of this size. A conservative watch-and-wait approach may be advocated by some in this asymptomatic patient, but given the increased cancer risk and her lack of significant comorbidities, surgery is the better option here. Furthermore, there is no evidence to support surveillance ultrasound in 12 months.

2. A 44-year-old woman with a history of recurrent urinary tract infections undergoes abdominal/renal ultrasound for further evaluation. She previously underwent laparoscopic cholecystectomy for gallbladder stones 20 years ago. Ultrasound reveals normal kidneys without hydronephrosis, but an incidental single 8-mm shadow suggesting a common bile duct stone is seen. The bile duct is 11 mm in diameter. She denies upper abdominal pain but has noted nausea and dark urine recently, which she has attributed to a recurrent urinary tract infection. She also notes diffuse pruritus. Her serum liver test results are as follows: aspartate transaminase 78 IU/L (normal, <40 IU/L), alanine transaminase 95 IU/L (<45 IU/L), alkaline phosphatase 235 IU/L (<125 IU/L), bilirubin 3.2 mg/dL (<1.0 mg/dL). Which of the following statements is the *most correct*?

A. Magnetic resonance cholangiopancreatography (MRCP) should be performed to confirm the finding of a bile duct stone, given the poor specificity of transabdominal ultrasound for detection of choledocholithiasis.
B. The bile duct stone is most likely to be a cholesterol stone.
C. In the absence of abdominal pain, a watch-and-wait approach is justified in this scenario.
D. Proceed directly to endoscopic retrograde cholangiopancreatography (ERCP) for attempted bile duct stone removal.
E. Perform a surgical exploration.

Answer: D Although transabdominal ultrasound has a low sensitivity for detection of common bile duct stones, its specificity is more than 90% in most series. MRCP is unnecessary in this case because the patient also has elevated liver test values and biliary dilation. Choledocholithiasis occurring 20 years after cholecystectomy probably represents a primary bile duct stone, which is likely to be a brown pigment stone rather than a cholesterol stone. Given the potential for severe, even life-threatening complications of bile duct stones (cholangitis, pancreatitis), a watch-and-wait approach is not recommended. Proceeding with ERCP and stone removal is the most appropriate next step in this scenario.

3. Regarding primary sclerosing cholangitis (PSC), which of the following statements is *true*?

A. Women outnumber men in a 2 : 1 ratio.
B. A liver biopsy demonstrating characteristic findings is necessary to establish the diagnosis of PSC in all affected patients.
C. Ursodeoxycholic acid is the only medical therapy that has been shown to slow the progression of PSC.
D. The risk for development of cholangiocarcinoma increases with the duration of the disease.
E. Current guidelines suggest that bone densitometry be performed in all PSC patients at diagnosis and every 2 to 3 years thereafter.

Answer: E Affected PSC patients are at risk for hepatic osteodystrophy, a metabolic bone disease that may be seen in chronic liver disease. The American Association for the Study of Liver Diseases (AASLD) has suggested that bone densitometry be performed to screen for this disorder. PSC is seen more commonly in men than in women in a 2 : 1 ratio. Cholangiography is necessary to establish the diagnosis of PSC because liver biopsy findings are nonspecific. Only in patients with "small-duct PSC," in which cholangiography is normal, may the liver biopsy suggest the diagnosis in association with other associated features (e.g., presence of inflammatory bowel disease, elevated alkaline phosphatase). No medical therapy has been shown to slow the progression of PSC, and recent AASLD guidelines do not support the routine use of ursodeoxycholic acid in these patients. Interestingly, there is an inverse relationship between incidence of cholangiocarcinoma and duration of PSC; PSC patients who have had the disease for a shorter time appear to be at increased risk for the development of this complication.

4. A 35-year-old woman presents to you for evaluation of a 6-month history of episodic, postprandial right upper quadrant pain. Her pain is identical to pain that previously responded to cholecystectomy 4 years ago for a functional gallbladder disorder. No gallbladder stones were identified at surgery. Aspartate transaminase and alanine transaminase were both found to be two to three times elevated during two recent emergency department visits but are normal on days when she is pain free. Alkaline phosphatase, bilirubin, amylase, and lipase levels are repeatedly normal. You obtain MRI/MRCP, which reveals a normal pancreas and biliary tree without intraductal filling defects or duct dilation. Findings on upper endoscopy and endoscopic ultrasound are normal. She continues to work full-time, without days missed from work, need for hospitalization, or use of narcotics. Which of the following statements is *correct*?

A. A common bile duct stone is the most likely diagnosis.
B. A liver biopsy is recommended as the next investigation.
C. She may have type III sphincter of Oddi dysfunction.
D. A trial of nifedipine or other antispasmodic is suggested before proceeding with ERCP.
E. This patient is at low-risk for the development of post-ERCP pancreatitis.

Answer: D Bile duct stones can occur after cholecystectomy but are rarely found in a patient who did not have gallbladder stones at cholecystectomy, especially in the absence of factors that increase the likelihood of bile stasis (see Bile Duct Stones, Pathobiology). Furthermore, normal findings on MRCP and endoscopic ultrasound make this diagnosis much less likely. Given the elevated liver test values associated with abdominal pain, but not during normal pain-free intervals, a biliary etiology is more likely than hepatocellular disease. Liver biopsy, therefore, is less likely to be helpful in the diagnostic evaluation of this patient. Biliary-type pain associated with elevated liver test values during attacks of pain and a nondilated bile duct on abdominal imaging defines type II sphincter of Oddi dysfunction. Medical trials with nonspecific antispasmodics or smooth muscle relaxants (e.g., calcium-channel blockers) should be considered in less severely symptomatic patients before proceeding with ERCP and sphincter of Oddi manometry as patients with suspected sphincter of Oddi dysfunction are at high risk for post-ERCP pancreatitis.

5. A 44-year-old woman with a remote history of alcohol abuse is brought to her local emergency department for evaluation after her family noted that her eyes and skin were yellow. She has been abstinent from alcohol for 15 years. She notes mild pruritus but is otherwise asymptomatic and has lost no weight. She previously had complained of episodic abdominal pain and was found to have a decreased gallbladder ejection fraction, leading to laparoscopic cholecystectomy 6 months earlier. Ultrasound did not reveal gallbladder stones. Her pain completely resolved postoperatively. Laboratory tests in the emergency department reveal a normal complete blood count, serum bilirubin 6.0 mg/dL (normal, <1 mg/dL), alkaline phosphatase 722 IU/L (<125 IU/L), aspartate transaminase 93 IU/L (<45 IU/L), alanine transaminase101 IU/L (<40 IU/L), amylase 64 IU/L (<125 IU/L), and CA19-9 367 U/mL (<37 U/mL). A computed tomography scan with intravenous administration of contrast material reveals dilated intrahepatic bile ducts proximal to the mid–hepatic duct, adjacent to the cholecystectomy clips, as well as a normal distal bile duct and pancreas, without stones or mass identified. Her pancreatic duct is normal. Which of the following statements is *true*?

A. Choledocholithiasis is the most likely diagnosis.
B. The elevated CA19-9 suggests that the most likely diagnosis is cholangiocarcinoma.
C. MRI/MRCP would be expected to identify a type I choledochal cyst in this scenario.
D. The presentation 6 months after cholecystectomy is consistent with a postoperative bile duct injury.
E. Your most likely diagnosis is a benign biliary stricture from chronic pancreatitis.

Answer: D Bile duct injuries after laparoscopic cholecystectomy may be manifested in the immediate postoperative period (e.g., bile duct transection or when a surgical clip is placed over the common bile duct) or delayed, when thermal or ischemic injury may result in a biliary stricture. Ischemic injuries most commonly occur within a few months of surgery but may be manifested longer than 1 year postoperatively. Although primary bile duct stones may occur after cholecystectomy, they are very unlikely when gallbladder stones are not identified before surgery, as in this case. An elevated CA19-9 may be seen in any tumor that causes biliary obstruction (not specific to cholangiocarcinoma) as well as in benign obstructing lesions (bile duct stones, primary sclerosing cholangitis). The computed tomography findings of biliary dilation adjacent to the surgical clips and a nondilated distal bile duct suggest an obstructing lesion in the proximal bile duct (e.g., stricture, stone), not a type I choledochal cyst. A patient with a history of alcohol abuse might present with cholestatic liver test results and painless jaundice from a biliary stricture secondary to chronic pancreatitis, but such a complication would be less likely in a patient who has been abstinent for 15 years. Furthermore, cross-sectional imaging or cholangiography would be expected to reveal a distal biliary stricture within the head of the pancreas or perhaps with a dilated bile duct proximally, not the imaging findings described in this case.

156 HEMATOPOIESIS AND HEMATOPOIETIC GROWTH FACTORS

KENNETH KAUSHANSKY

Hematopoiesis is the process by which bone marrow stem cells develop into all of the cell types present in the blood (erythrocytes, neutrophils, eosinophils, basophils, monocytes, platelets, T lymphocytes, B lymphocytes, natural killer cells) (Fig. 156-1). The regulation of the numbers of each cell type is carefully controlled by paracrine and endocrine hematopoietic growth factors, which exert antiapoptotic, proliferative, and differentiative effects on hematopoietic stem, progenitor, and maturing blood cells. Many of these glycoproteins are produced by recombinant DNA technology and have been among the most successful therapeutics in modern medicine.

HEMATOPOIETIC STEM AND PROGENITOR CELLS

Hematopoietic stem cells comprise one in 10^5 to 10^6 marrow cells and are not morphologically distinguishable from other progenitors or small lymphocytes but can be purified to homogeneity using physical characteristics and combinations of monoclonal antibodies (including $CD34^+$) to cell surface proteins.[1] The two critical characteristics of a hematopoietic stem cell are its ability to differentiate into all blood cell types and to self-renew. The decision to self-renew or differentiate is a stochastic process, at the stem cell stage and at the subsequent multipotent or unipotent stages of differentiation, that can be influenced by a number of cell extrinsic (growth factors and stromal proteins) and cell intrinsic (transcription factors) molecules. Hematopoietic stems cells reside in specialized microenvironments (niches) within the bone marrow. The complex and diverse stromal cell populations that comprise the stem cell niche provide signals that support critical stem cell properties such as maintenance, self-renewal capacity, and long-term multilineage repopulation ability.[2]

HEMATOPOIETIC CELL EXPANSION: HEMATOPOIETIC GROWTH FACTORS

A large number of transcription factors regulate stem cell number and differentiation state. Several molecular switches have been identified that determine hematopoietic cell fate.

Equally important to hematopoiesis is a group of hematopoietic growth factors that share structural homology and bind to nonredundant type I transmembrane proteins belonging to the cytokine receptor family. Many of these proteins are the physiological regulators of a specific lineage of blood cells (e.g., erythropoietin, granulocyte colony-stimulating factor, thrombopoietin); others appear to represent redundant hematopoietic growth-promoting activities of molecules essential for other biologic functions (e.g., interleukin-3 [IL-3], interleukin-11 [IL-11], granulocyte-macrophage colony-stimulating factor).

Erythropoietin is produced predominantly by the kidneys and to a lesser extent in the liver and acts on marrow erythroid progenitors to enhance their survival, proliferation, and differentiation. Levels of erythropoietin are inversely related to hemoglobin concentrations in the blood, as reflected in renal oxygen tension. In the presence of tissue (renal) hypoxia, the transcription factor hypoxia-induced factor (HIF) 1α is stabilized against proteasome-mediated destruction and drives erythropoietin transcription by binding to a critical hypoxia responsive element located in the 3′ untranslated region of the gene. Genetic elimination of erythropoietin or its receptor results in embryonic lethality, establishing that although other cytokines can influence erythropoiesis, red blood cell production is absolutely dependent on the hormone.

Granulocyte colony-stimulating factor stimulates the production of neutrophils from their marrow progenitors (Fig. 156-2). Levels of the hormone are also inversely related to neutrophil numbers but are regulated primarily by inflammatory stimuli, including tumor necrosis factor-α (TNF-α) and IL-1α acting on endothelial cells, fibroblasts, and macrophages. Similar to the action of erythropoietin on erythroid progenitors, granulocyte colony-stimulating factor acts to enhance the survival, proliferation, and differentiation of neutrophil progenitors. In addition, the cytokine acts to functionally activate the mature cells it helps to produce. Genetic elimination of granulocyte colony-stimulating factor or its receptor in mice reduces neutrophil levels to 25% of normal, the only hormone known to exert this great an impact on granulopoiesis.

Thrombopoietin, the primary regulator of platelet production, is produced in the liver and kidney and by marrow stromal cells and is regulated by both platelet receptor–mediated uptake and destruction and by transcriptional feedback inhibition of the thrombopoietin gene in marrow stromal cells by platelet granule proteins.[3] The plasma levels of free thrombopoietin are normally inversely related to bone marrow megakaryocyte (MK) and platelet mass. In a hypoproliferative thrombocytopenia, therefore, the resultant increase in plasma free thrombopoietin that is not bound to its MK and platelet receptors will drive a compensatory increase in thrombopoiesis. Similar to granulocyte colony-stimulating factor, thrombopoietin stimulates the survival, proliferation, and differentiation of its corresponding lineage, MKs and their precursors, and primes mature platelets to respond to platelet activation agonists. Genetic elimination of thrombopoietin or its receptor in mice or congenital nonsense or missense mutations in the gene for the thrombopoietin receptor in humans result in platelet levels approximately 10% of normal. Elimination of the thrombopoietin receptor in children leads to aplastic anemia by 1 to 2 years of age.

Other cytokine–receptor systems related to erythropoietin, granulocyte colony-stimulating factor, and thrombopoietin essential for one or more aspects of hematopoiesis include IL-7, critical for all types of lymphocyte production; IL-5, the primary regulator of eosinophil production; IL-4, responsible for immunoglobulin class switching in B lymphocytes; IL-15, essential for normal natural killer cell differentiation; and IL-2, a lymphocyte activation cytokine. They display modest effects on blood cell growth, but their genetic elimination fails to affect basal or stimulated production of those cells.

A second class of cytokines and receptors that influences hematopoiesis is exemplified by the c-kit receptor, a member of the receptor tyrosine kinase family of surface proteins, and its cognate ligand, stem cell factor (also termed steel factor or kit-ligand). Although the c-kit receptor is structurally distinct from members of the hematopoietic cytokine receptor family, possessing an intrinsic tyrosine kinase motif in its cytoplasmic domain, stem cell factor is structurally related to the cytokines that bind to members of the hematopoietic growth factor family. Genetic deletion of stem cell factor or the c-kit receptor results in the near complete elimination of hematopoietic stem cells, erythroid precursors and basophils, and mast cells. Two other hematopoietic members of this family of cytokines and receptors are Flt3 ligand and its receptor Flt-3 and monocyte colony-stimulating factor and its receptor, c-Fms. Similar to stem cell factor, both Flt3 ligand and monocyte colony-stimulating factor play nonredundant roles in hematopoiesis, inducing the formation of dendritic cells and monocytes, respectively.

The molecular mechanisms by which the hematopoietic growth factors affect blood cell survival, proliferation, and differentiation are becoming increasingly well understood. Binding of cognate ligand to each of the hematopoietic cytokine receptors results in activation of one or more tyrosine kinases, either tethered cytoplasmic kinases of the Janus (JAK) family for the hematopoietic cytokine receptor family or the intrinsic kinase of the cytokines that use the receptor tyrosine kinase class of receptors (stem cell factor, Flt3 ligand, and monocyte colony-stimulating factor). After activation, these kinases phosphorylate tyrosine residues within the cytoplasmic domains of each receptor, providing docking sites for cytoplasmic signaling intermediates possessing Src homology (SH)2 domains. Among the best characterized SH2 domain containing proteins that bind to hematopoietic receptors are nascent transcription factors, such as the signal transducers and activators of transcription (STAT) proteins; adapter proteins, including Grb2, Gab1, tensin2, and SHC; phosphatases, for example, SHP1 and SHP2; and the regulatory subunit (p85) of phosphoinositol-3-kinase (PI3K). When bound to one or more of the newly induced phosphotyrosine residues of the cytokine receptor or receptor tyrosine kinase, these secondary molecules are phosphorylated, either by JAK or other kinases, making them competent to bind additional molecules (e.g., the adapters that ultimately activate Ras, and p85 PI3K that binds its kinase [p110] subunit) or are activated as transcription factors (e.g., STATs). The downstream effector molecules then activated include a number kinases, transporter molecules, and transcription factors, ultimately leading to hematopoietic cell survival, proliferation, and differentiation.

CLINICAL USES OF HEMATOPOIETIC CELLS AND GROWTH FACTORS

The clinical development of erythropoietin, granulocyte colony-stimulating factor, and thrombopoietin mimetics represent some of the very best

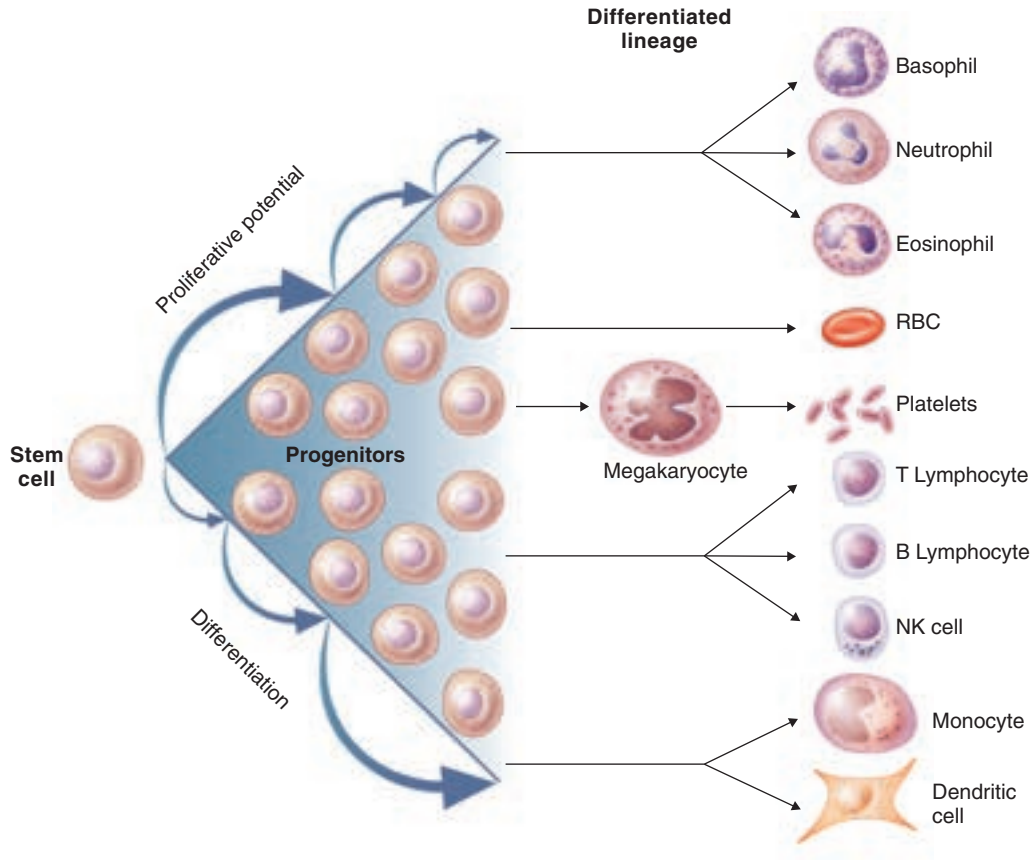

FIGURE 156-1. Hierarchical model of lymphohematopoiesis. NK = natural killer; RBC = red blood cell.

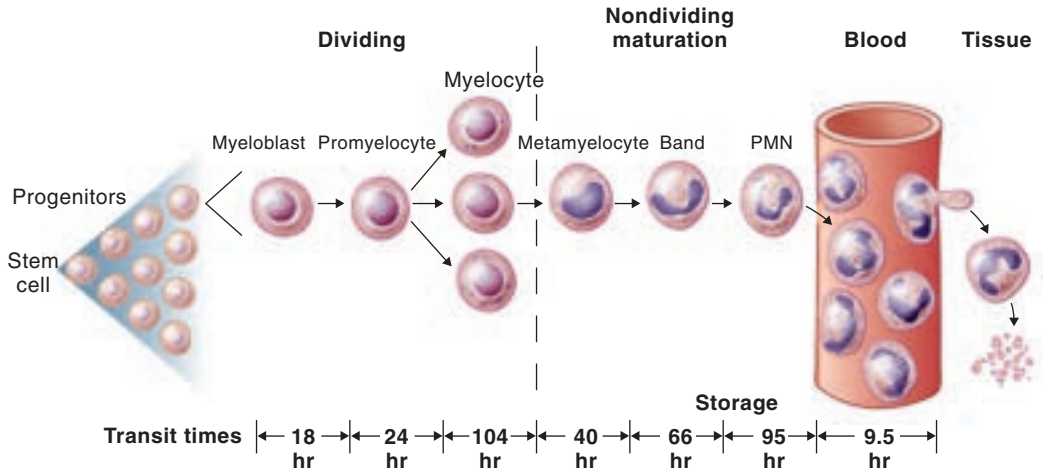

FIGURE 156-2. Neutrophil production system. PMN = polymorphonuclear leukocyte.

examples of harnessing recombinant DNA technology for therapeutic benefit. Patients with renal failure, widespread inflammation, or bone marrow replacement and those undergoing chemotherapy for cancer all experience variable degrees of anemia that are often very debilitating. Administration of erythropoietic stimulating agents almost invariably results in a rapid reticulocyte response and correction of the anemia. Most patients undergo an enhanced sense of well being as the blood hemoglobin concentration rises to 10 g/dL. Clinical trials have demonstrated the efficacy of these agents in patients with renal failure and with cancer, although recent analyses call into question the safety of these agents in some settings.[■■] For example, patients receiving higher levels of the drug for anemia secondary to kidney failure progressed to requiring dialysis more frequently and experienced increased cardiovascular events, such as myocardial infarction and stroke, than patients on low levels of the hormone sufficient to maintain their blood Hgb at 10 g/dL or lower.[■] And patients receiving erythropoietin for cancer also experienced increased relapses of their tumors than individuals not receiving the hormone. The only patients who regularly demonstrate a poor response are individuals with severe inflammation (Chapter 158). Overall, erythropoietic stimulating agents are safe and effective drugs for patients with anemia caused by a wide range of conditions, but their use and dose must be carefully considered.

Many patients undergoing cytotoxic therapy for cancer experience severe neutropenia and are thus at substantial risk for life-threatening infection. Clinical trials of recombinant granulocyte colony-stimulating factor in patients undergoing aggressive chemotherapy for leukemia and solid tumors resulted in the Food and Drug Administration (FDA) approval of the drug for use in patients undergoing chemotherapy of intensity sufficient to produce severe neutropenia (Chapter 167). The use of the drug is associated with the more rapid return of neutrophils to safe levels if administered soon after the inciting chemotherapy is completed but not at the nadir of neutrophil production and results in lower risk of severe infections. However, the use of granulocyte colony-stimulating factor has not enhanced survival in patients

with any tumor type. And similar to the use of erythropoietin in patients with cancer, the administration of granulocyte colony-stimulating factor to some patients receiving cytotoxic chemotherapy for cancer, (e.g., breast cancer) has been associated with a statistically significant increase in cancer recurrence, although this finding is controversial.

Thrombopoietin was cloned and characterized in 1994 and was quickly advanced to clinical trials following the model of granulocyte colony-stimulating factor use in patients undergoing cancer chemotherapy. Initial results with the intact hormone and a truncated version that included only the receptor-binding domain were mixed, and use of the truncated form of the drug, administration to healthy volunteer donors to improve platelet apheresis yields, resulted in a significant number of subjects developing anti-drug antibodies that cross-reacted with their native thrombopoietin, resulting in severe thrombocytopenia. This experience caused both manufacturers of thrombopoietin to cease clinical trials. Instead, a series of small molecule mimics have been developed that bind to the thrombopoietin receptor and stimulate thrombopoiesis. Two such drugs are approved by the FDA for use in patients with severe immune thrombocytopenia—one a peptibody that contains four copies of a thrombopoietin receptor binding peptide grafted onto an immunoglobulin scaffold and the other a small organic molecule that is orally bioavailable and binds to a spatially distinct site on the thrombopoietin receptor. The use of each drug results in a high rate of platelet responses into the normal range in patients with severe immune thrombocytopenia (Chapter 172) who were refractory to conventional therapies. However, because thrombopoietin mimics do not affect the underlying immune destruction of platelets and MKs in patients with this disease, discontinuation of each drug results in rapid relapse of thrombocytopenia. Current clinical trials in other settings of transient myelosuppression have led to improved platelet recovery, but such indications are not yet approved by the FDA.

Grade A References

A1. Palmer SC, Saglimbene V, Craig JC, et al. Darbepoetin for the anaemia of chronic kidney disease. *Cochrane Database Syst Rev*. 2014;3:CD009297.

A2. Tonia T, Mettler A, Robert N, et al. Erythropoietin or darbepoetin for patients with cancer. *Cochrane Database Syst Rev*. 2012;12:CD003407.

A3. Pfeffer MA, Burdmann EA, Chen CY, et al. A trial of darbepoetin alfa in type 2 diabetes and chronic kidney disease. *N Engl J Med*. 2009;361:2019-2032.

A4. Gurion R, Belnik-Plitman Y, Gafter-Gvili A, et al. Colony-stimulating factors for prevention and treatment of infectious complications in patients with acute myelogenous leukemia. *Cochrane Database Syst Rev*. 2012;6:CD008238.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Rector K, Liu Y, Van Zant G. Comprehensive hematopoietic stem cell isolation methods. *Methods Mol Biol*. 2013;976:1-15.
2. Anthony BA, Link DC. Regulation of hematopoietic stem cells by bone marrow stromal cells. *Trends Immunol*. 2014;35:32-37.
3. Deutsch VR, Tomer A. Advances in megakaryocytopoiesis and thrombopoiesis: from bench to bedside. *Br J Haematol*. 2013;161:778-793.

157

THE PERIPHERAL BLOOD SMEAR

BARBARA J. BAIN

With the development of sophisticated automated instruments to count and characterize blood cells, Romanowsky (Wright-Giemsa or May-Grünwald-Giemsa)–stained peripheral blood smears are now performed on only a minority of blood specimens received in a hematology laboratory. Nevertheless, the blood smear remains important for a number of reasons: it can (1) verify the result of an automated instrument, (2) provide an immediate specific diagnosis, or (3) indicate a narrow range of diagnostic possibilities, permitting a focused rather than indiscriminate investigation.[1-3] A blood film can provide a rapid diagnosis in cases in which speed is crucial, such as in acute promyelocytic leukemia, thrombotic thrombocytopenic purpura, Burkitt lymphoma, and certain infections.[4,5] Sometimes a smear provides unexpected information that is of value in patient management.

Usually, blood smears are initially interpreted by a laboratory scientist. In some countries, it is customary for clinicians to examine the blood smears of their own patients because the clinician has the final responsibility for integrating all information and making a diagnosis. However, the interpretation of a blood smear can be difficult, and a trained laboratory hematologist or hematopathologist has a major role in interpreting smears that may have been initially examined by a laboratory scientist. It is crucial that, when requesting a blood count, the clinician provides all the essential information needed to interpret the count and any associated smear. Regardless of whether the clinician examines the blood film, he or she must be able to interpret the written report issued by the laboratory. To do so, the clinician must be familiar with the terms generally used by laboratory staff and the possible significance of the abnormalities described. The most important of these terms are illustrated in Figures 157-1 through 157-20.

REASONS FOR PERFORMING A BLOOD SMEAR

A blood smear may be requested by a clinician or initiated by a laboratory scientist or a laboratory hematologist. Clinical findings that should lead a clinician to request a blood smear are summarized in Table 157-1.

Laboratory scientists and physicians may initiate a blood smear that has not been requested by the clinician if the clinical details indicate the possibility of a significant hematologic abnormality. However, they are also likely to evaluate a blood smear in response to abnormalities revealed by an automated instrument. These abnormalities may be quantitative or qualitative. Quantitative abnormalities that require evaluation include anemia, polycythemia, macrocytosis, microcytosis, neutrophilia, lymphocytosis, eosinophilia, thrombocytopenia, and thrombocytosis.

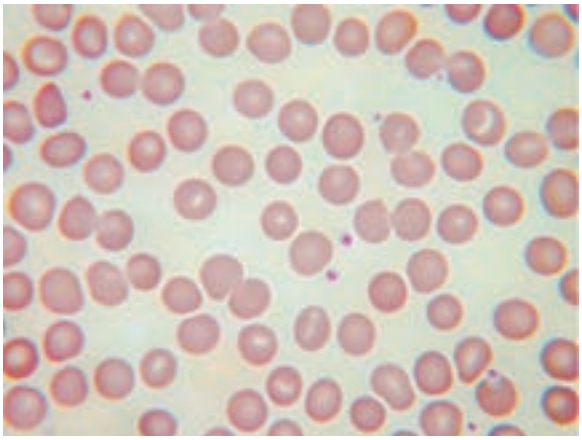

FIGURE 157-1. Normal peripheral blood smear. These normal red cells are described as *normocytic* (i.e., of normal size) and *normochromic* (i.e., their staining characteristics are normal). Normal erythrocytes are biconcave discs, causing them to have an area of central pallor that does not exceed one third the diameter of the cell. There are also scattered normal platelets (×1000).

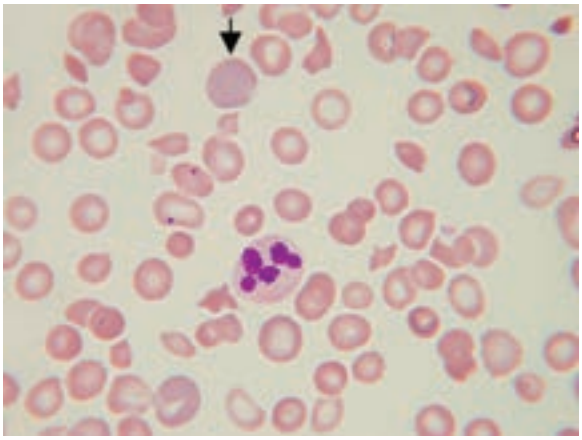

FIGURE 157-2. Smear showing multiple abnormalities. There is *anisocytosis,* defined as an increased variation in cell size; *poikilocytosis,* defined as an increased variation in cell shape; and *polychromasia,* defined as the presence of erythrocytes with a blue tinge to their cytoplasm—indicating a young cell recently released from the bone marrow in which the polychromasia is due to the presence of RNA. *Polychromatic cells* are erythrocytes with a blue tinge; because polychromatic cells are larger than normal mature erythrocytes, they are known as *polychromatic macrocytes (arrow)*. The film also shows two cells containing bluish purple *Howell-Jolly bodies*; these inclusions are nuclear remnants (×1000).

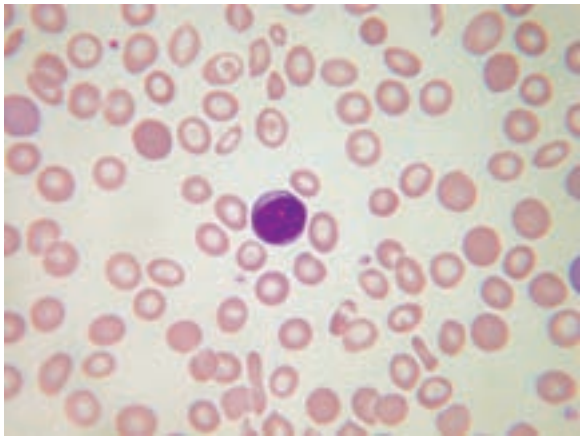

FIGURE 157-3. Microcytic red cells from a case of thalassemia minor. In a blood smear, a *microcyte* can be defined as a cell with a diameter less than that of the nucleus of a normal small lymphocyte. There are also some cells showing *hypochromia*, an area of central pallor that is larger than one third of the diameter of the red cell. In addition, there are two *target cells*, with a hemoglobinized area in the center of the area of pallor (×1000).

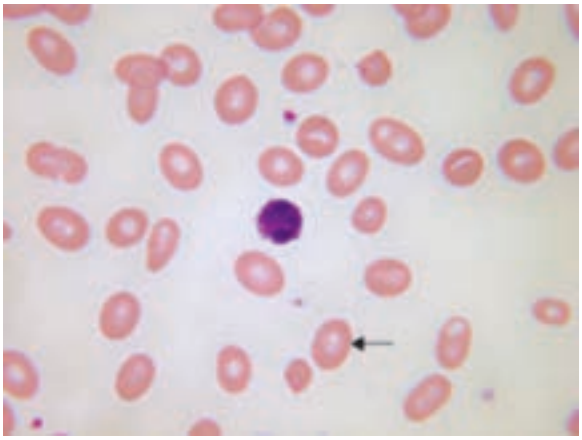

FIGURE 157-4. Macrocytic anemia. A *macrocyte* is recognized on a blood film as a cell with a diameter that is considerably greater than that of the nucleus of a small lymphocyte. In addition, this smear shows *oval macrocytes* (also known as *macro-ovalocytes*), defined as cells that are larger than normal and oval in shape *(arrow)*. They are of considerable diagnostic importance, being characteristic of megaloblastic anemia; they can also be seen in dyserythropoiesis (×1000).

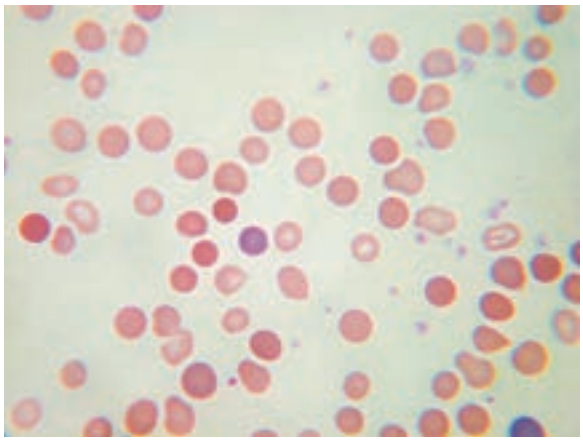

FIGURE 157-5. Hereditary spherocytosis. A *spherocyte* is a red cell that lacks central pallor because of its spherical shape. In hereditary spherocytosis, there are usually cells in which the central pallor is reduced rather than absent, and they are intermediate in shape between a spherocyte and a *discocyte,* which is an erythrocyte with the normal shape of a biconcave disc (×1000).

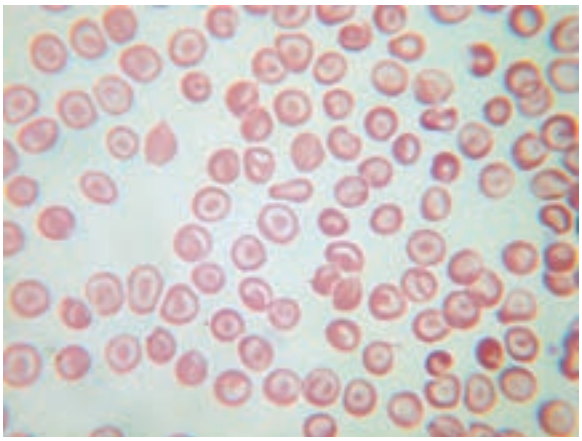

FIGURE 157-6. Target cells. A *target cell* is an erythrocyte with a hemoglobinized area in the middle of the normal area of central pallor (×1000).

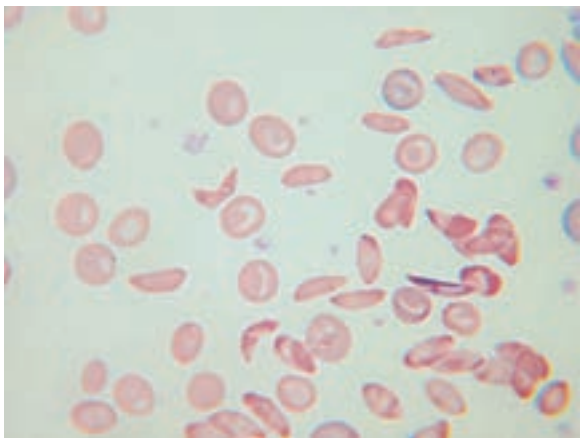

FIGURE 157-7. Sickle cells. A *sickle cell* is a cell with a sickle or crescent shape resulting from the polymerization of hemoglobin S. These cells are seen not only in sickle cell anemia (homozygosity for hemoglobin S) but also in compound heterozygous states such as sickle cell/hemoglobin C disease and sickle cell/β-thalassemia, which also lead to sickle cell disease. This smear also shows target cells and boat-shaped cells with a lesser degree of polymerization of hemoglobin S than in a classic sickle cell (×1000).

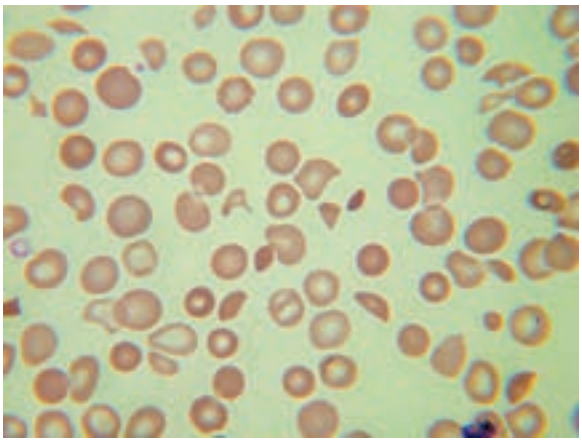

FIGURE 157-8. Red cell fragmentation. *Red cell fragments* or *schistocytes* are defined as fragments of erythrocytes. In addition to small angular fragments, there may be *microspherocytes,* cells of reduced size and spherical in form (also known as *spheroschistocytes*), and *keratocytes,* cells with two horn-like projections. Keratocytes can result from removal of a Heinz body, as well as from red cell fragmentation. Some schistocytes are referred to as *helmet cells* because of their typical shape. Schistocytes are seen in microangiopathic hemolytic anemias and in mechanical hemolysis (×1000).

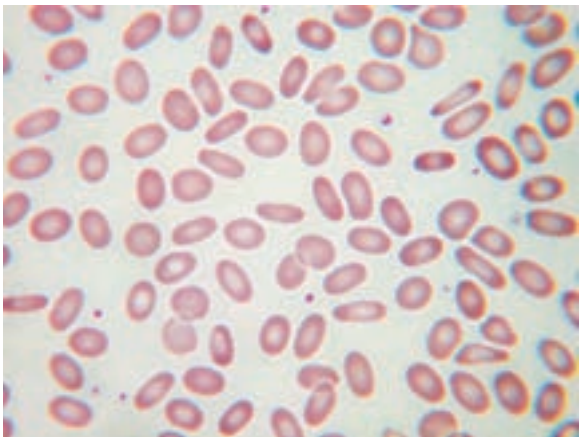

FIGURE 157-9. Hereditary elliptocytosis. An *elliptocyte* is an elliptical red cell. When seen in the numbers present in this smear, they are indicative of hereditary elliptocytosis; smaller numbers are seen in other conditions such as iron deficiency anemia, in which they are sometimes referred to as *pencil cells* (×1000).

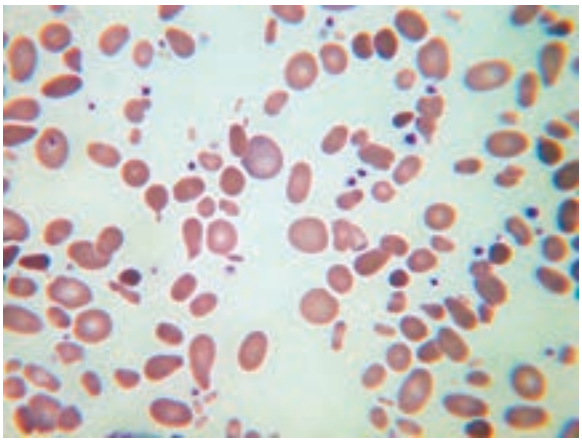

FIGURE 157-10. Hereditary pyropoikilocytosis. This smear shows striking poikilocytosis, including elliptocytes, microspherocytes and other fragments, and teardrop cells. This congenital condition, which is related to hereditary elliptocytosis, usually results from the inheritance of two different mutated genes from the two parents and is characterized by a severe hemolytic anemia (×1000).

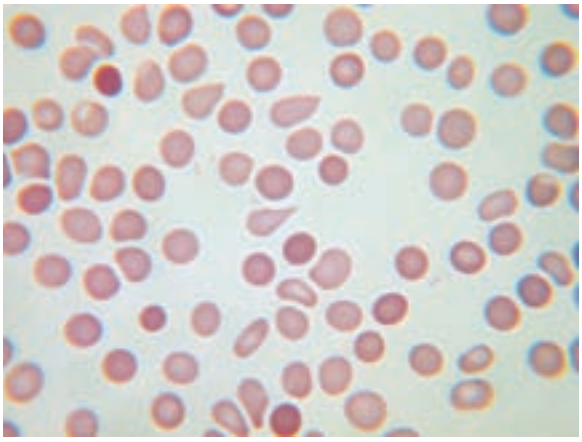

FIGURE 157-11. Teardrop poikilocytes. *Teardrop poikilocytes,* or *dacrocytes,* are teardrop-shaped red cells; they are characteristic of primary myelofibrosis but are also seen in megaloblastic anemia (×1000).

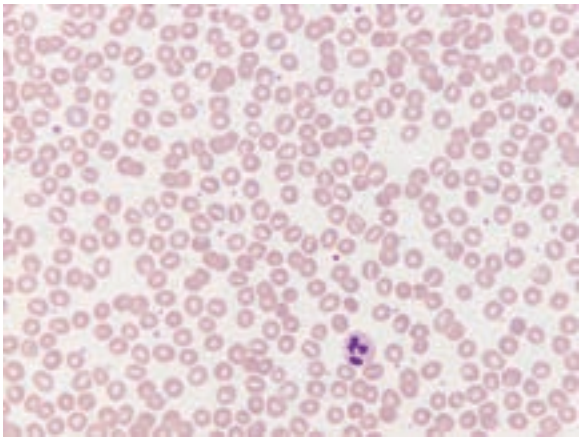

FIGURE 157-12. Stomatocytosis. A *stomatocyte* is a cell that appears to have a central mouth-shaped or slit-like stoma. Among the less common causes is hereditary stomatocytosis. Alcohol and hydroxycarbamide therapy are more common causes (×400).

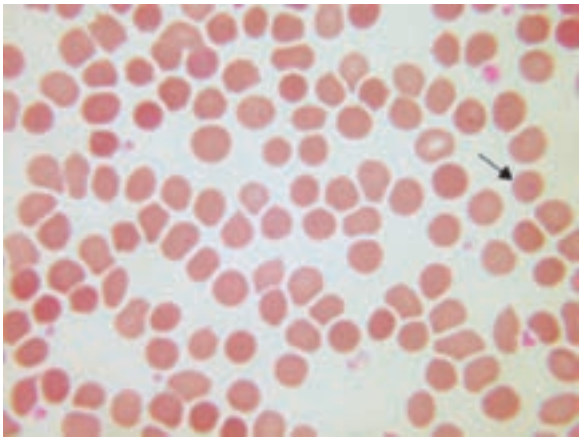

FIGURE 157-13. Numerous Pappenheimer bodies. A *Pappenheimer body* is an iron-containing red cell inclusion *(arrow)*. It is smaller and more angular than a Howell-Jolly body and stains navy blue rather than purple. Pappenheimer bodies are seen following splenectomy and in sideroblastic anemias (×1000).

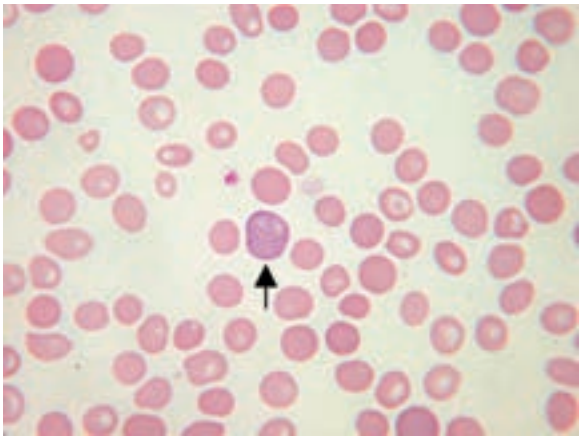

FIGURE 157-14. Basophilic stippling. *Basophilic stippling* means that there are fine (as in this case) or coarse purplish blue dots dispersed through the red cell *(arrow)*. They are a very nonspecific feature occurring in thalassemia trait, lead poisoning, pyrimidine 5′-nucleotidase deficiency, and dyserythropoiesis in general (×1000).

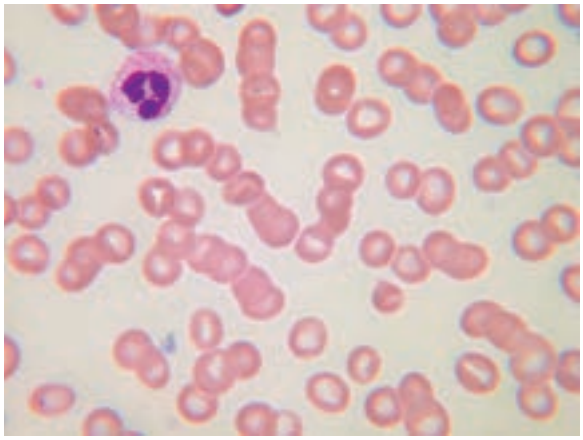

FIGURE 157-15. Rouleaux formation. *Rouleaux* are stacks of red cells, often compared to stacks of coins. They result from an increase of high-molecular-weight globulins in the plasma, either as a reactive change or as a result of secretion of a paraprotein in a plasma cell neoplasm (×1000).

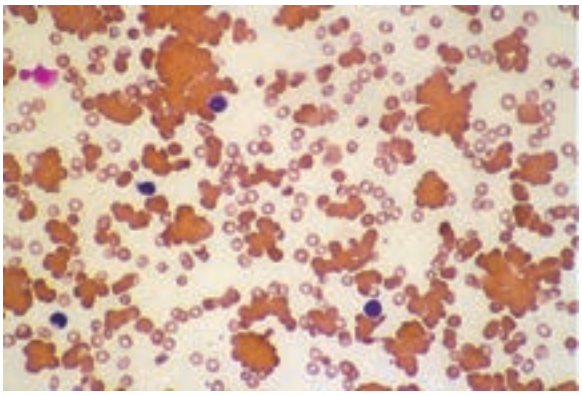

FIGURE 157-16. Red cell agglutination. *Red cell agglutinates* are irregular aggregates of red cells, as seen in *Mycoplasma pneumoniae* infection. They are also seen in other infections, such as infectious mononucleosis, and in chronic cold hemagglutinin disease (×100). (Courtesy of Jean Schafer.)

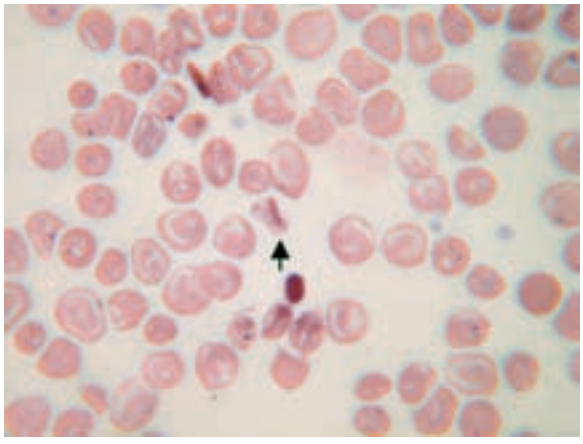

FIGURE 157-17. Homozygous hemoglobin C. Three *hemoglobin C crystals* are present, one indicated by an *arrow.* Hemoglobin C crystals are usually six-sided, with a long axis having parallel edges (×1000).

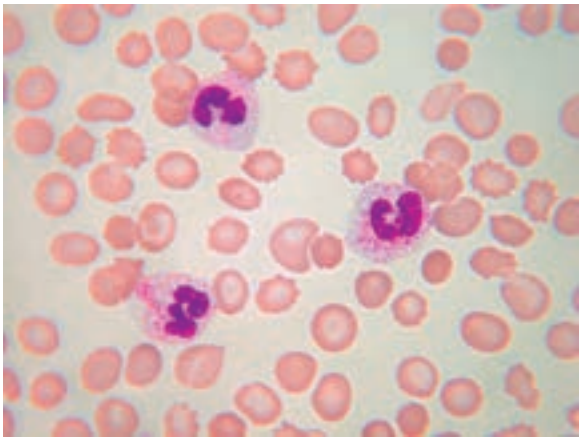

FIGURE 157-18. Toxic granulation. *Toxic granulation* refers to heavy staining of azurophilic granules of neutrophils. When accompanied by neutrophil vacuolation, it is often indicative of infection, but it can also result from inflammation, tissue damage, and normal pregnancy (×1000).

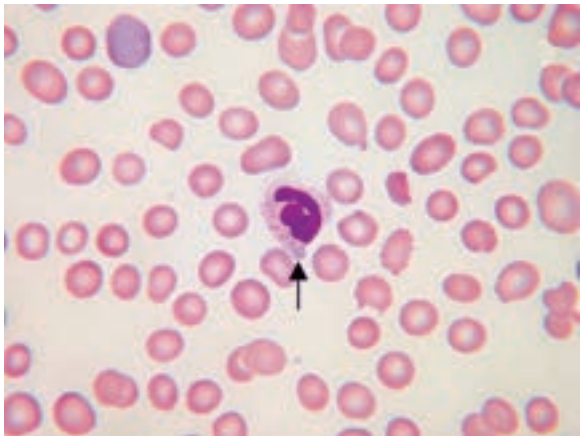

FIGURE 157-19. Döhle body. A *Döhle body (arrow)* is a pale blue-gray amorphous inclusion near the cell membrane of a neutrophil. Döhle bodies can result from infection and inflammation. Similar but different inclusions (larger and more angular) are seen in the May-Hegglin anomaly (×1000).

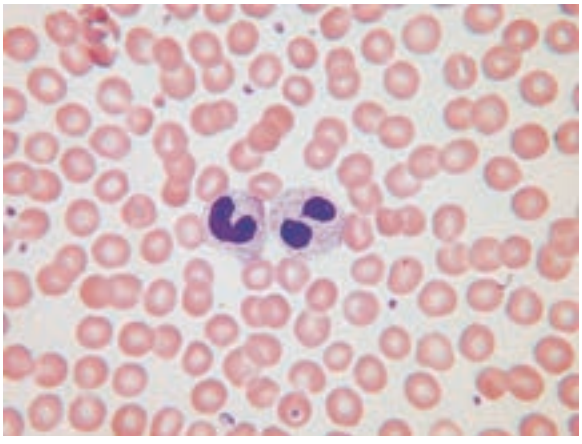

FIGURE 157-20. Pelger-Huët anomaly. A *Pelger-Huët anomaly* is a cytologic abnormality of neutrophils in which there is hypolobulation of nuclei and increased chromatin clumping. Nuclei may have a shape resembling a peanut or a pince-nez, as in the examples shown. The Pelger-Huët anomaly is inherited, but similar Pelger neutrophils are seen in myelodysplastic syndromes, in which the neutrophils may also be hypogranular (×1000). They can also be seen as an acquired reversible phenomenon in response to specific drugs, including tacrolimus and mycophenolate mofetil.

TABLE 157-1 CLINICAL FEATURES SUGGESTING THE NEED FOR A BLOOD SMEAR

CLINICAL FEATURE	REASON TO PERFORM A BLOOD SMEAR
Lymphadenopathy or splenomegaly	May be indicative of infectious mononucleosis or another reactive condition, or of leukemia or lymphoma
Clinically evident anemia	Helps in the differential diagnosis
Bruising or bleeding tendency, including unexplained retinal hemorrhages	May confirm thrombocytopenia or show morphologically abnormal platelets (which may have defective function); sometimes shows acute leukemia or other condition causing bone marrow failure
Acute renal failure	Hemolytic-uremic syndrome and thrombotic thrombocytopenic purpura should be confirmed or excluded
Jaundice and hypertension in a pregnant woman	May show schistocytes, supporting a diagnosis of HELLP syndrome
Bone pain	May indicate multiple myeloma, bone marrow infiltration, or sickle cell disease
Unexplained chest or abdominal pain or acute splenic enlargement in a child	Possible sickle cell disease
Unexplained hyperbilirubinemia	Assessment of possible hemolysis

HELLP = hemolysis, elevated liver enzymes, low platelets.

Modern automated instruments are able to "flag" the presence of qualitative abnormalities that require a blood film to be examined for confirmation of the abnormality or for further elucidation. "Flags" are generated in response to the electrical impedance or the light scattering characteristics of individual cells. Some instruments are dependent on cytochemical reactions of cells or on the cells' ability to polarize light. Most instruments can indicate the possibility of the presence of blast cells, reactive or other atypical lymphocytes, granulocyte precursors, or nucleated red blood cells. Some instruments can enumerate nucleated red blood cells. Instruments using a cytochemical reaction for peroxidase to help identify neutrophils, eosinophils, and monocytes may flag the appearance of large, unstained (i.e., peroxidase-negative) cells; such cells may indicate a harmless inherited peroxidase deficiency, but sometimes such cells are lymphoma cells, reactive lymphocytes, or leukemic blast cells. Instruments often flag the possibility of an erroneous platelet count, such as when there is an overlap in size between platelets and red cells or when light-scattering characteristics suggest the presence of platelet aggregates (a potential cause of pseudothrombocytopenia). A reported increase in basophil count should generally also be regarded as a flag because it often represents a pseudobasophilia, resulting from the presence of leukemia or lymphoma cells. Some instruments alert the instrument operator to the possible presence of malaria parasites.

International consensus guidelines indicate which automated instrument results require blood smear review. Whether a review is needed is determined in part by whether that specimen is the first one obtained from that patient and whether there has been a significant change from a previously validated result (referred to as a *delta check*). Laboratory computers can be programmed to indicate when a result meets the criteria for smear review.

Automated red cell measurements can, to some extent, replace examination of the blood film.[6] An increased red cell distribution width indicates the presence of anisocytosis. A decreased mean cell volume is usually a reliable indicator of microcytosis. A reduction in the mean cell hemoglobin concentration indicates hypochromia (for most instruments, however, this measurement is less sensitive than the human eye in the detection of hypochromia). An increased mean cell volume usually indicates macrocytosis, but examination of a smear is necessary both to confirm that the result is not artifactual and to elucidate the cause. Some instruments can measure the hemoglobin concentration in individual cells and thus flag the presence of hyperdense cells; however, a blood film is still necessary to distinguish spherocytes from irregularly contracted cells, sickle cells, and other cells that have an increased hemoglobin concentration. Most instruments produce a histogram of the size distribution of red cells, and some do the same for the distribution of hemoglobin concentration; either of these graphic representations may show dimorphic red cells (i.e., two populations of cells).

TABLE 157-2 USEFUL FEATURES FOR DETERMINING THE CAUSE OF MACROCYTIC ANEMIA

CAUSE	SMEAR FEATURES
Megaloblastic anemia (vitamin B_{12} or folic acid deficiency)	Oval macrocytes, teardrop poikilocytes, hypersegmented neutrophils; when severe, marked anisocytosis and poikilocytosis, which may include red cell fragments
Ethanol excess	Target cells and stomatocytes; anisocytosis and poikilocytosis less than in megaloblastic anemia
Liver disease	Target cells, stomatocytes
Myelodysplastic syndromes, including sideroblastic anemias	Other dysplastic features such as hypogranular and hypolobulated neutrophils; if erythropoiesis is sideroblastic, a population of hypochromic microcytic cells and Pappenheimer bodies
Chronic hemolytic anemia	Polychromasia; characteristic poikilocytes sometimes present (e.g., irregularly contracted cells if there is an unstable hemoglobin)

THE BLOOD SMEAR IN THE DIFFERENTIAL DIAGNOSIS OF ANEMIA

Microcytic Anemias

In microcytic anemias (Chapter 159), the automated count is of considerable importance and may permit a distinction between iron deficiency and thalassemia heterozygosity. In iron deficiency, there is initially a normocytic normochromic anemia; only when the deficiency becomes more severe is there microcytosis. Conversely, in β-thalassemia heterozygosity, there is usually a normal or near-normal hemoglobin concentration, but the red blood cell count is increased, and there is marked microcytosis (low mean cell volume) together with a parallel reduction in mean cell hemoglobin. The blood film provides supplementary information that can favor one diagnosis or the other. Iron deficiency is more likely to be associated with hypochromia and elliptocytes ("pencil cells"), whereas in β-thalassemia heterozygosity, there is microcytosis, hypochromia is less marked, and there are more likely to be target cells and basophilic stippling. Individuals with α-thalassemia involving the deletion of two α genes (–α/–α) or (– –/α α) have red cell indices similar to those of β-thalassemia heterozygosity; in this case, the blood film usually does not provide any additional diagnostically useful information, although individuals with nondeletional α-thalassemia due to hemoglobin Constant Spring have prominent basophilic stippling. When there is deletion of a single α gene, the blood count is either less abnormal or normal, and the blood smear does not provide any extra diagnostically useful information. The blood smear is, however, a useful supplement to the blood count in suggesting a diagnosis of hemoglobin H disease (Chapter 162). The count shows anemia, marked microcytosis (low mean cell volume and mean cell hemoglobin), and usually a reduction in the mean cell hemoglobin concentration. The smear usually shows marked poikilocytosis in addition to microcytosis, and there may be polychromasia, correlating with an elevated reticulocyte count. Iron deficiency anemia also needs to be distinguished from anemia of chronic disease. The blood counts may be quite similar, but in anemia of chronic disease, the smear often shows features of inflammation, such as increased rouleaux formation, background staining (as a result of increased plasma proteins), and sometimes neutrophilia. Other rare microcytic anemias that must be distinguished from iron deficiency include congenital sideroblastic anemia (Chapter 159) and lead poisoning. In congenital sideroblastic anemia, the film is dimorphic, with one population of hypochromic microcytes and another of normochromic normocytic cells. In lead poisoning, the presence of basophilic stippling and polychromasia in a patient with microcytosis can suggest the diagnosis.

Macrocytic Anemias

A blood film can be important in distinguishing true macrocytosis from factitious macrocytosis as a result of the presence of red cell agglutinates (see Fig. 157-16). Diagnostic features that can suggest the cause of macrocytosis are shown in Table 157-2.

A smear is particularly important in evaluating the possibility of a megaloblastic anemia (Chapter 164) (Figs. 157-4 and 157-21). Sometimes assays of

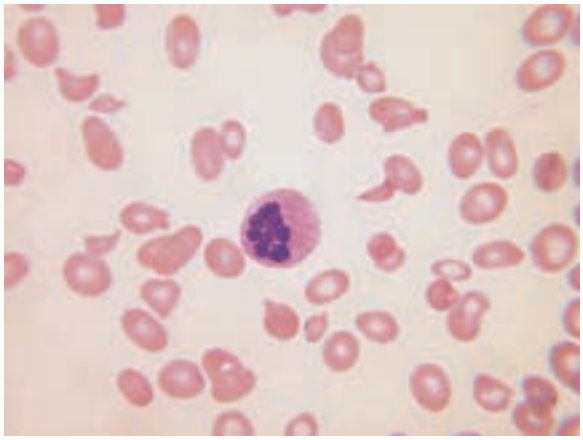

FIGURE 157-21. Hypersegmented neutrophils. A *hypersegmented neutrophil* is a neutrophil with more than five nuclear segments or lobes, as in this example from a patient with megaloblastic anemia. Neutrophil hypersegmentation is also said to be present if there are increased numbers of neutrophils with five lobes or if the median lobe count is increased (×1000).

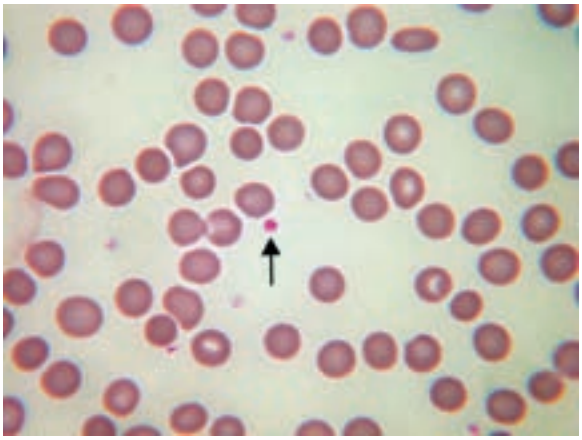

FIGURE 157-22. Normal-sized platelet *(arrow).* Platelets have central granules, although this is not apparent in this photomicrograph (×1000).

TABLE 157-3 BLOOD SMEAR FEATURES SUGGESTING A SPECIFIC CAUSE OF INHERITED OR ACQUIRED HEMOLYTIC ANEMIA

BLOOD SMEAR FEATURES	CONDITIONS SUGGESTED
Spherocytes	Hereditary spherocytosis, autoimmune hemolytic anemia, alloimmune hemolytic anemia (e.g., hemolytic disease of the newborn, delayed hemolytic transfusion reaction), drug-induced immune hemolytic anemia, *Clostridium perfringens* sepsis
Elliptocytes	Hereditary elliptocytosis
Oval macrocytes plus stomatocytes	South-East Asian ovalocytosis
Irregularly contracted cells	Glucose-6-phosphate dehydrogenase deficiency, oxidant damage from chemicals or drugs in individuals with normal red cell enzymes (e.g., dapsone administration), liver failure due to Wilson disease (release of copper from liver), unstable hemoglobin, hemoglobin C homozygosity
Sickle cells and boat-shaped cells	Sickle cell disease (e.g., sickle cell anemia or compound heterozygous states such as S/C, S/D-Punjab, S/O-Arab, S/β-thalassemia)
Target cells	Hemoglobin C homozygosity, other hemoglobinopathies, hereditary xerocytosis
Stomatocytes	Hereditary stomatocytosis
Acanthocytes	Liver failure (spur cell hemolytic anemia)
Basophilic stippling	Lead poisoning, pyrimidine 5′-nucleotidase deficiency
Red cell fragments (schistocytes)	Microangiopathic hemolytic anemia (including hemolytic-uremic syndrome, thrombotic thrombocytopenic purpura, HELLP syndrome, and sometimes hemolysis associated with disseminated intravascular coagulation), mechanical hemolytic anemia (e.g., defective cardiac prosthetic valve, march hemoglobinuria)

HELLP = hemolysis, elevated liver enzymes, low platelets.

vitamin B_{12} and folate are normal despite a deficiency, and only the blood film features suggest the true diagnosis and indicate the need for further investigation.

Normocytic Normochromic Anemia

A blood smear is only occasionally useful in determining the cause of a normocytic normochromic anemia. Signs of inflammation may be present in anemia of chronic disease. Increased rouleaux formation and background staining can also indicate multiple myeloma. Polychromasia suggests the possibility of young red cells as a result of recent blood loss or hemolysis. Small numbers of acanthocytes may indicate hypothyroidism or anorexia nervosa. Dysplastic features, such as hypogranular or pseudo-Pelger neutrophils, suggest a myelodysplastic syndrome.

Hemolytic Anemias

The possibility of hemolysis is suggested by the presence of polychromasia and macrocytosis. Specific causes of hemolysis are suggested by the presence of various poikilocytes, as shown in Table 157-3.

The distinction between spherocytes and irregularly contracted cells is important; both are dense cells with absent central pallor, but the differential diagnosis is quite different. Recognition of the features of oxidant damage is important in diagnosing glucose-6-phosphate dehydrogenase (G6PD) deficiency because sometimes an assay for G6PD performed during an acute hemolytic episode is normal (Chapter 161). In addition to irregularly contracted cells, there may be ghost cells, hemi-ghost cells ("blister cells"), and even Heinz bodies protruding from the red cells and confirmed on a Heinz body preparation. The observation of these features is an indication to repeat the assay when the acute hemolytic episode is over.

ASSESSMENT OF THROMBOCYTOPENIA, THROMBOCYTOSIS, AND PLATELET MORPHOLOGY

A blood smear is essential to validate the cell count whenever an automated count shows thrombocytopenia (e.g., a count $<60 \times 10^9$/L) (Chapter 172). This should be done quickly before patient management is altered, such as by postponing surgery or initiating further diagnostic workup. A platelet transfusion should never be given for an unexpected thrombocytopenia without microscopic confirmation of the count. A factitiously low platelet count is often the result of in vitro platelet aggregation (Chapter 171) and is occasionally the result of platelet satellitism, possibly also with platelet phagocytosis. To detect aggregates reliably, the edges and the tail of the smear should be examined. The presence of fibrin strands also suggests the activation of coagulation and an erroneous platelet count.

If a low platelet count is confirmed, the film may give clues to the cause (Chapter 172). Giant platelets (Figs. 157-22 and 157-23) occur in a number of inherited thrombocytopenias, including Bernard-Soulier syndrome and MYH9-related disorders (the May-Hegglin anomaly and related conditions). Small platelets are less common but are a feature of Wiskott-Aldrich syndrome and familial platelet disorder with propensity to myeloid malignancy. Agranular platelets occur in the gray platelet syndrome[7] and platelets with a reduced number of larger than normal granules are seen in Jacobsen/Paris-Trousseau syndrome. The presence of May-Hegglin inclusions (Döhle-like bodies; see Fig. 157-19) in neutrophils indicates that the cause of the thrombocytopenia is a mutation in *MYH9*. In acquired thrombocytopenias, increased platelet turnover is often accompanied by the presence of large platelets, whereas bone marrow failure is associated with platelets of normal size. It is important to look for red cell fragments to confirm or exclude a diagnosis of thrombotic thrombocytopenic purpura and atypical hemolytic-uremic syndrome in any patient with the apparent recent onset of

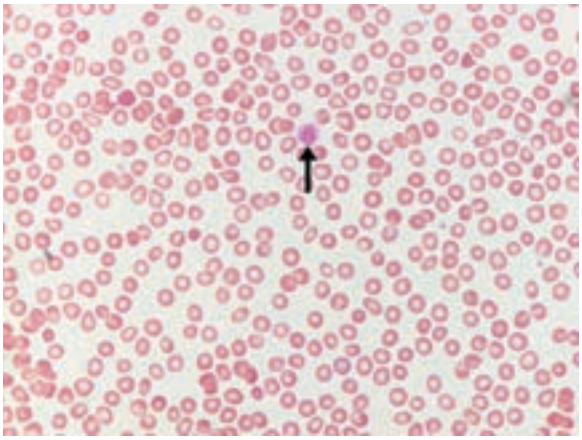

FIGURE 157-23. Giant platelet *(arrow).* Giant platelets are as large as or larger than normal red cells. Giant platelets can indicate increased platelet turnover or an inherited or acquired defect in thrombopoiesis (×1000).

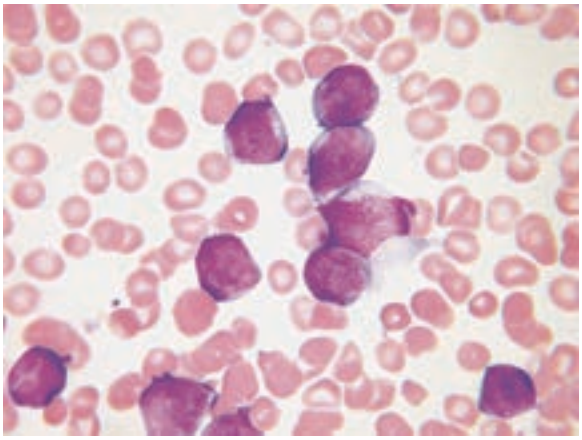

FIGURE 157-24. Acute lymphoblastic leukemia. Numerous agranular blast cells with a high nuclear-to-cytoplasmic ratio are present. Platelets are decreased in number (×1000).

thrombocytopenia; because platelet transfusions are usually contraindicated in these conditions, the smear should be examined before platelet transfusion is contemplated. The smear of any patient with the apparent recent onset of severe thrombocytopenia should be examined carefully for evidence of acute promyelocytic leukemia; the leukemic cells may be infrequent in the circulating blood. Hemorrhagic manifestations and a low platelet count can also be indicative of meningococcal septicemia; in some patients, organisms are seen in the blood smear and the diagnosis is confirmed; in other patients, only marked toxic changes in neutrophils are detected.

Thrombocytosis should also be confirmed on a smear. Factitiously elevated counts may be the result of the presence of red cell fragments (in microangiopathic or mechanical hemolytic anemia, burns, or accidental in vitro heating of the blood sample), white cell fragments (in acute leukemia and, less often, in lymphoma), cryoglobulin precipitates, or microorganisms (particularly *Candida* species). If the count is confirmed, the blood smear may be useful to indicate a likely cause (e.g., features of hyposplenism or the presence of basophilia in a myeloproliferative disorder).

It is sometimes necessary to examine a smear to confirm that an apparently normal platelet count is valid. This should always be done in patients with acute leukemia and an elevated white cell count; the presence of white cell fragments of a similar size to platelets can suggest that the platelet count is at a safe level when it is in fact dangerously low. Counting the ratio of platelets to other particles of similar size permits the count to be corrected. Any unexpectedly normal count should be confirmed; for example, the sudden rise of the automated platelet count in a patient being treated for a hematologic neoplasm may be the result of fungi that have colonized an indwelling intravenous line and are being shed into the blood stream.

LEUKOCYTOSIS AND LEUKOPENIA

The finding of leukocytosis is not necessarily an indication for a blood smear (Chapter 170). For example, this finding would be expected in a patient with infection or following surgery or trauma, in which case smear confirmation is not required. However, unexpected leukocytosis requires a smear. Artifactual elevation is unusual but can occur as a result of cryoglobulinemia, hyperlipidemia, or the presence of *Candida* species. Distinguishing reactive changes from leukemia on the basis of morphology is usually straightforward for myelocytosis but more difficult for lymphocytosis.[8] In reactive leukocytosis, there is usually toxic granulation, and Döhle bodies may be present (see Figs. 157-18 and 157-19). Vacuolation is particularly characteristic of bacterial infection, and there may be some degranulation of neutrophils. The hematologist should be aware of the changes induced by granulocyte colony-stimulating factor so that they are not confused with a response to infection; this cytokine can cause toxic granulation, Döhle bodies, vacuolation, and the presence of macropolycytes (giant neutrophils) and circulating neutrophil precursors. The changes typical of leukemia are discussed later.

Leukopenia (Chapter 167) usually requires a film for confirmation and elucidation. The exception is when it is expected in a given clinical context, such as when a patient has had recent chemotherapy. Rarely, an apparent leukopenia is artifactual, owing to the aggregation of neutrophils mediated by an autoantibody or resulting from infection-induced changes in adhesion molecules of the leukocyte surface membrane.

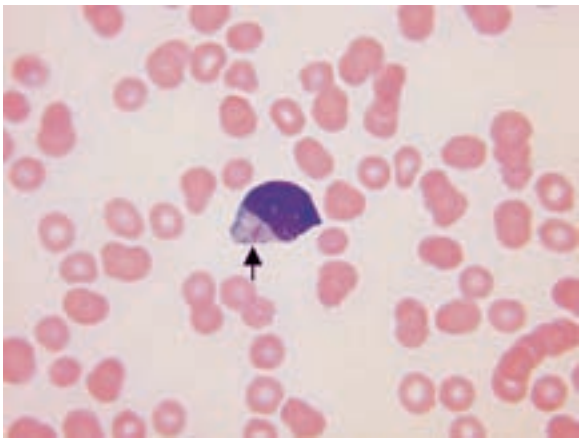

FIGURE 157-25. Auer rod. An Auer rod *(arrow)* is a rod-shaped inclusion in the cytoplasm of cells of myeloid lineage formed by the crystallization of azurophilic granule constituents. Auer rods are seen only in acute myeloid leukemia and high-grade myelodysplastic syndromes. They are usually seen in blast cells but are occasionally found in maturing cells (×1000).

With some automated instruments, it is necessary to confirm that neutropenia is real. If the automated count is based on peroxidase cytochemistry, the presence of an inherited deficiency will lead to an apparent neutropenia associated with an increase of large, unstained (i.e., peroxidase-negative) cells. The scatter plot is characteristic, but because the same features could be due to acute leukemia with neutropenia and circulating blast cells, a smear is needed for confirmation.

LEUKEMIAS AND LYMPHOMAS

The blood smear is critical in the diagnosis of leukemias and lymphomas. The lymphoblasts of acute lymphoblastic leukemia are usually medium-sized agranular cells with relatively scanty cytoplasm (Fig. 157-24), whereas in acute myeloid leukemia, blast cells are generally larger, with more plentiful cytoplasm that may contain granules or Auer rods (Fig. 157-25). Myeloid blast cells vary in appearance according to whether they are myeloblasts or monoblasts. Myeloblasts are usually medium sized and may have plentiful granules, scanty granules, or no visible granules; they may contain Auer rods. Monoblasts are much larger cells with plentiful cytoplasm containing few granules and, very rarely, Auer rods. Megakaryoblasts are present in some patients and are sometimes cytologically distinctive because of their tendency to form cytoplasmic blebs or develop platelet-type granules.

Chronic myelogenous leukemia has a very characteristic blood smear (Fig. 157-26) in which the most numerous cells are myelocytes and mature neutrophils. Eosinophils and basophils are also present. Dysplastic features are generally absent. In atypical Philadelphia chromosome–negative chronic myeloid leukemia, monocytosis is more frequent and dysplastic features are present. Chronic myelomonocytic leukemia is characterized by increased

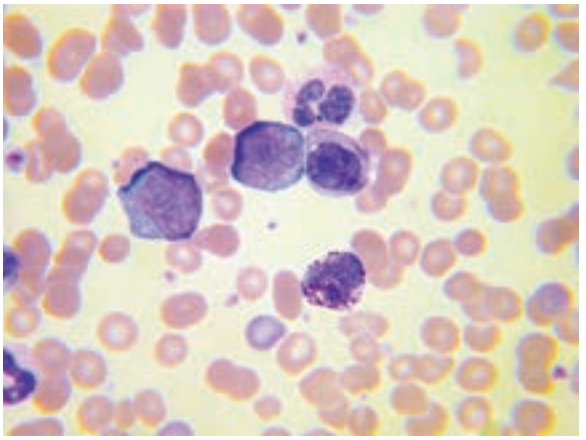

FIGURE 157-26. Chronic myelogenous leukemia. In this view there are myeloblasts, a myelocyte, a basophil, and a segmented neutrophil (×1000).

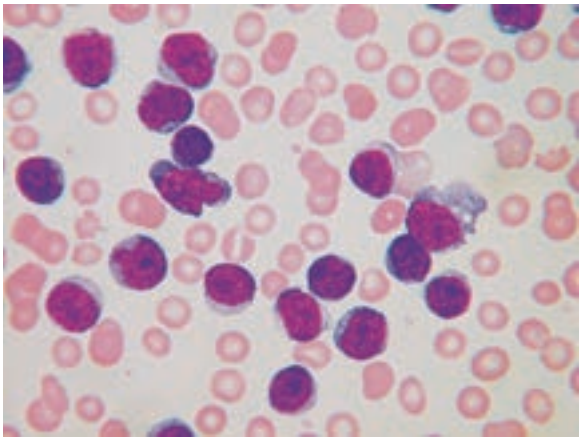

FIGURE 157-27. Chronic lymphocytic leukemia. There are large numbers of rather monotonous, mature small lymphocytes with chromatin clumping. Smear cells, reflecting the mechanical fragility of the cells, are characteristic but are not seen in this photomicrograph (×1000).

monocytes, some of them immature, with inconspicuous dysplastic features and infrequent granulocyte precursors.

Chronic lymphocytic leukemia also has a very characteristic blood film, with an increase of small, mature lymphocytes of rather uniform appearance. The chromatin is often irregularly clumped, creating a mosaic effect (Fig. 157-27). Smear cells are almost always increased in number but are not pathognomonic.

Lymphoma in leukemic phase often has cytologic features that aid in the diagnosis. Follicular lymphoma, Burkitt lymphoma, and splenic marginal zone lymphoma can all be distinctive.

THE INCIDENTAL DETECTION OF CLINICALLY SIGNIFICANT ABNORMALITIES

Sometimes the examination of a blood film reveals unexpected but clinically significant information. Examples are given in Table 157-4. Detection of microorganisms is particularly likely in patients with acquired immunodeficiency syndrome (AIDS), hyposplenic patients,[9] and patients with overwhelming sepsis, but sometimes they are detected in immunologically normal patients with only trivial symptoms. It should be noted that microorganisms in blood smears sometimes represent contaminants, particularly if specimens have been obtained by skin prick or from the umbilical cord and if there has been a delay in making the film.

CONCLUSION

Despite major advances in other diagnostic methods, the blood smear remains very useful in hematologic diagnosis. Sometimes it is critical either because it yields a diagnosis very quickly or because it provides information that is not available in any other way.

TABLE 157-4 INCIDENTAL BUT CLINICALLY RELEVANT BLOOD SMEAR OBSERVATIONS

OBSERVATION	POSSIBLE SIGNIFICANCE
Acanthocytes	Abetalipoproteinemia, neuroacanthocytosis (includes choreoacanthocytosis, McLeod phenotype, Huntington-like disease 2 and pantothenate-kinase associated neurodegeneration)
Howell-Jolly bodies, target cells, and acanthocytes	Hyposplenism (congenital, previous splenectomy, celiac disease, amyloidosis)
Cryoglobulin	Hepatitis C, multiple myeloma, Waldenström macroglobulinemia
Vacuolated lymphocytes	Inherited metabolic disorders
Parasites (e.g., malaria parasites, *Babesia*, microfilaria, trypanosomes, *Leishmania*)	Parasitic infection
Fungi (*Candida* spp, *Histoplasma capsulatum*, *Penicillium marneffei*, *Cryptococcus neoformans*, *Malassezia furfur*)	Disseminated fungal infection or, in the case of *Candida*, colonization of an indwelling intravenous line
Bacteria (e.g., pneumococcus, meningococcus, *Capnocytophaga canimorsus*, *Borrelia*, *Ehrlichia*, *Anaplasma*, *Yersinia pestis*)	Bacterial infection
Leukoerythroblastic blood film	Bone marrow infiltration (e.g., metastatic malignancy)

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bain BJ. Diagnosis from the blood smear. *N Engl J Med*. 2005;353:498-507.
2. Barnes PW, McFadden S, Machin SJ, et al. The International Consensus Group for Hematology Review. http://www.islh.org/web/index.php/2009/index.php?page=consensus_preface. Accessed May 28, 2014.
3. Froom P, Havis R, Barak M. The rate of manual peripheral blood smear reviews in outpatients. *Clin Chem Lab Med*. 2009;47:1401-1405.
4. Prokocimer M, Potasman I. The added value of peripheral blood cell morphology in the diagnosis and management of infectious diseases—part 1: basic concepts. *Postgrad Med J*. 2008;84:579-585.
5. Prokocimer M, Potasman I. The added value of peripheral blood cell morphology in the diagnosis and management of infectious diseases—part 2: illustrative cases. *Postgrad Med J*. 2008;84:586-589.
6. Ford J. Red blood cell morphology. *Int J Lab Hematol*. 2013;35:351-357.
7. Bain BJ, Bhavnani M. The gray platelet syndrome. *Am J Hematol*. 2011;86:1027.
8. Sun P, Kowalski EM, Cheng CK, et al. Predictive significance of absolute lymphocyte count and morphology in adults with a new onset peripheral blood lymphocytosis. *J Clin Pathol*. 2014;67:1062-1066.
9. Tay HS, Mills A, Bain BJ. Diagnosis from a blood film following dog-bite. *Am J Hematol*. 2012;87:915.

REVIEW QUESTIONS

1. A 72-year-old woman presents with gradual onset of digital paresthesia and difficulty with fine movement of the hands. More recently, she has noticed an electrical sensation running down her trunk and into her limbs when she bends her head forward. She is found to have a hemoglobin concentration of 9.8 g/dL and a mean cell volume of 107 fL. A blood film shows macrocytes, oval macrocytes, and hypersegmented neutrophils. Which of the following is the most likely diagnosis?

A. Neuroacanthocytosis
B. Vitamin B_{12} deficiency
C. Folic acid deficiency
D. Alcoholic liver disease
E. Copper deficiency

Answer: B The patient has described the features of peripheral neuropathy and Lhermitte syndrome. These neurologic features in a patient with macrocytic anemia and hypersegmented neutrophils are indicative of vitamin B_{12} deficiency causing megaloblastic anemia, peripheral neuropathy, and subacute combined degeneration of the spinal cord (Chapter 164). (Larkman N, Hulson O, Gilhooly M. Picture quiz: a man with tingling fingers. *BMJ.* 2013;346:37-38.)

2. A 67-year-old man presents with right heart failure and extensive bruising. The cardiac silhouette is normal on chest radiography. Electrocardiography shows reduced voltage of QRS complexes. Coagulation screen shows prolongation of the prothrombin time and activated partial thromboplastin time. A full blood count is normal apart from a slight increase of the platelet count to 487×10^9/L. A blood film shows Howell-Jolly bodies, target cells, occasional acanthocytes, and platelet anisocytosis. There is no relevant previous medical history. Which of the following is the most likely diagnosis?

A. Essential thrombocythemia
B. Hypothyroidism
C. AL (light-chain associated) amyloidosis
D. Acquired hemophilia
E. Disseminated intravascular coagulation

Answer: C The blood film features are those of hyposplenism, which also provides an explanation for the slight increase in the platelet count. The hyposplenism is the result of extensive amyloid deposition in the spleen. AL amyloidosis causes a variety of coagulation abnormalities, including acquired factor X deficiency, which could explain the prolonged prothrombin and activated partial thromboplastin times (Chapter 188). Acquired hemophilia would not explain the hyposplenism. (Gatt ME, Palladini G. Light chain amyloidosis 2012: a new era. *Br J Haematol.* 2013;160:582-598; and Gillmore JD, Wechalekar A, Bird J, et al. Guidelines on the diagnosis and investigation of AL amyloidosis. *Br J Haematol.* 2015;168:207-218.)

3. A young Greek woman presents to the emergency department because of the sudden onset of jaundice and fatigue. She has just returned from a brief trip to Greece for a family wedding. She is taking no medications. She is noted to be pale and has tachycardia. Her hemoglobin concentration is 8.7 g/dL and mean corpuscular volume is 109 fL. Reticulocyte count is increased to 280×10^9/L. Bilirubin is elevated and is found to be mainly unconjugated. Her blood film confirms marked anemia and shows polychromatic macrocytes, irregularly contracted cells, hemighosts (blister cells), and a few ghost cells. An assay for glucose-6-phosphate dehydrogenase is normal. Which of the following is the most likely diagnosis?

A. Acute liver failure
B. Autoimmune hemolytic anemia
C. Glucose-6-phosphate dehydrogenase deficiency
D. Hereditary spherocytosis
E. Pyruvate kinase deficiency

Answer: C Unconjugated hyperbilirubinemia does not suggest acute liver failure. The blood film features are those of oxidant-induced hemolysis, and a diagnosis of glucose-6-phosphate dehydrogenase (G6PD) deficiency must therefore be considered. The blood film features are not consistent with the other types of hemolytic anemia listed. Most patients with acute hemolysis due to G6PD deficiency are men because this is an X-linked condition. However, hemolysis can occur in women because a proportion of red cells will be deficient. If there is marked lyonization, hemolysis may be severe. The G6PD assay is likely to be normal because the remaining red cells are not deficient and the G6PD concentration will be increased by the presence of reticulocytes. In this circumstance, the blood film is very important in suggesting the right diagnosis. A dietary history should be taken. It is possible that the patient has recently eaten broad beans (Chapter 161). (Bain BJ. Sudden onset of jaundice in a Sardinian man. *Am J Hematol.* 2008;83:810; Bain BJ. A ghostly presence. *Am J Hematol.* 2010;85:271.)

4. A 35-year-old pregnant woman complains of severe fatigue and swollen ankles. Her blood count shows the following: red blood cells, 4.22×10^{12}/L; hemoglobin concentration, 7 g/dL; hematocrit, 0.29 L/L; mean corpuscular volume, 67 fL; mean corpuscular hemoglobin, 16.6 pg; mean corpuscular hemoglobin concentration, 24.5 g/dL. Her blood film shows hypochromia, microcytosis, and poikilocytosis including elliptocytes (pencil cells), target cells, and acanthocytes. Some Howell-Jolly bodies are present. Which of the following is the most likely diagnosis?

A. β-Thalassemia heterozygosity
B. Chronic blood loss
C. Celiac disease
D. Congenital sideroblastic anemia
E. Dietary iron deficiency

Answer: C The blood count is suggestive of iron deficiency, not β-thalassemia heterozygosity (in which moderately severe anemia would not be expected). The blood film shows the features of iron deficiency (hypochromia, microcytosis, and pencil cells) and of hyposplenism (acanthocytes, target cells, and Howell-Jolly bodies). The combination of hyposplenism and iron deficiency suggests a diagnosis of celiac disease with splenic atrophy (Chapter 159). (Croese J, Harris O, Bain BJ. Coeliac disease. Haematological features and delay in diagnosis. *Med J Aust.* 1979;ii:335-338.)

5. A 24-year-old woman who was hypertensive during pregnancy has an epileptiform convulsion 24 hours postpartum. She has jaundice, edema, and proteinuria. An urgent blood count shows leukocytosis, neutrophilia, anemia, and thrombocytopenia. Her blood film confirms the thrombocytopenia and in addition shows left shift, toxic granulation, and red cell fragments (schistocytes). Which of the following is the most appropriate diagnosis?

A. Disseminated intravascular coagulation
B. Hemolysis, elevated liver enzymes, and low platelets (HELLP) syndrome
C. Hemolytic-uremic syndrome
D. Preeclampsia
E. Thrombotic thrombocytopenic purpura

Answer: B This syndrome is a severe form of preeclampsia/eclampsia, occurring during pregnancy or within 48 hours of delivery. The hemolytic anemia is microangiopathic in nature, and the detection of schistocytes is diagnostically useful. (Bain BJ, Riches J. Help with HELLP. *Am J Hematol.* 2010;85:70; Zini G, d'Onofrio G, Briggs C, et al; International Council for Standardization in Haematology [ICSH]. ICSH recommendations for identification, diagnostic value, and quantitation of schistocytes. *Int J Lab Hematol.* 2012;34:107-116.)

158

APPROACH TO THE ANEMIAS

H. FRANKLIN BUNN

Anemia is defined as a significant reduction in the mass of circulating red blood cells. As a result, the oxygen binding capacity of the blood is diminished. Because blood volume is normally maintained at a nearly constant level, anemic patients have a decrease in the concentration of red cells or hemoglobin in peripheral blood. As shown in Table 158-1, hemoglobin and hematocrit levels vary with the age of the individual and, in adults, with gender. The values in women of childbearing age are 10% lower than those in men. At altitude, higher values are found, roughly in proportion to the elevation above sea level. Anemic patients' values are more than 1 standard deviation below the mean values for their gender. However, because of the wide range in normal hemoglobin and hematocrit levels, it is often difficult to document mild anemia. Anemia affects one fourth of the world's population, with a higher prevalence in low socioeconomic groups.[1]

Sometimes the diagnosis of anemia is confounded by a concomitant change in the plasma volume. For example, if a patient with a low red cell mass sustains a loss of plasma volume from dehydration, diarrhea, vomiting, or severe burns, the blood hemoglobin and hematocrit levels will be increased and may even be in the normal range. Another important example, discussed in detail later in this chapter, is acute hemorrhage, in which the loss of both red blood cells and plasma results in a false elevation of hemoglobin and hematocrit. In contrast, hemoglobin and hematocrit values may be falsely low in patients with an expanded plasma volume, such as in pregnancy or congestive heart failure.

PATHOBIOLOGY

Impact of Anemia on Oxygen Transport

In any organ or region of the body, the transport of oxygen is a product of three independent variables expressed in the Fick equation (Fig. 158-1). The middle variable—the oxygen carrying capacity of the blood—is, by definition, low in anemic patients. The two other variables in the Fick equation undergo compensatory changes that, as explained later, greatly enhance oxygen transport.[2]

TABLE 158-1 NORMAL VALUES FOR RED BLOOD CELL MEASUREMENTS

MEASUREMENT	UNIT	NORMAL RANGE (APPROXIMATE)*
Hemoglobin	g/dL	Males: 13.5-17.5 Females: 12-16
Hematocrit	%	Males: 40-52 Females: 36-48
Red blood cell (RBC) count	$\times 10^6$/μL of blood	Males: 4.5-6.0 Females: 4.0-5.4
Mean cell volume (MCV)	fL	81-99
Mean cell hemoglobin (MCH)	pg	30-34
Mean cell hemoglobin concentration (MCHC)	g/dL	30-36
RBC size distribution width RDW-CV† RDW-SD†	 % fL	 12-15 37-47
Reticulocyte count (absolute number)	No./μL of blood	20,000-100,000
Reticulocyte percentage	% of RBCs	0.5-1.5

*Actual normal ranges for many of these values may vary slightly, depending on factors such as the location and type of laboratory instruments used, altitude above sea level, and patient age.
†Depending on analyzer instrument used, the RDW (red cell distribution width) can be reported as coefficient of variation (CV) and/or standard deviation (SD), RDW-CV and/or RDW-SD, respectively.

O_2 Delivery = Blood Flow × Hb Concentration × (Asat – Vsat)

In anemia: ↑ Cardiac output, Altered flow distribution | ↑↑ Plasma Erythropoietin | ↑RBC 2,3-DPG, ↓RBC O_2 affinity

FIGURE 158-1. The Fick equation expresses the three independent variables that determine the transport of oxygen to a given organ or tissue. The impact of anemia on each of these variables is shown beneath the equation. Asat = oxygen saturation of arterial blood (oxyhemoglobin/oxyhemoglobin + deoxyhemoglobin); 2,3-DPG = 2,3-diphosphoglycerate (2,3-bisphosphoglycerate); Hb = hemoglobin; RBC = red blood cell; Vsat = oxygen saturation of venous blood.

Blood Flow

Anemia has a marked impact on blood flow, the left-hand variable in the Fick equation. In all anemic individuals, there is enhanced flow to vital organs, including the heart, brain, liver, and kidneys, at the expense of nonvital organs. Anemic patients are pale because blood is diverted away from the skin and mucous membranes to preserve oxygen supply to the critical organs. Resting cardiac output is normal in patients with mild or moderate anemia, but with exercise, it increases more than that of a normal individual. In severe anemia, resting cardiac output is increased, putting patients at risk for developing high-output cardiac failure, particularly those with coronary artery insufficiency or other types of preexisting cardiac disease.

Oxygen Binding to Hemoglobin

The variable on the right side of the Fick equation is the difference in fractional oxygenation between the arterial and venous blood. This difference in oxygen saturation is determined by the hemoglobin oxygen-binding curve. A comparison between a normal individual and an anemic patient is shown in Figure 158-2. As shown in part A, the curve is shifted to the right in an anemic patient. At any given oxygen tension (P_{O_2}), the oxygen saturation of hemoglobin is lower. Thus, red cells of anemic patients have decreased oxygen affinity. This change is due entirely to elevated levels of red cell 2,3-diphosphoglycerate (2,3-DPG) in red cells. This glycolytic intermediate is the principal determinant of oxygen affinity in human red cells. The P_{O_2} in arterial blood is normally approximately 95 mm Hg, resulting in nearly 100% oxygen saturation. During the transit of red cells from an artery through its capillary bed to its vein, oxygen is released to respiring cells. In normal individuals, at a normal venous P_{O_2} of about 40 mm Hg, the oxygen saturation is approximately 80%. Thus, as shown in Figure 158-2A, 20% of the oxygen in the blood is unloaded. In contrast, in patients with anemia and elevated red cell 2,3-DPG levels, the lower oxygen affinity of their red cells enables a much higher fraction of the oxygen (as much as 35%) to be unloaded. In Figure 158-2B, the oxygen-binding curves are depicted with the volume fraction of oxygen plotted on the *y*-axis. One gram of hemoglobin binds up to 1.34 mL of oxygen under standard conditions of temperature and pressure. Thus, in a normal individual having a hemoglobin of 15 g/dL, the oxygen-carrying capacity of the blood is 15 × 1.34, or 20 mL O_2/dL. As mentioned earlier, 20% of this oxygen will be unloaded, that is, 4 mL O_2/100 mL blood during arterial-venous transit. In contrast, an anemic patient with a hemoglobin of 7.5 g/dL has an oxygen-binding capacity that is half normal, or 10 mL O_2/dL. If this patient had red cells with normal oxygen affinity, 20%, or only 2 mL of oxygen, would be unloaded per 100 mL of blood. However, because the patient's red cells have a lower affinity for oxygen, 3.5 mL is unloaded, nearly as much as normal. Thus, the decrease in oxygen affinity is an important mechanism by which anemic patients compensate for the deficit in red cell mass.

Methemoglobinemia

In order for hemoglobin to reversibly bind oxygen, the iron atom in the heme must be in the reduced (Fe^{2+}) valence state. As red cells circulate, the heme iron slowly auto-oxidizes to Fe^{3+}, forming methemoglobin, which is incapable

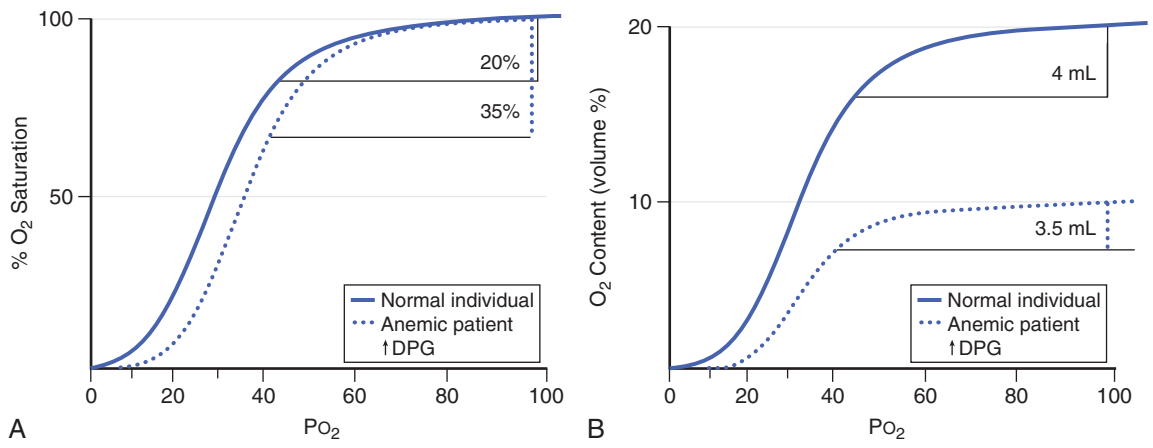

FIGURE 158-2. Oxygen-binding curves for hemoglobin of a normal individual and a patient with anemia. A, Conventional plot of percent of oxygen (O_2) saturation versus oxygen tension (P_{O_2}). **B,** The *y*-axis shows the volume of oxygen in milliliters per 100 mL of blood. DPG = diphosphoglycerate.

of binding oxygen. Normal red cells are endowed with a very efficient enzymatic pathway composed of cytochrome b_5, cytochrome b_5 reductase, and NADH that rapidly reduces the iron in methemoglobin back to its functional Fe^{2+} form. Thus, normal red cells contain less than 0.5% methemoglobin. However, either an inherited deficiency in cytochrome b_5 reductase or exposure to an oxidant drug or toxin can result in methemoglobinemia.[3] Laboratory samples of blood containing methemoglobin are dark brown, whereas patients with greater than 10% methemoglobinemia have cyanosis, a blue discoloration of the skin indistinguishable from that commonly seen in patients having normal hemoglobin but low oxygen saturation owing to pulmonary or cardiac disease. In many hospitals and large clinical laboratories, the instrument that measures oxygen saturation in blood samples also provides an accurate determination of methemoglobin.

Patients with congenital methemoglobinemia inherit an autosomal recessive deficiency in cytochrome b_5 reductase. Heterozygote relatives have low or undetectable methemoglobin levels, whereas affected individuals (homozygotes and compound heterozygotes) generally have 10 to 35% methemoglobin. These individuals are usually asymptomatic because the methemoglobin is distributed primarily in the older population of red cells. Nevertheless, many affected individuals have cosmetic concerns. Treatment with oral ascorbic acid or riboflavin is effective in lowering the level of methemoglobin below the threshold of detectable cyanosis.

A variety of drugs can cause methemoglobinemia, including acetaminophen (Tylenol), dapsone, nitroprusside, amyl nitrate, procaine congeners used for local anesthesia[4], and recreational drugs (volatile nitrites called "poppers" and cocaine). It is not clear why only a very small fraction of those using these drugs develop this complication, but some affected individuals have been shown to be heterozygous for cytochrome b_5 reductase deficiency. When these drugs are taken in prescribed doses, methemoglobinemia seldom reaches levels high enough to cause clinical concern.

In contrast, individuals exposed to industrial toxins such as nitrite, nitrate, or aniline may develop life-threatening levels of methemoglobin. The threshold at which symptoms occur is highly variable. Acute induction of 20% methemoglobin may cause fatigue; at 30%, individuals often develop tachycardia. When methemoglobin exceeds 50%, patients experience weakness, breathlessness, and confusion. At 70 to 80%, coma and death may occur. The toxicity of methemoglobinemia is not just because of the inability of oxidized hemes to bind oxygen; the remaining functional (Fe^{2+}) hemes in the hemoglobin tetramer have increased oxygen affinity and therefore, as suggested in Figure 158-2, are much less effective in releasing oxygen to tissues. Patients with toxic methemoglobinemia can be effectively treated with intravenous infusion of methylene blue (1 to 2 mg/kg).

Regulation of Erythropoiesis by Erythropoietin

Anemia also affects the middle component of the Fick equation (see Fig. 158-1). As mentioned earlier, hemoglobin levels are, by definition, low in anemic patients. The resultant decrease in the oxygen-carrying capacity of the blood causes cellular hypoxia. In all cells of the body, a molecular sensor detects even modest degrees of low oxygen tension and induces a hypoxia-inducible transcription factor called HIF. HIF upregulates expression of the hormone erythropoietin in the kidney and, to a lesser extent, in the liver. Erythropoietin (Chapter 156) binds to a specific receptor abundantly expressed on erythroid progenitor cells in the bone marrow and salvages these cells from apoptosis, thereby enhancing red blood cell production. Normal individuals maintain nearly constant levels of circulating red cells by finely tuned regulation of erythropoietin production. In anemic patients, the hypoxic signal in the kidneys and, to a lesser extent, in the liver results in a robust induction of erythropoietin expression. As the hematocrit falls, the plasma erythropoietin level rises markedly; in severely anemic patients, it may be 1000-fold higher than normal (Fig. 158-3). In patients with anemia due to impaired red cell production, the erythroid progenitors are unresponsive to such high levels of plasma erythropoietin. In contrast, in patients whose anemia is due to hemolysis or blood loss, elevated erythropoietin levels maximize red cell production.

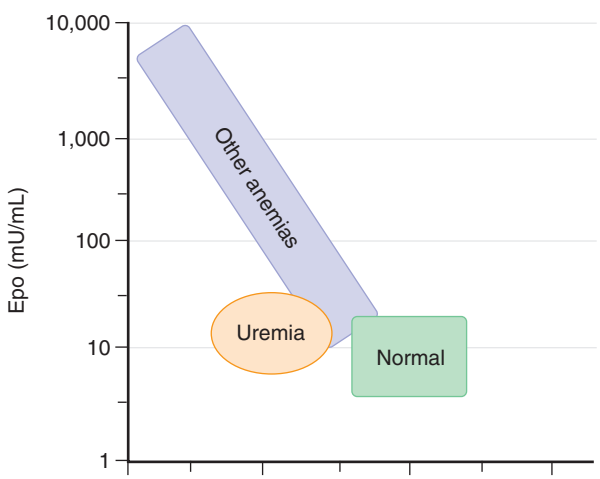

FIGURE 158-3. Plasma erythropoietin (Epo) levels in patients with different degrees of anemia. The subset of anemia patients with chronic renal disease (labeled "Uremia") have much lower plasma erythropoietin levels than those with other types of anemia. Values for normal individuals are also shown.

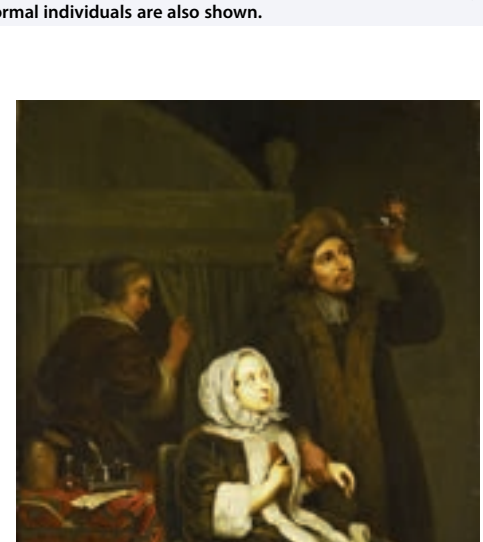

FIGURE 158-4. "The Sick Lady," a 17th-century painting attributed to Caspar Netscher, from the Royal Collection, Buckingham Palace.

CLINICAL MANIFESTATIONS

Figure 158-4 is a 17th-century painting of a pale young woman clutching her chest, apparently complaining of palpitations. Her physician is feeling her pulse, documenting her rapid, forceful heartbeat. These signs and symptoms, common in patients with very low hemoglobin levels, can be readily explained by the cardiovascular adjustments discussed in the preceding section. These clinical findings pertain to anemia per se, irrespective of cause, and are dependent on its severity and chronicity. The history and physical examination may reveal additional findings peculiar to specific causes of anemia or to other comorbid conditions. The degree to which symptoms occur in an anemic patient depends on several contributing factors. If the anemia has developed rapidly, there may not have been adequate time for compensatory adjustments to take place, and the patient may have more marked symptoms than if an anemia of equivalent severity had developed insidiously. Furthermore, the patient's complaints may depend on the presence of local vascular disease. For example, symptoms owing to ischemia in patients with angina pectoris, intermittent claudication, or transient cerebral episodes may be triggered by the development of anemia.

Symptoms

Many individuals with mild anemia have no complaints and are unaware that they have "tired blood." Others may complain of fatigue as well as dyspnea and palpitations, particularly following exercise. Patients with severe anemia are often symptomatic at rest and are unable to tolerate significant exertion. If the hemoglobin concentration falls below 7.5 g/dL, resting cardiac output is likely to rise, with an increase in both stroke volume and heart rate. The patient may be aware of this hyperdynamic state and complain of a rapid,

pounding sensation in the precordium. Patients with compromised myocardial reserve may develop complaints due to cardiac failure.

The symptoms of severe anemia often extend beyond the cardiac or circulatory system. Patients sometimes experience dizziness and headache and, less often, syncope, tinnitus, or vertigo. Many are irritable and have difficulty sleeping or concentrating. Because blood flow is shunted away from the skin, patients may complain of increased sensitivity to cold. In like manner, gastrointestinal symptoms such as indigestion, anorexia, or even nausea are attributable to the shunting of blood away from the splanchnic bed. Females commonly develop abnormal menstruation, either amenorrhea or increased bleeding. Males may experience impotence or loss of libido.

Physical Findings

Pallor is the most commonly encountered physical finding in patients with anemia. As mentioned earlier, this sign is due to the shunting of blood away from the skin and other peripheral tissues, permitting enhanced blood flow to vital organs. The usefulness of pallor as a physical finding is limited by other factors that affect the appearance of the skin. Blood flow to the skin may undergo wide fluctuations. Moreover, the skin's thickness and texture vary widely among individuals. Those with a fair complexion may appear pale even though they are not anemic, whereas pallor is difficult to detect in deeply pigmented individuals. The amount of melanin in the epidermis is an important determinant of skin color. Pallor may be difficult to detect in patients who have increased melanin pigmentation due to Addison disease or hemochromatosis. Nevertheless, even in blacks, the presence of anemia may be suspected by the color of the palms or of noncutaneous tissues such as oral mucous membranes, nail beds, and palpebral conjunctivas. When the creases of the palm are as pale as the surrounding skin, the patient usually has a hemoglobin of less than 7 g/dL.

In addition to tachycardia, wide pulse pressure, and hyperdynamic precordium, a systolic ejection murmur is often heard over the precordium, particularly at the pulmonic area. In addition, a venous hum may be detected over the neck vessels. These findings disappear when the anemia is corrected.

DIAGNOSIS

Laboratory Evaluation of the Patient with Anemia

In the clinical assessment of the anemic patient, it is important to proceed in a systematic way so that the diagnosis can be established with a minimum of laboratory tests and procedures. A thorough history and careful physical examination are critical in the initial evaluation of the anemic patient. For example, a family history that reveals a dominant inheritance pattern would reinforce the tentative diagnosis of hereditary spherocytosis. The presence of fever, a new heart murmur, and splenomegaly is an anemic patient suggests subacute bacterial endocarditis.

In evaluating the anemic patient, the clinician must first ask whether the anemia is caused by decreased production of red cells or by loss of blood cells as a result of hemorrhage or hemolysis (Table 158-2). Blood loss may be either the sole cause of the anemia or a significant contributor. Therefore, examination of the stool for occult blood is an indispensable part of the evaluation of all anemic patients.

The laboratory work-up of anemia includes a complete blood count, red cell indices, reticulocyte count, and microscopic examination of the blood smear. In addition, in many cases a bone marrow examination is a critical component of the initial laboratory assessment.

TABLE 158-2 INITIAL ASSESSMENT OF ANEMIA

Decreased red cell production
Usually acquired
Onset is insidious
Reticulocyte count is inappropriately low
Red cell indices (MCV, MCHC) are informative
Bone marrow examination is often required for diagnosis
Increased red cell destruction (hemolysis)
Often inherited
Onset may be abrupt or insidious
Reticulocyte count is increased
Red cell morphology on peripheral blood smear is usually informative
Bone marrow examination is usually not indicated
Blood loss—must be ruled out in any patient with anemia

MCHC = mean cell hemoglobin concentration; MCV = mean cell volume.

Complete Blood Count

Most hospitals and clinical laboratories use equipment that provides high-throughput analyses of red cell, platelet, and white cell counts and white cell differential, along with measurements of cell size. The mean red cell volume (MCV) is normally 81 to 99 fL. These instruments also provide accurate determinations of hemoglobin concentration. The hematocrit, or fraction of packed red cells over total blood volume, is determined indirectly from the red cell count and the MCV. The mean concentration of hemoglobin within the red cell population (MCHC) is the quotient of hemoglobin divided by hematocrit. The MCV is particularly useful in classifying the anemias caused by decreased red cell production. Microcytic anemias have low MCV values and often low MCHC. Microscopic examination reveals small and often pale red cells. The MCV in the macrocytic anemias is increased, and large, oval cells (macro-ovalocytes) are seen. In contrast to the anemias of underproduction, the hemolytic anemias are either normocytic or slightly macrocytic owing to the preponderance of young red cells that are relatively large. Severe forms of thalassemia (Chapter 162) are an exception; there, microcytic red cells may be accompanied by brisk hemolysis.

Reticulocyte Count

This simple and cost-effective test is extremely useful for distinguishing anemias secondary to decreased red cell production from those caused by hemolysis. With the application of an appropriate supravital stain, the 1- to 2-day-old red cells in the peripheral blood reveal a network of purple strands, which are aggregates of ribosomes. The reticulocyte count in normal individuals is about 1%, consistent with a red cell lifespan of approximately 120 days. An elevated reticulocyte count reflects the release of an increased number of young cells from the bone marrow. The rate of red cell production can be assessed more quantitatively by determining the absolute reticulocyte count, the product of the percentage of reticulocytes and the red cell count. Thus, normal blood contains about 50,000 reticulocytes/mm^3. In interpreting this test, one should consider the distribution of reticulocytes between the bone marrow and the peripheral blood. When erythropoiesis is robust, marrow reticulocytes enter the circulation prematurely. These "shift reticulocytes" appear larger than average on a routine (Wright-stained) blood smear and have a lavender hue, called polychromatophilia. Because the circulation of shift reticulocytes in the peripheral blood is prolonged, the reticulocyte count should be divided by two. This correction should always be made if normoblasts are encountered in the peripheral blood because this finding indicates the premature release of newborn red cells into the circulation.

A failure to produce red cells is reflected in an inappropriately low reticulocyte count. In contrast, a significant elevation of reticulocytes is suggestive of hemolysis. Exceptions include the following:

- The brisk reticulocyte response seen in patients with hemorrhage
- Reticulocytosis encountered in patients recovering from impaired erythropoiesis (e.g., an individual with pernicious anemia who received an injection of cobalamin 1 week earlier)
- Mild to moderate elevations in reticulocytes (3 to 7%) encountered in myelophthisic anemia (Chapter 157), in which the orderly release of cells is affected by alterations of the marrow stroma owing to tumor, fibrosis, or granuloma

These exceptions are generally appreciated in the initial evaluation of the patient.

A number of ancillary laboratory tests described later under Hemolytic Anemias are useful in determining both the cause and extent of hemolysis.

Examination of the Blood Smear

In the evaluation of any patient with unexplained anemia, the physician should take the time to examine a well-stained peripheral blood film (Chapter 157). Many subtleties escape the attention of the technologist, whose primary aim is to confirm or refine the white cell differential count provided by automated cell counters. The clinician approaches the specimen with a prepared mind and can scrutinize it for specific abnormalities. Examination of the blood film can confirm the size and color of red cells estimated by red blood cell indices. In contrast to the mean statistical values provided by automated cell counters, microscopic examination can reveal variations in red cell size (anisocytosis) or shape (poikilocytosis), abnormalities that are helpful in diagnosing specific anemias. Examination of the blood smear is particularly important in a patient with hemolysis. Many types of hemolytic anemia have characteristic abnormalities in red cell morphology. The presence of

abnormal white cells may be the first clue to a lymphoproliferative or primary bone marrow disorder.

Bone Marrow Examination

A microscopic examination of the bone marrow (aspirate with or without a core biopsy) is often useful and may be critical in the work-up of any *unexplained* anemia. Study of the bone marrow is informative in the diagnosis of anemias of underproduction, particularly those accompanied by abnormalities in white cells and/or platelets, suggesting disordered hematopoiesis. The more severe the anemia, the more likely it is that the procedure will be informative. An assessment of the quantity and quality of red cell precursors may identify a defect in cell production due to either hypoplasia or ineffective erythropoiesis. A marrow biopsy is required for estimating overall cellularity. The ratio of myeloid (M) to erythroid (E) precursors is normally about 2 : 1, but it may be artifactually increased by the inclusion of circulating leukocytes. The ratio is increased in patients with infection, a leukemoid reaction, or neoplastic proliferation of myeloid cells. Rarely, a high M/E ratio is due to selective aplasia of the red cell precursors. A decreased M/E ratio indicates erythroid hyperplasia. Erythroid maturation is normal in hemolysis and hemorrhage, but it is disordered when erythropoiesis is ineffective, such as in megaloblastic and sideroblastic anemias and in β-thalassemia major or intermedia. The bone marrow examination is also important in demonstrating the presence of cellular infiltrates such as those found in leukemia, lymphoma, or multiple myeloma. The demonstration of tumor, fibrosis, or granuloma usually requires a bone marrow biopsy, which provides information not available from bone marrow aspiration. A portion of the marrow specimen should be stained with Prussian blue. In addition to providing an assessment of iron stores, this iron stain is required for the identification of erythroid sideroblasts.

ANEMIA DUE TO BLOOD LOSS

The clinical presentation of anemia resulting from blood loss varies considerably, depending on the site, severity, and rapidity of the hemorrhage. At opposite extremes are acute fulminant bleeding producing hypovolemic shock and chronic occult blood loss leading to iron deficiency anemia.

Acute Blood Loss

Patients who have had a sudden hemorrhage present with clinical findings secondary to hypovolemia and hypoxia. Symptoms and signs depend on the severity of the process. The patient may experience weakness, fatigue, lightheadedness, or stupor and may appear pale, diaphoretic, and irritable. The vital signs reflect cardiovascular compensation for the acute blood loss. The degree of hypotension and tachycardia depends on the extent of the hemorrhage. Elicitation of postural signs is useful in the initial evaluation of a patient with acute blood loss. When a patient is lifted from a supine to a sitting position, an increase in the pulse of 25% or more or a fall in the systolic blood pressure of 20 mm Hg or more signifies significant hypovolemia (blood loss >1000 mL). Acute blood loss in excess of 1500 mL usually leads to cardiovascular collapse.

Following acute hemorrhage, the red cell mass and plasma volume are contracted in parallel; accordingly, there is often not a significant decrease in the hemoglobin or hematocrit level initially. This stress induces a moderate leukocytosis and a "shift to the left" in the white cell differential count. In both acute and chronic blood loss, the platelet count is often increased, particularly if the patient is already iron deficient. During the first few days after acute blood loss, there is usually an increase in reticulocytes. Severe hypoxia may trigger the release of nucleated red cells from the bone marrow into the peripheral blood. Because young red cells are larger than old ones, the MCV generally rises slightly. If significant blood loss continues, reticulocytosis will persist until iron stores have been exhausted. Internal bleeding is often accompanied by an increase in unconjugated bilirubin, reflecting an increase in the catabolism of heme from extravasated red cells. Patients with acute gastrointestinal blood loss sometimes have an elevation of blood urea nitrogen owing to impaired renal blood flow and perhaps to the absorption of digested blood protein.

These patients must be assessed promptly, and treatment must be initiated without delay. Patients with severe acute blood loss require transfusion of packed red cells, with central monitoring of the appropriate amount of volume replacement. The site or sites of bleeding should be emergently identified and controlled. In addition, an emergency coagulation profile should be obtained. The approach to the patient with shock is discussed in detail in Chapter 106.

Chronic Blood Loss

Chronic blood loss is usually due to lesions in the gastrointestinal tract or the uterus. Testing of stool specimens for occult blood is an essential but frequently overlooked part of the evaluation of anemia. It is sometimes necessary to examine serial specimens over a prolonged period because gastrointestinal bleeding may be intermittent. The hematologic manifestations of chronic blood loss are those of iron deficiency anemia, discussed in detail in Chapter 159.

ANEMIAS DUE TO DECREASED RED CELL PRODUCTION

As shown in Table 158-3, anemias caused by the underproduction of red cells can be conveniently classified according to red cell size: microcytic, macrocytic, and normocytic.

Microcytic Anemias

The presence of small red cells (MCV <77 fL) indicates a defect in the production of hemoglobin.[5] As shown in Figure 158-5, hemoglobin is composed of globin subunits into which heme is inserted. Heme is produced by the insertion of an iron atom into porphyrin (protoporphyrin IX). A defect in any of these three key components can cause microcytic anemia. Most individuals with microcytosis have either iron deficiency anemia (Chapter 159) or thalassemia (Chapter 162). A congenital or, more often, acquired impairment in porphyrin synthesis can lead to a buildup of excess iron in erythroid cells, resulting in the morphologic entity of ringed sideroblasts, which are identified in red cell precursors in the bone marrow (Chapter 159). Most patients with acquired sideroblastic anemia actually have a normal or somewhat elevated MCV but a broad distribution of red cell size, including a population of microcytes (because of this ambiguity, sideroblastic anemia appears in parentheses in Table 158-3). Iron deficiency anemia, the thalassemias, and sideroblastic anemia all involve some degree of ineffective erythropoiesis.

The anemias of chronic inflammation and malignancy, described in detail later, may be slightly microcytic owing to a defect in the availability of iron. However, these disorders are more often normocytic. Measurement of serum

TABLE 158-3 ANEMIAS DUE TO DECREASED RED CELL PRODUCTION

Microcytic	Abnormal erythroid maturation
Iron deficiency	
Thalassemias	
(Sideroblastic anemia)	
Macrocytic	
Megaloblastic	
Cobalamin deficiency	
Folic acid deficiency	
Other—hemolysis, acute blood loss, aplasia, ethanol	
Normocytic	Decreased erythroid progenitors
Primary bone marrow failure	
Aplasia	
Myelophthisis	
Secondary to a chronic disease	

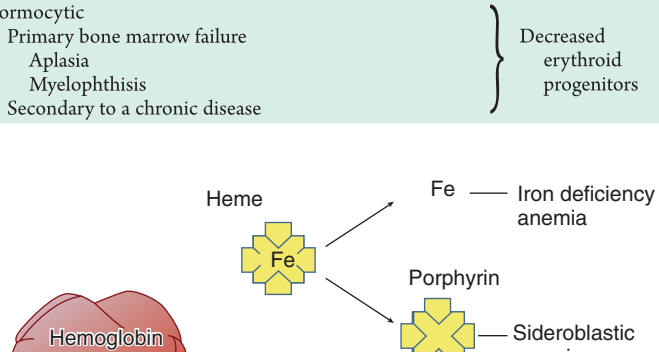

FIGURE 158-5. **Components of hemoglobin that are deficient in the microcytic anemias.**

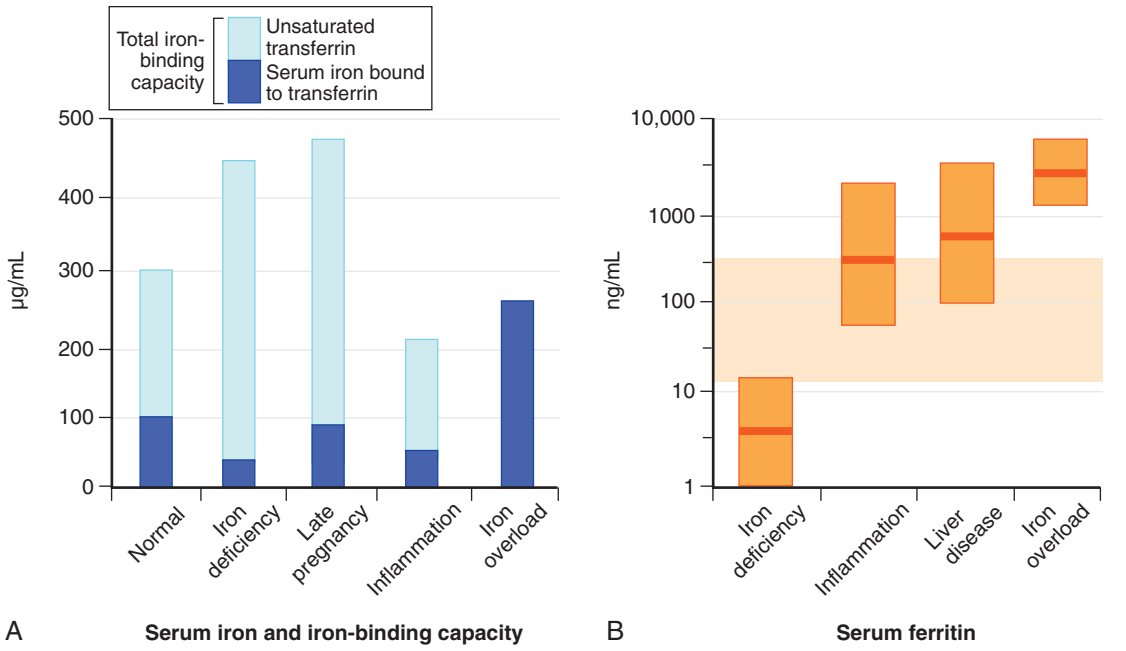

FIGURE 158-6. **A, Serum iron and transferrin saturation in different conditions. B, Serum ferritin in different conditions.** Note that the *y*-axis in panel B is on a log scale. The normal range (10 to 200 ng/mL) is shown by the beige shaded area.

iron and iron-binding capacity (Fig. 158-6A) and serum ferritin (Fig. 158-6B) particularly useful in distinguishing between iron deficiency and the anemia of chronic inflammation.

Macrocytic Anemias

A modest increase in red cell size is encountered in a variety of conditions, including liver disease, hypothyroidism, acute blood loss, hemolytic anemia, aplastic anemia, and alcoholism. Macrocytosis is so commonly seen in alcoholism that the MCV has been used as a clinical screen for abstinence from alcohol. Even in nonalcoholics, alcohol use can elevate the MCV. The macrocytes in liver disease and hypothyroidism may be related to an increased deposition of lipid in the red cell membrane. If the MCV exceeds approximately 105 fL, the patient is likely to be deficient in either cobalamin (vitamin B_{12}) or folic acid. The bone marrow reveals megaloblastic morphology, reflecting impaired replication of DNA. Because nuclear maturation lags behind cytoplasmic development, large red cells tend to be produced in the bone marrow. Megaloblastic anemias are discussed in detail in Chapter 164. Like the microcytic anemias, these disorders are maturation defects associated with ineffective erythropoiesis.

Normocytic Anemias

The normocytic anemias of underproduction are a diverse group of disorders. They can be conveniently divided into two categories: those due to intrinsic pathology within the bone marrow, and those secondary to some other underlying disease.

PRIMARY BONE MARROW DISORDERS

The primary disorders of the bone marrow, such as the leukemias (Chapters 183 and 184), myelodysplasia (Chapter 182), aplastic anemia (Chapter 165), and myelophthisis, are best approached by microscopic examination of a marrow aspirate and biopsy. This group of anemias is often accompanied by leukopenia and thrombocytopenia. A lesser degree of pancytopenia can also occur in hypersplenism and in the megaloblastic anemias.

ANEMIAS OF CHRONIC DISEASE

Among the most common anemias and the ones most prevalent in patients hospitalized on a medical service are those secondary to an underlying chronic disease. The diagnosis is usually quite straightforward. However, in some patients the predisposing illness may not be apparent. Thus, the presence of an unexplained normocytic anemia should prompt the search for the disorders listed in Table 158-4. Even if the presence of an underlying illness is established, the physician should investigate whether other factors such as blood loss or a nutritional deficiency are also contributing to the patient's anemia. Generally, the anemias due to chronic inflammation, an endocrinopathy, or liver disease are of only moderate severity. In contrast, the anemia of uremia is often severe.

TABLE 158-4 ANEMIAS SECONDARY TO CHRONIC DISEASE

Inflammation
Chronic infections
Cancer
Connective tissue disorders
Renal insufficiency
Endocrine disorders
Hypothyroidism
Hypoadrenalism
Hypopituitarism
Hypogonadism—males
Liver disease
Aging

ANEMIA OF CHRONIC INFLAMMATION

If a systemic inflammatory disorder persists for more than a few weeks, it is nearly always accompanied by anemia. As shown in Table 158-4, the most common causes of chronic inflammation are infection, tumor, or a connective tissue disorder. Many chronic infections can be responsible, including tuberculosis, lung abscess, subacute bacterial endocarditis, pyelonephritis, and osteomyelitis. The pathogenesis is more complex in some types of chronic infections. For example, in AIDS (Chapter 393), the human immunodeficiency virus can directly attack hematopoietic cells and suppress erythropoiesis. In malaria and babesiosis, the parasite enters circulating red cells and triggers their destruction.

There is considerable variability in tumors' ability to evoke an inflammatory response. Many tumors express inflammatory cytokines as part of their profile of abnormal gene expression. In some cases, impaired supply of oxygen or nutrients to the interior of the tumor mass can lead to necrosis and an inflammatory response. Red cell production may be further compromised by encroachment of the bone marrow with leukemia, lymphoma, or metastatic tumor.

Anemia is also a feature of a broad range of inflammatory conditions that are not associated with either infection or cancer. In some of these disorders, the autoimmune attack on the patient's cells and tissues is met with a robust inflammatory response. Rheumatoid arthritis (Chapter 264) is the most commonly encountered connective tissue disorder and gives rise to the prototypical anemia of chronic inflammation. Even more intense inflammation

and, accordingly, more severe anemia are seen in polymyalgia rheumatica and temporal arteritis (Chapter 271). In patients with systemic lupus erythematosus (Chapter 266), the anemia of chronic inflammation is often compounded by either immune hemolysis (Chapter 160) or renal insufficiency (discussed later).

PATHOBIOLOGY

Recently, the mechanism underlying the anemia of chronic inflammation has been elucidated by the discovery that plasma hepcidin levels are markedly increased as a result of induction by inflammatory cytokines. As shown in Figure 158-7, hepcidin blocks both iron absorption from the gut and the exit of iron from macrophages, thus explaining both reduced levels of serum iron and increased iron stores.

DIAGNOSIS

The anemia of chronic inflammation is associated with disordered iron homeostasis. Increased storage of iron in macrophages within the bone marrow, liver, and spleen results in elevated levels of serum ferritin (Fig. 158-6B). However, because of a block in the transfer of this excess iron into the plasma, serum iron is low (see Fig. 158-6A). The level of total transferrin in the serum is also low for unclear reasons. With the recent development of a reliable assay for hepcidin, elevated serum levels of this "master regulator" of iron homeostasis should become useful in the diagnosis of the anemia of chronic inflammation. Because of the impairment in iron availability, erythropoiesis is somewhat "iron deficient." The amount of cytoplasmic iron is decreased in erythroid precursors in the bone marrow, and the red cells that enter the circulation are slightly microcytic. This suppression of red cell production is earmarked by a low reticulocyte index. Because this block in iron utilization is subtle, the degree of anemia is seldom severe in patients with inflammatory disorders. If the hemoglobin is less than 8 g/dL, it is necessary to look for additional contributors such as hemolysis or bleeding.

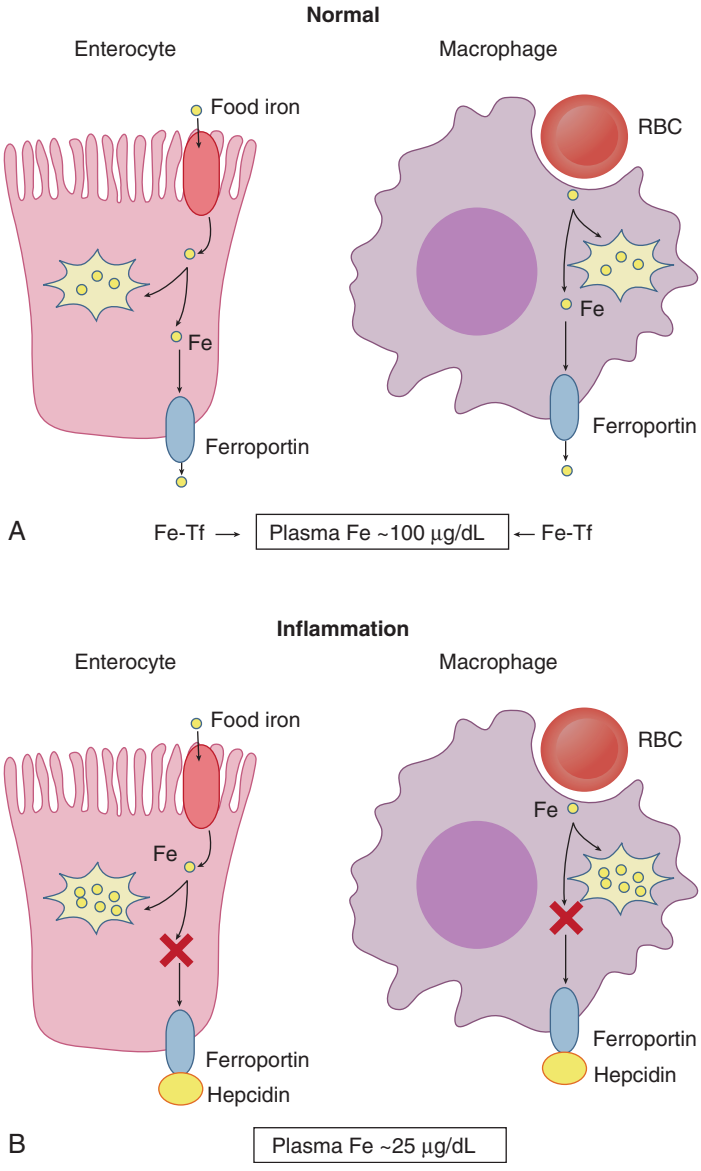

FIGURE 158-7. Pathogenesis of the block in iron availability in the anemia of chronic inflammation. The primary sources of iron in the plasma are from the breakdown of senescent red blood cells (RBCs) within macrophages and from duodenal absorption. **A,** In the presence of physiologically low levels of hepcidin in the plasma, there is efficient release of iron from the duodenal enterocyte and from macrophages through ferroportin. **B,** In patients with inflammation, the induction of plasma hepcidin by interleukin-6 and other cytokines results in the inactivation of ferroportin and the loss of iron egress from the duodenal enterocyte and from the macrophage.

TREATMENT

Rx

Because the anemia of chronic inflammation is not severe, patients seldom require red cell transfusions. Some patients may benefit from recombinant erythropoietin therapy. However, the anemia is not fully corrected unless the underlying disease is effectively treated.

ANEMIA OF RENAL INSUFFICIENCY

A normocytic anemia almost always accompanies uremia (Chapter 130). Although the hemoglobin level is highly variable, the severity of the anemia is roughly proportional to the degree of impaired renal function. The cause of the kidney failure usually has little bearing on the extent of anemia. However, for any level of serum creatinine, patients with polycystic disease tend to be less anemic than those with other types of renal disease. In contrast to the anemias associated with other chronic disorders, the anemia of uremia can be very severe, with hemoglobin levels as low as 4 g/dL.

Examination of the bone marrow seldom reveals any abnormalities. Red cell morphology is usually normal. In a minority of patients, the peripheral blood smear reveals so-called burr cells characterized by an evenly scalloped border. Neither the degree of anemia nor the red blood cell lifespan is influenced by the presence of burr cells. In most patients, the corrected reticulocyte count is low, and the red blood cell survival is only modestly decreased. Thus, the low red blood cell mass is due to decreased red blood cell production.

PATHOBIOLOGY

The primary basis for the anemia is the diseased kidneys' inability to secrete adequate amounts of erythropoietin. Plasma erythropoietin levels are lower than those of nonuremic patients with a comparable degree of anemia (see Fig. 158-3). Erythropoiesis is further impaired but not abolished in patients who have undergone bilateral nephrectomy. In addition, red blood cell production may be suppressed by the accumulation of substances that are normally cleared by the kidneys.

Other factors may aggravate the anemia of renal disease. Uremic patients have a propensity to hemorrhage, owing to a qualitative defect in platelet function. As in other patients, chronic gastrointestinal blood loss leads to iron deficiency. Folic acid deficiency may also occur, owing to the poor nutrition of many patients or to the loss of this vitamin during dialysis. Patients whose renal failure is due to thrombotic thrombocytopenic purpura or hemolytic-uremic syndrome (Chapter 172) have a severe form of microangiopathic hemolytic anemia, with characteristic abnormalities of red blood cell morphology (Chapter 157).

TREATMENT

Rx

Treatment of the anemia of uremia first focuses on reversing the renal failure. A prompt and dramatic correction of the anemia follows successful renal transplantation. Occasionally, polycythemia may be encountered after renal engraftment and may be a harbinger of impending rejection.

In patients who are not candidates for renal transplantation, the treatment of anemia of uremia has been revolutionized by the administration of recombinant human erythropoietin (rHuEPO). The rapid and complete responses that occur underscore the importance of erythropoietin in the pathogenesis of anemia. Figure 158-8 shows the hematologic response of one of the first patients treated with rHuEPO. Within a few days of initiating rHuEPO therapy, the hematocrit approached normal, necessitating a reduction in dose. Before rHuEPO treatment, this patient was overloaded with iron, as documented by increased serum ferritin and nearly full saturation of serum transferrin. As the

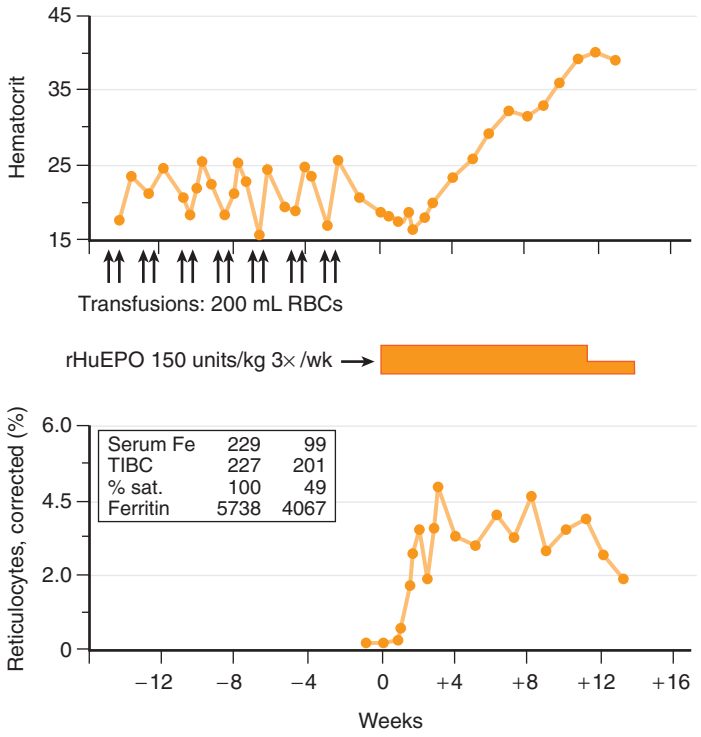

FIGURE 158-8. Response of a uremic patient to recombinant human erythropoietin (rHuEPO) therapy. Before therapy, the patient was severely anemic and transfusion dependent. Treatment with rHuEPO resulted in a reticulocytosis, followed by a progressive increase in hemoglobin. The dose of rHuEPO had to be lowered to prevent the hemoglobin from rising too high. Before rHuEPO therapy, the patient was severely iron overloaded. The marked increase in red cell mass following therapy was accompanied by a significant reduction in iron stores. RBC = red blood cell; TIBC = total iron-binding capacity. (From Eschbach JW, Egrie JC, Downing MR, et al. Correction of the anemia of end-stage renal disease with recombinant human erythropoietin: results of a combined phase I and II clinical trial. *N Engl J Med.* 1987;316:73-78.)

red cell mass increased rapidly following treatment, the robust utilization of iron stores resulted in a decline in serum ferritin and transferrin saturation. In contrast to this patient with iron overload, many uremic patients on dialysis therapy have normal or low iron stores before rHuEPO therapy and need the concomitant administration of iron to maximize the erythropoietic response. (See Erythropoietin Therapy in section on Approach to the Treatment of Anemia.)

ANEMIA OF ENDOCRINE HYPOFUNCTION

The in vitro proliferation of erythroid cells is stimulated by a number of hormones, including thyroxine, glucocorticoids, testosterone, and growth hormone. Therefore, it is not surprising that a mild to moderate normocytic anemia generally accompanies endocrine deficiency states, including hypothyroidism, Addison disease, hypogonadism, and panhypopituitarism.

In the anemia of hypothyroidism, erythropoiesis is suppressed, and the red blood cell lifespan is normal. A minority of patients have macrocytic red blood cells, sometimes owing to cobalamin deficiency. Patients with hypothyroidism have an increased incidence of pernicious anemia. Some patients, particularly females with menorrhagia, develop iron deficiency and a microcytic anemia. The anemia of hypothyroidism may be masked because of a reduction in plasma volume. Because the signs and symptoms of hypothyroidism are sometimes elusive (Chapter 226), this diagnosis should be considered in any patient with unexplained anemia.

The anemia of adrenal insufficiency, including Addison disease, may also be masked by a decrease in plasma volume. Untreated patients have an average hemoglobin level of about 13 g/dL. Upon hormone replacement, the plasma volume is rapidly reconstituted, and the hemoglobin level falls to 80% of its pretreatment value. With continued therapy, the red blood cell mass returns to normal.

Testosterone influences erythropoiesis in a physiologic manner. In males, the mean hemoglobin level increases from 13 to 15 g/dL during the transition from puberty to adulthood. Eunuchoid males usually have a mild anemia (hemoglobin ≈ 13 g/dL). Pituitary dysfunction or ablation is also associated with a mild anemia.

The anemias secondary to endocrine failure are all readily corrected when adequate hormone replacement is given.

ANEMIA OF CHRONIC LIVER DISEASE

Chronic liver disease, regardless of cause (Chapter 146), is usually accompanied by mild or moderate anemia that is normocytic or slightly macrocytic. An increased plasma volume may artificially lower the hematocrit, making the anemia seem worse than it is. Red cell morphology is normal, except for the presence of target cells (see Fig. 157-6) and occasional stomatocytes that have a slitlike rather than a circular area of central pallor. These morphologic features reflect an increased red cell membrane surface area due to increased deposits of cholesterol and phospholipid. The bone marrow is usually normal. Erythropoiesis fails to compensate for a modest shortening of the red cell lifespan. The mechanism underlying the anemia of chronic liver disease is not understood. The anemia is usually corrected if the patient regains normal hepatic function.

In patients with alcoholic liver disease (Chapters 152 and 153), the situation is much more complex. Many factors can contribute to the development of anemia. Alcohol in high doses suppresses not only erythropoiesis but also neutrophil and platelet production. In alcoholics who continue to drink up to the time of clinical evaluation, the bone marrow often reveals vacuoles in the cytoplasm of red and white blood cell precursors. In addition, ringed sideroblasts may be observed, especially if there is concurrent malnutrition. Folic acid deficiency is common in alcoholics because of both a suboptimal diet and an impairment of folate utilization. Moreover, the anemia in alcoholics is often compounded by gastrointestinal hemorrhage as a result of gastric erosions, duodenal ulcers, or esophageal varices. The risk for blood loss is further increased by the presence of thrombocytopenia and/or deficiencies in soluble clotting factors. Although alcoholics usually have increased iron stores, they may become iron deficient after prolonged gastrointestinal bleeding. Rarely, patients with alcoholic cirrhosis develop a severe hemolytic anemia accompanied by the appearance of rigid blood cells with irregular borders called acanthocytes or "spur" cells.

ANEMIA OF THE ELDERLY

As individuals age there is a slight and gradual fall in hemoglobin and hematocrit levels. Elderly individuals whose values fall below 2 standard deviations of normal have significantly enhanced morbidity and mortality. As people age, there is also an increased incidence of cancer, myelodysplasia, renal insufficiency, and chronic inflammatory disorders, all of which can suppress red cell production. Because of the high likelihood of comorbid conditions among the elderly,[6] it is not possible to affirm with any certainty whether aging per se is a cause of anemia. Nevertheless, a fall in hemoglobin in any patient, old or young, should prompt an investigation into the possible presence of one of these underlying disorders. Further studies will be needed to elucidate the molecular pathogenesis of anemia of elderly people, as well as to determine the appropriate hemoglobin concentrations for older adults in light of age, gender, race, and comorbidities.

HEMOLYTIC ANEMIAS

With the exception of sickle cell disease (Chapter 163), hemolytic anemias are encountered much less frequently than those caused by decreased red cell production. Although they are a diverse group, the hemolytic anemias share a number of clinical and laboratory features. Patients with moderate or severe hemolysis may have icterus owing to an elevation in nonconjugated (indirect) bilirubin. In addition, individuals with various types of hemolytic anemia often have splenomegaly, signifying the primary site of enhanced red cell destruction.

PATHOBIOLOGY

The presence of hemolysis is established by the laboratory tests outlined in Table 158-5. Further evaluation is required to establish the specific diagnosis. The clinician saves both time and money by using the available tests in a rational and orderly manner. Diagnosis of hemolytic anemias is greatly facilitated by the use of a logical and pathophysiologically based classification scheme. Table 158-6 groups these disorders going from the outside of the red cell into the cytoplasm, as well as by whether the defect is inherited or acquired. Hemolytic anemias due to environmental factors such as immune destruction or traumatic rupture (Chapter 160) are acquired. Abnormalities of red cell membrane proteins can also cause hemolysis. Mutations in

TABLE 158-5 LABORATORY FEATURES COMMON TO HEMOLYTIC ANEMIAS

Peripheral blood	
Increased reticulocyte count	
Polychromasia	
Bone marrow—erythroid hyperplasia	
Serum	
Increased nonconjugated (indirect) bilirubin	
Elevated lactate dehydrogenase (isoenzymes 1, 2, and 3)	
Decreased or absent haptoglobin	
Plasma hemoglobin	
Extravascular hemolysis: moderately increased	
Intravascular hemolysis: markedly increased	
Urine	
Hemoglobinuria	In intravascular hemolysis
Hemosiderin in urine sediment	In intravascular hemolysis

TABLE 158-6 CLASSIFICATION OF THE HEMOLYTIC ANEMIAS*

Environmental factors	Acquired
Antibody: immunohemolytic anemias	Acquired
Mechanical trauma: TTP, HUS, heart valve	Acquired
Toxins, infectious agents: malaria, etc.	Acquired
Membrane defects	Acquired
Paroxysmal nocturnal hemoglobinuria	Acquired
Spur cell anemia	Acquired
Hereditary spherocytosis, etc.	Congenital
Defects of cell interior	Congenital
Hemoglobinopathies: sickle cell, thalassemia	Congenital
Enzymopathies: G6PD deficiency, etc.	Congenital

*A more detailed differential diagnosis of the hemolytic anemias is presented in Table 160-1 in Chapter 160.
G6PD = glucose-6-phosphate dehydrogenase; HUS = hemolytic-uremic syndrome; TTP = thrombotic thrombocytopenic purpura.

proteins of the red cell cytoskeleton may cause hemolysis of varying severity. The most commonly encountered is hereditary spherocytosis (Chapter 161). Acquired red cell membrane defects are rare. Paroxysmal nocturnal hemoglobinuria is discussed in Chapter 160, and spur cell anemia was mentioned earlier in the section on anemias secondary to chronic liver disease. The proteins in the cytosol of the red cell include hemoglobin and enzymes. Mutations in these proteins can result in inherited hemolytic anemias. Sickle cell disease (Chapter 163) and the thalassemias (Chapter 162) are the most commonly encountered hemoglobinopathies. Homozygous SS disease and the compound heterozygous states SC disease and S/β-thalassemia are pure hemolytic anemias. In contrast, anemia in the clinically significant forms of β-thalassemia is primarily due to ineffective erythropoiesis. By far the most common red cell enzyme defect is glucose-6-phosphate dehydrogenase deficiency (Chapter 161).

DIAGNOSIS

A number of laboratory tests are used to establish the presence of accelerated breakdown of red cells (see Table 158-5). As mentioned in the section Laboratory Evaluation of the Patient with Anemia, the reticulocyte count is the simplest and most cost-effective way to distinguish between hemolytic anemias and those due to decreased red cell production. In this test, a supervital stain or a probe for RNA reveals strands of polyribosomes that are present for only 24 to 48 hours after red cells exit the bone marrow. On a routine Wright or Romanowsky stain, these cells often appear relatively large, with a blue-gray hue (so-called polychromasia). The reticulocyte count is nearly always elevated in patients with hemolysis (unless there is concomitant marrow suppression, such as by folic acid or iron deficiency). This test is a reliable index of red cell production. Thus, in patients with hemolytic anemia, the bone marrow nearly always exhibits erythroid hyperplasia. Because this result is predictable, a bone marrow examination is seldom helpful in patients with hemolytic anemia, unless there is suspicion that the hemolysis is due to an underlying lymphoma.

A number of serum and urine tests are available to confirm the presence of hemolysis and assess its magnitude. As mentioned earlier, serum nonconjugated bilirubin is elevated in proportion to the severity of the hemolysis. Lactate dehydrogenase (LDH) isoforms type 1 through 3 are released from red cells during hemolysis, resulting in increased serum LDH. Most kinds of hemolytic anemia are extravascular, with red cell destruction mediated by macrophages in the spleen, liver, and bone marrow. In these patients, a relatively small amount of hemoglobin is released from engulfed red cells into the plasma, where it binds specifically to haptoglobin. The hemoglobin-haptoglobin complex is rapidly cleared from the circulation. Measurement of serum haptoglobin is a useful test of hemolysis. Most patients with clinically significant hemolysis have very low or absent levels. Less often, patients have intravascular hemolysis with much higher levels of free hemoglobin in the plasma, sufficient to traverse renal glomeruli and exceed the tubular reabsorption capacity. These patients have red or brown urine that, after centrifugation, tests positive with a "dipstick" that detects heme protein. Hemoglobinuria can be readily distinguished from myoglobinuria. In the former, both the plasma and the urine are pigmented. In the latter, the plasma remains colorless because the smaller myoglobin molecule rapidly traverses the glomeruli. Hemoglobinuria is often transient. For a week or so after the episode has abated, the urine sediment will contain hemosiderin, which can be readily detected with the Prussian blue iron stain.

After these general laboratory tests confirm the presence of hemolysis, an array of specific tests is available to establish the specific cause (Chapters 160 through 163).

APPROACH TO THE TREATMENT OF ANEMIA

As in other disorders, the effective treatment of anemia is based on a thorough diagnostic evaluation. Hematinics such as iron, cobalamin, or folic acid should not be administered unless a specific deficiency has been demonstrated or is anticipated. Although indiscriminate treatment with cobalamin is not harmful per se, it lulls both the patient and the physician into a false sense of security in the absence of a firm diagnosis. In contrast, the inappropriate use of iron preparations over a prolonged period can be directly harmful, leading to a state of iron overload in some incorrectly diagnosed patients.

Many kinds of anemias can be corrected if a precipitating cause can be uncovered and reversed. If a drug or toxin is responsible, its withdrawal may allow full recovery. Correction of anemia secondary to a chronic disease usually depends on whether the underlying condition can be reversed. One of the most dramatic dividends of successful renal transplantation is the rapid correction of the anemia of uremia.

Erythropoietin Therapy

The administration of erythropoiesis-stimulating agents (ESA) (epoetin alfa and darbepoetin alfa) is remarkably effective in certain circumstances. In addition to those with the anemia of chronic renal failure, selected patients with other types of anemia may benefit from ESAs. Treatment can lower transfusion requirements in patients with cancer or HIV infection in whom anemia has been aggravated by chemotherapy. In comparison with patients with renal failure, higher doses are required for those with cancer or AIDS to achieve the same increase in red cell mass. Treatment with ESAs has also been effective in some patients with primary bone marrow disorders, particularly myelodysplasia (Chapter 182). Transfusion requirements in surgical patients, both perioperatively and postoperatively, may be reduced by prior short-term administration of ESAs. Treatment may also benefit rare patients who are unable to receive blood transfusions because of either antigen incompatibility or religious convictions.

A note of caution has emerged from a number of large studies suggesting that high doses of ESAs that drive the hemoglobin level above 12 g/dL are associated with a slight but significant increase in the risk for thrombosis and cardiovascular mortality. Systematic reviews and meta-analyses of treatment with ESAs in patients with chronic kidney disease that target higher hemoglobin levels increase the risk for vascular access thrombosis and cardiovascular complications (Chapters 130 and 131),[■■] and likewise for thromboembolic events in patients with cancer.[■]

Primary bone marrow disorders pose a formidable therapeutic challenge. Aplastic anemia (Chapter 165) can be cured by both immunosuppressive therapy and stem cell transplantation. Long-lasting remissions can be achieved in an increasing fraction of patients with acute leukemias by chemotherapy, often coupled with stem cell transplantation (Chapter 178). Other primary bone marrow disorders that are unresponsive to these interventions are treated with supportive measures such as transfusions of red cells and platelets.

Red Cell Transfusion

The decision whether to transfuse an anemic patient is often challenging. The risks and complications of the administration of blood products are discussed in Chapter 177. Patients with chronic or long-standing anemia are able to compensate in several ways, discussed earlier in this chapter. A considerable reduction in red cell mass can be surprisingly well tolerated, especially if the patient is young or sedentary. Transfusion is seldom indicated in a patient with chronic anemia whose hemoglobin is 9 g/dL or greater. Those who are expected to respond to the administration of a specific agent such as iron, folic acid, or vitamin B_{12} can usually be spared transfusions.

Current evidence does not support the benefit of liberal transfusions in patients with asymptomatic anemia and heart disease. Indeed the American College of Physicians Guidelines on Treatment for Anemia in Patients with Heart Disease warns against both red cell transfusion and the use of ESAs in patients with cardiovascular disease who have mild to moderate anemia.[7] However, if the anemia is severe and accompanied by myocardial or cerebral ischemia or by congestive heart failure, prompt but slow administration of packed red cells is indicated. Whole blood should be given only if the patient is hypovolemic.

Splenectomy

Splenectomy is indicated in the treatment of certain hemolytic anemias. The efficacy of splenectomy correlates with the degree to which the abnormal or defective red cells are destroyed or sequestered in the spleen. Splenectomy is curative in nearly all patients with hereditary spherocytosis (Chapter 161). The operation may be also beneficial in selected patients with immunohemolytic anemia, congestive splenomegaly, spur cell anemia, and certain hemoglobinopathies and enzymopathies. The operative morbidity and mortality from elective splenectomy are very low. The procedure can often be done by laparoscopy. Occasional patients develop postoperative left subphrenic abscess. Following splenectomy, young children are at risk for developing overwhelming septicemia. This complication can be partially prevented by vaccination against pneumococcus and meningococcus. Post-splenectomy sepsis occurs rarely in adults. The risk for sepsis can be circumvented by partial splenectomy. Thrombocytosis generally develops promptly following splenectomy. However, in most cases it is transient. In patients with continued hemolysis or with a myeloproliferative disorder (Chapter 166), the thrombocytosis usually persists and may occasionally be associated with thromboembolic complications.

Grade A References

A1. Vinhas J, Barreto C, Assuncao J, et al. Treatment of anaemia with erythropoiesis-stimulating agents in patients with chronic kidney disease does not lower mortality and may increase cardiovascular risk: a meta-analysis. *Nephron Clin Pract*. 2012;121:c95-c101.

A2. Palmer SC, Navaneethan SD, Craig JC, et al. Meta-analysis: erythropoiesis-stimulating agents in patients with chronic kidney disease. *Ann Intern Med*. 2010;153:23-33.

A3. Tonia T, Mettler A, Robert N, et al. Erythropoietin or darbepoetin for patients with cancer. *Cochrane Database Syst Rev*. 2012;12:CD003407.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Balarajan Y, Ramakrishnan U, Ozaltin E, et al. Anaemia in low-income and middle-income countries. *Lancet*. 2011;378:2123-2135.
2. Bunn HF, Aster JA. *Pathophysiology of Blood Disorders*. New York: McGraw Hill; 2010.
3. Skold A, Cosco DL, Klein R. Methemoglobinemia: pathogenesis, diagnosis, and management. *South Med J*. 2011;104:757-761.
4. Chowdhary S, Bukoye B, Bhansali AM, et al. Risk of topical anesthetic-induced methemoglobinemia: a 10-year retrospective case-control study. *JAMA Intern Med*. 2013;173:771-776.
5. DeLoughery TG. Microcytic anemia. *N Engl J Med*. 2014;371:1324-1331.
6. Price EA, Mehra R, Holmes TH, et al. Anemia in older persons: etiology and evaluation. *Blood Cells Mol Dis*. 2011;46:159-165.
7. Qaseem A, Humphrey LL, Fitterman N, et al. Treatment of anemia in patients with heart disease: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. 2013;159:770-779.

REVIEW QUESTIONS

1. Which of the following laboratory tests would be *least* informative for establishing the presence of ineffective erythropoiesis in a patient with anemia?

A. Serum lactate dehydrogenase (LDH)
B. Serum bilirubin
C. Reticulocyte count
D. Serum erythropoietin level
E. Bone marrow examination

Answer: D In patients with ineffective erythropoiesis, there is erythroid hyperplasia in the bone marrow. Therefore, the bone marrow examination (answer E) is informative. Moreover, these patients have markedly enhanced destruction of these erythroid precursors and, as a direct consequence, elevations of serum (nonconjugated) bilirubin (answer B) and LDH (answer A). Unlike patients with hemolytic anemia, the reticulocyte count (answer C) is low in ineffective erythropoiesis. In contrast to these informative tests, serum erythropoietin (answer D) is not helpful because it will be elevated in nearly all patients with anemia irrespective of pathogenesis.

2. What physiologic mechanism determines the regulation of erythropoietin levels in the plasma?

A. Sensing of arterial oxygen tension in the carotid body
B. Sensing of intracellular oxygen tension in the carotid body
C. Sensing of arterial oxygen tension in the kidney
D. Sensing of intracellular oxygen tension in the kidney
E. Sensing of blood viscosity in the kidney

Answer: C The kidney is the primary site of erythropoietin production in humans and other mammals. The transcriptional regulation of the erythropoietin gene depends on the sensing of oxygen tension within a subset of renal interstitial cells at the boundary of the cortex and medulla. Note that sensing of arterial oxygen tension in the kidney (answer D) would not benefit the patient with anemia because arterial Po_2 would likely be normal. The sensing of oxygen tension in the carotid body (answers A and B) regulates the rate of respiration. Regulation of the red cell mass by sensing blood viscosity would be maladaptive (answer E).

3. Which of the following causes of anemia is best explained by inadequate production of erythropoietin?

A. Uremia
B. Chronic liver disease
C. Iron deficiency
D. Renal cell carcinoma
E. Aplastic anemia

Answer: A In most patients with chronic renal failure (answer A), there is damage to cells at the medullary-cortical boundary that produce erythropoietin. Therefore, serum erythropoietin levels in these patients are inappropriately low. The liver (answer B) produces much less erythropoietin than the kidney, and therefore liver disease does not significantly affect erythropoietin production. Renal cell carcinoma (answer D) sometimes causes elevated levels of serum erythropoietin but never decreased levels. Serum erythropoietin is markedly elevated in patients with aplastic anemia.

4. Which of the following best characterizes the impact of hepcidin on iron homeostasis?

A. Decreased release of iron from the duodenal enterocyte and decreased release of iron from the macrophage
B. Increased release of iron from the duodenal enterocyte and decreased release of iron from the macrophage
C. Decreased release of iron from the duodenal enterocyte and increased uptake of iron into the bone marrow
D. Increased release of iron from the duodenal enterocyte and decreased uptake of iron into the bone marrow
E. Increased release of iron from the duodenal enterocyte and increased uptake of iron into the bone marrow

Answer: A Hepcidin binds to and inactivates ferroportin, the transmembrane protein responsible for the export of iron from mucosal epithelial cells in the duodenum as well as macrophages in the bone marrow and liver. As a result, there is impairment of iron absorption from the gut and release of iron from macrophages.

5. What is the most effective way a patient compensates for anemia due to impaired red cell production?

A. Increased heart rate and stroke volume
B. Increased red cell oxygen affinity
C. Decreased red cell oxygen affinity
D. Decreased peripheral vascular resistance
E. Increased production of erythropoietin

Answer: C By far the most effective way in which the anemic patient compensates for a reduction in red cell mass and oxygen carrying capacity is by elevation of red cell 2,3-DPG, which shifts the oxygen binding curve to the right and lowers oxygen affinity, thereby enhancing oxygen unloading to tissues. Erythropoietin production (answer E) is markedly enhanced in nearly all patients with moderate or severe anemia (except those with chronic renal failure), but the high levels of plasma erythropoietin are ineffective in boosting hemoglobin levels in those with impairment of red cell production. Increased resting heart rate and stroke volume (answer A) occur only in patients with severe anemia, and neither these cardiac changes nor decreased peripheral vascular resistance (answer D) are efficient modes of compensation.

159

MICROCYTIC AND HYPOCHROMIC ANEMIAS

GORDON D. GINDER

The oxygen-carrying hemoglobin molecule executes the principal function of the mature erythrocyte. The hemoglobin content of erythrocytes is determined by the coordinated production of globin protein, the heme porphyrin ring, and the availability of iron. A deficiency in any of these three critical components of hemoglobin results in hypochromic and/or microcytic anemia (see Fig. 158-5).[1] Microcytic anemia is typically reported initially by automated red blood cell (RBC) indices. Hypochromic microcytic anemia can be confirmed on the peripheral blood smear (Fig. 159-1). Disorders of globin protein production typically produce microcytosis but not hypochromia and are discussed elsewhere (Chapter 162).

IRON DEFICIENCY ANEMIA

EPIDEMIOLOGY

Iron deficiency is by far the most common cause of anemia worldwide and is among the most frequently encountered medical problems seen by primary care physicians in the United States. More than 1.6 billion people, almost one fourth of the world's population, are anemic, and iron deficiency accounts for about half of the world's anemia burden.[2] It is estimated that between 2 and 5% of adolescent girls in the United States have iron deficiency sufficient to cause anemia. Elsewhere, the prevalence of iron deficiency–induced anemia is much higher, with estimates that up to 10% of the world population, or more than 500 million people, are affected. The prevalence rates are especially high in developing countries where dietary insufficiency and intestinal parasites are prevalent.

PATHOBIOLOGY

Most of the approximately 4 g of iron in the adult human body is incorporated into hemoglobin (approximately 2100 mg) in erythrocytes or myoglobin (approximately 300 mg) in muscle. The remainder is chiefly present as storage iron in the liver (1000 mg) and in the reticuloendothelial macrophages of the bone marrow and spleen (600 mg) (Fig. 159-2). Only a small

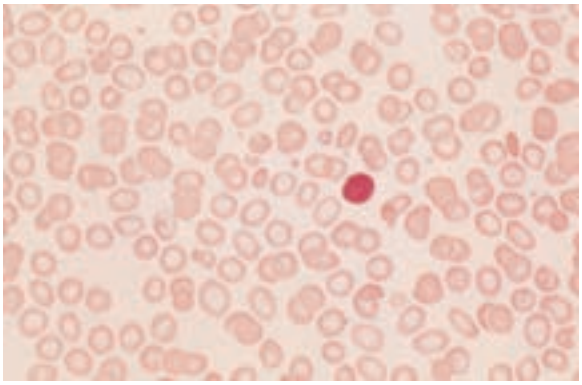

FIGURE 159-1. Iron deficiency anemia. Many of these red blood cells are microcytic (smaller than the nucleus of the normal lymphocyte near the center of the field) and hypochromic (with central areas of pallor that exceed half the diameter of the cells).

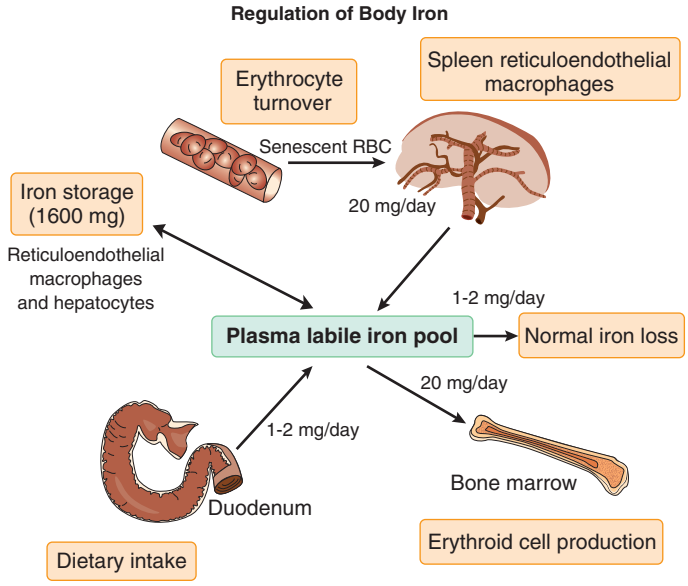

FIGURE 159-2. Iron homeostasis in normal humans. RBC = red blood cell.

amount of iron (3 to 7 mg) is freely circulating in plasma bound to transferrin, but this pool is kinetically very active, turning over every 3 to 4 hours. Because of the potent conservation mechanisms of iron recycling through the reticuloendothelial macrophage system, only an average of about 1 to 2 mg of iron is normally lost per day, largely through mucosal sloughing, desquamation, and, in females of reproductive age, menstruation.

The ability of iron to donate or accept electrons readily through conversion between the ferrous (Fe^{2+}) and ferric (Fe^{3+}) states makes it a critical component of the hemoglobin and myoglobin porphyrin rings that transport oxygen as well as cytochromes and various other vital enzymes. Free iron is extremely toxic owing to its capacity to catalyze the formation of free radicals, which lead to cellular damage. Therefore, the majority of body iron that is not stably incorporated into porphyrin rings is associated with proteins. Transferrin is the major protein associated with circulating plasma iron, and ferritin is the major protein associated with stored intracellular iron both in the cytoplasm and in mitochondria. Because the normal rate of iron loss is low, only about 1 to 2 mg/day of dietary iron is needed to maintain homeostasis. The average Western diet contains about 20 mg/day, and the efficiency of iron absorption in the duodenum is usually sufficient to maintain the amount of iron required for homeostatic balance.

Control of iron absorption by the duodenal crypt cells is critical because of the lack of any regulated physiologic mechanism for iron excretion. As a result, excessive iron intake can lead to deleterious iron overload, with concomitant organ damage (Chapter 212). Nonheme dietary iron is dissolved in part by the low pH of the stomach effluent. After reduction to the ferrous state by ferroreductase, iron is transferred across the apical crypt cell membrane by divalent metal transporter-1 (DMT-1). Several levels of regulation are involved in iron absorption by the intestine. One of these is modulated by dietary intake; thus, after a large influx of dietary iron, the duodenal absorptive capacity is diminished. A second regulatory mechanism modulates the iron absorption capacity based on total body iron stores. Finally, the so-called erythropoietic regulator modulates the capacity for enterocyte absorption based on the iron needed for erythropoiesis. A paradoxical increase in iron absorption through this mechanism occurs in certain types of anemia characterized by predominantly intramedullary destruction of erythroid cells (causing ineffective erythropoiesis): sideroblastic anemia, the thalassemias (Chapter 162), and congenital dyserythropoietic anemias.

Once inside the intestinal absorbing cell, iron is stored in complex with ferritin. Circulating plasma iron is complexed with the iron transport protein transferrin. The transferrin-iron complex is then taken up by erythroid precursors through the transferrin receptor. The high density of transferrin receptors on erythroid precursors ensures the preferential uptake of iron by these cells and explains why erythropoiesis proceeds normally until a critical deficiency of transferrin-bound iron, reflecting a depletion of total body iron, is present. The levels of transferrin, transferrin receptor, ferritin, and other proteins important in iron metabolism are regulated by the iron regulatory proteins IRP-1 and IRP-2.

Hepcidin, a 25-amino acid peptide, is the central regulator of iron homeostasis through its effects on intestinal absorption, macrophage recycling of iron from senescent RBCs, and iron mobilization from hepatic stores.[3] Thus, hepcidin affects all major sites of iron uptake and storage. Hepcidin is produced in the liver and acts as a negative regulator of iron absorption by the intestine and of iron release from storage in the macrophages and hepatocytes (see Fig. 158-7). It is believed that hepcidin binds to ferroportin, the major iron transporter in the membranes of the enterocyte, macrophage, and hepatocyte, causing the internalization and degradation of ferroportin. This process blocks the transport of iron across the membrane of the basolateral crypt cell, preventing its incorporation into transferrin-bound plasma iron. Likewise, loss of ferroportin function blocks the major export pathway of iron stores from macrophages and hepatocytes. Hepcidin production is upregulated by iron and downregulated by hypoxia, consistent with its homeostatic role. Because hepcidin is also upregulated by inflammatory cytokines, it is believed to play an important role in the paradoxical lack of transferrin-bound iron available for erythropoiesis in the face of adequate or even excess iron stores found in the anemia of chronic inflammation (see later and Chapter 158).

Blood Loss

Iron deficiency anemia results from an imbalance between available body iron for hemoglobin production and the minimal amount needed to sustain normal hemoglobin production during erythropoiesis (see Fig. 159-2). Because of the combined effectiveness of dietary absorption and retention of iron under normal circumstances, this mismatch is most often due to blood loss, with the gastrointestinal (GI) tract being the most common site (Chapter 135) in men and nonmenstruating women. In developed countries, the blood loss is usually secondary to benign or neoplastic lesions of the GI tract (Chapters 192 and 193) or chronic ingestion of drugs that cause GI mucosal damage (Chapter 139). The most common offending agents are alcohol and salicylates or other nonsteroidal anti-inflammatory agents. In developing countries, helminthic infections, including hookworm (Chapter 357) and schistosomiasis (Chapter 355), are among the most common causes of GI blood loss.

Genitourinary tract blood loss resulting in iron deficiency is most common in menstruating women. Less common are urinary tract malignancies (Chapter 197) and hemoglobinuria due to intravascular hemolysis (Chapter 160). Respiratory tract blood loss is far less common as a cause of iron deficiency.

Reproduction and Growth

In most cases, an increased iron requirement is due to blood loss; other causes include rapid growth in infancy and adolescence and pregnancy and lactation in adulthood. It is estimated that failure to satisfy the increased iron requirements during pregnancy with supplemental iron may result in a deficiency equivalent to a cumulative blood loss of up to 1500 mL.

Inadequate Iron Intake

The other major cause of iron deficiency is inadequate iron intake. Only diets that lack 1 to 2 mg/day fail to provide adequate iron. The average Western meal contains about 6 mg of iron, so dietary insufficiency is not a common cause of iron deficiency. Certain diets that lack iron or contain large quantities of phytates from cereals or tannate from tea, both of which inhibit intestinal iron absorption, may result in iron deficiency. Although iron is usually readily absorbed, primarily in the duodenum, pathologic states that can impair the process include generalized intestinal malabsorption (Chapter 140), atrophic gastritis (Chapter 139) with achlorhydria, and extensive gastric surgery. In contrast, chronic use of histamine-2 (H_2) receptor blockers or proton pump inhibitors does not appear to cause iron deficiency. In the United States, celiac disease (Chapter 140) is an increasingly common cause of iron deficiency, with resulting anemia.

CLINICAL MANIFESTATIONS

Because of compensatory physiologic mechanisms, patients with mild iron deficiency anemia may be asymptomatic. Iron deficiency in these patients may be recognized during the evaluation of an underlying disease process or as part of routine laboratory studies. The findings of microcytosis and hypochromia occur only after the hematocrit has fallen to approximately 30%, so neither finding may be present in early stages.

The anemia of iron deficiency, like other anemias, manifests with nonspecific symptoms such as weakness, pallor, dizziness, decreased exercise tolerance, or irritability. Because iron is a critical component of the porphyrin complex in muscle as well as many essential metabolic enzymes, its deficiency affects other organ systems besides the erythron, often resulting in a degree of fatigue, exercise intolerance, and weakness out of proportion to the hemoglobin level. Repletion of iron in iron-deficient individuals may improve cognitive and exercise performance. In addition, intravenous iron treatment in patients with heart failure and coexisting iron deficiency (to which heart failure patients are prone) can improve symptoms, functional capacity, and quality of life, irrespective of the presence or absence of anemia.

Rare patients, most frequently elderly women, may have dysphagia due to an esophageal stricture or web (Plummer-Vinson syndrome). A clinical manifestation unique to iron deficiency is pica, which is an unusual craving for certain non-nutritional substances. Pica may manifest as a craving for ice (pagophagia) or, less commonly, for clay (geophagia) or starch (amylophagia); pagophagia is believed to be the most specific for iron deficiency.

Physical findings that may be associated with the iron-deficient state include glossitis and angular stomatitis. Other less common but highly specific findings are spooning of the fingernails (koilonychia) and blue-tinged sclerae.

DIAGNOSIS

The diagnosis of iron deficiency anemia is made by laboratory testing. Because microcytic hypochromic RBCs are a sine qua non of this type of anemia, initial screening consists of a determination of hemoglobin levels, mean corpuscular volume, erythrocyte hemoglobin content, and reticulocyte

TABLE 159-1 LABORATORY FINDINGS FOR IRON STUDIES IN MICROCYTIC AND HYPOCHROMIC ANEMIAS

ANEMIA	SERUM IRON	TIBC	TRANSFERRIN SATURATION (%)	SERUM FERRITIN	SERUM TRANSFERRIN RECEPTOR	MARROW RE IRON	MARROW RINGED SIDEROBLASTS
Iron deficiency anemia	Low	High	0-15	Low (<30 μg/L)	High	Absent	Absent
Anemia of chronic disease	Low	Normal or low	5-15	Normal or high	Normal	Normal or high	Absent
Sideroblastic anemia	High	Normal	60-90	High	Normal or high	High	Present

RE = reticuloendothelial; TIBC = total iron-binding capacity.

count. In experienced hands, the peripheral blood smear (Chapter 157) is an excellent indicator of iron deficiency anemia. In iron deficiency anemia, most erythrocytes are smaller in diameter than the nucleus of a typical lymphocyte, and the area of central pallor is greater than 50% of the total diameter of the erythrocyte (see Fig. 159-1). Variability of erythrocyte size distinguishes iron deficiency anemia from other conditions that give rise to microcytosis; the calculated variability in RBC volume (the so-called RBC volume distribution width, or RDW) is elevated early in iron deficiency anemia.

The definitive diagnosis of iron deficiency anemia is made by tests that measure total body iron stores: the absence of iron stores that can be mobilized is unique to this microcytic hypochromic anemia. Transferrin and transferrin-bound iron levels may not be reliable indicators of iron deficiency because they are also abnormal in the anemia of chronic disease, despite adequate total body iron stores (see Fig. 158-6).

The serum ferritin level is the most reliable, noninvasive, and cost-effective indicator that is routinely available in most clinical laboratories (Table 159-1). In a large study of 259 anemic patients, a serum ferritin level less than 18 μg/L was diagnostic of iron deficiency with greater than 95% specificity and a 55% sensitivity. At a serum ferritin level of 45 μg/L, the sensitivity rose to approximately 70%, and a level higher than 100 μg/L in populations with a less than 40% prevalence of iron deficiency excluded a diagnosis of iron deficiency with more than 90% sensitivity. Although some recent studies have questioned the accuracy of the routine determination of stainable marrow iron, this test is still generally considered to be the "gold standard" for tests of iron deficiency. However, determination of total bone marrow iron stores is rarely necessary to diagnose iron deficiency anemia, except when there is some other complicating process.

One setting in which serum ferritin levels can be spuriously elevated is chronic inflammation or chronic disease. A meta-analysis involving 8796 subjects suggested that measurement of C-reactive protein (CRP) and α_1-acid glycoprotein (ACP) to assess for the level of inflammation can improve the accuracy of ferritin level as an indicator of iron deficiency.[4] The soluble serum transferrin receptor (sTfR) level is an excellent measure of total erythroid precursor mass. The sTfR is aberrantly elevated in the presence of iron deficiency, so it is considered a useful test for this condition. A number of studies have demonstrated the utility of the sTfR/ferritin ratio in distinguishing iron deficiency from the anemia of chronic disease. However, the lack of reliable standards and a meta-analysis of efficacy of the test indicating that additional data are needed to define the diagnostic accuracy of the sTfR[5] have prevented this assay from becoming routinely available in clinical practice.

TREATMENT

Rx

The treatment of iron deficiency anemia is replenishment of body iron stores. However, the underlying cause should always be investigated before treatment is begun because in many cases it is a correctable and potentially fatal GI lesion (Chapter 135).

Oral Administration

The preferred route of iron administration is oral. Oral iron is most readily absorbed in the absence of food, especially in the setting of decreased stomach acid production owing to atrophic gastritis, gastric surgery, or chronic suppression of gastric acid with an H_2 antagonist or proton pump inhibitor. The major obstacle to oral iron replacement is unacceptable side effects, chiefly epigastric discomfort or nausea; diarrhea or constipation also occurs in some patients. Reducing the dose often eliminates nausea and epigastric discomfort. Despite the development of a number of orally effective iron-containing compounds, the original salt, ferrous sulfate (325 mg three times daily), remains the most useful. Although some newer oral iron preparations, such as ferrous gluconate (300 mg two or three times daily) or ferrous fumarate (325 mg two or three times daily), may induce fewer GI side effects per milligram of iron, they are also less well absorbed, so there is no net advantage to these costlier formulations except for patients who cannot tolerate ferrous sulfate. Given both the low toxicity and the low cost of oral iron replacement, a therapeutic trial is an alternative means of confirming a diagnosis of iron deficiency anemia.

Parenteral Administration

In situations in which primary blood loss is uncontrollable, iron cannot be absorbed owing to severe malabsorption, or oral iron is not tolerated despite concerted efforts to minimize side effects, parenteral iron is an effective alternative treatment. Intramuscular dosing is limited to 100 mg/injection, so intravenous administration is recommended. Sodium ferric gluconate (given intravenously at a dose of 125 mg over 10 minutes) is the preferred form of parenteral iron owing to the low incidence of adverse reactions. A multiinstitutional, double-blind, randomized, placebo-controlled trial of more than 2500 patients showed similar adverse events in patients receiving sodium ferric gluconate versus those receiving placebo, and only one life-threatening complication occurred; in comparison, there were 23 such events among 3768 patients treated with iron dextran in a historical control arm. One limitation of sodium ferric gluconate is that the maximum dose deliverable in a single injection is approximately 125 mg, and a total dose of 500 to 2000 mg is usually required for adequate repletion. Although large doses of iron dextran can be delivered in a single intravenous injection, this is currently reserved for situations in which rapid iron replacement is required because of the life-threatening anaphylactic and delayed adverse reactions that occur in 0.6 and 2.5% of cases, respectively. If iron dextran is to be given intravenously, premedication with diphenhydramine and a slow test-dose injection of 30 to 40 mg diluted in normal saline are recommended.

Newer iron preparations for intravenous use include ferric carboxymaltose, iron isomaltoside 1000, and ferumoxytol. These preparations allow the administration of much higher doses of intravenous iron in a shorter time and can be used safely to replenish iron stores, sometimes even in a single treatment session.

The response to iron repletion therapy is usually quite rapid, with elimination of symptoms within a few days. Increased reticulocytosis usually begins within 4 to 5 days, and the hemoglobin level often rises within 1 week and reaches a normal level after 6 weeks of therapy if adequate iron replacement is achieved. The goal of therapy, which is to reach a serum ferritin level of greater than 50 mg/L, usually takes 4 to 6 months. Therapy must be continued after adequate replacement is achieved if the underlying cause of iron deficiency is not reversible. Because of the avidity of transferrin receptor–rich erythroid precursors for transferrin-bound iron, the serum ferritin level usually does not rise until hemoglobin levels reach normal.

Failure to Respond to Iron Therapy

An incomplete or lack of response to oral iron therapy, as determined by failure to normalize the hemoglobin level, usually means either that iron replacement has not been adequate (most commonly due to noncompliance with oral iron because of its side effects) or that iron deficiency is not the primary cause of the anemia (e.g., coexisting anemia of chronic disease). Less common causes of failure to respond to oral iron include iron malabsorption (e.g., celiac disease, atrophic gastritis) or blood loss in excess of iron replacement. Refractoriness to oral iron due to celiac disease can be screened by testing for anti-tissue transglutaminase (TTG) antibodies; autoimmune atrophic gastritis by serum gastrin, parietal cell, or intrinsic factor antibodies; and *Helicobacter pylori* infection and gastritis by antibody screening or fecal antigen and urease breath test.

TMPRSS6, also called matriptase-2, is a type II transmembrane serine protease that suppresses hepcidin production. Several types of mutations in *TMPRSS6* that are either sporadic or familial (usually autosomal recessive) have recently been described as causes of iron-refractory iron deficiency anemia (IRIDA). Associated with inappropriately increased levels of urinary hepcidin, the defect causes impaired iron absorption and recycling, leading to IRIDA. Characteristically, these patients exhibit no hematologic improvement in response to oral iron intake and are only partially responsive to parenteral iron because of abnormal iron utilization.[6,7]

PROGNOSIS

In most cases, iron deficiency anemia can be corrected rapidly by either oral or parenteral replacement, but the long-term prognosis ultimately depends on the clinical course of the underlying cause. It is critical that the patient undergo a full evaluation to determine the underlying cause of the iron deficiency, especially because an occult gastrointestinal lesion, often malignant, may be present, particularly in patients older than 50 years (Chapters 192 and 193).

ANEMIA OF CHRONIC DISEASE AND INFLAMMATION

DEFINITION

Anemia of chronic disease (or *anemia of chronic inflammation*) refers to anemia that occurs in the setting of a chronic disease state, usually one associated with elevated levels of inflammatory cytokines. Although anemia of chronic inflammation usually manifests as a normochromic normocytic process (Chapter 158), between 20 and 50% of cases are associated with microcytic RBC indices. The anemia is usually mild to moderate, and it may not be symptomatic.

EPIDEMIOLOGY

Anemia of chronic inflammation is believed to be the second most common cause of anemia, after iron deficiency. It is the most common type of anemia encountered among hospitalized patients. The wide spectrum of underlying diseases includes acute and chronic infections, inflammatory and autoimmune diseases, cancers, and chronic kidney diseases.

PATHOBIOLOGY

There are three major mechanisms of anemia of chronic inflammation, and all are believed to result from the effects of abnormal levels of inflammatory cytokines. The first is dysregulated iron homeostasis, manifested by low serum iron (hypoferremia) in the presence of normal or elevated serum ferritin levels and abundant reticuloendothelial macrophage iron stores. The functional consequence is a limited availability of iron for erythroid progenitor cells and resultant restriction of erythropoiesis. Pro-inflammatory stimuli, including lipopolysaccharides, interferon-γ, and tumor necrosis factor-α (TNF-α), upregulate DMT-1, which increases iron uptake by the reticuloendothelial cells. At the same time, these stimuli cause the downregulation of ferroportin expression; ferroportin is the protein required for the release of ferrous iron from storage cells and for the transport of dietary iron from duodenal enterocytes into the circulation.

Because hepcidin is an iron-regulated, acute phase reactant peptide that blocks both iron uptake in the gut and iron release from hepatocytes and macrophage stores, its upregulation by lipopolysaccharides, interleukin (IL)-6, and possibly IL-1 (indirectly, through induction of IL-6) results in another mechanism of anemia. Also, patients with hepatic adenomas that secrete high levels of hepcidin have iron-refractory anemia in the presence of normal or elevated ferritin and macrophage iron stores, despite the absence of elevated inflammatory cytokine levels. Elevated urinary hepcidin concentrations correlate with ferritin levels in patients with anemia of inflammation, iron overload, and iron deficiency.[8] These relationships are depicted in Figure 158-7.

A third pathophysiologic feature of anemia of chronic inflammation is the inhibition of erythroid progenitor expansion. Interferon-γ is the most potent inhibitory factor of erythropoiesis, but similar inhibition is believed to be mediated by IL-1, TNF-α, and interferon-β. These mediators of inflammation act to increase erythroid progenitor apoptosis, downregulate erythropoietin receptors, and antagonize pro-hematopoietic factors. The action of erythropoietin appears to be directly antagonized by these pro-inflammatory cytokines, which would explain why responsiveness to erythropoietin seems to be inversely related to the severity of the underlying chronic inflammation and the levels of interferon-γ and TNF-α. Finally, increased erythrophagocytosis in the presence of inflammation results in a modest shortening of RBC half-life.

CLINICAL MANIFESTATIONS

The clinical manifestations in patients with anemia of chronic inflammation are usually dominated by the underlying disease process. The anemia in this condition is usually mild, with hemoglobin levels in the range of 8 to 10 g/dL. However, supervening blood loss, absolute iron deficiency, or other aggravating factors can produce life-threatening anemia. Even mild to moderate anemia contributes to the debilitating effects of the underlying disease, adversely affecting performance status and quality of life. Moreover, the presence of anemia is associated with a poorer overall prognosis in many of the underlying chronic diseases, although correction of anemia has not been directly demonstrated to improve survival.

DIAGNOSIS

The clinical diagnosis of anemia of chronic disease presenting with microcytic hypochromic RBC indices is one of exclusion, based on low serum iron in the presence of normal or increased total body iron stores (see Table 159-1). Serum ferritin is the best single laboratory marker for assessing iron storage, and it is almost invariably normal or elevated in anemia of chronic disease. If both the serum iron and the transferrin saturation are reduced, reflecting dysregulation of iron homeostasis, the diagnosis of anemia of chronic disease can be made in the appropriate clinical setting after the exclusion of other causes of anemia, such as coexistent blood loss, thalassemia (Chapter 162), and drug-induced suppression of erythropoiesis. In the presence of inflammation, however, up to 30% of patients with true iron deficiency have serum ferritin levels greater than 100 μg/L, potentially obscuring the diagnosis of iron deficiency. Assays for sTfR are useful to diagnose iron deficiency in the presence of the inflammation associated with anemia of chronic disease, but problems with standardization have limited this test's availability in clinical practice. Examination of the bone marrow for reticuloendothelial macrophage iron stores (hemosiderin) and erythroblasts containing iron granules (sideroblasts) can provide definitive evidence of absent iron stores in the setting of anemia of chronic inflammation. A low serum erythropoietin level is also useful in supporting a diagnosis of anemia of chronic inflammation, but only when the hemoglobin level is less than 10 g/dL.

TREATMENT

Rx

Treatment of the Underlying Disease

The most effective treatment for anemia of chronic disease is successful treatment of the underlying inflammatory disease process, whether it is an acute or chronic infection, treatable cancer, renal failure, or rheumatoid arthritis. Even if definitive treatment is not possible, quality of life and perhaps prognosis can improve if symptomatic anemia is treated directly. Unfortunately, anemia of chronic inflammation remains undertreated, even in developed countries.

Blood Transfusion

Blood transfusion (Chapter 177) offers the immediate resolution of anemia, but it is indicated chiefly when the anemia is life threatening or seriously limits the patient's functioning. These situations almost always involve supervening blood loss or some other acute process that compounds the anemia of chronic disease. Transfusion is not recommended for the long-term treatment of mild or moderate anemia of chronic inflammation because of the secondary risks, which include transfusional iron overload, human leukocyte antigen (HLA) sensitization in the case of potential renal transplantation, and other side effects of transfusion.

Intravenous Iron and Erythropoietin Therapy

If iron replacement is needed for anemia of chronic inflammation, parenteral iron administration is usually required to replenish stores because of the block in intestinal absorption (see Parenteral Administration under Iron Deficiency Anemia). In hemodialysis patients receiving erythropoietin therapy, intravenous iron therapy improves anemia and increases both ferritin and transferrin saturation levels more than oral iron replacement; however, when intravenous iron replacement raises the transferrin saturation to greater than 20%, there appears to be an increased risk for developing bacteremia, underscoring the complex relationship between iron homeostasis and immunity.

Erythropoietin therapy is currently approved for use in patients with chronic kidney disease[9] or HIV infection, and in cancer patients who are undergoing myelosuppressive treatment. In patients with chronic kidney disease, the hemoglobin concentration goal should be 10-12 g/dL in chronic kidney disease[A2] (Chapter 130) and >9 g/dL but not >11 g/dL in end-stage renal disease patients on dialysis[A2] (Chapter 131).

Patients with demonstrated iron deficiency should receive supplemental iron with intravenous iron gluconate (see the earlier discussion) while being treated with erythropoietin. There is increasing evidence that addition of iron in this setting reduces the need for higher dose erythropoietin with its attendant risks.

In patients undergoing chemotherapy for cancer whose hemoglobin is less than 10 g/dL, erythropoietin therapy improves quality of life and performance status[A] but increases the risk of thromboembolic complications and death.[A] Current recommendations are for cautious use during chemotherapy and against routine use in inpatients not receiving chemotherapy.

PROGNOSIS

The overall prognosis of anemia of chronic inflammation is determined almost exclusively by the course of the underlying disease. It is well established that the degree of anemia correlates well with the severity of the underlying disease process and therefore with levels of inflammatory cytokines. In the absence of a supervening process, anemia of chronic inflammation is not life threatening, and treatment of the anemia per se has not been proved to affect overall survival.

SIDEROBLASTIC ANEMIAS

DEFINITION

This heterogeneous group of anemias is distinguished by the characteristic finding of excessive mitochondrial iron in erythroblasts, as manifested by iron-laden, ringed sideroblasts in the bone marrow in the presence of moderate to severe anemia. These disorders result from mitochondrial defects either in the biosynthesis of the heme porphyrin ring or in the metabolism of iron. Both hereditary and acquired types of sideroblastic anemia have been described. Although often characterized by microcytic and sometimes hypochromic anemia, these disorders can manifest with normochromic normocytic RBCs; if the anemia is associated with myelodysplasia (Chapter 182), macrocytic RBC indices may be present.

EPIDEMIOLOGY

Although acquired sideroblastic anemias are relatively rare, they are much more prevalent than hereditary forms. The true incidence of acquired sideroblastic anemia is not well established, in part owing to the heterogeneity of causes and clinical presentations. Hereditary X-linked sideroblastic anemias usually manifest in childhood or early adulthood.

PATHOBIOLOGY

Genetics

The pathophysiologic mechanisms of hereditary sideroblastic anemias are much better understood than those of the more common idiopathic, acquired variety associated with myelodysplasia.[10,11] Two main forms of X-linked hereditary sideroblastic anemia have been characterized, and both result from defects in the heme synthesis pathway (Fig. 159-3). The first type is caused by mutations in the gene coding for erythroid-specific δ-aminolevulinic acid synthase, known as *ALAS-2*, on the X chromosome. These mutations may affect the affinity of the enzyme for pyridoxal phosphate or its structural stability, catalytic site, or susceptibility to mitochondrial proteases. In those cases in which the affinity of ALAS-2 for pyridoxal phosphate is altered, pyridoxine supplementation can ameliorate the associated anemia. The other major group of X-linked sideroblastic anemias results from defects in the adenosine triphosphate binding cassette (ABC) protein known as hABC7. The hABC7 protein is believed to be involved in iron-sulfur [FeS] cluster formation. Because [FeS] cluster–associated proteins include ferrochelatase and the cytosolic IRP-1, defects in hABC7 are believed to result in defective iron metabolism or inadequate incorporation of iron into the heme porphyrin ring by ferrochelatase. This type of X-linked sideroblastic anemia is associated with ataxia.

In addition to the two X-linked causes, both autosomal dominant and recessive forms of hereditary sideroblastic anemia have been described. However, the exact mechanisms involved in these disorders are not known.

Other types of hereditary sideroblastic anemia are believed to result from mutations in the mitochondrial genome rather than in nuclear genes. The inheritance of these disorders is complex, owing to the exclusively maternal inheritance pattern of mitochondria; the ovum is the only source of embryonic mitochondria.

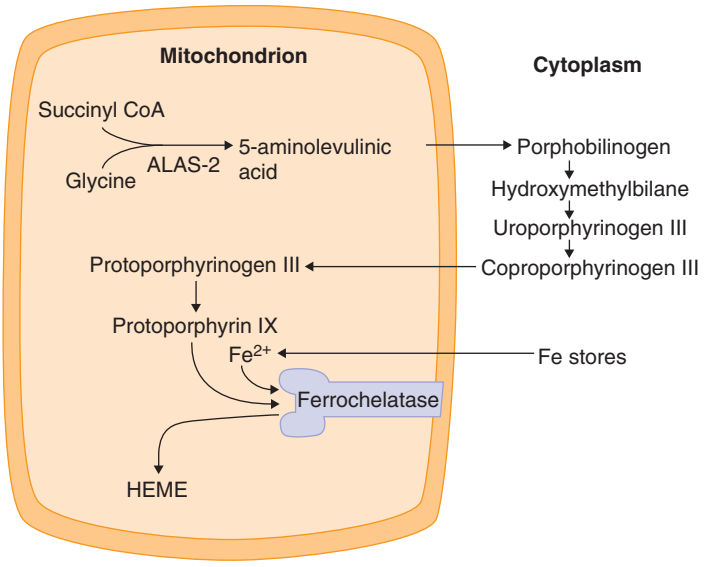

FIGURE 159-3. **The heme synthesis pathway.** ALAS-2 = δ-aminolevulinic acid synthase; CoA = coenzyme A.

Exposure to Drugs or Toxins

The most common form of acquired sideroblastic anemia results from nutritional deficiency or exposure to exogenous drugs or toxins, especially alcohol. Although sideroblastic anemia is not a common finding in alcoholism, the high incidence of alcohol abuse in Western cultures accounts for its frequency as a cause. Alcohol directly inhibits erythropoiesis, but sideroblastic anemia is usually seen only in the setting of concurrent alcoholism and nutritional deficiencies. Other well-documented drug exposures associated with sideroblastic anemia include isoniazid, chloramphenicol, and cycloserine. Sideroblastic anemia has also been attributed to lead exposure (Chapter 22), but there are limited primary data to support this association. Deficiency of pyridoxine causes sideroblastic anemia in animals and may also occur in the setting of alcoholism in humans, although ethanol is believed to be an antagonist of the interaction of pyridoxal phosphate with 5-aminolevulinic acid as a cofactor in the first step of heme biosynthesis. Copper deficiency, though rare, has also been associated with sideroblastic anemia, usually in the setting of an overdose of bivalent cation chelators such as penicillamine or trientine, used to treat the copper overloading found in Wilson disease (Chapter 211).

Idiopathic Forms

The major cause of acquired sideroblastic anemia is idiopathic, in association with myelodysplastic syndromes (Chapter 182). Refractory anemia with ringed sideroblasts is characterized by abnormalities in all three hematopoietic cell lineages, in addition to the presence of ringed sideroblasts.[12]

A second form, known as pure sideroblastic anemia, is less frequently associated with cytogenetic abnormalities, is characterized by dysplasia only in erythroid progenitors, and lacks cytopenias other than anemia. The prognosis in this type of acquired idiopathic sideroblastic anemia is much better than that in refractory anemia with ringed sideroblasts, in part because of a very low incidence (about 10%) of evolution to acute leukemia.

Because of the important differences in prognosis, it is imperative to evaluate cytogenetics and marrow morphology at the time of diagnosis. Recent evidence suggests that mitochondrial DNA mutations and attendant mitochondrial cytopathies account for many, if not all, cases.

CLINICAL MANIFESTATIONS

Because of the heterogeneous nature of the sideroblastic anemias, many of the clinical manifestations vary according to the underlying pathophysiologic cause. The anemia is usually moderate to severe, with hemoglobin levels in the range of 4 to 10 g/dL. The peripheral blood smear frequently reveals hypochromia, often with basophilic stippling. Microcytosis is often seen in hereditary forms, but normochromic, normocytic, or even macrocytic RBCs may be seen, especially in the setting of myelodysplasia or in a rare X-linked hereditary form known as Pearson syndrome.

DIAGNOSIS

The most useful diagnostic laboratory test for sideroblastic anemia is bone marrow morphology with Prussian blue iron staining, which reveals abnormally large and numerous bluish green siderosomes within at least 15% of erythroblasts, giving the characteristic appearance of ringed sideroblasts (Fig. 159-4). These ringed sideroblasts distinguish this disorder from iron deficiency anemia and anemia of chronic inflammation. Bone marrow findings in idiopathic acquired sideroblastic anemias include dyspoietic features of erythroid and/or myeloid and megakaryotic cell lineages.

Iron studies usually reveal normal iron stores or evidence of iron overload, which is caused by the ineffective erythropoiesis found in sideroblastic anemia as well as by the transfusion therapy often required for its treatment. Iron deficiency can occur coincident with sideroblastic anemia, complicating

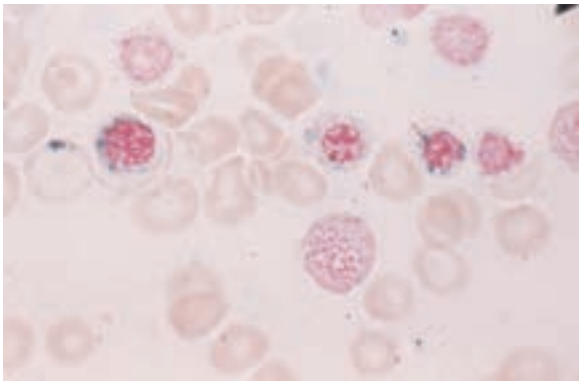

FIGURE 159-4. Sideroblastic anemia. Prussian blue iron stain of the bone marrow shows ringed sideroblasts, which are nucleated red blood cell precursors with perinuclear rings of iron-laden mitochondria.

the diagnosis owing to the lack of characteristic ringed sideroblasts, particularly in myelodysplastic syndromes in which thrombocytopenia leads to GI blood loss. If coexisting iron deficiency is suspected, a repeat bone marrow examination after iron repletion has failed to correct the anemia reveals the diagnostic ringed sideroblasts.

TREATMENT

Rx

Treatment of Underlying Disease

Most forms of sideroblastic anemia lack a specific therapy aimed at the underlying mechanism. Exceptions are those types caused by alcohol or drugs, for which removal of the offending agent usually results in resolution, or at least improvement, of the anemia. Abstinence from alcohol usually reverses the abnormalities in heme biosynthesis in 1 to 2 weeks, as evidenced by the disappearance of ringed sideroblasts in the marrow.

Pyridoxine markedly improves the relatively rare cases of nutritional deficiency, which are usually associated with alcoholism, and some forms of X-linked hereditary sideroblastic anemias in which the binding of pyridoxine by ALAS-2 is defective. Because of its low toxicity in moderate doses, a trial of pyridoxine, 100 to 200 mg/day orally for up to 3 months, is worthwhile in all patients. In responsive cases, reticulocytosis occurs within 2 to 3 weeks, and the hemoglobin level improves over several months. High-dose pyridoxine has been shown to overcome the defect in ALAS-2 activity in some patients with X-linked sideroblastic anemia, but prolonged high-dose therapy can be associated with peripheral neuropathy.

Transfusion

The mainstay of therapy for most severe sideroblastic anemias remains RBC transfusions. Because of the risks of long-term transfusion therapy, treatment should be aimed at achieving a normal performance status rather than a specific target hemoglobin level. Iron stores should be monitored regularly, and iron chelation therapy should be used in the setting of iron overload.

Erythropoietin

Therapy with erythropoietin, with or without granulocyte colony-stimulating factor (G-CSF), benefits a small percentage of patients with acquired sideroblastic anemia due to myelodysplasia. A meta-analysis of 17 studies in which 205 patients were treated with erythropoietin showed an overall response rate of only 16%. However, patients with a diagnosis other than refractory anemia with ringed sideroblasts who were not transfusion dependent had response rates greater than 50%, whereas none of the patients who had refractory anemia with ringed sideroblasts and a serum erythropoietin level greater than 200 U/L responded. Studies using a combination of erythropoietin and G-CSF showed somewhat higher response rates, although none of these studies was large or randomized. Allogeneic bone marrow transplantation (Chapter 178) benefits eligible patients whose myelodysplasia (Chapter 182) has a high risk for evolving into acute leukemia.

PROGNOSIS

As with the underlying pathophysiology, the prognosis in sideroblastic anemias is highly variable. Secondary acquired forms of the disease due to alcohol or toxins respond well to withdrawal of the offending agent, with rapid and often complete normalization of erythropoiesis. The pure sideroblastic anemia variant of myelodysplasia-associated sideroblastic anemia can usually be managed well for many years with transfusions and, if necessary, concordant iron chelation therapy. Other myelodysplasia-related sideroblastic anemias generally have a poor prognosis because of the frequent coexistence of pancytopenia and the relatively high incidence of progression to acute leukemia.

Grade A References

A1. Anker SD, Colet JC, Filippatos G, et al. Ferric carboxymaltose in patients with heart failure and iron deficiency. *N Engl J Med*. 2009;361:2436-2448.

A2. Pfeffer MA, Burdmann EA, Chen CY, et al. A trial of darbepoetin alfa in type 2 diabetes and chronic kidney disease. *N Engl J Med*. 2009;361:2019-2032.

A3. Drueke TB, Locatelli F, Clyne N, et al. Normalization of hemoglobin level in patients with chronic kidney disease and anemia. *N Engl J Med*. 2006;355:2071-2084.

A4. Singh AK, Szczech L, Tang KL, et al. Correction of anemia with epoetin alfa in chronic kidney disease. *N Engl J Med*. 2006;355:2085-2098.

A5. Ludwig H, Crawford J, Osterborg A, et al. Pooled analysis of individual patient-level data from all randomized, double-blind, placebo-controlled trials of darbepoetin alfa in the treatment of patients with chemotherapy-induced anemia. *J Clin Oncol*. 2009;27:2838-2847.

A6. Tonia T, Mettler A, Robert N, et al. Erythropoietin or darbepoetin for patients with cancer. *Cochrane Database Syst Rev*. 2012;12:CD003407.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. DeLoughery TG. Microcytic anemia. *N Engl J Med*. 2014;371:1324-1331.
2. Parischa S-R, Drakesmith H, Black J, et al. Control of iron deficiency anemia in low- and middle-income countries. *Blood*. 2013;121:2607-2617.
3. Ruchala P, Nemeth E. The pathophysiology and pharmacology of hepcidin. *Trends Pharmacol Sci*. 2014;35:155-161.
4. Thurnham DI, McCabe LD, Haldar S, et al. Adjusting plasma ferritin concentrations to remove the effects of subclinical inflammation in the assessment of iron deficiency: a meta-analysis. *Am J Clin Nutr*. 2010;92:546-555.
5. Infusino I, Braga F, Dolci A, et al. Soluble transferrin receptor (sTfR) and sTfR/log ferritin index for the diagnosis of iron-deficiency anemia: a meta-analysis. *Am J Clin Pathol*. 2012;138:642-649.
6. Hershko C, Camaschella C. How I treat unexplained refractory iron deficiency anemia. *Blood*. 2014;123:326-333.
7. De Falco L, Sanchez M, Silvestri L, et al. Iron refractory iron deficiency anemia. *Haematologica*. 2013;98:845-853.
8. Goodnough LT, Nemeth E, Ganz T. Detection, evaluation, and management of iron restricted erythropoiesis. *Blood*. 2010;116:4754-4761.
9. Koulouridis I, Alfayez M, Trikalinos TA, et al. Dose of erythropoiesis-stimulating agents and adverse outcomes in CKD: a metaregression analysis. *Am J Kidney Dis*. 2013;61:44-56.
10. Harigae H, Furuyama K. Hereditary sideroblastic anemia: pathophysiology and gene mutations. *Int J Hematol*. 2010;92:425-431.
11. Donker AE, Raymakers RA, Vlasveld LT, et al. Practice guidelines for the diagnosis and management of microcytic anemias due to genetic disorders of iron metabolism or heme synthesis. *Blood*. 2014;123:3873-3886.
12. Malcovati L, Cazzola M. Refractory anemia with ring sideroblasts. *Best Pract Res Clin Haematol*. 2013;26:377-385.

REVIEW QUESTIONS

1. A 60-year-old woman with a history of chronic arthritis that is currently under good control presents with symptoms of increased fatigue. She is noted to have a decreased hemoglobin of 8 g/dL and microcytic indices on her complete blood count. There is no evidence of blood loss by history or laboratory testing. The patient has a normal erythrocyte sedimentation rate (ESR) and a normal C-reactive protein (CRP) level. Which of the following routine laboratory tests is most likely to indicate whether this patient will respond to oral iron replacement with improvement in her hemoglobin level?

A. Transferrin saturation level
B. Serum ferritin level
C. Serum iron level
D. Vitamin B_{12} level
E. Reticulocyte count

Answer: B The ferritin level is generally the best clinically available indicator of available iron for erythropoiesis. Although inflammation associated with arthritis could falsely elevate the estimation of iron stores associated with ferritin, this patient's arthritis is not highly active, as shown by the normal ESR and CRP level. Elevated CRP and ESR are indicative of even subclinical inflammation. Because these are normal, a ferritin level in this individual would likely indicate whether there is iron deficiency. Moreover, there are criteria for correcting ferritin levels in the presence of low or moderate inflammation.

The use of soluble transferrin receptor (sTfR) assay and sTfR/ferritin ratio may provide a more accurate test for iron deficiency than ferritin level in the setting of inflammatory disease. However, this test is not routinely available, and a recent meta-analysis suggests that further data are needed to define its overall diagnostic usefulness. (See Diagnosis under Iron Deficiency Anemia.) (Thurnham DI, McCabe LD, Haldar S, et al. Adjusting plasma ferritin concentrations to remove the effects of subclinical inflammation in the assessment of iron deficiency: a meta-analysis. *Am J Clin Nutr.* 2010;92:546-555; Infusino I, Braga F, Dolci A, et al. Soluble transferrin receptor (sTfR) and sTfR/log ferritin index for the diagnosis of iron-deficiency anemia: a meta-analysis. *Am J Clin Pathol.* 2012;138:642-649.)

2. A 55-year-old man with metastatic colon cancer presents 6 months after completing chemotherapy with complaints of fatigue. His complete blood count shows a hemoglobin level of 10 g/dL with microcytic, normochromic indices, and his serum ferritin level is in the high-normal range. He has no history or physical findings of cardiovascular disease. He requests some kind of treatment for his anemia. Which of the following is the most appropriate recommendation?

A. Erythropoietin injections to increase his hemoglobin level
B. Begin red cell transfusions to maintain a hemoglobin of 12g/dL or higher
C. Advise that his fatigue may be somewhat affected by his hemoglobin level but that treatment to increase his hemoglobin level alone is unlikely to improve his symptoms and carries significant risks
D. Oral folate replacement
E. Oral pyridoxine therapy

Answer: C Although erythropoietin injections have been documented to improve quality of life and performance status in patients with cancer and a hemoglobin of less than 10 g/dL who are undergoing chemotherapy, strong data supporting this therapy in patients with a hemoglobin of 10 g/dL or higher is lacking. The same is true for red cell transfusions, in the absence of symptomatic cardiovascular disease. In cancer patients not on chemotherapy, large meta-analyses of randomized trials have shown increased mortality and cancer progression in those treated with erythropoietin for anemia. In the absence of documented folate deficiency or sideroblastic anemia, there is no clear evidence that either folic acid replacement or pyridoxine therapy is of benefit in this setting. (See Treatment under Anemia of Chronic Disease and Inflammation.) (Ludwig H, Crawford J, Osterborg A, et al. Pooled analysis of individual patient-level data from all randomized, double-blind, placebo-controlled trials of darbepoetin alfa in the treatment of patients with chemotherapy-induced anemia. *J Clin Oncol.* 2009;27:2838-2847; Bohlius J, Schmidlin K, Brillant C, et al. Recombinant human erythropoiesis-stimulating agents and mortality in patients with cancer: a meta-analysis of randomised trials. *Lancet.* 2009;373:1532-1542.)

3. An otherwise healthy 20-year-old woman with a history of excessive menstrual bleeding and increased fatigue was found to have a hemoglobin of 7 g/dL with microcytic, hypochromic indices, a low ferritin level, a low serum iron level, and reticulocyte count of 1.2%. Her menstrual bleeding pattern was normalized by hormone therapy, and she was begun on oral iron replacement with ferrous sulfate 350 mg three times daily. After 2 months, her hemoglobin was 8 g/dL, and there was no change in her other hematologic test results. After a full course of 2 g of intravenous iron gluconate, her hemoglobin, red cell indices, and iron level were partially corrected, but she remained somewhat anemic with a hemoglobin of 10 g/dL. A C-reactive protein (CRP) level and an ??$_1$-acid glycoprotein (ACP) level were both normal. Which of the following laboratory tests would be most helpful to potentially explain her lack of complete response to iron therapy?

A. Reticulocyte count
B. Serum vitamin B_{12} level
C. Bone marrow biopsy
D. Genetic analysis for mutation in the *TMPRSS6* gene
E. RDW on automated complete blood count

Answer: D This patient presents with classic routine laboratory findings of iron deficiency anemia. The failure to respond to oral iron replacement therapy suggests some reason for refractoriness to iron absorption. The patient is otherwise healthy, with no findings suggestive of malabsorption or inflammation or other reason to have high hepcidin levels that could account for this. However, the fact that intravenous iron supplementation only partially corrects the patient's anemia and iron deficiency suggests the presence of an excessive hepcidin level. Because the CRP and ACP levels were normal, demonstrating no evidence of inflammation to account for an elevated hepcidin level, it could be caused by a mutation in the *TMPRSS6* gene, which encodes the matriptase-2 protein that acts to suppress excess hepcidin production. Mutations in *TMPRSS6* can be either sporadic or familial. (See Failure to Respond to Iron Therapy under Iron Deficiency Anemia.) (Guillem F, Lawson S, Kannengiesser C, et al. Two nonsense mutations in the TMPRSS6 gene in a patient with microcytic anemia and iron deficiency. *Blood.* 2008;112:2089-2097; Finberg KE. Iron-refractory iron deficiency anemia. *Semin Hematol.* 2009;46:378-386.)

4. A healthy 25-year-old man who was in a minor auto crash was found to have a microcytic, hypochromic anemia (hemoglobin 10 g/dL), unrelated to the accident, because there was no blood loss. The remainder of the complete blood count was normal. The patient was noted to be mildly ataxic but had no alcohol or drugs detectable on screening, and there was no head trauma. Serum iron level was normal, and serum ferritin level was in the high normal range. Workup for evidence of tumor or subclinical inflammation was negative, as was any history of prior drug or toxin exposure. A bone marrow showed normal hematopoiesis but demonstrated ringed sideroblasts. Which of the following tests or procedures is most likely to be useful for determining the underlying cause of this patient's anemia?

A. Upper gastrointestinal endoscopy
B. Genetic testing for mutations of the *TMPRSS6* gene
C. Serum transferrin receptor level assay
D. Genetic testing for gene mutations of the genes coding for the adenosine triphosphate–binding cassette gene *(ABC-7)* and the *ALAS-2* gene
E. Serum folate level assay

Answer: D This patient has a moderate anemia that is not symptomatic, so it could have existed for many years undetected. Despite hypochromic and microcytic red cell features, there is no evidence of occult blood loss, and specific laboratory tests for diagnosing iron deficiency and inflammatory conditions were negative. The presence of ringed sideroblasts in an otherwise healthy young male with no history of drug or toxin exposure suggests that this may be a congenital X-linked sideroblastic anemia. The two major X-linked hereditary types of sideroblastic anemia result from mutations in either the erythroid-specific δ-aminolevulinic acid synthase gene, also known as *ALAS-2*, or the adenosine triphosphate–binding cassette gene *(ABC-7)*. X-linked sideroblastic anemias are often mild with minimal or no symptoms and therefore may go undetected until a supervening process results in symptomatic anemia or they are detected incidentally, such as in this case. The presence of mild ataxia in this young male would be most consistent with a subgroup of patients with *ABC-7* gene mutations who have associated ataxia.

(See Genetics under Sideroblastic Anemias.) (Camaschella C. Hereditary sideroblastic anemias: pathophysiology, diagnosis and treatment. *Semin Hematol.* 2009;46:371-377; Harigae H, Furuyama K. Hereditary sideroblastic anemia: pathophysiology and gene mutations. *Int J Hematol.* 2010;92: 425-431.)

5. A 60-year old man with chronic renal failure who is on hemodialysis has a falling hemoglobin level despite having been on erythropoietin (epo) therapy for a documented low serum epo level. His peripheral blood smear shows microcytic hypochromic red blood cells, and serum iron levels are low, but the serum ferritin level is 120 μg/dL. A bone marrow biopsy shows absent sideroblasts, and workup for occult blood loss is negative. Which of the following therapeutic interventions is likely to be the safest and most effective for increasing the patient's hemoglobin level?

A. Oral iron supplement with ferrous sulfate
B. Intravenous iron gluconate injections
C. Doubling the dose of erythropoietin
D. Intravenous iron dextran injections
E. Oral iron supplementation with ferrous fumarate.

Answer: B The clinical setting and laboratory findings in this patient are all consistent with coexistent anemia of chronic disease and iron deficiency. The normal ferritin level is most likely due to the chronic inflammation seen in patients with renal failure on hemodialysis. Increasing the dosage of erythropoietin in this setting in the absence of increasing available iron for erythropoiesis is unlikely to be beneficial and may introduce increased risk for thrombosis. Oral iron supplementation is usually ineffective in the setting of chronic inflammation due to elevated hepcidin levels that block iron uptake by the gastrointestinal mucosa. Intravenous iron therapy has been demonstrated to be effective in increasing the response to erythropoietin in patients with chronic renal failure. Although both iron dextran and iron gluconate are effective in this setting, iron gluconate has been shown to be safer. (See Parenteral Administration under Iron Deficiency Anemia.) (Michael B, Coyne DW, Fishbane S, et al; Ferrlecit Publication Committee. Sodium ferric gluconate complex in hemodialysis patients: adverse reactions compared to placebo and iron dextran. *Kidney Int.* 2002;61:1830-1839; Goodnough LT, Nemeth E, Ganz T. Detection, evaluation, and management of iron-restricted erythropoiesis. *Blood.* 2010;116:4754-4761; Albaramki J, Hodson EM, Craig JC, et al. Parenteral versus oral iron therapy for adults and children with chronic kidney disease. *Cochrane Database Syst Rev.* 2012;1:CD007857.)

160

AUTOIMMUNE AND INTRAVASCULAR HEMOLYTIC ANEMIAS

MARC MICHEL

DEFINITION

Hemolytic anemia (HA) is defined as anemia caused by a shortened lifespan of mature red blood cells (RBCs) in the peripheral circulation. Hemolysis and accelerated destruction of RBCs can take place within the vasculature (i.e., intravascular hemolysis) or mainly in the liver and the spleen (i.e., extravascular hemolysis). HA can be the consequence of an intrinsic and often genetically determined defect of the RBC membrane (discussed in Chapter 161) or an RBC constituent (hemoglobin [Hb] structure [Chapters 162 and 163] or enzyme machinery [Chapter 161]), or HA can result from an extrinsic and usually acquired disorder of the RBC membrane (immune, infectious, toxic) (Table 160-1).

Autoimmune hemolytic anemia (AIHA) is an acquired autoimmune disease in which autoantibodies directed against autologous RBC membrane antigens lead to their accelerated destruction. The diagnosis of AIHA is thus based on the presence of a positive result on the direct antiglobulin test (DAT), also known as the direct Coombs test, and on the absence of any other hereditary of acquired cause of hemolysis. In AIHA, hemolysis is mainly extravascular, but some features of concomitant intravascular hemolysis may also be present at onset. Beyond AIHA, other causes of immune-mediated HAs that do not involve autoantibodies may occur, as with certain drug-induced HAs or in a posttransfusional hemolytic reaction because of the presence of alloantibodies or other more complex mechanisms (Chapter 177). *Paroxysmal nocturnal hemoglobinuria (PNH)* is a rare and potentially life-threatening acquired clonal blood disorder with protean manifestations in which RBC membranes are highly vulnerable to damage by activated complement. The resulting chronic intravascular hemolysis is the hallmark of the classical hemolytic form of the disease, and the release of free Hb contributes to most of its clinical manifestations.

AUTOIMMUNE HEMOLYTIC ANEMIA

EPIDEMIOLOGY

Autoimmune hemolytic anemia can affect both children (mainly before the age 5 years) and adults and is estimated to have an overall (not age-adjusted) annual incidence of approximately one to three per 100,000 individuals. Whereas boys tend to be more frequently affected than girls, the female-to-male sex ratio is 1.5 to 2 in adults. Among adults, most patients are older than 40 years of age, and the peak incidence occurs around the seventh decade of

TABLE 160-1 PRINCIPAL CAUSES OF HEMOLYTIC ANEMIAS

INTRACORPUSCULAR	EXTRACORPUSCULAR
DISORDERS OF THE RBC MEMBRANE	**IMMUNOLOGIC**
Inherited	• Autoimmune HA
• Hereditary spherocytosis	• Alloimmunization
• Elliptocytosis	• Drug-induced HA
• Hereditary stomatocytosis	**MECHANICAL**
Acquired	• Thrombotic thrombocytopenic purpura
• Paroxysmal nocturnal hemoglobinuria	• Hemolytic-uremic syndrome
HEMOGLOBINOPATHY	• Other microangiopathies
Qualitative defect in hemoglobin	• HELLP syndrome
• Sickle cell disease	• Prosthetic heart valve dysfunction
• Unstable hemoglobins	• Acanthocytosis
Quantitative defect in hemoglobin	**INFECTIOUS**
• β-Thalassemia	• Malaria
• α-Thalassemia	• Babesiosis; *Clostridium perfringens*; gram-positive bacteria
RBC ENZYME ABNORMALITY	**TOXIC**
• G6PD deficiency	• **Exogenous:** thermal burns; industrial copper, arsine, lead poisoning, spider bite, snake bite, mushroom ingestion.
• Pyruvate kinase deficiency	• **Endogenous:** Wilson disease
• Others: pyrimidine 5′ nucleotidase deficiency	• **Drug related**

HA = hemolytic anemia; HELLP = hemolysis, elevated liver function tests, low platelets; RBC = red blood cell.

life. This age distribution may be related to the increased frequency of underlying lymphoproliferative malignancies in elderly adults, resulting in an age-related increase in secondary AIHA caused by lymphoma. Most cases of AIHAs develop sporadically; familial cases are very uncommon. AIHA usually occurs as an isolated immune "cytopenia" but can sometimes be associated simultaneously or sequentially with immune thrombocytopenia (ITP) as *Evans syndrome* or autoimmune neutropenia. Among AIHAs, the so-called warm AIHA (wAIHA; see Diagnosis and Classification section) accounts for 70% to 80% of all cases in adults and almost 90% of the cases in children. Whereas *paroxysmal cold hemoglobinuria (PCH)* is a very uncommon AIHA subtype seen almost exclusively in children, *cold agglutinin disease (CAD)* occurs almost exclusively in adults older than 50 years of age. AIHA can be primary (or idiopathic), or it can occur in association with or disclose an underlying disease (secondary AIHA).

PATHOBIOLOGY

The pathogenesis of AIHA is a complex multistep process involving not only the autoantibodies but also various effectors of the immune system, including the complement system, macrophages, and B and T lymphocytes. Whereas the mechanisms leading to hemolysis have been partially elucidated (antibody-dependent, cell-mediated cytotoxicity, and complement-dependent cytotoxicity being primarily involved), the mechanisms leading to the breakdown of self-tolerance are far from fully understood.

Red Blood Cell Antibodies

IgG anti–RBC autoantibodies mediate the destruction of RBCs mainly by the process of extravascular hemolysis. In contrast, when lytic components of the complement system participate in the process, the destruction of RBCs usually occurs directly within the circulation (intravascular hemolysis). The participation of lytic complement components in IgG-mediated AIHA is rare.

In warm (or warm-reactive) AIHA, the autoantibody targeting RBCs is mainly of the IgG1 isotype. It is able to bind macrophages via its Fc-γ receptors, thereby causing extravascular hemolysis to take place mainly in the spleen. The autoantigens on RBC membranes targeted by the autoantibody are, in decreasing frequency, the following: (1) peptides from the Rhesus system (~one third of the cases); (2) band 3 protein; and (3) glycophorin A, an RBC membrane glycoprotein. In other cases, the antibodies have specificity for antigens in the Kell or Duffy blood group system (very rarely the ABO antigens), and in less than 10% of cases of wAIHA, no specificity can be found. Cold agglutinins are IgM antibodies that react with polysaccharides on the RBC surface, mainly with the I/i antigens or less commonly with the Pr glycoprotein and sialylated polysaccharides. Whereas cold agglutinins associated with *Mycoplasma pneumoniae* infection have an anti-I specificity, those related to infectious mononucleosis have anti-i specificity. In CAD, the cold autoantibodies are almost always anti-I monoclonal IgM antibodies with a VH4-34 heavy chain, a heavy-chain shape of the antigen-binding surface that favors attachment to polysaccharides.

Mechanism of Antibody-Mediated Red Blood Cell Destruction

IgG anti–RBC autoantibodies are opsonins; when bound to autoantigens on RBC membranes, they instigate phagocytosis of the cells by macrophages. Using its Fcγ receptors, the macrophage can ingest an entire IgG-coated erythrocyte or transform it into a spherocyte (microspherocyte) by nibbling away its surface. Antibody-coated spherocytes are more vulnerable to osmotic forces than normal, unsensitized RBCs and ultimately surrender to macrophages, especially in the splenic sinusoids, where blood flows sluggishly. The rate of hemolysis in AIHA depends on the amount of autoantibody on the RBC surface, the affinity and avidity of autoantibodies for the RBC autoantigen, and the number of macrophages in the environment of the antibody-coated erythrocyte. Populations of autoantibodies with high avidity cause a higher rate of RBC destruction than populations with low avidity. Free (monomeric) IgG competes with antibody-coated RBCs for interaction with the Fc receptors of macrophages, but the IgG normally present in plasma has only a minor influence on the hemolytic rate. The importance of the subclasses of IgG is unclear, but IgG3 antibodies seem more potent than IgG1 antibodies in promoting phagocytosis.

The basis of RBC destruction in CAD is the ability of IgM antibodies to fix complement, with each IgM molecule having two binding sites for C1q. When blood cools sufficiently in the extremities, it allows the cold agglutinins to bind RBC. The adherent IgM attracts C1q, which initiates the generation of C3b and C4b on the RBC's surface. On entering the warmer visceral circulation, the RBC releases the cold agglutinin, but the C3 fragments remain engaged to the CR1 of macrophages, thereby enabling phagocytosis of the RBC.

The efficiency of this process depends on the amount of cold agglutinin on the RBC surface and the thermal amplitude of the cold agglutinin. These factors account for the great variability of severity in CAD. The abnormal production of autoantibody directed toward RBC antigens could be the consequence of different and nonmutually exclusive mechanisms: an immune response toward some cryptic antigens or molecular mimicry with cross reactivity between external antigens and autoantigens. Polyclonal activation of both B and T cells is likely to play a role in wAIHA. A positive DAT result is more frequently observed in patients with chronic infections with hypergammaglobulinemia such as HIV (Chapter 393) or leishmaniasis (Chapter 348). In other noninfectious settings with hypergammaglobulinemia and immune dysregulation such as in the *autoimmune lymphoproliferative syndrome (ALPS)* or in *angioimmunoblastic T-cell lymphoma*, a significant proportion of patients have a positive DAT result with or without active hemolysis. Regarding T-cell activation, there is a disequilibrium of the CD4+ T helper 1(Th1)/Th2 balance in patients with active AIHA compared with healthy control participants, with an increase of Th2 cells subsets and an increased expression of both interleukin-4 (IL-4) and IL-10 and a reduced expression of interferon-γ and IL-12. This "Th2 pattern" promotes the induction and proliferation of autoreactive B cell clones. More recently, it has been shown that the production of the effector cytokine IL-17 is strongly associated with AIHA compared with healthy donors and correlates with the severity of the disease. This observation suggests some future therapeutic potential for the use of anti-IL17 monoclonal antibodies in AIHA.

The role of a regulatory T cells (Tregs) defect in the loss of tolerance in wAIHA has been mainly suggested through the study of animal models (e.g., New Zealand Black mice). In humans, few data on the potential role of Tregs in AIHA is available. Some Tregs specific for autoantigens from the Rhesus system are able to inhibit the Th1 effector immune response in vitro through an IL-10–dependent mechanism. Indirect evidence suggesting that a decrease in the number or function of Tregs is likely to play a role in AIHA in humans comes from the **immune dysregulation polyendocrinopathy enteropathy X-linked** (IPEX) syndrome (Chapter 250). IPEX syndrome is a rare inherited disease linked to the dysfunction of the transcription factor FOXP3, widely considered to be the master regulator of the regulatory T-cell lineage. Patients diagnosed with an IPEX syndrome have a high risk of developing a number of autoimmune manifestations (enteropathy; endocrinopathies, including diabetes), including in a lesser extent AIHA.

CLINICAL MANIFESTATIONS

Clinical symptoms of AIHA are those of anemia (unusual fatigue, exertional dyspnea, tachycardia) or those attributable to active hemolysis (jaundice with or without dark urine). Moreover, specifically in patients with high-affinity cold agglutinins, exposure to cold can precipitate episodes of acrocyanosis by inducing massive agglutination of RBCs in the capillary circulation of the hands, feet, or both. Mild splenomegaly may be present, especially in wAIHA; splenomegaly is uncommon in CAD unless there is an underlying B-cell lymphoma. AIHA is associated with an increased risk of venous thromboembolism.[3]

The usual clinical and biologic features of AIHA are shown in Table 160-2. Elevated serum levels of indirect bilirubin and lactate dehydrogenase (LDH) and a reduced serum haptoglobin concentration are the usual albeit nonspecific signs of HA. Laboratory signs of intravascular destruction of RBCs (hemoglobinemia, hemoglobinuria, and hemosiderinuria) are unusual in the setting of AIHA. After HA has been recognized, the diagnosis of AIHA is usually rather easy. It is based first on the identification of anti-RBC autoantibody by means of the DAT, also known as the Coombs test, and then on the exclusion of other causes of HA. To rule out other causes of hereditary or acquired hemolysis, ethnicity must be taken into account, and the history should also focus on previous personal episodes of unexplained anemia with or without jaundice and any familial history of HA or splenectomy. Moreover, a careful analysis of the peripheral blood smear is essential, keeping in mind that an increase in the number of spherocytes can be observed in approximately 30% to 40% of AIHAs.

DIAGNOSIS

The Direct Antiglobulin (Coombs) Test

The DAT that reveals antibody-coated RBCs is central to the diagnosis of autoimmune HA. The indirect antiglobulin (Coombs) test was devised to test for the presence of incomplete antibodies in the patient's serum, not on RBC surfaces. As presently used, the standard antiglobulin reagent contains antibodies against all four classes of IgG and components of complement (usually C3 and C4).

A positive DAT result requires cautious interpretation when there are no other features of AIHA. False-positive test results are not unusual. The reported incidence of positive DATs in normal blood donors and general populations of hospitalized patients varies widely, from one in 100 to one in 15,000. Differences in the technique used to perform the test account for this variation. The usual reason for a false-positive DAT result is nonspecific, low-avidity adherence of IgG to RBCs. In rare cases, however, the result is not a false-positive result but a harbinger of the development of AIHA. False-negative DAT results in true AIHA are very rare and are usually due to low-affinity autoantibodies that spontaneously elute from the RBC in vitro or to amounts of erythrocyte-coating antibodies that are below the limit of detection by the antiglobulin test. Before considering the diagnosis of DAT-negative AIHA[4] as a possibility, every other potential cause of HA has to be excluded. The distinction between a true-positive and a false-positive DAT result can be made by eluting the antibody from the RBCs and testing its ability to bind to normal RBCs. In a false-positive reaction, the eluted antibody does not bind to normal RBCs, but binding does occur in a true-positive test.

Classification of Aiha: Warm versus Cold Autoimmune Hemolytic Anemia (see Table 160-2)

After the diagnosis of AIHA is made, the classification of the AIHA type is essential.

According to the DAT pattern and the optimal temperature at which the autoantibody reacts with human RBCs, AIHAs are typically subdivided in four major subtypes: (1) wAIHA, (2) cold AIHA (including the chronic CAD), (3) mixed-type AIHA, and (4) the exceptional PCH. The great majority of cold-reactive autoantibody are so-called "cold agglutinins," cold hemolysins being indeed much less common. A small proportion (~5%) of patients may exhibit both cold-reactive and warm-reactive autoantibodies defining mixed AIHA. The distinction between cold and wAIHA is very important because the mechanisms of RBC injury, the primary site of hemolysis, and especially the therapeutic approaches are significantly different. Table 160-2 summarizes the main features of the different types of AIHA.

In wAIHAs, which represent 70 to 80% of all AIHAs in adults, the DAT pattern is either solely IgG positive or IgG and C3 positive, autoantibodies are of IgG isotype (mainly IgG_1), and they bind and react optimally with RBCs at a temperature of 37° C (range, 35 to 40° C). In wAIHA, the hemolysis is mainly extravascular and occurs predominantly in the spleen. On the other hand, in AIHAs caused by cold autoantibodies, which account for 15% to 20% of AIHAs, the usual DAT pattern is IgG negative and C3 positive, and circulating cold agglutinins are detectable in the serum at a significant titer (i.e., >1/64). Cold autoantibodies that are also cold agglutinins are almost exclusively of IgM isotype. However, the RBC antigen-reactive IgM immunoglobulins themselves that are able to activate the classical complement pathway on RBC membranes in vivo are seldom found fixed on the RBCs membrane on laboratory testing because they are rapidly eluted ex vivo. Cold antibodies typically react with RBCs at temperatures below 30° C (optimum, 4° C) and lead to various degree of extravascular hemolysis, predominantly in the liver. In rare cases of AIHAs (~10%) known as mixed-type AIHAs, autoantibodies (mostly of IgM isotype) can react with RBCs within a wide

TABLE 160-2 MAIN CHARACTERISTICS OF VARIOUS TYPES OF AIHA

AIHA TYPE	EPIDEMIOLOGY/ TYPE OF HEMOLYSIS	SECONDARY FORMS	AUTOANTIBODY ISOTYPE	OPTIMAL TEMPERATURE	DAT PATTERN	ELUATE	AUTOANTIBODY SPECIFICITY (TARGETED RBC ANTIGENS)
Warm AIHA	~70%-80% of all AIHA; adults > children (mean age, 3 yr); EV hemolysis; subacute onset (rarely abrupt with associated IV hemolysis)	~50% of cases (see Table 160-3)	IgG ≫ IgA, IgM	37° C	IgG ± C3d	IgG	Antigens of the Rh system, band 3, glycophorin, A
Cold agglutinin syndrome	~20%-30% of all AIHA cases in adults; age >50 yr; EV hemolysis	CAD associated with monoclonal IgM κ gammopathy in 90% of cases ± features of definite clonal B-cell lymphoproliferative disorder	IgM ⋙ IgA or IgG; cold agglutinin titer >1/500	4° C	C3	Negative	I antigen > i ≫ Pr
Cold transient AIHA	Children, young adults; IV hemolysis	Infections (*Mycoplasma pneumoniae*, EBV, other viruses)	Polyclonal IgM; cold agglutinin titer ≥1/64	4° C	C3	Negative	I > i antigens
Paroxysmal cold hemoglobinuria	Children (rare); exceptional in adults; acute IV hemolysis	Infections (*M. pneumoniae,* virus)	IgG (Donath-Landsteiner hemolysin)	>30° C	C3	Negative	P + c antigens (biphasic hemolysins)
Mixed-type AIHA	Adult; mainly EV hemolysis	Mainly B-cell lymphoma	IgG, IgM	Wide range (4°-37° C)	IgG ± C3	IgG	Polyreactivity

AIHA = autoimmune hemolytic anemia; CAD = cold agglutinin disease; DAT = direct antiglobulin test; EBV = Epstein-Barr virus; EV = extravascular, Ig = immunoglobulin; IV = intravascular

range of temperatures (thermal amplitude from 4° to 37° C). Last, PCH is a very uncommon AIHA subtype that is seen almost exclusively in children. It is caused by an IgG hemolysin that reacts with autologous RBCs only at low temperatures (2° to 10° C) and can be detected only by means of the Donath Landsteiner test, a test available only in few referral laboratories. Besides these four distinct subtypes of AIHAs, drug-induced HAs must be regarded as a distinct entity.

In addition to the distinction between warm and cold AIHA, which is an essential step in diagnosis because it does influence the therapeutic strategy, AIHAs are also classified as (1) primary (or idiopathic) or (2) secondary, depending on the presence or absence of an underlying disease or condition promoting immune dysregulation. On average, 50% to 60% of wAIHAs are secondary (see Table 160-3 for associated conditions); therefore, a minimal workup must be performed at time of diagnosis in every patient to search for an underlying disease or condition (Table 160-4). It is important to emphasize that a presumably primary form of wAIHA can precede by many years the occurrence of a non-Hodgkin lymphoma, and therefore patients with wAIHA must be followed even after a remission of AIHA has been obtained.

Regarding AIHAs due to cold autoantibodies, in children and young adults they are mainly secondary to either bacterial (*M. pneumoniae* infection) or viral (Epstein-Barr virus [EBV] or cytomegalovirus) infection. In adults older than 50 years of age, although rare, chronic CAD is by far the most common cause of AIHA caused by cold autoantibodies.[5] Although chronic CAD in an older adult has long been considered as a primary (i.e., idiopathic) form of AIHA, it is now recognized that in about 90% of cases, it is associated with a monoclonal IgM κ and with other features of a clonal B-cell lymphoproliferative disorder in approximately 75% of cases. Thus, chronic CAD should be viewed more as a true lymphoproliferative disease characterized by clonal expansion of B cells with most often features of lymphoplasmacytic lymphoma in the bone marrow. HA is seldom very severe in the setting of chronic CAD, and it is classically exacerbated by exposure to cold or infections.

TABLE 160-3 MAIN DISORDERS OR CONDITIONS ASSOCIATED WITH SECONDARY WARM-REACTIVE AUTOIMMUNE HEMOLYTIC ANEMIAS

Hematologic disorders and lymphoproliferative diseases
- Chronic lymphocytic leukemia,* acute lymphoblastic leukemia,† LGL leukemia
- B-cell lymphoma,* Hodgkin lymphoma
- Angioimmunoblastic T-cell lymphoma
- Castleman disease
- Myelodysplasias, myelofibrosis

Solid tumors
- Thymoma
- Ovarian dermoid cyst
- Carcinomas

Autoimmune and inflammatory diseases
- Systemic lupus erythematosus, antiphospholipid syndrome
- Rheumatoid arthritis
- Inflammatory bowel disease
- Pernicious anemia, thyroiditis
- Myasthenia gravis
- Autoimmune hepatitis, giant cell hepatitis*
- Sarcoidosis
- Eosinophilic fasciitis

Infections
- Viruses: EBV,* hepatitis C, CMV
- Bacteria: tuberculosis, brucellosis, syphilis

Drugs

Primary immunodeficiencies
- Common variable immunodeficiency
- Hyper-IgM syndrome,† ALPS†
- IPEX syndrome,† APECED syndrome†

Others
- Pregnancy
- After allogeneic bone marrow transplantation or liver or small bowel transplant*
- Rosai-Dorfman disease

*Diseases that can also be associated with cold autoimmune hemolytic anemia and cold agglutinins.
†Seen almost exclusively in childhood.
ALPS = autoimmune lymphoproliferative syndrome; APECED = autoimmune polyendocrinopathy with candidiasis and ectodermal dystrophy; CMV = cytomegalovirus; EBV = Epstein-Barr virus; IPEX = immune dysregulation, polyendocrinopathy, enteropathy X-linked.
LGL = large granular lymphocytes.

TREATMENT

Rx

The management of AIHA is, regardless the subtype, mainly empirical or based on retrospective uncontrolled studies. Very few prospective studies have been reported (mainly in CAD), and only two randomized controlled studies have been performed in wAIHA to date.

Supportive Care and Transfusion

Regardless the type of AIHA, folic acid supplementation (5-10 mg/day) is warranted in patients with active AIHA to prevent further depletion of folate stores (caused by increased erythropoiesis), which may be misinterpreted as a treatment failure. RBC transfusions are also indicated in patients with disabling symptoms of anemia or a poor underlying cardiovascular condition (i.e., coronary artery disease or heart failure). Although younger patients may tolerate a stable Hb level as low as 6 g/dL, in patients with comorbidities, maintaining an Hb level of at least 8 g/dL is usually recommended. It is important for the managing physician to understand that no patient with symptomatic AIHA should be denied blood transfusions because of an "incompatible crossmatch." The blood bank should be informed of the patient's status. Indeed, the patient's positive DAT result almost always interferes with compatibility testing, so the role of the blood bank is to provide packed RBCs that are the "least incompatible" ones in regard to the specificity of the patient's autoantibody. It is possible that DNA-based methods might become clinically feasible to replace hemagglutination assays in these situations in the future.[6] Close communication and cooperation between the clinician and the specialist in transfusion medicine is therefore essential for reducing the risks associated with transfusion in patients with AIHA. Because transfused RBCs can be destroyed by the patient's autoantibodies, rapid transfusion of large volumes of RBCs must be avoided because they can have serious consequences. This risk is increased if the patient also has alloantibodies induced by previous transfusions or pregnancy. Packed RBCs units should therefore not be administered at a rate that exceeds 1 mL/kg/hour. There is no strong evidence supporting the efficacy of plasma exchange in enhancing the response to blood transfusions.

Although this is not supported by strong evidence-based data, in cold AIHA, it is usually recommended to transfuse prewarmed (at a temperature close to 37° C) packed RBCs by means of a specific warming device to minimize the risk of hemolysis.

Treatment of Acute Cold Autoimmune Hemolytic Anemias

In cases of transient cold AIHAs induced by an infection, the onset is usually abrupt, and the degree of anemia may be profound and sometimes life threatening, requiring transfusions of prewarmed packed RBCs, which must not be postponed. Besides supportive care, the use of antibiotics is justified in case of pneumonia caused by infection with *M. pneumoniae*. A short course of corticosteroids can sometimes be considered in case of severe cold AIHA secondary to a viral infection (EBV) to reduce the duration of anemia, but this approach is not evidence based and not uniformly recommended.

Treatment of Chronic Cold Agglutinin Disease

The treatment of CAD has long been only supportive and symptomatic. Avoidance of cold is still the mainstay in the management of CAD, especially to avoid cold-induced circulatory symptoms, but this measure is often not sufficient to avoid episodes of hemolysis. With cold exposure, in cases of infections and acute phase reactions, the level of both C_3 and C_4 increase because of enhanced production, resulting in exacerbation of complement-dependent

TABLE 160-4 RECOMMENDED WORKUP IN WARM-REACTIVE AUTOIMMUNE HEMOLYTIC ANEMIA

1. Antinuclear antibodies (± anti-DNA Abs if ANA positive)
2. Anticardiolipin antibodies and lupus anticoagulant*
3. Serum protein electrophoresis and immunoelectrophoresis
4. Immunophenotyping of peripheral lymphocytes
5. CT scan of the thorax, abdomen, and pelvis (in the absence of obvious SLE or APS)
6. Bone marrow biopsy: recommended only in the presence of hypogammaglobulinemia, or monoclonal gammopathy, or abnormal lymphadenopathy on the CT scan

*Especially in case of previous episode of venous and/or arterial thrombosis or in case of recurrent miscarriage in women, and systematically before splenectomy.
Abs = antibodies; ANA = antinuclear antibodies; APS = antiphospholipid syndrome; CT = computed tomography; SLE = systemic lupus erythematosus.

cytotoxicity caused by autoantibody directed toward autologous RBCs, leading to increased hemolysis as well as agglutination. Consequently, prompt treatment of febrile infections or illnesses as well as prevention of infection by means of influenza with or without pneumococcal vaccines is strongly recommended in patients with CAD. In case of exacerbation of hemolysis and severe anemia, transfusion of prewarmed packed RBCs must be considered because patients with CAD are often elderly with comorbid conditions. Whereas CAD is usually considered a rather indolent and slowly progressive lymphoproliferative disease with a relatively good prognosis, transfusion dependency may be observed in some patients with recurrent episodes of active hemolysis.

In patients with CAD who have active hemolysis and need to be treated, a number of single-agent therapies have shown little or no efficacy, and their use should therefore be avoided because they all have a significant toxicity. These include corticosteroids, alkylating agents, interferon-α, immunosuppressives, and cladribine. Oral cyclophosphamide and chlorambucil have shown favorable responses in a minority of patients with CAD. Because hemolysis takes place mainly in the liver in CAD, splenectomy should be avoided in these individuals because it is notoriously ineffective. In transfusion-dependent patients with normal or slightly elevated reticulocyte counts, the transient use of an erythropoiesis-stimulating agent off label may be useful as a transfusion-sparing strategy. Plasma exchanges can be temporarily helpful in cases of severe hemolysis or in preparation for surgery requiring cold exposure, such as lung–heart surgery. For patients with active or relapsing symptomatic episodes of HA, rituximab given either alone or in combination with fludarabine can be an option, especially in patients with an underlying B-cell lymphoma. An algorithm for the management of cold AIHAs is proposed in Figure 160-1. Last, because the hemolytic activity of cold agglutinins is complement dependent, the efficacy of eculizumab, an anti C5 humanized monoclonal antibody licensed in PNH (see later section Paroxysmal Nocturnal Hemoglobinuria), seems theoretically promising. Because only very few case reports have shown the efficacy of eculizumab, this approach should be only considered as a last-resort option in case of life-threatening exacerbation of HA not controlled by RBC transfusions.

Treatment of Primary Warm Autoimmune Hemolytic Anemias

First-line Treatment: Corticosteroids

Primary wAIHAs usually have a chronic course and, except for the very few patients who are unusual for having a mild compensated hemolysis with a normal or almost normal Hb level, treatment is needed in the large majority of the cases to improve RBC survival and significantly and durably increase the Hb level (Fig. 160-2). Corticosteroids are the cornerstone of therapy in wAIHAs, and they must be given as first-line treatment. As noted later under the discussion of rituximab, a recently reported randomized trial showed that rituximab and prednisolone combined, compared with prednisolone alone, increased the rate and duration of response as first-line treatment of warm antibody-reactive AIHA. Intravenous immunoglobulin (IgIV) has only little efficacy in wAIHAs, and because of their cost, they should be considered (at a total dose of 2 g/kg over 2 days) only in patients with severe, transfusion-dependent AIHA and in the absence of response to corticosteroids. The corticosteroid regimen is usually based on oral prednisone or prednisolone at the initial daily dose of 1 to 2 mg/kg. By analogy with other autoimmune diseases, the use of intravenous methylprednisolone at a dose of 250 to 1000 mg/day for 1 to 3 days may be considered in patients with profound anemia, although no clinical trials supporting their higher efficacy are available. The starting dose of oral prednisone is usually maintained for 3 to 4 weeks and then tapered progressively in case of at least partial initial response. There is no agreement on the total duration of treatment, but because the likelihood of early relapse is high when the treatment is prematurely stopped, corticosteroids should be maintained for at least 3 months after a complete response (defined by an increase of the Hb level back to normal and no active hemolysis) is achieved. Except for children, the use of alternate-day prednisone before stopping the treatment is not recommended. The effect of corticosteroids can take few days to 2 weeks, and one or several blood transfusions may be necessary at AIHA onset, especially in case of severe anemia in young children or elderly adults for maintaining a "safe" Hb level. After 2 to 3 weeks of treatment with corticosteroids, a clinically significant response is observed in 80% to 85% of the cases. Except for the few patients who are truly refractory to corticosteroids, the major issue that clinicians have to deal with when treating patients with wAIHA is that approximately 50% to 60% of them turn out to be dependant to corticosteroids. Thus, flares of relapses or wAIHA may occur either within weeks after withdrawal of corticosteroids or even while on treatment when the dose of daily prednisone is decreased below a threshold, which is usually between 10 to 15 mg, a dose that is associated with several adverse events in the long term. Overall, only one third of the patients can be considered to be in complete remission off treatment 1 year after disease onset.

Second-line Treatment

Danazol

In patients dependent on corticosteroids, the use of danazol, an attenuated androgen analog, at 400 to 800 mg/day can be helpful as a corticosteroid-sparing strategy in adults requiring a dose of daily prednisone greater than 15 mg to maintain a remission. However, because of its common masculinizing side effects, its use is usually rather limited in females, and its potential liver toxicity makes its long-term use also problematic in men.

Rituximab

The efficacy of rituximab, the well-known chimeric monoclonal antibody that targets CD20 antigen on B lymphocytes, was first shown in refractory wAIHA in children with a response rate reaching up to 100% in some studies. In adults, rituximab has also shown (both through retrospective but also a few prospective studies, including a randomized controlled trial) to be highly effective in primary wAIHAs, with an 80% to almost 100% response rate at 1 year. Rituximab can be effective in patients who have failed to respond to splenectomy (see later discussion),[7] and, conversely, patients who do not respond to rituximab are able to achieve a response after splenectomy. The

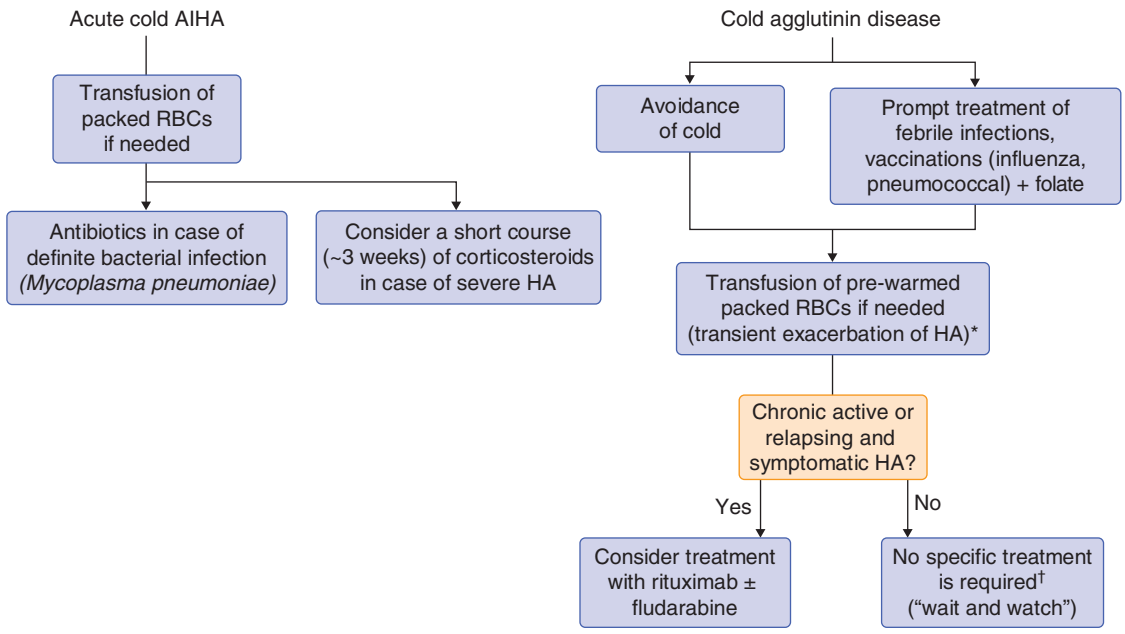

FIGURE 160-1. Proposed algorithm for the management of cold-reactive autoimmune hemolytic anemias (AIHAs). HA = hemolytic anemia; RBC = red blood cell. *The short-term use of an erythropoiesis stimulating agent can be considered off-label; †Unless the underlying B cell-clonal lymphoproliferative disorder requires any specific management.

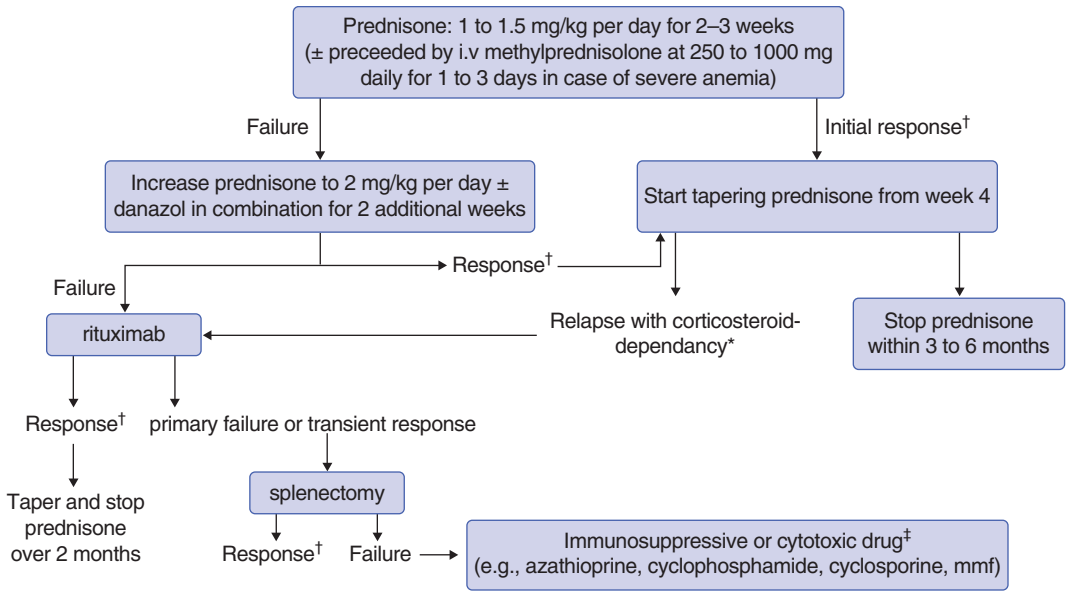

FIGURE 160-2. Proposed algorithm for the treatment of primary warm-reactive autoimmune hemolytic anemia (AIHA) in adults. IV = intravenous; mmf = mycophenolate mofetil. *Dose of prednisone ≥10 mg/day needed to maintain at least a partial response (i.e., Hb level >10g/dL with at least a 2 g increase from baseline without recent transfusion). †Partial response defined as a hemoglobilin level >10g/dL with at least a 2 g increase from baseline and complete response defined as a normal hemoglobin level withouth hemolysis. ‡In alphabetical order (no evidence for preferring any one of these drugs).

usual dose of rituximab is 375 mg/m^2 by intravenous infusion once a week for 4 weeks, but other regimens can be used. The safety profile is usually acceptable, although late-onset neutropenia and opportunistic infections, such as *Pneumocystis jiroveci* pneumonia, may rarely occur. Therefore, the systematic use of primary antibiotic prophylaxis should be considered in patients with AIHA treated with rituximab. Taken together the data from the literature clearly support, whenever possible, the use of rituximab (off-label use) in patients with chronic active or relapsing wAIHA who need to pursue a daily dose of prednisone (or prednisolone) of 15 mg or greater to maintain at least a partial remission. If rituximab is administered before splenectomy, vaccination against *Streptococcus pneumoniae* with or without *Haemophilus influenzae* type B or *Neisseria meningitides* must be systematically administered whenever possible 2 weeks before rituximab because splenectomy may be required thereafter.

Splenectomy

Splenectomy has long been the main and preferred second-line option for the treatment of primary wAIHAs. The rate of sustained response after splenectomy is approximately 60% to 70% according to the most recent data from the literature, but predicting factors of response are still lacking. The perioperative risk of laparoscopic splenectomy is low and acceptable with a mortality rate of less than 1%. The most feared complication remains the rare but unpredictable risk of overwhelming sepsis. Laparoscopy does not reduce the risk of postoperative thromboembolic complications, especially in the portal vein system. A systematic perioperative course of low-molecular-weight heparin is therefore recommended in patients with wAIHA who undergo splenectomy and especially in those who have positive antiphospholipid antibodies. The best time for splenectomy is controversial now that alternatives such as rituximab are available at least in some countries. In children younger than 5 to 7 years of age, this procedure should be avoided and delayed as long as possible. In adults, it must be considered early in the course of the disease in patients who fail to respond to corticosteroids (or need high and unacceptable doses to maintain at least a partial remission) and rituximab.

Other Treatment Lines

Immunosuppressive and Cytotoxic Agents

In patients with refractory wAIHA who have failed splenectomy and rituximab, the management is mainly based on the experience of the individual hematologist and on the few retrospective data available in the literature as opposed to prospective studies. The efficacy of azathioprine, cyclophosphamide, and to a lesser extent cyclosporine and mycophenolate mofetil has been reported in small cases series. The choice depend on the efficacy-to-safety ratio for each patient, and these drugs should be reserved for patients who have failed to respond to rituximab and to splenectomy or who, because of comorbidities, are not suitable candidates for splenectomy.

DRUG-INDUCED IMMUNE HEMOLYTIC ANEMIA

There are several mechanisms by which a drug can induce HA, and true drug-induced AIHAs are rare.[8] Many drugs or drug metabolites have the potential to elicit antidrug antibodies. Drugs that form covalent bonds with proteins in the RBC membrane can bind antidrug antibodies to the RBC surface, causing a positive DAT result and, in some cases, initiating antibody-mediated destruction of RBCs. Other drugs, such as the cephalosporins, can bind to RBC membranes and take up IgG nonspecifically from plasma. In these cases, there is no antidrug antibody. The diagnosis of drug-induced immune-mediated HA should be considered if the patient has a history of taking a suspected medication, there is acute complement-mediated hemolysis, only complement components are detectable on the RBC surface, or the patient's serum reacts with RBCs in the presence of the suspected drug. Some drugs can induce true autoantibodies against RBCs. Fludarabine, a purine nucleoside analogue used in the treatment of chronic lymphocytic leukemia, and monoclonal antibodies against tumor necrosis factor-α (infliximab and adalimumab), T cells (alemtuzumab), α4 integrin (natalizumab), and IL-2 receptor (daclizumab) also have this property, the cause of which is unknown. Notably, there is no definitive way of distinguishing drug-induced AIHA from primary AIHA.

PAROXYSMAL NOCTURNAL HEMOGLOBINURIA

DEFINITION

Paroxysmal nocturnal hemoglobinuria is a rare and potentially life-threatening clonal blood disorder with protean manifestations caused by an acquired somatic mutation in the phosphatidylinositol glycan *(PIG)-A* gene. In pluripotent hematopoietic stem cells, the mutation in *PIG-A* leads to a deficiency of glycosylphosphatidylinositol (GPI)-anchors and GPI-anchored membrane proteins, including the complement regulatory proteins CD55 and CD59 that are normally expressed on the surfaces of RBCs (and other blood cells). PNH RBCs are therefore highly vulnerable to the activation of complement (especially at times of fever, acidosis or hypoxia) and the unregulated formation of the membrane attack complex (MAC). The resulting chronic intravascular hemolysis is the hallmark of the classic hemolytic form of the disease and the release of free Hb contributes to most of its clinical manifestations (fatigue, dysphagia, recurrent abdominal pain, erectile dysfunction).

EPIDEMIOLOGY

Paroxysmal nocturnal hemoglobinuria can present at any age but most commonly between 10 and 50 years. The mean age at diagnosis is about 34 years

(median age is about 40 years), and the female-to-male ratio is close to 1. The median survival time after diagnosis is approximately 20 years. It is a rare disorder with an estimated prevalence in the population of one in 10^5 to one in 10^6. A family history of PNH is unusual.

PATHOBIOLOGY

Genetics

Paroxysmal nocturnal hemoglobinuria is caused by a somatic mutation that causes a defect in the RBC membrane.[9] The disease begins in a single hematopoietic stem cell in which the *PIGA* gene on the short arm of the active X chromosome acquires a somatic mutation. The *PIGA* gene encodes PIG-A, an enzyme that is essential for the synthesis of glycosylphosphatidylinositol (GPI). The lipid GPI forms a peptide link with the C-terminal amino acid of numerous proteins, normally anchoring them to the RBC membrane. The somatic mutation in a hematopoietic stem cell affects *PIG-A* in blood cells of all lineages. Almost 150 different mutations of *PIG-A* have been identified. Most of them inactivate *PIG-A* and cause total loss of the GPI anchor in the descendants of the affected hematopoietic stem cell. RBCs with complete deficiency of GPI are termed *PNH III erythrocytes*, and those with partial deficiency are called *PNH II erythrocytes*. The coexistence of PNH III and PNH II RBCs in the same patient indicates the presence of two mutant clones. A small number of hematopoietic stem cells in normal people bear the *PIG-A* mutation; they have no proliferative advantage and persist in small numbers. In normal blood, the frequency of PIG-A–deficient cells is about one in 50,000 RBCs. In contrast to those with PNH, however, the deficient cells in normal subjects arise from committed hematopoietic cells. The presence of PIG-A–deficient cells in normal subjects suggests that PNH involves not only the *PIG-A* mutation but also a second step, perhaps another mutation, that allows competitive expansion of the mutated clone.

Functional Consequences of Deficiency of GPI

The membrane inhibitor of reactive hemolysis (CD59, or protectin) and CD55, an inhibitor of C3 convertase, are two of the many proteins that GPI anchors to the RBC under normal circumstances. They prevent polymerization of C9, the final step in assembly of the MAC that begins with cleavage of C5 to C5b. Deficiencies of CD59 and CD55 on PNH RBC membranes thereby allow unimpeded assembly of the MAC on the erythrocyte surface (and the surfaces of other blood cells derived from the mutated hematopoietic stem cell clone), thereby initiating intravascular hemolysis. A variety of nonspecific factors, such as a reduction in the pH of blood, can activate complement. The morning hemoglobinuria of PNH is probably the result of subtle acidification of blood during sleep.

CLINICAL MANIFESTATIONS

Classically, a patient with PNH arises in the morning and passes dark urine. The typical but actually rarely described paroxysms of hemoglobinuria occur on a background of chronic, low-grade intravascular hemolysis that causes constant hemosiderinuria in PNH. About one third of cases evolve into aplastic anemia (Chapter 165). Transformation to acute myelogenous leukemia is a rare event. Abdominal pain, dysphagia, and erectile dysfunction are additional clinical features. The basis of these symptoms is probably the scavenging by free plasma Hb of nitric oxide, a vasodilatory regulator of vasomotor and smooth muscle tone. In about one third of cases, venous thrombosis occurs in unusual sites and can cause Budd-Chiari syndrome by obstructing the hepatic veins, portal vein thrombosis or less frequently cerebral vein thrombosis (Chapter 143). Splenomegaly is uncommon; hepatomegaly and ascites suggest the complication of hepatic vein thrombosis and the resultant portal hypertension. Hemosiderinuria is the result of chronic intravascular hemolysis. Subtle or overt signs of bone marrow damage (coexisting leukopenia and thrombocytopenia) are frequent. The extent of RBC destruction in PNH depends on the number of PNH (versus normal) RBCs in blood, the level of GPI on the RBC membrane (PNH III cells are devoid of GPI), and the degree of activation of complement at the cell surfaces. The anemia is often aggravated by iron deficiency caused by chronic urinary iron loss in the form of hemosiderinuria. Long-term, repeated episodes of hemoglobinuria may lead to iron deficiency. Therefore, a DAT-negative HA associated with iron deficiency should raise suspicion of PNH. The basis of the tendency to develop venous thrombosis[10] is unclear. Hypercoagulability caused by the release of prothrombotic materials from RBC and platelet membranes (the latter being also abnormal in PNH) and impaired fibrinolysis has been implicated. Nitric oxide scavenging by free plasma Hb may also cause vasoconstriction and endothelial cell dysfunction, leading to the activation and aggregation of platelets.

DIAGNOSIS

Often, the clinical picture is virtually diagnostic. The diagnosis can be established by demonstrating, by flow cytometry, a deficiency of CD59 on erythrocytes. Another reagent with utility in flow cytometry is Aerolysin, a bacterial protein that binds to the GPI anchor. The fluorescinated Aerolysin variant (FLAER) is also a very good and reliable reagent to study GPI-linked antigens on leukocytes, helpful for diagnosing PNH. Flow cytometry can also measure the proportions of PNH III and PNH II RBCs in blood, providing information about the severity of the disease. The usual cut-off being the presence more than 5% GPI-AP–deficient polymorphonuclear cells in the peripheral blood, the clone size is usually correlated with the degree of intravascular hemolysis.

TREATMENT

Rx

Eculizumab, a humanized monoclonal antibody against C5 that is essential for formation of the MAC, can reduce the signs of intravascular hemolysis, the requirement for transfusions, and the tendency to thrombosis. In a randomized trial,[A2,A3] the dose of the antibody was 600 mg every week for 4 weeks followed 1 week later by a 900-mg dose and then by 900 mg every other week for a total treatment period of 52 weeks. A thrombotic event is a strong indication for eculizumab treatment. A molecular basis for poor response to eculizumab was recently elucidated in a small population of Japanese patients who were found to have a missense mutation in the gene encoding C5 that made the complement factor incapable of binding to eculizumab and being blocked by it.[11] Peptide inhibitors of C3 activation have been shown to prevent hemolysis and C3 opsonization of PNH RBCs and are potential therapeutic candidates.[12]

Warfarin can also be used in patients with a history of a thrombosis. Oral iron can correct the iron deficiency; treatment with iron does not exacerbate the hemolysis. Transfusions are helpful in supportive care. Aplastic anemia (Chapter 165) has been treated successfully with immunosuppressive agents (antithymocyte globulin, usually at a dose of 1.5 mg/kg/day for 7-14 days) with or without cyclosporine (3-5 mg/kg for at least 3 months). Allogeneic bone marrow transplantation (Chapter 178) is risky but can be curative.

OTHER EXTRACORPUSCULAR HEMOLYTIC ANEMIAS

Hemolytic Transfusion Reactions

The cause of hemolytic transfusion reactions (Chapter 177) is intravascular lysis of the donor's RBCs by antibodies (alloantibodies or isoantibodies) in the recipient that bind to one or more blood group antigens on the transfused cells. The recipient's isoantibodies can be natural anti-A or anti-B antibodies, or they can be induced by previous transfusions or pregnancy. Whether IgM or IgG, the isoantibodies trigger the assembly of lytic complement components on the surface of the donor's RBC. The rapid formation of large amounts of C3a and C5a fragments causes hypotension and bronchial and smooth muscle spasm. Renal failure is a consequence of severe, prolonged hypotension; the main renal lesion is renal cortical ischemia secondary to shunting of blood away from the kidneys. Hb itself is not nephrotoxic. The signs and symptoms of a hemolytic transfusion reaction are nonspecific and include fever, back pain, urticaria, dyspnea, hypotension, and evidence of disseminated intravascular coagulation. These nonspecific signs appear and worsen during administration of the transfusion. Immediate steps must be taken to stop the transfusion, submit the transfused blood and a sample of the patient's blood to the blood bank, and order tests of plasma and urine for free Hb. Management is further discussed in Chapter 177.

Other Causes of Intravascular Hemolysis

Conditions in which vascular abnormalities, toxins, infections, or drugs damage RBCs and cause them to lose pieces of membrane and ultimately fragment into Hb-containing bits should be considered in the differential diagnosis of intravascular hemolysis (see Table 160-3). Most of these conditions are readily apparent from the history and physical examination. Treatment focuses on the underlying cause of the hemolysis.

Grade A References

A1. Birgens H, Frederiksen H, Hasselbalch HC, et al. A phase III randomized trial comparing glucocorticoid monotherapy versus glucocorticoid and rituximab in patients with autoimmune haemolytic anaemia. *Br J Haematol*. 2013;163:393-399.

A2. Hillmen P, Young NS, Schubert J, et al. The complement inhibitor eculizumab in paroxysmal nocturnal hemoglobinuria. *N Engl J Med*. 2006;355:1233-1243.

A3. Brodsky RA, Young NS, Antonioli E, et al. Multicenter phase 3 study of the complement inhibitor eculizumab for the treatment of patients with paroxysmal nocturnal hemoglobinuria. *Blood*. 2008;111:1840-1847.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bass GF, Tuscano ET, Tuscano JM. Diagnosis and classification of autoimmune hemolytic anemia. *Autoimmun Rev*. 2014;13:560-564.
2. Xu L, Zhang T, Liu Z, et al. Critical role of Th17 cells in development of autoimmune hemolytic anemia. *Exp Hematol*. 2012;40:994-1004.
3. Yusuf HR, Hooper WC, Beckman MG, et al. Risk of venous thromboembolism among hospitalizations of adults with selected autoimmune diseases. *J Thromb Thrombolysis*. 2014;38:306-313.
4. Segel GB, Lichtman MA. Direct antiglobulin ("Coombs") test-negative autoimmune hemolytic anemia: a review. *Blood Cells Mol Dis*. 2014;52:152-160.
5. Swiecicki PL, Hegerova LT, Gertz MA. Cold agglutinin disease. *Blood*. 2013;122:1114-1121.
6. El Kenz H, Efira A, Le PQ, et al. Transfusion support of autoimmune hemolytic anemia: how could the blood group genotyping help? *Transl Res*. 2014;163:36-42.
7. Maung SW, Leahy M, O'Leary HM, et al. A multi-centre retrospective study of rituximab use in the treatment of relapsed or resistant warm autoimmune haemolytic anaemia. *Br J Haematol*. 2013;163:118-122.
8. Garratty G. Immune hemolytic anemia caused by drugs. *Expert Opin Drug Saf*. 2012;11:635-642.
9. Parker CJ. Paroxysmal nocturnal hemoglobinuria. *Curr Opin Hematol*. 2012;19:141-148.
10. Hill A, Kelly RJ, Hillmen P. Thrombosis in paroxysmal nocturnal hemoglobinuria. *Blood*. 2013;121:4985-4996.
11. Nishimura J, Yamamoto M, Hayashi S, et al. Genetic variant in C5 and poor response to eculizumab. *N Engl J Med*. 2014;370:632-639.
12. Risitano AM, Ricklin D, Huang Y, et al. Peptide inhibitors of C3 activation as a novel strategy of complement inhibition for the treatment of paroxysmal nocturnal hemoglobinuria. *Blood*. 2014;123:2094-2101.

REVIEW QUESTIONS

1. A 68-year-old man is admitted to the intensive care unit because of acute respiratory failure secondary to clinically diagnosed community-acquired pneumonia. His temperature on admission is 41.2° C. Chest radiograph shows diffuse bilateral infiltrates. He is intubated and placed on a respirator, cultured, promptly started on broad-spectrum antibiotics, and placed on a cooling blanket. Intravenous hydration is begun. On the second day, the patient's condition deteriorates. Acrocyanosis develops in his fingertips and toes, the sclerae are noted to be icteric, and the urine from his Foley catheter has turned red (testing 4+ positive for heme but with only a small number of red blood cells). The admission hemoglobin of 14.2 has dropped to 7.8. What is the most likely cause of the acute anemia?

A. Gross hematuria possibly secondary to ureteral trauma or a previously undiagnosed genitourinary malignancy
B. Fulminant alveolar hemorrhage
C. Warm antibody-mediated autoimmune hemolytic anemia most likely secondary to previously undiagnosed, widely metastatic lung cancer
D. Drug-induced autoimmune hemolytic anemia, possibly caused by one of the antibiotics started on admission, with intravascular hemolysis
E. Acute cold antibody-mediated autoimmune hemolytic anemia and agglutination secondary to *Mycoplasma pneumoniae* infection

Answer: E Although unusual, community-acquired *M. pneumoniae* infection can present with bilateral pulmonary infiltrates and acute respiratory failure. *M. pneumoniae* infection (as well as certain viral infections) are associated with cold-reactive autoantibody-mediated cold agglutinin and simultaneously hemolytic disease. This patient's abrupt onset of massive intravascular hemolysis could have been triggered by artificially cooling his body temperature. The patient's intravascular cold autoimmune hemolytic anemia (AIHA) is manifested by hemoglobinuria and his acute intravascular cold agglutinin disease by the development of acrocyanosis. The patient does not have gross hematuria; he has gross hemoglobinuria, so genitourinary tract disease cannot be the cause. Fulminant alveolar hemorrhage could cause a precipitous decline in hemoglobin but is very unlikely in the absence of bloody airway secretions via his endotracheal tube. Warm antibody-mediated AIHA secondary to cancer would not be expected to begin suddenly on hospitalization and would not explain the acrocyanosis. It would be too early for drug-induced hemolysis to begin in this way on the second day of admission.

2. Which of the following statements about chronic cold agglutinin disease is incorrect?

A. It occurs mainly in young women, sometimes with other autoimmune disorders.
B. It is mediated by IgM autoantibody.
C. The direct antiglobulin (Coombs) test result is positive for only complement on the red blood cell surface.
D. It is associated with neoplastic B-cell lymphoproliferative disorders in most cases.
E. Splenectomy not effective in its treatment.

Answer: A Chronic cold agglutinin disease occurs in the great majority of cases in adults older than the age of 50 years. It is generally mediated by a complement-fixing monoclonal IgM autoantibody directed at the red blood cell (RBC) I antigen. However, when a direct antiglobulin test (DAT) is performed, the result is typically positive only for complement that remains fixed to the RBC membrane, and the associated IgM is readily eluted at room temperature before it can be detected in vitro. The great majority of patients with chronic cold agglutinin disease are now recognized to have an underlying B-cell lymphoproliferative neoplasm, such as non-Hodgkin lymphoma or chronic lymphocytic leukemia. The liver, not the spleen, is the major site of extravascular hemolysis in this disease, so splenectomy is not effective treatment.

3. A 45-year-old woman who has been treated in the past for presently clinically inactive systemic lupus erythematosus (SLE) presents with a recent onset of fatigue, lightheadedness, and yellowness of her eyes. Her hemoglobin is found to be 6.3 g/dL (with a baseline of about 12.5-13.0), reticulocytes are 15%, indirect bilirubin is 2.8 IU/L, and LDH is 840 mg/dL. A peripheral blood smear shows a large number of polychromatic red cells and spherocytes. Her direct antiglobulin test (Coombs test) result is positive for IgG. Her treatment at this time should include

A. plasmapheresis.
B. rituximab plus danazol.
C. intravenous methylprednisolone.
D. urgent splenectomy.
E. restarting her previous SLE treatment with azathioprine.

Answer: C This patient has severe warm antibody-mediated autoimmune hemolytic anemia (AIHA), which is probably related to her history of SLE. It appears to be sufficiently acute and severe to require transfusions of packed red blood cells, although the blood bank will likely have difficulty with typing and crossmatching. (The latter problem should not deter giving transfusions in a case like this in consultation with the transfusion medicine specialist on call, often having to resort to transfusing with not completely matched but the "best-matched" units of blood available.) High-dose corticosteroids are the mainstay of initial therapy. Rituximab, danazol, and splenectomy are all second-line treatment options. Plasmapheresis is usually ineffective because at least 50% of an individual's IgG (in this case, IgG autoantibody) is distributed into the extravascular space at any given time. Although azathioprine may have been effective in controlling this patient's other manifestations of SLE in the past, it is not good initial therapy for AIHA.

161 HEMOLYTIC ANEMIAS: RED BLOOD CELL MEMBRANE AND METABOLIC DEFECTS

PATRICK G. GALLAGHER

The mature erythrocyte differs from all other cells in the body. Lacking a nucleus, DNA, RNA, and ribosomes, it cannot synthesize RNA, DNA, or protein. It does not divide, it has no mitochondria, it cannot perform the Krebs cycle, and it lacks an electron transport system for oxidative phosphorylation. After enucleation, the reticulocyte, the precursor of the mature erythrocyte, leaves the marrow and enters the circulation equipped with a full complement of enzymes, transporters, signaling molecules, and all other proteins necessary to perform the essential functions of the red blood cell (RBC) during its lifespan.

The erythrocyte membrane accounts for only about 1% of the total weight of an RBC, yet it plays a critical role in the maintenance of normal RBC homeostasis through a number of mechanisms. These include retention of vital compounds and removal of metabolic waste, regulation of erythrocyte metabolism and pH, and import of iron required for hemoglobin (Hb) synthesis during erythropoiesis. The membrane maintains a slippery exterior so that erythrocytes do not aggregate or adhere to endothelial cells. The membrane skeleton, a network of proteins on the inner surface of the RBC, provides the strength and flexibility needed to maintain the normal shape and deformability of the erythrocyte.

The principal functions of erythrocyte metabolism in the mature erythrocyte include maintenance of adequate supplies of adenosine triphosphate (ATP), production of reducing substances to act as antioxidants, and control of oxygen affinity of Hb by production of adequate amounts of 2,3-diphosphoglycerate (2,3-DPG). Because the mature erythrocyte has lost its ability to perform oxidative phosphorylation, its energy is supplied by anaerobic glycolysis though the Embden-Meyerhof pathway, by oxidative glycolysis through the hexose monophosphate (HMP) shunt, and through nucleotide salvage pathways.

THE ERYTHROCYTE MEMBRANE

Composed of a lipid bilayer and an underlying cortical membrane skeleton (Fig. 161-1), the membrane provides the erythrocyte the deformability and stability required to withstand its travels through the circulation. In one circulatory cycle throughout the body, an erythrocyte is subjected to high sheer stress in the arterial system, dramatic size and shape changes in the microcirculation with capillary diameters as small as 7.5 μm and marked fluctuations in tonicity, pH, and Po_2.

Membrane Lipids

Red blood cell membrane lipids are asymmetrically distributed across the bilayer membrane, reflecting a steady state involving a constant exchange of phospholipids between the two bilayer hemileaflets. Glycolipids and cholesterol are intercalated between the phospholipids in the bilayer with their long axes perpendicular to the bilayer plane. Glycolipids, located in the external half of the bilayer with their carbohydrate moieties extending into the aqueous phase, carry several important RBC antigens and serve other important functions. Phospholipids are asymmetrically organized, with the choline phospholipids, phosphatidylcholine and sphingomyelin, primarily in the outer half of the bilayer, and the amino phospholipids, phosphatidylethanolamine and phosphatidylserine (PS), in the inner half of the bilayer. In pathologic states, such as thalassemia, sickle cell disease, and diabetes, loss of phospholipid asymmetry with externalization of PS leads to activation of blood clotting through conversion of prothrombin to thrombin and facilitates macrophage attachment to erythrocytes, marking them for destruction. Mature erythrocytes are unable to synthesize fatty acids, phospholipids, or cholesterol de novo and depend on lipid exchange and fatty acid acylation as mechanisms for phospholipid repair and renewal.

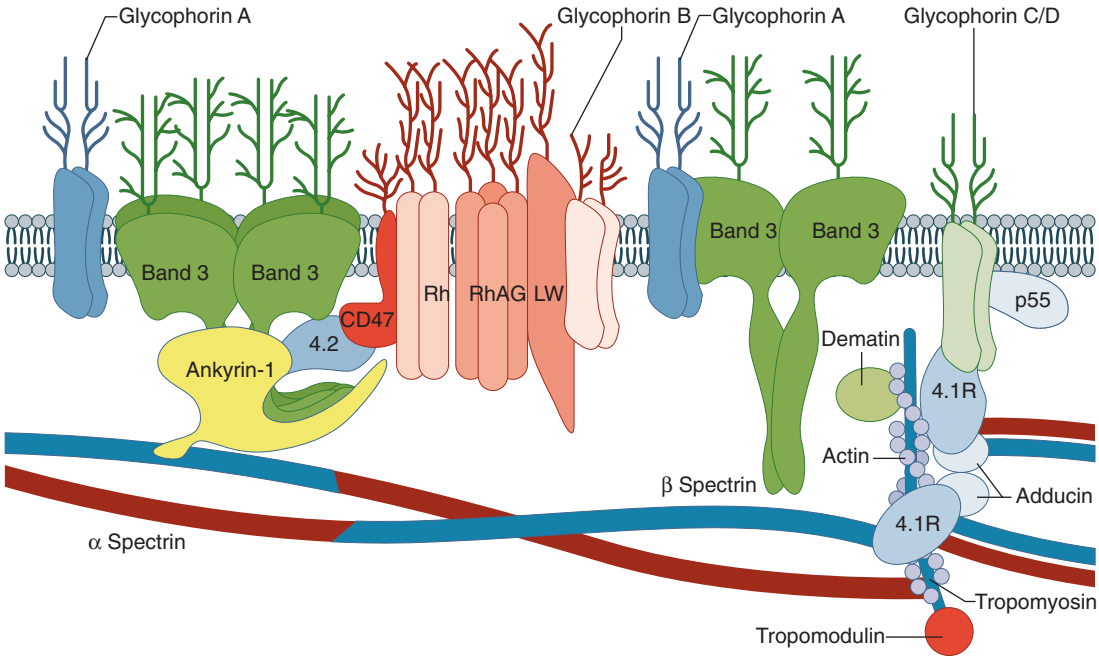

FIGURE 161-1. The erythrocyte membrane. A model of the major proteins of the erythrocyte membrane is shown: α and β spectrin, ankyrin, band 3 (the anion exchanger), 4.1 (protein 4.1) and 4.2 (protein 4.2), actin, and glycophorin. (From Perrotta S, Gallagher PG, Mohandas N. Hereditary spherocytosis. *Lancet.* 2008;372:1411-1426.)

Membrane Proteins

Membrane proteins are classified as *integral*, penetrating or crossing the lipid bilayer and interacting with the hydrophobic lipid core, or *peripheral*, interacting with integral proteins or lipids at the membrane surface but not penetrating into the bilayer core. Integral membrane proteins include the glycophorins, the Rh proteins, Kell and Duffy antigens, and transport proteins such as band 3 (AE1, anion exchanger 1, SLC4A1), Na^+,K^+-ATPase, Ca^{2+}-ATPase, and Mg^{2+}-ATPase. Numerous membrane receptors and antigens are present on integral membrane proteins. Peripheral membrane proteins are on the cytoplasmic membrane face and include enzymes such as glyceraldehyde-3-phosphate dehydrogenase and the structural proteins of the spectrin-actin–based membrane skeleton.

Integral Membrane Proteins

Band 3, the major integral protein of the RBC, has two primary functions, ion transport and maintenance of protein–protein interactions. Band 3 mediates chloride–bicarbonate exchange and provides a binding site for glycolytic enzymes, Hb, and the skeletal proteins ankyrin, protein 4.1, and protein 4.2. A single *N*-glycan chain attached to an Asn in the membrane spanning domain of band 3 is composed of *N*-acetyl-D-lactosamine units arranged in an unbranched, linear fashion in fetal erythrocytes (i antigen) and in a branched fashion in adult cells (I antigen).

The glycophorins are the next most abundant family of integral membrane proteins. They provide most of the negative surface charge required by RBCs to avoid sticking to each other and to the vascular wall. They are involved in transmembrane signaling and carry receptors for *Plasmodium falciparum*, a number of viruses and bacteria, and several blood group antigens.

Peripheral Membrane Proteins

Spectrin is the major component of the membrane skeleton. It is composed of two subunits, α and β spectrin, that are structurally related but functionally distinct. Spectrin is highly flexible and assumes a variety of conformations, an unusual property that may be critical for normal membrane pliancy. The spectrin-based membrane skeleton is linked to the plasma membrane through the actin–protein 4.1 junctional complex; through spectrin–ankyrin interactions; and through binding of a multiprotein complex containing Rh proteins, Rh-associated glycoproteins, CD47, LW, glycophorin B, and protein 4.2 to ankyrin. Protein 4.1, a protein necessary for normal membrane stability, interacts with spectrin, actin, and other proteins of the RBC membrane. Ankyrin serves as the primary linkage protein for the high-affinity binding of spectrin to the inner membrane through interactions with the cytoplasmic domain of band 3. Protein 4.2 is a peripheral membrane protein that helps link the skeleton to the lipid bilayer through interactions with ankyrin and band 3.

Erythrocyte membrane disorders result from alterations in the quantity or quality (or both) of individual proteins and their dynamic interactions with each other. Disruption of the vertical protein–protein interactions of the membrane, that is, the spectrin-ankyrin–band 3 linkage or the band 3–protein 4.2 interaction, leads to uncoupling of the membrane skeleton from the lipid bilayer. This leads to membrane instability with loss of lipids and some integral membrane proteins, resulting in loss of membrane surface area and the phenotype of spherocytosis. Disruption of the horizontal interactions of membrane skeleton proteins, including perturbation of spectrin self-association or junctional complex protein–protein interactions, leads to membrane instability, altered membrane deformability and mechanical properties, and the phenotype of elliptocytosis.

DISORDERS OF THE ERYTHROCYTE MEMBRANE

Hemolytic anemias caused by defects in the erythrocyte membrane comprise an important group of hereditary anemias. Hereditary spherocytosis (HS), hereditary elliptocytosis (HE), and hereditary pyropoikilocytosis (HPP) are the most common disorders among this group.[1] Detailed clinical studies carried out years ago have now been complemented by biochemical and genetic studies, providing both a better understanding of the pathogenesis of these disorders and a better understanding of the normal biology of the erythrocyte membrane.

Hereditary Spherocytosis

DEFINITION

Hereditary spherocytosis is a group of disorders characterized by spherical erythrocytes on the peripheral blood smear. Clinical, laboratory, and genetic heterogeneity characterize this group of disorders.

EPIDEMIOLOGY

Hereditary spherocytosis affects approximately one in 2000 to 3000 individuals of northern European ancestry. Found worldwide, it is much more common in whites than individuals of African ancestry.

PATHOBIOLOGY

The primary defect in HS is the loss of erythrocyte membrane surface area caused by defects in erythrocyte membrane proteins, including α spectrin, β spectrin, ankyrin, band 3, and protein 4.2. Qualitative or quantitative defects of one or more of these membrane proteins lead to membrane instability, which, in turn, leads to membrane loss. In approximately two thirds of HS patients, inheritance is autosomal dominant. In the remaining patients, inheritance is nondominant owing to a de novo mutation or autosomal recessive inheritance. Cases with autosomal recessive inheritance are caused by defects in either α spectrin or protein 4.2. Rare cases of homozygous HS have been reported, resulting in fetal death or severe hemolytic anemia. In most cases, HS mutations are "private," that is, each individual has a unique mutation, implying that there is no selective advantage to HS.

The spleen plays a critical, albeit secondary, role in the pathophysiology of HS. Splenic destruction of poorly deformable spherocytes is the primary cause of hemolysis experienced by HS patients. Abnormal erythrocytes are trapped in the splenic microcirculation and ingested by phagocytes. Moreover, the splenic environment is hostile to erythrocytes, with low pH, low glucose, and low ATP concentrations and high local concentrations of toxic free radicals produced by adjacent phagocytes, all contributing to membrane damage.

CLINICAL MANIFESTATIONS

The clinical manifestations of the spherocytosis syndromes vary widely.[2] The classic triad of HS is anemia, jaundice, and splenomegaly. Rarely, patients may have severe hemolytic anemia presenting in utero or shortly after birth and continuing through the first year of life. These patients may require multiple blood transfusions, and in some cases, splenectomy in the first year of life. Many patients with HS escape detection throughout childhood. In these patients, the diagnosis of HS may not be made until they are being evaluated for unrelated disorders later in life or when complications related to anemia or chronic hemolysis occur. Although the lifespan of an erythrocyte in these patients may be shortened to only 20 to 30 days, they adequately compensate for their hemolysis with increased bone marrow erythropoiesis.

Chronic hemolysis leads to the formation of bilirubinate gallstones, the most frequently reported complication in patients with HS. Although gallstones have been observed in early childhood, most appear in adolescents and young adults. Routine interval ultrasonography to detect gallstones should be performed even if patients are asymptomatic.

Other complications of HS include aplastic, hemolytic, and megaloblastic crises. Aplastic crises occur after virally induced bone marrow suppression and present with anemia, jaundice, fever, and vomiting. The most common etiologic agent in these cases is parvovirus B19 (Chapter 371). Hemolytic crises, usually associated with viral illnesses and occurring before 6 years of age, are generally mild and present with jaundice, increased spleen size, and a decrease in hematocrit. Megaloblastic crises occur in HS patients with increased folate demands, such as the pregnant patient, growing children, or patients recovering from an aplastic crisis.

Uncommon manifestations of HS include skin ulceration, gout, chronic leg dermatitis, cardiomyopathy, spinal cord dysfunction, movement disorders, and extramedullary erythropoiesis. In patients with untreated severe HS, poor growth and findings attributable to extramedullary hematopoiesis, such as hand and skull deformities, may be found.

DIAGNOSIS

Patients with HS may present at any age, usually with anemia, hyperbilirubinemia, or an abnormal blood smear. In evaluating a patient with suspected HS, particular attention should be paid to the family history, including questions about anemia, jaundice, gallstones, and splenectomy. The initial laboratory investigation should include a complete blood count with a peripheral smear, reticulocyte count, direct antiglobulin test (Coombs test), and serum bilirubin. When the peripheral smear or family history is suggestive of HS, an incubated osmotic fragility test or flow cytometric analysis of eosin-5-maleimide–labeled erythrocytes (EMA binding) (discussed later) should be obtained. Rarely, additional, specialized testing is required to confirm the diagnosis.

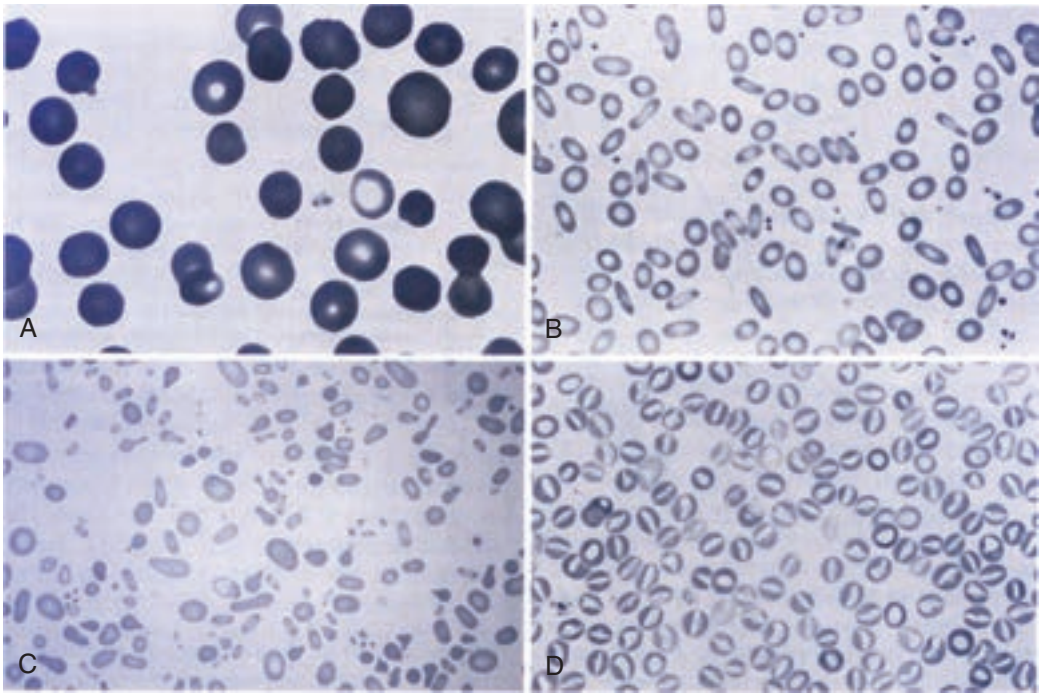

FIGURE 161-2. Peripheral blood smears in disorders of erythrocyte shape. A, Hereditary spherocytosis. Characteristic spherocytes lacking central pallor are seen. **B,** Hereditary elliptocytosis. Smooth, cigar-shaped elliptocytes are seen. **C,** Hereditary pyropoikilocytosis. Pronounced microcytosis, poikilocytosis, fragmentation of erythrocytes, and elliptocytes are seen. **D,** Hereditary stomatocytosis.

Overall, laboratory findings in HS are heterogeneous. Erythrocyte morphology is distinctive but not diagnostic (Fig. 161-2, *A*). Typical HS patients have blood smears with easily identifiable spherocytes lacking central pallor. Some patients present with only a few spherocytes on peripheral smear, but others present with numerous small, dense spherocytes and bizarre erythrocyte morphology. Specific morphologic findings have been identified in patients with certain membrane protein defects such as pincered erythrocytes (band 3) or spherocytic acanthocytes (β spectrin). When examining a smear in a case of suspected spherocytosis, it is important to have a high-quality smear with the erythrocytes well separated and some cells with central pallor in the field of examination because spherocytes are a common artifact on peripheral blood smears. The presence of spherocytosis on peripheral blood smear is not diagnostic of HS. Other disorders with spherocytes on peripheral blood smear are listed in Table 161-1.

The mean corpuscular hemoglobin concentration (MCHC) is increased (between 34.5 and 38) owing to relative cellular dehydration.[3] The mean corpuscular volume (MCV) is usually normal or slightly decreased.[4] Many cell counters provide a histogram of MCHCs claimed to be accurate enough to identify nearly all patients with HS.

In a normal erythrocyte, a redundancy of cell membrane gives the cell its characteristic discoid shape and provides it abundant surface area. In spherocytes, there is a decrease in surface area relative to cell volume, resulting in their abnormal shape. This change is reflected in the increased osmotic fragility found in these cells. Osmotic fragility is tested by adding increasingly hypotonic concentrations of saline to RBCs. Normal erythrocytes are able to increase their volume by swelling, but spherocytes, which are already at maximal volume for surface area, burst at higher saline concentrations than normal. Approximately one fourth of HS individuals will have a normal osmotic fragility on freshly drawn RBCs, with the osmotic fragility curve approximating the number of spherocytes seen on peripheral smear. However, after incubation at 37° C for 24 hours, HS RBCs lose membrane surface area more readily than normal because their membranes have become leaky and unstable. Thus, incubation accentuates the defect in HS erythrocytes and brings out the defect on osmotic fragility, making incubated osmotic fragility the standard test in diagnosing HS (Fig. 161-3, *bottom panel*). When the spleen is present, a subpopulation of very fragile erythrocytes that have been conditioned by the spleen form the tail of the osmotic fragility curve. This tail disappears after splenectomy. The osmotic fragility test suffers from poor sensitivity, with as many as 20% of mild cases of HS missed after incubation. It is unreliable in patients who have small numbers of spherocytes and in patients who have been recently transfused. It is abnormal in other conditions in which spherocytes are present.

TABLE 161-1 DISORDERS WITH SPHEROCYTES ON PERIPHERAL BLOOD FILM

Hereditary spherocytosis
Autoimmune hemolytic anemia
Thermal injuries
Microangiopathic and macroangiopathic hemolytic anemias
Hepatic disease
Clostridial septicemia
Transfusion reactions with hemolysis
Poisoning with certain snake, spider, and *Hymenoptera* venoms
Severe hypophosphatemia
Heinz body anemias
ABO incompatibility (neonates)

Eosin-5-maleimide binding is a flow cytometry–based test used in the diagnosis of HS.[5] EMA is a fluorescent dye that binds to band 3 and Rh-related proteins in the erythrocyte membrane. In HS, the mean fluorescence of EMA-stained erythrocytes is lower compared with control because of the reduction of band 3 and related proteins, typically decreased to approximately 65% of normal (Fig. 161-3, *top panel*). Although primary defects of band 3 protein are seen in only about 25% of HS patients, decreased fluorescence intensity is also observed in the erythrocyte membranes of HS patients with defects in other membrane proteins such as ankyrin and spectrin. This is thought to be attributable to transmission of long-range effects of mutant protein defects across the membrane lattice, ultimately influencing the amount of EMA binding to band 3. EMA binding has good sensitivity and specificity and is simple and rapidly performed.

Specialized testing is available for studying difficult cases or cases in which additional information is desired. Useful tests for these purposes include structural and functional studies of erythrocyte membrane proteins, such as protein quantitation, limited tryptic digestion of spectrin, spectrin, and ion transport. Membrane frigidity and fragility may be examined using an ektacytometer. Complementary DNA and genomic DNA analyses are available when a molecular diagnosis is desired.

Other laboratory manifestations in HS are manifestations of ongoing hemolysis. Increased serum bilirubin, increased lactate dehydrogenase,

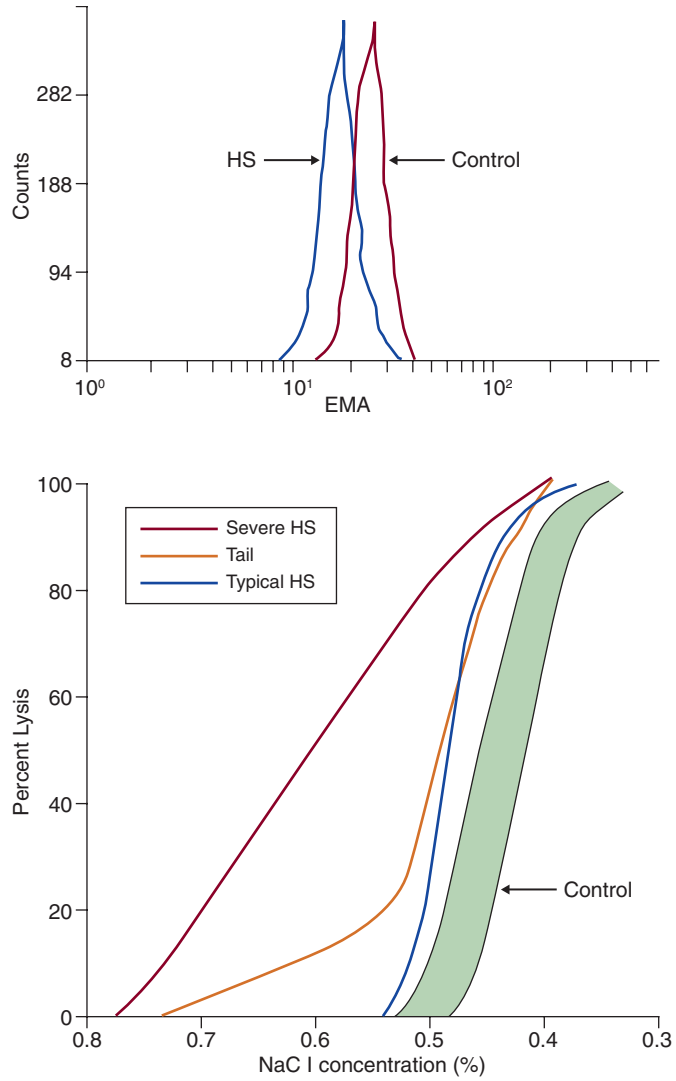

FIGURE 161-3. Testing in hereditary spherocytosis. ***Top panel,*** **Eosin-5-maleimide (EMA) binding. Histogram of fluorescence of EMA-labeled erythrocytes from a normal control and a patient with typical hereditary spherocytosis. Decreased fluorescence in observed from HS erythrocytes.** ***Bottom panel,*** **Osmotic fragility curves in hereditary spherocytosis. The *shaded region* is the normal range. Results representative of both typical, and severe spherocytosis are shown. A tail, representing fragile erythrocytes conditioned by the spleen, is common in spherocytosis patients prior to splenectomy. (From Gallagher PG. Abnormalities of the erythrocyte membrane. *Pediatr Clin North Am* 2013;60:1349-1352.)**

increased urinary and fecal urobilinogen, and decreased serum haptoglobin reflect increased erythrocyte destruction.

After diagnosing a patient with HS, family members should be examined for the presence of HS. This can be of great epidemiologic importance, particularly for very old and very young patients. Prenatal diagnosis of HS has been made in a few cases, but this is rarely necessary.

TREATMENT AND PROGNOSIS

Rx

Splenic sequestration and destruction is the primary determinant of erythrocyte survival in HS patients. Splenectomy cures or alleviates anemia in most patients, reducing or eliminating the need for transfusions. The risk for cholelithiasis is also decreased to nearly background levels. After splenectomy, spherocytes remain in the peripheral blood, but their lifespan becomes near normal.

In the past, splenectomy was routinely performed in all HS patients. However, the risk of overwhelming postsplenectomy infection; the emergence of penicillin-resistant pneumococci; and the growing recognition of the increased risk of postsplenectomy cardiovascular disease, particularly thrombosis and pulmonary hypertension, have led to reevaluation of the role of splenectomy in the treatment of HS. In addition, with growing globalization, the important role of the spleen in protection of individuals living in or traveling to geographic regions where parasitic diseases such as malaria or babesiosis occur has reemerged. When splenectomy is considered, health care providers, the patient, and family members must review and weigh the benefits of splenectomy against the immediate and long-term risks of the procedure. Considering the risks and benefits, a reasonable approach is to splenectomize all patients with severe spherocytosis and all patients who have significant signs or symptoms of anemia, including growth failure, skeletal changes, leg ulcers, and extramedullary hematopoietic tumors. Other candidates for splenectomy are older HS patients who have vascular compromise of vital organs. Whether patients with moderate HS and compensated, asymptomatic anemia should undergo splenectomy is controversial.[6]

When splenectomy is indicated, laparoscopic splenectomy has become the method of choice. This technique results in less postoperative discomfort, a quicker return to preoperative diet and activities, shorter hospitalization, decreased costs, and smaller scars. Even massive spleens can be removed laparoscopically because the spleen is placed in a large bag, diced intraoperatively, and eliminated through suction catheters. Partial splenectomy, initially advocated for infants and young children with significant anemia associated with erythrocyte membrane disorders to allow for palliation of hemolysis and anemia while maintaining some residual splenic immune function, is now being suggested by some for most HS patients. Updated UK guidelines,[7] which reflect changes in current opinion about surgical management, include (1) preference for a laparoscopic approach, (2) performance of splenectomy ideally after the age of 6 years, (3) no indication for extended thrombosis prophylaxis after splenectomy for HS, and (4) avoidance of splenectomy in patients with some forms of hereditary stomatocytosis because of an increased risk of venous thromboembolism.

Before splenectomy, patients should be immunized with vaccines against pneumococcus, *Haemophilus influenzae* type B, and meningococcus. Postsplenectomy care includes counseling of patients or parents to seek prompt medical care in case of febrile illness. Use of routine antibiotics after splenectomy for prevention of pneumococcal sepsis is controversial. Data are lacking to indicate or refute their prescription. Before splenectomy and, in severe cases, after splenectomy, HS patients should take folic acid (1 mg/day orally) to prevent folate deficiency.

Hereditary Elliptocytosis and Related Disorders

DEFINITION

Hereditary elliptocytosis is characterized by the presence of elliptical or oval cigar-shaped erythrocytes on peripheral blood smears of affected individuals (see Fig. 161-2, *B*).

EPIDEMIOLOGY

Hereditary elliptocytosis has been estimated to occur in approximately one in 2000 to 4000 individuals. The true incidence of HE is unknown because its clinical severity is heterogeneous, and many patients are asymptomatic. It is common in African Americans and people of Mediterranean ancestry, presumably because elliptocytes confer some resistance to malaria. In parts of Africa, the incidence of HE approaches one in 100.

PATHOBIOLOGY

The principal defect in HE is mechanical weakness or fragility of the erythrocyte membrane skeleton. Qualitative and quantitative defects in a number of RBC membrane proteins have been described in HE, including α spectrin, β spectrin, protein 4.1, and glycophorin C. Most defects occur in spectrin, the principal structural protein of the erythrocyte membrane skeleton. αβ Spectrin heterodimers self-associate into tetramers and higher order oligomers that are critical for erythrocyte membrane stability as well as erythrocyte shape and function. Most spectrin defects in HE impair the ability of spectrin dimers to self-associate into tetramers and oligomers, thereby disrupting the membrane skeleton. Structural and functional defects of protein 4.1 appear to disrupt the spectrin–actin contact in the membrane skeleton. Glycophorin C variants are also deficient in protein 4.1. The precise pathobiology of how elliptocytes are formed in these syndromes is unclear.

Genetically, HE is heterogeneous with multiple genetic loci. A wide variety of mutations have been described in the α spectrin, β spectrin, protein 4.1, and glycophorin C genes, including point mutations, gene deletions and insertions, and messenger RNA processing defects. Several mutations have been identified in a number of individuals of the same genetic background, suggesting a "founder effect" for these mutants, which supports the

hypothesis that there has been genetic selection for elliptocytosis because these RBCs confer some resistance to malaria. Most cases of HE are inherited in an autosomal dominant pattern, with rare cases of de novo mutations.

CLINICAL MANIFESTATIONS

The clinical presentation of HE is heterogeneous, ranging from asymptomatic carriers to patients with severe, life-threatening anemia. Most patients with HE are asymptomatic and are diagnosed incidentally during testing for unrelated conditions. Asymptomatic carriers have been identified who possess the same molecular defect as an affected HE relative but who have normal peripheral blood smears. The erythrocyte lifespan, normal in most patients, is decreased in only about 10% of patients. This subset of HE patients with decreased erythrocyte lifespan experience hemolysis, anemia, splenomegaly, and intermittent jaundice. Many of these patients have parents with typical HE and thus are homozygotes or compound heterozygotes for defects inherited from each of the parents. Symptoms may vary among members of the same family, indeed, they may vary in the same individual at different times.

Hereditary Pyropoikilocytosis

Hereditary pyropoikilocytosis is a rare cause of anemia with distinctive erythrocyte morphology on peripheral blood smear (see Fig. 161-2, *C*) and has a picture similar to that seen in patients with severe burns. Patients typically present in infancy with severe anemia and peripheral blood smear findings of elliptocytosis, poikilocytosis, pyknocytosis, and fragmentation. Microspherocytosis is common, and the MCV is usually very low (50-70 fL). Most patients are of African ancestry, and at least one third of HPP patients have a parent or sibling with typical HE. Patients with HPP tend to experience severe hemolysis and anemia in infancy that gradually improves, evolving toward typical HE later in life.

DIAGNOSIS

Cigar-shaped elliptocytes on peripheral blood smear are the hallmark of HE (see Fig. 161-2, *B*). These normochromic, normocytic elliptocytes vary in number from a few to 100%, with the likelihood of hemolysis not correlating with the number of elliptocytes present. Ovalocytes, spherocytes, stomatocytes, and fragmented cells may also be seen. In some cases, pyknocytes may be prominent. Elliptocytes may be seen in association with other disorders, including megaloblastic anemias, hypochromic microcytic anemias (iron deficiency anemia and thalassemia), myelodysplastic syndromes, and myelofibrosis; however, elliptocytes generally make up less fewer one third of RBCs in these conditions. History and additional laboratory testing usually clarify the diagnosis of these disorders. In typical cases, the incubated osmotic fragility is normal, but in severe HE and HPP, incubated osmotic fragility is increased, and EMA binding is decreased.

Other laboratory findings in HE are similar to those found in other hemolytic anemias and are nonspecific markers of increased erythrocyte production and destruction. The reticulocyte count generally is less than 5% but may be higher when hemolysis is severe.

Similar to HS, specialized laboratory procedures are available to study the erythrocyte membranes of HE and HPP patients. These studies are not routinely required to make the diagnosis of HE or HPP, but they may be helpful in studying problematic cases and in elucidating the underlying molecular defects.

TREATMENT

Rx

Therapy is rarely needed in patients with HE. In rare cases, occasional RBCs transfusions may be required. In cases of severe HE and HPP, splenectomy has been palliative because the spleen is the site of erythrocyte sequestration and destruction. Many practitioners think that the same indications for splenectomy in HS should be applied to patients with symptomatic HE or HPP. Post-splenectomy patients with HE or HPP experience increased hematocrits, decreased reticulocyte counts, and improvement in clinical symptoms.

Patients should be followed for signs of decompensation during acute illnesses. Interval ultrasonography to detect gallstones should be performed. In patients with significant hemolysis, folate should be administered daily.

Hereditary Stomatocytosis Syndromes

Red blood cell hydration is primarily determined by the intracellular concentration of monovalent cations. A net increase in sodium and potassium ions causes water to enter, forming *stomatocytes* (see Fig. 161-2, *D*) or *hydrocytes*, but a net loss of sodium and potassium produces dehydrated RBCs, or *xerocytes*. Numerous descriptions of congenital or familial hemolytic anemias associated with abnormal cation permeability and, in some cases, disturbed RBC hydration have been reported.[8] These span the range from severe hydrocytosis to severe xerocytosis. In many cases, the molecular bases of this group of disorders are unknown. An unusual characteristic of the stomatocytosis syndromes is a predisposition to thrombosis after splenectomy. Acquired stomatocytosis has been associated with acute alcoholism and hepatobiliary disease, vinca alkaloid administration, neoplasms, and cardiovascular disease. Stomatocytosis is also sometimes observed as a processing artifact.

OVERHYDRATED HEREDITARY STOMATOCYTOSIS (HYDROCYTOSIS)

This group of disorders is characterized by stomatocytes, erythrocytes with a mouth-shaped (stoma) area of central pallor on peripheral blood smear (see Fig. 161-2, *D*), severe hemolysis, macrocytosis (110-150 fL), elevated erythrocyte sodium concentration, reduced potassium concentration, and increased total Na^+ and K^+ content. The excess cations expand cell water, producing large, osmotically fragile cells with low MCHCs (24%-30%). The clinical severity of overhydrated hereditary stomatocytosis is variable; some patients experience hemolysis and anemia, but others are asymptomatic. Missense mutations in the Rh-associated glycoprotein (RhAG) have been identified in a subset of hydrocytosis patients.

DEHYDRATED HEREDITARY STOMATOCYTOSIS (XEROCYTOSIS)

Blood smears from patients with dehydrated hereditary stomatocytosis exhibit contracted and spiculated RBCs, dessicytes, a variable number of stomatocytes, and target cells. Most patients have nearly normal erythrocyte morphology, with only a few target cells and an occasional echinocyte or stomatocyte. The MCV (95-115 fL) and MCHC are increased, and the osmotic fragility is reduced (i.e., resistance to osmotic lysis). The characteristic biochemical abnormality is a decreased potassium concentration and total monovalent cation content. Dominantly inherited mutations in PIEZO1, encoded by the *FAM38A* gene, have been identified in xerocytosis patients. PIEZO proteins are the recently identified pore-forming subunits of channels that mediate mechanotransduction in mammalian cells. Association of PIEZO variants with changes in erythrocyte hydration suggest that these proteins play an important role in erythrocyte volume homeostasis.

INTERMEDIATE SYNDROMES AND HEREDITARY STOMATOCYTOSIS VARIANTS

Hydrocytosis and xerocytosis represent the extremes of a spectrum of RBC permeability defects. A number of families with features of both conditions have been reported. Some patients with severe permeability defects have little or no hemolysis. The proportion of stomatocytes and the degree of sodium influx do not correlate with each other, and neither correlates with the amount of hemolysis or anemia.

ERYTHROCYTE METABOLISM

The primary functions of the erythrocyte, gas transport and exchange, are maintained without a net change in energy state. However, several critical functions of the erythrocyte depend on the production and expenditure of energy. As erythrocytes age, glucose utilization and ATP levels fall, leading to decreased membrane deformability and, ultimately, a shortened lifespan. Lower potassium levels, higher sodium levels, and decreased membrane lipids are also seen in ATP-deficient, aging erythrocytes.

Erythrocytes do not undergo oxidative phosphorylation and do not store glycogen; thus, they must constantly catabolize glucose from the blood stream through the Embden-Meyerhof pathway and the HMP shunt as a source of energy (Fig. 161-4). Erythrocytes incorporate glucose from the plasma through facilitated transfer, with erythrocyte glucose levels rapidly equilibrating with changes in blood glucose levels. Glucose is the preferred carbohydrate of the RBC, but fructose and mannose are metabolized almost as readily. Inside the erythrocyte, glucose is converted to glucose-6-phosphate or to fructose by sorbitol. Glucose-6-phosphate follows one of three pathways: (1) most (~90%) enters the Embden-Meyerhof pathway, where it is converted into lactate, pyruvate, and ATP; (2) some (~5%-10%) enters the HMP shunt to produce reduced intermediates and ribulose 5-phosphate, the latter of which eventually enters the Embden-Meyerhof pathway; and (3) a tiny fraction (<1%) is converted to glucose-1-phosphate and then to glycogen.

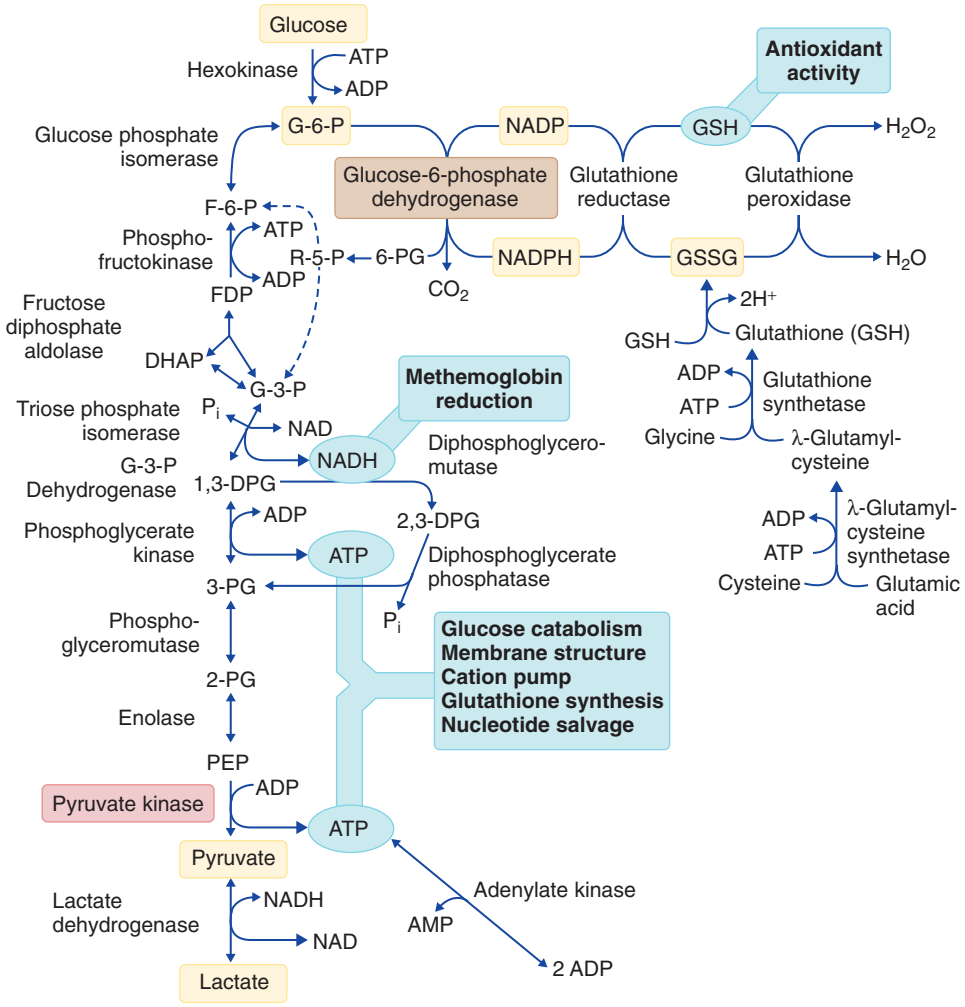

FIGURE 161-4. Pathways of energy metabolism in the erythrocyte. Glucose-6-phosphate may be degraded anaerobically to lactate through the Embden-Meyerhof pathway or oxidatively through the hexose monophosphate shunt. Pentose phosphates (R-5-P) can reenter anaerobic glycolysis as fructose-6-phosphate (F-6-P) and glyceraldehyde-3-phosphate (G-3-P) after conversion by enzymes of the terminal pentose phosphate pathway or as a product of adenosine or inosine degradation. 2,3-Diphosphoglycerate (2,3-DPG) may be generated instead of adenosine triphosphate (ATP) through diversion of triose through the Rapoport-Luebering shunt. Glutathione may be synthesized directly from constituent amino acids; its cycling from oxidized (GSSG) to reduced forms (GSH) depends on reduced pyridine cofactor (NADPH) generation. ADP = adenosine diphosphate; DHAP = dihydroxyacetone phosphate; FDP = fructose-1,6-diphosphate; NAD = nicotinamide adenine dinucleotide; NADP = nicotinamide adenine dinucleotide phosphate; NADPH = nicotinamide adenine dinucleotide phosphate, reduced form; PEP = polyestradiol phosphate.

Embden-Meyerhof Pathway

The Embden-Meyerhof pathway of glycolysis is the primary source of ATP, 2,3-DPG and nicotinamide adenine dinucleotide, reduced form (NADH) in erythrocytes (see Fig. 161-4). Most of the energy generated by erythrocytes is through the Embden-Meyerhof pathway followed by storage as high-energy phosphates such as ATP or as reducing energy in the form of glutathione or pyridine nucleotides (NADH and nicotinamide adenine dinucleotide phosphate, reduced form [NADPH]). This pathway metabolizes about 90% of erythrocyte glucose with the catabolism of 1 mole of glucose yielding 2 moles of ATP and 2 moles of lactate. Two moles of ATP per mole of metabolized glucose seems insignificant compared with the Krebs cycle of intermediary metabolism, in which 1 mole of glucose metabolized produces 38 moles of ATP. However, this ATP production is adequate to renew 150% to 200% of the total RBC ATP every hour.

The Embden-Meyerhof pathway is also the primary source of NADH, a necessary cofactor for NADH methemoglobin reductase, which maintains heme iron in the reduced state. Without this reaction, heme iron would be oxidized to methemoglobin, which is not a functional oxygen transporter.

Finally, the Rapoport-Luebering shunt of the Embden-Meyerhof pathway (see Fig. 161-4) produces 2,3-DPG, a compound found in high concentrations in erythrocytes but in low concentrations in other cells. After it is formed, under physiologic conditions of pH and solute concentrations, 2,3-DPG binds reversibly to tetramers of deoxyhemoglobin with greater affinity than it does to oxyhemoglobin. By binding to deoxyhemoglobin, it allosterically upregulates the release of the remaining oxygen bound to the Hb, enhancing the ability of erythrocytes to release oxygen near tissues that need it most.

Hexose Monophosphate Shunt (Pentose Phosphate Pathway)

In the HMP shunt (see Fig. 161-4), glucose-6-phosphate undergoes oxidation followed by a series of reactions to yield fructose-6-phosphate and glyceraldehyde-3-phosphate, intermediates in the glycolytic pathway. The HMP shunt is the primary source of erythrocyte NADPH, with 2 moles of NADPH produced for each mole of glucose metabolized. NADPH is required for the reduction of oxidized glutathione and some protein sulfhydryl groups.

Mature erythrocytes synthesize large amounts of reduced glutathione (GSH). GSH protects erythrocytes from oxidants, including hydrogen peroxide (H_2O_2), superoxide anions (O_2^-), and hydroxyl radicals (OH), which are produced as byproducts of the oxidation of heme by oxygen. Oxidants are also produced by activated phagocytes (e.g., during infection) and by erythrocytes after exposure to certain agents. When oxidants accumulate, they damage cellular proteins and lipids. Detoxification of H_2O_2 is significantly enhanced by glutathione peroxidase. GSH is converted to oxidized glutathione (GSSG) and to mixed disulfides with protein thiols. GSH levels are restored by glutathione reductase. In this process, NADPH is oxidized to nicotinamide adenine dinucleotide phosphate (NADP), which stimulates the HMP shunt to regenerate NADPH.[9] After oxidant stress, hypoxia, or acidosis, erythrocytes can increase the amount of glucose metabolized through the HMP shunt up to 10- to 20-fold to generate

increased amounts of reduced glutathione. The tight coupling of glutathione metabolism with the HMP shunt protects the mature erythrocyte from oxidative stress.

DISORDERS OF ERYTHROCYTE METABOLISM

Congenital nonspherocytic hemolytic anemia (CNSHA) traditionally includes erythrocyte disorders not due to defects of the RBC membrane or Hb, immune-mediated disease, or other diseases such as paroxysmal nocturnal hemoglobinuria. CNSHA is a heterogeneous group of disorders associated with various metabolic abnormalities of the erythrocytes, including enzymopathies of glucose, glutathione, and nucleotide metabolism. Similar to the membrane disorders, clinical, biochemical, and genetic heterogeneity are typical within the enzymopathies. Hemolysis may develop as a result of either enzyme or antioxidant deficiency or dysfunction (e.g., abnormal substrate or cofactor binding), altered activation or inhibition characteristics, or decreased stability or specific activity.

Peripheral blood smears in CNSHA, with the exception of pyrimidine 5′-nucleotidase (P5N) deficiency, are unremarkable. Osmotic fragility of fresh erythrocytes is normal. Response to splenectomy is variable. Inheritance is heterogeneous. A thorough family history is important and may be of assistance in determining the diagnosis. Manifestations of the metabolic defect are usually confined to the erythrocyte but may occasionally involve nonerythroid cells.

Definitive diagnosis of metabolic abnormalities of the RBCs depends on qualitative or quantitative assays of specific enzyme activity or identification of the specific genetic mutation by DNA analysis. Results of enzyme assays should be interpreted with caution because (1) they only sample surviving RBCs in the peripheral blood, and the metabolic milieu of these cells is not necessarily comparable to cells already hemolyzed; (2) in vitro enzyme assay conditions may not accurately reflect the in vivo environment; (3) transfusions before the assay may obscure the underlying metabolic defect, and (4) leukocyte contamination may lead to spurious results. Finally, average enzyme activity may not accurately reflect activity in subpopulations of erythrocytes. This is particularly true when there is reticulocytosis, which may yield artificially elevated mean enzyme activity owing to higher enzyme levels found in reticulocytes.

Disorders of the Embden-Meyerhof Pathway

Defects of the Embden-Meyerhof pathway are inherited in an autosomal recessive fashion, and usually hemolysis is seen only in homozygotes or compound heterozygotes.[10] Heterozygotes, whose erythrocytes contain less than normal amounts of mutant enzyme, are clinically normal.

An exception is phosphoglycerate kinase deficiency, an X-linked disorder with hemolysis found only in males. In this group of disorders, hemolysis is chronic, is not typically influenced by drugs or other inciting agents, and is attributed to insufficient levels of erythrocyte ATP. Splenomegaly from trapping of mutant erythrocytes is common. The hostile splenic environment contributes to the shortened erythrocyte lifespan. When performing specific diagnostic enzyme assays, measurement of glycolytic intermediates may assist in diagnosis because concentrations of intermediates are increased upstream of a defect and decreased downstream of a defect.

PYRUVATE KINASE DEFICIENCY

Pyruvate kinase (PK) deficiency accounts for approximately 90% of inherited defects of the Embden-Meyerhof pathway and is the second most common inherited erythrocyte enzymopathy associated with anemia after glucose-6-phosphate dehydrogenase (G6PD) deficiency (see later). PK deficiency is found worldwide, but it is most common in individuals of northern European descent.

PATHOBIOLOGY

Pyruvate kinase catalyzes the conversion of phosphoenolpyruvate (PEP) to pyruvate, generating ATP. Deficient or defective PK leads to decreased levels of erythrocyte ATP, disturbing many cellular processes such as signaling and maintenance of water and ion content, leading to energy failure and dehydration. Upstream catabolites accumulate in the erythrocyte, including 2,3-DPG, which shifts the oxygen dissociation curve to the right, enhancing tissue oxygenation and ameliorating some of the physiologic effects of anemia. Early PK-deficient reticulocytes retain the ability to use oxidative phosphorylation to produce ATP, bypassing their defect. This ability is lost as reticulocytes mature and is markedly dampened in the hypoxic environment of the spleen.

Pyruvate kinase deficiency is inherited in an autosomal recessive manner. Affected individuals are homozygous or compound heterozygotes for PK defects. Heterozygotes are clinically normal or exhibit very minimal hemolysis.

CLINICAL MANIFESTATIONS

Clinical manifestations in PK deficiency are heterogeneous, ranging from asymptomatic to transfusion-dependent hemolytic anemia.[11] More severely affected patients present in infancy or early childhood with anemia, jaundice, and splenomegaly. Occasionally, patients may escape detection until later in life when complications related to anemia and chronic hemolysis occur such as cholelithiasis or aplastic crisis or when the diagnosis is made during evaluation of the patient for another condition.

DIAGNOSIS

The peripheral blood smear demonstrates normocytic, normochromic erythrocytes, sometimes with spiculations (Fig. 161-5, *A*). Poikilocytes and acanthocytes may also be seen. Reticulocytosis is common. Osmotic fragility of fresh erythrocytes is usually normal. Occasional patients exhibit a population of osmotically fragile cells after incubation.

NADH fluorescence under ultraviolet light is a commonly used screening test for PK deficiency. PEP and NADH are mixed with the patient's blood, incubated, and spotted on filter paper, and fluorescence is measured. Direct enzyme assay, which uses PEP as substrate for PK, can be performed on leukocyte-free hemolysate to confirm abnormal fluorescence tests. Leukocytes must be carefully depleted from the samples because they contain more than 300 times the PK activity of erythrocytes.

TREATMENT

Rx

Most patients require only expectant management, with only rare transfusions, such as during an aplastic episode. In severe cases, patients may be transfusion dependent. In these cases, splenectomy typically lessens hemolysis and ameliorates the anemia. After splenectomy, some patients develop marked reticulocytosis, up to 50% to 70%. This paradoxical reticulocytosis is attributed to increased reticulocyte survival after removal of the hostile splenic environment.

OTHER DISORDERS OF THE EMBDEN-MEYERHOF PATHWAY

Other abnormalities of the Embden-Meyerhof pathway have been described. Hexokinase deficiency is quite uncommon, with great phenotypic variability in reported cases. Severely affected patients have had anemia beginning in infancy and may require blood transfusions. Glucose phosphate isomerase (GPI) deficiency is the third most common hemolytic enzymopathy. GPI deficiency usually presents in infancy or early childhood with moderate to severe hemolytic anemia. Rare cases of GPI deficiency may also be complicated by neurologic symptomatology. Phosphofructokinase deficiency may involve erythrocytes, muscle, or both. The presentation is usually in adolescence with exertional myopathy (Chapter 207). Hemolytic anemia has been described in isolated cases of 2,3-bisphosphoglycerate mutase deficiency and phosphoglycerate kinase deficiency.

Disorders of Nucleotide Metabolism

Mature erythrocytes lack the ability to synthesize purine and pyrimidine nucleotides de novo. However, they are able to form some nucleotides through salvage pathways.

PYRIMIDINE 5′-NUCLEOTIDASE DEFICIENCY

Pyrimidine 5′-nucleotidase degrades the pyrimidine nucleotides of RNA to cytidine and uridine, which can diffuse out of the cell. When P5N is deficient, nondiffusible, partially degraded RNAs accumulate, leading to the marked basophilic stippling characteristic of P5N-deficient erythrocytes (see Fig. 161-5, *B*). These accumulated pyrimidine nucleotides inhibit the transport of GSSG (oxidized glutathione) out of RBCs, leading to high levels of erythrocyte glutathione. Clinically, the patient has mild to moderate hemolytic anemia and splenomegaly. The cause of the hemolysis remains cryptic. Typically, splenectomy does not ameliorate the hemolysis and anemia.

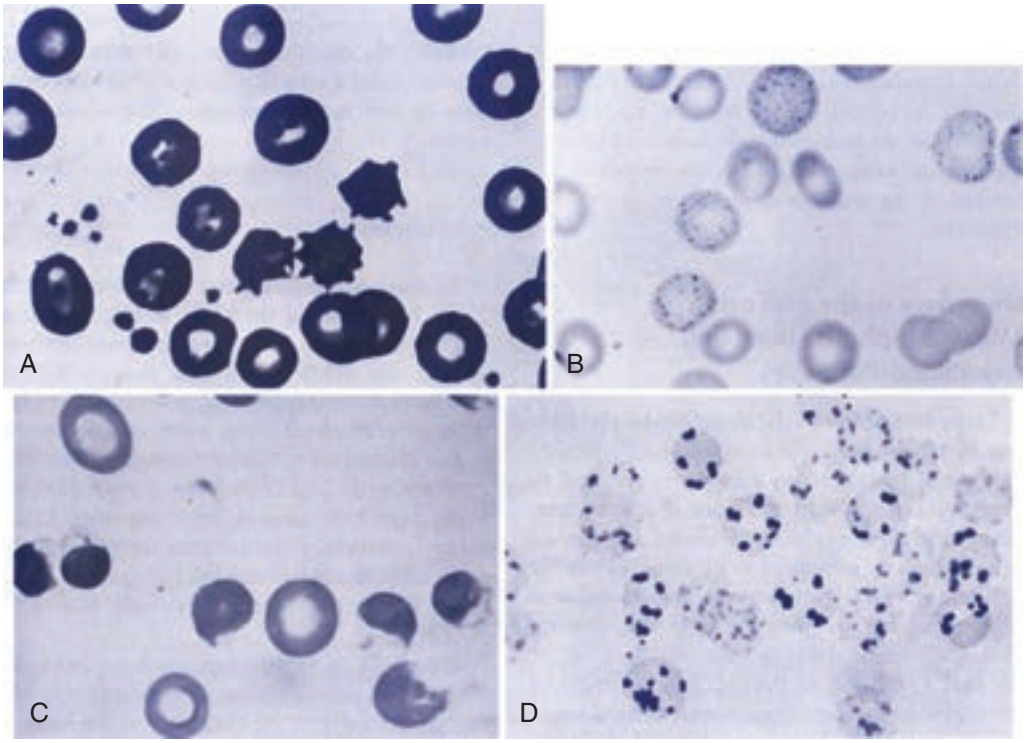

FIGURE 161-5, Peripheral blood smears in erythrocyte enzymopathies. A, Pyruvate kinase deficiency. **B,** Pyrimidine 5′-nucleotidase deficiency; **C,** Glucose-6-phosphate dehydrogenase (G6PD) deficiency. **D,** Heinz bodies in G6PD deficiency. (**B** from Paglia DE. Disorders of erythrocyte glycolysis and nucleotide metabolism. In: Handin RI, Lux SE, Stossel TP, eds. *Blood: Principles and Practice of Hematology*. Philadelphia: JB Lippincott; 1995:1877-1896.)

Disorders of the Hexose Monophosphate Shunt (Pentose Phosphate Pathway) and Associated Pathways

Disorders of the HMP shunt or of the glutathione metabolic pathways (see Fig. 161-4) compromise the ability of the RBC to respond adequately to oxidative stress. In normal erythrocytes, GSH detoxifies oxidants produced by various agents and infection. In G6PD-deficient erythrocytes, because of the inability to generate NADPH, GSH levels are inadequate, leaving the cell susceptible to oxidant stress. Oxidation of Hb sulfhydryl groups leads to the production of methemoglobin and intracellular Hb precipitates called *Heinz bodies*. Heinz bodies (see Fig. 161-5, *D*), usually visualized on peripheral blood smears with supravital stains such as methyl violet, attach to and damage the erythrocyte membrane. They induce clustering of immunoglobulins and band 3 protein, marking the erythrocyte for opsonization by phagocytes and eventual removal from the circulation. Heinz bodies are "pitted" from circulating cells by the spleen and are commonly seen on smears of patients after splenectomy. "Bite cells," erythrocytes with localized invaginations, possibly at the site of Heinz body injury or removal, are seen during acute hemolytic episodes. In addition to damage from Heinz body formation, GSH-deficient erythrocytes undergo peroxidation of membrane phospholipids and oxidative cross-linking of spectrin, decreasing membrane deformability and further promoting splenic trapping.

GLUCOSE-6-PHOSPHATE DEHYDROGENASE DEFICIENCY

G6PD deficiency is the most common inherited disorder of erythrocyte metabolism, affecting more than 400 million people worldwide.[12] The high prevalence of G6PD deficiency is thought to be attributable to genetic selection because G6PD-deficient erythrocytes have a selective advantage against invasion by the malaria parasite *Plasmodium falciparum*.

EPIDEMIOLOGY AND PATHOBIOLOGY

G6PD is the initial and rate-limiting step in the HMP shunt (see Fig. 161-4), which converts NADP into NADPH. NADPH is required for the generation of glutathione, a critical constituent in the prevention of oxidative damage to the cell. G6PD-deficient patients may develop acute hemolytic anemia after exposure to oxidative stress. Although G6PD is a ubiquitous enzyme, erythroid cells are particularly susceptible to oxidative stress because the HMP shunt is their only source of NADPH.

Hundreds of G6PD variants have been described, but only a few are common. Variants are classified on the basis of biochemical characteristics; electrophoretic mobility; ability to use substrate analogue, Km for NADP and G6PD; pH activity profile; and thermal stability. The normal enzyme, Gd^{B}, is present in 99% of white Americans and 70% of African Americans. A normal variant, Gd^{A+}, found in 20% of African Americans, has a faster electrophoretic mobility than Gd^{B}. Gd^{A-}, the most common variant associated with hemolysis, is found in about 10% of African Americans and in many Africans. Gd^{A-} has decreased catalytic ability compared with Gd^{A+}. Gd^{Med}, the second most common variant associated with hemolysis, is common in the Mediterranean area, in India, and in Southeast Asia, with a prevalence of up to 5% to 50%. Gd^{Med} exhibits markedly decreased catalytic activity. Gd^{Canton}, a variant common in Asian populations, produces a clinical syndrome similar to Gd^{A-}.[13,14]

Gd^{B} activity decreases as normal cells age, with a half-life of approximately 60 days. Despite very low levels of or no active G6PD, older erythrocytes maintain the ability to produce NADPH and maintain a GSH response to oxidative stress. The Gd^{A-} variant has a half-life of only 13 days, so young cells have a normal amount of enzyme activity, but older RBCs are grossly deficient. Because of this heterogeneity in G6PD levels, individuals with the Gd^{A-} variant experience only limited hemolysis after oxidant exposure.

More than 100 mutations in the *G6PD* gene, localized to Xq28, have been described. Most mutations are amino acid substitutions that influence enzyme kinetics, stability, or both, with a few rare deletions and splicing mutations described. Because it is X-linked, G6PD deficiency primarily affects males. Males have only one G6PD allele and express only one G6PD type. Females can express one or two G6PD types. The Lyon hypothesis specifies that only one X chromosome is active in any given cell; thus, any given cell in a heterozygous female is either normal or deficient. In females who are heterozygous for G6PD deficiency, average G6PD activity may be normal or mildly, moderately, or severely reduced, depending on the degree of lyonization. G6PD-deficient erythrocytes in heterozygous females are susceptible to the same oxidant stress as G6PD-deficient cells in males, but, typically, the overall degree of hemolysis is less because there is a smaller population of vulnerable cells.

CLINICAL MANIFESTATIONS

G6PD deficiency is divided into five classes based on clinical severity and degree of enzyme deficiency. Class I is characterized by CNSHA without precipitating cause and severe G6PD deficiency. Class II is characterized by intermittent hemolysis and severe G6PD deficiency. Class III is characterized by hemolysis after oxidant stress and mild G6PD deficiency. Class II and III

together represent more than 90% of G6PD variants. Classes IV and V are clinically asymptomatic. The most clinically significant syndromes of G6PD deficiency are acute hemolytic anemia (AHA); neonatal jaundice (NNJ); and rarely, CNSHA.

Acute hemolytic anemia is the most dramatic clinical presentation of G6PD deficiency with acute intravascular hemolysis after exposure to an oxidative stress.[15] Oxidative stresses include ingestion of certain drugs such as primaquine or sulfa-containing compounds, exposure to naphthalene (mothballs), ingestion of fava beans, or infection, the latter being the most common cause of hemolysis. Table 161-2 lists drugs that should be avoided in G6PD-deficient patients. Presenting symptoms include irritability, fever, nausea, abdominal pain, and diarrhea within 48 hours of oxidant exposure. Hemoglobinuria, jaundice, and anemia ensue. The spleen and liver may be enlarged and tender. Cases with severe anemia may precipitate congestive heart failure. Laboratory findings include a normochromic, normocytic anemia with anisocytosis and reticulocytosis. Poikilocytes and bite cells may be seen. Heinz bodies, a classic finding in G6PD deficiency, may be seen but are an inconsistent finding because these damaged cells are rapidly cleared from the circulation in the spleen. Additional laboratory findings may include hemoglobinuria and the presence of free Hb in the blood.

Another clinically significant syndrome of G6PD deficiency is NNJ. Jaundice is seldom present at birth, with the peak incidence of onset between days 2 and 3 of life. The severity of hyperbilirubinemia is variable. It may be severe, resulting in kernicterus or even death. In most cases, however, hyperbilirubinemia is adequately treated with phototherapy. In NNJ, it is important to note that the anemia is very rarely severe. The etiology of NNJ remains controversial. NNJ is increased in G6PD-deficient infants who also carry a polymorphism of the uridine diphosphoglucuronyl transferase (*UDPGT1*) gene associated with Gilbert syndrome.

Chronic nonspherocytic hemolytic anemia is associated with uncommon variants of G6PD deficiency, usually mutant enzymes unable to maintain basal NADPH production. Presentation may be in the neonatal period when NNJ is accompanied by anemia in a male. The degree of chronic anemia in CNSHA caused by G6PD deficiency has been variable. Some patients have compensated hemolysis, but others require intermittent transfusions. Transfusion dependence occurs in the most severe cases.

TABLE 161-2 AGENTS TO BE AVOIDED BY GLUCOSE-6-PHOSPHATE DEHYDROGENASE–DEFICIENT PATIENTS*

ANTIMALARIALS
Primaquine (people with the African A⁻ variant may take it at reduced dosage, under surveillance)
Pamaquine
Chloroquine (may be used under surveillance when required for prophylaxis or treatment of malaria)
SULFONAMIDES AND SULFONES
Sulfanilamide
Sulfapyridine
Sulfadimidine
Sulfacetamide (Albucid)
Acetyl sulfisoxazole (Gantrisin)
Salicylazosulfapyridine (Salazopyrin)
Dapsone
Sulfoxone
Glucosulfone sodium (Promin)
Sulfamethoxazole-trimethoprim (Septrin)
OTHER ANTIBACTERIAL COMPOUNDS
Nitrofurans—nitrofurantoin, furazolidone, nitrofurazone
[Nalidixic acid]
Chloramphenicol
p-Aminosalicylic acid
ANALGESICS
Acetylsalicylic acid (aspirin): moderate doses can be used
Acetophenetidin (Phenacetin)
Safe alternative: Paracetamol
ANTHELMINTICS
β-Naphthol
Stibophen
Niridazole
MISCELLANEOUS
Vitamin K analogues (1 mg of menaphthone can be given to babies)
Naphthalene (moth balls)
Probenecid
Dimercaprol (BAL)
Methylene blue
Arsine†
Phenylhydrazine†
Acetylphenylhydrazine†
Toluidine blue
Mepacrine

*Drugs in bold print should be avoided by people with all forms of glucose-6-phosphate dehydrogenase (G6PD) deficiency. Drugs in normal print should be avoided, in addition, by G6PD-deficient people of Mediterranean, Middle Eastern, and Asian origin. Items in normal print and within square brackets apply only to people with the African A⁻ variant.
†These drugs or chemicals may cause hemolysis in normal people if given in large doses. Many other drugs may produce hemolysis in certain individuals.

DIAGNOSIS

The G6PD reaction (glucose-6-phosphate + $NADP^+$ → 6-phosphogluconolactone + NADPH + H^+) reduces $NADP^+$ to NADPH. Formation of NADPH and NADH can be observed directly because they fluoresce in the visible spectrum when illuminated with long-wave ultraviolet light. Based on this observation, several simple screening tests performed using inexpensive long-wave ultraviolet light have been devised. These tests are semiquantitative, categorizing a sample as normal or deficient. They are unreliable after an acute hemolytic episode and do not typically detect female heterozygotes. Positive screening test results should be confirmed by spectrophotometric assay or DNA studies.

Definitive assay of the enzyme depends on direct spectrophotometric measurement of NADPH production. Although more sensitive than screening tests, this still requires 20% to 30% G6PD-deficient cells to obtain an abnormal result. Sensitivity can be increased by comparing the level of G6PD deficiency with levels of other age-dependent erythrocyte enzymes, especially when testing is temporally in close proximity to an acute hemolytic episode. The cyanide-ascorbate test measures the ability of erythrocytes to prevent the oxidation of Hb by ascorbate. Using intact erythrocytes, as few as 10% to 15% deficient cells can be detected, making this test useful for detecting female heterozygotes and males after a hemolytic episode. This test also detects other perturbations of the HMP shunt or glutathione metabolism.

TREATMENT

Rx

The best treatment for an individual with AHA is careful prescription of medications and avoidance of inciting agents (see Table 161-2). Outside of acute hemolytic episodes, these patients do not require any special therapy. AHA episodes are managed with particular attention to hematologic, cardiopulmonary, and renal complications of hemolysis. Management of NNJ does not differ from that recommended for other causes of neonatal hyperbilirubinemia. In CNSHA, management is expectant. Exposure to oxidant stresses should be avoided. Blood transfusions may be necessary during acute hemolytic episodes. In severe cases of CNSHA, splenectomy may ameliorate the anemia.

Disorders of Glutathione Metabolism

Defects of glutathione metabolism may be associated with hemolysis. Erythrocytes from patients lacking glutathione synthetase or γ-glutamylcysteine synthetase, enzymes involved in glutathione synthesis, have very low levels of GSH. Clinically, these disorders resemble G6PD deficiency. There is mild to moderate chronic hemolytic anemia with increased susceptibility to oxidant stress.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Da Costa L, Galimand J, Fenneteau O, et al. Hereditary spherocytosis, elliptocytosis, and other red cell membrane disorders. *Blood Rev*. 2013;27:167-178.
2. Gallagher PG. Abnormalities of the erythrocyte membrane. *Pediatr Clin North Am*. 2013;60: 1349-1362.
3. Brugnara C, Mohandas N. Red cell indices in classification and treatment of anemias: from M.M. Wintrobes's original 1934 classification to the third millennium. *Curr Opin Hematol*. 2013;20: 222-230.
4. Liao L, Deng ZF, Qiu YL, et al. Values of mean cell volume and mean sphered cell volume can differentiate hereditary spherocytosis and thalassemia. *Hematology*. 2014;19:393-396.
5. King MJ, Zanella A. Hereditary red cell membrane disorders and laboratory diagnostic testing. *Int J Lab Hematol*. 2013;35:237-243.
6. Casale M, Perrotta S. Splenectomy for hereditary spherocytosis: complete, partial or not at all? *Expert Rev Hematol*. 2011;4:627-635.
7. Bolton-Maggs PH, Langer JC, Iolascon A, et al. Guidelines for the diagnosis and management of hereditary spherocytosis—2011 update. *Br J Haematol*. 2012;156:37-49.
8. Gallagher PG. Disorders of red cell volume regulation. *Curr Opin Hematol*. 2013;20: 201-207.
9. Manganelli G, Masullo U, Passarelli S, et al. Glucose-6-phosphate dehydrogenase deficiency: disadvantages and possible benefits. *Cardiovasc Hematol Disord Drug Targets*. 2013;13:73-82.
10. Climent F, Roset F, Repiso A, et al. Red cell glycolytic enzyme disorders caused by mutations: an update. *Cardiovasc Hematol Disord Drug Targets*. 2009;9:95-106.
11. Rider NL, Strauss KA, Brown K, et al. Erythrocyte pyruvate kinase deficiency in an old-order Amish cohort: longitudinal risk and disease management. *Am J Hematol*. 2011;86:827-834.
12. Luzzatto L, Seneca E. G6PD deficiency: a classic example of pharmacogenetics with on-going clinical implications. *Br J Haematol*. 2014;164:469-480.
13. Howes RE, Dewi M, Piel FB, et al. Spatial distribution of G6PD deficiency variants across malaria-endemic regions. *Malar J*. 2013;12:418.
14. Howes RE, Battle KE, Satyagraha AW, et al. G6PD deficiency: global distribution, genetic variants and primaquine therapy. *Adv Parasitol*. 2013;81:133-201.
15. Monteiro WM, Franca GP, Melo GC, et al. Clinical complications of G6PD deficiency in Latin American and Caribbean populations: systematic review and implications for malaria elimination programmes. *Malar J*. 2014;13:70.

REVIEW QUESTIONS

1. A man recently diagnosed with glucose-6-phosphate dehydrogenase (G6PD) deficiency after an episode of hemolysis after taking chloroquine malaria prophylaxis before visiting Benin. He has many questions, including about starting a family. At the end of the visit, you provide several suggestions. These include all of the following except

A. always inform health care providers that he is G6PD deficient, particularly before they prescribe any medication.
B. seek medical attention if he is feeling tired, short of breath, and experiencing palpitations and if dark-colored urine is observed.
C. he should avoid eating fava beans and fava plant–derived food.
D. if his wife does not have G6PD deficiency, all sons will be unaffected, and all daughters will be carriers.
E. if his wife is a G6PD carrier, one of two daughters will be G6PD deficient, one out of two daughters will be G6PD carriers, and his sons will not be affected.

Answer: E Glucose-6-phosphate dehydrogenase (G6PD) deficiency is inherited in an X-linked manner. Thus, if his wife does not have G6PD deficiency, all sons will be unaffected, and all daughters will be carriers (answer D). However, if his wife is a G6PD carrier, one of two daughters will be G6PD deficient, one of two daughters will be G6PD carriers, one son of two will be G6PD deficient, and one son of two will be unaffected. With the diagnosis of G6PD deficiency, he should always tell his health care providers, especially when they are prescribing medications, and he should avoid fava beans and all fava-derived products. Finally, he should seek medical attention when experiencing the symptoms of a hemolytic reaction, tiredness, shortness of breath, palpitations, and dark-colored urine.

2. You have just diagnosed a 23-year-old patient with hereditary spherocytosis. Which of the following are appropriate actions?

A. Obtain ultrasonography of the spleen and gallbladder.
B. Counsel the patient to avoid oxidant-inducing foods and medications.
C. Determine if other family members could be affected and should be evaluated for the presence of hereditary spherocytosis.
D. Counsel regarding possible occurrence of hemolytic and aplastic crises.
E. A, B, and C
F. A, C, and D
G. All of the above

Answer: F Assessing spleen size and the biliary tract for the presence of cholelithiasis (answer A), determining if other family members could be affected after diagnosing a patient with hereditary spherocytosis (answer C), and counseling regarding hemolytic and aplastic crises (answer D) are all appropriate actions after diagnosing a patient with hereditary spherocytosis. Avoidance of oxidant-inducing food and medications (answer B), which is appropriate in patients with metabolic disease of the erythrocyte, especially glucose-6-phosphate dehydrogenase deficiency, is not necessary in hereditary spherocytosis.

3. You are asked to see a jaundiced 62-year-old man with cirrhotic liver disease complicated by portal hypertension and hepatosplenomegaly because of stomatocytes on peripheral smear. He is being evaluated for splenectomy. His total leukocyte count is 4000/mm^3 and his platelet count is 105,000/m^3. Hemoglobin and hematocrit are 8.4 g/dL and 26%, respectively. Peripheral blood smear shows decreased numbers of leukocytes and platelets. Erythrocyte morphology reveals nearly 100% stomatocytes, erythrocytes with a central bar, and a few target cells. Your most appropriate response would be

A. the presence of stomatocytes makes splenectomy an absolute contraindication.
B. obtain erythrocyte electrolyte determinations.
C. prominent stomatocytosis is common in patients with severe liver disease and no additional specific therapy is indicated.
D. obtain cholesterol and triglyceride levels to correlate with severity of stomatocytes and target cells and consider lipid lowering therapy.
E. obtain osmotic fragility or EMA binding test.

Answer: C Prominent stomatocytosis is common in patients with severe liver disease, and no additional specific therapy is indicated. Marked stomatocytosis is very common in advanced liver disease. Target cells may also be seen in advanced liver disease. Splenectomy is contraindicated in cases of hereditary stomatocytosis, not in cases of acquired stomatocytosis (answer A). Evaluation for hereditary stomatocytosis or hereditary spherocytosis via erythrocyte electrolytes (answer B), osmotic fragility or EMA binding (answer E) without additional historical data is not necessary. Although erythrocyte membrane lipids and cholesterol are perturbed in stomatocytosis, specific therapy is not warranted.

4. A 31-year old woman presents with a history of lifelong hemolytic anemia. Her parents are normal, but a sister also has anemia. She tells you she had a splenectomy as a teenager, but "it didn't work." Laboratory evaluation reveals hemoglobin of 10 g/dL, hematocrit of 31%, mean corpuscular volume of 96 fl, and reticulocyte count of 15%. The lactate dehydrogenase level is 750 IU/L. Peripheral blood smear shows evidence of hemolysis; many erythrocytes show prominent basophilic stippling. Her likely diagnosis is

A. pyruvate kinase deficiency.
B. glucose-6-phosphate dehydrogenase deficiency.
C. hereditary spherocytosis.
D. hereditary stomatocytosis.
E. pyrimidine 5′-nucleotidase deficiency.

Answer: E All of the disorders listed may be associated with a history of lifelong anemia. Pyruvate kinase and pyrimidine 5′-nucleotidase (P5N) deficiency are both recessively inherited (normal parents, affected sister.) Splenectomy failure is typical in P5N deficiency and some cases of hereditary stomatocytosis, in which splenectomy is contraindicated. Basophilic stippling of erythrocytes is a characteristic finding in P5N deficiency because of the accumulation of partially degraded, nondiffusable RNAs in the cell.

5. Evaluation of a patient with hemolytic anemia reveals the presence of spherocytes on peripheral blood smear. Which disorder is NOT in the differential diagnosis?

A. Heinz body hemolytic anemia
B. Autoimmune hemolytic anemia
C. Hereditary stomatocytosis
D. Hereditary spherocytosis
E. Liver disease

Answer: C Although the results of osmotic fragility testing may be similar in patients with hereditary stomatocytosis and hereditary spherocytosis, spherocytes are rarely seen on the peripheral blood smear in those with hereditary stomatocytosis. In addition to hereditary spherocytosis, spherocytes may be seen on the peripheral blood smear of patients with Heinz body hemolytic anemia, autoimmune hemolytic anemia, and liver disease.

162 THE THALASSEMIAS

MARIA DOMENICA CAPPELLINI

DEFINITION

The thalassemias, or more comprehensively the thalassemia syndromes, are a heterogeneous group of inherited hemolytic anemias characterized by deficient or absent production of one of the globin chains of hemoglobin. This leads to imbalanced globin chain synthesis that is the hallmark of all the thalassemia syndromes.

EPIDEMIOLOGY

Taken together, the thalassemias are the most common single-gene disorder in the world population, with estimated carrier numbers of more than 270 million, and more than 300,000 children are born each year with one of the thalassemia syndromes or one of the structural hemoglobin variants. The extremely high frequency of the hemoglobin disorders compared with other monogenic diseases reflects natural selection mediated by the relative resistance of carriers against *Plasmodium falciparum* malaria. Other factors that may be involved include the widespread practice of consanguineous marriage, increased maternal age in the poorer countries, and gene drift and founder effects. For these reasons, the thalassemias are most frequent in southeastern and southern Asia, in the Middle East, in the Mediterranean countries, and in northern and central Africa. However, as the result of mass migrations of African populations from high-prevalence areas, thalassemias are now encountered worldwide.

PATHOBIOLOGY

Normal adult red cells contain 97% adult hemoglobin (HbA: $\alpha_2\beta_2$), with approximately 2.5% of the minor component HbA_2 ($\alpha_2\delta_2$) and a small amount of fetal hemoglobin (HbF: $\alpha_2\gamma_2$). Because the stable tetramer $\alpha_2\beta_2$is the major component of hemoglobin after birth, there are two main forms of thalassemia: α-thalassemias and β-thalassemias. Because β-chain synthesis is fully activated only after birth, it follows that the β-thalassemias are not expressed as a disease in intrauterine life; they are manifested as γ-chain synthesis declines during the first year of life. In contrast, because α chains are shared by both fetal and adult hemoglobin, α-thalassemias are manifested in both fetal and adult life.

As knowledge about their genetic basis and pathophysiologic mechanisms has evolved, the thalassemia syndromes can now be classified at genetic and clinical levels (Table 162-1).[1]

TABLE 162-1 GENETIC AND CLINICAL CLASSIFICATIONS OF THE THALASSEMIAS

	GENETIC	CLINICAL
α-Thalassemias	α^0 α^+ Deletion (−α) Nondeletion (α^T)	α-Minor HbH disease Hydrops fetalis
β-Thalassemias	β^0 β^+ Variant with high HbA_2 Normal HbA_2 Silent Dominant Unlinked to β-gene cluster	β-Minor Thalassemia intermedia Thalassemia major
δβ-Thalassemia	$(\delta\beta)^0$ $(\delta\beta)^+$ $(^A\gamma\delta\beta)^0$	δβ-Minor Thalassemia intermedia
HPFH	Deletion Nondeletion Unlinked to β-gene cluster	Silent increase HbF

HbA = adult hemoglobin; HbF = fetal hemoglobin; HbH = hemoglobin H; HPFH = hereditary persistence of fetal hemoglobin.

Genetics

Six different types of globin chains (α,β,γ,δ,ε,ζ) are found in normal human hemoglobin at different stages of development. In the very early embryo, hemoglobin synthesis is restricted to the yolk sac and the production of hemoglobins Gower 1 ($\zeta_2\varepsilon_2$), Gower 2 ($\alpha_2\varepsilon_2$), and Portland ($\zeta_2\gamma_2$). Subsequently, at about 8 weeks of gestation, the fetal liver takes over, synthesizing predominantly HbF ($\alpha_2\gamma_2$) and a small amount (<10%) of HbA. Between about 18 weeks and birth, the liver is progressively replaced by bone marrow as the major site of red cell production; this is accompanied in the later stages of gestation by a reciprocal switch in production of HbF and HbA, which continues until, by the end of the first year, HbF production has dropped to less than 2%. The globin genes are encoded in separate gene clusters. The α cluster (ζ,α_2,α_1) lies at the telomere of chromosome 16; the β cluster ($\varepsilon,^G\gamma$ and $^A\gamma,\delta,\beta$) lies at chromosome 11p15.5. In both clusters, the genes are aligned 5′ to 3′ in the order in which they are expressed during development. Both sets of genes are under the regulation of enhancer-like elements (hypersensitive site [HS]–40 for α cluster and locus control region for β cluster) that lie some distance away at the 5′ end of the cluster. Deletion of these enhancer elements results in inactivation of any related globin gene. There are two α genes/alleles ($\alpha_2$and α_1) that differ by a few nucleotides in intron 2 and the 3′ untranslated region but produce identical protein products. The output of the α_2 gene exceeds that of the α_1 gene by two- to three-fold. The α cluster also contains pseudo-ζ and pseudo-α genes that are not translated into protein products. The region around a DNase1 hypersensitive site at 40 kilobases (kb) upstream of the ζ-globin gene (HS-40) is the major regulator of α-globin gene expression. The β cluster contains a single pseudo-β gene; the βcluster enhancer consists of five elements marked by erythroid-specific DNase1 hypersensitive sites lying 6 to 20 kb upstream of the ε-globin gene, each of which contains several binding sites for erythroid-specific and other transcription factors. All together, these elements are known as the locus control region, and each element contributes to the overall locus control region activity. In addition, there is an erythroid-specific hypersensitive site approximately 20 kb downstream of the β-globin gene; when the cluster is activated, the upstream and downstream hypersensitive sites are brought into proximity with the gene promoters to activate their transcription.

The individual globin genes share many general features; they consist of three coding sequence exons separated by two introns in identical position but of variable length for a total length of approximately 1500 nucleotides. This structure has been highly conserved throughout evolution. The upstream regions flanking the first exon contain a number of sequence motifs that are necessary for specifying correct transcriptional initiation. A TATA box is found at 30 base pairs upstream of the initiation site together with one or more CCAAT sites at 70 base pairs upstream. The gene promoters also contain a CACCC or CCGCCC box that binds erythroid Krüppel-like factor (EKLF) 1, and some have binding sites for erythroid transcription factor GATA-1. In model systems, mutations introduced into such sequences lead to reduction in the level of transcription.[2]

All the thalassemias have a similar pattern of inheritance; in most cases, the gene defects are transmitted in a mendelian autosomal fashion. Thus, the severe, symptomatic varieties usually result from the interaction of more than one genetic determinant. The inheritance of α-thalassemia is more complicated because it involves the products of the linked pairs of α genes (αα) (see Clinical Manifestations).

Molecular Basis of Thalassemias

The α- and β-thalassemias are divided into disorders in which no chains are produced from the affected chromosomes (α^0 and β^0) and those in which the output of the chains is reduced (α^+ and β^+). For the α-thalassemias, the most common molecular defects are deletions of one or both α genes, which are designated −α and − −, respectively. The single α gene is believed to have arisen by crossover between two misaligned α genes on the homologous chromosome that can give rise to chromosomes with either single (−α) or triplicated (ααα) α-globin genes. Depending on the point of crossover, deletions may remove between 2.5 and 5.3 kb of sequence, with the loss of 3.7 ($-\alpha^{3.7}$) or 4.2 ($-\alpha^{4.2}$) kb being the most prevalent. Full duplication of the α-globin gene locus, including the upstream regulatory element, has also been reported in subjects of different ancestry, suggesting that this type of homologous genetic recombination occurs relatively frequently in globin loci. To date, more than 20 different deletions that involve both α genes, resulting from illegitimate or nonhomologous recombination, have been reported. The lengths of deletion vary from 5.2 to more than 40 kb; the most

common are those from Southeast Asia, the Mediterranean, and the Philippines, designated $--^{SEA}$, $--^{MED}$, and $--^{FIL}$, respectively. Nondeletion types of α-thalassemia (α^{α}) are much less common than the deletion forms; in most cases, they result from single oligonucleotide mutations at regions of the α gene sequence that are critical for normal expression. Because expression of the α_2 gene is two to three times greater than that of the α_1 gene, it is not surprising that most of the nondeletion mutants predominantly affect expression of the α_2 gene. The mutations may affect the initiation codon or splicing signals, cause frame shifts, or introduce premature stop codons. At least five single-nucleotide variants affect the natural termination codon (TAA) of the α_2-globin gene. Among these, hemoglobin Constant Spring ($\alpha^{cs}\alpha$) is the most common and extensively studied. Finally, there are several α-globin variants that are so unstable that they undergo rapid, postsynthetic degradation. In such situations, β chains remain in excess within the red cell, and the patient carriers of these α-chain variants, by definition, have α-thalassemia. To date, 17 unstable α variants have been shown to produce the phenotype of α-thalassemia to a greater or lesser extent.[3]

Like the α-thalassemias, the β-thalassemias are classified as β^0 (in which no β-globin is produced) and β^+ (in which some β-globin is produced but less than normal). In some cases, the defects in β-chain production are so mild that they are designated β^{++}. So far, more than 200 different thalassemic mutations of the β-globin gene have been reported; most are point mutations within the gene or its immediate flanking sequence. A few β-thalassemia mutations that segregate independently of the β-globin gene cluster have been described, presumably involving *trans*-acting regulatory factors. The distribution of alleles is highly variable from one population to another, but within each population, there are only a few alleles that are common. The nondeletion forms of β-thalassemia account for most β-thalassemia alleles. An updated list of these mutations is accessible at the Globin Gene Server website (*http://globin.cse.psu.edu*). They include transcriptional mutations, RNA processing mutations, and mutations affecting translation.

Simple deletions of the β-globin gene are rare, ranging in size from 290 base pairs to more than 60 kb. The 619–base pair deletion at the 3′ end of the β gene is relatively common among Sind and Punjabi populations in India and Pakistan. The remaining deletions are restricted to single families, are necessarily β^0-thalassemias, and interestingly are associated with unusually high levels of HbA_2 in heterozygotes. Large deletions that affect the entire β-globin gene cluster $(\varepsilon\gamma\gamma\delta\beta)^0$ are rare and restricted to single families. Finally, some highly unstable β-chain variants may be manifested as a dominant form of β-thalassemia.[4]

The δβ-thalassemias and hereditary persistence of fetal hemoglobin (HPFH) are the result of deletions affecting various parts of the β-globin locus. These deletions are partially compensated by an increased expression of the γ genes that raises the level of HbF. The length of deletion accounts for different forms of δβ-thalassemia, including both $^{G}\gamma$ and $^{A}\gamma$ genes or only $^{A}\gamma$, and varies from 9 to 100 kb. Hemoglobin Lepore is a hybrid of δ and β chains resulting from a crossover between the two misaligned genes; this hemoglobin is synthesized inefficiently and gives rise to a form of δβ-thalassemia. Deletions of δ and β genes are also the molecular basis for many forms of HPFH that, however, usually have higher levels of compensatory HbF production than the δβ-thalassemias. Other HPFHs are due to point mutations in the promoter region upstream from the transcription start site in either the $^{G}\gamma$ or $^{A}\gamma$ genes that alter the binding of one or more transcription factors; they are known as nondeletion HPFHs. Genetic studies have identified three major quantitative trait loci that account for 20 to 50% of the common variation in HbF levels in patients with β-thalassemia and sickle cell disease as well as in healthy adults.

CLINICAL MANIFESTATIONS

The clinical manifestations (phenotype expression) of thalassemia syndromes are extremely variable and depend on the degree of globin chain imbalance.

α-Thalassemias

As previously mentioned, there are two major classes of α-thalassemias: α^0, in which both α genes are inactivated (− −/); and α^+, in which only one of the pair is defective because of α gene deletion or mutation (−α or $\alpha\alpha^T$).[5] The clinical spectrum of α-thalassemias correlates well with the number of the affected α genes, that is, from normal to the loss of all four genes. The inheritance of a normal allele (αα) with one of the α^+ or α^0 alleles most frequently results in α-thalassemia minor (− −/αα; −α/αα; $-\alpha^T/\alpha\alpha$; $\alpha^T\alpha$ /−α;−α/−α). In general, carriers of such genotypes have lower levels of total hemoglobin, mean corpuscular volume, and mean cell hemoglobin but higher red blood cell count than normal. The greatest differences are seen in mean cell hemoglobin, which is usually less than 26 pg. The peripheral blood smear is variable, showing various degrees of hypochromia with some target cells and occasional poikilocytes (Chapter 157). In carriers of α^0-thalassemia (− −/αα), it is possible to generate a few red cell HbH inclusions (β_4). The carriers of nondeletional forms ($\alpha\alpha^T/\alpha\alpha$) show slightly more marked hematologic changes than those for deletional forms. The hemoglobin constitution of adult carriers of α^+- or α^0-thalassemia is indistinguishable from normal but has slightly lower levels of HbA_2. Traces of hemoglobin Bart (γ_4) in the neonatal period are detectable in a large proportion of neonates with α-thalassemia, and they decline during the first 6 months after birth. The α-thalassemias are common in areas where β-thalassemias are also found at a high frequency. Thus, the coinheritance of α- and β-thalassemia trait may occur and even ameliorate the hematologic parameters. In some cases, for genetic counseling in families in which α- and β-thalassemias are present, genotype determination is essential. The unstable mutant Hb^{CT} causes a severe reduction in α_2-globin expression from the affected chromosome; therefore, the carriers and particularly the homozygotes have a more severe phenotype than α-thalassemia minor but not as severe as most cases of HbH disease.

HbH disease most frequently results from the interaction of α^+- and α^0-thalassemia, and not surprisingly most patients originate from the populations of southeastern Asia, the Mediterranean, and the Middle East. HbH disease is a diagnosis attributed to subjects older than 6 months having a sufficient globin imbalance to produce detectable levels of HbH (>1 to 2%) in their peripheral blood together with inclusion bodies (β_4 tetramers) in their red cells. The clinical phenotypes encompass a wide spectrum from mild clinical manifestations to thalassemia intermedia and are included in the so-called non–transfusion-dependent thalassemias.[6] The predominant features of HbH disease are a hypochromic, microcytic anemia with jaundice and hepatosplenomegaly. Because the main mechanism of the anemia is hemolysis rather than dyserythropoiesis, only a few patients have clinical evidence of an expanded erythron. The most common complication of HbH disease is the development of hypersplenism due to severe splenomegaly. Other complications include gallstones, leg ulcers, increased risk of infection, folic acid deficiency, and increased risk of venous thrombosis mainly after splenectomy. Hemoglobin levels range in different series from 3 to 12 g/dL, with fluctuations that may occur after exposure to an oxidant drug, infection, or transient aplasia possibly due to intercurrent viral infection. Rarely, patients with HbH disease require regular blood transfusions. The anemia is associated with reticulocytosis and typical thalassemic changes of the red cell indices. The relative amount of HbH varies from 1 to 40%. The values of HbA_2 are always reduced. The peripheral blood film shows hypochromia with variable anisopoikilocytosis, target cells, and basophilic stippling. The characteristic feature of HbH disease is that it is always possible to generate multiple inclusions in the red cells after incubation with brilliant cresyl blue. The bone marrow shows marked erythroid hyperplasia.[7]

The most severe form of α-thalassemia is hydrops fetalis, in which all four α-globin genes are deleted (genotype − −/− −). It is incompatible with life. In fact, because α-globin chains are absent during gestation, hemoglobin Bart (γ_4) becomes the dominant hemoglobin. Because of its high oxygen affinity, hemoglobin Bart is unable to deliver oxygen to tissues, and the intrauterine consequences are progressive severe anemia, severe ineffective erythropoiesis with marked extramedullary erythropoiesis, massive organomegaly, heart failure, severe hypoalbuminemia, and edema. Infants with hydrops fetalis syndrome die either in utero (30 to 40 weeks of gestation) or soon after birth. The hemoglobin levels range from 3 to 20 g/dL; the peripheral blood film is characterized by marked anisopoikilocytosis, large hypochromic macrocytes, and many nucleated red cells. The hemoglobin consists almost entirely of hemoglobin Bart (80 to 90%), with some remaining HbH and Portland. Mothers of these infants often have a history of previous neonatal deaths. Without medical care, women carrying these fetuses may have delivery and postpartum complications (e.g., retained placenta, eclampsia, sepsis) (Chapter 239).

There are several reports describing α-thalassemia in association with mental retardation (so-called ATR-16 syndrome). These conditions are mainly due to large deletions (one or two megabases) of the tip of chromosome 16 including the α-globin gene cluster. However, several cases have no deletions or other apparent abnormalities of the α-globin gene cluster. It has been shown that these patients with a peculiar phenotype characterized by severe mental retardation, dysmorphic facies, genital abnormalities, and

α-thalassemia have a disorder that maps to the X chromosome (ATR-X syndrome).

β–Thalassemias

The β-thalassemias include a considerably heterogeneous group of disorders of hemoglobin synthesis, all of which are characterized by a reduced output of the β chains of adult hemoglobin.[8] The clinical classification includes thalassemia major (TM, transfusion dependent), thalassemia intermedia (TI, of intermediate severity, non–transfusion dependent), and thalassemia minor (asymptomatic) (Fig. 162-1 and Table 162-1). The severity of the clinical manifestations correlates well with the degree of imbalance of globin chains; depending on the β-globin gene defects and their interaction, the production of β-globin chains is quantitatively reduced to different degrees, whereas the synthesis of α-globin continues as normal, resulting in accumulation of excess unmatched α-globin chains in the erythroid precursors. The free α-globin chains are not able to form stable tetramers; they therefore precipitate in the erythroid precursors, forming inclusion bodies that damage the red cell membrane, thereby causing premature destruction of erythroid precursors in the bone marrow (ineffective erythropoiesis). Ineffective erythropoiesis leads to a sequence of events responsible for bone marrow expansion, anemia, hemolysis, splenomegaly, and increased iron absorption. Any factor that reduces the degree of chain imbalance and the magnitude of α-chain excess, such as coinheritance of α-thalassemia or an innate ability to increase fetal hemoglobin, will ameliorate the clinical expression of the disease (Fig. 162-2).

The major forms of β-thalassemia (still sometimes called Cooley anemia) are disorders in which life can be sustained only by regular blood transfusions. This condition usually results from the homozygous state or from the compound heterozygous state for severe β-gene mutations (β^0). The typical forms of TM become manifested during the first year of life, during which γ chains are switched off but not replaced by β-chain synthesis. These infants, left untreated, are incapable of maintaining a hemoglobin level above 5 g/dL and show marked bone deformities and growth retardation. They develop the "thalassemic" facies due to frontal bossing of the skull and protrusion of the jaws and cheekbones. The magnitude of the increase in erythropoiesis may result in extramedullary masses arising usually from the sternum and ribs. Progressive hepatosplenomegaly is a constant finding leading to pancytopenia. The early childhood of the untreated or inadequately treated TM patients is interspersed with various complications including recurrent infections, spontaneous fractures, gallstones, and leg ulcers. The mortality rate in such patients was formerly high around puberty. Fortunately, this is no longer the case in children with well-treated TM. TM children well transfused to maintain a hemoglobin level above 9 g/dL have relatively normal growth and development, and their future course depends on whether they have received adequate iron chelation (see Treatment). Many children who are adequately transfused and are fully compliant with iron chelation therapy develop normally, enter puberty, and become sexually mature. At present, we are dealing with an adult population of TM patients who may suffer from the side effects of long-term treatment, namely, transfusion-associated infections (particularly hepatitis B and C and, in some populations, HIV infection) and organ damage (liver, heart, endocrine glands) due to unsatisfactory long-term iron chelation. The main causes of death in adult TM patients still remain cardiac complications, although liver cirrhosis is also increasing because of the prolongation of life.[9]

Heart failure in these patients is multifactorial, involving chronic anemia, remaining iron overload, myocarditis, pericarditis, and probably other

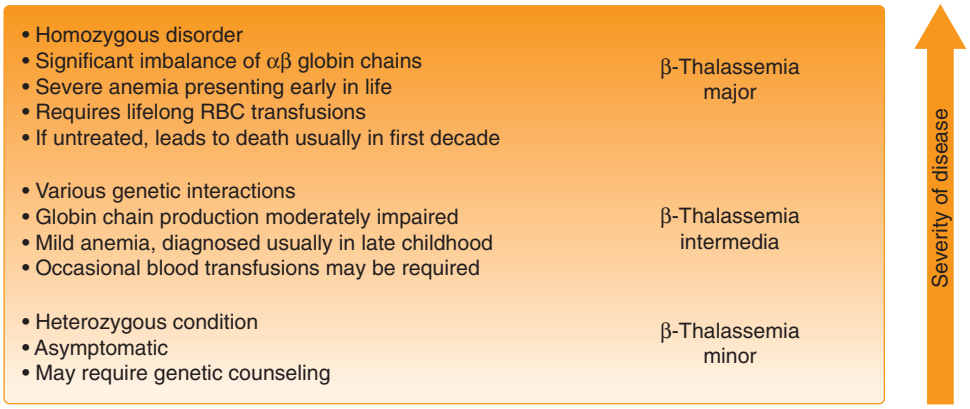

A

	Thalassemia major more likely	Thalassemia intermedia more likely
Clinical		
Presentation (years)	<2	>2
Hb levels (g/dL)	<7	7-10
Liver/spleen enlargement	Severe	Moderate to severe
Hematologic		
HbF (%)	>50	10-50 (may be up to 100%)
HbA_2 (%)	>3.5	<4-4.5
Genetic		
Parents	Both carriers of high HbA_2 β-thalassemia	One or both carriers: high HbF β-thalassemia, borderline HbA_2
Molecular		
Type of β-chain mutation	Severe	Mild/silent
Coinheritance of α-thalassemia	No/rare	Yes
Hereditary persistence of fetal hemoglobin	No	Yes
δβ-thalassemia	No	Yes
Gy Xmnl polymorphism	No	Yes

B

FIGURE 162-1. **A, Clinical classification of β-thalassemias, from β-thalassemia minor to β-thalassemia major, according to the disease severity. β-Thalassemia minor subjects are asymptomatic, whereas at the severe extreme, β-thalassemia major patients are transfusion dependent. In between is a wide spectrum of clinical phenotypes labeled thalassemia intermedia because of different genetic interactions characterized by moderate to severe transfusion-independent anemia. B, Tentative criteria to differentiate thalassemia major from thalassemia intermedia at presentation. HbA = adult hemoglobin; HbF = fetal hemoglobin; RBC = red blood cell.**

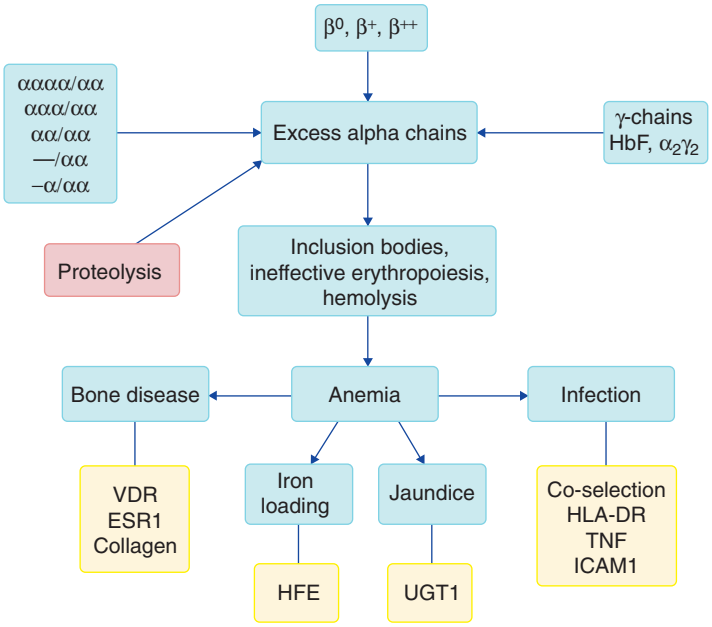

FIGURE 162-2. Pathophysiology of β-thalassemias and modifiers of globin chain imbalance. The severity of ineffective erythropoiesis is dependent on the degree of excess free α chains that result primarily from three different mechanisms: (1) inheritance of severe, mild, or silent β-chain mutations; (2) coinheritance of determinants associated with increased γ-chain production; and (3) coinheritance of α-thalassemia. A phenotype of thalassemia intermedia may result from the increased production of α-globin chains by a triplicated (ααα) or quadruplicated (αααα) α genotype associated with β heterozygosity. Inheritance of polymorphisms or mutations of genes involved in bone, iron, and bilirubin metabolism as well as in infection may contribute to modify the clinical course of the disease. ESR1 = estrogen receptor 1; HbF = fetal hemoglobin; HFE = hereditary hemochromatosis gene; HLA = human leukocyte antigen; ICAM1 = intercellular adhesion molecule 1; TNF = tumor necrosis factor; UGT1 = UDP-glucose : glycoprotein glucosyltransferase 1; VDR = vitamin D receptor.

mechanisms. Furthermore, besides the degree of globin chain imbalance that is dependent on genetic factors linked to the globin genes (coinheritance of α-thalassemia, innate increased synthesis of HbF), there are many other genetic modifiers that at the secondary level may affect the outcome of complications in different ways. For example, the presence of a polymorphic variant in the UGT1A1 promoter responsible for Gilbert syndrome (Chapter 147) may increase the predisposition to cholelithiasis, which is already a common complication in thalassemia. Likewise, polymorphisms in genes involved in iron homeostasis or bone metabolism may affect negatively or positively the degree of iron overload or the osteopenia and osteoporosis, respectively (see Fig. 162-2). Environmental factors, including social conditions, nutrition, and the availability of medical care, have also been implicated in the variable severity of TM clinical manifestations.

TI belongs to the non–transfusion-dependent group of thalassemia disorders; it is a clinical term used to describe patients with anemia and splenomegaly but without the full spectrum of clinical severity found in TM.[10] The clinical phenotypes of TI lie between those of thalassemia minor and major, encompassing a wide clinical spectrum. Mildly affected patients are almost completely asymptomatic until adult life, experiencing only mild anemia and spontaneously maintaining hemoglobin levels between 7 and 10 g/dL. Patients with more severe TI generally present between the ages of 2 and 6 years, and although they are able to survive without regular transfusion therapy, growth and development can be retarded. Most TI patients are homozygotes or compound heterozygotes for mild to moderate β-gene mutations (β^+/β^+;β^0/β^+); less commonly, only a single β-globin gene is affected. Because the clinical severity of the disease is dictated by the different extent of globin chain imbalance, at least three different mechanisms may promote the mild clinical characteristics of TI compared with TM: inheritance of mild or silent β-gene mutations; coinheritance of determinants associated with increased γ-chain production, which contributes to neutralizing the large proportion of unbound α chains; and coinheritance of α-thalassemia, which reduces the synthesis of α chains, thereby reducing the α/non-α chain imbalance.

Three main factors are responsible for the clinical sequelae in TI patients: rate of ineffective erythropoiesis, chronic anemia, and iron overload. The ineffective erythropoiesis primarily depends on the underlying molecular defects as already mentioned and is due to precipitation of free α chains in erythroid precursors in the bone marrow, causing membrane damage and premature cell death in the marrow. The degree of ineffective erythropoiesis is the primary determinant of the anemia of TI; peripheral hemolysis of mature, circulating red cells and an overall reduction in hemoglobin synthesis are secondary. Hemolysis and damaged red cells that expose negatively charged membrane phosphatidyl-serine residues have been linked to the development of the hypercoagulable state and increased risk of pulmonary hypertension in the TI patient population.[11]

Chronic anemia and chronic ineffective erythropoiesis lead to an inappropriate increase in gastrointestinal iron absorption, resulting in iron overload. In contrast, in TM, iron loading mainly results from transfusional iron infusion. It has been shown that chronic anemia and ineffective erythropoiesis, which are characteristic of TI, are associated with reduced expression of hepcidin, a hepatic peptide that plays a central role in iron homeostasis (Chapter 212). Moreover, the growth and differentiation factor 15 (GDF15) secreted by erythroid precursors and overexpressed in the presence of ineffective erythropoiesis may suppress hepcidin synthesis. More recently a molecule named erythroferrone, secreted by erythroblasts, has been identified and may also play a role in hepcidin regulation in TI. Taken together, ineffective erythropoiesis (leading to increased GDF15) and chronic anemia/hypoxia result in hepcidin suppression, increased dietary iron absorption from the gut, and increased release of recycled iron from the reticuloendothelial system, leading to an iron overload situation similar to that observed in patients with hereditary hemochromatosis syndromes (which are characterized by impaired hepcidin production)[12] (Chapter 212).

As a consequence of these pathophysiologic processes, several complications have been identified as unique in TI patients compared with TM patients, especially in splenectomized naïve patients. Cholelithiasis is much more common in TI than in TM because of ineffective erythropoiesis and peripheral hemolysis. Extramedullary hematopoiesis, as a compensatory mechanism, leads to the formation of erythropoietic tissue masses that primarily affect the spleen, liver, lymph nodes, and vertebrae. These masses may cause neurologic problems, such as spinal cord compression (sometimes causing paraplegia) and intrathoracic masses. Leg ulcers that are rare in well-transfused TM patients are common in adult TI patients; it remains unclear why at the same level of hemoglobin and at the same HbF level some patients develop leg ulcers and others do not. Thrombotic risk is definitely increased in TI patients. Several studies have collectively shown that the incidence of thromboembolic events is higher in TI splenectomized patients. Although stroke is rare in TI, asymptomatic brain damage including ischemia has been documented by magnetic resonance imaging (MRI) and computed tomography in TI patients. Pulmonary hypertension (Chapter 68) is prevalent in TI patients (approximately 60%) and is thought to be the primary cause of heart failure in this patient population.[13] Liver disease due to viral infection is less frequent than in TM; however, abnormal liver enzymes are frequently observed in TI patients, primarily because of hepatocyte damage resulting from iron overload. In some of the oldest patients with TI (>40 years), hepatocellular carcinoma (Chapter 196) due to long-term untreated iron accumulation has been detected with resultant cirrhosis in the liver as found in hereditary hemochromatosis (Chapter 212).[14]

Hypogonadism, hypothyroidism, and diabetes mellitus are rare. Although patients with TI generally experience puberty late, they have normal sexual development and are usually fertile. Women with TI may have spontaneous successful pregnancies, although complications during pregnancy may occur.[15]

Thalassemia minor is the heterozygous state of β-thalassemia. Subjects with thalassemia minor are "carriers" of a single β-globin gene defect, and they are usually asymptomatic except for a mild anemia of pregnancy. The carriers are usually identified as part of a family study, incidentally during an intercurrent illness, or as part of a population survey. The anemia is mild, microcytic and hypochromic, and associated with an elevated level of HbA_2. The blood smear shows characteristic microcytosis and hypochromia, with some variation in size and shape of the red cells. The presence of target cells is variable. The hematologic features are remarkably similar among different ethnic group. Carriers of β-thalassemia with normal HbA_2 have been observed in settings in which an individual with mild TI was found to have one parent with typical β-thalassemia minor with elevated HbA_2, whereas the other showed either minimal or no hematologic abnormalities and normal HbA_2. Subjects with normal HbA_2 are usually carriers of silent β^{++} mutations (β^{++}). Coincidental iron deficiency can lower elevated HbA_2 levels to the normal range.

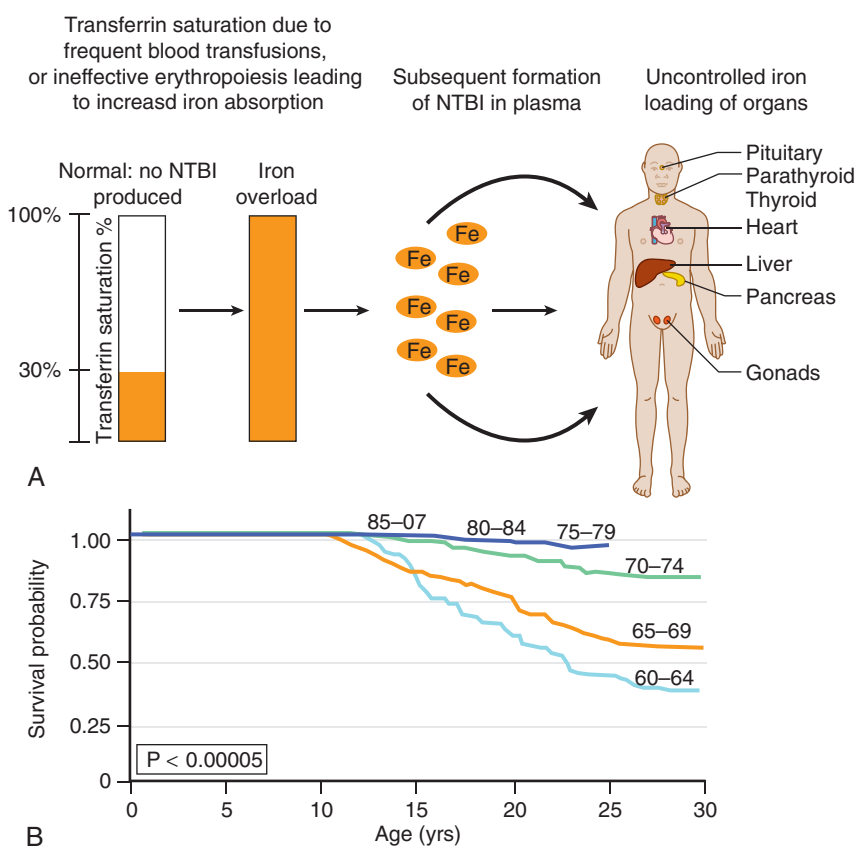

FIGURE 162-3. A, Transfusion iron overload and specific organ loading. B, Survival of Italian thalassemia major patients by cohort of birth. NTBI = non–transferrin-bound iron.

δβ-Thalassemia and Hereditary Persistence of Fetal Hemoglobin

The clinical manifestations of $(\delta\beta)^0$-thalassemias are similar to those of TI, whereas the heterozygotes are distinguished from β-thalassemia heterozygotes by normal levels of HbA_2 together with an increased HbF level of 5 to 20%. Homozygotes for Hb_{Lepore} may have phenotypes similar to either TM or TI. Subjects affected by deletion or nondeletion forms of HPFH are usually asymptomatic.

Any of the β-thalassemia defects may be coinherited with β-chain variants (HbS, HbC, HbE) and cause a clinically relevant β-thalassemia phenotype of different severity. These variants further illustrate that β-thalassemia syndromes have a wide clinical spectrum and that specific therapeutic approaches may completely change the clinical course and the natural history of these disorders (Fig 162-3).

Hemoglobin E and Hemoglobin E Thalassemias

Hemoglobin E (HbE) is caused by a substitution of glutamic acid by lysine at codon 26 of the β-globin gene. It is an common mutation, especially in the Indian subcontinent and Southeast Asia and among people who have emigrated from those regions. The frequency of HbE approaches 60% in many regions of Thailand, Laos, and Cambodia. HbE results in reduced synthesis of the β-E chain and therefore has the phenotype of a mild form of β-thalassemia. HbE trait has little clinical significance; it may be associated with slight microcytosis without anemia, but it could be mistaken for iron deficiency without appropriate laboratory testing. Homozygous HbEE individuals have mild anemia and microcytosis, occasional splenomegaly, and peripheral blood smears showing red cell changes similar to those seen in thalassemia traits, including target cells.

HbE interacts with different forms of α-thalassemia to produce a wide variety of clinical disorders. Its coinheritance with β-thalassemia, a condition called HbE/β-thalassemia, is the most common severe form of β-thalassemia in Asia and globally represents about 50% of the clinically severe β-thalassemias.[16]

DIAGNOSIS

The diagnosis of thalassemia may be required in a patient with an appropriate, suggestive clinical picture or for the identification of a heterozygote subject as part of a family study or population screening program. The general approach is common to any form of thalassemia, regardless of presentation. The primary evaluation is based on hematologic changes; the red cell indices by electronic cell counter and the red cell morphology examined on a well-stained blood film are sufficient to direct further investigations. Individuals with mean corpuscular volume below 80 fL and mean corpuscular hemoglobin below 27 pg with normal iron parameters need to be further investigated. The red blood cell number is usually higher than normal. In the presence of anemia with thalassemic red cell changes, the next step is the evaluation of hemoglobin fractions (HbA, HbA_2, HbF, or hemoglobin variants) by electrophoresis on cellulose acetate at alkaline pH or, even better, by high-performance liquid chromatography that enables the precise measurement of HbA_2, HbF, and HbA and the provisional identification of a large number of hemoglobin variants, including HbE. An HbA_2 level above 3.5% associated with hypochromic microcytic red cells is diagnostic of β-thalassemia minor. HbA_2 values between 3.2 and 3.5% (borderline) should be interpreted with care because they could be due to interaction of more than one thalassemic defect (α and β), a silent β mutation, or concomitant iron deficiency. If iron deficiency is present, it should be corrected and the HbA_2 estimation repeated. The majority of individuals with thalassemic red cell indices with normal or low HbA_2 and normal HbF will be α^0-thalassemia carriers or α^+-thalassemia homozygotes. Carriers of α^0-thalassemia may have a few red cells with HbH inclusions. Microcytosis with low or normal HbA_2 levels with elevated HbF (2 to 20%) indicates heterozygosity for δβ-thalassemia. Patients with HPFH usually have normal red blood cell indices but increased levels of HbF with different intercellular distribution (homogeneous or pancellular with the exception of heterocellular HPFH) compared with δβ-thalassemia (uneven or heterocellular).

A radioactive method for measuring the α/β-globin synthesis ratio was introduced in the mid to late 1960s, and it was largely directed at prenatal diagnosis in the pre-DNA era. Although it gives a quantitative assessment of globin production, today its use is limited to difficult cases because of interaction of different globin chain defects. The definitive diagnosis of the thalassemia syndromes involves the identification of the underlying mutations through DNA analysis. There are several methods available for the diagnosis of any particular mutation, such as polymerase chain reaction (PCR) restriction enzyme analysis, PCR allele-specific oligonucleotides, gap PCR, and

direct sequencing that at present is probably the easiest and most reliable method. For deletion forms of α-thalassemia, the multiplex ligation-dependent probe amplification is a recently introduced, useful method.

During their clinical course, patients affected by different forms of thalassemia develop several complications mainly due to iron overload, which requires monitoring to direct iron chelation therapy. The principal methods of determining body iron levels (Chapter 212) are measurements of serum ferritin level and assessment of liver iron concentration from biopsy tissue or, as an alternative noninvasive method, by R2 MRI. High serum ferritin levels (>2500 μg/L) and high liver iron concentration (>15 g/dry weight) indicate high risk of significant morbidity and mortality. Cardiac iron can be measured by a T2* MRI procedure that allows estimation of the cardiac iron load. MRI T2* values below 10 milliseconds are always associated with severe iron load and high risk of heart failure within 1 year. MRI T2* values above 20 milliseconds are considered normal, meaning no iron in the heart. Echocardiography may also be useful to evaluate functional changes. For other complications, including endocrinopathies, liver disease, lung disease, thrombophilia, and bone disease, the diagnostic approaches are similar to those used in clinical practice; these are performed with consideration of test cost, performance characteristics, and preferences of the patients, as described in the corresponding chapters.

TREATMENT

Conventional Treatment

No specific treatment is required for α- or β-thalassemia heterozygotes (carriers, thalassemia minor), but they should receive appropriate genetic counseling. During pregnancy, thalassemia-carrying women may become more anemic, so they should be observed carefully, mainly during the second and third trimesters, and supported with folic acid. When real iron deficiency is associated with thalassemia traits, iron supplementation should be provided, monitoring transferrin saturation and ferritin. A few cases of in utero blood transfusions have been reported with hemoglobin Bart hydrops fetalis syndrome; most of the infants have been delivered prematurely by cesarean section, subsequent development has been abnormal, and survivors required regular blood transfusions after birth. HbH patients in general have high hemoglobin levels (8 to 9 g/dL) and do not need regular blood transfusion. Supplementation with folic acid (2 to 5 mg/day) is generally recommended, especially in pediatric patients. The major complications in HbH disease are hemolytic crises that may occur during or after acute infections; in such cases, immediate intervention, including blood transfusions and treatment for infections, should be promptly administered.

The clinical management of TM and TI remains the major issue. The quality and duration of life of TM and TI patients have been transformed in this century, with life expectancy increasing well into the third and fourth decades. Nevertheless, prolongation of life is accompanied by several complications, partly due to the underlying disorder and partly as a consequence of the treatment with blood transfusions and iron overload. Moreover, we are starting to deal with aging-related complications in the context of a multiorgan disease that requires management by a team of clinicians who have specific knowledge of thalassemias in adults, working together with different specialists and well-trained nurses. The conventional treatment of TM patients includes regular transfusion therapy and iron chelation. The definition of the optimal transfusion and iron chelation regimen has been the most important advance in the management of TM patients, with the primary objective being to control the ineffective erythropoiesis, its consequences, and the body iron burden. The optimal transfusion regimen involves regular blood transfusions, usually administered every 2 to 5 weeks, to maintain the pretransfusion hemoglobin levels above 9 to 10.5 g/dL. The decision to initiate lifelong transfusion therapy should be based on a definitive diagnosis of severe thalassemia, taking into account the molecular defects, the severity of anemia on repeated measurement, the level of ineffective erythropoiesis, and the clinical criteria (such as failure to thrive or bone changes). It is advisable that TM patients receive leukoreduced packed red cells to minimize transfusion reactions and pathogen transmission. Adverse reactions to red blood cell transfusions may occur during or after transfusion and can be hemolytic and nonhemolytic. Transfusion-related acute lung injury is rare but severe and must be immediately managed (Chapter 177).

Many patients with TM require splenectomy because of hypersplenism. However, optimal clinical management may delay or even obviate the need for splenectomy that was common in the past. Splenectomy should be considered for patients whose annual blood consumption increases progressively and is responsible for significant increases in iron stores despite good chelation therapy or in the presence of symptoms due to spleen enlargement. Clinical problems related to leukopenia or thrombocytopenia due to hypersplenism could also be the reasons for considering splenectomy.[17] The major complication of splenectomy is severe and sometimes overwhelming infection. Because removal of the spleen may reduce the primary immune response to encapsulated organisms, it is advisable to delay splenectomy until patients are at least 5 years old. The mortality rate for postsplenectomy overwhelming infection in thalassemia patients is approximately 50% despite intensive supportive care. Therefore, it is mandatory to adopt preventive measures including immunoprophylaxis (vaccination against *Streptococcus pneumoniae*, pneumococcus, and meningococcus), chemoprophylaxis, and education of the parent and patient to recognize and to report febrile illnesses.

Iron overload is an inevitable and serious complication of long-term blood transfusion therapy and hyperabsorption of dietary iron that requires adequate treatment to prevent early death, mainly from iron-induced cardiac disease. Optimal chelation therapy extends complication-free survival (see Fig. 162-3). The standard chelation therapy for more than 40 years was deferoxamine (DFO), given for 10 to 24 hours daily as a continuous subcutaneous infusion 5 to 7 days per week. The long-term efficacy of DFO has been extensively documented in large cohorts of patients in Italy and elsewhere. Unfortunately, compliance with the rigorous regimen of daily subcutaneous infusions is a serious limiting factor, and life expectancy in noncompliant patients is not different from that in the pre-DFO era. This has been the rationale behind the intensive effort to identify alternative, orally effective iron chelators. At present, two oral iron chelators are on the market: deferiprone (DFP) and deferasirox (DFX). DFP is registered in Europe and more recently in the United States. By the guidelines of the EMEA countries (Europe, Middle East, and Africa), treatment with DFP at doses of 75 to 100 mg/kg/day is restricted to patients unable to use DFO or patients with an unsatisfactory response to DFO as judged by serum ferritin levels and by liver iron concentrations. Studies indicate that DFP may be more effective than DFO in protecting the heart from the accumulation of iron. A potential benefit of combined DFO and DFP therapy has been observed, and according to Thalassemia International Federation guidelines, a combination treatment (DFO and DFP) should be considered for patients with high levels of heart iron or cardiac dysfunction. The new orally effective iron chelator DFX has been shown to be effective and safe in removing excess iron from different organs, including the heart. DFX is now available in most countries throughout the world as first-line treatment. Its use has clearly demonstrated that iron chelation is not a standard treatment, but it should be individualized according to age, history of compliance with previous chelation, and other factors. Monitoring and adjustment of iron chelation by repeated measurements of ferritin, calculation of iron intake by transfusions, and, whenever possible, measurement of cardiac and liver iron at least once by MRI are mandatory.

The management of TI patients is more complicated because of the wide heterogeneity of TI phenotypes. A number of options are currently adopted for treatment of TI patients, including transfusion therapy, splenectomy, modulation of HbF production, and hematopoietic stem cell transplantation (HSCT). However, increasing evidence is documenting the benefit of transfusion therapy in decreasing the incidence of complications. Thus, although the common practice has been to initiate transfusion when complications ensue, it may be worthwhile to start transfusion therapy earlier as a preventive approach, which will also help alleviate the increased risk of alloimmunization with delayed initiation of transfusion. The initiation of iron chelation therapy in patients with TI depends not only on the amount of excess iron but also on the rate of iron accumulation, the duration of exposure to excess iron, and various other factors in individual patients.[18]

Bone Marrow Transplantation and Experimental Therapies

Allogeneic HSCT (Chapter 178) in thalassemia syndromes has been increasingly successful during the last 2 decades, mainly in β-thalassemia major.[19] Predictors of poor transplant outcome are hepatomegaly, history of irregular chelation, and hepatic fibrosis. Patients are categorized into three risk classes. Class 1 patients have none of these adverse risk factors, class 2 patients have one or two adverse risk factors, and class 3 patients have all three. In the most recent update of the Pesaro group's experience, the probability of thalassemia-free survival for patients younger than 17 years at the time of HSCT receiving the allograft from an HLA-identical relative was 87% and 85%, respectively, in patients belonging to class 1 and class 2 and much lower in young patients in class 3. The progressive adjustment of conditioning regimens in class 3 patients and in adults (>17 years) has significantly reduced the incidence of transplant-related mortality in patients in class 3. Only 25 to 30% of patients with diseases potentially curable by HSCT have a suitable HLA-compatible sibling. Bone marrow transplantation from unrelated donors increases significantly the incidence of acute and chronic graft-versus-host disease, particularly in thalassemia. A study from the Eurocord cooperative group reported the outcome of 33 patients with thalassemia belonging to class 1 and class 2 (Pesaro classification) who received cord blood hematopoietic stem cells from an HLA-identical sibling; no patient died of transplant-related complications, suggesting that related cord blood HSCT is a safe procedure for thalassemia patients.[20]

An alternative treatment of β-thalassemia consists of the pharmacologic stimulation of HbF synthesis. In humans, hemoglobin switch from HbF to HbA occurs in the period around birth as a result of γ- to β-globin gene switching. A number of pharmacologic agents able to reactivate HbF synthesis have been identified, including hypomethylating agents, histone deacetylase inhibitors, and hydroxyurea. Whereas the effect of these pharmacologic treatments (particularly hydroxyurea) in sickle cell disease is clear (Chapter 163), their benefit on the clinical course of β-thalassemia is presently limited. The discrepancy between these two conditions in the response to HbF inducers may be mainly related to the higher level of HbF required in β-thalassemia to achieve clinical results compared with those observed in sickle cell disease. The limited clinical response to γ-globin inducers observed in the majority of β-thalassemic patients may be also a reflection of the unfavorable effects of these agents on the other globin genes (i.e., increased α-globin synthesis). A new molecule, sotatercept (ACE-011), an activin-type IIA receptor fusion protein, has recently been shown to increase the release of mature erythrocytes into the circulation by acting mainly on late-stage erythropoiesis. Clinical data in healthy volunteers have shown that treatment with sotatercept results in increased red blood cell parameters. A phase IIa, multicenter, open-label, dose-finding study to determine a safe and active dose level of sotatercept in adult patients with β-thalassemia intermedia is ongoing.

Gene therapy (Chapter 44) is an attractive approach for thalassemia syndromes; however, this strategy poses major challenges in terms of controlling transgene expression, which should be erythroid specific and sustained over time. Treatment of β-thalassemia, sickle cell disease, and other disorders through lentivirus-mediated gene transfer has been reported in murine and primate models, but until now few patients have been treated.

UNSTABLE HEMOGLOBINOPATHIES

More than 80 rare mutant hemoglobins have been reported to cause hemolytic anemia by either amino acid replacements or deletions that significantly lower solubility. These mutant hemoglobins form intracellular precipitates that can be detected as so-called Heinz bodies when the blood smear is exposed to a supravital stain. Structural abnormalities include mutations that weaken the linkage between heme and globin, disrupt secondary (α-helical) structure, or introduce a charged or polar side group into the hydrophobic interior of the globin subunit.

This disorder, sometimes called congenital Heinz body hemolytic anemia, is inherited in an autosomal dominant manner. Severely affected individuals have jaundice, splenomegaly, and, on occasion, dark brown urine due to the release of heme and aberrant conversion to dipyroles. The instability of a few of these globin mutants is so extreme that they cannot be detected by routine laboratory methods. This results in a thalassemia phenotype with microcytosis and ineffective erythropoiesis. Like individuals with glucose-6-phosphate dehydrogenase deficiency (Chapter 161), those with unstable hemoglobin mutants often lack clinical symptoms and signs of hemolysis until they develop an infection or are exposed to an oxidant drug.

The diagnosis can be established by a combination of a positive Heinz body preparation and either abnormal hemoglobin electrophoresis or demonstration of a precipitate after exposure of a hemolysate to heat or isopropanol. Some clinics have access to a reference laboratory that can identify the specific mutation by α- and β-globin DNA sequencing.

Most individuals with this disorder do not require treatment; some are symptomatic from severe anemia. Splenectomy generally results in a significant increase in red cell mass. However, the fraction of Heinz body–positive red cells increases markedly after splenectomy, and these patients are now at significant risk for development of pulmonary hypertension and cor pulmonale.

Grade A References

A1. Fisher SA, Brunskill SJ, Doree C, et al. Desferrioxamine mesylate for managing transfusional iron overload in people with transfusion-dependent thalassaemia. *Cochrane Database Syst Rev*. 2013;8:CD004450.

A2. Pennell DJ, Berdoukas V, Karagiorga M, et al. Randomized controlled trial of deferiprone or deferoxamine in beta-thalassemia major patients with asymptomatic myocardial siderosis. *Blood*. 2006;107:3738-3744.

A3. Fisher SA, Brunskill SJ, Doree C, et al. Oral deferiprone for iron chelation in people with thalassaemia. *Cochrane Database Syst Rev*. 2013;8:CD004839.

A4. Tanner MA, Galanello R, Dessi C, et al. A randomized placebo-controlled, double-blind trial of the effect of combined therapy with deferoxamine and deferiprone on myocardial iron in thalassemia major using cardiovascular magnetic resonance. *Circulation*. 2007;115:1876-1884.

A5. Taher AT, Porter J, Viprakasit V, et al. Deferasirox reduces iron overload significantly in nontransfusion-dependent thalassemia: 1-year results from a prospective, randomized, double-blind, placebo-controlled study. *Blood*. 2012;120:970-977.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Forget BG, Bunn HF. Classification of the disorders of hemoglobin. *Cold Spring Harb Perspect Med*. 2013;3:a011684.
2. Katsumura KR, DeVilbiss AW, Pope NJ, et al. Transcriptional mechanisms underlying haemoglobin synthesis. *Cold Spring Harb Perspect Med*. 2013;3:a015412.
3. Higgs DR. The molecular basis of α-thalassemia. *Cold Spring Harb Perspect Med*. 2013;3:a011718.
4. Thein SL. The molecular basis of β-thalassemia. *Cold Spring Harb Perspect Med*. 2013;3:a011700.
5. Piel FB, Weatherall DJ. The α-thalassemias. *N Engl J Med*. 2014;371:1908-1916.
6. Musallam KM, Rivella S, Vichinsky E, et al. Non–transfusion-dependent thalassemias. *Haematologica*. 2013;98:833-844.
7. Karimi M, Cohan N, De Sanctis V, et al. Guidelines for diagnosis and management of Beta-thalassemia intermedia. *Pediatr Hematol Oncol*. 2014;31:583-596.
8. Joly P, Pondarre C, Badens C. Beta-thalassemias: molecular, epidemiological, diagnostical and clinical aspects. *Ann Biol Clin (Paris)*. 2014;72:639-668.
9. Kountouras D, Tsagarakis NJ, Fatourou E, et al. Liver disease in adult transfusion-dependent beta-thalassaemic patients: investigating the role of iron overload and chronic HCV infection. *Liver Int*. 2013;33:420-427.
10. Musallam KM, Rivella S, Vichinsky E, et al. Non-transfusion-dependent thalassemias. *Haematologica*. 2013;98:833-844.
11. Cappellini MD, Poggiali E, Taher AT, et al. Hypercoagulability in β-thalassemia: a status quo. *Expert Rev Hematol*. 2012;5:505-511.
12. Musallam KM, Cappellini MD, Taher AT. Iron overload in β-thalassemia intermedia: an emerging concern. *Curr Opin Hematol*. 2013;20:187-192.
13. Derchi G, Galanello R, Bina P, et al. Webthal Pulmonary Arterial Hypertension Group. Prevalence and risk factors for pulmonary arterial hypertension in a large group of β-thalassemia patients using right heart catheterization: a Webthal study. *Circulation*. 2014;129:338-345.
14. Maakaron JE, Cappellini MD, Graziadei G, et al. Hepatocellular carcinoma in hepatitis-negative patients with thalassemia intermedia: a closer look at the role of siderosis. *Ann Hepatol*. 2013;12:142-146.
15. Thompson AA, Kim HY, Singer ST, et al. Pregnancy outcomes in women with thalassemia in North America and the United Kingdom. Thalassemia Clinical Research Network. *Am J Hematol*. 2013;88:771-773.
16. Fucharoen S, Weatherall DJ. The hemoglobin E thalassemias. *Cold Spring Harb Perspect Med*. 2012;2:a011734.
17. Piga A, Serra M, Longo F, et al. Changing patterns of splenectomy in transfusion-dependent thalassemia patients. *Am J Hematol*. 2011;86:808-810.
18. Cappellini MD, Taher A, Musallam K. Guidelines for the clinical management of non–transfusion dependent thalassaemias. 2013. http://www.thalassaemia.org.cy/educational-programme/publications.shtml. Accessed February 6, 2015.
19. Mathews V, Srivastava A, Chandy M. Allogeneic stem cell transplantation for thalassemia major. *Hematol Oncol Clin North Am*. 2014;28:1187-1200.
20. Locatelli F, Kabbara N, Ruggeri A, et al. Outcome of patients with hemoglobinopathies given either cord blood or bone marrow transplantation from an HLA-identical sibling. Transplantation (EBMT) group. *Blood*. 2013;122:1072-1078.

REVIEW QUESTIONS

1. An 18-year-old woman from Pakistan has been observed for a lifelong severe, microcytic and hypochromic anemia. Her hemoglobin levels have generally been in the range of 7 to 10 g/dL, but she recalls needing blood transfusions only twice in her life, on both occasions after traumatic bone fractures. Throughout childhood, her parents had kept her from playing sports or participating in activities requiring significant physical exertion because they had been told by pediatricians that she had a "weak heart." She had markedly delayed puberty but now had regular menses. Physical examination shows marked hepatosplenomegaly; laboratory test abnormalities include indirect hyperbilirubinemia and hyperglycemia. What is the most likely diagnosis?

A. β-Thalassemia major
B. Homozygosity for hemoglobin E (HbEE)
C. Hemoglobin H (HbH) disease
D. Hydrops fetalis
E. Hereditary juvenile hemochromatosis

Answer: C The clinical picture is most consistent with thalassemia intermedia. In the case of α-thalassemia, this would be most like hemoglobin H disease (most frequently resulting from the interaction of α^+- and α^0-thalassemia), with the β_4 tetramers of hemoglobin H forming red cell inclusion bodies leading to chronic hemolysis. Patients with α- or β-thalassemia intermedia become severely iron overloaded from lifelong hyperabsorption of dietary iron, even without needing blood transfusions. A similar gut hyperabsorption of iron occurs in familial hemochromatosis states, but those individuals are not typically anemic. This patient's hyperglycemia, growth retardation, possibly her bone fragility, and even the history of a "weak heart" (perhaps due to cardiomyopathy) are all manifestations of systemic iron overload. Thalassemias are especially prevalent not only in people of Mediterranean and Middle Eastern origin but also in those from the Indian subcontinent. In contrast to thalassemia intermedia, a patient with β-thalassemia major would be expected to be transfusion dependent throughout life. Hydrops fetalis, in which all four α-globin genes are deleted, is incompatible with life. Hemoglobin E, even in its homozygous state, causes little if any anemia and microcytosis.

2. Which of the following statements is correct regarding hematopoietic stem cell transplantation for thalassemia?

A. It can lead to thalassemia-free survival of at least 85% in low-risk, younger patients.
B. It should not be administered with cord blood for transplantation.
C. It has been mostly effective in the α-thalassemias.
D. It is mostly reserved for thalassemia patients who are poorly iron chelated and have liver involvement.
E. It is associated with a reduced risk of graft-versus-host disease compared with comparable unrelated transplants for other hematologic diseases.

Answer: A The highest risk indicators in hematopoietic stem cell transplantation for thalassemia are hepatomegaly, hepatic fibrosis, and poor iron chelation history. On the other hand, the Pesaro group's experience has demonstrated 87% disease-free survival in thalassemia patients who are younger than 17 years at the time of transplantation, have none of the risk factors noted, and are allografted from an HLA-identical related donor. Early experience with cord blood transplantation has shown it to be a safe procedure for thalassemia patients. Hematopoietic stem cell transplantation has been mainly successful in β-thalassemia. The risk of graft-versus-host disease (acute and chronic) after bone marrow transplantation from unrelated donors is at least as high in thalassemia as it is in other indications for bone marrow transplantation.

3. Heterozygosity for hemoglobin E

A. Is a rare form of unstable hemoglobinopathy
B. Is associated with lifelong, low-grade hemolysis
C. Causes iron overload from hyperabsorption of dietary iron
D. Is frequently associated with iron deficiency
E. Is of little if any clinical significance

Answer: E Hemoglobin E is one of the most common mutations in the world, especially in the Indian subcontinent and Southeast Asia. Its phenotype is that of a mild thalassemia, not an unstable hemoglobinopathy. It is of little if any clinical importance in its heterozygous form (HbE trait); even individuals who are homozygous for hemoglobin E (HbEE) usually have only mild anemia and microcytosis, without any clinical sequelae. Hemoglobin E can be mistaken for iron deficiency in the absence of further laboratory testing, but it is not known to be associated with iron deficiency.

163

SICKLE CELL DISEASE AND OTHER HEMOGLOBINOPATHIES

MARTIN H. STEINBERG

SICKLE CELL DISEASE

DEFINITION

Sickle cell disease, caused by a mutation in the β-globin gene (*HBB*), consists of a group of chronic hemolytic anemias, all characterized by vaso-occlusive events, hemolytic anemia, vasculopathy, widespread acute and chronic organ damage, and premature mortality.

EPIDEMIOLOGY

The prevalences of the various forms of sickle cell disease and of the sickle cell trait (HbAS), which is not truly a form of sickle cell disease,[1] vary in the United States and worldwide (Table 163-1 and Fig. 163-1). The sickle hemoglobin mutation became prominent in equatorial Africa, the Middle East, and India several thousand years ago, when deforestation, the rise of agriculture, and stagnant pooling of water permitted *Plasmodium falciparum* infection to become endemic. Individuals with HbAS were more likely to survive to reproductive age and had a selective advantage where falciparum malaria was present. Slave trading and war spread this mutation from Africa and other sites of origin to the Americas, throughout the Mediterranean basin, and eastward to the Indian subcontinent. In some sites in Africa, half the population has HbAS. The partial protection provided from severe malaria by HbAS, HbC, HbE, α-thalassemia, and β-thalassemia along with other red blood cell traits is the source of the selective pressure maintaining the high prevalence of these carrier states.

PATHOBIOLOGY

Globin, the protein portion of hemoglobin, harbors the iron-containing porphyrin heme ring and permits the molecule to operate efficiently in oxygen transport and its other physiologic functions (Fig. 163-2). Mutations can alter the primary amino acid sequence of the globin polypeptide, sometimes resulting in clinically significant diseases called hemoglobinopathies, of which sickle cell disease is an example. Sickle hemoglobin (HbS: $\alpha_2\beta_2^S$) is caused by an adenine (A) to thymidine (T) substitution (GAG → GTG) in codon 6 of the β-globin gene (*HBB*), resulting in replacement of the normal glutamic acid residue by a valine (Glu6Val) (Fig. 163-3). HbS polymerizes when it is deoxygenated, a property only of hemoglobin variants that have the *HBB* Glu6Val substitution. Critical amounts of HbS polymer within sickle erythrocytes cause cellular injury and lead to the phenotype of sickle cell disease, which is recognized by hemolytic anemia and vaso-occlusion. Other hemoglobin variants, such as HbE and HbC, are also common. More than 1000 hemoglobin mutations are known, and they can occasionally affect the stability and function of hemoglobin and cause hemolytic anemia (Chapter 158), disordered oxygen transport Chapter 166), or methemoglobinemia (Chapter 158). However, most globin mutations are clinically insignificant. Thalassemias (Chapter 162) are also caused by mutations in globin genes, but these mutations affect globin gene expression, reducing or preventing the synthesis of globin, although the structure of any globin produced is usually normal.

Sickle cell anemia (homozygosity for HbS) is noteworthy for its clinical heterogeneity. Any patient can have almost all known disease complications; some have almost none but die suddenly; some skip one or more

TABLE 163-1 GENETIC AND LABORATORY FEATURES OF COMMON SICKLE HEMOGLOBINOPATHIES*

GENOTYPE	GENETICS	PREVALENCE AMONG AFRICAN AMERICANS†	HEMATOCRIT (%)	MCV (fL)	HbS (%)	HbA_2 (%)	HbF (%)	SEVERITY‡
Sickle cell anemia (HbSS)	Homozygous HbS	1 : 600	18-28	85-95	>85	2-3	2-15	4
HbSS-α-thalassemia	Homozygous HbS α^+-thalassemia	30% of HbSS patients	25-33	70-85	>85	4-6	2-15	4
HbSC disease (HbSC)	Compound heterozygous HbS, HbC	1 : 800	28-40	70-85	50	2-3	1-8	2
HbS-β^0-thalassemia (HbS-β^0-Thal)	Compound heterozygous HbS, β^0-thalassemia	1 : 1600	20-30	65-75	>85	4-6	5-15	4
HbS-β^+-thalassemia (HbS-β^+-Thal)	Compound heterozygous HbS, β^+-thalassemia	1 : 1600	30-40	60-70	70-95	4-6	2-10	1-3
HbSE disease (HbSE)	Compound heterozygous HbS, HbE	Rare§	30-45	70-80	60	2-3	1	1-2
HbS-HPFH	Compound heterozygous HbS and gene deletion HPFH	Rare	38-45	70-80	70	2	20-30	0
Sickle cell trait (HbAS)¶	Heterozygous HbS	1 : 12	38-50	80-90	35-40	2-3	<1	0
Normal (HbAA)	Homozygous HbA	—	38-50	80-90	0	2-3	<1	—

Hb = hemoglobin; HPFH = hereditary persistence of HbF; MCV = mean corpuscular volume.
*Many other abnormal globin genes can be found as compound heterozygotes with the HbS gene. The most common of these are α-thalassemia, HbD, HbO (Arabia), HbG (Philadelphia), HPFH, Hb Hope, and Hb Lepore. Average ranges of laboratory values are shown, but these can vary according to the patient's age.
†These figures differ by the prevalence of the involved genes in the population studied. In West and Central Africa, where the disease is most common, approximately 2% of all newborns have sickle cell disease. The prevalence of the HbC trait in African Americans is 3%, and that of the β-thalassemia trait is 1%. About 30% of African Americans carry an α-thalassemia gene, which can alter the phenotype of sickle cell disease by causing microcytosis, reduced cell density, and less hemolysis.
‡Severity of disease compared with sickle cell anemia, clinically the most severe genotype. This is a qualitative ranking of the clinical severity of each genotype; within each genotype, there is great clinical heterogeneity.
§Although this combination is still a rare genotype, the rising Asian population in the United States (HbE is a Southeastern Asian gene) will make it more frequent with time. With few cases reported compared with the other genotypes, the phenotype of HbSE disease is not totally defined. It may resemble HbS-β^+-thalassemia with symptoms appearing mainly in adults.
¶Sickle cell trait should not be classified as a form of sickle cell disease. About 8% of African Americans are carriers of HbS. Carriers are hematologically normal with a normal life expectancy. The few abnormalities traceable to the presence of HbS besides the renal lesions (see Table 163-3) include a four-fold increased risk of pulmonary embolism, an increased risk of splenic infarction at high altitude, and a higher risk of dying during the course of exertional heat illness.

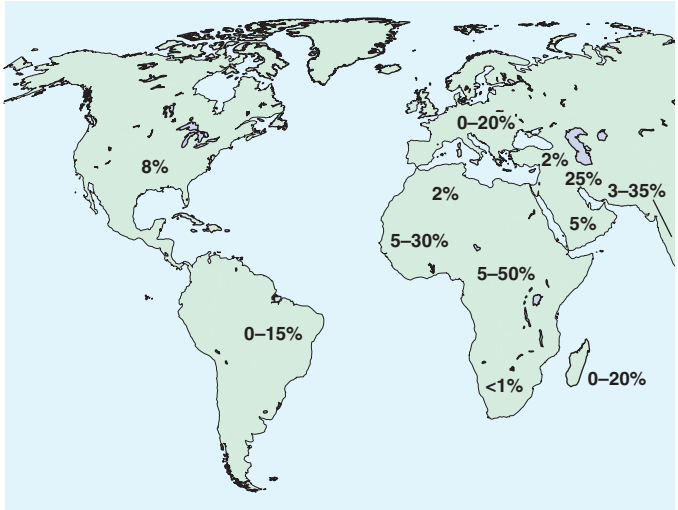

FIGURE 163-1. Worldwide prevalence of the sickle cell trait. Shown are the percentages of individuals with sickle cell trait in regions of the world where the hemoglobin S (HbS) gene is often present. In each geographic area, the prevalence of sickle cell trait can vary markedly according to racial or ethnic group, by historical migration patterns, and even from village to village. Not shown is the high concentration of the HbS gene in areas of Europe such as London, Manchester, and Paris, where migrants from Africa or Afro-Caribbean populations have settled.

complications of the disease but suffer intensely from others. Thus, this prototypical single-gene, mendelian disorder has exceptional phenotypic variability. Understanding of the pathobiology of the disease suggests many loci where the disease phenotype can be influenced by modifying genes. These genes affect the pathogenesis of sickle cell anemia by modulating fetal hemoglobin (HbF) concentration and mean corpuscular HbS concentration, and polymorphisms have also been noted in some genes that affect inflammation, oxidant injury, nitric oxide (NO) biology, vasoregulation, cell-cell interaction, blood coagulation, and hemostasis.[2]

HbF concentrations vary among patients with sickle cell anemia and among the erythrocytes of each individual.[3,4] Because HbF reduces HbS concentration and also directly inhibits HbS polymerization, its concentration within each cell and its distribution among all cells influence cellular heterogeneity. The sickle mutation is found on several different haplotypes of the β-globin gene cluster, reflecting different origins of the mutation in Africa and the Middle East; these haplotypes are associated with different HbF levels. The Senegal haplotype is associated with higher HbF levels than other African haplotypes are. The Arab-Indian haplotype, reflective of the HbS mutation that originates outside Africa, is associated with average HbF levels two to five times greater than in African haplotypes. Generally speaking, patients with higher HbF are more likely to have a less severe clinical course. Conversely, the Bantu haplotype, with lower HbF levels, may be associated with more disease complications. Nevertheless, there is great phenotypic heterogeneity within a particular haplotype group.

In sickle cell disease, erythrocytes are heterogeneous as a result of the highly variable cellular distribution and concentration of HbF and the varying increments of membrane damage. Cation homeostasis is impaired in some sickle cells.[5] A reduced capacity of sickle cells to maintain normal potassium gradients is mediated by activation of the Gardos, K^+/Cl^-, and other cotransport channels. As a result, sickle erythrocytes vary in their density and deformability. Irreversibly sickled cells (Fig. 163-4) always appear deformed because of permanent membrane damage, even though they may not contain HbS polymer. In some dense cells, the mean corpuscular hemoglobin concentration reaches 50 g/dL (normal, 27 to 38 g/dL), and deoxyHbS polymerization is exquisitely sensitive to the mean corpuscular hemoglobin concentration. Individuals with the highest numbers of irreversibly sickled cells and dense cells have the most hemolysis and anemia but not necessarily the highest incidence of acute vaso-occlusive events like painful episodes.

Hemolysis is mainly extravascular because of erythrophagocytosis by macrophages that recognize the damaged sickle erythrocyte. A variable amount of hemolysis is intravascular, and this liberates excessive amounts of hemoglobin into the circulation, thereby depleting haptoglobin and scavenging NO. This process promotes a vasoconstrictive, proinflammatory phenotype (Fig. 163-3). Certain complications of sickle disease, such as pulmonary

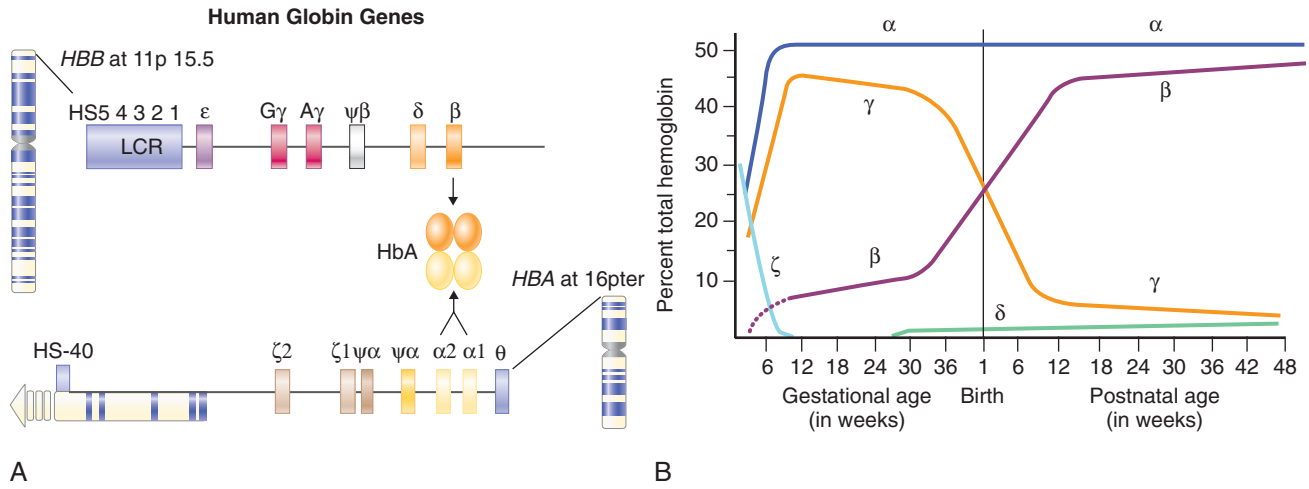

FIGURE 163-2. Human globin genes. A, The β-like globin gene cluster on chromosome 11 (*top*) and the α-like globin gene cluster (*bottom*) are shown. Two β-globin chains and two α-globin chains combine to form the normal hemoglobin A (HbA) tetramer, represented between the globin genes. Each globin chain contains one heme group, and oxygen transport takes place sequentially at the four iron-containing heme groups. Fetal hemoglobin (HbF) is composed of two α-globin and two γ-globin chains; the minor hemoglobin of adults, HbA_2, contains two α-globin and two δ-globin chains. Normally present at a level of only 2 to 3%, HbA_2 concentration is increased to 4 to 6% in most carriers of β-thalassemia. The ζ (*HBZ*) and ε (*HBE1*) genes are normally expressed only in the embryo. The 5′ ψα gene has been found to be expressed at a very low level and is now called the μ-globin gene (*HBM*). The θ-globin gene (*HBQ1*) is also expressed at low levels. Neither the θ nor the μ gene has been found in a functional hemoglobin. Any of the globin chains participating in hemoglobin formation may have a mutation altering its amino acid sequence. HbS, HbE, and HbC mutations affect the β-globin gene (*HBB*). Other mutations can affect the α-globin (*HBA1*, *HBA2*), γ-globin (*HBG1*, *HBG2*), and δ-globin (*HBD*) genes. LCR = locus control region. **B,** Expression of globin genes during development. The α chains are expressed throughout gestation and adult life, the fetal γ-globin chains are expressed predominantly in utero, and the β-globin and δ-globin genes are expressed mainly postnatally. This switching of gene expression patterns accounts for the different hemoglobins present in the embryo, fetus, and adult. It also accounts for the observation that disorders of α-globin can affect both fetus and adult, whereas β-globin chain diseases usually are not clinically apparent in the first months of life, when HbF levels are still high.

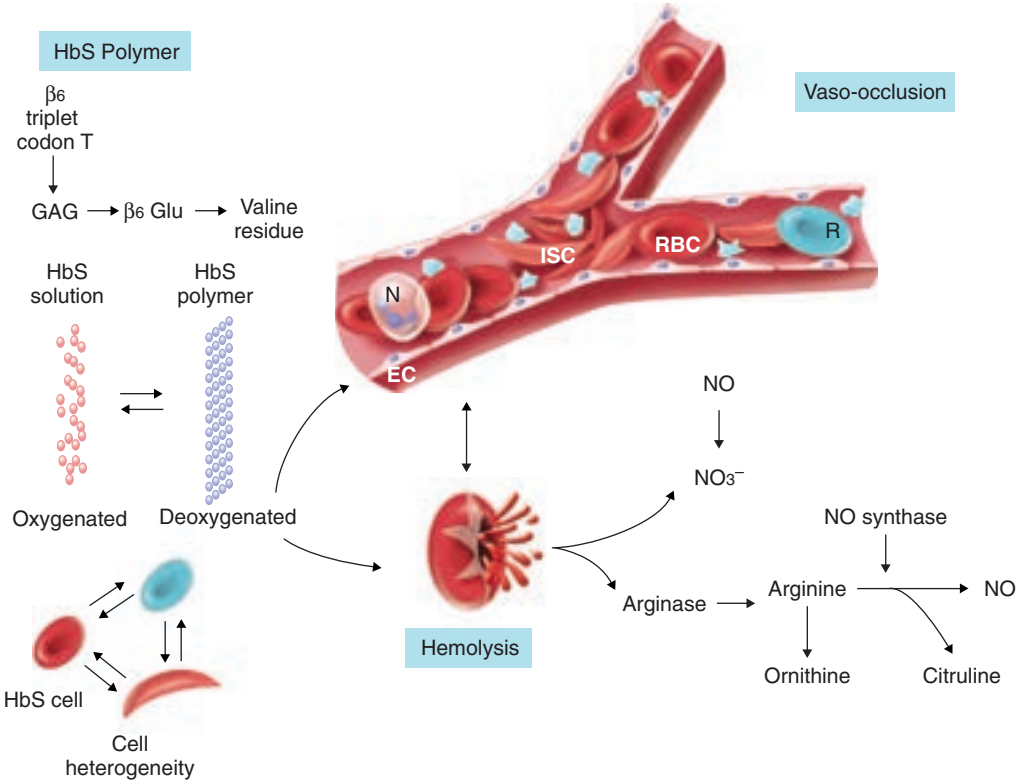

FIGURE 163-3. Pathophysiology of sickle cell disease. An adenine (A) to thymidine (T) transversion (A6T) at codon 6 in the β-hemoglobin gene on chromosome 11 (*HBB*) leads to the substitution of a glutamic acid codon by a valine codon. β^6 valine allows the hemoglobin S (HbS) molecule ($\alpha_2\beta_2^S$) to polymerize when it is deoxygenated. DeoxyHbS polymer injures the erythrocyte and leads to a heterogeneous population of sickle cells with damaged membrane cytoskeleton, reduced cation and water content, and altered distribution of membrane lipids. In the vasculature, sickle cells interact with endothelium and other blood cells to cause vaso-occlusion. Some damaged erythrocytes hemolyze intravascularly, thereby releasing heme into the plasma to scavenge nitric oxide (NO) and to reduce hemoglobin to methemoglobin and nitrate. NO, by binding soluble guanylate cyclase, converts cyclic guanosine triphosphate to guanosine monophosphate, thereby relaxing vascular smooth muscle and causing vasodilation. A state of reduced endothelial NO bioavailability in sickle cell disease impairs the homeostatic vascular functions of NO, such as inhibition of platelet activation and aggregation and transcriptional repression of genes transcribing cell adhesion molecules. Hemoglobin, heme, and heme iron catalyze the production of oxygen radicals and protein nitration, potentially further limiting NO bioavailability and activating endothelium. Lysed erythrocytes also liberate arginase, which destroys L-arginine, the substrate for NO production, providing another mechanism for endothelial NO deficiency. The normal balance of vasoconstriction versus vasodilation is therefore skewed toward vasoconstriction as well as endothelial activation and proliferation. EC = epithelial cell; ISC = irreversibly sickled cell; N = neutrophil; R = reticulocyte; RBC = red blood cell.

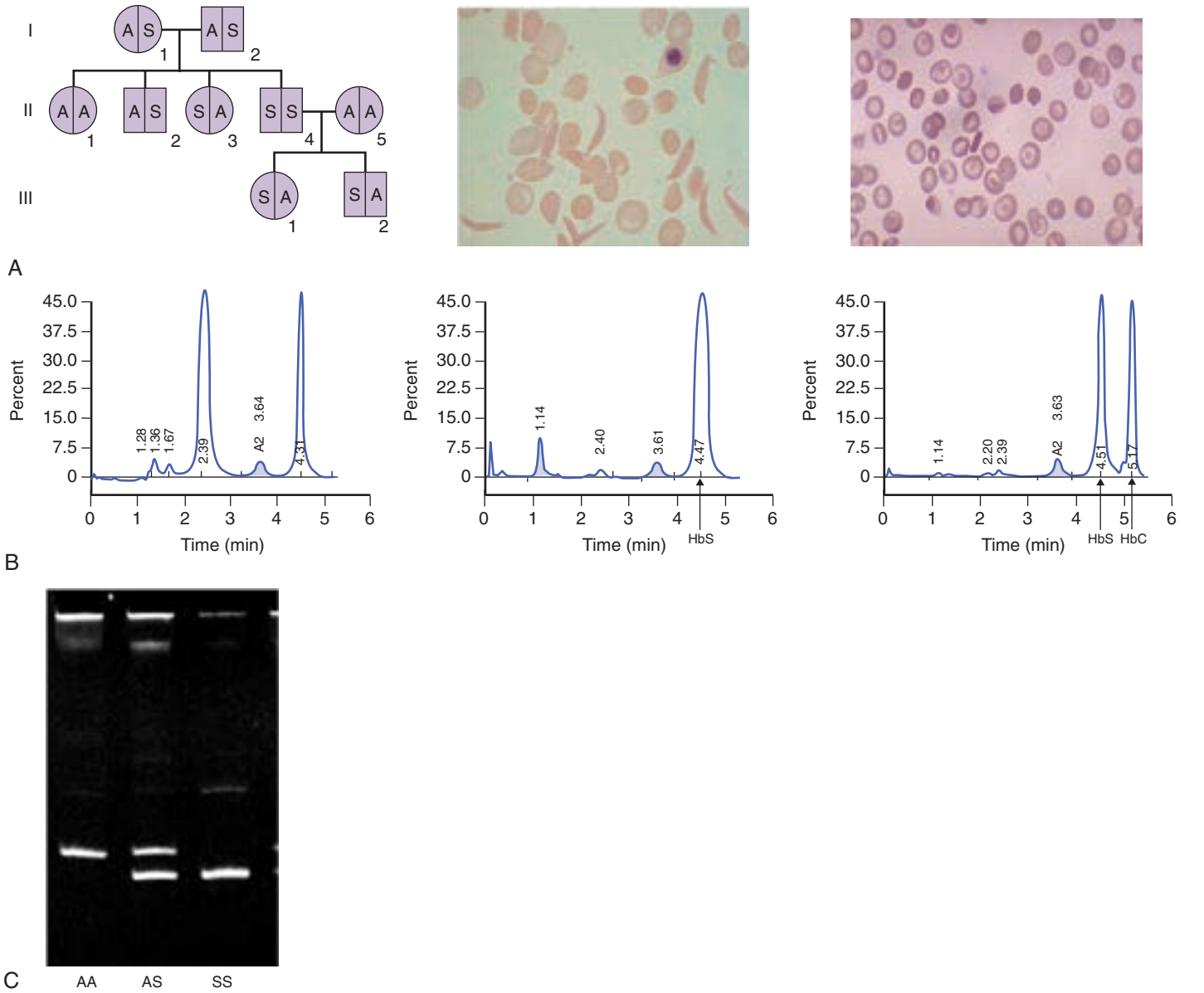

FIGURE 163-4. Diagnosis of sickle cell disease. A, A prototypical family structure in which both parents (I) have sickle cell trait and each offspring (II) has a 25% chance of having sickle cell anemia (SS). Each child of an affected parent (III) will have sickle cell trait (SA) if the other parent has a normal hemoglobin (AA) genotype. The blood films (*center and right*) are from patients with sickle cell anemia and HbSC disease, respectively. Note the irreversibly sickled cells in the former and the hemoglobin C (HbC) crystal and target cells in the latter. **B**, High-performance liquid chromatography profiles from patients with sickle cell trait (*left*), sickle cell anemia (*center*), and HbSC disease (*right*). **C**, Amplification refractory mutation system–based separation of the β-globin genes from a normal subject (AA), a carrier of sickle cell trait (AS), and a patient with sickle cell anemia (SS).

hypertension, priapism, leg ulcer, nephropathy, and stroke, are epidemiologically linked to the intensity of intravascular hemolysis, whereas other complications, such as painful episodes, acute chest syndrome, and osteonecrosis, are associated with high blood viscosity and the interactions among sickle cells, leukocytes, and the endothelium. Sickle vaso-occlusion and hemolysis are inextricably linked.

Vaso-occlusive events probably depend on features intrinsic to the sickle erythrocyte, such as polymer content and the degree of cellular damage, interacting with factors in the cell's environment, such as endothelial injury, vascular tone, and other blood cells. In the first hours of a painful episode, the number of dense erythrocytes falls; it rises again as pain resolves. These observations suggest the possibility that more deformable, more adherent cells might initiate vaso-occlusion, whereas dense cells become sequestered or destroyed in the microvasculature. Endothelial cells are responsive to many biologic modifiers that can be generated during sickle vaso-occlusive episodes and inflammation. Their activation and damage may be provoked by adherent sickle cells and shear stresses that cause release of oxidant radicals, expression of endothelin, and disturbed NO balance. Cellular damage enables adhesive interactions among sickle cells, endothelial cells, and leukocytes. Reperfusion injury can also induce endothelial activation and inflammation. The association of sickle and endothelial cells by a variety of adhesion molecules and their ligands may sufficiently delay cellular passage so that HbS polymerization, cell sickling, and vaso-occlusion happen before transit through the microvasculature is complete. Reticulocytes that are prematurely released from bone marrow display adhesive ligands that facilitate erythrocyte–endothelial cell interactions. Individuals with the greatest amount of hemolysis have the highest reticulocyte counts, and these adherent cells provide another link of hemolysis with vaso-occlusion. Neutrophils, which are modulators of inflammation and tissue damage, are increased in patients who have the acute chest syndrome, priapism, or stroke, and their numbers at baseline are a risk factor for survival.

CLINICAL MANIFESTATIONS

Sickle cell disease is a phenotype that results from different genotypes. Most patients have sickle cell anemia, HbSC disease, or HbS-β-thalassemia (see Table 163-1). Although the complications of disease are found in all genotypes, genotypes with higher cellular concentration of HbS are clinically more severe. Within milliseconds to seconds after HbS deoxygenation, depending on the intracellular concentration of HbS, HbS polymer appears in the sickle erythrocyte. In HbAS, each cell contains only 30 to 40% HbS, so polymer is not found under most conditions (see Fig. 163-4). Therefore, carriers have only subtle abnormalities and a normal life expectancy.

TABLE 163-2 FEATURES OF SICKLE CELL ANEMIA

Painful episodes: associated with higher hemoglobin and beneficially affected by high HbF
Acute chest syndrome: associated with higher hemoglobin and beneficially affected by high HbF
Stroke: associated with lower hemoglobin and little affected by HbF
Osteonecrosis: associated with higher hemoglobin and beneficially affected by high HbF
Priapism: associated with lower hemoglobin and little affected by HbF
Proliferative retinopathy: associated with higher hemoglobin and HbSC disease
Splenic infarction and sequestration more common in HbSC disease
Leg ulcers: associated with lower hemoglobin and beneficially affected by high HbF
Gallstones
Aplastic crisis due to B19 parvovirus
Osteopenia: bone marrow hyperplasia
Nutritional deficiencies: folic acid, zinc, calories
Pneumococcal disease and sepsis
Placental insufficiency

The features of sickle cell anemia change as life advances (Table 163-2). The switch from HbF to HbS underlies the clinical shift in life's first decade. This time is typified by acute problems: high risks of severe life-threatening infection, acute chest syndrome, splenic sequestration, and stroke. Chronic organ damage (renal failure, pulmonary hypertension, and late effects of previous cerebrovascular disease) becomes paramount in adults.

Most patients with sickle cell anemia have moderate anemia with a hematocrit between 25 and 30%. Some patients appear to have more severe hemolysis, with hematocrits less than 20%, marked reticulocytosis, and extreme elevation of serum lactate dehydrogenase. Patients with the most profound hemolysis appear more likely to have stroke, pulmonary hypertension, priapism, and leg ulcers (Fig. 163-5). Many patients with HbSC disease, especially adult men, have almost normal hematocrits and may have a higher incidence of sickle retinopathy, perhaps owing to their increased blood viscosity. Symptoms of anemia, such as weakness and dyspnea, are not the hallmarks of sickle cell disease, yet hemoglobin concentration can be a prognostic indicator for certain complications (see Table 163-2). A consequence of hemolysis is increased turnover of bile pigments, regulated in part by promoter polymorphisms in the uridine diphosphate glucuronosyltransferase 1A (*UGT1A*) gene, which is also associated with unconjugated hyperbilirubinemia and Gilbert syndrome (Chapter 147). As a result, more than half of all adults have cholelithiasis (Chapter 155). Hemolytic anemia places patients at risk for acute development of severe anemia when erythropoiesis is temporarily interrupted by parvovirus B19 infection (Chapter 371), which is the predominant cause of the aplastic crisis. Aplastic crisis is typified by a plummeting hematocrit, reticulocytopenia, and a bone marrow without erythroid precursors. It is a transient process, most common in children, and often requires blood transfusion to maintain circulatory competence until a spontaneous recovery follows. Rarely, if a patient's diet is inadequate, hemolysis-induced accelerated turnover of erythrocytes causes folic acid deficiency and megaloblastic anemia (Chapter 164).

Much of the epidemiologic data on the rate of complications in sickle cell disease discussed here antedates the widespread use of hydroxyurea in adults and its increasing use in very young children who have not yet developed complications of this disease. This, plus the use of chronic transfusions for prevention of stroke in many children, will change the phenotype of disease as these individuals advance to adulthood.

Ages 20 to 40 Years

Although any complication can occur at any age, certain events tend to predominate in different age groups. In the absence of hydroxyurea treatment or chronic transfusion, life's first decades are characterized by acute painful episodes, acute chest syndrome, and stroke. Delayed growth and sexual development, more severe in patients with sickle cell anemia than in those with HbSC disease, become major issues of concern to the adolescent, but sexual maturation is eventually achieved.

Psychosocial problems are common in adolescents with sickle cell disease. Difficulties with medical staff often begin in adolescence and frequently center on issues of pain management and inpatient stay.

The Painful Episode

Pain, presumed to be initiated by sickle vaso-occlusion, often starts in young children as the hand-foot syndrome or dactylitis: painful swelling of the hands and feet caused by inflammation of the metacarpal and metatarsal periosteum. Acute painful episodes are the most commonly encountered vaso-occlusive events in patients of all ages, but what triggers an acute painful episode is usually unknown. Commonly, painful episodes begin with little warning; some patients, however, may sense one in the offing. These episodes, which last hours to many days, can wax and wane in intensity and migrate from site to site. No useful laboratory test can tell whether a vaso-occlusive pain episode is occurring, and the history remains the best clue. Sickle cell pain is described as worse than postoperative or traumatic pain. Some women describe the pain of childbirth as paling in comparison with the pain experienced during painful episodes. These agonizing attacks of acute pain must be separated from chronic pain perhaps caused by osteoporosis of the spine, pain associated with osteonecrosis of the hips and shoulders, neuropathic pain, opioid-induced hyperalgesia, and the milder aches, pains, and soreness that are frequently present between severe episodes.

Almost all patients have acute painful episodes, but they vary greatly in number, severity, and frequency. Painful episodes are often stereotypical, affecting individuals in the same manner from episode to episode. Patients usually know whether the pain they are experiencing is different from their typical painful episode, and the wise physician should heed a patient's advice about the need for hospitalization or the likelihood that the pain has an alternative explanation.

In studies antedating the widespread use of hydroxyurea (see later), about 40% of patients did not have pain requiring a hospital visit in a given year, whereas 3% had more than six painful episodes per year. Having more than three pain episodes requiring hospitalization per year was associated with increased mortality among patients 20 years and older. Studies based on pain diaries suggest that pain is present on about half of all days and that most episodes are managed at home and not in a medical setting, so hospital visits underestimate the frequency of pain.

Most studies show that HbF levels are inversely related to the frequency of painful episodes. Concurrent α-thalassemia may increase the pain rate because it is associated with increased hematocrit. The day-to-day management of sickle cell disease often equates with the management of acute and chronic pain.

The pain accompanying acute chest syndrome, acute cholecystitis, splenic sequestration crisis, splenic infarction, or right upper quadrant syndrome may sometimes be mistaken for an uncomplicated pain episode. Acute painful episodes often precede the acute chest syndrome by 24 to 72 hours, and pain episodes occasionally end with multiorgan failure. Unexplained death can occur during acute painful episodes, perhaps as a result of an arrhythmia secondary to unrecognizable myocardial damage or perhaps as a sequela of pulmonary hypertension. Currently, it is not possible to foretell whether a "usual" pain episode will have an unexpected mortal outcome or presage acute chest syndrome, but the presence of atypically severe pain or an uncommonly high leukocyte count, low hematocrit, and thrombocytopenia should be cause for extra scrutiny.

Cerebrovascular Disease

A major complication of sickle cell anemia in early life is cerebrovascular disease that includes silent cerebral infarction and stroke caused by stenosis and occlusion of large vessels (Chapter 407).[6] Sickle erythrocytes and anemia-related high blood flow velocity lead to vascular damage. Hemorrhagic stroke in adults is a result of rupture of aneurysms or moyamoya disease, a proliferation of small vessels secondary to stenotic lesions (Chapter 407), and is associated with a mortality rate of more than 20%. Stroke is most common in HbS homozygotes, with much lower rates among those with HbSC disease or HbS-β^+-thalassemia. In the pre-hydroxyurea and pre–transcranial Doppler screening era, the incidence of stroke in sickle cell anemia was approximately 0.5 event per 100 patient-years until the age of 40 years, and the risk of having a first stroke was 11% by age 20 years, 15% by age 30 years, and 24% by age 45 years. These statistics have changed as a result of the use of transfusion to reduce the occurrence of stroke in children found to be at high risk by transcranial Doppler screening. Severe anemia, acute chest syndrome, and elevated systolic blood pressure are associated with ischemic strokes, whereas an elevated leukocyte count is a risk factor for hemorrhagic stroke. Concurrent α-thalassemia may protect patients with sickle cell anemia from stroke, perhaps because these patients have less hemolysis and a higher hematocrit. Subclinical neurologic events and silent cerebral infarction are even more common than stroke and are associated with decreased intellect and an increased likelihood of overt stroke. Neurologically

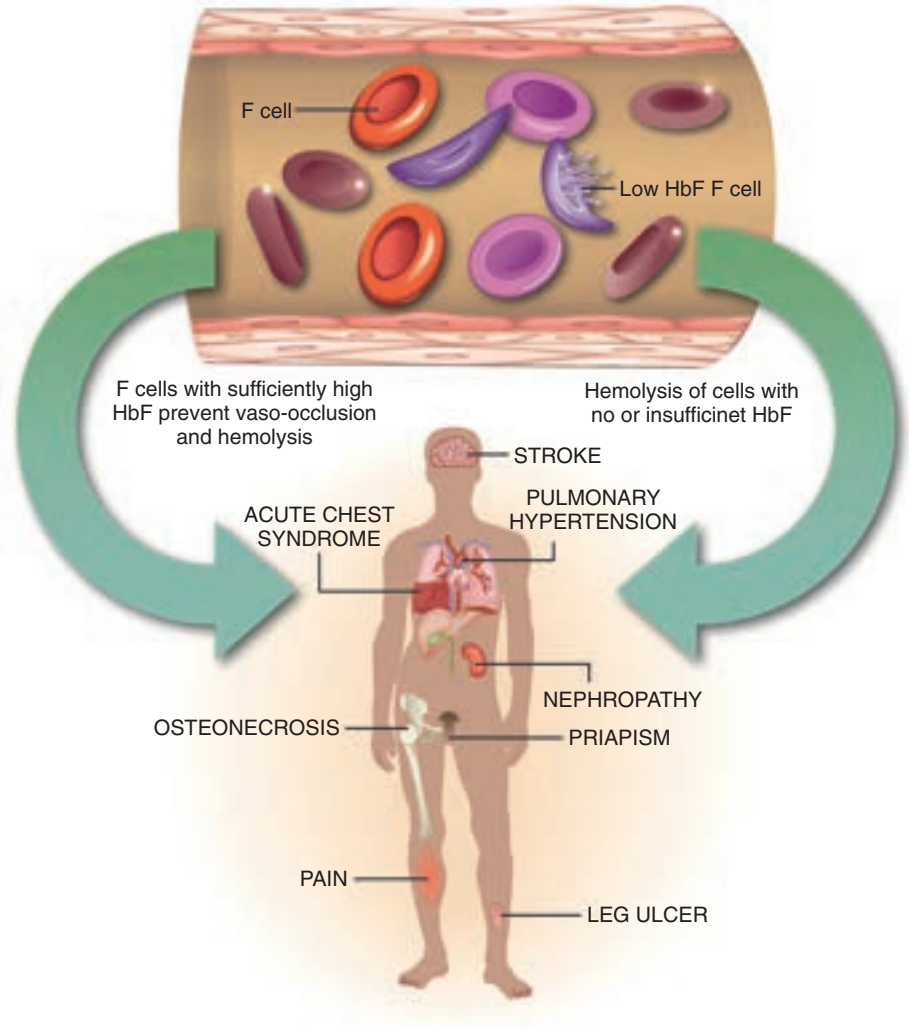

FIGURE 163-5. F cells and the subphenotypes of sickle cell anemia. The amount of HbF/F cell and the distribution of F cells vary among patients. In the example shown, some F cells have sufficiently high HbF concentration (bright red and magenta F cells), and their HbS will not polymerize or contains little polymer at physiologically relevant oxygen saturations. In some F cells (darker and brownish red), HbS polymerization will occur at venous oxygen saturations. Other cells with little or no HbF will have HbS polymers and sickle at high oxygen saturation; some of these cells will hemolyze intravascularly. Such cells, by releasing their hemoglobin intravascularly, can provoke certain subphenotypes of sickle cell disease. HbF = fetal hemoglobin.

intact adults without a stroke history have poorer cognitive performance than that of healthy controls.[7]

Acute Chest Syndrome

Acute chest syndrome, characterized by fever, chest pain, wheezing, cough, hypoxia, and a new lung infiltrate, is a sometimes lethal complication that affects more than half of all patients with sickle cell anemia.[8] It is the second most common reason for hospitalization and is a frequent cause of death in adults. The syndrome is more frequent in children, in whom its course is often mild, than in adults, in whom it tends to be more severe.

Commonly, acute chest syndrome develops after several days in individuals hospitalized for an acute painful episode (see earlier discussion). Acute chest syndrome also often occurs postoperatively, even when patients are properly prepared with blood transfusion. Other causes include rib infarction with atelectasis and regional hypoxia; fat emboli from the bone marrow (Chapter 98); infection with chlamydia, parvovirus B19, or other viral agents; microvascular or large-vessel in situ thrombosis; thromboembolic disease[9]; and vascular injury and inflammation. Fat embolism can be identified by finding lipid within pulmonary macrophages obtained by bronchopulmonary lavage, but this nonstandardized test is not recommended.

In most cases of acute chest syndrome, a cause cannot be found early enough to guide treatment.

Acute Anemia

Acute anemia can result from sequestration of blood in the spleen or liver; an aplastic crisis caused by parvovirus B19 infection (Chapter 371); or a severe vaso-occlusive event, such as acute chest syndrome or multiorgan failure. Transfusion may be needed. Megaloblastic arrest of erythropoiesis is uncommon if the diet is adequate in folic acid.

Infection

Because patients with sickle cell anemia are functionally asplenic early in life and hyposplenic later, they have increased susceptibility to infection with encapsulated bacteria. Persistent splenomegaly but not normal splenic function is common in patients with sickle cell disease in Africa, related to endemic malaria, and in Saudi Arabia, where half of the sickle cell disease population has α-thalassemia. Splenomegaly and splenic function often persist in patients with HbSC disease—hence the reduced incidence of infection but increased risk of splenic sequestration and infarction in adults. Prevention of mortality from pneumococcal infection is the basis for screening of newborns for sickle cell disease and the use of prophylactic oral penicillin in affected individuals. Pneumococcal vaccines (Chapter 18) are also recommended.

Pregnancy

There is no absolute contraindication to pregnancy for patients with sickle cell anemia, and fertility is probably normal. All approved methods of contraception can be used satisfactorily.

These are "high-risk" pregnancies, and close cooperation between the hematologist and obstetrician generally achieves good results; but the rates of pyelonephritis, pregnancy-induced hypertension (Chapter 239), and cesarean section are increased, and babies are more likely to have low birthweight. Limited data suggest that with good prenatal care, transfusions do not improve the outcome.

Osteonecrosis and Bone Diseases

Osteonecrosis of the hip and shoulder joints (Chapter 248) affects about half of all patients with sickle cell anemia or HbSC disease. Its onset is insidious but progressive, and most patients with early-stage disease progress to collapse of the femoral head within 2 years. Osteonecrosis of the hip usually is manifested with pain in and around the affected joint or at times with spasm of the surrounding muscles. Patients with higher hematocrits and with sickle cell anemia–α-thalassemia have the highest prevalence of osteonecrosis. Osteonecrosis can be detected very early by magnetic resonance imaging, but only more advanced disease is visible on plain radiographs.

Osteomyelitis (Chapter 272) is often difficult to distinguish from bone infarction. Osteomyelitis is usually caused by staphylococcal infection, but salmonella infection is a particular cause of sickle cell osteomyelitis.

Leg Ulcers

About 5 to 10% of patients with sickle cell anemia older than 10 years develop leg ulcers, but leg ulcers are rare in HbSC disease, in HbS-β^+-thalassemia, in Saudi Arabs with sickle cell disease, and in children before the age of 10 years. In the tropics, leg ulcers are more common. Small and superficial leg ulcers heal spontaneously with rest and careful local hygiene. Control of local inflammation and infection remains the mainstay of treatment. Dressing the ulcer with an Unna boot protects the involved area and is a reasonable method of conservative management. Deep, large, painful ulcers may require large doses of narcotic analgesics, prolonged bedrest, and even surgery.

Priapism

Priapism, a prolonged undesirable painful erection, may be seen in 40% of men with sickle cell anemia. Priapism in sickle cell anemia is usually bicorporal, with only the corpora cavernosa affected. Venous outflow is obstructed rather than arterial flow increased. In bicorporal priapism, the glans remains soft, and urination is normal. Recurrent, self-limited attacks of priapism can last for several hours with tolerable discomfort. These episodes have been termed stuttering priapism and usually have a nocturnal onset. Erectile function is usually preserved between attacks. Major episodes of priapism often but not always follow a history of stuttering attacks, last for days, and can be excruciatingly painful; they often end in impotency. Affected patients have a higher incidence of stroke, pulmonary hypertension, renal failure, leg ulcers, and premature death than individuals without priapism do, perhaps reflecting the severity of vasculopathy.

Digestive Diseases

Sickle hepatopathy, hepatic crisis, and right upper quadrant syndrome are terms applied to sickle cell–associated liver disease.[10] Liver disease may be related to intrahepatic and extrahepatic cholestasis, viral hepatitis (Chapter 148), cirrhosis, hypoxia, infarction, erythrocyte sequestration, iron overload (Chapter 212), or drug reactions (Chapter 150). Bilirubin levels can top 60 mg/dL, and such high levels are a poor prognostic sign portending liver failure. Differentiation among these potential causes can be difficult. Increased bilirubin levels (Chapter 147) are often a manifestation of hemolytic anemia and are related to polymorphisms of the UGT1A1 promoter.

Gallstones

Cholelithiasis (Chapter 155), a consequence of the accelerated bile pigment turnover typical of hemolytic anemia, can appear in the first decade of life, and more than half of all adults are affected. Depending on the degree of calcification, pigmented gallstones may be either radiopaque or radiolucent. Ultrasonography is the preferred means of detection, and laparoscopic cholecystectomy is the preferred method of dealing with symptomatic stones. Documented episodes of acute cholecystitis and typical obstructive jaundice are much less frequent than the presence of stones. If stones are asymptomatic or symptoms and laboratory findings suggesting cholecystitis are equivocal, careful observation may be the best course.

Beyond the Fifth Decade

Pulmonary Hypertension

Echocardiographic studies show that 30 to 43% of adults with sickle cell anemia have a tricuspid regurgitant jet velocity of more than 2.5 m/second (Chapter 55), and 5 to 10% of patients will have cardiac catheterization–documented pulmonary hypertension (Chapter 68). Both elevated tricuspid regurgitant jet velocity and true pulmonary hypertension are associated with a six- to ten-fold increased risk in mortality.[11] Increased tricuspid regurgitant jet velocity and the coexistence of relative systemic hypertension, renal disease, and intimal and smooth muscle proliferative changes in conduit vessels suggest the presence of a more widespread vasculopathy, which may be responsible for the observed mortality risk. Sickle cell anemia patients with pulmonary arterial hypertension have milder hemodynamics and symptoms compared with patients with idiopathic pulmonary arterial hypertension, particularly early in the course of disease, yet their survival is similar.

TABLE 163-3 RENAL ABNORMALITIES IN SICKLE CELL DISEASE*

Distal nephron
Impaired urine concentrating ability (hyposthenuria)
Impaired urine acidification—incomplete renal tubular acidosis
Impaired K^+ excretion
Hematuria
Papillary necrosis
Proximal tubule
Increased phosphate reabsorption
Increased β_2-microglobulin reabsorption
Increased uric acid secretion
Increased creatinine secretion
Hemodynamic changes
Increased glomerular filtration rate
Increased renal plasma flow
Decreased filtration fraction
Glomerular abnormalities
Proteinuria
Nephrotic syndrome with focal glomerular sclerosis
Chronic renal failure

*In carriers of sickle cell trait, because of its hypertonicity and oxygen content, HbS can polymerize in the renal medulla and lead to hyposthenuria, possibly an increased risk of urinary tract infection during pregnancy, papillary necrosis, and hematuria. A rare tumor, medullary carcinoma, arises from distal nephrons and is associated with sickle cell trait.

Nephropathy

Hyposthenuria is present in almost all patients with sickle cell anemia and even in most people with HbAS (Table 163-3). Clinically, the loss of urine concentrating ability is not important unless access to fluid is restricted. Isosthenuria, distal renal tubular acidosis, and impaired potassium excretion are signs of medullary dysfunction.

Glomerular hyperfiltration, increased creatinine secretion, and a very low serum creatinine concentration are characteristic of young patients with sickle cell anemia, so renal dysfunction can be present even with normal serum creatinine values. Glomerulopathy begins very early in life, but an increasing prevalence of renal failure is a hallmark of an aging population of sickle cell anemia patients. About 4% of patients with sickle cell anemia and 2% of those with HbSC develop renal failure, at median ages of 23 and 50 years, respectively. Among sickle cell anemia patients, 60% of those older than 40 years have proteinuria and 30% have renal insufficiency. Nephrotic syndrome is found in 40% of adults with creatinine levels above 1.5 mg/dL. Survival time for patients with sickle cell anemia after the diagnosis of sickle renal failure is 4 years, even with dialysis.

Eye Disease

Proliferative sickle retinopathy is present in less than 20% of patients with sickle cell anemia but in more than 40% of those with HbSC disease by the third decade of life. Vitreal hemorrhage and retinal detachment can occasionally lead to visual loss, but proliferative lesions may regress spontaneously. Screening for proliferative retinopathy by fluorescence angiography is recommended in patients with HbSC disease to guide possible laser photocoagulation.

Cardiovascular Complications

Cardiac complications of sickle cell disease are complex and can be manifested as both right and left ventricular systolic and diastolic dysfunction, elevated cardiac output, cardiomegaly, and myocardial ischemia. Progressive heart damage may result from iron overload in heavily transfused and poorly chelated patients, although this seems far less common than in β-thalassemia.

The heart is usually enlarged, with a hyperactive precordium and systolic ejection murmurs. Myocardial infarction is rare and when present suggests small-vessel disease.

Patients with sickle cell anemia usually have blood pressures that are in the normal range yet inappropriately high compared with controls who have similar levels of anemia. "Relative" hypertension in sickle cell anemia may reflect endothelial cell damage and increased NO scavenging by plasma hemoglobin. Survival is decreased, and the risk of stroke is increased as blood pressure rises. Treatment goals of 120/80 mm Hg or lower are generally the same as for other patients (Chapter 67).

DIAGNOSIS

Normocytic, microcytic, or macrocytic hemolytic anemia with reticulocytosis, increased levels of lactate dehydrogenase and aspartate aminotransferase, and a compatible clinical history should suggest the presence of sickle cell disease. Nevertheless, because of the multiple genotypes and considerable clinical heterogeneity of each genotype, milder cases may not be diagnosed for many years.

The Blood

In sickle cell anemia, the erythrocytes are normocytic or macrocytic, depending on the reticulocyte count. Microcytosis in a suspected case of sickle cell disease can be seen early in life, when iron deficiency has developed, or when β-thalassemia or α-thalassemia coexists with HbS. Sickled cells are usually present in the peripheral blood smear in sickle cell anemia and in HbS-β^0-thalassemia (see Fig. 163-4A) but are less common in other forms of sickle cell disease. In HbSC disease, target cells are prominent, and HbC crystallizes in some cells (see Fig. 163-4A). Irreversibly sickled cell numbers remain relatively constant over time, and their presence has no value for establishing whether a patient is experiencing a vaso-occlusive episode.

Hemoglobin Composition of the Blood

After 1 year of age, hemoglobin fractions are sufficiently stable to be relied on for diagnosis, but the high HbF concentrations of early infancy often make the results of hemoglobin analysis at that time difficult to interpret. In patients with sickle cell anemia, except in infancy, HbS almost always forms more than 80% of the hemolysate. HbS is best detected by high-performance liquid chromatography, which is also the method of choice for quantitation of the hemoglobin fractions in newborns and adults (see Fig. 163-4B). High-performance liquid chromatography provides excellent resolution of hemoglobin fractions, gives quantitative results, and is automated. HbF levels in adults average about 6% but can vary between 1 and 20%. DNA-based methods of detecting HbS are specific but usually are not needed for uncomplicated cases. Nevertheless, they are necessary for antenatal diagnosis and, sometimes, for genetic counseling (see Fig. 163-4C).

Family Studies

Hemoglobinopathies are inherited as codominant traits (see Fig. 163-2), implying that both normal and mutant alleles are expressed and are easily detectable. However, the sickle cell phenotype (see Table 163-1) is present only in homozygotes for HbS and in compound heterozygotes like HbSC disease. Family studies (see Fig. 163-4) can suggest a patient's hemoglobin genotype.

PREVENTION AND TREATMENT

Rx

Primary Prevention

In populations with a high prevalence of the HbS gene, heterozygote detection is simple, but there is little proven benefit of a broad screening effort to detect carriers. A preferred approach consists of educational programs about sickle cell disease and HbAS, followed by counseling for couples who are planning families. These couples are offered the choice of testing, after which the risks of having affected fetuses can be discussed and the availability of antenatal diagnosis presented.

General Measures

Sickle cell disease is a chronic disorder for which good nutrition and timely immunizations are critical. Work should be encouraged.

Children beyond the age of 5 years do not routinely need continued antibiotic prophylaxis. Neonatal screening to detect newborns with sickle cell disease allows early administration of prophylactic penicillin and antipneumococcal immunization. These measures reduce the incidence of and mortality from pneumococcal bacteremia in children younger than 5 years who have sickle cell anemia.

Because of increased rates of red blood cell production and inadequate nutrition, folic acid, 1 mg daily, is generally recommended but may not be necessary with a good dietary intake. There is no evidence that high concentrations of inhaled oxygen are of preventive value.

The transition from pediatric to adult care is a time of heightened vulnerability, and this period should be managed jointly by pediatric and adult providers in a structured program.

Hydroxyurea

Hydroxyurea, the sole drug approved by the U.S. Food and Drug Administration for treatment of sickle cell anemia, increases HbF in most patients. In a multicenter trial in adults with sickle cell anemia, hydroxyurea reduced the incidence of pain and acute chest syndrome by almost 50%, with little risk during more than 17.5 years of observation.[■] In follow-up studies, cumulative mortality was reduced almost 40%, and the favorable result was related to the ability of the drug to increase HbF and to reduce painful episodes and the acute chest syndrome.[■] Children have a more robust HbF response to hydroxyurea than adults do, and a study in children with a mean age of 13 months showed clinical results similar to those of the adult multicenter trial, whereas hemoglobin concentration and HbF levels were higher than those of placebo-treated controls.[■] Cancer and leukemia have been reported in adults with sickle cell disease treated with hydroxyurea, but whether the incidence is higher than in the general population is not known. Hydroxyurea should be used in nearly all adults and children with sickle cell anemia and HbS-β^0-thalassemia (Table 163-4).[12] However, not all patients who might benefit from this treatment receive it. A controlled clinical trial of hydroxyurea in HbSC disease has not been done.

Treatment of Common Complications

Painful Episodes

A decision as to whether hospitalization is needed can be made after an assessment of the duration and severity of the pain and the response to treatment. Associated factors, such as excessive tachycardia, hypotension, body temperature higher than 38.3°C (101°F), marked leukocytosis, fall in the hematocrit and platelet count, hypoxia, or new infiltrate on chest radiography, should prompt admission. On physical examination, there is sometimes localized swelling and pain over an involved bone. Low-grade fever and a mild increase in leukocytosis above baseline can accompany uncomplicated painful episodes, but higher temperature elevations may point to infection or extensive tissue damage.

The cornerstones of pain management (Chapter 30) are fluid replacement and opioid analgesics. Because almost every patient is hyposthenuric, urinary output in patients with sickle cell anemia may exceed 2 L/day, making them susceptible to dehydration. Pain is often accompanied by reduced fluid intake and increased water losses, so increased fluid intake is essential. Administration of 5% dextrose in water or 0.25 to 0.5 normal saline should be used for

TABLE 163-4 HYDROXYUREA TREATMENT IN SICKLE CELL ANEMIA*

BASELINE EVALUATION
Blood counts, RBC indices, HbF level, serum chemistries, pregnancy test, willingness to adhere to all recommendations for treatment, not receiving chronic RBC transfusions
INITIATION OF TREATMENT
Hydroxyurea 10-15 mg/kg/day or, for adults, 500 mg every morning for 6-8 wk
CONTINUATION OF TREATMENT
If counts are acceptable by CBC every 2 wk (granulocytes, ≥2000/mm^3; platelets, ≥80,000/mm^3), escalate dose in increments of 200 to 500 mg every 6-8 wk.
When a stable nontoxic dose of hydroxyurea is reached, CBC may be done at 4- to 8-wk intervals.
Most good responses require 1000 to 2000 mg/day, and a final dose of 30 mg/kg/day should be the maximum.
GOALS OF TREATMENT
Less pain, increase in HbF (usually measured every 6-8 wk) or MCV, increased hematocrit if severely anemic, acceptable toxicity
Failure of HbF to increase may be due to biologic inability to respond to treatment or, more often, to poor compliance with treatment. If compliance is documented, the dose can be increased cautiously to 2000-2500 mg/day (maximum dose, 30 mg/kg).

CBC = complete blood count; HbF = fetal hemoglobin; MCV = mean corpuscular volume; RBC = red blood cell.
*Special caution should be exercised in patients with compromised renal or hepatic function and in those who are habituated to narcotics. Contraception should be practiced by both men and women. Without chronic RBC transfusions or an intercurrent illness suppressing erythropoiesis, a trial period of 6 to 12 months is probably adequate.

initial fluid replacement. Although needs vary, hydration and serum electrolyte values should be monitored closely to avoid iatrogenic heart failure or electrolyte imbalance. The daily fluid intake should be approximately 3 to 5 L for adults and 100 to 150 mL/kg for children. Oxygen should be reserved for patients who are hypoxic or have acute respiratory distress. Infection should always be considered and treated early if it is present. Treatment with intravenous magnesium sulfate in addition to standard therapy for vaso-occlusive episodes in children aged 4 to 18 years has been found to have no effect on hospital length of stay, pain scores, or cumulative analgesia.

Analgesic management presupposes that other treatments, such as hydration, oxygen, and antimicrobial agents, are used if needed. The key to successful pain management is individualized treatment and dosing, taking into account prior pain management and prior use of opioids. Patient-controlled analgesia and a scheduled regimen of drug dosing are preferable, and analgesics on an as-needed basis should be avoided. Frequent reassessment of the effects of treatment is paramount so that opioid doses can be titrated for pain relief and tapered when relief is obtained. Morphine and hydromorphone are the principal opioids used; meperidine should be avoided because its metabolites (e.g., normeperidine) can cause central nervous system excitation. Pain management (Chapter 30) is complicated by the influence of learned pain behavior, pain memories, and pain therapy–induced pain. Management proves extremely difficult in perhaps 10% of all patients, and enormous doses of opioids are often required for relief. High opioid doses are associated with allodynia (pain produced by a non-noxious stimulus to the skin), opioid-induced hyperalgesia, and neuropathic pain.

Stroke

For new strokes due to cerebral infarction, after initial stabilization and transfusion, chronic red blood cell transfusion reduces the chance of recurrence. Long-term management of hemorrhagic stroke is unclear, and whether transfusion reduces its recurrence is unknown. Increased intracranial flow velocity, measurable by transcranial Doppler flow measurement only in children, increases the risk of stroke, but its sensitivity is only 10%, with a far from perfect specificity. Children found to be at risk for stroke by transcranial Doppler flow velocities should be started on chronic transfusions, but it is unclear whether or when such transfusions may be discontinued.

Acute Chest Syndrome

Routine treatments for acute chest syndrome initially include bronchodilators (e.g., albuterol nebulizer, 0.25 mL in 2.5 mL normal saline, or albuterol metered-dose inhaler during the acute episode), incentive spirometry, empirical antimicrobial agents in febrile patients as used for community-acquired pneumonia (e.g., ceftriaxone or azithromycin or levofloxacin for 5 to 7 days [Chapter 97]), and supplemental oxygen when hypoxia is noted by continuous or frequent monitoring of blood oxygen saturation. Opioids are often needed, but their dose should be titrated carefully to avoid respiratory depression and worsening of hypoxia.

Blood transfusion is the cornerstone of treatment when a patient becomes hypoxic, develops respiratory distress, has a clinically significant fall in the hematocrit and platelet count or increase in leukocyte count, or shows any sign of multiorgan failure, such as impaired mentation, rhabdomyolysis, renal failure, or liver failure. Both simple transfusion and exchange transfusions appear to reverse many adverse findings of the acute chest syndrome, but controlled studies have never tested the superiority of either method. Although the death rate in acute chest syndrome is less than 10%, a few patients have a rapidly deteriorating course with sudden development of the acute respiratory distress syndrome (Chapter 104), as manifested by increased oxygen requirements, extensive pulmonary opacification, and multiorgan failure. Excessive hydration, fat emboli, and widespread vaso-occlusion are potential contributing causes. Successful management of severe acute chest syndrome and acute respiratory distress syndrome requires close coordination among physicians and nurses. Some patients have repeated episodes of severe acute chest syndrome, and chronic transfusion can reduce the recurrence rate. Hydroxyurea also reduces the rate of acute chest syndrome by about 50%.

Osteonecrosis and Bone Disease

Treatment with reduced weight bearing, nonsteroidal anti-inflammatory drugs, and physical therapy is the mainstay of conservative management but does not retard progression of osteonecrosis and bone disease. Total hip arthroplasty can be successful, but about one third of prostheses fail within 4 to 5 years.

Diffuse osteoporosis (Chapter 243) is usually present, and osteomalacia (Chapter 244) due to vitamin D deficiency is common in both children and adults. If vitamin D deficiency is present, treatment with calcium (1000 mg PO daily) and vitamin D (50,000 IU PO every week for 2 months, then 50,000 IU PO every other week) is reasonable.

Priapism

Conservative treatment of priapism includes analgesics and hydration. No evidence supports the use of transfusion. Aspiration and irrigation of the corporeal bodies should be performed if more than 4 hours have elapsed from the onset of erection if the episode differs from prior episodes of stuttering priapism. Operative treatment, which should be considered after 24 to 48 hours of priapism, includes the creation of shunts between the corpora cavernosa and corpus spongiosum. Oral α-adrenergic agonists, such as etilefrine and pseudoephedrine, and phosphodiesterase-5 inhibitors, like sildenafil and tadalafil, have been used successfully to prevent severe acute attacks but are not useful for treating acute major priapism once it has occurred.[13]

Pulmonary Hypertension

Pulmonary arterial hypertension can be determined definitively only by right-sided heart catheterization, although echocardiography or measurement of blood N-terminal pro-brain natriuretic peptide levels can suggest its existence. As these patients are often asymptomatic early in the course of their disease, a consensus group convened by the American Thoracic Society recommended echocardiography for risk stratification in sickle cell disease adults every 1 to 3 years.[14]

Management of patients with few symptoms and minimally elevated tricuspid regurgitant jet velocity is not informed by clinical trials. For symptomatic individuals, optimization of hydroxyurea, transfusions, anticoagulation, bosentan, and epoprostenol have all been used. A controlled trial of sildenafil based on tricuspid regurgitant jet velocity and low exercise capacity was stopped early because of increased hospitalization for pain[A5] (Table 163-5).

Renal Disease

In a small randomized trial, 6 months' treatment with 25 mg/day of captopril caused a 37% reduction in microalbuminuria, compared with a 17% increase in placebo-treated patients; such treatment would be reasonable in patients with known microalbuminuria, but whether screening for this or long-term treatment is worthwhile is unknown. Nonsteroidal anti-inflammatory drugs reduce the glomerular filtration rate in sickle cell anemia and should be avoided in older individuals with creatinine levels of 1.2 mg/dL or higher. Dialysis and renal transplantation are used in end-stage sickle cell nephropathy, but with outcomes less favorable than in other types of renal failure.

Surgery and Anesthesia

Blood transfusion should be given before all surgeries requiring general anesthesia and selected other surgeries. Simple transfusion to a hematocrit of about 30% before surgery is as effective as exchange transfusion in preventing postoperative complications and causes fewer transfusion-related complications. In some low-risk surgeries, preoperative transfusion might not be necessary.

Implantable infusion ports and catheters have higher risks of complications in sickle cell anemia, including thrombosis of large veins and bacteremia. Low-dose warfarin may retard thrombosis of these devices.

Red Blood Cell Transfusion and Iron Chelation Therapy

Acute transfusions of packed red blood cells can be life-saving, and chronic transfusions reduce the incidence of stroke and severity of most complications of sickle cell disease. However, repeated transfusions produce iron overload, alloimmunization, loss of venous access, and viral infection. Whether exchange transfusion is preferable to simple transfusion in the acute chest syndrome or other acute complications has not been tested in clinical trials. For severe symptomatic anemia and stroke prophylaxis, simple transfusions are preferred. The customary level of chronic stable anemia alone is seldom

TABLE 163-5 PULMONARY COMPLICATIONS OF SICKLE CELL DISEASE

ACUTE CHEST SYNDROME
Diagnosis: chest pain, fever, cough, wheezing, new infiltrate on chest radiograph, hypoxemia
Management: simple or exchange transfusions, pain relief, oxygen if hypoxemic, bronchodilators, incentive spirometry, antibiotics (broad coverage)
PULMONARY HYPERTENSION
Screening evaluation: echocardiography with tricuspid regurgitant jet velocity ≥ 2.5 m/sec; elevated N-terminal pro-brain natriuretic peptide (≥160 pg/mL); reduced 6-minute walk distance (≤350 m)
Definitive diagnosis: screen by echocardiography beginning at age 18 yr; repeat echocardiography periodically according to symptoms or every 1 to 3 years. If tricuspid regurgitant jet velocity > 3.0 m/sec, refer for right-sided heart catheterization that provides a definitive diagnosis.
Management: Consider anticoagulation, treatment of iron overload and nocturnal hypoxemia. Optimize hydroxyurea, consider chronic transfusions and pulmonary vasodilator therapy.
ASTHMA
ABNORMAL PULMONARY FUNCTION

an indication for transfusion. With aging and the onset of renal failure, anemia worsens and can become symptomatic. Transfusion may become necessary, although judicious use of erythropoietin (e.g., darbepoetin, 0.45 µg/kg every 2 weeks, increased as needed) can often restore the hematocrit to prerenal failure levels only and should not be targeted at even higher levels because of the potential adverse effects of hyperviscosity.

It is advised that patients undergo erythrocyte phenotyping to determine their red blood cell antigens before embarking on a chronic transfusion program. Otherwise, alloimmunization occurs in about one fourth of frequently transfused patients. In the presence of multiple alloantibodies, it may be difficult to find compatible blood.[15]

With repeated transfusion, iron overload inevitably develops and can result in heart and liver failure and many other complications (Chapter 212). Serum ferritin is an inaccurate means of estimating tissue iron burden. For deciding when to begin iron chelation and for observing the effects of chelation, magnetic resonance imaging is now the standard, leaving little indication for liver biopsy to measure iron concentration.

Chelation of excessive iron can be achieved with deferasirox (20 mg/kg orally daily).■ Increases in serum creatinine concentration and proteinuria occur in about 40% of patients. Desferrioxamine is a parenteral chelator and is usually started at a dose of 25 to 30 mg/kg given subcutaneously five times per week as 8- to 12-hour continuous infusions (Chapter 212). Side effects include gastrointestinal and skin reactions, ototoxicity, retinal toxicity, bone and growth abnormalities, and *Yersinia* infection resulting from the sudden mobilization of iron. Another oral chelator, deferiprone, is less potent than deferasirox or desferrioxamine but is especially useful when magnetic resonance imaging shows high levels of cardiac iron. Side effects requiring discontinuation of deferiprone, seen in 5 to 10% of patients, include agranulocytosis, neutropenia, arthropathy, and gastrointestinal symptoms.[16]

Stem Cell Transplantation

Successful stem cell transplantation (Chapter 178) can cure sickle cell anemia, but only about 10% of patients have suitable donors. Myeloablative stem cell transplantation carries a 5 to 10% mortality rate and is poorly tolerated in adults. A nonmyeloablative regimen that uses total body irradiation and treatment with alemtuzumab and sirolimus has led to stable mixed chimerism in about 50% of adults transplanted with an HLA-identical family donor, with no graft-versus-host disease observed.[17] In another small study, similar results were observed for HLA-haploidentical transplants, thereby expanding the numbers of patients who could be offered this treatment.[18]

Future Directions

Experimental treatments to induce higher levels of HbF (thereby reducing HbS polymer), to decrease the adherence of sickle erythrocytes to endothelium, and to modulate the oxygen affinity of hemoglobin in the sickle cell are undergoing clinical trials. Where malaria is endemic, antimalaria prophylaxis reduces episodes of malaria and increases mean hemoglobin levels. Gene therapy, tried in a few patients with β-thalassemia syndromes with some early success, has not yet been reported in sickle cell disease, but trials should soon begin.

PROGNOSIS

Average life expectancy for patients with sickle cell anemia in the United States was reported to be between 50 and 60 years, but these figures do not reflect better supportive care or widespread use of hydroxyurea; patients with HbSC disease typically live 60 to 70 years. Patients with HbS-β^0-thalassemia are likely to have a life expectancy similar to that of those with sickle cell anemia, and the lifespan for patients with the HbS-β^+-thalassemia may resemble that of patients with HbSC disease. Death is often caused by pulmonary disease and infection, and another 20% of deaths are related to organ failure. However, death in adults often is unexpected, happening in the midst of an acute event such as an acute painful episode and occurring within the first 24 hours of hospitalization. In areas of the developing world without access to modern medical care, death in childhood is still common.

OTHER HEMOGLOBINOPATHIES

Hemoglobinopathies other than those associated with HbS, HbE (Chapter 162), and HbC rarely cause clinically recognizable disorders. HbC (*HBB* Glu6Lys) and HbE (*HBB* Glu26Lys) are common β-globin variants. HbC is present in about 2% of African Americans, and HbE is seen in Southeast Asia, where, in some areas, the gene frequency may reach 50%. HbE is a β-hemoglobin variant that is produced at a slightly reduced rate and hence has the phenotype of a mild form of β-thalassemia.[19] Heterozygotes with HbC or HbE are asymptomatic, although the blood of HbC trait carriers contains target cells, and HbE carriers may have mild anemia and microcytosis. Even homozygotes for HbC and HbE have virtually no clinical disease, only mild hematologic abnormalities such as microcytosis, target cells, and mild anemia. Screening testing can be done by high performance liquid chromatography.

Compound heterozygotes for HbE and β-thalassemia usually have the phenotype of transfusion-dependent β-thalassemia, although genotype-phenotype correlations are difficult to make because of the likelihood that other genes affect the expression of disease. Because of immigration from Southeast Asia, HbSE disease is increasingly common in North America and Europe. It resembles HbS-β^+-thalassemia.

Rare hemoglobinopathies may change the affinity of hemoglobin for oxygen, render it susceptible to oxidation, or cause molecular instability. Amino acid substitutions involving heme-binding residues may lead to irreversible iron oxidation, methemoglobinemia, and cyanosis (Chapter 158). Patients with these conditions have congenital pseudocyanosis but are usually asymptomatic and need no treatment.

Substitutions at contacts between globin subunits may alter the affinity of hemoglobin for oxygen. When hemoglobin-oxygen affinity is increased, less oxygen is available in tissues, erythropoietin production is enhanced, and erythrocytosis results (Chapter 166). No treatment is usually required because the erythrocytosis is mild. Hemoglobin-oxygen affinity may also be reduced by some mutations, resulting in anemia or cyanosis. Hemoglobin instability, produced by several molecular mechanisms (including introduction of proline residues into the α-helix, substitutions near the heme ring, and deletion or addition of amino acids), often causes loss of heme from the molecule and hemolytic anemia. Hemoglobin Köln is the most common example of this class of hemoglobinopathy, but more than 200 unstable variants have been described. Oxidant drugs sometimes provoke increased hemolysis. Splenectomy is sometimes an effective treatment when the anemia is severe.

Grade A References

A1. Charache S, Terrin ML, Moore RD, et al. Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. *N Engl J Med*. 1995;332:1317-1322.

A2. Steinberg MH, Barton F, Castro O, et al. Effect of hydroxyurea on mortality and morbidity in adult sickle cell anemia: risks and benefits up to 9 years of treatment. *JAMA*. 2003;289:1645-1651.

A3. Wang WC, Ware RE, Miller ST, et al. Hydroxycarbamide in very young children with sickle-cell anaemia: a multicentre, randomised, controlled trial (BABY HUG). *Lancet*. 2011;377:1663-1672.

A4. Goldman RD, Mounstephen W, Kirby-Allen M, et al. Intravenous magnesium sulfate for vaso-occlusive episodes in sickle cell disease. *Pediatrics*. 2013;132:e1634-e1641.

A5. Machado RF, Barst RJ, Yovetich NA, et al. Hospitalization for pain in patients with sickle cell disease treated with sildenafil for elevated TRV and low exercise capacity. *Blood*. 2011;118:855-864.

A6. Sasongko TH, Nagalla S, Ballas SK. Angiotensin-converting enzyme (ACE) inhibitors for proteinuria and microalbuminuria in people with sickle cell disease. *Cochrane Database Syst Rev*. 2013;3:CD009191.

A7. Howard J, Malfroy M, Llewelyn C, et al. The Transfusion Alternatives Preoperatively in Sickle Cell Disease (TAPS) study: a randomised, controlled, multicentre clinical trial. *Lancet*. 2013;381:930-938.

A8. DeBaun MR, Gordon M, McKinstry RC, et al. Controlled trial of transfusions for silent cerebral infarcts in sickle cell anemia. *N Engl J Med*. 2014;371:699-710.

A9. Meerpohl JJ, Schell LK, Rucker G, et al. Deferasirox for managing transfusional iron overload in people with sickle cell disease. *Cochrane Database Syst Rev*. 2014;5:CD007477.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Goldsmith JC, Bonham VL, Joiner CH, et al. Framing the research agenda for sickle cell trait: building on the current understanding of clinical events and their potential implications. *Am J Hematol*. 2012;87:340-346.
2. Steinberg MH, Sebastiani P. Genetic modifiers of sickle cell disease. *Am J Hematol*. 2012;87:824-826.
3. Akinsheye I, Alsultan A, Solovieff N, et al. Fetal hemoglobin in sickle cell anemia. *Blood*. 2011;118:19-27.
4. Steinberg MH, Chul DH, Dover GJ, et al. Fetal hemoglobin in sickle cell anemia: a glass half full? *Blood*. 2014;123:481-485.
5. Gallagher PG. Disorders of red cell volume regulation. *Curr Opin Hematol*. 2013;20:201-207.
6. Connes P, Verlhac S, Bernaudin F. Advances in understanding the pathogenesis of cerebrovascular vasculopathy in sickle cell anaemia. *Br J Haematol*. 2013;161:484-498.
7. Vichinsky EP, Neumayr LD, Gold JI, et al. Neuropsychological dysfunction and neuroimaging abnormalities in neurologically intact adults with sickle cell anemia. *JAMA*. 2010;303:1823-1831.
8. Desai PC, Ataga KI. The acute chest syndrome of sickle cell disease. *Expert Opin Pharmacother*. 2013;14:991-999.
9. Lim MY, Ataga KI, Key NS. Hemostatic abnormalities in sickle cell disease. *Curr Opin Hematol*. 2013;20:472-477.
10. Gardner K, Suddle A, Kane P, et al. How we treat sickle hepatopathy and liver transplantation in adults. *Blood*. 2014;123:2302-2307.
11. Mehari A, Gladwin MT, Tian X, et al. Mortality in adults with sickle cell disease and pulmonary hypertension. *JAMA*. 2012;307:1254-1256.
12. Yawn BP, Buchanan GR, Afenyi-Annan AN, et al. Management of sickle cell disease: summary of the 2014 evidence-based report by expert panel members. *JAMA*. 2014;312:1033-1048.
13. Olujohungbe A, Burnett AL. How I manage priapism due to sickle cell disease. *Br J Haematol*. 2013;160:754-765.
14. Klings ES, Machado RF, Barst RJ, et al. An official American Thoracic Society clinical practice guideline: diagnosis and treatment of pulmonary hypertension of sickle cell disease. *Am J Respir Crit Care Med*. 2014;189:727-740.
15. Chou ST, Liem RI, Thompson AA. Challenges of alloimmunization in patients with haemoglobinopathies. *Br J Haematol*. 2012;159:394-404.
16. Hoffbrand AV, Taher A, Cappellini MD. How I treat transfusional iron overload. *Blood*. 2012;120:3657-3669.
17. Hsieh MM, Fitzhugh CD, Weitzel RP, et al. Nonmyeloablative HLA-matched sibling. *JAMA*. 2014;312:48-56.
18. Bolanos-Meade J, Fuchs EJ, Luznik L, et al. HLA-haploidentical bone marrow transplantation with posttransplant cyclophosphamide expands the donor pool for patients with sickle cell disease. *Blood*. 2012;120:4285-4291.
19. Fucharoen S, Weatherall DJ. The hemoglobin E thalassemias. *Cold Spring Harb Perspect Med*. 2012;2:a011734.

REVIEW QUESTIONS

1. A 40-year-old man with sickle cell anemia (HbSS) requires open lung resection of a newly diagnosed and localized non–small cell lung cancer. He has had numerous hospitalizations in the past for sickle crises, including acute chest syndrome, and was taking hydroxyurea until about 3 years ago, when he discontinued it and was lost to follow-up. His complete blood count now is as follows: white blood cells, 11,000; hemoglobin, 9.5; hematocrit, 25.5; platelets, 420,000; reticulocytes, 4.5%. These values are similar to those he has had in the past. What preoperative management related to his sickle cell anemia should be recommended?

A. Transfusion of red blood cells to a hematocrit of about 30%
B. Exchange transfusions and initiation of iron chelation therapy
C. Hydroxyurea with a loading dose and then his previous maintenance dose
D. Erythropoietin therapy
E. None of the above but close hemodynamic and hematologic monitoring

Answer: A Cautious preoperative blood transfusion can improve surgical outcomes. Simple preoperative red cell transfusions to a hematocrit of about 30% are as effective as exchange transfusions to prevent postoperative complications, and they cause fewer transfusion-related complications. See the section Surgery and Anesthesia and reference A7 to the Transfusion Alternatives Preoperatively in Sickle Cell Disease ((TAPS) study, a randomized, controlled, multicenter clinical trial. Chronic iron chelation therapy is effective in the treatment of transfusional iron overload, but it is not urgent to begin it preoperatively, especially because he has not even had baseline magnetic resonance imaging evaluation for his level of iron overload (see the section Red Blood Cell Transfusion and Iron Chelation Therapy). Judicious use of erythropoietin therapy may be considered in sickle cell disease patients with renal failure, but even in those individuals, its effects on increasing the hematocrit are gradual in onset and unpredictable.

2. A 25-year-old woman with sickle cell anemia (HbSS) is brought to the emergency department after an episode of syncope. She has been observed regularly in the comprehensive sickle cell center of the hospital and has been taking hydroxyurea 500 mg daily for the past 4 years, which has been associated with a marked decrease in painful sickle crises. Blood counts have been regularly monitored during the hydroxyurea therapy, and the most recent complete blood count 1 week ago showed the following: white blood cells, 7500; hemoglobin, 9.9; hematocrit, 28.5; mean corpuscular volume, 106; platelets, 320,000; reticulocytes, 6.5%. On examination in the emergency department, she is sluggishly responsive, is hypotensive, and has a thread pulse with a rate of 140. There is mild, diffuse abdominal tenderness but no organomegaly or mass palpable. Stat complete blood count now shows the following: white blood cells, 8500; hemoglobin, 3.8; hematocrit, 17.2; reticulocytes, 0.2%; platelets, 410,000. What is the most likely diagnosis?

A. Acute vaso-occlusive crisis
B. Sequestration crisis
C. Marrow suppression by hydroxyurea
D. Aplastic crisis
E. Hyperhemolytic crisis

Answer D Acute anemia in patients with HbSS can have several causes, as described in the section Acute Anemia. A simple vaso-occlusive crisis should not be accompanied by an acute drop in hemoglobin and hematocrit. Sequestration crisis occurs in young children (peak age, 2 years) who still have a functioning spleen; they present with massive and painful enlargement of the spleen, not seen in this case, due to trapping of sickle cells in splenic sinusoids. Myelosuppression by hydroxyurea is always a concern, but her blood counts have been closely monitored and were unchanged just a week ago; and even if she happened to take an inadvertently high dose of the drug in the interval, her severe anemia should have been accompanied by major decreases in the white blood cell and platelet counts as well. Hemolytic crisis is essentially ruled out by the very low reticulocyte count. Aplastic crisis in HbSS patients is a rare but life-threatening emergency that is characterized by pure red cell aplasia typically as a result of parvovirus B19 infection. Parvovirus B19 has characteristic tropism for erythroid precursors and therefore causes severe erythroblastopenia in the bone marrow. Coupled with the ongoing brisk hemolysis of sickle cell disease, an abrupt shutdown of red cell production by the marrow leads to a precipitous fall in hemoglobin and hematocrit. In addition to aggressive but cautious red cell transfusions (to avoid fluid overload), intravenous immune globulin has been found to be effective by providing a large bolus of neutralizing antibodies to accelerate clearance of parvovirus B19.

164

MEGALOBLASTIC ANEMIAS

AŚOK C. ANTONY

DEFINITION

Megaloblastic anemias, a group of disorders characterized by a distinct morphologic pattern in hematopoietic cells, are commonly due to a deficiency of vitamin B_{12} (cobalamin) or folate. These anemias are globally prevalent and carry a significant burden of morbidity. Folate and cobalamin are both required to sustain one-carbon metabolism, which involves the transfer of one-carbon groups such as methyl-, formyl-, methylene-, methenyl-, and

formimino- in enzyme reactions essential for pyrimidine and purine biosynthesis, including the synthesis of three of the four nucleotides of DNA. Thus, a deficiency in cobalamin or folate results in the common biochemical feature of a defect in DNA synthesis along with lesser alterations in RNA and protein synthesis, leading to a state of unbalanced cell growth and impaired cell division. The majority of megaloblastic cells have DNA values between 2 and 4 N because of delayed cell division. This is morphologically expressed as larger-than-normal "immature" nuclei with finely particulate chromatin, whereas the relatively unimpaired RNA and protein synthesis results in large cells with greater "mature" cytoplasm and cell volume. The microscopic appearance of this nuclear-cytoplasmic asynchrony (or dissociation) is morphologically described as *megaloblastic*. Megaloblastic hematopoiesis commonly is manifested with anemia, the most easily recognized clinical manifestation of a global defect in DNA synthesis affecting all rapidly proliferating cells. Precise identification of the deficient vitamin and the cause of the deficiency (Table 164-1) dictates the dose and duration of replacement therapy.

EPIDEMIOLOGY

Cobalamin

Nutrition

Cobalamin is a red, water-soluble vitamin with a complex structure that generally resembles the heme molecule but with cobalt replacing iron in the center of the pyrrole ring. The recommended daily allowance of cobalamin is 2.4 µg for men and nonpregnant women, 2.6 µg for pregnant women, 2.8 µg for lactating women, and between 1.5 and 2 µg for children 9 to 18 years old. Cobalamin is produced in nature only by microorganisms, and humans receive cobalamin solely from the diet. Meat from parenchymal organs is richest in cobalamin (>10 µg/100 g wet weight); fish and animal muscle, milk products, and egg yolks have 1 to 10 µg/100 g wet weight. An average nonvegetarian Western diet with abundant meat, milk, and other dairy products and eggs contains 5 to 7 µg/day of cobalamin, which is adequate to sustain normal cobalamin equilibrium. Herbivores can receive minuscule amounts of cobalamin from nitrogen-fixing soil bacteria (genus *Rhizobium*) present in the roots and nodules of legumes; from fresh produce contaminated by tiny insects or cobalamin-rich manure; from dried seaweed varieties (nori, chlorella, and spirulina); and from tempeh (fermented soybean cake). However, these are not reliable sources of cobalamin. For vegetarians, the consumption of eggs, milk, and dairy products generally provides less than 0.5 µg/day of cobalamin and cannot sustain cobalamin balance.[1] Near-vegetarians who infrequently consume animal-source foods (often because of poverty) also have a cobalamin status that is only marginally better than that of lacto-ovovegetarians and are also at risk for cobalamin deficiency.

Cobalamin is stored exceptionally well in tissues. Of the total body content of 2 to 5 mg in adults, half is in the liver. With a daily loss of 1 µg, dietary

TABLE 164-1 ETIOPATHOPHYSIOLOGIC CLASSIFICATION OF COBALAMIN AND FOLATE DEFICIENCIES

COBALAMIN DEFICIENCY
Nutritional cobalamin deficiency (insufficient cobalamin intake): vegetarians, poverty-imposed near-vegetarians, breast-fed infants of mothers with pernicious anemia
Abnormal intragastric events (inadequate proteolysis of food cobalamin): atrophic gastritis, hypochlorhydria, proton pump inhibitors, H_2-blockers
Loss/atrophy of gastric oxyntic mucosa (deficient intrinsic factor molecules): total or partial gastrectomy, adult and juvenile pernicious anemia, caustic destruction (lye)
Abnormal events in the small bowel lumen
Inadequate pancreatic protease (R-factor-bound cobalamin not degraded, cobalamin not transferred to intrinsic factor)
Insufficient pancreatic protease: pancreatic insufficiency
Inactivation of pancreatic protease: Zollinger-Ellison syndrome
Usurping of luminal cobalamin (inadequate binding of cobalamin to intrinsic factor)
By bacteria: stasis syndromes (blind loops, pouches of diverticulosis, strictures, fistulas, anastomosis), impaired bowel motility (scleroderma), hypogammaglobulinemia
By *Diphyllobothrium latum* (fish tapeworm)
Disorders of ileal mucosa/intrinsic factor–cobalamin receptors (intrinsic factor–cobalamin not bound to intrinsic factor–cobalamin receptors [also known as cubam receptors])
Diminished or absent cubam receptors: ileal bypass, resection, fistula
Abnormal mucosal architecture or function: tropical/nontropical sprue, Crohn disease, tuberculous ileitis, amyloidosis
Cubam receptor defects: Imerslund-Gräsbeck syndrome, hereditary megaloblastic anemia
Drug effects: metformin, cholestyramine, colchicine, neomycin
Disorders of plasma cobalamin transport (TCII-cobalamin not delivered to TCII receptors): congenital TCII deficiency, defective binding of TCII-cobalamin to TCII receptors (rare)
Metabolic disorders (cobalamin not used by cells)
Inborn enzyme errors (rare)
Acquired disorders (cobalamin functionally inactivated by irreversible oxidation): nitrous oxide inhalation
FOLATE DEFICIENCY
Nutritional causes
Decreased dietary intake: poverty and famine, institutionalization (psychiatric facilities, nursing homes), chronic debilitating disease, prolonged feeding of infants with goat's milk, special slimming diets or fad foods (folate-rich foods not consumed), cultural or ethnic cooking techniques (food folate destroyed)
Decreased dietary intake and increased requirements
Physiologic: pregnancy and lactation, prematurity, hyperemesis gravidarum, infancy
Pathologic
Intrinsic hematologic diseases involving hemolysis with compensatory erythropoiesis, abnormal hematopoiesis, or bone marrow infiltration by malignant disease
Dermatologic disease: psoriasis
Folate malabsorption
With normal intestinal mucosa
Drugs: pyrimethamine, proton pump inhibitors (by inhibition of proton-coupled folate transporters); anticonvulsants (reduced absorption and induction of microsomal liver enzymes)
Hereditary folate malabsorption (mutations in proton-coupled folate transporters) (rare)
With mucosal abnormalities: tropical and nontropical sprue, regional enteritis
Defective CSF folate transport: cerebral folate deficiency (mutation or autoantibodies to folate receptors) (rare)
Inadequate cellular utilization
Folate antagonists (methotrexate)
Hereditary enzyme deficiencies involving folate
Drugs (multiple effects on folate metabolism): alcohol, sulfasalazine, triamterene, pyrimethamine, trimethoprim-sulfamethoxazole, phenytoin, barbiturates
MISCELLANEOUS MEGALOBLASTIC ANEMIAS NOT CAUSED BY COBALAMIN OR FOLATE DEFICIENCY
Congenital disorders of DNA synthesis
Orotic aciduria
Lesch-Nyhan syndrome
Congenital dyserythropoietic anemia
Acquired disorders of DNA synthesis
Deficiency: thiamine-responsive megaloblastic anemia (thiamine transporter 1 mutation)
Erythroleukemia; refractory sideroblastic anemias (pyridoxine responsive?)

CSF = cerebrospinal fluid; TCII = transcobalamin II.

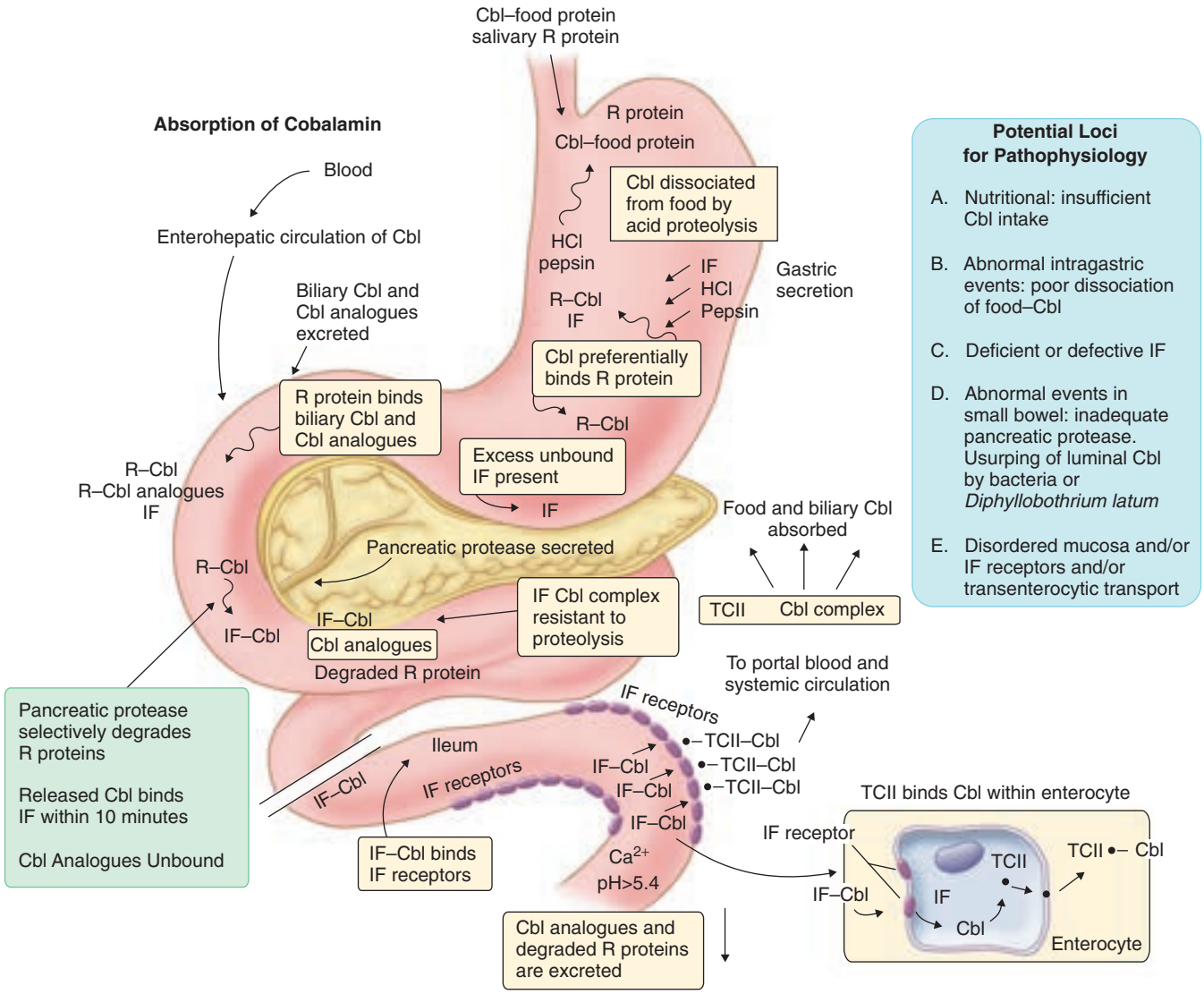

FIGURE 164-1. Components and mechanism of cobalamin absorption, with an indication of the locus for malabsorption. Cbl = cobalamin; IF = intrinsic factor; TCII = transcobalamin II. (From Antony AC. Megaloblastic anemias. In: Hoffman R, Benz EJ Jr, Silberstein LE, et al, eds. Hematology: Basic Principles and Practice. 6th ed. Philadelphia: Elsevier Saunders; 2013:473-504.)

cobalamin deficiency can take 5 to 10 years to become apparent. However, it takes about 3 to 4 years to deplete cobalamin stores if dietary cobalamin is abruptly malabsorbed (e.g., ileal resection), thereby interfering with an efficient enterohepatic circulation, which accounts for the turnover of 5 to 10 µg of cobalamin per day and reabsorption of 75% of cobalamin secreted into bile. Although cobalamin resists high-temperature cooking, it is unstable to light and can be converted to inactive analogues.

Folates

Nutrition

Folates are synthesized by microorganisms and plants. Rich food sources include green leafy vegetables (spinach, lettuce, broccoli), beans, fruit (bananas, melons, lemons), yeast, mushrooms, and animal protein (muscle, liver, kidney). The recommended daily allowance of folate is 400 µg for adult men and nonpregnant women, 600 µg for pregnant women, 500 µg for lactating women, and 300 to 400 µg for children 9 to 18 years of age. A balanced Western diet can prevent folate deficiency, but the net dietary intake of folate in many developing countries is more often grossly insufficient to sustain folate balance. Folates are susceptible to breakdown during prolonged cooking (boiling for more than 15 minutes), which can destroy 50 to 95% of folate.

PATHOBIOLOGY

Cobalamin

Normal Physiology

There are specialized protein chaperones that sequentially bind, sequester, and thereby protect cobalamin through its long odyssey—from the moment it is dissociated from food in the stomach to its final destination within cells as a cofactor for crucial enzyme reactions.

Absorption and Transport

Cobalamin in food is usually in coenzyme form (as deoxyadenosylcobalamin and methylcobalamin) and bound to proteins (Fig. 164-1). In the stomach, peptic digestion at low pH is a prerequisite for the release of cobalamin from food protein. Once it is released, cobalamin preferentially binds a high-affinity cobalamin-binding protein called R protein, which is secreted in salivary and gastric juice. These cobalamin–R protein complexes, along with unbound intrinsic factor, which is secreted by gastric parietal cells, pass into the duodenum, where pancreatic proteases degrade R proteins. This allows transfer of cobalamin to intrinsic factor.

The stable intrinsic factor–cobalamin complexes then pass through the jejunum to the ileum, where they specifically bind to intrinsic factor–cobalamin receptors (also called cubam receptors—a complex of two proteins, cubilin and amnionless) on the microvilli of ileal mucosal cells. Within enterocytes, cobalamin is transferred to transcobalamin II; this complex is then released into the circulation, from which it efficiently binds to high-affinity transcobalamin II receptors on cell surfaces. When cobalamin status is compromised, this is the fraction that is primarily reduced. There is, however, another protein, transcobalamin I, that binds the bulk (approximately 75%) of serum cobalamin but does not deliver it to cells, so it functions much like a storage protein for the cobalamin in blood. Because the cobalamin from this compartment of transcobalamin I–bound cobalamin—which is relatively invariant in serum—is also measured in serum cobalamin

assays, this accounts for the relative insensitivity of this test, which aims to discover if there is a reduction in the functionally relevant transcobalamin II–bound cobalamin.[1,2] A third minor protein, transcobalamin III, binds a wide spectrum of cobalamin analogues that are rapidly cleared by the liver into bile for efficient fecal excretion.

Cellular Processing

More than 95% of intracellular cobalamin is bound to two intracellular enzymes: methylmalonyl coenzyme A (CoA) mutase and methionine synthase. In mitochondria, deoxyadenosylcobalamin is a coenzyme for methylmalonyl-CoA mutase, which converts methylmalonyl-CoA to succinyl-CoA so that it can be easily metabolized. In the cytoplasm, methylcobalamin is a coenzyme for methionine synthase, which catalyzes the transfer of methyl groups from methylcobalamin to homocysteine to form methionine. The methyl group of 5-methyltetrahydrofolate (methyl-THF) is donated to regenerate methylcobalamin, thereby forming the THF that is essential to sustain one-carbon metabolism. The methionine so formed can be adenylated to *S*-adenosylmethionine, which donates its methyl group in a critical series of biologic methylation reactions involving more than 80 proteins, phospholipids, neurotransmitters, RNA, and DNA. The close functional interrelationship between cobalamin and folate within cells (involving the common enzyme methionine synthase, the single enzymatic reaction of which is a metabolic step for which both cobalamin and folate are essential) explains why cobalamin deficiency leads to a functional folate deficiency and is also the basis for similar clinical (hematologic) manifestations involving perturbed DNA.

Pathogenesis of Cobalamin Deficiency

Nutritional Cobalamin Deficiency

Severe cobalamin deficiency in the West is likely to be pernicious anemia.[2] However, vegetarianism and poverty-imposed near-vegetarianism are more common causes worldwide in all age groups. Up to three quarters of the population in resource-limited countries (particularly women and children), who subsist on a monotonous diet low in animal-source foods, have subtle evidence of cobalamin deficiency; many also have folate and iron deficiency.[3] Low maternal cobalamin status compromises the amount of cobalamin delivered to the fetus; moreover, low cobalamin content of breast milk predisposes one third of their infants to cobalamin deficiency, thereby establishing a vicious intergenerational circle of cobalamin (and multiple nutrient) deficiency. As large swaths of the populations are exposed to war or famine, those teetering on the brink of cobalamin (and folate) deficiency will eventually come to light clinically.

Inadequate Dissociation of Cobalamin from Food Protein

In affluent countries, failure to fully release food cobalamin because of chronic gastric atrophy and achlorhydria is common among one third to one half of elderly individuals with low cobalamin status. This is estimated to be 10-fold more common than pernicious anemia.

Absent Secretion of Acid and Intrinsic Factor

Total gastrectomy invariably leads to cobalamin deficiency in 2 to 10 years, thus warranting prophylactic cobalamin (and iron) replacement. After partial gastrectomy or bariatric gastric bypass surgery (Chapter 220) and in those receiving long-term H_2-blockers or proton pump inhibitors, multifactorial cobalamin deficiency may result from decreased secretion of intrinsic factor, hypochlorhydria, or intestinal bacterial overgrowth of cobalamin-consuming organisms.

In pernicious anemia, autoimmune gastritis involving destruction of the gastric parietal cell mass leads to atrophy of the fundus and body of the stomach, absence of intrinsic factor and hydrochloric acid, and eventually severe cobalamin malabsorption and deficiency. Pernicious anemia is found in persons of all ages, races, and ethnic origins; however, the precise global incidence is not known because population-based studies, which rely on finding diagnostic serum anti–intrinsic factor antibodies (found in serum of 60% and gastric juice of 75%), will miss a substantial number of patients. Nevertheless, with this approach, nearly 2% of free-living individuals older than 60 years in southern California had undiagnosed pernicious anemia, with minimal clinical manifestations of cobalamin deficiency; significantly, in this cohort, 4% of white and African American women had pernicious anemia. About 30% of patients have a positive family history, and there is an association with other autoimmune diseases (e.g., polyglandular autoimmune syndrome, Graves disease, Hashimoto thyroiditis, vitiligo, Addison disease, idiopathic hypoparathyroidism, myasthenia gravis, and type 1 diabetes).

Abnormal Events Precluding Absorption of Cobalamin

Although pancreatic insufficiency (Chapter 140) can preclude transfer of R protein–bound cobalamin to intrinsic factor, with the early use of pancreatic protease replacement, cobalamin deficiency is uncommon. Endogenous pancreatic protease can, however, be inactivated by massive gastric hypersecretion arising from a gastrinoma in Zollinger-Ellison syndrome (Chapter 195). Further, if the pH of the luminal contents in the ileum is less than 5.4, the binding of the intrinsic factor–cobalamin complex to cubam receptors will be precluded.

Bacterial overgrowth in the small bowel (arising from stasis, impaired motility, and hypogammaglobulinemia; Chapter 142) favors colonization by bacteria, which can usurp free cobalamin before it can bind to intrinsic factor. This can be reversed by a short course of antibiotics. Individuals heavily infested with the fish tapeworm *Diphyllobothrium latum* (acquired by consuming raw or partially cooked freshwater fish) can become cobalamin deficient when these long (10 m) adult worms in the jejunum avidly usurp cobalamin. After expulsion of worms (with an oral dose of praziquantel 10 to 20 mg/kg), cobalamin replenishment is curative.

Disorders of the Intrinsic Factor Receptors or Mucosa

Because the distal ileum has the greatest density of cubam receptors, the removal, bypass, or dysfunction of only 1 to 2 feet of terminal ileum will result in cobalamin malabsorption. Among drugs (see Table 164-1), metformin, when it is used in full doses for type 2 diabetes, is notorious for inducing cobalamin malabsorption and low cobalamin levels after 5 years or more.[4] Although it is preventable by calcium (1.2 g/day), many with metformin-induced low cobalamin levels will progress to cobalamin deficiency, warranting cobalamin replacement and prophylaxis.

Acquired Cobalamin Deficiency

Nitrous oxide (N_2O) irreversibly inactivates cobalamin and results in a state of functional intracellular cobalamin deficiency, which can be bypassed by the administration of 5-formyl-THF (leucovorin). N_2O exposure can induce megaloblastosis in those with marginal or low cobalamin stores, but chronic intermittent (surreptitious, accidental, or occupational) exposure more frequently leads to neuromyelopathic manifestations. Capsules used for making whipped cream are a cheap and easy source of N_2O, allowing for abuse in the community.

Folates

Normal Physiology

Specialized mechanisms also exist to ensure the digestion, absorption, and uptake of folates into cells—and across both the placenta to the fetus and the choroid plexus into the nervous system—to support intracellular enzymes that are critical for synthesis of DNA.

Absorption and Transport

In general, only half the folate in food, which is mainly in polyglutamylated form, is nutritionally available (bioavailable), whereas 85% of folic acid added to food or ingested as a supplement is bioavailable. The small intestine can absorb folic acid unchanged, but food folate polyglutamates must be hydrolyzed to monoglutamate by folate polyglutamate hydrolase at the brush border before transport into enterocytes, so food preparation (dicing, puréeing) can facilitate absorption. A jejunal luminal surface proton-coupled folate transporter, which has a low pH optimum, facilitates the efficient transport of folate into enterocytes, where it is reduced to THF and methylated before release by a cellular exporter protein into plasma as methyl-THF.[5] The serum folate level is maintained by dietary folate intake and an efficient enterohepatic circulation. From plasma, there is a rapid uptake of folates into tissues by cell membrane–associated folate receptors, which bind physiologically relevant methyl-THF, folic acid, and some newer antifolates with high affinity at concentrations found in serum. After folate receptor–mediated endocytosis, a proton-coupled folate transporter then helps export folate from acidified endosomes into the cytoplasm of cells. The sequentially coordinated function of folate receptors and proton-coupled folate transporters of the choroid plexus is also required to maintain the ratio of cerebrospinal fluid (CSF) folate to blood folate of up to 3 : 1. After glomerular filtration, folate receptors on the brush border membranes of proximal renal tubular cells bind luminal folate and transport it back into blood. Cells have an exquisite

molecular mechanism for sensing and responding to folate deficiency by upregulating folate receptors to restore folate homeostasis. Reduced-folate carriers are also folate transporters that primarily mediate the uptake of pharmacologic folates (methotrexate and folinic acid) into cells. Passive diffusion also operates to transport folate across biologic membranes at supraphysiologic folate concentrations.

Intracellular Metabolism and Cobalamin-Folate Interactions

After cellular uptake, methyl-THF must first be converted to THF by methionine synthase. Only then can the THF be polyglutamylated by folate polyglutamate synthase, which allows it to be retained intracellularly to play a central role in one-carbon metabolism. THF can be converted to 10-formyl-THF (for de novo biosynthesis of purines) and to methylene-THF.

The central role of methylene-THF is that it can be used either in the thymidylate cycle via thymidylate synthase for the synthesis of thymidine and DNA or in the methylation cycle via methionine synthase (but only after its conversion to methyl-THF by methylene-THF reductase). Inactivation of methionine synthase during cobalamin deficiency results in accumulation of the substrate methyl-THF, which cannot be polyglutamylated and thus leaks out of the cell, resulting in an intracellular THF deficiency and compromised one-carbon metabolism (E-Table 164-1).

Pathogenesis of Folate Deficiency

Folate deficiency can arise from decreased supply (reduced intake, absorption, transport, or utilization) or increased requirements (from metabolic consumption, destruction, or excretion). One individual may have multiple causes of folate deficiency, but specific tests to define each mechanism are not available clinically. Nutritional folate insufficiency is the most common cause of folate deficiency worldwide in all age groups, with women and children in resource-limited (developing) countries at highest risk. Indeed, more than 90% of pregnant women in resource-limited countries consume less than the estimated average requirement of folate, cobalamin, and iron. Although fortification of food with folate has dramatically reduced the prevalence of folate deficiency in the United States, it can still be found among some chronic alcoholics and the elderly, infirm, or socially isolated individuals who consume imbalanced diets.

Nutritional Causes (Decreased Intake or Increased Requirements)

With an abrupt reduction in folate consumption, body stores of folate are adequate for approximately 4 months. However, these stores are depleted faster in individuals in chronically negative folate balance who often have multiple nutritional deficiencies (and diseases) that tip them into frank folate deficiency. A seasonal reduction in folate-rich foods, poverty, cultural or ethnic diets that are intrinsically low in folates, cooking techniques that destroy folate, and anorexia that accompanies chronic illnesses are among the many reasons for folate deficiency.

Patients with intrinsic hematologic diseases involving increased cell proliferation or with increased compensatory erythropoiesis in response to chronic peripheral red blood cell destruction have increased requirements for folate. Indeed, folate deficiency in the face of chronic hemolysis can lead to an acute reticulocytopenic (aplastic) crisis, an unexpected increase in transfusion requirements, or a fall in platelets. Exfoliative skin diseases (Chapter 436) also cause folate deficiency when there is an increased demand from excess loss of skin cells.

Pregnancy and Infancy

Poor folate intake during pregnancy is a common cause of megaloblastic anemia in developing countries because pregnancy and lactation require additional folate for growth of the fetus and maternal tissues. Physiologic transplacental folate transport by folate receptors relies on the continued intake of adequate dietary folate by the mother; if it is compromised by poor nutrition or increased needs (short intervals between pregnancies, twin pregnancies), poor pregnancy outcomes will result. This can be manifested by premature, low-birthweight infants and other midline developmental abnormalities in the fetus ranging from neural tube defects (such as anencephaly, encephalocele, meningocele, and spina bifida) to neurocristopathies (such as ventricular septal defects, cleft lip, and cleft palate). Periconceptional folate supplementation studies suggest that mothers who fail to consume sufficient folate during pregnancy have children who exhibit subtle changes that are manifested in early childhood; these include behavioral abnormalities (hyperactive/inattentive with peer problems, emotionally reactive, aggressive, anxious/depressed symptoms, and somatic complaints or withdrawn[6] and autistic disorder[7]) as well as cognitive dysfunction,[8] poor academic performance, and delayed language acquisition.

Cerebral folate deficiency can be caused either by a congenital mutation in folate receptors or by autoantibodies to folate receptors, which perturbs folate transport into the CSF, leading to severe developmental regression in early childhood associated with movement disturbances, epilepsy, and leukodystrophy. Blocking autoantibodies to folate receptors, which develop against consumed bovine milk folate-binding proteins that share epitopes with human folate receptors, can also be found in two autism spectrum disorders: Rett syndrome and infantile low-functioning autism with neurologic abnormalities. When it is diagnosed early, cerebral folate deficiency responds to high doses of folinic acid, which normalizes CSF folates and induces a partial to complete recovery in 12 months. When it is induced by autoantibodies, a bovine milk–free diet will also help decrease the autoantibody titer.

Hereditary folate malabsorption, which is due to a mutation in the proton-coupled folate transporter, results in compromised intestinal folate absorption and folate transfer into the CSF. Patients present with folate deficiency anemia, hypoimmunoglobulinemia with recurrent infections, chronic diarrhea, neurologic abnormalities (seizures or mental retardation), and low to undetectable CSF folate levels. High parenteral doses of folinic acid can ensure passive diffusion into the CSF and lead to significant clinical improvement in these children.

Tropical and Nontropical (Celiac) Sprue

With the development of intestinal mucosal abnormalities, patients are at increased risk for folate malabsorption. In tropical sprue (Chapter 140), a dramatic response to a 4- to 6-month course of oral folic acid (5 mg/day) plus tetracycline (250 mg four times a day) can effect a cure in 60% or more of patients. When it is prolonged (>3 years), malabsorption of cobalamin can develop together with iron deficiency (Chapter 159), pellagra, and beriberi (Chapter 218).

Drugs

Excess alcohol consumption at the expense of a balanced diet is a common cause of folate deficiency in the United States. Inhibition of dihydrofolate reductase by trimethoprim and pyrimethamine or methotrexate can be acutely reversed by folinic acid. Pyrimethamine and proton pump inhibitors inhibit the proton-coupled folate transporter, whereas anticonvulsants can reduce folate absorption and induce microsomal liver enzymes. Antineoplastics and antiretroviral antinucleosides can also induce megaloblastosis by perturbing DNA synthesis.

CLINICAL MANIFESTATIONS

The finding of macrocytosis (increased mean corpuscular volume [MCV]) on a routine complete blood count may be the first clinical manifestation. In other patients, the findings may be dominated by the condition causing the deficiency of cobalamin or folate, such as malabsorption, alcoholism, or malnutrition (see Table 164-1).

The clinical manifestations of folate deficiency may include hematologic (pancytopenia with megaloblastic bone marrow), cardiopulmonary (secondary to anemia), gastrointestinal (megaloblastosis with or without malabsorption), dermatologic (hyperpigmentation of the skin, premature graying), infertility (sterility), and psychiatric (primarily a flat affect) symptoms. If such patients have additional neurologic findings, either associated cobalamin deficiency or other diseases that predispose to folate deficiency must be considered, such as alcoholism with thiamine deficiency, which may result in peripheral neuropathy (dry beriberi) with Wernicke-Korsakoff syndrome, with or without heart failure from cardiovascular disease (wet beriberi) (Chapters 218 and 416). Because megaloblastosis due to either folate or cobalamin deficiency results in functional folate deficiency, the hematologic manifestations of both deficiencies, including pancytopenia with megaloblastic bone marrow, are indistinguishable. However, only cobalamin deficiency results in a patchy but widespread demyelinating process, which is expressed clinically as cerebral abnormalities and subacute combined degeneration of the spinal cord (Chapter 416). Either hematologic or neurologic manifestations may dominate the clinical picture.

Chronic hyperhomocysteinemia (Chapter 209) is a risk factor for several diseases (Table 164-2), many of which are benefited by folate supplementation or homocysteine-lowering therapy with combined folic acid, cobalamin, and pyridoxine. Thus, by inference, patients can present with any of these conditions arising from long-standing hyperhomocysteinemia or poor folate status.

E-TABLE 164-1 SERUM FOLATE LEVELS ARE MISLEADINGLY ELEVATED IN COBALAMIN DEFICIENCY AND MALARIA, WHICH ARE BOTH COMMON IN RESOURCE-LIMITED SETTINGS*,†

	SERUM FOLATE LEVEL	RBC FOLATE LEVEL	SERUM COBALAMIN LEVEL
Pure folate deficiency	Low	Low	Normal/low*
Pure cobalamin deficiency‡	Normal/high*,‡	Low*,‡	Low
Folate plus cobalamin deficiency	Normal*	Low	Low
Pure malaria	Normal*/high*,§	High*,‖,¶	Normal
Malaria plus folate deficiency	Normal*	Normal*/high*/low	Normal
Malaria plus cobalamin deficiency	Normal*/high*,§	Normal*/high*,¶	Low
Malaria plus folate plus cobalamin deficiency	Normal*/high*,§	Low/normal*,¶	Low

RBC = red blood cell.
*The *asterisk* indicates misleading values in the clinical settings shown on the *left*.
†Both cobalamin deficiency and clinical malaria and other hemolytic states can complicate the diagnosis of folate deficiency with tests for serum or RBC folate.
‡Cobalamin deficiency is accompanied by inability to use folate, so RBC folate leaks out into serum.
§Release of the 30-fold excess folate from RBCs during hemolysis raises serum folate. An as yet unknown quantity of folate is released into blood when folate-rich hepatocytes are destroyed during the exoerythrocytic hepatic phase of malaria.
‖Hemolysis induces a compensatory reticulocytosis; these reticulocytes are richer in folate than mature RBCs.
¶*Plasmodium falciparum* can synthesize folate in RBC cultures in vitro and raises RBC folate in animal models with high levels of parasitemia; conversely, RBC folate has been normalized after therapy with chloroquine. Reticulocytopenia in severe *Plasmodium falciparum* malaria—due to combined cobalamin deficiency plus folate deficiency that can trigger a reticulocytopenic (megaloblastic) crisis, other nutrient deficiency, or cytokine-induced inhibition of hematopoiesis—will negate a rise in RBC folate.

DIAGNOSIS

Diagnostic Approach to the Patient

The general approach to a patient with megaloblastic anemia is first to recognize that megaloblastic anemia is present; then to distinguish whether folate, cobalamin, or combined folate and cobalamin deficiencies have led to the anemia; and finally to diagnose the underlying disease and mechanism causing the deficiency[9] (see Table 164-1).

TABLE 164-2 EFFECTS OF HOMOCYSTEINE-LOWERING THERAPY ON NONHEMATOPOIETIC SYSTEMS

FOLIC ACID, COBALAMIN, AND PYRIDOXINE SUPPLEMENTATION
Reduction in hip fracture
Reduction in progression of carotid intima-media thickness (a surrogate marker of early subclinical arteriosclerosis)
Reduction in age-related macular degeneration
Reduction in rate of brain atrophy
Reduction in rate of cognitive decline (folate plus cobalamin only)
FOLIC ACID SUPPLEMENTATION
Reduction in stroke
Reduction in age-related (sensorineural) hearing loss
Reduction in phenytoin-induced gingival hyperplasia
FOLIC ACID FORTIFICATION OF FOOD (POPULATION-BASED STUDIES)
Reduction in neural tube defects (anencephaly, spina bifida, encephalocele, meningocele, iniencephaly)
Reduction in cleft lip with or without cleft palate
Reduction in severe congenital heart disease (endocardial cushion defects, conotruncal defects)
Reduction in congenital pyloric stenosis, stenosis of the pelvic-ureteric junction, limb reduction defects, omphalocele
Reduction in stroke mortality
Decreased risk of preterm births, low-birthweight and small-for-gestational-age babies
No evidence of increase in cancer
DELETERIOUS EFFECT OF HOMOCYSTEINE-LOWERING THERAPY
Diabetic nephropathy

Although deficiencies of cobalamin and folate are only two of the many causes of macrocytosis (Fig. 164-2), they become increasingly more likely as the MCV increases. Because perturbed DNA synthesis from any cause (including folate and cobalamin deficiency) results in megaloblastosis of bone marrow precursor cells, the red cells released into the circulation have an MCV that is often greater than 110 fL; on the peripheral blood smear, these large cells appear oval (macro-ovalocytes). Compared with mature red cells, which normally have a central area of pallor that occupies about one third of the cell diameter, the central pallor of macro-ovalocytes is significantly reduced. By contrast, thin macrocytes, which contain an increased cell surface without a proportionate increase in volume, have a much larger area of central pallor (more than one-third the cell diameter). When reticulocytes, which are normally 20% larger than mature red cells, are prematurely released from the bone marrow during the stress of acute blood loss or hemolysis, these so-called shift reticulocytes are even larger. Therefore, assessment of the corrected reticulocyte count is a good starting point to distinguish whether the macrocytic anemia is due to increased reticulocytosis (see Fig. 164-2); if reticulocytopenia is evident, data from the history, physical findings, and peripheral smear showing either thin macrocytes or macro-ovalocytes will help narrow the differential diagnosis. The frequency with which a high MCV is found depends on the patient population studied. In U.S. hospitals, up to two thirds of cases of macrocytosis (MCV ≥100) can be due to chemotherapy or antiretroviral therapy, alcoholism, or liver disease. In resource-limited regions, insufficient intake of cobalamin, folate, and iron can result in a dimorphic anemia (with a large red cell distribution width from a mix of macro-ovalocytic and microcytic cells).

History and Physical Examination

The underlying condition predisposing to folate deficiency usually began within the previous 6 months and often dominates the overall clinical picture. By contrast, cobalamin deficiency takes several years to be manifested clinically. Therefore, the underlying condition is more chronic and symptoms develop more insidiously; defining the cause is a clinical challenge that necessitates a detailed history, physical findings, and the judicious use of laboratory studies.

The dietary history may be revealing (food faddism, vegetarianism), whereas the medical or family history may uncover gluten sensitivity, autoimmune diseases, epilepsy treated with an anticonvulsant, use of offending

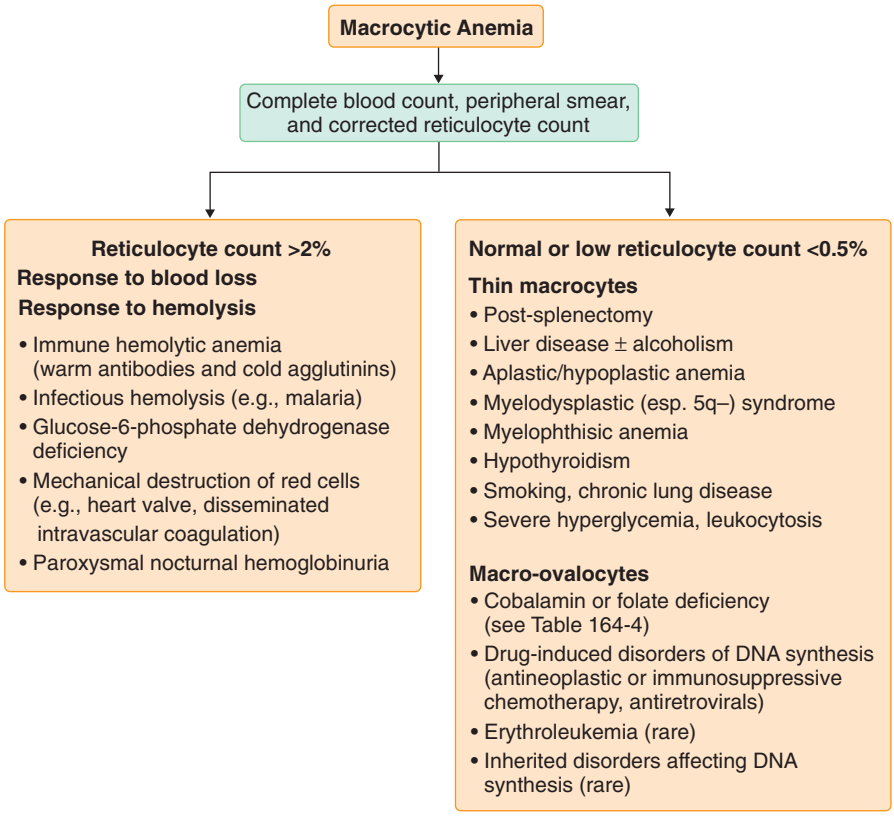

FIGURE 164-2. Algorithm for the evaluation of patients with macrocytosis.

drugs, hemolytic anemia, past surgery (e.g., gastrectomy, fistula, bowel resection), inhalation of N_2O, or travel history predictive for tropical sprue.

Physical examination of cobalamin-deficient vegetarians or those with pernicious anemia may reveal well-nourished individuals. By contrast, patients with folate deficiency are poorly nourished and may have other stigmata of multiple deficiencies from malabsorption (Chapter 140). Associated deficiency of iron and vitamins A, D, and K or protein-calorie malnutrition, or both, may give rise to angular cheilosis, bleeding mucous membranes, dermatitis, osteomalacia, and chronic infections. Varying degrees of pallor with lemon-tint icterus (a combination of pallor and icterus best observed in fair-skinned individuals) are common features of megaloblastosis. The skin may reveal either a diffuse brownish pigmentation or abnormal blotchy tanning. Premature graying is observed in both light- and dark-haired individuals. Examination of the mouth may reveal glossitis, with a smooth (depapillated), beefy-red tongue with occasional ulceration of the lateral surface. Thyromegaly may be observed with pernicious anemia and associated autoimmune thyroid disease. Heart failure from severe anemia may be accompanied by mild splenomegaly reflecting extramedullary hematopoiesis.

In prolonged cobalamin deficiency, neurologic examination reveals evidence of involvement of the posterior columns as well as of the pyramidal, spinocerebellar, and spinothalamic tracts. Posterior column dysfunction results in loss of position sense in the index toes (before great toe involvement) (Chapter 416) and loss of the ability to discern vibration of a high-pitched (256 cycles/second) tuning fork. Diminished vibratory sensation and proprioception of the lower extremities are the most common early objective signs. Neuropathic involvement of the legs precedes that of the arms. Upper motor neuron signs may be modulated by the subsequent involvement of peripheral nerves. A Romberg sign and a Lhermitte sign may be elicited. Loss of sphincter and bowel control or involvement of cranial nerves, such as optic neuritis, may be accompanied by other dysfunction of the cerebral cortex, including dementia, psychoses, and disturbances of mood. Cognitive impairment is not uncommon among vegetarians with cobalamin deficiency; a study reported that half had impaired recall and serial sevens, and one quarter had impaired naming. Objective findings of abnormal evoked potential (by the auditory oddball paradigm with P300 latency) are seen in half of these patients. The coexistence of folate deficiency with neurologic disease should prompt investigations to exclude cobalamin and other nutrient deficiencies arising from dietary insufficiency or malabsorption.

Nutritional cobalamin deficiency in resource-limited countries can be manifested as florid pancytopenia, mild hepatosplenomegaly, fever, and thrombocytopenia, with the neuropsychiatric syndrome developing as a later manifestation. Many of these individuals will have associated folate and iron deficiency. However, cobalamin-related neurologic disease has also been found in patients with only mild to moderate anemia secondary to cobalamin deficiency in both developing and developed countries. Indeed, in the United States, between 25 and 50% of patients who have neuropsychiatric abnormalities attributable to cobalamin deficiency may have a normal hematocrit and MCV. There is often an inverse correlation between the hematocrit and neurologic disease in cobalamin deficiency; most subjects have mild neurologic deficits, and 25% have only moderate deficits, with paresthesias or ataxia as the initial symptoms.

Laboratory Tests

Megaloblastosis

To establish the diagnosis of megaloblastosis, the evaluation begins with a complete blood count, MCV (which often reveals a steady increase during a period of several months or years), examination of the peripheral smear, and reticulocyte count; a low reticulocyte count with macro-ovalocytes suggests an underlying megaloblastic anemia (see Fig. 164-2). Classic megaloblastosis from cobalamin or folate deficiency may be accompanied by a hemoglobin level of less than 5 g/dL. Neutropenia and thrombocytopenia occur less commonly than anemia and are usually not severe. On occasion, neutrophil counts less than 1000/μL and platelet counts less than 50,000/μL can be seen. Additional abnormalities supporting intramedullary hemolysis include elevated levels of serum lactate dehydrogenase and bilirubin as well as decreased serum haptoglobin levels.

Peripheral Smear

In peripheral blood, the earliest manifestation of megaloblastosis is an increase in MCV with macro-ovalocytes (up to 14 μm). Nuclear hypersegmentation of neutrophils, diagnosed if more than 5% of polymorphonuclear leukocytes have five lobes or if 1% have six lobes on the smear (Fig. 164-3), strongly suggests megaloblastosis, especially in association with macro-ovalocytosis. However, neutrophil hypersegmentation is not sensitive for the diagnosis of mild cobalamin deficiency, and macrocytosis is absent in nearly 50% of cases. There may be associated teardrop-shaped erythrocytes and anisocytosis with leukopenia and thrombocytopenia.

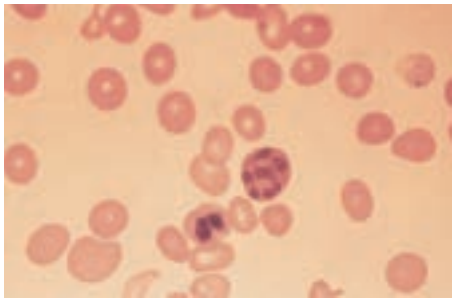

FIGURE 164-3. Megaloblastic anemia. The peripheral blood has oval macrocytes (large red blood cells) and marked neutrophil hypersegmentation.

TABLE 164-3 COMMON CONDITIONS PREDISPOSING TO MASKED MEGALOBLASTOSIS

Inadequate dietary intake of cobalamin, folate, and iron
Vegetarianism
Near-vegetarianism (in resource-limited settings)
Cobalamin *plus* iron deficiency
Post-gastrectomy
Pernicious anemia
Celiac disease (late manifestation)
Folate *plus* iron deficiency
Celiac disease (early manifestation)
Pregnancy in a woman not taking iron and folate supplements
Alcoholism with liver disease and gastrointestinal bleeding
Miscellaneous common clinical combinations
Rheumatoid arthritis with NSAIDs and gastrointestinal bleeding with methotrexate
Thalassemia with folate or cobalamin deficiency
Any patient with cancer receiving chemotherapy with chronic bleeding or iron deficiency
Antiretroviral therapy plus iron deficiency/thalassemia

NSAIDs = nonsteroidal anti-inflammatory drugs.

Megaloblastic anemia can be masked (Table 164-3) when there is a coexisting condition that neutralizes the tendency to generate large cells, such as iron deficiency (Chapter 159) or thalassemia (Chapter 162). In these situations, giant myelocytes and metamyelocytes in bone marrow and hypersegmented polymorphonuclear neutrophils in bone marrow and peripheral blood (see Fig. 164-3) are important clues to a masked megaloblastosis. This problem is clinically relevant because appropriate replacement with cobalamin or folate elicits a maximal hematologic response only when any associated iron deficiency is corrected. Conversely, if a combined iron and cobalamin deficiency (after gastrectomy) or iron and folate deficiency (with pregnancy) is treated with iron alone, megaloblastosis will be unmasked.

Cobalamin and Folate Levels

Laboratory evaluation of suspected cobalamin or folate deficiency begins with measurement of the serum levels of these vitamins. If any of these results are borderline, one can proceed to confirmatory tests by serum levels of metabolites (homocysteine and methylmalonic acid) (Table 164-4). Use of clinical information will improve the pretest probability of low serum cobalamin and folate levels in diagnosis of deficiency. Indeed, without detailed clinical information, the combined results of serum cobalamin, folate, and metabolite tests are not sufficiently unambiguous to diagnose cobalamin deficiency and to distinguish it from combined cobalamin plus folate deficiency.

Serum Cobalamin Levels

A low serum cobalamin level (<200 pg/mL) is a valuable clinical indicator of true tissue cobalamin deficiency in about 90% of patients. It is, however, relatively less sensitive compared with metabolite levels. Even among 173 unambiguously cobalamin-deficient patients, about 5% had normal cobalamin levels. Serum cobalamin is less than 300 pg/mL in 99% of patients

TABLE 164-4 STEPWISE APPROACH TO THE DIAGNOSIS OF COBALAMIN AND FOLATE DEFICIENCY

MEGALOBLASTIC ANEMIA OR NEUROLOGIC-PSYCHIATRIC MANIFESTATIONS CONSISTENT WITH COBALAMIN DEFICIENCY *PLUS* TEST RESULTS OF SERUM COBALAMIN AND SERUM FOLATE			
Cobalamin* (pg/mL)	**Folate† (ng/mL)**	**Provisional Diagnosis**	**Proceed with Metabolites?‡**
>300	>4	Cobalamin or folate deficiency unlikely	No
<200	>4	Consistent with cobalamin deficiency	No
200-300	>4	Rule out cobalamin deficiency	Yes
>300	<2	Consistent with folate deficiency	No
<200	<2	Consistent with combined cobalamin plus folate deficiency or with isolated folate deficiency	Yes
>300	2-4	Consistent with folate deficiency or with anemia unrelated to vitamin deficiency	Yes

TEST RESULTS OF METABOLITES: SERUM METHYLMALONIC ACID AND TOTAL HOMOCYSTEINE		
Methylmalonic Acid (normal = 70-270 nM)	**Total Homocysteine (normal = 5-14 µM)**	**Diagnosis**
Increased	Increased	Cobalamin deficiency confirmed; folate deficiency still possible (i.e., combined cobalamin plus folate deficiency possible)
Normal	Increased	Folate deficiency likely; <5% may have cobalamin deficiency
Normal	Normal	Cobalamin and folate deficiencies excluded

*Serum cobalamin levels: abnormally low, <200 pg/mL; clinically relevant low-normal range, 200-300 pg/mL. Methylmalonic acid is usually above 1000 nM in clinical cobalamin deficiency
†Serum folate levels: abnormally low, <2 ng/mL; clinically relevant low-normal range, 2-4 ng/mL. Homocysteine is usually above 25 µM in clinical folate deficiency. See E-Table 164-1 for additional limitations in the clinical interpretation of serum folate concentrations.
‡Any frozen-over sample from the serum folate or cobalamin determination can be subjected to metabolite tests.

with clinical hematologic or neurologic manifestations of cobalamin deficiency, whereas a cobalamin level above 300 pg/mL suggests folate deficiency or another cause of macrocytosis or neurologic disease. There have recently been several reports of patients who have had florid clinical evidence of cobalamin deficiency but whose serum cobalamin levels (by competitive binding luminescence assays) were either normal or in the high-normal range[10]; this is apparently due to the interference with the test by high titers of anti–intrinsic factor antibodies associated with pernicious anemia, a serious instance of a false-negative test result. This will lead to clinical manifestations if it is uncorrected. Therefore, if the clinical scenario is strongly suggestive of cobalamin deficiency but the test result is normal, the patient deserves a therapeutic trial with cobalamin. Confirmation of cobalamin deficiency with metabolites and detection of anti–intrinsic factor antibodies can allow retrospective diagnosis of pernicious anemia. Individuals with neuropsychiatric disorders attributed to cobalamin deficiency may not have anemia and only minimally depressed cobalamin levels. The serum cobalamin concentration can be falsely low in the absence of true cobalamin deficiency in patients with folate deficiency (one third of patients), pregnancy, multiple myeloma, transcobalamin I deficiency, and megadose vitamin C therapy.

An increase in cobalamin levels may be an unexpected finding in patients tested for cobalamin deficiency, for example, with high transcobalamin I and II levels (in myeloproliferative neoplasms, hepatic tumors, or active liver disease) and when transcobalamin II–producing macrophages are activated (in autoimmune diseases, monoblastic leukemias, and lymphomas). Approximately 10% of the U.S. population, especially the elderly, probably have metabolic evidence of true cobalamin deficiency.

Serum Folate Levels

In combination with a clinical picture of megaloblastic anemia and measurement of serum cobalamin, the serum folate level is the cheapest and best initial test for diagnosis of folate deficiency. Although red blood cell folate levels by earlier microbiologic assays correlated well with hepatic folate stores, the current assays for red blood cell folates are unreliable for routine clinical use.

The serum folate level is highly sensitive to the intake of a single folate-rich meal. Nutritional folate deficiency first leads to a decline in the serum folate level below normal (<2 ng/mL) in about 3 weeks; thus, it is a sensitive indicator of negative folate balance. Isolated reduction of serum folate in the absence of megaloblastosis (i.e., a false-positive result) occurs in one third of hospitalized patients with anorexia and acute alcohol consumption, in normal pregnancy, and in those using anticonvulsants. Because these groups are also at high risk for folate deficiency, additional testing with metabolites or an empirical therapeutic trial is indicated. Both cobalamin deficiency and malaria, which are common in resource-limited countries, can falsely elevate serum folate levels and mask the diagnosis of folate deficiency. In this scenario, the dietary history of folate intake is a better predictor of folate deficiency (see later).

Metabolite Levels: Homocysteine and Methylmalonic Acid

The serum methylmalonic acid (MMA) and homocysteine levels are highly sensitive tests, both of which rise in proportion to the severity of cobalamin deficiency; by contrast, the serum homocysteine level alone rises with folate deficiency (see Table 164-4). Therefore, an increase in both metabolites cannot differentiate between isolated cobalamin deficiency and combined cobalamin plus folate deficiency. These are too expensive to use as initial screening tests. Serum MMA levels are elevated in more than 95% of patients with clinically confirmed cobalamin deficiency (with median values of 3500 nM). Serum homocysteine concentrations are elevated in both cobalamin deficiency (median values of 70 µM) and folate deficiency (median values of 50 µM). Both homocysteine and MMA rise with dehydration or renal failure. Propionic acid, derived from anaerobic fecal bacterial metabolism, can raise MMA, which can be lowered by a course of metronidazole. The abnormally high metabolites return to normal in a week with the appropriate (deficient) vitamin replacement.

Clinicians can use serum MMA and homocysteine to assist in the diagnosis of patients with the following: borderline cobalamin and folate levels (see Table 164-4); existing conditions known to perturb folate and cobalamin tests, leading to difficulties in interpreting the results; low levels of both cobalamin and folate, in which case a high MMA level is useful to confirm cobalamin deficiency (rather than attributing the condition to folate deficiency alone); and low serum cobalamin levels when there is an alternative explanation for the syndrome that led to the test (e.g., a diabetic or alcoholic patient with peripheral neuropathy, or an alcoholic patient with a high MCV and low cobalamin level without anemia).

Bone Marrow Examination

In a plausible clinical setting for a patient with severe hematologic disease, the identification of nucleated red cells with megaloblastic features in the peripheral smear and classic macro-ovalocytes and hypersegmented polymorphonuclear neutrophils—which reflects the morphology in the bone marrow—can help clinch the diagnosis of megaloblastosis. If they are not found, a bone marrow aspirate can be invaluable in making a rapid diagnosis of megaloblastosis in a patient with severe anemia lacking classic findings. In the outpatient setting, when there is less urgency and anemia is only mild to moderate in a patient with a suggestive peripheral smear or when the clinical presentation is primarily neuropsychiatric, a good case can be made to initiate the sequence of diagnostic tests without bone marrow aspiration and to proceed with the measurement of serum levels of vitamins or metabolites (see Table 164-4 and Fig. 164-2).

On the bone marrow aspirate (Fig. 164-4), which is better than biopsy for observing megaloblastosis, the cells are actually proliferating very slowly despite what looks like exuberant cell proliferation with numerous mitotic figures. In early cobalamin or folate deficiency, normoblasts may dominate the marrow, with only a few megaloblasts seen, but the full spectrum of megaloblastic hematopoiesis is observed in severe deficiency and is accompanied by varying degrees of pancytopenia. In contrast to the normally dense chromatin of comparable normoblasts, megaloblastic erythroid precursors have an open, finely stippled, reticular, sieve-like pattern. The orthochromatic megaloblast, with its hemoglobinized cytoplasm, continues to retain its large, sieve-like immature nucleus, in sharp contrast to the clumped chromatin of orthochromatic normoblasts. Up to 90% of megaloblastic cells die in the bone marrow, with remnants of apoptotic cells giving an impression of myelodysplastic syndrome. These are scavenged by macrophages in a process

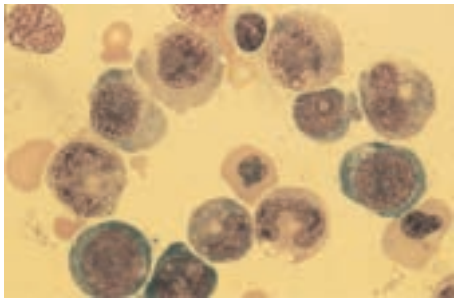

FIGURE 164-4. Megaloblastic anemia. A bone marrow aspirate shows red blood cell precursors that are giant megaloblasts with nuclear-cytoplasmic dissociation (nuclear maturation lagging behind cytoplasmic maturation). Megaloblastic changes in the leukocyte series are shown by the "giant metamyelocyte."

called ineffective erythropoiesis or intramedullary hemolysis. There is an increase in white blood cell production, with megaloblastic leukocytes also having a sieve-like chromatin. Giant (20 to 30 μm) metamyelocytes and "band" forms are pathognomonic for megaloblastosis. Hypersegmented polymorphonuclear leukocytes will be seen in the marrow and peripheral blood. Megakaryocytes may be normal or increased in number and can exhibit complex hypersegmentation, with liberation of fragments of cytoplasm and giant platelets into the circulation. The net output of platelets is decreased in severe megaloblastosis.

Determining the Cause of the Vitamin Deficiency

By the time that megaloblastic anemia is established, the cause of folate deficiency is usually clear from the history, physical examination, and clinical setting. With rare exceptions, adults with cobalamin deficiency have either cobalamin malabsorption or dietary cobalamin insufficiency. Whereas dietary cobalamin insufficiency is diagnosed by a dietary history and is easily prevented by small doses of prophylactic daily oral cobalamin, all other causes of cobalamin malabsorption respond to either parenteral cobalamin (warranting multiple loading doses followed by monthly maintenance) or daily high doses of oral cobalamin (1 to 2 mg/day). This raises a fundamental question: Why do we need to pursue the cause of cobalamin malabsorption? One practical reason is to determine the need for additional diagnostic tests (e.g., intestinal biopsy, examination of stool for malabsorption or *D. latum* infestation) to institute specific therapy (e.g., gluten-free diet, folate, antibiotics, anthelmintics). This evaluation, in turn, dictates whether cobalamin replacement should be lifelong.

The Schilling test, which was discontinued in 2003, provided such leads on the locus and mechanism of cobalamin malabsorption. Nowadays, without this test, the low sensitivity of serum anti–intrinsic factor antibodies allows only 60% of patients with pernicious anemia to be confidently identified through this measurement. Even poorer performance characteristics have rendered serum antiparietal cell antibodies diagnostically unhelpful. Although increased fasting levels of gastrin and low levels of pepsinogen I (a combination that suggests oxyntic gastric mucosal damage, like atrophy and hypochlorhydria) lack specificity, their sensitivity for detection of pernicious anemia is 90 to 92%. Therefore, some hematologists combine the specific but insensitive serum anti–intrinsic factor antibody test with the sensitive but nonspecific serum gastrin or pepsinogen I abnormalities to rule out pernicious anemia in patients with cobalamin deficiency in view of the loss of availability of the Schilling test.[11]

Because most of the conditions predisposing to cobalamin deficiency (see Table 164-1) should be clinically evident by the time that cobalamin deficiency is apparent, it is possible to identify several conditions through a detailed dietary history or past medical history. Such an exercise can point to esophagogastroduodenal disease, pancreatic insufficiency, impaired bowel motility, or other autoimmune diseases. The physical examination can also provide additional clues and suggest focused testing for rarer conditions (stool for ova; serum anti–tissue transglutaminase antibodies, lipase, or gastrin; intestinal biopsy; or radiographic contrast studies for stasis, strictures, or fistulas). Thus, by this classical medicine approach, it should be possible to identify the basis for cobalamin malabsorption and the duration of therapy in most instances (even without the Schilling test). For the younger patient with megaloblastic anemia, distinguishing juvenile pernicious anemia from congenital intrinsic factor deficiency warrants measurement of gastric juice for intrinsic factor and achlorhydria, and Imerslund-Gräsbeck syndrome requires DNA to detect mutations in cubam receptor (amnionless, cubilin) genes (see Table 164-1).

The utility of measurement of blood levels of holotranscobalamin (cobalamin-bound transcobalamin II) to evaluate cobalamin deficiency requires further clinical validation.

TABLE 164-5 CAUSES OF MEGALOBLASTIC ANEMIA NOT RESPONDING TO THERAPY WITH COBALAMIN OR FOLATE

Wrong diagnosis
Combined folate and cobalamin deficiencies being treated with only one vitamin
Associated iron deficiency
Associated hemoglobinopathy (e.g., sickle cell disease, thalassemia)
Associated anemia of chronic disease
Associated hypothyroidism

TREATMENT

If a severely anemic patient is decompensated or decompensation is imminent, after blood is drawn for diagnostic studies, the patient is transfused slowly with initially only 1 U of packed red cells under diuretic coverage, and both cobalamin and folate are immediately started at full doses (parenteral cobalamin 1 mg/day plus folic acid 1 mg/day), even before the type of vitamin deficiency has been established. If the patient is only moderately symptomatic, transfusions should be avoided because a dramatic improvement in well-being is likely to occur within 2 to 3 days of starting full-dose vitamin replacement, even before hematologic improvement. If the patient is well compensated or in the ambulatory setting, diagnostic testing should proceed in an orderly sequence (see Fig. 164-2 and Table 164-4) before therapy is initiated.

Drug Dosage

Established Cobalamin Deficiency

For patients with severe cobalamin deficiency from any cause, an aggressive scheme to replace cobalamin rapidly is 1 mg/day of intramuscular or subcutaneous cyanocobalamin (week 1), 1 mg twice weekly (week 2), 1 mg/week for 4 weeks, and then 1 mg/month for life. For mild cobalamin deficiency, one can employ oral cobalamin 2 mg/day for 3 months, which exploits passive intestinal diffusion, through which 1 to 2% of the oral dose is absorbed. Once stores are fully replenished, those with cobalamin malabsorption can either continue with monthly maintenance or be switched to daily oral cobalamin 2 mg/day. For nutritional cobalamin deficiency, lifelong daily oral cobalamin (5 to 10 μg), with either tablets or equivalent cobalamin-fortified foods, should be instituted, but only after cobalamin stores are replenished. It is worth reminding patients that food intake reduces absorption of oral cobalamin by 50%.

In cobalamin deficiency associated with either pernicious anemia or food cobalamin malabsorption, the malabsorption of iron (due to achlorhydria) requires oral iron coadministered with organic acids (e.g., ascorbic acid) or total-dose parenteral iron replacement (e.g., low-molecular-weight iron dextran) (Chapter 159) to elicit a complete hematologic response.

An incomplete response must trigger a search for additional conditions (Table 164-5).

Subclinical Cobalamin Deficiency

The entity of subclinical cobalamin deficiency is encountered when an apparently asymptomatic patient has serum cobalamin values that either hover or fluctuate just above or just below normal cutoff values for many months and sometimes even inexplicably return to the normal range. This can sometimes be due to a deficiency of transcobalamin I, which binds the bulk of serum cobalamin but is of little clinical significance. Some experts proceed to obtain metabolite studies and then ignore cobalamin values if the MMA level is normal. Others periodically observe these patients and treat with cobalamin only when patients exhibit symptoms or signs of frank cobalamin deficiency. Yet others err on the side of caution, preferring to preemptively replenish depleted cobalamin stores and thereby circumvent any potential for the patient to develop impaired cognitive function, which can be detected only by specialized neurophysiologic testing, or even frank cerebral atrophy, which has been documented by magnetic resonance imaging. Accordingly, they empirically treat those with borderline cobalamin levels for 6 months with oral cobalamin (1 to 2 mg/day) and reassess serum cobalamin levels to ensure that values have returned to a mid to upper end of the normal range. This approach warrants repeated testing of cobalamin levels 6 to 12 months later to evaluate whether there has been an interim fall in cobalamin levels. At that time, options of either instituting cobalamin for the long term or

alternating treatment for the first 6 months with no treatment for the next 6 months can be discussed with the patient.

Established Folate Deficiency

Oral folate (folic acid) at doses of 1 mg/day results in adequate absorption despite intestinal malabsorption of physiologic food folate. Therapy should be continued until complete hematologic recovery is documented; the subsequent duration of therapy is dictated by the cause. Folinic acid bypasses any block in one-carbon metabolism induced by methotrexate, trimethoprim-sulfamethoxazole, or N_2O.

Prophylaxis with Cobalamin or Folate

Several conditions are prevented by specific treatment designed either to lower homocysteine or to improve folate status (E-Table 164-2; see also Table 164-2). As one example, in elderly subjects with mild cognitive impairment, homocysteine lowering by a cocktail that included cobalamin and folate during 2 years did slow the rate of brain atrophy by almost 30%.[A4] For vegetarians, cobalamin-fortified foods (soy or rice beverages, fortified cereals, nutritional yeast) or oral cobalamin tablets 5 to 10 µg/day should suffice. For those with malabsorption of cobalamin from any other mechanism (see Table 164-1), cobalamin replacement is achieved with 1- to 2-mg tablets taken orally each day.

Periconceptional folate supplementation for all normal women (400 µg/day of folic acid) and for women who have previously delivered a baby with a neural tube defect (4 mg/day of folic acid) is now standard and prevents nearly three quarters of neural tube defects. Women of childbearing age who are taking anticonvulsant medications (phenytoin, phenobarbital, carbamazepine) are also at increased risk for delivering babies with neural tube defects and should routinely take 1 mg of folic acid daily. Moreover, administration of folic acid before initiation of phenytoin can eliminate the risk of gingival hyperplasia.[A5] Folate supplementation throughout pregnancy also helps prevent premature delivery of low-birthweight infants and is recommended for premature infants and lactating mothers. Folic acid supplements (1 mg/day orally) are taken by patients with hemolysis or myeloproliferative diseases and also to reduce the toxicity of methotrexate in patients with rheumatoid arthritis and psoriasis. Those individuals in whom cobalamin deficiency develops while they are receiving long-term folate replacement will present with a pure neurologic syndrome and not necessarily anemia.

The mandatory fortification of food (flour, enriched pasta, cornmeal, cereal foods) with 140 and 150 µg of folic acid per 100 g in the United States and Canada, respectively, has led to several beneficial effects (see Table 164-2). Indeed, 94% of U.S. adults who do not consume supplements or consume less than 400 µg/day of folic acid from supplements do not exceed the upper intake limit for folic acid (>1 mg/day), which has potential to mask cobalamin deficiency. Studies that have investigated adverse effects of fortification on development or progression of cancer have not identified that this is an issue.[A6] Moreover, folate deficiency–related anemia has nearly been eliminated in older U.S. adults after folate fortification.[12]

Prophylaxis in Resource-Limited Settings

Children of mothers from resource-limited countries with nutritional folate and cobalamin deficiency are at risk for cognitive dysfunction as toddlers and as young children[13]; hence, they deserve early prophylaxis with both folate and cobalamin. The serious and chronic problem of iron, folate, and cobalamin deficiency, especially among women in North India, which is due to poor dietary intake, and the high incidence of neural tube defects[14] warrant improving their nutrition through the use of iron, folate, and vitamin B_{12} supplements. Moreover, all pregnant women living in resource-limited malarious areas are at risk for iron, folate, and vitamin B_{12} deficiency[15] and malaria, so they should routinely be given intermittent preventive treatment according to the latest guidelines; currently, it is sulfadoxine-pyrimethamine (Chapter 345) together with insecticide-treated bed nets, plus 1 mg folic acid, prophylactic vitamin B_{12} (~10 to 25 µg), and replacement doses of oral iron. Likewise, women requiring antimalarial treatment should also be given iron, folate, and vitamin B_{12} to enable optimal hematopoiesis. Finally, although *Plasmodium falciparum* possesses two folate transporter proteins, in vitro as well as clinical studies using low-dose folic acid (1 mg/day) in 467 pregnant women with malaria have demonstrated benefit; by contrast, high-dose folic acid (5 mg/day) can "feed the parasite" and is contraindicated. Finally, optimizing the nutrition of the approximately 500 million adolescent girls before they become pregnant is a priority because maternal death is one of the most common causes of death for teenage girls between 15 and 19 years of age.

PROGNOSIS

The general response to cobalamin replacement is a dramatic improvement in well-being, with alertness, a good appetite, and resolution of a sore tongue. Megaloblastic hematopoiesis reverts to normal within 12 hours and resolves by 48 hours; the only persistent findings may be giant metamyelocytes in the bone marrow and hypersegmented neutrophils in the blood for up to 14 days. The reticulocyte count peaks by days 5 to 8, followed by a rise in the red cell count, hemoglobin level, and hematocrit. By the end of the first week, the white blood cell count rises, sometimes with a transient left shift, as does the platelet count; all three cell counts should normalize in 2 months.

With cobalamin replacement, the degree of reversal of neurologic damage is generally inversely related to the extent of disease and the duration of signs and symptoms. Most neurologic abnormalities improve in up to 90% of patients with documented subacute combined degeneration, and most signs and symptoms of less than 3 months' duration are reversible. With signs and symptoms of longer duration, there is invariably some residual neurologic dysfunction. The maximal response often takes up to 6 months, but recovery beyond 12 months is unusual.

Incorrect treatment of cobalamin deficiency with folate does not improve the neuropsychiatric abnormalities, which will continue to progress; hematologic improvements often occur, however. Alternatively, there may be associated iron deficiency or hypothyroidism that needs specific replacement or another hemoglobinopathy (e.g., sickle cell disease, thalassemia) that limits the normalization of hemoglobin values (see Tables 164-3 and 164-5).

In patients with pernicious anemia, subsequent iron deficiency anemia (Chapter 159), osteoporosis (Chapter 243) with fractures of the proximal end of the femur and vertebrae, gastrinoma and gastric cancer (Chapter 192), and cancer of the buccal cavity and pharynx can develop. Some experts recommend periodic upper endoscopic surveillance.

CLINICAL PRACTICE IN RESOURCE-LIMITED SETTINGS

Masking of Folate Deficiency with Associated Cobalamin Deficiency or Malaria

In resource-limited settings where vegetarianism or poverty-imposed near-vegetarianism is common, nutritional cobalamin as well as folate deficiency and hemolysis accompanying malaria often coexist. In such settings, the serum folate level is artificially raised in cobalamin deficiency and returns to baseline only after cobalamin replacement; likewise, the serum folate level is increased during hemolysis, which releases the extant 30-fold higher erythrocyte folate concentration into plasma. In this setting, measurement of the serum folate concentration will reveal normal to high values, predictably underestimate the tissue folate status, and completely miss the diagnosis of mild to moderate folate deficiency (see E-Table 164-1). By contrast, the dietary history or more formal assessment with a variety of methods (such as 24-hour recall, estimated/weighed record, or locally validated food frequency questionnaires), which are established methods to evaluate the quality and quantity of nutrients consumed, can predict the dietary folate and cobalamin intake of a population and identify those at risk for nutritional insufficiency.[15] Indeed, knowledge of dietary folate and cobalamin intake trumps the results of conventional blood tests for folate deficiency, which are seriously flawed in this clinical setting. In summary, all individuals at risk for nutritional anemia due to deficiency of iron, folate, and cobalamin (both adults and children) should be given prophylactic oral cobalamin, folate, and iron replacement; and in malarious zones, these should be combined with antimalarial drugs and insecticide-treated bed nets.

Grade A References

A1. Yang J, Hu X, Zhang Q, et al. Homocysteine level and risk of hip fracture: a meta-analysis and systematic review. *Bone*. 2012;51:376-382.

A2. Qin X, Xu M, Zhang Y, et al. Effect of folic acid supplementation on the progression of carotid intima-media thickness: a meta-analysis of randomized controlled trials. *Atherosclerosis*. 2012;222:307-313.

A3. Christen WG, Glynn RJ, Chew EY, et al. Folic acid, pyridoxine, and cyanocobalamin combination treatment and age-related macular degeneration in women: the Women's Antioxidant and Folic Acid Cardiovascular Study. *Arch Intern Med*. 2009;169:335-341.

A4. Smith AD, Smith SM, de Jager CA, et al. Homocysteine-lowering by B vitamins slows the rate of accelerated brain atrophy in mild cognitive impairment: a randomized controlled trial. *PLoS ONE*. 2010;5:e12244.

A5. Walker JG, Batterham PJ, Mackinnon AJ, et al. Oral folic acid and vitamin B-12 supplementation to prevent cognitive decline in community-dwelling older adults with depressive symptoms—the Beyond Ageing Project: a randomized controlled trial. *Am J Clin Nutr*. 2012;95:194-203.

A6. Wang X, Qin X, Demirtas H, et al. Efficacy of folic acid supplementation in stroke prevention: a meta-analysis. *Lancet*. 2007;369:1876-1882.

A7. Durga J, Verhoef P, Anteunis LJ, et al. Effects of folic acid supplementation on hearing in older adults: a randomized, controlled trial. *Ann Intern Med*. 2007;146:1-9.

A8. Arya R, Gulati S, Kabra M, et al. Folic acid supplementation prevents phenytoin-induced gingival overgrowth in children. *Neurology*. 2011;76:1338-1343.

A9. Vollset SE, Clarke R, Lewington S, et al. Effects of folic acid supplementation on overall and site-specific cancer incidence during the randomised trials: meta-analyses of data on 50,000 individuals. *Lancet*. 2013;381:1029-1036.

E-TABLE 164-2 INDICATIONS FOR PROPHYLAXIS WITH COBALAMIN OR FOLATE

PROPHYLAXIS WITH COBALAMIN
Infants on specialized diets*
Premature infants
Infants of mothers with pernicious anemia*
Infants and children of mothers with nutritional cobalamin deficiency
Vegetarianism and poverty-imposed near-vegetarianism*
Total gastrectomy†
PROPHYLAXIS WITH FOLIC ACID‡
All women contemplating pregnancy (at least 400 μg/day)
Pregnancy and lactation, premature infants
Mothers at risk for delivery of infants with neural tube defects
Hemolytic anemias/hyperproliferative hematologic states
Patients with rheumatoid arthritis or psoriasis receiving therapy with methotrexate§
Patients receiving antiepileptic drugs
Patients with ulcerative colitis

*For vegetarians, prophylaxis with cobalamin (5- to 10-μg tablet/day) orally should suffice. In all other conditions of cobalamin malabsorption, cobalamin tablets of 1000 μg/day should be administered to meet daily needs.
†Consider late development of cobalamin deficiency and iron malabsorption (prophylaxis with oral cobalamin and iron).
‡Ensure that the patient does not have a cobalamin deficiency before initiation of long-term folate prophylaxis.
§To reduce toxicity of the antifolate.

A10. House AA, Eliasziw M, Cattran DC, et al. Effect of B-vitamin therapy on progression of diabetic nephropathy: a randomized controlled trial. *JAMA*. 2010;303:1603-1609.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Pawlak R, Lester SE, Babatunde T. The prevalence of cobalamin deficiency among vegetarians assessed by serum vitamin B12: a review of literature. *Eur J Clin Nutr*. 2014;68:541-548.
2. Stabler SP. Clinical practice. Vitamin B_{12} deficiency. *N Engl J Med*. 2013;368:149-160.
3. Torheim LE, Ferguson EL, Penrose K, et al. Women in resource-poor settings are at risk of inadequate intakes of multiple micronutrients. *J Nutr*. 2010;140:2051S-2058S.
4. Beulens JW, Hart HE, Kuijs R, et al. Influence of duration and dose of metformin on cobalamin deficiency in type 2 diabetes patients using metformin. *Acta Diabetol*. 2015;52:47-53.
5. Zhao R, Diop-Bove N, Visentin M, et al. Mechanisms of membrane transport of folates into cells and across epithelia. *Annu Rev Nutr*. 2011;31:177-201.
6. Steenweg-de Graaff J, Roza SJ, Steegers EA, et al. Maternal folate status in early pregnancy and child emotional and behavioral problems: the Generation R Study. *Am J Clin Nutr*. 2012;95:1413-1421.
7. Suren P, Roth C, Bresnahan M, et al. Association between maternal use of folic acid supplements and risk of autism spectrum disorders in children. *JAMA*. 2013;309:570-577.
8. Strand TA, Taneja S, Ueland PM, et al. Cobalamin and folate status predicts mental development scores in North Indian children 12-18 mo of age. *Am J Clin Nutr*. 2013;97:310-317.
9. Devalia V, Hamilton MS, Molloy AM. Guidelines for the diagnosis and treatment of cobalamin and folate disorders. *Br J Haematol*. 2014;166:496-513.
10. Carmel R, Agrawal YP. Failures of cobalamin assays in pernicious anemia. *N Engl J Med*. 2012; 367:385-386.
11. Carmel R. Subclinical cobalamin deficiency. *Curr Opin Gastroenterol*. 2012;28:151-158.
12. Odewole OA, Williamson RS, Zakai NA, et al. Near-elimination of folate-deficiency anemia by mandatory folic acid fortification in older US adults: Reasons for Geographic and Racial Differences in Stroke study 2003-2007. *Am J Clin Nutr*. 2013;98:1042-1047.
13. Nguyen CT, Gracely EJ, Lee BK. Serum folate but not vitamin B-12 concentrations are positively associated with cognitive test scores in children aged 6-16 years. *J Nutr*. 2013;143:500-504.
14. Crider KS, Bailey LB, Berry RJ. Folic acid food fortification—its history, effect, concerns, and future directions. *Nutrients*. 2011;3:370-384.
15. Samuel TM, Duggan C, Thomas T, et al. Vitamin B_{12} intake and status in early pregnancy among urban South Indian women. *Ann Nutr Metab*. 2013;62:113-122.

REVIEW QUESTIONS

1. A 42-year-old Asian woman presents with an 8-month history of progressive lethargy and fatigue, anhedonia, forgetfulness, menorrhagia, abdominal bloating and occasional diarrhea, paresthesias, and staggering in the dark. Physical examination reveals anemia with lateral tongue ulcers, mild vitiligo, thyromegaly, delayed ankle reflexes, and reduced vibration sense and proprioception. The hemoglobin level is 5 g/dL, the white blood cell count is 3000/mm^3, and the platelet count is 130,000/mm^3. The peripheral smear reveals a dimorphic picture with macro-ovalocytes and hypochromic microcytic red cells with marked anisocytosis and occasional teardrop cells. Several five-lobed and one six-lobed polymorphonuclear neutrophils are noted. Serum ferritin level is 5 ng/mL, serum cobalamin level is 800 pg/mL, and serum folate level is 5 ng/mL (lower limit of normal, <2.4 ng/mL). The thyroid-stimulating hormone concentration is 15 mIU/L. Which of the following is an *incorrect* statement?

A. Vegetarianism, pernicious anemia with or without hypothyroidism, celiac disease, and pregnancy in women not taking supplemental iron or folate can give a similar picture of dimorphic red cells on a peripheral blood smear.
B. Even without the Schilling test, it is possible to make the diagnosis of pernicious anemia with a noninvasive blood test.
C. Antiparietal cell antibodies, which are diagnostic for pernicious anemia, can interfere with the current serum tests for cobalamin deficiency.
D. The possibility of folate deficiency cannot be ruled out, given the current serum folate level.
E. Treatment with a combination of iron and ascorbic acid plus L-thyroxine alone will not resolve this patient's anemia because it is likely she has an added dietary deficiency of cobalamin.

Answer: C Anti–intrinsic factor antibodies can interfere with the cobalamin-binding luminescence assays and lead to a failure in diagnosis of pernicious anemia in those with classic symptoms and sign (as in this patient). By contrast, antiparietal cell antibodies have no influence on this test. The identification of serum anti–intrinsic factor antibodies is highly specific and diagnostic for pernicious anemia; by contrast, antiparietal antibodies are only about 50% specific for pernicious anemia. The presence of cobalamin deficiency can falsely raise the serum folate level, which could drop below normal on replenishment of cobalamin. This patient probably has a cobalamin deficiency due to either vegetarianism (and even near-vegetarianism) or a combination of pernicious anemia with hypothyroidism and vitiligo (commonly associated autoimmune diseases); in addition to iron, replacement of cobalamin will be needed for a full hematologic response. A dimorphic picture due to iron deficiency and either folate or cobalamin deficiency is found in all the conditions listed in A.

2. An 18-year-old unemployed white woman is to be prescribed generic phenytoin for epilepsy. She is asymptomatic except for passing clots during her periods and "memory problems," and she has a hemoglobin level of 11 g/dL. She used to eat a balanced diet but has been on a vegetarian "low-carb diet" for the past 2 years ever since she learned two things from her favorite Hollywood actress: (1) her idol watches her weight by consuming small servings of salads and other organic-natural foods, and (2) she recently spoke out on a daytime television show about "our wanton exploitation of animals as a food source." The patient has lost 10 pounds in the past 8 months. She has had occasionally irregular periods and is concerned about pregnancy because her boyfriend does not like condoms. Which of the following statements is *not* true?

A. This patient's anemia is likely to be due to a combined deficiency of iron, cobalamin, and folate.
B. This patient is at risk for having a baby with neural tube defects because her diet is likely to be low in folate, and phenytoin can induce a folate deficiency.
C. Taking folate before and during administration of phenytoin will reduce the possibility that she will develop gingival hyperplasia.
D. The patient should be taking supplements of iron, cobalamin, and folate because her diet is likely to be low in these nutrients.
E. At this time, the patient's memory problems can be attributed to iron deficiency but not to either folate deficiency or cobalamin deficiency.

Answer: A Because she switched from a balanced diet to a vegetarian diet only 2 years ago, she is not at risk for dietary cobalamin deficiency; however, with prolonged vegetarianism for 5 years, she will need to be taking cobalamin supplements. Vegetarian diets are intrinsically low in bioavailable iron, but earlier she used to eat a balanced diet. Because carbohydrate (flour that makes bread and pasta) is the food source that is fortified with folic acid, her low-carb diet could be low in folate, even though she consumes "small servings of salads." Because she does not take folate supplements, the impending initiation of phenytoin, which interferes with folate absorption, places her at risk for having a baby with a neural tube defect. Recent studies show that taking folate before and during administration of phenytoin will reduce the possibility that she will develop gingival hyperplasia. Although folate deficiency does not cause neuropsychiatric problems in adults, iron deficiency is notorious for cognitive problems in young women with minimal (iron deficiency–induced) anemia.

3. A 29-year-old Kenyan postgraduate year 2 medical resident at a major university hospital in the Midwest with known thalassemia minor is referred to you for advice. She was admitted a month ago with high fever and chills with drenching sweats. She had recently returned from a trip home, where she also got married. She mentioned that she had been bitten several times by mosquitoes during her honeymoon. Her dietary history is significant only for the fact that she cooks all her meals and subsists "mostly on *ugali* (organic cornmeal without additives) and kale or other sautéed vegetables, occasional (once per week) small servings of meat, and lots of tea." In the hospital, a diagnosis of falciparum malaria was made on the basis of finding parasites on a peripheral smear. The accompanying laboratory results showed a hemoglobin level of 6 g/dL, mean corpuscular volume (MCV) of 90, and corrected reticulocyte count of 15%; lactate dehydrogenase was high, and haptoglobin was low. The serum folate level was well within the normal range (12 ng/mL; normal, >2 ng/mL), the cobalamin level was borderline low (202 pg/mL), and the serum ferritin level was high (300 ng/mL). Although antimalarials led to an eradication of malaria, the patient has still not significantly improved her hemoglobin level above 9 g/dL 2 months later. All of the following are true *except* which one?

A. The patient is at risk for deficiencies of folate, cobalamin, and iron because her native Kenyan diet is monotonous and relatively low in both animal-source foods and fresh produce.
B. The patient could have had a Coombs-positive as well as a Coombs-negative hemolytic anemia associated with falciparum malaria and should be prescribed both antimalarials and micronutrients to support compensatory hematopoiesis.
C. At presentation in the hospital, the folate value rules out the possibility of folate deficiency and argues against administration of folic acid.
D. If the patient is taking long-term folic acid in planning for a pregnancy, the development of neurologic symptoms should immediately trigger concerns of a cobalamin deficiency.
E. She should now be treated with iron, cobalamin, and folate to effect a complete resolution of the anemia and then be maintained with these supplements to optimize her nutrition.

Answer: C Red cells normally contain 30-fold more folate than what is present in the serum. Therefore, hemolysis associated with malaria will release the 30-fold excess folate from red cells into the circulation and result in a falsely high serum folate value. Hence, the serum folate level is a poor test to use in this setting because it is falsely elevated in both malaria and cobalamin deficiency. In addition, the patient is likely to have a diet that is low in all three nutrients (folate, cobalamin, and iron); food frequency questionnaire–derived estimates from resource-limited countries (including Kenya) indicate that most women consume less than what is optimum. Near-vegetarian and vegetarian diets predispose to cobalamin and iron deficiency. The borderline nutritional cobalamin deficiency can also contribute to a falsely elevated folate level. Because the cornmeal she purchases has no additives (i.e., no folic acid), she will need to take folate supplements in preparation for pregnancy.

4. A 29-year-old white woman is referred because of easy fatigue and anorexia. During the past 5 years, she has noticed abdominal bloating and diarrhea that are closely linked to consumption of gluten-containing foods. There is pallor, and the hemoglobin level is 9 with an MCV of 90 and a serum ferritin level of 10 ng/mL. The serum anti–tissue transglutaminase antibody test result is positive. All of the following are true *except* which one?

A. Treatment of the patient with total-dose parenteral iron will resolve the iron deficiency, but the anemia will likely persist.
B. Treatment with iron alone will result in a progressive increase in MCV, which reflects unmasking of a megaloblastic anemia.
C. The serum homocysteine and methylmalonic acid levels are likely to be increased and signify metabolic evidence of cobalamin deficiency.
D. Folate deficiency is unlikely to be present in a patient with elevated methylmalonic acid and homocysteine from cobalamin deficiency.
E. Folates are absorbed through the proton-coupled folate transporter in the jejunum, which can be involved in celiac disease.

Answer: D Because this patient has celiac disease, in the short term, folate deficiency can develop; but in the long term, with more extensive intestinal involvement that includes the ileum, cobalamin deficiency can eventually develop. The existence of high methylmalonic acid and homocysteine values, although confirming cobalamin deficiency, does not rule out an associated folate deficiency, which is associated with high homocysteine values. Despite low ferritin values consistent with iron deficiency from celiac disease, the normal MCV probably reflects an underlying folate or cobalamin deficiency. Treatment with parenteral iron will unmask the hematologic features of megaloblastic anemia.

5. A 58-year-old white man with type 2 diabetes for 15 years is evaluated for a 6- to 8-month history of slowly progressive fatigue, anemia, and mild peripheral neuropathy of both lower extremities. He had been well controlled with metformin 850 mg three times daily and only recently began taking a single dose of long-acting insulin at night because the latest hemoglobin A_{1c} level was 10. He does not take supplements, and his diet has been well balanced. He regularly makes his own bread every 2 or 3 days. He has recently been prescribed proton pump inhibitors and an H_2-blocker for dyspeptic symptoms. The hemoglobin level is 11 with an MCV of 108, with macro-ovalocytes and several five-lobed hypersegmented polymorphonuclear neutrophils seen on the peripheral blood smear. All of the following are true *except* which one?

A. The patient has features of megaloblastic anemia and should have a bone marrow examination to confirm the diagnosis.
B. The serum cobalamin level is likely to be low, but the serum folate level is likely to be in the high-normal range.
C. This patient could have food cobalamin malabsorption because of inhibition of acid in the stomach.
D. This patient could have cobalamin malabsorption at the level of cubam receptors in the ileum.
E. Upper endoscopy is indicated to look for gastrinomas or gastric cancer if anti–intrinsic factor antibodies are found.

Answer: A In view of the slowly progressive nature of the anemia, the characteristic findings of a megaloblastic anemia on the peripheral smear, and the known effect of metformin in interfering with trans-enterocytic transport of cobalamin (at the ileal level), there is no reason to proceed with a bone marrow examination. The serum cobalamin level is likely to be low after consumption of metformin for 5 years (this patient has been taking metformin for much longer—15 years); the serum folate level is likely to be high-normal among those consuming a balanced diet because of food fortification with folic acid (remember, he makes his own bread). The malabsorption of cobalamin in the stomach induced by a reduction of acid can, during several years, lead to symptoms of deficiency. Proton pump inhibitors and H_2-blockers can induce malabsorption. However, the primary mechanism of metformin is believed to be at the level of cubam receptors in the ileum, where supplemental calcium (1.2 g/day) taken with the drug can prevent this well-recognized, potentially serious but reversible, delayed adverse effect of metformin.

165

APLASTIC ANEMIA AND RELATED BONE MARROW FAILURE STATES

GROVER C. BAGBY

DEFINITION

Aplastic anemia is a life-threatening syndrome characterized by failure of the bone marrow to produce peripheral blood cells and their progenitors. Diverse diseases and environmental factors can cause this syndrome, but its hallmark is bone marrow hypocellularity and hypoplasia of the erythroid, myeloid, and megakaryocyte lines (Fig. 165-1).

EPIDEMIOLOGY

The annual incidence is 2 per 1 million persons in Europe and North America and 4 to 7 per 1 million persons in Asia. No age group is exempt, and although the syndrome occurs most often in young adults, the age distribution of newly diagnosed patients is bimodal, with peaks at 15 to 25 years and at 60 to 65 years.

PATHOBIOLOGY

Pathology

Peripheral blood pancytopenia is universally present in patients with aplastic anemia, but because other disorders can cause pancytopenia, bone marrow biopsy is required to establish the diagnosis. The diagnosis will be clear-cut if the biopsy specimen is of sufficient size and has been obtained from an anatomic site that has never been exposed to extensive trauma or irradiation. As shown in Figure 165-1B, some residual lymphoid cell populations can be found in marrow specimens. Although these lymphoid cells may be of pathophysiologic importance, it is the absence of *nonlymphoid* hematopoietic cells that is important in establishing the diagnosis of this syndrome. However, if the hematopoietic marrow has been suppressed (or "replaced") by the infiltration of neoplastic cells or fibroblasts, the diagnosis of aplastic anemia cannot be made. Therefore, the diagnosis requires not only a dearth of hematopoietic cells in the marrow but also an "empty" bone marrow.

Some bone marrow failure syndromes affect only one lineage. In those cases, only the marrow precursors of that lineage are missing. In patients with agranulocytosis, for example, there are rare neutrophils and neutrophil precursors present, and the ratio of erythroid to myeloid cells is very high. Likewise, in patients with the disorder known as pure red cell aplasia, few erythroid cells are detectable in the marrow, but the other lineages are well represented and functional. These two disorders are examples of bone marrow failure syndromes but are not examples of aplastic anemia, which involves global suppression of all hematopoietic lineages.

Pathophysiology

Marrow aplasia in a few patients (10 to 15%) can be attributed to an inherited bone marrow failure syndrome,[1] but most cases are acquired. In all cases, it is clear that the causative factors, genetic or environmental, injure pluripotent hematopoietic stem cells. This is in contrast to the case of the lineage-restricted disorders, wherein the causative agents and factors suppress the growth and development of unipotent progenitor cells committed to that particular lineage. Radiation, viral diseases, cytotoxic drugs, and chemicals are known causes of aplastic anemia, but the most common form of acquired aplastic anemia is immunologically mediated, and evidence is emerging that some marrow failure states attributed to viral infection or to idiosyncratic drug reactions may also result from immune suppression of hematopoiesis. The pathogenesis of the major causes of aplastic anemia and related bone marrow failure states is outlined in Table 165-1.

Pathogenesis

Autoimmune Aplastic Anemia

Acquired Aplastic Anemia

In patients with the most common form of acquired aplastic anemia, autologous T lymphocytes suppress the replicative activity and induce the death of hematopoietic stem and progenitor cells. Evidence supporting this model is found in studies demonstrating the following: removal of T lymphocytes from cultured bone marrow cells enhances hematopoiesis in vitro; patients with aplastic anemia can be effectively treated with immunosuppressive therapy alone[2]; oligoclonal T cells in the marrow and blood of aplastic anemia

TABLE 165-1 MAJOR CAUSES OF APLASTIC ANEMIA AND RELATED BONE MARROW FAILURE STATES

Autoimmune aplastic anemia
Acquired
Drug induced
Infections
Hepatitis
Epstein-Barr virus
Autoimmune-mediated failure of single hematopoietic lineages
Agranulocytosis
Pure red cell aplasia
Direct stem cell toxicity
Radiation
Chemicals
Drugs
Other aplastic states
Pregnancy
Paroxysmal nocturnal hemoglobinuria
Inherited bone marrow failure syndromes

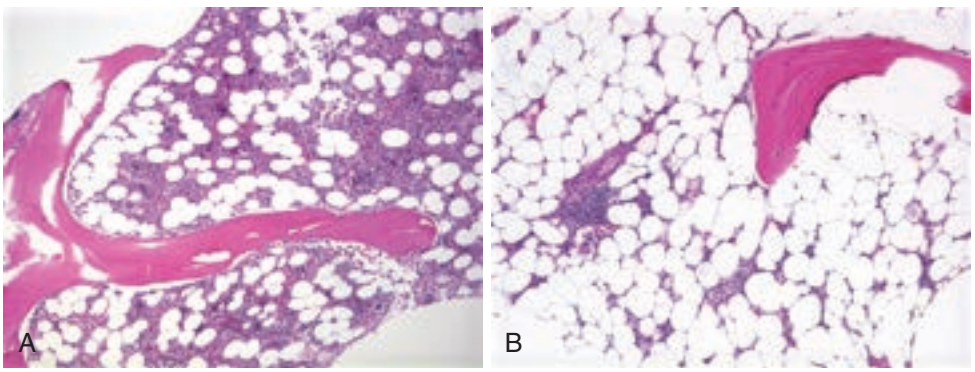

FIGURE 165-1. Two bone marrow biopsy samples from different patients. A, This specimen, from a normal individual, shows an abundance of hematopoietic cells, including myeloid and erythroid precursors and normal-appearing megakaryocytes. **B,** A specimen from a patient with severe aplastic anemia shows few detectable hematopoietic cells, and those that can be seen (one small nest) are largely lymphocytes. Lymphoid cells are often detectable and probably play an important pathophysiologic role in many cases of acquired idiopathic aplastic anemia, a disorder that is most often immunologically mediated. (Courtesy Dr. Ken Gatter, Oregon Health and Science University.)

patients contain high intracellular levels of the myelosuppressive cytokines interferon-γ and tumor necrosis factor-α; the interferon-γ/tumor necrosis factor-α–positive T-cell populations are suppressed in patients who respond to immunosuppressive therapy but are not suppressed in patients who do not respond; and the syndrome can be modeled in mice by infusing alloreactive lymphocytes that induce marrow failure. The mechanisms by which such autoinhibitory T-cell clones arise are unclear, but a loss of function in the regulatory T-cell population may play a role.[3]

Drug Induced

Although a wide variety of drugs have been associated with aplastic anemia, the association is loose. Much of the evidence is circumstantial, and apart from drugs that are known to be directly toxic to the marrow (e.g., chemotherapeutic agents; Table 165-2), the cases are not related to total dose of the suspect agent. These "idiosyncratic" reactions are likely to be autoimmune. Some agents, chloramphenicol being the classic example, are capable of inducing both types of injury. High doses can lead to myelosuppression in all treated patients (which abates after discontinuation of the drug), but even low doses can cause rare idiosyncratic aplastic responses as well (which do not remit after discontinuation of the agent). Drugs that have been repeatedly associated with aplastic anemia are listed in Table 165-2.

Infections

HEPATITIS. About 2 to 5% of patients with severe aplastic anemia have had viral hepatitis (Chapters 148 and 149). Some cases were associated with hepatitis A or B, but most patients with the hepatitis-aplasia syndrome have had hepatitis of unclear type (non-A, -B, -C, -E, or -G). Most patients with this syndrome are younger than 20 years. The natural course is rapid, with a 1-year mortality rate of more than 90%. The immune system is probably involved in the pathophysiologic mechanism of this syndrome because T-cell clonotypes are shared among patients with this type of aplasia[4] and immunosuppressive therapy has been reported to induce meaningful remissions.

EPSTEIN-BARR VIRUS. In rare patients with aplastic anemia, evidence of active Epstein-Barr virus (EBV) infection has been discovered (Chapter 377). Because the virus does not infect progenitor cells or stem cells, it is most likely that EBV induces an aberrant immune response that generates either immunoglobulin- or T-lymphocyte–mediated hematopoietic suppression. Because only a minority of EBV-infected aplastic patients describe a history of typical infectious mononucleosis, it is equally likely that the aplastic state came first and that EBV infection or reactivation was a second event.

TABLE 165-2 DRUGS AND TOXINS ASSOCIATED WITH APLASTIC ANEMIA

DOSE DEPENDENT

Antineoplastic Agents

Antimetabolites: fluorouracil, mercaptopurine, 6-thioguanine, methotrexate, cytosine arabinoside, gemcitabine, fludarabine, cladribine, pentostatin, hydroxyurea

Alkylating and cross-linking agents: busulfan, cyclophosphamide, chlorambucil, nitrogen mustard, melphalan, cisplatin, carboplatin, ifosfamide, nitrosoureas (BCNU and CCNU), mitomycin C

Cytotoxic antibiotics: daunorubicin, doxorubicin, mitoxantrone

Plant alkaloids: vinblastine, paclitaxel

Topoisomerase inhibitors: etoposide

Antimicrobial Agents

Chloramphenicol, dapsone, fluorocytosine

Anti-inflammatory and Antirheumatic Agents

Colchicine

Insecticides

Chlordane, chlorophenothane (DDT), lindane, parathion

Other Chemicals

Benzene

Benzene-containing chemicals: kerosene, chlorophenols, carbon tetrachloride

DOSE INDEPENDENT

Idiosyncratic, likely immune mediated

(*Note:* Most agents on this list should be considered to be possibly associated with aplastic anemia.)

Antimicrobial Agents

Chloramphenicol, dapsone, sulfonamides, tetracycline, methicillin, amphotericin, quinacrine, chloroquine, pyrimethamine

Anticonvulsants

Hydantoins, carbamazepine, phenacemide, primidone, ethosuximide

Anti-inflammatory Agents

Phenylbutazone, indomethacin, ibuprofen, oxyphenbutazone, sulindac, naproxen

Antiarrhythmic Drugs

Quinidine, tocainide, procainamide

Metals

Gold, arsenic, mercury, bismuth

Antihistamines

Cimetidine, ranitidine, chlorpheniramine, pyrilamine, tripelennamine

Diuretics

Acetazolamide, furosemide, chlorothiazide, methazolamide

Hypoglycemic Agents

Chlorpropamide, tolbutamide

Antithyroid Drugs

Propylthiouracil, potassium perchlorate, methylthiouracil, methimazole, carbimazole

Antihypertensive Agents

Methyldopa, enalapril, captopril

Sedatives

Chlordiazepoxide, chlorpromazine, meprobamate, prochlorperazine

Autoimmune-Mediated Failure of Single Hematopoietic Lineages

Agranulocytosis

Agranulocytosis is characterized by severe neutropenia and suppression of granulopoiesis (also see Chapter 167). This disorder can be an idiosyncratic reaction to certain drugs and most likely involves immune suppression of granulopoietic progenitor cells. The disease almost always abates when the offending drug is discontinued. Agranulocytosis also occurs in patients with established autoimmune diseases, including systemic lupus erythematosus, Sjögren syndrome, and rheumatoid arthritis. In some cases, the disorder is caused by myelosuppressive antibodies; in others, it is caused by T lymphocytes that suppress granulopoiesis. Immunosuppressive therapy is often effective in such patients and should be used in patients whose agranulocytosis is severe and associated with recurrent infections.

Pure Red Cell Aplasia

Severe normochromic, normocytic anemia (Chapter 158) with a marked decrease in reticulocyte number is sometimes associated with selective hypoplasia of the erythroid marrow without loss of megakaryocytes and myeloid precursor cells. In immunocompromised hosts and patients with chronic hemolytic diseases, this disorder, known as pure red cell aplasia, can be caused by parvovirus B19 infection (Chapter 371), an agent that infects erythroid precursor cells and likely generates an erythroid suppressive immune response. This disease can also be mediated by T lymphocytes or natural killer cells that suppress cells of the erythroid lineage and in more uncommon cases by antibody-dependent suppression of erythropoiesis. Pure red cell aplasia can also develop as a complication of thymoma (Chapter 99) and in these circumstances is likewise caused by oligoclonal T-cell expansion that specifically suppresses erythroid progenitor cells. This disorder can also be associated with drug exposure (e.g., isoniazid, chlorpropamide, and phenytoin), lymphoid neoplasms (chronic lymphocytic leukemia; Chapter 184), and myelodysplasia (Chapter 182). Rarely, adults with Diamond-Blackfan anemia present with isolated erythroid suppression, but the degree to which the erythroid marrow is suppressed in such patients rarely matches the profound suppression seen in acquired cases of immune-mediated pure red cell aplasia.

Direct Stem Cell Toxicity

Radiation

The severity of myelosuppression induced by radiation and the degree to which the marrow can recover from that injury depend on the radiation dose, the timing of exposure, and the fraction of hematopoietic tissues exposed. Low-dose total body radiation causes transient marrow suppression. High doses of total body radiation (700 to 1000 cGy) induce severe injury to the stem cell pool with persistent and life-threatening marrow failure. In the past, direct radiation injury to hematopoietic stem cells was the accepted explanation for stem cell injury, but recent evidence indicates that injury to other tissues (especially the gut) results in the release of endogenous factors that themselves suppress hematopoiesis.[5] When limited bone marrow sites are irradiated to very high doses (4000 cGy or more), the relatively radioresistant bone marrow stromal cells are eradicated, and thereafter that marrow space can never fully support hematopoietic activity.

Chemicals

Benzene suppresses the bone marrow in a dose-dependent manner, and chronic exposure to it has been linked with aplastic anemia and myeloid leukemogenesis. Benzene and many of its catabolites are directly toxic to stem

cells, damage DNA, suppress the supportive function of the bone marrow microenvironment, and accentuate the responsiveness of hematopoietic progenitor cells to intramedullary apoptotic cues that arise during the inflammatory response. Kerosene, carbon tetrachloride, and chlorophenols contain benzene, as do many other like products used for paint stripping, refinishing, and degreasing (see Table 165-2).

Drugs

Many agents in use for the treatment of malignant diseases are predictably myelosuppressive and can induce aplastic anemia because they are directly toxic to stem and progenitor cells in the marrow (see Table 165-2). These myelosuppressive responses are completely predictable and dose dependent. In practical terms, unless the patient receives a drug overdose or has an undiagnosed genetic disorder that predisposes the patient to respond to the agent in an exaggerated way (e.g., Fanconi anemia [FA]), most patients treated for neoplastic diseases develop reversible bone marrow aplasia or hypoplasia and recover bone marrow function within a matter of days.

Other Aplastic States

Pregnancy

Aplastic anemia can be diagnosed in pregnancy. In addition, some patients with aplastic anemia have become pregnant after the diagnosis. The prognosis in both instances is poor, and most fatal outcomes are due to bleeding complications. Some women who have been fortunate enough to recover bone marrow function post partum develop aplastic anemia with a subsequent pregnancy. The pathogenesis and causal relationship between pregnancy and aplastic anemia remain unknown but may reflect a high level of immunologic activation during pregnancy.

Paroxysmal Nocturnal Hemoglobinuria

Paroxysmal nocturnal hemoglobinuria (PNH) is an acquired disorder that results from the expansion of a clone of hematopoietic stem cells whose progeny are incapable of anchoring essential proteins to their membranes (Chapter 160).[6] The defect is caused by an inactivating somatic mutation of *PIGA*, an X-linked gene that encodes a protein essential for synthesis of the membrane anchor glycosyl phosphatidylinositol (GPI). Some of the GPI-anchored proteins (e.g., CD55 and CD59) are important in normal red cells to protect them from activated complement. Consequently, the loss of CD55 and CD59 results in chronic intravascular hemolysis. Two less clearly defined features of PNH are high relative risks of thromboembolism (Chapter 176) and aplastic anemia. The emergence of "PNH clones" is not uncommon in patients with acquired aplastic anemia, and experimental evidence supports the idea that the evolution of *PIGA*-deficient stem cells is an adaptive response to the immune attack. How the *PIGA*-deficient cells fend off cytotoxic T cells is less clear, but it is likely that their pool expands because they accomplish that task somehow. The most reasonable model is one in which, in the setting of an immune attack, *PIGA*-deficient stem cells are more fit than normal stem cells, but the complement-sensitive red cells to which they give rise as a result of this "tradeoff" have an unusually short lifespan.

Inherited Bone Marrow Failure Syndromes

Some inherited bone marrow failure syndromes can be manifested in adolescence and adulthood. They include dyskeratosis congenita, FA, and Diamond-Blackfan anemia. The molecular pathogenesis of the bone marrow failure seen in these diseases is under active investigation.[7] Although the genetic basis of these disorders has been well defined in the past decade, the canonical functions of the proteins encoded by these genes is distinct (Table 165-3) for each disease. For example, in dyskeratosis, the mutations occur in genes that encode proteins and RNAs involved in telomere maintenance. In FA, the mutations involve proteins involved in the DNA damage response, and in Diamond-Blackfan anemia, inactivating mutations involve ribosomal proteins known to be involved in ribosome biogenesis.[8] Whether loss of these canonical functions is linked with the pathogenesis of marrow failure is

TABLE 165-3 DISTINGUISHING CLINICAL FEATURES OF THE INHERITED BONE MARROW FAILURE SYNDROMES THAT MAY BE INITIALLY DIAGNOSED IN ADULTHOOD

DISTINGUISHING FEATURES	*DISEASES*		
	FANCONI ANEMIA	**DYSKERATOSIS CONGENITA**	**DIAMOND-BLACKFAN ANEMIA**
History	Skeletal and renal malformations, low birthweight, pancytopenia, family member with bone marrow failure, myelodysplasia, acute myelogenous leukemia, or squamous cell carcinoma at an early age Family member with Fanconi anemia	Intrauterine growth retardation, developmental delay, and short stature Family history of myelodysplasia, acute myelogenous leukemia, marrow failure, abnormal fingernails or toenails, leukoplakia, head and neck cancer, or pulmonary fibrosis	Low birthweight, arm and thumb deformities at birth
Physical findings	Thumb and radial malformations, hyperpigmented skin lesions (cafe au lait spots), short stature, myelodysplasia, acute myelogenous leukemia, squamous cell carcinoma at young age, renal and cardiac malformations, microcephaly, hypogonadism	Lacy reticular pigmentation of skin, dystrophic fingernails and toenails, premature graying of hair, hair loss, short stature, oral leukoplakia, squamous cell cancer of head and neck, pulmonary fibrosis, osteopenia, hypogonadism	Triphalangeal thumbs, short stature, arm anomalies
Genes inactivated	*FANCA, FANCB, FANCC, FANCD1* (also known as *BRCA2*), *FANCD2, FANCE, FANCF, FANCG* (also known as *BRCA2, FANCD2, FANCE, FANCF XRCC9*), *FANCI, FANCJ* (also known as *BRCA2, FANCD2, FANCE, FANCF BACH1,* and *BRIP1*), *FANCL* (also known as *PHF9* and *POG*), *FANCM* (also known as *Hef*), *FANCN* (also known as *PALB2*), *FANCO* (also known as *RAD51C*), and *FANCP* (also known as *BRCA2, FANCD2, FANCE, FANCF SLX4*) These genes encode proteins known to protect the genome from excessive damage induced by chemical cross-linking agents. These genes account for most cases of Fanconi anemia.	*DKC1, TERC, TERT, TINF2, NOLA2* (also known as *NHP2*) and *NOLA3* (also known as *NOP10*), and *WRAP53* These genes encode proteins known to participate in maintenance of telomeres. They account for only half of dyskeratosis cases, so there are additional genes to be discovered.	*RPS7, RPS10, RPS17, RPS19, RPS24, RPS26, RPL5, RPL11, RPL26,* and *RPL35A* These genes encode ribosomal proteins. They account for only half of cases, so there are additional genes to be discovered.
Screening and diagnostic tests	1. Chromosomal breakage test on skin fibroblasts or peripheral blood lymphocytes (in response to mitomycin C or diepoxybutane) 2. Complementation analysis (flow cytometric analysis of G_2 arrest in melphalan-exposed cells after transduction with retroviral vectors expressing normal Fanconi anemia genes) 3. Gene sequencing	1. Quantitative analysis of telomere length (flow-FISH) in lymphocytes 2. Gene sequencing	*Note:* Isolated erythroid failure is more common than full-blown aplastic anemia. 1. There are no screening tests, although serum ADA is often elevated. 2. Gene sequencing

ADA = adenosine deaminase; FISH = fluorescent in situ hybridization.

unclear. In fact, studies on hematopoietic cells have revealed noncanonical functions of some of these proteins. Some of the FA proteins, for example, participate directly or indirectly in stem cell survival signaling pathways. Interestingly, some of the pathways disrupted in mutant cells result in hyperactivation of precisely those same cytokine signaling pathways involved in the pathogenesis of acquired autoimmune aplastic anemia.[9]

Genetics

The genetic basis of inherited bone marrow failure syndromes is being rapidly solved. Some of these syndromes (e.g., Shwachman-Diamond syndrome, amegakaryocytic thrombocytopenia, and severe congenital neutropenia) are almost always diagnosed in early life. However, some may be first diagnosed in adulthood (e.g., dyskeratosis congenita, FA, and Diamond-Blackfan anemia). It is critically important to consider these three disorders early in the evaluation of adults with aplastic anemia because the treatment of such patients with conventional stem cell transplantation regimens is associated with high mortality rates (especially in dyskeratosis congenita and FA). Furthermore, current immunosuppressive therapy, an important therapeutic option in patients with acquired aplastic anemia, plays no role in these diseases. The clinical and laboratory manifestations of these three diseases and the findings that should prompt genetic testing in such patients are reviewed in Table 165-3. Whereas the stem cell defects can be identified during fetal development, adults who present with these diseases were not born with aplastic anemia. Instead, aplasia develops over time and can give rise to symptoms in adulthood. Prospective studies of children and adults who present with bone marrow failure have indicated that nearly 10% will have previously unsuspected FA. Diagnostic consideration of a hereditary form of aplastic anemia should not be limited to children.

CLINICAL MANIFESTATIONS

The natural course of aplastic anemia is influenced by its severity. Patients with hypoplastic bone marrows have severe aplastic anemia if they meet two of the following laboratory criteria: absolute neutrophil count of less than 0.5×10^9/L; platelet count of less than 20×10^9/L; or reticulocyte count of less than 20×10^9/L. Patients who do not have severe aplastic anemia often progress to severe aplasia, but the pace is slow (about 40% will have progressed in 5 years). Severe aplastic anemia is a life-threatening condition that, untreated, is associated with a mortality rate of 80% in the 24 months after diagnosis. Treatment of any cohort of patients with severe aplastic anemia will prolong life, and many patients, particularly those who have received stem cell transplants, will be cured.

History

Symptoms of this syndrome, cause notwithstanding, are almost always reflective of low blood counts. The most common presenting symptoms are those associated with thrombocytopenia and anemia. Low platelet counts are associated with bleeding, often epistaxis and bleeding gums, bruising with minor or no trauma, and menorrhagia. Anemia accounts for the nearly universal symptoms of fatigue and dyspnea on mild exertion. Some aplastic patients may present with intercurrent bacterial or fungal infection (because of severe neutropenia), but these cases are less common. Family histories that include any of the features listed in Table 165-3 should raise suspicion of an inherited bone marrow failure syndrome.

Physical Examination

Pallor and tachycardia at rest are common signs of anemia but can be absent or unnoticeable in younger patients and in patients whose aplasia is of recent onset. Hemorrhagic manifestations classic for thrombocytopenia are often found: petechiae (cutaneous or palatal), ecchymoses, and epistaxis. The most common form of aplastic anemia, autoimmune, is rarely associated with lymphadenopathy or hepatosplenomegaly, and when such findings are present, alternative diagnoses should be considered and painstakingly ruled out. Likewise, concerns of an inherited bone marrow failure disease should be triggered by the following: short stature; endocrinopathies; osteopenia; findings of developmental anomalies of the skin, nails, hands, or arms; and malformations of the heart, liver, or genitourinary tract (see Table 165-3).

Initial Laboratory Findings

Pancytopenia (anemia, leukopenia, and thrombocytopenia) is a universal presenting finding. The morphology of neutrophils, platelets, and red cells on peripheral blood smear is usually normal unless there is concurrent iron deficiency due to bleeding.

DIAGNOSIS

The evaluation of pancytopenic patients first requires examination of the peripheral blood smear. If there are morphologic or clinical signs of vitamin B_{12} or folic acid deficiency (e.g., hypersegmented neutrophils and oval macrocytes), those disorders should be ruled out because a bone marrow aspiration and biopsy would not be required in those conditions. In severe aplastic anemia, the peripheral blood smear will not show nucleated red blood cells or other signs that the marrow might be infiltrated with abnormal cells. All patients in whom vitamin B_{12} and folate deficiency have been ruled out (Chapter 164) require a bone marrow aspiration and biopsy. Obtaining both types of samples is important. The biopsy best assesses overall bone marrow cellularity and provides the most sensitive evidence for some infiltrative processes. The aspirated sample can be examined microscopically for the presence of abnormal cells but also provides cells for cytogenetic analyses (which can provide evidence supporting hypoplastic myelodysplasia and acute leukemia). Rarely in the early stages of aplasia, the biopsy finding can be somewhat cellular. A repeated biopsy in 1 to 2 weeks might be required to establish the diagnosis clearly. As summarized in Figure 165-2, once the diagnosis of aplastic anemia has been made, a series of additional tests must be obtained. In light of the life-threatening nature of this disease, the tests must be obtained simultaneously, but they serve three distinct purposes.

Ruling Out Aplastic Anemia Variants That Must Be Treated Differently

The best therapeutic option for many different aplastic states is often matched sibling donor stem cell transplantation, but there are some aplastic states that are managed differently. For example, a child with an inherited marrow failure syndrome might have a human leukocyte antigen (HLA)–identical sibling who also has the same genetic defect. If such a diagnosis has been overlooked in the recipient and donor, the recipient will most likely die because the donor cells are unfit for transplantation and the conditioning regimen will be too toxic. For example, patients with FA are highly intolerant of radiation and cross-linking agents used in conventional conditioning regimens, and patients with dyskeratosis suffer excessive post-transplantation morbidity and mortality. Patients with dyskeratosis congenita and children with the Shwachman-Diamond syndrome are also intolerant of conventional transplantation regimens and often suffer severe pulmonary and hepatic toxicity. Although patients with Diamond-Blackfan anemia are more tolerant of standard conditioning regimens, they are more apt to respond to glucocorticosteroid therapy, and if they are transplanted with stem cells from an undiagnosed affected sibling, they too will do poorly. Finally, patients with PNH should be identified with flow cytometric quantification of CD55- and CD59-deficient hematopoietic cells because more than half of patients with severe aplastic anemia will have PNH clones, and on treatment with immunosuppressive therapy, the PNH clone often expands, resulting in hemolysis and thrombosis.

Tests Helpful in Supportive Care

At some point during the course of the disease, red cell and platelet transfusions will be necessary. Irradiated and filtered blood products are used to prevent transfusion-associated graft-versus-host disease (GVHD), to reduce alloimmunization, and to reduce the complication of cytomegalovirus (CMV) infection. ABO and HLA typing are required. Infections can be of bacterial, viral, or fungal origin and must be quickly diagnosed and treated not only because the patients are often neutropenic but also because, once definitive therapy begins, the patient will be immunosuppressed. In the acutely infected, severely neutropenic patient, once culture and biopsy specimens are obtained, empirical antibiotic therapy should be given without waiting for the culture results. Post-transplantation CMV infection is best avoided in CMV-seronegative recipients by use of CMV-negative blood products (Chapter 177).

Evaluating the Patient as a Candidate for Stem Cell Transplantation

Timing of treatment depends on the severity of the aplastic anemia and the age of the patient (Table 165-4). Patients with mild marrow hypoplasia and mild bone marrow suppression can be observed closely to determine what the pace of hypoplasia might be. In patients up to 40 years of age with either severe acquired aplastic anemia or transfusion dependence, steps should be taken to evaluate them promptly for stem cell transplantation therapy by

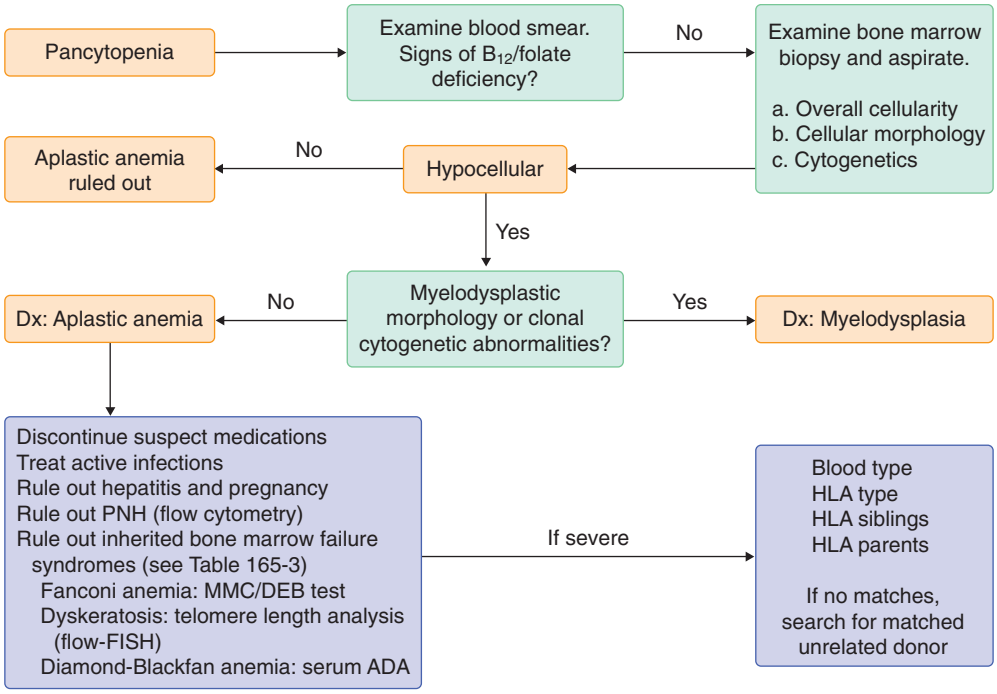

FIGURE 165-2. Diagnostic management of patients with aplastic anemia. Patients who present with pancytopenia will require bone marrow biopsy and aspiration unless there are signs of vitamin B_{12} or folate deficiency. If the bone marrow is as cellular as that shown in Figure 165-1A, the diagnosis of aplastic anemia has been effectively ruled out because the diagnosis requires bone marrow hypocellularity (see Fig. 165-1B). In all patients, regardless of severity, suspect medications should be discontinued, active infections should be treated without delay, and paroxysmal nocturnal hemoglobinuria (PNH) should be excluded, as should pregnancy and hepatitis. Microscopic evidence in the bone marrow sample of myelodysplastic changes and evidence of a clonal chromosomal abnormality indicate that the patient has "hypoplastic" myelodysplasia and should be treated accordingly. For patients with positive family histories or any one of the findings listed in Table 165-3 and for all patients younger than 40 years, inherited marrow failure syndromes should be ruled out with screening tests. Patients with severe aplastic anemia are those who have at least two of the following: absolute neutrophil count of less than 0.5×10^9/L, platelet count of less than 20×10^9/L, or reticulocyte count of less than 20×10^9/L. These patients must be treated with definitive therapy and should be evaluated for stem cell transplantation with human leukocyte antigen (HLA) typing (patient and family members) and, if necessary, a search for matched unrelated donors. ADA = adenosine deaminase; DEB = diepoxybutane; FISH = fluorescent in situ hybridization; MMC = mitomycin C.

TABLE 165-4 TREATMENT RECOMMENDATIONS FOR PATIENTS WITH SEVERE ACQUIRED IDIOPATHIC APLASTIC ANEMIA*

	AGE < 20 YEARS		AGE 20-40 YEARS		AGE > 40 YEARS	
HLA-identical sibling?	Yes	No	Yes	No	Yes	No
First-line treatment	MSBMT	IST	MSBMT	IST	IST	IST
Second-line treatment	IST or second MSBMT	IST or MUDT or CBT	IST	IST, eltrombopag, or MUDT	MSBMT	IST For failures, consider eltrombopag, clinical trials, or MUDT

*This general set of recommendations cannot be applied to patients with inherited bone marrow failure syndromes because they do not respond to immunosuppressive therapy.
CBT = umbilical cord blood transplantation; HLA = human leukocyte antigen; IST = immunosuppressive therapy; MSBMT = matched sibling bone marrow transplant; MUDT = matched unrelated donor transplant.
Modified from Bacigalupo A, Passweg J. Diagnosis and treatment of acquired aplastic anemia. *Hematol Oncol Clin North Am.* 2009;23:159-170.

seeking HLA-identical siblings and, in appropriate cases, searching for matched unrelated donors.

Differential Diagnosis

Most classic aplastic anemia patients have immunologically mediated disease. Notwithstanding strong evidence in support of this mechanism from some research laboratories during the past 30 years, there exists no validated screening tool or certified test that can either rule in or rule out immune-mediated disease, a fact complicated by the evidence that some cases of drug- and virus-induced disease are also immunologically mediated. Therefore, idiopathic autoimmune aplastic anemia remains a diagnosis of exclusion. For this reason, the obligation of the diagnostician is to consider disease induced by chemical or viral agents and radiation and disease associated with pregnancy. It is also most important to rule out PNH and inherited bone marrow failure syndromes (see Fig. 165-2).

Cytotoxic Drugs, Chemicals, and Radiation

A fastidiously obtained medication history is important. Any drug with the potential of inducing aplastic anemia should be discontinued. It is equally important to ask patients about alternative therapies that they might not consider to be "medicines." Some herbal remedies are known to contain molecules (e.g., phenylbutazone) not listed on the label but that are associated with aplasia (see Table 165-2). A history of exposure to radiation or to chemicals with myelosuppressive capacities (see Table 165-2) should likewise be obtained. Tests for benzene metabolites detect only acute exposure and are not reliable indicators of cumulative exposure in individual patients. Although it is intuitively obvious and prudent to discontinue the use of agents that might have inflicted severe stem cell injury, by the time the injury has progressed to the point of severe aplastic anemia, most of these patients are in need of the same types of therapy prescribed for patients with autoimmune aplastic anemia.

Idiosyncratic Drug Responses

If a medication history uncovers an exposure to an agent known to be associated with idiosyncratic (not related to dose) responses, the agent must likewise be discontinued. Because it is likely that these responses are immunologically mediated, the patient should be treated no differently from patients with severe idiopathic aplastic anemia, and the patient should be evaluated as a stem cell transplantation candidate. If, during the diagnostic evaluation of the patient and potential donors, there are signs that the marrow is recovering on its own, a more conservative approach can be taken.

Paroxysmal Nocturnal Hemoglobinuria

PNH can be ruled out by screening for CD55 and CD59 on the surface of granulocytes, monocytes, and red cells by flow cytometry. The proper diagnosis of PNH requires the absence of these or other GPI-anchored proteins on at least two hematopoietic cell types. Other characteristic features of this syndrome can be a high low-density lipoprotein level, high indirect bilirubin, low haptoglobin, and a positive urine hemosiderin test result.

Fanconi Anemia

FA should be considered in any adult of any age with a family history of aplastic anemia, acute myelogenous leukemia or myelodysplasia, or squamous cell carcinoma at an unusually young age. This disease should also be considered in any patient with any physical finding listed in Table 165-3 or in patients with a family member who has any of these findings. Unfortunately, some patients with FA meet none of these criteria, so some hematologists, including this author, advocate testing for FA in all patients with aplastic anemia younger than 40 years. This disease can be ruled out by obtaining a chromosomal breakage test (see Table 165-3). Here, either lymphocytes or skin fibroblasts are exposed to cross-linking agents (e.g., mitomycin C or diepoxybutane) for a period of 2 to 3 days, after which metaphase chromosomes are examined for chromosomal breaks and quadriradial forms (four-armed interchromosomal structures). If the clinical context is suggestive (see Table 165-3) but results of the lymphocyte chromosomal breakage test are negative or equivocal, testing of skin fibroblasts is required to rule out the diagnosis. Once the diagnosis is made, history, physical examination, blood counts, and chromosomal breakage tests should be performed on all immediate family members.

There are at least 15 different FA genes (see Table 165-3), and sequencing of them all with cells from every patient is not practical at this time. Fortunately, the involved gene can be first identified by a variety of more affordable complementation analyses. In this type of test, normal FA genes are introduced into primary cells of the patient in vitro, and the one gene that corrects the defect (i.e., reduces hypersensitivity to cross-linking agents [melphalan, mitomycin C, or diepoxybutane]) represents the gene of interest. That gene can then be fully sequenced to identify the precise mutation, information that will be of value to family members. In rare instances, it can also aid in applying preimplantation genetic diagnosis and in vitro fertilization, a process that has successfully resulted in unaffected offspring and ideal cord blood stem cell donors for transplantation of an affected sibling.

Dyskeratosis Congenita

Dyskeratosis congenita should be considered in any aplastic adult of any age with a family member who has had aplastic anemia. It should likewise be considered if either the patient or a family member has had acute myelogenous leukemia, myelodysplasia, nail dystrophy, lacy skin pigmentation, pulmonary fibrosis, oral leukoplakia, squamous cell carcinoma at an unusually young age, or any other physical finding listed in Table 165-3. This disease can be frequently ruled out by quantifying the length of telomeres in circulating white cells by a flow cytometric method. Because some of the physical findings overlap, FA should be ruled out in all patients being evaluated for dyskeratosis congenita. The dyskeratosis test quantifies telomere length with fluorescence in situ hybridization. Lymphocytes from dyskeratosis patients have extremely short telomeres (i.e., at or less than the first percentile). If two or three leukocyte types from a given patient are above the first percentile, the diagnosis of dyskeratosis is unlikely. If telomeres are in the diagnostic range, genetic testing is warranted. It is not yet known whether this test will become a "gold standard" test as reliable as the chromosomal breakage test is for FA because some investigators report that very short telomeres can be found in patients who have other causes of bone marrow failure. Unlike the genetic strategy employed with FA (at least today), complementation analyses are not routinely performed in dyskeratosis. The molecular diagnosis is based on gene sequencing. Once the diagnosis is made and even before the establishment of a molecular genetic diagnosis (see Table 165-3), all immediate family members of the patient should be seen individually, their history taken, and their physical examination performed along with peripheral blood counts and telomere length analysis.

Diamond-Blackfan Anemia

Diamond-Blackfan anemia is an inherited bone marrow failure syndrome that more often exhibits selective erythroid failure and is therefore an unusual cause of full-blown severe aplastic anemia. Although some patients can present in adulthood, most are discovered within the first year of life and present with anemia but less commonly neutropenia and thrombocytopenia. Phenotypic abnormalities like short stature and skeletal defects are the exception in this disease. Caused by mutations in one of at least 10 ribosomal proteins, there are no simple screening tests as reliable as those used for FA and dyskeratosis congenita. However, in patients with unexplained erythroid failure, the finding of an elevated adenosine deaminase level in the serum, although unexplained, is strongly suggestive of this disease, and genetic diagnosis should be considered. Because no validated screening test exists yet, there are no reliable or standardized complementation tests, and genetic diagnosis requires the application of gene sequencing methods (see Table 165-3).

Other Diagnostic Considerations

Aplastic anemia has been reported in recipients of organ allografts, in which mismatched T cells (either from the donated graft or from blood products that had not been irradiated) induce severe aplasia; in patients with myelodysplasia (Chapter 182); in patients with congenital and acquired immunodeficiency states (Chapter 250); and in patients with established autoimmune diseases including systemic lupus erythematosus (Chapter 266) and eosinophilic fasciitis[10] (Chapter 440), a disease characterized by painful swelling of the skin and subcutaneous tissue.

TREATMENT Rx

For patients with milder forms of aplastic anemia, aggressive therapy may not be indicated. A passive approach is more problematic in patients with FA and dyskeratosis congenita because transplantation early in life is better tolerated, immunosuppressive therapy is ineffective, and stem cell transplantation provides the only hope for cure of bone marrow failure. If there is evidence of an underlying autoimmune-mediated disease (e.g., isolated granulopoietic or erythroid failure in patients with rheumatic diseases or thymoma), immunosuppressive therapy alone is often highly effective. In fact, for severe aplastic anemia, immunosuppressive therapy either alone or associated with stem cell transplantation is required. Hematopoietic growth factors alone have been disappointingly ineffective until recently in studies using eltrombopag, a small-molecule agonist of the thrombopoietin receptor (Chapter 172). In a phase II study (starting with 50 mg daily and increasing as needed to a maximum of 150 mg daily, for a total of 12 weeks), this agent induced responses in at least one cell lineage in 44% of patients with refractory severe aplastic anemia,[11] and some responders have done well for two or more years.[12] Because of the theoretical potential for this agent to enhance clonal evolution, its role in the treatment of patients remains to be elucidated, and there are ongoing clinical trials being conducted to that end. Eltrombopag also can restore trilineage hematopoiesis in a subset of patients who are refractory to immunosuppressive therapy.

Hematopoietic Stem Cell Transplantation: Matched Sibling Donor

For patients with severe aplastic anemia, stem cell transplantation offers the advantages of immunosuppression, an infusion of new "healthy" stem cells, and the expectation that the lymphoid cells that suppressed the marrow in the first place will be replaced by more normal cells that have no myelosuppressive capacity. This expectation is supported by large retrospective studies suggesting that stem cell transplantation is superior to immunosuppressive therapy alone for the treatment of severe aplastic anemia, especially in patients younger than 40 years. In the past, this approach was relevant to a minority of patients because only 25 to 30% will have an HLA-matched sibling, but substantial improvements in outcomes for recipients of matched unrelated stem cell transplants have made such stem cell sources reasonable to consider, particularly in patients who have failed to respond to immunosuppressive therapy.[13] No perfectly designed study comparing "front-end" immunosuppressive therapy with stem cell transplantation has been conducted, so patients need to be clearly informed about the risks and benefits of each approach. For patients considering transplantation as a first-line therapy, the unique risks of early treatment-related mortality and later risk of GVHD should be reviewed (Chapter 178). Patients considering immunosuppressive therapy without transplant should be aware of the risks of recurrence and late life-threatening clonal evolution to myelodysplasia or acute leukemia.

Unless the donor is a twin (in which case, peripheral blood–derived stem cells may be preferable[14]), stem cells derived from the bone marrow, not the peripheral blood, should be the source of donor cells.[15] Studies testing the comparative effectiveness of peripheral blood as a source of stem cells have reported that the peripheral blood source is associated with excess mortality stemming from an increased incidence of chronic GVHD. The recipient (in whom congenital marrow failure syndromes have been ruled out) initially receives high-dose cyclophosphamide with either horse antithymocyte

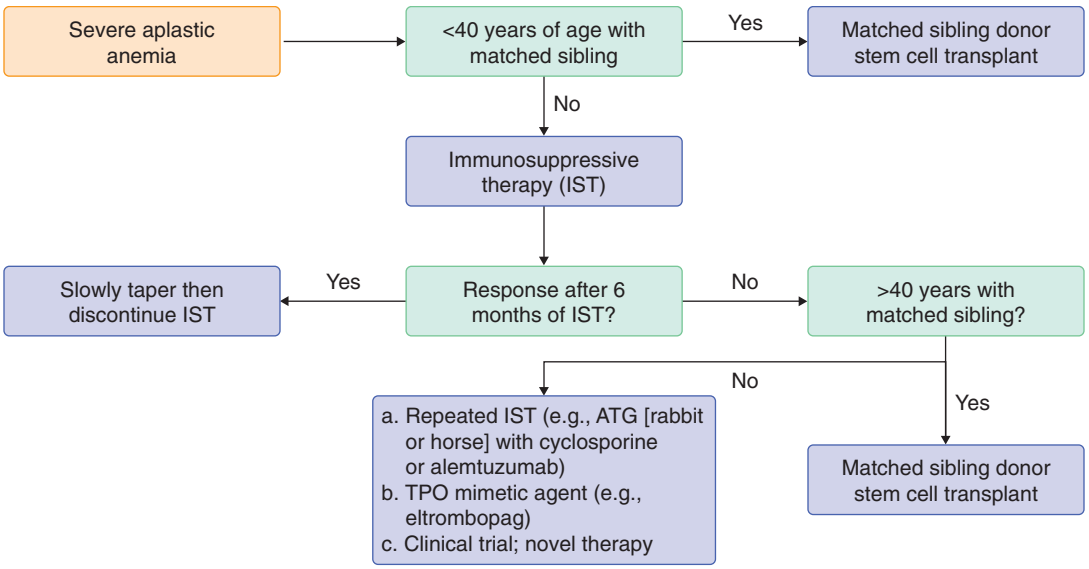

FIGURE 165-3. A therapeutic approach to management of severe aplastic anemia. There are no published results that clarify whether first-line therapy with immunosuppressive therapy (IST) alone is inferior to stem cell transplantation for patients with aplastic anemia. One accepted approach[18] takes into account the high incidence of complications in older patients receiving transplants and suggests matched sibling donor stem cell transplants for any patient younger than 40 years. The remainder receive IST. Because remissions can occur late, 6 months of therapy is required. If there is no response in a patient older than 40 years who also has a matched sibling, stem cell transplantation can be considered at that time. For those without a matched donor, reasonable alternatives are repeated IST, eltrombopag, and enrollment in a clinical trial of novel therapy. A similar approach focusing on the patient's age as a key element in the decision tree is outlined in Table 165-4. ATG = antithymocyte globulin; TPO = thrombopoietin.

globulin (ATG) or alemtuzumab[16,17] as the preparative regimen. Immunosuppressive therapy begins 2 to 4 days before infusion of stem cells. Common post-transplantation immunosuppression combines cyclosporine and methotrexate. The complication of graft failure with this approach is infrequently seen (<5%), grades II to IV acute GVHD is seen in 30 to 50% of recipients, 25% have chronic GVHD, and short-term (2-year) survival after transplantation is nearly 90%. Ten-year survival rates are highly age dependent: 83%, 73%, and 68% for recipients in the first, second, and third decades of life, respectively. In patients 40 years of age or older, the 10-year survival rate is only 51%. Treatment choices adjusted for age of the patient are presented in Table 165-4.

Hematopoietic Stem Cell Transplantation: Matched Unrelated Donor

The use of bone marrow from an HLA-matched unrelated donor should be considered for patients who have no HLA-identical siblings, have failed to respond to immunosuppressive therapy, and are refractory to transfused platelets. As it is with matched related donors, the use of peripheral blood–derived stem cells is associated with higher mortality than is seen with marrow-derived stem cells. Results have not been as favorable as those associated with the use of marrow from HLA-identical siblings, but this strategy has improved in the past decade in part because of more accurate selection of HLA-matched donors and in part because of adjustments made with pretransplantation conditioning regimens. In fact, in some pediatric populations, some centers are now reporting identical outcomes in children transplanted with marrow cells from HLA-matched related and matched unrelated donors. Conditioning regimens that exclude ATG are associated with inferior outcomes, and less toxic conditioning regimens appear to have improved survival (65% for patients older than 16 years and 75% for those 16 years of age or younger). The improvements have been encouraging enough to recommend strongly that unrelated donor searches be initiated at the time of diagnosis for any patient younger than 30 years. Adults older than 30 years should be considered candidates for alternative donor stem cell transplantation if two attempts at immunosuppressive therapy have failed because newer fludarabine-based conditioning regimens have improved outcomes of matched unrelated donor transplants substantially.

Immunosuppressive Therapy

Treatment with ATG alone prolongs survival compared with supportive care alone, and for patients with mild aplastic anemia (not severe), either no treatment or ATG alone is sufficient. In patients with severe aplastic anemia, however, the combination of ATG and cyclosporine is superior to treatment with ATG alone.[16] The combination reduces mortality and induces more rapid and higher overall response rates than does ATG as a single agent. Doses vary from center to center, but ATG is given at doses of 12 to 40 mg/kg/day for 4 to 5 days, and cyclosporine doses of 5 to 10 mg/kg/day (targeting blood levels of 150 to 250 ng/mL) are prescribed for 6 months, after which cyclosporine is slowly tapered during a period of 1 year or more. The median time to response is 120 days. Complete responses are defined as the resolution of pancytopenia and the development of normal blood counts. Relapses are uncommon in complete responders (10%). Partial responders are those whose counts do not normalize but who no longer require transfusion support. Forty percent to 60% of these patients will relapse in 5 years, but most will respond to a repeated course of immunosuppressive therapy. Some will require chronic immunosuppressive therapy with cyclosporine to remain transfusion independent. Those with unresponsive relapses have a poor prognosis. The humanized anti-CD52 monoclonal antibody alemtuzumab is also effective in severe acquired aplastic anemia, although best results are obtained in the relapsed and refractory settings. Its activity has been attributed to its lymphocytotoxic properties. In a conditioning regimen in transplantation trials, it has reduced the incidence of chronic GVHD, so its use is increasing in practice both for immunosuppressive therapy alone and for stem cell transplantation.

All age groups can benefit from immunosuppressive therapy. Long-term survival rates do vary with the age of the treated population. In patients younger than 20 years, 10-year survival rates are 80%, but these rates progressively decline to 25% in patients older than 70 years. However, when adjusted for the survival of an age-matched population by the standardized mortality ratio, the corrected risk for death is highest in young patients and declines as age increases. Advanced age is not a contraindication to the prescription of immunosuppressive therapy.

There are some caveats of importance related to the use of immunosuppressive therapy in aplastic anemia patients. First, patients may experience allergic reactions during infusions of ATG. This should not necessarily dissuade one from continuing this agent. Slowing the infusion after premedication with glucocorticosteroids and antihistamines often solves this problem. Fever and rigor can be signs of cytokine release from damaged cells in the T-cell pool, so they likewise should not be used as a reason to stop treatment. Second, if a responsive patient later relapses, a second course of ATG is frequently effective and can be used, although alemtuzumab is effective in relapsing patients as well. Third, it seems clear that a long interval between diagnosis of aplastic anemia and the initiation of immunosuppressive therapy is a negative predictor of response. Treatment should begin as soon as possible, certainly within 3 weeks of initial diagnosis. Fourth, the inclusion of hematopoietic growth factors during immunosuppressive therapy provides no benefit.

An approach to the management of severe aplastic anemia is shown in Figure 165-3, which complements Table 165-4.

Diamond-Blackfan Anemia

Although the mechanism by which glucocorticosteroids induce remissions in patients with Diamond-Blackfan anemia is unknown, nearly 80% of patients initially respond. Prednisone treatment is started at 2 mg/kg/day and is tapered after the hemoglobin increases to 10 g/dL. Most patients require a low every-other-day dose, but 15% remain in remission off steroids altogether. In those cases, survival to 40 years of age is nearly universal. Of patients who require ongoing steroid support, 75% reach the age of 40 years. Only half of the patients with steroid-resistant severe anemia survive to the age of 40 years. Pure red cell aplasia may respond to a synthetic erythropoietin receptor agonist.

Supportive Care

Platelet Transfusions

In the absence of bleeding or infection, platelet transfusions are commonly administered only when the platelet count declines to 10,000/mL or less (Chapters 172 and 177), but in the presence of active infection or bleeding, transfusion thresholds are often set at 20,000/mL or higher. If bleeding and infection coexist, it is prudent to rule out disseminated intravascular coagulation because fresh-frozen plasma or cryoprecipitate may be required along with platelet transfusions. Drugs (e.g., aspirin) that inhibit platelet function should be avoided, as should activities that might result in trauma. Menstrual activity should be suppressed with oral contraceptives or other agents.

Red Cell Transfusions

Filtered packed red blood cells should be provided to meet the needs of the patient's daily activities. Because there is substantial interindividual variation and because comorbidities influence exercise tolerance, there can be no firmly established hemoglobin target number for everyone. As a general rule, however, the number is higher in elderly patients than in young patients. Children with hemoglobin levels as low as 6 g/dL can compensate reasonably. Adults with underlying cardiopulmonary disease may have symptomatic anemia at 8 g/dL. If chronic transfusion therapy is required, the development of iron overload may require chelation therapy.

Management of Infections

The major infectious challenges result from the immunosuppression used (either as primary therapy or as a component of a stem cell transplantation regimen) more than from the neutropenia. For that reason, antibacterial, antiviral, and antifungal prophylaxis is routinely used in transplant recipients. In many centers, patients treated with immunosuppressive therapy alone are treated similarly for a 2- to 3-month period after ATG administration. Importantly, the onset of fever requires prompt clinical evaluation and antibiotic therapy as described in Chapter 281.

PREVENTION

Apart from public health measures controlling exposures to benzene, aromatic hydrocarbons, and radiation, little can be done to prevent acquired aplastic anemia. In the inherited bone marrow failure syndromes, prevention is achievable. Once the proband is identified, other affected family members, carriers, and siblings with no mutant allele can be identified, and genetic results can be applied in family planning even to the extent of preimplantation genetic diagnosis followed by in vitro fertilization. Because all somatic cells of children and adults with FA are hypersensitive to alkylating agents and oxidative stress, they must not receive standard doses of radiation or alkylating agents either for transplant conditioning or for treatment of squamous cell carcinoma.

Clonal evolution (e.g., to myelodysplasia and acute leukemia) occurs in 10 to 20% of patients with acquired aplastic anemia and up to 40% of children and adults with dyskeratosis congenita and FA. This complication likely evolves through a process of clonal selection and adaptation and for that reason is seen less commonly in patients completely responsive to treatment than in those with less than complete responses (who therefore have ongoing suppression of hematopoiesis). This suggests that more effective strategies of immunosuppressive therapy may better control ongoing marrow damage and decrease the incidence of clonal evolution.

PROGNOSIS

Acquired Idiopathic Aplastic Anemia

The severity of aplastic anemia and age are key determinants of long-term survival. Both these factors influence the choice of therapy (see Table 165-4). Early intervention is associated with a better prognosis, which should dictate the pace of diagnostic evaluation. Matched sibling donor transplants are associated with higher response rates, lower relapse rates, long-term survival rates approximating 80%, and lower incidence of clonal evolution. Immunosuppressive therapy is associated with long-term (10-year) survival rates of 70 to 75% in responders. Infection represents the most common cause of death in patients of any age treated with either immunosuppression alone or stem cell transplantation. Somatic mutations can identify patients with markedly increased risks of progressing to a myelodysplastic syndrome.[19]

Inherited Bone Marrow Failure Syndromes

Patients with dyskeratosis congenita and FA will not respond to immunosuppressive therapy. Stem cell transplantation with nonmyeloablative approaches has the potential of curing the marrow failure component of these diseases but does nothing to reduce the other common life-threatening complication of squamous cell carcinoma. There are good theoretical reasons for anticipating that successful transplantation will reduce the likelihood of clonal evolution to myelodysplasia and acute leukemia. Unfortunately, the decision to transplant is influenced by other key factors, as described earlier. In light of these complexities, no clear-cut when-to-transplant rule can be applied to all patients with these diseases, but it is generally accepted that transplantation early in childhood is associated with fewer short- and long-term complications. Taking these difficult issues into account, all patients should be evaluated early in an experienced transplantation center, and family members should be screened by hematologists, geneticists, and genetic counselors with use of specialty laboratories.

In patients with dyskeratosis, small case series report good outcomes after fludarabine-containing nonmyeloablative conditioning regimens, but studies of sufficient size are not available to permit concrete recommendations except that patients should be referred to an experienced center in light of the other organ systems at risk (e.g., lung). In patients with FA, with proper conditioning regimens, matched sibling donor transplant recipients have expected 3-year survival rates of about 85%, and matched unrelated recipients (when fludarabine-containing nonmyeloablative approaches are used) have 3-year survival rates of 50%. In patients with Diamond-Blackfan anemia, sibling donor transplant recipients have 3-year survival rates of about 80%, but disease-free survival rates after matched unrelated donor transplants have been poor (20 to 30%).

Future Treatments

For patients with acquired aplastic anemia, studies designed to selectively target the T-cell clones responsible for hematopoietic suppression may provide more effective and less toxic strategies for immunosuppressive therapy. For all patients with aplastic anemia, further improvements in matched unrelated transplantation should make this modality more widely available for patients who are not now considered optimal candidates. GVHD control will continue to improve. For children with nonsevere aplastic anemia, the 10-year progression-free survival rate is only about 25%, suggesting that prospective trials of early intervention are warranted. For patients with inherited bone marrow failure syndromes, the genes for which have been largely identified, the possibility of stem cell gene therapy holds enormous theoretical promise and has been nicely validated in murine models of the disease.

Grade A Reference

A1. Gafter-Gvili A, Ram R, Gurion R, et al. ATG plus cyclosporine reduces all-cause mortality in patients with severe aplastic anemia: systematic review and meta-analysis. *Acta Haematol.* 2008; 120:237-243.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Schneider M, Chandler K, Tischkowitz M, et al. Fanconi anaemia: genetics, molecular biology, and cancer—implications for clinical management in children and adults. *Clin Genet*. 2014;[Epub ahead of print].
2. Shin SH, Lee JW. The optimal immunosuppressive therapy for aplastic anemia. *Int J Hematol*. 2013;97:564-572.
3. Kordasti S, Marsh J, Al-Khan S, et al. Functional characterization of $CD4^+$ T cells in aplastic anemia. *Blood*. 2012;119:2033-2043.
4. Krell PF, Reuther S, Fischer U, et al. Next-generation-sequencing-spectratyping reveals public T-cell receptor repertoires in pediatric very severe aplastic anemia and identifies a β chain CDR3 sequence associated with hepatitis-induced pathogenesis. *Haematologica*. 2013;98:1388-1396.
5. Guinan EC, Barbon CM, Kalish LA, et al. Bactericidal/permeability-increasing protein (rBPI21) and fluoroquinolone mitigate radiation-induced bone marrow aplasia and death. *Sci Transl Med*. 2011;3:110ra118.
6. Parker CJ. Paroxysmal nocturnal hemoglobinuria. *Curr Opin Hematol*. 2012;19:141-148.
7. Garayoechea JI, Patel KJ. Why does the bone marrow fail in Fanconi anemia? *Blood*. 2014;123:26-34.
8. Horos R, von Lindern M. Molecular mechanisms of pathology and treatment in Diamond Blackfan anaemia. *Br J Haematol*. 2012;159:514-527.
9. Garbati MR, Hays LE, Keeble W, et al. FANCA and FANCC modulate TLR and p38 MAPK dependent expression of IL-1β in macrophages. *Blood*. 2013;122:3197-3205.
10. de Masson A, Bouaziz JD, Peffault de Latour R, et al. Severe aplastic anemia associated with eosinophilic fasciitis: report of 4 cases and review of the literature. *Medicine (Baltimore)*. 2013;92:69-81.
11. Olnes MJ, Scheinberg P, Calvo KR, et al. Eltrombopag and improved hematopoiesis in refractory aplastic anemia. *N Engl J Med*. 2012;367:11-19.
12. Desmond R, Townsley DM, Dumitriu B, et al. Eltrombopag restores trilineage hematopoiesis in refractory severe aplastic anemia that can be sustained on discontinuation of drug. *Blood*. 2014;123:1818-1825.
13. Bacigalupo A. Matched and mismatched unrelated donor transplantation: is the outcome the same as for matched sibling donor transplantation? *Hematology Am Soc Hematol Educ Program*. 2012;2012:223-229.
14. Gerull S, Stern M, Apperley J, et al. Syngeneic transplantation in aplastic anemia: pretransplant conditioning and peripheral blood are associated with improved engraftment—an observational study on behalf of the Severe Aplastic Anemia and Pediatric Diseases Working Parties of the European Group for Blood and Marrow Transplantation. *Haematologica*. 2013;98:1804-1809.
15. Bacigalupo A, Socie G, Schrezenmeier H, et al. Bone marrow versus peripheral blood as the stem cell source for sibling transplants in acquired aplastic anemia: survival advantage for bone marrow in all age groups. *Haematologica*. 2012;97:1142-1148.
16. Marsh JC, Pearce RM, Koh MB, et al. Retrospective study of alemtuzumab vs ATG-based conditioning without irradiation for unrelated and matched sibling donor transplants in acquired severe aplastic anemia: a study from the British Society for Blood and Marrow Transplantation. *Bone Marrow Transplant*. 2014;49:42-48.
17. Gandhi S, Kulasekararaj AG, Mufti GJ, et al. Allogeneic stem cell transplantation using alemtuzumab-containing regimens in severe aplastic anemia. *Int J Hematol*. 2013;97:573-580.
18. Scheinberg P, Young NS. How I treat acquired aplastic anemia. *Blood*. 2012;120:1185-1196.
19. Kulasekararaj AG, Jiang J, Smith AE, et al. Somatic mutations identify a subgroup of aplastic anemia patients who progress to myelodysplastic syndrome. *Blood*. 2014;124:2698-2704.

REVIEW QUESTIONS

1. A 20-year-old man with severe aplastic anemia is diagnosed by chromosomal breakage testing to have Fanconi anemia. Which of the following would be the preferred first-line treatment?

A. Antithymocyte globulin (ATG)
B. Bone marrow transplant from a matched related donor after fludarabine-based conditioning
C. Cord blood hematopoietic stem cell transplant after conditioning that includes irradiation
D. Supportive therapy with erythropoiesis-stimulating agents (ESAs) and granulocyte colony-stimulating factor (G-CSF)
E. Cyclosporine

Answer: B None of these is an entirely satisfactory first-line therapy for aplastic anemia caused by an inherited bone marrow failure syndrome (Fanconi anemia, dyskeratosis congenita, Diamond-Blackfan anemia), as in this case. These patients do not respond to ATG and other immunosuppressive therapies (including cyclosporine). Growth factors like ESAs and G-CSF are generally ineffective in severe aplastic anemia. Hematopoietic stem cell transplantation (HSCT) that includes irradiation in the conditioning regimen exposes the patient to an additional risk factor interacting with the main underlying biologic defect of Fanconi anemia, the DNA repair process. The use of specifically bone marrow as the source of HSCT and reduced-intensity conditioning regimens with fludarabine appear to have the most favorable outcome at this point, although it will not reduce the subsequent risk of a second malignant neoplasm. (Peffault de Latour R, Porcher R, Dalle JH, et al. Allogeneic hematopoietic stem cell transplantation in Fanconi anemia: the European Group for Blood and Marrow Transplantation experience. Blood. 2013;122:4279-4286).

2. Which of the following statements is not correct regarding inherited bone marrow failure syndromes?

A. Most of the known genes inactivated in Fanconi anemia encode proteins that protect the genome from excessive damage induced by chemical cross-linking agents.
B. Most of the known genes inactivated in dyskeratosis congenita encode proteins that participate in the maintenance of telomeres.
C. Most of the known genes inactivated in Diamond-Blackfan anemia encode ribosomal proteins.
D. The best screening test for Diamond-Blackfan anemia is the chromosomal breakage test.
E. The best screening test for dyskeratosis congenita is quantitative analysis of telomere length (flow-FISH) in lymphocytes.

Answer: D Gene mutations for these three syndromes mostly involve inactivations of specific genes that encode the key proteins involved in their molecular pathogenesis, respectively, as correctly indicated in choices A to C. These also form the basis of screening tests for Fanconi anemia and dyskeratosis congenita; however, there are no adequate screening tests for Diamond-Blackfan anemia. (For Diamond-Blackfan anemia, serum adenosine deaminase is often elevated, but this does not represent a screening test.) Gene sequencing can provide specific mutation analysis for all three of these syndromes, when needed.

3. In a patient with pancytopenia, which of the following abnormalities in the peripheral blood smear would provide a useful indicator that the cause of the pancytopenia is aplastic anemia?

A. Hypersegmented polymorphonuclear leukocytes
B. Nucleated red cells
C. Teardrop-shaped red cells
D. Giant platelets
E. None of the above

Answer: E The peripheral blood smear in aplastic anemia is generally normal, even in severe cases. This does not mean that the peripheral blood smear does not provide useful information about the etiology of pancytopenia in many cases. For example, hypersegmented polymorphonuclear leukocytes (A) suggest megaloblastic anemia (which in more advanced stages can cause pancytopenia, not just anemia). Nucleated (B) and teardrop-shaped (C) red cells suggest a "myelophthisic" process that involves invasion of the bone marrow by "foreign" elements (like metastatic cancer, fibrosis, or granulomas).

166

POLYCYTHEMIA VERA, ESSENTIAL THROMBOCYTHEMIA, AND PRIMARY MYELOFIBROSIS

AYALEW TEFFERI

DEFINITION

Polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF) belong to the category of myeloproliferative neoplasms (MPNs), under the 2008 World Health Organization (WHO) classification system for hematologic malignancies (Table 166-1). These disorders represent stem cell–derived clonal myeloproliferation with a propensity to evolve into acute myeloid leukemia (AML), also called blast-phase MPN. PV, ET, and PMF, together with chronic myelogenous leukemia (CML), used to be

TABLE 166-1 WORLD HEALTH ORGANIZATION CLASSIFICATION OF MYELOID MALIGNANCIES

1. Acute myeloid leukemia (AML) and related precursor neoplasms* (Chapter 183)
2. Myeloproliferative neoplasms (MPN)
   - 2.1. Classic MPN
     - 2.1.1. Chronic myelogenous leukemia, *BCR-ABL1* positive (CML)
     - 2.1.2. Polycythemia vera (PV)
     - 2.1.3. Primary myelofibrosis (PMF)
     - 2.1.4. Essential thrombocythemia (ET)
   - 2.2. Nonclassic MPN
     - 2.2.1. Chronic neutrophilic leukemia (CNL)
     - 2.2.2. Chronic eosinophilic leukemia, not otherwise specified (CEL-NOS)
     - 2.2.3. Mastocytosis
     - 2.2.4. Myeloproliferative neoplasm, unclassifiable (MPN-U)
3. Myelodysplastic syndromes (MDS) (Chapter 182)
   - 3.1. Refractory cytopenia† with unilineage dysplasia (RCUD)
     - 3.1.1. Refractory anemia (ring sideroblasts <15% of erythroid precursors)
     - 3.1.2. Refractory neutropenia
     - 3.1.3. Refractory thrombocytopenia
   - 3.2. Refractory anemia with ring sideroblasts (RARS; dysplasia limited to erythroid lineage and ring sideroblasts ≥15% of bone marrow erythroid precursors)
   - 3.3. Refractory cytopenia with multilineage dysplasia (RCMD; ring sideroblast count does not matter)
   - 3.4. Refractory anemia with excess blasts (RAEB)
     - 3.4.1. RAEB-1 (2%-4% circulating or 5%-9% marrow blasts)
     - 3.4.2. RAEB-2 (5%-19% circulating or 10%-19% marrow blasts or Auer rods present)
   - 3.5. MDS associated with isolated del(5q)
   - 3.6. MDS, unclassifiable
4. MDS/MPN
   - 4.1. Chronic myelomonocytic leukemia (CMML)
   - 4.2. Atypical chronic myeloid leukemia, *BCR-ABL1* negative
   - 4.3. Juvenile myelomonocytic leukemia (JMML)
   - 4.4. MDS/MPN, unclassifiable
     - 4.4.1. Provisional entity: refractory anemia with ring sideroblasts associated with marked thrombocytosis (RARS-T)
5. Myeloid and lymphoid neoplasms with eosinophilia and abnormalities of *PDGFRA*,‡ *PDGFRB*,‡ or *FGFR1*‡ (Chapter 170)
   - 5.1. Myeloid and lymphoid neoplasms with *PDGFRA* rearrangement
   - 5.2. Myeloid neoplasms with *PDGFRB* rearrangement
   - 5.3. Myeloid and lymphoid neoplasms with *FGFR1* abnormalities

*Acute myeloid leukemia–related precursor neoplasms include "therapy-related myelodysplastic syndrome" and "myeloid sarcoma."
†Either mono- or bicytopenia: hemoglobin level <10 g/dL, absolute neutrophil count $<1.8 \times 10^9$/L, or platelet count $<100 \times 10^9$/L. However, higher blood counts do not exclude the diagnosis in the presence of unequivocal histologic or cytogenetic evidence for myelodysplastic syndrome.
‡Genetic rearrangements involving platelet-derived growth factor receptor α/β (*PDGFRA*/*PDGFRB*) or fibroblast growth factor receptor 1 (*FGFR1*).

referred to as "myeloproliferative disorders." Because CML is invariably linked to the Philadelphia translocation (i.e., *BCR-ABL1*), the other three are operationally labeled "*BCR-ABL1*–negative MPN."

EPIDEMIOLOGY

According to a recent systematic review, reported annual incidence rates ranged from 0.01 to 2.61, 0.21 to 2.27, and 0.22 to 0.99 per 100,000 for PV, ET, and PMF, respectively.[1] The combined annual incidence rates for PV, ET, and PMF are 0.84, 1.03, and 0.47 per 100,000, respectively. Population-based studies suggest a median age at diagnosis of approximately 71 years when all three MPN variants are considered together and a male-to-female ratio of approximately 50%. Other studies indicate a lower age distribution for ET and a slight male preponderance for PMF. Family studies suggest a five- to sevenfold increased risk of MPN among first-degree relatives of patients with *BCR-ABL1*–negative MPN and the possibility of a hereditary component to disease susceptibility was further elaborated by *JAK2* haplotype studies. PMF has been associated with exposure to ionizing radiation (e.g., in Hiroshima survivors), heavy exposure to petroleum derivatives, and thorium dioxide (Thorotrast) contrast medium, but in the vast majority of cases, there is no such exposure history.

PATHOBIOLOGY

These diseases are clonal in nature, deriving from a genetically transformed hematopoietic, bone marrow stem cell that results in clonal myeloproliferation. In 2005, a *JAK2* gain-of-function mutation (*JAK2*V617F) was reported in *BCR-ABL1*–negative MPN. Subsequent studies using sensitive assays have revealed the presence of this mutation in approximately 95% of patients with PV and 60% of those with ET or PMF. Most of the remaining 5% of patients with PV carry another *JAK2* mutation (*JAK2* exon 12). In other words, virtually all patients with PV carry a *JAK2* mutation. Approximately 5% to 10% of *JAK2*V617F-negative patients with ET or PMF carry a *MPL*W515 mutation, which is equally JAK-STAT relevant (Table 166-2). In 2013, calreticulin (*CALR*) mutations were discovered in the majority of patients with ET or PMF who do not express *JAK2* or *MPL* mutations.[2] *CALR* mutations are relatively specific to ET and PMF and display mutational frequencies of approximately 20% and 25%, respectively. Additional mutations seen in MPN are listed in Table 166-2. Most of these latter mutations originate at the progenitor cell level, but they do not necessarily represent the primary clonogenic event, are not mutually exclusive, or follow a predictable hierarchy.

The above-listed molecular alterations in *BCR-ABL1*–negative MPN induce biologic changes that are demonstrated in animal models or ex vivo. For example, *JAK2* or *MPL* mutations induce PV-, ET-, or PMF-like disease in mice by experimental manipulation of mutant allele burden. These mutations are also believed to contribute to the growth factor independence or hypersensitivity of erythroid or megakaryocyte colony-forming progenitor cells. Bone marrow fibrosis, osteosclerosis, and angiogenesis in PMF are currently believed to be reactive and cytokine mediated.

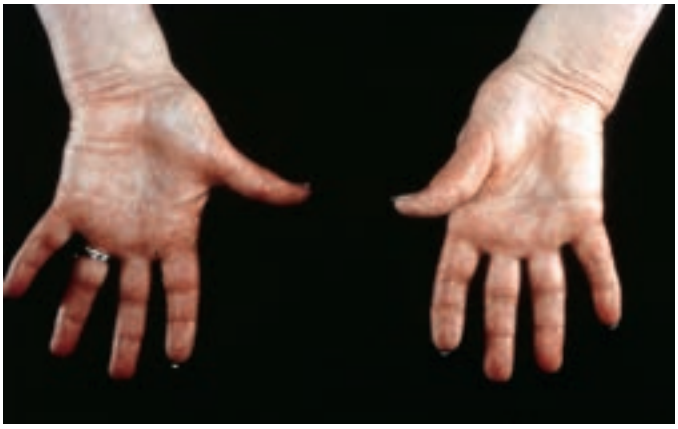

FIGURE 166-1. Erythromelalgia: painful red discoloration of the hands or, more commonly, the toes.

CLINICAL MANIFESTATIONS

Essential Thrombocythemia

At presentation, microvascular and vasomotor symptoms are found in 25% to 50% of ET patients. Major thrombosis is seen in 11% to 25% of patients at diagnosis and 10% to 22% during follow-up, and major hemorrhage is observed in 2% to 5% at diagnosis and 1% to 7% during follow-up. Vasomotor disturbances (e.g., headaches, lightheadedness, visual symptoms such as blurring and scotomata, palpitations, chest pain, erythromelalgia, and distal paresthesias) are not infrequent in ET and might be the result of abnormal platelet–endothelium interactions. Erythromelalgia (Fig. 166-1) is the most dramatic vasomotor symptom, characterized by erythema, warmth, and pain in the distal extremities; this symptom is rare but not entirely specific for ET. Life-threatening complications of ET include large-vessel thrombosis (both arterial and venous), hemorrhage, and transformation of the disease into either a fibrotic phase resembling PMF or AML. Venous thrombosis in ET occurs both in sites common to other thrombotic diatheses (e.g., pulmonary embolism and lower extremity deep vein thrombosis and pulmonary embolism) but also in more unusual sites (e.g., cerebral sinus thrombosis, retinal vein thrombosis, and hepatic and portal vein thrombosis).

Major hemorrhage in ET is most common in the gastrointestinal (GI) tract and may be precipitated by aspirin (ASA) or nonsteroidal antiinflammatory drug (NSAID) use. Hemorrhage also occurs in the central nervous system (CNS) and the retina, but such events are, fortunately, uncommon. Paradoxically, patients with extreme thrombocytosis may be at special risk for bleeding, in part related to the development of an acquired von Willebrand syndrome that is thought to be related to platelet adsorption of large multimers of von Willebrand protein (Chapter 173). Fibrotic and leukemic transformations of ET are rare events (<5% of patients) during the first 10 years after diagnosis, but the risk increases with time.

TABLE 166-2 SOMATIC MUTATIONS IN *BCR-ABL1*–NEGATIVE MYELOPROLIFERATIVE NEOPLASMS (MPN), INCLUDING POLYCYTHEMIA VERA (PV), ESSENTIAL THROMBOCYTHEMIA (ET), AND PRIMARY MYELOFIBROSIS (PMF). MUTATIONAL FREQUENCIES IN BLAST PHASE (BP) DISEASE ARE ALSO PROVIDED*

MUTATIONS	CHROMOSOME LOCATION	MUTATIONAL FREQUENCY	PATHOGENETIC RELEVANCE
JAK2 (Janus kinase 2): *V617F*	9p24	PV: ~96% ET:~55% PMF: ~65%	Contributes to abnormal myeloproliferation and progenitor cell growth factor hypersensitivity
JAK2 exon 12 mutation	9p24	PV: ~3%	Contributes to primarily erythroid myeloproliferation
CALR (Calreticulin): exon 9 deletions and insertions	19p13.2	PMF: ~25% ET: ~20% PV: 0%	Wild-type *CALR* is a multifunctional Ca^{2+} binding protein chaperone mostly localized in the endoplasmic reticulum
MPL (myeloproliferative leukemia virus oncogene): MPN-associated *MPL* mutations involve exon 10	1p34	ET:~3% PMF: ~10%	Contributes to primarily megakaryocytic myeloproliferation
LNK (as in links), or *SH2B3* (a membrane-bound adaptor protein): MPN-associated mutations are monoallelic and involve exon 2	12q24.12	PV: rare ET: rare PMF: rare BP-MPN: ~10%	Wild-type *LNK* is a negative regulator of JAK2 signaling.
TET2 (TET oncogene family member 2): mutations involve several exons	4q24	PV: ~16% ET: ~5% PMF: ~17% BP-MPN: ~17%	TET proteins catalyze conversion of 5mC to 5hmC, which favors demethylated DNA. Both TET1 and TET2 display this catalytic activity. *IDH* and *TET2* mutations might share a common pathogenetic effect.
ASXL1 (additional sex combs-like 1): exon 12 mutations	20q11.1	ET: ~3% PMF: ~13% BP-MPN: ~18%	Wild-type *ASXL1* is needed for normal hematopoiesis and might be involved in coactivation of transcription factors and transcriptional repression.
IDH1/IDH2 (isocitrate dehydrogenase): exon 4 mutations	2q33.3/15q26.1	PV: ~2% ET: ~1% PMF: ~4% BP-MPN: ~20%	*IDH* mutations induce loss of activity for the conversion of isocitrate to 2-KG and gain of function in the conversion of 2-KG to 2-HG. 2-HG might be the mediator of impaired TET2 function in cells with mutant *IDH* expression.
EZH2 (enhancer of zeste homolog 2): mutations involve several exons	7q36.1	PV: ~3% PMF: ~7% MDS: ~6%	Wild-type *EZH2* is part of a histone methyltransferase (polycomb repressive complex 2 associated with H3 Lys-27 trimethylation). MPN-associated *EZH2* mutations might have a tumor suppressor activity, which contrasts with the gain-of-function activity for lymphoma-associated *EZH2* mutations.
DNMT3A (DNA cytosine methyltransferase 3a): most frequent mutations affect amino acid R882	2p23	PV: ~7% PMF: ~7% BP-MPN: ~14%	DNA methyl transferases are essential In establishing and maintaining DNA methylation patterns in mammals
CBL (Casitas B-lineage lymphoma proto-oncogene): exon 8/9 mutations	11q23.3	PV: rare ET:rare MF: ~6%	CBL is an E3 ubiquitin ligase that marks mutant kinases for degradation. Transforming activity requires loss of this function.
IKZF1 (IKAROS family zinc finger 1): mostly deletions including intragenic	7p12	CP-MPN: rare BP-MPN: ~19%	IKZF1 is a transcription regulator and putative tumor suppressor
TP53 (tumor protein p53): exons 4 through 9	17p13.1	PMF: ~4% BP-MPN ~27%	A tumor suppressor protein that targets genes that regulate cell cycle arrest, apoptosis, and DNA repair
SF3B1 (splicing factor 3B subunit 1): mostly exons 14 and 15	2q33.1	PMF: ~7%	SF3B1 is a component of the RNA spliceosome. *SF3B1* mutations are closely associated with ring sideroblasts.
SRSF2 (serine/arginine-rich splicing factor 2): exon 2	17q25.1	PMF: ~17%	SRSF2 is a component of the RNA spliceosome, whose dysfunction promotes defects in alternative splicing.
U2AF1 (U2 small nuclear RNA auxiliary factor 1)	21q22.3	PMF: ~16%	U2AF1 is a subunit of the U2 small nuclear ribonucleoprotein auxiliary factor involved in pre-mRNA processing

*See text for references.
BP-MPN = blast phase MPN; CP-MPN= chronic phase MPN; ET = essential thrombocythemia; 5hmC = 5-hydroxymethylcytosine; 5mC = 5-methylcytosine; MF includes both PMF and post-ET/PV myelofibrosis; MPN = myeloproliferative neoplasms; PMF = primary myelofibrosis; PV = polycythemia vera; 2-KG = 2-ketoglutarate; 2-HG = 2-hydroxyglutarate.

Polycythemia Vera

Table 166-3 lists the typical clinical and laboratory features of PV. Increased red blood cell (RBC) mass in PV might result in blood hyperviscosity, which leads to a plethora of symptoms and signs. Headaches are frequent, but blurry vision, altered hearing, mucous membrane bleeding, shortness of breath, and malaise are also observed. At least two thirds of PV patients have splenomegaly. Thrombosis occurs, most commonly arterial thrombosis, in about 40% of patients. As in ET, venous thrombosis can occur in unusual sites, such as mesenteric, portal, or hepatic veins (the latter also being called Budd-Chiari syndrome). Bleeding, especially from the GI tract, is seen in PV but less often than thrombosis. Pruritus is common in PV and may be provoked by warm water ("aquagenic"). Erythromelalgia (described earlier under ET) might also trouble some patients with PV, as do other vasomotor symptoms, such as paresthesias and headaches.

TABLE 166-3 CLINICAL AND LABORATORY FEATURES OF POLYCYTHEMIA VERA

Persistent leukocytosis
Persistent thrombocytosis
Microcytosis secondary to iron deficiency
Increased red blood cell mass
JAK2 mutations
Increased leukocyte alkaline phosphatase
Splenomegaly
Generalized pruritus (usually after bathing)
Arterial and venous thrombosis, including unusual thrombosis (e.g., Budd-Chiari syndrome)
Erythromelalgia (acral dysesthesia and erythema; see Fig. 166-1)

Primary Myelofibrosis

Most patients with PMF present with anemia and marked splenomegaly. The anemia of PMF is multifactorial. Contributing factors include ineffective hematopoiesis and hypersplenism. Spleen and liver enlargement in PMF is secondary to extramedullary hematopoiesis (EMH) and may be associated with hypercatabolic symptoms, including profound fatigue, weight loss, night sweats, and low-grade fever. Patients also experience peripheral edema; diarrhea; early satiety; and, occasionally, complications of portal hypertension, including variceal bleeding and ascites.

Splenomegaly in PMF may be complicated by splenic infarction manifested by left upper quadrant abdominal pain and referred left shoulder pain. CT imaging in such cases can be unremarkable or may show wedge-shaped or rounded low-attenuation lesions in the spleen. EMH occurs in a great diversity of sites throughout the body. Common sites besides the spleen and liver include lymph nodes, skin, pleura, peritoneum, lung, and paraspinal and epidural spaces. The latter may result in spinal cord or nerve root compression, which is a medical emergency requiring corticosteroids to reduce edema and immediate radiotherapy. Other clinical features of PMF include diffuse and sometimes regional bone and joint pain.

DIAGNOSIS

At present, the 2008 WHO criteria are used for the diagnosis of ET, PV, and PMF (Table 166-4). These criteria are based on morphology and cytogenetic and molecular studies (Fig. 166-2). Almost all patients with PV carry a *JAK2* mutation (*JAK2*V617F or *JAK2* exon 12 mutations). *JAK2*V617F is, however, not specific to PV and is also found in ET (~60% of cases), PMF (~60% of cases), and other myeloid neoplasms (usually <5% of cases). *JAK2* exon 12 mutations are relatively specific to *JAK2*V617F-negative PV and occur in approximately 3% of all patients with PV. *JAK2* exon 12 mutation-positive PV patients are characterized by predominantly erythroid myelopoiesis, subnormal serum erythropoietin (Epo) levels, and younger age at diagnosis. Taken together, the presence of a *JAK2* mutation excludes secondary myeloproliferation (e.g., secondary polycythemia or reactive thrombocytosis; see later discussion), and its absence makes a diagnosis of PV very unlikely. Among the 30% to 40% of ET or PMF patients who do not harbor *JAK2* mutations, the majority (60% to 70%) carry *CALR* mutations, and 5% to 10% carry *MPL* mutations. In other words, 80% to 90% of patients with ET or PMF carry one of the three MPN-specific mutations: *JAK2, CALR,* or *MPL*.

As outlined in Figure 166-2, the diagnostic work-up for suspected MPN, including PV, ET, and PMF, should start with peripheral blood mutation screening for *JAK2*V617F.[3] The presence of the mutation confirms the presence of an underlying MPN, but it is not specific to any one of the three MPN variants. In the absence of *JAK2*V617F, the next step is to screen for *JAK2* exon 12 mutation for PV and *CALR* mutations for ET and PMF (see Fig. 166-2). *MPL* mutations are studied only when both *JAK2* and *CALR* mutations are absent in suspected cases of ET or PMF. Bone marrow examination is recommended to confirm the diagnosis in triple-negative cases (i.e., wild type for *JAK2, CALR,* and *MPL*) and to distinguish between ET and prefibrotic PMF. Clinically, patients with ET and PV present with thrombocytosis and erythrocytosis, respectively. Patients with PMF usually present with peripheral blood leukoerythroblastosis (i.e., presence of nucleated RBCs, metamyelocytes, or myelocytes [Fig. 166-3] and bone marrow fibrosis with morphologically atypical megakaryocytes [Fig. 166-4]). Bone marrow morphology can also help distinguish clonal from reactive myeloproliferation (Fig. 166-5).

Table 166-5 includes a comprehensive list of causes of erythrocytosis,[4] including congenital and secondary polycythemia. The diagnostic approach to congenital polycythemia should start with measurement of serum Epo level. The presence of a subnormal serum Epo level, in the absence of PV, suggests the presence of a germline mutation of the erythropoietin receptor. If the serum Epo level is normal or elevated, the next step is to measure the p50 (the oxygen tension at which hemoglobin is 50% saturated). Decreased p50 suggests the presence of either high oxygen-affinity hemoglobinopathy or 2,3-bisphosphoglycerate (2,3-BPG) deficiency. If the p50 is normal, then the possibility of *VHL* mutations (usually associated with increased serum Epo level) should be considered first because they constitute the most frequent mutations in congenital polycythemia. Gene expression studies can identify two different phenotypic expressions of polycythemia vera in terms of disease duration, hemoglobin level, splenomegaly, and the risk of thromboembolism.[5]

Table 166-6 outlines the different causes of thrombocytosis, and Figure 166-5 provides an algorithmic approach to its diagnosis. Figure 166-6 illustrates the value of peripheral blood smear examination in the differential diagnosis of thrombocytosis. Other causes of bone marrow fibrosis are outlined in Table 166-7. The diagnosis of post-PV or post-ET myelofibrosis requires full documentation of a previous morphologic diagnosis of PV or ET, respectively.

TABLE 166-4 THE 2008 WORLD HEALTH ORGANIZATION DIAGNOSTIC CRITERIA FOR POLYCYTHEMIA VERA, ESSENTIAL THROMBOCYTHEMIA, AND PRIMARY MYELOFIBROSIS

CRITERIA	POLYCYTHEMIA VERA*	ESSENTIAL THROMBOCYTHEMIA*	PRIMARY MYELOFIBROSIS*
Major criteria	1. Hb >18.5 g/dL (men); >16.5 g/dL (women) *or* Hb/Hct >99th percentile of reference range *or* Hb >17 g/dL (men) or>15 g/dL (women) if associated with a sustained increase of ≥2 g/dL from baseline, not otherwise explained *or* Elevated RBC mass >25% above mean normal predicted value	1. Platelet count ≥450 × 10^9/L	1. Megakaryocyte proliferation and atypia† accompanied by either reticulin or collagen fibrosis *or* In the absence of fibrosis, megakaryocyte changes accompanied by increased marrow cellularity and granulocytic proliferation
	2. Presence of *JAK2*V617F or similar mutation	2. Megakaryocyte proliferation with large and mature morphology; no or little granulocyte or erythroid proliferation	2. Not meeting WHO criteria for CML, PV, MDS, or other myeloid neoplasm
		3. Not meeting WHO criteria for CML, PV, PMF, MDS, or other myeloid neoplasm	3. Demonstration of *JAK2*V617F or other clonal marker *or* No evidence of reactive marrow fibrosis
		4. Demonstration of *JAK2*V617F or other clonal marker *or* No evidence of reactive thrombocytosis	
Minor criteria	1. BM trilineage myeloproliferation 2. Subnormal serum Epo level 3. EEC growth		1. Leukoerythroblastosis 2. Increased serum LDH 3. Anemia 4. Palpable splenomegaly

*Diagnosis of polycythemia vera (PV) requires meeting either both major criteria and one minor criterion or the first major criterion and two minor criteria. Diagnosis of essential thrombocythemia requires meeting all four major criteria. Diagnosis of primary myelofibrosis (PMF) requires meeting all three major criteria and two minor criteria.

†Small to large megakaryocytes with an aberrant nuclear-to-cytoplasmic ratio and hyperchromatic and irregularly folded nuclei and dense clustering.

CML = chronic myelogenous leukemia; EEC = endogenous erythroid colony; Epo = erythropoietin; Hct = hematocrit; Hb = hemoglobin; LDH = lactate dehydrogenase; MDS = myelodysplastic syndrome; RBC = red blood cell; WHO =World Health Organization.

With permission from Patnaik MM, Tefferi A. The complete evaluation of erythrocytosis: congenital and acquired. *Leukemia*. 2009;23:834-844.

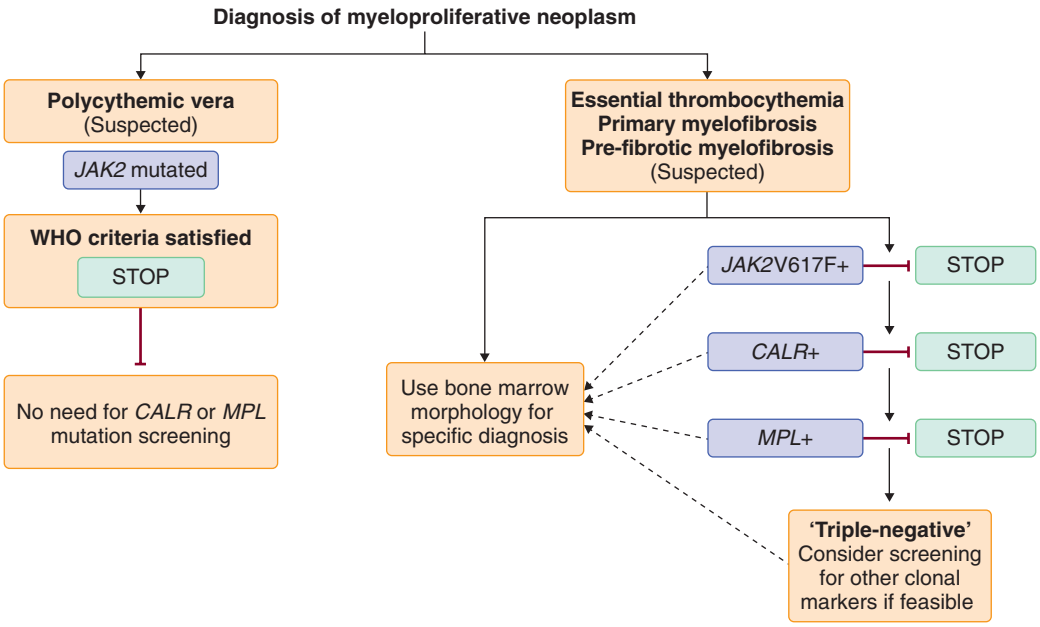

FIGURE 166-2. A diagnostic algorithm for polycythemia vera (PV), essential thrombocythemia (ET), primary myelofibrosis (PMF), and chronic myelogenous leukemia (CML) (With permission from Tefferi A, Pardanani A. Genetics: CALR mutations and a new diagnostic algorithm for MPN. *Nat Rev Clin Oncol*. 2014;11:125-126.)

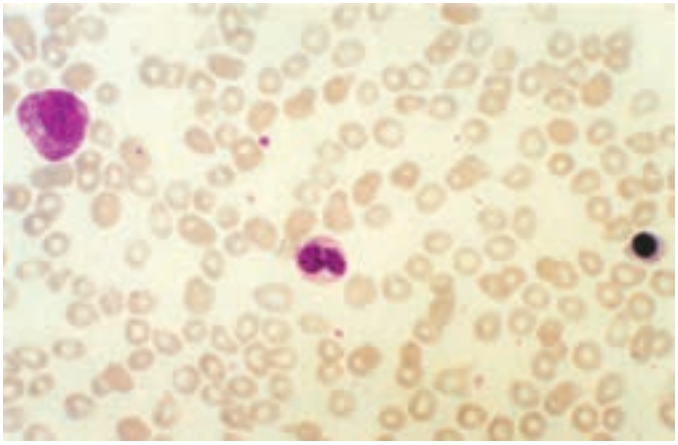

FIGURE 166-3. Myeloproliferative neoplasm. A peripheral blood smear from a patient with agnogenic myeloid metaplasia shows a leukoerythroblastic picture. The characteristic findings are teardrop-shaped red blood cells (dacryocytes), nucleated red blood cells (erythroblasts), and immature granulocyte precursors.

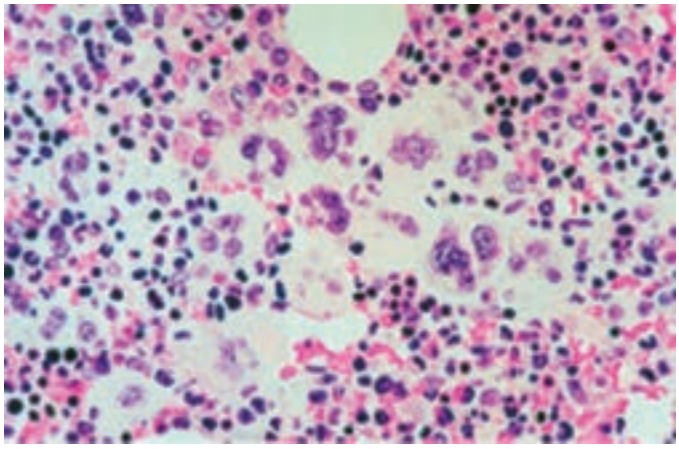

FIGURE 166-5. Myeloproliferative neoplasm. Bone marrow shows megakaryocytic clusters seen in essential thrombocythemia and other conditions associated with clonal thrombocytosis. These are not typically found in the bone marrow of individuals with other (secondary or reactive) types of thrombocytosis.

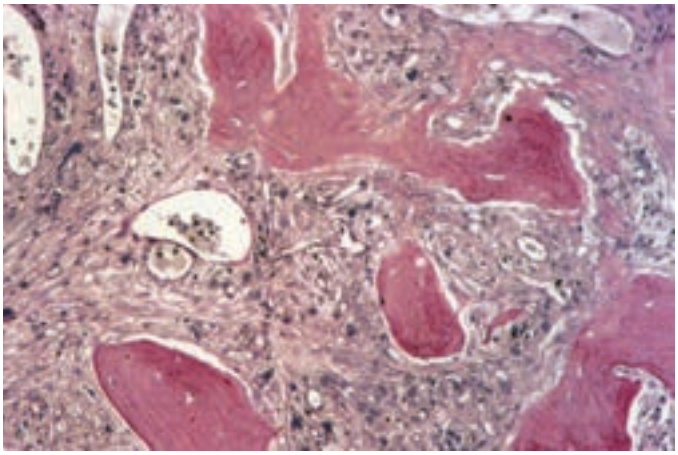

FIGURE 166-4. A bone marrow biopsy specimen from a patient with primary myelofibrosis shows reticulin fibrosis, osteosclerosis, and intrasinusoidal hematopoiesis.

TREATMENT

Essential Thrombocythemia and Polycythemia Vera

Drug therapy in ET or PV has not been shown to either improve survival or prevent disease transformation into post-ET or post-PV MF or blast-phase MPN. Instead, the main objective of specific therapy in ET or PV is either to prevent thrombosis in high-risk patients (i.e., age 60 years or older or presence of thrombosis history) or alleviate non–life-threatening symptoms, including microvascular disturbances (e.g., headaches, acral paresthesia, erythromelalgia), pruritus, or symptomatic splenomegaly. Microvascular symptoms are usually effectively treated with low-dose aspirin (81 mg/day). The cause of MPN-associated pruritus is poorly understood, but its dramatic response to JAK inhibitor therapy (see later discussion) suggests a causal relationship to cytokines that use JAK-STAT signaling. Other therapies for MPN-associated pruritus include antihistamines, selective serotonin reuptake inhibitors, interferon- (INF-α), and phototherapy. Symptomatic splenomegaly is usually managed by treatment with hydroxyurea (starting dose, 500 mg orally twice a day).

Observation alone is acceptable in asymptomatic young ET patients without a history of thrombosis[6]; the presence of microvascular symptoms,

TABLE 166-5 CLASSIFICATION OF ERYTHROCYTOSIS

1. Congenital erythrocytosis
   a. Associated with reduced P50 (partial pressure of oxygen at which 50% of hemoglobin is saturated with oxygen)
      i. High oxygen affinity hemoglobinopathy (usually autosomal dominant)
      ii. 2,3-Bisphosphoglycerate deficiency (usually autosomal recessive)
      iii. Methemoglobinemia
   b. Associated with normal P50
      i. *VHL* mutations, including Chuvash polycythemia (usually autosomal recessive)
      ii. *PHD2* mutations
      iii. *HIF2-α* mutations
      iv. *EPOR* mutations (usually autosomal dominant)
2. Acquired erythrocytosis
   a. Clonal (polycythemia vera)
   b. Secondary
      i. Hypoxia driven
         (1) Chronic lung disease
         (2) Right-to-left cardiopulmonary shunts
         (3) High-altitude habitat
         (4) Tobacco use or carbon monoxide poisoning
         (5) Sleep apnea or hypoventilation syndrome
         (6) Renal artery stenosis
      ii. Hypoxia independent
         (1) Use of androgen preparations or erythropoietin injection
         (2) After renal transplant
         (3) Cerebellar hemangioblastoma or meningioma
         (4) Pheochromocytoma, uterine leiomyoma, renal cysts, or parathyroid adenoma
         (5) Hepatocellular carcinoma or renal cell carcinoma

With permission from Patnaik MM, Tefferi A. The complete evaluation of erythrocytosis: congenital and acquired. *Leukemia*. 2009;23:834-844.

TABLE 166-6 CAUSES OF THROMBOCYTOSIS IN UNSELECTED COHORTS OF CONSECUTIVE PATIENTS

	Platelet Count (Approximate % of Patients)	
CONDITION	**ADULTS, >500,000/ML**	**>1 MILLION/ML**
Infection	22	31
Rebound thrombocytosis	19	3
Tissue damage (surgery)	18	14
Chronic inflammation	13	9
Malignancy	6	14
Renal disorders	5	<1
Hemolytic anemia	4	<1
Postsplenectomy status	2	19
Blood loss	NS	6
Essential (primary) thrombocythemia	3	14

NS = not stated.
From Tefferi A, Gilliland DG. Classification of chronic myeloid disorders: from Dameshek towards a semi-molecular system. *Best Pract Res Clin Haematol.* 2006;19:365-385.

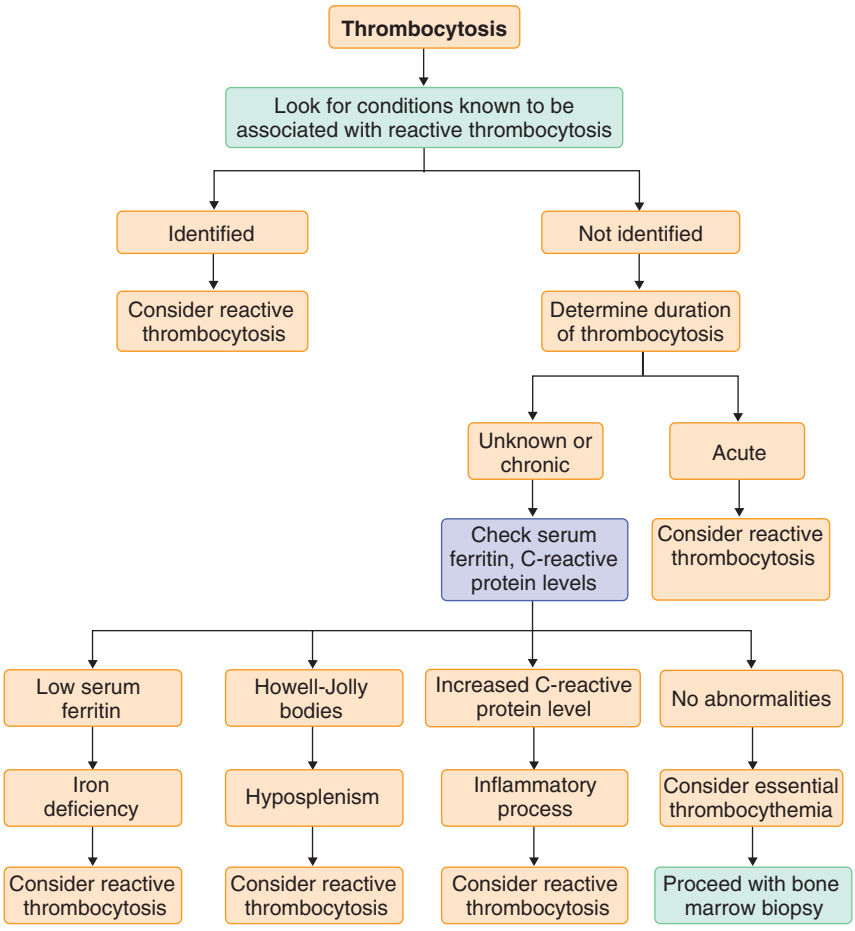

FIGURE 166-6. Diagnostic evaluation of thrombocytosis in routine clinical practice.

cardiovascular risk factors, or the *JAK2*V617F mutation justifies the use of once-daily aspirin therapy. Cytoreductive therapy for the prevention of thrombosis is indicated only in the presence of high-risk disease (Table 166-8). Randomized studies have demonstrated the value of low-dose aspirin (40-100 mg/day) in PV and hydroxyurea (starting dose, 500 mg orally twice a day) or anagrelide (starting at 1 mg daily) in high-risk patients with ET. Based on noncontrolled but prospective evidence, most experts agree that low-dose aspirin therapy (81 mg/day) should also be considered for ET and hydroxyurea therapy for high-risk PV (see Table 166-8). In addition, treatment with phlebotomy is required for all patients with PV, and the target hematocrit count in aspirin-treated patients should be less than 45% based on the most recent information from a controlled study.

In high-risk patients with ET in whom cytoreductive therapy is indicated, a platelet target of less than 400×10^9/L is reasonable and supported by data from retrospective studies. Hydroxyurea remains the first drug of choice for both high-risk ET and PV. High-risk PV or ET patients who are either intolerant or resistant to hydroxyurea are effectively managed by INF-α or busulfan.

Enthusiasm for the use of recombinant IFN in the MPNs was dampened by drug toxicity that often led to its discontinuation. The addition of a polyethylene glycol (PEG) moiety to IFN-α results in a longer half-life, allowing less frequent administration, more drug stability, less immunogenicity, and less toxicity, leading to the increased use of PEG-IFN-α instead of recombinant IFN.

Among the two second-line drugs, this author prefers the use of INF-α for patients younger than age 65 years and busulfan in the older age group. Because the long-term health effects of INF-α and impact on survival and disease complications are unknown, a controlled study is needed before INF-α is recommended for first-line therapy in either PV or ET. Busulfan is started at 4 mg/day orally and withheld in the presence of platelets below 100×10^9/L or white blood cell count below 3×10^9/L, and the dose is reduced to 2 mg/day if the corresponding levels are less than 150×10^9/L and less than 5×10^9/L.

There is unsubstantiated fear among primary caregivers regarding drug leukemogenicity with use of hydroxyurea or busulfan; there are no controlled studies in either PV or ET that shows this to be the case. A recent large retrospective study in PV, involving more than 1500 patients showed no evidence for leukemogenicity attached to either hydroxyurea or busulfan. Pregnancy in ET is associated with increased risk (~35% vs. approximately 15% in the control population) of first trimester spontaneous abortion. There is no association between the increased risk of spontaneous abortion and the degree of thrombocytosis. Low-risk pregnant patients with ET or PV are managed the same way as their nonpregnant counterparts, and the use of low-dose aspirin has been associated with a decreased prevalence of first trimester miscarriages in retrospective studies. High-risk pregnant women with ET (i.e., women with previous thrombosis) require cytoreductive therapy just like other high-risk patients. In such patients, there is anecdotal evidence on the safety of using IFN-α.

Major bleeding occurs in fewer than 10% of ET patients. Extreme thrombocytosis (i.e., platelet count >1 million/μL) appears to be a risk factor for

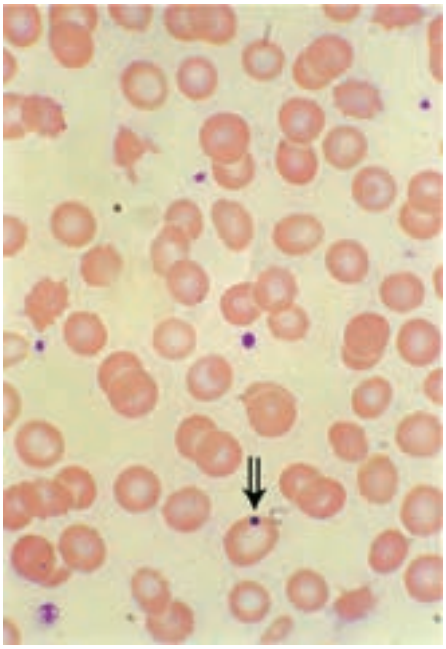

FIGURE 166-7. Peripheral blood smear showing Howell-Jolly bodies *(arrow)* in red blood cells, a finding typical of surgical or functional hyposplenism.

TABLE 166-7 CAUSES OF BONE MARROW FIBROSIS

MYELOID DISORDERS
Primary myelofibrosis
Metastatic cancer
Chronic myeloid leukemia
Myelodysplastic syndrome
Atypical myeloid disorder
Acute megakaryocytic leukemia
Other acute myeloid leukemias
Gray platelet syndrome
LYMPHOID DISORDERS
Hairy cell leukemia
Multiple myeloma
Lymphoma
NONHEMATOLOGIC DISORDERS
Connective tissue disorder
Infections (tuberculosis, kala-azar)
Vitamin D deficiency rickets
Renal osteodystrophy

TABLE 166-8 RISK STRATIFICATION AND RISK-ADAPTED THERAPY IN ESSENTIAL THROMBOCYTHEMIA, POLYCYTHEMIA VERA, AND PRIMARY MYELOFIBROSIS

RISK GROUPS: PV AND ET	MANAGEMENT: ET	MANAGEMENT: PV	MANAGEMENT: PMF	IPSS RISK GROUPS: PMF
Low risk (age <60 yr *and* no thrombosis history)	Low-dose aspirin	Low-dose aspirin + phlebotomy[§]	Observation	Low-risk (no risk factors[¶])
Low risk with extreme thrombocytosis[*]	Low-dose aspirin[†]	Low-dose aspirin[†] + phlebotomy	Conventional management[‖]	Intermediate-1 risk (one risk factor[¶])
High-risk (age ≥60 yr *or* thrombosis history)	Low-dose aspirin + hydroxyurea[‡]	Low-dose aspirin + phlebotomy + hydroxyurea[‡]	Allo-SCT if age <65 yr or experimental therapy Allo-SCT if age <65 years or experimental therapy Allo-SCT if age <65 yr or experimental therapy	Intermediate-1 risk with transfusions or unfavorable karyotype Intermediate-2 risk (2 risk factors[¶]) High-risk (≥3 risk factors[¶])

*Extreme thrombocytosis is defined as a platelet count of $>1000 \times 10^9$/L.
†Clinically significant acquired von Willebrand syndrome (ristocetin co-factor activity <30%) should be excluded before the use of aspirin in patients with extreme thrombocytosis because of bleeding risk.
‡In hydroxyurea-intolerant or -resistant patients, interferon-α (age < 60 years) or busulfan or pipobroman (age older than 60 years) might be used.
§In the presence of aspirin therapy, the hematocrit target can range between 38% and 50% and is set at a level that maintains best performance status.
‖Androgen preparations or thalidomide with prednisone for anemia; hydroxyurea for symptomatic splenomegaly.
¶The International Prognostic Scoring System (IPSS) uses five risk factors for inferior survival: age older than 65 years, hemoglobin <10 g/dL, leukocyte count $>25 \times 10^9$/L, circulating blasts ≥1%, and presence of constitutional symptoms.
Allo-SCT = allogeneic stem cell transplantation; ET = essential thrombocythemia; PMF = primary myelofibrosis; PV = polycythemia vera.

bleeding, in part because of the acquired von Willebrand disease syndrome. Measuring ristocetin cofactor activity in patients with extreme thrombocytosis and holding the use of aspirin therapy if the value is less than 30% have been recommended.

Primary Myelofibrosis

Conventional therapy for PMF is largely palliative and has not been shown to improve survival. Anemia and symptomatic splenomegaly are the main indications for treatment in PMF. Conventional drugs used for the treatment of PMF-associated anemia include androgen preparations (e.g., oral fluoxymesterone 10 mg two times a day), danazol (400-600 mg/day orally), prednisone (30-40 mg/day), and erythropoiesis-stimulating agents (ESAs).[7] The use of ESAs may exacerbate PMF-associated splenomegaly, and it is ineffective in transfusion-dependent patients. Prostate cancer screening in men is necessary when considering treatment with androgen preparations. Response rates to prednisone, androgen preparations, or danazol are in the vicinity of 20% and response durations average about one to two years.

Recent studies have shown the value of thalidomide and thalidomide-like drugs such as lenalidomide in the treatment of anemia in PMF. Thalidomide as a single agent (50-200 mg/day orally) is not as effective as the drug combined with prednisone (30 mg/day) where reported response rates range from 20% to 62% for anemia, 25% to 80% for thrombocytopenia, and 7% to 30% for splenomegaly. Response rates with lenalidomide are somewhat similar, but the drug works best in the presence of del(5q) cytogenetic abnormality (see Chapter 182 on myelodysplastic syndrome.) The occurrence of peripheral neuropathy with thalidomide and severe pancytopenia with lenalidomide therapy has limited their use in most patients with the disease.

The drug of choice for symptomatic splenomegaly in PMF is hydroxyurea (starting dose, 500 mg orally twice a day). Hydroxyurea-refractory cases have been often managed by splenectomy in the past since the value of other conventional drugs in this regard has been limited (see section below on experimental drug therapy). Indications for splenectomy in PMF have included mechanical discomfort, symptomatic portal hypertension (i.e., associated with ascites or variceal bleeding), and frequent RBC transfusions. The perioperative mortality rate of splenectomy in PMF is between 5% and 10%. Postsplenectomy complications occur in approximately 50% of the patients and include bleeding, thrombosis, hepatomegaly, extreme thrombocytosis, leukocytosis, and an increase in circulating blasts. Prophylactic therapy with hydroxyurea has been advised to prevent postsplenectomy thrombocytosis. Transjugular intrahepatic portosystemic shunt might be considered to alleviate symptoms of portal hypertension.

Extramedullary hematopoiesis is the main cause of hepatosplenomegaly in PMF and post-PV or post-ET MF. Nonhepatosplenic EMH also occurs in PMF and might involve the vertebral column (spinal cord compression) lymph nodes, lung (pulmonary hypertension), pleura (effusion), small bowel, peritoneum (ascites), urogenital tract, skin, or heart. Diagnosis is usually tissue based, but imaging studies are sometimes used; MF-associated pulmonary hypertension can be confirmed by a technetium-99m sulphur colloid scintigraphy, which shows diffuse pulmonary uptake.

Radiotherapy has been a treatment option in both hepatosplenic and nonhepatosplenic EMH. Splenic irradiation (100-500 cGy in 5-10 fractions) induces transient reduction in spleen size but might be associated with life-threatening pancytopenia. More recent studies suggest the relative safety and value of lower dose treatment (100 cGy total dose in four daily fractions of 25 cGy). Radiation therapy works best for nonhepatosplenic EMH. Single-fraction (100-400 cGy) involved field therapy has also been shown to benefit patients with MF-associated pulmonary hypertension or extremity pain.

Quality of life–relevant symptoms in MF include profound fatigue, night sweats, weight loss, pruritus, bone pain, and nonproductive cough. Conventional drug therapy is inadequate in the management of these symptoms. Recent studies of JAK inhibitor therapy in PMF or post-PV, or post-ET MF suggest major benefit in terms of constitutional symptoms and cachexia.

Experimental Drug Therapy in Myeloproliferative Neoplasms

JAK-inhibiting ATP mimetics currently represent the most popular investigational drug therapy in MF.[8] Ruxolitinib is now approved by the Food and Drug Administration for use in MF, and momelotinib is currently being compared with ruxolitinib in a randomized phase 3 study. Ruxolitinib is a JAK1/JAK2 inhibitor. Two randomized studies comparing ruxolitinib with either placebo or best supportive care have now been published. In the COMFORT-1 trial that compared the drug with placebo, the spleen size reduction response rate was approximately 42% for ruxolitinib versus less than 1% for placebo. In addition, about 46% of patients experienced substantial improvement in their constitutional symptoms, such as night sweats, pruritus, and fatigue. However, the use of the drug versus placebo was associated with anemia (31% vs. 13.9%) and thrombocytopenia (34.2% vs. 9.3%). The COMFORT-2 trial compared ruxolitinib with "best available therapy." The spleen size reduction response was 28.5% with ruxolitinib versus 0% otherwise, and the use of the drug was associated with thrombocytopenia (44.5% vs. 9.6%), anemia (40.4% vs. 12.3%), and diarrhea (24.0% vs. 11.0%). The long-term outcome of ruxolitinib is associated with high treatment discontinuation rate and the occurrence of severe withdrawal symptoms during ruxolitinib treatment discontinuation ("ruxolitinib withdrawal syndrome"), characterized by acute relapse of disease symptoms, accelerated splenomegaly, worsening of cytopenias, and occasional hemodynamic decompensation. The 3-year follow-up information on COMFORT-2 suggested a 55% drug discontinuation rate and a slight but significant improvement in survival without any evidence of drug effect on *JAK2*V617F allele burden or bone marrow fibrosis. Furthermore, reports have now associated ruxolitinib therapy with serious opportunistic infections.

Momelotinib (MMB, GS-0387, CYT387) is a JAK1 and JAK2 inhibitor. Among the first 60 patients treated in a phase 1 and 2 study,[9] Dose-limiting toxicities included grade 3 headache and hyperlipasemia (elevated lipase). Anemia and spleen responses were 59% and 48%, respectively. Most patients experienced constitutional symptoms improvement. The drug is currently undergoing phase 3 study and being compared with ruxolitinib.There is clearly a need to evaluate more drugs before making any conclusions regarding the value of anti-JAK2 therapy in MF or related MPN. It is also becoming evident that some of the salutary effects of these drugs might be the result of a potent anticytokine activity.

Allogeneic Stem Cell Transplantation

Allogenic stem cell transplant is a reasonable treatment modality in PMF, especially in the presence of high-risk disease.[10] However, there is a nontrivial risk of treatment-associated mortality and morbidity. In one study, for example, 5-year disease-free survival and treatment-related mortality rates were 33% and 35% for matched related and 27% and 50% for unrelated transplants, respectively. The respective chronic graft-versus-host disease and relapse rates for matched related transplants were 40% and 32%, respectively. Outcome did not appear to be favorably affected by reduced intensity conditioning (RIC). In another RIC transplant study, the 5-year disease-free survival rate was estimated at 51%, chronic graft-versus-host disease at 49%, and relapse at 29%. Splenectomy before transplant might not affect outcome. There is currently much interest in evaluating the use of JAK inhibitors before transplant with favorable and unfavorable experiences reported by different investigators, warranting the need for more studies before recommending such a strategy.

PROGNOSIS

Essential Thrombocythemia and Polycythemia Vera

Neither ET nor PV is currently a curable disease, and current therapy has not been shown to prolong survival. In a study of 1545 patients with PV, survival was adversely affected by older age, leukocytosis, abnormal karyotype, and venous thrombosis.[11] The first three also predicted leukemic transformation, which was 2.3% at 10 years and 5.5% at 15 years. Leukemic transformation was associated with treatment exposure to pipobroman or P32/chlorambucil but not to hydroxyurea or busulfan.

In WHO-defined ET, an international study identified age 60 years or older, thrombosis history, cardiovascular risk factors (CVR), leukocyte count greater than $11 \times 10(9)/L$, and the presence of *JAK2*V617F as independent predictors of arterial thrombosis and male gender as a predictor of venous thrombosis.[12] In a subsequent prognostic model,[13] thrombosis risk was lowest in the absence of these risk factors and intermediate or high otherwise. In ET, *CALR* mutations have been correlated with male sex, younger age, lower leukocyte count, lower hemoglobin level, higher platelet count, and longer thrombosis-free survival time.[14] Fibrotic or leukemic transformation in ET or PV is relatively infrequent (a combined rate of <10% in the first 15 years of disease).

Current risk stratification in PV and ET is designed to estimate the risk of thrombosis (see Table 166-8). Age 60 years or older and history of thrombosis are the two key risk factors used to classify patients with PV or ET into low- (no risk factors) and high- (one or two risk factors) risk groups. In addition, because of the potential risk for bleeding, low-risk patients with extreme thrombocytosis (platelet count $>1000 \times 10^9/L$) are considered separately. With the recent discovery of *CALR* mutations and their association with lower risk of thrombosis, the thrombogenic effect of *JAK2*V617F is increasingly being realized. Accordingly, more recent studies have identified the presence of *JAK2*V617F or cardiovascular risk factors as additional risk factors for thrombosis in ET.

Primary Myelofibrosis

The most robust prognostic model for PMF is the Dynamic International Prognostic Scoring System (DIPSS)-plus, which relies on eight adverse

parameters: age older than 65 years, hemoglobin less than 10 g/dL, leukocyte count greater than 25×10^9/L, circulating blasts 1% or greater, constitutional symptoms, unfavorable karyotype, RBC transfusion need, and platelet count less than 100×10^9/L. The four DIPSS-plus risk categories based on these risk factors are low (no risk factors), intermediate-1 (one risk factor), intermediate-2 (two or three risk factors), and high (four or more risk factors), with respective median survival periods of 15.4, 6.5, 2.9, and 1.3 years. More recently, *ASXL1* mutations have been identified as a DIPSS-plus independent risk factor. Furthermore, *CALR* mutations in PMF were associated with younger age; higher platelet count; and lower incidences of anemia, leukocytosis, and spliceosome mutations and longer overall survival time. The best survival in PMK was found in the presence of *CALR* and absence of *ASXL1* mutation (i.e., $CALR^{+}ASXL1^{-}$) and worst survival in patients with $CALR^{-}ASXL1^{+}$ mutational status.[15]

CONCLUDING COMMENTS

Despite the seminal discoveries of *JAK2, MPL,* and *CALR* mutations, the pathogenesis of *BCR-ABL1*–negative MPN remains complex and poorly understood. It is becoming increasingly evident that *JAK2* is not the whole story with these diseases, and we should therefore curb our expectations from anti-JAK2 treatment strategies and instead pay attention to additional pathogenetic insight from correlative laboratory studies. In the end, one must recognize the relatively indolent natural history of ET and PV and avoid unnecessary treatment, especially in low-risk patients. There is currently an unmet need for treatment in PMF, and this is where the real challenge is and where progress is needed. It is hoped that that next-generation sequencing will help identify additional new mutations and prognostically relevant molecular signatures in MPN.

Grade A References

A1. Squizzato A, Romualdi E, Passamonti F, et al. Antiplatelet drugs for polycythaemia vera and essential thrombocythaemia. *Cochrane Database Syst Rev*. 2013;4:CD006503.

A2. Harrison CN, Campbell PJ, Buck G, et al. Hydroxyurea compared with anagrelide in high-risk essential thrombocythemia. *N Engl J Med*. 2005;353:33-45.

A3. Gisslinger H, Gotic M, Holowiecki J, et al. Anagrelide compared with hydroxyurea in WHO-classified essential thrombocythemia: the ANAHYDRET Study, a randomized controlled trial. *Blood*. 2013;121:1720-1728.

A4. Marchioli R, Finazzi G, Specchia G, et al. Cardiovascular events and intensity of treatment in polycythemia vera. *N Engl J Med*. 2013;368:22-33.

A5. Verstovsek S, Mesa RA, Gotlib J, et al. A double-blind, placebo-controlled trial of ruxolitinib for myelofibrosis. *N Engl J Med*. 2012;366:799-807.

A6. Harrison C, Kiladjian JJ, Al-Ali HK, et al. JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis. *N Engl J Med*. 2012;366:787-798.

A7. Cervantes F, Vannucchi AM, Kiladjian JJ, et al. Three-year efficacy, safety, and survival findings from COMFORT-II, a phase 3 study comparing ruxolitinib with best available therapy for myelofibrosis. *Blood*. 2013;122:4047-4053.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Titmarsh GJ, Duncombe AS, McMullin MF, et al. How common are myeloproliferative neoplasms? A systematic review and meta-analysis. *Am J Hematol*. 2014;89:581-587.
2. Klampfl T, Gisslinger H, Harutyunyan AS, et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med*. 2013;369:2379-2390.
3. Tefferi A, Pardanani A. Genetics: CALR mutations and a new diagnostic algorithm for MPN. *Nat Rev Clin Oncol*. 2014;11:125-126.
4. Keohane C, McMullin MF, Harrison C. The diagnosis and management of erythrocytosis. *BMJ*. 2013;347:f6667.
5. Spivak JL, Considine M, Williams DM, et al. Two clinical phenotypes in polycythemia vera. *N Engl J Med*. 2014;371:808-817.
6. Tefferi A, Barbui T. Personalized management of essential thrombocythemia-application of recent evidence to clinical practice. *Leukemia*. 2013;27:1617-1620.
7. Tefferi A. Primary myelofibrosis: 2014 update on diagnosis, risk-stratification, and management. *Am J Hematol*. 2014;89:915-925.
8. Rosenthal A, Mesa RA. Janus kinase inhibitors for the treatment of myeloproliferative neoplasms. *Expert Opin Pharmacother*. 2014;15:1265-1276.
9. Pardanani A, Laborde RR, Lasho TL, et al. Safety and efficacy of CYT387, a JAK1 and JAK2 inhibitor, in myelofibrosis. *Leukemia*. 2013;27:1322-1327.
10. Salit RB, Deeg HJ. Role of hematopoietic stem cell transplantation in patients with myeloproliferative disease. *Hematol Oncol Clin North Am*. 2014;28:1023-1035.
11. Tefferi A, Rumi E, Finazzi G, et al. Survival and prognosis among 1545 patients with contemporary polycythemia vera: an international study. *Leukemia*. 2013;27:1874-1881.
12. Carobbio A, Thiele J, Passamonti F, et al. Risk factors for arterial and venous thrombosis in WHO-defined essential thrombocythemia: an international study of 891 patients. *Blood*. 2011;117:5857-5859.
13. Barbui T, Finazzi G, Carobbio A, et al. Development and validation of an International Prognostic Score of thrombosis in World Health Organization-essential thrombocythemia (IPSET-thrombosis). *Blood*. 2012;120:5128-5133.
14. Rumi E, Pietra D, Ferretti V, et al. JAK2 or CALR mutation status defines subtypes of essential thrombocythemia with substantially different clinical course and outcomes. *Blood*. 2014;123:1544-1551.
15. Tefferi A, Guglielmelli P, Lasho TL, et al. CALR and ASXL1 mutations-based molecular prognostication in primary myelofibrosis: an international study of 570 patients. *Leukemia*. 2014;28:1494-1500.

REVIEW QUESTIONS

1. A 65-year-old man with coronary artery disease, hypertension, chronic obstructive pulmonary disease, and an active 40 pack year smoking history is referred by his primary care physician because of gradually increasing hemoglobin and hematocrit over the past 5 years. On physical examination, his spleen is not palpable. His blood counts now are hemoglobin, 18.9 g/dL; hematocrit, 60.5%; MCV = 78; white blood cell count, 11,200/μL (normal differential); and platelets, 485,000/μL. You are asked to evaluate the cause of this patient's polycythemia. Which of the following tests is most likely to either make or rule out the diagnosis of polycythemia vera in this case?

A. Carboxyhemoglobin level
B. Abdominal ultrasonography findings for spleen size
C. PCR for *JAK2-V617F*
D. Bone marrow aspirate and biopsy with cytogenetics
E. Red blood cell mass

Answer: C Although it is not specific for polycythemia vera (PV), almost all patients with PV carry a *JAK2* mutation, and its absence makes the diagnosis questionable. (The *JAK2-V617F* mutation also occurs in about 60% of patients with essential thrombocythemia and 50% to 60% with primary myelofibrosis.) Differentiating PV from other secondary causes of polycythemia (see Table 166-5) can be challenging if the individual is negative for a *JAK2* mutation. Bone marrow aspirate and biopsy with cytogenetics can sometimes demonstrate characteristic findings of an myeloproliferative neoplasms, including trilineage hyperplasia (as opposed to just erythroid hyperplasia, the latter being what would be expected for the secondary polycythemias), as well as morphologic changes, but these are rarely absolutely diagnostic; furthermore, cytogenetic studies usually do not add to making or ruling out the diagnosis of PV. A red blood cell (RBC) mass determination is almost superfluous in this case because the great majority of patients with this degree of erythrocytosis, regardless of cause, will have elevated RBC masses. The finding of splenomegaly is neither specific nor sensitive for a diagnosis of PV, whether it is assessed by physical examination or by imaging. It would not be surprising to find an elevated level of carboxyhemoglobin in this smoker, but that would not be a conclusive finding for the diagnosis of his polycythemia. Bot mentioned among the choices, a serum erythropoietin would be a helpful ancillary test in a case like this; in general, it would be expected to be elevated in virtually all of the polycythemias but low in PV (the latter because of feedback inhibition of erythropoietin by the kidneys as a result of the primary myeloproliferative process of PV).

2. Which of the following is *not* a known complication of polycythemia vera (PV) and essential thrombocythemia (ET)?

A. Budd-Chiari syndrome
B. Aplastic anemia
C. Erythromelalgia
D. Marrow fibrosis
E. Acute leukemia

Answer: B Thrombosis is the major cause of morbidity and mortality in the myeloproliferative neoplasms, and it often presents in unusual intraabdominal sites such as hepatic vein thrombosis (Budd-Chiari syndrome) and portal and mesenteric vein thrombosis. Erythromelalgia is a striking and characteristic (but not specific) syndrome of microvascular vasomotor and occlusive disease, mostly affecting the dependent parts of the extremities, typically the feet and sometimes the hands. It is characterized by excruciating pain and tenderness in a patchy distribution, sometimes accompanied by corresponding patches of erythema and warmth. The large vessel peripheral pulses are usually intact. Erythromelalgia is generally very sensitive to prompt response to aspirin. Both marrow fibrosis and spontaneous transformation to acute leukemia are unusual but well-described feared complications of PV and ET. However, aplastic anemia is not associated with these diseases (in contrast to its association with a different bone marrow stem cell disease, paroxysmal nocturnal hemoglobinuria.)

3. Massive, painful splenomegaly develops in a 70-year-old woman with an established diagnosis of primary myelofibrosis (PMF). She is *JAK2-V617F* mutation negative but *CALR* mutation positive. She has become transfusion dependent and has been unresponsive to medical management of the splenomegaly. After much consideration of risks and benefits, it is decided to do a splenectomy. Which of the following are you most likely to find in the spleen on pathologic examination in this case?

A. Megakaryocytes
B. Fibrosis
C. Hemophagocytosis
D. Clonal lymphocytic infiltration
E. Storage cells

Answer: A A major cause of this PMF patient's massive splenomegaly is likely to be extensive "extramedullary hematopoiesis" (EMH) within the spleen. EMH occurs in myeloid metaplasia with or without myelofibrosis. It represents essentially "ectopic" sites of renew hematopoiesis, recapitulating the normal fetal state where hematopoiesis was not restricted to bone marrow but also occurred in other organs such as the spleen and liver. Therefore, histopathology of this patient's spleen is likely to demonstrate many blood cell precursors, including megakaryocytes. Fibrosis, hemophagocytosis, and infiltration with storage cells are not characteristic findings in PMF (or any of the other myeloproliferative neoplasms [MPNs]). Clonal lymphocytic infiltration can cause splenomegaly in the lymphomas but not in the MPNs, which are clonal myeloproliferative, not lymphoproliferative, disorders.

4. A 72-year-old woman has been followed by her hematologist for more than 15 years with the diagnosis of polycythemia vera (PV) (recently confirmed by the finding of the *JAK2-V617F* mutation). Her only treatments during this time have been periodic phlebotomies as needed to maintain a hematocrit below 45%, varying doses of hydroxyurea, and one aspirin daily. Her blood counts have been well controlled, and she has been largely asymptomatic. Over the past several months, despite lowering the hydroxyurea dose from 2 g/day to 1 g/day to 500 mg/day and now to 500 mg on alternate days, the patient has been noted to be actually anemic. Her only new symptoms are increasing fatigue. Her hematocrits have been ranging from 35% to 40% over this time. She reports no signs or symptoms of bleeding, her stools are negative for occult blood, and her serum ferritin level is within normal limits. What would you do now?

A. Carefully examine the peripheral blood smear.
B. Do evaluation for occult malignancy, including imaging of chest, abdomen, and pelvis.
C. Perform bone marrow aspirate and biopsy.
D. Check CALR (calreticulin gene) mutation status to rule out clonal evolution of disease.
E. Stop hydroxyurea completely and observe course of the anemia.

Answer: C The most likely diagnosis is that this patient's PV is now transforming to either myelofibrosis or acute leukemia (or myelodysplasia). This is the natural history of the disease in a small minority of patients with PV even in the absence of previous treatments with known leukemogenic agents (e.g., P32, alkylating agents). A peripheral blood smear is important and can provide important clues, including leukoerythroblastic or myelophthisic changes (e.g., teardrop-shaped red blood cells, nucleated red cells, and immature myeloid precursors) could indicate development of fibrosis; the presence of circulating blasts could indicate conversion to acute leukemia. However, whether or not one sees these changes on the blood smear, a bone marrow aspirate and biopsy are indicated as the definitive test. There is no clear evidence that patients with myeloproliferative neoplasms (MPNs) are predisposed to second solid tumor malignancies; therefore, other than conducting a thorough history and physical examination and then pursuing any leads found with further testing, general body imaging to look for occult malignancy is not indicated. Clonal evolution does occur in PV and the other MPNs, and new mutations can be acquired during the course of the disease. However, at this point, it appears that the *JAK2-V617F* and *CALR* mutations are mutually exclusive in the MPNs. Even if stopping the hydroxyurea slightly improves the anemia, the striking history of increasing sensitivity to hydroxyurea suggests that something unfavorable is happening in this patient's bone marrow.

167

LEUKOCYTOSIS AND LEUKOPENIA

NANCY BERLINER

The normal peripheral white blood cell count (WBC) ranges between 4500/μL and 10,000/μL, with a mean of 7500/μL, and is composed of neutrophils, lymphocytes, monocytes, basophils, and eosinophils. Because neutrophils usually represent about 60% of the peripheral WBC, derangement in the WBC usually reflects elevation or reduction in the absolute neutrophil count. Leukocytosis, an elevated WBC, and leukopenia, a depressed WBC, may be secondary to an underlying disease or exposure, or they may be manifestations of a primary hematologic disorder. This chapter outlines both the primary and secondary causes of leukocytosis and leukopenia, focusing particularly on neutrophilia and neutropenia.

NORMAL NEUTROPHIL DYNAMICS

Neutrophils arise from multipotent progenitor cells in the bone marrow that also give rise to erythrocytes, megakaryocytes, eosinophils, basophils, and monocytes. Neutrophil precursors in the marrow mature over 6 to 10 days to form a storage pool of mature neutrophils (Fig. 167-1). Together the marrow populations make up about 95% of the body's total granulocyte mass (20% neutrophil precursors, 75% mature bands and neutrophils). The circulating neutrophil pool thus represents only the remaining approximately 5% of the body's total neutrophils, just over half of which at any given time are adherent to the vascular endothelium and the spleen, a phenomenon termed *margination*. These marginated neutrophils are poised for immediate release into the circulation during times of stress. The remaining neutrophils circulate freely in the blood. The lifespan of neutrophils in the peripheral blood was thought to be very short, only 6 to 12 hours; however, newer in vivo studies suggest that they circulate for up to 3 to 4 days.[1] They subsequently migrate into tissues, where they can survive for 1 to 4 days. Changes in neutrophil number reflect these dynamics. Neutrophilia can occur as the result of increased marrow production, increased release of neutrophils from the storage pool, or mobilization of neutrophils from the marginated pool. Neutropenia, on the other hand, may be due to decreased marrow production, increased margination with or without sequestration by the spleen, or increased destruction of peripheral cells.

NEUTROPHILIA

Most cases of neutrophilia are reactive or secondary to an underlying inflammatory process. This includes neutrophilia due to infection, chronic inflammation, stress, drugs, nonhematologic malignancy, marrow stimulation (as in hemolysis or idiopathic thrombocytopenic purpura), or splenectomy. Primary causes of neutrophilia may be congenital, including hereditary neutrophilia, Down syndrome, and leukocyte adhesion deficiency (LAD), or acquired as in the case of chronic myelogenous leukemia and other myeloproliferative neoplasms (Table 167-1).

TABLE 167-1 DIFFERENTIAL DIAGNOSIS OF NEUTROPHILIA

- Primary hematologic etiologies
  - Congenital neutrophilia
    - Hereditary neutrophilia
    - Chronic idiopathic neutrophilia
    - Down syndrome
    - Leukocyte adhesion deficiency (LAD)
      - LAD-1
      - LAD-2
  - Acquired hematologic neoplasms
    - Myeloproliferative neoplasm
      - Chronic myelogenous leukemia
      - Polycythemia vera
- Secondary to other disease entities
  - Infection
    - Acute via release from marginated and storage pools
    - Chronic via increased myelopoiesis (e.g., tuberculosis, fungal infection, chronic abscess, other chronic infections)
  - Chronic inflammation
    - Rheumatic disease: juvenile rheumatoid arthritis, rheumatoid arthritis, Still disease, and others
    - Inflammatory bowel disease
    - Granulomatous disease
    - Chronic hepatitis
  - Cigarette smoking
  - Stress
  - Drug induced
    - Corticosteroids
    - β-Agonists
    - Lithium
    - Recombinant cytokine administration
  - Nonhematologic malignancy
    - Cytokine-secreting tumors (lung, tongue, kidney, urothelial tumors)
    - Marrow metastasis (myelophthisis)
  - Marrow stimulation
    - Hemolytic anemia, immune thrombocytopenia
    - Recovery from marrow suppression
    - Recombinant cytokine administration
  - Post-splenectomy

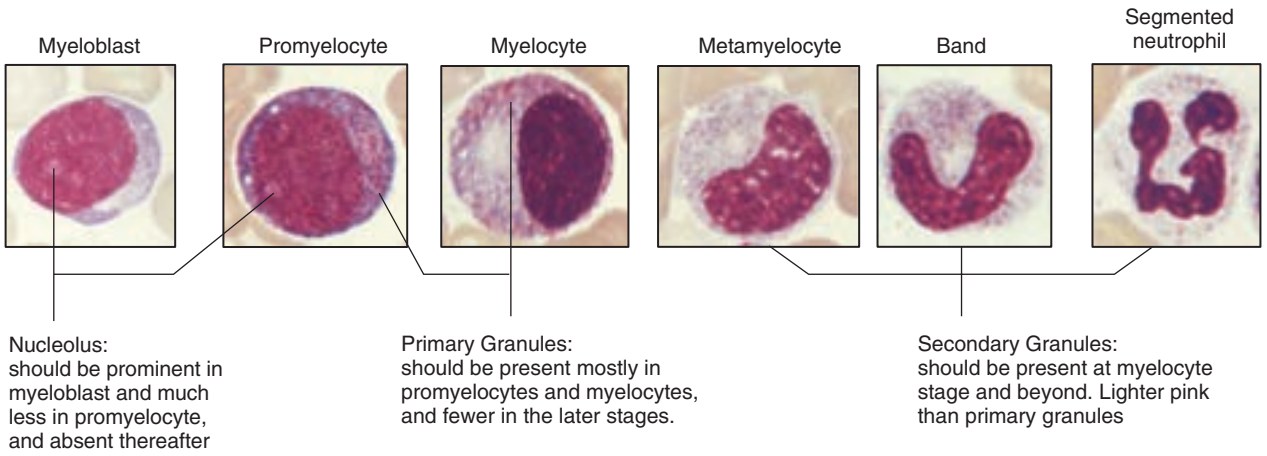

FIGURE 167-1. Myeloid maturation in the bone marrow. Nucleoli are prominent in myeloblasts, much less frequent in promyelocytes, and absent in more mature forms. Primary granules are present in the cytoplasm of promyelocytes and myelocytes, and secondary granules predominate beyond the myelocyte stage.

ETIOLOGY

Secondary Causes of Neutrophilia

Infection

Many acute bacterial infections can present with a modest leukocytosis with a "left shift," referring to the circulation of more immature myeloid cells. This left shift most commonly is restricted to release of an increased number of band forms; however, in severe stress, one may see circulating metamyelocytes and even earlier cells (see Fig. 167-1) in the peripheral blood. Leukocytosis occurs within minutes to hours of infection owing to release of neutrophils from both the marrow and marginated pools. Examination of these neutrophils on peripheral smear may reveal evidence of toxic granulation (see Fig. 157-18), Döhle bodies (see Fig. 157-19), and cytoplasmic vacuoles. Certain infections (e.g., *Clostridium difficile* or tuberculosis in particular) are known to cause elevations in the WBC to greater than 30,000/μL in about one fourth of infected patients and may result in a *leukemoid reaction*, defined as a WBC of greater than 50,000/μL with a pronounced left shift (Fig. 167-2).

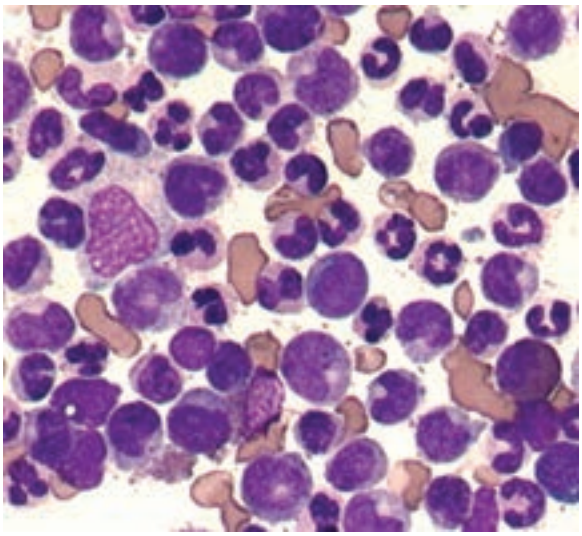

FIGURE 167-2. Peripheral blood from a patient with leukemoid reaction. From this smear, it would be impossible to distinguish this from chronic phase chronic myelogenous leukemia (Chapter 184). Distinction would depend on determination of presence or absence of *BCR-ABL* fusion.

Chronic Inflammation

Leukocytosis due to chronic inflammation results from increased leukocyte (specifically neutrophil and monocyte) production as opposed to altered neutrophil distribution. Mature neutrophil pools become depleted with ongoing inflammation, and the myeloid compartment of the marrow expands to increase neutrophil production. Myriad cytokines, including tumor necrosis factor-α (TNF-α), granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), macrophage inflammatory protein-1 (MIP-1), interleukin-1 (IL-1), IL-6, and IL-8, have been implicated in this marrow stimulation (Chapter 156). Chronic inflammatory conditions that are particularly associated with leukocytosis and neutrophilia include juvenile rheumatoid arthritis, rheumatoid arthritis, Still disease, Crohn disease, ulcerative colitis, granulomatous infection, and chronic hepatitis. The WBC and neutrophil elevation in these cases is typically more modest than that seen in acute infection or inflammation.

Cigarette Smoking

Cigarette smoking can cause a leukocytosis and neutrophilia in about 25 to 50% of chronic smokers that can persist for even up to 5 years after quitting smoking. The mechanism by which this occurs is unknown, although there is recent evidence that cigarette smoke slows neutrophil apoptosis.[2]

Stress

Within minutes of exercise, surgery, or stress, one can see an elevation in circulating neutrophils. This is presumed to be due to the effects of catecholamines on marginated neutrophils, with release of neutrophils into the circulation. Some cases of stress-induced neutrophilia can be prevented by pretreatment with β-adrenergic antagonists (e.g., propranolol), supporting the role of catecholamines in the process. Exercise-induced neutrophilia, however, is not blocked by propranolol, suggesting that it may instead be due to flow and mechanical perturbation of neutrophils in the lungs. An elevated WBC has also been noted in the setting of acute myocardial infarction, but whether this is a risk factor for cardiac ischemia or a result of inflammation is unclear.

Drug Induced

Probably the most well-known and widely used drugs associated with leukocytosis are corticosteroids. Other drugs that are associated with elevations in the neutrophil count include β-agonists and lithium. Lithium causes neutrophilia by increasing the production of endogenous colony-stimulating factors (CSFs). G-CSF or GM-CSF treatment likewise may result in neutrophilia, and although this is the desired effect, the neutrophilia can be quite pronounced if not appropriately monitored.

Nonhematologic Malignancy

Leukocytosis can be seen in a number of nonhematologic malignancies. Some tumors (lung, tongue, kidney, bladder) are thought to secrete G-CSF as an ectopic hematopoietic growth factor. Other tumors (lung, stomach, breast), when metastasized to the bone and bone marrow, can cause a *leukoerythroblastic reaction*, characterized by left-shifted leukocytosis, thrombocytosis, and red cell abnormalities including nucleated and teardrop-shaped red blood cells (Fig. 167-3). The presence of nonhematopoietic entities invading the bone marrow (metastatic cancer, fibrosis, granulomatous disease) is termed *myelophthisis*.

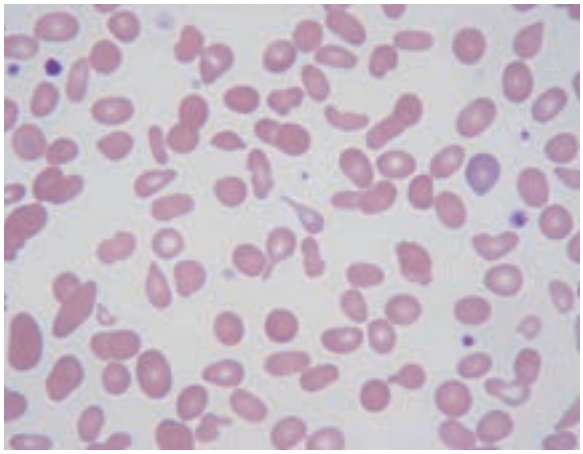

FIGURE 167-3. Myelophthisic changes in erythrocyte morphology. Note prominent teardrop forms. (From Rose M, Berliner N. Disorders of red blood cells. In: Andreoli TE, Benjamin IJ, Griggs RC, et al, eds. *Andreoli and Carpenter's Cecil Essentials of Medicine,* 8th ed. Philadelphia: Saunders; 2010:522, Fig. 49-2.)

Marrow Stimulation

Peripheral destruction of red cells and platelets, as seen with hemolytic anemia and idiopathic thrombocytopenic purpura, can result in stimulation of the bone marrow and a "spillover" leukocytosis. Recovery of cell counts following marrow suppression, as in the case of chemotherapy, can result in a rebound leukocytosis that may last several weeks.

Primary Causes of Neutrophilia

Hereditary Neutrophilia

Hereditary neutrophilia is an autosomal dominant genetic disease that is characterized by an elevated WBC in the 20,000 to 100,000/μL range with splenomegaly and widened diploe of the skull. The neutrophils in this disorder appear to function normally, and patients have no increased risk for bacterial infection or other sequelae. Hereditary neutrophilia caused by an autosomal-dominant *GCSF3* gene mutation has been reported, causing constitutive activation of the G-CSF receptor.

Chronic Idiopathic Neutrophilia

Chronic idiopathic neutrophilia is a condition marked by leukocytosis in the 11,000 to 40,000/μL range with a normal bone marrow biopsy. In one series with a 20-year follow-up, patients with this condition had no medical sequelae from this elevated WBC.

Pelger-Huët Anomaly

Patients with the Pelger-Huët anomaly (PHA) are often misdiagnosed as having a left-shifted WBC because many of their mature neutrophils are misinterpreted as band forms. Although these patients do not actually have leukocytosis, the anomaly often raises suspicion for an acute infection or inflammatory process because of this apparent left shift. PHA is due to a mutation in the lamin B receptor gene and manifests with mature neutrophils and condensed, clumped chromatin within a bilobed nucleus (see Fig. 157-20). These neutrophils, however, function normally. A number of drugs can reversibly induce pseudo-PHA, including colchicine, sulfonamides, ibuprofen, taxoids, and valproate. Pseudo-PHA is also seen in some patients with myelodysplasia (Chapter 182). Vitamin B_{12} or folate deficiency can cause increased nuclear lobation of neutrophils in patients with PHA, perhaps leading to a missed diagnosis. With correction of the vitamin deficiency, however, the aberrant neutrophil nuclear morphology returns.

Down Syndrome

Up to 10% of patients with Down syndrome develop transient myeloproliferative disorder (TMD) related to peripheral blood leukocytosis with blasts in association with an accumulation of megakaryoblasts in the blood, liver, and marrow. Similar reactions have also been reported in patients with trisomy 21 mosaicism who are phenotypically normal. TMD resolves spontaneously in most patients but can progress to acute megakaryoblastic leukemia (AMKL) in 23 to 30% of affected patients. This disorder is attributable to acquisition of mutations in the *GATA1* gene, which encodes a key transcription factor for hematopoietic regulation, leading to loss of normal GATA-1 expression and expression of a truncated GATA-1 protein. Evidence supports that these mutations are acquired during fetal life, and patients present in early infancy with TMD. The pathogenesis of progression to AMKL presumably requires additional genetic events and sequential epigenetic changes, and is the focus of intense study.[3,4]

Leukocyte Adhesion Deficiency

Patients with leukocyte adhesion deficiency (LAD) (see also Chapter 169) have persistent leukocytosis, defects in stimulus-dependent activation of neutrophils, recurrent infections, and delayed separation of the umbilical cord. LAD is an abnormality of leukocyte adhesion reflecting the loss of surface adhesion molecules. LAD-1 is due to absence or marked reduction in the common β chain of β2 integrins, resulting in loss of expression of leukocyte function–associated antigen 1 (LFA-1), the C3bi receptor, and GP150;95. This results in a failure to ingest and kill microbes opsonized by C3bi. In LAD-2, neutrophils lack sialyl Lewis X, the ligand for L-selectin expressed on endothelial cells. Neutrophils appear morphologically normal but are defective in chemotaxis, adherence, and phagocytosis.[5]

Familial Cold Urticaria

Familial cold urticaria is marked by episodic fevers, leukocytosis, urticaria, rash, conjunctivitis, and muscle and skin tenderness with cold exposure. The rash is composed of infiltrating neutrophils. The syndrome appears to be related to decreased levels of C1-esterase inhibitor and is associated with mutations in the *CIAS1* gene on chromosome 1q.

Chronic Myelogenous Leukemia and Other Myeloproliferative Disorders

Chronic myelogenous leukemia (CML), chronic neutrophilic leukemia (CNL), and the other myeloproliferative neoplasms (namely polycythemia vera [PV] and essential thrombocythemia [ET]) are discussed in detail in Chapters 184 and 166. They are the principal acquired primary hematologic disorders associated with neutrophilia. They are marked by clonal expansion of myeloid precursors and increased release of both immature and mature myeloid cells into the peripheral blood. CML on presentation often has to be distinguished from a leukemoid reaction. In contrast to a leukemoid reaction, CML is characterized by the presence of abnormalities of other blood cell lines ("panmyelosis") and by the presence of specific abnormalities. Therefore, the peripheral blood smear in CML (but not leukemoid reaction) demonstrates increased numbers in all cells of the neutrophilic series, classically with a greater proportion of myelocytes to metamyelocytes, and may display concomitant basophilia, eosinophilia, anemia, and thrombocytosis. CNL, a rare myeloproliferative neoplasm, is characterized by hepato/splenomegaly and leukocytosis of at least 25,000/μL, with more than 80% of leukocytes being segmented neutrophil/band forms and less than 10% being immature granulocytes, in contrast to CML.[6] CML is characterized by the diagnostic presence of the Philadelphia chromosome [t(9;22)], which can be identified in the peripheral blood by the detection of the *BCR-ABL* translocation by fluorescence in situ hybridization (FISH) or reverse transcription–polymerase chain reaction (RT-PCR). At least 50% of patients with CNL, in contrast, harbor mutations in the receptor for CSF-3 (*CSF3R; GCSFR*).[6] PV and ET, on the other hand, are notable for also having a marked increase in red cell mass and a marked thrombocytosis, respectively, which is often accompanied by leukocytosis.

The leukocyte alkaline phosphatase (LAP) score was historically used in the laboratory evaluation of granulocytosis as a diagnostic marker for myeloproliferative neoplasm. The LAP score is very low (usually 0) in the setting of CML and elevated in PV. The LAP score has a very wide normal range and in practical terms was only definitively helpful in the setting of CML because "high" LAP scores can also be seen in infectious and inflammatory settings. With the availability of direct molecular genetic diagnosis of CML by assay for the *BCR-ABL* fusion gene, it can no longer be recommended as a diagnostic test.

Post-splenectomy

Patients develop leukocytosis following splenectomy, and this may be long-standing, reflecting the loss of a major site of neutrophil margination. This is of no clinical importance, except insofar as it leads to unnecessary evaluation for other pathology.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

As outlined previously, acquired leukocytosis is most commonly the result of acute or chronic infection or inflammation. When it occurs in the absence of

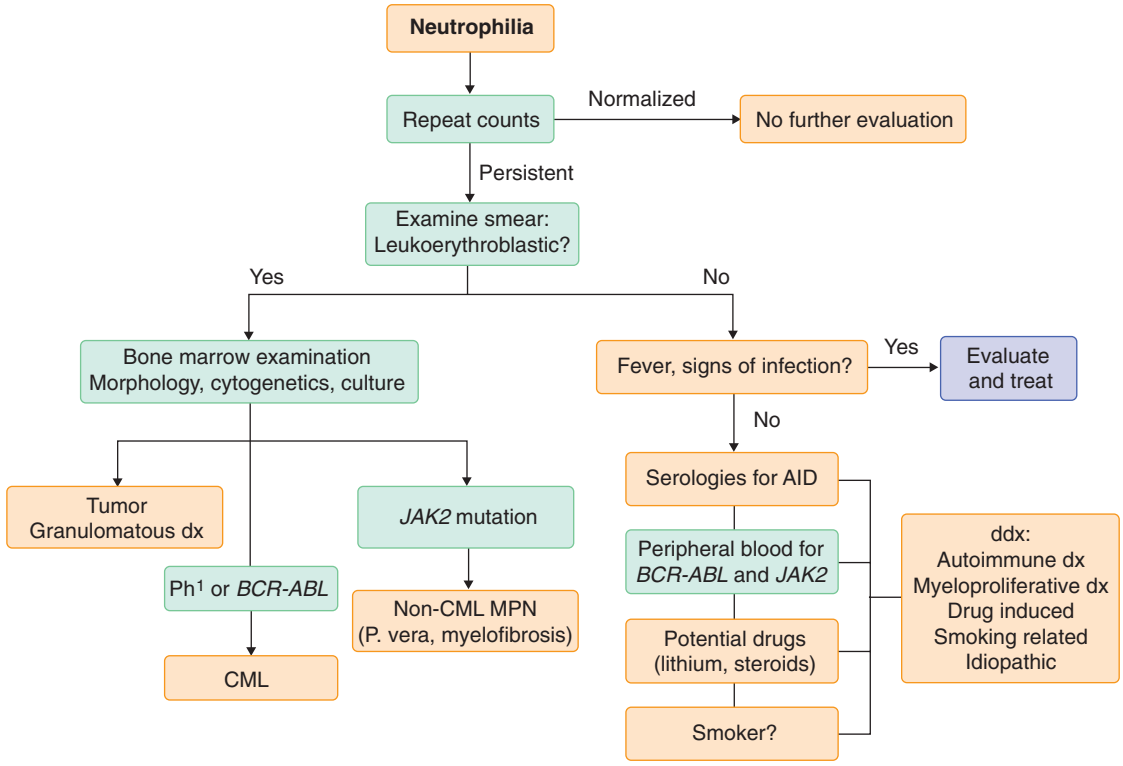

FIGURE 167-4. Diagnostic approach to neutrophilia. AID = autoimmune disease; CML = chronic myelogenous leukemia; ddx = differential diagnosis; dx = disease; MPN = myeloproliferative neoplasm; Ph[1] = Philadelphia chromosome; PV = P. vera, ET = essential thrombocytopenia.

fever, aberrations in acute phase reactants, effusions and edema, or other signs and symptoms of inflammation, it still may be secondary to drugs or an underlying nonhematologic malignancy. As such, it should be seen as the sign of a healthy hematopoietic system responding to an outside stress. Bone marrow evaluation is therefore rarely indicated. However, persistence of leukocytosis in the absence of signs and symptoms of inflammation or infection, nonhematologic malignancy, and offending drugs should prompt an evaluation for a primary myeloproliferative disease or clonal hematologic neoplasm, particularly when there is evidence of a leukoerythroblastic reaction. CML and other myeloproliferative neoplasms can be ruled out by molecular diagnosis on the peripheral blood, as described previously and in Chapter 166. In this setting, bone marrow examination is indicated to evaluate for marrow infiltration by infection, tumor, or fibrosis and should include cultures for tuberculosis and fungal infection as well as cytogenetics and flow cytometry (Fig. 167-4).

Leukocytosis Due to Expansion of Other Cell Lines

Monocytosis and lymphocytosis can also lead to elevations of the WBC. Monocytosis is defined by an absolute monocyte count of greater than 500/μL and usually occurs in the setting of chronic inflammation resulting from infections like tuberculosis, syphilis, or subacute bacterial endocarditis, autoimmune or granulomatous disease, and sarcoidosis. It can also be seen in malignancies, such as preleukemic states, nonlymphocytic leukemia including acute myelomonocytic and monocytic leukemia, histiocytosis, Hodgkin disease, non-Hodgkin lymphoma, and various carcinomas. Finally, it can be seen in the setting of chronic neutropenia, after splenectomy, and in the setting of recovery from marrow suppression (Table 167-2).

Lymphocytosis is defined by an absolute lymphocyte count of more than 5000/μL. The most common causes of an elevated lymphocyte count are viral infections such as Epstein-Barr virus and the hepatitis viruses. Although most bacterial infections cause neutrophilia, pertussis and cat-scratch disease due to *Bartonella henselae* can cause an impressive lymphocytosis. Other infections that may cause a secondary lymphocytosis include toxoplasmosis and babesiosis. Hypersensitivity reactions due to drugs or serum sickness may also be associated with lymphocytosis. Primary disorders that cause a lymphocytosis include chronic lymphocytic leukemia (CLL) and monoclonal B-cell lymphocytosis (Table 167-3; see also Table 184-2 and Chapter 184).

Eosinophilia is defined by an absolute eosinophil count of more than 400/μL. Eosinophils proliferate under the influence of IL-5 and play a role in phagocytosis and modulating toxicity due to mast cell degranulation in hypersensitivity reactions. Eosinophilia is therefore most often seen in the setting of drug reactions, allergy, atopy, and asthma. A variety of infections, particularly parasitic infections and, to a lesser degree, fungal infections, can be associated with an increased number of circulating eosinophils. Eosinophilia can also be the result of autoimmune and inflammatory conditions, as in Churg-Strauss vasculitis. Atheroembolic disease and adrenal insufficiency may also cause eosinophilia. A number of cancers have been associated with polytypic expansion of eosinophils, including lymphomas and solid tumors. There are also a number of clonal disorders of eosinophils that occur in the setting of some leukemias. Finally, there is a heterogeneous group of disorders termed *hypereosinophilic syndromes*. A *FIP1L1-PDGFRA* fusion gene has confirmed that some of these are primary clonal disorders of eosinophils; the clonality of other hypereosinophilic syndromes can be difficult to establish (see Table 170-1).

TABLE 167-2 DIFFERENTIAL DIAGNOSIS OF MONOCYTOSIS

Infection
Granulomatous disease (tuberculosis, fungal disease)
Endocarditis
Syphilis
Autoimmune diseases
Lupus, rheumatoid arthritis
Giant cell arteritis
Vasculitis
Inflammatory bowel disease
Sarcoid
Malignancy
Primary hematologic malignancy
Chronic myelomonocytic leukemia
Acute myelomonocytic leukemia
Lymphoma
Solid tumors
Neutropenia
Associated with chronic neutropenia
Recovery form marrow suppression
Post-splenectomy

TABLE 167-3 DIFFERENTIAL DIAGNOSIS OF LYMPHOCYTOSIS

- Infection
  - Viral infection
    - Epstein-Barr virus
    - Cytomegalovirus
    - Hepatitis
  - Bacterial infection
    - Pertussis
    - Bartonella
    - Tuberculosis
    - Syphilis
    - Rickettsia
    - Babesia
- Hypersensitivity reactions
  - Serum sickness
  - Drug hypersensitivity
- Primary hematologic disease
  - Chronic lymphocytic leukemia
  - Monoclonal B-cell lymphocytosis
  - Non-Hodgkin lymphoma

NEUTROPENIA

The risk for infection in the setting of neutropenia is highly dependent on the size of the neutrophil storage pool. Although neutropenia is defined by an absolute neutrophil count of less than 1500/μL, patients with neutrophil counts below this number may have different risks and rates of infection depending on the cause of neutropenia. For instance, patients who are neutropenic owing to chemotherapy, marrow failure, or marrow exhaustion experience infection at much higher rates than those with chronic neutropenic syndromes and immune-mediated neutropenia. Neutropenia may be congenital or acquired. The following sections first discuss congenital neutropenic disorders, the study of which has provided critical insights into normal myelopoiesis, and then the secondary causes of neutropenia (Table 167-4).

ETIOLOGY

Primary Causes of Neutropenia

Ethnic and Benign Familial (Constitutional) Neutropenia

The normal range of the neutrophil count is genetically determined and can be variable. A number of racial and ethnic groups have been observed to have a relatively large proportion of members who are neutropenic by comparison to the published normal range, usually based on young, largely white individuals. This is termed *constitutional neutropenia* and is seen among a variety of ethnic groups, including African Americans, Yemenite Jews, Falasha Jews, and African Bedouins. Single-nucleotide polymorphisms in the gene for the Duffy antigen receptor for chemokine *(DARC)* have been shown to associate with race and are a postulated candidate to explain racial and ethnic differences in neutrophil counts.[7] There is also an autosomal dominantly inherited condition called *benign familial neutropenia* that is characterized by neutrophil counts in the 800 to 1400/μL range. Neither ethnic nor benign familial neutropenia has been shown to be associated with *any* increased risk for infection.

Severe Congenital Neutropenia

First described by Rolf Kostmann in 1956, severe congenital neutropenia (SCN) is a disorder of severe neutropenia with neutrophil counts of less than 500/μL, presenting in the neonatal period with recurrent bacterial infections. These infections can occur as early as the first months of life and often include omphalitis and perirectal abscesses. There is often an increase in other myeloid cell lines, including monocytes and eosinophils. Bone marrow biopsy in SCN patients reveals a "maturation arrest" with an absence of mature neutrophil elements. SCN can follow autosomal dominant and recessive and X-linked patterns of inheritance and has been shown to be associated with mutations in a variety of genes, as summarized in Table 167-5.

Neutrophil elastase (ELA2, now ELANE) is a serine protease that is synthesized at high levels in the promyelocyte stage of neutrophil maturation and is packaged in primary granules. It was originally hypothesized that mutations in ELANE led to its defective cellular trafficking and cytoplasmic accumulation, subsequently triggering neutrophil apoptosis. Newer evidence, however, supports a mechanism by which accumulation of the mutant

TABLE 167-4 DIFFERENTIAL DIAGNOSIS OF NEUTROPENIA

- Congenital neutropenia
  - Ethnic and benign familial (constitutional) neutropenia
  - Severe congenital neutropenia
    - Autosomal dominant (*ELANE* mutation)
    - Autosomal recessive (Kostmann syndrome; *HAX2* mutation)
    - X-linked (*WASP* mutation)
    - Other rare defects (*G-CSFR* mutation, unknown)
  - Cyclic neutropenia
  - Shwachman-Diamond syndrome
  - Fanconi anemia
  - Dyskeratosis congenita
  - Glycogen storage disease type Ib
  - Myelokathexis
  - Chédiak-Higashi syndrome
  - Griscelli syndrome type II
  - Hermansky-Pudlak syndrome II
  - Barth syndrome
- Acquired neutropenia
  - Infection
  - Postinfection
  - Drug induced
  - Immune neutropenia
    - Primary immune neutropenia
    - Secondary to autoimmune disease
      - Rheumatoid arthritis
        - Felty syndrome
        - Large granular lymphocyte disease
      - Systemic lupus erythematosus
      - Wegener granulomatosis
      - Hyperthyroidism
    - Pure white cell aplasia associated with thymoma
    - Large granular lymphocyte disease
  - Primary bone marrow failure
    - Aplastic anemia
    - Myelodysplastic syndrome
    - Acute leukemia
  - Margination and hypersplenism
  - Vitamin and mineral deficiencies (including B_{12}, folate, copper)
  - Chronic idiopathic neutropenia in adults (CINA)

ELANE = neutrophil elastase; G-CSFR: granulocyte colony-stimulating factor receptor; WASP = Wiskott-Aldrich syndrome protein.

TABLE 167-5 CONGENITAL NEUTROPENIA SYNDROMES

SYNDROME	INHERITANCE PATTERN	GENE
SCN	Autosomal dominant	*ELANE* (~60%)
	Autosomal recessive	*HAX1* (~5%)
	X-linked	*WASP* (~5%)
	X-linked	*TAZ* (rare)
	Autosomal recessive	*G6PC3* (~2%)
	Autosomal dominant	*Gfi1* (rare)
	Autosomal dominant	*G-CSFR* (rare)
CYCLIC NEUTROPENIA	Autosomal dominant	*ELANE*
OTHER CONGENITAL SYNDROMES		
Shwachman-Diamond syndrome	X-Linked	*SBDS*
Fanconi anemia	Autosomal recessive	*FANCA-FANCO*
Dyskeratosis congenita	Variable	Telomerase and related genes
Glycogen storage disease type 1b	Autosomal recessive	*G-6-PT*
Myelokathexis	Autosomal dominant	*CXCR4*
Chédiak-Higashi syndrome	Autosomal recessive	*LYST*
Griscelli syndrome type 2	Autosomal recessive	*RAB27A*
Hermansky-Pudlak syndrome type 2	Autosomal recessive	*AP3B1*

ELANE = neutrophil elastase; G6PC3 = glucose-6-phosphatase catalytic subunit 3; G-6-PT = glucose-6-phosphatase translocase; G-CSFR = granulocyte colony-stimulating factor receptor; LYST = lysosomal trafficking regulatory gene; SBDS = Shwachman-Bodian-Diamond syndrome; SCN = severe congenital neutropenia; WASP = Wiskott-Aldrich syndrome protein.

neutrophil elastase in the endoplasmic reticulum activates the unfolded protein response, leading to apoptosis. SCN due to these mutations is inherited in an autosomal dominant fashion. Hax-1 is a mitochondrial protein with weak homology to bcl-2, and its absence results in mitochondrial-dependent apoptosis. It is the mutated protein implicated in the original autosomal recessive cases of SCN described by Kostmann. Although all SCN cases were originally referred to as *Kostmann syndrome*, that term is now reserved for this subgroup of autosomal recessive SCN. Wiskott-Aldrich syndrome protein (WASP) regulates actin polymerization in hematopoietic cells, and deficiency in this protein results in the Wiskott-Aldrich syndrome, characterized by small platelets in low number, sinopulmonary infections, and eczema. Another phenotype of mutated WASP, however, is X-linked thrombocytopenia and neutropenia. Mutations in the glucose-6-phosphatase catalytic subunit 3 (*G6PC3*) are the most recently discovered cause of a subset of SCN patients; homozygous loss of this metabolic enzyme also appears to lead to activation of the unfolded protein response and increased apoptosis of neutrophil precursors.[8]

SCN was formerly a disease of infancy and early childhood because few, if any, patients survived to adulthood. However, the advent of the availability of recombinant G-CSF and the observation that G-CSF is able to raise neutrophil counts and prevent infection in most patients have allowed these children to survive. Some patients require very high doses of G-CSF, but responses are seen in 80 to 90% of individuals. Increased survival led to the emerging realization that SCN predisposes to the development of myelodysplasia and acute leukemia (MDS/AML), with the development of MDS/AML at a rate of approximately 2% per year, and a cumulative risk of about 30% over 10 years.[9] Patients refractory to, or requiring very high doses of, G-CSF appear to have a higher risk for leukemic transformation. As discussed later, development of MDS/AML is often associated with the acquisition of a mutation in the gene encoding for the G-CSF receptor. Considerable controversy exists concerning the potential contribution of G-CSF administration to the risk for malignant transformation in children with SCN who receive lifelong treatment with G-CSF (see later under Treatment).

Two classes of G-CSF receptor mutations have been associated with SCN and either hyper-responsiveness or hyporesponsiveness of the receptor. Initial studies aimed at demonstrating that SCN was caused by mutation of the G-CSF or G-CSF receptor gene did not implicate such mutations in the pathogenesis of a significant number of patients with SCN. However, a small number of patients have been shown to have mutations in the G-CSF receptor gene that block ligand binding and produce a G-CSF-resistant form of SCN. More important, however, the studies identified an acquired missense mutation that introduces a stop codon and leads to the deletion of the distal intracellular domain of the receptor known to be responsible for differentiation signaling. This has been hypothesized to cause proliferation of hematopoietic progenitors at the expense of maturation and to be associated with hypersensitivity to G-CSF, suggesting that this mutation may play an important role in the development of MDS/AML in the setting of SCN. However, whether this mutation is in fact of pathogenetic importance to the development of MDS/AML and whether G-CSF influences the risk for developing the mutation and influencing leukemic transformation remain subjects of significant controversy.

Cyclic Neutropenia

Cyclic neutropenia is defined as periods of neutropenia (≤200/μL) lasting 3 to 5 days and occurring at approximately 21-day intervals. These periods of neutropenia may be marked by recurrent fevers, mouth sores, and infections of the skin, upper respiratory tract, and ears. The disorder can be dominantly inherited or sporadic. Congenital cyclic neutropenia has also been shown to be associated with mutations in the gene for neutrophil elastase in virtually all cases tested to date.[10] The diagnosis formerly required demonstration of transient neutropenia through frequent blood counts over a course of 6 weeks but now can be established by sequencing of the neutrophil elastase gene. It is successfully treated with G-CSF and is not associated with an increased risk for leukemic transformation. Rare cases of cyclic neutropenia acquired in adulthood have been associated with systemic diseases such as large granular lymphocytosis or T-cell lymphoma.

Other Congenital Syndromes with Associated Neutropenia

A number of other congenital syndromes are associated with neutropenia as one of the constellation of disease-associated abnormalities. These include Shwachman-Diamond syndrome, Fanconi anemia, dyskeratosis congenita, glycogen storage disease Ib, myelokathexis, Chédiak-Higashi syndrome, Griscelli syndrome II, and Hermansky-Pudlak syndrome II.[11]

Shwachman-Diamond syndrome usually begins as an isolated neutropenia but progresses to marrow failure and is also associated with pancreatic dysfunction and skeletal abnormalities. The responsible gene, the Shwachman-Bodian-Diamond syndrome gene (*SBDS*), is involved in the regulation of ribosomal RNA. These patients carry an increased risk for leukemic transformation.

Fanconi anemia is due to mutations in genes involved in DNA repair, and as such, it takes more time for marrow failure to develop in these patients (median age, 7 years). Patients with Fanconi anemia often have, in addition to marrow failure, short stature with upper limb anomalies and hyperpigmented cafe au lait spots, although about one third have no physical abnormalities. Screening is done by chromosomal fragility testing following exposure to diepoxybutane or mitomycin C as well as direct assessment for known Fanconi gene mutations. Stem cell transplantation is curative but carries a high risk for morbidity and mortality owing to the toxicity of preparative regimens.[12]

Dyskeratosis congenita (DKC) is a syndrome of nail dystrophy, leukoplakia, and skin pigmentation abnormalities with associated neutropenia or aplastic anemia, or both. It can be inherited in an autosomal dominant or recessive or X-linked fashion and has been shown to be associated with mutations in several genes that are implicated in telomere maintenance. It typically does not present until the second decade of life. Recent studies have demonstrated that DKC is one of several diseases associated with telomere abnormalities, with wide-ranging manifestations including pulmonary fibrosis and hepatic cirrhosis as well as bone marrow failure.

Glycogen storage disease Ib is inherited in an autosomal recessive fashion and characterized by intermittent neutropenia due to defects in the neutrophil respiratory burst with subsequent apoptosis of circulating neutrophils. Hepatomegaly and metabolic crises are also the hallmarks of this disease and are due to mutations in the gene for the glucose-6-phosphatase translocase enzyme.

The neutropenia of *myelokathexis* is due to retention of mature neutrophils in the bone marrow despite a low peripheral neutrophil count. During infection, however, patients with myelokathexis typically have a sudden rise in their neutrophil count, which makes their clinical course relatively more benign. There is an association between this condition and hypogammaglobulinemia and warts, the WHIM syndrome (*w*arts, *h*ypogammaglobulinemia, *i*mmunodeficiency, and *m*yelokathexis). It has been shown to be caused by heterozygous mutations in the gene encoding chemokine receptor CXCR4.

Chédiak-Higashi syndrome, Griscelli syndrome II, and *Hermansky-Pudlak syndrome II* are all syndromes of albinism and neutropenia due to defects in vesicular trafficking. Chédiak-Higashi syndrome is due to mutations in a lysosomal trafficking regulatory gene (*LYST*) and is characterized by oculocutaneous albinism, bleeding, progressive neurologic disease, and increased susceptibility to hemophagocytic syndrome. Patients with Griscelli syndrome II also have an increased susceptibility to hemophagocytic syndrome as well as albinism and periodic neutropenia; Griscelli syndrome II is caused by mutations in the gene encoding the small guanosine triphosphatase RAB27A, which is involved in the release of myeloperoxidase from the primary granules of neutrophils. Hermansky-Pudlak syndrome II is due to mutations in the *AP3B1* gene, which encodes a part of a protein transport complex that is involved in vesicular trafficking in many cell types and appears to be involved in the trafficking of neutrophil elastase. It is also marked by albinism, platelet abnormalities, and pulmonary fibrosis.

Barth syndrome is an X-linked autosomal recessive disorder characterized by neutropenia, cardiomyopathy, and growth retardation, with a high mortality rate through early childhood because of the heart disease. The causative mutation is in the *TAZ* gene, which encodes tafazzin protein that is critical to remodeling cardiolipin in the mitochondrial membrane.

Secondary Causes of Neutropenia

Infection-Related Neutropenia

Several viral infections have been shown to cause a transient neutropenia that typically resolves as the viremia abates. These include varicella, measles, rubella, hepatitis A and B, Epstein-Barr virus, influenza, parvovirus, and cytomegalovirus. The mechanisms are diverse and can involve redistribution, decreased production, and immune destruction of neutrophils. Human immunodeficiency virus and acquired immunodeficiency syndrome can likewise cause mulitfactorial leukopenia and neutropenia (Chapter 393). Patients

often have splenomegaly with increased sequestration, but more commonly the neutropenia reflects immune-mediated destruction. A myriad of atypical infections like *Mycobacterium tuberculosis*, ehrlichiosis, rickettsia, tularemia, brucellosis, and some staphylococcal infections can cause a moderate neutropenia. Any infection leading to overwhelming sepsis can cause neutropenia, but this is usually through consumption of the marrow neutrophil reserve and is typically seen in newborns and elderly patients and not in individuals with an otherwise healthy and mature marrow. There is also increased margination of neutrophils during sepsis due to systemic activation of complement, exacerbating the neutropenia.

Drug-Induced Neutropenia and Neutropenia Due to Marrow Injury

Drug-induced neutropenia is the most common cause of neutropenia. Multiple drugs have been implicated in neutropenia and agranulocytosis, in both predictable and idiosyncratic patterns. Drug-induced neutropenias may reflect either suppression of marrow granulopoiesis or increased destruction or clearance of peripheral neutrophils. Many drugs cause direct dose-dependent marrow suppression, which is predictable and often mild. Others incite an idiosyncratic immune-mediated destruction that can present with profound agranulocytosis. The typical pattern of drug-induced neutropenia is a marked decline in the neutrophil count that occurs after about 1 to 2 weeks of exposure to the drug with a recovery that begins within days of stopping the drug. However, atypical cases can present long after drug initiation, and others may be associated with a longer interval before recovery of the neutrophil count (Fig. 167-5). Patients with drug-induced agranulocytosis may present with acute sepsis, and it is associated with a significant risk for acute mortality. Recovery is often preceded by the appearance of monocytes and immature neutrophil forms. The more hypercellular the marrow is at diagnosis, the earlier marrow recovery may occur. Some common drugs known to cause neutropenia, in addition to antineoplastic, antiviral, and immunosuppressive agents, include clozapine, the antithryoidal thioamides including carbimazole, methimazole and propylthiouracil, quinidine, procainamide, sulfasalazine, and Levamisole, which is widely used as a cocaine adulterant, causing cocaine-associated neutropenia. Neutrophil recovery is speeded by G-CSF, although there are no definitive data that this improves survival in this setting.

Radiation can also result in marrow injury leading to an acute or chronic marrow failure state; in high doses, it is also a risk factor for the development of myelodysplasia and leukemia. These malignant hematopoietic diseases can themselves cause marrow failure because the malignant cells proliferate within the marrow-occupying space and can cause marrow fibrosis, both of which lead to cytopenias. These diseases are discussed in greater detail in Chapters 182 and 183, respectively. Likewise, metastatic carcinoma to the bone can also cause marrow failure because the marrow becomes increasingly replaced by the metastatic cells.

Immune Neutropenia

Infection and drugs cause immune-mediated neutrophil destruction. However, immune neutropenia can also occur as an isolated phenomenon (primary immune neutropenia) or as a manifestation of an underlying systemic autoimmune disease (secondary immune neutropenia). Primary autoimmune neutropenia is primarily a disease of children younger than 4 years; median age of onset is 6 to 12 months. Although infectious risk is increased, treatment is restricted to prophylactic antibiotics, with G-CSF reserved for acute infectious episodes. Ninety-five percent of patients undergo spontaneous remissions within 2 years. Nearly all of these patients have antineutrophil antibodies directed against antigens derived from the FcγIIIb receptor; these antibodies mediate neutrophil destruction by either sequestration in the spleen or complement-mediated neutrophil lysis.

Secondary autoimmune neutropenia is a disease of adults and can be seen in association with hyperthyroidism, Wegener granulomatosis, rheumatoid arthritis, and systemic lupus erythematosus (SLE). The role of antineutrophil antibodies in these patients is less clear. More than 50% of patients with SLE, for example, have antineutrophil antibodies, but many have normal neutrophil counts, and there is a poor correlation between the presence of the antibodies and neutrophil number.[13]

Felty syndrome and *large granular lymphocyte syndrome* deserve separate mention. Felty syndrome occurs in association with long-standing rheumatoid arthritis (RA) (Chapter 264) and is characterized by splenomegaly and profound neutropenia. Large granular lymphocyte syndrome often occurs in the setting of RA but can also occur as an isolated phenomenon. Both Felty syndrome and large granular lymphocyte syndrome are associated with a proliferation of large granular lymphocytes, with a characteristic surface phenotype ($CD3^+$, $CD8^+$, $CD16^+$, and $CD57^+$). In the setting of RA, these two syndromes were originally thought to be separate diseases, with Felty syndrome being polyclonal and large granular lymphocyte syndrome representing a monoclonal proliferation of larger granular lymphocytes. However, with increasing sensitivity of detection of monoclonal populations of lymphocytes, this distinction has become blurred. It has been observed that more than 90% of RA patients with either syndrome are human leukocyte antigen (HLA)-DR4-positive, leading to the postulate that the two entities represent the extremes of a single spectrum of disease. This HLA restriction is not found among non-RA patients with large granular lymphocytes. Both syndromes cause immune-mediated neutrophil destruction by a wide array of mechanisms, including antineutrophil antibodies and cell-mediated destruction. Some patients may also have G-CSF resistance mediated by inhibitory G-CSF antibodies.[14]

There are other rare forms of immune neutropenia. Isoimmune neonatal neutropenia is a moderate to severe neutropenia of the newborn due to transplacental passage of maternal immunoglobulin G antibodies against alleles inherited from the father, resulting in neutropenia in a manner similar to the development of anemia in Rh hemolytic disease. Pure white cell aplasia is a rare disease associated with severe pyogenic infections, and also with

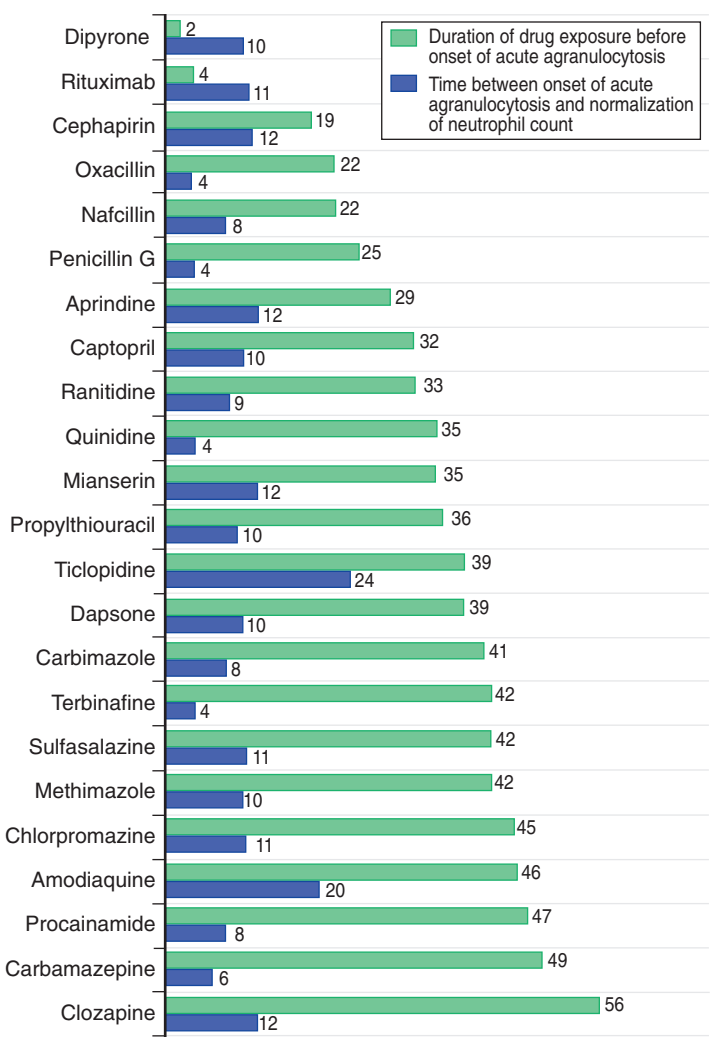

FIGURE 167-5. Median duration of treatment and neutropenia. Only drugs with more than three definite or probable reports of the time between onset of acute agranulocytosis and normalization of neutrophil count and the duration of treatment before onset of acute agranulocytosis were considered. (From Andersohn F, Konzen C, Garbe E. Systematic review: agranulocytosis induced by nonchemotherapy drugs. *Ann Intern Med.* 2007;146: 657-665.)

thymoma in more than two thirds of cases. It has also occurred following ibuprofen therapy. There is a complete absence of myeloid precursors on bone marrow examination. It is immune mediated, but removal of the thymoma in thymoma-associated cases may not be sufficient for remission. Adjuvant therapy with cyclophosphamide, corticosteroids, cyclosporine, and intravenous immunoglobulin may be needed.

Neutropenia Due to Increased Margination and Hypersplenism

Complement activation can result in both acute and chronic neutropenia as a result of increased margination of the circulating neutrophil pool. This is attributed to the fact that C5a renders neutrophils more adherent and thereby prone to aggregation within the pulmonary vasculature. This has been seen in patients suffering from burns and transfusion reactions. Complement activation may also lead to neutrophil destruction, as in paroxysmal nocturnal hemoglobinuria. The circulating neutrophil pool can also be diminished in association with hypersplenism, although this is typically less common and less pronounced than the anemia and thrombocytopenia seen in patients with an enlarged spleen.

Neutropenia Due to Nutritional Deficiency

Several vitamin and mineral deficiencies, particularly B_{12}, folate (Chapter 164), and copper, are associated with neutropenia. Copper deficiency is a particularly under-recognized cause of neutropenia (usually seen along with anemia), found especially in clinical situations like total parenteral nutrition without copper supplementation, protein-losing enteropathies, celiac disease, gastric bypass surgery and postgastrectomy, malabsorption syndromes, and zinc toxicity. In addition to low serum copper and ceruloplasmin levels, distinct morphologic findings include hypogranularity and hypolobation (PHA) on peripheral smear, and cytoplasmic vacuolization of myeloid as well as erythroid precursors and ringed sideroblasts.

Deficiencies in these vitamins and minerals result in ineffective myelopoiesis, maturation arrest, and megaloblastic changes with nuclear-cytoplasmic dyssynchrony. The characteristic finding in the setting of megaloblastic anemia is hypersegmentation of the neutrophil (Fig. 167-6).

Chronic Idiopathic Neutropenia

Chronic idiopathic neutropenia in adults (CINA) has been described as an acquired disorder of granulopoiesis characterized by prolonged neutropenia in the absence of any apparent underlying etiology.[15] Patients present with chronic neutropenia that is often an incidental finding on routine blood tests, making it impossible to know how long the neutropenia has been present. They have no evidence of autoimmune disease, nutritional deficiency, or myelodysplasia. The syndrome is heterogeneous, with a wide range of neutrophil counts. A group of patients with CINA from Greece were originally found to have increased production of transforming growth factor-β (TGF-β) and consequent suppression of granulopoiesis by the bone marrow. These patients tend to have very mild neutropenia, with an absolute neutrophil count (ANC) that is rarely less than 800. They may have an ethnic predisposition to neutropenia; indeed, the neutropenia has been demonstrated to be linked to a genetic polymorphism in the TGF-β locus. It seems likely, however, that these patients should be distinguished from other CINA patients, many of whom have an ANC below 200. The etiology of neutropenia in these patients is completely unknown. However, the natural history of CINA, even in the face of very low neutrophil counts, is generally benign. Most patients require no therapy, although those with very low counts must be treated with G-CSF when they develop fever in the setting of infections. Some patients with recurrent infections or troublesome aphthous ulcers require chronic G-CSF treatment. These patients typically respond to fairly low doses of G-CSF, and there is no reported increase in the development of MDS/AML.

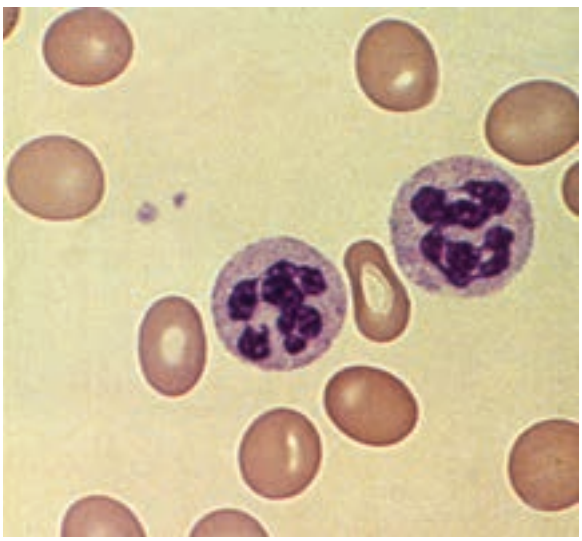

FIGURE 167-6. Peripheral blood with macrocytosis and hypersegmented neutrophils in megaloblastic anemia.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Neutropenia by itself is not associated with many clinical signs and symptoms other than those of the condition that may be causing it. It becomes clinically evident, however, when it results in infection. Although patients with an ANC below 1000/μL do have a slightly increased risk for infection, the risk is substantially increased once the neutrophil count falls below 500/μL. Given the lack of neutrophils, the signs and symptoms of infection may be attenuated; as such, pneumonia may be present with minimal infiltrate on chest radiograph, or a urinary tract infection may yield only a very mild pyuria. Given this, fever in any neutropenic patient must be considered an emergency with prompt acquisition of cultures and administration of empirical antibiotic therapy.

When fever, infection or sepsis, and neutropenia present concomitantly for the first time, it can be difficult to determine whether the neutropenia predated the infection or, conversely, if it is the result of the infection. Examination of the peripheral blood smear can be helpful in this regard because an elevation of band forms and evidence of toxic granulation suggest the latter.

Because drug-induced neutropenia is the most common cause of acquired neutropenia, a careful inventory of all drug and toxin exposures is warranted.[16] Likewise, there should be a careful evaluation for underlying malignant and inflammatory conditions as the precipitant of neutropenia. The time course of the neutropenia and infections can provide clues to the etiology of the neutropenia (acute versus chronic, persistent versus cyclic, neonatal versus childhood versus adult onset). Attention should be paid to the skin, bones, appendages, and nails because abnormalities in these may point toward one of the congenital neutropenia syndromes. Evaluation of the complete blood count, peripheral blood smear, and vitamin B_{12} and folate levels should also be performed.

When the neutropenia is not severe and is associated with anemia and thrombocytopenia, one should consider the possibility of hypersplenism. In many cases, the diagnosis can be made by the finding of palpable splenomegaly. However, especially in obese patients, abdominal imaging should be used to evaluate spleen size. Abdominal ultrasound allows the assessment of portal venous flow with Doppler studies. If splenic enlargement is confirmed, the etiology of the splenomegaly (Chapter 168) should be determined. It may reflect congestive splenomegaly secondary to portal hypertension (as a result of cirrhosis, fatty liver, or congestive heart failure, among others) or infiltrative splenomegaly due to a benign or malignant process. Felty syndrome should be considered in the setting of RA.

In patients with chronic neutropenia in the absence of a history of infection or drug or toxin exposure or an evident B_{12} or folate deficiency, bone marrow examination should be performed to rule out myelodysplasia, with assessment of morphology, flow cytometry for large granular lymphocyte syndrome, and cytogenetics. Once a normal marrow has been obtained, further bone marrow examination is not indicated in patients with CINA (Fig. 167-7).

TREATMENT AND MANAGEMENT

Rx

The management of neutropenia is dependent on the etiology of the depressed neutrophil count as well as its severity and the presence or absence of fever or infection.[17] The approach to the patient with fever and neutropenia is discussed in more detail in Chapter 281. Neutropenia with fever is a clinical emergency because these patients are at risk for hemodynamic collapse and septic shock. Therefore, these patients should be evaluated thoroughly and cultured promptly, and antibiotics should be administered within 30 to 60 minutes of presentation before obtaining the results of the cultures. The timely empirical administration of a combination of antipseudomonal antibiotics in

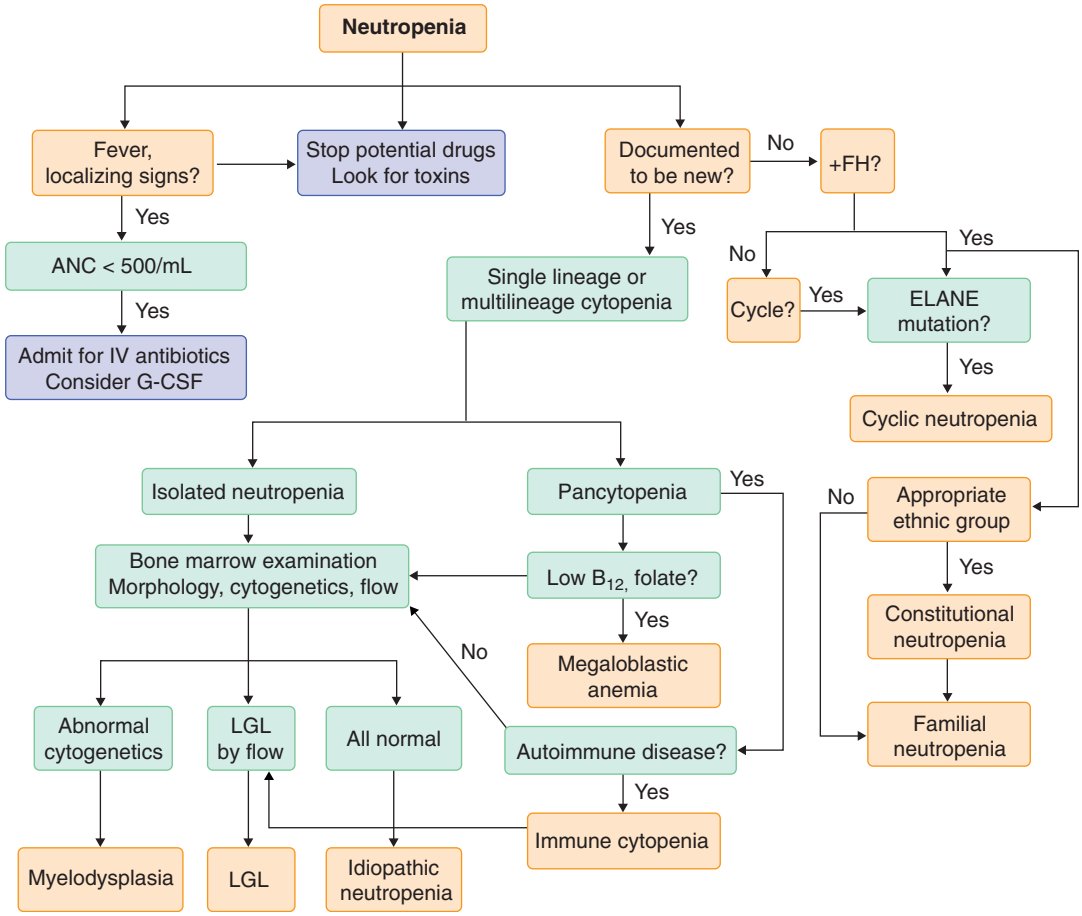

FIGURE 167-7. Diagnostic approach to neutropenia. ANC = absolute neutrophil count; ELANE = neutrophil elastase gene, FH = family history; G-CSF = granulocyte colony-stimulating factor; IV = intravenous; LGL = large granular lymphocyte syndrome.

patients with neutropenia at the onset of fever results in significant clinical benefit, with respect to both response and survival. With the advent of newer generation cephalosporins, studies have shown that monotherapy with a third- or fourth-generation cephalosporin at the onset of fever may be sufficient. A systematic review to assess the evidence for combination therapy versus monotherapy in cancer patients with febrile neutropenia in clinical trials has been recently updated.[■] It concluded that randomized controlled trials have demonstrated the survival superiority of β-lactam monotherapy compared with β-lactam-aminoglycoside combination therapy.

The addition of a second antipseudomonal agent, vancomycin, or antifungal agent is warranted in patients considered at risk for resistant pseudomonal infection, resistant gram-positive infection, or fungal infections, respectively, or in the face of a failure to defervesce within 3 to 5 days of antibiotic administration. These considerations are discussed in more detail in Chapter 281.

How to manage the uninfected and afebrile neutropenic patient is more nuanced and dependent on the etiology of the neutropenia. Patients with immune-mediated neutropenia are typically treated with immunosuppressive therapy, including steroids, antithymocyte globulin, or cyclosporine, or a combination of these, aimed primarily at treatment of the underlying autoimmune disease. Patients usually respond to G-CSF, although in the setting of RA, this may induce a flare of joint symptoms. Patients with large granular lymphocyte syndrome often respond to therapy with low-dose methotrexate (10 mg/m^2 orally once a week), cyclosporine (100 to 600 mg or 2 to 10 mg/kg orally daily), or low-dose cyclophosphamide (50 to 100 mg orally daily).[18]

Patients with congenital neutropenia, including idiopathic, severe congenital, or cyclic neutropenia, are usually successfully managed with G-CSF for years. Before the use of G-CSF, the mean age of death for patients with severe congenital neutropenia was 2 to 3 years. Since the advent of G-CSF, however, life expectancy has been extended by decades into adulthood. Therapy is daily and chronic, given by subcutaneous injection, with doses varying by the type of neutropenia and the individual responsiveness to therapy. It is usually well tolerated, although accelerated bone loss has been observed. Growth and development do not appear to be affected. Patients with SCN typically require the highest doses, whereas those with idiopathic neutropenia require the lowest, and patients with cyclic neutropenia fall somewhere in between. An increased rate of MDS/AML has been reported with the use of G-CSF in patients with SCN, but this has occurred coincidentally with improved survival. The increased incidence of MDS/AML may then be due to the fact that patients are living longer with a disease whose natural history includes a risk for developing MDS or AML. Certainly, the observation that acquired mutations in the G-CSF receptor are present in 65 to 80% of patients with SCN who develop MDS or AML suggests a potential mutagenic pathway toward leukemogenesis, but whether this is enhanced or accelerated by the administration of G-CSF has yet to be determined. Patients with idiopathic and cyclic neutropenia do not develop MDS or AML, despite therapy with G-CSF. However, recent evidence suggests that the risk for MDS/AML in SCN patients is much higher in those requiring high doses of G-CSF (>10 μg/kg/day), and patients with idiopathic and cyclic neutropenia are usually responsive to much lower G-CSF doses.

In patients with CINA, G-CSF should be reserved for acute febrile episodes unless the patient has recurrent infections. Patients with CINA treated with G-CSF may experience significant side effects, including fever, gastrointestinal symptoms, and splenomegaly. Consequently, in patients with CINA requiring chronic G-CSF administration, the cytokine should be administered at the lowest dose necessary to prevent infections; it is usually sufficient to treat to maintain the ANC in the range of 300 to 500.

For patients with an inflammatory, infectious, or drug-induced neutropenia, the recommendation is to treat the underlying condition or stop the offending agent. This, however, is not always immediately possible, as in the cases of HIV-infected patients with opportunistic infections or on antiretroviral therapy, solid organ and bone marrow transplant recipients on immunosuppression and antiviral prophylaxis and treatment, and cancer patients undergoing chemotherapy. Prophylactic use of G-CSF in these patients has been shown to be effective in improving the ANC and decreasing rates of infection and febrile neutropenia, but has not been associated with a proven or consistent survival advantage.[■] In light of these findings, many oncologic professional society guidelines recommend the use of prophylactic G-CSF in patients receiving chemotherapy who have a 20% or greater risk for developing febrile neutropenia based on age, comorbid illness, disease characteristics, and myelotoxicity of the chemotherapy regimen. In addition, it is recommended for use during hematopoietic stem cell transplantation, for patients receiving chemotherapy for non-Hodgkin lymphoma or dose-dense chemotherapy, and for patients with a history of febrile neutropenia receiving further chemotherapy.

The use of prophylactic antibiotics has also been investigated, predominantly in neutropenic patients receiving chemotherapy. Early studies in the 1980s and 1990s demonstrated an improvement in infection-related outcomes but not in infection-related or overall survival. Most recently, several randomized trials of prophylactic quinolones in patients receiving chemotherapy demonstrated an improvement in rates of infection and febrile neutropenia. A meta-analysis of studies investigating the use of prophylactic quinolones in patients receiving chemotherapy was the first to document an overall survival benefit as well, although most of these patients had hematologic as opposed to solid tumor malignancies.[A1] The current Infectious Diseases Society of America guidelines do not recommend the use of prophylactic antibiotics in cancer patients undergoing myelosuppressive chemotherapy, with the exception of trimethoprim-sulfamethoxazole in patients at risk for *Pneumocystis jirovecii* pneumonia. Antibiotic prophylaxis is not generally used outside of the stem cell transplantation setting. The use of antibiotics to prevent infection in patients with neutropenia due to other causes has not been extensively studied but is typically not recommended and should be based on clinical context.

Stem cell transplantation (Chapter 178) can be curative for a number of the congenital neutropenia and bone marrow failure syndromes. It is, however, not without risk and should therefore be reserved for patients with severe neutropenia complicated by recurrent infection definitively shown to be due to marrow failure.

Leukopenia Due to Deficiency of Other Cell Lines

Lymphocyte production takes place in a variety of anatomic sites, and lymphocyte trafficking from those sites is bidirectional, making it difficult to understand lymphocyte dynamics in the same way that we do for neutrophils. Despite this, the peripheral lymphocyte count seems to be maintained in a narrow range at 2000 to 4000/μL, 20% of which are B cells and 70% of which are T cells. Lymphocytopenia is a total lymphocyte count of less than 1500/μL. It can be the result of decreased production, defective trafficking, or increased loss or destruction. Decreased production can occur as a result of protein and calorie malnutrition; lymphocyte progenitor pool injury secondary to radiation, chemotherapy, or immunosuppressive agents; and congenital immunodeficiency states. Endogenous or exogenous glucocorticoid excess can cause lymphocytopenia by altering lymphocyte trafficking. This can also occur as the result of acute bacterial or fungal infections, certain viral infections, and granulomatous disease. Finally, many viruses can cause direct destruction of lymphocytes, as can antilymphocyte antibodies seen in patients with underlying autoimmune diseases. Lymphocytes can also be lost from intestinal lymphatics in cases of protein-losing enteropathy, primary disease of the gut or intestinal lymphatics, or gut edema secondary to severe heart failure. When lymphocytopenia is discovered, a comprehensive assessment of the immune system should be done, including lymphocyte subtyping, quantitative immunoglobulins, and skin testing to detect deficiencies of cell-mediated immunity. Treatment is typically aimed at the underlying disease, but intravenous immunoglobulin can be administered to patients who are hypogammaglobulinemic, and transplantation can be performed in patients with severe deficiencies of cell-mediated immunity due to impaired lymphocyte production and function.

Monocytopenia, eosinopenia, and basophilopenia can be seen in the setting of bone marrow failure syndromes or as a result of acute infection, malignancy, or severe injury. This is thought to be due to elevations in glucocorticoids, prostaglandins, and epinephrine. A rise in these humoral factors has the greatest impact on eosinophils such that a lack of eosinopenia in any of these settings should prompt suspicion for adrenal insufficiency, a primary myeloproliferative syndrome, parasitic infection, or primary hypereosinophilic syndrome. Monocytopenia is less frequently seen, probably owing to the diverse roles monocytes play in normal human physiology; prolonged and extreme monocytopenia may not be compatible with life.

Grade A References

A1. Paul M, Dickstein Y, Schlesinger A, et al. Beta-lactam versus beta-lactam-aminoglycoside combination therapy in cancer patients with neutropenia. *Cochrane Database Syst Rev*. 2013;6:CD003038.

A2. Cooper KL, Madan J, Whyte S, et al. Granulocyte colony-stimulating factors for febrile neutropenia prophylaxis following chemotherapy: systematic review and meta-analysis. *BMC Cancer*. 2011;11:404.

A3. Gafter-Gvili A, Fraser A, Paul M, et al. Meta-analysis: antibiotic prophylaxis reduces mortality in neutropenic patients. *Ann Intern Med*. 2005;142:979-995.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Pillay J, den Braber I, Vrisekoop N, et al. In vivo labeling with 2H2O reveals a human neutrophil lifespan of 5.4 days. *Blood*. 2010;116:625-627.
2. Xu Y, Li H, Bajrami B, et al. Cigarette smoke (CS) and nicotine delay neutrophil spontaneous death via suppressing production of diphosphoinositol pentakisphosphate. *Proc Natl Acad Sci U S A*. 2013;110:7726-7731.
3. Gamis AS, Smith FO. Transient myeloproliferative disorder in children with Down syndrome: clarity to this enigmatic disorder. *Br J Haematol*. 2012;159:277-287.
4. Malinge S, Chlon T, Doré LC, et al. Development of acute megakaryoblastic leukemia in Down syndrome is associated with sequential epigenetic changes. *Blood*. 2013;122:e33-e43.
5. Harris ES, Weyrich AS, Zimmerman GA. Lessons from rare maladies: leukocyte adhesion deficiency syndromes. *Curr Opin Hematol*. 2013;20:16-25.
6. Gotlib J, Maxson JE, George TI, et al. The new genetics of chronic neutrophilic leukemia and atypical CML: implications for diagnosis and treatment. *Blood*. 2013;122:1707-1711.
7. Paz Z, Nails M, Ziv E. The genetics of benign neutropenia. *IMAJ*. 2011;13:625-629.
8. Boztug K, Klein C. Genetics and pathophysiology of severe congenital neutropenia syndromes unrelated to neutrophil elastase. *Hematol Oncol Clin North Am*. 2013;27:43-60, vii.
9. Beekman R, Touw IP. G-CSF and its receptor in myeloid malignancy. *Blood*. 2010;115:5131.
10. Germeshausen M, Deerberg S, Peter Y, et al. The spectrum of ELANE mutations and their implications in severe congential and cyclic neutropenia. *Hum Mutat*. 2013;34:905-914.
11. Sokolic R. Neutropenia in primary immunodeficiency. *Curr Opin Hematol*. 2013;20:55-65.
12. Kee Y, D'Andrea AD. Molecular pathogenesis and clinical management of Fanconi anemia. *J Clin Invest*. 2012;122:3799-3806.
13. Newman K, Owlia MB, El-Hemaidi I, et al. Management of immune cytopenias in patients with systemic lupus erythematosus—old and new. *Autoimmun Rev*. 2013;12:784-791.
14. Liu X, Loughran TP Jr. The spectrum of large granular lymphocyte leukemia and Felty's syndrome. *Curr Opin Hematol*. 2011;18:254-259.
15. Gemetzi C, Mavroudi I, Koutala H, et al. Lymphopenia in patients with chronic idiopathic neutropenia is associated with decreased number of T-lymphocytes containing T-cell receptor excision circles. *Eur J Haematol*. 2012;88:210-223.
16. Hay D, Hill M, Littlewood T. Neutropenia in primary care. *BMJ*. 2014;349:g5340.
17. Gibson C, Berliner N. How we evaluate and treat neutropenia in adults. *Blood*. 2014;124:1251-1258.
18. Zhang D, Loughran TP Jr. Large granular lymphocytic leukemia: molecular pathogenesis, clinical manifestations, and treatment. *Hematology Am Soc Hematol Educ Program*. 2012;2012:652-659.

REVIEW QUESTIONS

1. A 20-year-old African American college student receives a scholarship to spend her junior year abroad in France. She needs to have a health form completed before her departure. A complete blood count (CBC) reveals neutropenia, and her physician advises her to postpone the scholarship and to see you for evaluation. She is healthy with no history of excessive infections. Her physical examination is normal. CBC shows: white blood cell count 3400 w/ 30% neutrophils, hematocrit 41, platelets 200,000. What should you advise her?

A. She should have a bone marrow to rule out a bone marrow process causing neutropenia.
B. She should have serial blood counts to rule out cyclic neutropenia.
C. She should have chromosome analysis of her peripheral blood.
D. She should be evaluated for an elastase mutation.
E. She should go to France without any further evaluation.

Answer: E The normal neutrophil count is partially determined by ethnic background. African American males frequently have a neutrophil count of 1000 to 1500, and African American females may have counts that are even lower. Many may actually have counts below 1000. With a total neutrophil count of 1120, a normal hematocrit and platelet count, and an unremarkable history and physical examination, the patient should be diagnosed with constitutional neutropenia. This requires no further evaluation, although if DARC testing is available, confirming she is Duffy negative will confirm the diagnosis. The patient should be reassured that these counts are normal for her and present no undue risk. She should leave for France as planned. (See Ethnic and Benign Familial [Constitutional] Neutropenia.)

2. A 1-month-old baby boy presents with fever and cough and is found to have pneumonia associated with severe neutropenia. His parents have normal blood counts but had lost a baby girl at age 3 months under similar circumstances 2 years before. Which of the following statements is *not* true?

A. The patient likely has a mutation in neutrophil elastase.
B. The child should be treated with granulocyte colony-stimulating factor (G-CSF).
C. The child is at risk for developing myelodysplastic syndrome and acute myeloid leukemia (MDS/AML).
D. If the child survives, his offspring will probably have normal neutrophil counts.
E. The child's white blood cell count likely has a 21-day cycle in and out of the normal range.

Answer: A This child has a disease that is apparent in neither parent but has a sibling that died of the same disease. This is consistent with an autosomal recessive form of severe congenital neutropenia (SCN). SCN related to neutrophil elastase mutation is an autosomal dominant disorder. Cyclic neutropenia is a milder autosomal dominant disorder, also due to mutations in *ELANE;* it is clinical mild and rarely presents in infancy. Although it can appear sporadically, the occurrence in a sibling makes that virtually impossible. This case is most consistent with Kostmann syndrome, which is due to mutations in the *HAX1* gene. Because it is a very rare recessive disorder, if the child lives to adulthood and has offspring, they will be obligate heterozygotes and, like the patient's parents, will have normal counts. Like patients with dominantly inherited SCN associated with neutrophil elastase mutations, these patients respond to G-CSF and have an increased risk for developing MDS/AML. (See Severe Congenital Neutropenia.)

3. A 45-year-old man presents to the emergency department after a motor vehicle crash. He is afebrile with a blood pressure of 160/95 mm Hg and a pulse of 110 beats per minute. He has several scrapes and bruises, but no major injuries. Admission complete blood count reveals: white blood cell count 16,000, hematocrit 45, and platelets 280,000. What is the most likely explanation for his leukocytosis?

A. Acute bacterial infection
B. Undiagnosed chronic myelogenous leukemia
C. Stress-induced demargination of neutrophils
D. Cytokine release stimulating increased marrow production and release of neutrophils
E. Unsuspected splenic rupture

Answer: C Acute stress is commonly associated with leukocytosis. Acute leukocytosis in almost any setting occurs by demargination, in this case in response to the release of catecholamines during the stress of the accident, also causing his tachycardia. Cytokine-induced marrow proliferation and release of neutrophils reflect more chronic inflammation and could not have occurred in less than about a week. Unsuspected splenic rupture would likely lead to hypotension and shock. There is no reason to consider a diagnosis of chronic myeloproliferative disease or sepsis in the absence of other evidence for a secondary condition. The patient should have his counts rechecked in several days or weeks to rule out any other pathology. (See Stress.)

4. A 3-year-old child presents with recurrent infections and a diagnosis of Chédiak-Higashi syndrome (CHS) is made. Which of the following is *not* true of his disease?

A. He is likely to have associated nystagmus.
B. His sibling has a 50% chance of having the disease.
C. He is likely to have abnormal neutrophil granules.
D. He is likely to develop hemophagocytic syndrome.
E. He is likely to have a defect in skin pigmentation.

Answer: A CHS is a generalized disorder of granule formation caused by abnormalities in the *LYST* gene, which is required for the correct trafficking of proteins into granules. Consequently, patients with CHS have other defects associated with abnormal granule formation, including oculocutaneous albinism, with decreased pigmentation and nystagmus, and abnormal-appearing neutrophil granules. As the disease is an autosomal recessive disorder, the heterozygous parents are usually totally unaffected, and siblings have a 25% chance of inheriting the disease. Patients with CHS often have a terminal phase of the disease associated with the development of hemophagocytic syndrome, manifesting as fever, splenomegaly, and pancytopenia, often as a result of Epstein-Barr virus infection. (See Other Congenital Syndromes with Associated Neutropenia.)

5. A 1-month-old infant presents with a temperature of 103° F and a white blood cell count (WBC) of 90,000. He has a history of delayed umbilical cord separation, poorly healing skin lesions, recurrent otitis media, and failure to thrive. He is admitted to the hospital, where he is documented to have methicillin-resistant *Staphylococcus aureus* (MRSA). He is placed on appropriate antibiotics, but he remains febrile, and his WBC is persistently elevated in the 80,000 to 100,000 range. This patient is likely to have a mutation in which of the following?

A. *ELANE*
B. *HAX1*
C. Integrin receptor β chain
D. *LYST*
E. *G-CSFR*

Answer: C The child has a congenital disorder of neutrophils that results in leukocytosis. The diagnosis is leukocyte adhesion deficiency. This disease arises from defects in leukocyte adhesion to endothelium. It can arise from defects in the integrin receptor common β chain (LAD-1), with loss of expression of leukocyte function–associated antigen 1 (LFA-1), C3bi receptor, and gp150;95, leading to a failure to ingest and kill microbes opsonized by C3bi. It can also arise from an abnormality of selectin glycosylation (LAD-2). *ELANE* and *HAX1* mutations are causes of severe congenital neutropenia, and affected patients present with profound neutropenia. *LYST* mutations are the cause of Chédiak-Higashi syndrome, which is also associated with neutropenia. Congenital *G-CSFR* mutations are very rare, although two or three patients in the literature have been described to have such mutations as the basis for G-CSF–resistant severe congenital neutropenia. More commonly, *G-CSFR* mutations are somatic mutations that arise in patients with severe congenital neutropenia and may be a harbinger of the development of myelodysplastic syndrome or acute myeloid leukemia. (See Leukocyte Adhesion Deficiency.)

168

APPROACH TO THE PATIENT WITH LYMPHADENOPATHY AND SPLENOMEGALY

JAMES O. ARMITAGE AND PHILIP J. BIERMAN

LYMPHADENOPATHY

PHYSIOLOGY AND ANATOMY

Lymph nodes are found throughout the body along the course of the lymphatic vessels, strategically located to allow the filtering of lymphatic fluid and the interdiction of microorganisms and abnormal proteins. Lymphatic fluid enters the node in afferent lymphatic vessels that empty into the subcapsular sinus. The fluid then transverses the node and exits in a single efferent lymphatic vessel. In doing so, the lymph and its contents are exposed to immunologically active cells throughout the node. Lymph nodes are populated predominantly by macrophages, dendritic cells, B lymphocytes, and T lymphocytes. B lymphocytes are located primarily in the follicles and perifollicular areas, whereas T lymphocytes are found principally in the interfollicular or paracortical areas of the lymph node. These cells function together to provide antigen processing, antigen presentation, antigen recognition, and proliferation of effector B and T lymphocytes as part of the normal immune response to microorganisms or foreign proteins.

Because the normal immune response leads to the proliferation and expansion of one or more cellular components of lymph nodes, it often results in significant lymph node enlargement. In young children, who are continuously being exposed to new antigens, palpable lymphadenopathy is the rule. In fact, the absence of palpable lymphadenopathy in them would be considered abnormal. In adults, lymph nodes larger than 1 to 2 cm in diameter are generally considered abnormal. However, lymph nodes 1 to 2 cm in diameter in the groin are sufficiently common to be considered normal.

Lymphoid proliferation is a normal response to exposure to foreign antigens. The location of the enlarged lymph nodes often reflects the site of invasion. For example, cervical lymphadenopathy would be typical in a patient with pharyngitis. Generalized immune proliferation and lymphadenopathy can occur with a systemic disorder of the immune system, disseminated infection, or disseminated neoplasia. Malignancies of the immune system might manifest as localized or disseminated lymphadenopathy.

DIAGNOSIS

Differential Diagnosis

The differential diagnosis of lymphadenopathy (Table 168-1) is vast, and the underlying causes are responsible for either proliferation of immunologically active cells or infiltration of the lymph node by foreign cells or substances. In practice, the cause of enlarged lymph nodes is often uncertain even in retrospect; in such cases, unrecognized infectious processes are generally blamed.

Infections by bacteria, mycobacteria, fungi, chlamydia, parasites, and viruses are the major causes of lymph node enlargement. Lymph nodes in the drainage area of essentially all pyogenic infections can enlarge. In certain infections, such as bubonic plague caused by *Yersinia pestis,* dramatic regional lymph node enlargement with fluctuant lymph nodes (i.e., buboes) can be a hallmark of the disease (Chapter 312). Other bacterial infections have lymph node enlargement as a prominent feature (e.g., cat-scratch disease; Chapter 315) and can mimic lymphoproliferative disorders. Mediastinal lymphadenopathy is seen in inhalational anthrax (Chapter 294). In some parts of the world, cervical lymphadenopathy is a sufficiently frequent manifestation of tuberculosis to lead to the institution of antituberculous therapy rather than biopsy. Disseminated lymphadenopathy can be seen in cases of infection by a variety of organisms such as *Toxoplasma,* Epstein-Barr virus (i.e., infectious mononucleosis), cytomegalovirus, and human immunodeficiency virus (HIV).

TABLE 168-1 CAUSES OF LYMPHADENOPATHY

Infection
Bacterial (e.g., all pyogenic bacteria, cat-scratch disease, syphilis, tularemia)
Mycobacterial (e.g., tuberculosis, leprosy)
Fungal (e.g., histoplasmosis, coccidioidomycosis)
Chlamydial (e.g., lymphogranuloma venereum)
Parasitic (e.g., toxoplasmosis, trypanosomiasis, filariasis)
Viral (e.g., Epstein-Barr virus, cytomegalovirus, rubella, hepatitis, HIV)
Benign disorders of the immune system (e.g., rheumatoid arthritis, systemic lupus erythematosus, serum sickness, drug reactions such as to phenytoin, Castleman disease, sinus histiocytosis with massive lymphadenopathy, Langerhans cell histiocytosis, Kawasaki syndrome, Kimura disease)
Malignant disorders of the immune system (e.g., chronic and acute myeloid and lymphoid leukemia, non-Hodgkin lymphoma, Hodgkin disease, angioimmunoblastic-like T-cell lymphoma, Waldenström macroglobulinemia, multiple myeloma with amyloidosis, malignant histiocytosis)
Other malignancies (e.g., breast carcinoma, lung carcinoma, melanoma, head and neck cancer, gastrointestinal malignancies, germ cell tumors, Kaposi sarcoma)
Storage diseases (e.g., Gaucher disease, Niemann-Pick disease)
Endocrinopathies (e.g., hyperthyroidism, adrenal insufficiency, thyroiditis)
Miscellaneous (e.g., sarcoidosis, amyloidosis, dermatopathic lymphadenitis, IgG4-related disease)

TABLE 168-2 MOST FREQUENT CAUSES OF LYMPHADENOPATHY IN ADULTS IN THE UNITED STATES

Unexplained
Infection
In drainage area of infection (e.g., cervical adenopathy with pharyngitis)
Disseminated (e.g., mononucleosis, HIV infection)
Immune disorders (e.g., rheumatoid arthritis)
Neoplasms
Immune system malignancies (e.g., leukemias, lymphomas)
Metastatic carcinoma or sarcoma

TABLE 168-3 FACTORS TO CONSIDER IN THE DIAGNOSIS OF LYMPHADENOPATHY

Associated systemic symptoms
Patient's age
History of infection, trauma, medications, travel experience, previous malignancy
Location: cervical, supraclavicular, epitrochlear, axillary, intrathoracic (hilar vs. mediastinal), intra-abdominal (retroperitoneal vs. mesenteric vs. other), iliac, inguinal, femoral
Localized vs. disseminated
Presence of tenderness or inflammation
Size
Consistency

A variety of nonmalignant disorders of the immune system can lead to localized or disseminated lymphadenopathy. Autoimmune diseases such as rheumatoid arthritis (Chapter 264) and systemic lupus erythematosus (Chapter 266) often have accompanying lymphadenopathy, which can pose a diagnostic challenge because of the increased incidence of lymphoma in patients with these disorders. In the lymphadenopathy that occurs as a reaction to drugs such as phenytoin, lymph node biopsy findings can sometimes be confused with lymphoma. Benign proliferative diseases of the immune system that can also be confused with lymphoma include Castleman disease (Chapters 185 and 393; angiofollicular lymph node hyperplasia), sinus histiocytosis with massive lymphadenopathy, and disorders seen more frequently in Asia, such as Kawasaki syndrome (Chapter 439) and Kimura disease.

All the cells in the immune system can become malignant. Several of these malignancies typically manifest as lymphadenopathy, and it can be seen in all of them. Lymphadenopathy as the initial manifestation is the rule for Hodgkin disease and non-Hodgkin lymphoma, and it is common in Waldenström macroglobulinemia and B-cell chronic lymphocytic leukemia; it is seen only occasionally in the myeloid leukemias (Chapters 183 through 186) and is rare in multiple myeloma. Malignancies of all organ systems can metastasize to the lymph nodes and cause lymphadenopathy, which is usually seen in the drainage area of the primary tumor—for example, axillary lymph nodes in patients with breast cancer, hilar and mediastinal lymph nodes in patients with lung cancer, and cervical lymph nodes in patients with head and neck cancer. However, widespread lymphadenopathy can also occur. Other disorders in which lymphadenopathy may be an initial finding include storage diseases such as Gaucher disease (Chapter 208), endocrinopathies such as hyperthyroidism (Chapter 226), sarcoidosis (Chapter 95), and dermatopathic lymphadenitis. Amyloidosis (Chapter 188) can cause lymphadenopathy in patients with multiple myeloma, hereditary amyloidosis, or amyloidosis associated with chronic inflammatory states.

Among patients with lymphadenopathy actually seen in practices in the United States, diagnoses are not determined in a high proportion (Table 168-2). In such cases, the lymphadenopathy is usually blamed on infection. When the lymphadenopathy is in the drainage site of a known infection (e.g., cervical lymphadenopathy in a patient with pharyngitis) or the patient has a known infection associated with lymphadenopathy (e.g., infectious mononucleosis; Chapter 377), this infectious assumption is usually correct. Alternatively, if a patient has an immunologic disorder that is known to cause lymphadenopathy, such as rheumatoid arthritis, this disorder is usually an acceptable explanation; however, progressive lymphadenopathy in such patients should trigger a biopsy because they are at increased risk for lymphoma. Localized, progressive lymphadenopathy, particularly when associated with fever, sweats, or weight loss, requires biopsy to exclude lymphoma.

Lymph Node Evaluation

Evaluation of a patient with lymphadenopathy includes a careful history, a thorough physical examination, laboratory tests, and sometimes imaging studies to determine the extent and character of the lymphadenopathy. The age of the patient and any associated systemic symptoms might be important clues (Table 168-3). Cervical lymphadenopathy in a child is much less worrisome than equally prominent lymphadenopathy in a 60-year-old adult. The occurrence of fever, sweats, or weight loss raises the possibility of a malignancy of the immune system. The explanation for the lymphadenopathy might become apparent with the identification of a site of infection, a particular medication, a travel history, or a previous malignancy.

Physical examination allows the identification of localized versus widespread lymphadenopathy. The particular sites of involvement can be important hints to the diagnosis because infection and carcinoma are likely to cause lymphadenopathy in the lymphatic drainage of the site of the disorder. In general, tender lymph nodes are more likely to be due to an infectious process, whereas painless adenopathy raises concern for malignancy. Lymph node consistency can also aid in the diagnosis: typically, lymph nodes containing metastatic carcinoma are rock hard, lymph nodes containing lymphoma are firm and rubbery, and lymph nodes enlarged in response to an infectious process are soft.

The larger the lymph node, the more likely it is that a serious underlying cause exists; lymph nodes greater than 3 to 4 cm in diameter in an adult are very worrisome. Physical examination to assess lymph node size is only marginally accurate and reproducible, although it is by far the most widely used method. More precise methods are available with various imaging techniques.

Imaging

Imaging studies, including routine radiographs, computed tomography (CT), ultrasonography, magnetic resonance imaging (MRI), and positron emission tomography (PET), can be used to assess the extent of lymphadenopathy in the chest and abdomen (Table 168-4). Chest radiographs are the most economical and easiest way to assess mediastinal and hilar lymphadenopathy but are not as accurate as CT of the chest. CT and ultrasonography are the most useful modalities for assessing abdominal and retroperitoneal lymphadenopathy. In most patients, CT is probably the most accurate approach, but ultrasonography has the advantage of being less expensive and not requiring radiation exposure. MRI and PET are not first-line studies for the assessment of lymphadenopathy. Although few randomized controlled trials have been published to date concerning the diagnostic accuracy of PET,[1] it is applied to the assessment of patients with lymphoma both at presentation and after treatment.[2] PET is usually positive in patients with Hodgkin disease and aggressive non-Hodgkin lymphomas and can be used to assess the presence of active lymphoma in patients with lymphadenopathy and a proven diagnosis; it is especially useful for re-evaluating patients after therapy because lymph nodes do not always regress to normal size after treatment, particularly those in the mediastinum and retroperitoneum, even though the malignancy has been eradicated.

TABLE 168-4 METHODS OF LYMPH NODE EVALUATION

Physical examination
Imaging
Chest radiography
Ultrasonography
Computed tomography
Magnetic resonance imaging
Positron emission tomography
Sampling
Needle aspiration
Cutting needle biopsy
Excisional biopsy

TABLE 168-5 APPROACH TO THE PATIENT WITH LYMPHADENOPATHY

Does the patient have a known illness that causes lymphadenopathy? Treat and monitor for resolution.
Is there an obvious infection to explain the lymphadenopathy (e.g., infectious mononucleosis)? Treat and monitor for resolution.
Are the nodes very large and/or very firm and thus suggestive of malignancy? Perform a biopsy.
Is the patient very concerned about malignancy and unable to be reassured that malignancy is unlikely? Perform a biopsy.
If none of the preceding is true, perform a complete blood cell count and, if unrevealing, monitor for a predetermined period (usually 2 to 8 weeks). If the nodes do not regress or if they increase in size, perform a biopsy.

TABLE 168-6 SOME DIAGNOSTIC CONSIDERATIONS FOR LOCALIZED LYMPHADENOPATHY

SITE	INFECTIONS	NEOPLASMS	OTHER
Cervical	Pharyngitis, other head and neck infections, mononucleosis, toxoplasmosis, TB	Head and neck cancers, thyroid cancer, lymphoma	
Supraclavicular		Intra-abdominal cancer (particularly left-sided nodes), lung cancer, lymphoma	
Axillary	Cat-scratch disease, distal infections, plague	Breast cancer, melanoma, lymphoma	Silicone implants
Mediastinal	TB, fungal infection, anthrax	Lymphoma, lung cancer, germ cell tumor	Sarcoidosis
Retroperitoneal	TB	Lymphoma, testicular cancer, kidney cancer, upper GI malignancy	Sarcoidosis
Mesenteric	Appendicitis, cholecystitis, diverticulitis, Whipple disease	Lymphoma, GI cancer	Inflammatory bowel disease, panniculitis
Inguinal	Distal or genital infection, plague, STDs	Lymphoma, melanoma, vulvar cancer	

GI = gastrointestinal; STD = sexually transmitted disease; TB = tuberculosis.

Interventional Evaluation

Lymph node aspiration or biopsy is often necessary for an accurate diagnosis of the cause of lymphadenopathy. Even then, morphologic findings alone may not be able to clearly distinguish between reactive changes in benign lymphadenopathies and neoplasm, particularly lymphoma; special studies on the specimens may be required to do so.[3] Fine-needle aspiration is currently popular and is an accurate means of diagnosing infection or carcinoma involving a lymph node. Although lymphomas can occasionally be diagnosed with this approach, it is inappropriate as an initial diagnostic maneuver for lymphoma. Cutting needle biopsy often provides sufficient material for an unequivocal diagnosis and subtyping of the lymphoma. However, excisional biopsy, which is most likely to provide the pathologist with adequate material to perform histologic, immunologic, and genetic studies, is also most likely to yield the correct diagnosis.

An Approach to the Patient with Lymphadenopathy

Patients with lymphadenopathy (Table 168-5) come to medical attention in several ways. Perhaps most common is a patient who feels a lymph node in the neck, axilla, or groin and seeks a physician's opinion. Lymphadenopathy might also be an unexpected finding on a routine physical examination or as part of an evaluation for another complaint. Finally, patients might be found to have unexpected lymphadenopathy on imaging studies of the chest or abdomen. When the nodes are multiple or larger than 2 to 3 cm, biopsy using mediastinoscopy, a paramediastinal incision, laparoscopy, or laparotomy is often required for diagnosis.

The approach to a patient complaining of newly discovered lymphadenopathy in the neck, axilla, or groin depends on the size, consistency, and number of enlarged lymph nodes and the patient's general health. In most cases, very large or very firm lymph nodes in the presence of systemic symptoms such as unexplained fever, sweats, or weight loss should lead to lymph node biopsy. Patients who have enlarged lymph nodes in the drainage area of a previously treated malignancy (e.g., neck nodes in a patient with a history of head and neck cancer) might be best approached by lymph node aspiration. Carcinoma can often be diagnosed in this manner, although it is a poor approach for the diagnosis of lymphoid malignancies. For cervical lymph nodes,[4] excisional biopsy should be delayed in a patient in whom head and neck cancer (Chapter 190) is a diagnostic consideration. These patients should initially undergo careful ear, nose, and throat examinations to avoid performing a biopsy that might complicate the patient's subsequent therapy. Some diagnostic possibilities for localized lymphadenopathy are presented in Table 168-6.

In the most common situation—that is, a lymph node is soft and is not larger than 2 to 3 cm and the patient has no obvious systemic illness—observation for a brief period is usually the best approach. Performance of a complete blood cell count and examination of a peripheral smear can be helpful in recognizing a systemic illness (e.g., infectious mononucleosis). These patients are often given antibiotics. If the lymph node does not regress over the course of a few weeks or if it gets bigger, a biopsy should be performed.

Part of the care of such patients involves the art of medicine and being responsive to the patient's particular needs. For example, biopsy might be performed more quickly in a patient who is very anxious about malignancy or who needs a definitive diagnosis expeditiously.

SPLENOMEGALY

DEFINITION

The spleen is the largest lymphatic organ in the body and is sometimes approached clinically as though it were a very large lymph node. Although it participates in the primary immune response to invading microorganisms and foreign proteins, the spleen has many other functions. It functions as a filter for the blood and is responsible for removing senescent red blood cells from the circulation, as well as blood cells and other cells coated with immunoglobulins. Blood enters the spleen, filters through the splenic cords, and is exposed to immunologically active cells in the spleen.

The splenic red pulp occupies more than half the volume of the spleen and is the site where senescent red cells are identified and destroyed and red blood cell inclusions are removed by a process known as *pitting*. In the absence of splenic function, basophilic inclusions known as *Howell-Jolly bodies* are seen in circulating red blood cells. The presence of Howell-Jolly bodies (Fig. 168-1) in peripheral blood indicates that the patient has undergone splenectomy or has a process that has rendered the spleen nonfunctional (e.g., sickle cell disease with repeated splenic infarcts, chronic graft-versus-host disease).

The white pulp of the spleen contains macrophages, B lymphocytes, and T lymphocytes; participates in the recognition of microorganisms and foreign proteins; and is involved in the primary immune response. Absence of this splenic function makes individuals particularly susceptible to certain

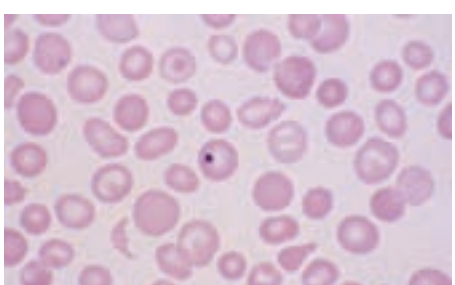

FIGURE 168-1. Howell-Jolly body in an erythrocyte. This is evidence of splenectomy or a nonfunctional spleen.

infections, including sepsis with encapsulated organisms such as *Streptococcus pneumoniae.* The risk for overwhelming sepsis is related to the age at the time of splenectomy or other cause of loss of splenic function. Children and young adults are at highest risk. If possible, all patients should undergo vaccination against *S. pneumoniae* and perhaps *Haemophilus influenzae* and *Neisseria meningitidis* before splenectomy. Some physicians have patients take oral penicillin (e.g., phenoxymethyl penicillin, 250 mg twice daily) indefinitely if splenectomy has been performed in childhood or adolescence.

PATHOBIOLOGY

As with lymphadenopathy, numerous conditions are associated with splenomegaly (Table 168-7). Certain bacterial infections such as endocarditis (Chapter 76), brucellosis (Chapter 310), and typhoid fever (Chapter 308) have splenomegaly as a frequent manifestation. Disseminated tuberculosis (Chapter 324) is often associated with splenomegaly, and splenomegaly can also be seen in cases of disseminated histoplasmosis (Chapter 332) and toxoplasmosis (Chapter 349). Splenomegaly is an almost constant accompaniment of malaria (Chapter 345). Rickettsial disorders such as Rocky Mountain spotted fever are frequently associated with splenomegaly. A wide variety of viral infections typically cause splenomegaly, including infectious mononucleosis associated with Epstein-Barr virus (Chapter 377) and viral hepatitis (Chapters 148 and 149). Splenomegaly can accompany HIV infection. Splenic abscesses, which are usually the result of hematogenous spread of pyogenic organisms, represent an unusual and difficult to diagnose cause of splenomegaly.

Splenomegaly is also seen in a variety of benign disorders of the immune system, including rheumatoid arthritis (Chapter 264); some of these patients have Felty syndrome and accompanying granulocytopenia. Splenomegaly can be detected in some patients with systemic lupus erythematosus (Chapter 266), certain drug reactions, and serum sickness.

Malignancies of the immune system and nonimmune organs can also lead to splenomegaly. Splenomegaly is usually seen in patients with chronic myeloid leukemia and is frequent in chronic lymphoid leukemia (Chapter 184). It can develop in patients with acute myeloid or lymphoid leukemia, non-Hodgkin lymphoma, Hodgkin disease, and Waldenström macroglobulinemia but is rare in multiple myeloma (Chapter 187). Isolated splenomegaly (i.e., without any enlarged lymph nodes) is characteristic of certain immune system malignancies, including hairy cell leukemia (Chapter 184), the prolymphocytic variant of chronic lymphocytic leukemia (Chapter 184), and splenic marginal zone lymphoma (Chapter 185). Metastasis of carcinomas and sarcomas to the spleen is unusual except for malignant melanoma; even with melanoma, however, palpable splenomegaly is an unusual finding.

Splenomegaly can develop as a result of increased pressure in the splenic circulation, especially in patients with portal hypertension caused by a variety of hepatic disorders, including alcoholic cirrhosis (Chapter 152). However, it also can be due to splenic or portal vein thrombosis. The first manifestation of an enlarged spleen in portal hypertension can be thrombocytopenia.

Hematologic disorders that can lead to palpable splenomegaly include autoimmune hemolytic anemia (Chapter 160), hereditary spherocytosis (Chapter 161), and a number of other anemias. The myeloproliferative neoplasms polycythemia vera (frequently), essential thrombocythemia (sometimes), and idiopathic myelofibrosis (usually) all can present with splenomegaly. In cases of idiopathic myelofibrosis, the spleen is frequently a site of extramedullary hematopoiesis (Chapter 166).

A variety of less common conditions can lead to splenomegaly. The storage disorder Gaucher disease (Chapter 208) usually manifests as splenomegaly.

TABLE 168-7 CAUSES OF SPLENOMEGALY

Infection
Bacterial (e.g., endocarditis, brucellosis, syphilis, typhoid, pyogenic abscess)
Mycobacterial (e.g., tuberculosis)
Fungal (e.g., histoplasmosis, toxoplasmosis)
Parasitic (e.g., malaria, leishmaniasis)
Rickettsial (e.g., Rocky Mountain spotted fever)
Viral (e.g., Epstein-Barr virus, cytomegalovirus, HIV, hepatitis)
Benign disorders of the immune system (e.g., rheumatoid arthritis with Felty syndrome, systemic lupus erythematosus, drug reactions such as to phenytoin, Langerhans cell histiocytosis, serum sickness)
Malignant disorders of the immune system (e.g., acute or chronic myeloid or lymphoid leukemia, non-Hodgkin lymphoma, Hodgkin disease, Waldenström macroglobulinemia, malignant histiocytosis)
Other malignancies (e.g., melanoma, sarcoma)
Congestive splenomegaly (e.g., portal hypertension secondary to liver disease, splenic or portal vein thrombosis)
Hematologic disorders (e.g., autoimmune hemolytic anemia, hereditary spherocytosis, thalassemia major, hemoglobinopathies, elliptocytosis, extramedullary hematopoiesis)
Storage diseases (e.g., Gaucher disease)
Endocrinopathies (e.g., hyperthyroidism)
Miscellaneous (e.g., sarcoidosis, amyloidosis, tropical splenomegaly, cysts)

TABLE 168-8 METHODS FOR EVALUATING THE SPLEEN

Physical examination
Imaging
Ultrasonography
Computed tomography
Liver-spleen scanning
Positron emission tomography
Biopsy
Needle aspiration
Splenectomy
Laparotomy (total or partial splenectomy)
Laparoscopy

Splenomegaly can be seen in endocrinopathies such as hyperthyroidism (Chapter 226). Sarcoidosis (Chapter 95) and amyloidosis (Chapter 188) can manifest as splenomegaly. *Tropical splenomegaly* is a term used to describe the palpable spleens found in patients who live in tropical areas, for which there may be numerous causes.

DIAGNOSIS

Evaluation of Spleen Size and Function

Physical Examination

The ability to perform an accurate physical examination and determine the presence of an enlarged spleen (Table 168-8) is an important skill, but it is not easily learned. Physical examination of the spleen can be performed with the patient supine or in the right lateral decubitus position. Inspection, percussion, auscultation, and palpation are all important parts of an accurate assessment. It is rare for a spleen to be so large that it is visible and can be seen to move with respiration. However, in patients with such a condition, it is possible to miss splenomegaly by failing to start palpation sufficiently low to find the edge. Occasionally, percussion of the left upper quadrant helps identify an area of dullness that moves with respiration and can lead to the identification of splenomegaly. Spleen size is generally recorded as the number of centimeters the spleen descends below the left costal margin in the midclavicular line on inspiration. Although auscultation is not usually a regular part of splenic examination, the existence of a splenic rub on inspiration can lead to the diagnosis of splenic infarction. The left kidney is sometimes confused with the spleen on physical examination, but its failure to move with respiration in the manner typical for the spleen usually allows its distinction.

Laboratory Evaluation

Laboratory studies are frequently valuable in assessing splenic function. In patients with an absent or nonfunctional spleen,[5,6] Howell-Jolly bodies will be seen in circulating red blood cells (see Fig. 168-1). Splenic hyperfunction (a condition often referred to as *hypersplenism*) is associated with cytopenias: the spleen is the normal reservoir for a significant proportion of platelets, and

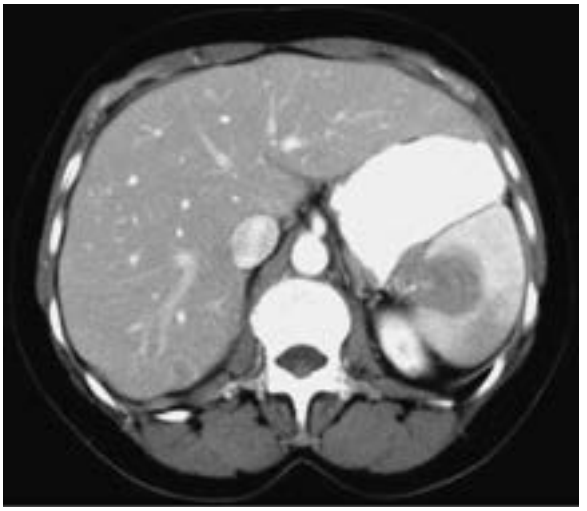

FIGURE 168-2. Enlarged spleen with metastatic adenocarcinoma.

this reservoir function can lead to thrombocytopenia in patients with splenomegaly. Patients with autoimmune hemolytic anemia usually have palpable splenomegaly, but patients with idiopathic (immune) thrombocytopenic purpura usually do not.

The spleen can be imaged with ultrasonography, CT,[7] traditional radionuclide scans, and PET (Fig. 168-2). Ultrasonography can provide an accurate determination of spleen size and is easy to repeat. CT frequently gives a better view of the consistency of the spleen and can identify splenic tumors or abscesses that would otherwise be missed. PET can aid in evaluating focal lesions in the spleen. The technetium-labeled liver-spleen scan can be important in identifying liver disease as the cause of splenomegaly; in patients with cryptogenic cirrhosis who are found to have thrombocytopenia, a technetium liver-spleen scan that shows higher activity in the spleen than in the liver might be the first hint of liver disease.

Because of the spleen's location and its propensity to bleed, needle aspiration or cutting needle biopsy of the spleen is rarely performed. In general, splenic "biopsy" involves splenectomy, which can be performed at the time of laparotomy or by laparoscopy. However, performing splenectomy laparoscopically usually leads to maceration of the organ and can reduce the diagnostic information. In very young children, in whom splenectomy leads to a high risk for serious infections such as pneumococcal septicemia, partial splenectomy can sometimes be performed. Patients who undergo splenectomy at the time of splenic trauma and rupture may have seeding of splenic cells to other sites in the abdomen (i.e., splenosis). Some patients have additional small or accessory spleens. Persistent, functional splenic tissue can be the explanation for recurrent immune thrombocytopenia after splenectomy and might be recognized by the absence of Howell-Jolly bodies in circulating red blood cells. Patients whose spleen has been removed often have thrombocytosis.

An Approach to the Patient with Splenomegaly

Patients with splenomegaly (Table 168-9) may come to medical attention for a variety of reasons. Patients may complain of left upper quadrant pain or fullness or early satiety. A splenic infarct, which typically manifests as left upper quadrant pain that sometimes radiates to the left shoulder, can be the first clue to the existence of an enlarged spleen. Rarely, splenomegaly initially manifests with the catastrophic symptoms of splenic rupture. Some patients are found to have splenomegaly as a result of evaluation for unexplained cytopenia. Splenomegaly can also be discovered incidentally on physical examination. In recent years, splenomegaly has frequently been discovered on imaging studies of the abdomen performed for other purposes.

The presence of a palpable spleen on physical examination is almost always abnormal. The one exception to this rule is a palpable spleen tip in a slender young woman. In general, the presence of a palpable spleen should be considered a serious finding, and an explanation should be sought. It is less clear whether the same rules apply to borderline splenomegaly discovered incidentally on routine imaging studies.

The approach to a patient with an enlarged spleen should focus initially on excluding a systemic illness that could explain the splenomegaly. Infectious mononucleosis, leukemia or lymphoma, rheumatoid arthritis, sarcoidosis,

TABLE 168-9 APPROACH TO THE PATIENT WITH SPLENOMEGALY

Does the patient have a known illness that causes splenomegaly (e.g., infectious mononucleosis)? Treat and monitor for resolution.
Search for an occult infection (e.g., infectious endocarditis), hematologic disorder (e.g., hereditary spherocytosis), occult liver disease (e.g., cryptogenic cirrhosis), autoimmune disease (e.g., systemic lupus erythematosus), or storage disease (e.g., Gaucher disease). If found, manage appropriately.
If systemic symptoms are present and suggest malignancy, focal replacement of the spleen is seen on imaging studies, and no other site is available for biopsy, splenectomy is indicated.
If none of the above is true, monitor closely and repeat studies until the splenomegaly resolves or a diagnosis becomes apparent.

cirrhosis of the liver, malaria, and a host of other illnesses would be reasonable explanations for the splenomegaly. The systemic condition should be treated, and then the spleen should be re-evaluated. If the systemic illness can be treated successfully, the spleen should regress to normal size over time.

Patients with no obvious explanation for an enlarged spleen present a difficult diagnostic problem. Careful follow-up of these patients sometimes reveals occult liver disease or an autoimmune process that initially defied diagnosis. Concerns about malignancy, particularly in patients with systemic symptoms such as fever, sweats, or weight loss or in whom imaging studies show a focal abnormality, are sometimes indications for splenectomy. However, in the absence of such findings, it is generally preferable to monitor patients closely with repeated attempts to establish the diagnosis by approaches other than splenectomy. It is particularly important to avoid splenectomy in a patient with occult liver disease and portal hypertension.

Splenectomy was once performed routinely as part of the staging evaluation for Hodgkin disease or other lymphomas. Today, this procedure is rarely needed to choose the correct therapy, and it should be avoided. Splenectomy[8,9] can be an effective therapy for immune thrombocytopenic purpura (Chapter 172) and autoimmune hemolytic anemia (Chapter 160), and it is occasionally an appropriate therapy to relieve cytopenias in other conditions such as advanced myelofibrosis (Chapter 166). Radiofrequency ablation is an alternative to surgical removal.■ Splenectomized patients are at increased risk for infections, especially Gram-positive infections, and require careful follow-up[10,11] (Chapter 281).

Grade A References

A1. Feng K, Ma K, Liu Q, et al. Randomized clinical trial of splenic radiofrequency ablation versus splenectomy for severe hypersplenism. *Br J Surg*. 2011;98:354-361.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Scheibler F, Zumbé P, Janssen I, et al. Randomized controlled trials on PET: a systematic review of topics, design, and quality. *J Nucl Med*. 2012;53:1016-1025.
2. Ramsay AD. 30 Years of lymph node pathology: biomarkers and other advances. *Appl Immunohistochem Mol Morphol*. 2013;21:103-109.
3. Weiss LM, O'Malley D. Benign lymphadenopathies. *Modern Pathol*. 2013;26:588-596.
4. Bryson TC, Shah GV, Srinivasan A, et al. Cervical lymph node evaluation and diagnosis. *Otolaryngol Clin North Am*. 2012;45:1363-1383.
5. Lammers AJJ, de Porto APNA, Bennink RJ, et al. Hyposplenism: comparison of different methods for determining splenic function. *Am J Hematol*. 2012;87:484-489.
6. Di Sabatino A, Carsetti R, Corazza GR. Post-splenectomy and hyposplenic states. *Lancet*. 2011;378:86-97.
7. Harris A, Kamishima T, Hao HY, et al. Splenic volume measurements on computed tomography utilizing automatically contouring software and its relationship with age, gender, and anthropometric parameters. *Eur J Radiol*. 2010;75:e97-e101.
8. Zhan XL, Ji Y, Wang YD. Laparoscopic splenectomy for hypersplenism secondary to liver cirrhosis and portal hypertension. *World J Gastroenterol*. 2014;20:5794-5800.
9. Fan Y, Wu SD, Kong J, et al. Feasibility and safety of single-incision laparoscopic splenectomy: a systematic review. *J Surg Res*. 2014;186:354-362.
10. Rubin LG, Schaffner W. Clinical practice. Care of the asplenic patient. *N Engl J Med*. 2014;371:349-356.
11. Yong M, Thomsen RW, Schoonen WM, et al. Mortality risk in splenectomised patients: a Danish population-based cohort study. *Eur J Intern Med*. 2010;21:12-16.

REVIEW QUESTIONS

1. Which of the following is the most likely diagnosis in a patient with a 4-cm firm, nubby, nontender right lower cervical lymph node in an otherwise well 26-year-old female?

A. Infectious mononucleosis
B. Systemic lupus erythematosus (SLE)
C. HIV infection
D. Hodgkin lymphoma
E. Cat-scratch disease

Answer: D When lymphadenopathy occurs in mononucleosis, HIV infection, or SLE, it tends to be generalized. The lymphadenopathy of cat-scratch disease can mimic that of a lymphoproliferative disorder but tends not to be as large and firm as in the case described here. Hodgkin lymphoma is the most likely diagnosis among these choices in an otherwise healthy young patient with a single prominent and firm enlarged node. (See Differential Diagnosis under Lymphadenopathy.)

2. After splenectomy, in addition to Howell-Jolly bodies, a peripheral smear is most likely to show which of the following?

A. Thrombocytosis
B. Toxic granulation of the neutrophils
C. Spherocytes
D. Dutcher bodies
E. Nucleated red blood cells

Answer: A A reactive thrombocytosis is often seen after splenectomy, and it may persist for months or years. Toxic granulations in neutrophils are unrelated to removal of the spleen; they are seen in infections and inflammatory states regardless of spleen status. The other morphologic findings in blood cells are not modified according to whether or not there is an intact spleen. (See Laboratory Evaluation under Splenomegaly; also see Figure 168-1.)

3. Lymph node enlargement would be unusual in a patient with which of the following?

A. Follicular lymphoma
B. Tuberculosis
C. Castleman disease
D. Waldenström macroglobulinemia
E. Multiple myeloma

Answer: E Localized or disseminated lymphadenopathy can be caused by malignant lymphoproliferative diseases (including follicular lymphoma and Waldenström macroglobulinemia; benign proliferative diseases of the immune system (including Castleman disease); and infections by bacteria, mycobacteria (including tuberculosis, fungi, chlamydia, parasites, and viruses). In contrast, multiple myeloma is a plasma cell dyscrasia (as opposed to Waldenström macroglobulinemia, which is actually a malignant lymphoproliferative disorder associated with a paraprotein) and is not characterized by lymphadenopathy. (See Differential Diagnosis under "Lymphadenopathy.)

4. Fine-needle aspiration (FNA) of a lymph node is a good way to diagnose which of the following?

A. Castleman disease
B. Hodgkin lymphoma
C. Sarcoidosis
D. Metastatic tonsil carcinoma
E. Post-transplantation lymphoproliferative disorder

Answer: D FNA is a popular and usually accurate means of diagnosing an infection or carcinoma involving a lymph node, like metastatic tonsil carcinoma. However, unequivocal diagnosis and histopathologic subtyping of lymphomas (including Hodgkin and non-Hodgkin lymphoma, such as post-transplantation lymphoproliferative disorder) or granulomatous disease (such as the noncaseating granulomas of sarcoidosis) require cutting needle or excisional lymph node biopsy, which preserves histologic architecture. (See Interventional Evaluation under Lymphadenopathy.)

5. Splenectomy early in life makes one particularly vulnerable to which of the following?

A. Pneumococcal sepsis
B. Bacterial endocarditis
C. Salmonella sepsis
D. Herpetic encephalitis
E. Severe influenza

Answer: A Splenectomy early in life makes individuals particularly susceptible to sepsis with encapsulated organisms such as *Streptococcus pneumoniae.* However, there is little if any strong evidence for susceptibility to overwhelming systemic infections with other types of bacteria or viruses. This is the rationale for the recommendations for vaccination against *S. pneumoniae* and possibly also *Haemophilus influenzae* and *Neisseria meningitidis* before splenectomy. (See Definition under Splenomegaly.)

169

DISORDERS OF PHAGOCYTE FUNCTION

MICHAEL GLOGAUER

Neutrophils and monocyte-macrophages are the key phagocytes of the innate immune system. Their principal innate immune role is to recognize and eliminate microorganisms that make their way past primary physical barriers, such as the epithelium and body secretions that protect the external and lining surfaces of the body. The neutrophil-pathogen interaction generates an army of antimicrobial mediators that results in efficient killing of pathogens.[1] Phagocytes identify foreign invaders through a series of pattern recognition receptors, most of which belong to the toll-like receptor family (Chapters 45 and 48). Whereas macrophages carry out sentinel duty looking for microbes in healthy tissue and act as a bridge between the innate and

adaptive immune systems, neutrophils appear only in infected or damaged tissue after being recruited by inflammatory mediators released from activated macrophages and endothelial cells or by chemical signals released by invading microorganisms themselves (Table 169-1). After accumulation of these key immune cells at sites of infection, the microbes are eliminated through the process of phagocytosis, which is defined as the engulfment, internalization, and degradation of extracellular material.

NEUTROPHILS

Neutrophils develop in the bone marrow from myeloid precursors, migrate into the circulation, and, if required, make their way into infected or damaged tissue (Fig. 169-1). Their travels are essentially one-way trips because once they leave a compartment, they do not return. After release from the bone marrow compartment, a mature neutrophil has a blood half-life of 10 hours and may survive up to an additional 48 hours within infected or damaged tissue.

TABLE 169-1 PRIMARY IMMUNE ROLES OF MONOCYTE-MACROPHAGES AND NEUTROPHILS

MONONUCLEAR PHAGOCYTE FUNCTIONS
Elimination of invading pathogens
Elimination of cellular debris from sites of tissue damage and the blood stream
Wound healing and remodeling of normal tissue
Amplification of the innate immune response: release of immune regulators
Bridge to the adaptive immune system: presentation of antigens to lymphocytes
NEUTROPHIL FUNCTIONS
Elimination of invading pathogens

The Bone Marrow Compartment: The Site of Granulopoiesis

Neutrophils are the most abundant white blood cell and account for up to 70% of circulating leukocytes. Neutrophil numbers can increase rapidly by as much as 5- to 10-fold during periods of acute infection with a rate of continuous supply by the bone marrow of $5 \times 10^{10} - 10 \times 10^{10}$ neutrophils/day. Because these cells have a very short half-life in blood, the bone marrow compartment provides a steady supply of mature neutrophils with the capability to upregulate cell production rapidly during times of infection. Neutrophils originate in the bone marrow from a common population of hematopoietic stem cells through a 10- to 14-day process of proliferation, differentiation, and maturation (Chapter 156).

Steps in Granulopoiesis

The stages of neutrophil granulopoiesis in the bone marrow (Fig. 169-2) are identified by the major transitions from the pluripotent stem cell to the mature neutrophil. The *myeloblast* is the first recognizable progenitor cell committed to granulopoiesis. This proliferating cell is characterized by its large nucleus and agranular cytoplasm. The *promyelocyte* follows and displays the initial development of primary granules. *Myelocytes* occupy the next stage of neutrophil maturation and are characterized by development of the first specific or "secondary" (peroxidase-negative) granules. *Metamyelocytes,* which follow myelocytes, are incapable of further mitosis and are readily identifiable by their now numerous cytoplasmic granules. The functional maturation of metamyelocytes results in the development of *band* cells, which are usually slightly larger than mature neutrophils and have a horseshoe-shaped nucleus and a moderate to abundant supply of specific granules. Band cells can be found in the circulation during periods of acute infection. The final mature *neutrophil,* which is released into the circulation, has a diameter of approximately 10 μm with a characteristic nucleus that is segmented and multilobed and occupies about 20% of the cell's volume; the remaining

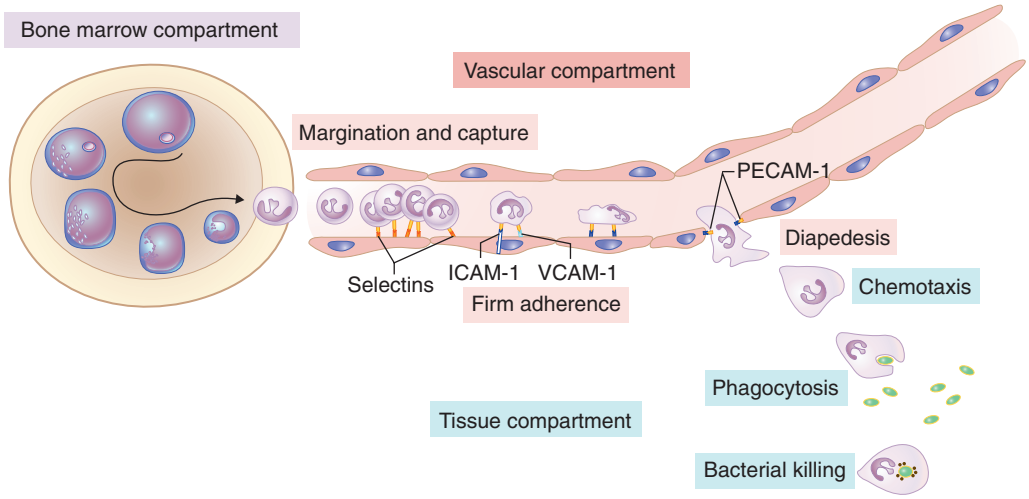

FIGURE 169-1. Life cycle of the neutrophil. The three major neutrophilic compartments (bone marrow, vascular, and tissue compartments) and the various steps involved in recruiting neutrophils to sites of infection are shown. ICAM = intercellular adhesion molecule; PECAM = platelet endothelial cell adhesion molecule; VCAM = vascular cell adhesion molecule.

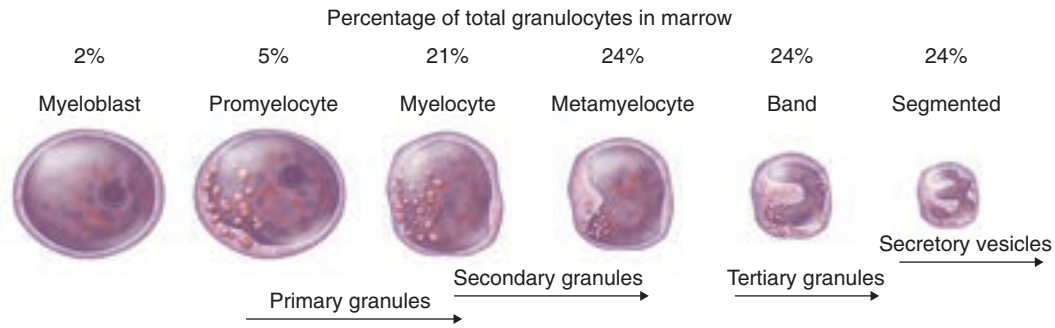

FIGURE 169-2. Cellular stages of granulopoiesis in bone marrow.

cytoplasm is taken up by granules. The distinctiveness of the granules reflects differences in content; as a result, granules formed at different stages carry specific types of matrix and membrane proteins.

Defects in granulopoiesis are manifested clinically as low circulating levels of neutrophils (neutropenia; Chapter 167). Verification of the stage at which neutrophil developmental arrest occurs can be determined by a bone marrow biopsy to assess the cellularity and characteristics of the neutrophil precursors present in the marrow space.

Regulation of Granulopoiesis

Granulopoiesis is driven by hematopoietic growth factors (Chapter 156). These factors, which are synthesized by a variety of cells, including fibroblasts and endothelial cells, are known to work together with other regulatory molecules, such as cytokines, to regulate hematopoiesis. Hematopoietic growth factors such as *interleukin-3* (IL-3), *granulocyte-macrophage colony-stimulating factor* (GM-CSF), and *granulocyte colony-stimulating factor* (G-CSF) bind to their target cells through specific receptors and are critical for the hematopoietic system to respond rapidly to infection or inflammation by dramatically increasing the production of leukocytes.

G-CSF is a potent cytokine that influences the proliferation, survival, maturation, and functional activation of cells from the neutrophil-granulocyte lineage. In normal individuals, circulating levels of G-CSF are very low (<100 pg/mL). However, in conditions of stress, G-CSF levels can rise to 20 times baseline levels, thereby resulting in a rapid increase in circulating neutrophils. G-CSF may regulate this increased granulopoiesis by increasing the mitotic pool at the promyelocyte and myelocyte stages and shortening neutrophil transit time in bone marrow.

Neutrophil Granules

One of the major mechanisms used by neutrophils to eliminate bacteria is a remarkable arsenal of antimicrobial proteins that are packed into cytoplasmic granules (Table 169-2). These antimicrobial proteins are securely contained within their respective granules and are released only when granules fuse with phagosomes or directly with the plasma membrane. Granulogenesis begins between the myeloblast and promyelocyte stages of neutrophil development and continues throughout the differentiation and maturation process of the cell. *Azurophilic* granules, which make up 30% of granules in a mature neutrophil, are the first to appear at the promyelocyte stage; they contain hydrolytic enzymes, microbicidal peptides, and myeloperoxidase. During phagocytosis, azurophil granule degranulation is restricted to internalization of phagocytic vacuoles. The *secondary,* or specific, granules appear later, beginning at the metamyelocyte stage; they are twice as abundant in the cytoplasm as azurophilic granules and contain proteins such as collagenase and lactoferrin. The *gelatinase-containing,* or *tertiary,* granules also appear at the metamyelocyte stage. A fourth group of granules, *secretory vesicles,* appears at the very final stages of neutrophil maturation, immediately before release of the cell into the circulation. All the granule types contain membrane proteins such as CR1, CR3, CD45, CD11c, and fMLP (*N*-formyl-methionyl-leucyl-phenylalanine) receptors, which are rapidly transported to the plasma membrane during activation to enhance neutrophil microbicidal activity.

A number of clinical conditions result from specific defects in granule development and formation. Initial assessment for granule defects can be made by microscopic evaluation of a peripheral blood smear (Chapter 157). Examples of obvious clinical diagnoses made with the peripheral smear include specific granule deficiency, which is characterized by bilobed nuclei in more than 80% of the neutrophils, and a significant decrease in cytoplasmic granularity. Abnormally large cytoplasmic granules are seen in individuals with Chédiak-Higashi syndrome.

The Vascular Compartment

Mature neutrophils are released from the postmitotic bone marrow compartment into the circulation, where they have an approximate lifespan of 8 to 12 hours and either circulate within the center of the blood vessel or attach to its endothelial lining, a process termed *margination*. Marginated neutrophils on the vessel walls are able to detach and reenter the circulation when required. For example, corticosteroids and epinephrine induce a rapid increase in circulating neutrophils by releasing neutrophils from the marginated pool. Neutrophils circulate until they are recruited to a site of infection. The initial phase of recruitment involves changes in the endothelial cell surface receptors lining the capillary beds closest to the site of infection or tissue damage. These critical changes in endothelial cells are mediated by immune regulators released by tissue macrophages, which initially detect the tissue damage or bacterial invasion. Emigration of circulating neutrophils from the vasculature to the site of infection or tissue damage requires three steps (see Fig. 169-1): capture and margination, firm adhesion to the endothelial wall, and diapedesis.

Margination and Capture

The marginated pool of neutrophils consists of neutrophils transiently retained against the walls of pulmonary capillaries and postcapillary venules. In the 20-μm diameter of a postcapillary venule, the smaller and faster circulating red blood cells displace the slower moving and larger neutrophils, which move to the vessel margins, where a low-affinity molecular interaction occurs between surface adhesion molecules of the neutrophil and the endothelial cells. This interaction results in neutrophil rolling and capture along the vessel walls, an event that requires the specific neutrophil receptors *leukocyte selectin* (L-selectin) and the corresponding endothelial ligand, sialyl Lewis (sLe). L-selectin is constitutively expressed in neutrophils, with highest expression in young circulating neutrophils and gradual decline with a cell's age, probably because previous margination events have depleted the receptor. The endothelial ligand for L-selectin, sLe, is a sialylated carbohydrate linked to a mucin-like molecule that can be upregulated by bacterial lipopolysaccharide or other mediators of inflammation. The selectin-ligand interactions are reversible and serve to promote and maintain accumulation of circulating neutrophils on inflamed endothelium.

TABLE 169-2 MEMBRANE AND MATRIX COMPONENTS OF NEUTROPHILIC GRANULES

COMPONENT	AZUROPHIL GRANULES (PRIMARY; PEROXIDASE POSITIVE)	SPECIFIC GRANULES (SECONDARY; PEROXIDASE NEGATIVE)	GELATINASE GRANULES (TERTIARY; PEROXIDASE NEGATIVE)	SECRETORY VESICLES
Antimicrobial proteins	Defensins Lysozyme Elastase Myeloperoxidase Cathepsin G	Lysozyme Lactoferrin	Lysozyme	
Membrane proteins and receptors	CD63 CD68 Alkaline phosphatase	CD11b fMLP-R Cytochrome b_{558} CR3	CD11b fMLP-R Cytochrome b_{558} CR3 CD45	CD11b fMLP-R Cytochrome b_{558} CR1 CD14 CD16
Matrix proteins	β-Glucuronidase	Collagenase Gelatinase Laminin	Gelatinase	Albumin

Modified from Edwards SW. *Biochemistry and Physiology of the Neutrophil.* Cambridge, UK: Cambridge University Press; 2005:55.

Adherence to the Endothelial Wall

Low-affinity, selectin-mediated transient interactions must be replaced by high-affinity, adhesive contacts between neutrophils and endothelial cells. During an acute inflammatory event, mediators derived from bacteria, damaged host cells, complement activation, or other immune cells are released from the site of infection and diffuse to the capillary beds, where they induce an immediate and transient vascular response that results in vascular leakage, which further encourages neutrophil margination. Endothelial cells adjacent to the site of inflammation, as well as the activated neutrophils that are bound to them, express integrin receptors that lead to high-affinity attachments between the neutrophils and endothelial cells. These high-affinity connections occur between neutrophil *β2 integrins* and their endothelial counterparts, the intercellular adhesion molecules (ICAMs). Integrins, which are a receptor family of heterodimeric transmembrane glycoproteins made up of an α- and β-subunit, are integral for cell adhesion. Neutrophil β2 integrins consist of three different α-subunits (CD11a, CD11b, and CD11c) that bind to a common β-subunit (CD18). The cytoplasmic tails of these transmembrane receptors possess phosphorylation sites for attachment of signal transduction and cytoskeletal proteins. The neutrophil integrins that mediate this adhesion step are *macrophage antigen-1* (Mac-1; CD11b/CD18) and *lymphocyte-associated function antigen-1* (LFA-1; CD11a/CD18). The receptors are stored in the neutrophil granule compartments to facilitate quick transfer to the plasma membrane during cell stimulation. The integrins bind to endothelial ICAM-1 and ICAM-2 and *vascular cell adhesion molecule-1* (VCAM-1), which are upregulated on the endothelial cell membranes when a cell is exposed to inflammatory cytokines. L-selectin receptors on neutrophils are concentrated on microvillus projections of the cell membrane, whereas the integrins are restricted to the body of the neutrophil. As a result, soon after initial contact during rolling interactions, the projections retract, thereby allowing integrins to interact with their ligands.

Diapedesis

Firm adherence through the L-selectin and integrin receptors facilitates transendothelial migration, or diapedesis, which marks the "point of no return" in the process of neutrophil recruitment to the site of injury. Unlike rolling and firm adhesion, which require heterophilic interactions between one class of molecules on the neutrophil and another class of molecules on the endothelial cell, diapedesis involves homophilic interactions between the same class of molecules on both cells—the *platelet-endothelial cell adhesion molecule-1* (PECAM-1 or CD31). PECAM-1 is expressed evenly on the surface of neutrophils and is concentrated at endothelial cell junctions. When they are firmly bound to the endothelial cell surface, neutrophils migrate between the closest tricellular endothelial cell junctions through interactions with PECAM-1 receptors. The neutrophil has now entered the tissue compartment, where it is primed for its final critical role in the elimination of microorganisms and cellular debris.

Laboratory Evaluation of Margination and Firm Adhesion

A defect in neutrophil margination or adhesion to the endothelial lining of the vascular compartment results in neutrophilia (elevated circulating neutrophil levels). This condition is usually associated with leukocyte adhesion deficiency (LAD), which is the result of a lack of CD11/CD18 receptor surface expression in peripheral blood neutrophils. If LAD is suspected, surface expression of these receptors can be measured with a flow cytometer and specific antibodies to CD11, CD18, or CD15 receptors.

The Tissue Compartment

Chemotaxis

Chemotaxis is the directed movement of cells up a chemical concentration gradient of a *chemoattractant.* Chemoattractants are soluble proteins or peptides, including bacterial products, complement factors, and chemokines produced by both inflammatory and noninflammatory cells, that are released from damaged or infected tissue. A concentration difference of 1% at the opposite ends of the neutrophil is sufficient to activate neutrophil chemotaxis. After a chemoattractant binds to its corresponding neutrophil membrane receptor, a series of cytoplasmic signaling pathways leads to activation of the neutrophil cytoskeleton. This activation results in the neutrophil assuming a polarized state characterized by an actin-rich leading lamella or pseudopod that drives cell motility.

Directional cell crawling, the intrinsic basis of chemotaxis, can be broken down into smaller processes, including extension of the cell membrane, adhesion to the tissue matrix, and contraction of the cell body in an organized and reversible manner. The actin-dependent protrusion of the leading edge, which is a sheetlike structure rich in actin filaments, is critical for normal neutrophil motility. The actin filaments within these lamellar regions are assembled into highly organized structures that push the membrane forward. These structures are formed by different collections of actin-binding proteins under the regulation of specific signal transduction cascades linking chemotactic receptors with cell movement. Defects in actin assembly also result in defects in chemotaxis and recurrent infections.

Actin Assembly Biology

Actin filaments are polar structures, with each end differing in its equilibrium-binding constant for actin monomers (Fig. 169-3). Filaments grow at the high-affinity or barbed end, whereas depolymerization occurs at the low-affinity or pointed end. This difference, generated by the ability of actin to bind and hydrolyze adenosine triphosphate, provides a physical polarity that regulatory proteins use to drive filament dynamics with high temporal and spatial precision. Three classes of proteins regulate the availability of high-affinity actin filament ends: filament-nucleating proteins (e.g., ARP2/3 de novo nucleation), filament-capping proteins (e.g., gelsolin), and filament-severing proteins (e.g., cofilin). Actin-nucleating factors bind actin monomers under conditions otherwise unfavorable for assembly and generate a new filament with a free high-affinity end available for assembly. Actin filament-capping proteins bind to the high-affinity filament end and regulate the addition of monomers by their presence or absence at the end of the filament. Actin-binding proteins are regulated by various second messengers,

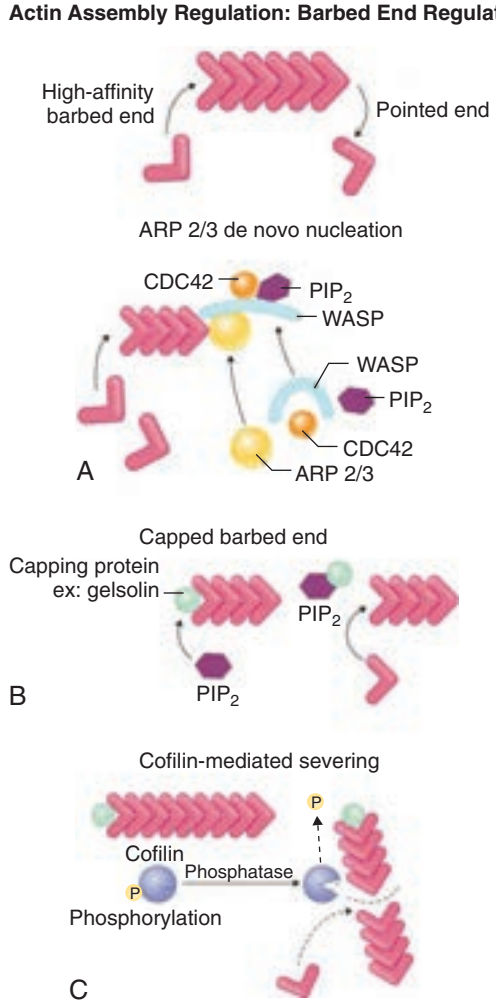

FIGURE 169-3. Regulation of actin assembly through the generation of free barbed ends by actin-binding proteins. A, The components below join together to form a nucleation complex (above). **B,** PIP_2 binds the capping protein, leading to its removal from the high-affinity end, allowing for addition and filament growth. **C,** A phosphatase removes the P from cofilin, thereby allowing it to sever the actin filament and leaving a free high-affinity end. PIP_2 = phosphatidylinositol 4,5-biphosphate; WASP = Wiskott-Aldrich syndrome protein.

including calcium. On stimulation, localized changes in the intracellular Ca^{2+} concentration lead to the rapid initiation of actin assembly and disassembly. The changes in actin filament length and the extent of cross-linkage between the filaments may account for the directional extension of actin-rich lamellae and contraction of the tail-like uropod at the other end of the cell. Movement in the neutrophil is therefore the result of lamellar protrusions resulting from the growth of actin filaments. Actin-rich lamellae will continue to be maintained as long as the neutrophil detects the chemoattractant gradient.

Laboratory Evaluation of Chemotaxis

A defect in neutrophil chemotaxis can be measured in the laboratory with a Boyden chamber, which uses a porous membrane to separate isolated neutrophils from a chemoattractant. A chemical gradient develops across the porous membrane and activates the neutrophils to crawl through the membrane toward the compartment containing the chemoattractant. Defects in chemotaxis can be determined by a lack of neutrophil transmigration through the membrane compared with control neutrophils from a healthy donor.

Phagocytosis

Phagocytosis is the process whereby neutrophils engulf and internalize invading pathogens into membrane compartments called *phagosomes*. Bacterial targets are "highlighted" or opsonized by antibodies (immunoglobulin G) or products from the classical complement pathway that coat the target and serve to mediate phagocytic adhesion. Neutrophilic phagocytosis involves two separate classes of receptors: *Fcγ receptors* (CD32 and CD16) for antibody-coated targets and *complement receptors* (CR1 and CR3) for complement-coated targets. CD32 and CR3 are functional receptors directly involved in neutrophilic phagocytosis, whereas CD16 and CR1 are coreceptors that assist their mate in completing binding and internalization. Activation of Fcγ receptors brings about phosphorylation of their cytoplasmic *immunoreceptor tyrosine-based activation motifs* (ITAMs) through activation of *Src family kinases;* the result is transduction of signals that induce extension of pseudopods, including signaling to the small Rho family of small guanosine triphosphatases (GTPases). These GTPases are responsible for the assembly of actin filaments, thereby leading to remodeling of the plasma membrane and the formation of actin-rich pseudopods, which are essential for the ingestion of particles and formation of phagosomes.

Laboratory Evaluation of Phagocytosis

Neutrophils can be incubated with fluorescently labeled bacteria after opsonization with serum from either the patient or a control. Phagocytosis is assessed by flow cytometry, which measures the increase in neutrophilic fluorescence after uptake of the fluorescently tagged bacteria.

Bacterial Killing

Phagocytes use two potent mechanisms for killing bacteria within the membrane-bound phagosome. The first involves fusion of the previously described storage granules with the phagosome to deliver microbicidal and lytic enzymes into the membrane compartment that contains the ingested microorganisms. The second mechanism uses a multiprotein enzyme complex to generate microbicidal oxidants through partial reduction of oxygen. The multiprotein enzyme complex known as reduced nicotinamide adenine dinucleotide phosphate (NADPH) oxidase generates oxidants by means of oxygen consumption, hence the term *respiratory burst*.

The NADPH enzyme system is made up of four essential polypeptide subunits that are denoted by their molecular weight (kD) and the superscript phox, which denotes phagocyte oxidase. Within the cytoplasmic membrane, the subunits $p22^{phox}$ and $gp91^{phox}$ bind the electron-carrying components of the oxidase (NADPH, a flavin adenine dinucleotide, and two nonidentical hemes) and form the cytochrome b_{558} redox center of the oxidase complex. Cellular activation by inflammatory mediators results in the addition of two cytosolic components, $p47^{phox}$ and $p67^{phox}$, to the complex along with the Rac small guanosine triphosphatase (GTPase).

The membrane-bound electron transport chain NADPH oxidase catalyzes the reduction of molecular oxygen to superoxide (O_2^-). The superoxide generated by this process is in turn catalytically converted to hydrogen peroxide and serves as a cosubstrate for myeloperoxidase to oxidize halides and to produce hypochlorous acid (HOCl), a very potent antimicrobial agent. These oxidants are able to kill bacteria within the phagosomes by oxidizing their cellular constituents.

Laboratory Evaluation of the Respiratory Burst and Bacterial Killing

Flow cytometry, a rapid and effective method for quantitatively assessing the respiratory burst, measures the fluorescence generated by cytoplasmic fluorescent probes such as dihydrorhodamine, which is converted to rhodamine by H_2O_2. The nitroblue tetrazolium (NBT) test is still used for rapid assessment of the respiratory burst when flow cytometry is not available.

Bacterial killing assays using a patient's neutrophils with either the patient's or control serum and bacteria such as *Staphylococcus aureus* or *Escherichia coli* are a definitive method to determine whether a given patient's neutrophils have an intracellular killing defect. Neutrophils from a healthy control subject phagocytose and kill approximately 95% of the bacteria within 2 hours. In assays in which an intracellular killing defect is present, neutrophils kill less than 10% of bacteria over a 2-hour period. It is necessary to confirm that there is no phagocytic defect before performing the bacterial killing assay to be sure that any defect in bacterial killing is not due to an internalization defect.

Neutrophil Extracellular Traps

Neutrophils use an extracellular process to contain and kill bacteria. Neutrophil extracellular traps (NETs) are formed by the release of chromatin and antimicrobial proteins from the neutrophil cytoplasm and granules. The chromatin forms a netlike meshwork that traps the bacteria and brings them in closer proximity to the antimicrobial elements adhered to the chromatin. Activation of NET formation requires simultaneous activation by at least two different receptors, and reactive oxygen species are essential to the process. IL-8 has been shown to be a potent activator of NET formation. The importance of NETs was highlighted by work showing that DNase expressing strains of group A streptococcus (GAS) and *Streptococcus pneumoniae* are more virulent than their non-DNase-expressing counterparts because of their ability to escape NETs.

CLINICAL MANIFESTATIONS

In addition to fever and recurrent infections, the most common findings in patients with phagocytic defects are oral infections resulting in gingival inflammation, periodontal bone loss, mobile or loose teeth, and premature loss of teeth (Table 169-3). An oral examination should be performed at the initial evaluation, followed by a full dental examination, depending on the findings. The history and laboratory tests can differentiate among the various clinical causes of disordered phagocytosis (Table 169-4).

DEFECTS IN LEUKOCYTE ADHESION

A defect in neutrophil adhesion to the endothelial lining leads to neutrophilia—an accumulation of neutrophils in the circulation, with very few neutrophils at sites of infection. Defects in neutrophil adhesion can be induced by drugs or due to a genetic defect. Drugs such as corticosteroids and epinephrine result in a transient leukocyte adhesive defect that results in an apparent dramatic increase in circulating neutrophils because of release of the marginated neutrophil pool. The major genetic disease that results in an adhesion deficiency is termed *leukocyte adhesion deficiency.*

Leukocyte Adhesion Deficiency

LAD-1

PATHOBIOLOGY

LAD-1 is an autosomal recessive inherited disorder in which patients have a mutation in the gene encoding CD18. The result is a deficiency of β2 integrin

TABLE 169-3 SYMPTOMS SUGGESTIVE OF A PHAGOCYTIC DISORDER

Recurrent infections that fail to resolve with conventional treatment
Recurrent infections of unusual severity
Recurrent infections in the lung, liver, or bone
Normally nonpathogenic bacteria or fungi identified in cultures from the infection sites
Aphthous ulcers
Severe periodontal diseases, including gingivitis
Lymphadenopathy or hepatosplenomegaly
Severe recurrent cutaneous infections with *Staphylococcus aureus*
Recurrent mycobacterial infections

receptors, which are required for neutrophil migration from the vasculature into the tissues, thereby impairing the binding of neutrophils to C3bi and endothelial ICAM-1 and ICAM-2. Indications of the disorder are high resting neutrophil counts accompanied by frequent dissemination, sepsis, and recurrent infections.[2]

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Clinical manifestations in patients diagnosed with LAD-1 include delayed separation of the umbilical cord, bacterial and fungal infections, delayed wound healing, impaired pus formation, and severe destructive periodontitis with rapid tooth loss. Patients usually die during childhood. Flow cytometry is used to measure CD11/CD18 surface expression levels on neutrophils.

TREATMENT

Rx

Treatment is mainly supportive of early intervention for periodontal disease with prophylactic antibiotics in patients with recurrent infections. In severe cases, bone marrow transplantation is the treatment of choice.[3]

LAD-2 AND LAD-3

LAD-2, a variant of LAD-1, is associated with neutrophilia, the Bombay (hh) blood phenotype, dwarfism, and mental retardation. This disorder is due to a mutation in the guanosine diphosphate-fucose transporter gene, which

TABLE 169-4 DISORDERS OF PHAGOCYTIC FUNCTION

DISORDER	ETIOLOGY	IMPAIRED FUNCTION	CLINICAL CONSEQUENCE
DEGRANULATION ABNORMALITIES			
Chédiak-Higashi syndrome	Autosomal recessive; disordered coalescence of lysosomal granules Responsible gene found at 1q42-45. The encoded protein (LYST) has structural features homologous to a vacuolar sorting protein	Decreased neutrophilic chemotaxis, degranulation, and bactericidal activity; platelet storage pool defect; impaired NK function; failure to disperse melanosomes	Neutropenia, recurrent pyogenic infections, propensity for the development of marked hepatosplenomegaly in the accelerated phase, partial albinism
Specific granule deficiency	Autosomal recessive; abnormal regulation of various myeloid granule genes by a transacting factor	Impaired chemotaxis and bactericidal activity; bilobed nuclei in neutrophils; reduced content of neutrophil defensins, gelatinase, collagenase, vitamin B_{12}–binding protein, and lactoferrin	Recurrent infections, especially sinopulmonary and skin infections
ADHESION ABNORMALITIES			
Leukocyte adhesion deficiency type 1	Autosomal recessive; absence of CD11/CD18 surface adhesive glycoprotein (β2-integrins) on leukocyte membranes, most commonly arising from failure to express CD18 mRNA	Decreased binding of C3bi to neutrophils and impaired adhesion to ICAM-1 and ICAM-2	Neutrophilia, recurrent bacterial infection associated with lack of pus formation
Leukocyte adhesion deficiency type 2	Autosomal recessive; absence of neutrophil sialyl-$Lewis^x$	Decreased adhesion to activated endothelium expressing ELAM	Neutrophilia, recurrent bacterial infection without pus
Leukocyte adhesion deficiency type 3	Autosomal recessive; defects in activation of β1, β2, and β3 integrins	Severe leukocyte adhesion dysfunction; abnormal platelet aggregation	Neutrophilia, recurrent bacterial infection without pus, severe bleeding tendency
Neutrophil actin dysfunction	Altered polymerization of neutrophil cytoplasmic actin, perhaps arising from the presence of an inhibitor to F-actin formation	Impaired neutrophil adhesion, chemotaxis, and bacterial killing	Neutrophilia, recurrent bacterial infections without pus
DISORDERS OF CELL CHEMOTAXIS			
Hyperactive Chemotaxis			
Familial Mediterranean fever (FMF)	Autosomal recessive gene responsible for FMF on chromosome 16, which encodes for a protein called pyrin; pyrin may modify neutrophil activation	Excessive accumulation of neutrophils at inflamed sites	Recurrent fever, peritonitis, pleuritis, arthritis, amyloidosis
Depressed Chemotaxis			
Intrinsic defects of the neutrophil, e.g., leukocyte adhesion deficiency, Chédiak-Higashi syndrome, specific granule deficiency, neutrophil actin dysfunction, neonatal neutrophils	In the neonatal neutrophil, there is diminished ability to express β2 integrins and a qualitative impairment in β2 integrin function	Diminished chemotaxis	Propensity for the development of pyogenic infections
Direct inhibition of neutrophil mobility, e.g., drugs	Ethanol, glucocorticoids, cyclic AMP	Impaired locomotion and ingestion, impaired adherence	Possible causes of frequent infections; neutrophilia seen with epinephrine is the result of cyclic AMP release from the endothelium
Immune complexes	Bind to Fc receptors on neutrophils in patients with rheumatoid arthritis, systemic lupus erythematosus, and other inflammatory states	Impaired chemotaxis	Recurrent pyogenic infections
Hyperimmunoglobulin E syndrome	Disorders of cytokine signaling, most commonly due to autosomal dominant mutations in the *STAT3* gene	Impaired chemotaxis, impaired IgG opsonization of *Staphylococcus aureus*	Recurrent skin and sinopulmonary infections
DEFECTS OF MICROBICIDAL ACTIVITY			
Chronic granulomatous disease (CGD)	X-linked and autosomal recessive; failure to express functional $gp91^{phox}$ (in the phagocyte membrane) and $p22^{phox}$ (autosomal recessive). Other autosomal recessive forms of CGD arise from failure to express protein $p47^{phox}$ or $p67^{phox}$	Failure to activate neutrophil respiratory burst leading to failure to kill catalase-positive microbes	Recurrent pyogenic infections with catalase-positive microorganisms
G6PD deficiency	Less than 5% of normal activity of G6PD	Failure to activate NADPH-dependent oxidase	Infections with catalase-positive microorganisms

TABLE 169-4 DISORDERS OF PHAGOCYTIC FUNCTION—cont'd

DISORDER	ETIOLOGY	IMPAIRED FUNCTION	CLINICAL CONSEQUENCE
Myeloperoxidase deficiency	Autosomal recessive; failure to process modified precursor protein arising from missense mutation	H_2O_2-dependent antimicrobial activity not potentiated by myeloperoxidase	None
Deficiencies of glutathione reductase and glutathione synthetase	Failure to detoxify H_2O_2	Excessive formation of H_2O_2	Minimal problems with recurrent pyogenic infections
IMPAIRED MACROPHAGE FUNCTION			
Defects in the interferon-γ–IL-12 axis	Interferon-γ receptor ligand-binding chain, interferon-γ receptor signaling chain, IL-12 receptor β_1 chain, IL-12 p40 deficiency; the interferon-γ receptor abnormalities may be autosomal dominant or recessive; the IL-12 receptor and IL-12 abnormalities are autosomal recessive	Impaired killing of microorganisms. Fatal BCG infection secondary either to an inability to produce IL-12 by dendritic cells and macrophages or to depressed bactericidal activity of macrophages lacking normal function of the interferon receptor	Infection with atypical mycobacteria, *Salmonella*, and *Listeria*
Hemophagocytic lymphohistiocystosis (HLH) and macrophage activation syndrome (MAS)	Primary inherited form with mutations in perforin gene and genes involved in exocytosis; secondary acquired forms associated with infections, malignancies, and (in MAS) rheumatologic disorders.	Hyperinflammatory state and hypercytokinemia; impairment of NK and cytotoxic T-cell function.	Fever, hepatosplenomegaly, pancytopenia, pulmonary and neurologic complications, increased serum ferritin and triglycerides

AMP = adenosine monophosphate; BCG = bacille Calmette-Guérin; ELAM = endothelial leukocyte adhesion molecule; G6PD = glucose-6-phosphate dehydrogenase; ICAM = intracellular adhesion molecule; IL-12 = interleukin-12; NADPH = nicotinamide adenine dinucleotide phosphate; NK = natural killer; phox = phagocyte oxidase.
Modified from Boxer LA. Quantitative abnormalities of granulocytes. In: Beutler E, Lichtman MA, Coller BS, et al, eds. *Williams Hematology*, 6th ed. New York: McGraw-Hill; 2001:836.

results in impaired expression of CD15s and other selectin ligands. Symptoms are similar to those of LAD-1, and the diagnosis is confirmed by flow cytometry for CD15s.

In the most recently described LAD-3,[4] there is a primary activation defect in all three β integrins (β1, β2, and β3), and mutations have been found in kindlin-3, which binds the cytoplasmic tail of integrin. Clinical manifestations include defects in platelet activation and severe bleeding tendency. Treatment includes blood transfusion during a bleeding episode.

DEFECTS IN NEUTROPHILIC CHEMOTAXIS

After phagocytes enter the tissue compartment from the vascular pool, they migrate up the concentration gradients of various chemoattractants to the site of focal infection. A number of chemotactic defects result in severe recurrent infections.

Hyperimmunoglobin E Syndrome

PATHOBIOLOGY

Hyperimmunoglobin E syndrome, or hyper-IgE syndrome, also referred to as Job syndrome, is a group of genetically diverse, multisystem disorders of cytokine signaling. The most common form of the syndrome involves dominant mutations in the gene for signal transducer and activator of transcription-3 (*STAT3*). Hyper-IgE syndrome is prevalent among white, Asian, and African populations, with equal frequency among males and females.[5]

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The neutrophilic disorder is characterized by recurrent skin abscesses, pneumonia, and periodontal diseases. After birth, patients usually have moderate to severe dermatitis, eczematous skin eruptions, nonerythematous abscesses, pneumatoceles, and severe osteoporosis that can result in bone fractures. The organisms most commonly present at infected sites are *Staphylococcus aureus, Haemophilus influenzae, Escherichia coli,* and *Candida albicans.* Patients have elevated IgE levels (typically >1000 IU/mL) and eosinophilia. The defect in neutrophilic chemotaxis is less severe than that in Chédiak-Higashi syndrome (see later). Genetic testing of the *STAT3* gene, in conjunction with characteristic manifestations, such as immunologic and infectious complications and involvement of skeletal and connective tissue, is imperative to confirm diagnosis.[6] With implementation of prophylactic measures, accompanied by IgG infusion, good long-term prognosis can be seen.

TREATMENT

Rx

Treatment includes prophylactic antibiotics, antifungal prophylaxis, IgG infusions, and aggressive treatment of infections.[7] In severe cases, hematopoietic stem cell transplantation may be considered.

Familial Mediterranean Fever

PATHOBIOLOGY

Familial Mediterranean fever, further discussed in Chapter 261, also known as recurrent polyserositis, is an autosomal recessive autoinflammatory disease that is widespread among people of Mediterranean descent, including Arabs, Armenians, and Sephardic Jews.[7] The genetic defect is a missense mutation in the *MEFV* gene, which encodes the protein pyrin. Pyrin is believed to be a transcription factor involved in downregulating inflammation, possibly through an effect on chemotaxis in neutrophils and monocytes. The *MEFV* mutation results in a hyperinflammatory response characterized by abundant neutrophilic infiltration into the peritoneal, pleural, and joint spaces.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The most common findings include acute, self-limited attacks of fever accompanied by pleuritis, peritonitis, arthritis, pericarditis, and erythematous skin lesions. Although first attacks may be observed during infancy, onset of clinical disease usually occurs in childhood or adolescence, with a relatively small number of adult-onset cases.

Leukocytosis has been observed during attacks, but the leukocyte count is normal between episodes. Genetic testing is available for the most common mutations. This disease can be fatal if renal failure develops as a result of amyloidosis (Chapter 188), which occurs in up to 25% of those affected.

TREATMENT

Rx

The hyperinflammatory attacks can be reduced significantly and even the complications of amyloidosis can be prevented[8] with prophylactic colchicine, 0.6 mg orally two or three times daily, up to a maximum dose of 2.0 to 2.4 mg if needed and if tolerated. The prognosis is generally good for most affected individuals maintained with colchicine. Rilonacept, an IL-1 decoy receptor, has been found to reduce frequency of attacks and is a treatment option for patients with colchicine-resistant or -intolerant disease.

DISORDERS OF NEUTROPHILIC DEGRANULATION

Granules supply key membrane proteins, including receptors required for phagocytosis. Granule-related defects result in profound abnormalities in bacterial killing.

Chédiak-Higashi Syndrome

PATHOBIOLOGY

Chédiak-Higashi syndrome is a rare autosomal recessive disorder of the *LYST* gene, which encodes a protein responsible for lysosomal trafficking. Defective targeting of granules to the membrane results in large cytoplasmic

granules that are unable to target to the plasma membrane in neutrophils, monocytes, and lymphocytes.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Symptoms are recurrent bacterial infections of the skin, mouth, and respiratory tract; partial albinism; peripheral neuropathy; and mild bleeding disorders as a result of a deficiency in serotonin- and adenosine phosphate–containing granules in platelets. Defects in myelopoiesis result in neutropenia. Death usually occurs by 7 years of age because of infection. Advanced disease is characterized by lymphocytic tissue infiltrates and pancytopenia.

Giant cytoplasmic granules are seen in the peripheral blood smear. Neutrophil function testing shows defects in chemotaxis and bacterial killing.

TREATMENT Rx

Prophylactic antibiotics should be used to prevent infections. Bone marrow transplantation from an HLA-matched donor may be successful if performed before the disease becomes advanced.

Specific Granule Deficiency

PATHOBIOLOGY

Specific granule deficiency (SGD) is an autosomal recessive disorder that manifests during infancy as the recurrent appearance of deep and superficial skin infections, respiratory infections, and abscesses. Azurophilic granules in neutrophils lack lactoferrin, defensins, gelatinase, collagenase, cytochrome *b*, and vitamin B_{12}-binding protein. Neutrophils are morphologically altered and have a bilobed rather than a trilobed nucleus.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

This disorder is characterized by impaired neutrophil chemotaxis, reduced respiratory burst, and a defect in bacterial killing. Infections are commonly caused by *S. aureus, Pseudomonas aeruginosa,* and *C. albicans.* Flow cytometry is used to determine deficiency in lactoferrin and vitamin B_{12}-binding protein for diagnosis.

TREATMENT Rx

Treatment includes administration of parenteral antibiotics and drainage for infections. With aggressive treatment of infections, survival into adulthood is possible.

DISORDERS OF OXYGEN-DEPENDENT BACTERIAL KILLING

A genetic defect in any component of the respiratory burst results in delayed or ineffective bacterial killing.

Chronic Granulomatous Disease

PATHOBIOLOGY

Chronic granulomatous disease (CGD) is a genetic disease that occurs in about 1 in 200,000 live births. Neutrophils and macrophages cannot generate superoxide and are therefore unable to kill catalase-positive organisms. This condition results from mutations in one of the four structural genes of the NADPH oxidase complex. The most common genetic defect occurs in the 91-kD component of cytochrome b_{558}, which is coded on the X chromosome. The other mutations are autosomal recessive and have been detected in the 22-, 47-, and 67-kD structural proteins.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Children are prone to infections or granulomatous lesions in the lungs, skin, and liver. *S. aureus* is the most common organism, but other organisms include *Serratia marcescens, Burkholderia cepacia, Aspergillus* species, and *Nocardia* species. Staphylococcal liver abscesses are pathognomonic of CGD. Flow cytometry is used to measure the increase in fluorescence generated when dihydrorhodamine is converted to rhodamine by H_2O_2.

TREATMENT Rx

Abscesses can be removed by surgery. Trimethoprim-sulfamethoxazole prophylaxis (5 mg/kg/day divided into two equal doses) and antifungal prophylaxis with itraconazole (100 mg/day for <50 kg, 200 mg/day for >50 kg) have been shown to reduce the frequency of infections in these patients. Interferon-γ (50 μg/m^2 subcutaneously three times per week) prophylaxis is now considered "standard of care" in many centers. Bone marrow transplantation can also be considered for patients with refractory infections.[9] Gene therapy for CGD by gene-modified autologous hematopoietic stem cell transplantation has resulted in transient immune restoration but also genomic instability, monosomy 7, and clonal progression toward myelodysplasia. Future trials will be required to determine the role of gene therapy in clinical care.

Myeloperoxidase Deficiency

PATHOBIOLOGY

Myeloperoxidase (MPO) deficiency is a relatively common disorder (1 in 4000) in which the enzyme for conversion of neutrophilic hydrogen peroxide to HOCl is absent. This deficiency is not associated with increased susceptibility to infections, probably because of the accumulation of hydrogen peroxide, which is also bactericidal. Several genes have been reported to cause MPO deficiency, including *R569W*, *Y173C*, and *M251T*. The most common gene mutation related to MPO deficiency is the *R569W* gene, associated with the absence of MPO in neutrophils. Less frequently, patients with a mutation in *Y173C* gene have MPO that cannot fully mature. Differently, patients with a mutation in the *M251T* gene have fully mature MPO that are not functional.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

MPO is usually asymptomatic, although patients with diabetes mellitus may occasionally experience candidal infections. The diagnosis is made by observation of a negative peroxidase stain of the peripheral blood smear.

TREATMENT Rx

Symptomatic patients may be treated with prophylactic antibiotics with routine control of blood glucose in those who have diabetes mellitus.

Glutathione Synthetase Deficiency

PATHOBIOLOGY

Glutathione synthetase (GSS) deficiency is a autosomal recessive disorder of glutathione metabolism with approximately 70 cases reported worldwide. Glutathione, which is a potent antioxidant found in granulocytes, is required for a normal respiratory burst and bacterial killing.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

GSS deficiency can exist in mild, moderate, and severe forms. Patients with GSS deficiency typically have recurrent otitis and hemolytic anemia. Mildly affected patients typically present with hemolytic anemia; in the moderate form, metabolic acidosis occurs; and patients with the severe form in addition develop central nervous system impairment, such as seizures, mental retardation, and ataxia. The diagnosis is confirmed by verifying low or no glutathione synthetase in red blood cells, high levels of 5-oxoproline in the urine (up to 1 g/kg/day), and mutations in the glutathione synthetase (*GSS*) gene. Onset of symptoms ranges from birth to infancy and childhood.

TREATMENT Rx

Treatment goals include supplementation with antioxidants, correction of acidosis with bicarbonate, and blood transfusion.

Severe Glucose-6-Phosphate Dehydrogenase Deficiency

PATHOBIOLOGY

Glucose-6-phosphate dehydrogenase (G6PD) deficiency is an X-linked disorder (>400,000 people) distributed throughout Africa, Asia, the Mediterranean, and the Middle East. The prevalence of G6PD deficiency is associated

with the distribution of malaria and provides partial protection against malarial infection. White individuals with a severe reduction in G6PD activity are subject to recurrent infections, whereas Asians or blacks with similarly reduced G6PD levels are not. G6PD is crucial for regulating the availability of NADPH for the respiratory burst. Gene mutations associated with G6PD are located on the distal arm of the X chromosome and are frequently identified as missense mutations.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

G6PD deficiency results in recurrent bacterial infections, hemolytic anemia (Chapter 161), and jaundice. The diagnosis can be made with flow cytometry to assess the respiratory burst and to demonstrate the absence of G6PD in all blood cells. The production of NADPH from NADP is detected thorough rapid fluorescent spot testing.

TREATMENT Rx

Treatment goals include abstaining from oxidative stressors and the consumption of fava beans. In severe cases of anemia, a blood transfusion is warranted.

MACROPHAGE-RELATED ABNORMALITIES

Accumulation of monocyte-macrophages at sites of infection occurs after the major influx of neutrophils. Macrophages have a critical role in antigen presentation to lymphocytes, thereby activating the adaptive arm of the immune system. A critical defect in macrophage signaling results in susceptibility to mycobacterial infection.

Interferon-γ Receptor-1 Defects

PATHOBIOLOGY

When macrophages phagocytose mycobacteria, they produce IL-12, which in turn stimulates T cells to produce interferon-γ (IFN-γ). IFN-γ is critical to the killing of mycobacteria and other intracellular bacteria. Patients with recurrent and severe mycobacterial infections who are not infected with human immunodeficiency virus should be assessed for abnormalities in pathways that lead to the generation and utilization of IFN-γ.

Patients with autosomal recessive mutations in the IFN-γ receptors typically have a complete loss of function of the IFN-γ receptors. Autosomal dominant mutations in the IFN-γ receptors result in normal ligand binding but defective intracellular signal transduction because of a cytoplasmically truncated form of the receptor.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Recessive mutations typically manifest as severe disseminated infections and poor formation of granulomas. Multifocal mycobacterial osteomyelitis is pathognomonic of an autosomal dominant mutation in the IFN-γ receptor. Flow cytometry confirms the absence of membrane expression of IFN-γ receptor-1 in the autosomal recessive form and up to 10-fold higher membrane expression levels of the cytoplasmically truncated receptor in the autosomal dominant form.

TREATMENT Rx

For patients with autosomal dominant mutations, subcutaneous IFN-γ is effective. For autosomal recessive patients completely lacking IFN-γ receptor function, hematopoietic stem cell transplantation should be considered. Long-term antibiotic prophylaxis against mycobacterial infections with azithromycin or clarithromycin is recommended.

Hemophagocytic Lymphohistiocytosis and Macrophage Activation Syndrome

PATHOBIOLOGY

Hemophagocytic lymphohistiocytosis (HLH), also known as *hemophagocytic syndrome*,[10] and the related *macrophage activation syndrome* are rare, often life-threatening syndromes of diverse etiologies. Their clinical manifestations reflect a state of extreme systemic inflammation and unregulated immune activation. These disorders have been traditionally classified as either (1) primary HLH that has an inherited basis and is seen in children,[11] or (2) secondary HLH that is seen in adults and is triggered by a variety of acquired disorders like infections, malignancies, and rheumatologic diseases. Macrophage activation syndrome has been considered to represent mainly the latter group, the rheumatologic forms of secondary HLH. However, the distinction between primary and secondary HLH is becoming increasingly blurred because genetic defects are also being discovered in adults with apparently acquired disease who have presentations of the syndrome that occur later in life. Genetic defects underlying HLH (and macrophage activation syndrome) generally cause impairment of natural killer (NK) cell and cytotoxic T-cell function. Mutations tend to occur in the perforin gene or in genes important for the exocytosis of cytotoxic granules.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The clinical manifestations of HLH result from its underlying hyperinflammatory state with hypercytokinemia. They include fever (often presenting as "fever of unknown etiology"), hepatosplenomegaly, various cutaneous manifestations, pulmonary involvement including acute respiratory failure with alveolar or interstitial infiltrates, various neurologic manifestations, bilineage or trilineage cytopenia, abnormal liver function tests, hypertriglyceridemia, and hypofibrinogenemia.

Conditions associated with secondary HLH include viral infections (herpes simplex virus, Epstein-Barr virus, cytomegalovirus, human immunodeficiency virus, parvovirus, influenza, and post vaccination); other infections (mycoplasma, bacterial, protozoal, fungal, and mycobacterial); malignancies (leukemia, Hodgkin and non-Hodgkin lymphoma, solid tumors like germ cell tumors); and immune deficiency states (including CGD and stem cell transplantation). Macrophage activation syndrome is now essentially considered to be HLH associated with rheumatic diseases, classically as a potentially lethal complication of systemic juvenile rheumatoid arthritis, but also with systemic lupus erythematosus, scleroderma, Sjögren syndrome, mixed connective tissue disorders, and Kawasaki disease.

A diagnostic hallmark of HLH, although sometimes not found early in the course of the disease, is histopathologic evidence of hemophagocytosis in the bone marrow.[12] Laboratory findings reflecting the hyperinflammatory state include extremely high serum levels of ferritin and soluble CD25 (i.e., the soluble IL-2 receptor, IL-2Rα).

TREATMENT Rx

A high index of suspicion is required to make an early diagnosis of HLH or macrophage activation syndrome so that treatment can be initiated promptly to attempt to prevent irreversible tissue damage. Most important is the identification and specific treatment of the underlying cause of the syndrome in any individual patient. Hematopoietic stem cell transplantation with reduced-intensity conditioning regimens is being increasingly considered when a genetic cause is identified. Combination therapy with (1) dexamethasone, etoposide with or without cyclosporine, or (2) corticosteroids, cyclosporine and antithymocyte globulin has been attempted, especially as a bridge to stem cell transplantation. Alemtuzumab, a monoclonal antibody to CD52-bearing lymphocytes, appears to be effective as a salvage agent for refractory HLH, leading to improvement and survival to transplantation in pediatric and adult patients.[13]

ASSESSING PHAGOCYTE FUNCTION: MAKING THE DIAGNOSIS

If a phagocyte functional disorder may be the underlying cause of recurrent infections in a patient, a complete blood count (CBC) and peripheral smear guide subsequent definitive testing (Fig. 169-4). Cultures from infected areas allow antimicrobial targeting and also provide critical diagnostic information. If the defect is a result of abnormal neutrophil development and maturation, the CBC will show neutropenia; a bone marrow biopsy might be required. Repeated CBC (twice per week for 6 weeks) is indicated if cyclic neutropenia is suspected because of a periodicity of the infections (Chapter 167).

If the CBC reveals neutrophilia, a defect in the recruitment of neutrophils into tissues is suggested. An assessment of the receptors required for transmigration by flow cytometry and specific antibodies to the surface receptors is indicated.

If circulating levels of phagocytes are normal yet the patient is experiencing recurrent infections, a phagocytic defect within the infected tissue is likely. Laboratory testing to evaluate chemotaxis, phagocytosis, and bacterial killing is indicated.

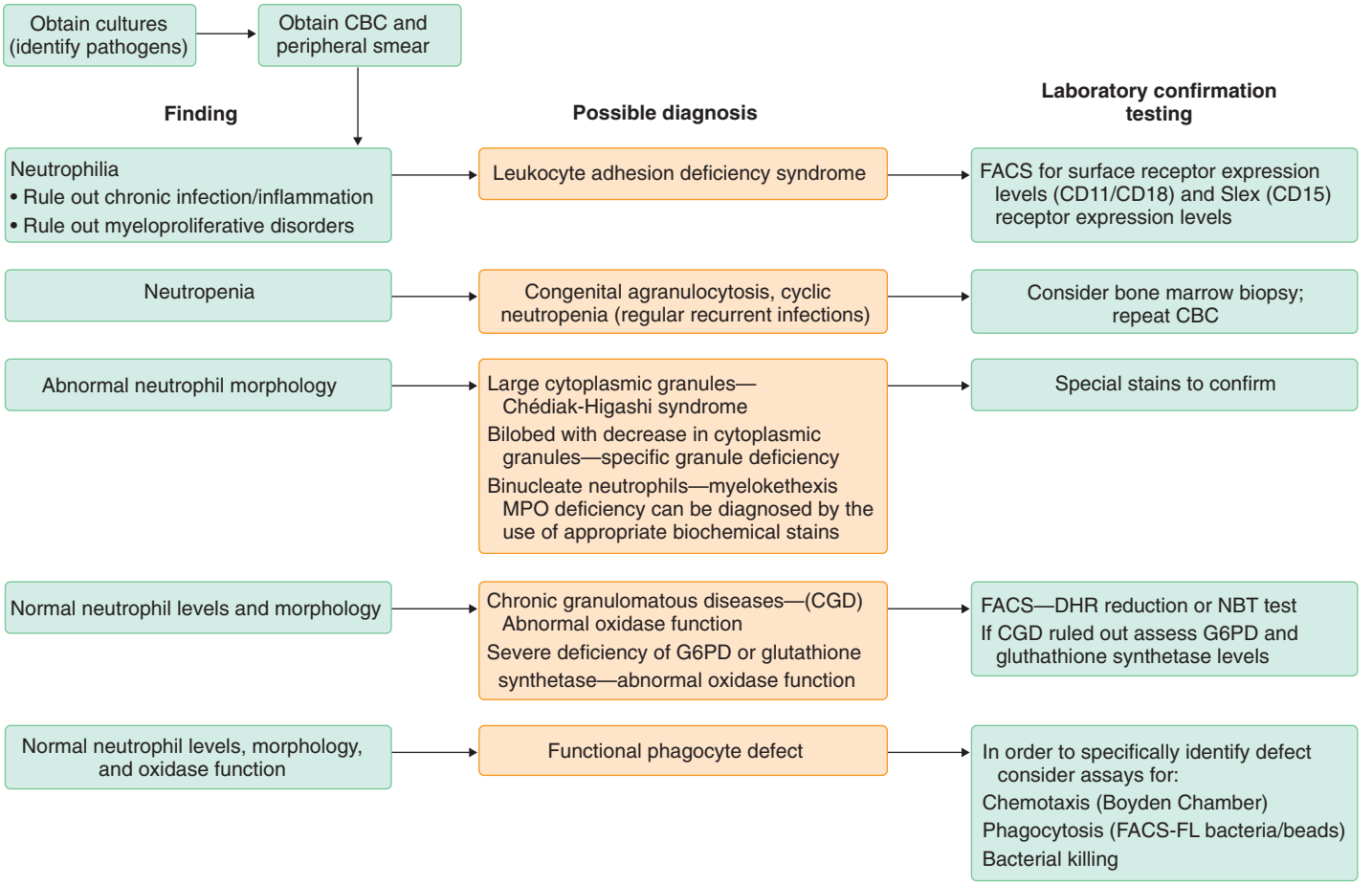

FIGURE 169-4. Approach to diagnosing a suspected phagocytic defect. CBC = complete blood count; DHR = dihydrorhodamine; FACS = flow cytometry; FL = fluorescent; G6PD = glucose-6-phosphate dehydrogenase; MPO = myeloperoxidase; NBT = nitroblue tetrazolium.

Grade A References

A1. Ozaltin F, Bilginer Y, Gulhan B, et al. Diagnostic validity of colchicine in patients with Familial Mediterranean fever. *Clin Rheumatol*. 2014;33:969-974.

A2. Hashkes PJ, Spalding SJ, Giannini EH, et al. Rilonacept for colchicine-resistant or -intolerant familial Mediterranean fever: a randomized trial. *Ann Intern Med*. 2012;157:533-541.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Nordenfelt P, Tapper H. Phagosome dynamics during phagocytosis by neutrophils. *J Leukoc Biol*. 2011;90:271-284.
2. van de Vijver E, van den Berg TK, Kuijpers TW. Leukocyte adhesion deficiencies. *Hematol Oncol Clin North Am*. 2013;27:101-116.
3. Elhasid R, Kilic SS, Ben-Arush M, et al. Hematopoietic stem cell transplantation in neutrophil disorders: severe congenital neutropenia, leukocyte adhesion deficiency and chronic granulomatous disease. *Clin Rev Allergy Immunol*. 2010;38:61-67.
4. Stepensky PY, Wolach B, Gavrieli R, et al. Leukocyte adhesion deficiency type III: clinical features and treatment with stem cell transplantation. *J Pediatr Hematol Oncol*. 2014; [Epub ahead of print].
5. Sowerwine KJ, Holland SM, Freeman AF. Hyper-IgE syndrome update. *Ann N Y Acad Sci*. 2012;1250:25-32.
6. Chandesris MO, Melki I, Natividad A, et al. Autosomal dominant STAT3 deficiency and hyper-IgE syndrome: molecular, cellular, and clinical features from a French national survey. *Medicine (Baltimore)*. 2012;91:e1-e19.
7. Caso F, Rigante D, Vitale A, et al. Monogenic autoinflammatory syndromes: state of the art on genetic, clinical, and therapeutic issues. *Int J Rheumatol*. 2013;2013:513782.
8. Ter Haar NM, Frenkel J. Treatment of hereditary autoinflammatory diseases. *Curr Opin Rheumatol*. 2014;26:252-258.
9. Güngör T, Teira P, Slatter M, et al. Reduced-intensity conditioning and HLA-matched haemopoietic stem-cell transplantation in patients with chronic granulomatous disease: a prospective multicentre study. *Lancet*. 2014;383:436-448.
10. Janka GE, Lehmberg K. Hemophagocytic syndromes—an update. *Blood Rev*. 2014;28:135-142.
11. Zhang M, Behrens EM, Atkinson TP, et al. Genetic defects in cytolysis in macrophage activation syndrome. *Curr Rheumatol Rep*. 2014;16:439.
12. Lehmberg K, Ehl S. Diagnostic evaluation of patients with suspected haemophagocytic lymphohistiocytosis. *Br J Haematol*. 2013;160:275-287.
13. Marsh RA, Allen CE, McClain KL, et al. Salvage therapy of refractory hemophagocytic lymphohistiocytosis with alemtuzumab. *Pediatr Blood Cancer*. 2013;60:101-109.

REVIEW QUESTIONS

1. What is the probable diagnosis of an 8-year-old boy with severe dermatitis, recurrent bacteria skin infections, and pneumonia for evaluation of immunodeficiency? Initial tests reveal normal complete blood count and platelets, 40,000 IU of immunoglobulin E (IgE), and normal IgA, IgM, and IgG levels.

A. Familial Mediterranean fever
B. Chédiak-Higashi syndrome
C. Hyper-IgE syndrome
D. Chronic granulomatous disease
E. Leukocyte adhesion deficiency

Answer: C This patient has recurrent bacterial infections and pneumonia, which suggest the possibility of an immunodeficiency. Significant levels of IgE present with normal complete blood count and platelets indicate hyper-IgE syndrome. Genetic testing of the *STAT3* gene can be used to confirm diagnosis. Good prognosis seen in patients with administration of antibiotics, antifungal prophylaxis, and IgG infusions.

2. A 3-year-old girl with abnormally light pigmentation of the hair and skin has a history of recurrent respiratory infection. Blood smear shows giant cytoplasmic granules and low neutrophil counts. What is the mostly likely diagnosis?

A. Chédiak -Higashi syndrome (CHS)
B. Specific granule deficiency
C. Chronic granulomatous disease
D. Interferon-γ receptor-1 defect
E. Myeloperoxidase deficiency

Answer: A The low neutrophil count is an indication of neutropenia, in addition to giant cytoplasmic granules in the blood smear, suggesting CHS. Significant indication of CHS includes partial albinism as seen in patients light pigmentation and defects in bacterial killing and chemotaxis, which can be confirmed through neutrophil function testing. Infections can be treated with prophylactic antibiotics.

3. A 6-year-old girl has a history of recurrent otitis and hemolytic anemia. Testing indicates significant excretion of 5-oxoproline in the urine and low amounts of glutathione synthetase in red blood cells. What is the mostly likely diagnosis?

A. Glutathione synthetase deficiency
B. Macrophage activation syndrome
C. Chronic granulomatous disease
D. Myeloperoxidase deficiency
E. Hyper-IgE syndrome

Answer: A This patient's low levels of glutathione synthetase suggest a mild form of glutathione synthetase deficiency. Diagnosis can be confirmed with 5-oxoproline buildup and mutations in the *GSS* gene. Antioxidant supplementations are suggested for treatment.

4. The attraction of neutrophils and other white blood cells to an inflammatory site is called which of the following?

A. Chemotaxis
B. Margination
C. Diapedesis
D. Phagocytosis
E. Extracellular trap

Answer: A Chemotaxis is the attraction and oriented movement of neutrophils to an inflammatory site after the release of chemoattractants.

5. A 15-month-old female infant is admitted for recurrent deep skin infections. Flow cytometry confirmed deficiency in vitamin B_{12}. Blood smear displays bilobed neutrophil nucleuses. What is the likeliest diagnosis?

A. Specific granule deficiency
B. Chronic granulomatous disease
C. Interferon-γ receptor-1 defect
D. Myeloperoxidase deficiency
E. Macrophage activation syndrome

Answer: A Morphologically altered neutrophils are a strong indication of specific granule deficiency. This condition usually manifests during infancy, including recurrent skin infections and vitamin B_{12} deficiency. Treatment options include parenteral antibiotics and drainage of infections.

170 EOSINOPHILIC SYNDROMES

MARC E. ROTHENBERG

DEFINITION

Eosinophilic syndromes are a heterogeneous group of disorders that involve eosinophilia, which is defined as the accumulation of eosinophils in peripheral blood and/or tissues. Circulating eosinophils normally account for only 1 to 3% of peripheral blood leukocytes, and the upper limit of the normal range is 350 cells/mm^3 of blood. Eosinophilia occurs in a variety of disorders (Table 170-1) and is usually arbitrarily classified according to the degree of blood eosinophilia: mild (351 to 1500 cells/mm^3), moderate (>1500 to 5000 cells/mm^3), or severe (>5000 cells/mm^3). Tissue eosinophilic disorders, such as eosinophil-associated gastrointestinal disorders and eosinophilic fasciitis, are not necessarily associated with blood eosinophilia, so their diagnosis is based on the microscopic identification of eosinophil-rich inflammatory infiltrates associated with tissue damage.

Historically, hypereosinophilic syndromes were generally classified as idiopathic and were defined by (1) the presence of eosinophilia (>1500 cells/mm^3 for at least 6 months) that remained unexplained despite a comprehensive evaluation for known causes of eosinophilia (such as drug reactions and infections) and (2) evidence of organ dysfunction directly attributable to the eosinophilia. Now, however, it is known that in some patients, an acquired genetic etiology is responsible. These include the (a) *FIP1L1*–platelet-derived growth factor receptor-α (*PDGFRA*) fusion gene associated with a microdeletion on 4q24; and (b) other abnormalities of 4q12 (*PDGFRA* fusion partners instead of *FIP1L1*), 5q31-33 (*PDGFRB*), 8p11-13 (*FGFR1*), 9p24 (*JAK2*), and 13q12 (*FLT3*). These are detected by conventional cytogenetics, fluorescent in situ hybridization (FISH), and reverse transcription–polymerase chain reaction (RT-PCR).[1] Identification of these diseases has important therapeutic implications because they can be treated with targeted agents like imatinib, a tyrosine kinase inhibitor.

EPIDEMIOLOGY

The most common cause of eosinophilia worldwide is helminth infections, which affect hundreds of millions of people. The most frequent cause in industrialized nations is atopic disease, which affects 10 to 30% of the population. Hypereosinophilic disorders such as *FIP1L1-PDGFRA*–associated disease and Churg-Strauss syndrome (Chapter 270) are very rare. For example, Churg-Strauss syndrome affects 4 to 6 cases per million per year, whereas true idiopathic hypereosinophilic syndromes may affect only 4000 to 5000 people worldwide. Other syndromes such as eosinophil-associated gastrointestinal disorders are more common, with a prevalence of approximately 1 in 2000 individuals.

TABLE 170-1 CAUSES OF EOSINOPHILIA

REACTIVE EOSINOPHILIA
Allergic diseases—asthma, atopic dermatitis, allergic rhinitis
Drug reactions—including cytokine infusions; drug reaction (rash) with eosinophilia and systemic symptoms (DRESS) syndrome
Infection—viral (human immunodeficiency virus) or fungal (allergic bronchopulmonary aspergillosis, coccidioidomycosis)
Parasitic infection—mostly helminths
EOSINOPHILIA ASSOCIATED WITH OTHER DISEASES
Eosinophil-associated gastrointestinal disorders—eosinophilic esophagitis, gastroenteritis
Skin—bullous pemphigoid, urticaria, eosinophilic cellulitis, episodic angioedema
Pulmonary—eosinophilic pneumonia, allergic bronchopulmonary aspergillosis
Neurologic—eosinophilic meningitis
Autoimmune—Churg-Strauss syndrome, eosinophilic fasciitis
Primary immunodeficiency—hyperimmunoglobulin E syndrome, Omenn syndrome
Post-transplantation status—liver (in association with immunosuppression)
Transplant rejection—lung, kidney, liver
Malignancy—Hodgkin disease, solid tumors
Hypoadrenalism—Addison disease, adrenal hemorrhage
Renal—drug-induced interstitial nephritis, eosinophilic cystitis, dialysis
PRIMARY AND CLONAL EOSINOPHILIAS*
Myeloid and lymphoid neoplasms with *PDGFRA* rearrangement
Myeloid neoplasms with *PDGFRB* rearrangement
Myeloid and lymphoid neoplasms with *FGFR1* abnormalities
Chronic eosinophilia leukemia not otherwise specified (CEL-NOS)
Idiopathic hypereosinophilic syndrome (HES)
Idiopathic hypereosinophilia

*These represent the revised World Health Organization (WHO) classification (Gottlib J. World Health Organization-defined eosinophilic disorders: 2011 update on diagnosis, risk stratification, and management. *Am J Hematol* 2011;86:677-688.)

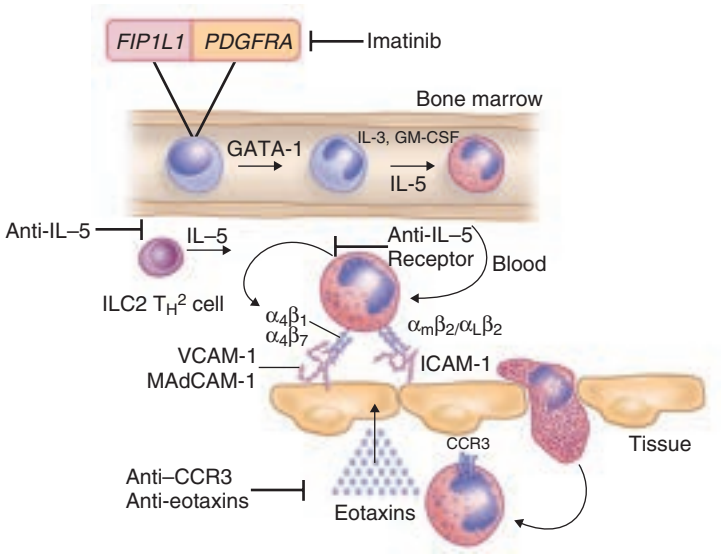

FIGURE 170-1. Schematic representation of eosinophil development, tissue recruitment, and therapeutic intervention. Eosinophil lineage development is specified by the GATA-1 transcription factor and promoted by the cytokines interleukin-3 (IL-3), IL-5, and granulocyte-macrophage colony-stimulating factor (GM-CSF). IL-5 is most selective to the eosinophil lineage and regulates eosinophil movement from the bone marrow into the peripheral blood. Eosinophil adhesion is mediated by β1, β2, and β7 integrins and their interaction with the endothelial adhesion molecules intercellular adhesion molecule 1 (ICAM-1), vascular cell adhesion molecule 1 (VCAM-1), and mucosal address in cell adhesion molecule 1 (MAdCAM-1). Recruitment of eosinophils into tissue is regulated by the eotaxin chemokines that stimulate eosinophilic chemoattraction and activation through their receptor CCR3. Hypereosinophilic syndromes can develop after an 800-kilobase microdeletion on chromosome 4 results in fusion of the *FIP1L1* and *PDGFRA* genes, thereby resulting in activation of an imatinib-sensitive tyrosine kinase. Targeted therapeutic intervention for eosinophilic syndromes includes anti-IL–5 and anti–CCR3/anti-eotaxins, which are currently in clinical development.

PATHOBIOLOGY

Eosinophils are multifunctional leukocytes that are produced in the bone marrow from pluripotential hematopoietic stem cells under regulation of the transcription factor GATA-1 and the cytokines interleukin-3 (IL-3), IL-5, and granulocyte-macrophage colony-stimulating factor (GM-CSF) (Fig. 170-1). They are capable of producing a wide variety of pro-inflammatory mediators and immunomodulatory molecules.[2] Eosinophils are under the regulation of helper type 2 T cells (T_H2) and type 2 innate lymphoid cells (ILC2) that secrete IL-4, IL-5, and IL-13. Notably, IL-5 is a cytokine that specifically regulates the selective differentiation of eosinophils, their release from bone marrow into the peripheral circulation, and their survival. IL-5 activity is counterbalanced by paired immunoglobulin-like receptors (PIR) expressed on eosinophils, which counterbalance cellular activation and inhibition. A humanized anti-IL–5 drug markedly lowers blood eosinophilia and reduces tissue eosinophilia more modestly. Recent preliminary studies in patients with severe asthma have shown that anti-IL–5 therapy improves asthma control, including exacerbations, and allows steroid reduction. Similarly, anti-IL–5 has a steroid-sparing effect in hypereosinophilic syndromes. Humanized anti-IL–5 therapy and cytotoxic anti-IL–5 receptor therapy are currently in clinical testing for a variety of indications, including eosinophilic esophagitis, asthma, and hypereosinophilic syndromes. IL-4 and IL-13 induce eosinophil recruitment and survival, expression of critical adhesion molecules on the endothelium that bind to the β1 and β2 integrins on eosinophils (such as intercellular adhesion molecule 1 [ICAM-1] and vascular cell adhesion molecule 1 [VCAM-1]), and eosinophil-active chemokines such as the eotaxins. The eotaxins are three structurally related eosinophil chemoattractant and activating proteins that signal exclusively through the eosinophil-selective receptor CCR3. In addition to regulating the baseline homing of eosinophils to the various tissues in which they normally predominantly reside under the regulation of ILC2, such as the gastrointestinal tract, the eotaxins are induced by T_H2-associated inflammatory triggers (e.g., IL-13) and thereby promote tissue accumulation of eosinophils. Humanized antibodies against the eotaxins and small-molecule inhibitors against CCR3 are promising new approaches for treating eosinophilic disorders that are in clinical development.

Eosinophil granules contain a crystalloid core composed of major basic protein (MBP-1 and MBP-2), as well as a matrix composed of eosinophil cationic protein (ECP), eosinophil-derived neurotoxin (EDN), and eosinophil peroxidase (EPO). MBP, EPO, and ECP have cytotoxic effects on a variety of tissues in concentrations similar to those found in biologic fluids from patients with eosinophilia. Additionally, ECP and EDN belong to the ribonuclease A superfamily and possess antiviral and ribonuclease activity. ECP can insert voltage-insensitive, ion-nonselective toxic pores into the membranes of target cells, and these pores may facilitate the entry of other toxic molecules. MBP directly increases smooth muscle reactivity by causing dysfunction of vagal muscarinic M_2 receptors, and this process has been postulated to contribute to the airway hyperresponsiveness associated with asthma. MBP also triggers degranulation of mast cells and basophils. Triggering of eosinophils by engagement of receptors for cytokines, immunoglobulins, and complement can lead to the generation of a wide range of inflammatory cytokines, including IL-1, IL-3, IL-4, IL-5, IL-13, GM-CSF, transforming growth factor α/β, tumor necrosis factor α, RANTES, macrophage inflammatory protein 1α (MIP-1α), and the eotaxins, thus indicating that eosinophils have the potential to modulate multiple aspects of the immune response. Additionally, eosinophils can directly activate T cells by antigen presentation and help polarize dendritic cells to promote a T_H2 phenotype.[3] Further eosinophil-mediated tissue damage is caused by toxic hydrogen peroxide and halide acids generated by EPO and by superoxide generated by the respiratory burst oxidase enzyme pathway in eosinophils. Eosinophils also generate large amounts of cysteinyl leukotriene C_4 (LTC_4), which is metabolized to LTD_4 and LTE_4. These three lipid mediators increase vascular permeability and mucus secretion and are potent stimulators of smooth muscle contraction. Finally, bipyramidal Charcot-Leyden crystals are derived from a nongranule lysophospholipase in eosinophils and are frequently found in sputum, feces, and tissues infiltrated by eosinophils.

CLINICAL MANIFESTATIONS

Hypereosinophilia is often recognized on a routine blood count in a patient who is asymptomatic or being evaluated for unrelated or nonspecific signs or symptoms. On other occasions, the possibility of eosinophilia may be specifically investigated in a patient with gastrointestinal or respiratory symptoms because helminthic disease or allergic causes are suspected. The clinical signs and symptoms of hypereosinophilic syndromes are heterogeneous because of the diversity of the causes and potential organ involvement. Common signs and symptoms include dermatitis, heart failure, neuropathy, and abdominal pain. One of the most serious complications of hypereosinophilia

is cardiac disease secondary to endomyocardial thrombus formation and restrictive fibrosis (Chapter 60). Mitral and tricuspid valve regurgitation may result from progressive fibrotic damage to the chordae tendineae, and resultant heart failure can develop from valvular insufficiency and endomyocardial fibrosis. Cardiac involvement can occur in association with chronic eosinophilia from diverse causes, including parasitic infections. Hypereosinophilic syndromes can result in cerebral emboli from cardiac disease, diffuse encephalopathy, and peripheral neuropathy.

DIAGNOSIS

Differential Diagnosis

The differential diagnosis of eosinophilia includes reactive eosinophilia, eosinophilia associated with other primary disorders, and eosinophilia associated with clonal hematopoiesis (see Table 170-1). Evaluation of patients is based on their history and clinical characteristics (Fig. 170-2). The initial goal is to determine whether the eosinophilia is secondary to a reactive cause (i.e., in response to another primary trigger such as allergy, infection, solid tumor, vasculitis). If reactive causes are not identified, further evaluation should determine whether the eosinophilia is secondary to a clonal hematologic disorder. If no evidence of clonality is determined, the patient is considered to have an idiopathic hypereosinophilic syndrome. The diagnosis of idiopathic hypereosinophilic *syndrome* requires an absolute eosinophil count of greater than 1500 cells/mm^3 *and* evidence of organ involvement and dysfunction.[4] These diagnostic criteria have been challenged for reasons noted in the Treatment section. Eosinophilia of an as yet unknown etiology that does not meet these criteria should be called *idiopathic hypereosinophilia*.

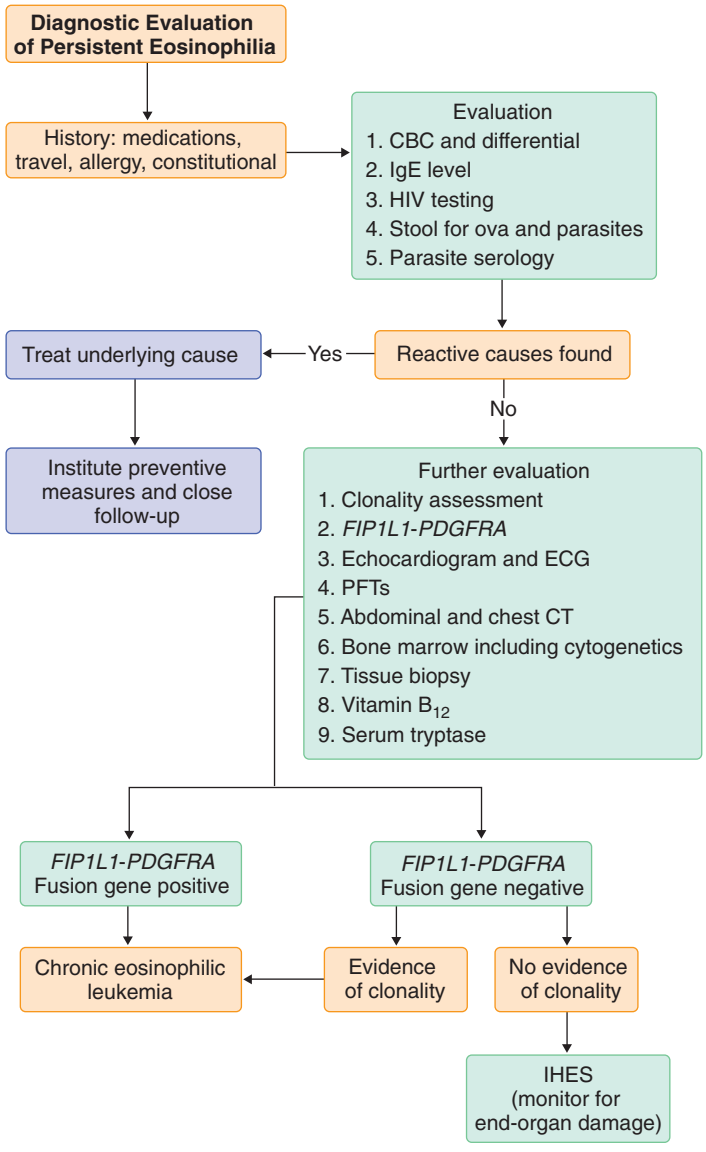

FIGURE 170-2. Diagnostic evaluation of persistent eosinophilia. CBC = complete blood count; CT = computed tomography; ECG = electrocardiogram; IgE = immunoglobulin E; IHES = idiopathic hypereosinophilic syndrome; HIV = human immunodeficiency virus; PDGFRA = platelet-derived growth factor receptor-α; PFTs = pulmonary function tests.

The differential diagnosis of eosinophilia requires a review of the patient's history, which may reveal wheezing (Chapter 87), rhinitis (Chapter 251), or eczema (indicating atopic causes); travel to areas where helminth infections (e.g., schistosomiasis [Chapter 355]) are endemic; the presence of a pet dog (indicating possible infection with *Toxocara canis* [Chapter 357]); symptoms of cancer; or drug ingestion (indicating a possible hypersensitivity reaction [Chapter 254]). Eosinophilia caused by drugs (Chapter 254) is usually benign but can sometimes be accompanied by tissue damage, as in hypersensitivity pneumonitis (Chapter 97) and in *DRESS syndrome* (drug reaction or rash with eosinophilia and systemic symptoms) (also see Chapter 440).[5] In most cases, the eosinophilia resolves when use of the drug ceases, but in some cases, such as eosinophilia-myalgia syndrome secondary to the ingestion of contaminated L-tryptophan, the disease can persist despite withdrawal of the drug.

The presence of abnormal morphologic features of eosinophils, an increase in immature and dysplastic cells in the bone marrow or blood, elevated levels of vitamin B_{12}, and splenomegaly raises suspicion of a clonal hypereosinophilic syndrome. In such cases, evidence of clonality (e.g., by analysis of X-chromosome inactivation patterns in female patients), an elevated level of mast cell tryptase (elevated in myelodysplastic variants of hypereosinophilic syndrome), the presence of aberrant lymphocyte phenotypes (elevated in lymphocytic variants of hypereosinophilic syndrome), abnormal cytogenetics, and the possible presence of specific fusion genes such as *FIP1L1-PDGFRA* should be investigated.

Other eosinophilic syndromes, such as Churg-Strauss syndrome, which is now referred to as eosinophilic granulomatosis with polyangiitis [EGPA] (Chapter 270), should be considered in patients with a history of worsening asthma, sinus disease, neuropathy, or blood eosinophilia and the presence of abnormal laboratory findings associated with inflammation and autoimmunity, such as an elevated erythrocyte sedimentation rate, C-reactive protein, and antineutrophil cytoplasmic antibodies.

An accumulation of eosinophils that is limited to specific organs is characteristic of particular diseases, such as eosinophilic cellulitis (Wells syndrome), eosinophilic esophagitis (Chapter 138), eosinophilic pneumonias (e.g., Löffler syndrome [Chapter 92]), and eosinophilic myositis (which now includes a genetic etiology caused by recessive mutations in calpain3 *[CAPN3]*).

Diagnostic Evaluation

Diagnostic studies that should be performed in patients with moderate to severe eosinophilia and considered in patients with persistent mild eosinophilia include morphologic examination of a blood smear, human immunodeficiency virus (HIV) screen, serial stool examinations for ova and parasites, parasite serology, and plasma immunoglobulin E (IgE) level.[6] Parasitic infections that cause eosinophilia are usually limited to helminthic parasites, with the exception of two enteric protozoans, *Isospora belli* (Chapter 353) and *Dientamoeba fragilis* (Chapter 353). *Strongyloides stercoralis* (Chapter 358) infection is important to diagnose because it can cause disseminated fatal disease in immunosuppressed patients; detection of such infection often requires serologic testing. Other infections to consider include trichinosis (Chapter 358), *T. canis* infection (Chapter 357), and HIV infection (Chapter 393).

Patients with sustained hypereosinophilia should be monitored closely for the subsequent development of cardiac disease. A pathologically similar disease, Löffler endomyocarditis (Chapter 60), has been noted in tropical regions, where antecedent parasite-elicited eosinophilia may be responsible for the cardiac damage. There should be a low threshold for a bone marrow analysis and testing for the presence of the *FIP1L1-PDGFRA* fusion gene in patients with hypereosinophilia. Testing for the presence of other activated tyrosine kinases (e.g., *PDGFRA* and *FGFR1*) should generally be reserved for individuals with bone marrow cytogenetic abnormalities.

TREATMENT

Reactive Hypereosinophilia and Hypereosinophilia Associated with Other Diseases

Treatment of reactive hypereosinophilia and hypereosinophilia associated with other diseases centers around identifying the cause and then treating the

underlying disease process.[7,8] For example, reactive eosinophilia typically responds by removal of the inciting triggers (e.g., allergens, parasites, and medications). Eosinophilia associated with other disease processes typically improves after treatment of the underlying disease, such as dietary manipulation in patients with allergic eosinophilic gastroenteritis.

FIP1L1-PDGFRA–Positive Disease

As noted in the next section, other hypereosinophilic syndromes, treatment should be started as soon as possible to prevent potentially serious eosinophilia-mediated organ damage. Imatinib should be considered as first-line therapy in patients in whom the *FIP1L1-PDGFRA* fusion gene has been demonstrated and in selected patients with the characteristic clinical, laboratory, and molecular features of this myeloproliferative subtype of hypereosinophilic syndrome (e.g., male gender, tissue fibrosis, elevated serum vitamin B_{12} and tryptase levels). Clinical responses to imatinib in *FIP1L1-PDGFRA*–positive patients are rapid, with normalization of eosinophil counts generally occurring within 1 week of initiation of treatment and reversal of the signs and symptoms occurring within 1 month. Doses of imatinib as low as 100 mg daily appear to be effective in controlling symptoms and eosinophilia in most patients, but some recommend beginning imatinib treatment at 400 mg daily to achieve molecular remission and then decreasing the dose slowly while monitoring the patient closely for evidence of molecular relapse. In imatinib-resistant patients, sorafenib may be effective. The utility of imatinib therapy in hypereosinophilic patients without a demonstrable *FIP1L1-PDGFRA* mutation remains controversial, although some patients have responded. Nonmyeloablative allogeneic bone marrow transplantation (Chapter 178) has also been used successfully in several patients with hypereosinophilia.

Other Hypereosinophilic Syndromes

The current diagnostic criteria for hypereosinophilic syndromes that require at least 6 months of persistent eosinophilia of more than 1500 cells/mm^3 have been challenged because prompt diagnosis and treatment are required (before 6 months have passed) to prevent potentially serious end-organ damage. New diagnostic criteria have been proposed to address this limitation.[9]

Corticosteroids, which have been used for decades in the treatment of idiopathic hypereosinophilic syndromes, remain the first-line treatment for most patients, except those with *PDGFRA*-associated hypereosinophilia. The most appropriate initial corticosteroid dose and the duration of steroid therapy have not been subjected to randomized trials, but a general recommendation is to start with a moderate to high dose (≥40 mg prednisone equivalent) and taper very slowly while monitoring the eosinophil count closely. With this approach, most patients will respond initially, and some can be maintained on low doses of corticosteroids for prolonged periods.

Monoclonal anti-IL–5 antibody therapy (e.g., mepolizumab or reslizumab) for hypereosinophilia and eosinophil-associated asthma has a number of unique advantages related to the specificity of IL-5 for the eosinophil lineage. In patients treated with this agent, eosinophil counts are twice as likely to fall below 600/μL (95% vs. 45%, $P < .001$) with significantly lower prednisone doses. Of the cytotoxic therapies that have been used for steroid-refractory hypereosinophilia, hydroxyurea has been the most extensively studied at doses of 1 to 3 g/day. Vincristine at a dose of 1 to 2 mg intravenously can rapidly lower eosinophilia in patients with extremely high eosinophil counts (>100,000/mm^3) and may be useful for the treatment of children whose aggressive disease is unresponsive to other therapies. In patients who have corticosteroid-refractory hypereosinophilic syndromes or who develop intolerable side effects of steroid treatment, immunomodulatory agents that are sometimes helpful include interferon-α, cyclosporine, and alemtuzumab. Responses can often be achieved with relatively low doses of interferon-α (1 to 2 × 10^6 U/day) and may persist for prolonged periods. Because the effects of interferon-α on eosinophil numbers in peripheral blood may not become evident for several weeks, escalation to an effective dose may require several months. Rarely, patients have remained in remission for extended periods after cessation of interferon-α therapy, suggesting that interferon-α may be curative in a small subset of individuals. Low-dose (500 mg daily) hydroxyurea appears to act synergistically with interferon-α to lower the eosinophil count without increasing side effects.

PROGNOSIS

The prognosis of hypereosinophilic syndromes depends on the primary cause. Whereas *FIP1L1-PDGFRA*–positive disease and other forms of clonal disorders have a poor prognosis (25 to 50% 5-year mortality rate if responsiveness to therapeutic intervention is not achieved), the prognosis of hypereosinophilia from reactive and other causes is usually better and continues to improve.

A recent retrospective review of 247 cases of hypereosinophilic syndrome seen at the Mayo Clinic over a period of 19 years showed that only 23 patients died during this time, with the most common causes of death being cardiac dysfunction, infection, unrelated malignancy, and thromboembolic and vascular disease.[10] It was noted that targeted monitoring of at-risk end organs, combined with early treatment, may further improve survival and reduce morbidity.

FUTURE DIRECTIONS

Treatments on the horizon for hypereosinophilic disorders include targeted therapy against the eotaxin chemokines and their receptor CCR3, as well as anti-IL–5 and anti-IL–5 receptor–based antibody therapy.[11]

Grade A References

A1. Nair P, Pizzichini MM, Kjarsgaard M, et al. Mepolizumab for prednisone-dependent asthma with sputum eosinophilia. *N Engl J Med*. 2009;360:985-993.

A2. Haldar P, Brightling CE, Hargadon B, et al. Mepolizumab and exacerbations of refractory eosinophilic asthma. *N Engl J Med*. 2009;360:973-984.

A3. Castro M, Mathur S, Hargreave F, et al. Reslizumab for poorly controlled, eosinophilic asthma: a randomized, placebo-controlled study. *Am J Respir Crit Care Med*. 2011;184:1125-1132.

A4. Rothenberg ME, Klion AD, Roufosse FE, et al. Treatment of patients with the hypereosinophilic syndrome with mepolizumab. *N Engl J Med*. 2008;358:1215-1228.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Noel P, Mesa R. Eosinophilic myeloid neoplasms. *Curr Opin Hematol*. 2013;20:157-162.
2. Kita H. Eosinophils: multifunctional and distinctive properties. *Int Arch Allergy Immunol*. 2013;161(suppl 2):3-9.
3. Nussbaum JC, Van Dyken SJ, von Moltke J, et al. Type 2 innate lymphoid cells control eosinophil homeostasis. *Nature*. 2013;502:245-248.
4. Valent P, Klion AD, Rosenwasser LJ, et al. ICON: eosinophil disorders. *World Allergy Org J*. 2012;5:174-181.
5. Husain Z, Reddy BY, Schwartz RA. DRESS syndrome: part 1. Clinical perspectives. *J Am Acad Dermatol*. 2013;68:693.e1-693.e14.
6. Roufossue F, Weller PF. Practical approach to the patient with hypereosinophilia. *J Allergy Clin Immunol*. 2010;126:39-44.
7. Fulkerson PC, Rothenberg ME. Targeting eosinophils in allergy, inflammation and beyond. *Nat Rev Drug Discov*. 2013;12:117-129.
8. Davis BP, Rothenberg ME. Eosinophils and cancer. *Cancer Immunol Res*. 2014;2:1-8.
9. Gotlib J. World Health Organization-defined eosinophilic disorders: 2014 update on diagnosis, risk stratification, and management. *Am J Hematol*. 2014;89:325-337.
10. Podjasek JC, Butterfield JH. Mortality in hypereosinophilic syndrome: 19 years of experience at Mayo Clinic with a review of the literature. *Leuk Res*. 2013;37:392-395.
11. Bochner BS, Book W, Busse WW, et al. Workshop report from the National Institutes of Health Taskforce on the Research Needs of Eosinophil-Associated Diseases (TREAD). *J Allergy Clin Immunol*. 2012;130:587-596.

REVIEW QUESTIONS

1. A 40-year-old man presents with blood eosinophilia and rash. The chest radiograph is normal, and the blood eosinophil count is 5000 cells/mm^3 on two occasions over the course of 1 month. He has no history of allergies and no risk factors for parasite infection. A bone marrow biopsy shows elevated eosinophils with no blasts. The next step would be which of the following?

A. Lymphocyte phenotyping and serum immunoglobulin levels
B. Determining the presence of the *FIP1L1-PDGFRA* fusion gene
C. Following the eosinophil count over the course of 6 months to document true hypereosinophilia
D. Determining IgE and HIV titers
E. All of the above

Answer: B The patient is presenting with classic signs of hypereosinophilic syndrome and should be evaluated for *FIP1L1-PDGFRA* because its presence would lead to imatinib therapy. Delay of 6 months exposes the patient to the potentially unnecessary risk for eosinophilia-mediated organ damage.

2. An asthmatic young man presents with progressive difficulty swallowing that is refractory to proton pump inhibitor therapy for the past 2 months. The peripheral blood count and differential are normal, specifically without eosinophilia. Which of the following is true regarding eosinophilic esophagitis?

A. It is not indicated because the patient does not have blood eosinophilia.
B. It should only be considered after repeating the blood eosinophil count over a period of time.
C. It can be assumed and treated with anti-IL–5 therapy.
D. It should be ruled out with a referral to a gastroenterologist for endoscopic evaluation.
E. It should only be considered after checking for the presence of the *FIP1L1-PDGFRA* fusion gene.

Answer: D Eosinophilic esophagitis is a tissue-specific eosinophilic disorder that can only be diagnosed by endoscopic biopsy and does not have to be accompanied by peripheral blood eosinophilia.

3. A middle-age women presents with progressive weight loss and diarrhea and is found to have blood eosinophil count of 5000 cells/mm^3. She is a nonatopic individual and has recently returned from a trip to Brazil. Which of the following would be the next step in evaluating this person?

A. Bone marrow analysis and cytogenetics
B. HIV titer
C. Stool culture and evaluation for parasites
D. Colonoscopy
E. Serologic testing for inflammatory bowel disease

Answer: C The patient has an acute presentation of eosinophilia following travel to a region likely endemic for parasites such as schistosomiasis. The next step would be a stool evaluation for parasites.

4. An 18-year-old male presents with a generalized erythematous eruption and peripheral blood eosinophilia of 2500 cells/mm^3. The history is significant for recent antibiotic use for a cellulitis in the leg. The cellulitis has not improved despite 10 days of antibiotics. The next step in evaluating this patient would include consideration of which of the following?

A. The patient likely has a simple drug allergy and should switch antibiotics.
B. The patient may have an underlying hypereosinophilic syndrome and should undergo skin biopsy, particularly at the site of the cellulitis.
C. The patient should have a chest and abdominal computed tomography scan to rule out solid tumor malignancy.
D. Referral should be made to a cardiologist to rule out cardiac involvement.
E. It is critical to rule out parasitic infection as the next step.

Answer: B The most common organ involved in hypereosinophilic syndrome is the skin. Although this patient may have a drug reaction, the persistent cellulitis is concerning for an underlying primary inflammatory process.

5. A well-appearing 60-year old women is found to have persistent blood eosinophil counts of 3500 cells/mm^3. She denies risk factors for parasite infection, is nonatopic, and has started no new medication, although she has been taking a statin and antihypertensive agent for several years. Which of the following considerations should take place?

A. Because this is a well-appearing female patient, there is little concern for hypereosinophilic syndrome.
B. Complete work-up is indicated that includes a computed tomography scan of the abdomen and bone marrow analysis.
C. Consideration of drug-induced eosinophilia is not appropriate because the patient has not started any new medications.
D. The patient should be simply followed by serial complete blood counts because she is clinically well.
E. A stool sample for ova and parasite should be the next step.

Answer: E The next logical step would be ruling out an infection. Serial complete blood counts should be performed but not in isolation because it has already been established that the patient's eosinophilia is persistent.

171

APPROACH TO THE PATIENT WITH BLEEDING AND THROMBOSIS

ANDREW I. SCHAFER

MECHANISMS OF HEMOSTASIS AND THROMBOSIS

Normal Hemostasis

The coagulation system is normally quiescent, and blood fluidity is maintained by the actions of a continuous monolayer of endothelial cells that line the intimal surface of the vasculature throughout the circulatory tree. At a site of vascular damage, the antithrombotic properties of endothelium are lost, and thrombogenic constituents of the subendothelial vessel wall become exposed to circulating blood. The result is rapid formation of a hemostatic clot that consists of platelets and fibrin and is localized to the area of vascular injury. Activation of platelets and formation of fibrin occur essentially simultaneously and interdependently to effect hemostasis. Subsequently, vascular repair is accomplished by thrombolysis and recanalization of the occluded site.[1]

Platelet activation at a site of vascular injury begins with the adhesion of platelets to the locally de-endothelialized intimal surface (platelet–vessel wall interaction). Platelet adhesion is mediated by von Willebrand factor, which sticks circulating platelets to the area of damaged vessel wall by binding to its receptors located in platelet membrane glycoprotein Ib. The adherent platelets then undergo a "release reaction," during which they discharge constituents of their storage granules, including adenosine diphosphate (ADP), and simultaneously elaborate thromboxane A_2 from arachidonic acid through the aspirin-inhibitable cyclooxygenase reaction. ADP, thromboxane A_2, and other components of the release reaction act in concert to recruit and activate additional platelets from the circulation to the site of vascular injury. These activated platelets expose binding sites for fibrinogen by forming the surface membrane glycoprotein IIb/IIIa complex. In the process of platelet aggregation (platelet-platelet interactions), fibrinogen (or von Willebrand factor under conditions of high shear stress) mediates the final formation of an occlusive platelet plug.

Fibrin, which anchors the hemostatic platelet plug, is formed from soluble plasma fibrinogen by the action of the potent protease enzyme thrombin (Fig. 171-1). The fibrin mesh is stabilized by covalent cross-linking mediated by factor XIII. Thrombin is formed from its inactive (zymogen) plasma

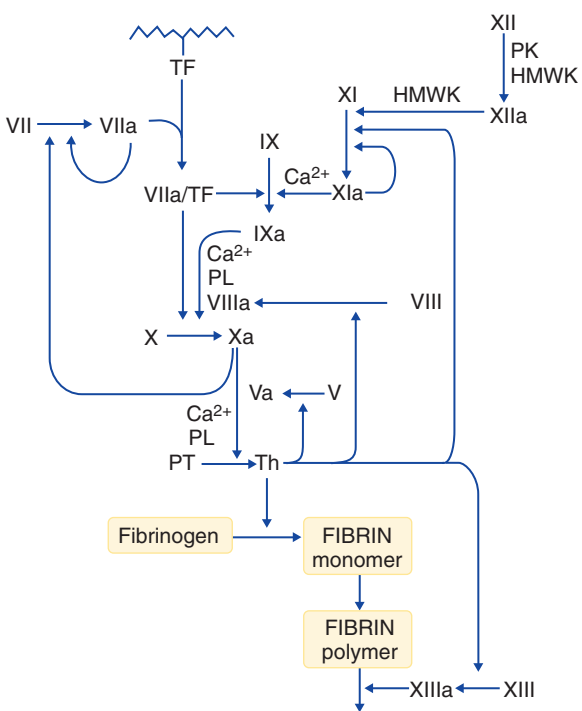

FIGURE 171-1. Coagulation cascade. This scheme emphasizes an understanding of (1) the importance of the tissue factor pathway in initiating clotting in vivo, (2) the interactions among pathways, and (3) the pivotal role of thrombin in sustaining the cascade by feedback activation of coagulation factors. HMWK = high-molecular-weight kininogen; PK = prekallikrein; PL = phospholipid; PT = prothrombin; TF = tissue factor; Th = thrombin. (From Schafer AI. Coagulation cascade: an overview. In: Loscalzo J, Schafer AI, eds. *Thrombosis and Hemorrhage.* Cambridge, MA: Blackwell Scientific Publications; 1994:3-12.)

precursor, prothrombin, by the action of activated factor X (Xa) and its cofactor, factor Va. This sequence of reactions has classically been referred to as the *common pathway* of coagulation. Factor X can be activated by either the *tissue factor (extrinsic) pathway* or the *contact activation (intrinsic) pathway* of coagulation. The tissue factor pathway is now considered to be the major physiologic initiator of coagulation activation. It is triggered by the formation of the complex of tissue factor, which is exposed on the surfaces of activated vascular and blood cells, with activated factor VII (VIIa). The contact activation pathway involves a series (or cascade) of zymogen-protease reactions that are initiated by factor XII, high-molecular-weight kininogen, and prekallikrein. Activated factor XII (XIIa) converts factor XI to XIa, which in turn activates factor IX to IXa. Factor IXa is the enzyme that converts factor X to Xa, a reaction that requires factor VIIIa as a cofactor.

Physiologic Antithrombotic Mechanisms

Intact, normal endothelium promotes blood fluidity by inhibiting platelet activation. It likewise plays a crucial role in preventing fibrin accumulation. Among the physiologic antithrombotic systems that produce this latter effect are (1) antithrombin III, (2) protein C and protein S, (3) tissue factor pathway inhibitor (TFPI), and (4) the fibrinolytic system. Antithrombin is the major protease inhibitor of the coagulation system[2]: it inactivates thrombin and other activated coagulation factors. Heparin functions as an anticoagulant by binding to antithrombin and greatly accelerating these reactions. Heparin and heparin sulfate proteoglycans are naturally present on endothelial cells, so antithrombin inactivation of thrombin and other coagulation proteases most likely occurs physiologically on vascular surfaces rather than in fluid plasma. Activated protein C, with its cofactor protein S, functions as a natural anticoagulant by destroying factors Va and VIIIa, two essential cofactors of the coagulation cascade. Thrombin itself is the activator of protein C, and this reaction occurs rapidly only on the surfaces of intact vascular endothelial cells, where thrombin binds to the glycosaminoglycan thrombomodulin. TFPI is a plasma protease inhibitor that specifically quenches tissue factor–induced coagulation. Finally, what little fibrin can be produced, despite these potent physiologic antithrombotic mechanisms, is digested rapidly by the endogenous fibrinolytic system. Fibrinolysis is mediated by the protease plasmin, which is generated from plasminogen in plasma by the action of endothelium-derived plasminogen activators.

EVALUATION OF THE PATIENT WITH A POSSIBLE BLEEDING DISORDER

History and Physical Examination

A thorough history is paramount in evaluating a patient for a possible systemic bleeding disorder. In addition to asking the patient about spontaneous bleeding episodes in the past, the responses to specific hemostatic challenges should be recorded. A bleeding tendency may be suspected if a patient previously experienced excessive hemorrhage after surgery or trauma, including common events such as circumcision, tonsillectomy, labor and delivery, menses, dental procedures, vaccinations, and injections. Conversely, a history of normal blood clotting after such specific challenges in the recent past is just as important to note. It may be a better test of the integrity of systemic hemostasis than any laboratory measurement can provide.

In a patient with a history of excessive or unexplained bleeding, the initial goal is to determine whether the cause is a systemic coagulopathy or localized anatomic or mechanical problem with a blood vessel. This situation is encountered most frequently in patients with excessive postoperative bleeding, which could be due to either local surgical trauma or a coagulation abnormality. A history of prior bleeding suggests a coagulopathy, as does the finding of bleeding from multiple sites. However, this is not always the case. Even diffuse bleeding may arise from anatomic rather than hemostatic abnormalities. An example of this is recurrent mucosal hemorrhage in patients with hereditary hemorrhagic telangiectasia (Chapter 173). Conversely, a single episode of bleeding from an isolated site may be the initial manifestation of a coagulopathy.

The history must include a survey of coexisting systemic diseases and drug ingestions that could affect hemostasis. Renal failure and the myeloproliferative disorders are associated with impaired platelet–vessel wall interactions and qualitative platelet abnormalities, connective tissue diseases and lymphomas are associated with thrombocytopenia, and liver disease causes a complex coagulopathy (Chapter 175). Ingestion of aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) that cause nonselective inhibition of cyclooxygenase leads to platelet dysfunction; these drugs are often contained in over-the-counter preparations that patients may neglect to report without specific questioning. Other drugs, such as antibiotics, also may be associated with a bleeding tendency by causing abnormal platelet function or thrombocytopenia. Finally, it is important to elicit a family history of bleeding problems. Although a positive history provides an important clue to a possible inherited coagulopathy, a negative history does not exclude a familial cause; for example, 20% of patients with classic hemophilia have a completely negative family history of bleeding.

Mild bleeding events are commonly reported by patients with and without subsequently laboratory-documented bleeding disorders, sometimes making it difficult for hematologists to define a "significant bleeding history." Using a web-based questionnaire, 25% of subjects in a healthy population reported epistaxis, 18% easy bruising (more commonly in women), 18% prolonged bleeding after dental extraction, and 47% of women heavy menstrual bleeding.[3] More precise quantification of bleeding symptoms is being attempted by using "bleeding score" instruments such as the Vicenza bleeding score to help discriminate, in conjunction with laboratory testing, between healthy subjects and those with mild bleeding disorders.[4]

Patterns of clinical bleeding, as revealed by the history and physical examination, may be characteristic of certain types of coagulopathy (Table 171-1). In general, patients with thrombocytopenia or qualitative platelet or vascular disorders present with bleeding from superficial sites in the skin and mucous membranes. These may involve petechiae, which are pinpoint cutaneous hemorrhages that appear particularly over dependent extremities (characteristic of severe thrombocytopenia), ecchymoses (common bruises), purpura, gastrointestinal and genitourinary tract bleeding, epistaxis, and hemoptysis. In these types of disorders, bleeding tends to occur spontaneously or immediately after trauma. In contrast, patients with inherited or acquired coagulation factor deficiencies, such as hemophilia, or those on anticoagulant therapy tend to bleed from deeper tissue sites (e.g., hemarthroses, deep hematomas, retroperitoneal hemorrhage) and in a delayed manner after trauma.

Laboratory Testing

A few simple screening tests have traditionally been used in the initial evaluation of patients with a suspected coagulopathy: platelet count, bleeding

TABLE 171-1 CHARACTERISTIC PATTERNS OF BLEEDING IN SYSTEMIC DISORDERS OF HEMOSTASIS

TYPE OF DISORDER	SITES OF BLEEDING				ONSET OF BLEEDING	CLINICAL EXAMPLES
	General	*Skin*	*Mucous Membranes*	*Other*		
Platelet-vascular disorders	Superficial surfaces	Petechiae, ecchymoses	Common: oral, nasal, gastrointestinal, genitourinary	Rare	Spontaneous or immediately after trauma	Thrombocytopenia, functional platelet disorder, vascular fragility, disseminated intravascular coagulation, liver disease
Coagulation factor deficiency	Deep tissues	Hematomas	Rare	Common: joint, muscle, retroperitoneal	Delayed after trauma	Inherited coagulation factor deficiency, acquired inhibitor, anticoagulation, disseminated intravascular coagulation, liver disease

time, prothrombin time (PT) (also reported as the international normalized ratio, or INR), activated partial thromboplastin time (aPTT), and thrombin time (TT).[5] The North American Specialized Coagulation Laboratory Association (NASCOLA) has reported that most coagulation laboratories currently perform these tests, with the exception of the bleeding time, as their "bleeding disorder panels."

Thrombocytopenia, reported by electronic particle counting, should be verified by examination of the peripheral smear. Pseudothrombocytopenia, a laboratory artifact of ex vivo platelet clumping, may be caused by the ethylenediaminetetraacetic acid (EDTA) anticoagulant used in tubes for blood cell counts, by other anticoagulants, or by nonphysiologic cold agglutinins acting at room temperature. It should be suspected whenever a very low platelet count is unexpectedly reported in a patient who does not exhibit any clinical bleeding. Pseudothrombocytopenia is indicated by the finding of platelet clumps on the peripheral smear, and the diagnosis is supported by the finding of simultaneously normal platelet counts in blood samples obtained by finger stick, in tubes containing other anticoagulants, or in a tube maintained at 37° C before platelet counting. Examination of the blood smear can also reveal clues to the cause of real thrombocytopenia, such as fragmented red blood cells in thrombotic thrombocytopenic purpura.

The bleeding time was a widely used clinical screening test for disorders of platelet–vessel wall interactions. It measures the time to cessation of bleeding after a standardized incision over the volar aspect of the forearm. However, the test is prone to problems related to quality control, reproducibility, sensitivity, and specificity. Therefore, because von Willebrand disease is the most common genetic cause of abnormal platelet–vessel wall interactions, most experts now recommend replacing the bleeding time with specific tests for von Willebrand disease (VWD screen) in the initial evaluation of patients with a suspected coagulopathy (Chapter 173). As an additional replacement for the bleeding time, especially when a functional (qualitative) abnormality of platelets is suspected by characteristic mucocutaneous bleeding or bruising, a global assay of platelet function can be appended to the panel of screening tests. This is most commonly a platelet function analyzer (PFA) closure time. The PFA and other platelet function studies (see later) must be performed after discontinuation of drugs that interfere with platelet function (e.g., aspirin and other NSAIDs).

The PT measures the integrity of the extrinsic and common pathways of coagulation (factors VII, X, and V; prothrombin; and fibrinogen) (Fig. 171-2). The aPTT measures the integrity of the intrinsic and common pathways of coagulation (high-molecular-weight kininogen; prekallikrein; factors XII, XI, IX, VIII, X, and V; prothrombin; and fibrinogen). The sensitivity of the PT and aPTT in detecting coagulation factor deficiencies may vary with the reagents used to perform these tests, and each laboratory must determine its own reference standards. The TT is a screen for quantitative deficiencies and qualitative defects of plasma fibrinogen.

With a few notable exceptions, normal results for all these screening tests of hemostasis essentially exclude any clinically significant systemic coagulopathy. However, patients with factor XIII deficiency may have a serious bleeding diathesis but normal screening tests; specific tests for factor XIII deficiency should be performed if this disease is suspected (Chapter 174). The PT and aPTT detect only the more severe deficiencies of coagulation factors, usually at levels less than 30% of normal; specific factor levels should be determined if a mild coagulation factor deficiency is suspected. Rare disorders of fibrinolysis also may be associated with normal screening tests, necessitating more specialized tests when indicated (Chapter 174).

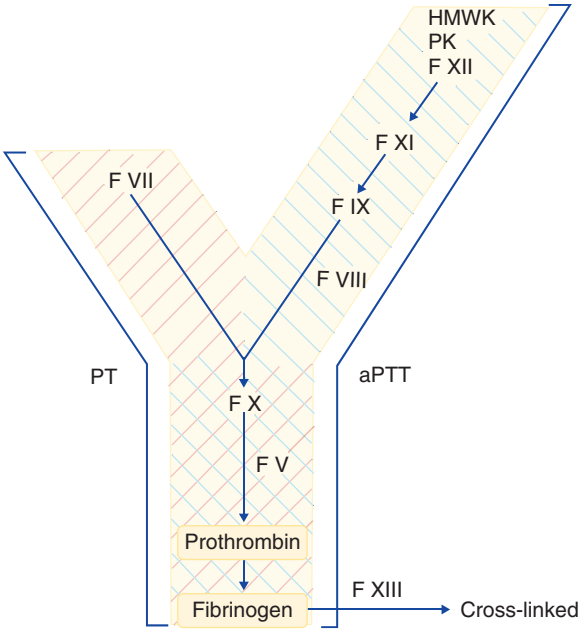

FIGURE 171-2. Classic coagulation cascade. The prothrombin time (PT) measures the integrity of the extrinsic and common pathways, whereas the activated partial thromboplastin time (aPTT) measures the integrity of the intrinsic and common pathways. Factor (F) XIII deficiency is not detected by PT or aPTT. HMWK = high-molecular-weight kininogen; PK = prekallikrein.

Abnormalities in these screening tests of hemostasis may be pursued by more specialized tests to establish a specific diagnosis (Fig. 171-3); other similar algorithms have been published.[6]

An abnormal VWD screen should be followed up with more specialized tests, including von Willebrand factor (VWF) multimer analysis, to identify the type of von Willebrand disease involved. Abnormalities in the PFA closure time should be followed by more specialized light transmission aggregometry (LTA) of platelets using a panel of agonists (ADP, epinephrine, collagen, arachidonic acid, ristocetin) that induce characteristic changes in light transmission (or optical density) in stirred suspensions of freshly isolated platelets in plasma (platelet-rich plasma, PRP). LTA is still considered the gold standard for platelet function testing[7] (see Fig. 171-3C and Chapter 173.)

The finding of a prolonged PT and/or aPTT indicates either a deficiency of one or more coagulation factors or the presence of an inhibitor (Fig. 171-4), usually an antibody, directed at one or more components of the coagulation system (see Fig. 171-3). These two possibilities can be distinguished by performing a simple inhibitor screen, which involves a 1 : 1 mix of the patient's plasma and normal plasma. The premise of the test is that even if the patient's plasma is completely deficient (0% level) in a certain factor, mixing it 1 : 1 with normal plasma (100% level) should bring the concentration of that factor to 50% in the mixture; this is sufficient to correct the prolonged PT or aPTT. If correction occurs with the inhibitor screen, individual coagulation factor levels should be assayed for a specific deficiency state. If the 1 : 1 mix fails to correct the prolonged PT and/or aPTT, an inhibitor is likely to be

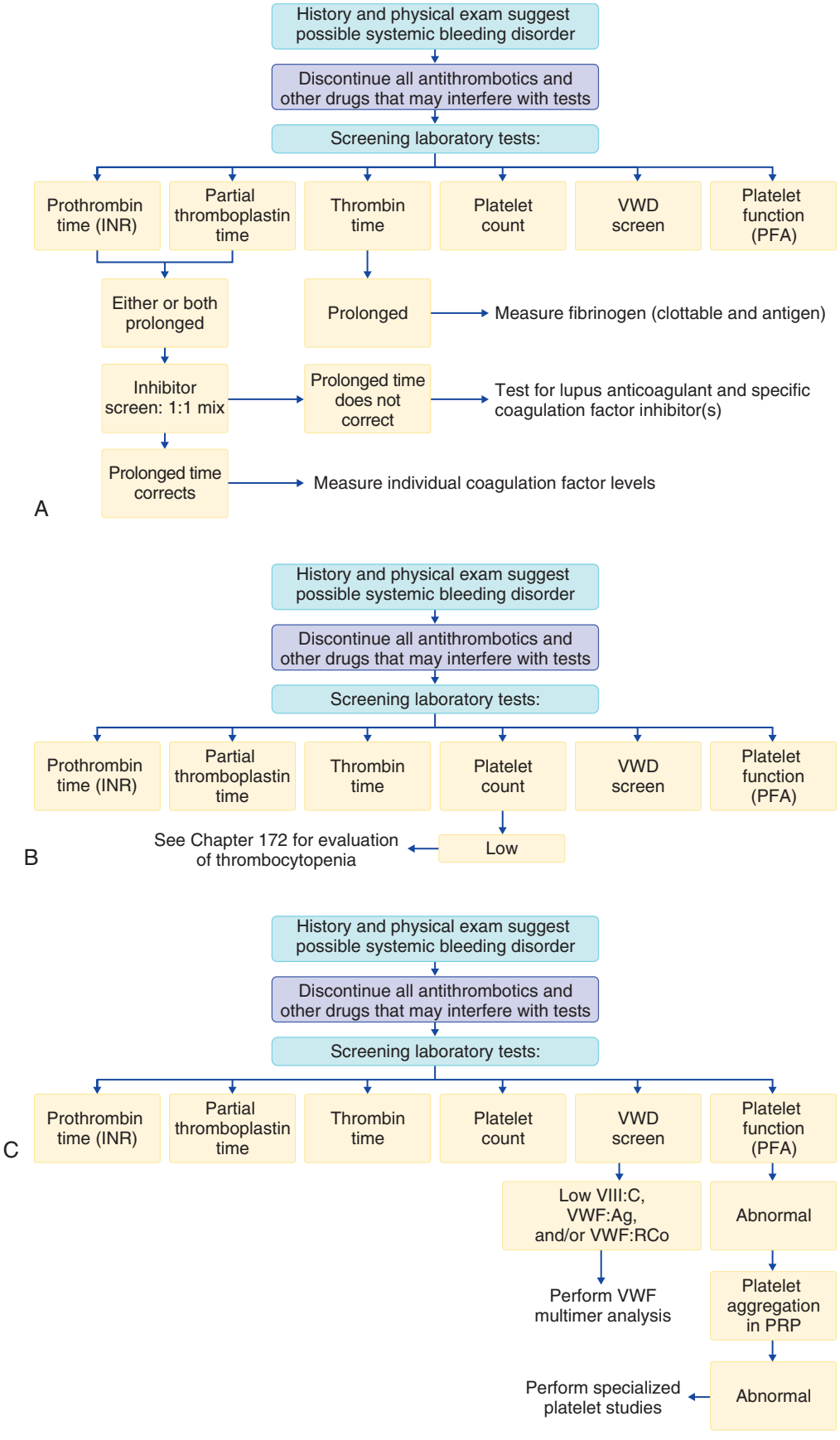

FIGURE 171-3, A to C, Algorithm for a clinical and laboratory approach to the diagnosis of a patient with a suspected systematic bleeding disorder (coagulopathy). The critical importance of a thorough personal and family history and physical examination is emphasized before initiating a laboratory work-up. INR = international normalized ratio; PFA = platelet function analyzer; VWF = von Willebrand factor; VIII:C = Factor VIII coagulant activity; VWF:Ag = von Willebrand factor antigen; VWF:RCo = von Willebrand factor: ristocetin cofactor activity; PRP = platelet-rich plasma.

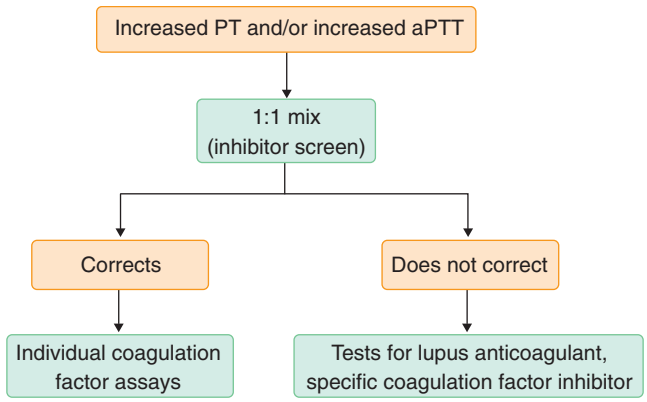

FIGURE 171-4. Approach to evaluating patients with prolonged prothrombin time (PT) or activated partial thromboplastin time (aPTT).

present, and it is interfering with coagulation in vitro in both the patient's plasma and the normal plasma. Specific assays should then be performed to determine whether there is a true inhibitor against a coagulation factor (e.g., factor VIII antibody) or whether the inhibitor is a lupus anticoagulant. A prolonged TT, with or without prolongation of the PT and/or aPTT, suggests the presence of a quantitative deficiency or a qualitative defect of fibrinogen. These can then be distinguished by following up with simultaneously performed clottable (functional) and antigenic assays of fibrinogen: a disproportionately low clottable fibrinogen points to a functional fibrinogen abnormality (dysfibrinogenemia), whereas proportionately reduced levels of clottable and antigenic fibrinogen suggest a quantitative deficiency (hypofibrinogenemia or afibrinogenemia) (see Fig. 171-3A and Chapter 174).

Newer coagulation tests, known as global assays, have the potential to offer fuller evaluation of a patient's overall clot-forming capability. These global assays, including thrombin generation tests and viscoelastic assays, have not yet advanced from experimental research to routine clinical practice.[8,9]

EVALUATION OF THE ASYMPTOMATIC PATIENT WITH ABNORMAL COAGULATION TESTS

In asymptomatic individuals who are discovered incidentally to have abnormalities in screening laboratory tests of hemostasis, the first critical question to ask is whether the findings are clinically relevant. Patients with inherited deficiencies of one of the contact activation coagulation factors (factor XII, high-molecular-weight kininogen, prekallikrein) characteristically have a markedly prolonged aPTT, yet they do not have a clinical bleeding tendency. Likewise, patients with lupus anticoagulants typically have prolongation of the aPTT and sometimes also the PT; they more often have thrombotic rather than bleeding complications. In patients with heparin-induced thrombocytopenia, a marked decrease in the platelet count is sometimes associated with arterial and venous thrombosis rather than bleeding. It is crucial to view the clinical setting, history, physical examination, and screening laboratory tests as complementary facets of the approach to patients with suspected coagulopathies.

EVALUATION OF THE PREOPERATIVE PATIENT

Routine screening of all preoperative patients with a platelet count, bleeding time, PT, and aPTT not only is uninformative but also may be counterproductive if follow-up testing causes unnecessary expense and delays in surgery (Chapter 431). Preoperative bleeding time, PT, and aPTT do not predict surgical bleeding risk in patients who are not found to be at increased risk on clinical grounds, so a thorough clinical assessment should guide the need for these preoperative screening tests. Laboratory testing and possibly further specialized tests of coagulation are indicated for patients whose bleeding histories are suspicious for a hemostatic abnormality. Preoperative screening tests of coagulation are probably warranted for patients who cannot cooperate with an adequate clinical assessment and for those who will be undergoing procedures in which even minimal postoperative hemorrhage could be hazardous. Recent guidelines for preinterventional and preoperative hemostatic evaluation have likewise concluded that bleeding risk requires a detailed personal and family history of hemorrhagic events and physical examination; but individuals with a negative history and without conditions that may interfere with systemic hemostasis should not undergo coagulation testing.[10]

TABLE 171-2 CLINICAL CHARACTERISTICS OF PATIENTS WITH INHERITED HYPERCOAGULABLE STATES (THROMBOPHILIA)

Venous thromboembolism (>90% of cases)
Deep vein thrombosis and/or pulmonary embolism most common
Mesenteric, cerebral vein thrombosis rare but characteristic
Frequent family history of thrombosis
Typically autosomal dominant
First episode of thrombosis typically in young adulthood (<40 years)
Often apparently unprovoked (but an acquired thrombosis trigger is frequently identified with a careful history)
More common thrombophilias (factor V Leiden, prothrombin 20210G→A) are associated with a lower thrombosis risk; rarer thrombophilias (antithrombin III, protein C, protein S deficiency) are associated with a higher thrombosis risk
Warfarin-induced skin necrosis or neonatal purpura fulminans with protein C or protein S deficiency (very rare)

EVALUATION OF THE PATIENT WITH A POSSIBLE HYPERCOAGULABLE STATE

Most patients with venous thromboembolism (VTE) have an inherited basis for hypercoagulability (Chapter 176). Patients with inherited hypercoagulable states (or thrombophilia) typically present with an initial episode of VTE in early adulthood, but thrombotic manifestations may begin at any time from early childhood to old age. Patients usually have deep vein thrombosis of the lower extremities or pulmonary embolism, but other uncommon sites of venous thrombosis may be involved. Arterial thrombosis is not characteristically associated with inherited hypercoagulable states. Arterial thrombosis that occurs prematurely or in the absence of apparent risk factors should trigger a different line of investigation, possibly including evaluation for vasculitis, myeloproliferative neoplasms, hyperhomocysteinemia, antiphospholipid syndrome, and potential sources of systemic embolization.

The primary or hereditary hypercoagulable states (see Table 176-1) result from specific mutations or polymorphisms that lead to decreased levels of physiologic antithrombotic proteins or increased levels of procoagulant proteins. In contrast, the secondary or acquired hypercoagulable states (see Fig. 176-3) are a heterogeneous group of disorders that predispose to thrombosis by complex mechanisms. VTE is often precipitated by the combination of an underlying hypercoagulable genotype and an acquired prothrombotic, hypercoagulable state such as pregnancy, immobilization, or the postoperative state. Certain clinical characteristics suggest the presence of an inherited hypercoagulable state (Table 171-2). Patients with recurrent thrombosis should be tested for these disorders and, in most cases, committed to lifelong prophylactic anticoagulation. It is not clear whether it is essential to order these tests after a single episode of VTE. Counterintuitively, it has been found that most of the common inherited hypercoagulable states—which, to varying extents, clearly increase the risk for a *first* episode of VTE—are at best only weak predictors of *recurrent* VTE. It has been argued, therefore, that making a specific thrombophilia diagnosis after an initial episode of VTE will not influence the decision about the duration of prophylactic anticoagulation. However, the increased risk for recurrent VTE with the more strongly prothrombotic mutations (e.g., antithrombin III, protein C or S deficiency), combined inherited thrombophilias, or antiphospholipid syndrome may necessitate long-term anticoagulation, and therefore their diagnosis after a first episode of VTE would in fact alter decisions about the duration of anticoagulation. An algorithm for thrombophilia testing after a first episode of VTE is presented in Figure 176-2. Even if they are not maintained on long-term anticoagulation, patients with diagnosed primary hypercoagulable states should receive prophylactic anticoagulation during situations that pose a high risk for thrombosis, such as the peripartum period. There is no simple screening test for primary hypercoagulable states, and the timing of obtaining these tests is crucial to avoid erroneous diagnoses. Acute thrombosis itself can cause transient decreases in the levels of antithrombin, protein C, and protein S. Heparin therapy can cause a decrease in plasma antithrombin activity. Warfarin therapy lowers the functional levels of protein C and protein S. Inherited deficiency states can be diagnosed spuriously under these conditions. Use of one of the newer anticoagulants can also interfere with laboratory tests, especially for lupus anticoagulant.

Patients who present with VTE have an increased risk for harboring an occult malignancy. This association is increased further in patients with recurrent and unprovoked thrombosis. There are conflicting opinions on whether an evaluation for occult malignancy in these patients must be exhaustive.

Most recommend that evaluation can be limited to a thorough history, physical examination, routine complete blood cell count and chemistries, test of fecal occult blood, urinalysis, mammogram (in women), and chest radiograph, with further testing guided by any abnormalities found in this initial evaluation. Others argue for routine computed tomography scanning of the chest, abdomen, and pelvis[11,12] (see algorithm in Fig. 176-3).

In addition to classic deep vein thrombosis and pulmonary embolism, certain characteristic types of thrombosis may provide important clues to the cause and trigger a more directed evaluation. Migratory, superficial thrombophlebitis (Trousseau syndrome) or nonbacterial thrombotic endocarditis strongly suggests the presence of an occult malignancy (Chapter 179). Hepatic vein thrombosis (Budd-Chiari syndrome; Chapter 143) or portal vein thrombosis may indicate a myeloproliferative neoplasm (Chapter 166) or paroxysmal nocturnal hemoglobinuria (Chapter 160). Extensive inferior vena cava thrombosis may occur with renal cell carcinoma (Chapter 197). Warfarin-induced skin necrosis strongly suggests underlying protein C or protein S deficiency. Recurrent, spontaneous miscarriages are characteristic of the antiphospholipid syndrome (Chapter 174), although they are also associated with other thrombophilias.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Versteeg HH, Heemskerk JW, Levi M, et al. New fundamentals in hemostasis. *Physiol Rev*. 2013;93:327-358.
2. Esmon CT, Esmon NL. The link between vascular features and thrombosis. *Ann Rev Physiol*. 2011;73:503-514.
3. Mauer AC, Khazanov NA, Levenkova N, et al. Impact of sex, age, race, ethnicity and aspirin use on bleeding symptoms in healthy adults. *J Thromb Haemost*. 2011;9:100-108.
4. O'Brien SH. Bleeding scores: are they really useful? *Hematol Am Soc Hematol Edu Program*. 2012;2012:152-156.
5. Lippi G, Favaloro EJ, Franchini M. Dangers in the practice of defensive medicine in hemostasis testing for investigation of bleeding or thrombosis: part I—routine coagulation testing. *Semin Thromb Hemost*. 2014;40:812-824.
6. Klein K, Hartman SK, Teruya J, et al. An algorithmic approach to coagulation testing. *Dis Mon*. 2012;58:431-439.
7. Harrison P, Mackie I, Mumford A, et al. Guidelines for the laboratory investigation of heritable disorders of platelet function. *Br J Haematol*. 2011;155:30-44.
8. Choi JL, Li S, Han JY. Platelet function tests: a review of progresses in clinical application. *Biomed Res Int*. 2014;2014:456569.
9. Young G, Sørensen B, Dargaud Y. Thrombin generation and whole blood viscoelastic assays in the management of hemophilia: current state of art and future perspectives. *Blood*. 2013;121:1944-1950.
10. Levy JH, Szlam F, Wolberg AS, et al. Clinical use of the activated partial thromboplastin time and prothrombin time for screening: a review of the literature and current guidelines for testing. *Clin Lab Med*. 2014;34:453-477.
11. Keeling D, Alikhan R. Management of venous thromboembolism: controversies and the future. *Br J Haematol*. 2013;161:755-763.
12. Lauw MN, van Doormaal FF, Middeldorp S, et al. Cancer and venous thrombosis: current comprehensions and future perspectives. *Semin Thromb Hemostas*. 2013;39:507-514.

REVIEW QUESTIONS

1. You are asked to clear a 48-year-old man for elective surgery for a herniated disc because he told his surgeon that he has a history of some kind of "blood problem." The patient tells you he bled heavily one day after a dental extraction 5 years ago, requiring blood transfusions in his local hospital's emergency department. He also recalls his parents telling him that he "almost died" of bleeding after a tonsillectomy in early childhood. There have been no other bleeding problems, even though he was an All-American soccer player in college and proudly says that he didn't bruise any more than his teammates. Family history is negative for bleeding except for his maternal grandfather, who had serious postoperative bleeding after coronary artery bypass graft surgery. The patient's screening coagulation test results now are normal: platelets = 380,000; prothrombin time (PT) = 11.4 (international normalized ration [INR] = 1.0); partial thromboplastin time (PTT) = 27.5. What would you do now?

A. Clear him for surgery with close postoperative observation for bleeding.
B. Order a bleeding time and platelet function tests.
C. Order factor VIII and factor IX levels.
D. Order liver function tests.
E. Order von Willebrand factor (VWF) multimer analysis.

Answer: C The type of bleeding complications this patient describes (i.e., delayed, deep bleeding, especially at oropharyngeal sites with surgery) is more characteristic of a coagulation factor defect than a platelet-vascular defect. His family history is suggestive of an X-linked inherited coagulopathy, with his mother being the silent carrier. Factor VIII deficiency (classic hemophilia) and factor IX deficiency fit that description. The PTT is not a sensitive screening test to pick up less severe forms of a coagulopathy (i.e., when the factor level is >25%), so *mild* hemophiliacs (with either factor VIII or IX deficiency) may have a normal PTT and yet still have serious bleeding complications when provoked with surgery. Therefore, factor VIII and IX levels must be obtained here to rule out the mild hemophilia, which would require prophylactic preoperative (as well intraoperative and postoperative) replacement therapy. The bleeding time, a method to test the integrity of platelet-vascular interactions, is now largely an obsolete test; and the kinds of bleeding problems this patient describes (as well as the notable absence of pathologic bruising with trauma) argues strongly against a qualitative platelet abnormality. Liver function tests could be done but are unlikely to shed additional light on this diagnostic problem; furthermore, this patient appears to have an inherited, not acquired coagulopathy. The possibility of von Willebrand disease is not unreasonable (even though it is usually autosomal dominant), although, again patients with von Willebrand disease tend to bleed superficially like patients with platelet problems. In any case, multimer analysis is not the first step to diagnosing von Willebrand disease: it should be preceded by a von Willebrand panel to assay levels of functional VWF, antigenic VWF, and factor VIII activity.

2. A 68-year-old man with coronary artery disease, diabetes mellitus, hypertension, and a history of stroke followed by seizures is now 3 days after an uncomplicated coronary artery bypass graft (CABG) surgery, and you are asked to consult for severe thrombocytopenia. The patient is on multiple medications for his chronic conditions and is now on postoperative subcutaneous heparin for thromboprophylaxis. There have been no bleeding problems postoperatively, and physical examination shows no signs of bleeding from external orifices and no ecchymoses or petechiae. The platelet count today is 5000, which triggered the STAT consult; it had been previously normal, including a count of 280,000 yesterday. Which of the following should be the first step?

A. Examine the peripheral blood smear for platelet clumping.
B. Stop heparin.
C. Stop heparin and all other medications.
D. Order STAT platelet transfusions.
E. Take a thorough history and do a comprehensive physical examination.

Answer: A The first step is to promptly determine whether the apparently sudden drop in platelet count is real or artifactual. The abrupt drop (when counts were normal as recently as 1 day ago), accompanied by the striking absence of symptoms or signs of bleeding in this patient raises a high level of suspicion for "pseudothrombocytopenia." Seeing large clumps of platelets on the peripheral smear, which should be examined by the hematologist, would confirm the suspicion, which should be then promptly documented by ordering repeat platelet counts in both the customary purple-top tube containing EDTA as the anticoagulant along with tubes containing other anticoagulants (heparin or citrate). (The platelet count in the EDTA tube should be again very low, but not in blood collected into test tubes with the other anticoagulants.) Occasionally, pseudothrombocytopenia occurs as a result of the presence of a nonphysiologic "cold agglutinin," which clumps platelets in the test tube while sitting at room temperature before counting. This can be ruled out by having the laboratory do simultaneous platelet counts on a tube kept at 37° C versus one that has been intentionally left standing at room temperature: the former should have a normal platelet count if this is the case. These findings represent pseudothrombocytopenias in which the platelet counts are artifactually low and of no clinical consequence. These tests can be all done very quickly without stopping important medications, including heparin, and before undertaking further work-up and treatment of the thrombocytopenia if they remain necessary.

3. Which of the following statements is correct regarding the relationship between cancer and thrombosis?

A. In a patient being treated for gastric cancer, secondary prevention following a first episode of venous thromboembolism is best provided by warfarin for 3 to 6 months at doses to target an international normalized ratio of 2.0 to 3.0.
B. A first episode of apparently unprovoked deep vein thrombosis in a previously healthy individual requires comprehensive search for an occult malignancy, including computed tomography of the chest, abdomen, and pelvis.
C. Pulmonary embolism in a patient being actively treated for cancer requires the placement of an inferior vena cava (IVC) filter to prevent recurrence.
D. A patient with breast cancer who completed treatment 6 months ago and is now in complete remission, with no evidence of disease, is not at increased risk for venous thromboembolism.
E. In a previously healthy individual, the spontaneous development of extensive IVC thrombosis is highly suggestive of an occult renal cell carcinoma.

Answer: E IVC thrombosis, a rare and dramatic form of venous thrombosis that can extend into the right atrium, is associated with renal cell carcinoma. Other characteristic (but not diagnostic) cancer-associated thrombosis syndromes include Budd-Chiari syndrome (hepatic vein thrombosis) or portal vein thrombosis in patients with myeloproliferative neoplasms, and superficial, migratory thrombophlebitis (Trousseau syndrome) with occult cancers of the gastrointestinal tract and pancreatic cancer. Subcutaneous low-molecular-weight heparin is superior to dose-adjusted warfarin for secondary prophylaxis after an episode of venous thromboembolism in a cancer patient, even when in remission, and it should be continued indefinitely. Although the incidence of occult malignancy in a previously healthy person who presents with unprovoked venous thromboembolism is definitely increased, most experts agree that a comprehensive history and physical examination, stool for occult blood, chest radiograph, blood counts and basic chemistries, and a mammogram in women is a sufficient initial screen for occult cancer, as long as any abnormalities picked up on that screen are more aggressively pursued. Increased risk for thrombosis in a patient with cancer continues for months or even years after complete remission has been achieved. An IVC filter is inadequate to prevent recurrent pulmonary embolism in patients with or without cancer.

172 THROMBOCYTOPENIA

CHARLES S. ABRAMS

Thrombocytopenia is defined as a platelet count of less than the normal range, typically below 140,000/μL. In the absence of qualitative platelet defects (Chapter 173), excessive bleeding does not occur in thrombocytopenic patients following trauma or surgery unless the platelet count is lower than 75,000/μL. In otherwise hemostatically normal patients, spontaneous hemorrhage typically does not occur with platelet counts above 30,000/μL. Patients with platelet counts of less than 5000/μL to 10,000/μL are at a high risk for spontaneous, life-threatening hemorrhage. However, there is no absolute threshold for spontaneous bleeding due to thrombocytopenia. Bleeding may occur at higher counts when fever, sepsis, severe anemia, and other hemostatic defects are present, or when platelet function is impaired by medication. Notably, a prolonged cutaneous bleeding time test (Chapter 171) does not accurately predict the risk for clinical bleeding. Therefore, it is critical that the physician consider the gamut of possibilities when diagnosing and managing a patient with a low platelet count.

Typically, the first step in diagnosing a patient with a low platelet count is to determine whether the thrombocytopenia is attributable to one of the three broad mechanisms for developing thrombocytopenia: (1) increased destruction of platelets, such as seen in immune-mediated causes; (2) decreased production of platelets, usually due to an underlying bone marrow disorder; or (3) sequestration of platelets within the spleen, as occurs in conditions that cause splenomegaly (hypersplenism).

Determining the cause of the patient's thrombocytopenia is essential to select the most appropriate therapy and to avoid unnecessary procedures. Because there is no easy test to differentiate among the three possibilities, clinical evaluation is critical. Therefore, a thorough history and physical examination, with attention to possible alternative explanations for thrombocytopenia, are mandatory. Particular attention should be paid to the duration of symptoms, which helps to determine whether the patient has acute or chronic thrombocytopenia. The clinician should also focus on the patient's recent exposure to new medications that might induce thrombocytopenia as a side effect.

Preliminary laboratory tests include a complete blood count (CBC) with a differential, along with an examination of the peripheral blood smear. Abnormalities in the number or morphology of the leukocytes or erythrocytes may indicate a systemic inflammatory disorder or a pathology within the bone marrow. Clumping of the platelets on the peripheral blood smear may indicate that the patient has pseudothrombocytopenia (Chapter 171). This anomaly is usually due to an antibody that binds to and thereby agglutinates platelets only in the presence of ethylenediaminetetraacetic acid (EDTA) or when the blood sample is allowed to cool to room temperature. Repeating the platelet count in a blood sample that has been anticoagulated with heparin and kept warm until it is analyzed can help exclude this spurious diagnosis. Any additional testing to determine the cause of thrombocytopenia is based solely on the available clinical information derived from the history and examination of the patient. For example, testing for human immunodeficiency virus (HIV) would be prudent in a patient with risk factors. A bone marrow aspirate and biopsy are typically not required for a patient with thrombocytopenia who has an otherwise normal CBC and peripheral blood smear. However, a bone marrow aspirate and biopsy should be considered in older patients and in patients for whom standard therapy has not been effective.

With the widespread use of the "routine CBC," asymptomatic individuals with incidentally discovered mild thrombocytopenia are being increasingly recognized. The clinical significance of mild thrombocytopenia (platelet counts between 100,000/μL and 150,000/μL) can be difficult to ascertain. Some of these individuals represent outliers of the normal distribution of platelet counts. However, others have thrombocytopenia that represents the early manifestation of an unrecognized disease. In the absence of other signs, many of these individuals are diagnosed with idiopathic (immune) thrombocytopenic purpura (ITP)—a disease that will be discussed later in this chapter. Studies of apparently healthy individuals who were incidentally discovered to have borderline thrombocytopenia have yielded some insight into their prognosis.[1] After a decade, approximately 10% will have developed ITP or another autoimmune disorder. However, approximately 90% of the subjects continued to manifest borderline thrombocytopenia without developing another disorder during this period of time. It can be concluded that most individuals with isolated and asymptomatic chronic mild thrombocytopenia simply represent the lower end of the normal platelet count distribution. Consequently, international consensus groups have set the threshold for the diagnosis of ITP at a platelet count of less than100,000/μL.

PATHOBIOLOGY

See Figure 172-1 for an approach to evaluating patients with chronic thrombocytopenia.

Increased Platelet Destruction

In the absence of splenomegaly, the presence of megakaryocytes in an otherwise normal bone marrow implies that thrombocytopenia is due to increased platelet destruction. An acute drop in a patient's platelet count also implies that a peripheral destructive process is the likely cause. For example, a patient who develops abrupt thrombocytopenia during hospitalization most likely has a low platelet count caused by an infection or by the introduction of a new medication. Both nonimmune and immune processes can lead to a shortened platelet lifespan. Nonimmune reasons for the accelerated destruction of platelets include sepsis, disseminated intravascular coagulation (DIC), thrombotic thrombocytopenic purpura (TTP), hemolytic-uremic syndrome (HUS), preeclampsia or eclampsia, cardiopulmonary bypass, and giant cavernous hemangioma. Thrombocytopenia that occurs in these circumstances usually resolves with treatment of the underlying disorder, and platelet transfusions are rarely necessary. Thrombocytopenia due to TTP, HUS, or heparin-induced thrombocytopenia (HIT) is more characteristically associated with microvascular occlusion or thrombosis than with bleeding (Table 172-1), so platelet transfusions are rarely necessary. In addition, reports have noted the clinical deterioration of patients with TTP or HIT following platelet transfusion, leading to the controversial suggestion that platelets should not be given to most patients with these disorders.

Immune-mediated platelet destruction can result from drugs, alloimmune sensitization, or autoimmunity. Medications should always be considered as a possible cause of acute thrombocytopenia. The list of potential offending agents is long, but drugs with strong evidence of antibody-mediated platelet destruction include quinine, quinidine, sulfonamides, and gold salts. In addition to stopping the offending medication, platelet transfusions might be required for severe thrombocytopenia.

Platelet Sequestration

Approximately 30% of the circulating platelet mass is normally sequestered within the spleen. Enlargement of the spleen due to portal hypertension or

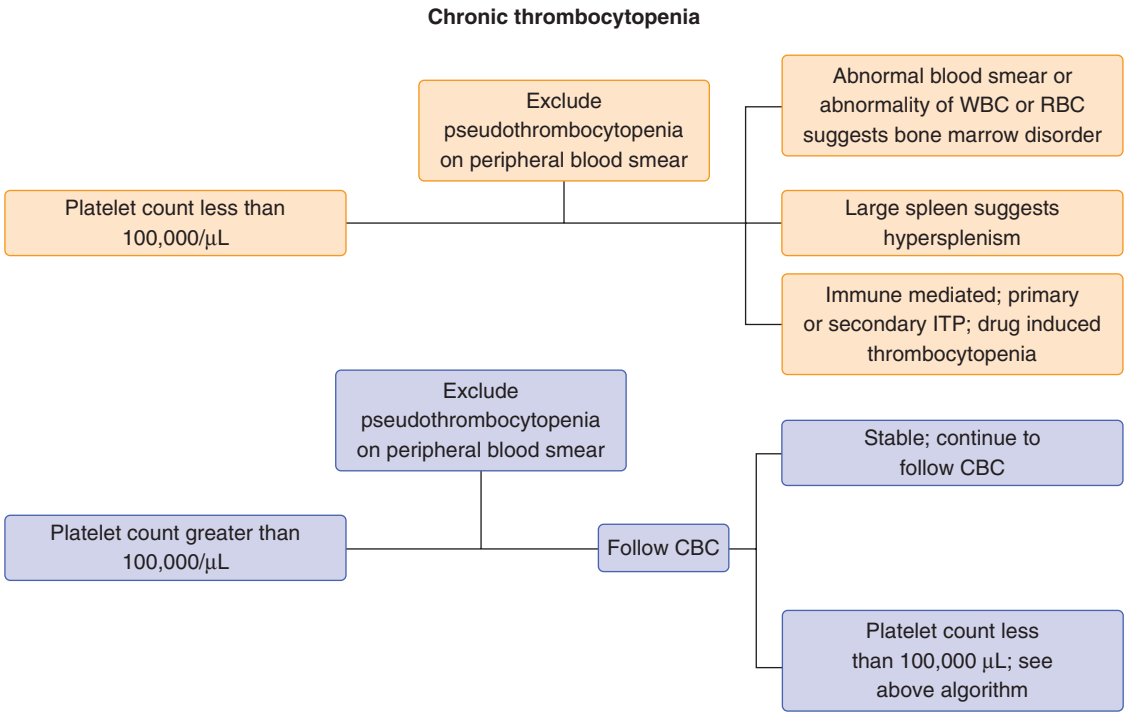

FIGURE 172-1. **Systematic approach to the evaluation of chronic thrombocytopenia.**

TABLE 172-1 DISORDERS THAT CAUSE BOTH THROMBOCYTOPENIA AND THROMBOSIS

Thrombotic thrombocytopenic purpura (TTP)
Hemolytic-uremic syndrome (HUS)
Heparin-induced thrombocytopenia (HIT)
Disseminated intravascular coagulation (DIC)
Paroxysmal nocturnal hemoglobinuria (PNH)
Vasculitis (such as systemic lupus erythematosus)
Antiphospholipid antibody syndrome (APS)

infiltration is accompanied by expansion of the splenic platelet pool. Because hypersplenism by itself rarely causes a platelet count of less than 40,000/μL to 50,000/μL, bleeding due to thrombocytopenia from hypersplenism alone is unusual. This in turn can result in moderate thrombocytopenia.

Decreased Platelet Production

Decreased platelet production occurs in primary diseases of the bone marrow, such as acute leukemia and aplastic anemia; myelophthisic processes in which bone marrow is affected by metastatic carcinoma, fibrosis, or other clonal hematopoietic disorders; following chemotherapy and/or radiation therapy; with ethanol toxicity; and during infections with viruses such as HIV, cytomegalovirus (CMV), Epstein-Barr virus (EBV), and varicella. Thrombocytopenia also occurs when normal megakaryocyte proliferation is impaired by myelodysplasia.

Thrombocytopenia due to decreased platelet production is often accompanied by abnormalities of both the white blood cells and red blood cells. This should be readily apparent by inspection of the CBC and review of the peripheral blood smear. When thrombocytopenia in this setting is accompanied by bleeding, the bleeding usually requires treatment by platelet transfusion. Additionally, when decreased platelet production cannot be readily reversed, prophylactic platelet transfusion to prevent bleeding is often considered. However, as will be discussed in the following section, prophylactic platelet transfusion is problematic because of the short lifespan of platelets and the potential to develop alloantibodies that will limit the effectiveness of future platelet transfusions.

PLATELET TRANSFUSIONS

Clinicians usually consider "wet" bleeding to be much more ominous than "dry" bleeding. Signs of wet bleeding include epistaxis, gingival bleeding, gastrointestinal bleeding, genitourinary bleeding, and bleeding around

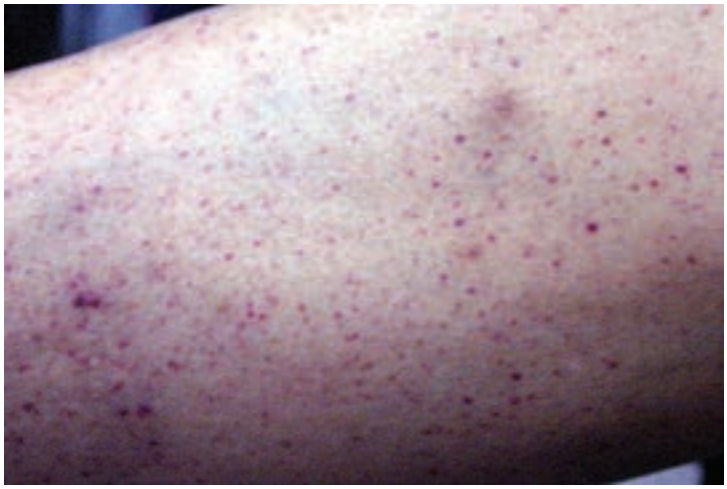

FIGURE 172-2. **Petechiae. Multiple pinpoint, nonblanching, erythematous macules found predominantly on skin of dependent extremities in patients with severe thrombocytopenia.**

intravenous sites. Dry bleeding is defined as ecchymosis or petechiae (Fig. 172-2). Overt wet bleeding that is clearly due to thrombocytopenia is usually treated with platelet transfusion (Chapter 177). Prophylactic platelet transfusion for patients who are not bleeding is controversial. When making the decision whether to treat a nonbleeding patient with thrombocytopenia, the practitioner must consider the short lifespan of platelets (10 days), the 5-day shelf-life of stored platelets, and the potential for the patient to develop transfusion-induced platelet alloantibodies. In patients undergoing treatment for acute leukemia, recent clinical trials have supported the clinical practice of prophylactic platelet transfusions when the platelet count is below 5000/μL or 10,000/μL, rather than waiting until bleeding occurs in the severely thrombocytopenic patient. The indications for prophylactic platelet transfusions before interventional procedures is predominantly based on clinical experience and expert opinions, rather than on clinical trials. These recommendations are summarized in Table 172-2. Platelet transfusions should continue for several days following procedures where the risk for bleeding or the complications resulting from bleeding are high. For extremely high-risk procedures that involve the nervous and ocular systems, the prophylactic platelet transfusions should be continued for at least 7 to 10 days.

TABLE 172-2 SUGGESTED MINIMUM PLATELET COUNTS BEFORE INVASIVE PROCEDURES

VERY-HIGH-RISK PROCEDURES—75,000/μL TO 100,000/μL
Neurosurgery Ocular surgery (except cataract extraction) Thyroid surgery Prostatectomy
MODERATE-RISK PROCEDURES—50,000/μL
Liver Biopsy Dental extraction Most surgical procedures
LOW-RISK PROCEDURES—30,000/μL
Endoscopy Bronchoscopy Lumbar puncture (with scrupulous technique)
VERY-LOW-RISK PROCEDURES—NO PLATELET TRANSFUSIONS NECESSARY
Bone marrow biopsy Cataract extraction

A donated unit of blood contains approximately 50 billion platelets. Infusing this number of platelets into a patient should increase his or her platelet count by 20,000/μL divided by the patient's body surface area in square meters. Therefore, transfusion into an average-sized patient should increase the recipient's platelet count by 10,000/μL to 12,000/μL. Bags of platelets used for transfusions are typically obtained by pooling platelets derived from four to eight blood donors. Therefore, a "unit" at one hospital may be derived from more donors than a "unit" obtained at another hospital.

Large quantities of platelets can be derived from a single donor by using apheresis technology (plateletpheresis) (Chapter 177). Because there is minimal loss of red blood cells with this technique, plateletpheresis allows one individual to donate very large numbers of platelets (equivalent to the number of platelets obtained from 6 to 10 units of donated blood). As discussed in the following section, the risk for platelet alloimmunization is partially dependent on the number of individuals donating platelets to a patient. Therefore, single-donor plateletpheresis can prevent or at least minimize alloimmunization.

Several trials have addressed how many platelets should be transfused into thrombocytopenic patients to maintain hemostasis and to minimize exposure to blood products (also see Chapter 177). The largest of these trials was the multicenter Platelet Dose (PLADO) Trial.■ In this study, patients with thrombocytopenia were randomized to receive transfusions of low-dose (110 billion/m^2), medium-dose (220 billion/m^2), or high-dose (440 billion/m^2) platelets while recovering from hematopoietic stem cell transplantation. Patients who received the low-dose therapy required more frequent platelet transfusions. However, overall they received fewer transfused platelets over the course of the study. Importantly, bleeding complications were identical with all platelet transfusion strategies. This study demonstrates that low doses of platelets (110 billion/m^2 of body surface area) are as safe as larger doses in patients with hypoproliferative thrombocytopenia who are receiving prophylactic platelet transfusion therapy.

Several complications of platelet transfusion therapy merit mention (also see Chapter 177). First, some patients can become alloimmunized against platelet antigens and can become refractory to future platelet transfusion therapy (discussed in detail later). Second, bacterial contamination of stored platelets is a much more common complication than the infectious risk associated with red blood cell transfusions. Unlike red blood cells, which are stored frozen after being harvested from donated blood, platelets are sensitive to cool temperatures and therefore need to be stored at room temperature. This method of storage results in more bacterial overgrowth in the transfusion bags. Transfusion-associated bacterial sepsis is one of the most frequently reported causes of transfusion-induced mortality in the United States, and platelet transfusions are the most common blood products associated with sepsis. It should also be noted that room temperature storage of platelets contributes to their short shelf life of approximately 5 to 7 days. Consequently, unlike the blood-derived products that can be frozen (such as red blood cells and plasma), there is a constant need for platelet donations year-round. In contrast to the risk for bacterial infection, the risk for viral infection from platelet transfusions is no higher than it is for red blood cell transfusions. From a single unit of platelets, the risk for being exposed to HIV is less than 1 in 2.5 million, the risk for exposure to hepatitis B virus is less than 1 in 750,000, and the risk for exposure to hepatitis C virus is less than 1 in 1 million.

Platelet transfusions have been considered contraindicated for thrombocytopenia caused by TTP or HIT. However, this recommendation is driven entirely by case reports of thrombotic events that have occurred soon after platelet transfusions have been given to such patients. This recommendation is not based on randomized trials designed to address this issue. Because thrombotic events are a known complication of both TTP and HIT, it is difficult to determine whether the thrombi that occurred in these patients were due to the platelet transfusions or merely to the intrinsic prothrombotic risk for TTP and HIT. In any case, there is usually no indication for platelet transfusions in patients with TTP and HIT because thrombosis is a much greater risk than hemorrhage in these disorders.

The Transfusion-Refractory Patient

Many patients do not have an optimal increase in their platelet count following transfusion. Although there are various definitions of platelet refractoriness, it is simplest to base this definition on the timing of platelet loss following transfusion. Patients who do not have the predicted rise in their platelet count (between 10,000/μL and 12,000/μL for every unit of donated platelets) should have their platelet count analyzed before their next platelet transfusion, 1 hour after that transfusion, and again 24 hours later. If their platelet count rises appropriately 1 hour after the transfusion but falls substantially 24 hours later, the patient has ongoing platelet consumption. This is often seen in patients with sepsis, DIC, or severe active hemorrhage, or as a result of the drug-mediated immune destruction of platelets. In these situations, the best therapy is to treat the underlying cause and to continue to support the patient with platelet transfusions as clinically indicated. Alternatively, some patients fail to have a significant increase in platelets even 1 hour after a transfusion. These patients may have (1) hypersplenism, (2) an autoantibody that eliminates not only endogenous platelets but also allogeneic platelets (as seen in patients with ITP), or (3) alloantibodies that react with antigens on the transfused platelets.

Alloantibodies develop in approximately 20% of patients who are repeatedly exposed to platelet transfusions, and these patients present some of the most challenging management issues in transfusion medicine (Chapter 177). The alloantibodies may be directed against human leukocyte antigen (HLA) class I antigens (HLA-A and HLA-B) or against platelet-specific antigens present on the surface of the platelets. Because it appears that HLA class II antigens present on the surface of leukocytes are essential for the development of antibodies directed against HLA class I antigens, efforts to remove the contamination of leukocytes in platelet transfusion preparations can minimize alloantibody formation. Therefore, filters that trap leukocytes during transfusions are useful to prevent patients from becoming refractory to future platelet transfusions. After a patient develops alloantibodies against transfused platelets, the clinician and the blood bank should attempt to identify whether the alloantibody is directed against an HLA class I antigen or against a platelet-specific antigen. If the alloantibody can be identified, platelets derived from donors matched for an HLA class I antigen or the platelet-specific antigen can be used. Corticosteroids, intravenous immunoglobulin, splenectomy, and recombinant factor VIIa are probably of no value in maintaining hemostasis in platelet-refractory patients. Antifibrinolytic agents (e.g., ε-aminocaproic acid) might be helpful for patients with bleeding that is predominantly in the oral cavity or in the genitourinary tract. However, antifibrinolytic agents are contraindicated in patients with DIC (Chapter 175) because these agents can precipitate thrombotic events.

SPECIFIC CAUSES OF THROMBOCYTOPENIA

Drug-Induced Thrombocytopenia

PATHOBIOLOGY

Drug-induced thrombocytopenia is one of the most frequent causes of cytopenias evaluated by physicians. The typical drug-induced thrombocytopenia is the result of an immune reaction elicited by either the drug or by one of its metabolites. Sometimes the drug is simply deposited on the platelet surface, and antidrug antibodies lead to accelerated platelet clearance, primarily in the spleen. More frequently, the drug binds to a protein on the platelet surface and induces a neoantigen that is ultimately recognized by the immune system.

In the absence of the drug or drug metabolites, this platelet neoantigen disappears, and the thrombocytopenia slowly resolves as new platelets are released from the bone marrow. This mechanism explains most drug-induced thrombocytopenias, but there are some exceptions. For example, chemotherapy and other bone marrow toxins can decrease platelet production and thereby induce thrombocytopenia by a mechanism independent of the immune system.

CLINICAL MANIFESTATIONS

Patients with drug-induced thrombocytopenia can have mild, moderate, or even severe forms. Almost all types of drug-induced thrombocytopenia predispose the patient to hemorrhagic but not thrombotic complications. The notable exception to this rule is heparin-induced thrombocytopenia (discussed separately in the next section). In acute or subacute cases of thrombocytopenia, it may be especially difficult to distinguish between drug-induced thrombocytopenia and the thrombocytopenia of an infection (viral or bacterial). In very ill patients, both possibilities must be considered in the management plan (see later section on Sepsis). Another difficult diagnostic dilemma is to discern drug-induced thrombocytopenia from ITP. For this distinction, the patient's medical history may be helpful. A gradually progressive thrombocytopenia is more consistent with ITP than with one induced by a medication. Conversely, the rapid onset of thrombocytopenia following the initiation of a new medication indicates that the drug is the most likely culprit. The time between the initiation of the offending drug and the development of thrombocytopenia has not been well documented for most medications. However, thrombocytopenia typically begins days to a few weeks after the administration of most medications with this type of toxicity. Presumably, this lag time represents the period required for the patient to develop an immune response against the drug-platelet complex. Medications that have been taken safely for years before the onset of thrombocytopenia are unlikely to be the offending agents. In addition to an assessment of the patient's prescribed medications, a careful and comprehensive evaluation of the patient's over-the-counter medications is necessary. Agents that contain quinine lead the list of often nonprescribed drugs that can induce life-threatening thrombocytopenia. Quinine is frequently contained in over-the-counter pills for leg cramps. Even the quinine in tonic water can induce severe thrombocytopenia ("gin and tonic purpura") when ingested by certain patients.

DIAGNOSIS

Some medications are much more likely to induce thrombocytopenia than others. Table 172-3 contains a partial list of drugs that frequently induce this type of toxicity. One particularly useful online database (www.ouhsc.edu/platelets/ditp.html) lists and periodically updates the level of evidence for specific drugs that may induce thrombocytopenia. With the exception of tests for HIT, laboratory tests for drug-induced thrombocytopenia do not have widespread applicability. Usually, drug-induced thrombocytopenia is proved only in retrospect, when the platelet count improves after the discontinuation of the suspected medication. Even then, the diagnosis is not conclusive unless thrombocytopenia recurs after rechallenging the patient with the offending drug (a practice that is almost never recommended).

TREATMENT

Rx

The most efficacious method of treating drug-induced thrombocytopenia is to stop all suspected offending medications. Usually, no additional therapy is indicated. The thrombocytopenia typically begins to resolve without further intervention within days to a week of stopping the drug. The notable exceptions involve drugs with particularly long half-lives. Sulfonamides, quinine, and quinidine are particularly notorious for inducing severe or even life-threatening thrombocytopenia. In patients with profound thrombocytopenia (<10,000/μL to 15,000/μL) or at a high risk for bleeding, platelet transfusions are indicated. If ITP cannot be confidently excluded, and the thrombocytopenia is life threatening, specific treatment for ITP can also be initiated (see later in this chapter).

Heparin-Induced Thrombocytopenia

EPIDEMIOLOGY

A special case of drug-induced immune-mediated thrombocytopenia associated with arterial and venous thrombosis, rather than bleeding, is HIT.[2] It is seen in 2 to 5% of patients exposed to unfractionated heparin, and in 0.7% of patients given low-molecular-weight heparin. HIT almost never occurs in patients who are exposed only to fondaparinux.

TABLE 172-3 DRUGS THAT ARE STRONGLY ASSOCIATED WITH THROMBOCYTOPENIA

ANTIBIOTICS AND ANTIVIRALS
Quinine, quinidine
Penicillin
Cephalosporin
Vancomycin
Trimethoprim-sulfamethoxazole
Sulfonamides, sulfonylureas
Linezolid
Valaciclovir
Ganciclovir
Indinavir
CARDIOVASCULAR MEDICATIONS
Abciximab
Tirofiban
Eptifibatide
Salicylates
Digoxin
Furosemide
MISCELLANEOUS
Cimetidine
Ranitidine
Famotidine
Valproate
Interferon

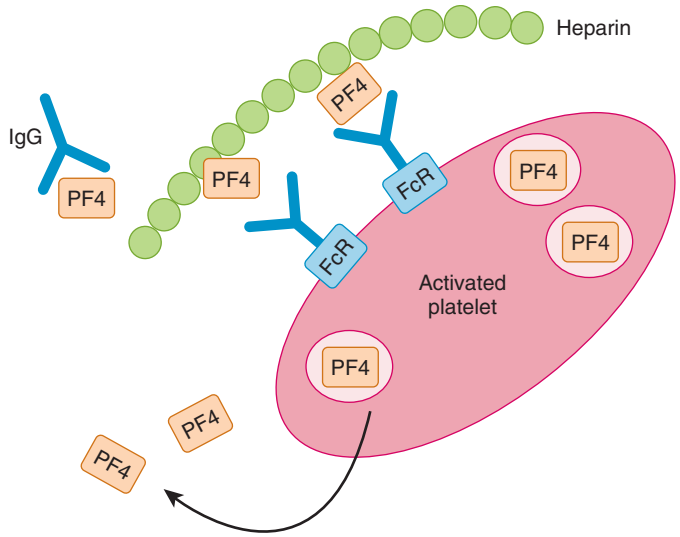

FIGURE 172-3. Pathophysiology of heparin-induced thrombocytopenia. Platelet factor 4 (PF4) is a protein that is stored in platelet granules and that is secreted from platelets upon their activation. Heparin can bind to PF4 and induce a conformational change within PF4 that in some patients produces an antibody-mediated immune response. Large complexes of immunoglobulin (Ig), heparin, and PF4 can accumulate on the platelet surface and stimulate the platelet when the Fc portion of the antibody interacts with the platelet's FcγRII receptor (FcR). Once activated, the platelets contribute to thrombosis. Additionally, the activated platelets secrete more PF4, which continues the process.

PATHOBIOLOGY

To understand this disease, one must know that platelets can secrete a protein called platelet factor 4 (PF4) that can bind to heparin. HIT is caused by an antibody that binds to this PF4-heparin complex (Fig. 172-3). Large complexes of antibodies directed against heparin-bound PF4 accumulate on the surface of platelets. At times, these anti-PF4 immunoglobulins bind to the Fc receptor that is also present on the platelet surface, leading to activation of the platelet, release of more PF4, and a cycle of events that causes the stimulation of even more platelets. Ultimately, this also leads to the activation of the

coagulation cascade. The activation of platelets and the clotting cascade leads to the formation of both thrombi and thrombocytopenia.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

When thrombocytopenia is detected in a hospitalized patient, HIT must always be considered. Patients with this syndrome often have multiple potential causes of thrombocytopenia (such as other medications or infections). Therefore, it is important to exclude these other causes. Large retrospective studies of HIT have suggested that the onset of thrombocytopenia relative to the initiation of heparin therapy is useful in establishing or excluding this diagnosis. HIT typically occurs 4 to 14 days after patients are given heparin by any route (it even occurs after subcutaneous injections of "minidose" heparin or following extremely low doses by heparin flush). This lag time between heparin exposure and the appearance of HIT is due to the time it takes for the immune response to generate the requisite antibodies against the heparin-PF4 complex. Some patients who have been exposed to heparin within the past several months may already have preexisting antibodies against the heparin-PF4 complex in their circulation. Therefore, when these patients are re-exposed to heparin, they may develop an acute onset of HIT even within the first day of the reinitiation of heparin.

The unique timing of acute-onset HIT within hours of heparin re-exposure and the more typical development of HIT within 4 to 14 days of initial heparin exposure emphasize the importance of conducting a careful and critical history in establishing the diagnosis. HIT should be suspected in any patient who develops thrombocytopenia while on heparin therapy. It is important to note that the normal platelet count varies widely (140,000/μL to 450,000/μL), and some patients may have a substantial decrease in their platelet count but still remain within the normal range. Therefore, a greater than 50% decrease in the platelet count of a patient on heparin should raise the suspicion of this syndrome. HIT should also be suspected in any patient who develops a thrombotic event while on heparin therapy.

Although antibodies against the heparin-PF4 complex are almost always present in patients with HIT, these antibodies are also frequently present in heparin-treated patients who do not have this disorder (i.e., those patients with neither thrombocytopenia nor thrombosis). There are two general categories of laboratory assays for the diagnosis of HIT: functional assays and immunologic assays (Table 172-4). Functional assays analyze whether the combination of heparin and the patient's plasma can induce normal platelets to aggregate or to secrete serotonin; these assays have very high specificity but relatively low sensitivity. Immunologic assays test the patient's plasma for antibodies that bind to the heparin-PF4 complex; these assays have very high sensitivity but lack specificity. Consequently, a *negative immunologic assay* is useful in excluding this diagnosis, and a *positive functional assay* is useful in confirming the diagnosis of HIT. Several days may elapse before the test results are available. Therefore, in practice, these laboratory assays provide only confirmatory information. Urgent clinical decisions should *not* be deferred until such test results are available. The timing of the thrombocytopenia with respect to the heparin exposure and the degree of the platelet drop are the most important pieces of information needed by the physician when evaluating a patient suspected of having HIT.

Because platelets are consumed as they become activated, the clinical presentation of HIT is thrombocytopenia. However, it is unusual for the thrombocytopenia to actually cause bleeding in patients with HIT. Instead, HIT is a highly prothrombotic disorder. Both venous and arterial thromboses are common in HIT. Patients with HIT who do not have thrombosis on initial presentation still have a 20 to 30% chance of developing a thrombus within the next month. Nevertheless, the incidence of HIT-related clinical events is greatest immediately after diagnosis. More importantly, clinical studies report that the cessation of heparin alone frequently fails to prevent the development of new thrombotic events.

TABLE 172-4 LABORATORY ASSAYS FOR HEPARIN-INDUCED THROMBOCYTOPENIA

ASSAY	SENSITIVITY (%)	SPECIFICITY (%)	POSITIVE PREDICTIVE VALUE (%)	NEGATIVE PREDICTIVE VALUE (%)
Functional assay (e.g., serotonin release assay)	88	≈100	≈100	81
PF4/heparin enzyme immunoassay (ELISA)	95-98	86	93	95

ELISA = enzyme-linked immunosorbent assay; PF4 = platelet factor 4.

TREATMENT

Rx

If a patient has HIT, all heparin should be stopped immediately. This includes subcutaneous injections of "minidose" heparin, heparin flushes of intravenous lines, and low-molecular-weight heparin; even heparin-coated intravenous catheters should be withdrawn. Alternative anticoagulation, such as a direct thrombin inhibitor like argatroban (Chapter 38), should be administered, at least until the platelet count normalizes. Although not as well studied, fondaparinux appears to be a viable alternative. In contrast, low-molecular-weight heparin should not be used because this drug can react with the pathologic HIT antibodies. Warfarin should also not be used initially in cases of acute HIT because of its delayed therapeutic effect and its association with venous limb gangrene. When the platelet count has returned to a normal level after an acute episode of HIT, however, warfarin can be slowly introduced at a dose of 5 mg or less daily and gradually increased to achieve an international normalized ratio (INR) of 2 to 3 for a duration of at least 4 to 6 weeks because HIT is a risk factor for subsequent venous thromboembolism.[3] Because patients rarely become profoundly thrombocytopenic as a result of HIT alone, platelet transfusions are typically not required. In fact, some reports suggest that platelet transfusions can actually precipitate thrombotic complications, although this premise remains controversial.

Sepsis

Along with the exposure to certain drugs, bacterial and viral infections are among the most common causes of acute thrombocytopenia in hospitalized patients. Acute thrombocytopenia can be caused by the deposition of antibody-antigen complexes on the platelet surface through an "innocent bystander" phenomenon. These antibody-coated platelets are then cleared from the circulatory system by Fc receptor–expressing macrophages in the spleen. Thrombocytopenia associated with infections can also be due to DIC. The treatment of sepsis-induced thrombocytopenia is directed at treating the underlying cause, along with the administration of platelet transfusions as clinically indicated.

Idiopathic (Immune) Thrombocytopenic Purpura

DEFINITION

ITP is an autoimmune disorder caused by circulating antiplatelet autoantibodies. An ITP-like picture can also be found in patients with autoimmune diseases, such as in systemic lupus erythematosus (Chapter 266); in low-grade lymphoproliferative disorders, such as chronic lymphocytic leukemia (Chapter 184); and in HIV infection (Chapter 393).

CLINICAL MANIFESTATIONS

ITP was originally thought to be a disease of young women. Although this description is appropriate for many individuals with this disorder, more recent data indicate that ITP can occur in patients of either sex and at any age. ITP is a chronic, recurring disorder in most adults with this disease. This is in stark contrast to pediatric patients, who usually suffer from acute ITP and rarely have the chronic variant of this disorder.

In contrast to patients with coagulation factor deficiencies who present with bleeding deep within their tissues, individuals with ITP (or other platelet disorders) typically have excessive mucocutaneous bleeding. Consequently, the clinician should inquire whether the patient has noticed epistaxis, gingival bleeding, easy bruising, hematuria, melena, or hematochezia. Female patients should also be asked about inappropriate or excessive vaginal bleeding. The physical examination should pay particular attention to signs of mucocutaneous bleeding. The patient should be thoroughly examined for petechiae (see Fig. 172-2) and ecchymosis as well as for evidence of hemorrhage within the conjunctiva, retina, and central nervous system.

DIAGNOSIS

In patients with ITP, the CBC is usually normal except for the thrombocytopenia, and their peripheral blood smear is remarkable only for a decreased number of platelets, some of which may be larger than normal. Splenomegaly is absent unless the ITP is due to an underlying disorder, such as lymphoma that is associated with splenomegaly. A bone marrow examination is usually not necessary in the absence of findings that suggest a different disease such as myelodysplasia.[4] If performed, the bone marrow aspirate and biopsy of ITP patients typically show normal or increased numbers of megakaryocytes. However, the rest of their bone marrow is otherwise normal. These bone marrow findings are similar to what is observed in the bone marrows of patients who have other forms of destructive thrombocytopenia. Antiplatelet antibody assays are insufficiently sensitive or specific to be clinically useful.

TREATMENT

Rx

Because platelet production is assumed to be increased in patients with ITP as a result of the immune-mediated accelerated platelet destruction, traditional therapy has focused on moderating this immune response. For most of the latter half of the 20th century, splenectomy and corticosteroids were the sole therapies for ITP. Corticosteroids in the form of either prednisone (1 mg/kg orally daily) or high-dose dexamethasone (4-day pulses of 40 mg intravenously per day every 28 days for four to six cycles) are effective. Although corticosteroids remain first-line therapeutic modalities, immunomodulating agents, such as intravenous immunoglobulin (IVIg) and anti-D, as well as alternative immunosuppressives, have been introduced for the therapy of ITP. Alternatively, other immunosuppressives, such as cyclophosphamide, azathioprine, cyclosporine, mycophenolate mofetil, dapsone, interferon, and etanercept, might be useful.

For years, the basic mechanism of and rationale for ITP treatment revolved around altering the immune system by immunosuppressives, splenectomy, or immune modulators, such as IVIG or anti-D. A more recent addition to the ITP treatment regimen is rituximab, a "humanized" murine monoclonal antibody against CD20, which is a B-cell antigen. Although rituximab has not been approved by the U.S. Food and Drug Administration (FDA) for the treatment of ITP, it is currently widely used off label in patients who are unresponsive to splenectomy and corticosteroids. Response rates vary significantly between studies, ranging from 28% to 44% in larger trials. In a recent study of 133 newly diagnosed adult ITP patients with follow-up of up to 4 years, the combination of rituximab (375 mg/m^2 intravenously weekly for 4 weeks) plus dexamethasone (40 mg per day orally for 4 days) induced higher response rates and longer time to relapse, although also increased the incidence of grade 3 to 4 adverse events, than monotherapy with dexamethasone alone at the same doses.

Although most ITP patients have a compensatory increase in megakaryopoiesis in response to their rapid platelet destruction, plasma derived from some ITP patients was unexpectedly found to inhibit platelet production. This prompted a re-evaluation of whether impaired megakaryopoiesis contributes to the development of thrombocytopenia in this disease. Thrombopoietin (TPO) is a potent megakaryocyte colony-stimulating factor, and along with other cytokines, it increases the size and number of megakaryocytes (Chapter 156). TPO levels are not markedly elevated in ITP, suggesting that supplemental TPO could help to increase platelet production and correct the thrombocytopenia.

Early experiments demonstrated that recombinant as well as truncated forms of TPO significantly increased the platelet count in some refractory ITP patients. However, this also induced autoimmune autoantibodies against endogenous TPO, which resulted in profound and persistent thrombocytopenia. Therefore, both recombinant TPO and its truncated form were withdrawn from clinical trials. Subsequently, peptides were developed that bear no structural resemblance to TPO but still bind to and activate the TPO receptor; these agents are called *TPO receptor agonists*. Because these recombinant drugs bear little structural similarity to native TPO, they should not trigger autoimmune anti-TPO antibodies. The FDA has approved two of these drugs, and several more are being used in clinical trials. The first of these, called romiplostim (N-plate), is composed of several copies of a TPO receptor–binding peptide spliced into a recombinant antibody. This peptide agonist competes with TPO for binding to the TPO receptor and activates the receptor in an identical fashion as would endogenous TPO. The second FDA-approved TPO receptor agonist is eltrombopag (Promacta). It is an oral drug that activates the TPO receptor by binding to the receptor's transmembrane region.

Both subcutaneously administered romiplostim and orally administered eltrombopag are capable of increasing platelet counts in approximately 70% of patients with ITP. Remarkably, these drugs can also increase the platelet count in patients with ITP that is refractory to other treatment modalities, including splenectomy. However, adverse events, including bone marrow fibrosis and thromboembolism, have been reported. It should be noted that the drugs' effects disappear soon after their discontinuation.

General Principles of ITP Therapy

Initial management of ITP is guided by both symptoms and platelet count.[4] Asymptomatic patients with platelet counts of greater than 30,000/μL can be followed without treatment. If the patient is bleeding and/or has a platelet count of less than 30,000/μL, treatment with prednisone is recommended (Table 172-5). Refractory patients may require splenectomy, other immunosuppressive medications, or one of the new thrombopoiesis-stimulating agents (Table 172-6). Splenectomy has a long history of success in this disorder, and durable complete response rates are approximately 66 to 70%. Approximately half of the remaining patients who do not have normal platelet counts following splenectomy achieve a partial response that is clinically meaningful. Unfortunately, 10 to 15% of patients derive no benefit from splenectomy, and there is no test to predict whether a particular patient will respond to this treatment.

ITP patients with severe thrombocytopenia (<5000/μL) and/or those who have internal bleeding should be promptly treated with high doses of pulse corticosteroids and IVIG. Platelet transfusions may be given concurrently with IVIG for critical bleeding. In Rh-positive patients who have not undergone a splenectomy, anti-D immune globulin may be substituted for IVIg. However, some patients develop autoimmune hemolysis from this treatment.

TABLE 172-5 THERAPY FOR THE INITIAL MANAGEMENT OF IDIOPATHIC THROMBOCYTOPENIC PURPURA

ORAL PREDNISONE	
The effect is dose dependent—approximately 80% of patients respond to 1 mg/kg/day. Toxicity also increases with the dose and duration of treatment, and side effects include glucose intolerance, immunosuppression, osteoporosis, and cataracts. Relapse is typical once therapy is discontinued.	
INTRAVENOUS IMMUNOGLOBULIN	
The effect is more rapid than daily prednisone. Administered at a dose of 1 g/kg/day for 2 consecutive days or 0.4 g/kg/day for 5 consecutive days. The response rates are approximately 80%, and the effects typically last 2 to 4 weeks. Toxicity includes headache, allergic reactions, and rarely, thrombosis.	
ANTI-D IMMUNOGLOBULIN	
It is administered at a dose of 50 to 75 μg/kg IV. The response rates are dose dependent but can approach 75 to 80%. Hemolysis is a common toxicity but is usually mild. Rarely, hemolysis can be life threatening and can be associated with disseminated intravascular coagulation, renal failure, and end-organ infarction.	

TABLE 172-6 THERAPY FOR THE MANAGEMENT OF REFRACTORY IDIOPATHIC THROMBOCYTOPENIC PURPURA

ORAL PREDNISONE	
The effect is dose dependent and rapidly dissipates after discontinuation of the medication. Some patients can be maintained on a very low and tolerable daily dose (e.g., 5 mg). Long-term use is associated with infections, diabetes, osteoporosis, avascular necrosis, weight gain, and cataracts.	
ORAL DEXAMETHASONE	
This is administered at a dose of 40 mg/day for 4 consecutive days, repeated every 2 to 4 weeks for several months. Sustained response rates of 29 to 42% have been reported, although response rates are widely believed to be lower. Toxicity is similar to that of oral prednisone.	
SPLENECTOMY	
Durable (often lifelong), significant responses are seen in 65 to 70% of patients who undergo this procedure. More modest benefits are seen in another 10 to 15% of patients. There is no useful way to predict who will respond. Splenectomy is associated with surgical morbidity and some mortality (≈1-2%). Splenectomy produces lifelong immunosuppression to encapsulated and gram-positive organisms.	
RITUXIMAB	
This is given at a dose of 375 mg/m^2/wk IV for a total of 4 weeks. Significant responses are seen in 28 to 44% of patients, and these responses typically last for months. Toxicity includes reactivation of hepatitis B, immunosuppression, and rarely, progressive multifocal leukoencephalopathy.	
THROMBOPOIETIN RECEPTOR AGONISTS	
These are administered daily (eltrombopag) or weekly (romiplostim). An effect is typically seen in 2 to 3 weeks and disappears shortly after the discontinuation of the medication. Toxicity from long-term use is not well known but may include excessive thrombosis and bone marrow fibrosis.	

Thrombotic Thrombocytopenic Purpura

DEFINITION

TTP is a consumptive thrombocytopenia associated with a mechanical hemolytic anemia.[5] The classic pentad of symptoms, which are thrombocytopenia, anemia, fever, neurologic problems, and renal abnormalities, are fully present in only a minority of patients. Many patients today have only hemolytic anemia and thrombocytopenia. The poor specificity of the clinical features of this disease and the nonspecific laboratory abnormalities makes TTP difficult to diagnose. Historically, patients with untreated TTP had a mortality rate of 90% within 3 months. With modern therapy, patients with TTP have a mortality rate of approximately 10 to 20%.

PATHOBIOLOGY

Some patients with recurrent TTP have very large multimers of von Willebrand factor (Chapter 173). Because larger multimers of von Willebrand factor are more efficient at binding and activating platelets than smaller multimers, it was speculated that TTP could be due to a deficiency of a protease that normally cleaves large multimers of von Willebrand factor into smaller, less sticky multimers. A genome-wide linkage analysis of patients with inherited and recurrent TTP (Upshaw-Schulman syndrome) demonstrated a deficiency of a metalloproteinase that has now been identified as ADAMTS13 (*a d*isintegrin-like *a*nd *m*etalloprotease with *t*hrombospondin type 1 motif, member *13*). Many patients with adult-onset TTP have an antibody against ADAMTS13 that thus causes an acquired deficiency of ADAMTS13. Inherited or acquired ADAMTS13 deficiency causes the accumulation of ultralarge multimers of von Willebrand factor in plasma, which leads to platelet activation and the microvascular thrombosis that is characteristic of this disease (Fig. 172-4). Because many patients can have a partial deficiency of ADAMTS13 without having TTP, clinical assays of ADAMTS13 enzymatic activity have not been useful in diagnosing TTP. Levels of ADAMTS13 below 5% of normal are hypothesized to be highly suggestive of TTP, but levels greater than this are not very helpful. This hypothesis still needs to be validated in large clinical trials. Consequently, analysis of ADAMTS13 enzymatic activity should not be used to diagnose TTP outside of a research setting at this time.

Some patients reportedly developed TTP soon after taking an antiplatelet drug of the thienopyridine class. Approximately 1 of 2000 patients taking ticlopidine will develop TTP. In several individuals with ticlopidine-induced TTP, an antibody against ADAMTS13 was identified. A few studies have also suggested that clopidogrel can induce TTP at an incidence as rare as 1 in 250,000. Because the incidence of TTP in the general population is approximately 1 in 100,000, it is difficult to determine whether there is a significant risk for TTP in patients who are prescribed clopidogrel. The thienopyridine-derivative prasugrel has also been implicated. It should be noted that some patients can also develop TTP after exposure to medications that directly injure the vascular endothelium (such as cyclosporin or tacrolimus). Most of these patients do not have TTP but have the related disorder HUS, discussed later in this chapter.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

In TTP, patients can occasionally present with symptoms of excessive mucocutaneous bleeding, but they might present with signs of a thrombotic event (including phlebitis, myocardial infarction, or stroke). Many patients complain of abdominal pain, which is presumably due to intestinal ischemia. Signs of central nervous system disease, somnolence, and even coma can also be seen on presentation.

The two major hallmarks of TTP are microangiopathic hemolytic anemia and thrombocytopenia. Microangiopathic hemolytic anemia is a non-immune-mediated hemolytic anemia caused by red cell fragmentation. Patients with this disorder have typical laboratory findings of hemolytic anemia, including a decreasing hemoglobin concentration, a high lactate dehydrogenase (LDH) level, elevated indirect bilirubin, and an increased reticulocyte count. Examination of the peripheral blood smear (Chapter 157) shows torn red blood cells (schistocytes), and frequently also shows, early red blood cell precursor cells (nucleated red blood cells). The thrombocytopenia may be mild if the disease is diagnosed at an early stage, but advanced cases of TTP can have platelet counts of less than 10,000/μL. TTP should be considered in any patient who has evidence of hemolysis accompanied by thrombocytopenia.

Unlike most patients with typical thrombocytopenia, who tend to bleed excessively, patients with TTP have few hemorrhagic complications. Instead, they are markedly predisposed to thrombosis. In fact, a thrombotic complication in a thrombocytopenic patient is another clue that TTP might be the cause of the thrombocytopenia (see Table 172-1). Characteristically, the thrombotic occlusions are in the terminal arterioles and capillaries and are composed mainly of platelets within the damaged vascular lumen. In contrast to most blood clots, these occlusions contain very little fibrin and are referred to as *hyaline thrombi*. Before plasmapheresis was used in the treatment of this disease, TTP typically progressed and caused renal disease, neurologic symptoms, and fever. These symptoms are believed to be due to ischemia and infarction of the affected organs. Today, TTP is often diagnosed in its early stages, so patients may have only microangiopathic hemolytic anemia and thrombocytopenia.

TREATMENT Rx

The prompt initiation of plasma exchange (plasmapheresis with plasma replacement) reduces the mortality rate associated with TTP from 90% to approximately 15%. The mechanism of this benefit is not entirely known. A randomized trial demonstrated that plasma exchange is more beneficial than plasma infusion for the treatment of TTP. This implies that some of the benefit of plasmapheresis is attributable to the removal of a pathologic substance from the patient's plasma. Assuming that the pathogenesis of most acquired forms of TTP is the antibody-mediated inhibition of ADAMTS13, plasma exchange might have two benefits. First, it helps to remove the pathogenic antibody from the plasma. Second, the normal plasma infused into the patient during plasmapheresis repletes the deficiency of ADAMTS13. Both of these

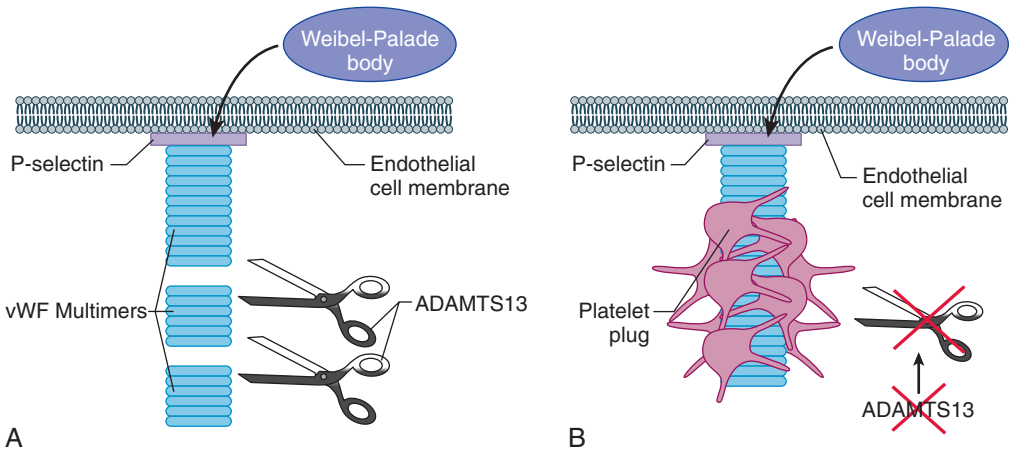

FIGURE 172-4. Pathophysiology of thrombotic thrombocytopenia purpura (TTP). A, Von Willebrand factor (vWF) is synthesized in endothelial cells and stored in Weibel-Palade bodies. The vWF is assembled into ultralarge multimers that are cleaved after they are released into the blood stream by the protease ADAMTS13. The resultant smaller multimers of vWF can bind to platelets to participate in normal hemostasis. **B,** Some patients with TTP have a deficiency of ADAMTS13. This results in the accumulation of ultralarge multimers of vWF within the circulation, which promotes the excessive adhesion of platelets. This produces large hyaline plugs of vWF and platelets that cause vascular occlusions.

benefits probably play a role in the efficacy of plasma exchange in the treatment of this disease.

Because patients with TTP can develop sudden thrombotic events (including stroke and myocardial infarction), and because patients can deteriorate quickly, it is prudent to initiate plasma exchange as rapidly as possible. Given the morbidity and mortality of untreated TTP, and given the relatively low risk of administering plasmapheresis, this therapy should be initiated even when the diagnosis is not certain. As discussed in the section on HUS (later in this chapter), plasmapheresis is not beneficial in children with Shiga toxin–induced microangiopathic hemolytic anemia or in patients who develop HUS following their exposure to endothelium-toxic drugs, such as chemotherapy.

Plasma exchange is administered to replace one entire plasma volume and is usually repeated once daily. An average-sized individual requires 20 to 30 units of fresh-frozen plasma for each plasmapheresis session. Therapy is usually administered daily, and patients are monitored for signs of improvement in their thrombocytopenia, hemolysis, neurologic symptoms, fever, and renal disease. Appropriate daily laboratory tests to monitor the patient are CBC, LDH, reticulocyte count, and creatinine. After the thrombocytopenia and hemolysis have been corrected for a few days, the daily plasmapheresis can either be discontinued or continued every other day for several more days. A typical duration of therapy is 1 to 2 weeks, although some patients require treatment far beyond this usual time course.

Approximately one third of patients relapse quickly after plasmapheresis is stopped. These patients require a reinitiation of plasma exchange and perhaps the addition of an immunosuppressive drug, such as a glucocorticoid. Some reports indicate that rituximab shows promise in reducing the incidence of relapse. However, given that this drug does not decrease antibody production for weeks to months after its administration, it is likely that if rituximab has any benefit in the treatment of TTP, it is probably only to minimize the incidence of late relapses.

Major complications of plasmapheresis do occur and are often attributable to the use of a central venous catheter. Hypotension, bacteremia, hemorrhage, and thrombosis are among the most common life-threatening complications of plasmapheresis therapy. Two specific complications deserve emphasis. First, patients taking an angiotensin-converting enzyme (ACE) inhibitor are susceptible to plasmapheresis-induced hypotension. This is due to these drugs' interference with bradykinin catabolism. Second, a patient who develops recurrent thrombocytopenia and a fever several days to a week into treatment may not be having an exacerbation of TTP but rather may be developing central line–induced sepsis. Therefore, a vigorous investigation for infection should be initiated in patients who appear to be relapsing early.

Hemolytic-Uremic Syndrome

DEFINITION

HUS is another cause of microangiopathic hemolytic anemia, and it may be difficult to distinguish from TTP. When acute renal failure is predominant, and there are no neurologic symptoms, many clinicians consider the syndrome to be HUS, not TTP.[6] In contrast to TTP, HUS is not caused by a deficiency of ADAMTS13. However, at this point, it is not clear whether ADAMTS13 levels will be clinically useful to distinguish between HUS and TTP. Because there is such overlap between the signs and symptoms of TTP and those of HUS, the distinction and diagnosis may be impossible to make in some cases.

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

A toxin that directly damages endothelial cells can cause HUS. These direct endothelial toxins are either infectious or the result of a drug. Perhaps the best-characterized infectious agent that damages endothelial cells is Shiga toxin, which produces enterohemorrhagic *Escherichia coli* (Chapter 304). In Shiga toxin–induced HUS, the renal disease and microangiopathic hemolytic anemia occur after an episode of diarrhea that is often bloody. *E. coli* O157:H7 is the cause of many of the cases in the United States, and *E. coli* O104:H4 causes some of the other cases.[7] The toxin damages endothelial cells within the glomeruli and promotes the adhesion of platelets and the trapping of red blood cells in the kidneys.

Several drugs can also induce HUS by directly damaging the endothelial cells. These include calcineurin inhibitors, such as tacrolimus and cyclosporine; sometimes, lowering the dose of these drugs can reverse the microangiopathic hemolytic anemia. Cytotoxic drugs such as mitomycin C, cisplatin, and bleomycin can also induce HUS by directly damaging the endothelial cells.

Inherited forms of HUS have been described, and they usually involve the dysregulation of the complement cascade (Chapter 50). The best-described genetic mutations cause the deficiency of factor H, which is a complement regulatory protease. HUS-inducing mutations in other components of complement regulation include those found in C3, CD46, and factor I.

TREATMENT AND PREVENTION

Rx

Reducing the spread of Shiga toxin–producing *E. coli* is vital because antibiotics and antimotility drugs do not lower the risk for developing HUS symptoms. In contrast to TTP, there is little evidence that plasma exchange is beneficial. The clinical course and outcome of *E. coli* O104:H4–induced HUS are similar to infections with the more common O157:H7-induced HUS. Aggressive antibiotic therapy appears to be beneficial for reducing seizures and death. However, neither plasmapheresis nor glucocorticoids seem to be helpful. In a pilot study of 12 patients, *E. coli* O104:H4–associated HUS appeared to respond to immunoadsorption. The infusion of plasma or plasmapheresis is reportedly beneficial in the treatment of the atypical forms of HUS that are caused by the dysregulation of complement. Recent evidence suggests that some patients with atypical forms of HUS respond to eculizumab, an antibody that inhibits the terminal steps of the complement cascade.[8]

Disseminated Intravascular Coagulation

DIC is a pathologic condition that depletes components of the coagulation system, including platelets. Therefore, in contrast to TTP and HUS, the thrombocytopenia of DIC is associated with the consumption of coagulation factors and increased fibrinolysis. This often leads to prolongations of the prothrombin time (PT) and activated partial thromboplastin time (aPTT), as well as to increased levels of D-dimers. In its fully manifested state, DIC can be associated with arterial and venous thrombi, hemorrhage, microangiopathic hemolysis, thrombocytopenia, excessive fibrinolysis, and the deficiency of coagulation factors, such as fibrinogen. For a more complete review of DIC and related disorders, see Chapter 175.

Thrombocytopenia during Pregnancy

Thrombocytopenia occurs in approximately 10% of pregnant women. It may be due to a normal variant of pregnancy (gestational thrombocytopenia), a pregnancy-specific condition (preeclampsia and HELLP [*h*emolysis, *e*levated *l*iver enzymes, and *l*ow *p*latelets]) syndrome, or a condition exacerbated by pregnancy (ITP, vasculitis, TTP). A recent review estimated that thrombocytopenia occurs in approximately 7 to 10% of pregnancies, and about 75% of these cases are due to gestational thrombocytopenia, 15 to 20% are secondary to hypertensive disorders, 3 to 4% are due to an immune process, and the remaining 1 to 2% are rare constitutional thrombocytopenias, infections, and malignancies.[9] The prognosis and treatment vary tremendously based on the underlying cause.

Gestational thrombocytopenia (incidental thrombocytopenia of pregnancy) is a mild, asymptomatic thrombocytopenia that typically occurs late in pregnancy. There is no association with fetal thrombocytopenia, and this maternal thrombocytopenia resolves spontaneously after delivery. Other than thrombocytopenia during previous pregnancies, women with gestational thrombocytopenia have no prior history of a low platelet count. This lack of previous thrombocytopenia helps to distinguish gestational thrombocytopenia from ITP. Additionally, in contrast to other causes of maternal thrombocytopenia, gestational thrombocytopenia does not produce a platelet count lower than 70,000/μL. Consequently, deviation from standard obstetric care is usually not required for a patient with gestational thrombocytopenia.

Preeclampsia is a syndrome that causes the gradual development of proteinuria and hypertension during the later stages of pregnancy (Chapter 239). A minority of women with preeclampsia progress to have seizures; when this occurs, the syndrome is called *eclampsia*. Approximately 15% of women with preeclampsia have thrombocytopenia, and 5% of patients with preeclampsia have platelet counts of less than 50,000/μL. Some patients have a more severe form of preeclampsia associated with HELLP syndrome. Preeclampsia and HELLP syndrome are thought to result from a factor produced within the placenta because these syndromes usually resolve rapidly and spontaneously after delivery. These syndromes can also present postpartum, but they still resolve within a few days of delivery. If these syndromes do not resolve spontaneously within 3 days, alternative diagnoses, such as ITP and TTP, should be considered. Consequently, immediate delivery of the child is recommended, if possible, because the microangiopathic process stops soon thereafter.

Most patients with ITP or vasculitis during pregnancy have a history of thrombocytopenia before pregnancy. The antiplatelet IgG antibodies that cause this disease in the mother can cross the placenta and lead to thrombocytopenia in the fetus. However, the platelet count in the unborn child does

not correlate well with the platelet count of the mother, and attempts to sample fetal blood to monitor platelet counts are associated with significant risk. Because most children born to women with ITP have a platelet count high enough for a successful delivery, the management of ITP focuses on keeping the maternal platelet count acceptable for the mother's safety. In the early stages of pregnancy, a platelet count of 30,000/μL to 50,000/μL is considered safe. In preparation for delivery, a platelet count of 50,000/μL to 80,000/μL is more desirable. This can be managed by the use of oral glucocorticoids, and if necessary, by the administration of IVIG. The safety of rituximab and thrombopoiesis-stimulating agents during pregnancy is not known. Splenectomy performed during the first trimester has a risk for inducing miscarriage, and this operation is technically difficult during the later stages of pregnancy because of uterine enlargement. There is no evidence that a cesarean section is safer than vaginal delivery.

The diagnosis of TTP during pregnancy can be difficult to make because many of the symptoms are identical to those of preeclampsia. If the symptoms develop during the early stages of pregnancy, when preeclampsia is unlikely, the diagnosis of TTP is more certain. The presence of hyperuricemia, hypoproteinemia, elevated liver transaminases, and direct bilirubin is more consistent with preeclampsia. The development of microangiopathic hemolytic anemia and thrombocytopenia in the peripartum period is presumed to be attributable to preeclampsia, and if it is safe to do so, immediate delivery of the fetus is desirable. Typically, the thrombocytopenia starts to resolve rapidly after delivery. If no improvement is seen by the third postpartum day, standard management for TTP, including plasma exchange, is recommended.

Post-transfusion Purpura and Neonatal Alloimmune Thrombocytopenia

Alloimmune thrombocytopenia is due to the sensitization to alloantigens, such as Pl^{A1} (HPA-1a). These alloimmune antibodies can cause thrombocytopenia approximately a week after a blood transfusion (post-transfusion purpura, or PTP).[10] A similar alloimmune antibody can cause thrombocytopenia in a fetus when it is produced by the mother (neonatal alloimmune thrombocytopenia, or NAIT). PTP causes profound thrombocytopenia 7 to 10 days after exposure to the small amount of platelets that contaminate most red blood cell transfusions. It can be treated with IVIG or by plasma exchange. NAIT can cause severe thrombocytopenia and bleeding in neonates, and it is treated with the transfusion of platelets derived from a Pl^{A1}-negative donor (most conveniently the newborn's mother), corticosteroids, and IVIg.

Vasculitis

Autoimmune diseases such as systemic lupus erythematosus, rheumatoid arthritis, and other forms of vasculitis can also cause thrombocytopenia. This can be due to an ITP-like process that occurs in patients with a propensity for the development of autoantibodies. The treatment for these patients is immunosuppression, similar to the treatment for other patients with ITP. Some patients with active vasculitis and endothelial inflammation consume circulating platelets by promoting their adhesion to the damaged vessel wall. This creates a clinical picture that is difficult to distinguish from that of TTP without a tissue diagnosis that documents vasculitis. Because of the difficulty of discerning TTP from a flare-up of vasculitis, these patients are often treated with both plasma exchange and immunosuppression.

Dilutional Thrombocytopenia

Thrombocytopenia does not typically occur after blood loss, possibly in part owing to the release of platelets from the splenic pool. Thrombocytopenia occasionally occurs after a massive hemorrhage. However, this is only when the hemorrhage is severe enough to necessitate the replacement of 1.5 to 2 times the total blood volume, which usually requires the transfusion of at least 15 to 20 units of packed red blood cells over a short period of time. Even in this circumstance, the thrombocytopenia is usually only mild to moderate. Therefore, the clinician should remain vigilant for another cause of the thrombocytopenia (such as DIC resulting from hemorrhage-induced hypotension and shock). When indicated, dilutional thrombocytopenia can be corrected with platelet transfusions.

Congenital Thrombocytopenias

Lifelong thrombocytopenia can be due to an inherited defect that affects platelet production or survival. These disorders can be autosomal dominant (May-Hegglin anomaly and Sebastian syndrome), autosomal recessive (Bernard-Soulier disease, Fanconi's anemia, gray platelet syndrome, and thrombocytopenia with absent radius syndrome), or X-linked (Wiskott-Aldrich syndrome). Many of these disorders are associated with other abnormalities in addition to thrombocytopenia. A congenital cause of thrombocytopenia should be suspected in a patient who has long-standing moderate thrombocytopenia that was presumed to be treatment-refractory ITP.

Grade A References

A1. Stanworth SJ, Estcourt LJ, Powter G, et al. A non-prophylaxis platelet transfusion strategy for hematologic cancers. *N Engl J Med*. 2013;368:1771-1780.

A2. Slichter SJ, Kaufman RM, Assmann SF, et al. Dose of prophylactic platelet transfusions and prevention of hemorrhage. *N Engl J Med*. 2010;362:600-613.

A3. Gudbrandsdottir S, Biryens HS, Fredericksen H, et al. Rituximab dexamethasone vs dexamethasone monotherapy in newly diagnosed patients with primary immune thrombocytopenia. *Blood*. 2013;121:1976-1981.

A4. Kuter DJ, Bussel JB, Lyons RM, et al. Efficacy of romiplostim in patients with chronic immune thrombocytopenic purpura: a double-blind randomised controlled trial. *Lancet*. 2008;37:395-403.

A5. George JN, Mathias SD, Go RS, et al. Improved quality of life for romiplostim-treated patients with chronic immune thrombocytopenic purpura: results from two randomized, placebo-controlled trials. *Br J Haematol*. 2009;144:409-415.

A6. Bussel JB, Provan D, Shamsi T, et al. Effect of eltrombopag on platelet counts and bleeding during treatment of chronic idiopathic thrombocytopenic purpura: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2009;373:641-648.

A7. Cheng G, Saleh MN, Marcher C, et al. Eltrombopag for management of chronic immune thrombocytopenia (RAISE): a 6-month, randomised, phase 3 study. *Lancet*. 2011;377:393-402.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Lieberman L, Bercovitz RS, Sholapur NS, et al. Platelet transfusions for critically ill patients with thrombocytopenia. *Blood*. 2014;123:1146-1151.
2. McKenzie SE, Sachais BS. Advances in the pathophysiology and treatment of heparin-induced thrombocytopenia. *Curr Opin Hematol*. 2014;21:380-387.
3. Kelton JG, Arnold DM, Bates SM. Nonheparin anticoagulant for heparin-induced thrombocytopenia. *N Engl J Med*. 2013;368:737-744.
4. Provan D, Stasi R, Newland AC, et al. International consensus report on the investigation and management of primary immune thrombocytopenia. *Blood*. 2010;115:168-186.
5. George JN, Nester CM. Syndromes of thrombotic microangiopathy. *N Engl J Med*. 2014;371:654-666.
6. Mele C, Remuzzi G, Noris M. Hemolytic uremic syndrome. *Semin Immunopathol*. 2014;36:399-420.
7. Menne J, Nitschke M, Stingele R, et al. Validation of treatment strategies for enterohaemorrhagic Escherichia coli O104:H4 induced haemolytic uraemic syndrome: case-control study. *BMJ*. 2012;345:e4565.
8. Legendre CM, Licht C, Muus P, et al. Terminal complement inhibitor eculizumab in atypical hemolytic-uremic syndrome. *N Engl J Med*. 2013;368:2169-2181.
9. Myers B. Diagnosis and management of maternal thrombocytopenia in pregnancy. *Br J Haematol*. 2012;158:3-15.
10. Heikal NM, Smock KJ. Laboratory testing for platelet antibodies. *Am J Hematol*. 2013;88:818-821.

REVIEW QUESTIONS

1. You are asked to see a patient who had coronary bypass surgery 6 days ago. The patient has abruptly developed a cold right foot, and his platelet count decreased from 350,000/μL yesterday to 155,000/μL today. He is afebrile, and his only medication is subcutaneous heparin for deep vein thrombosis (DVT) prophylaxis. You are suspicious that the patient may have heparin-induced thrombocytopenia. What should you recommend?

A. Discontinue heparin and start argatroban.
B. Discontinue heparin and observe platelet trend.
C. Discontinue heparin and start aspirin.
D. Send heparin-PF4 enzyme-linked immunosorbent assay (ELISA) and base treatment on results.
E. Change heparin to low-molecular-weight heparin.

Answer: A Stop the heparin and administer argatroban. This patient developed a probable thrombosis in his foot and had a greater than 50% decrease in his platelet count 6 days after starting to receive heparin. The time course and the magnitude of the decrease in the platelet count, along with the new thrombosis, places this patient at very high risk for the diagnosis of heparin-induced thrombocytopenia and at a very high immediate risk for a recurrent thrombotic event, even if the heparin was discontinued (ruling out answers B, C, and D). Consequently, this patient needs to be started on an alternative anticoagulant that is different enough from heparin to not be recognized by the immune response directed against heparin (ruling out answer E). Aspirin alone is not sufficient to prevent recurrent thrombi. The only acceptable approach would be to stop the heparin and to immediately start Argatroban. See Heparin-induced Thrombocytopenia, Table 172-1, Figure 172-2, and reference 2.

2. You have just met a patient with a platelet count of 115,000/μL. She has no other medical problems, and the reminder of her complete blood count (CBC) and blood smear are normal. Review of her old CBC indicates that her platelet count 1 year ago was 120,000/μL. What should you recommend?

A. Refer patient for bone marrow biopsy.
B. Repeat CBC in 1 week.
C. Repeat CBC in several months.
D. Analyze patient for antiplatelet antibodies.
E. Initiate folate and B_{12} therapy.

Answer: C Repeat CBC in several months. The patient has asymptomatic and stable mild thrombocytopenia. Her platelet count is probably at the lower end of the normal platelet count distribution, although she has a small chance of having mild idiopathic thrombocytopenic purpura (ITP) (see reference 1). Because her white and red blood cell counts are normal, and her peripheral blood smear is normal, there is no need for a bone marrow biopsy. There is no indication to repeat her CBC in a week because her platelet count is already documented as having been stable for the past year. Antiplatelet antibody assays have inadequate sensitivity and specificity to be clinically useful. In the absence of anemia and a documented deficiency of folate or B_{12}, there is no indication for the supplementation of these vitamins. See Figure 172-1 for further details on evaluating patients with chronic thrombocytopenia.

3. You are managing a patient with ITP, who typically has a platelet count of 60,000/μL to 90,000/μL. Today, she complains of symptoms consistent with an upper respiratory infection. Her CBC demonstrates that her platelet count is 42,000/μL today but is otherwise normal. Her blood smear is normal. She is on no new medications, and on physical examination her nares appear congested, but she has no ecchymosis, petechiae, or splenomegaly. What should you recommend?

A. Refer patient for bone marrow biopsy.
B. Repeat CBC in 1 week .
C. Start corticosteroid therapy.
D. Start intravenous immunoglobulin.
E. Start romiplostin.

Answer: B Repeat CBC in 1 week. The patient has an exacerbation of her ITP that is potentially a result her viral syndrome. Her otherwise normal CBC and blood smear exclude the need to consider diagnoses that would be demonstrated on a bone marrow biopsy (answer A). In the absence of excessive bleeding or a platelet count of less than 30,000/μL, there is no indication for therapy (excluding answers C, D, and E.) If her platelet count in a week continues to decrease, she might require a brief course of an ITP therapy, such as corticosteroids. See Idiopathic (Immune) Thrombocytopenic Purpura and reference 4.

4. One week ago, you prescribed trimethoprim-sulfamethoxazole for a patient with cellulitis. Her cellulitis is now resolved, but you order a CBC because her leukocyte count was 13,500/μL last week. Today's CBC demonstrates that her leukocyte count is now 9,200/μL with a normal differential, her hemoglobin is13.9 g/dL, and her platelet count is 77,000/μL. Looking back at the older laboratory results, you notice that this is the first time she has ever been thrombocytopenic. The patient has no symptoms or signs of excessive bleeding, and her spleen is not enlarged. Which of the following should you recommend?

A. Corticosteroids
B. Bone marrow biopsy
C. Hospitalization of patient for close observation
D. Discontinuing antibiotics and repeating CBC in 1 week

Answer: D Discontinue antibiotics and repeat the CBC in 1 week. The patient is asymptomatic and has an abrupt onset of moderate thrombocytopenia. Because her platelet count was normal a week ago, it makes it extremely likely that her thrombocytopenia is due to her exposure to a drug or infection (excluding answer B). Her current platelet count does not place her at a high risk for hemorrhage (excluding answers A and C.) See Drug-induced Thrombocytopenia and Table 172-3.

5. A patient is admitted with confusion and abdominal pain. She has a body temperature of 101.1° F, and her laboratory studies demonstrate the following: leukocyte count 4,400/μL, hemoglobin 10.5 g/dL, platelet count 11,000/μL, D-dimer 0.2 μg/mL, and creatinine 1.1 mg/dL. The peripheral blood smear shows nucleated red blood cells and fragmented red blood cells (schistocytes). Which of the following should you recommend?

A. Intravenous immunoglobulin
B. Platelet transfusion
C. Platelet transfusion and red blood cell transfusion
D. Plasmapheresis
E. Antibiotics

Answer: D Plasmapheresis. The patient has thrombocytopenia and anemia associated with schistocytes on her blood smear. This indicates that she has a microangiopathic hemolytic anemia. The presence of fever and neurologic symptoms suggests that the patient's diagnosis is thrombotic thrombocytopenic purpura (TTP). The normal D-dimer excludes the other likely diagnosis of disseminated intravascular coagulation. TTP requires immediate treatment with plasmapheresis. This treatment alone will correct the problem and increase the platelet count (this excludes answers A and B). Her anemia is mild, so there is no indication for red blood cell transfusion at this time (answer C). Because TTP frequently causes fever, there is no indication for antibiotics at this time (answer E). See Thrombotic Thrombocytopenic Purpura and Figure 172-4.

173

VON WILLEBRAND DISEASE AND HEMORRHAGIC ABNORMALITIES OF PLATELET AND VASCULAR FUNCTION

WILLIAM L. NICHOLS

VON WILLEBRAND DISEASE

DEFINITION AND EPIDEMIOLOGY

Von Willebrand disease (VWD), a usually autosomal dominantly inherited condition affecting both males and females of all ethnicities, is the most common hereditary bleeding disorder worldwide, with prevalence estimates dependent on case definition. It may also occur less frequently as an acquired disorder (acquired von Willebrand syndrome [AVWS]). Von Willebrand factor (VWF) plasma levels are variably decreased (minimally to substantially) in at least 2.5% of humans, but approximately 0.1% (one in 1000) have definite VWD with medically significant bleeding symptoms related to low VWF, and approximately 0.01% (one in 10,000) have more severe forms of VWD that often result in referral to tertiary care centers, including hemophilia centers.[1-3]

PATHOBIOLOGY

Von Willebrand disease reflects deficiency or dysfunction of VWF, a multimeric plasma glycoprotein that mediates platelet adhesion and aggregation at sites of vascular injury (Chapter 171) and that also carries and stabilizes blood coagulation factor VIII (FVIII) in the circulation. The *VWF* gene is located on chromosome 12, with a partial pseudogene on chromosome 22 that potentially complicates DNA-based mutation detection (which remains primarily a research application, with evolving clinical diagnostic potential). The protein is synthesized by vascular endothelium as approximately 270-kD subunits (protomers) that are dimerized, processed and polymerized into very large (≤~20,000 kD) hemostatically active VWF multimers, and finally secreted into the blood. Vascular endothelial cells additionally provide a storage reservoir of multimerized VWF (in intracellular Weibel-Palade

bodies) from which it can be released by stress or by drugs such as DDAVP (desmopressin). VWF is also synthesized and multimerized by bone marrow megakaryocytes and stored in circulating blood platelet α granules, from which it is secreted with platelet activation. Platelet-stored VWF represents about 10% of total blood VWF.

Multimerization of VWF is essential for its hemostatic activity that is mainly mediated by the higher-molecular-weight (larger) multimers. Circulating plasma VWF normally does not interact with platelets. However, when VWF binds to injured (deendothelialized) blood vessel walls, reflecting its collagen-binding activity, multimers can be stretched and unfolded by high intravascular shear forces, such as in the microvasculature, exposing and activating previously cryptic platelet-binding domains to promote platelet adhesion and aggregation to vessel-bound VWF. Simultaneously, activated (stretched) VWF multimers expose protomer cleavage sites for proteolysis by the circulating enzyme, ADAMTS13 (A Disintegrin And Metalloprotease domain with ThromboSpondin type 1 motifs, member 13), ultimately decreasing the size of VWF multimers and thereby downregulating VWF hemostatic function. Severe deficiency of ADAMTS13 is associated with the pathologic microangiopathy of thrombotic thrombocytopenic purpura (TTP) (Chapter 172), in which the microvascular thrombosis is mediated by ultralarge VWF multimers in the circulation.

Von Willebrand disease is classified into three major types that vary in severity and differ in clinical management, such that defining the VWD subtype is important. Type 1 VWD reflects partial quantitative deficiency of normally functioning VWF, of variable severity, and comprises 75% to 80% of individuals with symptomatic VWD. Type 2 VWD reflects qualitative VWF deficiency, with four subtypes (A, B, M, N), and comprises 20% to 25% of persons with VWD. Type 3 VWD reflects the virtual absence of VWF, with secondary near-absence of FVIII, and is rare (<1% of VWD). Inheritance of VWD is usually autosomal dominant, except that type 3 VWD is typically autosomal recessive in inheritance. Table 173-1 summarizes the classification of VWD subtypes and their basic pathophysiology.

The FVIII deficiency that often accompanies different types of VWD (see Table 173-1) is not genetically related to the disease; the gene that encodes FVIII is on the X chromosome, and its primary deficiency causes classic hemophilia (Chapter 174). FVIII deficiency in VWD is attributable to a primary deficiency or defect in the VWF molecule, which normally carries and stabilizes FVIII in the circulation. Therefore, FVIII is more rapidly cleared from the circulation, leading to its secondary deficiency in VWD.

TABLE 173-1 CLASSIFICATION OF VON WILLEBRAND DISEASE

TYPE	DESCRIPTION
1	Partial quantitative deficiency of VWF (~75%-80% of VWD)
2	Qualitative VWF defect (~20-25% of VWD)
2A	Caused by VWF mutations that decrease the proportion of large functional VWF multimers, leading to decreased VWF-dependent platelet adhesion and aggregation
2B	Caused by VWF mutations that increase platelet–VWF binding, resulting in depletion of large, functional VWF multimers. Circulating platelets are coated with mutant VWF, which may impair platelet adhesion and aggregation at sites of injury. Thrombocytopenia, persistent or intermittent, is observed in most cases. Distinguishing type 2B VWD may require RIPA to detect increased (abnormal) platelet aggregation response to low-dose ristocetin; the latter may also reflect platelet-type (pseudo) VWD caused by rare mutations in the platelet VWF receptor.
2M	Caused by VWF mutations that decrease VWF-dependent platelet adhesion and aggregation but do not deplete the large VWF multimers. Distinguishing between types 2A and 2M VWD requires VWF multimer gel electrophoretic analysis.
2N	Caused by VWF mutations that impair binding to FVIII, thereby shortening FVIII survival and lowering FVIII levels so that type 2N VWD can masquerade as an autosomal recessive form of hemophilia A. Discrimination from hemophilia A may require assays of VWF-FVIII binding.
3	Virtually complete VWF deficiency associated with markedly decreased FVIII (<1% of VWD)

FVIII = factor VIII coagulant activity; RIPA = ristocetin-induced platelet aggregation/aggregometry; VWD = von Willebrand disease; VWF = von Willebrand factor.
Adapted and modified from National Heart, Lung, and Blood Institute. *The Diagnosis, Evaluation, and Management of von Willebrand Disease.* Bethesda, MD: National Institutes of Health Publication 08-5832. December 2007 (released February 29, 2008), and Yawn BP, Nichols WL, Rick ME. Diagnosis and management of von Willebrand disease: guidelines for primary care. *Am Fam Physician.* 2009;80:1261-1268, 1269-1270.

CLINICAL MANIFESTATIONS

Individuals with VWD usually experience mucocutaneous bleeding symptoms such as easy bruising, prolonged or excessive bleeding from minor cuts or other injuries, nosebleeds or other mucosal bleeding such as gastrointestinal (GI) hemorrhage, or heavy menstrual bleeding in women, and they may be at increased risk for bleeding after surgery or invasive procedures, dental extractions, traumatic injury, or childbirth. Symptoms can range from relatively mild or infrequent bleeding in type 1 VWD to severe, life-threatening bleeding in type 3 VWD. As a group, women with VWD may be more affected by bleeding symptoms than men because of hemostatic challenges of menstruation and childbirth.

DIAGNOSIS

Clinical Assessment

The diagnosis of VWD (as in other bleeding disorders [Chapter 171]) begins with clinical assessment to evaluate for a personal and possible family history of abnormal bleeding, accompanied by focused physical examination to detect signs or symptoms of bleeding (e.g., petechiae, ecchymoses, hematomas, anemia), as well as findings that may suggest other causes of increased bleeding such as liver disease (e.g., hepatosplenomegaly, jaundice), joint or skin laxity (e.g., Ehlers-Danlos syndrome), telangiectasia (e.g., hereditary hemorrhagic telangiectasia [HHT]), or anatomic lesions on gynecologic examination. Because bleeding symptoms are common in apparently normal individuals, with a prevalence of certain symptoms as high as 25% to 50%, clinical assessment for VWD or other possible bleeding disorders can be challenging (Chapter 171).

The most important parts of the medical history include (1) family history of a known or suspected bleeding disorder; (2) review of personal surgical or other hemostatic challenges throughout life and recently, including dental extractions and traumatic injuries, and whether abnormal bleeding occurred and its severity; (3) mucocutaneous bleeding symptoms (e.g., bruising, nosebleeds, ecchymoses, GI bleeding, menorrhagia), including frequency, severity, and spontaneity of events; and (4) assessment for medical conditions that can increase the risk for bleeding, such as use of certain drugs that can impede normal hemostasis, including aspirin or other nonsteroidal antiinflammatory drugs (NSAIDs), clopidogrel, or warfarin or heparin or newer direct-acting oral anticoagulants; the presence of liver or kidney disease (e.g., cirrhosis, uremia); and a history of a low or high platelet count and blood or bone marrow disorders.[4]

Bleeding history scoring tools are being developed and applied for studying populations with VWD or other bleeding disorders but are not yet validated for routine clinical use.[5] However, in general, an increasing number of positive or abnormal clinical bleeding history items increases the likelihood that an individual has a bleeding disorder, including VWD, and may merit undergoing appropriate laboratory evaluation (Chapter 171).

Laboratory Evaluation

Because no simple, single laboratory test is available to screen for VWD, the initial laboratory evaluation requires measurements of plasma: (1) VWF antigen (VWF:Ag), (2) ristocetin cofactor activity (VWF:RCo), and (3) factor VIII coagulant activity (FVIII). All three tests are recommended for initial evaluation, and the results may not only establish the diagnosis but also suggest the type and severity of VWD if it is present.[6] These three tests are also used for monitoring therapy. Tests such as the bleeding time (BT) or platelet function analyzer (PFA-100, Siemens) assay lack sufficient sensitivity and specificity and are not recommended for routine screening.

If one or more of the three test results are abnormally low or if the ratio of VWF:RCo to VWF:Ag is below 0.5 to 0.7, additional laboratory evaluation may include selecting one or more of the following tests, based on the pattern of initial results together with clinical assessment and experience: (1) repeating the three initial VWD tests with timing and procedures optimized for conditions of the patient, the blood sample, and laboratory testing (see later discussion for additional information about these conditions); (2) VWF multimer analysis to help differentiate or exclude types 2A, 2B, or 2M VWD; (3) ristocetin-induced platelet aggregometry (RIPA), including low-dose ristocetin testing to evaluate for type 2B VWD or platelet-type (pseudo) VWD; (4) VWF-FVIII binding assay (or molecular DNA-based VWF testing) to evaluate for type 2N VWD; and (5) other tests such as VWF:CB (collagen-binding activity) or immunoassays reflecting VWF-platelet binding activity to supplement VWF:RCo testing. The latter tests (VWF:CB and VWF-platelet binding immunoassays) do not replace the VWF:RCo assay but can be useful as screening or supplementary tests.[7] Some of these tests

TABLE 173-2 PATTERNS OF LABORATORY TEST RESULTS FOR VON WILLEBRAND DISEASE DIAGNOSIS AND CLASSIFICATION*

CONDITION	VWF:RCo (IU/dL)	VWF:Ag (IU/dL)	FVIII (IU/dL)	VWF:RCo/VWF:Ag Ratio¶	RIPA	VWF MULTIMERS
Type 1 VWD	<30†	<30†‡	↓ or Normal	>0.5-0.7	Often normal	Normal
Type 2A VWD	<30†	<30-200†‡	↓ or Normal	<0.5-0.7	↓ or Normal	↓ HMW
Type 2B VWD	<30†	<30-200†‡	↓ or Normal	Usually <0.5-0.7	↑ (Low dose)	↓ HMW
Type 2M VWD	<30†	<30-200†‡	↓ or Normal	<0.5-0.7	↓ or Normal	No ↓ HMW
Type 2N VWD	30-200	30-200	↓↓	>0.5-0.7	Normal	Normal
Type 3 VWD	<3	<3	↓↓↓ (1-9)	NA	Absent	NA
"Low VWF"§	30-50§	30-50§	Normal	>0.5-0.7	Normal	Normal
Normal	50-200	50-200	Normal	>0.5-0.7	Normal	Normal

*Values in the table represent prototypical cases without additional VWF (or other disease) abnormalities. Exceptions occur, and repeat testing and clinical experience may be necessary for clarification and interpretation of laboratory test results.
†VWF values <30 IU/dL (or %) are designated as the level for definite diagnosis of VWD (especially type 1 VWD) because (1) there is a high frequency of blood type O that is associated with "low VWF" levels but not necessarily VWD; (2) bleeding symptoms are reported by a significant proportion of individuals with no disease; and (3) no abnormality in the VWF gene has been identified in many individuals who have only mildly to moderately decreased VWF levels. VWF values of 30 to 50 IU/dL include both apparently normal persons and those with mild VWD.
‡VWF:Ag is <50 IU/dL in most persons with types 2A, 2B, or 2M VWD.
§Diagnosis of VWD is not precluded for persons with VWF:RCo of 30 to 50 IU/dL if there is supporting clinical or family evidence for VWD, nor is the use of agents to increase VWF levels precluded in those who have VWF:RCo of 30 to 50 IU/dL and who may be at risk for bleeding.
¶Until more laboratories clearly define a reference range, the VWF:RCo/VWF:Ag ratio of <0.5 to 0.7 is recommended to distinguish type 1 vs. type 2 VWD variants (A, B, or M).
FVIII = factor VIII coagulant activity; HMW = high-molecular-weight VWF multimers; IU/dL = international units per deciliter (e.g., 100 IU/dL = 100% of mean normal level); NA = not applicable; RIPA = ristocetin-induced platelet aggregation/aggregometry; VWD, von Willebrand disease; VWF = von Willebrand factor; VWF:Ag = von Willebrand factor antigen; VWF:RCo = von Willebrand factor ristocetin cofactor activity; ↓, ↓↓, ↓↓↓, ↑, refer to varying degrees of decrease, or an increase, of the test result compared to the laboratory reference range.
Adapted and modified from National Heart, Lung, and Blood Institute. *The Diagnosis, Evaluation, and Management of von Willebrand Disease.* Bethesda, MD: National Institutes of Health Publication 08-5832. December 2007 (released February 29, 2008), and Nichols WL, Rick ME, Ortel TL, et al. Clinical and laboratory diagnosis of von Willebrand disease: a synopsis of the 2008 NHLBI/NIH guidelines. *Am J Hematol.* 2009;84:366-370.

TABLE 173-3 VARIABLE CONDITIONS OF THE PATIENT, BLOOD SAMPLING, AND LABORATORY TESTING AFFECTING LABORATORY EVALUATION FOR VON WILLEBRAND DISEASE

Phlebotomy conditions: An atraumatic blood draw limits the exposure of tissue factor from the site and the activation of clotting factors, minimizing falsely high or low values. Lipemia should be avoided because it may interfere with photo-optical testing methods, especially some used for VWF:RCo assay.
Patient stress level: Undue stress, such as struggling or crying in children or anxiety in adults, may falsely elevate VWF and FVIII levels. Very recent exercise can also elevate VWF levels.
Additional conditions in the individual: The presence of an acute or chronic inflammatory illness may elevate VWF and FVIII levels, as may pregnancy or administration of estrogen or oral contraceptives. Individuals with blood group O have VWF levels approximately 25% lower than those of other ABO blood groups. African Americans have higher VWF levels than whites.
Sample processing: To prevent cryoprecipitation of VWF and other proteins, blood samples for VWF assays should be transported to the laboratory at room temperature. Plasma should be separated from blood cells promptly at room temperature, and the plasma should be centrifuged thoroughly to remove platelets. If plasma samples will be assayed within 2 hours, they should be kept at room temperature. Frozen plasma samples should be carefully thawed at 37° C and kept at room temperature for <2 hours before assay.
Sample storage: Plasma samples that will be stored or transported to a reference laboratory must be frozen promptly at or below –40° C and remain frozen until assayed. A control sample that is drawn, processed, stored, and transported under the same conditions as the tested person's sample may be helpful in indicating problems in the handling of important test samples. Activity of FVIII typically is 10% to 20% lower in frozen-thawed plasma than in fresh (nonfrozen) plasma and can be even lower if blood processing or storage conditions are suboptimal.
Laboratory testing: Calibrators for assays of VWF:Ag, VWF:RCo, and FVIII should be referenced to the WHO plasma standard. These three tests have relatively high coefficients of variation (CVs of 10%-30%), especially the VWF:RCo assay. The quality of laboratory testing also varies considerably among laboratories (high interlaboratory CV). Test results can be reported as international units per deciliter (IU/dL) rather than as a percentage (%) of mean normal, if WHO-linked calibrators are used. Referencing VWF testing results to the population reference range, rather than to ABO-stratified reference ranges, may be clinically useful.

CV = coefficient of variation; FVIII = coagulation factor VIII; VWF = Von Willebrand factor; VWF:Ag = VWF antigen; VWF:RCo = VWF ristocetin cofactor activity; WHO = World Health Organization.
Adapted and modified from National Heart, Lung, and Blood Institute. *The Diagnosis, Evaluation, and Management of von Willebrand Disease.* Bethesda, MD: National Institutes of Health Publication 08-5832. December 2007 (released February 29, 2008), and Nichols WL, Rick ME, Ortel TL, et al. Clinical and laboratory diagnosis of von Willebrand disease: a synopsis of the 2008 NHLBI/NIH guidelines. *Am J Hematol.* 2009;84:366-370.

have limited availability and are mainly performed in reference laboratories. Consultation with a hemostasis specialist can help guide test selection.

Multimer analysis visualizes the distribution of plasma VWF multimers, is technically complex, is qualitatively interpreted in conjunction with results of the initial three tests and available clinical information, and is used to help determine the VWD subtype. Therefore, VWF multimer analysis is not recommended for initial VWD screening and should only be performed if initial VWD testing identifies an abnormal result (e.g., abnormally low VWF:RCo or ratio of VWF:RCo to VWF:Ag) or clinical information suggests a high likelihood of abnormal VWF multimer analysis.

Table 173-2 provides prototypical laboratory values for VWD subtypes. Diagnosis, especially for individuals with mildly decreased VWF (30%-50% IU/dL), requires correlation of clinical assessment (personal and family history of bleeding) and results of laboratory testing, the latter preferably performed or repeated in the absence of conditions associated with elevation of baseline VWF and with careful attention to blood specimen collection, processing, transportation, and storage.

The laboratory evaluation of a person with possible VWD or AVWS is relatively complex, particularly because results of the laboratory tests can be influenced by certain conditions of the patient and by variables in the blood sample and laboratory testing methodology (Table 173-3). In interpreting test results, therefore, it is important to be aware of these variabilities and the status of the patient at the time of evaluation.

TREATMENT

Management of patients with VWD is focused on treatment or prevention of bleeding episodes. Three main approaches to treatment of VWD are used individually or in combination: (1) increasing plasma concentration of VWF and FVIII by releasing endogenous VWF stores through stimulation of endothelial cells with DDAVP (desmopressin: 1-desamino-8-D-arginine vasopressin); (2) replacing or supplementing VWF and FVIII by using human plasma-derived, viral-inactivated concentrates; and (3) promoting hemostasis using hemostatic agents that work by mechanisms other than increasing VWF and FVIII. Regular prophylaxis is seldom required, and treatment is given primarily before and after planned invasive procedures or in response to episodes of bleeding. Table 173-4 outlines usual durations of treatment for a variety of surgical and other invasive procedures.

TABLE 173-4 SUGGESTED DURATIONS OF VON WILLEBRAND FACTOR REPLACEMENT FOR DIFFERENT TYPES OF SURGICAL PROCEDURES

MAJOR SURGERY 7-14 DAYS*	MINOR SURGERY 1-5 DAYS*	OTHER PROCEDURES: IF UNCOMPLICATED, SINGLE VWF TREATMENT
Cardiothoracic	Biopsy: breast, cervical	Cardiac catheterization
Cesarean section	Complicated dental extractions	Cataract surgery
Craniotomy	Gingival surgery	Endoscopy (without biopsy)
Hysterectomy	Central line placement	Liver biopsy
Open cholecystectomy	Laparoscopic procedures	Lacerations
Prostatectomy		Simple dental extractions

*Individual cases may need longer or shorter duration depending on the severity of Von Willebrand disease and the type of procedure.
VWF = Von Willebrand factor.
Adapted and modified from National Heart, Lung, and Blood Institute. *The Diagnosis, Evaluation, and Management of von Willebrand Disease.* Bethesda, MD: National Institutes of Health Publication 08-5832. December 2007 (released February 29, 2008).

TABLE 173-5 INITIAL DOSING RECOMMENDATIONS FOR VON WILLEBRAND FACTOR CONCENTRATE REPLACEMENT FOR PREVENTION OR MANAGEMENT OF BLEEDING

MAJOR SURGERY/BLEEDING
Loading dose*: 40-60 U/kg
Maintenance dose†: 20-40 U/kg every 8 to 24 hours
Monitoring: VWF:RCo and FVIII trough and peak at least daily
Therapeutic goal: trough VWF:RCo and FVIII >50 IU/dL for 7-14 days
Safety parameter: do not exceed VWF:RCo >200 IU/dL or FVIII >250-300 IU/dL
May alternate with DDAVP for latter part of treatment
MINOR SURGERY/BLEEDING
Loading dose*: 30-60 U/kg
Maintenance dose†: 20-40 U/kg every 12 to 48 hours
Monitoring: VWF:RCo and FVIII trough and peak at least once
Therapeutic goal: trough VWF:RCo and FVIII >50 IU/dL for 3-5 days
Safety parameter: do not exceed VWF:RCo >200 IU/dL or FVIII >250-300 IU/dL
May alternate with DDAVP for latter part of treatment

*Loading dose is in VWF:RCo IU/dL.
†Dosing intervals reflect approximately 12-hour average half-lives of plasma VWF and FVIII without conditions resulting in shortened survival or enhanced clearance such as bleeding or surgery.
DDAVP = desmopressin: 1-desamino-8-D-arginine vasopressin; FVIII = factor VIII; VWF = Von Willebrand factor; VWF:RCo = VWF ristocetin cofactor activity.
Adapted and modified from National Heart, Lung, and Blood Institute. *The Diagnosis, Evaluation, and Management of von Willebrand Disease.* Bethesda, MD: National Institutes of Health Publication 08-5832. December 2007 (released February 29, 2008).

Desmopressin (DDAVP)

DDAVP is a synthetic derivative of the antidiuretic hormone and causes release of preformed VWF from endothelial cells. It is mainly used in type 1 VWD; is contraindicated for type 2B VWD (because it can cause thrombocytopenia); has limited efficacy for types 2A, 2M, or 2N VWD; and has no efficacy for type 3 VWD. DDAVP is generally administered for short time periods (48-72 hours) and usually not more frequently than at 24- to 48-hour intervals, owing to tachyphylaxis and side effects. If it is required for longer periods or more frequently, the patient should be monitored for fluid and electrolyte problems because DDAVP may cause symptomatic hyponatremia. DDAVP therapy is potentially mildly thrombogenic, especially for persons who are at increased risk for atherosclerotic cardiovascular disorders such as stroke or myocardial infarction.

The hemostatic dose of DDAVP is 0.3 µg/kg body weight infused intravenously over about 30 minutes or administered subcutaneously; however, a concentrated formulation for the latter is not available in the United States. For outpatients, DDAVP can also be administered nasally using a concentrated spray (Stimate, CSL Behring) that delivers 150 µg per nostril (300 µg total dose for persons ≥50 kg body weight). Peak increments of plasma VWF and FVIII occur about 1 to 2 hours after DDAVP, typically reaching two- to fourfold higher levels than at baseline, and decline toward baseline during the next 24 hours, reflecting VWF and FVIII plasma half-lives that are about 12 hours on average. However, plasma half-lives depend on the specific VWF subtype and phenotype and also vary considerably among individuals. Because of variability in response, before therapy with DDAVP, it is important to perform a treatment trial with pharmacokinetic monitoring of VWF and FVIII, measuring baseline (pre-DDAVP) and peak (~1 hour post-DDAVP) levels, supplemented with testing about 4 to 6 hours after DDAVP if a shortened survival of endogenous VWF is a consideration. A DDAVP treatment trial with extended monitoring can also help diagnose VWD variants with intrinsically heightened VWF clearance. Because of the problem of rapid tachyphylaxis, the trial of DDAVP should be performed at least 1 to 2 weeks before an elective procedure.

Factor Replacement Therapy

Alphanate SD/HT (Grifols), Humate-P (CSL Behring), and Wilate (Octapharma) are U.S. Food and Drug Administration–approved plasma-derived VWF concentrates for treatment of VWD by intravenous infusions. They also contain FVIII but differ in VWF/FVIII ratios and in content of large (high-molecular-weight) VWF multimers. Other VWF-containing concentrates include Koate-DVI (Talecris) and Wilfactin (LFB, France), which lacks FVIII (endogenous levels of which should rise secondarily within a few hours after infusion) and is not available in the United States. Cryoprecipitate is no longer recommended for VWF (or for FVIII) replacement, and its use should be limited to urgent situations when VWF concentrates are not available.

Table 173-5 outlines dosing and laboratory monitoring recommendations for treatment or prevention of bleeding in patients with VWD. Whenever possible, particularly for individuals with more severe forms of VWD, major surgeries or bleeding events should be managed in hospitals with appropriate laboratory capability and clinical staff, including a hematologist and a surgeon skilled in the management of bleeding disorders.

Other Hemostatic Agents

Treatment with combined oral contraceptive pills can ameliorate menorrhagia in women with VWD, mediated partly by hormonal effects on uterine tissues as well as by elevating blood levels of VWF and FVIII. Contraceptive hormonal skin patches have similar effects, and either type of therapy can also increase the risk for thromboembolic events. The levonorgestrel-releasing intrauterine device is an alternative to oral or skin-patch hormonal agents. For pregnancy and delivery in women with VWD, it is important to ensure that VWF and FVIII levels are normalized at the time of delivery, either spontaneously (which often occurs) or by therapy, including before administration of epidural anesthesia and postpartum.

ACQUIRED VON WILLEBRAND SYNDROME

Acquired von Willebrand syndrome refers to deficiencies or defects in VWF concentration, structure, or function that are not inherited but are consequences of other medical disorders. AVWS is less common than congenital (hereditary) VWD and is typically associated with several different mechanisms and medical conditions (Table 173-6).

Laboratory findings in AVWS are similar to those in congenital VWD (types 1, 2A, or 3). Although VWF:RCo values are typically decreased in AVWS (with variably decreased VWF:Ag or FVIII), sometimes only VWF multimer analysis is abnormal, with mild to moderate reduction or loss of the highest-molecular-weight (largest) multimers. This latter situation is more likely for AVWS associated with enhanced VWF proteolysis reflecting shear-induced conformational changes in VWF leading to increased proteolysis of VWF by ADAMTS13, such as with severe aortic valvular stenosis and other conditions causing abnormally high shear forces somewhere in the circulation (see Table 173-6).[8] Heyde's syndrome refers to AVWS caused by severe aortic stenosis and accompanied by GI bleeding from arteriovenous malformations (AVMs).

Acquired von Willebrand syndrome and disorders causing it should be considered in individuals found to have abnormal VWF test results and bleeding symptoms without a personal or family history consistent with hereditary VWD. Conversely, when bleeding occurs in association with one of the known causative conditions listed in Table 173-6, AVWS should be considered and initial VWD testing performed if indicated.

TREATMENT

Rx

Treatment of AVWS should be focused first on elimination or amelioration of the associated causative disorder, if amenable to treatment (e.g., aortic valve replacement or repair for Heyde's syndrome). Survival of both endogenous

TABLE 173-6 CAUSES OF ACQUIRED VON WILLEBRAND SYNDROME

PATHOPHYSIOLOGIC CATEGORY*	DISEASE OR ASSOCIATION
Antibodies to VWF	Monoclonal gammopathies, lymphoproliferative disorders, or autoimmune diseases such as SLE
Shear-induced VWF conformational changes leading to increased proteolysis of VWF	Aortic valvular stenosis, VSD, hypertrophic obstructive cardiomyopathy, LVAD, or primary pulmonary hypertension
Markedly elevated blood platelet count	Essential thrombocythemia, polycythemia vera, myeloid metaplasia with myelofibrosis, or other myeloproliferative neoplasms
Removal of VWF from circulation by aberrant binding to tumor cells	Wilms' tumor and certain lymphoproliferative or plasma cell proliferative disorders
Decreased VWF synthesis	Hypothyroidism
Drugs associated with AVWS	Ciprofloxacin, valproic acid, hydroxyethyl starch, griseofulvin

*Pathophysiologic categories are listed in descending order of approximate prevalence.
AVWS = acquired von Willebrand syndrome; LVAD = left ventricular assist device; SLE = systemic lupus erythematosus; VWF, von Willebrand factor; VSD = ventricular septal defect.
Adapted and modified from Nichols WL, Rick ME, Ortel TL, et al. Clinical and laboratory diagnosis of von Willebrand disease: a synopsis of the 2008 NHLBI/NIH guidelines. *Am J Hematol.* 2009;84:366-370.

and infused VWF is often shortened in AVWS, and replacement therapy should be monitored with measurement of plasma VWF (VWF:RCo and VWF:Ag) and FVIII. For AVWS caused by monoclonal gammopathy of undetermined significance, intravenous immunoglobulin infusion may temporarily normalize plasma VWF and FVIII and stop abnormal bleeding by abrogating heightened VWF clearance. However, such treatment reflects an off-label product use.

BLEEDING CAUSED BY QUALITATIVE PLATELET DISORDERS

Hereditary Platelet Bleeding Disorders

Hereditary defects in platelet function can occur at all stages of the linked sequence of events involved in hemostatic platelet activation at sites of vascular injury, as described in Chapter 171: (1) platelet adhesion (Bernard-Soulier syndrome [BSS]), (2) platelet release reaction (storage pool disorders), and (3) platelet aggregation (Glanzmann thrombasthenia [GT]).[9,10] They lead to varying levels of severity of bleeding.

BERNARD-SOULIER SYNDROME AND GLANZMANN THROMBASTHENIA

These rare, autosomally recessive platelet hypofunctional disorders typically manifest moderately severe (minimally provoked or spontaneous) mucocutaneous bleeding symptoms during childhood and beyond as well as abnormal bleeding with hemostatic challenges such as surgery. BSS results from deficiency or dysfunction of the platelet membrane glycoprotein (GP) Ib-IX-V complex that is the principal receptor for binding VWF, mediating platelet adhesion to injured blood vessels. (In some ways, therefore, BSS is the platelet counterpart of VWD as an adhesion defect, the latter being caused by an abnormality in the plasma VWF rather than the platelets.) GT reflects deficiency or dysfunction of the platelet GPIIb-IIIa complex that is the principal receptor for binding plasma fibrinogen, mediating platelet aggregation. The platelet count and morphology are normal in GT, but BSS demonstrates moderate thrombocytopenia and enlarged (giant) platelets. BT or PFA-100 test results are typically markedly abnormal in both disorders but are nonspecific. Platelet aggregometry findings are usually diagnostic. BSS demonstrates an absence of aggregation response to ristocetin, but responses to other agonists are relatively normal. Conversely, GT demonstrates relatively normal response to ristocetin, contrasting with complete absence of aggregation response to other agonists, such as adenosine diphosphate, collagen, arachidonic acid, and epinephrine. Quantitative analysis of platelet membrane GPs by flow cytometry is evolving as a supplemental diagnostic tool for BSS and GT. Mutational DNA-based testing is primarily a research tool.

PLATELET STORAGE POOL DISORDERS

Platelet dense (δ) granule storage pool deficiency (DG-SPD) is more common than BSS and GT. Inheritance is typically autosomal recessive or occasionally dominant, with evolving understanding of different mutational causes. Bleeding symptoms are usually mild but may be clinically significant. Hermansky-Pudlak syndrome (HPS) is oculocutaneous albinism associated with platelet DG-SPD, with a propensity to develop pulmonary fibrosis or granulomatous colitis. DG-SPD can accompany some other hereditary disorders such as the X-linked Wiskott-Aldrich syndrome, in which it is associated with microthrombocytopenia, or Chédiak-Higashi syndrome with partial albinism and leukocyte inclusions. Isolated DG-SPD, including HPS, demonstrates normal platelet count and morphology by peripheral blood smear review. Results of BT, PFA-100, or platelet aggregometry testing may or may not be abnormal. Platelet content and secretion of adenosine triphosphate are decreased in DG-SPD, and platelet electron microscopy can diagnostically confirm absence or marked decrease of dense granules; however, these tests have limited availability.

Platelet α-granule deficiency manifests as the gray platelet syndrome, which is typically autosomal recessive in inheritance and may result in mild bleeding symptoms. Moderate thrombocytopenia is present, and peripheral blood smear review is usually diagnostic, demonstrating enlarged platelets that are "gray" (absence of granulomere staining, reflecting absence of α granules and their contents that include VWF). X chromosome–linked mutations of *GATA1*, the gene encoding a transcription factor essential for erythropoiesis and megakaryocytopoiesis, can cause a gray platelet syndrome–like disorder affecting males.

OTHER HEREDITARY PLATELET HYPOFUNCTIONAL DISORDERS

Hereditary platelet secretion disorders (defective release of α- and δ-granule contents without granule deficiency) comprise a variety of abnormalities affecting different platelet receptors or mechanisms of signal transduction and platelet activation, including defective platelet procoagulant activity, with or without thrombocytopenia or other syndromic features, and with variable bleeding propensity and symptoms. Some of these disorders are reviewed in Chapter 172.

TREATMENT Rx

For treating or preventing major bleeding events (e.g., certain surgical challenges), judicious use of platelet transfusion may be indicated for patients with more severe hereditary platelet bleeding disorders such as BSS and GT. To reduce the risk for platelet alloimmunization, single-donor apheresis platelet concentrates are preferred when available, and HLA-matched transfusions may sometimes be indicated. DDAVP administration has been reported to improve hemostasis in patients with DG-SPD, BSS, and some other hereditary platelet disorders. Treatment with recombinant activated coagulation factor VII (NovoSeven, NovoNordisk) has been reported as an alternative or salvage therapy for bleeding in GT and certain other platelet hypofunctional disorders. Antifibrinolytic therapy with epsilon-aminocaproic acid or tranexamic acid can be useful for dental extractions or for surgical procedures involving other tissues with high intrinsic fibrinolytic activity (e.g., nose, mouth and throat, extraocular tissues).[11]

Acquired Platelet Bleeding Disorders

DRUGS

Drugs are a relatively common cause of platelet hypofunction that can result in mild bleeding symptoms, particularly if other bleeding propensities are present, such as thrombocytopenia.[12] Aspirin, thienopyridines (e.g., clopidogrel, prasugrel, ticagrelor), dipyridamole, and inhibitors of platelet GPIIb/IIIa function are used therapeutically or prophylactically for atherosclerotic cardiovascular disorders such as coronary artery or cerebrovascular disease. In addition to aspirin, other NSAIDs can inhibit platelet arachidonic acid metabolism (Chapter 37) and contribute to bleeding. Other agents that may sometimes impair platelet function include certain selective serotonin reuptake inhibitors or antibiotics and some herbal or nutritional supplements. Laboratory testing for platelet hypofunction is not often indicated and may not yield diagnostic findings.

OTHER ACQUIRED PLATELET BLEEDING DISORDERS

Myelodysplastic and myeloproliferative disorders can sometimes manifest intrinsic platelet hypofunction disproportionate to thrombocytopenia or

thrombocytosis when they are present. Cirrhosis or liver failure can cause platelet hypofunction as well as thrombocytopenia. Uremia may induce platelet hypofunction that can be ameliorated by dialysis and erythropoietin therapy. Antibodies causing autoimmune thrombocytopenia (Chapter 172) can sometimes cause acquired platelet hypofunction (e.g., acquired GT or BSS) in addition to the low platelet count.

TREATMENT

Rx

Review of patient medications and medical conditions and modifications of them when feasible and indicated can often suffice for recognition and management of acquired platelet hypofunctional disorders.

VASCULAR HEMORRHAGIC DISORDERS

Hereditary Vascular Hemorrhagic Disorders

Hereditary hemorrhagic telangiectasia, also called Osler-Weber-Rendu syndrome, is an autosomal dominant vascular disorder characterized by development of telangiectases and AVMs in the skin, mucous membranes, and certain viscera (especially the central nervous system, lung, and liver), with a propensity for severe recurring nosebleeds and GI bleeds resulting in chronic anemia and iron deficiency.[13] The diagnosis primarily relies on physical examination (Fig. 173-1). Several mutations in two genes have been described: endoglin in HHT type 1 and activin receptor-like kinase-1 (*ALK1*) in HHT type 2. These genes encode proteins that modulate transforming growth factor-β signaling in vascular endothelial cells, and mutations in them in HHT lead to the development of fragile telangiectatic vessels and AVMs. Molecular DNA-based testing has limited availability. Supportive treatment includes supplemental iron therapy for iron deficiency anemia. Among 24 patients with HHT, severe hepatic vascular malformations, and high cardiac output, bevacizumab (5 mg/kg every 14 days for a total of six injections) normalized the cardiac index in 20 of 24 patients (complete response in three; partial response in 17), with concomitant reductions in epistaxis and improvements in quality of life.[14] Tranexamic acid (1 gram three times daily) is effective for reducing recurrent epistaxis, and intranasal administration is also being studied. In a small randomized trial, tamoxifen (20 mg per day for 6 months) also significantly reduced bleeding in patients with recurrent epistaxis.

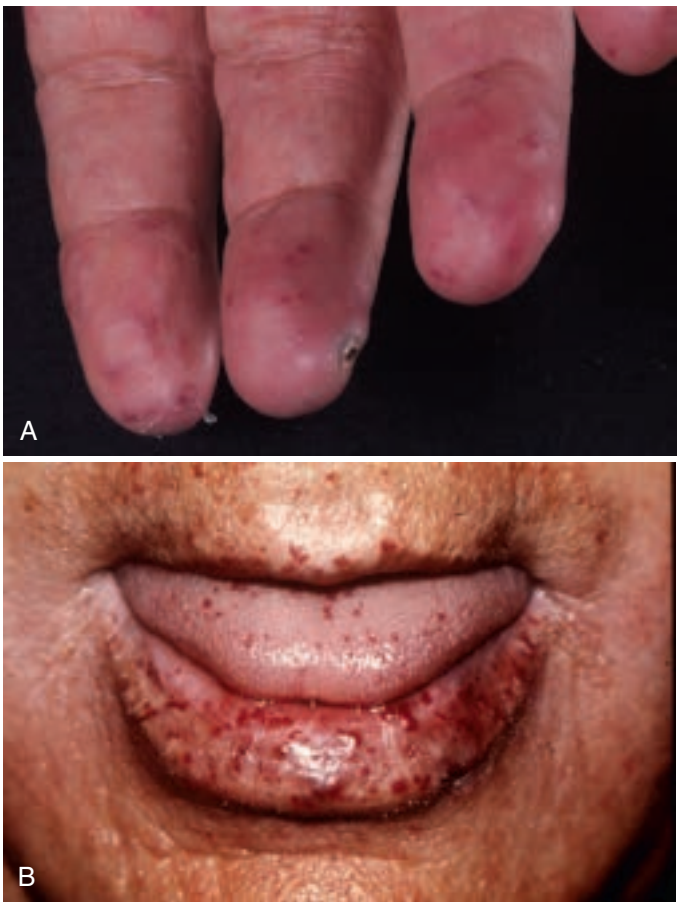

FIGURE 173-1. Hereditary hemorrhagic telangiectasia (HHT). Telangiectasias commonly occur on the fingers (**A**); face, lips, and tongue (**B**); and in other areas, including the nasal and gastrointestinal mucosa, and may develop in certain other internal organs. Skin or mucous membrane lesions typically blanch with pressure in contrast to petechiae, which do not. (**A,** Copyrighted and used with permission of the Mayo Foundation for Medical Education and Research, all rights reserved. **B,** Courtesy of Dr. Andrew Schafer.)

Ehlers-Danlos syndrome (Chapter 260) is caused by mutations in genes encoding fibrillar collagen or related genes and includes at least six subtypes that vary in clinical manifestations, severity, and prognosis.[15] Inheritance is primarily autosomal dominant. Principal manifestations of Ehlers-Danlos syndrome include skin or joint hyperextensibility, with an associated bleeding tendency including easy bruisability (purpura), surgical bleeding, poor wound healing, and menorrhagia in women. Diagnosis depends mainly on physical examination, supplemented with molecular DNA-based testing that currently has limited availability.

Acquired Vascular Hemorrhagic Disorders

Systemic amyloidosis, Waldenström macroglobulinemia, and cryoglobulinemia are dysproteinemic disorders reflecting monoclonal gammopathies or hyperglobulinemias (Chapters 187 and 188) that can manifest purpuric or other bleeding symptoms caused by either vascular deposition of immunoglobulin fragments (amyloidosis) or inhibition of platelet-vessel hemostatic functions by immunoglobulins.[16] Henoch-Schönlein purpura (Chapter 270) is primarily a transient disease of children that typically presents with a palpable purpuric rash of the lower extremities, arthralgias, abdominal pain and renal symptoms (hematuria, proteinuria) in some cases, manifestations of a systemic vasculitis characterized by deposition of immunoglobulin A–containing immune complexes in the skin, GI tract, and kidneys. Palpable purpura can also result from systemic sepsis or disseminated intravascular coagulation. Scurvy (Chapter 218) is caused by the effects of vitamin C deficiency on collagen structure and often presents with bruising or petechiae, including tiny perifollicular hemorrhages. Senile purpura or purpura simplex is thought to reflect partly the age-related changes of skin vascular structure and is typified by easy bruisability of the dorsa of the hands, forearms, and lower legs. Psychogenic purpuras are controversial entities that include Gardner-Diamond syndrome, which is characterized by recurring focal pain preceding development of ecchymoses, believed to reflect poorly understood psychosomatic mechanisms. It can be difficult to differentiate this rare syndrome from factitious or self-induced purpura and bleeding.

Grade A References

A1. Gaillard S, Dupuis-Girod S, Boutitie F, et al. Tranexamic acid for epistaxis in hereditary hemorrhagic telangiectasia patients: a European cross-over controlled trial in a rare disease. *J Thromb Haemost*. 2014;12:1494-1502.

A2. Yaniv E, Preis M, Hadar T, et al. Antiestrogen therapy for hereditary hemorrhagic telangiectasia: a double-blind placebo-controlled clinical trial. *Laryngoscope*. 2009;119:284-288.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Favaloro EJ, Bodo I, Israels SJ, et al. von Willebrand disease and platelet disorders. *Haemophilia*. 2014;20(suppl 4):59-64.
2. Laffan MA, Lester W, O'Donnell JS, et al. The diagnosis and management of von Willebrand disease: a United Kingdom Haemophilia Centre Doctors Organization guideline approved by the British Committee for Standards in Haematology. *Br J Haematol*. 2014;167:453-465.
3. Lillicrap D. von Willebrand disease: advances in pathogenetic understanding, diagnosis and therapy. *Blood*. 2013;122:3735-3740.
4. American Society of Hematology. Clinical Practice Guideline on the Evaluation and Management of von Willebrand Disease (VWD). 2012. http://www.hematology.org/Clinicians/Guidelines-Quality/Quick-Ref/528.aspx, Mobile device application available at: http://www.hematology.org/Clinicians/Guidelines-Quality/538.aspx. Accessed February 2, 2015.
5. Rydz N, James PD. The evolution and value of bleeding assessment tools. *J Thromb Haemost*. 2012;10:2223-2229.
6. Castaman G, Hillarp A, Goodeve A. Laboratory aspects of von Willebrand disease: test repertoire and options for activity assays and genetic analysis. *Haemophilia*. 2014;20(suppl 4):65-70.
7. Favaloro EJ, Bonar R, Chapman K, et al. Differential sensitivity of von Willebrand factor (VWF) "activity" assays to large and small VWF molecular weight forms: a cross-laboratory study comparing ristocetin cofactor, collagen-binding and mAb-based assays. *J Thromb Haemost*. 2012;10:1043-1054.
8. Blackshear JL, Wysokinska EM, Safford RE, et al. Indexes of von Willebrand factor as biomarkers of aortic stenosis severity. *Am J Cardiol*. 2013;111:374-381.
9. Nurden AT, Nurden P. Congenital platelet disorders and understanding of platelet function. *Br J Haematol*. 2014;165:165-178.
10. Monteferrario D, Bolar NA, Marneth AE, et al. A dominant-negative GFI1B mutation in the gray platelet syndrome. *N Engl J Med*. 2014;370:245-253.
11. Nurden AT, Freson K, Seligsohn U. Inherited platelet disorders. *Haemophilia*. 2012;18(suppl 4):154-160.
12. Konkle BA. Acquired disorders of platelet function. *Hematology*. 2011;1:1-11.
13. Faughnan ME, Palda VA, Garcia-Tsao G, et al. International guidelines for the diagnosis and management of hereditary haemorrhagic telangiectasia. *J Med Genet*. 2011;48:73-87.
14. Dupuis-Girod S, Ginon I, Saurin JC, et al. Bevacizumab in patients with hereditary hemorrhagic telangiectasia and severe hepatic vascular malformations and high cardiac output. *JAMA*. 2012;307:948-955.
15. De Paepe A, Malfait F. The Ehlers-Danlos syndrome, a disorder with many faces. *Clin Genet*. 2012;82:1-11.
16. Coppola A, Tufano A, Di Capua M, et al. Bleeding and thrombosis in multiple myeloma and related plasma cell disorders. *Semin Thromb Hemost*. 2011;37:929-945.

REVIEW QUESTIONS

1. A 40-year-old woman complains of easy bruising and menorrhagia present for a few years. Bruises have mainly occurred on the lower extremities, often after recollected minor trauma, and typically have been small (<1-3 cm in diameter) and without induration (hematomas). Her menses have been heavier in recent years but without passage of blood clots, and there is no history of anemia. A previous pregnancy and full-term vaginal delivery were uncomplicated by bleeding with delivery or postpartum. During childhood, she had occasional nosebleeds that did not require medical attention. She has not undergone surgical challenges nor had dental extractions or significant injuries. Currently, she is not using oral contraceptive or iron supplementation, nor taking aspirin or other nonsteroidal antiinflammatory drug (NSAIDs). Family history review is negative for bleeding disorder. Physical examination discloses two small superficial bruises (1-2 cm diameter) on the lower extremities but is otherwise unremarkable, including no evidence of oronasal or skin telangiectasias nor unusual skin or joint laxity. Recent laboratory testing demonstrates normal complete blood count (CBC) results, including platelet and erythrocyte measurements, and normal prothrombin time (PT) and activated partial thromboplastin time (aPTT). From previous ABO blood group testing, she is known to have blood group O. Concerning additional evaluation for possible von Willebrand disease (VWD), which *one* of the following is the most appropriate recommendation for initial laboratory blood sample testing?

A. Platelet function analyzer (PFA-100) closure time
B. Von Willebrand factor (VWF) antigen (VWF:Ag)
C. VWF:Ag and VWF ristocetin cofactor activity (VWF:RCo)
D. VWF:Ag, VWF:RCo, and coagulation factor VIII activity (FVIII)
E. VWF:Ag, VWF:RCo, FVIII, and VWF multimer analysis

Answer: D The medical and bleeding history review raises the possibility of a mild bleeding disorder such as VWD (that would not be excluded by normal CBC, PT, and aPTT test results). PFA-100 or bleeding time (BT) testing are not recommended for initial screening to detect or exclude VWD because they lack sufficient sensitivity and are nonspecific tests. The three recommended tests to screen for or exclude VWD are VWF:Ag, VWF:RCo, and coagulation FVIII. Abnormal results on one or more of these three tests may also help identify the subtype of VWD if present and its severity, and supplemental or follow-up VWD tests may be indicated for clarification. VWF multimer analysis is not recommended for initial screening for VWD unless there is high clinical suspicion of abnormal VWF multimers such as can occur in association with acquired Von Willebrand syndrome (AVWS).

2. The following blood plasma test results are obtained for the 40-year-old woman described in question 1.

VWF:Ag = 42% (reference range, 50%-200%)
VWF:RCo = 42% (reference range, 50%-200%)
FVIII = 60% (reference range, 55%-200%)
What is the most appropriate interpretation of these test results in the context of the clinical information?

A. Type 1 VWD
B. Possible type 1 VWD
C. "Low VWF"
D. Type 2A VWD
E. B and C

Answer: E The mildly decreased levels of VWF:Ag and VWF:RCo (with low normal FVIII) suggest the possibility of mild type 1 VWD vs. "low VWF" such as can occur in apparently normal individuals with blood group O. The latter can have VWF levels as low as approximately 40% of mean normal. The patient's mild bleeding history (childhood nosebleeds not requiring medical interventions, ease of bruising with small bruises only, menorrhagia that apparently has not caused anemia or iron deficiency, absence of other bleeding symptoms or events) is not definitive for VWD or another bleeding disorder at this time, although raising possibility of such. The concordant VWF:RCo and VWF:Ag levels (ratio, 1.0; reference range, ≥0.5-0.7) are not suggestive of type 2A VWD. If clinical suspicion of VWD remains, it could be appropriate to repeat the VWD panel testing in the absence of conditions associated with elevation of baseline VWF and FVIII levels (see Table 173-3).

Nichols WL, Hultin MB, James AH, et al. Von Willebrand disease (VWD): evidence-based diagnosis and management guidelines, the National Heart, Lung, and Blood Institute (NHLBI) Expert Panel report (USA). *Haemophilia.* 2008;14(2):171-232.

American Society of Hematology. *2012 Clinical Practice Guideline on the Evaluation and Management of von Willebrand Disease (VWD).* www.hematology.org/policy/resources/guidelines/vwd.

3. A 47-year-old woman is establishing medical care, having recently relocated, and has a history of von Willebrand disease (VWD) but is otherwise apparently in good health. She is unsure of the specific diagnosis such as VWD subtype and is in the process of obtaining her previous medical records. She reports a lifelong bleeding tendency, including recurring nosebleeds that have sometimes required nasal packing or cautery, excessive bleeding with cuts or other minor trauma, and troublesome bleeding after a dental extraction (prolonged oozing managed with packing). Since menarche, she has had menorrhagia that was initially managed by oral contraceptive therapy beginning at age 15 years in association with the diagnosis of VWD. Her only successful pregnancy at age 38 years was managed by cesarean section supplemented with infusions of VWF-FVIII concentrate, and she did not have bleeding complications. Family history discloses that neither of her parents had a history of abnormal bleeding, nor did two brothers or one sister nor their progeny. Her son has had troublesome nosebleeds and tonsillar bleeding and has been diagnosed as having VWD; however, no additional details or laboratory test results are currently available. Her physical examination is essentially unremarkable, including no pallor.

Initial laboratory testing results include a normal complete blood count (CBC), prothrombin time (PT) and activated partial thromboplastin time (aPTT). A VWD test profile with reflexive supplemental VWF testing of plasma yielded these results:
VWF:Ag = 24% (reference range, 50%-200%)
VWF:RCo = <12% (reference range, 50%-200%)
FVIII = 34% (reference range, 55%-200%)
VWF-platelet binding activity (latex immunoassay screen) = 7% (reference range, 50%-200%)
VWF multimer analysis = abnormal, with absence of higher molecular weight multimers and increased abundance of lower-molecular-weight multimers (normal = "normal multimer distribution")
What is the most appropriate interpretation of the VWD profile test results in the context of the clinical information?

A. Type 1 VWD
B. "Low VWF"
C. Type 2A VWD
D. Type 2B VWD
E. Type 3 VWD

Answer: C The clinical history identifies a variety of mucocutaneous bleeding episodes lifelong and altogether yields an elevated bleeding history assessment "score" that is highly suggestive of hereditary or congenital VWD or a similar bleeding disorder. The family history of VWD (son) provides additional probability that the woman has VWD. Values for plasma VWF:Ag and FVIII are substantially decreased, and VWF:RCo is undetectably low, altogether consistent with VWD. The unmeasurably low value for VWF:RCo was reflexively evaluated using an automated immunoassay based on a latex-coupled monoclonal antibody to the VWF-platelet glycoprotein (GP) Ib binding domain to indirectly measure VWF-platelet binding activity and confirms a very low but measureable value (7%). The discordantly lower values for VWF:RCo activity and VWF "immunoactivity," relative to the higher value for VWF:Ag, yield decreased activity to antigen ratios (<0.5; reference range, ≥0.7-0.8) and are indicative of a qualitative defect of VWF-platelet binding, consistent with types 2A, 2B, or 2M VWD. Supplemental VWF multimer analysis demonstrates absence of higher-molecular-weight multimers consistent with types 2A or 2B VWD and not consistent with type 2M VWD (nor types 1 or 3 VWD). The blood platelet count (CBC) is normal, providing no suggestion of type 2B VWD (or platelet-type "pseudo-VWD"), conditions that typically manifest persisting or intermittent thrombocytopenia. However, supplemental testing would be indicated to definitively exclude these rare VWD subtypes: ristocetin-induced platelet aggregation (RIPA), including demonstration of no abnormal response to low-dose ristocetin.

Nichols WL, Hultin MB, James AH, et al. Von Willebrand disease (VWD): evidence-based diagnosis and management guidelines, the National Heart, Lung, and Blood Institute (NHLBI) Expert Panel report (USA). *Haemophilia.* 2008;14(2):171-232.

American Society of Hematology. 2012 Clinical Practice Guideline on the Evaluation and Management of von Willebrand Disease (VWD). http://www.hematology.org/policy/resources/guidelines/vwd.

4. A 64-year-old man has had intermittent gastrointestinal (GI) bleeding episodes (melena or hematochezia) for about 7 years, sometimes requiring red blood cell transfusions. Repeated upper and lower GI endoscopies and capsule imaging studies have failed to identify a source of bleeding. However, recently a capsule study and subsequent extended GI endoscopy identified several focal vascular ectasias (arteriovenous malformations) in the duodenum and proximal jejunum which were photoablated. He has not had other types or episodes of bleeding, including having undergone several surgical procedures (lithotripsy for nephrolithiasis, partial parathyroidectomy for hyperparathyroidism, dental extraction) before onset of the recurrent GI bleeding. There is no family history of a bleeding disorder. Additionally, the patient has a 13-year history of exertional dyspnea with occasional syncopal episodes and 5 years ago was diagnosed as having hypertrophic obstructive cardiomyopathy (HOCM) with aortic stenosis–like left ventricular outflow obstruction that has gradually worsened by clinical and echocardiographic assessments.

Recent hematologic testing included a normal complete blood count (CBC) and platelet count and normal prothrombin time (PT) and activated partial thromboplastin time (aPTT). A von Willebrand disease (VWD) test profile with reflexive supplemental von Willebrand factor (VWF) multimer analysis yielded these plasma test results:

VWF:Ag = 159% (reference range, 50%-200%)
VWF:RCo = 116% (reference range, 50%-200%)
FVIII = 119% (reference range, 55%-200%)
VWF multimer analysis = abnormal, with absence of the highest-molecular-weight multimers and no increased abundance of lower molecular weight multimers (normal = "normal multimer distribution")

What is the most appropriate clinical diagnosis, considering the clinical and laboratory test results, including the VWD profile test results?

A. No evidence of VWD
B. Acquired Von Willebrand syndrome (AVWS)
C. Heyde's syndrome
D. B and C
E. Type 2M VWD

Answer: D The clinical history identifies recurring GI bleeding for several years but no antecedent abnormal bleeding history nor family history of such, and suggests the possibility of focal (GI) anatomic abnormalities or an acquired bleeding diathesis. Recent GI endoscopy has identified small bowel vascular ectasias or angiodysplasias (arteriovenous malformations) as the probable anatomic source of GI bleeding. The association of GI bleeding and severe aortic stenosis is known as Heyde's syndrome. Many patients with Heyde's syndrome have GI angiodysplasias (AVMs), as well as acquired abnormalities of VWF (selective absence of the highest-molecular-weight VWF multimers) that reflects effects of the abnormally high blood shear flow stress associated with severe aortic stenosis. It is thought that the VWF abnormality predisposes to bleeding from GI AVMs and may also contribute to AVM formation (dysregulated angiogenesis) and results in AVWS, which is a bleeding propensity reflecting acquired VWF changes.

In this context, answer D (Heyde's syndrome + AVWS) is the most appropriate diagnosis. However, because the patient has HOCM rather than classic aortic valvular stenosis, one might consider the diagnosis to be "Heyde's syndrome variant" or a similar appellation. Although the patient's levels of VWF:RCo and VWF:Ag are normal and adequately concordant (ratio, 0.73; reference range, ≥0.7), it is only the VWF multimer analysis that allows detection of the VWF abnormality. Alternative evolving methods for measuring VWF activity may have greater sensitivity for detecting subtle VWF multimeric abnormalities (abnormal VWF activity to antigen ratio) compared with the VWF:RCo assay. This vignette is based on a published case; the patient underwent cardiac septal myectomy and has remained free of GI bleeding for more than 7 years of follow-up.

Blackshear JL, Schaff HV, Ommen SR, et al. Hypertrophic obstructive cardiomyopathy, bleeding history and acquired von Willebrand syndrome: response to septal myectomy. *Mayo Clin Proc.* 2011:86:219-224.

5. A 75-year-old man is referred for evaluation of an 18-month history of recurring severe nosebleeds and an earlier episode of gastrointestinal (GI) bleeding. He had three episodes of severe and prolonged nosebleeds requiring nasal packing or cautery, and red blood cell transfusions were given on one occasion. One year before the onset of nosebleeds, he had melena that was evaluated with upper and lower endoscopies that identified arteriovenous malformations (AVMs) in the stomach and duodenum that were cauterized, and later a GI capsule imaging study identified a possible jejunal AVM. He had also noted some increased bruising that improved after stopping prophylactic aspirin. The patient has not previously experienced abnormal bleeding and had several surgical challenges, including bilateral knee replacements, radical prostatectomy, inguinal hernia repair, and childhood tonsillectomy and adenoidectomy. His general health has otherwise been relatively good (treated hypertension and hyperlipidemia). Family history is not suggestive of a hereditary or congenital bleeding diathesis. Physical examination is unremarkable for age.

Initial laboratory evaluation identified borderline low blood hemoglobin (12.2 g/dL) with normocytic erythroid indices, and normal blood platelets and leukocyte count and differential. Serum ferritin was decreased (12 mcg/L; reference range, 24-336 mcg/L). Hemostatic testing included normal prothrombin time (PT), thrombin time, and fibrinogen level, however the activated partial thromboplastin time (aPTT) was mildly prolonged (48 sec; reference range, 26-36 sec), and a 1 : 1 mixing study with normal plasma (nonincubated) demonstrated correction. Coagulation factor activity assays demonstrated substantially decreased factor VIII (FVIII, 8%; reference range, 55%-200%), but factors IX, XI and XII were normal, as was a factor XIII screening assay. Supplemental testing was negative for a time-dependent FVIII inhibitor. Von Willebrand disease (VWD) testing demonstrated markedly decreased plasma VWF:Ag (9%; reference range, 50%-200%) and undetectably low VWF:RCo (<12%; reference range, 50%-200%), with supplemental VWF immunoactivity assay confirming low value (11%; reference range, 50%-200%). VWF multimer analysis revealed barely detectable VWF multimers, however, all multimeric species appeared to be present with no reduction of higher–molecular-weight multimers.

Considering a preliminary diagnosis of probable acquired Von Willebrand syndrome (AVWS), which *one* of the following investigations would be most likely to uncover the etiologic cause of AVWS?

A. VWF inhibitor testing
B. Transthoracic echocardiography (TTE)
C. Serum thyroid stimulating hormone (TSH) assay
D. Serum protein electrophoresis and immunoelectrophoresis
E. Bone marrow biopsy

Answer: D The broad differential diagnostic categories of disorders pathophysiologically associated with AVWS include (1) antibodies to VWF, especially monoclonal gammopathies (e.g., monoclonal gammopathy of undetermined significance [MGUS], myeloma, macroglobulinemia) and less commonly autoimmune disorders such as (systemic lupus erythematosus [SLE]); (2) lymphoproliferative disorders such as chronic lymphocytic leukemia (CLL) and lymphoma; (3) myeloproliferative disorders associated with marked thrombocytosis (e.g., essential thrombocythemia, polycythemia vera); (4) aberrant VWF binding to tumor cells (e.g., certain lymphoproliferative disorders or rarely other tumors); (5) cardiovascular disorders associated with shear-induced proteolysis of VWF, including aortic valvular stenosis, hypertrophic obstructive cardiomyopathy (HOCM), ventricular septal defect, ventricular assist devices (VADs), and so on; (6) decreased VWF synthesis such as in association with hypothyroidism; and (7) certain drugs (e.g., ciprofloxacin, valproic acid, hydroxyethyl starch). Considering the patient's age, essentially normal CBC findings, overall medical history, and the physical examination without evidence of hepatosplenomegaly or lymphadenopathy, nor aortic stenosis or similar conditions, MGUS would be high on the list of conditions to consider and evaluate. In this case, the patient was found by serum protein electrophoresis and immunoelectrophoresis to have a monoclonal IgG kappa monoclonal protein along with increased free

kappa light chain in the serum, but total serum IgG level was normal, all consistent with MGUS in the overall clinical context (Chapter 187). Monoclonal immunoglobulins can have specificity for VWF and cause clearance of immune complexes, typically without inhibiting VWF function, and causing very low VWF levels with secondary deficiency of FVIII. Inhibitors of VWF activity (VWF:RCo), either paraproteins or autoantibodies, are rarely detected by currently available testing, and VWF inhibitor test results were negative. Essentially normal CBC findings (except for mild anemia reflecting iron deficiency), together with physical examination, effectively excluded other causes of AVWS, including myeloproliferative or other lymphoproliferative or malignant disorders. The patient was not receiving drugs that can cause AVWS.

This patient received a trial infusion of VWF-FVIII concentrate (50 VWF:RCo units/kg body weight), and testing of blood samples showed transient normalization of VWF and FVIII levels that returned to baseline values within 24 hours, suggesting shortened survival. A trial infusion of intravenous immunoglobulin (IVIG, 1 g/kg body weight) was given, and serial testing of blood samples documented normalization of VWF and FVIII levels within 24 hours, with sustained normal levels for more than 5 days. It was recommended that either the former or the latter of these treatments be considered for future episodes of serious bleeding. There have been several case reports and small case series describing that infusion of IVIG can transiently normalize decreased VWF and FVIII levels in persons with IgG isotype monoclonal gammopathies.

174

HEMORRHAGIC DISORDERS: COAGULATION FACTOR DEFICIENCIES

MARGARET V. RAGNI

COAGULATION DEFICIENCIES

Severe coagulation factor deficiencies, or coagulopathies, are typically characterized by spontaneous or traumatic bleeding, such as during surgery or trauma, and may result in life- or limb-threatening complications. By contrast, moderate and mild coagulopathies may remain clinically silent until they are detected coincidentally on routine laboratory screening tests (e.g.,

prothrombin time [PT], activated partial thromboplastin time [aPTT]) or when these tests are ordered to evaluate the cause of abnormal bleeding or bruising. Much of the morbidity of coagulopathies can be minimized or avoided altogether by prophylactic replacement of the deficient clotting factor proteins. In contrast to the lifelong clinical manifestations of hereditary or congenital coagulopathies, acquired coagulation deficiencies usually appear acutely in previously asymptomatic individuals; they may not be suspected and often remit spontaneously or after eradication of an inciting disease state or withdrawal of an offending medication. Acquired coagulation deficiencies may be associated with more severe bleeding than congenital deficiencies, in part because of the delay in diagnosis. In general, coagulation factor deficiencies may result from inadequate synthesis of coagulation factor proteins or from inhibition of activated clotting factor proteins by acquired antibodies or anticoagulant medications. Finally, qualitative defects of coagulation factors, either congenital or acquired, may also result in bleeding.

Hereditary Hemophilias

DEFINITION

The hemophilias include hemophilia A and hemophilia B, caused by deficiencies or defects in clotting factor VIII (antihemophilic factor) and factor IX (antihemophilic factor B, or Christmas factor), respectively. A deficiency of either of these intrinsic coagulation pathway proteins results in inadequate formation of thrombin at sites of vascular injury.

EPIDEMIOLOGY

Hemophilia A and B are sex-linked recessive disorders estimated to occur in 1 in 5000 and 1 in 30,000 male births, respectively. The higher incidence of hemophilia A may be due to the greater amount of DNA "at risk" for mutation in the factor VIII gene (186,000 base pairs) compared with the factor IX gene (34,000 base pairs). Hemophilia A and B are observed in all racial and ethnic groups, and in the United States, more than 20,000 individuals are affected. Although carrier testing, genetic counseling, and prenatal diagnosis are widely available in the United States through the network of federally funded hemophilia treatment centers (HTCs), fecundity rates remain high, and few confirmed carriers elect to terminate their pregnancies, even if an affected fetus is detected in utero. These decisions are likely influenced by the wide availability of safe and effective coagulation factor replacement concentrates and by the prospect of an eventual cure for the hemophilias through gene transfer. A substantial proportion (30%) of hemophilia cases arise as new, spontaneous mutations. Overall, the hemophilias are much more common than the autosomal recessive coagulation factor deficiencies (see later), which often affect progeny from consanguineous relationships and require the inheritance of two defective alleles for the bleeding manifestations to become evident.

PATHOBIOLOGY

Genetics

As with other sex-linked recessive diseases, the genes for factor VIII and factor IX are located on the long arm of the X chromosome. Males with a defective allele on their single X chromosome transmit this gene to all their daughters, who are obligate carriers, but to none of their sons. Because the offspring of female carriers inherit one affected X chromosome, half of their sons develop the coagulation disorder, and half of their daughters are obligate carriers. Female carriers may manifest bleeding symptoms, particularly postpartum bleeding, if the alleles on the X chromosome are unequally inactivated (lyonization); the defective hemophilic allele is expressed in preference to the normal allele, plasma factor VIII levels are below 50%, and phenotypic hemophilia results. Female hemophilia may arise as a result of mating between a hemophilic male and a female carrier (homozygous for the defective factor VIII or IX gene) or in carrier females who have the 45,XO karyotype (Turner syndrome) and are hemizygous for the defective hemophilia gene.

No single mutation is responsible for the hemophilias. Many missense and nonsense point mutations, deletions, and inversions have been described. Severe molecular defects predominate, with 40 to 50% of all cases of severe hemophilia A evolving from a unique inversion of intron 22 (the largest of the factor VIII introns). This inversion results from the recombination and translocation of DNA within intron 22 of the factor VIII gene, with areas of extragenic but homologous "nonfunctional" DNA located at a distance from intron 22. Other less common severe molecular defects include large gene deletions (5 to 10% of cases) and nonsense mutations (10 to 15% of cases). The encoded proteins resulting from these mutations are defective and do not express any factor VIII activity. Mild or moderate hemophilia A is commonly associated with point mutations and deletions. In contrast, factor IX mutations are more diverse, and severe hemophilia B is more likely caused by large deletions. Mutated clotting factor genes responsible for the hemophilias may also encode for the production of defective nonfunctional proteins that circulate in the plasma and are detected at normal quantitative levels by immunoassays but not by functional assays. A listing of the mutations that cause the hemophilias can be accessed through the Human Gene Mutation Database (www.hgmd.org).

CLINICAL MANIFESTATIONS

Hemophilia A and hemophilia B are said to be clinically indistinguishable, with clinical severity corresponding inversely to the circulating levels of plasma coagulant factor VIII or IX activity, respectively. However, several studies have found a higher bleeding frequency, greater factor use, and more frequent hospitalizations in hemophilia A, suggesting greater clinical severity than hemophilia B. Individuals with less than 1% of normal factor VIII or IX activity are classified as having "severe" disease, characterized by frequent spontaneous bleeding events in joints (hemarthrosis) and soft tissues and by profuse hemorrhage with trauma or surgery. Although spontaneous bleeding is uncommon in mild deficiencies (>5% normal activity), excess bleeding typically occurs with trauma or surgery. A moderate clinical course is associated with factor VIII or IX levels between 1 and 5%. Approximately 60% of all cases of hemophilia A are clinically severe, whereas only 20 to 45% of cases of hemophilia B are severe.

Severe hemophilia is typically suspected and diagnosed during infancy in the absence of a family history. Among newborns, intracranial hemorrhage is the leading cause of morbidity and mortality, with a cumulative incidence of 3.8%, according to data collected by the Centers for Disease Control and Prevention. Intracranial hemorrhage does not appear to be related to the mode of delivery, although half of such hemorrhages occur in the newborn period. Current guidelines suggest cesarean section delivery be considered for any known male infants with severe hemophilia A. Vacuum extraction may increase cephalohematoma formation and is discouraged. Circumcision within days after birth is accompanied by excessive bleeding in less than half of severely affected boys. The first spontaneous hemarthrosis in severely affected hemophiliacs usually occurs between 9 and 18 months of age, when ambulation begins; in moderately affected individuals, it generally does not occur until 2 to 5 years of age. The knees are the most prominent sites of spontaneous bleeds, followed by the elbows, ankles, shoulders, and hips; wrists are less commonly involved.

Acute hemarthroses (Fig. 174-1) originate from the subsynovial venous plexus underlying the joint capsule and produce a tingling or burning sensation, followed by the onset of intense pain and swelling. On physical examination, the joint is swollen, hot, and tender to palpation, with erythema of the overlying skin. Joint mobility is compromised by pain and stiffness, and the joint is usually maintained in a flexed position. Immediate or early replacement of the deficient clotting factor to normal hemostatic levels rapidly reverses the pain. Delayed treatment results in excess pain, morbidity, and joint damage. Optimal management includes *r*est, *i*ce, factor *c*oncentrate, and *e*levation (RICE). Swelling and joint immobility improve as the intra-articular hematoma resolves. Intra-articular needle aspiration of fresh blood is not recommended because of the risk for introducing infection. Short courses of oral corticosteroids may be helpful to reduce the acute joint symptoms in children but are rarely used in adults.

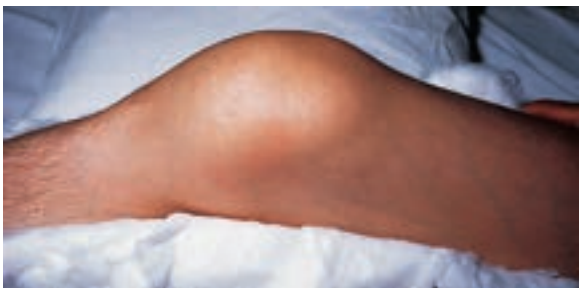

FIGURE 174-1. Acute hemarthrosis of the knee is a common complication of hemophilia. It may be confused with acute infection unless the patient's coagulation disorder is known because the knee is hot, red, swollen, and painful. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine*, 3rd ed. London: Mosby; 2003.)

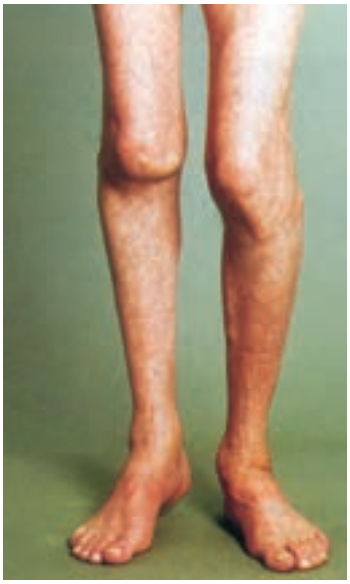

FIGURE 174-2. Severe chronic arthritis in hemophilia. The knee is the most commonly affected joint. Both knees are severely deranged in this patient. Note that he is unable to stand with both feet flat on the floor. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine*, 3rd ed. London: Mosby; 2003.)

Recurrent or untreated bleeds result in chronic synovial hypertrophy and, eventually, damage to the underlying cartilage, with subsequent subchondral bone cyst formation, bony erosion, and flexion contractures. Abnormal mechanical forces from weight bearing can produce subluxation, misalignment, loss of mobility, and permanent deformities of the lower extremities (Fig. 174-2). These changes are accompanied by chronic pain, swelling, arthritis, and disability. Plain radiographs and clinical examination of chronic hemarthroses often underestimate the extent of bone and joint damage; serial magnetic resonance imaging is superior to radiography or computed tomography and is the most sensitive and specific means of detecting and monitoring early and progressive disease.

Intramuscular hematomas account for about 30% of hemophilia-related bleeding events and are rarely life threatening. They are usually precipitated by physical or iatrogenic trauma (e.g., after intramuscular injections) and may compromise sensory and motor function or arterial circulation if they entrap and compress vital structures in closed fascial compartments. The latter occurrence, termed *compartment syndrome*, presents with the rapid onset of swelling and severe pain in an extremity, unrelieved by factor infusion and standard analgesics. This is considered a medical emergency and may require fasciotomy to preserve tissue and provide pain relief. Retroperitoneal hematomas may be confused clinically with appendicitis or hip bleeds but should be suspected in a patient who is hunched over and unable to stand erect owing to the pain of muscle extension in the presence of hematoma. Unless these bleeding episodes are treated immediately and aggressively, permanent anatomic deformity, such as flexion contracture, nerve damage, or pseudotumor formation (expanding hematomas that erode and destroy adjacent skeletal structures), may occur. Bleeding from mucous membranes is very common and may be exaggerated by the degradation of fibrin clots by fibrinolytic enzymes contained in secretions. Bleeding involving the tongue or the retropharyngeal space may rapidly produce life-threatening compromise of the airway. Gastrointestinal hemorrhage typically originates from anatomic lesions proximal to the ligament of Treitz and may be exacerbated by esophageal varices secondary to cirrhosis and portal hypertension or by the use of nonsteroidal anti-inflammatory drugs (NSAIDs) for the treatment of hemarthroses. Spontaneous bleeding in the genitourinary tract secondary to hemophilia is a diagnosis of exclusion after ruling out renal stones and infection. Ureteral blood clots produce renal colic, which may be confused with nephrolithiasis and may be worsened by the use of antifibrinolytic agents. A short course of steroids may be helpful, especially in children, to hasten their resolution. Ninety percent of hemophiliacs experience at least one episode of gross hematuria or hemospermia.

Intracranial bleeds occur in 10% of patients, are usually induced by trauma, and may be fatal in 30% of cases. The risk for developing an intracranial hemorrhage is approximately 2% per year. Neuromuscular defects, seizure disorders, and intellectual deficits may ensue.

Individuals with hemophilia cared for at HTCs have lower mortality and reduced costs of care compared with those receiving treatment elsewhere. The chronic care model practiced at HTCs emphasizes prevention to reduce joint disease and complications, optimization of factor dosing, and counseling regarding safe sports and the avoidance of aspirin and other drugs that inhibit platelet function.

DIAGNOSIS

The diagnosis of hemophilia[1] is suspected on the basis of a family and personal bleeding history and laboratory detection of prolongation of the aPTT (with normal PT). It is confirmed by significantly reduced plasma factor VIII or IX activity. As noted in Chapter 171, the aPTT is not a sufficiently sensitive screening test to be prolonged in mild hemophiliacs, in whom the factor VIII level is sometimes greater than 30% of normal. Severe hemophilia is usually recognized in infancy, with circumcision bleeding; by contrast, moderate or mild disease is recognized later in life after trauma or surgery. Hemophilia can be distinguished from von Willebrand disease (VWD; Chapter 173) by normal ristocetin cofactor and von Willebrand factor (VWF) antigen levels. In type 2N VWD, factor VIII is significantly lower than ristocetin cofactor and VWF antigen levels because of reduced factor VIII binding; this variant of VWD may require genotyping to distinguish it from hemophilia A. Other congenital intrinsic factor deficiencies (e.g., factor XI and XII) can be determined by coagulation factor–specific assays. Vitamin K deficiency (see later and Chapter 175) can be detected by the associated PT prolongation; deficiencies of factors II, VII, IX, and X; and resolution of the coagulation defect with vitamin K. The presence of heparin can be confirmed by correction of the aPTT after running the sample over a heparin absorption column. Failure of the aPTT to correct in a 1 : 1 mix with normal plasma suggests the presence of an inhibitor; specific inhibitors are associated with a single decreased factor level (usually factor VIII; see Alloantibody Inhibitors to Factors VIII and IX under Treatment), whereas blocking inhibitors cause nonspecific factor level changes associated with a positive hexagonal lipid assay (see Antiphospholipid Syndrome and Lupus Anticoagulant later).

Although existing laboratory tests quantify the amount of factor in plasma, which is useful diagnostically and prognostically, these assays are limited in their ability to fully evaluate a patient's clot-forming capability. Newer, so-called global assays have the potential to more objectively measure the hemophilic phenotype as well as the response to treatment, especially in patients who develop inhibitors and in those for whom traditional coagulation tests cannot fully measure laboratory response to bypassing agents (see later); global assays such as thrombin-generation tests and viscoelastic assays await full validation in clinical practice.[2]

TREATMENT

Rx

Treatment and prevention of acute bleeding events in hemophilia A and B are based on replacement of the missing or deficient clotting factor protein to restore adequate hemostasis (Table 174-1). The morbidity, mortality, and overall cost of care for individuals with hemophilia are reduced significantly if care is provided by comprehensive HTCs that have the multispecialty expertise and laboratory capabilities to coordinate and monitor specific patient needs.

The goal of replacement therapy (Table 174-1) is to achieve plasma factor VIII and IX activity levels of 25 to 30% for minor spontaneous or traumatic bleeds (e.g., hemarthroses, persistent hematuria), at least 50% clotting factor activity for the treatment or prevention of severe bleeds (e.g., major dental surgery, maintenance replacement therapy after major surgery or trauma), and 80 to 100% activity for any life-threatening or limb-threatening hemorrhagic event (e.g., major surgery, trauma). After major trauma or if visceral or intracranial bleeding is suspected, replacement therapy adequate to achieve 100% clotting factor activity should be administered *before* diagnostic procedures are initiated. To calculate the initial dose, plasma factor VIII activity generally increases about 2% (0.02 IU/mL) for each unit of factor VIII administered per kilogram of body weight, and factor IX activity increases about 1% (0.01 IU/mL) for each unit of factor IX administered per kilogram of body weight. Therefore, a 70-kg individual with severe hemophilia A or B (factor VIII or IX activity <1% of normal) who requires replacement to 100% activity for major surgery should initially receive 3500 IU of factor VIII or 7000 IU of factor IX concentrate. The circulating half-lives of factors VIII and IX require subsequent dosing at half the initial dose every 8 to 12 hours and every 18 to 24 hours, respectively. However, this empirical dosing (based on calculations) should be individualized according to the peak recovery increment within 10 to 15 minutes after bolus infusion, as well as trough activity levels. The frequency of repeat dosing is also determined by the rapidity of pain relief, recovery of joint function, and resolution of active bleeding. Replacement is usually maintained

TABLE 174-1 FDA-APPROVED COAGULATION PROTEINS AND REPLACEMENT THERAPIES AVAILABLE IN THE UNITED STATES

COAGULATION PROTEIN DEFICIENCY	INHERITANCE PATTERN	PREVALENCE	MINIMUM HEMOSTATIC LEVEL	REPLACEMENT SOURCES
Factor I (fibrinogen) Afibrinogenemia Dysfibrinogenemia	 Autosomal recessive Autosomal dominant or recessive	 Rare (<300 families) Rare (>300 variants)	50-100 mg/dL	Cryoprecipitate, FFP, fibrinogen concentrate
Factor II (prothrombin)	Autosomal dominant or recessive	1 in 2 million births	30% of normal	FFP, factor IX complex concentrates
Factor V (labile factor)	Autosomal recessive	1 in 1 million births	25% of normal	FFP
Factor VII	Autosomal recessive	1 in 500,000 births	25% of normal	Recombinant factor VIIa (15-20 μg/kg), FFP, factor IX complex concentrates
Factor VIII (antihemophilic factor)	X-linked recessive	1 in 5000 male births	80-100% for surgery/life-threatening bleeds, 50% for serious bleeds, 25-30% for minor bleeds	Factor VIII concentrates (recombinant preferred)
Von Willebrand disease (also see Chapter 173) Type 1 and 2 variants Type 3	 Usually autosomal dominant Autosomal recessive	 1% prevalence 1 in 1 million births	>50% VWF antigen and ristocetin cofactor activity	DDAVP for mild to moderate disease (except type 2B; variable response to 2A); cryoprecipitate and FFP (not preferred, except in emergencies); factor VIII/VWF concentrates, viral attenuated, intermediate purity (preferred for surgery, for disease unresponsive to DDAVP, and for type 3
Factor IX (Christmas factor)	X-linked recessive	1 in 30,000 male births	25-50% of normal, depending on extent of bleeding, surgery	Factor IX concentrates (recombinant preferred); FFP not preferred except in dire emergencies
Factor X (Stuart-Prower factor)	Autosomal recessive	1 in 500,000 births	10-25% of normal	FFP or factor IX complex concentrates
Factor XI (hemophilia C)	Autosomal dominant; severe type is recessive	4% of Ashkenazi Jews; 1 in 1 million general population	20-40% of normal	FFP or factor XI concentrate
Factor XII (Hageman factor), prekallikrein, high-molecular-weight kininogen	Autosomal recessive	Not available	No treatment necessary	—
Factor XIII (fibrin stabilizing factor)	Autosomal recessive	1 in 3 million births	5% of normal	FFP, cryoprecipitate, or viral-attenuated factor XIII concentrate

DDAVP = desmopressin; FDA = U.S. Food and Drug Administration; FFP = fresh-frozen plasma; VWF = von Willebrand factor.

for 10 to 14 days after major surgery to allow proper wound healing. Bolus dosing typically results in wide fluctuations in clotting factor activity levels and requires frequent laboratory monitoring to avoid suboptimal troughs. A continuous infusion regimen, consisting of 1 to 2 IU of factor VIII or IX concentrate per kilogram per hour after a bolus dose, maintains a plateau level without the need for frequent laboratory testing. Continuous infusion also reduces total concentrate consumption by 30 to 75% in surgical settings. Ongoing and completed phase I/II/III clinical trials of long recombinant factors VIII and IX, pegylated or fused to albumin or to the Fc domain of immunoglobulin G1 (IgG1), indicate these proteins are safe, effective, and prolong half-life 1.5- to 2-fold for factor VIII and 2- to 4-fold for factor IX, suggesting they may allow simpler dosing schedules and fewer intravenous factor infusions. A completed phase III clinical trial of recombinant factor IX-Fc fusion protein (rFIXFc) in adults with severe hemophilia B has demonstrated safety, efficacy, and a 3-fold longer circulating half-life than recombinant factor IX, such that some patients were able to dose every 10 to 14 days.[3] Similarly, recombinant factor VIII-Fc (rFVIIIFc) is safe, effective, and has a 1.5-fold longer half-life than recombinant factor VIII, resulting in dosing every 3 to 5 days.

Because of the potential thrombogenicity associated with the repeated administration of prothrombin complex concentrates to replace factor IX deficiency, high-purity, plasma-derived, or genetically engineered factor IX concentrates, which lack activated vitamin K–dependent clotting factors, are preferred in hemophilia B.

Cryoprecipitate (a cold precipitate of fresh-frozen plasma [FFP] after thawing at 4°C) and FFP contain factor VIII, but only FFP contains factor IX. However, neither cryoprecipitate nor FFP is an optimal replacement product for either hemophilia A or hemophilia B because these agents may transmit blood-borne pathogens and require large-volume infusion. All clotting factor concentrates available in the United States (see Table 174-1), whether plasma derived or genetically engineered, are equally efficacious and are considered extremely safe; none has been implicated in the transmission of blood-borne viral pathogens or prions. The second- and third-generation recombinant factor VIII and IX concentrates are manufactured free of added human or animal proteins in the culture medium or in the final formulation, eliminating the theoretical risks for transmission of prions or murine viruses.

Hemarthroses

The moderate or severe pain that accompanies acute hemarthroses responds to immediate analgesic relief, temporary immobilization, restraint from weight bearing, and clotting factor replacement. Narcotic analgesics, such as codeine or synthetic derivatives of codeine, should be prescribed alone or combined with doses of acetaminophen that are low enough to avoid hepatic toxicity in patients with chronic hepatitis. Although these medications do not possess significant anti-inflammatory activity, they are preferable to NSAIDs or aspirin, which can exacerbate bleeding complications through their inhibition of platelet aggregation.

Strategies intended to prevent end-stage joint destruction should be initiated at an early age. Although prophylaxis beginning soon after the first bleed is the recommended approach (see later), most adults have not had the benefit of early prophylaxis and thus have advanced arthropathy with reduced motion, disability, and pain, for which surgery may be recommended. Synovectomy through open surgery or arthroscopy removes the inflamed tissue and should result in substantially decreased pain and less recurrent bleeding. Nonsurgical synovectomy (synoviorthosis), which involves the intra-articular administration of a radioisotope, is particularly useful for high-risk patients and for those with alloantibody inhibitors to factor VIII or IX (see later). The occurrence of leukemia in several hemophilic children undergoing radioisotopic synoviorthosis has raised concerns about potential leukemogenesis, especially given the low background rate of cancers in individuals with hemophilia. Neither synovectomy nor synoviorthosis reverses joint damage, but both procedures may delay its progression. Non-weight-bearing exercises, such as swimming and isometrics, are important to periarticular muscle development and maintenance of joint stability for ambulation. Intractable pain and severe joint destruction secondary to repeated hemorrhage require prosthetic replacement. Chronic ankle pain responds best to open surgical or arthroscopic fixation and fusion (arthrodesis).

The ultimate strategy to minimize or eliminate progressive joint destruction by recurrent hemarthroses is predicated on the concept of prophylaxis—the scheduled administration of clotting factor concentrates several times weekly (twice a week for factor IX, three times a week for factor VIII) at doses to maintain trough factor activity levels greater than 1 to 2% of normal. In a

prospective, randomized clinical trial, the Joint Outcomes Trial, prophylaxis with factor VIII (25 IU/kg every other day) was superior to episodic therapy (on demand) in young children with severe hemophilia in reducing joint bleeding and joint damage as shown by magnetic resonance imaging and radiography.[3] These findings were confirmed in older children in the ESPRIT trial,[3] with the best outcomes when prophylaxis was initiated before 3 years of age. In several adult studies of two- or three-times-weekly prophylaxis at 20 to 80 IU/kg, there were significantly fewer joint bleeds and pain and better quality of life than in those on on-demand therapy.[3] Although prophylaxis uses more factor product at greater expense, the benefits of long-term prophylaxis to promote joint health and avert disability have led to the recommendation that it be initiated in young children with severe disease at the time of the first bleed. Compliance with prophylaxis is not easy because it requires intravenous factor, which is invasive, burdensome, and in small children may require central venous access devices, which may be complicated by infection.

With the development of long-acting factor VIII and IX,[4] currently in clinical trials, infusion frequency and bleeds may decrease and access devices may be avoided entirely. Some unanswered questions include the minimal effective dose for prophylaxis in children, and the role of generic versus personal pharmacokinetic studies to optimize dosing. The risk for bleeding is thought to be related to the time spent at nadir factor levels less than 1% (0.01 IU/mL) between dosing. It is generally agreed that spontaneous bleeding may be averted by maintaining factor levels at 1% (0.01 U/mL) or greater.

Factor Concentrate–Transmitted Viral Infections

In contrast to other at-risk groups, individuals with hemophilia were exposed at a young age to transmissible agents through clotting factors. These include hepatitis C virus (HCV), leading to infection in 90% of those transfused from the late 1970s through the mid-1980s, and human immunodeficiency virus (HIV) infection in 80% of those transfused with factor VIII products and 50% of those transfused with factor IX products from 1978 through the mid-1980s. Overall, of those with HCV infection (Chapter 148), 40% have coinfection with HIV, and hepatitis C remains the leading cause of death in hemophilia. With the implementation of viral inactivation and recombinant technologies, viral transmission has been virtually eliminated in hemophiliacs born since the 1990s. Nevertheless, among those exposed to HCV, 25% have liver fibrosis, and those with HIV coinfection have a 1.4-fold greater fibrosis rate. Combined antiretroviral therapy (cART) has slowed the rate of HCV progression in coinfected patients to rates observed in HCV-monoinfected hemophilic men.

GB virus C (GBV-C), formerly known as hepatitis G virus, observed in 15 to 25% of hemophiliacs, is susceptible to current viral attenuation procedures. Hepatitis A and B vaccination have rendered viral infections with these agents rare in those with hemophilia, and although parvovirus B19 (Chapter 371) seroprevalence approaches 80% in older adult hemophiliacs exposed to plasma-derived products, the long-term clinical consequences remain unclear. Cadaver and living-donor liver transplantation (Chapter 154) has improved the survival of hemophilic men with chronic hepatitis-induced liver failure and has also resulted in the phenotypic cure of hemophilia, confirming that the liver is the predominant source of normal synthesis of factors VIII and IX. Liver transplantation is also performed successfully in HIV/HCV-coinfected patients on cART therapy, although HCV recurrence remains a universal problem.

Ancillary and Other Therapies

Ancillary treatment strategies for the hemophilias include antifibrinolytic agents, such as ε-aminocaproic acid (50 mg/kg 3 to 4 times daily) or tranexamic acid (3 or 4 g given orally daily in divided doses), to minimize mucous membrane bleeding, and the application of fibrin glue to bleeding sites. Desmopressin (DDAVP), which is also used in VWD (Chapter 173), can be administered by nasal insufflation 2 hours before a scheduled surgical procedure (one spray per nostril, to provide a total dose of 300 µg; or, in patients weighing less than 50 kg, 150 µg administered as a single spray); alternatively, DDAVP can be administered intravenously (dissolved in 50 mL normal saline) over 30 minutes at a dose of 0.3 µg/kg. DDAVP is useful in patients with mild hemophilia A because an adequate incremental rise in factor VIII activity can circumvent the use of clotting factor concentrates. Repeated administration of DDAVP (intravenously or by intranasal spray) may be complicated by facial flushing, tachyphylaxis, hyponatremic seizures (primarily in children), and, rarely, angina.

Alloantibody Inhibitors to Factors VIII and IX

Alloantibodies—that is, antibodies to "foreign" infused factor VIII or, less frequently, factor IX—are usually detected in childhood after a median of 9 to 12 days of exposure to clotting factor. These inhibitors occur preferentially in those with a family history of inhibitors; those with large, multidomain factor VIII and factor IX gene deletions; blacks; and Hispanics. Among blacks, gene sequence data suggest that mismatched factor VIII transfusions may account for the high inhibitor risk. Although hemophiliacs with major deletions of the factor VIII gene have a very high (up to 90%) prevalence of inhibitors, its prevalence in those with factor VIII gene missense mutations is low (<10%), as it is in those with intron 22 inversion (20%) even though the latter have severe hemophilia. Algorithms have been developed to stratify inhibitor risk for individuals and subpopulations. The incidence of factor VIII alloantibodies among hemophilia A patients is 15 to 25%, whereas the incidence of factor IX alloantibodies among hemophilia B patients is 1.5 to 3%. The latter is most common in Scandinavians and is also associated with anaphylaxis and nephrotic syndrome. Increasing evidence from the RODIN study suggests that although inhibitor risk appears to be unrelated to factor type (e.g., recombinant vs. plasma-derived), it is significantly associated with early high-intensity factor exposure (i.e., >3 to 5 days), especially in those with a family history or an at-risk mutation (e.g., a large deletion mutation).[5,6]

The development of an alloantibody inhibitor is suspected when replacement therapy is ineffective in controlling bleeding symptoms. These antibodies, typically of the IgG4 subclass, completely neutralize clotting factor activity and prevent or reduce any increment in factor VIII or IX levels following bolus infusions of concentrate. Characterized as time and temperature dependent, inhibitors are quantitated in Bethesda units (BU): by definition, 1 BU is the amount of inhibitor that neutralizes 50% of the specific clotting factor activity in normal plasma. "High responders," or patients with high-titer inhibitors (>5 BU), mount an anamnestic antibody response to factor VIII clotting factor protein, usually within 5 to 7 days after subsequent exposure, and are no longer responsive to infused factor VIII. By contrast, "low responders," or patients with low-titer inhibitors (≤5 BU), do not manifest anamnesis, and such low-titer inhibitors can easily be overwhelmed by large amounts of human factor VIII or factor IX concentrate and can be successfully treated with three to four times the usual factor dose.

Management of patients with high-titer inhibitors against factor VIII or IX is difficult, and no single approach is uniformly successful.[7] There are two components: first, assurance of hemostasis using "bypass therapy," and second, eradication of inhibitor formation. Hemostasis can be provided by "bypass" agents that are used to treat bleeding episodes (see Table 174-1); specifically, the activated prothrombin complex concentrate FEIBA VH (75 to 100 IU/kg initially, then 50 to 100 IU/kg every 6 to 8 hours) and recombinant factor VIIa (90 µ/kg every 2 to 3 hours) can be administered until bleeding is controlled. In studies of congenital hemophilia A patients with alloantibody inhibitors, one dose of FEIBA VH or two doses of recombinant factor VIIa controlled hemarthrosis episodes 81 and 79% of the time, respectively. The activated and unactivated prothrombin complex concentrates contain activated vitamin K–dependent clotting factors that "bypass" the intrinsic pathway (factor VIII or IX) inhibitor. As a result, repeated administration over a short time may be complicated by potential thrombogenicity: the aPTT and clotting factor assays are not helpful in monitoring hemostasis in these cases. In patients with high-titer inhibitors to factor VIII or IX, recombinant factor VIIa may achieve effective hemostasis, but its use is limited by the need for frequent intravenous dosing, usually every 2 hours to start. Although continuous dosing has been used in some patients, general experience dictates that for surgery or procedures in which significant bleeding is likely, bolus dosing is required to achieve optimal hemostasis—a so-called "thrombin burst." The product has also been effective in patients who experience anaphylactic reactions or nephrotic syndrome after exposure to factor IX–containing replacement products or fresh-frozen plasma (FFP). Several studies have shown that prophylactic treatment in adult inhibitor patients with either rFVIIa, 90 µg/kg daily or 270 µg/kg daily, or FEIBA, 85 IU/kg weekly, can significantly reduce bleeding rates compared with on-demand treatment.[3]

Eradication of inhibitors is usually attempted with "immune tolerance induction" regimens,[8] which are generally effective if initiated within 12 months of the inhibitor's detection. Tolerance regimens consist of daily doses of factor concentrates to accomplish desensitization to infused factor, a process associated with a 68% success rate. In an international randomized trial of high-dose factor VIII (200 IU/kg daily) vs. low dose (50 IU/kg three times weekly) for immune tolerance induction (ITI) in young inhibitor patients, there was no difference in inhibitor eradication, but there was greater bleeding (mostly hematomas) in the low-dose arm.[3] These findings suggest improved hemostasis may be possible in the high-dose arm, despite inhibitor detection. The authors also initiated ITI only after waiting for the inhibitor titer to drop below 10 BU because tolerance took longer to achieve the higher the inhibitor titer.[3] After tolerance was achieved, prophylaxis with factor VIII or IX concentrate two or three times weekly (at 20 to 30 IU/kg) appears necessary to maintain tolerance. Individuals with inhibitors since childhood for whom immune tolerance was not possible before adulthood are unlikely to respond to ITI. Alternative single or combination immunosuppressive agents (e.g., rituximab, mycophenolate, or cyclosporine) may be used as alternative therapies but appear to be variably effective.[9]

PREVENTION

Carrier Detection and Prenatal Diagnosis

Carrier detection and prenatal diagnosis have become technically feasible, very sensitive, and widely available. Noninvasive prenatal diagnosis by use of

microfluidic digital PCR of maternal plasma DNA is an emerging technique that is based on relative mutation dosage from a blood sample of a pregnant carrier with or without an affected fetus, which may detect hemophilia as early as the 11th week of gestation. The application of these diagnostic tools, however, is influenced by ethical, cultural, religious, economic, educational, and personal considerations. For instance, carrier detection is particularly useful in identifying women who may be at risk for hemorrhagic complications during the delivery process and male offspring who are particularly vulnerable to intracerebral bleeds at birth. Alternatively, these techniques can provide important information used to make difficult reproductive decisions. Genetic testing should be performed only following genetic counseling and with no patient coercion to accept such testing.

PROGNOSIS

The life expectancy of individuals with severe hemophilia is approaching that of the normal population. The age-matched death rate in hemophilia is 2.7-fold greater than that in the general population, although ischemic heart disease mortality is nearly 60% lower than in the general population. Life expectancy is related to the severity of hemophilia: the mortality rate of severely affected patients is 4- to 6-fold greater than that of patients with mild deficiency. Among those with alloantibody inhibitors, mortality rates are significantly higher than in noninhibitor patients. The three leading causes of death are hepatitis C, HIV/AIDS, and central nervous system bleeding. Hepatitis C, the leading cause of death, accounts for more than half of the deaths, whereas deaths from HIV have declined with the availability of cART therapy (Chapter 388). Among those with HIV/HCV coinfection, cART also slows the progression of HCV-related liver disease. Predictors of hepatitis C disease progression in hemophilia include alcohol use, the use of acetaminophen for pain relief of hemophilic arthropathy, hepatitis B surface antigenemia, and HIV coinfection. Bleeding also accounts for causes of death: the lifetime risk for intracranial hemorrhage is 2 to 8%, and although it is the leading cause of morbidity and mortality in the newborn period, prospective monitoring of central nervous system function continues to be a priority as the population ages. Finally, with the essential elimination of disease transmission through blood products in this population, the usual problems of aging are increasingly being recognized in those with hemophilia, including atherosclerosis, hypertension, hyperlipidemia, obesity, and diabetes. The impact of these conditions on the natural history of hemophilia, given the recognized lower mortality from ischemic heart disease, remains to be seen. More data are needed to help guide clinical management of these patients.

FUTURE DIRECTIONS

Gene Therapy for Hemophilia a and B

The hereditary hemophilias are model diseases for gene therapy because they are caused by specific, well-defined gene mutations; a small, incremental rise in clotting factor synthesis can lead to substantially improved treatment and quality of life; and inadvertent overexpression by successful gene transfer would not be detrimental. Successful gene transfer techniques have been developed to provide long-term therapeutic benefits in hemophilic mice and dog models, and early success is now being reported in humans. In one study of gene transfer using serotype-8 pseudotyped adeno-associated virus vector (AAV8) carrying a codon-optimized factor IX transgene, factor IX expression of 1 to 6% was achieved in hemophilia B patients, persisting up to the present, up to 2 years later.[10] An immune response to the AAV capsid, associated with increases in alanine transaminase (ALT) and aspartate transaminase (AST) in some of the subjects, resolved with a 4-week course of oral steroids, but approaches to avert this immune response are underway, including modification to neutralize excess AAV. Other approaches include design of less immunogenic vectors and use of innovative delivery systems, such as platelets. Novel therapeutic approaches to enhance clotting factor replacement therapy include RNAi silencing of antithrombin III, a thrombin inhibitor, to promote hemostasis through thrombin generation; oral delivery systems using bioencapsulated proteins; and molecular modifications to enhance the desirable properties of clotting factor.

Acquired Hemophilias

EPIDEMIOLOGY AND PATHOBIOLOGY

Autoantibody inhibitors occur spontaneously in individuals with previously normal hemostasis (nonhemophiliacs). In contrast to alloantibody inhibitors in hemophilic men, which are directed against foreign infused clotting factor, autoantibody inhibitors are directed against a "self" clotting factor, most commonly factor VIII. These autoantibodies typically arise in individuals with no past bleeding history; thus the diagnosis may be missed until a prolonged aPTT and aPTT mix tests are obtained (Chapter 171). Although half of those with autoantibody inhibitors have no obvious underlying cause, autoimmune diseases, lymphoproliferative disorders, idiosyncratic drug reactions, pregnancy, and advanced age are associated in the other half.[11]

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Massive hemorrhagic events, even more severe than in hemophilia patients with alloantibodies, and more commonly in soft tissues, may occur in those with autoantibodies because of a delay in diagnosis and treatment. The laboratory expression of autoantibodies is similar to that of alloantibodies, except that clotting factor activity is not completely neutralized. Residual clotting factor activities between 3 and 20% of normal are frequently observed in patients with autoantibodies.

TREATMENT

Rx

The same principles of replacement therapy for alloantibodies apply to these acquired autoantibody inhibitors.[12] There are two goals of treatment: (1) stop the bleeding and assure hemostasis, and (2) eradicate the inhibitor. For hemostasis, recombinant factor VIIa, activated prothrombin complex concentrate (aPCC), or factor VIII bypass activity is commonly used, in similar doses used to achieve hemostasis with alloantibody. Data from the EACH2 registry (European Acquired Haemophilia Registry) indicate that optimal bleeding control is with bypass treatment (rFVIIa or aPCC), which was effective in 93%, compared with FVIII or DDAVP, 68%. Thrombotic events were reported in only 3.6% for all of these hemostatic agents combined. Eradication of the inhibitor was significantly better in those treated with steroids and cyclophosphamide (80%), than those treated with steroids alone (58%); and the combination is more effective in autoantibody patients than in alloantibody patients; this includes corticosteroids (prednisone 1 mg/kg/day orally), cytotoxic agents (e.g., cyclophosphamide 150 mg/day orally or 500 to 750 mg/m^2 intravenous bolus every 3 to 4 weeks), or a combination, with dose titration based on inhibitor levels and complicating cytopenias. Rituximab (anti-CD20 antibody)-based regimens (375 mg/m^2 intravenously weekly for 4 weeks) are also effective in 61%. High-dose intravenous gamma globulin (IVIG)-based regimens may be effective in 45%. These agents are tapered and discontinued after the autoantibody has disappeared. The time to complete remission is about 5 weeks for steroids with or without cyclophosphamide, whereas rituximab-based regimens require twice as long to achieve remission.

PROGNOSIS

Several large series of patients with acquired hemophilia reveal a substantial mortality rate of 15 to 25%, which is considerably higher than that observed with alloantibody factor VIII inhibitors. A large meta-analysis found that overall survival in acquired hemophilia was influenced primarily by the achievement of a complete remission, age younger than 65 years at diagnosis, and related diseases (malignancy vs. postpartum vs. others). As many as 17% of the deaths were associated with sepsis, and 71% of those arose as a complication of cyclophosphamide-induced neutropenia. Hemorrhagic complications were the primary cause of death, but these could be reduced if the inhibitor could be eradicated.

Von Willebrand Disease

The most common congenital bleeding disorder is von Willebrand disease (VWD). This disorder is inherited in an autosomal dominant fashion and affects both sexes, with a prevalence of 1 to 3% and no ethnic predominance. Homozygous patients are rare and carry a recessive mutant gene. VWF, the protein that is decreased or defective in VWD, is a large, multimeric glycoprotein encoded by the *VWF* gene, located on chromosome 12. A personal history of mucocutaneous bleeding, a family history, and decreased functional VWF constitute a diagnostic triad. Treatment is accomplished with DDAVP or VWF concentrates. Detailed discussion of VWD is provided in Chapter 173.

Factor XI Deficiency

EPIDEMIOLOGY

Factor XI deficiency has a prevalence of 1 in 1 million in the general population and 1 in 500 births in Ashkenazi Jewish families.[13] Factor XI is the only component of the contact phase (factor XII, prekallikrein, and high-molecular-weight kininogen) of the intrinsic pathway of coagulation

that is associated with excessive bleeding complications when a deficiency exists.

PATHOBIOLOGY

Factor XI deficiency is predominantly an autosomal recessive trait, although some mutations may have a dominant transmission pattern. The factor XI gene (*FXI*) is located on chromosome 4, and the protein circulates as a homodimer, with each FXI monomer composed of 4 apple domains encoded by exons 3 through 10 and a protease domain encoded by exons 11 through 15. The Glu117 stop mutation in *FXI* is the most common cause of factor XI deficiency; it is secondary to poor secretion or stability of the truncated protein or decreased levels of messenger RNA. To date, more than 170 mutations of *FXI* have been identified in factor XI–deficient patients (see www.factorxi.org), but close correlation between hemorrhagic phenotype and genotype is lacking. Most are missense or nonsense mutations and are located across all 4 apple domains and the serine protease region. In Ashkenazi Jewish individuals, factor XI deficiency is common, with a heterozygote frequency of 8 to 10%; two predominant gene mutations occur with equal frequency and are designated type II (a stop codon in exon 5) and type III (a single base defect in exon 9). The most severe clinical disease is observed in patients homozygous for type II, who usually have less than 1% factor XI activity. Homozygous type III individuals also manifest severe symptoms, but typically less severe than those of type II patients; they have slightly higher factor XI levels of about 10 to 20%. Compound heterozygotes, type II/III, make up the bulk of factor XI–deficient patients; they have clinically mild disease, with factor XI levels between 30 and 50%. In non-Jewish individuals, the mutations are more variable, although Cys128 stop has been described in several kindreds, and Cys38Arg has been found in several French Basque families. One third of those who develop inhibitors are homozygous for Glu117 stop, which results in an absent *FXI*. Genotypic identification of affected patients is determined by measuring factor XI levels rather than by defining the specific gene defect.

CLINICAL MANIFESTATIONS

The clinical bleeding tendency in factor XI deficiency is less severe than that observed in severe hemophilia A or B and, in contrast to the hemophilias, does not correlate with the severity of the deficiency. Most individuals with less than 20% of normal factor XI activity experience excessive bleeding after trauma or surgery; however, a few do not bleed. In contrast, bleeding has been observed in approximately 35 to 50% of mildly affected patients with factor XI levels between 20 and 50% of normal. Spontaneous hemorrhagic episodes, hemarthroses, and intramuscular and intracerebral bleeds are unusual; traumatic and surgical bleeds typically involve the mucous membranes. Patients undergoing tonsillectomy, prostatectomy, or dental extraction are at highest risk for bleeding unless replacement therapy is administered. Women may experience significant menorrhagia, and it has been recommended that women with menorrhagia be screened for both VWD and factor XI deficiency. Patients with mild factor XI deficiency and coincident mild VWD have an increased risk for bleeding.

DIAGNOSIS

Factor XI deficiency is diagnosed in the laboratory by a prolonged aPTT, normal PT, and decreased factor XI activity ascertained in a specific quantitative clotting assay (normal range, 60 to 130%).

TREATMENT Rx

Not all individuals with factor XI deficiency bleed, so it may be reasonable to monitor with no treatment, especially if there is no family history of bleeding. For surgical or other major bleeding, FFP 15 to 20 mL/kg may be given, although potential complications include fluid overload and infection risk, which can be reduced with the use of pathogen-inactivated FFP, if available. Use of factor XI concentrate, which is available in Europe but not the United States, may be complicated by thrombosis, which occurs in approximately 10% of patients, particularly in older individuals with preexisting cardiovascular disease or malignancy. Replacement dosing levels should not exceed 70% of factor XI activity. Repeat dosing with FFP or factor XI concentrate should take into consideration the long (60- to 80-hour) biologic half-life of factor XI in vivo.

The decision to treat heterozygotes with factor XI at levels greater than 20% is empirical and should be based on individual history of bleeding after trauma or surgery. There is no clear evidence of benefit with the use of DDAVP. Because hemorrhagic complications originate most commonly from mucous membrane surfaces, antifibrinolytic agents such as ε-aminocaproic acid or tranexamic acid are frequently helpful as adjunctive therapy. In women with menorrhagia or postpartum hemorrhage, testing for VWD is recommended because both diseases may be present.

Alloantibody inhibitors, which neutralize the hemostatic effects of exogenously administered factor XI replacement, can develop in patients with severe factor XI deficiency who have been exposed to plasma or factor XI concentrate. Recombinant factor VIIa can prevent bleeding during or after surgery in these patients.

Deficiencies of Contact Activation Factors

Although factor XI plays an important role in the activation of factor IX in the intrinsic pathway generation of thrombin, it is only one of the four components of the contact phase of coagulation. Deficiencies in any of the other three factors (factor XII, prekallikrein, and high-molecular-weight kininogen) produce in vitro laboratory abnormalities. Even among those with severe factor XII deficiency (<1% activity), there is no clinical bleeding; however, up to 8 to 10% with severe factor XII deficiency actually experience venous thromboembolic events, which are occasionally fatal. This finding has led to speculation that factor XII deficiency may lead to hypercoagulability through defective participation of the contact phase proteins in the activation of fibrinolysis.

DIAGNOSIS

Deficiencies of each of these factors prolong the aPTT, often markedly, which may normalize after prolonged incubation of the patient's plasma at 37° C with a negatively charged activator of the aPTT assay (i.e., kaolin or celite). Specific assays are also available to quantitate each of the contact factors.

TREATMENT Rx

Deficiencies of the contact activation factors, however severe they are and however prolonged the associated aPTT may be, do not cause clinical bleeding problems, even in response to surgery or trauma. Therefore, no therapy is indicated for factor XII deficiency, prekallikrein deficiency, or high-molecular-weight kininogen deficiency. Routine anticoagulation regimens are used to treat thrombogenic events.

Factor XIII (Fibrin-Stabilizing Factor) Deficiency

Factor XIII is a transglutaminase that is activated by thrombin and functions to cross-link fibrin to protect it from lysis by plasmin. It is also involved in wound healing and tissue repair and seems to be crucial for maintaining a viable pregnancy. Homozygous severe deficiency states are rare and are inherited in an autosomal recessive manner, with a prevalence of 1 per 3 million births. Consanguinity is common. Most patients are deficient in the A subunit (FXIII-A). Half of the molecular defects that account for A-subunit deficiency are missense mutations.

CLINICAL MANIFESTATIONS

Heterozygous carriers are usually asymptomatic, but homozygotes have lifelong bleeding that typically starts shortly after birth with persistent bleeding around the umbilical stump. Intracranial bleeding events, usually precipitated by minimal trauma, occur commonly enough in infants (25%) to justify the initiation of a primary prophylaxis regimen of replacement therapy. Delayed bleeding after surgery and trauma is the hallmark of the disease; however, easy bruising, poor wound healing with defective scar formation and dehiscence, and hemarthroses are characteristic. Spontaneous abortions are increased in severely affected women.

DIAGNOSIS

The diagnosis is usually suspected on clinical grounds, given that factor XIII deficiency is not detected by conventional screening coagulation assays (i.e., aPTT, PT). Most laboratories use a rapid screening assay that assesses the ability of a fibrin clot to remain intact with incubation in 5 mol/L of urea or 1% monochloroacetic acid. With factor XIII levels less than 1% of normal, the clot dissolves within 2 to 3 hours.

TREATMENT Rx

Replacement therapy for prophylaxis or the treatment of acute bleeds in factor XIII deficiency can be accomplished by administering cryoprecipitate, FFP, or, preferably, plasma-derived factor XIII concentrate (Corifact, CSL Behring), which is pasteurized for viral safety. Normal hemostasis is achieved with a factor XIII level of only 5% of normal. The dose is 10 to 20 U/kg intravenously, and because it has a long half-life (10 days), prophylactic replacement can be given every 3 to 4 weeks. Acquired alloantibody inhibitors can develop in severely affected individuals. Autoantibodies also occur, usually in association with systemic lupus erythematosus. Preliminary data from an ongoing clinical trial of recombinant factor XIII indicates it is safe and effective in preventing bleeds in factor XIII–deficient patients.[14]

Dysfibrinogenemia and Afibrinogenemia

Approximately 300 abnormal fibrinogens have been described, but few cause hemostatic symptoms. Abnormal fibrinogens are rare autosomally inherited proteins. Quantitative fibrinogen deficiencies (afibrinogenemia and hypofibrinogenemia) may result from mutations affecting fibrinogen synthesis or processing, whereas qualitative defects (dysfibrinogenemia) are caused by mutations that lead to abnormal polymerization, defective cross-linking, or defective assembly of the fibrinolytic system.

CLINICAL MANIFESTATIONS

More than 50% of the dysfibrinogenemias are asymptomatic, 25% are associated with a mild hemorrhagic tendency (commonly caused by defective release of fibrinopeptide A), and 20% predispose individuals to thrombophilia (usually caused by impaired fibrinolysis) (Chapter 176). Concurrent bleeding and thrombosis also may occur. The prevalence of dysfibrinogenemia in patients with a history of thromboembolic episodes approaches 0.8%. A high prevalence of dysfibrinogenemia has been reported among patients with chronic thromboembolic pulmonary hypertension, implicating changes in the molecular structure of fibrin in the development of this disorder (Chapter 68). Women experience a high incidence of pregnancy-related complications, such as spontaneous abortion and postpartum thromboembolic events. Thrombin times and reptilase times (plasma-based clotting times with the substitution of reptilase snake venom for thrombin) are not helpful in predicting whether an abnormal fibrinogen will be prothrombotic, prohemorrhagic, or asymptomatic. However, clinical history, fibrinopeptide release studies, and fibrin polymerization studies may be useful. Clinically insignificant dysfibrinogenemias may be acquired in association with hepatocellular carcinoma (Chapter 196).

In contrast to the hepatic synthesis of a qualitatively abnormal protein in dysfibrinogenemia, congenital afibrinogenemia, an autosomal recessive disorder, represents the markedly deficient production of a normal protein. Severe life-threatening hemorrhagic complications can occur at any site, beginning at birth with umbilical bleeding. Intracranial hemorrhage is a frequent cause of death. Poor wound healing is characteristic. All coagulation-based assays that depend on the detection of a fibrin clot end point are markedly prolonged. Afibrinogenemia is usually detectable by specific functional or immunologic assays. Platelet dysfunction may accompany afibrinogenemia and exacerbate bleeding.

DIAGNOSIS

Abnormalities are usually detected incidentally when routine coagulation screening assays reveal decreased fibrinogen concentrations and prolonged thrombin clotting times. On further evaluation, discordance between functional and immunologic fibrinogen levels (>50 mg/dL more antigenic than functional) is observed; clotting times using snake venom (reptilase or ancrod) are variably prolonged.

TREATMENT Rx

Deficiencies of fibrinogen can be corrected by the administration of FFP, cryoprecipitate, or a viral-attenuated (pasteurized), plasma-derived fibrinogen concentrate (Riastap, CSL Behring). The target replacement goal is a plasma level of 100 mg/dL, and given a circulating biologic half-life of at least 96 hours, treatment every 3 to 4 days is adequate. Primary prophylaxis regimens may be useful in afibrinogenemia, with on-demand or prophylactic replacement for trauma or surgery. Individuals with thrombophilic manifestations should receive anticoagulation long term, depending on risk-benefit assessment (Chapter 176). Riastap is not recommended for use in dysfibrinogenemia. Solvent/detergent plasma (Octaplas, Octapharma), a pooled human plasma that reduces allergic reactions and transfusion-related acute lung injury (TRALI), or psoralen-treated FFP (when licensed by the U.S. Food and Drug Administration [FDA]) may provide an alternative therapy.

Factor V Deficiency

Factor V is a component of the prothrombinase complex that assembles factors Va and Xa on the phospholipid membrane of the platelet for prothrombin (factor II) activation to thrombin (Chapter 171).

CONGENITAL FACTOR V DEFICIENCY

Deficiency of factor V is a rare, autosomal recessive disorder (1 in 1 million births). The factor V Leiden protein, which is responsible for resistance to activated protein C and thrombophilia, does not affect factor V coagulant activity (Chapter 176). The severity of plasma factor V deficiency correlates less well with the risk for clinical bleeding than does the factor V content in platelet α-granules (which cannot be measured in the clinical laboratory). This observation illustrates the critical role of the platelet in promoting adequate hemostasis at bleeding sites and explains why transfusions of normal platelets may be preferred over FFP for the treatment of hemorrhagic episodes secondary to congenital or acquired factor V deficiency. Hemostasis can be maintained without correcting plasma factor V activity (to >25% of normal).

COMBINED DEFICIENCIES OF FACTORS V AND VIII

Factors V and VIII are structurally homologous proteins, and combined deficiencies of these factors occur as an autosomal recessive disorder with a prevalence of 1 in 100,000 births among Jews of Sephardic origin. The severity of bleeding is determined by the levels of these factors, which usually range from 5 to 30% of normal. Replacement therapy should be aimed at normalizing both clotting protein activities.

ACQUIRED FACTOR V DEFICIENCY

Acquired factor V deficiency has been described in individuals exposed to bovine factor V, which contaminates the thrombin preparations used topically to control bleeding during cardiovascular surgery. This abnormality probably represents the development of anti–bovine factor V antibodies that cross-react with the human factor V protein. Profuse bleeding accompanies this complication.

Deficiencies of Vitamin K–Dependent Coagulation DEFICIENCIES OF FACTORS II, VII, AND X

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

The coagulopathies of vitamin K deficiency and liver failure are discussed in Chapter 175. Congenital deficiencies of factors II (prothrombin), VII, and X are rare autosomally inherited disorders. Heterozygotes (with factor levels approximately 20% of normal) are typically asymptomatic except in the immediate newborn period, when physiologic vitamin K deficiency exacerbates the underlying clotting factor deficiency. Homozygotes with clotting factor levels less than 10% of normal manifest variable symptoms. As with other coagulopathies, these deficiencies are usually suspected after neonatal umbilical stump bleeding. Thereafter, unless replacement or prophylactic therapy is provided, these patients are subject to mucosal bleeding from epistaxis, menorrhagia, and dental extractions; hemarthroses and intramuscular hematomas; and bleeding after surgery or trauma.

Acquired factor VII deficiency has been associated with Dubin-Johnson and Gilbert syndromes (Chapter 147). Acquired factor IX deficiency has been associated with Gaucher disease because factor IX binds to glucocerebroside (Chapter 208). Acquired factor X deficiency and amyloidosis are discussed later.

DIAGNOSIS

In the coagulation laboratory, factor VII deficiency is associated with a prolonged PT and a normal aPTT. This pattern localizes the deficiency to the extrinsic pathway. In contrast, deficiencies of factors II (prothrombin) and X prolong both the PT and the aPTT, with the defects localized to the common pathway of coagulation. A Russell viper venom–based clotting assay can

differentiate between these two deficiencies; as a direct activator of factor X, the assay is prolonged with factor X deficiency but not with factor II deficiency. Mixing patient plasma with normal plasma results in a correction of these assays, and specific clotting assays using plasma deficient in the coagulation protein to be studied can confirm the diagnosis.

TREATMENT Rx

Replacement therapy is indicated for acute symptomatic bleeds and for prophylaxis before surgery. In addition to FFP, which has the potential to transmit blood-borne viruses, factor IX complex concentrates can be administered to achieve hemostatic levels of any of these vitamin K–dependent factors (to >25 to 30% of normal).

Bleeding complications caused by acquired IgG autoantibodies directed against any coagulation factor protein can be reversed rapidly, albeit temporarily, by extracorporeal immunoadsorption over a Sepharose-bound polyclonal antihuman IgG or staphylococcal A column, with concomitant replacement therapy and initiation of immunosuppression. Recombinant factor VIIa is increasingly used for surgery in those with rare congenital coagulation deficiencies but typically at lower doses than with hemophilia inhibitors—generally 10 to 15 µg/kg/day or less, for up to several days.

FACTOR DEFICIENCY IN AMYLOIDOSIS

Severe acquired deficiency of factor X, often accompanied by deficiencies of other vitamin K–dependent factors, occasionally occurs in individuals with systemic amyloidosis (Chapter 188). Because amyloid fibrils in the reticuloendothelial system bind endogenous and exogenous sources of factor X, replacement therapy with FFP or factor IX complex concentrates, even in large quantities, may not be sufficient. Recombinant factor VIIa concentrate has been used to reverse acute bleeding. Splenectomy may ameliorate recurrent bleeding complications.

Other Acquired Coagulation Abnormalities

ANTIPHOSPHOLIPID SYNDROME AND LUPUS ANTICOAGULANT

EPIDEMIOLOGY

The antiphospholipid antibody syndrome (APS) is associated with recurrent thrombosis or recurrent pregnancy loss with autoantibodies directed against phospholipids. Antiphospholipid antibodies (APAs) may include lupus anticoagulant (LA), anticardiolipin antibody (CL), or anti-β2-glycoprotein-I (anti-β2-GPI). Thrombocytopenia is also a common finding in this prothrombotic disorder. Rarely, multiorgan failure may arise from widespread thrombosis, termed *catastrophic APS*.

PATHOBIOLOGY

Our understanding of the role of individual antibodies in diagnosing APS continues to be enhanced with clinical studies, which may also contribute to the development of better diagnostic assays.[15] It has been suggested that the prothrombotic effects of APAs may derive from their ability to complex with β2-GPI in vivo (thereby negating their modulatory phospholipid-binding function) or through APA inhibition of protein C activation, interference with antithrombin III activity, and/or disruption of the annexin V "shield," thereby preventing normal clot breakdown (fibrinolysis). Although complement and inflammation play a role in fetal loss in a murine model of APS, the pathogenic mechanism of APS-associated pregnancy loss remains unclear. Although non-β2-GPI antibodies detected by CL enzyme-linked immunosorbent assay (ELISA) may play a role in early obstetric APS, and anti-β2-GPI antibodies with LA activity play a role in late miscarriage, additional studies are needed to better classify, understand, and manage this syndrome.

CLINICAL MANIFESTATIONS

APAs prolong coagulation in in vitro assays but are not generally associated with clinical bleeding. Rarely, clinical bleeding may occur when APA interacts with factor II (prothrombin), producing an acquired prothrombin (factor II) deficiency associated with accelerated clearance of LA-prothrombin complexes from the circulation. Bleeding tendencies also may arise when LA targets platelet membranes and produces quantitative and/or qualitative platelet abnormalities. Other clinical findings in APS include livedo reticularis (Chapter 80), valvular heart lesions, and nephropathy. Clinically, although both arterial and venous thrombosis (Chapter 176) may occur in APS, the most relevant is venous thrombosis and stroke in young adults. A diagnostic work-up in those with arterial thrombosis should include transesophageal echocardiography to exclude a cardiac source for arterial clots. In obstetric patients, other causes of miscarriage should be excluded. A case-control study has indicated that a history of recurrent spontaneous abortion associated with antiphospholipid syndrome is a risk factor for subsequent venous thromboembolism in the long term.[16] Nonpregnant individuals with APS-associated thrombotic manifestations have a 50% risk for experiencing recurrent events over a 5-year period. Typically, recurrent thrombotic and vascular episodes occur in a pattern consistent with the initial findings (e.g., venous recurrence following an initial deep vein thrombosis).

DIAGNOSIS

By international consensus, a diagnosis of APS is based on both clinical and laboratory criteria.[17] Clinical criteria include (1) arterial, venous, or small vessel thrombosis in any tissue or organ and (2) pregnancy morbidity. The latter includes one or more pregnancy losses before 10 weeks' gestation; one or more pregnancy losses before 34 weeks due to eclampsia, preeclampsia, or placental insufficiency; or three or more spontaneous abortions before 10 weeks. APS may be suspected in the setting of recurrent spontaneous miscarriages or pregnancy-related thromboembolic events; with the detection of cerebral arterial thromboses in young adults; in those with systemic lupus erythematosus (20 to 40%; Chapter 266) or other autoimmune diseases or lymphoproliferative malignancies; in those receiving psychotropic medications (e.g., chlorpromazine); or when incidental coagulation assays reveal aPTT prolongation.

Laboratory criteria[18] include one or more high-titer APAs (LA, CL, anti-β2-GPI) on at least two occasions at least 12 weeks apart. APAs can be detected by three assays: the LA assay, the CL ELISA assay, and the β2-GPI ELISA assay. The LA assay includes a screening test, based on a prolonged phospholipid-dependent clotting time; a mixing test that distinguishes the inhibitor (fails to correct in a 1 : 1 mix) from a deficiency state (corrects in a 1 : 1 mix); and a confirmatory test based on platelet neutralization to demonstrate that the inhibitor is phospholipid dependent (clotting time corrects) or a modified aPTT reagent based on the binding of hexagonal phase phospholipids to LA. The CL ELISA measures antibodies in dilute sera that bind to CL-coated plates, including those that bind to CL alone or bind to bovine β2-GPI and both IgG and IgM, but it may miss CL bound to human β2-GPI. The β2-GPI ELISA detects antibodies that bind to β2-GPI–coated irradiated plates, but it has reduced specificity owing to binding by nonpathogenic antibodies. Positivity in multiple assays is strongly associated with thrombosis and miscarriage: the risk for first thrombosis among asymptomatic individuals positive for LA, CL, and β2-GPI antibodies, so-called triple-positive patients, is 5.3% per year. Of the three assays, the β2-GPI ELISA is most strongly associated with thrombosis and early recurrent miscarriage; the LA assay is associated with venous thrombosis, stroke, and late miscarriage. The greatest weight is given to a positive LA test. From a practical standpoint, testing for all three antibodies at diagnosis is recommended to guide follow-up because oral anticoagulants may interfere with the LA test.

TREATMENT Rx

Because of the high risk for recurrent thromboembolism, patients with APS should receive antithrombotic therapy. The approach is based on achieving a balance between thrombosis risk and bleeding complications, especially in those with thrombocytopenia. Because of the limited information from randomized clinical trials, the recommendations for APS are similar to those for patients without APS—that is, to use unfractionated heparin or low-molecular-weight heparin (LMWH), with a 4- to 5-day overlap with warfarin, for acute venous thromboembolism (VTE), followed by long-term warfarin to a target international normalized ratio (INR) of 2.0 to 3.0. This recommendation is based on a randomized trial indicating no difference in bleeding rates with high-intensity (INR >3.0) versus low-intensity (INR 2.0 to 3.0) anticoagulation. In individuals with thrombocytopenia, more frequent monitoring of the INR may be warranted to avoid bleeding complications. The duration of anticoagulation in those both with and without APS is based on a balance between VTE recurrence and bleeding.

For women with APS and pregnancy loss but no past thrombosis, the goal is to prevent recurrent pregnancy loss. Based on prospective studies indicating higher birth rates with aspirin plus heparin, recommendations include low-dose aspirin (81 mg) in combination with prophylactic heparin or LMWH in the antepartum period.

Among those with APS-associated thrombocytopenia and a platelet count below 20 to 30 $\times 10^9$/L, in the setting of bleeding or when bleeding potential exceeds the risk for bleeding with VTE treatment, management is similar to

that for idiopathic thrombocytopenic purpura (Chapter 172). This includes steroids, intravenous immunoglobulin, and immunosuppressive agents (e.g., cyclophosphamide, azathioprine, off-label rituximab). Case reports indicate that danazol, dapsone, aspirin, and chloroquine may be helpful adjuncts. In pregnant women with APS-associated thrombocytopenia, treatment is advised before epidural anesthesia or cesarean delivery. In the latter case, early delivery under cover of intravenous immunoglobulin is recommended.

Treatment of APS-associated bleeding is dependent on the site and severity of bleeding. For anticoagulation bleeding, the anticoagulant should be withheld and an antidote (e.g., protamine) given, if available. For thrombocytopenia bleeding, platelet transfusions should be given, along with red blood cell transfusions, if necessary. If both thrombosis and bleeding occur, treatment should be directed at the component that is most life threatening. In those with a high bleeding risk, anticoagulants should be withheld or given at a lower dosage; those with a high thrombotic risk, despite thrombocytopenia, should be anticoagulated and given agents to boost platelet count simultaneously. For those who receive prolonged heparin, supplemental calcium and vitamin D should be administered to minimize the risks for osteoporosis.

In the life-threatening multisystem complication of catastrophic antiphospholipid syndrome, which is characterized by histopathologic evidence of small vessel thrombosis and dysfunction of multiple organs (e.g., lung, brain, heart, kidneys, skin, and/or gastrointestinal tract) over a short period of time, maximal anticoagulation, plasma exchange, and immunosuppressive therapy have been used. A recent report showed that eculizumab, an inhibitor of terminal complement, was effective in a patient with recurrent catastrophic antiphospholipid syndrome.[19]

Grade A References

A1. Mahlangu J, Powell JS, Ragni MV, et al. Phase 3 study of recombinant factor VIII Fc fusion protein in severe hemophilia A. *Blood*. 2014;123:317-325.

A2. Manco-Johnson MJ, Abshire TC, Shapiro AD, et al. Prophylaxis versus episodic treatment to prevent joint disease in boys with severe hemophilia. *N Engl J Med*. 2007;357:535-544.

A3. Gringeri A, Lundin B, von Mackensen S, et al. A randomized clinical trial of prophylaxis in children with hemophilia A (the ESPRIT Study). *J Thromb Haemost*. 2011;9:700-710.

A4. Valentino LA, Mamonov V, Hellmann A, et al. A randomized comparison of two prophylaxis regimens and a paired comparison of on-demand and prophylaxis treatments in hemophilia A management. *J Thromb Haemost*. 2012;10:359-367.

A5. Leissinger C, Gringeri A, Antmen B, et al. Anti-inhibitor coagulant complex prophylaxis in hemophilia with inhibitors. *N Engl J Med*. 2011;365:1684-1692.

A6. Hay CRM, DiMichele DM. The principal results of the International Immune Tolerance Study: a randomized dose comparison. *Blood*. 2012;119:1335-1344.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. MacCallum P, Bowles L, Keeling D. Diagnosis and management of heritable thrombophilias. *BMJ*. 2014;349:g4387.
2. Young G, Sørensen B, Dargaud Y, et al. Thrombin generation and whole blood viscoelastic assays in the management of hemophilia: current state of art and future perspectives. *Blood*. 2013;121:1944-1950.
3. Powell JS, Pasi KJ, Ragni MV, et al. Phase 3 study of recombinant factor IX Fc fusion protein in hemophilia B. *N Engl J Med*. 2013;369:2313-2323.
4. Shapiro A. Development of long-acting recombinant F VIII and F IX Fc fusion proteins for the management of hemophilia. *Expert Opin Biol Ther*. 2013;13:1287-1297.
5. Gouw SC, van den Berg HM, Fischer K, et al. Intensity of factor VIII treatment and inhibitor development in children with severe hemophilia A: the RODIN study. *Blood*. 2013;121:4046-4055.
6. Gouw SC, van der Bom HM, Ljung R, et al. Factor VIII products and inhibitor development in severe hemophilia A. *N Engl J Med*. 2013;368:231-239.
7. Oldenburg J, Brackmann HH. Prophylaxis in adult patients with severe haemophilia A. *Thromb Res*. 2014;134(Suppl 1):S33-S37.
8. Scott DW, Pratt KP, Miao CH. Progress toward inducing immunologic tolerance to factor VIII. *Blood*. 2013;121:4449-4456.
9. Leissinger C, Josephson CD, Granger S, et al. Rituximab for treatment of inhibitors in haemophilia A. A Phase II study. *Thromb Haemost*. 2014;112:445-458.
10. Walsh CE, Batt KM. Hemophilia clinical gene therapy: brief review. *Trans Res*. 2013;161:307-312.
11. Chang HH, Chiang BL. The diagnosis and classification of autoimmune coagulopathy: an updated review. *Autoimmun Rev*. 2014;13:587-590.
12. Baudo F, Collins P, Huth-Kühne A, et al. Management of bleeding in acquired hemophilia A: results from the European Acquired Hemophilia (EACH2) Registry. *Blood*. 2012;120:39-46.
13. Hsieh L, Nugent D. Rare factor deficiencies. *Curr Opin Hematol*. 2012;19:380-384.
14. Inbal A, Oldenburg J, Carcao M, et al. Recombinant factor XIII: a safe and novel treatment for congenital factor XIII deficiency. *Blood*. 2012;119:5111-5117.
15. Giannakopoulos B, Krilis SA. The pathogenesis of the antiphospholipid syndrome. *N Engl J Med*. 2013;368:1033-1044.
16. Martinez-Zamora MA, Peralta S, Creus M, et al. Risk of thromboembolic events after recurrent spontaneous abortion in antiphospholipid syndrome: a case-control study. *Ann Rheum Dis*. 2012;71:61-66.
17. Pengo V. APS—controversies in diagnosis and management, critical overview of current guidelines. *Thromb Res*. 2011;127(suppl 3):S51-S52.
18. Favaloro EJ, Wong RC. Laboratory testing for the antiphospholipid syndrome: making sense of antiphospholipid antibody assays. *Clin Chem Lab Med*. 2011;49:447-461.
19. Shapira I, Andrade D, Allen SL, et al. Induction of sustained remission in recurrent catastrophic antiphospholipid syndrome via inhibition of terminal complement with eculizumab. *Arthritis Rheum*. 2012;64:2719-2723.

REVIEW QUESTIONS

1. A 7-month-old boy with severe hemophilia A, who is receiving daily postoperative factor replacement for major surgery 10 days ago, has developed several large painful hematomas over his extremities. Despite factor replacement, his hematomas have not improved. Which of the following approaches is most appropriate for this child?

A. Draw trough factor VIII (before next dose), activated partial thromboplastin time (aPTT) mix, and anti–factor VIII.
B. Draw peak factor VIII (after the next dose), aPTT mix, and anti–factor VIII.
C. Draw aPTT, aPTT mix, and anti–factor VIII.
D. Draw random factor VIII.
E. Dose with 100% factor VIII and apply ice to hematomas.

Answer: B This child likely has developed an inhibitor because there was long duration factor early in life, and despite ongoing factor treatment, hematoma formation has occurred, indicating lack of response. To determine whether an inhibitor is present, the simplest approach is to draw a factor VIII level just after a 100% dose (i.e., peak); in the case of an inhibitor, the factor VIII level will be neutralized and lower than expected. A random factor VIII could also be drawn, but because children have a large volume of distribution, it is not as helpful as a peak factor VIII. Prolongation of both the aPTT and aPTT mix confirms the presence of an inhibitor, and the anti–factor VIII quantitates how high it is in Bethesda units. Given that the hematomas occurred while on treatment, it is unlikely that increasing the dose will improve the hematomas. Ice is a good approach, but determining whether the inhibitor is present is the most pressing need because, if present, a bypass agent may be needed, that is, recombinant factor VIIa (rFVIIa) or prothrombin complex concentrate (PCC).

2. A 26-year-old woman who is 20 weeks pregnant is referred for management recommendations at delivery. She has had menorrhagia in the past and oozing with dental extractions. A maternal uncle has hemarthroses. What is the most appropriate next step in the evaluation of this patient?

A. Screen for von Willebrand disease.
B. Screen for hemophilia carrier status and von Willebrand disease.
C. Screening is not helpful because von Willebrand factor (VWF) and factor VIII increase during pregnancy.
D. Obtain VWF and factor VIII in the 8th month of pregnancy.
E. Plan DDAVP (desmopressin) for delivery and neuranesthesia.

Answer: D This woman may have a congenital bleeding disorder, with both a personal and a family bleeding history. Menorrhagia is a common symptom of von Willebrand disease, but it could also indicate hemophilia A or B carrier status. The history of a hemarthrosis (bleeding in a joint) in her uncle suggests he may have the X-linked congenital bleeding disorder, hemophilia, either A or B, and if inherited, gives her a 50-50 chance of being a carrier. Hemarthrosis is uncommon in other bleeding disorders but may occur in severe type 3 von Willebrand disease. Testing for von Willebrand disease or hemophilia carrier status is complicated by the well-known increase in VWF and factor VIII (but not factor IX) during pregnancy, potentially masking diagnosis of VWD or hemophilia carrier, but it is critical to screen VWF, factor VIII, and factor IX at the 8th month of pregnancy to plan for peripartum and neuranesthesia management. DDAVP is contraindicated at delivery because of major fluid loss and intravenous replacement, which could precipitate hyponatremia and seizures; VWF concentrate is the drug of choice for VWD, and recombinant factor VIII or IX is indicated for symptomatic factor VIII (or IX) for carriers.

3. A 47-year-old woman weighing 150 kg with factor VII deficiency is planning to undergo gastric bypass surgery. She is referred to you for hemostatic management of her surgery. What is the most appropriate management for her surgery?

A. Fresh-frozen plasma (FFP) 15-20 U/kg before and for up to 1-2 days postoperatively
B. PCC 50 IU/kg before and every 6 hours for 2 days postoperatively
C. PCC 25 IU/kg before and every 6 hours for 2 days postoperatively
D. Recombinant VIIa 90 μg/kg before and every 3-4 hours for 2 days postoperatively
E. Recombinant VIIa 15 μg/kg before and on day 1 and 2 postoperatively

Answer: E The current treatment for congenital factor VII deficiency is low-dose factor VIIa: for surgery, it is given preoperatively and for several days postoperatively. After the tissue factor VIIa–triggered extrinsic coagulation pathway is activated, several days of treatment are usually sufficient to maintain the clot. Because obesity and surgery are both associated with risk for thromboembolism, the patient should be mobilized early in the postoperative course. FFP and PCC are alternative treatments, but levels of VII are low in FFP and in three-factor PCC, and the latter may increase thrombosis risk. rFVIIa in this patient would be less likely to increase thrombosis risk, given the patient's congenital deficiency (associated with lack of thrombin burst).

4. A 17-year-old high school senior with congenital factor XI deficiency is referred for management of his upcoming wisdom teeth extraction. His family history is negative for bleeding, and the patient has no bleeding history but has not had surgery. What is the most appropriate surgical management for this patient?

A. FFP 15-20 U/kg preoperatively
B. Amicar 50 mg/kg PO every 6 hours preoperatively
C. DDAVP 0.3 μg/kg IV over 30 minutes preoperatively, and days 1 and 2 postoperatively
D. Stimate 300 μg by nasal spray, given as 150 μg in each nostril preoperatively
E. No preoperative treatment, but FFP 14-20 U/kg for postoperative bleeding

Answer: E At least half of those with factor XI deficiency have no personal or family history of bleeding, even with factor XI levels below 1%. Bleeding tendency is consistent within families. In the absence of family history, no treatment is given preoperatively, but if postoperative bleeding occurs, FFP is recommended. In Europe, a factor XI concentrate is available, but it not approved in the United States. Although Amicar, DDAVP, and Stimate may provide hemostasis, they are not specific for factor XI deficiency. For mucosal bleeding postoperatively, Amicar is a reasonable alternative.

5. A 50-year-old man with the lupus anticoagulant test (LAC) and positive β2-glycoprotein I (β2-GPI) antibody, as well as a recent lower extremity deep vein thrombosis, comes for a second opinion regarding need for long-term anticoagulation. In 2 weeks, he will complete a 6-month course of Coumadin 7.5 mg alternating with 5 mg to maintain an international normalized ratio (INR) of 2 to 3. What is the optimal management of this patient?

A. Coumadin to maintain INR of 2.0 to 3.0 as long as benefits outweigh risks
B. Coumadin to maintain INR of 2.0 to 3.0 plus low-dose acetylsalicylic acid (ASA; 81 mg) as long as benefits outweigh risks
C. Coumadin to maintain INR of >3.0 as long as benefits outweigh risks
D. Coumadin to maintain INR of >3.0 plus low-dose ASA (81 mg) as long as benefits outweigh risks.
E. Stop Coumadin after 6 months, with careful monitoring for venous thromboembolism (VTE) recurrence.

Answer: A The occurrence of β2-GPI antibody and LAC with deep vein thrombosis is consistent with the antiphospholipid syndrome (APS). Individuals with APS are at increased risk for recurrent VTE, and thus anticoagulation should be continued as long as the benefits of anticoagulation outweigh bleeding risk. Several studies have demonstrated that standard Coumadin to an INR of 2.0 to 3.0 is adequate VTE prophylaxis for individuals with APS, with no added benefit to maintain INR above 3 for patients with antiphospholipid syndrome. ASA, other than in the setting of pregnancy, appears to provide no added benefit in clot protection in individuals with APS.

175

HEMORRHAGIC DISORDERS: DISSEMINATED INTRAVASCULAR COAGULATION, LIVER FAILURE, AND VITAMIN K DEFICIENCY

ANDREW I. SCHAFER

DISSEMINATED INTRAVASCULAR COAGULATION

DEFINITION

Disseminated intravascular coagulation (DIC), also referred to as consumptive coagulopathy or defibrination, is caused by a wide variety of serious disorders (Table 175-1). In most patients, the underlying process dominates the clinical picture, but in some cases (e.g., occult malignant neoplasm, envenomation), DIC may be the initial or predominant manifestation of the disorder. DIC never occurs in isolation, without an inciting cause.

TABLE 175-1 MAJOR CAUSES OF DISSEMINATED INTRAVASCULAR COAGULATION

INFECTIONS
Gram-negative bacterial sepsis
Other bacteria, fungi, viruses, Rocky Mountain spotted fever, malaria
IMMUNOLOGIC REACTIONS
Transfusion reactions (ABO incompatibility)
Transplant rejection
OBSTETRIC COMPLICATIONS
Amniotic fluid embolism
Retained dead fetus
Abruptio placentae
Toxemia, preeclampsia
Septic abortion
MALIGNANT NEOPLASMS
Pancreatic carcinoma
Adenocarcinomas
Acute promyelocytic leukemia
Other neoplasms
LIVER FAILURE
ACUTE PANCREATITIS
ENVENOMATION
RESPIRATORY DISTRESS SYNDROME
TRAUMA, SHOCK
Brain injury
Crush injury
Burns
Hypothermia or hyperthermia
Fat embolism
Hypoxia, ischemia
Surgery
VASCULAR DISORDERS
Giant hemangioma (Kasabach-Merritt syndrome)
Aortic aneurysm
Vascular tumors

PATHOBIOLOGY

DIC is pathophysiologically a thrombotic process. However, its clinical manifestation may be widespread hemorrhage in acute cases. The basic pathophysiologic mechanism (Fig. 175-1), regardless of cause, is entry into the circulation of procoagulant substances that trigger systemic activation of the coagulation system and platelets. This results in disseminated deposition of fibrin-platelet thrombi within the microvasculature.[1] In most cases, the procoagulant stimulus is tissue factor, a lipoprotein that is not normally exposed to blood. In DIC, tissue factor gains access to blood by tissue injury, its elaboration by malignant cells, or its expression on the surface of monocytes and endothelial cells by inflammatory mediators. Components of the inflammatory response and the coagulation system are reciprocally activated in some forms of DIC, such as sepsis. Tissue factor triggers generation of the coagulation protease thrombin, which induces fibrin formation and platelet activation. In some specific cases of DIC, procoagulants other than tissue factor (e.g., a cysteine protease or mucin in certain malignant neoplasms) and proteases other than thrombin (e.g., trypsin in pancreatitis, exogenous enzymes in envenomation) provide the procoagulant stimulus.

In acute, uncompensated DIC, coagulation factors are consumed at a rate in excess of the capacity of the liver to synthesize them, and platelets are consumed in excess of the capacity of bone marrow megakaryocytes to release them. The resulting laboratory manifestations under these circumstances are a prolonged prothrombin time (PT) and activated partial thromboplastin time (aPTT) and thrombocytopenia. Increased fibrin formation in DIC stimulates a heightened process of secondary fibrinolysis, in which plasminogen activators generate plasmin to digest fibrin (and fibrinogen) into fibrin(ogen) degradation products (FDPs). FDPs are potent circulating anticoagulants that further contribute to the bleeding manifestations of DIC. Intravascular fibrin deposition can cause fragmentation of red blood cells and lead to the appearance of schistocytes in blood smears; however, frank microangiopathic hemolytic anemia is unusual in DIC. Occlusive microvascular thrombosis in DIC can compromise the blood supply to some organs and

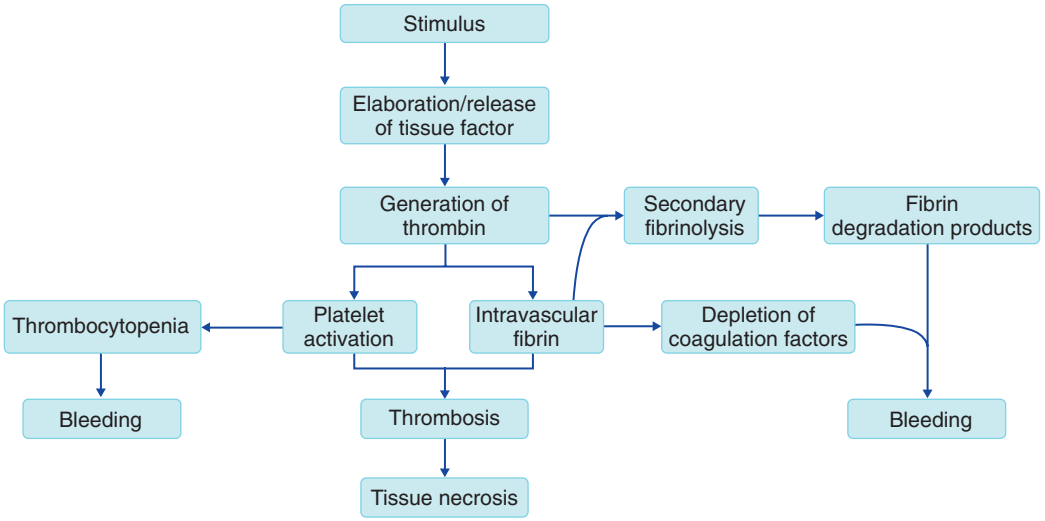

FIGURE 175-1. **Pathophysiologic process of bleeding, thrombosis, and ischemic manifestations in patients with disseminated intravascular coagulation.**

lead to multiorgan failure, particularly when it is accompanied by systemic hemodynamic and metabolic derangements.

Underlying Causes

DIC always has an underlying cause that generally must be identified and eliminated if the coagulopathy is to be managed successfully. The development of DIC in many of these disorders is associated with an unfavorable outcome. Infection is the most common cause of DIC. The syndrome is particularly associated with gram-negative or gram-positive sepsis (Chapter 108), although it can be triggered by a variety of other bacterial, fungal, viral, rickettsial, and protozoal microorganisms.

The placenta and uterine contents are rich sources of tissue factor and other procoagulants that are normally excluded from the maternal circulation; a spectrum of clinical manifestations of DIC may accompany obstetric complications when this barrier is breached, especially in the third trimester. These syndromes range from acute, fulminant, and often fatal DIC in amniotic fluid embolism to chronic or subacute DIC with a retained dead fetus. Other obstetric problems associated with DIC include abruptio placentae, toxemia, and septic abortion.

Chronic forms of DIC are caused by a variety of malignant neoplasms, particularly pancreatic cancer (Chapter 194) and mucin-secreting adenocarcinomas of the gastrointestinal tract (Chapter 193), in which thrombotic rather than bleeding manifestations predominate. Treatment with all-*trans*-retinoic acid has greatly reduced the incidence of severe DIC in patients with acute promyelocytic leukemia (Chapter 183). It is not known whether liver failure (see later) can cause DIC or whether its coexistence merely exacerbates intravascular coagulation because of impaired clearance of activated clotting factors, plasmin, and FDPs. Snake venom contains a variety of substances that can affect coagulation and endothelial permeability. Bites from rattlesnakes and other vipers can induce profound DIC by introduction of these exogenous toxins and release of endogenous tissue factor through tissue necrosis.

The likelihood and degree of DIC caused by trauma, surgery, and shock (Chapter 106) are related to the extent of tissue damage and the organs involved. The brain is a particularly rich source of tissue factor, so traumatic brain injury (Chapter 399) can precipitate acute DIC. Large aortic aneurysms (Chapter 78), giant hemangiomas, and other vascular malformations can cause subclinical or clinical DIC that is initiated locally within the abnormal vasculature but can "spill" into the systemic circulation.

CLINICAL MANIFESTATIONS

The clinical manifestations of DIC are determined by the nature, intensity, and duration of the underlying stimulus. The coexistence of liver disease exacerbates DIC of any cause. Low-grade DIC is often asymptomatic and diagnosed only by laboratory abnormalities. Thrombotic complications of DIC occur most often with chronic underlying diseases, as exemplified by Trousseau's syndrome in cancer (Chapter 176). DIC can be manifested as acrocyanosis and gangrene of the digits in critically ill, hemodynamically compromised patients receiving vasopressors. Hemorrhagic necrosis of the skin (Fig. 175-2) and purpura fulminans may also be manifestations of DIC.

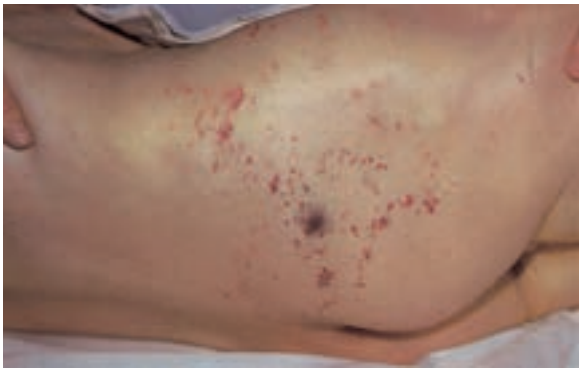

FIGURE 175-2. **Disseminated intravascular coagulation resulting from staphylococcal septicemia.** Note the characteristic skin hemorrhage ranging from small purpuric lesions to larger ecchymoses. (From Forbes CD, Jackson WF. Color Atlas and Text of Clinical Medicine. 3rd ed. London: Mosby; 2003.)

Bleeding is the most common clinical finding in acute, uncompensated DIC. Bleeding can be limited to sites of intervention or anatomic abnormalities, but it tends to be generalized in more severe cases, including widespread ecchymoses and diffuse oozing from mucosal surfaces and orifices.

DIAGNOSIS

The laboratory diagnosis of severe, acute DIC is not usually difficult. Consumption and inhibition of the function of clotting factors cause prolongation of the PT, aPTT, and thrombin time. Consumption of platelets causes thrombocytopenia. Secondary fibrinolysis generates increased titers of FDPs, which can be measured by latex agglutination or D-dimer assays. Some schistocytes may be seen in the peripheral blood smear, but this finding is neither sensitive nor specific for DIC. Chronic or compensated forms of DIC are more difficult to diagnose, with highly variable patterns of abnormalities in "DIC screen" coagulation tests.[2,3] Increased D-dimers and a prolonged PT are generally more sensitive measures than are abnormalities of the aPTT and platelet count. Overcompensated synthesis of consumed clotting factors and platelets in some chronic forms of DIC may actually cause shortening of the PT and aPTT or thrombocytosis (or both), even though elevated levels of D-dimers indicate secondary fibrinolysis in such cases.

The most difficult differential diagnosis of DIC occurs in patients who have coexisting liver disease. The coagulopathy of liver failure (see later section on liver failure) is often indistinguishable from that of DIC, partly because advanced hepatic dysfunction is in fact accompanied by a state of DIC. In liver failure, the combination of decreased synthesis of clotting factors, impaired clearance of activated clotting factors, secondary fibrinolysis, and thrombocytopenia from portal hypertension and hypersplenism may make the coagulopathy virtually impossible to differentiate from DIC. Thrombotic microangiopathies including thrombotic thrombocytopenic purpura and hemolytic-uremic syndrome, the syndrome of "hemolysis, elevated liver enzymes, and low platelet count" (HELLP) in obstetric patients,[4] and other

forms of platelet consumption and thrombocytopenia (Chapter 172) are not accompanied by activation of clotting factors or secondary fibrinolysis; therefore, the PT, aPTT, thrombin time, and D-dimers are generally normal in these disorders. Schistocytes, often with frank hemolysis, are much more prominent in the peripheral smear in thrombotic thrombocytopenic purpura and hemolytic-uremic syndrome (Chapter 172) than in DIC.

Primary hyperfibrinolysis is disputed as a distinct entity.[5] Some patients with a serious clinical bleeding diathesis, however, have laboratory evidence of predominantly fibrinolysis, including high levels of FDPs (D-dimers) and severe hypofibrinogenemia, with relatively little consumption of coagulation factors and normal or nearly normal platelet counts. These unusual findings, which approximate the findings expected with fibrinolytic therapy, are encountered occasionally, particularly in patients with prostate cancer.

TREATMENT

Successful treatment of DIC (Table 175-2) requires that the underlying cause be identified and eliminated. All other therapies, including hemodynamic support, replacement of coagulation factors and platelets, and pharmacologic inhibitors of coagulation and fibrinolysis, are just temporizing measures. Because of the difficulty of testing this complex syndrome with multiple causes by controlled, randomized clinical trials, current guidelines for treatment are generally not based on high-grade evidence.[6]

In many patients with asymptomatic, self-limited DIC who have only laboratory manifestations of the coagulopathy, no treatment may be necessary. In patients with DIC who are actively bleeding or who are at high risk for bleeding, the blood component treatments of choice are transfusions of platelets to improve the thrombocytopenia and fresh-frozen plasma to replace all consumed coagulation factors and to correct the prolonged PT and aPTT. Large volumes of plasma (e.g., >6 U/24 hours) may be required to ameliorate bleeding in severe cases. In some patients who have particularly profound hypofibrinogenemia, the additional transfusion of cryoprecipitate, a plasma concentrate that is enriched in fibrinogen, may be useful. The theoretical concern that these blood products could "fuel the fire" and exacerbate the DIC has not been supported by clinical experience.

The use of pharmacologic inhibitors of coagulation and fibrinolysis in DIC is controversial. Heparin is of theoretical benefit because it blocks thrombin activity and quenches intravascular coagulation and the resultant secondary fibrinolysis. In practice, heparin might exacerbate the bleeding tendency in acute DIC. Heparin is usually reserved for special forms of DIC, including those manifested by thrombosis or acrocyanosis and forms that accompany cancer, vascular malformations, retained dead fetus, and possibly acute promyelocytic leukemia, in which active bleeding is not present. In cases of DIC in which thrombosis or acral ischemia predominates, unfractionated heparin should be used by continuous infusion because of its short half-life and reversibility in the event of increased bleeding. Monitoring of the aPTT in the presence of DIC may be problematic, so heparin infusion in this setting should be followed mainly by clinical response and improvement in results of other tests of coagulation (e.g., thrombocytopenia). Antifibrinolytic agents, including ε-aminocaproic acid and tranexamic acid, are generally contraindicated in DIC. By blocking the secondary fibrinolytic response to DIC, these drugs cause unopposed fibrin deposition and may precipitate thrombosis. Antifibrinolytic agents may be effective in decreasing life-threatening bleeding in DIC, however, particularly in extreme cases in which aggressive blood component replacement fails to control the hemorrhage; in such situations, simultaneous infusion of low doses of heparin may reduce the risk for thrombosis. Bleeding and DIC laboratory parameters may also be improved with recombinant activated factor VII (rFVIIa) in patients whose bleeding is not controlled by standard measures.

TABLE 175-2 TREATMENT OF DISSEMINATED INTRAVASCULAR COAGULATION

Identify and eliminate the underlying cause
No treatment if mild, asymptomatic, and self-limited
Hemodynamic support, as indicated, in severe cases
Blood component therapy
Indications: active bleeding or high risk for bleeding
Fresh-frozen plasma
Platelets
In some cases, consider cryoprecipitate, antithrombin III
Drug therapy
Indications: heparin for DIC manifested by thrombosis or acrocyanosis and without active bleeding; antifibrinolytic agents are generally contraindicated except with life-threatening bleeding and failure of blood component therapy

DIC = disseminated intravascular coagulation.

Recombinant human activated protein C (rhAPC) was approved in 2001 for use in severe sepsis or septic shock, which is often accompanied by DIC. The mechanisms by which rhAPC is thought to modify disease course are through its actions on dysregulated coagulation and subsequent microvascular thrombosis as well as by its possible anti-inflammatory effects. However, subsequent trials failed to confirm the mortality reduction benefit of rhAPC in patients with septic shock or any of its prescribed subgroups or in groups that were less acutely ill, with lower risk of death from sepsis (Acute Physiology and Chronic Health Evaluation or APACHE II scores <25 or single-organ failure), and with extended infusions or in pediatric patients. Therefore, rhAPC was withdrawn by its manufacturer in 2011. Nevertheless, controversy has persisted with the subsequent publication of favorable meta-analyses and observational trials.[7] Although recombinant tissue factor pathway inhibitor, antithrombin III concentrates, and human soluble thrombomodulin may have some efficacy in improving laboratory parameters of DIC, the overall survival benefit with these agents in patients with DIC has not been demonstrated.

LIVER FAILURE

Bleeding complications in patients with advanced liver disease (Chapters 153 and 154) can be severe and even fatal and directly account for about 20% of the deaths associated with hepatic failure. The extent of the bleeding tendency depends on the severity and type of liver disease involved. About one third of deaths in patients undergoing liver transplantation are attributable to perioperative hemorrhage.

PATHOBIOLOGY

The pathophysiologic mechanism of bleeding in liver failure is complex and multifactorial.[8] Anatomic abnormalities resulting from portal hypertension are frequently the major cause of gastrointestinal bleeding in patients with liver disease. Upper gastrointestinal bleeding can be caused by esophageal varices or hemorrhagic gastritis (congestive gastropathy), whereas lower gastrointestinal bleeding, although seldom life-threatening, can be due to hemorrhoids.

The complexity of the systemic coagulopathy of liver failure is not surprising inasmuch as the liver is the principal organ site for the synthesis of coagulation and fibrinolytic factors as well as their protein inhibitors (Table 175-3). Therefore, the impaired hemostasis of liver failure is accompanied by opposing prothrombotic alterations. Hepatocytes produce all of the clotting factors except von Willebrand factor, and advanced parenchymal liver disease results in impaired synthesis of these proteins. Liver disease can also cause impairment in vitamin K–dependent γ-carboxylation of the procoagulant factors II, VII, IX, and X as well as the anticoagulant proteins C and S. Functional abnormalities of fibrinogen, termed dysfibrinogenemias, are frequently found in various forms of liver disease, particularly in hepatocellular carcinoma. Most forms of advanced liver disease are accompanied by some degree of DIC caused by impaired synthesis of inhibitors of blood coagulation and defective hepatocellular clearance of activated coagulation factors. DIC and bleeding risk are exacerbated by the enhanced fibrinolytic activity (hyperfibrinolysis) of liver disease caused by increased levels of tissue plasminogen activator accompanied by decreased synthesis of inhibitors of plasminogen activator and plasmin.

Quantitative and qualitative abnormalities of platelets also contribute to the bleeding diathesis of liver failure. Congestive splenomegaly secondary to portal hypertension causes increased pooling of platelets in the spleen (hypersplenism). The resultant thrombocytopenia, the degree of which generally correlates with spleen size, rarely causes a reduction in the platelet count to less than 30,000/mm^3. In alcoholic patients, suppression of bone marrow thrombopoiesis by the acute toxic effects of alcohol or folate deficiency may contribute to the thrombocytopenia. Qualitative platelet abnormalities have also been described in patients with liver disease.

Liver transplantation (Chapter 154) poses special problems to the coagulation system. During the anhepatic stage of surgery, which lasts about 2 hours, the complete cessation of synthesis of coagulation factors causes further prolongation of the PT and aPTT. Release of tissue plasminogen activator from the newly grafted liver leads to increased fibrinolysis and transient exacerbation of bleeding risk in the postoperative period.

CLINICAL MANIFESTATIONS

The most common hemorrhagic complication of liver disease is gastrointestinal bleeding, which is usually caused by anatomic abnormalities and exacerbated by the systemic coagulopathy of liver failure. Bleeding from other

TABLE 175-3 COAGULATION ABNORMALITIES IN LIVER DISEASE

ABNORMALITIES IN COAGULATION
Decreased synthesis of coagulation factors
Impaired vitamin K–dependent γ-carboxylation
Dysfibrinogenemia
Disseminated intravascular coagulation
Increased fibrinolytic activity
ABNORMALITIES IN PLATELETS
Thrombocytopenia (hypersplenism)
Abnormal platelet function

mucosal sites, extensive ecchymoses, or more serious hemorrhage into the retroperitoneum or central nervous system generally indicates more significant derangements of the coagulation system.

The severe coagulopathy in patients with liver disease makes liver biopsy a potentially hazardous procedure. The PT and platelet count may be the best guides to bleeding risk, but they too lack reliability as predictors of risk of bleeding with liver biopsy. In general, liver biopsy can be performed safely if the PT and aPTT do not exceed 1.5 times control values and the platelet count is higher than 50,000/mm^3. The American Association for the Study of Liver Diseases has concluded that the bleeding risk-benefit ratio of liver biopsy must be carefully considered on a case-by-case basis.[9]

DIAGNOSIS

Although the PT and aPTT are often prolonged in patients with advanced liver disease, the PT tends to be a more sensitive assay early in the course; a disproportionate prolongation of the aPTT should raise suspicion of a coexisting coagulation abnormality, such as a lupus anticoagulant or clotting factor inhibitor. A prolonged PT is also a useful prognostic indicator of poor outcome in patients with cirrhosis, acute acetaminophen hepatotoxicity, and acute viral hepatitis; in acute viral hepatitis, it is a better index of prognosis than are serum albumin and transaminases. A disproportionate prolongation of the thrombin time should suggest the presence of dysfibrinogenemia. Hypersplenism (Chapter 168), possibly associated with nutritional folate deficiency or the acute toxic effects of alcohol on bone marrow, often causes mild to moderate thrombocytopenia in patients with liver disease; however, consideration should be given to other coexisting causes of thrombocytopenia if the platelet count is significantly less than 30,000/mm^3.

The coagulopathy of liver failure is often indistinguishable from that of DIC, in part because some degree of DIC is in fact a necessary accompaniment of advanced liver disease. In general, patients with DIC have more marked decreases in levels of factor VIII and increases in D-dimer than do patients with liver failure. The simplistic notion that a prolonged PT (and sometimes also aPTT) as well as a thrombocytopenia creates a state of natural "auto-anticoagulation" in patients with liver failure ignores the complex pathophysiologic mechanism of the coagulopathy in this setting, which involves a delicate balance of not only antihemostatic abnormalities (leading to a bleeding tendency) but also prohemostatic abnormalities (leading to a simultaneous thrombotic tendency).[10] In fact, the risk of venous thromboembolism is two-fold higher in patients with liver failure than in controls, even in the presence of a prolonged international normalized ratio. The thrombotic risk is not just a localized one, involving portal vein and other abdominal venous thrombosis, but also a systemic one, involving lower extremity deep venous thrombosis and pulmonary embolism.

TREATMENT

Rx

Therapy for the coagulopathy of liver disease may be directed at preventing the hemorrhagic complications of invasive procedures or treating active bleeding. The most effective treatment is blood component therapy with fresh-frozen plasma (which contains all the coagulation factors) and platelet transfusions (Chapter 177). Some patients require large volumes of fresh-frozen plasma (15 to 20 mL/kg) to lower the prolonged PT; rarely, plasmapheresis with plasma exchange is required to avoid fluid overload in these situations. Because of the short half-lives of some clotting factors, fresh-frozen plasma may have to be administered every 8 to 12 hours to maintain acceptable coagulation test parameters. In some patients, especially those with cholestasis, parenteral administration of vitamin K can at least partially reverse the coagulation abnormalities; however, in patients with advanced hepatocellular failure, vitamin K is largely ineffective. Prothrombin complex concentrates are relatively contraindicated in patients with liver failure, as in those with DIC, because of the risk of thrombotic complications. Because of immediate pooling of transfused platelets in the enlarged spleen of patients with hypersplenism, a higher than calculated dose of platelet concentrates is usually required to increase circulating platelet counts significantly. The oral thrombopoietin receptor agonist eltrombopag (25 to 100 mg daily) has been shown to increase platelet numbers in thrombocytopenic patients with hepatitis C virus infection complicated by advanced fibrosis and cirrhosis, allowing otherwise ineligible patients to use antiviral therapy, leading to increases in sustained virologic responses.[A2]

A Cochrane database meta-analysis of randomized clinical trials of recombinant human activated factor VII (rFVIIa) found no evidence to support or to reject its administration to patients with liver disease and upper gastrointestinal bleeding.[A3] Routine use of rFVIIa for the coagulopathy of liver disease cannot be recommended at this time. Desmopressin (DDAVP), which can shorten the bleeding time of patients with cirrhosis, may be considered as ancillary therapy in patients undergoing invasive procedures.

VITAMIN K DEFICIENCY

Vitamin K is required for γ-carboxylation of glutamic acid residues of the procoagulant factors II (prothrombin), VII, IX, and X and the anticoagulant factors protein C and protein S. This post-translational modification normally renders these proteins functionally active in coagulation. The PT is more sensitive than the aPTT in detecting vitamin K deficiency states because factor VII, the only vitamin K–dependent factor that is in the extrinsic pathway of coagulation, is the most labile of these proteins.

The two major sources of vitamin K are dietary intake and synthesis by the bacterial flora of the intestine. In the absence of malabsorption, nutritional deficiency alone rarely causes clinically significant vitamin K deficiency. The condition can arise, however, when eradication of gut flora is combined with inadequate dietary intake. This situation typically occurs in critically ill patients in intensive care units who have no oral intake and are receiving broad-spectrum antibiotics for prolonged periods. Vitamin K deficiency can also develop in patients receiving total parenteral nutrition unless the infusions are supplemented with vitamin K.

Vitamin K is absorbed predominantly in the ileum and requires the presence of bile salts. Clinically significant vitamin K deficiency occurs with malabsorption of fat-soluble vitamins secondary to obstructive jaundice (Chapter 155) or with malabsorption caused by intrinsic small bowel diseases, including celiac sprue, short-bowel syndrome, and inflammatory bowel disease (Chapters 140 and 141).

Warfarin (Chapter 38) acts as an anticoagulant by competitive antagonism of vitamin K. Rare cases of hereditary deficiency of the vitamin K–dependent coagulation factors may cause a lifelong bleeding tendency.

Correction of vitamin K deficiency, when it is clinically significant, can be achieved with oral supplementation, unless malabsorption is present, in which case parenteral vitamin K (10 mg subcutaneously daily) should be administered. Emergency treatment of bleeding caused by vitamin K deficiency is transfusion of fresh-frozen plasma.

Grade A References

A1. Ranieri VM, Thompson BT, Barie PS, et al. Drotrecogin alfa (activated) in adults with septic shock. *N Engl J Med*. 2012;366:2055-2064.

A2. Afdhal NH, Dusheiko GM, Giannini EG, et al. Eltrombopag increases platelet numbers in thrombocytopenic patients with HCV infection and cirrhosis, allowing for effective antiviral therapy. *Gastroenterology*. 2014;146:442-452.

A3. Marti-Carvajal AJ, Karakitsiou DE, Salanti G. Human recombinant activated factor VII for upper gastrointestinal bleeding in patient with liver disease. *Cochrane Database Syst Rev*. 2012;3:CD004887.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hook KM, Abrams CS. The loss of homeostasis in hemostasis: new approaches in treating and understanding acute disseminated intravascular coagulation in critically ill patients. *Clin Transl Sci*. 2012;5:85-92.
2. Hunt BJ. Bleeding and coagulopathies in critical care. *N Engl J Med*. 2014;370:847-859.
3. Wada H, Matsumoto T, Yamashita Y, et al. Disseminated intravascular coagulation: testing and diagnosis. *Clin Chim Acta*. 2014;436:130-134.
4. Abildgaard U, Heimdal K. Pathogenesis of the syndrome of hemolysis, elevated liver enzymes, and low platelet count (HELLP): a review. *Eur J Obstet Gynecol Reprod Biol*. 2013;166:117-123.
5. Bennani-Baiti N, Daw HA. Primary hyperfibrinolysis in liver disease: a critical review. *Clin Adv Hematol Oncol*. 2011;9:250-251.
6. Di Nisio M, Baudo F, Cosmi B, et al. Diagnosis and treatment of disseminated intravascular coagulation: guidelines of the Italian Society of Haemostasis and Thrombosis (SISET). *Thromb Res*. 2012;129:e177-e184.
7. Lai PS, Thompson BT. Why activated protein C was not successful in severe sepsis and septic shock: are we still tilting at windmills? *Curr Infect Dis Rep*. 2013;15:407-412.
8. Northup PE, Caldwell SH. Coagulation in liver disease: a guide for the clinician. *Clin Gastroenterol Hepatol*. 2013;11:1064-1074.
9. Rockey DC, Caldwell SH, Goodman ZD, et al. Liver biopsy. *Hepatology*. 2009;49:1017-1044.
10. Habib M, Roberts LN, Patel RK, et al. Evidence of rebalanced coagulation in acute liver injury and acute liver failure as measured by thrombin generation. *Liver Int*. 2014;34:672-678.

REVIEW QUESTIONS

1. Which of the following statements is correct regarding disseminated intravascular coagulation (DIC)?

A. There is always an underlying cause for DIC.
B. Laboratory screening test results showing normal to increased platelet counts or normal to shortened (not prolonged) prothrombin time (PT) and activated partial thromboplastin time (aPTT) essentially rule out DIC.
C. The primary pathophysiologic mechanism of DIC is bleeding; thrombosis can sometimes occur as a secondary event.
D. Large numbers of schistocytes (fragmented red cells) are characteristic blood smear findings in DIC.
E. The prolongation of PT and aPTT in DIC is a function of depletion of coagulation factors.

Answer: A DIC is a syndrome with many possible causes and never occurs in isolation as a primary event without an underlying cause. The results of laboratory screening tests of DIC (PT, aPTT, platelet count, D-dimers) can be highly variable; in fact, in chronic forms of so-called overcompensated DIC, most commonly due to malignant disease, the PT and aPTT can be actually shortened and the platelet count can even be elevated. The primary pathophysiologic mechanism of DIC, regardless of cause, is activation of the coagulation cascade leading to increased fibrin generation and microvascular fibrin deposition; therefore, DIC is a primarily thrombotic process. The systemic bleeding that is more often seen in acute DIC is due to a combination of depleted ("consumed") coagulation factors and platelets as well as the circulating anticoagulant effect generated by increased levels of D-dimers. Schistocytes (fragmented red cells) are nowhere nearly as prominent in the peripheral smear of a patient with DIC as they are in thrombotic thrombocytopenic purpura. As noted before, the prolongation of PT and aPTT in DIC is caused not just by the depletion of clotting factors ("consumption") but also by the powerful circulating anticoagulant effects of fibrin(ogen) degradation products.

2. Which of the following statements is correct regarding the coagulopathy of liver failure?

A. In considering anticoagulation in a patient with liver failure, it should be taken into account that the already prolonged PT (and sometimes also aPTT) intrinsic to liver failure has already made the patient at least partially "auto-anticoagulated."
B. The degree of decrease in serum albumin is a better prognostic marker than the PT (or international normalized ratio [INR]) in a patient with acute viral hepatitis who has fulminant liver failure.
C. Patients with the coagulopathy of liver failure (prolonged PT and sometimes aPTT, with thrombocytopenia) have a serious bleeding risk but not thromboembolic risk.
D. In general, liver biopsy can be safely performed in a patient with liver failure when the PT and aPTT do not exceed 1.5 times control values and the platelet count is above 50,000.
E. In case of bleeding in a liver failure patient, prothrombin complex concentrates and recombinant human activated factor VII (rFVIIa) are first-line hemostatic agents.

Answer: D The American Association for the Study of Liver Diseases has stated that the bleeding risk-benefit ratio of liver biopsy must be carefully considered on a case-by-case basis. Nevertheless, most experts think that in general, biopsy can be performed without excessive risk of bleeding with the PT and aPTT both being less than 1.5 times control and the platelet count above 50,000. The notion that a patient in liver failure who already has a prolonged PT and thrombocytopenia is auto-anticoagulated is misleading. It ignores the complex pathophysiologic process of the coagulopathy of liver failure, which also involves important prothrombotic abnormalities. Patients with liver failure have a two-fold increase in the risk of venous thromboembolism compared with control individuals, even when the PT (INR) is prolonged. Prothrombin complex concentrates and rVIIa are effective "general" hemostatic agents but carry a risk of thrombosis, especially in liver failure patients, who may have an impaired ability to clear those activated clotting factors. A prolonged PT is a useful prognostic indicator in various forms of liver disease and is actually a better index of prognosis than the serum albumin in acute viral hepatitis.

3. A 62-year-old man with known cirrhosis with portal hypertension and splenomegaly as well as alcoholic cardiomyopathy with poorly controlled heart failure is admitted to the intensive care unit with septic shock secondary to gram-negative sepsis. He is still intubated on hospital day 6. A hematology consultation is requested now because it is noted that his INR (PT) has been progressively rising since admission, from a baseline of 1.2 to 1.9. Which of the following is *not* likely to be contributing to this patient's rising INR?

A. Cirrhosis
B. Splenomegaly with hypersplenism
C. Sepsis
D. Heart failure
E. Vitamin K deficiency

Answer: B Worsening liver function could be contributing to the rising INR in this setting but not the associated portal hypertension, splenomegaly, and hypersplenism. Sepsis could be causing DIC, and worsening right-sided heart failure could be causing passive hepatic congestion, both of which might be likewise contributing. The two major sources of vitamin K are its dietary intake and its synthesis by bacterial flora in the intestine. Both of these sources of vitamin K have been essentially eliminated in this patient because he has not been eating and his gut flora has been largely sterilized with the use of broad-spectrum antibiotics for sepsis; this is a prototypical setting for vitamin K deficiency. (Elimination of only one source of vitamin K, that is, poor nutrition or use of antibiotics, is usually not sufficient to cause clinically significant vitamin K deficiency.)

4. Which of the following is *not* a coagulation abnormality in advanced liver disease?

A. Dysfibrinogenemia
B. Disseminated intravascular coagulation
C. Acquired factor VIII inhibitor
D. Increased fibrinolytic activity
E. Decreased synthesis of coagulation factors

Answer: C The pathophysiologic mechanism of bleeding in liver failure is complex and multifactorial. Impaired liver function reduces the synthesis of all coagulation factors because that organ is the major site of their synthesis. Because the liver is also the major site of clearance of trace amounts of circulating activated clotting factors, liver failure is associated with increased levels of these, leading to essentially a state of DIC. (It is why it is so difficult and sometimes impossible to distinguish the coagulopathy of liver failure from the coagulopathy of DIC; they may in fact often coexist in liver failure.) Enhanced fibrinolytic activity, partly secondary to DIC, is part of the picture in liver disease, as is the development of dysfibrinogenemia (functional abnormality of fibrinogen).

176 THROMBOTIC DISORDERS: HYPERCOAGULABLE STATES

ANDREW I. SCHAFER

The hypercoagulable states, also referred to as thrombophilias, encompass a group of inherited or acquired conditions that are associated with an increased risk for thrombosis.

The primary hypercoagulable states are quantitative or qualitative abnormalities in specific coagulation proteins that induce a prothrombotic state. Most of these disorders involve inherited mutations and polymorphisms that lead to either a deficiency of a physiologic antithrombotic factor (typically associated with a loss-of-function mutation) or an increased level of a prothrombotic factor (typically associated with a gain-of-function mutation) (Table 176-1). Particularly when they are combined with other inherited prothrombotic mutations (multigene interactions), these primary hypercoagulable states are associated with a lifelong predisposition to thrombosis. The secondary hypercoagulable states, a diverse group of mostly acquired conditions, cause a thrombotic tendency by more complex, often multifactorial mechanisms. The trigger for a discrete, clinical thrombotic event is often the development of one of the acquired, secondary hypercoagulable states superimposed on an inherited state of hypercoagulability (Fig. 176-1).

PRIMARY HYPERCOAGULABLE STATES

Antithrombin III Deficiency

EPIDEMIOLOGY AND PATHOBIOLOGY

Inherited quantitative or qualitative deficiency of antithrombin III leads to increased fibrin accumulation and a lifelong propensity to thrombosis (Chapter 171). Antithrombin is the major physiologic inhibitor of thrombin and other activated coagulation factors; therefore, its deficiency leads to unregulated protease activity and fibrin formation.[1]

The frequency of asymptomatic heterozygous antithrombin deficiency in the general population may be 1 in 350. Most of these individuals have clinically silent mutations and never have thrombotic manifestations. The frequency of symptomatic antithrombin deficiency in the general population has been estimated to be between 1 in 2000 and 1 in 3000. Among all patients seen with venous thromboembolism (VTE), antithrombin deficiency is detected in only about 1 to 2%, but it is found in approximately 2.5% of selected patients with recurrent thrombosis or onset of thrombosis at a younger age (<45 years old).

Patients with type I antithrombin deficiency have proportionately reduced plasma levels of antigenic and functional antithrombin that result from a quantitative deficiency of the normal protein. Impaired synthesis, defective secretion, or instability of antithrombin in type I antithrombin-deficient individuals is caused by major gene deletions, single nucleotide changes, or short insertions or deletions in the antithrombin gene. Patients with type II antithrombin deficiency have normal or nearly normal plasma antigen accompanied by low activity levels, indicating a functionally defective molecule. Type II deficiency is usually caused by specific point mutations leading to single amino acid substitutions that produce a dysfunctional protein. More than 250 different mutations causing type I or type II antithrombin deficiency have been recognized to date.

The pattern of inheritance of antithrombin deficiency is autosomal dominant. Most affected individuals are heterozygotes whose antithrombin levels are typically about 40 to 60% of normal. These individuals may have the full clinical manifestations of lifelong hypercoagulability. Rare homozygous antithrombin-deficient patients generally have type II deficiency with reduced heparin affinity, a variant that is associated with a low risk for thrombosis in its heterozygous form; other forms of homozygous antithrombin deficiency are probably incompatible with life.

TABLE 176-1 PRIMARY HYPERCOAGULABLE STATES

DEFICIENCY OF ANTITHROMBOTIC FACTORS
Antithrombin (III) deficiency
Protein C deficiency
Protein S deficiency
INCREASED PROTHROMBOTIC FACTORS
Factor Va (activated protein C resistance; factor V Leiden)
Prothrombin (prothrombin G20210A mutation)
Factor IX Padua (factor IX R338L mutation)
Factors VII, XI, IX, VIII; von Willebrand factor; fibrinogen (epidemiologic data; molecular mechanisms unknown)

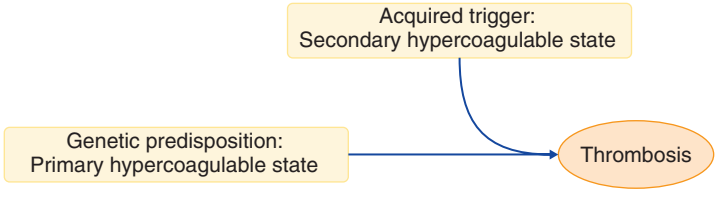

FIGURE 176-1. General scheme of thrombosis pathogenesis. A clinical episode of thrombosis is triggered by an acquired trigger, often one of the secondary hypercoagulable states, superimposed on a genetic predisposition caused by a primary hypercoagulable state. The magnitude of the acquired trigger varies from major (e.g., knee or hip surgery) to minor (e.g., lengthy air travel) to subclinical (not overt and identifiable). Likewise, the magnitude of the lifelong genetic predisposition varies from major (e.g., heterozygous antithrombin III, protein C, or protein S deficiency; homozygous factor V Leiden or prothrombin gene mutation; or two or more primary hypercoagulable states) to minor (e.g., heterozygous factor V Leiden or prothrombin gene mutation).

Protein C Deficiency

Protein C deficiency leads to unregulated fibrin generation because of impaired inactivation of factors VIIIa and Va, two essential cofactors in the coagulation cascade.

EPIDEMIOLOGY

The prevalence of heterozygous protein C deficiency in the general population is about 1 per 200 to 500. Protein C deficiency is found in 2 to 5% of all patients with VTE.

PATHOBIOLOGY

As with antithrombin deficiency, two general forms of protein C deficiency are recognized: type I, in which quantitative deficiency of the protein is associated with a proportionate decrease in protein C antigen and activity; and type II, in which qualitative defects in protein C are associated with disproportionately reduced protein C activity relative to antigen. More than 270 mutations are known to cause protein C deficiency. In the more common type I deficiency, frameshift, nonsense, or missense mutations cause premature termination of synthesis or loss of protein C stability. In type II deficiency, different mutations can cause abnormalities in protein C activation or function. The mode of inheritance of protein C deficiency is autosomal dominant. As in antithrombin deficiency, most affected individuals are heterozygotes. Neonatal purpura fulminans, a very rare complication involving widespread and sometimes fatal thrombosis, occurs in homozygous protein C– or protein S–deficient individuals.

Protein S Deficiency

Protein S is the principal cofactor of activated protein C (APC), and its deficiency mimics that of protein C in causing loss of regulation of fibrin generation by impaired inactivation of factors VIIIa and Va. A population-based study showed that low levels of free protein S and total protein S could only marginally identify subjects at risk for venous thrombosis. Only when cutoff levels for free protein S were far below the normal range, or when unprovoked venous thrombosis was considered an outcome event, was even just a two-fold to five-fold increased risk found.[2]

EPIDEMIOLOGY

Protein S deficiency is estimated to occur in about 1 in 500 in the general population. Its frequency in all patients evaluated for VTE (1 to 3%) is comparable to that of protein C deficiency.

PATHOBIOLOGY

Protein S circulates in plasma partly in complex with C4b-binding protein; only free protein S, which normally constitutes about 35 to 40% of total protein S, can function as a cofactor of APC. As in antithrombin and protein

C deficiencies, quantitative (type I) and qualitative (type II) forms of inherited protein S deficiency are known. In addition, type III protein S deficiency is characterized by normal plasma levels of total protein S but low levels of free protein S.

More than 220 mutations of the protein S gene have been found to cause a deficiency state to date. Most involve frameshift, nonsense, or missense point mutations.

Activated Protein C Resistance (Factor V Leiden)

EPIDEMIOLOGY

The factor V Leiden mutation is remarkably frequent (3 to 8%) in healthy white populations of European ancestry but is far less prevalent in certain black and Asian populations. It causes APC resistance. In various studies, APC resistance was found in a wide range of frequencies (10 to 64%) in patients with VTE.

PATHOBIOLOGY

Almost all subjects with functional APC resistance have a single, specific point mutation in the gene for factor V, which is a critical target of the physiologic anticoagulant action of APC. In this mutation, termed factor V Leiden, guanine is replaced with adenine at nucleotide 1691 (G1691A), which leads to the amino acid substitution of Arg504 by Gln and renders factor Va incapable of being inactivated by APC. Heterozygosity for the autosomally transmitted factor V Leiden mutation increases the risk for thrombosis by a factor of 5 to 10, whereas homozygosity increases the risk by a factor of 50 to 100.

Prothrombin Gene Mutation (Prothrombin G20210A)

The substitution of G for A at nucleotide 20210 of the prothrombin gene has been associated with elevated plasma levels of prothrombin and an increased risk for venous thrombosis. The allele frequency for this gain-of-function mutation is 1 to 6% in white populations, but it is much less prevalent in other racial groups. The prothrombin G20210A mutation is found in 3 to 8% of all patients with VTE.

Other Primary Hypercoagulable States

The first X-linked thrombophilia was reported in a family with a gain-of-function mutation in the gene for coagulation factor IX (factor IX Padua). Juvenile thrombophilia in this family was associated with a five- to ten-fold increase in factor IX clotting activity (with normal quantitative protein levels). Hereditary thrombosis in a Japanese family was associated with a missense mutation in the prothrombin gene (prothrombin Yukuhashi) that causes impaired inhibition of its mutant thrombin product by antithrombin. Elevated levels of factor VIII coagulant activity are a significant risk factor for venous thrombosis and its recurrence, and family studies suggest that high factor VIII levels are often genetically determined.[3] Increased levels of factor VII, factor IX, factor XI, fibrinogen, von Willebrand factor, and thrombin-activatable fibrinolysis inhibitor as well as very low levels of tissue factor pathway inhibitor may also confer increased risk. Many other inherited abnormalities of specific physiologic antithrombotic systems may be associated with a thrombotic tendency. Most of these conditions are limited to case reports or family studies, their molecular genetic bases are less well defined, and their prevalence rates are unknown but are probably much lower than those of the disorders described earlier. The other primary hypercoagulable states include heparin cofactor II deficiency, dysfunctional thrombomodulin, and many fibrinolytic disorders that lead to impaired fibrin degradation, including hypoplasminogenemia, dysplasminogenemia, plasminogen activator deficiency, and certain dysfibrinogenemias that cause a thrombotic rather than a bleeding diathesis.

CLINICAL MANIFESTATIONS

The primary hypercoagulable states are associated with predominantly venous thromboembolic complications (see Table 171-2). Deep venous thrombosis (DVT) of the lower extremities and pulmonary embolism (PE) are the most frequent clinical manifestations. Venous thromboses occurring in more unusual sites include superficial thrombophlebitis and mesenteric and cerebral venous thrombosis (see Table 171-2). Arterial thrombosis involving the coronary, cerebrovascular, and peripheral circulations is not generally linked to any of the primary hypercoagulable states. However, venous thrombosis can result in arterial occlusion by paradoxical embolism across a patent foramen ovale. Also, epidemiologic studies have revealed increased overall risk of arterial thrombotic and atherosclerotic vascular disease in patients with a history of VTE and vice versa.[4] Recurrent pregnancy loss is probably increased in primary hypercoagulable states, but this association is not as strongly established as it is in antiphospholipid syndrome (see later under Secondary Hypercoagulable States).

The initial episode of VTE can occur at any age in patients with primary hypercoagulable states, but it typically takes place in early adulthood. Positive family histories of thrombosis can frequently be elicited. The risk for thrombosis varies among the individual primary hypercoagulable states and is highest in patients with deficiencies of antithrombotic factors (Table 176-2); it is markedly increased with the coexistence of multiple prothrombotic mutations. Patients with homozygous deficiency states tend to have more severe thrombotic complications. Warfarin-induced skin necrosis (Fig. 176-2) infrequently complicates the initiation of oral anticoagulant therapy in patients with heterozygous protein C or protein S deficiency. Because both these proteins depend on vitamin K for normal function, their plasma levels in patients with inherited deficiency states may drop to nearly zero within a few days of starting therapy with warfarin, a vitamin K antagonist, and lead to a transient prothrombotic imbalance and skin necrosis caused by dermal vascular thrombosis. Nevertheless, oral anticoagulation does provide effective long-term antithrombotic prophylaxis in these individuals.

In most patients with primary hypercoagulable states, discrete clinical thrombotic complications appear to be precipitated by acquired prothrombotic events (e.g., pregnancy, use of oral contraceptives, surgery, trauma, immobilization), many of which are the secondary hypercoagulable states discussed subsequently (see Fig. 176-1). For example, thrombosis complicates pregnancy, especially during the puerperium, in about 30 to 60% of women with antithrombin deficiency, 10 to 20% with protein C or protein S deficiency, and almost 30% with APC resistance (factor V Leiden) unless prophylactic anticoagulation is administered during this period.

TABLE 176-2 INHERITED HYPERCOAGULABLE STATES (THROMBOPHILIAS): EPIDEMIOLOGY AND RISK OF VENOUS THROMBOEMBOLISM (VTE)

	PREVALENCE (%)		RELATIVE RISK	
THROMBOPHILIA	*General Population**	*Unselected VTE*	*First VTE*	*Recurrent VTE*
Antithrombin deficiency	0.02-0.3	1-2	5-8	2.5
Protein C deficiency	0.2-0.5	2-5	5-8	2.5
Protein S deficiency	0.5	1-3	1.7-8	2.5
Factor V Leiden	3-8	10-65	5-10†	1.3
Factor II G20210A	1-6	3-8	1.5-3.8	1.4
Factor V Leiden and factor II G20210A	0.01	—	20-60	2.5

*Data refer to white populations. The prevalence of factor V Leiden and factor II G20210A is less than 0.1% in African, African American, and Asian populations.
†Relative risk of first VTE with homozygous factor V Leiden is up to 10-fold higher (50-100).Modified from Coppola A, Tufano A, Cerbone AM, et al. Inherited thrombophilia: implications for prevention and treatment of venous thromboembolism. *Semin Thromb Hemost.* 2009;35:683-694.

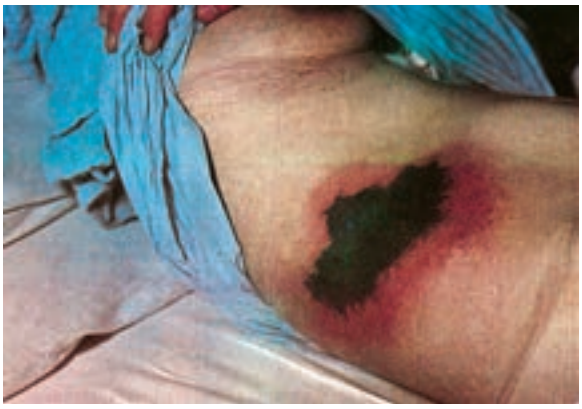

FIGURE 176-2. Acute skin necrosis in a patient with protein C deficiency who was treated with heparin and warfarin for deep venous thrombosis that occurred after elective hip surgery. Warfarin treatment was withdrawn and anticoagulation continued with heparin. Skin grafting of the affected area was required. (From Forbes CD, Jackson WF. Color Atlas and Text of Clinical Medicine. 3rd ed. London: Mosby; 2003.)

Although DVT and PE have been considered to simply reflect the clinical spectrum of manifestations of the same single entity (i.e., VTE), some distinct patterns for the two clinical events have become evident. When patients initially present with PE, regardless of cause, they are more likely to have PE again if VTE recurs, and likewise for DVT if VTE recurs.[5] In addition, certain hypercoagulable states have been noted to preferentially be manifested with either DVT or PE. Factor V Leiden poses a clearly higher risk for DVT than for PE (the so-called factor V Leiden paradox); conversely, individuals with underlying pulmonary disease (e.g., chronic obstructive pulmonary disease, pneumonia, or sickle cell disease) have a disproportionately higher likelihood of presenting with PE than with DVT as the primary manifestation of VTE.[6] The reasons that some individuals with VTE appear to be more embolism prone than others are not known.

DIAGNOSIS

Laboratory diagnosis (Chapter 171) of the primary hypercoagulable states requires testing for each of the disorders individually because no general screening test is currently available to determine whether a patient may have such a condition.[7] Factor V Leiden can be diagnosed by a DNA-based assay or by a functional test for APC resistance; the DNA-based assay is required to distinguish between heterozygous and homozygous states. DNA-based assay is required to identify the prothrombin G20210A mutation. In contrast, many different mutations have been found in the genes for antithrombin, protein C, and protein S, so DNA-based tests are not practical for the diagnosis of these inherited deficiency states. Therefore, they are detected by functional and immunologic tests. Because type II (qualitative) deficiencies of antithrombin, protein C, and protein S exhibit normal immunologic levels, functional assays for these proteins are better screening tests.

The diagnostic approach to a patient with a documented episode of VTE includes a thorough history, physical examination, and basic laboratory tests to search for possible precipitating conditions or triggers for the acute event and to rule out underlying acquired causes of hypercoagulability, including occult malignant disease, as discussed later in the section on secondary hypercoagulable states.

As noted in Chapter 171 (section under Evaluation of the Patient with a Possible Hypercoagulable State), the lack of evidence of clearly increased risk of thrombosis *recurrence* in patients with a history of VTE associated with most of the primary hypercoagulable states (thrombophilias) has made thrombophilia testing after a first episode debatable. This controversy is reflected by current practice guidelines.[8-11] An algorithm for thrombophilia testing is shown in Figure 176-3.

A reasonable diagnostic approach at this time is to screen at least all "strongly thrombophilic" patients after an initial episode of VTE: individuals with a documented event before 40 years of age, a positive family history, or unprovoked or recurrent VTE. Thrombophilia testing is not routinely recommended, however, indiscriminately in all patients after a first VTE episode or in those with overt provoking factors. Analysis of results from a randomized trial showed that recurrent VTE was not increased, at least during warfarin therapy, in the presence of factor V Leiden, prothrombin G20210A mutation, antithrombin deficiency, or elevated levels of factor VIII, factor XI, or homocysteine.[■] However, there have been no controlled clinical trials to date that have assessed the benefit of testing for thrombophilia on the risk of recurrent VTE. Although indefinite anticoagulation is recommended for patients who have had two or more VTE events, regardless of whether a primary hypercoagulable state is found, testing of these individuals for thrombophilia is still useful to guide family screening strategies (see Fig. 176-3). Individuals with arterial thrombosis generally should not be tested for any of these disorders because primary hypercoagulable states (see Table 176-1) are not clearly associated with an increased risk for arterial thrombosis. In contrast, some of the secondary hypercoagulable states, including hyperhomocysteinemia and the antiphospholipid syndrome (see later), are associated with an increased risk for arterial as well as venous thrombosis.

In general, testing for primary hypercoagulable states is not recommended immediately after a major thrombotic event. Optimally, it should be performed in clinically stable patients at least 2 weeks after completion of oral anticoagulation following a thrombotic episode. This is because active thrombosis may transiently consume and deplete some of the antithrombotic proteins in plasma and lead to the erroneous diagnosis of inherited antithrombin, protein C, or protein S deficiency. In addition to acute thrombosis, pregnancy, estrogen use, liver disease, and disseminated intravascular coagulation may cause acquired deficiencies of antithrombin, protein C, or protein S. Anticoagulation may also interfere with some of the functional tests for primary hypercoagulable states. Heparin treatment can cause a decline in antithrombin levels to the deficiency range even in normal individuals. In contrast, warfarin can elevate antithrombin levels into the normal range in patients who do have an inherited deficiency state. Warfarin therapy, by its very mode of action, also predictably reduces the functional levels and, less prominently, the immunologic levels of protein C and protein S. This warfarin action thereby potentially leads to a misdiagnosis of inherited deficiency. When testing is indicated in patients in whom interruption of prophylactic oral anticoagulation is considered to be too risky, protein C and protein S levels can be determined after warfarin therapy has been discontinued under heparin coverage for at least 2 weeks.

As noted previously, functional assays are the best screening tests for antithrombin, protein C, and protein S deficiencies because they detect both quantitative and qualitative defects; antigenic (immunologic) assays detect only quantitative deficiencies of these proteins. Functional coagulation assays for protein C and protein S may yield spuriously low values, however, if APC resistance is present. APC resistance can be diagnosed by newer high-sensitivity and high-specificity coagulation assays or by DNA analysis of peripheral blood mononuclear cells for the factor V Leiden mutation.

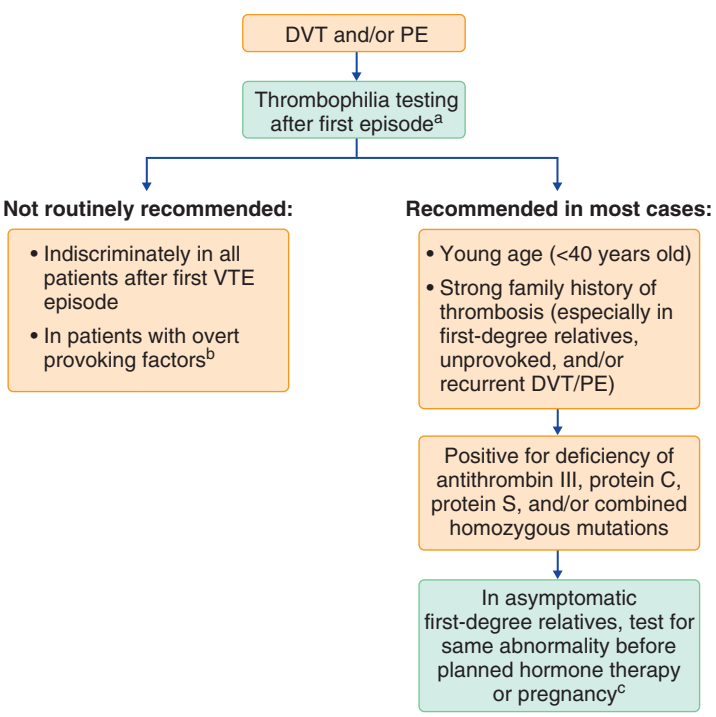

FIGURE 176-3. Algorithm for thrombophilia testing after a first episode of deep venous thrombosis (DVT) or pulmonary embolism (PE) and in asymptomatic, first-degree relatives. In the absence of strong evidence-based guidelines (see text), decisions about thrombophilia testing should be individualized, ranging from "not routinely recommended" to "recommended in most cases." VTE = venous thromboembolism. [a]Thrombophilia testing to include antithrombin III, protein C, protein S, factor V Leiden, prothrombin 20210A gene mutation, and lupus anticoagulant/antiphospholipid antibodies. [b]Provoking factors include active malignant disease, postoperative state, immobilization, trauma, active chronic inflammatory disease (e.g., extensive psoriasis), myeloproliferative neoplasm, pregnancy, oral contraceptives, and hormone replacement therapy. [c]In asymptomatic first-degree relatives of such individuals, avoidance is an alternative to testing.

TREATMENT

The initial treatment of acute venous thrombosis or PE in patients with primary hypercoagulable states is not different from that in patients without genetic defects (Chapters 38 and 81). As in patients without known thrombophilia, thrombolytic therapy should be considered after massive venous thrombosis or PE. Acute management is initiated with at least 5 days of unfractionated or low-molecular-weight heparin or fondaparinux. Oral anticoagulation with warfarin can be started on the first day of parenteral anticoagulation use and continued for at least 6 months in patients with VTE in the absence of triggering factors (e.g., postoperative state), with regulation of the dose to maintain an international normalized ratio of the prothrombin time between 2.0 and 3.0.

TABLE 176-3 LONG-TERM MANAGEMENT OF PATIENTS WITH PRIMARY HYPERCOAGULABLE STATES*

RISK CLASSIFICATION	MANAGEMENT
High risk ≥2 spontaneous thromboses 1 spontaneous life-threatening thrombosis 1 spontaneous thrombosis at an unusual site (e.g., mesenteric, cerebral venous) 1 spontaneous thrombosis in the presence of antiphospholipid syndrome, antithrombin deficiency, or more than a single hypercoagulable state	Indefinite or lifelong anticoagulation
Moderate risk 1 thrombosis with an acquired prothrombotic stimulus	Vigorous prophylaxis during high-risk situations
Asymptomatic	

*Beyond initial period of thromboprophylaxis.
Modified from Bauer K. Approach to thrombosis. In: Loscalzo J, Schafer AI, eds. Thrombosis and Hemorrhage. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2003:330-342.

Continuing oral anticoagulant prophylaxis beyond the initial 6 to 12 months after an acute episode of VTE must be weighed against continued exposure of the individual patient to the significant risk for bleeding complications. Patients with primary hypercoagulable states who have had two or more thrombotic events should receive indefinite or lifelong prophylactic oral anticoagulation (Chapter 38). Indefinite or lifelong anticoagulation is probably indicated for individuals with recurrent thrombosis even in the absence of identifiable primary hypercoagulable states.

The decision to continue prophylactic oral anticoagulation beyond the initial period after the first episode of thrombosis is more difficult (Table 176-3). After a single episode of thrombosis, patients with inherited hypercoagulable states should probably receive indefinite or lifelong anticoagulation if their initial episodes were life-threatening or occurred in unusual sites (e.g., mesenteric or cerebral venous thrombosis) or if they have more than one prothrombotic genetic abnormality. Some authorities also recommend indefinite or lifelong anticoagulation after an initial venous thromboembolic event in patients whose risk of recurrence likewise appears to be increased: those with isolated heterozygous deficiencies of antithrombin, protein C, or protein S and patients with homozygous factor V Leiden. In the absence of these characteristics, particularly if the initial episode was precipitated by a transient acquired prothrombotic situation (e.g., pregnancy, postoperative state, immobilization), it is reasonable at this time to discontinue warfarin therapy after 3 months and to administer subsequent prophylactic anticoagulation only during high-risk periods.

Asymptomatic individuals with known thrombophilia who have not had previous thrombotic complications do not require prophylactic anticoagulation except during periods of high risk for thrombosis. Because about half of the first-degree relatives of a patient with a primary hypercoagulable state should be affected, these individuals should be counseled about the implications of testing them and potentially making a diagnosis.

Management of pregnancy in women with primary hypercoagulable states requires special consideration because of the high risk for thrombosis, particularly during the puerperium.[12] Women with thrombophilia who have previously had thrombosis—and probably also asymptomatic women with thrombophilia—should receive prophylactic anticoagulation throughout pregnancy and for at least 6 weeks post partum, a particularly high-risk period. Coumarin derivatives cross the placenta and have the potential to cause both bleeding and teratogenic effects in the fetus; therefore, oral anticoagulants should not be used during pregnancy. Heparin does not cross the placenta and does not cause these fetal complications. Fixed-dose, low-molecular-weight heparin (instead of unfractionated heparin) is the anticoagulant of choice during pregnancy. Neither warfarin nor heparin induces an anticoagulant effect in a breast-fed infant when the drug is given to a nursing mother, so either can be given safely when indicated in the postpartum period.

Because warfarin-induced skin necrosis (see Fig. 176-2) is a rare problem, screening of all patients for inherited protein C or protein S deficiency, conditions that are known to predispose to this complication, is not indicated before starting warfarin therapy. Most cases can be avoided by not initiating warfarin therapy with high loading doses and by concomitant coverage with heparin. When the complication does occur, as manifested by painful red and subsequently dark, necrotic skin lesions (see Fig. 176-2) within a few days of starting warfarin, such therapy must be discontinued immediately, vitamin K administered, and heparin started (Chapter 38). The use of fresh-frozen plasma or purified protein C concentrate to normalize protein C levels rapidly can improve results. Despite this rare complication, warfarin is an effective, long-term prophylactic anticoagulant in patients with inherited protein C or protein S deficiency.

Antithrombin concentrate purified from normal human plasma or human recombinant antithrombin may be a useful adjunct to anticoagulation in "heparin-resistant" patients, who represent unusual cases of type II antithrombin deficiency, and in antithrombin-deficient patients with recurrent thrombosis despite adequate anticoagulation. Infusion of antithrombin concentrate can also be considered in some perioperative or obstetric settings in which anticoagulation poses an unacceptable bleeding risk.

SECONDARY HYPERCOAGULABLE STATES

DEFINITION

The secondary hypercoagulable states are diverse, mostly acquired disorders that predispose patients to thrombosis by complex, multifactorial pathophysiologic mechanisms. Many of these conditions also represent the acquired precipitating stimuli for clinical thrombotic events in individuals with a genetic predisposition (primary hypercoagulable states). Although each disorder causes thrombosis primarily through abnormalities in blood flow (rheology), the composition of blood (coagulation factors and platelet function), or the vessel wall, multiple overlapping mechanisms are operative in many of them.

Hyperhomocysteinemia

Hyperhomocysteinemia is an elevated blood level of homocysteine, a sulfhydryl amino acid derived from methionine by a transmethylation pathway (E-Fig. 176-1). Homocysteine is remethylated to methionine or catabolized to cystathionine. The major remethylation pathway requires folate and cobalamin (vitamin B_{12}) and involves the action of methylenetetrahydrofolate reductase (MTHFR); a minor remethylation pathway is mediated by betaine–homocysteine methyltransferase. Alternatively, homocysteine is converted to cystathionine in a trans-sulfuration pathway catalyzed by cystathionine β-synthase (CBS), with pyridoxine used as a cofactor.

Homozygous CBS deficiency states that lead to severe hyperhomocysteinemia (homocystinuria) (Chapter 209) cause premature arterial atherosclerotic disease and VTE as well as mental retardation, neurologic defects, lens ectopy, and skeletal abnormalities. By comparison, adults with heterozygous CBS deficiency, with resultant mild to moderate hyperhomocysteinemia, may have only venous or arterial thrombotic manifestations. Hyperhomocysteinemia resulting from inherited remethylation pathway defects usually involves reduced activity of MTHFR. In homozygous individuals with the autosomal recessive C677T mutation of the *MTHFR* gene, which occurs in 15% of certain populations, moderate hyperhomocysteinemia may occur and is correctable with folic acid, but it does not appear to be related to risk of venous thrombosis.[13] Acquired causes of hyperhomocysteinemia in adults most commonly involve nutritional deficiencies of the cofactors required for homocysteine metabolism, including pyridoxine, cobalamin, and folate.

Acquired and inherited hyperhomocysteinemia is a probable risk factor for both arterial and venous thrombosis. The mechanism of homocysteine-induced thrombosis and atherogenesis involves complex, probably multifactorial effects on the vessel wall. Homocysteine can cause vascular endothelial injury, conversion of the endothelial surface of blood vessels from an antithrombotic to a prothrombotic state, and smooth muscle cell proliferation. These toxic effects of homocysteine on the vessel wall may be mediated by oxidant stress.

Vitamin supplementation with folate, pyridoxine, and cobalamin can normalize elevated blood levels of homocysteine. However, several prospective clinical trials of homocysteine-lowering therapy have failed to show reduced rates of vascular events in patients with established vascular disease. It remains to be determined whether this disappointing lack of clinical benefit with homocysteine-lowering vitamin therapy indicates that homocysteine is not a direct atherogenic factor, or that vitamin therapy in this setting might have other, offsetting deleterious effects, or that possibly other mechanisms are operative.

Malignant Disease

Multiple abnormalities of hemostasis are involved in the hypercoagulable state in cancer patients, many of which initiate a systemic process of chronic disseminated intravascular coagulation (Chapter 175). The thrombotic tendency of patients with cancer may also be related to mechanical factors, such as immobility, indwelling central venous catheters, or a bulky tumor mass compressing vessels, and to comorbid conditions, such as sepsis, surgery,

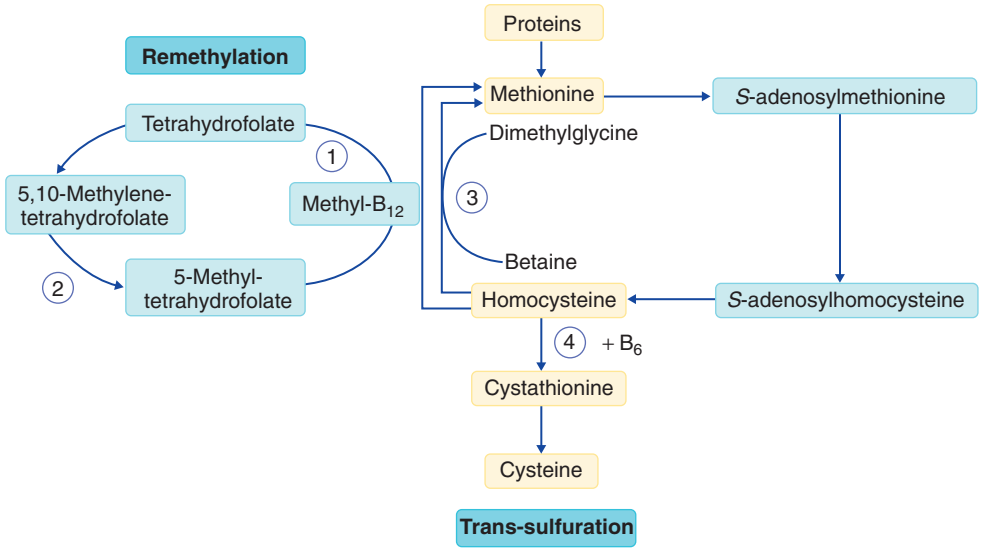

E-FIGURE 176-1. Intracellular metabolism of homocysteine occurs through remethylation to methionine or trans-sulfuration to cysteine. Numbered circles indicate the principal enzymes involved: (1) methionine synthase; (2) 5,10-methylenetetrahydrofolate reductase; (3) betaine–homocysteine methyltransferase; (4) cystathionine β-synthase. (Modified from De Stefano V, Finazzi G, Mannucci PM. Inherited thrombophilia: pathogenesis, clinical syndromes, and management. *Blood.* 1996;87:3531-3544.)

liver dysfunction secondary to metastases, and the prothrombotic effects of certain antineoplastic agents.

The incidence of thrombotic complications in cancer patients depends in part on the type of malignant disease. Hypercoagulability is most prominent in patients with pancreatic cancer (Chapter 194), adenocarcinoma of the gastrointestinal tract (Chapters 192 and 193), lung (Chapter 191), ovarian cancer (Chapter 199), and hematologic malignant neoplasms. The presence of underlying malignant disease compounds the independent risk for thrombosis in the postoperative state. There is a two-fold increase in risk of postoperative thrombosis (see later section, Postoperative State, Immobilization, and Trauma) in cancer patients compared with noncancer patients. Thrombosis most commonly occurs in patients with established or concurrently diagnosed malignant disease. In these patients, the risk of venous thrombosis has been found to be highest in the first few months after the diagnosis of malignant disease, then decreases progressively during the subsequent 15 years. The same large case-control study showed that the risk of venous thrombosis in cancer patients is approximately 12- to 17-fold increased in those who also have the factor V Leiden or the prothrombin G20210A mutation.

The increased risk of harboring an undiagnosed, usually occult malignant neoplasm in patients who present with VTE is now well established. The overall prevalence of undiagnosed cancer in patients with unprovoked VTE is approximately 6% at the time of thrombosis and 10% within the first year after thrombosis. Beyond thorough history, physical examination, and routine laboratory tests and radiography, the benefit of extensive evaluation for occult cancer including advanced imaging studies is not established, however.[14] An algorithm for evaluating a patient for occult malignant neoplasm after an episode of VTE is shown in Figure 176-4.

The most frequent thrombotic manifestations in patients with neoplasms are DVT and PE, but more unusual and distinctive thrombotic complications are also found. Trousseau's syndrome, characterized by migratory superficial thrombophlebitis of the upper or lower extremities, is strongly linked to cancer. Nonbacterial thrombotic endocarditis involves fibrin-platelet vegetations on heart valves, which produce clinical manifestations by systemic embolization (Chapter 60). Of patients with nonbacterial thrombotic endocarditis, 75% have underlying malignant neoplasms at autopsy. Trousseau's syndrome and nonbacterial thrombotic endocarditis are highly associated with adenocarcinomas. The occurrence of either syndrome in patients without known cancer demands a more vigorous search for occult malignant disease than in patients with DVT or PE. Thrombotic microangiopathy (Chapter 172), characterized by hemolysis with red blood cell fragmentation, thrombocytopenia, and microvascular thrombosis with involvement of target organs, occurs in about 5% of patients with metastatic carcinomas, most commonly those with gastric (Chapter 192), lung (Chapter 191), and breast (Chapter 198) primary sites.

Some antineoplastic agents themselves increase the risk of thrombosis in cancer patients. These include antiangiogenic agents (VTE with thalidomide and lenalidomide; arterial thrombosis with bevacizumab, sunitinib, sorafenib). In multiple myeloma (Chapter 187), the thrombotic risk of thalidomide is increased 10-fold when the drug is combined with high-dose dexamethasone and an anthracycline. Other thrombogenic chemotherapeutic agents include all-*trans*-retinoic acid and arsenic trioxide, particularly when used as induction chemotherapy for acute promyelocytic leukemia (Chapter 183); cisplatinum (causing both venous and arterial thrombosis); L-asparaginase (which can also cause bleeding); and methotrexate.[15]

TREATMENT

Rx

Treatment of acute VTE in cancer patients should be initiated as in other patients, but subsequent prophylactic anticoagulation generally should be continued while active malignant disease is present.[16] In some patients, however, complete resolution of the VTE may allow treatment to be stopped safely. Anticoagulation can be difficult in many cancer patients; these patients may be resistant to warfarin prophylaxis. Anticoagulation can also be complicated by bleeding into tumors. Long-term treatment of cancer patients with low-molecular-weight heparin (LMWH) after VTE (Chapter 38) reduces recurrences and possibly decreases bleeding complications compared with treatment with warfarin. LMWH also reduces the risk of a first VTE in patients with cancer and is more effective than warfarin in doing so, although it has not been shown to increase overall survival in this setting. Evidence-based practice guidelines have been published for the prevention and treatment of VTE in cancer patients.

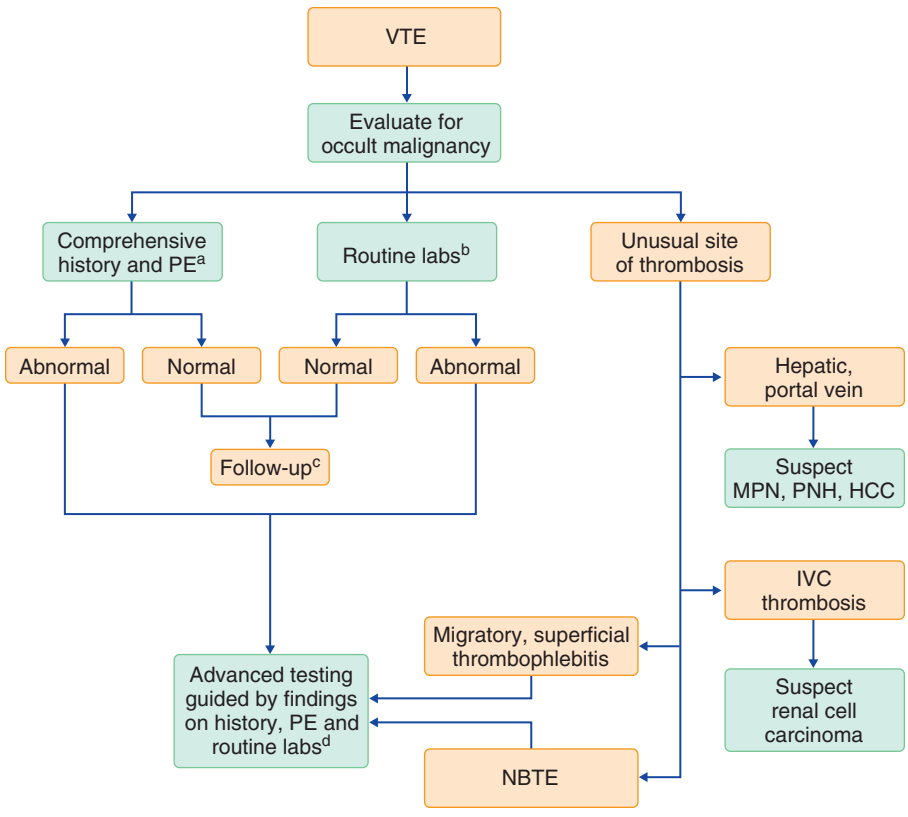

FIGURE 176-4. Algorithm for evaluating a patient for occult malignant neoplasm after an unprovoked episode of deep venous thrombosis or pulmonary embolism or thrombosis at an unusual site. HCC = hepatocellular carcinoma; IVC = inferior vena cava; MPN = myeloproliferative neoplasm; NBTE = nonbacterial thrombotic endocarditis; PE = physical examination; PNH = paroxysmal nocturnal hemoglobinuria; VTE = venous thromboembolism. [a]Physical examination to include rectal (with stool sample for occult blood) and breast and pelvic examinations for women. [b]Routine laboratory evaluations to include chest radiography, complete blood count, basic chemistries, and calcium concentration. [c]Follow-up suggested at 6 to 12 months. [d]Advanced testing could include advanced imaging, endoscopies, cytologies, and the like.

Myeloproliferative Neoplasms and Paroxysmal Nocturnal Hemoglobinuria

Thrombosis and, apparently paradoxically, bleeding are major causes of morbidity and mortality in patients with chronic myeloproliferative neoplasms (Chapter 166) and the related bone marrow stem cell disorder paroxysmal nocturnal hemoglobinuria (Chapter 160). In uncontrolled polycythemia vera (Chapter 166), increased whole blood viscosity contributes to the thrombotic tendency. Thrombocytosis, abnormal platelet function, and other less well understood factors are also probably involved in the hemostatic defect of the myeloproliferative neoplasms and paroxysmal nocturnal hemoglobinuria.

In addition to DVT and PE, some distinctive thrombotic manifestations are seen. Hepatic vein thrombosis (Budd-Chiari syndrome) and portal and other intra-abdominal venous thromboses (Chapter 143) are associated with myeloproliferative neoplasms and paroxysmal nocturnal hemoglobinuria (Chapter 160) and may be the initial manifestations of the disease. Myeloproliferative neoplasms, particularly polycythemia vera (Chapter 166) and essential thrombocythemia (Chapter 166), may cause erythromelalgia, a syndrome of microvascular thrombosis manifested by intense pain accompanied by warmth, duskiness, and mottled erythema, sometimes resembling livedo reticularis, in a patchy distribution in the extremities, most prominently in the feet; digital microvascular ischemia progressing to vascular insufficiency and gangrene may ensue (Chapter 80). A wide spectrum of neurologic manifestations may be caused by cerebrovascular ischemia, especially in patients with essential thrombocythemia.

TREATMENT

Rx

Treatment of VTE in patients with the myeloproliferative neoplasms and paroxysmal nocturnal hemoglobinuria should be initiated as in patients without these hematologic disorders. In patients with thrombosis associated with polycythemia vera, the hematocrit should be maintained in the normal range with phlebotomies or chemotherapy, or with both (Chapter 166). Patients maintained at a hematocrit target of less than 45% have a significantly lower rate of cardiovascular death and major thrombosis than do those with hematocrit targets of 45 to 50%. Low-dose aspirin (100 mg daily) can prevent thrombotic complications without increasing the incidence of major bleeding in patients with polycythemia vera who have no contraindications to such treatment. In patients with essential thrombocythemia, cytoreduction of the elevated platelet count should be achieved with chemotherapy (Chapter 166).

Antiphospholipid Syndrome

Antiphospholipid syndrome (Chapter 174) is characterized by venous and arterial thrombosis, recurrent spontaneous pregnancy loss (which may also be due to thrombosis), thrombocytopenia, and a variety of neuropsychiatric manifestations.[17] The syndrome is associated with a heterogeneous group of autoantibodies that bind to anionic phospholipid-protein complexes, a protein cofactor of which is β_2-glycoprotein I. Patients with this syndrome have any combination of positive responses to tests to detect different plasma antiphospholipid-protein antibodies (e.g., anticardiolipin antibodies) and phospholipid-based clotting tests (lupus anticoagulants) (Chapter 257). The predominant prothrombotic effects of these antibodies are probably directed to the vessel wall.

DVT and PE are the most frequent venous thrombotic events in these patients. Cerebrovascular events are the most common arterial thrombotic complications and are manifested as stroke, transient ischemic attacks (Chapter 407), multi-infarct dementia (Chapter 402), or retinal artery occlusion. Peripheral and intra-abdominal vascular occlusion is encountered more rarely. About one third of these patients have nonbacterial heart valve vegetations (Libman-Sacks endocarditis). The most prominent obstetric complications are recurrent spontaneous pregnancy loss and fetal growth retardation, which are probably due to thrombosis of placental vessels. Patients are occasionally seen with "catastrophic" antiphospholipid syndrome involving a series of acute and sometimes fatal vascular occlusive events, or "thrombotic storm." Thrombotic complications are limited largely to patients with primary antiphospholipid syndrome and patients in whom the antibodies are associated with collagen vascular disease, not with drugs or infections.

TREATMENT AND PREVENTION

Rx

Acute management of thrombosis in these patients is essentially the same as in other individuals. Monitoring of heparin anticoagulation is difficult in patients with a lupus anticoagulant because they already have a prolonged activated partial thromboplastin time at baseline; the use of low-molecular-weight heparin, which does not require monitoring, can circumvent this problem. Warfarin is effective in preventing recurrent thrombosis but usually requires prolonged or indefinite therapy with doses to achieve an international normalized ratio of 2.0 to 3.0. Whereas laboratory evidence of antiphospholipid antibodies in patients with a first episode of VTE is still usually considered an indication for indefinite anticoagulant therapy, a systematic review showed that the strength of this association is uncertain because of low-quality evidence supporting it.[18] No established treatment of women with antiphospholipid syndrome has been shown to prevent recurrent fetal loss. Treatment with prednisone and aspirin during pregnancy is not effective in promoting live birth and may increase the risk for prematurity.

Asymptomatic carriers of a "high-risk profile" of antiphospholipid antibodies (confirmed triple positive for anticardiolipin, lupus anticoagulant, and anti–β_2-glycoprotein I) are at considerable risk for a first thromboembolic event; some such individuals should even be considered for primary thromboprophylaxis.[19]

Pregnancy, Oral Contraceptives, and Hormone Replacement Therapy

Activation of the coagulation system is initiated locally in the maternal uteroplacental circulation, where the placenta is the source of increased thrombin generation. Platelet activation and increased platelet turnover also occur during normal pregnancy, and about 8% of healthy women have mild thrombocytopenia at term. Simultaneously, the fibrinolytic system is progressively blunted throughout pregnancy because of the action of placental plasminogen activator inhibitor type 2. The net effect of these coagulation changes is creation of a state of hypercoagulability that makes pregnant women vulnerable to thrombosis, particularly in the puerperium. These systemic alterations are compounded by prothrombotic mechanical and rheologic factors in pregnancy, including venous stasis in the legs caused by the gravid uterus, pelvic vein injury during labor, and the trauma of cesarean section. Oral contraceptives induce a prothrombotic state by increasing procoagulant effects and decreasing physiologic anticoagulant effects.

The use of oral contraceptives is associated with an increased risk for venous thrombosis, myocardial infarction, stroke, and peripheral arterial disease, particularly during the first year of use (Chapter 238). Unexpectedly, third-generation oral contraceptives, which contain less estrogen and a different progestin, double the risk for VTE in comparison to second-generation preparations. Postmenopausal hormone replacement increases the risk for deep VTE by a factor of 2 to 3.5, at least during the first year. Hormone replacement therapy has no beneficial effect and possibly even a detrimental effect on the risk for arterial disease (Chapter 240).

DVT and PE are the most common thrombotic complications of pregnancy and the use of oral contraceptives or hormone replacement therapy. Coexisting primary hypercoagulable states are an at least additive risk factor in all these settings. In the absence of a clear family history of clinical VTE or a strongly thrombophilic mutation in a first-degree relative, there is little justification, however, to screen for prothrombotic mutations with pregnancy or before starting hormone replacement therapy or oral contraceptives. Increasing age, increasing parity, cesarean delivery, prolonged bedrest or immobilization, obesity, and previous thromboembolism are additional prothrombotic risk factors in pregnant women. Most thrombotic events associated with pregnancy occur in the peripartum period, especially after delivery. Special considerations for anticoagulation in the setting of pregnancy are noted in the section on treatment of primary hypercoagulable states.

Systemic Inflammation

Systemic inflammation may be involved in the pathogenesis of VTE through vessel wall inflammation and damage as well as systemic hypercoagulability. Patients with autoimmune connective tissue diseases, particularly rheumatoid arthritis,[20] juvenile rheumatoid arthritis, and systemic lupus erythematosus, are at increased risk of VTE. Psoriasis, an immunoinflammatory disease that is associated with cardiovascular risk and atherothrombotic

events, has also been shown in a cohort study to carry an increased risk of VTE; the risk was found to be highest in younger patients with severe psoriasis. Other skin diseases that are autoimmune in etiology were not associated with VTE risk in another population-based case-control study.

Postoperative State, Immobilization, and Trauma

Postoperative thrombosis (Chapter 433) is caused by a combination of local mechanical factors, including decreased venous blood flow in the lower extremities, and systemic changes in coagulation. The level of risk for postoperative thrombosis depends largely on the type of surgery performed. It is compounded by coexisting risk factors, such as an underlying inherited primary hypercoagulable state or malignant disease (see earlier sections), advanced age, and prolonged procedures. Postoperative DVT and PE, the most common thrombotic complications, are often asymptomatic and detectable only by noninvasive studies. The incidence of DVT after general surgical procedures is about 20 to 25%, with almost 2% of these patients having clinically significant PE. Without prophylaxis, the risk for DVT after hip surgery and knee reconstruction ranges from 45 to 70%, and clinically significant PE occurs in 20% of patients undergoing hip surgery. Although the process of thrombosis generally begins intraoperatively or within a few days of surgery, the risk for this complication can be protracted beyond the time of discharge from the hospital, particularly in hip replacement patients.

Patients who are bedridden or experiencing prolonged air travel are at increased risk for VTE. VTE is also one of the most common causes of morbidity and mortality in survivors of major trauma, and asymptomatic DVT of the lower extremities has been detected by venography in more than 50% of hospitalized trauma patients. The risk for venous thrombosis after trauma is increased by advanced age, need for surgery or transfusions, and presence of lower extremity fractures or spinal cord injury.

Mechanical methods of prophylaxis against VTE should be considered in high-risk postoperative patients and bedridden patients with medical conditions, either in combination with anticoagulant prophylaxis or instead of it in patients who have an unusually high risk for bleeding with anticoagulation. Such methods include graduated compression stockings, intermittent pneumatic compression devices, and venous foot pumps.

For long-distance travelers at increased risk of VTE, including those with previous VTE or hypercoagulable states, current recommendations include frequent ambulation, nonalcoholic hydration, calf muscle exercises, and the use of properly fitted, below-knee graduated compression stockings providing 15 to 30 mm Hg of pressure at the ankle during travel; compression stockings are not recommended for other long-distance travelers. The use of aspirin or anticoagulants to prevent VTE is not recommended for long-distance travelers at increased risk of VTE.[21]

Grade A References

A1. Kearon C, Julian JA, Kovacs MJ, et al. Influence of thrombophilia on risk of recurrent venous thromboembolism while on warfarin: results from a randomized trial. *Blood*. 2008;112:4432-4436.

A2. Napolitano M, Saccullo G, Malato A, et al. Optimal duration of low molecular weight heparin for the treatment of cancer-related deep vein thrombosis: The Cancer-DACUS Study. *J Clin Oncol*. 2014;32:3607-3612.

A3. Ben-Aharon I, Stemmer SM, Leibovici L, et al. Low molecular weight heparin (LMWH) for primary thrombo-prophylaxis in patients with solid malignancies—systematic review and meta-analysis. *Acta Oncol*. 2014;53:1230-1237.

A4. Di Nisio M, Porreca E, Otten HM, et al. Primary prophylaxis for venous thromboembolism in ambulatory cancer patients receiving chemotherapy. *Cochrane Database Syst Rev*. 2014;8:CD008500.

A5. Sanford D, Naidu A, Alizadeh N, et al. The effect of low molecular weight heparin on survival in cancer patients: an updated systematic review and meta-analysis of randomized trials. *J Thromb Haemost*. 2014;12:1076-1085.

A6. Farge D, Debourdeau P, Beckers M, et al. International clinical practice guidelines for the treatment and prophylaxis of venous thromboembolism in patients with cancer. *J Thromb Haemost*. 2013;11:56-70.

A7. Marchioli R, Finazzi G, Specchia G, et al. Cardiovascular events and intensity of treatment in polycythemia vera. *N Engl J Med*. 2013;368:22-33.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hepner M, Karlaftis V. Antithrombin. *Methods Mol Biol*. 2013;992:355-364.
2. Pintao MC, Ribeiro DD, Bezemer ID, et al. Protein S levels and the risk of venous thrombosis: results from the MEGA case-control study. *Blood*. 2013;122:3210-3219.
3. Jenkins PV, Rawley O, Smith OP, et al. Elevated factor VIII levels and risk of venous thrombosis. *Br J Haematol*. 2012;157:653-663.
4. Franchini M, Mannucci PM. Association between venous and arterial thrombosis: clinical implications. *Eur J Intern Med*. 2012;23:333-337.
5. Baglin T, Douketis J, Tosetto A, et al. Does the clinical presentation and extent of venous thrombosis predict likelihood and type of recurrence? A patient-level meta-analysis. *J Thromb Haemost*. 2010;8:2436-2442.
6. Van Langevelde K, Flinterman LE, van Hylckama Vlieg A, et al. Broadening the factor V Leiden paradox: pulmonary embolism and deep-vein thrombosis as 2 sides of the spectrum. *Blood*. 2012;120:933-946.
7. Johnson NV, Khor B, Van Cott EM. Advances in laboratory testing for thrombophilia. *Am J Hematol*. 2012;87(suppl 1):S108-S112.
8. Kearon C, Akl EA, Comerota AJ, et al. Antithrombotic therapy for VTE disease: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141(suppl):e419S-e494S.
9. Chong LY, Fenu S, Stansby G, et al. Guideline Development Group. Management of venous thromboembolic disease and the role of thrombophilia testing: summary of NICE guidance. *BMJ*. 2012;345:43-45.
10. Baglin T, Gray E, Greaves M, et al. Clinical guidelines for testing for heritable thrombophilia. *Br J Haematol*. 2010;149:209-220.
11. De Stefano V, Rossi E. Testing for inherited thrombophilia and consequences for antithrombotic prophylaxis in patients with venous thromboembolism and their relatives. A review of the Guidelines from Scientific Societies and Working Groups. *Thromb Haemost*. 2013;110:697-705.
12. Bates SM, Greer IA, Middeldorp S, et al. VTE, thrombophilia, antithrombotic therapy, and pregnancy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141(suppl):e691S-736S.
13. Hickey SE, Curry CJ, Toriello HV. ACMG Practice Guideline: lack of evidence for MTHFR polymorphism testing. *Genet Med*. 2013;15:153-156.
14. Lauw MN, van Doormaal FF, Middeldorp S, et al. Cancer and venous thrombosis: current comprehensions and future perspectives. *Semin Thromb Hemost*. 2013;39:507-514.
15. McMahon BJ, Kwaan HC. Thrombotic and bleeding complications associated with chemotherapy. *Semin Thromb Hemost*. 2012;38:808-817.
16. Connors JM. Prophylaxis against venous thromboembolism in ambulatory patients with cancer. *N Engl J Med*. 2014;370:2515-2519.
17. Lim W. Thrombotic risk in the antiphospholipid syndrome. *Semin Thromb Hemost*. 2014;40:741-746.
18. Garcia D, Akl EA, Carr R, et al. Antiphospholipid antibodies and the risk of recurrence after a first episode of venous thromboembolism: a systematic review. *Blood*. 2013;122:817-824.
19. Pengo V, Ruffatti A, Legnani C, et al. Incidence of a first thromboembolic event in asymptomatic carriers of high-risk antiphospholipid antibody profile: a multicenter prospective study. *Blood*. 2011;118:4714-4718.
20. Lee JJ, Pope JE. A meta-analysis of the risk of venous thromboembolism in inflammatory rheumatic diseases. *Arthritis Res Ther*. 2014;16:435.
21. Kahn SR, Lim W, Dunn AS, et al. Prevention of VTE in nonsurgical patients: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141(suppl):e195S-e226S.

REVIEW QUESTIONS

1. A 45-year-old man was diagnosed 3 days ago to have a left lower extremity deep venous thrombosis (DVT) by ultrasound. No provoking factors were identified. He was discharged from the emergency department without further work-up but with instructions to begin subcutaneous injections of therapeutic doses of a low-molecular-weight heparin together with daily warfarin at a dose of 7.5 mg/day, to be followed up by you in your primary care office. Which of the following would you do on this visit?

A. Initiate imaging work-up for diagnosis of occult malignant disease.
B. Obtain a prothrombin time (international normalized ratio).
C. Repeat the lower extremity ultrasound.
D. Obtain pulmonary embolism protocol chest computed tomography.
E. Test for most common primary hypercoagulable states (thrombophilias).

Answer: B The emergency department physicians correctly initiated therapeutic heparin while at the same time beginning longer term warfarin oral anticoagulation to prevent recurrent venous thromboembolism (VTE). It is now time to begin regulating the warfarin dose on the basis of the prothrombin time, with a target international normalized ratio of 2.0 to 3.0. Although several studies have shown some increased risk of otherwise healthy individuals with VTE harboring an occult malignant neoplasm, a thorough history and physical examination, including stool samples for occult blood and basic studies like chest radiography and mammography in women, are considered to be sufficient. Advanced imaging studies to look for cancer are indicated only if abnormalities are found in this basic work-up. Another lower extremity ultrasound examination at this point is not indicated unless there is some uncertainty about the original diagnosis. The finding of a simultaneous, silent pulmonary embolism will not make any difference in your diagnostic and therapeutic plan. In fact, a large percentage (up to 50%) of patients with symptomatic lower extremity DVT would be found to have silent pulmonary emboli on imaging. Whatever might be found on blood tests for a primary hypercoagulable state at this point would not alter therapy, and the test results may be misleading right after an acute thrombotic event.

2. You are the primary care physician of a 50-year-old woman with a history of pregnancy-associated DVT 20 years ago. She has had no work-up for hypercoagulable states (thrombophilias). The patient has had no recurrent VTE and is not taking anticoagulants, but she is planning a nonstop flight from New York to Hong Kong. Which of the following would you recommend?

A. Start aspirin 325 mg daily until return from Hong Kong.
B. Prescribe a single dose of subcutaneous injection of low-molecular-weight heparin immediately before embarking on flights to and from Hong Kong.
C. Recommend nothing except frequent ambulation and hydration on flights.
D. Do thrombophilia testing before flights.
E. Strongly recommend interrupting flights with layovers.

Answer: C A single episode of DVT in the past that was clearly related to a provoking event (i.e., pregnancy) does not require any special preventive measures before a long-haul flight in addition to the recommendation for ample hydration (with nonalcoholic beverages), frequent ambulation, and leg exercises. There is no evidence that prophylactic aspirin or low-molecular-weight heparin coverage for the flights would significantly reduce risk of flight-related VTE, even if a thrombophilia is found on testing in this situation. There is a very rough relationship between duration of long-haul air travel and risk of VTE, but the risk in any case is very small, and there is no evidence to suggest that it requires interruption of nonstop flights.

3. Which of the following statements is incorrect with regard to cancer-associated venous thromboembolism (VTE)?

A. VTE may precede detection of an occult malignant neoplasm by months or years.
B. Risk of cancer-associated VTE is further increased by surgery.
C. Risk of cancer-associated VTE is further increased by some chemotherapies.
D. VTE risk is increased with localized as well as with advanced metastatic cancers.
E. Risk of recurrent VTE is best prevented by use of warfarin for as long as the cancer is active.

Answer: E Although prophylactic anticoagulation is recommended in most cancer patients after one or more thrombotic events, low-molecular-weight heparin has been demonstrated to be superior to warfarin in this particular setting. VTE may be an indicator of an occult malignant neoplasm months or even years before its discovery (see algorithm in Figure 176-4). Major surgery clearly increases the risk of thrombosis in a patient with cancer. The same is true for certain antineoplastic agents, including antiangiogenic agents (thalidomide, lenalidomide, bevacizumab, sunitinib, sorafenib), all-*trans*-retinoic acid and arsenic trioxide (used for treatment of acute promyelocytic leukemia), L-asparaginase, cisplatinum, and methotrexate. Thrombotic complications are more frequent in cancer patients with advanced disease but can certainly also develop in those with localized tumors.

177

TRANSFUSION MEDICINE

LAWRENCE T. GOODNOUGH

DEVELOPMENT OF TRANSFUSION MEDICINE

Concepts in blood transfusion and blood conservation include blood safety, blood inventory, and blood costs. The formation of clinical practice guidelines in patient blood management is based on effective blood utilization, new strategies in blood conservation, including pharmacologic alternatives to blood transfusion, and patients' informed consent for transfusion. The broad-based multidisciplinary constituency of transfusion medicine includes blood collection facilities, hospital-based transfusion services, research laboratories, and the commercial sector.

Patient-focused blood management is described in the Circular of Information[1] as "a professional judgment based on clinical evaluation that determines the selection of components, dosage, rate of administration. ..." Patient blood management therefore encompasses an evidence-based medical and surgical approach with preventive strategies that are emphasized to identify, evaluate, and manage anemia (e.g., pharmacologic therapy and reduced iatrogenic blood losses from diagnostic testing); to optimize hemostasis (e.g., pharmacologic therapy and point-of-care testing); and to establish decision thresholds (e.g., guidelines) for the appropriate administration of blood therapy.[2,3]

The medical director of the blood bank is responsible for issues related to blood inventory and safety, through the oversight of laboratory policies and procedures and quality management (Fig. 177-1). Management of the transfusion service includes coordinating blood transfusion and blood conservation activities and serving as a consultant to clinicians whose patients are undergoing massive transfusions, apheresis, or transplantation or who are having difficulty finding compatible blood products. Finally, the transfusion medicine specialist supervises quality assurance to satisfy regulatory and accreditation requirements.[4]

Blood Availability

Blood centers must be able to supply blood in response to acute crises. Sporadic shortages of blood and blood products (e.g., packed red cells, platelet products, albumin, intravenous immunoglobulin, and clotting factor concentrates) are potentially life threatening. Such shortages have been attributed to disruptions in production, increasingly strict criteria for accepting donors, product recalls, increased use (including off-label use), and stockpiling or other market issues.

In the United States, the Department of Health and Human Services monitors sentinel community blood services and hospital transfusion services to track blood collection and transfusion activity. Blood transfusion and collection activities peaked in 1986 and then declined until 1994. However, blood transfusion and collection subsequently increased 8.0% in 1997, 10.2% in 1999, and an estimated 4% to 5% annually thereafter.

Within the American Red Cross, which supplies about half of all blood products in the United States, a 3- to 4-day supply of red blood cell products is typical, but some independent centers have higher reserves. Use of frozen red cells as a hedge against inventory shortages has generally not been practical because the shelf life of thawed units is only 24 hours; however, an automated, functionally closed system for the glycerolization and deglycerolization processes allows a 2-week post-thaw shelf life. With such a system, many blood centers and transfusion services can expand their available red cell inventories as reserves. After disasters—human or natural—on-hand blood supplies are adequate in the short term and can be rapidly mobilized over great distances. Because of the need to maintain reserves, blood collection must routinely surpass the anticipated need for blood transfusion.[5]

Red Blood Cell Transfusion

Worldwide, more than 75 million units of whole blood are estimated to be donated every year, with yearly transfusion of 24 million blood units in the United States.

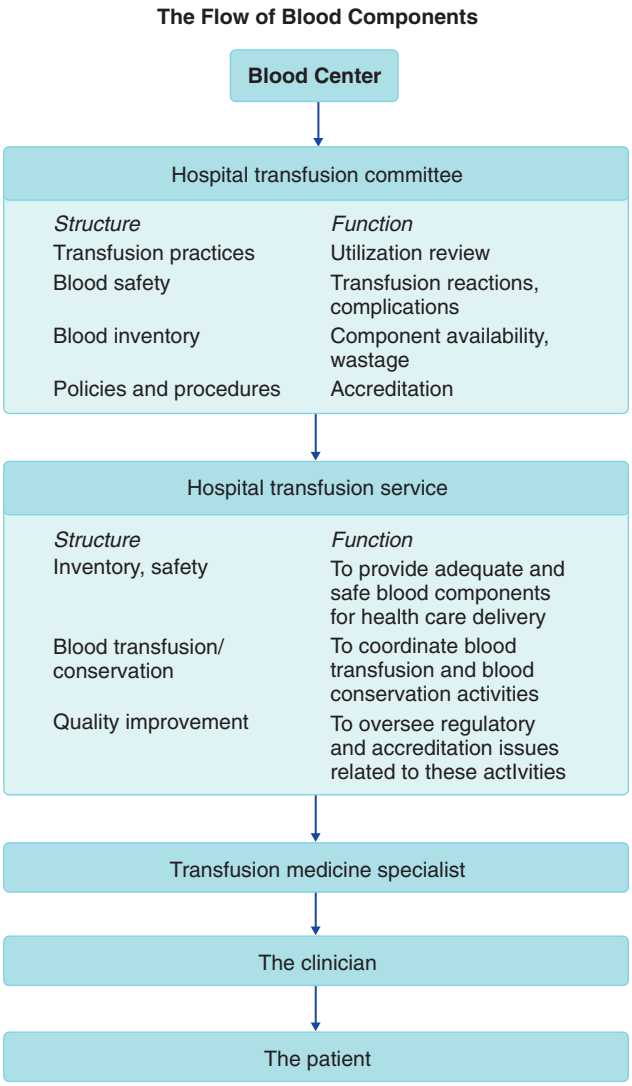

FIGURE 177-1. **A hospital's transfusion committee and transfusion service can help treating physicians manage the availability, safety, and use of blood products. (From Goodnough LT. What is a transfusion medicine specialist? *Transfusion.* 1999;39: 1031-1033.)**

Whole Blood

A unit of blood is collected as a donation of 450 mL ± 10% into a citrate anticoagulant that also contains phosphate and dextrose. The red cell and hemoglobin content is variable and dependent on the donor's hematocrit and the precise volume bled.

Whole blood is stored at 4 ± 2°C to diminish red cells' utilization of adenosine triphosphate and to preserve their viability, which should be at least 70% at the end of a shelf life of 35 days. After 10 days of storage, all predonation 2,3-diphosphoglycerate content in red cells is lost, but up to 50% is regenerated within 8 hours after transfusion.

In Western countries, whole blood is rarely used because within a few hours or days, some coagulation factors (especially factors V and VIII) and platelets decrease in quantity or lose viability. After a 7-day hold at 4°C, factor VIII levels fall to 0.32 ± 0.09 IU/mL, and factor V levels fall to 0.78 ± 0.15 IU/mL. At 4°C, platelets undergo a shape change from discoid to spherical that is irreversible after 8 hours, and their in vivo survival is reduced to 2 days.

Red Cell Components

Red cells are provided in various formats that differ with respect to the presence of additive solutions and the extent to which white cells are removed. Solutions that contain combinations of saline, adenine, phosphate, bicarbonate, glucose, and mannitol provide better red cell viability during storage and allow up to a 42-day shelf life. Red cells and red cells in additive solution can be used interchangeably, with the exception that red cells in additive solution are not recommended for exchange or massive transfusions in neonates.

Red cells should be refrigerated until the time of the transfusion because of the risk for bacterial proliferation within the pack at room temperature. Red cells that have been out of refrigeration for 30 minutes or longer cannot be returned to stock. A unit of red cells should be infused over a maximum period of 4 hours.

Irradiated Cellular Blood Components

Gamma-irradiated cellular blood components are used to prevent the occurrence of transfusion-associated graft-versus-host disease (see later).

Leukocyte-Reduced Blood Components

Blood components are depleted of leukocytes by means of filtration. A leukocyte-reduced component is defined as one with less than 5×10^6 residual white cells per liter, and 100% of tested units should meet this standard. Leukocyte removal should be performed while the cells are still intact by filtering the blood as soon as possible after collection.

Leukocyte reduction reduces the incidence of human leukocyte antigen (HLA) alloimmunization in multitransfused recipients, but immunization rates of 10 to 25% are still seen in women who have already been exposed to HLA alloantigens through previous pregnancy. Leukocyte reduction has less impact on refractoriness to random platelet donations because of the importance of nonimmune factors in the poor response to transfused platelets.

Leukocyte reduction also lowers the risk for nonhemolytic febrile reactions after the transfusion of red cells or platelets. Several studies suggest that leukocyte reduction has a beneficial impact on the rate of postoperative infection. The impact of leukocyte reduction on the incidence of transfusion-associated graft-versus-host disease is unknown.

Preventing Transmission of Cytomegalovirus

A small number of studies of prestorage leukocyte reduction have suggested its efficacy in preventing transfusion-transmitted cytomegalovirus (CMV) infection (Chapter 377), which occurs in 4% of cases, even with seronegative components. Prestorage leukocyte depletion reduces the infection rate to less than 1%, suggesting that this technology may be at least equivalent to serologic testing for CMV antibodies for the prevention of CMV transmission; however, CMV testing of leukocyte-reduced components continues to be performed by most transfusion services. Indications for CMV-seronegative components include transfusion in pregnancy, intrauterine transfusion, transfusions to neonates less than 37 weeks' gestation, transfusions to CMV-seronegative patients who are potential or actual recipients of allogeneic bone marrow or peripheral blood progenitor transplants when the donor is also CMV seronegative, and patients infected with human immunodeficiency virus (HIV).

Immunomodulatory Effects

Previous transfusion has an immunomodulatory benefit for the survival of renal allografts, even with the use of potent immunosuppressive drugs, although the exact mechanism of this benefit remains unknown. Retrospective studies suggest an adverse effect of transfusion on the rate of perioperative infections or recurrent cancers.

Typing and Crossmatching

Blood and blood components for transfusion must be compatible with the same blood type as the patient. Obtaining an accurate ABO/Rh grouping for a patient is the most significant serologic test performed before transfusion. When type-specific blood and components are unavailable or emergency circumstances do not allow their identification or use, type O-negative red cells should be used. Group O is the only choice for group O recipients and is the alternative choice for groups A, B, and AB.

Red cell antigens other than ABO and D are not routinely considered when selecting *donor* blood products for transfusion; the exception is when unexpected, clinically significant red cell antibodies are present in the patient, as determined by an antibody screen or previous identification. Red cell alloantibodies are produced by exposure to foreign red cell antigens through previous transfusion, pregnancy, or both. For an antibody to be considered clinically significant, it must be associated with a hemolytic transfusion reaction or decreased survival of transfused incompatible red cells. Most of the clinically significant antibodies are optimally reactive at 37°C or are detected by the antiglobulin test. If an antibody screen is negative in the *recipient,* the probability is greater than 99% that an ABO and Rh crossmatch will also be compatible. If no unexpected, clinically significant antibodies are detected and there is no record of their previous detection, only serologic testing for

ABO incompatibility is required (i.e., antiglobulin testing is not required when the crossmatch is performed).

Red Blood Cell Transfusion

If a transfusion is appropriate, a benefit should occur. In a large study of Jehovah's Witnesses (whose religious beliefs preclude the use of blood transfusion), the risk for death was higher in surgical patients with cardiovascular disease than in those without. A follow-up analysis of a subset of these patients reported that the odds of death in patients with a postoperative hemoglobin level less than 7 g/dL increased 2.5 times for each 1 gram decrease in the hemoglobin level; although no deaths occurred in 98 patients with postoperative levels of 7.1 to 8.0 g/dL, 34.4% of 32 patients with postoperative levels of 4.1 to 5.0 g/dL died. These data suggest that in surgery-induced anemia, survival is improved if blood transfusion is administered to maintain the hemoglobin concentration at greater than 7 g/dL. Data from randomized trials consistently support more restrictive policies for the transfusion of red blood cells and platelets (Table 177-1).

In a large, randomized trial of elderly patients undergoing repair of hip fracture, transfusion at a hemoglobin threshold of 10 g/dL was no better than transfusion for symptoms of anemia or at physician discretion for a hemoglobin level of less than 8 g/dL. In a multi-institutional study, 418 critical care patients received red cell transfusions when their hemoglobin levels dropped below 7 g/dL and had their levels maintained between 7 and 9 g/dL, whereas another 420 patients received transfusions when their hemoglobin levels dropped below 10 g/dL and had their levels maintained between 10 and 12 g/dL. Thirty-day mortality rates were not significantly different in the two groups, suggesting that a transfusion threshold as low as 7 g/dL may be as safe as a higher threshold of 10 g/dL in critically ill patients. A follow-up analysis found that the more restrictive strategy of red blood cell transfusion also appeared to be safe in most patients with cardiovascular disease.

A retrospective study analyzed the relationships among anemia, blood transfusion, and mortality in nearly 80,000 patients older than 65 years hospitalized for acute myocardial infarction. Anemia, defined as a hematocrit below 39%, was present on hospital admission in 44% of patients and was 33% or less in 10% of patients; blood transfusion in patients with hematocrit levels lower than 33% at admission was associated with a significantly lower 30-day mortality. On the basis of this study, transfusion to maintain hematocrit levels above 30% has been recommended in patients with acute myocardial infarction. In patients with heart failure, transfusion to maintain a hemoglobin level above 10 g/dL appears to improve outcomes. Among patients undergoing cardiac surgery, a restrictive perioperative transfusion strategy is as good as a more liberal strategy.

Clinical Practice Guidelines

The number of published clinical practice guidelines for red blood cell, platelet, and plasma transfusions attests to the increasing interest and importance of appropriate blood utilization by professional societies and health care institutions (Table 177-2). The guidelines acknowledge the necessity of considering patient covariables or other patient-specific criteria for making transfusion decisions. Among published guidelines, it is generally agreed that transfusion is not of benefit when the hemoglobin is greater than 10 g/dL but may be beneficial when the hemoglobin is less than 6 to 7 g/dL.[6] The selection of a discrete hemoglobin as a trigger for transfusion has been controversial. For example, the initial guidelines by the American Society of Anesthesiology (ASA) in 1996 identified a hemoglobin level of less than 6 g/dL as a transfusion trigger for acute blood loss, whereas the updated ASA guidelines in 2006 noted that "although multiple trials have evaluated transfusion thresholds on patient outcome, the literature is insufficient to define a transfusion trigger in surgical patients with substantial blood loss." Three guidelines have specified a hemoglobin threshold for patients with acute bleeding only, whereas the concept of an empirical hemoglobin concentration as a transfusion trigger has been refuted by a number of other published clinical practice guidelines. Arbitrary laboratory values are inadequate to define when red blood cell transfusions are appropriate, so that each patient must be evaluated individually and patient-specific anemia management strategies employed. Implementing real-time clinical decision support and best practice alerts into physician order entry for blood transfusions has resulted in improved blood utilization and improved patients outcomes.

Platelet Transfusion

The use of intensive chemotherapy regimens and bone marrow or stem cell transplantation has increased the demand for platelet products, particularly in patients with severe thrombocytopenia or bleeding complications. The use of apheresis platelet transfusions (i.e., a platelet unit collected from a dedicated donor through an apheresis procedure) has increased substantially, driven by the need for platelet inventories to support cardiac surgery, oncology, and stem cell transplantation programs. Emerging issues in platelet transfusion therapy include re-evaluation of the platelet threshold for prophylactic transfusion and modification of the dose of platelet transfusions. The problem

TABLE 177-1 KEY CLINICAL TRIALS IN RED BLOOD CELL AND PLATELET THERAPY

RED BLOOD CELL TRANSFUSION

CLINICAL SETTING	HEMOGLOBIN THRESHOLD (g/dL)	PATIENTS TRANSFUSED	DEVIATION FROM TRANSFUSION PROTOCOL	MEAN HEMOGLOBIN AT TRANSFUSION (g/dL)	PARTICIPATION OF ELIGIBLE PATIENTS	REFERENCE
Intensive care	7.0 vs. 10.0	67% 99%	1.4% 4.3%	8.5 ± 0.7* 10.7 ± 0.7*	41%	A2
Cardiothoracic surgery	8.0 vs. 10.0	47% 78%	1.6% 0%	9.1 (9.0-9.2) 10.5 (10.4-10.6)	75%	A3
Hip surgery	8.0 vs. 10.0	41% 97%	9.0% 5.6%	7.9 ± 0.6 9.2 ± 0.5	56%	A4
Acute upper GI bleeding	7.0 vs. 9.0	49% 86%	9% 3%	7.3 ± 1.4 8.0 ± 1.5	93%	A5

PLATELET TRANSFUSION

TYPE OF TRIAL	COMPARISONS	PATIENTS TRANSFUSED	PATIENTS TRANSFUSED OFF PROTOCOL	MEDIAN PLATELET COUNT AT TRANSFUSION ($\times 10^9$/L) (RANGE)	PATIENTS WITH GRADE 2 OR GREATER BLEEDING†	REFERENCE
Trigger for prophylactic platelet transfusion	Platelet count of 10×10^9/L vs. 20×10^9/L	135 120	5.4% 2.0%	9 (1-89) 14 (0-64)	21.5% 20%	A6
Platelet dose	Low vs. medium vs. high dose	417 423 432	21% 8% 14%	9 (7-16) 9 (7-19) 9 (7-12)	71% 69% 70%	A7
Prophylactic vs. therapeutic transfusion	Prophylactic vs. therapeutic	197 194	11% 22%	Not provided Not provided	19% 42%	A8
Prophylactic vs. therapeutic transfusion	Prophylactic vs. therapeutic	299 301	23% 14%	Not provided Not provided	43% 50%	A9

*Average daily hemoglobin.
†Different grading systems were used for documenting bleeding

TABLE 177-2 MEDICAL SOCIETY CLINICAL PRACTICE GUIDELINES

RED BLOOD CELL TRANSFUSION

YEAR	SOCIETY	RECOMMENDATIONS	REFERENCE
1988	NIH Consensus Conference	<7 g/dL (acute)	*JAMA* 1988;260:2700
1992	American Colleges of Physicians (ACP)	No number	*Ann Intern Med* 1992;116:393-402
1996/2006	American Society of Anesthesiologists (ASA)	<6 g/dL (acute) No number	*Anesthesia* 1996;84:732-747 *Anesthesia* 2006;105:198-208
1997/1998	Canadian Medical Association (CMA)	No number	*Can Med Assoc J* 1997;156:S1-24 *J Emerg Med* 1998;16:129-131
1998	College of American Pathologists (CAP)	6 g/dL (acute)	*Arch Pathol Lab Med* 1998;122:130-138
2001/2012	British Committee for Standards in Haematology	No number 7-8 g/dL*	*Br J Haematol* 2001;113:24-31 http://www.bcshguidelines.com/documents/BCSH_Blood_Admin_-_addendum_August_2012.pdf
2001	The NHMRC/Australasian Soc Blood Trans	7 g/dL	https://www.nhmrc.gov.au/guidelines-publications/cp78
2007/2011	Society of Thoracic Surgeons and Society of Cardiovascular Anesthesiologists	7 g/dL or 8 g/dL*	*Ann Thorac Surg* 2007;83:S27-86 *Ann Thorac Surg* 2011;91:944-982
2009	American College of Critical Care Medicine Society of Critical Care Medicine	7 g/dL 7 g/dL	*Crit Care Med* 2009;37:3124-157 *J Trauma* 2009;67:1439-1442
2011	Society for the Advancement of Blood Medicine	8 g/dL	*Trans Med Rev* 2011;25:232-246
2012	National Blood Authority, Australia	No number	http://www.nba.gov.au/guidelines/review.html
2012	American Association of Blood Banks	7-8 g/dL or 8 g/dL†	*Ann Intern Med* 2012;157:49-58
2012	Kidney Disease Improving Global Outcomes	No number	*Kidney Int* 2012;2:311-316
2012	National Cancer Center Network	7 g/dL	*JNCCN* 2012;10:628-653

PLATELET TRANSFUSION

YEAR	SOCIETY	RECOMMENDATIONS FOR TRIGGER FOR PROPHYLACTIC PLATELET TRANSFUSIONS‡	REFERENCE
1992	British Committee for Standards in Haematology	10×10^9/L	*Transfus Med* 1992;2:311-318
1994	College of American Pathologists	5×10^9/L	*JAMA* 1994;271:777-781
1998	Consensus Conference, Royal College of Physicians, Edinburgh	10×10^9/L	*Transfusion* 1998;38:796-797
2001	American Society of Clinical Oncology	10×10^9/L	*J Clin Oncol* 2001;19:1519-1538
2003	British Committee for Standards in Haematology	10×10^9/L	*Br J Haematol* 2003;122:10-23
2009	Italian Society of Transfusion Medicine and Immunohaematology	10×10^9/L	*Blood Transfus* 2009;7:132-150

PLASMA TRANSFUSION

YEAR	SOCIETY	PRINCIPAL INDICATIONS	REFERENCE
1984	Consensus Conference, National Institutes of Health	Replacement of clotting factor deficiencies Reversal of warfarin coagulopathy Massive blood transfusion Treatment of TTP Antithrombin III deficiency Immunodeficiencies	*JAMA* 1985;253:551-553
1992	British Committee for Standards in Haematology	1. Replacement of clotting factor deficiencies where factor concentrates are unavailable 2. Immediate reversal of warfarin effect 3. DIC 4. TTP	*Transfus Med* 1992;2:57-63
1994	College of American Pathologists	1. Coagulopathy (inherited or acquired) with bleeding or before an invasive procedure 2. Massive hemorrhage and transfusion 3. Reversal of warfarin coagulopathy 4. Antithrombin III deficiency 5. Immunodeficiencies 6. TTP	*JAMA* 1994;271:777-781
1997	Canadian Medical Association Expert Working Group	1. Acquired coagulation factor deficiencies (e.g., vitamin K deficiency, warfarin, liver disease) with bleeding or before an invasive procedure 2. Acute DIC 3. Massive hemorrhage and transfusion 4. TTP 5. Replacement of coagulation factor deficiencies where factor concentrates are unavailable	*Can Med Assoc J* 1997;156(11 Suppl):S1-S24

TABLE 177-2 MEDICAL SOCIETY CLINICAL PRACTICE GUIDELINES—cont'd

YEAR	SOCIETY	PRINCIPAL INDICATIONS	REFERENCE
2004	British Committee for Standards in Haematology	1. Replacement of coagulation factor deficiencies where factor concentrates are unavailable 2. DIC 3. TTP 4. Reversal of warfarin coagulopathy 5. Massive hemorrhage and blood transfusion	*Br J Haematol* 2004;126:11-28
2009	Italian Society of Transfusion Medicine and Immunohaematology	1. Correction of clotting factor deficiencies (factor concentrate unavailable) when the PT or aPTT is >1.5 (e.g., liver disease, warfarin coagulopathy, DIC, massive hemorrhage, and blood transfusion) 2. TTP 3. Reconstitution of whole blood for exchange transfusions 4. Hereditary angioedema where C1-esterase inhibitor is not available	*Blood Transfus* 2009;7:132-150
2010	American Association of Blood Banks[§]	1. Massive transfusion in trauma patients 2. Warfarin coagulopathy and intracranial hemorrhage	*Transfusion* 2010;50:1227-1239

*For patients with acute blood loss.
†For patients with symptoms of end-organ ischemia.
‡Consider higher threshold for patients with additional risk factors for bleeding.
§Only six questions relating to plasma use in specific scenarios were considered.
aPTT = activated partial thromboplastin time; DIC, disseminated intravascular coagulation; PT = prothrombin time; TTP = thrombotic thrombocytopenic purpura.
Modified from Goodnough LT, Levy JH, Murphy MF. Concepts of blood transfusion in adults. *Lancet.* 2013;381:1845-1854.

of platelet alloimmunization and the platelet transfusion–refractory patient is discussed in detail in Chapter 177.

Threshold for Platelet Transfusion

The appropriate threshold for prophylactic platelet transfusion depends on the clinical situation (Fig. 177-2). Prospective, randomized studies indicate that a platelet transfusion threshold of 10,000 cells/μL is as safe and effective as higher thresholds in patients undergoing chemotherapy or stem cell transplantation.

For consumptive thrombocytopenias such as disseminated intravascular coagulation (Chapter 175), platelet therapy is supportive but not effective until the underlying cause is treated. Platelet transfusions are generally not indicated in patients with idiopathic thrombocytopenic purpura or the thrombotic microangiopathies, including thrombotic thrombocytopenic purpura and hemolytic-uremic syndrome (Chapter 172).

Platelet Dose

The AABB standards require that 75% of apheresis platelet products contain more than 3×10^{11} platelets; however, no consensus exists for a standardized platelet dose, and clinical trials have used a broad range of doses. In general, higher platelet doses result in greater incremental increases in the platelet count and prolonged time to the next transfusion; however, the estimated platelet half-life is similar, and there are no other differences in outcomes.

Low-dose platelet therapy (i.e., <1 platelet unit) may be more beneficial in thrombocytopenic patients who are receiving prophylactic platelet transfusions. The fixed platelet requirement for hemostasis is estimated at 7100/μL^3/day, and platelet needs above this threshold are mainly a result of platelet consumption. For patients who become thrombocytopenic as a result of myeloablative therapy, platelet survival decreases with increasing severity of thrombocytopenia. A trial of prophylactic platelet transfusions in patients with hypoproliferative thrombocytopenia found that low doses (1.1×10^{11} per square meter) of platelets led to a decreased number of transfusions but an increased number of platelets transfused (and donor exposures). At doses between 1.1×10^{11} and 4.4×10^4 platelets per square meter, no effect on the incidence of bleeding was demonstrated.

Plasma Transfusion

Plasma therapy provides a source of clotting factors for patients with inherited or acquired coagulation disorders. Patients with inherited disorders such as hemophilia (Chapter 174) or von Willebrand disease (Chapter 173) are now treated primarily with clotting factor concentrates, which are blood derivatives from processed and treated commercial lots from pooled donors. For patients with acquired coagulopathies, there is little systematically derived, evidence-based guidance to inform plasma transfusion decisions. For patients with mild prolongation of coagulation assays such as prothrombin time and partial thromboplastin time (Chapter 171), plasma therapy has little or no value in prophylaxis for invasive procedures. Plasma therapy is indicated in patients undergoing massive hemorrhage (trauma, postpartum hemorrhage, or gastrointestinal bleeding) or for emergent reversal of warfarin-associated coagulopathy (Chapter 38), such as in patients with intracranial hemorrhage.[7] Recommended dosage in these settings is 15 to 30 mL/kg.

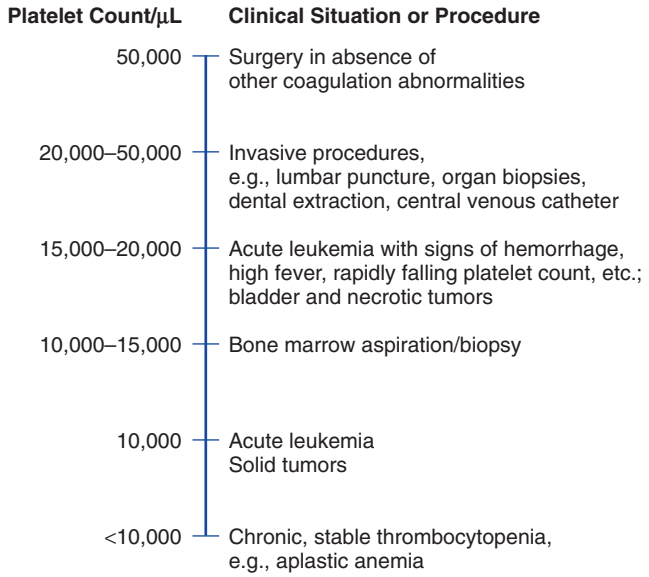

FIGURE 177-2. Threshold for providing platelet transfusions in thrombocytopenic patients.

BLOOD SAFETY

Adverse reactions associated with transfusion are listed in Table 177-3.

Errors in Transfusion Medicine

The mistransfusion rate (blood transfused to other than the intended recipient) is about 1 in 14,000 to 28,000 units. About two thirds of the errors occur in the clinical arena (patient misidentification and/or specimen mislabeling at the time of drawing blood for type and screen/crossmatch, incorrect identification of the recipient of the blood unit, and failure to recognize a transfusion reaction). Implementing a two-blood-specimen requirement to verify a patient's blood type before transfusion can minimize the issuing of mismatched blood products owing to wrong blood in the tube specimens. About

TABLE 177-3 TRANSFUSION-ASSOCIATED ADVERSE REACTIONS

ADVERSE REACTION	RISK PER UNIT INFUSED
Delayed serologic reactions	1 in 1600
Delayed hemolytic reactions	1 in 6700
Transfusion-related acute lung injury	1 in 8000
Graft-versus-host disease	Rare
Fluid overload	1 in 20
Febrile, nonhemolytic transfusion reactions	1 in 20-200
Allergic reactions	1 in 30-100
Anaphylactic reactions	1 in 150,000
Iron overload	After 80-100 U
Post-transfusion purpura	Rare
ABO-incompatible blood transfusions	1 in 30,000-60,000
Fatalities	1 in 600,000
Storage lesions	Unknown
Immunosuppressive effects	Unknown

Modified from Goodnough LT. Current issues in transfusion medicine. *Clin Adv Hematol Oncol.* 2005;3:614-616.

half of mistransfusions are ABO incompatible, and about 10% of these are fatal. The frequency of death as a result of ABO error can therefore be estimated at approximately 1 per 600,000 blood units, a risk that is higher than the current risks for transmitting HIV or hepatitis C virus (see later).

Transfusion Reactions

Acute Hemolytic Transfusion Reaction

An acute hemolytic transfusion reaction is most commonly defined as hemolysis of donor red cells within 25 hours of transfusion by preformed alloantibodies in the recipient's circulation. Life-threatening acute hemolytic transfusion reactions are most commonly due to ABO-incompatible blood being transfused to a recipient with naturally occurring ABO alloantibodies (anti-A, anti-B, anti-A, B). Clerical errors (mislabeling blood or misidentifying patients) account for 80% of such reactions.

Signs and symptoms of an acute intravascular hemolytic transfusion reaction may develop when as little as 10 to 15 mL of ABO-incompatible blood has been infused. Fever, which is the most common initial manifestation, is frequently accompanied by chills. The patient may complain of a general sense of anxiety or uneasiness or pain at the infusion site or in the back or chest (or both). The most serious sequela is acute renal failure. In an unconscious or anesthetized patient, diffuse bleeding at the surgical site may be the first indication of intravascular hemolysis and may be accompanied by hemoglobinuria and hypotension.

Treatment begins with immediate cessation of the transfusion. The risk for renal failure may be reduced by the administration of crystalloid fluids, including sodium bicarbonate (250 to 500 mg intravenously over a 1- to 4-hour period), to maintain urine pH at 7.0 and by diuresis with 20% mannitol (100 mL/m^2 in 30 to 60 minutes, followed by 30 mL/m^2/hour for 12 hours) or furosemide (40 to 120 mg intravenously).

Febrile Nonhemolytic Transfusion Reactions

Febrile nonhemolytic transfusion reactions are common and are estimated to occur in 0.5% of all red cell transfusions and up to 15% of platelet transfusions. A febrile transfusion reaction is defined as a rise in temperature greater than 1°C, which may be accompanied by chills, rigor, or both.

These reactions are thought to be due to a reaction of HLA or leukocyte-specific antigens on transfused lymphocytes, granulocytes, or platelets in the donor unit with antibodies in previously alloimmunized recipients. Multiply transfused individuals and multiparous women are most likely to experience this type of transfusion reaction. Febrile nonhemolytic transfusion reactions, especially those associated with platelet transfusions, may be caused by the infusion of biologic response modifiers, such as cytokines, that have accumulated in the platelet concentrate during storage.

Symptoms may occur during the transfusion or may not manifest until 1 to 2 hours after its completion. The diagnosis of a febrile nonhemolytic transfusion reaction is generally made by excluding other causes of fever (e.g., underlying patient condition, bacterial contamination of the unit, acute hemolytic transfusion reaction).

Febrile nonhemolytic transfusion reactions in susceptible populations can often be prevented by administering antipyretics before the transfusion of blood components. Prestorage leukocyte reduction is recommended to prevent reactions resulting from the accumulation of cytokines during storage.

Allergic Reactions

Allergic reactions can be mild, moderate, or life threatening and are associated with the amount of plasma transfused. From 1 to 5% of all blood transfusion recipients experience mild allergic reactions.

Anaphylactic transfusion reactions are sometimes associated with antibodies to immunoglobulin (Ig) A, which are common and have an incidence of approximately 1 in 700 individuals. However, the incidence of anaphylactic transfusion reactions is much lower—1 in 20,000 to 50,000.

Urticarial reactions are not well understood but are believed to be an interaction between antibodies in the recipient's plasma and plasma proteins in the donor blood. There is usually no specific identifiable antigen to which the patient is reacting. Symptoms are generally mild and include localized urticaria, erythema, and itching.

Anaphylactic or anaphylactoid reactions, which can occur after the transfusion of only a few milliliters of blood or plasma, include skin flushing, nausea, abdominal cramps, vomiting, diarrhea, laryngeal edema, hypotension, shock, cardiac arrhythmia, cardiac arrest, and loss of consciousness. Fever is notably absent. In some cases, there may be symptoms indicative of airway involvement, such as hoarseness, wheezing, dyspnea, and substernal pain. Management begins with discontinuation of the transfusion. Treatment is diphenhydramine (25 to 50 mg intravenously), but more severe episodes may require aggressive therapy (Chapters 252 and 253).

Patients who experience recurrent allergic or urticarial reactions can be treated with antihistamines and/or histamine-2 receptor antagonists (H_2 blockers) before transfusion. Washed red blood cells may be indicated for patients who experience repeated severe urticarial reactions. Severe allergic or urticarial reactions may require treatment with corticosteroids.[8]

Bacterial Contamination

Bacterial contamination may be introduced into a unit of blood through skin contaminants during venipuncture or from donors with asymptomatic bacteremia. Multiplication of bacteria can occur in blood and blood components stored at refrigerated temperatures but is more likely to occur in platelets stored at room temperature.

Bacterial contamination of red cells is most often due to *Yersinia enterocolitica,* followed by *Serratia liquefaciens,* whereas platelets are most often contaminated with *Staphylococcus* and Enterobacteriaceae. The incidence of bacterial contamination of red cells has been estimated to be 1 in 60,000, with an overall fatality rate of 1 in 1 million. The incidence of bacterial contamination of platelets was estimated to be 1 in 5000 before the initiation of bacterial detection systems in 2004, but it is now estimated to be 50% lower (1 in 10,000).

Recipients of units with low bacterial counts may have relatively mild symptoms such as fever and chills, but those receiving units with high bacterial counts may have severe or fatal reactions. Clinically, the patient may experience high fever, shock, hemoglobinuria, renal failure, and disseminated intravascular coagulation. The blood transfusion must be stopped immediately, the patient's blood and any untransfused blood must be cultured, and broad-spectrum antibiotics (Chapter 108) should be started.

Circulatory Overload

Acute pulmonary edema, caused by the circulatory system's inability to handle an increased fluid volume, can occur in any patient who is transfused too rapidly. Although the true frequency of this type of transfusion reaction is unknown, prospective surveillance studies estimate that 8% of patients have transfusion-associated circulatory overload (TACO). Susceptible populations are primarily very young patients, elderly patients, and patients with a small total blood volume or cardiopulmonary disease.[9,10] Treatment is the same as for heart failure (Chapter 59).

Delayed Reactions

A delayed hemolytic transfusion reaction generally occurs 3 to 7 days after transfusion of the implicated unit. Hemolysis is usually extravascular, and red cells are destroyed in the recipient's circulation by antibody produced as a result of an immune response to the transfusion. These reactions are most commonly due to an anamnestic response (secondary exposure to a red cell antigen) in a patient with a negative antibody screen despite a low level of antibody as a result of previous exposure to a foreign red cell antigen through

either pregnancy or transfusion. Exposure to the same antigen a second time may cause IgG antibody to reappear within hours or days of the transfusion. This subsequent exposure to the antigen produces an anamnestic antibody response, resulting in increased production of IgG antibodies that are capable of reacting with any transfused cells present.

In most cases, anamnestic production of antibody does not result in acute hemolysis, but red cell destruction does occur between 3 days and 2 weeks after the transfusion. Patients are generally asymptomatic, and hemolysis may be noted only by a more rapid decline than usual in the patient's hemoglobin level or by an absence of the expected rise in hemoglobin. Fever, the most common initial symptom, is occasionally noted, along with jaundice; renal failure is rare. Prednisone (1 to 2 mg/kg/day) is indicated for more severe reactions.

Transfusion-Associated Graft-versus-Host Disease

Transfusion-associated graft-versus-host disease results from the transfusion of immunologically competent lymphocytes into an immunologically incompetent host. An individual's risk depends on whether the recipient is immunocompromised (and to what degree), the degree of HLA similarity between the transfusion donor and recipient, and the number of transfused T lymphocytes capable of multiplying and engrafting. The engrafted lymphocytes mount an immunologic response against the recipient's tissue, resulting in pancytopenia with bleeding and infectious complications. Symptoms usually appear within 12 days of transfusion. Transfusion-associated graft-versus-host disease is rare, but it is fatal in approximately 90% of affected patients.[11]

Transfusion-Related Acute Lung Injury

Transfusion-related acute lung injury (TRALI; Chapter 94) is an acute respiratory distress syndrome (Chapter 104) that occurs within 6 hours after transfusion and is characterized by dyspnea and hypoxia secondary to noncardiogenic pulmonary edema.[12,13] Although the actual incidence is almost certainly underreported, the estimated frequency is approximately 1 in 8000 transfusions. In a prospective nested case-control study, 16 of 668 (2.4%) of cardiac surgery patients developed TRALI, suggesting that the incidence of TRALI is particularly high in this population. In approximately 50% of cases, blood donor antibodies with HLA or neutrophil antigenic specificity can be shown to react with the recipient's leukocytes, leading to increased permeability of the pulmonary microcirculation.

Most recently, reactive lipid products from donor blood cell membranes that arise during the storage of blood products have been implicated in the pathophysiology of TRALI. Such substances are capable of neutrophil priming, with subsequent damage to the pulmonary-capillary endothelium of the recipient, particularly in patients who receive massive transfusions during cardiac surgery or for trauma or in patients receiving chemotherapy for malignancy. In each of these settings, the true incidence of TRALI may be underreported because the findings may be blamed on the underlying disease process or the surgical procedure. Similar to other causes of acute respiratory distress syndrome, therapy is supportive, and 90% of patients recover. Ongoing initiatives to reduce TRALI risks include recruiting male plasma donors or female donors with no prior history of pregnancy and screening female apheresis platelet donors for HLA antibodies and limiting donation to those who are HLA antibody negative.

Transmission of Blood Pathogens

Categories of transfusion-transmitted agents,[14] as well as those currently screened, are listed in Table 177-4. The implementation of nucleic acid testing of multiple minipools (donation samples, test well) from blood donations has markedly reduced the transmission of HIV and hepatitis C virus during the infectious window period. Current estimates of the risk per unit of blood are 1 in 1.4 million to 2.4 million for HIV and 1 in 872,000 to 1.7 million for hepatitis C virus (Fig 177-3). Restrictive transfusion strategies also reduce health care-associated infections.

Only 43% of the World Health Organization's 191 member states test blood for HIV, hepatitis C virus, and hepatitis B virus, so at least 13 million units of blood donated every year are not tested for these transmissible viruses. In the poorest countries, the cost of testing ($40 to $50 per blood

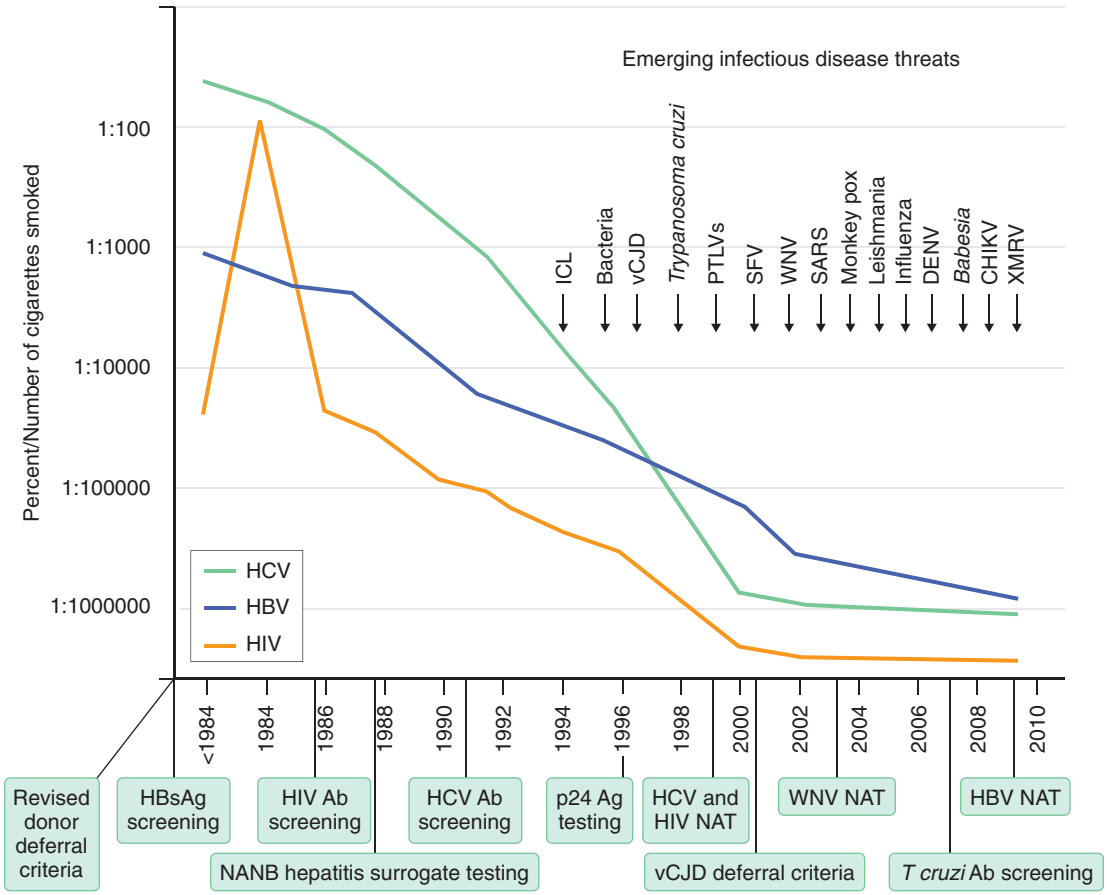

FIGURE 177-3. Risks for major transfusion-transmitted viruses (TTVs) linked to interventions, and accelerating rate of emerging infectious diseases (EIDs) of concern to blood safety. Evolution of the risks for transmission by blood transfusion of human immunodeficiency virus (HIV), hepatitis B virus (HBV), and hepatitis C virus (HCV). Major interventions to reduce risks are indicated below the time line on the *x*-axis. Emerging infectious disease threats over the past 20 years are indicated in the top right quadrant of the figure. Ab = antibody; Ag = antigen; CHIKV = Chikungunya virus; DENV = dengue virus; HBsAg = hepatitis B surface antigen; ICL = idiopathic CD4+ T lymphocytopenia; NANB = non-A, non-b hepatitis; NAT = nuclear acid testing; PTLV = primate T-lymphotopic virus; SARS = severe acute respiratory syndrome; SFV = simian foamy virus; vCJD = variant Creutzfeldt-Jakob disease; WNV = West Nile virus; XMRV = xenotropic murine leukemia virus–related virus.

TABLE 177-4 POTENTIAL RISKS OF BLOOD TRANSFUSION

1. Infectious agents
   - Transfusion-transmitted disease for which donors are tested*
     - Hepatitis B virus (HBV; 1970 [surface antigen]; 1986-1987 [core antibody]); 2009 [nucleic acid])
     - Human immunodeficiency virus (HIV; 1985 [antibody]; 1999 [nucleic acid])
     - Hepatitis C virus (HCV; 1986-1987 [alanine aminotransferase]; 1990 [antibody];1999 [nucleic acid])
     - Human T-cell lymphotropic virus (HTLV; 1988 [antibody])
     - West Nile virus (WNV; 2003 [nucleic acid])
     - Bacteria (in platelets only; 2004)
     - *Trypanosoma cruzi* (2007 [antibody])
     - Cytomegalovirus (CMV)
     - Syphilis
   - Transfusion-transmitted disease for which donors are not routinely tested
     - Hepatitis A virus (HAV)
     - Parvovirus B19
     - Dengue fever virus (DFV)
     - Malaria
     - *Babesia* sp
     - *Plasmodium* sp
     - *Leishmania* sp
     - *Brucella* sp
     - New variant Creutzfeldt-Jakob disease (nvCJD) prions
     - Unknown pathogens
2. Transfusion reactions
3. Alloimmunization
4. Medical errors (wrong blood to patient due to mislabeled specimen or patient misidentification)
5. Transfusion-associated acute lung injury (TRALI)
6. Transfusion-associated circulatory overload (TACO)
7. Iron overload
8. Immunomodulation
9. Storage lesions: age of blood

*The target of the screening assay (antibody, microbial antigen, or microbial nucleic acid) and the year of assay implementation are indicated in parentheses.From Goodnough LT. Blood management: transfusion comes of age. *Lancet.* 2013;381:1791-1792.

donation) is prohibitive. Every year, unsafe transfusions are estimated to account for 8 million to 16 million hepatitis B infections, 2.3 million to 4.7 million hepatitis C infections, and 80,000 to 160,000 HIV infections.

EMERGING ISSUES AND EVOLVING TECHNOLOGIES

Problems related to blood component storage have long been known, such as depletion of 2,3-diphosphoglycerate in red cell units during storage at 4°C for longer than 35 to 42 days. Some evidence suggests that in patients undergoing coronary artery bypass surgery, transfusion of older red cells is associated with a significantly increased risk for postoperative complications, as well as reduced survival. A prospective, randomized multicenter trial comparing the transfusion of blood less than 10 days old versus blood 21 days or older in adult patients undergoing coronary artery bypass surgery is currently underway in the United States. Erythropoiesis-stimulating agents, including recombinant human erythropoietin, modified erythropoietin molecules, and erythropoietin receptor agonists, along with artificial oxygen carriers, can potentially decrease the need for red cell transfusion.[15] It is important that oversight of these biotechnology products and other specialized blood products (e.g., solvent- or detergent-treated plasma, leukoreduced blood products, irradiated blood components, cytomegalovirus-negative blood components) be placed under the auspices the transfusion medicine committee. Otherwise, promotion of such products by the commercial sector directly to consumers undermines both institutional oversight and evidence-based rationales for their use.

Grade A References

A1. Carson JL, Carless PA, Hebert PC. Transfusion thresholds and other strategies for guiding allogeneic red blood cell transfusion. *Cochrane Database Syst Rev*. 2012;4:CD002042.

A2. Hebert PC, Wells G, Blajchman MA, et al. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. Transfusion Requirements in Critical Care Investigators, Canadian Critical Care Trials Group. *N Engl J Med*. 1999;340:409-417.

A3. Hajjar LA, Vincent JL, Galas FR, et al. Transfusion requirements after cardiac surgery: the TRACS randomized controlled trial. *JAMA*. 2010;304:1559-1567.

A4. Carson JL, Terrin ML, Noveck H, et al. Liberal or restrictive transfusion in high-risk patients after hip surgery. *N Engl J Med*. 2011;365:2453-2462.

A5. Villanueva C, Colomo A, Bosch A, et al. Transfusion strategies for acute upper gastrointestinal bleeding. *N Engl J Med*. 2013;368:11-21.

A6. Rebulla P, Finazzi G, Marangoni F, et al. The threshold for prophylactic platelet transfusions in adults with acute myeloid leukemia. Gruppo Italiano Malattie Ematologiche Maligne dell'Adulto. *N Engl J Med*. 1997;337:1870-1875.

A7. Slichter SJ, Kaufman RM, Assmann SF, et al. Dose of prophylactic platelet transfusions and prevention of hemorrhage. *N Engl J Med*. 2010;362:600-613.

A8. Wandt H, Schaefer-Eckart K, Wendelin K, et al. Therapeutic platelet transfusion versus routine prophylactic transfusion in patients with haematological malignancies: an open-label, multicentre, randomised study. *Lancet*. 2012;380:1309-1316.

A9. Stanworth SJ, Estcourt LJ, Powter G, et al. A no-prophylaxis platelet-transfusion strategy for hematologic cancers. *N Engl J Med*. 2013;368:1771-1780.

A10. Diedrich B, Remberger M, Shanwell A, et al. A prospective randomized trial of a prophylactic platelet transfusion trigger of $10 \times 10(9)$ per L versus $30 \times 10(9)$ per L in allogeneic hematopoietic progenitor cell transplant recipients. *Transfusion*. 2005;45:1064-1072.

A11. Rohde JM, Dimcheff DE, Blumberg N, et al. Health care-associated infection after red blood cell transfusion: a systematic review and meta-analysis. *JAMA*. 2014;311:1317-1326.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. AABB, American Red Cross, America's Blood Centers, and Armed Services Blood Program. Circular of Information for the use of human blood and blood components. http://www.aabb.org/tm/coi/Documents/coi1113.pdf; Accessed November 11, 2014.
2. Goodnough LT, Shander AS. Patient blood management. *Anesthesiology*. 2012;116:1367-1376.
3. Goodnough LT, Levy JH, Murphy MF. Concepts of blood transfusion in adults. *Lancet*. 2013;381:1845-1854.
4. Goodnough LT, Shieh L, Hadhazy E, et al. Improved blood utilization using real-time clinical decision support. *Transfusion*. 2014;54:1358-1365.
5. Wiliamson LM, Devine DV. Challenges in the management of the blood supply. *Lancet*. 2013;381:1866-1875.
6. Goodnough LT, Maggio P, Hadhazy E, et al. Restrictive blood transfusion practices are associated with improved patient outcomes. *Transfusion*. 2014;54:2753-2759.
7. Goodnough LT, Shander A. How we treat: management of warfarin-associated coagulopathy in patients with intracerebral hemorrhage. *Blood*. 2011;117:6091-6099.
8. Hiryama F. Current understanding of allergic transfusion reactions: incidence, pathogenesis, laboratory tests, prevention and treatment. *Br J Haematol*. 2013;160:434-444.
9. Menis M, Anderson SA, Forshee RA, et al. Transfusion-associated circulatory overload (TACO) and potential risk factors among the inpatients US elderly as recorded in Medicare administrative databases during 2011. *Vox Sang*. 2014;106:144-152.
10. Murphy EL, Kwaan N, Looney MR, et al. Risk factors and outcomes in transfusion-associated circulatory overload. *Am J Med*. 2013;126:357.e29-357.e38.
11. Fast LD. Developments in the prevention of transfusion-associated graft-versus-host disease. *B J Haematol*. 2012;158:563-568.
12. Shaz BH, Stowell SR, Hillyer CD. Transfusion-related acute lung injury: from bedside to bench and back. *Blood*. 2011;117:1463-1471.
13. Vlaar AP, Juffermans NP. Transfusion-related acute lung injury: a clinical review. *Lancet*. 2013;382:984-994.
14. Perkins H, Busch M. Transfusion-associated infections: 50 years of relentless challenges and remarkable progress. *Transfusion*. 2010;50:2080-2099.
15. Spahn DR, Goodnough LT. Alternatives to blood transfusion. *Lancet*. 2013;381:1855-1865.

REVIEW QUESTIONS

1. A 65-year-old, generally healthy man is transfused with 2 units of packed red cells in the recovery room after uncomplicated hip replacement surgery. Two hours later, he complains of acute onset of shortness of breath, and he is found to be hypertensive, cyanotic, and hypoxic. He is afebrile. Blood counts are normal. Chest radiograph shows bilateral opacities. A bedside echocardiogram is normal, as is the serum brain natriuretic peptide level. He is quickly transferred to the intensive care unit for imminent intubation and ventilatory support. Which of the following is the most likely diagnosis?

A. Transfusion-associated graft-versus-host disease
B. Bacterial contamination of the blood products
C. Delayed transfusion reaction
D. Transfusion-associated circulatory overload (TACO)
E. Transfusion-related acute lung injury (TRALI)

Answer: E TRALI is now the leading cause of transfusion-related mortality. It typically presents as this case, with the patient becoming acutely symptomatic after having been transfused with any blood product within the previous 3 to 6 hours. Transfusion-associated graft-versus-host disease results from the transfusion of immunologically competent lymphocytes into an immunocompromised host, which this patient is not, and leads mainly to pancytopenia. A delayed transfusion reaction occurs much later, usually 3 to 7 days after the transfusion, and leads to hemolysis of red cells in the recipient by antibody produced as a result of the immune response to the transfusion. TACO can be difficult to distinguish from TRALI, but in this case, the normal echocardiogram and serum brain natriuretic peptide would argue against that (though not rule it out). Finally, septic transfusions are typically accompanied by hypotension and fever, which this patient did not have. Bacterial contamination of blood products occurs much more frequently with the transfusion of platelets, which are stored at room temperature: this patient received packed red cells. Very importantly, however, septic transfusions can mimic or even coexist with TRALI and, whenever there is any doubt, should be aggressively diagnosed and treated as such. The management of TRALI by itself is largely supportive.

2. Each of the following patients needs transfusion of 2 units of pack red cells for various reasons. Which one of them requires cytomegalovirus (CMV)-negative blood?

A. 30-year-old woman in the third trimester of an uncomplicated pregnancy
B. 50-year-old man in remission from acute myelogenous leukemia
C. 87-year-old woman hospitalized for exacerbation of chronic heart failure
D. 45-year-old woman with systemic lupus erythematosus, maintained on prednisone
E. 55-year-old man, homeless and alcoholic, admitted with upper gastrointestinal bleed

Answer: A Blood product recipients who are CMV naïve (anti–CMV antibody negative) and have profoundly impaired immune systems are at increased risk for primary CMV infection when they receive CMV-positive blood products. CMV-negative blood products are required for pregnant mothers to prevent congenital CMV infection in the fetus. Profoundly immunocompromised individuals would include those receiving allogeneic or peripheral stem cell transplantations and untreated AIDS patients. In general, most institutions do not screen for CMV-negative blood products for patients needing transfusion who are receiving steroids or chemotherapy. Other indications in recipients for having CMV-seronegative blood transfused are intrauterine transfusion and neonates less than 37 weeks' gestation. Prestorage leukocyte depletion of blood products (red cells, platelets) significantly reduces the infection risk, but even for those products, most blood banks will still do CMV testing. (See Preventing Transmission of Cytomegalovirus.)

3. Which of the following statements regarding delayed transfusion reactions is correct?

A. Delayed hemolytic transfusion reactions typically occur 48 to 72 hours after transfusion of the implicated unit.
B. Delayed transfusion reactions mostly occur in patients who have very low titers of preexisting antibodies from previous transfusions or through pregnancy and therefore had a negative antibody screen before transfusion; they are thus anamnestic responses.
C. Most patients present with fevers, chills, and back and flank pain.
D. Steroids should be avoided in the management of delayed transfusion reactions.
E. The hemolysis of delayed transfusion reactions is mostly intravascular.

Answer: B Delayed hemolytic transfusions (1) characteristically occur 3 to 7 days after transfusion of the implicated unit; (2) typically present without symptoms, with just a more rapid decline than usual in the recipient's hemoglobin (or failure to increase the hemoglobin after transfusion) accompanied by hemolytic chemistries including hyperbilirubinemia; and (3) should be treated with prednisone (1 to 2 mg/k/day) in more severe reactions. The hemolysis in delayed transfusion reactions is mostly extravascular and results from an anamnestic response in recipients previously exposed to the antigen(s) through blood transfusions or pregnancy but having sufficiently low baseline titers of antibody to escape detection by routine screening. (See Delayed Reactions.)

178

HEMATOPOIETIC STEM CELL TRANSPLANTATION

ARMAND KEATING AND MICHAEL R. BISHOP

INTRODUCTION

Hematopoietic stem cell transplantation (HSCT) is a procedure by which hematopoietic stem cells (HSCs) are infused intravenously to restore hematopoiesis and immune function after high-dose chemotherapy with or without radiation therapy. Stem cells used for HSCT are of hematopoietic origin, in contrast to the more primitive pluripotent stem cells (embryonic stem cells), which are of growing interest for use in regenerative therapy because of their ability to differentiate into virtually any somatic cell. The term *HSCT* has replaced the term *bone marrow transplantation* because HSCs can be derived from a variety of sources, including bone marrow, the peripheral blood, and umbilical cord blood. In addition to treating blood cancers and other malignancies, HSCT can also provide clinically significant enzyme replacement. HSCs can give rise to all blood elements and may be genetically modified to enhance their function. Furthermore, HSCs may be manipulated ex vivo and used in the burgeoning field of regenerative medicine. Taken together, HSCT is a major component and foundation of the broader field of cellular therapy, which many believe forms the "third pillar" of medicine, complementing small molecule drugs and biologic agents.

The clinical development and application of HSCT originated in the observations of the severe myelosuppressive effects of radiation among survivors of the Hiroshima and Nagasaki nuclear bombings. There were intensive research efforts over the subsequent decade to develop methods to protect against and reverse the myelosuppressive effects of radiation, including the infusion of bone marrow. These efforts were primarily hindered by graft rejection. Among animals that did not reject their marrow grafts, a syndrome of weight loss, alopecia, diarrhea, and eventually death, referred to as "runting" disease, was commonly observed. This syndrome is now referred to clinically as graft-versus-host disease (GVHD) and is discussed in more detail later in this chapter. The subsequent determination and understanding of the major histocompatibility complex (MHC) and human leukocyte antigens (HLA), the major determinants of graft rejection, and of GVHD, enabled HSCT to be clinically applied. The first successful reports of clinical bone marrow transplantation, used for patients with severe combined immunodeficiency disorders, severe aplastic anemia, and advanced acute leukemias, occurred in the late 1960s and early 1970s. Subsequently in the 1980s, it was demonstrated that the administration of high-dose chemotherapy followed by HSCs obtained from the bone marrow or peripheral blood of the patients themselves improved outcomes and survival of patients with a

variety of hematologic and solid tumors. HSCT is a standard treatment option for many malignancies, immunodeficiency states, metabolic disorders (e.g., Hurler syndrome), and defective hematopoietic states (e.g., severe aplastic anemia, thalassemia).

The three types of HSCT are allogeneic, autologous, and syngeneic. HSCs obtained from someone other than the patient are referred to as allogeneic, while HSCs obtained from the patient are referred to as autologous. In syngeneic HSCT, patients receive HSCs from an identical twin, a relatively rare event (<1%). The determination of the type of HSCT that a patient receives (i.e., allogeneic vs. syngeneic vs. autologous) is based on the disease to be treated, disease state (e.g., initial treatment vs. treatment of recurrent disease), the urgency to treat with HSCT, availability of a donor, and urgency in obtaining a donor.

As previously mentioned, HSCs may be obtained from the bone marrow, the peripheral blood, and umbilical cord blood. HSCs can be obtained directly from bone marrow by serial aspirations from the posterior iliac crests while the patient is under general anesthesia. Alternatively, HSCs can be obtained from peripheral blood after stimulation of the donor with hematopoietic growth factors such as granulocyte colony-stimulating factor (G-CSF) and plerixafor followed by leukapheresis. HSCs from bone marrow are used in both autologous and allogeneic HSCT, although less frequently than in the past. Peripheral blood HSCs are currently used in approximately 90% of autologous HSCT and in approximately 70% of allogeneic HSCT. The greater use of peripheral blood HSCs is related to their relative ease of procurement and moderate advantage in the rate of hematopoietic recovery after infusion compared with HSCs derived from bone marrow, but the outcomes of the two approaches are similar. The HSCs may be infused fresh or cryopreserved. Cryopreservation involves processing the cells in culture medium containing dimethyl sulfoxide (DMSO), placing them in specialized plastic bags, and then storing the cells indefinitely in the vapor phase of liquid nitrogen until needed. Cord blood HSCs are collected at the time of delivery, cryopreserved, and stored. For allogeneic donation, HSCs may also be infused on the day of collection. For autologous donation, HSCs are always cryopreserved.

ALLOGENEIC HEMATOPOIETIC STEM CELL TRANSPLANTATION

The distinctive characteristics of allogeneic HSCT are that the stem cell graft is free of contamination by malignant cells and contains lymphocytes, primarily T and natural killer (NK) cells that are capable of mediating an immunologic reaction against foreign antigens.[1] This latter characteristic can be a major advantage if the immunologic response is directed against malignant cells, referred to as the graft-versus-leukemia or graft-versus-tumor effect, which can potentially eradicate disease and reduce the chance of disease relapse. However, if the immunologic response is directed against antigens present on normal tissues, the graft-versus-host response, the destruction of normal organs can result and is described clinically as GVHD. The risk of both graft rejection (host-versus-graft reaction) and GVHD rises with the degree of HLA disparity between the donor and the recipient.

The graft-versus-leukemia effect first was recognized in animal models and subsequently was noted among patients undergoing allogeneic HSCT for acute and chronic leukemias. The clinical importance of the role that immunocompetent donor T cells play in mediating a graft-versus-leukemia effect was provided by initial observations of increased rates of relapse in patients who received allogeneic stem cell grafts from which T cells had been removed (T-cell depletion), an inverse correlation between relapse and severity of GVHD, and increased rates of relapse after syngeneic or autologous HSCT using the same myeloablative conditioning regimen. These data suggested that T cells within the allograft were involved directly in eradicating leukemia. Finally, the most compelling evidence for the importance of T cells mediating the graft-versus-leukemia effect arose from the clinical observation that infusion of allogeneic T cells alone, termed a donor lymphocyte infusion (DLI), at a time remote from the transplant conditioning regimen, successfully eradicated leukemia which had persisted or recurred after allogeneic HSCT. However, there is wide variability in the clinical effectiveness of the graft-versus-leukemia effect against different malignancies after allogeneic HSCT.

Sources of Donor Hematopoietic Stem Cells

As previously emphasized, HSCs are obtained from a donor other than the recipient or an identical twin in allogeneic HSCT. Because of the immunologic barriers that can result in graft rejection and GVHD, allogeneic HSCT requires that the donor and the recipient share key genes. The most important of these are HLA, which are derived from the MHC located on chromosome 6. The most important HLAs include HLA-DR, HLA-DP, and HLA-DQ loci. A single set of MHC alleles, described as a haplotype, is inherited from each parent, resulting in HLA pairs. A clinical "match" means that HLA in the donor match perfectly with those in the patient.

The choice of donor for an allogeneic HSCT takes into account several factors, including the patient's disease, disease state, and urgency in obtaining a donor. A fully HLA-matched sibling is the preferred donor source because of the sibling's availability and the observation that the risks of graft rejection and GVHD are lowest with these allogeneic HSCs. The probability of having a HLA match from a sibling is approximately 25% and increases with the number of siblings within a specific family. The probability can be estimated using the following formula: chance of having an HLA-matched sibling = $1 - (0.75)^n$, where n is the number of potential sibling donors. There is approximately a 1% chance of a crossover event (i.e., genetic material switched between chromosomes during meiosis), primarily between the HLA-A and the HLA-B loci. The clinical outcomes for allogeneic HSCT using a sibling with a single HLA mismatch are similar to those with a fully HLA-matched sibling.

For patients who lack an HLA-identical sibling donor, the alternative sources for allogeneic HSCs include a fully HLA-matched volunteer unrelated donor, a fully or partially HLA-matched cord blood unit, or a partially HLA-matched first-degree family member. The genes encoding HLAs are numerous, and the odds that any two unrelated individuals are HLA identical for main loci are less than one in 10,000. More than 22 million volunteer donors are available worldwide; a donor can be found for about 50% of patients for whom a search is initiated. The probability of identifying a suitable volunteer is highly dependent upon race because of varying degrees of HLA diversity. For example, for persons of northern European extraction, the probability of a match can be as high as 90%. It usually takes about 2 to 4 months to locate an unrelated donor, which may be too long for some patients with rapidly progressive malignancies. When a suitable volunteer is not available or a donor is needed more urgently, HSCs from alternative donors can be considered, including partially HLA-matched ("haploidentical") relatives or umbilical cord blood. Haploidentical HSCs are readily available but have increased risks of graft rejection and GVHD. These risks can be reduced by T-cell depletion methods but lead to delayed immune reconstitution and competency posttransplant, leading to increased risks of infection and disease recurrence. Umbilical cord blood HSCs are taken at the time of delivery, stored in cord blood banks, and are readily available when needed.[2] Because of the unique immature biology of lymphocytes in umbilical cord blood, these transplants are associated with less GVHD; consequently, the HLA matching requirements are less strict, and two or three HLA mismatches may be acceptable. However, the small volume (50-150 mL) of cord blood that can be collected results in a limited number of HSCs, which often prohibits their use in adults because successful engraftment correlates with the number of HSCs per patient body weight. The limitation of cell dose can be overcome by the use of more than one cord blood unit. Cord blood HSCs are associated with delayed times to engraftment and immune reconstitution leading to an increased rate of infection. The other significant disadvantage is that after the cord blood unit is used, there is no possibility to obtain additional cells in the event of graft failure or if a DLI is required.

Conditioning and Preparation of Recipient

After an allogeneic stem cell source has been identified, the recipient patients then receive regimens with the intent of "conditioning" or "preparing" them for the infusion of HSCs (Fig. 178-1). These conditioning regimens are designed to be adequately immunosuppressive to overcome the host-versus-graft reaction and permit engraftment. They also are designed for tumor eradication in patients with an underlying malignancy. Most conditioning regimens use a combination of radiation and chemotherapy. Doses of total body and total lymphoid irradiation vary between 200 and 1440 cGy. The most commonly used chemotherapy agents in conditioning regimens are alkylating agents (e.g., cyclophosphamide). Conditioning regimens also may contain monoclonal antibodies that target T cells (e.g., alemtuzumab). The choice of a specific conditioning regimen depends on the disease that is being treated. The doses of chemotherapy and radiation used in these regimens are highly variable. When the doses result in a degree of myelosuppression and immunosuppression that is nearly universally fatal without the infusion of HSCs as a rescue product, they are referred to as "myeloablative." Allogeneic HSCT with myeloablative conditioning regimens has been performed successfully in patients older than 60 years of age; however, survival after these

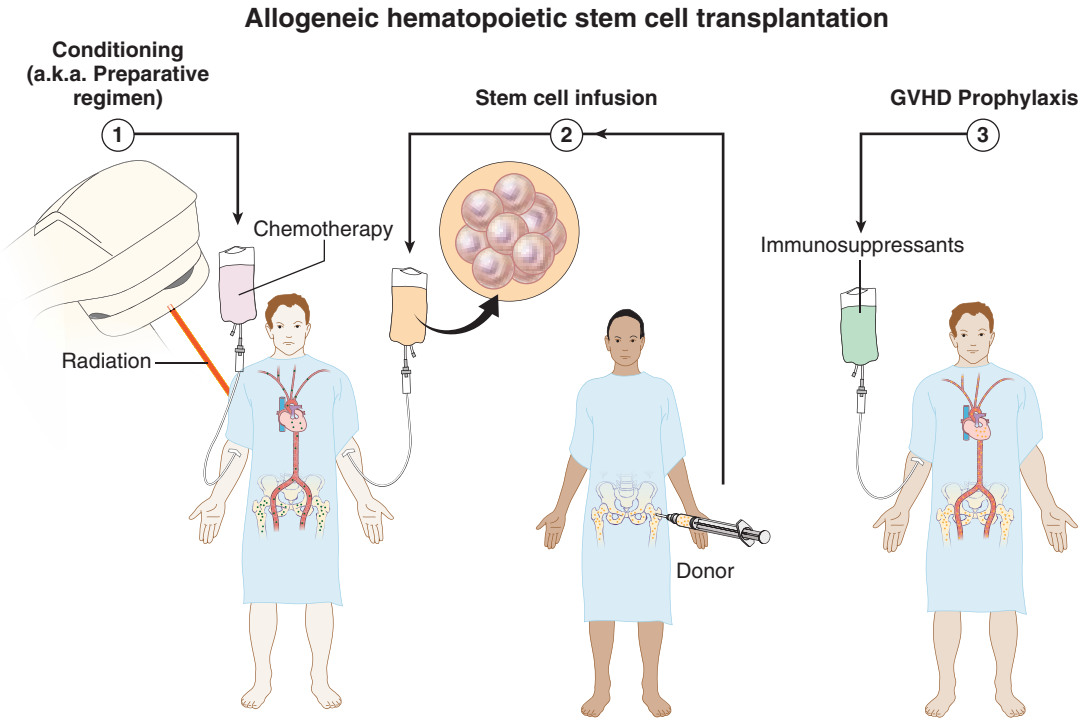

FIGURE 178-1. The process of allogeneic hematopoietic stem cell (HSC) transplantation. (1) The patient receives a conditioning regimen of chemotherapy or radiation therapy (or both) that is used to both eliminate the underlying malignancy and suppress the immune system to prevent rejection of the allogeneic HSCs. (2) Allogeneic HSCs are then collected directly from the bone marrow or from the blood of a related or unrelated stem cell donor and infused intravenously into the donor. (3) After the infusion of the HSCs, the patient receives immunosuppressant drugs to help prevent the development of graft-versus-host disease (GVHD).

transplants declines with increasing age, limiting the application of allogeneic transplantation to a minority of patients who potentially could benefit from this procedure. However, the demonstration that an immune-mediated graft-versus-leukemia effect plays a central role in the therapeutic efficacy of allogeneic HSCT led to the hypothesis that myeloablative conditioning regimens were not essential for tumor eradication. This idea subsequently led investigators to develop less intensive conditioning regimens, which were adequately immunosuppressive to permit the engraftment of donor HSCs but were associated with decreased toxicities compared with myeloablative regimens. These "nonmyeloablative" and "reduced-intensity" conditioning regimens increase the potential option of allogeneic HSCT for older patients and patients with preexisting comorbidities. However, the reduced doses in radiation and chemotherapy result in decreased antitumor activity and are associated with higher rates of disease recurrence after transplant.

Graft-Versus-Host Disease

After engraftment has occurred, patients are at risk for the development of GVHD, which represents the most important clinical challenge with allogeneic HSCT. GVHD is described as either acute, generally presenting within the first 100 days after transplant, or chronic, generally presenting after the first 100 days after transplant. However, clinical features of both acute and chronic GVHD may be observed at any time point after allogeneic HSCT. Current opinion holds that clinical manifestations rather than time after transplantation determine whether the clinical GVHD syndrome is considered acute or chronic. The broad category of acute GVHD includes "classic" acute GVHD, occurring within 100 days after transplantation or DLI and "persistent," "recurrent," or "late" acute GVHD, occurring beyond 100 days after transplantation or DLI. Risk factors for acute GVHD include HLA mismatch, a female donor (particularly a multiparous donor), more advanced age in the patient and the donor, and cytomegalovirus (CMV) seropositivity of the donor or patient, use of an unrelated donor or the use of a T cell–replete versus T cell–depleted graft.

Acute GVHD can often be diagnosed on the basis of clinical findings and is manifest by symptoms in several organ systems, but it primarily affects the skin, gastrointestinal tract, and liver. The skin manifestations range from a maculopapular rash to generalized erythroderma or desquamation. The severity of gastrointestinal GVHD is based on the quantity of diarrhea per day, and the severity of liver GVHD is scored on the basis of the bilirubin level. Organs may be involved in isolation or simultaneously. Histologic confirmation can be valuable in excluding other possibilities such as infection. Mild GVHD of the skin may demonstrate vacuolar degeneration and infiltration of the basal layer by lymphocytes. With more advanced disease, histologic findings of necrotic dyskeratotic cells with acantholysis may progress to frank epidermolysis. In the liver, early GVHD may be difficult to distinguish from hepatitis of other causes. A clinical grading system (Table 178-1) correlates with clinical outcome. Severity is described as grade I (mild) to grade IV (severe).

TABLE 178-1 CLASSIFICATION OF PATIENTS WITH ACUTE GRAFT-VERSUS-HOST DISEASE

	CLINICAL STAGING		
STAGE	**SKIN**	**LIVER**	**GUT**
+	Rash <25% BSA	Total bilirubin, 2-3 mg/dL	Diarrhea, 500-1000 mL/day
++	Rash 25%-50% BSA	Total bilirubin, 3-6 mg/dL	Diarrhea, 1000-1500 mL/day
+++	Generalized erythroderma	Total bilirubin, 6-15 mg/dL	Diarrhea >1500 mL/day
++++	Desquamation and bullae	Total bilirubin >15 mg/dL	Pain with or without ileus

	CLINICAL GRADING STAGE			
GRADE	**SKIN**	**LIVER**	**GUT**	**PS**
0 (none)	0	0	0	0
I	+ to ++	0	0	0
II	+ to +++	+	+	+
III	++ to +++	++ to +++	++ to +++	++
IV	++ to ++++	++ to ++++	++ to ++++	+++

BSA = body surface area; PS = performance status.

The best therapy for GVHD is prophylaxis.[3] The prophylactic use of a calcineurin inhibitor (e.g., cyclosporine, tacrolimus) in combination with methotrexate is effective in reducing the incidence of acute GVHD as well as improving the survival of transplant patients and is the most commonly used

form of GVHD prophylaxis. Cyclosporine is a cyclic polypeptide that prevents T-cell activation by inhibiting interleukin-2 (IL-2) production and IL-2 receptor expression. Although effective as GVHD prophylaxis, cyclosporine imparts significant toxicities, including hypertension, nephrotoxicity, hypomagnesemia, a risk for seizures, hypertrichosis, gingival hyperplasia, tremors, and anorexia. Tacrolimus is a macrolide lactone that closely resembles cyclosporine in mechanism of action, spectrum of toxicities, and pharmacologic interactions. The combination of tacrolimus and methotrexate was demonstrated to be superior to cyclosporine and methotrexate in reducing grade II to IV acute GVHD when used as prophylaxis. For allogeneic HSCT, prophylactic immunosuppression is not lifelong; when immunologic tolerance is established, immunosuppressive agents can be slowly withdrawn and discontinued. The incidence of clinically significant GVHD (grades II-IV) in recipients of HLA-matched sibling grafts (T-cell replete) using tacrolimus and methotrexate for GVHD prophylaxis is approximately 30% to 40%. The incidence of grade II to IV GVHD in recipients of HLA-matched unrelated donor grafts is approximately 50% to 80%. Another strategy to prevent GVHD is to deplete the donor's T cells from the graft; the disadvantage of this approach is its association with increased rates of disease relapse and infection, and overall survival does not seem to be improved.

Moderate to severe GVHD (grades II-IV) requires treatment; the mainstay of therapy is corticosteroids. Methylprednisolone at a dose of 1 to 2 mg/kg/day achieves responses in 40% to 60% of patients. Higher doses of steroids are not of greater benefit. Steroid-refractory GVHD responds poorly to second line therapies such as antithymocyte globulin or various monoclonal antibodies (e.g., daclizumab, infliximab) and are associated with increased mortality rates. In general, whereas acute GVHD of the skin is most responsive to treatment, GVHD of the liver is least responsive. The fatality rate for acute GVHD may be as high as 50%.

Chronic GVHD occurs in 20% to 50% of long-term survivors. Chronic GVHD occurs most commonly between 100 days and 2 years from the transplant and has polymorphic features similar to a number of autoimmune diseases. Chronic GVHD is classified as either classic chronic GVHD, which consists only of manifestations that can be ascribed to chronic GVHD or acute and chronic GVDH overlap syndrome, in which features of both acute and chronic GVHD appear together. These distinctions are clinically relevant as "late-onset" acute GVHD and overlap syndrome subsets have been associated with poor survival rates in some studies. Chronic GVHD is most likely to develop in older patients who also had acute GVHD or received peripheral blood rather than bone marrow grafts. In 20% of cases, there is no history of prior acute GVHD. Adverse prognostic factors include thrombocytopenia, a progressive clinical presentation, extensive skin involvement, and an elevated bilirubin. Common manifestations include the sicca syndrome, lichen planus–like skin rash, scleroderma-like skin changes, esophageal and intestinal fibrosis, obstructive lung disease with or without pneumonitis, and elevated alkaline phosphatase with or without hyperbilirubinemia. Underlying immunologic deficiencies, including hypogammaglobulinemia, are common, placing patients at increased risk for infectious events.

Historically, chronic GVHD has been considered as limited or extensive. Whereas limited disease implies localized skin involvement with minimal or no liver involvement, extensive disease suggests generalized skin involvement with or without other organ involvement. This classification is relatively poorly reproducible and does not always provide useful prognostic information. A new chronic GVHD clinical staging system is now recommended for scoring of individual organs (scale, 0-3) that describes the severity for each affected organ or site at any given time and also measures functional impact. Treatment for chronic GVHD is guided by the extent of disease. Initiation of therapy before functional impairment is of critical importance. Treatments for chronic GVHD include corticosteroids, cyclosporine, thalidomide, ultraviolet light treatments, monoclonal antibodies, or other immunosuppressive agents. Alternative treatments include azathioprine, psoralen ultraviolet A, and extracorporeal photopheresis. Given that the most common cause of death in patients with chronic GVHD remains infection, all patients with GVHD should receive prophylactic antibiotics with or without intravenous immunoglobulin depending on levels.

SYNGENEIC HEMATOPOIETIC STEM CELL TRANSPLANTATION

Syngeneic HSCT uses stem cells from an identical twin. Because the HSCs are genetically identical with the recipient, the major advantage of a syngeneic HSCT is that it is not associated with GVHD or graft rejection, resulting in a relatively low risk of treatment-related morbidity and mortality. Another advantage of syngeneic HSCT, shared with allogeneic HSCT, is a lack of contamination of the graft by malignant cells. The major disadvantage of syngeneic HSCT is that it does not provide the graft-versus-leukemia effect associated with allogeneic HSCT. However, fewer than 1% of patients have an identical twin, and consequently, this is not an option for most patients.

AUTOLOGOUS HEMATOPOIETIC STEM CELL TRANSPLANTATION

The major rationale for autologous HSCT, using the patient's own HSCs, is that certain malignancies, such as leukemias and lymphomas, have a steep dose-response curve to chemotherapy and, to a relative degree, radiation.[4] However, the major limitation to the administration of higher doses of chemotherapy or radiation is the myelosuppressive effects of these therapies. Autologous HSCs (and for that matter, allogeneic and syngeneic HSCs) permit the administration of high-dose chemotherapy with or without radiation by restoring hematopoiesis. The major advantages of autologous HSCT compared with allogeneic HSCT are that (1) the patient can serve as his or her own donor, and (2) it may be performed in older patients with significantly lower morbidity and mortality because of the absence of GVHD as a major complication. However, autologous HSCT can be associated with more morbidity than conventional doses of chemotherapy. Although patients undergoing autologous transplantation have higher relapse rates than those undergoing allogeneic transplantation, the lower rate of other complications seems to translate into similar long-term outcomes.

INDICATIONS FOR TRANSPLANTATION

An estimated 60,000 HSC transplants are performed annually worldwide. They are used primarily for the treatment of hematologic malignancies, bone marrow failure states, and immune and enzyme deficiencies (Fig. 178-2).

Acute Myeloid Leukemia

Acute myeloid leukemia (AML) (Chapter 183) was one of the first malignancies in which all forms of HSCT were demonstrated to be effective. Allogeneic HSCT is capable of resulting in the long-term survival of 10% of patients with refractory AML. Long-term survival and an apparent cure rate of 20% to 40% have been achieved in patients treated in second or subsequent complete remission, and cure rates of 40% to 70% have been reported in patients given transplants in first complete remission.[5] Randomized controlled trials comparing autologous and allogeneic HSCT with conventional chemotherapy in patients with AML in first complete remission have demonstrated improved leukemia-free survival with both forms of HSCT. However, the data vary regarding an improvement in overall survival. Several meta-analyses demonstrate improved overall survival with allogeneic HSCT for AML in first complete remission, compared with nonallogeneic treatments in patients with intermediate- and high-risk cytogenetics but not for good-risk AML.[6] More recent studies indicate the importance of the presence of molecular mutations. For example, intermediate-risk AML with a normal karyotype represents a highly heterogeneous group with respect to prognosis based on molecular mutation status. Approximately one third of patients with normal karyotype AML harbor the *FLT3*-internal tandem mutation, which carries a poor prognosis, and may benefit from HLA-matched related HSCT transplantation, regardless of the presence of other mutations.

Acute Lymphocytic Leukemia

The results of conventional chemotherapy for acute lymphocytic leukemia (ALL) in children are excellent, except for ALL associated with the Philadelphia chromosome (Ph). In adults, however, although remission is frequently attained after intensive induction therapy, the probability of relapse is high is this disease, and some form of consolidation therapy is recommended, including allogeneic HSCT. Adverse prognostic factors include the presence of Ph, high white blood cell count, advancing age, and the presence of minimal residual disease. Studies suggest that HSCT improves overall survival compared with conventional chemotherapy alone, particularly for those who are Ph positive. The overall prognosis for both pediatric and adult patients with relapsed ALL is relatively poor, and the general treatment strategy is to obtain a second complete remission and then proceed to an allogeneic HSCT.

Myelodysplastic Syndrome

The only known curative treatment for myelodysplastic syndrome (MDS) (Chapter 182) is allogeneic HSCT. Unfortunately, the vast majority of patients are not considered candidates for this therapy for many reasons, including advanced age, comorbidities, donor availability, access to relatively

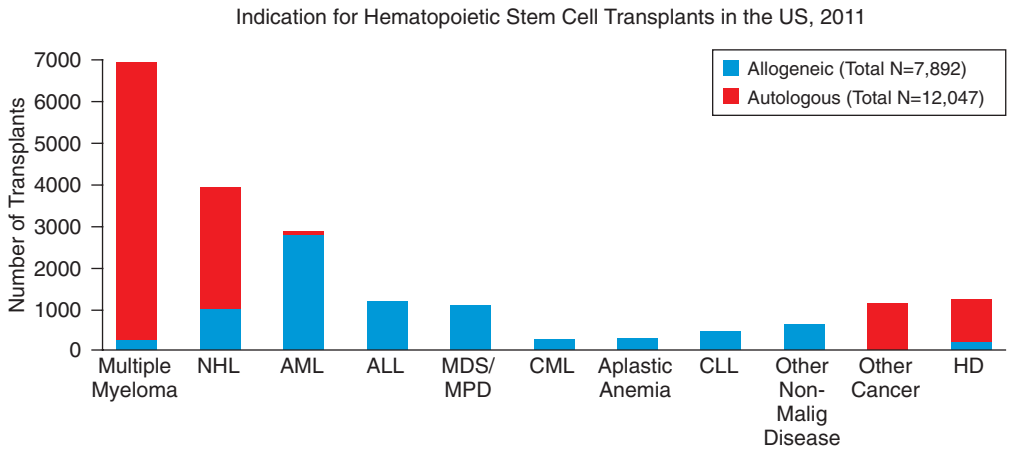

FIGURE 178-2. Indications for hematopoietic stem cell transplantation in the United States as Reported to the Center for International Blood & Marrow Research. AA = aplastic anemia; ALL = acute lymphocytic leukemia; AML = acute myeloid leukemia; CLL = chronic lymphocytic leukemia; HD = Hodgkin disease; MDS/MPD = myelodysplastic syndrome/ myeloproliferative disorders; MM = multiple myeloma; NHL = non-Hodgkin lymphoma. (From Dunn, R. Current Uses and Outcomes of Hematopoietic Stem Cell Transplantation 2013: Summary Slides. Center for International Blood & Marrow Transplant Research. http://www.cibmtr.org/referencecenter/slidesreports/summaryslides/documents/2013 summary slides- final web version v2 4.14.2014.pptx)

well tolerated agents such as azacytidine or lenalidomide, and clinician or patient preferences. The best results have been obtained in relatively younger patients, who are earlier in their disease course and have not received any prior therapy. There is increasing evidence that reduced-intensity allogeneic HSCT may benefit even considerably older patients with MDS, and this is of importance given that the median age at diagnosis is in the seventh to eighth decades of life. The use of autologous HSCT for MDS remains investigational.

Chronic Myeloid Leukemia

Previously, allogeneic HSCT was the treatment of choice for chronic myeloid leukemia (CML) (Chapter 184); however, currently most CML patients are successfully treated with a tyrosine kinase inhibitor (TKI), such as imatinib. Consequently, in most cases, allogeneic HSCT is not performed in patients with CML in chronic phase, but it is reserved for patients in accelerated phase (AP) or blast crisis (BC) CML, those who are intolerant of or fail TKIs, and those with TKI-resistant mutations of *BCR-ABL*. Although the results of allogeneic HSCT in AP or BC CML are poor, results of posttransplant treatment with TKIs appear promising. In the TKI era, autologous stem cell collections, even in patients with hematopoiesis shown to be *BCR-ABL* negative by reverse transcription polymerase chain reaction, are rarely performed.

Myeloproliferative Neoplasms

Myeloproliferative neoplasms (Chapter 166), such as primary myelofibrosis, polycythemia vera, and essential thrombocythemia, usually are chronic in nature but can progress to a "spent" phase and develop myeloid metaplasia, which is characterized by bone marrow fibrosis and a generally poor prognosis with transformation into acute leukemia and a median survival time of less than 3 years. Conventional treatment options in this disease state are limited, and an accepted standard of care for myeloid metaplasia/myelofibrosis is allogeneic HSCT with increasing reports of the efficacy of nonmyeloablative allogeneic HSCT for these disorders in this older age population.

Non-Hodgkin Lymphoma

Although allogeneic, syngeneic, and autologous HSCT have all been reported to yield long-term, disease-free survival and an apparent cure for patients with advanced non-Hodgkin lymphomas (NHL) (Chapter 185), the current standard of care for patients with primary refractory or chemotherapy-sensitive relapsed NHL of specific histologies, including diffuse large B-cell lymphoma, remains autologous HSCT.

Autologous HSCT also has been used to treat patients with indolent follicular NHL resulting in disease-free survival rates as high as 60%. However, late relapses and the long overall survival period observed with conventional therapy make long-term follow-up necessary to document the efficacy of this approach. The demonstration of a potent graft-versus-leukemia effect against NHL is less clear. Consensus is currently lacking regarding the most appropriate role of HSCT (whether autologous or allogeneic) for patients with follicular lymphoma. Recent studies, however, suggest the potential value of nonmyeloablative allogeneic HSCT regimens in the context of clinical trials that incorporate rituximab or radioimmunoconjugates.

In the case of mantle cell lymphoma (MCL), further prospective trials are required, although autologous HSCT appears to improve progression-free and possibly overall survival in patients when used as part of front-line therapy. Patients with high-risk MCL are candidates for trials investigating the role of nonmyeloablative allogeneic HSCT.

Hodgkin Lymphoma

Based on a small prospective randomized trial conducted in the early 1990s, autologous HSCT has become the standard of care for patients with primary refractory and relapsed Hodgkin lymphoma (Chapter 186). The usual approach is to first treat these patients with second-line chemotherapy followed by the high-dose therapy and autologous HSCT. Allogeneic HSCT has had a limited role because of the efficacy of autologous HSCT and the significant treatment-related toxicities associated with myeloablative allogeneic HSCT. Data are accumulating on the efficacy of reduced-intensity allogeneic HSCT, especially in patients who relapse after autologous HSCT.

Multiple Myeloma

Although the advent of new agents, such as immunomodulatory derivatives and proteasome inhibitors, for the treatment of multiple myeloma (MM) (Chapter 187) is leading to a reappraisal of the role of autologous HSCT in the early treatment of the disease, current opinion indicates that standard of care remains induction therapy with novel agents followed by autologous HSCT.[7] Prospective comparisons of single versus tandem autologous transplants give conflicting results, although overall survival rates appear similar. Comparisons of tandem autologous versus autologous followed by allogeneic HSCT indicate improved complete remission rates and event-free survival but no increased overall survival and significantly worse nonrelapse mortality rates with the latter. Consensus appears to be to reserve a second transplant for patients who relapse after the first autologous HSCT, provided the duration of response was more than 1 year after the initial transplant.

Solid Tumors

High-dose chemotherapy plus transplantation has had success in the treatment of some chemotherapy-sensitive solid tumors, including germ cell tumors and childhood cancers such as neuroblastoma and Wilms tumor. In patients with germ cell tumors for whom platinum-based chemotherapy regimens fail to result in a cure, the use of high-dose chemotherapy and autologous HSCT has resulted in prolonged disease-free survival time, including patients with refractory disease. After initial promising results of autologous HSCT for advanced and metastatic breast cancer, several randomized trials failed to demonstrate an overall survival benefit, and the procedure is no longer performed for these indications.

Nonmalignant Conditions

Hematopoietic stem cell transplantation is also effective for treatment of nonmalignant disorders, including aplastic anemia, thalassemias, sickle cell disease[8], immunodeficiency disorders, and enzyme deficiency states. These latter indications are primarily for children and young adults. Allogeneic

HSCT can lead to long-term, disease-free survival times in more than 50% of patients with severe aplastic anemia (Chapter 165). When compared with standard immunosuppressive therapy, allogeneic transplantation is more likely to produce a complete and durable reversal of the hematologic abnormalities. Patients with aplastic anemia who are less heavily transfused have better outcomes with allogeneic transplantation. For patients with less severe aplastic anemia, patients older than 40 years, and those without a matched sibling donor, a trial of immunosuppression therapy is usually appropriate before consideration of allogeneic transplantation.

The hemoglobinopathies can be cured only by allogeneic HSCT, and the most extensive experience is with β-thalassemia (Chapter 162). Best results are obtained with HLA-identical sibling donor transplantations and in pediatric patients. Adults tend to have more advanced disease with greater iron overload and more organ dysfunction; hence, they have high treatment-related mortality rates. Studies with nonmyeloablative regimens are ongoing. There is considerably less experience with sickle cell disease (Chapter 163), partly because of reluctance to contemplate allogeneic HSCT in patients with an unpredictably variable clinical course.

Autologous HSCT has been investigated for almost 20 years for the treatment of severe autoimmune disease, including systemic lupus erythematosus, rheumatoid arthritis, scleroderma, and multiple sclerosis. Results have been promising in nonrandomized trials with approximately one third of patients exhibiting clinically significant responses, including medication-free remissions. The strategy has been to reconstitute the immune system without the presence of autoreactive T-cell clones. There has been a reluctance to consider allogeneic HSCT, with its attendant morbidity and mortality, because in general patients with autoimmune diseases have a low probability of death from their underlying disorder, but HSCT can confer a long-term survival benefit in patients with diffuse cutaneous systemic sclerosis.

COMPLICATIONS AFTER HEMATOPOIETIC STEM CELL TRANSPLANTATION

Early complications associated with all forms of HSCT include direct organ toxicities from the conditioning regimen (e.g., mucositis) and prolonged cytopenias (7-14 days) resulting in infections and bleeding. As described earlier in this chapter, allogeneic HSCT can be associated with acute and chronic GVHD. Long-term complications, particularly secondary malignancies, require that patients continue to be monitored for the remainder of their lives. Through improved supportive measures, there has been steady improvement in survival after HSCT over time.[9] A simple index, based on pretransplant comorbidities, has been developed that reliably predicts nonrelapse mortality and survival.[10] This comorbidity index is useful for patient counseling before HSCT. The late toxicities of HSCT must always be kept in mind when considering the option of HSCT for patients.

Graft Rejection

Graft rejection occurs when immunologically competent cells of host origin destroy the transplanted cells of donor origin. This complication occurs more commonly in patients who receive transplants from alternative or HLA-mismatched donors, in T cell–depleted transplants, and in patients with aplastic anemia. Graft rejection is rarely, if ever, observed in patients undergoing autologous or syngeneic HSCT.

Infections

Infections are a major cause of morbidity and mortality after all forms of HSCT, but especially after allogeneic transplantation because of prolonged use of immunosuppression for the prevention or treatment of GVHD. Bacterial infections are frequently related to central venous catheters. Prophylactic antibiotics significantly reduce the incidence of infection but not mortality. *Aspergillus* infections typically occur in patients receiving prolonged high-dose steroids for the treatment of GVHD.[11] Viral infections include reactivation of CMV, human herpes virus 6, and Epstein-Barr virus (EBV) infections. Brincidofovir, an investigational oral nucleotide analog, and letermovir, a novel antiviral agent, reduce CMV events in recipients of HSCT. Post-HSCT patients are also susceptible to seasonal respiratory viruses. Revaccinations for common childhood infections are required after transplantation.

Cardiac Toxicity

Most transplant centers screen potential patients for underlying cardiac abnormalities that would place them at a potentially increased risk during the procedure. Despite this screening, however, a small number of patients experience cardiotoxicity, either acutely during the transplant or at a later time, manifest as a cardiac arrhythmia, congestive heart failure, or cardiac ischemia because of the large volumes of fluids administered during the procedure or from the added physiological stress. Complications associated with a pericardial effusion can be seen in some patients during or after transplant and are more common in patients with disease near that area and those receiving radiation therapy in that field. An idiosyncratic cardiomyopathy, associated with the administration of high doses of cyclophosphamide, has been documented in a small number of patients. Viral cardiomyopathies also can be seen.

Engraftment Syndrome

Engraftment syndrome occurs during neutrophil recovery after both autologous and allogeneic HSCT. It consists of a constellation of symptoms and signs that may include fever, erythrodermatous skin rash, and noncardiogenic pulmonary edema, and, in its most extreme forms, acute renal failure and diffuse alveolar hemorrhage. These clinical findings reflect the manifestations of increased capillary permeability and extensive endogenous cytokine. Corticosteroid therapy is often dramatically effective for engraftment syndrome, particularly for the treatment of the pulmonary manifestations.

Pulmonary Toxicities

Pulmonary toxicities are common during and after transplantation. Patients who receive certain chemotherapeutic agents, such as 1,3-bis (2-chloroethyl)-1-nitrosourea (BCNU; carmustine) have an increased incidence of chemotherapy-induced lung tissue damage after transplant, which usually can occur several weeks after transplantation and can be treated successfully with the prompt initiation of corticosteroid therapy. In addition to these complications, patients who are undergoing allogeneic HSCT are at increased risk for pneumonitis caused by CMV and fungal infections due to the patient's increased immunosuppression, and adult respiratory distress syndrome or interstitial pneumonia of unknown etiology. Chronic GVHD also can manifest as bronchiolitis obliterans in the lung.

Liver Toxicity

The most common liver complication associated with transplantation is veno-occlusive disease (VOD)/sinusoidal obstruction syndrome (SOS) of the liver. Symptoms associated with VOD/SOS include jaundice, tender hepatomegaly, ascites, and weight gain. Progressive hepatic failure and multiorgan system failure can develop in the most severe cases. Predisposing factors appear to be previous hepatic injury, use of estrogens, and high-dose intensity conditioning.

Renal Toxicity

Acute renal failure requiring dialysis during the transplant occurs infrequently, although patients with underlying renal dysfunction are at risk for this complication. An idiopathic or cyclosporine-induced hemolytic-uremic syndrome can be a serious complication after allogeneic HSCT and poses a high mortality risk or can result in end-stage renal disease. Recently, nephrotic syndrome and membranous nephropathy have been described in long-term survivors; these complications seem to be associated more commonly with chronic GVHD and nonmyeloablative conditioning.

Secondary Malignancies

One complication of the chemotherapy or radiation therapy (or both) used to treat malignancy is the development of a secondary malignancy. There have been several reports of the development of secondary AML or MDS after autologous transplantation. Some studies have suggested that total-body irradiation may increase the risk for these complications. After allogeneic transplantation, the overall incidence of secondary malignancies is 2.2% at 10 years and 6.7% at 15 years after transplantation. Within the first 1 to 2 years, the most common malignancies are EBV-related lymphoproliferative disorders; solid tumors are more likely to occur more than 3 years after transplantation. Risk factors include the use of antithymocyte globulin to treat GVHD, the use of a T cell–depleted allogeneic graft, HLA incompatibility, and perhaps total-body irradiation.

Infertility and Hypogonadism

Many of the preparative regimens used for transplant are associated with a high incidence of permanent sterility, particularly regimens containing total-body irradiation. However, successful pregnancies have occurred in some patients after other regimens, particularly in younger patients. Gynecomastia

occasionally occurs in males. A reproductive endocrinologist should be consulted before transplantation in patients for whom future fertility is important.

Endocrine Dysfunction

Iatrogenic Cushing syndrome and diabetes can occur and are commonly caused by long-term steroid therapy for chronic GVHD. Particularly disabling are steroid-induced myopathy, avascular necrosis of the hip, and osteoporosis. Because many patients take steroids for many months, tapering can be associated with malaise, nausea, hypotension, and musculoskeletal pains. In these situations, slower tapering over several months or reintroduction of physiological replacement doses (e.g., 5 to 7.5 mg/day of prednisone) is appropriate. Hypothyroidism is typically related to the use of total-body irradiation or local irradiation of the head and neck for lymphoma or other cancers. Osteoporosis occurs in 50% to 60% of patients after HSCT. The major contributing causes include hypogonadism, secondary hyperparathyroidism caused by low serum calcium, and posttransplant steroid therapy. Bone mineral density should be evaluated before and after transplantation; osteopenia should be treated as appropriate with bisphosphonates, calcium, vitamin D, estrogen, and testosterone (Chapter 243).

CONCLUSION

There has been much progress in increasing the safety of HSCT and in expanding the application of this treatment to more patients. Areas currently under development that may further improve the use and efficacy of transplantation include continuous improvements in supportive care for transplant patients and the broadened use of alternative donors. Better treatments are necessary for the treatment of GVHD, and innovative studies are ongoing. Future progress depends on the ability to identify safer and better-targeted antitumor therapies that can be incorporated in transplantation regimens without increasing toxicity or attenuating graft-versus-tumor responses. However, novel cellular therapies, including genetic modification of lymphocytes to enhance cancer-killing activity, hold considerable promise.

Grade A References

A1. Holtick U, Albrecht M, Chemnitz JM, et al. Bone marrow versus peripheral blood allogeneic haematopoietic stem cell transplantation for haematological malignancies in adults. *Cochrane Database Syst Rev*. 2014;4:CD010189.

A2. van Laar JM, Farge D, Sont JK, et al. Autologous hematopoietic stem cell transplantation vs intravenous pulse cyclophosphamide in diffuse cutaneous systemic sclerosis: a randomized clinical trial. *JAMA*. 2014;311:2490-2498.

A3. Kimura S, Akahoshi Y, Nakano H, et al. Antibiotic prophylaxis in hematopoietic stem cell transplantation. A meta-analysis of randomized controlled trials. *J Infect*. 2014;69:13-25.

A4. Marty FM, Winston DJ, Rowley SD, et al. CMX001 to prevent cytomegalovirus disease in hematopoietic-cell transplantation. *N Engl J Med*. 2013;369:1227-1236.

A5. Chemaly RF, Ullmann AJ, Stoelben S, et al. Letermovir for cytomegalovirus prophylaxis in hematopoietic-cell transplantation. *N Engl J Med*. 2014;370:1781-1789.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Gyurkocza B, Rezvani A, Storb R. Allogeneic hematopoietic cell transplantation: the state of the art. *Expert Rev Hematol*. 2010;3:285-299.
2. Ballen KK, Gluckman E, Broxmeyer HE. Umbilical cord blood transplantation: the first 25 years and beyond. *Blood*. 2013;122:491-498.
3. Pavletic SZ, Fowler DH. Are we making progress in GVHD prophylaxis and treatment? *Hematology Am Soc Hematol Educ Program*. 2012;2012:251-264.
4. Hamadani M. Autologous hematopoietic cell transplantation: An update for clinicians. *Ann Med*. 2014:1-14.
5. Liu H, Stock W, Bishop MR. Expanded Indications for allogeneic stem cell transplantation in patients with myeloid malignancies. *Curr Opin Hematol*. 2013;20:115-122.
6. Thomas X. Current indications of allogeneic stem cell transplant in adults with acute myeloid leukemia. *Bull Cancer*. 2014;101:856-865.
7. Gahrton G, Krishnan A. Allogeneic transplantation in multiple myeloma. *Expert Rev Hematol*. 2014;7:79-90.
8. Hsieh MM, Fitzhugh CD, Weitzel RP, et al. Nonmyeloablative HLA-matched sibling allogeneic hematopoietic stem cell transplantation for severe sickle cell phenotype. *JAMA*. 2014;312:48-56.
9. Gooley TA, Chien JW, Pergam SA, et al. Reduced mortality after allogeneic hematopoietic-cell trans- plantation. *N Engl J Med*. 2010;363:2091-2101.
10. Sorror ML. How I assess comorbidities before hematopoietic cell transplantation. *Blood*. 2013;121:2854-2863.
11. Girmenia C, Ferretti A, Barberi W. Epidemiology and risk factors for invasive fungal diseases in hematopoietic stem cell transplantation. *Curr Opin Hematol*. 2014;21:459-465.

REVIEW QUESTIONS

1. The distinctive characteristic(s) of allogeneic hematopoietic stem cell transplantation include(s):

A. The stem cell graft is free of contamination by malignant cells
B. Associated with decreased complications compared with autologous and syngeneic hematopoietic stem cell transplantation
C. Contains lymphocytes that are capable of mediating an immunologic reaction against foreign antigens
D. Both A and C
E. A, B, and C

Answer: D Both allogeneic and syngeneic grafts are free of tumor cell contamination. Allogeneic hematopoietic stem cell transplantation (HSCT) is associated with higher rates of treatment-related morbidity and mortality because of the potential for the development of graft-versus-host disease and need for immunosuppressive agents leading to higher risks of infection. Allogeneic transplantation contains lymphocytes that can mediate immunologic reactions against foreign antigens, including those on tumors leading to lower rates of relapse compared with either autologous or allogeneic HSCT.

2. The cells that contribute most to the graft-versus-tumor effects observed with allogeneic stem cell transplantation are

A. neutrophils.
B. monocytes.
C. dendritic cells.
D. T lymphocytes.
E. B lymphocytes.

Answer: D Although neutrophils and monocytes play important roles in eliminating infections, they contribute little to observed graft-versus-tumor effect. B lymphocytes and dendritic cells have been demonstrated to play roles in the graft-versus-tumor effects observed with allogeneic hematopoietic stem cell transplantation ; however, T lymphocytes play the most important role. This has been most clearly evident clinically by the observation that infusion of donor T lymphocytes (i.e., donor lymphocyte infusion) after recurrence of cancer following allogeneic stem cell transplantation can alone result in sustained complete remissions.

3. All of the following are disadvantages for the clinical use of cord blood hematopoietic stem cells (HSCs) except

A. clinical outcomes are in general inferior to fully HLA-matched sibling and volunteer unrelated donors.
B. because of the unique immature biology of lymphocytes in umbilical cord blood, these transplants are associated with more GVHD.
C. because of the small volume (50-150 mL) of cord blood, only a limited number of HSCs can be collected, which often prohibits their use in adults because successful engraftment correlates with the number of HSCs per patient body weight.
D. cord blood HSCs are associated with delayed times to engraftment and immune reconstitution, leading to an increased rate of infection.
E. after a cord blood unit is used, there is no possibility to obtain additional cells in the event of graft failure or if a donor lymphocyte infusion is required.

Answer: B Cord blood transplantation is generally associated with less graft-versus-host disease even when the donor and recipient are human leukocyte antigen (HLA) disparate. This is the major advantage of cord blood stem cells and permits the application of hematopoietic stem cell transplantation to a broader group of patients. Outcomes, however, are generally not as good as those observed with either fully HLA-matched sibling or volunteer donors in similar patients. The limited number of HSCs, delayed hematopoietic recovery and immune reconstitution, and availability of only a single unit for each patient are major clinical disadvantages.

4. Risk factors for the development of acute graft-versus-host disease (GVHD) include all except which of the following?

A. HLA mismatch
B. Cytomegalovirus seronegativity of the donor or patient
C. More advanced age in the patient and the donor
D. Female donor (particularly a multiparous donor)
E. Use of an unrelated donor

Answer: B Seropositivity, not negativity, in either the recipient or the donor is associated with a greater risk of developing acute GVHD. All of the other factors have also been associated with higher risks of developing acute GHVD. When a patient has more than one potential human leukocyte antigen–matched donors, these factors are used in the donor selection.

5. All of the following are considered standard indications for hematopoietic stem cell transplantation (HSCT) except

A. allogeneic HSCT from HLA-matched donors for patients with intermediate- and high-risk acute myeloid leukemia in first remission.
B. high-dose chemotherapy and autologous HSCT for patients with primary refractory and relapsed diffuse large B-cell lymphoma that remains chemotherapy sensitive.
C. allogeneic HSCT from HLA-matched sibling in adult patients with advanced sickle cell disease.
D. high-dose chemotherapy and autologous HSCT for patients with newly diagnosed multiple myeloma after induction therapy.
E. allogeneic HSCT for patients with therapy-related acute myeloid leukemia.

Answer: C In all of the other clinical scenarios, either allogeneic or autologous HSCT has been demonstrated to significantly improve progression-free or overall survival. Several meta-analyses have demonstrated the benefit of allogeneic HSCT from human leukocyte antigen–matched donors for patients with intermediate- and high-risk acute myeloid leukemia in first remission. The benefit in patients with normal cytogenetics is being further delineated by molecular studies (e.g., FLT3-ITD). In relapsed, chemotherapy-sensitive diffuse large B-cell lymphoma, autologous HSCT has been demonstrated to result in improved progression-free and overall survival in randomized trials compared with conventional therapy. Although the continuing introduction of novel agents is changing clinical practice, autologous transplantation remains a standard component in the initial management of multiple myeloma patients. In the case of therapy-related acute myeloid leukemia, allogeneic stem cell transplantation is the only option that can provide long-term survival and possibly cure. Although there have been promising results with allogeneic HSCT in adults with advanced sickle cell disease, it still is considered investigational because ongoing trials attempt to identify the most appropriate patients in which to use this treatment.

179 APPROACH TO THE PATIENT WITH CANCER

JAMES H. DOROSHOW

INTRODUCTION TO THE CANCER PATIENT

Conveying or receiving an initial diagnosis of cancer, or the knowledge that cancer has recurred, are among the most difficult of human enterprises, and no amount of either specialized training or forewarning can adequately assuage the intensity of the emotions associated with these encounters. Patients often experience a storm of feelings that may limit useful discussion immediately following the receipt of a diagnosis of cancer. Unquestionably, it is difficult for even the most well-informed patient to process the complexities of his or her individual situation, including the range of additional diagnostic tests that may be required and the potentially vast array of treatment options and potential outcomes that lie ahead. And yet, at some point prior to the initiation of treatment, the physician and the patient must discuss the diagnosis, its implications, and therapeutic alternatives. Because of the wide range of potential prognoses (from curable disseminated testicular cancer to the limited lifespan of patients with locally advanced gastric or pancreatic cancer), it is often useful for family members or close friends to be present in the consulting room when detailed discussions of the complexities of either disease or therapy are conducted, both to provide emotional support and to be another "set of ears" during the visit. It is helpful to ask patients directly: "What do you understand about your diagnosis and treatment?" Family members or close associates of the patient may also be especially helpful in developing a written or digital record of the questions posed to and answered by health care providers; many patients find such a record to be especially helpful for later reference.

If the physician is not familiar with the latest treatment options, prompt referral to a specialist, whether a surgical oncologist, radiation oncologist, or medical oncologist, is imperative. The generalist should not be a therapeutic nihilist unless he or she is intimately involved in the field and is well versed in the potential risks and benefits of currently available therapies and clinical trials.

DIAGNOSIS

Diagnostic possibilities are protean for the wide range of human malignancies that may be discovered either in the presence of nonspecific but foreboding symptoms or signs (severe weight loss, hematuria, jaundice) or in asymptomatic individuals (e.g., during a routine physical examination). The importance of the medical history and physical examination, however, must be emphasized *whether or not* a pathologic diagnosis of cancer has already been confirmed. One of the most important considerations that underlies the approach to both the diagnostic work-up and choice of cancer treatment (surgical, radiation, or systemic therapy) is the patient's basic physiologic condition or "performance status" (Table 179-1). For example, a past medical history of prolonged tobacco smoking is not only relevant to a possible diagnosis of lung cancer but to the ability of a patient to tolerate potentially curable multimodality treatment. Underlying evidence of excessive alcohol consumption may play a role in both the tolerance for and metabolism of systemic chemotherapeutic agents. Family histories of cancer can provide prognostic indicators as well as suggest molecularly guided treatment approaches in, for example, women with possible *BRCA1*-related breast or ovarian cancer. It is also critical to assess the environment of care, including all of the patient's support systems, that may be sorely tested by the experience of a cancer diagnosis and the interventions that ensue therefrom. Finally, the physical examination will define the extent of certain sites of measurable malignancy—lymph node, splenic, or hepatic enlargement, for example—as well as the patient's muscle strength, the presence of possible malignant effusions, and potential central or peripheral neuropathies reflective of metastatic disease or paraneoplastic syndromes. But most importantly, it provides the treating physician with an initial sense of the patient's well-being, or lack thereof, prior to the initiation of therapy.

Diagnostic Procedures

A lesion that has been found on physical examination, or following radiographic studies prompted by abnormal laboratory results, will often undergo a percutaneous biopsy for pathologic evaluation. It is critical that the biopsy be representative of the entire tumor and be robust enough in size that appropriate investigations (e.g., special immunohistologic stains, flow cytometry, cytogenetics, hormone assays) can be performed before treatment is initiated. If there is a question whether the lesion is benign or malignant or about its proper classification, consideration should be given to additional biopsies, and consultation with a reference pathologist may be indicated. The recent emergence of a wide range of molecularly targeted systemic therapeutic agents active in solid tumors (see Table 179-4) has focused renewed attention on obtaining sufficient tissue from patients for performance of the essential molecular studies (DNA sequencing, RNA expression analyses, FISH [fluorescence in situ hybridization]) necessary to determine treatment choice. Although tumors removed by surgical resection are routinely of sufficient size to permit the full range of diagnostic examinations, active involvement of a skilled interventional radiologist or endoscopist is often required to produce both the size and number of tumor biopsies required for modern cancer therapeutic decision making. Finally, there is seldom a need for such rapid therapy that appropriate pretreatment evaluations cannot be performed. For some tumor sites such as the colon (Chapter 193), there is one predominant histology but critical molecular features of the tumor (presence or absence of a mutant *Ras* oncogene) that can define therapy; in others, such as the lung (Chapter 191), the distinction between small cell lung cancer and non–small cell lung cancer is critical for treatment. For breast cancer (Chapter 198), the treating physician is interested in a variety of factors, such as histology, tumor grade, the presence (and its degree) or absence of estrogen and progesterone receptor proteins, and the presence of HER2/neu overexpression. The rapidly increasing sophistication of molecular diagnostics has, furthermore, begun to improve the potential to localize cancers of unknown primary origin (Chapter 204).

Staging and Multidisciplinary Evaluation

After a tissue diagnosis has been established, staging follows to determine the extent of disease. The American Joint Committee on Cancer staging system is considered the standard in the United States and is based on the TNM (tumor, node, metastasis) system that is anatomically and pathologically based. The approach to staging from a clinical perspective depends on the type of cancer, but it commonly includes computed tomography (CT), magnetic resonance imaging (MRI), radionuclide scans, and, increasingly, positron emission tomography (PET). These studies are supplemented by routine hematologic and chemistry profiles, tumor markers (when appropriate), and in some cases, bone marrow aspiration and biopsy.

The goal of tumor staging is to define the extent of a patient's disease. Tumor stage provides critical prognostic information that will inform the therapeutic approach that is most appropriate. Accurate staging involves delineating the magnitude of the tumor determined by imaging procedures, as well as confirming the pathologic limits of disease spread from tissues removed at surgery. In essence, for most solid tumor patients, the tumor stage will establish whether the treatment will focus on a local (usually confined to an organ), regional, or disseminated pattern of malignancy and can determine whether the expected outcome of therapy is curative or palliative. On the other hand, hematopoietic malignancies are often disseminated at diagnosis and demand their own prognostic classifiers.

The focus of the staging work-up is to identify potential sites of metastases and to establish indicator lesions with which to monitor therapy. For most solid tumors, CT scans can accomplish both goals; however, in some circumstances, other imaging modalities are more appropriate for therapeutic monitoring (e.g., MRI when central nervous system [CNS] metastases are likely [small cell lung cancer], or combined PET/CT imaging to establish that a given lesion is likely to be malignant, or in diseases where early metabolic responses to treatment can be confirmed (such as for gastrointestinal stromal tumors). In patients with established advanced disease, indicator lesions for therapeutic monitoring should be carefully chosen prior to the initiation of treatment, be well documented in the medical record, and should be evaluated with the minimum frequency of imaging procedures required for accurate follow-up, consistent with evidence-based medical practice or the clinical protocol on which the patient is entered.

The consulting medical oncologist may often be advised by a local tumor board composed of other medical, surgical, and radiation oncologists,

TABLE 179-1 KARNOFSKY AND ZUBROD PERFORMANCE SCALES

KARNOFSKY PERFORMANCE STATUS SCALE

VALUE	LEVEL OF FUNCTIONAL CAPACITY
100	Normal, no complaints, no evidence of disease
90	Able to carry on normal activity, minor signs or symptoms of disease
80	Normal activity with effort, some signs or symptoms of disease
70	Cares for self, unable to carry on normal activity or to do active work
60	Requires occasional assistance but is able to care for most needs
50	Requires considerable assistance and frequent medical care
40	Disabled, requires special care and assistance
30	Severely disabled; hospitalization is indicated, although death is not imminent
20	Hospitalization is necessary; very sick, active supportive treatment necessary
10	Moribund; fatal processes progressing rapidly
0	Dead

EASTERN COOPERATIVE ONCOLOGY GROUP (ZUBROD) PERFORMANCE SCALE

PERFORMANCE STATUS	DEFINITION
0	Asymptomatic
1	Symptomatic; fully ambulatory
2	Symptomatic; in bed < 50% of day
3	Symptomatic; in bed > 50% of day
4	Bedridden

pathologists, radiologists, and members of the cancer care team (oncology nurses, social workers, and palliative care specialists). In such a multidisciplinary environment, the patient's overall prognosis can be reviewed and alternatives for care, including standard therapy, possible clinical trials, a second opinion, or no treatment, can be considered. The outcome of such an evaluation is usually viewed by patients and families as an important component in the coordinated development of an overall plan for either further diagnostic procedures or therapy. Finally, many oncologists actively participate in clinical trials that may make investigational drugs or other investigational procedures available for patients, or they may suggest referral to a tertiary cancer center where disease-specific clinical trials are available, as appropriate.

TREATMENT

Therapeutic Plan

Intention of Treatment

Based on the multidisciplinary evaluation of an accurately staged patient, it should be possible to define whether the intent of treatment is curative or palliative. Whether or not this is done during a formal tumor board or multidisciplinary case conference, clarity must be reached regarding the specific choice of therapeutic options (and their potential risks and benefits, overall goals, and alternatives) when considered in the context of the wishes of the patient and family. This is particularly true when the side effects of treatment are substantial (such as for the multidisciplinary management of non–small cell lung cancer or esophageal cancer). For cancers amenable to surgery, resection is often an initial alternative if the patient is a suitable candidate for anesthesia (Chapter 432) and is otherwise in acceptable condition in terms of concomitant or comorbid illnesses. Determination of the patient's performance score (see Table 179-1) is a simple means of assessing functional status. If life expectancy is limited or if the patient is not a good candidate for surgery, more limited approaches to palliative radiation or systemic therapy may be appropriate. There are now substantial data suggesting that the extent of surgery required for an optimal long-term outcome may be reduced for certain solid tumors through the use of presurgical neoadjuvant chemotherapy, often as part of an "organ-sparing" approach. Concomitant and/or sequential multimodality treatments also have the potential to produce long-term disease-free remissions. When the best therapeutic outcomes require the combined skills of surgical, radiation, and medical oncologists, the need for coordination of care among a variety of specialists becomes paramount, and the use of predefined treatment regimens critical.

Therapeutic Paradigm and Therapeutic Index

The therapeutic paradigm in oncology, although still directed at improving the multidisciplinary care model, has begun to change from one focused on delivering treatments at the "maximally tolerated dose" for normal tissues, to therapy that is personalized based on both the molecular characteristics of the patient's tumor as well as any individual germline features that could modify treatment tolerance (Chapter 181). Based on the rapid expansion of our understanding of somatic mutations in human malignancies, and the ability to produce therapeutic molecules that can target specific deficiencies in tumoral DNA repair, growth factor signaling, or energy balance, for example, the currently developing approach to cancer therapeutics involves the employment of predictive molecular markers to guide all modalities of cancer care for the benefit of unique cancer patients. Hence, the focus of oncologists today is on the elaboration of treatments that can be administered with a high therapeutic index, defined as the comparison of the amount of treatment that is effective to the amount that causes toxicity; in this era of personalized cancer medicine, the goal of cancer therapy is to minimize normal tissue toxicity while preserving quality of life by advancing therapies or procedures that are targeted only to specific molecular dependencies in tumors.

Surgical Therapy

Surgery is used to biopsy a suspected lesion, remove the primary tumor, bypass obstructions, provide palliation, and prevent cancers in patients at very high risk because of genetic predispositions or chronic inflammatory states. Surgical staging also establishes the extent of disease. For example, patients with ovarian cancer (Chapter 199) benefit from surgical "debulking" to remove all visible disease, leaving minimal residual tumor, a process that may enhance the effectiveness of systemic treatment. Placement of a venous access device at the time of surgery, if considered proactively, may eliminate the need for a second surgical anesthesia.

Surgery remains the most common method to cure localized cancers such as breast cancer (Chapter 198), colorectal cancer (Chapter 193), and lung cancer (Chapter 191), but it is limited by the location of the tumor, its extension, and distant metastases. Even if a tumor cannot be removed, surgical biopsy provides confirmation of the diagnosis and additional tissue for molecular analysis. Occasionally, an obstructing lesion can be bypassed to provide palliation.

In specific circumstances when the primary tumor has been controlled, removal of a single metastasis (metastasectomy) can result in long-term survival; an example is resection of a solitary liver metastasis found at the time of colectomy for colorectal cancer. A variety of surgical techniques, such as radiofrequency ablation or cryoablation, can also be used to treat hepatic metastases in carefully selected patients. Adjuvant chemotherapy is often given after surgery in this situation to treat microscopic metastases.

The careful application of reconstructive surgery after a disfiguring procedure is critical to long-term physical and emotional well-being. Examples include postmastectomy breast reconstruction (Chapter 198) and plastic surgical procedures to correct deformities following head and neck surgery (Chapter 190).

Radiation Therapy

Ionizing radiation (Chapter 20) can be delivered using beams of high-energy rays, known as *teletherapy*, via a linear accelerator; by brachytherapy, through the application of sealed radioactive implants, seeds, wires, or plaques; and intravenously by using radioisotopes, either directly or attached to antibodies or other targeting molecules. Radiation interacts with water molecules to induce free radical species, including hydroxyl radicals, which damage DNA, proteins, and lipid membranes, leading to cell death. Like chemotherapy, radiation therapy is most effective against rapidly dividing cells that are well oxygenated.

The utility of radiation therapy is limited by the inapparent extension of disease outside a local treatment field, by the location of tumors next to normal structures that must be preserved, and by the presence of distant metastases. Normal tissue tolerance, which varies across different organs and tissues, often prevents the use of radiation doses that could uniformly eradicate cancers. Radiation therapy is also limited by tumor hypoxia: large, bulky tumors are frequently relatively radioresistant, whereas well-oxygenated tumors can be more effectively treated at lower doses. In addition to acute radiation-related toxicities (Chapter 20), late effects of radiation therapy include second malignancies, such as breast cancers that may occur decades after administering thoracic radiation fields during curative treatment for Hodgkin disease.

Radiation therapy can be used as primary treatment, as part of multimodality therapy, in the adjuvant setting, and for palliation. As a single modality, radiation therapy can be curative for early-stage malignancies such as laryngeal cancer (Chapter 190), cervical cancer (Chapter 199), and prostate cancer

(Chapter 201). Breast-conserving surgery (Chapter 198) requires the use of radiation to treat the remaining breast. Partial irradiation techniques using three-dimensional planning with external beam radiation have recently been developed and used in selected patients with appropriately placed and sized breast cancers. For localized prostate cancer (Chapter 201), implanted radioactive seeds of gold or palladium offer an alternative to surgery or external beam radiation therapy in certain patients.

Newer techniques, such as intensity-modulated radiation therapy (IMRT), permit more exact tailoring of the dose to the target, thereby reducing damage to the surrounding normal tissues. Stereotactic radiation therapy or gamma knife techniques allow the treatment of primary or metastatic brain tumors (Chapter 189) measuring up to 3 cm with enhanced accuracy, minimizing damage to normal brain. Particle-based treatment with protons has expanded in use, particularly for limited-stage prostate cancer, based on the potential to deliver higher radiation doses locally. However, there are no randomized studies demonstrating its superiority over other approaches using computational techniques to improve the specificity of radiation delivery (such as IMRT); it is also used for some uveal melanomas, base-of-skull tumors, and a few pediatric malignancies.

Low- to moderate-dose palliative radiation is used to ameliorate symptomatic cancer when cure is no longer the goal. For instance, radiation therapy can improve symptoms from brain metastases (Chapter 189), relieve pain from bone lesions, relieve some obstructing lesions, and sometimes improve hemoptysis caused by lung cancer (Chapter 191) or bleeding from a gynecologic malignancy (Chapter 199). Bone-seeking radioisotopes of samarium, strontium, or radium may relieve pain from bone metastases in prostate cancer (Chapter 201) or breast cancer (Chapter 198).

Systemic Therapy

Cancer Pharmacology

Principles

The fundamental goal of cancer pharmacology is the development of treatments that can be matched to the intrinsic sensitivity of specific tumors, and that can be delivered in concentrations that affect the molecular target of interest with an acceptable therapeutic index. Three properties of all cancer therapeutic agents underlie their clinical utility: (1) pharmacogenetics of the drug (how the germline or somatic expression of genes alters normal tissue toxicity or antitumor efficacy); (2) action of the drug (pharmacodynamics, or what the drug does to the tumor/body); and (3) delivery of the drug (pharmacokinetics, or what the body does to the drug).

Pharmacogenetics

Pharmacogenetics, the study of inherited interindividual differences in drug disposition and effects, is important in cancer therapy because genetic polymorphisms in drug-metabolizing enzymes may be responsible for variations in efficacy and toxicity observed with many chemotherapeutic agents. Drugs potentially affected by polymorphisms identified to date include the thiopurines, 5-fluorouracil, irinotecan, taxanes, and the platinum agents. In patients who are heterozygous or homozygous for deficiencies in metabolizing enzymes, toxicity can be dramatically enhanced. Testing for pharmacogenetic variations that predict for altered normal tissue tolerance is currently available for thiopurines and irinotecan.

Molecular Targeting and Pharmacodynamics

Molecular diagnostic testing to predict antitumor activity for hormonal and anti-HER2 therapeutics has been part of routine oncologic practice for the past two decades, but the recent past has witnessed a remarkable increase in the use of molecular diagnostic testing in cancer drug development and a consequent increase in the number of molecularly targeted anticancer agents that are prescribed only after demonstration of specific molecular abnormalities (predictive of response) in primary or metastatic tumor tissues.[1,2] Examples of such diagnostic/drug pairs include: EGFR mutations in lung adenocarcinomas and erlotinib; ALK tyrosine kinase translocations in lung adenocarcinomas and crizotinib; and $BRAF^{V600E}$ mutations in melanoma and vemurafenib. Recent advances in the process of cancer drug discovery, furthermore, have focused on demonstrating engagement of presumed molecular targets of drug action early in the development process (e.g., evidence of enzyme inhibition, protein dephosphorylation, or DNA damage) as the first step toward implementation of a predictive molecular test for the drug in clinical practice.

Pharmacokinetics and Drug Delivery

Dose selection in oncology is a critical issue primarily because anticancer agents have among the smallest therapeutic ratios in all of medicine. When tumors are responsive to treatment, higher doses may or may not be more effective but are likely to be more toxic to normal tissues. The clearance of a drug from both the systemic circulation and, potentially, from specific physiologic compartments (CNS, thoracic, or peritoneal effusions) is the most important determinant of the dose chosen for a particular patient. It is a composite of all the routes and mechanisms by which the drug can be eliminated from the body and thus determines dose and dose adjustments in the face of changes in drug metabolism, transport, or altered organ function. Although it is important to develop a clear understanding of drug half-lives to establish initial dosing schedules, drug clearance will be the actual determinant of the dose of drug that can be safely administered. Another aspect of drug delivery is the use of body surface area dosing versus flat dosing (using fixed amounts of a drug) in oncologic practice. Although dosing based on body surface area has a long history in oncology (unlike other areas of internal medicine), very little data support this approach; for most agents in common use, empirically determined pharmacokinetic variability from patient to patient is far greater than that which could reasonably be ameliorated by dosing based on weight or surface area.

Routes of Drug Dosing

Prior to the approval of imatinib (administered orally) for the treatment of chronic myelogenous leukemia in 2001, most anticancer agents were developed for parenteral use. However, although intravenous administration remains important, most new anticancer agents are developed for oral use; 7 of the 11 cancer drugs approved by the U.S. Food and Drug Administration (FDA) in 2012 are given by mouth. This change in the route of administration for new drugs brings to the forefront many drug delivery issues that had previously not been routinely considered in the oncology clinic, such as: treatment adherence (does the patient take his/her medication), variability in absorption due to food effects, emesis prior to drug absorption, first-pass metabolism in the liver or intestine, and difficulties in swallowing. For example, whether the orally administered anti-HER1/2 drug lapatinib is taken with food or on an empty stomach can alter its absorption by as much as 10-fold.

In addition to intravenous or oral dosing, anticancer drugs may be used: for intrathecal delivery (to overcome the blood-brain barrier) for the treatment or prevention of meningeal spread of leukemia; for intravesicle therapy of early-stage bladder cancer; for intra-arterial delivery of fluoropyrimidines or other agents to treat hepatic metastases from colon cancer or hepatocellular carcinoma; and for intraperitoneal administration of drugs such as platinum agents for ovarian cancer therapy, which provides a survival advantage for these agents compared with intravenous treatment. In almost all cases, administration of anticancer agents by other than the oral or intravenous routes requires intimate involvement of an oncologic specialist.

Clinical Trials

Clinical trials in oncology are defined by the steps in which new diagnostic approaches, therapeutic agents, or procedures are tested to determine whether they will become part of the standard care for cancer patients. During interventional (rather than observational) clinical trials, specific treatments or procedures are assigned to participants, and the effects of the interventions are measured. For trials of new oncologic drugs, the FDA recognizes several states in the drug development process. Exploratory studies of new drugs examined for the first time in humans (phase 0 trials) may be conducted in on a limited number of patients to define a drug's mechanism of action or biodistribution and to guide subsequent dosing in larger studies. Phase 1 trials define, using a variety of dose-escalation strategies, the maximum dose that can be safely administered to humans and the most appropriate schedule for drug administration, as well as the pharmacokinetic (and more recently pharmacodynamics) profile of the drug. Phase 2 trials often enroll 50 to 150 subjects and focus on testing the drug to determine its effectiveness and side-effect profile in a specific malignancy (or for a specific cancer-related molecular abnormality). If a drug demonstrates anticancer activity in a phase 2 study, phase 3 trials are performed to compare the usefulness of an investigational treatment to a control group receiving the standard of care; patients in most phase 3 studies are randomly assigned to the new or standard treatment to avoid a biased assessment of the results of the study. Finally, after approval of a new drug by the FDA, usually based on the results of phase 3 studies, phase 4 trials may be performed to collect safety information on larger patient populations to define the prevalence of side effects that may be rare but serious.

Drug Interactions

Many drug interactions affect the toxicity profile of anticancer agents, for the most part because concomitant administration of a second drug changes the clearance of the cancer therapeutic, potentially enhancing side effects if the clearance is decreased or diminishing efficacy because of reduced drug exposure. Most often these interactions occur because one drug affects the metabolism of the other by inhibiting or enhancing the activity of cytochrome P-450 isoforms (such as CYP3A) in the liver. This is particularly true for agents administered orally, such as imatinib, crizotinib, enzalutamide, pazopanib, and lapatinib. Concomitant drug-related as well as pharmacogenetic variations in proteins that affect the transport of anticancer agents across tumor and normal cell membranes, such as efflux pumps, also affect the sensitivity or resistance of many classes of cancer drugs. For example, certain drugs that are well-known inducers of hepatic metabolism (phenytoin and rifampin) also induce the expression of drug transport proteins.

Combination Therapy

Virtually all curative chemotherapy regimens developed for hematologic malignancies or solid tumors use combinations of active agents. Combination

chemotherapy is usually superior to the use of single agents in adjuvant and neoadjuvant therapy as well. The improved results achieved by combination chemotherapy can be explained in several ways. Mechanisms of resistance to any particular single agent are almost always present in the tumor genome at diagnosis, even in clinically responsive tumors.[3,4] Tumors that are initially "sensitive" to systemic therapy rapidly acquire resistance to single agents, either as a result of selection of a preexisting clone of resistant tumor cells or because of a variety of potential acquired molecular changes (e.g., increased drug efflux, enhanced DNA repair, insensitivity to apoptosis) leading to clinical drug resistance. Combination therapy may address these phenomena by providing a broader range of mechanisms of drug action against initially resistant tumor cells, preventing or slowing the selection of resistant clones.

The development of combination systemic therapy regimens follows a set of principles (Table 179-2). For standard cytotoxic agents, each drug in the combination must be active against the tumor, and all drugs must be given at an optimal dose and on an appropriate schedule. The drugs should have different mechanisms of antitumor activity as well as different toxicity profiles, and the drugs should be given at consistent intervals for the shortest possible treatment time. The use of molecularly targeted agents in combination requires that each agent engages its specific target and that dual target inhibition produces a complementary enhancement of tumor growth inhibition. Toxicities of targeted agents should be moderate to allow prolonged administration and maximum target inhibition.

Therapeutic Settings

Systemic therapy is used in a variety of settings with or without and before, during, or after surgery and radiation therapy (Table 179-3). Considerable experimental evidence suggests that cancers are most sensitive to chemotherapy during early stages of growth because of higher growth fractions and shorter cell cycle times. Thus, a given dose of a cytotoxic drug may exert a greater therapeutic effect against a rapidly growing tumor than against a larger quiescent tumor.

Neoadjuvant therapy, also called primary or induction systemic therapy, is used before surgery or radiation therapy to decrease the size of locally advanced cancers, thereby permitting a more complete surgical resection or eradicating undetectable metastases. It also affords an opportunity to evaluate the effectiveness of treatment by histologic and molecular analysis of resected tissue. This approach is most often used for locally advanced breast cancer (Chapter 198).

TABLE 179-2 PRINCIPLES OF COMBINATION THERAPY

CYTOTOXIC AGENTS	MOLECULARLY TARGETED DRUGS
Drugs are each active against the tumor	Agent has therapeutic effect on molecular pathway in vivo
Drugs have different mechanisms of action	Agents have complementary effects on the same target or other targets in the same pathway or pathways that cross-talk to control tumor growth
Drugs have different clinical toxicities to allow full doses of each to be administered	Toxicities do not overlap with cytotoxics and are moderate to low to allow prolonged administration. Consider physiologic consequences of target engagement in relation to toxicity profile
Intermittent intensive therapy preferred to continuous treatment for cytoreduction and to reduce immunosuppression	Schedule chosen to maximize target inhibition

TABLE 179-3 COMMON EXAMPLES OF THERAPEUTIC SETTINGS

ADJUVANT THERAPY	NEOADJUVANT THERAPY	ORGAN-SPARING THERAPY	COMBINATION CHEMOTHERAPY
Stage I and II breast cancer Stage III colorectal cancer Stage II lung cancer	Stage III breast cancer	Anal cancer Laryngeal cancer Esophageal cancer	Metastatic solid tumors* Hematologic malignancies

*Usually palliative.

Organ-sparing therapy is another use of chemotherapy, radiation therapy, or both, to salvage organs that would have been surgically removed if cure were the intended result. This technique is often effective in patients with cancers of the larynx (Chapter 190), esophagus (Chapter 192), and anus (Chapter 193).

Adjuvant chemotherapy is used in patients whose primary tumor and all evidence of cancer (e.g., regional lymph nodes) have been surgically removed or treated definitively with radiation, but in whom the risk of recurrence is high because of involved lymph nodes or certain morphologic or biologic characteristics of the cancer. Common examples include cancers of the breast (Chapter 198) and colon (Chapter 193). The typical end points of chemotherapy, such as shrinkage of measurable tumor on serial radiographic studies, are not available in this situation; instead, relapse-free survival and overall survival are the principal measures of treatment effect. For an individual patient receiving adjuvant therapy, there is no way to determine whether such therapy is beneficial; hence, decisions are generally based on evidence from clinical trials.

Assessment of Response

Assessment of the response to therapy (usually performed using RECIST [Response Evaluation Criteria in Solid Tumors]) depends largely on tumor size, determined by either direct measurement or diagnostic imaging studies, using predefined categories. The categories of response are "complete response," with total absence of tumor and correction of tumor-associated changes measured twice at least 4 weeks apart; "partial response," defined as 30% or greater reduction in the sum of the longest diameters of up to 5 target lesions per organ confirmed by repeat measurement 4 weeks later; "progressive disease," characterized by either 20% or greater increase over the smallest sum of the longest diameters of target lesions or the development of new tumors; and "stable disease," defined as meeting criteria for neither partial response nor progressive disease. Leukemias are assessed by bone marrow biopsies and molecular diagnostic tests for residual disease, and multiple myeloma is typically assessed by the measurement of monoclonal proteins, peripheral blood counts, and percentages of malignant plasma cells in bone marrow samples, as well as imaging of bone lesions. Accurate assessment of response following systemic therapy is essential because of the tight relationship between the degree of response and the duration of disease control.

Classes of Therapeutic Agents

Cytotoxic Agents, Targeted Small Molecules, and Antibodies

The pharmacologic properties of the most commonly used cytotoxic and molecularly targeted chemotherapeutic agents approved by the FDA are described in Table 179-4, as well as their most common therapeutic indications. In all cases, current information from the manufacturer should be sought before therapy is initiated.

Administration of chemotherapy is best done by specifically trained individuals because of the dual acute risks of hypersensitivity reactions and extravasation. No doses or schedules are suggested in Table 179-4 because these agents are often used in combination, and the doses of each drug may need to be reduced when the compounds are combined. The treatment of special populations, including patients with significant obesity, during pregnancy, the elderly, and those with abnormal end-organ function, are addressed later in this chapter.

Unless otherwise specified, most cytotoxic chemotherapeutic agents are capable of producing some degree of nausea and vomiting, myelosuppression, alopecia, mucositis, and/or diarrhea after treatment; many agents are also teratogenic, mutagenic, and carcinogenic. Drugs used routinely to prevent agent-specific toxicities are also included in Table 179-4.

Over the past decade, several dozen small molecule anticancer agents with more precisely targeted mechanisms of action have become a standard part of oncologic practice (see Table 179-4). Although the molecular dependencies within tumor cells against which these drugs are targeted are broad in scope, including tyrosine kinase growth factors or their receptors, they are functionally much more specific than prior generations of systemic cancer therapies. This allows for a better appreciation of the clinical situations wherein certain drugs might be beneficial, as well as the possibility of developing agents for use in specific tumors based on their genetic susceptibilities. It is also noteworthy that the toxicity profiles of molecular targeted agents most often reflect alterations produced in biochemical pathways that control normal organ function, rather than a general pattern of toxicity consistent with injury to rapidly growing tissues, such as the bone marrow or gastrointestinal tract. Current research aims to clarify specific mutational profiles in the clinic that can be used to prospectively select patients for therapy (Chapter 181).

The development of monoclonal antibodies directed against antigens found on cancer cells represents an additional approach to molecular targeting of systemic therapy. Examples include cetuximab (targeting the epidermal growth factor receptor), rituximab (targeting the B-cell CD20 surface antigen), and trastuzumab (which blocks HER2). These monoclonal antibodies can be used alone, or labeled with a radioactive molecule, or conjugated to another

Text continued on p. 1218

TABLE 179-4 U.S. FDA-APPROVED DRUGS COMMONLY USED FOR THE SYSTEMIC TREATMENT OF CANCER

DRUG NAME	DRUG CLASS AND MECHANISM OF ACTION	PHARMACOKINETICS AND METABOLISM	TOXICITY	INDICATIONS
ALKYLATING AGENTS				
Bendamustine (Treanda)	Alkylating agent; bifunctional, with both alkylating and purine-like antimetabolite action	Biotransformation in liver; decrease dose for hematologic toxicity	Nausea, vomiting, and bone marrow suppression	CLL and B-cell non-Hodgkin lymphoma
Carboplatin (Paraplatin)	Platinum coordination compound; produces intrastrand and interstrand DNA cross-links leading to introduction of DNA breaks during replication	Rapidly cleared largely unchanged by kidney; patients with decreased CrCl experience greater thrombocytopenia	Thrombocytopenia; nephrotoxicity substantially less than cisplatin	Ovarian cancer, testicular cancer, lung cancer, head and neck cancer, breast cancer
Chlorambucil (Leukeran)	Bifunctional alkyl forms interstrand DNA cross-links with resultant inactivation of DNA; cell cycle nonspecific acting agent	Highly orally bioavailable; hepatic biotransformation	Dose-limiting myelosuppression; mucosal toxicity mild	CLL, Waldenström macroglobulinemia, non-Hodgkin lymphomas
Cisplatin (Platinol)	Platinum coordination compound; produces interstrand and intrastrand DNA cross-links leading to DNA breaks; cell cycle nonspecific	Rapid distribution and DNA binding to tissues; over 90% of drug is protein bound	Nephrotoxicity is dose-limiting; significant nausea and vomiting require expert management; ototoxicity; peripheral neuropathy; hypomagnesemia and potassium wasting	Testicular cancer, other germ cell tumors, ovarian cancer, bladder cancer, lung cancer, sarcomas, cervical cancer, endometrial cancer, gastric cancer, breast cancer, head and neck cancer
Cyclophosphamide (Cytoxan, Neosar)	Alkylating agent; cross links DNA, decreasing macromolecular synthesis; immunosuppressive	Hepatic biotransformation of parental pro-drug to alkylating species; metabolites excreted in urine	Bone marrow suppression moderate at standard doses; alopecia; hemorrhagic cystitis, SIADH, and pulmonary fibrosis infrequent	Breast cancer, Hodgkin and non-Hodgkin lymphomas, leukemias, neuroblastoma, retinoblastoma, other sarcomas, osteogenic sarcoma, Wilms tumor
Ifosfamide (Ifex)	Alkylating agent; alkylated metabolites interact with DNA; cell cycle nonspecific	Hepatic biotransformation; renal elimination	Myelosuppression; hemorrhagic cystitis (requires co-administration of the uroprotector mesna); nephrotoxicity; CNS toxicity (lethargy, stupor)	Germ cell tumors, sarcomas, non-Hodgkin lymphomas
Melphalan (Alkeran)	Alkylating agent; forms interstrand, intrastrand, or DNA protein cross-links; cell cycle nonspecific	Unpredictable absorption by GI tract; highly protein bound; hydrolyzed in plasma; partially eliminated by kidney	Myelosuppression; vomiting when used at high dose; associated with secondary leukemias when used chronically	Multiple myeloma, rhabdomyosarcoma, bone marrow ablation for stem cell transplantation
Oxaliplatin (Eloxatin)	Platinum coordination compound; produces interstrand DNA cross-links; not identical to other platins	Renal elimination following active metabolism; no excretion by liver and safe to use in face of liver dysfunction	Cumulative sensory neuropathy is dose-limiting and worse in the cold; fatigue and nausea common	Colorectal cancer
Temozolomide (Temodar)	Nonclassic alkylating agent that is affected by DNA methylation status	Oral bioavailability good; penetrates blood-brain barrier	Myelosuppression may be cumulative; treatable nausea; fatigue	Melanoma, brain tumors
ANTIMETABOLITES				
5-Azacitidine (Vidaza)	Antimetabolite; pyrimidine nucleoside analogue of cytidine; hypomethylates DNA, producing gene activation; directly incorporated into DNA, with cytotoxicity at higher doses	Hepatic metabolism; renal excretion	Dose-limiting myelosuppression; transient liver function abnormalities; nausea, vomiting, abdominal pain	Myelodysplastic syndrome
Capecitabine (Xeloda)	Pyrimidine antimetabolite; pro-drug form of 5-fluorouracil; inhibits DNA and RNA synthesis	Well absorbed after oral administration; metabolized to 5-fluorouracil in liver and in tumor	Myelosuppression, hand-foot syndrome, diarrhea, stomatitis, fatigue	Breast cancer, colorectal cancer
Cladribine (Leustatin), 2-chloro-2-deoxy-D-adenosine	Purine nucleoside antimetabolite; inhibits DNA synthesis and repair	Excreted primarily in urine as unchanged parent drug; excellent oral bioavailability	Bone marrow suppression, fever	Hairy cell leukemia, CLL, non-Hodgkin lymphoma
Clofarabine (Clofar)	Purine nucleoside antimetabolite; inhibits DNA synthesis and repair; activates apoptosis	Excreted in urine	Bone marrow suppression, hepatotoxicity, capillary leak syndrome	Relapsed acute lymphoblastic leukemia
Cytarabine* (Cytosar-U, Tarabine PFS)	Antimetabolite activated to cytarabine triphosphate in tissues; inhibits DNA synthesis; cell cycle specific, S phase	Deaminated in blood and tissues, with short terminal half-life	Bone marrow suppression, stomatitis, pancreatitis; with high doses, cerebral dysfunction, GI damage, hepatotoxicity, pulmonary edema, corneal damage, Ara-C syndrome	Acute myelocytic leukemia

Decitabine (Dacogen)	Antimetabolite that inhibits DNA methyltransferase, producing hypomethylation and gene activation	Deaminated in liver, blood, and GI tract; short half-life	Bone marrow suppression, nausea and vomiting, fatigue, abnormal liver function and blood sugar	Myelodysplastic syndrome
Fludarabine phosphate (Fludara)	Purine nucleotide antimetabolite; 2-fluoro-ara-ATP inhibits DNA synthesis by inhibition of ribonucleotide reductase and DNA polymerases	After IV dosing, 2-fluoro-ara-A widely taken up by tissues; drug and metabolites excreted by kidney	Neurotoxicity, including somnolence and demyelinating lesions, dose-limiting; myelosuppression (lymphopenia)	CLL; also used for low-grade non-Hodgkin lymphomas
Fluorouracil (5-FU, Adrucil)	Pyrimidine antimetabolite; inhibitor of thymidylate synthase; also alters RNA synthesis	Primarily metabolized by dyhydropyrimidine dehydrogenase in liver; remainder metabolized in tumor and other tissues to active species; renal excretion of remaining drug	Mucositis and diarrhea, myelosuppression; hand-foot syndrome when used as an IV infusion; rare cerebellar ataxia or cardiac ischemia	Colorectal cancer, other GI cancers, breast cancer, head and neck cancer
Gemcitabine (Gemzar)	Nucleoside analogue antimetabolite that inhibits ribonucleotide reductase and is incorporated into DNA following intracellular metabolism, leading to DNA synthesis inhibition	Metabolized by deamination in a variety of tissues and excreted both as parent drug and metabolites by the kidneys; elevated bilirubin requires dose reduction; increased creatinine enhances drug toxicity	Myelosuppression, nausea and vomiting, elevated transaminases, fever	Pancreatic cancer, breast cancer, non–small cell lung cancer, bladder cancer, ovarian cancer
Methotrexate (Folex, Mexate)	Folic acid analogue antimetabolite; inhibition of dihydrofolate reductase blocks nucleotide synthesis	Mainly renal excretion with minimal hepatic metabolism; strict attention to renal function required for dosing; third-space accumulation, including pleural effusions and ascites, with prolonged release and associated toxicity	Myelosuppression is dose-limiting; stomatitis and diarrhea common, especially if delayed excretion; renal toxicity can also be severe if drug elimination impaired	Acute leukemias, especially in children; non-Hodgkin lymphoma, breast cancer, head and neck cancer, sarcomas
Pemetrexed (Alimta)	Folic acid analogue antimetabolite; inhibits multiple enzymes in folate pathway, leading to altered DNA and RNA synthesis	Excreted as unchanged drug by kidney	Myelosuppression, nausea, diarrhea, rash, fatigue; must be given with folic acid and vitamin B_{12} to reduce toxicity	Mesothelioma, non–small cell lung cancer
Pralatrexate (Fotolyn)	Folic acid analogue antimetabolite	Renal excretion	Myelosuppression, mucositis, pyrexia; must be given with folic acid and vitamin B_{12} to reduce toxicity	Relapsed peripheral T-cell lymphoma
DIFFERENTIATING AGENTS				
All-*trans*-retinoic acid (ATRA)	Retinoid; induces cellular differentiation and/or apoptosis	Conjugated to glucuronic acid, with subsequent biliary excretion and enterohepatic circulation	Mucocutaneous, ocular, musculoskeletal, neurologic, hepatic toxicity; hyperlipidemia	Acute promyelocytic leukemia
Arsenic trioxide (Trisenox)	Arsenical differentiating agent	Hepatic metabolism; excreted in urine	Prolonged QT interval; acute promyelocytic leukemia differentiation syndrome (leukocytosis, fever, dyspnea, chest pain, hypoxia) that can be treated with corticosteroids; peripheral neuropathy	Acute promyelocytic leukemia
ENZYMES				
L-Asparaginase (Elspar)	Enzyme that hydrolyzes 1-asparagine to aspartic acid and ammonia, resulting in cellular deficiency of 1-asparagine, critical for tumor cells that lack asparagine synthetase; interferes with protein, DNA, and RNA synthesis	Metabolized in the vasculature by proteolysis	Hypersensitivity reactions; inhibitory effects on protein synthesis, with resultant decreases in hepatic synthesis of coagulation factors, pancreatitis, hyperglycemia, CNS depression, hepatotoxicity, transient renal dysfunction	Acute lymphoblastic leukemia
DNA-ACTIVE DRUGS WITH PLEIOTROPIC MECHANISMS OF ACTION				
Bleomycin (Blenoxane)	Antitumor antibiotic; produces free radical–related DNA strand breaks	Metabolized partially by intracellular aminopeptidases; renal excretion	Dose-related pulmonary fibrosis, fever, hypersensitivity reactions, skin toxicity, including Raynaud phenomenon	Testicular cancer and other germ cell tumors, Hodgkin and non- Hodgkin lymphomas; sclerosing agent for pleural effusions
Carmustine (BiCNU, BCNU)	Nitrosourea-class alkylating agent; alkylates DNA and may affect carbamoylation of amino acids	Taken up by CNS; metabolized in liver and excreted by kidney	Myelosuppression may be slow in onset and cumulative; severe nausea and vomiting, renal toxicity, interstitial lung disease	Glioblastoma
Daunorubicin (Cerubidine)	Anthracycline antibiotic; pleiotropic effects, including free radical formation, inhibition of topoisomerase II, altered mitochondrial metabolism, activation of pro-apoptotic signal transduction	Hepatic metabolism with biliary excretion; small amount of renal excretion leads to "orange-red" urine	Myelosuppression and mucositis are dose-limiting; alopecia, cumulative dose-related cardiotoxicity, extravasation injury	Acute myelocytic leukemia, acute lymphoblastic leukemia

TABLE 179-4 U.S. FDA-APPROVED DRUGS COMMONLY USED FOR THE SYSTEMIC TREATMENT OF CANCER—cont'd

DRUG NAME	DRUG CLASS AND MECHANISM OF ACTION	PHARMACOKINETICS AND METABOLISM	TOXICITY	INDICATIONS
Doxorubicin (Adriamycin, Rubex)	Anthracycline antibiotic; pleiotropic effects, including free radical formation, inhibition of topoisomerase II and DNA strand breaks, altered mitochondrial metabolism, activation of pro-apoptotic signal transduction	Hepatic metabolism to both active and inactive species, with 50% biliary excretion; "orange-red" urine	Myelosuppression and mucosal injury are dose-limiting; alopecia, cumulative dose-related cardiotoxicity (cardiomyopathy), nausea and vomiting, severe extravasation injury risk, secondary AML	Acute myelocytic leukemia, acute lymphoblastic leukemia, breast cancer, small cell lung cancer, Hodgkin and non-Hodgkin lymphomas, sarcomas, endometrial cancer, Wilms tumor, neuroblastoma
Doxorubicin liposomal (Doxil)	Anthracycline antibiotic with pleiotropic mechanisms of action	Liver	Myelosuppression, dose-related cardiotoxicity, extravasation injury, hand-foot syndrome	Ovarian cancer, breast cancer, Kaposi sarcoma
Epirubicin (Ellence)	Anthracycline antibiotic with pleiotropic mechanisms of action similar to others in class	Liver	Similar to doxorubicin; modestly less cardiotoxic; extravasation	Breast cancer adjuvant therapy
Etoposide (VP-16, VePesid)	Epipodophyllotoxin plant alkaloid; inhibits topoisomerase II, leading to decreased DNA synthesis	Extensively protein bound; hepatic metabolism; excreted as metabolites in bile and unchanged in urine	Dose-limiting myelosuppression, nausea and vomiting, stomatitis when used at high dose, secondary AML	Small cell lung cancer, germ cell tumors, lymphomas; high-dose therapy conditioning regimens
Idarubicin (Idamycin)	Anthracycline antibiotic with pleiotropic mechanisms of action similar to others in class	Hepatic metabolism with biliary excretion	Similar to doxorubicin; modestly less cardiotoxic; extravasation	Acute myelocytic leukemia
Irinotecan (Camptosar)	Semisynthetic camptothecin analogue that poisons DNA topoisomerase I, preventing religation of replication-related single-strand breaks	Partially inactivated in plasma; metabolized by esterases to active species SN-38, which is excreted into bile; patients with elevated bilirubin have increased toxicity	Myelosuppression; early and late diarrhea may be severe; flushing, alopecia	Colorectal cancer
Topotecan (Hycamtin)	Semisynthetic camptothecin analogue that inhibits DNA topoisomerase I, inhibiting transcription	Excreted primarily unchanged in urine	Severe myelosuppression is dose-limiting (both granulocytes and platelets); mild nausea, fatigue, diarrhea	Relapsed ovarian and small cell lung cancer
INHIBITORS OF MITOTIC FUNCTION				
Docetaxel (Taxotere)	Mitotic spindle poison that stabilizes tubulin polymers, leading to mitotic tumor cell death	Biliary excretion	Myelosuppression and alopecia, hypersensitivity reactions, fluid retention syndrome, peripheral sensorimotor neuropathy	Breast cancer, non–small cell lung cancer, ovarian cancer, gastric cancer, head and neck cancer
Eribulin mesylate (Halaven)	Semisynthetic natural product; microtubule inhibitor	Excreted primarily as unchanged parent compound in feces	Neutropenia, peripheral neuropathy, QT prolongation, fatigue, nausea	Metastatic breast cancer
Ixabepilone (Ixempra)	Microtubule inhibitor; analogue of epothilone B that binds to β-tubulin, suppressing function of microtubules and causing mitotic cell death	Metabolized by liver and excreted in feces; dose reduction required in face of liver dysfunction	Cumulative peripheral neuropathy, neutropenia, hypersensitivity reactions, fatigue, myalgias, stomatitis	Breast cancer
Paclitaxel (Taxol)	Taxane natural product; mitotic spindle poison that inhibits tubulin depolymerization	Hepatic metabolism, biliary excretion	Myelosuppression, mucositis, hypersensitivity reactions, cumulative peripheral neuromyopathy with arthralgias, cardiovascular toxicity with hypotension, arrhythmias	Non–small cell lung cancer, ovarian cancer, breast cancer, esophageal cancer, gastric cancer, head and neck cancer
Paclitaxel protein-bound particles (Abraxane)	Albumin nanoparticle-bound form of paclitaxel; same mechanism of action	Hepatic metabolism, biliary excretion	Hypersensitivity reactions, myelosuppression, neuropathy, arthralgias/myalgias, cardiotoxicity	Metastatic breast cancer, pancreatic cancer, non–small cell lung cancer
Vincristine (Oncovin)	Vinca alkaloid natural product; inhibits tubulin polymerization, arresting cells in metaphase	Hepatic metabolism, biliary excretion	Extravasation injury, dose-limiting peripheral neurotoxicity, constipation	Acute lymphocytic leukemia, neuroblastoma, Wilms tumor, Hodgkin and non-Hodgkin lymphomas, rhabdomyosarcoma
Vinorelbine (Navelbine)	Semisynthetic vinca alkaloid; inhibits tubulin polymerization during mitosis	Hepatic metabolism, biliary excretion	Myelosuppression, extravasation, milder neurotoxicity than other vincas	Non–small cell lung cancer, breast cancer

HORMONAL AGENTS				
Abiraterone acetate (Zytiga)	Inhibitor of androgen biosynthesis	Parent drug and metabolites excreted in stool; systemic clearance decreased in patients with liver dysfunction	Joint swelling, edema, hepatotoxicity	Castration-resistant metastatic prostate cancer
Anastrozole (Arimidex)	Nonsteroidal aromatase inhibitor; inhibits conversion of adrenal androgens to estrogens	Metabolized in liver, excreted into bile and urine	Hot flashes, headache, arthralgias	Adjuvant and metastatic breast cancer in postmenopausal women
Bicalutamide (Casodex)	Nonsteroidal antiandrogen that binds to prostatic androgen receptor	Hepatic metabolism	Worsening bone pain, hot flashes, gynecomastia	Prostate cancer (usually in conjunction with LHRH antagonist)
Degarelix (Firmagon)	GnRH receptor antagonist; binds to GnRH receptors in pituitary, decreasing gonadotropin release	Biliary excretion of parent and metabolites; lesser renal excretion of unchanged parent drug	Injection site reactions, hot flashes, weight gain, increased LFTs, QT prolongation	Prostate cancer
Enzalutamide (Xtandi)	Nonsteroidal antiandrogen; inhibits androgen receptor-mediated signal transduction	Hepatic metabolism, eliminated primarily in urine	Fatigue, arthralgias, dizziness, bone pain	Metastatic castration-resistant prostate cancer
Exemestane (Aromasin)	Steroidal aromatase inhibitor; binds and irreversibly inhibits aromatase, inhibits synthesis of estrogens by preventing conversion of adrenal androgens to estrogens	Metabolized in liver	Hot flashes, fatigue, arthralgias	Metastatic breast cancer in postmenopausal women
Flutamide (Eulexin)	Nonsteroidal antiandrogen; inhibition of nuclear binding of androgen in target tissues; its interference with testosterone at cellular level complements "medical castration" produced by LHRH analogues	Metabolized to active and inactive metabolites in liver	Worsening bone pain, hot flashes, gynecomastia, impotence	Prostate cancer (usually in conjunction with LHRH antagonist)
Fulvestrant (Faslodex)	Estrogen receptor antagonist; binds to estrogen receptor, causing degradation of estrogen receptor protein	Metabolized by liver, excreted in bile	Hot flashes, nausea, peripheral edema and weight gain, fatigue, arthralgias	Recurrent breast cancer in postmenopausal women
Goserelin (Zoladex)	Synthetic decapeptide analogue of LHRH; suppresses pituitary gonadotropins, with fall of serum testosterone into castrate range	Slowly released from depot injection site; not extensively metabolized; urinary excretion	Worsening bone pain, hot flashes, impotence, gynecomastia, breakthrough vaginal bleeding	Prostate cancer, breast cancer
Letrozole (Femara)	Nonsteroidal competitive inhibitor of aromatase; inhibits estrogen synthesis by blocking conversion of adrenal androgens to estrogens	Metabolized in liver, excreted by kidney	Hot flashes, fatigue, arthralgias	Adjuvant and metastatic breast cancer in postmenopausal women
Leuprolide (Lupron, Lupron Depot)	Synthetic LHRH analogue; suppresses secretion of GnRH, with resultant fall in testosterone secretion, producing "medical castration"	Metabolized to inactive peptide fragments; minor renal excretion	Increased bone pain, hot flashes, gynecomastia, lethargy, thromboembolic phenomena	Prostate cancer, breast cancer
Octreotide (Sandostatin)	Synthetic octapeptide analogue of somatostatin; suppresses secretion of serotonin and GI peptides; blocks carcinoid flush, decreases serum 5-HIAA, and controls other symptoms associated with carcinoid syndrome	Hepatic metabolism; renal excretion following hydrolysis in plasma	Hyper/hypoglycemia, hepatic dysfunction, diarrhea	Palliative treatment of carcinoid tumors and vasoactive intestinal peptide tumors (VIPomas)
Tamoxifen (Nolvadex)	Nonsteroidal antiestrogen; competes with estradiol for estrogen receptor protein; also has non–estrogen receptor–dependent effects on tumor cells	Metabolized in liver but not excreted in bile or urine	Hot flashes, nausea/vomiting, vaginal bleeding or discharge, endometrial hyperplasia, thrombophlebitis, hypercalcemia, visual disturbances	Adjuvant and metastatic estrogen receptor–positive breast cancer; also approved for chemoprevention of breast cancer in high-risk individuals

TABLE 179-4 U.S. FDA-APPROVED DRUGS COMMONLY USED FOR THE SYSTEMIC TREATMENT OF CANCER—cont'd

DRUG NAME	DRUG CLASS AND MECHANISM OF ACTION	PHARMACOKINETICS AND METABOLISM	TOXICITY	INDICATIONS
BIOLOGIC AND IMMUNOLOGIC MODIFIERS				
Aldesleukin (Human Recombinant IL-2, Proleukin)	Cytokine that supports T-cell proliferation, augments natural killer cell cytotoxicity, induces lymphokine-activated killer (LAK) cell development, and participates in activation of monocytes and B cells	Catabolized by proteolysis in many tissues; minimal renal or biliary excretion	Toxicities associated with continuous infusion (and to a lesser extent with bolus dosing) include: capillary leak syndrome, fever and chills, hypotension, edema, arrhythmias, nephrotoxicity, pulmonary edema, abnormal liver function, endocrinopathies, dermatologic complications, CNS toxicity, myelosuppression, sepsis	Renal cancer, melanoma
Erythropoietin (Aranesp, Epogen, Procrit)[†]	Hematopoietic growth factor; stimulates division and differentiation of committed erythroid progenitors in bone marrow	Proteolytically degraded in vasculature, with minimal excretion of intact peptide	Increased risk of thrombosis, stroke, myocardial infarction; headache, hypertension, and possible seizures; allergic reactions; can produce iron deficiency with prolonged use—concomitant iron dosing may enhance efficacy	Correction of anemia due to chronic renal failure or HIV-related infection, and symptomatic chemotherapy-induced anemia (however, risk of tumor progression limits use for patients being treated with chemotherapy for cure)
Filgrastim (G-CSF, Neupogen)	Hematopoietic growth factor; binds to specific cell surface receptors on progenitor cells to stimulate proliferation and differentiation of neutrophils	Elimination by proteolysis in vasculature and by the kidney	Pain at site of subcutaneous injection, allergic reactions, bone pain, low-grade fever, myalgia, arthralgia	Decreases incidence of infection after myelosuppressive chemotherapy; enhances myeloid engraftment after BMT; enhances peripheral progenitor cell yield prior to BMT
Interferon-α (Intron-A, Roferon)[‡]	Interferon antiviral immunostimulant with both antiproliferative and immunomodulatory properties	Catabolized in renal tubules	Fever and flulike symptoms, fatigue, myelosuppression, cardiotoxicity, depression, neurotoxicity	Hairy cell leukemia, Kaposi sarcoma
Ipilimumab (Yervoy)	CTLA-4-blocking monoclonal antibody that enhances the anticancer effect of activated T cells	No clear pharmacokinetic effects of altered renal or hepatic function	Fatigue, diarrhea and immune-mediated colitis, hepatitis, rash, pruritus, endocrinopathy	Metastatic melanoma
Sargramostim (GM-CSF, Leukine, Prokine)	Hematopoietic growth factor; binds to specific cell surface receptors to stimulate proliferation and differentiation of granulocytes and macrophages; not lineage specific	Metabolized in liver and kidney	Fever and capillary leak syndrome, pain at site of subcutaneous injection, allergic reactions, arthralgias, bone pain	Decreases incidence of infection after myelosuppressive chemotherapy; enhances myeloid engraftment after BMT; enhances peripheral progenitor cell yield
MOLECULARLY TARGETED AGENTS				
Aflibercept (Zaltrap)	VEGF inhibitor; decoy receptor that binds circulating VEGF thus inhibiting activation of VEGFR	Metabolized by proteolysis	Hemorrhage, GI fistulas and perforation, hypertension, altered wound healing, stomatitis, fatigue, myelosuppression	Colon cancer in combination with chemotherapy
Axitinib (Inlyta)	Multikinase inhibitor, including VEGFR1, 2, and 3; inhibits angiogenesis	Metabolized by hepatic P-450 enzymes; excreted in feces>urine; multiple interactions with CYP 3A4/5 inducers	Hypertension, hand-foot syndrome, fatigue, asthenia, stomatitis, hypothyroidism	Advanced renal cancer
Bevacizumab (Avastin)	Recombinant humanized monoclonal antibody against VEGF; inhibits angiogenesis by binding to VEGF, blocking receptor binding and subsequent stimulation of blood vessel growth	Prolonged (20-day) half-life after IV infusion	Fatigue, nausea, delayed wound healing, hypertension, proteinuria, thromboembolic phenomena and hemorrhage	Metastatic colorectal cancer, lung cancer, kidney cancer, breast cancer, ovarian cancer
Bortezomib (Velcade)	Reversible inhibitor of 26S proteasome; blocks breakdown of ubiquitinated intracellular proteins and disrupts ubiquitin-proteasome pathway	Undergoes oxidative metabolism (deboronation) as well as cytochrome P-450-dependent metabolism; dose reduction to 0.7 mg/m^2 required in face of moderate to severe liver dysfunction; adverse event profile unchanged in myeloma patients with CrCl < 50 mL/min	Myelosuppression, peripheral neuropathy, asthenia, diarrhea, hypotension	Multiple myeloma, non-Hodgkin lymphoma

Cetuximab (Erbitux)	Chimeric monoclonal antibody targeted against EGFR; blocks growth factor binding to EGFR, preventing cell signaling by tyrosine kinase phosphorylation	Half-life 5-7 days, with minimal renal or hepatic clearance	Hypersensitivity reactions (fever, dyspnea), acneiform rash, diarrhea, hypomagnesemia	Metastatic colorectal cancer (*K-Ras* wild-type); head and neck cancer in combination with radiation
Carbozantanib (Cometriq)	Multikinase inhibitor that targets RET, MET, VEGFR1-3, KIT, AXL, TIE-2, and FLT-3	Metabolized by hepatic CYP 3A4 and subject to interactions related to CYP3A4 substrates; excreted into urine and feces	Hypertension, GI perforation, diarrhea, stomatitis, hemorrhage, fatigue	Metastatic medullary thyroid cancer, prostate cancer, bone metastases
Crizotanib (Xalkori)	Tyrosine kinase inhibitor active against ALK, ROS, MET	Hepatic P-450-mediated metabolism; excreted in feces and urine	Hepatotoxicity, pneumonitis, QT prolongation, vision disturbance, nausea, fatigue	Non–small cell lung cancer positive for anaplastic lymphoma kinase (ALK) rearrangement
Dasatinib (Sprycel)	Multitargeted tyrosine kinase inhibitor; inhibits BCR-ABL, SRC, and multiple other kinases	Metabolized in liver by P-450 CYP3A4; multiple interactions with CYP3A4 inducers or inhibitors	Myelosuppression, fluid retention, diarrhea, rash, musculoskeletal pain	Resistant/refractory CML
Denosumab (Prolia)	IgG2 monoclonal antibody that targets RANKL, a ligand for the RANK receptor on osteoclasts, and decreases bone resorption	Prolonged elimination (>20 days) typical of monoclonal antibodies	Musculoskeletal and bone pain, arthralgias, abdominal pain, peripheral edema, hypersensitivity, hypocalcemia	Prevention of malignant bone fractures
Erlotinib (Tarceva)	Small molecule inhibitor of tyrosine kinase domain of EGFR	Hepatic metabolism with excretion in feces; CYP3A4 inducers and inhibitors may alter metabolism, and dose should be reduced by 50% in patients with hepatic dysfunction; renal dysfunction does not alter drug tolerability	Acneiform rash, fatigue, diarrhea, interstitial lung disease, weight gain, hepatic toxicity	Non–small cell lung cancer; pancreatic cancer in combination with gemcitabine
Everolimus (Affinitor)	mTOR inhibitor that inhibits signal transduction through PTEN/AKT pathway	Metabolized by CYP3A4 in liver	Pneumonitis, stomatitis, renal dysfunction, myelosuppression, hyperglycemia, altered lipids	Renal cell carcinoma; hormone receptor–positive breast cancer in combination with exemestane; neuroendocrine tumors
Ibrutinib (Imbruvica)	Oral inhibitor of Bruton tyrosine kinase; blocks signaling through the B-cell antigen receptor important for B-cell chemotaxis and adhesion	Metabolized in the liver by CYP3A (avoid use with strong CYP3A inhibitors such as ketoconazole or grapefruit); primarily eliminated unchanged in feces	Myelosuppression, renal dysfunction, bleeding	Mantle cell lymphoma
Imatinib (Gleevec)	Inhibits BCR-ABL tyrosine kinase	Hepatic metabolism, excreted in feces; inhibits CYP3A4 and CYP2D6; mild hepatic dysfunction requires dose reduction to 500 mg/day; mild to moderate renal dysfunction alters kinetics but does not affect drug tolerance	Myelosuppression, hypophosphatemia, fluid retention, nausea, fatigue, hemorrhage, myelosuppression, hepatotoxicity	CML; GIST
Lapatinib (Tykerb)	Inhibitor of both EGFR and HER2 tyrosine kinases	Metabolized by CYP3A4 in liver	Diarrhea, hand-foot syndrome, hepatotoxicity, decreased cardiac function	HER2-positive breast cancer in combination with capecitabine
Lenalidomide[§] (Revlimid)	Immunomodulatory and antiangiogenic agent, in part through downregulation of VEGF, TNF-α, and IL-6, and T-cell and NK cell activation	Renal excretion	Neutropenia, thrombocytopenia, diarrhea, pruritus, rash, fatigue, leg cramps	Myelodysplasia, multiple myeloma
Nilotinib (Tasigna)	Inhibits ATP site of BCR-ABL kinase	Metabolized in liver	Prolonged QT interval, rash, fatigue, myalgias, nausea, myelosuppression	Resistant/intolerant CML
Ofatumumab (Arzerra)	Monoclonal antibody against CD-20	Not known if dose requires modification for patients with renal impairment	Infusion reactions, tumor lysis syndrome, myelosuppression, hepatitis B reactivation, infection	Resistant CLL
Pazopanib (Votrient)	Inhibits multiple tyrosine kinases including VEGFR-1, VEGFR-2, PDGFR, FGF, Kit, Lck, and cFms	Hepatic P-450 metabolism; excreted in feces	Hepatotoxicity, QT prolongation, hypertension, fatigue, nausea, decreased cardiac function	Renal cell cancer, soft tissue sarcoma
Panitumumab (Vectibix)	Monoclonal antibody that binds EGFR, inhibiting ligand interaction	Prolonged (7-day) half-life	Acneiform rash (may be severe), diarrhea, infusion reaction, hypomagnesemia, pulmonary fibrosis	EGFR-expressing colorectal cancer

TABLE 179-4 U.S. FDA-APPROVED DRUGS COMMONLY USED FOR THE SYSTEMIC TREATMENT OF CANCER—cont'd

DRUG NAME	DRUG CLASS AND MECHANISM OF ACTION	PHARMACOKINETICS AND METABOLISM	TOXICITY	INDICATIONS
Pertuzumab (Perjeta)	Monoclonal antibody against HER2; blocks HER2-mediated signaling; associated with antibody-dependent cell-mediated cytotoxicity	Prolonged half-life	Decreased cardiac ejection fraction (particularly in patients previously exposed to anthracyclines), hypersensitivity reactions, nausea, diarrhea, rash	HER2-positive metastatic breast cancer in combination with trastuzumab and docetaxel for patients without prior anti-HER2 therapy
Ponatinib (Iclusig)	BCR-ABL kinase inhibitor; also inhibits BCR-ABL carrying the T315I resistance mutation; also inhibits other tyrosine kinases including VEGFR, FGFR, SRC, PDGFR, and FLT3	Hepatic metabolism with fecal excretion	Significant risk of arterial thrombosis, stroke, myocardial infarction; myelosuppression, hypertension, hepatotoxicity, fatigue, rash	Because of thrombotic risk, usage limited to patients with CML resistant to or intolerant of first- or second-line BCR-ABL inhibitors
Romidepsin (Istodax)	Histone deacetylase inhibitor; catalyzes removal of acetyl groups from lysines on histone proteins, enhancing transcription from less condensed chromatin	Metabolized by hepatic P-450 system	Nausea and vomiting, T-wave changes and QT prolongation on ECG, diarrhea, infections, hypomagnesemia, hypotension, myelosuppression, hypersensitivity reactions	Cutaneous T-cell lymphoma
Rituximab (Rituxan)	Chimeric antibody targeting B-cell CD20 surface antigen on lymphocytes	Proteolysis without substantial excretion	Hypersensitivity reactions, lymphopenia	Relapsed low-grade CD20-positive non-Hodgkin lymphomas
Rogorafenib (Stivarga)	Multikinase inhibitor; targets VEGFR2, TIE-2; inhibits angiogenesis	Hepatic metabolism with excretion into feces>urine	Hypertension, hepatotoxicity, fatigue, abdominal pain	Colorectal cancer, GIST
Ruxolitinib (Jakafi)	JAK1 and 2 kinase inhibitor; blocks JAK-dependent immune and hematopoietic signal transduction pathways	Hepatic metabolism with excretion into urine >feces	Myelosuppression, bruising, headache, weight gain, liver function abnormalities	Myelofibrosis
Sorafenib (Nexavar)	Multikinase inhibitor; inhibits RAF kinase, as well as VEGF and PDGF receptors; antiangiogenic	Metabolized in liver, excreted in feces; patients with severe liver or kidney dysfunction do not tolerate drug well	Rash, hand-foot syndrome, fatigue, diarrhea, hair loss, hypertension, arthralgias, myelosuppression, cardiac ischemia and QT prolongation	Renal cell carcinoma, hepatocellular carcinoma, thyroid cancer
Sunitinib maleate (Sutent)	Multitargeted tyrosine kinase inhibitor with activity against VEGFR1-3, FLT-3, PDGFR; antiangiogenic	Metabolized by P-450's in liver, with excretion in feces	Bleeding, decreased cardiac function, prolonged QT interval, hypertension, myelosuppression, nausea, rash, liver function abnormalities	Renal cell cancer, GIST, pancreatic neuroendocrine tumors
Temosirolimus (Torisel)	Inhibits mTOR kinase-dependent cell signaling	Metabolized in liver	Hypersensitivity reaction, bowel perforation, interstitial lung disease	Renal cell cancer
Thalidomide (Thalomid)	Immunomodulatory and antiangiogenic agent; inhibits TNF-α production; alters endothelial cell proliferation and cytokine production	Nonenzymatic hydrolysis; eliminated in urine	Teratogenicity, sedation, constipation, peripheral neuropathy, rash	Multiple myeloma
Trastuzumab (Herceptin)	Recombinant monoclonal antibody against HER2; downregulates expression of HER2 pathways; immune-mediated effects	Minimal renal or hepatic clearance	Hypersensitivity reactions, fever, and chills; nausea; enhances anthracycline cardiac toxicity	Metastatic or adjuvant HER2-expressing breast cancer or HER2-positive gastric cancer

Vemurafenib (Zelboraf)	Inhibits B-RAF kinase carrying V600E mutation	Metabolized by liver, excreted into feces	QT prolongation, liver function abnormalities, photosensitivity, alopecia, arthralgias, fever, rash, hyperkeratosis and skin papillomas	Malignant melanoma carrying B-RAF V600E mutation
Vismodegib (Erivedge)	Hedgehog pathway inhibitor; binds and inhibits Smoothened, a transmembrane G-protein receptor important for signal transduction in the hedgehog pathway	Hepatic metabolism, excretion in feces	Muscle spasms, fatigue, alopecia, weight loss, diarrhea	Advanced basal cell carcinoma
Vorinostat (Zolinza)	Histone deacetylase inhibitor; catalyzes removal of acetyl groups from lysines on histone proteins, enhancing transcription from less condensed chromatin	Metabolized in liver, eliminated in urine	Deep venous thrombosis, diarrhea, fatigue, alopecia, myelosuppression,	Cutaneous T-cell lymphoma
Zoledronic acid (Zometa)	Bisphosphonate inhibitor of osteoclastic bone resorption	Renal elimination	Bone pain, arthralgias and muscle pain, fever, fatigue, abnormal renal function, osteonecrosis of jaw, atypical subtrochanteric femoral fractures	Hypercalcemia of malignancy, multiple myeloma; prevention of bone fractures for patients with advanced breast and prostate cancer in concert with standard systemic therapy
DRUGS THAT AMELIORATE CHEMOTHERAPY SIDE EFFECTS				
Dexrazoxane (Zinecard)	Anthracycline protective agent that chelates iron and protects the heart by inhibiting formation of anthracycline-induced reactive oxygen species	Hepatic metabolism, excreted in urine	Myelosuppression, nausea and vomiting, stomatitis	Reduces cumulative cardiotoxicity when administered with anthracyclines; also ameliorates anthracycline-induced extravasation injury
Leucovorin (folinic acid, citrovorum factor, Wellcovorin)	Water-soluble folate vitamin; increases body and tumor pool of reduced folates; enhances 5-FU metabolite-mediated inhibition of thymidylate synthase	Renal excretion	Well tolerated by itself; occasional nausea	Prophylaxis and treatment of hematopoietic side effects of folic acid antagonists; enhanced efficacy of 5-FU for colon cancer and other GI malignancies
Mesna (Mesnex)	Synthetic sulfhydryl compound; metabolite, mesna disulfide, reacts chemically with urotoxic ifosfamide metabolites, resulting in their detoxification	Renal	Bad taste, diarrhea	Prophylaxis of cyclophosphamide/ifosfamide-induced hemorrhagic cystitis

*An intrathecal formulation, DepoCyt, is used for the treatment of carcinomatous meningitis.
†Dosing differs among agents.
‡Dosages differ among brands.
§An analogue of thalidomide, which is a severe human teratogen; restricted prescribing.
AML = acute myelogenous leukemia, ATP = adenosine triphosphate; bFGF, basic fibroblast growth factor; BMT = bone marrow transplantation; CrCe = creatine clearance, CHF = congestive heart failure; CLL = chronic lymphocytic leukemia; CML = chronic myelogenous leukemia; CNS = central nervous system; CSF = colony-stimulating factor; CTLA-4 = cytotoxic T-lymphocyte-associated protein 4; EGFR = epidermal growth factor receptor; ERBB2 = HER2/neu; FDA = U.S. Food and Drug Administration; FSH = follicle-stimulating hormone; G-CSF = granulocyte colony-stimulating factor; GM-CSF = granulocyte-macrophage colony-stimulating factor; GI = gastrointestinal; GIST = gastrointestinal stromal tumor; GnRH = gonadotropin-releasing hormone; 5-HIAA = 5-hydroxyindolacetic acid; HIV = human immunodeficiency virus; IL = interleukin; LFT = liver function test; LH = luteinizing hormone; LHRH = luteinizing hormone–releasing hormone; MAO = monoamine oxidase; mTOR = mammalian target of rapamycin; NSAID = nonsteroidal anti-inflammatory drug; PDGF = platelet-derived growth factor; SIADH = syndrome of inappropriate secretion of antidiuretic hormone; TKI = tyrosine kinase inhibitor; TNF = tumor necrosis factor; VEGF = vascular endothelial growth factor.

cytotoxin to enhance cell killing.[■] Radioimmunoconjugate approaches have been most effective in the treatment of non-Hodgkin lymphoma (Chapter 185) and chronic lymphocytic leukemia (Chapter 184). The effectiveness of monoclonal antibodies in specific tumor types is not identical to small molecules developed against the same target, in part because of the induction of immunologically mediated mechanisms of tumor cell killing that are unique to antibodies.

Hormonal Therapies

Endocrine or hormonal therapy for cancer, the earliest form of systemic therapy, is almost entirely limited to breast cancer (Chapter 198) and prostate cancer (Chapter 201). Many premenopausal breast cancers are thought to be under the influence of estrogens, and hormonal deprivation (ablation) may produce long-term responses in properly selected patients (those with estrogen and/or progesterone receptor positivity who have predominantly soft tissue or bone disease). The antiestrogen tamoxifen is effective against breast cancer, and it may decrease the incidence of contralateral breast cancers in both premenopausal and postmenopausal women with breast cancer. It also has an estrogen-like activity that is responsible for an increased rate of endometrial cancers. Postmenopausal women who are candidates for hormonal therapy may also respond to tamoxifen; however, aromatase inhibitors (e.g., anastrozole, letrozole, exemestane), which decrease the conversion of metabolites in fat and muscle into estrogen, have been found to be more effective than tamoxifen as first-line therapy in both the adjuvant and metastatic settings.

Prostate cancer (Chapter 201) is usually androgen dependent, and androgen deprivation can produce meaningful responses. The recent introduction of more potent inhibitors of androgen biosynthesis (abiraterone) and androgen receptor-mediated signal transduction (enzalutamide)[■] has further enhanced the range and effectiveness of androgen deprivation therapy for this disease.

The corticosteroids (Chapter 35), typically prednisone or dexamethasone, are widely used in the treatment of hematologic and oncologic cancers. In Hodgkin disease (Chapter 186), the non-Hodgkin lymphomas (Chapter 185), and multiple myeloma (Chapter 187), corticosteroids have antitumor activity. In solid tumor patients, they are used as antiemetics and for symptomatic relief of cerebral edema in cases of CNS metastases (Chapter 189), or as an adjunct to radiation therapy for spinal cord metastases.

Immunotherapy

Recently, several new approaches to improving cancer immunotherapy by blocking the negative effects on the immune system produced by tumors have yielded dramatic clinical benefits for patients with a variety of advanced cancers, including melanoma, kidney, and lung cancers. The survival of men and women with metastatic melanoma (Chapter 203) was significantly increased following treatment with an antibody (ipilumumab, anti-CTLA-4) that neutralizes proteins that protect tumor cells against destruction by the immune system.[■] Additional antibodies that target other immunologic checkpoints (anti-PD-1 and anti-PD-L1) are in advanced stages of clinical investigation and are likely to provide substantive clinical benefits for patients with melanoma, renal cancer (Chapter 197), and non–small cell lung cancer.

Drugs for Prevention of Toxicity

In addition to the hematopoietic growth factors used to reduce the adverse effects of systemic cancer therapies on the bone marrow (discussed later under Management of Complications), there are drugs that have been developed to ameliorate important side effects of cytotoxic chemotherapy (see Table 179-4). These include dexrazoxane, an iron chelating agent that can prevent the cardiac toxicity of the anthracyclines (doxorubicin and daunorubicin); leucovorin, which can diminish the hematologic side effects of folic acid antagonists; and mesna, a thiol-containing compound that blocks damage to the bladder mucosa from metabolites of cyclophosphamide.

Bone Marrow or Hematopoietic Stem Cell Transplantation

Because the major dose-limiting toxicity of most chemotherapeutic agents is myelosuppression, approaches have been developed to harvest the pluripotent stem cells found in bone marrow, peripheral blood, or, less often, cord blood before marrow-damaging chemotherapy so that the stem cells can be reinfused later (Chapter 178). This technique is most effective for acute leukemias (Chapter 183), relapsed lymphomas (Chapter 185), and germ cell tumors (Chapter 200). The effectiveness of the approach is limited more by the inability to eradicate cancer cells than by the inability to achieve engraftment. Transplants may be syngeneic (from an identical twin), autologous (from oneself), allogeneic (from a matched donor such as a sibling or parent), or from a matched unrelated donor. Nonablative hematopoietic transplants that do not completely abolish myelopoiesis reduce toxicity and allow the treatment of older and medically infirm patients. Hematopoietic stem cell transplantation is discussed in detail in Chapter 178.

Special Treatment Populations

Obesity

Studies of practice patterns indicate that up to 40% of obese patients receive limited doses of chemotherapy that are not based on actual body weight. Concerns about toxicity or overdosing based on the use of actual body weight in obese patients with cancer are unfounded. The American Society of Clinical Oncology (ASCO) has published evidence-based practice guidelines that recommend that full cytotoxic chemotherapy doses be used to treat obese patients with cancer, especially when the goal of treatment is cure.[5]

Pregnancy

Cancer during pregnancy is not uncommon, with breast, cervical, ovarian, and thyroid cancers, melanoma, and hematologic malignancies being most common. This is an emotionally charged time, and clinical decision making is complicated by ethical, moral, cultural, and religious issues. If surgery can be safely accomplished, this may be the best course, even if it is only a temporizing measure. Radiation therapy carries the very real risk of radiation exposure to the fetus, and staging is almost always suboptimal and confined to ultrasound examinations. When the disease requires chemotherapy, changes in both the mother and fetus must be taken into account; for instance, there are major changes in drug clearance during pregnancy, along with gastrointestinal absorption and placental transfer, not to mention fetal pharmacokinetics and placental excretion. Many commonly used chemotherapeutic drugs are classified by the FDA as category D (positive human fetal risk, but the benefits in pregnant women may be acceptable despite the risk) or category X (studies in humans and animal have shown fetal malformations or there is evidence of fetal risk based on human evidence). If the mother's condition permits, it is advisable to defer chemotherapy (including anthracyclines and taxanes) during the first trimester and to treat life-threatening situations during the third trimester after extensive counseling with the parents.

Geriatrics

An increasing proportion of cancers occur in the older population. The physiologic changes that develop with age include: decreased excretion of drugs and metabolites from the kidneys, decreased volume of distribution of water-soluble drugs, and increased susceptibility to myelosuppression, cardiomyopathy, and neuropathy, related in part to comorbid conditions. As a general rule, the suitability of an older patient for therapy can be determined by a comprehensive geriatric assessment (CGA) that evaluates the patient's function, comorbidity, nutrition, medications, and resources. Geriatric assessment is discussed in detail in Chapter 24. By itself, age is not a barrier to surgery; rather, the patient's performance status and the CGA should determine the likelihood of a good recovery. Tolerance of radiation therapy seems to remain largely intact with increasing age. Chemotherapy decisions are also based on the performance status and CGA. Dosage adjustments are also made for individual glomerular filtration rates for patients aged 65 and older, where appropriate. The use of lower chemotherapy doses based on age alone is not advisable and may result in ineffective treatment.

Organ Dysfunction

Alterations in drug clearance play a critical role in the safe administration of anticancer agents. Over the past decade, pharmacokinetic studies have begun to detail the landscape of how specific levels of carefully defined renal or hepatic dysfunction alter the clearance and tolerance of many of the most commonly used drugs for the systemic treatment of cancer. For each new agent, prospective investigations are required to define usage parameters for each clinically defined level of organ dysfunction. It should be pointed out that alterations in pharmacokinetic parameters per se may occur with or without important changes in toxicity. Where evidence exists, applicable recommendations for chemotherapeutic drug use in the setting of renal or hepatic dysfunction are outlined in Table 179-4.

Management of Complications

Supportive Care

Nutritional Support

Nutrition is always a concern for patients newly diagnosed with cancer, even if they have not experienced weight loss. In fact, significant weight loss is an adverse prognostic factor for several cancers, especially lung cancer. Patients are often concerned about whether their diet contributed to development of the cancer and whether diet can influence the results of therapy. In most settings, neither of these scenarios is the case. Malnourished patients should be evaluated by a dietitian to determine whether they are ingesting sufficient calories and whether dietary supplements might be needed. Nutritional assessment is discussed in detail in Chapter 214. Some patients, such as those with head and neck cancers (Chapter 190) or esophageal cancers (Chapter 192), may require parenteral nutrition through a percutaneous gastrostomy tube. Total parenteral nutrition (Chapter 217) is rarely indicated. Larger-than-recommended doses of vitamins are also not helpful and may be toxic. It is important to determine whether over-the-counter and/or

alternative medications (Chapter 39) are being contemplated or used by the patient because of the potential for drug interactions.

Psychosocial Support

Patients with a recent cancer diagnosis have increased risks of death from cardiovascular causes, especially during the first week after diagnosis. The need for continuing psychosocial support in the face of ongoing cancer treatment, and the associated anxiety, depression, and fear experienced by many patients, is substantive and may be beyond the ability of the immediate family to fulfill. In this setting, patients often benefit from participation in support groups or from direct one-on-one counseling, and from efforts to improve communication across all levels of care and support systems.

Hematopoietic Growth Factors

Growth factors, such as granulocyte colony-stimulating factor (G-CSF) and granulocyte-macrophage colony-stimulating factor (GM-CSF), speed recovery from white blood cell count depression, permitting chemotherapy to be given on schedule, without reducing the dosage in many cases (Chapter 156). However, such therapy does not decrease hospitalizations or improve survival. It is possible to determine which individuals are at greatest risk for febrile neutropenia (Chapter 167) and to treat them in advance, based on published guidelines.[6] Correction of anemia with erythropoiesis-stimulating agents (ESAs) (Epoetin alfa and Darbepoietin alfa), while possible, may be associated with complications of adverse cardiovascular events and even potentially tumor progression (Chapter 158).[7]

Prevention of Pathologic Bone Fractures

The bisphosphonates pamidronate and zoledronate are very effective not only for the treatment of tumor-induced hypercalcemia but also to reduce pathologic fractures in bones with metastatic lesions, particularly from breast cancer (Chapter 198), prostate cancer (Chapter 201), and myeloma (Chapter 187). They are also used to treat osteoporosis caused by chemotherapy-induced premature menopause in young women with breast cancer (Chapter 243). Denosumab is a human monoclonal antibody that binds to RANK ligand, a protein found on osteoclasts that is involved in bone breakdown. Some clinical trials have found denosumab to be superior to zoledronic acid for the prevention of skeletal-related events in cancer patients with bone metastases.

Symptom Management

Effective management of symptoms is critical to successful delivery of either curative or palliative treatment and maintenance of a patient's quality of life.

Nausea and Vomiting

Patients continue to fear chemotherapy because of the risk of nausea and vomiting. New antiemetics, used in combination, have made this side effect much less debilitating. Chemotherapeutic drugs can be ranked according to their probability of causing nausea and vomiting, with prophylactic treatment given accordingly. The availability of the serotonin 5-hydroxytryptamine type 3 (5-HT_3) receptor antagonists (dolasetron, granisetron, ondansetron) has dramatically improved our ability to completely control nausea and vomiting. More emetogenic regimens require combination therapy with a corticosteroid (usually dexamethasone), a 5-HT_3 antagonist, and a benzodiazepine (e.g., lorazepam) or the neurokinin-1 receptor antagonist, aprepitant. Aprepitant is particularly useful for the treatment/prevention of delayed nausea and vomiting. A double-blind randomized clinical trial of four combination regimens for controlling delayed nausea concluded that the addition of dexamethasone on days 2 and 3 was particularly effective.

Pain Control

Pain control[8] (Chapter 30) can be accomplished with a variety of analgesics, both non-narcotic and narcotic. Oncologists use a variety of scales for the evaluation of pain and start treatment with nonsteroidal anti-inflammatory drugs (NSAIDs) such as aspirin and acetaminophen, progress through ibuprofen and related drugs, and then through combinations of NSAIDs and narcotics to stronger narcotics. Newer narcotics are available in both short-duration and long-duration forms; some dermal patches last 72 hours, which is ideal for patients who have severe pain and are unable to take oral medications. Oral transmucosal fentanyl is more effective than standard-release morphine in this setting. Painful oral mucositis, a common complication of intensive therapy for hematologic malignancies, can be treated with local measures or with recombinant human keratinocyte growth factor. Oral anti-*Candida* drugs that are absorbed or partially absorbed from the gastrointestinal tract can help prevent pain from oral candidiasis. American Pain Society standards for pain management in cancer recommend both pharmacologic[9] and psychosocial[10] interventions as complementary approaches. A recently published meta-analysis of randomized controlled studies of various psychosocial interventions among adult cancer patients (e.g., relaxation training, cognitive behavioral therapy, and other education- and skills-based approaches) demonstrated medium-sized effects on both pain severity and interference with daily activities.

Malignant Effusions

Accumulations of fluid and malignant cells in the pleural, peritoneal, or pericardial spaces are common complications of epithelial and hematopoietic malignancies that frequently produce a significant array of symptoms, either at the time of diagnosis or accompanying tumor progression. Malignant pleural effusions (Chapter 99) are most commonly associated with cancers of the lung and breast or lymphomas, may be the result of lymphatic obstruction or direct invasion of pleural membranes, and can produce significant degrees of dyspnea, cough, or pain that require therapy. Diagnostic thoracentesis of sufficient volume (>60 mL), with cytologic analysis of the pleural effusion, has a reasonably high diagnostic yield for malignancy (60 to 90%). In patients with previously untreated lymphoma, breast cancer, or small cell lung cancer, objective response to the initiation of systemic chemotherapy may provide long-term symptomatic relief. However, in patients with recurrent lung or breast cancer, for example, pleural effusions that are confirmed to contain malignant cells may present difficult ongoing therapeutic challenges. For symptomatic patients, therapeutic thoracentesis, usually under ultrasound guidance, is required and may need to be repeated to reduce dyspnea. When frequent thoracenteses over short intervals are needed, a pleurodesis procedure is often performed, encompassing drainage of the pleural space with a chest thoracostomy and the instillation of a sclerosing compound (talc, doxycycline) that will initiate an inflammatory response of sufficient magnitude to obliterate the pleural space. Pleurodesis is at least temporarily successful in preventing fluid recurrence in most patients; when it is not, placement of an indwelling pleural catheter may provide long-term symptomatic relief of dyspnea.

Malignant ascites (peritoneal effusion) occurs most frequently in patients with intra-abdominal malignancies (gastric, ovarian, pancreatic, and primary peritoneal cancers) but can be observed as well in patients with advanced breast and lung cancers or lymphoma. Malignant ascites may be caused in part by increased permeability of the tumor vasculature that is a result of vascular endothelial growth factor overexpression, by inflammatory cytokine overproduction in the peritoneal space, or by lymphatic blockade secondary to carcinomatosis. Ultrasound-guided paracentesis provides relief of bloating, dyspnea, and the pain of abdominal distension, but will often need to be repeated, which carries the risk of dehydration, protein loss, electrolyte imbalance, bleeding, infection, and kidney dysfunction. A requirement for paracentesis at frequencies less than 1 week should prompt consideration of placement of a permanent catheter to allow self-drainage, although these devices carry a significant risk of infection.

Malignant pericardial effusions (Chapter 77) are most commonly related to direct extension or metastatic spread from lung or breast cancers, melanomas, and hematologic malignancies. As is the case for other malignant effusions, image-guided pericardiocentesis with cytologic examination of the fluid that has been evacuated will frequently provide diagnostic confirmation of malignancy; furthermore, even the removal of a relatively modest amount of fluid (<50 mL) may, at least partially, relieve the hemodynamic compromise produced by the effusion. The approach to a patient with malignant pericardial effusions is dictated by hemodynamic status (which can drive the choice between emergency pericardiocentesis or elective pericardiostomy) and by the predicted sensitivity of the inciting tumor to systemic therapy (untreated lymphoma versus chemotherapy-resistant lung cancer, for example).

ENDOCRINE MANIFESTATIONS OF CANCER

Clinical syndromes associated with ectopic hormone production may pose special diagnostic dilemmas, can produce a significant degree of morbidity or even death in cancer patients, and may be difficult to treat (Table 179-5). Management of these syndromes involves the simultaneous treatment of both the cancer and the syndrome caused by excessive hormone production. Many of the endocrine manifestations of cancer are caused by the production of small polypeptide hormones by tumors, some of which are derived from specific types of neuroendocrine cells. These cells are widely dispersed in a wide variety of organs, are often of neural crest origin, and can produce biogenic amines. The hormones produced from these tumors include adrenocorticotropic hormones (corticotropin, ACTH), calcitonin, vasoactive intestinal peptide, growth hormone–releasing hormone, corticotropin-releasing hormone (CRH), somatostatin, and other peptides. A second group of tumors, generally derived from squamous epithelium, produces parathyroid hormone–related proteins (PTHrP) and vasopressin.

Hypercalcemia of Malignancy

Humoral hypercalcemia is one of the most common endocrine syndromes related to an underlying malignancy. There are several different underlying mechanisms related to this pathophysiologic process, including ectopic production of PTHrP with activation of the PTH receptor to increase osteoclast

TABLE 179-5 SOME CLINICAL SYNDROMES OF ECTOPIC HORMONE PRODUCTION

Humoral hypercalcemia
Parathyroid hormone–related protein
Squamous cell carcinoma
Breast cancer
Neuroendocrine tumors
Renal cell cancer
Melanoma
Prostate cancer
Increased calcitriol
Lymphoma
Benign conditions: sarcoid, berylliosis, tuberculosis, fungal infections
Corticotropin
Proopiomelanocortin
Small cell lung cancer
Pulmonary carcinoid
Medullary thyroid cancer
Islet cell tumor
Pheochromocytoma
Ganglioneuroma
Corticotropin-releasing hormone
Medullary thyroid cancer
Paraganglioma
Prostate cancer
Islet cell tumors
Human chorionic gonadotropin
Choriocarcinoma
Testicular embryonal cell carcinoma
Seminoma
Hypoglycemia
Insulinoma
Sarcomas or large retroperitoneal tumors
Inappropriate antidiuretic hormone secretion
Small cell lung cancer
Squamous cell head and neck cancer
Erythropoietin
Renal cell cancer
Hepatoma
Pheochromocytoma
Benign conditions: cerebellar hemangioblastoma, uterine fibroids

TABLE 179-6 EVALUATION AND DIAGNOSIS OF PARANEOPLASTIC SYNDROMES

Characterize abnormality; obtain laboratory studies and biopsy as necessary.
Carefully elicit any additional symptoms and signs.
Eliminate common causes.
If there is no obvious etiology, consider a paraneoplastic syndrome.
If findings are consistent with a known syndrome, screen for underlying malignancy.
If signs and symptoms are consistent with a known paraneoplastic syndrome, undertake a search for an unknown primary cancer or recurrence or progression of a known primary tumor.
Screening should include a careful physical examination with breast, gynecologic, and prostate evaluations; basic hematology, chemistry, and urine studies; chest radiograph; and mammogram.
Computed tomography (CT) of the abdomen and pelvis or positron emission tomography (PET) scan is indicated if there are any suspicious symptoms, signs, or laboratory abnormalities.
Antibody testing for paraneoplastic neurologic syndromes and/or skin biopsy should be performed as indicated.
Consider treatment of cancer and/or appropriate palliative treatment, including immunosuppressive therapy for paraneoplastic symptoms when possible.

differentiation and bone resorption, with consequent hypercalcemia. Ectopic PTHrP (rather than PTH) production by several different types of cancer, most characteristically in squamous cell, breast, renal cell, and prostate cancer, as well as neuroendocrine tumors and melanoma (see Table 179-5), is one of the most common causes of hypercalcemia of malignancy.[11] Increased production of calcitriol, which increases calcium absorption with suppression of serum PTH levels, is another cause of malignant hypercalcemia that is most commonly observed in patients with lymphoma. Bone metastases, particularly in patients with breast cancer and myeloma, may produce hypercalcemia due to increased local production of PTHrP or other cytokines that increase bone resorption.

The treatment of malignant hypercalcemia is similar to that caused by hyperparathyroidism (Chapter 245) in that reversal of dehydration and the initiation of a saline diuresis should begin early; patients with a serum calcium in excess of 13 mg/dL should be treated with a bisphosphonate initially, with extended use of the bisphosphonate, as described earlier, for the prevention of bone fractures and recurrence of hypercalcemia (Chapter 245).

Other Ectopic Hormone Syndromes

Inappropriate secretion of ACTH is rare but resembles pituitary Cushing disease (Chapter 224); tumors that produce CRH include medullary thyroid cancer, prostate cancer, and islet cell neoplasms. Ectopic ACTH syndrome may become manifest as classic Cushing syndrome, with easy bruisability, centripetal obesity, muscle wasting, hypertension, diabetes, and metabolic alkalosis, although many patients with ectopic ACTH-producing cancers progress too quickly to develop prominent cushingoid manifestation clinically. Profound hypokalemia may predominate without all the classic features of Cushing syndrome in patients with small cell lung cancer.

Tumor-associated hypoglycemia, although uncommon, may be the result of: insulin overproduction by islet cell tumors; insufficient hepatic gluconeogenesis related to loss of functional hepatic mass by metastatic disease; and overexpression of insulin-like growth factor II, which can activate the insulin receptor in patients with large retroperitoneal sarcomas or hepatocellular carcinomas. In each of these cases, treatment with frequent small feedings can be prescribed; however, successful symptomatic management of hypoglycemia may be difficult without control of the primary tumor mass or metastases.

The clinical syndrome of inappropriate secretion of antidiuretic hormone is caused by ectopic production of vasopressin, primarily in patients with small cell lung cancer or squamous cancers of the head and neck, and occasionally in those with primary brain tumors. It is characterized by hyponatremia, hypo-osmolality, excessive urine sodium excretion, an inappropriately high urine osmolality for the low serum osmolality, and normal kidney, adrenal, and thyroid function (Chapter 116). Fluid (free water) restriction can provide adequate short-term management of symptomatic hyponatremia; however, treatment with demeclocycline, which blocks the effects of vasopressin on the kidney, provides more effective long-term therapy.

PARANEOPLASTIC SYNDROMES

The term *paraneoplasia*, which means "alongside cancer," has been commonly used to denote remote effects of cancer that cannot be attributed either to direct invasion or to distant metastases. These syndromes may be the first sign of a malignancy and affect up to 15% of patients with cancer (Table 179-6). However, if patients with cachexia are excluded, the incidence probably drops to only a few percent. Paraneoplastic syndromes may be the initial presenting sign or symptom of an underlying malignancy. Up to two thirds of paraneoplastic syndromes arise before an associated malignancy is diagnosed. In some cases, the paraneoplastic syndrome may be associated with relatively small tumors; recognition of these associations may lead to earlier diagnosis and possibly more effective therapy. Furthermore, one of the hallmarks in defining a paraneoplastic syndrome is that the course of the syndrome generally parallels the course of the tumor. Therefore, effective treatment of the underlying malignancy is often accompanied by improvement or resolution of the syndrome. Conversely, recurrence of the cancer may be heralded by the return of systemic symptoms. The numerous neurologic paraneoplastic syndromes[12] are reviewed in Chapter 411.

Dermatologic Paraneoplastic Syndromes

Associations between cutaneous syndromes and underlying malignancies may be difficult to confirm. Generally, the skin condition and cancer follow a parallel course, and the two diagnoses should be made at about the same time. Some skin lesions are almost always associated with malignancy. Others, however, are nonspecific and are most commonly seen with nonmalignant conditions, making it difficult or impossible to connect the skin disease with the underlying malignancy. In addition, biopsies of the skin lesion are usually nonspecific, showing features identical to those when the same lesion is seen without a malignant condition. The formation of tumor-related autoantibodies has rarely been associated with dermatologic paraneoplastic syndromes, although inflammatory cell infiltration may be seen.

Recognition of cutaneous manifestations of malignancy can be critical for the early diagnosis and successful treatment of cancer, but some syndromes

are seen only with advanced, incurable disease. Cutaneous manifestations include direct involvement of the skin with tumor as well as the remote effects of cancer.[13,14] Both specific and nonspecific dermatologic adverse effects are also seen with cytotoxic chemotherapeutic agents, including alkylating agents, antimetabolites, anthracyclines, and antitumor antibiotics.

One of the best-known paraneoplastic syndromes is acanthosis nigricans, the pathogenesis of which is unclear. The tumor may produce factors that activate insulin-like growth factors or the insulin receptor in skin. Many tumors are known to produce transforming growth factor-α (TGF-α), which might activate epidermal growth factor receptors in skin, causing hyperpigmentation and thickening. The skin lesions arise as velvety, verrucous hyperpigmentation of the neck, axilla, groin, and mucosal membranes, including the lips, periocular area, and anus. Although acanthosis nigricans clearly occurs as a benign entity associated with obesity and endocrinopathy, its appearance in older adults, especially when it includes mucosal lesions, has been highly associated with malignancies of the gastrointestinal tract as well as other adenocarcinomas. The lesions often regress with successful treatment of the underlying tumor.

Rheumatologic Paraneoplastic Syndromes

In patients who have rheumatic disorders with atypical clinical presentation—particularly older patients, those with coexisting systemic symptoms, and patients who respond unexpectedly poorly to usual antirheumatic treatments—the possibility of an underlying occult malignancy should be considered.[15] Chemotherapeutic agents can also cause rheumatic adverse effects.[16]

One of the more common and specific rheumatologic paraneoplastic syndromes is hypertrophic osteoarthropathy, which arises as an oligoarthritis or polyarthritis of the distal joints, with clubbing, tender periostitis of the distal long bones, and noninflammatory synovial effusions (also see Chapter 275). Hypertrophic osteoarthropathy may affect up to 10% of patients with adenocarcinoma of the lung. It is also seen with a variety of other pulmonary malignancies, including lung metastases from other primary sites. The etiology is unknown. Laboratory studies often reveal an elevation in the erythrocyte sedimentation rate; bone radiographs show linear ossification of the distal long bones separated by a radiolucent zone from the underlying cortex (Fig. 179-1). Treatment is symptomatic with anti-inflammatory agents; successful treatment of the underlying tumor may also improve the signs and symptoms of this syndrome.

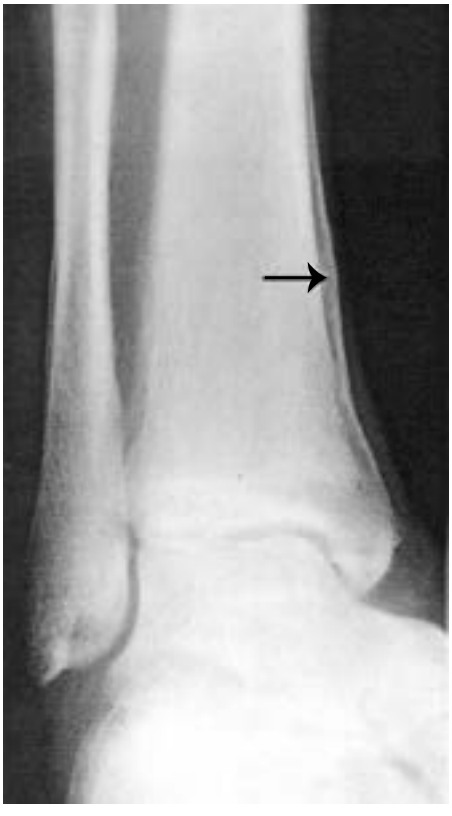

FIGURE 179-1 Hypertrophic pulmonary osteoarthropathy characterized by periosteal elevation of the tibia (*arrow*). (Courtesy Dr. Lynne S. Steinbach.)

Fever and Cachexia

Fever (Chapter 280), night sweats, and cachexia are nonspecific symptoms that, when seen in the absence of infection or a known disorder, suggest the diagnosis of an underlying malignancy. Cytokines clearly play a pathogenetic role in inducing both fever and cachexia. TNF-α, interleukins (particularly IL-1 and IL-6), and interferon-γ are produced directly by the tumor or by tumor-associated host inflammatory cells, such as macrophages, which results in a catabolic state. Cytokines may produce fever directly by acting at the level of the hypothalamic thermoregulatory center. In addition to the burden of tumor and the production of cytokines, cachexia may be caused or worsened by the side effects of cancer treatment, by intestinal blockage or malabsorption caused by tumor infiltration, and by depression.

Fever is generally cyclic and may be associated with drenching night sweats. Symptoms resolve with successful treatment of the underlying tumor, and return of fever usually heralds relapse. When treatment of the tumor is not possible or is ineffective, NSAIDs or steroids given around the clock significantly improve quality of life. Although cancer-related fever is most commonly seen in association with malignant lymphoproliferative disease (Chapters 185 and 186), renal cell carcinoma (Chapter 197), and leukemias (Chapters 183 and 184), it may also occur with other cancers, particularly in the face of extensive hepatic metastases.

Cachexia, or the cancer wasting syndrome, is probably the single most common paraneoplastic syndrome, eventually affecting up to 80% of patients with cancer. This syndrome is characterized by anorexia, muscle wasting, loss of subcutaneous fat, and fatigue. It appears to be caused by a combination of protein wasting, malabsorption, immune dysregulation, and increased glucose turnover in the setting of tumor-induced increases in energy expenditure. Successful treatment of the underlying tumor reverses the process; symptomatic treatment for patients with advanced disease is modestly successful at best. Megestrol acetate given in high concentrations in liquid form (400 to 800 mg/day) can improve appetite and result in weight gain, but at the cost of fluid retention.

SURVIVORSHIP AND FOLLOW-UP

Approximately 4% of the United States population (≈14 million people) is living with a history of cancer; about 60% of cancer survivors are age 65 or older.[17] Thus, there is a large and increasing number of individuals who are living longer during the period of cancer survivorship—from the end of active treatment to the point of recurrence or death from another condition. There has been a growing appreciation of the specific care needs of these patients, including: a defined program of surveillance to detect recurrence or second cancers and the late effects of cancer treatment; intervention to treat the consequences of the cancer and its treatment (such as lymphedema, fatigue, and psychosocial distress); prevention of new cancers through changes in diet, behavior, and physical activity; and the institution of a coordinated program of care for cancer survivors that may require a variety of specialty services. Long-term follow-up can be optimized by the provision of a survivorship care plan for patients that provides a comprehensive summary of all diagnostic and therapeutic procedures undergone, toxicities experienced, and therapeutic outcomes, as well as a specific program of individualized follow-up care. Although definitive follow-up regimens do not exist for most cancers, evidence-based templates for common malignancies have been developed by the American Society of Clinical Oncology and the National Comprehensive Cancer Network. Lastly, it must be remembered that issues of survivorship affect caregivers, who often experience a high degree of psychological distress, along with the patient, during and after the period of active treatment.

Grade A References

A1. Maemondo M, Inoue A, Kobayashi K, et al. Gefitinib or chemotherapy for non-small cell lung cancer with mutated EGFR. *N Engl J Med*. 2010;362:2380-2388.

A2. Verma S, Miles D, Gianni L, et al. Trastuzumab emtansine for HER2-positive advanced breast cancer. *N Engl J Med*. 2012;367:1783-1791.

A3. Scher HI, Fizazi K, Saad F, et al. Increased survival with enzalutamide in prostate cancer after chemotherapy. *N Engl J Med*. 2012;367:1187-1197.

A4. Hodi FS, O'Day SJ, McDermott DF, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med*. 2010;363:711-723.

A5. Mhaskar R, Clark OA, Lyman G, et al. Colony-stimulating factors for chemotherapy-induced febrile neutropenia. *Cochrane Database Syst Rev*. 2014;10:CD003039.

A6. Fizazi K, Carducci M, Smith M, et al. Denosumab versus zoledronic acid for treatment of bone metastases in men with castration-resistant prostate cancer: a randomized, double-blind study. *Lancet*. 2011;377:785-786.
A7. Basch E, Prestrud AA, Hesketh PJ, et al. Antiemetics: American Society of Clinical Oncology clinical practice guidelines update. *J Clin Oncol*. 2011;29:4189-4198.
A8. Roscoe JA, Heckler CE, Morrow GR, et al. Prevention of delayed nausea: a University of Rochester Cancer Center Community Clinical Oncology Program study of patients receiving chemotherapy. *J Clin Oncol*. 2012;30:3389-3395.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Huang M, Shen A, Ding J, et al. Molecularly targeted cancer therapy: some lessons from the past decade. *Trends Pharm Sci*. 2014;35:41-50.
2. Ohashi K, Maruvka YE, Michor F, et al. Epidermal growth factor receptor tyrosine kinase inhibitor-resistant disease. *J Clin Oncol*. 2013;31:1070-1080.
3. Park SR, Davis M, Doroshow JH, et al. Safety and feasibility of targeted agent combinations for solid tumors. *Nat Rev Clin Oncol*. 2013;10:154-168.
4. Rebucci M, Michiels C. Molecular aspects of cancer cell resistance to chemotherapy. *Biochem Pharmacol*. 2013;85:1219-1226.
5. Griggs JJ, Mangu PB, Anderson H, et al. Appropriate chemotherapy dosing for obese adult patients with cancer: American Society of Clinical Oncology clinical practice guideline. *J Clin Oncol*. 2012;30:1553-1561.
6. Bennett CL, Djulbegovic B, Norris LB, et al. Colony-stimulating factors for febrile neuropenmia during cancer therapy. *N Engl J Med*. 2013;368:1131-1139.
7. Adamson JW. Iron, erythropoietic stimulating agents, and anemia in cancer. *Crit Rev Oncog*. 2013;18:471-483.
8. Hui D, Bruera E. A personalized approach to assessing and managing pain in patients with cancer. *J Clin Oncol*. 2014;32:1640-1646.
9. Pasternak GW. Opiate pharmacology and relief of pain. *J Clin Oncol*. 2014;32:1655-1661.
10. Syrgala KL, Jensen MP, Mendoza ME, et al. Psychological and behavioral approaches to cancer pain management. *J Clin Oncol*. 2014;32:1703-1711.
11. Meng QH, Wagar EA. Laboratory approaches for the diagnosis and assessment of hypercalcemia. *Crit Rev Clin Lab Sci*. 2014:1-13.
12. Muppidi S, Vernino S. Paraneoplastic neuropathies. *Continuum (Minneap Minn)*. 2014;20:1359-1372.
13. Yuste-Chaves M, Unamuno-Perez P. Cutaneous alerts in systemic malignancy: part I. *Actas Dermosifiliogr*. 2013;104:285-298.
14. Yuste Chaves M, Unamuno Perez P. Cutaneous manifestations of systemic malignancies: part 2. *Actas Dermosifiliogr*. 2013;104:543-553.
15. Ashouri JF, Daikh DI. Rheumatic manifestations of cancer. *Rheum Dis Clin North Am*. 2011;37:489-505.
16. Alias A, Rodriguez EJ, Bateman HE, et al. Rheumatology and oncology: an updated review of rheumatic manifestations of malignancy and anti-neoplastictherapy. *Bull NYU Hosp Jt Dis*. 2012;70:109-114.
17. Rowland JH, Kent EE, Forsythe LP, et al. Cancer survivorship in Europe and the United States: where have we been, where are we going, and what can we learn from each other? *Cancer*. 2013;119(suppl 11):2094-2108.

REVIEW QUESTIONS

1. A 65-year-old man reports symptoms of fatigue and loss of appetite. Physical examination demonstrates jaundice and an abdominal mass. CT scan of the abdomen reveals a mass at the head of the pancreas, with surrounding nodal involvement. The most appropriate next step in this patient's care is:

A. Examination of serum biomarker studies including CEA and α-fetoprotein
B. Image-guided biopsy of the abdominal mass
C. Initiation of combination chemotherapy for advanced pancreatic cancer
D. Referral for surgical resection and subsequent pathologic staging

Answer: B The diagnosis of locally advanced pancreatic cancer can often be made by CT-guided biopsy. In this patient with regionally advanced disease, surgery is unlikely to be beneficial; neither are serum biomarkers of use in making a tissue diagnosis, which is the first step in planning further treatment.

2. An 82-year-old woman is found to have a stage III left-sided invasive ductal carcinoma of the breast. She has well-controlled inflammatory bowel disease and adult-onset diabetes, lives alone, and has no immediate family. The most important factor in choosing a program for further diagnostic evaluation and treatment is her:

A. Age
B. Comorbid illnesses
C. ECOG performance score
D. Lack of family support

Answer: C A patient's performance status on the ECOG scale has been demonstrated to correlate well with tolerance of treatment and long-term outcome across a wide variety of malignancies to a much greater degree than age or individual comorbid illnesses.

3. In choosing a therapeutic regimen for systemic administration, of the following, which is the most important consideration for an adult cancer patient?

A. Drug pharmacokinetics
B. The ratio of the amount of treatment that can be delivered to the toxicity expected at that dose
C. Route of drug administration
D. The range of potential toxicities that might accompany drug administration

Answer: B Pharmacokinetic profiles and routes of administration are important factors in early drug development; however, the "therapeutic index" relating efficacy to toxicity is of paramount consideration in deciding which treatment should be employed.

4. Altered hepatic function would require substantive dose reduction of which one of the following anticancer agents?

A. 5-Fluorouracil
B. Imatinib
C. Oxaliplatin
D. Carboplatin

Answer: B Only imatinib, which unlike the other three drugs is metabolized by the hepatic cytochrome P-450 system, requires significant dose reduction for liver dysfunction.

5. A 52-year-old man with chronic hepatitis C develops a multifocal hepatocellular carcinoma, accompanied by malignant ascites and thrombocytopenia. The most appropriate approach to the control of the patient's malignant effusion is:

A. Fluid restriction and diuresis
B. Regular large-volume paracentesis in the clinic
C. Cryosurgery for the major hepatic lesions
D. Insertion of a permanent catheter for self-drainage

Answer: D This patient's tumor is not curable surgically; fluid restriction will lead to hyponatremia without producing symptomatic benefit; and large-volume paracentesis without ultrasound guidance in this patient carries an excessive risk of bleeding.

180 EPIDEMIOLOGY OF CANCER

DAVID J. HUNTER

EPIDEMIOLOGY

Overview

The International Agency for Research on Cancer (IARC) estimates that over 14.1 million new incident cases of cancer (excluding non-melanoma skin cancer) occurred in 2012, of which over 8 million (57%) occurred in less developed regions (defined as Africa, Asia [excluding Japan], Latin America, and the Caribbean, plus Melanesia, Micronesia, and Polynesia in Oceania). Over 8.2 million deaths occurred, with over 5.3 million (65%) in less developed regions.[1] This is predicted to increase to 21.7 million incident cases and 13 million deaths by 2030 (Fig. 180-1), with the majority of this increase being due to the aging of populations, as well as increased total population.

The 10 most common cancers diagnosed worldwide in 2012 were lung (13%), breast (12%), colon/rectum (10%), prostate (8%), stomach (7%), liver (6%), cervix (4%), esophagus (3%), bladder (3%), and non-Hodgkin lymphoma (3%) (Fig. 180-2A), with 8 of these being among the 10 most common causes of cancer death: lung (19%), stomach (9%), liver (9%), colon/rectum (9%), breast (6%), esophagus (5%), prostate (4%), pancreas (4%), cervix (3%), and leukemia (3%) (Fig. 180-2B). These global estimates mask very large differences in regional incidence rates of specific cancers, with age-standardized incidence rates varying over five-fold between low- and high-incidence regions (Fig. 180-3); for some cancers there is a 20-fold difference between the lowest and highest incidence rates in individual countries. The burden of cancer on individual countries is a function of age-specific rates and population size. Thus, the country with the largest population in the world faces the largest number of cases of cancer, but the United States, with the world's third largest population, has the second largest number of cases of cancer, owing to high age-specific rates, and a high proportion of people in older age groups (Fig. 180-4).

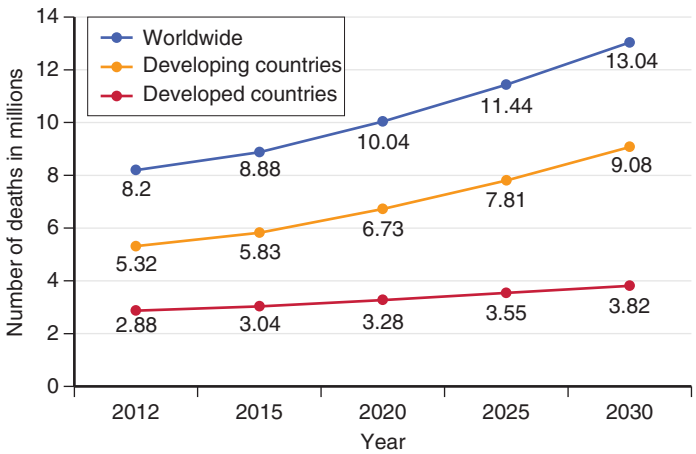

FIGURE 180-1. Projected Number of Cancer Deaths Worldwide, 2012-2030. (Data from Ferlay J, Soerjomataram I, Ervik M, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr. Accessed February 18, 2015.)

In the United States, the most common cancers predicted for 2014 for men are prostate, (27%), lung (14%), colorectal (8%), bladder (7%), melanoma (5%), renal (5%), non-Hodgkin lymphoma (4%), oropharyngeal (4%), leukemia (4%), and liver (3%), with 8 of these being among the 10 most common causes of cancer death: lung (28%), prostate (10%), colorectal (8%), pancreas (7%), liver (5%), leukemia (5%), esophagus (4%), bladder (4%), non-Hodgkin lymphoma (3%), and renal (3%), with the addition of pancreas (7%) and esophagus (4%). For women, these are breast cancer (20%), lung (13%), colorectal (8%), uterine (6%), thyroid (6%), non-Hodgkin lymphoma (4%), melanoma (4%), renal (3%), pancreas (3%), and leukemia (3%), with 7 of these being the among the top 10 causes of mortality: lung (26%), breast (15%), colon/rectum (9%), pancreas (7%), leukemia (4%), uterine (3%), and non-Hodgkin lymphoma (3%), with the addition of ovary (5%), liver (3%), and brain and other nervous system (2%).[2]

Demographic Factors

Age

The rates of most cancers increase with age, often in a log-linear (exponential) fashion; any population that experiences increases in life expectancy in the proportion of the population in older age groups will almost inevitably see an increase in the numbers of cases of cancer. Some cancers have different age-incidence curves, notably cancers that occur mostly in the first few years of life, such as retinoblastoma and neuroblastoma, or in young adulthood, such as testicular cancer. Hodgkin lymphoma has a bimodal age incidence curve with peaks in younger and older adults. Breast cancer rates increase with age in premenopausal women but plateau or increase more slowly postmenopausally.

Sex

Male breast cancer occurs at less than 1% of the rate of female breast cancer in most countries. For cancers that occur in both sexes, age-specific rates are often two- to three-fold higher in men than women; for smoking-related cancers, this is often due to higher prevalence and duration of smoking among men.

Causes of Cancer

People who migrate from low to high organ-specific cancer incidence countries tend to acquire the cancer incidence profile of their new country within one to three generations; for example, breast cancer risk among Asian American women with grandparents born in the United States is actually higher than for white U.S. women. These data suggest that the large international differences in cancer incidence rates are caused by environment and lifestyle differences between countries, rather than ethnic-specific differences in genetic susceptibility. Substantial changes in cancer incidence within countries over time, such as the rise in U.S. lung cancer incidence in the last 50 years (while stomach cancer incidence has fallen steadily) also suggest changes in the environmental determinants of these diseases. Thus, a high proportion of cancer cases are due to environment and lifestyle, and cancer epidemiologists have tried to establish these causative factors over the last 60 years.

Smoking

Smoking is the major modifiable cause of cancer in many countries, estimated to cause one third of cancer deaths in the United States and 21% of cancer deaths worldwide. While smoking prevalence has fallen in the United States over the past several decades, it has increased in many other countries, including the world's largest, China[3]; the number of cancer deaths from tobacco is likely to be much higher in the 21st century than in the 20th century. Although lung cancer dominates the spectrum of smoking-related cancers, cancers at many other anatomic sites have been convincingly related to smoking, including those of the oropharynx, larynx, esophagus, stomach, liver, pancreas, kidney and ureter, cervix, bladder, and colon/rectum, as well as acute myeloid leukemia.[4] Second-hand (or passive) smoking has been associated with lung cancer. Smoking cessation results in rapidly reduced rates of lung cancer; cessation before age 30 reduces lifetime risk by more than 90% compared with continuing tobacco smoking. However, the latency between smoking initiation and cancer occurrence is several decades, so the emergence of large numbers of smoking-related cancers occurs decades after smoking prevalence increases.

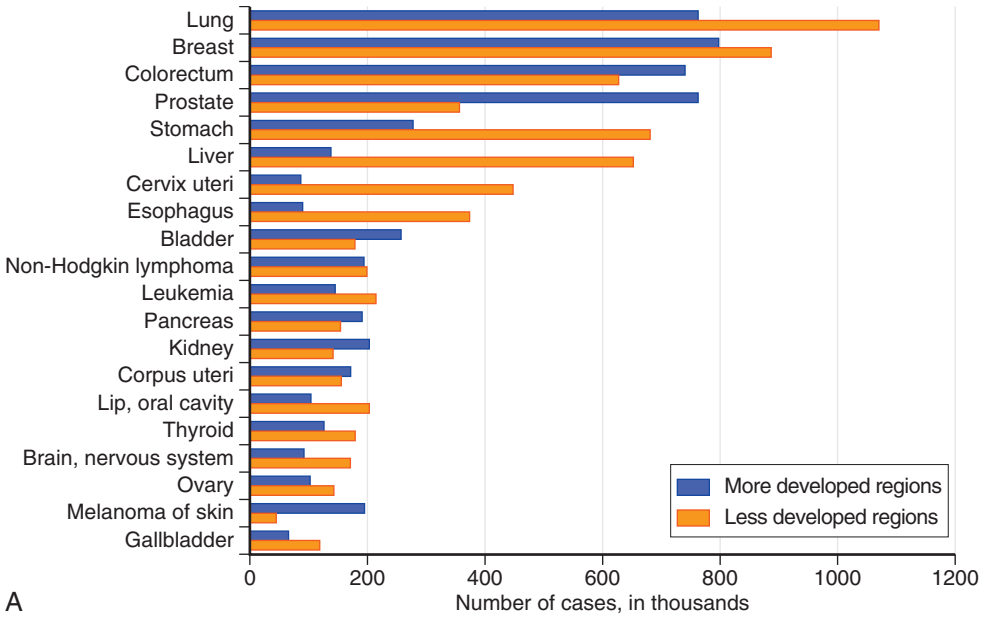

FIGURE 180-2A. Global Annual Cancer Incidence, Both sexes, All ages. (Data from Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr. Accessed February 18, 2015.)

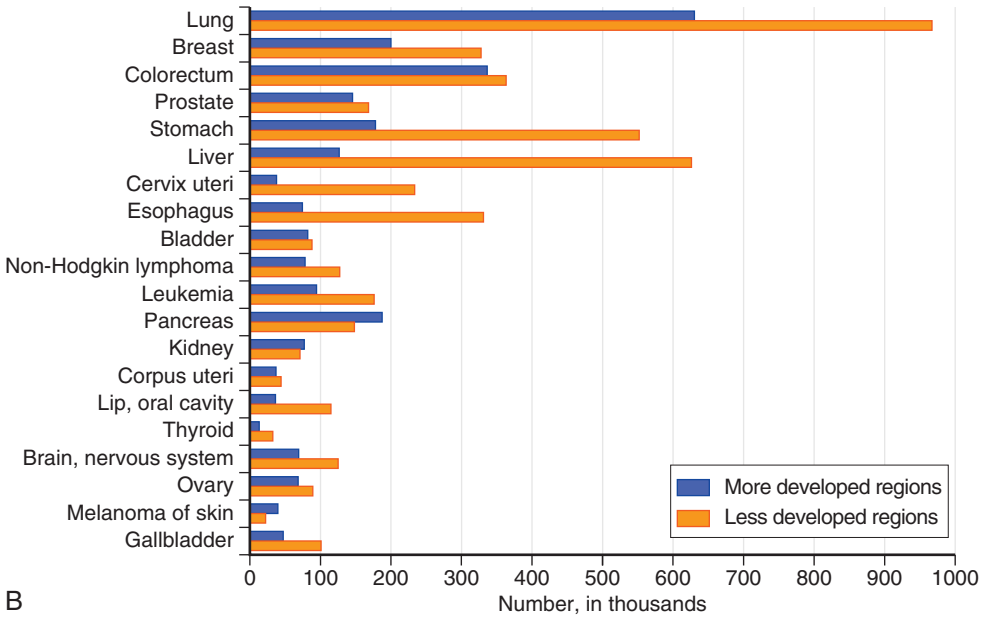

FIGURE 180-2B. Global Annual Cancer Mortality, Both sexes, All ages. (Data from Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr. Accessed February 18, 2015.)

Infections

A substantial fraction of cancers, estimated to be about 16% globally, are caused by infectious agents, particularly in less developed countries (estimated to be ≈ 23%).[5] On the other hand, only 3% of cancers were estimated to be due to infections in the United Kingdom.[6] The major infectious agents are *Helicobacter pylori* (stomach cancer), human papillomavirus (HPV [cervical cancer]), and hepatitis B and C viruses (liver cancer).

Diet

Perhaps no field of epidemiology has been as complex and controversial as the relationship between diet and cancer. Authoritative sources have estimated that a large fraction of cancer incidence is associated with dietary factors; however, a definitive understanding of the mechanisms involved has been difficult, mainly because of difficulties in obtaining valid estimates of intake of specific foods and nutrients and dietary patterns. Foods or nutrients associated with one type of cancer may not be associated with other types of cancer; it is also difficult to distinguish between true etiologic differences between cancer at different sites and false-positive results for a specific cancer type. The World Cancer Research Fund has issued two large-scale summaries of the evidence, most recently in 2007,[7] and the critical recommendations on diet were to: (1) limit consumption of energy-dense foods (avoid sugary drinks); (2) eat mostly foods of plant origin; (3) limit intake of red meat, and avoid processed meat; and (4) limit consumption of salt, and avoid moldy cereals (grains) and legumes.

Part of the difficulty in making generalizations about the percent of cancer attributable to diet is the relationship between diet and body size, and whether to "count" this influence as a "dietary" or "anthropometric" factor. Well-nourished, more affluent societies tend to have taller and heavier populations, both factors positively related to risk of cancer at several sites.

Body Weight and Height

The role of overweight and obesity as risk factors for a wide variety of cancers has become more apparent in the last decade. Even this relationship can be

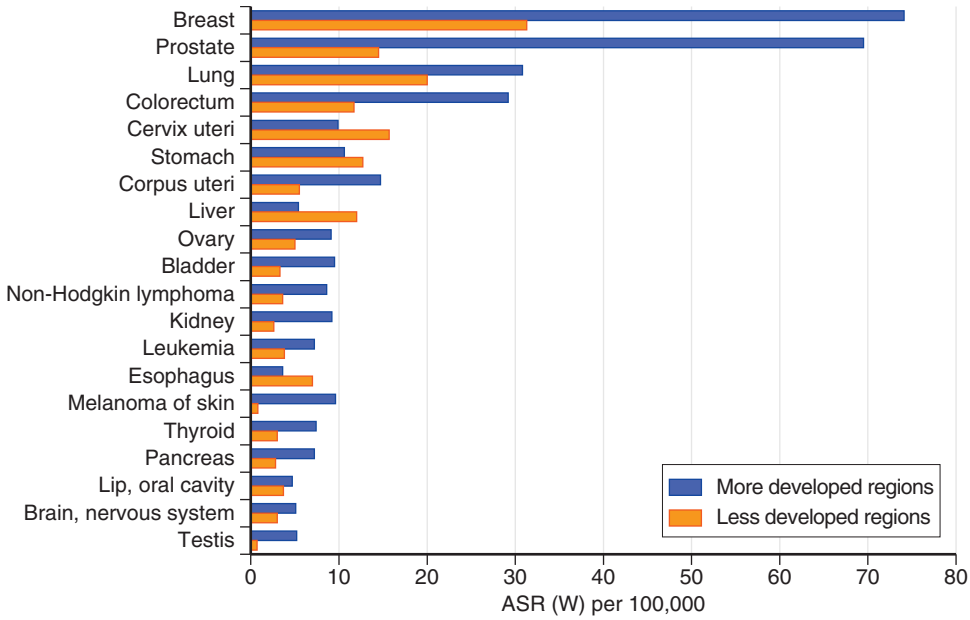

FIGURE 180-3. Age-standardized Rate of Cancer Incidence, Both Sexes, All ages. (Data from Bray F, Ren JS, Masuyer E, Ferlay J. Estimates of global cancer prevalence for 27 sites in the adult population in 2008. Int J Cancer. 2013 Mar 1;132(5):1133-45. doi: 10.1002/ijc.27711. Epub 2012 Jul 26. Available from: http://globocan.iarc.fr. Accessed February 18, 2015.)

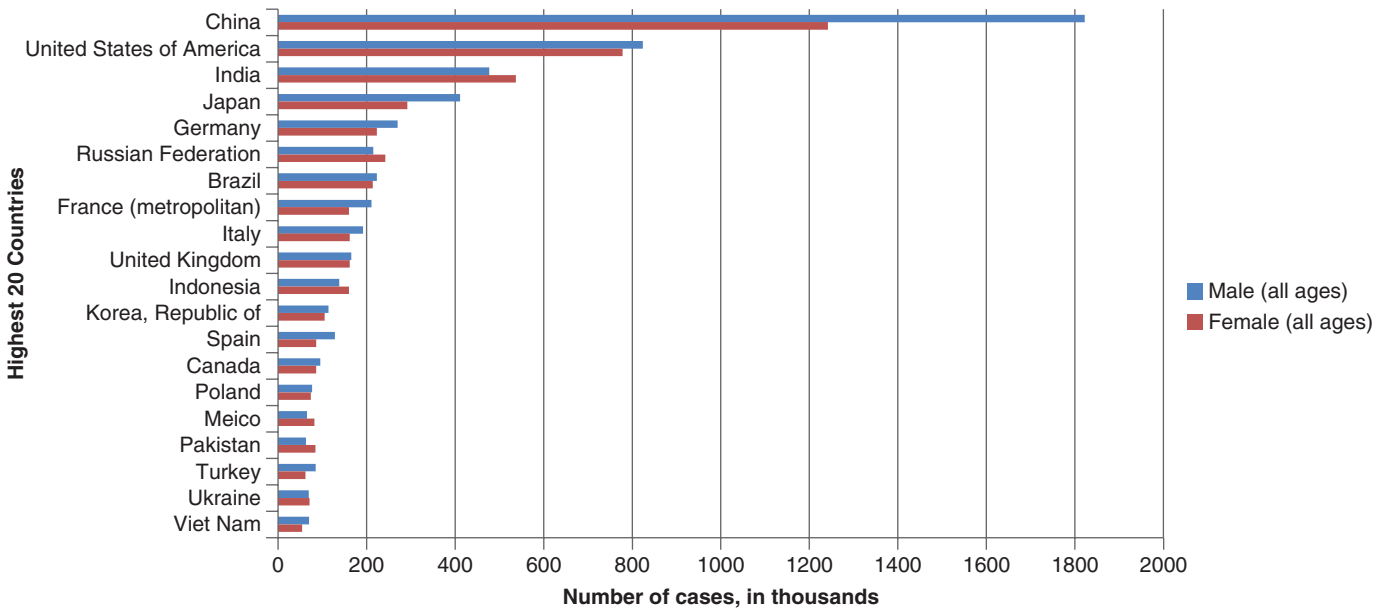

FIGURE 180-4. Annual cancer incidence (excluding non-melanoma skin cancer) by number of cases, highest 20 countries. (Data from Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr. Accessed February 18, 2015.)

complex. For example, obesity is inversely related to the incidence of premenopausal breast cancer but positively related to postmenopausal breast cancer. Greater height is associated with modest increases in the risk of cancer at many sites. The secular trends of increasing height and weight in more affluent societies explain a substantial fraction of the increases in cancer rates over time.

Physical Inactivity

An independent role for physical activity in cancer causation is difficult to dissect from the fact that sedentary lifestyles are associated with overweight and obesity. Physical activity has occupational, recreational, and activities of daily living components and can be difficult to measure in epidemiologic studies. A majority of studies suggest that higher levels of physical activity are associated with lower colorectal cancer risk, independent of body weight, and many studies have also reported an inverse relationship with breast cancer. Whether or not this is due to uncontrolled confounding by body weight, a more active lifestyle is a key component of prevention of weight gain, and thus is recommended for prevention of at least these two types of cancer.

Alcohol

Consumption of alcohol is associated with cancer at several sites, notably cancers of the mouth, pharynx and larynx, esophagus, liver, breast, and rectum. Globally, about 5% of cancers were estimated to be due to alcohol, and for the United Kingdom the estimate is 4%. For breast cancer, the risk appears to increase linearly with increasing alcohol consumption, whereas for the aerodigestive cancers, such as esophageal cancer, the risk is most apparent for heavy drinking. Most authorities recommend no more than one alcoholic drink per day for women, and two for men.

Ionizing Radiation

Ionizing radiation is a clear cause of leukemia and thyroid cancer; however, minimization of exposure means that it is a relatively infrequent cause of cancer, estimated to account for about 2% of cancers in Western societies.

Ultraviolet Radiation

Ultraviolet radiation is the major cause of non-melanoma skin cancers and melanoma, and is estimated to account for 3 to 4% of cancers in the Western world. Avoidance of sunburns in early life may be particularly important for reducing the risk of melanoma.

Occupational Exposures

Starting with the work of Sir Perceval Potts on scrotal cancer in chimney sweeps, it has been appreciated that cancer at certain sites is more frequent in specific occupations. A wide variety of industrial chemicals, more common in certain occupations, are associated with risk of specific cancers. In advanced economies, efforts to minimize exposure usually ensure that risks are minimal, although it was estimated that almost 4% of cancers in the United Kingdom are still due to occupational exposures. In less developed economies, risks may be higher, owing to migration of "dirty" industries to less regulated environments, and a lower level of appreciation of the risks and need to protect workers from carcinogenic exposures.

Exogenous Hormones

Use of oral contraceptives increases the relative risk of breast cancer by about 30% while women are taking them; however, since most women taking oral contraceptives are in the 15- to 40-year age group, when breast cancer incidence rates are low, the absolute number of additional cases is small.[8] Ten or more years of use of oral contraceptives reduces ovarian cancer risk by more than 50%, as well as reducing risk of endometrial cancer.[9]

Use of postmenopausal hormones, particularly those combining estrogen and progestins, increases risk of breast cancer, although risk declines quickly after cessation[10] and decreases risk of colon cancer.[11] After the publication of the Women's Health Initiative findings of a positive association, use of postmenopausal hormones declined dramatically, and the incidence rate of breast cancer (mostly estrogen receptor positive) declined over the next few years, one of the few examples in which a rapid change in prevalence of a risk factor can be linked to a short-term change in cancer incidence.

Reproductive Factors

A variety of reproductive factors have been associated with risk of cancer in women. These include early age at menarche, late age at first birth, nulliparity or low parity, and late age at menopause, short duration of lactation (all increase risk of breast cancer), and nulliparity or low parity (increased risk of ovarian and endometrial cancer).

PREVENTION

Definitions of what risk factors are modifiable vary, but consensus estimates in the 1980s, 1990s, and the 2000s[12] agree that the majority of cancers are theoretically preventable by various combinations of risk factor reduction, and immunization against cancer-causing infectious agents.

Primary Prevention

Primary prevention refers to the strategy of reducing risk factor prevalence, and thus reducing the incidence of cancer. Reducing the prevalence of smoking and the use of other tobacco products is the single factor with the greatest potential for reducing cancer risk. Unfortunately, while per capita consumption of cigarettes has approximately halved in several high-income countries since the 1970s, global consumption is still increasing, fueled by younger age population structures in less developed countries, as well as the tobacco industry's marketing in countries where cigarette consumption was previously low.

Next to smoking, there is consensus that a diet that minimizes weight gain through the course of life is the next major potentially preventive factor. Although the precise content of the diet that is preventive remains unclear, consensus panels stress diets high in fruits and vegetables and low in meat and processed meat products. Minimal to moderate alcohol exposure reduces cancer risk.

Avoidance of sunburns and long periods of ultraviolet exposure will reduce risk of skin cancers. Hepatitis B and HPV vaccination are underused strategies globally; clinical interventions against *H. pylori* would probably reduce risk of stomach cancer, but this is unproven. Occupational health regulations and enforcement are needed to reduce risk of exposure to ionizing radiation and workplace carcinogens. Reduction in use of postmenopausal hormones has already been shown to reduce rates of estrogen-positive breast cancer. Reproductive factors such as age at menarche, parity, and age at first birth, although theoretically modifiable, are difficult to change (e.g., extreme physical activity delays menarche) or are socially determined in a way that tends to supersede concerns about future cancer risk.

Secondary Prevention

Secondary prevention aims to prevent cancer death in those diagnosed with cancer or a premalignant lesion, usually by treating at an early stage. The major success in cancer screening has been organized cervical cancer screening, which is responsible for dramatic reduction in cervical cancer deaths in countries such as the United States, where most cases occur in women with a history of suboptimal screening.[13] A decline in U.S. colorectal cancer incidence is at least partly due to screening for early cancers and premalignant colorectal adenomas, with fecal occult blood testing, sigmoidoscopy, and colonoscopy having different tradeoffs in terms of patient acceptance and test performance. The use of PSA screening for prostate cancer and mammography for breast cancer remains controversial. In both cases, there is plausibly a reduction in the death rate from each cancer in men or women screened, respectively. However, better appreciation of the harms of screening and "overdiagnosis" (i.e., detection and treatment of lesions that might never have come to clinical attention) has meant that recommendations have been changed in recent years. In the case of breast cancer, better treatment of the disease and better awareness by women of the need to seek early assessment of any lumps or changes in the breast may mean that the effectiveness of mammography in reducing death rates has changed over time.[14] A report for the U.S. Preventive Services Taskforce describes the evidence as "strong" that low-dose computed tomography screening reduces risk of lung cancer and lung cancer death, although the evidence is based on a single, large, good-quality study.[15]

Although many cancers can be cured if detected early enough, the sheer volume of cancers in less developed countries with limited infrastructure, plus the cost of frequent screening, means that organized screening activities will take time to develop. The exception may be cervical screening, for which HPV testing combined with gynecologic follow-up may provide a means of screening large numbers of women relatively rapidly, although this is by no means a trivial endeavor. Greater awareness of cancer symptoms and signs, combined with appropriate and prompt follow-up, will remain a mainstay of attempts to downstage cancers in less developed regions, and developing adequate diagnostic and treatment services may be required before instituting screening services.

TREATMENT

Rx

Once cancer is diagnosed, appropriate treatment is not available to many in less developed countries, resulting in substantially higher case-fatality rates. Many of the most efficacious chemotherapy drugs are off-patent and relatively inexpensive; other modalities such as radiation can be offered at low cost per patient once a facility has been built and staff trained to use it. It is estimated that 80% of patients in the world have no access to opiates at the end of life,[16] a situation that could be remedied at low cost if international treaties that limit opiate export, as well as concerns about their abuse, could be addressed.

FUTURE DIRECTIONS

As for other non-communicable diseases, a broad range of actions need to be taken to prevent, detect, and treat cancers as incidence rates of many cancers increase and absolute rates increase faster because of the aging of populations around the world. Fortunately, several non-communicable diseases share similar risk factors,[17] particularly tobacco use and sedentary lifestyles/obesity; hence, efforts to limit the spread of these risk factors may be important in the control of cardiovascular disease, diabetes, and respiratory diseases, as well as cancer.[18] Access to even low-cost treatments requires major investments in health systems, but without these investments the human and economic costs of cancer will be greater still.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ferlay J, Soerjomataram I, Ervik M, et al. *GLOBOCAN 2012 v1.0,* Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 *[Internet]*. Lyon, France: International Agency for Research on Cancer; 2013 Available from: http://globocan.iarc.fr; Accessed February 2, 2015.
2. American Cancer Society. Cancer Facts & Figures 2014. Atlanta: American Cancer Society; 2014.
3. Jha P, Peto R. Global effects of smoking, of quitting, and of taxing tobacco. *N Engl J Med*. 2014;370:60-68.
4. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The Health Consequences of Smoking-50 Years of Progress: A Report of the Surgeon General. Atlanta: Centers for Disease Control and Prevention (US); 2014.
5. de Martel C, Ferlay J, Franceschi S, et al. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncol*. 2012;13:607-615.
6. Parkin DM, Boyd L, Walker LC. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br J Cancer*. 2011;105:S77-S81.
7. World Cancer Research Fund/American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective. Washington, D.C.: AICR; 2007.
8. Hunter DJ, Colditz GA, Hankinson SE, et al. Oral contraceptive use and breast cancer: a prospective study of young women. *Cancer Epidemiol Biomarkers Prev*. 2010;19:2496-2502.
9. Havrilesky LJ, Gierisch JM, Moorman PG, et al. Oral contraceptive use for the primary prevention of ovarian cancer. *Evid Rep Technol Assess (Full Rep)*. 2013;212:1-514.
10. Narod SA. Hormone replacement therapy and the risk of breast cancer. *Nat Rev Clin Oncol*. 2011;8:669-676.
11. Farguhar C, Marjoribanks J, Lethaby A, et al. Long term hormone therapy for perimenopausal and postmenopausal women. *Cochrane Database Syst Rev*. 2012;7:CD004143.
12. Schottenfeld D, Beebe-Dimmer JL, Buffler PA, et al. Current perspective on the global and United States cancer burden attributable to lifestyle and environmental risk factors. *Annu Rev Public Health*. 2013;34:97-117.
13. U.S. Preventive Services Task Force. Screening for cervical cancer: recommendation statement. *Am Fam Physician*. 2012;86:555-559.
14. Kalager M, Adami HO, Bretthauer M. Too much mammography. *BMJ*. 2014;348:g1403. doi:10.1136/bmj.g1403.
15. Humphrey LL, Deffebach M, Pappas M, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive Services Task Force recommendation. *Ann Intern Med*. 2013;159:411-420.
16. Palliative care: a peaceful, humane global campaign is needed. *Lancet*. 2014;383:487.
17. Norheim OF, Jha P, Admasu K, et al. Avoiding 40% of the premature deaths in each country, 2010-30: review of national mortality trends to help quantify the UN Sustainable Development Goal for health. *Lancet*. 2014;S140-S6736.
18. Beaglehole R, Bonita R, Ezzati M, et al. NCD Countdown 2025: accountability for the 25x25 NCD mortality reduction target. *Lancet*. 2014;384:105-107.

REVIEW QUESTIONS

1. Which of the following types of cancer is not related to smoking?

A. Bladder
B. Esophagus
C. Pancreas
D. Acute lymphocytic leukemia
E. Head and neck

Answer: D Acute lymphocytic leukemia is primarily a disease of children, and even in adults, it appears to be unrelated to smoking.

2. Excessive alcohol intake is associated with which one of the following malignancies?

A. Laryngeal cancer
B. Soft tissue sarcoma
C. Neuroblastoma
D. Melanoma
E. Basal cell skin cancer

Answer: A Cancers of the head and neck, aerodigestive tract, and liver are well known to be associated with excessive alcohol consumption; no such association exists for pediatric malignancies or melanoma or for non-melanoma skin cancers. (See section on "Alcohol" under "Causes of Cancer.")

3. Which of the following malignancies is not associated with exposure to an infectious agent?

A. Cervical cancer
B. Hepatocellular cancer
C. Head and neck cancer
D. Chronic lymphocytic leukemia
E. Gastric cancer

Answer: D Cervical and head and neck cancers are associated with the human papillomavirus, and hepatocellular cancer with chronic hepatitis B and/or C infection. Gastric cancer is associated with *Helicobacter pylori* infection. Chronic lymphocytic leukemia is not associated with exposure to an infectious agent. (See section on "Infections" under "Causes of Cancer.")

4. "Overdiagnosis" in the context of cancer screening efforts can be defined as:

A. The excessive use of noninvasive imaging tests for asymptomatic patients
B. The detection and treatment of lesions that might never have come to clinical attention
C. The use of clinical algorithms for cancer diagnosis that have not been validated
D. Use of plasma biomarkers of cancer to justify initiation of systemic treatment

Answer: B Treating tumors that are of no clinical significance (e.g., asymptomatic prostate cancers in men > 80 years of age) can produce morbidity with no therapeutic benefit.

5. The factor that is not related to cancer incidence is:

A. age
B. physical inactivity
C. diet
D. exposure to cold weather

Answer: D Although excessive sun (and concomitant UV) exposure is a well-known risk for the development of melanoma, no such relationship exists for exposure to colder climates. The rates of most cancers increase with age, often in a nonlinear (exponential) fashion. Physical inactivity is indirectly (if not directly) linked to cancer causation by its association with obesity. A large fraction of cancer incidence is acknowledged to be associated with dietary factors, although a definitive understanding of the mechanisms involved has been challenging. (See section on "Causes of Cancer.")

181 CANCER BIOLOGY AND GENETICS

ADRIAN R. BLACK AND KENNETH H. COWAN

CANCER AS A GENETIC DISEASE

Cancer Development Is Driven by Random Mutations

Cancer develops from mutations in genes that regulate normal cellular processes; thus, cancer can be thought of as a genetic disease. Mutations that lead to cancer involve two types of genes: mutant genes that enhance the development of cancer are known as oncogenes, whereas genes that inhibit tumorigenesis are known as tumor suppressor genes. Alterations in the function of oncogenes and tumor suppressors allow cells to escape the normal controls that regulate tissue homeostasis, leading to the outgrowth of mutant cells and the subsequent development of cancer. Except in rare cases, such as seen with the *BCR-ABL* fusion gene in chronic myelogenous leukemia (CML), a single mutation is not sufficient to lead to cancer. Instead, cancer develops in a multistep process that requires accumulation of mutations that drive progression of normal tissues through benign precancerous lesions to aggressive malignant disease.[1]

The acquisition of these mutations is random, and tumor evolution follows a model of natural selection where mutations that bestow a growth and/or survival advantage allow for clonal expansion of the population of cells that carry the mutation. During this process, many mutations occur that are disadvantageous, and cells that carry these mutations die out. The timing of mutations is also a factor in the evolution of cancer. Some mutations that help drive cancer development are deleterious to normal cells but can provide a growth/survival advantage in cells in which other mutations have occurred. For example, activation of oncogenes in normal cells can lead to cell death through apoptosis (see later) or induce a permanent growth arrest known as senescence, and their oncogenic activity is only manifested if mutations that negate these effects have already occurred in the cell. The need for mutations to occur in the correct order represents an important brake on cancer progression. The sequential multistep nature of cancer is reflected in the preferential association of specific mutations with different stages of tumor progression, as has been elegantly demonstrated for the development of colon cancer (Fig. 181-1).[2]

Mutation Rate and Cancer Risk Factors

Because mutations occur randomly within the genome, it is impossible to predict who will get cancer or to know definitively why it developed in any particular patient. However, factors that increase the rate of mutation increase the risk of tumorigenesis. For example, exposure to environmental chemicals that damage DNA (mutagens) has long been recognized as a major risk for developing cancer, as exemplified by the association of exposure to tobacco smoke—which contains multiple mutagenic compounds—with lung, oral, and other cancers. Similarly, ionizing radiation, which causes DNA strand breaks and crosslinks, is also closely associated with cancer risk. This is seen in the high risk of leukemia in survivors of the nuclear bombs in Hiroshima and Nagasaki and in the association of exposure to ultraviolet radiation in sunlight and on tanning beds with melanoma and other skin cancers. Physiologically, increased cell proliferation also increases the risk of cancer because it enhances the rate at which mutations from DNA damage become fixed in the genome during replication, increases replication-associated mutations and chromosomal rearrangements, and accelerates expansion of the pool of mutated cells. This effect is seen in the association of endometrial cancer with hormone replacement therapy, tamoxifen therapy, and obesity, which are all mitogenic in the endometrium owing to increases in estrogenic signaling.

Another major risk factor for cancer is chronic inflammation[3], such as that found in the increased risk of colon cancer for patients with inflammatory bowel disease and of stomach cancer for patients harboring *Helicobacter pylori* infections. Inflammatory cells increase the mutation rate by producing reactive oxygen species and other immune effectors that damage DNA, as well as through mitogenic cytokines that can directly increase tumor cell proliferation. Because mutations accumulate over time, increasing age itself is a major risk factor for most cancer types.

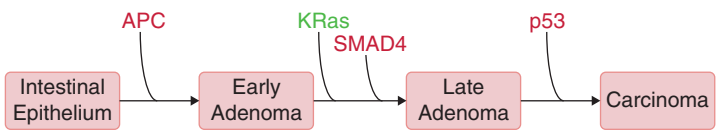

FIGURE 181-1. Genetic changes associated with different stages of colon cancer development. Scheme for accumulation of genetic changes in colon cancer in familial adenomatous polyposis patients, adapted from that originally proposed by Dr. Burt Vogelstein and colleagues at the Johns Hopkins University. Proto-oncogenes/oncogenes are indicated in green; tumor suppressor genes are indicated in red.

Genomic Instability and Cancer Development

Although the genome is continuously being damaged, cells have efficient mechanisms that keep the overall mutation rate low. Mechanisms for repair of replication errors and damaged nucleotides include mismatch repair, base-excision repair, and nucleotide-excision repair, whereas more severe damage and double strand breaks can be repaired by homologous recombination and non-homologous end joining. Mutations that disrupt these mechanisms dramatically increase the rate of mutation. Because cancer development requires ordered mutation of selected genes, efficient DNA repair means that the chances that appropriate mutation will accumulate in a cell to produce cancer within the lifetime of an individual are very low. Thus, it is not surprising that mutations that disrupt DNA repair are almost universally seen in cancer cells. Although these mutations do not directly provide an advantage to cells (and may even decrease survival), they dramatically increase the chance that cells will accumulate appropriate mutations for cancer development.

The contribution of DNA repair defects to cancer development[4] can be seen in the number of inherited cancer susceptibility syndromes that are due to mutations in DNA repair genes (Table 181-1). These mutations are also found in corresponding sporadic cancers. Mutations in mismatch repair genes (e.g., *MLH1*, *MSH2*) that cause the Lynch syndromes (also known as hereditary nonpolyposis colon cancer [HNPCC]) are commonly found in sporadic colon cancer. Mutations in *BRCA1* (involved in homologous recombination repair, non-homologous end joining, and nucleotide excision repair) or *BRCA2* (homologous recombination repair) that lead to familial breast and ovarian cancer are also common in sporadic disease.

DNA repair is essential for cell survival, and mutations in DNA repair genes can thus predispose to "synthetic lethality" in cancer cells. For example, inhibition of poly-ADP ribose polymerase (PARP) induces double strand breaks in cells with *BRCA1* or *BRCA2* mutations, but not in BRCA1/2 wild-type cells. The potential clinical relevance of this finding is supported by results from clinical trials.

Two Types of Cancer Genes

As mentioned, genes whose mutation contributes to the development of cancer are traditionally divided into two groups, oncogenes and tumor suppressor genes. The products of oncogenes, which are generally proteins referred to as oncoproteins, are positive drivers of cancer development. In contrast, tumor suppressor genes help to prevent the development of cancer in normal tissues and often counter the effects of oncogenes. Because mutations in oncogenes are activating, only one copy of the gene needs to be affected for the phenotype to be manifested; thus, oncogene mutations are dominant. The presence of two copies of tumor suppressor genes in the normal diploid genome means that both copies of these genes have to be inactivated to remove tumor suppressive activity. Thus, tumor suppressor mutations are recessive. This need for two "hits" for inactivation of tumor suppressors represents another brake on cancer development.

Familial Cancer Susceptibility Syndromes

Although most tumors are sporadic, there are a number of inherited syndromes that are associated with the risk of cancer. The vast majority of these syndromes are due to a germline mutation in one copy of a tumor suppressor gene (see Table 181-1). Because tumor suppressor genes are recessive, they remain phenotypically silent until the wild-type copy of the gene is mutated. However, the germline mutations provide the first "hit" for the loss of a tumor suppressor and vastly increase the risk of developing cancer. The dominant nature of mutations in oncogenes means that they are generally incompatible with normal development. Thus, although some familial syndromes are associated with oncogenes, these are relatively rare (see Table 181-1). Individuals with family histories of cancers that point to the possibility of an inherited disease should be offered genetic screening and counseling.

TABLE 181-1 HEREDITARY CANCER RISK SYNDROMES

SYNDROME	GENE	ASSOCIATED TUMORS
TUMOR SUPPRESSOR MUTATIONS		
Cowden	*PTEN*	Breast, thyroid, endometrial, colorectal, kidney, melanoma
Familial adenomatous polyposis	*APC*	Colorectal, stomach, intestinal, hepatoblastoma, thyroid, pancreatic, adrenal, bile duct, medulloblastoma
Familial gastric carcinoma	*CDH1* (E-Cadherin)	Stomach, breast
Familial malignant melanoma	*CDKN2* ($p16^{INK4A}$ and $p14^{ARF}$), *CDK4*	Melanoma, pancreatic
Familial Wilms tumor	*WT1*	Wilms (kidney)
Hereditary retinoblastoma	*RB1* (pRB)	Retinoblastoma
Hereditary paraganglioma	*SDHA, SDHB, SDHC, SDHD* or *SDHAF2* (succinate dehydrogenase)	Paraganglioma, pheochromocytoma
Li-Fraumeni	*TP53* (p53)	Osteosarcoma, soft tissue sarcoma, leukemia, breast, brain, adrenal, melanoma, Wilms, stomach, colorectal, pancreatic, esophageal, lung, gonadal germ cells
Multiple endocrine neoplasia type 1	*MEN1*	Parathyroid, pituitary, islet cell, adrenal, carcinoid
Neurofibromatosis type 1	*NF1*	Neurofibroma, glioma, leukemia, breast, rhabdomyosarcoma, gastrointestinal stromal
Neurofibromatosis type 2	*NF2*	Neuromas, schwannomas, meningiomas, astrocytomas, gliomas
Peutz-Jeghers syndrome	*LKB1*	Colon, intestine, stomach, pancreas, cervix, ovary, testis, breast, thyroid
Von Hippel-Lindau	*VHL*	Kidney, adrenal, pheochromocytoma
DNA REPAIR GENE MUTATIONS		
Ataxia telangiectasia	*ATM*	Leukemia, lymphoma
Bloom	*BLM*	Leukemia, multiple solid tumors
Fanconi anemia	*FANCA, FANCC, FANCD2, FANCE, FANCF or FANCG*	Leukemia, multiple solid tumors
Hereditary breast and ovarian cancer	*BRCA1*or *BRCA2*	Breast, ovary, pancreatic, fallopian tube
Lynch (hereditary nonpolyposis colon cancer)	*MLH1, MSH2, MSH6 or PMS2*	Colorectal, endometrial, stomach, breast, ovarian, intestinal, pancreatic, prostate, urinary tract, liver, kidney
MYH-associated polyposis	*MUTYH*	Colorectal, intestinal
Xeroderma pigmentosum	*XPA, XPC, ERCC2, ERCC3, ERCC4, ERCC5,* or *DDB2*	Skin (basal cells, squamous), melanoma, tongue, eye
ONCOGENE MUTATIONS		
Costello	*HRAS*	Papilloma, rhabdomyosarcoma, neuroblastoma
Noonan	*PTPN11, SOS1, KRAS, NRAS, RAF1 or BRAF* (Ras-Erk pathway regulators)	Leukemia, rhabdomyosarcoma, neuroblastoma
Familial malignant melanoma	*CDK4*	Melanoma
Multiple endocrine neoplasia type 2	*RET*	Thyroid, parathyroid, pheochromocytoma

Oncogenes and Proto-oncogenes

Oncogenes are activated forms of normal cellular genes that are termed *proto-oncogenes*. The products of proto-oncogenes are involved in the regulation of multiple aspects of physiology, including proliferation, survival, and cell migration, and play an important role in development and tissue homeostasis. Mutations that activate oncogenes lead to loss of the normal control that governs the activity of proto-oncogenes and gives a growth or survival advantage to the cell. A single oncogene can affect multiple aspects of tumor progression and in rare cases may be all that is needed to support oncogenic transformation (e.g., the *BCR-ABL* oncogene in CML). In keeping with their diverse roles, proto-oncogenes perform diverse functions in the cell. They commonly function as signal transduction molecules, acting as extracellular ligands/growth factors (e.g., KGF, SIS, INT-2, WNT1), growth factor receptors (e.g., EGFR, HER2, FGFR), and signaling intermediates (e.g., the PI-3 kinase [PI-3K] catalytic subunit *PIK3CA*, PKCι, B-Raf, and the Ras proteins K-Ras, N-Ras, and H-Ras). They can also be involved in direct regulation of gene expression, acting as transcription factors (e.g., myc, fos, jun, myb), translation factors (e.g., eIF4E), or regulators of protein degradation (e.g., MDM2, Cbl). Other common roles are in promoting cell survival (e.g., Bcl2, MCL1) and cell cycle progression (e.g., cyclin D1, CDK4, cyclin E).

Oncogenes can be activated by multiple mechanisms, and a single oncogene can be activated by different mechanisms. Activation can result from mutations that lead to a qualitative change that affects the activity or regulation of the encoded oncoprotein, or by a quantitative effect that increases the levels of an oncoprotein in the cell (through increased gene expression or reduced protein turnover). Qualitative effects are generally due to point mutations that lead to loss of negative regulatory domains or, more commonly, to single amino acid changes that affect the activity or interactions of an oncoprotein. These mutations can (1) directly increase kinase activity (e.g., *B-Raf* mutations in melanoma and colon cancer, or *PIK3CA* mutations in colorectal, brain, and gastric cancers), (2) decrease negative regulation (e.g., C-terminal truncation of Src), (3) alter autoregulation (e.g., K-Ras, H-Ras, and N-Ras, seen in 50% of colon cancers), and (4) reduce protein degradation (e.g., phosphorylation site mutations of β-catenin, seen in colon cancer). These mutations often occur at a single site or limited region of the gene, so-called hot spots that localize to important functional regions of the protein encoded by the oncogene. For example, oncogenic mutations of *B-Raf* are commonly missense mutations of the valine residue in its activation loop (V600).

In addition to point mutations, gross chromosomal abnormalities can lead to activation of oncogenes. Increased gene copy number through gene amplification is a mechanism by which oncogenes become overexpressed. Gene amplification is extremely common in cancer and represents a major mechanism by which oncogenes are activated. Amplifications are detected histologically by karyotypic abnormalities such as double-minute chromosomes (DMs) and homogeneous staining regions (HSRs). Chromosomal translocations can also lead to overexpression of oncogenes by locating the gene next to a strong transcriptional element. For example, the t(8;14) translocation in Burkitt lymphoma transcriptionally activates the *MYC* oncogene by placing it in the immunoglobulin heavy chain locus. In many cases, translocations not only lead to increased expression of an oncoprotein but can also alter the

protein itself. The t(7;9) translocation in T-cell acute lymphoblastic leukemia (ALL) places the *NOTCH1* gene (*TAN1*) next to the T-cell receptor promoter. However, the translocation is only partial and leads to overexpression of the transcriptionally active NOTCH1-IC fragment, and thus constitutively activates Notch signaling. Translocations can also lead to the production of fusion proteins that have distinct properties from the original proto-oncogene. For example, the t(9;22) translocation seen in CML fuses the B-cell receptor gene with *Abl* tyrosine kinase genes. The resultant BCR-ABL fusion protein is not only overexpressed and a constitutively active kinase, it is also aberrantly localized to the cytoplasm and thus phosphorylates different substrates from Abl. Although oncogenic chromosomal translocations are particularly prevalent in leukemias and lymphomas, they are also seen in solid tumors such as Ewing sarcoma, papillary thyroid cancer, and rhabdomyosarcoma (producing the *EWS-FLI*, *RET-PCP*, and *PAX3-FKHR* gene fusions, respectively).

Viral Infection and Cancer

Viral infection has been associated with up to 20% of cancers worldwide and is thus a major concern in oncology (Chapter 180). Known oncoviruses include: human papillomaviruses HPV-16 and HPV-18 (the main causative factors for cervical cancer), hepatitis B and hepatitis C viruses (linked to hepatocellular carcinoma), Epstein-Barr virus (linked to lymphomas and nasopharyngeal carcinoma), HTLV-I (linked to T-cell leukemia), Kaposi sarcoma herpesvirus (Kaposi sarcoma), and Merkel cell polyomavirus (Merkel cell carcinoma). Viral reproduction requires replication of viral DNA in the host cell. Oncoviruses achieve this by enlisting the host replicative machinery through expression of oncogenes (e.g., E6 and E7 of human papillomavirus and LMP1 and BARF1 of Epstein-Barr virus), which results in an increased risk of cancer in the infected tissue.

Tumor Suppressors

Whereas the term *oncogene* refers to a mutated version of a proto-oncogene, tumor suppressor genes are normal cellular genes whose inactivation contributes to the development of cancer. Although mutations of tumor suppressors are almost always recessive, occasionally mutation in a tumor suppressor gene can give rise to a "dominant negative" protein that can block the activity of the wild-type protein; such mutated forms of the tumor suppressor act as an oncogene (e.g., some c-cbl and p53 mutations).

As with proto-oncogenes, tumor suppressors are involved in multiple aspects of cellular physiology. They can be cell surface receptors and downstream mediators of growth inhibitory signaling [e.g., TGFβ receptor II (*TGFBR2*), SMAD4 (*DPC4*)], negative regulators of mitogenic signaling [e.g., APC, NF1, PTEN], negative regulators of the cell cycle [e.g., the retinoblastoma protein (pRB, *RB1*), $p16^{INK4a}$ (*CDKN2A*), p53 (*TP53*)], promoters of apoptosis (p53, *RASSF1A*), regulators of cell adhesion [e.g., E-cadherin (*CDH1*)], transcription factors [e.g., p53, pRB, Wilms tumor suppressor 1 (*WT1*)] or regulators of protein degradation (e.g., Von Hippel-Lindau protein (*VHL*)]. Genes involved in DNA damage repair and maintenance of genome stability are also considered tumor suppressors (see Table 181-1). As discussed earlier, although DNA repair defects do not directly contribute to the transformed phenotype, they greatly increase the chances that cancer will develop.

Inactivation of tumor suppressors can occur through point mutations in the gene. These can be missense mutations that lead to alterations in amino acid sequences (hot spots) that are critical for the activity of a tumor suppressive protein (e.g., Trp248 and His273 mutations in p53) or nonsense mutations that lead to a truncated protein (e.g., APC truncations in colon cancer). Tumor suppressor genes can also be inactivated by chromosomal deletion or rearrangement (e.g., inactivation of PML by the t[15;17] translocation in acute promyelocytic leukemia, which results in the RAR-PML fusion protein).

In addition to genetic changes, tumor suppressor genes can be inactivated epigenetically.[5] DNA methylation of the promoter region of genes, which results in transcriptional repression, is a normal regulatory mechanism that ensures correct gene expression and stable silencing during development. Aberrant promoter methylation of tumor suppressor genes is very common in cancer and represents a major mechanism by which these genes are inactivated. Reversal of this transcriptional repression allows for reexpression of the tumor promoters and forms the rationale for the use of drugs that inhibit DNA methylation, such as decitabine.

Once one copy of a tumor suppressor gene is disrupted, the most common mechanism for inactivation of the remaining wild-type copy is through loss of heterozygosity (LOH). LOH can occur when the wild-type gene is lost through deletion, chromosomal loss, or unequal recombination, which results in a single mutant copy remaining in the cell. Alternatively, copy neutral LOH can occur, in which the wild-type gene is replaced by the mutant gene through loss and reduplication or mitotic recombination.

miRNAs as Oncogenes and Tumor Suppressors

Although most of our knowledge of tumor suppressors and oncogenes is concentrated on genes that encode for protein, an important role for non-coding RNAs is rapidly emerging.[6] microRNAs (miRNAs) regulate the expression of other genes by binding to their mRNAs in a sequence specific manner and causing their degradation or inhibiting their translation. miRNAs that inhibit the expression of oncogenes act as tumor suppressors, whereas those that inhibit the expression of tumor suppressors have oncogenic properties and are known as oncomirs. Recent research is identifying miRNA profiles that are associated with particular cancer types, underlining the importance of these non-coding RNAs in tumor development.

Activities of Oncogenes and Tumor Suppressors Are Interrelated

Oncogenes and tumor suppressors are commonly part of signal transduction pathways (Fig. 181-2). A pathway that is activated in many cancers is the Ras/Raf/Erk signaling cascade. In addition to the Ras and Raf proteins (e.g., K-Ras, B-Raf), growth factor receptors that activate Ras (e.g., EGFR, HER2, PGDF, KIT) and downstream targets of the pathway (e.g., the transcription factors Fos and Jun) also act as oncoproteins to activate this pathway. In contrast, the NF1 tumor suppressor (see Table 181-1) negatively regulates the pathway by inactivating Ras. Another pathway that is activated in cancer is the PI-3K/AKT/mTOR pathway, in which growth factor receptors activate the Akt protein kinase through PI-3K-mediated phosphorylation of phosphatidylinositol (see Fig. 181-2). Although both PI-3K (PIK3A) and Akt can act as oncogenes in this pathway, the PTEN tumor suppressor negatively regulates the pathway by dephosphorylating phosphatidylinositol-(3,4,5) triphosphate. Thus, in addition to activation by direct mutation, proto-oncogenes can be activated indirectly through mutation of upstream oncogenes or tumor suppressors. Because only one oncogene or tumor suppressor needs to be mutated to activate a pathway, it is rare that two mutations are seen in the same pathway. An example can be seen in the APC/β-catenin pathway: although mutations in either APC or β-catenin are commonly found in colon cancer, mutation of these genes is mutually exclusive.

Signal transduction pathways in a cell do not act simply as linear pathways but participate in interlinked networks. Oncogenes and tumor suppressors

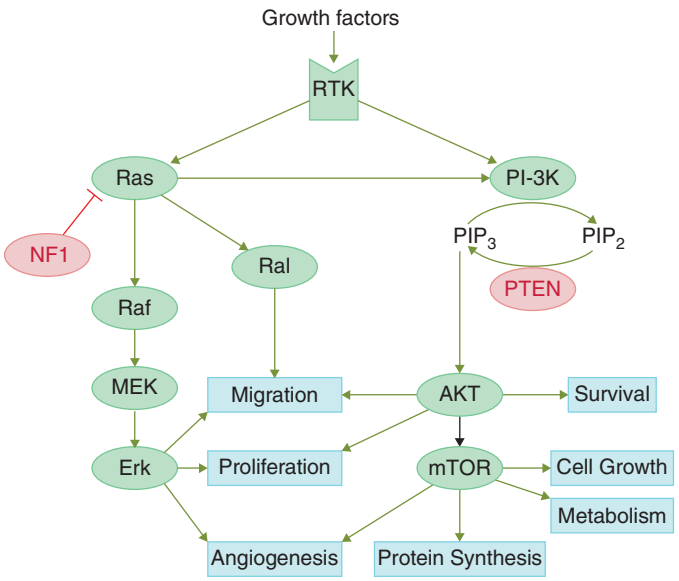

FIGURE 181-2. Ras/Erk and PI-3K/Akt pathways. Schematic diagram of the relationship between critical components of the Ras/Erk and PI-3K/Akt pathways are shown, along with downstream effects of pathway activation. Proto-oncogenes/oncogenes are indicated in green; tumor suppressors are indicated in red. Green arrows indicate activation, and the red T indicates inhibition. NOTE: mutations that affect oncogenes or tumor suppressors at any point along a signaling pathway can stimulate the pathway and have similar effects on downstream targets. RTK = receptor tyrosine kinase (e.g., EGFR, HER2, PDGFR). For details see section on "Activities of Oncogenes and Tumor Suppressors Are Interrelated"

are involved in this signaling crosstalk. For example, Ras activates both the Raf-Erk and the PI-3K/Akt pathways, and Ras-Erk signaling inactivates the *LKB1* tumor suppressor that is a negative regulator of AMPK signaling (see Table 181-1).

Oncogenes and Tumor Suppressors Are Not Restricted by Disease Site

Certain oncogenes and tumor suppressors are associated with particular cancer types, as can be seen in many familial syndromes. This often reflects the roles of particular signaling pathways in the normal tissue from which the cancer arises. For example, in keeping with the critical role of the Wnt/APC/β-catenin signaling pathway in regulating proliferation and stem cell maintenance in the normal intestinal epithelium, activation of this pathway through loss of *APC* or activation of β-catenin is found in the vast majority of colon cancers. Because the same signaling pathways can be important in multiple tissues, cancers in different disease sites can have mutations in the same oncogenes and tumor suppressors. For example, in addition to colon cancer, disruption of the Wnt/APC/β-catenin pathway is found in other cancer types including those arising in the pancreas, breast, and thyroid. This lack of disease site specificity for tumor suppressors can clearly be seen in the array of cancers that are associated with familial cancer susceptibility syndromes (see Table 181-1).

There are two pathways that appear to be of universal importance and are disrupted in most, if not all, cancers. These involve the tumor suppressor genes *RB1*, encoding the retinoblastoma protein (pRB), and *TP53*, encoding p53 (Fig. 181-3). pRB was the first tumor suppressor to be identified and is responsible for hereditary retinoblastoma. It is a central player in a pathway that senses the extracellular growth environment and controls whether or not a cell will proliferate. Disruption of this pathway in cancer cells thus divorces proliferation from the normal controls that maintain tissue homeostasis. pRB inhibits proliferation by repressing transcription of growth-related genes through binding to E2F transcription factors. The activity of pRB is regulated by cyclin-dependent kinases (cdks). When bound to cyclins, cdks phosphorylate pRB, and this blocks its ability to bind E2Fs and allows for transcription of E2F-regulated genes, many of which are proto-oncogenes. Expression of D-type cyclins (e.g., cyclin D1), regulatory molecules whose binding to cdks initiates the phosphorylation of pRB, is regulated by extracellular signaling and thus provides a link between cell proliferation and the extracellular growth environment. Activity of this pathway is negatively regulated by a number of cdk inhibitory proteins, including $p21^{CIP1}$, $p27^{KIP1}$, and $p16^{INK4A}$. Although mutations in pRB itself are relatively rare, mutations in multiple members of the pathway that mediate inactivation of pRB are a common feature of cancer (Table 181-2). In addition, oncogenes and tumor suppressors feed into this pathway through regulation of cyclins and cdk inhibitory protein.

p53 is the most commonly mutated tumor suppressor gene, with mutations seen in about 50% of all human cancers. Germline mutations in p53 are responsible for Li-Fraumeni syndrome (see Table 181-1). p53 is known as "the guardian of the genome" because it is a central player in the system that assesses DNA damage and the internal fitness of the cell to divide. DNA damage leads to transcriptional activation p53, which has two major effects dependent on the severity of the damage. If DNA damage is not too severe, p53 can induce cell cycle arrest by increasing expression of the cdk inhibitor $p21^{CIP1}$, which activates the pRB growth inhibitory pathway. This allows time for DNA to be repaired prior to the initiation of replication. With more severe DNA damage, p53 induces apoptosis through upregulation of proapoptotic proteins such as PUMA and Bax. In addition to p53 mutations, activation of oncogenes such as *MDM2* or loss of tumor suppressors like $p14^{ARF}$ can disrupt p53 activity in the cancer cell (see Fig. 181-3).

The importance of the p53 and the pRB-E2F pathway to cancer development can be seen in the targeting of these pathways by DNA tumor viruses. For example, the E6 and E7 oncogenes of human papillomavirus target p53 for degradation and prevent pRB binding to E2F, respectively.

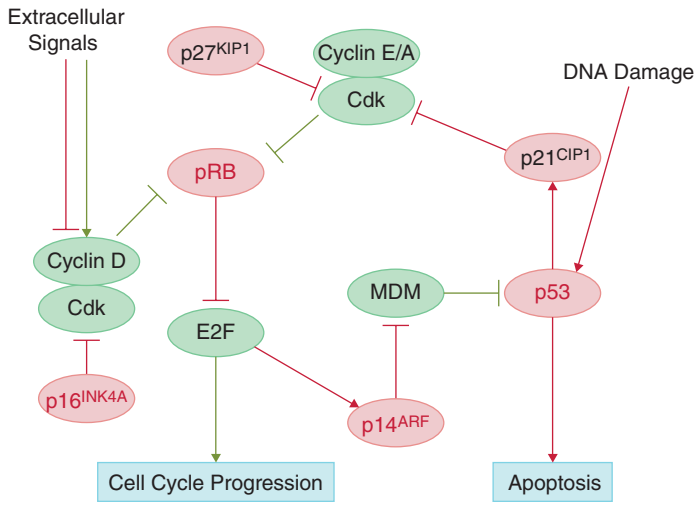

FIGURE 181-3. Interaction of the pRB/E2F and p53 pathways. Proto-oncogenes/oncogenes are indicated in green; tumor suppressors are indicated in red. Green and red connectors indicate growth/survival promoting and inhibitory effects, respectively. For details, see section on "Oncogenes and Tumors Are Not Restricted by Disease Site."

CRITICAL FEATURES OF THE CANCER PHENOTYPE

As cells accumulate oncogenic mutations, they progressively gain the ability to circumvent the normal mechanisms that control tissue homeostasis. Although diverse sets of mutations can lead to cancer, there are a set of general characteristics that must be acquired for tumors to develop. These phenotypic changes have been termed "hallmarks of cancer" and include the ability to sustain growth and resist growth inhibitory signals, resist apoptosis, stimulate angiogenesis, acquire replicative immortality, metastasize, deregulate cellular energetics, and avoid immune destruction (Table 181-3). These core characteristics, along with other factors that have a direct impact on cancer development and its clinical management, are discussed below.

Sustained Cell Proliferation

Normal tissue homeostasis is regulated by the balance between signals that promote proliferation and those that inhibit proliferation and induce differentiation. Multiple oncogenic mutations free cancer cells from this regulation by lessening their dependence on growth factors and weakening their sensitivity to negative signals. As noted earlier, direct disruption of cell

TABLE 181-2 pRB-E2F PATHWAY DISRUPTION IN CANCER

PROTEIN	GENE	COMMON CHANGES IN CANCER
pRB	*RB1* (tumor suppressor)	Mutation, deletion, methylation
p130	*RB2* (tumor suppressor)	Mutation
Cyclin D1	*CCND1* (oncogene)	Transcriptional activation, amplification, chromosome rearrangement
Cyclin E	*CCNE1* (oncogene)	Amplification
Cdk4	*CDK4* (oncogene)	Amplification, mutation
$p16^{INK4a}$	*CDKN2A* (tumor suppressor)	Deletion, methylation
$p27^{KIP1}$	*CDKN1B* (tumor suppressor)	Post-transcriptional downregulation

TABLE 181-3 THE HALLMARKS OF CANCER

- Sustaining proliferative signaling
- Evading growth suppressors
- Activating invasion and metastasis
- Enabling replicative immortality
- Inducing angiogenesis
- Resisting cell death
- Deregulating cellular energetics
- Avoiding immune destruction

The "hallmarks of cancer" comprise these biological capabilities acquired during the multistep development of human tumors. Underlying these hallmarks are "enabling characteristics," including genome instability and mutation and tumor-promoting inflammation. In addition to the cancer cells themselves, tumors recruit a repertoire of ostensibly normal cells that contribute to the acquisition of hallmark traits by creating a "tumor microenvironment."

Adapted from Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144:646-674.

cycle regulation through the pRB-E2F pathway, and activation of growth promoting signaling pathways through oncogene activation and tumor suppressor inactivation, isolates cancer cells from the requirement for growth factors and the effects of negative growth signals. Similarly, loss of p53 contributes to continued proliferation in the presence of genetic damage that would normally cause growth arrest.

Replicative Immortality

Even in the presence of continuous positive growth signals, the majority of cells in normal tissues have a limited replicative potential. Once this limit is reached, cells either enter senescence or die through apoptosis (see later). Therefore, for a cancer to grow to a size where it becomes clinically relevant, this limit must be overcome. One of the major factors that limit replicative potential is the shortening of chromosomes during replication. Genetic material at the end of chromosomes is protected from this loss by repetitive DNA sequences known as telomeres. Telomeres also form specialized structures that protect chromosomes from cellular exonucleases and from fusion due to non-homologous end joining. Germline cells and some stem cells express an enzyme, telomerase, that is able to lengthen their telomeres following replication. However, for most cells in the body, replication leads to a progressive shortening of telomeres until they become unable to protect the chromosomes; the DNA ends are then recognized as DNA damage, and the cell becomes senescent or enters apoptosis. Tumor cells evade this fate by acquiring mechanisms to maintain telomere length above the critical minimum. This is mainly by reexpression of telomerase or, less commonly, through a recombination-based mechanism. Loss of telomeres can also contribute to genetic instability, because defects in the DNA damage response can allow continued replication in the presence of chromosomal fusions associated with telomere loss.

Apoptosis

Apoptosis is a physiologic mechanism of programmed cell death that involves a protease cascade, leading to activation of executioner caspases. It is induced by both external and internal signals and is integral to development and control of tissue damage. Many of the alterations that lead to cancer also trigger apoptosis, mainly through the intrinsic pathway. These include oncogene activation, uncontrolled E2F activity due to inactivation of the pRB pathway, and DNA damage. Thus, cancer progression requires mechanisms to overcome these apoptotic responses. A major mechanism of apoptosis that results from intrinsic cellular abnormalities involves the release of cytochrome c from mitochondria, which leads to activation of cytoplasmic caspases. The tumor suppressors Bax and Bim mediate the release of cytochrome c, whereas the oncogene *Bcl2* inhibits its release. Another mechanism by which apoptosis is reduced in cancer cells involves upregulation of inhibitors of apoptotic proteins (e.g., survivin, XIAP, cIAP) that can directly suppress caspase activity. Changes in the expression of survival proteins can result from mutation of oncogenes and tumor suppressors, or from oncogenic activation of prosurvival signaling pathways such as those involving Akt and the NF-κB transcription factors. As noted, loss of p53 activity is a major contributor to apoptosis resistance in cancer cells.

Although these changes help to reduce the effect of pro-apoptotic changes associated with tumor development, apoptosis is prevalent in many tumors, and cancer cells remain susceptible to apoptotic signals. This susceptibility forms much of the basis for the selectivity of cytotoxic chemotherapeutic agents, and changes affecting apoptosis represent a mechanism of drug resistance in cancer.

Stress Responses

The accumulation of genetic damage and mutations, along with the hostile environment in which tumors develop, places significant stress on cancer cells. Cells respond to stress by eliciting a stress response characterized by upregulation of chaperone proteins. Chaperones, such as the heat shock proteins hsp90 and hsp70, play a central role in regulation of protein folding and are expressed at low levels under normal conditions. Under stress, these proteins become limiting, and their upregulation protects cells from the deleterious effects of protein denaturation and reduces apoptosis. Cancer cells characteristically have a constitutive stress response and are dependent on elevated levels of chaperone proteins for survival. Notably, many oncogenes are highly dependent on chaperones for their activity, making these proteins potential targets for cancer therapy. While no antichaperone therapies have been approved to date, several—especially those targeting hsp90—have shown promise in clinical trials.

Angiogenesis

Without an ability to recruit new blood vessels through angiogenesis, tumors are reliant on diffusion of oxygen and nutrients from surrounding tissue and can only grow to a few millimeters in diameter. Angiogenesis is a normal physiologic response that is essential for normal development and wound healing but is dormant in most adult tissues. It is regulated by a balance of pro-angiogenic and anti-angiogenic signals; several factors switch this balance toward angiogenesis in cancers. Hypoxia, which develops in tumors as they outgrow the diffusion limit of oxygen, is a major physiologic inducer of angiogenesis. In addition, activation of oncogenic signaling and loss of tumor suppressors directly leads to secretion of angiogenic factors such as vascular endothelial growth factor (VEGF) and basic fibroblast growth factor (bFGF), and to downregulation of anti-angiogenic factors such as Tsp-1. Many tumors also recruit significant numbers of inflammatory cells that secrete pro-angiogenic factors. Angiogenesis in tumors differs from that in normal tissues; the tumor vasculature is not properly organized. Blood vessels in tumors are leaky to both water and macromolecules, which leads to an increase in intratumoral pressure. These vessels are also prone to collapse, leading to large areas of tumors that remain hypoxic. The properties of tumor vessels can present a problem for cancer therapy, particularly for those that require oxygen (e.g., ionizing radiation).

Angiogenesis is an attractive target for cancer treatment, and agents that target this process have entered the clinic. Bevacizumab, a monoclonal antibody that inhibits angiogenesis by binding VEGF-A, has been approved for treatment of refractory glioblastoma. Several approved multitargeted kinase inhibitors that block VEGF receptors, such as pazopanib, sorafenib, and sunitinib, may also partially act through effects on angiogenesis.

Invasion and Metastasis

For most solid tumors, the main clinical threat comes from their ability to metastasize. Metastasis is a complex multistep process that involves local invasion, intravasation, survival in the circulatory system, extravasation, and colonization. A key mechanism underlying this process is epithelial-to-mesenchymal transition (EMT), characterized by the loss of cell-cell junctions and polarity, together with the gain of mesenchymal markers such as vimentin. EMT, which is normally involved in developmental cell dispersal and tissue regeneration, is subverted in cancer cells, allowing them to gain enhanced migratory capacity, invasiveness, and increased resistance to apoptosis. EMT has also been linked to the acquisition of stem cell characteristics (see later) and may thus contribute to the population of the metastatic site. Multiple transcription factors, including Snail, Slug and Zeb1, as well as miRNAs, cooperate with oncogenic signaling changes to induce EMT in cancer cells. In addition to cell-inherent changes, environmental factors including hypoxia and tissue interactions modulate EMT.

A major change that drives EMT is downregulation of E-cadherin, a component of adherens junctions and a tumor suppressor in the Wnt/β-catenin pathway. In many cases, tumor cells only undergo partial EMT, with the degree of EMT varying both within a tumor and between different tumors. Furthermore, EMT is reversible, and cells can undergo mesenchymal-to-epithelial transition (MET), which contributes to the establishment of tumors at metastatic sites.

Tumor Microenvironment

A tumor does not solely consist of transformed cancer cells but represents a complex tissue that also includes a wide array of stromal cells. The importance of stromal components in tumor phenotype is highlighted by the fact that four out of the seven genes in the genetic profile used by the recently developed Oncotype DX Colon Cancer Assay to predict the risk of recurrence in colon cancer are stromal factors. Cancer progression occurs within the context of complex interactions between multiple cell types, which can be major drivers of the evolution of tumor cells. For example, the ability to recruit immune cells can engender a significant advantage to tumor cells because intratumoral inflammatory cells secrete cytokines and growth factors that promote tumor cell proliferation and survival as well as angiogenesis. Recruitment of immune suppressor cells also helps tumors evade immune surveillance (see later). As a result, many cancers evolve to recruit inflammatory cells, and infiltration by these cells is often seen in tumor biopsies.

Other cellular components of the stroma, such as fibroblasts, can also secrete growth factors that support survival and proliferation of cancer cells. Direct interaction of integrins on cancer cells with the extracellular matrix

and other cells also enhances tumorigenesis by activation of focal adhesion kinase (FAK) and several oncogenic signaling pathways, including Mek/Erk and PI-3K/Akt signaling (see Fig. 181-2).

The Immune Response to Cancer

Accumulating evidence indicates that immune surveillance and removal of immunogenic transformed cells presents a physiologic barrier to cancer development. Cancer cells can evolve a number of mechanisms to evade the immune system. They can reduce immune recognition by downregulation of major histocompatibility complex (MHC) class I antigens or by expressing immune suppressive proteins such as PD-L1. Cancer cells also have direct effects on the immune system. They secrete factors such as TGF-β, VEGF, and IL-10 that reduce the immune response, potentiate immune checkpoints, and recruit T_{reg} cells and other immune suppressor cells. Exploitation of the immune response for cancer therapy is gaining momentum in the clinic.[7] A major strategy in immune therapy is to boost the antitumor response by targeting immune checkpoints. Ipilimumab, a humanized antibody approved for the treatment of melanoma, enhances T-cell activation through blockade of CTLA-4-mediated co-inhibition. Similar therapies targeting PD-1/PD-L1–mediated checkpoints have also recently demonstrated substantial clinical effectiveness. Cancer cells also express antigens that can be used in vaccine therapies. These include cancer-testis antigens that are normally only expressed on germline cells (e.g., NY-ESO-1) and tissue-specific antigens. Sipuleucel-T is a cancer vaccine against prostatic acid phosphatase (PAP); it is approved for metastatic hormone-refractory prostate cancer. Evidence indicates that immune responses may also play a role in the therapeutic effect of antibodies designed to target specific oncogenes (see following and Table 181-4).

Energy Metabolism

Cancer cells evolve changes in energy metabolism known as the Warburg effect. This is characterized by a vast increase in the uptake of glucose, which is preferentially channeled to glycolysis rather than aerobic respiration. Oncogenic alterations that underlie these changes in glucose uptake and metabolism include increased PI-3K/Akt/mTor signaling, as well as upregulation of c-Myc and hypoxia-inducible factor-1α (HIF-1α). A direct role for altered metabolism in carcinogenesis[8] is supported by mutations affecting succinate dehydrogenase in hereditary paraganglioma (see Table 181-1), succinate dehydrogenase or fumarate hydratase in some renal cell carcinomas, and isocitrate dehydrogenase-1 and -2 (*IDH1*, *IDH2*) in glioma. Although the exact advantage conferred by the Warburg effect remains to be established, a likely rationale is that increased glycolysis provides many of the intermediates required for macromolecular synthesis, and thus may be required to sustain tumor growth. In some cases, it may also reflect an adaptation to hypoxia in tumors. Altered energy metabolism has attracted attention as a potential biomarker and target for cancer therapy, with 2-deoxyglucose and inhibitors of 6-phosphofructo-2-kinase, and pyruvate kinase M2 already being tested in clinical trials. The increased glucose uptake in cancer cells also underlies the use of positron emission tomography for visualizing tumors in vivo.

Cancer Stem Cells

Normal proliferating tissues contain pluripotent self-renewing stem cells that have high proliferative potential and can differentiate into all of the cell types in the tissue. Stem cell division gives rise to daughter stem cells and to more differentiated, transit-amplifying progenitor cells. Transit amplifying cells then continue to divide and progressively differentiate to give rise to the full spectrum of a tissue's cellular components. Stem cells are rare within tissues and usually divide very slowly, with the bulk of the proliferation responsible for tissue renewal carried out by the transit-amplifying population. Stem cells can also be found in tissues that are traditionally considered to be nonproliferative, where they remain quiescent until required for tissue repair.

Increasing evidence indicates that cancer cells are not phenotypically identical cells but are organized similarly to proliferative tissues. Analysis of solid tumors and hematopoietic cancers has revealed the presence of populations of cells that have a high potential to regenerate cancers, whereas the bulk of cancer cells have a negligible regenerative potential. Moreover, these cells can regenerate tumors that recapitulate the heterogeneity of the original malignancy. Such cells are analogous to normal tissue stem cells with regard to their capacity for self-renewal and "tissue" regeneration, and are thus referred to as cancer stem cells. This analogy is strengthened by the fact that cancer stem cells express many of the markers that differentiate stem cells in the tissues from which the cancer arose. In addition to their ability for self-renewal, cancer stem cells differ from the bulk of cancer cells in that they are slowly dividing and relatively resistant to chemotherapy. Despite these differences, the precise relationship between the different populations of cancer cells remains to be established. For example, EMT can induce a cancer stem cell–like phenotype, opening the possibility that acquisition of "stemness" in cancer cells is a dynamic process, with cells able to enter and exit a stem cell state in response to microenvironmental cues or other stimuli.

Whatever the nature of cancer stem cells, the presence of a population of drug-resistant cells with self-renewing capacity has required a reevaluation of the criteria by which cancer treatments are assessed. Treatments that do not kill the cancer stem cell population are destined to result in relapse, no matter how efficiently they debulk the cancer. Future development of treatment regimens will have to consider efficacy against the minor population of cancer stem cells that may be difficult to detect after treatment.

Tumor Heterogeneity

In addition to the phenotypic heterogeneity associated with the presence of stem cells, cancers show considerable genetic and morphologic heterogeneity both within and between patients.[9] The genetic instability associated with cancer development, together with the stochastic nature of oncogenic mutations, supports genetic drift and the coevolution of cancer cells with different genetic lesions within a tumor. Morphologic heterogeneity within tumors is particularly noticeable in the stroma, which is affected not only by the ability of cancers to recruit different cell types but also by random factors such as tumor depth, variations in vascularization, and differences in the surrounding tissue. The resultant variation in microenvironment produces different evolutionary pressures that select for regional heterogeneity within the cancer cells of a tumor. Even in apparently homogeneous cancers that arise from clonal expansion of genetically identical cells, small variations in the microenvironment may allow for the survival of subpopulations of genetically distinct cells.

Heterogeneity across tumors has implications for therapy that are analogous to those seen with stem cells. Genetically distinct populations in different microenvironments would be expected to differ in their susceptibility to treatment. Thus, heterogeneous cancers are likely to have subpopulations of resistant cells that may evade therapy and lead to relapse. This is a particular concern for metastases, which are exposed to environmental pressures that are distinct from those of the original tumor.

EXPLOITATION OF TUMOR BIOLOGY FOR TARGETED THERAPIES

Targeted Therapies

Knowledge of the biology that underlies cancer development is being leveraged in the clinic,[10] with an increasing number of drugs that target molecular changes in cancer being evaluated in patients (Table 181-4). The promise of these therapies is that by targeting changes specific to cancer, highly effective treatments can be developed that do not have the side effects associated with traditional cytotoxic chemotherapy. The majority of agents in the clinic directly target oncogenic pathways, with most approved agents being either small molecules that target protein kinases or antibodies that disrupt signaling from cell surface receptors (see Table 181-4). Another strategy is to exploit the potential "synthetic lethality" that results from oncogenic changes, as seen with the susceptibility of cells with BRCA1 or BRCA2 mutations to inhibitors of PARP (see "Genome Instability and Cancer Development" section).

By their nature, targeted therapies are only effective in tumors with relevant mutations. Specific mutations may be common in tumors from a particular disease site, leading to targeted therapies being associated with specific cancer types (e.g., HER2 inhibitors and breast cancer, EGFR inhibitors in lung and colon cancer); however, only a subpopulation of tumors at a specific disease site will be susceptible to any given targeted therapy. For example, although breast cancers with overexpression of HER2 are responsive to trastuzumab (see Table 181-4), they represent only 20 to 25% of all breast cancers; furthermore, the majority of these HER2-expressing breast tumors do not respond to HER2 inhibitors. Tumors can also have mutations in downstream components of a signaling pathway that negate the action of a drug. For example, B-Raf is downstream of EGFR, and thus EGFR inhibitors are ineffective for tumors carrying mutations activating B-Raf, even if they overexpress EGFR. As a result, application of targeted therapies requires testing of tumors to determine if they have the correct genomic profile to respond to any given treatment. Optimal use of targeted agents requires that

TABLE 181-4 TARGETED CANCER THERAPIES

THERAPEUTIC AGENT	TYPE	TARGET	DISEASE	INDICATION	COMPANION TEST
Imatinib Dasatinib Nilotinib	Small molecule	BCR-ABL	CML, ALL	Philadelphia chromosome positive	Cytogenetic analysis, FISH, PCR
Ponatinib Bosutinib	Small molecule		CML and AML resistant to prior tyrosine kinase inhibitor therapy	Philadelphia chromosome positive	Cytogenetic analysis, FISH, RT-PCR
Imatinib	Small molecule	c-Kit	Gastrointestinal stromal tumors	CD117 (c-Kit) positive	Immunohistochemistry (c-Kit PharmDx)
Cetuximab Panitumumab	Monoclonal antibody	EGFR	Colorectal cancer, head and neck	*KRAS* wild-type in colon cancer	PCR (*therascreen* KRAS RGQ Kit)
Gefitinib Erlotinib	Small molecule	EGFR	NSCLC	Mutant EGFR	PCR (*therascreen* EGFR RGQ PCR Kit, cobas EGFR Mutation Test)
Tramatenib Dabrafenib Vemurafenib	Small molecule	B-Raf	Melanoma	V600 mutant BRAF	PCR (THxID, cobas 4800 BRAF V600 Mutation Test)
Trastuzumab	Monoclonal antibody	Her2	Breast, stomach, gastroesophageal	Her2 positive	Immunohistochemistry (PATHWAY, InSite, Bond Oracle, HercepTest), FISH (Inform, PathVysion, SPOT-Light, HER2 CISH PharmDx)
Lapatinib	Small molecule	Her2, EGFR	Breast	Her2 positive, trastuzumab resistant	
Everolimus	Small molecule	mTOR	Breast, pancreatic, renal cell, astrocytoma		
Bevacizumab	Monoclonal antibody	VEGF	NSCLC, metastatic colorectal, metastatic kidney, glioblastoma		
Ipilimumab	Monoclonal antibody	CTLA-4	Melanoma		

Companion tests in parentheses are U.S. Food and Drug Administration–stipulated tests for certain uses of the corresponding drug.
ALL = acute lymphoblastic leukemia; CML = chronic myelogenous leukemia; FISH = fluorescence in situ hybridization; NSCLC = non–small cell lung cancer; PCR = polymerase chain reaction.

the appropriate mutation or other molecular abnormality be present (see Table 181-4). Many new targeted drugs are being approved by the U.S. Food and Drug Administration, together with a companion diagnostic test for identification of relevant therapeutic targets.

Use of genetic testing and targeted therapies is blurring the classification of cancer based on disease site. The fact that similar genetic changes can occur in cancers from different organs means the same drug may be appropriately used to target the same molecular change in cancers originating from different histologies (see Table 181-4). This, together with the observation that only a subset of cancers respond to a targeted treatment, suggests that in the future the classification of tumors based on the genetic abnormalities they contain may be more relevant for the choice of therapy than the tissues from which they arise.

Efficacy and Resistance Associated with Targeted Therapies

When appropriate genetic changes are present, targeted therapies can be remarkably effective, particularly in homogeneous cancers that are driven by a single well-defined genetic lesion. This can be seen with Philadelphia chromosome–positive CML, which arises because of the aberrant tyrosine kinase activity of the *BCR-ABL* oncogene (Chapter 184). Imatinib, an inhibitor of Abl (along with PDGFRα and c-Kit) is highly effective in treating chronic-phase CML patients, turning it from a uniformly fatal disease to one with a 7-year survival of about 90%. It is important to note that despite its clinical efficacy, the majority of patients treated with imatinib harbor residual disease, and recurrence is observed when treatment is stopped. This is true for most targeted therapies which are labeled for continuous treatment until disease recurrence. Thus, even successful targeted therapies may not cure cancer but turn it into a chronic condition that requires ongoing treatment for continuing patient benefit.

Because of their specificity, most targeted therapies are particularly vulnerable to the development of resistance resulting from mutations that produce changes in the target protein. Mutations that affect drug binding or induce overexpression (such as amplification) can render a targeted therapy ineffective. The genetic instability present in tumors, along with the need for chronic treatment with these agents, exacerbates this problem. Resistance arising from mutation of the target can be overcome by use of agents that are insensitive to the mutation. For example, tyrosine kinase inhibitors such as dasatinib, nilotinib, or ponatinib can be effective against imatinib-resistant BCR-ABL. Where resistance is the result of mutations in another protein that reduces reliance on a particular target, the other proteins can be targeted if known. For example, the B-Raf inhibitor dabrafenib (see Table 181-4) is being tested in combination with EGFR inhibitors for treatment of tumors that are resistant to EGFR inhibitors owing to the V600-activating mutation of B-Raf (see Fig. 181-2).

Although targeted therapies have proven highly effective for homogeneous cancers, they have more limited efficacy in the majority of more heterogeneous tumors. This can be seen with CML, where imatinib has limited efficacy in patients experiencing blast crisis where additional molecular lesions have accumulated (Chapter 184). Targeted therapies are available that prolong survival of appropriately selected patients, but in most cases these treatments give rise to transient responses with rapid emergence of resistant disease. Treatment failure may arise from de novo mutations or from preexisting subpopulation(s) of resistant cells present in heterogeneous tumors. The rapid emergence of resistant tumors indicates that tumor heterogeneity is a particular concern. Use of combination therapies, where targeted agents are combined with either conventional chemotherapy or with other targeted therapies, is one strategy that can increase the efficacy of these agents (e.g., pertuzumab, a monoclonal antibody that targets HER2, has been approved for combined therapy with trastuzumab and docetaxel in breast cancer).

Side Effects of Targeted Therapies

Although targeted therapeutics may lack the systemic toxicity of traditional chemotherapy, they are not without side effects. With traditional chemotherapies, the therapeutic window is based on the increased susceptibility of cancer cells to their cytotoxic effects, and treatment is limited by their maximum tolerated dose (MTD). However, the concept of MTD is less appropriate for targeted therapies where side effects are a direct consequence of the intended action of the drug. Most oncogenes have normal cellular equivalents (proto-oncogenes) that regulate homeostasis in non-tumor tissues. Since proto-oncogenes can be inhibited by the targeted agents, disruption of normal tissue homeostasis may, in some cases, be an unavoidable consequence of targeted therapies. In many situations, such as the rash seen with EGFR inhibitors, side effects can provide an indication of whether the drug is altering its target. Severe side effects of targeted therapeutics, such as cardiac problems associated with HER2 inhibitors, are rare and generally reverse upon cessation of treatment.

Genetic Profiling and Personalized Medicine

Development of molecularly targeted cancer therapies opens the possibility that treatment could be tailored to the specific defect in the cancer of an individual patient. Such a personalized approach to cancer therapy has become feasible with recent technologic advances in DNA sequencing that allow mutational profiles of individual tumors to be determined over a short time interval. Technologies currently under development include array-based platforms for analysis of mutations in genes of interest, deep sequencing of known drivers of tumorigenesis (i.e., repeated reads of the sequence of interest), and whole genome or exome sequencing (Chapter 43). Each approach has its advantages and disadvantages. For example, deep sequencing can identify mutations in subpopulations of cells that may be missed by other methods, but this approach is more costly and time consuming. Approaches that rapidly test a limited number of genes, on the other hand, may miss changes that would be identified in whole genome/exome sequencing. Thus, the best strategy for personalized medicine remains to be determined. Because analysis is normally performed on biopsies of primary tumors, the problem of regional heterogeneity in tumors and potential genetic differences in metastases also needs to be addressed for successful implementation of more personalized treatments.

Genetic profiling of tumors can also be used to gain prognostic information. Our increasing knowledge of the biology and genetics of cancer has led to the development of mutation and expression signatures that are associated with various cancer phenotypes. As a result, several commercial tests that use the mutational status of a panel of genes to predict the risk of recurrence and/or aggressiveness for breast (e.g., MammaPrint, Oncotype DX), colon (e.g., Oncotype DX), and prostate (e.g., Prolaris, Oncotype DX) cancers are now available for use in the clinical setting.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144:646-674.
2. Vogelstein B, Papadopoulos N, Velculescu VE, et al. Cancer genome landscapes. *Science*. 2013;339:1546-1558.
3. Kidane D, Chae WJ, Czochor J, et al. Interplay between DNA repair and inflammation, and the link to cancer. *Crit Rev Biochem Mol Biol*. 2014;49:116-139.
4. Menck CF, Munford V. DNA repair diseases: what do they tell us about cancer and aging? *Genet Mol Biol*. 2014;37(1 suppl):220-233.
5. Tsai HC, Baylin SB. Cancer epigenetics: linking basic biology to clinical medicine. *Cell Res*. 2011;21:502-517.
6. Palanichamy JK, Rao DS. miRNA dysregulation in cancer: towards a mechanistic understanding. *Front Genet*. 2014;5:54. doi:10.3389/fgene.2014.00054; eCollection 2014.
7. Lizée G, Overwijk W, Radvanyi W, et al. Harnessing the power of the immune system to target cancer. *Annu Rev Med*. 2013;64:71-90.
8. Frezza C, Pollard P, Gottlieb E. Inborn and acquired metabolic defects in cancer. *J Mol Med (Berl)*. 2011;89:213-220.
9. Almendro V, Marusyk A, Polyak K. Cellular heterogeneity and molecular evolution in cancer. *Annu Rev Pathol*. 2013;8:277-302.
10. Huang M, Shen A, Ding J, et al. Molecularly targeted cancer therapy: some lessons from the past decade. *Trends Pharmacol Sci*. 2014;35:41-50.

REVIEW QUESTIONS

1. A 36-year-old woman is found to have acute myelogenous leukemia. Family history obtained on diagnosis reveals that the patient's mother died from a glioblastoma multiforme detected at the age of 40. The patient's older brother died from recurrent osteogenic sarcoma at the age of 39, and her maternal grandmother succumbed to metastatic breast cancer before the onset of menopause. Mutations in which one of the following genes is associated with this familial cancer syndrome?

A. *NF1*
B. *APC*
C. *TP53*
D. *MLH1*
E. *WTI*

Answer: C This family demonstrates the varied range of expression of the Li-Fraumeni syndrome, caused by mutations in the TP53 tumor suppressor. Although *NF1* mutations can be associated with breast cancer and leukemia, glioblastoma multiforme does not occur with this mutation. *MLH1* and *APC* mutations are associated with gastrointestinal malignancies for the most part. *WTI* mutations are specifically associated with familial Wilms tumor (See Table 181-1.)

2. Viral infections are not associated with which one of the following malignancies:

A. Cervical cancer
B. Merkel cell carcinoma
C. B-cell non-Hodgkin's lymphoma
D. Colorectal carcinoma
E. Hepatocellular carcinoma

Answer: D The major etiologic factor in the causation of cervical cancer is the human papillomavirus. Merkel cell carcinomas are associated with the Merkel cell polyomavirus. Epstein-Barr virus causes B-cell non-Hodgkin's lymphoma, and hepatocellular carcinomas are associated with hepatitis B and C viruses. There is no virus known to be associated with the development of colon cancer.

3. One mechanism related to the unlimited replicative potential of tumor cells is:

A. Low threshold for necrotic cell death
B. Increased apoptotic gene expression
C. Tumor cell heterogeneity
D. Overexpression of telomerase
E. Deregulation of cellular energetics

Answer: D High levels of tumor cell telomerase limit tumor cell senescence, promoting cellular longevity. Cells with a low threshold for necrosis or apoptosis have a diminished lifespan. Tumor cell heterogeneity is not directly related to the ability of tumor cells to replicate.

4. There is increasing evidence that specific genetic abnormalities in certain tumors are associated with alterations in tumor cell energy metabolism. Which of the following such associations is correct?

A. Mutations in the mitochondrial fumarate hydratase gene and hereditary renal cancer
B. Mutations in membrane-bound NADPH oxidase 1 and colon cancer
C. Mutations in mitochondrial complex I and breast cancer
D. Mutations in ribonucleotide reductase and CML

Answer: A Hereditary renal cancers have been associated with mutations in the fumarate hydratase gene that produce tumors that upregulate HIF-1α and are highly vascular, related in part to enhanced oxidative glycolysis.

182

MYELODYSPLASTIC SYNDROMES

DAVID P. STEENSMA AND RICHARD M. STONE

DEFINITION

The myelodysplastic syndromes (MDS) are a heterogeneous group of bone marrow failure syndromes collectively characterized by ineffective hematopoiesis resulting in peripheral blood cytopenias, most commonly anemia. MDS have a tendency to evolve over time, and there is a risk of progression to acute myeloid leukemia (AML), which is arbitrarily defined by the World Health Organization (WHO) as the presence of 20% or more immature myeloid "blast" cells in the blood or bone marrow.

Although some investigators feel that MDS should not be considered a form of cancer because some cases remain stable for many years, and a few patients have a pathophysiology dominated by autoimmune suppression of hematopoietic progenitor cells and resembling aplastic anemia, MDS are classified as myeloid neoplasms by the WHO and other organizations. Many of the somatic gene mutations that are recurrent in MDS are also present in AML and other myeloid neoplasms, and there are typically one or just a few mutant clones in the bone marrow contributing to most of the marrow cellularity.[1]

The minimal diagnostic criteria for MDS as defined by the WHO include the presence of a meaningful and persistent cytopenia (i.e., hemoglobin <11 g/dL, absolute neutrophil count <1500/mm^3, or platelet count <100,000/mm^3) together with at least one of the following features: (1) greater than 5% blasts in the bone marrow, (2) an abnormal marrow karyotype with an acquired chromosome anomaly that is consistent with the diagnosis of MDS, (3) other evidence of clonal hematopoiesis besides marrow karyotype (e.g., abnormal result using a fluorescence in situ hybridization [FISH] panel of probes directed at common MDS-associated chromosomal abnormalities), or (4) greater than 10% cells in any given hematopoietic linage that appear dysplastic (i.e., morphologically abnormal, described further below) in a patient in whom other causes of dysplastic cell morphology such as nutritional deficiency have been ruled out. Patients sometimes have persistent cytopenias and do not quite meet these diagnostic criteria, yet no other diagnosis is apparent after detailed investigation. These disorders are sometimes termed *idiopathic cytopenias of undetermined significance* (ICUS), and such patients should be followed over time because they have a risk of developing overt MDS.[2]

EPIDEMIOLOGY

The incidence and prevalence of MDS have been difficult to estimate accurately, in part because of varying definitions and imprecise terminology historically, and also because until recently, MDS cases were not captured by most global cancer registries. The U.S. National Cancer Institute's Survey, Epidemiology, and End Results (SEER) data indicate that 10,000 to 12,000 new cases are diagnosed in the United States each year. However, analysis of insurance claims data, including Medicare claims, suggests that the actual incidence of MDS is much higher—at least 40,000 new U.S. cases per year, several times higher than the incidence of AML.[3] In addition, many geriatric patients who have cytopenias of uncertain etiology and may have MDS are incompletely evaluated (e.g., they do not undergo bone marrow aspiration) and never receive an MDS diagnosis.

The primary risk factor for development of MDS is aging, with disease a consequence of cumulative acquisition of somatic mutations in marrow stem cells. The rate at which MDS is diagnosed is slightly increased in workers in certain industries such as the petroleum industry and agricultural sector, probably owing to exposure to genotoxic hydrocarbons, accelerating acquisition of such mutations. There is a slight male predominance overall, perhaps representing patterns of occupational exposure. In Asia and Eastern Europe, MDS is typically diagnosed at younger ages than in the West; in the United States, the median age at which MDS is diagnosed is 71 years, whereas in China it is closer to 50 years.

Individuals who have been exposed to ionizing radiation or certain types of cytotoxic chemotherapy (e.g., alkylating agents such as chlorambucil or mephalan, or topoisomerase II inhibitors such as etoposide or doxorubicin) have an increased risk of developing MDS subsequently. The peak incidence of MDS diagnosis after alkylating agent or radiation exposure is 5 to 10 years later, whereas after a topoisomerase inhibitor exposure, it is 1 to 3 years. The risk never goes away entirely; even 50 years after the Hiroshima and Nagasaki atomic bomb events in Japan, radiation-exposed individuals continued to be diagnosed with MDS at a higher rate than the unexposed population. Patients with a chemotherapy or radiation exposure history are said to have therapy-related MDS (t-MDS) or "secondary" MDS.

Although familial MDS is rare, germline gene mutations in several genes, including *RUNX1* and *GATA2*, predispose to MDS.[4] *RUNX1* mutations are associated with a prodrome of thrombocytopenia that can be mistaken for immune thrombocytopenic purpura, whereas *GATA2* mutations are associated with a history of mycobacterial infections and monocytopenia.

PATHOBIOLOGY

MDS are clonal disorders, meaning that one or several stem or progenitor cells in the marrow that bear somatic gene mutations expand and proliferate at the expense of healthy stem cells and come to dominate the bone marrow. These clones can then acquire additional mutations that give rise to subclones that contribute to progression of disease. The process of somatic mutation acquisition can be accelerated through exposure to radiation or DNA-damaging chemicals, as noted earlier.

Marrow cellularity is usually normal or increased in MDS, but excessive apoptosis (i.e., programmed cell death) of hematopoietic cells within the bone marrow accounts for the peripheral blood cytopenias. Excessive apoptosis within the marrow characterizes earlier MDS, but as the disease progresses toward AML and one clone begins to dominate, intramedullary apoptosis may decrease. In addition, hematopoietic differentiation, which is disordered but still can proceed in earlier stages of the disease, may become completely impaired late in the disease, such that most developing hematopoietic cells are arrested at the myeloblast stage. If 20% or more of the cells in the marrow or blood are blast cells, this is considered AML; MDS by definition is associated with less than 20% blast cells.

Large gains and losses of chromosomal material within diseased cells, including monosomies, trisomies, and large chromosomal deletions, are detectable by routine metaphase karyotyping present in one half of patients

with de novo MDS and more than 80% of patients with t-MDS.[5] Newer genetic techniques such as array-based comparative genomic hybridization may uncover cryptic chromosomal deletions in patients with MDS who have normal karyotype results.

More than 25 genes are now known to be recurrently mutated in MDS.[6,7] No one mutation dominates; MDS are associated with considerable genetic heterogeneity. These genes and the probable function of the proteins they encode are listed in Table 182-1. More than 90% of patients have at least one detectable clonal somatic mutation, and most patients with MDS have several such mutations detectable in their marrow cells. Mutations affecting epigenetic patterning (e.g., DNA methylation) and chromatin conformation (e.g., histone-DNA interactions) are common in MDS, and this may be why agents that affect DNA methylation, such as azacitidine and decitabine, are effective in some patients. In addition, mutations affecting pre-mRNA splicing are common, especially in patients with MDS subtype refractory anemia with ring sideroblasts, but the specific genes that are misspliced and give rise to an MDS phenotype are unclear.

A subset of patients with pathologic findings that resemble MDS will have autoimmune suppression of hematopoiesis, which may be driven by a clonal T-lymphocyte population.[8] Patients are more likely to have an immune component to their marrow failure if they are younger than 55 years and have a hypocellular marrow similar to that seen in aplastic anemia, a normal

TABLE 182-1 GENE MUTATIONS ASSOCIATED WITH MDS AND FUNCTIONAL CLASS OF THE ASSOCIATED PROTEIN

RECURRENTLY MUTATED GENES	MUTATION FREQUENCY IN MDS	ADDITIONAL NOTES
SPLICEOSOME COMPONENTS (ALTER PRE-mRNA SPLICING)		
SF3B1	20-25%	Strongly associated with presence of ring sideroblasts (60-80% of RARS cases), often co-occurs with *DNMT3A** mutations
*U2AF1**	5-10%	Often co-occurs with del(20q)
SRSF2	5-10%	More frequent in CMML (25-30%), often co-occurs with *RUNX1** or *ASXL1** mutations
ZRSR2	5-10%	
U2AF2, SFRA1, PRPF40B, SF1	<2% each	
EPIGENETIC PATTERN & CHROMATIN CONFORMATION MODIFIERS		
TET2	20%	More frequent in CMML (≈40%)
*DNMT3A**	10-15%	Associated with poor outcome after transplant
SETBP1	5-10%	More common in CMML; germline mutations cause Schinzel–Giedion syndrome
*ASXL1**	10-20%	More frequent in CMML (≈40%)
*EZH2**	6%	More frequent in CMML (≈12%)
KDM6A	<2%	Rare in MDS, more frequent in CMML
IDH1, IDH2	<2%	More frequent in AML than MDS
PHF6	<1%	
ATRX	Rare	Associated with acquired α-thalassemia (hemoglobin H inclusions in RBCs)
TRANSCRIPTION FACTORS		
*RUNX1**	10-15%	Germline mutations cause familial platelet disorder with propensity to AML (FPD-AML)
*ETV6**	<5%	Translocations common in AML; formerly known as *TEL*
MYBL2	<1%	
GATA2	Rare	Germline mutations cause familial MDS with monocytopenia
CEBPA	Rare	More frequently mutated in AML, germline mutations cause familial AML without preceding dysplasia
WT1	Rare	More frequently mutated in AML
GENOME STABILITY		
*TP53**	5-10%	Associated with very poor prognosis; associated with t-MDS and with complex karyotype
TYROSINE KINASE SIGNALING		As a group, these genes are more often mutated in myeloid malignancies with a proliferative component
NRAS, KRAS, BRAF*	5-10% collectively	*NRAS* mutations more common in CMML and AML; first mutations described in MDS (1987)
JAK2	<5%	More frequent in RARS with thrombocytosis (RARS-T) (50%)
CBL, CBLB	<5%	More frequently mutated in CMML than MDS
FLT3, MPL, KIT	<1%	*MPL* mutated in (5%) of RARS-T, *FLT3* and *KIT* mutations are rare in MDS and much more common in AML
NF1	<1%	Germline mutations cause neurofibromatosis type 1
PTPN11	<1%	Rare in MDS, more frequent in JMML (30%)
OTHERS		
NPM1	<2%	Much more frequently mutated in AML
GNAS	<1%	
BCOR, BCORL1	<1%	
UMODL1	<1%	
ZSWIM4	<1%	
Cohesins *(RAD21, STAG1, STAG2, SMC3, SMC1A)*	Rare	

*Those mutations that appear to be independently associated with poor outcomes in MDS are denoted by an asterisk. (Bejar R. Clinical and genetic predictors of prognosis in myelodysplastic syndromes. *Haematologica* 2014;99:956-964.) AML = acute myelogenous leukemia; CMMS = chronic myelomonocytic leukemia; JMML = juvenile myelomonocytic leukemia; RARS = refractory anemia with ring sideroblasts; t-MDS = therapy-related MDS. Data based on Bejar R, Levine R, Ebert BL. *J Clin Oncol.* 2011;29:504-515, and others.

chromosome pattern or trisomy 8 rather than a more complex karyotype, a paroxysmal nocturnal hemoglobinuria cell population (Chapter 160) detectable by flow cytometry, and HLA-DR15 tissue type. Such patients may respond to immunosuppressive therapy with anti–T-cell therapies.

Most patients with MDS come to medical attention because they have anemia (which is present in 95% cases and is often macrocytic); one half of patients also have neutropenia, thrombocytopenia, or both at the time of diagnosis. The peripheral blood and bone marrow exhibit characteristic "dysplastic" cell morphologic abnormalities. In the peripheral blood, these include hypogranular neutrophils, hypolobated neutrophils such as the two-lobed pseudo-Pelger-Hüet cell, platelet size and granularity abnormalities, poorly hemoglobinized red cells, and circulating early myeloid cells such as small numbers of myeloblasts. Maturation abnormalities in the bone marrow include megaloblastoid erythroid maturation, multi-nucleated erythroid cells, ring sideroblasts (i.e., erythroid precursors with abnormal peri-nuclear iron-containing mitochondria visible with Perls' Prussian blue reaction), nuclear karyorrhexis, myeloid lineage abnormalities such as left-shifted myelopoiesis and the presence of hypogranular white cells, and megakaryocytic abnormalities such as micromegakaryocytes, very large hypernucleated megakaryocytes, or hypolobated megakaryocytes (Fig. 182-1). A mild increase in reticulin fibrosis is common in MDS, but severe reticulin fibrosis or collagen fibrosis are rare. Extramedullary hematopoiesis is uncommon in MDS, and the presence of splenomegaly or hepatomegaly should prompt consideration of another cause.

CLINICAL MANIFESTATIONS

The natural history of MDS is highly variable. Approximately half of patients will die of complications of the cytopenias, particularly infection due to both neutropenia and to functional neutrophil defects (i.e., impaired bactericidal activity due to hypogranularity). Among lethal infections in MDS, pneumonia and bacteremia due to gastrointestinal or urinary infections are the most common. Bleeding is the second leading common cause of death. In addition, anemia may exacerbate cardiovascular or cerebrovascular disease. Approximately 25% of patients will progress to AML, which is usually a fatal complication. Finally, because MDS is often diagnosed in elderly patients or in those with a history of cancer, unrelated causes are the chief contributor to death in at least one third of patients.

The clinical course in MDS is dominated by complications of the cytopenia, including mucosal bleeding, recurrent infections, and exertional dyspnea and fatigue. Fatigue correlates poorly with the degree of anemia and may be due in part to cytokine release by the abnormal clonal cells. Paraneoplastic manifestations including neutrophilic dermatosis (Sweet syndrome, Chapter 440), inflammatory arthritis, and other rheumatologic syndromes occur in 10 to 15% of patients. Repeated transfusions of red cells can lead to transfusional hemosiderosis and organ dysfunction, including hepatic and cardiac abnormalities. The frequency with which actual clinical complications of iron overload occur, as well as the importance of chelation therapy in MDS, are controversial topics.

DIAGNOSIS

The diagnosis of MDS is usually suspected when the patient is discovered to have cytopenias, especially macrocytic anemia in an older person that is not due to B_{12} or folate deficiency, medication effect (e.g., methotrexate or azathioprine), or alcohol abuse. The blood count abnormality may be an incidental finding, but most patients with MDS are symptomatic with fatigue, dyspnea, and other cytopenia-associated symptoms.

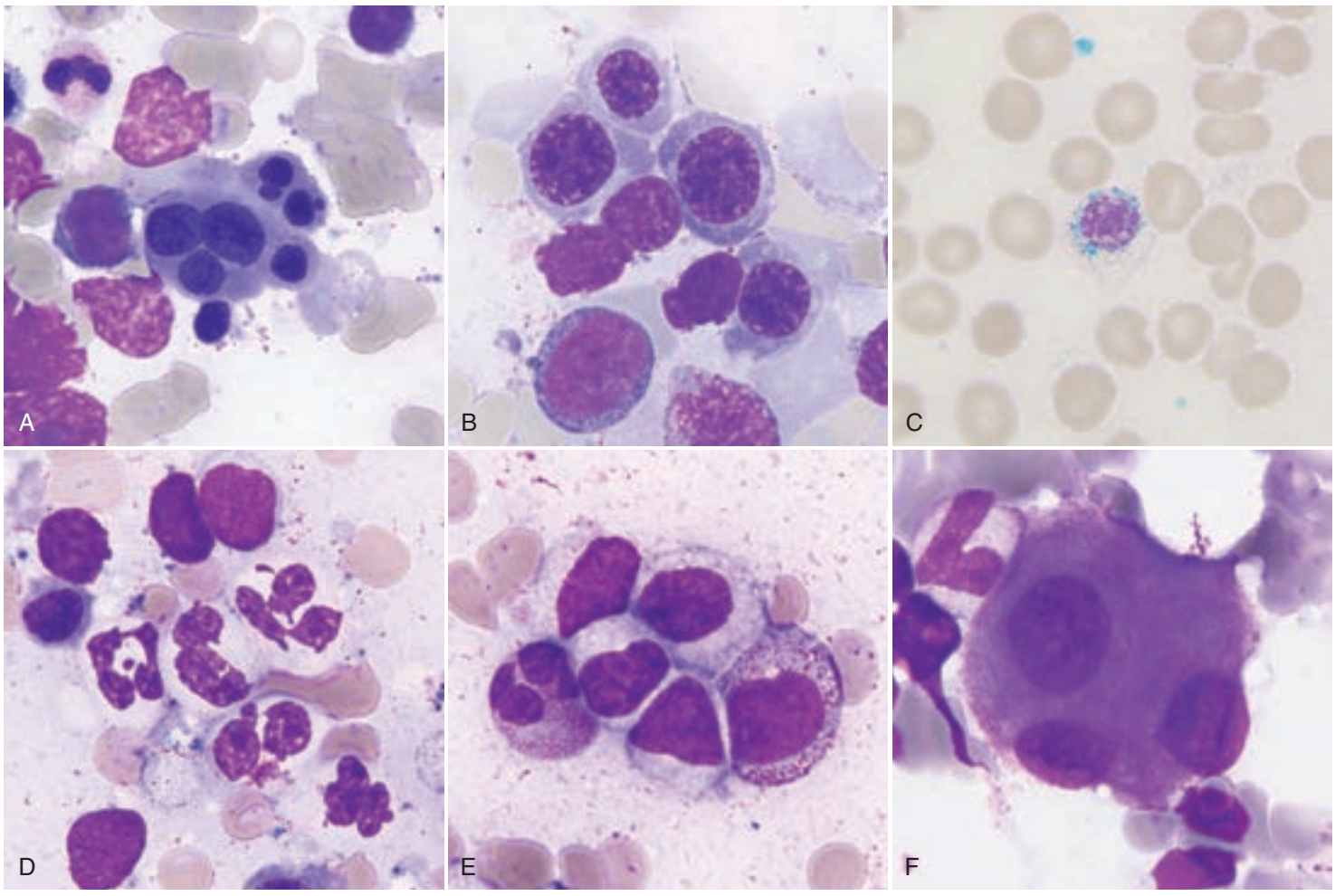

FIGURE 182-1. Dysplastic cell morphologic abnormalities commonly observed in the peripheral blood or marrow of patients with myelodysplastic syndrome (MDS). A, Multinucleated erythroid precursor. **B,** Megaloblastoid maturation of erythroid precursors with "open" nuclear chromatin. **C,** Ring sideroblast (pathologic erythroid precursor). **D,** Hypogranular neutrophils. **E,** Pseudo-Pelger-Huët anomaly (hypolobated neutrophil). **F,** Abnormal nuclear lobation in a megakaryocyte—a "Pawn Ball" cell with three separated nuclei/nuclear lobes. *A, B, D,* and *F* are Wright-Giemsa–stained marrow aspirate; *C* is Perls' Prussian blue reaction of marrow aspirate; *E* is Wright-Giemsa–stained peripheral blood. (Figures courtesy Elizabeth A. Morgan, MD, PhD, Brigham & Women's Hospital.)

A bone marrow aspirate is required to make the diagnosis, and bone marrow core (trephine) biopsy is also useful and can provide information about the marrow architecture and overall marrow cellularity.[9] Metaphase cytogenetics is essential, but FISH is only necessary if karyotyping fails because of an unsuccessful aspiration. FISH abnormalities are rarely detected in patients for whom at least 20 metaphases can be counted during karyotyping.

Not all that is dysplastic is MDS, and it is important to consider other potential causes for cytopenias and morphologic abnormalities. The differential diagnosis includes nonclonal disorders such as B_{12} or folate deficiency, copper deficiency, iron deficiency, HIV infection, medication exposure (especially cytotoxic chemotherapy agents), or an immune disorder such as T-cell large granular lymphocyte (T-LGL) leukemia. In patients with isolated sideroblastic anemia, it is important to rule out congenital sideroblastic anemias, including those due to germline mutations in *ALAS2*, which can present late in life.

In addition, MDS must be separated from aplastic anemia (Chapter 165), which can be oligoclonal but is not typically associated with Pelger-Hüet cells or other morphologic abnormalities, and the chromosome karyotype in aplastic anemia is usually normal. MDS can also be mistaken for AML or a myeloproliferative neoplasm (MPN) such as primary myelofibrosis. Cases in which MDS features, such as cytopenias, overlap with MPN features (Chapter 166), such as leukocytosis and splenomegaly, include chronic myelomonocytic leukemia (CMML). The WHO categorizes MDS/MPN overlap syndromes separately from MDS, and these cases have a unique mutational spectrum.

Several different MDS subtypes of varying risk of progression to leukemia are recognized by the WHO; these are listed in Table 182-2. If a patient has cytopenias but does not have characteristic MDS-associated morphologic abnormalities, yet has a typical MDS-associated marrow cytogenetic abnormality such as loss of chromosome 7 or deletion of the long arm of chromosome 5, the patient can be considered to have "unclassifiable MDS" (MDS-U) and does have a risk of progression.

TREATMENT

Given the heterogeneity of MDS and the wide spectrum of functional status and comorbidities present in the typically older population with MDS, treatment must be individualized. Currently, lack of a specific well-characterized molecular target in most patients with MDS precludes a true biologically based approach to therapy. Algorithms for treatment (Fig. 182-2) are based primarily on the patient's prognosis, as discussed further below.

General Approach to the Patient

When choosing a potential therapy for a patient with MDS, clinicians should consider the patient's age, comorbidity and functional status, disease biology (currently based on assessment of bone marrow and peripheral blood findings plus cytogenetic analysis), and pace of evolution of the disease. Some patients with MDS have mild cytopenias and minimal or no symptoms, and they can reasonably be observed without therapy. For higher-risk patients (e.g., International Prognosis Scoring System [IPSS] intermediate 2 or high risk), immediate therapy is usually indicated because patients are usually severely cytopenic and have a life expectancy of less than 2 years. For lower-risk patients (IPSS low and intermediate 1 risk) a more cautious approach is appropriate. The only curative modality for MDS is allogeneic stem cell transplantation, so this option should be considered in developing a therapeutic plan if the patient is young enough and lacks other major medical problems.

Supportive Care

Hematopoietic Growth Factors and Other Approaches to Cytopenias

Because most patients with MDS are anemic, erythropoiesis-stimulating agents (ESA) including epoetin and darbepoetin are often used, even though these are not specifically U.S. Food and Drug Administration (FDA) approved for treatment of MDS. The benefits of ESAs are generally modest and of limited duration, and no prospective data confirm that they prolong survival. Patients with lower baseline serum erythropoietin levels (<500 U/L, and especially <100 U/L) are more likely to respond to ESAs than patients with higher baseline serum erythropoietin levels. An inappropriately low erythropoietin

TABLE 182-2 WORLD HEALTH ORGANIZATION (WHO) SUBTYPES OF MYELODYSPLASTIC SYNDROME

MDS SUBTYPE	DYSPLASIA	BLAST COUNT	OTHER FINDINGS

Refractory cytopenia with unilineage dysplasia (RCUD)	• Present in only 1 lineage (>10% of cells in that lineage):
	  • Erythroid (subtype: refractory anemia)
	  • Myeloid (subtype: refractory neutropenia)
	  • Megakaryocytic (subtype: refractory thrombocytopenia)
		• Peripheral blood: <1%
		• Bone marrow: <5%
			• Cases with bilineage cytopenias may be included in this category, but marrow dysplasia must be limited to 1 lineage
			• Ring sideroblasts represent <15% of erythroid precursors

Refractory anemia with ring sideroblasts (RARS)	• Present in the erythroid lineage (>10% of erythroid precursors)	• Peripheral blood: none • Bone marrow: <5%	• Anemia (normocytic/macrocytic) • Ring sideroblasts comprise ≥ 15% of erythroid precursors
Refractory cytopenia with multilineage dysplasia (RCMD)	• Present in 2 or more lineages (>10% of cells in each affected lineage)	• Peripheral blood: <1% • Bone marrow: <5%	• One or more cytopenias • No Auer rods • May or may not have ≥ 15% ring sideroblasts
Refractory anemia with excess blasts-1 (RAEB-1)	• Present in 1 or more lineages (>10% of cells in each affected lineage)	• Peripheral blood: <5% • Bone marrow: 5-9%	• One or more cytopenias • No Auer rods • 2-4% blasts in the peripheral blood → RAEB-1
Refractory anemia with excess blasts-2 (RAEB-2)	• Present in 1 or more lineages (>10% of cells in each affected lineage)	• Peripheral blood: 5-19% • Bone marrow: 10-19%	• One or more cytopenias • Presence of Auer rods for any blast count <20% → RAEB-2
MDS with isolated deletion of chromosome 5q (5q– syndrome)	• Increased numbers of megakaryocytes, many small and with hypolobated/nonlobated nuclei • Dysplasia in other lineages less common	• Peripheral blood: no or rare blasts (<1%) • Bone marrow: <5%	• Anemia (often macrocytic) with or without other cytopenias/thrombocytosis • No Auer rods • Interstitial or terminal deletion of the long arm of chromosome 5
MDS, unclassifiable (MDS-U)	• Unequivocal, but present in <10% of cells of one or more lineages	• Peripheral blood: ≤1% • Bone marrow: <5%	• May progress to a specific MDS • Can also include cases otherwise classified as RCUD or RCMD but with 1% blasts in peripheral blood; RCUD but with pancytopenia • Cases with an MDS-associated chromosome abnormality (other than loss of Y chromosome or trisomy 8 or deletion of chromosome 20) but without dysplasia can be included in this category

Data based on Swerdlow SH, Campo E, Harris NL, et al, eds. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. 4th ed. Lyon: IARC Press; 2008.

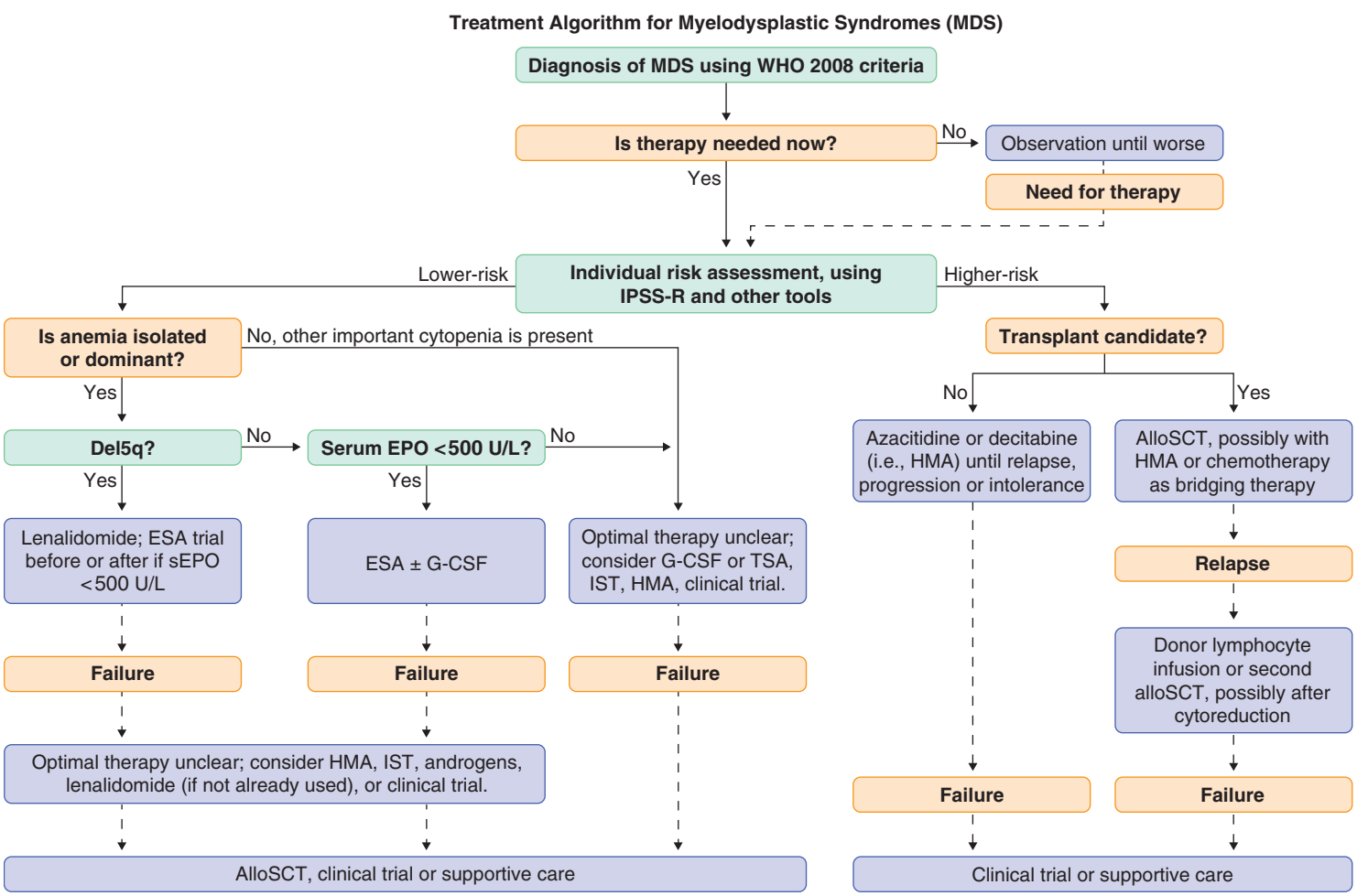

FIGURE 182-2. Suggested treatment algorithm for myelodysplastic syndrome (MDS). AlloSCT = allogeneic stem cell transplant; Del5q = deletion of long arm of chromosome 5 (5q–); EPO = erythropoietin; ESA = erythropoiesis-stimulating agent; G-CSF = granulocyte colony-stimulating factor; HMA = hypomethylating agent; IST = immunosuppressive therapy; sEPO = serum erythropoietin; TSA = thrombopoiesis-stimulating agent; WHO = World Health Organization.

response to a given degree of anemia is more likely in older patients because of the higher frequency of renal insufficiency. If no response is observed after 2 to 3 months of ESA therapy, the trial should be concluded. For patients with refractory anemia with ring sideroblasts (RARS), the addition of granulocyte colony-stimulating factor (G-CSF) to the ESA may potentiate the erythropoietic response.[10] A few patients will benefit from treatment with androgens, but the risk of prostate enlargement in men and other complications such as hepatotoxicity must be considered.

Although myeloid growth factors such as G-CSF can ameliorate the neutropenia often seen in MDS patients, such an approach (probably not leukemogenic, as was once feared) is not likely to reduce the risk of infections or provide other clinical benefit, probably because of the neutrophil dysfunction that may occur in addition to neutropenia. There is no clear role for prophylactic antibiotics.

The thrombopoiesis-stimulating agents romiplostim and eltrombopag have been approved to treat immune thrombocytopenic purpura. These agents have undergone limited study in MDS and are not FDA approved in MDS. They probably should not be used in higher-risk patients with excess blasts, because they could potentially promote leukemogenesis, but a therapeutic trial could be justified in a patient with lower-risk disease with severe thrombocytopenia or bleeding. Platelet transfusions (Chapters 172 and 177) should be used judiciously in MDS patients owing to the risk of alloimmunization, with prophylactic transfusion appropriate only in those with platelet counts less than 5000 to 10,000/mm^3. Patients with mucosal bleeding and refractory thrombocytopenia may benefit from treatment with antifibrinolytic agents such as aminocaproic acid.

Iron Chelation Therapy

The multiple red blood cell (RBC) transfusions needed by some patients with MDS may cause iron overload and tissue injury, but the frequency with which this complication occurs with clinical significance is unclear. Patients with higher serum ferritin levels fare less well than patients with lower values, both in the transplant and nontransplant settings. However, whether the ferritin is simply a marker of inflammation or more advanced disease, or whether the iron deposition in marrow, liver, heart, and pancreas directly results in this adverse outcome, remains uncertain. That uncertainty coupled with the rarity of reported deaths in MDS secondary to iron overload, as well as the toxicity of available chelators, makes the role of iron chelation therapy with agents such as deferasirox or deferoxamine in this disease quite controversial. Although it is reasonable to recommend a trial of iron chelation therapy in patients with lower-risk disease and high transfusional burdens (e.g., >20 to 40 RBC units transfused), there are no prospective data yet to suggest a clear-cut benefit in this setting. Retrospective data suggest that chelated patients live longer than nonchelated patients, but these studies are confounded by patient selection factors.

Disease-Modifying Therapy

Immunosuppressive Therapy

As described earlier, some patients with MDS display a pathophysiology that seems similar to aplastic anemia (Chapter 165), including T-lymphocyte–driven autoimmune attack against hematopoietic cells. Such patients may respond to anti–T-cell therapies, including antithymocyte globulin and calcineurin inhibitors such as cyclosporine or tacrolimus. However, although younger patients with normal karyotype, trisomy 8, paroxysmal nocturnal hemoglobinuria clones, or HLA-DR15 status seem to have a higher likelihood of benefitting, there are currently no good predictive strategies for selecting patients for immunosuppressive therapy.

Immunomodulatory Agents

Early reports showing improvements in anemia in some MDS patients during thalidomide therapy prompted a trial of the immunomodulatory lenalidomide in patients with lower-risk disease. An initial trial and a large phase II trial that followed clearly showed that patients with loss of chromosome 5q (5q–), either alone or in conjunction with other chromosomal abnormalities, had high response rates to lenalidomide, including a nearly 70% transfusion independence rate and better than 30% cytogenetic normalization rate, lasting a median of more than 2 years. Therefore, in the United States, lenalidomide has become the treatment of choice for such

5q– lower-risk MDS patients. Lenalidomide has also been used in higher-risk MDS and AML with 5q– chromosomal abnormalities, but it is less effective.[11] A large randomized trial of two doses of lenalidomide versus placebo shows that AML progression was not increased in lower-risk 5q– MDS with lenalidomide therapy and that responses to a 10-mg dose were more frequent than to a 5-mg dose. In patients with lower-risk MDS without a chromosome 5 abnormality, the response rate to lenalidomide is approximately 25%, which is about the same as one might expect with hypomethylating agents; responses last a median of 8 to 9 months. Important adverse effects of lenalidomide include cytopenias, diarrhea, and rash.

DNA Hypomethylating Agents

Low doses of the nucleoside analogue cytarabine (Ara-C) were used in the 1980s as cytoreductive and so-called differentiating therapy in patients with MDS, with limited effect. Use of low-dose Ara-C has largely been supplanted by azacitidine and decitabine, two azanucleoside analogues that irreversibly inhibit the enzyme DNA methyltransferase, thereby reducing the cytosine methylation "epigenetic" status of DNA and altering gene expression. Whether or not this putative epigenetic mechanism of action accounts for the response to these agents is unproven, but hypomethylating agents have become the mainstay of therapy for most patients with higher-risk MDS, as well as for patients with lower-risk MDS who are refractory to other therapies.

A randomized trial showed that the use of azacitidine at a dose of 75 mg/m^2 subcutaneously daily for 7 days every month (given until disease progression or toxicity), compared to observation, resulted in a delay in time to progression to AML and improved quality of life. Subsequently, an international multicenter trial randomized patients to receive either 7-day azacitidine or conventional care therapies (i.e., doctor's choice of either AML induction chemotherapy, low-dose Ara-C, or supportive care alone) and found a 9-month survival prolongation (from 15 to 24 months) in patients receiving azacitidine, compared with the control arm. These results made azacitidine the standard of care for patients with higher-risk MDS. Cytopenias and gastrointestinal upset are the most common adverse effects of this class of drugs.

Decitabine is also an active agent in patients with MDS. However, a randomized trial of decitabine compared to observation failed to show survival benefit, probably because of a suboptimal dose and schedule of decitabine.

The complete and partial response rates associated with these drugs are low, but at least half of the patients who receive a hypomethylating agent do experience an improvement in at least one cytopenia. Despite these relatively modest responses, a survival benefit is seen even in patients who do not have a complete response.[12] The major problem is that these agents are not curative, and once they fail, the patient's life expectancy is less than 6 months. There are numerous therapies in development that are being tested, but at present there is no standard of care for patients whose disease worsens while undergoing therapy with azacitidine or decitabine.

Stem Cell Transplantation

Because allogeneic stem cell transplantation (Chapter 178) is potentially curative, this modality should be considered in all patients with higher-risk MDS who are approximately 75 years old or younger and lack major comorbid conditions. Reduced-intensity conditioning stem cell transplantation is feasible in adults up to that age, whereas myelosuppressive stem cell transplants are generally reserved for those younger than approximately age 55. Studies considering both types of stem cell transplants suggest that patients with higher-risk disease should be referred immediately for stem cell transplantation if feasible.[13] The optimal conditioning regimen is unclear.

Relapses are common after stem cell transplantation. It is uncertain whether there is benefit from the administration of so-called bridging therapy with azacitidine or decitabine before a planned transplant in an effort to reduce the leukemic burden. Because most patients with MDS are older and uncommonly have a younger sibling who is a suitable stem cell donor, azacitidine or decitabine may be useful to stabilize a patient prior to transplant while an unrelated donor search is ongoing. Although the long-term outcome of patients undergoing stem cell transplant varies widely according to disease and host features,[14] approximately one third of patients with MDS can expect to be long-term disease-free survivors after a stem cell transplant at a cost of 20 to 30% in treatment-related mortality, a figure that may be dropping with time.

Summary

For higher-risk patients, if feasible, urgent transplantation should be the goal, either directly or after exposure to a hypomethylating agent to decrease marrow blasts. If transplant is not the goal, prolonged administration of hypomethylating agents is generally indicated. In lower-risk patients who have a chromosome 5 abnormality, the immunomodulatory agent lenalidomide is the treatment of choice. For other lower-risk patients, supportive care alone, hematopoietic growth factors, immunosuppressive therapy, immunomodulatory therapy, or hypomethylating agents can be considered, depending on the clinical situation.

PREVENTION

At present there is no clear way MDS can be prevented, with the exception of avoiding known precipitants (e.g., using lower doses of radiotherapy or avoiding radiation altogether, and avoiding exposure to alkylating agents or topoisomerase II inhibitors). In the future, it is likely that emerging clonal hematopoiesis following treatment for cancer or arising de novo may be recognized at an early stage. Initiation of immune-based therapies designed to break the lymphocyte tolerance of such abnormal clones may become useful.

PROGNOSIS

Because the natural history of MDS varies widely among patients, several prognostic tools have been derived to help clinicians distinguish patients with a high risk of progression to AML and death from cytopenias within a few months, from those patients whose disease is likely to be more indolent and stable for several years. Among these tools is the 1997 International Prognosis Scoring System (IPSS), which assessed patients' risk by scoring the number of cytopenias, the chromosome pattern (some anomalies are associated with a better prognosis than others), and the proportion of blast cells in the marrow. In 2012, a revised version of the IPSS (IPSS-R) was published; it included a broader range of cytogenetic abnormalities than the original IPSS and weighted abnormal cytogenetics more heavily.[15] The IPSS-R (Table 182-3) defines five different subgroups of MDS with varying risks of death and progression to AML.

Comorbid conditions also influence prognosis independently of the MDS. Other relevant factors that are not included in current prognostic models include serum ferritin and lactate dehydrogenase levels, aberrant expression of certain cell surface markers on myeloid cells detected by flow cytometry (e.g., CD5 or CD56 on myeloid cells), rarer cytogenetic abnormalities not included in the IPSS, and most importantly, the presence of certain molecular abnormalities.[16,17] Several molecular changes, including mutations in *EZH2*, *TP53*, *ASXL1*, *RUNX1*, and *ETV6* are associated with a poorer prognosis than would have been predicted by the IPSS; in patients with lower-risk disease, only *EZH2* retains independent prognostic value. Additionally, the presence of either a *TP53* or a *DNMT3A* mutation predicts for a worse outcome after stem cell transplant.

TABLE 182-3 2012 REVISED INTERNATIONAL PROGNOSTIC SCORING (IPSS-R) FOR MYELODYSPLASTIC SYNDROMES

RISK GROUP	INCLUDED KARYOTYPES	MEDIAN SURVIVAL, YEARS	25% OF PATIENTS TO AML, YEARS	PROPORTION OF PATIENTS IN THIS GROUP
Very good	del(11q), –Y	5.4	N/R	4%
Good	Normal, del(20q), del(5q) alone or with 1 other anomaly, del(12p)	4.8	9.4	72%
Intermediate	+8, del(7q), i(17q), +19, any other single or double abnormality not listed	2.7	2.5	13%
Poor	Abnormal 3q, –7, double abnormality include –7/del(7q), complex with 3 abnormalities	1.5	1.7	4%
Very poor	Complex, with >3 abnormalities	0.7	0.7	7%

TABLE 182-3 2012 REVISED INTERNATIONAL PROGNOSTIC SCORING (IPSS-R) FOR MYELODYSPLASTIC SYNDROMES—cont'd

PARAMETER	CATEGORIES AND ASSOCIATED SCORES				
Cytogenic risk group	Very good 0	Good 1	Intermediate 2	Poor 3	Very poor 4
Marrow blast proportion	≤2% 0	>2-<5% 1	5-10% 2	>10% 3	
Hemoglobin	≥10 g/dL 0	8-<10 g/dL 1	<8 g/dL 1.5		
Absolute neutrophil	≥0.8 × 10^9/L 0	<0.8 × 10^9/L 0.5			
Platelet count	≥100 × 10^9/L 0	50 -100 × 10^9L 0.5	<50 × 10^9/L 1		

Possible range of summed scores: 0-10.

IPSS-R (see: http://www.mds-foundation.org/ipss-r-calculator/)

RISK GROUP	POINTS	% PATIENTS (*n* = 7012; AML DATA ON 6485)	MEDIAN SURVIVAL, YEARS	MEDIAN SURVIVAL FOR PATIENTS UNDER 60 YEARS	TIME UNTIL 25% OF PATIENTS DEVELOP AML, YEARS
Very low	0-1.5	19%	8.8	Not reached	Not reached
Low	2.0-3.0	38%	5.3	8.8	10.8
Intermediate	3.5-4.5	20%	3.0	5.2	3.2
High	5.0-6.0	13%	1.5	2.1	1.4
Very high	>6.0	10%	0.8	0.9	0.7

Top panel, Cytogenetic (chromosome) classification used in IPSS-R. *Middle panel,* Table for calculation of IPSS-R risk group. *Bottom panel,* Outcomes by risk group. Cytogenic (chromosome) classification used in IPSS-R.
Data from Steensma DP. The changing classification of myelodysplastic syndromes: what's in a name? *Hematology* 2009;2009:645-655.

Grade A References

A1. Meerpohl JJ, Schell LK, Rücker G, et al. Deferasirox for managing iron overload in people with myelodysplastic syndrome. *Cochrane Database Syst Rev*. 2014;10:CD007461.
A2. Fenaux P, Giagounidis A, Selleslag D, et al. A randomized phase 3 study of lenalidomide versus placebo in RBC transfusion-dependent patients with low-/intermediate-1-risk myelodysplastic syndromes with del5q. *Blood*. 2011;118:3765-3776.
A3. Silverman LR, Demakos EP, Peterson BL, et al. Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. *J Clin Oncol*. 2002;20:2429-2440.
A4. Fenaux P, Mufti G, Hellstrom-Lindberg E, et al. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. *Lancet Oncol*. 2009;10:223-232.

GRADE REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Walter MJ, Shen D, Ding L, et al. Clonal architecture of secondary acute myeloid leukemia. *N Engl J Med*. 2012;366:1090-1098.
2. Valent P, Bain BJ, Bennett JM, et al. Idiopathic cytopenia of undetermined significance (ICUS) and idiopathic dysplasia of uncertain significance (IDUS), and their distinction from low risk MDS. *Leuk Res*. 2012;36:1-5.
3. Cogle CR, Craig BM, Rollison DE, et al. Incidence of the myelodysplastic syndromes using a novel claims-based algorithm: high number of uncaptured cases by cancer registries. *Blood*. 2011;117:7121-7125.
4. Hahn CN, Chong CE, Carmichael CL, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet*. 2011;43:1012-1017.
5. Schanz J, Steidl C, Fonatsch C, et al. Coalesced multicentric analysis of 2,351 patients with myelodysplastic syndromes indicates an underestimation of poor-risk cytogenetics of myelodysplastic syndromes in the international prognostic scoring system. *J Clin Oncol*. 2011;29:1963-1970.
6. Bravo GM, Lee E, Merchan B, et al. Integrating genetics and epigenetics in myelodysplastic syndromes: advances in pathogenesis and disease evolution. *Br J Haematol*. 2014;166:646-659.
7. Bejar R, Stevenson K, Abdel-Wahab O, et al. Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med*. 2011;364:2496-2506.
8. Olnes MJ, Sloand EM. Targeting immune dysregulation in myelodysplastic syndromes. *JAMA*. 2011;305:814-819.
9. DeZern AE, Sekeres MA. The challenging world of cytopenias: distinguishing myelodysplastic syndromes from other disorders of marrow failure. *Oncologist*. 2014;19:735-745.
10. Malcovati L, Cazzola M. Refractory anemia with ring sideroblasts. *Best Pract Res Clin Haematol*. 2013;26:377-385.
11. Garcia-Manero G. Myelodysplastic syndromes: 2014 update on diagnosis, risk-stratification, and management. *Am J Hematol*. 2014;89:97-108.
12. Adès L, Itzykson R, Fenaux P. Myelodysplastic syndromes. *Lancet*. 2014;383:2239-2252.
13. Koreth J, Pidala J, Perez WS, et al. Role of reduced-intensity conditioning allogeneic hematopoietic stem-cell transplantation in older patients with de novo myelodysplastic syndromes: an international collaborative decision analysis. *J Clin Oncol*. 2013;31:2662-2670.
14. Bejar R, Stevenson KE, Caughey B, et al. Somatic mutations predict poor outcome in patients with myelodysplastic syndrome after hematopoietic stem-cell transplantation. *J Clin Oncol*. 2014;32:2691-2698.
15. Greenberg PL, Tuechler H, Schanz J, et al. Revised international prognostic scoring system for myelodysplastic syndromes. *Blood*. 2012;120:2454-2465.
16. Giagoundis A, Haase D. Morphology, cytogenetics and classification of MDS. *Best Pract Res Clin Haematol*. 2013;26:337-353.
17. Nybakken GE, Bagg A. The genetic basis and expanding role of molecular analysis in the diagnosis, prognosis, and therapeutic design for myelodysplastic syndromes. *J Mol Design*. 2014;16:145-158.

REVIEW QUESTIONS

1. A 69-year-old man presents with complaints of dyspnea on exertion, intermittent palpitations, and excessive fatigue. His history is significant only for gastroesophageal reflux and hypertension, and his medications include omeprazole and metoprolol. There is no family history of cancer or hematologic disorders. Physical examination reveals generalized pallor and regular tachycardia (106 beats/minute.) A complete blood count is as follows:

White blood cell count: 6.1×10^9/L

51% neutrophils, 2% bands, 17% monocytes, 28% lymphocytes, 1% basophils, 1% eos

Hemoglobin 8.1 g/dL; mcv 92 fL

Platelets: 216×10^9/L

B_{12} and folate levels are within normal limits, and serum ferritin is 710 ng/mL. The bone marrow is 70% cellular and exhibits erythroid dysplasia with <5% myeloblasts. Prussian blue reaction demonstrates 55% ring sideroblasts, consistent with a diagnosis of myelodysplastic syndrome, subtype refractory anemia with ring sideroblasts. The karyotype is normal male (46,XY). Which of the following genes is most likely to be mutated in this patient's marrow and blood cells?

A. *BCR-ABL*
B. *SF3B1*
C. *GATA2*
D. *TP53*
E. *JAK2*

Answer: B This patient has refractory anemia with ring sideroblasts (RARS), according to the WHO classification of myelodysplastic syndromes (see Table 182-2). *SF3B1* mutations are strongly associated with RARS (occurring in 60-80% of cases; see Table 182-1). Mutations in the transcription factor *GATA2* are rare germline mutations in patients with familial MDS with monocytopenia; however, this patient has no family history of cancer or hematologic disorders. *TP53* mutation has a very poor prognosis and occurs in treatment-related MDS and is associated with a complex karyotype; however, this patient has not had such drug, toxic, or radiation exposures, and his karyotype is normal (46,XY). The *BCR-ABL* mutation is characteristic of Philadelphia chromosome–positive chronic myelogenous leukemia (CML); however, this patient does not have that clinical phenotype and does not have the Philadelphia chromosome on cytogenetics. *JAK2* mutations (specifically JAK2-V617F) are found in a large percentage of chronic myeloproliferative neoplasms, especially polycythemia vera, and can also be seen in RARS with thrombocytosis; however, the patient has a normal platelet count.

2. A 51-year-old elementary school teacher develops fatigue, dyspnea on exertion, and cold/tingling extremities. Exam reveals generalized pallor and mildly decreased proprioception and hyperreflexia. Past history is significant for hypertension, impaired glucose tolerance, and bariatric surgery 3 years ago for obesity. Her medications include metoprolol, aspirin, a multivitamin, and zinc and echinacea supplements.

The complete blood count is as follows: hemoglobin 10.2 g/dL, MCV 88 fL, white blood cell count 2.9×10^9/L, absolute neutrophil count 380/mm^3, and platelet count 218×10^9/L. Peripheral blood smear shows some hypochromic red cells and neutropenia but is otherwise unrevealing.

Serum B_{12}, red cell folate, serum ferritin, and homocysteine levels are within normal limits. A serum chemistry panel is unremarkable. The bone marrow is normocellular for age, with erythroid dysplasia, 40% ring sideroblasts, granulocytic dysplasia with vacuolization of neutrophil precursors, and no increase in myeloblasts. The karyotype is normal female. What is the most likely cause of her cytopenias?

A. Congenital sideroblastic anemia
B. Myelodysplastic syndrome, subtype refractory anemia with ring sideroblasts
C. Vitamin/mineral deficiency
D. Alcohol abuse
E. Myeloproliferative neoplasm

Answer: C The differential diagnosis in this patient from among the choices offered is quite challenging. Vitamin and mineral deficiencies like advanced folate or B_{12} deficiency can cause dysplastic changes in blood cell precursors in the bone marrow that might be difficult to distinguish morphologically from MDS. In particular, copper deficiency is being increasingly recognized today as a cause of sideroblastic anemia (which this patient has, with 40% ring sideroblasts in the marrow) in individuals who have had bariatric surgery (as in this case). Copper deficiency also causes neuropathy, which appears to be present here. Therefore, the picture is most consistent with copper deficiency secondary to bariatric surgery. MDS (refractory anemia with ringed sideroblasts [RARS]) has many of the features present in this case, including dysplastic changes in both myeloid and erythroid precursors in the marrow, but only drawing a blood copper level and giving a therapeutic trial of copper replacement would rule out the working diagnosis. Congenital sideroblastic anemia can present in adulthood, but it is much less likely here because these inherited anemias typically have changes only in the erythroid series. There is nothing to support the diagnosis of alcohol abuse or a myeloproliferative neoplasm in this case presentation. (See: Steensma DP. Dysplasia has a differential diagnosis: distinguishing genuine myelodysplastic syndromes [MDS] from mimics, imitators, copycats and impostors. *Curr Hematol Malig Rep.* 2012;7:310-320.)

3. A 58-year-old woman is found to be anemic during an evaluation for excessive fatigue. Her history is significant only for cholecystectomy and quiescent rheumatoid arthritis. Other than generalized pallor and borderline regular tachycardia, her exam is unrevealing. A complete blood count is as follows:

Hemoglobin 6.0 g/dL

Mean cell volume 102 fL

White blood count 4.8×10^9/L

Absolute neutrophil count 2.55×10^9/L

Platelet count 466×10^9/L

The reticulocyte count is 0.7%. Serum chemistries and serum B_{12} and red blood cell folate levels are normal. A peripheral blood smear is nondiagnostic. The serum erythropoietin level is measured at 702 U/L.

The patient is admitted to the hospital, where red cell transfusions are administered. There is no evidence of gastrointestinal bleeding. Bone marrow aspirate/biopsy shows a hypercellular marrow with erythroid dysplasia. Megakaryocytes are increased in number, and many are dysplastic with occasional small monolobated forms. The karyotype is 46,XX del(5)(q13q33) [18]/46,XX [2]. Which of the following is the most appropriate initial treatment for this patient?

A. Darbepoetin alfa
B. Lenalidomide
C. Azacitidine
D. Antithymocyte globulin plus cyclosporine
E. Referral for allogeneic stem cell transplantation

Answer: B This patient has the 5q− syndrome, which has a relatively favorable prognosis among the myelodysplastic syndromes. It was the first form of MDS found to be significantly responsive to lenalidomide therapy. Initial trials of lenalidomide in these patients demonstrated high response rates, including a nearly 70% transfusion independence rate and better than 30% cytogenetic normalization rate lasting a median of more than 2 years. (See section on "Immunomodulatory Agents" under "Disease-Modifying Therapy" in the Treatment box.) Therefore, lenalidomide should be first-line treatment for this patient.

4. A 79-year-old man with a history of congestive heart failure due to multivessel coronary artery disease (most recent echocardiogram showed a left ventricular ejection fraction of 30%), renal insufficiency due to hypertension, and two hospitalizations for exacerbations of chronic obstructive pulmonary disease within the last year has developed worse fatigue and dyspnea. His medications include furosemide, lisinopril, metoprolol, aspirin, inhaled salmeterol and fluticasone, and fluvastatin. A complete blood count is as follows:

Hemoglobin 8.8 g/dL

Mean cell volume 103 fL

White blood cell count 1.9×10^9/L

Absolute neutrophil count 0.48×10^9/L

Platelet count 77×10^9/L

Serum vitamin B_{12} and folate levels are within normal limits. The creatinine is 1.8 mg/dL, and blood urea nitrogen is 52; a chemistry group is otherwise unremarkable.

The bone marrow is hypercellular for age with multilineage dysplasia, increased reticulin fibrosis, and 12% blasts, and the marrow karyotype is complex with 4 different chromosome abnormalities and 2 separate clones. Which of the following is the most appropriate treatment for this patient?

A. Induction chemotherapy with daunorubicin and cytarabine
B. Lenalidomide
C. Epoetin alfa
D. Azacitidine
E. Allogeneic stem cell transplantation

Answer: D Azacitidine. This patient has higher-risk MDS, associated with both increased blasts and a high-risk, complex karyotype. He is too old and has too many comorbid conditions to be a viable candidate for allogeneic stem cell transplantation or intensive induction chemotherapy. Azacitidine, a DNA methyltransferase inhibitor, has been demonstrated in a randomized trial to improve survival compared to conventional care (i.e., supportive care or low-dose cytarabine). Patients who have thrombocytopenia and lack chromosome 5q deletion are less likely to respond to lenalidomide.

5. A 19-year-old man was recently found to have pancytopenia (hemoglobin 9.3 g/dL, absolute neutrophil count of 0.45×10^9/L, and platelet count of 23×10^9/L) during routine follow-up of a pediatric cancer. At age 15, he was treated for locally advanced osteosarcoma of the left tibia, with limb sparing surgery, radiotherapy, and adjuvant combination chemotherapy. He sees his pediatric oncologist annually, and at his last visit 9 months ago there was no evidence of recurrence of the sarcoma. Other than fatigue, he is asymptomatic.

His bone marrow aspirate and core biopsy show trilineage dysplasia and 6% myeloblasts, consistent with a diagnosis of MDS, subtype refractory anemia with excess blasts-1. The karyotype is complex and includes 2 marker chromosomes and monosomy 7 in 16 of 20 examined metaphases. He has 3 siblings, and one is fully HLA matched. Which of the following treatments is most appropriate for this patient?

A. Allogeneic stem cell transplantation
B. Lenalidomide
C. Azacitidine
D. Decitabine
E. Epoetin alfa and filgrastim

Answer: A Allogeneic stem cell transplantation. Unfortunately, this patient has developed a late complication of curative therapy for his sarcoma: secondary, therapy-related MDS (t-MDS). t-MDS has a poor prognosis and often progresses rapidly to AML. This patient would benefit from proceeding to allogeneic stem cell transplant as soon as possible, because this is the only potentially curative therapy for his MDS. Since he does not have >10% blasts, there is no clear role for pre-transplant cytoreductive therapy.

183

THE ACUTE LEUKEMIAS

FREDERICK R. APPELBAUM

DEFINITION

Normal hematopoiesis (Chapter 156) requires tightly regulated proliferation and differentiation of pluripotent hematopoietic stem cells that become mature peripheral blood cells. Acute leukemia is the result of a malignant event or events occurring in an early hematopoietic precursor. Instead of proliferating and differentiating normally, the affected cell gives rise to progeny that fail to differentiate but continue to proliferate in an uncontrolled fashion. As a result, immature myeloid cells in acute myeloid leukemia (AML) or lymphoid cells in acute lymphoblastic leukemia (ALL)—often called blasts—rapidly accumulate and progressively replace the bone marrow, diminishing the production of normal red cells, white cells, and platelets. This loss of normal marrow function in turn gives rise to the common clinical complications of leukemia: anemia, infection, and bleeding. With time, the leukemic blasts pour out into the blood stream and eventually occupy the lymph nodes, spleen, and other vital organs. If untreated, acute leukemia is rapidly fatal; most patients die within several months after diagnosis. With appropriate therapy, however, the natural history of acute leukemia can be markedly altered, and many patients can be cured.

EPIDEMIOLOGY

Incidence

There were 14,590 new cases of AML and 6075 new cases of acute ALL in the United States in 2013, leading to 10,320 deaths from AML and 1430 deaths from ALL. The incidence of acute leukemia has remained relatively stable over the past 3 decades. Although acute leukemia accounts for only about 2% of cancer deaths, the impact of leukemia is heightened because of the young age of some patients. For example, with a maximum incidence between ages 2 and 10 years, ALL is the most common cancer in children younger than 15 years and accounts for one third of all childhood cancer deaths. The incidence of AML gradually increases with age, without an early peak. The median age at diagnosis of AML is about 60 years.

Determinants

In most cases, acute leukemia develops for no known reason, but sometimes a possible cause can be identified.

Genetic Predisposition

The concordance rate is virtually 100% in identical twins if one twin develops leukemia during the first year of life. Single germline mutations in *RUNX1*, *CEBPA*, and *GATA2* cause rare syndromes leading to acute leukemia without other manifestations. The incidence of acute leukemia is markedly increased in syndromes involving defective DNA repair, such as Fanconi's anemia and Bloom's syndrome, and in bone marrow failure syndromes associated with ribosomal abnormalities (Chapter 165), including Diamond-Blackfan syndrome, Shwachman-Diamond syndrome, and dyskeratosis congenita. Germline mutations in P53 (Li-Fraumeni syndrome) and abnormalities in chromosome number, as in Down and Klinefelter's syndromes, are also associated with an increased incidence of acute leukemia.

Radiation

Ionizing radiation (Chapter 20) is leukemogenic. The incidence of ALL, AML, and chronic myeloid leukemia (CML) is increased in patients given therapeutic radiation and among survivors of the atomic bomb blasts at Hiroshima and Nagasaki. The magnitude of the risk depends on the dose of radiation, its distribution in time, and the age of the individual. Greater risk results from higher doses of radiation delivered over shorter periods to younger patients. In areas of high natural background radiation (often from radon), chromosomal aberrations are reportedly more frequent, but an increase in acute leukemia has not been consistently found. Concern has been raised about the possible leukemogenic effects of extremely low-frequency nonionizing electromagnetic fields emitted by electrical installations. If such an effect exists at all, its magnitude is small.

Oncogenic Viruses

The search for a viral cause of leukemia has been pursued intensely, but only two clear associations have been found. Human T-cell lymphotropic virus type I (HTLV-I), an enveloped, single-stranded RNA virus, is considered the causative agent of adult T-cell leukemia (Chapter 378). This distinct form of leukemia is found within geographic clusters in southwestern Japan, the Caribbean basin, and Africa. Because HTLV-I seropositivity was found with increasing frequency among heavily transfused patients and intravenous drug users, screening of blood products for antibodies to HTLV-I is now routine practice in blood banks in the United States. Epstein-Barr virus (Chapter 377), the DNA herpes family virus that causes infectious mononucleosis, is associated with the endemic African form of Burkitt's lymphoma/leukemia (Chapter 185).

Chemicals and Drugs

Heavy occupational exposure to benzene and benzene-containing compounds such as kerosene and carbon tetrachloride may lead to marrow damage, which can take the form of aplastic anemia, myelodysplasia, or AML. A link between leukemia and tobacco use has been reported.

With the increasing use of chemotherapy and radiotherapy to treat other malignancies, as much as 10% of AMLs and a smaller percentage of ALLs are likely the consequence of prior therapy.[1] Prior exposure to alkylating agents such as melphalan and the nitrosoureas is associated with an increased risk for secondary AML, which often manifests initially as a myelodysplastic syndrome (Chapter 182), frequently with abnormalities of chromosomes 5, 7, and 8 but with no distinct morphologic features. These secondary AMLs typically develop 4 to 6 years after exposure to alkylating agents, and their incidence may be increased with greater intensity and duration of drug exposure. Secondary AML associated with exposure to topoisomerase II inhibitors, including the epipodophyllotoxins (teniposide or etoposide) and doxorubicin, tends to have a shorter latency period (1 to 2 years), lacks a myelodysplastic phase, has a monocytic morphology, and involves abnormalities of the long arm of chromosome 11 (band q23) or chromosome 21 (band q22). Recently, concern has been raised about an increased risk for second cancers including myeloid malignancies in patients receiving lenalidomide as maintenance therapy in multiple myeloma. Because patients frequently receive combination chemotherapy, it is often difficult to identify a single causative agent.

PATHOBIOLOGY

Clonality and Cell of Origin

The acute leukemias are clonal disorders, and all leukemic cells in a given patient are descended from a common progenitor. The clonal nature of acute leukemia suggests that there are leukemic stem cells capable of both self-renewal and proliferation. Leukemic stem cells in AML are rare among the leukemic mass, with a frequency of 0.2 to 10 per 10^6, and are within the primitive $CD34^{2+}$ $CD38^{-}$ fraction. Less is known about the ALL stem cell.

Classification

The World Health Organization (WHO) classification of acute leukemias is based on clinical, morphologic, immunophenotypic, cytogenetic, and molecular features (Table 183-1).

TABLE 183-1 WORLD HEALTH ORGANIZATION CLASSIFICATION OF ACUTE LEUKEMIAS

CLASSIFICATION SUBTYPES (2008)
Acute Myeloid Leukemia (AML) and Related Neoplasms
AML with recurrent genetic abnormalities
AML with t(8;21)(q22;q22); *RUNX1-RUNX1T1*
AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11*
Acute promyelocytic leukemia (APL) with t(15;17)(q22;q12); *PML-RARA*
AML with t(9;11)(p22;q23); *MLLT3-MLL*
AML with t(6;9)(p23;q34); *DEK-NUP214*
AML with inv(3)(q21q26.2) or t(3;3)(q21;q26.2); *RPN1-EVI1*
AML (megakaryoblastic) with t(1;22)(p13;q13); *RBM15-MKL1*
Provisional entity: AML with mutated *NPM1*
Provisional entity: AML with mutated *CEBPA*
AML with myelodysplasia-related changes
Therapy-related myeloid neoplasms
AML, not otherwise specified
AML with minimal differentiation
AML without maturation
AML with maturation
Acute myelomonocytic leukemia
Acute monoblastic/monocytic leukemia
Acute erythroid leukemia
Pure erythroid leukemia
Erythroleukemia, erythroid/myeloid
Acute megakaryoblastic leukemia
Acute basophilic leukemia
Acute panmyelosis with myelofibrosis
Myeloid sarcoma
Myeloid proliferations related to Down syndrome
Transient abnormal myelopoiesis
Myeloid leukemia associated with Down syndrome
Blastic plasmacytoid dendritic cell neoplasm
B-Lymphoblastic Leukemia (ALL)/Lymphoma
B-lymphoblastic leukemia/lymphoma, not otherwise specified
B-lymphoblastic leukemia/lymphoma with recurrent genetic abnormalities
B-lymphoblastic leukemia/lymphoma with t(9;22)(q34;q11.2); *BCR-ABL1* (Philadelphia chromosome–positive ALL)
B-lymphoblastic leukemia/lymphoma with t(v;11q23); *MLL* rearranged
B-lymphoblastic leukemia/lymphoma with t(12;21)(p13;q22); *TEL-AML1* (*ETV6-RUNX1*)
B-lymphoblastic leukemia/lymphoma with hyperdiploidy
B-lymphoblastic leukemia/lymphoma with hypodiploidy
B-lymphoblastic leukemia/lymphoma with t(5;14)(q31;q32); *IL3-IGH*
B-lymphoblastic leukemia/lymphoma with t(1;19)(q23;p13.3); *TCF3-PBX1*
T-Lymphoblastic Leukemia (ALL)/Lymphoma

ALL = acute lymphoblastic leukemia.

Morphology

Leukemic cells in AML are typically 12 to 20 nm in diameter, with discrete nuclear chromatin, multiple nucleoli, and cytoplasm that usually contains azurophilic granules (Fig. 183-1). Auer rods, which are slender, fusiform cytoplasmic inclusions that stain red with Wright-Giemsa stain, are virtually pathognomonic of AML (Fig. 183-2). The French-American-British (FAB) morphologic system divides AML into eight subtypes: M0, M1, M2, and M3 reflect increasing degrees of differentiation of myeloid leukemic cells; M4 and M5 leukemias have features of the monocytic lineage; M6 has features of the erythroid cell lineage; and M7 is acute megakaryocytic leukemia. The WHO system also recognizes acute basophilic leukemia and acute leukemia with predominant myelofibrosis.

The leukemic cells in ALL tend to be smaller than AML blasts and relatively devoid of granules (see Fig. 183-1). ALL can be divided by FAB criteria into L1, L2, and L3 subgroups. L1 blasts are uniform in size, with homogeneous nuclear chromatin, indistinct nucleoli, and scanty cytoplasm with few, if any, granules. L2 blasts are larger and more variable in size and may have nucleoli. L3 blasts are distinct, with prominent nucleoli and deeply basophilic cytoplasm with vacuoles.

Immunophenotyping

Immunophenotyping by multiparameter flow cytometry is used to determine lineage involvement of newly diagnosed acute leukemias and to detect aberrant immunophenotypes, allowing the measurement of minimal residual disease after therapy. Most cases of AML express antigens seen on normal immature myeloid cells. The most immature forms of AML express CD34,

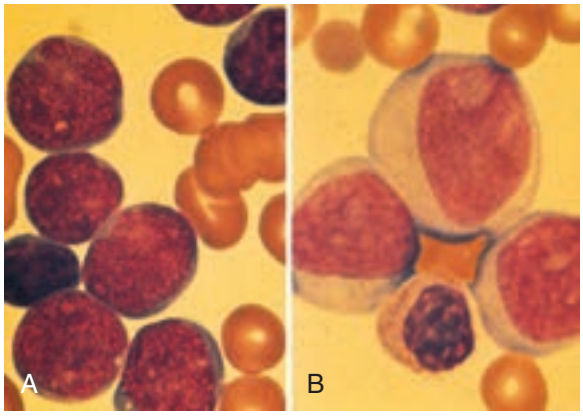

FIGURE 183-1. Acute leukemia. A, Acute lymphoblastic leukemia (ALL). **B,** Acute myeloid leukemia (AML). Lymphoblasts in ALL are smaller, with a higher ratio of nuclear to cytoplasmic material and less distinct nucleoli than in the myeloblasts in AML. The nucleoli in the myeloblasts are clear and "punched out."

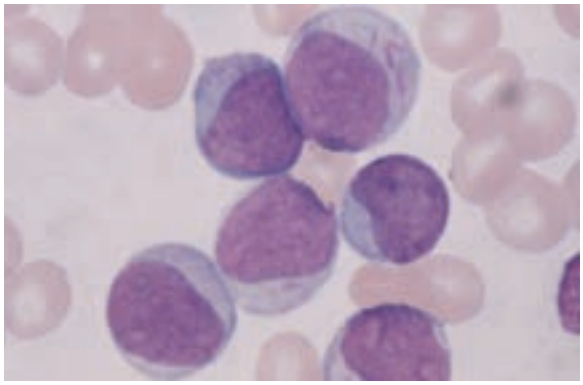

FIGURE 183-2. Acute myeloid leukemia. The myeloblasts in the smear show Auer rods as cytoplasmic inclusions.

CD117 and HLA-DR, whereas more differentiated forms express CD13 and CD33. CD14, CD15, and CD11b are expressed by AMLs with monocytic features, erythroid leukemias express CD36 and CD71, and megakaryocytic AMLs express CD41a and CD61. In 10 to 20% of patients, otherwise typical AML blasts also express antigens usually restricted to B- or T-cell lineage. Expression of a single lymphoid antigen by AML cells does not change either the natural history or the therapeutic response of these leukemias.

Approximately 75% of cases of ALL express B-lineage antigens and can be subdivided into four categories. The most immature group, pro-B ALL, expresses CD19 and/or CD 22 but not CD10 and represents about 10% of cases of ALL. Approximately 50 to 60% of cases of ALL express the early B-cell antigens CD19 and/or CD22 along with the common ALL antigen (CALLA, or CD10), a glycoprotein that is also found occasionally on normal early lymphocytes. CALLA-positive ALL is thought to represent an early pre-B-cell differentiation state. Approximately 10% of cases of ALL have intracytoplasmic immunoglobulin and are termed pre-B-cell ALL. Mature B-cell ALL is signified by the presence of surface immunoglobulin and accounts for less than 5% of cases of ALL. In general, the best therapeutic outcomes among B-cell ALL types are with early pre-B-cell (CALLA positive) ALL. The 25% of cases of ALL that express T-lineage antigens can be separated into three groups: (1) early T-precursor ALL expressing CD7 but not CD1a or CD3, (2) thymic T-ALL expressing CD1a but not surface CD3, and (3) mature T-ALL expressing surface CD3. The prognosis for thymic T-cell ALL is superior to that of the other forms of T-ALL. In about 25% of patients with ALL, the leukemic cells may also express a myeloid antigen, but with current therapies, this does not affect outcome.

Acute leukemias of ambiguous lineage are rare cases with no evidence of lineage differentiation (i.e., acute undifferentiated leukemia [AUL]) or those with blasts that express definitive markers of more than one lineage (i.e., mixed phenotype acute leukemia [MPAL]). MPAL can contain either distinct blast populations of different lineages (bilineal) or a single population expressing features of both lineages (biphenotypic). In general, the prognosis of patients with AUL or MPAL is poor when treated with standard chemotherapy.[2]

Cytogenetics and Molecular Biology

In most cases of acute leukemia, an abnormality in chromosome number or structure is found. These abnormalities are clonal, involving all the malignant cells in a given patient; they are acquired and are not found in the normal cells of the patient; and they are referred to as "nonrandom" because specific abnormalities are found in multiple cases and are associated with distinct morphologic or clinical subtypes of the disease. These abnormalities may be simply the gain or loss of whole chromosomes, but more often they include chromosomal translocations, deletions, or inversions. When patients with acute leukemia and a chromosomal abnormality receive treatment and enter into complete remission, the chromosomal abnormality disappears; when relapse occurs, the abnormality reappears. In many cases, these abnormalities have provided clues into the pathobiology of acute leukemia.

The most common cytogenetic abnormalities seen in AML can be categorized according to their underlying biology and prognostic significance.[3] The translocation t(8;21) and the inversion inv(16) result in abnormalities of a transcription factor made up of core binding factor-α (CBF-α) and CBF-β. The t(8;21) results in the fusion of CBF-α on chromosome 21 with the *MTG8* gene on chromosome 8, whereas inv(16) results in the fusion of CBF-β on the q arm of chromosome 16 with the *MYH11* gene on the p arm. Both of these "core binding factor" AMLs are characterized by a high complete response rate and relatively favorable long-term survival. An additional translocation with a favorable prognosis, t(15;17), involves two genes, *PML* and *RAR-α* (a gene encoding the α-retinoic acid receptor), and is invariably associated with acute promyelocytic leukemia (APL), the M3 subtype of AML. Translocations involving the *MLL* gene, located at chromosome band 11q23, are seen in 5-7% of AMLs. *MLL* is perhaps the most promiscuous oncogene partner in oncology, with more than 30 fusion partners identified. The prognosis of MLL-associated AML depends on the fusion partner, with t(9;11) and t(11;19) predicting an intermediate prognosis and all others considered unfavorable. Trisomy 8 is among the most common nonrandom cytogenetic abnormalities seen in AML; it accounts for 9% of cases and carries an intermediate prognosis. Trisomies of chromosome 21, chromosome 11, and other chromosomes are sometimes seen as well. Deletions of part or all of chromosome 5 or 7 each account for 6 to 8% of cases of AML. These abnormalities are seen with greater frequency in older patients and in patients with AML secondary to myelodysplasia or prior exposure to alkylating agents, and are associated with an unfavorable prognosis. The presence of multiple (more than three) cytogenetic abnormalities in individual cases of AML defines "complex cytogenetics" and also is associated with an unfavorable prognosis.

The identification of recurrent chromosomal abnormalities in acute leukemia, including translocations, inversions, and gene duplications, led to the identification and cloning of the involved genes. More recently, directed and genome-wide assays have provided a better understanding of the genomic landscape of AML.[4] AML cells appear to carry, on average, a total of approximately 13 mutations per cell, far less than found in epithelial cancers. Of these, on average 5 mutations are in genes recurrently mutated in AML (so called driver mutations), with the remainder being considered as passenger mutations. Among the recurrently mutated driver mutations, several are of significant prognostic importance and thus are part of the standard evaluation of AML. *CEBPA*, a gene encoding a leucine zipper transcription factor involved in myeloid differentiation, is mutated in 4 to 15% of cases of AML and is associated with a more favorable prognosis. *NPM1* encodes a nucleolar phosphoprotein with multiple functions. Mutations in *NPM1* are found in approximately 30% of AML cases and are also associated with a more favorable prognosis. *FLT3* is a receptor tyrosine kinase and is mutated in 30 to 35% of AML patients, one fourth of the time as a point mutation and three fourths of the time as an internal tandem duplication. Mutations in *FLT3* are associated with a poorer clinical outcome. Other driver mutations recurrently mutated in AML include *DNMT3A, IDH 1* and *2, NRAS* and *KRAS, RUNX1, TET2,* and *TP53*. Although assays of the mutational status of these genes has not yet become standard, increasing numbers of studies are suggesting their possible utility both for prognosis and for treatment selection.[5]

Whole-genome sequencing of paired samples of skin and bone marrow in individuals who transform from myelodysplastic syndromes to secondary

AML shows that the genetic evolution of secondary AML is a dynamic process shaped by multiple cycles of mutation acquisition and clonal selection. The preexisting myelodysplastic syndrome–founding clone persists with transformation to AML. With the acquisition of each new set of mutations, all the preexisting mutations are carried forward, resulting in daughter subclones that contain increasing numbers of mutation during evolution.[6]

The most common cytogenetic abnormality seen in adults with ALL is the Philadelphia (Ph) chromosome, or t(9;22).[7] This translocation results in fusion of the *BCR* gene on chromosome 22 to the *ABL* tyrosine kinase gene on chromosome 9. This results in the constitutive activation of *ABL*, but the precise mechanism by which this activity leads to leukemia is unclear. The *BCR-ABL* fusion is associated with both ALL and CML (Chapter 184), with a difference in the breakpoint of *BCR* distinguishing the two. A slightly smaller 190-kD fusion protein is usually found in ALL, whereas a larger 210-kD protein is characteristic of CML. The frequency of t(9;22) in ALL increases with age: it is found in approximately 5% of childhood cases and 25% of adults. Before the development of specific tyrosine kinase inhibitors, ALL with t(9;22) had a poor prognosis; newer regimens combining tyrosine kinase inhibitors with chemotherapy are providing improved outcomes. The most common translocation seen in childhood ALL is t(12;21), which involves the genes *TEL* and *AML1*. Like the AML-associated t(8;21) and inv(16), t(12;21) is thought to result in abnormal DNA transcription by interfering with the normal function of CBF. Although t(12;21) is difficult to diagnose by routine cytogenetics, by molecular studies it has been shown to account for 25% of childhood ALL and 4% of adult ALL and has a favorable prognosis. Partial deletions in 9p, seen in 5 to 7% of adults with ALL, are also associated with a favorable outcome. Other abnormalities sometimes seen in B-cell ALL include t(8;14) and t(8;22), which result in translocation of the *MYC* gene on chromosome 8 and immunoglobulin enhancer response genes on chromosomes 14 or 22; they are associated with a poor therapeutic outcome. T-cell ALLs are frequently associated with abnormalities of chromosome 7 or 14 at the sites of T-cell receptor enhancer genes on these chromosomes. The leukemia cells in about 20% of patients with ALL have a propensity to gain chromosomes, sometimes reaching an average of 50 to 60 chromosomes per cell. Patients with such hyperdiploid leukemias tend to respond well to chemotherapy.

As in AML, directed and genome-wide evaluation of cases of ALL have revealed recurrent mutations in addition to those already identified through cytogenetics.[8] Among the most common are mutations in *PAX5* and *IKZF1* seen in 30 and 25% of cases of B-cell ALL respectively, and mutations in *NOTCH1* seen in 35% of cases of T-cell ALL.

CLINICAL MANIFESTATIONS

The signs and symptoms of acute leukemia are usually rapid in onset, developing over a few weeks to a few months at most; they result from decreased normal marrow function and invasion of normal organs by leukemic blasts. Anemia is present at diagnosis in most patients and causes fatigue, pallor, headache, and, in predisposed patients, angina or heart failure. Thrombocytopenia is usually present, and approximately one third of patients have clinically evident bleeding at diagnosis, usually in the form of petechiae, ecchymoses, bleeding gums, epistaxis, or hemorrhage. Most patients with acute leukemia are significantly granulocytopenic at diagnosis. As a result, approximately one third of patients with AML and slightly fewer patients with ALL have significant or life-threatening infections when initially seen, most of which are bacterial in origin.

In addition to suppressing normal marrow function, leukemic cells can infiltrate normal organs. In general, ALL tends to infiltrate normal organs more often than AML does. Enlargement of lymph nodes, liver, and spleen is common at diagnosis. Bone pain, thought to result from leukemic infiltration of the periosteum or expansion of the medullary cavity, is a common complaint, particularly in children with ALL. Leukemic cells sometimes infiltrate the skin and result in a raised, nonpruritic rash, a condition termed *leukemia cutis*. Leukemic cells may infiltrate the leptomeninges and cause leukemic meningitis, typically manifested by headache and nausea. As the disease progresses, central nervous system (CNS) palsies and seizures may develop. Although less than 5% of patients with ALL have CNS involvement at diagnosis, the CNS is a frequent site of relapse; therefore, CNS prophylaxis is an essential component of ALL therapy. Because the incidence of CNS disease is low in AML, there is no proven benefit to CNS surveillance or prophylaxis. Testicular involvement is seen in ALL, and the testicles are a frequent site of relapse. In AML, collections of leukemic blast cells, often referred to as *chloromas* or *myeloblastomas*, can occur in virtually any soft tissue and appear as rubbery, fast-growing masses.

Certain clinical manifestations are unique to specific subtypes of leukemia. Patients with acute promyelocytic leukemia (APL) of the M3 type commonly have subclinical or clinically evident disseminated intravascular coagulation (DIC; Chapter 175) caused by tissue thromboplastins released by the leukemic cells. Acute monocytic or myelomonocytic leukemias are the forms of AML most likely to have extramedullary involvement. M6 leukemia often has a long prodromal phase. Patients with T-cell ALL frequently have mediastinal masses.

DIAGNOSIS

Abnormalities in peripheral blood counts are usually the initial laboratory evidence of acute leukemia. Anemia is present in most patients. Most are also at least mildly thrombocytopenic, and up to one fourth have severe thrombocytopenia (platelets <20,000/μL). Although most patients are granulocytopenic at diagnosis, the total peripheral white cell count is more variable; approximately 25% of patients have very high white cell counts (>50,000/μL), approximately 50% have white cell counts between 5000 and 50,000/μL, and about 25% have low white cell counts (<5000/μL). In most cases, blasts are present in the peripheral blood, although in some patients the percentage of blasts is quite low or absent.

The diagnosis of acute leukemia is typically established by marrow aspiration and biopsy, usually from the posterior iliac crest. Marrow aspirates and biopsy specimens are usually hypercellular and contain 20 to 100% blast cells, which largely replace the normal marrow (see Figs. 183-1 and 183-2). Occasionally, in addition to the blast cell infiltrate, other findings are present, such as marrow fibrosis (especially with M7 AML) or bone marrow necrosis. Marrow samples should also be evaluated by immunophenotyping and cytogenetics. If AML is suspected, samples should be evaluated for the presence of mutations in *FLT3*, *NPM1*, and *CEBPA*. A diagnostic lumbar puncture is generally recommended in suspected cases of ALL, but not in asymptomatic cases of AML.

The prothrombin and partial thromboplastin times are sometimes elevated. In APL, reduced fibrinogen and evidence of DIC are often seen. Other laboratory abnormalities frequently present are hyperuricemia, especially in ALL, and increased serum lactate dehydrogenase. In cases of high cell turnover and cell death (e.g., L3 ALL), evidence of *tumor lysis syndrome* may be noted at diagnosis, including hypocalcemia, hyperkalemia, hyperphosphatemia, hyperuricemia, and renal insufficiency. This syndrome, which is more commonly seen shortly after therapy is begun, can be rapidly fatal if untreated.

Differential Diagnosis

The diagnosis of acute leukemia is usually straightforward but can occasionally be difficult. Both leukemia and aplastic anemia (Chapter 165) can manifest with peripheral pancytopenia, but the finding of a hypoplastic marrow without blasts usually distinguishes aplastic anemia. An occasional patient has hypocellular marrow and a clonal cytogenetic abnormality, which establishes the diagnosis of myelodysplasia (Chapter 182) or hypocellular leukemia. A number of processes other than leukemia can lead to the appearance of immature cells in the peripheral blood. Although other small round cell neoplasms can infiltrate the marrow and mimic leukemia, immunologic markers are effective in differentiating the two. Leukemoid reactions to infections such as tuberculosis can result in the outpouring of large numbers of young myeloid cells, but the proportion of blasts in marrow or peripheral blood almost never reaches 20% in a *leukemoid reaction* (Chapter 167). Infectious mononucleosis (Chapter 377) and other viral illnesses can sometimes resemble ALL, particularly if large numbers of atypical lymphocytes are present in the peripheral blood and the disease is accompanied by immune thrombocytopenia or hemolytic anemia.

TREATMENT Rx

With the development of effective programs of combination chemotherapy and advances in hematopoietic cell transplantation, many patients with acute leukemia can be cured. These therapeutic measures are complex and are best carried out at centers with appropriate support services and experience in treating leukemia. Because leukemia is a rapidly progressive disease, specific antileukemic therapy should be started as soon after diagnosis as possible, usually within 72 hours. The goal of initial chemotherapy is to induce a complete remission (CR) with restoration of normal marrow function. In general, induction chemotherapy is intensive and is accompanied by significant

toxicities. Therefore, patients should be stabilized to the extent possible before specific antileukemic therapy is begun.

Preparing the Patient for Therapy

Severe bleeding usually results from thrombocytopenia, which can be reversed with platelet transfusions (Chapter 177). Once thrombocytopenic bleeding is stopped, continued prophylactic transfusions of platelets is warranted to maintain the platelet count higher than 10,000/μL. Occasionally, patients also have evidence of DIC, usually associated with the diagnosis of M3 AML. If M3 AML is suspected as the cause, all-*trans*-retinoic acid (ATRA) should be started without waiting for molecular confirmation of the diagnosis; the drug can be discontinued if the diagnosis is not M3 AML. If active bleeding is due to DIC (Chapter 175), low doses of heparin (50 U/kg) given intravenously every 6 hours can be of benefit. Platelets and fresh-frozen plasma (or cryoprecipitate) should be transfused to maintain the platelet count higher than 50,000/μL and the fibrinogen level greater than 100 mg/dL until the DIC abates. Whether heparin should be given prophylactically to patients with laboratory evidence of DIC but no active bleeding is an often debated but unsettled question.

Blood cultures should be obtained in patients with fever and granulocytopenia; while awaiting culture results, infection should be assumed and broad-spectrum antibiotics begun empirically (Chapters 167 and 281). It is preferable to bring an infection under control before starting initial chemotherapy if the patient has an adequate granulocyte count. However, patients often have infection but essentially no granulocytes; in this situation, delaying chemotherapy is unlikely to be beneficial.

Patients with very high blast counts may develop symptoms attributable to the effect of masses of these immature cells on blood flow (leukostasis). Although randomized trials are lacking, many experts suggest immediate leukapheresis and administration of hydroxyurea (3 g/m^2/day orally for 2 or 3 days) in an effort to rapidly lower counts in asymptomatic patients with AML presenting with while blood cell counts greater than 100,000/μL and for ALL patients presenting with counts greater than 300,000/μL.[9] The most often affected organs are the CNS and lungs. If CNS symptoms occur, immediate whole-brain irradiation (600 cGy in one dose) should be added. If pulmonary symptoms occur, high-dose corticosteroids are often used.

Before treatment, management in all patients should be aimed at preventing the tumor lysis syndrome. Patients should be hydrated and given allopurinol 100 to 200 mg orally or intravenously three times a day before chemotherapy is initiated. Allopurinol prevents the conversion of xanthine and hypoxanthine to uric acid, but does not affect already formed uric acid. Patients presenting with very high white cell counts may have uremia and anuria secondary to greatly increased serum uric acid levels and intratubular crystallization, even before starting therapy. These patients should be treated with rasburicase 0.20 mg/kg/day for up to 5 days intravenously, given over 30 minutes. Rasburicase promotes the catabolism of already formed uric acid to allantoin. Urinary alkalinization, a past standard, is no longer recommended because, although it increases uric acid solubility, it also increases the possibility of xanthine and calcium phosphate precipitation in the kidney.

The diagnosis of leukemia usually comes as a profound psychological shock to the patient and family. Therefore, in addition to stabilizing the patient hematologically and metabolically, it is worthwhile to have at least one formal conference in which the patient and family are advised about the meaning of the diagnosis of leukemia and the consequences of therapy before treatment is initiated.

Treatment of Acute Lymphoblastic Leukemia

After the patient's condition has been stabilized, antileukemic therapy should be started as soon as possible.[10] Treatment of newly diagnosed ALL can be divided into three phases: remission induction, postremission therapy, and CNS prophylaxis.

Remission Induction

The initial goal of treatment is to induce CR, defined as the reduction of leukemic blasts to undetectable levels and the restoration of normal marrow function. A number of different chemotherapeutic combinations can be used to induce remission; all include vincristine and prednisone, and most add L-asparaginase and daunorubicin, administered over a period of 3 to 4 weeks. With such regimens, CR is achieved in 90% of children and 80 to 90% of adults (Table 183-2). Because vincristine, prednisone, and L-asparaginase are relatively nontoxic to normal marrow precursors, the disease often enters CR after a relatively brief period of myelosuppression. Failure to achieve CR is usually due to either the leukemic cells' resistance to the drugs or progressive infection. These two complications occur with approximately equal frequency.

Postremission Chemotherapy

If no further therapy is given after induction of CR, relapse occurs in almost all cases, usually within several months. Chemotherapy after CR can be given in a variety of combinations, dosages, and schedules. The term *consolidation chemotherapy* refers to short courses of further chemotherapy given at doses similar to those used for initial induction (requiring rehospitalization). Usually, different drugs are selected for consolidation chemotherapy than were used to induce the initial remission. In the case of ALL, such drugs include high-dose methotrexate, cyclophosphamide, and cytarabine, among others. Most regimens include six to eight courses of intensive consolidation therapy. Maintenance involves the administration of low-dose chemotherapy on a daily or weekly outpatient basis for long periods. The most commonly used maintenance regimen in ALL combines daily 6-mercaptopurine and weekly or biweekly methotrexate. The optimal duration of maintenance chemotherapy is unknown, but it is usually given for 2 to 3 years. Maintenance is most beneficial for patients with pro- and pre-B-cell ALL, less so for T-cell ALL, and of no apparent benefit for patients with mature B-cell ALL.

TABLE 183-2 COMMON REGIMENS FOR COMMON FORMS OF ACUTE LEUKEMIA

I. MANAGEMENT OF NEWLY DIAGNOSED ACUTE MYELOID LEUKEMIA

A. Induction—daunorubicin 60-90 mg/m^2/day for 3 days (or idarubicin 10-12 mg/m^2/day for 3 days) and cytarabine 200 mg/m^2/day for 7 days
B. Postremission
   1. Favorable risk—cytarabine 3 g/m^2 over 3 hr q12h on days 1, 3, and 5 every month for 4 mo
   2. Intermediate risk—as for favorable risk; or, if HLA-matched related or unrelated donor exists, allogeneic hematopoietic cell transplantation
   3. Unfavorable risk—proceed to allogeneic transplantation if possible; if not, treat as for intermediate risk

II. MANAGEMENT OF NEWLY DIAGNOSED ACUTE PROMYELOCYTIC LEUKEMIA

A. Induction—ATRA 45 mg/m^2/day until complete remission plus daunomycin 45-60 mg/m^2/day for 3 days and cytarabine 200 mg/m^2/day for 7 days
B. Consolidation #1—arsenic trioxide 0.15 mg/kg/day 5 days/wk for 5 wk; repeat course after 2-wk rest
   Consolidation #2—ATRA 45 mg/m^2/day for 7 days and daunomycin 50 mg/m^2/day for 3 days; repeat course 1 mo later
C. Maintenance—ATRA 45 mg/m^2/day for 15 days every 3 mo plus 6-MP 100 mg/m^2/day and MTX 10 mg/m^2/wk for 2 yr

Or, if intermediate/good risk, consider:

   A. Induction: Arsenic trioxide 0.15 mg/kg/day plus ATRA 45 mg/m^2/day until complete remission
   B. Consolidation: Arsenic trioxide 0.15 mg/kg/day 5 days/wk, 4 wk on, 4 wk off for 4 cycles
      ATRA 45mg/m^2/day 2 wk on, 2 wk off for 7 cycles

III. MANAGEMENT OF NEWLY DIAGNOSED ADULT PH-NEGATIVE ACUTE LYMPHOID LEUKEMIA

A. Induction (and courses 3, 5, 7)—cyclophosphamide 300 mg/m^2 over 3 hr q12h for 6 doses on days 1, 2, 3; doxorubicin 50 mg/m^2 on day 4; vincristine 2 mg/day on days 4 and 11; and dexamethasone 40 mg/day on days 1-4 and days 11-14
B. Consolidation (courses 2, 4, 6, 8)—MTX 200 mg/m^2 over 2 hr, followed by 800 mg/m^2 over 22 hr on day 1; high-dose cytarabine 3 g/m^2 over 2 hr q12h for 4 doses on days 2 and 3
C. Four intrathecal treatments of MTX 12 mg alternating with cytarabine 100 mg are given during the first four courses of systemic therapy

ATRA = all-*trans*-retinoic acid; HLA = human leukocyte antigen; 6-MP = 6-mercaptopurine; MTX = methotrexate; Ph = Philadelphia chromosome [t(9;22)].

Central Nervous System Prophylaxis

Most chemotherapeutic agents that are given intravenously or orally do not penetrate the CNS well, and if no form of CNS prophylaxis is given, at least 35% of adults with ALL will develop CNS leukemia. With prophylaxis, relapse in the CNS as an isolated event occurs in less than 10% of patients. Systemic chemotherapy with high-dose methotrexate (e.g., 200 mg/m^2 intravenously over 2 hours, followed by 800 mg/m^2 over 22 hours) and cytarabine (e.g., 3 g/m^2 over 2 hours every 12 hours for four doses) can achieve therapeutic drug levels within the CNS. Alternatives are intrathecal methotrexate, intrathecal methotrexate combined with 2400 cGy radiation to the cranium, or 2400 cGy to the craniospinal axis.

Treatment of Burkitt-like ALL

Burkitt-like ALL (also called FAB L3 or mature B-cell ALL) is characterized by the presence of monoclonal surface immunoglobulin, cytogenetics showing t(8;14), and the constitutive expression of the *MYC* oncogene. Burkitt-like ALL, which accounts for 3 to 5% of adult cases of ALL, responds well to regimens that incorporate short, intensive courses of high-dose methotrexate (1.5 g/m^2 over 24 hours with leucovorin), cytarabine (3 g/m^2 over 2 hours every 12 hours for four doses), and cyclophosphamide (200 mg/m^2/day for 5 days); this regimen yields high rates of complete response and cures in about 50% of

patients. Recent results suggest that the addition of rituximab may further improve outcomes.

Treatment of Philadelphia Chromosome–Positive ALL

Approximately 5% of pediatric cases and 25% of adult cases of ALL have cytogenetics showing t(9;22), the Ph chromosome. Historically, such patients had CR rates slightly lower than those seen in Ph-negative ALL and markedly reduced remission durations, averaging less than a year; few if any such patients were cured with conventional chemotherapy. Therefore, the general recommendation has been that such patients receive an allogeneic transplant during the first remission, if possible. With this approach, approximately 50% of patients can be cured. More recently, the addition of the tyrosine kinase inhibitor imatinib mesylate to conventional chemotherapeutic regimens has increased complete response rates, equaling those seen in Ph-negative ALL, but the impact of the addition of imatinib mesylate on the duration of remission is not yet known. Dasatinib, a second-generation BCR-ABL tyrosine kinase inhibitor, has demonstrated efficacy in relapsed/refractory adult Ph-positive ALL. Therapy with dasatinib and prednisone alone without chemotherapy appears sufficient to allow the majority of older patients with Ph-positive ALL to achieve an initial hematologic remission.[11]

Prognosis of ALL after Initial Chemotherapy

A number of factors are predictive of outcome in ALL, the most important of which are younger age, a lower white cell count at diagnosis, and favorable cytogenetics. With currently available treatment regimens, 80 to 85% of children and 35 to 40% of adults who initially achieve CR maintain that state for more than 5 years, and these patients are probably cured of their disease.

Treatment of Relapsed ALL

Most relapses occur within 2 years after diagnosis, and most occur in the marrow. Occasionally, relapse is initially found in an extramedullary site such as the CNS or testes. Extramedullary relapse is usually followed shortly by systemic (marrow) relapse and should be considered part of a systemic recurrence. With the use of chemotherapeutic regimens similar to those used for initial induction, 50 to 70% of patients achieve at least a short-lived second remission. A small percentage of patients for whom the initial remission was longer than 2 years may be cured with salvage chemotherapy. If the CNS or testes is the initial site of the relapse, specific therapy to that site is also required, along with systemic retreatment. Because the prognosis of relapsed leukemia treated with chemotherapy is so poor, hematopoietic stem cell transplantation is usually recommended in this setting. Newer agents active in recurrent ALL include nelarabine for T-cell ALL and the anti-CD19 targeted agent blinatumomab and the anti-CD22 immunoconjugate inotuzumab in B-cell ALL. Complete responses have also recently been reported in early trials using chimeric antigen receptor T cells targeting CD19.[12]

Hematopoietic Stem Cell Transplantation

The use of high-dose chemoradiotherapy followed by hematopoietic stem cell transplantation (Chapter 178) from an HLA-identical sibling can cure 20 to 40% of patients with ALL who fail to achieve an initial remission or who have a relapse after an initial CR; it can cure 50 to 60% of patients who undergo transplantation during a first remission. Although there is still considerable debate, several recent studies have reported improved survival for adults with high-risk or standard-risk ALL who receive a hematopoietic stem cell transplant during a first remission rather than being treated with standard chemotherapy.■■ The major limitations of transplantation are graft-versus-host disease, interstitial pneumonia, and recurrence of disease. If an HLA-identical sibling is not available, transplantation from a matched unrelated donor or transplantation of cord blood from a partially matched unrelated donor can be conducted, with results that approach those seen with matched related donors.

Treatment of Acute Myeloid Leukemia

Remission Induction

Treatment with a combination of an anthracycline and cytarabine (100 to 200 mg/m^2/day for 7 days) leads to CR in 60 to 80% of patients with AML. Prospective randomized trials have demonstrated that for patients aged 65 years or less idarubicin (10 to 12 mg/m^2/day for 3 days) or a higher dose of daunorubicin (60 to 90 mg/m^2/day for 3 days) is superior to the conventional daunorubicin dose of 45 mg/m^2/day for 3 days.■■ Profound myelosuppression always follows when these agents are used at doses capable of achieving CR. Failure to achieve CR is usually due to either drug resistance or fatal complications of myelosuppression. In a randomized study of adults with AML, five doses of intravenous gemtuzumab ozogamicin (3 mg/m^2 on days 1, 4, and 7 during induction and day 1 of each of the two consolidation chemotherapy courses) doubled the probability of event-free survival at 2 years.■

Postremission Therapy

Intensive consolidation chemotherapy with repeated courses of daunorubicin and cytarabine at doses similar to those used for induction, high-dose cytarabine (1 to 3 g/m^2/day for 3 to 6 days), or other agents prolongs the average remission duration and improves the chances for long-term disease-free survival. The best results reported to date have generally been achieved with repeated cycles of high-dose cytarabine.[13] Unlike the situation with ALL, low-dose maintenance therapy is of limited benefit after intensive consolidation treatment. In AML, leukemic recurrence occurs less often in the CNS (approximately 10% of cases), most commonly in patients with the M4 or M5 variant. There is no evidence that CNS prophylaxis improves survival in AML.

Prognosis of AML after Initial Chemotherapy

Among patients in whom CR is achieved, 20 to 40% remain alive in continuous CR for more than 5 years, suggesting a probable cure. As with ALL, younger patients and those with a low white cell count at diagnosis have a more favorable outcome. The European LeukemiaNet recognizes four risk categories of AML based on cytogenetics and molecular testing.[14] The favorable group includes those with t(8;21) or inv(16) and those with normal cytogenetics but mutated *CEBPA* or mutated *NPM1* and wild-type *FLT3*. Intermediate I patients are those with normal cytogenetics but without mutations in *NPM1* and *CEBPA*, or with mutations in *FLT3*. Intermediate II includes those with t(9;11), t(11;19), or those with cytogenetics abnormalities not classified as favorable or unfavorable. Finally, the unfavorable group includes those with inv(3), t(6:9), –5 or del5, –7 or del7, and those with complex cytogenetics (Fig. 183-3). Patients with *DNMT3A* and *NPM1* mutations and those with *MLL* translocations appear to selectively benefit from higher dose induction.[15] Patients who have a preleukemic phase before their condition evolves into acute leukemia and those whose leukemia is secondary to prior exposure to chemotherapy have a poorer prognosis. Increased expression of the multidrug resistance gene 1 *(MDR1)* is also associated with a worse outcome.

Treatment of Recurrent AML

Patients whose AML recurs after initial chemotherapy can achieve a second remission in about 50% of cases after retreatment with daunorubicin-cytarabine or high-dose cytarabine. The likelihood of achieving a second

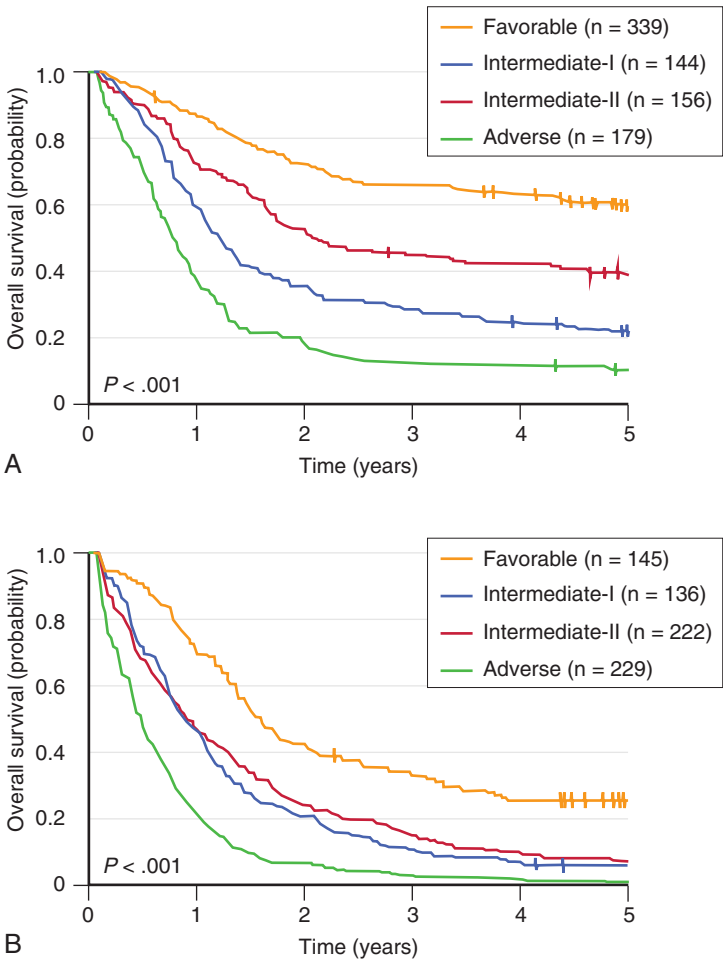

FIGURE 183-3. Overall survival in acute myeloid leukemia according to European LeukemiaNet risk groups. A, Overall survival in patients younger than 60 years. **B,** Overall survival in patients 60 years and older. (From Mrózek K, et al. Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. *J Clin Oncol.* 2012;30:4515-4523).

remission is predicted by the duration of the first remission: 70% in patients whose first remission persisted beyond 2 years, compared with less than 15% in those whose first remission lasted less than 6 months. Older patients may benefit from gemtuzumab ozogamicin (9 mg/m^2 intravenously on days 1 and 15), a form of antibody-targeted chemotherapy. Second remissions tend to be short-lived, however, and few patients in whom relapse occurs after first-line chemotherapy are cured by salvage chemotherapy.

Treatment of Acute Promyelocytic Leukemia

CR can be induced in at least 90% of patients with APL by using ATRA (45 mg/m^2/day orally until CR is achieved) in combination with an anthracycline.[16] Patients treated with ATRA usually have their coagulation disorders corrected within several days. A unique toxicity of ATRA in the treatment of APL is the development of hyperleukocytosis accompanied by respiratory distress and pulmonary infiltrates. The syndrome responds to temporary discontinuation of ATRA and the addition of corticosteroids. By combining ATRA with anthracyclines for induction and consolidation and then using ATRA as maintenance therapy, approximately 70% of patients can be cured. Arsenic trioxide (0.15 mg/kg/day intravenously until CR is achieved) is effective in patients with recurrent APL and appears to improve overall survival if used as part of consolidation therapy for patients in their first CR. The combination of ATRA plus arsenic trioxide without any chemotherapy has recently been shown to be as effective as the standard ATRA-chemotherapy combination for patients with low- or-intermediate risk APL (i.e., those who present with white blood cell counts <10,000/μL).

Hematopoietic Stem Cell Transplantation

For patients with AML in whom an initial remission cannot be achieved or for those who relapse after chemotherapy, hematopoietic stem cell transplantation (Chapter 178) from an HLA-identical sibling offers the best chance for cure. Fifteen percent of patients with end-stage disease can be saved by this treatment. If the procedure is applied earlier, the outcome is better: approximately 30% of patients who undergo hematopoietic stem cell transplantation at first relapse or second remission are cured, and 50 to 60% of patients are cured if hematopoietic stem cell transplantation is performed during the first remission. A large number of studies have prospectively compared the outcome of allogeneic hematopoietic stem cell transplantation with that of chemotherapy in patients with AML in first remission. The trends in all these studies have been toward higher treatment-related mortality but improved disease-free and overall survival time with allogeneic transplantation. Meta-analyses conclude that survival is improved with allogeneic transplantation from a matched sibling in first remission when compared with continued chemotherapy. This improvement is clearest in patients with unfavorable or intermediate-risk disease and is not seen in those in the favorable-risk category. The major limitations of allogeneic hematopoietic stem cell transplantation are lack of a matched sibling donor, graft-versus-host disease, interstitial pneumonia, and disease recurrence. Because transplant-related toxicities increase with patient age, some centers limit hematopoietic stem cell transplantation to those 65 years of age or younger. However, recent studies of reduced-intensity or nonmyeloablative allogeneic transplantation have shown encouraging results in patients with AML in remission at ages up to 75 years. Allogeneic transplantation using matched unrelated donors results in survival essentially equivalent to that seen using matched siblings, although there is a higher incidence of complications.[17] Autologous hematopoietic stem cell transplantation offers an alternative for patients without matched siblings to serve as donors. In randomized trials, the use of autologous transplantation after consolidation chemotherapy significantly prolonged the duration of disease-free survival for patients with AML in first remission but did not alter overall survival.[18]

Treatment of AML in Older Patients

The benefits of intensive consolidation chemotherapy in younger patients do not translate to patients older than 65 years, in part because older patients are less able to tolerate therapy, but also because AML in older patients is more often associated with unfavorable cytogenetics (particularly abnormalities of chromosomes 5 and 7) and more often overexpresses P-glycoprotein, resulting in the multidrug resistance phenotype. Accordingly, long-term survival rates of only 10 to 15% are seen with chemotherapy in patients older than 65 years. Otherwise healthy older patients can usually tolerate intensive chemotherapy and should be offered this treatment. However, intensive chemotherapy can cause more harm than good in older patients with poor performance status. In prospective randomized trials, low-dose cytarabine prolonged survival in older patients unfit for intensive therapy. Other alternative therapies for this group of patients include the demethylating agents azacitidine or decitabine or entry onto clinical trials.[19]

MANAGEMENT OF COMPLICATIONS

Treatment of acute leukemia, especially AML, is accompanied by a number of complications, the two most serious and frequent being infection and bleeding. During the granulocytopenic period that follows induction and consolidation chemotherapy, the risk for bacterial infection is high. A Cochrane review examined results of antibiotic prophylaxis compared with placebo in afebrile neutropenic patients. Antibiotic prophylaxis significantly decreased infection-related deaths, with the best results seen with quinolones. Despite prophylaxis, many patients are febrile while neutropenic, and patients can still develop important infections. The most common bacterial species vary somewhat from medical center to medical center, but *Staphylococcus* (primarily *S. epidermidis*) and *Enterococcus* species are the most frequent gram-positive organisms, whereas *Pseudomonas aeruginosa* and enteric organisms such as *Escherichia coli* and *Klebsiella* species are the most common gram-negative organisms isolated.

Even if no cause for fever is found, bacterial infection should be assumed, and, in general, all patients with fever and neutropenia should receive broad-spectrum antibiotics (Chapters 167 and 281). Monotherapy with an intravenous antipseudomonal agent, such as a carbapenem (e.g., imipenem-cilastatin), a cephalosporin (e.g., cefepime), or an antipseudomonal penicillin (e.g., piperacillin-tazobactam), is recommended as empirical therapy. Vancomycin should not be part of standard coverage in most patients but may be used in those with suspected catheter infections or severe mucositis and in patients with hemodynamic instability or altered mental status. Combination therapy with other gram-negative agents (e.g., aminoglycosides) may be needed. Once begun, antibiotics should be continued until patients recover their granulocyte counts, even if they become afebrile first. If documented bacteremia persists despite appropriate antibiotics, the physician should consider removal of indwelling catheters.

Invasive fungal infections are also common following chemotherapy for acute leukemia and are associated with significant morbidity and mortality. A review of randomized trials found a significant reduction in death from fungal infection in patients given antifungal prophylaxis. Posaconazole is considered by many to be more effective than fluconazole or itraconazole. In addition to being granulocytopenic, patients undergoing induction chemotherapy for leukemia have deficient cellular and humoral immunity, at least temporarily, and therefore are subject to infections common in other immunodeficiency states, including *Pneumocystis jirovecii* (formerly *Pneumocystis carinii*) infection and a variety of viral infections. *P. jirovecii* infection can be prevented by the prophylactic use of trimethoprim-sulfamethoxazole. In cytomegalovirus (CMV) -seronegative patients, CMV-seronegative or leukocyte-reduced blood products should be used to prevent primary infection (Chapter 376). Herpes simplex (Chapter 374) can often complicate existing mucositis and can be prevented with prophylactic acyclovir. Acyclovir is also useful for the prevention and treatment of herpes zoster (Chapter 375).

Myeloid growth factors (granulocyte or granulocyte-macrophage colony-stimulating factor; Chapter 156), if given shortly after the completion of chemotherapy, shorten the period of severe myelosuppression by, on average, 4 days. In most studies, this accelerated recovery has resulted in fewer days with fever and less use of antibiotics, but it has not improved the complete response rate or altered survival.

The platelet count that signals a need for platelet transfusion has been the subject of debate. Traditionally, platelet transfusions from random donors were used to maintain platelet counts greater than 20,000/μL, but more recently it has been demonstrated that lowering this threshold to 10,000/μL is safe in patients with no active bleeding. In contrast, a no-prophylaxis strategy results in more days with bleeding and a shorter time to first bleeding episode and therefore is not recommended. In 30 to 50% of cases, patients eventually become alloimmunized and require the use of HLA-matched platelets (Chapter 177). Transfusion-induced graft-versus-host disease (Chapter 177), manifesting as a rash, low-grade fever, elevated values in liver function tests, and decreasing blood counts, can be prevented by irradiating all blood products before transfusion.

Grade A References

A1. Goldstone AH, Richards SM, Lazarus HM, et al. In adults with standard-risk acute lymphoblastic leukemia, the greatest benefit is achieved from a matched sibling allogeneic transplantation in first complete remission, and an autologous transplantation is less effective than conventional consolidation/maintenance chemotherapy in all patients: final results of the International ALL Trial (MRC UKALL XII/ECOG E2993). *Blood*. 2008;111:1827-1833.

A2. Ram R, Gafter-Gvili A, Vidal L, et al. Management of adult patients with acute lymphoblastic leukemia in first complete remission: systematic review and meta-analysis. *Cancer*. 2010;116:3447-3457.

A3. Fernandez HF, Sun Z, Yao X, et al. Anthracycline dose intensification in acute myeloid leukemia. *N Engl J Med*. 2009;361:1249-1259.

A4. Lowenberg B, Ossenkoppele GJ, van Putten W, et al. High-dose daunorubicin in older patients with acute myeloid leukemia. *N Engl J Med*. 2009;361:1235-1248.
A5. Castaigne S, Pautas C, Terré C, et al. Effect of gemtuzumab ozogamicin on survival of adult patients with de-novo acute myeloid leukaemia (ALFA-0701): a randomised, open-label, phase 3 study. *Lancet*. 2012;379:1508-1516.
A6. Lo-Coco F, Avvisati G, Vignetti M, et al. Retinoic acid and arsenic trioxide for acute promyelocytic leukemia. *N Engl J Med*. 2013;369:111-121.
A7. Koreth J, Schlenk R, Kopecky KJ, et al. Allogeneic stem cell transplantation for acute myeloid leukemia in first complete remission: a systematic review and meta-analysis of prospective clinical trials. *JAMA*. 2009;301:2349-2360.
A8. Gafter-Gvili A, Fraser A, Paul M, et al. Meta-analysis: antibiotic prophylaxis reduces mortality in neutropenic patients. *Ann Intern Med*. 2005;142:979-995.
A9. Stanworth SJ, Estcourt LJ, Powter G, et al. A no-prophylaxis platelet-transfusion strategy for hematologic cancers. *N Engl J Med*. 2013;368:1771-1780.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Klimek VM. Recent advances in the management of therapy-related myelodysplastic syndromes and acute myeloid leukemia. *Curr Opin Hematol*. 2013;20:137-143.
2. Matutes E, Pickl WF, Van't Veer M, et al. Mixed-phenotype acute leukemia: clinical and laboratory features and outcome in 100 patients defined according to the WHO 2008 classification. *Blood*. 2011;117:3163-3171.
3. Grimwade D, Hills RK, Moorman AV, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood*. 2010;116:354-365.
4. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013;369:98.
5. Marcucci G, Haferlach T, Dohner H. Molecular genetics of adult acute myeloid leukemia: prognostic and therapeutic implications [review]. *J Clin Oncol*. 2011;29:475-486.
6. Walter MJ, Shen D, Ding L, et al. Clonal architecture of secondary acute myeloid leukemia. *N Engl J Med*. 2012;366:1090-1098.
7. Gruber M, Wu CJ. Evolving understanding of the CLL genome. *Semin Hematol*. 2014;51:177-187.
8. Roberts KG, Li Y, Payne-Turner D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *N Engl J Med*. 2014;371:1005-1015.
9. Zuckerman T, Ganzel C, Tallman MS, Rowe JM. How I treat hematologic emergencies in adults with acute leukemia. *Blood*. 2012;120:1993-2002.
10. Inaba H, Greaves M, Mullighan CG. Acute lymphoblastic leukaemia. *Lancet*. 2013;381:1943-1955.
11. Foà R, Vitale A, Vignetti M, et al. Dasatinib as first-line treatment for adult patients with Philadelphia chromosome-positive acute lymphoblastic leukemia. *Blood*. 2011;118:6521-6528.
12. Maude SL, Frey N, Shaw PA, et al. Chimeric antigen receptor T cells for sustained remissions in leukemia. *N Engl J Med*. 2014;371:1507-1517.
13. Lowenberg B, Pabst T, Vellenga E, et al. Cytarabine dose for acute myeloid leukemia. *N Engl J Med*. 2011;364:1027-1036.
14. Mrózek K, Marcucci G, Nicolet D, et al. Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. *J Clin Oncol*. 2012;30:4515-4523.
15. Patel JP, Gonen M, Figueroa ME, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med*. 2012;366:1079-1089.
16. Sanz MA, Lo-Coco F. Modern approaches to treating acute promyelocytic leukemia [review]. *J Clin Oncol*. 2011;29:495-503.
17. Gupta V, Tallman MS, He W, et al. Comparable survival after HLA-well-matched unrelated or matched sibling donor transplantation for acute myeloid leukemia in first remission with unfavorable cytogenetics at diagnosis. *Blood*. 2010;116:1839-1848.
18. Pfirrmann M, Ehninger G, Thiede C, et al. Prediction of post-remission survival in acute myeloid leukaemia: a post-hoc analysis of the AML96 trial. *Lancet Oncol*. 2012;13:207-214.
19. Walter RB, Estey EH. Management of older or unfit patients with acute myeloid leukemia. *Leukemia*. 2014;[Epub ahead of print].
20. Fenaux P, Mufti GJ, Hellstrom-Lindberg E, et al. Azacitidine prolongs overall survival compared with conventional care regimens in elderly patients with low bone marrow blast count acute myeloid leukemia. *J Clin Oncol*. 2010;28:562-569.

REVIEW QUESTIONS

1. A 66-year-old man presents with fatigue and is found to have pancytopenia with a hematocrit of 31%, a white blood cell count of 2300/μL with peripheral blasts, and a platelet count of 9000/μL. A bone marrow study shows 60% myeloblasts consistent with M1 acute myelogenous leukemia (AML). The patient had previously been in outstanding health. The choice is made to attempt induction therapy with daunorubicin and cytarabine. Which of the following statements is correct concerning prophylaxis against bleeding during induction?

A. No platelet prophylaxis is necessary. It is sufficient to wait until clinical evidence of bleeding (petechiae or ecchymoses) for transfusing platelets.
B. Prophylactic platelet transfusions from random donors should be given when platelet counts drop below 10,000/μL.
C. Prophylactic platelet transfusions from random donors should be given when platelet counts drop below 20,000/μL.
D. Prophylactic platelet transfusions from HLA-matched donors should be given when platelet counts drop below 10,000/μL.
E. Prophylactic platelet transfusions from HLA-matched donors should be given when platelet counts drops below 20,000/μL.

Answer: B See Stanworth SJ, Estcourt LJ, Powter G, et al. A no-prophylaxis platelet-transfusion strategy for hematologic cancers. *N Engl J Med*. 2013;368:1771-1780.

2. A 24-year-old woman presents with bleeding gums and easy bruising. She is found to have a hematocrit of 29%, a white blood cell count of 4800/μL with 5% peripheral promyelocytes, and a platelet count of 38,000/μL. A bone marrow study shows morphologic and molecular evidence of promyelocytic leukemia. She is begun on ATRA and daunorubicin. Infection prophylaxis includes a quinolone and Bactrim. Six days after starting therapy, the patient presents with fever and increasing shortness of breath that has developed over several days. Her white blood cell count has risen to 15,000/μL with 20% promyelocytes and metamyelocytes. A chest radiograph shows diffuse pulmonary infiltrates. What would be the most appropriate next step?

A. Add daunorubicin to therapy.
B. Add arsenic trioxide to therapy.
C. Perform a bronchoscopy.
D. Add posaconazole for antifungal coverage.
E. Temporarily discontinue the ATRA and add dexamethasone.

Answer: E See Sanz MA, Lo-Coco F. Modern approaches to treating acute promyelocytic leukemia [review]. *J Clin Oncol.* 2011;29:495-503.

3. An 84-year-old man presents with increasing angina and is found to have a hematocrit of 21%, a white blood cell count of 22,000/μL, and a platelet count of 78,000/μL. A bone marrow examination shows AML with complex cytogenetics including monosomy 7. The patient has a known past history of coronary artery disease, hypertension, and peripheral vascular disease. His ECOG performance status is 2. Which of the following would you recommend?

A. Supportive care only
B. Supportive care plus therapy with low-dose cytarabine
C. Imatinib
D. Conventional induction with cytarabine and daunorubicin
E. Conventional induction with cytarabine but with reduced-dose daunorubicin

Answer: B See Fenaux P, Mufti GJ, Hellstrom-Lindberg E, et al. Azacitidine prolongs overall survival compared with conventional care regimens in elderly patients with low bone marrow blast count acute myeloid leukemia. *J Clin Oncol.* 2010;28:562-569.

4. A 44-year-old woman presents with increasing bone pain and pallor and is found to have a hematocrit of 24%, a white blood cell count of 32,000/μL with circulating myeloblasts, and a platelet count of 45,000/μL. A bone marrow study shows AML with t(4;11). Three years earlier, she had been diagnosed with breast cancer and had been treated with surgery, radiation, and adjuvant chemotherapy including cyclophosphamide and doxorubicin. She is treated for her AML with daunorubicin and cytarabine and after one cycle achieves a complete remission. She does not have a matched sibling or a matched unrelated donor. What would you recommend for her further therapy?

A. Three cycles of consolidation therapy using high-dose cytarabine
B. Three cycles of consolidation therapy using high-dose cytarabine combined with gemtuzumab ozogamicin
C. Cord blood transplantation as soon as possible
D. Autologous stem cell transplantation as soon as possible
E. Three cycles of consolidation therapy followed by autologous transplantation

Answer: C See Lowenberg B, Pabst T, Vellenga E, et al. Cytarabine dose for acute myeloid leukemia. *N Engl J Med.* 2011;364:1027-1036.

5. A 74 year-old woman presents with increasing peripheral edema and is found to have a hematocrit of 26%, a white blood cell count of 38,000/μL with circulating lymphoblasts, and a platelet count of 110,000/μL. A bone marrow shows Ph+ acute lymphoblastic leukemia. She has a prior history of compensated congestive heart failure. What therapy would you recommend for her?

A. Supportive care only
B. Supportive care plus low-dose cytarabine
C. Dasatinib plus dexamethasone
D. Vincristine, prednisone, L-asparaginase, daunorubicin, and imatinib
E. Gemtuzumab ozogamicin

Answer: C See Bassan R, Hoelzer D. Modern therapy of acute lymphoblastic leukemia [review]. *J Clin Oncol.* 2011;29:532-543.

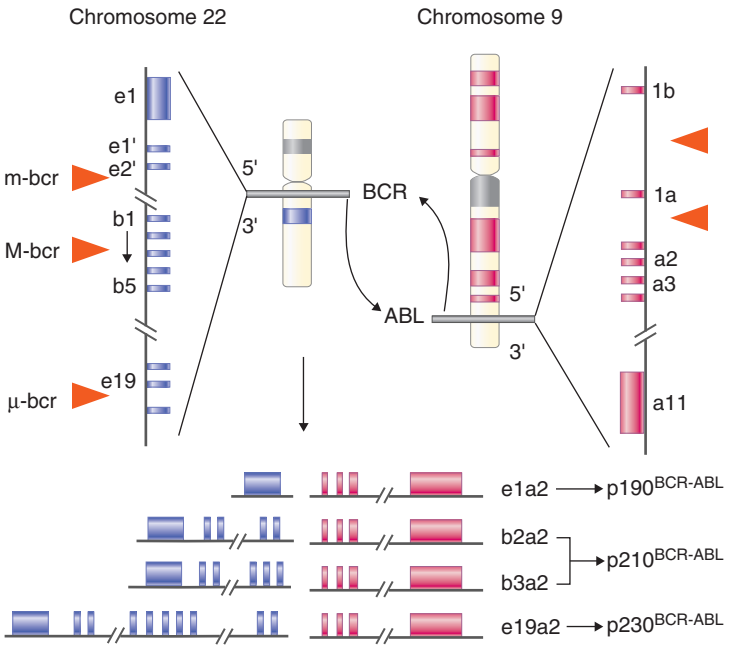

FIGURE 184-1. The Philadelphia chromosome. Originally described as a shortened long arm of chromosome 22, the Philadelphia chromosome (Ph) was later found to be the result of a balanced translocation of genetic material between the long arms of chromosomes 9 and 22: t(9;22)(q34;qll.2). This results in the juxtaposition of *ABL1* to *BCR,* producing a hybrid *BCR-ABL1* oncogene. Depending on the breakpoint on *BCR,* three oncoproteins may be produced: $p210^{BCR-ABLI}$, which is associated with 98% or more of the cases of Ph-positive chronic myelogenous leukemia (CML); $p190^{BCR-ABL1}$, which is associated with 60 to 80% of cases of Ph-positive acute lymphocytic leukemia (the other 20 to 40% of cases are $p210^{BCR-ABL1}$); and $p230^{BCR-ABL}$, which is associated with rare cases of Ph-positive CML.

184

THE CHRONIC LEUKEMIAS

SUSAN O'BRIEN AND ELIAS JABBOUR

CHRONIC MYELOGENOUS LEUKEMIA

DEFINITION

Chronic myelogenous leukemia (CML), also called chronic myeloid leukemia, is a clonal myeloproliferative neoplasm of the primitive hematopoietic stem cell that is characterized by overproduction of cells of the myeloid series, resulting in marked splenomegaly and leukocytosis. Basophilia and thrombocytosis are common. A characteristic cytogenetic abnormality, the Philadelphia (Ph) chromosome, which produces the fusion oncogene *BCR-ABL,* is present in the bone marrow cells in more than 90% of cases. Most patients (85 to 90%) present in the chronic phase. Eventually, if poorly controlled, CML evolves into the accelerated and blastic phases.

EPIDEMIOLOGY

CML constitutes one fifth of all cases of leukemia in the United States. It is diagnosed in 1 or 2 persons per 100,000 per year and has a slight male preponderance. This incidence of 4800 to 5000 cases annually has not changed significantly in the past few decades. The incidence of CML increases with age; the median age at diagnosis is 50 to 55 years. Ph-positive *BCR-ABL*-positive CML is uncommon in children and adolescents. No familial association of CML has been noted; for example, the risk is not increased in monozygotic twins or in relatives of patients with CML. Because of the availability of effective therapy, the annual mortality has been reduced from 15 to 20%, before 2000, to 1 to 2% currently.[1] Thus, the prevalence of CML is predicted to increase gradually, from 15,000 to 20,000 cases before 2000 up to 180,000 cases by 2030 in the United States.

Usually, no etiologic agent is incriminated in CML. Exposure to ionizing radiation (e.g., in survivors of the atomic bomb explosions in Japan in 1945, in those undergoing radiation treatment for ankylosing spondylitis or cervical cancer) increases the risk for CML; the peak incidence occurs 5 to 12 years after exposure and is dose related. No increase in the risk for CML has been demonstrated among individuals working in the nuclear industry. Radiologists working without adequate protection before 1940 were more likely to develop myeloid leukemia, but no such association has been found in recent studies. Benzene exposure increases the risk for acute myelogenous leukemia (AML) but not of CML. CML is not a frequent secondary leukemia after treatment of other cancers with radiation, alkylating agents, or both.

PATHOBIOLOGY

Molecular Pathogenesis

The Ph chromosome abnormality, present in more than 90% of patients with typical CML (Fig. 184-1), results from a balanced translocation of genetic material between the long arms of chromosomes 9 and 22: t(9;22)(q34;q11.2). The breakpoint at band q34 of chromosome 9 results in translocation of the cellular oncogene *ABL1* (previously *c-ABL)* to a region on chromosome 22 coding for the major breakpoint cluster region *(BCR). ABL1* is a homologue of *v-ABL,* the Abelson virus that causes leukemia in mice. This translocation allows juxtaposition of a 5′ portion of a *BCR* and 3′ position of *ABL;* the two genetic sequences produce a new hybrid oncogene *(BCR-ABL1),* which codes for a novel BCR-ABLl oncoprotein with a molecular weight of 210 kD ($p210^{BCR-ABLI}$). The $p210^{BCR-ABLI}$ oncoprotein results in uncontrolled kinase activity of BCR-ABLl, which triggers the excessive proliferation and reduced apoptosis of CML cells, thereby giving CML cells a growth advantage over normal cells and suppressing normal hematopoiesis. Although in most cases 100% of the metaphases on cytogenetic analysis show *BCR-ABL1,* normal stem cells emerge on long-term bone marrow culture and after treatment with interferon-α (IFN-α), imatinib, and other BCR-ABLl-selective tyrosine kinase inhibitors (TKIs).

The constitutive activation of BCR-ABLI results in autophosphorylation and activation of multiple downstream pathways that affect gene transcription, apoptosis, cytoskeletal organization, cytoadhesions, and degradation of inhibitory proteins. The signal transduction pathways implicated involve RAS, mitogen-activated-protein (MAP) kinases, signal transducers and activators of transcription (STAT), phosphatidyl inositol 3-kinase (PI3K), MYC, and others. Many of these interactions are mediated through tyrosine phosphorylation and require binding of the BCR-ABLI to adapter proteins such as GRB-2, CRK, CRK-like protein (CRKL), and SCR homology-containing proteins (SHC). Although imatinib and new-generation TKIs (nilotinib, dasatinib, bosutinib, ponatinib) have been extremely successful at targeting BCR-ABLI, understanding of the pathophysiology of the downstream events of BCR-ABLI is important for the future development of agents that may target these events.

In Ph-positive acute lymphocytic leukemia (ALL), the breakpoint in *BCR* is proximal, in the minor *BCR,* resulting in a smaller *BCR* gene apposing *ABL1*; the resulting fusion gene, messenger RNA, and BCR-ABLI oncoprotein ($p190^{BCR-ABLI}$) are smaller. A third rare, "micro" *BCR* breakpoint distal to the major *BCR* produces a $p230^{BCR-ABLI}$ hybrid oncoprotein, which is associated with a more indolent CML course.

What induces this molecular rearrangement is unknown. Molecular techniques that amplify detection of *BCR-ABL1* have demonstrated its presence in the marrow cells of 25 to 30% of healthy volunteers and 5% of infants, but not in cord blood. Because clinical CML develops in only 1 to 2 of 100,000

individuals (i.e., 1 to 2 per 25,000 to 30,000 individuals who express *BCR-ABL* in their bone marrow), immune regulatory processes or additional molecular events presumably contribute to the development of CML.

BCR-ABL1 is found only in hematopoietic cells and has its origin close to the pluripotent stem cell. The Ph chromosome occurs in erythroid, myeloid, monocytic, and megakaryocytic cells; less commonly in B lymphocytes; rarely in T lymphocytes; and not at all in marrow fibroblasts. The fusion *BCR-ABL1* gene and the p210 protein can be found in cases of morphologically typical CML in which no cytogenetic abnormality occurs or in which changes other than the typical t(9;22) (q34;q11.2) are identified. These patients have a survival rate and a response to therapy similar to those of patients with Ph-positive CML. Patients with atypical CML (usually older and more frequently exhibiting anemia, thrombocytopenia, monocytosis, and dysplasia) who are Ph negative and *BCR-ABL1* negative have a worse prognosis than those who are either Ph positive or Ph negative and *BCR-ABL1* positive; they more closely resemble patients with myelodysplastic syndrome (Chapter 182). Thus, three groups of patients with CML can be identified: (1) those who are positive for Ph and *BCR-ABL1;* (2) those who are Ph negative but *BCR-ABL1* positive; and (3) those who are negative for Ph and *BCR-ABL1*. *PDGFB* (previously *SIS),* which codes for platelet-derived growth factor (PDGF) and is the homologue of the simian sarcoma virus, is also translocated from chromosome 22 to chromosome 9 in CML, but it is distant from the breakpoint and is not expressed.

The pathophysiologic mechanisms underlying CML resistance to TKIs is a fascinating topic that has now been replicated with other targeted therapies in other hematologic and solid tumors. Several mechanisms of resistance to TKIs have been identified; the most common are mutations in the BCR-ABLl kinase domain. More than 100 different mutations have been reported and can involve any of the important domains in the BCR-ABLl structure, including the P-loop (the area where adenosine triphosphate [ATP] binds), the activation loop, and the catalytic domain, as well as the amino acids where imatinib makes contact with BCR-ABLl. The different mutations have considerable variability with respect to resistance to imatinib and other TKIs. Some mutations are overcome by higher concentrations of imatinib than required to inhibit the wild-type form; others are completely insensitive to imatinib. Mutational analysis is useful in patients with imatinib resistance to identify those with the T3151 mutation, who do not respond to imatinib or the second-generation TKIs (dasatinib, nilotinib, bosutinib) but do respond to ponatinib therapy. Knowledge of the sensitivity of the different mutations, as determined by the IC_{50} for particular agents, can help select the TKIs.

CLINICAL MANIFESTATIONS

About 40 to 50% of patients diagnosed with CML do not have symptoms, and the disease is found on routine physical examinations or blood tests. In these patients, the white blood cell (WBC) count may be relatively low at diagnosis. The degree of leukocytosis correlates with tumor burden, as defined by spleen size.

The symptoms of CML, when present, are due to anemia and splenomegaly; they include fatigue, weight loss, malaise, easy satiety, and left upper quadrant fullness or pain. Rarely, bleeding or thrombosis occurs. Other rare presentations include gouty arthritis (from elevated uric acid levels), priapism (usually with marked leukocytosis or thrombocytosis), retinal hemorrhages, and upper gastrointestinal ulceration and bleeding (from elevated histamine levels due to basophilia). Headaches, bone pain, arthralgias, pain from splenic infarction, and fever are uncommon in the chronic phase but more frequent as CML progresses. Symptoms of leukostasis, such as dyspnea, drowsiness, loss of coordination, or confusion, which are due to leukocyte sludging in the pulmonary or cerebral vessels, are uncommon in the chronic phase despite WBC counts exceeding 50×10^9 cells/μL, but these symptoms appear more frequently in the accelerated or blastic phases of the disease.

Splenomegaly, the most consistent physical sign in CML, occurs in 30 to 50% of cases. Hepatomegaly is less common (10 to 20%) and usually minor. Lymphadenopathy is uncommon, as is infiltration of skin or other tissues. If present, these findings suggest Ph-negative CML or the accelerated or blastic phase of CML.

DIAGNOSIS

The diagnosis of typical CML is not difficult. Patients with untreated CML usually have leukocytosis ranging from 10 to 500×10^9/μL. The predominant cells are neutrophils, with a left shift extending to blast cells. Basophils and eosinophils are commonly increased. Monocytes may be slightly increased in some cases that overlap with chronic myelomonocytic leukemia (CMML; see later discussion). Thrombocytosis is common, whereas thrombocytopenia is rare and, if present, suggests a worse prognosis. A hemoglobin level of less than 11 g/dL is present in one third of patients. Some patients demonstrate a cyclic oscillation of the WBC count. The presence of unexplained myeloid leukocytosis (Fig. 184-2) with splenomegaly should lead to a bone marrow examination and cytogenetic and molecular analysis.

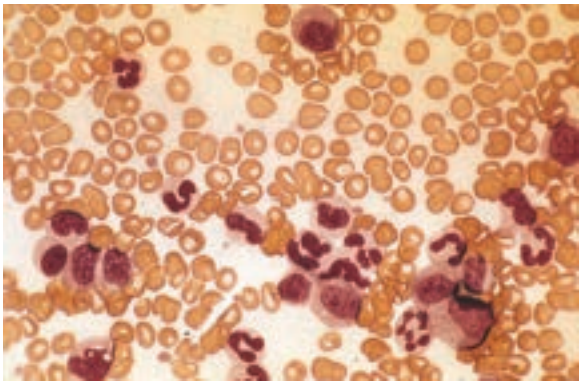

FIGURE 184-2. Chronic myelogenous leukemia, chronic phase. Peripheral smear shows leukocytosis, with representation by the entire spectrum of leukocyte differentiation, ranging from myeloblasts to mature neutrophils. (Courtesy Andrew Schafer, MD.)

Bone Marrow

The bone marrow is hypercellular, with marked myeloid hyperplasia and, at times, evidence of increased reticulin or collagen fibrosis. The myeloid-erythroid ratio is 15 : 1 to 20 : 1. About 15% of patients have 5% or more blast cells in the peripheral blood or bone marrow at diagnosis.

Cytogenetics

The presence of the t(9; 22) (q34; q11.2) abnormality establishes the diagnosis of CML. If the Ph chromosome is not found in a patient with suspected CML, molecular studies for the presence of the hybrid *BCR-ABL1* gene should be performed. About 25 to 30% of patients with a typical morphologic picture of CML who are Ph negative have the *BCR-ABL1* rearrangement. The Ph chromosome is usually present in 100% of metaphases, often as the sole abnormality. Between 10 and 15% of patients have additional chromosomal changes (loss of the Y chromosome, trisomy 8, an additional loss of material from 22q, or double Ph). Some patients have complex chromosomal changes involving chromosome 9 or chromosome 22 (Ph variants, three-way translocations).

Differential Diagnosis

CML must be differentiated from leukemoid reactions (Chapter 167), which usually produce WBC counts lower than 50×10^9/μL, and toxic granulocytic vacuolation, Döhle bodies in the granulocytes, absence of basophilia, and normal or increased leukocyte alkaline phosphatase (LAP) levels (which are typically low in CML). The clinical history and physical examination generally suggest the origin of the leukemoid reaction. Corticosteroids can rarely cause extreme neutrophilia with a left shift, but this abnormality is self-limited and of short duration.

CML may be more difficult to differentiate from other myeloproliferative or myelodysplastic syndromes (Chapters 166 and 182). Patients with myeloid metaplasia with or without myelofibrosis frequently have splenomegaly, neutrophilia, and thrombocytosis. Polycythemia vera with associated iron deficiency, which causes normal hemoglobin and hematocrit values, can manifest with leukocytosis and thrombocytosis. Such patients usually have a normal or increased LAP score, a WBC count less than 25×10^9/μL, and no Ph chromosome or *BCR-ABL1* rearrangement.

The greatest diagnostic difficulty lies with patients who have splenomegaly and leukocytosis but do not have the Ph chromosome. In some, the *BCR-ABL* hybrid gene can be demonstrated despite a normal or atypical cytogenetic pattern. Patients who are Ph negative and *BCR-ABLI* negative are considered to have Ph-negative CML or CMML (see later discussion). Isolated megakaryocytic hyperplasia can be seen in essential thrombocythemia (Chapter 166), with marked thrombocytosis and splenomegaly. Some patients who present with clinical characteristics of essential thrombocythemia (with marked thrombocytosis but without leukocytosis) have CML; cytogenetic

and molecular studies showing the Ph chromosome, the *BCR-ABLI* rearrangement, or both lead to the appropriate diagnosis and treatment.

Rarely, patients have myeloid hyperplasia, which involves almost exclusively the neutrophil, eosinophil, or basophil cell lineage. These patients are described as having chronic neutrophilic, eosinophilic, or basophilic leukemia and do not have evidence of the Ph chromosome or the *BCR-ABLI* gene but may have other molecular abnormalities. Most patients with *chronic neutrophilic leukemia*, characterized clinically by sustained, mature neutrophilic leukocytosis, hepatosplenomegaly, and bone marrow granulocytic hyperplasia, have oncogenic mutations in the gene for colony-stimulated factor 3 receptor (*CSF3R*).[2] Patients with *chronic eosinophilic leukemia* (or clonal hypereosinophilic syndrome) (Chapter 170) have mutations in the genes for platelet-derived growth factor receptor-aα (*PDGFRA*) or *PDGFRB* or *FGFR1*; the prototypical abnormality is the *FIP1L1-PDGFRA* gene fusion.[3] Peripheral blood or bone marrow can be tested for these genetic markers by reverse transcription–polymerase chain reaction (RT-PCR) or interphase/metaphase fluorescent in situ hybridization (FISH), as can the *BCR-ABL1* rearrangement for CML.

Clinical Course

Evolution to Accelerated and Blastic Phases

More than 90% of patients present with CML in the benign or chronic phase. If symptomatic at presentation, it becomes asymptomatic once the disease is controlled. Death rarely occurs during the chronic phase of CML. When poorly controlled, CML evolves into an accelerated phase, usually defined by the presence of 15% or more blasts, 30% or more blasts plus promyelocytes, 20% or more basophils, thrombocytopenia (platelets $<100 \times 10^9/\mu L$) unrelated to therapy, or cytogenetic clonal evolution. The accelerated phase can be also characterized by worsening anemia; increasing splenomegaly or hepatomegaly; infiltration of nodes, skin, bones, or other tissues; and fever, malaise, and weight loss. In the accelerated phase, bone marrow studies may show dysplastic changes, increased percentages of blasts and basophils, myelofibrosis, and chromosomal abnormalities in addition to the Ph chromosome (clonal evolution). About 5 to 10% of patients present in the accelerated phase.

Before the era of imatinib therapy, the risk for developing accelerated or blastic phase CML was 10% per year in the first 2 years after diagnosis and 15 to 20% per year thereafter, unless therapies such as IFN-α or allogeneic hematopoietic stem cell transplantation (AHSCT) were used. With imatinib, the annual incidence of progression of CML from the chronic to the accelerated or blastic phase has been 2% in the first 10 years of observation (Fig. 184-3). Before imatinib therapy, the median survival of accelerated phase CML was 18 months or less, but survival has now increased to 4 years or more. In de novo accelerated phase CML, the estimated 8-year survival rate with TKI therapy is 75%.

The blastic phase of CML is diagnosed when 30% or more blast cells are present in the bone marrow and/or peripheral blood or when extramedullary blastic disease is present. Most patients develop features of the accelerated phase before progressing to the blastic phase, but 20% of patients evolve quickly into a blastic phase without warning. Most patients in the accelerated or blastic phase have additional chromosomal abnormalities (clonal evolution) such as duplication of the Ph chromosome, trisomy of chromosome 8, or development of an isochromosome 17. The extramedullary blastic phase of CML can occur in the spleen, lymph nodes, skin, meninges (especially in the lymphoid blastic phase), bones, and other sites; extramedullary transformation is usually followed shortly by evidence of marrow involvement. Blastic phase CML is associated with a very poor median survival time of 5 months. About 25% of patients develop a lymphoid blastic phase. Ph-negative and *BCR-ABL1*-negative CML often appear to overlap clinically with CMML in their behavior, progress, and response to therapy and seem to resemble the myelodysplastic syndromes (Chapter 182) more than Ph-positive CML. A male preponderance and older age are noted; splenomegaly is common (60 to 70%).

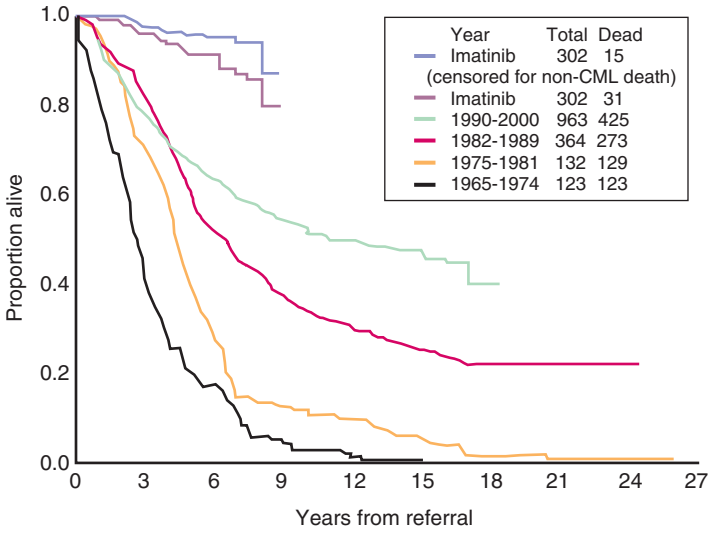

FIGURE 184-3. Survival of patients with Ph-positive chronic myelogenous leukemia (CML) in chronic phase CML. In single-institutional studies, industry-sponsored trials, and national reports where TKI therapy has full penetration, the estimated 8- to 10-year survival rates are high. The survival rates are lower where TKI treatment penetration is not optimal. **A,** From Bjorkholm M, Ohm L, Eloranta S, et al. Success story of targeted therapy in chronic myeloid leukemia: a population-based study of patients diagnosed in Sweden from 1973-2008. *J Clin Oncol.* 2011;29:2514-2520. **B,** From Chen Y, Wang H, Kantarjian H, et al. Trends in chronic myeloid leukemia incidence and survival in the United States from 1975-2009. *Leuk Lymphoma.* 2013;54:1411-1417.

TREATMENT

Today, there are five TKIs that have been approved for the treatment of CML. Imatinib mesylate was approved in 2001 by the U.S. Food and Drug Administration (FDA) for CML salvage and in 2002 for CML firstline therapy. Nilotinib, a more potent selective BCR-ABL1 TKI, was approved in 2007 for CML salvage and in 2010 for first-line CML therapy. Dasatinib, a dual BCR-ABL TKI, was approved for CML salvage therapy in 2006 and for first-line therapy in 2010. In 2012, two additional TKIs were approved for CML salvage: bosutinib, a dual SRC-ABL1 inhibitor, and ponatinib, a pan BCR-ABL1 kinase inhibitor with selective potency against the resistant T315I mutation. In 2012, omacetaxine mepesuccinate, a semisynthetic cephalotaxine that inhibits protein synthesis, was also approved for the treatment of CML following failure of two or more TKIs.

First-line therapy for CML today includes imatinib, nilotinib, or dasatinib.[4,5] Patients who demonstrate CML resistance or intolerance to treatment may be offered salvage TKI therapy with any of the other available TKIs. The choice of second-line therapy with a TKI versus allogeneic hematopoietic stem cell transplantation (AHSCT) depends on several factors: (1) the patient's age and general condition; (2) the availability of acceptable donors (related or matched unrelated); (3) whether a patient has intolerance or CML resistance to first-line TKI therapy; (4) the emerging mutation in the resistant CML clone; (5) whether there is additional clonal evolution at the time of relapse; (6) the response to second-line therapy with the new TKI; (7) the estimated safety and success of the AHSCT; and (8) additional comorbid conditions of the patients (e.g., diabetes, pulmonary conditions, cardiac status, prior history of pancreatitis or pulmonary hypertension) (Chapter 178).

Patients who present with or develop accelerated or blastic phase should receive second-line TKIs to reduce the disease burden and should be offered AHSCT as soon as possible (the exception possibly being de novo CML accelerated phase which may respond durably to first-line TKI therapy, particularly with achievement of a complete cytogenetic response). Patients who develop intolerance to first-line TKI in chronic phase could be offered second-line TKIs as durable therapy, particularly if they achieve complete cytogenetic response. Patients who develop resistance to a first-line TKI in chronic phase are offered a second-line TKI based on their mutation analysis. A T3151 mutation in the CML clone requires therapy with ponatinib and (as of today) early consideration of AHSCT until the results of ponatinib therapy mature. Mutations involving V299L, T315A, or F317LN/I/C are sensitive to nilotinib therapy. Mutations involving Y253H, E255KN, or F359N/C/I are sensitive to dasatinib and bosutinib therapy. Patients who harbor clonal evolution in the CML cells (additional chromosomal abnormalities in the Ph positive cells) or mutations at the time of second-line therapy, or those who do not achieve a complete cytogenetic response by 1 year of second-line TKI therapy, should be considered early for AHSCT. However, if there is no clonal evolution or mutations at the time of second-line therapy, and if the patient achieves complete cytogenetic response with second-line TKI therapy, responses are durable, and TKIs can be continued until evidence of cytogenetic relapse before AHSCT is considered as third-line therapy. Older patients (e.g., 65 to 70 years or older) may forgo a curative option of AHSCT in favor of several years of good disease control. In such patients, TKI therapy with or without additional (older) agents (hydroxyurea, cytarabine, decitabine, 6-mercaptopurine) may sustain less than a complete cytogenetic response (partial, minor) or a complete hematologic response for many years with a good quality of life, and without the risk for mortality or morbidities associated with AHSCT, particularly if the donor is not

optimal (unrelated, mismatched). Treatment options and choice and timing of AHSCT in CML are detailed in Tables 184-1 and 184-2.

Recent trials with different TKIs have indicated that early and "deep" cytogenetic and molecular responses are predictive of improved progression-free and overall survival.[6]

Imatinib Mesylate

Since its discovery in 1999, imatinib mesylate has become standard therapy for CML. Imatinib is a 2-phenylaminopyrimidine derivative that binds to the canonical ATP lining the groove between the N and C lobes of the ABL1 kinase domain, thus blocking the phosphorylation of tyrosine residues on substrate protein. Blocking of ATP binding inactivates the ABLl kinase because it cannot transfer phosphate to its substrate. By inhibiting phosphorylation, imatinib prevents the activation of signal transduction pathways that induce the leukemic transformation processes that cause CML (Fig. 184-4). Imatinib inhibits several tyrosine kinases, including $p210^{BCR\text{-}ABLl}$, $p190^{BCR\text{-}ABLl}$, v-ABL, c-ABL, c-Kit, and PDGF receptor.

In a randomized trial of 1106 patients with newly diagnosed CML, imatinib 400 mg/day orally provided significantly higher rates of major cytogenetic response (87% vs. 35%) and complete cytogenetic response (76% vs. 14%), as well as lower rates of progression (8% vs. 26%) and transformation (3% vs. 9%) after 12 months of therapy, compared with the prior standard nontransplantation therapy (a combination of IFN-α and cytosine arabinoside).[7] The longer term follow-up results continue to demonstrate excellent outcomes with imatinib therapy (Table 184-3; see Fig. 184-3); with a median follow-up of 8 years, the complete cytogenetic response rate (occurring at least once during therapy) is 83%, the estimated 8-year event-free survival rate is 81%, and the transformation-free survival rate 92%. The estimated 8-year survival rate is 85% (93% if only CML-related deaths are included). The annual rate of transformation was 1.5 to 2.8% in the first 3 years and decreased to less than 1% in the subsequent 5 years among patients who continued on imatinib therapy. Therapy with high-dose imatinib or with combinations of imatinib and other agents (e.g., peg-IFN-α2) did not show convincingly improved results compared with standard imatinib 400 mg daily.

Imatinib has a 5% or lower rate of serious side effects, which include nausea, vomiting, diarrhea, skin rash, muscle cramps, bone pain, periorbital or leg edema, weight gain, and rarely, hepatic, renal, or cardiopulmonary dysfunction; most of these are manageable with dose reduction or treatment interruption. Drug-related myelosuppression occurs in 10 to 20% of patients with newly diagnosed CML and is manageable with brief treatment interruptions, dose modifications, or both, or with the administration of growth factors (erythropoietin for anemia, granulocyte colony-stimulating factor for neutropenia). Imatinib and other TKIs may prolong the cardiac QTc interval; medications that contribute to QTc prolongation should be avoided. Hypophosphatemia associated with altered bone metabolism can occur, and the serum phosphate level should be monitored. Chromosomal abnormalities may appear in the Ph-negative diploid cells in 5 to 10% of responding patients, probably owing

TABLE 184-1 THERAPY OF CHRONIC MYELOID LEUKEMIA

First line	Imatinib 400 mg daily Nilotinib 300 mg twice daily Dasatinib 100 mg daily
Second/third line	Nilotinib, dasatinib, bosutinib, ponatinib Omacetaxine Allogeneic stem cell transplantation
Other	Decitabine, pegylated interferon Hydroxurea, cytarabine, decitabine Combinations of tyrosine kinase inhibitor–based regimens Investigational

TABLE 184-2 ROLE AND TIMING OF ALLOGENEIC STEM CELL TRANSPLANTATION IN CHRONIC MYELOID LEUKEMIA

CHRONIC MYELOID LEUKEMIA STATUS	TYROSINE KINASE INHIBITOR THERAPY	ALLOGENEIC STEM CELL TRANSPLANTATION
Accelerated or blastic phase	Interim therapy to minimal residual disease	As soon as possible
Imatinib failure in chronic phase with T3151 mutation	Ponatinib interim therapy to minimal residual disease	As soon as possible if no good response obtained
Imatinib failure in chronic phase; no clonal evolution, no mutations, good initial response	Long-term second-line tyrosine kinase inhibitors	Third-line post second tyrosine kinase inhibitors failure
Imatinib failure in chronic phase with clonal evolution or mutation or no cytogenetic response to second-line tyrosine kinase inhibitors	Interim therapy to minimal residual disease	Second line
Older (>65-70) post-imatinib failure in chronic phase	Long-term	May forgo allogeneic stem cell transplantation for many years of quality of life

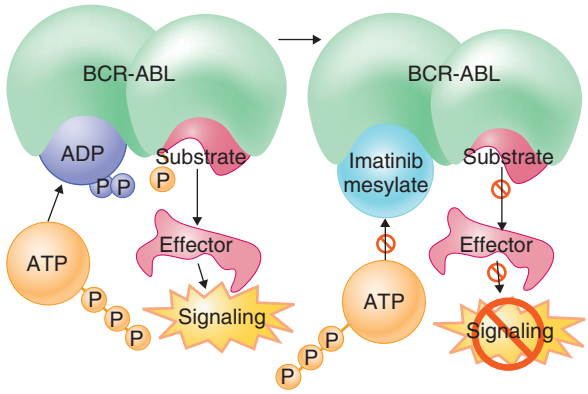

FIGURE 184-4. Mechanism of action of imatinib. By occupying the adenosine triphosphate (ATP)-binding pocket of the ABL kinase domain, imatinib prevents substrate phosphorylation and downstream activation of signals, thus inhibiting the leukemogenic effects of BCR-ABL on cells in chronic myelogenous leukemia. ADP = adenosine diphosphate; P = phosphate group.

TABLE 184-3 RESULTS OF TYROSINE KINASE INHIBITOR THERAPY IN CHRONIC PHASE CHRONIC MYELOGENOUS LEUKEMIA

THERAPY	LEUKEMIA STATUS	COMPLETE CYTOGENETIC RESPONSE (%) (AT INDICATED YEAR OF TREATMENT)	MAJOR/COMPLETE MOLECULAR RESPONSES (%) (AT INDICATED YEAR OF TREATMENT)	SURVIVAL (%) (AT INDICATED YEAR AFTER INITIATION OF TREATMENT)
Imatinib	First line	65 (5)	40/20 (5)	85 (8-10)
Nilotinib	First line	85-87 (4)	73-76/37-40 (4)	95-97 (4)
Dasatinib	First line	86 (2)	74/34 (4)	93 (4)
Dasatinib	Salvage	50 (5)	43 (6)	71 (6)
Nilotinib	Salvage	44 (4)	30-40 (4)	85 (3)
Ponatinib	Salvage	45-65 (2)	30-50 (2)	90 (2)
Bosutinib	Salvage	40 (2)	30 (2)	92 (2)
Omacetaxine	Salvage	10 (2)	—	85-90 (2)

to unmasking of a fragile stem cell prone to the development of CML or to chromosomal instability; such changes disappear spontaneously in 70% of cases and rarely evolve into a myelodysplastic syndrome or acute myeloid leukemia, probably as part of the natural course of CML.

Nilotinib

Nilotinib, a selective BCR-ABLI TKI 30 times more potent than imatinib, was initially approved for the treatment of CML after imatinib failure. In CML chronic phase after imatinib failure, nilotinib 400 mg orally twice daily was associated with complete cytogenetic response rates of 40 to 50%. The major molecular response rates, defined as BCR-ABL1 transcripts, 0.1% by International Scale [IS], are 30 to 40%, and the estimated 3-year survival rate is 80%. Subsequent studies compared nilotinib to imatinib in newly diagnosed patients with CML. In a three-arm randomized study,[6] patients received imatinib 400 mg daily, nilotinib 400 mg twice a day, or nilotinib 300 mg twice daily. With a minimum follow-up of 4 years, the two arms of nilotinib demonstrate better early results compared with imatinib. There was no difference in the estimated 4-year survival rates (94%, 97%, 93%, respectively). Nilotinib therapy was associated with lower rates of fluid retention, diarrhea, headaches, muscle cramps, nausea and vomiting, and neutropenia. However, it was associated with higher rates of headache, rash, pruritus, and hyperglycemia, and with a low but notable incidence of pancreatitis (<2%), ischemic heart disease (4 to 5% vs. 1% with imatinib), and peripheral arterial occlusive disease (1.4 to 1.8% vs. 0%).

Dasatinib

Dasatinib, a dual SRC-ABL1 inhibitor, is 300 times more potent than imatinib. In chronic phase CML after imatinib failure, dasatinib therapy was associated with complete cytogenetic response rates of 45 to 60%, major molecular response rate of 43%, and estimated 6-year survival rate of 71%. A first-line study comparing dasatinib to imatinib (DASISION trial)[6] randomized patients to receive either imatinib 400 mg daily or dasatinib 100 mg daily. With a minimum follow-up of 48 months, the incidence of complete cytogenetic response by 24 months was 86% with dasatinib and 82% with imatinib. The incidence of major molecular response was 74% vs. 60%. The rate of transformation to accelerated or blastic phase was 4.6% vs. 7%. The estimated 4-year progression-free survival rates were similar, 90%. The estimated 4-year survival rates were 93% and 92%, respectively. Dasatinib therapy was associated with lower rates of fluid retention, edema, myalgia, nausea, vomiting, and rashes. However, it was associated with higher rates of pleural effusions (about 10 to 15%) and cytopenias, particularly thrombocytopenia, as well as a low but notable incidence of pulmonary hypertension (<2 to 3%).

Bosutinib

Bosutinib, a dual SRC-ABLI inhibitor (similar to dasatinib), is 30 to 200 times more potent than imatinib. It has minimal inhibitory activity against c-Kit and PDGF receptor and therefore is expected to produce less myelosuppression and fewer pleural effusions. In studies of patients with chronic phase CML after imatinib failure treated with bosutinib 500 mg orally daily, the major cytogenetic response rate was 53%, the complete cytogenetic response rate was 41%, and the estimated 2-year survival rate was 92%.[6] Grade 3 to 4 toxicities included diarrhea (8%), rashes (9%), and thrombocytopenia (5 to 10%).

Ponatinib

Ponatinib is a pan-BCR-ABLI TKI with potent activity against native and mutated BCR ABLI kinases, including T3151. In a phase II study of patients with chronic, accelerated, or blastic (CML) or Ph-positive acute lymphocytic leukemia, all of whom were resistant or intolerant to several TKIs, ponatinib showed high activity. In these chronic phase CML patients, the complete cytogenetic response rate was 44%, and the major molecular response rate was 30%. In the subset of patients with T3151 mutation, the major cytogenetic response rate was 70%, the complete cytogenetic response rate 66%, and the major molecular response rate 50%.[7] Significant side effects included pancreatitis (5 to 10%), thrombocytopenia (30%), and skin rash (30%). Because of the cumulative incidence of serious thrombotic events since the approval of ponatinib, the FDA has restricted its use to patients resistant to other TKIs. Additional clinical trials assessing the safety profile and different dose schedules of ponatinib are being considered.

Omacetaxine

Omacetaxine mepesuccinate, a semisynthetic analogue of homoharringtonine, is a first-in-class cephalotaxine that acts as a protein synthesis inhibitor that induces apoptosis in leukemic cells by reducing levels of multiple oncoproteins, including BCR-ABL1. Data pooled from two phase II trials of subcutaneous omacetaxine, 1.25 mg/m^2 twice daily for 2 weeks every 4 weeks until response, then for 1 week every 4 weeks, in patients with chronic phase CML after failure of two TKIs, a major cytogenetic response was reported in 20% and complete cytogenetic response in 10%.[8] Grade 3 to 4 side effects included cytopenias in 37 to 67% of patients, which were reversible. This led to FDA approval of omacetaxine for patients whose disease has progressed despite treatment with two TKIs.

Allogeneic Hematopoietic Stem Cell Transplantation

AHSCT (Chapter 178), a potentially curative therapy in selected patients with CML, is most effective during the chronic phase, when it is associated with a 20-year survival rate of 40 to 50%. Transplant-related mortality rates range from 5 to 50%, depending on the patient's age, whether the donor is related or unrelated, the degree of matching, and other, less important factors such as positivity for cytomegalovirus, preparative and post-transplantation regimen, and institutional expertise. Disease-free survival rates with related allogeneic stem cell transplantation are 40 to 80% in chronic phase, 15 to 40% in accelerated phase, and 5 to 20% in blastic phase. In chronic phase CML, patients younger than 30 to 40 years have disease-free survival rates of 60 to 80%, compared with only 30 to 40% for patients older than 50 years. A major limitation of allogeneic stem cell transplantation is the availability of related donors. Human leukocyte antigen (HLA)-compatible unrelated donors can be found for 50% of patients; the median time from initiation of the donor search to transplantation is 3 to 6 months.

Nonmyeloablative preparative regimens have expanded the indications for AHSCT to older patients and have reduced transplant-related mortality and complications (Chapter 178). Early results show acceptable degrees of engraftment, less mortality and organ damage, more persistent residual disease, and similar degrees of graft-versus-host disease. Patients whose CML recurs after AHSCT may respond to imatinib or new-generation TKIs, donor lymphocyte infusions, IFN-α, or a second AHSCT.

AHSCT can produce an estimated cure rate of 40% at 20 years. However, it is associated with a 1-year mortality rate of 5 to 40% and with morbidities such as cataracts, infertility, second cancers (5 to 10%), immune-mediated complications, and chronic graft-versus-host disease. Delaying AHSCT beyond 1 to 3 years after diagnosis may be associated with worse results and with occasional sudden blastic transformation, which may not be salvageable. The outcome of AHSCT may be even better after exposure to TKIs.

Choice of First-Line Chronic Myelogenous Leukemia Therapy

Currently, all three TKIs (imatinib, nilotinib, and dasatinib) are acceptable first-line therapies for CML. The choice of TKI may depend on patient and physician preferences and patient prior history and comorbidities (e.g., diabetes, pancreatitis, cardiopulmonary conditions, and pulmonary hypertension). Current oncology practice trends appear to increasingly favor nilotinib and dasatinib over imatinib as initial therapy because of their better early results, particularly the lower early incidence of CML transformation. However, the costs of TKIs may shift treatment paradigms in some emerging nations to using a particular TKI over others, or even to consider first-line AHSCT (total one-time procedure cost of $30 to $100,000) in situations in which patients or the national health care system cannot afford the burden of the TKI. Imatinib may become available in generic formulations in 2015. The price of generic imatinib is unknown but may be lower than that of other agents ($2000 to $10,000 per year). The choice of first-line TKI therapy may then depend on the differential pricing of generic imatinib versus dasatinib and imatinib and on the maturing long-term data (5 to 8 years) for survival, transformation-free survival, and event-free survival with the three TKIs. With an estimated 8-year survival rate of 93% with imatinib (considering only CML-related deaths) and the high efficacy of new-generation TKIs as salvage therapies, the survival benefit with dasatinib or nilotinib may or may not be apparent compared with imatinib first-line therapy, careful monitoring for cytogenetic relapse, and rapid institution of second-line TKI therapies at that time.

Treatment of Chronic Phase Chronic Myelogenous Leukemia after Failure of Therapy with Imatinib or Other TKIs

In several studies, the achievement of a complete cytogenetic response (Ph-positive metaphases 0%; BCR-ABLI transcripts 1%) at 12 months or later on TKI therapy was associated with significant survival benefit compared with achievement of lesser degrees of response. Therefore, achievement of complete cytogenetic response is now the primary end point of TKI therapy. The achievement of complete molecular response (nonmeasurable BCR-ABLI transcripts) offers the possibility of treatment discontinuation in the clinical trial setting. Lack of achievement of major molecular response or of complete molecular response should not be interpreted as a need to change TKI therapy or to consider AHSCT. Response assessments at earlier times on first-line TKI therapy (3 to 6 months) have shown better outcomes, with achievement of a major cytogenetic response by 3 to 6 months on imatinib therapy (Ph-positive metaphases 35%, or BCR-ABLI transcripts 10%). Although this is interpreted to mean that a change to a second TKI therapy may be considered if such an outcome is not obtained, no studies have shown that changing therapy from imatinib to second TKI for this indication has improved patient outcome. When nilotinib or dasatinib is used for first-line therapy, achievement of complete cytogenetic response by 3 to 6 months of TKI therapy has been associated with improved outcomes.

Currently, imatinib failure (requiring a change of therapy) should be strictly defined as failure to achieve a major cytogenetic response after 6 months of

imatinib therapy and a complete cytogenetic response after 12 months or cytogenetic or hematologic relapse at any later time, on an optimal imatinib dose schedule (adjusting dose for significant side effects or for intolerance and checking for treatment compliance). With the use of second-generation TKIs in the first-line setting, failure of TKI therapy has been suggested to be lack of achievement of complete cytogenetic response or BCR-ABLI transcript levels of 1% by 3 to 6 months of therapy. Such patients (<10% of the denominator) have a worse event-free survival, although their survival at 3 to 5 years remains in the range of 90%, better or equivalent to what would be achievable with AHSCT. Thus, although the early surrogate response parameters at 3 to 6 months on first-line TKI therapy predict for differences in outcome, a change of therapy at that point in time has not been proved to improve longer term prognosis.

Patients with CML whose disease progresses on imatinib therapy may be treated with a newer generation TKI or with AHSCT, as discussed earlier. Patients with CML and failure on first-line dasatinib or nilotinib therapy may possibly be salvaged with ponatinib if the CML clones exhibit a T3151 mutation. If no such mutation is detected, they could be considered for other TKI therapies, AHSCT, treatment with omacetaxine, or combined-modality therapies including TKIs and older agents (hydroxyurea, cytarabine, decitabine). The choice of AHSCT as second-line versus later salvage therapy was discussed earlier.

Treatment of Accelerated and Blastic Phase Chronic Myelogenous Leukemia

Patients with accelerated or blastic phase CML may receive initial therapy with TKIs (newer generation TKIs like dasatinib or ponatinib are preferred over imatinib) to reduce the CML burden and may be considered for early AHSCT. Response rates with combinations of TKIs and chemotherapy are 40% in nonlymphoid blastic phase CML and 70 to 80% in lymphoid blastic phase CML. Median survival times are 6 to12 months and 12 to 24 months, respectively. The addition of TKIs to chemotherapy has improved the response rates and prolonged the median survival time in blastic phase CML.

At present, AHSCT is the only curative therapy for accelerated and blastic phase CML; overall cure rates are in the range of 15 to 40% and 5 to 20%, respectively. Patients with cytogenetic clonal evolution as the only accelerated phase criterion have a long-term event-free survival rate of about 60%. Otherwise, TKIs provide hematologic responses in 80% of patients and an estimated 4-year survival rate of 40 to 55% in accelerated phase CML, but only a 40% response rate and a median survival of 9 to 12 months in blastic phase CML. Patients in the accelerated or blastic phase should be encouraged to participate in investigational strategies to improve their prognosis. Patients with de novo CML accelerated phase have a better outcome with first-line TKI therapy than patients who evolve from chronic to accelerated phase. The estimated 6- to 8-year survival rates with TKI therapy in de novo accelerated phase CML are 60 to 80%. Such patients may continue on TKI therapy as their long-term treatment if they achieve a complete cytogenetic response on TKI therapy.

Special Therapeutic Considerations

Patients with severe leukocytosis and manifestations of leukostasis may benefit from initial leukapheresis. Severe thrombocytosis uncontrolled with anti-CML measures may respond to anagrelide, thiotepa, IFN-α, 6-mercaptopurine, 6-thioguanine, hydroxyurea, and platelet pheresis. CML during pregnancy may be controlled with pheresis in the first trimester and then with hydroxyurea until delivery. Use of IFN-α during pregnancy has been reported anecdotally to be safe. An analysis of 125 babies delivered to women with CML on imatinib therapy (who discontinued imatinib once the pregnancy was known) showed most babies to be healthy. However, the study demonstrated imatinib therapy to be associated with a syndrome of ocular, skeletal, and renal abnormalities in three babies delivered. Therefore, imatinib (and presumably other TKIs, although there is little experience with them) should be discontinued immediately once pregnancy is documented, but abortion is not recommended because fetal malformations are rare. Partners of men with CML on TKI therapy who become pregnant have delivered normal babies. Splenectomy can be useful as a palliative measure in patients with massive, painful splenomegaly, hypersplenism, or thrombocytopenia.

Monitoring Response to Therapy in Chronic Myelogenous Leukemia

With TKI therapy, complete cytogenetic response, major molecular response, and even complete molecular response have been achieved. These response rates improve with continued therapy and are higher with new-generation TKIs (dasatinib, nilotinib) compared with imatinib as first-line therapy (see Table 184-3). Techniques have been developed to measure these responses more accurately (rather than relying on only 20 metaphases by cytogenetic analysis), with less tedious and less painful procedures (peripheral blood rather than marrow studies), and at levels below the level of detection by routine cytogenetic evaluations. FISH studies with improved probes can measure 200 interphase cells using peripheral blood, and they have false-positive rates of less than 2 to 3%. Quantitative PCR tests usually measure the BCR-ABLI transcript levels (ratio of the abnormal message, BCR-ABLI, to a normal message, such as ABL1). BCR-ABLI transcript levels of 0.1% [IS], about a 3-log reduction of disease, have been associated with a very low risk for CML relapse on TKI therapy. This is referred to as a *major molecular response*. Undetectable BCR-ABL1 transcript levels, (usually <0.0032% [IS], or 4.5 log of reduction) are sometimes referred to as *complete molecular response*. The percentage of patients achieving complete molecular response is significantly higher with new-generation TKIs compared with imatinib.

In monitoring the response to TKI-based therapies, patients require a bone marrow analysis before treatment (to determine the percentages of blasts and basophils and clonal evolution) and FISH and quantitative PCR analyses. Quantitative PCR can be falsely negative at diagnosis in 5 to 8% of patients with unusual breakpoints and messages (e.g., b2a3 or b3a3) if proper procedures are not used. Thus, knowledge of the pretreatment BCR-ABL1 transcript levels avoids the false assumption of complete molecular response because of the false negativity. Bone marrow analysis may be useful at 6 and 12 months (to assess cytogenetic response and confirm complete cytogenetic response) and once every 1 to 3 years in patients with stable, durable complete cytogenetic responses (to look for chromosomal abnormalities in both Ph-positive and Phnegative cells). Monitoring in patients with confirmed durable complete cytogenetic responses can be continued with either FISH or quantitative PCR studies every 6 months (or more often, such as every 3 months, if there are concerns about significant and consistent increases in BCR-ABLI transcripts levels). Some CML experts have shifted from monitoring by marrow studies to monitoring by peripheral blood studies using molecular analysis, with or without FISH studies. Among patients achieving major molecular response, molecular analysis without FISH studies is sufficient.

Resistance to imatinib therapy (discussed earlier) requires a change of therapy to other TKIs, TKI combinations with chemotherapy, or consideration of AHSCT. It is important to emphasize that many patients with apparent CML resistance to a TKI therapy may be noncompliant with the treatment. This should be discussed clearly with patients when they exhibit signs of CML progression by either molecular or FISH studies. If they are noncompliant, they may continue on the same TKI treatment with emphasis on compliance and evaluated 3 to 6 months later, before CML resistance is declared.

Among patients in complete cytogenetic response on a particular TKI, failure to achieve a major molecular response does not, at present, indicate resistance to the particular TKI or a need to change therapy. Mutational studies are recommended in patients who develop cytogenetic or hematologic resistance or relapse on a particular TKI therapy, when considering changing therapy to another TKI. The detection rate of mutations in this situation is 30 to 50%. Mutational studies are not recommended in patients in complete cytogenetic response on a particular TKI because the detection rate of mutations are then very low (<3 to 5%).

FUTURE DIRECTIONS

In 2014, patients with CML have multiple treatment options, including several TKIs, omacetaxine (protein synthesis inhibition), and several older agents (hydroxurea, IFN-α, busulfan, 6-mercaptopurine, cytarabine, decitabine). Most patients with CML would be expected to live their normal functional life and to be functionally, although not molecularly, cured, as long as they continue therapy with TKI-based regimens, are compliant with the treatment, and are monitored closely for signs of resistance in order to change therapy in a timely manner and/or consider AHSCT before CML progression. Future directions will focus on the potential molecular cure of CML (i.e., achievement of a durable complete molecular response and its persistence after discontinuation of TKI therapy). This is not a trivial issue because, with effective TKI therapy and full treatment penetration worldwide (to 100% of all diagnosed patients and continuation of TKI therapy without interruptions), the prevalence of CML would increase annually and plateau in about 2030 to 2040 at a rate 35 times the incidence. This figure is estimated to be close to 160,000 patients with CML in the United States and about 3 million patients worldwide. This may represent a considerable burden on patients and the health care systems in relation to drug availability, compliance, potential development of long-term side effects, and costs. Therefore, it is critical to continue research into therapies that increase the rates of durable complete molecular responses. This may be achievable with the current more potent new-generation TKIs alone or in combination with other available (pegIFN-α, omacetaxine, decitabine) or investigational therapies (JAK2 inhibitors, hedgehog inhibitors, stem cell poisons, vaccines). Such strategies may improve the eradication of minimal residual disease, potentially obviating the need for indefinite therapy with TKIs. Further understanding of the pathophysiologic events downstream of BCR-ABLI may help in the development of new strategies to target them.

PROGNOSIS

Treatment of CML with imatinib and other TKIs has revolutionized the outcome of the disease. In patients with newly diagnosed CML, imatinib therapy is associated with an estimated 8- to 10-year survival rate of 85% (93% if non-CML deaths are censored). If this favorable trend continues with longer follow-up, the median survival time in CML may exceed 25 years. The annual mortality rate of CML with TKIs in the first decade of experience has been reduced from the historical rate of 10 to 20%, down to 2% (1% if only CML deaths are counted). Many well established poor prognostic factors in CML (e.g., older age, splenomegaly, presence of marrow fibrosis, deletion of 9q) have lost much of their prognostic importance since the advent of TKI therapy. With AHSCT, cures can be expected in 40 to 80% of patients with chronic phase CML, 15 to 40% of those with accelerated phase CML, and 5 to 20% of those with blastic phase CML.

CHRONIC MYELOMONOCYTIC LEUKEMIA AND ATYPICAL CHRONIC MYELOID LEUKEMIAS

DEFINITION AND EPIDEMIOLOGY

Although superficially resembling CML in its clinical and morphologic presentation, CMML should be considered a separate entity because of its particular clinical, therapeutic, and prognostic aspects. CMML is a hybrid entity manifesting as a proliferation of the myeloid monocytic series and dysplasia of the erythroid-megakaryocytic series. Patients with CMML are older (median age, 65 to 70 years) than most patients with CML.

PATHOBIOLOGY

The cytogenetic findings in patients with CMML are either normal or involve an additional chromosome 8 or findings other than the Ph chromosome. Patients with CMML have *RAS* mutations in 40 to 60% of cases.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients often present with symptoms related to anemia and thrombocytopenia (fatigue, bleeding). Other typical features include splenomegaly, leukocytosis, and monocytosis. Organ infiltration (lymph nodes, skin, liver) is less common. Basophilia and thrombocytosis are not presenting features. High-frequency mutations in the granulocyte colony-stimulating factor 3 receptor gene (*CSF3R*) in *chronic neutrophilic leukemia* (CNL) and in some patients with *atypical chronic myeloid leukemia* (aCML) have been identified. In addition, recurrent mutations in *SETBP1* (Set binding protein) have been identified in 25% of aCML patients.

TREATMENT AND PROGNOSIS

Rx

AHSCT (Chapter 178), which is the only curative modality, should be considered first-line therapy in candidate patients. Other therapies include hydroxyurea to control leukocytosis, erythropoietin to improve anemia, azacitidine or decitabine (both approved by the FDA for the treatment of CMML), topotecan and cytarabine or other intensive anti-AML (Chapter 183) regimens for CMML transformation, splenectomy for symptomatic splenomegaly and/or hypersplenism, and investigational agents. Inhibition of Janus kinase 2 or SRC kinase signaling downstream of mutated CSF3R is being explored therapeutically.

Poor prognostic factors include the presence of anemia (hemoglobin <10 g/dL), thrombocytopenia, and more than 5% blasts. Median survival is 12 to 18 months.

HAIRY CELL LEUKEMIA

DEFINITION AND EPIDEMIOLOGY

Hairy cell leukemia (HCL) is an uncommon and indolent B-cell leukemia (1 to 2% of all leukemias). The median age at diagnosis is 50 years, and there is a 4 : 1 male preponderance.

PATHOBIOLOGY

The cell of origin of HCL is the B lymphocyte, as documented by the demonstration of heavy and light chain immunoglobulin gene rearrangements. In a series of 47 patients, all had a *BRAF V600E* activating mutation. Hairy cells express CD19, CD20, CD11C, CD103, FMC7, and CD22, but not CD21, CDS, CD10, or CD23. The cells demonstrate a κ or λ light chain phenotype. The cells also express CD25 (TAC), the low-affinity interleukin-2 (IL-2) receptor, and CD103, a unique hairy cell antigen. High levels of soluble IL-2 receptor (more than five times normal) are present in the sera of almost all patients with HCL, with extremely high levels noted in many cases. Immune dysfunction is wide ranging in HCL. Monocytopenia is universal; B and T lymphocytes are decreased in number; the CD4/CD8 (helper T/suppressor T) ratio is often inverted; and skin test reactivity to recall antigens is impaired, as is antibody-dependent cellular cytotoxicity. Humoral immunity is relatively preserved, with normal immunoglobulin levels. Marrow failure in HCL may be due in part to inhibitory factors (e.g., tumor necrosis factor) produced by the leukemic infiltrate; the pancytopenia is often more marked than would be anticipated from the degree of leukemic infiltration.

CLINICAL MANIFESTATIONS

Most patients present with pancytopenia and splenomegaly. Patients may also have fatigue, fever, weight loss, and infection secondary to granulocytopenia or monocytopenia. Leukocytosis is uncommon, and lymphadenopathy is rare. Anemia is present in up to 85% of patients, whereas leukopenia and thrombocytopenia are present in 60 to 75%. The cytopenias are caused by a combination of bone marrow failure due to leukemic infiltration and hypersplenism. Patients may experience repeated infections and, rarely, a systemic vasculitis resembling polyarteritis nodosa. Although bacterial infections occur, as would be expected with neutropenia, patients with HCL have a predilection to develop tuberculosis, atypical mycobacterial infections, and fungal infections, perhaps related to the severe monocytopenia that is characteristic of this disorder.

DIAGNOSIS

In conjunction with the clinical features, careful examination of the peripheral blood smear may demonstrate the occasional typical cells with cytoplasmic projections, giving rise to the name *hai1y cell leukemia* (Fig. 184-5). The hairy cells are 10 to 15 mm in diameter, with pale blue cytoplasm, a nucleus with a loose chromatin structure, and one or two indistinct nucleoli. Bone marrow aspiration is often inadequate owing to increased deposition of reticulum, collagen, and fibrin; bone marrow biopsy is usually necessary. Bone marrow involvement is interstitial or patchy, and the infiltrate is characterized by widely spaced nuclei due to the abundant cytoplasm, giving rise to the commonly described fried-egg appearance.

Hairy cells exhibit a strong acid phosphatase (isoenzyme 5) cytochemical reaction in 95% of cases, a reaction that is resistant to the inhibitory effect of tartaric acid (TRAP). Other lymphoproliferative diseases are rarely TRAP positive. Electron microscopy clearly demonstrates the microvillar projections. Often, ribosomal-lamellar complexes, which are characteristic but not diagnostic of HCL, can be identified. The peroxidase stain is negative, and lysozyme activity is absent in hairy cells, thereby differentiating these cells from monocytes.

Differential Diagnosis

The differential diagnosis must distinguish HCL from non-Hodgkin lymphoma (Chapter 185) or chronic lymphocytic leukemia (CLL) (see later), which can manifest with predominant splenomegaly and minimal lymphadenopathy. Some patients with a myelodysplastic syndrome (Chapter 182) or

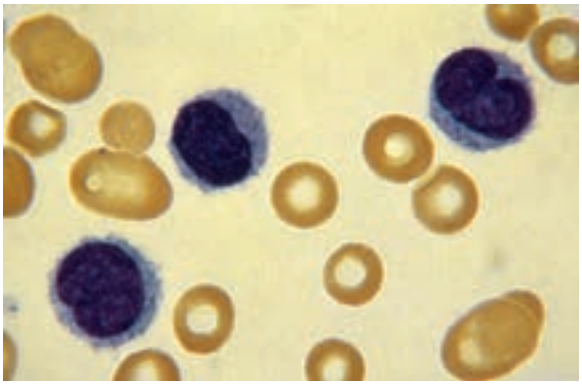

FIGURE 184-5. Hairy cell leukemia. Peripheral smear shows hairy cells with blue-gray cytoplasm; fine, hair-like projections (resembling ruffles); and oval or slightly indented nuclei with loose chromatin and indistinct nucleoli. (Courtesy Andrew Schafer, MD.)

a chronic myeloproliferative neoplasm (Chapter 166) have splenomegaly and pancytopenia with only a few atypical cells. Patients with other diseases, such as systemic lupus erythematosus (Chapter 266) and other autoimmune diseases, B-cell and T-cell prolymphocytic leukemias (see later), infiltrative splenomegaly (Chapter 168), or tuberculosis (Chapter 324), may have splenomegaly and cytopenia, but these diagnoses can usually be made by history, physical examination, and appropriate blood and bone marrow tests. Splenomegaly, cytopenia, and nonaspirable marrow in a middle-aged man should create a high index of suspicion for HCL. Splenectomy or lymph node biopsy is sometimes necessary to establish the diagnosis in difficult cases. Cases of HCL variant manifest with higher WBC counts, are TRAP negative, have prominent nucleoli, and are only occasionally positive for antibodies against CD25. HCL variant does not respond as well to the agents that are usually effective in the management of typical HCL.

TREATMENT

Rx

A small proportion (<5%) of patients with HCL do not require therapy. These patients have mild cytopenias, are not transfusion dependent, have no history of infections, and have a low level of marrow infiltration by hairy cells. 2-Chlorodeoxyadenosine (cladribine), an adenosine analogue that *is* resistant to deamination by adenosine deaminase, produces complete remission in more than 80% of HCL patients after a single course of 0.1 mg/kg/day for 7 days given by continuous intravenous infusion, and it is now the recommended first-line therapy.[9] It can also be given at 0.14 mg/kg/day for 5 days as a short daily intravenous infusion. Remissions are durable, and patients who relapse can often attain a second remission after retreatment with cladribine. The drug is well tolerated, with a low infection rate. Despite long-lasting suppression of $CD4^+$ lymphocyte counts, there does not appear to be an increase in late opportunistic infections or second malignancies. Partial response to purine analogs is regarded as a poor prognostic factor, and a second course of purine analog therapy is recommended if patients do not enter complete remission, with the addition of rituximab to be considered. Rituximab in combination with a purine analogue is often used in the treatment of relapsed disease.

Deoxycoformycin (pentostatin; 4 mg/m^2 weekly or every 2 weeks for up to 6 months), an adenosine deaminase inhibitor, produces complete remission in 70 to 80% of patients. The response to treatment is rapid. Toxicity includes nausea and vomiting, infection, renal and hepatic dysfunction, conjunctivitis, and photosensitivity, albeit mild in most cases.

Human leukocyte interferon (HuIFN), or recombinant interferon-α (r-IFN-α), rapidly improves granulocyte, platelet, and hemoglobin levels (within 1 to 3 months); reduces spleen size; and decreases marrow infiltration. Peripheral blood cell counts return to normal in 80% of cases, but complete remission is uncommon. In addition, when treatment is discontinued, relapse occurs within 1 to 2 years. Rituximab, the monoclonal antibody targeting CD20, also produces responses; eight weekly infusions appear to be more effective than four. Two immunotoxins can produce responses in refractory patients. LMB2 is composed of the Fc portion of the anti-TAC antibody linked to a *Pseudomonas* exotoxin. Moxetumomab also contains a *Pseudomonas* exotoxin linked to an antibody targeting CD22. The B-RAF inhibitor, vemurafenib, has successfully been used to treat a patient with HCL, and a clinical trial in relapsed HCL is underway.[10] Splenectomy is recommended mainly for patients with splenic infarcts or massive splenomegaly.

PROGNOSIS

More than 85 to 90% of patients treated with cladribine or pentostatin are expected to be alive at 10 years.

CHRONIC LYMPHOCYTIC LEUKEMIA

DEFINITION

CLL is a neoplasm characterized by the accumulation of monoclonal lymphocytes of B-cell origin. The cells accumulate in the bone marrow, lymph nodes, liver, spleen, and occasionally other organs. After decades of chemotherapy-based treatment of CLL, recent progress has turned attention to mechanism-driven therapy with targeting of the B-cell receptor signaling pathway.[11]

EPIDEMIOLOGY

CLL is the most common leukemia (one third of all cases) in the Western world and is twice as common as CML. The disease occurs rarely in those younger than 30 years; most patients with CLL are older than 60 years. CLL increases in incidence exponentially with time; by age 80 years, the incidence rate is 20 cases per 100,000 persons per year. The male-to-female ratio is approximately 2 : 1. The incidence of CLL among Asians in Japan and China is only 10% of that in the United States and other Western countries. Intermediate incidence rates are seen in persons of Hispanic origin.

The cause of CLL is unknown. Ionizing radiation and viruses have not been associated with CLL, although hepatitis C infection has recently been associated with splenic lymphoma with villous lymphocytes (another indolent B-cell disorder). Familial clustering in CLL is more common than in other leukemias; first-degree relatives of patients have a two- to four-fold higher risk and develop CLL at a younger age compared with the general population. Farmers have a higher incidence of CLL than do those in other occupations, raising the possibility of an etiologic role for herbicides or pesticides. Agent Orange, the defoliating agent used in Vietnam, has been associated with the development of CLL.

PATHOBIOLOGY

Leukemia cells in CLL are homogeneous and have the appearance of normal mature lymphocytes. However, clonality can be documented by the presence of immunoglobulin gene rearrangements and the restriction to either κ or λ light chains on the cell surface. The cells express low-intensity monoclonal surface immunoglobulin (Smig; usually immunoglobulin [Ig] M ± IgD) and the pan-B-cell antigens CD19, CD20, CD23, and CD24 in almost all cases, as well as CD21 (which includes the receptor for the Epstein-Barr virus and the C3d component of complement) in more than 75% of cases. Almost all cells exhibit Ia antigen and receptors for the Fc fragment of IgG and spontaneously form rosettes with mouse erythrocytes. In addition to B cell antigens, CLL cells express CDS (Leu 1, Tl, and TlOl), a pan-T-cell antigen. Other T-cell antigens are absent. CD25 (TAC, IL-2 receptor) antigen is positive in about 25% of cases. T cells are increased in number at diagnosis, and the CD4/CD8 ratio is often inverted, owing to a relatively greater increase in $CD8^+$ cells. The CD4/CD8 ratio declines as the disease progresses and after therapy. The T cells have a blunted response to T-cell mitogens and decreased delayed hypersensitivity reactions to recall antigens. However, these T-cell functions are impaired by factors produced by the CLL cells because purified T cells have a normal response to T-cell mitogens.

GENETICS

Genes mutated in CLL include *TP53* (15% of patients), *SF3B1* (15%), *ATM* (9%), *MYD88* (10%), and *NOTCH1* (4%). Standard cytogenetic analysis identifies abnormalities in 40 to 50% of cases of CLL, but CLL cells have low mitotic activity. By FISH, the likelihood of detecting abnormalities increases to 80%. A 13q deletion is the most common abnormality; other abnormalities include llq deletion (15 to 20%), trisomy 12 (15 to 20%), and 17p deletion (5 to 10%). The 17p deletion increases in frequency as the disease progresses, recurs after therapy, and is associated with a very poor prognosis. The llq deletion also is associated with a poorer prognosis, whereas the 13q deletion, if present as the sole abnormality, is associated with a favorable prognosis.

CLINICAL MANIFESTATIONS

Most patients with CLL do not have symptoms, and the disease is diagnosed when absolute lymphocytosis is noted in the peripheral blood (Fig. 184-6)

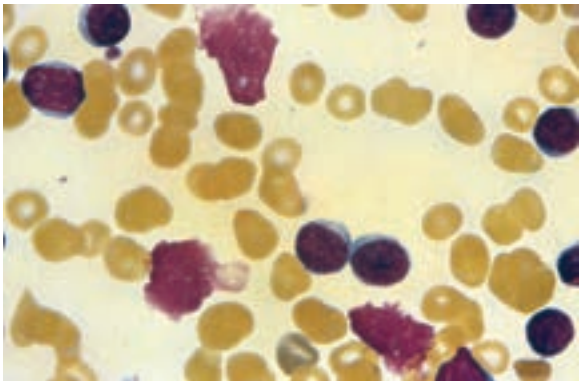

FIGURE 184-6. Chronic lymphocytic leukemia. Peripheral smear shows that the predominant leukocytes are "normal," mature-appearing lymphocytes, with occasional "smudge" cells. (Courtesy Andrew Schafer, MD.)

during evaluation for other illnesses or when the patient undergoes a routine physical examination. Symptoms such as fatigue, lethargy, loss of appetite, weight loss, and reduced exercise tolerance are nonspecific. Many patients have enlarged lymph nodes. B symptoms (fever, night sweats, weight loss) are rarely present initially, and their presence in later stages of the disease suggests transformation to large cell lymphoma (Richter transformation). The most common infections are sinopulmonary. As the disease progresses, the frequency of neutropenia, T-cell deficiency, and hypogammaglobulinemia increases, resulting in infections with gram-negative bacteria, fungi, and viruses such as herpes zoster and herpes simplex.

The major physical findings relate to infiltration of the reticuloendothelial system. Lymphadenopathy with discrete, soft, mobile lymph nodes is present in two thirds of patients at diagnosis. Later, as the lymph nodes enlarge, they can become matted. Enlargement of the liver or spleen is less common at diagnosis (approximately 10% and 40% of cases, respectively) but occurs more frequently with progression. Organ failure resulting from infiltration with CLL is uncommon. Infiltration of the central nervous system in CLL is rare, and central nervous system symptoms are more likely to be caused by opportunistic infections such as cryptococcosis or listeriosis.

DIAGNOSIS

CLL is characterized by absolute lymphocyte counts that typically range from 5000 to 600,000 $\times 10^9/\mu L$ in the peripheral blood. Even with markedly elevated WBC counts, hyperviscosity symptoms rarely occur, probably because of the small size and pliability of the cells. Anemia (hemoglobin <11 g/dL) is present in 15 to 20% of patients at diagnosis and thrombocytopenia (platelet count $<100 \times 10^9/\mu L$) in 10%. However, bone marrow replacement and hypersplenism, which are seen with progressive disease, increase the frequency of anemia and thrombocytopenia. The anemia is usually normochromic and normocytic, and the reticulocyte count is normal unless the patient has autoimmune hemolytic anemia (Chapter 160), which usually results from the development of a warm-reacting IgG antibody. The diagnosis of autoimmune hemolytic anemia, which occurs in 10% of cases, is confirmed by a positive direct Coombs (DAT) test (80 to 90% of cases), reticulocytosis, a low serum haptoglobulin concentration, and an elevated unconjugated serum bilirubin level. In such patients, reactive erythroid hyperplasia as a response to the hemolysis may be masked in the bone marrow by the marked lymphocytic infiltration. Cold agglutinin hemolysis occurs rarely in CLL. Autoimmune thrombocytopenia (immune thrombocytopenic purpura; Chapter 172) can be diagnosed in 10 to 15% of cases. The antibodies causing red cell and platelet destruction are not produced by the CLL cells, and the mechanisms for the associated autoimmune diseases are not known. Pure red cell aplasia (Chapter 165) is an additional, underappreciated cause of anemia in CLL.

The lymphocytes in CLL are indistinguishable on light or electron microscopy from normal small B lymphocytes (see Fig. 184-6). On bone marrow aspiration, the proportion of lymphocytes is greater than 30% and may be up to 100%. Four patterns of lymphocyte infiltration on bone marrow biopsy occur: nodular (15%), interstitial (30%), mixed nodular and interstitial (30%), and diffuse (35%). Most early-stage cases have one of the first three patterns; diffuse histology is common in advanced-stage disease and becomes more prominent as the disease evolves. A diffuse histologic pattern confers a poor prognosis regardless of the stage of disease.

Differential Diagnosis

There are many diseases that can cause lymphocytosis, including pertussis (Chapter 313), cytomegalovirus (Chapter 370), Epstein-Barr virus mononucleosis (Chapter 377), tuberculosis (Chapter 324), toxoplasmosis (Chapter 349), chronic inflammatory disorders, and autoimmune syndromes. These diseases are seldom confused with B-cell CLL, largely because the lymphocytosis in these conditions is usually less than $15 \times 10^9/\mu L$ and is not sustained. If doubt persists, immunophenotypic or molecular studies can distinguish the monoclonal lymphocytosis in CLL from the T-cell or polyclonal B-cell proliferation in the other disorders.

In individuals 62 to 80 years old, monoclonal CLL-phenotype B cells are found in about 5% of individuals with normal blood counts. In patients with greater than 4000 lymphocytes/mL, about 45% have CLL, about 40% have reactive lymphocytosis, and about 15% have monoclonal CLL-phenotype B cells that confer a 1.1% per year risk for developing CLL. The latter is a more recently recognized entity that has been termed *monoclonal B-cell lymphocytosis*. In patients ultimately diagnosed with CLL, B-cell clones were previously present in peripheral blood in 98% of patients, sometimes many years before diagnosis.

Other Chronic Lymphocytic Leukemias

The more difficult differential diagnosis is distinguishing CLL from other lymphoproliferative disorders such as prolymphocytic leukemia (PLL), splenic lymphoma with villous lymphocytes, HCL (see earlier section), the leukemic phase of mantle cell lymphoma, and Waldenström macroglobulinemia (Chapter 187). Although certain clinical features are more common in some of these disorders (e.g., marked splenomegaly with minimal or no lymphadenopathy in PLL, splenic lymphoma, and HCL vs. extensive lymphadenopathy with or without splenomegaly in CLL), none of these clinical features is specific. The differential diagnosis therefore depends largely on histopathologic and, more specifically, immunophenotypic features (Table 184-4).

Prolymphocytic Leukemia

PLL is an uncommon disease (incidence <5% that of CLL), and its characteristics of massive splenomegaly, minimal lymphadenopathy, and markedly elevated WBC count (often $>100 \times 10^9/\mu L$), with 10 to 90% of the cells being prolymphocytes, distinguish this disease from typical B-cell CLL. Prolymphocytes are larger cells that have a distinct nucleolus and express FMC-7. The male-to-female ratio is 4 : 1, and the median age at diagnosis is 70 years. Survival is shorter than in CLL (median, 3 years), and response to therapies usually applied in CLL is poor. A serum paraprotein, typically IgG or IgA, is present in one third of cases. The immunoglobulin on the surface of the cells is occasionally IgG or IgA, not IgM ± IgD, as in CLL. Several karyotypic abnormalities have been reported in PLL, including t (11; 14) (q13; q32). Deletions of llq3, 23, and 17p are more common in B-cell PLL than in CLL. Abnormalities in the *TP53* oncogene are found in 75% of cases of B-cell PLL. One fifth of PLL cases express a T-cell phenotype.

Small Lymphocytic Lymphoma

Small lymphocytic lymphoma (SLL) shares histopathologic and immunophenotypic features with CLL, differing only in the lack of lymphocytosis in

TABLE 184-4 DIFFERENTIAL DIAGNOSIS OF INDOLENT LYMPHOPROLIFERATIVE DISORDERS

DISEASE	LYMPHADENOPATHY (%)	SPLENOMEGALY(%)	CELL OF ORIGIN (BIT)	POSITIVE MARKERS			
				Smlg	*CD5*	*CD19, CD20 (%)*	*Other*
Chronic lymphocytic leukemia (CLL)	75	50	B (20 : 1)	Weak	>90%	90	Mouse red blood cell receptors
Prolymphocytic leukemia (PLL)	33	95	B (4 : 1)	Bright	T-cell PLL	75	FMC-7
Hairy cell leukemia	<10	80	B (T rare)	Bright	—	>90	CD25, CD11C, CD103
Lymphoma (leukemic phase)	90	90	B (T rare)	Bright	Some	>90	CD10
Splenic lymphoma with villous lymphocytes	10	80	B	Bright	20%	>90	FMC-7, CD22
Waldenström macroglobulinemia	33	33	All B	Weak	Some	Many	CD38, PCA-1
Large granular lymphocytosis	10	10	All T	Absent	-	-	CD2, CD3, CD8

CD2 = pan-T cell; CD3 = pan-mature T cell; CDS = pan-T cell, B-cell CLL; CD8 = T cell (suppressor cytotoxic); CD10 = early B cell; CD11C = hairy cell, activated T cell, NK cell; CD19 =early pan-B cell; CD20 = pan-B cell; CD25 = low-affinity interleukin-2 receptor; CD38 = activated B cell, thymocyte, plasma cell; FMC-7 = PLL, hairy cell leukemia; PCA-1 = plasma cell; Smlg = monoclonal surface immunoglobulin.

the peripheral blood. The bone marrow in SLL may or may not have more than 30% lymphocytes. LFA-1 adhesion protein is much more commonly expressed on SLL cells than on CLL cells. Other lymphomas, such as follicular and mantle cell lymphomas (Chapter 185), occasionally manifest a leukemic phase on initial presentation. Follicular lymphoma cells are often cleaved on light microscopy, have bright staining for Smlg, and are positive for FMC-7 and CD10. Lymph node biopsy should be performed to identify these cases with greater precision. The presence of lymphoma cells in the blood in follicular lymphoma is more common with advanced disease. Follicular lymphoma can usually be identified by the presence of the translocation t(14;18) and consequent *BCL2* rearrangement, both of which are rare in CLL. The WBC count in Waldenström macroglobulinemia (Chapter 187) is usually much lower than in CLL ($<10 \times 10^9/\mu L$), and many patients are leukopenic. The cells have a plasmacytoid appearance, CD38 and PCA-1 positivity, and more Smlg and cytoplasmic Ig. A serum monoclonal IgM peak is present in almost all cases of Waldenström macroglobulinemia but is uncommon in CLL.

T-Cell Leukemias

The predominant clinical manifestation of *Sézary syndrome* (a $CD4^+$ T-cell malignant disorder related to *mycosis fungoides*) is chronic exfoliative erythroderma with a low number of circulating monoclonal T cells. The clinical and laboratory differentiation from CLL is not difficult.

Other T-cell malignant disorders with peripheral blood involvement are *adult T-cell leukemia-lymphoma* and *large granular lymphocytosis*, also referred to as *large granular lymphoproliferative disorder*, T-cell lymphocytosis with neutropenia, or T-γ lymphocytosis syndrome. Adult T-cell leukemia-lymphoma is associated with a retrovirus (human T-lymphotropic virus I) and is common in Japan and the Caribbean. It frequently manifests with lytic bone lesions and hypercalcemia. In T-cell large granular lymphoproliferative disorders, the absolute lymphocyte count is usually low ($<5 \times 10^9/\mu L$). The disease is defined by clonal amplification of either $CD3^+$ cytotoxic T-lymphocytes or $CD3^-$ natural killer (NK) cells.[12] These patients often have splenomegaly, neutropenia, and rheumatoid arthritis–like symptoms. A subset, called NK-cell large granular lymphocytosis, has an NK-cell phenotype ($CD16^-$) and no molecular evidence of T-cell receptor rearrangement. The lymphocytes in large granular lymphoproliferative disorder have abundant cytoplasm with azurophilic granules. Most patients have a benign course, although repeated infections can occur.

Staging and Prognostic Factors

The natural history of CLL is heterogeneous, with survival times ranging from 2 to 20 years after diagnosis. Either of two validated clinical staging systems can be used. The Rai staging system (1975) defines five stages and is most frequently used in the United States, whereas the Binet system (1981) defines three stages and is most frequently used in Europe (Table 184-5). Patients with anemia and thrombocytopenia (Rai III and IV, Binet C) have the worst prognosis; patients with lymphocytosis alone (Rai 0, some Binet A patients) have an excellent prognosis. A group of patients with a lymphocyte count of less than $30 \times 10^9/\mu L$, hemoglobin greater than 13 g/dL, platelet count greater than $100 \times 10^9/\mu L$, fewer than three involved node areas, and lymphocyte doubling time greater than 12 months has been described as having "smoldering" CLL, with survival equal to that of an age- and sex-matched control population. Patients tend to progress through stages, with many patients developing more sites of involvement over time and eventually experience marrow failure; however, anemia and thrombocytopenia can develop abruptly, even without antibody-mediated destruction or markedly increased tumor burden.

Other adverse factors include a diffuse pattern of lymphocytic infiltration observed on bone marrow biopsy; molecular abnormalities, including deletion of llq or 17p; advanced age; male sex; elevated serum levels of thymidine kinase, β_2-microglobulin, and soluble CD23; rapid lymphocyte doubling time; an increased proportion of large or atypical lymphocytes in the peripheral blood; and lack of somatic mutation of the *VH* gene within the B-CLL cell or the presence of the ZAP-70 protein or CD38 on the CLL cell surface.

TREATMENT

The major therapeutic questions are when to treat and which therapeutic regimen to use. Recent progress has transformed the treatment of CLL, initially from chemotherapy to chemoimmunotherapy and now to mechanism-driven drugs that target the B-cell receptor signaling pathway.[13] Patients with CLL are usually older, and the prognosis of the disease is variable (with some early-stage cases being stable for 10 to 20 years). Treatment of early-stage CLL (Rai 0, Binet A) is delayed until the disease progresses. In randomized trials, early treatment with alkylating agents did not prolong survival and was associated with a heightened risk for developing second malignant tumors. Treatment of Rai stages III and IV (Binet stage C) is recommended at the time of diagnosis because of the morbidities associated with cytopenias and the poor survival time of these patients. Treatment of intermediate-stage disease (Rai I and II, Binet B) is recommended if symptomatic disease (fever, sweats, weight loss, severe fatigue) or massive lymphadenopathy, with or without hepatosplenomegaly, is present.

Medical Therapy

Chemotherapy

Fludarabine monophosphate (25 mg/m^2/day for 5 days every 4 weeks), an adenosine analogue, is approved by the FDA for treatment of relapsed CLL; it produces an overall response rate of 50 to 60%. The dose-limiting toxicity is myelosuppression. In a large randomized trial of initial therapy for CLL, fludarabine was compared with chlorambucil, the historical standard therapy; it resulted in higher overall and complete remission rates, longer duration of remission, and improved response rates on crossover. Ten-year follow-up showed a survival benefit in the fludarabine arm. Cladribine is used primarily in Europe, where it appears to have efficacy similar to fludarabine. Pentostatin has not been as widely studied in CLL as in HCL.

The combination of fludarabine and cyclophosphamide (FC) was a logical attempt to improve on the efficacy of fludarabine alone by combining it with the other most active agent in this disease, an alkylating agent. The FC combination has been compared with fludarabine alone in three randomized trials. All these trials consistently showed a higher complete response rate, higher overall response rate, and longer progression-free survival with the FC combination.

TABLE 184-5 RAI AND BINET STAGING SYSTEMS IN CHRONIC LYMPHOCYTIC LEUKEMIA

STAGE	LYMPHOCYTOSIS	LYMPHADENOPATHY	HEPATOMEGALY OR SPLENOMEGALY	HEMOGLOBIN (g/dL)	PLATELETS $\times 10^3$/mL
RAI STAGING SYSTEM					
0	+	–	–	≥11	≥100
I	+	+	–	≥11	≥100
II	+	±	+	≥11	≥100
III	+	±	±	<11	≥100
IV	+	±	±	Any	<100
BINET STAGING SYSTEM					
A	+	±	± (<3 lymphatic groups* positive)	≥10	≥100
B	+	±	± (≥3 lymphatic groups* positive)	≥10	≥100
C	+	±	±	<10†	<100†

*The three lymphatic groups are (1) cervical, axillary, and inguinal nodes; (2) liver; and (3) spleen. Each group is considered one group whether unilateral or bilateral.
†The criterion is hemoglobin <10 g/dL and/or platelets $<100 \times 10^3$/mL.

TABLE 184-6 DEFINITION OF REMISSION IN CHRONIC LYMPHOCYTIC LEUKEMIA: THE INTERNATIONAL WORKSHOP IN CHRONIC LYMPHOCYTIC LEUKEMIA–NATIONAL CANCER INSTITUTE WORKING GROUP CRITERIA

CRITERION	COMPLETE REMISSION	PARTIAL REMISSION
Physical examination		
Nodes	None ≥1.5 cm	≥50% decrease
Liver/spleen	Not palpable	≥50% decrease
Symptoms	None	N/A
Peripheral blood		
Neutrophils	≥1500 m/L	≥1500/mL or ≥50% increase from baseline
Platelets	>100,000/μL	>100,000/μL or ≥50% increase from baseline
Hemoglobin	>11 g/dL	11 g/dL or >50% increase from baseline
Lymphocytes	≤4000 /mL	>50% decrease
Bone marrow	<30%, no B-lymphoid nodules	50% reduction in infiltrate or B-lymphoid nodules

After therapy, many patients remain stable for months to years before progressive disease indicates the need for further treatment. The goal of treatment is to achieve complete response (Table 184-6).

Bendamustine is a potent alkylating agent that has some structural similarity to nucleoside analogues, but preclinical data suggest that it does not function as a nucleoside analogue. Bendamustine was recently approved by the FDA for the treatment of CLL based on a randomized trial comparing this agent to chlorambucil as initial therapy for patients with CLL. Complete and overall response rates were higher with bendamustine, and progression-free survival was longer. The main side effect is myelosuppression.

Monoclonal Antibodies Alone and with Chemotherapy

Rituximab, a monoclonal antibody targeting the CD20 antigen, is associated with response rates of about 50% when given at the standard dose (375 mg/m^2/week for 4 weeks) as initial therapy for CLL and significantly lower rates when used in the salvage setting; complete responses are rare in either situation. The major benefit of this antibody appears to be its use in combination with chemotherapy. Fludarabine combined with rituximab produces better responses than those seen historically with fludarabine alone. In addition, progression-free and overall survival rates are better than those seen in the historical cohort. A three-drug regimen of fludarabine, cyclophosphamide, and rituximab (FCR) appears to produce the best and most durable complete remission rates when used as first-line therapy. A randomized trial compared the activity of FC to that of FCR. The complete response rate and the overall response rate were significantly higher with FCR, and progression-free and overall survival rates were significantly longer than with chemotherapy alone. A trial of previously untreated CLL patients with coexisting conditions tested the efficacy of chlorambucil alone versus chlorambucil plus rituximab versus chlorambucil plus obinutuzumab, a glycoengineered anti-CD20 monoclonal antibody. Combining an anti-CD20 antibody with chemotherapy improved outcomes, and obinutuzumab was found to be superior to rituximab when each was combined with chlorambucil.

Alemtuzumab (Campath-IH; 30 mg intravenously three times a week for 4 to 12 weeks), a monoclonal antibody that binds to the CD52 antigen, was originally approved for the treatment of fludarabine-refractory CLL. One third of such patients can achieve remission. Recently, alemtuzumab was compared with chlorambucil as initial therapy for symptomatic patients with CLL. Alemtuzumab produced higher complete and overall response rates and longer progression-free survival. In the first-line treatment of high-risk CLL, the addition of subcutaneous alemtuzumab to oral FC chemotherapy (fludarabine plus cyclophosphamide) improved progression-free survival and overall survival (the latter only in patients younger than 75 years) compared with FC chemotherapy alone. Alemtuzumab was withdrawn from the market in the United States and Europe in 2012 but is available free of charge from the company by compassionate request.

Ofatumumab is a humanized monoclonal antibody that binds to CD20, but to a different epitope than rituximab. This drug was recently approved by the FDA for the treatment of fludarabine- and alemtuzumab-refractory CLL. In the pivotal trial, the drug was given intravenously for 8 weeks and then monthly for 4 months. The overall response rate in this highly refractory population was 58%; the median progression-free survival was 6 months, with a median overall survival of 13.7 months. As with other monoclonal antibodies, the predominant side effects are infusion reactions, which tend to be more common with the initial doses.

B-Cell Receptor Signaling Pathway Inhibitors

A new category of targeted agents, called B-cell receptor signaling pathway inhibitors, is providing significant increased efficacy in CLL. Several inhibitors are being tested, targeting different kinases in the B-cell receptor pathway. Recent data have demonstrated that B-cell receptor signaling inhibitors are very active, particularly for the treatment of relapsed CLL. Ibrutinib is a first-in-class, oral covalent inhibitor of Bruton tyrosine kinase, an essential enzyme in B-cell receptor signaling, homing, and adhesion. It has been granted accelerated approval by the FDA for use in CLL. A randomized trial compared the effect of ibrutinib with that of the anti-CD20 antibody ofatumumab, both used alone, in the treatment of patients with relapsed or refractory CLL or small lymphocytic lymphoma. Ibrutinib significantly improved both progression-free and overall survival, compared with ofatumumab. Idelalisib is an inhibitor of phosphatidylinositol 3-kinase δ (PI3Kδ). Signaling through the B-cell receptor is mediated in part by activation of PI3Kδ. The δ isoform of PI3K is highly expressed in lymphoid cells, and it is the most critical isoform involved in the malignant phenotype of CLL. In a randomized, phase 3 study, patients with relapsed CLL who had clinically significant coexisting medical conditions that made them less able to undergo standard chemotherapy were randomized to receive eight intravenous infusions of rituximab plus either idelalisib 150 mg orally twice daily or placebo. The combination of idelalisib and rituximab, compared with rituximab alone, significantly improved response rate, progression-free survival, and overall survival.

Chimeric Antigen Receptor–Directed T Cells

Chimeric antigen receptors (CARs) are fusion proteins that combine antigen moieties and costimulatory T-cell receptors to redirect T cells toward malignant cells. All CARs targeting CLL have thus far directed the T cells through recognition of CD19.[14] Complete remissions, including absence of residual disease by sensitive testing, has been produced in refractory patients. *Cytokine release syndrome*, which can be life-threatening, is the initial toxicity. (Cytokine release syndrome, a rare phenomenon, also occurs in graft-versus-host disease after transplantation, severe infections, hemophagocytic lymphohistiocytosis or macrophage activation syndrome, and monoclonal antibody therapy; cytokines trigger an acute systemic inflammatory response that leads to endothelial and organ damage, microvascular leakage, and heart failure.[15]) Late toxicity includes the need for immunoglobulin replacement due to simultaneous eradication of normal B cells. CARs targeting other antigens in CLL are being investigated.

Stem Cell Transplantation

Autologous stem cell transplantation has no proven benefit in terms of survival or longterm disease control in CLL. Data on allogeneic stem cell transplantation are limited to young patients with refractory disease, in whom a long-term control rate of 40 to 55% has been reported. Nonmyeloablative stem cell transplantation, which works mainly by its graft-versusleukemia effect, has been used in older patients with CLL with some success.

Radiation Therapy

In CLL, radiation therapy is used palliatively to shrink unsightly or painfully enlarged nodal masses or an enlarged spleen.

Autoimmune and Infectious Manifestations

Autoimmune hemolytic anemia and immune-mediated thrombocytopenia do not correlate closely with the activity of CLL. Prednisone (60 to 100 mg/day) is indicated as treatment for autoimmune hemolytic anemia (Chapter 160) and for some cases of immune-mediated thrombocytopenia (Chapter 172) in CLL. If there is no response in 3 to 4 weeks, the treatment has failed, and the dose should be tapered over 1 to 2 weeks. If a response is obtained, the dose is reduced by 25% each week over 4 weeks. Patients for whom corticosteroids fail often respond to low-dose oral cyclosporine 100 mg three times a day. Other therapeutic options include splenectomy, intravenous immunoglobulin, rituximab, and alemtuzumab. Intravenous immunoglobulin (400 mg/kg every 3 to 4 weeks) significantly decreases the incidence of infections in patients with recurrent infections and hypogammaglobulinemia. However, the cost of this therapy is substantial.

PROGNOSIS

Approximately one third of patients who present with early-stage CLL never require therapy and have the same survival as age-matched controls. Frequent characteristics of such patients include WBC less than 30×10^9/L, hemoglobin greater than 13 g/dL, nondiffuse pattern on bone marrow biopsy, and slow lymphocyte doubling time.

Factors associated with shorter time to treatment failure as well as either poorer response to at least chemotherapy-based regimes, or shorter remission durations, include an unmutated *IgVH* gene, presence of 17p or llq

deletions, presence of ZAP70 and CD38, and mutations in *TP53, SF3Bl, ATM,* and *NOTCH*. A poor response to therapy is an adverse factor in all phases of the disease. As CLL progresses, the development of a prolymphocytic transformation (10% of cases) or transformation to large cell lymphoma (Richter transformation) portends a median survival time of less than 6 months. Other factors that may suggest transformation are the development of B symptoms (fevers, night sweats, weight loss), a markedly elevated lactate dehydrogenase level, or fluorodeoxyglucose-avid disease on positron emission tomography. A high incidence of second malignant tumors (10 to 20% of patients) either precedes or follows the diagnosis of CLL; the roles of therapy versus impaired immune surveillance as causative factors are unclear. Skin cancer, including melanoma, as well as colorectal and lung cancers particularly are common. CLL tends to develop in older people; in indolent cases, death occurs from other intercurrent illnesses seen in this age group. Almost all patients younger than 60 years and those with progressive disease die as a result of CLL, primarily from infections. Gram-positive organisms usually cause nonfatal infections early in CLL, but most deaths due to infection are associated with gram-negative bacterial or fungal infections. Infection with other opportunistic organisms, such as *Mycobacterium tuberculosis,* herpes virus, and *Pseudomonas jiroveci*, may also be fatal.

Grade A References

A1. Jabbour E, Kantarjian HM, Saglio G, et al. Early response with dasatinib or imatinib in chronic myeloid leukemia: a 3-year follow-up from a randomized phase 3 trial (DASISION). *Blood*. 2014;123:494-500.

A2. Hughes TP, Saglio G, Kantarjian H, et al. Early molecular response predicts outcomes in patients with chronic myeloid leukemia in chronic phase treated with frontline nilotinib or imatinib. *Blood*. 2014;12:1353-1360.

A3. Shah NP, Guilhot F, Cortes JE, et al. Long-term outcome with dasatinib after imatinib failure in chronic-phase chronic myeloid leukemia: follow-up of a phase 3 study. *Blood*. 2014;123: 2317-2324.

A4. Druker B, Guilhot F, O'Brien S, et al. Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N Engl J Med*. 2006;355:2408-2417.

A5. Larson RA, Hochhaus A, Hughes TP, et al. Nilotinib vs imatinib in patients with newly diagnosed Philadelphia chromosome-positive chronic myeloid leukemia in chronic phase: ENESTnd 3-year follow-up. *Leukemia*. 2012;26:2197-2203.

A6. Kantarjian HM, Shah NP, Cortes JE, et al. Dasatinib or imatinib in newly diagnosed chronic- phase chronic myeloid leukemia: 2-year follow-up from a randomized phase 3 trial (DASISION). *Blood*. 2012;119:1123-1129.

A7. Cortes JE, Kim DW, Kantarjian HM, et al. Bosutinib versus imatinib in newly diagnosed chronic-phase chronic myeloid leukemia: results from the BELA trial. *J Clin Oncol*. 2012;30:3486-3492.

A8. Knauf WU, Lissichkov T, Aldaoud A, et al. Phase III randomized study of bendamustine compared with chlorambucil in previously untreated patients with chronic lymphocytic leukemia. *J Clin Oncol*. 2009;27:4378-4384.

A9. Hallek M, Fischer K, Fingerle-Rowson G, et al. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet*. 2010;376:1164-1174.

A10. Goede V, Fischer K, Busch R, et al. Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions. *N Engl J Med*. 2014;370:1101-1110.

A11. Geisler CH, van T' Veer MB, Jurlander J, et al. Frontline alemtuzumab with fludarabine and cyclophosphamide prolongs progression-free survival in high-risk CLL. *Blood*. 2014;123:3255-3262.

A12. Wierda WG, Kipps TJ, Dürig J, et al. Chemoimmunotherapy with O-FC in previously untreated patients with chronic lymphocytic leukemia. *Blood*. 2011;117:6450-6458.

A13. Byrd JC, Brown JR, O'Brien S, et al. Ibrutinib versus ofatumumab in previously treated chronic lymphocytic leukemia. *N Engl J Med*. 2014;371:213-223.

A14. Furman RR, Sharman JP, Coutre SE, et al. Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *N Engl J Med*. 2014;370:997-1007.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Chin Y, Wang H, Kantarjian H, et al. Trends in chronic myeloid leukemia incidence and survival in the United States from 1975-2009. *Leuk Lymphoma*. 2013;54:1411-1417.
2. Elliott MA, Tefferi A. Chronic neutrophilic leukemia 2014: update on diagnosis, molecular genetics, and management. *Am J Hematol*. 2014;89:651-658.
3. Gotlib J. World Health Organization-defined eosinophilic disorders: 2014 update on diagnosis, risk stratification, and management. *Am J Hematol*. 2014;89:325-337.
4. Baccarani M, Deininger MW, Rosti G, et al. European LeukemiaNet recommendations for the management of chronic myeloid leukemia. *Blood*. 2013;122:872-874.
5. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: chronic myelogenous leukemia. Version 4. Available at: http://www.nccn.org/professionals/physician; 2013. Accessed February 2, 2015.
6. Marin D. Patient with chronic myeloid leukemia in complete cytogenetic response: what does it mean, and what does one do next? *J Clin Oncol*. 2014;32:379-384.
7. Cortes JE, Kim DW, Pinilla-Ibarz J, et al. A phase 2 trial of ponatinib in Philadelphia chromosome-positive leukemias. *N Engl J Med*. 2013;369:1783-1796.
8. Cortes J, Lipton JH, Rea D, et al. Phase 2 study of subcutaneous omacetaxine mepesuccinate after TKI failure in patients with chronic-phase CML with T315I mutation. *Blood*. 2012;120:2573-2580.
9. Rosenberg JD, Burian C, Waalen J, et al. Clinical characteristics and long-term outcome of young hairy cell leukemia patients treated with cladribine: a single-institution series. *Blood*. 2014;123:177-183.
10. Kreitman RJ. Hairy cell leukemia: new genes, new targets. *Curr Hematol Malig Rep*. 2013;8:184-195.
11. Foá R. Changes in the treatment landscape for chronic lymphoid leukemia. *N Engl J Med*. 2014;371:273-274.
12. Zhang D, Loughran TP Jr. Large granular lymphocytic leukemia: molecular pathogenesis, clinical manifestations, and treatment. *Hematology Am Soc Hematol Educ Program*. 2012;2012:652-659.
13. Jones JA, Byrd JC. How will B-cell-receptor-targeted therapies change future CLL therapy? *Blood*. 2014;123:1455-1460.
14. Porter D, Levine B, Kalos M, et al. Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia. *N Engl J Med*. 2011;725-733. *Complete remission in a refractory patient after CD19-directed CAR*.
15. Xu X-J, Tang Y-M. Cytokine release syndrome in cancer immunotherapy with chimeric antigen receptor engineered T cells. *Cancer Lett*. 2014;343:172-178.

REVIEW QUESTIONS

1. 35-year-old man was diagnosed with chronic phase chronic myelogenous leukemia (CML) following a regular check-up. He had no significant comorbidities or medical history and was initiated on imatinib 400 mg daily. At 12 months, he achieved a complete cytogenetic response with transcripts level of 0.6% (IS). At 18 and 24 months, his transcripts levels were 0.5% and 0.4% (IS), respectively. Which of the following would be the next step in treatment?

A. Continue imatinib.
B. Switch to nilotinib.
C. Switch to dasatinib.
D. Switch to ponatinib
E. Consider allogeneic stem cell transplantation

Answer: A "Resistance" to imatinib therapy requires a change of therapy to other tyrosine kinase inhibitors (TKIs), TKI combinations with chemotherapy, or consideration of allogeneic stem cell transplantation. Resistance can be defined as persistent 100% Ph positivity after 6 months of therapy, Ph positivity of more than 35% after 12 months of therapy, or cytogenetic or hematologic relapse. It is important to emphasize that many patients with CML resistance to a TKI therapy may be noncompliant with the treatment. This should be discussed clearly with patients when they exhibit signs of CML progression by either molecular or fluorescent in situ hybridization studies. If they are noncompliant, they may continue on the same TKI treatment with emphasis on compliance and evaluated 3 to 6 months later, before CML resistance is declared. Among patients in complete cytogenetic response on a particular TKI, failure to achieve a major molecular response in a patient with a complete cytogenetic response does not, at present, indicate imatinib resistance to the particular TKI or the need to change therapy. The rate of resistance to imatinib in chronic phase CML is less than 4% per year; for those who achieve a complete cytogenetic response, the rate of resistance beyond year 3 of imatinib therapy is 1% or less, suggesting the durable stability of a complete cytogenetic response on imatinib and the predictability of the CML course after such a response is obtained.

2. A 68-year-old woman was diagnosed with chronic phase CML after a regular check-up. She was initiated on dasatinib 100 mg daily. At 6 months, she achieved a major cytogenetic response with transcripts level of 9% (IS). At 12 months, she lost her cytogenetic response with transcripts levels of 30% (IS). A T315I mutation was identified. Which of the following would be the next step in treatment?

A. Increase dasatinib dose to 140 mg daily.
B. Switch to nilotinib.
C. Switch to bosutinib.
D. Switch to ponatinib.
E. Switch to omacetaxine.

Answer: D Patients who develop resistance to first-line TKI in chronic phase are offered second-line or third-line TKI based on their mutation analysis. A T315I mutation in the CML clone requires therapy with ponatinib and (as of today) early consideration of allogeneic stem cell transplantation until the results of ponatinib therapy mature. Mutations involving V299L, T315A, or F317LIV/1/C are sensitive to nilotinib therapy. Mutations involving Y253H, E255K/V, or F359/V/C/I are sensitive to dasatinib and bosutinib therapy. Omacetaxine is approved for patients who had failed two TKIs.

3. A 34-year-old woman was diagnosed with chronic phase CML and was initiated on imatinib 400 mg daily. She achieved a complete cytogenetic response by 12 months of therapy. Her last transcripts level was reported to be 0.5% (IS). She comes to see you for advice regarding her desire to become pregnant. Which of the following would be the next step in treatment?

A. Stay on imatinib and try to conceive.
B. Switch to dasatinib and try to conceive.
C. Switch to nilotinib and try to conceive.
D. Switch to ponatinib and try to conceive.
E. You should not conceive while you are taking any TKIs.

Answer: E CML during pregnancy may be controlled with pheresis in the first trimester and then with hydroxyurea until delivery. Use of interferon-α during pregnancy has been reported anecdotally to be safe. An analysis of 125 babies delivered to women with CML on imatinib therapy (who discontinued imatinib after the pregnancy was known) showed most babies to be healthy. However, the study showed imatinib therapy to be associated with syndrome of ocular, skeletal, and renal abnormalities in three babies delivered. Therefore, imatinib (and presumably other TKIs, although there is little experience with them) should be discontinued immediately when pregnancy is documented, but abortion is not recommended because fetal malformations are rare. Partners of men with CML on TKI therapy who become pregnant have delivered normal babies.

185 NON-HODGKIN LYMPHOMAS

PHILIP J. BIERMAN AND JAMES O. ARMITAGE

DEFINITION

Lymphomas are solid tumors of the immune system. Increasing knowledge of the biology of the immune system has led to a corresponding increase in the understanding of these malignancies. In addition to better systems of classification and clinical evaluation, this new knowledge has led to the development of new therapies. Beneficial treatment is available for essentially every patient with non-Hodgkin lymphoma. The overall survival of lymphoma patients has increased steadily over the past 30 years, and many patients can be cured.

EPIDEMIOLOGY

In the United States, approximately 70,000 new cases of non-Hodgkin lymphoma are diagnosed annually, and about 19,000 people are estimated to die each year of this disease. Non-Hodgkin lymphomas account for about 4% of new cancers in the United States and result in about 3% of cancer deaths. The U.S. lifetime risk for developing non-Hodgkin lymphoma is estimated to be 2.4% (1 in 42) for men and 1.90% (1 in 52) for women. In 2010 the U.S. age-adjusted incidence rate for non-Hodgkin lymphoma was about 26.82 per 100,000 for men and 17.39 per 100,000 for women.[1] The incidence rate increases dramatically with age and is higher in whites than in other ethnic groups.

Geographic differences in the incidence of non-Hodgkin lymphomas vary as much as five-fold. The highest rates are seen in the United States, Europe, and Australia, whereas lower rates are seen in Asia. Even more striking are geographic differences in the incidence of certain types of non-Hodgkin lymphoma, such as Burkitt lymphoma, follicular lymphoma, extranodal natural killer (NK)/T-cell nasal lymphoma, and adult T-cell leukemia/lymphoma (see later).

Between 1950 and the beginning of the 21st century, the incidence rate for non-Hodgkin lymphomas in the United States increased by about 3 to 4% yearly. Since then, the incidence rate for non-Hodgkin lymphoma has reached a plateau. Increases occurred among both men and women in all parts of the world. The increase in incidence was partially related to the aging population (Fig. 185-1) and to the acquired immunodeficiency syndrome (AIDS) epidemic (Chapter 393). Occupational and environmental exposures (e.g., agricultural chemicals) may also explain some of the increase. Finally, some of the increase may be explained by improvements in the ability of pathologists to diagnose lymphoma and by improvements in imaging techniques.

PATHOBIOLOGY

For most cases of non-Hodgkin lymphoma, the cause is unknown, although genetic, environmental, and infectious agents have been implicated (Table 185-1).[2-4]

Genetic Factors

Familial non-Hodgkin lymphoma clusters have been described, and there is a slightly higher risk for non-Hodgkin lymphoma among siblings and first-degree relatives of patients with lymphoma or other hematologic malignancies. There is increasing recognition that host genetics are involved in the development of lymphomas. The incidence of non-Hodgkin lymphoma has

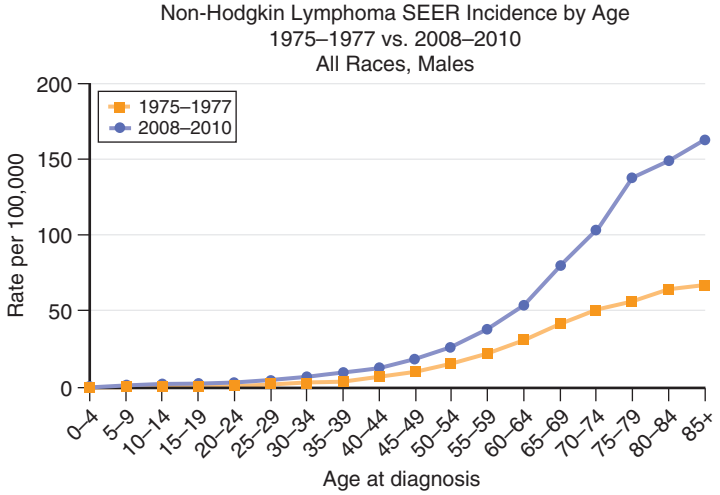

FIGURE 185-1. Non-Hodgkin lymphoma incidence by age in men, 1975-1977 versus 2008-2010. (From the Surveillance, Epidemiology, and End Results [SEER] Program of the National Cancer Institute, http://seer.cancer.gov/csr/1975_2010/.)

TABLE 185-1 FACTORS ASSOCIATED WITH THE DEVELOPMENT OF NON-HODGKIN LYMPHOMA

Inherited immune disorders
Severe combined immunodeficiency disease
Common variable immunodeficiency disease
Wiskott-Aldrich syndrome
Ataxia-telangiectasia
X-linked lymphoproliferative disorder
Autoimmune lymphoproliferative syndrome
Acquired immune disorders
Solid organ transplantation
Acquired immunodeficiency syndrome (AIDS)
Methotrexate therapy for autoimmune disorders
Rheumatoid arthritis and systemic lupus erythematosus
Sjögren syndrome
Hashimoto thyroiditis
Infectious agents
Epstein-Barr virus
Human T-lymphotropic virus type 1
Human herpesvirus type 8
Hepatitis C virus
Helicobacter pylori
Borrelia burgdorferi
Chlamydia psittaci
Campylobacter jejuni
Occupational and environmental exposure
Herbicides
Organic solvents
Hair dyes
Ultraviolet light
Diet
Smoking
Drugs

been associated with polymorphisms in a variety of genes related to immunity, including chemokines, tumor necrosis factor, interleukin-10 (IL-10), and lymphotoxin-α. Polymorphisms in other genes related to the cell cycle and apoptosis have also been associated with an increased risk for developing lymphoma.

Immune System Abnormalities

Several inherited disorders increase the risk for developing non-Hodgkin lymphoma as much as 250-fold (see Table 185-1). In some of these conditions, the lymphoma may be related to Epstein-Barr virus (EBV; Chapter 377). For example, patients with X-linked lymphoproliferative disorder have mutations in the *SH2D1A* gene, which encodes proteins that regulate the host immune response against EBV-infected cells. Patients may develop fatal infectious mononucleosis or non-Hodgkin lymphoma after primary exposure to EBV. Acquired immunodeficiency states are also associated with an increased risk for non-Hodgkin lymphoma. For example, post-transplantation lymphoproliferative disorders occur in as many as 20% of solid organ transplant recipients, related to the proliferation of B lymphocytes that have been transformed during immunosuppressive therapy. The risk for non-Hodgkin lymphoma is also increased more than 100-fold in patients infected with the human immunodeficiency virus (HIV). Almost all central nervous system (CNS) lymphomas and approximately 50% of other lymphomas in patients with AIDS are related to EBV. Some studies have shown a two-fold increase in the incidence of non-Hodgkin lymphomas among patients with rheumatoid arthritis (Chapter 264), and the risk for marginal zone lymphomas is increased approximately 30- to 40-fold in patients with Sjögren syndrome (Chapter 268). Increases in the incidence of thyroid lymphoma are seen in patients with Hashimoto thyroiditis (Chapter 226). Enteropathy-type T-cell lymphomas are associated with celiac disease (Chapter 140). Patients with the autoimmune lymphoproliferative syndrome, associated with mutations in the *FAS* gene, also appear to be at higher risk for developing lymphoma.

Infectious Agents

EBV is associated with the majority of post-transplantation lymphoproliferative disorders and many AIDS-associated lymphomas. This viral genome is detectable in more than 95% of cases of endemic Burkitt lymphoma and in approximately 15 to 35% of cases of sporadic Burkitt lymphoma and AIDS-associated lymphomas. This virus is also associated with EBV-positive diffuse large B-cell lymphoma of elderly people, plasmablastic lymphoma, and extranodal NK/T-cell lymphoma.

The human T-lymphotropic virus type 1 (HTLV-1; Chapter 378) is detectable in virtually all cases of adult T-cell leukemia/lymphoma. The risk for lymphoma is approximately 3% in patients infected with HTLV-1. In endemic areas, up to 50% of all non-Hodgkin lymphomas may be related to HTLV-1.

Human herpesvirus-8 (HHV-8, Kaposi sarcoma–associated herpesvirus; Chapter 393), which is associated with expansion of the B-cell population, is also associated with primary effusion lymphoma (see later) in immunocompromised patients and with multicentric Castleman disease. Patients with primary effusion lymphoma are often coinfected with EBV.

Epidemiologic evidence has linked hepatitis C virus (Chapter 149) to lymphoplasmacytic lymphoma associated with type II cryoglobulinemia, nodal marginal zone lymphoma, and splenic marginal zone lymphoma. Chronic antigenic stimulation from this virus may lead to the emergence of malignant B-cell clones.

Helicobacter pylori is associated with gastric lymphoma (Chapter 192) of extranodal marginal zone/mucosa-associated lymphoid tissue (MALT). Colonized patients develop gastritis from chronic antigenic stimulation mediated by T cells, which respond to *H. pylori*–specific antigens and emergence of malignant B-cell clones. *Borrelia burgdorferi* (Chapter 321) has been associated with marginal zone B-cell lymphoma of the skin. Evidence also links *Chlamydia psittaci* (Chapter 318) with ocular adnexal lymphomas and *Campylobacter jejuni* with immunoproliferative small intestinal disease (Chapter 303).

Environmental and Occupational Exposure

Agricultural chemicals have been associated with an increased risk for developing non-Hodgkin lymphomas, and the strongest associations involve phenoxy herbicides such as 2,4-dichlorophenoxyacetic acid (2,4-D), which was also a component of Agent Orange (Chapter 19). An increased risk has also been associated with ionizing radiation (Chapter 20), organic solvents, hair dyes, and nitrates in drinking water, although contradictory results have been reported. Some studies have also linked non-Hodgkin lymphomas to high-fat diets and ultraviolet radiation (Chapter 180). The risk for non-Hodgkin lymphomas is increased approximately 20-fold after treatment for Hodgkin lymphoma (Chapter 186). Heavy smokers (Chapter 32) have an increased risk for developing follicular lymphoma. Low levels of vitamin D have been associated with increased risk for recurrent lymphoma, as well as poor outcome. Anti–tumor necrosis factor (anti-TNF) agents might be associated with an increased risk for developing lymphoma—particularly hepatosplenic T-cell lymphoma. There is also a reported association of breast implants with the development of anaplastic large cell lymphoma.

Pathology

Non-Hodgkin lymphomas are derived from cells of the immune system at varying stages of differentiation. In some cases, the cell of origin is directly linked to the morphology, immunophenotype, and clinical behavior of the lymphoma (Fig. 185-2 and Table 185-2).

The transformation of cells from the normal immune system into malignant lymphoma reflects the acquisition of specific genetic abnormalities. In many cases, cytogenetic studies can identify chromosomal translocations that underlie the development or progression of the lymphoma. In most cases of non-Hodgkin lymphoma, the activation of proto-oncogenes is the major abnormality, but occasionally chromosomal translocations can lead to fusion genes that code for chimeric proteins. In addition, some cases are associated with deletion of tumor suppressor genes. Specific genetic abnormalities are associated with some specific subtypes of non-Hodgkin lymphoma (Table 185-3). It is becoming clear that the tumor microenvironment from cells of the host immune system is important in tumor cell survival and response to therapy.[5,6]

Classification

Recognition of the Reed-Sternberg cell approximately 100 years ago made it possible to define Hodgkin lymphoma (Chapter 186) as a distinct entity, whereas other lymphomas were included under the heading "non-Hodgkin lymphomas." In the 1990s, a classification system incorporating morphologic, immunologic, genetic, and clinical information (the Revised

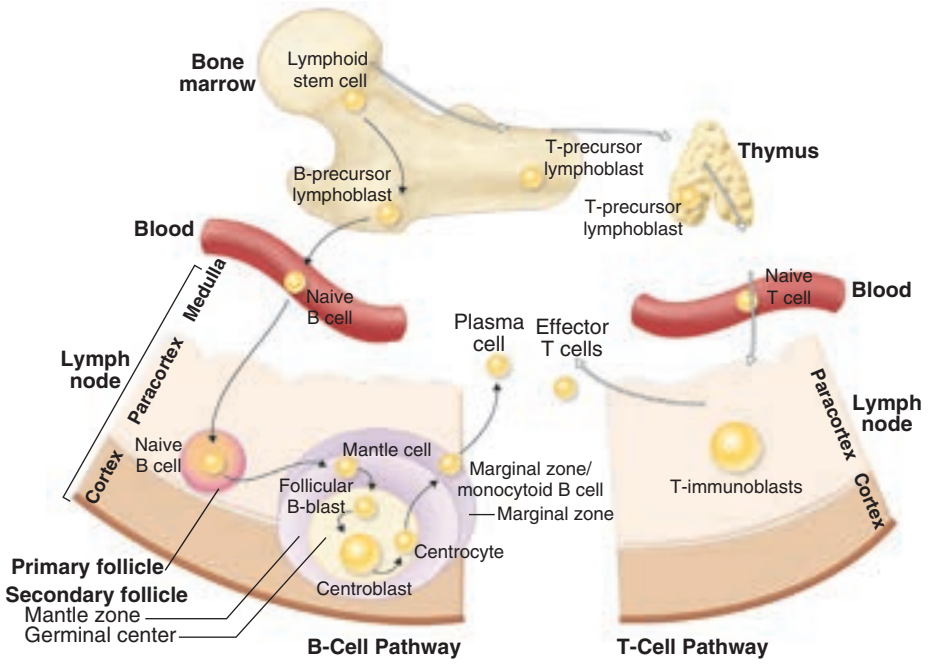

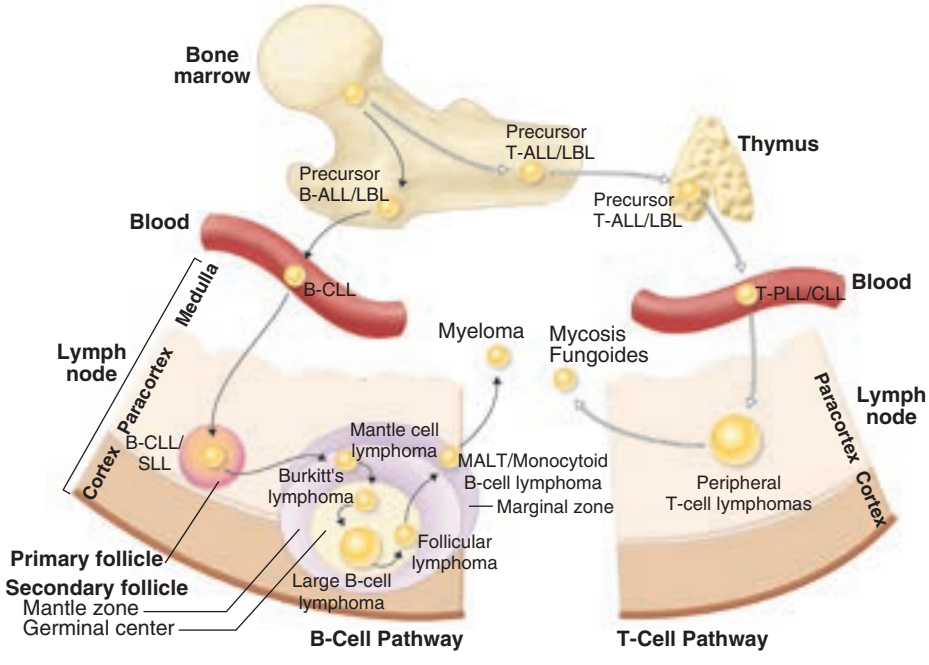

FIGURE 185-2. Postulated normal counterparts of currently recognized B- and T-cell malignancies. A, Schema of normal B- and T-cell differentiation. Bone marrow–derived lymphoid stem cells differentiate into committed B-cell precursors or into T-cell precursors that undergo further maturation in the thymus. These B- and T-cell precursors mature into naïve B or T cells that circulate to lymph nodes. After antigen exposure, normal B blasts proliferate and undergo further differentiation in the germinal center of the secondary follicle. The germinal center is surrounded by a mantle zone and a marginal zone. Antigen-specific B cells generated in the germinal center leave the follicle and reappear in the marginal zone. Thereafter, immunoglobulin-producing plasma cells accumulate in the lymph node medulla and subsequently exit to the periphery. Antigen-dependent T-cell proliferation occurs in the lymph node paracortex. After antigen exposure, mature T cells become immunoblasts and, subsequently, antigen-specific effector T cells that exit to the periphery. The postulated normal counterparts of many currently recognized T- and B-cell neoplasms are shown. **B,** T- and B-cell malignancies derived from the postulated normal counterparts shown in **A.** ALL = acute lymphoblastic leukemia; CLL = chronic lymphocytic leukemia; LBL = lymphoblastic lymphoma; MALT = mucosa-associated lymphoid tissue; PLL = prolymphocytic leukemia; SLL = small lymphocytic lymphoma.

European-American Lymphoma classification, or REAL) was developed to identify distinct clinicopathologic subgroups representing diseases that can be recognized by clinicians. This system was subsequently adopted as the World Health Organization (WHO) classification of lymphomas in 2008 (Table 185-4).[7]

The WHO classification divides lymphomas on the basis of B-cell or T/NK-cell origin and whether they are derived from primitive precursor cells or from more mature "peripheral" cells. Specific clinical and pathologic entities are recognized within each of these groupings. In the United States and Europe, 85 to 90% of non-Hodgkin lymphomas are B cell in origin.

The most frequent type is diffuse large B-cell lymphoma, which represents 31% of all non-Hodgkin lymphomas worldwide. The next most frequent type is follicular lymphoma, which represents 22% of cases. Follicular lymphoma is relatively more frequent in North America and Western Europe and less frequent in Asia. Less common types, each representing between 5 and 10% of all non-Hodgkin lymphomas, are extranodal marginal zone/MALT

TABLE 185-2 TYPICAL IMMUNOPHENOTYPES OF COMMON NON-HODGKIN LYMPHOMAS

LYMPHOMA	CD20	CD3	CD10	CD5	CD23	OTHER
Small lymphocytic	+	–	–	+	+	
Lymphoplasmacytic	+	–	–	–	–	Cytoplasmic Ig^+
Extranodal marginal zone MALT	+	–	–	–	–	
Nodal marginal zone	+	–	–	–	–	
Follicular	+	–	+	–		
Mantle cell	+		–	+	–	Cyclin $D1^+$
Diffuse large B cell	+	–				
Mediastinal large B cell	+	–				
Burkitt	+	–	+	–		TdT^-
Precursor T lymphoblastic	–	+/–				TdT^+, $CD1a^{+/-}$, $CD7^+$
Anaplastic large T cell	–	+/–				$CD30^+$, $CD15^-$, EMA^+, ALK^+
Peripheral T cell	–	+/–				Other pan-T variable

ALK = anaplastic lymphoma kinase; EMA = epithelial membrane antigen; MALT = mucosa-associated lymphoid tissue; TdT = terminal deoxynucleotidyl transferase.

TABLE 185-3 CHROMOSOMAL TRANSLOCATIONS CHARACTERISTIC OF NON-HODGKIN LYMPHOMA

LYMPHOMA SUBTYPE	TRANSLOCATION	GENES INVOLVED	FREQUENCY (%)
Diffuse large B cell	t(3q27) t(14;18)(q32;q21) t(18;14)(q24;q32)	*BCL6* *IgH, BCL2* *MYC (c-Myc), IgH*	35 15-20 <5
Burkitt	t(8;14)(q24;q32) t(8;22)(q24;q11) t(2;8)(p12;q24)	*MYC, IgH* *MYC, IgL* *IgK, MYC*	100% have one of these, most commonly t(8;14)
Follicular	t(14;18)(q32;q21)	*IgH, BCL2*	~90
Mantle cell	t(11;14)(q13;q32)	*BCL1, IgH*	>90
ALCL	t(2;5)(p23;q35)	*ALK, NPM*	>80 of ALK+ ALCLs
MALT	t(11;18)(q21;q21) t(14;18)(q21;q32) t(1;14)(p22;q32)	*API2, MALT1* *IgH, MALT1* *BCL10, IgH*	35 20 10

ALCL = anaplastic large cell lymphoma; ALK = anaplastic lymphoma kinase; MALT = mucosa-associated lymphoid tissue.

lymphomas, peripheral T-cell lymphomas, small lymphocytic lymphoma, and mantle cell lymphoma. Other types each represent less than 2% of non-Hodgkin lymphomas seen in the United States.

The non-Hodgkin lymphomas recognized in the WHO classification have clinically distinctive characteristics (Table 185-5), such that an experienced hematopathologist can accurately classify 85% or more of patients by WHO criteria when adequate material is available. Some diagnoses, such as follicular lymphoma, can be made with a high degree of accuracy without immunologic or genetic studies. The diagnosis of T-cell lymphomas cannot be made accurately without immunophenotyping. Cytogenetic studies and molecular genetic studies (fluorescent in situ hybridization, or FISH) can help resolve difficult differential diagnoses. For example, the presence of a t(8;14) translocation supports the diagnosis of Burkitt lymphoma, whereas a t(11;14) with cyclin D1 overexpression can confirm the diagnosis of mantle cell lymphoma (see Tables 185-2 and 185-3).

The use of DNA microarrays has allowed the identification of distinct subsets of patients with diffuse large B-cell lymphoma. Patients with histologically identical lymphomas can be divided into those with gene expression patterns resembling normal germinal center B cells (GCB), those whose tumors resemble activated post–germinal center B cells (ABC), or those with patterns resembling that seen in Hodgkin lymphoma. The last pattern is most frequently found in young women who present with large mediastinal masses. Use of immunohistochemistry to subdivide diffuse large B-cell lymphoma into GCB and ABC subtypes is widely used but does not correlate perfectly with DNA microarray results.

CLINICAL MANIFESTATIONS

The most common presentation of non-Hodgkin lymphoma is lymphadenopathy (Fig. 185-3; Chapter 168). In many cases, patients notice cervical, axillary, or inguinal adenopathy and seek a physician's advice. In general, lymph nodes containing lymphoma are firm, nontender, and not associated with a regional infection. In other patients, lymphadenopathy occurring in sites such as the mediastinum or retroperitoneum causes symptoms that bring the patient to the physician. Chest pain, cough, superior vena cava syndrome, abdominal pain, back pain, spinal cord compression, and symptoms of renal insufficiency associated with ureteral compression are characteristic.

Non-Hodgkin lymphomas are often associated with systemic symptoms that may lead to the diagnosis. The most obvious symptoms are fever, night sweats, and unexplained weight loss. Any of these symptoms without an obvious cause should lead a physician to consider the diagnosis of lymphoma. Other, less characteristic symptoms include fatigue, which is frequently present at the time of diagnosis if the patient is questioned carefully, and pruritus.

Non-Hodgkin lymphomas can involve essentially any organ in the body, and malfunction of that organ can cause symptoms that lead to the diagnosis. Examples include neurologic symptoms with primary brain lymphoma (Chapter 189), shortness of breath with MALT lymphomas in the lung, epigastric pain and vomiting with gastric MALT lymphomas or diffuse large B-cell lymphomas (Chapter 192), bowel obstruction with small bowel lymphomas (Chapter 193), testicular masses with diffuse large B-cell lymphoma (Chapter 200), and skin lesions with cutaneous lymphomas (Chapter 440). Many lymphomas involve the bone marrow and occasionally cause extensive myelophthisis (Chapter 165) and bone marrow failure. These patients may present with infections, bleeding, and anemia.

Non-Hodgkin lymphomas can also manifest with a variety of immunologic abnormalities. For example, autoimmune hemolytic anemia (Chapter 160) and immune thrombocytopenia (Chapter 172) can be the presenting manifestations of non-Hodgkin lymphoma, especially small lymphocytic

TABLE 185-4 WORLD HEALTH ORGANIZATION CLASSIFICATION OF NON-HODGKIN LYMPHOMA (2008)

PRECURSOR LYMPHOID NEOPLASMS
B-lymphoblastic leukemia/lymphoma
T-lymphoblastic leukemia/lymphoma
MATURE B-CELL NEOPLASMS
Chronic lymphocytic leukemia/small lymphocytic lymphoma
B-cell prolymphocytic leukemia
Splenic marginal zone lymphoma
Hairy cell leukemia
Splenic lymphoma/leukemia, unclassifiable
Lymphoplasmacytic lymphoma
Heavy chain diseases
Plasma cell neoplasms
Extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue (MALT lymphoma)
Nodal marginal zone lymphoma
Follicular lymphoma
Primary cutaneous follicle center lymphoma
Mantle cell lymphoma
Diffuse large B-cell lymphoma, NOS (see Table 185-10 for variants and subtypes)
Burkitt lymphoma
B-cell lymphoma, unclassifiable, with features intermediate between diffuse large B-cell lymphoma and Burkitt lymphoma
B-cell lymphoma, unclassifiable, with features intermediate between diffuse large B-cell lymphoma and classical Hodgkin lymphoma
MATURE T- AND NK-CELL NEOPLASMS
Adult T-cell leukemia/lymphoma
Extranodal NK/T-cell lymphoma, nasal type
Enteropathy-associated T-cell lymphoma
Hepatosplenic T-cell lymphoma
Subcutaneous panniculitis-like T-cell lymphoma
Mycosis fungoides
Sézary syndrome
Primary cutaneous $CD30^+$ T-cell lymphoproliferative disorders
Peripheral T-cell lymphoma, NOS
Angioimmunoblastic T-cell lymphoma
Anaplastic large cell lymphoma, ALK positive
Anaplastic large cell lymphoma, ALK negative

ALK = anaplastic lymphoma kinase; MALT = mucosa-associated lymphoid tissue; NK = natural killer; NOS = not otherwise specified.

lymphoma/chronic lymphocytic leukemia as well as other subtypes, including diffuse large B-cell lymphoma. Peripheral neuropathies (Chapter 420), often associated with overproduction of a monoclonal protein, can be seen in a variety of subtypes but are most characteristic of lymphoplasmacytic lymphoma; sometimes they are also seen with POEMS syndrome (polyneuropathy, organomegaly, endocrinopathy, M protein, skin changes; Chapter 187). Paraneoplastic neurologic complications of non-Hodgkin lymphoma include demyelinating polyneuropathy, Guillain-Barré syndrome, autonomic dysfunction, and peripheral neuropathy. Paraneoplastic syndromes (Chapter 179) associated with non-Hodgkin lymphoma can affect the skin (e.g., pemphigus), kidney (e.g., glomerulonephritis), and miscellaneous organ systems (e.g., vasculitis, dermatomyositis, cholestatic jaundice).

The differential diagnosis in patients with non-Hodgkin lymphoma is broad. Any cause of lymphadenopathy or splenomegaly can potentially be confused with non-Hodgkin lymphoma (Chapter 168). However, this confusion is resolved by an appropriate biopsy. It is extremely important to recognize that the diagnosis of non-Hodgkin lymphoma must be considered in patients with compatible clinical presentations and then confirmed by an adequate biopsy read by an experienced hematopathologist. The diagnosis should never be inferred, and patients should not be treated until the diagnosis is confirmed by biopsy. This is also true for patients who achieve a

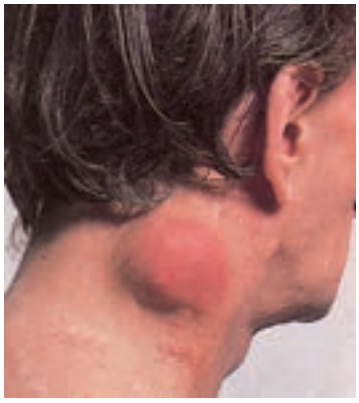

FIGURE 185-3. Non-Hodgkin lymphoma. Despite the redness of the skin over the enlarged lymph node in this patient, the lesion was completely painless. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine,* 3rd ed. London: Mosby; 2003.)

TABLE 185-5 PRESENTING CLINICAL CHARACTERISTICS OF COMMON NON-HODGKIN LYMPHOMA SUBTYPES

TYPE OF LYMPHOMA	MEDIAN AGE (yr)	MALE (%)	STAGE (%) I	STAGE (%) IV	B SYMPTOMS (%)	BONE MARROW INVOLVED (%)
B-CELL LYMPHOMAS						
Small lymphocytic	65	53	4	83	33	72
Lymphoplasmacytic	63	53	7	73	13	73
Extranodal marginal zone MALT	60	48	39	31	19	14
Nodal marginal zone	58	42	13	40	37	32
Follicular	59	42	18	51	28	42
Mantle cell	63	74	13	71	28	64
Diffuse large B cell	64	55	25	33	33	16
Mediastinal large B cell	37	34	10	31	38	3
Burkitt	31	89	37	38	22	33
PRECURSOR B/T-CELL LYMPHOMAS						
Precursor T lymphoblastic	28	64	0	75	21	50
T-CELL LYMPHOMAS						
Anaplastic large T cell	34	69	19	39	53	13
Peripheral T cell	61	55	8	65	50	36

B symptoms = fevers, night sweats, and weight loss; MALT = mucosa-associated lymphoid tissue.
Adapted from Armitage JO, Weisenburger DD, for the Non-Hodgkin Lymphoma Classification Project. New approach to classifying non-Hodgkin lymphomas: clinical features of the major histologic subtypes. *J Clin Oncol.* 1998;16:2780-2795.

complete remission with initial therapy; they should not be treated for presumed relapse on the basis of symptoms or abnormal images without a biopsy.

DIAGNOSIS

Each new patient with a non-Hodgkin lymphoma should be thoroughly evaluated in a systematic manner (Table 185-6). Because subtle pathologic distinctions may alter therapy, the most important issue in managing non-Hodgkin lymphoma is to establish an accurate diagnosis. Core needle biopsies can occasionally be used for a primary diagnosis if the specimen is handled properly. Fine-needle aspirates should not be used to diagnose lymphoma, and they preclude an accurate diagnosis of the specific subtype of non-Hodgkin lymphoma. In most cases, an excisional biopsy is necessary (and is always preferred) for the initial diagnosis. Another biopsy should be performed if sufficient material is not obtained. Review by an experienced hematopathologist is essential.

Staging and Prognostic Systems

After diagnosis, a meticulous staging evaluation is necessary to estimate prognosis and determine therapy. Staging requires a careful history and physical examination; complete blood count; renal and hepatic function tests; serum lactate dehydrogenase (LDH) level; computed tomography (CT) of the chest, abdomen, and pelvis; and bone marrow biopsy. Positron emission tomography (PET) can identify initial sites of involvement and, after treatment, can distinguish persisting lymphoma from residual fibrosis in masses seen on CT. The most common staging system is the Ann Arbor classification, which separates patients into four stages based on anatomic sites of disease (Table 185-7). In addition, each stage is subdivided into A (no defined general symptoms) and B (unexplained weight loss >10% of body weight in the previous 6 months, unexplained temperature >38° C, or night sweats) categories. Known sites of disease can be reexamined later to evaluate the response to therapy.

TABLE 185-6 TYPICAL EVALUATION OF A PATIENT NEWLY DIAGNOSED WITH NON-HODGKIN LYMPHOMA

1. Biopsy to establish diagnosis
2. Careful history and physical examination
3. Laboratory evaluation
   A. Complete blood count
   B. Chemistry screen, including lactate dehydrogenase
4. Imaging studies
   A. Computed tomography of chest, abdomen, and pelvis
   B. Positron emission tomography
5. Additional biopsies
   A. Bone marrow
   B. Any other suspicious site if results would change therapy

TABLE 185-7 STAGING OF NON-HODGKIN LYMPHOMA

STAGE	DESCRIPTION
I	Involvement of a single lymph node region (I) or a single extralymphatic organ or site (I_E)
II	Involvement of two or more lymph node regions on the same side of the diaphragm (II) or localized involvement of an extralymphatic organ or site and one or more lymph node regions on the same side of the diaphragm (II_E)
III	Involvement of lymph node regions on both sides of the diaphragm (III), which may also be accompanied by localized involvement of an extralymphatic organ or site (III_E) or by involvement of the spleen (III_S) or both (III_{SE})
IV	Diffuse or disseminated involvement of one or more extralymphatic organs or tissues with or without associated lymph node enlargement
Subtypes	
A	No B symptoms
B	B symptoms: unexplained weight loss ≥10% of body weight in prior 6 months, unexplained fever with temperature >38° C, or night sweats

Adapted from Carbone PP, Kaplan HS, Musshoff K, et al. Report of the Committee on Hodgkin Disease Staging Classification. *Cancer Res.* 1971;31:1860-1861.

Although a wide variety of patient factors (e.g., age, symptoms, LDH level) and tumor factors (e.g., bulk, gene expression pattern, proliferation rate) can affect treatment outcomes, two prognostic systems can help in choosing therapy and determining an accurate prognosis. The International Prognostic Index (IPI; Table 185-8) is the most widely used method to predict treatment outcome and survival. The IPI is based on five adverse factors (age >60 years, performance status ≤2, elevated serum LDH level, two or more extranodal sites of disease, Ann Arbor stage III or IV), which are summed to give the score. This index was developed for patients with diffuse aggressive lymphoma (predominantly diffuse large B-cell lymphoma), but it can be used to predict treatment outcome with any subtype. For young patients, an abbreviated index that uses only reduced performance status, elevated serum LDH level, and high stage can be applied. Because patients with follicular lymphoma rarely have a reduced performance status or a large number of extranodal sites, an alternative index termed the *Follicular Lymphoma International Prognostic Index* (FLIPI) was developed. It substitutes more than four nodal areas of involvement and a hemoglobin less than 12 g/dL as staging criteria and does a somewhat better job of predicting treatment outcome in follicular lymphoma.

Other prognostic indices have been developed for mantle cell lymphoma and the peripheral T-cell lymphomas.

TREATMENT

Rx

Lymphomas may behave in an indolent or an aggressive manner. The behavior of many of these neoplasms is distinctive, but within each category, behavior is frequently influenced by disease site, tumor bulk, and performance status of the patient. Some lymphomas can be managed, at least initially, with observation, whereas other situations are medical emergencies, such as spinal cord compression (Chapter 189). It is important to consider three questions before starting therapy: (1) Does treatment have curative potential? (2) Can treatment prolong survival? (3) Will treatment alleviate symptoms?

SPECIFIC TYPES OF NON-HODGKIN LYMPHOMAS

Precursor T-Cell and B-Cell Lymphomas

These tumors are nodal or other solid tissue infiltrates of cells that are morphologically and immunophenotypically identical to the immature cells seen in B-cell or T-cell acute lymphoblastic leukemia (Chapter 183). Patients who have predominantly nodal disease with minimal or no involvement of the bone marrow are frequently classified as having *lymphoblastic lymphoma,* whereas those with more than 25% neoplastic cells in the marrow are classified as having *lymphoblastic leukemia.* These distinctions are arbitrary and reflect the stage of disease rather than different diagnoses. These neoplasms are more common in children than in adults.

B-cell precursor lymphomas frequently manifest as solid tumors with involvement of the skin and bones, whereas T-cell neoplasms typically

TABLE 185-8 INTERNATIONAL PROGNOSTIC INDEX

CATEGORY	SCORE (NO. OF RISK FACTORS)
ALL PATIENTS*	
Low	0 or 1
Low intermediate	2
High intermediate	3
High	4 or 5
AGE-ADJUSTED INDEX, PATIENTS ≤60 YEARS†	
Low	0
Low intermediate	1
High intermediate	2
High	3

*Adverse factors for all patients: age >60 yr, ↑LDH, performance status 2-4, >1 extranodal site, Ann Arbor stage III or IV.
†Adverse factors for patients ≤60 yr: ↑LDH, performance status 2 –4, Ann Arbor stage III or IV.
LDH = lactate dehydrogenase.
Adapted from Shipp M, Harrington D, Anderson J, et al. A predictive model for aggressive non-Hodgkin lymphoma. *N Engl J Med.* 1993;329:987-994.

manifest as a mediastinal mass in a young male. Involvement of the CNS is common. Approximately 90% of patients who present with lymphoblastic lymphoma have a T-cell phenotype, whereas about 85% of patients who present with acute lymphoblastic leukemia have a B-cell phenotype. Adverse prognostic characteristics include CNS involvement, stage IV disease, and elevated LDH level.

TREATMENT Rx

Patients with either T-cell lymphoblastic lymphoma or precursor B-cell lymphoblastic lymphoma are typically treated with regimens modeled after those used for acute lymphoblastic leukemia (Chapter 183). These regimens frequently contain cytarabine and high-dose methotrexate, and they often include maintenance therapy. CNS prophylaxis with intrathecal chemotherapy, high-dose methotrexate, or cranial irradiation is often a component of these regimens.

Mature B-Cell Lymphomas

SMALL LYMPHOCYTIC LYMPHOMA AND CHRONIC LYMPHOCYTIC LEUKEMIA

Small lymphocytic lymphoma is defined as a lymph node or other tissue infiltrate that is morphologically and immunophenotypically identical to chronic lymphocytic leukemia (Chapter 184). Patients frequently are symptom free, and the diagnosis is often made when blood counts are performed for other reasons. Patients frequently have lymphadenopathy or splenomegaly. Fatigue is common. Hypogammaglobulinemia can occur and may lead to an increased susceptibility to infection.

A poor prognosis is associated with advanced stage and systemic symptoms, expression of high levels of CD38 and ZAP-70 on tumor cells, lack of rearranged immunoglobulin heavy chain genes, and genetic abnormalities such as del(17p) and del(11q). As many as 10% of patients exhibit transformation to diffuse large B-cell lymphoma (Richter syndrome), which is associated with a poor prognosis.

The median survival time is more than 10 years for patients without adverse characteristics, and these patients can often be managed initially with observation. Therapy is necessary for patients who have systemic symptoms, for those with rapidly progressive or symptomatic lymphadenopathy or splenomegaly, and for those who develop cytopenias.

TREATMENT Rx

Management must be individualized because therapy is unlikely to be curative, and patients are often elderly. A regimen consisting of fludarabine in combination with cyclophosphamide and rituximab (FCR) (Table 185-9) has a high complete remission rate and is frequently used in the United States for relatively young, fit patients. Fludarabine plus rituximab or the combination of chlorambucil and obinituzumab is used in elderly patients. Bendamustine plus rituximab (B-R) is an active combination. It is better tolerated but has a lower response rate than FCR. The immune modulator lenalidomide is also an active agent. The B-cell receptor pathway inhibitors ibrutinib and idelalisib are new highly active oral agents that may transform the management of this lymphoma.[8] Allogeneic hematopoietic stem cell transplantation may be curative, but few patients are candidates for this approach.

Patients may develop autoimmune thrombocytopenia (Chapter 172), autoimmune neutropenia (Chapter 167), and red blood cell aplasia (Chapter 165). These disorders may respond to treatment with corticosteroids, intravenous immune globulin, rituximab, or splenectomy, as used in patients without underlying lymphoma.

EXTRANODAL MARGINAL ZONE LYMPHOMA OF MUCOSA-ASSOCIATED LYMPHOID TISSUE (MALT LYMPHOMA)

MALT lymphomas are indolent tumors that originate in association with epithelial cells and are seen most commonly in the gastrointestinal tract, salivary glands, breast, thyroid, orbit, conjunctiva, skin, and lung. The majority of cases are stage I or II at diagnosis, although in some series as many as 30% disseminate to bone marrow or other sites. These lymphomas tend to remain localized for extended periods. Local treatment with surgery or radiation therapy cures a high proportion of localized neoplasms. Disseminated disease is treated similarly to follicular lymphoma (see later).

TABLE 185-9 COMBINATION CHEMOTHERAPY REGIMENS FOR NON-HODGKIN LYMPHOMA

REGIMEN	DOSE	DAYS OF ADMINISTRATION	FREQUENCY
R-CHOP			**EVERY 21 DAYS**
Cyclophosphamide	750 mg/m^2 IV	1	
Doxorubicin	50 mg/m^2 IV	1	
Vincristine	1.4 mg/m^2 IV*	1	
Prednisone, fixed dose	100 mg PO	1-5	
Rituximab	375 mg/m^2 IV	1	
R-EPOCH§			**EVERY 21 DAYS**
Etoposide	50 mg/m^2/d IV¶	1-4	
Doxorubicin	10 mg/m^2/d IV¶	1-4	
Vincristine	0.4 mg/m^2/d IV¶	1-4	
Cyclophosphamide	750 mg/m^2 IV	5	
Prednisone	60 mg/m^2 bid PO	1-5	
Rituximab	375 mg/m^2 IV	1	
R-CVP			**EVERY 21 DAYS**
Cyclophosphamide	1000 mg/m^2 IV	1	
Vincristine	1.4 mg/m^2 IV*	1	
Prednisone, fixed dose	100 mg PO	1-5	
Rituximab	375 mg/m^2 IV	1	
FCR			**EVERY 28 DAYS**
Fludarabine	25 mg/m^2 IV	1-3	
Cyclophosphamide	250 mg/m^2 IV	1-3	
Rituximab	375 mg/m^2	1	
B-R			**EVERY 28 DAYS**
Bendamustine	90 mg/m^2 IV	1-2	
Rituximab	375 mg/m^2 IV	1	

*Vincristine dose is often capped at 2 mg total.
§Doses are adjusted based on myelosuppression with previous cycle.
¶Continuous infusion.

Gastric MALT lymphomas are frequently associated with infection by *H. pylori*. About 75% of patients achieve remission after eradication of *H. pylori*, and more than 90% remain in remission for prolonged periods of time. Nonresponders are more likely to have submucosal invasion by endoscopic ultrasonography. Approximately 25% progress, of which about 25% develop diffuse large B-cell lymphoma. Response to antibiotic therapy is less likely if invasion is deeper, lymph node metastases are found, or the t(11;18) chromosomal translocation is present.

TREATMENT Rx

Patients may have tumors in more than one extranodal site, and these locations can sometimes be successfully treated with local therapy. Patients without symptoms can be monitored closely without therapy until symptoms progress. Patients with symptoms can be treated with rituximab, single-agent chemotherapy, or combinations. Gastric MALT lymphoma that does not respond to antibiotics can be treated with radiation, rituximab as a single agent (similar to its use in follicular lymphoma), or several traditional combination chemotherapy regimens (see Table 185-9).[9]

FOLLICULAR LYMPHOMA

Follicular lymphoma accounts for the majority of indolent or "low-grade" lymphomas in the United States. Follicular lymphoma is divided into three grades based on the proportion of large, transformed cells (centroblasts).

Patients with follicular lymphoma are frequently asymptomatic. The most common presenting complaint is painless lymphadenopathy. Some patients have cough or dyspnea related to pulmonary or mediastinal involvement or pleural effusions. Other patients have symptoms of abdominal pain or

fullness related to subdiaphragmatic or splenic disease. A minority of patients have systemic symptoms of fevers, night sweats, or weight loss.

The clinical behavior and treatment of follicular lymphoma grades 1 and 2 are the same and are discussed in this section. Some grade 3 follicular lymphomas have a more aggressive clinical course and are treated similarly to diffuse large B-cell lymphoma (see later). This distinction between grade 3a (indolent behavior) and 3b (aggressive behavior) is based on the presence or absence of sheets of centroblasts, although this separation may not be reproducible among pathologists. Because of this, we favor treating all patients with grade 3 follicular lymphoma like diffuse large B-cell lymphoma.

TREATMENT Rx

Localized Disease

Approximately 5 to 15% of patients have localized disease (stage I or minimal stage II disease) at diagnosis. These lymphomas are usually treated with involved-field radiation, and most series report 10-year disease-free survival rates of approximately 50% and overall survival rates of 60 to 70%.[10] Some retrospective series have reported improved outcomes when chemotherapy plus rituximab is combined with radiation.[11]

Advanced Disease

Most patients with follicular lymphoma have extensive disease at diagnosis. The median survival time of these patients has improved after the addition of rituximab and is now 10 to 20 years. Spontaneous regression has been described. As many as 30 to 50% of patients experience transformation to a more aggressive histology—usually diffuse large B-cell lymphoma. Transformation is frequently associated with new systemic symptoms and rapidly progressive lymphadenopathy, an aggressive clinical course, and a poor prognosis.

Asymptomatic patients, especially elderly patients and those with other medical illnesses, are frequently managed with a "watch and wait" approach. Prospective trials have demonstrated that this approach does not influence overall survival, and patients can sometimes be observed for long periods before treatment is required.

Most patients with follicular lymphoma eventually require treatment because of systemic symptoms, symptomatic or progressive lymphadenopathy, splenomegaly, effusions, or cytopenias. In elderly patients, those who are poor candidates for intensive chemotherapy regimens, and those who want to avoid the side effects of chemotherapy, single-agent rituximab (375 mg/m^2 intravenously, given weekly for 4 consecutive weeks) yields an objective response rate of well over 50%. The median duration of response is approximately 1 to 2 years for patients who receive no additional therapy, but the response can be extended with ongoing administration of rituximab once every 2 or 3 months or by repeating the initial four doses every 6 months. When rituximab is combined with standard chemotherapy regimens (see Table 185-9), the response rate, duration of response, and survival are increased compared with the chemotherapy regimen alone. Maintenance rituximab is widely used and extends the duration of remission.

Salvage Therapy

Most patients respond to initial chemotherapy. However, follicular lymphoma recurs in the majority of patients with advanced-stage disease. Patients who relapse usually respond to additional therapy, often with the same agents, although the duration of response becomes progressively shorter with repeated courses of therapy. Some patients respond to the radiolabeled antibodies tositumomab or ibritumomab. Radiation therapy may also be useful for patients with a localized site of symptomatic disease.

Prolonged remissions have been observed after autologous hematopoietic stem cell transplantation. Allogeneic hematopoietic stem cell transplantation may cure some patients with relapsed follicular lymphoma.

MANTLE CELL LYMPHOMA

Mantle cell lymphoma is a B-cell neoplasm composed of small lymphoid cells that may resemble small lymphocytic lymphoma or follicular lymphoma. It is most common in elderly patients and is usually at an advanced stage at the time of diagnosis. Males are more frequently affected, and extranodal disease, especially involvement of the bone marrow, Waldeyer's ring, and the gastrointestinal tract, is common. Mantle cell lymphoma is the most common cause of multiple lymphomatous polyposis, and many oncologists recommend that the gastrointestinal tract be evaluated with endoscopy during the initial evaluation.

Some patients present with involvement of the peripheral blood as well as the bone marrow, a clinical picture that resembles chronic lymphocytic leukemia (Chapter 184). The lymphocytes in both disorders are CD5$^+$, but the t(11;14) and overexpression of cyclin D1 seen in mantle cell lymphoma usually allow an accurate diagnosis.

The median survival of mantle cell lymphoma has significantly improved with new therapies. Occasional patients may have an indolent course without initial therapy—particularly those with a leukemic presentation. Patients have a poor outcome when treated with the same regimens used for diffuse large B-cell lymphoma. Regimens similar to those used for Burkitt lymphoma and lymphoblastic lymphoma are effective. The combination of bendamustine plus rituximab and regimens that incorporate high-dose cytarabine are efficacious. Maintenance rituximab seems to prolong remission duration. Autologous hematopoietic stem cell transplantation for patients in their first remission is widely used.[12] Several new agents, including lenalidomide, bortezomib, and the Bruton tyrosine kinase inhibitor ibrutinib,[13] are very active and may become part of standard therapy. Allogeneic transplantation can be curative but is associated with considerable morbidity and mortality.

DIFFUSE LARGE B-CELL LYMPHOMA

These tumors are the most common type of non-Hodgkin lymphoma, but their morphology and genetic features are heterogeneous. Signs and symptoms are similar to those of other subtypes, although patients are more likely to have B symptoms or symptoms from the local tumor than are patients with follicular lymphoma.

The WHO classification of non-Hodgkin lymphomas has identified several variants and subtypes of diffuse large B-cell lymphoma (Table 185-10). Many of these are histologic or genetic subtypes or variants for which treatment is the same as the standard approach for diffuse large B-cell lymphoma. Other subtypes present unusual clinical syndromes or specific therapeutic problems.

TREATMENT Rx

Localized Disease

As many as 30% of patients with diffuse large B-cell lymphoma have stage I or minimal stage II disease. These patients are occasionally cured with radiation therapy alone, but initial treatment with chemotherapy is more effective. In the United States, most patients are treated with R-CHOP. Some patients can be treated with fewer courses of chemotherapy if radiotherapy is included. For patients with bulky disease, a complete course of chemotherapy is usually used. Some physicians advocate the use of PET scans to shorten the duration of therapy or eliminate the need for radiotherapy in complete responders.

TABLE 185-10 DIFFUSE LARGE B-CELL LYMPHOMA VARIANTS AND SUBTYPES

Diffuse large B-cell lymphoma, not otherwise specified (NOS)
Common morphologic variants
Centroblastic
Immunoblastic
Anaplastic
Immunohistochemical subgroups
CD5$^+$ DLBCL
Germinal center B-cell like (GCB)
Non–germinal center B-cell like (non-GCB)
Diffuse large B-cell lymphoma, subtypes
T-cell/histiocyte-rich large B-cell lymphoma
Primary DLBCL of the CNS
Primary cutaneous DLBCL, leg type
EBV-positive DLBCL of elderly people
Other large B-cell lymphomas
Primary mediastinal (thymic) large B-cell lymphoma
Intravascular large B-cell lymphoma
DLBCL associated with chronic inflammation
Lymphomatoid granulomatosis
ALK-positive large B-cell lymphoma
Plasmablastic lymphoma
Large B-cell lymphoma arising in HHV-8-associated multicentric Castleman disease
Primary effusion lymphoma

ALK = anaplastic lymphoma kinase; CNS = central nervous system; DLBCL = diffuse large B-cell lymphoma; EBV = Epstein-Barr virus; HHV = human herpesvirus.

Advanced Disease

R-CHOP is the most widely used regimen for adults of all ages with advanced-stage diffuse large B-cell lymphoma. Among patients older than 60 years, 75% achieve a complete response, with a 10-year progression-free survival rate of 36.5% and a 10-year overall survival rate of 43.5%. In younger patients, the outcome is better.[] In younger patients, R-EPOCH and the intensive regimen R-ACVBP may yield superior results.[] Autologous transplantation in first remission may benefit very high-risk patients.[]

Salvage Therapy

A variety of chemotherapy regimens have been developed for patients who relapse after attaining a remission with initial chemotherapy. These regimens commonly contain agents such as cisplatin, cytarabine, etoposide, carboplatin, and ifosfamide. Response rates exceeding 50% can be observed with these combinations, although no more than 10% of patients achieve long-term disease-free survival. High-dose therapy followed by autologous hematopoietic stem cell transplantation has become accepted therapy for patients with relapsed diffuse large B-cell lymphoma. Approximately 20 to 50% of these patients attain long-term disease-free survival, depending on their response to conventional salvage chemotherapy.[] Several new drugs such as lenalidomide and ibrutinib seem to particularly benefit patients with ABC subtype.

Subtypes of Diffuse Large B-Cell Lymphoma

Primary mediastinal large B-cell lymphoma originates in the thymus and is most common in young women. This subtype shares genetic features with classic Hodgkin lymphoma. This entity is distinguished by the presence of a mediastinal mass, which usually causes symptoms of cough, chest pain, or superior vena cava syndrome. A very large mass (>10 cm) or the existence of a malignant pleural effusion is associated with a worse prognosis. Mediastinal large B-cell lymphoma is treated with the same chemotherapy regimens used for diffuse large B-cell lymphoma, followed, in some cases, by consolidative radiation therapy.[14] The prognosis is similar to that of other diffuse large B-cell lymphomas. Relapses often occur in extranodal sites such as the CNS, lungs, gastrointestinal tract, liver, ovaries, and kidneys.

Intravascular large B-cell lymphoma is an aggressive lymphoma caused by cells that infiltrate the lumens of small blood vessels. Widespread extranodal involvement is common. Focal neurologic deficits and mental status changes are frequent. Cases are often diagnosed at autopsy, although durable responses to combination chemotherapy have been described.

Primary effusion lymphoma is associated with HHV-8 and is seen in HIV-infected and other immunosuppressed patients. Effusions occur in serous body cavities. Peripheral lymphadenopathy is not seen. Prognosis is poor despite chemotherapy.

Plasmablastic lymphoma is most often seen in patients with HIV infection and frequently presents with involvement of the head and neck. This tumor does not express CD20 and thus does not benefit from treatment with rituximab.

Primary cutaneous diffuse large B-cell lymphoma of the leg (leg type) is one of two presentations of B-cell lymphoma in the skin. This tumor occurs mostly in older patients and follows an aggressive course. This neoplasm must be distinguished from primary cutaneous follicle center lymphoma, which might also be diagnosed as a cutaneous diffuse large B-cell lymphoma but follows an indolent course and requires only local therapy.

Double-hit and gray-zone lymphomas are newly described entities. Double-hit lymphomas have rearrangements of *Myc* in combination with *Bcl-2* and/or *Bcl-6*. They have a poor outcome when treated with standard chemotherapy regimens. Gray-zone lymphomas include those with features intermediate between diffuse large B-cell lymphoma and Burkitt lymphoma and those with features intermediate between mediastinal diffuse large B-cell lymphoma and classic Hodgkin lymphoma.[15]

BURKITT LYMPHOMA

Burkitt lymphoma is a highly aggressive B-cell lymphoma that is more common in children and immunosuppressed individuals than in healthy adults (see later). Widespread extranodal involvement is common. The endemic form of Burkitt lymphoma is seen most frequently in children who reside in equatorial Africa. Involvement of bones of the jaw is common in this form. The sporadic form of Burkitt lymphoma is seen most commonly in children in the United States. Males are more frequently affected. Both children and adults frequently have bulky abdominal disease, sometimes with involvement of the kidneys, ovaries, and breasts. Bone marrow involvement is seen in about one third of cases.

TREATMENT Rx

Tumors may progress extremely rapidly, so therapy should be started as soon as possible. Tumor lysis syndrome may occur because of the frequent presence of bulky disease, the high rate of tumor proliferation, and the extreme sensitivity of the tumor to chemotherapy. Patients are usually treated with specialized high-intensity regimens, including rituximab, of relatively short duration.[16] Treatment with the R-CHOP regimen used for diffuse large B-cell lymphoma has a poor outcome. CNS prophylaxis with intrathecal chemotherapy or high-dose methotrexate is required. Cure rates well in excess of 50% are typical with appropriate therapy. Excellent results have been described with R-EPOCH.

RARE TYPES OF B-CELL LYMPHOMA

Several rare types of lymphoma have distinct clinical features.

Lymphoplasmacytic lymphoma is an indolent lymphoma that frequently involves the bone marrow, peripheral blood, and spleen. Patients frequently have an immunoglobulin M (IgM) paraprotein (and therefore the disease could be called Waldenström macroglobulinemia) that may lead to symptoms of hyperviscosity, autoimmune phenomena, or neuropathies (Chapter 187). Plasmapheresis can reduce symptoms of hyperviscosity. Chemotherapy with alkylating agents, combination chemotherapy, or fludarabine all in combination with rituximab may be used.[] Newer agents such as bortezomib and ibrutinib are active.

Splenic marginal zone lymphoma is an indolent lymphoma that usually manifests with splenomegaly and lymphocytosis. A monoclonal gammopathy is frequently seen. Peripheral lymphadenopathy is unusual. Anemia and thrombocytopenia may respond to splenectomy. Chemotherapy with single agents or anthracycline-based combinations may be useful, and when the lymphoma is associated with hepatitis C, antiviral therapy may be effective. Responses to interferon have been described. This lymphoma appears to be particularly responsive to rituximab.

Nodal marginal zone B-cell lymphoma is an indolent disorder that is usually associated with generalized lymphadenopathy. The clinical course and prognosis are similar to those of follicular lymphoma, and it is usually treated in a similar manner.

Small intestinal immunoproliferative disease, a disorder seen most frequently in the Middle East, begins as a polyclonal process and can progress to a large B-cell lymphoma. The process is often associated with *C. jejuni* infection. Early in the disease, patients may respond to antibiotics, and frank lymphoma may respond to combination chemotherapy regimens.

Mature T-Cell Lymphomas (Peripheral T-Cell Lymphomas)

Peripheral (or mature) T-cell lymphomas are neoplasms of post-thymic T cells. These include relatively indolent disorders such as mycosis fungoides and $CD30^+$ cutaneous lymphoproliferative disorders, but most patients diagnosed with peripheral T-cell lymphoma have an aggressive neoplasm. Peripheral T-cell lymphomas represent only 10% of the non-Hodgkin lymphomas occurring in the United States. Unfortunately, the treatment for these lymphomas has not progressed as rapidly as the treatments for B-cell lymphomas.[17]

MYCOSIS FUNGOIDES

Mycosis fungoides (often referred to as *cutaneous T-cell lymphoma*) is an indolent malignancy that is most common in middle-aged and older adults.[18] The clinical course is usually a slow progression from isolated patches or plaques to thickened, more widespread plaques and then to multiple cutaneous tumors that may ulcerate (Chapter 440). A subset of patients presents with generalized erythroderma and circulating tumor cells, called *Sézary syndrome*. Lymph node and visceral involvement may occur late in the course of the disease.

TREATMENT Rx

Cutaneous radiation therapy may be curative for patients with limited patch or plaque disease. Patients with early-stage disease (<10% body surface area) are frequently treated with skin-directed therapy that may include ultraviolet radiation, topical steroids, or topical nitrogen mustard.

Patients with more advanced disease frequently benefit from total skin electron beam therapy or extracorporeal photopheresis. Medical treatments

include interferon-α, retinoids, monoclonal antibodies, histone deacetylase inhibitors (vorinostat, depsipeptide), the fusion toxin denileukin diftitox, and traditional cytotoxic chemotherapeutic agents. However, these treatments are usually only palliative. Results with autologous hematopoietic stem cell transplantation are usually poor, although allogeneic stem cell transplantation has yielded promising results in some cases.

ADULT T-CELL LYMPHOMA/LEUKEMIA

Adult T-cell lymphoma/leukemia, which is associated with HTLV-1 infection (Chapter 378), is most commonly seen in southern Japan and the Caribbean. Most infected patients are asymptomatic, and the lifetime risk for developing adult T-cell lymphoma/leukemia is approximately 3%.

Patients may have acute leukemia, aggressive lymphoma, or an indolent lymphoproliferative disease. Patients with aggressive disease present with generalized lymphadenopathy, hepatosplenomegaly, cutaneous infiltration, and hypercalcemia. Many patients have characteristic circulating tumor cells with a "flower" or "cloverleaf" nucleus.

TREATMENT Rx

Patients with indolent disease can sometimes be monitored without therapy. Aggressive disease is usually treated with combination chemotherapy, but there is no consensus on the best regimen. Historically, the 5-year survival rate has been less than 10%, although recent trials have reported better outcomes.

CD30⁺ CUTANEOUS LYMPHOPROLIFERATIVE DISORDERS

These disorders represent a spectrum of diseases that may have an identical histologic appearance and overlapping clinical manifestations. Treatment decisions are often based on the clinical behavior of the lesions. These lymphomas express CD30 but do not express the anaplastic lymphoma kinase (ALK) protein (see later).

Lymphomatoid papulosis is a "histologically malignant" clonal disorder consisting of erythematous or skin-colored papules that frequently undergo spontaneous ulceration and necrosis over a period of weeks. The prognosis is excellent, although patients may eventually develop lymphoma.

Primary cutaneous anaplastic large cell lymphoma occurs most commonly in older men and also undergoes frequent spontaneous regression. The 5-year survival rate is greater than 90%. Treatment usually consists of local measures (surgery or radiation), although chemotherapy may be required.

PRIMARY SYSTEMIC ANAPLASTIC LARGE CELL LYMPHOMA

Anaplastic large cell lymphoma (ALCL) is an aggressive, CD30⁺, T-cell non-Hodgkin lymphoma that is seen most frequently in young males. B-cell lymphomas with similar morphology can occur, but they have clinical features identical to other diffuse large B-cell lymphomas and are not considered part of this disease. A morphologically similar but biologically unrelated and clinically distinct neoplasm, primary cutaneous ALCL, occurs predominantly in older adults and represents part of the spectrum of cutaneous CD30⁺ lymphoproliferative disorders (see earlier). Primary systemic ALCL frequently has a t(2;5) chromosomal translocation that leads to overexpression of ALK, a protein not normally detectable in lymphoid cells.

Patients usually have lymphadenopathy, and involvement of the skin, bone, and gastrointestinal tract may be observed.

A newly described entity of anaplastic large cell lymphoma of the breast is associated with breast implants. When localized, these patients may be cured with surgery including removal of the implant.

TREATMENT Rx

Patients are usually treated with chemotherapy regimens such as CHOP with or without the addition of etoposide.[19] Patients whose tumors express ALK have an excellent outcome, and 5-year survival rates of 70 to 90% have been observed. ALK-negative ALCL is more common in older patients and is associated with an inferior response rate and shorter survival time. Autologous hematopoietic stem cell transplantation may be curative for patients who relapse. The anti-CD30 antibody-drug conjugate brentuximab vedotin is highly active for patients with relapsed disease and may be incorporated into primary therapy. Patients whose lymphomas express ALK usually respond to the ALK inhibitor crizotinib.

PERIPHERAL T-CELL LYMPHOMAS, UNSPECIFIED

The largest group of patients with peripheral T-cell lymphomas is defined in the WHO classification as having "peripheral T-cell lymphoma, unspecified." These patients' signs and symptoms are similar to those of patients with aggressive B-cell lymphomas, although systemic symptoms (fevers, night sweats, and weight loss) and extranodal involvement are frequent. The diagnosis of peripheral T-cell lymphoma requires immunophenotyping to demonstrate the T-cell origin.

TREATMENT Rx

Patients are generally treated with the same regimens used for diffuse large B-cell lymphomas (e.g., CHOP with or without etoposide), often in combination with upfront autologous hematopoietic stem cell transplantation, although the outcome is substantially worse.[20] Patients who relapse after complete remission can sometimes be cured with hematopoietic stem cell transplantation.

UNUSUAL SUBTYPES OF T-CELL LYMPHOMA

Angioimmunoblastic T-cell lymphoma is associated with generalized lymphadenopathy, fever, weight loss, skin rash, and polyclonal hypergammaglobulinemia. Results of therapy are similar to those for peripheral T-cell lymphoma, unspecified.

Extranodal NK/T-cell lymphoma usually occurs in extranodal sites, especially the nose, palate, and nasopharynx. Involvement of the nose and face leads to the syndrome that was previously called lethal midline granuloma. This disorder is unusual in the United States, but it is frequent in Southeast Asia and Latin America. The prognosis is extremely poor, although patients with localized disease can sometimes be cured with aggressive combination radiation therapy and chemotherapy.

Hepatosplenic T-cell lymphoma is characterized by sinusoidal infiltration of the spleen, liver, and bone marrow, which leads to hepatosplenomegaly, systemic symptoms, and cytopenias. Lymphadenopathy is unusual. Patients are typically young males, and this disease can occur in allograft recipients and in the setting of immune dysfunction. The prognosis is poor, and remissions are rarely observed with chemotherapy.

Enteropathy-type T-cell lymphoma is usually seen in patients with gluten-sensitive enteropathy (Chapter 140). Patients typically present with abdominal pain and diarrhea and sometimes with bowel perforation. Treatment of celiac disease with a gluten-free diet may reduce the risk for lymphoma. The prognosis in these often undernourished patients is poor.

Subcutaneous panniculitis-like T-cell lymphoma manifests with multiple subcutaneous nodules and is often misdiagnosed as panniculitis. Patients with disseminated disease can have a syndrome consisting of fevers, weight loss, hepatosplenomegaly, pancytopenia, and phagocytosis of blood cells (hemophagocytic syndrome). Patients sometimes respond to combination chemotherapy regimens used for diffuse large B-cell lymphoma, interferon, and radiation therapy, but long-term disease-free survival is unusual.

SPECIAL CLINICAL SITUATIONS

The diagnosis and management of patients with the various types of non-Hodgkin lymphoma can be profoundly influenced by the site of origin of the lymphoma and by certain clinical characteristics of the patients. Examples of the latter include pregnant patients with lymphoma, elderly patients with lymphoma, and lymphoma in patients who are severely immunosuppressed.

Specific Primary Sites of Diffuse Large B-Cell Lymphoma

Approximately 30% of diffuse large B-cell lymphomas originate in extranodal sites. Presentation in certain extranodal sites is associated with unique clinical behaviors that may necessitate diagnostic studies or additional therapy beyond that used for patients with nodal presentations.

Patients with primary CNS lymphoma (Chapter 189) commonly have ocular involvement, and all patients with this diagnosis should have a slit lamp

examination. Surgical resection of primary CNS lymphoma is usually not performed, and the primary role of surgery is for diagnosis. Primary lymphomas of the CNS are very sensitive to corticosteroids, but the best results have been observed with chemotherapy regimens that use high-dose methotrexate alone or in combination with other agents such as cytarabine and temozolomide. By comparison, conventional chemotherapy regimens, such as CHOP, are of little benefit. Whole brain irradiation is also effective therapy, although the incidence of leukoencephalopathy is extremely high, especially in elderly patients. Radiation therapy is frequently reserved for relapse rather than being used as adjunctive treatment with primary chemotherapy.

Treatment of primary testicular lymphoma, the most common testicular cancer in men older than 60 years (Chapter 200), usually consists of orchiectomy followed by combination chemotherapy. Relapse in the contralateral testicle is common, and most oncologists recommend adjuvant radiation to the scrotum. CNS involvement is common, and prophylactic intrathecal chemotherapy is usually recommended. Late relapses occur frequently.

Diffuse large B-cell lymphoma of the stomach and gastrointestinal tract is treated differently from gastric MALT lymphoma, even if there is a history of prior MALT lymphoma. Patients can be cured with surgery and adjunctive radiation or chemotherapy, although surgery is rarely performed for gastric lymphomas because of the morbidity associated with gastric resection. Patients should be treated with chemotherapy regimens used for other diffuse large B-cell lymphomas. Radiation therapy is sometimes used after chemotherapy, although the role of combined-modality treatment is not defined.

Lymphoma in AIDS and Post-transplantation Lymphoproliferative Disorders

Non-Hodgkin lymphoma is an AIDS-defining illness in HIV-infected individuals (Chapter 393), and the risk for developing a non-Hodgkin lymphoma is increased more than 150-fold after the diagnosis of another AIDS-defining illness. Most cases are diffuse large B-cell lymphomas or Burkitt lymphomas. AIDS-associated lymphomas behave aggressively and frequently involve the CNS and other unusual sites, such as the gastrointestinal tract, anus, rectum, skin, and soft tissue (Chapter 393). Factors associated with poor survival include low CD4 counts, poor performance status, older age, and advanced stage. The incidence of AIDS-associated non-Hodgkin lymphoma has declined, and the prognosis has improved considerably. Patients are generally treated in the same manner as non-immunocompromised patients, and the prognoses are similar. Intrathecal prophylaxis is generally recommended because of a higher risk for CNS involvement.

The risk for developing a non-Hodgkin lymphoma is also markedly increased in patients who have received a solid organ transplant. The histologic appearance of these lymphomas is variable, but they frequently resemble aggressive lymphomas in non-immunocompromised patients. Similar disorders can be seen in patients who are treated with methotrexate and other drugs for autoimmune disorders and in recipients of allogeneic hematopoietic stem cell transplants, especially if the transplants are T-cell depleted. These post-transplantation lymphoproliferative disorders, which may develop within weeks after surgery, are more common in patients who receive aggressive immunosuppression after transplantation. Involvement of extranodal sites is common, and lymphoma frequently involves the transplanted organ. Post-transplantation lymphoproliferative disorders may respond to reduction or withdrawal of immunosuppression. Some investigators have advocated the use of acyclovir or ganciclovir because these lymphomas are usually related to EBV, but this practice is controversial. High response rates are also seen with rituximab. Other patients require treatment with combination chemotherapy regimens.

Non-Hodgkin Lymphoma in Elderly Patients

More than 50% of patients who develop non-Hodgkin lymphomas are older than 60 years, and the prognosis is generally worse for elderly patients. These poorer outcomes are related to increased toxicity of drug therapy, lower remission rates, increased rates of relapse, and higher death rates from cardiovascular disease and causes other than the lymphoma itself. Older patients are more likely to have other adverse prognostic characteristics (see Table 185-8), which also contribute to poorer outcomes. The practice of arbitrary dose reductions based solely on age should be discouraged if patients have a good performance status and no comorbid illnesses.

Non-Hodgkin Lymphoma and Pregnancy

Non-Hodgkin lymphoma in pregnancy involves major clinical and ethical issues, and a multidisciplinary approach is needed (Chapter 239). Although chest radiographs are generally considered safe, ultrasound examination is usually used instead of CT for staging in the abdomen and pelvis.

Treatment can occasionally be delayed until after delivery; however, most women have a tumor that is potentially curable, and treatment delays may decrease the chance for cure. Other patients have conditions such as superior vena cava syndrome that require immediate treatment. After the first trimester, full-dose standard therapy such as R-CHOP-R may be used. Several studies indicate high probabilities of cure without significantly increased adverse long-term physical or intellectual deficits for the child. Although it is reasonable to offer therapeutic abortion for women in the first trimester, chemotherapy may also be successful in this situation. Regimens using methotrexate or abdominal radiation must be avoided.

DISEASES SOMETIMES CONFUSED WITH LYMPHOMA

A variety of conditions are associated with lymphadenopathy that can be confused with lymphoma.[21] The most common atypical lymphoid proliferations that can be confused with lymphoma are florid reactions to immune stimulation. Follicular hyperplasia with diffuse proliferation of B cells and T cells can be seen in a variety of autoimmune diseases (e.g., Sjögren syndrome, systemic lupus erythematosus, rheumatoid arthritis) and infectious processes (e.g., EBV, cytomegalovirus, cat-scratch disease) (Chapter 168). If the definitive diagnosis of lymphoma cannot be made even after immunologic and molecular studies, the patient should be closely observed. The clinical course or subsequent biopsies can usually resolve the confusion.

Castleman disease, or angiofollicular lymph node hyperplasia, usually appears with a hyaline vascular pattern of lymphoid proliferation, but a subset of patients has hyperplastic lymphoid follicles and sheets of plasma cells. Patients with Castleman disease often present with a localized lymphoid mass, but some patients have a systemic illness with fevers, night sweats, weight loss, and fatigue. Frequently, the systemic symptoms of Castleman disease are related to excessive production of IL-6. Castleman disease in HIV-infected patients is frequently associated with HHV-8. Patients with disseminated and plasma cell–rich forms of Castleman disease may occasionally progress to lymphoma. Patients with localized Castleman disease can be treated with surgical removal or radiation therapy. Patients with systemic disease may respond to treatment with high-dose corticosteroids. Patients with overexpression of IL-6 frequently benefit from treatment with an anti-IL-6 antibody. If other treatments fail, patients sometimes benefit from combination chemotherapy regimens, autologous or allogeneic hematopoietic stem cell transplantation, or both.

Sinus histiocytosis with massive lymphadenopathy, also known as *Rosai-Dorfman disease,* manifests as bulky lymphadenopathy in children and young adults. Extranodal sites such as the skin, upper airways, gastrointestinal tract, and CNS can be involved. The disease is usually self-limited, but it has been associated with autoimmune hemolytic anemia.

Kikuchi's disease (histiocytic necrotizing lymphadenitis) is a disease of unknown origin that most commonly affects young women. Symptoms most commonly consist of painless cervical lymphadenopathy that is often accompanied by fever, flu-like symptoms, and rash. Treatment is symptomatic, and manifestations usually resolve within weeks or months.

IgG4-related disease is a systemic immune-mediated fibrosing inflammatory disease in which various organs are infiltrated by IgG4-positive plasma cells. Patients often have accompanying asymptomatic lymphadenopathy that usually responds to corticosteroids.

Grade A References

A1. Hallek M, Fischer K, Fingerle-Rowson G, et al. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomized, open-label, phase 3 trial. *Lancet.* 2010;376:1164-1174.

A2. Goede V, Fischer K, Busch R, et al. Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions. *N Engl J Med.* 2014;370:1101-1110.

A3. Eichhorst B, Fink A-M, Busch R, et al. Chemoimmunotherapy with fludarabine (F), cyclophosphamide (C), and rituximab (R) (FCR) versus bendamustine and rituximab (BR) in previously untreated and physically fit patients (pts) with advanced chronic lymphocytic leukemia (CLL): results of a planned interim analysis of the CLL10 trial, an international, randomized study of the German CLL Study Group (GCLLSG). *Blood.* 2013;122(abstr. 526).

A4. Zucca E, Conconi A, Laszlo D, et al. Addition of rituximab to chlorambucil produces superior event-free survival in the treatment of patients with extranodal marginal-zone B-cell lymphoma: 5-year analysis of the IELSG-19 randomized study. *J Clin Oncol.* 2013;31:565-572.

A5. Ardeshna KM, Qian W, Smith P, et al. Rituximab versus a watch-and-wait approach in patients with advanced-stage, asymptomatic, non-bulky follicular lymphoma: an open-label randomised phase 3 trial. *Lancet Oncol.* 2014;15:424-435.

A6. Bachy E, Houot R, Morschhauser F, et al. Long-term follow up of the FL2000 study comparing CHVP-interferon to CHVP-interferon plus rituximab in follicular lymphoma. *Haematologica*. 2013;98:1107-1114.
A7. Rummel MJ, Niederle N, Maschmeyer G, et al. Bendamustine plus rituximab versus CHOP plus rituximab as first-line treatment for patients with indolent and mantle-cell lymphomas: an open-label, multicenter, randomized, phase 3 non-inferiority trial. *Lancet*. 2013;381:1203-1210.
A8. Salles G, Seymour JF, Offner F, et al. Rituximab maintenance for 2 years in patients with high tumour burden follicular lymphoma responding to rituximab plus chemotherapy (PRIMA): a phase 3, randomised controlled trial. *Lancet*. 2011;377:42-51.
A9. Kluin-Nelemans HC, Hoster E, Hermine O, et al. Treatment of older patients with mantle-cell lymphoma. *N Engl J Med*. 2012;367:520-531.
A10. Coiffier B, Thieblemont C, Van Den Neste E, et al. Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. *Blood*. 2010;116:2040-2045.
A11. Récher C, Coiffier B, Haioun C, et al. Intensified chemotherapy with ACVBP plus rituximab versus standard CHOP plus rituximab for the treatment of diffuse large B-cell lymphoma (LNH03-2B): an open-label randomised phase 3 trial. *Lancet*. 2011;378:1858-1867.
A12. Stiff PJ, Unger JM, Cook JR, et al. Autologous transplantation as consolidation for aggressive non-Hodgkin's lymphoma. *N Engl J Med*. 2013;369:1681-1690.
A13. Gisselbrecht C, Glass B, Mounier M, et al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *J Clin Oncol*. 2010;28:4184-4190.
A14. Leblond V, Johnson S, Chevret S, et al. Results of a randomized trial of chlorambucil versus fludarabine for patients with untreated Waldenström macroglobulinemia, marginal zone lymphoma, or lymphoplasmacytic lymphoma. *J Clin Oncol*. 2012;31:301-307.
A15. Whittaker S, Ortiz P, Dummer R, et al. Efficacy and safety of bexarotene combined with psoralen-ultraviolet A (PUVA) compared with PUVA treatment alone in stage IB-IIA mycosis fungoides: final results from the EORTC Cutaneous Lymphoma Task Force phase III randomized clinical trial (NCT00056056). *Br J Dermatol*. 2012;167:678-687.
A16. Tsukasaki K, Utsunomiya A, Fukuda H, et al. VCAP-AMP-VECP compared with biweekly CHOP for adult T-cell leukemia-lymphoma: Japan Clinical Oncology Group study JCOG9801. *J Clin Oncol*. 2007;25:5458-5464.
A17. Thiel E, Korfel A, Martus P, et al. High-dose methotrexate with or without whole brain radiotherapy for primary CNS lymphoma (G-PCNSL-SG-1): a phase 3, randomized, non-inferiority trial. *Lancet Oncol*. 2010;11:1036-1047.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Howlader N, Noone AM, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2011, National Cancer Institute. Bethesda, MD. http://seer.cancer.gov/csr/1975_2011/; based on November 2013 SEER data submission, posted to the SEER web site, April 2014. Accessed January 15, 2015.
2. Leechawengwongs E, Shearer WT. Lymphoma complicating primary immunodeficiency syndromes. *Curr Opin Hematol*. 2012;19:305-312.
3. Bassig BA, Lan Q, Rothman N, et al. Current understanding of lifestyle and environmental factors and risk of non-Hodgkin lymphoma: an epidemiological update. *J Cancer Epidemiol*. 2012;2012:978930.
4. Vockerodt M, Yap LF, Shannon-Lowe C, et al. The Epstein-Barr virus and the pathogenesis of lymphoma. *J Pathol*. 2015;235:312-322.
5. Lenz G, Staudt LM. Aggressive lymphomas. *N Engl J Med*. 2010;362:1417-1429.
6. Nogai H, Dörken B, Lenz G. Pathogenesis of non-Hodgkin's lymphoma. *J Clin Oncol*. 2011;29:1803-1811.
7. Swerdlow SH. Lymphoma classification and the tools of our trade: an introduction to the 2012 USCAP Long Course. *Mod Pathol*. 2013;26:S1-S14.
8. Woyach JA, Johnson AJ, Byrd JC. The B-cell receptor signaling pathway as a therapeutic target in CLL. *Blood*. 2012;120:1175-1184.
9. Olszewski AJ, Castillo JJ. Comparative outcomes of oncologic therapy in gastric extranodal marginal zone (MALT) lymphoma: analysis of the SEER-Medicare database. *Ann Oncol*. 2013;24:1352-1359.
10. Kuruvilla J, Assouline S, Hodgson D, et al. A Canadian evidence-based guideline for the first-line treatment of follicular lymphoma: joint consensus of the Lymphoma Canada Scientific Advisory Board. *Clin Lymphoma Myeloma Leuk*. 2015;15:59-74.
11. Friedberg JW, Byrtek M, Link BK, et al. Effectiveness of first-line management strategies for stage I follicular lymphoma: analysis of the National LymphoCare study. *J Clin Oncol*. 2012;30:3368-3375.
12. Bhatt VR, Vose JM. Hematopoietic Stem cell transplantation for non-Hodgkin lymphoma. *Hematol Oncol Clin North Am*. 2014;28:1073-1095.
13. Wang ML, Rule S, Martin P, et al. Targeting BTK with ibrutinib in relapsed or refractory mantle-cell lymphoma. *N Engl J Med*. 2013;369:507-516.
14. Dunleavy K, Pittaluga S, Maeda LS, et al. Dose-adjusted EPOCH-rituximab therapy in primary mediastinal B-cell lymphoma. *N Engl J Med*. 2013;368:1408-1416.
15. Aukema SM, Siebert R, Schuuring E, et al. Double-hit B-cell lymphomas. *Blood*. 2011;117:2319-2331.
16. Linch DC. Burkitt lymphoma in adults. *Br J Haematol*. 2012;156:693-703.
17. Coiffier B, Federico M, Caballero D, et al. Therapeutic options in relapsed or refractory peripheral T-cell lymphoma. *Cancer Treat Rev*. 2014;40:1080-1088.
18. Wilcox RA. Cutaneous T-cell lymphoma: 2014 update on diagnosis, risk-stratification, and management. *Am J Hematol*. 2014;89:837-851.
19. Sibon D, Fournier F, Brière J, et al. Long-term outcome of adults with systemic anaplastic large-cell lymphoma treated within the Groupe d'Étude des Lymphomes de L'Adulte Trials. *J Clin Oncol*. 2012;30:3939-3946.
20. Dearden CE, Johnson R, Pettengell R, et al. Guidelines for the management of mature T-cell and NK-cell neoplasms (excluding cutaneous T-cell lymphoma). *Br J Haematol*. 2011;153:451-485.
21. O'Malley DP, Grimm KE. Reactive lymphadenopathies that mimic lymphoma: entities of unknown etiology. *Semin Diagn Pathol*. 2013;30:137-145.

REVIEW QUESTIONS

1. A 66-year old man develops cough and chest pain. A chest radiograph shows a mediastinal mass. A biopsy of the mass shows diffuse large B-cell lymphoma. Additional studies reveal enlarged lymph nodes in the para-aortic and mesenteric regions. His bone marrow biopsy is negative. His performance status is otherwise normal, and his serum lactate dehydrogenase level is normal. How should this man be treated?

A. Observation (watch and wait)
B. CHOP (cyclophosphamide, doxorubicin, vincristine, prednisone)
C. CHOP + rituximab
D. Radiation to the chest and abdomen
E. CHOP + rituximab followed by autologous stem cell transplantation

Answer: C This man has a stage III diffuse large B-cell lymphoma. His IPI score is II (low-intermediate risk). A watch and wait approach is reasonable for some patients with indolent lymphoma, but not for patients with diffuse large B-cell lymphoma. There is no role for radiation, alone, in this situation. Treatment with CHOP would yield a high response rate, although randomized trials show improved survival with the addition of the anti-CD20 antibody rituximab (Coiffier B, Thieblemont C, Van Den Neste E, et al. Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. *Blood.* 2010;116:2040-2045). There is no evidence that upfront autologous stem cell transplantation is beneficial for patients in low-risk groups.

2. A 27-year old woman has an unexplained 20-lb weight loss, drenching night sweats, and bulky bilateral cervical lymphadenopathy. A chest radiograph shows bilateral cervical lymphadenopathy. Fine-needle aspirate (FNA) of a cervical lymph node reveals cells consistent with a B-cell non-Hodgkin lymphoma. What should you do next?

A. Performa FNA on another cervical lymph node.
B. Prescribe antibiotics for 10 days and reevaluate.
C. Start ABVD (doxorubicin, bleomycin, vinblastine, dacarbazine).
D. Perform an excision biopsy of a cervical lymph node.
E. Start CHOP (cyclophosphamide, doxorubicin, vincristine, prednisone) + rituximab.

Answer: D This woman has an obvious systemic illness. Empirical antibiotic therapy would be inappropriate. All non-Hodgkin lymphomas are treatable, and many are curable. It is essential to have an accurate diagnosis before starting therapy so that appropriate therapy is administered. Core needle biopsies can occasionally be used for a primary diagnosis if the specimen is handled properly. Fine-needle aspirates should not be used to diagnose lymphoma, and they preclude an accurate diagnosis of the specific subtype of non-Hodgkin lymphoma. In most cases, an excisional biopsy is necessary (and is always preferred) for the initial diagnosis. Her diagnosis from FNA is inadequate. An excisional biopsy with review by an expert hematopathologist is necessary for appropriate treatment (Swerdlow SH. Lymphoma classification and the tools of our trade: an introduction to the 2012 USCAP Long Course. *Mod Pathol.* 2013;26:S1-S14). It would be inappropriate to start CHOP + rituximab before an accurate diagnosis is established. The ABVD regimen is used for Hodgkin lymphoma.

3. Patients with mantle cell lymphoma are most likely to have the following cytogenetic abnormality in their lymph node biopsy?

A. t(11;14)
B. t(11;18)
C. t(14;18)
D. t(8;14)
E. t(2;5)

Answer: A Non-Hodgkin lymphomas are often associated with cytogenetic abnormalities. These abnormalities often consist of translocations that result in overproduction of fusion proteins with result in cell proliferation or inhibit apoptosis (Nogai H, Dörken B, Lenz G. Pathogenesis of Non-Hodgkin lymphoma. *J Clin Oncol.* 2011;29:1803-1811). Certain subtypes of non-Hodgkin lymphoma are associated with specific cytogenetic abnormalities. The t(11;18) is associated with MALT lymphoma, the t(14;18) is associated with follicular lymphoma, the t(8;14) is associated with Burkitt's lymphoma, and the t(2;5) is associated with anaplastic large cell lymphoma. The t(11;14) is characteristic of mantle cell lymphoma. This translocation results in overproduction of cyclin D1 and resultant cell proliferation.

4. A 40-year old woman has a history of a mediastinal large B-cell lymphoma. She was treated several years ago with R-CHOP followed by radiation therapy involving the chest and neck. She comes to you with complaints of fatigue, trouble concentrating at work, and feeling cold. Her examination reveals no lymphadenopathy. Her chest radiograph is unchanged from 1 year ago. Laboratory tests are normal except for high cholesterol. What should you do?

A. Order a computed tomography (CT) scan of the chest, abdomen, and pelvis.
B. Check serum vitamin B_{12} and folic acid levels.
C. Perform a bone marrow aspirate and biopsy.
D. Check a thyroid-stimulating hormone (TSH) level.
E. Order a positron emission tomography (PET) scan.

Answer: D It is important to monitor patients for late effects of therapy. Although disease recurrence should always be considered, there is no evidence of this. It would be inappropriate to order additional imaging at this time. There is no evidence that she is B_{12} or folic acid deficient. Hypothyroidism is a frequent complication of radiation therapy and should always be considered. Her history increases her risk for cardiovascular disease, and strict control of modifiable risk-factor, such as cholesterol, is necessary (http://www.nccn.org/professionals/physician_gls/pdf/survivorship.pdf).

5. Which patient is least likely to have a lymphoma associated with the Epstein-Barr virus (EBV)?

A. A 42-year old man with a history of a liver transplantation
B. A 31-year old woman with chronic HIV infection
C. A 45-year old woman with an extranodal NK/T-cell lymphoma
D. A 68-year old woman with follicular lymphoma
E. A child from sub-Saharan Africa with Burkitt's lymphoma

Answer: D EBV, as well as other infectious agents, is associated with several types of non-Hodgkin lymphoma *Semin Diagn Pathol.* 2011;28:178-187). Follicular lymphoma is not usually associated with EBV. (De Falco G, Rogena EA, Leoncini L. Infectious agents and lymphoma.

186

HODGKIN LYMPHOMA

JOSEPH M. CONNORS

DEFINITION

Hodgkin lymphoma, formerly called Hodgkin disease, is one of the B-cell lymphomas. It consists of two major types: classic Hodgkin lymphoma, with a characteristic neoplastic cell, the Hodgkin-Reed-Sternberg cell; and the much less common (~5% of cases) nodular lymphocyte-predominant Hodgkin lymphoma, with a characteristic lymphocyte-predominant cell. Both types have a distinct natural history, and most important, an excellent response to treatment, with the large majority of patients being cured. Its management, which requires careful multidisciplinary cooperation, serves as a paradigm for the successful application of modern oncologic concepts. Highly effective multiagent chemotherapy is the cornerstone of treatment. Carefully selected patients may require the addition of radiation or, if the lymphoma recurs after primary treatment, high-dose chemoradiation therapy and autologous hematologic stem cell transplantation (HDC/HSCT). The major challenge to clinicians managing this neoplasm is to cure the disease while minimizing long-term toxicity.

EPIDEMIOLOGY

The incidence of Hodgkin lymphoma varies substantially around the world. The highest rates occur in the United States, Canada, Switzerland, and northern Europe. Intermediate rates are seen in southern and eastern Europe and low rates in eastern Asia. No clear explanation for this variation in incidence has been found. Postulated reasons include differences in the age at onset or genotype of any associated Epstein-Barr virus (EBV) infection; crowding during childhood as a result of lower socioeconomic status, predisposing to passage of an as yet undiscovered infectious vector; and intrinsic genetic differences in susceptibility.

Approximately 20,000 new cases are seen annually in North America and Europe. The age-adjusted incidence of Hodgkin lymphoma declined modestly over the 20 years before 1990 at a rate of approximately 0.9% per year but has leveled off since then and is now approximately 2.7 per 100,000. Age-adjusted annual mortality is 0.5 per 100,000. Hodgkin lymphoma occurs slightly more often in men and is seen more frequently in whites than African Americans and much less frequently in Asian populations. Much of the difference in incidence between whites and blacks in North America can be attributed to the higher incidence seen in higher socioeconomic classes. The cumulative lifetime risk for development of Hodgkin lymphoma is approximately 1 in 250 to 1 in 300 in North America.

The incidence of Hodgkin lymphoma is bimodal in distribution, rising from very low in childhood to an early peak in young adulthood (25 to 30 years) and a later peak in late adulthood (>50 years). In the Western world, only about 5% of cases occur in persons younger than 15 years and 5% in individuals older than 70 years. In contrast, the age distribution in the Indian subcontinent is strongly shifted into childhood.

PATHOBIOLOGY

The cause of Hodgkin lymphoma remains unclear. Hodgkin lymphoma is not associated with exposure to radiation, chemicals, biocidal agents, working in health care–related professions, or previous tonsillectomy. The leading suspect remains EBV, based on much suggestive evidence but no definitive proof.[1]

Epstein-Barr Virus

EBV is a large B-lymphocyte tropic herpesvirus (Chapter 377). Approximately 90% of the general population acquires infection with EBV by early adulthood. In the developing world, this infection usually occurs in childhood, but in developed countries, infection is often delayed into the teens, when it is associated with the syndrome of infectious mononucleosis in up to 30% of new cases. A history of infectious mononucleosis increases the likelihood for subsequent Hodgkin lymphoma three-fold. Antibodies to the viral capsid antigen reach higher levels in patients with Hodgkin lymphoma than in controls, and these higher levels appear several years before the neoplasm. In situ hybridization studies have demonstrated that the Hodgkin-Reed-Sternberg cells in approximately 50% of cases of Hodgkin lymphoma contain EBV-encoded small RNA (EBER), and in these cases, virtually all the Hodgkin-Reed-Sternberg cells are positive for the virus. The EBV genome is amplified 50-fold or more in Hodgkin-Reed-Sternberg cells and is monoclonal in an individual patient's Hodgkin-Reed-Sternberg cells. In some populations, virtually all cases of Hodgkin lymphoma occur in EBV-positive individuals, but up to 50% of patients in developed countries do not have EBV in their Hodgkin-Reed-Sternberg cells. Thus, although EBV may play an important role in the development of Hodgkin lymphoma, that role is neither straightforward nor universal.

Genetic Factors

Circumstantial evidence for a genetic contribution to the etiology of Hodgkin lymphoma comes from studies showing that Hodgkin lymphoma is almost 100-fold more likely to develop in the monozygotic twin of an affected individual than in a dizygotic twin. First-degree relatives of individuals with the disease have up to a five-fold increased risk for development of the lymphoma. Perhaps genetically predisposed individuals react differently to EBV, thereby increasing the chance that a lymphoid neoplasm will develop.

Polymerase chain reaction–based genotypic analysis has demonstrated the clonal derivation of Hodgkin-Reed-Sternberg cells, including identical *p53* mutations from multiple Hodgkin-Reed-Sternberg cells extracted from a single biopsy specimen, thereby unequivocally establishing clonality. The presence of clonal immunoglobulin gene rearrangements from multiple cells in the same biopsy specimen also confirms a B-cell origin. Only a few rare cases with a T-cell genotype have been reported, but these are exceptional. The presence of clonal somatic mutations provides proof of the germinal center origin of the neoplastic cells. Finally, identification of cells with identical immunoglobulin gene rearrangements both at diagnosis and at relapse verify that the B-cell clonality of the disease is preserved over time.

Despite their B-cell origin, the neoplastic cells of Hodgkin lymphoma are incapable of making intact antibodies, perhaps because they lack the ability to make the transcription factors necessary to activate the immunoglobulin promoter. B cells that are incapable of manufacturing antibody should undergo apoptosis, but the Hodgkin-Reed-Sternberg cells avoid

self-destruction. The observation that the antiapoptotic nuclear transcription factor NFκB is constitutively activated in these cells may provide an explanation.

Classic cytogenetics has been unrevealing in Hodgkin lymphoma. Aneuploidy and hyperploidy consistent with the multinucleated nature of Hodgkin-Reed-Sternberg cells are frequent, but no consistent translocations have been detected. A germline *NPAT* mutation and multiple possible sites on chromosome 6 at 6p21.3[2,3] have been associated with an increased risk for Hodgkin lymphoma.

CLINICAL MANIFESTATIONS

Hodgkin lymphoma is usually manifested as lymphadenopathy (Chapter 168), typically in the cervical, axillary, or mediastinal areas, and only about 10% of patients present with nodal disease below the diaphragm. Although peripherally located nodes seldom reach large size, very large mediastinal masses or, less often, retroperitoneal masses can develop with only modest symptoms. Lymph node involvement in Hodgkin lymphoma is usually painless, but an occasional patient notes discomfort in involved nodal sites immediately after drinking alcohol.

Approximately 25% of patients with Hodgkin lymphoma have constitutional symptoms. The classic B symptoms, significant weight loss (>10% of baseline), night sweats, and persistent fever, usually signal widespread or locally extensive disease and imply a need for systemic treatment. Generalized pruritus, occasionally severe, can antedate the diagnosis of Hodgkin lymphoma by up to several years. Some patients have symptoms suggestive of a growing mass lesion, such as cough or stridor as a result of tracheobronchial compression from mediastinal disease or bone pain secondary to metastatic involvement. Because Hodgkin lymphoma can involve the bone marrow extensively, an occasional patient presents with symptomatic anemia or incidentally noted pancytopenia. Paraneoplastic neurologic or endocrine syndromes have been reported with Hodgkin lymphoma but are rare.

DIAGNOSIS

The diagnosis of classic Hodgkin lymphoma is based on recognition of Hodgkin-Reed-Sternberg cells (Fig. 186-1) or Hodgkin cells (or both) in an appropriate cellular background in tissue sections from a lymph node or extralymphatic organ, such as bone marrow, lung, or bone. Fine-needle aspiration biopsy is not adequate for the diagnosis of Hodgkin lymphoma. Open biopsy and standard histochemical staining are required to establish the diagnosis unequivocally and to determine the histologic subtype. Immunohistochemical studies can prove helpful in difficult cases or to distinguish special subtypes such as the lymphocyte-rich classic and nodular lymphocyte-predominant types. In classic Hodgkin lymphoma, scattered large Hodgkin-Reed-Sternberg cells either are multinucleated or have large polyploid nuclei. Variations include mononuclear cells that are similar to the usual polylobated or multinuclear cells but have only one large nucleus with a prominent nucleolus, as well as lacunar cells, which are Hodgkin-Reed-Sternberg variants with abundant cytoplasm that has retracted as an artifact of formalin fixation. The infrequent Hodgkin-Reed-Sternberg cells are usually present in a background mixture of polyclonal lymphocytes, eosinophils, neutrophils, plasma cells, fibroblasts, and histiocytes. A high number of associated macrophages have been demonstrated to be a strong predictor of treatment resistance. Occasionally, granulomas form with a prominent histiocytic component.

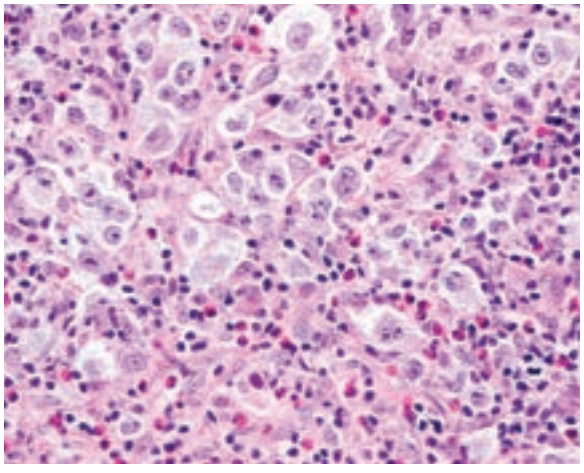

FIGURE 186-1. Nodular sclerosing Hodgkin lymphoma. This figure shows a typical case of classic nodular sclerosing Hodgkin lymphoma with many lacunar cells, occasional diagnostic Hodgkin-Reed-Sternberg cells, and the characteristic background of lymphocytes and eosinophils. (Photomicrograph courtesy of Randy D. Gascoyne, MD, British Columbia Cancer Agency.)

Hodgkin lymphoma can typically be classified into well-described subtypes (Table 186-1). Reproducibility of the distinctions among these subtypes has been confirmed in the current widely accepted World Health Organization classification of lymphoid neoplasms. With addition of the new category of lymphocyte-rich classic Hodgkin lymphoma, this newest classification scheme permits confident identification of nodular lymphocyte-predominant Hodgkin lymphoma as a separate entity. The most common subtype is nodular sclerosing, which has characteristic coarse bands of sclerosis surrounding nodules composed of typical Hodgkin-Reed-Sternberg cells in the usual background mixture of reactive and inflammatory cells.

The immunophenotype of the neoplastic cells in Hodgkin lymphoma can help identify the specific subtype. Typically, the Hodgkin-Reed-Sternberg cells stain positively for CD30 (80 to 100% of cases), CD15 (75 to 85% of cases), and B-cell-specific activating protein (BSAP), which is the product of the *PAX5* gene (>90% of cases). However, often only a minority of the malignant cells stain positively for the CD15 and BSAP markers. CD20, a generally reliable marker of B-cell lineage, is positive in about 40% of cases of classic Hodgkin lymphoma, but usually only in a minority of cells, and the staining can be weak. In contrast, nodular lymphocyte-predominant Hodgkin lymphoma almost always stains strongly positive for CD20 and for the specialized B-cell markers CD79a and CD45, but it is negative for CD30 and CD15. Anaplastic large cell lymphoma (Chapter 185) is positive for CD30 but reliably negative for CD15, CD20, and CD79a.

Differential Diagnosis

Depending on the site of occurrence and associated symptoms, the differential diagnosis of Hodgkin lymphoma includes non-Hodgkin lymphoma (Chapter 185), germ cell tumors (Chapter 200), thymoma (Chapter 422), sarcoidosis (Chapter 95), and tuberculosis (Chapter 324). However, the specific diagnosis is readily determined by obtaining an adequate biopsy specimen for review by an experienced hematopathologist. Proceeding to such a biopsy early in the assessment of patients with lymphadenopathy (Chapter 168), especially of the mediastinum, often saves time and spares the patient needless testing and delay in diagnosis.

With computed tomography (CT) and appropriate biopsy procedures to investigate enlarged central thoracic or intra-abdominal lymph nodes, the diagnosis of Hodgkin lymphoma seldom presents difficulty. The immunophenotype helps distinguish Hodgkin lymphoma from other diseases. For example, T-cell-rich B-cell lymphoma (Chapter 185) is distinguished from classic Hodgkin lymphoma by being CD30 and CD15 negative but positive for CD20 and CD45. However, T-cell-rich B-cell lymphoma (Chapter 185) can be very difficult to distinguish from nodular lymphocyte-predominant Hodgkin lymphoma because both are negative for CD30 and CD15 but positive for CD45. This distinction is best made by focusing on the histologic pattern of the neoplastic cells. In fact, the combination of appropriate immunohistopathologic evaluation by an expert hematopathologist and clinical assessment has virtually eliminated difficulties with the differential diagnosis. Problems mostly arise when inadequate or improperly processed material is all that is available for diagnosis.

TABLE 186-1 WORLD HEALTH ORGANIZATION CLASSIFICATION OF HODGKIN LYMPHOMA SUBTYPES

SUBTYPE NAME	FREQUENCY (%)*
Classic Hodgkin lymphoma	
Nodular sclerosis	70
Lymphocyte rich	3
Mixed cellularity	10
Lymphocyte depleted	1
Nodular lymphocyte-predominant Hodgkin lymphoma	7
Hodgkin lymphoma, not otherwise classifiable	9

*Frequency based on all new cases ($N = 1043$) seen in British Columbia since January 1998 when the category of lymphocyte-rich classic Hodgkin lymphoma became well established.

Staging

Physical Examination

Given its tendency to spread in an orderly fashion, usually from initially involved lymph nodes, the stage of Hodgkin lymphoma can be established by using readily available imaging and laboratory tests (Fig. 186-2 and Table 186-2). The evaluation should start with a careful history to search for the presence of localizing signs, such as bone pain, or the constitutional symptoms of fever, weight loss, or night sweats. The history may also reveal comorbid conditions that may affect the safe delivery of planned treatment. The physical examination may identify lymphadenopathy or organomegaly.

Laboratory Testing

Laboratory testing should include blood cell counts and the erythrocyte sedimentation rate, assessment of liver and renal function, serum albumin level, serum protein electrophoresis, and serologic testing for hepatitis B. Also, the patient should be tested for hepatitis C if liver enzyme abnormalities are detected and human immunodeficiency virus (HIV) antibody if the history indicates an increased risk or if the sites of extranodal disease are unusual. Bone marrow aspiration and biopsy are only useful for the minority of patients with constitutional (B) symptoms or those with lower than normal peripheral blood counts at diagnosis and may be rendered entirely unnecessary by fluorodeoxyglucose positron emission tomography if current studies are validated.[4]

TABLE 186-2 TESTS REQUIRED FOR STAGING OF HODGKIN LYMPHOMA

Complete history to search for B symptoms (fever, weight loss, night sweats) or other symptomatic problems suggesting more advanced disease	
Physical examination for lymphadenopathy or organomegaly	
Complete blood count	
Serum creatinine, alkaline phosphatase, lactate dehydrogenase, bilirubin, and protein electrophoresis (including serum albumin level)	
Chest radiograph, posteroanterior and lateral views	
Computed tomography scan of the neck, thorax, abdomen, and pelvis	
Fluorodeoxyglucose positron emission tomography with computed tomography (PET/CT)	
Certain tests are required only for specific manifestations	
MANIFESTATION/CONDITION	**TEST**
B symptoms or WBC count <4.0 × 10^9/L, Hgb <120 g/L (women) or 130 g/L (men), or platelets <125 × 10^9/L	Bone marrow biopsy and aspiration
Stage IA or IIA disease with upper cervical lymph node involvement (suprahyoid)	ENT examination

ENT = ear, nose, and throat; Hgb = hemoglobin; WBC = white blood cell.

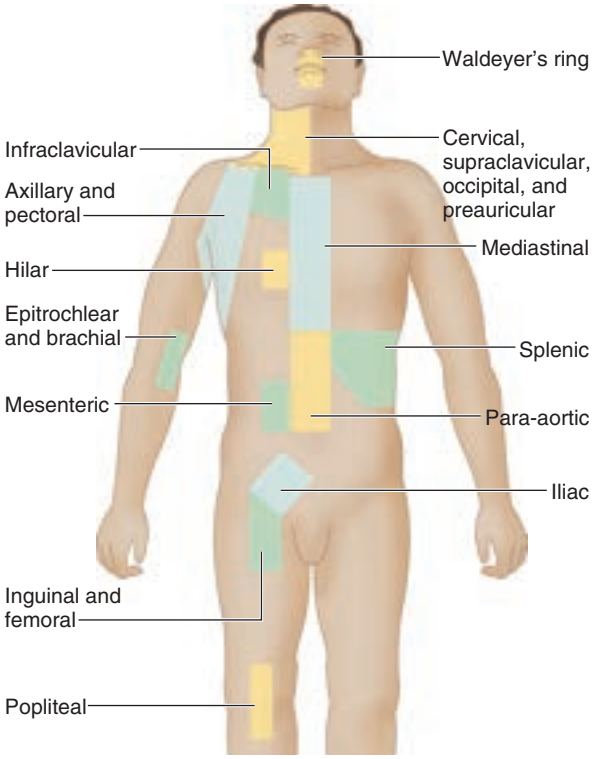

FIGURE 186-2. Anatomic definition of lymph node regions for staging of Hodgkin disease. (From Kaplan HS, Rosenberg SA. The treatment of Hodgkin disease. *Med Clin North Am*. 1966;50:1591-1610.)

Imaging

Imaging techniques to evaluate Hodgkin lymphoma continue to evolve (Fig. 186-3). All patients should undergo contrast-enhanced CT scanning of the neck, thorax, abdomen, and pelvis with slices at intervals of 1 cm or less. Magnetic resonance imaging is occasionally useful when the extent of bone or soft tissue involvement must be determined precisely or for a patient with an absolute contraindication to the use of intravenous contrast agents.

Positron Emission Tomography

Fluorodeoxyglucose positron emission tomography (PET) is more sensitive and specific than CT or gallium scanning both for staging and for assessment of residual masses after treatment and is now considered mandatory for staging of Hodgkin lymphoma.[5] It may also be useful for the assessment of residual masses during or after planned treatment, allowing better selection of patients who should receive altered or additional therapy, especially radiotherapy.

Staging System

The Ann Arbor staging system with the Cotswold modification (Table 186-3) categorizes patients into four stages. The first three indicate the expanding extent of lymph node disease (see Fig. 186-2): stage I, a single nodal area; stage II, two or more nodal areas but still on one side of the diaphragm; and stage III, nodal disease on both sides of the diaphragm. The spleen and the lymphoid tissue of Waldeyer's ring each count as nodal sites in this system. Stage IV is reserved for extranodal disease, which for all practical purposes is disease in the bone marrow, lung, bone, or liver. Hodgkin lymphoma at any other extranodal site should prompt questioning of the diagnosis or a search for HIV infection.

Bulky disease is defined as the presence of any tumor mass with the largest diameter greater than 10 cm or a mediastinal mass with a transverse diameter exceeding one third of the largest transverse transthoracic diameter. With CT scanning, use of the mediastinal mass ratio is obsolete, and the term *bulky* is best assigned to tumors exceeding 10 cm in largest single diameter.

The E lesion designation identifies patients whose limited extranodal extension of Hodgkin lymphoma could be included in a reasonable involved field of irradiation. As part of staging, patients are further subdivided into those with or without fever, night sweats, or weight loss (B symptoms).

TREATMENT

Over the past 70 years, Hodgkin lymphoma has been transformed from a nearly uniformly fatal illness to one that is usually cured. This remarkable success has provided a paradigm on which much of modern oncologic treatment is based. The principles underlying combined-modality treatment and multiagent chemotherapy, the mainstays of today's successful treatment of many malignancies, were first demonstrated to be effective with Hodgkin lymphoma. The essential involvement of a multidisciplinary team, including pathologists, experts in diagnostic imaging, medical and radiation oncologists, nurses, and support staff, has served as a model for all cancer. The necessity to balance greater efficacy of initial treatment, which often requires an increase in intensity and therefore toxicity, against troublesome and occasionally fatal late complications has encouraged a long-term perspective and careful monitoring for late treatment-related side effects.

From a practical therapeutic viewpoint, patients with stage III or IV disease or bulky disease are defined as having advanced disease, whereas patients without these characteristics have limited-stage disease. In Europe, patients with limited disease are further subdivided into those with favorable and unfavorable outcomes. Cure rates exceed 90 to 95% for patients with nonbulky stage IA or IIA (limited) disease. However, for patients with advanced-stage disease, independent predictors of progression include gender, age, stage, hemoglobin level, white blood cell count, lymphocyte count, and serum albumin level (Table 186-4). Based on results obtained in the 1980s, the 80% of patients with fewer than four factors had a progression-free survival of 70%,

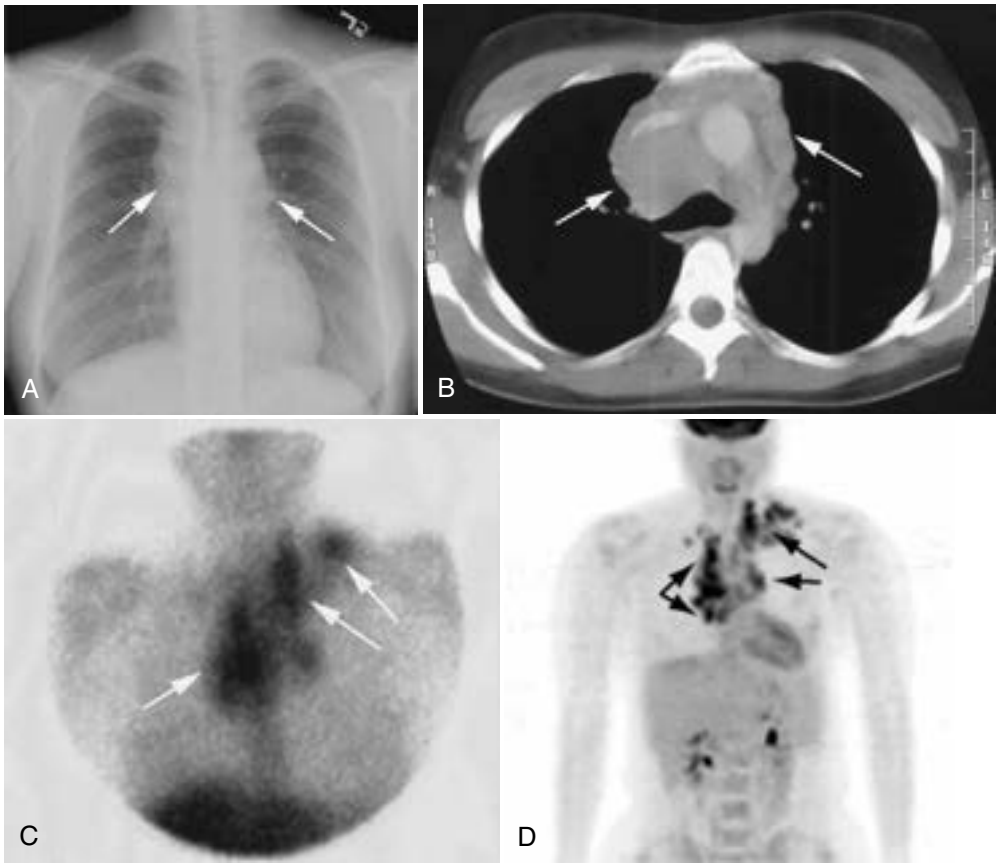

FIGURE 186-3. Imaging of Hodgkin lymphoma. Bulky Hodgkin disease as seen on chest radiograph (**A**), computed tomography (CT) of the chest (**B**), gallium scan (**C**), and positron emission tomography (PET) (**D**). The *arrows* indicate sites of disease. Note that the PET and CT scans provide more detailed information than the chest radiograph and gallium scan.

TABLE 186-3 MODIFIED ANN ARBOR STAGING SYSTEM FOR HODGKIN LYMPHOMA

STAGE	INVOLVEMENT
I	Single lymph node region (I) or one extralymphatic site (I_E)
II	Two or more lymph node regions, same side of the diaphragm (II), or local extralymphatic extension plus one or more lymph node regions, same side of the diaphragm (II_E)
III	Lymph node regions on both sides of the diaphragm (III); may be accompanied by local extralymphatic extension (III_E)
IV	Diffuse involvement of one or more extralymphatic organs or sites
A	No "B" symptoms
B	Presence of at least one of the following: Unexplained weight loss >10% of baseline during 6 mo before staging Recurrent unexplained fever >38° C Recurrent night sweats

TABLE 186-4 RATES OF PROGRESSION IN 5 YEARS IN PATIENTS WITH ADVANCED-STAGE HODGKIN LYMPHOMA

NO. OF FACTORS*	FREQUENCY (%)	PERCENTAGE WITH PROGRESSION-FREE SURVIVAL AT 5 YEARS
0-3	81	70
4-7	19	47

*Male sex, age older than 45 years, stage IV, hemoglobin level less than 10.5 g/dL, white blood cell count greater than 15,000/mL, lymphocyte count less than 600/mL or less than 8% of the white cell count, or serum albumin less than 4 g/dL.

TABLE 186-5 TREATMENT PLAN FOR HODGKIN LYMPHOMA BASED ON STAGE

STAGE	TREATMENT
IA or IIA, no bulky disease*	ABVD† × 4 if CR after 2 cycles, or ABVD × 2 + IRRT
IB, IIB, or any stage III or IV or bulky disease, any stage	ABVD† until 2 cycles past CR (minimum 6, maximum 8) or escalated BEACOPP†

*Bulky refers to disease with the largest diameter of any single mass equal to or greater than 10 cm.
†See text for drugs in each regimen. Optimal dosing must be individualized.
CR = complete response.

but for the 20% who had four or more factors, the progression-free survival rate fell to less than 50%. However, with better attention to dose delivery, more current results indicate a spread of only 80% to 60% between these two groups.[6] A straightforward plan of treatment for the 90% of patients in whom Hodgkin lymphoma is diagnosed between the ages of 16 and 70 years can be based on clinical stage, the presence of B symptoms, and bulk of the largest tumor mass (Table 186-5).

Treatment of Limited-Stage Hodgkin Lymphoma

More than 95% of the one third of patients with Hodgkin lymphoma initially found to have limited-stage disease can be cured regardless of the site of occurrence, the presence of disease above or below the diaphragm, or the histologic subtype. The challenge is to achieve this goal with the least toxicity and cost.

Two cycles of ABVD (Adriamycin [doxorubicin], bleomycin, vinblastine, and dacarbazine) chemotherapy followed by involved-field irradiation cures nearly 95% of patients with limited-stage disease, nearly completely eliminates the risk for infertility, premature menopause, and leukemia, and minimizes cardiopulmonary toxicity.[7] However, chemotherapy alone has been shown to be

a viable alternative for patients with a good initial response to treatment and further reduces the risks of exposure to radiation.[7] More intensified approaches have been tested for early-stage but unfavorable Hodgkin lymphoma with adverse prognostic factors such as bulky mediastinal mass, extranodal extension, elevated erythrocyte sedimentation rate (ESR), or three or more lymph node areas involved. In such patients, two cycles of BEACOPP (bleomycin, etoposide, Adriamycin, cyclophosphamide, vincristine, procarbazine, and prednisone), followed by two cycles of ABVD before involved-field radiotherapy, significantly improves tumor control compared with only four cycles of ABVD before involved-field radiotherapy, but long-term results do not differ.[8] The chemotherapy in the combined-modality treatment of limited-stage Hodgkin lymphoma eradicates subclinical disease and allows smaller fields of irradiation to be used. However, a substantial proportion of the excess, long-term mortality in patients with limited-stage Hodgkin lymphoma is due to cardiovascular disease and second neoplasms that are closely related to the use of irradiation.[9] In a randomized trial that compared four to six cycles of ABVD chemotherapy alone versus irradiation, either alone or augmented with two cycles of ABVD chemotherapy, the strategy of chemotherapy alone proved equivalent to irradiation-based treatment in terms of event-free and overall survival, although the irradiation-based approach did produce a modest improvement in progression-free survival. At 12 years, however, ABVD therapy alone without subtotal nodal radiation therapy is associated with a better overall survival (94% vs. 87%) owing to a lower rate of death from other causes. These data suggest that patients with limited-stage Hodgkin lymphoma can be successfully treated with four to six cycles of ABVD alone; for the minority whose lymphoma does not completely regress after two cycles, probably best assessed by PET scanning, the addition of radiation may be optimal.

Treatment of Advanced-Stage Hodgkin Lymphoma

In advanced-stage Hodgkin lymphoma (stages IIIA, IIIB, IVA, and IVB), ABVD has become the most widely used regimen. The addition of radiation therapy significantly improves progression-free survival at 10 years in patients with advanced-stage Hodgkin lymphoma, but it does not improve overall survival. The adverse long-term effects of radiation therapy and its lack of improvement in overall survival appear to outweigh any benefits for the usual patient with advanced-stage disease. A negative PET scan at the end of chemotherapy distinguishes between residual fibrosis and persistent lymphoma and identifies the three fourths of patients with a residual mass who do not require radiation therapy.[7]

Recently devised regimens for patients with advanced Hodgkin lymphoma are the Stanford V regimen (doxorubicin, vinblastine, mechlorethamine, etoposide, vincristine, bleomycin, and prednisone) and escalated BEACOPP (bleomycin, etoposide, doxorubicin, cyclophosphamide vincristine, procarbazine, and prednisone) (see Table 186-5). As originally described, both include post-chemotherapy irradiation to sites of initial or residual tumor bulk (≥5 cm), but radiation does not appear to be routinely necessary after BEACOPP. Although initial results appeared quite promising, the Stanford V regimen is no more effective than standard ABVD.[7] Assessing the role of escalated BEACOPP for advanced-stage Hodgkin lymphoma requires consideration of potential secondary treatments for those not cured by primary therapy. About 50% of patients who are not cured by primary chemotherapy can be effectively treated with HDC/HSCT (Chapter 178). Thus, although escalated BEACOPP offers better initial disease control, randomized trials have failed to demonstrate its superiority over ABVD for overall survival.[7,8]

Management of Refractory or Relapsed Hodgkin Lymphoma

HDC/HSCT has become the established treatment for most patients whose Hodgkin lymphoma persists or recurs despite primary chemotherapy. However, the treatment-related mortality, high levels of toxicity, and cost associated with HDC/HSCT demand that it be reserved for patients in whom it clearly increases the chance of cure over alternative treatments;[10] such patients include those whose disease progresses during or within 3 months of initial multiagent chemotherapy (refractory Hodgkin lymphoma) and those who relapse more than 3 months after multiagent chemotherapy (relapsed Hodgkin lymphoma). For relapsed lymphoma, controversy remains, however, for two special subgroups: patients who relapse solely in originally involved but unirradiated lymph nodes and without B symptoms or extranodal disease, who may obtain up to a 40 to 50% cure rate with wide-field irradiation, and patients who relapse without B symptoms more than 1 year after completion of primary chemotherapy, who may achieve up to a 30 to 40% cure rate with additional chemotherapy with or without irradiation. However, even these two subgroups may achieve up to an 80% 10-year disease-free survival rate after HDC/HSCT. Thus, data suggest that standard treatment for patients with progressive Hodgkin lymphoma after primary chemotherapy for advanced-stage disease should be HDC/HSCT regardless of the characteristics of the relapse.

Management of Complications

Follow-up and Late Complications of Treatment

Most adult patients with Hodgkin lymphoma are cured and experience minimal long-term toxicity from their treatment. However, the risk for certain predictable and rare and less predictable late effects warrants careful but not intrusive follow-up and selective intervention (Table 186-6). At the conclusion of treatment, a thorough reassessment of the initial sites of lymphoma should be completed to provide post-treatment baseline measurements. Patients should be seen by a specialist knowledgeable in the management of lymphoma, preferably about every 3 months for 2 years, then every 6 months for 3 years, then annually. Patients should be strongly encouraged to refrain from smoking, to perform careful breast and skin examinations on a regular basis, and to undergo regular immunizations for influenza annually, pneumococcus at diagnosis and 5 years after treatment, and diphtheria and tetanus every 10 years (Chapter 18). Patients who have received radiation to the head or neck area should follow a vigorous program of dental prophylaxis in anticipation of the deleterious effect of reduced saliva production and should have their thyroid-stimulating hormone (TSH) level checked annually in recognition of the 50% risk for eventual hypothyroidism.

Other potential long-term sequelae of Hodgkin lymphoma treatment[11] include second malignancies. Radiation therapy–related solid tumors, most commonly nonmelanoma skin cancers, lung, breast, and colorectal malignancies, have a median latency period of more than 14 years after treatment. Chemotherapy-related myelodysplasia and acute myelogenous leukemia are more likely to be caused by alkylating agents that are included in regimens like MOPP and BEACOPP compared with ABVD and have shorter latency periods (median of 3 years) after treatment.[11] Long-term (usually >10 years after treatment) cardiovascular complications are coronary artery disease (related to mediastinal irradiation) and cardiomyopathy (caused by the cardiotoxicity of cumulative doses of anthracycline chemotherapy). Chemotherapy, particularly alkylating agents in MOPP and BEACOPP but not ABVD regimens, can cause permanent infertility. Long-term survivors of childhood Hodgkin lymphoma (median age at diagnosis of 15 years) are at increased risk for neurocognitive impairment.[12]

TABLE 186-6 MONITORING AFTER SUCCESSFUL PRIMARY TREATMENT OF HODGKIN LYMPHOMA

RISK/PROBLEM	INCIDENCE/RESPONSE
Relapse	Ten to 30% of patients experience relapse. Careful attention should be directed to lymph node sites, especially if previously involved with disease and not treated with radiation. New persistent focal symptoms such as bone pain should be investigated with appropriate laboratory and imaging studies.
Dental caries	Neck or oropharyngeal irradiation may cause decreased salivation. Patients should have regular dental care and should make their dentist aware of the previous irradiation.
Hypothyroidism	After external beam thyroid irradiation at doses sufficient to cure Hodgkin lymphoma, at least 50% of patients eventually become hypothyroid. All patients who have been exposed to neck irradiation should have an annual TSH level determined. Patients whose TSH level becomes elevated should be treated with lifelong thyroxine replacement in doses sufficient to suppress TSH levels to low normal (Chapter 226).
Infertility	ABVD is not known to cause permanent gonadal toxicity, although temporary oligospermia or irregular menses may persist for 1 to 2 years after treatment. Direct or scatter radiation to gonadal tissue may cause infertility, amenorrhea, or premature menopause, but this adverse event seldom occurs with the current fields used for the treatment of Hodgkin lymphoma. In general, women who continue menstruating are fertile, but men require semen analysis to provide a specific answer.
Secondary neoplasms	Although uncommon, certain secondary neoplasms occur with increased frequency in patients who have been treated for Hodgkin lymphoma: acute myelogenous leukemia; thyroid, breast, lung, cervical, and upper gastrointestinal carcinoma; and melanoma. It is appropriate to "be vigilant" for these neoplasms for the remainder of the patient's life because they may have a lengthy induction period.

ABVD = doxorubicin (Adriamycin), bleomycin, vinblastine, and dacarbazine; TSH = thyroid-stimulating hormone.

Special Problems in the Management of Hodgkin Lymphoma

Hodgkin Lymphoma during Pregnancy

Between 0.5 and 1.0% of cases of Hodgkin lymphoma occur coincident with pregnancy (Chapter 239). When the lymphoma is discovered during pregnancy, it is almost always possible to keep it under control and allow the pregnancy to go to full term.

Standard staging tests (see Table 186-2) should be completed, except that imaging requiring radiation must be minimized. For example, abdominal ultrasonography can identify bulky retroperitoneal disease, and a single posteroanterior radiograph of the chest, with proper shielding, can identify bulky mediastinal disease.

Patients can often continue the pregnancy to term without any treatment of the lymphoma.[13] If symptomatic or progressive disease develops, systemic chemotherapy can be given in the second and third trimester with very low risk for injuring the fetus. An attractive alternative to multiagent chemotherapy is intermittent single-agent vinblastine, given in the lowest dose that can control symptoms until delivery, followed by a full course of six to eight cycles of multiagent chemotherapy after delivery.

Hodgkin Lymphoma and Acquired Immunodeficiency Syndrome

In patients with HIV infection, the incidence of Hodgkin lymphoma is increased as much as 5- to 10-fold, and the lymphoma manifests differently and pursues a more aggressive natural history (Chapter 393). Hodgkin lymphoma in HIV-positive individuals is almost always associated with EBV within Hodgkin-Reed-Sternberg cells. The histology is much more likely to be mixed cellularity or lymphocyte depleted. The disease most commonly develops in extranodal sites, especially the bone marrow. More than 80% of patients have advanced-stage disease, and most patients have B symptoms.

Patients are prone to opportunistic infections, and the interaction of chemotherapeutic agents with other medications may compromise the patient's ability to tolerate treatment. The best approach is a combination of highly active antiretroviral agents (Chapter 389), vigorous supportive care with antiherpetic and antifungal agents and neutrophil-stimulating growth factors, and standard multiagent chemotherapy. With appropriate supportive care, regimens such as ABVD can be delivered. However, more severe than normal toxicity must be anticipated, and although cure rates for the lymphoma are comparable to those seen in HIV-negative individuals, median overall survival is much shorter than that seen in the non-HIV-infected patients, typically 3 to 4 years.[14]

Hodgkin Lymphoma in the Elderly Population

Elderly patients with Hodgkin lymphoma have a worse outcome. For example, the 5-year overall survival rate falls from 80% in patients younger than 65 years to less than 50% in patients older than 65 years. Explanations include more advanced stage at diagnosis, comorbid diseases, delay in diagnosis, incomplete staging, inadequate adherence to treatment protocols, and failure to maintain full dose intensity.

Of note is that elderly patients achieve outcomes equivalent to those of younger patients when they receive similar doses of chemotherapy. The best approach for elderly patients is to attempt to treat them in a manner similar to younger patients, with vigorous supportive care and the addition of neutrophil growth factors if necessary to enable safe delivery of full doses. For patients with preexisting pulmonary or cardiac disease, it might be necessary to reduce or eliminate bleomycin or doxorubicin, respectively.

FUTURE DIRECTIONS

The ability to profile multigene expression patterns and identify genetic polymorphisms associated with specific malignancies may provide better insight into the molecular genesis of Hodgkin and other lymphomas.[15] Therapeutic agents more specifically targeted at the malignant Hodgkin-Reed-Sternberg cells, such as those coupling an antibody to the CD30 antigen with a cellular toxin, have proved remarkably effective and hold substantial promise to improve treatment outcome.

Grade A References

A1. Meyer RM, Gospodarowicz MK, Connors JM, et al. ABVD alone versus radiation-based therapy in limited-stage Hodgkin lymphoma. *N Engl J Med*. 2012;366:399-408.

A2. Engert A, Haverkamp H, Kobe C, et al. Reduced-intensity chemotherapy and PET-guided radiotherapy in patients with advanced stage Hodgkin lymphoma (HD15 trial): a randomised, open-label, phase 3 non-inferiority trial. *Lancet*. 2012;379:1791-1799.

A3. Gordon LI, Hong F, Fisher RI, et al. Randomized phase III trial of ABVD versus Stanford V with or without radiation therapy in locally extensive and advanced-stage Hodgkin lymphoma: an intergroup study coordinated by the Eastern Cooperative Oncology Group (E2496). *J Clin Oncol*. 2013;31:684-691.

A4. Mounier N, Brice P, Bologna S, et al. ABVD (8 cycles) versus BEACOPP (4 escalated cycles ≥ baseline): final results in stage III-IV low-risk Hodgkin lymphoma (IPS 0-2) of the LYSA H34 randomized trial. *Ann Oncol*. 2014;25:1622-1628.

A5. Viviani S, Zinzani PL, Rambaldi A, et al. ABVD versus BEACOPP for Hodgkin lymphoma when high-dose salvage is planned. *N Engl J Med*. 2011;365:203-212.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Vockerodt M, Yap LF, Shannon-Lowe C, et al. The Epstein-Barr Virum and the Pathogenesis of Lymphoma. *J Pathol*. 2014;235:312-322.
2. Cozen W, Li D, Best T, et al. A genome-wide meta-analysis of nodular sclerosing Hodgkin lymphoma identifies risk loci at 6p21.32. *Blood*. 2012;119:469-475.
3. King RL, Howard MT, Bagg A. Hodgkin lymphoma: pathology, pathogenesis, and a plethora of potential prognostic predictors. *Adv Anat Pathol*. 2014;21:12-25.
4. Khan AB, Barrington SF, Mikhaeel NG, et al. PET-CT staging of DLBCL accurately identifies and provides new insight into the clinical significance of bone marrow involvement. *Blood*. 2013;122:61-67.
5. Kostakoglu L, Evens AM. FDG-PET imaging for Hodgkin lymphoma: current use and future applications. *Clin Adv Hematol Oncol*. 2014;12:20-35.
6. Moccia AA, Donaldson J, Chhanabhai M, et al. International Prognostic Score in advanced-stage Hodgkin lymphoma: altered utility in the modern era. *J Clin Oncol*. 2012;30:3383-3388.
7. Engert A, Plutschow A, Eich HT, et al. Reduced treatment intensity in patients with early-stage Hodgkin lymphoma. *N Engl J Med*. 2010;363:640-652.
8. von Tresckow B, Plutschow A, Fuchs M, et al. Dose-intensification in early unfavorable Hodgkin lymphoma: final analysis of the German Hodgkin Study Group HD14 trial. *J Clin Oncol*. 2012;30:907-913. PMID 22271480.
9. Andre MP. Combination chemoradiotherapy in early Hodgkin lymphoma. *Hematol Oncol Clin North Am*. 2014;28:33-47.
10. Rancea M, von Tresckow B, Monsef I, et al. High-dose chemotherapy followed by autologous stem cell transplantation for patients with relapsed or refractory Hodgkin lymphoma: a systematic review with meta-analysis. *Crit Rev Oncol Hematol*. 2014;92:1-10.
11. Thompson CA, Mauck K, Havyer R, et al. Care of the adult Hodgkin lymphoma survivor. *Am J Med*. 2011;124:1106-1112.
12. Krull KR, Sabin ND, Reddick WE, et al. Neurocognitive function and CNS integrity in adult survivors of childhood Hodgkin lymphoma. *J Clin Oncol*. 2012;30:3618-3624.
13. Evens AM, Advani R, Press OW, et al. Lymphoma occurring during pregnancy: antenatal therapy, complications, and maternal survival in a multicenter analysis. *J Clin Oncol*. 2013;31:4132-4139.
14. Montoto S, Shaw K, Okosun J, et al. HIV status does not influence outcome in patients with classical Hodgkin lymphoma treated with chemotherapy using doxorubicin, bleomycin, vinblastine, and dacarbazine in the highly active antiretroviral therapy era. *J Clin Oncol*. 2012;30:4111-4116.
15. Scott DW, Chan FC, Hong F, et al. Gene expression-based model using formalin-fixed paraffin-embedded biopsies predicts overall survival in advanced-stage classical Hodgkin lymphoma. *J Clin Oncol*. 2013;31:692-700.

REVIEW QUESTIONS

1. A 23-year-old man has noticed bilateral lower neck firm masses eventually reaching about 3 cm in greatest diameter over the past 3 months, which have not resolved after 2 more weeks of observation during which he took a 7-day course of oral antibiotics. A fine-needle aspiration biopsy reveals predominantly small lymphocytes and granulocytes, small numbers of eosinophils, and occasional monoclonal large multinucleated cells that are positive for CD30 and variably positive for CD20 and are judged consistent with Hodgkin-Reed-Sternberg cells and a diagnosis of probable Hodgkin lymphoma. What is the most appropriate next step in this patient's evaluation?

A. Computed tomography (CT) of the neck, chest, abdomen and pelvis
B. Whole body positron emission tomography/computed tomography (PET/CT)
C. Bone marrow biopsy
D. Whole lymph node excisional biopsy of an abnormal neck lymph node
E. A 14-day course of ciprofloxacin

Answer: D This patient probably has Hodgkin lymphoma; however, a firm diagnosis cannot be made based on needle aspiration biopsy. Other conditions that could be present and cannot be distinguished from Hodgkin lymphoma using only a needle aspiration biopsy include anaplastic large cell lymphoma, primary mediastinal B-cell lymphoma, T-cell-rich B-cell lymphoma, or unusual atypical lymphoid hyperplasia. Accurate diagnosis of Hodgkin lymphoma requires at least an incisional biopsy of involved nodal tissue, preferable an entire enlarged lymph node. Initiating extensive evaluation or further delay in assessment without a firm diagnosis is inappropriate.

2. A 26-year-old schoolteacher seeks a second opinion from you. She has been found to have neck and mediastinal nodular sclerosing Hodgkin lymphoma based on an excisional biopsy of a left supraclavicular fossa lymph node, PET/CT scanning showing definitely enlarged lymph nodes in the lower neck bilaterally, the mediastinum and closely contiguous left hilar lymph nodes with standardized uptake values (SUVs) ranging from 4.2 to 18.9, but no disease outside the neck and thorax. The largest nodal mass is 4.5 cm in diameter. Peripheral blood cell counts, serum creatinine, and screening liver function tests including alkaline phosphatase are all normal except for a mildly elevated neutrophil count. She has received a recommendation that she be treated with six cycles of ABVD chemotherapy (Adriamycin [doxorubicin], bleomycin, vinblastine, and dacarbazine) followed by involved-field radiation if residual disease remains abnormal on PET/CT scanning. What is your recommendation?

A. Omit the chemotherapy and proceed with irradiation to the neck, axillae. and mediastinum using the mantle technique.
B. Shorten the ABVD chemotherapy to two cycles and follow that with mantle radiation.
C. Shorten the ABVD chemotherapy to two cycles and follow that with involved-field radiation.
D. Perform a PET/CT scan after two cycles of ABVD chemotherapy and stop treatment if it is negative.
E. Agree with the previous specialist's recommendation.

Answer: C Treatment with two cycles of ABVD followed by involved-field radiation has been shown be equivalent to more chemotherapy plus involved-field radiation. Answer B is incorrect because it is not necessary to use a larger radiation field, and the larger field would be expected to increase the risk for serious late complications. Answer D is incorrect because it has not been proved that radiation can be omitted based on a negative PET/CT scan after only two cycles of chemotherapy. The previous consultant recommended more chemotherapy than is necessary for limited-stage Hodgkin lymphoma.

3. An 18-year-old college student is referred to you to investigate new palpable inguinal lymphadenopathy. He has been experiencing drenching night sweats on at least 4 nights each week for the past month and has lost 7 kg of weight on a baseline weight of 84 kg. You document palpable abnormal lymph nodes in the left neck, left axilla, and bilateral inguinal/femoral areas. An excisional biopsy of a left supraclavicular fossa lymph node demonstrates nodular sclerosing Hodgkin lymphoma, and PET/CT scanning shows definite enlarged lymph nodes in the lower neck bilaterally, mediastinum, retroperitoneum, common iliac, and inguinal/femoral areas and there are multiple nodules within the spleen, with SUVs ranging from 3.2 to 28.9. The largest nodal mass is in the mediastinum and measures 15.5 cm in diameter. Peripheral blood cell counts show hemoglobin 9.2 g/L, white blood cell count 13.5×10^9/L, normal differential except for increased neutrophils, and normal serum creatinine. Alkaline phosphatase and lactate dehydrogenase are both twice the upper limit of normal. Bone marrow biopsy demonstrates involvement with Hodgkin lymphoma. What is your recommendation?

A. Initiate ABVD chemotherapy with the plan to continue through six cycles if the lymphoma shows continuous response through that time.
B. Give escalated BEACOPP chemotherapy (bleomycin, etoposide, Adriamycin, cyclophosphamide, vincristine, procarbazine, and prednisone) for eight cycles followed involved-field radiation.
C. Give four cycles of ABVD followed by high-dose chemotherapy and autologous peripheral blood stem cell transplantation.
D. Give two cycles of ABVD followed by six cycles of escalated BEACOPP chemotherapy if the PET/CT scan remains abnormal after the ABVD.
E. Give involved-field radiation followed by six cycles of ABVD.

Answer: A ABVD has emerged as the best chemotherapy overall because of its balance of high efficacy and moderate toxicity. Although escalated BEACOPP produces a higher initial response rate, overall survival is not improved compared with ABVD. Neither regimen must be followed by radiation if the chemotherapy induces a complete response as measured by PET/CT scanning. Planned escalation to high-dose chemotherapy and autologous peripheral blood stem cell transplantation has not been shown to improve overall survival for Hodgkin lymphoma. Altering treatment based on interim PET/CT scanning, although an attractive concept, has not been shown to improve ultimate disease control for Hodgkin lymphoma. It is never appropriate to initiate treatment of Hodgkin lymphoma with radiation because that will compromise safe delivery of necessary chemotherapy without improving outcome.

4. Two years ago, you treated a 47-year-old housewife with six cycles of ABVD for stage IIIA mixed-cellularity Hodgkin lymphoma, and the patient has been off treatment and well for 18 months. She now returns with new palpable bilateral neck lymph nodes, and a biopsy confirms recurrent Hodgkin lymphoma. Complete reassessment confirms recurrent disease in the neck and retroperitoneum. What is your recommendation?

A. Initiate MOPP/ABVD chemotherapy (mechlorethamine, vincristine, procarbazine, and prednisone alternating with ABVD) with the plan to continue through six cycles if the lymphoma shows continuous response through that time.
B. Give escalated BEACOPP chemotherapy (bleomycin, etoposide, Adriamycin, cyclophosphamide, vincristine, procarbazine, and prednisone) for six cycles.
C. Give mantle followed by inverted-Y extended-field radiation.
D. Give two cycles of ABVD followed by high-dose chemotherapy and autologous peripheral blood stem cell transplantation.
E. Give two cycles of cisplatin- or carboplatin-based multiagent chemotherapy followed by high-dose chemotherapy and autologous peripheral blood stem cell transplantation.

Answer: E Choices A, B and D are incorrect because repeating ABVD chemotherapy by itself or with addition of other agents is inappropriate since this patient's lymphoma has already demonstrated resistance to ABVD by relapsing. Although wide-field radiation may be able to cure some patients with lymph node–only Hodgkin lymphoma relapsing after ABVD, it is much less likely to do so than high-dose chemotherapy and autologous peripheral blood stem cell transplantation and is likely to contribute substantially to potentially lethal toxicity if treatment is necessary for another relapse. Two to three cycles of cisplatin- or carboplatin-based multiagent chemotherapy followed by high-dose chemotherapy and autologous peripheral blood stem cell transplantation is the treatment of choice for Hodgkin lymphoma that has relapsed after primary treatment with ABVD for advanced-stage disease.

5. Which of the following statements about Hodgkin lymphoma occurring in an individual known to harbor HIV infection is *true*?

A. Coincident HIV infection can be ignored when found in a patient with Hodgkin lymphoma.
B. HIV-associated Hodgkin lymphoma is incurable and thus best approached with palliative intent.
C. Stage IV is uncommon in patients with HIV-associated Hodgkin lymphoma.
D. Patients with HIV-associated Hodgkin lymphoma should be treated with standard ABVD.
E. Patients with HIV-related Hodgkin lymphoma are at high risk for central nervous system (CNS) involvement with the lymphoma.

Answer: D Although patients with HIV-associated Hodgkin lymphoma can and should be treated curatively with standard chemotherapy, such as ABVD (B is *false*; D is *true*), these patients require substantial special supportive care with antiviral, antifungal, and anti-*Pneumocystis* antibiotics, coincident treatment with highly active antiretroviral therapy (HAART) and often neutrophil growth factors (A is *false*). HIV-associated Hodgkin lymphoma causes stage IV disease much more often than non-HIV-associated disease (C is *false*). Although HIV-associated non-Hodgkin lymphoma is linked to a markedly increased risk for CNS involvement, the same has not been noted with Hodgkin lymphoma.

187

PLASMA CELL DISORDERS

S. VINCENT RAJKUMAR

Plasma cell disorders are neoplastic or potentially neoplastic diseases associated with the clonal proliferation of immunoglobulin-secreting plasma cells (Table 187-1). They are characterized by the secretion of electrophoretically and immunologically homogeneous (monoclonal) proteins that represent intact or incomplete immunoglobulin molecules. Monoclonal proteins are commonly referred to as M proteins, myeloma proteins, or paraproteins.

Syndromes associated with plasma cell disorders and monoclonal proteins include premalignant disorders (monoclonal gammopathy of undetermined significance, smoldering multiple myeloma), malignant neoplasms (multiple myeloma, Waldenström macroglobulinemia), and disorders primarily related to the unique properties of the secreted monoclonal protein (cryoglobulinemia, immunoglobulin light chain [AL] amyloidosis, light chain deposition disease) (see Table 187-1).[1]

Serum Immunoglobulins

Intact immunoglobulins consist of two heavy (H) polypeptide chains of the same class and subclass and two light (L) polypeptide chains of the same type (Chapter 45). The heavy polypeptide chains are designated by Greek letters: γ in immunoglobulin G (IgG), α in immunoglobulin A (IgA), μ in immunoglobulin M (IgM), δ in immunoglobulin D (IgD), and ε in immunoglobulin E (IgE). The light chain types are kappa (κ) and lambda (λ). Both heavy chains and light chains have constant and variable regions with respect to the amino acid sequence. The class specificity of each immunoglobulin is defined by a series of antigenic determinants on the constant regions of the heavy chains (γ, α, μ, δ, and ε) and the two major classes of light chains (κ and λ). The amino acid sequence in the variable regions of the immunoglobulin molecule corresponds to the active antigen-combining site of the antibody.

In the majority of clonal plasma cell disorders, *intact* immunoglobulin molecules are secreted as monoclonal (M) proteins. In addition, there can also be abnormal secretion of excess monoclonal *free* light chains that are released without being bound to immunoglobulin heavy chains. In some patients, heavy chain expression is completely lost, and only monoclonal free light chains (commonly referred to as Bence Jones proteins) are secreted. Even less frequently, only heavy chains are secreted, resulting in heavy chain diseases (HCDs). Rare patients with multiple myeloma secrete no identifiable immunoglobulin (nonsecretory myeloma).

TABLE 187-1 PLASMA CELL PROLIFERATIVE DISORDERS

I. Premalignant monoclonal gammopathies
  A. Monoclonal gammopathy of undetermined significance (MGUS)
  B. MGUS in association with chronic lymphocytic leukemia and non-Hodgkin lymphoma
  C. Biclonal and triclonal gammopathies of undetermined significance
  D. Idiopathic Bence Jones proteinuria and light chain MGUS
  E. Smoldering multiple myeloma
II. Malignant monoclonal gammopathies
  A. Multiple myeloma and related malignant neoplasms (IgG, IgA, IgD, IgE, and free light chains)
    1. Symptomatic multiple myeloma
    2. Plasma cell leukemia
    3. Osteosclerotic myeloma (including POEMS syndrome)
    4. Solitary plasmacytoma of bone
    5. Solitary extramedullary plasmacytoma
  B. Waldenström macroglobulinemia (IgM)
III. Heavy chain diseases (HCDs)
  A. γ-HCD
  A. α-HCD
  A. μ-HCD
IV. Cryoglobulinemia (types I, II, and III)
V. Immunoglobulin light chain amyloidosis

Ig = immunoglobulin; POEMS = polyneuropathy, organomegaly, endocrinopathy, M protein, and skin changes.

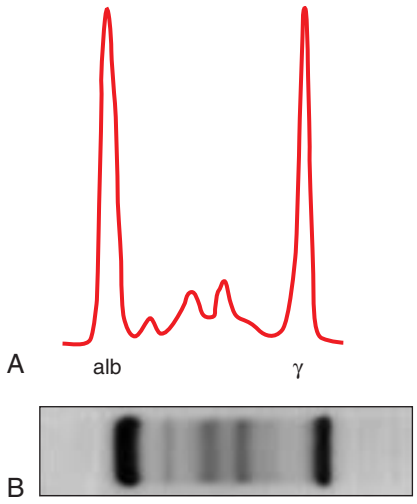

FIGURE 187-1. Serum protein electrophoresis showing a monoclonal (M) protein. **A,** Monoclonal pattern of serum protein as traced by a densitometer after electrophoresis on agarose gel: tall, narrow-based peak of γ mobility. **B,** Monoclonal pattern from electrophoresis of serum on agarose gel (anode on the left): dense, localized band representing monoclonal protein of γ mobility. (From Kyle RA, Katzmann JA. Immunochemical characterization of immunoglobulins. In: Rose NR, Conway de Macario E, Folds JD, et al, eds. *Manual of Clinical Laboratory Immunology*. 5th ed. Washington, DC: ASM Press; 1997:156, with permission of the American Society for Microbiology.)

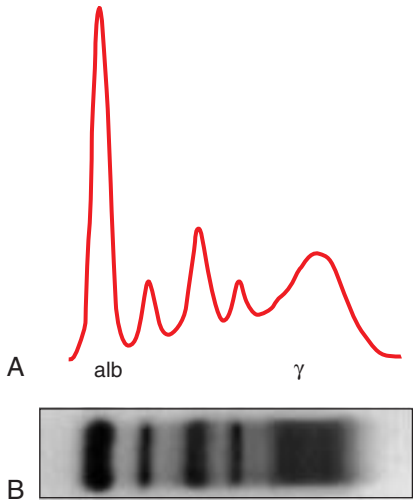

FIGURE 187-2. Serum protein electrophoresis showing increased polyclonal immunoglobulins. **A,** Polyclonal pattern from a densitometer tracing of agarose gel: broad-based peak of γ mobility. **B,** Polyclonal pattern from electrophoresis of agarose gel (anode on the left). The band at the right is broad and extends throughout the γ area. (From Kyle RA, Katzmann JA. Immunochemical characterization of immunoglobulins. In: Rose NR, Conway de Macario E, Folds JD, et al, eds. *Manual of Clinical Laboratory Immunology*. 5th ed. Washington, DC: ASM Press; 1997:156, with permission of the American Society for Microbiology.)

Identification of Monoclonal Proteins

Protein electrophoresis of the serum and urine detects M protein as a narrow peak (like a church spire) on the densitometer tracing or as a dense, discrete band on agarose gel (Fig. 187-1). Electrophoresis also permits quantitation of M proteins. Monoclonal light chains (Bence Jones proteinemia) are rarely seen on serum electrophoresis but are easily detected on urine electrophoresis. Urine electrophoresis requires a 24-hour urine collection.

Immunofixation of the serum and urine is performed when a peak or band is first seen on protein electrophoresis to identify the heavy and light chain types of the M protein. Immunofixation is also a more sensitive test than protein electrophoresis, and it should always be performed in conjunction with electrophoresis when multiple myeloma or related disorders are first suspected to detect small, unmeasurable M proteins that may be missed on electrophoresis. This is particularly important in oligosecretory myeloma, primary amyloidosis, and solitary plasmacytoma and after successful treatment of multiple myeloma or macroglobulinemia. In these instances, a small M protein can be concealed in the normal β or γ areas of the electrophoresis gel and may be overlooked.

Monoclonal proteins must be distinguished from an excess of polyclonal immunoglobulins (one or more heavy chain types and both κ and λ light chains, usually limited to the γ region), which produce a broad-based peak or broad band (Fig. 187-2). This finding is associated with chronic infectious or inflammatory states, including chronic liver disease.

Detection of Serum Free Light Chains

The serum free light chain assay measures the level of free κ and λ immunoglobulin light chains (i.e., light chains that are not bound to intact immunoglobulin). An abnormal κ/λ free light chain ratio (normal range, 0.26 to 1.65) indicates an excess of one light chain type versus the other and is interpreted as representing a monoclonal elevation of the corresponding light chain type. The serum free light chain assay is more sensitive than electrophoresis or immunofixation in detecting free monoclonal light chains and is useful in the diagnostic evaluation of plasma cell disorders and in risk stratification.

MONOCLONAL GAMMOPATHY OF UNDETERMINED SIGNIFICANCE

DEFINITION

Monoclonal gammopathy of undetermined significance (MGUS; formerly called benign monoclonal gammopathy) is a premalignant clonal plasma cell disorder characterized by the presence of a serum M protein in persons who lack evidence of multiple myeloma, macroglobulinemia, amyloidosis, or other related diseases. MGUS is defined by a serum M protein concentration lower than 3 g/dL; less than 10% clonal plasma cells in the bone marrow; and absence of lytic bone lesions, anemia, hypercalcemia, and renal insufficiency that can be attributed to a plasma cell disorder. The main clinical significance of MGUS is its lifelong risk of transformation to myeloma or related malignant disease at a fixed but unrelenting rate of 1% per year.

EPIDEMIOLOGY

More than 50% of patients in whom a serum M protein is detected have MGUS (Fig. 187-3). The prevalence of MGUS in the general population increases with age, from approximately 1% in persons 50 to 60 years old to more than 5% in those older than 70 years.[2] The age-adjusted prevalence is higher in men than in women and is twice as high in blacks compared with whites. There is an increased prevalence of MGUS as well as of multiple myeloma among blood relatives of individuals with monoclonal gammopathies.

PATHOBIOLOGY

MGUS represents a limited, nonmalignant expansion of monoclonal plasma cells. The etiology of MGUS is unknown, but age, male gender, family history, immunosuppression, and exposure to certain pesticides are known risk factors. It is hypothesized that infection, inflammation, or other antigenic stimuli, acting in concert with the development of cytogenetic abnormalities in the plasma cells, are the initiating pathogenetic events in most patients. Approximately 40% of MGUS is associated with plasma cell translocations involving the immunoglobulin heavy chain (IgH) locus on chromosome 14q32 (IgH-translocated MGUS), 40% with trisomies involving odd-numbered chromosomes (hyperdiploid MGUS), 15% with both trisomies and IgH translocations, and the remaining with other cytogenetic abnormalities. The primary IgH translocations seen in MGUS commonly involve one of five recurrent partner chromosome loci: 11q13 (*CCND1* [cyclin D1 gene]), 4p16.3 (*FGFR3* and *MMSET*), 6p21 (*CCND3* [cyclin D3 gene]), 16q23 (c-*maf*), and 20q11 (*mafB*).

CLINICAL MANIFESTATIONS

MGUS is asymptomatic and is usually diagnosed incidentally on laboratory testing. Patients with MGUS progress to multiple myeloma or related malignant disease at a rate of approximately 1% per year. The interval from the time of recognition of the M protein to the diagnosis of serious disease ranges from 1 to 32 years (median, 10.6 years), and the relative risk versus a control population is 25.0 for progression to multiple myeloma, 8.4 for primary amyloidosis, 46.0 for Waldenström macroglobulinemia, 2.4 for the development of other forms of non-Hodgkin lymphoma, and 8.5 for plasmacytoma.

DIAGNOSIS

MGUS is differentiated from multiple myeloma and smoldering multiple myeloma by the size of the M protein; the bone marrow plasma cell

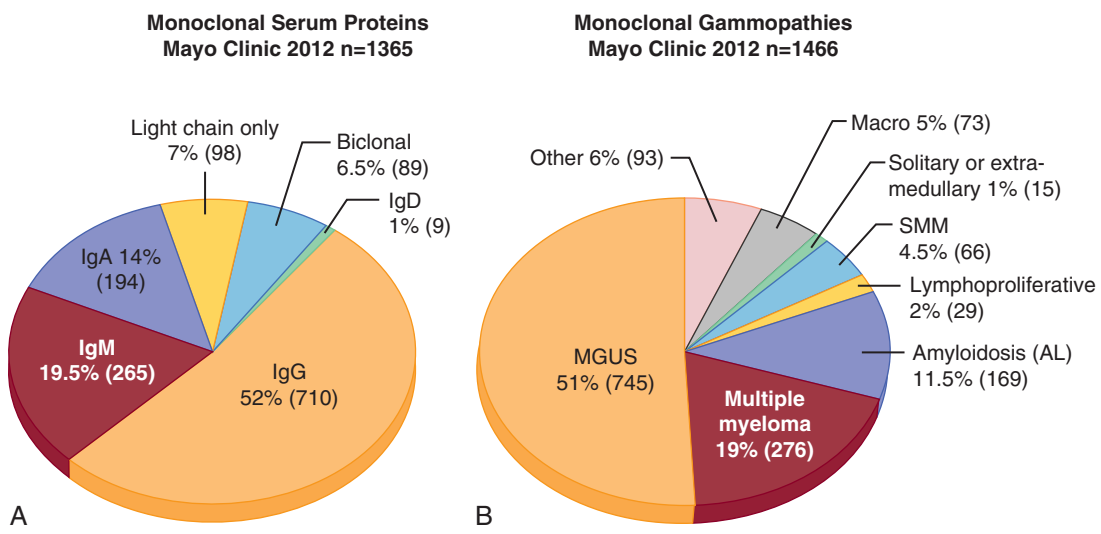

FIGURE 187-3. Monoclonal gammopathy. A, Distribution of serum monoclonal proteins in patients seen at the Mayo Clinic. **B,** Diagnoses in cases of monoclonal gammopathy seen at the Mayo Clinic. Ig = immunoglobulin; Macro = Waldenström macroglobulinemia; MGUS = monoclonal gammopathy of undetermined significance; SMM = smoldering multiple myeloma.

percentage; and the presence or absence of anemia, renal failure, hypercalcemia, or lytic bone lesions (Table 187-2). Because anemia and renal insufficiency are relatively common in the elderly population with MGUS, the causes of these conditions should be carefully investigated with adequate laboratory studies. For example, in a patient with anemia, tests to exclude iron, vitamin B_{12}, or folate deficiency must be performed. In certain instances, such as unexplained renal failure, a renal biopsy may be needed. Only patients with strong evidence of end-organ damage thought to be directly related to a plasma cell disorder can be considered to have myeloma or a related malignant disease.

Although a reduction in the levels of the immunoglobulin classes other than M protein (i.e., the normal polyclonal or background immunoglobulins) is more frequently seen in multiple myeloma or Waldenström macroglobulinemia, such reductions also occur in almost 40% of patients with MGUS.

Association of MGUS with Other Diseases

MGUS is associated with numerous diseases. However, because 3% of the general population older than 50 years has MGUS, it is often difficult to determine whether these reported associations are causal or coincidental. Some associations have been verified on the basis of epidemiologic studies; these include peripheral neuropathy (Chapter 420), proliferative glomerulonephritis, deep venous thrombosis (Chapter 81), osteoporosis (Chapter 243), and lymphoproliferative disorders (Chapter 185). A secondary form of MGUS also occurs with immunosuppression after organ transplantation (Chapter 49) and autologous or allogeneic stem cell transplantation (Chapter 178). M proteins also occur in the sera of some patients with chronic lymphocytic leukemia (Chapter 184) but have no recognizable effect on the clinical course.

Approximately 5% of patients with sensorimotor peripheral neuropathy of unknown cause (Chapter 420) have an associated monoclonal gammopathy (monoclonal gammopathy–associated neuropathy). In half of such patients, the M protein binds to myelin-associated glycoprotein. These patients have a slowly progressive sensory neuropathy more than motor neuropathy, beginning in the distal ends of the extremities and extending proximally. The clinical and electrodiagnostic manifestations of MGUS neuropathy resemble those of a chronic inflammatory demyelinating polyneuropathy. A causal relationship is usually assumed in younger patients and those without other conditions known to cause neuropathy in whom the neuropathy is severe and progressive. Therapeutic approaches include plasmapheresis and, occasionally, chemotherapy (similar to myeloma for IgG or IgA monoclonal proteins, and rituximab or rituximab-based regimens for IgM monoclonal proteins; see the later section on the treatment of multiple myeloma).

Monoclonal gammopathy is also thought to be the underlying cause of approximately 50% of idiopathic proliferative glomerulonephritis, including membranoproliferative glomerulonephritis and C3 glomerulopathy (Chapter 121). Certain skin disorders are also known to be associated with MGUS. Lichen myxedematosus (papular mucinosis, scleromyxedema) is associated with an IgG γ protein. Pyoderma gangrenosum (Chapter 440) and necrobiotic xanthogranuloma are other associated skin disorders.

PREVENTION AND TREATMENT

Rx

No treatment is necessary for MGUS. Low-risk patients (Table 187-3) can be evaluated when symptoms suggestive of myeloma or related disorders occur. In all other patients with MGUS, the M protein level in serum and urine should be measured serially, together with periodic reevaluation of clinical and other laboratory findings, to determine whether multiple myeloma or another related disorder is present. In general, electrophoresis, complete blood count, and creatinine and calcium levels should be repeated at 6 months and, if stable, yearly thereafter.

PROGNOSIS

Differentiating a patient with MGUS in whom the disorder will remain stable for life from one in whom multiple myeloma, macroglobulinemia, or a related disorder will eventually develop is difficult when the M protein is first recognized. The size and type of the M protein at diagnosis of MGUS and an abnormal serum free light chain ratio are prognostic factors for progression (see Table 187-3). A study of 728 Swedish cases of MGUS observed for up to 30 years showed a cumulative risk of 15.4% for development of lymphoid disorder and a cumulative risk of 10.6% for progression to multiple myeloma (approximately 0.5% annual risk). Three factors were significantly associated with progression: (1) abnormal free light chain ratio (<0.26 or >1.65); (2) M-protein level of 1.5 g/dL and higher; and (3) reduction of one or two noninvolved immunoglobulin isotype levels (immunoparesis).[3] The first two of these confirm the factors considered by the Mayo Clinic group (see Table 187-3).

BICLONAL GAMMOPATHIES

Biclonal gammopathies occur in at least 5% of patients with clonal plasma cell disorders. A biclonal gammopathy of undetermined significance (analogous to MGUS) accounts for about two thirds of such patients. The remainder have multiple myeloma, macroglobulinemia, or other lymphoproliferative diseases. Rarely, triclonal gammopathies may occur.

LIGHT CHAIN MGUS AND IDIOPATHIC BENCE JONES PROTEINURIA

The diagnosis of typical MGUS requires expression of an intact heavy chain type. In some patients, a premalignant clonal plasma cell disorder characterized by the presence of monoclonal immunoglobulin light chains without expression of heavy chains can occur (light chain MGUS).[4] By definition, these patients should not have evidence of end-organ damage attributable to the light chain, and the clonal bone marrow plasma cell percentage should be

TABLE 187-2 CRITERIA FOR THE DIAGNOSIS OF PLASMA CELL DISORDERS

DISORDER	DISEASE DEFINITION
Monoclonal gammopathy of undetermined significance (MGUS)	All 3 criteria must be met: 1. Serum monoclonal protein (IgG, IgA, or IgM) <3 g/dL 2. Clonal bone marrow plasma cells <10% 3. Absence of end-organ damage, such as hypercalcemia, renal insufficiency, anemia, and bone lesions, that can be attributed to the plasma cell proliferative disorder (or, in the case of IgM MGUS, no evidence of anemia, constitutional symptoms, hyperviscosity, lymphadenopathy, or hepatosplenomegaly that can be attributed to the underlying lymphoproliferative disorder)
Light chain MGUS	All 6 criteria must be met: 1. Abnormal FLC ratio (<0.26 or >1.65) 2. Increased level of the appropriate involved light chain (increased κ FLC in patients with ratio >1.65 and increased λ FLC in patients with ratio <0.26) 3. No immunoglobulin heavy chain expression on immunofixation 4. Absence of end-organ damage that can be attributed to the plasma cell proliferative disorder 5. Clonal bone marrow plasma cells <10% 6. Urinary monoclonal protein <500 mg/24h
Smoldering multiple myeloma (also referred to as asymptomatic multiple myeloma)	Both criteria must be met: 1. Serum monoclonal protein (IgG or IgA) ≥3 g/dL (or urinary monoclonal protein ≥500 mg/24 hours) and/or clonal bone marrow plasma cells 10-60% 2. Absence of myeloma defining events or amyloidosis

Multiple myeloma	Both criteria must be met:
	1. Clonal bone marrow plasma cells ≥10% or biopsy-proven bony or extramedullary plasmacytoma
	2. Any one or more of the following myeloma defining events:
	• Evidence of end organ damage that can be attributed to the underlying plasma cell proliferative disorder, specifically:
	  • Hypercalcemia: serum calcium >1 mg/dL (>0.25 mmol/L) higher than the upper limit of normal or >11 mg/dL (>2.75 mmol/L)
	  • Renal insufficiency: creatinine clearance <40 mL per minute or serum creatinine >2 mg/dL (>177 μmol/L)
	  • Anemia: hemoglobin value of >2 g/dL below the lower limit of normal, or a hemoglobin value <10 g/dL
	  • Bone lesions: one or more osteolytic lesions on skeletal radiography, CT, or PET-CT
	• Any one or more of the following biomarkers of malignancy:
	  • Clonal bone marrow plasma cell percentage ≥60%
	  • Involved: uninvolved serum free light chain ratio ≥100 (involved free light chain level must be ≥100 mg/L)
	  • >1 focal lesions on MRI studies (at least 1mm in size)

Waldenström macroglobulinemia	Both criteria must be met: 1. IgM monoclonal gammopathy (regardless of the size of the M protein) 2. >10% bone marrow lymphoplasmacytic infiltration (usually intertrabecular) by small lymphocytes that exhibit plasmacytoid or plasma cell differentiation and a typical immunophenotype (surface IgM^{+}, $CD5^{+/-}$, $CD10^{-}$, $CD19^{+}$, $CD20^{+}$, $CD23^{-}$) that satisfactorily excludes other lymphoproliferative disorders, including chronic lymphocytic leukemia and mantle cell lymphoma
Smoldering Waldenström macroglobulinemia (also referred to as indolent or asymptomatic Waldenström macroglobulinemia)	Both criteria must be met: 1. Serum IgM monoclonal protein ≥3 g/dL and/or bone marrow lymphoplasmacytic infiltration ≥10% 2. No evidence of end-organ damage, such as anemia, constitutional symptoms, hyperviscosity, lymphadenopathy, or hepatosplenomegaly, that can be attributed to a lymphoplasma cell proliferative disorder
Solitary plasmacytoma	All 4 criteria must be met: 1. Biopsy-proven solitary lesion of bone or soft tissue with evidence of clonal plasma cells 2. Normal bone marrow with no evidence of clonal plasma cells 3. Normal skeletal survey and either MRI of spine and pelvis or PET computed tomography (except for the primary solitary lesion) 4. Absence of end-organ damage, such as hypercalcemia, renal insufficiency, anemia, or bone lesions, that can be attributed to a plasma cell proliferative disorder
POEMS syndrome	All 4 criteria must be met: 1. Presence of a monoclonal plasma cell disorder (almost always λ type) 2. Peripheral neuropathy 3. Any one of the following 3 major features: sclerotic bone lesions, Castleman disease, elevated levels of vascular endothelial growth factor 4. Any one of the following 7 minor features: organomegaly, edema, endocrinopathy (excluding diabetes mellitus or hypothyroidism), typical skin changes, papilledema, thrombocytosis, polycythemia The features should have a temporal relationship to one another, with no other attributable cause

CT = computed tomography; FLC = free light chain; MRI = magnetic resonance imaging; PET = positron emission tomography.
Derived from Rajkumar SV, et al. International Myeloma Working Group updated Criteria for the diagnosis of multiple myeloma. *Lancet Oncology* 2014;15:e538-e548; and Kyle RA, Rajkumar SV. Criteria for diagnosis, staging, risk stratification and response assessment of multiple myeloma. *Leukemia.* 2009;23:3-9.

TABLE 187-3 RISK OF PROGRESSION OF MONOCLONAL GAMMOPATHY OF UNDETERMINED SIGNIFICANCE TO MYELOMA OR RELATED DISORDERS

RISK GROUP	RELATIVE RISK	CUMULATIVE ABSOLUTE RISK OF PROGRESSION AT 20 YEARS (%)*	CUMULATIVE ABSOLUTE RISK OF PROGRESSION AT 20 YEARS ACCOUNTING FOR DEATH AS A COMPETING RISK (%)†
Low risk: serum M protein <1.5 g/dL, IgG subtype, normal free light chain ratio (0.26-1.65)	1	5	2
Low-intermediate risk: any 1 factor abnormal	5.4	21	10
High-intermediate risk: any 2 factors abnormal	10.1	37	18
High risk: all 3 factors abnormal	20.8	58	27

*Estimates in this column represent the risk of progression assuming that patients do not die of other causes during this period.
†Estimates in this column represent the risk of progression calculated by use of a model that accounts for the fact that patients can die of unrelated causes during this time.
Ig = immunoglobulin.
Modified from Rajkumar SV, Kyle RA, Therneau TM, et al. Serum free light chain ratio is an independent risk factor for progression in monoclonal gammopathy of undetermined significance (MGUS). *Blood.* 2005;106:812-817. © The American Society of Hematology.
Also see reference 3.

less than 10%. No therapy is indicated unless progression to malignancy occurs.

MULTIPLE MYELOMA

DEFINITION

Multiple myeloma is a malignant neoplasm of plasma cells characterized by bone marrow infiltration and extensive skeletal destruction resulting in anemia, bone pain, and fractures. Multiple myeloma (commonly referred to as myeloma) is defined by the presence of 10% or more clonal plasma cells on bone marrow examination or biopsy-proven plasmacytoma; and evidence of one or more myeloma defining events (see Table 187-2).[5] Patients with multiple myeloma must be differentiated from those with MGUS and smoldering multiple myeloma.

EPIDEMIOLOGY

Multiple myeloma accounts for 1% of all malignant disease and slightly more than 10% of hematologic malignant neoplasms in the United States. The annual incidence of multiple myeloma is 4 per 100,000. Its incidence in blacks is almost twice that in whites. Multiple myeloma is slightly more common in men than in women. The median age of patients at the time of diagnosis is about 65 years; only 2% of patients are younger than 40 years.

PATHOBIOLOGY

The cause of multiple myeloma is unclear. Exposure to radiation, benzene, and other organic solvents, herbicides, and insecticides may play a role. Multiple myeloma has been reported in familial clusters of two or more first-degree relatives and in identical twins.

Almost all cases of myeloma evolve from a premalignant MGUS phase, although the MGUS is clinically recognized before the diagnosis of myeloma in only a small minority of patients. The progression of MGUS to myeloma suggests a simple, random, two-hit genetic model of malignant transformation in which the risk of progression is fixed (approximately 1% per year) regardless of the duration of MGUS. Unfortunately, the precise mechanisms of progression are unknown, although several potentially pathogenetic abnormalities have been described in the clonal plasma cells. These include *RAS* and *p53* mutations, p16 methylation, *MYC* abnormalities, and secondary translocations. Changes in the bone marrow microenvironment may also play a role in the pathogenesis, including induction of angiogenesis and abnormal paracrine loops involving cytokines such as interleukin (IL)–6, which serves as a major growth factor for plasma cells.

The lytic bone lesions, osteopenia, hypercalcemia, and pathologic fractures in patients with myeloma are a result of abnormal osteoclast activity induced by the neoplastic plasma cells as well as inhibition of osteoblast differentiation. Osteoclasts are activated by stimulation of the transmembrane receptor RANK (receptor activator of nuclear factor κB), which belongs to the tumor necrosis factor receptor superfamily. The ligand for this receptor (RANKL) also has a decoy receptor, osteoprotegerin (OPG). In myeloma, there is an increase in RANKL expression by osteoblasts (and possibly plasma cells), accompanied by a reduction in the level of OPG. The resultant increase in the RANKL/OPG ratio causes osteoclast activation and increased bone resorption and turnover (Chapter 242). Other factors that may play a role in osteoclast activation include increased levels of macrophage inflammatory protein 1α, stromal cell–derived factor α, IL-3, IL-1β, and IL-6. In addition to these changes that promote osteoclast activation, there is simultaneous suppression of osteoblasts mediated by increased levels of IL-3, IL-7, and dickkopf 1 (DKK1). This combination leads to the pure osteolytic bone disease that is the hallmark of multiple myeloma.

Cytogenetic Abnormalities

As discussed earlier (see pathobiology of MGUS), primary translocations involving the IgH loci (chromosome 14q32) are seen in up to 40% of patients with multiple myeloma (IgH-translocated or nonhyperdiploid myeloma). Approximately 40% of patients do not have IgH translocations but have evidence of trisomies (hyperdiploid myeloma), 15% have both IgH translocations and trisomies, and 5% have other abnormalities. Although primary IgH translocations and trisomies originate at the MGUS stage, response to therapy and prognosis of myeloma are affected by the specific underlying abnormality (Table 187-4). Besides these cytogenetic abnormalities, other secondary cytogenetic abnormalities occur as late events during the course of symptomatic myeloma; these include activating mutations of N- and K-*RAS*, inactivating mutations of *p53*, and dysregulation of c-*MYC*. Complete or partial deletions of chromosome 13 are well described in myeloma and have prognostic value, but they also occur at the MGUS stage.

TABLE 187-4 STAGING AND PROGNOSTIC FACTORS IN MULTIPLE MYELOMA

STAGE/RISK FACTOR	MEDIAN SURVIVAL
INTERNATIONAL STAGING SYSTEM	
Stage I (serum β_2-microglobulin <3.5 mg/L and serum albumin ≥3.5 g/dL)	62 months
Stage II (neither stage I nor stage III)	44 months
Stage III (serum β_2-microglobulin ≥5.5 mg/L)	29 months
RISK STRATIFICATION*	
High-risk myeloma (any one of the following in the absence of trisomies): Translocations t(14;16), t(14;20) Deletion 17p	24-36 months
Intermediate-risk myeloma Translocation t4;14	Similar to standard-risk myeloma with bortezomib-based induction, transplantation, and maintenance
Standard-risk myeloma Translocations t(11;14), t(6;14) Trisomies	84-120 months
OTHER ADVERSE PROGNOSTIC FACTORS	
Elevated lactate dehydrogenase level	
Poor performance status	
Increased circulating plasma cells	
Plasmablastic morphology	
Increased plasma cell labeling index ≥1%	

*Typically detected in clonal plasma cells by fluorescence in situ hybridization of plasma cells.

TABLE 187-5 MAJOR CLINICAL MANIFESTATIONS OF MULTIPLE MYELOMA

CLINICAL FINDINGS	APPROXIMATE PERCENTAGE OF PATIENTS WITH ABNORMALITY AT DIAGNOSIS
Skeletal involvement: pain, reduced height, lytic bone lesions, pathologic fractures	80
Anemia (hemoglobin ≤12 g/dL): caused mainly by decreased erythropoiesis; produces weakness and fatigue	75
Renal insufficiency (serum creatinine ≥2 mg/dL): caused mainly by "myeloma kidney" from light chains or hypercalcemia, rarely from amyloidosis	20
Hypercalcemia (≥11 mg/dL)	15
Light chain amyloidosis	10
Evidence of monoclonal protein by immunofixation and serum free light chain assay	97
Evidence of clonal plasma cells ≥10% in bone marrow	96

CLINICAL MANIFESTATIONS

History

Bone pain, particularly in the back or chest and less often in the extremities, is present at the time of diagnosis in more than two thirds of patients (Table 187-5). The patient's height may be reduced by several inches because of vertebral collapse. Weakness and fatigue are common and are often associated with anemia. Fever is rare and, when present, is generally from an infection; in some patients, the infection itself is the initial feature. Other symptoms may result from renal insufficiency, hypercalcemia, nephrotic syndrome, radiculopathy, or amyloidosis (Chapter 188).

Physical Examination

Pallor is the most frequent physical finding. The liver is palpable in about 5% of patients and the spleen in 1%. Tenderness may be noted at sites of bone involvement. Radiculopathy may be caused by spinal compression fractures. On occasion, extramedullary plasmacytomas are palpable.

DIAGNOSIS

Laboratory Findings

A normocytic, normochromic anemia (Chapter 158) is present initially in approximately 75% of patients, but it eventually occurs in nearly every patient with multiple myeloma. Serum protein electrophoresis shows an M protein in 80% of patients. With serum immunofixation, an M protein can be detected in 93% of patients. When these serum studies are combined with urine electrophoresis plus immunofixation, an M protein can be detected in 97% of patients with myeloma. The serum free light chain assay is more convenient and can be used in place of urine studies in the diagnostic evaluation. The type of M protein is IgG in 52%, IgA in 21%, light chain only (Bence Jones proteinemia) in 16%, IgD in 2%, and biclonal gammopathy in 2%; the light chain type is κ in 65% of cases and λ in 35%. In 3% of patients, no secreted M protein can be identified; these patients are considered to have nonsecretory myeloma.

In the bone marrow, clonal plasma cells account for more than 10% of all nucleated cells in 96% of patients (Fig. 187-4). In 4% of patients, bone marrow examination shows less than 10% plasma cells, even though the patient otherwise meets the criteria for myeloma; because bone marrow involvement in myeloma may be focal rather than diffuse, repeated bone marrow examinations or biopsy of a discrete bone or extramedullary lesion may be required. In most cases, the plasma cells in myeloma are cytoplasmic Ig^+, $CD38^+$, $CD45^-$, $CD138^+$, $CD56^+$, and $CD19^+$; only a minority express CD10 and HLA-DR, and 20% express CD20. The clonality of the plasma cells is established by the κ/λ ratio, which is abnormal in myeloma (either >4 : 1, indicating a clonal κ population, or <1 : 2, indicating a clonal λ population). This is helpful for differentiation of monoclonal plasma cell proliferation in multiple myeloma from reactive plasmacytosis related to connective tissue disease, metastatic carcinoma, liver disease, and infection.

Radiologic Findings

Conventional radiographs reveal abnormalities consisting of punched-out lytic lesions (Fig. 187-5), osteoporosis, or fractures in nearly 80% of patients. The vertebrae, skull, thoracic cage, pelvis, and proximal ends of the humerus and femur are the most frequent sites of involvement. Technetium Tc99m bone scanning is inferior to conventional radiography and should not be used. Positron emission tomography (Fig. 187-6) and magnetic resonance imaging are increasingly used to evaluate patients in whom there is doubt about the magnitude of the disease burden, in those who have skeletal pain but no abnormality on radiographs, and for monitoring of the response to therapy.

Organ Involvement

Renal

At diagnosis, the serum creatinine value is increased initially in almost half of patients and is more than 2 mg/dL in 20%.

The two major causes of renal insufficiency are light chain cast nephropathy (*myeloma kidney*) and hypercalcemia. Light chain cast nephropathy is characterized by the presence of large, waxy, laminated casts in the distal and collecting tubules. The casts are composed mainly of precipitated monoclonal light chains. The extent of cast formation correlates directly with the amount of free urinary light chain and with the severity of renal insufficiency. Dehydration may precipitate acute renal failure.

Hypercalcemia (Chapter 245), which is present in 15 to 20% of patients initially, is a major and treatable cause of renal insufficiency. It results from destruction of bone. Hyperuricemia may contribute to renal failure. Besides light chain cast nephropathy and hypercalcemia, there are other mechanisms by which renal dysfunction can occur in myeloma. For example, light chain amyloidosis (Chapter 188) occurs in nearly 10% of patients and may produce nephrotic syndrome, renal insufficiency, or both. Acquired Fanconi syndrome (Chapter 122), characterized by proximal tubular dysfunction, results in glycosuria, phosphaturia, and aminoaciduria. Deposition of monoclonal light chains in the renal glomerulus (light chain deposition disease) may also produce renal insufficiency and nephrotic syndrome.

Neurologic

Radiculopathy (Chapter 400), the single most frequent neurologic complication, usually occurs in the thoracic or lumbosacral area and results from compression of the nerve by the vertebral lesion or by the collapsed bone itself. Compression of the spinal cord occurs in up to 10% of patients. Peripheral neuropathy (Chapter 420) is uncommon in multiple myeloma and, when present, is generally caused by amyloidosis. Rarely, myeloma cells diffusely infiltrate the meninges. Intracranial plasmacytomas almost always represent extensions of myelomatous lesions of the skull.

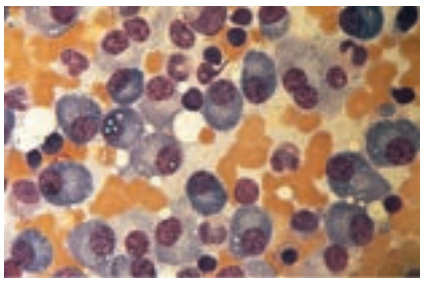

FIGURE 187-4. Multiple myeloma. A bone marrow aspirate shows a predominance of plasma cells.

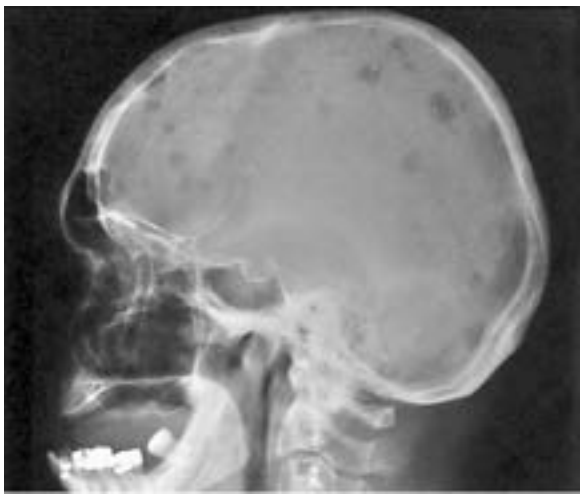

FIGURE 187-5. Skull radiograph of a patient with multiple myeloma showing multiple lytic lesions.

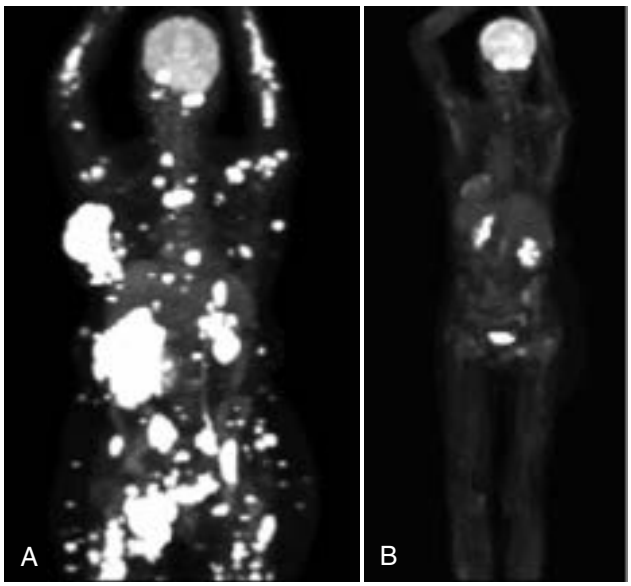

FIGURE 187-6. Positron emission tomography in multiple myeloma. A, Extensive bone and extramedullary disease. **B,** Significant improvement after systemic chemotherapy for myeloma.

Other Systemic Involvement

Hepatomegaly from plasma cell infiltration is uncommon. Plasmacytomas of the ribs are common and arise either as expanding bone lesions or as soft tissue masses. The incidence of infections is increased in patients with multiple myeloma. Historically, *Streptococcus pneumoniae* and *Staphylococcus aureus* have been the most frequent pathogens, but gram-negative organisms now account for more than half of all infections. The propensity for infection results from impairment of the antibody response, deficiency of normal immunoglobulins, and neutropenia. Bleeding from coating of the platelets by M protein may occur. Myeloma patients have an increased risk of deep venous thrombosis, particularly in relation to its therapy (see later).

PREVENTION AND TREATMENT

Rx

Patients with MGUS or smoldering multiple myeloma should not be treated until evidence of multiple myeloma develops. The approach to treatment of multiple myeloma is illustrated in Figure 187-7.

Initial Therapy for Patients Who Are Candidates for Autologous Stem Cell Transplantation

In the approximately 50% of patients with newly diagnosed multiple myeloma who are considered candidates for autologous stem cell transplantation on the basis of good performance status, no or limited comorbid conditions, and younger physiologic age (<65 to 70 years), autologous peripheral blood stem cell transplantation (Chapter 178) with high-dose chemotherapy improves overall survival in comparison to conventional chemotherapy.[■] Currently, it is not possible to eradicate myeloma cells completely with conditioning regimens, and reinfused autologous stem cells are usually contaminated by myeloma cells or their precursors. As a result, autologous transplantation is not curative, but it prolongs event-free and overall survival.

Initial therapy for stem cell transplant candidates typically consists of a non–melphalan-containing induction regimen for approximately 4 months followed by the stem cell collection. Most modern induction regimens have not been compared against each other in randomized trials, and the choice of regimen is dependent on availability and costs.[6] Common induction regimens include lenalidomide plus low-dose dexamethasone (Rd); bortezomib, thalidomide, plus dexamethasone (VTD); bortezomib, lenalidomide, plus dexamethasone (VRD); and bortezomib, cyclophosphamide, plus dexamethasone (VCD) (Table 187-6). VTD is associated with superior response rates and progression-free survival compared with thalidomide-dexamethasone (TD).[■] In a randomized trial, lenalidomide plus low-dose dexamethasone (40 mg once a week) was associated with superior overall survival compared with lenalidomide and high-dose dexamethasone (40 mg on days 1 to 4, 9 to 12, and 17 to 20). As a result, high-dose pulse dexamethasone is no longer recommended in the context of initial therapy.[■] Toxicities of lenalidomide include deep venous thrombosis, and all patients must be treated with prophylactic aspirin or an anticoagulant.

After induction therapy, peripheral blood stem cells adequate for one or two stem cell transplants are collected with the use of granulocyte colony-stimulating factor, with or without plerixafor or cyclophosphamide to aid in mobilization. Autologous stem cell transplantation (Chapter 178) is performed with melphalan 200 mg/m^2 as the conditioning regimen, followed by infusion of the peripheral blood stem cells. Patients who do not achieve a complete or very good partial response with the first autologous transplant can be considered for a second autologous transplant.[■]

An alternative approach in patients with newly diagnosed disease is to cryopreserve stem cells for future use after initial therapy. Patients then continue initial therapy, such as lenalidomide plus low-dose dexamethasone, until progression or achievement of a plateau phase, with stem cell transplantation reserved for the first relapse. Data from randomized trials comparing early versus delayed transplantation indicate no significant difference in survival between the two strategies. The choice is based on the patient's preferences and other clinical conditions, but early transplantation is often preferred

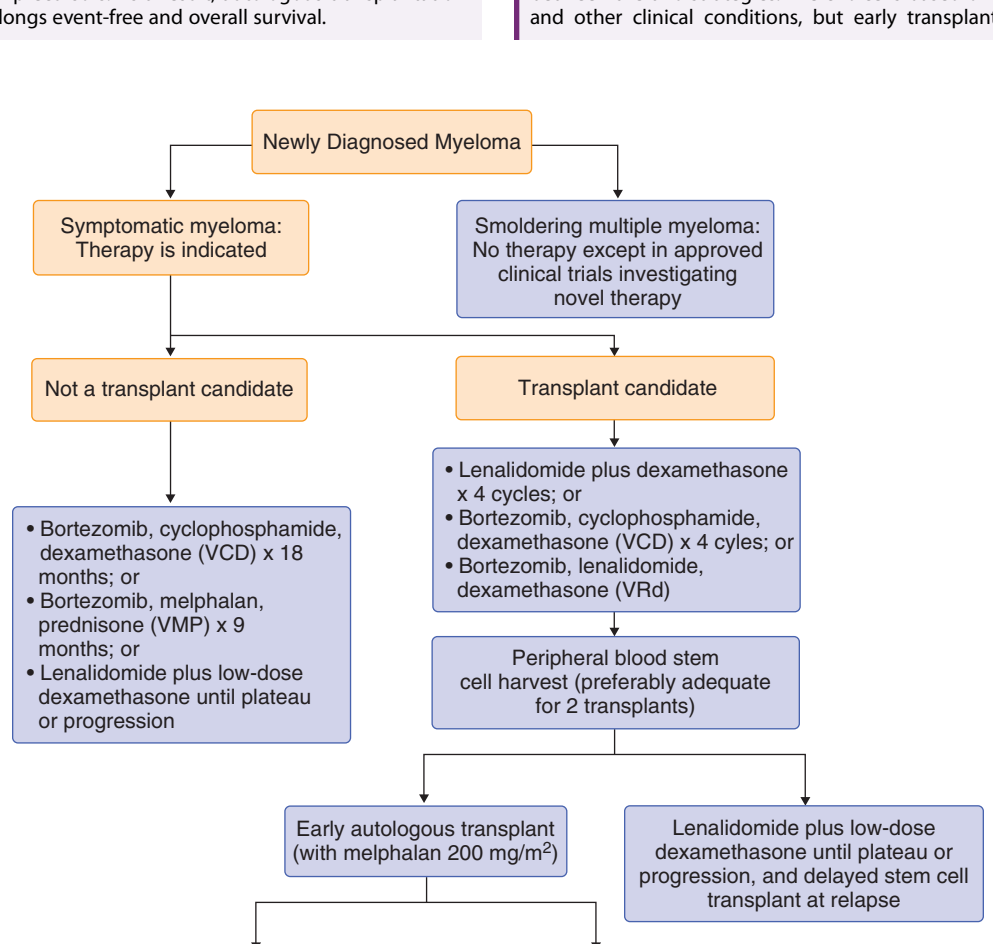

FIGURE 187-7. Therapeutic approach to newly diagnosed multiple myeloma.

TABLE 187-6 TREATMENT REGIMENS IN NEWLY DIAGNOSED MULTIPLE MYELOMA

REGIMEN	SUGGESTED STARTING DOSES*	OVERALL RESPONSE RATE (%)
REGIMENS FOR TRANSPLANT-ELIGIBLE AND TRANSPLANT-INELIGIBLE PATIENTS		
Lenalidomide-dexamethasone (Rd)	Lenalidomide, 25 mg orally, on days 1-21 every 28 days Dexamethasone, 40 mg orally, on days 1, 8, 15, 22 every 28 days Repeated every 4 weeks	70
Bortezomib-thalidomide-dexamethasone* (VTD)	Bortezomib, 1.3 mg/m^2 IV, on days 1, 8, 15, 22 Thalidomide, 100-200 mg orally, on days 1-21 Dexamethasone, 20 mg on day of/after bortezomib (or 40 mg on days 1, 8, 15, 22) Repeated every 4 weeks	95
Bortezomib-cyclophosphamide-dexamethasone* (VCD)	Cyclophosphamide, 300 mg/m^2 orally, on days 1, 8, 15 and 22 Bortezomib, 1.3 mg/m^2 IV, on days 1, 8, 15, 22 Dexamethasone, 40 mg orally, on days 1, 8, 15, 22 Repeated every 4 weeks	90
Bortezomib-lenalidomide-dexamethasone* (VRD)	Bortezomib, 1.3 mg/m^2 IV, on days 1, 8, 15 Lenalidomide, 25 mg orally, on days 1-14 Dexamethasone, 20 mg on day of and day after bortezomib (or 40 mg on days 1, 8, 15, 22) Repeated every 3 weeks	100
REGIMENS FOR TRANSPLANT-INELIGIBLE PATIENTS		
Melphalan-prednisone-thalidomide (MPT)	Melphalan, 0.25 mg/kg orally, on days 1-4 (use 0.20 mg/kg/day orally on days 1-4 in patients older than 75 years) Prednisone, 2 mg/kg orally, on days 1-4 Thalidomide, 100-200 mg orally, on days 1-28 (use 100-mg dose in patients >75 years) Repeated every 6 weeks	70
Bortezomib-melphalan-prednisone* (VMP)	Bortezomib, 1.3 mg/m^2 IV, on days 1, 8, 15, 22 Melphalan, 9 mg/m^2 orally, on days 1-4 Prednisone, 60 mg/m^2 orally, on days 1-4 Repeated every 35 days	70

*Doses of dexamethasone and bortezomib reduced from initial trial reports to once-weekly schedules.
Reproduced from Rajkumar SV. Treatment of multiple myeloma. *Nat Rev Clin Oncol.* 2011;8:479-491.

because its mortality is low (<1%), and it avoids the inconvenience, cost, and potential side effects of prolonged chemotherapy.

After stem cell transplantation, a short course of bortezomib administered as consolidation has been shown to improve response rates and progression-free survival. Similarly, studies suggest that long-term outcome may be improved by the administration of prolonged maintenance therapy after autologous stem cell transplantation. In randomized trials, lenalidomide maintenance (10 mg/day for the first 3 months, increased to 15 mg if tolerated) significantly prolongs progression-free survival but has the potential for more toxicity and second cancers. There are emerging data that bortezomib maintenance administered every 2 weeks may also provide a similar benefit. At present, the routine use of consolidation and maintenance in all patients after transplantation remains controversial because of lack of clear overall survival benefit and concerns about toxicity, cost, and impact on quality of life. Consolidation and maintenance should be considered, however, in intermediate- and high-risk myeloma (bortezomib maintenance preferred) and in patients not achieving a very good partial response or better with transplantation (lenalidomide maintenance preferred) (see Table 187-4 for definitions of intermediate- and high-risk myeloma).

Role of Allogeneic Bone Marrow Transplantation

Most patients with multiple myeloma cannot undergo allogeneic bone marrow transplantation because of their age, lack of an HLA-matched sibling donor, or inadequate renal, pulmonary, or cardiac function (Chapter 178). There are no clear data showing the benefit of either conventional myeloablative allogeneic transplantation or nonmyeloablative (mini) allogeneic transplantation compared with autologous stem cell transplantation, and results of randomized trials are conflicting. The treatment-related mortality is approximately 20%. Allogeneic transplantation for myeloma is best performed in the context of clinical trials or as second-line salvage therapy in selected high-risk patients who are willing to accept the high treatment-related mortality rate associated with the procedure.

Initial Therapy for Patients Who Are Not Candidates for Transplantation

Approximately 50% of newly diagnosed patients are not considered candidates for stem cell transplantation because of advanced age, poor performance status, or associated comorbidities. For decades, the oral administration of melphalan and prednisone was the standard of care. Randomized trials have shown that the addition of thalidomide or bortezomib to the standard regimen of melphalan plus prednisone improves event-free and overall survival compared with melphalan plus prednisone alone in patients with newly diagnosed myeloma who are not candidates for transplantation. On the basis of these data, melphalan and prednisone plus either thalidomide (MPT) or bortezomib (VMP) are two treatments for this population of patients. More recently, non–melphalan-containing regimens such as Rd, VRD, and VCD used in patients who are candidates for stem cell transplantation are being increasingly preferred over melphalan-based regimens in this group of patients as well. In a large randomized trial, Rd administered until progression was associated with superior progression-free and overall survival compared with MPT. By contrast, the addition of lenalidomide to melphalan and prednisone does not improve overall survival and is not recommended.

Regimens such as VCD, VRD, MPT, and VMP are typically given for approximately 12-18 months. Rd can be given until progression or for approximately 18 months, based on tolerability.

Treatment of Relapsed Refractory Myeloma

Almost all patients with multiple myeloma eventually relapse. Single-agent dexamethasone, alkylating agents, thalidomide, lenalidomide, and bortezomib, administered alone or in combination, are options for the treatment of relapsed refractory myeloma. Methylprednisolone, 2 g three times a week intravenously for a minimum of 4 weeks, then reduced to once or twice a week if there is a response, is helpful for patients with pancytopenia and may be associated with fewer side effects than with dexamethasone.

Thalidomide (50 to 200 mg/day orally) produces an objective response, with a median duration of about 1 year, in about a third of patients with refractory myeloma. Side effects are sedation, constipation, peripheral neuropathy, rash, bradycardia, and thrombotic events. The addition of dexamethasone to thalidomide increases the response rate to approximately 50%, and combinations of thalidomide, dexamethasone, and alkylating agents produce response rates exceeding 70% in patients with relapsed refractory disease.

Lenalidomide, an analogue of thalidomide, is better tolerated and produces objective benefits in approximately 40% of patients with relapsed refractory myeloma as a single agent; in combination with dexamethasone, 60% of patients benefit. Lenalidomide plus dexamethasone significantly prolongs time to progression and overall survival compared with dexamethasone alone. The starting dose of lenalidomide is 25 mg orally on days 1 to 21, every 28 days. Lenalidomide has significantly fewer nonhematologic toxicities than thalidomide does; myelosuppression is the most common adverse event.

Bortezomib, an inhibitor of the ubiquitin-proteasome pathway, acts through multiple mechanisms to arrest tumor growth, tumor spread, and angiogenesis. It produces objective responses in about a third of patients with refractory myeloma and is superior to single-agent dexamethasone. Bortezomib is usually combined with dexamethasone and other active agents (e.g., lenalidomide, thalidomide, or cyclophosphamide) to increase response rates. The usual dose is 1.3 mg/m^2 administered subcutaneously on days 1, 8, 15, and 22

every 28 days. The once-weekly subcutaneous dosing is associated with significantly lower neuropathy than the twice-weekly intravenous schedule. The most common adverse events are gastrointestinal side effects, fatigue, and neuropathy.

Options for the treatment of patients with myeloma refractory to lenalidomide and bortezomib include pomalidomide (an analogue of lenalidomide) and carfilzomib (a novel keto-epoxide tetrapeptide proteasome inhibitor).[7,8] Pomalidomide and carfilzomib have a response rate of approximately 25% in this population of patients and can be combined with other active agents to improve response rates. In a randomized trial, the addition of carfilzomib to lenalidomide and dexamethasone significantly improved progression-free survival.

Patients with relapsed refractory myeloma should also be considered for clinical trials. Promising investigational agents with single-agent activity include MLN 9708 (an oral proteasome inhibitor), marizomib (proteasome inhibitor), ARRY-520 (kinesin spindle protein inhibitor), monoclonal antibodies to CD38, and cyclin-dependent kinase inhibitors. Additional agents with potential activity in combination with standard anti-myeloma agents include panobinostat (histone deacetylase inhibitor) and elotuzumab (an anti–CS-1 antibody).

Role of Radiation Therapy

Palliative radiation in a dose of 20 to 30 Gy should be limited to patients who have multiple myeloma with disabling pain and a well-defined focal process that has not responded to chemotherapy and to patients with spinal cord compression from a plasmacytoma. Analgesics in combination with chemotherapy can usually control the pain (Chapter 30).

Management of Complications

Hypercalcemia

Hypercalcemia, present in 15 to 20% of patients at diagnosis, should be suspected in those with anorexia, nausea, vomiting, polyuria, polydipsia, constipation, weakness, confusion, or stupor. If hypercalcemia is untreated, renal insufficiency may develop. Hydration, preferably with isotonic saline plus prednisone (25 mg four times/day), usually relieves the hypercalcemia. Bisphosphonates, such as zoledronic acid or pamidronate, are recommended and will correct hypercalcemia in almost all patients (Chapter 243).

Renal Insufficiency

The most common cause of acute renal failure is light chain cast nephropathy in patients who have excess excretion of monoclonal protein in urine (myeloma kidney). Aggressive treatment of acute renal failure due to light chain cast nephropathy is critical for long-term overall survival. If the patient is not oliguric, intravenous fluids and furosemide are needed to maintain a high urine flow rate (100 mL/hour). If the underlying cause is thought to be light chain cast nephropathy on the basis of clinical findings (e.g., serum free light chains >150 mg/dL) or renal biopsy, plasmapheresis is recommended daily for 5 days to reduce the levels of circulating light chains. Hemodialysis is necessary for symptomatic azotemia. The mainstay of therapy is aggressive treatment of myeloma with a regimen such as bortezomib, thalidomide, and dexamethasone (VTD) or bortezomib, cyclophosphamide, and dexamethasone (VCD). Allopurinol is necessary if hyperuricemia is present.

Infection

Prompt, appropriate therapy for bacterial infections is necessary. Prophylactic antibiotics, such as trimethoprim-sulfamethoxazole, should be considered in patients taking high-dose corticosteroids. Acyclovir should be given as prophylaxis against herpes zoster in patients receiving bortezomib. Intravenously administered gamma globulin is reserved for patients with hypogammaglobulinemia and recurrent severe infections. Pneumococcal and influenza immunizations (Chapter 18) should be given to all patients.

Skeletal Lesions

Patients should be encouraged to be as active as possible but to avoid trauma. Pamidronate (90 mg infused intravenously during a 4-hour period every 4 weeks) or zoledronic acid (4 mg intravenously during at least 15 minutes every 4 weeks) reduces the incidence of bone pain, pathologic fractures, and spinal cord compression; such prophylaxis is now routinely recommended for all patients with myeloma bone disease and may improve overall survival. After 1 to 2 years, the dosing can be reduced to once every 3 months in patients who are stable to minimize the risk of osteonecrosis of the jaw, which is a complication of long-term bisphosphonate therapy.

Spinal cord compression from an extramedullary plasmacytoma (Chapter 400) should be suspected in patients who have severe back pain, weakness or paresthesias of the lower extremities, or bladder or bowel dysfunction. Initial treatment is with dexamethasone-based therapy or radiation therapy. If the neurologic deficit increases, surgical decompression is necessary.

Miscellaneous Complications

Symptomatic hyperviscosity (see later) is less common than in Waldenström macroglobulinemia. Anemia that persists despite adequate treatment of underlying myeloma often responds to erythropoietin.

PROGNOSIS

Multiple myeloma is considered incurable at present, but survival has improved significantly in recent years. The median survival is approximately 5 years, but it varies widely according to clinical stage and risk stratification factors (see Table 187-4).[9] In some patients, an acute or aggressive terminal phase is characterized by rapid tumor growth, pancytopenia, soft tissue subcutaneous masses, decreased M protein levels, and fever; survival in this subset is generally only a few months.

FUTURE DIRECTIONS

Future efforts must be directed toward identifying new active agents and developing effective combinations of active drugs. Studies are under way to improve the conditioning regimen used in autologous stem cell transplantation and to better integrate novel therapies with stem cell transplantation.

VARIANT FORMS OF MULTIPLE MYELOMA

Smoldering Multiple Myeloma

Smoldering (asymptomatic) multiple myeloma is defined by the presence of an M protein level higher than 3 g/dL in serum or 10 to 60% clonal plasma cells in bone marrow in the absence of myeloma defining events and amyloidosis.[10] Patients with smoldering multiple myeloma are biologically similar to those with MGUS but carry a much higher risk for progression to myeloma or related malignant disease: 10% per year for the first 5 years, 5% per year for the next 5 years, and 1 to 2% per year thereafter. As a result, patients must be observed more closely (every 3 to 4 months), but they should not be treated unless progression to symptomatic multiple myeloma occurs. A small randomized trial found improved survival with the use of Rd as preventive therapy in patients with high-risk smoldering multiple myeloma, but additional data are needed before this approach can be recommended as routine practice. However, patients with ultrahigh-risk features (such as serum free light chain ratio ≥100 or presence of one or more focal lesions on magnetic resonance imaging) are candidates for therapy similar to that for symptomatic myeloma because they are at imminent risk of progression.

Plasma Cell Leukemia

Patients with plasma cell leukemia have more than 20% plasma cells in the peripheral blood and an absolute plasma cell count of 2000/μL or higher.[11] Plasma cell leukemia is classified as primary when it is diagnosed in the leukemic phase (60%) or as secondary when there is leukemic transformation of a previously recognized multiple myeloma (40%). Patients with primary plasma cell leukemia are younger and have a greater incidence of hepatosplenomegaly and lymphadenopathy, higher platelet count, fewer bone lesions, smaller serum M protein component, and longer survival (median, 6.8 vs. 1.3 months) than do patients with secondary plasma cell leukemia. Treatment of plasma cell leukemia is unsatisfactory. An aggressive initial treatment regimen such as bortezomib, dexamethasone, thalidomide, cisplatin, doxorubicin, cyclophosphamide, and etoposide (VDT-PACE) for two cycles, followed by autologous stem cell transplantation and subsequent maintenance therapy with a bortezomib-based regimen, is a reasonable strategy if the patient's clinical condition permits such an approach. Secondary plasma cell leukemia rarely responds in a durable manner to chemotherapy because the patients have already received chemotherapy and are resistant.

Nonsecretory Myeloma

Patients with nonsecretory myeloma have no M protein in either serum or urine and account for only 3% of cases of myeloma. For the diagnosis to be made, the clonal nature of bone marrow plasma cells should be established by immunoperoxidase, immunofluorescence, or flow cytometric methods. Treatment and survival are similar to those of patients with typical myeloma. The serum free light chain assay is abnormal in more than 60% of patients and can be used to monitor the response to therapy.

Osteosclerotic Myeloma (POEMS Syndrome)

This syndrome is characterized by polyneuropathy, organomegaly, endocrinopathy, M protein, and skin changes (POEMS) (see Table 187-2). The major clinical features are a chronic inflammatory-demyelinating polyneuropathy with predominantly motor disability and sclerotic skeletal lesions. The bone marrow usually contains less than 5% plasma cells, and hypercalcemia and renal insufficiency rarely occur. Almost all patients have a λ-type

M protein. The diagnosis is confirmed by identification of monoclonal plasma cells obtained at biopsy of an osteosclerotic lesion.

If the lesions are in a limited area, radiation therapy substantially improves the neuropathy in more than 50% of patients. If the patient has widespread osteosclerotic lesions, treatment is with autologous stem cell transplantation or other systemic therapy similar to that used for myeloma.

Solitary Plasmacytoma (Solitary Myeloma) of Bone

The diagnosis of solitary bone plasmacytoma is based on histologic evidence of a solitary tumor consisting of monoclonal plasma cells identical to those in multiple myeloma. In addition, complete skeletal radiographs and magnetic resonance imaging of the spine and pelvis must show no other lesions of myeloma, and the bone marrow aspirate must contain no evidence of clonal plasma cells. An M protein may be present in serum or urine at diagnosis, but persistence of the M protein after radiation therapy is associated with an increased risk for progression to multiple myeloma. Treatment consists of radiation in the range of 40 to 50 Gy. Almost 50% of patients who have a solitary plasmacytoma are alive at 10 years, and disease-free survival rates at 10 years range from 15 to 25%. Progression to myeloma, when it occurs, usually takes place within 3 years, but patients must be monitored indefinitely. There is no convincing evidence that adjuvant chemotherapy decreases the rate of conversion to multiple myeloma.

Extramedullary Plasmacytoma

Extramedullary plasmacytomas outside the bone marrow are most commonly found in the upper respiratory tract (80% of cases), especially in the nasal cavity and sinuses, nasopharynx, and larynx. Extramedullary plasmacytomas may also occur in the gastrointestinal tract, central nervous system, urinary bladder, thyroid, breast, testes, parotid gland, or lymph nodes. Extramedullary plasmacytomas may be solitary, or they may occur in the context of existing myeloma. The diagnosis of solitary extramedullary plasmacytoma is based on detection of a plasma cell tumor in an extramedullary site, absence of clonal plasma cells on bone marrow examination, and absence of other bone or extramedullary lesions on radiographic studies. Treatment of solitary extramedullary plasmacytoma consists of either complete surgical resection or tumoricidal irradiation. The plasmacytoma may recur locally, metastasize to regional nodes, or, rarely, develop into multiple myeloma.

WALDENSTRÖM MACROGLOBULINEMIA (PRIMARY MACROGLOBULINEMIA)

DEFINITION

Waldenström macroglobulinemia is the result of the uncontrolled proliferation of lymphocytes and plasma cells in which an IgM M protein is produced.[12] The cause is unknown; familial clusters have been reported. The median age of patients at the time of diagnosis is about 65 years, and approximately 60% are male. The diagnostic criteria are IgM monoclonal gammopathy (regardless of the size of the M protein), 10% or greater bone marrow infiltration (usually intertrabecular) by clonal lymphocytes that exhibit plasmacytoid or plasma cell differentiation, and a typical immunophenotype (e.g., surface IgM^+, $CD5^{+/-}$, $CD10^-$, $CD19^+$, $CD20^+$, $CD23^-$) that would satisfactorily exclude other lymphoproliferative disorders, including chronic lymphocytic leukemia (Chapter 184) and mantle cell lymphoma (Chapter 185). A recurrent mutation of the *MYD88* gene (MYD88 L265P) has recently been shown to be present in most patients with Waldenström macroglobulinemia and is thought to be relatively specific for this disease.[13]

CLINICAL MANIFESTATIONS

Weakness, fatigue, and bleeding (especially oozing from the oronasal area) are common initial symptoms. Blurred or impaired vision, dyspnea, weight loss, neurologic symptoms, recurrent infections, and heart failure may occur. In contrast to multiple myeloma, lytic bone lesions, renal insufficiency, and amyloidosis are rare. Physical findings include pallor, hepatosplenomegaly, and lymphadenopathy. Retinal hemorrhages, exudates, and venous congestion with vascular segmentation ("sausage" formation) may occur. Sensorimotor peripheral neuropathy is common. Pulmonary involvement is manifested by diffuse pulmonary infiltrates and isolated masses.

Laboratory Evaluation

Almost all patients have moderate to severe normocytic, normochromic anemia. The serum electrophoretic pattern is characterized by a tall, narrow peak or dense band that is of the IgM type on immunofixation. Quantitative IgM levels are high. A monoclonal light chain is detected in the urine of 80% of patients, but the amount of urinary protein is generally modest.

The bone marrow aspirate is often hypocellular, but the biopsy specimen is hypercellular and extensively infiltrated with lymphoid cells and plasma cells. The number of mast cells is frequently increased. Rouleau formation is prominent (Chapter 157), and the sedimentation rate is markedly increased. About 10% of cases may have an associated type I cryoglobulinemia (see later).

DIAGNOSIS

Diagnosis requires the combination of typical symptoms and physical findings, the presence of an IgM M protein, and 10% or greater lymphoplasmacytic infiltration of the bone marrow. The lymphoplasmacytic cells express CD19, CD20, and CD22, whereas expression of CD5 and CD10 occurs in a minority. Asymptomatic patients with 10% or greater lymphoplasmacytic infiltration of the bone marrow are considered to have smoldering Waldenström macroglobulinemia. Multiple myeloma, chronic lymphocytic leukemia, and MGUS of the IgM type must be excluded.

Patients meeting the diagnostic criteria for Waldenström macroglobulinemia but who have less than 3 g/dL IgM protein at diagnosis have sometimes been classified as having lymphoplasmacytic lymphoma with an IgM M protein (Chapter 185). However, except for hyperviscosity, the clinical picture, therapy, and prognosis for these patients do not differ from those of patients with an IgM level of 3 g/dL or higher; thus, these patients are also considered to have Waldenström macroglobulinemia by the current definition.

PREVENTION AND TREATMENT

Rx

Patients should not be treated unless they have anemia, constitutional symptoms (such as weakness, fatigue, night sweats, or weight loss), hyperviscosity, or significant hepatosplenomegaly or lymphadenopathy. Rituximab, a chimeric anti-CD20 monoclonal antibody (Chapter 36), produces a response in at least 50% of untreated patients. The most common regimen used as front-line therapy is the combination of rituximab with cyclophosphamide and dexamethasone (RCD).[14] This combination is highly active and also preserves the ability to mobilize stem cells for transplantation, if necessary. Alternatives include bendamustine plus rituximab (BR); rituximab, bortezomib plus dexamethasone; and cladribine with or without rituximab.[15]

In general, for minimally symptomatic patients, rituximab as a single agent is an excellent choice for initial therapy. For patients with more advanced symptoms, including severe anemia or hyperviscosity, combination approaches such as RCD or BR are preferred.

For relapse, the agents used as initial therapy can be given alone or in combination. Autologous stem cell transplantation can be considered for eligible patients with relapsed disease.

Spuriously low hemoglobin and hematocrit levels may occur because of the increased plasma volume from the large amount of intravascular M protein. Consequently, transfusions should not be given solely on the basis of the hemoglobin or hematocrit value. Symptomatic hyperviscosity should be treated by plasmapheresis. The median survival of patients with macroglobulinemia is 5 years.

HYPERVISCOSITY SYNDROME

Hyperviscosity syndrome occurs in patients with Waldenström macroglobulinemia who have high levels of serum IgM M protein (>5 g/dL) and occasionally in those with myeloma, especially of the IgA type. Hyperviscosity is disproportionately more common relative to the same serum concentration of IgM and IgA M proteins compared with IgG M proteins because of the inherent tendency of IgM and IgA molecules to polymerize. Typically, IgM forms pentamers, whereas IgA forms dimers or sometimes trimers, resulting in high-molecular-weight complexes. Chronic nasal bleeding and oozing from the gums are the most frequent symptoms of hyperviscosity, but postsurgical or gastrointestinal bleeding may also occur. Retinal hemorrhages are common, and venous congestion with sausage-like segmentation and papilledema may be seen (Fig. 187-8). The patient occasionally complains of blurring or loss of vision. Dizziness, headache, vertigo, nystagmus, decreased hearing, ataxia, paresthesias, diplopia, somnolence, and coma may occur. Hyperviscosity can precipitate or exacerbate heart failure. Most patients have symptoms when the relative viscosity is greater than 4 cP, but the relationship between serum viscosity and clinical manifestations is not precise. There have

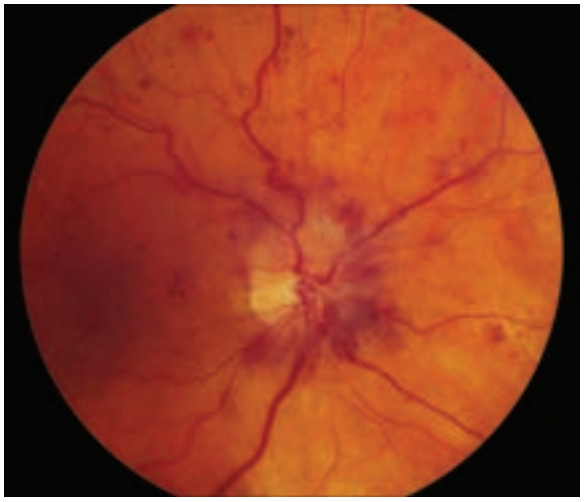

FIGURE 187-8. Hyperviscosity syndrome. Right eye retinal image in a patient with Waldenström macroglobulinemia and hyperviscosity syndrome showing sausaging (focal venular dilations), intraretinal hemorrhages, microaneurysms, and peripapillary cotton-wool spots and disc swelling (papilledema).

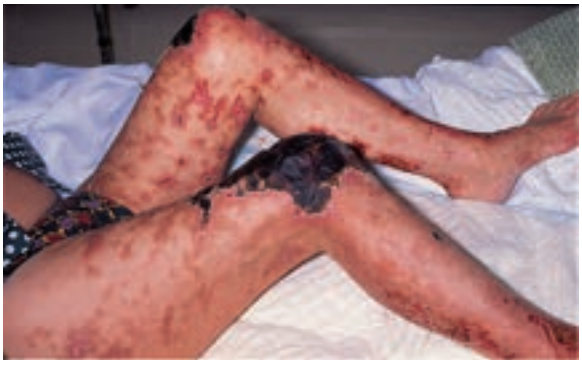

FIGURE 187-9. Skin infarction in cryoglobulinemia. The skin has a reticulated pattern as a result of leakage of red blood cells from damaged skin capillaries. Necrosis and ulceration have occurred in peripheral sites because of vessel blockage. This patient eventually required plastic surgery. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine*. 3rd ed. London: Mosby; 2003.)

been no randomized trials on management of hyperviscosity syndrome. Patients with symptomatic hyperviscosity should be treated with plasmapheresis and with chemotherapy to treat the underlying malignant disease. Plasma exchange of 3 to 4 L with albumin should be performed daily until the patient is asymptomatic. Plasma exchange is rapidly effective (two or three exchanges) in the case of IgM M proteins, which are primarily distributed in the intravascular space; with IgG M proteins, in contrast, multiple attempts may be needed because a significant amount of IgG can exist in the extravascular space.

HEAVY CHAIN DISEASES

The HCDs are characterized by the presence of an M protein consisting of a portion of the immunoglobulin heavy chain in serum, urine, or both. These heavy chains are devoid of light chains and represent a lymphoplasma cell proliferative process. There are three major types: γ-HCD, α-HCD, and μ-HCD.

γ-HCD

Patients with γ-HCD often initially have a lymphoma-like illness, but the clinical findings are diverse and range from an aggressive lymphoproliferative process (Chapter 185) to an asymptomatic state. Hepatosplenomegaly and lymphadenopathy occur in about 60% of patients. Anemia is found in approximately 80% initially and in nearly all eventually. The electrophoretic pattern often shows a broad-based band more suggestive of a polyclonal increase than an M protein. The diagnosis depends on the identification of an isolated monoclonal γ heavy chain on serum immunofixation, without evidence of either monoclonal κ or λ light chain expression.

Treatment is indicated only for symptomatic patients and consists of chemotherapy with melphalan plus prednisone or regimens used to treat non-Hodgkin lymphoma (Chapter 185), such as cyclophosphamide, vincristine, and prednisone. The prognosis of γ-HCD is variable and ranges from a rapidly progressive downhill course of a few weeks' duration to the asymptomatic presence of a stable monoclonal heavy chain in serum or urine.

α-HCD

α-HCD is the most common form of HCD and occurs in patients from the Mediterranean region or the Middle East, generally in the second or third decade of life. About 60% are men. The gastrointestinal tract is most commonly involved, and severe malabsorption with diarrhea, steatorrhea, and weight loss is noted (Chapter 140). Plasma cell infiltration of the jejunal mucosa is the most frequent pathologic feature. Immunoproliferative small intestinal disease is restricted to patients with small intestinal lesions who have the pathologic features of α-HCD but do not synthesize α heavy chains.

The serum protein electrophoretic pattern is normal in half the cases; in the remainder, an unimpressive broad band may appear in the α_2 or β region. The diagnosis depends on identification of an isolated monoclonal α heavy chain on serum immunofixation, without evidence of either monoclonal κ or λ light chain expression. The amount of α heavy chain in urine is small.

In the absence of therapy, α-HCD is typically progressive and fatal. The usual treatment consists of antibiotics, such as tetracyclines, and the eradication of any concurrent parasitic infection. Patients who do not respond adequately to antibiotics are given chemotherapy similar to that used to treat non-Hodgkin lymphoma, for example, the cyclophosphamide, hydroxydaunomycin, vincristine (Oncovin), and prednisone (CHOP) regimen (Chapter 185).

μ-HCD

This disease is characterized by the demonstration of an isolated monoclonal μ chain fragment on serum immunofixation, without evidence of either monoclonal κ or λ light chain expression.

The serum protein electrophoretic pattern is usually normal, except for hypogammaglobulinemia. Bence Jones proteinuria has been found in two thirds of cases. Lymphocytes, plasma cells, and lymphoplasmacytoid cells are increased in the bone marrow. Vacuolization of the plasma cells is common and should suggest the possibility of HCD. The course of μ-HCD is variable, and survival ranges from a few months to many years. Treatment is with corticosteroids and alkylating agents.

CRYOGLOBULINEMIA

Cryoglobulins are plasma proteins that precipitate when cooled and dissolve when heated. They are designated idiopathic or essential when they are not associated with any recognizable disease. Cryoglobulins are classified into three types: type I (monoclonal), type II (mixed monoclonal plus polyclonal), and type III (polyclonal).

Type I Cryoglobulinemia

Type I (monoclonal) cryoglobulinemia is most commonly of the IgM or IgG class, but IgA and Bence Jones cryoglobulins have been reported. Most patients, even with large amounts of type I cryoglobulin, are completely asymptomatic from this source. Others with monoclonal cryoglobulins in the range of 1 to 2 g/dL may have evidence of vasculitis with pain, purpura, Raynaud phenomenon, cyanosis, and even ulceration and sloughing of skin and subcutaneous tissue (Fig. 187-9) on exposure to cold because their cryoglobulins precipitate at relatively high temperatures. Type I cryoglobulins are associated with macroglobulinemia, multiple myeloma, or MGUS. Therapy for patients with symptomatic type I cryoglobulinemia and significant symptoms is similar to that for Waldenström macroglobulinemia for the IgM type and multiple myeloma for the non-IgM type.

Type II Cryoglobulinemia

Type II (mixed) cryoglobulinemia typically consists of an immune complex of IgM M protein and polyclonal IgG, although monoclonal IgG or monoclonal IgA may also be seen with polyclonal IgM. Serum protein electrophoresis generally shows a normal pattern or a diffuse, polyclonal hypergammaglobulinemic pattern. The quantity of mixed cryoglobulin is

usually less than 0.2 g/dL. Despite the monoclonal component, most patients do not have a clonal plasma cell disorder; rather, they have serologic evidence of infection with hepatitis C virus (Chapter 149). At present, hepatitis C is thought to be the cause of most cases of type II cryoglobulinemia.

Most clinical manifestations are related to the development of vasculitis and include palpable purpura, livedo reticularis, polyarthralgias, and neuropathy. Involvement of the joints is symmetrical, but joint deformities rarely develop. Raynaud phenomenon, necrosis of the skin, and neurologic involvement may be present. In almost 80% of renal biopsy specimens, glomerular damage can be identified. Nephrotic syndrome may result, but severe renal insufficiency is uncommon.

Early administration of corticosteroids is the most frequent therapy. Treatment should also target underlying hepatitis C infection with interferon alfa-2 or ribavirin (Chapter 149). Agents to treat the monoclonal component, such as cyclophosphamide, chlorambucil, azathioprine, or rituximab, are used if there is no response. Plasmapheresis (with a warmed circuit) is helpful in the acute management of symptoms by removal of circulating immune complexes.

Type III Cryoglobulinemia

Type III (polyclonal) cryoglobulinemia does not have a monoclonal component and is not associated with a clonal plasma cell proliferative disorder. Type III cryoglobulins are found in many patients with chronic infections or inflammatory diseases and are usually of no clinical significance unless they are associated with hepatitis C infection.

Grade A References

A1. Palumbo A, Cavallo F, Gay F, et al. Autologous transplantation and maintenance therapy in multiple myeloma. *N Engl J Med*. 2014;371:895-905.

A2. Cavo M, Tacchetti P, Patriarca F, et al. Bortezomib with thalidomide plus dexamethasone compared with thalidomide plus dexamethasone as induction therapy before, and consolidation therapy after, double autologous stem-cell transplantation in newly diagnosed multiple myeloma: a randomised phase 3 study. *Lancet*. 2010;376:2075-2085.

A3. Rajkumar SV, Jacobus S, Callander NS, et al. Lenalidomide plus high-dose dexamethasone versus lenalidomide plus low-dose dexamethasone as initial therapy for newly diagnosed multiple myeloma: an open-label randomised controlled trial. *Lancet Oncol*. 2010;11:29-37.

A4. Attal M, Harousseau JL, Facon T, et al. Single versus double autologous stem-cell transplantation for multiple myeloma. *N Engl J Med*. 2003;349:2495-2502.

A5. Cavo M, Pantani L, Petrucci MT, et al. Bortezomib-thalidomide-dexamethasone is superior to thalidomide-dexamethasone as consolidation therapy after autologous hematopoietic stem cell transplantation in patients with newly diagnosed multiple myeloma. *Blood*. 2012;120:9-19.

A6. Attal M, Lauwers-Cances V, Marit G, et al. Lenalidomide maintenance after stem-cell transplantation for multiple myeloma. *N Engl J Med*. 2012;366:1782-1791.

A7. McCarthy PL, Owzar K, Hofmeister CC, et al. Lenalidomide after stem-cell transplantation for multiple myeloma. *N Engl J Med*. 2012;366:1770-1781.

A8. Sonneveld P, Schmidt-Wolf IG, van der Holt B, et al. Bortezomib induction and maintenance treatment in patients with newly diagnosed multiple myeloma: results of the randomized phase III HOVON-65/GMMG-HD4 trial. *J Clin Oncol*. 2012;30:2946-2955.

A9. Krishnan A, Pasquini MC, Logan B, et al. Autologous haemopoietic stem-cell transplantation followed by allogeneic or autologous haemopoietic stem-cell transplantation in patients with multiple myeloma (BMT CTN 0102): a phase 3 biological assignment trial. *Lancet Oncol*. 2011;12:1195-1203.

A10. Hulin C, Facon T, Rodon P, et al. Efficacy of melphalan and prednisone plus thalidomide in patients older than 75 years with newly diagnosed multiple myeloma: IFM 01/01 Trial. *J Clin Oncol*. 2009;27:3664-3670.

A11. San Miguel JF, Schlag R, Khuageva NK, et al. Bortezomib plus melphalan and prednisone for initial treatment of multiple myeloma. *N Engl J Med*. 2008;359:906-917.

A12. Benboubker L, Dimopoulos MA, Dispenzieri A, et al. Lenalidomide and dexamethasone in transplant-ineligible patients with myeloma. *N Engl J Med*. 2014;371:906-917.

A13. Palumbo A, Hajek R, Delforge M, et al. Continuous lenalidomide treatment for newly diagnosed multiple myeloma. *N Engl J Med*. 2012;366:1759-1769.

A14. Stewart AK, Rajkumar SV, Dimopoulos MA, et al. Carfilzomib, lenalidomide, and dexamethasone for relapsed multiple myeloma. *N Engl J Med*. 2015;372:142-152.

A15. Morgan GJ, Davies FE, Gregory WM, et al. First-line treatment with zoledronic acid as compared with clodronic acid in multiple myeloma (MRC Myeloma IX): a randomised controlled trial. *Lancet*. 2010;376:1989-1999.

A16. Mateos M-V, Hernández M-T, Giraldo P, et al. Lenalidomide plus dexamethasone for high-risk smoldering multiple myeloma. *N Engl J Med*. 2013;369:438-447.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Rajkumar SV. Multiple myeloma: 2013 update on diagnosis, risk-stratification, and management. *Am J Hematol*. 2013;88:225-235.
2. Wadhera RK, Rajkumar SV. Prevalence of monoclonal gammopathy of undetermined significance: a systematic review. *Mayo Clinic Proc*. 2010;85:933-942.
3. Turesson I, Kovalchik SA, Pfeiffer RM, et al. Monoclonal gammopathy of undetermined significance and risk of lymphoid and myeloid malignancies: 728 cases followed up to 30 years in Sweden. *Blood*. 2014;123:338-345.
4. Dispenzieri A, Katzmann JA, Kyle RA, et al. Prevalence and risk of progression of light-chain monoclonal gammopathy of undetermined significance: a retrospective population-based cohort study. *Lancet*. 2010;375:1721-1728.
5. Rajkumar SV, Dimopoulos MA, Palumbo A, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol*. 2014;15:e538-e548.
6. Ludwig H, Miguel JS, Dimopoulos MA, et al. International Myeloma Working Group recommendations for global myeloma care. *Leukemia*. 2014;28:981-992.
7. Lacy MQ, McCurdy AR. Pomalidomide. *Blood*. 2013;122:2305-2309.
8. Kortuem KM, Stewart AK. Carfilzomib. *Blood*. 2013;121:893-897.
9. Munshi NC, Anderson KC, Bergsagel PL, et al. Consensus recommendations for risk stratification in multiple myeloma: report of the International Myeloma Workshop Consensus Panel 2. *Blood*. 2011;117:4696-4700.
10. Kyle RA, Larson DR, Therneau TM, et al. Clinical course of light-chain smouldering multiple myeloma (idiopathic Bence Jones proteinuria): a retrospective cohort study. *Lancet Haematol*. 2014;1:e28-e36.
11. Fernandez de Larrea C, Kyle RA, Durie BG, et al. Plasma cell leukemia: consensus statement on diagnostic requirements, response criteria and treatment recommendations by the International Myeloma Working Group. *Leukemia*. 2013;27:780-791.
12. Sahin I, Leblebjian H, Treon SP, et al. Waldenström macroglobulinemia: from biology to treatment. *Expert Rev Hematol*. 2014;7:157-168.
13. Treon SP, Xu L, Yang G, et al. MYD88 L265P somatic mutation in Waldenström's macroglobulinemia. *N Engl J Med*. 2012;367:826-833.
14. Gertz MA. Waldenström macroglobulinemia: 2011 update on diagnosis, risk stratification, and management. *Am J Hematol*. 2011;86:411-416.
15. Dimopoulos MA, Kastritis E, Owen RG, et al. Treatment recommendations for patients with Waldenström macroglobulinemia (WM) and related disorders: IWWM-7 consensus. *Blood*. 2014;124:1404-1411.

REVIEW QUESTIONS

1. A 63-year-old man is found to have a serum monoclonal protein level of 1.6 g/dL during work-up of joint pains. The monoclonal protein is IgG κ on immunofixation. He has no anemia, hypercalcemia, renal failure, or bone lesions. Bone marrow biopsy shows 8% plasma cells. The most likely diagnosis is

A. Multiple myeloma
B. Smoldering multiple myeloma
C. Monoclonal gammopathy of undetermined significance (MGUS)
D. Idiopathic Bence Jones proteinuria
E. Waldenström's macroglobulinemia

Answer: C This patient has a small monoclonal protein, without evidence of end-organ damage. This could be either MGUS or smoldering myeloma. The bone marrow plasma cell percentage is less than 10%, and the serum monoclonal protein level is less than 3 g/dL. Thus the diagnosis is MGUS. In contrast to MGUS, smoldering multiple myeloma requires either 10-60% plasma cells on bone marrow or 3 g/dL or more M spike on serum protein electrophoresis. The diagnosis of myeloma requires presence of one or more myeloma defining events. (See Table 187-2.)

2. Which of the following are risk factors for progression in monoclonal gammopathy of undetermined significance?

A. Size of the M protein
B. Type of the M protein
C. Abnormal free light chain ratio
D. A and C
E. A, B, and C

Answer: E The best predictors of progression of MGUS are the size of the M protein, the type of the M protein, and an abnormal baseline free light chain ratio. Patients with a serum M protein of less than 1.5 g/dL, IgG type, and normal free light chain ratio have a 5% probability of progression during 20 years. In contrast, if all three factors were abnormal, the risk of progression during that time is in excess of 50%. Another predictor of progression has recently been reported: reduction of one or two noninvolved immunoglobulin isotype levels (immunoparesis). (See section on prognosis under MGUS.)

3. Autologous stem cell transplantation is commonly used in the treatment of multiple myeloma in eligible patients. Which of the following statements is true about this procedure for the treatment of myeloma?

A. It is less commonly used more recently as therapy for myeloma compared with nonmyeloablative allogeneic transplantation.
B. It is not curative but prolongs overall and event-free survival.
C. Patients typically receive four to six cycles of melphalan, prednisone, and bortezomib (VMP) before transplantation.
D. Patients typically receive four to six cycles of thalidomide, melphalan, and prednisone (MPT) before transplantation.
E. The typical conditioning regimen is total body irradiation plus high-dose melphalan.

Answer: B Autologous stem cell transplantation for myeloma is not curative. Transplantation prolongs overall and event-free survival by about 12 to 18 months. It is much more commonly used than nonmyeloablative allogeneic transplantation, which should still be considered investigational for this disease. Melphalan should be avoided before induction because it may affect stem cells and prevent adequate mobilization; therefore, VMP and MPT are not good options before transplantation. The typical conditioning regimen is melphalan 200 mg/m^2.

4. Which of the following agents is associated with a high risk of peripheral neuropathy?

A. Bortezomib
B. Lenalidomide
C. Pomalidomide
D. Carfilzomib
E. Melphalan

Answer: A Of the agents listed, bortezomib is the drug associated with the highest risk of neuropathy. The major adverse effects with lenalidomide and pomalidomide include thrombosis, rash, low blood counts, and fatigue. Carfilzomib is a new proteasome inhibitor that unlike bortezomib appears to have less risk of neuropathy. Melphalan is an alkylating agent that is associated with low blood counts and risk of myelodysplastic syndrome. The rate of bortezomib neuropathy and its severity can be limited by use of a once-weekly dosing schedule and the subcutaneous route of administration. (See section on treatment of relapsed refractory myeloma under prevention and treatment of multiple myeloma.)

5. Which of the following is *not* true of Waldenström macroglobulinemia?

A. The type of monoclonal protein in Waldenström macroglobulinemia tends to form a pentamer.
B. Most patients have a mutation of the *MYD88* gene (MYD88 L265P).
C. Hyperviscosity is an important feature of the disease.
D. Osteosclerotic bone lesions are seen in 25% of patients.
E. Rituximab, cyclophosphamide, dexamethasone (RCD) is a commonly used treatment regimen.

Answer: D Osteosclerotic bone lesions are a feature of the POEMS syndrome, not of Waldenström's macroglobulinemia. Unlike with myeloma, in which osteolytic bone disease is present in most patients, patients with Waldenström's macroglobulinemia do not have bone disease. The type of monoclonal protein is IgM; this molecule tends to pentamerize, and the high molecular weight leads to hyperviscosity. Rituximab, cyclophosphamide, dexamethasone (RCD) is a commonly used treatment regimen for initial therapy. (See section on Waldenström macroglobulinemia.)

188

AMYLOIDOSIS

MORIE A. GERTZ

DEFINITION

Immunoglobulin light chain amyloidosis is characterized by a clonal population of bone marrow plasma cells that produces a monoclonal light chain of the κ or λ type, as either an intact molecule or a fragment. The light chain protein, instead of conforming to the α-helical configuration of most proteins, misfolds and forms a β-pleated sheet.[1] This insoluble protein is deposited in tissues and interferes with organ function. The β-pleated sheet configuration is responsible for the tinctorial properties; when the protein is stained with Congo red and viewed under polarized light, apple-green birefringence is demonstrated and is required for the diagnosis. Systemic light chain amyloidosis (AL) must be distinguished from the much less common amyloidosis associated with chronic infection and inflammatory arthropathies (secondary amyloidosis or AA) or with inherited amyloid cardiomyopathies and neuropathies (familial amyloidosis or AF).

CLINICAL MANIFESTATIONS

Amyloidosis is particularly difficult to diagnose and is a challenge for internists. The presenting symptoms can be diverse and are mimicked by far more common disorders (Table 188-1). The signs include tongue enlargement with dental indentations (Fig. 188-1) and "pinch" or periorbital purpura (Fig. 188-2), a result of vascular fragility. The signs are specific but lack sensitivity in that they are present in no more than 20% of patients. No single imaging procedure or laboratory study is diagnostic for the disease. The clinician must therefore be aware of the possibility of amyloidosis, or it may be overlooked. The kidney is commonly involved in amyloidosis (50% of cases). The diagnosis should be suspected in any patient who presents with nondiabetic nephrotic-range proteinuria (Chapter 121).[2] One third of patients with amyloidosis have nephrotic syndrome that is manifested with dramatic increases in the blood cholesterol level (median, 270 mg/dL), and urinalysis for proteinuria should be done in patients with a sudden increase in the serum cholesterol level. A patient with nondiabetic proteinuria may receive an empirical course of corticosteroids for possible minimal-change glomerulopathy. This treatment delays the diagnosis of amyloidosis and allows other organs to become involved.[3] Ten percent of renal biopsy specimens from patients with nondiabetic nephrotic syndrome are subsequently shown to be involved by amyloidosis. The incidence of bleeding after percutaneous renal biopsy is not increased in patients with AL.

The heart is involved in approximately 50% of patients with amyloidosis, and the presentation is subtle because fatigue is often the only manifestation.[4] Because amyloid heart disease (Chapter 60) is a disorder of diastolic failure, the typical findings of cardiomyopathy (enlarged cardiac silhouette on chest radiography, depressed ejection fraction by echocardiography, and pulmonary vascular redistribution) are absent. The effect of amyloidosis on the heart is poor filling during diastole. Patients have low end-diastolic volume and, as a consequence, poor stroke volume, despite a completely

TABLE 188-1 SYMPTOMS, SIGNS, AND SYNDROME OF AMYLOIDOSIS

SYMPTOMS AND SIGNS
Common symptoms: fatigue, edema, dyspnea, anorexia, paresthesias
Rare symptoms: claudication, joint pain and stiffness, sicca syndrome
Common signs: periorbital purpura, glossomegaly, hepatomegaly
Rare signs: waxy infiltration of eyelids, shoulder pad sign
SYNDROMES
Nondiabetic nephrotic syndrome
Nonischemic cardiomyopathy with an echocardiogram showing "hypertrophy"
Hepatomegaly or increased alkaline phosphatase with no imaging abnormality
Peripheral neuropathy with monoclonal gammopathy of undetermined significance or chronic inflammatory demyelinating polyneuropathy with autonomic features
Atypical myeloma with monoclonal light chains and modest marrow plasmacytosis

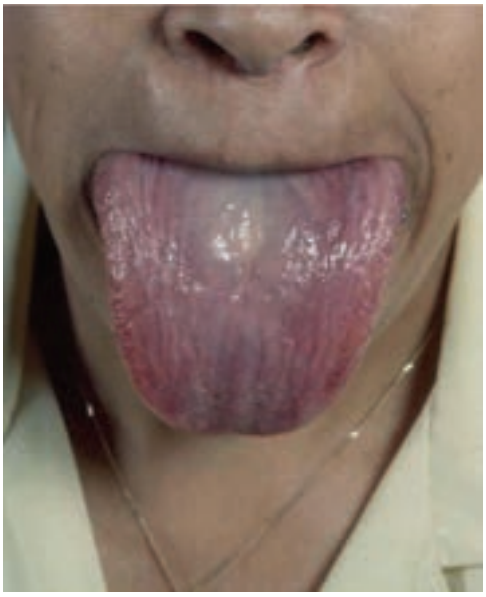

FIGURE 188-1. Macroglossia (or glossomegaly) in a patient with amyloidosis. (From Esplin BL, Gertz MA. Current trends in diagnosis and management of cardiac amyloidosis. *Curr Probl Cardiol.* 2013;38:53-96.)

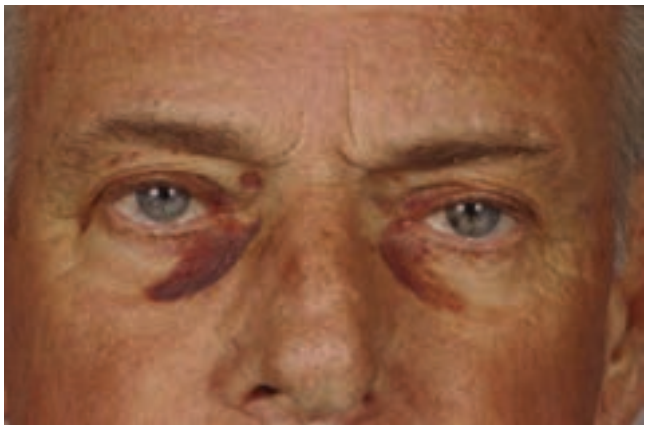

FIGURE 188-2. Periorbital purpura in amyloidosis. (From Kitchens CS. Purpura and other hematovascular disorders. In: Kitchens CS, Konkle BA, Kessler CM, eds. *Consultative Hemostasis and Thrombosis.* 3rd ed. Philadelphia: Elsevier; 2012.)

normal ejection fraction. Electrocardiography frequently shows a pseudoinfarct pattern, which can be interpreted as demonstrating silent ischemic infarction; this finding leads to coronary angiography, which is invariably negative (unless there is coincidental coronary artery disease). Echocardiography, which shows thickening of the heart walls due to amyloid infiltration, is frequently interpreted as showing left ventricular hypertrophy, and the cause of heart failure can be ascribed to silent hypertension or, alternatively, hypertrophic cardiomyopathy. Restrictive cardiomyopathy has been confused with pericardial disease, and patients have undergone unnecessary pericardiectomy, without clinical benefit. The classic granular sparkling appearance on the echocardiogram is not a useful diagnostic finding. Patients with amyloidosis rarely have symptoms of ischemic heart disease. Enhancement on magnetic resonance imaging with gadolinium is delayed in 69% of patients with cardiac amyloidosis. Myocardial enhancement is associated with increased ventricular mass and impaired left ventricular systolic function.

The liver is involved in 13% of patients. The typical presentation is hepatomegaly and an increased serum alkaline phosphatase value. Increased transaminase values and hyperbilirubinemia are late signs. Imaging is not helpful, and liver uptake is homogeneous. Many patients undergo evaluation for metastatic malignant disease. Liver biopsy is not associated with an increased rate of bleeding and is not contraindicated in the presence of hepatic amyloidosis. Rarely, patients present with spontaneous splenic rupture. Acquired deficiency of coagulation factor X (Chapter 174) is specific to AL and can be associated with clinically severe hemorrhage. Levels of coagulation factor X improve with effective therapy because the cause of low levels of circulating factor X is binding of factor X to the amyloid fibrils.

The peripheral neuropathy (Chapter 420) associated with amyloidosis begins in the lower extremities, is symmetrical, and is generally sensory or mixed sensorimotor. When a monoclonal protein is recognized, frequent diagnoses are chronic inflammatory demyelinating polyneuropathy and neuropathy associated with monoclonal gammopathy of undetermined significance because amyloidosis has not been considered in the differential diagnosis. Associated autonomic neuropathy occurs in approximately 4% of patients and can be characterized by orthostatic hypotension, which may be misattributed to cardiac failure. Autonomic dysmotility of the bowel is a common associated finding (Chapter 136). It can be upper intestinal, leading to pseudo-obstruction and recurrent emesis, or lower intestinal, characterized by alternating obstipation and fecal incontinence. Diarrhea caused by autonomic failure has been misdiagnosed as collagenous colitis when eosinophilic deposits are found in the bowel mucosa on hematoxylin-eosin staining in the absence of Congo red staining. Carpal tunnel syndrome (Chapter 420) occurs in approximately 13% of patients; it is not clinically distinguishable from the syndrome associated with repetitive stress injury but frequently fails to improve after surgical release. Rarely, interstitial lung disease, pseudoclaudication, periarticular deposits, and unexplained weight loss are presenting symptoms.

Systemic amyloidosis can be confused with early multiple myeloma (Chapter 187). Patients who present with vague symptoms of fatigue and edema are found to have a monoclonal protein in the urine, and a bone marrow biopsy specimen shows a clonal plasmacytosis with a median of 5% plasma cells in the bone marrow; however, a quarter of patients have more than 10% plasma cells in the bone marrow, a finding that qualifies as a diagnosis of multiple myeloma. These patients are often considered to have atypical multiple myeloma when the underlying amyloid syndrome is undetected. In these patients, a misdiagnosis of myeloma kidney or demyelinating neuropathy is made when an adequate diagnostic evaluation to exclude amyloidosis has not been performed.

DIAGNOSIS

Screening for Amyloidosis

In a patient with nondiabetic proteinuria, cardiomyopathy without ischemic risk factors, unexplained hepatomegaly, peripheral or autonomic neuropathy, or carpal tunnel syndrome, amyloidosis is not the most likely cause. The disorder occurs in only 8 per million persons per year, and routine biopsy is not appropriate whenever consistent symptoms are found. The classic physical finding of periorbital purpura occurs in only 10% of patients, is often limited to petechial eruptions over the eyelids, and is easily overlooked. Enlargement of the tongue occurs in 10 to 15% of patients; therefore, although it is specific, it is not sensitive for the diagnosis. Amyloidosis in patients with enlarged tongues may be unrecognized, or these patients may be evaluated for acromegaly or undergo unnecessary tongue biopsies because of the suspicion of squamous cell cancer. If biopsy is not an appropriate screening technique, what algorithm should be used to recognize AL?

By definition, amyloidosis is a plasma cell dyscrasia (Chapter 187); therefore, virtually all patients have a detectable immunoglobulin abnormality by immunofixation of the serum or urine, or they have abnormal results on a serum immunoglobulin free light chain assay. When a patient presents with a compatible clinical syndrome, these diagnostic studies should be completed before invasive diagnostic studies are performed (Fig. 188-3). Simple electrophoresis without immunofixation is inadequate because the monoclonal proteins are quantitatively very small in most patients and will not cause a detectable peak on serum protein electrophoresis. When these three diagnostic studies are used in combination, the sensitivity is 100%. If a monoclonal protein is detected, further investigations for amyloid should proceed, as described later. If a monoclonal protein is not found, three possibilities exist: (1) the patient does not have amyloidosis; (2) if the patient is known to have amyloidosis, it may be localized rather than systemic; or (3) if the patient is known to have systemic amyloidosis, it may be the senile systemic or familial type rather than the light chain type (see later section on other forms of amyloidosis).

Confirming the Diagnosis of Amyloidosis

In view of the grave prognosis associated with AL, the diagnosis must be confirmed by biopsy (with Congo red staining) in all cases.[5] Although it is reasonable to biopsy the kidney when proteinuria is the presenting symptom,

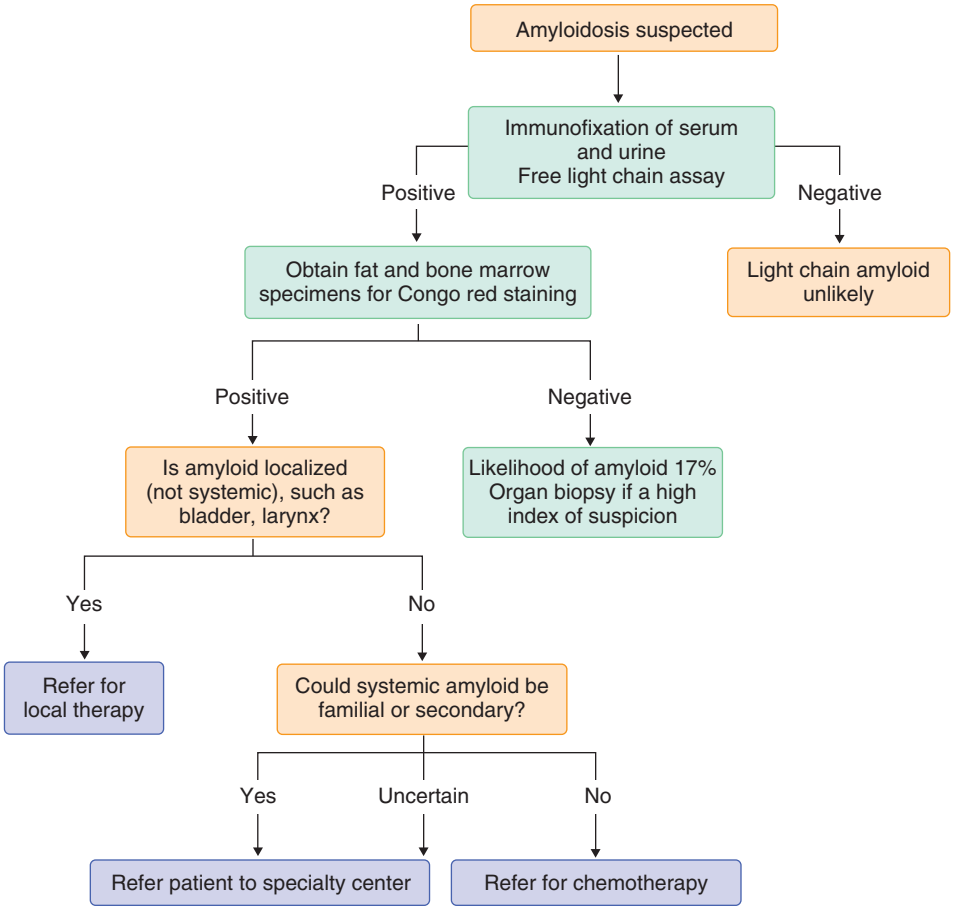

FIGURE 188-3. **Algorithm for the cost-effective pursuit of a diagnosis of systemic light chain amyloidosis.**

the heart when cardiomyopathy is recognized, the liver when there is hepatomegaly and increased alkaline phosphatase, or the nerve when there is a sensorimotor functional loss, these invasive and occasionally risky procedures are not required. Subcutaneous fat aspiration is an outpatient procedure that has a 24-hour turnaround time and recognizes amyloid deposits in 70% of patients. Bone marrow is a second convenient biopsy site, and this test is often required to exclude the possibility of associated multiple myeloma. Bone marrow biopsy is positive in 50% of patients. When both subcutaneous fat aspiration and bone marrow biopsy are done, amyloid is detected in 83% of patients. The remaining patients should have biopsy of the appropriate organ.

Once amyloid deposits are detected in tissues, further diagnostic evaluation is required. The presence of a monoclonal protein in the serum or urine and the presence of congophilic deposits in tissue do not verify that amyloidosis is light chain in origin. Further diagnostic studies are essential to classify the type of amyloid before therapy is initiated. Immunohistochemical studies on the tissue may be useful, but misfolding of the amyloid light chain often prevents epitopes from being recognized by commercial antisera; thus, false-negative results are common. Mass spectrometric analysis of the amyloid deposit can be done on paraffin-embedded tissue and validates the type of amyloid by direct amino acid sequencing, leaving no question about the origin of the amyloid protein as an immunoglobulin light chain. The incidence of monoclonal gammopathies in the elderly ranges from 3 to 5%. Therefore, that fraction of patients with senile systemic, localized, and familial amyloidosis (see section on other forms of amyloidosis) could be expected to have an associated monoclonal gammopathy, which would be misleading. Mass spectrometric analysis is feasible on subcutaneous fat tissue.

TREATMENT

Amyloidosis was previously thought to be untreatable and invariably fatal. With current therapy, response rates of about 70% regularly occur, and the median duration of survival is reportedly upward of 5 years.[6] Agents to reverse the misfolding of the protein and render it soluble would be ideal, but they are not available. The source of the immunoglobulin light chain is the clonal plasma cell population in the bone marrow. All known therapies are directed at destruction of the plasma cell clone.

The two treatment choices are generally traditional-dose chemotherapy and high-dose chemotherapy with autologous stem cell transplantation. Most patients are not candidates for high-dose therapy because of age, advanced cardiac dysfunction, or renal insufficiency. High-dose melphalan is a feasible approach in selected patients with cardiac AL and is associated with a high rate of hematologic and organ responses that lead to prolonged survival. Current therapies include combinations of melphalan, cyclophosphamide, dexamethasone, bortezomib, and lenalidomide. Treatment with a combination of bortezomib (1.5 mg/m^2 weekly or 1.3 mg/m^2 on days 1, 4, 8, and 11 every 28 days), cyclophosphamide (300 mg/m^2 orally weekly), and dexamethasone (40 mg weekly) has produced rapid and complete hematologic responses in the majority of patients with AL with few side effects.[7] Effective therapy has been associated with resolution of nephrotic syndrome, cardiac failure, and hepatomegaly. Imaging has shown amyloid deposits to regress after the suppression of light chain synthesis.

A systematic review and meta-analysis indicated that autologous hematopoietic stem cell transplantation does not appear to be superior to conventional chemotherapy in improving overall survival in patients with AL amyloidosis, although the quality of evidence was deemed low.

Assessing the Effect of Therapy

The serum immunoglobulin free light chain assay has been cited as a useful screening test for patients with a compatible clinical syndrome. This assay is also used to measure the therapeutic effect of intervention because the light chain level is quantifiable and reproducible. On the basis of current hematologic response criteria, successful therapy is characterized by a 50% reduction in the abnormal free light chain level.[8,9] Because the tissue toxicity associated with amyloid is related to the deposition of small amounts of light chain, it is unclear whether a successful outcome requires complete eradication of the light chain product. Studies have shown that patients achieving complete normalization of the free light chain have a better outcome, but it is uncertain whether patients who do not achieve this level of response should be subjected to more intensive treatment in an effort to remove this pathogenic amyloid serum precursor.

PROGNOSIS

The outcome of patients with AL depends on the extent of cardiac involvement (Chapter 60). With the advent of routine hemodialysis for this population, death due to renal failure is uncommon. The greater the involvement of the heart, the shorter a patient's survival. Echocardiography provides useful information about the ejection fraction, the thickness of the ventricular septum and left ventricular free wall, and the strain percentage (the rate at which wall shortening occurs). Doppler echocardiography allows quantitative measurements of diastolic function and reflects the slowing of blood flow into the ventricular chamber as the noncompliant left ventricle fills. This "stiffness," measured by the deceleration time, provides useful information and correlates well with survival.

Cardiac biomarkers are extremely sensitive measures of myocardial function, are reproducible, and can be used not only for prognosis but also to follow cardiac response after effective therapy. The serum troponin value is a powerful predictor of survival in patients with amyloidosis, and the N-terminal pro–brain natriuretic peptide value predicts survival after a diagnosis of amyloidosis. A staging system has been developed with these two cardiac biomarkers and the difference between involved and uninvolved free light chain levels to accurately predict survival.

OTHER FORMS OF AMYLOIDOSIS

Localized amyloidosis can be confused with systemic amyloidosis, but it has a much better prognosis. Localized AL amyloidosis represents a true plasma cell neoplasm and not a pseudotumor.[10] However, patients with localized amyloidosis generally do not require systemic therapy; management can be supportive or localized to the deposition. The location of the amyloid deposits can be a clue to the localized nature. Typical sites for localized amyloid deposition include the ureter, bladder, urethra, and prostate. Therapy entails cystoscopic resection or intravesical instillation of dimethyl sulfoxide. Most forms of cutaneous amyloidosis are localized, although nodular cutaneous amyloidosis has occurred in systemic AL. Tracheobronchial and laryngeal amyloidosis and nodular pulmonary amyloidosis are localized, are not associated with a plasma cell dyscrasia, and generally require only local therapy. Nodular pulmonary amyloidosis is often diagnosed after thoracotomy for a presumed malignant pulmonary nodule. Most cases of laryngeal amyloidosis are found when the patient presents to an otorhinolaryngologist with hoarseness and amyloid deposits are found on endoscopic biopsy. Patients with localized amyloidosis do not have a demonstrable monoclonal protein in the serum or urine and have a normal free light chain ratio. The localized amyloidosis found in Alzheimer disease is chemically unrelated to AL, and AL patients have no increased risk of dementia.

Senile systemic amyloidosis results from the deposition of a normal serum protein, transthyretin (TTR), in the myocardium. It has a much better prognosis than cardiac AL and generally necessitates endomyocardial biopsy for diagnosis; most patients are older than 70 years. When a monoclonal protein is present, it is incidental, and confirmation of type generally requires analysis of the amyloid-laden tissues. Therapy is supportive. The clinical presentation of senile systemic amyloidosis is not distinguishable from that of cardiac AL.

AF is uncommon in the United States and represents only 3% of cases of systemic amyloidosis. Patients present with the full clinical spectrum associated with amyloidosis, including cardiomyopathy, peripheral neuropathy, and proteinuria. Patients do not have a monoclonal protein because the deposited precursor is a mutant form of TTR, fibrinogen, lysozyme, or apolipoprotein A. Diflunisal (250 mg twice dail) can slow the rate of progression of the AF associated polyneuropathy.■ Selected patients with AF with polyneuropathy have benefited from liver transplantation because this disease is characterized by systemic accumulation of polymerized TTR in the peripheral nerves and systemic organs and liver transplantation stops the major production of amyloidogenic TTR.[11] In one hospital, the estimated probability of 10-year survival in patients with familial amyloid polyneuropathy was 100% after liver transplantation compared with 56% for the nontransplantation group, with the survival curves diverging at 6 years.[12]

One important form of AF in the United States is associated with an allele of the normal serum protein TTR in which isoleucine is substituted for valine at position 122 (TTR Val-122-Ile). The prevalence of this mutation in blacks in the United States is as high as 3.9%. Heterozygous inheritance of this mutant TTR is associated with late-onset cardiomyopathy in this population. Wall thickening is found on echocardiography. Heart failure is often mild at onset. The prognosis is far better than that of cardiac AL, and its recognition has important implications for genetic counseling.

SECONDARY AMYLOIDOSIS

Systemic AA is the rarest form in Western countries.[13] Previously, it was a consequence of uncontrolled sustained inflammation, usually infectious, and causes included tuberculosis and osteomyelitis. Cystic fibrosis, bronchiectasis, decubitus ulcers, and skin abscesses related to subcutaneous injection of illicit drugs are modern-day infectious causes. Today it occurs primarily in patients with difficult-to-control inflammatory syndromes, including Crohn's disease, juvenile arthritis, and ankylosing spondylitis. Organ damage results from the extracellular deposition of proteolytic fragments of the acute phase reactant serum amyloid A (SAA) as amyloid fibrils. However, because only a minority of patients with chronic inflammatory disorders actually develop this complication, disease-modifying factors must be also operative. The best characterized of these is the *SAA1* genotype. There has been a recent decline in numbers of patients presenting with AA amyloidosis due to rheumatic diseases, at least in part due to the use of disease-modifying antirheumatic therapy.[14]

The large majority of these patients present with proteinuria, nephrotic syndrome, or renal dysfunction. Diarrhea related to intestinal involvement (22%) and thyromegaly (9%) also occur.

Suppression of the inflammatory process results in regression of tissue amyloid deposits. Early diagnosis and rapid control of the underlying inflammatory disease are critical to prevent irreversible organ damage and to improve survival of patients with AA amyloidosis. Therefore, monitoring of patients with chronic, active inflammatory disease by serial testing of SAA, C-reactive protein, microalbuminemia, proteinuria, and other indicators of amyloid development, as well as possibly determination of the *SAA1* genotype, can guide approach to management. Anti– tumor necrosis factor-α agents have markedly reduced the incidence of AA from inflammatory arthritis. There are familial forms of AA associated with familial periodic fever syndromes, the most common being familial Mediterranean fever (Chapter 261) due to mutations in the tumor necrosis factor receptor. Interleukin-1 inhibitors (anakinra) have been used successfully in these inherited periodic fever syndromes.

Grade A References

A1. Mhaskar R, Kumar A, Behera M, et al. Role of high-dose chemotherapy and autologous hematopoietic cell transplantation in primary systemic amyloidosis: a systematic review. *Biol Blood Marrow Transplant*. 2009;15:893-902.

A2. Berk J, Suhr O, Obici L, et al. Repurposing diflunisal for familial amyloid polyneuropathy: a randomized clinical trial. *JAMA*. 2013;310:2658-2667.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Blancas-Mejia LM, Ramirez-Alvarado M. Systemic amyloidoses. *Ann Rev Biochem*. 2013;82:745-774.
2. Jazbeh S, Said A, Haddad RY, et al. Renal amyloidosis. *Dis Mon*. 2014;60:489-493.
3. Kourelis TV, Kumar SK, Go RS, et al. Immunoglobulin light chain amyloidosis is diagnosed late in patients with preexisting plasma cell dyscrasias. *Am J Hematol*. 2014;89:1051-1054.
4. Gertz MA, Dispenzieri A, Sher T. Pathophysiology and treatment of cardiac amyloidosis. *Nat Rev Cardiol*. 2015;12:91-102.
5. Gertz MA. Immunoglobulin light chain amyloidosis: 2013 update on diagnosis, prognosis, and treatment. *Am J Hematol*. 2013;88:416-425.
6. Gatt ME, Palladini G. Light chain amyloidosis 2012: A new era. *Br J Haematol*. 2013;160:582-598.
7. Mikhael JR, Schuster SR, Jimenez-Zepeda VH, et al. Cyclophosphamide-bortezomib-dexamethasone (CyBorD) produces rapid and complete hematologic response in patients with AL amyloidosis. *Blood*. 2012;119:4391-4394.
8. Palladini G, Dispenzieri A, Gertz MA, et al. New criteria for response to treatment in immunoglobulin light chain amyloidosis based on free light chain measurement and cardiac biomarkers: impact on survival outcomes. *J Clin Oncol*. 2012;30:4541-4549.
9. Kumar S, Dispenzieri A, Lacy MQ, et al. Revised prognostic staging system for light chain amyloidosis incorporating cardiac biomarkers and serum free light chain measurements. *J Clin Oncol*. 2012;30:989-995.
10. Westermark P. Localized AL amyloidosis: a suicidal neoplasm? *Ups J Med Sci*. 2012;117:244-250.
11. Yamashita T, Ando Y, Okamoto S, et al. Long-term survival after liver transplantation in patients with familial amyloid polyneuropathy. *Neurology*. 2012;78:637-643.
12. Benson MD. Liver transplantation and transthyretin amyloidosis. *Muscle Nerve*. 2013;47:157-162.
13. Real de Asúa D, Costa R, Galván JM, et al. Systemic AA amyloidosis: epidemiology, diagnosis, and management. *Clin Epidemiol*. 2014;6:369-377.
14. Obici L, Merlini G. AA amyloidosis: basic knowledge, unmet needs and future treatments. *Swiss Med Wkly*. 2012;142:w13580.

REVIEW QUESTIONS

1. A 61-year-old man presents to an otorhinolaryngologist with hoarseness. During laryngoscopy, a mass is found on the left vocal cord. A biopsy specimen shows dense deposits of amyloid. The patient is referred to you for subsequent evaluation. What is the most likely next step?

A. Cardiac biopsy to assess suitability for high-dose chemotherapy and stem cell transplantation
B. Initiation of bortezomib-based chemotherapy
C. Bone marrow biopsy to determine whether this represents coexistent multiple myeloma
D. Referral back to the otorhinolaryngologist for therapy
E. Bronchoscopy to assess status of mainstem bronchi

Answer: D Laryngeal and vocal cord amyloid is always localized (not associated with systemic amyloid). Therefore, the treatment is local and usually involves endoscopic laser therapy of the amyloid deposits. Recurrence is common, and the patient is likely to require ongoing monitoring. Pulmonary function testing would be reasonable to detect flow obstruction. Computed tomography imaging would not add value to the evaluation.

2. A 76-year-old black man is referred to you by a cardiologist after an endomyocardial biopsy specimen showed amyloid deposits. The patient was seen for dyspnea on exertion that evolved during 24 months. The echocardiogram was clinically suggestive of infiltrative cardiomyopathy. The patient has congestive heart failure. Serum and urine immunofixation assays were negative, and a free light chain assay showed a normal ratio. What should be your next step in the evaluation?

A. Classify the amyloid by mass spectrometry.
B. Chemotherapy is urgent in the presence of progressive congestive heart failure.
C. If a bone marrow assay is negative for amyloid deposits, the patient is eligible for cardiac transplantation.
D. Computed tomography imaging of the chest, abdomen, and pelvis should be performed to screen for occult amyloid deposits.

Answer: A The absence of serum and urine light chains and a normal light chain ratio make light chain amyloidosis highly unlikely. This patient is more likely to have non–immunoglobulin-related amyloidosis. Three percent of black men have a mutation in transthyretin, which causes late-onset familial amyloid cardiomyopathy. At 76 years of age, senile cardiac (systemic) amyloidosis is a definite possibility. There is a 1% possibility that this is light chain amyloidosis. The evaluation cannot move forward until the type of amyloid is identified. Mass spectrometry is the method most likely to provide this information (see reference 2).

3. Your patient with renal amyloidosis read that cerebral amyloid angiopathy is found in most patients with Alzheimer disease. The patient is concerned that dementia will soon follow. What is your next step?

A. Referral to a neurologist for neurocognitive studies
B. Magnetic resonance imaging to look for cerebral amyloid angiopathy
C. Aspirin therapy to prevent subcortical infarcts
D. Tell the patient not to worry

Answer: D The amyloid in Alzheimer disease is composed of amyloid-β protein, which is chemically unrelated to immunoglobulin light chains. There is no increased risk of Alzheimer disease in the patient.

4. A 58-year-old man is referred to you with endomyocardial biopsy–proven amyloidosis. The λ monoclonal protein in the serum is 0.6 mg/dL. The κ/λ free light chain ratio is 0.01. You perform a bone marrow biopsy that shows 7% plasma cells. The serum troponin level is 0.08 ng/mL. The N-terminal pro–brain natriuretic peptide value is 7800 pg/mL. What is your recommendation?

A. Referral to a transplant center for high-dose therapy
B. Referral for cardiac transplantation
C. Initiation of conventional chemotherapy
D. Initiation of digoxin to prevent atrial fibrillation
E. Referral for consultation in palliative care

Answer: C This patient's cardiac biomarkers suggest a degree of cardiac dysfunction that is associated with an excessive risk of death during a transplant procedure. This patient cannot be considered for cardiac transplantation until the underlying hematologic disorder is under control. Even patients with advanced cardiac failure can have improvement of cardiac function with appropriate systemic chemotherapy. Digoxin is contraindicated in cardiac amyloidosis.

5. Which is (are) the most important test(s) for assessing prognosis in a patient with light chain amyloidosis?

A. Troponin
B. Free light chains
C. N-terminal pro–brain natriuretic peptide
D. β_2-Microglobulin
E. Fluorescence in situ hybridization (genetic test)
F. A through C
G. B and D
H. A and C
I. A through D
J. A through E

Answer: F The cardiac biomarkers (troponin and N-terminal pro–brain natriuretic peptide) determine the prognosis in this disease and are the most important tests to perform. Testing for free light chains would complete the evaluation of prognosis as a measure of tumor mass. β_2-Microglobulin and albumin are not important markers for light chain amyloidosis, although they are important in multiple myeloma. The role of genetics is presently undefined in amyloidosis, and very few patients carry the adverse t(4;14) and −17p chromosomal abnormalities seen in myeloma.

189

TUMORS OF THE CENTRAL NERVOUS SYSTEM

LISA M. DEANGELIS

INTRACRANIAL TUMORS

General Approach to Brain Tumors

EPIDEMIOLOGY

About 23,000 new primary brain tumors and nervous system cancers are diagnosed annually in the United States, making central nervous system (CNS) tumors more than twice as common as Hodgkin disease and approximately one third as common as melanoma. There is no definitive evidence that any of them are linked to cell phone use. In contrast, intracranial metastases are five times more common than primary brain tumors. More than 120 types of primary brain tumors arise from the different cells that make up the CNS (Table 189-1). In addition to classifying tumors by their cell of origin, in clinical practice it is often useful to classify a tumor by its intracranial site as well, such as pineal region tumors or pituitary and suprasellar tumors.

TABLE 189-1 WORLD HEALTH ORGANIZATION CLASSIFICATION OF BRAIN TUMORS

TUMORS OF NEUROEPITHELIAL TISSUE
Astrocytic tumors
Astrocytoma
Anaplastic (malignant) astrocytoma
Glioblastoma
Pilocytic astrocytoma
Pleomorphic xanthoastrocytoma
Subependymal giant cell astrocytoma
Oligodendroglial tumors
Oligodendroglioma
Anaplastic (malignant) oligodendroglioma
Ependymal tumors
Ependymoma
Anaplastic (malignant) ependymoma
Myxopapillary ependymoma (spinal tumor)
Subependymoma
Mixed gliomas
Oligoastrocytoma
Anaplastic (malignant) oligoastrocytoma
Choroid plexus
Choroid plexus papilloma
Choroid plexus carcinoma
Neuronal and mixed neuronal—glial tumors
Gangliocytoma
Dysembryoplastic neuroepithelial tumor
Ganglioglioma
Anaplastic (malignant) ganglioglioma
Central neurocytoma
Pineal parenchymal tumors
Pineocytoma
Pineoblastoma
Embryonal tumors
Medulloblastoma
Primitive neuroectodermal tumor
TUMORS OF CRANIAL AND SPINAL NERVES
Schwannoma
Neurofibroma
TUMORS OF MENINGES
Meningioma
Hemangiopericytoma
Hemangioblastoma
PRIMARY CENTRAL NERVOUS SYSTEM LYMPHOMAS
GERM CELL TUMORS
Germinoma
Embryonal carcinoma
Yolk sac tumor (endodermal sinus tumor)
Choriocarcinoma
Teratoma
Mixed germ cell tumors
CYSTS AND TUMOR-LIKE LESIONS
Rathke cleft cyst
Epidermoid cyst
Dermoid cyst
Colloid cyst of the third ventricle
TUMORS OF THE SELLAR REGION
Pituitary adenoma
Pituitary carcinoma
Craniopharyngioma
METASTATIC TUMORS

Abridged and modified from World Health Organization classification.

PATHOBIOLOGY

In contrast to tumors arising elsewhere in the body, there is little distinction between benign and malignant tumors when they occur in the brain. The growth of brain tumors is restricted to the CNS; they rarely if ever metastasize to other organs. In the CNS, a malignant tumor is characterized by aggressive pathologic features, including local tissue invasion, neovascularity, regional necrosis, and cytologic atypia. These features confer a growth advantage to malignant cells and lead to rapid expansion and, frequently, to early regrowth after treatment. Tumors lacking these aggressive histologic features are preferably classified as low grade rather than benign. Many low-grade tumors continue to grow within the CNS, causing progressive neurologic disability, and some may acquire a more malignant phenotype over time. The low-grade tumors that transform into high-grade neoplasms are primarily the intra-axial tumors that cannot be cured by resection because of their diffuse infiltration of brain. Almost all truly benign CNS tumors are extra-axial tumors, such as meningiomas and acoustic neuromas that can be cured with complete surgical resection.

CLINICAL MANIFESTATIONS

A patient with a brain tumor can present with one or both of two types of symptoms and signs. *Generalized symptoms,* which typically reflect the increased intracranial pressure (ICP) that often accompanies cerebral tumors include headache, lethargy, personality change, nausea, and vomiting. *Localizing symptoms,* which reflect the specific location of the tumor include hemiparesis, hemisensory deficits, aphasia, visual field impairment, and seizures (Table 189-2).

Most patients have symptoms that progress during a week to a few months. A sudden intensification of symptoms may precipitate the patient's initial visit to the physician; however, a careful history usually reveals symptoms that predated the acute deterioration and slowly worsened over time. Two exceptions are the new appearance of a seizure in a previously asymptomatic individual (Chapter 403) and sudden hemorrhage into a tumor.

Symptoms of brain tumors can be produced by tumor invading brain parenchyma, tumor and edema compressing brain tissue, cerebrospinal fluid (CSF) obstruction caused directly by the tumor or by a shift of brain tissue, and herniation. Invasion and compression typically produce focal symptoms, many of which can be relieved if the compression is reduced. Obstruction of CSF flow and herniation are frequently a consequence of elevated ICP and typically produce generalized symptoms of headache, nausea, and vomiting, but they can also cause false localizing signs, such as an abducens nerve palsy as a result of diffuse increased ICP.

Headache (Chapter 398) is a presenting symptom of approximately 35% of brain tumors. It is more common in younger than in older patients and more common in patients who have rapidly growing tumors than in those whose tumors have evolved slowly (Fig. 189-1). Mental and cognitive abnormalities may be a reflection of local tumor (e.g., aphasia, alexia, agnosia) or of general impairment (e.g., lethargy, confusion, word finding difficulty, apathy). Seizures affect approximately one-third of patients with brain tumors, and they are especially common as the presenting and only symptom of a low-grade tumor. The seizures, which are focal because they originate at the site of the tumor, may remain restricted (e.g., focal motor seizures), or they may generalize secondarily, producing loss of consciousness, sometimes so quickly that the focal signature is missed by the patient or even an observant witness.

DIAGNOSIS

Imaging

Magnetic resonance imaging (MRI) is far superior to computed tomography (CT) and should be used in all cases of suspected intracranial tumor. MRI should be performed both without and with intravenous administration of gadolinium. A well performed MRI scan identifies any intracranial tumor, and a normal finding on MRI effectively excludes a neoplasm. The MRI of some extra-axial tumors (e.g., acoustic neuromas, meningiomas) is so characteristic that histologic confirmation is not required. A non–contrast-enhancing infiltrative lesion that is visible primarily on T2-weighted or fluid-attenuated inversion recovery images is most consistent with a low-grade glioma (Fig. 189-2), whereas a contrast-enhancing lesion with an area of central necrosis and surrounding edema is most likely to be a glioblastoma or possibly a brain metastasis. Although these diagnoses must be confirmed histologically, the preoperative diagnostic possibilities affect the surgical approach to the lesion.

Perfusion MRI after rapid infusion of gadolinium can measure the relative cerebral blood volume and neovascularity associated with a tumor; high perfusion is associated with higher grade of malignancy. This technique can help estimate the tumor grade preoperatively and guide the planning of treatment.

Magnetic resonance spectroscopy noninvasively assesses tissue composition. High-grade primary brain tumors are associated with a decrease in *N*-acetylaspartate and an increase in choline. More malignant tumors are

TABLE 189-2 FOCAL CLINICAL MANIFESTATIONS OF BRAIN TUMORS

Frontal lobe
Generalized seizures
Focal motor seizures (contralateral)
Expressive aphasia (dominant side)
Behavioral changes
Dementia
Gait disorders, incontinence
Hemiparesis (contralateral)
Basal ganglia
Hemiparesis (contralateral)
Movement disorders (rare)
Parietal lobe
Receptive aphasia (dominant side)
Spatial disorientation (nondominant side)
Cortical sensory dysfunction (contralateral)
Hemianopia (contralateral)
Agnosias
Occipital lobe
Hemianopia (contralateral)
Visual disturbances (unformed)
Temporal lobe
Complex partial (psychomotor) seizures
Generalized seizures
Behavioral changes
Olfactory and complex visual auras
Language disorder (dominant side)
Visual field defect
Corpus callosum
Dementia (anterior)
Memory loss (posterior)
Behavioral changes
Asymptomatic (middle)
Thalamus
Sensory loss (contralateral)
Behavioral changes
Language disorder (dominant side)
Midbrain/pineal
Paresis of vertical eye movement
Pupillary abnormalities
Precocious puberty (boys)
Sella/optic nerve/pituitary
Endocrinopathy
Bitemporal hemianopia
Monocular visual defects
Ophthalmoplegia (cavernous sinus)
Pons/medulla
Cranial nerve dysfunction
Ataxia, nystagmus
Weakness, sensory loss
Spasticity
Cerebellopontine angle
Deafness (ipsilateral)
Loss of facial sensation (ipsilateral)
Facial weakness (ipsilateral)
Ataxia
Cerebellum
Ataxia (ipsilateral)
Nystagmus

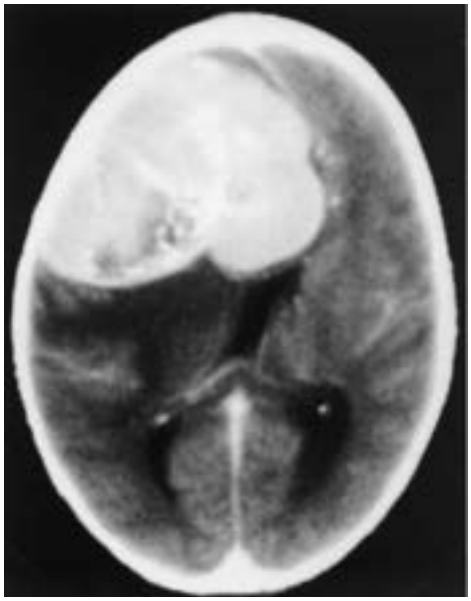

FIGURE 189-1. Meningioma. Computed tomography scan with contrast enhancement of a meningioma in a patient who presented with mild cognitive deficits, illustrative of the size that a slow-growing tumor can attain in the brain. The tumor was completely resected.

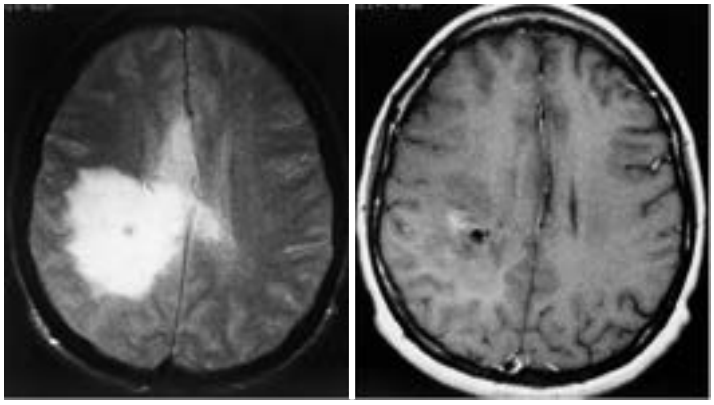

FIGURE 189-2. Glioma. Magnetic resonance imaging of a low-grade glioma. *Left*, T2-weighted image. *Right*, T1-weighted image, gadolinium contrast with minimum enhancement. The images are typical of this tumor, which is being detected with increasing frequency by magnetic resonance imaging in seizure patients. Many are invisible on computed tomography scans.

associated with a greater choline/*N*-acetylaspartate ratio and frequently contain areas with elevation of lactate and lipid.

Surgical resection is a major objective in the treatment of almost every kind of brain tumor, but resection must be balanced against possible damage to adjacent normal brain. The development of functional MRI (fMRI), which measures cerebral blood flow when areas of cortex are activated, has greatly enhanced the ability to localize critical neurologic functions and their relationship to the tumor preoperatively. When the fMRI is fused with the anatomic MRI, essential functions can be identified in relationship to the patient's tumor, and a safer and more complete resection may be planned.

On positron emission tomography (PET), high-grade tumors are usually hypermetabolic, whereas low-grade tumors are hypometabolic. New technologies using ^{11}C-methionine PET may differentiate low- from high-grade gliomas much more efficiently than deoxyglucose PET.

CT, without and with intravenous administration of contrast material, should be used only for patients who cannot undergo MRI. A CT scan, even with the administration of contrast material, may miss low-grade tumors and tumors in the posterior fossa.

Angiography no longer has a role in the diagnosis of intracranial tumors. However, angiographic embolization is occasionally useful preoperatively to reduce the vascularity of some meningiomas, thereby making a complete resection safer and more feasible.

Other Tests

Electroencephalography is rarely needed in the diagnosis or management of brain tumors. An electroencephalogram can occasionally be useful in a patient who has prolonged or unexplained stupor and in whom nonconvulsive status epilepticus is a consideration. Intraoperative monitoring is also used frequently to help guide resection of epileptogenic cortex adjacent to or within brain tumor tissue.

CSF analysis has little role in the diagnosis of most intracranial neoplasms. In primary CNS lymphoma (Chapter 185), the diagnosis may be established on CSF cytologic examination in about 15% of patients. The sensitivity of CSF cytology in the diagnosis of CNS lymphoma increases when it is combined with flow cytometry and is further enhanced by immunophenotypic and molecular genetic analyses of the CSF.[1] Rarely, a lumbar puncture is

TABLE 189-3 DIFFERENTIAL DIAGNOSIS OF INTRACRANIAL TUMORS

- Infection
  - Brain abscess
    - Bacterial
    - Fungal
    - Parasitic (e.g., cysticercosis)
  - Herpes encephalitis
- Vascular disease
  - Stroke
  - Intracranial hemorrhage
- Inflammatory conditions
  - Granuloma (sarcoid)
  - Multiple sclerosis: tumefactive single large lesion
- Vascular malformations
  - Cavernous angiomas
  - Venous angiomas
- Congenital abnormalities
  - Cortical dysplasia
  - Heterotopia

TABLE 189-4 TREATMENT OF BRAIN TUMORS

SYMPTOMATIC

- Glucocorticoids
- Antiepileptics
- Venous thromboembolism prophylaxis and treatment

DEFINITIVE

- Surgery
  - Goal is gross total excision
- Radiation therapy
  - Standard external beam
    - Fractionated
    - Usually focal
  - Stereotactic radiosurgery
- Chemotherapy
  - Limited by intrinsic drug resistance and blood-brain barrier

required to exclude inflammatory conditions or other processes that may be confused with a primary brain tumor. Lumbar puncture must be avoided in patients with cerebellar tumors because the release of pressure through the spinal needle may result in herniation of the cerebellar tonsils through the foramen magnum.

Differential Diagnosis

Patients who present with symptoms of raised ICP or the new onset of central neurologic symptoms, such as hemiparesis or seizure, should be evaluated rapidly. Prompt neuroimaging discloses a mass, and the radiographic features narrow the differential diagnosis (Table 189-3). Extra-axial tumors, such as a meningioma or acoustic neuroma, can be confused with a dural metastasis. Low-grade intra-axial tumors, which are nonenhancing on MRI, have been confused with infections such as herpes encephalitis when they involve the temporal lobe. Contrast-enhancing intra-axial tumors can be confused with a stroke, brain abscess, or focal plaque of demyelination. Subacute infarction can show brisk contrast enhancement, usually in a gyral pattern, unlike brain tumors in which enhancement is primarily in the white matter; however, the two are occasionally indistinguishable radiographically. Brain abscesses typically have a thinner enhancing wall than a malignant tumor and have restricted diffusion. Despite careful evaluation, patients thought to have a malignant glioma occasionally are found at surgery to have a brain abscess. A single large plaque of demyelination can also be confused radiographically with a brain tumor, and sometimes the diagnosis can be established only by biopsy.

When MRI suggests a primary brain tumor, there is no need for an extensive systemic search for a possible source of metastasis. Brain metastases are more common than primary brain tumors, but most occur in patients with known cancer, typically with active systemic disease. If an obvious systemic cancer is not revealed by a thorough physical examination, chest radiograph, routine blood tests, and urinalysis, the patient should proceed to craniotomy. Even if a brain metastasis is found at surgery, resection of a single brain metastasis is the appropriate treatment, and the pathologic examination of the lesion guides the subsequent search for the primary tumor.

TREATMENT

The treatment for all brain tumors can be divided into two main categories: symptomatic and definitive (Table 189-4). Symptomatic treatment addresses the associated problems, such as cerebral edema, seizures, and thromboembolic disease, which can contribute substantially to clinical symptoms. Definitive treatment addresses the tumor itself.

Symptomatic Treatment

Symptomatic management includes the use of corticosteroids, anticonvulsants, and prophylaxis for deep venous thrombosis (Chapter 38). Corticosteroids decrease the vasogenic edema that surrounds primary and metastatic brain tumors. Blood vessels associated with tumor formation are leaky and do not share the normal morphologic and physiologic features that form the blood-brain barrier; corticosteroids effectively reconstitute the blood-brain barrier by decreasing the abnormal permeability of these neovessels. Clinical improvement may begin within minutes, and frequently patients are dramatically improved within 24 to 48 hours.

Dexamethasone is the most commonly used glucocorticoid because it has the least mineralocorticoid activity. The usual starting dose is 12 to 16 mg/day, but this can be adjusted to find the lowest possible dose that alleviates neurologic symptoms. After definitive treatment is instituted, many patients can be tapered off their corticosteroid completely. Chronic high-dose corticosteroid therapy is associated with substantial side effects (Chapter 35) and should be avoided if possible. Patients who will be taking glucocorticoids for 6 weeks or longer should receive prophylaxis against *Pneumocystis jiroveci* (formerly *Pneumocystis carinii*; Chapter 341).

Anticonvulsants are administered to any patient who has had a seizure, but prophylactic anticonvulsants should not be prescribed for patients who have never had a seizure, except in the immediate perioperative period.[A1] A taper should begin 2 to 3 weeks after craniotomy.

Venous thromboembolism, which occurs in about 25% of patients with brain tumors, can occur early in the illness or at any time during treatment. All patients undergoing neurosurgery should have pneumatic compression boots in the postoperative period to reduce the incidence of venous thromboembolism. Prophylactic anticoagulants have also been used successfully in the immediate postoperative period without increasing postoperative hemorrhage. Appropriately regulated anticoagulation (Chapter 38) is the optimal therapy for deep venous thrombosis and is not associated with an increased risk of intracerebral hemorrhage in patients with intracranial tumors. Inferior vena cava filters can be used for patients who have deep vein thrombi or pulmonary emboli and who cannot be fully anticoagulated.

Definitive Treatment

Surgery

Complete excision is the goal for a primary brain tumor. Surgical excision can often be accomplished for primary extra-axial tumors, such as meningiomas and acoustic neuromas, unless their intracranial location makes resection impossible. Tumors of the skull base are particularly difficult to remove, and partial resection for decompression is often performed to preserve neurologic function. The safe boundaries for resecting cortical lesions while preserving function can often be elucidated by preoperative fMRI and intraoperative cortical mapping. However, lesions involving critical structures, such as the brainstem or thalamus, cannot be excised safely.

Lesions that cannot be resected are still amenable to biopsy for diagnostic purposes. Stereotactic biopsy can reach lesions in almost any area of the brain with minimal morbidity. The risks of stereotactic biopsy include inadequate tissue sample to make a diagnosis; a tissue sample that does not accurately reflect the most malignant grade of the tumor; and a procedure-related complication, such as hemorrhage. Hemorrhage that causes neurologic impairment occurs in only 2% of stereotactic biopsies, typically in patients with glioblastoma.

Complete excision can cure an extra-axial primary brain tumor and is associated with prolonged survival and better neurologic outcome even in patients with primary intra-axial tumors. Gross total excision, as measured by postoperative neuroimaging, is associated with prolonged survival in patients with malignant gliomas and probably in those with low-grade gliomas as well. However, most low-grade gliomas are not amenable to gross total excision, and usually only partial excision is feasible. Macroscopic tumor can frequently be removed completely in patients with high-grade gliomas, but there is always remaining microscopic disease that infiltrates surrounding brain.

Some tumors, such as brainstem gliomas, are in such critical locations that biopsy is not attempted. Their characteristic radiographic appearance permits diagnosis and initiation of medical treatment.

Radiation Therapy

A course of external beam radiation therapy is delivered in small daily fractions to a total cumulative dose usually between 45 and 60 Gy. Dividing the treatment into small daily fractions permits sublethal repair in normal tissues and markedly reduces neurologic toxicity associated with cerebral irradiation. External beam irradiation, which is the most effective nonsurgical treatment of brain tumors, doubles median survival time of patients with malignant primary brain tumors or metastatic lesions. It can also be useful for recurrent meningiomas and acoustic neuromas. However, it only rarely cures any of these lesions, and most patients develop recurrent disease despite maximal radiation therapy.

Stereotactic radiosurgery has been developed to deliver high fractions of focused radiation therapy that spare normal surrounding tissue. The technique is limited to tumors that are 3 cm in diameter or smaller and is less useful for malignant gliomas because of their infiltrative nature.

The neurologic complications of radiation therapy, which are usually observed in patients months to years after completion of treatment, include radionecrosis, dementia, and leukoencephalopathy. The incidence is reported as less than 5%, and most patients die of their brain tumor before the delayed consequences of treatment can be observed. However, in long-term survivors (e.g., patients with low-grade glioma or children with medulloblastoma), the late consequences of radiation therapy are important. Dementia accompanying radiation-induced leukoencephalopathy can progress and result in severe neurologic impairment. Radionecrosis can mimic recurrent tumor with a large contrast-enhancing lesion on MRI. Corticosteroids can reduce the edema and sometimes are sufficient to treat small areas of radionecrosis. However, if the lesion is sufficiently large, resection may be required to decompress the mass and reduce the steroid requirements.

Chemotherapy

Chemotherapy for brain tumors has usually been disappointing because of the intrinsic resistance of these tumors to most conventional agents. Carboplatin and cisplatin are active agents against medulloblastoma, even when the tumor is disseminated in the CSF. Temozolomide (150 to 200 mg/m^2 for 5 days every 4 weeks) is active in all gliomas, and high-dose methotrexate (3 to 8 g/m^2 for 3 to 12 months) is effective for primary CNS lymphoma. For patients with glioblastoma, polymers impregnated with carmustine (BCNU) and placed in a resection cavity offer modest benefit compared with no chemotherapy, but they are associated with local tissue injury and edema.

Specific Types of Brain Tumors

PRIMARY EXTRA-AXIAL TUMORS

The most common primary extra-axial tumors are meningiomas, pituitary adenomas, and acoustic neuromas. These tumors arise within the intracranial cavity but are not tumors of brain tissue. Almost all are benign; because the brain is rarely invaded, complete excision often enables cure with full recovery of neurologic function. These tumors produce neurologic symptoms and signs by compressing the underlying brain; however, edema of the underlying brain is infrequent, so glucocorticoids have a limited role.

Meningiomas

EPIDEMIOLOGY

Meningiomas are usually benign. Between 5 and 10% of meningiomas are atypical or malignant variants with a more aggressive course.[2] Meningiomas are more common in women, may be multiple in about 10% of patients with sporadic meningioma, and are occasionally part of a familial syndrome. They occur with increased frequency in patients with neurofibromatosis type 2. *NF2* inactivation is seen in approximately 50% of sporadic tumors, and mutations in *AKT1*, *SMO*, and *TRAF7* have also been identified.

DIAGNOSIS

Meningiomas grow slowly and produce symptoms that are insidious in onset and typically slowly progressive. Tumors can reach a considerable size, but they grow so slowly that the brain accommodates to the progressive compression. Meningiomas typically occur in specific locations: over the convexity, along the falx and parasagittal area, in the olfactory groove, at the base of the skull near the sphenoid bone, in the cavernous sinus (Fig. 189-3), in the cerebellopontine angle, and in the foramen magnum. Cortical and parasagittal tumors typically are manifested with seizures or progressive hemiparesis. Tumors in the anterior cranial fossa can cause slowly progressive changes in personality and cognition. Meningiomas at the base of the skull are manifested with cranial neuropathies and gait difficulties when there is brainstem compression. Frequently, tumors are completely asymptomatic and are identified on neuroimaging done for another purpose, such as head trauma.

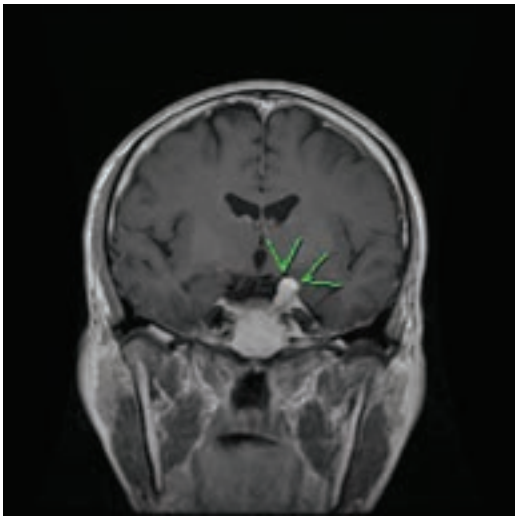

FIGURE 189-3. Skull base meningioma. A post-gadolinium coronal magnetic resonance image demonstrating a small left meningioma arising from the clinoid.

On MRI, meningiomas have a characteristic appearance consisting of a diffusely enhancing, dural-based lesion that is associated with a thin enhancing dural tail extending from the tumor. The radiographic features are often so characteristic that surgery is performed for therapeutic purposes only. The radiographic differential diagnosis includes the less common hemangiopericytoma and dural metastasis. Most meningiomas are not accompanied by significant edema, but marked edema is seen with high-grade malignant lesions or the secretory variant.

If small meningiomas are discovered in the absence of clinical symptoms or the symptoms are minor, lesions may be monitored with serial images because growth can be so slow.

TREATMENT

If treatment is indicated, complete resection is often curative, but even completely resected benign tumors may recur (as many as 20% in some series), so radiologic follow-up is essential.[3] Tumors at the base of the skull often cannot be resected completely and tend to recur despite successive attempts at surgical resection. Stereotactic radiosurgery may be an alternative to surgery if the lesion is small or there is progressive or residual tumor. External beam radiation therapy may slow progression of recurrent lesions and is essential for the treatment of malignant meningiomas. No effective chemotherapy has yet been identified.[4]

Acoustic Neuromas

Acoustic neuromas (Chapter 428), better called vestibular schwannomas, are benign tumors that arise from the eighth cranial nerve. Acoustic neuromas are twice as common in women as in men; the peak age is between 40 and 60 years. Sporadic vestibular schwannomas are unilateral; bilateral acoustic neuromas are pathognomonic of neurofibromatosis type 2 (Chapter 417).

Acoustic neuromas usually arise from the vestibular portion of the nerve and typically are manifested with unilateral hearing loss, sometimes preceded or accompanied by tinnitus and a sensation of dizziness or unsteadiness but not true vertigo. The slow, progressive enlargement of the tumor produces ipsilateral facial numbness or weakness by compressing the fifth or seventh cranial nerve, respectively. Tumors originate within the internal auditory meatus but grow out of the acoustic canal and into the cerebellopontine angle, where they can compress the brainstem and cause ataxia and ipsilateral cerebellar signs. Cranial MRI with gadolinium delineates even small acoustic neuromas with ease (Fig. 189-4).

Treatment is often surgical; stereotactic radiosurgery may be an alternative for lesions smaller than 3 cm. It is preferable to treat the tumors when they are small to preserve facial nerve function and hearing.

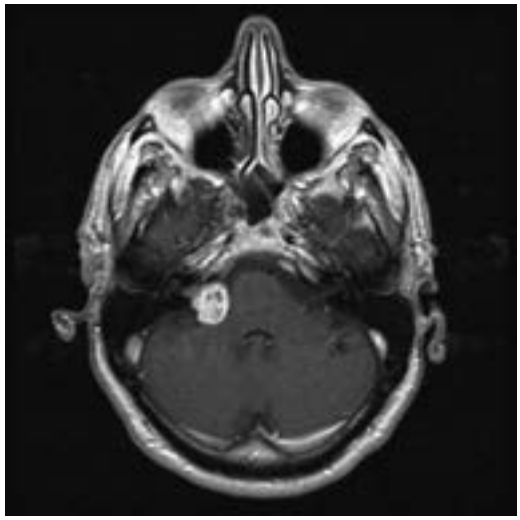

FIGURE 189-4. Acoustic neuroma. A post-gadolinium magnetic resonance image demonstrating a large right acoustic neuroma. The origin can be seen in the acoustic canal, but the tumor has grown into the cerebellopontine angle, where it is compressing the brainstem.

Pituitary Adenomas

Pituitary adenomas (Chapter 224) can be classified according to their size as microadenomas (<1 cm in diameter) or macroadenomas; by the presence or absence of endocrine function; and by the endocrinologic or neurologic syndromes caused by tumor compression. Microadenomas typically manifest with endocrine symptoms as described in Chapter 224. As pituitary tumors enlarge and become macroadenomas, they compress the surrounding neural structures, including the optic chiasm and optic nerves, typically causing bitemporal hemianopia and occasionally causing unilateral visual loss. Macroadenomas are frequently nonsecreting but destroy pituitary tissue, causing panhypopituitarism. Rarely, pituitary tumors manifest with the abrupt onset of headache, ophthalmoplegia, unilateral blindness, and even a depressed level of alertness or coma—a syndrome of *pituitary apoplexy* caused by hemorrhage or infarction.

Cranial MRI, particularly with coronal images and gadolinium administration, can completely outline the pituitary tumor and surrounding neural structures. All microadenomas and some macroadenomas can be treated with transsphenoidal pituitary surgery, which is associated with minimal morbidity. On occasion, residual or recurrent tumor necessitates radiation therapy. Some hormone-secreting tumors, particularly prolactinomas or growth hormone–secreting tumors, can be treated medically with cabergoline or somatostatin analogues such as octreotide, respectively (Chapter 224). These medications not only correct the hormonal excess but also shrink the tumor; they must be taken for life.

Other tumors in the pituitary and suprasellar region include craniopharyngiomas, suprasellar epidermoid cysts, Rathke cleft cysts, germinomas (discussed later), and lymphocytic hypophysitis, which is a benign inflammatory condition that usually is manifested with diabetes insipidus (Chapter 225). MRI frequently differentiates these conditions, which are usually suprasellar and erode into the pituitary fossa only secondarily. Some of these lesions also have characteristic radiographic features. These lesions are benign. Except for hypophysitis, which resolves completely with corticosteroid treatment (e.g., methylprednisolone 120 mg daily for 2 weeks and then tapered for 1 additional week), complete surgical excision is the curative therapy.

Other Extra-Axial Tumors

Pineal region tumors all have a characteristic clinical presentation that includes Parinaud syndrome, which consists of paresis of upward gaze, poor pupillary reaction to light with brisk reaction on accommodation, impairment of convergence, and convergence-retraction nystagmus. Some of these lesions may also cause hydrocephalus and symptoms of increased ICP. Pineal region tumors include pineal parenchymal tumors, such as pineocytomas and the more aggressive pineoblastomas, and germ cell tumors, including germinomas and nongerminomatous germ cell tumors. Germinomas can be completely cured with focal radiation therapy, whereas nongerminomatous germ cell tumors are more aggressive and frequently relapse despite chemotherapy plus cranial irradiation.

Chordomas are rare tumors of residual notochordal tissue. They usually occur at the base of the skull, are locally invasive, and are characterized by multiple recurrences despite surgery and radiation therapy. Chordomas are characterized by overexpression of the transcription factor brachyury. They also have activation of several receptor tyrosine kinases with overactivation of the downstream pathways, specifically the PI3K/AKT mTOR pathway. Inhibitors of the epidermal growth factor receptor and platelet-derived growth factor β have each been reported to produce responses and clinical benefit in patients with recurrent disease.[5] PI3K/AKT/mTOR inhibitors are under investigation.

Lipomas are benign tumors that can occur in midline structures, particularly near the corpus callosum. They can be cured by complete removal.

Arachnoid cysts are not tumors but can manifest with headache, seizures, or focal neurologic symptoms if they become large enough to compress underlying brain tissue. Many are completely asymptomatic and are found incidentally on neuroimaging. Only symptomatic cysts require removal.

PRIMARY INTRA-AXIAL TUMORS

Most primary intra-axial brain tumors are gliomas, including the astrocytomas, oligodendrogliomas, and ependymomas. Less common are medulloblastomas, other rare neuroectodermal tumors, and primary CNS lymphomas. All of these tumors have a tendency to invade brain tissue, and none can be completely excised surgically.

Glioma

DEFINITION

Astrocytomas, which are the most common glioma, are classified into one of four World Health Organization categories: grade I, the pilocytic astrocytoma; grade II, the fibrillary astrocytoma; grade III, the anaplastic astrocytoma; and grade IV, the glioblastoma. Pilocytic astrocytomas (grade I) are extremely low-grade focal tumors that are more common in children and may be associated with neurofibromatosis type 1; they are often cured by complete surgical excision. Fibrillary astrocytomas, anaplastic astrocytomas, and glioblastomas are diffuse tumors that infiltrate widely into brain; even grade II tumors progress over time, and most acquire the histologic features and growth patterns of grade III and IV tumors.

EPIDEMIOLOGY

Gliomas occur at any age, but the peak age is 20 to 30 years for an astrocytoma, 40 years for anaplastic astrocytoma, and 55 to 60 years for glioblastoma. Age is the single most important prognostic factor; younger patients live substantially longer than older patients. Histology is also critical; patients with glioblastoma do significantly worse than patients with lower-grade lesions. Performance status, duration of symptoms, and whether a complete resection has been achieved are also strong predictors of improved outcome and prolonged survival. For all grades of glioma, men are more frequently affected than women, and whites are significantly more frequently affected than blacks. Gliomas are typically single lesions, but multifocal disease is seen in approximately 5% of patients with high-grade tumors. A variant of gliomas, called gliomatosis cerebri, causes widespread infiltration of the entire brain; most patients have relatively low-grade pathologic findings on biopsy, but focal regions of high-grade transformation can exist.

PATHOBIOLOGY

At least 95% of gliomas are sporadic, and only 5% occur in patients with a family history of brain tumor. Furthermore, patients with a familial history of glioma usually do not fall into a well-recognized hereditary syndrome. However, neurofibromatosis 1 (von Recklinghausen disease; Chapter 417) is associated with an increased incidence of gliomas, particularly in the optic pathway, hypothalamus, and brainstem. Gliomas also occur with increased frequency in Turcot and Lynch syndrome, in which colorectal neoplasms are seen in association with a variety of CNS tumors. Somatic mutations of the isocitrate dehydrogenase 1 and 2 genes (*IDH1* and *IDH2*) have been identified in most low-grade gliomas and secondary glioblastomas; these patients have a better outcome than do those with wild-type *IDH* genes.[6] Molecular profiling has identified four subclasses of glioblastoma: (1) classical, defined by overactivation of the epidermal growth factor receptor pathway; (2) proneural, defined by *IDH* mutation and alterations of *PDGFRA* and expression of neural markers; (3) mesenchymal, defined by *NF1* loss and expression of mesenchymal markers; and (4) neural, characterized by expression of neuronal markers. These categories

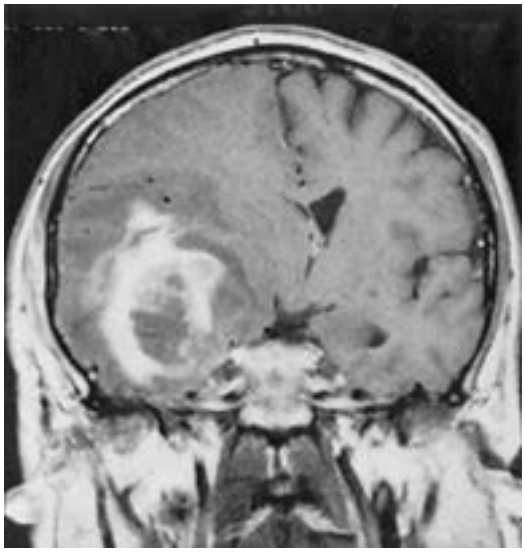

FIGURE 189-5. Temporal lobe glioblastoma. This T1-weighted gadolinium-enhanced magnetic resonance image shows a typical ring configuration of contrast material with central necrosis and marked mass effect.

define distinctly different pathways to histologically identical glioblastomas. The prognostic and therapeutic implications of these distinctions are under investigation.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients with gliomas often present with seizures, headache, and lateralizing signs such as hemiparesis, aphasia, or a visual field deficit.[7] On MRI, low-grade gliomas typically appear as diffuse, nonenhancing lesions with a propensity to occur in the frontal lobe and insular cortex. High-grade gliomas, which typically enhance with contrast material, occur in the cortical white matter and are accompanied by significant surrounding edema. Glioblastomas frequently have regions of central necrosis (Fig. 189-5), and hemorrhage can occur in 5 to 8% of patients.

TREATMENT

Rx

For all gliomas, treatment frequently involves surgery, radiation therapy, and chemotherapy. The surgical goal of complete removal of all visible disease is often impossible. A prospective, randomized trial provided evidence for the use of intraoperative MRI guidance in optimizing the extent of resection of gliomas.[A2] The adequacy of resection is best assessed on a postoperative MRI study, without and with gadolinium, performed within 72 to 96 hours after surgery. Glioma resections with intraoperative stimulation mapping are associated with fewer severe neurologic deficits and more extensive resection. Surgical removal usually improves neurologic function and reduces dependency on corticosteroids.

Anaplastic Astrocytoma and Glioblastoma

All anaplastic astrocytomas and glioblastomas should be treated with postoperative radiation therapy to a dose of approximately 60 Gy. In a randomized trial of patients with glioblastoma, the alkylating agent temozolomide (75 mg/m^2 daily), administered concurrently with radiation therapy and followed by adjuvant temozolomide (150 to 200 mg/m^2 for 5 consecutive days every 4 weeks for six cycles), significantly improved survival (median, 14.6 months) compared with radiation therapy alone (median, 12.1 months; $P < .001$), and the 2-year survival rate more than doubled to 26.5%.[A3] Patients whose tumors contained a methylated promoter of the O^6-methylguanine-DNA methyltransferase (*MGMT*) DNA repair gene benefited most from the addition of temozolomide. On the basis of these data, combined treatment is the current standard of care for patients with glioblastoma. Chemotherapy is generally well tolerated and associated with minimal toxicity. Elderly patients often do poorly, but a randomized trial showed that radiation therapy (compared with supportive care only) results in a modest improvement in survival, without reducing quality of life or cognition, in elderly patients with glioblastoma.[A4] In the Nordic randomized phase III trial, standard radiation therapy was associated with poor outcomes in elderly patients with glioblastoma, especially those older than 70 years. Both temozolomide and hypofractionated radiation therapy should be considered as standard treatment options in elderly individuals with glioblastoma.[A5] The addition of bevacizumab (an anti-vascular endothelial growth factor molecule, initially at 10 mg/kg intravanously every 2 weeks) may further improve progression-free survival.[A6][A7] Recurrences can be treated with re-resection, additional chemotherapy or, occasionally, stereotactic radiosurgery or a combination of these. Despite aggressive treatment, disease recurs in almost all patients, and the median survival time is 15 months for glioblastoma. Patients with anaplastic gliomas (including anaplastic oligodendrogliomas) had an identical median survival of about 7 years whether the initial treatment was radiation therapy alone or chemotherapy alone.[A8] However, some young patients with anaplastic astrocytoma can survive much longer before the tumor recurs.

Optic and Brainstem Glioma

Optic gliomas, which can involve the optic nerve or optic chiasm, are usually associated with neurofibromatosis type 1. These gliomas are typically pilocytic tumors that can have an indolent course including rare spontaneous regression. They are often not amenable to surgical resection, and they can have a stuttering clinical course, with periods of visual loss punctuated by prolonged periods of visual stability. When necessary, radiation therapy or even chemotherapy may be useful, but often no treatment is required. Brainstem gliomas usually involve the pons, less often the medulla or midbrain. Brainstem gliomas are most commonly seen in children in the first decade of life but can be found even in elderly people; they can have a low-grade or high-grade histology, but outcome is primarily determined by the location of the tumor. In general, most brainstem gliomas have a dismal outcome with survival of 1 year or less, but relatively benign variants occasionally occur.

Low-Grade Astrocytoma

Low-grade astrocytomas have a variable course. In patients who present with isolated seizures that can be controlled easily with antiepileptics, treatment with radiation therapy or chemotherapy immediately after surgery may prolong progression-free but not overall survival, and such patients can be monitored until there is clinical or radiographic evidence of tumor progression. However, resection should be performed at diagnosis, if feasible.[8] Patients with progressive neurologic symptoms or cognitive impairment require immediate treatment after diagnosis, and focal radiation therapy to a total of about 54 Gy is the optimal choice. For low-grade gliomas, progression-free and overall survival appear to be better when chemotherapy is added to radiation therapy. An astrocytoma can progress as a low-grade tumor or transform to a higher-grade malignant neoplasm, a change that typically is associated with the appearance of contrast enhancement on MRI. Resection or a biopsy may be necessary in these patients, followed by radiation therapy if they have not received it previously; chemotherapy with temozolomide (150 to 200 mg/m^2 for 5 days every 4 weeks for anywhere from 6 to 24 cycles) is also used. Patients with an astrocytoma have a median survival of about 5 years, but the range is wide.

Oligodendroglioma

Oligodendrogliomas occur as low-grade tumors and, less commonly, as anaplastic lesions. Treatment of these tumors differs from that of their astrocytic counterparts because oligodendrogliomas are uniquely chemosensitive due to their characteristic loss of chromosomes 1p and 19q. As with the low-grade astrocytomas, treatment should be withheld in patients with low-grade oligodendrogliomas who have no symptoms other than well-controlled seizures. Patients with progressive neurologic symptoms or radiographic progression require treatment, and initial therapy is often chemotherapy, usually with single-agent temozolomide (150 to 200 mg/m^2 for 5 days every 4 weeks for 6 to 24 cycles) or the combination of procarbazine, lomustine, and vincristine (PCV). Radiation therapy is withheld until chemotherapy fails.

By comparison, all anaplastic oligodendrogliomas require immediate treatment. The standard approach includes focal radiation therapy. Adjuvant chemotherapy significantly prolongs disease-free and overall survival in patients with codeleted tumors but not in those with intact 1p and 19q chromosomes.[A9] However, there is a growing movement toward treatment of high-grade tumors with chemotherapy alone initially[9]; this can be considered only in neurologically healthy patients. Tumor progression should be treated with re-resection, radiation therapy if it has not been previously administered, and additional chemotherapy. Patients with low-grade oligodendrogliomas have a median survival time in excess of 15 years. Median survival is about 14 years

for patients with a 1p/19q codeleted anaplastic oligodendroglioma compared with the median survival of about 3 years for patients with an intact 1p/19q.

Medulloblastoma

Medulloblastomas usually occur in the vermis of the cerebellum and principally affect children and young adults. Boys outnumber girls by about 2 : 1, and peak onset is 7 years of age; medulloblastoma in adulthood is rare and usually affects the cerebellar hemisphere.

Transcription profiles have identified four distinct subclasses that include: (1) the Wnt subtype driven by a stabilizing mutation in *CTNNB1* (β-catenin), which has an excellent prognosis; (2) the SHH subtype with mutations in *PTCH1*, *SMO*, *GLI2*, or *SUFU*, which has an intermediate prognosis; (3) group 3, which has an increased expression of *MYC* and has the worst prognosis; and (4) group 4, which is characterized by isochromosome 17q and marked male predominance. Medulloblastomas have a characteristic clinical presentation, with ataxia (due to cerebellar and brainstem involvement) and headache, nausea, and vomiting (due to increased ICP from obstructive hydrocephalus). Aggressive surgery with complete excision is strongly associated with improved outcome. Surgery is always followed by neuraxis radiation therapy. Chemotherapy with vincristine, etoposide, carboplatin, and cyclophosphamide significantly improves 5-year event-free survival from 60 to 74% but has not significantly prolonged overall survival, which is about 70 to 80% at 5 years when all patients are considered together. However, clinical high-risk patients do worse and standard-risk patients do better. It is unknown if molecular subclasses should supersede clinical stratification or dictate treatment selection. This vigorous therapy often results in delayed complications in survivors, including intellectual deficits, growth impairment, and endocrinologic dysfunction.[10] Late relapses as well as secondary neoplasms compromise long-term outcome.

Ganglioglioma

Gangliogliomas, as the name implies, possess both a glial component and a neoplastic neural component (ganglion cell). Some low-grade gangliogliomas are indolent and do not require additional treatment after surgical extirpation. Patients with anaplastic tumors may fare better than patients with malignant gliomas, but recurrence is the rule despite surgery and radiation therapy.

Primary Central Nervous System Lymphoma

Primary CNS lymphomas are associated with immunodeficiency states, particularly acquired immunodeficiency syndrome and organ transplantation, and are seen with increased frequency among the apparently immunocompetent population, in whom the median age at diagnosis is about 60 years. These tumors are usually diffuse large B-cell non-Hodgkin's lymphomas identical to systemic diffuse large B-cell lymphoma (Chapter 185). The tumor can involve the CSF, eye, and brain, where it is multifocal in about 40% of patients at presentation (Fig. 189-6). In contrast to all other brain tumors, surgical resection may not be associated with improved survival and can cause significant neurologic morbidity; therefore, biopsy is usually the preferred surgical approach. Chemotherapy is the primary treatment, and high-dose methotrexate (3 to 8 g/m^2 on alternate weeks for 3 to 12 months) is the most important chemotherapeutic agent. In most patients, radiation therapy is avoided because the necessary whole brain irradiation causes significant cognitive impairment when it is combined with chemotherapy and does not prolong survival.[A10] Corticosteroids (e.g., dexamethasone 8 to 16 mg/day), which are frequently used as part of the chemotherapeutic regimen, not only help manage the associated cerebral edema but also can cause tumor regression. With the use of multiagent chemotherapy, with or without cranial irradiation, median survival times is 3 to 5 years.

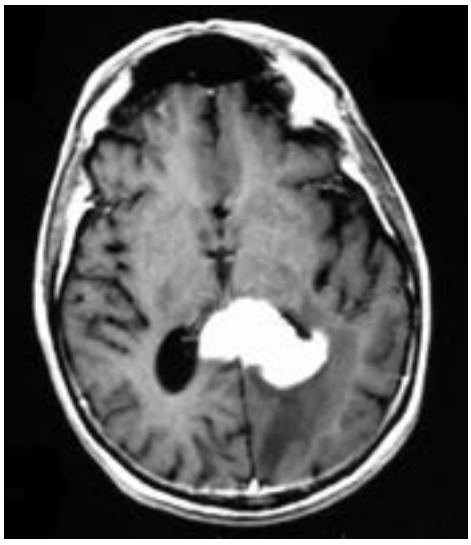

FIGURE 189-6. Primary CNS lymphoma. A post-gadolinium magnetic resonance image demonstrating a diffusely enhancing splenial lesion. The periventricular location and absence of central necrosis are characteristic of primary CNS lymphoma.

Other Intra-Axial Tumors

Rare, intra-axial cerebral tumors include the *ependymoma*, which is optimally treated with surgical excision followed by radiation therapy. *Choroid plexus papillomas* and *carcinomas* may be manifested with hydrocephalus or lateralizing signs. Resection may be sufficient for the benign papilloma, but carcinomas rapidly recur even when postoperative radiation therapy is also used. *Colloid cysts* of the third ventricle are benign tumors that can cause obstructive hydrocephalus; they may be treated with a third ventriculostomy or with resection by use of an intraventricular endoscope. *Hemangioblastomas* occur primarily in the cerebellum but can also occur in the spinal cord and the hemispheres. About 15% of patients with a hemangioblastoma have the autosomal dominant disorder von Hippel–Lindau disease (Chapter 417), which is characterized by hemangioblastomas in the CNS and retina, renal cell carcinoma, pheochromocytoma, endolymphatic sac tumors, and cysts in a variety of visceral organs. Hemangioblastomas are treated by surgical excision and require radiation therapy only for recurrence. Complete removal usually results in cure.

METASTATIC TUMOR

Brain Metastasis

DEFINITION AND EPIDEMIOLOGY

Every systemic cancer is capable of metastasizing to the brain. Melanoma (Chapter 203) has the greatest propensity to spread to the CNS, but the most common causes of CNS metastases are cancers of the breast (Chapter 198) and lung (Chapter 191), followed by cancers of the colon (Chapter 193) and kidney (Chapter 197). CNS metastases are being seen with greater frequency as patients with systemic cancers have prolonged survival with better treatments. In most patients with brain metastases, CNS disease develops late in the course of their illness, but a brain metastasis may be the initial presentation of a systemic cancer. In most of these patients, lung cancer is the primary site; in some, however, a primary site is never identified (Chapter 204).

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients with brain metastases present with progressive neurologic symptoms and signs that typically include headache, seizures, and lateralizing signs. Metastases are best diagnosed by cranial MRI with gadolinium (Fig. 189-7). All lesions can be clearly seen by MRI, which is better than CT for visualizing the posterior fossa. Metastases, which are usually well-circumscribed lesions at the gray matter–white matter junction, are often associated with extensive edema. Hemorrhage into a metastasis occurs most frequently with metastases from melanoma, renal cancer, or thyroid cancer; however, because brain metastases from lung cancer are so common, they are the type most commonly associated with hemorrhage. Sometimes, hemorrhage into a brain metastasis is best visualized by noncontrast head CT.

TREATMENT

Rx

Because brain metastases do not widely infiltrate into brain tissue and tend to have a pseudocapsule around them, they can be completely excised surgically. In randomized controlled studies, complete removal of a single brain metastasis substantially prolonged life and maintained neurologic function for a longer period. Postoperative whole brain radiation therapy significantly improves control of CNS disease after resection of a single brain metastasis, but it does not prolong survival because patients die of progressive systemic tumor. Consequently, the use of postoperative whole brain radiation therapy is frequently decided on an individual basis. If multiple lesions can be completely resected, these patients do as well as those with a single lesion that has been removed.

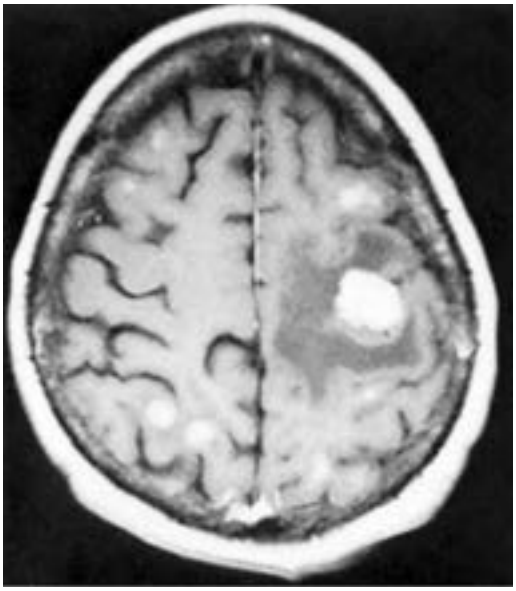

FIGURE 189-7. Brain metastasis. Multiple metastases from breast carcinoma are seen on this T1-weighted gadolinium-enhanced magnetic resonance image. The multiple smaller tumors were not visible on computed tomography, even after a contrast agent was given.

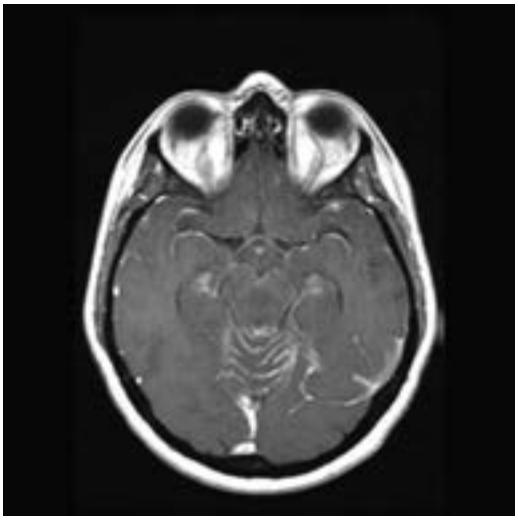

FIGURE 189-8. Leptomeningeal metastases. A post-gadolinium cranial magnetic resonance image demonstrating enhancement of all sulci of the cerebellar vermis representing tumor cells in the subarachnoid space of this patient with breast cancer.

Most patients with multiple brain metastases are best treated with a course of whole brain radiation therapy, most commonly 3 Gy in 10 fractions for a total of 30 Gy.[11] Some patients with single brain metastasis are also treated with whole brain irradiation if they are in poor general condition, have uncontrolled systemic disease, or are not good candidates for surgical treatment.

Stereotactic radiosurgery, with either a gamma knife that delivers gamma radiation from multiple cobalt sources or a linear accelerator that delivers x-rays to a highly focused area involving the tumor, has been effective for the treatment of one or a few brain metastases.[12] Most patients tolerate radiosurgery without difficulty, but the procedure is occasionally complicated by seizures or acute swelling that causes more neurologic dysfunction. Approximately 20 to 30% of patients develop radionecrosis, which may be indistinguishable clinically and on MRI from recurrent tumor. One advantage of stereotactic radiosurgery is that most of the normal brain is not exposed to the radiation.

Chemotherapy is used to treat brain metastases from only a few chemosensitive primary cancers, such as choriocarcinoma, small cell lung cancer, and, to a lesser extent, breast cancer. Because few patients have a significant response to chemotherapy, it is usually used as a last resort, although it has increasingly been employed in asymptomatic patients in whom brain metastases are identified at diagnosis on a screening MRI examination and who require chemotherapy for their systemic disease. The planned chemotherapy is often administered and the brain metastases will frequently respond in a fashion comparable to other systemic sites of disease. The choice of agents is based on the primary cancer type and the patient's prior drug exposures. Some targeted therapies have also been effective against brain metastases, such as in *BRAF* mutant melanoma, in which responses are seen after vemfurafenib.[13]

Leptomeningeal Metastasis

The brain is the most common intracranial site of metastases, but systemic cancer can spread to the dura and the leptomeninges as well. Dural metastases most commonly arise from breast (Chapter 198) or prostate (Chapter 201) cancer, frequently from a metastasis in the overlying calvaria. Metastasis to the leptomeninges often is manifested as multifocal neurologic symptoms and signs. These metastases involve the cranial nerves to cause diplopia or bulbar palsy (Fig. 189-8); the cervical and lumbar roots to cause limb pain or weakness; and the intracranial space to cause headache, nausea, vomiting, and elevated ICP. The diagnosis is established by the presence of tumor cells in the CSF, by cytologic examination or novel techniques to identify isolated cancer cells,[14] or by neuroimaging that definitively outlines tumor in the subarachnoid space (Fig. 189-9). Treatment frequently involves radiation therapy to symptomatic sites; intrathecal chemotherapy, usually through an intraventricular cannula (Ommaya reservoir); or systemic chemotherapy with agents or doses that penetrate into the CSF.

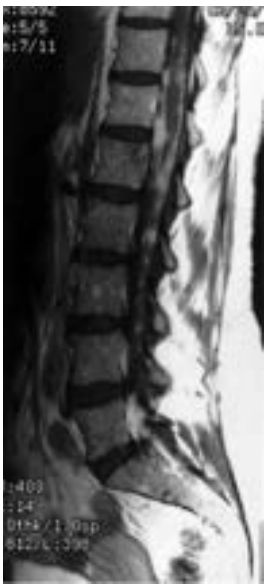

FIGURE 189-9. Leptomeningeal metastases. Gadolinium-enhanced magnetic resonance imaging of the lumbosacral spine in a patient with leptomeningeal metastases from melanoma. Multiple enhancing nodules are seen on the cauda equina, and the conus medullaris and lower spinal cord are encased by tumor.

SPINAL TUMORS

Tumors involving the spine can be classified according to the anatomic area they involve: extradural, intradural extramedullary, and intramedullary tumors (Table 189-5). Extradural tumors typically arise from the bone elements of the spine and cause neurologic symptoms and signs by spinal cord compression. Intradural but extramedullary tumors arise from the pachymeninges or nerve roots (meningiomas or schwannomas) and can cause either radicular symptoms or spinal cord compression. Intramedullary spinal cord tumors are rare; they arise from the spinal cord parenchyma and have a biology similar to that of brain tumors.

Extradural Tumors

EPIDEMIOLOGY AND CLINICAL MANIFESTATIONS

Most extradural tumors originate from a metastasis to the bone elements of the spine, typically the vertebral body and occasionally the vertebral lamina or spinous process. Less common are primary tumors of the spine, including chordoma, osteogenic sarcoma, plasmacytoma, and chondrosarcoma. Expansile growth of the bone tumor impinges on the spinal canal and, if untreated, compresses the spinal cord or the nerve roots as they exit the intervertebral foramina. Whereas most of these lesions arise from bone

TABLE 189-5 SPINAL TUMORS

Extradural
Metastasis
Primary bone tumors arising in the spine
Chordoma
Osteogenic sarcoma
Chondrosarcoma
Plasmacytoma
Intradural extramedullary
Meningiomas
Neurofibromas
Schwannomas
Lipomas
Arachnoid cysts
Epidermoid cysts
Metastasis
Intramedullary
Ependymoma
Glioma
Hemangioblastoma
Lipoma
Metastasis

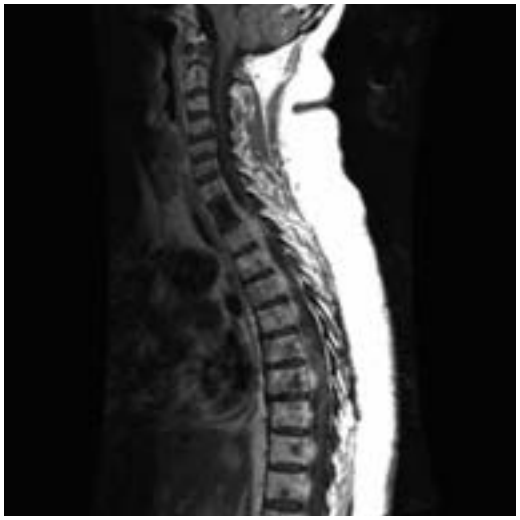

FIGURE 189-10. Multilevel epidural spinal cord compression. This patient with metastatic breast cancer has multiple areas of ventral epidural tumor seen throughout the thoracic spine on this post-gadolinium sagittal magnetic resonance image.

metastases, extradural tumors can also arise from paravertebral metastases that can grow through the intervertebral foramina and into the epidural space without affecting surrounding bone structures; very rarely, a direct metastasis to the epidural space is also seen. The most common primary cancers that cause extradural metastases are prostate cancer (Chapter 201), breast cancer (Chapter 198), and lung cancer (Chapter 191) as well as the lymphomas (Chapter 185). Hematologic malignant neoplasms may also be associated with paravertebral disease that grows through the intervertebral foramina.

Whether the mass is a primary bone tumor or a metastasis from a distant source, 98% of patients present with pain that is usually localized to the site of the tumor. Because there are more thoracic than cervical or lumbar vertebrae, the tumor and pain are likely to be in the middle or high back, a less common site for benign pain (Chapter 400). Motor impairment and sensory symptoms are present in about 50% of patients, whereas sphincter disturbances are found in only about 25% of patients. Back pain often precedes the development of any other neurologic symptom or sign, frequently by weeks and occasionally by months.

DIAGNOSIS

Severe back pain in a patient with cancer should be evaluated by MRI which does not require intravenous administration of contrast material. Plain films of the spine, bone scans, or even CT scans may show bone disease, but epidural tumor can be seen only on MRI (Fig. 189-10). Furthermore, MRI is the only technique that can reveal paravertebral or direct epidural metastasis. Patients who cannot have an MRI study should be imaged by CT with sagittal reconstruction images.

Differential Diagnosis

The differential diagnosis of extradural tumors includes epidural abscess (Chapter 413), acute or subacute epidural hematomas (Chapter 400), herniated intervertebral discs (Chapter 400), spondylosis (Chapter 400), epidural lipomatosis, and, rarely, extramedullary hematopoiesis. On occasion, a percutaneous needle biopsy or decompressive laminectomy is required to make a definitive diagnosis.

TREATMENT

Rx

Epidural metastases require immediate treatment because patients can develop acute and unpredictable neurologic deterioration resulting in paraplegia. Patients should be started on high-dose corticosteroids (usually >20 mg IV dexamethasone), which rapidly relieve pain and may contribute to neurologic recovery. Surgery followed by postoperative radiation therapy is superior to radiation therapy alone in preserving the ability to walk and may prolong survival in a wide population of patients with metastatic spinal cord compression, but its advantage may be lost in patients 65 years of age and older.[A11] It is much easier to preserve neurologic function than to reverse impairment, so clinically silent areas of extradural tumor that are detected on MRI should be treated before neurologic compromise develops. Patients with epidural metastasis can have a good neurologic outcome if they are treated before the onset of severe neurologic compromise, but their overall survival is usually short because of the presence of widespread metastatic disease. Patients whose primary tumor arises in the spine, such as an osteogenic sarcoma (Chapter 202), should undergo definitive surgery; the need for postoperative radiation therapy is based on the tumor's histology.

Intradural Extramedullary Tumors

Meningiomas

Most intradural extramedullary tumors are benign. Meningiomas are benign, slow-growing tumors that occur primarily in middle-aged women and are predominantly located in the thoracic region. Back pain is a common symptom, but about 25% of patients have no pain and present with slowly progressive neurologic dysfunction, typically a gait disorder that has been progressing, frequently for years. Spinal MRI with gadolinium clearly delineates the lesion. Surgical resection is curative, and a complete resection can usually be accomplished easily.

Nerve Sheath Tumors

Nerve sheath tumors include schwannomas and neurofibromas. Both typically arise from the dorsal root, and the first symptom is often radicular pain that precedes symptoms of spinal cord compression by months or even years. Some patients with spinal neurofibroma or schwannoma have neurofibromatosis type 1 (Chapter 417), but most do not. The diagnosis is clearly established by gadolinium-enhanced MRI of the spine. The treatment is surgical, and complete removal results in cure.

Metastases

Metastasis to the spinal leptomeninges can be manifested as an intradural extramedullary lesion. A single large tumor nodule can cause focal symptoms and signs referable to that spinal level, but in most patients, multiple levels of the neuraxis are involved, causing multifocal neurologic symptoms and signs. Cervical and lumbosacral radicular pain as well as sensory and motor loss is seen in more than half of patients. The diagnosis is established by gadolinium-enhanced MRI showing multifocal nodules or sometimes a layer of cells coating the spinal cord or nerve roots (see Fig. 189-9). If imaging is negative, the diagnosis can be established by demonstrating tumor cells in the CSF. Treatment is complicated and frequently requires radiation therapy to symptomatic sites of disease, intrathecal chemotherapy best administered through an intraventricular cannula (Ommaya device), and occasionally systemic chemotherapy. Radiation therapy can ameliorate neurologic symptoms, particularly pain, but the disease often has a relentless progressive course, resulting in death in 3 to 6 months despite aggressive therapy. Because of the diffuse nature of the disease, surgery is not an option.

Intramedullary Tumors

Intramedullary spinal cord tumors are similar to neoplasms that arise in brain parenchyma. The most common spinal cord tumors are ependymomas and

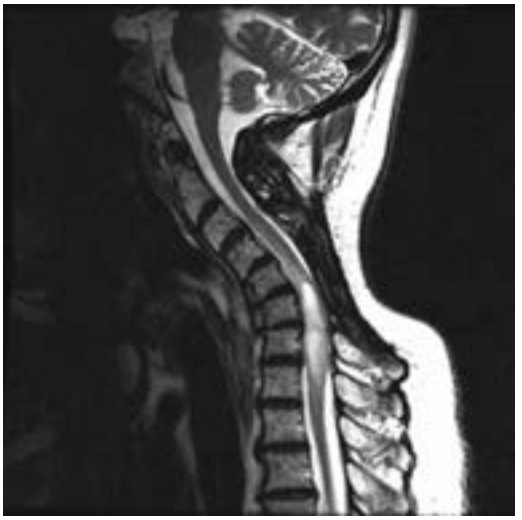

FIGURE 189-11. Spinal cord astrocytoma. T2-weighted sagittal magnetic resonance image demonstrating a lower cervical intramedullary low-grade astrocytoma; no enhancement was evident.

astrocytomas; hemangioblastomas (particularly in association with von Hippel–Lindau disease; Chapter 417), lipomas, and, rarely, intramedullary metastases are also seen.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

All intramedullary tumors have a similar clinical presentation, and pain is a common initial symptom. Signs of spinal cord dysfunction subsequently ensue and reflect the location of the lesion. In addition, some intramedullary tumors are accompanied by a syrinx (Chapter 417), which can contribute to symptoms. The classic signs of intramedullary spinal cord lesions, such as dissociated sensory loss, sacral sparing, and early sphincter problems, are not sufficiently reliable to distinguish intramedullary from extramedullary lesions on the basis of clinical findings. The diagnosis is established by gadolinium-enhanced and T2-weighted MRI (Fig. 189-11).

TREATMENT

Rx

Surgery is the first therapeutic intervention, both to obtain a definitive diagnosis and to resect the lesion. Complete resection of spinal cord tumors is possible, particularly in the case of ependymomas and hemangioblastomas. However, spinal cord tumors are rare, and only neurosurgeons experienced in removal of this type of lesion should perform the procedure. High-grade gliomas and residual ependymomas should be treated with postoperative radiation therapy.[15] Low-grade astrocytomas of the spinal cord can be treated with radiation therapy when the patient develops symptomatic neurologic impairment, but presymptomatic treatment does not prevent the development of impairment nor necessarily delay it. Intramedullary metastases do not require surgery because the diagnosis is usually straightforward; radiation therapy provides limited benefit because these patients typically have other CNS metastases.

Grade A References

A1. Mikkelsen T, Paleologos NA, Robinson PD, et al. The role of prophylactic anticonvulsants in the management of brain metastases: a systematic review and evidence-based clinical practice guideline. *J Neurooncol*. 2010;96:97-102.

A2. Senft C, Bink A, Franz K, et al. Intraoperative MRI guidance and extent of resection in glioma surgery: a randomised, controlled trial. *Lancet Oncol*. 2011;12:997-1003.

A3. Stupp R, Hegi ME, Mason WP, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol*. 2009;10:459-466.

A4. Keime-Guibert F, Chinot O, Taillandier L, et al. Radiotherapy for glioblastoma in the elderly. *N Engl J Med*. 2007;356:1527-1535.

A5. Malmström A, Gronberg BH, Marosi C, et al. Temozolomide versus standard 6-week radiotherapy versus hypofractionated radiotherapy in patients older than 60 years with glioblastoma: the Nordic randomized, phase 3 trial. *Lancet Oncol*. 2012;13:916-926.

A6. Chinot OL, Wick W, Mason W, et al. Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. *N Engl J Med*. 2014;370:709-722.

A7. Gilbert MR, Dignam JJ, Armstrong TS, et al. A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N Engl J Med*. 2014;370:699-708.

A8. Wick W, Hartmann C, Engel C, et al. NOA-04 randomized phase III trial of sequential radiochemotherapy of anaplastic glioma with procarbazine, lomustine, and vincristine or temozolomide. *J Clin Oncol*. 2009;27:5874-5880.

A9. Cairncross G, Wang M, Shaw E, et al. Phase III trial of chemoradiotherapy for anaplastic oligodendroglioma: long-term results of RTOG 9402. *J Clin Oncol*. 2013;31:337-343.

A10. Thiel E, Korfel A, Martus P, et al. High-dose methotrexate with or without whole brain radiotherapy for primary CNS lymphoma (G-PCNSL-SG-1): a phase 3, randomized, non-inferiority trial. *Lancet Oncol*. 2010;11:1036-1047.

A11. Chi JH, Gokaslan Z, McCormick P, et al. Selecting treatment for patients with malignant epidural spinal cord compression—does age matter? Results from a randomized clinical trial. *Spine*. 2009;34:431-435.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Scott BJ, Douglas VC, Tihan T, et al. A systematic approach to the diagnosis of suspected central nervous system lymphoma. *JAMA Neurol*. 2013;70:311-319.
2. Baldi I, Engelhardt J, Bonnet C, et al. Epidemiology of meningiomas. *Neurochirurgie*. 2014; [Epub ahead of print].
3. Fathi AR, Roelcke U. Meningioma. *Curr Neurol Neurosci Rep*. 2013;13:337.
4. Marosi C. Light at the end of the tunnel: towards an effective drug therapy for surgery- and radiation-refractory meningioma. *Neuro-Oncol*. 2015;17:7-8.
5. Stacchiotti S, Tamborini E, Lo Vullo S, et al. Phase II study on lapatinib in advanced EGFR-positive chordoma. *Ann Oncol*. 2013;24:1931-1936.
6. Verhaak RG, Hoadley KA, Purdom E, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010;17:98-110.
7. Omuro A, DeAngelis LM. Glioblastoma and other malignant gliomas: a clinical review. *JAMA*. 2013;310:1842-1850.
8. Jakola AS, Myrmel KS, Kloster R, et al. Comparison of a strategy favoring early surgical resection vs a strategy favoring watchful waiting in low-grade gliomas. *JAMA*. 2012;308:1881-1888.
9. Panageas KS, Iwamoto FM, Cloughesy TF, et al. Initial treatment patterns over time for anaplastic oligodendroglial tumors. *Neuro-Oncol*. 2012;14:761-767.
10. Ris MD, Walsh K, Wallace D, et al. Intellectual and academic outcome following two chemotherapy regimens and radiotherapy for average-risk medulloblastoma: COG A9961. *Pediatr Blood Cancer*. 2013;60:1350-1357.
11. Gaspar LE, Mehta MP, Patchell RA, et al. The role of whole brain radiation therapy in the management of newly diagnosed brain metastases: a systemic review and evidence-based clinical practice guideline. *J Neurooncol*. 2010;96:17-32.
12. Burke D, Mascott C, Rock L, et al. Stereotactic radiosurgery for the treatment of brain metastases; results from a single institution experience. *Ir J Med Sci*. 2013;182:481-485.
13. Soffietti R, Trevisan E, Ruda R. Targeted therapy in brain metastasis. *Curr Opin Oncol*. 2012;24:679-686.
14. Nayak L, Fleisher M, Gonzalez-Espinoza R, et al. Rare cell capture technology for the diagnosis of leptomeningeal metastasis in solid tumors. *Neurology*. 2013;80:1598-1605.
15. Oh MC, Ivan ME, Sun MZ, et al. Adjuvant radiotherapy delays recurrence following subtotal resection of spinal cord ependymomas. *Neuro-Oncol*. 2013;15:208-215.

REVIEW QUESTIONS

1. A 78-year-old woman is brought to the hospital by her family, who are complaining that she has lost interest in her usual activities and appears depressed during the past 4 to 6 months. She was started on an antidepressant by her internist 3 months ago without any response. They have also noticed that her walking has deteriorated in the past month, and in the last week she has developed incontinence. On examination, you find she is indifferent and has a flat affect and bilateral frontal release signs. She shuffles when she walks and her balance is poor. You order magnetic resonance imaging (MRI), which shows a large bifrontal diffusely enhancing tumor compressing the frontal lobes without much underlying edema; a dural tail is evident. The tumor type is likely to be

A. Glioblastoma
B. Meningioma
C. Metastasis
D. Primary CNS lymphoma
E. Anaplastic oligodendroglioma

Answer: B Meningiomas can grow to a huge size without causing problems because they grow so slowly and the underlying brain accommodates the compression. Subacute onset of social withdrawal in an older patient should always be suggestive of an underlying neurologic process, particularly when other findings, such as a gait disorder and incontinence, are apparent; all these symptoms can be caused by compression of both frontal lobes. Resection can completely remove this tumor and restore the patient to normal function, and she is unlikely to experience recurrence. (Rocha H, Cerejo A, Garrett MC, Massano J. Reversible parkinsonism due to a large intracranial tumour. *BMJ Case Rep*. 2012; pii: bcr2012007823.)

2. A 56-year-old man with lung cancer comes to the emergency department complaining of progressively severe upper back pain that can occasionally radiate around his chest. The pain increases with cough or Valsalva maneuver. He has no sensory or motor symptoms. His examination reveals no neurologic deficits. The next step in evaluating his pain is

A. Spine radiograph
B. Chest radiograph
C. Bone scan
D. Treat with pain medications and observe
E. Spine MRI

Answer: E This patient has a high likelihood of having metastases to the spinal column. The pain is located in the upper back, which is atypical for the common musculoskeletal causes of back pain, and the pain radiating around his chest suggests compression of the thoracic nerve roots; exacerbation of the pain with cough or Valsalva maneuver indicates compression on the thecal sac. Imaging is essential, so observation is not an option, and plain radiographs or bone scan cannot visualize tumor extending into the epidural space. A complete spine MRI study was obtained and demonstrated marked epidural tumor with spinal cord compression at T6. Because it was a single site of disease and the patient had few comorbidities, he had a complete resection followed by focal radiation therapy and did well.

3. A 72-year-old man has a history of prior myocardial infarction, congestive heart failure, atrial fibrillation with anticoagulation, and metastatic renal cell cancer. He was at home when he had the sudden onset of left hand twitching that progressed to involve his whole arm and the left side of his face. It stopped after 45 seconds, and he had mild weakness of his left arm and hand, which resolved completely during 12 hours. He went to the local emergency department, where MRI revealed a single enhancing mass 1×0.8 cm in the right frontal lobe with surrounding edema, consistent with a metastasis. What is the best therapeutic approach in this man with poor renal function and severe cardiac disease?

A. Supportive care only
B. Surgical resection
C. Whole brain radiation therapy
D. Stereotactic radiosurgery
E. A change in chemotherapy

Answer: D This is a single brain metastasis in a patient with a primary tumor type that is relatively radioresistant. Furthermore, he is older and likely to experience greater neurologic toxicity if his whole brain is irradiated, so whole brain radiation therapy is a poor option. Resection could be considered, but his medical condition is too poor for surgery. Chemotherapy is unlikely to be effective for a brain metastasis from renal cancer. Stereotactic radiosurgery is an excellent option. The tumor is small and therefore a good candidate for this approach. Relatively radioresistant primaries respond better to the single high-dose fraction of stereotactic radiosurgery; it is also easy to administer, and with a good response the patient is more likely to discontinue steroids and avoid their attendant toxicities. Supportive care could be a reasonable option, but given that his brain metastasis has never been treated and stereotactic radiosurgery has a high probability of being efficacious with minimal toxicity, it is the best option.

4. A 67-year-old right-handed woman comes to your office complaining of several episodes during the past 2 months of the abrupt onset of an inability to speak. She knows what she wants to say but is unable to get the words out. These episodes can last for about 1 minute and then resolve and she is normal. In addition, in the past 2 to 3 weeks, she has noticed progressive word finding difficulty and occasional word substitutions and very mild headache during the past few days. MRI reveals a contrast-enhancing mass in the left temporal lobe. She has a surgical resection, and the pathologic examination reveals a glioblastoma. The next steps in her treatment are

A. Radiation therapy with concurrent and adjuvant temozolomide
B. Radiation therapy alone
C. Temozolomide alone
D. No further treatment after resection
E. Radiation therapy, temozolomide, and bevacizumab

Answer: A Despite a complete resection, a glioblastoma requires additional therapy after surgery because there are always infiltrative cells left behind even when postoperative MRI shows that all the enhancing disease was removed. A large phase III trial has established that the standard of care for a newly diagnosed glioblastoma includes focal radiation therapy with concurrent and adjuvant temozolomide. This is clearly superior to either modality alone. Two recent studies demonstrated that the addition of bevacizumab to radiation therapy and temozolomide failed to improve survival, so it is not included in the initial regimen.

190 HEAD AND NECK CANCER

MARSHALL R. POSNER

DEFINITION

The principal cancers of the head and neck include squamous cell cancers arising from the mucosal surfaces of the upper aerodigestive tract and a diverse group of salivary gland neoplasms. Unique cancers of the region include Epstein-Barr virus (EBV)–associated nasopharyngeal carcinoma, human papillomavirus (HPV)–associated oropharynx cancer (HPVOPC), thyroid malignant neoplasms (Chapter 226), esthesioneuroblastoma, and sinonasal undifferentiated carcinoma. A variety of other cancers arise from structures and tissues in the head and neck, including the more common skin cancers (Chapter 203), lymphomas (Chapters 185 and 186), and sarcomas (Chapter 202).

EPIDEMIOLOGY

Squamous cell carcinomas account for 95% of all malignant neoplasms of the head and neck, whereas salivary gland cancers represent nearly all of the remaining 5%. They represent 4% of all malignant neoplasms in the United States. Squamous cell cancers of the head and neck can be divided into two distinct groups based on pathogenesis, biology, and prognosis. Environmentally related cancers, caused principally by tobacco and alcohol, have been declining in incidence; however, they remain common. There has been an increasing incidence of HPV-related oropharynx cancer.[1] HPVOPC now represents about 75% of oropharynx cancers seen in the United States and Europe.[2] HPVOPC affects a younger population (50 to 60 years) than environmental cancers do (55 to 65 years). HPVOPC patients are also generally healthier and are not prone to comorbid illnesses and second cancers seen in environmentally related squamous cancers.

The mucosal surfaces of the head and neck are divided into six anatomic regions: the oral cavity, oropharynx, hypopharynx, larynx, nasopharynx, and paranasal sinuses. The site of anatomic origin for a squamous cell carcinoma of the head and neck has important albeit imperfectly defined implications for diagnosis, pathogenesis, spread, prognosis, and treatment. This is because of intrinsic differences in the biology of the mucosal cells and subsequent cancers at the sites of origin, carcinogenic viral tropisms, and differences in lymphatic drainage patterns and proximity to other structures in this compact region.

Oral Cavity

The oral cavity includes the floor of the mouth, anterior or oral aspect of the tongue, lips, buccal surfaces, hard palate, retromolar trigone, and gums. This region is easily appreciable by physical examination, and thus tumors in this area can frequently be detected early in their course. Tumors of the oral cavity, which are strongly related to the use of smokeless tobacco and other

oral tobacco products (Chapter 32) and, in southern Asia, to betel nut and pan chewing, appear on the mid oral tongue and buccal and gingival surfaces in the sites where these products are held in contact with the mucosa for long periods. Anterior tongue cancers are more common in smokers. Lip cancers are particularly prevalent in transplant recipients and can be caused by DNA damage from solar ultraviolet light.

Oropharynx

The oropharynx consists of the tongue from the circumvallate papillae posteriorly to the epiglottis, the tonsils, the associated pharyngeal walls, and the soft palate. The oropharynx has become the most common location for head and neck tumors in the United States and is a common site of origin in Europe. This is due to a high rate of HPVOPC, which continues to increase in incidence.[3] HPVOPC is caused almost exclusively by HPV-16, a high-risk HPV type associated with cervical, anal, and vulvar cancers. Other high-risk HPV types account for 10 to 15% of new diagnoses. High-risk HPV types are transmitted in body fluids and infect squamous mucosal surfaces of the anogenital tract and the oropharynx. Whereas smoking does not increase the risk of HPVOPC, 50% of smokers with oropharynx cancer have HPV as a cause. Compared with environmental cancers, HPVOPC often presents with lower primary T stage (T1 and T2) and higher nodal stage (N2 and N3), and it is frequently a cause of cancer of unknown primary origin (Chapter 204) because of small and difficult-to-identify primary tumors.[4]

Hypopharynx

The hypopharynx comprises the piriform sinuses, the lateral and posterior pharyngeal walls, and the posterior surfaces of the larynx. These structures surround the larynx posteriorly and laterally. Tumors in this region can be difficult to detect because of the recesses and spaces surrounding the larynx. As a result, primary hypopharyngeal tumors may be asymptomatic and, like oropharyngeal tumors, may initially be recognized in an advanced state or diagnosed as an unknown primary (Chapter 204). These tumors are associated with tobacco (Chapter 32) and alcohol (Chapter 33) use.

Larynx

The larynx includes the vocal cords, the subglottis, and the supraglottic larynx as well as the thyroid, cricoid, and arytenoid cartilages. Tumors arising in the true vocal cords are frequently symptomatic early and rarely spread beyond the confines of the larynx, whereas subglottic and supraglottic cancers can be relatively asymptomatic and have a much higher and earlier risk of spread to the lymphatics and regional sites. Laryngeal cancers are strongly associated with smoking (Chapter 32).

Nasopharynx

The nasopharynx includes the mucosal surfaces and structures of the cavity behind the nasal passages. Nasopharyngeal cancers are common in the Pacific Rim, northern Africa, and the Middle East. In some areas of China and Southeast Asia, nasopharyngeal cancers occur with a frequency that rivals that of lung cancer. In North America, there are about 2000 cases each year, but numbers are increasing as high-risk ethnic populations settle in North America. Nasopharyngeal cancers are frequently associated with the presence of latent infection of the epithelial tumor cells by EBV, the etiologic agent of infectious mononucleosis (Chapter 377). Nasopharyngeal cancers are also associated with both environmental and genetic factors in susceptible populations that have migrated to North America and remain at high risk for this disease. Unlike other squamous cell carcinomas of the head and neck, nasopharyngeal cancers can occur at an early age, with a distinct peak in adolescence and young adulthood. Nasopharyngeal cancers are categorized into three histologic subtypes by the World Health Organization (WHO): the undifferentiated (WHO III) and nonkeratinizing forms (WHO II) are latently infected with EBV in 95% of cases and represent the majority of cases in North America and worldwide; the well-differentiated (WHO I) form is rarer and represents about 5% worldwide but about 15 to 25% of all nasopharyngeal cancers in North America, and it is usually associated with traditional risk factors such as smoking. Nasopharyngeal cancers have a high risk of early regional lymph node involvement, a prolonged natural history, and a very high risk of spread to distant sites.

Paranasal Sinuses

The paranasal sinuses comprise the maxillary, ethmoid, sphenoid, and frontal sinuses as well as the nasal cavity. These are relatively rare locations for tumors of the head and neck in North America, but there is an unexplained higher rate of malignant sinus disease among the Japanese. Squamous maxillary sinus cancers are more common in smokers. Up to 50% of cancers of the sinuses may be of salivary gland origin. Adenocarcinomas have often been related to exposure to dust from woodworking, tanning, or leather working (Chapter 19). On occasion, neuroendocrine tumors and the rare sinonasal undifferentiated carcinoma are found. Sinus cancers are frequently diagnosed late in their course at the time of symptomatic invasion of surrounding structures, including the orbit, nasal cavity, base of the skull, and cranial nerves.

Salivary Glands

Salivary glands occur in all the regions described as well as in the trachea and esophagus. Tumors can develop in all of the major and minor salivary glands with an incidence that is roughly proportional to the quantity of glandular tissue. The most common single site is the parotid. Although tumors can develop at any age, including childhood, the peak incidence is between 55 and 65 years of age. Salivary gland cancers have diverse histologic findings and manifest different behavior on the basis of their histologic classification. A substantial fraction of parotid salivary tumors can be benign. Risk factors for salivary gland cancers are poorly understood, but previous radiation therapy in adjacent areas increases the risk (Chapter 20).

PATHOBIOLOGY

Tobacco products and alcohol are major etiologic and risk factors for squamous cell carcinoma of the head and neck. Both show a clear dose response. Any irritating smoked product increases the risk for local cancer, but nicotine in tobacco as well as in other tobacco leaf components directly affects the oral mucosa and increases the risk for squamous cancer (Chapter 32). Alcohol is also a carcinogen, and alcohol consumption (Chapter 33) as well as its direct application in mouthwashes is associated with an increased risk. Moreover, alcohol affects local and systemic detoxification enzymes and may increase the carcinogenic potential of other environmental carcinogens. Other environmental risk factors include radiation exposure and solar radiation; welding, metal refining, diesel and wood stove exhaust, and asbestos exposure; chronic irritants; vitamin A deficiency; and immunosuppression.

Carcinogenic viruses are responsible for the increasing incidence of head and neck cancer in the United States and Europe. Pathogenic HPV subtypes, most frequently HPV-16, independently account for approximately 75% of oropharyngeal cancer cases. HPV is transmitted by epithelial contact and body fluids. There is increased risk of HPVOPC associated with increasing numbers of individual sexual partners, although the majority of HPVOPC patients do not have sexual activity above the average. HPV DNA can be found in the tumor cells, and the oncogenic viral proteins E6 and E7 are frequently expressed and are responsible and necessary for the growth and survival of the cancer cells by affecting critical signaling pathways. Men have a three-fold greater incidence of HPVOPC than women do, although the cause for this imbalance in incidence is not known. Of great importance, HPVOPC patients have a three-fold better prognosis than that of patients with environmentally related cancers. This is due to three factors: fewer comorbid illnesses in patients with HPVOPC; few second cancers caused by tobacco and alcohol; and increased sensitivity of HPVOPC tumors to treatments. The majority of nasopharyngeal cancer is caused by another more complex virus that infects keratinocytes, EBV (Chapter 377). The classic nasopharyngeal cancer, which is also called *lymphoepithelioma*, is associated with a brisk lymphoid infiltrate that can be confused with a lymphoma. Careful examination reveals the malignant epithelial cells, with EBV detectable in the tumor cells. EBV is also rarely associated with epithelial tumors of the oropharynx, tonsil, and salivary gland.

Several inherited diseases and genetic abnormalities are associated with the development of head and neck cancer.[5,6] Fanconi anemia (Chapter 165), a rare disorder of a family of related gene products, has been linked to the development of tongue cancer, as has Cowden syndrome (Chapter 181), which is associated with mutation of the *PTEN* gene. *NOTCH1* may function as a tumor suppressor gene rather than as an oncogene in head and neck squamous cell carcinoma. Finally, common inherited allelic variants of the alcohol dehydrogenase and P-450 genes may be associated with increased susceptibility to alcohol and other environmental carcinogens.

The development of environmentally related squamous cell carcinoma of the head and neck is a multistep process in which early genetic changes evolve into frank malignancy. In environmentally related cancers, an abnormal premalignant clone of mucosal cells may be localized to a single site within the head and neck, or clones may occur independently in many sites. The

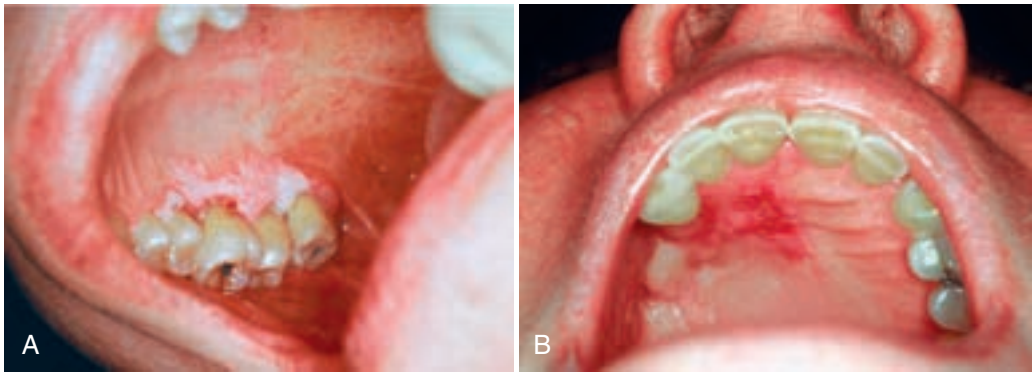

FIGURE 190-1. High-risk early mouth lesions. A, Oral leukoplakia. **B,** Oral erythroplakia.

pathogenesis of HPVOPC is less well understood. Second cancers are rare with HPVOPC in the short term, and long-term risk is unknown. In contrast, about 20% of patients with environmentally caused cancers will develop a second primary cancer, most commonly in the head and neck, lung, and esophagus; 5% of patients are initially seen with a synchronous second primary. In environmentally related cancers, the cell cycle is dysregulated by the early loss of p16, an inhibitor of cyclin D1, or by upregulation of cyclin D1; p53 is disabled through a number of mechanisms preventing programmed cell death; mitogenic signaling is enhanced by upregulation of epidermal growth factor (EGF) receptor function; cyclooxygenase 2 is overexpressed, thereby inhibiting apoptosis and promoting angiogenesis; and chromosomal instability with aneuploidy develops. Many of these early molecular and functional changes occur without obvious alteration in the physical appearance of the oral mucosa, although leukoplakia can occur. In HPVOPC, the RB and p53 pathways are inactivated by the HPV oncogenic proteins E6 and E7. As a result, the p16 protein is upregulated as a biomarker of HPV tumor origin. As a consequence of RB and p53 inactivation, these patients have dysregulation of cell growth and DNA damage control/programmed cell death, respectively. The differences between the genetic and molecular determinants of cancer in HPVOPC and environmentally related cancers can be differentially targeted for therapy.

In environmentally related cancers, high-risk early lesions can occasionally be identified as leukoplakia and erythroplakia. Leukoplakia (Fig. 190-1A) is diagnosed clinically as a white patch of mucosal tissue in the oral mucosa or larynx. It can unpredictably progress to cancer during a period of several years in approximately 30% of patients. Erythroplakia (Fig. 190-1B), a red hyperkeratotic change in the mucosa, is a more advanced premalignant lesion with an approximately 60% rate of progression to oral cancer. Surgical resection of leukoplakia or erythroplakia has no effect on the subsequent development of invasive cancer. There is no proven chemopreventive therapy for persons with oral premalignant lesions. Because continued smoking or alcohol consumption increases the risk for recurrence and second primaries dramatically, patients with prior environmentally related cancers should stop alcohol and tobacco use.

CLINICAL MANIFESTATIONS

The symptoms and clinical manifestations of tumors in the head and neck can vary broadly and are related to the structures at the site of the primary tumor as well as regional lymph node drainage. Small tumors of the oral cavity and larynx can be easily appreciated because of physical self-discovery or early compromise of the function of a critical structure. As a result of the propensity for squamous cell carcinoma of the head and neck to remain a local and regional disease, it is unusual for this cancer to be associated with abnormalities outside the head and neck. Salivary gland malignant neoplasms are less constrained and frequently spread distantly; however, because the primary tumors are also frequently accessible to direct physical examination and early discovery, it is still uncommon to identify these tumors as a result of metastatic spread outside the region.

Clinical manifestations of *tumors in the oral cavity* include a painless lump, a painful mass or ulcer, or simple thickening of the mucosa. Small lesions in the lateral aspect of the tongue and the floor of the mouth can cause pain referred to the mandible, gums, and ear because of the shared sensory nerves supplying these areas. Antibiotics can relieve symptoms and even reduce the size of a tumor or lymph nodes when superficial infection and inflammation are contributing to the pain; however, recurrent or continued pain in an adult should trigger early suspicion about more ominous disease. Speech may be affected late if the tumor causes restricted tongue motion or cranial nerve XII dysfunction. Gingival tumors can loosen teeth and invade the mandible along tooth sockets.

In true vocal cord cancer, hoarseness and other forms of voice change are common and expected early symptoms, but they may be later manifestations of *supraglottic and subglottic laryngeal tumors*, which become relatively large without affecting the voice. Tumors of the piriform sinus can affect the voice when they become large and impair the recurrent laryngeal nerve or are associated with deep local invasion; pain in the ear or pain on swallowing referred to the ear is also a common and important feature of these tumors. Adults with ear pain or persistent hoarseness should be referred to an otolaryngologist for evaluation. Because this posterior area is difficult to assess directly, primary tumors are frequently missed in routine office examinations. *Tumors of the supraglottic region, subglottic cancers, and cancers of the piriform sinus* can also be manifested as acute, emergency airway obstruction. Frequently, patients have a history of wheezing and mild upper airway distress in the period leading up to the emergency situation. On occasion, such findings are confused with adult-onset asthma.

A middle ear infection or effusion in an adult should also prompt an ear, nose, and throat evaluation. Nasopharyngeal cancer may be manifested as an ear infection in young adults. Hemoptysis or epistaxis may be the only clue to a nasopharyngeal cancer or a *paranasal sinus tumor*. Cranial nerve findings from deep invasion of the base of the skull are late events and include lateral gaze abnormalities, diplopia, facial pain, or facial nerve paralysis. *Sinus tumors* can also be associated with these later findings, although nasal stuffiness occurs frequently and can be confused with sinusitis. New and persistent symptoms of sinusitis or facial pain should raise suspicion of sinus cancer and prompt an evaluation.

Tumors in the tonsil or base of the tongue can cause local pain and referred ear pain; however, they are frequently asymptomatic and can attain a large size before becoming evident as a result of changes in speech ("hot potato voice"), a sense of globus, trismus, or restriction of tongue movement. Manifestation as a painless lump in the neck is increasingly common with the increased incidence of HPVOPC. Tumors of the tonsil or base of the tongue may also lose their mucosal component, not be seen or felt on direct inspection, and occur as a solid or cystic neck mass. Isolated neck masses can wax and wane with antibiotics. A mass, especially a cystic mass, in the neck in an adult is cancer and specifically HPVOPC until proved otherwise and should prompt an ear, nose, and throat evaluation and positron emission tomography (PET)/computed tomography (CT) imaging, fine-needle aspiration, and examination under anesthesia before excisional biopsy.

The staging of squamous cell carcinoma of the head and neck is based on the TNM (tumor, node, metastasis) staging system, and prognosis is related primarily to the N and T stages. The risk of the cancer's spreading to lymph nodes is directly related to the location of the primary and secondarily to the size of the primary. Tumors of the oropharynx have a high risk for nodal metastases, followed in risk by the supraglottic larynx and piriform sinus (hypopharynx), oral portion of the tongue, soft palate, oral cavity and floor of the mouth, and larynx. Nasopharyngeal cancer is often associated with extensive nodal spread, whereas paranasal sinus cancers rarely spread to the lymph nodes. The location of lymph node spread is determined in part by site. Nasopharyngeal cancer spreads to the posterior cervical lymph nodes as

well as to the high cervical nodes. Oropharynx, larynx, and piriform sinus tumors spread to the high cervical nodes. Nodal metastases from these locations can be bilateral. Oral cavity tumors spread to the submental nodes and submandibular nodes. Spread tends to be orderly from the submandibular nodes to the midcervical nodes. Oral cavity cancers can have as high as a 20% risk of clinically unappreciated contralateral spread.

DIAGNOSIS

The relative accessibility of the head and neck to direct inspection makes physical examination critical for diagnosis and staging. Patients with localized symptoms or a sign such as an ulcer or a small mass should have a thorough head and neck office examination performed by their primary physician and by a specialist, including inspection of the visible structures and palpation of the base of the tongue and tonsil areas as well as the neck. Specialized office examination with fiberoptics should be included in the preliminary assessment. Regardless of whether cancer is suspected, excisional biopsies should be discouraged because margins are frequently violated and inadequate, thereby leading to larger re-excisions. A simple punch biopsy is sufficient for diagnosis, particularly in the oral portion of the tongue where tumors can spread readily through lymphatics.

When cancer is highly suspected and before definitive surgical intervention, PET/CT of the body and a high-resolution CT scan from the base of the skull to the clavicles, preferably with the spiral technique, are indicated. Magnetic resonance imaging (MRI) provides added information in evaluating soft tissue involvement, especially in the base of the tongue and the parapharyngeal spaces and for sinus tumors. MRI can distinguish between soft tissue masses and retained secretions, whereas PET/CT and high-resolution CT are more helpful in assessing nodal spread in the neck, and CT is effective in identifying extracapsular nodal extension and bone invasion, which are important prognostic and clinical findings. PET scanning provides an important adjunct to CT scanning and can identify occult disease. PET is particularly helpful when the patient has an "unknown primary tumor," before biopsy, to guide the evaluation and to reduce the risk of inadequate diagnostic and premature therapeutic procedures (Fig. 190-2).

When a biopsy indicates cancer or cancer is highly suspected, an examination under anesthesia with endoscopy can be performed to stage the primary tumor before definitive therapy is undertaken. This procedure, which may be part of definitive therapy, provides information about the extent of disease, the appropriateness of the planned definitive procedure, and the presence of second primaries. It is an absolute requirement before definitive therapy can be completed. Endoscopy and palpation under anesthesia can identify unexpected local spread or a synchronous second primary (found synchronously in about 5% of patients with environmentally related cancers), which can substantially alter the treatment plan.

Approach to the Patient with an Unknown Primary Site

Patients frequently seek care from their primary physician because of an enlarged lymph node, a cystic mass, or a collection of lymph nodes in the upper part of the neck (Fig. 190-3). Such masses in an adult should be considered cancer until proved otherwise. Masses in the supraclavicular areas usually derive from primary tumors below the clavicles, and masses in the midneck and cervical regions are almost always from the head and neck. Identification of a primary site is critical to focus therapy, to reduce morbidity, and to determine prognosis.

The most common primary sites for painless lumps are the oropharynx (base of the tongue and tonsil) and piriform sinus. Oropharynx cancers are frequently due to HPV, and a positive HPV or EBV-encoded RNA (EBER) finding in the biopsy specimen or fine-needle aspirate is presumptive evidence of oropharynx or nasopharynx origin, respectively. Salivary gland cancers, lymphomas, melanomas, and skin cancers can also occur in this manner. Bilateral nodal disease or nodal disease with systemic symptoms may suggest lymphoma. By comparison, pain, warmth, and erythema may suggest an infectious etiology. Intraparotid nodes most likely represent metastases from skin malignant neoplasms. Physical examination should include a careful investigation for primary skin cancers. CT, PET, and MRI should be part of the initial evaluation. Fine-needle aspiration with HPV and EBER

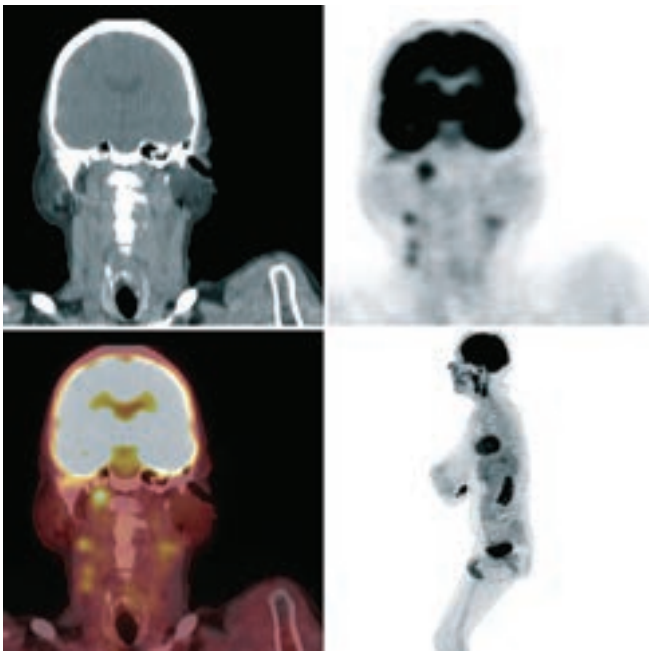

FIGURE 190-2. Positron emission tomography and computed tomography fused images. A primary human papillomavirus–positive base of tongue cancer and neck adenopathy are shown.

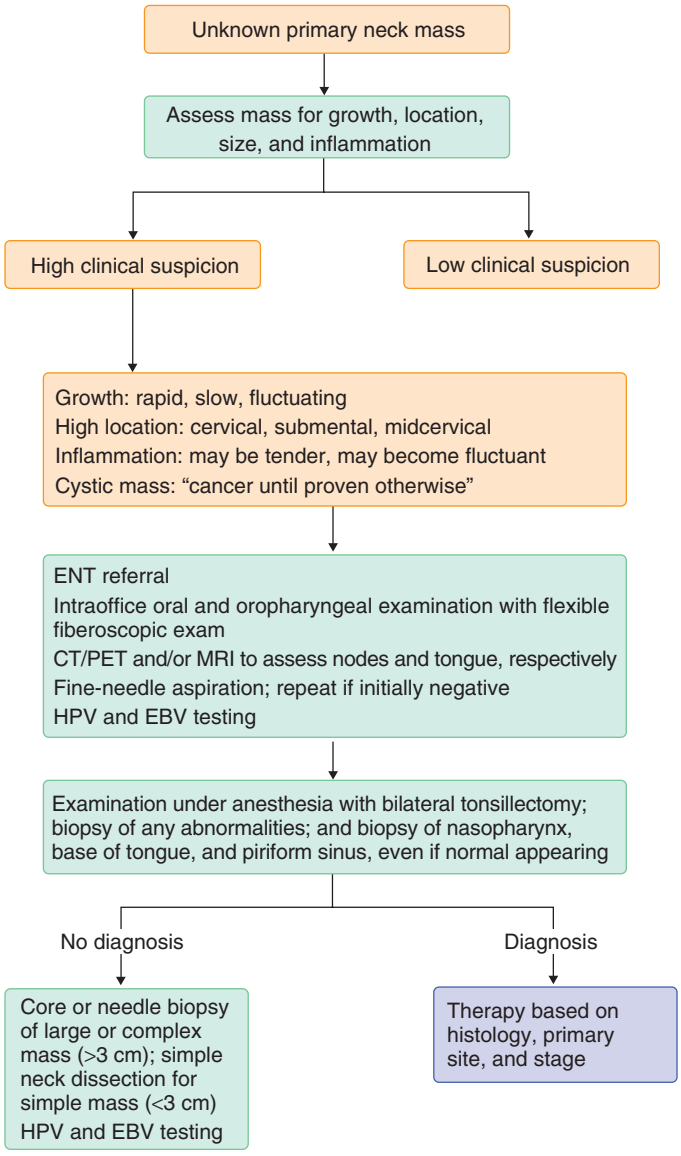

FIGURE 190-3. Evaluation of an unknown primary neck mass. CT = computed tomography; EBV = Epstein-Barr virus; ENT = ear, nose, and throat; HPV = human papillomavirus; MRI = magnetic resonance imaging; PET = positron emission tomography.

testing for squamous tumors should be performed. CT-guided biopsy may be indicated if the mass is difficult to approach. If squamous cells are identified, the tumor is most likely a squamous cell carcinoma of the head and neck. Next, endoscopy under anesthesia should be performed with bilateral tonsillectomy and directed biopsies of any abnormalities, areas of firmness, and the base of the tongue, nasopharynx, and ipsilateral piriform sinus, even if they appear normal. Core or excisional (single node <3 cm) biopsy of the lymph node should be performed if the pathologic findings are equivocal and a primary site is not confirmed. Neck dissection can be accomplished if a primary site is not identified and the patient has an N1 or small N2a/b manifestation. Some unknown primaries with squamous histology are never identified. Currently, HPV and EBV are the only molecular markers known to distinguish head and neck cancer from skin or salivary gland squamous cancer. EBV positivity indicates a nasopharyngeal cancer, and HPV an oropharyngeal primary. Although p16 immunohistochemistry is often used as a surrogate for HPV testing, it is not adequate for a final therapeutic decision to be made and may be positive in up to 20% of non-HPV cancers.

In contrast to squamous cell carcinoma of the head and neck, salivary gland cancers are heterogeneous in their natural history and treatment. The three most common histologic types are adenoid cystic carcinoma, mucoepidermoid cancer, and adenocarcinoma. Other histologic types include the aggressive salivary ductal cancer and squamous cell cancers, whereas less aggressive histologic varieties include adenocarcinoma ex pleomorphic adenoma and acinic cell carcinoma. Because adenoid cystic carcinoma travels along nerves and can spread hematogenously, careful assessment of the cranial nerves and the chest by CT is indicated before major surgery is undertaken. Patients should also be evaluated for bone and liver metastases. PET scan may not be positive in acinic cell carcinoma because of slow proliferation and metabolism in the malignant cells. Formal lymph node dissection is not indicated. Ethmoid and sphenoid sinus adenoid cystic carcinomas are locally and regionally aggressive and require specialized surgery and radiation therapy techniques for local and regional control. The behavior of mucoepidermoid carcinoma is determined by histology. Low- and intermediate-grade lesions rarely metastasize. Isolated high-grade tumors spread to local lymph nodes and by hematogenous routes and carry a high risk for the development of lung metastases. Local therapy should be directed at local and regional control with lymph node dissection. Radiation therapy is indicated for close microscopic margins or lymph node involvement. Adenocarcinoma, salivary ductal cancers, and squamous cell carcinoma are poor prognosis lesions with aggressive local and distant behavior. These tumors should be evaluated in the same fashion as aggressive mucoepidermoid carcinomas. Salivary ductal carcinomas may be positive for overexpression of EGFR2 or androgen receptor and should be tested for these markers to guide therapy with targeted agents. Acinic cell carcinoma and carcinoma ex pleomorphic adenoma are relatively rare. They have a propensity for local and regional recurrence if they are not removed in toto. Metastases are rare in acinic cell carcinoma and tend to be slow growing.

Other Tumors of the Head and Neck

Lymphomas in the head and neck frequently are manifested either as nodal disease in the neck or as tumor involving the lymphoid tissues of Waldeyer's ring (Chapters 185 and 186). A primary head and neck cancer may later develop in patients with lymphoma as a consequence of past exposure to tobacco, radiation therapy, or immunosuppression. The tonsil is a preferred site for mantle cell and undifferentiated lymphomas. Mucosa-associated lymphoid tissue lymphomas can affect the salivary glands.

In the context of an isolated neck mass, a systematic evaluation (see Fig. 190-3) should be undertaken, even in young adults without a smoking history. The sinonasal T-cell and natural killer cell lymphomas, also known as *lethal midline granulomas*, represent a unique family of lymphomas of the head and neck. These lymphomas are associated with EBV infection (Chapter 377). Solitary, extramedullary plasmacytoma can also occur in the nasopharynx or paranasal sinuses (Chapter 187).

Sarcomas that arise in the head and neck include osteogenic sarcomas (Chapter 202) and nerve sheath tumors. Paragangliomas, which are rare malignant tumors of the chief cells of nerve paraganglia, can be extensive, multicentric, and vascular. Rhabdomyosarcomas, which have a predilection for the orbit and sinuses, occur in younger persons; the prognosis tends to be better for tumors of the head and neck than for other locations. Olfactory neuroblastomas or esthesioneuroblastomas invade the nasal cavity and base of the skull.

Many skin tumors, including melanoma and squamous cell cancer, can be accompanied by adenopathy of the neck or parotid area (Chapter 203). An unusual skin appendage tumor, Merkel cell cancer, can be confused with other neuroendocrine epithelial tumors. Merkel cell tumors are associated with HIV infection and may, in up to 50% of cases, be caused by Merkel cell polyomavirus.

PREVENTION AND TREATMENT

Rx

Selection of a treatment program for an individual patient is based on three factors: (1) the primary site and stage of the tumor; (2) the patient's comorbid conditions, including performance status and preferences; (3) and the biology of the tumor (Table 190-1). Early-stage lesions, T1N0 and T2N0, are defined by their size, and their prognosis is site specific. For example, early *larynx cancer* involving the true vocal cords has an excellent prognosis and can be treated by local excision. Voice-preserving larynx conservation surgery is effective for selected patients. Radiation therapy is equally effective for early cancer. When there is a risk of lymph node spread, radiation therapy must be given postoperatively, and the primary value of surgery is significantly diminished. Intensity-modulated radiation therapy allows radiation to be delivered in a more conformal manner to the tumor and areas at risk while sparing critical structures such as the spinal cord and noncritical but important structures such as the salivary glands and swallowing structures. Intensity-modulated radiation therapy is now a standard of care for almost all patients with head and neck cancer.[7]

Oral tongue, piriform sinus, and environmentally related oropharynx tumors have a poor prognosis and are difficult to stage accurately because of submucosal spread or lymphatic involvement. Stage I and stage II cancers are cured with local and regional surgery or radiation therapy in 70 to 90% of cases. Surgery may be preferred for oral cavity and anterior lesions. In surgically treated patients, those with a positive margin, two or more positive lymph nodes, or extracapsular spread have a significantly poorer survival rate (<30%) at 5 years. Perineural invasion and lymphovascular invasion may also be associated with a poor prognosis. Postoperative cisplatin-based chemoradiotherapy improves local and regional control as well as has a trend to increased survival, and it should be given to patients with a poor prognosis if their condition permits. At present, aside from HPV status and p16 immunohistochemistry, no molecular or immunohistochemical finding definitively adds to the information gleaned from pathology, staging, and performance status.[8]

TABLE 190-1 GENERAL APPROACH TO SQUAMOUS CELL HEAD AND NECK CANCER

STAGE	TNM	DISEASE-SPECIFIC SURVIVAL	TREATMENT APPROACH	SPECIAL CONDITIONS
I	T1N0	85-95%	Surgery or radiation therapy	Consider organ function and long-term toxicity
II	T2N0	75-90%	Surgery, radiation therapy, or chemoradiotherapy	Consider organ function Combined modality treatment for high-volume tumor Postoperative chemoradiotherapy for poor prognostic findings on pathologic staging
III	T3N0 T1-3N1	50-75%	Combined modality treatment	Primary chemoradiotherapy or TPF induction therapy or sequential therapy for organ function Postoperative chemoradiotherapy More aggressive approach (sequential therapy) for high-volume disease or hypopharynx tumors
IV	T1-3N2-3 T4N0-3 Any M1	20-60%	Combined modality treatment	Combined modality therapy Limited surgery Postoperative chemoradiotherapy Palliative therapy for M1 (curative therapy for isolated lung metastases)

When organ preservation and function are issues for stage III or stage IV cancers or when radiation therapy is required regardless of surgical outcome, primary chemoradiotherapy or sequential therapy should be considered. The curative treatment of intermediate (stage III, T1-3N1, T3N0) and locally advanced (stage IV, T1-3N2-3, T4) disease remains controversial. Long-term (3 years) survival rates in patients with stage III disease are generally between 50 and 75%, whereas only 15 to 50% of stage IV patients survive for 3 years. Intermediate-stage tumors are usually resectable, but organ preservation may be an important consideration. In many of these cases, a combined modality approach that includes chemotherapy and radiation therapy is the standard of care.

Patients with anterior lesions may do better with initial surgical treatment. The oral cavity is easy to access and is relatively forgiving for surgery and reconstruction; postoperative radiation therapy or chemoradiotherapy can be moderated in the absence of poor prognostic features. For intermediate and advanced tumors of the oral cavity, newer microvascular surgical techniques can substantially improve functional outcome and may help increase local-regional control. Radiation therapy or chemoradiotherapy remains a necessary adjunct to prevent recurrence. In contrast, tumors of oropharyngeal base of tongue or hypopharynx are almost always more extensive than is clinically appreciated, and functional outcome can be compromised by surgery followed by chemoradiotherapy. These may be more suited to a nonsurgical regional and systemic approach. In addition, patients with rapidly growing tumors are more suitable for a combined modality approach. Patients with extensive N2 or N3 nodal disease (stage IV) should be considered relatively unresectable because of a poor prognosis from regional recurrence and distant metastases. Certain locations, such as the nasopharynx and posterior pharynx, should also be considered for definitive radiation therapy, sequential therapy, or chemoradiotherapy.

Surgical therapy has changed radically in the last 5 years. Microvascular reconstructive techniques have improved outcomes in the oral cavity and have substantially reduced functional morbidity and permitted previously morbid resections to be accomplished with good functional outcomes. Transoral laser microdissection and transoral robotic surgery have created the opportunity to operate on previously "inoperable" tumors of the oropharynx and hypopharynx. Surgery with these technologies is performed without bystander tissue damage, which often leads to complications and required prolonged hospitalizations. These technologies are being integrated into combined modality approaches.

Radiation therapy has been proved by randomized trials to yield better local control and disease-free survival if it is given in twice-daily fractionated treatments rather than as daily therapy. However, the absolute benefit of hyperfractionated radiation therapy at 5 years is only 3 to 4%, and a twice-daily schedule is not advantageous with chemotherapy or better than chemotherapy and standard once-daily approaches. Studies strongly support the notion that altered fractionation radiation therapy with chemotherapy is substantially less efficacious or more toxic than standard fraction chemotherapy and that chemoradiotherapy with standard fractionation is more efficacious than altered fractionation alone. Proton beam radiation therapy has become available for tumors of the base of skull or those close to the eyes or the optic chiasm and is suitable in that context.

Induction chemotherapy is the delivery of chemotherapy before definitive local-regional treatment. Sequential therapy adds chemoradiotherapy (see later) to induction chemotherapy. Induction chemotherapy with docetaxel (75 mg/m^2), cisplatin (75 to 100 mg/m^2 by intravenous bolus), plus 5-fluorouracil (750 to 1000 mg/m^2/day for 4 to 5 days by intravenous infusion), repeated every 3 or 4 weeks (TPF), is an effective, standard regimen. For patients with advanced oropharynx, larynx, and hypopharynx tumors, sequential chemotherapy with chemoradiotherapy using concomitant carboplatin and once-daily radiation therapy improves survival and preserves function compared with radiation, surgery, or cisplatin and 5-fluorouracil chemotherapy (PF).

Chemoradiotherapy integrates chemotherapy and radiation therapy together and has led to significant improvements in overall survival in patients with advanced disease compared with radiation therapy alone. For example, patients with unresectable disease who received cisplatin (100 mg/m^2 by intravenous bolus) every 3 weeks during radiation therapy have significantly better survival than do those treated with radiation therapy alone. In a trial of patients with oropharyngeal carcinoma, those treated with carboplatin and 5-fluorouracil plus simultaneous radiation therapy had significantly better survival than did those treated by radiation therapy alone. Cetuximab plus radiation therapy has also proved effective in improving survival in patients with locally advanced head and neck cancer compared with radiation therapy alone in a single trial. Compelling data for an improvement of cetuximab or equivalence with cisplatin-based chemoradiotherapy in either toxicity or survival has not been produced. Conventional chemoradiotherapy results in more favorable outcomes than chemotherapy with accelerated radiation therapy or very accelerated radiation therapy alone in patients with locally advanced head and neck carcinoma.

Patients with locally advanced or unresectable disease (or both) should receive chemotherapy and radiation therapy as part of a combined modality approach. Surgery may be integrated into this approach. Organ preservation should be offered to patients who can tolerate the treatment and participate in the post-treatment rehabilitation.

Treatment of *tumors of the paranasal sinuses* is a special case. They rarely metastasize, and treatment should focus on surgical resection with postoperative radiation therapy or chemoradiotherapy for resectable stage III and stage IV disease and on chemoradiotherapy for local and regional control of unresectable disease. Proton beam irradiation may be more suited for tumors in and around the base of the skull, optic chiasm, orbits, and brain.

Follow-up

Patients need lifelong follow-up. Surveillance examinations for second primaries and recurrences should be performed monthly to bimonthly in the first year and then less frequently over time. PET/CT scan can provide evidence of local-regional persistence approximately 12 to 16 weeks after completion of radiation therapy and can be used to guide early salvage surgery and neck management. Biannual PET/CT scan as postoperative surveillance can identify early metastases or recurrence and increase the rate and success of salvage surgery. Treatment failure after 5 years is uncommon, but second primaries and distant metastases may continue to be identified in environmentally related cancers and HPVOPC, respectively. It is important to counsel these patients to avoid tobacco products.

During radiation therapy and immediately after radiation therapy, patients benefit from pain medications, local anesthetics, mucolytics, and saline mouthwash. Patients must avoid alcohol-containing preparations or irritants. Long-acting agents such as fentanyl or time-release narcotics should be added when needed (Chapter 30). A percutaneous endoscopic gastrostomy feeding tube is often effective for maintaining weight, improving healing, and managing nutrition during radiation therapy. Because depression is a major problem, psychiatric support and antidepressants may be very helpful. Salivary function improves during more than 4 years after radiation therapy, but most improvement occurs in the first 2 years. Pilocarpine and cevimeline (Evoxac) are effective stimulants of salivary flow in about 20% of patients.

Long-term sequelae of radiation therapy include dependence on a feeding tube in patients treated with aggressive chemoradiotherapy or radiation therapy alone. Attention should be paid to preservation of swallowing function by means of training in speech and swallowing as well as by dilation in selected patients. Hypothyroidism occurs in up to 50% of patients and as early as 3 months after treatment. Patients should be monitored by determining serum thyroid-stimulating hormone levels at regular intervals and then treated as appropriate (Chapter 226). Dental failure is a common problem. Patients must be counseled to see their dentists regularly for cleanings and to obtain fluoride therapy daily for dental preservation. Patients are at substantial lifelong risk for complications from dental manipulations after radiation therapy. Bone necrosis is painful, can be confused with recurrent tumor, and requires vigorous antibiotic therapy, débridement, and possibly hyperbaric oxygen to promote healing. Late vascular compromise of the carotid artery should lead to routine carotid ultrasound beginning about 10 years after radiation therapy.

Patients with recurrent disease, a second primary, or metastatic disease must be evaluated for potential curability. Symptomatically, persistent pain may be the most important indicator of a recurrence, and repeated biopsy should be considered when a suspicious lesion is observed. If patients have a recurrence or second primary, curative treatment options are defined by their current stage, their previous therapy, and the interval from their original therapy. Patients who have previously been treated with surgery but not radiation therapy can undergo surgery and chemoradiotherapy as part of a curative treatment plan. Patients with a surgically treatable recurrence in an irradiated field should undergo surgery as appropriate. Surgical salvage may cure as many as 30% of patients with recurrent oral cavity, larynx, or hypopharyngeal tumors. The surgery must encompass the entire recurrence. A repeated course of radiation therapy or chemoradiotherapy is also acceptable in selected patients.

Patients who are incurable can be managed effectively with palliative therapy to improve quality of life and survival (e.g., tracheostomy for airway control, laryngectomy for pain and aspiration, percutaneous endoscopic gastrostomy tube for feeding). These maneuvers can improve comfort and care in appropriate patients.

Palliative chemotherapy can provide meaningful benefit to some patients. Response rates with single agents are generally poor, and combination therapy offers higher response rates (30 to 50%). The combination of a platinum (cisplatin or carboplatin) plus 5-fluorouracil with an anti-EGF receptor antibody, cetuximab, significantly improves survival, response rate, and progression-free survival compared with the same chemotherapy without cetuximab.

Salivary Gland Tumors

In contrast to squamous cell carcinoma of the head and neck, salivary gland cancers are heterogeneous in their natural history and treatment; however, the mainstays of therapy for these tumors are surgery and radiation therapy. Early symptoms of local-regional recurrence include cranial nerve dysfunction

and progressive pain. A PET scan may be useful in distinguishing recurrence from the neuropathy that may result from radiation therapy.

There are no highly active agents or combinations for treatment of metastatic salivary gland tumors with the exception of anti-Her2 therapy for Her2-positive tumors. Local therapy can include surgical removal of isolated metastases, radio frequency ablation, and radiation therapy. Response rates are generally in the range of 20 to 35%, but prolonged responses are occasionally seen.

Future Directions

Antibodies that target newly identified molecular targets are being evaluated and may improve local and regional control as well as survival when they are delivered in combination with other therapies. Therapeutic vaccines may improve outcomes in EBV nasopharyngeal carcinoma and HPVOPC, and preventive vaccines in adolescents may prevent later malignant disease. New immune checkpoint blockade inhibitors appear to be showing promise in viral and environmentally caused head and neck cancers.

PROGNOSIS

The prognosis for patients with squamous cell carcinoma of the head and neck (see Table 190-1) is directly related to the presence of HPV, stage, and performance status. The risk for recurrence declines dramatically at 3 years after definitive treatment, and survival and possible cure can be defined after 5 years. HPV status[9] and then N (nodal) stage are the most important prognostic indicators of potential recurrence, with T (tumor) stage and smoking history being next.[10] Stage I patients (T1N0) have a 90% likelihood of tumor control, whereas stage II patients (T2N0) have greater than 70 to 85% tumor control. Tumor control in stage III patients (T1-2N1, T3N0-1) is site dependent, is HPV status and smoking history dependent, and varies from 35 to 95%. Patients with stage IVa and IVb environmentally related cancers (T1-3, N2-3, or T4NX) have a 20 to 50% tumor-specific 5-year survival rate compared with 60 to 90% 5-year survival for HPVOPC. Poor prognostic signs in advanced-stage (IVb) patients are related to N3 nodal disease, extracapsular extension, and invasion of basic structures (carotid artery encasement, base of the skull, pterygoid muscles). Patients with M1 disease are categorized as stage IVc. Patients with single lung metastases, whether as a second primary or as an isolated recurrence, can be cured. Cures with metastatic disease caused by HPVOPC may be seen after aggressive management. Patients with recurrent disease and no curative options have a median survival of 6 to 9 months.

Distant metastases occur in about 15 to 20% of patients, but this rate is increasing as better local and regional control prolongs survival in patients with locally advanced disease. Oropharyngeal, tonsil, and piriform sinus tumors have the highest risk for distant metastases. A single synchronous lung metastasis in a patient at initial evaluation or at follow-up can be cured in about 20% of cases.

Salivary gland cancers vary in behavior, depending on their histology. Adenocarcinoma, salivary ductal cancer, salivary squamous cell cancer, and high-grade mucoepidermoid cancer not only spread to lymph nodes but also spread rapidly hematogenously. The presence of lymph node metastases signals a high risk for distant metastases. Adenoid cystic carcinoma infrequently involves lymph nodes but spreads along nerves. Regional recurrences along cranial nerves are frequent and associated with "skip" lesions. Adenoid cystic carcinoma is also associated with the late development of lung metastases, but these patients can have a prolonged lifespan lasting more than 20 years. Low-grade mucoepidermoid cancer and acinic cell carcinoma have little risk of distant spread and are more notable for local recurrence if they are not completely removed.

Grade A References

A1. Cooper JS, Zhang Q, Pajak TF, et al. Long-term follow-up of the RTOG 9501/intergroup phase III trial: postoperative concurrent radiation therapy and chemotherapy in high-risk squamous cell carcinoma of the head and neck. *Int J Radiat Oncol Biol Phys*. 2012;84:1198-1205.

A2. Forastiere AA, Zhang Q, Weber RS, et al. Long-term results of RTOG 91-11: a comparison of three nonsurgical treatment strategies to preserve the larynx in patients with locally advanced larynx cancer. *J Clin Oncol*. 2013;31:845-852.

A3. Pointreau Y, Garaud P, Chapet S, et al. Randomized trial of induction chemotherapy with cisplatin and 5-fluorouracil with or without docetaxel for larynx preservation. *J Natl Cancer Inst*. 2009;101:498-506.

A4. Blanchard P, Bourhis J, Lacas B, et al. Taxane-cisplatin-fluorouracil as induction chemotherapy in locally advanced head and neck cancers: an individual patient data meta-analysis of the meta-analysis of chemotherapy in head and neck cancer group. *J Clin Oncol*. 2013;31:2854-2860.

A5. Bonner JA, Harari PM, Giralt J, et al. Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. *Lancet Oncol*. 2010;11:21-28.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Panwar A, Batra R, Lydiatt WM, et al. Human papillomavirus positive oropharyngeal squamous cell carcinoma: a growing epidemic. *Cancer Treat Rev*. 2014;40:215-219.
2. Garbuglia AR. Human papillomavirus in head and neck cancer. *Cancers (Basel)*. 2014;6: 1705-1726.
3. Gillison ML, Broutian T, Pickard RK, et al. Prevalence of oral HPV infection in the United States, 2009-2010. *JAMA*. 2012;307:693-703.
4. D'Souza G, Gross ND, Pai SI, et al. Oral human papillomavirus (HPV) infection in HPV-positive patients with oropharyngeal cancer and their partners. *J Clin Oncol*. 2014;32:2408-2415.
5. Rothenberg SM, Ellisen LW. The molecular pathogenesis of head and neck squamous cell carcinoma. *J Clin Invest*. 2012;122:1951-1957.
6. Szyfter K, Wierzbicka M, Hunt JL, et al. Frequent chromosomal aberrations and candidate genes in head and neck squamous cell carcinoma. *Eur Arch Otorhinolaryngol*. 2014;[Epub ahead of print].
7. Gregoire V, Jeraj R, Lee JA, et al. Radiotherapy for head and neck tumours in 2012 and beyond: conformal, tailored, and adaptive? *Lancet Oncol*. 2012;13:e292-e300.
8. Schache AG, Liloglou T, Risk JM, et al. Evaluation of human papilloma virus diagnostic testing in oropharyngeal squamous cell carcinoma: sensitivity, specificity, and prognostic discrimination. *Clin Cancer Res*. 2011;17:6262-6271.
9. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med*. 2010;363:24-35.
10. Gillison ML, Zhang Q, Jordan R, et al. Tobacco smoking and increased risk of death and progression for patients with p16-positive and p16-negative oropharyngeal cancer. *J Clin Oncol*. 2012;30: 2102-2111.

REVIEW QUESTIONS

1. The most important prognostic biomarker in squamous cell cancer of the head and neck is

A. Cytomegalovirus
B. Human papillomavirus
C. Herpes simplex virus
D. Merkel cell tumor virus
E. All of the above

Answer: B Human papillomavirus (HPV) is the single most important prognostic biomarker in squamous cell cancer of the head and neck. Patients with HPV cancers have a three-fold better survival than that of patients with cancers caused by environmental factors. Patients with HPV-related cancers also have fewer comorbid illnesses and less risk of second primaries.

2. The patient and the physician should discuss the possibility of organ-preserving therapy for patients with hypopharynx and larynx cancers. Some of the accepted strategies for organ preservation are

A. Chemoradiotherapy
B. Partial laryngectomy
C. Sequential therapy
D. Radiation therapy
E. All of the above

Answer: E All three treatments are reasonable larynx-preserving treatments, depending on stage, comorbidities, and lymph node involvement. For early-stage tumors in which radiation therapy may be avoided, partial surgery is indicated. For more advanced cancers, radiation therapy only or a combination of chemotherapy and radiation therapy is indicated. Laryngectomy is a reasonable therapy when the patient cannot tolerate treatment, is noncompliant, or already has a nonfunctional larynx.

3. Postoperative adjuvant cisplatin-based chemoradiotherapy is absolutely indicated for

A. Lymphovascular invasion
B. A positive margin
C. Perineural invasion
D. Extracapsular lymph node extension
E. Both B and D

Answer: E After surgery, pathologic findings of a positive margin and extracapsular lymph node extension are absolute indications for postoperative cisplatin-based adjuvant chemoradiotherapy if the patient can tolerate therapy. Lymphovascular invasion and perineural invasion are less robust criteria and can justify adjuvant chemoradiotherapy. Postoperative chemoradiotherapy is associated with a significant improvement in locoregional control and a trend to improved survival at 10 years.

4. What therapeutic biomarker is associated with prognosis in salivary ductal carcinoma?

A. EGFR1 expression
B. Her2 expression
C. Androgen receptor
D. *PTEN* deletion
E. Both B and C

Answer: E Salivary ductal carcinomas can express either Her2 or androgen receptor (AR). Her2-positive salivary ductal carcinoma will respond to Her2-specific therapies and combinations with long survival, and similarly there are data to support use of AR antagonists in AR-positive salivary ductal carcinoma. No other biomarkers have been identified in salivary gland tumors.

5. An adult presenting with a persistent cervical neck mass should be presumed to have

A. A throat infection
B. A cancer until proven otherwise
C. A branchial cleft cyst
D. Mononucleosis
E. Both B and C

Answer: B A neck mass, cystic or otherwise, presenting in an adult should be presumed to be cancer until proven otherwise. Even a normal finding on fine-needle aspiration is insufficient to rule out malignant disease. Patients should be worked up extensively before neck dissection.

191 LUNG CANCER AND OTHER PULMONARY NEOPLASMS

FADLO R. KHURI

BRONCHOGENIC LUNG CANCER

DEFINITION

Lung cancer, or bronchogenic carcinoma, is a proliferative malignant neoplasm arising from the primary respiratory epithelium. Lung cancer is generally divided into two major histologic groups: *non–small cell lung cancer* (NSCLC), which accounts for approximately 85% of all lung cancers, and *small cell lung cancer* (SCLC). There are several other less common pulmonary neoplasms including carcinoid tumors, primary soft tissue sarcomas of the lung, pulmonary blastomas, and lymphoma.

EPIDEMIOLOGY

Lung cancer is by far the leading cause of cancer-related mortality globally, with an estimated 1.3 million new cases diagnosed worldwide each year, accounting for nearly 12% of all cancers and an estimated 1.1 million deaths each year. Among men, lung cancer is the most common malignant neoplasm (incidence rate of 35.5 per 100,000), whereas in women, lung cancer incidence (12.1 per 100,000) is next only to breast, cervix, and colon cancers. The incidence and mortality related to lung cancer in men have declined during the last two decades in Western countries but continue to increase in the developing world; in women, lung cancer deaths are increasing in most regions of the world. The most dramatic increases in lung cancer incidence and death globally are in China, which has experienced a 465% increase in lung cancer–related deaths during the past 30 years.

Risk Factors

Cigarette smoking is the most common risk factor for lung cancer, with roughly 85% of lung cancer patients having a tobacco-smoking history and approximately 50% being former smokers (defined as free from smoking for at least 12 months before diagnosis). The risk for development of lung cancer correlates with the number of cigarettes smoked per day and the cumulative duration of smoking time. Patients with a smoking history of at least 20 to 30 pack-years (defined as 1 pack per day of cigarettes for 20 to 30 years) are at substantially increased risk for development of lung cancer. Since the release of the first U.S. Surgeon General's Report on the Hazards of Smoking in 1964, the prevalence of cigarette smoking has declined considerably in the United States but continues to increase at an alarming rate in developing and third world countries. As a result, the number of cases of lung cancer diagnosed annually is likely to rise during the next few decades, and it is estimated that the majority of lung cancer cases will occur outside the United States and Europe by the year 2030. Smoking cessation is associated with a gradual reduction in risk for development of lung cancer, although it does not reach that of a never-smoker. Second-hand exposure to smoke is another risk factor that contributes to nearly 1% of all cases of lung cancer.

Because only about 11% of heavy smokers develop lung cancer, genetic susceptibility to lung cancer also appears to play a role. Patients with a family history of early lung cancer (before 60 years of age) have a two-fold higher risk for development of the disease. Women appear to be at a higher risk for development of lung cancer at the same smoking exposure level as that of men, but the reasons behind this remain unclear. In recent years, an increasing number of never-smokers have been diagnosed with lung cancer.

TABLE 191-1 HISTOLOGIC MARKERS IN NON–SMALL CELL LUNG CANCER

	PERCENTAGE IHC-POSITIVE CASES AMONG HISTOLOGIC CARCINOMA SUBTYPES			POSITIVE AND NEGATIVE PREDICTIVE VALUE OF ANTIBODY PANEL	
	Adenocarcinoma (n = 215)	*Squamous Cell Carcinoma (n = 123)*	*Large Cell Carcinoma (n = 22)*	*Adenocarcinoma*	*Squamous Cell Carcinoma*
p63	7.0	99.2	52		88.9, 99.5
Cytokeratin 5/6	9.8	99.2	68		84.9, 99.5
TTF-1	83.5	3.4	23	97.7, 76.9	
Cytokeratin 7	97.2	23.5	77	88.4, 93.6	
Mucin	43.4	13.4	0		

Modified from Sterlacci W, Savic S, Schmid T, et al. Tissue-sparing application of the newly proposed IASLC/ATS/ERS classification of adenocarcinoma of the lung shows practical diagnostic and prognostic impact. *Am J Clin Pathol.* 2012;137:946-956.
IHC = immunohistochemically; TTF-1 = thyroid transcription factor-1.

The etiology behind this is unclear at this time. These individuals are more likely to harbor certain genetic alterations in the tumor, such as mutations in the gene encoding epidermal growth factor receptor (EGFR) and rearrangement in the gene encoding anaplastic lymphoma kinase (ALK). Occupational exposure to asbestos leads to an estimated four-fold higher risk of lung cancer, with cigarette smoking having an additive effect on risk. There is a latency of several decades between asbestos exposure and the development of lung cancer, and risk is related to the duration of exposure as well as to the quantity and the type of asbestos fiber. The Environmental Protection Agency and the World Health Organization consider all forms of asbestos to be carcinogenic; accordingly, the use of asbestos is banned in nearly 50 countries.

Radon exposure has also been implicated in the development of 5 to 8% of lung cancer cases.[1] Household exposure to radon, which results from the radioactive decay of uranium, is high in certain geographic regions. The Environmental Protection Agency recommends that the household radon level be less than 4 picocuries/liter of air, and simple remedial methods are available to reduce radon exposures above this threshold. Exposure to ionizing radiation in the form of therapeutic radiation or frequent diagnostic radiographic tests is also associated with a higher risk for development of lung cancer, as to a lesser degree are exposures to metals such as arsenic, nickel, and chromium as well as to silica and general air pollution, including biomass fuels such as coal and wood smoke. A study estimated that human immunodeficiency virus is associated with an increased risk for development of lung cancer with a hazard ratio of 3.6.

PATHOBIOLOGY

Pathology

Lung cancer is broadly subdivided into NSCLC and SCLC on the basis of the distinct biologic behavior and response to chemotherapy of these two subsets. NSCLC comprises *adenocarcinoma*, *squamous cell carcinoma*, and *large cell carcinoma* subtypes. In the past several years, distinct differences between the various histologic subtypes that comprise NSCLC have been recognized, with an increasing emphasis placed on the identification of histologic subtype from diagnostic specimens.

Adenocarcinoma is now the most common histologic subtype of lung cancer; never-smokers who develop lung cancer most frequently have adenocarcinoma. It has gradually increased in incidence, surpassing squamous cell cancer during the past two decades, and now represents nearly 50% of all newly diagnosed cases of lung cancer in the United States. Adenocarcinoma has a higher predilection for distant metastasis compared with squamous cell histology. In 2011, a new classification system for lung adenocarcinoma was developed dividing adenocarcinomas into preinvasive, minimally invasive, and invasive types.[2] *Atypical adenomatous hyperplasia* refers to a localized proliferative lesion consisting of atypical type II pneumocytes or Clara cells and measuring less than 5 mm. *Adenocarcinoma in situ* refers to lesions smaller than 3 cm that lack any invasive characteristics. This entity was previously referred to as bronchioloalveolar carcinoma or noninvasive adenocarcinoma. Lesions 3 cm or smaller with a predominantly lepidic pattern and with invasion of less than 5 mm in greatest dimension are referred to as *minimally invasive adenocarcinoma*. Adenocarcinoma in situ and minimally invasive adenocarcinoma have a more than 95% 5-year survival rate when they are treated with surgical resection, making the establishment of a precise pathologic diagnosis of great significance. Invasive adenocarcinoma represents nearly 90% of all cases of adenocarcinoma. Based on the predominant characteristic features, it is categorized as lepidic, acinar, papillary, micropapillary, or solid predominant with mucin production.

Squamous cell lung cancer is decreasing in incidence in the United States, most likely because of the changing smoking habits of the population. Squamous tumors of the lung are generally centrally located and are almost always seen in patients with a significant smoking history. Squamous dysplasia and squamous cell carcinoma in situ are preinvasive lesions that can develop into invasive cancers.

In addition to morphologic features, immunohistochemical studies are important in establishing NSCLC histologic subtype. Adenocarcinoma specimens usually stain positive for cytokeratin 7 and thyroid transcription factor-1 (TTF-1) and are negative for cytokeratin 20. The majority of squamous cell tumors stain positive for p40 and p63, members of the p53 family of proteins, whereas adenocarcinomas occasionally stain positive for p63. On the basis of these findings, a panel of markers including TTF-1, p63, and p40 is frequently evaluated in diagnostic specimens of patients with lung cancer to accurately identify the histologic subtype (Table 191-1). Large cell carcinoma represents 3 to 4% of NSCLC and is characterized by a high mitotic rate, necrosis, and morphologic features of NSCLC. Large cell tumors stain positively for neuroendocrine markers such as chromogranin A and synaptophysin. Because this histologic subtype is often difficult to accurately diagnose owing to an abundance of necrotic tissue and a poor degree of differentiation, diagnosis requires an adequate tissue specimen. Large cell carcinoma is often associated with an aggressive clinical course and poor survival rates, even when it is found in the setting of early-stage disease. Large cell carcinoma is strongly associated with a history of prior smoking.

SCLC is diagnosed in approximately 13% of lung cancer cases in the United States, and its incidence has gradually declined during the past three decades. SCLC is strongly associated with smoking and is rare in never-smokers. Pathologic diagnosis can be challenging because of an abundance of necrotic tissue but is established by characteristic features, such as a high degree of mitosis and necrosis. Diagnostic work-up of SCLC includes immunostaining for TTF-1, chromogranin, synaptophysin, and CD56. Approximately 15% of SCLC specimens have mixed morphology with components of NSCLC.

Molecular Pathology

In recent years, a number of molecular abnormalities have been identified in lung cancer. Many of these represent novel targets for therapy, strengthening the rationale for obtaining adequate tumor tissue to conduct molecular studies as an essential component of the diagnostic work-up for lung cancer. With modern genomic techniques, a greater understanding of the molecular features that account for the long-recognized clinical heterogeneity of lung cancer is leading to individualized treatment approaches.[3]

Oncogenes

In lung adenocarcinoma, nearly two thirds of patients harbor an oncogenic mutation that can potentially be targeted with specific agents (Fig. 191-1A).[4] The most common are mutations involving *KRAS*, *EGFR*, *BRAF*, *HER2*, and *PIK3CA* and gene rearrangements involving *ALK*, *RET*, and *ROS1*. *KRAS* mutations are present in approximately 25% of lung adenocarcinoma patients and are usually associated with cigarette smoking. The most common sites of mutation in *KRAS* include codons 12, 13, and 61, resulting in amino acid substitutions, which cause impaired GTPase activity and constitutive

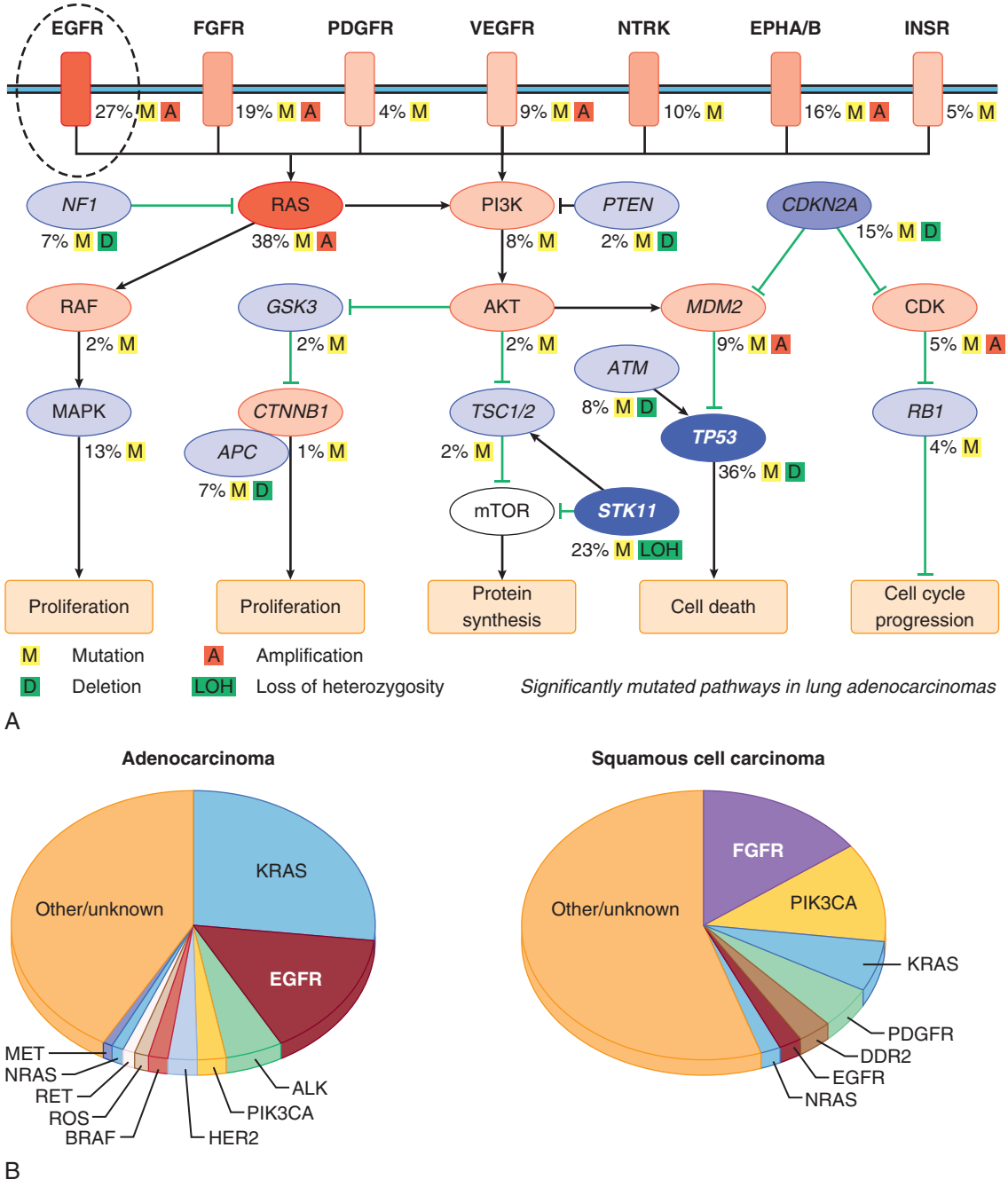

FIGURE 191-1. Altered signaling networks in lung cancer. A, Significantly mutated pathways in lung adenocarcinoma. (Modified from Ding L, Getz G, Wheeler DA, et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*. 2008;455:1069-1075.) **B,** Genetic profiles by histologic subtype. (From the Lung Cancer Mutation Consortium and modified from Sequist LV, Heist RS, Shaw AT, et al. Implementing multiplexed genotyping of non-small-cell lung cancers into routine clinical practice. *Ann Oncol*. 2011;22:2616-2624; Bergethon K, Shaw AT, Ou SH, et al. ROS1 rearrangements define a unique molecular class of lung cancers. *J Clin Oncol*. 2012;30:863-870; Weiss J, Sos ML, Seidel D, et al. Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. *Sci Transl Med*. 2010;2:62ra93; Hammerman PS, Sos ML, Ramos AH, et al. Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer. *Cancer Discov*. 2011;1:78-89.)

activation of RAS signaling. The prognostic value of *KRAS* mutation in patients with lung cancer is controversial.

Mutations in *EGFR* are observed in nearly 15% of white and almost 40% of Asian lung adenocarcinoma patients. Deletion mutations in exon 19 and a point mutation in exon 21 are located in the tyrosine kinase–binding domain of the receptor and result in constitutive activation of the signaling pathway, leading to proliferation, evasion of apoptosis, and enhanced angiogenesis. Patients with *EGFR*-activating mutations can derive robust and durable clinical benefit from treatment with EGFR tyrosine kinase inhibitors (TKIs). However, most of the benefit is limited in duration, and within 12 to 24 months, nearly 60% of these patients will develop a secondary mutation in exon 20 that confers resistance to EGFR TKI therapy. This mutation can also be found de novo in certain patients with lung adenocarcinoma along with an exon 19 or 21 mutation before exposure to EGFR TKI therapy. Another common mechanism of acquired resistance to EGFR TKI therapy is amplification of the growth factor receptor c-Met. In approximately 5% of patients with lung adenocarcinoma, gene rearrangement involving *ALK* is observed. Clinical features associated with the *ALK* gene rearrangement include never-smokers, adenocarcinoma histology, signet ring features on histopathologic evaluation, and younger age. The fusion gene results from inversion or translocation of portions of the *echinoderm microtubule-associated protein-like 4 (EML4)* with the *ALK* gene and leads to activation of downstream signals that can be inhibited by specific ALK kinase inhibitors. Crizotinib, an ALK inhibitor, induces objective tumor response in nearly two thirds of patients. *ALK* gene rearrangement is detected by fluorescent in situ hybridization, often in addition to immunohistochemistry. Other fusion abnormalities involving the *RET* and *ROS1* genes are each observed in 1% of lung adenocarcinoma specimens. It is noteworthy that *EGFR* and *KRAS* mutations

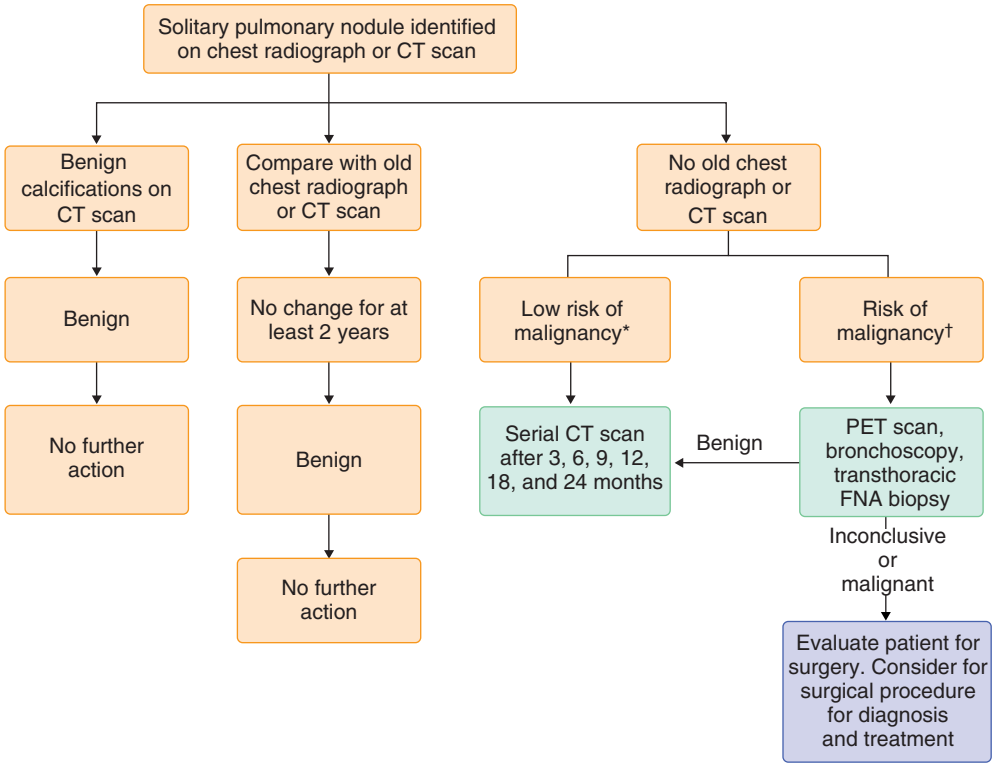

FIGURE 191-2. Evaluation of a patient with a solitary pulmonary nodule. *Patient with a minimal or absent history of smoking and other known risk factors for the development of lung cancer and a nodule 8 mm or smaller. †Patient with a history of smoking and other risk factors for the development of lung cancer and a nodule 8 mm or larger. CT = computed tomography; FNA = fine-needle aspiration; PET = positron emission tomography.

and *ALK* gene rearrangements are usually mutually exclusive. Squamous cell carcinoma has an entirely different spectrum of molecular abnormalities (Fig. 191-1B). Recent studies from the Cancer Genome Atlas (TCGA) project indicate common mutations in *p53*, *PTEN*, *PIK3CA*, *KEAP1*, *DDR2*, and *RB1*. Amplification of the gene for *fibroblast growth factor receptor* (FGFR) is also noted in 10 to 20% of squamous cell lung cancers. Many of these abnormalities provide potential opportunities for targeted therapies. The availability of highly sophisticated genomic sequencing has permitted the elucidation of hitherto unidentified molecular abnormalities, uncovering new therapeutic targets for lung cancer. Increasingly, the performance of "multiplex" testing for a number of molecular markers simultaneously, using limited amounts of tumor tissue, is changing the therapeutic paradigm for NSCLC. Guidelines from the International Association for the Study of Lung Cancer recommend routine testing for *EGFR* mutation and *ALK* translocation for all newly diagnosed patients with lung adenocarcinoma.[5] For patients with tumors of squamous cell histology, routine molecular testing is not yet recommended, although randomized clinical trials using this approach are in progress. In the absence of Food and Drug Administration–approved, molecularly directed therapies for squamous cancers, the use of standard chemotherapies is recommended for this disease at present.

Tumor Suppressor Genes

The function of multiple tumor suppressor genes is frequently lost in lung cancer, including *p53*, *Rb*, *LKB1*, and a number of genes found on the short arm of chromosome 3 (*3p*). *p53* mutation or loss correlates with cigarette smoking and has been detected in some preneoplastic lesions of the lung. Mutations of *p53* are common in both NSCLC (~50%) and SCLC (~80%). Mutations in *LKB1* are also common in NSCLC. The *STK11/LKB1* gene, which encodes a *serine/threonine kinase*, regulates cell polarity and functions as a tumor suppressor. One of the earliest genetic abnormalities in lung cancer occurs during the deletion of genetic material on chromosome 3p (p14-p23). The deletion occurs in approximately 50% of NSCLC and 90% of SCLC patients. The *FHIT* (fragile histidine triad) gene (3p14.2), which is abnormal in many lung cancers, may function as a tumor suppressor gene by limiting tumor growth and enhancing apoptosis. The Rb protein is not expressed in 90% of SCLC because of mutation or deletion. In NSCLC, Rb is normally expressed, but when Rb is phosphorylated, uncontrolled cell division can occur.

Epigenetics

Epigenetics refers to a change in gene expression that is heritable but does not involve a change in DNA sequence. Epigenetic modifications involving changes in DNA methylation are common in lung cancer and include hypomethylation, dysregulation of DNA methyltransferase I, and hypermethylation. Genes that are methylated in NSCLC include *p16*, *RARB*, *RASSFIA*, *MGMT* (methylguanine-methyltransferase), and death-associated protein kinase (DAP-kinase). Hypermethylation in lung cancer can often silence tumor suppressor genes, thereby promoting dysregulated cell growth. Silencing of tumor suppressor genes in histologically normal lymph nodes in patients with resectable NSCLC is associated with higher likelihood of disease relapse.

CLINICAL MANIFESTATIONS

Lung cancers grow from a single abnormal cell or small group of abnormal cells to develop into large macroscopic masses that may be several centimeters in diameter. Most lung cancers originate from the bronchial epithelium and are termed carcinomas. Primary noncarcinoma lung cancers are less common and include carcinoid, pulmonary blastomas (more common in younger patients), and sarcomas. Early lung cancers often are manifested as pulmonary nodules, defined as "a rounded opacity, well or poorly defined, measuring up to 3 cm in diameter" (see Fig. 191-2 for evaluation of a patient with a solitary pulmonary nodule). Abnormal lung tissues range in histologic grade from mildly atypical cells to aggressive cancers. Lesions such as atypical adenomatous hyperplasia are considered preinvasive lesions, with a continuum of cellular atypia through adenocarcinoma (Fig. 191-3).

Currently, only 15% of patients with lung cancer are asymptomatic when they are initially diagnosed. Diagnosis in these patients is often made incidentally on a chest radiograph obtained for other reasons (e.g., a preoperative study). The work-up for suspected lung cancers is dependent on the probability that the lesion in question is malignant or the stage of disease at presentation (see Fig. 191-2). Pulmonary nodules often are due to current or prior infection, although they may be the manifestation of early cancer. Results from the National Lung Screening Trial (NLST) showed that of all nodules detected, more than 95% were false positives and were noncancerous. However, most patients have symptoms and signs that are (1) caused by the pulmonary lesion itself—these include local tumor growth, invasion, and

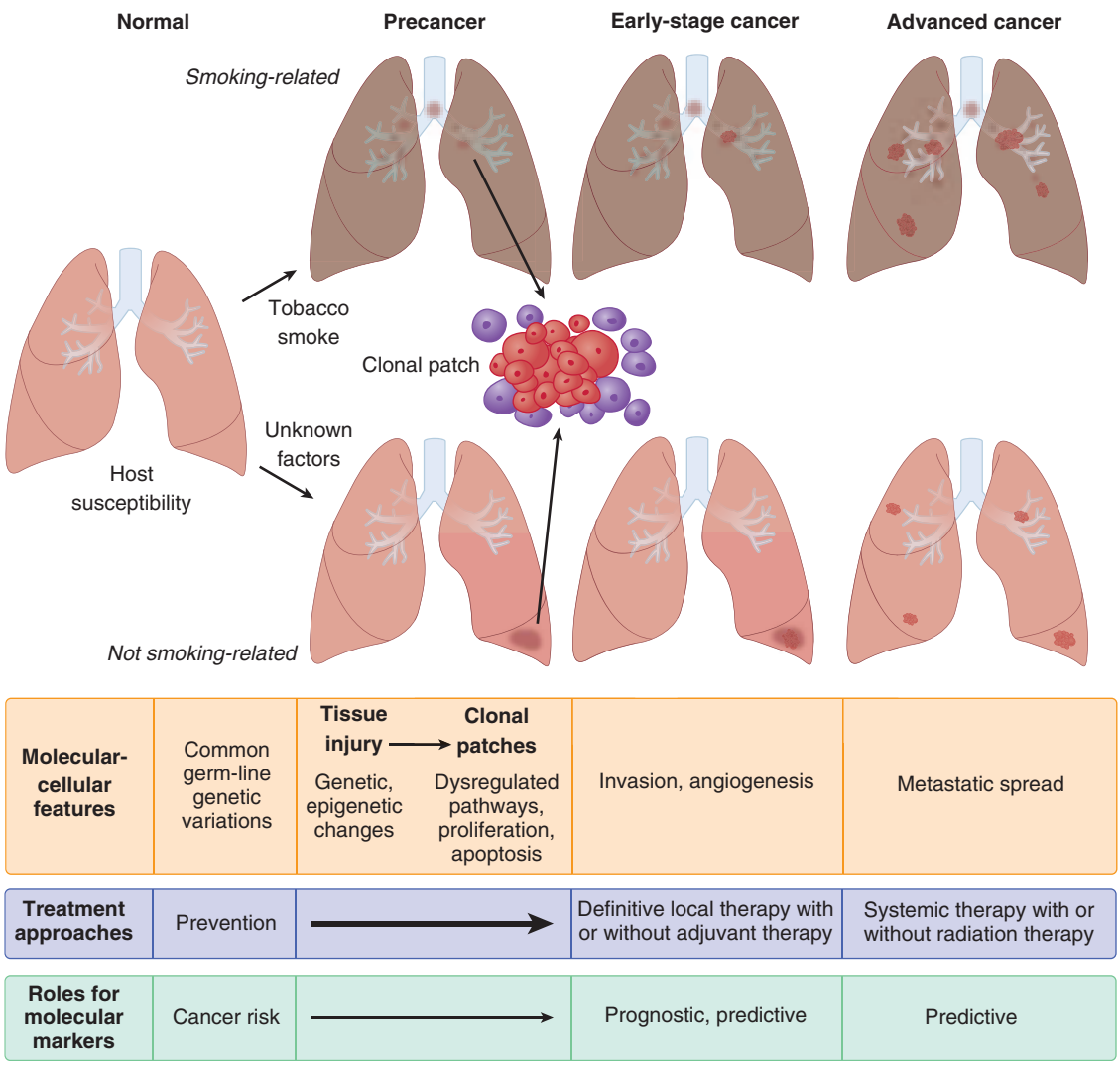

FIGURE 191-3. Carcinogenesis of lung cancer. (From Herbst RS, Heymach JV, Lippmann SM. Lung cancer. *N Engl J Med* 2008;359:1367-1380.)

obstruction; (2) intrathoracic-regional tumor spread to lymph nodes and adjacent structures; (3) extrathoracic-distant spread of disease; or (4) paraneoplastic syndromes. Common presenting symptoms of lung cancer include cough, dyspnea, pain, hemoptysis, and weight loss; anorexia occurs in about 30% of patients, fatigue in one third of patients, and anemia and fever in 10 to 20% of patients. More than 80% of patients initially have three or more symptoms or signs as a result of the lung cancer. Because the majority of patients with lung cancer have other tobacco-related cardiopulmonary diseases, such as emphysema/chronic obstructive pulmonary disease, ischemic heart disease, and others, these overlapping symptoms often result in a delay in diagnosis of the underlying malignant disease. Symptoms could also result from local invasion or metastasis of the tumor, such as headache, bone pain, airway obstruction, cough, and hemoptysis. Paraneoplastic syndromes associated with lung cancer include the syndrome of inappropriate antidiuretic hormone (Chapter 116), hypercalcemia (Chapter 245), pulmonary hypertrophic osteoarthropathy (Chapter 275 and Fig. 275-1), Eaton-Lambert myasthenic syndrome (Chapter 422), and Cushing's syndrome (Chapters 179 and 227). Hypercalcemia is common in squamous cell histology, whereas the syndrome of inappropriate antidiuretic hormone, Eaton-Lambert myasthenic syndrome, and Cushing's syndrome are most commonly associated with SCLC.

DIAGNOSIS

With the advent of computed tomography (CT) screening, it is anticipated that a greater subset of patients with lung cancer will be diagnosed before the onset of symptoms.[6] In patients with clinical or radiographic findings suggestive of lung cancer, CT scans of the chest and abdomen are indicated to determine the location of the primary tumor, involvement of mediastinal lymph nodes (Fig. 191-4), and spread to other anatomic sites.

Diagnostic Procedures

Accurate diagnostic characterization of lung cancers is essential because the presence or absence of mediastinal nodal metastases is crucial in determining prognosis, assessing resectability, and selecting the appropriate treatment strategy for primary lung cancer. Enlarged lymph nodes identified by CT or positron emission tomography (PET) require histologic confirmation. It is debatable whether all patients require invasive mediastinal staging before surgical resection or other local treatment modality, such as stereotactic body radiation therapy (SBRT). Only 5 to 15% of patients with peripheral T1 tumors with a negative mediastinum by CT or PET have mediastinal nodal metastases.

Invasive Diagnostic Procedures

Important advances in surgical diagnosis have been established and refined in the last few years. Transbronchial needle aspiration (TBNA) allows staging of the mediastinum during diagnostic bronchoscopy. Sensitivity of TBNA is dependent on lymph node size, location, and needle size, being best suited for large, clinically positive lymph nodes. On-site cytopathologic analysis increases the likelihood of obtaining a malignant diagnosis. Linear array ultrasound technology combined with TBNA allows endobronchial ultrasound (EBUS) fine-needle aspiration (FNA) of mediastinal and hilar lymph node stations. EBUS-FNA is superior in performance to TBNA, with overall sensitivity approaching 90%. Among the diseases that are also associated with enlarged and metabolically active mediastinal lymph nodes are sarcoidosis, tuberculosis, and multiple infectious causes, generally of a fungal, atypical, or viral nature (histoplasmosis, tuberculosis, and coccidioidomycosis among the differential diagnoses), making it mandatory to obtain nodal tissue.

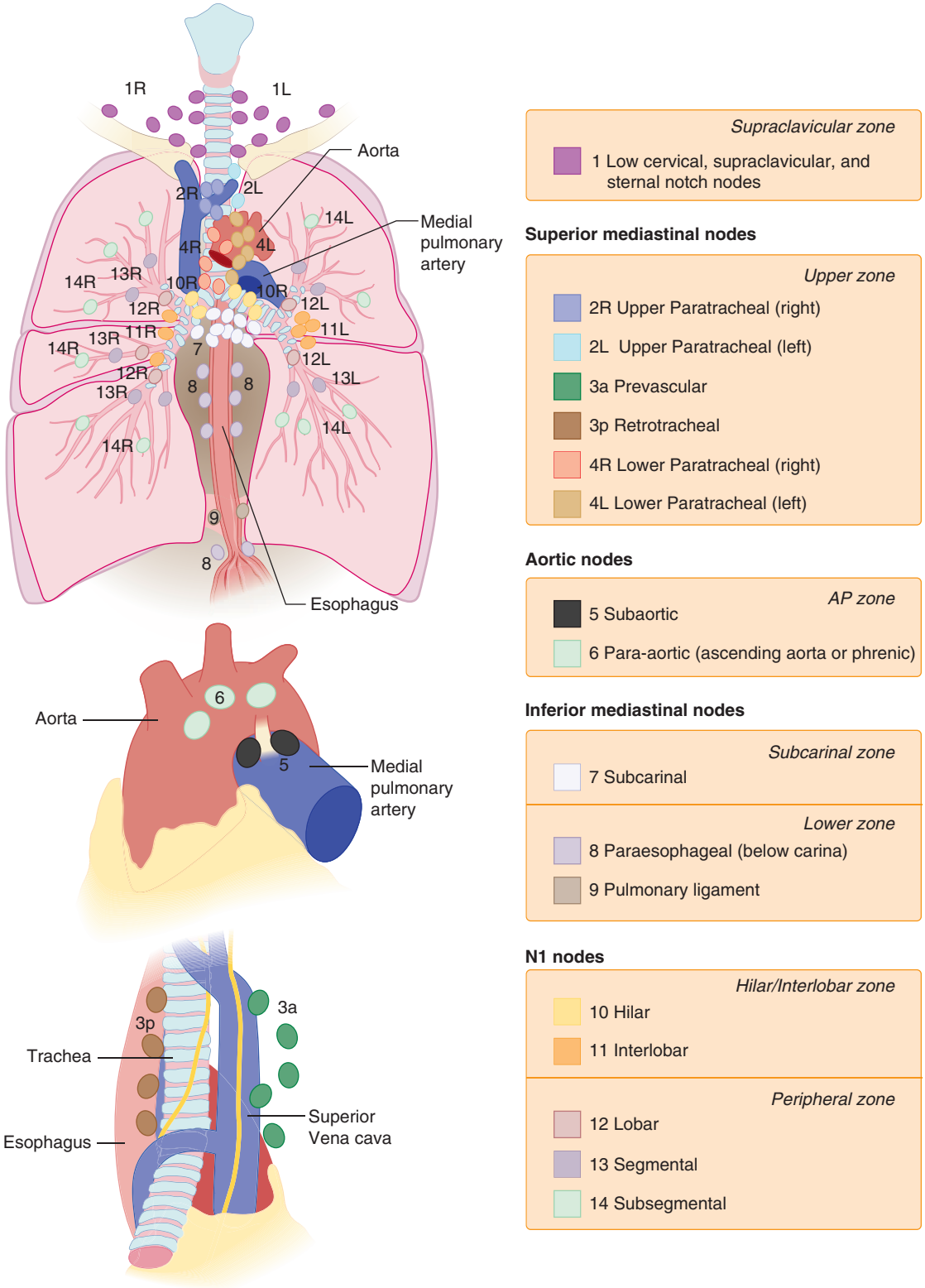

FIGURE 191-4. The International Association for the Study of Lung Cancer (IASLC) lymph node map. Included is the proposed grouping of lymph node stations into "zones" for the purposes of prognostic analyses. (From Rusch VW, Asamura H, Watanabe H. The IASCL lung cancer staging project. A proposal for a new international lymph node map in the forthcoming seventh edition of the TNM classification for lung cancer. *J Thorac Oncol.* 2009;4:568-577.)

Esophageal endoscopic ultrasonography (EUS), or EUS-FNA, allows sampling of inferior pulmonary ligament, periesophageal, and subcarinal lymph node stations (9, 8, and 7). Stations 2 and 4 are difficult to sample. The combination of EBUS-FNA and EUS-FNA can exceed 90% yield in obtaining a tissue diagnosis of lung cancer when it is present.

Mediastinoscopy involves surgical assessment of mediastinal lymph nodes for determination of tumor involvement. Cervical mediastinoscopy is the standard and allows sampling or removal of lymph node stations 2, 4, 7, and often 10. The complication rate is 2%, with few life-threatening complications. With video mediastinoscopy, sensitivity and specificity exceed 97%.

An anterior mediastinotomy (Chamberlain procedure) provides access to stations 5 and 6 (aortic and aortopulmonary window), generally not accessible with cervical mediastinoscopy. This is performed through an incision in the left second or third intercostal space or through excision of the second costal cartilage.

A biopsy is necessary to establish diagnosis and, in recent years, to conduct molecular studies that can further guide therapy. The most accessible site with the least invasive method is the preferred approach to obtaining diagnostic tissue. Whereas an FNA procedure is often adequate to establish diagnosis and can be accomplished by a transthoracic approach or by

bronchoscopy, the yield is often inadequate to conduct molecular studies. Therefore, a core needle biopsy to obtain sufficient tissue is recommended for patients with suspected lung cancer. For patients presenting with pleural or pericardial effusions, transthoracic aspiration of fluid is sufficient for diagnosis and staging. Cell blocks prepared from the fluid can be used to conduct molecular studies, although the success rate depends on the number of viable cancer cells in the specimen. The diagnostic yield of pleural fluid in patients with malignant effusion is approximately 50 to 70%. In instances in which repeated aspiration of pleural fluid is nondiagnostic, a video-assisted thoracoscopy procedure might be necessary to establish diagnosis. For patients with localized lung tumors that are suggestive of cancer, it is reasonable to proceed with surgical resection without a diagnostic biopsy if all other potential causes are excluded.

With the recent use of molecularly targeted therapies, understanding of the mechanisms of resistance is an important determinant of subsequent therapies. Therefore, it is recommended to obtain additional tumor biopsy specimens at various time points during the course of treatment.

Diagnostic Imaging

In conjunction with standard thoracic imaging procedures, additional sites of disease may require evaluation based on presenting complaints. The most common sites of lung cancer metastasis include contralateral lung, liver, adrenal gland, bones, and brain. Imaging of the brain is recommended to evaluate for metastasis in patients with suggestive symptoms and signs or those with lung adenocarcinoma larger than 3 cm and evidence of mediastinal nodal involvement. Magnetic resonance imaging (MRI) and CT scan with contrast enhancement are both acceptable modalities to evaluate for brain metastasis, although MRI is preferred for its superior sensitivity. Radionuclide study of the bones is indicated in patients with bone pain or an unexplained elevation in serum alkaline phosphatase level.

PET using [^{18}F]fluorodeoxyglucose (FDG) is included as part of staging in patients with localized lung cancer; however, the use of FDG-PET scan to assess response to anticancer therapy and in surveillance after curative therapy is controversial at this time and is not recommended. MRI scan of the chest may be useful in determination of invasion of surrounding structures, such as the brachial plexus, in patients with tumors involving the superior sulcus of the lung, but its use in staging is generally restricted to preoperative settings.

Solitary Pulmonary Nodule

The management of pulmonary nodules differs by whether they are solid (soft tissue attenuation) or subsolid (less than soft tissue attenuation without obscuring of the underlying lung architecture on CT). For larger nodules, CT characteristics suggestive of malignancy include irregular margins, spiculation, invasion of adjacent structures, lymphadenopathy, and distant metastases. Asymptomatic single pulmonary nodules less than 3 cm in diameter, the "solitary nodule," with normal surrounding lung architecture are found incidentally in up to 0.2% of chest radiographs; 10 to 70% are malignant. The potential for these lesions to be malignant increases with the patient's age, nodule size (< 4 mm vs. > 8 mm), growth rate, definite history of smoking, and size changes compared with prior imaging studies. PET scans may be helpful in defining abnormal mediastinal nodes in these patients. A pulmonary nodule that has not changed in size for more than 2 years is probably benign. For lesions larger than 8 mm, serial high-resolution CT scans are appropriate. Suspicious nodules should undergo definitive biopsy (see Fig. 191-2).

Staging

Stage is the most important determinant of prognosis in patients with lung cancer. The seventh edition of the *American Joint Committee on Cancer (AJCC) Cancer Staging Manual* is currently in use (Table 191-2). This staging system uses new T and M descriptors to determine lung cancer stage based on the TNM (tumor, node, and metastasis) profile of the patient. Individual T descriptors are defined on the basis of tumor size of less than 2, 2 to 3, 3 to 5, 5 to 7, and more than 7 cm. In the previous system, tumors were categorized on the basis of size of less than or more than 3 cm. Currently, satellite nodules in the same lobe as the primary tumor are categorized as T3 and nodules in another lobe of the ipsilateral lung as T4. Malignant pleural or pericardial effusion constitutes M1 disease. Presence of metastases within the thorax constitutes M1a and extrathoracic disease is classified as M1b because patients with M1a disease have a slightly better prognosis than those with M1b. Nodal descriptors were not changed from the previous system.

The TNM staging system is also now recommended for SCLC, given its ability to ascertain prognosis in a more accurate manner. SCLC was previously classified as limited or extensive stage on the basis of the ability to administer radiation therapy to the tumor with a single port. Treatment guidelines for localized SCLC have not been affected.

TABLE 191-2 TREATMENT BY AMERICAN JOINT COMMITTEE ON CANCER (AJCC) STAGE

	T	N	M	MEDICALLY FIT	MEDICALLY UNFIT
Stage IA	1	0	0	Surgery	SBRT
Stage IB	2	0	0	Surgery plus adjuvant chemotherapy if T ≥4 cm	SBRT plus chemotherapy at progression
Stage IIA/B	1-2	1	0	Surgery plus adjuvant concurrent chemoradiation	Thoracic radiation plus chemotherapy
Stage IIIA	X-3	2	0		Concurrent or sequential chemoradiation
	4	X	0		
Stage IIIB	X-4	3	0	Concurrent chemoradiation	
Stage IVA	X	X	1a	Chemotherapy (if *EGFR* mutation, treat with EGFR TKI)	
Stage IVB	X	X	1b	Chemotherapy (if *EGFR* mutation, treat with EGFR TKI); radiation for palliation	

T = tumor; N = node; M = metastasis; SBRT = stereotactic body radiation therapy; EGFR = epidermal growth factor receptor; TKI = tyrosine kinase inhibitor. (From AJCC Cancer Staging Manual. 7th ed. New York: Springer-Verlag; 2018)

TREATMENT

Non–Small Cell Lung Cancer

Surgery

Surgical management plays the major role in the treatment of patients with stage I, stage II, and selected stage III NSCLC. However, nearly 40% of patients with early-stage lung cancer are not candidates for surgery because of limiting comorbid conditions. The commonly used parameters for inoperability include pulmonary functions with baseline forced expiratory volume in the first second of expiration (FEV_1) of less than 40%, predicted postoperative FEV_1 of less than 30%, and severely limited diffusion capacity. Such patients are referred to as medically inoperable despite the presence of localized disease and may be candidates for SBRT, as discussed later.

The first step in managing localized lung cancer is to stage the mediastinal lymph nodes. For peripheral tumors that are not associated with mediastinal adenopathy and do not have FDG uptake in the nodes, many surgeons advocate proceeding with surgical resection and sampling mediastinal nodes intraoperatively. However, for patients with nodes that are positive on PET scan, sampling is strongly recommended before surgery. The false-positive rate for PET scan in the mediastinum for patients with localized lung cancer is approximately 20%. The likelihood of nodal involvement in patients with negative PET scan is approximately 5 to 15%.

Lobectomy is the standard surgical procedure for patients with localized lung cancer who are medically fit.[7] If anatomic resection cannot be achieved with lobectomy, bilobectomy or pneumonectomy might be necessary. Sleeve resection refers to removal of the tumor along with the bronchus and anastomosis of the remaining ends of the bronchial tree. Surgical resection can be achieved by performing an open thoracotomy or by video-assisted thoracic surgery (VATS). VATS is gaining wider use because of lower morbidity, faster recovery from surgery, and better ability to administer postoperative systemic therapy. The ability to achieve an R0 resection is critical, and surgery should not be attempted if this is not deemed feasible during preoperative work-up. For patients with positive surgical margins, re-resection should be attempted whenever it is feasible. If not, postoperative radiation therapy should be administered. Robotic resection of lung cancers, including robotic lobectomy and pneumonectomy, has increased in use but is not as widely employed as VATS approaches.

Sublobar resections are not recommended because of the higher risk of local recurrence. An exception to this rule is for patients with peripheral tumors smaller than 2 cm, for which studies have demonstrated excellent outcomes. An ongoing study is comparing sublobar resection to standard lobectomy and will likely provide definitive answers to this important question.

A randomized comparison of mediastinal lymph node dissection to nodal sampling demonstrated comparable outcomes for patients with NSCLC. Another study compared sublobar resection followed by placement of ^{125}I brachytherapy to the tumor bed with surgery alone in patients who are not candidates for standard lobectomy. There was no difference in overall survival between the two groups, and therefore the brachytherapy approach is not recommended. Tumors involving the superior sulcus are managed with preoperative chemoradiotherapy to enhance tumor resection and to gain local and distant tumor control. The decision to perform surgery for these anatomically challenging tumors depends on the extent of local invasion, involvement of the brachial plexus, and mediastinal lymph node involvement.

The role of surgery in the management of stage III NSCLC with mediastinal nodal involvement continues to be controversial. Surgery alone is associated with a poor outcome. In a randomized study, patients with N2-positive disease who underwent chemoradiotherapy followed by surgery did not have improved survival compared with chemoradiotherapy alone and had an unacceptably high rate (almost 30%) of postoperative mortality. Therefore, trimodality therapy is not recommended for patients who require pneumonectomy. For patients with multistation N2 disease or bulky nodal disease, surgical resection is not recommended. Clearance of mediastinal nodes after induction therapy might be the most important predictor of benefit from surgical resection, which calls for restaging of the mediastinum after induction therapy if surgery is contemplated.

The role of surgery in patients with oligometastatic disease can be considered under certain situations. Surgical resection of both the primary and a solitary brain metastasis has resulted in 5-year survival rates of approximately 20%. There are also limited data with oligometastatic disease to the adrenal glands, but similar approaches with solitary metastasis at other distant sites are not recommended. This approach cannot be recommended for patients with mediastinal nodal involvement.

Radiation Therapy

Radiation therapy is an important part of multimodality therapy for NSCLC. It plays a major role in curative therapy for stage III disease and palliation of stage IV disease and has been successfully tested for patients with medically unresectable stage I disease. Significant improvements in the delivery of radiation therapy in the past two decades allow use of smaller radiation field size, reducing exposure of normal tissue to radiation and facilitating more effective treatment of tumor. Respiratory gating techniques allow the delivery of radiation therapy to the tumor regardless of the phase of respiration. SBRT involves the delivery of high-dose radiation to a limited tumor volume after stereotactic localization.

Stage I and Stage II NSCLC

SBRT[8] has emerged as an effective treatment option for patients with T1 and T2 tumors who are node negative but are medically inoperable because of comorbid illness. In one study, delivery of SBRT over three to five fractions resulted in nearly 90% local control rate, prompting studies of SBRT in medically fit patients and in combination with systemic therapy for early-stage NSCLC. SBRT has shown considerable efficacy for peripheral tumors; studies are ongoing to examine its use in centrally located tumors.

Radiation therapy is indicated for patients with positive surgical margins after surgery for early-stage NSCLC but not for those with negative surgical margins. A meta-analysis reported a detrimental effect for patients treated with postoperative radiation therapy, especially for those with N0 and N1 disease. Patients with involved mediastinal nodes (or N2 disease) demonstrated favorable survival with radiation therapy. This has also been observed in an analysis of the U.S. Surveillance, Epidemiology and End Results (SEER) database. A prospective study is under way in Europe to compare postoperative radiation therapy with observation in patients with surgically resected N2 disease.

Stage III NSCLC

Whereas surgery is appropriate for patients with T3N1 disease, administration of radiation therapy results in improved outcomes for patients with involvement of the mediastinal lymph nodes. A subset of N2-positive patients might benefit from multimodality therapy involving neoadjuvant chemoradiation followed by surgical resection, including stage IIIA patients with single-station or microscopic lymph node involvement and disease amenable to resection with lobectomy or bilobectomy. Preoperative radiation consists of 45 Gy once daily, and a dose of 60 Gy has been piloted with acceptable safety results.

For patients with stage III disease that is not appropriate for surgical resection, thoracic radiation therapy to a dose of 60 to 66 Gy in once-daily fractions with concurrent (rather than sequential) chemotherapy is the recommended treatment. This category includes patients with bulky mediastinal disease, involvement of contralateral or supraclavicular nodes (N3), and direct invasion of major structures such as the vertebrae, trachea, major blood vessel, or esophagus by the primary tumor (T4). A 5-year survival rate of 20 to 25% has been reported with combined chemoradiotherapy in this setting. The main adverse events include esophagitis and pneumonitis, the latter depending on the extent of normal lung tissue and the dose of radiation received by normal lung tissue. Radiation-related pneumonitis can occur immediately after radiation therapy or after 6 to 9 months.

Several efforts to improve on standard chemoradiotherapy have been undertaken in the past two decades. Hyperfractionated radiation therapy with administration of two or three fractions per day has demonstrated favorable results over once-daily fractionation, particularly in squamous cell carcinoma. However, logistical constraints have limited the adoption of this approach. Single-arm studies showed the use of higher doses of up to 74 Gy in once-daily fractions to be a promising approach. However, randomized studies have failed to show benefit to dose escalation to 70 Gy or higher in this population of patients. Therefore, 60 to 66 Gy remains the standard radiation dose for stage III NSCLC.

Stage IV NSCLC

In patients with advanced-stage NSCLC, radiation therapy can be effective for palliation of spinal cord compression, brain metastasis, airway obstruction, hemoptysis, and pain.

Spinal cord compression is an emergency situation, and neurosurgical evaluation should be initiated to assess whether surgery is advised, which can be the case when there is a large tumor burden and rapidly deteriorating motor and sensory function. Surgical decompression is used when neurologic compromise is early and the patient has well-controlled systemic disease and it is followed by radiation therapy. Spinal cord compression is usually managed with external beam radiation therapy to 30 Gy and corticosteroids. For brain metastases, resection of oligometastatic disease has been associated with better outcomes. Whole brain radiation therapy of 30 to 37.5 Gy given in 10 to 15 fractions can be administered when multiple metastases are present. Stereotactic radiosurgery can be used instead of whole brain radiation therapy for patients with low-volume brain metastasis limited to one to three lesions and brain lesions that progress after whole brain radiation therapy. Pain control in sites of bone metastasis or chest wall involvement can be achieved by a short course of radiation therapy. Palliative radiation therapy is also administered to 30 to 45 Gy during 2 to 3 weeks before the initiation of systemic therapy in patients with systemic disease who present with hemoptysis or postobstructive pneumonia.

Systemic Therapy

Systemic therapy refers to the use of cytotoxic or molecularly targeted agent therapy or both. Although it was initially developed for patients with advanced-stage lung cancer, the high propensity for metastasis of lung cancer cells has extended the use of systemic therapy to patients with earlier stages of the disease. A number of effective and well-tolerated cytotoxic agents have been developed during the past three decades that are used for the routine care of patients with lung cancer. In addition, many targeted agents are under clinical investigation; six are currently approved for use in NSCLC (Table 191-3): the EGFR TKIs erlotinib, gefitinib, and afatinib; the vascular endothelial growth factor (VEGF) monoclonal antibody bevacizumab; and crizotinib, an inhibitor of MET, ALK, and ROS1 tyrosine kinases.

Systemic Therapy in Early-Stage NSCLC

Even in the setting of optimal surgery, patients with early-stage NSCLC remain at high risk for disease recurrence or metastases due to the presence of micrometastasis, which can increasingly be quantified by the evaluation of circulating tumor cells or circulating tumor DNA. A consistent benefit of 5 to 15% in 5-year survival rate observed across multiple trials has resulted in the adoption of four cycles of adjuvant cisplatin-based two-drug combination regimens as the standard of care for stage II and stage IIIA NSCLC. In stage IA disease, however, the potential benefits of chemotherapy are largely outweighed by the risks, and there is an overall detrimental effect. For patients with stage IB disease, post hoc analyses have revealed that improvement in survival with adjuvant therapy was restricted to patients with tumors larger than 4 cm, but this observation has yet to be validated in prospective trials.

Whereas cisplatin-vinorelbine has been the most commonly used combination in clinical trials of adjuvant therapy, the use of newer, better tolerated agents effective in the treatment of advanced NSCLC, such as taxanes, gemcitabine, and pemetrexed, is increasing. It is hoped that the future use of adjuvant chemotherapy may be tailored to patients at high risk for recurrence, based on genomic or proteomic markers, with circulating tumor cells or circulating tumor DNA among the most promising tools at present.[9]

Locally Advanced NSCLC

For patients with stage III disease that is not amenable to surgical resection, concomitant administration of chemotherapy with radiation therapy has

TABLE 191-3 TARGETED AGENTS IN LUNG CANCER THERAPY

MOLECULAR TARGET	FDA-APPROVED AGENTS	INVESTIGATIONAL AGENT
EGFR	Erlotinib Gefitinib Afatinib	Dacomitinib Cetuximab
ALK translocation	Crizotinib Certinib	
ROS	Crizotinib	
VEGF	Bevacizumab	
B-Raf		Dabrafenib
HSP90		Ganetespib

FDA = Food and Drug Administration.

consistently demonstrated efficacy superior to that of sequential therapy. Both cisplatin- and carboplatin-based regimens are associated with meaningful survival results but have not been compared in this setting. The combination of cisplatin and etoposide allows administration of full systemic dose of chemotherapy with radiation therapy. The carboplatin and paclitaxel regimen involves administration of lower "radiosensitizing" doses of the two agents with radiation therapy followed by consolidation therapy with two cycles at regular doses. This approach has a favorable tolerability profile compared with cisplatin-based regimens. The use of induction or consolidation chemotherapy in other settings has not resulted in improved survival. With modern combined chemoradiotherapy, cure rates of nearly 20 to 25% are achieved in locally advanced NSCLC, with esophagitis and pneumonitis being the main toxicities.

Several systemic targeted agents, including cetuximab (a monoclonal antibody against EGFR), bevacizumab (a monoclonal antibody against VEGF), and ALK inhibitors, are under evaluation for the treatment of stage III disease. Ongoing studies continue to evaluate the role of EGFR TKIs in patients with activating *EGFR* mutations and locally advanced disease.

Advanced-Stage NSCLC

In patients with advanced-stage NSCLC, randomized trials have shown platinum-based systemic therapy to improve overall survival and quality of life compared with supportive care alone. Cisplatin treatment is associated with nausea, emesis, nephrotoxicity, and neurotoxicity, although the availability of highly effective antiemetic agents has greatly improved its tolerability. Carboplatin is a better tolerated alternative to cisplatin and is associated with ease of outpatient administration. The dose-limiting toxicity of carboplatin is thrombocytopenia. Two-drug combination regimens have proved superior to monotherapy with cisplatin alone or a nonplatinum compound, leading to the adoption of combination chemotherapy as the recommended treatment of advanced NSCLC.

A meta-analysis of randomized trials compared the efficacy of cisplatin to carboplatin in advanced-stage NSCLC and demonstrated comparable survival but a numerically higher incidence of treatment-related deaths with cisplatin-based regimens. Although cisplatin-based regimens have a narrow efficacy advantage, carboplatin-based regimens have found wider adoption because of their favorable therapeutic index.

A number of partner agents for platinum have demonstrated similar efficacy in advanced NSCLC in randomized trials, including etoposide, vinblastine, vindesine, vinorelbine, taxanes, gemcitabine, irinotecan, and pemetrexed. On the basis of several randomized studies, the choice of chemotherapy agent for frontline treatment is made on consideration of toxicity, preference of the patient, schedule, and cost. Recent trials involving superior supportive care have shown median survival for combinations of cisplatin with docetaxel, gemcitabine, or pemetrexed in the 10- to 11-month range. Combinations of three cytotoxic agents are not recommended because of a higher toxicity burden and lack of meaningful incremental benefit.

In randomized studies of patients with an activating *EGFR* mutation, treatment with gefitinib or erlotinib has been associated with improvements in progression-free survival and quality of life over platinum-based chemotherapy. This did not translate into a survival benefit because the majority of patients treated with chemotherapy were subsequently crossed over to receive an EGFR inhibitor on disease progression. The importance of molecular testing before initiation of EGFR inhibitor therapy in first-line treatment is highlighted by the inferior outcomes in wild-type patients treated with targeted therapy.[■] Afatinib, an irreversible EGFR TKI, has recently demonstrated superiority over chemotherapy in patients with an activating *EGFR* mutation, leading to its approval in this setting.

Initial studies of the VEGF antibody bevacizumab for first-line therapy for advanced NSCLC were promising, although squamous histology was associated with a higher incidence of life-threatening hemoptysis. Further studies limited to patients with nonsquamous histology demonstrated a significant improvement in overall survival and progression-free survival with the addition of bevacizumab to carboplatin and paclitaxel chemotherapy, leading to the Food and Drug Administration approval of bevacizumab in this setting. Notable adverse events included bleeding, hypertension, proteinuria, and neutropenia.

The MET, ALK, and ROS1 TKI crizotinib demonstrated a response rate of nearly 60% and a clinical benefit rate of 90% in patients with ALK-positive advanced-stage NSCLC, with a median progression-free survival of 10 months.[■] Accordingly, crizotinib is approved in this setting in the United States and Europe. The second-generation ALK inhibitor ceritinib also is useful and approved for the treatment of patients with ALK-positive metastatic NSCLC with disease progression or intolerance to crizotinib.[■]

Role of Histology in Choice of Chemotherapy

Until recently, chemotherapy regimens were considered to be suitable for all histologic subtypes of NSCLC. This notion was dispelled in a randomized phase III study demonstrating superior survival for patients with nonsquamous advanced-stage NSCLC treated with cisplatin-pemetrexed and superior survival for patients with squamous histology receiving cisplatin-gemcitabine. Nanoparticle albumin-bound paclitaxel (nab-paclitaxel) is associated with a favorable response rate in patients with advanced NSCLC, which is restricted to patients with squamous histology. The biologic basis for the histologically specific therapeutic activity of these drugs is not presently known; however, the variable efficacy of pemetrexed and nab-paclitaxel based on histology should be considered when chemotherapy is selected for first-line treatment of advanced NSCLC.

Maintenance Therapy

In the first-line treatment of advanced-stage NSCLC, continuation of combination, multi–cytotoxic drug treatment beyond four to six cycles is associated with cumulative toxicities but no tangible benefit. Recently, single-agent maintenance therapy has proved to be useful in patients who have a demonstrated clinical benefit from four cycles of platinum-based combination chemotherapy. Current therapeutic strategies employ either a "switch maintenance" approach, which uses an alternative cytotoxic or targeted agent that has not been previously administered, or "continuation maintenance," which involves continuing the nonplatinum agent beyond four cycles.

Pemetrexed is the only cytotoxic agent that provides a survival advantage as maintenance therapy in advanced NSCLC. It has demonstrated similar benefit in both continuation and switch maintenance strategies in randomized trials for patients with advanced nonsquamous histology and has thus been approved for maintenance therapy in the United States and Europe. The EGFR TKI erlotinib also extends survival when it is used as maintenance therapy in patients who have experienced stable disease with a platinum-based combination for four cycles, with greater benefit in patients with an activating *EGFR* mutation. The use of bevacizumab maintenance has been adopted for patients who receive it as part of their initial treatment regimen because all pivotal randomized trials performed with bevacizumab used it as maintenance therapy after six cycles of combination treatment.

Docetaxel, pemetrexed, and erlotinib are also efficacious when they are used as salvage therapy for patients with advanced NSCLC who experience disease progression during or after platinum-based chemotherapy. Therefore, the relative merits of using the last two agents as maintenance therapy versus use after disease progression has become controversial. Careful discussion with patients with nonsquamous lung cancers who have responded to frontline chemotherapy regarding the value of maintenance therapy versus close observation is recommended.

Salvage Therapy

Virtually all patients with advanced NSCLC will eventually experience disease progression regardless of the extent of benefit from first-line chemotherapy. Salvage therapy for these patients provides a modest but real improvement in survival. Docetaxel, at a dose of 75 mg/m^2 every 3 weeks, was the first agent proven in this setting to improve survival compared with best supportive care or first-generation cytotoxic agents. Pemetrexed has efficacy as salvage therapy similar to that of docetaxel and a more favorable toxicity profile, but its use is restricted to patients with nonsquamous histology. Targeted therapy with erlotinib has also been demonstrated to enhance survival and progression-free survival in the salvage setting. Overall, the presently available salvage therapy options are associated with response rates of less than 10%, with a median progression-free survival of 3 months and overall survival of 8 months.

Crizotinib significantly improves progression-free survival and response rate compared with chemotherapy in the salvage setting for patients whose tumors exhibit an *ALK* rearrangement. The use of bevacizumab as salvage therapy for NSCLC is currently being examined in ongoing clinical trials. Combinations of cytotoxic or targeted agents for salvage therapy have not resulted in improved survival, and these are therefore not recommended.

Investigational Targeted Agents

The use of targeted therapies has gained considerable momentum in recent years, and many investigational targeted agents are in preclinical and clinical development. The irreversible EGFR inhibitor dacomitinib demonstrated a favorable efficacy profile compared with erlotinib in a randomized phase II study in an unselected population of patients and is presently being evaluated in phase III trials. Several new agents and combinations are under study for managing resistance to EGFR TKIs. Cetuximab, a monoclonal antibody against EGFR, produces a modest improvement in overall survival when it is combined with cisplatin and vinorelbine for first-line treatment of advanced NSCLC. The identification of predictive biomarkers to select patients is necessary for the use of this combination in routine care.

LDK378, a potent ALK inhibitor, produced a response rate of 60% in patients who developed disease progression during crizotinib therapy. Other novel ALK inhibitors are also under development for management of crizotinib resistance or as primary therapy. Heat shock protein 90 (HSP90) plays a critical chaperone function for ALK, and HSP90 inhibitors have recently been found to possess single-agent activity in this subgroup of patients.

Besides the VEGF monoclonal antibody bevacizumab, other strategies to inhibit angiogenesis, including small-molecule VEGF TKIs and vascular disrupting agents, have not been successful to date in advanced NSCLC. Efforts

to identify biomarkers to predict benefit with bevacizumab and other antiangiogenic agents have also been unsuccessful and have consequently restricted optimal use of these agents.

Advances in genomic technology have made it possible to prospectively identify novel mutations that play a critical role in the growth of lung cancers. In adenocarcinoma, a fusion gene involving *ROS1*, observed in 1% of patients, also confers sensitivity to treatment with crizotinib. Another fusion involving the *RET* gene has been identified in 0.5 to 1% of patients. Patients with mutations in *BRAF* appear to respond to therapy with dabrafenib, a BRAF inhibitor. Recent DNA sequencing data from the Cancer Genome Atlas project in a cohort of patients with squamous cell lung carcinoma, together with the efforts of the Lung Cancer Mutation Consortium, have demonstrated several novel, potentially targetable mutations in this disease. Routine testing of tumor specimens of patients for molecular targets is increasingly seen as a strategy to personalize treatment options for lung cancer.

Management of Special Populations of Patients

Elderly patients represent a growing subset of lung cancer patients. In the United States, the median age at diagnosis of lung cancer is 70 years. Aging is associated with declines in physiologic and vital organ function that affect a patient's tolerance for systemic therapy, making current physical function and prospective quality of life considerations important factors in the choice of systemic treatment. Single-agent chemotherapy has a demonstrated survival benefit compared with supportive care in elderly patients; and for those with good performance status, platinum-based combinations are superior to single-agent therapy. The appropriate use of targeted agents may have clinical benefit in the elderly. The addition of bevacizumab to chemotherapy for nonsquamous cancers in older patients is, however, associated with a narrow therapeutic index and is recommended only with great caution.

A substantial percentage of NSCLC patients present with reduced performance status, which limits their ability to tolerate combination chemotherapy. The median survival for advanced NSCLC patients with a performance status of 2 or worse (Eastern Cooperative Oncology Group scale) is dismal at less than 4 months; however, studies conducted exclusively in patients with a poor performance status indicate that chemotherapy may be beneficial for highly selected patients. It is important to consider the underlying cause of poor performance status in making treatment plans for this population of patients. For those with limiting comorbid conditions, a less aggressive approach with single-agent chemotherapy might be more appropriate. For those with a targetable mutation, appropriate targeted therapy can be administered regardless of the performance status in view of the greater potential for benefit.

Small Cell Lung Cancer

SCLC is characterized by enhanced initial sensitivity to systemic chemotherapy, although disease recurrence is common regardless of the extent of initial response. Approximately two thirds of patients present with extensive-stage SCLC, defined as the presence of metastatic disease outside the chest or large-volume thoracic disease that cannot be treated with radiation therapy. The median survival of untreated extensive-stage SCLC is less than 2 months, and the overall goal of treatment is palliation. The use of platinum-based chemotherapy results in response rates of 50 to 70% and median survival of 9 to 11 months. The regimen of four to six cycles of cisplatin and etoposide is considered the standard approach for the treatment of SCLC, extending to six cycles in responding patients, with no proven role for maintenance therapy. Carboplatin is considered an acceptable alternative in the treatment of extensive-stage disease. Despite the extent of initial response, disease recurrence develops in a median of 4 to 6 months. Disease that progresses during or within 90 days of administration of cisplatin-based chemotherapy is referred to as refractory relapse. Recurrence outside this time window represents a "sensitive" subgroup that might benefit from salvage treatment. Other approaches, such as high-dose chemotherapy, alternating chemotherapy regimens, dose-dense therapy, and three-drug combination regimens, have been unsuccessful in improving survival. In the Japanese population but not in Western patients, the regimen of cisplatin and irinotecan has demonstrated superior results over cisplatin and etoposide.

Salvage therapy has yielded only modest results in sensitive relapsed SCLC. Topotecan is the only agent to demonstrate clinical benefit in relapsed SCLC, with a response rate of 20% and a favorable symptom improvement profile but no improvement in overall survival in randomized studies. Several novel agents, including molecularly targeted agents against known targets, are presently being studied in efforts to improve the outcomes for patients with SCLC.

For patients with limited-stage SCLC, radiation therapy is used in combination with chemotherapy and can achieve cure for approximately 20 to 30% of patients. Earlier initiation of radiation therapy appears superior to a delayed approach and has been adopted as the standard in fit patients. A randomized study demonstrated superior survival of twice-daily radiation therapy fractions compared with the same dose given once daily along with concomitant chemotherapy, and this is being further evaluated in an ongoing trial.

Prophylactic cranial irradiation is associated with a modest improvement in 5-year survival for patients with limited-stage SCLC who achieve complete remission after combined modality therapy. This is due to the high risk of brain recurrence observed in patients with SCLC. Similarly, studies have demonstrated that prophylactic cranial irradiation results in modest improvement in overall survival and reduced risk of brain recurrence in patients with extensive-stage disease who achieve a favorable response to combination chemotherapy.

The role of surgery is limited to the less than 5% of SCLC patients presenting with peripheral lung lesions without mediastinal nodal involvement. In 10 to 15% of patients with SCLC, a mixed histology with NSCLC features is observed. These patients may demonstrate local progression after combined modality therapy resulting from the NSCLC component. Therefore, these individuals could be considered for surgical resection in selected situations.

Treatment advances in SCLC have, unfortunately, lagged behind those for NSCLC during the past two decades; consequently, survival outcomes for SCLC have not changed considerably. Concerted efforts to develop appropriate preclinical models to test new agents, genomic subcategorization of SCLC, and discovery of new systemic anticancer agents are necessary to improve outcomes for this aggressive disease.

SURVIVORSHIP AND SURVEILLANCE

As outcomes for lung cancer have improved in recent years, with a concomitant increase in the number of survivors after surgery or chemoradiotherapy, an important part of lung cancer care has become defining the optimal surveillance, follow-up for second primary disease, and management of long-term consequences of chemoradiotherapy. The importance of smoking cessation cannot be overemphasized, given the high risk of enhanced complications of chemotherapy and the higher incidence of second primary tumors in lung cancer survivors. Patients should be provided with appropriate opportunities to receive counseling and behavioral therapy and other treatment modalities to assist their efforts to discontinue smoking (Chapter 32).

At present, there is no standard approach to optimal radiographic and clinical follow-up in patients who undergo surgical resection or chemoradiotherapy. CT scans are commonly used for follow-up of these patients. However, the relative merits of CT scan versus chest radiography, the frequency of evaluation, and the role of FDG-PET scans are unclear and will be answered only by prospective clinical trials. For patients with advanced-stage disease, CT scans are used to assess response to therapy every two or three cycles of treatment. In view of the proven role for salvage therapy, patients who are in follow-up after combination chemotherapy should be closely observed for development of new symptoms or clinical deterioration in addition to periodic radiographic studies.

Respiratory therapy should be offered and strongly recommended to patients with dyspnea after surgery or chemoradiotherapy. Because a high proportion of these patients also have smoking-related pulmonary diseases, referral to a pulmonologist should be considered in symptomatic patients. Overall, a team approach that includes supportive care personnel, oncologic psychiatrists, nutritionists, oncologists, and appropriate additional specialists should be used to ensure the return of lung cancer survivors to normalcy to the fullest extent possible.

SUPPORTIVE CARE

Patients with lung cancer have a cure rate of only 16%, even in the most advanced Western health systems, because the majority of patients present with locally advanced or metastatic disease. Early discussion and institution of palliative and supportive care are therefore critical for the patient to ensure maximal quality of life. Studies indicate that early integration of supportive care helps maintain and improve quality of life, with one randomized trial finding that the addition of early palliative care to chemotherapy for patients with advanced disease can produce a meaningful impact on overall survival,[■] a surprising but reaffirming finding.

EARLY DETECTION AND PREVENTION

Decades of research into the screening of high-risk individuals for earlier detection of lung cancer have recently met with success.[10] The NLST randomized more than 50,000 subjects between 55 and 74 years of age with a history of cigarette smoking of at least 30 pack-years to screening with low-dose CT scan or chest radiographs at baseline and 1 and 2 years after enrollment. The rate of adherence to scans was more than 90% in both arms. Positive results were observed in nearly 25% and 7% of subjects with CT screening and chest radiograph, respectively. Among patients with a positive CT scan, 96.4% were deemed false positive after additional evaluation. Adverse events were uncommon. Screening with annual low-dose CT scans in high-risk

individuals was associated with a reduction of 20% in lung cancer mortality and 6.7% in all-cause mortality. Nearly 80% of patients diagnosed with lung cancer with low-dose CT had stage I, II, or IIIA disease amenable to curative therapy. Subsequent analysis of the NLST data showed that screening with low-dose CT prevented the greatest number of deaths from lung cancer among participants who were at highest risk and prevented very few deaths among those at lowest risk. These results have led to the adoption of low-dose CT for early detection of lung cancer by major health organizations including the U.S. Preventive Services Task Force.[11,12] For lung nodules detected on low-dose CT screening, it has been shown that predictive tools based on patient and nodule characteristics can be used to accurately estimate the probability that they are malignant.[13]

Primary prevention of lung cancer focuses on ways to prevent individuals from smoking and promotion of smoking cessation, which remain the most effective means to prevent lung cancer. Smoking cessation efforts are often sporadic; but most studies indicate that patients who successfully quit smoking benefit from pharmacologic, psychological, and physician support.

Trials of supplemental β-carotene and vitamin E, stimulated by epidemiologic evidence of lower serum levels of these antioxidants in patients with lung cancer, not only have been unsuccessful but actually produced a higher risk of lung cancer in smokers. High concentrations of selenium in the blood are associated with a lower risk of lung cancer, but a phase III trial in patients with resected stage I lung cancer failed to show any benefit for selenium supplementation. These and other trials have indicated that lung cancer patients who continue to smoke have the highest incidence of recurrence and second primary tumor development.

OTHER PULMONARY NEOPLASMS

Malignant Mesothelioma

Malignant pleural mesotheliomas[14] are generally related to asbestos exposure, with a peak risk for development of the disease 30 to 35 years after the initial exposure. Additional possible risk factors include radiation and SV40 virus. Mesothelioma is generally diagnosed in the fifth to seventh decade of life (median age, 60 years), with a male-to-female preponderance of 5 : 1. Common symptoms include dyspnea (60%) and chest wall pain or discomfort (60%). Chest radiographs usually reveal the presence of a unilateral pleural effusion. Tumor progression and symptoms are generally the result of local progression, with symptomatic distant metastases a late occurrence. Cytologic evaluation of pleural fluid to establish diagnosis is difficult and often inaccurate. The diagnosis is generally made by a biopsy procedure under CT guidance or thoracoscopically, including VATS if necessary. Several staging systems have been proposed, but none has achieved complete acceptance. Treatment, depending on the extent of disease, includes surgery (thoracoscopy with sclerosis, pleurectomy, extrapleural pneumonectomy), radiation therapy, and, often, chemotherapy. These three modalities individually have not significantly improved survival rates, and chemotherapy or radiation therapy does not provide long-term palliation or relief of symptoms. Combination chemotherapy with pemetrexed and cisplatin is associated with symptomatic relief in patients with advanced, unresectable malignant mesothelioma and with a significant response rate in advanced disease. It is occasionally used as preoperative or induction chemotherapy before extrapleural pneumonectomy followed by radiation therapy. Multimodality and molecularly targeted agents are being evaluated. Mesothelioma unfortunately remains a nearly universally fatal illness, with a median survival of less than 12 months from the time of diagnosis.

Neuroendocrine and Other Lung Tumors

Neuroendocrine lung tumors are classified into four types: carcinoid tumors, atypical carcinoids, SCLC, and large cell neuroendocrine carcinomas. *Carcinoid tumors* (Chapter 232) are low-grade neuroendocrine tumors with a 10-year survival rate greater than 90%. Atypical carcinoid tumors are intermediate-grade tumors, significantly more aggressive than carcinoid, with 10-year survival of below 20%. *Large cell neuroendocrine carcinoma* is an aggressive neuroendocrine tumor that does not meet the criteria for carcinoid, atypical carcinoid, or SCLC. Considerable activity for everolimus and other inhibitors of the mammalian target of rapamycin (mTOR) has been demonstrated in low-grade carcinoid tumors.

Carcinoid tumors of the lung account for 1 to 2% of all lung neoplasms. These neuroendocrine tumors trace their origin to the Kulchitsky cell present in bronchial epithelium. Typical and atypical carcinoids differ in the number of mitoses (<2 per 10 high-power fields vs. 2 to 10 per 10 high-power fields, respectively), nuclear pleomorphism (absent vs. present), and regional lymph node metastases (5 to 15% vs. 20 to 28%). Distant metastases at initial evaluation are rare (<20%). Patients with typical carcinoid tumors usually live for many years, whereas patients with atypical carcinoid tumors have a 5-year mortality rate of 61 to 88%. Carcinoid tumors are not associated with cigarette smoking, are twice as common in women as in men, usually occur in patients younger than 40 years, and arise in the perihilar area of the lung. Treatment of bronchial carcinoid tumors is based on the stage of the disease. Mediastinal staging is usually followed by surgical resection. With mediastinal lymph node involvement, radiation therapy is recommended for typical carcinoid tumors if surgery cannot be performed. For atypical carcinoid or metastatic disease, chemotherapy (etoposide plus cisplatin every 3 weeks for four cycles) plus radiation therapy is commonly used, but there is no evidence for the benefit of one therapy over another.

Salivary gland carcinomas include mucoepidermoid carcinoma and adenoid cystic carcinoma, which represent approximately 0.2% of lung cancers. These slow-growing neoplasms arise from the bronchial glands and are usually treated with surgery.

Primary sarcomas of the lung are very rare and include malignant fibrous histiocytoma, fibrosarcoma, leiomyosarcoma, rhabdomyosarcoma, epithelioid hemangioendothelioma, angiosarcoma, and liposarcoma (Chapter 202). The primary treatment is surgery, but depending on the size and grade of the tumor and whether the margins are clear, radiation therapy or chemotherapy may also be given. Primary lymphomas of the lung are also very rare, accounting for approximately 0.3% of all primary lung cancers. The most common type is a low-grade small lymphocytic lymphoma, for which surgery and chemotherapy are the usual treatments (Chapter 185).

Grade A References

A1. Curran WJ Jr, Paulus R, Langer CJ, et al. Sequential vs. concurrent chemoradiation for stage III non–small cell lung cancer: randomized phase III trial RTOG 9410. *J Natl Cancer Inst*. 2011;103:1452-1460.

A2. Lee JK, Hahn S, Kim DW, et al. Epidermal growth factor receptor tyrosine kinase inhibitors vs conventional chemotherapy in non-small cell lung cancer harboring wild-type epidermal growth factor receptor: a meta-analysis. *JAMA*. 2014;311:1430-1437.

A3. Shaw AT, Kim DW, Nakagawa K, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med*. 2013;368:2385-2394.

A4. Shaw AT, Kim DW, Mehra R, et al. Ceritinib in ALK-rearranged non-small-cell lung cancer. *N Engl J Med*. 2014;370:1189-1197.

A5. Temel JS, Greer JA, Muzikansky A, et al. Early palliative care for patients with metastatic non-small-cell lung cancer. *N Engl J Med*. 2010;363:733-742.

A6. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395-409.

A7. Kovalchik SA, Tammemagi M, Berg CD, et al. Targeting of low-dose CT screening according to the risk of lung-cancer death. *N Engl J Med*. 2013;369:245-254.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Lantz PM, Mendez D, Philbert MA. Radon, smoking, and lung cancer: the need to refocus radon control policy. *Am J Public Health*. 2013;103:443-447.
2. Kerr KM. Clinical relevance of the new IASLC/ERS/ATS adenocarcinoma classification. *J Clin Pathol*. 2013;66:832-838.
3. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012;489:519-525.
4. Kris MG, Johnson BE, Berry LD, et al. Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *JAMA*. 2014;311:1998-2006.
5. Lindeman NI, Cagle PT, Beasley MB, et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. *Arch Pathol Lab Med*. 2013;137:828-860.
6. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med*. 2013;369:910-919.
7. Lackey A, Donington JS. Surgical management of lung cancer. *Semin Intervent Radiol*. 2013;30:133-140.
8. Nagata Y. Sterotactic body radiotherapy for early stage lung cancer. *Cancer Res Treat*. 2013;45:155-161.
9. Heitzer E, Auer M, Ulz P, et al. Circulating tumor cells and DNA as liquid biopsies. *Genome Med*. 2013;5:73.
10. Gould MK. Clinical practice. Lung-cancer screening with low-dose computed tomography. *N Engl J Med*. 2014;371:1813-1820.
11. Prosch H, Schaefer-Prokop C. Screening for lung cancer. *Curr Opin Oncol*. 2014;26:131-137.
12. Moyer VA. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160:330-338.
13. Patz EF Jr, Campa MJ, Gottlin EB, et al. Biomarkers to help guide management of patients with pulmonary nodules. *Am J Respir Crit Care Med*. 2013;188:461-465.
14. van Zandwijk N, Clarke C, Henderson D, et al. Guidelines for the diagnosis and treatment of malignant pleural mesothelioma. *J Thorac Dis*. 2013;5:E254-E307.

REVIEW QUESTIONS

1. Which of the following individuals should not consider computed tomography (CT) screening for lung cancer because of an absence of increased risk?

A. Individuals who are older than 55 years and have at least 35 pack-years of smoking cigarettes
B. A woman with a heavy smoking history (more than 50 pack-years) and evidence for emphysema/chronic obstructive pulmonary disease
C. A former smoker who is 70 years old with a 60 pack-year history but who quit smoking 5 years ago
D. A patient with a prior history of laryngeal cancer treated by chemotherapy and radiation therapy, with a 40 pack-year history, who is down to one or two cigarettes per day
E. A 35-year-old man whose father died of lung cancer after a long smoking history

Answer: E There is no evidence that nonsmoking individuals who have family histories of lung cancer, particularly lung cancer that is attributable to smoking, will benefit by being screened with CT scans. The best evidence to date for the utility of CT screening is for individuals older than 55 years with at least a 30 to 35 pack-year smoking history. These individuals are more likely to have a nodule detected by CT scan than by chest radiography, as shown by the NLST study. Thus, individuals A, B, C, and D, all of whom are older than 55 years and have at least a 30 to 35 pack-year smoking history, are at increased risk for lung cancer, and all could potentially benefit from screening with CT scan.

2. Which of the following is most consistent with a tissue diagnosis of a lung adenocarcinoma?

A. TTF-1 positive, CK7 positive, EGFR driver mutation, and CK20 negative and p63 negative
B. TTF-1 negative, p63 positive, p40 positive, FGFR amplification, and EGFR wild type
C. Synaptophysin and chromogranin positive, p53 mutation and Rb mutation
D. Tumor positive for vimentin and melanin, negative for TTF-1 and for all cytokeratins
E. CD20 positive, negative for TTF-1 and for all cytokeratins

Answer: A Patients with lung adenocarcinoma are likely to have a TTF-1–positive tumor, with CK7 expressed. CK20 is more commonly expressed in colorectal cancers than in lung tumors. The fact that p63 is negative in this case makes this less likely to be a squamous cell carcinoma. The clearest indicator here is that the patient has an EGFR driver mutation. Mutations in exons 19 or 21 of EGFR are pathognomonic for lung adenocarcinoma, particularly in never-smokers. The other three cases are all less likely to be adenocarcinomas. Case B is p63 positive, which makes it more likely to be squamous cell cancer, and all the other features, including being p40 positive and FGFR amplification, are more frequently seen in squamous cell cancers. EGFR is mutated almost exclusively in lung adenocarcinomas as opposed to squamous cell carcinomas, although the fact that this case B is wild type does not make it less likely to be an adenocarcinoma. Case C is more likely to be a small cell carcinoma, as synaptophysin and chromogranin are positive for neuroendocrine tumors, and the combination of the p53 and Rb mutations makes this likely to be a small cell lung cancer. In case D, the tumor is positive for vimentin and melanin and negative for TTF-1 and all cytokeratins; this is not an epithelial tumor and is likely to be a melanoma or poorly differentiated sarcoma. The tumor in E that is CD20 positive is likely to be a lymphoma, and this is again negative for cytokeratins that are expressed on epithelial tumors as well as for TTF-1, which would make this highly unlikely to be an epithelial tumor in general and a lung cancer in particular.

3. A 63-year-old woman with a heavy smoking history presents with a diagnosis of a squamous cell lung cancer, staged radiographically as T3N2M0. She has been referred to a thoracic surgeon who is planning to perform a mediastinoscopy. If the mediastinoscopy confirms the presence of tumor cells in her mediastinal lymph nodes, the best recommendation for the multidisciplinary clinic is

A. Surgical resection of her tumor by a pneumonectomy and clearing of her mediastinum followed by radiation therapy to any positive margins
B. Induction or preoperative chemotherapy followed by radiation therapy followed by surgery
C. Concurrent chemotherapy and radiation therapy
D. Induction chemotherapy followed by radiation therapy
E. Palliative chemotherapy followed by hospice referral

Answer: C This is the presentation of a patient with stage 3A non–small cell lung cancer. A patient such as this with a good performance status would best be treated with concurrent chemotherapy and radiation therapy, which would give a 15 to 20% chance of 5-year survival. Surgical resection of the tumor by pneumonectomy and clearing of the mediastinum followed by radiation therapy is a highly morbid procedure and a particularly poor choice, given that the mediastinal lymph nodes are involved. This patient is at least as likely to fail at distant sites as in the mediastinum, and chemotherapy is as essential as radiation therapy to this patient. Option B is suboptimal in that randomized trials have shown that concurrent chemotherapy and radiation therapy is superior to induction or preoperative chemotherapy followed by radiation therapy. Following chemotherapy and radiation therapy with surgery is controversial, with evidence that patients receiving this trimodality therapy with all but the most limited mediastinal lymph nodes are more likely to do poorly. Option D is again inferior to option C as randomized trials have shown over time. Answer E is the most inappropriate of all. This is a patient with a 15 to 20% chance for a cure, and it is inappropriate to treat her with palliative chemotherapy followed by hospice referral, as long as the patient's performance status is good and the patient is willing to be treated fully in an attempt at cure.

4. A 70-year-old man with lung adenocarcinoma, diagnosed as stage IVB (T2N2M1) with metastases to the lymph nodes, bilateral adrenal glands, and liver, presents with an excellent performance status, an Eastern Cooperative Oncology Group (ECOG) status 1. He has never received prior chemotherapy for his lung cancer. He has no evidence of a driver mutation on genomic analysis of his tumor. Which of the above is *not* an acceptable approach to his management?

A. Combination chemotherapy with carboplatin, paclitaxel, and bevacizumab
B. Combination therapy with carboplatin and pemetrexed
C. Combination chemotherapy with supportive care initiated at the time of the first visit
D. Single-agent paclitaxel plus erlotinib, the EGFR tyrosine kinase inhibitor
E. A second opinion followed by participation in a frontline clinical trial combining chemotherapy and immunotherapy

Answer: D This question describes a patient with good performance status who does not have a driver mutation that would increase the likelihood of benefiting from frontline targeted therapy. As such, there are many reasonable options, all of them starting with platinum-based chemotherapy. ECOG-5597 showed that the combination of carboplatin, paclitaxel, and bevacizumab was associated with a median survival of more than 12 months, and studies with combination therapy with cisplatin or carboplatin and pemetrexed indicate similar survivals. Option C may be the best option of all, as recent data indicate that combination chemotherapy with early initiation of supportive care appears to be associated with a survival advantage. Finally, option E is also an excellent one with evidence that patients who do not have curable disease should strongly consider participating in a frontline clinical trial, in this case, combining immunotherapy and chemotherapy. In fact, the only inappropriate option is D, which is the correct answer to this question. Patients with lung cancer benefit from combination chemotherapy, and there is little evidence that combining paclitaxel with a small-molecule EGFR inhibitor such as erlotinib provides equivalent benefit to combination chemotherapy alone. Therefore, in the absence of a driver *EGFR* mutation, all options are acceptable, making D the only wrong choice.

5. A 50-year-old never-smoking woman presents with a headache followed by a seizure. Her emergent work-up reveals a solitary brain lesion, which is resected and reveals lung adenocarcinoma containing an EGFR driver mutation in exon 19. After consideration of radiation therapy to the brain, the optimal treatment approach would be

A. Combination chemotherapy with carboplatin, paclitaxel, and bevacizumab
B. Combination therapy with carboplatin and pemetrexed
C. Combination chemotherapy with supportive care initiated at the time of the first visit
D. Initiation of erlotinib, the EGFR tyrosine kinase inhibitor
E. A second opinion followed by participation in a frontline clinical trial combining chemotherapy and immunotherapy

Answer: D This question describes a woman who has an *EGFR* mutation–driven lung cancer. She presents with a solitary brain lesion and is treated. The patient may well benefit from radiation therapy, either sterotactic radiation to the residual lesion or whole brain irradiation, but it is clear that the optimal therapy for her, based on the randomized trial published by Mok along with an abundance of other data, is to initiate her treatment with erlotinib or gefitinib, first-generation tyrosine kinase inhibitors, which, in this population of patients, are far more active than chemotherapy combinations. Therefore, option D, given the documented *EGFR* mutation, is superior to options A, B, and C. Finally, whereas E is reasonable from the perspective that all patients with lung cancer should strongly consider a second opinion, this patient is far more likely to benefit from erlotinib or other EGFR tyrosine kinase inhibitors than from chemotherapy and immunotherapy.

192

NEOPLASMS OF THE ESOPHAGUS AND STOMACH

ANIL K. RUSTGI

NEOPLASMS OF THE ESOPHAGUS

DEFINITION

The esophagus is a hollow tubular organ with primary physiologic functions related to contraction to permit propulsion of solid and liquid food contents into the stomach. The mucosa is a stratified squamous epithelium that covers the submucosa and muscle; the latter is skeletal muscle in the proximal esophagus and smooth muscle in the mid-distal esophagus. Cancers of the esophagus may be classified broadly into epithelial versus nonepithelial.

There are benign epithelial tumors referred to as squamous cell papillomas. Malignant epithelial tumors are classified into two main subtypes: esophageal squamous cell carcinoma and esophageal adenocarcinoma. Other, less common, esophageal epithelium-derived cancers include verrucous squamous cell carcinoma, adenosquamous carcinoma, adenoid cystic carcinoma, and mucoepidermoid carcinoma. Benign nonepithelial tumors include leiomyoma, granular cell tumors, fibrovascular polyp, hemangioma, lymphangioma, lipoma, and fibroma. Malignant nonepithelial tumors include leiomyosarcoma and other sarcomas, metastatic carcinoma (originating from breast, lung), and lymphoma.

Esophageal Squamous Cell Carcinoma

EPIDEMIOLOGY

Esophageal squamous cell carcinoma is the more common type of esophageal cancer worldwide and represents a leading cause of cancer-related mortality in men. Esophageal squamous cell carcinoma may have rates of up to 100 per 100,000 population in what is often termed the *Central Asian belt*, including regions around the Caspian Sea, Iran, India, and China; other areas of high incidence include some Mediterranean countries and South Africa. In the United States, esophageal squamous cell carcinoma is more common among African American males than white males, with risks of 15.1 per 100,000 compared with 2.9 per 100,000, respectively. Overall, although the U.S. incidence of esophageal squamous cell carcinoma is low in males or females younger than 50 years, it does increase with advancing age.

Risk Factors

Cancers in general are viewed in the context of hereditary or inherited forms versus sporadic or seemingly random diseases that are related to age, environmental exposures, and genetic alterations (Table 192-1). That being said, the hereditary basis for esophageal squamous cell carcinoma is exceedingly rare, consisting of a desquamating condition termed *tylosis palmaris et plantaris*. As implied, the desquamation most dramatically affects the hands and feet, but this extends to the esophagus as well. Another uncommon condition, Plummer-Vinson syndrome or Paterson-Brown Kelly syndrome, entails glossitis, cervical esophageal webs, and iron deficiency anemia. In both conditions, it is likely that chronic inflammation triggers the cascade of events that culminate in esophageal dysplasia and esophageal squamous cell carcinoma.

The preponderance of esophageal squamous cell carcinoma cases are attributable to cigarette smoking or alcohol, but especially so in combination, because there appear to be synergistic deleterious effects of various chemical carcinogens in both, including *N*-nitroso compounds, polycyclic aromatic hydrocarbons, and aromatic amines. The relative risk for esophageal squamous cell carcinoma is 6.2 in those who smoke more than 25 cigarettes on a daily basis. Cessation of cigarette smoking is helpful in attenuating risk after 10 years of abstinence. Cigarette smokers who partake in beer and whiskey have a 10- to 25-fold enhanced risk of developing esophageal squamous cell carcinoma. Indeed, it is the type of alcohol and the manner of distillation that are most critical. In endemic areas of the world, deficiencies of vitamins A, B_{12}, C, and E, folic acid, and certain minerals (zinc, selenium, molybdenum) are important risk factors. All these vitamins and minerals exert direct or indirect antioxidant effects, and their deficiencies impair epithelial and tissue homeostasis and regeneration.

Other risk factors for esophageal squamous cell carcinoma[1] include achalasia (Chapters 136 and 138), a disorder that involves agangliosis of Auerbach's plexus, resulting in dysphagia, chest pain, and weight loss, among other symptoms. The emergence of esophageal squamous cell carcinoma may be observed 10 to 20 years after the identification of achalasia in patients. Because head and neck squamous cell carcinoma (HNSCC) (Chapter 190) shares many of the environmental and lifestyle risk factors with esophageal squamous cell carcinoma, particularly alcohol and tobacco smoking, HNSCC and esophageal squamous cell carcinoma may occur synchronously or metachronously. In different parts of the world, esophageal squamous cell carcinoma is also associated with chronic esophageal stricture due to lye ingestion, consumption of maté (a hot herb-based beverage), celiac sprue, human papillomavirus (HPV) infection (especially genotypes HPV-16, HPV-18, and HPV-33), and radiation injury.

TABLE 192-1 RISK FACTORS FOR ESOPHAGEAL CANCER

Esophageal squamous cell cancer
Tylosis palmaris et plantaris
Achalasia
Plummer-Vinson syndrome
Cigarette smoking
Alcohol
Chronic lye ingestion
Human papillomavirus infection
Radiation injury
Celiac sprue
Esophageal adenocarcinoma
Gastroesophageal acid reflux
Bile reflux
Obesity
Barrett esophagus

PATHOBIOLOGY

Esophageal squamous cell carcinoma involves the transition from normal squamous epithelium to squamous dysplasia to cancer. Esophageal squamous cell carcinoma initiation, progression, and metastasis are associated with a number of genetic alterations. Among these genetic alterations are overexpression of epidermal growth factor receptor *(EGFR)* and cyclin D1 oncogenes, and inactivation of *TP53*, *p16INK4A*, *E-cadherin*, and p120-catenin *(p120ctn)* tumor suppressor genes. The frequency of these changes varies greatly based upon various studies, but the oncogenic alterations generally appear early in dysplasia and early esophageal squamous cell carcinoma, whereas the inactivation of tumor suppressor genes appear as later events in established primary and metastatic esophageal squamous cell carcinoma. From a genomic viewpoint, the SOX-2 transcription factor, important in the pluripotent capacity of somatic cells, has been shown to be an important gene involved in esophageal squamous cell carcinoma pathogenesis and transformation by virtue of SOX-2 amplification. The ability to model esophageal squamous cell carcinoma in vitro and in vivo has witnessed great strides in recent years through the advent and characterization of three-dimensional organotypic culture models, xenograft transplantation mouse models, and genetically engineered mouse models. For example, conditional knockout of the *p120ctn* tumor suppressor gene in the esophagus of mice results in invasive esophageal squamous cell carcinoma.

CLINICAL MANIFESTATIONS

Symptoms and Signs

Although esophageal squamous dysplasia is typically not associated with symptoms, esophageal squamous cell carcinoma, which has a predilection for the proximal to midesophagus, may be associated with dysphagia, odynophagia, atypical or typical chest pain, gastrointestinal bleeding, nausea, vomiting, weight loss, and malnutrition. Esophageal squamous cell carcinoma may metastasize to local lymph nodes, lung, liver, and bone. Symptoms attributable to metastatic esophageal squamous cell carcinoma may involve bone-related pain, dyspnea, and evidence of jaundice and liver failure, depending upon the extent of metastatic disease.

Physical Examination

The patient should be evaluated for changes in hair, skin integrity, and nail beds as a reflection of malnutrition. Weight loss may result in general cachexia and muscle wasting. There may be lymphadenopathy in the anterior cervical and superclavicular regions. Hepatomegaly and complications of liver disease may be present with metastatic disease to the liver.

Laboratory Studies

There may be progressive iron deficiency anemia due to chronic indolent upper gastrointestinal bleeding. Additional abnormalities may be reflected in metabolic disturbances, such as metabolic alkalosis due to vomiting and hypernatremia due to dehydration. Liver enzyme abnormalities, both hepatocellular and cholestatic, may reflect metastasis to the liver. There are no specific markers for esophageal squamous cell carcinoma, but an elevated carcinoembryonic antigen (CEA) level may be used to aid in monitoring disease recurrence after therapy.

DIAGNOSIS

Barium swallow radiography is useful for the diagnosis of esophageal squamous cell carcinoma, with depiction of a filling defect due to the mucosal lesion or impaired transit of barium due to luminal growth (Chapter 138). However, definitive diagnosis involves direct visualization with upper endoscopy (Chapter 134); once the mass is visualized, biopsies are necessary for confirmation by histopathology and immunohistochemistry for

TABLE 192-2 TNM STAGING SYSTEM FOR CANCER OF THE ESOPHAGUS (AMERICAN JOINT COMMITTEE ON CANCER CRITERIA)

PRIMARY TUMOR (T)*	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	High-grade dysplasia†
T1	Tumor invades lamina propria, muscularis mucosae, or submucosa
T1a	Tumor invades lamina propria or muscularis mucosae
T1b	Tumor invades submucosa
T2	Tumor invades muscularis propria
T3	Tumor invades adventitia
T4	Tumor invades adjacent structures
T4a	Resectable tumor invading pleura, pericardium, or diaphragm
T4b	Unresectable tumor invading other adjacent structures, such as aorta, vertebral body, trachea, etc.
*(1) At least maximal dimension of the tumor must be recorded, and (2) multiple tumors require the T(m) suffix. †High-grade dysplasia includes all noninvasive neoplastic epithelia that was formerly called carcinoma in situ.	
LYMPH NODE (N)*	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in 1-2 regional lymph nodes
N2	Metastasis in 3-6 regional lymph nodes
N3	Metastasis in 7 or more regional lymph nodes
*Number must be recorded for total number of regional nodes sampled and total number of reported nodes with metastasis.	
DISTANT METASTASIS (M)	
MX	Metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis
From *AJCC Cancer Staging Manual*. 7th ed. New York: Springer-Verlag; 2010.	

cytokeratins associated with proliferation and differentiation. Esophageal squamous cell carcinoma may involve local lymph nodes, which are best detected by endoscopic ultrasound (EUS); as needed, samples can then be analyzed by cytopathology following fine-needle aspiration (FNA). In high-volume centers in the United States, the cytopathologist will be in the procedure room with the gastroenterologist to provide an initial evaluation of the specimens obtained through FNA. Evaluation of metastatic disease involves chest and abdominal computed tomography (CT) scans. Bone scan might be useful in patients who are symptomatic with bone-related pain. Positron emission tomography (PET) has become increasingly used in some settings. In totality, these diagnostic modalities also allow for staging of esophageal squamous cell carcinoma (Table 192-2), which is important in guiding therapeutic options.[2]

TREATMENT

Surgical Therapy

Surgery is the cornerstone of therapy for curative intent. Technical advances have led to improvements in both operative mortality and postoperative morbidity. The different surgical techniques include transthoracic, transhiatal, and radical en bloc resections. Depending upon the location of the esophageal squamous cell carcinoma, either total esophagectomy or subtotal esophagectomy is pursued. For the latter, jejunal or colonic interposition can be done. Although there is currently a lack of high-quality studies comparing minimally invasive esophagectomy (MIE) to conventional approaches, MIE can achieve equivalent or better perioperative mortality, morbidity, and oncologic outcomes compared to open surgery in selected patients.[3]

Medical Therapy

Depending upon the stage of disease, there is some variation in whether to proceed with preoperative (neoadjuvant) chemoradiation therapy (preferred for early stage) or postoperative (adjuvant) chemoradiation therapy.

A study in which patients were randomized to receive surgery alone or surgery plus postoperative chemotherapy with 5-fluorouracil and leucovorin and concurrent radiation therapy revealed that the median survival was 36 months for patients in the adjuvant arm compared with 26 months for those in the surgery-only arm. The 3-year overall survival rates were 50% (surgery plus adjuvant therapy) compared with 40% (surgery alone), respectively.

PROGNOSIS

The 5-year survival for treated esophageal squamous cell carcinoma is dependent upon stage and types of therapies used. For stages T1 and T2 without lymph node involvement, surgery alone may be curative in more than 60% of cases. The occurrence of major postoperative complications, which occur in about one third of patients, exerts a long-lasting negative effect on health-related quality of life in patients who survive for 5 years after esophagectomy for cancer. Dyspnea, fatigue, eating restriction, sleep difficulty, and gastroesophageal reflux progressively worsen more during follow-up in those who suffer such major postoperative complications compared with those without major surgical complications. For patients with metastatic disease, therapy is palliative, involving endoscopically placed expandable prosthetic stents to open the nearly obstructed lumen for passage of food contents, percutaneous endoscopic gastrotomy tubes for delivery of nutrition to the stomach distal to the mass lesion, total parenteral nutrition, pain control, and systemic chemotherapy.

Esophageal Adenocarcinoma

EPIDEMIOLOGY

Esophageal adenocarcinoma affects whites more than African Americans and males much more than females (3 : 1 to 5.5 : 1); it increases in incidence after the age of 40 years. The age-adjusted incidence annually is 1.3 per 100,000. In this chapter, esophageal adenocarcinoma is discussed as a separate entity from gastroesophageal (GE) adenocarcinomas (so-called GE junctional cancer) and gastric cardia adenocarcinomas, although there has been an increasing tendency to think of these in aggregate. The incidence of esophageal adenocarcinoma is increasing dramatically in developed countries, especially in the United States (by 4 to 10% annually) and Western/Northern Europe.

Etiology

Obesity (central) is an important risk factor for esophageal adenocarcinoma. This may be related to either mechanical factors that foster greater gastroesophageal reflux disease (GERD) or the release of proinflammatory cytokines and chemokines that track to the esophagus, or both. The major recognized precursor of esophageal adenocarcinoma is Barrett esophagus. Barrett esophagus is the replacement of the normal stratified squamous epithelium by an incomplete small intestinal epithelium (metaplasia) in the distal esophagus, projecting from the GE junction in a distal-proximal gradient. In turn, it has been demonstrated that Barrett esophagus is fostered by GERD but also by an admixture of bile acids in the acid refluxate. Patients with scleroderma (Chapter 267) may be at increased risk for Barrett esophagus. A small subset of Barrett esophagus patients may progress to esophageal adenocarcinoma through intermediate stages of low-grade and high-grade dysplasia. In Barrett esophagus, one case of esophageal adenocarcinoma is estimated to arise in 55 to 441 patient-years, which corresponds to an approximately 125-fold increased risk for esophageal adenocarcinoma compared with that in the general population.

PATHOBIOLOGY

Barrett esophagus involves transdifferentiation from normal esophageal epithelium to an epithelium of the small intestine with columnar enterocytes and secretory goblet cells, but without Paneth cells and enteroendocrine cells—hence the designation of incomplete intestinal metaplasia. By itself, the metaplasia of Barrett esophagus cannot become esophageal adenocarcinoma. However, if and when Barrett esophagus transitions to low-grade and high-grade dysplasia, there is the aforementioned risk for esophageal adenocarcinoma. Barrett esophagus is associated with abnormal DNA ploidy based on flow cytometry analysis, and certain genetic alterations in epidermal growth factor receptor signaling, *TP53*, and *p16INK4A*. Microsatellite instability may be noted as well. Whole genome approaches are revealing gains and losses of chromosomal regions that might lead to identification of known and previously unknown genes critical in the pathogenesis of esophageal

adenocarcinoma.[4] Recent advances in genetically engineered mouse models have allowed the phenocopying of Barrett esophagus and esophageal adenocarcinoma through the direct targeting of interleukin (IL)-1β to the esophagus; and in another approach, through the global knockout of p63 (an important marker of stem cells and progenitor cells), Barrett esophagus is evident in the postnatal period.

CLINICAL MANIFESTATIONS

Symptoms and Signs

It is estimated that 5 to 15% of GERD patients may develop Barrett esophagus, but such population-based studies are difficult to pursue because vast millions of people are affected with GERD, and most GERD patients do not undergo upper endoscopy. Patients with Barrett esophagus may or may not have symptoms related to GERD. Chronic GERD with Barrett esophagus may be associated with distal esophageal strictures. With esophageal adenocarcinoma, patients may suffer from dysphagia, odynophagia, upper gastrointestinal bleeding, chest pain, nausea, vomiting, early satiety, weight loss, and malnutrition.

Physical Examination

Examination of the patient may reveal signs consistent with malnutrition and weight loss. Lymph adenopathy should be explored. There may be hepatomegaly. Paraneoplastic syndromes are unusual with esophageal adenocarcinoma (as well as with esophageal squamous cell carcinoma). Nevertheless, it is important to ensure that esophageal adenocarcinoma is not mistaken for a benign entity such as a primary esophageal motility disorder.

Laboratory Studies

Patients with esophageal adenocarcinoma may have iron deficiency anemia, metabolic derangements, and abnormal liver enzyme tests owing to metastatic disease. CEA may be elevated as a tumor serologic marker.

DIAGNOSIS

Barium swallow radiography may lead one to the suspicion of Barrett esophagus and can diagnose luminal mass lesions consistent with esophageal adenocarcinoma in the distal esophagus. However, the mainstay of diagnosis is upper endoscopy. At that time, a characteristic salmon-colored mucosa is visualized at the GE junction, with frondlike projections in a proximal direction. If the extent of Barrett esophagus is 3 cm or less, it is termed *short-segment Barrett esophagus*; if it is more than 3 cm, it is referred to as *long-segment Barrett esophagus*. This distinction is important in that the risk for esophageal adenocarcinoma in long-segment Barrett esophagus is greater than in short-segment Barrett esophagus. Noting that the normal esophageal mucosa is more pinkish-white in hue, one can visually distinguish the two different types of epithelia, with the caveat that the gastric cardia mucosa at the GE junction should not be mistaken for Barrett esophagus. Endoscopic mucosal biopsies from the Barrett esophagus region (with control biopsies from the normal esophagus and gastric cardia) are required for histopathologic diagnosis. Features of dysplasia are best appreciated in the absence of reflux-related esophagitis that can lead to nuclear architectural distortion; hence, suppression of acid production with proton pump inhibitor therapy for 6 to 8 weeks is needed, with a view to repeat biopsies.

If the patient has Barrett esophagus metaplasia, upper endoscopy should be repeated every 3 years. However, low-grade dysplasia (with confirmation by an expert pathologist) requires surveillance endoscopy every 6 to 12 months. High-grade dysplasia, if properly evaluated by the pathologist, may require reconfirmation, but then leads to either medical (radio frequency ablation [RFA], endoscopic mucosal resection) or surgical intervention because of the possibility of missed contiguous esophageal adenocarcinoma. EUS may be helpful in discriminating between high-grade dysplasia and esophageal intramucosal adenocarcinoma.

TREATMENT

Rx

The principles are very similar to those applied to esophageal squamous cell carcinoma in terms of surgery. Preoperative chemoradiotherapy (carboplatin titrated to achieve an area under the curve of 2 mg/mL/min and paclitaxel 50 mg/m^2 for 5 weeks) and concurrent radiotherapy (at 41.4 Gy in 23 fractions, 5 days per week) significantly improves median survival from 24 months to 49 months among patients with potentially curable esophageal or esophagogastric-junction cancer.[■] Overall prognosis for esophageal adenocarcinoma is not too dissimilar from that noted in esophageal squamous cell carcinoma. Major improvements in the treatment options for Barrett esophagus, the main precursor of esophageal adenocarcinoma, have witnessed dramatic growth (Chapter 138). RFA reduces progression of Barrett esophagus with low-grade dysplasia or possibly high-grade dysplasia to cancer.[■] Endoscopic mucosal resection (EMR) is used for Barrett esophagus–related high-grade dysplasia or Barrett esophagus associated with intramucosal esophageal adenocarcinoma. For advanced, unresectable tumors, self-expanding stents, often with localized brachytherapy, can provide palliation.[5]

NEOPLASMS OF THE STOMACH

DEFINITION

Gastric neoplasms are predominantly malignant, and nearly 90 to 95% of these tumors are adenocarcinomas. Less frequently observed malignant diseases include lymphomas, especially non-Hodgkin's lymphoma, and sarcomas such as leiomyosarcoma. Benign gastric neoplasms include leiomyomas, carcinoid tumors, and lipomas.

Adenocarcinoma of the Stomach

EPIDEMIOLOGY

The great geographic variation in the incidence of gastric cancer worldwide indicates that environmental factors influence the pathogenesis of gastric carcinogenesis. Further support for this notion comes from observations that groups emigrating from high-risk to low-risk areas, such as Japanese individuals moving to Hawaii and Brazil, acquire the low risk of the area into which they emigrate, presumably because of adoption of the endogenous lifestyle and exposure to different environmental factors.

Gastric adenocarcinoma was the most frequently observed malignant disease in the world until the mid-1980s, and it remains extremely common among men in certain regions such as tropical South America, some parts of the Caribbean, and Eastern Europe. Regardless of gender, it remains one of the most common malignancies in Japan and China.

Whereas gastric cancer was the most common cancer in the United States in the 1930s, its annual incidence has steadily decreased. The annual incidence is now fewer than 10,000 new cases per year. However, although the incidence of gastric adenocarcinoma localized to the distal stomach has declined, the incidence of proximal gastric and GE junctional adenocarcinomas has been steadily increasing in the United States, a finding that perhaps reflects differences in pathogenic factors. Typically, gastric cancer occurs between the ages of 50 and 70 years and is uncommon before age 30. The rates are higher in men than in women by 2 : 1. Five-year survival is less than 20%.

Risk Factors

Risk factors for the development of gastric adenocarcinoma[6] can be divided into environmental and genetic factors as well as precursor conditions (Table 192-3). For example, *Helicobacter pylori* infection is significantly more common in patients with gastric cancer than in matched control groups. Epidemiologic studies of high-risk populations have also suggested that genotoxic agents such as *N*-nitroso compounds may play a role in gastric tumorigenesis. *N*-nitroso compounds can be formed in the human stomach by nitrosation of ingested nitrates, which are common constituents of the diet. High nitrate concentrations in soil and drinking water have been observed in areas with high death rates from gastric cancer. Atrophic gastritis (Chapter 139), with or without intestinal metaplasia, is observed in association with gastric cancer, especially in endemic areas. Pernicious anemia (Chapter 139) is associated with a several-fold increase in gastric cancer. Atrophic gastritis and gastric cancer have certain environmental risk factors in common. It is likely that atrophic gastritis and intestinal metaplasia represent intermediary steps to gastric cancer. The achlorhydria associated with gastritis related to *H. pylori* infection, pernicious anemia, or other causes favors the growth of bacteria capable of converting nitrates to nitrites. The nitrosamine *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine causes a high rate of induction of adenocarcinoma in the glandular stomach of rats. At the same time, most patients with atrophic gastritis do not develop gastric cancer, a finding suggesting that neither atrophic gastritis nor achlorhydria alone is responsible.

Benign gastric ulcers do not appear to predispose patients to gastric cancer. However, patients who have a gastric remnant after subtotal gastrectomy for benign disorders have an increased relative risk for gastric cancer of 1.5 to 3.0 by 15 to 20 years after surgery.

TABLE 192-3 CONDITIONS PREDISPOSING TO OR ASSOCIATED WITH GASTRIC CANCER

ENVIRONMENTAL
Helicobacter pylori infection
Dietary: excess of salt (salted pickled foods), nitrates/nitrites, carbohydrates; deficiency of fresh fruit, vegetables, vitamins A and C, refrigeration
Low socioeconomic status
Cigarette smoking
GENETIC
Familial gastric cancer (rare)
Associated with hereditary nonpolyposis colorectal cancer
Blood group A
PREDISPOSING CONDITIONS
Chronic gastritis, especially atrophic gastritis with or without intestinal metaplasia
Pernicious anemia
Intestinal metaplasia
Gastric adenomatous polyps (>2 cm)
Postgastrectomy stumps
Gastric epithelial dysplasia
Ménétrier's disease (hypertrophic gastropathy)
Chronic peptic ulcer

PATHOBIOLOGY

Gastric adenocarcinomas can be divided into two types based on the Lauren classification: intestinal and diffuse. The intestinal type is typically in the distal stomach with ulcerations, is often preceded by premalignant lesions, and is declining in incidence in the United States. By contrast, the diffuse type involves widespread thickening of the stomach, especially in the cardia, and it often affects younger patients; this form may present as linitis plastica, a nondistensible stomach with the absence of folds and a narrowed lumen caused by infiltration of the stomach wall with tumor. Diffuse-type gastric cancers harbor mucin-producing cells. Other conditions may result in linitis plastica, such as lymphoma (Chapter 185), tuberculosis (Chapter 324), syphilis (Chapter 319), and amyloidosis (Chapter 188). The prognosis is generally worse in the diffuse type.

Key histopathologic features of gastric cancer include its degree of differentiation, invasion through the gastric wall, lymph node involvement, and the presence or absence of signet ring cells within the tumor itself. Other pathologic manifestations include a polypoid mass, which may be difficult to distinguish from a benign polyp. Early gastric cancer, a condition that is common in Japan and has a relatively favorable prognosis, consists of superficial lesions with or without lymph node involvement. Here, the Borrmann classification scheme is helpful: I, polypoid; II, fungating ulcer with sharp raised margins; III, ulcer with poorly defined infiltrative margins; and IV, infiltrative, mostly intramural lesion, not well demarcated.

The leading hypothesis explaining the way in which *H. pylori* predisposes to gastric cancer risk is the induction of an inflammatory response, in which IL-1β may be pivotal. Chronic *H. pylori* infection also leads to chronic atrophic gastritis with resulting achlorhydria, which in turn favors bacterial growth that can convert nitrates (dietary components) to nitrites. These nitrites, in combination with genetic factors, promote abnormal cellular proliferation, genetic mutations, and eventually cancer. In a mouse model of gastric cancer, *H. pylori* infection may play a role in the recruitment of bone marrow–derived stem cells that facilitate gastric carcinogenesis. Animal models can now recapitulate the cardinal features of gastric adenocarcinoma, either through the use of carcinogens or through genetic approaches.

Genetics

It is clear that genetic factors play a role in gastric cancer. For example, blood group A is associated with a higher incidence rate of gastric cancer, even in nonendemic areas. A three-fold increase in gastric cancer has been reported among first-degree relatives of patients with the disease. Furthermore, germline or inherited mutations in the genes for E-cadherin and α-catenin, albeit rare, have been described in diffuse hereditary gastric cancer, which is seen in young patients. In addition, in Lynch syndrome (Chapter 193), patients have associated extracolonic cancers, including gastric cancer. Patients with familial adenomatous polyposis (FAP) have an increased risk of distal (antral) gastric adenocarcinoma.

It now appears that several genetic mechanisms are important in gastric cancer: oncogene activation, tumor suppressor gene inactivation, and DNA microsatellite instability. For example, loss of heterozygosity of the *APC* (adenomatous polyposis coli) gene has been observed in gastric cancers. The p53 tumor suppressor gene product regulates the cell cycle at the G_1-S phase transition and probably also functions in DNA repair and apoptosis (programmed cell death). The *p53* gene is mutated not only in gastric cancer but also in gastric precancerous lesions, a finding suggesting that mutation of the *p53* gene is an early event in gastric carcinogenesis. Microsatellite DNA alterations or instability in dinucleotide repeats occur frequently in sporadic gastric carcinoma. Mutations in genes may accumulate as a result of DNA microsatellite instability.

CLINICAL MANIFESTATIONS

Symptoms and Signs

In its early stages, gastric cancer may often be asymptomatic or may produce only nonspecific symptoms that make early diagnosis difficult. Later symptoms include bloating, dysphagia, epigastric pain, or early satiety. Early satiety or vomiting may suggest partial gastric outlet obstruction, although gastric dysmotility may contribute to the vomiting in patients with nonobstructive cases. Epigastric pain reminiscent of that associated with peptic ulcer (Chapter 139) occurs in about one fourth of patients, but in most patients with gastric cancer, the pain is not relieved by food or antacids. Pain that radiates to the back may indicate that the tumor has penetrated the pancreas. When dysphagia is associated with gastric cancer, this symptom suggests a more proximal gastric tumor at the GE junction or in the fundus.

Signs of gastric cancer include bleeding, which can result in iron deficiency anemia that produces the symptoms of weakness, fatigue, and malaise, as well as (rarely) more serious cardiovascular and cerebrovascular consequences. Perforation related to gastric cancer is unusual. Gastric cancer metastatic to the liver can lead to right upper quadrant pain, jaundice, and fever. Lung metastases can cause cough, hiccups, and hemoptysis. Peritoneal carcinomatosis can lead to malignant ascites unresponsive to diuretics. Gastric cancer can also metastasize to bone.

Physical Examination

In the earliest stages of gastric cancer, the physical examination may be unremarkable. At later stages, patients become cachectic, and an epigastric mass may be palpated. If the tumor has metastasized to the liver, hepatomegaly with jaundice and ascites may be present. Portal or splenic vein invasion and thrombosis can cause splenomegaly. Lymph node involvement in the left supraclavicular area is termed *Virchow's node*, and periumbilical nodal involvement is called *Sister Mary Joseph's node*. The fecal occult blood test may be positive. Metastasis to the ovary is termed *Krukenberg's tumor*.

Paraneoplastic syndromes may precede or occur concurrently with gastric cancer. Examples include the following: Trousseau syndrome (Chapter 176), which is recurrent migratory superficial thrombophlebitis indicating a possible hypercoagulable state; acanthosis nigricans, which arises as raised and hyperpigmented skin lesions of flexor areas, neck, axilla, groin, and mucosal membranes; neuromyopathy with involvement of the sensory and motor pathways; and central nervous system syndromes with altered mental status and ataxia.

Laboratory Studies

Laboratory studies may reveal iron deficiency anemia. Microangiopathic hemolytic anemia has been reported. Abnormalities in liver tests generally indicate metastatic disease. Hypoalbuminemia is a marker of malnutrition. Protein-losing enteropathy is rare but can be seen in Ménétrier's disease, another predisposing condition. Serologic test results, such as those for CEA and CA72.4, may be abnormal. Although these tests are not recommended for initial diagnosis, they may be useful for monitoring disease after surgical resection.

DIAGNOSIS

The diagnostic accuracy of upper endoscopy with biopsy and cytologic examination approaches 95 to 99% for both types of gastric cancer.[7] Cancer may arise as a small mucosal ulceration, a polyp, or a mass (Fig. 192-1). In some patients, gastric ulceration may first be noted in an upper gastrointestinal barium contrast study. A benign gastric ulcer is associated with a smooth, regular base, whereas a malignant ulcer is associated with a surrounding mass, irregular folds, and an irregular base. Although these and other radiographic characteristics historically helped to predict benign versus malignant disease, upper gastrointestinal endoscopy with biopsy and cytologic examination is mandatory whenever a gastric ulcer is found in the radiologic study, even if the ulcer has benign characteristics.

Endoscopic ultrasonography (EUS) is very helpful in both diagnosis and staging of gastric cancer (Table 192-4). The extent of tumor, including gastric wall invasion and local lymph node involvement, can be assessed by EUS (Fig. 192-2), which provides information complementary to that obtained from CT scans. EUS can help guide aspiration biopsies of lymph nodes to determine their malignant features if any. CT of the chest, abdomen, and pelvis should be performed to document lymphadenopathy and extragastric organ (especially lung and liver) involvement. In some centers, staging of gastric cancer entails bone scans because of the proclivity of gastric cancer to metastasize to bone.

TREATMENT

Rx

Surgical Therapy

The only chance for cure of gastric cancer remains surgical resection, which is possible in 25 to 30% of cases. If the tumor is confined to the distal stomach, subtotal gastrectomy is performed, with resection of lymph nodes in the porta hepatis and the pancreatic head. By contrast, tumors of the proximal stomach merit total gastrectomy to obtain an adequate margin and to remove lymph nodes; distal pancreatectomy and splenectomy are usually also performed as part of this procedure, which carries with it higher mortality and morbidity rates. The addition of para-aorta nodal dissection does not improve survival. Even if a curative procedure is not possible because of metastasis, limited gastric resection may be necessary for patients with excessive bleeding or obstruction. If cancer recurs in the gastric remnant, limited resection may again be necessary for palliation. Most recurrences in both types of gastric cancer are in the local or regional area of the original tumor.

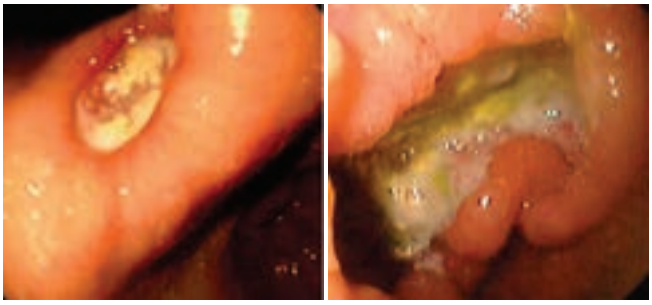

FIGURE 192-1. Benign (*left*) and malignant (*right*) gastric ulcer. Note the shaggy, thickened, and overhanging edges of the cancer. (Courtesy Pankaj Jay Pasricha, MD.)

TABLE 192-4 TNM STAGING OF STOMACH CANCER

PRIMARY TUMOR (T)	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma in situ: intraepithelial tumor without invasion of the lamina propria
T1	Tumor invades lamina propria, muscularis mucosa, or submucosa
T1a	Tumor invades lamina propria or muscularis mucosa
T1b	Tumor invades submucosa
T2	Tumor invades muscularis propria*
T3	Tumor penetrates subserosal connective tissue without invasion of visceral peritoneum or adjacent structures†,‡
T4	Tumor invades serosa (visceral peritoneum) or adjacent structures†,‡
T4a	Tumor invades serosa (visceral peritoneum)
T4b	Tumor invades adjacent structures
REGIONAL LYMPH NODES (N)	
NX	Regional lymph nodes(s) cannot be assessed
N0	No regional lymph nodes metastasis§
N1	Metastasis in 1-2 regional lymph nodes
N2	Metastasis in 3-6 regional lymph nodes
N3	Metastasis in 7 or more regional lymph nodes
N3a	Metastasis in 7-15 regional lymph nodes
N3b	Metastasis in 16 or more regional lymph nodes
DISTANT METASTASIS (M)	
M0	No distant metastasis
M1	Distant metastasis

**Note:* A tumor may penetrate the muscularis propria with extension into the gastrocolic or gastrohepatic ligaments, or into the greater or lesser omentum, without perforation of the visceral peritoneum covering these structures. In this case, the tumor is classified T3. If there is perforation of the visceral peritoneum covering the gastric ligaments or the omentum, the tumor should be classified T4.
†The adjacent structures of the stomach include the spleen, transverse colon, liver, diaphragm, pancreas, abdominal wall, adrenal gland, kidney, small intestine, and retroperitoneum.
‡Intramural extension to the duodenum or esophagus is classified by the depth of the greatest invasion in any of these sites, including the stomach.
§*Note:* A designation of pN0 should be used if all examined lymph nodes are negative, regardless of the total number removed and examined.
From *AJCC Cancer Staging Manual*. 7th ed. New York: Springer-Verlag, 2010.

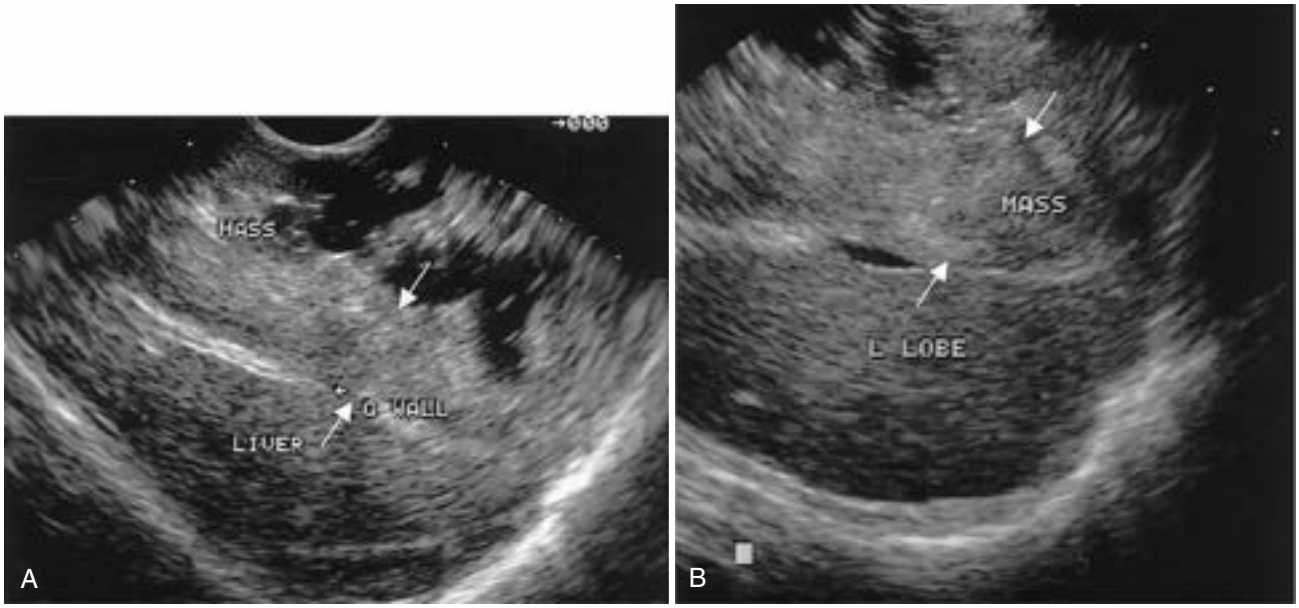

FIGURE 192-2. Gastric mass. Endoscopic ultrasonography depicting a large gastric mass that is compressing the liver and gallbladder wall (A) and, on a different view, the left lobe of the liver (B).

Medical Therapy

Gastric cancer is one of the few gastrointestinal cancers that is somewhat responsive to chemotherapy.[8] In patients with gastric cancer who undergo gastrectomy and extended lymph node dissection with curative intent, S-1 (an oral fluoropyrimidine, 80 mg daily for 4 weeks followed by 2 weeks off, repeated in 6-week cycles for 1 year starting within 6 weeks after surgery) significantly improves 3-year survival from 70 to 80%. Chemotherapy with the combination of epirubicin, cisplatin, and fluorouracil, given both preoperatively and postoperatively, significantly improves 5-year survival from 23 to 36% in patients with resectable GE cancer. Similarly, the combination of chemotherapy (fluorouracil and leucovorin) with radiation therapy has been shown to improve median survival from 27 to 36 months compared with surgery alone in patients with adenocarcinoma of the stomach or GE junction. With a more than 10-year median follow-up, overall survival and relapse-free survival demonstrate continued strong benefit from postoperative radiochemotherapy.

Single-agent chemotherapy treatment, which provides partial response rates of 20 to 30%, is reserved for patients with a poor performance status. Combination regimens that can yield partial response rates of 35 to 50% include the following: ECF, which is most popular in Europe (epirubicin, 50 mg/m^2 on day 1; cisplatin, 60 mg/m^2 on day 1; 5-fluorouracil, 200 mg/m^2/day as a continuous infusion through a central venous access device [CVAD] given throughout treatment, repeated every 21 days for a maximum of 8 cycles); CF (5-fluorouracil infusion, 1000 mg/m^2/day for 4 days; cisplatin, 75 to 100 mg/m^2 on day 1, every 4 weeks); or TCF (docetaxel [Taxotere], 75 mg/m^2 on day 1; cisplatin, 75 mg/m^2 on day 1; and 5-fluorouracil, 750 mg/m^2/day for 5 days, every 3 weeks), or capecitabine plus cisplatin. A randomized trial of triple chemotherapy for advanced esophagogastric cancer showed that oral capecitabine is at least as effective as infused fluorouracil, and that oxaliplatin (which does not require hydration) is at least as effective as cisplatin (which does require hydration) with respect to overall survival. Bevacizumab, a monoclonal antibody targeting vascular endothelial growth factor A (VEGF-A), administered at a dose of 7.5 mg/kg by intravenous infusion every 3 weeks, was found to increase progression-free survival and overall response rate when it was given along with capecitabine-cisplatin chemotherapy as first-line treatment of advanced gastric cancer. Ramucirumab (a VEGF receptor antagonist) given as 8 mg/kg intravenously every two weeks can improve median survival from 3.8 to 5.2 months in advanced gastric cancer. Trastuzumab (8 mg/kg intravenously once, then 6 mg/kg every 3 weeks) can increase survival of HER2-positive advanced gastric or GE junction cancer from 11.1 months to 13.8 months. Radiation therapy alone is ineffective and is generally employed only for palliative purposes in the setting of bleeding, obstruction, or pain. Gene therapy and immune-based therapy are currently only investigational in animal models.

General Methods

Implicit in the management of the patient with gastric cancer is meticulous attention to nutrition (jejunal enteral feedings or total parenteral nutrition), correction of metabolic abnormalities that arise from vomiting or diarrhea, and treatment of infection from aspiration or spontaneous bacterial peritonitis. *H. pylori* eradication treatment reduces the risk for metachronous gastric carcinoma by about two thirds. To maintain lumen patency, endoscopic laser treatment or prosthesis placement can be used in a palliative fashion.

PROGNOSIS

Approximately one third of patients who undergo a curative resection are alive after 5 years. In aggregate, the overall 5-year survival rate in patients with gastric cancer is less than 10%. Prognostic factors include anatomic location and nodal status. Distal gastric cancers without lymph node involvement have a better prognosis than proximal gastric cancers with or without lymph node involvement. Other prognostic factors include depth of penetration and tumor cell DNA aneuploidy. Linitis plastica and infiltrating lesions have a much worse prognosis than polypoid disease or exophytic masses. In the subset of mostly Japanese patients with early gastric cancer that is confined to the mucosa and submucosa, surgical resection may be curative and definitely improves the 5-year survival rate to more than 50%. In fact, when early gastric cancer is confined to the mucosa, endoscopic mucosal resection may be an alternative.

Lymphoma of the Stomach

EPIDEMIOLOGY

Gastric lymphoma represents about 5% of all malignant gastric tumors and is increasing in incidence. Most gastric lymphomas are non-Hodgkin's lymphomas (Chapter 185), and the stomach is the most common extranodal site for non-Hodgkin's lymphomas. Patients with gastric lymphoma are generally younger than those with gastric adenocarcinoma, but the male predominance remains.

CLINICAL MANIFESTATIONS

Patients commonly present with symptoms and signs similar to those of gastric adenocarcinoma. Lymphoma in the stomach can be a primary tumor, or it can be secondary to disseminated lymphoma.

B-cell lymphomas (Chapter 185) of the stomach are most commonly large cell with a high-grade type. Low-grade variants are noted in the setting of chronic gastritis and are termed *mucosa-associated lymphoid tissue* (MALT) lymphomas. MALT lesions are strongly associated with *H. pylori* infection.

DIAGNOSIS

Radiographically, gastric lymphomas usually arise as ulcers or exophytic masses; diffusely infiltrating lymphoma is more suggestive of secondary lymphoma. Thus, upper gastrointestinal barium studies usually show multiple nodules and ulcers for a primary gastric lymphoma and typically have the appearance of linitis plastica (see earlier) with secondary lymphoma. As with gastric adenocarcinoma, however, upper endoscopy with biopsy and cytologic examination are required for diagnosis and have an accuracy of nearly 90%. Apart from conventional histopathologic analysis, immunoperoxidase staining for lymphocyte markers is helpful in diagnosis. As with gastric adenocarcinoma, proper staging of gastric lymphoma involves EUS, chest and abdominal/pelvic CT scans, and bone marrow biopsy as needed.

TREATMENT

Rx

Treatment of gastric diffuse large B-cell lymphoma is best pursued with combination chemotherapy with or without radiation therapy (Chapter 185). For MALT lesions, eradication of *H. pylori* infection with antibiotics should be attempted[9] (Chapter 139), but patients with refractory lesions that are confined to the stomach can sometimes be cured with chemoradiotherapy (Chapter 185).

Other Malignant Tumors of the Stomach

Leiomyosarcoma, which constitutes approximately 1% of all gastric cancers, usually occurs as an intramural mass with central ulceration. Symptoms may include bleeding accompanied by a palpable mass. Leiomyosarcomas are often relatively indolent; surgical resection yields a 5-year survival rate of about 50%. Metastasis can occur to lymph nodes and the liver. Other gastric sarcomas include liposarcomas, fibrosarcomas, myosarcomas, and neurogenic sarcomas.

Most gastrointestinal stromal tumors (GISTs)[10] have been associated with activating mutations in the C-*kit* gene; a subset of such tumors is associated with mutations in the platelet-derived growth factor receptor *(PDGFR)* gene. C-*kit* mutations are also found in chronic and acute myelogenous leukemia (Chapters 183 and 184), and approximately 50% of GISTs respond to imatinib mesylate, which should be continued for at least 3 years. If imatinib is not successful, sunitinib increases survival. If both fail, regorafenib, a multikinase inhibitor, can be tried.[11] Carcinoid tumors (Chapter 232) may begin in the stomach and are curable by removal if they have not yet spread to the liver.

Primary tumors can also spread to the stomach. In addition to lymphomas, other tumors found in the stomach include primary lung (Chapter 191) and breast (Chapter 198) cancers as well as malignant melanoma (Chapter 203).

Leiomyomas and Benign Tumors

Leiomyomas, which are smooth muscle tumors of benign origin, occur with equal frequency in men and women and are typically located in the middle and distal stomach. Leiomyomas can grow into the lumen, with secondary ulceration and consequent bleeding. Alternatively, they can expand to the serosa with extrinsic compression. Endoscopy may reveal a mass that has overlying mucosa or mucosa replaced by ulceration. On upper gastrointestinal series, leiomyomas are usually smooth with an intramural filling defect, with or without central ulceration. However, benign leiomyomas can be difficult to distinguish from their malignant counterparts radiographically or endoscopically; tissue diagnosis is imperative. Symptomatic leiomyomas should be removed, but those without associated symptoms do not require therapy.

Other benign gastric tumors include lipoma, neurofibroma, lymphangioma, ganglioneuroma, and hamartoma, the last associated with Peutz-Jeghers syndrome (Chapter 193) or juvenile polyposis (when restricted to the stomach).

Adenomas

Gastric adenomas and hyperplastic polyps are unusual but may be found in middle-aged and elderly patients. Polyps may be sessile or pedunculated. Although isolated gastric adenomatous polyps are generally asymptomatic, some patients may have dyspepsia, nausea, or bleeding. Gastric adenomas and hyperplastic polyps are smooth and regular on upper gastrointestinal series, but the diagnosis must be confirmed by upper endoscopy with biopsy. Pedunculated polyps that are larger than 2 cm or that have associated symptoms should be removed by endoscopic snare cautery polypectomy, whereas large sessile gastric adenomatous polyps may merit segmental surgical resection. If polyps progress to an intermediary stage of severe dysplasia or culminate in cancer, treatment is the same as for gastric adenocarcinoma.

Grade A References

A1. van Hagen P, Hulshof MC, van Lanschot JJ, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med*. 2012;366:2074-2084.

A2. Phoa KN, van Vilsteren FG, Weusten BL, et al. Radiofrequency ablation vs endoscopic surveillance for patients with Barrett esophagus and low-grade dysplasia: a randomized clinical trial. *JAMA*. 2014;311:1209-1217.

A3. Smalley SR, Benedetti JK, Haller DG, et al. Updated analysis of SWOG-Directed Intergroup Study 0116: a phase III trial of adjuvant radiochemotherapy versus observation after curative gastric cancer resection. *J Clin Oncol*. 2012;30:2327-2333.

A4. Cunningham D, Starling N, Rao S, et al. Capecitabine and oxaliplatin for advanced esophagogastric cancer. *N Engl J Med*. 2008;358:36-46.

A5. Ohtsu A, Shah MA, Van Custem E, et al. Bevacizumab in combination with chemotherapy as first-line therapy in advanced gastric cancer: a randomized, double-blind, placebo-controlled phase III study. *J Clin Oncol*. 2011;29:3968-3976.

A6. Fuchs CS, Tomasek J, Yong CJ, et al. Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD): an international, randomised, multicentre, placebo-controlled, phase 3 trial. *Lancet*. 2014;383:31-39.

A7. Joensuu H, Eriksson M, Sundby Hall K, et al. One vs three years of adjuvant imatinib for operable gastrointestinal stromal tumor: a randomized trial. *JAMA*. 2012;307:1265-1272.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Zhang Y. Epidemiology of esophageal cancer. *World J Gastroenterol*. 2013;19:5598-5606.
2. Varghese TK Jr, Hofstetter WL, Rizk NP, et al. The society of thoracic surgeons guidelines on the diagnosis and staging of patients with esophageal cancer. *Ann Thorac Surg*. 2013;96:346-356.
3. Mallipeddi MK, Onaitis MW. The contemporary role of minimally invasive esophagectomy in esophageal cancer. *Curr Oncol Rep*. 2014;16:374.
4. Levine DM, Ek WE, Zhang R, et al. A genome-wide association study identifies new susceptibility loci for esophageal adenocarcinoma and Barrett's esophagus. *Nat Genet*. 2013;45:1487-1493.
5. Dai Y, Li C, Xie Y, et al. Interventions for dysphagia in oesophageal cancer. *Cochrane Database Sys Rev*. 2014;10:CD005048.
6. De Martel C, Forman D, Plummer M. Gastric cancer: epidemiology and risk factors. *Gastroenterol Clin North Am*. 2013;42:219-240.
7. Thrumurthy SG, Chaudry MA, Hochhauser D, et al. The diagnosis and management of gastric cancer. *BMJ*. 2013;347:f6367.
8. Kasper S, Schuler M. Targeted therapies in gastroesophageal cancer. *Eur J Cancer*. 2014;in press.
9. Fischbach W. MALT lymphoma: forget surgery? *Dig Dis*. 2013;31:38-42.
10. Joensuu H, Hohenberger P, Corless CL, et al. Gastrointestinal stromal tumour. *Lancet*. 2013;382:973-983.
11. Demetri GD, Reichardt P, Kang YK, et al. Efficacy and safety of regorafenib for advanced gastrointestinal stromal tumours after failure of imatinib and sunitinib (GRID): an international, multicentre, randomised, placebo-controlled, phase 3 trial. *Lancet*. 2013;381:295-302.

REVIEW QUESTIONS

1. What is a leading risk factor for the development of Barrett's esophagus and esophageal adenocarcinoma?

A. Obesity
B. Nutritional deficiencies
C. Childhood gastroesophageal reflux disease
D. High-fiber diet

Answer: A Obesity, specifically central adiposity, has been associated with increased risk of Barrett's esophagus and esophageal adenocarcinoma in separate studies. Mechanistically, there is secretion of cytokines and chemokines from adipocytes that likely produce pro-inflammatory states in the esophagus in both conditions.

2. What is a key risk factor for the development of gastric adenocarcinoma?

A. Obesity
B. *Helicobacter pylori* infection
C. History of pancreatic cancer
D. Blood group O

Answer: B *Helicobacter pylori*, a common organism worldwide that affects nearly 2 billion individuals, causes gastric atrophy with achlorhydria and subsequent intestinal metaplasia in the gastric epithelium. This can evolve to dysplasia (low grade and high grade) and adenocarcinoma, so-called intestinal type. Note that only a subset of chronically *H. pylori*–infected patients progress to gastric adenocarcinoma, suggesting other cofactors are important.

3. Therapy of esophageal or gastric cancer may involve all the following modalities *except*:

A. Surgery
B. Chemoradiation therapy
C. Palliative chemotherapy
D. Bone marrow transplantation

Answer: D Bone marrow transplantation is not an option. The modalities in A, B, and C are all used and dictated by the stage of esophageal or gastric cancer.

4. What hereditary condition predisposes to gastric adenocarcinoma?

A. Germline *E-cadherin* mutation
B. Germline *TP53* mutation
C. Germline *p16INK4a* mutation
D. Germline *p120* catenin mutation

Answer: A Germline E-cadherin mutation predisposes to early-age onset of diffuse hereditary gastric adenocarcinoma and may also be associated with lobular-type breast cancer.

5. What is a likely treatment option for Barrett's esophagus with high-grade dysplasia and no surrounding esophageal adenocarcinoma?

A. Neoadjuvant chemoradiation therapy
B. Laser therapy
C. Endoscopic mucosal resection (EMR)
D. Endoscopic prosthetic stent

Answer: C Endoscopic mucosal resection (EMR) is a viable option for high-grade dysplasia, as well as intramucosal esophageal adenocarcinoma. Radiofrequency ablation via endoscopy is used for low-grade dysplasia Barrett's esophagus and may be used for high-grade dysplasia Barrett's esophagus under certain circumstances.

193 NEOPLASMS OF THE SMALL AND LARGE INTESTINE

CHARLES D. BLANKE AND DOUGLAS O. FAIGEL

NEOPLASMS OF THE SMALL INTESTINE

EPIDEMIOLOGY

The small intestine accounts for the majority of the length (≈75%) and absorptive surface (≈90%) of the gastrointestinal (GI) tract. Nonetheless, it is a rare site for the development of neoplastic disease, because only 1 to 2% of primary GI tumors originate in the duodenum, jejunum, or ileum. In fact, half of all small bowel neoplasms represent metastatic disease from other sites, particularly elsewhere in the GI tract. Although small bowel malignancies constitute less than 0.5% of all cancers, the incidence of small bowel tumors (especially carcinoids) has increased dramatically over the last several decades, which may possibly be a reflection of better diagnostic techniques. Overall, the mean age at diagnosis of a small bowel tumor is about 67 years (younger for those with sarcomas and lymphomas); neoplasms are more common in males, and they occur more frequently in African Americans than whites.

PATHOBIOLOGY

Adenocarcinomas, arising from mucosal glands, were formerly the most frequent primary small bowel tumor. They now constitute 25 to 33% of small bowel neoplasms, including benign growths, and 40% of all malignant tumors. Adenocarcinomas most commonly arise in the duodenum (65% of all small intestinal adenocarcinomas), even though the duodenum represents only a tiny fraction of the length of the small bowel. They occur less commonly in the jejunum and least commonly in the ileum. The majority are well or moderately differentiated.

Small bowel carcinoids, which derive from enterochromaffin cells in the crypts of Lieberkühn, are now the most common small bowel tumors, accounting for up to 44% of malignancies. They tend to be quite well differentiated. In contrast to glandular tumors, small intestinal carcinoids tend to arise in the distal ileum, and up to 30% are multifocal. Other small bowel neuroendocrine tumors are occasionally seen, including biochemically active neoplasms such as gastrinomas and somatostatinomas. Very rarely, high-grade true small-cell carcinomas occur.

Malignant connective tissue tumors account for 10 to 17% of small bowel neoplasms. Gastrointestinal stromal tumors (GISTs), which derive from the interstitial cells of Cajal or a common precursor, account for approximately 85% of these neoplasms (Fig. 193-1). GISTs, like adenocarcinomas, disproportionately arise in the duodenum, and the small bowel itself is the second most common primary site for these mesenchymal neoplasms (33% derive from small bowel). Morphologically, GISTs often resemble leiomyosarcomas (Fig. 193-2), but they can be differentiated by the expression of the Kit protein (CD117). Other small bowel sarcomas, such as true leiomyosarcomas, are seen more rarely (Chapter 202).

Primary GI lymphomas are the most common extranodal lymphomatous variation, and the small intestine is the second most common GI site for such tumors (Chapter 185). Lymphomas account for approximately 8% of small bowel neoplasms. The ileum, rich in submucosal lymphoid follicles, is the most common small bowel site. Tumors may be low or higher grade and can arise from precursor B or T lymphocytes. The overwhelming majority are non-Hodgkin tumors (Fig. 193-3). Lymphomas involving the small bowel may also be a manifestation of true systemic disease.

Malignant melanoma (Chapter 203) can develop as a primary mucosal small bowel tumor, probably arising from schwannian neuroblasts associated with GI innervation. In addition, the small bowel is the most common GI site for melanoma metastases.

Finally, a variety of common benign tumors may originate in the small bowel, including adenomas, leiomyomas, and lipomas. Desmoids (most often seen in patients with familial adenomatous polyposis [FAP]), hamartomas, and hemangiomas are relatively rare. Benign growths are more common in the distal small intestine.

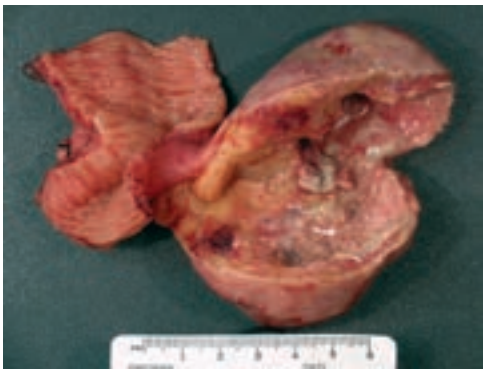

FIGURE 193-1. Small bowel gastrointestinal stromal tumor with ulceration.

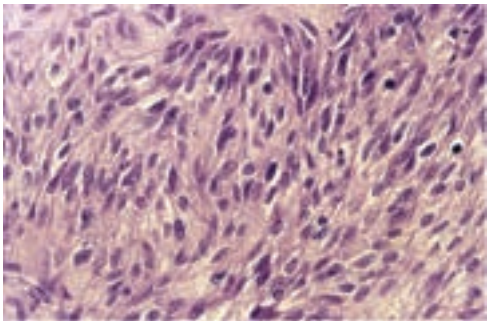

FIGURE 193-2. Histopathology of gastrointestinal stromal tumor.

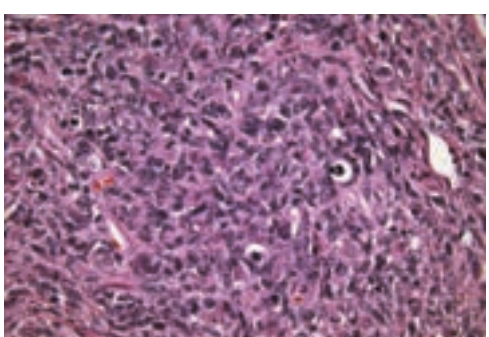

FIGURE 193-3. Histopathology of small bowel non-Hodgkin lymphoma. (Courtesy Dr. M.K. Washington.)

The small intestine can be involved with advanced cancers from other sites through direct invasion, extension of peritoneal metastases, or hematogenous spread. As noted, the small bowel is the most common GI site for melanoma metastases; involvement is also fairly common with ovarian, breast, lung, and other GI neoplasms.

Predisposing Conditions

Inflammatory bowel disease (Chapter 141) and some environmental factors (e.g., salt-cured foods, alcohol) predispose to small bowel adenocarcinomas. Data regarding tobacco exposure and obesity are conflicting. Additionally, many of the polyposis syndromes are associated with small bowel neoplasms. Most notably, FAP (see later) is associated with adenomas and carcinomas of the duodenum and jejunum, but especially in the ampullary and periampullary region. At least 90% of FAP patients develop duodenal adenomas, and up to 10% develop cancer. The risk of cancer is related to the number of polyps, their size and histologic type, and the presence of high-grade dysplasia. Patients with FAP should undergo regular screening for duodenal neoplasia with both forward- and side-viewing endoscopes beginning around the time of colectomy and repeated at 1- to 5-year intervals, depending on the presence and degree of duodenal polyposis. Patients with *MUTYH*-associated polyposis likewise develop duodenal neoplasia and should undergo screening. Patients with hereditary nonpolyposis colon cancer (HNPCC) are also at increased risk for small intestine adenocarcinoma, which may be the first manifestation of their disease. HNPCC-associated small bowel cancer may present at a young age (median, 39 years) and occurs with decreasing frequency from the duodenum to the ileum, with about 50% of occurrences in the duodenum. Screening may be considered beginning at age 30. Patients with sprue are at increased risk for small bowel lymphomas, and they have an almost 35-fold increased risk for developing adenocarcinomas.

CLINICAL MANIFESTATIONS

The most common presenting symptom of small bowel tumors is abdominal pain, especially for those that are true cancers. Weight loss, nausea, GI bleeding, and symptoms related to perforation are less common. Approximately 25% of patients have GI obstruction, and duodenal periampullary tumors can lead to obstructive jaundice. Most malignancies are symptomatic, whereas benign tumors may be asymptomatic in up to half of patients. Carcinoid tumors in the small bowel are often asymptomatic, although in the setting of advanced disease, they may secrete bioactive amines, leading to flushing, diarrhea, wheezing, and eventually symptoms of right heart failure (related to valvular fibrosis) (Chapter 232). This occurs more commonly with tumors originating in the jejunum and ileum. Benign tumors tend to be found incidentally, although intraluminal growth may eventually cause symptoms of obstruction, and some may grow large enough to ulcerate and bleed.

DIAGNOSIS

The physical examination in patients with small bowel tumors is often unremarkable, although a palpable mass and, in more advanced cases, ascites may be present. As discussed earlier, patients with periampullary neoplasms may be jaundiced and/or icteric. Obstructive signs such as hyperperistalsis may be present, and patients with lymphoma may also have splenomegaly or other signs of systemic involvement, such as lymphadenopathy. Laboratory findings may include iron deficiency anemia or increased hepatic enzymes (the latter is especially common in those with liver metastases or biliary obstruction). Serum levels of the carcinoembryonic antigen (CEA) tumor marker may be elevated in small bowel adenocarcinoma, especially in advanced cases, but it is neither sensitive nor specific enough for routine diagnostic use. Patients with neuroendocrine tumors may demonstrate elevated levels of serotonin, chromogranin A, tumor-specific bioactive amines (e.g., gastrin), or urinary 5-hydroxyindoleacetic acid. Variants of intestinal lymphomas may show heavy chain immunoglobulin A fragments in serum and urine.

Proper imaging is crucial for both diagnosis and staging of small bowel neoplasms, but no one method is clearly the best. Standard radiographic techniques of value include upper GI with small bowel follow-through (helpful in demonstrating both masses and potential mucosal defects), angiography (which may show a site of bleeding or tumor blush with specific neoplasms), and computed tomography (CT) or magnetic resonance imaging (MRI) enteroclysis (double-contrast studies are both sensitive and specific for small bowel masses). Transabdominal ultrasound and standard CT may indicate a primary mass as well as metastases; MRI appears to be superior to CT in detecting and characterizing liver metastases.

Primary neuroendocrine tumors and their metastases are often apparent on indium-111 octreotide scanning. A wide variety of histologies may have uptake on positron emission tomography (PET) scanning. PET does not yet have a well-defined role in the diagnosis of most small bowel malignancies, although PET scans are useful in those with GISTs, to monitor response to systemic therapy (see later). Plain films rarely are specific enough to lead to a diagnosis, but they may demonstrate intestinal obstruction.

Capsule endoscopy uses a wireless endoscopic device that allows minimally invasive imaging of the small intestine. The system consists of the capsule camera—a swallowable, self-contained, battery-operated device that transmits two images per second—a receiver worn on the patient's belt, and a computerized work station for downloading and viewing the images. The primary indications for capsule endoscopy are the evaluation of obscure GI bleeding and Crohn disease. Tumors are found in about 2 to 3% of patients undergoing capsule endoscopy for obscure GI bleeding, and they may be more common in younger patients. Tumors detected include lymphoma, adenocarcinoma, metastatic disease, carcinoid tumor, and GIST. Capsule endoscopy has also been used to assess the small intestines of patients with FAP and Peutz-Jeghers syndrome (PJS), although its clinical utility for routine screening of these patients has not been established.

Deep enteroscopy describes a group of related techniques to facilitate deep intubation of the small bowel with long endoscopes. The small bowel may be examined antegrade (through the mouth) or retrograde (through the colon), allowing the majority of the small bowel to be examined. Although more invasive than capsule endoscopy, deep enteroscopy has a similar diagnostic yield but allows for tissue sampling (biopsy) and therapy including polypectomy and control of bleeding.

Surveillance

Patients with FAP, *MUTYH* polyposis, PJS, and probably HNPCC require regular surveillance of the small intestine, with specific recommendations as noted earlier. Patients with sporadic small intestine adenomas and possibly carcinoid tumors should undergo colonoscopy because they are at increased risk for colonic neoplasia. No specific guidelines exist for following patients with resected small bowel adenocarcinomas.

Staging

The staging systems for small bowel tumors vary by histology. Adenocarcinomas and, more recently, neuroendocrine cancers and GISTs are staged using different classifications within the American Joint Committee on Cancer's TNM malignant tumors system. Intestinal non-Hodgkin lymphomas, whether primary or part of a systemic process, are staged in accordance with a modified Ann Arbor system originally used in Hodgkin disease (Chapter 185).

TREATMENT

Rx

In general, surgical excision is the treatment of choice for most localized small bowel tumors. The extent of excision necessary depends on the tumor's location and histology. Adenocarcinomas involving the first and second portions of the duodenum require pancreaticoduodenectomy, whereas those in the more distal small intestine may be treated with segmental or wide local resection, including regional lymph nodes. Low-grade neuroendocrine tumors

should be managed with en bloc resection, again including regional nodes. GISTs, which very rarely spread to regional nodes, may be treated with excision without lymphadenectomy (except in cases with gross involvement of nodes). Primary surgery for lymphomas may be offered for low-stage malignancies and may also be required for complications of disease (e.g., intussusception). Local management of benign small bowel tumors varies from observation (incidentally discovered lipomas) to endoscopic polypectomy (small adenomas) to pancreaticoduodenectomy (periampullary villous adenomas) (Video 193-2).

The need for and types of adjuvant therapy vary as well. Fully resected benign tumors require no additional therapy. Adenocarcinomas are often treated according to the principles developed for colorectal cancer, with some experts advocating fluoropyrimidine-based systemic chemotherapy, at least for patients with nodal involvement. Older retrospective studies demonstrated no benefit from adjuvant systemic therapy, but its use has still increased almost three-fold over the last two decades. Chemoradiotherapy has occasionally been recommended for those with more locally advanced duodenal adenocarcinomas. So far, no randomized trials have proven either strategy offers any advantage. Fully excised well- to moderately differentiated neuroendocrine cancers do not require adjuvant therapy. Postoperative imatinib mesylate treatment clearly postpones the recurrence of high-risk GISTs (Grade A)■; a survival benefit has been demonstrated for those with gastric and nongastric primaries. Questions remain regarding dose and ideal duration of therapy. There is no defined role for the postoperative treatment of other mesenchymal tumors. Lymphomas treated with excision alone have high rates of recurrence, and systemic chemotherapy is advocated for high-grade variants; some experts also recommend chemotherapy for low-grade subtypes.

Patients with advanced small bowel adenocarcinomas are often treated with systemic chemotherapy regimens known to be effective against cancers of similar histology originating in the colon. However, the data supporting any specific regimen are scant and are not derived from randomized trials. Approximately 90% of patients with incurable GISTs have durable disease control when treated with the tyrosine kinase inhibitor imatinib mesylate, and the median survival for patients with metastatic disease has recently improved from approximately 18 months to 5 years.

Patients with small bowel lymphomas may be treated with chemotherapy that is effective in similar tumors originating outside the GI tract.

PROGNOSIS

Patients with small bowel adenocarcinomas generally do worse than those with similarly staged colonic glandular tumors. Also, patients with duodenal primaries may have poorer outcomes than those with more distal small bowel cancers. In general, 5-year survival rates range from 4% for those with metastases to 80% for those with very early disease confined to the small bowel wall. Five-year survival in patients with small bowel carcinoids exceeds 50%. In the pre-imatinib era, patients with surgically resected small bowel GISTs had recurrence rates ranging to 90% or higher, depending on the tumor's size, precise location, and mitotic rates; those with more distal tumors had a worse prognosis than those with duodenal primaries. Patients with recurrences almost invariably died within 2 years because salvage surgery and systemic chemotherapy were ineffective. True life expectancy in the era of postoperative imatinib use is unknown, but the median survival of patients with advanced GISTs likely exceeds 5 years. Small intestinal lymphomas have 5-year survival rates surpassing 60%, although this is highly variable and depends on the histologic subtype.

NEOPLASMS OF THE LARGE INTESTINE

Colorectal cancer (CRC) is the third most common cancer in the United States. Disease that has spread beyond regional lymph nodes is, for the most part, incurable, and CRC in general remains the second leading cause of neoplastic death. The lifetime risk of developing CRC for the average individual is about 1 in 18 to 20.

EPIDEMIOLOGY

Almost three quarters of large bowel cancers arise proximally (i.e., are of colonic origin). Although CRC is primarily a disease of the elderly (median age ≈ 73 years), about 10% of cases occur in those aged 50 or younger. CRC incidence and mortality have decreased overall recently, although the incidence has been increasing in the young. Incidence rates for right-sided cancers have also been decreasing, possibly but not solely owing to effective distal large bowel screening with flexible sigmoidoscopy. CRC is slightly more common in men than in women and in African Americans than in whites. Men develop CRC an average of 5 to 10 years earlier than women; similarly, large bowel cancers seem to arise an average of 5 to 10 years earlier in African Americans than in whites. Significant geographic variation in incidence occurs, probably based more on environmental factors than on genetic ones, as suggested by migration studies.

PATHOBIOLOGY

Between 96 and 98% of CRCs are adenocarcinomas. Rarely seen histologies include neuroendocrine cancers, epidermoid carcinomas, lymphomas, and sarcomas (including GISTs). Composites, particularly adenocarcinomas with neuroendocrine differentiation, are frequently encountered.

Adenocarcinomas derive from colonic columnar glandular epithelium in the colorectal mucosal lining. They are equally common in males and females and are most frequently reported in the sigmoid colon. Adenocarcinomas most commonly present at a localized or regional (nodal) stage. About two thirds are of moderate grade. The majority are nonmucinous, although the mucinous phenotype constitutes up to one fifth of all CRCs. Another variant is the true signet ring carcinoma, identified by large quantities of single tumor cells with nuclear displacement by intracytoplasmic mucin. Data are controversial as to whether mucinous tumors have a worse prognosis, whereas signet ring histology is clearly associated with advanced disease and/or worse outcome.

Neuroendocrine cancers can have a variety of histologies ranging from bland, well-differentiated carcinoids to high-grade small-cell carcinomas. True carcinoids are the second most common histologic colorectal subtype. They are more common in nonwhites and make up the vast majority of non-adenocarcinoma epithelial cancers. Distal bowel carcinoids are hormonally inactive. Noncarcinoid neuroendocrine cancers tend to be high grade and commonly present with hepatic and other distant metastases.

Epidermoid carcinomas are rare overall but still account for up to one fourth of CRCs. Most are squamous cell subtypes. They are more common in women and Hispanic patients. Epidermoid carcinomas are located in the rectum more than 90% of the time, and they are usually moderately or poorly differentiated. Interestingly, they commonly present as localized cancers, regardless of their degree of differentiation.

Medullary carcinomas, which are more often right-sided and seen in older female patients, tend to have a lower incidence of lymph node involvement. The majority exhibit microsatellite instability, and they tend to have a relatively better prognosis.

Primary colorectal lymphomas are fairly rare, constituting 10 to 20% of all GI lymphomas but less than 1% of CRCs. They are much more common in males and in the elderly. The cecum is the most common site of origin. Tumors are usually of B-cell origin.

Sarcomas of the large bowel have no gender or racial predilections. More than 50% have been classified as leiomyosarcomas, and they are most commonly found in the rectum. Sarcomas are usually diagnosed at a localized stage, regardless of grade, although about 40% are actually poorly differentiated. Kaposi sarcomas and GISTs are other sarcoma histologies found in the large bowel; many of the distal tumors called "leiomyosarcomas" in older registries were likely true GISTs.

Predisposing Conditions

Predisposing conditions for colonic neoplasia include age (discussed previously), gender, race, inflammatory bowel disease, family history,[1] and defined inherited syndromes. Defined genetic cancer syndromes, however, account for only a small percentage of CRCs (see later). Patients with first-degree relatives who also have colon neoplasia (adenomas or carcinomas) are commonly seen. Individuals with a first-degree relative with CRC face a 2- to 3-fold increased risk for malignancy, and this risk rises to 5- or 6-fold if two first-degree relatives are affected. Patients whose relatives have adenomas face a 1.8-fold increased risk for CRC, and this rises to 2.6 if the relative is younger than 60 years.

Patients with ulcerative colitis and Crohn disease (Chapter 141) are at increased risk for CRC in proportion to the amount of bowel involved and the duration of illness.[2] For example, adenocarcinoma of the colon is 10 to 20 times more common in persons with ulcerative colitis than in the general population. Between 2 and 4% of all patients with long-term ulcerative colitis develop this malignancy, and the cumulative incidence over a 25-year period is approximately 12%. Overall, the incidence of colorectal adenocarcinoma is 60% higher in persons with inflammatory bowel disease than in the general population and has been stable over time. Patients with both ulcerative colitis and primary sclerosing cholangitis seem to be at even greater risk. For those with Crohn colitis, patients with extensive disease involving more than one third of the colon are at increased risk (six- to eight-fold), similar to those

VIDEO 193-2. **Laparoscopic-assisted double balloon enteroscopy with polypectomy of a jejunal adenoma, followed by surgical oversew of the polypectomy site.**

TABLE 193-1 GENERAL FEATURES OF INHERITED COLORECTAL CANCER SYNDROMES

SYNDROME	POLYP HISTOLOGY	POLYP DISTRIBUTION	AGE OF ONSET	RISK OF COLON CANCER	GENETIC LESION	CLINICAL MANIFESTATIONS	ASSOCIATED LESIONS
Familial adenomatous polyposis	Adenoma	Large intestine, duodenum	16 yr (range, 8-34 yr)	100%	5q (*APC* gene)	Rectal bleeding, abdominal pain, bowel obstruction	Desmoids, CHRPE
Peutz-Jeghers syndrome	Hamartoma	Large and small intestine	First decade	Slightly above average	19p *(STK11)*	Possible rectal bleeding, abdominal pain, intussusception	Orocutaneous melanin pigment spots, other tumors
MUTYH-associated polyposis	Adenoma	Large intestine, duodenum	45-50 yr (range, 13-60 yr)	75% (range, 50-100%)	1p (*MUTYH* gene)	Rectal bleeding, abdominal pain, bowel obstruction	CHRPE, osteomas
Juvenile polyposis	Hamartoma (rarely adenoma)	Large and small intestine	First decade	≈9%	*PTEN*, *SMAD4*, *BMPR1*	Possible rectal bleeding, abdominal pain, intussusception	Pulmonary AVMs
Hereditary nonpolyposis colon cancer	Adenoma	Large intestine	40 yr (range, 18-65 yr)	30%	Mismatch repair genes*	Rectal bleeding, abdominal pain, bowel obstruction	Other tumors (e.g., ovary, uterus, pancreas, stomach)

*Including *hMSH2*, *hMSH3*, *hMSH6*, *hMLH1*, *hPMS1*, and *hPMS2*.
AVM = arteriovenous malformation; CHRPE = congenital hypertrophy of the retinal pigment epithelium.

with ulcerative colitis. Isolated proctitis is not a risk factor. Patients with ulcerative colitis or extensive Crohn disease should undergo screening colonoscopy every 1 to 2 years beginning 8 to 10 years after disease onset. At colonoscopy, a dye spray may be used to identify suspicious areas, and multiple biopsies (at least 32 for pan-colitis) are obtained with targeting of suspicious lesions. The purpose of this is to identify the presence of dysplasia. The presence of high-grade dysplasia, any dysplasia in a mass or lesion that cannot be excised endoscopically, or multifocal low-grade dysplasia should prompt colectomy.

Patients who have had ureterocolostomy and those with acromegaly are also at increased risk. Case-control studies suggest that obesity, low physical activity, smoking, excessive alcohol, high-fat diet, and lack of dietary fiber increase the CRC risk. Patients with *Streptococcus bovis* bacteremia or endocarditis have increased rates of CRC and should undergo colonoscopy.

Polyposis Syndromes

Several defined dominant and recessive genetic conditions have been identified that convey an increased risk of CRC (Table 193-1). These include FAP, HNPCC, *MUTYH*-associated polyposis, PJS, juvenile polyposis, *PTEN* hamartoma syndrome, and Cronkhite-Canada syndrome.

FAMILIAL ADENOMATOUS POLYPOSIS

FAP is an autosomal dominant condition characterized by the development of hundreds to thousands of adenomatous polyps and CRC by age 40 (Fig. 193-4). Estimates of disease prevalence are 1 in 8000 to 15,000 births.

PATHOBIOLOGY

FAP is inherited as an autosomal dominant disease with incomplete penetrance. It has been mapped to the adenomatous polyposis coli *(APC)* gene located on the long arm of chromosome 5 (5q21). *APC* is a tumor suppressor gene that is a critical regulator of intestinal epithelial cell growth. Inherited mutations generally result in a truncated gene product. Patients with the familial syndrome inherit one mutant copy of *APC;* when a loss-of-function mutation develops in the other *APC* allele, mucosal epithelial cell growth is no longer controlled normally, and polyps develop. Variable phenotypes can be partly attributed to differences in the location of the *APC* mutation, with attenuated FAP being seen in mutations at the 5′ and 3′ ends of the gene.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Adenomas begin to appear early in the second decade of life, and GI symptoms begin to appear in the third or fourth decade. Polyps are distributed relatively evenly throughout the colon, although a slight predominance has been noted in the distal colon. Almost all patients with FAP develop frank colorectal carcinoma by age 40 years if the condition is left untreated. Gastric polyps (mostly nonadenomatous) occur in 30 to 100% of patients, and duodenal adenomas are found in 45 to 90%. Periampullary duodenal cancer develops in approximately 10% of cases. Small bowel lesions that are distal to the duodenum rarely progress to malignancy. In attenuated FAP, fewer than 100 colonic adenomas develop, there is a right colon predominance, and cancer develops approximately 10 years later. Genetic testing may identify the mutation in up to 85% of affected individuals and is useful for family screening.

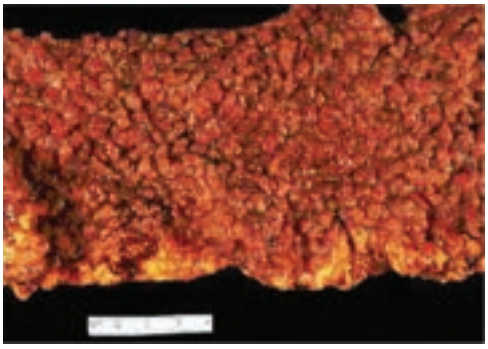

FIGURE 193-4. Resected colon lined with hundreds of adenomatous polyps in a patient with familial adenomatous polyposis.

TREATMENT Rx

The primary treatment option in FAP patients is total proctocolectomy with conventional ileostomy or ileoanal (pouch) anastomosis. Individuals with *APC* mutations and those with no identified mutation but clinical FAP in their families should be screened with annual sigmoidoscopy beginning at age 10 to 12 years. In families with known *APC* mutations, individuals who test negative do not require heightened surveillance but should undergo routine risk screening. FAP patients should be screened for duodenal polyposis with upper endoscopy beginning at age 20, with subsequent surveillance depending on polyp burden and histology. Cyclooxygenase-2 inhibition with sulindac or celecoxib may be considered in patients with small bowel adenomas or adenomas in the remnant rectum. Eicosapentaenoic acid supplementation has also been shown to decrease polyps in the remnant rectum.

GARDNER SYNDROME

Gardner syndrome is a phenotypic subtype of FAP that is also caused by mutations in the *APC* gene. It is distinguished by the presence of extraintestinal manifestations, including osteomas (particularly mandibular), soft tissue tumors (including lipomas, sebaceous cysts, and fibrosarcomas), supernumerary teeth, desmoid tumors, mesenteric fibromatosis, and congenital hypertrophy of the retinal pigment epithelium. The phenotypic differences between Gardner syndrome and FAP appear to result from variations in the location of the *APC* mutation, the presence of modifying genes, and environmental factors. Adenomatous polyps in Gardner syndrome have the

same malignant potential as those found in FAP, and CRC screening and treatment recommendations are the same.

TURCOT SYNDROME

A hallmark of Turcot syndrome is the combination of colorectal polyposis and malignant diseases of the central nervous system. Mutations in the *APC* gene account for two thirds of cases, and the remaining one third result from mutations in the DNA mismatch repair genes that are also mutated in HNPCC. Central nervous system manifestations include medulloblastomas, glioblastomas, and ependymomas.

HEREDITARY NONPOLYPOSIS COLON CANCER

HNPCC, also known as Lynch syndrome, is the most common hereditary CRC syndrome and accounts for approximately 2% of all cases of CRC. It is an autosomal dominant trait and is highly penetrant. Clinically, HNPCC has been defined by the presence of all three of the following: (1) three or more relatives with histologically verified HNPCC-associated cancer (CRC or cancer of the endometrium, small bowel, ureter, or renal pelvis), one of whom is a first-degree relative of the other two, in the absence of FAP; (2) CRC involving at least two generations; and (3) one or more family members with cancer diagnosed before age 50.

PATHOBIOLOGY

HNPCC is caused by loss-of-function germline mutations in a set of genes involved in the repair of DNA base pair mismatches that occur during DNA replication (also known as the mutation mismatch repair system).

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The median age for diagnosis of HNPCC is the mid-40s. Although several adenomas may be present, the diffuse polyposis characteristic of FAP is not found. Colonic neoplasia has a right-sided predominance (proximal to the splenic flexure). Although the cancers tend to be poorly differentiated, they generally have a better prognosis than similar sporadic CRCs. Synchronous and metachronous CRC is common. Patients with HNPCC are also at high risk for other malignant diseases, especially endometrial carcinoma, as well as cancers of the ovary, stomach, small bowel, hepatobiliary tract, ureter, and pancreas.[3] The Muir-Torre syndrome variant is associated with cutaneous lesions and visceral malignancies. Screening for HNPCC may begin with testing of the tumor for microsatellite instability, performing immunohistochemical staining for products of mismatch repair genes (including *hMSH2*, *hMSH6*, *hMLH1*, and *hPMS2*). A positive screen does not definitely indicate HNPCC, because up to 15% of sporadic tumors may have these features; this should be followed with germline testing.

TREATMENT Rx

Persons potentially affected with HNPCC should undergo a colonoscopy every 2 years beginning at age 21 and annually beginning at age 40. Patients with CRC or large adenomas should undergo subtotal colectomy. Women in HNPCC-affected families should have pelvic examinations every 1 to 3 years beginning at age 18; annual pelvic examinations, transvaginal ultrasonography, and endometrial biopsy have been recommended beginning at age 25. Prophylactic total abdominal hysterectomy with bilateral salpingo-oophorectomy may also be considered at the time of colectomy. Chemoprophylaxis with aspirin 600 mg daily may be considered; one randomized controlled trial did show benefit.

MUTYH-ASSOCIATED POLYPOSIS

MUTYH-associated polyposis is a recently described autosomal recessive syndrome caused by mutations in the *MUTYH* gene (also called *MYH* [mutY homologue]). It is characterized by colonic polyposis and a high rate of CRC. Approximately 0.4% to 0.7% of CRC patients are homozygous for *MUTYH* mutations.

PATHOBIOLOGY

MUTYH-associated polyposis is caused by a biallelic inherited defect in the *MUTYH* gene located on chromosome 1p. Inherited as an autosomal recessive trait, this leads to defects in base excision repair and acquired mutations of the *APC* gene and other genes, such as *KRAS*. This results in adenoma formation and the subsequent development of adenocarcinoma.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Phenotypically, affected patients are similar to those with attenuated FAP. Patients have five to hundreds of adenomas. Multiple hyperplastic polyps and serrated adenomas may also be seen. The onset is later than in classic FAP, with cancers more likely to be right sided and to occur at age 45 to 50. Associated extracolonic features include gastroduodenal polyps, duodenal carcinoma, breast and ovarian cancer in female carriers, bladder cancer, skin cancer, congenital hypertrophy of the retinal pigment epithelium, and osteoma. Monoallelic carriers do not appear to have an increased cancer risk.

The diagnosis is suggested by the presence of colonic polyposis in the absence of FAP, or when there appears to be recessive inheritance. In these cases, genetic testing for *MUTYH* gene mutations should be considered. Whether heterozygote carriers are at increased risk has not been established, but screening similar to that for individuals with first-degree relatives with CRC may be advisable (i.e., colonoscopy at age 40 and then every 5 years).

TREATMENT Rx

Patients with numerous polyps should undergo colectomy. Patients with mild disease and a relatively small number of polyps may be considered for colonoscopy with polypectomy and regular surveillance. Colonoscopy should be performed beginning at age 18 to 20 and repeated every 1 to 2 years. Regular endoscopic surveillance for duodenal polyps should also be performed beginning at age 25 to 30.

PEUTZ-JEGHERS SYNDROME

PJS is an intestinal hamartomatous polyposis of the upper and lower GI tract that is associated with characteristic mucocutaneous pigmentation. The average age at diagnosis is in the mid-20s. PHS predisposes to both intestinal and extraintestinal malignancies.

PATHOBIOLOGY

PJS is a rare autosomal dominant syndrome with high penetrance. The prevalence is between 1 in 8300 and 1 in 29,000. The gene responsible for the syndrome is the serine-threonine kinase *(STK11)* gene located on chromosome 19p; a mutation in *STK11* is found in approximately 60% of patients with this syndrome. The hamartomatous polyps in PJS are located predominantly in the small intestine (64 to 96%), stomach (24 to 49%), and colon (60%). Histologically, these polyps are benign; they are unique in that a layer of muscle that extends into the submucosa or muscularis propria may surround the glandular tissue. Adenomatous and hyperplastic polyps may also be found.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The most common symptoms are small bowel intussusception, obstruction, and GI bleeding that may require surgery and may be recurrent. PJS is associated with an increased risk of cancer, with an estimated 47% of patients developing a malignancy by age 65. The most common cancers are those of the small intestine, stomach, colon, pancreas, testes, breast, ovary, cervix, and uterus. More than 95% of patients have a characteristic pattern of melanin spots on the lips, buccal mucosa, and skin (Fig. 193-5). Because genetic testing is not widely available, first-degree relatives should be screened beginning at birth with an annual history, physical examination, and evaluation for melanotic spots, precocious puberty, and testicular tumors.

TREATMENT Rx

Standard medical care for patients with PJS involves an annual physical examination that includes evaluation of the breasts, abdomen, pelvis, and testes, as well as a complete blood cell count. Surveillance for cancer includes small bowel radiography every 2 years, esophagogastroduodenoscopy and colonoscopy every 2 years, and endoscopic ultrasound of the pancreas every 1 to 2 years. For women, annual Pap smear, transvaginal ultrasound, CA125, and mammography are recommended. Polyps larger than 1 cm should be removed endoscopically. Laparotomy and resection are recommended for recurrent or persistent small intestinal intussusception, obstruction, or intestinal bleeding.

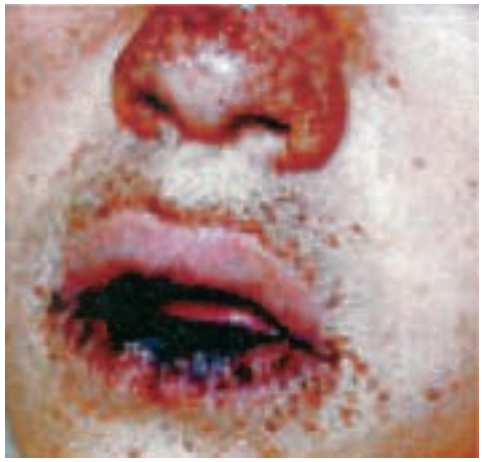

FIGURE 193-5. Mucosal pigmentation in a patient with Peutz-Jeghers syndrome.

JUVENILE POLYPOSIS

Familial juvenile (non-neoplastic, hamartomatous) polyposis is a rare (<1 in 100,000 births) syndrome characterized by 10 or more non-neoplastic hamartomatous polyps throughout the GI tract, or any number of polyps in a patient with a family history of juvenile polyposis. The syndrome is inherited in an autosomal dominant manner with high penetrance and is caused by mutations in the *SMAD4*, *PTEN*, or *BMPR1A* gene. The hamartomas are histologically distinct from the polyps seen in PJS. Patients generally present with rectal bleeding, anemia, abdominal pain, or intestinal obstruction in childhood or early adolescence. Extraintestinal symptoms include pulmonary arteriovenous malformations in some probands. The risk of malignancy in juvenile polyposis is reportedly as high as 20% and occurs in adulthood (median age, 37 years). Affected individuals should undergo regular colonoscopic surveillance. Patients with numerous, large, or high-grade dysplastic polyps may be considered for subtotal colectomy. Family members should be screened with colonoscopy every 3 to 5 years beginning at age 12 to 15 until age 40.

PTEN HAMARTOMA SYNDROME

PTEN hamartoma syndrome is a rare autosomal dominant syndrome consisting of multiple hamartomatous polyps of the skin and mucous membranes, including GI polyps, facial tricholemmomas, oral papillomas, and keratoses of the hands and feet. It was previously referred to as Cowden syndrome and Bannayan-Riley-Ruvalcaba syndrome. The causative genetic lesion has been mapped to the *PTEN* tumor suppressor gene. The rate of associated malignancy is high, particularly in the thyroid, breast, and reproductive organs. The polyps in *PTEN* hamartoma syndrome are benign. The incidence of CRC is approximately 10 times higher than in the general population.

CRONKHITE-CANADA SYNDROME

Cronkhite-Canada syndrome is a rare, sporadic, acquired condition characterized by multiple hamartomatous polyps throughout the GI tract, along with alopecia, dermal pigmentation, and atrophy of the nail beds. Symptoms include diarrhea, protein-losing enteropathy, GI bleeding, intussusception, and rectal prolapse. It carries a poor prognosis, with 5-year mortality rates as high as 55%. Patients are at risk for gastric and CRC, and endoscopic surveillance has been recommended.

MISCELLANEOUS GENETIC SYNDROMES

Recently, several new syndromes have been identified that predispose to CRC. In the serrated polyposis syndrome, affected individuals have multiple (≥5), large (at least 2 ≥ 10 mm), serrated adenomas, affected family members, and an increased CRC risk. Patients with defects in the epithelial cell adhesion molecule *(EPCAM)* gene have the phenotypic features of Lynch syndrome without mutations of the mismatch repair genes. Germline defects in the *POLE* and *POLD1* genes leads to DNA proofreading errors, polyposis, and/or CRC. The hereditary mixed polyposis syndrome is an autosomal dominant condition causing polyps of multiple and mixed morphologies and CRC; the causative genetic defect (a duplication spanning the 3′ end of the *SCG5* gene and a region upstream of the *GREM1* locus) has recently been discovered.

Polyps of the Colon

A polyp is defined as a grossly visible mass of epithelial cells that protrudes from the mucosal surface into the lumen of the intestine. A polyp may be sessile, flat, or pedunculated when it is attached by a stalk. Polyps are classified as either non-neoplastic or neoplastic (adenomatous). Polyps may rarely cause symptoms such as bleeding, prolapse, or obstruction. Neoplastic polyps have the potential to become malignant.

NON-NEOPLASTIC POLYPS

Non-neoplastic polyps account for approximately half of all mucosal polyps detected in the large bowel of average-risk individuals older than age 50 years. These polyps, which are also termed *nonadenomatous polyps*, can be subcategorized into hyperplastic, inflammatory, lymphoid, and juvenile polyps. Most non-neoplastic polyps are hyperplastic polyps that arise as a result of abnormal maturation of the mucosal epithelial cells; these polyps are usually small in diameter and are found predominantly in the distal sigmoid colon and rectum. Hyperplastic polyps are not malignant and are not thought to be associated with any measurable increase in malignant potential. Patients with inflammatory bowel disease may develop inflammatory pseudopolyps, which may require biopsy or removal to distinguish them from neoplastic polyps. Lymphoid polyps are regions of the mucosa that contain exaggerated intramucosal lymphoid tissue. Juvenile polyps usually develop in the rectum of children younger than 5 years and are termed *hamartomatous* because they are focal malformations that resemble tumors but are caused by abnormal development of the lamina propria; these polyps require no therapy unless they cause symptoms (e.g., obstruction, severe bleeding) or are part of a genetic syndrome.

ADENOMATOUS POLYPS

DEFINITION

Adenomatous polyps (or adenomas) are neoplastic polyps with malignant potential. They are benign glandular tumors that exhibit either low- or high-grade dysplasia under microscopy. Their anatomic distribution parallels that of colorectal adenocarcinoma. Adenomatous polyps manifest in a range of sizes and may be sessile, flat, or pedunculated in morphology. They are believed to be the precursor lesion to colorectal adenocarcinoma, a process that occurs along the adenoma-carcinoma sequence. Evidence supporting the adenoma-carcinoma sequence comes from several sources. Patients with genetic conditions that predispose to adenoma formation (e.g., FAP) develop cancer at high rates. Animal studies in which adenomas are induced by either carcinogens or genetic manipulation show carcinoma formation. Correlative evidence includes the observations that the epidemiology is similar for adenomas and carcinomas, that both lesions are more common in the same anatomic locations, and that adenomatous tissue can often be found in small adenocarcinomas. Intervention studies have shown that the removal of adenomatous polyps leads to a significant decrease in the risk for CRC.

EPIDEMIOLOGY

Adenomatous polyps are relatively common, particularly in elderly populations. Among healthy screening populations older than 50 years, adenomas are found in more than 15% of women and 25% of men. The prevalence of adenomas tends to be high in regions of the world where CRC is common. The importance of genetic risk factors is clear in the hereditary polyposis syndromes, and sporadic adenomas have a familial component; for example, individuals with a positive first-degree family history have a four-fold greater risk of developing adenomatous polyps. African Americans in the United States have an increased risk for developing adenomas and carcinomas relative to whites; the risk in Asian and Hispanic individuals is similar to that in whites.

PATHOBIOLOGY

The layer of epithelial cells lining the surface of the normal large bowel undergoes continuous self-renewal, with a turnover period of 3 to 8 days. Undifferentiated stem cells located at the base of invaginated crypts give rise to cells that migrate toward the lumen as they differentiate further into specialized enterocytes; these cells are subsequently removed by apoptosis, by extrusion, or by phagocytes underlying the epithelial layer. The development of adenomatous polyps is associated with a sequence-specific accumulation

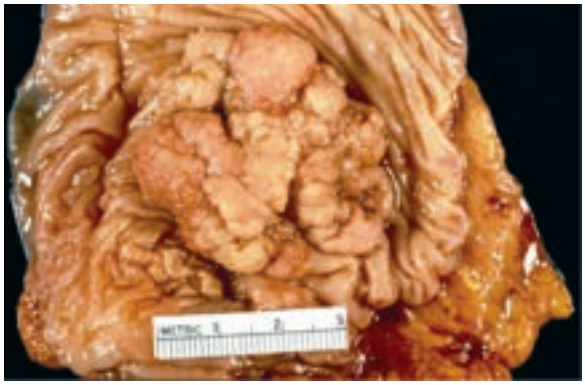

FIGURE 193-6. Villous adenoma. Large and sessile villous adenomas of the large bowel with finger-like projections into the gut lumen. (Courtesy Dr. M.K. Washington.)

of genetic lesions that cause an imbalance between epithelial cell proliferation and cell death. As a result, cells accumulate at the luminal surface, where they remain undifferentiated and continue to undergo cell division, eventually leading to the abnormal development of a mass of adenomatous tissue.

Adenomas are classified into three main histologic subtypes: tubular adenomas, villous adenomas (Fig. 193-6), and tubulovillous adenomas. Tubular adenomas account for 70 to 85% of all adenomas removed at colonoscopy. They are often small and pedunculated, and they consist of dysplastic tubular glands that divide and branch out from the mucosal surface; they rarely contain concomitant high-grade dysplasia or carcinoma. In contrast, villous adenomas (<5% of all adenomas) are generally large and sessile and are composed of strands of dysplastic epithelium that project, finger-like, into the lumen of the gut; they have a much higher prevalence of high-grade dysplasia or carcinoma. Tubulovillous adenomas (10 to 25% of all adenomas) have a mixture of tubular and villous architecture. Advanced adenomas are defined as those that measure 1 cm or greater or have any villous histology or high-grade dysplasia. Patients with advanced adenomas or multiple adenomas (three or more) are at much greater risk for synchronous (developing simultaneously) or metachronous (developing after a time interval) CRC.

CLINICAL MANIFESTATIONS

Patients with adenomatous polyps generally remain asymptomatic, but they may present with an asymptomatic positive stool occult blood test or with evident hematochezia. The lifetime incidence of additional adenomas in a patient with one known adenoma is 30 to 50%. Fewer than 5% of all adenomas eventually develop into carcinomas. Two critical factors that determine the likelihood of an adenoma developing into an invasive lesion are the size of the polyp and the grade of dysplasia. For polyps less than 1 cm, the risk for carcinoma is 1 to 3%; polyps between 1 and 2 cm have a 10% risk of becoming cancerous; and more than 40% of polyps greater than 2 cm progress to an invasive lesion. All adenomatous polyps contain some degree of dysplasia, but they can be further categorized as low- or high-grade to indicate the degree of dysplasia and the corresponding risk for invasive carcinoma. High-grade dysplasia is associated with a 27% rate of eventual transformation into carcinoma.

DIAGNOSIS

Adenomatous polyps in the colon and rectum can be diagnosed by endoscopy, barium radiography, or CT scanning (CT colography or virtual colonoscopy). Colonoscopy is the preferred method for diagnosing adenomas because of its high accuracy and the ability to immediately biopsy and resect most polyps. Barium enema, as assessed in the National Polyp Study, missed 52% of polyps measuring 1 cm or more. CT colography has good sensitivity for detecting large (>1 cm) polyps (>85%) and for detecting cancers (96%), but it is less sensitive and specific for smaller polyps. CT colography requires bowel preparation, exposes the patient to ionizing radiation, and cannot remove polyps. Colonoscopy may miss 6 to 12% of large (≥1 cm) polyps and 5% of cancers. Of note is that nonpolypoid (flat and depressed) colorectal neoplasms are found in about 9% of asymptomatic and symptomatic adults on colonoscopy and are more likely to contain a carcinoma than are polypoid lesions; these lesions may not be visible on barium radiography or CT.

Flexible sigmoidoscopy, which is often used to screen asymptomatic persons at average risk for colorectal adenocarcinoma, detects 50 to 60% of all polyps and cancers. Generally, patients who have polyps detected by barium radiography, CT colography, or flexible sigmoidoscopy should undergo colonoscopy to remove the lesion and search for additional polyps. In one study in which patients with polyps discovered by flexible sigmoidoscopy underwent subsequent colonoscopy, there was an 80% reduction in the incidence of CRC.

TABLE 193-2 COLONOSCOPY SURVEILLANCE INTERVALS

MOST ADVANCED FINDING	INTERVAL
No polyps or small (<10 mm) hyperplastic polyps	10 yr
1-2 adenomas, <1 cm	5-10 yr
3-10 adenomas or adenoma with villous features, ≥1 cm, or with high-grade dysplasia	3 yr
>10 adenomas	<3 yr
Sessile adenoma ≥ 2 cm, piecemeal excision	2-6 mo
Serrated lesions:	
<10 mm, no dysplasia	5 yr
≥10 mm or dysplasia	3 yr

SERRATED POLYPS

Sessile serrated adenomas/polyps (SSA/P) are a recently recognized neoplastic lesion with malignant potential. In the past, these have been confused with large, benign, hyperplastic polyps. SSA/P are characterized microscopically as having a disorganized and distorted crypt growth pattern. SSA/P are sessile or flat, may be difficult to distinguish on endoscopy and have a right colon predominance. The risk of progression to cancer is at least as high as for conventional adenomas. Pathogenesis is via hypermethylation of CpG islands ("CIMP-high") and *MLH1*, and have *BRAF* mutations, leading to polyps and tumors with microsatellite instability. Traditional serrated adenomas are a rare subtype that are histologically distinct from SSA/P and have a high prevalence of high-grade dysplasia and carcinoma in situ.

TREATMENT

Rx

The goal of treatment for neoplastic polyps is to remove or destroy the lesion during endoscopy. This recommendation is based on overwhelming evidence that endoscopic polypectomy reduces the subsequent incidence and mortality of CRC. Pedunculated adenomas are generally removed by snare polypectomy (Video 193-1), with subsequent submission of the tissue for pathologic analysis. Piecemeal snare resection may be required to remove sessile polyps. Surgical resection of a polyp is indicated when endoscopic resection of an advanced adenoma is not possible. The biopsied polyp must be evaluated histologically to determine the presence or absence of carcinoma; if a malignant lesion is found, its histologic grade, vascular and lymphatic involvement, and proximity to the margin of resection should be determined. Unfavorable histopathologic factors that should prompt surgical resection include poorly differentiated histology, vascular invasion, lymphatic invasion, and incomplete endoscopic resection. Malignant pedunculated polyps with cancer confined to the submucosa, with no evidence of unfavorable histologic features, can be definitively treated with endoscopic resection, without the need for surgical resection. Whether similar malignant sessile polyps can be managed nonoperatively is controversial. In these cases, the risk of surgery versus the risk of recurrence or lymphatic metastases needs to be balanced.

PROGNOSIS

Patients who have undergone resection of an adenomatous or sessile serrated polyp are at increased risk for the subsequent development of adenoma and colorectal adenocarcinoma. This risk is influenced by the size, histology, and number of adenomas, and the surveillance intervals differ (Table 193-2). Low-risk patients—those with only 1 or 2 small tubular adenomas—should undergo colonoscopy in 5 to 10 years. Patients with multiple (>2) adenomas, large (≥1 cm) adenomas, or adenomas with villous or high-grade histology should undergo colonoscopy in 3 years. Patients with numerous (>10) adenomas should undergo colonoscopy within 3 years. Patients who have had polypectomy of a large (≥2 cm) adenoma or an adenoma that had to be

VIDEO 193-1. **Snare polypectomy of a colon adenoma.**

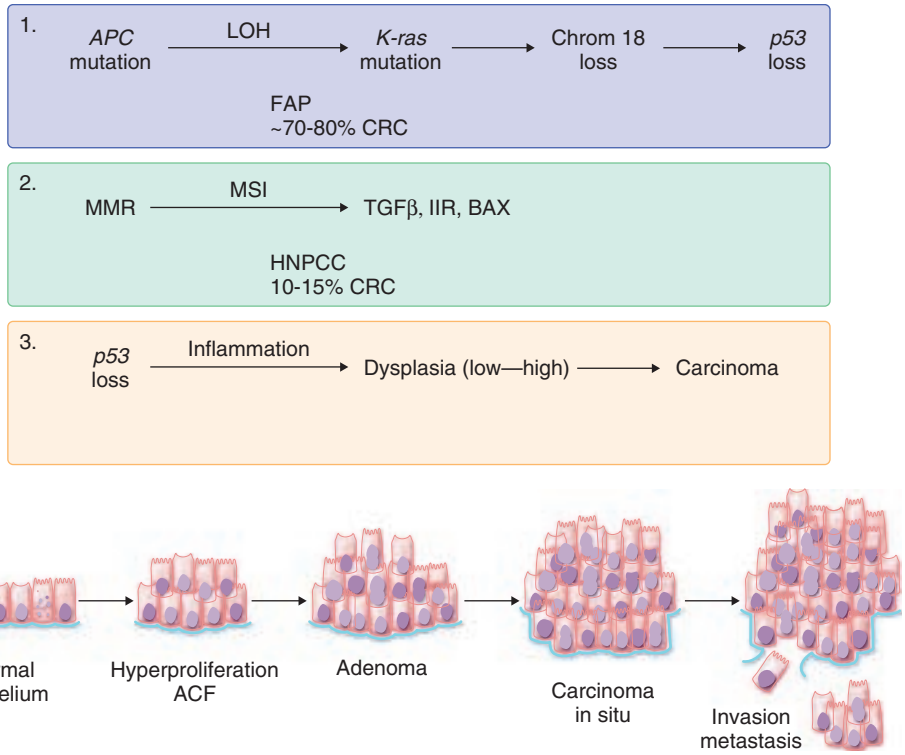

FIGURE 193-7. The molecular basis of colorectal cancer. Sequence-specific genetic lesions result in the transition from normal large bowel mucosa to invasive carcinoma. ACF = aberrant crypt foci; BAX = apoptosis-related protein; CRC = colorectal cancer; HNPCC = hereditary nonpolyposis colorectal cancer; IIR = type II receptor; LOH = loss of heterozygosity; MMR = mutation mismatch repair; MSI = microsatellite instability; TGFβ = transforming growth factor-β.

removed in pieces (piecemeal resection) should undergo colonoscopy within 6 months to evaluate the completeness of the resection. Patients with sessile serrated polyps smaller than 10 mm without dysplasia should undergo surveillance colonoscopy at 5 years. Patients with sessile serrated polyps 10 mm or larger, high-grade dysplasia, or a traditional serrated adenoma should undergo colonoscopy at 3 years (Video 193-3).

Adenocarcinoma of the Colon and Rectum

PATHOBIOLOGY

Colorectal cancer is caused by the accumulation of multiple genetic lesions over time. Except for hypermutated tumors, colon and rectal primaries have similar patterns of alterations. Both the tissue architecture and the cellular genotype change as the disease progresses (Fig. 193-7).[4] Three distinct molecular pathways have been recognized: chromosomal instability, microsatellite instability, and CpG island methylator phenotype (CIMP). These pathways are not mutually exclusive, and tumors may exhibit features of more than one.

The chromosomal instability pathway is the most common, accounting for up to 70% of sporadic CRC. The most common gene mutations are in the *APC* gene (a tumor suppressor gene) and *KRAS* (a proto-oncogene involved in the transduction of mitogenic signals across cell membranes). Germline mutations in *APC* are the cause of FAP. Chromosomal instability leads to aneuploidy (imbalance in chromosome number), genomic amplifications, and loss of heterozygosity where cells have only one allele of a gene owing to the loss of individual chromosomes during mitosis. Additional important affected genes include the mutated in colon cancer *(MCC)* gene (a tumor suppressor gene), *p53* (a regulator of the cell cycle), *VEGF*, *MYC*, *MET*, *LYN*, *PTEN*, and others. Many of the genetic changes affect the Wnt signaling pathway, which appears to be important for initiation and progression of CRC.

The microsatellite instability (MSI) pathway is caused by defects in DNA mismatch repair. Microsatellites are short, repeating nucleotide sequences that are prone to errors owing to their repetitive nature. MSI-high tumors have genetic defects in the mismatch repair genes, especially *MLH1*, *MSH2*, *MSH6*, and *PMS2*. Germline mutations in these genes are the cause of HNPCC (Lynch syndrome). Hypermethylation silencing of *MLH1* also results in MSI-high cancers. MSI-high tumors are more common in the right colon, in women, and have a lymphocytic infiltration and poor differentiation. They are associated with improved survival, despite being less responsive to some chemotherapeutic agents such as 5-fluorouracil.

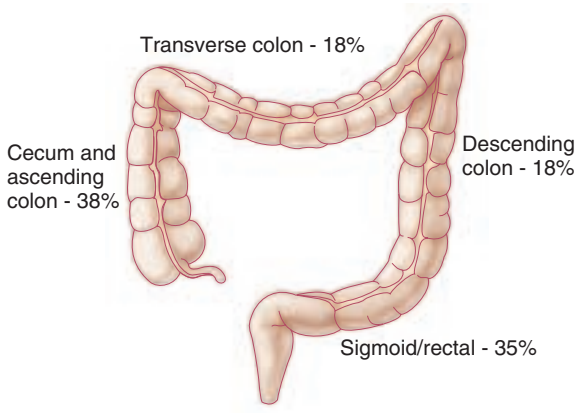

FIGURE 193-8. Sites of development of large bowel adenocarcinoma.

In the CIMP pathway, hypermethylation of DNA promoter regions leads to gene silencing, and silencing of tumor suppressor genes leads to carcinogenesis. In colorectal carcinogenesis, these include *APC*, *MCC*, *MLH1*, *MGMT*, and others. CIMP-high tumors are often poorly differentiated with mucinous or signet ring morphology, are MSI-high, and have *BRAF* mutations. The sessile serrated adenoma is likely the precursor lesion.

CLINICAL MANIFESTATIONS

The majority of patients with CRC present with symptoms; these may be emergent. Common symptoms related to primary disease include rectal bleeding with or without manifestations of anemia, abdominal pain, and change in bowel function. Patients with systemic disease may exhibit anorexia, weight loss, and symptoms related to hepatic dysfunction, such as jaundice, icterus, and ascites (the last may also be seen with peritoneal metastases). Symptoms vary depending on the primary site (Fig. 193-8). Proximal lesions are more likely to present with bleeding and associated symptoms; more distal disease has a higher risk of obstruction and perforation. Rectal cancers can also manifest with tenesmus and changes in stool caliber. They involve sacral nerve plexi, causing significant neuropathic pain. Patients with PJS or

VIDEO 193-3. **Endoscopic mucosal resection using saline lift polypectomy of a colon adenoma, followed by closure of the mucosal defect with clips.**

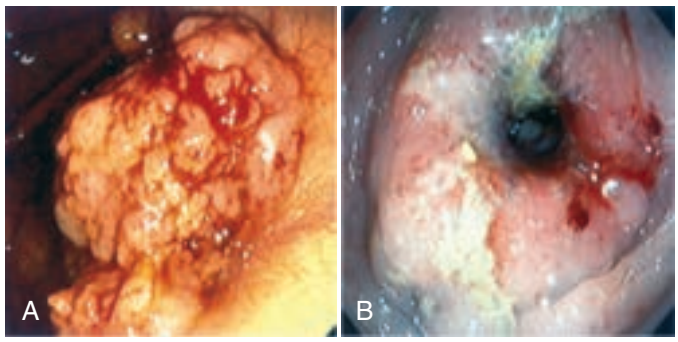

FIGURE 193-9. Two manifestations of large bowel adenocarcinoma. A, Exophytic growth within the lumen. B, Stricturing ("apple core") lesion.

Gardner syndrome may exhibit extraintestinal manifestations. Patients with *Streptococcus bovis* bacteremia or endocarditis are at increased risk of harboring CRC and should undergo colonoscopy.

TABLE 193-3 OPTIONS FOR COLORECTAL NEOPLASIA SCREENING*

TEST	INTERVAL
TESTS THAT DETECT ADENOMATOUS POLYPS OR CANCER	
Colonoscopy	Every 10 yr
Flexible sigmoidoscopy†	Every 5 yr
CT colography†	Every 5 yr
Double-contrast barium enema†	Every 5 yr
TESTS THAT DETECT PRIMARILY CANCER†	
Fecal occult blood tests	
High-sensitivity guaiac-based fecal occult blood test	Annually
Fecal immunochemical test	Annually
Stool DNA	Interval uncertain

*Beginning at age 50 for average-risk individuals.
†Positive test should prompt full colonoscopy.
CT = computed tomography.

DIAGNOSIS

The history, physical examination, and judicious use of both laboratory and radiologic tests are important in diagnosing CRC. The history should include the possibility of prior CRC or adenomatous polyps, inflammatory bowel disease, and family history of colonic neoplasia. On physical examination, extraintestinal lesions characteristic of PJS or Gardner syndrome may be noticed. Metastatic disease may be suggested by enlargement of left supraclavicular lymph nodes (Virchow's nodes) or the liver, or by the presence of an umbilical mass (Sister Mary Joseph's node) or ascites. The digital rectal examination may reveal a distal rectal cancer or spread of the tumor to the rectal shelf or pelvis (Blumer's shelf). The stool shows evidence of frank or occult blood in 40 to 80% of advanced cases. Iron deficiency anemia or an elevation in liver enzymes may aid in the diagnosis. The CEA level may be elevated, but it cannot be relied on for diagnosis, owing to inadequate sensitivity.

Methods for diagnosing CRC are similar to those used to detect adenomatous polyps. Colonoscopy is the procedure of choice for all patients who have occult blood in their stools, unexplained iron deficiency anemia, or signs and symptoms suggestive of CRC (Fig. 193-9). Colonoscopy is more accurate than barium radiographic studies for the detection of colorectal neoplasms of all sizes and has the advantage of enabling the clinician to detect synchronous cancers and to obtain tissue for histologic analysis.

Additionally, accurate *local* staging of rectal cancers is paramount. Endoscopic ultrasound combines high-frequency ultrasonography with videoendoscopy. It is superior to CT and allows an accurate determination of the degree of invasion and detection and sampling of enlarged lymph nodes. Endoscopic ultrasound is also highly sensitive for the detection of rectal cancer recurrence after local resection or low anterior resection. MRI using either endorectal or phased array coils can also provide accurate local staging of rectal cancer. Local staging of nonrectal large bowel cancer is generally not performed preoperatively, because this information is not used to guide therapy.

Many expert consensus guidelines now recommend CT scans of the abdomen and pelvis for CRC patients because advanced liver metastases would preclude resection of an asymptomatic primary. Chest imaging with plain films or CT is recommended as well. PET scans have specific uses in defined cases of CRC, such as precluding additional systemic disease before resection of a solitary metastatic site. They may also be used to further describe abnormalities seen on CT. However, there is no role for PET in the routine work-up. Plain abdominal films are mostly useful in diagnosing obstruction.

Screening

The purpose of screening is to reduce CRC-related mortality by removing the precursor adenomas and detecting prevalent cancers at earlier, more curable stages.[5] The long latency between adenoma development and subsequent cancer, on the order of 10 to 20 years, makes CRC a preventable disease through colonoscopy with polypectomy. It is an age-associated disease, and most patients should begin screening at age 50.[6] There is a range of options for average-risk individuals (Table 193-3). These can be divided into two categories: stool tests, which include tests for occult blood and abnormal DNA, and structural tests, which include colonoscopy, flexible sigmoidoscopy, CT colography, and double-contrast barium enema. Stool tests are best suited for detecting prevalent cancers (and some advanced adenomas), whereas structural tests detect both cancers and adenomas. In a randomized trial of asymptomatic adults 50 to 69 years of age, one-time colonoscopy and fecal immunochemical testing were equally good at finding prevalent colon cancer, but more adenomas were identified by colonoscopy.[6] These tests may be used alone or in combination. Current multisociety guidelines recognize the multiple screening options but encourage the use of structural tests that have the ability to both detect and prevent CRC. For example, the American College of Physicians recommends that clinicians screen for CRC in average-risk adults starting at age 50, and in high-risk adults starting at age 40 (or 10 years younger than the age at which the youngest affected relative was diagnosed with CRC), using a stool-based test, flexible sigmoidoscopy, or optical colonoscopy in patients who are at average risk, but optical colonoscopy in patients who are at high risk. Clinicians should stop screening for CRC in adults older than 75 years or in adults with a life expectancy of less than 10 years.

Tests for fecal occult blood detect hemoglobin in the stool from bleeding tumors. These tests are either guaiac based or immunochemical based. The guaiac tests detect blood in stool through the pseudoperoxidase activity of heme or hemoglobin. The immunochemical tests react with human globin and are therefore more specific. Guaiac testing consists of collecting two samples from three consecutive stools. To improve test accuracy, individuals undergoing guaiac testing are instructed to avoid aspirin, nonsteroidal anti-inflammatory drugs (NSAIDs), vitamin C, red meat, poultry, fish, and some vegetables. Three large prospective randomized trials demonstrated that guaiac-based testing decreased CRC mortality by 15 to 33%, and this benefit appears to persist for 30 years.[7] One large U.S. study also showed a 20% decrease in CRC incidence, attributed to the relatively higher rates of colonoscopy in the study group. Guaiac-based tests have sensitivities for cancer from 35 to 80%. Immunochemical-based tests do not rely on a peroxidase reaction and therefore may have fewer false positives and false negatives. They are done essentially the same way as the traditional guaiac, but they do not require the dietary restrictions of guaiac-based tests and only one stool specimen is collected. The immunochemical tests have better adherence, find 2.5 times as many cancers and advanced adenomas as guaiac-based tests, and are now the preferred fecal occult blood test for screening. Fecal occult blood tests are repeated annually, and any positive test should prompt colonoscopy.

Stool DNA tests rely on the observation that both adenomas and carcinomas contain altered DNA, and this DNA is shed in the stool. Tests contain a multiple marker panel designed to detect exfoliated DNA markers. Collection kits are designed to facilitate stool collection (at least 30 g) and mailing. First-generation stool DNA tests had better sensitivity for CRC (52%) than a guaiac-based test (unrehydrated Hemoccult II; 13% sensitivity) but poor sensitivity for advanced adenomas (15%). Next-generation stool DNA testing with an updated marker panel can identify about 90% of patients with CRC and about 40% of patients with advanced adenomas, with an 87% specificity.[8] Stool DNA testing is more expensive than fecal occult blood testing, and the optimal screening interval has not been defined.

Flexible sigmoidoscopy is capable of examining the distal 60 cm of the colon, or roughly the splenic flexure to the rectum. It can be done with minimal bowel preparation, does not require sedation, and can be performed

by primary care providers, nurses, or physician's assistants. Flexible sigmoidoscopy can detect distal cancers or polyps as well as colonoscopy can.[9] The detection of an adenoma should prompt referral for full colonoscopy, owing to the high prevalence of synchronous neoplasia in the unexamined proximal colon. In one randomized trial, a single flexible sigmoidoscopy performed once between ages 55 to 64 years reduced CRC incidence by 23% and mortality by 31%.[■] In another randomized trial, screening with flexible sigmoidoscopy led to a significant 29% decrease in CRC incidence in both the distal and proximal colon, and a significant 50% decrease in mortality from cancers of the distal colon only.[■] In a third trial, CRC incidence was reduced by 20% and death by 12%.[■] Sigmoidoscopy should be repeated every 5 years and may be combined with yearly fecal occult blood testing.

Colonoscopy allows complete examination of the colon as well as adenoma removal, and thus cancer prevention. In the United States, it is generally done under sedation after an oral bowel preparation, and by physicians with specific training in colonoscopy and polypectomy. Large cohort studies have found an up to 1% cancer rate, 5 to 10% advanced adenoma rate (size > 10 mm, villous histology, or high-grade dysplasia), and at least 20% with 1 or more adenomas.[10] Major complication rates are less than 0.5%, with more than half due to polypectomy. There are no direct randomized studies assessing the efficacy of colonoscopy for screening, although there is a large amount of indirect evidence. Colonoscopy is generally used as the gold standard in assessing other screening methods. In a large randomized Minnesota study of fecal occult blood testing, a decrease in cancer incidence was observed that could be explained only by the use of colonoscopy and polypectomy at a higher rate in the screened population. Randomized and case-control studies of sigmoidoscopy demonstrate a CRC mortality benefit that should also apply to colonoscopy. The National Polyp Study, which followed patients after polypectomy, found that the incidence of CRC was reduced 76 to 90% compared to three reference populations. In another study of patients who had adenomas detected and removed by colonoscopy, the risk of death from CRC at 16 years was only 50% of what was expected in the general population,[11] with most of the reduced incidence found among patients who had low-risk adenomas removed.[12] The efficacy of colonoscopy is influenced by the quality of the exam, with the most important quality markers being the cecal intubation rate (indicating a complete exam) and the adenoma or polyp removal rate. Studies have demonstrated superior CRC prevention when the colonoscopy is performed by a physician with a high adenoma or polyp removal rate.[13] Physician experience and specialty are also important, with gastroenterologists preventing more CRC than other specialties.[14] If the initial screening colonoscopy is normal, further screening can be deferred for 10 years.[15]

Double-contrast barium enema evaluates the entire colon by coating the mucosal surface with high-density barium and insufflating with air via a rectal tube. Multiple radiographs are obtained while varying the patient's position under fluoroscopy. It requires a colon preparation and exposes the patient to a small amount of ionizing radiation. There are no studies of double-contrast barium enema as a screening test. It is 85 to 97% sensitive for colon cancer, but its sensitivity for detecting large adenomas is only 48 to 73%. It should be repeated every 5 years. The finding of a polyp larger than 5 mm should prompt colonoscopy.

CT colography (also known as virtual colonoscopy) uses multidetector CT technology to obtain two- and three-dimensional images of the entire colon. It requires an adequate colon preparation and gaseous distention of the bowel using a rectal tube. Tagging of residual stool with barium or iodinated contrast material is frequently used. CT scanning is performed with the patient in the supine and prone positions. No studies have been done to evaluate the efficacy of CT colography in decreasing CRC incidence or mortality. CT colography has been compared with standard colonoscopy for the detection of neoplasia in screening populations. The sensitivity of CT colography for large (≥10 mm) polyps is 85 to 92%, with a specificity of 83 to 86%. For polyps 5 mm or larger, CT colography is about 65% sensitive and 89% specific. Sensitivity for CRC (96%) is similar to that of colonoscopy. Laxative-free CT colography is accurate in detecting adenomas 10 mm or larger but less so for smaller lesions.[16] Although the optimal interval for testing has not been established, it is recommended that CT colography screening begin at age 50; if negative, it should be repeated at 5-year intervals. If a polyp larger than 5 mm is found, colonoscopy should be performed.

Staging

Accurate staging of CRC is of the utmost importance in determining both prognosis and the most relevant and effective therapy. The versions of the TNM classification system for large bowel cancers used by the American Joint Committee on Cancer and the International Union Against Cancer are identical (Table 193-4). Differing from many solid tumors (although not those of other GI origin, except for anal cancer), CRCs are not staged according to size. Stage I cancers penetrate into but not through the bowel wall (T1–2N0M0), whereas stage II cancers penetrate through the wall and can involve nearby organs without spreading to regional lymph nodes. About 40% of patients present with stage I or II disease in the United States. Stage III cancers involve regional lymph nodes and constitute about 40% of presenting cases. Stage IV colorectal tumors (distant metastasis or metastases) commonly involve liver, lung, distant nodes, and peritoneum, with about 20% of patients presenting with this stage. Rectal primaries, because of early access to the systemic circulation, may involve the lungs without liver metastases; this pattern of spread is distinctly unusual in proximal large bowel cancers.

TABLE 193-4 STAGING FOR COLORECTAL ADENOCARCINOMAS

STAGE	TUMOR	NODE	METASTASIS
0	Tis	N0	M0
I	T1, T2	N0	M0
IIA	T3	N0	M0
IIB	T4a	N0	M0
IIC	T4b	N0	M0
IIIA	T1, T2 T1	N1 N2a	M0 M0
IIIB	T3, T4a T2, T3 T1, T2	N1 N2a N2b	M0 M0 M0
IIIC	T4a T3, T4a T4b	N2a N2b N1, N2	M0 M0 M0
IVA	Any T	Any N	M1a
IVB	Any T	Any N	M1b

Tis = in situ; T1 = submucosa; T2 = muscularis propria; T3 = subserosa, pericolorectal tissues; T4a = visceral peritoneum; T4b = other structures.
N1a = 1 regional node; N1b = 2-3 regional nodes; N1c = satellite(s) without regional nodes; N2a = 4-6 regional nodes; N2b = ≥7 regional nodes.
M1a = 1 organ; M1b = >1 organ, peritoneum.

TREATMENT

Chemoprevention

NSAIDs, including aspirin, are believed to reduce adenoma formation and inhibit colon cancer development by inhibiting cyclooxygenase and subsequent prostaglandin generation (Chapter 37). Prostaglandins (e.g., E_2) promote cell proliferation and tumor growth. The NSAIDs sulindac and celecoxib cause regression of existing adenomas and inhibit the formation of new adenomas in patients with FAP. Epidemiologic studies have shown decreased CRC rates in regular users of NSAIDs. Randomized trials of aspirin have shown 20 to 40% reductions in adenoma recurrence. A secondary analysis combining data from four European vascular event prevention trials reported up to a 70% reduction in the incidence of CRC when at least 75 mg of aspirin daily was continued for 5 or more years.[17] However, the 4 individual studies and two large American trials (Women's Health Trial, Physicians Health Study) did not find an aspirin benefit. Calcium supplementation (1200 mg/day) was shown to decrease the rate of metachronous adenomas by 20%. However, the large Women's Health Trial (36,000 participants) found that calcium (1000 mg) plus vitamin D (400 IU) had no effect in reducing CRC incidence. Currently, the routine use of aspirin and calcium for CRC prevention is not recommended, but it may be considered on an individual basis.

Dietary Prevention

Epidemiologic studies have reported correlations between CRC and obesity, smoking, inactivity, excessive alcohol use, and diets high in fat and low in fruits, vegetables, and fiber. These observations suggest that lifestyle modifications may decrease CRC risk. Unfortunately, three randomized intervention trials of modest dietary changes (10% less fat, 25 to 75% more fiber, 50% more fruits and vegetables) found no significant reductions in adenomas or CRC over 3 to 8 years of follow-up. Fish consumption is inversely associated with CRC

incidence, and eicosapentaenoic acid supplementation decreases polyp burden in FAP.

Surgery

Resection is the primary treatment modality for patients with regionally confined CRC. Highly selected patients with metastatic disease may also undergo surgery with curative intent. The goal of curative surgery for colonic adenocarcinoma is margin-negative elimination of the tumor, plus en bloc removal of the primary feeding arterial vessel and corresponding lymphatics for that segment of bowel. A minimum of 12 lymph nodes should be retrieved for microscopic examination to assure staging accuracy. Synchronous colon cancers may be removed individually or with subtotal colectomy, and tumors adherent to adjacent structures should be resected en bloc. Prophylactic oophorectomy is no longer recommended, but women with one ovary grossly involved with cancer should undergo bilateral oophorectomy because of the relatively high risk of involvement of the other side. Laparoscopic resection is currently thought to be as effective as open resection and requires a modestly shorter recovery time. Very preliminary data suggest elderly patients undergoing a laparoscopic procedure have a lower chance of being discharged to a nursing home (as opposed to their own residences) than those treated with standard resection.

In general, surgical considerations for rectal primaries are similar. Total mesorectal excision (en bloc removal of the lymphovascular and fatty envelope surrounding the rectum) is recommended for distal cancers, whereas tumor-specific mesorectal excision (en bloc removal of the mesorectum 5 cm distal to the tumor) should suffice for upper rectal tumors. Local transanal excision is acceptable for selected low rectal cancers thought to have minimal risk of nodal involvement. Selection criteria include the following: T1 disease, size less than 3 cm, low grade (well differentiated), location within 8 cm of the anal verge, no lymphovascular invasion, and less than one third circumferential. The wide spectrum of symptoms that occur in most patients after resection and reconstruction of the rectum, ranging from increased bowel frequency to fecal incontinence or evacuatory dysfunction, has been termed *anterior resection syndrome*.

Resection of the primary formerly was recommended for patients presenting with synchronous CRC and unresectable metastases, regardless of whether the primary caused symptoms. This practice pattern obviously had the potential to delay systemic treatment in patients who were markedly more likely to die from their metastases before suffering significant complications from the intact large bowel tumor. Recent data suggest that stage IV patients with an asymptomatic primary tumor can safely begin systemic therapy without undergoing surgery, with only a small chance of developing serious complications requiring urgent operative interaction. Patients with rectal primaries may be at slightly higher risk for developing complications than those with tumors originating in the proximal large bowel.

Obstructing tumors that can be fully removed should be resected, with bowel anastomosis usually being acceptable in this setting. Proximal diversion alone, especially in the setting of very locally advanced unresectable cancer, may be necessary; if the primary tumor responds to the point where it can later be removed, resection followed by ostomy closure is reasonable. Endoscopic stenting may be useful to relieve acute obstruction. Perforated bowel is usually resected, with the choice of anastomosis, with or without diversion, depending on a number of factors, including degree of fecal contamination and general health of the patient.

Radiation Therapy

Radiation therapy may be used as curative or palliative treatment of large bowel cancers. In general, it is employed much more commonly in treating rectal versus colonic primaries. Single-institution trials have shown that irradiation improves local control following resection of high-risk proximal large bowel (colonic) cancers, but these findings were not confirmed in a randomized intergroup trial that closed early owing to slow accrual. Current recommendations for adjuvant radiation to the tumor bed following colon cancer resection include positive margins and localized perforation. Some authorities also advocate its use in colon cancers at particularly high risk of local recurrence (T4, T3N1-2 tumors in the ascending or descending colon), but that recommendation is not universally accepted. Irradiation may still play a role in treating colon cancer metastases to bone, brain, liver, and lung, as well as in cases of bleeding, obstruction, and locally advanced unresectable disease.

A major use for radiation in the definitive treatment of large bowel adenocarcinoma involves perioperative therapy for resectable rectal cancer. It is also commonly employed with chemotherapy for unresectable locally advanced invasive tumors, which occasionally may downsize and be surgically removed after therapy. As with colon primaries, irradiation may be used to palliate bleeding, obstruction, or selected metastases from rectal cancers.

Systemic and Combination Therapies

The backbone of CRC treatment in both the adjuvant and metastatic settings is a fluorinated pyrimidine. The most commonly used drug is 5-fluorouracil (5-FU), although oral prodrugs are increasingly being used. 5-FU targets the enzyme thymidylate synthase, inhibiting DNA synthesis and/or repair. It also may be incorporated into RNA, interfering with further processing. Although not particularly effective as a single agent (see later), 5-FU's efficacy can be enhanced by changing its means of administration (prolonging infusion) and by administering it with a variety of biochemical modulators, most commonly leucovorin. Other chemotherapeutic agents commonly used in treating advanced CRC include irinotecan, a topoisomerase I inhibitor, and oxaliplatin, a later-generation platin that forms bulky DNA adducts that inhibit replication. Recent effective biologics[18] include agents targeting vascular endothelial growth factor (bevacizumab) or circulating VEGF (aflibercept), the epidermal growth factor receptor (cetuximab, panitumumab), or a combination of the VEGF and various tyrosine kinase receptors (regorafenib).

Patients with resected stage I colon cancers have a high cure rate, and this cannot easily be improved on with systemic therapy. Stage II patients have a higher chance of relapse (event-free survival ≈ 76% at 3 years), and systemic therapy with 5-FU and leucovorin improved that figure by about 3% for an unselected group of node-negative patients. Patients harboring highly microsatellite-unstable tumors have a better prognosis in general but may actually have worse outcomes with 5-FU treatment, although that is controversial. Treatment of stage II patients remains controversial in general: some experts suggest that the proportional benefit of systemic therapy is as great as it is in stage III disease, while others point out the low absolute magnitude of benefit and do not advocate its use. Risk stratification using molecular markers has been attempted but has not been demonstrated to have predictive capabilities, with the exception of microsatellite instability and its described resistance to fluoropyrimidines. Some clinical categories of stage II disease are believed to be at particularly high risk for recurrence (e.g., obstruction). These patients usually receive postoperative chemotherapy, often with regimens commonly prescribed for patients with stage III cancer. Radiation has been tested in patients thought to be at higher-than-average risk of local relapse (T4 tumors), but its use is not standard. Other drugs useful in metastatic disease and resected node-positive patients (specifically oxaliplatin) are not routinely recommended nor used in those with stage II cancers.

Five-year disease-free survival for patients with stage III colon cancer ranges from 45 to 85%, depending on substage. Barring significant comorbidities or other confounding factors, all patients with node-positive disease receive adjuvant systemic therapy. Combinations of 5-FU and oxaliplatin (usually FOLFOX [fluorouracil, leucovorin, oxaliplatin]) clearly improve long-term survival rates[19] and represent standard care. Regimens containing irinotecan and biologics (bevacizumab, cetuximab) are highly effective in the treatment of advanced disease (see later), but oddly they do not clearly benefit patients when they are given in the postoperative setting. Patients who are not candidates for combination chemotherapy may be offered capecitabine, an oral prodrug activated to 5-FU in sequential enzymatic steps. Capecitabine has also been combined effectively and relatively safely with oxaliplatin for adjuvant use. An important outstanding question regarding adjuvant chemotherapy is whether a shorter duration (3 months vs. the standard 6) might be equally effective, as suggested in one small randomized trial.

Rectal Adenocarcinoma

Therapeutic considerations are slightly different for patients with primary rectal adenocarcinomas, owing to the difficulty of achieving negative circumferential (radial) margins. Specifically, the risk of local recurrence is much more significant. Adjuvant chemotherapy for rectal cancer is still controversial.[20] In general, rectal cancer patients undergoing standard resection for stage I tumors do not receive additional treatment. However, higher-risk patients (T2 disease, T1 with poorly differentiated histology, perineural or lymphovascular invasion, or close margins) treated with local excision should receive postoperative pelvic irradiation with or without 5-FU chemotherapy, or they should return to the operating room for total mesorectal excision. Irradiation is standard for those with stage II and III rectal adenocarcinomas to decrease local relapse rates, increase the chance of sphincter preservation (when used in selected preoperative settings for low-lying cancers), and possibly improve survival. The major considerations are timing (pre- or postoperative use), course (short or long), and whether to combine it with fluoropyrimidine-based chemotherapy. Short-course (5-day) preoperative irradiation without chemotherapy may be considered and used if tumor downsizing is not necessary. When short-course radiation is used, patients still require postoperative systemic therapy with either a fluoropyrimidine alone (stage II) or a fluoropyrimidine-oxaliplatin combination regimen (node-positive disease). Long-course irradiation (≈5.5 weeks) is particularly important when tumor response is necessary to make surgery easier or more feasible. It is usually combined with continuous-infusion 5-FU or capecitabine, and it can be given pre- or postoperatively (if the patient did not receive preoperative short-course radiation). Including oxaliplatin preoperatively does not improve results and is not recommended independent of a clinical trial, owing to its higher toxicity profile. Oxaliplatin may be used postoperatively if the pathology specimen demonstrates nodal involvement. With long-course irradiation, preoperative (compared with postoperative) treatment is less toxic and may offer improved local control.[21] However, it requires accurate preoperative staging (to avoid treating patients who might have stage I disease). Highly

selected rectal cancer patients thought to be at low risk of local recurrence (T3N0 or T1-2N1) may receive total mesorectal excision plus best systemic therapy without irradiation, usually postoperatively.

Metastatic Colorectal Cancer

Metastatic disease is treated identically, regardless of the site of origin (colon vs. rectum). Patients with incurable metastatic CRC have a median survival of approximately 6 months with best supportive care alone; however, incremental gains made through the adoption of new systemic therapies have extended that time to nearly 2 years or more. For example, treatment with single-agent fluoropyrimidines leads to median survival in the 10- to 13-month range; adding one more effective chemotherapy drug (either irinotecan or oxaliplatin) affords survival of about 15 to 20 months, and adding both drugs to 5-FU has been reported to extend life to 23 months. Interestingly, long-term results seem to be similar regardless of which drug (oxaliplatin or irinotecan) is added to 5-FU first (although the toxicity pattern varies, depending on which drug is given), or even regardless of whether 5-FU is used first and combination chemotherapy is used subsequently, as long as patients are eventually exposed to all active agents.

Additional improvements have arisen through the development of drugs that are more convenient and/or less toxic than standard agents, although they may not be more effective. Capecitabine, an oral prodrug activated to 5-FU in three sequential enzymatic steps, can be substituted for that agent alone and in combination with oxaliplatin (although patients still need intravenous access for the latter drug).

Recent breakthroughs have mostly been related to use of biologic agents (Chapter 36). Drugs used successfully to date include several monoclonal antibodies and one oral tyrosine kinase inhibitor. Bevacizumab is a humanized monoclonal antibody directed against vascular endothelial growth factor, an important mediator of angiogenesis. Bevacizumab has little single-agent activity against CRC, but it improves the interval without progression when added to irinotecan- or oxaliplatin-containing chemotherapy. Bevacizumab with FOLFOX or FOLFIRI (fluorouracil, leucovorin, irinotecan) now represents first-line treatment for patients with advanced large bowel cancer in the United States,[A1] and many oncologists continue its use along with second-line fluoropyrimidine-based chemotherapy. Aflibercept, or VEGF trap, is a fusion protein with VEGF-binding portions from VEGFR and the Fc portion of IgG1. It prevents VEGF-A and B from binding to receptors. A second-line trial after oxaliplatin failure showed aflibercept with irinotecan- and fluoropyrimidine-based chemotherapy (FOLFIRI), improves survival over chemotherapy with placebo. It is unknown whether it is better to continue bevacizumab versus switching to aflibercept in the setting of second-line therapy.

The epidermal growth factor receptor (EGFR) is another important target in advanced CRC. Cetuximab and panitumumab are monoclonal antibodies (chimeric and human, respectively) directed against the EGFR. They have single-agent activity against CRCs, and both may be combined with irinotecan-based chemotherapy to improve progression-free survival. Panitumumab also improves progression-free survival front-line when combined with oxaliplatin-containing chemotherapy. Although both agents appear effective in front-line or later use, efficacy is restricted to patients whose tumors harbor wild-type *KRAS*.

Regorafenib is an oral inhibitor of angiogenic (VEGFR-1, -2, -3, and TIE-2), stromal (PDGFR-B, FGFR), and oncogenic (KIT, RET, BRAF) tyrosine kinases. It has activity in advanced CRC, improving progression-free and overall survival after failure of standard therapies.[A8]

Locally Directed Treatment of Metastatic Disease

Complete resection of hepatic or pulmonary metastases may result in long-term survival and is the standard of care for selected patients with CRC. The majority of data exist for hepatic metastasectomy. Actuarial 5-year survival rates have been in the 25 to 60% range. However, relapse after 5 years still occurs, suggesting that this percentage does not reflect the number actually *cured* with surgery. That figure, calculated from 10-year survival rates, is probably between 17 and 25%, which still compares quite favorably with the survival rate for patients with CRC metastatic to the liver who do not undergo surgery. The role of preoperative systemic therapy in potentially resectable patients remains controversial; if used, a fluoropyrimidine with either oxaliplatin or irinotecan +/– panitumumab (for wild-type *KRAS*) or irinotecan-based chemotherapy +/– cetuximab (for wild-type *KRAS*), or triple chemotherapy with a fluoropyrimidine, oxaliplatin, and irinotecan may be employed. In those undergoing metastasectomy, many experts advocate 6 months of postoperative fluoropyrimidine-based chemotherapy, often with oxaliplatin. Importantly, bevacizumab can impair wound healing, and most experts avoid its use in the immediate perioperative period.

Small liver metastases that are not resectable because of anatomic location or in a frail patient unable to undergo hepatic resection may be treated with radio frequency ablation, which uses alternating electric current to generate heat, destroying malignant cells through protein coagulation. Although it has never been directly compared with resection in a randomized trial, radio frequency ablation appears to offer inferior local control of disease. Some advocate the placement of a hepatic artery pump to infuse fluoropyrimidine-based chemotherapy to treat unresectable liver-predominant metastases. Other techniques useful in noncolorectal liver-based tumors (e.g., hepatocellular carcinoma), such as chemoembolization, have no proven role for large bowel liver metastases.

Surveillance

Surveillance should be undertaken in patients fit enough to undergo metastasectomy or systemic therapy for recurrent CRC. The American Society of Clinical Oncology recommends annual CT of the chest and abdomen for 3 years following resection of high-risk primary tumors, with pelvic CT added in cases of rectal origin. Colonoscopy should also be done at 3 years and then every 5 years, with flexible proctosigmoidoscopy offered to rectal cancer patients who have not received irradiation. Physical examinations should be performed every 3 to 6 months for 3 years, then biannually for at least 2 more years. CEA levels should be checked every 3 months for at least 3 years, although the benefits for improving outcome are uncertain.[A9]

PROGNOSIS

The prognosis for patients with CRC depends primarily on stage. Five-year survival for patients with proximal (colonic) adenocarcinomas ranges from a low of 6 to 8% for those with metastatic disease to approximately 95% for those with stage I resected tumors. Corresponding rates for rectal cancers are similar to slightly inferior overall, ranging from 4 to 72%. Besides TNM stage, additional factors that are prognostic for poorer outcomes in patients undergoing potentially curative resection include signet ring histologic subtype (see preceding discussion under Pathobiology), lymphovascular and perineural invasion, absence of host lymphoid response, presence of clinical obstruction preoperatively, high preoperative serum levels of the CEA tumor marker, positive margins, high tumor grade, and microsatellite-stable disease. Differing from many solid tumors originating outside the GI tract, CRCs do not have different prognoses based on size.

Genetic factors are important as well, even in the metastatic setting. Patients with tumors harboring *BRAF* mutations appear to have a worse outcome. *KRAS* mutations were formerly thought to be prognostic but now appear to be only predictive for lack of benefit from certain types of systemic therapy.

Colorectal cancer remains a significant problem despite the fact that most cases can be prevented. Large gains have been made in terms of overall survival, but the vast majority of patients with advanced disease still succumb to their malignancy. Minimal tailoring can currently be offered to patients (e.g., selecting a chemotherapeutic agent based on toxicity or not using an anti-EGFR antibody in those with *KRAS*-mutated tumors); the dream of truly individualized therapy remains elusive but is under active study.

Grade A References

A1. Joensuu H, Eriksson M, Hall KS, et al. One versus three years of adjuvant imatinib for operable gastrointestinal stromal tumor: a randomized trial. *JAMA*. 2012;307:1265-1272.

A2. Burn J, Gerdes AM, Macrae F, et al. Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial. *Lancet*. 2011; 378:2081-2087.

A3. Quintero E, Castells A, Bujanda L, et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. *N Engl J Med*. 2012;366:697-706.

A4. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet*. 2010;375:1624-1633.

A5. Schoen RE, Pinsky PF, Weissfeld JL, et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *N Engl J Med*. 2012;366:2345-2357.

A6. Holme O, Loberg M, Kalager M, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. *JAMA*. 2014;312:606-615.

A7. Sauer R, Liersch T, Merkel S, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer: results of the German CAO/ARO/AIO-94 randomized phase III trial after a median follow-up of 11 years. *J Clin Oncol*. 2012;30:1921-1933.

A8. Grothey A, Van Cutsem E, Sobrero A, et al. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international multicenter, randomized, placebo-controlled, phase 3 trial. *Lancet*. 2013;381:303-312.

A9. Loupakis F, Cremolini C, Masi G, et al. Initial therapy with FOLFOXIRI and bevacizumab for metastatic colorectal cancer. *N Engl J Med*. 2014;371:1609-1618.

A10. Primrose JN, Perera R, Gray A, et al. Effect of 3 to 5 years of scheduled CEA and CT follow-up to detect recurrence of colorectal cancer: the FACS randomized clinical trial. *JAMA*. 2014;311: 263-270.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Samadder NJ, Curtin K, Tuohy TM, et al. Increased risk of colorectal neoplasia among family members of patients with colorectal cancer: a population-based study in Utah. *Gastroenterology*. 2014;147:814-821.
2. Herrinton LJ, Liu L, Levin TR, et al. Incidence and mortality of colorectal adenocarcinoma in persons with inflammatory bowel disease from 1998 to 2010. *Gastroenterology*. 2012;143: 382-389.
3. Win AK, Lindor NM, Young JP, et al. Risks of primary extracolonic cancers following colorectal cancer in Lynch syndrome. *J Natl Cancer Inst*. 2012;104:1363-1372.
4. Akhtar-Zaidi B, Cowper-Sal-lari R, Corradin O, et al. Epigenomic enhancer profiling defines a signature of colon cancer. *Science*. 2012;336:736-739.
5. Lieberman D. Colorectal cancer screening: practice guidelines. *Dig Dis*. 2012;30(suppl 2):34-38.
6. Qaseem A, Denberg TD, Hopkins RH Jr, et al. Screening for colorectal cancer: a guidance statement from the American College of Physicians. *Ann Intern Med*. 2012;156:378-386.
7. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. *N Engl J Med*. 2013;369:1106-1114.
8. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med*. 2014;370:1287-1297.
9. Nishihara R, Wu K, Lochhead P, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N Engl J Med*. 2013;369:1095-1105.
10. Pox CP, Altenhofen L, Brenner H, et al. Efficacy of a nationwide screening colonoscopy program for colorectal cancer. *Gastroenterology*. 2012;142:1460-1467.
11. Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med*. 2012;366:687-696.
12. Loberg M, Kalager M, Holme O, et al. Long-term colorectal-cancer mortality after adenoma removal. *N Engl J Med*. 2014;371:799-807.
13. Kaminski MF, Regula J, Kraszewska E, et al. Quality indicators for colonoscopy and the risk of interval cancer. *N Engl J Med*. 2010;362:1795-1803.
14. Baxter NN, Warren JL, Barrett MJ, et al. Association between colonoscopy and colorectal cancer mortality in a US cohort according to site of cancer and colonoscopist specialty. *J Clin Oncol*. 2012;30:2664-2669.
15. Lieberman DA, Rex DK, Winawer SJ, et al. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US multi-society taskforce on colorectal cancer. *Gastroenterology*. 2012;143:844-857.
16. Zalis ME, Blake MA, Cai W, et al. Diagnostic accuracy of laxative-free computed tomographic colonography for detection of adenomatous polyps in asymptomatic adults: a prospective evaluation. *Ann Intern Med*. 2012;156:692-702.
17. Rothwell PM, Wilson M, Elwin CE, et al. Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomized trials. *Lancet*. 2010;376:1741-1750.
18. Patel A, Sun W. Ziv-aflibercept in metastatic colorectal cancer. *Biologics*. 2014;8:13-25.
19. Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet*. 2014;383:1490-1502.
20. Beets GL, Glimelius BL. Adjuvant chemotherapy for rectal cancer still controversial. *Lancet Oncol*. 2014;15:130-131.

REVIEW QUESTIONS

1. A 55-year-old average-risk patient undergoes a screening colonoscopy. There is no family history of polyps or colorectal cancer, and the patient is asymptomatic. A single 6-mm tubular adenoma was found in a well-prepared and otherwise normal colonoscopy. The next colonoscopy should be performed at:

A. 1 year
B. 2 years
C. 3 years
D. 5-10 years
E. Never. Surveillance is not indicated.

Answer: D For average-risk patients with otherwise normal colonoscopies, the finding of a single nonadvanced tubular adenoma (<10 mm, no villous histology or high-grade dysplasia) requires a next colonoscopy to be performed at 5-10 years. (See section on "Prognosis" under "Adenomatous Polyps.")

2. A patient with colorectal cancer (CRC) is found to have microsatellite instability on testing of the resected tumor. There is a strong family history of CRC, including a parent and a sibling. Germline mutation analysis is performed and finds a mutation in one of the mismatch repair genes *(MLH1)*. This patient most likely has:

A. Hereditary nonpolyposis colorectal cancer (Lynch syndrome)
B. Familial adenomatous polyposis
C. *MUTYH*-associated polyposis
D. CpG island methylation (CIMP)
E. *BRAF* mutation

Answer: A Mutations in the mismatch repair genes (*hMSH2*, *hMSH3*, *hMSH6*, *hMLH1*, *hPMS1*, and *hPMS2*) result in errors in DNA replication. These errors can be detected as microsatellite instability in the resected tumor. Errors in these genes result in hereditary nonpolyposis colorectal cancer (HNPCC), also known as Lynch syndrome. (See section on "Hereditary Nonpolyposis Colon Cancer.")

3. The most important factor in deciding whether to offer adjuvant chemotherapy following potentially curative resection of colon cancer is:

A. Tumor size
B. Tumor grade
C. Patient age
D. Patient gender
E. Tumor stage

Answer: E Although multiple factors are prognostic and/or predictive, stage is the most important determinant in the decision regarding adjuvant systemic therapy. (See section on "Systemic and Combination Therapies" under "Treatment" of "Adenocarcinoma of the Colon and Rectum.")

4. Reasonable agents to offer a patient with resected stage III colon cancer include:

A. 5-Fluorouracil
B. Oxaliplatin
C. Irinotecan
D. A and B
E. A, B, and C

Answer: D All the listed agents have activity in metastatic disease, but irinotecan has not been proven to offer any benefit in the adjuvant setting. (See section on "Systemic and Combination Therapies" under "Treatment" of "Adenocarcinoma of the Colon and Rectum.")

5. A major difference between colonic and rectal adenocarcinoma is:

A. Treatment of metastatic disease
B. Underlying genetic makeup
C. Frequency of use of irradiation
D. Frequency of complications of an intact primary in the setting of treated metastatic disease
E. C and D

Answer: E Colonic and rectal cancers are for the most part biologically similar, including sensitivity to systemic agents. Because of anatomic concerns, treatment of higher-risk rectal cancer almost always involves irradiation, whereas that modality is seldom used in treating colon cancer. Similarly, rectal cancers have a slightly higher chance of bleeding or obstructing if not undergoing primary treatment.

194

PANCREATIC CANCER

DANIEL LAHERU

DEFINITION

Pancreatic cancer usually refers to ductal adenocarcinomas of the pancreas, because more than 90% of pancreatic tumors arise from the ductal epithelium. Other major tumors of the pancreas include endocrine malignancies (Chapter 195), carcinoid tumors (Chapter 232), lymphomas (Chapter 185), and a variety of rare sarcomas.

EPIDEMIOLOGY

Pancreatic ductal adenocarcinoma (PDAC) has one of the highest incidence-to-mortality ratios of any disease. Although it represents the tenth leading cause of cancer in the United States, it is the fourth leading cause of cancer-related deaths,[1] because the vast majority of patients will die from their disease. Annually, approximately 40,000 individuals will die in the United States from PDAC or its complications. The incidence of pancreatic cancer is slowly increasing based on the changing demographics of the U.S. population. The risk of developing PDAC increases with age, with a mean age at onset of 71 years; the risk in men and women is equivalent. The average lifetime risk for developing PDAC is about 1 in 78 for both men and women. Globally, 70% of all pancreatic cancer cases occur in people living in advanced economies, with over 270,000 deaths per year worldwide. Some PDACs occur in association with other cancers or diseases, but most do not occur in association with a defined syndrome. The overwhelming majority of PDAC cases are sporadic—that is, occurring without a history of the disease in first-degree relatives. Smoking tobacco, as well as passive exposure to tobacco smoke in the environment, contributes significantly to the development of PDAC. Occupational hazards that have been associated with an enhanced risk of developing PDAC include exposure to chlorinated hydrocarbon solvents and heavy metals.

Approximately 10% of PDACs occur in families with a history of PDAC; in these patients, the risk of developing PDAC is increased seven-fold compared with the general population. Premalignant cystic lesions of the pancreas also occur in pancreatic cancer families. In addition to the recent discovery of the *PALB2* gene in 3% of these families, mutations in *ATM* or *BRCA2*, critical partners in the DNA damage repair pathway, as well as *p16*, have also been discovered.

Chronic pancreatic inflammation from alcohol misuse or genetic anomalies significantly enhances the risk of being diagnosed with PDAC. Although the association between chronic pancreatitis and the development of PDAC has been well known for decades, only recently have studies clarified how pro-inflammatory cytokines contribute to the progression from premalignant lesion to advanced tumor. In addition to chronic pancreatitis, the role of diabetes mellitus and obesity in the development of PDAC has been emphasized. Long-standing type 1 and type 2 diabetes mellitus may represent increased risk of PDAC, but the cause-and-effect relationship between pancreatic cancer and diabetes is complex. Recently, epidemiologic studies have focused on the development of PDAC in patients with type 3c diabetes (diabetes related to pancreatic disease, or pancreatogenic diabetes), a major subset of diabetes characterized by a severe deficiency of all glucoregulatory hormones (Chapter 229). Patients with type 3c diabetes appear to have the highest associated risk of developing PDAC, especially in the setting of coexisting chronic pancreatitis. Type 3c diabetes is also a consequence of PDAC in approximately 30% of patients. The increase in obesity in the U.S. population and the concomitant increase in associated diabetes mellitus are strongly associated with an enhanced lifetime risk of developing PDAC.

PATHOBIOLOGY

Understanding the molecular characteristics of cancers of the exocrine pancreas is critical for the development of targeted therapies. Pancreatic cancer is caused by inherited (germline) and acquired (somatic) mutations in cancer-causing genes. Several oncogenes and tumor suppressor genes have been demonstrated to be involved in the development of pancreatic cancer, both by contributing to the growth of the tumor itself as well as to the surrounding microenvironment. Oncogenes, typically inactive in normal cells, cause uncontrolled cell proliferation by inhibiting apoptosis and activating the cell cycle when mutations render them constitutively active. The *KRAS* oncogene, located on chromosome 12, is the most frequently mutated oncogene in pancreatic cancer in (>90% of tumors). It encodes a membrane-bound protein that has GTP-ase activity and is involved in signal transduction.[2] When activated by mutation, typically a point mutation in codon 12, the functions of *KRAS* are independent of growth factor control, leading to chronic activation of its downstream signaling partners, *PI3K*, *MAPK*, and *RAF*, causing inhibition of apoptosis and activation of the cell cycle, migration, angiogenesis, cytoskeletal remodeling, and unchecked proliferation. Tumor suppressor genes, when functioning normally, act in the opposite manner by enhancing apoptosis and inhibiting cell proliferation. The tumor suppressor *CDKN2A/TP16*, a cell cycle control gene, is commonly inactivated in pancreatic cancer, with 80 to 95% loss of activity leading to increased cell cycle progression. *TP53* is activated by DNA damage to stop cell cycle progression and repair damaged DNA or initiate apoptosis. Mutations in this tumor suppressor gene are commonly found in 50 to 75% of pancreatic tumors. Inactivation of *SMAD4* (DPC4), involved in regulating cell cycle progression through the transforming growth factor (TGF)-β pathway, is observed in over 50% of pancreatic cancers and is associated with worse prognosis and the development of metastases. Inactivation of *RB1* (in < 10% of pancreatic cancers) and *STK11* (responsible for Peutz-Jeghers syndrome) is also observed.

There are several major signaling pathways involved in pancreatic tumorigenesis. Hedgehog (Hh) signaling, critical in embryogenesis, regulates the cell cycle and apoptosis, aids in the formation of tumor stroma, and is often upregulated and abnormal in pancreatic cancers. The NOTCH pathway, also important in normal embryogenesis to prevent terminal differentiation of cells until appropriate, can be abnormally activated in pancreatic cancer, allowing cells to remain in an undifferentiated state that contributes to tumor growth. When the Wnt pathway is activated, β-catenin is stabilized and migrates into the nucleus, where it activates its target genes. The epidermal growth factor receptor (EGFR), TGF-β, and JAK/STAT pathways have also been found to be abnormal in pancreatic cancer cells, leading to the promotion of cell growth, proliferation, differentiation, and cell survival. Epigenetic modification, the process by which gene expression is altered by mechanisms other than changes in actual DNA sequence, also has a role in pancreatic cancer tumorigenesis. Telomere shortening and overexpression of microRNAs lead to chromosomal instability and dysregulation of gene expression, respectively, and are also observed.

In addition to cancer cells themselves, the tumor microenvironment is composed of stromal cells, inflammatory cells, and endothelial cells that all play a particularly important role in pancreatic cancer growth. Cancer cells secrete growth factors, including insulin-like growth factor (IGF)-1, fibroblast growth factor (FGF), TGF-β, vascular endothelial growth factor (VEGF), and platelet-derived growth factor (PDGF), that stimulate pancreatic stellate cells (also called myofibroblasts) to secrete excess amounts of extracellular matrix. The matrix and its stromal cells, in turn, secrete cytokines and growth factors that promote cancer cell growth, invasion, and dissemination and protect pancreatic cancer cells from apoptosis, as well as generating a desmoplastic reaction that interferes with the delivery of chemotherapy to the tumor site.[3] The local presence of TGF-β also leads to decreased helper T-cell activity, which suppresses the body's immune reaction against pancreatic cancer cells.

Precursor Lesions

A major advance in understanding the development of pancreatic cancer has been the appreciation that the majority of pancreatic adenocarcinomas progress sequentially from histologically normal ductal epithelium to low-grade pancreatic intraepithelial neoplasia (PanIN), to high-grade PanIN, to invasive carcinoma. This process is associated with the accumulation of specific gene alterations (Fig. 194-1).

CLINICAL MANIFESTATIONS

Symptoms of early PDAC are often subtle and include nonspecific gastrointestinal complaints (nausea, vague abdominal pain), fatigue, and weight loss of undetermined etiology.[4] Epigastric pain and obstructive jaundice often prompt the initial diagnostic work-up of the biliary tree, but are frequently late symptoms that are associated with advanced local or regionally disseminated disease. Because approximately 75% of pancreatic carcinomas are located in the head of the pancreas, it is not unexpected that clinical presenta-

tions are often related to compression or invasion of the biliary tree or pancreatic ducts. Deep or superficial venous thrombosis (Trousseau syndrome) is not infrequent, either early or late in the presentation of PDAC (Chapter 176). Observation of a palpably distended gallbladder (from obstruction of the distal common bile duct), or Courvoisier's sign, is uncommon.

The laboratory abnormalities that accompany PDAC at presentation include anemia and elevations of serum bilirubin, alkaline phosphatase, and aminotransferases. A majority of patients eventually develop signs of obstructive jaundice as well as hyperglycemia, reflective of associated diabetes mellitus. Unfortunately, pre-neoplastic cystic lesions of the pancreas often remain asymptomatic until discovered following acute symptoms (due to ductal obstruction) that precipitate an abdominal computed tomography (CT) scan. Early PanIN lesions are asymptomatic.

DIAGNOSIS

Differential Diagnosis

The differential diagnosis of PDAC includes conditions that can present as a solid pancreatic mass, including acute (or an exacerbation of chronic) pancreatitis, ampullary or distal cholangiocarcinomas with associated biliary obstruction and jaundice, and non-neoplastic cystic pancreatic neoplasms.

Cystic neoplasms of the pancreas are frequent, detectable in 2% of abdominal magnetic resonance imaging (MRI) examinations. Whereas serous cystadenomas are benign abnormalities that do not connect to pancreatic ducts and warrant surgery only if symptomatic, mucinous cystadenomas (MCN) are precursors of PDAC that often occur in the tail of the pancreas, more frequently in women in their 40s; they are overtly malignant in 15% of patients and require surgical resection. Intraductal papillary mucinous neoplasms (IPMN) of the pancreas occur in the head of the organ, are often polycystic, and contain malignant elements in 40% of patients; all patients with this lesion require surgery.

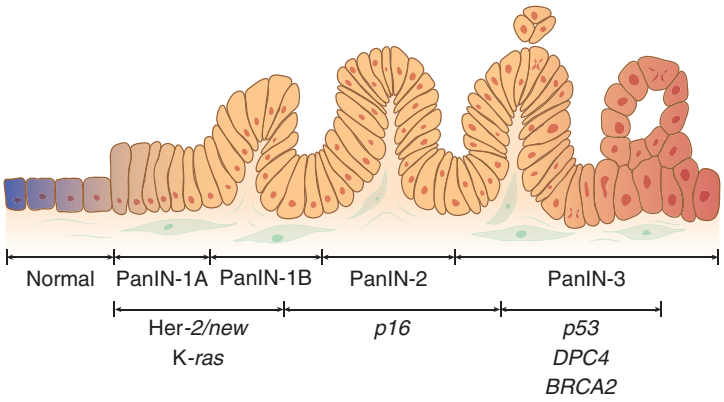

FIGURE 194-1. Genetic progression model of pancreatic adenocarcinoma. Molecular changes in the development of pancreatic ductal adenocarcinoma begin with early changes that include telomere shortening and development of mutations in *KRAS*; these are followed by loss of p16 function and expression of cyclin D1. Late changes include expression of p53 and loss of SMAD4/DPC4. PanIN = pancreatic intraepithelial neoplasia.

Imaging

CT scan with dynamic contrast and thin cuts through the pancreas is the imaging procedure of choice when PDAC is suspected. CT scans can provide information regarding the presence of metastatic disease, vascular invasion, and potential for resection. Endoscopic ultrasound (EUS) may provide additional information useful for preoperative assessment, including fine-needle aspiration biopsy under endoscopic guidance. It is important to point out that the typical desmoplastic reaction that can encase PDACs increases the possibility of false-negative biopsy findings. Patients with locally advanced or metastatic disease should have their diagnosis confirmed pathologically by fine-needle biopsy of the primary site or of a metastasis.

Biomarkers and Screening

A recent preliminary study identified two diagnostic panels based on microRNA expression in whole blood, with the potential to distinguish patients with pancreatic cancer from healthy controls.[5] However, so far, no serum- or tumor-based biomarkers or biomarker panels have been established that are both sensitive and specific enough for accurate early detection in clinical practice. CA19.9 is the most commonly used tumor biomarker for monitoring therapeutic progress in PDAC, but the lack of specificity of the assay is a concern, and CA19.9 therefore cannot be used for early detection. EUS is useful for the evaluation of cystic lesions and permits sampling of cystic fluid for genetic markers associated with the development of PDAC in cancer-prone families, as well as for cytologic studies.

TREATMENT

Resectable Pancreatic Cancer

Surgery offers the only chance to cure pancreatic cancer. Tumors are considered resectable if a clear fat plane is present around the celiac and superior mesenteric arteries, and if the mesenteric and portal veins are patent. Unfortunately, tumor will recur in the majority of patients who undergo resection. Adjuvant therapy is indicated to decrease the risk and delay the timing of locoregional and metastatic recurrence. It is typically started 1 to 2 months after surgery to allow the patient to recover from the complications associated with the underlying cancer as well as from surgery itself. Although no regimen has been proven substantially more effective than others, 6 months of adjuvant therapy with 5-fluorouracil (5-FU) or gemcitabine-based chemotherapy is an appropriate standard (Table 194-1). A meta-analysis has shown that chemotherapy with 5-FU or gemcitabine is the optimum adjuvant treatment for pancreatic adenocarcinoma and reduces mortality after surgery by about a third. Chemoradiation plus chemotherapy is less effective in prolonging survival and is more toxic than chemotherapy.

TABLE 194-1 SELECTED ADJUVANT STUDIES

STUDY	TREATMENT	1-YEAR SURVIVAL (%)	2-YEAR SURVIVAL (%)	5-YEAR SURVIVAL (%)	DISEASE-FREE SURVIVAL (MO)	MEDIAN SURVIVAL (MO)
ESPAC-1 (2004) N = 289	2 × 2 design: Observation v. Chemorad v. Chemo v. Chemorad plus chemo			 11 v. 7 v. 29 v. 13	 Chemorad, 10.7 v. no chemorad, 15.2 Chemo, 15.3 v. no chemo, 9.4	 16.9 v. 13.9 v. 21.6 v. 19.9
RTOG-9704 (2008) N = 451	Gem/XRT v. 5-FU/XRT	69 v. 65	35 v. 39	20 v. 20	NR	20.5 v. 16.9
ESPAC-3 (2010) N = 1088	Gem v. 5-FU	70 v. 70	40 v. 40	NR	14.3 v. 14.1	23.6 v. 23
Johns Hopkins and Mayo Clinic Retrospective (2010)[6] N = 1272	Observation v. chemorad	58. v. 80	34.6 v. 44.7	16.1 v. 22.3		15.5 v. 21.1 ($P < .001$)
ACOSOG (2011)[7] N = 89	IFN-cisplatin-5-FU plus XRT	80	60	NR	14.1	25.4
CONKO-1 (2013) N = 354	Observe v. gem	72 v. 72	42 v. 47	11 v. 22	6.7 v. 13.4 ($P < .001$)	10.4 v. 21.7 ($P = .06$)

5-FU = 5-flourouracil; chemo = chemotherapy; chemorad = chemoradiation therapy; gem = gemcitabine; IFN = interferon; NR = not reported; XRT = radiation therapy.

TABLE 194-2 SELECTED METASTATIC PANCREATIC CANCER STUDIES

STUDY	CHEMOTHERAPY	RESPONSE RATE (% PARTIAL RESPONSE)	MEDIAN SURVIVAL (MO)	1-YEAR SURVIVAL (%)
Burris (1997) N = 126	5-FU v. gem	NR	4.4 v. 5.6 ($P = .0025$)	2 v. 18
Cunningham (2009) N = 533	Gem ± capecitabine	14 v. 7 ($P = .008$)	7.4 v. 6 ($P < .05$)	26 v. 19
Conroy (2011) N = 343	Gem v. FOLFIRINOX	9 v. 32	6.8 v. 11.1 ($P < .0001$)	17 v. 36
Van Hoff (2013) N = 861	Gem ± nab-paclitaxel	23 % v. 7%	8.5 v. 6.7 ($P < .001$)	35 v. 22 ($P < .001$)

5-FU = 5-fluorocil; FOLFIRINOX = combination chemotherapy with oxaliplatin, irinotecan, fluorouracil, and leucovorin; gem = gemcitabine; NR = rot reported.

Neoadjuvant (presurgical) therapy remains an option in the treatment of pancreatic cancer. It has the potential to downstage patients of borderline resectability who achieve a partial response with therapy, allowing them to undergo surgery with a higher likelihood of complete resection. In fact, 15 to 40% of patients who initially present with borderline or unresectable tumors may eventually be deemed appropriate to undergo surgery. In addition, neoadjuvant therapy spares some patients the risks and stress of a complex surgical procedure, because rapidly developing metastatic disease may be detected by routine restaging following the completion of neoadjuvant treatment. Unfortunately, chemotherapy-resistant cancer can be demonstrated in 15 to 35% of patients initially considered for surgery in this setting. The use of neoadjuvant treatment also guarantees that almost all patients will receive some form of chemotherapy and/or chemoradiation because they do not have any postoperative complications from which to recover. Studies show that 73 to 100% of patients are able to complete the majority of their neoadjuvant regimens. The chemoradiation component of neoadjuvant therapy also decreases local recurrence rates in patients who undergo surgery.

However, surgery is required for cure, and neoadjuvant therapy does delay potentially curable surgery. Although the same percentage of patients ultimately undergo surgery, there are no randomized trials favoring neoadjuvant over adjuvant treatment. Therefore, at this time neoadjuvant therapy is usually reserved for borderline resectable patients, whereas resectable patients are taken to surgery immediately, with adjuvant therapy administered following recovery. Patients treated with either adjuvant or neoadjuvant therapy have similar survival rates when resection can be completed successfully. Decisions regarding initial treatment should, if possible, be made in a multidisciplinary manner to achieve the most timely and coordinated therapy.

Metastatic Disease

For many years, 5-FU was the standard of care to treat advanced PDAC. In 1997, a phase III trial demonstrated a survival benefit for gemcitabine over bolus 5-FU, with a median survival of 5.65 months compared to 4.41 months, and 1-year survival of 18% versus 2%, favoring treatment with gemcitabine. Gemcitabine also had a superior clinical benefit response, described as improvement in pain, performance status, or weight in 24% of patients versus 5% in the 5-FU group; it is therefore appropriate treatment for patients with moderate performance status.

Gemcitabine has also been combined with targeted therapies; however, addition of the EGFR inhibitor erlotinib has been the only molecularly targeted agent shown to improve overall survival, albeit modestly. Newer combination regimens, both with improved efficacy based on phase III clinical trials for metastatic PDAC patients, include FOLFIRINOX (oxaliplatin, irinotecan, fluorouracil, and leucovorin) and the nab-paclitaxel and gemcitabine doublet (Table 194-2).

PROGNOSIS AND SUPPORTIVE CARE

The overall 5-year survival for all patients with pancreatic cancer is less than 5%. This is in part because there is no appropriate screening test for the general population, and presenting symptoms are vague. Once patients are diagnosed, only 15 to 20% present with resectable, and thus potentially curable, disease. In these patients, the somatostatin analogue pasireotide (900 mg subcutaneously twice daily for 7 days beginning in the morning of surgery) can reduce the risk of fistula leak with abscess by 50%. However, even in patients with early-stage disease, median survival is 20 to 24 months, with a 5-year survival of only 15 to 20% because the majority will eventually recur despite surgery and adjuvant or neoadjuvant therapy. Patients with locally advanced (≈25 to 30% at presentation) and metastatic disease (≈50 to 60% at presentation) have median survivals of 8 to 14 months and 4 to 6 months, respectively. Because many patients suffer from biliary obstruction, diarrhea, pain, and malnutrition, palliative care can provide great benefit. Surgery can also play an important palliative role; patients found to be unresectable may be improved symptomatically by a biliary or gastric bypass procedure, depending on the site of obstruction. Placement of both biliary and duodenal stents in the absence of a surgical approach may also improve pruritus, pain, or other complications of biliary tract obstruction. Palliative care for patients with PDAC necessitates scrupulous attention to pain management that often requires a multidisciplinary approach, as well as maintenance of hydration and adequate nutritional status.

Grade A References

A1. Neoptolemos JP, Stocken DD, Friess H, et al. A randomized trial of chemoradiotherapy and chemotherapy after resetion of pancreatic cancer. *N Engl J Med*. 2004;350:1200-1210.

A2. Regine WF, Winter KA, Abrams RA, et al. Fluorouracil vs gemcitabine chemotherapy before and after fluorouracil-based chemoradiation following resection of pancreatic adenocarcinoma: a randomized controlled trial. *JAMA*. 2008;299:1019-1026.

A3. Neoptolemos JP, Stocken DD, Bassi C, et al. Adjuvant chemotherapy with fluorouracil plus folinic acid vs. gemcitabine following pancreatic cancer resection: a randomized controlled trial. *JAMA*. 2010;304:1073-1081.

A4. Oettle H, Neuhaus P, Hochhaus JT, et al. Adjuvant chemotherapy with gemcitabine and long-term outcomes among patients with resected pancreatic cancer: the CONKO-001 randomized trial. *JAMA*. 2013;310:1473-1481.

A5. Liao WC, Chien KL, Lin YL, et al. Adjuvant treatments for resected pancreatic adenocarcinoma: a systematic review and network meta-analysis. *Lancet Oncol*. 2013;14:1095-1103.

A6. Conroy T, Desseigne F, Ychou M, et al. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *N Engl J Med*. 2011;364:1817-1825.

A7. Von Hoff DD, Ervin T, Arena FP, et al. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *N Eng J Med*. 2013;369:1691-1703.

A8. Burris HA 3rd, Moore MJ, Andersen J, et al. Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. *J Clin Oncol*. 1997;15:2403-2413.

A9. Cunningham D, Chau I, Stocken DD, et al. Phase III randomized comparison of gemcitabine plus capecitabine in patients with advanced pancreatic cancer. *J Clin Oncol*. 2009;27:5513-5518.

A10. Allen PJ, Gönen M, Brennan MF, et al. Pasireotide for postoperative pancreatic fistula. *N Engl J Med*. 2014;370:2014-2022.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Wolfgang CL, Herman JM, Laheru DA, et al. Recent progress in pancreatic cancer. *Cancer J Clin*. 2013;63:318-348.
2. di Magliano MP, Logdson CD. Roles for Kras in pancreatic cancer tumor development and progression. *Gastroenterology*. 2013;144:1220-1229.
3. Feig C, Gopinathan A, Neese A, et al. The pancreas cancer microenvironment. *Clin Cancer Res*. 2012;18:4266-4276.
4. Ryan DP, Hong TS, Bardeesy N. Pancreatic adenocarcinoma. *N Engl J Med*. 2014;371:2140-2141.
5. Schultz NA, Dehlendorff C, Jensen BV, et al. MicroRNA biomarkers in whole blood for detection of pancreatic cancer. *JAMA*. 2014;311:392-404.
6. Hsu CC, Herman JM, Corsini MM, et al. Adjuvant chemoradiation for pancreatic adenocarcinoma: the Johns Hopkins Hospital–Mayo Clinic collaborative study. *Ann Surg Oncol*. 2010;17:981-990.
7. Picozzi VJ, Abrams RA, Decker PA, et al. Multicenter phase II trial of adjuvant therapy for resected pancreatic cancer using cisplatin, 5-fluorouracil, and interferon-alfa-2b-based chemoradiation: ACOSOG Trial Z05031. *Ann Oncol*. 2011;22:348-354.

REVIEW QUESTIONS

1. Which of the following is *not* a known risk factor for pancreatic cancer:

A. Chronic proton pump inhibitor therapy
B. Positive family history
C. Cigarette smoking
D. Obesity
E. Chronic pancreatitis

Answer: A Gastrectomy, much more commonly performed in the past for refractory peptic ulcer disease, was considered to be a risk factor for pancreatic cancer; however, there is no evidence that medical therapy for it (e.g., with proton pump inhibitors) poses a risk. Approximately 10% of pancreatic cancer cases occur in families with a positive history. Cigarette smoking, even second-hand, has been clearly linked to pancreatic cancer risk. Obesity is associated with increased risk for a variety of cancers, including pancreatic. The long-known risk of pancreatic cancer with chronic pancreatic inflammation is now being elucidated at the molecular level as a function of the action of pro-inflammatory cytokines on progression from pre-malignant lesions to advanced tumor. (See section on "Epidemiology.")

2. Which of the following is *not* known to be a clinical manifestation of pancreatic cancer:

A. Diabetes mellitus
B. Courvoisier's sign
C. Superficial thrombophlebitis
D. Splenic infarction
E. Obstructive jaundice

Answer: D The symptoms of early pancreatic cancer are often subtle, and there may be no abnormal findings on physical exam. Although extensive pancreatic cancer could potentially cause portal vein or splenic vein thrombosis, compromise of the spleen's arterial supply to cause splenic infarction is not characteristic. Hyperglycemia and even frank clinical type 2 diabetes mellitus may be a harbinger of occult pancreatic cancer, and its onset may antedate the cancer diagnosis by months. Courvoisier's sign, a palpably distended gallbladder, is a striking physical finding resulting from distal common bile duct obstruction due to pancreatic cancer, although it is very rare. Obstructive jaundice that is often painless is, however, not an uncommon presenting manifestation because 75% of pancreatic cancers are located in the head of the pancreas. Superficial thrombophlebitis, particularly when it is migratory in nature (called Trousseau syndrome) is frequently a sign of occult cancer, with pancreatic cancer being among the most common.

3. Which of the following approaches provides the best results in increasing survival after surgery for pancreatic cancer:

A. Radiation therapy
B. Combined chemoradiation therapy and chemotherapy
C. Gemcitabine or 5-fluorouracil
D. Biologic agents
E. Unfortunately, no current therapy increases survival after surgery.

Answer: C A recently published meta-analysis of several clinical trials has shown that chemotherapy with 5-fluoruracil or gemcitabine is the currently optimum adjuvant treatment for pancreatic adenocarcinoma and reduces mortality after surgery by about a third. Adjuvant radiation therapy alone or combined chemoradiation and chemotherapy are less effective in prolonging survival and are more toxic than chemotherapy with 5-FU or gemcitabine. Biologic agents are not known to be effective at this time.

195

PANCREATIC NEUROENDOCRINE TUMORS

ROBERT T. JENSEN

DEFINITION

Pancreatic neuroendocrine tumors (pNETs) are also called islet cell tumors, but because the cell of origin of most of these tumors is unknown, the general term *pNET* is preferred. This term is also a misnomer, however, because pNETs can occur outside the pancreas. Eleven pNETs are well established (Table 195-1).[1] Other functional pNET syndromes have been rarely reported: pNETs secreting renin causing hypertension, pNETs secreting erythropoietin

TABLE 195-1 PANCREATIC NEUROENDOCRINE TUMORS

NAME OF TUMOR	NAME OF SYNDROME	MAIN SIGNS OR SYMPTOMS	LOCATION	MALIGNANCY (%)	HORMONE CAUSING SYNDROME
I. FUNCTIONAL pNET					
Gastrinoma	Zollinger-Ellison syndrome	Abdominal pain, diarrhea, esophageal symptoms	Pancreas—30% Duodenum—60% Other—10%	60-90	Gastrin
Insulinoma	Insulinoma	Hypoglycemic symptoms	Pancreas—100%	5-15	Insulin
Glucagonoma	Glucagonoma	Dermatitis, diabetes/glucose intolerance, weight loss	Pancreas—100%	60	Glucagon
VIPoma	Verner-Morrison, pancreatic cholera, WDHA	Severe watery diarrhea, hypokalemia	Pancreas—90% Other—10% (neural, adrenal, periganglionic tissue)	80	Vasoactive intestinal peptide (VIP)
Somatostatinoma	Somatostatinoma	Diabetes mellitus, cholelithiasis, diarrhea	Pancreas—56% Duodenum/jejunum—44%	60	Somatostatin
GRFoma	GRFoma	Acromegaly	Pancreas—30% Lung—54% Jejunum—7% Other—13% (adrenal, foregut, retroperitoneum)	30	Growth hormone–releasing factor (GRF)
ACTHoma	ACTHoma	Cushing's syndrome	Pancreas—4-16% of all ectopic Cushing's cases	>95	Adrenocorticotropic hormone (ACTH)
pNET causing carcinoid syndrome	pNET causing carcinoid syndrome	Diarrhea, flushing	Pancreas—<1% of all carcinoids	60-90	Serotonin, tachykinins
pNET causing hypercalcemia	pNET causing hypercalcemia	Signs/symptoms of hypercalcemia	Pancreas (rare cause of hypercalcemia)	>85	PTHrP, other unknown
CCKoma	CCKoma	Diarrhea, peptic ulcer, gallbladder disease,	Pancreas (1 case)	100	Cholecystokinin
II. NONFUNCTIONAL pNET					
Nonfunctioning PPoma	Nonfunctional PPoma	Weight loss, abdominal mass, hepatomegaly	Pancreas—100%	60-90	None: pancreatic polypeptide, chromogranin released, but no known symptoms due to hypersecretion

pNET = pancreatic endocrine tumor; PP = pancreatic polypeptide; PTHrP = parathormone-related peptide; WDHA = watery diarrhea, hypokalemia, and achlorhydria.

causing polycythemia, pNETs secreting luteinizing hormone that causes virilization, pNETs secreting insulin-like growth factor (IGF)-II or glucagon-like peptide (GLP)-1 causing hypoglycemia, and pNETs secreting enteroglucagon causing small intestinal hypertrophy1. In addition, pNETs synthesizing neurotensin, calcitonin, and ghrelin have been reported, but no distinct syndromes related to them have been generally accepted).

pNETs frequently are classified as functional or nonfunctional (see Table 195-1), depending on whether a clinical syndrome resulting from the autonomously released hormone is present).[2,3] Nonfunctional pNETs frequently release hormones and peptides (pancreatic polypeptide, neurotensin, α- and β-subunits of human chorionic gonadotropin, neuron-specific enolase, chromogranin A and breakdown products) that cause no distinct clinical syndromes.

EPIDEMIOLOGY

pNETs are uncommon, having a prevalence of less than 10 cases per 1 million population. Insulinomas, gastrinomas, and nonfunctional pNETs are the most common, with an incidence of 1 to 3 new cases per 1 million population.[4]

PATHOBIOLOGY

All pNETs share certain features. pNETs are classified as APUDomas (amine precursor uptake and decarboxylation), which share cytochemical features with carcinoid tumors, melanomas, and other endocrine tumors (pheochromocytomas, medullary thyroid cancer).[5] Except for insulinomas, these tumors are frequently malignant. All pNETs appear similar histologically, with few mitotic figures. Ultrastructurally, they have dense granules containing peptides, amines, and products of neuroendocrine differentiation (neuron-specific enolase, chromogranins, synaptophysin). The presence of chromogranin immunoreactivity in the tumor is now widely used to identify these tumors as endocrine tumors.

Molecular studies show that pNETs have a different pathogenesis than common gastrointestinal adenocarcinomas because they infrequently demonstrate mutations in common tumor suppressor genes (e.g., retinoblastoma gene, *p53*) or common oncogenes (*ras*, *c-Jun*, *c-Fos*).[6] Recent studies show important alterations are found in the *MEN1* gene, gene alterations affecting p53 and retinoblastoma activity, mutations in the DAXX-ATRX complex (important for transcription/chromatin remodeling), and the mTOR pathway.[7] Alterations in $p16^{INK4a}$, the *MEN1* gene, and the expression of growth factors, as well as chromosomal losses (1q, 3p, 3q, 6p, X) and gains (17p, 17q, 20q), have been associated with a worse prognosis in numerous studies.[8] A number of other factors have prognostic significance, the most important of which is the presence of liver metastases. Recently, classification systems including a TNM classification and a grading system (World Health Organization, ENETS [European Neuroendocrine Tumor Society], AJCC/UICC [American Joint Committee on Cancer/Union for International Cancer Control]) have been proposed for pNETs based on tumor size, presence of metastases, invasiveness, and proliferative indices; these classification systems have been recently shown to have prognostic value in a number of studies. Furthermore, they are important in choosing the proper treatment approach.[9]

Four autosomal dominant inherited disorders are associated with an increased occurrence of pNETs: multiple endocrine neoplasia type 1 (MEN 1; 80 to 100% develop pNETs), von Hippel-Lindau disease (VHL; 10 to 17% have pNETs), von Recklinghausen disease (neurofibromatosis [NF]-1; 12% develop duodenal somatostatinomas), and tuberous sclerosis (<1% develop pNETs).[10]

FUNCTIONAL PANCREATIC NEUROENDOCRINE TUMOR SYNDROMES

Zollinger-Ellison Syndrome (Gastrinomas)

DEFINITION AND EPIDEMIOLOGY

Zollinger-Ellison syndrome (ZES) is a clinical syndrome caused by a gastrin-releasing endocrine tumor usually located in the pancreas or duodenum and characterized by clinical symptoms and signs resulting from gastric acid hypersecretion (peptic ulcer disease, diarrhea, esophageal reflux disease).

ZES is diagnosed most frequently between the ages of 35 and 65 years and is slightly more common in men (60%).[11]

PATHOBIOLOGY

In recent surgical series, gastrinomas occur two to five times more frequently in the duodenum than in the pancreas. Duodenal gastrinomas are generally small (<1 cm), whereas pancreatic gastrinomas are generally larger. Occasionally ZES results from a gastrinoma in the splenic hilum, mesentery, stomach, or only in a lymph node or an ovary. Extrapancreatic gastrinomas producing ZES have been reported in the heart and as a result of small-cell lung cancer. As with other pNETs, malignancy can be reliably determined only by demonstrating the presence of metastatic disease, and no light microscopic or ultrastructural finding can clearly establish malignant behavior.

Gastrin stimulates parietal cells to secrete acid and also has a growth effect on cells of the gastric mucosa. Chronic hypergastrinemia thus leads to increased gastric mucosal thickness, prominent gastric folds, and increased numbers of parietal cells and gastric enterochromaffin-like cells. Patients with gastrinomas have increased basal and maximal acid outputs. *Helicobacter pylori* appears not to be important in the pathogenesis of the ulcer disease in ZES, in contrast to that of routine peptic ulcers (Chapter 139). Diarrhea is common because the large-volume gastric acid output leads to structural damage to the small intestine (inflammation, blunted villi, edema), interference with fat transport, inactivation of pancreatic lipase, and precipitation of bile acids. These same mechanisms, if prolonged, can lead to steatorrhea. If acid hypersecretion is controlled, the diarrhea will stop.

Twenty to 25% of ZES patients have MEN 1(MEN1/ZES) (Chapter 231). These patients have hyperplasia or tumors of multiple endocrine glands (parathyroid hyperplasia [>90%], pituitary tumors [60%], and pNETs [80 to 100%]). In 80 to 95% the gastrinomas are in the duodenum, frequently small (<0.5 cm), almost always multiple, and in 40 to 60% associated with lymph node metastases.

CLINICAL MANIFESTATIONS

Abdominal pain resulting from a peptic ulcer is the most common symptom (>80%). Most ulcers occur in the duodenum (>85%), but they occasionally occur in the postbulbar area, jejunum, or stomach, or in multiple locations. Initially, the pain is usually similar to that of patients with typical peptic ulcers (Chapter 139). With time, the symptoms become persistent and, in general, respond poorly to treatments aimed at eliminating *H. pylori* and to conventional doses of histamine-2 receptor antagonists, as well as to the now rarely used surgical treatments for ulcer disease. By comparison, conventional doses of proton pump inhibitors (PPIs) (e.g., omeprazole, lansoprazole, pantoprazole, esomeprazole, rabeprazole) can mask the symptoms of most patients with ZES and can also cause hypergastrinemia as seen in ZES. The widespread use of PPIs has delayed the diagnosis of ZES.[12]

Heartburn is also common (20%). Diarrhea (60 to 70%) occurs frequently and may be the first symptom (10 to 20%). In MEN1, ZES is the most common functional pNET syndrome (54%), although patients typically first develop renal stones related to hypercalcemia from hyperparathyroidism or have elevated prolactin levels resulting from pituitary tumors and only later develop ZES. However, studies show that 20 to 40% of patients with MEN1/ZES initially present with ZES symptoms.

In ZES patients, almost all the symptoms result from the effects of gastric acid hypersecretion, but late in the disease, patients can have tumor-related symptoms. Approximately one third of patients have metastatic liver disease at presentation, but less than 20% of other patients develop metastatic disease to the liver during a 10-year follow-up period.

Up to 5% of patients with ZES develop Cushing syndrome (Chapter 227) as a result of adrenocorticotropic hormone (ACTH) secretion by the gastrinoma. These patients usually have a metastatic gastrinoma in the liver, ZES without MEN 1, and a poor prognosis.

DIAGNOSIS

ZES should be suspected in any patient whose peptic ulcer disease is accompanied by diarrhea, is recurrent, does not heal with treatment, is not associated with *H. pylori* infection, is associated with a complication (bleeding, obstruction, esophageal stricture), is multiple or occurs in unusual locations, or is associated with a pancreatic tumor. ZES should also be suspected in patients with chronic secretory diarrhea (Chapter 140), peptic ulcer disease associated with large gastric folds, a family or personal history of renal stones or endocrinopathies, or the laboratory finding of hypercalcemia, hypergastrinemia, or gastric acid hypersecretion.

When suspected, the initial test is a fasting serum gastrin level, which is elevated in 99 to 100% of ZES patients. Recent studies report up to 60% of commercial gastrin assays are unreliable (over/underestimate true value), so a reliable assay should be used. Besides ZES, other causes of fasting hypergastrinemia include renal failure, *H. pylori* infections, and a physiologic response to achlorhydria or hypochlorhydria because of pernicious anemia, atrophic gastritis, or the use of PPIs. If the serum gastrin level is elevated, the fasting gastric pH should be determined. If the serum gastrin is more than 1000 pg/mL (normal, <100) and the gastric pH is less than 2.0, the patient almost certainly has ZES; approximately 40% of patients have this combination. If the gastrin is increased less than 10-fold and the gastric pH is higher than 2.0, basal acid output and a secretin test should be performed. Basal acid output is increased in patients with ZES, and more than 95% have a value greater than 15 mEq/hour if no previous gastric acid–reducing surgery has been performed. Because of their long duration of action, PPIs must be stopped for at least 1 week, if possible, to ensure that the cause of the hypergastrinemia is not the drug itself. Stopping a PPI in an undiagnosed ZES patient can lead to complications, so it needs to be done with care, and it is best to consult a group well versed in making the diagnosis.

Differential Diagnosis

A secretin test can exclude *H. pylori* infection, retained gastric antrum syndrome, antral G-cell hyperfunction or hyperplasia, chronic renal failure, and gastric outlet obstruction that may mimic ZES. Physiologically normal individuals show a less than 120 pg/mL increase in the gastrin level after intravenous secretin, whereas 94% of ZES patients with a fasting gastrin level that is elevated less than 10-fold above normal have a positive test. No false-positive results are reported except in patients with achlorhydria. In all patients with ZES, evaluation must exclude MEN 1 syndrome by searching for other endocrinopathies and assessing family history.

Imaging and Endoscopy

All patients should have imaging studies to localize the tumor. A cross-sectional imaging study such as triphasic computed tomography (CT) or magnetic resonance imaging (MRI) with contrast is usually the initial study because of their widespread availability.[13] Somatostatin receptor scintigraphy using single-photon emission CT (SPECT) after injection of indium-111–[diethylenetriamine pentaacetic acid-D-phenylalanine-1] octreotide is the most sensitive modality; it identifies 60% of primary gastrinomas and more than 90% of patients with metastatic liver disease, with a sensitivity equal to all conventional imaging studies (MRI, CT, ultrasound, angiography) combined.[14] For pancreatic gastrinomas, endoscopic ultrasound is particularly sensitive. Small duodenal gastrinomas (<1 cm) are frequently not detected by an imaging modality but can be found at surgery if routine duodenotomy is performed. Recent studies show that two new imaging techniques may be useful for small gastrinomas and other pNETs: the use of hybrid scanning with CT or MRI and somatostatin receptor scintigraphy (SRS) and the use of positron emission tomographic scanning, especially with gallium-68-labeled somatostatin analogues.

TREATMENT Rx

Medical Therapy

Patients need medical therapy directed at controlling the gastric acid hypersecretion and, if possible, surgical therapy to remove the gastrinoma. PPIs are now the drugs of choice. Because of their long duration of action, acid hypersecretion can be controlled in all patients with once- or twice-daily doses. The recommended starting dose for omeprazole is 60 mg once a day. In 30% of patients, higher doses are needed, particularly in patients with complicated disease (MEN 1), previous gastric surgery, or a history of severe esophageal reflux. With time the omeprazole dose can be reduced in most patients with uncomplicated disease to 20 to 40 mg/day. Patients must be treated indefinitely unless surgically cured. Long-term therapy is safe, and patients have been treated for up to 20 years with omeprazole without loss of efficacy, although reduced vitamin B_{12} levels may occur and require vitamin B_{12} supplementation (Chapter 164). Histamine-2 receptor antagonists are also effective, but frequent (every 4 to 6 hours) high doses are needed. Total gastrectomy, the historical treatment, is now performed only for patients who cannot or will not take oral antisecretory medications.[15] Selective vagotomy reduces acid secretion, but many patients continue to require a low dose of drug, and it is now rarely used. Parathyroidectomy should be performed in MEN 1 patients with hyperparathyroidism and ZES, because it reduces acid secretion and increases the sensitivity to antisecretory drugs.

Surgical Therapy

Surgical exploration for cure is recommended in all patients without unresectable liver metastases, MEN 1, or complicating medical conditions that limit life expectancy. Tumors are found by experienced endocrine surgeons in 95% of patients at surgery. Surgical resection decreases the metastatic rate, increases survival, and results in a 5-year cure rate of 30%. Patients with metastatic gastrinoma in the liver have a poor prognosis, with a 5-year survival rate of 30%.

Metastatic Disease

If the metastatic disease can be resected (>90%), surgery should be considered (5 to 15% of cases). If the metastatic disease is nonresectable and slowly increasing in size or is symptomatic, treatment with octreotide (100 to 450 µg two to three times daily) alone or in combination with interferon-α (1 to 5 million U, 3 to 7 days/week) is effective in inhibiting further tumor growth in 50 to 60% of patients. In a randomized multinational study, the somatostatin analog lanreotide was associated with significantly prolonged survival among patients with somatostatin-positive metastatic enteropancreatic neuroendocrine tumors of grade 1 or 2. If this treatment fails or if the tumor is rapidly growing, chemotherapeutic agents (streptozotocin, 5-fluorouracil, doxorubicin) or treatment with the mTOR inhibitor everolimus or the tyrosine kinase inhibitor sunitinib is recommended. Everolimus and sunitinib also have been shown in prospective randomized studies to more than double the progression-free survival time. For patients with extensive metastatic disease, somatostatin receptor–directed radiation therapy using analogues labeled with yttrium-90, lutetium-177, or indium-111 is increasingly used, but these therapies are still not approved.[16] Newer chemotherapy treatments using temozolomide show some promise in a small number of gastrinomas and other pNET patients. In advanced cases where metastases are confined to the liver, liver-directed therapies (embolization, chemoembolization, radioembolization) and local ablative methodologies such as radio frequency ablation, are increasingly used. Liver transplantation is occasionally performed in the rare patient with metastases limited to the liver.

PROGNOSIS

Approximately 25% of gastrinomas show aggressive growth. The most important prognostic predictor is the development of liver metastases. The presence of a large primary tumor, a pancreatic tumor, bone metastases, development of ectopic Cushing syndrome, or a high fasting gastrin level is associated with aggressive growth.

Glucagonomas

DEFINITION

Glucagonomas are endocrine tumors of the pancreas that ectopically secrete glucagon, causing a specific syndrome (see Table 195-1).

PATHOBIOLOGY

Glucagon hypersecretion explains the glucose intolerance. The exact origin of the rash is unclear; some studies report that prolonged glucagon infusions can cause the characteristic skin lesions. A role for possible zinc deficiency has been proposed because of the similarity of the rash to that seen with zinc deficiency (acrodermatitis enteropathica) and because the rash improves in some patients who are given zinc. The hypoaminoacidemia is thought to be secondary to the effect of glucagon on amino acid metabolism by altering gluconeogenesis. The wasting and weight loss are intrinsic parts of the glucagonoma syndrome, and recent studies suggest that a novel anorectic substance distinct from glucagon is responsible.

CLINICAL MANIFESTATIONS

The cardinal clinical features are a distinct dermatitis (necrolytic migratory erythema, seen in 70 to 90%) (Fig. 195-1), diabetes mellitus and glucose intolerance (40 to 90%), weight loss (70 to 96%), anemia (30 to 85%), hypoaminoacidemia (80 to 90%) with deficiencies of essential fatty acids, thromboembolism (10 to 25%), diarrhea (15 to 30%), and psychiatric disturbances (0 to 20%). The characteristic rash is usually found at intertriginous and periorificial sites, especially in the groin and buttocks (see Fig. 195-1). It is initially erythematous, becomes raised, and develops central bullae whose tops detach, with the eroded areas becoming crusty. Healing occurs with hyperpigmentation.

DIAGNOSIS

The diagnosis is established by demonstrating elevated plasma glucagon levels. Normal levels are 150 to 200 pg/mL; in patients with glucagonomas, levels usually (>90%) are higher than 1000 pg/mL. However, in some recent studies, up to 40% of patients had plasma glucagon values of 500 to 1000 pg/mL. Increased plasma glucagon levels also occur in renal insufficiency, acute pancreatitis, hypercorticism, hepatic diseases, celiac disease, severe stress and prolonged fasting, in patients treated with danazol, and in familial hyperglucagonemia. In these conditions, the level is usually less than 500 pg/mL except in patients with hepatic diseases or in those with familial hyperglucagonemia. Recently, two new glucagonoma-related syndromes have been described: Mahvash disease (characterized by mutations in the glucagon receptor, glucagon cell hyperplasia, hyperglucagonemia, but no symptoms of the glucagonoma syndrome) and glucagon cell adenomatosis (characterized by glucagon cell hyperplasia and occasional symptoms mimicking the glucagonoma syndrome).

Glucagonomas are generally large when discovered (mean, 5 to 10 cm), and they most frequently occur in the pancreatic tail (>50%). Liver metastases are commonly present at the time of diagnosis (45 to 80%).

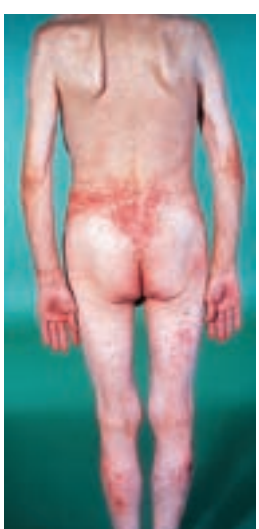

FIGURE 195-1. **A patient with a glucagonoma with the characteristic rash (necrolytic migratory erythema). The rash is usually at intertriginous areas or periorificial sites and shows various stages of erythema, blistering, and crusting. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine*. 3rd ed. London: Mosby; 2003.)**

TREATMENT Rx

Subcutaneous administration of the synthetic long-acting somatostatin analogue octreotide (100 to 400 µg two to three times daily) controls the rash in 80% of patients and improves weight loss, diarrhea, and hypoaminoacidemia,[17] but it usually does not improve the diabetes mellitus. Increasingly, long-acting depot formulations of octreotide (octreotide-LAR) or lanreotide autogel are being given by monthly injection. Zinc supplementation and infusions of amino acids or fatty acids, or both, can diminish the severity of the rash. After tumor localization, surgical resection is preferred; even debulking of metastatic tumor may be of benefit. For advanced disease, treatment is similar to outlined for advanced nonresectable gastrinomas.

PROGNOSIS

The prognosis is now largely determined by the growth of the tumor per se, because the symptoms of glucagon excess can be largely controlled by somatostatin analogues. This is particularly true with glucagonomas; in many series, more than 50 to 80% are metastatic at presentation, and patients usually present late with large primary tumors. The mean 5-year survival rate is 50%; however, extended survivals (>15 years) are reported in some patients with treatment with somatostatin analogues and other tumor-directed therapies.

VIPomas

DEFINITION

The VIPoma syndrome, also called the Verner-Morrison syndrome, pancreatic cholera, and the WDHA syndrome (for watery diarrhea, hypokalemia, and achlorhydria), results from an endocrine tumor, usually in the pancreas, that ectopically secretes vasoactive intestinal polypeptide (VIP).

EPIDEMIOLOGY AND PATHOBIOLOGY

VIPomas in adults are found in the pancreas in 80 to 90% of cases; rare cases result from intestinal carcinoids, ganglioneuromas, ganglioneuroblastomas, and pheochromocytomas. In children younger than 10 years and rarely in adults (<5%), the VIPoma syndrome is caused by ganglioneuromas or ganglioneuroblastomas at extrapancreatic sites. VIPomas are usually large and solitary; 50 to 75% of these tumors occur in the pancreatic tail, and 40 to 70% have metastasized at diagnosis. VIPomas frequently secrete both VIP and peptide histidine methionine, but VIP is responsible for the symptoms. VIP is a potent stimulant of secretion in the small and large intestine, which causes the cardinal clinical features of the VIPoma syndrome. VIP also causes relaxation of gastrointestinal smooth muscle, and this may contribute to the dilated loops of bowel that are common in this syndrome, as well as a dilated atonic gallbladder that is sometimes seen. Hypochlorhydria is thought to result from the inhibitory effect of VIP on acid secretion, the flushing is related to the vasodilatory effects of VIP, and the hyperglycemia is caused by the glycogenolytic effect of VIP. The mechanism of the hypercalcemia remains unclear.

CLINICAL MANIFESTATIONS

The cardinal clinical feature is severe, large-volume, watery diarrhea (>1 L/day) (100%), which is secretory and occurs during fasting. Hypokalemia (67 to 100%) and dehydration (83%) commonly occur because of the volume of the diarrhea. Achlorhydria is occasionally noted, but hypochlorhydria is usually found (34 to 72% of cases). Flushing occurs in 20% of patients, hyperglycemia in 25 to 50%, and hypercalcemia in 41 to 50%. Steatorrhea is uncommon (16%) despite the volume of diarrhea.

DIAGNOSIS

The diarrhea of VIPomas characteristically persists during fasting and is large in volume (>3 L/day in 70 to 80%); the diagnosis is excluded when fasting stool volume is less than 700 mL/day. To differentiate VIPomas from other causes of large-volume fasting diarrhea, fasting plasma VIP levels should be determined. The normal value in most laboratories is less than 190 pg/mL, and elevated levels are present in 90 to 100% of patients in various series. The differential diagnosis of large-volume fasting diarrhea (>700 mL/day) includes ZES, diffuse islet cell hyperplasia, surreptitious use of laxatives, the pseudopancreatic cholera syndrome, and rarely, HIV infection (Chapter 390). Serum gastrin levels identify patients with ZES, and plasma VIP levels are normal in most patients who abuse laxatives, in 82% of patients with pancreatic islet cell hyperplasia, and in patients with HIV-induced secretory diarrhea.

TREATMENT Rx

The symptoms caused by the VIP are controlled initially in more than 85% of patients by daily doses of octreotide (50 to 400 µg once to three times daily) or by monthly injections of the depot form, octreotide-LAR or lanreotide autogel, but increased doses may be needed over time. Before the availability of octreotide, small numbers of patients were reported to respond to a variety of agents, including high-dose prednisone (60 to 100 mg/day; 40 to 50%), clonidine, lithium carbonate, indomethacin, loperamide, metoclopramide, and phenothiazines. After tumor localization studies, surgical resection should be attempted if it is possible to remove all visible tumor; however, more than 50% of patients have generalized liver metastases at diagnosis, so complete resection may not be possible. For patients with unresectable advanced disease, treatment is similar to outlined for advanced nonresectable gastrinomas.[18]

PROGNOSIS

The prognosis is now largely determined by the growth of the tumor per se, because the symptoms of VIP excess can be largely controlled by somatostatin analogues. This is particularly true in patients with VIPomas, because they frequently (>50%) present with advanced metastatic disease. The mean 5-year survival is 50 to 70%.

Somatostatinomas

Definition and Pathobiology

Somatostatinomas are endocrine tumors that occur in the pancreas or upper small intestine and ectopically secrete somatostatin. In the gastrointestinal tract, somatostatin inhibits basal and stimulated acid secretion, pancreatic secretion, intestinal absorption of amino acids, gallbladder contractility, and release of numerous hormones, including cholecystokinin and gastrin.

CLINICAL MANIFESTATIONS

Most reported somatostatinomas are diagnosed histologically as an endocrine tumor containing somatostatin-like immunoreactivity and are not associated with a distinct clinical syndrome (the somatostatinoma syndrome). The somatostatinoma syndrome includes diabetes mellitus, gallbladder disease, diarrhea, steatorrhea, and weight loss. Sixty percent of somatostatinomas occur in the pancreas, and 40% are found in the duodenum or jejunum. Pancreatic somatostatinomas occur in the pancreatic head in 60 to 80% of cases; 70 to 92% will have metastasized at diagnosis, and they are usually large (mean, 5 cm) and solitary. In contrast, duodenal somatostatinomas are smaller (mean, 2.4 cm), frequently associated with psammoma bodies on histologic examination (11%), and less frequently have metastases at diagnosis (30 to 40%).

The somatostatinoma syndrome occurs much more commonly (80 to 95% of all cases) in patients with pancreatic than duodenal or intestinal somatostatinomas. Duodenal somatostatinomas occur in up to 10% of patients with von Recklinghausen disease and are usually asymptomatic.

DIAGNOSIS

Somatostatinomas are usually found by accident, particularly during exploratory laparotomy for cholecystectomy, during endoscopy, or on imaging studies. The presence of psammoma bodies on histologic examination of a duodenal endocrine tumor or any duodenal lesions in patients with von Recklinghausen disease should raise the suspicion of a duodenal somatostatinoma. The diagnosis of the somatostatinoma syndrome requires the demonstration of increased concentrations of somatostatin-like immunoreactivity in the plasma and the resected tumor. However, other tumors outside the pancreas or intestine, such as small-cell lung cancer, medullary thyroid carcinoma, pheochromocytomas, and paragangliomas, may also have elevated concentrations of somatostatin-like immunoreactivity. Somatostatinomas can be imaged using somatostatin receptor scintigraphy or, if needed, other conventional imaging studies to assess the tumor's location.

TREATMENT Rx

Treatment with octreotide or lanreotide can improve symptoms. Surgery, if possible, should be performed. For patients with unresectable advanced disease, treatment is similar to outlined for advanced nonresectable gastrinomas.[19]

PROGNOSIS

Patients with intestinal somatostatinomas, which uncommonly cause the somatostatinoma syndrome and are less malignant, have an excellent prognosis (5-year survival rate, >80%), whereas those with pancreatic somatostatinomas, which frequently cause the somatostatinoma syndrome and present with metastatic disease (>70%), have a much reduced 5-year survival rate (<50%).

GRFomas

DEFINITION

GRFomas are endocrine tumors that frequently originate in the pancreas but also occur in other extrapancreatic locations and ectopically release growth hormone–releasing factor (GRF). The GRF causes acromegaly that is clinically indistinguishable from that caused by a pituitary adenoma.

PATHOBIOLOGY

GRFomas most commonly occur in the lung (54%). Most of the remainder occur in the gastrointestinal tract, including 30% in the pancreas. Pancreatic GRFomas are usually large (mean, 6 cm), 39% are metastatic at diagnosis, 40% occur in combination with ZES, and 33% are in patients with MEN 1.

DIAGNOSIS

GRFomas are an uncommon cause of acromegaly. These tumors occurred in none of 177 unselected patients with acromegaly in one study. However, any patient with acromegaly and abdominal complaints, with acromegaly but no pituitary tumor (Chapter 224), or with acromegaly and hyperprolactinemia (which occurs in 70% of GRFomas) should be suspected of having a GRFoma. The intra-abdominal features of GRFomas result from its metastases and are typical of any malignant pNET. The diagnosis is confirmed by performing a plasma assay for GRF and growth hormone.

TREATMENT

The effects of the GRF can be controlled with octreotide or lanreotide in more than 90% of patients. Treatment should be directed at the GRFoma per se, as described for the other more common pNETs. For patients with unresectable advanced disease, treatment is similar to that outlined for advanced nonresectable gastrinomas.

Nonfunctional Pancreatic Neuroendocrine Tumors

DEFINITION

Nonfunctional pNETs are endocrine tumors that originate in the pancreas and either secrete no peptides or secrete products that do not cause clinical symptoms.

PATHOBIOLOGY

Frequently secreted nonfunctional peptides include chromogranin A (100%), pancreatic polypeptide (60%), and the α-subunit (40%) and β-subunit of human chorionic gonadotropin. Immunocytochemically, even higher percentages contain these peptides as well as insulin (50%), glucagons (30%), and somatostatin (13%).

CLINICAL MANIFESTATIONS

Nonfunctional pNETs are frequently diagnosed only late in the course of disease after the patient presents with symptoms or signs of metastatic disease and a liver biopsy reveals metastatic pNET. Any symptoms or signs result from the tumor per se and include abdominal pain (36 to 56%), abdominal mass or hepatosplenomegaly (8 to 40%), weight loss or cachexia (8 to 46%), and jaundice (27 to 40%). In 20% of asymptomatic patients, tumors are found incidentally at surgery.

DIAGNOSIS

Any patients with a long survival (>5 years) after a diagnosis of metastatic pancreatic adenocarcinoma should be suspected of having a nonfunctional pNET. Most primary tumors are large (70% are > 5 cm), and 70% occur in the pancreatic head. Liver metastases are frequent (38 to 62%) at presentation. An elevated plasma chromogranin A or pancreatic polypeptide level or a positive somatostatin receptor scintigraphic scan is strong evidence that a pancreatic mass is a pNET. Malignancy correlates with vascular or perineural invasion, a proliferative index of more than 2%, a mitotic rate of 2 or higher, a size of at least 4 cm, capsular penetration, nuclear atypia, lack of progesterone receptors, and the presence of calcitonin immunoreactivity in the tumor.

TREATMENT

Tumor localization is needed in all patients. Survival is better in patients with smaller tumors, patients who are asymptomatic at presentation, patients with no metastases, and patients in whom surgical resection can be performed.

Surgical resection should be performed whenever possible. For patients with unresectable advanced disease, treatment is similar to that outlined for advanced nonresectable gastrinomas.

PROGNOSIS

The overall 5-year survival rate varies in different series from 30 to 70%, but it is heavily dependent on the extent of the disease at diagnosis, with survival rates of 96% in patients without metastases at presentation, decreasing to 30 to 50% for those with metastatic disease.[20]

ACTHomas and Other Uncommon Tumors

pNETs that ectopically secrete ACTH cause 4 to 16% of the cases of ectopic Cushing syndrome. Cushing syndrome (Chapter 227) occurs in 5% of all cases of ZES, but in patients with sporadic ZES it is a late feature, occurring with metastatic liver disease. Its development is associated with a poor prognosis, and the response to chemotherapy is generally poor; however, occasional patients benefit from the use of long-acting somatostatin analogues (octreotide, lanreotide).

Paraneoplastic hypercalcemia (Chapter 245) can result from a pNET that releases parathormone-related peptide or an unknown hypercalcemic substance. Tumors are generally large, with metastatic liver disease at diagnosis. Somatostatin analogues may help control the hypercalcemia, but surgery, chemotherapy, hepatic embolization, and chemoembolization are the mainstays of treatment.

pNETs causing the carcinoid syndrome (Chapter 232) are usually large, and 68 to 88% are malignant. Octreotide may control the symptoms. Surgery, chemotherapy, hepatic embolization, chemoembolization, or molecular targeted therapy (everolimus, sunitinib) may be helpful.

A single case of a pNET that secreted renin manifested with severe hypertension; the tumor was localized with somatostatin receptor scintigraphy, and the patient's symptoms improved significantly after tumor resection. A single case of an erythropoietin-secreting pNET resulting in polycythemia, and a single case of a pNET secreting IGF-II or GLP-1 causing hypoglycemia have been described.

Two symptomatic cases of pNETs that secreted luteinizing hormone have been described; virilization occurred in the female patient, whereas the male patient had increased acne and a rash. In both cases, the tumors were resectable, and symptoms improved postoperatively. A single case of a pNET secreting cholecystokinin (CCKoma) has been recently described, with the patients demonstrating peptic ulcer disease, gallbladder disease, diarrhea, and weight loss.[21] A case of a pNET secreting enteroglucagon causing small intestinal hypertrophy has also been described.

Grade A References

A1. Caplin ME, Pavel M, Cwikla JB, et al. Lanreotide in metastatic enteropancreatic neuroendocrine tumors. *N Engl J Med*. 2014;371:224-233.

A2. Yao JC, Shah MH, Ito T, et al. Everolimus for advanced pancreatic neuroendocrine tumors. *N Engl J Med*. 2011;364:514-523.

A3. Raymond E, Dahan L, Raoul JL, et al. Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. *N Engl J Med*. 2011;364:501-513.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ito T, Igarashi H, Jensen RT. Pancreatic neuroendocrine tumors: clinical features, diagnosis and medical treatment: advances. *Best Pract Res Clin Gastroenterol*. 2012;26:737-753.
2. Singh S, Dey C, Kennecke H, et al. Consensus recommendations for the diagnosis and management of pancreatic neuroendocrine tumors: guidelines from a Canadian national expert group. *Ann Surg Oncol*. 2014;[Epub ahead of print].
3. Jensen RT, Cadiot G, Brandi ML, et al. ENETS Consensus guidelines for the management of patients with digestive neuroendocrine neoplasms: functional pancreatic endocrine tumor syndromes. *Neuroendocrinology*. 2012;95:98-119.
4. Fraenkel M, Kim MK, Faggiano A, et al. Epidemiology of gastroenteropancreatic neuroendocrine tumours. *Best Pract Res Clin Gastroenterol*. 2012;26:691-703.
5. Shi C, Klimstra DS. Pancreatic neuroendocrine tumors: pathologic and molecular characteristics. *Semin Diagn Pathol*. 2014;31:498-511.
6. Gebauer N, Schmidt-Werthern C, Bernard V, et al. Genomic landscape of pancreatic neuroendocrine tumors. *World J Gastroenterol*. 2014;20:17498-17506.
7. Oberg K, Casanovas O, Castano JP, et al. Molecular pathogenesis of neuroendocrine tumors: implications for current and future therapeutic approaches. *Clin Cancer Res*. 2013;19:2842-2849.
8. Jiao Y, Shi C, Edil BH, et al. DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science*. 2011;331:1199-1203.
9. Ito T, Igarashi H, Jensen RT. Therapy of metastatic pancreatic neuroendocrine tumors (pNETs): recent insights and advances. *J Gastroenterol*. 2012;47:941-960.
10. Oberg K. The genetics of neuroendocrine tumors. *Semin Oncol*. 2013;40:37-44.
11. Krampitz GW, Norton JA. Pancreatic neuroendocrine tumors. *Curr Probl Surg*. 2013;50:509-545.
12. Ito T, Cadiot G, Jensen RT. Diagnosis of Zollinger-Ellison syndrome: increasingly difficult. *World J Gastroenterol*. 2012;18:5495-5503.
13. Kunz PL, Reidy-Lagunes D, Anthony LB, et al. Consensus guidelines for the management and treatment of neuroendocrine tumors. *Pancreas*. 2013;42:557-577.
14. Sundin A. Radiological and nuclear medicine imaging of gastroenteropancreatic neuroendocrine tumours. *Best Pract Res Clin Gastroenterol*. 2012;26:803-818.
15. Ito T, Igarashi H, Uehara H, et al. Pharmacotherapy of Zollinger-Ellison syndrome. *Expert Opin Pharmacother*. 2013;14:307-321.
16. Bergsma H, van Vliet EI, Teunissen JJ, et al. Peptide receptor radionuclide therapy (PRRT) for GEP-NETs. *Best Pract Res Clin Gastroenterol*. 2012;26:867-881.
17. Toumpanakis C, Caplin ME. Update on the role of somatostatin analogs for the treatment of patients with gastroenteropancreatic neuroendocrine tumors. *Semin Oncol*. 2013;40:56-68.
18. Pavel M, Baudin E, Couvelard A, et al. ENETS Consensus Guidelines for the management of patients with liver and other distant metastases from neuroendocrine neoplasms of foregut, midgut, hindgut, and unknown primary. *Neuroendocrinology*. 2012;95:157-176.
19. Castellano D, Grande E, Valle J, et al. Expert consensus for the management of advanced or metastatic pancreatic neuroendocrine and carcinoid tumors. *Cancer Chemother Pharmacol*. 2014;[Epub ahead of print].
20. Lombardi M, De Lio N, Funel N, et al. Prognostic factors for pancreatic neuroendocrine neoplasms (pNET) and the risk of small non-functioning pNET. *J Endocrinol Invest*. 2014;[Epub ahead of print].
21. Rehfeld JF, Federspiel B, Bardram L. A neuroendocrine tumor syndrome from cholecystokinin secretion. *N Engl J Med*. 2013;368:1165-1166.

REVIEW QUESTIONS

1. Which is not a typical pathologic feature of most pancreatic neuroendocrine tumors (pNETs)?

A. Stain for chromogranin
B. Stain for synaptophysin
C. Stain for neuron-specific enolase
D. Have features of APUDomas
E. Show moderate to increased mitoses

Answer: E Most pNETs show few mitoses, whereas most stain for all the other listed items, and they are historically classified as APUDomas owing to their staining characteristics.

2. Which molecular change is not generally involved in the pathogenesis of pNETs?

A. Mutations of oncogenes
B. Mutations in the *MEN1* gene
C. Mutations in the DAXX-ATRX complex genes
D. Mutations/alterations in the mTOR pathway
E. Alterations in the *p16* gene

Answer: A In contrast to adenocarcinomas of the pancreas, mutations in common oncogenes (*ras*, *c-Jun*, etc.) or in common tumor suppressor genes (*p53*, retinoblastoma) are uncommon in pNET. In contrast, each of the other changes occurs in pNETs.

3. A 49-year-old man with abdominal pain is referred to you, and on laboratory evaluation is found to have a serum gastrin level of 35,000 pg/mL (normal, <100). Which of the following is true?

A. Diagnosis of Zollinger-Ellison syndrome (ZES) is likely.
B. Diagnosis of ZES syndrome is certain.
C. Diagnosis of ZES is possible.
D. Imaging studies should be performed next.
E. Upper GI endoscopy should be the next study to establish ZES.

Answer: C No level of elevation of fasting gastrin alone establishes the diagnosis of ZES, no matter how high. Other diseases, particularly atrophic gastritis or pernicious anemia, which are much more frequent, can give fasting gastrin levels in this range. The next study to perform to rule out these latter possibilities is to measure the gastric pH and establish that it is 2 or less when the fasting serum gastrin is elevated.

4. Diarrhea is not a presenting symptom of which pNETs:

A. Zollinger-Ellison syndrome
B. VIPoma
C. Glucagonoma
D. Somatostatinoma
E. GRFoma

Answer: E GRFomas characteristically present with acromegaly due to ectopic secretion of GRF, and diarrhea is not a feature. Each of the other pNETs not infrequently present with diarrhea as a feature.

5. Which of the following treatments is not commonly used to treat metastatic progressive pNETs?

A. Streptozotocin ± 5-fluoruracil ± doxorubicin
B. Somatostatin analogues
C. Surgery if most (usually > 90%) of the metastatic tumor can be removed
D. High-dose radiation to the hepatic metastases
E. Liver-directed therapies such as embolization

Answer: D Radiation to the liver is not used, whereas each of the other modalities is used at different points to treat patients with progressive metastatic pNETs.

196

LIVER AND BILIARY TRACT CANCERS

ROBIN K. KELLEY AND ALAN P. VENOOK

Malignancies arising from the liver parenchyma and biliary ductal epithelium are a heterogeneous group of cancers with generally poor prognosis, whose treatment is complicated by underlying liver injury, cirrhosis, or biliary obstruction in the majority of cases. These cancers collectively represent the third leading cause of cancer death worldwide and are rising in incidence, requiring clinician awareness of their presentation, evaluation, and treatment.

DEFINITION

Cancers arising from the liver and biliary tract include hepatocellular carcinoma, cholangiocarcinoma, and gallbladder adenocarcinoma, as well as other less common malignant histologies (Table 196-1, Fig. 196-1). Cholangiocarcinomas are further subclassified by their anatomic location along the biliary tract: intrahepatic (also known as peripheral) or extrahepatic, which includes hilar (also known as Klatskin tumors) and distal locations. The liver is a frequent site of metastasis from other primary cancers, requiring consideration of metastatic disease in the differential diagnosis of liver tumors. This chapter focuses on primary malignant tumors of the liver and biliary tract; metastatic disease to the liver is presented in other chapters, according to primary tumor of origin.

EPIDEMIOLOGY

Hepatocellular Carcinoma

Hepatocellular carcinoma is the most common primary tumor of the liver and the second leading cause of cancer death worldwide, accounting for approximately 782,000 new cases and 746,000 deaths worldwide each year based on estimates from 2012.[1] Over 80% of hepatocellular carcinoma cases

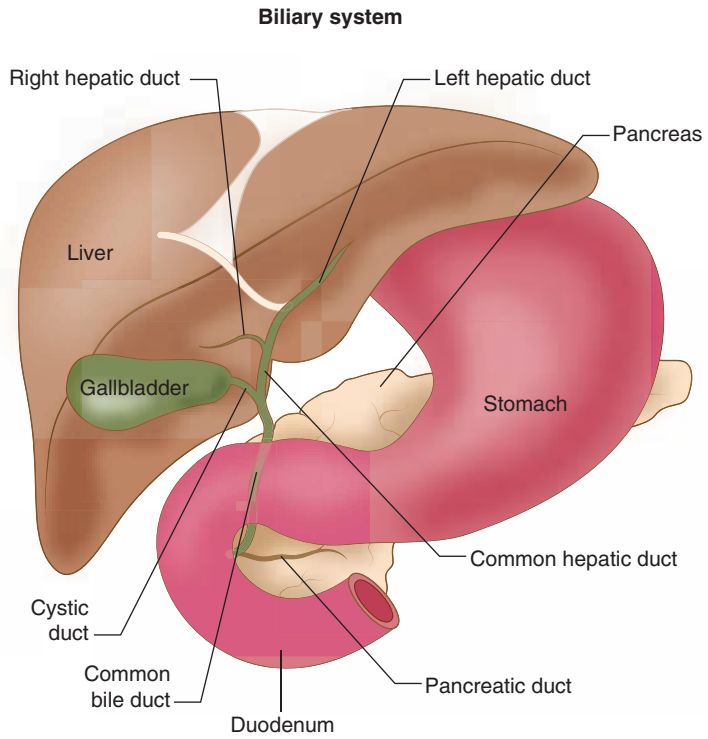

FIGURE 196-1. Anatomy of the liver and biliary tract. Hepatocellular carcinoma arises from liver parenchyma, and cholangiocarcinomas arise from the biliary ductal epithelium within (intrahapetic or peripheral) or outside of (extrahepatic) the liver. Tumors arising from the confluence of the right and left bile duct at the hepatic hilum are called hilar cholangiocarcinoma or Klatskin tumors.

TABLE 196-1 PRIMARY MALIGNANT TUMORS OF THE LIVER AND BILIARY TRACT

TUMOR TYPE	APPROXIMATE ANNUAL INCIDENCE
Hepatocellular carcinoma	782,000 cases/yr worldwide*
Gallbladder carcinoma	145,000 cases/yr worldwide
Cholangiocarcinoma	9,810 cases/yr in United States†
Mixed hepatocellular-cholangiocarcinoma	Rare
Fibrolamellar hepatocellular carcinoma	Rare
Cystadenocarcinoma	Rare
Hepatic endothelioid hemangioendothelioma	Rare
Angiosarcoma and other sarcomas	Rare

*Includes cases of intrahepatic cholangiocarcinoma.
†Includes cases of gallbladder cancer and excludes cases of intrahepatic cholangiocarcinoma.

occur in developing countries (Fig. 196-2), which is attributed to an increased prevalence of hepatocellular carcinoma risk factors, particularly hepatitis B virus (HBV) and hepatitis C virus (HCV) (Chapters 148 and 149, Table 196-2). HBV and HCV together account for over 80% of all hepatocellular carcinoma cases worldwide.[2,3] HBV is endemic in regions including Asia and sub-Saharan Africa and is vertically transmitted from mother to fetus in utero. HCV, which may be acquired by intravenous drug abuse or from blood transfusion, is another important risk factor for hepatocellular carcinoma, leading to an estimated hepatocellular carcinoma incidence of 2 to 8% per year among patients with cirrhosis caused by HCV. Obesity, diabetes, and the metabolic syndrome are risk factors for nonalcoholic fatty liver disease, a condition of rising incidence, particularly in Western populations, which can lead to steatohepatitis, cirrhosis, and hepatocellular carcinoma. Nonalcoholic fatty liver disease, alcohol-related disorders, and HCV are leading causes of hepatocellular carcinoma in the United States. Other risk factors for hepatocellular carcinoma include hepatitis D virus (which requires coinfection with HBV for pathogenicity), hereditary hemochromatosis, α_1-antitrypsin deficiency, primary biliary cirrhosis, autoimmune hepatitis, male gender, and dietary exposure to fungal aflatoxins.

Biliary Tract Cancers

Although classified as a rare cancer, cholangiocarcinoma is the most common primary biliary tract malignancy and the second most common primary cancer of the liver after hepatocellular carcinoma.[4] The incidence of intrahepatic cholangiocarcinoma appears to be increasing. Although most cases are sporadic, there is an increased risk for cholangiocarcinoma with the following conditions or exposures: primary sclerosing cholangitis, inflammatory bowel disease, viral hepatitis infection, congenital choledochal cysts or other structural abnormalities of the biliary tract, chronic pancreatitis, obesity, fluke infections and other causes of chronic cholangitis, bile duct adenomas, Caroli disease, and diabetes mellitus.

Gallbladder adenocarcinomas are a rare tumor type in the United States but account for 145,000 cases and 109,000 deaths per year worldwide based on estimates from 2008. There is significant regional variation in incidence, with higher incidence reported in India and areas of South America and Asia. Risk factors for gallbladder adenocarcinoma include female gender, chronic cholelithiasis, chronic cholecystitis, a history of gallbladder polyps, and abnormalities of the common bile duct. Calcification of the gallbladder ("porcelain gallbladder") is often considered a risk factor, but it is only weakly associated with development of cancer.

Other Primary Liver Cancers

Other primary tumors of the liver are much less common, with poorly understood risk factors. Chronic inflammatory conditions, including viral hepatitis, may be associated with the development of combined-histology liver tumors. Exposure to polyvinyl chloride has been implicated in the development of hepatic angiosarcomas. The fibrolamellar variant of hepatocellular carcinoma may occur more commonly in females. Hepatoblastoma tumors occur in infants and children, but almost never in adults.

PATHOBIOLOGY

Hepatocellular Carcinoma

Hepatocellular carcinoma is an epithelial neoplasm that arises from malignant transformation of liver hepatocytes (Fig. 196-3A and B). The pathogenesis of hepatocellular carcinoma is thought to be a multistep process triggered by underlying liver injury (such as from viral hepatitis, alcohol, iron overload, or aflatoxin exposure) in the majority of cases. Subsequent inflammation, necrosis, regeneration, cell turnover, and proliferation result in the progressive accumulation of genetic damage and somatic (acquired)

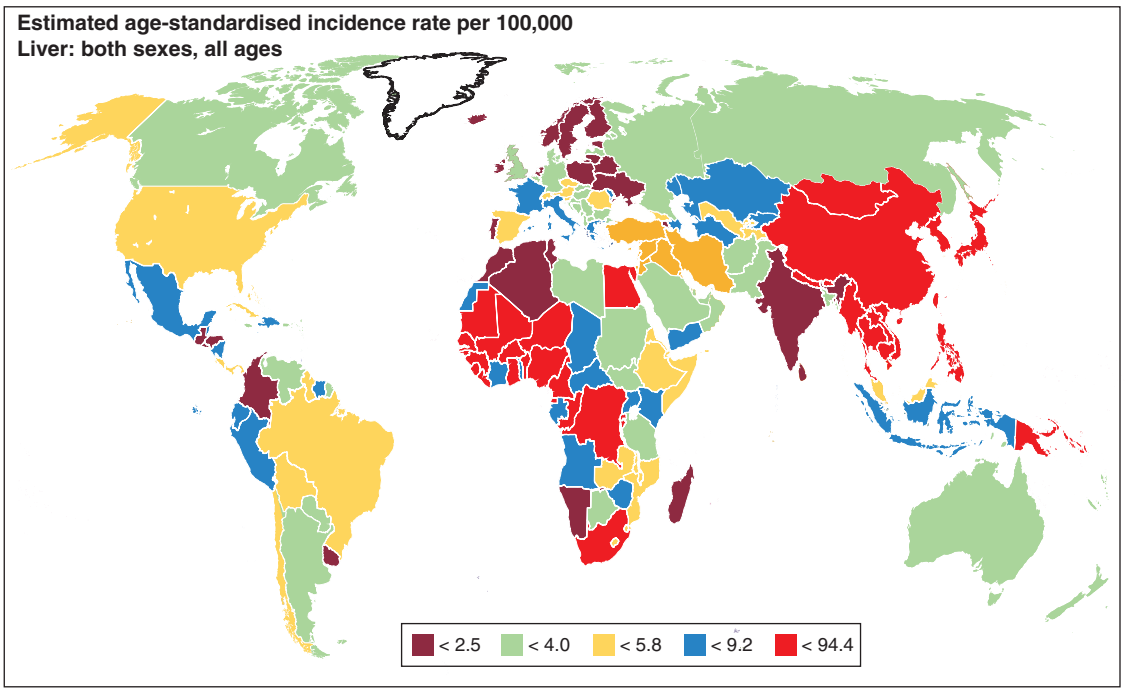

FIGURE 196-2. Global incidence of liver cancer. The incidence of liver cancer is highest in less developed regions, particularly Eastern and South-Eastern Asia and Middle and Western Africa.

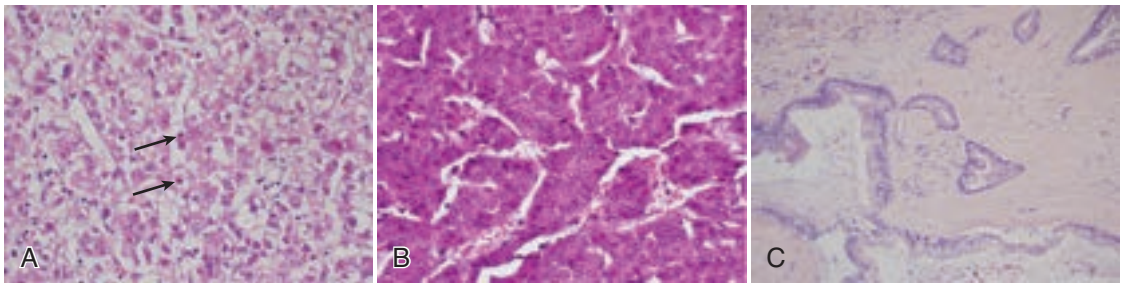

FIGURE 196-3. Histology of hepatocellular carcinoma and cholangiocarcinoma. A, Hepatocellular carcinoma (40×) with clear cell features and Mallory hyaline inclusions (*black arrow*). **B,** Hepatocellular carcinoma (40×) with trabecular pattern and small cell change. **C,** Hilar cholangiocarcinoma (20×) with dense fibrous stromal reaction. (Courtesy Dr. Linda Ferrell, Department of Pathology, University of California, San Francisco.)

TABLE 196-2 RISK FACTORS FOR HEPATOCELLULAR CARCINOMA

RISK FACTOR	INCIDENCE OF HEPATOCELLULAR CARCINOMA*
Asian male HBV carriers over age 40	0.4-0.6%/yr
Asian female HBV carriers over age 50	0.3-0.6%/yr
HBV carrier with family history of HCC	Higher incidence than without family history
African/North American Blacks with HBV	HCC occurs at younger age
HBV carriers with cirrhosis	3-8%/yr
HCV cirrhosis	3-5%/yr
Stage 4 primary biliary cirrhosis	3-5%/yr
Hereditary hemochromatosis with cirrhosis	Unknown
α_1-Antitrypsin deficiency with cirrhosis	Unknown
Other cause cirrhosis	Unknown

Modified from Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology.* 2011;53:1020-1022.
*Data from Nordenstedt H, White DL, El-Serag HB. The changing pattern of epidemiology in hepatocellular carcinoma. *Digest Liver Dis.* 2010;42(Suppl 3):S206-S214.
HCC = hepatocellular carcinoma.

mutations.[5] Activation of oncogenes or inactivation of tumor suppressor genes, dysplasia, and subsequently carcinoma can arise. The most well-described mutations in hepatocellular carcinoma are point mutations or deletions resulting in inactivation of the tumor suppressor *TP53* in over 50% of cases and mutations in β-catenin *(CTNNB1)* in approximately 30% of cases. Alterations in Wnt, cell cycle, and chromatin-remodeling pathways have been described and may be associated with the etiology of underlying liver injury. In HBV-associated hepatocellular carcinoma, a unique mechanism of malignant transformation is direct viral DNA integration into the host genome, which appears to favor specific loci. A recurrent locus for HBV integration is the *TERT* gene which encodes telomerase reverse transcriptase. HBV integration can activate *TERT,* resulting in malignant transformation and immortalization in a subset of HBV-associated hepatocellular carcinomas.

Biliary Tract Cancers

Cholangiocarcinomas are a histologically diverse group of epithelial cancers that may arise from multiple different cell types within the liver, including biliary epithelial cells or hepatic progenitor cells, and are often surrounded by a dense stroma with cancer-associated fibroblasts (see Fig. 196-3C). Cholangiocarcinomas are associated with underlying inflammation and cholestasis, which activate growth factors and a proliferative response. Overexpression of Notch1 and AKT have been implicated in a process of hepatocyte conversion into cholangiocyte precursors of intrahepatic

cholangiocarcinoma. Molecular and genomic subtypes of biliary tract cancers can also be defined by mutational status (such as isocitrate dehydrogenase-1 mutation in approximately 15%) and gene expression profiles.

CLINICAL MANIFESTATIONS

General Considerations

Regardless of histologic subtype, primary tumors of the liver can manifest with right upper quadrant pain, mass effect causing early satiety or obstructive symptoms, nausea, bleeding, and biliary obstruction. All of these tumors have metastatic potential; thus presentation with constitutional symptoms (such as weight loss, fevers, or night sweats) and signs or symptoms of metastatic disease (such as bone pain or a pathologic fracture) are also possible, though less common. Paraneoplastic syndromes are rarely a manifesting symptom in hepatobiliary cancers but can include erythrocytosis from erythropoietin production by tumor and hypercalcemia of malignancy.

Hepatocellular Carcinoma

The clinical presentation of hepatocellular carcinoma can vary according to extent of tumor and underlying liver dysfunction. Some patients may be asymptomatic, particularly when diagnosed by surveillance imaging and/or α-fetoprotein (AFP) tumor marker elevation. In some cases, patients present with symptoms of worsening liver function and portal hypertension (Chapter 153), such as new ascites, encephalopathy, gastrointestinal bleeding, or jaundice, as a result of hepatic decompensation triggered by a growing tumor. This presentation is more common if there is associated portal vein thrombosis. In other cases, there may be chronic progressive upper abdominal pain because of tumor involvement of the sensitive liver capsule, sudden onset acute pain from tumor bleeding or rupture, or a palpable mass leading to the diagnosis of hepatocellular carcinoma. Constitutional symptoms such as cachexia, fatigue, and weight loss may be present with advanced stages. On physical examination, patients with hepatocellular carcinoma may have an enlarged liver with tenderness. Particularly in large and rapidly growing tumors, a bruit may be auscultated over the liver surface. Ascites, jaundice, signs of portal hypertension such as caput medusae and splenomegaly, and asterixis may be variably present if there is associated decompensation in hepatic function (Chapter 153). The tumor marker AFP is elevated in approximately 70% of cases but is not diagnostic (see later discussion).

Biliary Tract Cancers

Extrahepatic cholangiocarcinomas most often manifest with signs and symptoms of biliary obstruction, such as jaundice, pruritus, pale stools, dark urine, anorexia, nausea, and weight loss. Intrahepatic cholangiocarcinomas can also cause obstruction, but generally only when there is extensive disease present. Less commonly, complications such as biliary fistulae and hemobilia can occur. Gallbladder cancers are often diagnosed incidentally during cholecystectomy, but in some cases can be associated with right upper quadrant pain, biliary colic, or a tender, palpable mass. Infections of the biliary tract can produce symptoms including right upper quadrant pain, fever, chills, nausea, vomiting, and jaundice. Cholangitis can result from biliary obstruction by tumors throughout the biliary tract and can lead to complications such as abscess, bacteremia, or sepsis syndrome.

Diagnosis and Staging

General Considerations

The approach to patients with the finding of a liver mass requires assessment of risk factors and extent of underlying liver disease, if present. The diagnostic evaluation, staging, and treatment options are guided by whether underlying liver disease is present as well as by its extent. The differential diagnosis for hepatocellular carcinoma and biliary tract cancers includes benign liver lesions (such as hemangiomas, adenomas, abscesses, and regenerative nodules), malignant tumors of other primary liver histologies, mixed-histology tumors, and metastatic disease. It is recommended that diagnostic evaluation and procedures in patients with hepatobiliary tumors be performed in expert centers because the unique interplay of underlying liver disease, disease-specific imaging findings and staging, and risks for biopsy.

Hepatocellular Carcinoma

In patients with known cirrhosis or other risk factors for hepatocellular carcinoma (see Table 196-2), a liver mass may be identified during a program of surveillance by imaging and/or elevated serum AFP level, in the absence of other signs or symptoms of cancer. A randomized, controlled trial of surveillance by ultrasound and AFP every 6 months compared to no surveillance in a large, predominantly HBV-positive Chinese population showed a survival benefit from surveillance, although the benefit has not been proved in randomized trials in other populations.[■] In other patients, clinical manifestations as described earlier may prompt imaging that identifies a liver mass and leads to further diagnostic evaluation.

Hepatocellular carcinoma is unique in oncology in that a diagnosis can be made radiographically without tumor tissue sampling, in the appropriate clinical context. The requirements for a radiographic diagnosis of hepatocellular carcinoma without biopsy are that known underlying liver disease (see Table 196-2) is present as a risk factor and that imaging is performed using a hepatocellular carcinoma protocol that entails contrast-enhanced cross-sectional imaging during multiple phases of contrast administration (including arterial, portal venous, and delayed phases). In patients at risk for hepatocellular carcinoma, a nodule of at least 1 cm featuring arterial phase enhancement with decreased enhancement (known as "washout") during the portal venous phase of contrast is sufficient for a radiographic diagnosis of hepatocellular carcinoma (see diagnostic algorithm in Fig. 196-4). The bright arterial enhancement of hepatocellular carcinoma lesions on contrast-enhanced imaging studies is due to the propensity of this tumor to parasitize blood supply from the hepatic artery, while the normal hepatic parenchyma derives the majority of its blood supply from the portal vein. This results in hepatocellular carcinomas "washing out" as the background liver brightens during the later portal venous phase. Figure 196-5 depicts the classic arterial enhancement (part A) and portal venous washout (part B) of a hepatocellular carcinoma tumor within a cirrhotic liver. Of note, an elevated AFP level is not sufficiently sensitive or specific for diagnosis of hepatocellular carcinoma in cirrhotic patients with a liver mass. A biopsy is warranted to confirm the diagnosis of hepatocellular carcinoma if either arterial enhancement or portal venous washout is not present. In patients with small lesions who may be eligible for curative surgery or transplantation, consultation with a hepatologist and/or experienced liver surgeon, should be obtained before performing percutaneous biopsy, because of risk for tumor seeding (Fig. 196-4).

Once a patient has been radiographically and/or pathologically diagnosed with hepatocellular carcinoma, the staging for extent of disease requires cross-sectional imaging of the chest, abdomen, and pelvis, AFP measurement, and a bone scan, if symptoms or signs of bone metastases (such as bone pain or markedly elevated alkaline phosphatase value) are present. The staging of hepatocellular carcinoma also requires a thorough assessment of underlying liver function, which affects prognosis and treatment options independent of tumor extent. Several hepatocellular cancer–specific, joint tumor staging and liver disease scoring systems have been developed, including the Barcelona Clinic Liver Cancer (BCLC) staging system, the Okuda classification, and the Cancer of the Liver Italian Program (CLIP) scoring system, though there is no consensus regarding which is superior.[6]

Other Liver Tumors and Biliary Tract Cancers

For patients without known risk factors for hepatocellular carcinoma, or if diagnostic enhancement features of hepatocellular carcinoma are not met, pathologic confirmation by biopsy or cytologic examination is required for diagnosis of liver tumors. As is the case with suspected hepatocellular carcinomas, however, there is concern for tumor seeding of the needle track by percutaneous or endoluminal biopsy. Referral to an expert center is recommended to guide the diagnostic evaluation in potential surgical candidates.

Endoluminal approaches such as by endoscopic retrograde cholangiopancreaticogram (ERCP) cytologic brushing, or endoscopic ultrasound fine needle aspiration are preferred in patients with early stages of disease that may be amenable to curative surgery or transplantation. Staging of biliary tract and other rarer types of liver cancer generally includes cross-sectional imaging of the chest, abdomen, and pelvis. A mass in the gallbladder fossa suggests primary gallbladder cancer. Unlike hepatocellular carcinoma, cholangiocarcinomas characteristically display progressive enhancement during the portal venous phase of contrast imaging (Fig. 196-7). Biliary obstruction and ipsilateral hepatic lobe atrophy and contralateral hypertrophy may be present and can sometimes obscure identification of the actual tumor mass. For biliary tract cancers, CA-19-9 and carcinoembryonic antigen (CEA) tumor markers may be elevated and can help to monitor response to treatment, although levels can be confounded if biliary obstruction and hyperbilirubinemia are present. Upper and lower endoscopy are indicated to exclude metastatic disease in patients diagnosed with intrahepatic cholangiocarcinoma, which otherwise can be difficult to discriminate from metastatic disease radiographically and histologically. Cholangiography (by ERCP or magnetic resonance cholangiogram) may be indicated for both

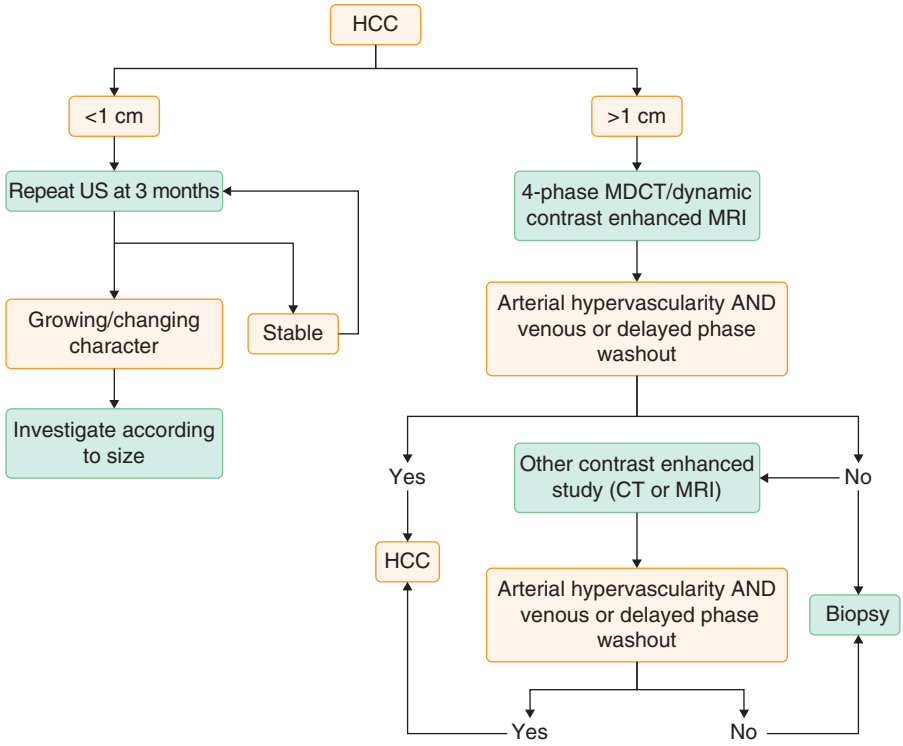

FIGURE 196-4. Hepatocellular carcinoma (*HCC*) diagnostic algorithm. American Association for the Study of Liver Diseases (AASLD) diagnostic algorithm for suspected hepatocellular carcinoma. CT = computed tomography; MDCT = multidetector CT; MRI = magnetic resonance imaging; US = ultrasound.

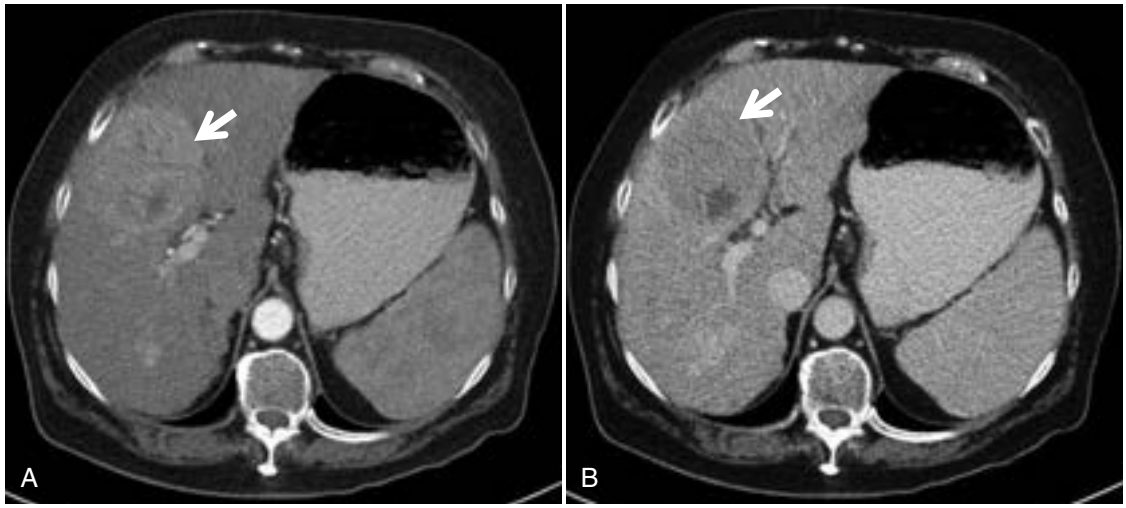

FIGURE 196-5. Imaging of hepatocellular carcinoma. In this contrast-enhanced computed tomography scan of the liver of a patient with hepatocellular carcinoma, a tumor in the right hepatic lobe demonstrates arterial phase enhancement (*white arrow,* panel A) followed by "wash-out" in the portal venous phase (white arrow, panel B).

diagnostic and therapeutic purposes (such as stent placement), particularly if biliary obstruction is present. A diagnostic laparoscopy to exclude peritoneal disease should be considered before undertaking a laparotomy for curative resection in patients with newly diagnosed gallbladder adenocarcinoma and cholangiocarcinoma, because of the propensity for radiographically occult peritoneal metastases, which would have an impact on surgical decision making. Staging of biliary tract cancers and other liver tumors follows the Tumor, Node, Metastasis (TNM) system of the American Joint Committee on Cancer.

TREATMENT

General Considerations

The treatment of patients with hepatobiliary cancers requires management of underlying liver disease as well as the tumor itself. In patients with active HBV infection or prior exposure, close monitoring of liver function and viral load is advisable; antiviral therapy may be required to prevent reactivation, which can occur with immunosuppressive therapy (Chapter 35).[7] Management of liver tumors in patients with cirrhosis may necessitate treatment for complications of portal hypertension or liver dysfunction (Chapter 153). In patients with biliary tract cancers, biliary obstruction is a common complication that often requires endoscopic stent placement, percutaneous drainage, or antibiotic therapy if cholangitis develops. Chemotherapy and supportive care medications may need dose adjustments depending on the degree of liver dysfunction.

Hepatocellular Carcinoma

The treatment of early stages of hepatocellular carcinoma depends on the degree of liver dysfunction. In patients with preserved liver function and without significant portal hypertension, surgical resection can be curative and is better than catheter-based options.[8] Among patients with increasing degrees of portal hypertension, however, surgical outcomes are significantly poorer than in patients without portal hypertension. For individuals with contraindications to surgery or inadequate projected future liver remnant

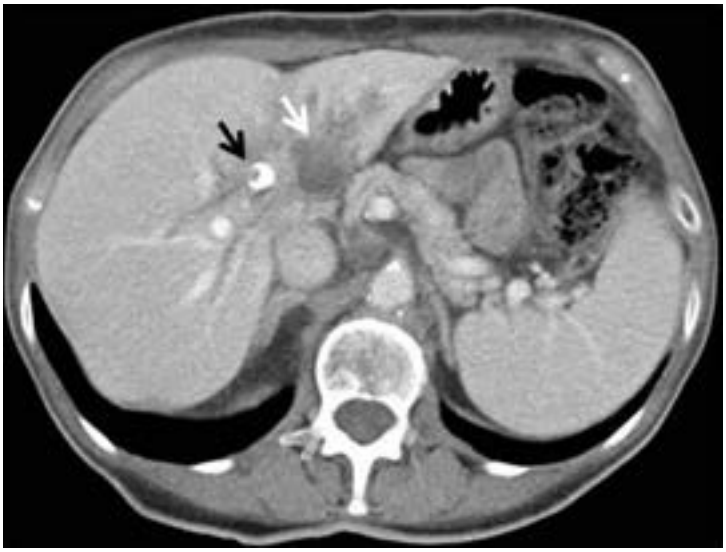

FIGURE 196-6. Imaging of cholangiocarcinoma. This computerized tomography image with contrast in portal venous phase depicts an infiltrative hilar cholangiocarcinoma (Klatskin tumor) extending into the left hepatic lobe (*white arrow*) with an endobiliary stent in place (*black arrow*).

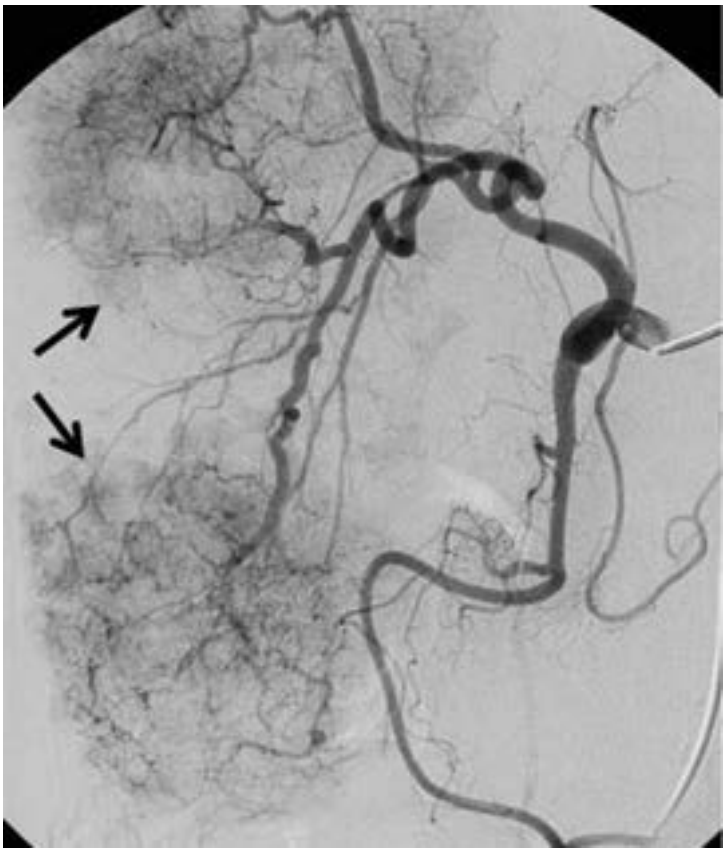

FIGURE 196-7. Trans-arterial chemoembolization (TACE) of hepatocellular carcinoma. This common hepatic artery angiogram from a 42-year-old woman with hepatitis B shows two distinct areas of "tumor blush" corresponding to underlying hepatocellular carcinoma lesions (*black arrows*) that were treated by TACE. (Courtesy Dr. Nicholas Fidelman, Department of Interventional Radiology, University of California, San Francisco.)

function, ablation of small tumors using probes that convey radiofrequency or microwaves, or by ethanol injection, can provide long-term control, and is sometimes curative.

For patients with early stages of hepatocellular carcinoma by the Milan Criteria (one lesion ≤ 5 cm or up to 3 lesions ≤ 3 cm each, without any evidence of vascular involvement or extrahepatic spread), orthotopic liver transplantation is an approved treatment with potential for long-term survival, achieving survival outcomes similar to patients undergoing transplantation for cirrhosis without cancer present. Extended criteria with parameters including larger tumor sizes may be accepted for transplantation at selected centers, often accompanied by liver-directed treatments such as embolization or ablation to control tumor burden during the period patients are waiting for transplant.[8]

When the hepatocellular carcinoma tumor burden exceeds criteria for transplantation or surgery but remains limited to the liver (BCLC intermediate stage), liver-directed therapies are commonly employed to delay progression and prolong survival, although these treatments are not likely to be curative. Transarterial chemoembolization (TACE) is the most common approach, which delivers embolic material, usually mixed with chemotherapeutic agents, via arterial catheters directly to vascular tumors within the liver. Randomized, clinical trials have demonstrated a survival benefit from TACE for intermediate-stage hepatocellular carcinoma. The optimal embolic material and chemotherapeutic agents for TACE have not yet been defined. Radioembolization using yttrium-90 bound to glass or resin microspheres is another arterially delivered therapy that may be employed in intermediate-stage hepatocellular carcinoma.[9]

For patients with advanced disease characterized by vascular involvement or extrahepatic spread, systemic therapy with sorafenib, a multikinase inhibitor whose targets include Raf kinase and vascular endothelial growth factor receptor isoforms, significantly prolonged survival compared to placebo in two randomized, phase III clinical trials. Conventional cytotoxic chemotherapy agents have not produced significant improvements in survival in hepatocellular carcinoma, unlike most other cancers, although a randomized trial of sorafenib combined with the anthracycline chemotherapeutic agent doxorubicin compared to doxorubicin alone suggested improved survival for the combination.[10] Sorafenib therapy is associated with greater absolute prolongation of survival in (1) patients with underlying HCV than in those with underlying HBV infection, (2) Western than in Asian populations, and (3) patients with Child-Pugh A than those with Child-Pugh B or poorer liver function, underscoring the clinical and biologic heterogeneity of hepatocellular carcinoma. In a recently reported phase III trial, sorafenib was found to be superior to sunitinib in overall survival, and also less toxic, in patients with hepatocellular carcinoma.

Biliary Tract Cancers

Surgical resection is the definitive therapy for patients with localized biliary tract cancers,[11] including gallbladder cancer. For patients with distal cholangiocarcinoma, a Whipple pancreaticoduodenectomy may be required. Liver transplantation for early-stage hilar cholangiocarcinoma after neoadjuvant chemotherapy and/or chemoradiation may be an option at selected centers.[12] In gallbladder cancer, cholecystectomy with en bloc hepatic resection and porta hepatis lymphadenectomy is recommended for patients with a gallbladder mass identified preoperatively on imaging or intraoperatively. A staging laparoscopy may be performed in advance to exclude occult peritoneal carcinomatosis. For patients with early-stage gallbladder cancer diagnosed incidentally on review of surgical pathologic findings after a cholecystectomy performed for benign causes, patients with tumors invading no deeper than the lamina propria (T1a) and negative surgical margins may be treated with observation only. Those found to have invasion to the muscular layer (T1b) or beyond may require additional hepatic resection and lymphadenectomy but should be referred to a center with expertise in the management of biliary tract cancers for evaluation and treatment.[13]

After surgical resection for patients with cholangiocarcinoma and gallbladder cancers, adjuvant therapy with chemotherapy and/or radiation, is often employed in fit patients, although randomized data are lacking. A large meta-analysis including 6710 patients from 20 studies suggested higher survival rates with the use of adjuvant chemotherapy, chemoradiation, or radiation.[14] In contrast, however, a randomized phase III trial comparing adjuvant treatment with gemcitabine or fluorouracil plus folinic acid versus observation in patients with periampullary cancers did not show a benefit for adjuvant therapy in a subset analysis of patients enrolled with a diagnosis of biliary tract cancers, although interpretation was limited by small sample size.

In patients with advanced biliary tract cancers not amenable to resection, combination chemotherapy with gemcitabine plus cisplatin improved survival compared to gemcitabine alone in a randomized phase III trial.

END-OF-LIFE CARE

The majority of patients diagnosed with primary hepatobiliary cancers will succumb to their cancer or complications thereof within a relatively short period of time; therefore, palliative care and end-of-life care play an integral role in management. As with most advanced cancers, pain control, treatment of nausea and constipation, and family and social support are essential. Bone metastases may require palliative radiation or stabilization procedures. Patients with advanced stages of hepatobiliary cancers, particularly hepatocellular carcinoma, are also at risk for developing complications of end-stage liver disease, such as intractable ascites, jaundice, pruritus, encephalopathy,

infections, and gastrointestinal bleeding (Chapter 135). Diuretic therapy, therapeutic paracentesis, and endoscopic management of gastrointestinal bleeding may be required. In biliary tract cancers, biliary obstruction and recurrent cholangitis also may require endoscopic or percutaneous biliary drainage and antibiotic therapy for palliation. Hospice referral may be appropriate when patients have progressed on standard anticancer therapies and/or are ineligible for further anticancer therapy because of extent of disease, liver dysfunction, poor performance status, or patient preferences.

In the United States, hepatocellular carcinoma is associated with a high incidence of health disparity, including immigrant status, racial or ethnic minority, and lower socioeconomic status. Providers must have an awareness of cross-cultural issues surrounding disclosure of diagnosis, pain control, use of alternative therapies, and end-of-life care. Pain management in patients with active substance abuse or a history of substance abuse (a risk factor for hepatocellular carcinoma) may be complicated by tolerance and/or dependency. In some cases, patients and caregivers may require providers' reassurance before using opiates or other analgesics for pain control because of concerns about addiction. Palliative care specialists and social workers provide important ancillary services in the end-of-life care for patients with hepatobiliary cancers.

PREVENTION

Screening and surveillance for hepatocellular carcinoma is associated with improved outcomes in selected populations, although a role for routine screening and surveillance has not been established across populations at risk and remains controversial. HBV vaccination programs have also been shown to reduce the incidence of hepatocellular carcinoma. Effective antiviral therapy for underlying HBV or HCV in patients with active viral hepatitis may be associated with a reduced risk for developing hepatocellular carcinoma.[15-17] Prevention and treatment of alcohol-related disorders, obesity, and other conditions associated with nonalcoholic fatty liver disease are appropriate measures to mitigate hepatocellular carcinoma risk factors and minimize ongoing liver injury, although prospective evidence for cancer risk reduction is limited.

There is limited evidence to support preventive measures or screening for biliary tract cancers. For patients with primary sclerosing cholangitis at increased risk for cholangiocarcinoma, periodic screening by noninvasive imaging and serum CA-19-9 marker measurements may be considered, although with limited supporting data.

PROGNOSIS

The prognosis of patients with hepatocellular carcinoma and biliary tract cancers is generally poor. For patients with hepatocellular carcinoma or intrahepatic cholangiocarcinoma, the overall 5-year survival across stages is approximately 20%. Among the minority of patients with hepatocellular carcinoma or intrahepatic cholangiocarcinoma diagnosed with localized stages of disease, the 5-year relative survival rate approaches 30%, although this figure is substantially higher in patients with early stages of disease who undergo curative ablation, surgery, or transplantation, among whom 5-year cause-specific survival rates generally approximate or exceed 60%.[18,19] For patients diagnosed with advanced stages of hepatocellular carcinoma or cholangiocarcinoma, the 5-year survival rate declines to approximately 3%, and the median overall survival remains less than a year with or without treatment.

For patients with gallbladder adenocarcinoma, the 5-year overall survival is approximately 15% across stages, with high rates of metastatic recurrence even among those with resectable disease. The prognosis of patients with rarer types of primary liver cancers such as fibrolamellar hepatocellular carcinoma, hepatic endothelioid hemangioendothelioma, or hepatic angiosarcoma is extremely heterogeneous and generally based on limited data from retrospective case series.

Grade A References

A1. Zhang BH, Yang BH, Tang ZY, et al. Randomized controlled trial of screening for hepatocellular carcinoma. *J Cancer Res Clin Oncol*. 2004;130:417-422.

A2. Yin L, Li H, Li AJ, et al. Partial hepatectomy vs. transcatheter arterial chemoembolization for resectable multiple hepatocellular carcinoma beyond Milan Criteria: a RCT. *J Hepatol*. 2014;61:82-88.

A3. Liu Z, Gao F, Yang G, et al. Combination of radiofrequency ablation with transarterial chemoembolization for hepatocellular carcinoma: an up-to-date meta-analysis. *Tumour Biol*. 2014;35:7407-7413.

A4. Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med*. 2008;359:378-390.

A5. Cheng AL, Kang YK, Chen Z, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. *Lancet Oncol*. 2009;10:25-34.

A6. Cheng AL, Kang YK, Lin DY, et al. Sunitinib Versus Sorafenib in Advanced Hepatocellular Cancer: Results of a Randomized Phase III Trial. *J Clin Oncol*. 2013;31:4067-4075.

A7. Neoptolemos JP, Moore MJ, Cox TF, et al. Effect of adjuvant chemotherapy with fluorouracil plus folinic acid or gemcitabine vs observation on survival in patients with resected periampullary adenocarcinoma: the ESPAC-3 periampullary cancer randomized trial. *JAMA*. 2012;308:147-156.

A8. Valle J, Wasan H, Palmer DH, et al. Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. *N Engl J Med*. 2010;362:1273-1281.

A9. Liaw YF, Sung JJ, Chow WC, et al. Lamivudine for patients with chronic hepatitis B and advanced liver disease. *N Engl J Med*. 2004;351:1521-1531.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. DeSantis CE, Lin CC, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin*. 2014;64:252-271.
2. Nordenstedt H, White DL, El-Serag HB. The changing pattern of epidemiology in hepatocellular carcinoma. *Digest Liver Dis*. 2010;42(suppl 3):S206-S214.
3. Bosetti C, Turati F, La Vecchia C. Hepatocellular carcinoma epidemiology. *Best Pract Res Clin Gastroenterol*. 2014;28:753-770.
4. Rizvi S, Gores GJ. Pathogenesis, diagnosis, and management of cholangiocarcinoma. *Gastroenterology*. 2013;145:1215-1229.
5. Han ZG. Functional genomic studies: insights into the pathogenesis of liver cancer. *Annu Rev Genomics Hum Genet*. 2012;13:171-205.
6. Li X, Dong M, Lin Q, et al. Comparison of current staging systems for advanced hepatocellular carcinoma not amendable to locoregional therapy as inclusion criteria for clinical trials. *Asia Pacif J Clin Oncol*. 2013;9:86-92.
7. Ling WH, Soe PP, Pang AS, et al. Hepatitis B virus reactivation risk varies with different chemotherapy regimens commonly used in solid tumours. *Br J Cancer*. 2013;108:1931-1935.
8. Roberts JP, Venook A, Kerlan R, et al. Hepatocellular carcinoma: ablate and wait versus rapid transplantation. *Liver Transplant*. 2010;16:925-929.
9. Salem R, Lewandowski RJ, Kulik L, et al. Radioembolization results in longer time-to-progression and reduced toxicity compared with chemoembolization in patients with hepatocellular carcinoma. *Gastroenterology*. 2011;140:497-507.
10. Abou-Alfa GK, Johnson P, Knox JJ, et al. Doxorubicin plus sorafenib vs doxorubicin alone in patients with advanced hepatocellular carcinoma: a randomized trial. *JAMA*. 2010;304:2154-2160.
11. Razumilava N, Gores GJ. Cholangiocarcinoma. *Lancet*. 2014;383:2168-2179.
12. Bridgewater J, Galle PR, Khan SA, et al. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol*. 2014;60:1268-1289.
13. Wernberg JA, Lucarelli DD. Gallbladder cancer. *Surg Clin North Am*. 2014;94:343-360.
14. Horgan AM, Amir E, Walter T, et al. Adjuvant therapy in the treatment of biliary tract cancer: a systematic review and meta-analysis. *J Clin Oncol*. 2012;30:1934-1940.
15. Gordon SC, Lamerato LE, Rupp LB, et al. Antiviral therapy for chronic hepatitis B virus infection and development of hepatocellular carcinoma in a U.S. population. *Clin Gastroenterol Hepatol*. 2014;12:885-893.
16. Morgan RL, Baack B, Smith BD, et al. Eradication of hepatitis C virus infection and the development of hepatocellular carcinoma: a meta-analysis of observational studies. *Ann Intern Med*. 2013;158:329-337.
17. Hosaka T, Suzuki F, Kobayashi M, et al. Long-term entecavir treatment reduces hepatocellular carcinoma incidence in patients with hepatitis B virus infection. *Hepatology*. 2013;58:98-107.
18. Salgia RJ, Singal AG, Fu S, et al. Improved post-transplant survival in the United States for patients with cholangiocarcinoma after 2000. *Dig Dis Sci*. 2014;59:1048-1054.
19. Altekruse SF, McGlynn KA, Dickie LA, et al. Hepatocellular carcinoma confirmation, treatment, and survival in surveillance, epidemiology, and end results registries, 1992-2008. *Hepatology*. 2012;55:476-482.

REVIEW QUESTIONS

1. Which of the following descriptions is sufficient for the diagnosis of HCC without biopsy?

A. There is no exception to the requirement for histologic diagnosis for HCC. A biopsy is always required to confirm malignant cells are present.
B. An Asian man with known hepatitis B virus (HBV) infection without cirrhosis who is found to have a 3-cm, space-occupying lesion in the liver and an α-fetoprotein (AFP) value of 400 ng/mL.
C. A patient with Child-Pugh C cirrhosis as a result of long-standing HCV infection who develops hepatic decompensation along with the finding of an infiltrative, ill-defined mass on a liver ultrasound and α-fetoprotein value of 43 ng/mL compared to baseline AFP of 18 ng/mL 1 year earlier.
D. The presence of a tumor at least 1 cm in size arterial phase enhancement and hypoenhancement ("wash-out") in portal venous phase on contrast-enhanced CT or MRI scan in a patient with known liver disease.

Answer: D A nodule greater than 1 cm in patients with risk factors for HCC may be diagnosed radiographically as HCC without biopsy if there is arterial phase enhancement with decreased enhancement during the portal venous phase. AFP is not sufficient or necessary for diagnosis. (Reference: Bruix J, Sherman M, Management of hepatocellular carcinoma: an update. *Hepatology.* 2011;53:1020-1022.)

2. A 54-year-old man has Child-Pugh B cirrhosis as a result of HCV infection and a recent history of esophageal variceal bleeding requiring transfusion. He is radiographically diagnosed with a 2-cm HCC in the right lobe of the liver that demonstrates arterial enhancement and portal venous phase "wash out" on a contrast-enhanced CT scan. There is no evidence of vascular invasion or extrahepatic spread on imaging. Which of the following is the most appropriate management?

A. Percutaneous liver biopsy to confirm the diagnosis of HCC followed by referral to the interventional radiology department for transarterial chemoembolization (TACE)
B. Ablation or resection of the tumor followed by whole liver adjuvant radiation therapy
C. Referral to an expert center to evaluate for liver transplant
D. Surgical resection followed by adjuvant sorafenib
E. Watchful waiting

Answer: C This patient with Child-Pugh B liver dysfunction has a history of variceal bleeding suggesting significant portal hypertension, which indicates he is not an appropriate candidate for liver resection because of inadequate remnant liver function. There is no proved role for adjuvant radiation after resection or ablation. Watchful waiting is not an appropriate option in this patient at high risk for complications of liver dysfunction and/or tumor progression. Outcomes of liver transplant for cirrhotic patients with small HCC lesions are equivalent to those of cirrhotic patients without HCC. (References: Mazzaferro V, Regalia E, Doci R, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med.* 1996;334:693-699; Pelletier SJ, Fu S, Thyagarajan V, et al. An intention-to-treat analysis of liver transplantation for hepatocellular carcinoma using organ procurement transplant network data. *Liver Transplant.* 2009;15:859-868. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology.* 2011;53:1020-1022.)

3. Sorafenib is currently the standard of care for patients with advanced stages of hepatocellular carcinoma. Which of the following statements is true about this drug?

A. Sorafenib is a selective inhibitor of c-Met tyrosine kinase that achieves prolonged time to progression in patients with advanced HCC whose tumors are positive for overexpression of c-Met by immunohistochemistry.
B. Sorafenib is a multikinase inhibitor, whose targets include VEGFR2 and Raf kinases, that has been shown to prolong survival in patients with incurable HCC in randomized, double-blinded, placebo-controlled phase III trials.
C. Sorafenib is a multikinase inhibitor that inhibits viral replication and can reverse the progression of cirrhosis, with increased benefit in patients with Child-Pugh B and C cirrhosis compared to Child-Pugh A cirrhosis.
D. Sorafenib demonstrates improved outcomes in Asian patients with HBV-associated liver disease by comparison to non-Asian patients with underlying HCV infection as the cause of liver disease.
E. Sorafenib is an immune-modulatory drug that stimulates effector T-cell responses against tumor antigens.

Answer: B Sorafenib is not a c-Met inhibitor nor an immune-modulatory drug. Patients with worse hepatic dysfunction experience greater toxicity and shorter survival on sorafenib therapy. Absolute survival outcomes on sorafenib appear poorer in Asian patients with underlying HBV than in non-Asian patients with underlying HCV. (References: Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med.* 2008;359:378-390. Cheng AL, Kang YK, Chen Z, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. *Lancet Oncol.* 2009;10:25-34.)

4. A 47-year-old woman of South American descent undergoes laparoscopic cholecystectomy for a clinical diagnosis of cholelithiasis. After surgery, her pathology specimen is found incidentally to harbor adenocarcinoma of the gallbladder invading into the muscle layer (T1b) with negative margins. Staging imaging is negative for metastatic disease. Which of the following should be performed?

A. No further treatment or observation is required after laparoscopic cholecystectomy.
B. She should be followed with close observation including cross-sectional imaging and CA-19-9 tumor marker levels at approximately 6 month intervals for up to 5 years.
C. She should be referred to a center with expertise in biliary tract cancers for hepatic resection and lymphadenectomy.
D. She should be treated with adjuvant gemcitabine-based chemotherapy for 6 months.
E. She should receive adjuvant radiation to the surgical resection bed.

Answer: C (Reference: Fuks D, Regimbeau JM, Le Treut YP, et al. Incidental gallbladder cancer by the AFC-GBC-2009 Study Group. *World J Surg.* 2011;35:1887-1897.)

5. A 71-year-old man previously in good health and without any known liver disease is found to have mildly elevated AST, ALT, and alkaline phosphatase, but normal bilirubin. He is asymptomatic without jaundice. An ultrasound shows a 4-cm right lobe liver mass along with two smaller right lobe lesions and one left lobe lesion, without significant biliary dilatation. Contrast-enhanced CT scans of the chest, abdomen, and pelvis show the same four hypoenhancing liver lesions during arterial phase, with no other abnormalities noted. A percutaneous CT-guided biopsy shows adenocarcinoma. Which is the most appropriate next step in management?

A. Referral to a gastroenterologist for esophagogastroduoedenoscopy (EGD) and colonoscopy to complete diagnostic evaluation.
B. Referral to radiation oncology department for stereotactic beam radiation therapy for a diagnosis of unresectable, multifocal intrahepatic cholangiocarcinoma.
C. Referral to an expert center to evaluate for liver transplant for intrahepatic cholangiocarcinoma.
D. Referral to interventional radiology department for chemoembolization for a diagnosis of intermediate-stage bilobar hepatocellular carcinoma exceeding Milan criteria for transplantation.

Answer: A The clinical scenario describes multifocal adenocarcinoma lesions that could be due to intrahepatic cholangiocarcinoma with intrahepatic metastases, versus metastatic adenocarcinoma of another primary site, most commonly the gastrointestinal tract. Other primary adenocarcinoma tumors within the gastrointestinal tract that commonly metastasize to the liver (such as gastric and colon cancers) are not adequately evaluated by cross-sectional imaging and require EGD and colonoscopy to complete staging and guide treatment decision making. Although there may be a role for radiation for intrahepatic cholangiocarcinomas, the diagnosis must be confirmed before proceeding with radiation or chemotherapy. Liver transplantation is not an appropriate treatment for intrahepatic cholangiocarcinoma with multifocal liver lesions, nor for metastatic disease to the liver. Choice D is incorrect because the clinical vignette describes adenocarcinoma, not hepatocellular carcinoma, which has a histology different from that of adenocarcinoma.

197 TUMORS OF THE KIDNEY, BLADDER, URETERS, AND RENAL PELVIS

DEAN F. BAJORIN

RENAL CELL CARCINOMA

DEFINITION

Cancers of the kidney are a heterogeneous group of neoplasms, the majority of which are of epithelial origin and malignant. Renal cell carcinoma, classically referred to as clear cell carcinoma or hypernephroma, is not a single malignancy. Rather, renal cell carcinoma comprises a group of distinguishable entities, each with a strong relationship between its morphologic and genetic features.[1] The World Health Organization recognizes these biologic and histologic differences in its classification system of kidney cancers (Table 197-1). The metastatic potential depends on the histologic subtype and ranges from the most virulent conventional clear cell carcinomas (65% of total tumors but accounting for 90% of the metastases), to the more indolent papillary and chromophobe carcinomas (25% of the total but only 10% of the metastases), and to the benign oncocytomas (10% of all tumors).

EPIDEMIOLOGY

There will be over 63,000 new cases of kidney and renal pelvis tumors in the United States in 2014, resulting in approximately 13,000 deaths.[2] These cancers represent the sixth most common form of cancer in men and the eighth most common in women. The increase in incidence of renal cell cancers may be in part related to early detection as a consequence of computed tomography (CT) and magnetic resonance imaging (MRI) of the abdomen for other medical conditions. The ratio of males to females is approximately 2 : 1 to 3 : 1, and the incidence is highest in African Americans

TABLE 197-1 CLASSIFICATION OF RENAL CELL NEOPLASMS

BENIGN	MALIGNANT
Oncocytoma	Clear cell (conventional) renal cell carcinoma
Papillary (chromophil) adenoma	Papillary (chromophil) renal cell carcinoma
Metanephric adenoma	Chromophobe renal cell carcinoma
Nephrogenic adenofibroma	Collecting duct carcinoma
	Medullary carcinoma
	Mixed tubular and spindle cell carcinoma
	Renal cell carcinoma, unclassified

Modified from Storkel S, Eble JN, Adlakha K, et al. Classification of renal cell carcinoma: Workgroup No. 1. Union Internationale Contre le Cancer (UICC) and the American Joint Committee on Cancer (AJCC). *Cancer.* 1997;80:987-989.

TABLE 197-2 HISTOLOGIC SUBTYPES, GENETICS, AND SYNDROMES

HISTOLOGIC SUBTYPE	PERCENT	MAJOR GENETIC/ MOLECULAR DEFECTS	OTHER GENETIC/MOLECULAR DEFECTS	ASSOCIATED SYNDROMES
Conventional clear cell	75	LOH 3p Mutation of 3p25 (VHL)	+5q, −8p, −9p, −14q *TP53* mutation, *c-erB-1* oncogene expression	Von Hippel-Lindau Hereditary RCC
Papillary 1	5	*C-Met* gene mutation 7q31	Trisomy 7, −4q, −6q, −9p, −13q, +12, +16, +20	Hereditary papillary renal cell carcinoma (HPRCC)
Papillary 2	10	Fumarate hydratase 1q42	−9p, −11q, −14q, −17p, −21q *PRCC-TFE3* gene fusion	Hereditary leiomyomatosis renal cell carcinoma (HLRCC)
Chromophobe	5	Birt-Hogg Dubé 17p11	−1p, −2p, −6p, −13q, −21q, −Y *TP53* mutation	Birt-Hogg Dubé
Oncocytoma	9.7	Birt-Hogg Dubé 17p11	−1, −Y, 11q Rearrangement	Familial oncocytoma Birt-Hogg Dubé
Collecting duct	0.4	−18, −Y	−1q, −6p, −8p, −11, −13q, −21q *c-erB-1* oncogene expression	Renal medullary carcinoma

Modified from Zambrano N, Histopathology and molecular genetics of renal tumors. *J Urol.* 1999;162:1246-1258.

and lowest in Asians and Pacific Islanders. The mean age at diagnosis is in the sixth to seventh decade of life. Aside from genetic predisposition, risk factors associated with renal cell carcinoma include cigarette smoking, obesity, hypertension, and the use of diuretics. Cigarette smoking has been associated with greater risk in both men and women. The risk may decrease after smoking cessation but requires about 20 years. Obese persons have an increased risk for renal cell carcinoma, and the risk rises with increasing body mass index. Although there is an elevated risk associated with diuretic use, this association is hard to distinguish from the increased risk associated with hypertension. Renal cell carcinoma is more prevalent in patients with preexisting renal conditions such as polycystic kidney disease, horseshoe kidney, and chronic renal failure requiring hemodialysis.

PATHOBIOLOGY

The classification system for renal cell carcinomas permits a better understanding of the cell of origin for the various subtypes and their chromosomal abnormalities (Table 197-2). The classic clear cell carcinoma constitutes approximately 65% of tumors and is believed to be derived from the proximal convoluted tubule. It is generally solitary and well circumscribed, with a golden yellow color resulting from the abundant cytoplasmic lipid. Higher grade tumors contain less lipid and glycogen. Approximately half of the tumors exhibit either a solid or acinar growth pattern characterized by solid sheets of tumor cells accompanied by a rich capillary vascular network. Papillary renal cell carcinomas comprise from 7 to 14% of primary epithelial renal neoplasms. The majority of patients present with unilateral tumors. Multifocality, either bilateral or multifocal lesions in the same kidney, is present in approximately 45% of cases. The majority of these tumors exhibit a broad morphologic spectrum, including papillary, papillary-trabecular, and papillary-solid areas; associated necrosis is a common finding. The classic papillary pattern is characterized by discrete papillary fronds lined by neoplastic epithelial cells and containing a central fibrovascular core, easily recognized on low magnification. These tumors are divided into type 1 and type 2 lesions, based on cytologic features and genetic differences. Chromophobe renal cancers account for 6 to 11% of renal epithelial tumors. Characteristically, these tumors are solitary and discrete but not encapsulated. The typical histologic findings consist of large round-to-polygonal cells with well-defined cell borders and pale basophilic cytoplasm admixed with a smaller population of polygonal cells with eosinophilic cytoplasm. These tumors may be quite large at diagnosis, with resectable tumors reported as big as 23 cm.

Clear cell carcinoma is characterized by the loss of genetic material from the short arm of chromosome 3 (3p) and mutations in the von Hippel-Lindau *(VHL)* gene. In patients with von Hippel-Lindau (VHL) disease, these losses and mutations occur in virtually all cases. The more common sporadic tumors also have somatic mutations and hypermethylation in the same region in approximately 75 to 80% of cases. Conventional clear cell tumors have a mutation in the VHL gene, which is inactivated by a point mutation or by epigenetic gene silencing by promoter methylation. The loss of *VHL*, responsible for ubiquination and degradation of hypoxia-inducible factor (HIF), leads to upregulation of HIF-responsive genes responsible for angiogenesis and cell growth. Two of these upregulated genes are platelet derived growth factor (PDGF) and vascular endothelial growth factor (VEGF), which are pro-angiogenic proteins thought to induce the neovascularity in both primary and metastatic clear cell cancers. Patients with VHL more commonly develop tumors at an earlier age and frequently have multiple tumors. Other tumors associated with the syndrome include central nervous system hemangioblastomas, pancreatic neuroendocrine tumors, pheochromocytomas, retinal angiomas, and epididymal cystadenomas. More recent molecular characterization of renal cell carcinoma shows alterations in genes responsible for maintenance of chromatin states such as *PBRM1,* the SWI/SNF chromatin remodeling complex including *ARID1A* and *SMARCA4,* and members of the PI3K/AKT pathway.[3]

The majority of sporadic papillary renal cell carcinomas are characterized by trisomy of chromosomes 7 and 17 and loss of chromosome Y. Chromophobe renal cell cancers have genetic loss on chromosomes 1 and Y, as well as combined chromosomal losses affecting chromosomes 1, 6, 10, 13, 17, and 21. Hereditary papillary renal cell cancer is a result of germline mutations and activation of the *MET* proto-oncogene, which is located on chromosome 7p. These cells have aberrant hepatocyte growth factor receptors that are unable to deactivate after binding by the growth factor. Somatic *MET* gene amplifications also have been observed in approximately 10% of sporadic papillary renal cancer. Hereditary leiomyomatosis renal cell carcinoma, characterized by alteration of the gene fumarate hydratase, is associated with uterine leiomyomas (more common) or leiomyosarcoma (rare), cutaneous nodules (leiomyomas), and type 2 papillary renal cell carcinoma, which is often solitary and frequently develops metastases. Birt-Hogg Dubé syndrome is a rare disorder predominantly associated with chromophobe renal cancers but in which clear cell and chromophobe/oncocytic tumors can develop. Birt-Hogg-Dubé syndrome is characterized by fibrofolliculomas, pulmonary cysts, pneumothorax, and bilateral renal tumors. The gene associated with Birt-Hogg Dubé syndrome has been mapped to 17p and expresses a novel protein, folliculin, whose function is not yet characterized.

CLINICAL MANIFESTATIONS

Although renal cell carcinoma has a high propensity for metastases and is associated with paraneoplastic syndromes, the majority of patients are asymptomatic at presentation. Historically, renal cell carcinoma was characterized by the presenting triad of hematuria, a palpable mass, and pain in as many as 10% of patients. However, there has been a stage migration resulting in the detection of tumors at earlier stages with the increased use of abdominal imaging for unrelated medical conditions in modern series. Up to 48% of tumors may be discovered in this manner, and less than 5% of patients have a palpable mass at presentation. The more common manifesting symptoms are anemia, weight loss, malaise, and anorexia (Table 197-3). Patients presenting with renal cell carcinoma frequently have associated paraneoplastic syndromes. Hypercalcemia has been observed in approximately 20% of patients and can be due to the secretion of parathyroid hormone, parathyroid hormone–like peptide, and interleukin-6 (IL-6), which have been shown to stimulate osteoclastic bone resorption. Other associated syndromes include hypertension, erythrocytosis (from ectopic erythropoietin production), and the rare Stauffer's syndrome, which is the presence of liver dysfunction without the presence of hepatic metastases; the hepatic dysfunction resolves after surgical resection of the tumor.

TABLE 197-3 PRESENTING SYMPTOMS AND SIGNS OF RENAL CELL CARCINOMA (BOTH LOCALIZED AND METASTATIC DISEASE)

SYMPTOMS AND SIGNS	PERCENT
Anemia	52
Hepatic dysfunction	32
Weight loss	23
Hypoalbuminemia	20
Malaise	19
Hypercalcemia	13
Anorexia	11
Thrombocytosis	9
Night sweats	8
Fever	8
Hypertension	3
Erythrocytosis	4
Chills	3

Modified from Kim HI, Belldegrun AS, Freitas DG, et al. Paraneoplastic signs and symptoms of renal cell carcinoma: implications for prognosis. *J Urol.* 2003;170:1742-1746.

DIAGNOSIS

The complete evaluation for patients with suspected renal cell carcinoma should include a complete blood count, a chemistry profile, a bone scan, and a CT scan of the chest, abdomen, and pelvis. CT is the most reliable method for detecting and staging of renal cell carcinoma. The "ideal" CT scan for renal masses can be divided into four phases, including the pre-contrast images, the arterial phase (~25 seconds after injection), the nephrographic phase (~90 seconds into the injection), and the excretory phase. The most important phases for imaging renal tumors are the pre-contrast and nephrographic images because renal lesions appear low in density in contrast to the uniformly enhanced renal parenchyma. The arterial phase is helpful for identifying renal arteries and small hypervascular masses. The excretory phase aids in assessing the collecting system and the renal pelvis. The CT scan is also helpful in detecting regional metastases, and three dimensional CT imaging is now possible in cases in which nephron-sparing surgery or partial nephrectomy is planned. The additional use of ultrasonography and MRI can help distinguish benign from malignant lesions of the kidney and in treatment planning. Ultrasound is used when distinguishing cysts from solid lesions. MRI has the advantage of imaging tumors in patients with poor renal function in whom intravenous contrast may be contraindicated. MRI is also helpful for delineating any thrombi that may be extending into the renal vein or inferior vena cava, and magnetic resonance angiography can be used to determine the number and location of renal arteries in patients who are candidates for partial nephrectomy. Once the evaluation is complete, the clinical stage is assessed using the (Tumor, Node, Metastasis (TNM) system (Table 197-4).

TABLE 197-4 TNM STAGING OF RENAL CELL CARCINOMA

TUMOR, NODES, AND METASTASES CLINICAL CLASSIFICATION	
Primary Tumor (T)	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
T1	Tumor 7 cm or less in greatest dimension, limited to the kidney
T1a	Tumor 4.0 cm or less in greatest dimension limited to kidney
T1b	Tumor more than 4 cm but not more than 7 cm in greatest dimension; limited to kidney
T2	Tumor more than 7 cm in greatest dimension, limited to kidney
T2a	Tumor more than 7 cm but less than or equal to 10 cm in greatest dimension, limited to the kidney
T2b	Tumor more than 10 cm, limited to the kidney
T3	Tumor extends into major veins or the perinephric tissues but into the ipsilateral adrenal gland and not beyond Gerota's fascia
T3a	Tumor grossly extends into renal vein(or its segmental (muscle containing) branches, or tumor invades perirenal and/or renal sinus fat but not beyond Gerota's fascia
T3b	Tumor grossly extends into the vena cava below the diaphragm
T3c	Tumor grossly extends into vena cava above diaphragm or invades the wall of the vena cava
T4	Tumor invades beyond Gerota's fascia (including contiguous extension into the ipsilateral adrenal gland)
N—Regional Lymph Nodes	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in regional lymph node(s)
M—Distant Metastasis	
Mx	Distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis
STAGE GROUPING	
Stage I	T1 N0 M0
Stage II	T2 N0 M0
Stage III	T1 N1 M0
	T2 N1 M0
	T3a N0 or N1 M0
	T3b N0 or N1 M0
	T3c N0 or N1 M0
Stage IV	T4 N0 M0
	T4 N1 M0
	Any T Any N M1

From *AJCC Staging Manual,* 7th ed. New York: Springer-Verlag; 2010.

TREATMENT

Rx

Localized Disease

The historical standard of care for patients with a renal cell carcinoma is a radical nephrectomy. Kidney cancers routinely selected for radical nephrectomy include large and centrally localized tumors that have effectively replaced the majority of the normal renal parenchyma, tumors associated with regional adenopathy (of benign or malignant etiology), those with inferior vena cava or right atrial extension, and even those in which metastatic disease is evident. Nephrectomy can be performed through a flank, transperitoneal, or transthoracic incision. The ipsilateral adrenal gland is also removed, but a regional lymph node dissection is optional and controversial. The increasing percentage of small tumors has resulted in a corresponding decrease in patients undergoing radical nephrectomy with excellent long-term survival.[4] Both open and laparoscopic approaches can be used for partial nephrectomy to control disease and preserve renal function. Laparoscopic nephrectomy offers a minimally invasive alternative to the classic radical nephrectomy. Partial nephrectomy for tumors of 7 cm or less, whether performed by open or minimally invasive laparoscopic technique, accomplishes rates of local tumor control and survival similar to radical nephrectomy. Partial nephrectomy reduces the risk for renal insufficiency over time.[5] Management with a partial nephrectomy is further supported by the fact that approximately 35% of renal cortical tumors are the indolent papillary or chromophobe carcinomas.

Renal cell carcinomas are resistant to both radiation therapy and cytotoxic chemotherapy; hence, those modalities of treatment have no role in the adjuvant setting after nephrectomy. Immunotherapy agents and approved drugs targeting VEGF or mammalian target of rapamycin (mTOR), beneficial in metastatic disease, have not been shown to affect survival favorably after nephrectomy.

Metastatic Disease

Approximately 30% of patients with renal cell carcinoma present with metastatic disease, and an additional 20 to 30% of patients with surgically resected primary tumors will relapse with metastases. Complications of metastatic disease include pain from either an unresectable primary tumor or skeletal metastases. Radiation therapy is frequently used for palliation of bone

metastases and for patients with multiple brain metastases. Palliative nephrectomy is sometimes used to provide symptomatic relief of pain. Surgical resection of the primary tumor is considered a mainstay of treatment even in the patient with metastatic disease and has been shown to extend survival in patients with metastatic disease. Surgical resection of metastatic sites of disease (metastasectomy) may also extend survival and even cure a subset of patients. The patients most likely to benefit from surgical resection of metastatic disease are those with a disease-free interval of greater than 1 year, those with a solitary site of metastasis, and those with lung metastases. Long-term survival has been observed when the solitary site of resection was the lung (up to 45%) and even the brain (up to 20%). Renal cell carcinoma is resistant to most conventional chemotherapy agents, with responses seen in less than 10% of patients.

Immunotherapy with either IL-2 or interferon-α (IFN-α) has been the historical standard treatment for patients with metastatic disease. High-dose intravenous IL-2, a potentially curative therapy, requires a dedicated inpatient setting because of its severe toxicities, including hypotension, pulmonary edema, renal failure, and central nervous system toxicity. However, most toxicities are reversible and complete or partial responses are seen in approximately 15 to 20% of patients; approximately 4% of patients achieve long-term, disease-free survival.[6] IFN-α therapy is less toxic than IL-2 and has an overall response rate of approximately 15%, but long-term survival is not observed. Reversible toxicities of IFN-α treatment include flulike symptoms, including fever, chills, myalgias, mild myelosuppression, and mild hepatic dysfunction.

Renal cell carcinoma has been an ideal candidate for the development of drugs targeting the downstream effects of *VHL* mutations.[7] Clinical trials have shown the benefit of tyrosine kinase inhibitors (TKIs), such as sunitinib,■ axitinib,■ pazopanib,■ and sorafenib,■ which block the actions of VEGF and PDGF; all are approved for the treatment of metastatic disease. Common side effects include fatigue, diarrhea, hypertension, and hand-foot syndrome, a condition in which blisters appear at areas of contact. The combination of IFN plus bevacizumab, an antibody that blocks the VEGF receptor, is superior to IFN alone and also has been approved for first-line treatment.■ Side effects include hypertension and an increased risk for bleeding. Two drugs targeting the mTOR pathway are also approved for the treatment of renal cell carcinoma both as first-line treatment and for patients whose disease has progressed despite TKI treatment. Temsirolimus, an intravenous mTOR inhibitor, improves survival of patients with untreated poor risk disease (those with more than three risk factors, see later).■ Everolimus, an oral mTOR inhibitor, improves the outcomes of patients who have been previously treated with sunitinib, sorafenib, or bevacizumab.■ Common side effects include fatigue, skin rash, and mouth sores.

PROGNOSIS

Progression-free survival and overall survival rates for resected nonmetastatic renal cortical tumors substantially differ according to multiple factors including age, size, grade and pathologic state (Figure 197-1, and also E-Figures 197-1, 197-2, and 197-3). The prognosis declines considerably for patients with more advanced disease, with long-term survival seen in only 20% of stage III patients and 5% or less in stage IV patients. Of the more common histologic subtypes of renal cell carcinoma, the prognosis of clear cell carcinoma is less favorable than that of papillary renal cell carcinoma; chromophobe renal cell carcinoma is the most favorable. For patients with metastatic disease, five clinical features associated with shorter survival are low performance status, high lactate dehydrogenase, low hemoglobin, high calcium, and absence of prior nephrectomy. Three strata groups have been defined using survival data from patients treated with immunotherapy: (1) favorable (zero risk factors) with a median survival of 20 months; (2) intermediate (1 or 2 risk factors), with a median survival of 10 months; and (3) poor (three or more risk factors) with a median survival of 4 months. IL-2 immunotherapy and surgical resection of solitary metastases can result in long-term survival of a small percentage of patients with renal cell cancer. Drugs targeting VEGF and the mTOR pathway are now the standard of care for patients with metastatic disease.

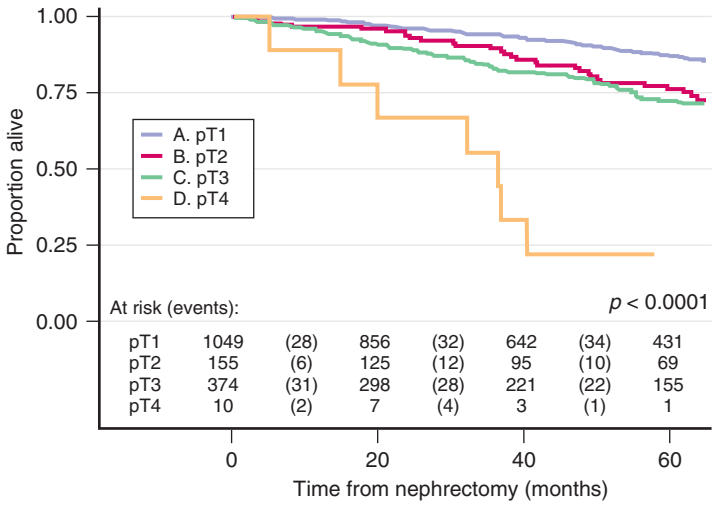

FIGURE 197-1. **Overall survival after resection of localized kidney cancer according to pathologic tumor (pT) classification. Curve A indicates pT1 tumors; curve B, pT2 tumors; curve C, pT3 tumors, curve D, pT4 tumors. (The "pT" categories correspond to the "T" categories in the TNM staging categories shown in Table 197-4). (From Russo P, Jang TL, Pettus JA, et al. Survival rates after resection for localized kidneys cancer: 1989-2004. *Cancer*. 2008;113:84-96.)**

BLADDER CANCER

DEFINITION

A spectrum of tumors arise from the urothelial lining of the bladder, renal pelvis, ureters, and urethra, of which transitional cell carcinoma is the most common. The vast majority of tumors arise from the bladder, with a minority arising from the upper tracts (renal pelvis and ureters) and even less frequently from the proximal urethra. Although transitional cell cancers possess a variable natural history, they have a proclivity for multifocality, high recurrence rates, and progression to higher pathologic stages. These tumors are generally grouped into the three broad categories of non–muscle invasive, muscle-invasive, and metastatic disease, each of which differs in clinical behavior, prognosis, and primary management. For non–muscle invasive tumors, the aim is to prevent recurrences and progression to a more advanced stage. In muscle-invasive disease, the medical challenge is to integrate the modalities of surgery, chemotherapy, and/or radiation to optimize cure and minimize morbidity. For metastatic disease, chemotherapy is used to palliate the symptoms of most patients, but there is a subset of patients in which combination chemotherapy may result in long-term cure. Long-term cure is directly related to stage and grade, ranging from 99% for low-grade Ta tumors to up to 15% for metastatic disease.

EPIDEMIOLOGY

An estimated 74,000 new cases of bladder cancer will be diagnosed in the United States in 2014, of which approximately 15,000 patients are expected to succumb to their disease. The ratio of males to females is 3 : 1, similar in all racial groups; it is the fourth most common cancer in men. The 5-year survival for all stages is 78%, resulting in a high prevalence (>500,000 people in the United States living with bladder cancer); it is twice as prevalent in whites as in African Americans and is less frequently observed in Asians. The vast majority of patients (90%) are over 50 years of age at diagnosis, with a median age of 73 at diagnosis. There is a lifetime risk of 2.4% of men and women developing bladder cancer.

Carcinogens or their metabolites implicated in the carcinogenesis of bladder cancer are believed to be excreted in the urine, where they can act directly on the urothelial lining. The latency period from initial exposure to the development of cancer is almost 20 years, making it difficult to establish a definitive cause and effect relationship between a putative carcinogen and the development of disease. Cigarette smoking is the leading risk factor for bladder cancer, believed to contribute to half of the cancers in men and one quarter of the cancers in women. A longer duration of exposure is associated with a higher risk than a more intense exposure (in cigarettes/day) over a shorter time period. Overall, smokers have a two- to four-fold higher relative risk for bladder cancer than nonsmokers. Smoking is associated with cellular atypia of the urothelium; individuals who never smoked show atypia in only 4% of cases in contrast to a 50% incidence of atypia in smokers.

Polycyclic aromatic hydrocarbons such as 2-naphthylamine, 4-aminobiphenyl and benzidine, and benzene or exhausts from combustion gases are associated with an increased risk for bladder cancer. Occupations reported to be at higher risk include aluminum workers, dry cleaners, manufacturers of preservatives and polychlorinated biphenyls, and pesticide applicators. Arylamines, also implicated in carcinogenesis, are metabolically activated to electrophilic compounds by N-hydroxylation in the liver by cytochrome P-450 IA2 and detoxified by N-acetylation; studies suggest that individuals with a fast oxidizer and slow acetylator phenotype are at highest risk.

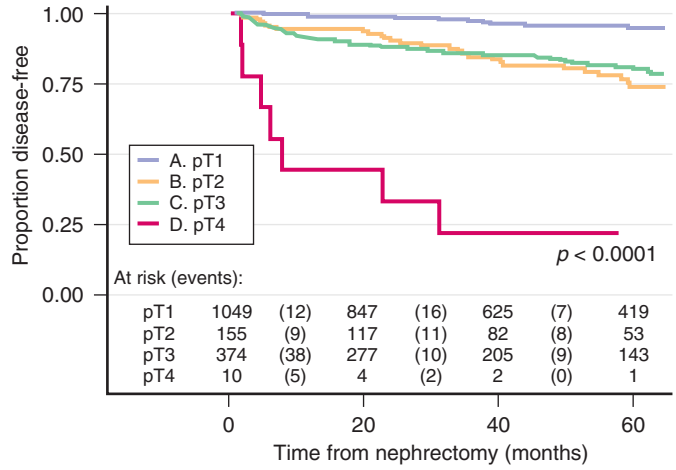

E-FIGURE 197-1. Progression-free survival according to pathologic tumor (pT) classification. Curve A indicates pT1 tumors; curve B, pT2 tumors; curve C, pT3 tumors; curve D, pT4 tumors. (The "pT" categories correspond to the "T" categories in TNM staging categories shown in Table 197-4.) (From Russo P, Jang TL, Pettus JA, et al. Survival rates after resection for localized kidney cancer: 1989-2004. *Cancer* 2008;113:84-96.)

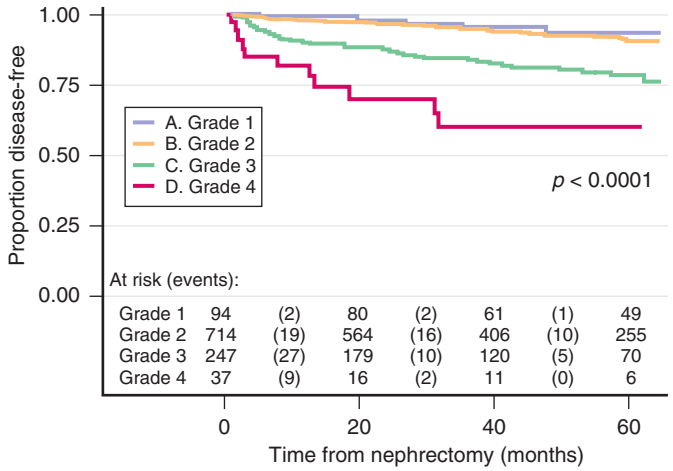

E-FIGURE 197-2. Progression-free survival according tumor grade. Curve A indicates grade 1; curve B, grade 2; curve C, grade 3; curve D, grade 4.) (From Russo P, Jang TL, Pettus JA, et al. Survival rates after resection for localized kidney cancer: 1989-2004. *Cancer* 2008;113:84-96.)

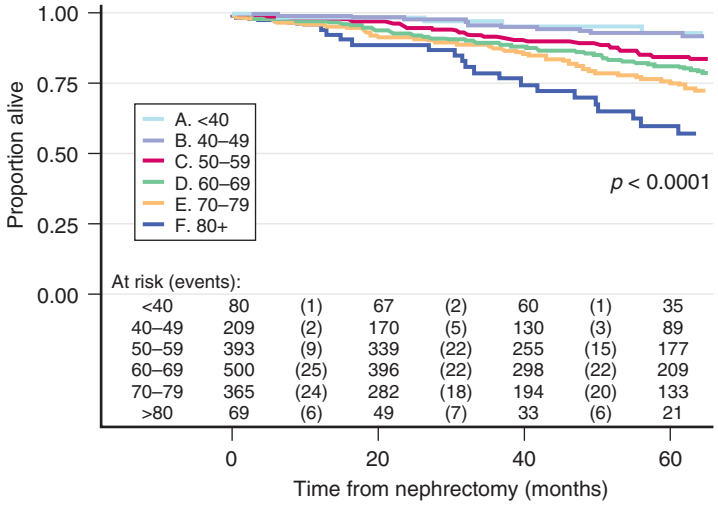

E-FIGURE 197-3. Overall survival according to patient age. Curve A indicates <40 years; curve B, ages 40 to 49 years; curve C, ages 50-50 years; curve D, ages 60-69 years; curve E, ages 70-79 years; curve F, ages ≥ 80 years. (From Russo P, Jang TL, Pettus JA, et al. Survival rates after resection for localized kidney cancer: 1989-2004. *Cancer* 2008; 113:84-96.)

Occupations associated with a higher exposure to arylamines such as workers in the dye, rubber, or leather manufacturing industries are believed to be at higher risk for developing bladder cancer. *Schistosoma haematobium* infection enhances formation of carcinogenic N-nitroso compounds and results in an increased risk for both squamous and transitional cell carcinomas of the bladder. An association has been observed between squamous cell carcinoma (but not transitional cell tumors) and the presence of chronic urinary tract infections seen in paraplegics and patients with chronic bladder stones and indwelling Foley catheters. The chemotherapy agent cyclophosphamide can increase the risk for bladder cancer nine-fold when used chronically, and phenacetin-containing compounds have been implicated in the development of renal pelvis and ureteral tumors.

PATHOBIOLOGY

Urothelial tumors occur anywhere along the urinary tract, including the renal pelvis, ureters, bladder, and the urethra. Over 90% of tumors originate in the bladder, 8% in the renal pelvis, and the remaining 2% originate in the ureter and urethra. Transitional cell carcinomas comprise 90 to 95% of urothelial tumors; squamous cell (keratinizing) tumors (3%), adenocarcinomas (2%), and small cell tumors (1%) are the remainder. Mixed-histology tumors, consisting of predominantly transitional cell carcinoma with areas of squamous, adenocarcinomatous, or neuroendocrine elements are frequently observed. Squamous cell tumors are more frequent in the distal urethra, and adenocarcinomas occur in the embryonal remnant of the urachus on the dome of the bladder and in periurethral tissues. In endemic areas of *S. haematobium* infections (such as Egypt), 40% of tumors are squamous cell carcinomas. Rare tumors of the bladder include lymphoma, sarcoma, and melanoma.

The majority (70-80%) of newly detected bladder cancers are classified as non–muscle invasive tumors and include exophytic papillary tumors confined to the mucosa (Ta), tumors invading the lamina propria (T1), and carcinoma in situ (CIS). Non–muscle invasive bladder tumors are typically graded according to the World Health Organization/International Society of Urologic Pathology (WHO/ISUP) grading system as low-grade and high-grade. If the grading system is not specified, a numeric system can be used: well differentiated (G1), moderately differentiated (G2), poorly differentiated (G3), and undifferentiated (G4). Grading is more important for noninvasive Ta tumors because almost all invasive bladder tumors (T1 or greater) are high grade. Primary CIS, or Tis, without a concurrent Ta or T1 tumor, constitutes 1 to 2% of new bladder cancer cases. More frequently, Tis is found in the presence of multiple papillary tumors, either immediately adjacent to another lesion or involving remote mucosa in the bladder. Tis is, by definition, high-grade disease; it is regarded as a precursor to more invasive tumors because 60% of untreated tumors develop more invasive disease within 5 years. T1 tumors are an aggressive, invasive malignancy. Virtually all T1 tumors are high grade, and 50% have associated Tis. Disease in 50% of patients recurs by 1 year and in 90% within 5 years. A minority of primary tumors at diagnosis is found to invade the muscularis propria (T2), extend to perivesicular fat (T3), or extend into immediately adjacent organs (T4); all primary tumors stage T2 or higher are high grade.

The natural history of a urothelial tumor is to recur either at the same location or at a separate site in the urothelial tract and at the same or a more advanced stage. Several studies support the controversial concept that these recurrences are clonal in origin. The epidermal growth factor receptor (EGFR) is highly expressed (~80%) in bladder cancers; the Her2/Neu growth factor receptor is less frequently expressed (~50-70%). Studies suggest that higher expression of these receptors is associated with a more advanced and/or more aggressive phenotype of disease. Bladder cancer has a very high somatic mutation rate (7.7 per megabase) compared to that of other cancers, exceeded only by lung cancer and melanoma.[8] Epigenetic modifying genes *MLL2, ARID1A, KDM6A,* and *EP300* are significantly mutated in bladder cancer; approximately 75% have at least one inactivating mutation. Genes that regulate the cell cycle are also frequently mutated, including *TP53* in 49% and *RB1* in 13% of tumors. Amplifications of *ERBB2, MDM2,* and *EGFR* occur in a minority of tumors and represent potential therapeutic targets.

CLINICAL MANIFESTATIONS

Hematuria is the manifesting symptom in 80 to 90% of bladder cancer cases, but other patients may present with a urinary tract infection. Individuals over 40 years of age who develop hematuria should have an evaluation for the presence of urothelial cancer that includes urinary cytology, cystoscopy, and imaging of the urinary tract by either an ultrasound or CT scan. Screening of asymptomatic individuals for hematuria increases the probability of diagnosing bladder cancer at an earlier stage but does not improve survival; thus, it is not routinely recommended. Urinary frequency and nocturia may be present either as a consequence of irritative symptoms or a reduced bladder capacity. Pain, when present, typically reflects the location of the bladder tumor. Lower abdominal pain may occur as a result of a bladder mass, and rectal discomfort and perineal pain can result from tumors invading the prostate or pelvis. Tumors of the renal pelvis, ureter, or bladder in which the ureteral orifice is obstructed can cause hydronephrosis, reduced renal function, and flank pain. Patients with more advanced disease can present with anorexia, fatigue, weight loss, or pain from a metastatic bone lesion. The physical examination is frequently unremarkable in patients presenting with bladder tumors because the vast majority of patients have organ-confined tumors.

DIAGNOSIS

The mainstay of bladder cancer diagnosis and staging is the cystoscopic evaluation. The procedure includes examination under anesthesia to determine if a palpable mass (either mobile or not) is present. A nonmobile tumor mass is indicative of disease invading the pelvic sidewall that is unlikely to be resectable. Urine is obtained to evaluate for the presence of malignant cells. A cystoscope is inserted to visually inspect the bladder and detail the size, number, location, and growth pattern (papillary or solid) of all lesions. All visible disease undergoes transurethral resection of the bladder tumor to determine the histologic subtype and depth of invasion. Adequate evaluation, particularly in large tumors that may be invasive, requires that muscle is identified in the pathologic specimen. Repeat biopsy of the resected area is occasionally required to ensure that no muscle invasion is present, because invasion into muscle requires consideration of surgical removal of the bladder rather than endoscopic resection of the tumor. Biopsies from any areas of erythema are performed to assess for CIS. The urethra is inspected during withdrawal of the cystoscope, and biopsies are taken if clinically indicated. Patients with a positive cytologic findings but no apparent tumor within the bladder undergo selective retrograde catheterization of the ureters up to the renal pelvices to determine whether upper tract disease is present.

The decision whether to obtain images of the abdomen and pelvis is based on the cystoscopy results and the pathology of the tumor. Either a CT or magnetic resonance urogram can evaluate the upper urinary tracts, and a CT or MRI may distinguish whether a tumor extends to the perivesical fat (T3), prostate, or vagina (T4) and whether regional lymph nodes are involved (N+). In the case of larger, invasive tumors, the presence or absence of distal metastases can be documented with physical examination, CT of the abdomen and pelvis, a chest radiograph, and radionuclide bone scan.

All patients with carcinoma of the bladder or related sites are staged using the TNM system, advocated by the American Joint Committee on Cancer (AJCC) (Table 197-5). The TNM system categorizes the depth of invasion of the primary tumor, nodal metastases in the pelvis (or retroperitoneum for upper tract disease) on the basis of the number and size of regional nodal involvement, other nonregional lymph node sites, and any visceral sites of disease.

TREATMENT

Non–Muscle Invasive Disease

The standard treatment for non–muscle invasive tumors is a complete endoscopic resection. The majority of patients develop new tumors, 30% of which progress to a higher stage, mandating vigilant surveillance at 3-month intervals with cystoscopy, urine cytology, and repeat transurethral resection when indicated. Additional treatment in the form of adjuvant intravesical therapy depends on the number of lesions, the size, the depth of invasion, and the number of prior tumors in that individual. Prophylactic or adjuvant intravesical therapy is typically instituted in the setting when a patient has shown either a repeated tendency to develop new lesions in the bladder or is at high risk for recurrence or progression de novo. Intravesical therapy is not warranted for the first Ta tumor that is low grade. Instances of high recurrence and progression warranting intravesical therapy include multifocal or large lesions, high-grade papillary lesions, T1 tumors, CIS, or a combination of these. It is never advised for muscle-invasive tumors because agents instilled in the bladder do not penetrate beyond a few layers of cells. After allowing sufficient time for healing after the endoscopic resection, intravesical therapy is most frequently initiated with the immunologic agent bacillus Calmette-Guérin (BCG) weekly for 6 weeks, followed by a prolonged maintenance

TABLE 197-5 TNM DEFINITIONS FOR CANCERS OF THE BLADDER, URETER AND RENAL PELVIS

PRIMARY TUMORS OF THE BLADDER (T)	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Ta	Noninvasive papillary carcinoma
Tis	Carcinoma in situ (i.e., flat tumor)
T1	Tumor invades subepithelial connective tissue
T2	Tumor invades muscularis propria
pT2a	Tumor invades non-muscle invasive muscularis propria (inner half)
pT2b	Tumor invades deep muscularis propria (outer half)
T3	Tumor invades perivesical tissue
pT3a	Microscopically
pT3b	Macroscopically (extravesical mass)
T4	Tumor invades any of the following: prostatic stroma, seminal vesicles, uterus, vagina, pelvic wall, abdominal wall
T4a	Tumor invades the prostatic stroma, uterus, vagina
T4b	Tumor invades the pelvic wall, abdominal wall
REGIONAL LYMPH NODES FOR UROTHELIAL TUMORS OF THE BLADDER (N)	
NX	Regional lymph nodes cannot be assessed
N0	No lymph node metastasis
N1	Single regional lymph node metastasis in the true pelvis (hypogastric, obturator, external iliac, or presacral node)
N2	Multiple regional lymph node metastases in the true pelvis (hypogastric, obturator, external iliac, or presacral node)
N3	Lymph node metastasis to the common iliac lymph nodes
PRIMARY TUMORS OF THE URETER AND RENAL PELVIS (T)	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Ta	Papillary noninvasive carcinoma
Tis	Carcinoma *in situ*
T1	Tumor invades subepithelial connective tissue
T2	Tumor invades the muscularis
T3	(For renal pelvis only) Tumor invades beyond muscularis into peripelvic fat or the renal parenchyma
T3	(For ureter only) Tumor invades beyond muscularis into periureteric fat
T4	Tumor invades adjacent organs or through the kidney into perinephric fat

REGIONAL LYMPH NODES FOR UROTHELIAL TUMORS(N) OF URETER AND RENAL PELVIS	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in a single lymph node, ≤2 cm in greatest dimension
N2	Metastasis in a single lymph node, >2 cm but ≤5 cm in greatest dimension; or multiple lymph nodes, ≤5 cm in greatest dimension
N3	Metastasis in a lymph node, >5 cm in greatest dimension
DISTANT METASTASIS FOR ALL UROTHELIAL TUMORS (M)	
MX	Distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis
AJCC STAGE GROUPINGS FOR BLADDER CANCER	
0a	Ta, N0, M0
0is	Tis, N0, M0
I	T1, N0, M0
II	T2a, N0, M0
	T2b, N0, M0
III	T3a, N0, M0
	T3b, N0, M0
	T4a, N0, M0
IV	T4b, N0, M0
	Any T, N1-3, M0
	Any T, any N, M1
AJCC STAGE GROUPINGS FOR CANCER OF THE RENAL PELVIS AND URETER	
0a	Ta, N0, M0
0is	Tis, N0, M0
I	T1, N0, M0
II	T2, N0, M0
III	T3, N0, M0
IV	T4, N0, M0
	Any T, N1, M0
	Any T, N2, M0
	Any T, N3, M0
	Any T, any N, M1

From *AJCC Staging Manual*, 7th ed. New York: Springer-Verlag; 2010.

schedule.[9] Occasionally, other chemotherapeutic agents or cytokines are used when BCG is contraindicated. Treatment outcome is assessed at the 3- and 6-month evaluations following treatment to determine whether the bladder has been rendered tumor-free. If disease persists, either a repeat course of BCG treatment or even an immediate cystectomy may be recommended. Bladder toxicity caused by urothelial irritation can occur and includes bladder irritability or spasms, hematuria, and pain on urination. A rare complication of BCG is development of a systemic tuberculosis infection requiring treatment with systemic antituberculosis agents (Chapter 324). BCG is highly effective in eradicating Tis, with 70% of patients disease-free at 1 year and 40% at 10 years. Selected tumors in the ureter or renal pelvis can be managed by ureteroscopic resection, in some cases by instillation of BCG through the renal pelvis, or nephroureterectomy. Tumors of the prostatic urethra are frequently managed by cystoprostatectomy, particularly if a complete resection cannot be accomplished.

Muscle-Invasive Tumors

For patients with tumors infiltrating the muscularis propria, the standard of care in the United States is a radical cystectomy and pelvic lymphadenectomy because of the high incidence of cancer extending into the perivesicular fat or into regional lymph nodes. A prostatectomy is also performed in men; in women, the urethra, uterus, fallopian tubes, ovaries, and anterior vaginal wall are removed. Urinary flow can be directed through either a conduit diversion or a continent reservoir. With a conduit diversion, urine is drained directly from the ureters to a loop of small bowel that is anastomosed to the skin surface with no internal reservoir. Urine is collected in an external appliance. Alternatively, a low-pressure continent reservoir can be created from a detubularized segment of bowel attached to the abdominal wall with a continent stoma that can be self-catheterized at regular intervals. Low-pressure reservoirs also can be anastomosed to the urethra, creating an internal orthotopic neobladder that permits the patient to void via the normal urethra. The standard pelvic lymphadenectomy includes the distal common iliac, external iliac, obturator, and hypogastric nodes; improved survival and decreased local recurrence are associated with an increased number of lymph nodes removed. Complications of cystectomy include recurrent urinary infections, hyperchloremic acidosis, oxalate stones, incontinence, and impotence. Perioperative chemotherapy in addition to surgery is a standard of care for patients with muscle-invasive bladder cancer.[10] Neoadjuvant chemotherapy before cystectomy for muscle-invasive bladder cancer increases survival.[■] This survival benefit is achieved only with cisplatin-based combinations, which require that the patient has normal renal function and a good performance status. Some physicians prefer immediate cystectomy followed by adjuvant chemotherapy for tumors with a high risk for relapse. However, no evidence from prospective, randomized trials exists to support this approach in contrast to a proved survival benefit from neoadjuvant chemotherapy. Radical cystectomy is effective at providing long-term disease control in 75 to 80% of patients with organ-confined

disease; approximately 50% of those with tumors extending into the perivesical tissues; and up to a third of patients with regional lymph node involvement. Metastatic disease in the pelvic lymph nodes despite a normal preoperative CT scan is very frequent. Bilateral pelvic lymph node dissection in addition to cystectomy improves survival.[11] Some patients prefer a nonsurgical, bladder-sparing approach using radiation treatment rather than a cystectomy. Radiation with chemotherapy sensitization is preferred over radiation alone because of better tumor control. The best candidates for this approach are patients with a solitary early-stage lesion and no evidence of hydronephrosis. This tri-modality treatment for bladder preservation first requires a successful, near-complete transurethral resection of tumor followed by concurrent chemotherapy and radiation. External beam treatments are typically delivered in five daily fractions per week, ranging from 2.0 to 2.5 Gy, to a total treatment dose of approximately 65 Gy. Toxicities include inflammation of the skin, impotence, fatigue, and irritative symptoms from the bladder and bowel; persistent proctitis is rare. In this approach, cystectomy is reserved for patients whose disease failed to achieve complete response. The 5-year disease-free survival with this approach is 50%, with the majority of patients retaining a normally functioning bladder.

Metastatic Disease

Patients with metastatic disease are treated predominantly with chemotherapy. Cisplatin-based chemotherapy is the standard of care, and the two most commonly used regimens are gemcitabine plus cisplatin (GC) and the four-drug regimen of methotrexate, vinblastine, doxorubicin, and cisplatin (MVAC); six cycles of therapy are given over a 6-month period. The most frequently observed toxicities include anemia, thrombocytopenia, neutropenic fever, mucositis, and fatigue. The GC regimen is better tolerated and has less severe toxicities than MVAC. The median survival of patients treated with both regimens is approximately 14 months, and the 5-year survival is 15% or less. Patients with a good performance status and whose metastatic disease is limited to the lymph nodes (i.e., no visceral metastases) have the highest likelihood of response, and a 20 to 33% chance of 5-year disease-free survival.[12] The addition of paclitaxel to the GC doublet has not improved survival for all patients with metastatic bladder cancer but can be integrated with post-chemotherapy surgery in patients with local-regional metastases resulting in long-term cure.[13] Based on the high vascularity of bladder cancer and prior studies suggesting the benefit of bevacizumab which blocks the effects of VEGF, a national intergroup randomized trial is comparing the GC doublet plus either bevacizumab or placebo.[14] Multiple medical conditions preclude the use of cisplatin-based chemotherapy.[15] Patients with impaired renal function are treated with carboplatin-based therapy rather than cisplatin because it is less toxic to the kidneys. Carboplatin plus gemcitabine is the standard of care in patients ineligible for cisplatin.

PROGNOSIS

Cancer of the urinary bladder is a common but heterogeneous disease. Non–muscle invasive TaG1 lesions, easily treated with endoscopic resection alone, almost never progress. At the other end of the spectrum of non–muscle invasive disease, aggressive transitional cell carcinoma in situ requires intravesical immunotherapy with BCG in addition to endoscopic resection. This intravesical treatment can substantially reduce recurrence and progression, with 5-year disease-free survival rates of 60%. Muscle-invasive disease is most frequently cured with an integrated approach of systemic chemotherapy for micrometastases followed by cystectomy and pelvic lymphadenectomy; cure rates for T2 tumors can be as high as 80% with this multimodality approach. Bladder-sparing approaches associated with an improved quality of life are possible using external beam radiation. Metastatic urinary bladder cancer is a fast-growing and often lethal malignancy; despite aggressive chemotherapy only a small proportion (~15%) of patients are disease-free at 5 years.

CANCERS OF THE RENAL PELVIS AND URETERS

Approximately 10% of transitional cell carcinomas occur in the ureters and the renal pelvis. These tumors can arise either de novo or in the setting of prior tumors; the risk for developing an upper tract tumor in patients with multifocal carcinoma in situ of the bladder approaches 25% by 10 years. These tumors are morphologically similar to the tumors in the bladder and behave in a similar manner. Hematuria is the most common manifesting symptom, although patients with large tumors and/or ureteral obstruction can present with flank pain. A CT scan or MRI is used to stage the extent of primary disease and detect regional metastases. Low-grade tumors can be treated endoscopically, but high-grade tumors are most commonly treated with a nephroureterectomy. In contrast to cystectomy, a regional lymphadenectomy is not routinely performed. In renal pelvis tumors, the ureter is removed in addition to a nephrectomy because of the high risk for multifocal tumors along the entire upper tract and the inability to monitor the ureteral stump with accuracy. Systemic chemotherapy is used for unresectable primary tumors, patients with regional adenopathy, or recurrent tumors. Upper tract transitional cell carcinomas are staged according to the TNM system (see Table 197-5). Treatment for advanced, nonsurgical disease is with chemotherapy; these urothelial tumors have the same sensitivity to chemotherapy as bladder cancer, with similar response rates and 5-year survival rates. Cisplatin-based chemotherapy is used in patients with normal renal function, and carboplatin-based chemotherapy is considered if there is renal insufficiency from obstruction, a prior nephroureterectomy, or medical comorbidity.

Grade A References

A1. Motzer RJ, Hutson TE, Tomczak P, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med*. 2007;356:115-124.
A2. Rini BI, Escudier B, Tomczak P, et al. Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomized phase 3 trial. *Lancet*. 2011;378:1931-1939.
A3. Motzer RJ, Hutson TE, Cella D, et al. Pazopanib versus sunitinib in metastatic renal-cell carcinoma. *N Engl J Med*. 2013;369:722-731.
A4. Escudier B, Eisen T, Stadler WM, et al. Sorafenib for treatment of renal cell carcinoma: final efficacy and safety results of the phase III treatment approaches in renal cancer global evaluation trial. *J Clin Oncol*. 2009;27:3312-3318.
A5. Escudier B, Pluzanska A, Koralewski P, et al. AVOREN Trial investigators. Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *Lancet*. 2007;370:2103-2111.
A6. Hudes G, Carducci M, Tomczak P, et al. Global ARCC Trial. Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *N Engl J Med*. 2007;356:2271-2281.
A7. Motzer RJ, Escudier B, Oudard S, et al. RECORD-1 Study Group. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet*. 2008;372:449-456.
A8. International Collaboration of Trialists. International phase III trial assessing neoadjuvant cisplatin, methotrexate, and vinblastine chemotherapy for muscle-invasive bladder cancer: long-term results of the BA06 30894 trial. *J Clin Oncol*. 2011;29:2171-2177.
A9. James ND1, Hussain SA, Hall E, et al. Radiotherapy with or without chemotherapy in muscle-invasive bladder cancer. *N Engl J Med*. 2012;366:1477-1488.
A10. De Santis M, Bellmunt J, Mead G, et al. Randomized phase II/III trial assessing gemcitabine/carboplatin and methotrexate/carboplatin/vinblastine in patients with advanced urothelial cancer who are unfit for cisplatin-based chemotherapy: EORTC study 30986. *J Clin Oncol*. 2012;30:191-199.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Delahunt B, Cheville JC, Martignoni G, et al. The International Society of Urological Pathology (ISUP) grading system for renal cell carcinoma and other prognostic parameters. *Am J Surg Pathol*. 2013;37:1490-1504.
2. Siegel R, Ma J, Zou Z, et al. A Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64:9-29.
3. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*. 2013;499:43-49.
4. Daugherty M, Bratslavsky G. Compared with radical nephrectomy, nephron-sparing surgery offers a long-term survival advantage in patients between the ages of 20 and 44 years with renal cell carcinomas (≤4cm): An analysis of the SEER database. *Urol Oncol*. 2014;32:549-554.
5. Sprenkle PC, Power N, Ghoneim T, et al. Comparison of open and minimally invasive partial nephrectomy for renal tumors 4-7 centimeters. *Eur Urol*. 2012;61:593-599.
6. Shablak A1, Sikand K, Shanks JH, et al. High-dose interleukin-2 can produce a high rate of response and durable remissions in appropriately selected patients with metastatic renal cancer. *J Immunother*. 2011;34:107-112.
7. Albiges L, Choueiri T, Escudier B, et al. A systematic review of sequencing and combinations of systemic therapy in metastatic renal cancer. *Eur Urol*. 2015;67:100-110.
8. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*. 2014;507:315-322.
9. Zhu S, Tang Y, Li K, et al. Optimal schedule of bacillus Calmette-Guérin for non-muscle-invasive bladder cancer: a meta-analysis of comparative studies. *BMC Cancer*. 2013;13:332-347.
10. Meeks JJ, Bellmunt J, Bochner BH, et al. A systematic review of neoadjuvant and adjuvant chemotherapy for muscle-invasive bladder cancer. *Eur Urol*. 2012;62:523-533.
11. Tilki D, Brausi M, Colombo R, et al. Lymphadenectomy for bladder cancer at the time of radical cystectomy. *Eur Urol*. 2013;64:266-276.
12. Apolo AB, Ostrovnaya I, Halabi S, et al. Prognostic model for predicting survival of patients with metastatic urothelial cancer treated with cisplatin-based chemotherapy. *J Natl Cancer Inst*. 2013;105:499-503.
13. Bellmunt J, von der Maase H, Mead GM, et al. Randomized phase III study comparing paclitaxel/cisplatin/gemcitabine and gemcitabine/cisplatin in patients with locally advanced or metastatic urothelial cancer without prior systemic therapy: EORTC Intergroup Study 30987. *J Clin Oncol*. 2012;30:1107-1113.
14. Balar AV, Apolo AB, Ostrovnaya I, et al. Phase II study of gemcitabine, carboplatin, and bevacizumab in patients with advanced unresectable or metastatic urothelial cancer. *J Clin Oncol*. 2013;31:724-730.
15. Galsky MD, Hahn NM, Rosenberg J, et al. A consensus definition of patients with metastatic urothelial carcinoma who are unfit for cisplatin-based chemotherapy. *Lancet Oncol*. 2011;12:211-214.

REVIEW QUESTIONS

1. A patient with muscle-invasive urothelial cancer of the bladder wishes to have bladder-sparing treatment with chemotherapy plus radiation. Which of the following components is critical to the success of the therapy?

A. The radiation port includes the bladder, pelvic nodes, and a boost to the bladder.
B. Brachytherapy to the primary tumor, followed by a boost to the entire bladder.
C. Complete or near-complete transurethral resection of the bladder tumor (TURBT).
D. Adjuvant chemotherapy for three cycles of cisplatin-based chemotherapy.
E. Both A and C.

Answer: E A complete or near-complete TURBT is critical for success of this tri-modality approach (surgery, radiation, chemotherapy) to bladder-sparing treatment. Radiation is generally given to the bladder and pelvis, with a boost to the bladder cancer primary tumor.

2. A 56-year-old male patient with transitional cell carcinoma presents with carcinoma in situ involving the trigone of the bladder. The patient is treated with BCG therapy for two 6-week intervals without response. Repeat transurethral resection at 3 months shows persistent carcinoma in situ, along with new tumor invading the lamina propria but not the muscularis propria. The standard of care is:

A. Radical cystectomy with preservation of the prostate and neurovascular bundle to preserve sexual function.
B. Cystoprostatectomy and bilateral pelvic lymph node dissection.
C. Intravesical interferon or interleukin-2.
D. Systemic chemotherapy with a cisplatin-containing regimen.
E. External beam radiation with a chemotherapy radiosensitizer.

Answer: B The standard of care for refractory carcinoma in situ is cystoprostatectomy and bilateral pelvic lymph node dissection. Refractory disease is defined in this instance as progression to a greater stage within 6 months of therapy.

3. Interleukin-2 (IL-2) therapy for patients with renal cell carcinoma differs from drugs that target vascular endothelial growth factor (VEGF) in that:

A. IL-2 therapy can result in a 4% long-term cure in patients with metastatic disease.
B. IL-2 therapy has a greater response rate than VEGF inhibitors.
C. IL-2 therapy is given immediately after nephrectomy in contrast to VEGF inhibitors, which are given only for metastatic disease.
D. IL-2 can be given in patients with poor renal function.
E. IL-2 therapy is reserved for patients with brain metastases.

Answer: A The approval for IL-2 therapy by the U.S. Food and Drug Administration was, in part, based on this durability of complete responses in approximately 4% of patients, which is not seen with any other systemic treatments for metastatic renal cell carcinoma.

4. The preferred surgical management for a patient with a 7-cm renal cell carcinoma in one kidney and a simple cyst in the contralateral kidney includes:

A. Bilateral nephrectomy.
B. Open partial nephrectomy for the kidney with the 7-cm lesion.
C. Nephrectomy for the kidney with the 7-cm lesion and retroperitoneal lymph node dissection
D. Laparoscopic partial nephrectomy for the kidney with the 7-cm lesion.
E. Answers B and D.

Answer: E Partial nephrectomy to preserve renal function as much as possible is the preferred surgical management in patients with renal cell carcinoma. Simple cysts of the kidney are observed, and a retroperitoneal lymph node dissection is not the surgical standard of care for renal cell carcinoma.

5. A patient presenting with benign hair follicle tumors (fibrofolliculomas), pulmonary cysts with a history of pneumothorax in the past, and bilateral renal tumors has which of the following syndromes?

A. Hereditary papillary renal cell carcinoma (HPRCC)
B. Hereditary leiomyomatosis renal cell carcinoma (HLRCC)
C. Birt-Hogg-Dubé syndrome
D. Familial oncocytoma
E. Lynch syndrome

Answer: C Explanation: Birt-Hogg-Dubé syndrome is characterized by fibrofolliculomas, pulmonary cysts, pneumothorax, and bilateral of the renal tumors. Hereditary papillary renal cell carcinoma is associated with bilateral, multifocal tumors but not the other features. Hereditary leiomyomatosis renal cell carcinoma is associated with uterine leiomyomas (more common) or leiomyosarcoma (rare), cutaneous nodules (leiomyomas), and type 2 papillary renal cell carcinoma, which is frequently solitary and frequently develops metastases.

198

BREAST CANCER AND BENIGN BREAST DISORDERS

NANCY E. DAVIDSON

Invasive breast cancer, the most common nonskin cancer in women in the United States, will be diagnosed in approximately 232,000 women in 2014 and will result in approximately 40,000 deaths. Incidence and mortality from breast cancer appear to be dropping in the United States and parts of Western Europe. This decline is believed to reflect early detection by screening mammography and widespread use of adjuvant systemic therapy, as well as decreased use of hormone replacement therapy.

BREAST CANCER

EPIDEMIOLOGY AND PATHOBIOLOGY

Multiple risk factors for the development of breast cancer have been identified (Table 198-1). The principal risk factor is gender. Breast cancer is largely a disease of women, although it does occur in men at an incidence of approximately 1% that seen in women. A second critical risk factor is age. Approximately 75% of breast cancer cases in the United States are diagnosed in women older than 50 years of age.

Family history is a third critical risk factor. Approximately 20% of breast cancer occurs in women with a family history of breast cancer; increased risk

TABLE 198-1 RISK FACTORS FOR BREAST CANCER

RISK FACTOR	RELATIVE RISK
Any benign breast disease	1.5
Postmenopausal hormone replacement (estrogen with progestin)	1.5
Menarche at < 12 yr	1.1-1.9
Moderate alcohol intake (two to three drinks/day)	1.1-1.9
Menopause at > 55 yr	1.1-1.9
Increased bone density	1.1-1.9
Sedentary lifestyle and lack of exercise	1.1-1.9
Proliferative breast disease without atypia	2
Age at first birth > 30 yr or nulliparous	2-4
First-degree relative with breast cancer	2-4
Postmenopausal obesity	2-4
Upper socioeconomic class	2-4
Personal history of endometrial or ovarian cancer	2-4
Significant radiation to chest	2-4
Increased breast density on mammogram	2-4
Older age	>4
Personal history of breast cancer (in situ or invasive)	>4
Proliferative breast disease with atypia	>4
Two first-degree relatives with breast cancer	5
Atypical hyperplasia and first-degree relative with breast cancer	10

is associated with a diagnosis of breast cancer in first-degree relatives younger than 50 years. Of breast cancer cases, 5 to 8% occur in high-risk families. Several familial breast cancer syndromes with associated molecular abnormalities have been identified. Chief among them is the breast-ovarian cancer syndrome, which is linked to germline mutations in the breast cancer susceptibility genes, *BRCA1* and *BRCA2*. These mutations are inherited in an autosomal dominant fashion and can therefore be transmitted through the maternal or paternal line. Extensive studies suggest that a germline mutation in either of these genes is associated with a 50 to 85% lifetime risk for developing breast cancer. Testing for *BRCA1* and *BRCA2* mutations is now viewed as a standard option for women with clinical features suggestive of a hereditary breast cancer syndrome; these include multiple family members with early-onset breast or ovarian cancer, bilateral breast cancer, or Ashkenazi Jewish heritage. Careful counseling about the implications of a positive or negative test and about the limitations of testing is a prerequisite for testing.

Other hereditary cancer syndromes (Chapter 181) include germline loss-of-function mutations in *PALB2*,[1] the Li-Fraumeni syndrome (which is linked with germline mutations in the *p53* tumor suppressor gene), and Cowden syndrome (which is associated with inherited mutations in the *PTEN* gene). Finally, in addition to these high-penetrance genetic susceptibility syndromes, recent results from genome-wide association studies have identified a number of low-penetrance genetic associations, including single nucleotide polymorphisms in a variety of genes. If or how to incorporate these low-penetrance traits into clinical practice remains to be established. Whole-genome sequencing of breast cancers is exposing the scope of tumor diversity and helping to pinpoint avenues for precise diagnostics and targeted therapy.[2]

Reproductive risk factors include early menarche, late menopause, nulliparity, and late first pregnancy. In aggregate, these factors result in prolonged estrogen exposure of the breast. The emerging association between postmenopausal obesity and breast cancer likely reflects estrogen exposure as well. Certain types of breast pathology, including atypical hyperplasia and lobular carcinoma in situ, are also associated with increased risk. The possibility that increased breast density as assessed by mammography is a risk factor has also been raised. Finally, much interest has focused on the possibility that exogenous environmental factors predispose to breast cancer. Among the factors that appear to enhance breast cancer risk are ionizing radiation during adolescence, prolonged use of hormone replacement therapy, ongoing use of oral contraceptives, and alcohol consumption. Large studies have failed to show any convincing association between exposure to estrogenic pesticides or a high-fat diet and breast cancer.

CLINICAL MANIFESTATIONS

Breast cancer usually manifests as a mammographic abnormality or a physical change in the breast, including a mass or asymmetrical thickening, nipple discharge, or skin or nipple changes. Two unusual clinical manifestations include Paget disease of the nipple and inflammatory breast cancer. The former is a form of adenocarcinoma involving the skin and ducts and is manifested as nipple excoriation. The latter is recognized as a constellation of redness, warmth, and edema that often reflects tumor cell infiltration of dermal lymphatics of the breast; it should not be mistaken for simple mastitis.

Nipple discharge may be associated with breast malignancy. Although milky discharge is seldom associated with a malignant diagnosis, patients with a clear or bloody nipple discharge require breast examination and mammography and often excisional biopsy of any suspicious area. Ductography and sometimes ductoscopy may be used to identify the inciting lesion. A bloody discharge is frequently caused by an intraductal papilloma.

Breast pain is common, especially as a premenstrual symptom in premenopausal women. But it may also be associated with an underlying malignancy. Patients with localized noncyclic breast pain should undergo breast examination and bilateral mammography. If these are normal, ultrasound or magnetic resonance imaging (MRI) may be used to exclude the small possibility of a malignancy.

DIAGNOSIS

Diagnostic evaluation is generally triggered by suspicious findings on a screening mammogram or detection of a palpable breast abnormality by the patient or health care provider. For both clinically occult and clinically apparent lesions, pathologic evaluation is mandatory to establish a diagnosis. Today, fine-needle aspiration and core needle biopsy have replaced incisional or excisional biopsy as the standard diagnostic measures. These procedures can be performed in the office in patients with suspicious palpable lesions. For women with nonpalpable lesions, biopsy guided by mammography, ultrasonography, or MRI is now standard. Stereotactic- or ultrasound-guided core needle biopsies are almost as accurate as, and associated with lower complication rates than, open surgical biopsy. These technologies permit an accurate diagnosis that can be followed by definitive treatment planning. It is axiomatic, however, that further evaluation must be undertaken for suspicious lesions that give an equivocal diagnosis after needle aspiration or core biopsy. Finally, bilateral breast imaging is always recommended to identify any unsuspected lesions in the contralateral breast that may also require evaluation.

Staging and Prognostic and Predictive Markers

Although staging originally reflected the clinical assessment of tumor size, nodal status, and evidence of metastatic disease, pathologic staging is the most accurate estimate of tumor involvement and prognosis. The staging system for breast cancer was revised in 2010 (Tables 198-2 and 198-3).

Most patients with breast cancer present with stage I or II disease in the absence of symptoms. In these patients, laboratory studies can be limited to blood counts, chemistry panel, and chest radiograph, and more extensive radiologic evaluation is not warranted because of low yield. In contrast, women with clinical evidence of stage III or IV disease should undergo more intensive evaluation of common sites for metastases, including lung, liver, and bone, through computed tomography and radionuclide scanning.

The two most important determinants of prognosis for early-stage breast cancer are pathologic lymph node status and tumor size. Other factors that contribute to prognosis are the expression of the estrogen receptor-α (ER), progesterone receptor (PR), and HER2 proteins; these are conventionally measured by immunohistochemistry (IHC), although in situ hybridization (ISH) for *HER2* gene amplification is also employed. Poor prognosis is associated with high lymph node burden, poor histologic grade, large tumor size, absence of ER and PR expression, and overexpression of HER2.

Recently, the focus has been on the development of predictive markers to guide selection of therapy. The three established predictive markers for breast cancer are ER, PR, and HER2, and these should be routinely evaluated in every invasive cancer. Many tumors that express ER or PR, or both, are responsive to endocrine therapy, whereas those that lack ER and PR expression seldom respond to such therapy. Overexpression of the HER2 protein by IHC or *HER2* gene amplification by ISH is associated with response to the HER2-targeted therapies. Evidence that links expression of ER, PR, or HER2 to chemotherapy efficacy is equivocal.

TABLE 198-2 AMERICAN JOINT COMMITTEE ON CANCER TUMOR, NODE, METATASIS STAGING SYSTEM FOR BREAST CANCER

TNM STAGING

PRIMARY TUMOR (T)

Definitions for classifying the primary tumor (T) are the same for clinical and pathologic classification. If the measurement is made by the physical examination, the examiner uses the major headings (T1, T2, or T3). If other measurements, such as mammographic or pathologic measurements, are used, the subsets of T1 can be used. Tumors should be measured to the nearest 0.1-cm increment.

TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis Tis (DCIS) Tis (LCIS) Tis (Paget)	Carcinoma in situ Ductal carcinoma in situ Lobular carcinoma in situ Paget disease of the nipple with no tumor
Note: Paget disease associated with a tumor is classified according to the size of the tumor.	
T1	Tumor 2 cm or less in greatest dimension
T1mic	Microinvasion 0.1 cm or less in greatest dimension
T1a	Tumor more than 0.1 cm but not more than 0.5 cm in greatest dimension
T1b	Tumor more than 0.5 cm but not more than 1 cm in greatest dimension
T1c	Tumor more than 1 cm but not more 2 cm in greatest dimension
T2	Tumor more than 2 cm but not more than 5 cm in greatest dimension
T3	Tumor more than 5 cm in greatest dimension
T4	Tumor of any size with direct extension to (a) chest wall or (b) skin, only as described below
T4a	Extension to chest wall, not including pectoralis muscle
T4b	Edema (including peau d'orange) or ulceration of the skin of the breast or satellite skin nodules confined to the same breast
T4c	Both T4a and T4b
T4d	Inflammatory carcinoma

REGIONAL LYMPH NODES (N)

Clinical

NX	Regional lymph nodes cannot be assessed (e.g., previously removed)
N0	No regional lymph node metastasis
N1	Metastasis to movable ipsilateral axillary lymph node(s)
N2	Metastases in ipsilateral axillary lymph nodes fixed to one another (matted) or in clinically detected ipsilateral internal mammary nodes in the absence of clinically evident axillary lymph node metastasis
N2a	Metastases in ipsilateral axillary lymph nodes fixed to one another (matted) or to other structures
N2b	Metastases only in clinically apparent ipsilateral internal mammary nodes and in the absence of clinically evident axillary lymph node metastasis
N3	Metastasis in ipsilateral infraclavicular lymph node(s), with or without axillary lymph node involvement, or in clinically apparent ipsilateral internal mammary lymph node(s) and in the presence of clinically evident axillary lymph node metastasis; or metastasis in ipsilateral supraclavicular lymph node(s), with or without axillary or internal mammary lymph node involvement
N3a	Metastasis in ipsilateral infraclavicular lymph node(s)
N3b	Metastasis in ipsilateral internal mammary lymph node(s) and axillary lymph node(s)
N3c	Metastasis in ipsilateral supraclavicular lymph node(s)

Pathologic (pN)

pNX	Regional lymph nodes cannot be assessed (e.g., previously removed or not removed for pathologic study)
pN0	No regional lymph node metastasis histologically
pN0(i–)	No regional lymph node metastasis histologically, negative IHC
pN0(i+)	Malignant cells in regional lymph node(s) 0.2 mm or less (detected by IHC including isolated tumor cell(s)
pN0(mol–)	No regional lymph node metastasis histologically, negative molecular findings (RT-PCR)
pN0(mol+)	No regional lymph node metastasis by IHC or histology, positive molecular findings (RT-PCR)
pN1	Metastasis in 1 to 3 axillary lymph nodes, or in internal mammary nodes with microscopic disease detected by sentinel lymph node dissection, but not clinically detected
pN1mi	Micrometastasis (greater than 0.2 mm and/or more than 200 cells, none greater than 2.0 mm)
pN1a	Metastasis in 1 to 3 axillary lymph nodes with a least 1 metastasis greater than 2 mm
pN1b	Metastasis in internal mammary nodes with micrometastases or macrometastases detected by sentinel lymph node dissection but not clinically detected
pN1c	Metastasis in 1 to 3 axillary lymph nodes and in internal mammary nodes with micrometastases or macrometastases detected by sentinel lymph node dissection, but not clinically detected
pN2	Metastasis in 4 to 9 axillary lymph nodes or in clinically detected internal mammary lymph nodes in the absence of axillary lymph node metastasis
pN2a	Metastasis in 4 to 9 axillary lymph nodes (at least 1 tumor deposit greater than 2.0 mm)
pN2b	Metastasis in clinically detected internal mammary lymph nodes in the absence of axillary lymph node metastasis
pN3	Metastasis in 10 or more axillary lymph nodes, or in infraclavicular lymph nodes, or in clinically detected ipsilateral internal mammary lymph nodes in the presence of 1 or more positive axillary lymph nodes; or in more than 3 axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected; or in ipsilateral supraclavicular lymph nodes
pN3a	Metastasis in 10 or more axillary lymph nodes (at least 1 tumor deposit greater than 2.0 mm), or metastasis to the infraclavicular lymph nodes
pN3b	Metastasis in clinically detected ipsilateral internal mammary lymph nodes in the presence of 1 or more positive axillary lymph nodes; or in more than 3 axillary lymph nodes and in internal mammary lymph nodes with micrometastases or macrometases detected by sentinel lymph node dissection, but not clinically detected
pN3c	Metastasis in ipsilateral supraclavicular lymph nodes

DISTANT METASTASIS (M)

MX	Distant metastasis cannot be assessed
M0	No clinical or radiographic evidence of distant metastasis
cM0(+)	No clinical or radiographic evidence of distant metastases but deposits of molecularly or microscopically detected tumor cells in blood, bone marrow, or nonregional nodes without symptoms or signs of metastases
M1	Distant detectable metastases as determined by classic clinical and radiologic means and/or histologically proved greater than 0.2 mm

STAGE GROUPING

Stage 0	Tis	N0	M0
Stage IA	T1	N0	M0
Stage IB	T0 or T1	N1mi	M0
Stage IIA	T0	N1	M0
	T1	N1	M0
	T2	N0	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T0	N2	M0
	T1	N2	M0
	T2	N2	M0
	T3	N1	M0
	T3	N2	M0
Stage IIIB	T4	N0	M0
	T4	N1	M0
	T4	N2	M0
Stage IIIC	Any T	N3	M0
Stage IV	Any T	Any N	M1

Note: Stage designation may be changed if postsurgical imaging studies show distant metastases, provided that the studies are carried out within 4 months of diagnosis in the absence of disease progression and provided that the patient has not received neoadjuvant therapy.

IHC = immunohistochemistry; RT-PCR = reverse-transcriptase polymerase chain reaction; TNM = tumor, node, metastasis.

TABLE 198-3 TUMOR, NODE, METASTASIS (TNM) STAGE AND SURVIVAL

STAGE	TNM CATEGORY*	RECURRENCE FREE AT 10 YEARS (NO SYSTEMIC ADJUVANT THERAPY) (%)
0	TisN0M0	98
I	T1N0M0	80 (all stage I patients)
	T ≤ 1 cm	90%
	T > 1-2 cm	80-90
IIA	T0N1M0; T2N0M0	60-80
IIA	T1N1M0	50-60
IIB	T2 N1M0	5-10 worse than IIA and based on node status
IIB	T3N0M0	30-50
IIIA	T0 or T1 or T2N2M0; or T3N1 or N2M0	10-40
IIIB	T4N0 or N1 or N2M0	5-30
IIIC	Any T, N3M0	15-20
IV	Any T, any NM1	<5

*See Table 198-2 for TNM definitions.
TNM = tumor, node, metastasis.

TABLE 198-4 CARCINOMA IN SITU: DUCTAL VERSUS LOBULAR

FEATURE	LOBULAR CARCINOMA IN SITU	DUCTAL CARCINOMA IN SITU
Age	Younger	Older
Palpable mass	No	Uncommon
Mammographic appearance	Not detected on mammography	Microcalcifications, mass
Immunophenotype	E-cadherin negative	E-cadherin positive
Usual manifestation	Incidental finding on breast biopsy	Microcalcifications on mammography or breast mass
Bilateral involvement	Common	Uncertain
Risk and site of subsequent breast cancer	25% risk for invasive breast cancer in either breast over remaining lifespan	At site of initial lesion; 0.5% risk/yr of invasive breast cancer in opposite breast
Prevention	Consider tamoxifen or raloxifene or aromatase inhibitor	Consider tamoxifen or raloxifene if estrogen receptor positive
Treatment	Yearly mammography and breast examination	Lumpectomy ± radiation; mastectomy for large or multifocal lesions

Modern molecular techniques have provided further insight into molecular classification of breast cancer. Integrating data derived from analysis by genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA sequencing, and reverse-phase protein arrays, multiple genetically distinct types of breast cancer have been identified.[3,4] Transcriptional profiling has suggested that breast cancers can be divided into at least four molecular subsets: luminal A and B, HER2, and basal. The luminal subtypes frequently express ER, but luminal A appears to be associated with a better prognosis and higher likelihood of response to endocrine therapy than luminal B. The basal subtype is dominated by tumors that lack expression of ER, PR, and HER2—the so-called triple-negative breast cancer that lacks a readily identified molecular target. Multigene assays that evaluate these gene expression patterns are under investigation, and several multigene assays are available in clinical practice. One such assay, Oncotype Dx, may assist in the identification of women with early-stage steroid receptor–positive breast cancer who would benefit from the addition of chemotherapy to tamoxifen. A second assay, Mammaprint, may be useful to identify young women with breast cancer with poor prognosis. A number of other assays are under development, and both the Oncotype Dx and Mammaprint assays are the subject of large randomized trials to refine the conditions for their optimal use.

TREATMENT

Local Treatment of Early-Stage Breast Cancer

In Situ Carcinoma

Thanks to heightened breast cancer awareness and use of screening mammography, in situ carcinomas now account for 20 to 25% of newly diagnosed cases of breast cancer (Table 198-4). Most of these are ductal carcinoma in situ (DCIS). These lesions are associated with approximately a 30% risk for subsequent invasive breast cancer in the same breast. The risk for metastatic breast cancer with a diagnosis of DCIS is extremely small. As a consequence, management decisions are centered on the involved breast, and axillary lymph node evaluation is not routinely performed. Total mastectomy, the traditional therapy, has a high likelihood of cure, but studies suggest that breast conservation is appropriate for many women with DCIS. The major contraindications include poor cosmesis, extensive disease, or patient preference. Several models have suggested that size and grade of lesion and surgical margin status are important determinants of local outcome. Excision to obtain tumor-free margins is critical. Careful mammographic examination of the specimen and postexcision mammography of the breast are crucial to confirm that the DCIS has been adequately excised. A large randomized trial showed that radiotherapy plus lumpectomy decreased the likelihood of in situ or invasive recurrence compared with lumpectomy alone. Other data sets suggest that some women with favorable histologic findings who are willing to undergo close surveillance are candidates for local excision alone. In addition, the use of tamoxifen for 5 years can reduce ipsilateral breast cancer recurrence and contralateral breast cancer diagnosis by approximately 50%.

Controversy continues over whether lobular carcinoma in situ (LCIS) is truly a malignant lesion. LCIS is usually an incidental finding on a breast biopsy done for other indications, and it appears to be associated with a 25% risk for development of invasive breast cancer in either breast. Women with LCIS are generally managed expectantly with regular breast examination and mammography. Bilateral total mastectomy is sometimes considered for women with LCIS who have other risk factors or extreme anxiety. Finally, these women are candidates for tamoxifen or raloxifene or an aromatase inhibitor as a risk-reduction strategy based on the results of several large breast cancer chemoprevention trials.

Invasive Breast Cancer

Surgery

Although radical mastectomy (removal of the breast, axillary contents, and underlying chest musculature) was the mainstay for breast cancer treatment for many years, it is seldom performed today. Multiple randomized trials have consistently shown that breast conservation therapy (BCT) with lumpectomy plus radiotherapy provides identical survival rates to modified radical mastectomy (removal of the breast and lymph nodes) for women with stages I and II breast cancer. Medical contraindications to BCT include multifocal disease, previous radiotherapy, ongoing pregnancy that precludes the timely use of radiotherapy, poor cosmesis, and patient preference. Although the number of patients who receive BCT has increased substantially, there are wide geographic differences within the United States. Patients who undergo mastectomy should be counseled about the availability of a number of autologous tissue and implant options for reconstruction, either at the time of surgery or any time thereafter.

Because the likelihood of distant micrometastatic spread is highly correlated with the number of pathologically involved axillary lymph nodes, axillary dissection has traditionally been used to provide prognostic information. A drive toward limiting axillary surgery to minimize the incidence of postoperative lymphedema (see later) has led to the development of sentinel node techniques. A radioactive tracer, blue dye, or both are injected into the area around the primary breast tumor. The injected substance tracks rapidly to the dominant axillary lymph node—the sentinel node—which can be located and removed by the surgeon. If the sentinel node is tumor free, the remaining nodes are likely to be tumor free as well and no further axillary surgery is required. Currently, women with palpable axillary nodes and those with more extensively involved sentinel nodes are counseled to undergo axillary dissection. For women with small tumors and a clinically negative axilla, large randomized trials of sentinel node management and traditional axillary dissection suggest similar outcomes. Data suggest that it may also be possible to omit further axillary surgery for patients with a low burden of disease in sentinel nodes.

Adjuvant Radiotherapy

Radiotherapy has been a cornerstone in breast conservation therapy because women who undergo lumpectomy alone have a breast cancer recurrence rate of up to 40%, whereas the rate of recurrence is less than 10% with whole-breast radiotherapy. As a result, radiotherapy to the conserved breast reduces the breast cancer death rate by approximately 15%. Attempts to identify women whose tumors are so favorable that radiotherapy can be

withheld are ongoing. One large trial suggested that women older than 70 years with small ER-positive tumors who receive tamoxifen gain little with radiotherapy. Current research is also focused on the possibility that radiotherapy can be delivered safely and effectively to a smaller field (partial breast radiotherapy) or over a shorter period.

The role of postmastectomy radiotherapy continues to be a matter of debate. Based on the results of individual randomized trials and a meta-analysis suggesting a survival advantage, many radiation oncologists recommend postmastectomy radiation to women with more than three involved nodes and discuss its use for those with involvement of one to three nodes, in whom a smaller benefit is seen.

Adjuvant Systemic Therapy for Early-Stage Breast Cancer

Adjuvant systemic therapy is defined as the use of chemotherapy, endocrine therapy, or biologic therapy, or a combination of these, after definitive local therapy for early breast cancer. Its goal is to suppress or eradicate clinically occult micrometastases. Because current testing does not permit the definitive identification of the patient with micrometastases, recommendations for adjuvant systemic therapy are based on menopausal status, lymph node status, tumor size, and the extent of expression of the ER, PR, and HER2 proteins in breast cancer cells. The treatment algorithms that are currently used are the result of more than 50 years of clinical trials; the results of these trials have been compiled in sequential overview analyses that have evaluated the worldwide experience with use of endocrine therapy and chemotherapy. These analyses have shown that adjuvant therapy results in a proportional reduction in risk for recurrence across all patients regardless of risk for recurrence; this implies that the absolute benefit of adjuvant systemic therapy is greatest for individuals with the greatest risk for recurrence. Tools to assist the clinician and patient to make decisions about the use of adjuvant therapy include guidelines based on evidence and expert consensus such as the National Comprehensive Care Network (NCCN) and St. Gallen conference guidelines, as well as web-based algorithms such as Adjuvant Online (Table 198-5).[5]

Adjuvant Endocrine Therapy

Tamoxifen (20 mg/day for 5 years) has been the most widely used endocrine therapy. It improves outcomes in women of all ages with ER- or PR-positive breast cancer. Its side-effect profile includes an increased risk for thromboembolic events and uterine cancer, especially in postmenopausal women, because of its estrogen agonist properties. Potential benefits include promotion of bone density and lowering of cholesterol. An active area of research is the effects on outcome of altered activity of the tamoxifen metabolizing enzyme, CYP2D6, by either coadministration of pharmacologic inhibitors or by single-nucleotide polymorphism variants in the *CYP2D6* gene.

In recent years, the role of estrogen deprivation has been the subject of intense scrutiny—ovarian suppression or ablation for premenopausal women and aromatase inhibition for postmenopausal women. Ovarian ablation through surgery or radiotherapy is the oldest form of systemic therapy for breast cancer. More recent work has focused on the use of luteinizing hormone–releasing hormone (LHRH) agonists as a means of effecting a temporary and reversible ovarian suppression. Two large meta-analyses sought to define the role of these approaches. A meta-analysis of trials addressing the efficacy of LHRH agonists in women with early-stage ER-positive breast cancer suggested the following: (1) monotherapy with LHRH agonist has significant activity; (2) the efficacy of LHRH monotherapy is similar to that of certain chemotherapy regimens; and (3) LHRH agonists appear to add benefit to adjuvant chemotherapy, especially in women younger than 40 years, who are less likely than older women to become postmenopausal as a consequence of adjuvant chemotherapy. Unfortunately, these trials did not routinely include tamoxifen because its value in premenopausal breast cancer was recognized only after accrual for these ovarian suppression studies was completed. In sum, however, it appears that ovarian ablation or suppression is a viable strategy for premenopausal women with steroid receptor–positive breast cancer.

In premenopausal women, the combination of LHRH agonist plus aromatase inhibitor may be better than LHRH agonist plus tamoxifen. For example, in a recent randomized trial of hormone-receptor-positive early breast cancer in premenopausal women, adjuvant treatment with the aromatase inhibitor exemestane plus ovarian suppression, as compared with tamoxifen plus ovarian suppression, significantly reduced recurrence.

In postmenopausal women, the primary source of estrogen is the conversion of androgens synthesized by the adrenal glands to estrogen through the activity of CYP19 or aromatase in peripheral tissues such as mammary and adipose tissues. The aromatase inhibitors (anastrozole, letrozole, and exemestane) specifically inhibit this conversion, leading to further estrogen deprivation in older women. Randomized trials have shown that efficacy of aromatase inhibitors is similar or superior to that of tamoxifen and that these drugs have an acceptable side-effect profile. Multiple trials have compared monotherapy with tamoxifen, aromatase inhibitor, or sequential therapy; in aggregate, they suggest that the use of an aromatase inhibitor at some point should be considered for most postmenopausal women with steroid receptor–positive invasive breast cancer.[6] Side effects include postmenopausal symptoms, osteoporosis and fractures, and arthralgias. Aromatase inhibitors are not useful for receptor-negative breast cancer, nor should they be used as monotherapy in premenopausal women. The simultaneous administration of tamoxifen plus aromatase inhibitor does not improve outcome over aromatase inhibitor alone.

Duration of endocrine therapy appears to be quite important. Direct evidence suggests that at least 5 years of adjuvant endocrine therapy is associated with better outcomes than shorter periods. Two large, recently reported trials suggest that 10 years of tamoxifen is better than 5 years. Optimal duration of aromatase inhibitor therapy is under study.

Adjuvant Anti-HER2 Therapy

Increased understanding of growth and death pathways of breast cancer led to the identification of critical nonendocrine pathways that are potential targets for therapy. The transmembrane HER2/neu protein is overexpressed in approximately 20% of breast cancers, generally because of gene amplification. The efficacy and safety of the monoclonal antibody trastuzumab in the treatment of women with HER2-overexpressing metastatic breast cancer laid the foundation for several adjuvant trials that in aggregate showed that the addition of 1 year of trastuzumab to chemotherapy reduced the risk for recurrence by approximately 50% in women with high-risk HER2-positive early breast cancer. Thus, use of trastuzumab is considered for many women with

TABLE 198-5 ADJUVANT TREATMENT GUIDELINES FOR PATIENTS WITH EARLY-STAGE INVASIVE BREAST CANCER*

PATIENT GROUP*	TREATMENT
FAVORABLE HISTOLOGY (TUBULAR OR COLLOID)	
ER- and/or PR-Positive Breast Cancer	
<1 cm and pN0 or pN1mi	No adjuvant therapy
1-2.9 cm and pN0 or PN1mi	Consider adjuvant hormonal therapy†
≥3 cm or node-positive	Adjuvant hormonal therapy ± adjuvant chemotherapy†
ER- and PR-Negative Breast Cancer	
<1 cm and pN0	No adjuvant therapy
1-2.9 cm	Consider adjuvant chemotherapy
≥3 cm or node-positive	Adjuvant chemotherapy
HORMONE RECEPTOR–POSITIVE (ER- AND/OR PR-POSITIVE) BREAST CANCER	
Lymph Nodes Negative	
≤0.5 cm	Consider adjuvant hormonal therapy
0.6-1.0 cm well differentiated and no unfavorable features‡	Consider adjuvant hormonal therapy†
>1 cm	Adjuvant hormonal therapy ± adjuvant chemotherapy†
Lymph Nodes Positive	
	Adjuvant hormonal therapy + adjuvant chemotherapy
HORMONE RECEPTOR–NEGATIVE (ER- AND PR-NEGATIVE) BREAST CANCER	
≤0.5 cm and pN0	No adjuvant therapy
0.6-1.0 cm and pN0 or pNmi	Consider chemotherapy
>1 cm or lymph-node positive	Adjuvant chemotherapy
HER2 POSITIVE	
	Trastuzumab should be added to the suggested treatment above for all node-positive patients and considered for pN0 tumors that are > 5 mm

Modified from National Comprehensive Cancer Network Guidelines. Available at http://www.nccn.org.
*Data are insufficient to make chemotherapy recommendations for patients 70 years and older. Treatment should be individualized for these patients based on life expectancy and comorbidity.
†In ER-positive or PR-positive patients, decisions regarding the added value of chemotherapy in addition to hormonal therapy alone in N0 patients can be aided by the use of the Oncotype Dx assay to assess recurrence score.
‡Unfavorable characteristics include high-grade tumor, blood vessel or lymphatic invasion by tumor, and high tumor proliferation rate (high S phase by flow cytometry or high Ki-67 value by immunohistochemistry) or HER2-positive status or high recurrence score.
ER = estrogen receptor; PR = progesterone receptor.

HER2-positive tumors. More recent trials suggest that 1 year of therapy is better than 6 months and 2 years is no better than 1 year.[■] Important questions remain regarding long-term risks and benefits, the use of trastuzumab in the absence of chemotherapy, and the role of other anti-HER2 agents such as lapatinib, pertuzumab, or trastuzumab, and emtansine in addition to or in place of trastuzumab.

Adjuvant Chemotherapy

Individual trials and the Early Breast Cancer Trialists Collaborative Group meta-analysis have shown the benefit of adjuvant chemotherapy. Benefit varies by age and nodal status such that, in the meta-analysis, the absolute benefit is greatest in women younger than 50 years, in whom 15-year breast cancer mortality decreased from 42% to 32%, whereas it decreased from 50% to 47% for women aged 50 to 69 years.

These trials established several principles that guide chemotherapy use. Combination therapy appears to be more effective than single-agent chemotherapy. Effective agents include anthracyclines, taxanes, antimetabolites, and cyclophosphamide. Randomized trials demonstrated that 3 to 6 months of therapy is preferred over longer durations. Dose reduction below the standard level is associated with inferior outcome, but dose escalation through the use of colony-stimulating factors or autologous stem cells support leads to excess toxicity without improved outcome. Regimens that use colony-stimulating factors to accelerate the schedule of chemotherapy administration have been more successful.

Increased use of adjuvant chemotherapy and longer survival led to concerns about toxicity. Acute side effects of therapy are nausea and vomiting, bone marrow suppression, and hair loss; all are reversible, and the first may be mitigated by the use of modern antiemetics. Careful use of colony-stimulating agents can minimize complications of neutropenia, but current evidence argues against the use of erythroid-stimulating agents for chemotherapy-induced anemia. Induction of menopause is a common concern for premenopausal women. Its likelihood is related to the type and duration of chemotherapy and the age of the patient; most women older than 40 years will suffer drug-induced menopause. Doxorubicin-related cardiomyopathy is noted in approximately 1% of women who received doxorubicin-containing adjuvant chemotherapy. Use of standard adjuvant chemotherapy regimens results in a very small increase in the incidence of acute leukemia, but there is no evidence of increased incidence of other second tumors. Effect of chemotherapy on cognitive function is an area of study.

Sequencing of Adjuvant Therapy

Women frequently receive several adjuvant interventions, including chemotherapy, radiotherapy, endocrine therapy, and/or anti-HER2 therapy. A logical question is how best to sequence these therapies.

Because a large randomized trial showed that concurrent chemotherapy plus tamoxifen led to worse outcome than chemotherapy followed by tamoxifen, most practitioners delay the administration of endocrine therapy until after completion of chemotherapy. Conversely, the benefit of trastuzumab appears to be greater when it is coadministered with taxane chemotherapy rather than following completion of taxane. A randomized trial showed no clear difference between the sequence of chemotherapy followed by radiotherapy compared with radiotherapy followed by chemotherapy.

A number of trials have tested the concept that administration of systemic therapy before primary surgery would improve outcome over the standard sequence of surgery followed by systemic therapy. Together they suggest that, compared with adjuvant therapy, preoperative (also termed *neoadjuvant*) systemic therapy improves the rate of breast conservation but does not enhance disease-free or overall survival. The possibility that preoperative therapy can provide an in vivo assessment of tumor response to therapy is suggested by the correlation between the finding of a pathologic complete response (absence of invasive cancer in the surgical specimen) and long-term disease-free survival in some studies.

Follow-Up of Early-Stage Breast Cancer Survivors

A critical question is how longitudinal medical follow-up should be conducted in women who have received appropriate local and systemic therapy for early breast cancer. Randomized trials have addressed the question of the value of serial laboratory and radiology testing, as well as the role of primary care versus oncology specialist follow-up. On the basis of these and other studies, the American Society of Clinical Oncology has published evidence-based guidelines for follow-up of asymptomatic survivors of early-stage breast cancer. These guidelines are summarized in Table 198-6.

Stage III Breast Cancer

Locally advanced or inoperable stage III breast cancer accounts for approximately 10% of breast cancers. It is characterized by large primary tumor, fixed tumor or lymph nodes, or neoplastic invasion of the skin or chest wall. Inflammatory breast cancer falls into this category. It has a clinical presentation of breast swelling, warmth, and erythema and may or may not be associated with a mass. Because as many as one third of women with locally advanced breast cancer have distant metastases at the time of diagnosis, many oncologists perform an evaluation for distant disease even in asymptomatic patients.

TABLE 198-6 FOLLOW-UP GUIDELINES FOR PATIENTS WITH EARLY-STAGE BREAST CANCER: AMERICAN SOCIETY OF CLINICAL ONCOLOGY GUIDELINES

PROCEDURE OR TEST	FREQUENCY
History and physical examination* (eliciting of symptoms of breast cancer)	Every 3-6 mo for first 3 yr, every 6-12 mo for next yr, then yearly
MAMMOGRAPHY	
Mastectomy patients	Yearly
Lumpectomy patients	Yearly
Pelvic examination	Age appropriate
Breast self-examination	Monthly
Complete blood cell counts and chemistry studies	The literature does not support the use of these tests
Chest radiography, bone scans, PET scans, breast MRI, liver imaging, and tumor marker studies	Not recommended for routine follow-up in asymptomatic patients
Patient education regarding signs and symptoms of recurrence	Each visit

Modified from Khatcheressian JL, Hurley P, Bantug E, et al. Breast cancer follow-up and management after primary treatment: American Society of Clinical Oncology clinical practice guideline update. *J Clin Oncol.* 2013;31:961-965.
*Limited evaluation: assess for pain, dyspnea, weight loss, and other major changes in function. The limited examination should include an assessment of nodes, axillae, lumpectomy or mastectomy site, chest, and abdomen. Patients should be instructed regarding symptoms of recurrence.
MRI = magnetic resonance imaging; PET = positron emission tomography.

Diagnosis is usually established by fine-needle aspiration or core biopsy, and combined-modality therapy is used to maximize control of local disease and distant micrometastases. Several months of preoperative endocrine therapy or chemotherapy results in tumor regression in most patients, thereby allowing definitive breast surgery of some type to be performed. Postoperative radiotherapy is generally employed to enhance local control, and some studies suggest that administration of further chemotherapy, hormone therapy, anti-HER2 therapy, or a combination thereof (depending on the features of the cancer) is then desirable. Multimodality therapy results in a 5-year disease-free survival rate of approximately 50%.

Stage IV or Metastatic Breast Cancer

Although seldom curable, advanced breast cancer is a highly treatable illness. Palliation or prevention of symptoms without excess toxicity is the primary goal of treatment. The median survival after diagnosis of metastatic breast cancer is 2 to 3 years, although the range is great, and a small cadre of long-term survivors has been described. Several recent clinical trials have documented small improvements in survival with some of the newer therapies.

Most women with metastatic breast cancer present with symptoms or abnormalities on physical examination. Less than 10% of women present initially with metastatic disease; rather, advanced disease is normally diagnosed in women with a previous diagnosis of early breast cancer for which they received treatment. Common sites for metastases include bone, soft tissues, lung, liver, and brain. If metastatic disease is suspected, relevant hematologic, biochemical, and radiographic evaluation is indicated to assess location and severity of involvement. Because of the import of the diagnosis, pathologic confirmation is preferred. This permits verification of recurrent disease, exclusion of other diagnoses, and reassessment of biologic features such ER, PR, and HER2. Elevation of tumor markers (e.g., CA-27-29, carcinoembryonic antigen) or the presence of circulating tumor cells is not diagnostic of recurrent disease, although these markers may be useful adjuncts in the assessment of the effects of therapy.

The role of surgery in metastatic breast cancer is limited. It is useful in certain circumstances such as resection of a chest wall nodule or solitary brain metastasis or orthopedic stabilization to treat or prevent a long-bone fracture. Radiotherapy is a mainstay in the management of advanced disease. It may be used at any time during the patient's course to treat localized disease such as chest wall recurrence, brain metastases, or painful bony metastases. Systemic treatment is the primary mode for management of disseminated disease. Key principles for selection of therapy include maximal palliation of symptoms, minimization of treatment-related toxicity, and prevention of disease complications. An algorithm for treatment of stage IV breast cancer is shown in Figure 198-1.

Endocrine Therapy

Endocrine therapy is preferred as the first intervention for metastatic breast cancer whenever feasible because of its favorable therapeutic index. Factors that support the use of endocrine therapy include the expression of hormone

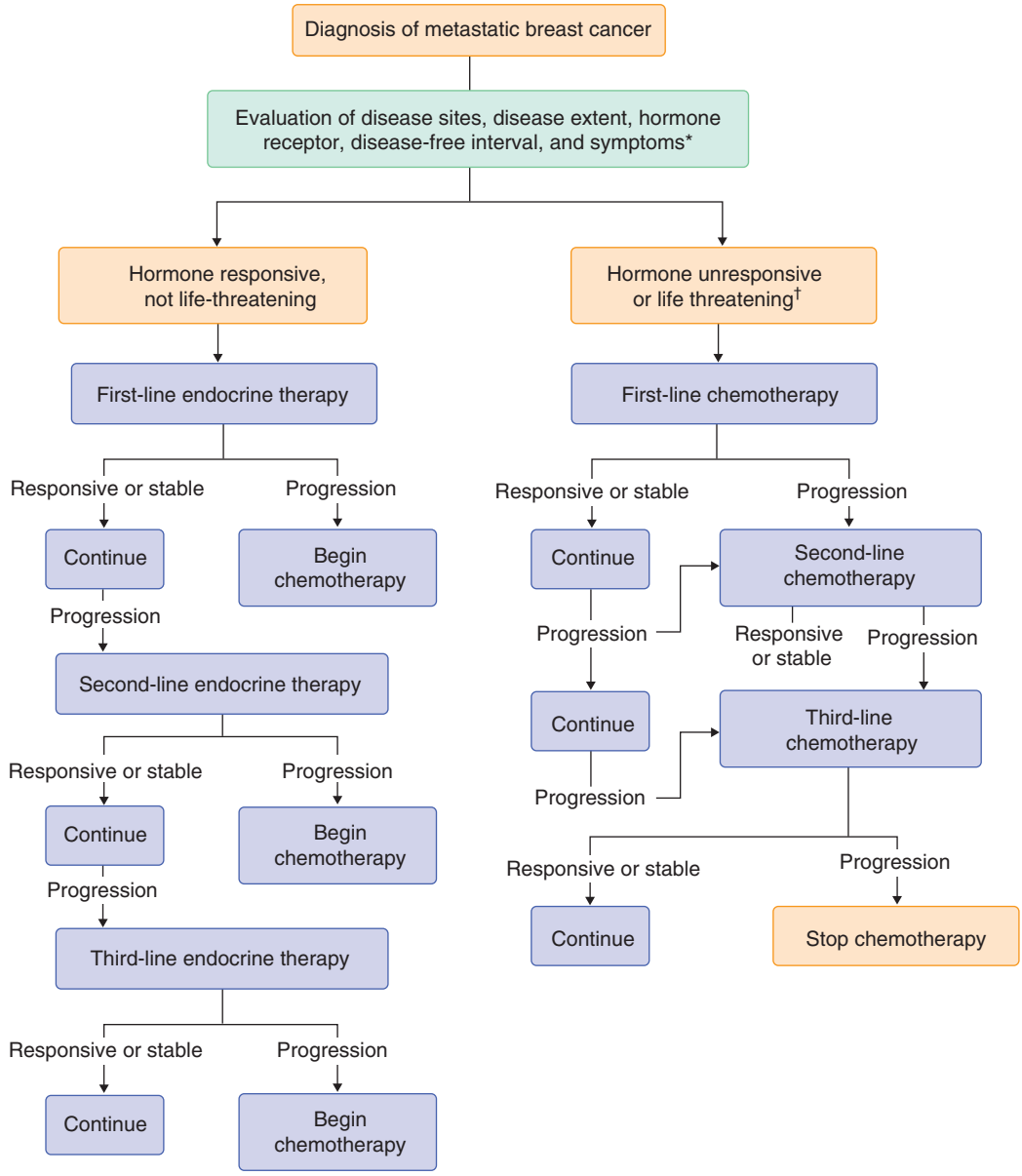

FIGURE 198-1. An algorithm used for the systemic treatment of stage IV breast cancer. *Consider use of bisphosphonate or denosumab if bone involvement. †Consider integration of anti-HER2 therapy if tumor is HER2 overexpressing.

receptors, a long disease-free interval, and absence of symptoms or visceral disease. More than half of the women who meet these criteria respond to an initial course of hormone therapy, with a median response duration of 9 to 12 months. Length of response is a good predictor for the likelihood of response to a second course of endocrine therapy when the first agent fails. A second course of endocrine therapy is less likely to be successful, and the duration of response is shorter; again, duration of response predicts for likelihood of success with third-line therapy. Successful application of this algorithm can result in good disease control with little toxicity for several years in some women.

A number of types of hormone therapy are now available. These include the selective estrogen receptor modulator (SERM) tamoxifen, the selective estrogen receptor–degrading agent fulvestrant, the aromatase inhibitors, and ovarian suppression by oophorectomy or LHRH agonists. Selection is usually made on the basis of efficacy, toxicity, and menopausal status. Serial administration of agents is the norm, although one study in estrogen receptor–positive postmenopausal metastatic breast cancer showed that the combination of the aromatase inhibitor anastrozole with fulvestrant was superior to anastrozole alone or sequential anastrozole and fulvestrant.[■] Other combination therapies (with the possible exception of ovarian suppression plus tamoxifen) do not improve outcomes. Recent work has also demonstrated the value of everolimus but not temsirolimus combined with an aromatase inhibitor to improve progression-free survival compared with an aromatase inhibitor alone in patients with hormone receptor–positive advanced breast cancer previously treated with nonsteroidal aromatase inhibitors.[■]

Several months of therapy are needed before the efficacy of a newly introduced endocrine therapy can be assessed. Patients and clinicians should be aware of the possibility of *treatment-related tumor flare*—a syndrome of worsening symptoms and increased circulating tumor markers—that can occur within the first few weeks of treatment. This response is usually of short duration and should not be confused with disease progression.

Chemotherapy

Most women with advanced breast cancer experience endocrine-unresponsive disease at some point and become candidates for palliative chemotherapy. Serial chemotherapy is the norm. Patients receive two to four cycles of therapy and then are evaluated for disease stabilization or improvement. Duration of therapy is variable for those who are responding to therapy. Several trials compared the approach of continuing therapy until time of disease progression with the approach of administering therapy, followed by a "drug holiday," with resumption of therapy at the time of disease progression. In sum, these studies suggest that survival is the same with these two approaches but that quality of life, as judged by the patients, is often better with continued therapy. Thus, decisions about continuation or cessation of a therapy are driven by perception of side effects and benefits by the patient and clinician.

Many active agents for breast cancer are now available. These include well-established drugs such as cyclophosphamide, doxorubicin, and methotrexate, as well as newer agents such as paclitaxel, docetaxel, vinorelbine, capecitabine, gemcitabine, carboplatin, ixabepilone, pegylated doxorubicin, and nanoalbumin–bound paclitaxel. All these agents are active individually and in combination. There is considerable debate about the value of combination therapy compared with sequential single-agent therapy for metastatic breast cancer. Current guidelines suggest the use of serial monotherapy unless the patient

has highly symptomatic disease or extensive visceral disease. In addition, much attention has been focused on schedule of administration. For example, weekly paclitaxel regimens appear to be more effective and better tolerated than regimens of every 3 weeks for many women. As with endocrine therapy, response rates and duration diminish with each successive change in therapy. As with early breast cancer, high-dose chemotherapy combined with autologous stem cell or bone marrow support has not shown benefit. A difficult question for patient and doctor is when to stop chemotherapy. No fixed rules exist, but many patients and physicians move to a program of supportive care if two successive chemotherapy regimens fail to produce a tumor response or disease stabilization.

Biologic Agents

Enhanced understanding of breast cancer biology has resulted in the identification of new targets for therapy beyond the estrogen receptor pathway. The first agent brought into clinical practice was the monoclonal antibody, trastuzumab (Herceptin), which is active against the transmembrane HER2/neu protein. Administration of trastuzumab monotherapy to women with metastatic HER2-overexpressing breast cancer led to partial or complete tumor regression in about 30% of women. Concurrent trastuzumab with paclitaxel increased response rate and duration and survival compared with paclitaxel alone for women with newly diagnosed HER2-overexpressing metastatic breast cancer. Similar results were seen with concurrent administration of trastuzumab and doxorubicin, but this combination was associated with a 20% incidence of congestive heart failure. This unexpected finding demonstrates the need for careful evaluation of new biologic agents as they enter the clinic. Dose and schedule of trastuzumab have been studied, and it appears that administration every 3 weeks is active with less patient inconvenience. Three other anti–HER2 agents are now available for treatment of advanced breast cancer in the United States, the oral small molecular inhibitor lapatinib; the monoclonal antibody pertuzumab, which inhibits HER dimerization; and the conjugate of a cytotoxic to trastuzumab, trastuzumab emtansine. How to combine and sequence these agents in the treatment of metastatic breast cancer is the subject of several clinical trials; in addition these agents are being tested in the neoadjuvant and adjuvant settings.

Strategies to block tumor angiogenesis have been explored. One such agent is bevacizumab, a monoclonal antibody that targets the VEGF. This agent has only modest activity as monotherapy in metastatic breast cancer. Its addition to taxane-based chemotherapy for women with newly diagnosed stage IV breast cancer appears to improve progression-free survival very modestly without a major impact on survival. Although two neoadjvuant studies have suggested that use of bevacizumab with chemotherapy can increase the rate of pathologic complete response in the surgical specimen compared with chemotherapy alone, trials of its use in the adjuvant setting are thus far negative.

Many other targeted agents are also under development. Of recent interest are the poly-ADP-ribose polymerase (PARP) inhibitors. These small molecule inhibitors of DNA repair show particular activity in women with *BRCA* mutant breast cancer in early studies; more definitive testing is in progress.

Supportive Care

Bone Health

Because palliation of symptoms and prevention of complications of metastatic disease are the primary goals of treatment for advanced breast cancer, careful attention to supportive care is vital. Bone is the most common site of metastasis in breast cancer, and bone disease can be a source of significant morbidity. Several studies have shown that regular administration of a bisphosphonate such as zoledronate or pamidronate or the anti-RANKL monoclonal antibody denosumab, in addition to endocrine therapy or chemotherapy, can reduce pain and lower the incidence of skeletal complications of disease. Although such therapy is now the norm, several issues remain unaddressed, including the optimal treatment interval and duration of therapy in metastatic breast cancer.

Given the propensity of breast cancer therapy to spread to bone and to lead to estrogen deprivation and osteopenia or osteoporosis (Chapter 243), the use of bisphosphonates in the adjuvant setting is under evaluation. Although individual trials have not shown improved outcomes with the use of bisphosphonates in the adjuvant setting, a recently presented meta-analysis suggests that these agents may be useful in the setting of estrogen deprivation such as in postmenopausal patients.

Postmenopausal Symptoms

Postmenopausal symptoms as a consequence of therapy or natural aging are common in breast cancer survivors. In general, hormone replacement therapy should be avoided in women with a history of breast cancer. A short course of treatment can be considered in patients with early-stage breast cancer who have truly disabling symptoms. Topical estrogens are considered for women with vaginal dryness that does not respond to lubricants. Vasomotor symptoms may be reduced with the use of certain antidepressants of the selective serotonin release inhibitor family. Multiple studies have failed to show consistent benefit from many alternative therapies.

Lymphedema

Lymphedema develops in the ipsilateral arm in up to 15% of women after treatment for early breast cancer. Its incidence is lower with sentinel lymph node procedures and meticulous radiotherapy planning. Prevention includes avoidance of trauma and possibly exercises. Early recognition of symptoms is key. Affected patients should be referred to specialists for consideration of treatment such as manual drainage or the use of compression stockings or pumps.

Special Circumstances

The risk for breast cancer is slightly increased during and just after pregnancy. Suspicious breast findings during pregnancy should be investigated vigorously. Surgical therapy for breast cancer can be safely performed after the first trimester of pregnancy, but radiotherapy should be delayed until after delivery. Current data suggest that certain adjuvant chemotherapy regimens can be safely administered in the second and third trimesters of pregnancy, and development is normal in children of mothers who receive this chemotherapy. Administration of antimetabolites should be avoided because of the potential for damage to the placenta. Initiation of endocrine therapy is generally delayed until after delivery. Pregnancy after a diagnosis of early breast cancer does not appear to increase the risk for metastatic disease in limited data sets. The major considerations for women contemplating pregnancy should be its timing with regard to endocrine therapy, which may require 5 or more years to complete, as well as the underlying risk for disease recurrence.

Men account for not more than 1% of breast cancer cases. Failure of the patient or health care provider to diagnose the disease means that it may be diagnosed at a later stage. Most breast cancers in men express ER, and treatment recommendations for these men are generally similar to those for postmenopausal women.

PREVENTION AND SCREENING

There is enormous interest in the development of breast cancer prevention strategies. Currently these include surgical, chemopreventive, and lifestyle modification approaches.

Prophylactic Mastectomy and Oophorectomy

Prophylactic mastectomy appears to reduce the risk for developing breast cancer by approximately 90% in individuals who are at high risk because of a strong family history or carriage of a germline *BRCA1* or *BRCA2* mutation. Prophylactic oophorectomy in *BRCA* mutation carriers has been shown to decrease breast cancer incidence by approximately 50%, presumably because of reduction in ovarian steroids.[7] It is critical that women who opt for prophylactic mastectomy be counseled about the possibility that cancer may develop in remnants of breast tissue that remain after prophylactic mastectomy.

Chemoprevention

The observation that adjuvant endocrine therapy with the SERM tamoxifen also decreased contralateral breast cancer led to evaluation of tamoxifen as a chemopreventive for well women who are at high risk for breast cancer. Meta-analysis of four randomized trials of tamoxifen versus placebo for high-risk women confirmed a 38% reduction in invasive breast cancer with tamoxifen. The largest trial, NSABP P01, randomized 13,388 high-risk women to receive either tamoxifen or placebo for 5 years. Risk factors used to determine eligibility were age 60 years or older, a diagnosis of lobular carcinoma in situ, or age 35 to 59 years with a constellation of risk factors that, when combined, resulted in a 1.67% or greater risk for breast cancer within 5 years. This study showed a 50% decrease in the diagnosis of breast cancer across all age groups at a cost of increased risk for endometrial cancer and thromboembolic events in women older than 50 years. These data led to the approval of tamoxifen to reduce the incidence of breast cancer in women at high risk as defined by the eligibility criteria for this prevention trial.

Other SERMs have been studied as chemopreventive agents.[8,9] Two trials have documented the utility of raloxifene as a breast cancer risk-reduction strategy. In one trial, raloxifene administration reduced the incidence of breast cancer in postmenopausal women of average breast cancer risk and elevated cardiovascular risk. A second trial showed that raloxifene was less effective than tamoxifen in preventing invasive breast cancer in postmenopausal women at high risk for breast cancer. Raloxifene use was associated with fewer uterine cancers than tamoxifen but carries the same risk for thromboembolic events. It is approved by the FDA for risk reduction

in high-risk postmenopausal women. Two large studies of an aromatase inhibitor (exemestane or anastrozole) show that either of these drugs lowers the incidence of breast cancer in moderate- to high-risk postmenopausal women by approximately 50%.[■■] A vitamin A derivative, fenretinide, showed some impact as a chemopreventive agent for secondary prevention in premenopausal breast cancer survivors but has not found a place in routine practice. Finally, large epidemiologic studies have raised the possibility that aspirin or certain statins might also decrease breast cancer risk, but these agents have not been prospectively tested.

Lifestyle Modification

Prevention strategies that involve lifestyle alterations have been suggested. The Women's Health Initiative did not show a clear role for a low-fat diet as a means of breast cancer prevention. Regular exercise, especially during adolescence, may be associated with reduced breast cancer risk. By extrapolation from the epidemiologic studies mentioned previously, abstinence from alcohol might slightly reduce the risk for breast cancer.

Screening

Screening strategies for breast cancer have traditionally included the triad of breast self-examination (BSE), clinical breast examination (CBE) by a health care professional, and screening mammography in well women. Although widely promulgated as an important component of early detection, two large randomized trials of conventional BSE versus observation failed to show any clinical advantage with BSE. As a result, many experts now promote breast awareness rather than regular BSE. The independent value of CBEs has not been rigorously assessed. Rather, it has been studied in conjunction with screening mammography, in which case the two interventions appear to decrease mortality from breast cancer by 25 to 30% in women older than 50 years. Considerable controversy continues over the value of screening mammography in women 40 to 50 years of age and those over 70 years of age as well, as the optimal interval between mammograms for women aged 50 to 70 years. Currently, the American Cancer Society recommends annual screening mammography for women older than 40 years of age who are at standard risk for breast cancer. In contrast, the U.S. Preventive Services Task Force recommends that women between 40 and 50 years of age be counseled about the risks and benefits of screening mammography and that screening mammography can be used at 2-year intervals for women aged 50 to 75 years (Table 198-7). However, women aged 40 to 49 years with a two-fold increased risk for breast cancer have similar benefit-to-harm ratios for biennial screening mammography, as do average-risk women aged 50 to 74 years. A study of digital versus conventional film screen mammography failed to show an overall advantage for digital mammography but suggests that digital mammography may be more useful for women with dense breasts. The survival benefits of mammography come at the cost of false positive results that lead to anxiety and further evaluation. According to various estimates, more than half of the women screened by mammography have false-positive results over 10 years of screening, and 7 to 9% get recommendations for breast biopsy.[10] Other studies have estimated an overdiagnosis rate of 3.3% for invasive cancer and from 18 to 32% for carcinoma in situ.[11] In general, screening is unlikely to be worthwhile for women with a life expectancy of less than ten years.[12]

The knowledge that screening mammography fails to diagnose approximately 10 to 15% of breast cancers has led to the evaluation of other imaging modalities. Of these, MRI is the most mature.[13] MRI has been promoted as a useful screening tool for women at high risk by virtue of the *BRCA* mutation; indeed, the American Cancer Society has recommended consideration of screening MRI for women whose predicted risk for breast cancer exceeds 20%. MRI has been shown to detect breast cancer in the contralateral breast in 3% of women with a newly diagnosed breast cancer whose contralateral mammogram showed no abnormality. Use of MRI in the general population is limited by the fact that it is highly sensitive but lacks specificity. As a result, MRI has not translated into improved selection of surgical treatments or a reduction in the number of operations. Insufficient information exists about other breast imaging modalities such as ultrasound and radionuclide imaging to support their use in screening asymptomatic women.

TABLE 198-7 BREAST CANCER SCREENING

TEST	ACS	USPSTF
Mammography		
40-<49 yr	Annual	Insufficient evidence to support
>49 yr	Annual if healthy	Every 2 yr, 50-75 yr
Clinical breast examination	Every 3 yr, 20-39 yr Annually, ≥40 yr	Insufficient evidence to support
Breast self-examination	Optional	Not recommended

ACS = American Cancer Society; USPSTF = U.S. Preventive Services Task Force.

BENIGN BREAST LESIONS

Benign breast disease includes mastalgia (pain and tenderness), mastitis (including infectious and noninfectious inflammatory conditions), trauma, and benign tumors. In addition, mastalgia can be caused by extramammary conditions, such as myocardial ischemia, pneumonia, pleural irritation, esophageal spasm, costochondritis, rib fracture, and varicella-zoster virus (preceding the skin eruption). After these conditions have been ruled out, mastalgia is considered to be a benign condition that is self-limited. It can be cyclical or noncyclical in timing (where it is thought to be related to hormonal activity), and it may occur in postmenopausal women, even in the absence of hormone replacement therapy.

Mastitis is due to breast inflammation or infection, and it can occur in either nonlactating or lactating women. The typical presentation in nonlactating women is occurrence in their 40s with the acute onset of severe breast pain and tenderness followed by erythema and swelling that tends to be localized in the nipple-areolar area. The cause in most cases is considered to be rupture of dilated subareolar ducts that leads to an inflammatory response to leakage of intraductal contents into periductal tissue. Infection is difficult to rule out, and, in practice, patients are often treated with empirical antibiotics; if symptoms do not resolve within 7 to 10 days on antibiotics, ultrasound is required to rule out abscess. If the latter is found, incision and drainage are required. The most important differential diagnosis with mastitis is inflammatory carcinoma, as noted previously. Therefore, failure to improve with antibiotics or to find an abscess should lead to evaluation by a breast surgeon. Mastitis in lactating women is often due to infection caused by a break in the skin of the nipple or milk stasis. *Staphylococcus aureus, Staphylococcus albus,* and sometimes *Escherichia coli* or streptococci are the most common pathogens. Treatment requires antibiotics and milk removal.

Nonproliferative and Proliferative Benign Breast Lesions

Nonproliferative benign breast lesions include (1) inflammatory fat necrosis, which follows surgical or blunt trauma, and generally resolves spontaneously; (2) lymphocytic mastitis, which may be seen in diabetic patients; and (3) granulomatous mastitis, associated with foreign body reactions (e.g., silicone and paraffin for breast augmentation and reconstruction after cancer surgery), sarcoid, or certain infections. Other nonproliferative benign breast lesions appear as tumor-like processes, including (4) fibroadenoma, a very common (~25% of women), usually solitary, sharply demarcated, smooth lesion in younger age groups; (5) phyllodes tumor (previously known as cystosarcoma phyllodes); (6) intraductal papilloma, solitary lesions that may be accompanied by bloody nipple discharge; (7) fibrocystic breast disease, which is now more appropriately called *fibrocystic changes* because it is observed clinically in up to 50% and histologically in 90% of women, composed of varying amounts of fibrosis and cysts sometimes associated with calcifications and inflammation; and (8) simple or complex cysts, which should be aspirated with ultrasound guidance and, when the fluid is not clear, sent for cytologic analysis.

Proliferative benign breast lesions include ductal or lobular hyperplasia; atypical hyperplasias are associated with increased risk for breast cancer.

Breast Cancer Risk

The increasing use of mammography has increased the frequency of breast biopsies, which, in turn, has increased the finding of benign breast lesions, the most common findings on biopsy. The benign breast lesions listed in the previous section encompass the general histologic spectrum of (1) nonproliferative lesions, (2) proliferative lesions without atypia, and (3) atypical hyperplasia, listed in ascending order of risk for breast cancer. A large number of retrospective and prospective studies have shown an overall relative risk for breast cancer of 1.5 to 1.6 for women with biopsy-proved benign breast disease compared with women in the general population. A study of 9087 women with all types of benign histologic findings, followed for a median of 15 years at the Mayo Clinic, found that 707 of them developed breast cancers. Increased risk for cancer persisted for at least 25 years after the original biopsy. Of the three broad histologic categories, the relative risk for cancer development was 4.24 associated with atypia, 1.88 with proliferative changes

without atypia, and 1.27 with nonproliferative lesions. It is not known whether the finding of a benign breast lesion with atypical histologic findings represents an actual precursor lesion for cancer or if it is only a marker for a general tendency to develop breast cancer. The observations that approximately half of breast cancers in these patients arise in the contralateral breast suggests the latter hypothesis.

Grade A References

A1. Rao R, Euhus D, Mayo HG, et al. Axillary node interventions in breast cancer: a systematic review. *JAMA*. 2013;310:1385-1394.

A2. Early Breast Cancer Trialists' Collaborative Group (EBCTCG), et al. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. *Lancet*. 2011;378:1707-1716.

A3. McGale P, Taylor C, Correa C, et al. Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials. *Lancet*. 2014;383:2127-2135.

A4. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet*. 2012;379:432-444.

A5. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamosifen: patient-level meta-analysis of randomized trials. *Lancet*. 2011;378:771-784.

A6. Pagani O, Regan MM, Walley BA, et al. Adjuvant exemestane with ovarian suppression in premenopausal breast cancer. *N Engl J Med*. 2014;371:107-118.

A7. Dowsett M, Cuzick J, Ingle J, et al. Meta-analysis of breast cancer outcomes in adjuvant trials of aromatase inhibitors versus tamoxifen. *J Clin Oncol*. 2010;28:509-518.

A8. Davies C, Pan H, Godwin J, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oesteroen receptor-positive breast cancer: ATLAS, a randomized trial. *Lancet*. 2013;381:805-816.

A9. Perez EA, Romond EH, Suman VJ, et al. Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2-positive breast cancer: planned joint analysis of overall survival from NSABP B-31 and NCCTG N9831. *J Clin Oncol*. 2014;32:3744-3752.

A10. Goldhirsch A, Gelber RD, Piccart-Gebhart MJ, et al. 2 years versus 1 year of adjuvant trastuzumab for HER2-positive breast cancer (HERA): an open-label, randomised controlled trial. *Lancet*. 2013;382:1021-1028.

A11. Mehta RS, Barlow WE, Albain KS, et al. Combination anastrozole and fulvestrant in metastatic breast cancer. *N Engl J Med*. 2012;367:435-444.

A12. Baselga J, Campone M, Piccart M, et al. Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N Engl J Med*. 2012;366:520-529.

A13. Baselga J, Cortés J, Kim SB, et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med*. 2012;366:109-119.

A14. Verma S, Miles D, Gianni L, et al. Trastuzumab emtansine for HER2-positive advanced breast cancer. *N Engl J Med*. 2012;367:1783-1791.

A15. Coleman RE, Marshall H, Cameron D, et al. Breast-cancer adjuvant therapy with zoledronic acid. *N Engl J Med*. 2011;365:1396-1405.

A16. Vogel VG, Costantino JP, Wickerham DL, et al. Update of the National Surgical Adjuvant Breast and Bowel Project Study of Tamoxifen and Raloxifene (STAR) P-2 Trial: preventing breast cancer. *Cancer Prev Res*. 2010;3:696-706.

A17. Goss PE, Ingle JN, Alés-Martínez JE, et al. Exemestane for breast-cancer prevention in postmenopausal women. *N Engl J Med*. 2011;364:2381-2391.

A18. Cuzick J, Sestak I, Forbes JF, et al. Anastrozole for prevention of breast cancer in high-risk postmenopausal women (IBIS-II): an international, double-blind, randomised placebo-controlled trial. *Lancet*. 2014;383:1041-1048.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Antoniou AC, Casadei S, Heikkinen T, et al. Breast-cancer risk in families with mutations in PALB2. *N Engl J Med*. 2014;371:497-506.
2. Couch FJ, Nathanson KL, Offict K. Two decades after *BRCA*: setting paradigms in personalized cancer care and prevention. *Science*. 2014;343:1466-1470.
3. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490:61-70.
4. Ellis MJ, Perou CM. The genomic landscape of breast cancer as a therapeutic roadmap. *Cancer Discov*. 2013;3:27-34.
5. Goldhirsch A, Winer EP, Coates AS, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol*. 2013;24:2206-2223.
6. Burstein HJ, Prestrud AA, Seidenfeld J, et al. American Society of Clinical Oncology clinical practice guideline: update on adjuvant endocrine therapy for women with hormone receptor–positive breast cancer. *J Clin Oncol*. 2010;28:3784-3796.
7. Domchek SM, Friebel TM, Singer CF, et al. Association of risk-reducing surgery in *BRCA1* or *BRCA2* mutation carriers with cancer risk and mortality. *JAMA*. 2010;304:967-975.
8. Cuzick J, Sestak I, Bonanni B, et al. Selective oestrogen receptor modulators in prevention of breast cancer: an updated meta-analysis of individual participant data. *Lancet*. 2013;381:1827-1834.
9. Visvanathan K, Hurley P, Bantug E, et al. Use of pharmacologic interventions for breast cancer risk reduction: American Society of Clinical Oncology clinical practice guideline. *J Clin Oncol*. 2013;31:2942-2962.
10. Pace LE, Keating NL. A systematic assessment of benefits and risks to guide breast cancer screening decisions. *JAMA*. 2014;311:1327-1335.
11. Kalager M, Adami HO, Bretthauer M, et al. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. *Ann Intern Med*. 2012;156:491-499.
12. Walter LC, Schonberg MA. Screening mammography in older women: a review. *JAMA*. 2014;311:1336-1347.
13. Morrow M, Waters J, Morris E. MRI for breast cancer screening, diagnosis, and treatment. *Lancet*. 2011;378:1804-1811.

REVIEW QUESTIONS

1. A 36-year-old woman presents with a palpable 2-cm right upper outer quadrant breast mass. There is no palpable axillary, cervical, or supraclavicular adenopathy, nor any skin or nipple changes. Bilateral mammography shows only the index 2-cm right breast lesion, and ultrasound-guided biopsy shows infiltrating ductal cancer with positive estrogen and progesterone receptor expression and absence of *HER2* expression. Options for local therapy could include breast conservation (lumpectomy, sentinel node removal, and radiotherapy) or mastectomy with or without reconstruction. Which factor would be an absolute contraindication to breast conservation therapy?

A. Patient is 35 weeks pregnant
B. Previous radiation therapy for a diagnosis of cervical cancer
C. *BRCA1* mutation carrier
D. Inability to obtain negative margins of the lumpectomy specimen even after re-excision
E. Family history of breast cancer in maternal grandmother

Answer: D Randomized clinical trials have demonstrated that breast conservation therapy and mastectomy are equally effective for most women with early-stage breast cancer. Thus, breast conservation therapy is the preferred approach for local therapy for most women. The primary reasons to favor mastectomy include the inability to excise all tumor in the breast (positive margins on re-excision or multifocal breast cancer) or inability to deliver radiotherapy (e.g., during an ongoing pregnancy or after previous mantle radiotherapy for treatment for Hodgkin's lymphoma). In this patient, radiation therapy could be safely delayed for several weeks to permit completion of the pregnancy. Because the breast does not fall in the treatment area for radiotherapy for cervical cancer, this patient could safely receive breast radiotherapy. Family history of breast cancer may contribute to breast cancer risk but does not have an impact on selection of therapy for breast cancer once diagnosed. Breast conservation therapy remains an option for women with a germline *BRCA1* or *BRCA2* mutation, although some patients will opt for bilateral mastectomy. Finally, patient preference should always take precedence after a detailed informed consent discussion has taken place.

2. The patient in Question 1 asks if she should consider genetic testing for a hereditary breast cancer syndrome because of the diagnosis of breast cancer at an early age. What additional element in her history would support the recommendation to proceed with testing?

A. Personal history of cervical cancer
B. Patient is of Irish Catholic extraction
C. Personal history of Hodgkin's lymphoma
D. Family history of breast cancer in paternal grandmother and ovarian cancer in paternal aunt
E. Family history of endometrial cancer in mother and postmenopausal breast cancer in maternal grandmother

Answer: D Germline mutations in the *BRCA1* or *BRCA2* gene leading to hereditary breast and ovarian cancer syndromes account for a few percent of breast cancer diagnoses in women. These mutations are autosomal dominant and can thus be inherited through the maternal or paternal line. Careful ascertainment of family history for cancer diagnosis on both sides of the family is critical. Clinical hallmarks of *BRCA1*- or *BRCA2*-related malignancies include the predisposition to breast and/or ovarian cancer (but not endometrial or cervical cancer), bilateral disease, and early onset of disease. Three founder mutations in *BRCA1* or *BRCA2* are found in 1 to 2% of Ashkenazi Jews but have not been described in Irish Catholics.

3. A 70-year-old postmenopausal woman underwent breast conservation surgery and radiotherapy for management of a 1.7-cm, grade 2 infiltrating ductal carcinoma with highly positive estrogen and progesterone receptor and absent *HER2* expression by immunohistochemistry. Sentinel lymph node biopsy was negative. Her general health is good. She takes no medications. Family history is negative for malignancy. A 5-year course of adjuvant treatment with the aromatase inhibitor, anastrozole, is recommended. She should be counseled that a potential side effect of anastrozole is:

A. Neutropenia.
B. Arthralgias.
C. Neuropathy.
D. Uterine cancer.
E. Risk of thromboembolic events.

Answer: B Two common types of adjuvant endocrine therapy, tamoxifen and the aromatase inhibitors such as anastrozole, are considered for postmenopausal women with hormone receptor–positive invasive breast cancer. Current guidelines suggest that use of an aromatase inhibitor is preferred for most postmenopausal women. These agents suppress aromatase, the enzyme that converts androgens synthesized in the adrenal glands to estrogens in adipose and breast tissues. Therefore, this patient should be counseled that letrozole could lead to signs and symptoms of estrogen deprivation such as vasomotor symptoms, bone loss, and elevated cholesterol. An increased incidence of arthralgias is also seen in placebo-controlled trials of aromatase inhibitors; the underlying mechanism is not well understood. Because of its estrogen suppressive action, anastrozole administration is not associated with the development of uterine cancer, a malignancy that is associated with unopposed estrogen, nor is it associated with increased risk for thromboembolic events that are seen with tamoxifen or hormone replacement therapy. Finally, unlike cytotoxics, endocrine therapy such as letrozole is not associated with neutropenia or neuropathy.

4. The patient in Question 3 asks what type of breast cancer–specific testing is required during her follow-up in the absence of symptoms. Which is the most appropriate strategy?

A. Regular history and physical examination without any testing in the absence of symptoms
B. Regular history and physical examination and annual mammography
C. Regular history and physical examination and annual mammography and breast MRI
D. Regular history and physical examination, annual mammography, and routine blood testing, including CBC, chemistry panel, and tumor markers
E. Regular history and physical examination; annual mammography; routine blood testing including CBC, chemistry panel and tumor markers, and annual PET/CAT

Answer: B Evidence-based guidelines for the follow-up of asymptomatic survivors of early breast cancer recommend regular physical examination and annual mammography. The goal of follow-up in such individuals is to facilitate the early diagnosis of a new breast cancer in the contralateral breast or a recurrence in the ipsilateral breast so that appropriate management can be undertaken. There is no information to support the use of MRI in addition to mammography for this purpose, and bilateral mammography is the preferred approach. A third potential goal is to assess patients for metastatic disease. The primary goal of therapy for metastatic breast cancer is palliation and prevention of symptoms. Randomized trials failed to show a benefit for routine blood tests or whole-body imaging to facilitate this goal or prolong survival. The preferred approach is regular history and physical examination and education of patients about symptoms that might suggest distant recurrence. Breast cancer survivors also should undertake a program of age-appropriate general health measures (e.g., pelvic examination, Pap smear, colon cancer screening, etc.). In addition, a diagnostic evaluation should be undertaken to investigate any suspicious symptoms or findings on physical examination in a survivor of early-stage breast cancer.

5. Multiple hereditary cancer syndromes and their associated genes have been identified. In addition to *BRCA1* and *BRCA2* germline mutations, germline mutations in which gene can predispose to breast cancer?

A. *p53*
B. *RET*
C. *APC*
D. *VHL*
E. *HER2*

Answer: A Germline mutations of the *p53* tumor suppressor gene define the Li-Fraumeni hereditary cancer syndrome, which is characterized by predisposition to multiple types of cancer, including breast cancer. Germline mutations of the *RET* gene are associated with multiple endocrine neoplasia (MEN) syndrome 2; its hallmarks are medullary thyroid cancer, pheochromocytoma, and hyperparathyroid gland abnormalities. Inherited mutations of the *APC* tumor suppressor gene characterize familial adenomatous polyposis of the colon and predispose to colon cancer but not breast cancer. Mutated *VHL* is seen in a variety of benign and malignant tumors, including clear cell renal cancers and breast cancer. The *HER2* gene is amplified in 15 to 20% of breast cancers; this somatic change is associated with sensitivity of the tumor to *HER2*-targted therapy; germline mutations of the *HER2* gene have not been described.

199

GYNECOLOGIC CANCERS

DAVID SPRIGGS

GENERAL CONSIDERATIONS

The role of the internist or primary care practitioner in the care of women with gynecologic cancers has three important components: awareness and prompt diagnosis of gynecologic cancers, primary care of the patient with active cancer, and care and surveillance of cancer survivors. Diagnosis of these cancers relies on both the collection of appropriate clinical history and routine gynecologic examination. None of these malignancies are common, but in every case, early detection and treatment leads to superior outcomes. Both demographic trends and the increased effectiveness of cancer treatment will contribute to a growing population of cancer survivors receiving care in the non-oncologist's practice.

CERVICAL CANCER

EPIDEMIOLOGY

Cervical cancer is an important worldwide health problem, often affecting younger women in the prime of life. Newly diagnosed cervical cancer will affect over 500,000 women annually and will lead to over 250,000 deaths, making it the third most common cancer in women around the world. Because of the long natural history of premalignant and early cervical cancer, annual cytologic screening (Pap smear) has dramatically reduced the incidence of advanced cervical cancer in the developed world and areas with good medical infrastructure. Cervical cancer remains an important cause of cancer death in the developing world, often striking young women in their most productive years. It is primarily a disease afflicting patients of lower socioeconomic status whose access to advanced medical care is limited. The overall 5-year survival rate for cervical cancer is approximately 67% in the United States (~12,000 new cases and roughly 4000 deaths in 2014) but depends on the group prevalence of annual screening to detect early-stage disease.[1] Younger women and white women have better outcomes than older women and black and Hispanic women based on many factors, including early diagnosis of localized cancers. Other factors linked to higher incidence of cervical cancer include age of first intercourse, parity (more live births linked to higher risk), current cigarette smoking, and male human papillomavirus (HPV)-related factors. Regions with a higher frequency of penile cancer and lower rates of male circumcision have a higher incidence of cervical cancer diagnoses. High prevalence of other sexually transmitted diseases is also a known risk factor, including *Chlamydia trachomatis* and herpes simplex virus infections. Finally, host immunity appears to play a major role. Both human immunodeficiency virus (HIV) infection and immune suppression related to transplantation are associated with dramatic increases in the incidence of cervical cancer.

PATHOBIOLOGY

Infection with specific, carcinogenic strains of HPV has been established as the necessary causal event for nearly all cervical cancer.[2] This includes both the most common squamous cell cancers (85%) and the less common, more difficult to detect, adenocarcinomas. Not all HPV strains are linked to cervical cancer, and the most common cancer-related subtypes include 16, 18, 32, 33, 35, 45, 52, and 58. Although the prevalence of the genotype will show substantial regional variation, over 70% of all cervical cancers can be linked to HPV 16, 18, and 45. Many other HPV strains can cause infection but will not lead to cervical cancer. HPV infection tends to occur soon after sexual initiation, and in most women will be cleared by the immune system within 24 months. In some cases the HPV DNA is incorporated into the host DNA and can continue to produce viral proteins (persistence). Often the HPV DNA is silenced and will go through a period of latency but can be reactivated later. The HPV protein E7 mediates immortalization through abrogation of the G1/S transition through interaction with the Rb protein. Other targets of E7 include a variety of cyclin-dependent kinases and cyclins that are critical for cell cycle regulation. The E6 protein binds to p53 and promotes its degradation, leading to decreased capacity for DNA damage repair. E6 also upregulates the cellular telomerase complex and contributes to immortalization of HPV-infected epithelial cells. As a consequence of inactivation of Rb and p53, there is loss of cell cycle regulation and a variety of subsequent mutations appear to accumulate. Mutational load appears to play a dominant role in the eventual development and progression of cervical cancer.

SCREENING

HPV infection of cervical epithelial cells leads to a failure of cellular differentiation and to cytologic abnormalities of the cervical epithelium referred to as cervical intraepithelial neoplasia (CIN). Whereas early, or low-grade, CIN often resolves, higher grades of CIN (CIN 2/3) appear to be linked to viral integration into host DNA and a failure to clear the infection. Persistent CIN can be detected by Pap cytologic testing during its long preinvasive phase, and this fact has led to the success of screening strategies in developed countries. Current recommendations mandate initiation of cytologic (Pap) testing at age 21 and every 2 years until age 30, when women with persistently negative screening tests can be screened every 3 years until age 65.[3,4] For women in the 30- to 65-year age group, a combined screening strategy of cytologic and HPV testing is preferred. The combination of HPV testing and

liquid-based cytology has a sensitivity of 96.7%; patients testing positive for HPV 16/18 should all be referred for colposcopy, regardless of their cytologic result. At age 65, women with multiple negative cytologic screening tests may discontinue screening. Any woman with a cytologic screen revealing CIN 2 or other abnormalities should be referred to a gynecologist for colposcopy and appropriate subsequent management. CIN 2/3 implies probable viral integration, so a woman with a history of CIN 2/3 should undergo more frequent and extended screening.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Early cervical cancer is generally asymptomatic, and thus detection is highly dependent on screening and routine gynecologic care. In advanced stages, patients may present with vaginal discharge or bleeding, pelvic pain, or abnormalities of bowel, bladder, or sexual function. The diagnosis of cervical cancer is usually based on positive cytologic status and direct inspection by vaginal speculum examination. Early cervical cancers are detected primarily by colposcopy and direct biopsy. The staging of cervical cancer is based on both expert clinical evaluation of resectability and imaging studies of both the pelvis and potential sites of distant metastases (lymph nodes, lungs, and abdominal sites).

TREATMENT

Rx

The primary treatment of localized cervical cancer is stage dependent. Staging of cervical cancer depends on the clinical evaluation of a skilled gynecologic oncologist; the International Federation of Gynecology and Obstetrics (FIGO) staging system is summarized in Table 199-1. In general, treatment of early-stage cervical cancer is exclusively surgical. The appropriate procedure will depend on patient age, comorbidities, stage, and tumor size. Radical hysterectomy with nodal dissection is the most common procedure, although younger patients with low risk for metastatic disease may undergo simple hysterectomy or even radical trachelectomy for highly selected patients who desire preservation of fertility. Patients with low-stage cervical cancer who have certain high-risk pathologic findings after surgery may benefit from radiation or chemotherapy.

There is some variability in treatment recommendations for patients with FIGO stage IB to IIA. Younger women with smaller tumors may be offered radical hysterectomy, whereas bulkier tumors and older patients may be better served by definitive chemoradiation. Patients with FIGO stage IIB to IV should all be treated with combinations of cisplatin and external beam radiation.[■] It is important to note that the combination of cisplatin and external beam radiation can provide an opportunity for cure and long-term survival, even in stage IV cancer of the cervix.

Patients with recurrent cancer of the cervix or patients with distant metastases at diagnosis are very rarely curable. Death is usually related to local recurrence, but distant metastatic spread to the lung, peritoneum, or bone is also common in cervical cancer. Platinum-based chemotherapy remains the mainstay of palliative treatment. The addition of bevacizumab, a humanized anti-vascular endothelial growth factor (VEGF) monoclonal antibody, to combination chemotherapy in patients with recurrent, persistent, or metastatic cervical cancer has been associated with an improvement of 3.7 months in median overall survival.[■] Recurrent or persistent tumors within the prior radiation field are particularly difficult to manage with chemotherapy. Local problems such as pain, bleeding, and fistulas between the bladder, vagina, and bowel are common in the advanced stages of this disease.

TABLE 199-1 APPROACH TO CERVIX CANCER BY STAGE*

STAGE	BRIEF DEFINITION	USUAL TREATMENT
0	Carcinoma in situ	Conization, hysterectomy
IA1	Microscopic stromal invasion < 3 mm	Conization, hysterectomy
IA2	Microscopic stromal invasion 3 to 5 mm	Radical hysterectomy
IB1	Visible lesion < 4 cm in greatest dimension	Radical hysterectomy
IB2	Visible lesion > 4 cm in greatest dimension	Chemoradiation
IIA	Tumor beyond the uterus, no parametrial involvement	Radical hysterectomy (selected)
IIB	Tumor beyond the uterus with parametrial extension	Chemoradiation
IIIA	Tumor involves lower third vagina but no pelvic wall extension	Chemoradiation
IIIB	Tumor extends to pelvic wall or hydronephrosis or regional lymph nodes	Chemoradiation
IVA	Involvement of mucosa of bowel or bladder	Chemo-radiation
IVB	Distant metastatic disease	Chemotherapy only

*Cervical cancer is staged clinically and not by either radiographic or pathologic findings. Chemoradiation generally includes weekly cisplatin treatment. Metastatic disease chemotherapy is generally based on cisplatin doublet treatment.

PROGNOSIS AND CARE OF SURVIVORS

Patients with a history of CIN 2/3 or other noninvasive disease require regular cytologic examinations, because the onset of cancer in patients with HPV 16/18 infection can occur years after primary infection. Early-stage cancer of the cervix is highly curable. Although perioperative complications can occur for any major operation, late complications of radical hysterectomy are not common but may include lymphedema and persistent bladder or rectal dysfunction. Patients with more advanced stage disease appropriately treated with chemoradiation combinations have long-term survival rates from 30 to over 80% percent, based on stage and underlying health. The late complications of chemoradiation are similar but probably more common, affecting 5 to 10% of survivors. Fistulas can occur between pelvic organs, including bladder, vagina, small bowel, and rectum. Other bladder problems may include urgency, incontinence, and chronic cystitis with loss of bladder capacity. Rectal problems include pain, diarrhea or constipation, urgency, and incontinence. Skeletal complications can include pelvic insufficiency fractures, associated with pain in as many as 10% of patients after radiation. Vaginal stenosis, dryness, and dyspareunia are common but usually can be managed with vaginal dilators and appropriate lubricants. Late complications of cisplatin-based chemotherapy include neuropathy, as well as renal insufficiency with chronic wasting of potassium and magnesium. Multimodality therapy regimens including radiation, surgery, and chemotherapy pose a higher risk for late complications and impaired patient quality of life decades after primary curative treatment.

PREVENTION

Our understanding of the pathobiology of HPV-derived cervical cancer supported the development of vaccines for the prevention of the most common oncogenic subtypes of HPV infection (Chapter 373). There are now licensed vaccines that immunize against HPV 16, 18, and selected other high-risk types. They are both safe and effective in substantially reducing the risk for cervical dysplasia and carcinoma.[■■] HPV vaccines require administration before sexual activity and can be highly effective in reducing the risk for cervical cancer. Both the approval trials and subsequent follow-up trials reveal a high level of safety and durable type-specific immunity that seems to persist for more than 5 years after a series of three initial vaccinations.[■] Additional HPV type coverage may eventually be available, but it is already clear from population-based studies that the vaccine can reduce the incidence of precursor lesions such as cervical dysplasia. Beyond immunization, prevention is absolutely linked to regular screening for early HPV-related lesions and appropriate management during the premalignant phase of the disease.

ENDOMETRIAL CANCER

EPIDEMIOLOGY

Endometrial cancer arises from the epithelium of the uterine lining, in distinction to uterine sarcomas that have their tissue of origin in the smooth muscle of the myometrium. Endometrial cancer is the most common gynecologic cancer in the United States, with over 52,000 new cases each year. Death from endometrial cancer occurs in approximately 8600 women each year. The risk factors for endometrial cancer include advancing age, estrogen exposure, obesity, and multiparity. As populations age across the world, the incidence of endometrial cancer is generally rising. The median age for the development of endometrial cancer is 60 years, and most endometrial cancers are diagnosed after age 50. Exposure to exogenous or endogenous estrogens appears to be a strong risk factor. Early menarche, late menopause, nulliparity, and unopposed estrogen in hormone replacement therapy are all implicated in higher endometrial cancer risk. Obesity is associated with higher levels circulating estrogens related to peripheral conversion of androstenedione to estrogen. Even the low rates of tamoxifen (a selective estrogen receptor modulator)-induced endometrial cancer are probably related to weak estrogenic effects on the endometrium.

PATHOBIOLOGY

Endometrial cancers can be divided into two large categories for understanding risk and biology. The more common tumors (type 1 cancers; 85%) are distinguished by endometrioid histology, a more differentiated grade, and a relationship to unopposed estrogen exposure. These type 1 malignancies tend to present as stage I or II cancers (confined to the uterus) and generally have a better prognosis than the type 2 cancers. They often retain expression of progesterone receptors.[5] On a molecular level, these tumors are more likely to be diploid, have wild-type *p53,* and be associated with loss of *PTEN* and microsatellite instability. In contrast, type 2 endometrial cancers are poorly differentiated, often have serous or clear cell histology, and often manifest at advanced stages with higher risk for dissemination. At the molecular level, type 2 endometrial cancers are aneuploid, carry mutations or loss of *p53* and lack *PTEN* deletion or other abnormalities of the PI3 kinase/AKT pathway.[6]

Most endometrial cancers are sporadic, but small percentages (3-5%) of type 1 cancers are associated with a cancer family history consistent with Lynch syndrome (hereditary nonpolyposis, colorectal cancer [HNPCC]). This familial cancer syndrome is associated with early-onset colon, ovary, renal, and endometrial cancers, related to germline mutations in one of the mismatch repair genes *MLH1, MSH2,* and *MSH6*. The development of endometrial cancer before 50 years of age with an appropriate family history or histologic appearance should prompt genetic testing of the patient, to guide both proband and family preventive management.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The 80% survivorship in endometrial cancer is predominantly related to the high likelihood of detection while the cancer is confined to the uterus. Abnormal uterine bleeding in postmenopausal women always should be regarded as suspicious. Abnormal vaginal discharge or nonspecific gastrointestinal symptoms also may be present at the time of diagnosis. It is important to recall that the routine Pap smear is not an appropriate diagnostic procedure to exclude endometrial cancer. The classic dilatation and curettage (D&C) has largely been replaced by an office endometrial biopsy. The D&C and hysteroscopy are now usually reserved for uterine bleeding problems that are difficult to diagnose. Common sites of spread for endometrial cancer at the time of diagnosis include pelvic or periaortic lymph nodes, ovary, and peritoneal implants. The recently revised FIGO staging system for endometrial cancer is shown in Table 199-2.

TABLE 199-2 APPROACH TO ENDOMETRIAL CANCER BY STAGE*

STAGE	BRIEF DEFINITION	USUAL TREATMENT
IA G1/2	Invasion to less than half the myometrial thickness	TAH and BSO, no XRT
IA G3	Invasion to less than half the myometrial thickness, poorly differentiated, serous cancer	TAH and BSO, consider chemotherapy, no XRT
IB G1/2	50% or more invasion of myometrial thickness	TAH and BSO, no XRT
IB G3	50% or more invasion of myometrial thickness, poorly differentiated	TAH and BSO, consider chemotherapy, no XRT
II	Uterine and cervical involvement	TAH and BSO, consider VBT
IIIA	Invades corpus serosa or adnexa	TAH and BSO, tumor-directed chemo-XRT and chemotherapy
IIIB	Vaginal or parametrial invasion	TAH and BSO, tumor-directed chemo-XRT and chemotherapy
IIIC1	Pelvic lymph node involvement	TAH and BSO, tumor-directed chemo-XRT and chemotherapy
IIIC2	Paraaortic lymph node involvement	TAH and BSO, tumor-directed chemo-XRT and chemotherapy
IVA	Invasion of bowel or bladder	TAH and BSO, tumor-directed chemo-XRT and chemotherapy
IVB	Distant metastatic disease	Chemotherapy only

*Endometrial cancer (FIGO 2009) staging is based on pathologic findings. Stages I and II endometrial cancer do not show a survival benefit from adjuvant radiation therapy. High-risk disease may benefit from adjuvant chemotherapy, but the studies are ongoing. For stage III and IV disease, complete debulking is preferred and platinum-based chemotherapy is generally needed. The use of volume-directed radiotherapy for gross residual disease is under investigation.
TAH and BSO = total abdominal hysterectomy and bilateral salpingo-oophorectomy ; XRT = radiation therapy; VBT = vaginal brachytherapy irradiation.

TREATMENT Rx

For the vast majority of patients with endometrial cancer, surgical resection represents the primary approach to management. In many cases, surgery resection (total hysterectomy with bilateral salpingo-oophorectomy, lymph node assessment) will be sufficient for both staging and treatment. Laparoscopic or robotic surgery has replaced open hysterectomy in many situations and appears to be a less morbid procedure. The role and extent of lymph node dissection and the use of sentinel node biopsy are areas of active surgical research. These advances continue to decrease the morbidity and length of hospitalization associated with primary surgical treatment. Following primary surgery, patients are staged by pathologic findings and stratified by risk factors to assign appropriate adjuvant therapy (see Table 199-2). Important risk factors for recurrence include stage (as determined by metastatic deposits, lymph node involvement, and depth of myometrial invasion), tumor histology, cytologic grade, and lymphovascular space involvement. General treatment approaches, by stage, are shown in Table 199-2. High-risk patients with stage 1 disease historically received adjunctive radiotherapy with significant reduction in local recurrence, but there is no evidence supporting a long-term survival impact of radiation. In type 2 endometrial cancers of serous histology, patients with even the lowest stage disease are at risk for distant recurrence and adjuvant chemotherapy is increasingly used.

For patients with higher stage (Stage III or IV) disease, complete surgical resection of gross disease appears to be the important determinant of survival. Platinum-based chemotherapy has been increasingly identified as effective and has replaced whole-abdominal radiation therapy as the mainstay of adjuvant therapy. Combinations of carboplatin and paclitaxel are generally well tolerated and will reduce recurrence in both high-risk and advanced stage disease groups. Even patients with substantial residual disease after surgery can sometimes achieve long remissions following platinum-based treatment. As with most solid tumors, recurrent cancer is difficult to cure. An exception appears to be local recurrence in patients who did not receive prior radiation. In that select group, radiation therapy should be given with curative intent. Otherwise, metastatic endometrial cancer can be managed with other systemic chemotherapies; the patterns of drug sensitivities are similar to those seen in recurrent ovarian cancer. Common sites of metastatic disease include the peritoneal cavity, liver, bone, lung, and occasionally brain. The comorbidities commonly associated with age and obesity can play an important role in the treatment choice for endometrial cancer. For patients whose overall health precludes a surgical procedure, curative radiation therapy can be offered. Although the long-term complications of definitive radiation treatment can be significant, it may be the best choice for patients with a compromised performance status.

PREVENTION

For sporadic endometrial cancers, the best prevention strategy appears to be a high degree of suspicion in the context of postmenopausal bleeding. Because of its location, these tumors are rarely asymptomatic late in their natural history and surgical extirpation is usually curative. In families with HNPCC, a total hysterectomy with bilateral salpingo-oophorectomy should be strongly considered in confirmed carriers who have completed their childbearing.

PROGNOSIS

Stage I and II endometrial cancers have an excellent cancer-specific prognosis. The common comorbidities of obesity and age (diabetes, hypertension, cardiovascular disease) complicate their overall management, but most of these patients will survive their cancers without recurrent disease. A substantial minority of patients with more advanced disease can still be cured at the time of primary therapy with combinations of surgery, chemotherapy, and sometimes radiation. Patients with persistent or recurrent disease will generally succumb to their illness within a 2- to 3-year period.

Late complications of surgical resection tend to be quite limited; some mild lymphedema is the most common late toxicity. Pelvic radiation has more reported late complications, including bowel, bladder, and sexual dysfunction, but recent advances in computer-based treatment plans may provide a better long-term risk-to-benefit ratio. The long-term adverse effects of adjuvant chemotherapy are rare but can include neuropathy and occasional renal injury.

UTERINE SARCOMAS

Gynecologic sarcomas most often arise in the uterus. The two most common malignant histologic types are carcinosarcoma of the uterus and leiomyosarcoma. Carcinosarcoma of the uterus is generally sporadic and is managed with a combination of complete surgical resection and adjuvant combination chemotherapy. Like endometrial cancer, even advanced stage carcinosar-

coma can be successfully treated by multimodality therapy, but recurrent disease is rapidly fatal. Leiomyosarcoma is a malignant smooth muscle tumor arising in the myometrium. Surgical resection is the mainstay of treatment, and the role of adjuvant chemotherapy or radiotherapy is unproved. Recurrent leiomyosarcoma can be treated with surgery (for localized disease) and chemotherapy.

CANCERS OF THE FALLOPIAN TUBE AND OVARY

EPIDEMIOLOGY

Ovarian cancer is a generic term for a family of diseases, including epithelial ovarian cancers, germ cell tumors of the ovary, and stromal cancers arising in the ovary. Cancer originating in the epithelium of the fallopian tube and ovarian surface is the fourth most common lethal cancer in American women. Approximately 22,000 new cases of ovarian cancers will be diagnosed each year, and because most are diagnosed at advanced stages, roughly 15,000 of these women will eventually die of their disease despite current therapy. Internationally, the incidence of ovarian cancer is highest in white women, and societies characterized by low parity, high-fat diets, and older populations. Ovarian cancer incidence is highest in North America and Europe, with lower incidence in Asia and Sub-Saharan Africa. In the United States, risk factors for ovarian cancer include family history of breast or ovarian cancer, early menarche, delayed menopause, and nulliparity, whereas oral contraceptives, aspirin, and breast-feeding are consistently shown to reduce risk. The median age of onset for ovarian cancer is approximately 60 years, although familial ovarian cancers occur roughly 10 years earlier than the sporadic form.

The indolent borderline tumors of the ovary (or tumors with low malignant potential) form a distinct group of epithelial ovarian cancers characterized by slow growth, rare metastatic spread, and a very different pattern of genetic alterations. These rare borderline tumors are primarily managed by surgical resection as necessary.

PATHOBIOLOGY

The histologic appearance (and accompanying genetic characteristics) allows epithelial ovarian cancers to be divided into serous, mucinous, clear cell, and endometrioid cancers. Although distinct on histologic and molecular grounds, the clinical management of these ovarian cancers has not yet diverged. High-grade serous ovarian cancer (HGSOC) comprises over 80% of the total number of patients and is the best characterized subset of ovarian tumors.[7]

At least 85% of ovarian cancers are sporadic and not associated with heritable abnormalities. However, approximately 10 to 15% of HGSOCs are linked to families carrying germline mutations that are inherited in an autosomal dominant fashion. The affected genes include *BRCA1* and *BRCA2,* which are critical cell cycle checkpoint regulators involved in the maintenance of DNA integrity, particularly in the homologous recombination pathway that is important for double-strand DNA repair. Certain ethnic populations founded from small ancestral populations such as the Ashkenazi Jews have a high frequency of mutations in these genes. In addition to these genes, other less common germline mutations in DNA repair–linked genes, including *CHEK2, ATM,* and *PALB2,* also can increase ovarian cancer risk. Affected members of these breast/ovarian cancer families have an increased risk for both early-onset breast and ovarian cancer and a high lifetime risk for these cancers. In sporadic ovarian cancers, these same genes may be lost through somatic mutation or gene silencing. It is now estimated that 50% or more of all ovarian cancers have a BRCA-like phenotype. Interestingly, *BRCA1/2* mutated ovarian cancer appears to have a superior survival compared to sporadic ovarian cancer.[8] Mutations in *BRCA1* or *BRCA2* confer a "collateral sensitivity" to cisplatin and other chemotherapy agents that induce double-stranded DNA damage. This collateral weakness has been targeted through the development of poly-ADP-ribose polymerase inhibitors. A smaller group of familial ovarian cancers is linked to families with Lynch syndrome (HNPCC) and impairment of DNA mismatch repair.

It is now accepted that most serous ovarian cancers arise through a well-defined series of mutations in the fallopian tube epithelium. Through sequential losses of the homologous DNA repair competence (via loss *of BRCA1, BRCA2, PALB2,* or other genes in the DNA repair pathway) and *p53* function, a defined sequence of events leads to serous carcinoma in situ and eventually carcinoma involving the ovary, lymph nodes, and peritoneal surfaces of the abdominal organs. The lifetime risk for women carrying an altered, high-risk allele for one of these genes may approach 70% or more, depending on environmental factors and modifying alleles when the second, nonpathogenic allele is lost.

CLINICAL MANIFESTATIONS

The diagnosis of ovarian cancer is complicated by the nonspecific nature of the initial symptoms.[9] Unfortunately, small adnexal masses are usually silent, but as a mass increases in size, symptoms include pelvic fullness, constipation, urinary frequency, and dyspareunia. In more advanced disease, the most common symptoms include fatigue, malaise, early satiety, bloating, and loss of appetite. Abdominal girth increase is often noted as malignant ascites accumulates.

DIAGNOSIS

Ultrasonography or computed tomography of the abdomen and pelvis will usually strongly support the diagnosis and lead to early surgical intervention. Although peritoneal fluid cytology can sometimes be diagnostic, percutaneous biopsy of adnexal masses should *not* be part of the initial workup, to avoid contamination of the peritoneal cavity by an otherwise confined tumor. CA-125 is a normal mucin antigen from MUC16 that is often found at abnormally high levels in the serum of patients with ovarian cancer (and sometimes in other cancers, including lung, pancreas, and uterus). However, serologic findings can be supportive but is not diagnostic in affirming or excluding the diagnosis of ovarian cancer because a variety of benign conditions also can elevate the CA-125 value.

TREATMENT

Rx

The treatment of ovarian cancer begins at the time of exploratory laparotomy, and adequate primary surgery remains the most important determinant of ovarian cancer survival. The surgery should be performed by a trained gynecologic oncologist in a highly experienced center to achieve the best outcome. Adequate initial surgery should include a total abdominal hysterectomy with bilateral salpingo-oophorectomy, omentectomy, and lymph node dissection. The upper abdomen and diaphragm should be inspected and biopsy samples obtained. Every effort should be made to remove any visible cancer. If complete resection is not achievable, cytoreduction to tumor bulk less than 1 cm is still useful and will permit intraperitoneal chemotherapy. If a successful complete resection is not possible, neoadjuvant chemotherapy and deferred surgery is an acceptable choice, although the projected survival is inferior to complete primary resection.

At the completion of primary surgery, a stage is assigned to the patient's tumor, based on extent of involvement as shown in Table 199-3. Using a combination of residual tumor bulk, tumor histology, grade and stage, one of four treatment plans is chosen, as illustrated in Figure 199-1. For stage IA or IB, grade 1 or 2 tumors, observation alone is the most acceptable treatment because these patients have a tumor recurrence rate below 5%. For all stage IC cancers, all grade 3 tumors, and the rare stage IIA tumors, the recommendation should be for three to six cycles of platinum-based chemotherapy. For stage IIB to IIIC tumors, with residual disease less than 1 cm in greatest dimension, six cycles of platinum-based intraperitoneal treatment is preferred, and bulkier disease or stage IV disease should receive intravenous platinum-based therapy.[10] These strategies are indicated in Table 199-3.

Following primary chemotherapy, reassessment of extent of residual disease by CA-125 and computed tomography is performed. For patients with persistent or resistant disease, other palliative chemotherapy can be offered but the outlook for such patients is quite poor. Patients in complete remission can be observed expectantly. Nearly all relapses will occur within a 3-year period, and regular quarterly follow-up is recommended in the United States. Over 75% of patients with advanced ovarian cancer will relapse in 12 to 30 months after diagnosis; although patients can be successfully treated for years, curative treatment is not yet available. Choice of chemotherapy at the time of recurrence depends on treatment-free interval, comorbidities, and residual toxicities. Active agents for retreatment include with carboplatin, gemcitabine, liposomal doxorubicin, pemetrexed, and bevacizumab.[11]

SCREENING AND PREVENTION

Although screening procedures have been aggressively sought for 25 years, there are no screening strategies that currently can be recommended, even for high-risk patients and definitely not for the general population of postmenopausal women. Combinations of serum CA-125 screening, transvaginal ultrasound testing, and panels of circulating serologic markers have all failed to achieve the requisite sensitivity and specificity in large-scale testing. Testing for germline and somatic mutations is increasingly important to

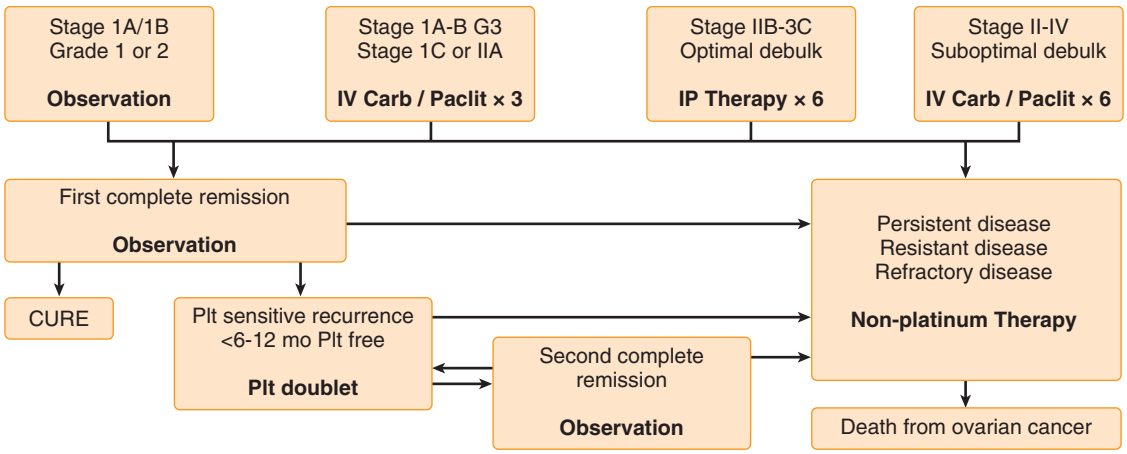

FIGURE 199-1. Ovarian cancer disease states model. The bulk of ovarian cancer management occurs after relapse. Primary therapy is allocated as indicated in the upper portion of the figure. The subsequent treatment depends on the elapsed time since the last exposure to a platinum complex drug. Patients with a platinum-free interval exceeding 6 to 12 months usually will receive a platinum-containing doublet and may attain a second (or greater) complete remision. Shorter intervals of platinum-free time will go on to nonplatinum single-agent treatment and have a median life expectancy of 1 to 2 years. Carb = carboplatin; Paclit = paclitexel; Plt = platinum.

TABLE 199-3 APPROACH TO OVARIAN CANCER BY STAGE*

FIGO STAGE	BRIEF DEFINITION	USUAL TREATMENT
IA G1/2	One ovary, no surface involvement	Staging laparotomy, observation only
IA G3	One ovary, no surface involvement; poorly differentiated	Staging laparotomy, chemotherapy for 3 cycles
IB G1/2	Both ovaries, no surface involvement	Staging laparotomy, observation only
IB G3	50% or more invasion of myometrial thickness, poorly differentiated	Staging laparotomy, chemotherapy for 3 cycles
IC	Surface involvement, positive cytology, or intraoperative spillage	Staging laparotomy, chemotherapy for 3 cycles
IIA	Extension to tubes or uterus	Staging laparotomy, chemotherapy for 3-6 cycles
IIB	Other pelvic organ extension	Staging laparotomy, chemotherapy for 3-6 cycles
IIC	Pelvic extension plus surface involvement, positive cytology, or intraoperative spillage	Staging laparotomy, chemotherapy with IP route preferred
IIIA	Microscopic tumor outside the true pelvis	Staging laparotomy, chemotherapy with IP route preferred
IIIB	Tumor < 2 cm in dimension outside the pelvis	Staging laparotomy, chemotherapy with IP route preferred
IIIC	Tumor > 2 cm outside the true pelvis or positive lymph nodes	Staging laparotomy, chemotherapy with IP route preferred
IV	Involvement of liver parenchyma, extension beyond the abdomen, cytology positive pleural effusion, inguinal lymph nodes	Chemotherapy

*Staging laparotomy includes total abdominal hysterectomy, bilateral salpingo-oophorectory and omentectomy. A pelvic lymph node dissection and upper abdominal exploration is also required. In well-staged, early-stage disease (IA/IB), the need for chemotherapy is based primarily on histologic grade. In advanced disease, optimal debulking to < 1 cm residual disease permits the use of intraperitoneal treatment (IP). All primary therapy should include a platinum drug and a second, non–cross-reactive drug such as paclitaxel or liposomal doxorubicin.

predict the response to therapy and prognosis.[12] The routine genetic testing of women with newly diagnosed ovarian cancer is recommended, in the appropriate circumstance, to discover families bearing *BRCA1/2* mutations who will require further monitoring.

Prevention strategies for ovarian cancer remain limited. Oral contraceptive use for 5 years or more can reduce ovarian cancer incidence by approximately 50%.[13] For high-risk women with mutations in *BRCA1* or *BRCA2*, a prophylactic salpingo-oophorectomy by age 40 can reduce ovarian cancer risk by more than 90% and will also reduce breast cancer incidence. Low-dose aspirin use has also been associated with reduced ovarian cancer risk.[14]

PROGNOSIS AND SURVIVOR MANAGEMENT

Women who remain disease-free for 3 years are likely to be cured of their ovarian cancer. Among patients with recurrence, median survival seems highly dependent on their initial stage and success of their primary debulking surgery. Patients with optimally debulked stage IIIC disease have a median survival in excess of 5 years, whereas patients with stage IV and patients with suboptimal debulking surgery have median survivals of 30 to 48 months. The terminal phase of ovarian cancer is generally characterized by progressive inanition and eventual intestinal obstruction. Ovarian cancer survivors may have intestinal adhesions from surgery and chronic neuropathy or electrolyte wasting from primary chemotherapy. Those ovarian cancer survivors living with chronic ovarian cancer are often able to continue to have productive work and family lives for several years after the diagnosis of ovarian cancer.

VULVAR CANCER

Cancer of the vulva is a relatively rare gynecologic malignancy that represents less than 5% of gynecologic cancer and is primarily seen in postmenopausal women. Like cervical cancer, the etiology of vulvar cancer is prior infection with HPV. Patients with vulvar cancer complain of itching, pain, and local discomfort, often arising in the presence of atrophic changes of the vulvar epithelium. The treatment of vulvar cancer is surgical excision with good margins and possible dissection of one or both inguinal lymph node regions in advanced disease. Recurrent or unresectable vulvar cancer also can be treated with external beam radiation, often in combination with cisplatin in radiation sensitizing doses.

UTERINE FIBROIDS

Uterine fibroids are the most common female pelvic tumor, with a prevalence as high as 40% during the reproductive years and as high as 70 to 80% by age 50. Fibroids, which are almost always benign monoclonal tumors, arise from disordered smooth-muscle cells in the uterine myometrium. A variety of somatic mutations, especially of the *MED12* gene on the X chromosome, have been described. Fibroids depend on estrogen and progesterone and usually shrink after menopause.[15] In early pregnancy and the postpartum period, however, they can grow rapidly.

The majority of fibroids are asymptomatic, but they can cause pelvic pressure, pain, and heavy uterine bleeding. Diagnosis is typically made by physical examination and confirmed by ultrasound. Although fibroids are not premalignant tumors, a uterine sarcoma can have similar symptoms at presentation.

Noninvasive, hormonally based anti-progesterone therapies such as oral ulipristal acetate (5 or 10 mg per day for 13 weeks) can reduce the size of fibroids, decrease dysfunctional uterine bleeding and pain, and improve overall quality of life. For persistent symptoms, surgical removal of the fibroid or of the uterus itself can be performed during laparotomy or by laparoscopic surgery. Power morcellation, a process by which the uterus is divided into smaller fragments that can be removed in stages laparoscopically, is no longer routinely recommended because of the small but finite risk of disseminating pieces of potentially malignant tissue found in a small percentage of fibroid uteruses.[16] Robotically assisted and laparoscopic hysterectomy have similar morbidity profiles, but the use of robotic technology results in

substantially higher costs. Other strategies for the treatment of symptomatic uterine fibroids include uterine artery embolization, transvaginal temporary uterine artery occlusion, and MRI-guided focused ultrasound.

OTHER GYNECOLOGIC CANCERS

Any structure of the müllerian tract can undergo malignant transformation. Other types of gynecologic cancer include germ cell tumors of the ovary (homologs of testicular cancer), stromal tumors of the ovary, and cancers of the vagina, vulva, and vulvar adnexa. These are all much less common than the tumors described earlier. Because of the rarity of these cancers, early referral to experts in gynecologic oncology is essential. Survivors of these rare forms of gynecologic cancer will share the residual side effects that affect survivors of cervical, endometrial, and ovarian cancers, such as potential alterations of bowel, bladder, and sexual function, as well as lymphedema and the consequences of early menopause.

Grade A References

A1. DiSilvestro PA, Ali S, Craighead PS, et al. Phase III randomized trial of weekly cisplatin and irradiation versus cisplatin and tirapazamine and irradiation in stages IB2, IIA, IIB, IIIB, and IVA cervical carcinoma limited to the pelvis: a Gynecologic Oncology Group study. *J Clin Oncol*. 2014;32:458-464.

A2. Tewari KS, Sill MW, Long HJ 3rd, et al. Improved survival with bevacizumab in advanced cervical cancer. *N Engl J Med*. 2014;370:734-743.

A3. Munoz N, Manalastas R Jr, Pitisuttithum P, et al. Safety, immunogenicity, and efficacy of quadrivalent human papillomavirus (types 6, 11, 16, 18) recombinant vaccine in women aged 24-45 years: a randomised, double-blind trial. *Lancet*. 2009;373:1949-1957.

A4. Paavonen J, Naud P, Salmeron J, et al. Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. *Lancet*. 2009; 374:301-314.

A5. Lu B, Kumar A, Castellsague X, et al. Efficacy and safety of prophylactic vaccines against cervical HPV infection and diseases among women: a systematic review & meta-analysis. *BMC Infect Dis*. 2011;11:13.

A6. Walker JL, Piedmonte MR, Spirtos NM, et al. Recurrence and survival after random assignment to laparoscopy versus laparotomy for comprehensive surgical staging of uterine cancer: Gynecologic Oncology Group LAP2 Study. *J Clin Oncol*. 2012;30:695-700.

A7. Kong A, Johnson N, Kitchener HC, et al. Adjuvant radiotherapy for stage I endometrial cancer: an updated Cochrane systematic review and meta-analysis. *J Natl Cancer Inst*. 2012;104:1625-1634.

A8. Nout RA, van de Poll-Franse LV, Lybeert ML, et al. et al. Long-term outcome and quality of life of patients with endometrial carcinoma treated with or without pelvic radiotherapy in the post operative radiation therapy in endometrial carcinoma 1 (PORTEC-1) trial. *J Clin Oncol*. 2011;29: 1692-1700.

A9. Nout RA, Putter H, Jurgenliemk-Schulz IM, et al. Five-year quality of life of endometrial cancer patients treated in the randomised Post Operative Radiation Therapy in Endometrial Cancer (PORTEC-2) trial and comparison with norm data. *Eur J Cancer*. 2012;48:1638-1648.

A10. Ledermann J, Harter P, Gourley C, et al. Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. *N Engl J Med*. 2012;366:1382-1392.

A11. Vergote I, Trope CG, Amant F, et al. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. *N Engl J Med*. 2010;363:943-953.

A12. Burger RA, Brady MF, Bookman MA, et al. Incorporation of bevacizumab in the primary treatment of ovarian cancer. *N Engl J Med*. 2011;365:2473-2483.

A13. Buys SS, Partridge E, Black A, et al. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *JAMA*. 2011;305:2295-2303.

A14. Donnez J, Tatarchuk TF, Bouchard P, et al. Ulipristal acetate versus placebo for fibroid treatment before surgery. *N Engl J Med*. 2012;366:409-420.

A15. Donnez J, Tomaszewski J, Vazquez F, et al. Ulipristal acetate versus leuprolide acetate for uterine fibroids. *N Engl J Med*. 2012;366:421-432.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Siegel R, Ma J, Zou Z, et al. Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64:9-29.
2. Crosbie EJ, Einstein MH, Franceschi S, et al. Human papillomavirus and cervical cancer. *Lancet*. 2013;382:889-899.
3. Jemal A, Simard EP, Dorell C, et al. Annual Report to the Nation on the Status of Cancer, 1975-2009, featuring the burden and trends in human papillomavirus (HPV)-associated cancers and HPV vaccination coverage levels. *J Natl Cancer Inst*. 2013;105:175-201.
4. Partridge EE, Abu-Rustum N, Giuliano A, et al. Cervical cancer screening. *J Natl Compr Canc Netw*. 2014;12:333-341.
5. Setiawan VW, Yang HP, Pike MC, et al. Type I and II endometrial cancers: have they different risk factors? *J Clin Oncol*. 2013;31:2607-2618.
6. Cancer Genome Atlas Research Network, Kandoth C, Schultz N, Cherniack AD, et al. Integrated genomic characterization of endometrial carcinoma. *Nature*. 2013;497:67-73.
7. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011;474:609-615.
8. Bolton KL, Chenevix-Trench G, Goh C, et al. Association between *BRCA1* and *BRCA2* mutations and survival in women with invasive epithelial ovarian cancer. *JAMA*. 2012;307:382-390.
9. Rossing MA, Wicklund KG, Cushing-Haugen KL, et al. Predictive value of symptoms for early detection of ovarian cancer. *J Natl Cancer Inst*. 2010;102:222-229.
10. Jayson GC, Kohn EC, Kitchener HC, et al. Ovarian cancer. *Lancet*. 2014;384:1376-1388.
11. Aghajanian C, Blank SV, Goff BA, et al. OCEANS: a randomized, double-blind, placebo-controlled phase III trial of chemotherapy with or without bevacizumab in patients with platinum-sensitive recurrent epithelial ovarian, primary peritoneal, or fallopian tube cancer. *J Clin Oncol*. 2012;30:2039-2045.
12. Pennington KP, Walsh T, Harrell MI, et al. Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. *Clin Cancer Res*. 2014;20:764-775.
13. Hatzipetros I, Gocze PM, Farkas B. Oral contraceptive pills as primary prevention for ovarian cancer: a systematic review and meta-analysis. *Obstet Gynecol*. 2013;122:1114.
14. Trabert B, Ness RB, Lo-Ciganic WH, et al. Aspirin, nonaspirin nonsteroidal anti-inflammatory drug, and acetaminophen use and risk of invasive epithelial ovarian cancer: a pooled analysis in the Ovarian Cancer Association Consortium. *J Natl Cancer Inst*. 2014;106:djt431.
15. Bulun SE. Uterine fibroids. *N Engl J Med*. 2013;369:1344-1355.
16. Wright JD, Tergas AI, Burke WM, et al. Uterine pathology in women undergoing minimally invasive hysterectomy using morcellation. *JAMA*. 2014;312:1253-1255.

REVIEW QUESTIONS

1. Which gynecologic cancer is responsible for the most deaths in the United States each year?

A. Cervical cancer
B. Endometrial cancer
C. Uterine sarcomas
D. Ovarian cancer
E. Vulvar cancer

Answer: D As noted in the chapter, endometrial cancer is more than twice as common as ovarian cancer, yet ovarian cancer is more likely to be discovered at an advanced stage. The number of deaths in the United States is roughly twice the number of endometrial cancer deaths. Cervix cancer is much more common in other parts of the world where vaccination and screening are less prevalent.

2. Which of the following women can stop routine cytologic screening for cervical cancer?

A. A 21-year-old woman immunized at age 19 against HPV
B. A 38-year-old woman who has completed her childbearing
C. A *BRCA1* carrier who underwent a risk-reducing salpingo-oophorectomy 5 years ago
D. A 48-year-old woman who underwent a total hysterectomy for benign uterine bleeding 3 years ago
E. A 26 year old graduate student with 3 prior negative Pap tests.

Answer: D The immunization for HPV needs to occur before sexual initiation and generally must occur in the early teens at the latest. HPV vaccination has no effect on patients previous infected with HPV. Neither pregnancy nor parity is a major risk factor for cervical cancer, and salpingo-oophorectomy has no benefit for this purpose. However, a woman who underwent a total hysterectomy has no cervical tissue and no risk.

3. Which endometrial cancer stage grouping includes patients most likely to receive a survival benefit from adjuvant whole-pelvic radiation therapy?

A. Stage IA, grade 1
B. Stage IB, grade 2
C. Stage IIIC1, grade 3
D. Stage IVB, grade 3
E. Stage IA, grade 3

Answer: C Patients with early-stage (stage I) disease will have decreased local recurrence rates, but no overall survival advantage has been demonstrated to date. A patient with stage III disease with positive pelvic nodes may benefit. Patients with stage IVB will not have disease that can be encompassed in a single field.

4. Which ovarian cancer patient can have postsurgical observation without chemotherapy?

A. Stage IA, grade 3 serous ovarian cancer
B. Stage IB, grade 1 serous ovarian cancer
C. Stage 1C, grade 2 cancer (based on positive peritoneal cytology)
D. Stage 2A, grade 1 optimally debulked ovarian cancer
E. Stage IIIA, endometrioid ovarian cancer

Answer: B Only stage IA or IB, grade 1 cancers are suitable for observation.

5. Which of the following is the most common complication of chemoradiation for cervical cancer?

A. Vascular insufficiency fracture of the pubic ramus
B. Transitional cell cancer of the bladder
C. Intermittent diarrhea
D. Chronic anemia
E. Rectovaginal fistula

Answer: C Diarrhea can be seen in 10% or more of patients following pelvic radiation. Insufficiency fractures occur less frequently, and radiation-induced tumors are quite rare.

200

TESTICULAR CANCER

LAWRENCE H. EINHORN

EPIDEMIOLOGY

Testicular tumors are relatively uncommon and account for only 1% of male malignancies in the United States. The incidence of testicular cancer is increasing globally, although mortality rates remain low.[1] The primary age group is 15 to 35 years for nonseminomatous tumors and a decade older for seminomas.

Male patients with a history of cryptorchidism have a 10- to 40-fold increased risk for the development of testicular cancer. The normally descended testis in these men is also at higher risk, suggesting a dysgenetic abnormality.

PATHOBIOLOGY

More than 95% of tumors of the testis originate from germ cells.[2] Germ cell tumors can be seminomas or nonseminomatous germ cell tumors. Seminomas are more likely to be confined to the testis (stage I) and small-volume metastases to the retroperitoneal lymph nodes are exquisitely sensitive to radiation therapy. Pure seminomas never have elevated serum α-fetoprotein (AFP) levels. Nonseminomatous germ cell tumors consist of embryonal cell carcinomas, choriocarcinomas, yolk sac tumors, or teratomas, alone or mixed with other elements. Teratomas do not secrete human chorionic gonadotropin (hCG) or AFP and do not usually metastasize; they grow by local extension and are completely resistant to radiation therapy and chemotherapy.

Most germ cell testicular cancers in adults are associated with the cytogenetic abnormality i12p—an isochromosome of the short arm of chromosome 12—which is a highly specific finding in germ cell tumors. Genome-wide associated studies (GWAS) have identified association with variants in several genes, most strongly with *KITLG*.[3] Sertoli cell tumors, Leydig cell tumors, and lymphomas are the most common non–germ cell tumors. In men older than 60 years, most tumors are non-Hodgkin lymphoma (Chapter 185), with a predilection for bilateral involvement.

CLINICAL MANIFESTATIONS

Most patients with testicular cancer are initially evaluated because of testicular pain or because of a mass in or enlargement of one testis. Others are asymptomatic, and the cancer is first detected during a routine physical examination. Less commonly, the diagnosis is made during an evaluation for infertility, in part because testicular cancer can cause oligospermia.

Metastatic spread is either lymphatic or hematogenous. Lymphatic metastases usually go initially to the ipsilateral retroperitoneal lymph nodes, where they may be associated with flank pain. Lymphatic metastases may continue in a superior direction to the posterior mediastinum and eventually to the left supraclavicular lymph nodes. A large retroperitoneal mass or a supraclavicular lymph node may be palpable on physical examination. Hematogenous spread usually occurs first to the pulmonary parenchyma bilaterally. Pulmonary symptoms such as chest pain, shortness of breath, dyspnea on exertion, coughing, or hemoptysis are seen only with extensive pulmonary metastases. Other sites of hematogenous spread include the liver, bone, or brain. Significant elevation of the serum hCG level may produce gynecomastia.

DIAGNOSIS

Patients with a palpable mass in the testis should be suspected of having testicular cancer, especially if there is a history of cryptorchidism. Other causes of testicular and scrotal abnormalities may be included in the differential diagnosis. Acute pain in the testis suggests torsion. Painful enlargement

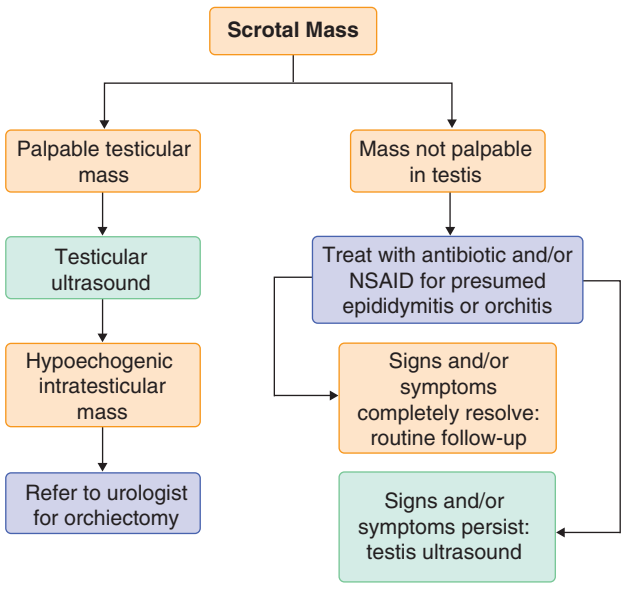

FIGURE 200-1. Management of a scrotal mass. NSAID = nonsteroidal anti-inflammatory drug.

may be due to a hydrocele, which may be caused by an underlying primary testicular malignancy. Pain and tenderness adjacent to the testis may be due to epididymitis or a varicocele. Tenderness of the testis itself on physical examination may reflect orchitis. However, an underlying neoplasm should always be considered.

Any testicular symptoms, including pain or a suspected mass, require evaluation. Testicular ultrasound is the test of choice in all suspicious cases. A hypoechogenic mass within the testis must be presumed to be testicular cancer and requires referral to a urologist (Fig. 200-1).

When orchiectomy reveals the diagnosis of testicular cancer, a staging evaluation is performed to determine the extent of disease and appropriate therapy. Clinical stage I disease is confined to the testis; stage II disease reflects spread to the retroperitoneal lymph nodes; and stage III is supradiaphragmatic disease, with either nodal metastases to the posterior mediastinum or supraclavicular region or hematogenous spread, especially to the lungs.

In addition to a full history and physical examination, serum hCG and AFP levels should be determined. Because the serum half-life is 1 day for hCG and 5 days for AFP, an AFP level of 1000 may take more than a month to normalize after orchiectomy, even if the tumor has been completely removed and there are no metastases. Imaging studies to define the extent of disease include abdominal and pelvic computed tomography (CT) and a chest radiograph. If the chest radiograph does not show pulmonary metastases, a chest CT scan should be performed. Bone scans and head CT scans can be reserved for patients with symptoms suggestive of osseous and central nervous system metastases, respectively.

PREVENTION AND TREATMENT

Rx

A careful testicular examination is a mandatory part of the physical examination in men, especially young men (Chapter 15), and is the key means of detecting tumors at an early stage.[4] With the patient lying supine or standing up, the testis is gently palpated with the thumb and second and third fingers; the entire anterior, posterior, and lateral surfaces of the testis should be examined. Men ages 15 to 34 years should be taught to perform the examination on themselves.[5]

Local and Regional Disease

Seminomas

Approximately 70% of seminomas are initially diagnosed at clinical stage I.[6] Although the cure rate with orchiectomy alone is 85%, treatment also can include 2000 cGy para-aortic irradiation or adjuvant carboplatin. The preferred option, however, is surveillance, which avoids unnecessary therapy in 85% of patients as well as potential late consequences of therapy.

Twenty percent of patients with seminoma are initially found to have stage II disease (positive abdominal CT scan). For these patients, radiation therapy has a 90% cure rate; in patients who are not cured by radiation therapy, subsequent combination chemotherapy (cisplatin combined with etoposide, with or without bleomycin) is usually curative. If the transverse diameter of the tumor is greater than 3 cm, if there are multiple anatomic levels of nodal metastases, or if stage III disease is present, the preferred initial treatment is cisplatin combination chemotherapy without irradiation; the cure rate is 90 to 100% unless there are non-pulmonary visceral metastases.[7]

Nonseminomatous Germ Cell Tumors

Management of clinical stage I nonseminomatous germ cell tumors begins with orchiectomy, which has a 70% cure rate. This is followed by either retroperitoneal lymph node dissection; one course of bleomycin, etoposide, and cisplatin (BEP); or close surveillance (which can detect metastases early and guide curative chemotherapy). Most relapses occur in the first year, during which meticulous surveillance should include history and physical examination, serum markers, and chest radiograph every 2 months and abdominal CT scans every 4 months. All studies are performed every 4 months during the second year, every 6 months during the third to fifth years of surveillance, and annually thereafter. However, abdominal CT scans are performed just every 6 months during the second year of surveillance, but annually in years 3-5, and then are discontinued. The physical examination should include palpation of the remaining testis, because patients have a 1 to 2% chance of developing a contralateral primary tumor. The major complication of retroperitoneal lymph node dissection is inadvertent severing of the sympathetic plexus, with resultant retrograde ejaculation or failure of ejaculation. Nerve-sparing retroperitoneal lymph node dissection can retain antegrade ejaculation in more than 95% of patients. Some centers advocate primary chemotherapy for high-risk clinical stage I disease (embryonal predominant with vascular invasion) with cisplatin combination chemotherapy, but surveillance is still an option in these patients.

For clinical stage II disease with persistently elevated serum markers or a transverse tumor diameter greater than 3 cm, chemotherapy is preferred. Other patients with stage II disease are treated with retroperitoneal lymph node dissection, often followed by close surveillance (as described earlier, but without abdominal CT scans) or adjuvant chemotherapy. Testicular cancer has a higher cure rate with surgery alone, despite nodal metastases, than any other cancer.

TABLE 200-1 DEFINITION OF POOR RISK DISEASE (ALL NONSEMINOMATOUS GERM CELL TUMOR)

Presence of any of the following:
- Human chorionic gonadotropin (hCG) > 50,000 mIU/mL
- α-Fetoprotein (AFP) > 10,000 ng/mL
- Non-pulmonary visceral metastases (e.g., bone liver, brain)
- Primary mediastinal nonseminomatous germ cell tumor

Chemotherapy for Disseminated or Persistent Disease

The combination of bleomycin, etoposide, and cisplatin (BEP) repeated every 3 weeks for three or four courses cures 70% of patients with metastatic disease and is the standard chemotherapy for disseminated testicular cancer. Poor-risk disease (Table 200-1) has a 50 to 60% cure rate with standard three-drug therapy, intermediate-risk disease (hCG 5000 to 50,000 IU/mL or AFP 1000 to 10,000 ng/mL) has a 75% cure rate, and all other forms of metastatic disease (good risk) have a 90 to 100% cure rate. Approximately 60% of patients receiving chemotherapy have good-risk disease, and just three courses of BEP is an adequate duration of therapy.

In the 30% of metastatic germ cell tumors that are not cured by initial combination chemotherapy, the use of standard-dose salvage therapy (ifosfamide, cisplatin, and either vinblastine or paclitaxel)[8] or high-dose carboplatin and etoposide therapy, followed by peripheral blood stem cell transplantation, can cure 25 to 70% of refractory cases, depending on patient characteristics. Late relapse beyond 2 years from completion of chemotherapy occurs in 2 to 3% of patients with metastatic disease. This usually is manifested by a rise in serum AFP. Optimal approach is evaluation for site(s) of metastases and surgical resection if feasible.

PROGNOSIS

Sequelae in Long-Term Survivors

Testicular cancer is the most curable solid tumor, with a 10-year survival rate of more than 95%. However, the typically normal life expectancy of individuals with the disease who were treated at a young age has led to the emergence of significant morbidities.

Testicular cancer survivors, particularly those who were initially treated with a combination of chemotherapy and irradiation, are at increased risk for developing second malignancies. These may include a wide variety of solid tumors and hematologic malignancies. Etoposide and cisplatin are associated with cumulative dose-dependent development of myelodysplastic syndrome (Chapter 182) and secondary leukemia.[9] Patients are also at increased risk for developing metachronous contralateral testicular cancer.

Increased cardiovascular risk in long-term survivors manifests most prominently as hyperlipidemia (Chapter 206) and metabolic syndrome (Chapter 229). The risk for metabolic syndrome is especially increased in survivors with testosterone levels in the lowest quadrille.[10] The clinical significance of low-grade hypogonadism among testicular cancer survivors has not been well studied. Although 10-year paternity rate in testicular cancer survivors is reduced by 30% compared with that in the general population, the majority of individuals who attempt paternity after treatment will become biologic fathers without medical assistance. Other long-term complications in survivors may include peripheral neuropathy, ototoxicity, and chronic renal insufficiency.

Grade A References

A1. Jones WG, Fossa SD, Mead GM, et al. Randomized trial of 30 versus 20 Gy in the adjuvant treatment of stage I testicular seminoma. *J Clin Oncol*. 2005;23:1200-1208.

A2. Oliver RTD, Mason MD, Mead GM, et al. Radiotherapy versus single-dose carboplatin in adjuvant treatment of stage I seminoma: a randomised trial. *Lancet*. 2005;366:293-300.

A3. Nichols C, Roth B, Albers P, et al. Active surveillance is the preferred approach to clinical stage I testicular cancer. *J Clin Oncol*. 2013;31:3490-3493.

A4. Alberts P, Siener R, Krege S, et al. Randomized phase III trial comparing retroperitoneal lymph node dissection with one course of bleomycin and etoposide plus cisplatin in the adjuvant treatment of clinical stage I non-seminomatous germ cell tumors. *J Clin Oncol*. 2008;26:2966-2972.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Shanmugalingam T, Soultati A, Chowdhury S, et al. Global incidence and outcome of testicular cancer. *Clin Epidemiol*. 2013;5:417-427.
2. Hanna N, Einhorn LH. Testicular cancer: a reflection on 50 years of discovery. *J Clin Oncol*. 2014;32:3085-3092.
3. Nallu A, Mannuel HD, Hussain A. Testicular germ cell tumors: biology and clinical update. *Curr Opin Oncol*. 2013;25:266-272.
4. U.S. Preventive Services Task Force. Screening for testicular cancer: U.S. Preventive Services Task Force reaffirmation recommendation statement. *Ann Intern Med*. 2011;154:483-486.
5. Akar SZ, Bebiş H. Evaluation of the effectiveness of testicular cancer and testicular self-examination training for patient care personnel: intervention study. *Health Educ Res*. 2014;29:966-976.
6. Hanna N, Einhorn LH. Testicular cancer—discoveries and updates. *N Engl J Med*. 2014;371:2005-2016.
7. Fosså SD, Cvancarova M, Chen L, et al. Adverse prognostic factors for testicular cancer-specific survival: a population-based study of 27,948 patients. *J Clin Oncol*. 2011;29:963-970.
8. O'Carrigan B, Grimison P. Current chemotherapeutic approaches for recurrent or refractory germ cell tumors. *Urol Oncol*. 2014;[Epub ahead of print].
9. Abouassaly R, Fossa SD, Giwercman A, et al. Sequalae of treatment in long-term survivors of testis cancer. *Eur J Urol*. 2011;60:516-626.
10. Willemse PM, Burggraaf J, Hamdy NA, et al. Prevalence of the metabolic syndrome and cardiovascular risk in chemotherapy-treated testicular germ cell tumour survivors. *Br J Cancer*. 2013;109:60-67.

REVIEW QUESTIONS

1. A 25-year-old man presents with a mass in his left testis, elevated hCG and α-fetoprotein, 10-cm retroperitoneal mass, and multiple bilateral pulmonary metastases. An orchiectomy revealed embryonal cell carcinoma + teratoma. He is treated with bleomycin + etoposide + cisplatin (BEP) and achieves a complete remission except for a persistent 3-cm retroperitoneal mass. This is subsequently resected and revealed mature teratoma. Seven years later, routine follow-up revealed an elevated serum α-fetoprotein of 150 ng/mL, normal hCG, and normal history and physical examination. His chest CT is normal, but abdominal CT revealed a left-sided pelvic mass measuring 5 cm in size. You now recommend:

A. Fine-needle aspiration of pelvic mass.
B. Chemotherapy with paclitaxel + ifosfamide + cisplatin (TIP).
C. Chemotherapy with vinblastine + ifosfamide + cisplatin (VeIP).
D. High-dose chemotherapy with peripheral blood stem cell support.
E. Surgical resection of pelvic mass.

Answer: E This patient has experienced a late relapse. This happens in approximately 2% of patients seemingly cured with initial chemotherapy. Typically, this is first diagnosed during annual follow-up after 5 years from initial chemotherapy and is manifested by an elevated serum AFP and nodal metastasis. The optimal curative strategy is surgery, not salvage chemotherapy.

2. Using the International Staging System for non-seminomatous germ cell tumors, all of the following constitute advanced disease (poor risk) *except:*

A. Histology of pure choriocarcinoma.
B. Serum α-fetoprotein greater than 10,000 ng/mL.
C. Presence of osseous metastases.
D. Primary mediastinal yolk sac tumor with 8-cm anterior mediastinal mass as only evidence of disease.

Answer: A Histology is not an independent prognostic variable. Any non-pulmonary visceral metastasis, AFP > 10,000 or any primary mediastinal non-seminomatous germ cell tumor constitutes poor-risk disease.

3. A 27-year-old man underwent a left orchiectomy with pathologic findings revealing embryonal cell carcinoma. His pre-orchiectomy serum hCG was 90 MIU/mL and AFP 130 ng/mL. Chest CT was normal, as was physical examination post-orchiectomy. Abdominal CT revealed a 3-cm left para-aortic node. One week later, hCG normalized, but AFP rose to 180 ng/mL. The correct recommendation would be:

A. Repeat serum AFP 2 weeks later
B. Fine-needle biopsy of left retroperitoneal lymph node
C. Three courses of bleomycin + etoposide + cisplatin (BEP)
D. Four courses of bleomycin + etoposide + cisplatin (BEP)

Answer: C He has a rising serum AFP and requires appropriate chemotherapy for his good-risk metastatic disease with three courses of BEP.

4. Which statement about testicular seminoma is incorrect?

A. Slight elevation of AFP may be present
B. Slight elevation of hCG may be present
C. Most patients present with clinical stage I disease
D. Radiation therapy is an option for clinical stage II disease with abdominal nodes < 3 cm

Answer: A Most seminomas are clinical stage I. Slight elevations in serum hCG may be observed, but seminomas are never associated with elevated AFP. Radiation therapy remains an option for clinical stage II seminoma.

5. A 35-year-old patient presents with an enlarged left supraclavicular lymph node. An excisional biopsy and complete resection reveals pure seminoma. A testis ultrasound is abnormal, disclosing a small tumor in the right testis. Orchiectomy pathology is also determined to be pure seminoma. Abdominal and pelvic CT scans abnormal with a 2- × 2-cm inter-aortocaval node. Chest CT is normal. What is the preferred treatment for this patient?

A. Radiotherapy to abdominal nodes and left supraclavicular fossa
B. Radiotherapy to abdominal nodes
C. Retroperitoneal lymph node dissection (RPLND)
D. Chemotherapy with BEP × 3

Answer: D He has stage III seminoma. The left supraclavicular metastasis was removed, but this is still stage III disease and radiotherapy is inappropriate. RPLND is never an option for initial therapy of any stage of seminoma.

201 PROSTATE CANCER

ERIC J. SMALL

DEFINITION

Prostate cancer is the most common noncutaneous malignant neoplasm in men in the United States, where it results in about 32,000 deaths each year, making it the second most common cause of cancer death in men. Prostate cancer is a single histologic disease with marked clinical heterogeneity ranging from indolent, clinically unimportant disease to a virulent, rapidly lethal phenotype.

EPIDEMIOLOGY

The incidence of clinically diagnosed prostate cancer reflects the effects of screening by the prostate-specific antigen (PSA) assay. Before PSA testing was available, about 19,000 new cases of prostate cancer were reported each year in the United States; this number reached 84,000 by 1993 and peaked at about 300,000 new cases in 1996. Since 1996, the reported annual incidence of prostate cancer in the United States has declined to about 190,000, a number that may more closely estimate the true incidence of clinically detectable disease. The death rate from prostate cancer has declined by about 1% per year since 1990. The age-specific decrease in the mortality rate has been greatest in men younger than 75 years. Men older than 75 years still account for two thirds of all prostate cancer deaths. Whether this decline is due to early detection (screening) or to improved therapy has not been established.

Risk factors for prostate cancer[1] include increasing age, family history, African American ethnicity, obesity, and dietary factors. Epidemiologic studies have suggested that nutritional factors such as reduced fat intake and increased soy protein may have a protective effect against the development of prostate cancer. The incidence of prostate cancer among African Americans is nearly twice that observed among white Americans. Prostate cancer is diagnosed in African Americans at a more advanced stage, and disease-specific survival is lower in African Americans. The relative contributions of biologic, genetic, and environmental differences, as well as differences in health care access, are not well established. Prior vasectomy and benign prostatic hypertrophy (BPH) (Chapter 129) do not increase the risk. Prostatic intraepithelial neoplasia, particularly when it is high grade, is recognized as a premalignant lesion, so its presence on biopsy increases the likelihood of subsequent malignancy.

PATHOBIOLOGY

Prostate cancer is more common among relatives of men with early-onset prostate cancer. However, although many genetic abnormalities with both loss and gain of function have been identified, none occur in more than 10% of patients with prostate cancer. For example, germline mutation (G84E) of $H0XB_{13}$, a homeobox transcription factor gene that is important in prostate development, has been associated with significantly increased risk of hereditary prostate cancer.[2] Consistent patterns of changes associated with an increased likelihood for the development of prostate cancer have not been identified.

Approximately half of prostate cancers demonstrate genetic rearrangements,[3,4] including fusion of promoters or enhancers of androgen-responsive genes such as *TMPRSS2* (transmembrane protease, serine 2) with oncogenic *ETS (E-26)* transcription factors such as *ERG (ETS-related gene)*. Gene fusions lead to overexpression of these oncogenic transcription factors.

Testosterone is required for maintenance of a normal, healthy prostatic epithelium, but it is also a prerequisite for the development of prostate cancer. Prostate cancers express robust levels of androgen receptor (AR), and signaling through the AR results in growth, progression, and invasion of prostate cancer. Inhibition of signaling, typically by the surgical or pharmacologic reduction of testosterone levels, results in prostate cancer apoptosis and involution. Ultimately, however, androgen-deprivation therapy (ADT) loses clinical efficacy. The biologic events surrounding the clinical development of "androgen-deprivation–resistant prostate cancer," also called castration-resistant prostate cancer (CRPC), are not well delineated, but amplification of the AR, which is a common event in these patients, presumably makes the cancer sensitive to minute levels of androgen or other ligands of the AR. Androgens produced through accessory pathways by the adrenal gland and upregulation of enzymatic regulators of androgen synthesis pathways within CRPC cells provide additional sources of ligand. The identification of AR splice variants that are constitutively active and ligand independent raises this as a potential mechanism by which true hormone resistance develops. The development of resistance to potent androgen signaling inhibition may be associated with the emergence of aggressive, lethal CRPC with neuroendocrine differentiation. Whether this reflects a process of transdifferentiation or clonal selection is not known.

CLINICAL MANIFESTATIONS

Most patients with early-stage, organ-confined disease are asymptomatic. Obstructive voiding symptoms (hesitancy, intermittent urinary stream, decreased force of stream) generally reflect locally advanced disease with growth into the urethra or bladder neck, although these symptoms can be indistinguishable from BPH (Chapter 129). Locally advanced tumors can also result in hematuria and hematospermia. Prostate cancer that has spread to the regional pelvic lymph nodes occasionally causes edema of the lower extremities or discomfort in the pelvic and perineal areas. Metastasis occurs most commonly to bone, where it is frequently asymptomatic, but it can also cause severe and unremitting pain. Bone metastasis can result in pathologic fractures or spinal cord compression. Although visceral metastases are rare as presenting features of prostate cancer, there is an increasing incidence of pulmonary, hepatic, pleural, peritoneal, and central nervous system metastases that appear to be treatment emergent.

DIAGNOSIS

More than 60% of patients with prostate cancer are asymptomatic, and the diagnosis is made solely because of an elevated screening PSA level. A palpable nodule on digital rectal examination (DRE), which is the next most common clinical presentation, generally prompts biopsy. Much less commonly, prostate cancer is diagnosed because of advanced disease that causes obstructive voiding symptoms, pelvic or perineal discomfort, lower extremity edema, or symptomatic bone lesions.

Although the DRE has a low sensitivity and specificity for the diagnosis of prostate cancer, biopsy of a nodule or area of induration reveals cancer 50% of the time, suggesting that prostate biopsy should be undertaken in all men with palpable nodules. The PSA level has a far better sensitivity but a low specificity because conditions such as BPH and prostatitis can cause false-positive PSA elevations (Chapter 129). By use of a PSA threshold of 4 ng/mL, 70% to 80% of tumors are detected. Far greater accuracy is achieved with age-specific PSA thresholds. Thus, for men age 40 to 49 years, a PSA greater than 2.5 is considered abnormal; for men 50 to 59 years, a PSA greater than 3.5 is abnormal; for men 60 to 69 years, a PSA greater than 4.5 should prompt further evaluation; and patients age 70 to 79 years should have a PSA of 6.5 or less. The positive predictive value for a single PSA level above 10 ng/mL is greater than 60% for cancer, but the positive predictive value for a PSA level between 4 and 10 ng/mL is only about 30%. Assays of the PSA fraction that circulates unbound (percentage of free PSA) may help distinguish prostate cancer from benign processes; in patients with PSA levels of 4 to 10 ng/mL, the percentage of free PSA appears to be an independent predictor of prostate cancer, and a cutoff value of free PSA less than 25% can detect 95% of cancers while avoiding 20% of unnecessary biopsies.

Transrectal ultrasonography with biopsies is indicated when the PSA level is elevated, when the percentage of free PSA is less than 25%, or when an abnormality is noted on DRE. Extended field specimens (preferably up to six biopsies on each side) are generally obtained. Seminal vesicles are sampled in high-risk patients. A bone scan is warranted only in patients with PSA levels above 10 ng/mL, and abdominal and pelvic computed tomography or magnetic resonance imaging is usually unrevealing in patients with PSA levels below 10 to 20 ng/mL.

The prognosis of patients with prostate cancer correlates with histologic grade and extent (stage) of disease. More than 95% of prostate cancers are adenocarcinomas, and multifocality is common. Although uncommon, a neuroendocrine variant is increasingly being identified, because of increased awareness but also as a manifestation of the development of resistance to potent androgen signaling inhibition. The histologic (Gleason) grade of adenocarcinomas range from 3 to 5. The Gleason score, which refers to the sum of the two most common histologic patterns seen on each tissue specimen, ranges from 6 (3 + 3) to 10 (5 + 5). In general, tumors are classified as

well differentiated (Gleason score of 6), of intermediate differentiation (Gleason score of 7), or poorly differentiated (Gleason score of 8-10). Neuroendocrine differentiation is a histologic diagnosis that is confirmed by staining for chromogranin A or synaptophysin.

Clinical stage is defined by the extent of disease based on the physical examination, imaging studies, and pathology. Stage T1 is nonpalpable prostate cancer detected only on pathologic examination, noted either incidentally after transurethral resection for benign hypertrophy (T1a and T1b) or on a biopsy specimen obtained because of an elevated PSA level (T1c, the most common clinical stage at diagnosis). Stage T2 is a palpable tumor that appears to be confined to the prostate gland (T2a in one lobe or T2b in two lobes), and stage T3 is tumor with extension through the prostatic capsule (T3a if it is focal or T3b if seminal vesicles are involved). T4 tumors are those with invasion of adjacent structures, such as the bladder neck, external urinary sphincter, rectum, levator muscles, or pelvic sidewall. Nodal metastases can be microscopic and detectable only by biopsy or lymphadenectomy, or they can be visible on imaging studies. Distant metastases are predominantly to bone, but occasional visceral metastases occur.

PREVENTION

The precise role for screening remains controversial.[5-7] Currently, many organizations recommend screening with PSA, but the U.S. Preventive Services Task Force recommends against screening.[8] Of two large randomized screening trials using PSA levels, one reported a reduction in prostate cancer–specific mortality,[■] but neither found an overall reduction in mortality rate.[■■] Overall, PSA screening reduces prostate cancer death at 11 years by 21% (absolute reduction, 0.10 deaths per 1000 persons-years or 1.07 deaths per 1000 men) but not all-cause mortality.[■] Although overall mortality from prostate cancer has fallen during the screening era, there is no direct evidence that there is a causal relationship.[■] Screening of men at high risk for developing prostate cancer (family history and African Americans) has not been specifically tested. Randomized trials have shown that vitamins C and selenium are not effective in preventing prostate cancer,[9] and vitamin E supplementation increases prostate cancer by 17%.[10] The use of 5α-reductase inhibitors (both finasteride and dutasteride) unambiguously reduces the risk for development of prostate cancer.[■■] However, this approach has not been widely adopted, primarily because of attendant side effects, most notably sexual dysfunction.

TREATMENT Rx

Localized Prostate Cancer

Principles of Therapy

The principal therapeutic options for men with localized prostate cancer include (1) active surveillance[11]; (2) retropubic or perineal radical prostatectomy, with or without postoperative radiation therapy to the prostate margins and pelvis; (3) external-beam radiation therapy; and (4) brachytherapy (either permanent or temporary radioactive seed implants), with or without external-beam radiation therapy to the prostate margins and pelvis.

Treatment options require individualization, taking into account the patient's comorbidity, life expectancy, likelihood of cure, and personal preferences based on an understanding of the potential morbidity associated with each treatment. A multidisciplinary approach to integrate surgery, radiation therapy, and androgen deprivation is increasingly recommended. For higher risk patients with a greater likelihood of nodal micrometastases, androgen deprivation is often combined with radiation therapy encompassing both the prostate and the pelvis. In patients at extremely high risk of micrometastatic disease or with comorbidities, systemic therapy alone without concurrent local therapy may be appropriate.

Prostate-specific antigen screening has led to the early detection of a large number of nonpalpable tumors, for which conventional clinical means of staging are inadequate. Thus, less emphasis is being placed on clinical stage, and more emphasis is being placed on Gleason score, PSA values, and other predictors of outcome. Careful risk assessment is required to identify patients who are appropriate candidates for definitive local treatment.

Several studies have confirmed that serum PSA level, clinical stage, and biopsy Gleason score can be used to predict the final pathologic stage after prostatectomy and that these are independent predictors of clinical outcome. For example, in a radiation therapy series, clinical stage T3 or higher, PSA level above 10 ng/mL, and biopsy Gleason score of 7 or higher were risk factors for poor outcome (death or PSA elevation); the 5-year survival rate without PSA elevation was 85% for patients with none of these adverse features (good risk), 65% for patients with one adverse feature (intermediate risk), and 35% for patients with two or three adverse features (poor risk). Similar statistics are cited in radical prostatectomy series. The percentage of biopsy specimens that are positive and the rate of increase in the PSA value are each independent predictors of outcome after radical prostatectomy and can be used to counsel patients about their therapeutic options. A number of multivariable prognostic models have been developed and validated and have been used to develop simple nomograms or online risk calculators.

Low- to Intermediate-Risk Disease

In one randomized trial of patients younger than 75 years with clinical stage T1b, T1c, or T2 prostate cancer, radical prostatectomy compared with no therapy significantly reduced the relative risk of death caused by prostate cancer by about 40% (an 11% absolute risk reduction) and the overall mortality rate by a similar absolute amount at 18 years.[■] Reductions in progressive disease and metastases were also significant. The adverse effects on quality of life differed between the two strategies—more sexual dysfunction and urinary leakage after radical prostatectomy and more urinary obstruction with active surveillance—but were of similar magnitude.[■] Nerve-sparing radical prostatectomy was not routinely performed in this study, and many patients already had palpable disease, so the implications for less advanced disease with newer surgical techniques are not known. In another randomized trial of men with localized prostate cancer, radical prostatectomy did not significantly reduce all-cause or prostate-cancer mortality compared with observation through at least 12 years of follow-up, although patients with a PSA greater than 10 ng/mL and possibly patients with intermediate-risk or high-risk tumors may benefit.[■]

Active surveillance is an increasingly important option for men with low-risk disease. Careful observation of PSA and serial biopsies identifies patients who will never need local therapy. In large active surveillance series, anywhere from 50% to 75% of patients never require local therapy.

Nerve-sparing procedures and careful dissection techniques have decreased the risk of postoperative urinary incontinence and impotence. Postoperative urinary incontinence is reported to occur in fewer than 10% of cases. Postoperative impotence depends on a variety of factors, including the patient's age, preoperative erectile function, extent of cancer, and whether a nerve-sparing procedure was performed. In general, impotence rates of 10% to 50% are cited. Robotic-assisted laparoscopic prostatectomies have gained popularity but have not been shown to result in better outcomes. After a radical prostatectomy, the PSA should become undetectable; a detectable PSA implies the presence of cancer cells, either locally or at a metastatic site. Immediate (adjuvant) postoperative radiation therapy improves biochemical progression-free survival and local control in patients with one or more pathologic risk factors (capsule penetration, positive surgical margins, invasion of a seminal vesicle) after radical prostatectomy.

Conventional external-beam radiation therapy is being replaced by three-dimensional conformal radiation therapy or intensity-modulated radiation therapy, which permits higher doses to the target tissue with less toxicity. Randomized trials suggest a benefit with higher doses of radiation. Brachytherapy, which is the placement of permanent or temporary radioactive seeds directly into the prostate, is adequate for intracapsular disease with no more than minimal transcapular extension; otherwise, it should be combined with external-beam radiation therapy.

High-Risk Disease

Patients with adverse risk features (Gleason score of 8 to 10, PSA >10, stage T3) are at high risk of nodal and micrometastatic disease and are generally treated with aggressive local therapy in combination with androgen deprivation, which is synergistic with radiation therapy.[■■] Taken in the aggregate, trials suggest that 4 months of androgen deprivation with radiation therapy can improve local control and prolong progression-free survival in patients with intermediate-risk features, and long-term androgen deprivation (up to 3 years) prolongs local control, progression-free survival, and overall survival in patients with high-risk features compared with radiation therapy alone.[■■] Patients with stage T3 disease and Gleason scores of 7 have intermediate outcomes, with 8-year survival rates of about 70%; patients with stage T3 disease and Gleason scores of 8 to 10 have 8-year survival rates after radiation therapy of about 50%. Compared with ADT alone, the addition of radiation therapy to ADT improved the overall survival rate at 7 years from 66% to 74% in patients with locally advanced prostate cancer.[■] Several randomized controlled trials suggest that patients with high-risk disease who are treated surgically and have capsule penetration, positive margins, or seminal vesicle involvement should receive immediate adjuvant radiation therapy.

Recurrent Disease

Between 30% and 50% of men treated with radiation therapy or prostatectomy have evidence of disease recurrence, as defined by a climbing PSA level. PSA doubling time is predictive of survival, and a short PSA doubling time (<3-6 months) is associated with a higher likelihood of systemic disease. For selected patients with clear local recurrences, low PSA levels, and prolonged PSA doubling times, local salvage therapy (surgery for patients previously treated with radiation therapy, radiation therapy for patients previously treated with surgery, and androgen deprivation) can be considered. Although

ADT readily controls PSA levels, it is unknown whether it prolongs life in patients with PSA-only recurrent disease.

Advanced Disease

In patients whose radical prostatectomy surgery reveals microscopic involvement of lymph nodes, immediate androgen deprivation prolongs survival compared with deferment of androgen deprivation until osseous metastases are detected.[12] Similarly, patients who are at high risk of nodal invasion and who undergo external-beam radiation benefit from concurrent short-term hormonal therapy.

In patients with newly diagnosed metastatic prostate cancer, ADT is the mainstay of treatment and results in symptomatic improvement and disease regression in approximately 80% to 90% of patients. ADT can also be achieved by orchiectomy or by medical castration with a luteinizing hormone–releasing hormone (LHRH) agonist (leuprolide acetate, goserelin acetate). Intermittent ADT, typically consisting of 12 months of therapy followed by time off therapy before resuming ADT is an option in patients with nonmetastatic disease and is potentially useful in patients with metastatic prostate cancer.

Some LHRH agonists cause a transient worsening of signs and symptoms during the first week of therapy as a result of a surge in luteinizing hormone and testosterone, which peaks within 72 hours; an antiandrogen (flutamide, bicalutamide, or nilutamide) should be given with the first LHRH injection to prevent a tumor flare. Medical castration occurs within 4 weeks. The duration of hormone sensitivity is 5 to 10 years for node-positive or high-risk localized (or recurrent) prostate cancer, but it is closer to 24 months in patients with overt metastatic disease. The most common side effects of androgen ablation are loss of libido, impotence, hot flashes, weight gain, fatigue, anemia, and osteoporosis. Bisphosphonates and denosumab reduce bone mineral loss associated with androgen deprivation.

Castration-Resistant Prostate Cancer

Typically, the first manifestation of resistance to androgen deprivation is a climbing PSA level in the setting of anorchid levels of testosterone.[13] In about 15% of patients, discontinuation of antiandrogen therapy (flutamide, bicalutamide, nilutamide) while continuing treatment with LHRH agonists results in a PSA decline that can be associated with symptomatic improvement and can persist for 4 months or more. If antiandrogen withdrawal fails, treatment with secondary hormonal manipulations, such as ketoconazole or estrogens, is appropriate. Sipuleucel-T is an autologous dendritic cell product that has been shown to prolong life and is appropriate for patients with castration-resistant metastatic prostate cancer who do not have cancer-associated pain, visceral metastases, rapidly progressive disease, or the need for systemic steroids.

As noted earlier, AR amplification is a common event in castration-resistant prostate cancer. Consequently, novel agents that target the AR axis have demonstrated dramatic improvements in response proportion, progression-free survival, and overall survival. Agents that target either the ligand (e.g., the androgen biosynthesis inhibitor abiraterone acetate), as well as agents that target the receptor (e.g., enzalutamide, a direct AR antagonist) in this pathway have been shown to increase survival in metastatic castration-resistant prostate cancer patients, have been approved for use by the Food and Drug Administration, and have dramatically changed the therapeutic landscape. The optimal sequencing or combination of these agents is under investigation, as is the treatment of metastatic castration-resistant prostate cancer resistant to these agents, including treatment-emergent neuroendocrine variants.

Radium 223 is an α-emitting agent that localizes to metastatic prostate cancer lesions in bone and has been shown to provide a survival advantage in patients with bone-predominant disease and no visceral metastases. The optimal sequencing and combination of this agent with others is under investigation.

Thereafter, treatment with chemotherapeutic regimens, such as docetaxel plus corticosteroids or mitoxantrone plus corticosteroids, may be effective. Randomized phase III trials have demonstrated a survival advantage of approximately 25% in patients receiving taxol-based therapy compared with mitoxantrone, and docetaxel-prednisone is now considered a standard therapeutic approach in patients with metastatic, androgen deprivation–resistant prostate cancer. After therapy with docetaxel, patients who remain candidates for further chemotherapy can be treated with cabazitaxel, an agent shown to prolong life in this group of patients. In general, serial PSA levels are the best (albeit imperfect) way to follow up with patients, and a decline of 30% to 50% is associated with improved survival. Zoledronic acid or denosumab is indicated in castration-resistant prostate cancer patients with bone metastases because each reduces the incidence of skeletal-related events.

Palliative Care

Many patients with advanced prostate cancer have bone pain or functional impairments that adversely affect quality of life, and the provision of appropriate palliative care is an integral component of their management. In addition to the usual analgesics, glucocorticoids serve as anti-inflammatory agents and can alleviate bone pain. For patients with widespread bone metastases and pain not easily controlled with analgesics or local irradiation, samarium-153 or radium-223 can be administered intravenously; they are selectively concentrated in bone metastases and are effective in alleviating pain in many patients.

The approach to the treatment of patients with prostate cancer is detailed in Table 201-1.

TABLE 201-1 APPROACH TO THE TREATMENT OF PROSTATE CANCER

EXTENT OF CANCER	THERAPEUTIC OPTIONS
Organ confined: low risk (usually T1 or T2, GS = 7, PSA <10 ng/mL)	Active surveillance Radical prostatectomy External-beam radiation therapy to prostate Brachytherapy
Organ confined: intermediate risk (usually T2, GS = 7, PSA = 10-20 ng/mL)	Active surveillance Radical prostatectomy External-beam radiation therapy to prostate, possibly to pelvis, with or without ADT Brachytherapy
Organ confined: high risk (usually T3, GS >7, PSA >20 ng/mL)	Radical prostatectomy (with adjuvant radiation therapy, if needed) External-beam radiation therapy to prostate and pelvis (usually with ADT) Brachytherapy plus radiation therapy (usually with ADT)
Climbing PSA level after local therapy	ADT: antiandrogen monotherapy or combined ADT Salvage radiation therapy (for patients with prior prostatectomy) Salvage radical prostatectomy (for patients with prior radiation therapy) Surveillance Investigational therapy
Node positive	ADT Pelvic or prostate radiation therapy + ADT Investigational therapy
Metastatic: untreated hormone-refractory prostate cancer	ADT Second-line hormones Sipuleucel-T immunotherapy Chemotherapy Investigational therapy

ADT = androgen-deprivation therapy; GS = Gleason score; PSA = prostate-specific antigen.

PROGNOSIS

In general, the 10-year PSA progression-free survival rate is 70% to 80% with well-differentiated tumors, whether treatment is with radiation therapy or surgery; 50% to 70% for intermediate risk tumors; and 30% for high risk tumors. For patients with a climbing PSA level after radical prostatectomy, time to detectable PSA, Gleason score at the time of prostatectomy, and PSA doubling time are important prognostic variables. The likelihood of bone metastases at 7 years ranges from 20% for good-prognosis patients to 80% for poor-prognosis patients. Patients require periodic surveillance and careful comprehensive medical care.[14]

For patients with microscopic nodal disease, the 10-year survival rate approaches 80% in men treated with androgen deprivation. The median survival period in men treated with androgen deprivation for established metastatic disease ranges from 2 to 6 years. The median survival period for men with metastatic castration-resistant prostate cancer approaches 3 years, a dramatic improvement from the survival reported even 5 years ago.

FUTURE DIRECTIONS

Molecular markers can not only identify patients at risk for the development of progressive disease but also act as therapeutic targets. In addition, the genomic characterization of prostate cancer subtypes will lead to risk-adapted therapy. Enhanced understanding of AR biology may permit the development of specific hormonal therapies and guide the more rational use of existing agents.

Grade A References

A1. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med*. 2009;360:1320-1328.

A2. Andriole GL, Grubb RL 3rd, Buys SS, et al. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med*. 2009;360:1310-1319.

A3. Schröder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. *N Engl J Med*. 2012;366:981-990.

A4. Ilic D, Neuberger MM, Djulbegovic M, et al. Screening for prostate cancer. *Cochrane Database Syst Rev*. 2013;1:CD004720.

A5. Andriole GL, Bostwick DG, Brawley OW, et al. Effect of dutasteride on the risk of prostate cancer. *N Engl J Med*. 2010;362:1192-1202.

A6. Thompson IM Jr, Goodman PJ, Tangen CM, et al. Long-term survival of participants in the prostate cancer prevention trial. *N Engl J Med*. 2013;369:603-610.

A7. Bill-Axelson A, Holmberg L, Garmo H, et al. Radical prostatectomy or watchful waiting in early prostate cancer. *N Engl J Med*. 2014;370:932-942.

A8. Steineck G, Helgesen F, Adolfsson J, et al. Quality of life after radical prostatectomy or watchful waiting. *N Engl J Med*. 2002;347:790-796.

A9. Wilt TJ, Brawer MK, Jones KM, et al. Radical prostatectomy versus observation for localized prostate cancer. *N Engl J Med*. 2012;367:203-213.

A10. D'Amico AV, Chen MH, Renshaw AA, et al. Androgen suppression and radiation vs radiation alone for prostate cancer: a randomized trial. *JAMA*. 2008;299:289-295.

A11. Jones CU, Hunt D, McGowan DG, et al. Radiotherapy and short-term androgen deprivation for localized prostate cancer. *N Engl J Med*. 2011;365:107-118.

A12. Bolla M, van Poppel H, Collette L, et al. Postoperative radiotherapy after radical prostatectomy: a randomized controlled trial (EORTC trial 22911). *Lancet*. 2005;366:572-578.

A13. Roach M 3rd, DeSilvio M, Lawton C, et al. Phase III trial comparing whole-pelvic versus prostate-only radiotherapy and neoadjuvant versus adjuvant combined androgen suppression: radiation Therapy Oncology Group 9413. *J Clin Oncol*. 2003;21:1904-1911.

A14. Warde P, Mason M, Ding K, et al. Combined androgen deprivation therapy and radiation therapy for locally advanced prostate cancer: a randomised, phase 3 trial. *Lancet*. 2011;378:2104-2111.

A15. Bria E, Cuppone F, Giannarelli D, et al. Does hormone treatment added to radiotherapy improve outcome in locally advanced prostate cancer?: meta-analysis of randomized trials. *Cancer*. 2009;115:3446-3456.

A16. Kantoff PW, Higano CS, Shore ND, et al. Sipuleucel-T immunotherapy for castration-resistant prostate cancer. *N Engl J Med*. 2010;363:411-422.

A17. de Bono JS, Logothetis CJ, Molina A, et al. Abiraterone and increased survival in metastatic prostate cancer. *N Engl J Med*. 2011;364:1995-2005.

A18. Ryan CJ, Smith MR, de Bono JS. Abiraterone in metastatic prostate cancer without previous chemotherapy. *N Engl J Med*. 2013;368:138-148.

A19. Scher HI, Fizazi K, Saad F, et al. Increased survival with enzalutamide in prostate cancer after chemotherapy. *N Engl J Med*. 2012;367:1187-1197.

A20. Beer TM, Armstrong AJ, Rathkopf DE, et al. Enzalutamide in metastatic prostate cancer before chemotherapy. *N Engl J Med*. 2014;371:424-433.

A21. Parker C, Nilsson S, Heinrich D, et al. Alpha emitter radium-223 and survival in metastatic prostate cancer. *N Engl J Med*. 2013;369:213-223.

A22. de Bono JS, Oudard S, Ozguroglu M, et al. Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. *Lancet*. 2010;376:1147-1154.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Roobal MJ, Carllson SV. Risk stratification in prostate cancer screening. *Nat Rev Urol*. 2013;10:38-48.
2. Ewing CM, Ray AM, Lange EM, et al. Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med*. 2012;366:141-149.
3. Roychowdhury S, Chinnaiyan AM. Advancing precision for prostate cancer through genomics. *J Clin Oncol*. 2013;31:1866-1873.
4. Berger MF, Lawrence MS, Demichelis F, et al. The genomic complexity of primary human prostate cancer. *Nature*. 2011;470:214-220.
5. Hayes JH, Barry MJ. Screening for prostate cancer with the prostate-specific antigen test: a review of current evidence. *JAMA*. 2014;311:1143-1149.
6. Heijnsdijk EA, Wever EM, Auvinen A, et al. Quality-of-life effects of prostate-specific antigen screening. *N Engl J Med*. 2012;367:595-605.
7. National Collaborating Centre for Cancer. Prostate Cancer: Diagnosis and Treatment. Cardiff UK; 2014.
8. Moyer VA, LeFevre ML, Siu AL, et al., on behalf of the U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2012;157:120-134.
9. Sandhu GS, Nepple KG, Tanagho YS, et al. Prostate cancer chemoprevention. *Semin Oncol*. 2013;33:4163-4174.
10. Klein EA, Thompson IM Jr, Tangen CM, et al. Vitamin E and the risk of prostate cancer: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *JAMA*. 2011;306:1549-1556.
11. Whitson JM, Porten SP, Carroll PR. Prostate cancer: reducing overtreatment: active surveillance in low-risk disease. *Nat Rev Urol*. 2011;8:124-125.
12. Mottet N, Bellmunt J, Bolla M, et al. EAU guidelines on prostate cancer. Part II: treatment of advanced, relapsing, and castration-resistant prostate cancer. *Eur Urol*. 2011;59:572-583.
13. Basch E, Loblaw DA, Oliver TK, et al. Systemic therapy in men with metastatic castration-resistant prostate cancer: American Society of Clinical Oncology and Cancer Care Ontario clinical practice guideline. *J Clin Oncol*. 2014;32:3436-3448.
14. Skolarus TA, Wolf AM, Erb NL, et al. American Cancer Society prostate cancer survivorship care guidelines. *CA Cancer J Clin*. 2014;64:225-249.

REVIEW QUESTIONS

1. Which of the following is not a risk factor for prostate cancer?

A. African American race
B. Benign prostatic hypertrophy
C. Obesity
D. Family history
E. Dietary factors

Answer: B Neither benign prostatic hypertrophy (BPH) nor a prior history of vasectomy is associated with an increased risk of prostate cancer. The incidence of prostate cancer among African Americans is nearly twice that observed among white Americans. Obesity is associated with several malignancies, and prostate cancer is among the prominent (along with breast cancer, gynecologic malignancies, esophageal adenocarcinoma, colorectal, and renal cell carcinoma). Family history is a significant risk, especially involving relatives of men with early age of onset of prostate cancer. Some dietary factors are likewise considered to enhance or reduce the risk of prostate cancer; the protective nutritional factors include reduced fat intake and increased soy protein (see the Epidemiology section).

2. A 65-year-old man in previously good health generally is found to have an area of induration on the prostate palpable on routine digital rectal examination. The next step should be

A. prostate biopsy.
B. PSA level.
C. imaging for visceral metastases.
D. cystoscopy.
E. sigmoidoscopy.

Answer: A Although the digital rectal examination has a low sensitivity and specificity for the diagnosis of prostate cancer, biopsy of a nodule or an area of induration reveals cancer 50% of the time, suggesting that prostate biopsy should be done in all such cases. The PSA (if it has not been already done) will be useful as a parameter of disease to follow if prostate cancer is pathologically diagnosed, but whatever its level is at this time it would not obviate the need for a biopsy. Imaging or procedures to evaluate the local extent of disease or metastases is premature before a pathological diagnosis is made (see the Diagnosis section).

3. Which of the following is effective in preventing prostate cancer?

A. Vitamin C
B. Selenium
C. Nonsteroidal anti-inflammatory drugs (NSAIDs)
D. Vitamin E
E. Finasteride

Answer: E Androgen inhibition by the 5α-reductase inhibitors finasteride and dutasteride has been shown to unambiguously reduce the risk of development of overall prostate cancers, although there was a subsequently disputed finding of an increase in the number of cases with specifically higher grade disease with these drugs. Randomized trials have shown that vitamin C and selenium are not effective in preventing prostate cancer. Vitamin E has not been demonstrated to be effective. NSAIDs have demonstrated effectiveness in prevention of some cancers, such as colorectal cancer, but there is no convincing evidence that its actions likewise apply to the prevention of prostate cancer (see the Prevention section).

202

MALIGNANT TUMORS OF BONE, SARCOMAS, AND OTHER SOFT TISSUE NEOPLASMS

JAMES H. DOROSHOW

PRIMARY BONE TUMORS

DEFINITION

Primary bone tumors arise from cells that are normal components of bone tissues and that have the potential to metastasize. They are relatively uncommon malignancies (accounting for 0.2% of all neoplasms in the Surveillance, Epidemiology, and End Results [SEER] database, with 1.8 new cases per 100,000 population per year). Malignant bone tumors must be distinguished from a variety of more common benign bone lesions, such as osteochondromas and enchondromas.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients with primary malignant and benign bone tumors present with pain, swelling, and occasionally pathologic fracture of the involved bone. If radiologic studies suggest a malignant primary bone tumor (see characteristics of each subtype described later), an orthopedic oncologist should be consulted before carrying out a biopsy because improper biopsy technique may compromise subsequent surgical care, particularly limb-sparing surgery. Staging of patients with bone tumors generally requires computed tomography (CT) scans of the chest, abdomen, and pelvis to evaluate whether metastatic disease is present. Characterization of the primary bone tumor may benefit from magnetic resonance imaging (MRI) assessment of soft tissue extension or CT scan assessment of cortical bone involvement, or both.

MAJOR PRIMARY MALIGNANT BONE TUMORS

Myeloma, the most common primary bone malignancy, is covered in Chapter 187.

Osteosarcoma

Osteosarcoma is the most common malignant sarcoma of bone, representing about 35% of cases. It has a bimodal age distribution with the highest incidence in patients younger than 20 years of age, most likely related to the normal rapid bone growth that occurs during adolescence. In this age group, most tumors arise in the metaphyseal areas of the long bones of the extremities, particularly around the knee. Males are affected more commonly than females at a ratio of 3 : 2. A second peak of incidence occurs in adults older than 60 years. The sites of origin in these older patients are somewhat more heterogeneous, with craniofacial and pelvic bones each accounting for 20% of tumors. Osteosarcomas are classified based on location, cell type, and tumor grade. All osteosarcomas contain varying amounts of osteoid, with most also containing some cartilage and fibrous tissue.[1,2] Radiographically, osteosarcomas usually present as mixed osteoblastic and osteolytic lesions, although pure forms of either appearance can occur. Periosteal elevation (Codman's triangle), cortical destruction, and tumor extension into soft tissue are common on plain radiographs or MRI.

The incidence of osteosarcoma is increased in families that carry germline deletions of retinoblastoma (*RB*), *TP53* (Li-Fraumeni), or *RecQ* DNA helicase (Rothman-Thompson, Werner, or Bloom syndrome) genes. Consistent with these observations, although most younger patients with osteosarcomas have no apparent predisposing factor or family history of bone tumors, alterations in the *TP53* and *RB* genes of such sporadic tumors occur in 40% and 60% of patients, respectively. In older patients, a variety of conditions may predispose to osteosarcoma, most convincingly antecedent Paget disease or prior radiation therapy.

TREATMENT

Rx

Osteosarcoma is a highly proliferative neoplasm that metastasizes rapidly, most often by hematogenous spread; the most common site of metastasis is the lung. Despite aggressive surgical resection of the primary bone tumor, the incidence of recurrence with metastatic disease is high in the absence of systemic treatment, consistent with the concept that most patients present with clinically inapparent micrometastatic disease. The development of effective systemic chemotherapy with doxorubicin and cisplatin, with or without methotrexate, has had a profoundly positive effect on treatment outcome, with 5-year disease-free survival rates exceeding 65% in patients younger than 40 years with nonmetastatic extremity tumors. However, subsequent progress has been less striking.[3] Most patients are managed with initial neoadjuvant chemotherapy, delayed resection of the primary tumor, and then further postoperative chemotherapy. Serum alkaline phosphatase levels, often elevated in patients with osteosarcoma, can be used to monitor disease status.

Modern surgical techniques have allowed resection of most extremity osteosarcomas without amputation. Although resection of lung metastases can be curative in about 20% of selected patients, detection of radiographically-apparent metastatic disease at presentation significantly worsens prognosis. Long-term disease control in older adults with osteosarcoma is substantially lower than in younger patients, with a 5-year overall survival rate of 22% in one series of patients older than 65 years, most likely because of fundamental differences in the underlying molecular pathophysiology of tumors in older adults. Although osteosarcoma is generally considered to be relatively radiation resistant, radiation can play a palliative role in selected patients.

Chondrosarcoma

Chondrosarcoma is a malignant tumor characterized by hyaline cartilage differentiation; it is the second most common sarcoma of bone, representing 25% of bone sarcomas. The peak incidence is in the fifth to seventh decades of life. The most common primary sites are in the pelvis, proximal femur, and proximal humerus. Patients present with long-standing complaints of swelling, pain, or both. Radiographically, chondrosarcoma usually appears as a mixed lytic and sclerotic lesion.[4] MRI provides the optimal modality for determining the extent of marrow replacement by conventional intramedullary chondrosarcoma. Distinguishing low-grade chondrosarcomas from benign central enchondromas can be difficult; location in the axial skeleton and size larger than 5 cm favors malignancy.

Up to 15% of chondrosarcomas arise from preexisting peripheral osteochondromas and, similar to their benign counterparts, harbor mutations in the exostosin *(EXT)* gene. The remaining 85% of chondrosarcomas arise in a central location, some in preexisting enchondromas. Chondrosarcomas are divided into three grades, with higher grade tumors characterized by greater cellularity and cellular atypia. In one series, 61% of patients had grade 1 tumors; only 4% of such patients developed metastases. In contrast, 36% of patients had grade 2, and 3% grade 3 tumors; among this combined group, 29% developed metastases.

TREATMENT

Unlike osteosarcoma and Ewing sarcoma, chondrosarcomas generally grow slowly, metastasize less commonly, and have an excellent prognosis after adequate surgical resection. Although chondrosarcomas are considered relatively radiation resistant, radiation therapy may provide palliation for patients with large or recurrent, unresectable central chondrosarcomas.

Ewing Sarcoma

Ewing sarcoma and primitive neuroectodermal tumors (PNET) are a family of small round cell sarcomas that represent 16% of primary bone sarcomas. The molecular hallmark of Ewing sarcoma is the translocation between the Ewing sarcoma protein (EWS) and an ETS (E26 transformation-specific or E-twenty-six) family transcription factor. In 85% of cases, the t(11;22)(q24;q12) translocation between *EWSR1* and *FLI1* is detected, although other fusion genes have also been described.[5,6]

As with osteosarcoma, the peak incidence occurs during the second decade of life, but unlike osteosarcoma, the incidence of Ewing sarcoma is unimodal, being distinctly unusual in older adults and in nonwhites. Ewing sarcoma tends to arise in the diaphyseal region of long bones, in the pelvis, or in the ribs. Ewing tumors are characterized radiologically by a permeative or "moth-eaten" appearance of the affected bone, with a multilayered "onion-skin" periosteal reaction. MRI studies frequently document a significant soft tissue mass associated with the bone lesion. Unlike other sarcomas of bone, Ewing sarcoma may present with symptoms of an inflammatory systemic illness, with intermittent fevers, anemia, leukocytosis, and an increased sedimentation rate.

Eighty-five percent of Ewing's family sarcomas contain a t(11;22)(9q24;q12) chromosomal translocation that juxtaposes the EWS gene on chromosome 22 with FLI1, an ETS family transcription factor. Another 15% contain a variant in which EWS is juxtaposed to ERG, another ETS family member on chromosome band 21q22. Because Ewing sarcoma resembles other small round cell tumors microscopically, reverse transcription–polymerase chain reaction and fluorescent in situ hybridization studies that document such translocations play a critical role in confirming the diagnosis. Ewing sarcoma/PNET cells characteristically express the CD99/MIC2 cell membrane glycoprotein.

TREATMENT

Patients with localized disease treated with multimodality therapy can achieve a 5-year event-free survival rate of 70%, but the 5-year overall survival rate of patients who present with overt bone or bone marrow metastatic disease at diagnosis is less than 20%.[5] The development of effective systemic chemotherapy regimens has substantially improved long-term control of Ewing sarcoma.[7] After completion of staging procedures, patients are treated with neoadjuvant chemotherapy. One highly active regimen alternates cycles of vincristine, doxorubicin, and cyclophosphamide with ifosfamide and etoposide. After 3 months of chemotherapy, the primary tumor is resected, radiated, or both, depending on the location and extent of the primary tumor. Chemotherapy is then resumed for a total of up to 1 year of treatment. Using such an approach, the mean 5-year event-free survival rate for patients who present with nonmetastatic disease is 73%.[8] Insulin-like growth factor-1 receptor antagonists have demonstrated clinical activity in recent clinical trials for chemotherapy-refractory disease.

METASTATIC TUMORS TO BONE

Tumors metastatic to bone are important causes of cancer-related morbidity. Effective prevention and treatment of skeleton-related metastases are important parts of clinical care for many cancer patients. The most common tumors that metastasize to bone are breast cancer in women and prostate cancer in men followed by cancers of the lung, kidney, gastrointestinal tract, and thyroid.

Bone metastases typically present with localized or referred pain and less commonly as a new bone fracture. Plain radiographs may demonstrate blastic or lytic lesions. Although bone metastases from prostate cancer are often blastic and multiple myeloma usually lytic, most other tumors have a mixed appearance. In patients with one documented bone metastasis or in patients with widely metastatic disease and bone pain, a radiologic survey can identify bone metastases that may ultimately place the patient at risk for a pathologic fracture. Radionuclide bone scans are useful to delineate the extent of bone metastases and in following response to therapy. However, a negative study result must be interpreted cautiously because tumors that are potentially purely lytic (particularly multiple myeloma) may not be detectable by bone scan. In such tumors, a plain skeletal survey or CT or MRI scan is preferable. Routine screening for bone metastases is not indicated for cancer patients with no symptoms or signs of bone involvement.

TREATMENT Rx

In the absence of fracture or impending fracture, painful bone metastases are treated with external-beam radiation therapy. In patients with numerous bone metastases, systemic chemotherapy or endocrine therapy can play an important palliative role. Pathologic fractures or imminent fractures are generally managed by operative internal fixation.

PREVENTION

In patients with breast cancer, prostate cancer, and multiple myeloma, bisphosphonate therapy increases the time to a first skeletal event.[8] Bisphosphonate therapy also appears to prolong survival in patients with metastatic breast cancer. The optimal schedule and duration of administration of bisphosphonates to maximize benefit and minimize potential complications remain to be established.

SARCOMAS AND OTHER CONNECTIVE TISSUE NEOPLASMS

DEFINITION

Sarcomas are tumors of mesenchymal origin that make up approximately 1% of human cancers.[9] They are a heterogeneous group of malignant neoplasms of connective tissues, including bone and soft tissue, comprising more than 50 histologic subtypes. Mesenchymal cells (derived from mesoderm), as well as neural crest cells (from ectoderm), give rise to connective tissues and provide critical functions such as support and nourishment to neural tissues. When growth, differentiation, or survival of these cells is aberrant, tumors arise, and this is the family of neoplasms to which sarcomas belong. Sarcomas include a vast array of tumor types related to muscle, stromal tissue, adipose tissue, blood and lymphatic vessels, nerves and nerve sheaths, cartilage, bone, and other fibrous tissues.

EPIDEMIOLOGY

Although very rare in adults, true sarcomas represent a disproportionately large number of cancers in the pediatric population (≈15% of pediatric cancers). The overall incidence of sarcomas of soft tissue and bone is approximately 15,000 cases per year in the United States. The prevalence of sarcomas significantly exceeds the incidence because sarcomas can be cured with expert multidisciplinary care. Hence, the initial evaluation and management of patients suspected of having sarcomas should be performed by an experienced team with relevant expertise and interdisciplinary capabilities (including expertise in sarcoma pathology, surgical specialization, and radiotherapeutic

experience and judgment, as well as access to the latest systemic therapeutic agents).[10]

Certain patients are at high risk of developing sarcomas, most notably individuals in families with Li-Fraumeni syndrome and those with neurofibromatosis (at risk for malignant peripheral nerve sheath tumors and gastrointestinal stromal tumors [GISTs]) or familial polyposis (at risk for intra-abdominal desmoid tumors). Other risk factors include exposure to radiation (including radiation therapy for other cancers, such as patients with prior irradiated breast cancer or survivors of retinoblastoma). Chemical carcinogens can also increase the risk of sarcoma development, such as the increased incidence of sarcomas in Vietnam veterans exposed to Agent Orange or the greatly increased risk of hepatic angiosarcomas associated with occupational exposure to polyvinyl chloride. However, the vast majority of sarcomas appear to be sporadic, with no evident inciting risk factors.

PATHOBIOLOGY

Sarcomas and other connective tissue neoplasms are a heterogeneous mixture of diseases with a wide range of clinical behaviors and outcomes. Some soft tissue neoplasms, such as localized tenosynovial giant cell tumors, can be cured by expert resection, but more advanced tumors of this type (referred to as pigmented villonodular tenosynovitis) often lead to debilitating amputations or even death because of metastatic disease. Expert pathological review is necessary for the diagnosis of specific sarcoma subtypes; unfortunately, interobserver variability can impair even the most elegant diagnostic categories, such as those promulgated by the World Health Organization. Increasingly, knowledge of the molecular pathways that drive sarcomas has provided more objective diagnostic tools, including novel immunohistochemical staining patterns and genetic markers. In general, sarcomas may exhibit differentiation patterns consistent with defined connective tissues (e.g., well-differentiated liposarcoma may appear as only slightly bizarre fat cells under the microscope), or they may be unclassifiable. In any case, as diagnostic tools have evolved, it has become possible to place poorly differentiated tumors more accurately into certain histopathologic categories based on the expression of lineage-related proteins (e.g., smooth muscle actin expression may help categorize tumors as leiomyosarcomas) or on the basis of genomic markers (e.g., overexpression of chromosome 12 material or the *MDM2* gene locus is most consistent with a dedifferentiated liposarcoma). Recent studies have revealed that mesenchymal-to-epithelial transition (MET) may be operative in sarcomas, and MET may be an important basic biological and clinical process in tumors of mesenchymal original in general.[11]

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Given the variety of sarcoma subtypes, it is understandable that the clinical course of these diseases can range from rapidly evolving and immediately life-threatening to indolent lesions that can take decades to evolve (e.g., atypical lipomatous tumors, also known as well-differentiated liposarcomas). Most patients with sarcomas present with a mass, often nontender, with a history of abnormal growth over time. For extremity tumors of soft tissues, it is important to note that many benign tumors (e.g., lipomas) cannot be easily distinguished from more worrisome neoplasms or even from frankly malignant sarcomas. Therefore, it is important to include sarcoma in the differential diagnosis of any mass.

The initial biopsy or surgical approach to a sarcomatous lesion is often the most important, and a poorly oriented biopsy or a suboptimal surgical procedure can make the difference between cure with full limb function and disease recurrence with the need for amputation or mutilating surgical re-resection. The National Comprehensive Cancer Network has developed expert consensus guidelines for clinical practice that emphasize the importance of expert management from the moment a suspected sarcoma presents. The initial diagnosis includes appropriate imaging studies of relevant anatomic areas, including plain radiographs, CT, or MRI to define the anatomic area of the mass and surrounding tissue, as well as systemic staging because sarcomas can spread in well-defined patterns to distal sites such as the lung or liver. The decision to proceed to diagnostic biopsy, with optimal orientation, is an important one, and for certain lesions, forgoing incisional biopsy and proceeding directly to expert surgical excision may be justified. The most important consideration is to make the correct diagnosis, and there must be sufficient amounts of properly prepared and expertly oriented tissues for optimal diagnostic analysis. In certain tumors with pathognomonic molecular markers (e.g., the translocation between chromosomes X and 18 that characterizes synovial sarcoma or the balanced translocation between chromosomes 12 and 16 that defines myxoid and round cell liposarcoma), molecular analyses such as fluorescence in situ hybridization may help make the diagnosis. New molecular subtypes of sarcoma enter the pathology literature frequently; these new diagnostic categories may lead to the use of novel, molecularly targeted therapies. Nowhere has this been more evident than in the rapid evolution of effective therapy against the major pathobiologic cause of GISTs (see below).

The diagnosis of a soft tissue sarcoma is made by evaluating biopsy material in a clinical context, which includes understanding the tumor's anatomic location and imaging characteristics. Such contextual diagnostics are critical to understanding whether a lesion may represent a primary sarcoma or whether it may be the first presentation of metastases from an occult primary tumor located elsewhere. Given the broad spectrum of sarcomas, the diagnostic considerations are quite far reaching, especially because many benign pathophysiologic conditions can mimic sarcomas.

TREATMENT

Rx

The most important element of treatment is expert multidisciplinary care. The range of options is too broad to categorize simply, and the specific details of each patient's anatomy, comorbidities, functional status, and personal preferences must be taken into account when defining treatment options and management plans. Therefore, the care of virtually all sarcoma patients should be managed by an expert multidisciplinary team with expertise in advanced surgical or orthopedic oncology techniques, radiation therapy, reconstructive surgery, physical therapy and rehabilitation medicine, systemic therapies such as conventional cytotoxic chemotherapy, hormonal therapy, and modern molecularly targeted therapy with agents such as kinase inhibitors, and psychosocial support and specialized nursing care. Therefore, appropriate referral to obtain expert opinion and to define the diagnostic and treatment options for patients with suspected sarcomas is recommended.

For most localized masses for which sarcoma is in the differential diagnosis, the first step is to obtain the correct diagnosis in a manner that does not compromise patient outcome or function. The need for biopsy must be considered first because some small, localized sarcomas are best approached through definitive surgical excision following careful staging and expert review of imaging studies. For suspected sarcomas in deeper locations, such as within muscle compartments, or for large visceral lesions, biopsy may be necessary to ascertain that the process is in fact a sarcoma, as well as to fully characterize the histopathologic subtype. This may make the difference between initial management with chemotherapy, as might be appropriate for a highly chemosensitive disease, versus surgery, which might be appropriate for a less chemosensitive disease such as dedifferentiated liposarcoma. Expert opinion varies regarding the utility and timing of adjuncts to surgical resection, such as radiation therapy or systemic cytotoxic chemotherapy.

In general, expert surgical resection is the first-line of therapy for localized sarcomas. Preoperative systemic chemotherapy may be appropriate for some rhabdomyosarcomas. Many expert teams favor preoperative radiation therapy for certain sarcomas; irradiation of a large primary tumor can deliver smaller doses to surrounding normal tissues preoperatively compared with postoperatively, but this is a matter of personal preference. A randomized trial of preoperative versus postoperative radiation therapy for large sarcomas of the extremity was performed in Canada, and outcomes were similar, with subtle differences: patients who received preoperative radiation had a higher incidence of serious postoperative wound complications, but the long-term functional outcomes were slightly more favorable.

Many sarcoma centers disagree about the relative value of cytotoxic chemotherapy, although there is no doubt that chemotherapy has greatly improved disease control rates and cure rates of certain subtypes of aggressive sarcomas such as rhabdomyosarcoma. For other forms of sarcoma originating in bone (e.g., chondrosarcoma) or soft tissue (e.g., leiomyosarcoma, synovial sarcoma), there is no strong evidence that systemic chemotherapy increases cure rates or long-term clinical outcomes, although there may be some improvement in local disease control and recurrence-free survival. This has led to discordant expert opinion: many experts believe that the risks and toxicities of aggressive chemotherapy justify the benefit of longer disease-free survival, but others believe that such toxicities are not reasonable without a major improvement in overall survival. Limited series of postoperative adjuvant therapy are often contradictory owing to the relatively small patient groups under study, as well as divergent patient selection factors, such as the inclusion of those with a lower risk of recurrence or death from metastatic disease. Inclusion of a sizable percentage of sarcoma patients with low-risk disease no doubt dilutes the results of even a reasonably effective therapy and runs the risk of making the study result negative.

It is also critical to recognize that treatment may differ radically depending on the histologic diagnosis. The best example of this is GIST (also see Chapters 192 and 193), a form of sarcoma that, in more than 95% of cases, is driven by aberrant tyrosine kinase signaling.[12] Routine systemic chemotherapy is completely ineffectual against this disease; however, tyrosine kinase inhibitor

therapy (e.g., with imatinib mesylate or sunitinib malate) produces dramatic tumor regressions and tumor control in more than 85% of patients. Fortunately, pathologists are increasingly able to recognize this disease histopathologically, and immunohistochemistry to detect CD117 (the Kit receptor tyrosine kinase) and DOG1 (a membrane antigen that is reasonably specific for GIST), as well as tumor genotyping via molecular genetics have significantly increased the accuracy of GIST diagnosis in the past decade. Imatinib has been approved by the Food and Drug Administration to decrease the risk of disease recurrence following resection of GISTs with significant potential for relapse. However, patients with a low risk of relapse have not derived substantive benefits from adjuvant imatinib because they have a very good chance of being cured by surgery alone. GIST is an excellent example of a sarcoma in which critical therapy decisions must be made in the context of the proper molecular diagnosis.[13]

For soft tissue sarcomas other than GIST, radiation therapy can play a meaningful role in preventing disease recurrence, especially for lesions that arise in the extremities. Radiation therapy can also provide significant palliation of unresectable disease, and it can be surprisingly effective in certain tumors such as desmoid tumor. However, radiation-associated sarcomas are increasingly common after therapeutic radiation therapy (e.g., in patients cured of breast cancer with radiation therapy), and an increasing incidence of poorly differentiated sarcomas or vascular sarcomas has been reported in patients after irradiation for other diseases.

As noted earlier, traditional cytotoxic chemotherapy is effective at increasing disease control and cure rates for certain sarcomas (especially those most prevalent in pediatric patients, such as rhabdomyosarcoma). It is less effective in improving long-term cure rates for patients with most other forms of soft tissue and bone sarcomas, such as liposarcoma, leiomyosarcoma, synovial sarcoma, and other subtypes. Nonetheless, appropriate use of chemotherapy can lead to objective responses in certain patients and can palliate those with metastatic disease with disease control and prolongation of progression-free survival. Targeted therapies hold potential promise for some sarcomas. For example, pazopanib, a multitargeted tyrosine kinase inhibitor at 800 mg once daily, improves the overall survival period from 10.7 to 12.5 months in patients who have metastatic nonadipocytic soft tissue sarcoma and who have had previous chemotherapy. Furthermore, a recent study demonstrated the first effective treatment for an uncommon, vasculogenic sarcoma of young adults, alveolar soft part sarcoma, with a multikinase inhibitor (cediranib) that primarily targets the vascular endothelial growth factor.

Sarcoma experts often disagree about the relative value and toxicity of combination chemotherapy as opposed to sequential single-agent chemotherapy. In patients with very aggressive and highly symptomatic sarcomas, clinicians may choose a combination chemotherapy regimen, even if there is a greater risk of toxicity, to ensure some measure of rapid disease control or even a greater chance of disease regression. In contrast, in patients with asymptomatic metastatic disease (e.g., a sarcoma patient with indolent pulmonary metastases and no symptoms), the optimal choice might be single-agent chemotherapy to avoid toxicity and to maximize the choice of subsequent chemotherapeutic agents after the benefit of the first agent is fully realized. There are no definitive data from properly powered randomized trials demonstrating that any chemotherapy for metastatic sarcoma of soft tissue (other than GIST) improves overall survival. All experts agree that owing to the complexity of the trials, the diseases, and the clinical settings, one must allow for variations in interpretations for individual patients. This explains the greatly discordant practice patterns observed across the country based on referrals, provider experience, and patient characteristics and preferences.

PROGNOSIS

Discussing prognosis for sarcomas overall is difficult because they represent such a variety of diseases with widely divergent natural histories. It is estimated that approximately 50% of patients with localized sarcomas can be cured, and the risk of recurrence is related to variables such as tumor grade (low-grade tumors have a lower risk of recurrence or metastasis compared with intermediate- or high-grade tumors), tumor size, and tumor location or depth. For patients with recurrent or metastatic sarcoma, outcome depends on many factors, including the time from initial diagnosis to the first appearance of metastatic disease; a longer disease-free interval is associated with longer survival, probably indicating a slower rate of tumor proliferation. Another factor that may determine outcome is the number of metastatic lesions; it is possible that oligoclonal metastases, with few lesions, can be surgically resected, which itself may be associated with improved survival.

Advances in targeted therapies have also affected survival, as documented by the dramatic improvements in survival and disease control for GIST and other kinase-driven sarcomas, such as dermatofibrosarcoma protuberans, giant cell tumor of bone, perivascular epithelioid cell–oma (PEComa), tenosynovial giant cell tumor, and alveolar soft part sarcoma. The natural history of these diseases will almost certainly be changed by a more mechanistic understanding of the underlying neoplasia-promoting signals that cause the transformation and maintenance of the sarcoma. Expert multidisciplinary management of sarcomas is critical to improving outcomes, and ongoing translational and therapeutic research will provide dividends far beyond the relatively low incidence and prevalence of these mesenchymal cell neoplastic disorders.

Grade A References

A1. Bernthal NM, Federman N, Eilber FR, et al. Long-term results (>25 years) of a randomized, prospective clinical trial evaluating chemotherapy in patients with high-grade, operable osteosarcoma. *Cancer*. 2012;118:5888-5893.

A2. Womer RB, West DC, Krailo MD, et al. Randomized controlled trial of interval-compressed chemotherapy for the treatment of localized Ewing sarcoma: a report from the Children's Oncology Group. *J Clin Oncol*. 2012;30:4148-4154.

A3. Barrett-Lee P, Casbard A, Abraham J, et al. Oral ibandronic acid versus intravenous zoledronic acid in treatment of bone metastases from breast cancer: a randomized, open-label, non-inferiority phase 3 trial. *Lancet Oncol*. 2014;15:114-122.

A4. Gastrointestinal Stromal Tumor Meta-Analysis Group (MetaGIST). Comparison of two doses of imatinib for the treatment of unresectable or metastatic gastrointestinal stromal tumors: a meta-analysis of 1640 patients. *J Clin Oncol*. 2010;28:1247-1253.

A5. DeMatteo RP, Ballman KV, Antonescu CR, et al. Placebo-controlled randomized trial of adjuvant imatinib mesylate following the resection of localized, primary gastrointestinal stromal tumor (GIST). *Lancet*. 2009;373:1097-1104.

A6. van der Graaf WT, Blay JY, Chawla SP, et al. Pazopanib for metastatic soft-tissue sarcoma (PALETTE): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet*. 2012;379:1879-1886.

A7. Sharma S, Takyar S, Manson SC, et al. Efficacy and safety of pharmacological interventions in second- or later-line treatment of patients with advanced soft tissue sarcoma: a systematic review. *BMC Cancer*. 2013;13:385.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Fox MG, Trotta BM. Osteosarcoma: review of the various types with emphasis on recent advancements in imaging. *Semin Musculoskelet Radiol*. 2013;17:123-136.
2. Botter SM, Neri D, Fuchs B. Recent advances in osteosarcoma. *Curr Opin Pharmacol*. 2014;16:15-23.
3. Luetke A, Meyers PA, Lewis I, et al. Osteosarcoma treatment—where do we stand? A state of the art review. *Cancer Treat Rev*. 2014;40:523-532.
4. Logie CI, Walker EA, Forsberg MD, et al. Chondrosarcoma: a diagnostic imager's guide to decision making and patient management. *Semin Musculoskelet Radiol*. 2013;17:101-115.
5. Arnaldez FI, Helman LJ. New strategies in Ewing's sarcoma: lost in translation? *Clin Cancer Res*. 2014;20:3050-3056.
6. Paronetto MP. Ewing's sarcoma protein: a key player in human cancer. *In J Cell Biol*. 2013;2013:642853.
7. Cote GM, Choy E. Update in treatment and targets in Ewing's sarcoma. *Hematol Oncol Clin North Am*. 2013;27:1007-1019.
8. Ganjoo KN, Patel S. The treatment outcome for adult patients with Ewing's sarcoma. *Curr Oncol Rep*. 2013;15:372-377.
9. Forscher C, Mita M, Figlin R. Targeted therapy for sarcomas. *Biologics*. 2014;8:91-105.
10. Benson C, Judson I. Role of expert centres in the management of sarcomas—a UK perspective. *Eur J Cancer*. 2014;50:1951-1956.
11. Yang J, Du X, Wang G, et al. Mesenchymal to epithelial transition in sarcomas. *Eur J Cancer*. 2014;50:593-601.
12. Joensuu H, Hohenberger P, Corless CL. Gastrointestinal stromal tumour. *Lancet*. 2013;382:973-983.
13. Serrano C, George S. Recent advances in the treatment of gastrointestinal stromal tumors. *Ther Adv Med Oncol*. 2014;6:115-127.

REVIEW QUESTIONS

1. A 72-year-old woman with metastatic breast cancer to multiple sites who has been treated with an aromatase inhibitor complains of severe left hip pain. Plain radiographs of the pelvis and left femur demonstrate greater than 90% lytic destruction of the femoral neck. Her performance status except for the pain is excellent. The most appropriate therapy, in addition to adequate pain control, to consider at this time is

A. initiation of combination chemotherapy with doxorubicin and paclitaxel.
B. intravenous bisphosphonate therapy.
C. immediate palliative radiation therapy to the left hip.
D. operative procedure to stabilize the hip.
E. bed rest followed by reassessment in 6 weeks.

Answer: D Patients with breast cancer at risk of immediate hip fracture should be referred for an orthopedic procedure to stabilize the hip; radiation therapy will follow the surgery, as will consideration of a change in systemic treatment.

2. Which of the following malignancies is associated with germline abnormalities of the *TP53* gene?

A. Chondrosarcoma
B. Ewing's sarcoma
C. Angiosarcoma
D. Osteogenic sarcoma
E. Gastrointestinal stromal tumor (GIST)

Answer: D Li-Fraumeni families are characterized by hereditary osteogenic sarcomas that are associated with germline mutations in *TP53*.

3. Gastrointestinal stromal tumors (GISTs) are highly responsive to tyrosine kinase inhibitors that target which of the following proteins?

A. c-Kit
B. Epidermal growth factor receptor (EGFR)
C. BCL-2
D. Focal adhesion kinase
E. PDGFRA (platelet-derived growth factor receptor-alpha)

Answer: A Imatinib, which targets both the c-Kit and bcr-abl kinase, is highly active in the treatment of GISTs (as well as chronic myelogenous leukemia).

4. Which of the following statements concerning soft tissue sarcomas is **incorrect**?

A. Soft tissue sarcomas are a relatively homogeneous group of malignancies at the molecular level because they all originate from primordial mesenchyme.
B. Classification of soft tissue sarcoma subtypes is difficult at the level of light microscopy.
C. Traditional cytotoxic chemotherapy is only useful for specific types of soft tissue sarcoma.
D. Radiation to the chest wall as part of a combined modality approach to patients with breast cancer can lead to secondary sarcomas in the radiated field.
E. Many benign tumors (e.g., lipomas) cannot be easily distinguished clinically from frankly malignant sarcomas.

Answer: A Soft tissue sarcomas are among the most heterogeneous groups of malignancies currently known.

5. Which of the following sarcomas has the best prognosis after definitive surgery?

A. Ewing's sarcoma
B. Liposarcoma
C. Chondrosarcoma
D. Leiomyosarcoma
E. Osteosarcoma

Answer: C Definitive surgery is curative in many patients with chondrosarcoma; the recurrence rate after surgery with each of the other diseases is considerably higher.

203

MELANOMA AND NONMELANOMA SKIN CANCERS

LYNN M. SCHUCHTER

MELANOMA

EPIDEMIOLOGY

Current estimates are that one of 37 men and one of 56 women will be diagnosed with melanoma during their lifetimes. Each year in the United States, approximately 76,690 new cases of invasive melanoma are detected, and 9480 patients die of melanoma. The explanation for the rising incidence is thought to be increasing sun exposure, especially early in life. Melanoma is the leading cause of death from cutaneous malignant disease, and it accounts for 1% to 2% of all cancer deaths in the United States. Melanoma affects all age groups; the median age at diagnosis is 50 years. Melanoma is largely a disease of whites, with a very low incidence in African Americans, Asians, and Hispanics.

PATHOBIOLOGY

Exposure to sunlight, especially ultraviolet (UV) radiation (Chapter 20), has been strongly implicated as a causative factor in the development of melanoma. Melanomas originate from melanocytes, which are located predominantly in the basal cell layer of the epidermis and use the enzyme tyrosinase to synthesize melanin pigment, which serves to protect against UV damage (Chapter 435). Worldwide, the incidence of melanoma in whites generally correlates inversely with latitude; that is, rates are generally higher closer to the equator and become progressively lower near the poles.

Risk Factors

Risk factors for melanoma include family history of melanoma, prior melanoma or nonmelanoma skin cancer, inherited genetic susceptibility, and sun

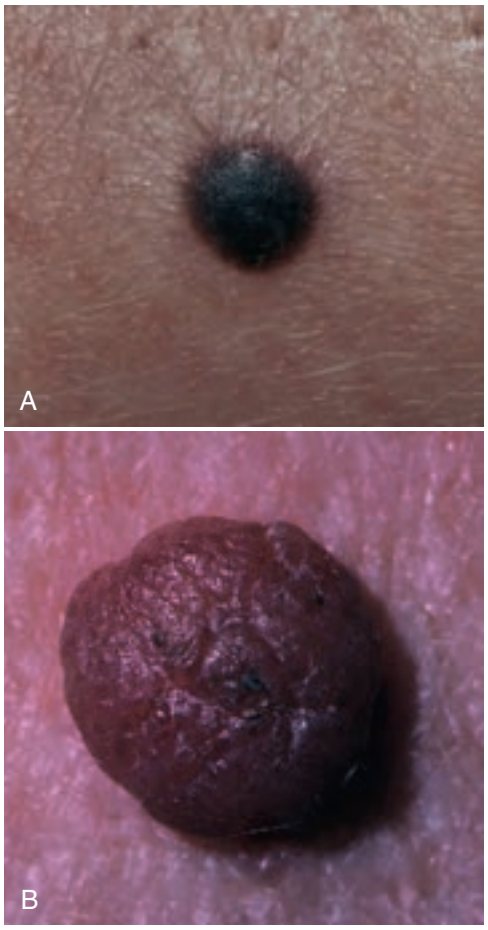

FIGURE 203-1. Nevi. A, Common benign nevus. B, Dermal nevus.

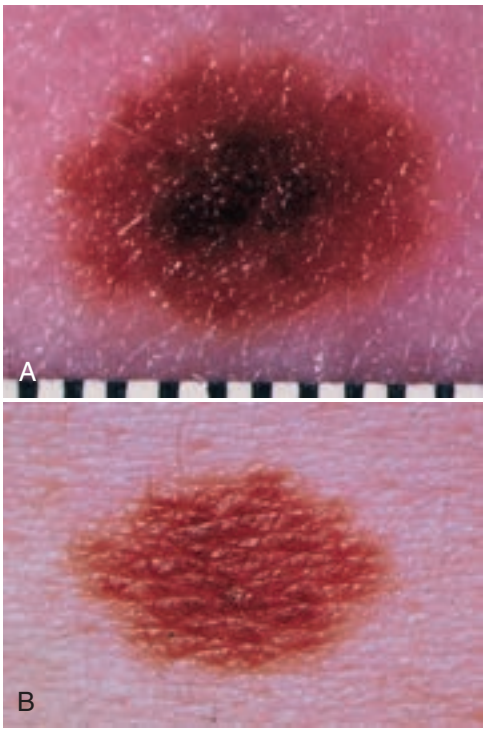

FIGURE 203-2. Dysplastic nevi. A and B, Examples of dysplastic nevi.

exposure. Artificial exposure to UV radiation by indoor tanning is likewise a risk factor for melanoma. Individuals with fair complexions, blond or red hair, blue eye color, and freckles, who have a tendency to burn rather than tan, have higher rates of melanoma. The pattern of sun exposure may also be important; intermittent intense exposure, rather than long-term exposure, may carry a higher risk of melanoma.

Individuals with an increased number of typical or benign moles, atypical moles, or dysplastic nevi (Figs. 203-1 and 203-2) also have an increased risk for melanoma. Atypical moles or dysplastic nevi are important precursor lesions of melanoma and serve as markers for increasing risk. For example, individuals with dysplastic nevi have a 6% lifetime chance of developing melanoma, and this risk increases to as high as 80% in individuals who have dysplastic nevi and a strong family history of melanoma.

Genetics

Approximately 10% of patients with melanoma have a family history of melanoma.[1] Several chromosomal loci determine susceptibility to melanoma, the most important of which is *p16/CDKN2A*, a gene located on chromosome 9p21. This gene is a member of a class of molecules that play a central role in cell cycle regulation. Of the members of melanoma-prone families, 25% to 40% have mutations in this gene. The risk of developing cutaneous melanoma in an individual who is a *CDKN2A* carrier is between 30% and 90% by age 80 years and varies by geographic location. Testing for mutations in the *p16/CDKN2A* locus is commercially available, but its clinical utility is unclear at this time. Genetic variability in melancortin-1 receptor (MC1R) plays a key role in pigmentation of skin and hair and more recently has been implicated in melanoma predisposition.

Somatic mutations in primary and metastatic melanoma primarily involve the mitogen activated protein kinase pathway. Activating mutations in *B-RAF* can be found in approximately 50% of melanomas, and 20% of melanomas are associated with a mutation in *N-RAS*. Recent studies have found that melanoma on mucous membranes, acral skin (soles, palms), and skin with chronic sun damage (i.e., lentigo maligna melanoma) have frequent mutations in *c-kit*. Thus, distinct patterns of genetic alterations are found in primary melanomas based on anatomic location and extent of sun exposure. Uveal melanoma is associated with mutations in *GNAQ/GNA11*. Monosomy of chromosome 3 and somatic mutations in the gene encoding BRCA-1 associated protein (BAP) on chromosome 3 have been associated with worse outcome and the development of metastatic melanoma. The discovery of somatic mutations in melanoma and associated aberrant signal transduction pathways has provided leads for the development of molecularly targeted therapy for patients with advanced melanoma.

CLINICAL MANIFESTATIONS

Early detection and recognition of melanoma are key to improving survival. The signs of early melanoma are based on the clinical appearance of the pigmented lesion and a change in the shape, color, or surface of an existing mole. Most patients report a preexisting mole at the site of the melanoma. Itching, burning, or pain in a pigmented lesion should increase suspicion, although melanomas often are not associated with local discomfort. Bleeding and ulceration are signs of a more advanced melanoma. Most melanomas are varying shades of brown, but they may be black, blue, or pink. The ABCDEs for the recognition of melanoma are asymmetry, border irregularity, color variation, diameter greater than 6 mm, and evolution or a change in a skin lesion. The "ugly duckling" sign is recognizing a pigmented lesion that looks different from other skin lesions and is therefore suspicious.

Cutaneous melanoma has been divided into four subtypes. Superficial spreading melanoma, which accounts for 70% of all melanomas, can be located on any anatomic site (Fig. 203-3). Lentigo maligna melanoma, which represents 4 to 10% of all melanomas, tends to occur more commonly in chronically sun-exposed skin in older patients, frequently on the head and neck (Fig. 203-4); clinically, it appears as a macular (flat) lesion, arising in a lentigo maligna. Nodular melanoma (Fig. 203-5) accounts for 15% to 30% of melanomas and manifests as a rapidly enlarging elevated or polypoid lesion, often blue or black. The ABCDE rule does not always apply as well to nodular melanomas. Acral lentiginous melanoma manifests as a darkly pigmented, flat to nodular lesion on the palm, on the sole, or subungually; sunlight is not thought to play a causative role in this form of melanoma. Histologic subtype does not directly correlate with clinical behavior. However, recent data suggest that histologic subtype may correlate with specific genetic abnormalities.

Ocular melanomas arise from the pigmented layer of the eye. Uveal melanoma is the most common intraocular malignancy of adults. Melanomas

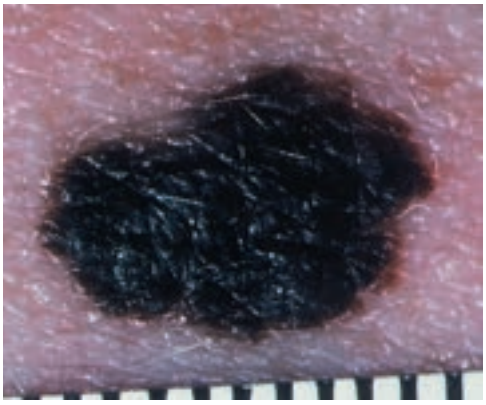

FIGURE 203-3. Superficial spreading melanoma.

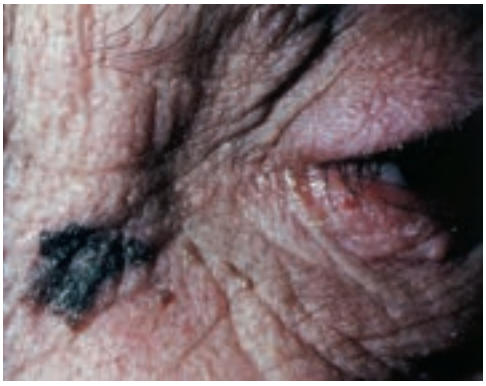

FIGURE 203-4. Lentigo maligna melanoma.

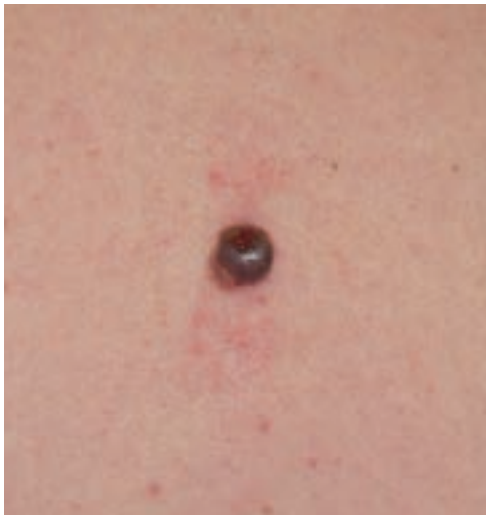

FIGURE 203-5. Nodular melanoma.

can also arise from noncutaneous sites, including mucosal epithelium in the gastrointestinal tract, anorectal area, genitourinary tract, and nasal and nasopharyngeal mucosa. Melanomas of the vulva and vagina are relatively rare. In general, mucosal melanomas are diagnosed at a more advanced stage of disease. The mainstay of treatment is surgical.

DIAGNOSIS

Any skin lesion suggestive of melanoma should be sampled using biopsy with complete excision, including a 1- to 2-mm margin of normal skin and some subcutaneous fat. An incisional biopsy may be necessary for lesions too large for complete excision. The role of sentinel lymph node biopsy (SLNB) is explained in detail in the Treatment section. Shallow shave biopsies, curettage, cryosurgery, laser, and electrodessication are contraindicated in lesions suggestive of melanoma. Other lesions that can be confused with melanoma include blue nevi, pigmented basal cell carcinoma (BCC), seborrheic keratosis, and hemangiomas (Table 203-1).

TABLE 203-1 CLINICAL FEATURES OF COMMON NEVI, DYSPLASTIC NEVI, AND MELANOMAS

DISEASE	CHARACTERISTICS
Common acquired nevi (moles)	These tend to be small, flat, and round; the border is regular, smooth, and well defined; the color is homogeneous, usually no more than two shades of brown; any site is affected; lesions are usually <6 mm.
Dysplastic nevi (atypical moles)	These occur predominantly on the trunk; they tend to be large, usually >5 mm, with a flat component; the border is characteristically fuzzy and ill defined. The shape can be round, oval, or misshapen. The color is usually brown but can be mottled with dark brown, pink, and tan. Some individuals have only one to five moles; others have more than 100.
Melanoma	The border is more irregular; lesions tend to be larger, often >6 mm; substantial heterogeneity of color is noted, ranging from tan-brown, dark brown, black, pink, red, gray, blue, or white.

Prognostic Factors

As with most malignancies, the outcome of melanoma depends on the stage and extent of disease at presentation. For localized melanoma, the most important prognostic factor is involvement of regional lymph nodes. The most important prognostic factor related to the primary tumor is the depth of invasion (Breslow's thickness) of the melanoma, which is measured in millimeters from the top of the epidermis to the underlying dermis. Increasing thickness is associated with an increased risk for recurrence, regional lymph node involvement, and death from melanoma. Whereas patients whose melanomas are smaller than 1 mm thick have about an 80% to 90% 10-year survival rate, patients whose melanomas are greater than 4 mm thick have only a 40% to 50% 10-year survival rate. Other poor prognostic factors related to the primary melanoma include the presence of ulceration, an increasing level of invasion (Clark's level), a high mitotic rate, and the presence of microscopic satellites. Regional lymph node involvement (stage III) has a major impact on survival, with 5-year survival rates ranging from 20% to 70%, depending on the number of involved lymph nodes. Melanomas that arise on the extremity tend to have a better prognosis, and women tend to do better than men.

Staging System for Melanoma

Staging and prognosis for melanoma are based on the TNM system, in which T refers to tumor, N to nodes, and M to metastasis, which was updated in 2009 (E-Table 203-1). Stages I and II indicate clinically localized primary melanoma, stage III indicates regional involvement (lymph nodes or in-transit metastases), and stage IV is metastatic disease beyond the regional lymph nodes (i.e., lung, liver, brain).

Patient Evaluation

The initial evaluation of a patient with melanoma includes a personal history, a family history, a total skin examination, and palpation of regional (draining) lymph nodes. The focus is to identify risk factors, signs or symptoms of metastases, dysplastic nevi, and additional melanomas. A chest radiograph and liver enzyme tests may be performed at the discretion of the physician. Most patients who present with melanoma do not have distant metastatic disease at presentation; therefore, extensive evaluations with computed tomography (CT) to search for distant metastases have an extremely low yield and are not indicated in asymptomatic patients. More extensive staging evaluation with CT or positron emission tomography (PET) can be considered in patients with high-risk disease (primary melanoma >4 mm thick or node-positive disease), in whom the risk of distant metastatic disease is higher.

E-TABLE 203-1 TUMOR, NODE, METASTASIS STAGING CATEGORIES FOR CUTANEOUS MELANOMA

CLASSIFICATION	THICKNESS (MM)	ULCERATION STATUS OR MITOSES
T		
Tis	NA	NA
T1	≤1.00	a: Without ulceration and mitosis <1/mm^2 b: With ulceration or mitoses ≥1/mm^2
T2	1.01-2.00	a: Without ulceration b: With ulceration
T3	2.01-4.00	a: Without ulceration b: With ulceration
T4	>4.00	a: Without ulceration b: With ulceration
	NO. METASTATIC NODES	**NODAL METASTATIC BURDEN**
N		
N0	0	NA
N1	1	a: Micrometastasis* b: Macrometastasis†
N2	2-3	a: Micrometastasis* b: Macrometastasis† c: In transit metastases or satellites without metastatic nodes
	SITE	**SERUM LDH**
M		
M0	No distant metastases	NA
M1a	Distant skin, subcutaneous, or nodal metastases	Normal
M1b	Lung metastases	Normal
M1c	All other visceral metastases	Normal
	Any distant metastasis	Elevated

LDH = lactate dehydrogenase; NA = not applicable.
*Micrometastases are diagnosed after sentinel lymph node biopsy.
†Macrometastases are defined as clinically detectable nodal metastases confirmed pathologically.
From Balch CM, Gershenwald JE, Soong S, et al. Final version of 2009 AJCC Melanoma Staging and Classification. *J Clin Oncol*. 2009;27:6199-6206.

TREATMENT

Primary Melanoma

After melanoma is diagnosed, the standard treatment is surgical excision. Several prospective randomized trials have been conducted to define the optimal surgery for primary melanoma. The extent of the surgery depends on the thickness of the primary melanoma. Large surgical excisions are no longer required, and most wide excisions can be performed with primary closure. For melanoma in situ, excision with a 0.5-cm border of clinically normal skin is sufficient. For melanomas less than 1 mm thick, a 1-cm margin is recommended. If the thickness is between 1 and 4 mm, a 1- to 2-cm margin is recommended. For melanomas thicker than 2 mm, 2-cm resection margin also is sufficient and safe. In cosmetically sensitive areas (face) or anatomically difficult areas (ear, hands), it may be difficult to achieve the desired margin, but at least a 1-cm margin should be obtained whenever possible.

Management of Regional Lymph Nodes

Clinically Normal Regional Lymph Nodes

In approximately 10% to 20% of patients who do not have clinically apparent lymph node involvement, lymph nodes contain occult micrometastases. The risk for occult lymph node involvement rises with increasing tumor thickness. Results from randomized trials fail to show a survival benefit from elective or prophylactic lymph node dissections in patients with clinically negative lymph nodes.

Sentinel lymph node biopsy is a technique that accurately evaluates whether microscopic melanoma cells involve regional lymph nodes. The technique relies on the concept that specific regions of the skin drain specifically to an initial lymph node within the regional nodal basin through an organized pathway of afferent lymphatic channels. This technique is performed by injecting the primary melanoma site with blue dye (isosulfan blue), radiolabeled colloid, or both. When both modalities are used in combination, a sentinel node can be identified in 98% of patients; biopsy of the node accurately determines whether melanoma cells have metastasized to that specific lymph node basin. The sentinel node technique also promotes a more comprehensive histologic examination of lymph nodes because limited amounts of pathologic material are submitted.

Sentinel lymph node biopsy allows earlier identification of metastases and is an important staging tool. The likelihood of detecting melanoma in SLNB increases with thickness of the primary lesion. SLNB is recommended to patients with melanoma 1 mm thick or thicker. The use of this technique for patients with thinner melanomas, that is, less than 1 mm thick can be considered if the primary has high risk features, is controversial. The SLNB is generally performed at the same time as the wide excision of the primary tumor. If the SLNB result is negative for melanoma, no further lymph node surgery is required. If melanoma is detected by the SLNB, complete lymph node dissection remains the standard of care. A recently reported randomized trial showed that the use of SLNB in patients with intermediate-thickness of thick primary melanomas provides accurate and important staging information; enhances regional disease control; and among patients with nodal metastases, improved melanoma-specific survival.

Clinically Apparent Regional Lymph Nodes

Surgical (therapeutic) lymphadenectomy is the preferred treatment of cytologically (fine-needle aspiration) positive or pathologically proven regional lymph node involvement with melanoma. The goal is to provide long-term, disease-free survival and reduce local morbidity of enlarged lymph nodes.

Adjuvant Therapy

Postsurgical adjuvant therapy can be considered for patients at high risk for recurrence (melanomas ≥4 mm thick or node-positive disease). These patients have at least a 25% to 75% chance of dying of melanoma. Adjuvant treatment options include interferon-α (IFN-α), enrollment in a clinical trial, or observation. High-dose IFN-α is the only U.S. Food and Drug Administration (FDA)—approved adjuvant therapy for patients with melanoma. Randomized clinical trials have shown that therapy with IFN-α can prolong disease-free survival but has not consistently demonstrated improvement in overall survival. The treatment is given for 1 year and is associated with considerable side effects, which require close monitoring. Intermediate and low doses of IFN-α as well as pegylated IFN-α have also been evaluated in a series of clinical trials. Pegylated IFN-α was recently approved by the FDA for treatment of patients with stage III melanoma. Numerous vaccine studies are ongoing but are considered experimental at present. Ipilimumab, a monoclonal antibody targeting CTLA-4, has been shown to prolong survival in patients with metastatic melanoma. Currently, ipilimumab is being studied in multiple phase III trials as an adjuvant therapy in patients with high risk melanoma.

Treatment and Course of Advanced Melanoma (Stage IV)

Melanoma can metastasize to virtually any organ, especially the lung, skin, liver, and brain. Until recently, the overall survival period for patients with metastatic melanoma has ranged from 5 to 11 months, with a median survival period of 9 months. However, new approaches with immunotherapy and molecularly targeted therapy based on somatic mutation profile have led to several recent FDA approvals of new agents that have redefined the standard of care for patients with metastatic melanoma.[2]

New immunotherapy approaches include ipilimumab, which blocks cytotoxic T-lymphocyte antigen 4 (CTLA-4).[3] Immunotherapy with ipilimumab has been shown to be better than dacarbazine alone (3-year survival rate, 21% vs. 12%) for previously untreated metastatic melanoma. Ipilimumab, 3 mg/kg intravenously every 3 weeks for a total of four doses, is FDA approved for patients with unresectable stage III or stage IV melanoma. Treatment with ipilimumab results in immune-mediated adverse reactions, including enterocolitis, hepatitis, dermatitis, and endocrinopathies such as hypopituitarism and hypothyroidism. Dose interruption of ipilimumab and corticosteroids are the main stays of treatment for this side effect mediated by T-cell activation and proliferation.

A second investigational approach to inhibit regulation of T-cell activation is to block the PD-1/PD-L1 (programmed cell death) pathway. Monoclonal antibodies targeting both PD-1 and PD-L1 are now in clinical trials. The first-in-class anti PD-1 antibody nivolumab, a fully human IgG4 monoclonal PD-1 antibody, has shown impressive clinical activity in patients with advanced melanoma. Another PD-1 inhibitor recently FDA approved is lambrolizumab, which also has significant clinical activity (response rate >40%) with fewer immune adverse events than traditionally seen with ipilimumab. Combining anti-CTLA-4 with anti-PD-1 immunotherapy has been encouraging in early non-randomized studies.[4]

The discovery of somatic genetic mutations in melanoma has provided leads for the development of molecularly targeted therapies. The MAP (mitogen-activated protein) kinase pathway, which is activated in most melanomas because of mutations in *BRAF*, *NRAS*, and *c-kit*, has been the focus of most clinical investigations of signal transduction (kinase) inhibitors. In patients with metastatic melanoma whose tumors harbor the *V600E BRAF* mutation, vemurafenib (960 mg orally twice daily), trametinib (2 mg orally twice daily), and dabrafenib (150 mg orally twice daily) each can increase 6-month survival rate significantly compared with dacarbazine. Squamous cell carcinoma (SCC) and keratoacanthomas develop in approximately 20% of patients with melanoma treated with vemurafenib because of paradoxical activation of MAPK signaling. Nevertheless, vemurafenib provides a median overall survival time of about 16 months in treated patients with *BRAF V600*–mutant metastatic melanoma. Recently, combination targeted therapy with dabrafenib (BRAF inhibitor) and trametinib (MEK inhibitor)[5] has been FDA approved based on results showing a 70% response rate and an acceptable safety profile. Combination therapy with several different regimens has now been shown to provide better progression free survival, better overall survival, and no increase in side effects in patients with metastatic melanoma. Preliminary results also show that imatinib can induce regression in patients whose melanomas are driven by *KIT* mutations.

Surveillance and Follow-up

Patients should be educated on the clinical characteristics of melanoma, the importance of safe sun exposure strategies, and the performance of monthly self-examinations of the skin. Patients should be followed regularly for evidence of local or regional recurrence, distant metastatic disease, and a second primary melanoma. The intensity of the surveillance and the extent of the investigation are influenced by risk for recurrence with more frequent follow-up visits in patients who have thicker tumors or node-positive disease because these patients are at greater risk for recurrence.

For patients with low-risk melanoma (≤1 mm), visits are recommended every 6 months for 2 years and then annually. The surveillance guidelines for patients with high-risk melanoma include evaluation every 3 to 4 months for 2 years and then every 6 months for 3 years. After 5 years, patients are seen once a year. Patients are generally followed for 10 years. However, lifelong dermatologic examination is recommended, particularly for patients with dysplastic nevi or a family history of melanoma. In general, a history and physical examination are performed at each visit. Periodic chest radiographs, laboratory studies, and other imaging studies are performed at the discretion of the treating physician. The physical examination should include a thorough skin examination because at least 3% of patients develop an additional primary cutaneous melanoma within 3 years. Regional lymph nodes should be thoroughly examined, especially in patients without prior nodal surgery. For the remainder of the examination, one should keep in mind the frequency of metastases to lung, liver, and brain. Follow-up studies may include a complete blood cell count and chemistry studies, including liver enzyme tests. An elevated lactate dehydrogenase level suggests metastatic melanoma.

PREVENTION

The most important measures to prevent melanoma are to reduce excessive sun exposure, particularly to the midday sun, and to avoid sunburns. Sunscreen products with a sun protection factor (SPF) of 15 or greater and protective clothing are recommended, and one randomized trial found

that regular sunscreen use reduced incident melanoma by 50% and invasive melanoma by 75%.[6] Sunscreens block primarily UVB rays, which are considered to be the major causative agent of cutaneous cancers. Newer sunscreen products also block UVA rays, which may contribute to the risk of melanoma.

Screening for skin cancer, whether by self-examination or by a health care provider, is controversial (Chapter 15). Many public health experts do not recommend screening for adults in the general population, but some organizations do. On the basis of the type and number of nevi, family history of melanoma, prior melanoma, and history of severe sunburns, clinicians can identify patients who are at high risk for melanoma and who may benefit from screening programs. In several population studies, screening has detected melanomas at an earlier, curable stage. Physicians, other health care providers, and the public should be educated regarding the early signs of melanoma and the need for prompt biopsy of a suspicious pigmented lesion.

Patients with clinically atypical nevi (see Fig. 203-2), particularly if they have a family history of melanoma, require a regular dermatologic surveillance program. Regular skin examinations should be performed every 6 to 12 months, preferably assisted by the use of serial photography.

Recent studies have focused on the impact of vitamin D levels on the risk for melanoma, and results have been conflicting. The potential health benefits of vitamin D continue to be evaluated, both in terms of melanoma prevention and risk.

BASAL AND SQUAMOUS CELL SKIN CANCER

Nonmelanoma skin cancer (BCC and SCC) is the most common malignant disease in the United States.[7] Although national statistics are imprecise, an estimated 900,000 to 1,200,000 of nonmelanoma skin cancers are diagnosed annually in the United States. SCC accounts for 20%, and most of the remainder are BCC. SCC is associated with a higher absolute mortality rate; most of the 2300 annual deaths from nonmelanoma skin cancer in the United States arise from this tumor.

EPIDEMIOLOGY

Overall, skin cancer incidence rates are rising because of increased recreational sun exposure, longer life expectancy, and depletion of the ozone layer. More than 99% of nonmelanoma skin cancers occur in whites. These skin cancers are most commonly seen in elderly persons, especially those with fair skin and long-standing sun exposure. However, nonmelanoma skin cancers are increasingly being seen in people in their 30s and 40s. The lifetime risk of developing BCC is 30%.

PATHOBIOLOGY

Basal cell carcinoma arises from a pluripotential stem cell within the skin. Acquired mutations in the patched gene 1 *(PTCH1)*, a tumor suppressor gene in the hedgehog signaling pathway, have been identified in cases of sporadic BCC. Sporadic BCC are also associated with mutations in the genes encoding p53 and ras.

Squamous cell carcinoma of the skin is a malignant disease of epidermal keratinocytes. Many such carcinomas are derived from actinic keratosis, a precursor that appears as a rough, scaly, often erythematous papule, which often is more apparent on palpation than on visual examination. Estimates of the likelihood of progression of actinic keratosis to SCC range from 0.025% to as high as 20%. Mutations in the gene encoding the p53 protein and in the *RAS* oncogene have been found in both actinic keratosis and SCCs. Mutations in *p16* have also been reported in SCCs.

Risk Factors

The most important risk factor is exposure to UV radiation from sunlight. The most clearly established association is with UVB radiation, but increasing evidence suggests that UVA is probably carcinogenic as well (Chapter 20). The timing and pattern of sun exposure are associated with different types of skin cancer. In general, SCC is associated with cumulative sun exposure and occurs most frequently in areas maximally exposed to the sun (e.g., the face, back of hands, and forearms). Intermittent, intense exposure to the sun, particularly in childhood, is associated with an increased risk for BCC. There is evidence for a dose-response relationship between artificial UV radiation exposure by use of tanning beds and the risk of skin cancers, especially BCC, and the association is stronger for individuals with a younger age at exposure.[8] Individuals who have fair skin, light-colored eyes, red hair, a tendency to burn rather than tan, and a history of severe sunburns are at increased risk for nonmelanoma skin cancers. Other risk factors, primarily for SCC, include chronic arsenic exposure, therapeutic radiation, chronic inflammatory skin conditions, psoralen plus UVA (PUVA) treatment for psoriasis and other diseases, and immunosuppression. Most cases in African American patients are associated with scarring or burns rather than UV exposure. Human papillomavirus infection (Chapter 373) has also been implicated in some SCCs, particularly in the autosomal dominant disorder epidermodysplasia verruciformis.

Basal cell carcinoma can be seen in association with several conditions, including the basal cell nevus syndrome (also called nevoid basal cell carcinoma syndrome or Gorlin syndrome), albinism, and xeroderma pigmentosum. The basal cell nevus syndrome is a rare autosomal dominant disorder caused by germline mutations in the patched gene *(PTCH)*.

CLINICAL MANIFESTATIONS

Basal Cell Carcinoma

Approximately 90% of BCCs occur on sun-exposed areas such as the face, neck, ears, scalp, and arms. The nose is the most common site. Typical BCC appears as slowly growing, shiny, skin-colored to pink translucent papules with telangiectasia and a "pearly," rolled border (Fig. 203-6). As the tumor grows, the center may become ulcerated and bleed, although there is usually no associated pain or tenderness. BCC rarely metastasizes and is usually curable with a variety of treatments. Although the mortality rate is low, these cancers may result in significant morbidity owing to invasive local growth with potential disfigurement and destruction of skin, bone, and cartilage. Clinical trials with inhibitors of the hedgehog pathway for patients with advanced BCC have demonstrated very encouraging clinical activity.

Squamous Cell Carcinoma

This type of skin cancer usually appears on areas of skin that are heavily damaged by sun exposure. The most common sites include the head or neck, back, forearms, and dorsum of the hand. Clinically, SCC occurs as a discrete scaly erythematous papule on an indurated base that can develop on normal-appearing skin or on an actinic keratosis (Fig. 203-7). The lesion may grow

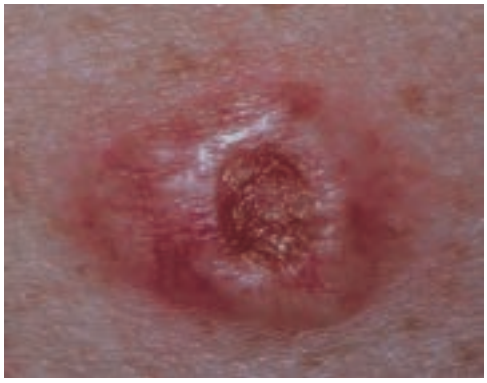

FIGURE 203-6. Basal cell carcinoma.

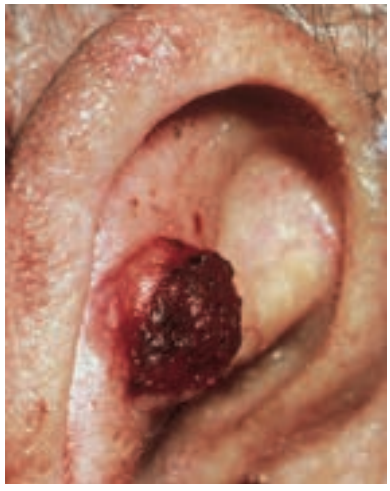

FIGURE 203-7. Squamous cell carcinoma of the skin.

over time and may become ulcerated, itchy, or painful and may bleed. Keratoacanthoma is a variant that is characterized by rapid growth and a crateriform appearance with a central plug. Bowen disease, or SCC in situ, manifests as an erythematous, scaly, sharply defined plaque.

Untreated SCC may cause significant local destruction. However, unlike BCC, SCC carries a 0.5% to 5% risk for metastasis. Higher risk lesions are those that are larger than 2 cm, are moderately or poorly differentiated, have perineural involvement, are located on the ear or the lip, arise in scars, or occur in immunosuppressed patients. Most metastases develop in regional lymph nodes, although metastases may also occur in lung, liver, brain, skin, or bone. For patients with lymph node metastases, the 5-year survival rate is less than 50%.

DIAGNOSIS

The diagnosis of BCC and SCC is frequently suspected by inspection alone, but histologic confirmation is usually indicated. Either a shave or a punch biopsy technique is acceptable (Chapter 436). Care should be taken to include the base of the lesion if a shave biopsy technique is used.

TREATMENT

Rx

Basal Cell Carcinoma

Basal cell carcinomas are classified as low or high risk based on their clinical features, location, and histology. Treatment options includes cryotherapy (liquid nitrogen), electrosurgery (i.e., curettage and electrodessication), topical treatment (i.e., 5-fluorouracil, photodynamic therapy, or imiquimod), surgical excision, Mohs' surgery, or radiation therapy. Mohs' microsurgery involves serial excisions of a skin cancer with subsequent microscopic examinations for residual tumor, providing histologic control of the surgical margins to achieve the lowest recurrence rate while maximally preserving uninvolving tissue.[9] The procedure should be considered when treating recurrent cases; microscopically aggressive forms, such as the morpheaform subtype; lesions greater than 2 cm in greatest diameter; and tumors of the ears, eyelids, nose, nasolabial folds, and lips. Cure rates for BCC range between 90% and 99%.

Treatment of advanced BCC and metastatic BCC can now be approached using targeted therapy. Vismodegib, an inhibitor of the hedgehog signaling pathway, has recently been FDA approved.[10] In a nonrandomized study, treatment with vismodegib provided objective tumor responses in 30% to 43% of patients and a complete response in 20% of patients with locally advanced or metastatic basal-cell carcinoma. Side effects associated with therapy include muscle spasms, taste abnormalities, diarrhea, and fatigue.

Squamous Cell Carcinoma

As with BCC, SCC can also be cured by traditional surgical excision or Mohs' surgery, cryotherapy, topical therapies, and radiation therapy. Topical 5-fluorouracil, photodynamic therapy, and imiquimod have roles in the management of in situ SCC. The optimal approach for a specific patient requires consideration of likelihood of the lesion recurring or metastatic potential, cosmetic factors, and the expertise of the treating physicians.

Mohs' micrographic surgery provides the lowest recurrence rate, with cure rates greater than 90%. Mohs' microsurgery is especially useful for recurrent tumors or lesions that have an increased risk of metastasis. Cetuximab, a monoclonal antibody that targets epidermal growth factor receptor (EGFR), has some antitumor activity in patients with advanced SCC of the skin.

Follow-up

Patients with BCC and SCC require ongoing follow-up to detect local recurrences and to recognize new skin cancers. The likelihood of developing a second BCC or SCC has been estimated to be 15% over 3 years. In addition, these patients have an increased risk of developing cutaneous melanoma. Patient education regarding modification of risk factors (i.e., sun exposure) is an important component of follow-up.

PREVENTION

Primary prevention strategies are aimed at reducing long-term sun exposure.[11] Public education and patient education should encourage the regular use of sunscreens with a SPF of 15 or greater, especially in childhood, and sun-protective clothing (e.g., a broad-brimmed hat). Avoidance of tanning parlors and minimizing of total sun exposure, especially to the midday sun, is recommended. The thinning of the ozone layer has been linked to increased UV radiation and increases in the incidence of nonmelanoma skin cancers. Currently, no evidence indicates that total-body skin examination is effective at reducing mortality or morbidity from nonmelanoma skin cancer.

Grade A References

A1. Gillgren P, Drzewiecki KT, Niin M, et al. 2-cm versus 4-cm surgical excision margins for primary cutaneous melanoma thicker than 2 mm: a randomised, multicentre trial. *Lancet*. 2011;378:1635-1642.

A2. Morton DL, Thompson JF, Cochran AJ, et al. Final trial report of sentinel-node biopsy versus nodal observation in melanoma. *N Engl J Med*. 2014;370:599-609.

A3. Robert C, Thomas L, Bondarenko I, et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med*. 2011;364:2517-2526.

A4. Robert C, Long GV, Brady B, et al. Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med*. 2015;372:320-330.

A5. Chapman PB, Hauschild A, Robert C, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med*. 2011;364:2507-2516.

A6. Robert C, Karaszewska B, Schachter J, et al. Improved overall survival in melanoma with combined dabrafenib and trametinib. *N Engl J Med*. 2015;372:30-39.

A7. Larkin J, Ascierto PA, Dréno B, et al. Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. *N Engl J Med*. 2014;371:1867-1876.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hill VK, Gartner JJ, Samuels Y, et al. The genetics of Melanoma. *Annu Rev Genomics Hum Genet*. 2013;14:257-279.
2. Eggermont AM, Spatz A, Robert C. Cutaneous melanoma. *Lancet*. 2014;383:816-826.
3. Snyder A, Makarov V, Merghoub T, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med*. 2014;371:2189-2199.
4. Wolchok JD, Kluger H, Callahan MK, et al. Nivolumab plus ipilimumab in advanced melanoma. *N Engl J Med*. 2013;369:122-133.
5. Flaherty KT, Robert C, Hersey P, et al. Improved survival with MEK inhibition in BRAF-mutated melanoma. *N Engl J Med*. 2012;367:107-114.
6. Green AC, Williams GM, Logan V, et al. Reduced melanoma after regular sunscreen use: randomized trial follow-up. *J Clin Oncol*. 2011;29:257-263.
7. Madan V, Lear JT, Szeimies RM. Non-melanoma skin cancer. *Lancet*. 2010;375:673-685.
8. Zhang M, Qureshi AA, Geller AC, et al. Use of tanning beds and incidence of skin cancer. *J Clin Oncol*. 2012;30:1588-1593.
9. Ad Hoc Task Force, Connolly SM, Baker DR, Coldiron BM, et al. AAD/ACMS/ASDSA/ASMS 2012 appropriate use criteria for Mohs micrographic surgery: a report of the American Academy of Dermatology, American College of Mohs Surgery, American Society for Dermatologic Surgery Association, and the American Society for Mohs Surgery. *J Am Acad Dermatol*. 2012;67:531-550.
10. Sekulic A, Migden MR, Oro AE, et al. Efficacy and safety of vismodegib in advanced basal-cell carcinoma. *N Engl J Med*. 2012;366:2171-2179.
11. Lin JS, Eder M, Weinmann S. Behavioral counseling to prevent skin cancer: a systematic review for the U.S. Preventive Services task force. *Ann Intern Med*. 2011;154:190-201.

REVIEW QUESTIONS

1. A 62-year-old woman was recently diagnosed with a 0.77-mm-thick melanoma. There was no ulceration, it was a Clark's level III, and the mitotic count was 1.0/mm^2. The melanoma arose from her left posterior arm. The patient had blood work performed, which included a CBC, chemistry panel, and LDH, which were normal. Which of the following imaging studies is appropriate for this patient?

A. No imaging studies
B. CT scan of the chest, abdomen, and pelvis and MRI of the brain
C. PET/CT scan and MRI of the brain
D. Bone scan; CT scan of the chest, abdomen, and pelvis; and brain MRI
E. CT scan of the chest, abdomen, and pelvis

Answer: A The staging evaluation of a patient with a low risk melanoma should include a physical examination, including a skin examination. Routine blood work and imaging is not routinely recommended. The majority of patients who present with melanoma do not have distant metastatic disease at presentation; therefore, extensive evaluations with computed tomography (CT) scans to search for distant metastases have an extremely low yield and consequently are not indicated in asymptomatic patients. More extensive staging evaluation with CT scans of the chest, abdomen, and pelvis can be considered in patients with high-risk disease (thick primary melanoma >4 mm thick or node-positive disease) in whom the risk of distant metastatic disease is higher.

National Comprehensive Cancer Network (NCCN) guidelines. Available at http://www.nccn.org.

2. A 55-year-old man presents to your office with a right axillary lump. Two years ago, he was treated for a malignant melanoma on his right upper arm. The melanoma was 1.3 mm thick, and there was no ulceration present. The melanoma was treated with wide excision, which showed no residual melanoma, and the patient was then monitored. He now presents to your office with an enlarged right axillary lymph node, which measures 2.5 × 2.0 cm. The patient has no symptoms, no fever, and the remainder of the physical examination findings are unremarkable. Which of following would you recommend to this patient now?

A. Arrange for fine-needle aspiration (FNA) of the lymph node
B. Surgical excision of the lymph node
C. Prescribe of a 10-day course of antibiotics
D. Referral to radiation therapy to begin radiation therapy to right axillary region
E. Continue monitoring and have the patient return to the office in 3 months because this likely represents a postoperative seroma

Answer: A In this patient with a history of melanoma, an enlarged regional lymph node is highly suspicious for disease recurrence. Obtaining an FNA, often with ultrasound guidance, is the most appropriate next step to confirm diagnosis. The physical examination does not suggest this represents an infection, and postoperative seroma would be highly unusual this far out from surgery. After melanoma is confirmed, then staging evaluation should be completed with CT scans of the chest, abdomen, and pelvis or PET scan. Additional surgery with complete lymph node dissection and plans for adjuvant systemic therapy would follow providing no distant metastatic disease was identified.

3. A 35-year-old woman presents with a mild cough, fatigue, and some weight loss. She was treated for melanoma on her left lower leg 3 years ago. Imaging studies reveal multiple pulmonary nodules. A CT-guided biopsy confirms diagnosis of metastatic melanoma. Analysis of the tumor for which of the following mutations would be most useful in determining treatment options?

A. *KIT*
B. *BRAF*
C. *EGFR*
D. *ALK*
E. *KRAS*

Answer: B *BRAF* V600 is the most common somatic mutation in melanoma, occurring in 50% of patients. It also tends to occur in younger patients with melanoma, so this is the most likely mutation in this patient. Importantly, BRAF inhibitors (vemurafenib, dabrafenib) are effective therapies for patients with *BRAF* mutant melanoma and are now approved by the Food and Drug Administration based on survival benefit demonstrated in randomized phase III clinical trials. *KIT* is uncommonly mutated in melanoma (<3% of melanomas). *KIT* mutations are most frequent in acral (arising on palms and soles or subungual), mucosal melanoma or cutaneous melanomas associated with chronic sun exposure. In this patient, testing for *BRAF* mutation first is most appropriate. *EGFR* mutation is not associated with melanoma; it is mutated in lung cancer. *ALK* and *KRAS* are not somatic mutations that occur in melanoma.

4. Which of the following patients has the highest risk for developing melanoma?

A. A 35-year-old woman with red hair and blue eyes who has had more than five blistering sunburns
B. A72-year-old farmer who has had chronic sun exposure for more than 60 years
C. A68-year-old woman who has had several basal cell and squamous cell skin cancers
D. A29-year-old woman with multiple dysplastic nevi and an uncle and cousin who had melanoma

Answer: D All of these patients are at increased for melanoma. Individuals with red hair, blue eyes, history of blistering sun burns, chronic sun exposure, and history of nonmelanoma skin cancers (basal cell and squamous cell) are at increased risk. However, the highest risk patient is the woman with dysplastic nevi and a family history of melanoma. In this patient, a careful family history should be obtained to determine if additional relatives have melanoma or other cancers, and all should be referred for dermatology evaluation. This patient could be part of familial melanoma/dysplastic nevus syndrome, which is associated with very high risk of developing melanoma and associated with germline mutation in p16.

5. A 44-year-old man was recently diagnosed with stage IV melanoma. Staging workup showed multiple liver metastases, lung metastases, and bone metastases. Symptoms at this time include mild shortness of breath, cough, significant bone pain, and an 8-lb weight loss. Mutation testing for *BRAF* confirmed *V600E BRAF* mutation. What treatment would you recommend now?

A. Dabrafenib
B. Trametinib
C. Imatinib
D. Sorafenib

Answer: A In this patient with symptomatic metastatic melanoma, the most appropriate choice would be a highly potent and selective *BRAF* inhibitor. Dabrafenib treatment is associated with an approximately 60% response rate. Clinical benefit has been seen within 72 hours of administration. Trametinib is a recently approved MEK inhibitor; however, clinical activity is less than that reported for dabrafenib. Similarly, sorafenib, although it targets BRAF, has little activity in melanoma. Imatinib targets *c-kit*, so its selection is not appropriate.

Ascierto P, Minor D, Ribas A, et al. Phase II trial (BREAK-2) of the BRAF inhibitor dabrafenib (GSK118436) in patients with metastatic melanoma. *J Clin Oncol.* 2013;31:3205-3211.

6. A 63-year-old woman was recently diagnosed with stage IV melanoma. Staging workup showed lung and bone metastases. She has no symptoms at this time. Her oncologist has recommended treated with ipilimumab. Which of the following statement best describes the mechanism of action of ipilimumab?

A. Blocks cytotoxic T-lymphocyte antigen 4 (CTLA-4)
B. Inhibits DNA replication
C. Inhibits the MAP kinase pathway
D. Directly stimulates T cells activity
E. Interferes with CD 20 on B cells

Answer: A Ipilimumab is a negative regulator of T-cell activation. It interferes with the ability of CTLA-4 to interact with its ligand, which is an inhibitory signal. Thus, CTLA-4 blockade with ipilimumab results in T-cell activation and proliferation. Ipilimumab, a type of immunotherapy, is approved by the Food and Drug Administration (FDA) for the treatment of unresectable stage III and IV melanoma. Chemotherapy agents such as cyclophosphamide interfere with DNA replication. Molecularly targeted agents such as BRAF inhibitors work through inhibiting the MAP kinase pathway. Interleukin-2, FDA approved for the treatment of metastatic melanoma, directly stimulates T cells. Monoclonal antibodies, such as rituximab, induces destruction of B cells, which express CD 20. This therapy is used to treat patients with chronic lymphocytic leukemia and non-Hodgkin's lymphoma.

204

CANCER OF UNKNOWN PRIMARY ORIGIN

JOHN D. HAINSWORTH AND F. ANTHONY GRECO

DEFINITION

The first signs or symptoms of cancer are frequently the result of metastases to visceral or nodal sites. In most such patients, the clinical diagnosis of metastatic cancer is evident and is confirmed by biopsy of a metastatic lesion. Subsequent clinical evaluation with a comprehensive history, physical examination, complete blood cell count, screening chemistries, chest and abdominal computed tomography (CT) scans, and directed radiologic studies based on specific symptoms or signs identifies the primary tumor site as well as the extent of metastatic disease. Patients who have no primary tumor located after this clinical evaluation are defined as having *cancer of unknown primary origin*. Further clinical and pathologic evaluation identifies the primary site in only a few patients, and approximately 80% never have a primary site identified during their subsequent clinical course.

EPIDEMIOLOGY

In patients whose primary site of cancer remains undetectable, the primary site has presumably remained small or, less likely, regressed spontaneously. Before the routine use of CT or magnetic resonance imaging (MRI) for diagnosis, large autopsy series identified primary sites (usually <2 cm in diameter) in 85% of patients with cancer of unknown primary origin, usually in the pancreas, lung, and various other gastrointestinal (GI) sites. With the use of CT and MRI for diagnosis, however, primary sites are identified at autopsy in only 50% to 70% of patients.[1]

Approximately 4% of all patients with cancer have metastatic disease without a known primary site; the annual incidence is approximately 80,000 cases in the United States. Cancer of unknown primary site occurs with approximately equal frequency in men and women, and it increases in incidence with advancing age.

DIAGNOSIS

The initial clinical and pathologic evaluations should focus on identifying a primary site, when possible, and on identifying patients for whom specific treatment is indicated.

Biopsy and Pathologic Evaluation

The diagnosis of metastatic cancer should be confirmed by biopsy of the most accessible metastatic lesion. A fine-needle aspiration is usually sufficient to confirm the diagnosis of metastatic cancer but does not provide adequate material for optimum pathologic evaluation. Therefore, a larger biopsy (surgical or core needle) should be performed if technically feasible.

TABLE 204-1 RECOMMENDED EVALUATION AFTER INITIAL LIGHT MICROSCOPIC DIAGNOSIS

DIAGNOSIS	CLINICAL EVALUATION*	SPECIAL PATHOLOGIC STUDIES
Adenocarcinoma (or poorly differentiated adenocarcinoma)	PET CT of the chest and abdomen Men: serum PSA Women: mammography, breast MRI Colonoscopy (patients with colon cancer "profile") Additional directed radiologic or endoscopic studies to evaluate abnormal symptoms, signs, laboratory values	Men: PSA stain Women: estrogen and progesterone receptor stains (if clinical features suggest metastatic breast cancer) Molecular tumor profiling
Poorly differentiated carcinoma	PET CT of the chest and abdomen Serum hCG and AFP Additional directed radiologic or endoscopic studies to evaluate abnormal symptoms, signs, and laboratory values	Immunoperoxidase staining Molecular tumor profiling Electron microscopy (if other studies are indeterminate or conflicting)
SCC, cervical nodes	PET Direct laryngoscopy with visualization; biopsy of the nasopharynx, pharynx, hypopharynx, and larynx Fiberoptic bronchoscopy (if laryngoscopy results are negative)	—
SCC, inguinal nodes	PET Complete examination of perineal area (including pelvic examination) Anoscopy Cystoscopy	—

*In addition to a history, physical examination, complete blood cell counts, chemistry profile, and chest radiograph.
AFP = α-fetoprotein; CT = computed tomography; hCG = human chorionic gonadotropin; MRI = magnetic resonance imaging; PET = positron emission tomography; PSA = prostate-specific antigen; SCC = squamous cell carcinoma.

The initial light microscopic evaluation identifies adenocarcinoma in approximately 60% of patients with cancer of unknown primary site. Other diagnoses obtained by initial light microscopy include poorly differentiated carcinoma (25%), squamous carcinoma (10%), and poorly differentiated neoplasm (inability to distinguish among carcinoma, lymphoma, melanoma, and sarcoma; 5%).[3]

Additional pathologic evaluation is essential in all poorly differentiated tumors. Immunochemical (IHC) stains can usually identify the tumor lineage (e.g., carcinoma vs. lymphoma vs. sarcoma) when the histologic diagnosis is "poorly differentiated neoplasm." Because more than 50% of these patients have lymphoma, this distinction is important. When the histologic diagnosis is "poorly differentiated carcinoma," IHC staining sometimes identifies germ cell tumors or neuroendocrine carcinoma.

In patients with adenocarcinoma, it is seldom possible for the pathologist to determine a primary site by light microscopic examination of histology. IHC stains can narrow the diagnostic spectrum, particularly when interpreted in conjunction with clinical features. In several situations, IHC stains are quite specific, including prostate-specific antigen (PSA) for prostate cancer (Chapter 201), estrogen and progesterone receptors for breast cancer (Chapter 198), and leukocyte common antigen for non-Hodgkin lymphoma (Chapter 185). Other diagnoses suggested by immunoperoxidase staining include neuroendocrine carcinomas, melanomas (Chapter 203), and sarcomas (Chapter 202).

Occasionally, electron microscopy or analysis for tumor-specific chromosomal abnormalities (i12p in germ cell tumors, Chapter 200; t11:22 in Ewing tumor; immunoglobulin gene rearrangements in non-Hodgkin lymphoma, Chapter 185) are useful in the evaluation of poorly differentiated tumors if results of other pathologic studies are inconclusive.

Molecular tumor profiling is a new diagnostic method that is in the process of changing the management of patients with cancer of unknown primary origin. Gene expression profiles differ in various normal body tissues, and most cancers retain expression profiles specific to their tissue of origin. Molecular profiling assays can correctly predict the tissue of origin in 85% to 90% of metastatic cancers by detecting tissue-specific gene expression patterns.[4,5] Although the accuracy of the profiling predictions is more difficult to quantitate in patients with cancer of unknown primary origin (because most primaries never become manifest), current evidence indicates a similar high accuracy rate.[6,7] Therefore, molecular tumor profiling is a valuable addition to the pathologic evaluation, and is indicated when histologic examination and IHC do not identify a tissue of origin.

Search for the Primary Site

After completion of the brief, directed initial evaluation necessary to make the diagnosis of cancer of unknown primary origin (see the Definition section), further diagnostic studies should be limited (Table 204-1). The value of positron emission tomography (PET) in identifying a primary site is debated; in the only prospective study to date, PET results did not augment results of CT scanning.[8] Other routine radiologic and endoscopic evaluations of asymptomatic areas are rarely useful in identifying a primary site and therefore are not recommended. Levels of serum tumor markers, including carcinoembryonic antigen, CA-125, CA-19-9, and CA-15-3, are frequently increased in patients with carcinoma of unknown primary site; however, these elevations are nonspecific and should not be used to infer a primary site even though they can be useful in monitoring response to treatment.

Specific clinical presentations should trigger additional diagnostic evaluation. In all men with metastatic adenocarcinoma, serum PSA should be measured. Mammography and breast MRI should be considered in women with metastatic adenocarcinoma, particularly if clinical features are consistent with metastatic breast cancer (e.g., axillary node involvement, pleural effusion, lytic or blastic bone metastases). In patients younger than 50 years of age with poorly differentiated carcinoma, serum human chorionic gonadotropin and α-fetoprotein (AFP) levels should be measured to screen for germ cell tumors. Patients with metastatic squamous carcinoma involving cervical lymph nodes should have a thorough endoscopic evaluation of the head and neck, including visualization of the structures from the nasopharynx to the larynx and biopsy of any suspicious areas (Chapter 190). Fiberoptic bronchoscopy should also be considered in patients who have low cervical adenopathy and in whom no head or neck primary site is established by endoscopic examination. In patients with metastatic squamous carcinoma involving inguinal lymph nodes, all perineal structures should be carefully inspected, including by anoscopy, a urologic evaluation, and a pelvic examination in women.

TREATMENT

Management of Specific Treatable Subsets

Because patients with cancer of unknown primary site have advanced disease, therapeutic nihilism has been common. However, several subsets of patients can benefit from specific treatment, and they can be identified on the basis of clinical and pathologic features (Table 204-2). These patients are important to recognize and treat appropriately because some individuals in each group have the potential for long-term survival.

Adenocarcinoma

Women with Axillary Lymph Node Metastases

Metastatic breast cancer should be suspected in women who have axillary lymph node involvement with adenocarcinoma, particularly when other metastatic sites are not evident. In these patients, pathologic evaluation of the

TABLE 204-2 SPECIFIC PATIENT SUBSETS AND RECOMMENDED TREATMENT

SUBSET-IDENTIFYING FEATURES		
HISTOLOGIC	**CLINICAL**	**TREATMENT RECOMMENDATIONS**
Adenocarcinoma	Women, isolated axillary adenopathy	Treat as stage II breast cancer
Adenocarcinoma, poorly differentiated carcinoma	Women, peritoneal carcinomatosis (Occasionally men?)	Treat as stage III ovarian cancer
Adenocarcinoma	Men, elevated PSA or blastic bone metastases	Treat as advanced prostate cancer
Adenocarcinoma, poorly differentiated carcinoma	Single metastatic lesion	Definitive local therapy (resection or radiation therapy [or both]) with or without chemotherapy
Adenocarcinoma	Colon cancer profile	Treat as metastatic colorectal cancer
Squamous	Cervical adenopathy	Treat as locally advanced head or neck cancer
Squamous	Inguinal adenopathy	Definitive local therapy (node dissection with or without radiation therapy) with or without chemotherapy
Poorly differentiated carcinoma	Young men with midline tumor or elevated hCG or AFP	Treat as extragonadal germ cell tumor
Neuroendocrine carcinoma, poorly differentiated	Diverse clinical presentations	Treat as advanced stage small cell lung cancer
Neuroendocrine carcinoma, well differentiated	Usually liver metastases	Treat as metastatic carcinoid tumor

AFP = α-fetoprotein; hCG = human chorionic gonadotropin; PSA = prostate-specific antigen.

initial lymph node biopsy should include staining for estrogen and progesterone receptors and for *HER-2* expression; elevated levels provide strong evidence for the diagnosis of breast cancer.

When no other metastases are identified, these women should be treated as if they had stage II breast cancer, which is potentially curable with appropriate therapy (Chapter 198). Modified radical mastectomy identifies a breast primary site in 44% to 82% of women even when the breast examination and mammographic findings are normal. Axillary lymph node dissection followed by radiation therapy to the breast appears to give results similar to those of mastectomy, although these two options for primary therapy have not been compared directly. Adjuvant systemic therapy should follow standard guidelines for the treatment of women with stage II breast cancer.

Women with Peritoneal Carcinomatosis

Adenocarcinoma involving the peritoneum in women usually originates from the ovary (Chapter 199), although carcinomas arising in the GI tract or breast can occasionally produce this syndrome. However, diffuse peritoneal carcinomatosis occasionally occurs in women who have histologically normal ovaries or who have had previous bilateral oophorectomy. The peritoneum is frequently the only site of tumor involvement, and serum CA-125 levels are usually elevated. When histologic features suggest ovarian cancer, this syndrome has been called *peritoneal papillary serous carcinoma* or *primary extraovarian serous carcinoma*.

Even when the histologic features are not typical, women with adenocarcinoma of unknown primary site involving the peritoneum often have cancers with biologic characteristics similar to those of ovarian cancer (Chapter 199). Treatment of these patients should follow guidelines for stage III ovarian cancer. When feasible, a full laparotomy with maximal surgical cytoreduction should be performed followed by combination chemotherapy with a taxane/platinum–containing regimen. Measurement of serial serum CA-125 levels provides an accurate assessment of the efficacy of treatment. A few of these patients may have complete responses and long-term survival, particularly when initial surgical cytoreduction leaves minimal residual disease. A similar syndrome of peritoneal carcinomatosis that is responsive to chemotherapy for ovarian cancer has rarely been reported in men.

Men with Skeletal Metastases or Elevated Serum Prostate-Specific Antigen Levels

Metastatic prostate cancer (Chapter 201) should be suspected in men with adenocarcinoma predominantly involving bone, particularly if the metastases are blastic. An elevated serum PSA level or tumor immunostaining for PSA confirms the diagnosis of prostate cancer. Occasionally, men with adenocarcinoma of unknown primary site and patterns of metastasis unusual for prostate cancer (e.g., lung metastases, mediastinal lymph node metastases) are found to have elevated serum PSA levels. These patients should be treated according to guidelines for advanced prostate cancer. Androgen ablation produces excellent responses and substantial palliation in most patients.

Single Metastatic Lesion

Occasionally, a single metastatic lesion containing adenocarcinoma or poorly differentiated carcinoma is identified, and a complete evaluation reveals no other evidence of disease. Such presentations can include a single lymph node or subcutaneous site or single lesions in various visceral sites, including bone, liver, lung, brain, and adrenal gland. The possibility of an unusual primary site mimicking a metastatic lesion should be considered (e.g., a subcutaneous nodule from a primary apocrine or sebaceous carcinoma rather than a metastasis), but this possibility can usually be excluded on the basis of clinical or pathologic features. PET is useful in excluding other metastatic lesions.

For patients with only a single identifiable lesion, definitive local therapy is recommended, guided by the site of tumor involvement. Such therapy may include surgical resection, radiation therapy, or a combination of these modalities. Although most of these patients eventually develop other metastatic sites, a significant disease-free interval is often experienced, and local treatment provides substantial palliation. The role of systemic chemotherapy in addition to definitive local therapy is not well defined; younger patients with poorly differentiated carcinoma or poorly differentiated adenocarcinoma are often treated with a short course of a taxane/platinum–based regimen.

Colon Cancer Profile

The accurate recognition of patients likely to respond to treatment for advanced colon cancer has become increasingly important because the efficacy of colon cancer treatment has improved substantially. In patients with cancer of unknown primary origin, a "colon cancer profile" includes (1) metastases predominantly in the liver, peritoneum, or both; (2) adenocarcinoma with histologic features typical of GI origin; and (3) typical immunohistochemical staining (CK20 positive, CK7 negative, and CDX-2 positive).[9] Patients with this profile should be treated according to guidelines for metastatic colorectal cancer (Chapter 193).

Squamous Cell Carcinoma

Cervical Adenopathy

Squamous cell carcinoma (SCC) of unknown primary site is relatively uncommon. Most patients with this syndrome have involvement of cervical lymph nodes, usually in the upper or midcervical area. Often, patients with this syndrome are middle aged or elderly and have a history of substantial tobacco or alcohol use or both. A primary site in the head and neck region should be suspected (Chapter 190); however, complete endoscopic evaluation fails to identify a primary site in approximately 15% of these patients. Even if other test results are negative, PET identifies a primary site in the head and neck region in approximately 25% of such patients and should be part of the initial evaluation.

Even when no primary site is identified, these patients should receive treatment with concurrent chemotherapy and radiation therapy,[10] as is currently recommended for locally advanced SCC arising in the head and neck. Combined modality therapy produces 5-year disease-free survival rates of 50% to 60%; multiple involved lymph nodes or nodes larger than 2 cm are adverse prognostic features (Chapter 190).

Inguinal Adenopathy

Occasionally, metastatic squamous cell cancer is found in inguinal lymph nodes. In most of these patients, a primary site can be located in the perineal or anorectal area. For the occasional patient in whom no primary site is identified, long-term survival can result from local therapy with inguinal lymph node dissection, with or without radiation therapy. Recently, combined-modality treatment with concurrent chemotherapy and radiation therapy has improved

cure rates in patients with several SCCs arising in this region (e.g., cervix, anus, bladder). Although data are incomplete, a reasonable approach is the addition of chemotherapy with a platinum–5-fluorouracil regimen, as described for locally advanced carcinoma of the cervix.

Poorly Differentiated Carcinoma

Extragonadal Germ Cell Cancer Syndrome

Young men with clinical features of extragonadal germ cell tumors, including tumors in the mediastinum or retroperitoneum or those associated with elevated serum levels of human chorionic gonadotropin or AFP, should be treated according to guidelines for extragonadal germ cell tumors (Chapter 200). Some of these patients can be proven to have germ cell tumors by identifying an i12p chromosomal abnormality even when the diagnosis is not possible with other standard pathologic techniques. Approximately 30% to 40% of such patients achieve complete responses and long-term survival after chemotherapy with cisplatin, etoposide, and bleomycin as used for advanced germ cell tumors.

Anaplastic Lymphoma

An appropriate initial pathologic evaluation should identify most histologically atypical lymphomas. Occasionally, IHC staining for leukocyte common antigen is negative or cannot be adequately performed in patients with anaplastic lymphoma. The disease in some of these patients can be recognized using other IHC stains (e.g., Ki-1, CD-30), molecular genetic analysis (detection of immunoglobulin gene rearrangements), or molecular tumor profiling. All patients with lymphomas identified by special pathologic studies should be treated using standard guidelines for aggressive non-Hodgkin lymphoma (Chapter 185).

Neuroendocrine Carcinoma

In approximately 10% of poorly differentiated carcinomas, neuroendocrine features are identified by either IHC staining or electron microscopy. Treatment of these patients is discussed later (see Neuroendocrine Carcinoma).

Other Poorly Differentiated Carcinomas

In most patients with poorly differentiated carcinoma, there are no clinical or pathologic features that allow their assignment to any of the treatable subsets. These patients have a prognosis similar to patients with adenocarcinoma of unknown primary site, and treatment should follow similar guidelines.

Neuroendocrine Carcinoma

Poorly Differentiated Neuroendocrine Carcinoma or Small Cell Anaplastic Carcinoma

These high-grade neuroendocrine tumors are now reliably identified using widely available IHC stains. Although the origin of these tumors remains obscure, they are often highly sensitive to combination chemotherapy; platinum–etoposide chemotherapy as used in the treatment of small cell lung cancer, produces an overall response rate of approximately 60%, and 15% to 20% of patients have complete responses. In patients with locoregional disease, the addition of radiation therapy after chemotherapy is reasonable.

Low-Grade (Carcinoid-Type) Neuroendocrine Tumors

Occasionally, low-grade neuroendocrine tumors are found at a metastatic site. In almost all cases, the liver is the site of involvement, and the histologic features suggest a carcinoid (Chapter 232) or islet cell tumor of GI origin (Chapter 195). Various clinical syndromes caused by the secretion of vasoactive peptides (e.g., serotonin, vasoactive intestinal peptide, gastrin) have been described. Similar to other carcinoid tumors, these tumors often have indolent biologic characteristics, and patients can frequently survive for several years despite multiple liver metastases. Unlike poorly differentiated neuroendocrine tumors, these tumors are relatively resistant to chemotherapy, and intensive combination regimens should usually be avoided. Management of these patients should follow guidelines for metastatic carcinoid tumors (Chapter 232) and may include the use of somatostatin analogues, local ablative procedures (e.g., surgical resection, radiofrequency ablation, chemoembolization), targeted agents (e.g., sunitinib, everolimus), or fluorouracil-based chemotherapy regimens.

Empiric Chemotherapy

Approximately 70% to 80% of patients with adenocarcinoma or poorly differentiated carcinoma of unknown primary site do not fit into any of these defined clinical subsets. In these patients, treatment with empiric chemotherapy (usually taxane–platinum or gemcitabine–platinum combinations) has been of modest benefit, producing response rates of 30% to 45% and a median survival time of 9 to 11 months. However, empiric chemotherapy regimens, designed at a time when there was substantial overlap in the systemic therapy of different tumor types, are no longer able to provide adequate "coverage" as therapy for different cancers becomes more specific. At present, the standard of care is shifting from empiric chemotherapy to site-specific treatment, directed by the molecular tumor profiling prediction of the site of tumor origin.

Site-Specific Therapy Directed by Molecular Tumor Profiling

Molecular tumor profiling results in a prediction of the primary site in more than 90% of patients with cancer of unknown origin. Site-specific therapy has several potential advantages over empiric therapy, including the ability to appropriately incorporate tumor-specific targeted agents and to avoid unnecessary treatment in patients with unresponsive tumor types. In a group of 194 patients with cancer of unknown primary, assay-directed site-specific treatment produced a median survival period of 12.5 months.[11] Patients predicted to have more responsive tumor types (and therefore likely to derive the most benefit from site-specific treatment) had a median survival period of 13.4 months compared with 7.6 months for less treatment-responsive tumor types. In general, treatment responses and survival were consistent with the cancer type predicted; patients predicted to have breast or ovarian tumors had median survival periods of more than 24 months with site-specific treatment.[12] As additional treatment results become available, it is likely that molecular assay-directed, site-specific therapy will become the new treatment standard, with empiric chemotherapy reserved for patients whose tumors are unclassifiable by molecular profiling.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Varadhachary GR, Raber MN. Carcinoma of unknown primary site. *N Engl J Med*. 2014;371:2040.
2. Brewster DH, Lang J, Bhatti LA, et al. Descriptive epidemiology of cancer of unknown primary site in Scotland, 1961-2010. *Cancer Epidemiol*. 2014;38:227-234.
3. Morris GJ, Greco FA, Hainsworth JD, et al. Cancer of unknown primary site. *Semin Oncol*. 2010;37:71-79.
4. Erlander MG, Ma XJ, Kesty NC, et al. Performance and clinical evaluation of the 92-gene real time PCR assay for tumor classification. *J Mol Diagn*. 2011;13:493-503.
5. Meiri E, Mueller WC, Rosenwald S, et al. A second-generation microRNA-based assay for diagnosing tumor tissue origin. *Oncologist*. 2012;17:801-812.
6. Greco FA, Lennington WJ, Spigel DR, et al. Molecular profiling diagnosis in unknown primary cancer: accuracy and ability to complement standard pathology. *J Natl Cancer Inst*. 2013;105:782-790.
7. Hainsworth JD, Greco FA. Gene expression profiling in patients with carcinoma of unknown primary site: from translational research to standard of care. *Virchows Arch*. 2014;64:393-402.
8. Moller AK, Loft A, Berthelsen AK, et al. A prospective comparison of 18F-FDG PET/CT and CT as diagnostic tools to identify the primary site in patients with extracervical cancer of unknown primary site. *Oncologist*. 2012;17:1146-1154.
9. Varadhachary GR, Karanth S, Qiao W, et al. Carcinoma of unknown primary with a gastrointestinal profile: Immunohistochemistry and survival data for this favorable subset. *Int J Clin Oncol*. 2014;19:479-484.
10. Cerezo L, Raboso E, Ballesteros AI. Unknown primary cancer of the head and neck: a multidisciplinary approach. *Clin Trans Oncol*. 2011;13:88-97.
11. Hainsworth JD, Rubin MS, Spigel DR, et al. Molecular gene expression profiling to predict the tissue of origin and direct site-specific therapy in patients with carcinoma of unknown primary site: a prospective trial of the Sarah Cannon Research Institute. *J Clin Oncol*. 2013;31:217-223.
12. Greco FA, Lennington WJ, Spigel DR, et al. Carcinoma of unknown primary site: outcomes in patients with a colorectal molecular profile treated with site-specific chemotherapy. *J Cancer Ther*. 2012;3:37-43.

REVIEW QUESTIONS

1. A 65-year-old man, former cigarette smoker, presents with an asymptomatic right neck mass. He feels well otherwise. Physical examination is normal except for a firm, nontender, movable 2 cm lymph node in the right jugulodigastric area. CT scan of the neck shows the palpable right cervical node and two adjacent nodes, each measuring 1.5 cm, deeper in the right anterior cervical area. CT scans of the chest and abdomen are normal. Needle biopsy of the palpable node shows squamous carcinoma. The most likely diagnosis is

A. metastatic lung cancer.
B. metastatic gastric cancer.
C. locally advanced base of tongue cancer.
D. metastatic esophageal cancer.
E. metastatic squamous skin cancer.

Answer: C Squamous cell carcinomas (SCCs) arising from various sites in the head and neck (nasopharynx, oropharynx, hypopharynx, larynx) frequently metastasize to ipsilateral upper cervical lymph nodes. In 85% of patients who present with metastatic SCC in the upper cervical nodes, appropriate evaluation identifies the head or neck primary site. SCCs of the lung or esophagus occasionally present with a palpable neck node but almost always in the lower cervical or supraclavicular area. Metastatic gastric cancer sometimes spreads to the left supraclavicular nodes (Virchow's nodes) but has adenocarcinoma histology. Metastases from SCC skin cancers are rare.

2. In the patient described in question 1, the procedure most likely to identify the primary site is

A. PET scan.
B. ENT endoscopy (nasopharynx, oropharynx, hypopharynx, larynx).
C. fiberoptic bronchoscopy.
D. upper GI endoscopy.
E. right radical neck dissection.

Answer: B Endoscopy with visualization of the mucosal surfaces of the nasopharynx, oropharynx, hypopharynx, and larynx identifies a primary site in a large majority of such patients. PET scan can also identify an occult head or neck primary site but much less consistently. Bronchoscopy and upper GI endoscopy are not usually productive because most often the primary site is in the head or neck. Radical neck dissection before further evaluation is inappropriate.

3. A previously healthy 30-year-old man presents with fatigue, a 10-lb weight loss, and severe abdominal and back pain that has steadily increased during the past month. Abdominal CT scan shows a 15 × 10 × 8 retroperitoneal mass, with the superior aspect adjacent to the pancreas but not clearly invading the pancreas or other organs. A CT-guided core needle biopsy shows poorly differentiated carcinoma. A chest CT scan shows a single 0.5-cm noncalcified right lung nodule of questionable significance. Further evaluation should include all of the following except

A. immunohistochemical staining of the biopsy specimen (by pathologist) to better characterize the cancer type.
B. GU history, particularly any history of cryptorchidism.
C. testicular ultrasonography.
D. serum hCG and AFP levels.
E. endoscopic retrograde cholangiopancreatography (ERCP).

Answer: E Because of this patient's age, tumor location in the retroperitoneum, and biopsy showing poorly differentiated carcinoma, an extragonadal germ cell tumor (EGCT) is a likely diagnosis and is definitely the most treatable carcinoma in this setting. Additional pathologic evaluation with immunohistochemical stains (OCT4, PLAP, hCG) may support a germ cell tumor diagnosis, and high-grade non-Hodgkin's lymphoma (occasionally mistaken for a poorly differentiated carcinoma) should be ruled out with an LCA stain. Elevated serum hCG or AFP in this clinical setting would provide strong evidence for an EGCT. The risk of germ cell tumor is increased in patient with cryptorchidism even after surgical correction. Pancreatic cancer is an unlikely diagnosis in this situation (size and location of mass, age of patient, histology), and ERCP would prolong the evaluation and delay initiation of treatment.

4. A 60-year-old woman presents with a 2-month history of anorexia, 15-lb weight loss, and mild RUQ abdominal discomfort. Physical examination results are normal. Blood counts are normal; the chemistry profile reveals levels of AST and ALT three times the upper limit of normal. Abdominal CT scan shows six liver lesions consistent with metastases (the largest is 4 cm). Chest CT scan is normal. Colonoscopy is normal. Liver biopsy (core needle) shows metastatic adenocarcinoma compatible with a GI primary tumor; immunohistochemical staining shows CK20 positive, CK7 negative, and CDX2 positive. Molecular tumor profile predicts a colorectal site of tumor origin. Best management includes

A. treat with chemotherapy for metastatic colon cancer.
B. repeat the colonoscopy.
C. laparotomy to search for primary site and resect liver metastases if possible.
D. treat with empiric chemotherapy for cancer of unknown primary site.
E. advise the patient that treatment in this situation will probably be ineffective and discuss hospice referral.

Answer: A This patient has a carcinoma of unknown primary site with a colon cancer "profile": metastases to liver, typical adenocarcinoma histology, and suggestive immunohistochemical staining. In addition, the molecular tumor profiling prediction is colon cancer. Site-specific treatment of metastatic colon cancer is currently associated with median survival time of 20 to 24 months, which is substantially better than the expected 9- to 11-month median survival time with empiric CUP chemotherapy and long enough that hospice referral is not appropriate. A negative colonoscopy is frequent in patients with this syndrome; repeat colonoscopy will not change treatment, even if a small primary colon cancer is found. Resection of six liver metastases is very unlikely to be beneficial even if technically feasible. However, resection of residual hepatic metastases may have a role in patients who have an excellent response to first-line chemotherapy.

5. A 45-year-old premenopausal woman presents with a new mass in her right groin. She feels well and denies any recent urinary or gynecologic symptoms. Physical examination reveals a firm, nontender 2.5-cm right inguinal lymph node, with an adjacent 1.5-cm node, also firm and nontender. CT scans of the chest, abdomen, and pelvis reveal no additional findings. A needle biopsy of the largest inguinal lymph node shows metastatic squamous cell cancer. In this patient, procedures indicated in a search for the primary site may include all except

A. pelvic examination with colposcopy and biopsy of suspicious lesions.
B. cystoscopy.
C. PET scan.
D. breast MRI.
E. anoscopy.

Answer: D Female patients who present with squamous cell cancer metastatic to inguinal nodes usually have a primary site that is detectable in the vagina, cervix, bladder, or anorectal area. Evaluation should search for these primary sites. If metastases from any of these primary sites are limited to ipsilateral inguinal lymph nodes, appropriate treatment is sometimes curative. This presentation (location of metastases, squamous histology) is not suggestive of breast cancer, and a breast MRI is an unnecessary procedure in this situation.

205

APPROACH TO INBORN ERRORS OF METABOLISM

OLAF A. BODAMER

DEFINITION

The term *metabolism* (Greek: *metabolé*, "change") refers to the network of chemical reactions that sustain the human organism through the digestion, absorption, transport, and utilization of nutrients. Inborn errors of metabolism are genetic disorders that affect these intrinsic metabolic pathways through deficiencies of enzymes, membrane transporter proteins, signaling peptides, or structural proteins. The resulting clinical phenotype follows a spectrum of different organ manifestations that may be progressive, fluctuating, or stationary in nature and may be manifested at any age. Any inborn error of metabolism can principally be manifested during adolescence or adulthood, although severe presentations are typically recognized during infancy and childhood.

HISTORY

Archibald Garrod pioneered the field of inborn errors of metabolism after recognizing alkaptonuria as one of the first metabolic conditions due to homozygosity of mutant alleles in 1902. He had the foresight to recognize the autosomal recessive inheritance of additional inborn errors of metabolism, including cystinuria, pentosuria, and albinism, and to speculate about "chemical individuality" as one of the driving forces of selection and evolution. However, it was not until the early 1950s that the deficiency of homogentisate 1,2-dioxygenase (HGD) was recognized as the underlying cause of alkaptonuria, and it took many more years to identify pathogenic mutations in the *HGD* gene.

The advent of novel analytical techniques led to the molecular and biochemical characterization of known inborn errors of metabolism and the delineation and recognition of new clinical phenotypes, some of which were previously not presumed to be due to inborn errors of metabolism. The completion of the first human genome in 2001 and the following "genomics" revolution laid the foundation for the successive identification of many additional inborn errors of metabolism through next-generation sequencing, bringing the total number of catalogued inborn errors of metabolism to more than 1500 (March 2014).

The initiation of population-based newborn screening in 1964 through Robert Guthrie resulted in its recognition as an important public health measure to prevent morbidity and mortality of inborn errors of metabolism. More than 4 million newborn infants are screened annually in the United States for 31 core conditions, including mostly inborn errors of metabolism. As a consequence, approximately 12,500 newborn infants are diagnosed each year through newborn screening. Rarely, mothers with an inborn error of metabolism are diagnosed through newborn screening of their infants subsequent to placental transfer of pathognomonic metabolites.

EPIDEMIOLOGY

Inborn errors of metabolism occur in all populations, although their incidence and prevalence rates may vary considerably because of differences in carrier rates. These variations are readily explained by the presence of founder mutations, for example, in individuals of Ashkenazi Jewish or Amish ancestry, or by an increased rate of parental consanguinity that leads to a relative increase in mutant allele frequency (Table 205-1). Knowledge of the increased carrier frequencies is instrumental for preconception genetic counseling and targeted carrier screening.

PATHOBIOLOGY

The complexity of human metabolism and its spatial relationship with the human proteome, genome, and methylome are poorly understood. Naturally occurring variants in human nucleotide sequences may or may not result in variation of amino acid sequences in peptides and proteins. It is now well established from whole exome and genome sequencing that individuals may carry in excess of 10,000 nucleotide variants; most variants are silent, *single-nucleotide polymorphic variants*. Up to 4% of variants may be pathogenic in either recessive or dominant genes. These variants in particular will lead to functional changes in proteins that may render the affected individual susceptible to disease, increase the risk for undesired side effects on treatment with certain drugs, or increase the risk for genetic conditions in future generations.

Variation of human peptides and proteins is not merely explained through genomic sequence variation. Post-transcriptional alternative splicing will generate tissue-specific isoforms of proteins that are adapted to their functional needs through post-translational modification and conformational plasticity.

Genetics

Inborn errors of metabolism are monogenic conditions that follow autosomal recessive or dominant, X-linked recessive or dominant, or mitochondrial inheritance patterns. Of note is the existence of genetic or environmental modifiers that contribute to the interindividual and intrafamilial variability of phenotypic expression, although for most inborn errors of metabolism, these modifiers remain elusive. In case of mitochondrial inheritance, heteroplasmy (the random distribution and expression of mitochondrial mutations in different organs) may explain by itself the striking variability of clinical symptoms in mitochondrial conditions. The concept of synergistic heterozygosity (i.e., heterozygosity for pathogenic mutations affecting different enzymes simultaneously within the same pathway) may explain why some individuals with symptoms reminiscent of inborn errors of metabolism are not formally diagnosed.

Pathophysiology

The severity of any given inborn error of metabolism depends on the degree of enzyme deficiency and the complex interaction of the underlying pathogenic mutations, genetic modifiers, and environment. Hypomorphic mutations may not lead to overt disease until adulthood, whereas severe mutations in the same gene may lead to infantile-onset disease associated with significant morbidity and mortality. The underlying pathophysiologic mechanisms may contribute individually or in combination to the disease state (Table 205-2). Complete blockage of a catabolic pathway may result in accumulation of toxic substrates, activation of secondary minor pathways, or a relative shortage of downstream products. As a consequence, different organs may be affected by the same metabolic defect. An example is homocystinuria due to mutations in the gene for cystathionine β-synthase, which causes lens dislocation and intellectual disabilities and increases the risk for cardiovascular disease. Accumulation of homocysteine contributes to the vascular risk, whereas lack of the downstream product cysteine is an important factor in the dislocation of the lens through loosening of the zonular fibers (Table 205-3).

Clinical Phenotype

Inborn errors of metabolism typically affect multiple organs and, in more than 50% of cases, the central and peripheral nervous systems and muscles. One or more organ manifestations may dominate the clinical phenotype, although oligosymptomatic cases may occur. The clinical phenotype represents a continuous clinical spectrum ranging from the severe end, presenting during infancy, to the mild end of the spectrum, presenting during adolescence or adulthood. Some affected individuals may never come to medical attention because of almost complete absence of symptoms or atypical presentation. Recent data from newborn screening programs suggest much higher incidence rates for some inborn errors of metabolism due to the detection of a high rate of mild cases in which disease-related signs or symptoms may never develop. Some clinical signs are pathognomonic for an inborn error of metabolism, whereas others should raise the suspicion for the presence of an inborn error of metabolism (Table 205-4).

Classification

Inborn errors of metabolism can be classified on the basis of the underlying pathomechanism (see Table 205-2), the nature or localization of the protein involved (see Table 205-3), or the clinical phenotype (see Table 205-4). The most logical classification is based on the nature or localization of the affected protein and pathway.

INBORN ERRORS OF METABOLISM

Disorders of Protein Metabolism

These conditions are due to cytosolic or mitochondrial enzyme deficiencies affecting mostly catabolic pathways (Table 205-5). Disorders of protein

TABLE 205-1 INCIDENCE OF INBORN ERRORS OF METABOLISM

DISORDER	GENE	INCIDENCE*	CARRIER RATE	POPULATION
Familial hypercholesterolemia	*LDLR*	1 : 500	1 : 500	All
Phenylketonuria	*PAH*	1 : 4000 <1 : 120,000 1 : 15,000	1 : 32 <1 : 173 1 : 61	Ireland Finland, Japan United States
Gaucher's disease	*GBA*	1 : 20,000 1 : 450	1 : 71 1 : 11	United States† Ashkenazi Jews
Canavan's disease	*ASPA*	Unknown 1 : 6000	Unknown 1 : 39	United States Ashkenazi Jews
Glycogen storage disease IA	*G6PC*	1 : 100,000 1 : 1225	1 : 158 1 : 18	United States Ashkenazi Jews
Mucolipidosis IV	*MCOLN1*	Unknown 1 : 3000	Unknown 1 : 27	United States Ashkenazi Jews
Niemann-Pick disease A	*SMPD1*	<1 : 250,000 1 : 40,000	<1 : 250 1 : 100	United States Ashkenazi Jews
Tay-Sachs disease	*HEXA*	1 : 300,000 1 : 3500	1 : 274 1 : 30	United States Ashkenazi Jews

*Per live births.
†Includes Ashkenazi Jews.

TABLE 205-2 PATHOPHYSIOLOGIC MECHANISMS IN INBORN ERRORS OF METABOLISM

MECHANISM	DISORDER
Accumulation of toxic substrates through primary blockage of catabolic pathway	Organic acidopathies (MMA, PA) MSUD, tyrosinemia type I
Accumulation of nontoxic macromolecules through blockage of catabolic pathway	Lysosomal storage disorders (MPS)
Energy failure through primary blockage of pathway relevant for ATP synthesis	Fatty acid oxidation defects Glycogen storage disorder types I and III Respiratory chain enzyme deficiencies
Impairment of post-translational glycosylation	Congenital disorders of glycosylation
Deficiency of end product through primary blockage of anabolic pathway	Albinism, orotic aciduria, scurvy, disorders of creatine metabolism
Lack of detoxification through primary blockage of catabolic pathway	Urea cycle defects

ATP = adenosine triphosphate; MMA = methylmalonic aciduria; MPS = mucopolysaccharidoses; MSUD = maple syrup urine disease; PA = propionic aciduria.

TABLE 205-3 SELECTED PROTEINS AND INBORN ERRORS OF METABOLISM

PROTEIN	LOCALIZATION	FUNCTION	DISORDER
Enzyme	Cytosolic	Urea production	Urea cycle defects
Enzyme	Mitochondrial	Fatty acid oxidation	Disorders of fatty acid oxidation
		ATP synthesis	Respiratory chain enzyme deficiencies
Transporter	Cellular membrane	Creatine transport	Creatine deficiency
		Folic acid transport	Folic acid deficiency
	Mitochondrial membrane	Carnitine transport	CACT deficiency
Transporter	Blood	Transport of cobalamin	Cobalamin deficiency

ATP = adenosine triphosphate; CACT = carnitine acylcarnitine translocase.

TABLE 205-4 CHARACTERISTIC SIGNS OF INBORN ERRORS OF METABOLISM

ORGAN	CLINICAL SIGN	DISORDER
Eye (cornea)	Cornea verticillata	Fabry's disease
Skeletal system	Ochronosis, black urine	Alkaptonuria
Connective tissue	Carpal tunnel syndrome	MPS I, II, VI, and VII
Central nervous system	Ataxia	Respiratory chain enzyme deficiency
Muscle	Hypotonia	Pompe's disease, GSD V Disorders of creatine metabolism
Liver	Hepato(-spleno)megaly Fibrosis, cirrhosis	MPS I, II, VI, and VII GSD I, III GSD IV, IXb/c, LAL deficiency
Kidney	Renal insufficiency	Cystinosis, Fabry's disease
Skin	Angiokeratomas	Fabry's disease

GSD = glycogen storage disease; LAL = lysosomal acid lipase; MPS = mucopolysaccharidosis.

metabolism may also be due to defects of plasma membrane protein transport (Table 205-6).

PHENYLKETONURIA

Phenylalanine is an essential amino acid important for growth and production of thyroid hormone, neurotransmitters, and melanin. Phenylketonuria (PKU) is caused by deficiency of tetrahydrobiopterin (BH4)–dependent phenylalanine hydroxylase that catalyzes the conversion of phenylalanine to tyrosine. PKU may also be caused by deficiency of enzymes that are required for BH4 synthesis. PKU is one of the "traditional" inborn errors of metabolism and the first to be included in newborn screening programs, demonstrating that early diagnosis and continued therapy consistently result in normal intellectual development.[1]

TYROSINEMIA

There are three types of tyrosinemia due to different enzyme deficiencies within the catabolic pathway of tyrosine. Deficiency of fumarylacetoacetase in tyrosinemia type I leads to production of toxic byproducts of a minor pathway. These byproducts are primarily toxic to the liver, resulting in acute liver failure and renal Fanconi's tubulopathy if it is left untreated. Succinylacetone, one of the toxic byproducts, serves as a diagnostic metabolite. Tyrosinemia type II results in significant elevations of tyrosine concentration in all tissues, leading to painful corneal lesions, hyperkeratosis of the palms and soles, and mild intellectual disability.

MAPLE SYRUP URINE DISEASE

The clinical phenotype in maple syrup urine disease (MSUD) is due to accumulation of toxic compounds including oxoisocaproic acid resulting from deficiency of the branched-chain α-keto acid dehydrogenase complex. This multienzyme complex, similar to the pyruvate dehydrogenase complex, consists of four different subunits, E_{1a}, E_{1b}, E_2, and E_3. Deficiency of any subunit or a combination of subunits will cause MSUD. Mild forms of MSUD may

TABLE 205-5 DISORDERS OF PROTEIN METABOLISM

DISORDER	ENZYME DEFECT	METABOLITES	CLINICAL PHENOTYPE
Argininemia	Arginase	Arginine	Hyperammonemia, neurologic disease
Argininosuccinic aciduria	Argininosuccinate lyase	Argininosuccinate*	Hyperammonemia, liver cirrhosis
Citrullinemia	Argininosuccinate synthetase	Citrulline, orotic acid*	Hyperammonemia, liver cirrhosis
Homocystinuria	Cystathionine β-synthase	Homocysteine, methionine	Marfanoid habitus, intellectual disability, lens dislocation
Maple syrup urine disease	Branched-chain α-keto acid dehydrogenase complex	Alloisoleucine,* leucine, valine, isoleucine	Encephalopathy, ataxia, metabolic decompensation
Ornithine transcarbamylase deficiency	Ornithine transcarbamylase	Orotic acid,* ornithine, arginine	Severe hyperammonemia, X-linked inheritance
Phenylketonuria	Phenylalanine hydroxylase	Phenylalanine	Intellectual disability,† seizures†
Tyrosinemia type I	Fumarylacetoacetase	Succinylacetone,* tyrosine	Acute liver failure, tubulopathy
Tyrosinemia type II	Tyrosine aminotransferase	Tyrosine, phenylalanine	Corneal lesions, hyperkeratosis of the skin, mild intellectual disability

*Diagnostic compound.
†If untreated.

TABLE 205-6 INBORN ERRORS OF METABOLISM CAUSED BY DEFECTS IN PLASMA MEMBRANE TRANSPORTER PROTEINS

DISORDER	TISSUE AFFECTED	SUBSTRATE	MODE OF INHERITANCE	CLINICAL PHENOTYPE
Cobalamin malabsorption	Ileum	Cobalamin	Autosomal recessive	Pernicious anemia
Carnitine uptake deficiency	Kidney, small intestine	Carnitine	Autosomal recessive	Hypoglycemia, hypotonia
Cystic fibrosis	Apical epithelia	Chloride	Autosomal recessive	Pneumonia, ileus
Cystinuria	Kidney, small intestine	Cystine, lysine, arginine, ornithine	Autosomal recessive	Urolithiasis
Familial hypophosphatemic rickets	Kidney, small intestine	Phosphate	X-linked dominant	Rickets
Folate deficiency	Lymphocyte, erythrocyte	Methyltetrahydrofolate	Autosomal recessive	Aplastic anemia
GLUT1 deficiency	Blood-brain barrier, erythrocyte	Glucose	Autosomal recessive	Seizures, microcephaly
Hartnup's disease	Kidney, small intestine	Neutral amino acids	Autosomal recessive	Nicotinic acid deficiency
Hereditary renal hypouricemia	Kidney	Uric acid	Autosomal recessive	Urolithiasis
Hereditary spherocytosis	Erythrocyte	Sodium	Autosomal dominant or recessive	Hemolytic anemia
Iminoglycinuria	Kidney, small intestine	Glycine, proline, hydroxyproline	Autosomal recessive	Benign, pancreatitis?
Isolated lysinuria	Kidney, small intestine	Lysine	Autosomal recessive	Growth failure, seizures
Lysinuric protein intolerance	Kidney, liver, intestine	Lysine, arginine, ornithine	Autosomal recessive	Growth failure, intellectual disability, hyperammonemia
Renal glycosuria	Kidney	Glucose	Autosomal recessive	Benign
Renal tubular acidosis type 1	Distal renal tubule	H^+ secretion, citrate, calcium	Autosomal dominant	Hypokalemia, growth failure, nephrocalcinosis
Renal tubular acidosis type 2	Proximal renal tubule	Bicarbonate	Autosomal recessive	Hyperchloremic metabolic acidosis

Modified from Elsas L II: Approach to inborn errors of metabolism. In: Goldman L, Schafer AI: Goldman's Cecil Medicine. 24th ed. Philadelphia: Elsevier Saunders; 2012:1340-1346.

present with fluctuating neurologic symptoms as well as episodes of ketoacidosis.

Disorders of the Urea Cycle

The role of the urea cycle is to convert ammonium as a byproduct of amino acid metabolism to nontoxic urea that is readily excreted in urine and to synthesize arginine and ornithine. Arginine is an important precursor for the nitric oxide pathway, and it is substrate for creatine/creatine phosphate synthesis. Several mitochondrial and cytosolic enzymes as well as transporters are required for the function of the urea cycle. Individuals with any of the disorders of the urea cycle are at risk for hyperammonemia during catabolic episodes when there is an increased rate of protein breakdown.[2] All conditions are inherited as an autosomal recessive trait with the exception of ornithine transcarbamylase deficiency, which is inherited as an X-linked recessive trait (see Table 205-5).

Disorders of Membrane Transport

Many inborn errors are due to mutations in genes encoding plasma membrane transporter proteins (see Table 205-6). These transporter proteins are responsible for the active transport of small molecules including vitamins, carnitine, glucose, amino acids, and electrolytes across different membranes. The blood-brain barrier plays a particular role in the protection and transport of nutrients to and from the central nervous system. GLUT1 facilitates the transport of glucose across many cellular membranes, including the blood-brain barrier. Deficiency of GLUT1 results in low cerebrospinal fluid glucose concentrations despite normal glucose levels in the blood, a circumstance that is used diagnostically. Affected individuals develop seizures that may be treated with ketone bodies. Other glucose transporters GLUT 2, 3, and 4 are expressed in the liver, kidneys, neurons, and skeletal and heart muscles, with more than one glucose transporter expressed by most cells.

Organic Acidurias

Organic acidurias are disorders due to mitochondrial enzyme deficiencies and accumulation of potentially toxic substrates, activation of alternative pathways, and lack of downstream products. Although the typical clinical presentation is during infancy or childhood, adult cases with mild or atypical symptoms have been reported in the medical literature. Although individuals with organic acidurias are at risk for metabolic decompensation during catabolic episodes, this risk is somewhat lower during adulthood (Table 205-7).

TABLE 205-7 ORGANIC ACIDURIAS

DISORDER	ENZYME DEFECT	INHERITANCE	URINE METABOLITES
Glutaric aciduria type I	Glutaryl-CoA dehydrogenase	Autosomal recessive	3-Hydroxyglutaric acid, glutaric acid
Holocarboxylase synthetase deficiency	Holocarboxylase synthetase	Autosomal recessive	β-Hydroxyisovaleric acid, β-methylcrotonylglycine, β-hydroxypropionic acid, 3-methylcitrate
Isobutyric aciduria	Isobutyryl-CoA dehydrogenase	Autosomal recessive	Isobutyric acid
Isovaleric aciduria	Isovaleryl-CoA dehydrogenase	Autosomal recessive	Isovaleric acid
Methylmalonic aciduria	Methylmalonyl-CoA dehydrogenase	Autosomal recessive	Methylmalonic acid
Mevalonic aciduria	Mevalonate kinase	Autosomal recessive	Mevalonic acid
Propionic aciduria	Propionyl-CoA carboxylase	Autosomal recessive	3-Methylcitrate, propionic acid, 3-hydroxypropionic acid

TABLE 205-8 INBORN ERRORS OF PEROXISOMES

DISORDER	GENE DEFECT	METABOLITES
Zellweger's spectrum	*PEX 1,2,3,5,6,10,12,13,14,16,19,26; DLP1*	Very long chain fatty acids, bile acids, phytanic and pipecolic acids, erythrocyte plasmalogen
Rhizomelic chondrodysplasia punctate type I	*PEX 7*	Phytanic acid, erythrocyte plasmalogen
X-linked adrenoleukodystrophy	*ABCD1*	Very long chain fatty acids
Acyl-CoA oxidase deficiency	*ACOX1*	Very long chain fatty acids
Methylacyl-CoA racemase deficiency	*AMACR*	Bile acid intermediates, phytanic and pristanic acids
Dihydroxyacetone phosphate acyltransferase deficiency	*GNPAT*	Erythrocyte plasmalogen
Refsum's disease	*PHYH*	Phytanic acid
Hyperoxaluria type I	*AGXT*	Oxalic acid

Lysosomal Storage Disorders

The lysosomal storage disorders comprise a heterogeneous group of more than 50 distinct disorders due to genetic defects in lysosomal enzymes and membrane proteins or transporters, resulting in lysosomal accumulation of specific substrates. The accumulation in tissues and organs is progressive, ultimately causing deterioration of cellular and tissue function. Many lysosomal disorders affect the central nervous system, and most patients have a decreased lifespan and significant morbidity. Lysosomal storage disorders may be categorized on the basis of the type of substrate stored. Disorders of glycosaminoglycan metabolism include the mucopolysaccharidoses (MPS): MPS I, Hurler's syndrome, Scheie's syndrome; MPS II, Hunter's syndrome; MPS IIIA-D, Sanfilippo's syndrome A-D; MPS IVA, IVB, Morquio's syndrome A and B; MPS VI, Maroteaux-Lamy syndrome; MPS VII, Sly's syndrome.[3] Disorders of ganglioside metabolism include Fabry's disease, Gaucher's disease, Niemann-Pick disease, Tay-Sachs disease, I-cell disease, fucosidosis, mannosidosis, sialidosis, and aspartylglycosaminuria. Danon's and Pompe's diseases are two lysosomal storage disorders resulting in storage of glycogen in different types of muscle cells.

Peroxisomal Disorders

Peroxisomes are cell organelles that are metabolically very active. They are the site of plasmalogen, cholesterol, and bile acid synthesis. Additional pathways include gluconeogenesis from amino acids, the formation of oxalic acid, and the breakdown of hydrogen peroxide, purines, polyamines, phytanic acid, pipecolic acid, and very long chain fatty acids. Disorders of peroxisomal metabolism may result in elevated very long chain fatty acids and phytanic and pipecolic acids but low plasmalogen concentrations (Table 205-8).

Disorders of Respiratory Chain Complexes

Thirteen proteins of the five different mitochondrial respiratory chain enzyme complexes are encoded in the mitochondrial gene, with the remainder being nuclear encoded. Complex II is entirely encoded by the nuclear genome. Disorders that affect any or a combination of the mitochondrial complexes will result in a broad clinical phenotype affecting multiple organs with marked intrafamilial and interfamilial variability due to heteroplasmy in case of mitochondrial inheritance.

Congenital Disorders of Glycosylation

Congenital disorders of glycosylation (CDG) are due to defects in proteins that are involved in post-translational glycosylation of peptides and proteins. More than 30 different forms of CDG with a broad clinical spectrum affecting multiple organs are recognized. The traditional classification of CDG disorders is based on pathophysiologic considerations, whereas the newer classification uses the underlying molecular defects. Secondary changes in glycosylation pattern may be observed in galactosemia and hereditary fructose intolerance. The most common form of CDG, CDG type Ia, is due to deficiency in the *PMM2* gene encoding phosphomannomutase. Most of the patients with CDG Ia present during infancy and childhood with developmental delay, severe infections, bleeding diathesis, and liver impairment, although older patients with milder phenotypes have been described.[4]

DIAGNOSIS

Biochemical and Molecular Testing

The path to a diagnosis of an inborn error of metabolism begins with ascertainment of the medical and family history as well as an in-depth clinical evaluation. The majority of inborn errors of metabolism can be diagnosed through analysis of small molecules (metabolites, peptides, and hormones) in appropriate body fluids (serum, whole blood, urine, and cerebrospinal fluid), followed by enzyme testing in tissues (dried whole blood, lymphocytes, leukocytes, fibroblasts, and muscle tissue). Tissue biopsies for histology and histochemistry are still of value in some cases. Selected metabolic tests for diagnosis of inborn errors of metabolism include analysis of amino acids in plasma, dried blood spots, urine, and cerebrospinal fluid; analysis of acylcarnitine species in plasma and dried blood spots; analysis of total and free carnitine in plasma and urine; analysis of succinylacetone in dried blood spots and urine; and analysis of orotic acid in plasma and urine.

Molecular confirmation is warranted for prediction of phenotype and is a prerequisite for family planning, including preimplantation diagnosis, prenatal testing, and carrier testing for the partner (Table 205-9).

Next-Generation Sequencing

The completion of the first draft sequence of the human genome in 2001 laid the foundation for a new era that came to fruition after the advent of next-generation sequencing technologies. Next-generation sequencing enables the rapid and accurate sequencing of whole human genomes and exomes that represent the 1 to 2% of the genome that is translated into proteins. The continuous refinement of next-generation sequencing technologies has led to a rapid decline of sequencing cost, thereby facilitating the sequencing of tens of thousands of individuals in both research and clinical settings.

Clinical whole exome or genome sequencing in CLIA- and CAP-accredited laboratories may aid in the diagnosis of rare mendelian disorders including

TABLE 205-9 DIAGNOSTIC TESTS FOR INBORN ERRORS OF METABOLISM

METABOLITE	BIOLOGIC MATRIX	METHOD	DISORDER
Amino acids	Plasma, urine, CSF	HPLC, MS-MS	Disorders of amino acid metabolism including PKU, MSUD, tyrosinemia, urea cycle defects, lysinuric protein intolerance
Organic acids	Urine, CSF	GC-MS	Organic acidopathies including MMA, PA, IVA Lactic acidosis, mitochondrial disorders including disorders of the Krebs cycle, respiratory chain enzymes, fatty acid oxidation defects
Acylcarnitine species	Plasma, DBS	MS-MS	Fatty acid oxidation defects, organic acidopathies
Total/free carnitine	Plasma, DBS, urine	MS-MS	Carnitine transporter deficiency, CPT I/II deficiencies (in conjunction with acylcarnitine species), secondary carnitine deficiency
Orotic acid	Plasma, urine	GC-MS, MS-MS	Orotic aciduria, OTC deficiency, citrullinemia type I
Succinylacetone	Plasma, DBS, urine	GC-MS	Tyrosinemia type I
Glycosaminoglycans	Urine	TLC, MS-MS	Mucopolysaccharidoses

CPT I/II = carnitine palmitoyltransferase I/II; CSF = cerebrospinal fluid; DBS = dried blood spots; GC-MS = gas chromatography–mass spectrometry; HPLC = high-pressure liquid chromatography; IVA = isovaleric aciduria; MMA = methylmalonic aciduria; MS-MS = tandem mass spectrometry; MSUD = maple syrup urine disease; OTC = ornithine transcarbamylase; PA = propionic aciduria; PKU = phenylketonuria; TLC = thin-layer chromatography.

TABLE 205-10 THERAPEUTIC STRATEGIES FOR INBORN ERRORS OF METABOLISM

LEVEL	THERAPEUTIC APPROACH	DISORDER
Gene	Solid organ transplantation	Urea cycle defects, tyrosinemia type I
	Stem cell transplantation	Adrenoleukodystrophy, MPS I, GSD IA (experimental)
	Gene therapy	Pompe's disease (phase I/II trial)
	Read-through therapy	Duchenne's muscular dystrophy (phase III), cystic fibrosis (phase III)
Enzyme	Recombinant enzyme infusion	Gaucher's, Pompe's, and Fabry's diseases; MPS I, II, IVA, VI Phenylketonuria (phase III)
	Chaperone	Fabry's and Pompe's diseases (phase III)
Substrate	Substrate reduction	Phenylketonuria, maple syrup urine disease
	Substrate inhibition	Gaucher's disease, Tay-Sachs disease

GSD IA = glycogen storage disease type IA; MPS I = mucopolysaccharidosis type I.

inborn errors of metabolism, provided all other diagnostic avenues are exhausted. Although sequencing may be relatively straightforward and inexpensive, analytical challenges remain. An individual may carry a large number of gene variants requiring a dedicated analytical pipeline to identify the pathogenic variants of interest. The diagnostic yield may be as high as 25% in preselected cases.[5]

TREATMENT

The individual treatment strategy follows these general principles: enhancement of enzyme activity through cofactor administration; stabilization of enzyme structure through chaperone therapy; enzyme replacement therapy; reduction of substrate through dietary intervention; substrate inhibition through blockage of the reverse enzyme reaction; replacement of the affected organ (e.g., liver); and stem cell transplantation or therapy.[6,7] Other therapeutic approaches including "read-through" and gene therapies are in clinical trials. Read-through therapy refers to a small molecule (ataluren) that renders ribosomes less sensitive to premature stop codons and allows read-through. Ataluren has been tested in individuals with Duchenne's muscular dystrophy and cystic fibrosis due to nonsense mutations (Table 205-10).

The choice of therapy is guided by the underlying diagnosis, but additional factors warrant consideration. A curative approach may be preferred whenever possible, although this may be rarely an option in adults with inborn errors of metabolism, as in the case of bone marrow or stem cell transplantation in metachromatic leukodystrophy or X-linked adrenoleukodystrophy. Therapeutic agents that address central nervous system manifestations have to cross the blood-brain barrier to be effective, thereby limiting the use of larger molecules, including enzymes, for treatment of neurologic manifestations.

Individuals with inborn errors of metabolism should always be managed by a multidisciplinary team of biochemical geneticists, internists, and genetic counselors at a tertiary center with significant expertise in the management of these disorders. Ideally, a biochemical genetics laboratory to facilitate immediate sample testing should be on site.

Genetic Counseling

Genetic counseling through board-certified genetic counselors is an integral part of the evaluation and management of patients with any familial condition including complex inborn errors of metabolism. A three-generation family history and thorough medical history are the prerequisites for a focused clinical and diagnostic evaluation. Genetic counseling communicates the limitations and implications of genetic testing and associated test results to the patient and the family at large.

Enzyme Therapy

The therapeutic goal of enzyme therapy is to increase endogenous enzyme activity. Individuals with cofactor (vitamin)–responsive enzyme deficiencies may benefit from supraphysiologic doses of the respective vitamin. An example is BH4-responsive PKU due to mutations in the *PAH* gene affecting the binding site of phenylalanine hydroxylase (PAH); 20 mg/kg of BH4 will result in stabilization of the PAH through a chaperone-like effect and, as a consequence, increased PAH activity. Patients with BH4-responsive PKU will experience improved phenylalanine tolerance and metabolic control when taking BH4 supplementation.

Enzyme replacement therapy has been available for treatment of Gaucher's disease for more than 15 years. Initially, glucocerebrosidase, the enzyme deficient in Gaucher's disease, had been purified from human placentas and administered intravenously to affected patients. More recently, glucocerebrosidase has been overexpressed in Chinese hamster ovary cells and produced in large quantities in bioreactors. Recombinant enzyme replacement therapies are now available for Pompe's disease (α-glucosidase), Fabry's disease (α-galactosidase), mucopolysaccharidosis type I (α-iduronidase), mucopolysaccharidosis type II (α-iduronate sulfatase), mucopolysaccharidosis type IVA (galactosamine-6-sulfatase), and mucopolysaccharidosis type VI (arylsulfatase B). Additional enzyme therapies are currently in different phases of clinical trials (see Table 205-10).

Nutritional Therapy

The therapeutic goal of nutritional therapy is the correction of the metabolic imbalance through reduced substrate accumulation, promoting protein synthesis through anabolism and the prevention of episodes of metabolic decompensation. In addition, supplementation of a reduced product may be needed. An example is the therapy for PKU. The mainstay of its therapy is the reduction of phenylalanine intake through low-protein food and the simultaneous supplementation of phenylalanine-free amino acids for sustained growth and development. This regimen will reduce the phenylalanine levels in plasma and, most important, in the brain to nontoxic, nearly normal levels that facilitate age-appropriate intellectual development. On occasion, tyrosine supplementation is needed when tyrosine levels are low. Similar dietary strategies apply to other disorders of amino acid metabolism and organic acidopathies, although natural protein restriction may be more pronounced to reduce the risk for metabolic decompensation during a catabolic episode (intercurrent illness).

Another approach to reduce substrate accumulation is through inhibition of the reverse enzyme reaction that leads to substrate synthesis. This concept has been shown to be effective in reducing glucosylceramide in mild to moderate Gaucher's disease type I. Miglustat is an inhibitor of the enzyme glucosylceramide synthase that catalyzes the reverse direction of glucocerebrosidase, the enzyme that is deficient in Gaucher's disease type I.

Individuals with inborn errors of metabolism at risk for metabolic decompensation, such as urea cycle disorders, organic acidopathies, and disorders of fatty acid oxidation, should always carry an "emergency letter" detailing the diagnosis, symptoms of decompensation, emergency treatment, and contact information of the tertiary metabolic center.

Vitamin Therapy

There are a number of inborn errors of metabolism that affect transport or metabolism of vitamins. These conditions typically benefit from supraphysiologic doses of vitamins. An example is thiamine-responsive megaloblastic anemia, which is due to mutations in the thiamine transporter gene *SLC19A2*. Affected individuals develop sensorineural deafness, vision loss, diabetes mellitus, and megaloblastic anemia. Diabetes and anemia respond to high doses of thiamine. Other examples include disorders of vitamin B_{12} (cobalamin) absorption, transport, and metabolism and disorders of biotin and folic acid metabolism.

Organ, Bone Marrow, and Stem Cell Transplantation

Liver transplantation has been done in patients with tyrosinemia type I, urea cycle defects, methylmalonic aciduria, propionic aciduria, lysosomal lipase deficiency, and glycogen storage diseases affecting the liver. The effect of liver transplantation in these conditions is two-fold. First, metabolic control will be improved in those conditions in which the diseased liver is the main contributor to the overall lack of sufficient metabolic control. However, the intrinsic defect will not be corrected elsewhere after liver transplantation, and an affected individual may continue to have significant neurologic disease in methylmalonic aciduria, for example. Second, liver function will be restored in those conditions that lead to chronic liver disease, including liver fibrosis and cirrhosis, or in which there is a significant risk for malignant transformation. Kidney transplantation may be required in conditions that affect kidney function, as is the case in methylmalonic aciduria or cystinosis.

Bone marrow or stem cell transplantation has limited benefits in patients with inborn errors of metabolism. Examples include presymptomatic bone marrow or stem cell transplantation in severe mucopolysaccharidosis type I, metachromatic leukodystrophy, and X-linked adrenoleukodystrophy. Stem cell therapy is currently under investigation for glycogen storage disease type I.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Camp KM, Parisi MA, Acosta PB, et al. Phenylketonuria Scientific Review Conference: state of the science and future research needs. *Mol Genet Metab*. 2014;112:87-122.
2. Butterworth RF. Pathophysiology of brain dysfunction in hyperammonemic syndromes: the many faces of glutamine. *Mol Genet Metab*. 2014;113:113-117.
3. Archer LD, Langford-Smith KJ, Bigger BW, et al. Mucopolysaccharide diseases: a complex interplay between neuroinflammation, microglial activation and adaptive immunity. *J Inherit Metab Dis*. 2014;37:1-12.
4. Wolthuis DF, Janssen MC, Cassiman D, et al. Defining the phenotype and diagnostic considerations in adults with congenital disorders of N-linked glycosylation. *Expert Rev Mol Diagn*. 2014;14:217-224.
5. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013;369:1502-1511.
6. Sirrs SM, Lehman A, Stockler S, et al. Treatable inborn errors of metabolism causing neurological symptoms in adults. *Mol Genet Metab*. 2013;110:431-438.
7. Sokal EM. Treating inborn errors of liver metabolism with stem cells: current clinical development. *J Inherit Metab Dis*. 2014;37:535-539.

REVIEW QUESTIONS

1. Which of the following statements is correct regarding disorders caused by inborn errors of metabolism?

A. They present clinically in neonates and early childhood because the metabolic pathways disrupted in these disorders affect vital organ functions.
B. They are typically autosomal recessive diseases caused by founder mutations and parental consanguinity.
C. They can affect multiple organs with the exception of the central and peripheral nervous systems.
D. They may present with pathognomonic clinical signs.
E. They are characterized by mitochondrial inheritance.

Answer: D Inborn errors of metabolism may exhibit clinical signs that are diagnostically pathognomonic, such as ochronosis and black urine in alkaptonuria or angiokeratomas and cornea verticillata in Fabry's disease (see Table 205-4). The clinical phenotypes of inborn errors of metabolism range widely from severe disease presenting in infancy to mild symptoms and signs recognized only in adulthood as well as many cases that are asymptomatic and consistent with normal life expectancy. They are monogenic conditions that follow autosomal recessive or dominant, X-linked recessive or dominant, or mitochondrial inheritance patterns. More than 50% of clinical overt cases affect the central or peripheral nervous system. (See section on pathobiology.)

2. Which of the following therapeutic modalities is *not* currently in use for the clinical management of patients with inborn errors of metabolism?

A. Intravenous enzyme replacement therapy
B. Stem cell therapy
C. Liver transplantation
D. Nutritional therapy
E. None of the above has been shown to be effective in any inborn errors of metabolism.

Answer: B Examples of effective treatments include intravenous enzyme (glucocerebrosidase) replacement therapy (A) for Gaucher's disease; liver transplantation (C) for tyrosinemia type I, urea cycle defects, methylmalonic aciduria, propionic aciduria, lysosomal lipase deficiency, and glycogen storage diseases; and nutritional therapy (D) for phenylketonuria by reduction in phenylalanine intake through low-protein food along with simultaneous supplementation with phenylalanine-free amino acids. Stem cell therapy is currently under investigation for glycogen storage disease type I. (See section on treatment.)

3. Which of the following is *not* currently considered first-line diagnostic testing for inborn errors of metabolism?

A. Whole exome sequencing
B. Biochemical analysis of metabolites in body fluids
C. Enzyme testing in tissues
D. Tissue biopsies for histology and histochemistry
E. Prenatal and carrier testing

Answer: A Most inborn errors of metabolism can be diagnosed by analysis of small-molecule metabolites in appropriate body fluids like whole blood, serum, urine, and cerebrospinal fluid (B). This is usually followed by appropriate enzyme testing in tissues like dried whole blood, leukocytes, fibroblasts, and muscle tissue (C). Tissue biopsies for histology and histochemistry (D) are still of value in some cases. Molecular confirmation is available for prediction of phenotype and is a prerequisite for family planning, including preimplantation diagnosis, prenatal testing, and carrier testing for the partner (E). Clinical whole genome or exome sequencing in CLIA- and CAP-accredited laboratories may aid in the diagnosis of rare mendelian disorders like inborn errors of metabolism, but analytical challenges remain and should be considered at this time only if all other diagnostic avenues are exhausted. (See section on diagnosis.)

4. A 35-year-old woman is brought to the emergency department with a history of acute-onset nausea, headaches, and stupor after a protein-rich meal. She has a history of similar events in the past. Her brother died as an infant of an encephalopathic crisis. The plasma ammonium level on admission is 400 μmol/L (reference: <40 μmol/L). What is the most appropriate initial treatment?

A. Enzyme replacement therapy
B. Liver transplantation
C. Intravenous fluids only
D. Alternative pathway therapy with intravenous sodium benzoate or sodium phenylbutyrate
E. Hydroxycobalamin IM

Answer: D The 35-year-old woman suffers from ornithine transcarbamylase (OTC) deficiency. OTC is a urea cycle enzyme that is part of the urea cycle responsible for conversion of nitrogen to urea. Deficiency of any of the five urea cycle enzymes and three transporter proteins will lead to variable elevations of plasma ammonium during episodes of intercurrent illnesses, catabolism, or high-protein meals. OTC deficiency is inherited as an X-linked recessive trait, explaining why the affected brother died as an infant during a hyperammonemic crisis. Female carriers of OTC deficiency may show symptoms only when they are exposed to a high-protein meal. The goal of therapy is to normalize ammonium levels quickly through alternative pathway therapy, hemofiltration, or hemodialysis. Outcome is poor if ammonium levels are not corrected in a timely manner.

5. A 32-year-old man from Pakistan with a seizure disorder has been recently diagnosed with an inborn error of metabolism. His family history is remarkable for an affected sister and consanguinity of the healthy parents. There is a distant male cousin with the same condition. Given this family history, what is the most likely mode of inheritance for this condition?

A. Mitochondrial inheritance
B. X-linked recessive inheritance
C. Autosomal recessive inheritance
D. Autosomal dominant inheritance with complete penetrance
E. Complex inheritance including different genetic and environmental factors

Answer: C Due to consanguinity of the healthy parents and absence of affected family members in other generations, autosomal recessive inheritance is the most likely mode of inheritance for the condition in question. Mitochondrial (maternal) inheritance of mutations in mitochondrial genes may be observed in respiratory chain disorders, with variable expressivity of disease due to heteroplasmy. X-linked recessive inheritance as in OTC deficiency results in severely affected males and mildly or unaffected carrier females. Disorders with an autosomal dominant inheritance with complete penetrance affect multiple family members at each generation. The recurrence risk is 50% for future offspring of affected individuals. Complex inheritance may be observed in cancer, glaucoma, and Alzheimer's disease because of different contributions from genetic and environmental factors.

6. A 23-year-old woman and her 25-year-old husband, both of Ashkenazi Jewish ancestry, are seen for genetic counseling regarding their risk of having a child with Gaucher's disease or other conditions. What is the most appropriate response regarding their risk?

A. Their risk of having a child with Gaucher's disease is negligible.
B. Carrier screening of both parents for Gaucher's disease through analysis of common mutations in the *GBA* gene is indicated.
C. Prenatal diagnosis is recommended during early pregnancy through analysis of glucocerebrosidase activity in amniocytes.
D. Carrier screening of both parents should be comprehensive and include an Ashkenazi Jewish disease panel and SMA (spinal muscular atrophy) for both parents and fragile X for the mother.
E. Only karyotyping for the mother is indicated.

Answer: D The carrier frequency for a number of genetic conditions including Gaucher's disease is markedly increased in individuals of Ashkenazi Jewish ancestry. Carrier screening should therefore not be restricted to Gaucher's disease but also include Tay-Sachs disease, Canavan's disease, spinal muscular atrophy, and many others for both parents as well as fragile X for the mother.

7. A 19-year-old man has been diagnosed with a disorder of fatty acid oxidation. He is at increased risk for complications due to which pathophysiologic mechanism?

A. Accumulation of toxic substrates due to blockage of pathway
B. Energy failure due to primary blockage of a pathway that is important for ATP synthesis
C. Progressive accumulation of nontoxic substrates due to a defect in a catabolic pathway
D. Defect in post-translational glycosylation
E. Blockage of an anabolic pathway

Answer: B Disorders of fatty acid oxidation lead to reduced β-oxidation of fatty acids of various chain lengths, depending on the exact enzyme deficiency. As a consequence, fewer acetyl-CoA equivalents are produced that would be available for ATP or ketone body synthesis. Clinically, patients with disorders of fatty acid oxidation have a reduced fasting tolerance and are at risk for hypoglycemia and energy failure in skeletal and cardiac muscles.

8. The diagnosis of a mucopolysaccharidosis type I is suspected in a 21-year-old woman with hepatosplenomegaly, corneal clouding, and carpal tunnel syndrome. What is the first-line diagnostic test to confirm the diagnosis?

A. Urine glycosaminoglycans
B. Urine organic acids
C. Plasma acylcarnitine profile
D. Plasma amino acids
E. Whole exome sequencing

Answer: A Mucopolysaccharidosis type I is characterized through increased excretion of dermatan and heparan sulfates (glycosaminoglycans).

206

DISORDERS OF LIPID METABOLISM

CLAY F. SEMENKOVICH

As Western lifestyles become more pervasive, disorders of lipid metabolism remain among the most common problems faced by clinicians. Ischemic heart disease, the most common cause of global disability, is caused in part by abnormal lipids. Stroke, a common cause of death in the United States, is also related to disorders of lipid metabolism. Both are likely to dominate the clinical landscape of a world where obesity and diabetes are ubiquitous. Appropriate management of common lipid disorders should be a part of the skill set for everyone who provides clinical care to adults.

COMPONENTS OF LIPID TRANSPORT

Cholesterol and Triglycerides

Cholesterol is a critical constituent of eukaryotic cell membranes and the precursor for the synthesis of steroid hormones such as cortisol, vitamin D, progestins, estradiol, and testosterone. Triglycerides carry fatty acids, nutrients that are used preferentially by muscle tissue and are especially important as an energy source in the fasting state. Because both cholesterol and triglycerides are essentially insoluble in water, the lipid transport system evolved to transport fats from one site to another through an aqueous environment.

Lipoproteins

Cholesterol and triglycerides are transported in lipoproteins (Table 206-1), spherical particles that differ in size and composition, depending on their site of origin. Each particle is composed of a central core consisting of cholesteryl esters (the product of an esterification reaction between the polar cholesterol molecule and a fatty acid) and triglycerides, both nonpolar compounds. Free cholesterol, phospholipids, and apolipoproteins are found on the particle surface.

Chylomicrons and their remnants are the largest lipoproteins. Produced by the intestine, these particles carry fats that are absorbed from the diet. Their residence time in the circulation after a meal is short, on the order of minutes in healthy people. Chylomicrons are large and light; that is, their density is low. Because fat floats on water, particles with high fat and low protein content have lower density. Very low density lipoprotein (VLDL) is a triglyceride-rich particle produced by the liver. The removal of triglycerides from VLDL converts this particle to intermediate-density lipoprotein (IDL), which is subsequently metabolized to yield low-density lipoprotein (LDL, known as bad cholesterol). A covalent modification of the apolipoprotein (apo) in LDL, apo B100, results in the formation of lipoprotein (a). High-density lipoprotein (HDL, or good cholesterol) is formed in the blood as a byproduct of the metabolism of triglyceride-rich lipoproteins and the acquisition of esterified cholesterol from peripheral tissues.

Apolipoproteins

Apolipoproteins are amphipathic molecules capable of interacting with both the lipids of the lipoprotein core and the aqueous environment of the plasma. They function as biochemical keys, allowing lipoprotein particles access to specific sites for the delivery, acceptance, or modification of lipids. Major apolipoproteins, their chromosomal locations with sequence accession numbers, and functions are shown in Table 206-2. Serum measurements of apolipoproteins may have clinical utility. For example, increased levels of apo B and decreased levels of apo AI are associated with vascular disease. Apo B48, specific for gut-derived particles, derives its name from the fact that it is about 48% of the size of apo B100. Apo B100 and apo B48 are products of the same gene, with B48 resulting from the post-transcriptional introduction of a premature stop codon in the apo B messenger RNA by apobec1, a cytidine deaminase. Apolipoproteins can be associated with well-defined disorders. For example, genetic variation at the lipoprotein (a) locus is associated with aortic valve calcification and clinical aortic stenosis.[1]

Receptors and Proteins

Several receptors and proteins required for normal lipid transport are listed in Table 206-3.

TABLE 206-1 LIPOPROTEIN CHARACTERISTICS

LIPOPROTEIN	APOLIPOPROTEIN CONTENT	MAJOR LIPIDS	SIZE (DIAMETER, nm)	DENSITY (g/mL)
Chylomicrons, chylomicron remnants	Apo B48, apo E, apo AI, apo AII, apo AIV, apo CII, apo CIII	Triglycerides from diet	80-500	≪1.006
VLDL	Apo B100, apo E, apo CII, apo CIII	Triglycerides from liver	30-80	<1.006
IDL	Apo B100, apo E	Cholesteryl esters, triglycerides	25-35	1.006-1.019
LDL	Apo B100	Cholesteryl esters	18-25	1.019-1.063
HDL	Apo AI, apo AII, apo AV	Cholesteryl esters, phospholipids	5-12	1.063-1.210
Lp(a)	Apo B100, apo(a)	Cholesteryl esters	~30	1.055-1.085

Apo = apolipoprotein; HDL = high-density lipoprotein; IDL = intermediate-density lipoprotein; LDL = low-density lipoprotein; Lp(a) = lipoprotein (a); VLDL = very low density lipoprotein.

TABLE 206-2 MAJOR APOLIPOPROTEINS

APOLIPOPROTEIN	CHROMOSOMAL LOCATION, GENBANK SEQUENCE IDENTIFICATION	FUNCTIONS
Apo B100	2p24-p23, M14162	Structural component of atherogenic lipoproteins (VLDL, IDL, LDL); VLDL secretion; ligand for LDL receptor; elevated levels associated with vascular disease
Apo B48	Same as apo B100	Chylomicron secretion from intestine
Apo E	19q13.31, K00396	Ligand for binding of triglyceride-rich particles to LDL receptor and LRP; potential roles in Alzheimer's disease and neuronal injury
Apo AI	11q23-q24, X02162	Structural component of HDL; activates LCAT; elevated levels associated with protection from vascular disease
Apo AII	1q21-Q23, NM_001643	Genetically and biochemically associated with familial combined hyperlipidemia
Apo AIV	11q23-qter, NM_000482	Potential role in regulating food intake
Apo AV	11q23, AF202889	Required for normal lipolysis of triglyceride-rich lipoproteins
Apo CII	19q13.2, X00568	Activator of LPL
Apo CIII	11q23-qter, X01388	Inhibitor of LPL
Apo (a)	6q26-q27, X06290	Covalent bond with apo B100 forms Lp(a) and renders particle resistant to uptake by LDL receptor; genetically and biochemically associated with valvular calcification and aortic stenosis

Apo = apolipoprotein; HDL = high-density lipoprotein; IDL = intermediate-density lipoprotein; LCAT = lecithin–cholesterol acyltransferase; LDL = low-density lipoprotein; Lp(a) = lipoprotein (a); LPL = lipoprotein lipase; LRP = LDL receptor–related protein; VLDL = very low density lipoprotein.

TABLE 206-3 IMPORTANT RECEPTORS AND PROTEINS IN LIPID TRANSPORT

PROTEIN	CHROMOSOMAL LOCATION, GENBANK SEQUENCE IDENTIFICATION	FUNCTIONS
LDL receptor	19p13.3, AY114155	Clearance of apo B100 and apo E–containing lipoproteins; activity increased by statin drugs; deficiency causes familial hypercholesterolemia
PCSK9	1p32.3, NC_000001.10	Degrades LDL receptor; deficiency decreases LDL levels
LDL receptor–related protein (LRP)	12Q13-Q14, NM_000014	Clearance of apo E–containing lipoproteins
Scavenger receptor B1 (SR-B1)	12q24.32, Z22555	HDL receptor
Lipoprotein lipase (LPL)	8p22, NM_000237	Rate limiting for triglyceride metabolism; deficiency causes chylomicronemia syndrome
Lecithin–cholesterol acyltransferase (LCAT)	16q22.1, NM_000229	Esterifies cholesterol in HDL to increase HDL cholesterol levels; deficiency decreases HDL levels
Cholesteryl ester transfer protein (CETP)	16q13, NM_000078	Exchanges cholesteryl ester in HDL for triglycerides in apo B–containing lipoproteins; deficiency increases HDL levels
ABCA1	9q31, AJ12376	Transfers cholesterol in tissues to nascent HDL particles; deficiency causes Tangier disease

ABCA1 = ATP-binding cassette A1; apo = apolipoprotein; HDL = high-density lipoprotein; LDL = low-density lipoprotein; PCSK9 = proprotein convertase subtilisin-like/kexin type 9.

Low-Density Lipoprotein Receptor

The LDL receptor mediates the removal of LDL as well as some VLDL and IDL particles by binding to apo B100 and apo E. The most important site of LDL receptor expression is the liver, where its regulation is controlled by sterol regulatory element–binding proteins (SREBPs). SREBPs are found in inactive forms in the endoplasmic reticulum. Cholesterol levels in the cell are sensed by SCAP (SREBP cleavage-activating protein), which interacts with SREBPs. SCAP is capable of transporting SREBPs to the Golgi and subsequently to a compartment where they are cleaved by proteases. These proteases result in the release of the SREBP N terminus, allowing this molecule to migrate to the nucleus to stimulate the expression of genes involved in cholesterol synthesis. When intracellular cholesterol levels are high, the SCAP/SREBP complex does not move to the Golgi, SREBPs are not processed, and cholesterol synthesis stops. When cholesterol levels are low, the SCAP/SREBP complex moves to the Golgi, SREBPs are converted to active forms, and genes important for cholesterol synthesis and acquisition (such as the LDL receptor) are transcribed. Statin drugs effectively lower cholesterol. They inhibit 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, the rate-limiting enzyme in cholesterol synthesis. When statins inhibit HMG-CoA reductase, intracellular cholesterol levels fall, SCAP shepherds SREBPs to the Golgi for activation, an active SREBP stimulates transcription of the LDL receptor gene, and increased levels of the LDL receptor protein on the surface of the hepatocyte bind and remove LDL particles from the circulation.

Proprotein Convertase Subtilisin-like/Kexin Type 9

Proprotein convertase subtilisin-like/kexin type 9 (PCSK9) is a secreted enzyme that binds to the LDL receptor and increases its degradation, resulting in elevated levels of LDL cholesterol. PCSK9 deficiency in humans is associated with low LDL levels and less atherosclerosis. Pharmacologic antagonism of PCSK9 strikingly lowers LDL cholesterol levels in humans treated with a statin.[2] Potential therapies targeting PCSK9 to affect clinical outcomes are being actively pursued.

Low-Density Lipoprotein Receptor–Related Protein

The LDL receptor–related protein (LRP), also called the chylomicron remnant receptor, participates in the removal of intestine-derived lipoproteins by interacting with apo E. Chylomicron remnants carry apo B48, which is missing the LDL receptor–binding domain, but these particles are also cleared by the LDL receptor through apo E binding.

Scavenger Receptor B1

Scavenger receptor B1 (SR-B1) is a protein expressed in liver that binds HDL. Unlike the LDL receptor that endocytoses LDL particles, the SR-B1 protein does not internalize HDL particles but instead facilitates the transfer of cholesteryl ester from HDL to the liver. Its genetic manipulation in mice has raised clinically relevant questions about the significance of elevated HDL levels. Inactivation of SR-B1 elevates HDL cholesterol levels but promotes atherosclerosis, presumably because of disruption of the transport of cholesterol from peripheral cells, where it can cause disease to the liver, where it is excreted. These results suggest that it is not the level of HDL but the flux of cholesterol through HDL that affords protection from vascular disease.

Lipoprotein Lipase

Lipoprotein lipase (LPL) is rate limiting for the metabolism of triglyceride-rich lipoproteins and is required for the generation of HDL particles because

HDL is absent from LPL-deficient mice. Deficient LPL activity thus provides a physiologic explanation for the common association between high triglyceride levels and low HDL cholesterol.

Niemann-Pick C1-like Protein

The Niemann-Pick C1-like protein (NPC1L1) in the small intestine and liver helps transport dietary cholesterol from the intestinal lumen to intestinal enterocytes. Heterozygous inactivating mutations of the *NPC1L1* gene are associated with 12-mg/dL lower LDL levels and a 50% lower risk of coronary disease.[3] Ezetimibe inhibits the NPC1L1 protein.

Lecithin–Cholesterol Acyltransferase

Lecithin–cholesterol acyltransferase (LCAT) is associated with HDL in the circulation, where it esterifies free cholesterol to form cholesteryl esters that are easily stored in the nonpolar core of the lipoprotein. LCAT deficiency, a rare disorder, is characterized by low HDL as well as by anemia and renal failure, clinical features probably related to disruption of normal membrane function by the accumulation of excess unesterified cholesterol.

Cholesteryl Ester Transfer Protein

Cholesteryl ester transfer protein (CETP) exchanges one molecule of cholesteryl ester in HDL for one molecule of triglyceride in apo B–containing particles such as VLDL. The resulting HDL particle is triglyceride enriched, enhancing its clearance (especially by an enzyme related to LPL, hepatic lipase) and lowering HDL. Inhibition of CETP activity increases HDL levels.

ATP-Binding Cassette A1

ATP-binding cassette A1 (ABCA1) is a cell membrane protein that mediates the transfer of cholesterol and phospholipids from cells to lipid-poor apo AI, a process that promotes HDL formation. ABCA1 in the liver contributes to the genesis of HDL, and overexpression of ABCA1 in macrophages may diminish atherosclerosis. Heterozygous ABCA1 deficiency is responsible for isolated low HDL cholesterol levels that occur in some families. Rare homozygotes for ABCA1 mutations have Tangier disease, characterized by the accumulation of cholesteryl esters in macrophages and resulting in distinctive features, including orange-yellow tonsils, neuropathy, and hepatosplenomegaly. HDL is very low to absent. Atherosclerosis is probably increased in these patients, but its extent may be moderated by concomitantly low LDL levels.

EXOGENOUS LIPID METABOLISM

Animal products containing cholesterol and triglycerides are eaten regularly by most people. Dietary fats broken down in the gut into individual components are transported across cell membranes into the enterocyte, where they are re-esterified into cholesteryl ester and triglycerides, then packaged onto apo B48. These particles gain access to the plasma through the thoracic duct and acquire other apolipoproteins in part by transfer from HDL. These mature chylomicrons circulate to peripheral tissues. LPL, bound to the capillary endothelium in tissues such as adipose tissue and muscle, is activated by apo CII on chylomicrons, and fatty acids hydrolyzed from triglycerides by LPL are released and transported into adipose tissue for storage or into muscle for energy. This process also requires apo AV, an apolipoprotein transported in HDL that appears to facilitate the interaction between LPL and triglyceride-rich lipoproteins, as well as glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1 (GPIHBP1), a recently discovered protein that forms a platform for triglyceride metabolism at the endothelium.

Progressive hydrolysis of triglyceride converts chylomicrons into chylomicron remnants, which are enriched in cholesteryl esters. Chylomicron remnants are removed in the liver by species that bind apo E: LRP, the LDL receptor, and cell surface glycosaminoglycans. Chylomicrons are large, and it is unlikely that they contribute to atherosclerosis. Chylomicron remnants are small enough to enter the subendothelial space, where they are taken up by macrophages. Remnants are atherogenic and may promote atherosclerosis after meals. This process is missed by the standard practice of measuring fasting lipoproteins.

Chylomicrons are not soluble. Their presence causes the "tomato soup" appearance of blood drawn after a fatty meal. Because they are mostly triglycerides, they float to the top of serum that is refrigerated overnight, leaving a layer of "cream" on top of the sample. The detection of chylomicrons in fasting serum has clinical relevance because it indicates a risk for pancreatitis and other elements of the chylomicronemia syndrome.

ENDOGENOUS LIPID METABOLISM

Fats deposited in the liver are metabolized into component lipid species, re-esterified as cholesteryl ester and triglycerides, and stored in hepatocytes or exported as lipoproteins (Fig. 206-1). The liver produces the triglyceride-rich VLDL. Its rate of production depends on the availability of triglycerides. Apo B100 is the major apolipoprotein of VLDL, but regulation of the apo B gene does not appear to control VLDL synthesis. Production of the apo B100 protein depends on its cotranslational stabilization. As the message is translated into protein, the presence of triglyceride stabilizes the peptide and allows the continued addition of amino acids. In the absence of triglycerides, the apo B molecule is degraded. The transfer of triglycerides to the growing apo B peptide is mediated by microsomal transfer protein (MTP). Mutations in MTP cause abetalipoproteinemia, a rare disease characterized by the absence of circulating apo B. In the absence of apo B, the metabolism of fat-soluble vitamins (normally carried in lipoproteins) is disrupted, and patients with abetalipoproteinemia suffer from multisystem defects, including severe neurologic dysfunction and retinopathy that are presumably caused by deficiency of vitamins E and A. Drugs that interfere with MTP function lower lipids but, not surprisingly, cause the accumulation of triglyceride in the liver. The apo B gene is normal in patients with abetalipoproteinemia. Mutations in the apo B gene cause another condition known as hypobetalipoproteinemia, caused by shortened forms of the apo B protein. Subjects with hypobetalipoproteinemia have very low but not absent levels of circulating lipids and appear to be healthy.

Nascent VLDL containing one apo B100 molecule per particle is secreted into the plasma, where it acquires apo E, apo CII, and apo CIII. In a process analogous to that occurring with chylomicrons, apo CII on VLDL activates LPL, and fatty acids hydrolyzed from triglycerides by LPL are released in capillary beds and transported into tissues. With continued hydrolysis, VLDL

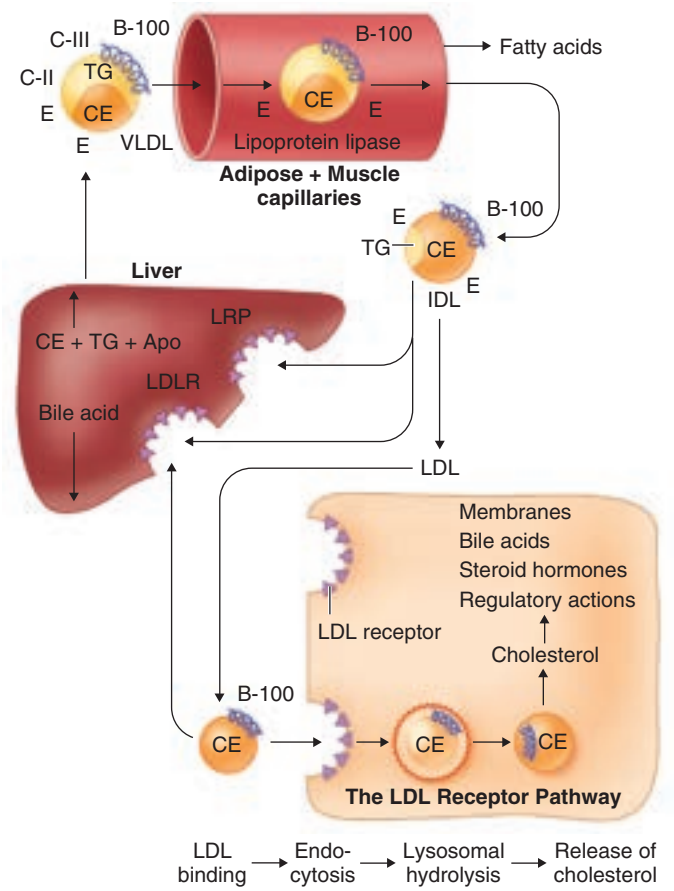

FIGURE 206-1. Endogenous lipid metabolism. In the liver, triglycerides (TG), cholesteryl esters (CE), and apolipoprotein B100 are packaged as very low density lipoprotein (VLDL) particles. TG is hydrolyzed by lipoprotein lipase to generate intermediate-density lipoprotein (IDL), which is further metabolized to generate low-density lipoprotein (LDL). This particle can be removed by the liver or by peripheral cells. Cholesterol derived from LDL regulates several processes and can be used for the synthesis of bile acids, steroid hormones, and cell membranes. LDLR = low-density lipoprotein receptor.

is converted to IDL, a cholesteryl ester–rich particle with an apolipoprotein complement of only apo B and apo E. These particles, like chylomicron remnants, are atherogenic. Unlike chylomicron remnants, IDLs are included in current management schemes because reporting of LDL cholesterol levels by most clinical laboratories includes IDL. IDL can be taken up by either the LRP or the LDL receptor in the liver. In the presence of a normal apo E molecule, IDLs are converted to LDL, consisting of one molecule of apo B100 per particle and cholesteryl esters with essentially no triglycerides. The majority of LDL is removed from the plasma by the LDL receptor pathway in the liver. LDL uptake is followed by migration of LDL particles to lysosomes, where cholesterol is released for (depending on the cell type) plasma membrane localization, bile acid synthesis, steroid hormone synthesis, and interaction with SCAP for the control of SREBP activation. Some LDL enters the subendothelial space of the vascular wall, where its modification by oxidation or other processes promotes its uptake by macrophages in atherosclerotic lesions.

Most VLDL particles are large and are not thought to promote vascular disease. Some small VLDL particles as well as IDL and LDL are atherogenic. Because VLDL is mostly triglyceride, patients with elevated fasting triglyceride levels have either increased numbers of VLDL particles or an increased triglyceride content in VLDL. LDL has a plasma half-life of 2 to 5 days. The detection of elevated fasting cholesterol levels usually reflects the presence of either increased numbers of LDL particles or increased cholesteryl ester in LDL. LDL also exists in a range of sizes. Small, dense LDL tends to occur in the setting of concomitant hypertriglyceridemia. This type of lipoprotein is thought to have greater atherogenic potential than larger LDL species, perhaps because of easy access to the vascular wall and greater susceptibility to oxidative modification. Lipoprotein particle size and number can be quantified by nuclear magnetic resonance techniques, but it is not clear that these data provide diagnostic advantages beyond the determination of total cholesterol, triglycerides, LDL cholesterol, and HDL cholesterol.

REVERSE CHOLESTEROL TRANSPORT AND HIGH-DENSITY LIPOPROTEIN METABOLISM

Lipid metabolism is dynamic. Lipoprotein particles interact with the vasculature and with one another, exchanging surface materials, apolipoproteins, and nonpolar lipids. HDL is an important reservoir for components cast off during the metabolism of other lipoproteins as well as lipids discarded by cells. Nascent HDL is generated by the liver and intestine as a phospholipid disc containing apo AI and apo AII. It accepts unesterified (free) cholesterol and phospholipids shed from cells. This unesterified cholesterol is converted to cholesteryl ester by the action of LCAT and stored in the center of the disc, allowing it to become spherical. As the core triglycerides of VLDL are metabolized by LPL, the VLDL particle collapses, leaving redundant surface lipids (phospholipid in the form of lecithin and unesterified cholesterol) and excess apolipoproteins such as apo CII, apo CIII, and apo E, which are transferred to HDL.

Reverse cholesterol transport is the beneficial process by which cholesterol in peripheral cells, such as foam cells in an atherosclerotic lesion, is transported back to the liver for excretion. There are at least two well-defined pathways mediating this transfer (Fig. 206-2). First, after accepting cholesterol from peripheral cells and esterifying it through the action of LCAT, HDL can interact directly with the liver by binding to SR-B1 and transferring cholesteryl ester to the hepatocyte. Second, HDL can transfer cholesteryl ester to apo B100–containing lipoproteins such as VLDL through the action of CETP. This cholesteryl ester can ultimately be transported to the liver after conversion of VLDL to IDL to LDL and uptake by the LDL receptor. This pathway is not direct because the transfer of cholesteryl ester to apo B–containing lipoproteins results in cholesterol-enriched particles that may be taken up by foam cells in atherosclerotic plaques before being cleared by the liver. Humans with genetic defects in CETP have high HDL levels and appear to be healthy. The benefits of raising HDL with medications are uncertain. Increasing HDL cholesterol and apo AI through the use of the CETP inhibitor dalcetrapib did not improve outcomes in patients after an acute coronary syndrome.[4] Increasing HDL cholesterol through the use of niacin does not provide clinical benefit to patients with atherosclerotic vascular disease intensively treated with a statin.

LIPID-ACTIVATED NUCLEAR RECEPTORS AND LIPID METABOLISM

Nuclear receptors are transcription factors that are activated by ligand binding to increase the expression of specific sets of genes. Several nuclear receptors

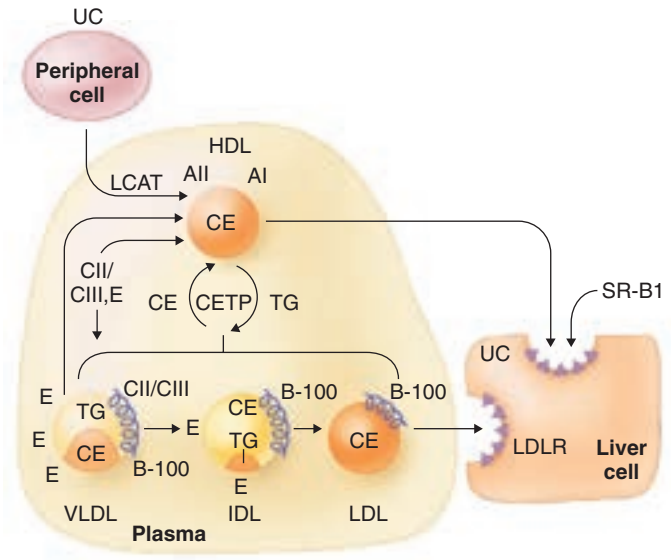

FIGURE 206-2. Reverse cholesterol transport and high-density lipoprotein (HDL) metabolism. Unesterified cholesterol (UC) in peripheral cells can be transferred to HDL and esterified by lecithin-cholesterol acyltransferase (LCAT). This cholesterol ester (CE) in HDL can be transferred to the liver directly through scavenger receptor B1 (SR-B1). Alternatively, it can be transferred to apolipoprotein B100–containing lipoproteins in exchange for triglycerides (TG) through the action of cholesteryl ester transfer protein (CETP). IDL = intermediate-density lipoprotein; LDL = low-density lipoprotein; LDLR = low-density lipoprotein receptor; VLDL = very low density lipoprotein.

are activated by lipids, play important roles in systemic lipid metabolism, and are current or potential targets of medications for altering lipids in patients.

Peroxisome proliferator-activated receptors (PPARs) are thought to be activated by fatty acids or their derivatives, such as phospholipids. There are at least three types: PPARα, PPARγ, and PPARδ. PPARα stimulates the expression of genes mediating fatty acid oxidation and those promoting the formation of HDL. Fibrate drugs such as gemfibrozil and fenofibrate work by activating PPARα. They lower triglycerides by accelerating the oxidation of fatty acids in the liver, so that less lipid is available for stabilizing apo B100 in VLDL secretion, and they elevate HDL by increasing the expression of apo AI, LPL, and other genes. PPARγ is expressed mostly in adipose tissue and macrophages. It increases the expression of genes promoting the development of fat tissue and appears to suppress chronic inflammation. Thiazolidinedione drugs such as pioglitazone work by activating PPARγ. They lower blood glucose concentration in people with diabetes by decreasing insulin resistance, a complex process that also results in multiple lipid effects, including lower triglycerides and higher HDL (expected to be beneficial) and higher LDL (expected to be detrimental). Their use is complicated by volume expansion and an increased risk of heart failure. PPARδ is expressed widely and has multiple effects, including increased fatty acid oxidation.

Two other nuclear receptors are activated by lipids and modulate lipid physiology. Liver X receptors (LXRα and LXRβ) are activated by oxidized derivatives of cholesterol. In the liver, they promote the synthesis of fatty acids and triglycerides in addition to stimulating both the conversion of cholesterol into bile acids and the excretion of bile acids into the gut. In the intestine, they suppress the absorption of cholesterol. The farnesoid X receptor (FXR) is activated by bile acids. FXR stimulates the secretion of bile acids into bile as well as the reabsorption of bile acids from the intestine.

Together, these receptors help orchestrate two important, futile cycles in lipid metabolism. In one, fatty acids exported from the liver (in the form of triglycerides within VLDL) and from the intestine (in chylomicron triglycerides) are released to peripheral tissues by the action of LPL. Some are taken up by muscle, where their activation of PPARα accelerates their oxidation in mitochondria, yielding adenosine triphosphate. Others enter adipose tissue, where they are re-esterified into triglycerides. From there, fatty acids are released in a process stimulated by catecholamines and glucagon. This process is complicated, involving hormone-sensitive lipase, adipose triglyceride lipase, and the remodeling of proteins that coat lipid droplets to alter their accessibility to lipases. After lipolysis, fatty acids bind to albumin and return to the liver, where they can fuel the production of more VLDL particles. Nicotinic acid improves lipids in part by blocking the release of fatty acids from adipose tissue through a process that is incompletely understood.

In another futile cycle, bile acids (formed from cholesterol and constituting the major pathway for the excretion of cholesterol from the body) are secreted into the intestine through events stimulated by LXR and FXR. Bile acids are reabsorbed in the terminal ileum. Treatment with bile acid sequestrants such as colesevelam interrupts this enterohepatic circulation, and the increased excretion of cholesterol (in the form of bile acids) depletes cholesterol content in the liver, leading to the induction of LDL receptors and thereby lowering circulating levels of LDL. This treatment also tends to elevate triglyceride levels, the result of de-repression of several processes mediated by FXR that decrease fatty acids and triglycerides. Through effects likely involving both LXR and FXR, colesevelam lowers blood glucose concentration in people with type 2 diabetes.

IMPORTANT CLINICAL DISORDERS OF LIPID METABOLISM

Familial Hypercholesterolemia

EPIDEMIOLOGY

Familial hypercholesterolemia (FH) is an autosomal dominant form of hypercholesterolemia caused by defects in LDL receptor activity. The majority of affected patients have mutations in the LDL receptor gene. Heterozygotes for LDL receptor mutations occur at a frequency of perhaps 1 in 500 in the population, but they account for up to 5% of premature myocardial infarctions (those occurring in men younger than 55 years and women younger than 65 years).

PATHOBIOLOGY

In addition to defects in the LDL receptor gene, other mutations can cause autosomal dominant hypercholesterolemia that is clinically indistinguishable from FH. These include familial defective apo B, caused by mutations that interfere with the ability of LDL to bind the LDL receptor, and variants in PCSK9.

CLINICAL MANIFESTATIONS

Total cholesterol levels are usually higher than 300 mg/dL, with LDL cholesterol levels higher than 200 mg/dL. Triglyceride levels are generally normal. Clinical features include thickening of the Achilles tendon as well as xanthomas at the extensor tendons of the knees and hands, reflecting the infiltration of lipid-laden macrophages at these sites. Clinically apparent tendon xanthomas may occur in a minority of patients with FH. Arthralgias are common, perhaps because of the presence of macrophage-mediated inflammation; these tend to improve with cholesterol lowering. Other features include xanthelasmas and corneal arcus, although the latter is also seen in elderly people and certain ethnic populations independent of cholesterol levels. Homozygotes (about 1 in 1 million) have total cholesterol levels in the range of 800 to 1000 mg/dL and usually do not survive to adulthood without liver transplantation to provide LDL receptors. Children and adolescents with this disease may develop aortic valve disease because of macrophage infiltration at the aortic origin.

The penetrance of cardiovascular events in heterozygous FH is variable. Some subjects present with sudden death or accelerated disease in their 20s, and others (generally women without other risk factors) survive beyond menopause without clinically evident disease.

TREATMENT

Rx

Treatment should address any associated risk factors for vascular disease, such as smoking, hypertension, and diabetes. Patients should be instructed in maintaining a diet low in saturated fat and cholesterol, and most require more than one cholesterol-lowering medication to achieve goals. It is not uncommon for FH patients to be treated with a statin drug, a bile acid sequestrant such as colesevelam, and nicotinic acid. Some with aggressive disease require LDL apheresis, which involves perfusing blood through a column that extracts apo B–containing lipoproteins. Lomitapide, an MTP inhibitor, was recently approved to treat homozygous FH. This agent prevents the assembly of apo B–containing lipoproteins, which allows improvement in circulating lipids in those who have LDL receptor deficiency and thus are unable to increase LDL receptor expression through the use of statins. A serious side effect of lomitapide is fatty liver. Mipomersen, an antisense oligonucleotide targeting apo B, was also recently approved to treat homozygous FH. Hepatic toxicity is also a concern with this agent. The exact role for these therapies is uncertain, and both require careful monitoring to ensure safe use.

Familial Combined Hyperlipidemia

EPIDEMIOLOGY

Familial combined hyperlipidemia is an autosomal dominant form of hyperlipidemia that is present in up to 2% of the general population. It accounts for as many as 20% of cases of premature coronary artery disease.

PATHOBIOLOGY

The specific molecular defect is uncertain. Familial combined hyperlipidemia also appears to be associated with the metabolic syndrome (Chapter 229). The disorder is characterized by the primary overproduction of apo B. VLDLs secreted by the liver are small. Small, dense LDLs, thought to be particularly atherogenic, accumulate. For a given concentration of LDL cholesterol, patients with familial combined hyperlipidemia have greater numbers of LDL particles and an increased apo B concentration.

DIAGNOSIS

The diagnosis is made in the setting of a family history of premature coronary disease with different lipid phenotypes combined in the same family. Affected family members may have elevated triglycerides, elevated LDL cholesterol, or both, or they may have hypertriglyceridemia with low HDL cholesterol. Lipid phenotypes commonly change over time.

Familial Hypertriglyceridemia

Familial hypertriglyceridemia is also a common autosomal dominant disorder, occurring in 1 to 2% of the general population. Affected family members have isolated elevated triglycerides. A unifying molecular mechanism is lacking. The phenotype is stable, with affected family members consistently showing isolated hypertriglyceridemia on repeated analyses. The disorder is characterized by primary overproduction of triglyceride. Lipoprotein particles tend to be large, consisting of increased amounts of triglyceride relative to apo B. For a given level of cholesterol, these patients have lower numbers of lipoprotein particles and a decreased apo B concentration. The relationship between this disorder and cardiovascular risk is uncertain. Affected kindreds do not appear to have a propensity for premature vascular disease, but these individuals are at risk for the chylomicronemia syndrome when an additional stimulus for hypertriglyceridemia, such as uncontrolled diabetes, is present.

Chylomicronemia Syndrome

PATHOBIOLOGY

This syndrome occurs when triglycerides are extremely elevated, usually higher than 2000 mg/dL. Individuals with homozygous defects in the *LPL* gene can present with the syndrome in infancy, although the penetrance of clinical sequelae is extremely variable. Some individuals suffer repeated episodes of the syndrome throughout life, whereas others with triglyceride levels consistently above 2000 mg/dL remain completely asymptomatic. The basis for the wide spectrum of symptoms despite similar degrees of severe hypertriglyceridemia is unknown. *LPL* gene therapy in the form of alipogene tiparvovec (the human *LPL* gene in an adeno-associated virus vector) was recently approved by the European Commission to treat patients with LPL deficiency. This represents the first gene therapy approved for use in the Western world. Other molecular defects responsible for the chylomicronemia syndrome include mutations in apo CII and apo AV.

Although defects in LPL, apo CII, and apo AV can cause the syndrome, the disorder is most likely to occur when patients with a common predisposition to hypertriglyceridemia (familial combined hyperlipidemia or familial hypertriglyceridemia) develop another defect associated with elevated triglycerides (such as uncontrolled diabetes, obesity, and treatment with glucocorticoids, estrogens, or other drugs).

CLINICAL MANIFESTATIONS

Clinical features include eruptive xanthomas (see Fig. 51-12) on the back, buttocks, knees, and elbows; lipemia retinalis (a white appearance of retinal blood vessels, usually seen with triglyceride levels >4000 mg/dL); severe abdominal pain and pancreatitis (which can be life-threatening); hepatosplenomegaly; dyspnea; lymphadenopathy; and neurologic dysfunction, such as memory loss and peripheral neuropathy.

DIAGNOSIS

Evaluation can be complicated by the presence of extreme hypertriglyceridemia, which interferes with the determination of amylase (pancreatic lipase

should be measured when pancreatitis is suspected) and artifactually lowers serum glucose, sodium, and other analytes.

TREATMENT Rx

Treatment consists of intravenous hydration and other supportive care for pancreatitis, complete elimination of dietary fat (which usually causes striking decreases in lipids within 24 to 48 hours), and appropriate blood glucose control with insulin in the setting of diabetes.

Dysbetalipoproteinemia

This rare disorder is caused by a mutation in apo E. There are three common variants of the apo E protein: E2, E3 (considered normal), and E4. Subjects with one or more E4 alleles are at risk for Alzheimer's disease. Subjects with two E2 alleles are at risk for dysbetalipoproteinemia. The frequency of this genotype is about 1% in the general population, but dysbetalipoproteinemia is rare, requiring an additional poorly defined factor. Hypothyroidism is known to precipitate the disorder. Subjects classically present with equal elevations of triglycerides and cholesterol in the range of about 300 to 600 mg/dL and xanthomas, especially in the palmar creases of the hands. They are at substantial risk for coronary artery disease. Unlike patients with FH, patients with dysbetalipoproteinemia are also at risk for severe peripheral vascular disease. Atherosclerosis occurs in part because of the presence of elevated concentrations of chylomicron remnants, which are not removed normally because of the presence of apo E2, and IDL particles, which are not converted normally to LDL in the presence of apo E2. It is unknown why the remnants that accumulate in dysbetalipoproteinemia cause both peripheral vascular disease and coronary disease, whereas the LDL particles that accumulate in FH tend to cause only coronary disease.

Diabetic Dyslipidemia

PATHOBIOLOGY

Insulin is a critical regulator of lipid metabolism, and because diabetes represents impaired insulin signaling, lipid disorders are common in both type 1 and type 2 diabetes (Chapter 229). Hypertriglyceridemia is the hallmark of diabetic dyslipidemia. This is driven by two mechanisms. First, LPL is insulin dependent. In the absence of insulin or in the presence of insulin resistance, LPL enzyme activity is deficient, and triglyceride-rich lipoproteins cannot be metabolized appropriately. Second, insulin suppresses the release of free fatty acids from adipose tissue stores. Insulin deficiency or resistance results in unabated release of free fatty acids, and these return to the liver, where they stabilize apo B synthesis and increase VLDL production.

TREATMENT Rx

The lack of insulin in patients with type 1 diabetes causes diabetic ketoacidosis, in which elevated triglycerides can be severe and are corrected by reinstitution of insulin therapy (Chapter 229). Intensive insulin therapy in type 1 diabetes decreases triglycerides, LDL, and often lipoprotein (a). Patients gain weight with intensive insulin therapy, and increased adiposity tends to lower HDL levels, which may explain why HDL does not always increase in intensively treated type 1 patients. Patients with type 2 diabetes usually have increased triglycerides and decreased HDL. These abnormalities improve but seldom normalize with better glycemic control. Most of the improvement occurs with initial pharmacologic glucose-lowering therapy, regardless of its mechanism of action.

Diabetes is frequently classified as a secondary cause of lipid disorders. However, lipid abnormalities are intrinsic to diabetes, and people with diabetes who lack clinical evidence of vascular disease have the same risk for cardiovascular events as those with established coronary disease. Therefore, every person with diabetes should be evaluated for lipid disorders, and adequate control of lipid levels often requires treatment with a statin drug. In patients with diabetes, cholesterol-lowering therapy with a statin reduces major cardiovascular events on the basis of primary prevention trials as well as secondary prevention trials. Statin treatment should not be delayed while glycemic control is optimized.

EVALUATION AND THERAPY OF LIPID DISORDERS

DIAGNOSIS

The initial evaluation should include a complete history and physical examination, with careful attention to potential secondary causes of lipid disorders (Table 206-4). Diabetes, obesity, hypothyroidism, and excess alcohol intake are probably the most common secondary contributors to abnormal lipid metabolism.

Among prescription medications, β-adrenergic blocking agents are frequent contributors to abnormal lipid profiles. These agents have proven beneficial effects after myocardial infarction but also tend to promote weight gain, to elevate triglycerides, and to decrease HDL. Many clinicians are using lower doses of these agents and relying on more lipid-neutral drugs (angiotensin-converting enzyme inhibitors and angiotensin receptor blockers) for blood pressure control.

It is now common for internists to encounter patients with human immunodeficiency virus (HIV) infection with hyperlipidemia. Some series estimate that more than half of HIV-infected patients treated with protease inhibitors for 2 years develop dyslipidemia, frequently with a redistribution of fat resembling that in genetic lipodystrophy syndromes (Chapter 389). Newer protease inhibitors appear to have fewer metabolic sequelae than older agents such as ritonavir. In addition to appropriate management of other risk factors, lipid-lowering therapy should be tailored to the ongoing HIV drug regimen. Pravastatin is least likely to interact with protease inhibitors but is less effective at lowering LDL cholesterol. Agents such as simvastatin that are substantially metabolized by the cytochrome P-450 3A4 system should not be used in patients treated with protease inhibitors because of delayed

TABLE 206-4 SECONDARY CAUSES OF LIPID DISORDERS

CONDITION OR MEDICATION	COMMENTS
Diabetes	Common contributor to dyslipidemia; abnormal lipids are seldom normalized by glycemic control alone
Obesity	Increased triglycerides and decreased HDL are common; LDL may be elevated in some and decrease with weight loss
Hypothyroidism	Thyroid hormone regulates multiple steps in lipid metabolism, including LDL receptor expression and LPL activity
Alcohol	Can cause hypertriglyceridemia in susceptible patients, but mild intake is linked to decreased risk of vascular disease
Renal disease	Increased LDL in nephrotic syndrome, hypertriglyceridemia in end-stage renal disease
Obstructive liver disease	Can be associated with very high cholesterol levels; some evidence that diseases such as primary biliary cirrhosis are not associated with increased vascular events despite dyslipidemia
Diuretics	Increased LDL with high doses; current practice of using low doses of thiazides decreases vascular events and has minimal effect on lipids
β-Adrenergic receptor blockers	Increased triglycerides and decreased HDL, probably by inhibiting LPL
Anabolic steroids	Can result in very low HDL (<10 mg/dL)
Estrogens	Exacerbate hypertriglyceridemia when given orally; this effect is not seen with topical estrogen therapy
Protease inhibitors	Increased triglycerides and decreased HDL, especially in the setting of HIV-associated lipodystrophy
Glucocorticoid excess	Increased triglycerides and decreased HDL, probably related to exacerbation of insulin resistance
Antipsychotics	Increased triglycerides and decreased HDL, probably related to increased adiposity and insulin resistance
Retinoids	Increased triglycerides
Systemic lupus erythematosus	Chronic inflammation may increase risk of vascular disease independent of effects on lipid metabolism
Acute intermittent porphyria	Many agents used to treat lipid disorders reported to provoke episodes of abdominal pain

HDL = high-density lipoprotein; HIV = human immunodeficiency virus; LDL = low-density lipoprotein; LPL = lipoprotein lipase.

clearance of the statin. PPARα agonists such as fenofibrate can be used to lower triglycerides, but the effects are limited because their clearance is accelerated by protease inhibitor treatment.

Antipsychotics have complex effects on metabolism. Some of the newer agents used to treat schizophrenia promote hyperlipidemia, obesity, and insulin resistance. Although therapeutic decisions should be based on psychiatric responses, substituting for agents with prominent metabolic side effects, such as olanzapine, can improve lipid profiles.

Laboratory Evaluation

After an 8- to 12-hour fast (tell patients that drinking water and other calorie-free beverages during this period is acceptable), total triglycerides, total cholesterol, HDL cholesterol, and LDL cholesterol should be measured. In most clinical laboratories, LDL cholesterol is still calculated by this formula: LDL = total cholesterol − HDL cholesterol − (triglycerides/5). This formula is not valid when triglyceride levels are higher than 400 mg/dL, however. It is also possible to measure LDL directly, which is sometimes useful for monitoring the therapeutic effects on LDL alone, and this determination does not require patients to fast. Excessive lipid determinations may not be helpful. Biologic and random variability in cholesterol levels is considerable, and for patients with levels that are 19 mg/dL or more below their goal, serial monitoring is more likely to detect false-positive than true-positive increases during a period of 3 years.

What constitutes "normal" lipid levels is unknown. Prior expert consensus panels have identified a normal triglyceride level as lower than 150 mg/dL. Values from 150 to 199 mg/dL are borderline high, 200 to 499 mg/dL is considered high, and above 500 mg/dL is very high. In patients with fasting triglyceride levels higher than 500 mg/dL, triglycerides are a primary target of therapy to decrease the risk of pancreatitis and chylomicronemia syndrome. These individuals should be managed in consultation with specialists in lipid disorders. Treatment usually involves a very low fat diet, an exercise program, a weight loss regimen in the setting of obesity, glycemic control in the setting of diabetes, and either nicotinic acid or a fibrate drug alone or in combination with fish oils.

Risk Assessment

In 2013, the American College of Cardiology and the American Heart Association released new guidelines for the treatment of cholesterol.[5,6] These recommendations differ from the recommendations of the National Cholesterol Education Program Adult Treatment Panel III report from 2002 in several important ways. Previously, the number of risk factors present was determined and used in conjunction with calculated risk to identify specific therapeutic goals for LDL cholesterol levels. Risk factors for coronary heart disease include smoking, hypertension (or treatment for hypertension), family history of premature coronary heart disease (younger than 55 years in men, 65 years in women), low HDL (<40 mg/dL), and age (older than 45 years in men, 55 years in women). These risk factors can still be useful in the evaluation and management of patients, but their use is de-emphasized in the most recent guidelines. Also de-emphasized is the notion of specific LDL targets. Previously, the LDL goal was lower than 100 mg/dL (or optionally lower than 70 mg/dL) in high-risk patients with known atherosclerotic cardiovascular disease. An alternative metric was non-HDL cholesterol. Goals for non-HDL cholesterol (calculated as total cholesterol minus HDL cholesterol) were 30 mg/dL higher than those for LDL. For example, in a high-risk patient with an LDL goal of less than 100 mg/dL, the corresponding non-HDL cholesterol goal was less than 130 mg/dL. These targets may still be useful in certain clinical settings, but they were not recommended in the most recent guidelines because their utility was not validated in randomized clinical trials.

The latest guidelines have not been uniformly endorsed by professional organizations representing those involved in the care of patients with lipid disorders. However, they are likely to influence clinical practice.[7] They are also subject to modifications as new data become available.

In short, the current American College of Cardiology/American Heart Association guidelines identify groups of individuals for which statin therapy is recommended and other groups for which statin therapy is not recommended. These are summarized in Table 206-5. High-intensity statin therapy to decrease LDL by 50% or more than with the use of higher doses of atorvastatin or rosuvastatin is recommended. Moderate-intensity therapy to decrease LDL by 30% to less than 50% with the use of lower doses of atorvastatin, rosuvastatin, or other statins is an option for patients who cannot tolerate high-intensity therapy or for patients with diabetes and a 10-year risk of atherosclerotic cardiovascular disease of less than 7.5%.

TABLE 206-5 TREATMENT RECOMMENDATIONS BASED ON 2013 ACC/AHA CHOLESTEROL TREATMENT GUIDELINES

STATIN THERAPY RECOMMENDED	STATIN THERAPY NOT RECOMMENDED
Patients with clinically evident atherosclerotic cardiovascular disease	Adults with end-stage renal disease
Adults with LDL cholesterol > 190 mg/dL	Patients with NYHA class II, III, or IV heart failure
Patients with type 1 or type 2 diabetes and LDL cholesterol ≥ 70 mg/dL	Adults > 75 years of age without clinical evidence of atherosclerotic cardiovascular disease
Adults with LDL cholesterol ≥ 70 mg/dL and 10-year risk of atherosclerotic cardiovascular disease ≥ 7.5%*	

*Determined by risk calculator found at *http://my.americanheart.org/cvriskcalculator*.
ACC/AHA = American College of Cardiology/American Heart Association; LDL = low-density lipoprotein; NYHA = New York Heart Association.

A risk calculator for estimating 10-year risk of atherosclerotic cardiovascular disease is found at *http://my.americanheart.org/cvriskcalculator*. This can be downloaded to devices such as smart phones, allowing the provision of individualized information to patients at the point of care.

TREATMENT Rx

General Measures

Intensive Lowering of Low-Density Lipoproteins

Many randomized clinical trials support the concept that intensive LDL lowering decreases cardiovascular event rates in patients with coronary heart disease and its equivalents. In the Heart Protection Study, those with an LDL level of less than 100 mg/dL benefited from further lowering of LDL with simvastatin. In the Pravastatin or Atorvastatin Evaluation and Infection Therapy trial, patients with acute coronary syndrome derived greater benefit with an LDL level of 62 mg/dL reached with atorvastatin than with an LDL level of 95 mg/dL reached with pravastatin. In the Treating to New Targets study, patients with stable coronary heart disease had fewer clinical events with an LDL level of 77 mg/dL on 80 mg of atorvastatin than with an LDL level of 101 mg/dL on 10 mg of atorvastatin. In the Incremental Decrease in End Points through Aggressive Lipid-Lowering study, patients with stable coronary heart disease with an LDL level of 81 mg/dL on 80 mg of atorvastatin had fewer events than those with an LDL level of 104 mg/dL on 20 mg of simvastatin. In healthy individuals with an LDL level less than 130 mg/dL and C-reactive protein level higher than 2 mg/dL, rosuvastatin (20 mg/day) reduced major cardiovascular events by 44% and all-cause mortality by 20%. Moderate-dose statin therapy increases the risk of diabetes by 2 persons per 1000 person-years, but it reduces the risk of cardiovascular events by 6.5 persons per 1000 person-years. The benefits of statin therapy exceed the diabetes hazard, even in patients at high risk for development of diabetes.[8]

Lifestyle Changes

Therapeutic lifestyle changes should be recommended as part of the treatment regimen. These include smoking cessation, weight reduction, exercise on most days of the week, reduced intake of dietary cholesterol to less than 200 mg/day, reduction of saturated fat to less than 7% of total calories, increased soluble fiber intake (to at least 10 g/day), and consumption of plant stanols or sterols (2 g/day). However, the role of these changes alone in reducing the risk of cardiovascular disease is unclear. In overweight or obese people with diabetes, an intensive lifestyle intervention did not decrease cardiovascular events.[9]

Medical Therapy

Medications commonly used to treat lipid disorders are listed in Table 206-6.

Statins

Statins are effective. An analysis of results from 90,000 participants in statin trials suggested that the 5-year incidence of cardiovascular events decreases by about 1% for each decrease of 2 mg/dL in LDL. The benefit appears to be independent of a patient's baseline lipid values.

Statins are safe, a notion confirmed in numerous studies during two decades. When statins were introduced, concerns were raised about the possibility of increased mortality with these agents due to noncardiovascular causes such as cancer. Eleven-year total follow-up of patients in the Heart Protection Study, a trial that compared simvastatin with placebo, confirmed long-term benefits with statin use without effects on noncardiovascular mortality or cancer incidence. In the entire Danish population, those using statins had reduced cancer mortality compared with those who never used statins for each of 13 different cancer types.[10]

TABLE 206-6 MEDICATIONS USED TO TREAT LIPID DISORDERS

CLASS	SIDE EFFECTS	LIPID EFFECTS	SPECIFIC AGENTS (DOSE)
Statins (HMG-CoA reductase inhibitors)	Mildly increased liver enzymes, myalgias without evidence of muscle disease, constipation, insomnia, rhabdomyolysis (rare)	↓↓LDL, 18-55% ↑HDL, 5-15% ↓TG, 7-30%	Simvastatin (20-40 mg/day) Atorvastatin (10-80 mg/day) Pravastatin (20-80 mg/day) Fluvastatin (20-80 mg/day) Lovastatin (20-80 mg/day) Rosuvastatin (10-40 mg/day) Pitavastatin (1-4 mg/day)
Nicotinic acid	Flushing, nausea, diarrhea, hyperglycemia, hyperuricemia, hepatotoxicity (rare)	↑↑HDL, 15-35% ↓↓TG, 20-50% ↓LDL, 5-25% ↓Lp(a), variable	Extended-release or crystalline niacin (1-2 g/day)
Fibrates	Mildly increased liver enzymes, dyspepsia, gallstones, hepatotoxicity (rare), rhabdomyolysis (rare)	↓↓TG, 20-90% ↑HDL, 10-20%	Gemfibrozil (1.2 g/day) Fenofibrate (34-200 mg/day)
Cholesterol absorption inhibitor	Hepatitis, abdominal pain, back pain, arthralgias	↓LDL, 18%	Ezetimibe (10 mg/day)
Bile acid sequestrants	Constipation, decreased absorption of some drugs	↓LDL, 15-30% ↑TG, variable ↑HDL, 3-5%	Colesevelam (2.5-3.75 g/day); this agent also approved for lowering glucose in type 2 diabetes Cholestyramine (4-16 g/day) Colestipol (2-16 g/day)
Fish oils	Eructation, dyspepsia	↓TG, variable ↑ LDL, variable	Omega-3-acid ethyl esters (variable)

HDL = high-density lipoprotein; HMG-CoA = 3-hydroxy-3-methylglutaryl coenzyme A; LDL = low-density lipoprotein; Lp(a) = lipoprotein (a); TG = triglyceride.

Statins are well tolerated. Up to 5% of patients have mildly elevated serum transaminases with therapy, which is usually asymptomatic and resolves spontaneously. Statins cause important liver injury in 1.2 per 100,000 users, usually 3 to 4 months after therapy is started. Atorvastatin is mostly associated with cholestatic liver injury, whereas hepatocellular injury is more common with simvastatin. Therapy should be discontinued if elevations exceed three times the upper limit of normal or if patients have symptoms of liver dysfunction (especially fatigue and weight loss). Rhabdomyolysis is rare and more likely in the setting of concurrent treatment with azole antifungals, erythromycin, cyclosporine, and several other agents. Common but generally mild side effects include constipation, abdominal pain, and difficulty sleeping. Some patients report cognitive difficulties with statins, but it is difficult to appreciate the exact role of statin use in the development of these symptoms.

A substantial minority of patients treated with statins develop myalgias without physical findings or laboratory evidence of muscle dysfunction. The cause of this side effect is unknown. Rarely, statins cause clinical myopathy, which may be related to genetic variations in the capacity to synthesize creatine.[11] Curiously, clinical trials have not documented a difference in muscle symptoms in statin-treated compared with placebo-treated subjects. Clinicians should not minimize the impact of statin intolerance on quality of life for people with lipid disorders. However, muscle aches and joint pain are experienced by all patients at some time, and these common symptoms are often incorrectly attributed to statins. It may be appropriate to consider hypothyroidism, vitamin D deficiency, rheumatologic disorders such as polymyalgia rheumatica, or depression in these patients. The problem may be resolved by stopping the medication, allowing symptoms to resolve, restarting at a much lower dose with gradual increases during weeks, and using a different statin. In one health care system, more than half of patients discontinued statins at least temporarily, 17% had statin-related events documented, and more than 90% of those who were rechallenged with a statin were still taking the medication 1 year later.[12]

Combinations of statins with other agents such as amlodipine and extended-release niacin can enhance compliance and decrease copayment costs.

Nicotinic Acid

Nicotinic acid is a B complex vitamin that in high doses lowers triglycerides, elevates HDL, and modestly lowers LDL. The most common side effect is flushing, which can be diminished with aspirin, by taking the medication with a small snack, and by avoiding hot beverages. It may be useful in patients with very high triglyceride levels who are at risk for pancreatitis. In those already achieving low levels of LDL with a statin, adding niacin does not improve clinical outcomes even though it lowers LDL and raises HDL, and it increases serious adverse events.

Fibrates

Fibrates lower triglycerides in patients with very high levels who are at risk for pancreatitis, especially those who cannot tolerate niacin. These drugs increase HDL but also tend to increase LDL in patients with high triglyceride levels. Fibrates in combination with a statin increase risk of rhabdomyolysis. In patients with type 2 diabetes, the addition of fenofibrate to simvastatin did not reduce cardiovascular end points compared with simvastatin alone.[13]

Other Agents

Ezetimibe lowers both LDL and triglycerides, with minimal effects on HDL. It is effective at achieving additional LDL lowering when it is used in combination with a statin. In a large randomized trial, the addition of 10 mg ezetimibe to high-dose statin therapy further reduced LDL from 70 mg/dL to 53 mg/dL and reduced adverse cardiovascular outcomes by an absolute 2% without adverse effects in patients who were stable after an acute coronary syndrome.

Proprotein convertase subtilisin/kexin type 9 (PCSK9), a serine protease that is produced primarily in the liver, controls levels of LDL cholesterol by binding to hepatic LDL receptors and promoting their degradation. PCSK9 antagonists lower LDL cholesterol levels by about 50% or more, regardless of whether they are added to diet therapy, statin therapy, or combined statin plus ezetimibe.[14] Ongoing clinical trials will assess whether they have an equivalent effect on cardiovascular outcomes.

Bile acid sequestrants lower LDL and decrease cardiovascular event rates. They are especially effective in combination with statins. Acceptance by patients is limited because of gastrointestinal side effects. These agents also tend to elevate triglycerides and decrease the absorption of some drugs. The sequestrant colesevelam may have less of an effect on the absorption of other drugs and has been demonstrated to lower glucose concentration in patients with type 2 diabetes.

Fish oils may be used as an adjunct in the therapy of patients with very high triglyceride levels. These omega-3 fatty acids, which occur naturally in cold-water fish, may work by activating PPARα. They may have beneficial effects on cardiovascular risk, but low-dose fish oil treatment of patients with diabetes or at risk for diabetes, most of whom were receiving statins, did not decrease cardiovascular events.[15]

PRIMARY PREVENTION

The results of clinical trials strongly support lipid lowering for secondary prevention, which is decreasing events in those with known disease. Data are less compelling for primary prevention, which is decreasing events in those without known disease. Use of a statin to lower lipids is clearly effective for primary prevention in at least three groups: patients with multiple risk factors, individuals with type 2 diabetes, and middle-aged and older people with relatively low LDL levels in the setting of elevated C-reactive protein. In the Anglo-Scandinavian Cardiac Outcomes Trial–Lipid-Lowering Arm, subjects with hypertension and at least three other risk factors had fewer events with an LDL level of 90 mg/dL achieved with atorvastatin than with an LDL level of about 130 mg/dL in control subjects. In the Collaborative Atorvastatin Diabetes Study, people with type 2 diabetes had fewer events with an LDL level of about 75 mg/dL achieved with atorvastatin compared with placebo-treated subjects with an LDL level of about 119 mg/dL. In the Justification for the Use of Statins in Prevention: An Intervention Trial Evaluating Rosuvastatin, men older than 50 years and women older than 60 years had fewer events and were less likely to die with an LDL level of about 55 mg/

dL achieved with rosuvastatin compared with placebo-treated subjects with an LDL level of about 108 mg/dL. An analysis of 18 primary trials involving more than 56,000 participants concluded that treatment of subjects with no evidence of cardiovascular disease with statins reduced all-cause mortality and vascular events with no evidence of excess adverse events.

The 2013 guidelines recommend statin therapy for primary prevention in individuals with an LDL cholesterol level above 190 mg/dl, those with diabetes, and those with a 10-year risk of atherosclerotic cardiovascular disease of at least 7.5% (see Table 206-5).

There does not appear to be a threshold effect of LDL on atherogenesis. In people with established coronary heart disease and those with diabetes, reaching very low LDL cholesterol levels is desirable. The absolute value at which maximal benefit is reached is unknown, and higher doses of statins are associated with more side effects (especially increased serum transaminases), but some analyses suggest that an LDL level of 40 mg/dL represents no increased risk of vascular disease.

Grade A References

A1. de Vries FM, Denig P, Pouwels KB, et al. Primary prevention of major cardiovascular and cerebrovascular events with statins in diabetic patients: a meta-analysis. *Drugs*. 2012;72:2365-2373.

A2. Mills EJ, O'Regan C, Eyawo O, et al. Intensive statin therapy compared with moderate dosing for prevention of cardiovascular events: a meta-analysis of >40,000 patients. *Eur Heart J*. 2011;32:1409-1415.

A3. Preiss D, Seshasai SR, Welsh P, et al. Risk of incident diabetes with intensive-dose compared with moderate-dose statin therapy: a meta-analysis. *JAMA*. 2011;305:2556-2564.

A4. Heart Protection Study Collaborative Group. Effects on 11-year mortality and morbidity of lowering LDL cholesterol with simvastatin for about 5 years in 20,536 high-risk individuals: a randomised controlled trial. *Lancet*. 2011;378:2013-2020.

A5. Landray MJ, Haynes R, Hopewell JC, et al. Effects of extended-release niacin with laropiprant in high-risk patients. *N Engl J Med*. 2014;371:203-212.

A6. Špinar J, Špinarová L, Vítovec J. IMProved Reduction of Outcomes: Vytorin Efficacy International Trial (studie IMPROVE-IT). *Vnitr Lek*. 2014;12:1095-1101.

A7. Taylor F, Huffman MD, Macedo AF, et al. Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev*. 2013;1:CD004816.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Thanassoulis G, Campbell CY, Owens DS, et al. Genetic associations with vascular calcification and aortic stenosis. *N Engl J Med*. 2013;368:503-512.
2. Roth M, McKenney JM, Hanotin C, et al. Atorvastatin with or without an antibody to PCSK9 in primary hypercholesterolemia. *N Engl J Med*. 2012;367:1891-1900.
3. The Myocardial Infarction Genetics Consortium Investigators. Inactivating mutations in *NPC1L1* and protection from coronary heart disease. *N Engl J Med*. 2014;371:2072-2082.
4. Schwartz GG, Olsson AG, Abt M, et al. Effects of dalcetrapib in patients with a recent acute coronary syndrome. *N Engl J Med*. 2012;367:2089-2099.
5. Stone NJ, Robinson J, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association task force on practice guidelines. *Circulation*. 2014;129:S1-S45.
6. Stone NJ, Robinson JG, Lichtenstein AH, et al. Treatment of blood cholesterol to reduce atherosclerotic cardiovascular disease risk in adults: synopsis of the 2013 American College of Cardiology/American Heart Association cholesterol guideline. *Ann Intern Med*. 2014;160:339-343.
7. Keany JF, Curfman GD, Jarcho JA. A pragmatic view of the new cholesterol treatment guidelines. *N Engl J Med*. 2014;370:275-278.
8. Ridker PM, Pradhan A, MacFadyen JG, et al. Cardiovascular benefits and diabetes risks of statin therapy in primary prevention: an analysis from the JUPITER trial. *Lancet*. 2012;380:565-571.
9. The Look AHEAD Research Group. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *N Engl J Med*. 2013;369:145-154.
10. Nielsen SF, Nordestgaard BG, Bojesen SE. Statin use and reduced cancer-related mortality. *N Engl J Med*. 2012;367:1792-1802.
11. Mangravite LM, Engelhardt BE, Medina MW, et al. A statin-dependent QTL for *GATM* expression is associated with statin-induced myopathy. *Nature*. 2013;502:377-380.
12. Zhang H, Plutzky J, Skentzos S, et al. Discontinuation of statins in routine care settings: a cohort study. *Ann Intern Med*. 2013;158:526-534.
13. The ACCORD Study Group. Effects of combination lipid therapy in type 2 diabetes mellitus. *N Engl J Med*. 2010;362:1563-1574.
14. Blom DJ, Hala T, Bolognese M, et al. A 52-week placebo-controlled trial of evolocumab in hyperlipidemia. *N Engl J Med*. 2014;370:1809-1819.
15. The ORIGIN Trial Investigators. n-3 fatty acids and cardiovascular outcomes in patients with dysglycemia. *N Engl J Med*. 2012;367:309-318.

REVIEW QUESTIONS

1. A 46-year-old man presents with hypercholesterolemia. He underwent coronary artery bypass grafting at the age of 43 years. He has a 20 pack-year history of smoking but quit tobacco after his bypass surgery. His father died suddenly of unknown causes at the age of 45 years, and an older brother had a myocardial infarction in his 50s. The patient is currently taking amlodipine, metoprolol, and aspirin. He took some type of statin in the past but did not like the way it made him feel. His examination is notable for a body mass index of 25, blood pressure of 150/92 mm Hg, arcus corneae, a fourth heart sound on cardiac examination, no evidence of heart failure, and thickened Achilles tendons. His laboratory results are notable for normal renal and liver function, normal fasting glucose concentration, total cholesterol level of 290 mg/dL, low-density lipoprotein (LDL) level of 232 mg/dL, high-density lipoprotein (HDL) level of 44 mg/dL, and triglyceride level of 135 mg/dL. Which course of action might be appropriate?

A. Noninvasive imaging to assess the status of his bypass grafts before deciding on a medical regimen.
B. Referral to a dietitian (with a goal of decreasing dietary cholesterol and saturated fat) and initiation of statin therapy with rapid escalation to maximum recommended doses.
C. Initiation of fibrate therapy, given his apparent intolerance of a statin.
D. Referral to a dietitian, optimization of blood pressure control, and gradual introduction of a statin in combination with one or more additional agents for LDL lowering.

Answer: D This patient has heterozygous familial hypercholesterolemia, caused by the presence of a single copy of a deficient allele for the LDL receptor. Whereas the development of atherosclerosis can vary in familial hypercholesterolemia, this individual is known to have vascular disease, and imaging will be unlikely to affect his lipid management. Dietary advice is warranted, but a statin alone or a fibrate alone will be inadequate for LDL lowering. Therapeutic lifestyle changes, management of other risk factors for atherosclerosis (such as hypertension), and initiation of lipid therapy that includes a statin as well as agents with complementary mechanisms of action (such as a bile acid sequestrant and niacin) are most appropriate.

2. A 39-year-old man is referred for management of high cholesterol. He was recently hospitalized for uncontrolled diabetes and severe abdominal pain. Review of the electronic medical record from that admission revealed no previous history of diabetes, peak glucose concentration of 390 mg/dL, no evidence of diabetic ketoacidosis, pancreatitis detected on abdominal computed tomography scanning, initial total cholesterol level of 945 mg/dL, triglyceride level of 4732 mg/dL, and hemoglobin A_{1c} level of 10.5%. He was made NPO, was treated with insulin, and improved during several days. He was discharged with a total cholesterol level of 371 mg/dL and triglyceride level of 947 mg/dL, with a laboratory notation that the sample was lipemic. His discharge medications included a high dose of rosuvastatin, long-acting insulin, and metformin. Which of the following statement(s) about this presentation is (are) true?

A. The elevated total cholesterol level in this patient indicates a very high risk of cardiovascular disease.
B. Skin lesions, transient hepatosplenomegaly, mild cognitive impairment, and abnormal eye findings are associated with this condition.
C. Elevated triglycerides in this condition should be treated with plasmapheresis.
D. Appropriate therapy may result in nearly normal lipid levels in this patient.

Answers: B and D This patient has the chylomicronemia syndrome, usually due to a genetic predisposition to hypertriglyceridemia in combination with uncontrolled diabetes. These patients may have xanthomas (especially on the trunk), organomegaly, confusion, and lipemia retinalis on funduscopic examination. The presence of chylomicrons is suggested by the laboratory detection of lipemia. The very high levels of cholesterol probably do not indicate a high risk of vascular disease because they reflect the presence of chylomicrons that are thought to be too large to enter the vascular wall. Pheresis is not usually indicated. Potent statins such as rosuvastatin are not appropriate initial therapy. Optimized glycemic control and some combination of a fibrate, niacin, and fish oils are indicated to maintain fasting triglyceride levels below 500 mg/dL, a value thought to be associated with less risk of lipid-induced pancreatitis. Many of these patients are obese, and significant weight loss can result in near-normalization of lipid levels.

3. A 44-year-old woman presents with elevated triglycerides. Except for childbirth, she has never been hospitalized. She admits to being sedentary but wants to lose weight. She denies chest pain or other symptoms. She has had abnormal lipids for several years and previously tried niacin but could not tolerate the side effect of flushing. She does not drink alcohol or use tobacco and takes no medications except for ibuprofen for menstrual cramps. He father died at the age of 58 years, around the time of an elective surgical procedure. She has three sisters, and all have elevated triglycerides, although none is known to have heart disease. There is no family history of pancreatitis. Her examination findings are normal except for a body mass index of 34. Her fasting lipids are notable for triglyceride level of 318 mg/dL, total cholesterol level of 243 mg/dL, LDL cholesterol level of 126 mg/dL, and HDL cholesterol level of 34 mg/dL. Appropriate management could include which of the following?

A. Noninvasive imaging to assess the possibility of coronary artery disease.
B. Therapeutic lifestyle changes for 6 months followed by fibrate therapy, given her intolerance of niacin by history.
C. Therapeutic lifestyle changes for 6 months followed by initiation of prescription fish oil therapy at low doses with gradual escalation to the maximum recommended dose.
D. Therapeutic lifestyle changes and initiation of statin therapy at this visit.

Answer: D This patient has familial combined hyperlipidemia, perhaps the most common inherited lipid disorder. It is associated with premature coronary disease. She has no symptoms, so noninvasive cardiac imaging may not be useful, but a treadmill stress test might be helpful as she considers starting an exercise program as part of therapeutic lifestyle changes. Statin therapy should be recommended. Whereas her LDL level is not particularly high, her non-HDL cholesterol level is 209 mg/dL. An appropriate goal might be a non-HDL cholesterol level of less than 130 mg/dL.

4. A 67-year-old man presents to you with statin intolerance. He underwent coronary artery bypass grafting at the age of 55 years and has taken simvastatin followed by atorvastatin for many years with no symptoms. Recent angiography demonstrated patent grafts. Several months ago, he developed vague aching and cramping in his lower extremities that was attributed to his statin. His atorvastatin was discontinued and the symptoms improved, but they did not resolve. Other statins were prescribed, and all were associated with increased muscle cramping, especially noticed at night. He has now been off statins for a month and currently complains of muscle pain, constipation, and recent weight gain of about 10 pounds that he attributes to physical inactivity due to muscle pain. His examination is notable for a blood pressure of 142/94 mm Hg with two antihypertensives, pulse of 58, periorbital edema, no evidence of heart failure, mild diffuse muscle tenderness without fasciculations, and trace lower extremity edema. His laboratory results reveal an LDL cholesterol level of 157 mg/dL and a creatine kinase level of 351 μg/L (normal, <120 μg/L). What is the best next step for this patient?

A. Admission to the hospital for management of rhabdomyolysis
B. Muscle biopsy
C. Initiation of ezetimibe to lower LDL cholesterol
D. Assessment of renal and thyroid function

Answer: D This patient has hypothyroidism. He does not have rhabdomyolysis, which is associated with extreme elevations of muscle enzymes, severe muscle pain, dark urine, and renal dysfunction. A muscle biopsy is not indicated at this point. Ezetimibe might be a useful adjunct, but its use does not address his muscle symptoms. The current level of creatine kinase elevation is modest, but it is prudent to assess renal function because this patient may have had greater degrees of muscle dysfunction while taking statins. In this patient, thyroid-stimulating hormone was elevated and free thyroxine was decreased, consistent with primary hypothyroidism. This condition is common, increases the risk for statin-induced myopathy, and causes myopathy independent of statins. Once thyroid status is corrected with thyroid hormone, statin therapy, reintroduced cautiously, can be well tolerated.

5. A 56-year-old woman with type 2 diabetes presents to your office for routine care. She is postmenopausal and has intermittent vasomotor symptoms but does not wish to take estrogens because of concerns about side effects. She smoked less than a pack of cigarettes a day in her 20s but quit more than 20 years ago. Her most recent hemoglobin A_{1c} level is 7.4%, she does not have microalbuminuria, and there is no personal history of heart disease, but both of her parents died of cardiovascular causes after the age of 65 years. She has no complaints except for knee pain due to osteoarthritis. Her current medications include losartan, exenatide, and metformin. Except for a body mass index of 29, her examination findings are normal. Her most recent lipid panel is notable for triglyceride level of 145 mg/dL, HDL level of 50 mg/dL, and LDL level of 96 mg/dL. She realizes that cardiovascular disease is the most common cause of death in diabetes but does not wish to take statins because her LDL is already below 100 mg/dL and she is wary of statins, especially since they are reported to cause diabetes. Which of the following represents appropriate advice for this patient?

A. Begin statin therapy with a goal of decreasing LDL by 30 to 40% from the current level of 96.
B. Begin insulin therapy with a goal of lowering the hemoglobin A_{1c} to a nearly normal level.
C. Institute an aggressive program of weight loss through diet and exercise to decrease the risk of cardiovascular disease.
D. Begin nutritional supplements including chromium, antioxidant vitamins, and low-dose fish oils.

Answer: A Even if lipid levels are initially fairly low, patients at risk for coronary disease, like this woman with diabetes, benefit from statin treatment. Aggressive attempts to improve glycemia in this population of patients do not clearly provide benefit and may cause harm. A controlled trial of weight loss did not decrease cardiovascular events in subjects similar to this patient. Supplements including minerals, vitamins, and fish oils have not been shown to provide cardiovascular benefit in individuals like this patient.

207

GLYCOGEN STORAGE DISEASES

DAVID A. WEINSTEIN

DEFINITION

Glycogen, a highly branched polymer of glucose, is the storage form of glucose in mammals. The major sites of glycogen deposition are skeletal muscle and liver. Several other tissues and organs, including the heart, smooth muscle, kidney, and intestine, are sites of glycogen synthesis that can be impaired in the glycogen storage diseases (GSDs).

EPIDEMIOLOGY

The overall frequency of the GSDs is approximately 1 case per 20,000 to 25,000 births. Sixteen distinct types have been identified, which are referred to either by the deficient enzyme or by a numbering system that reflects the historical sequence of their description. They are all uncommon and some are extremely rare. Six types account for approximately 97% of GSD cases: GSD I (25%), GSD II (15%), GSD III (24%), GSD IV (3%), and GSD VI and IX (30%). It is likely, however, that the mild forms of GSD are under-recognized.

PATHOBIOLOGY

Glucose transporter type 2 (GLUT2) predominates in the liver (and pancreatic beta cells) and has a high K_m (≈15 to 20 mmol/L); consequently, the free glucose concentration in hepatocytes increases in direct proportion to the increase in plasma glucose concentration. Glucose is rapidly phosphorylated by glucokinase to form glucose 6-phosphate, which is converted to glucose 1-phosphate, the starting point for glycogen synthesis (Fig. 207-1). Hepatic glycogen synthase catalyzes the formation of α-1,4 linkages that elongate the chains of glucose molecules. A branching enzyme leads to formation of α-1,6 linkages at branch points along the chain. The concentration of GLUT4 in the plasma membrane of skeletal muscle increases markedly after exposure

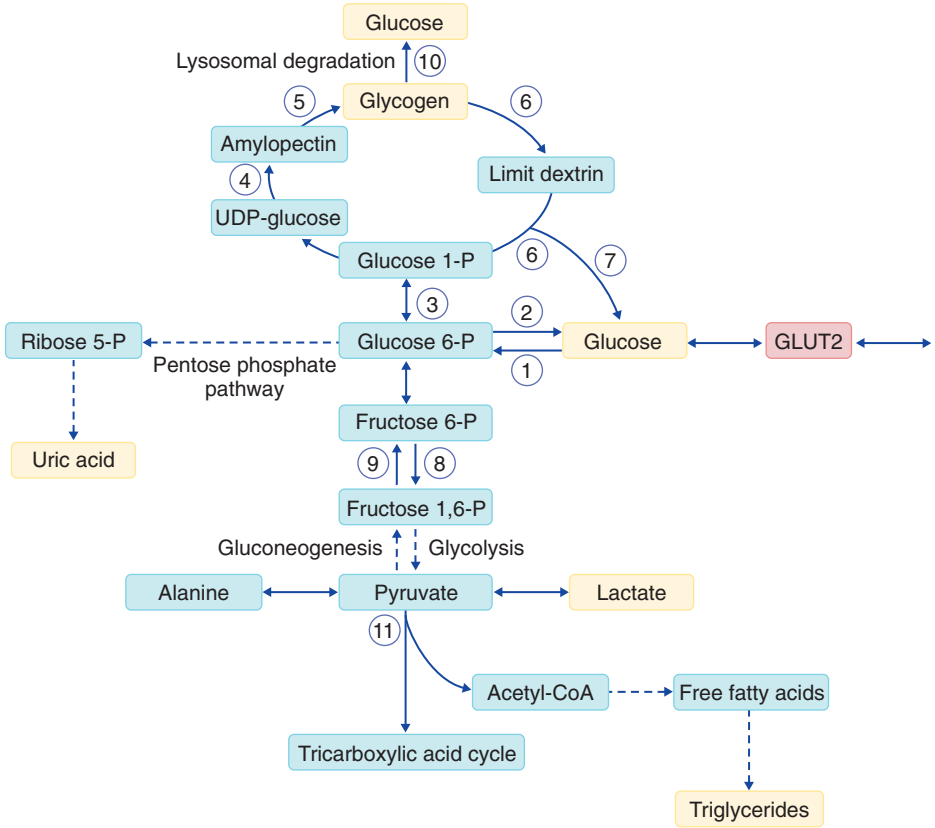

FIGURE 207-1. Simplified scheme of glycogen synthesis and degradation in the liver. Note that in skeletal muscle, GLUT4 transports glucose across the cell membrane and glucose-6-phosphatase is absent. UDP-glucose = uridine diphosphoglucose; 1, hexokinase/glucokinase; 2, glucose-6-phosphatase; 3, phosphoglucomutase; 4, glycogen synthase; 5, branching enzyme; 6, glycogen phosphorylase; 7, debranching enzyme; 8, phosphofructokinase; 9, fructose-1,6-bisphosphatase; 10, acid maltase; 11, pyruvate dehydrogenase.

to insulin and in response to exercise, resulting in increased glucose transport into skeletal muscle, where it is either oxidized to provide energy for contracting muscle or converted to glycogen.

In the intervals between meals and during the overnight fast, a cascade of enzymatic reactions (including adenylate cyclase, phosphorylase *b* kinase, and cyclic adenosine monophosphate–dependent protein kinase) activate hepatic glycogen phosphorylase, the rate-limiting enzyme in glycogenolysis, leading to the formation of glucose 6-phosphate. Glucose-6-phosphatase catalyzes the terminal reaction of both glycogenolysis and gluconeogenesis, the hydrolysis of glucose 6-phosphate, thereby allowing glucose to be released from the liver into the systemic circulation. This process is critically important for the maintenance of glucose homeostasis. Because muscle lacks glucose-6-phosphatase, it cannot release glucose for systemic use. Muscle glycogen is used to meet the energy requirement of contracting muscle and is a source of lactate, pyruvate, and alanine for gluconeogenesis early in starvation. The rate of glycogenolysis in muscle is most rapid during the first 5 to 10 minutes of exercise. As exercise continues and blood flow to muscle increases, blood-borne substrates (glucose and free fatty acids) become increasingly important sources of energy.

The GSDs or glycogenoses comprise several inherited disorders of glycogen synthesis or degradation. All are autosomal recessive with the exception of a subtype of GSD IX that is X-linked. The GSDs are caused by mutations in the genes that code for enzymes involved in the synthesis or degradation of glycogen and may involve the liver, skeletal muscle, and kidney. They are all characterized by an abnormal tissue concentration or structure (or both) of the glycogen molecule.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Hepatomegaly and hypoglycemia[1] are the principal clinical manifestations of the hepatic glycogenoses; muscle cramps, exercise intolerance, easy fatigability, and progressive weakness are the major manifestations of the muscle glycogenoses. Features of the most common GSDs are shown in Table 207-1. Molecular genetic testing performed on DNA extracted from blood or saliva is used to diagnose all the common forms of GSD.

TREATMENT

Rx

The goal of treatment of the hepatic forms of GSD is to prevent hypoglycemia and glucose counter-regulation. The specific details of therapy principally depend on whether normal gluconeogenesis can occur. In GSD I, abnormal glucose-6-phosphatase activity impairs both glycogenolysis and gluconeogenesis. In contrast, gluconeogenesis is intact in the other liver forms of GSD, allowing protein to be used as a substrate for endogenous glucose production. Fatty acid oxidation is also intact in all types of GSD except for GSD I, resulting in ketone formation during periods of hypoglycemia.[2]

TABLE 207-1 PRINCIPAL FEATURES OF THE COMMON GLYCOGEN STORAGE DISEASES

TYPE AND DEFECTIVE ENZYME	CHARACTERISTIC CLINICAL FEATURES	HIGH-RISK POPULATIONS	THERAPY
0 *Hepatic glycogen synthase*	Liver small or normal in size, fasting ketotic hypoglycemia, postprandial hyperglycemia and hyperlactatemia	French Canadian Italians	UCS, especially at bedtime, with high-protein diet
Ia *Glucose-6-phosphatase* *von Gierke's disease*	Hepatomegaly, failure to thrive, growth retardation, severe hypoglycemia, lactic acidosis, hyperuricemia, hyperlipidemia	Ashkenazi Jews Mexicans Chinese Japanese	UCS during the day and night or continuous overnight intragastric feeding
Ib *Glucose 6-phosphate transporter*	Same as type Ia; also neutropenia, recurrent bacterial infections, and inflammatory bowel disease	Native Americans Iranian Jews Italians	UCS as for Glycogen Storage Disease Ia; granulocyte colony-stimulating factor
II *Lysosomal acid maltase (α-glucosidase)* *Pompe's disease*	Infantile form is characterized by severe generalized hypotonia, muscle weakness, and hypertrophic cardiomyopathy leading to cardiorespiratory failure usually by 1 year of age. Skeletal myopathy with slowly progressing muscle weakness is the primary clinical manifestation of the juvenile- and adult-onset forms. Serum creatine kinase is markedly increased.	None	Intravenous enzyme replacement with recombinant human α-glucosidase
III *Debranching enzyme* *Cori's or Forbes' disease*	Hepatomegaly, moderate to severe ketotic hypoglycemia, muscle weakness and wasting, hypertrophic cardiomyopathy (IIIa), increased transaminases; without muscle involvement (IIIb)	Faroe Islanders First Nation (Canada) Indian subcontinent	High-protein diet with low-dose UCS supplementation
IV *Glycogen branching enzyme* *Andersen's disease*	A clinically heterogeneous disorder The typical presentation is liver disease in early childhood progressing to lethal cirrhosis. The less common neuromuscular presentation is distinguished by age at onset into 4 groups: perinatal, congenital, childhood, and adult.	None	No specific treatment Liver transplantation has resulted in decreased glycogen storage in heart and skeletal muscle.
V *Muscle glycogen phosphorylase* *McArdle's disease*	Symptoms usually begin in adolescence or early adulthood with exercise intolerance, fatigue, myalgia, muscle cramps, and muscle swelling. Transient myoglobinuria due to rhabdomyolysis may occur after exercise. Severe myoglobinuria may lead to acute renal failure. Later in adult life, persistent and progressive muscle weakness and atrophy with fatty replacement occur. Serum creatine kinase is increased.	None	High-protein diet (50% carbohydrate and 25-30% protein) Oral sucrose before sustained aerobic exercise may be beneficial.
VI *Hepatic glycogen phosphorylase* *Hers' disease*	Hepatomegaly; growth retardation; moderate ketotic hypoglycemia; increased serum transaminases, cholesterol, and triglycerides	Mennonites	UCS dosed to prevent hypoglycemia and ketosis
VII *Muscle phosphofructokinase* *Tarui's disease*	Manifested in childhood with fatigue, muscle cramps, exercise intolerance; rhabdomyolysis and myoglobinuria with strenuous exertion; increased serum creatine kinase; may have mild hemolytic anemia and mild hyperbilirubinemia; hyperuricemia	None	No specific treatment; avoid strenuous exercise
IX *Phosphorylase b kinase*	Hepatomegaly, mild ketotic hypoglycemia, growth retardation, increased serum transaminases, hypercholesterolemia, hypertriglyceridemia May be X-linked or autosomal recessive	None	UCS dosed to prevent hypoglycemia and ketosis; high-protein diet to normalize prealbumin

UCS = uncooked cornstarch.

Treatment of GSD I consists of providing a continuous dietary source of glucose to maintain blood glucose levels at 75 to 90 mg/dL before meals and overnight. Glucose concentrations must be maintained above 70 mg/dL to prevent counter-regulation, which causes shunting of glucose 6-phosphate into alternative pathways, resulting in hyperlactacidemia, hyperuricemia, and hypertriglyceridemia. In infants, continuous glucose can be provided by frequent feeds during the day and continuous intragastric feeds at night through a nasogastric or gastrostomy tube. Beginning at 6 to 12 months of age, uncooked cornstarch (UCS), which is slowly digested and absorbed into the circulation as glucose, can be used as an alternative method of continuously providing glucose. Initially, UCS is given every 3 hours. As children age and guided by the results of periodic blood glucose and lactate monitoring, the interval between feeds eventually is increased to 4 to 6 hours. An extended-release cornstarch preparation allows many older patients to sleep through the night without awakening.[3] Galactose and fructose must be restricted because they cannot be converted to glucose, and consumption of large quantities may exacerbate the biochemical derangements. Optimal care usually ameliorates all biochemical abnormalities; however, if optimal dietary management does not lower serum uric acid and triglycerides to acceptable levels, treatment with allopurinol and gemfibrozil, respectively, is indicated. Neutropenia (Chapter 167) in type Ib responds well to low-dose granulocyte colony-stimulating factor (G-CSF) therapy; however, untoward effects of G-CSF may include splenomegaly and very rare cases of leukemia.[4] The recommended starting dose (2.5 µg/kg/day) is therefore lower than in other conditions, and the lowest possible dose that prevents infections is used. An enterocolitis that resembles Crohn's disease (Chapter 141) occurs almost universally in GSD type Ib, and mesalamine (Pentasa) is the first-line therapy because small intestinal disease predominates.

Patients with the other forms of GSD use a high-protein diet (2 or 3 g/kg) supplemented with complex carbohydrates and UCS, which, typically, is administered every 6 to 8 hours to maintain glucose concentrations above 75 mg/dL. Because β-oxidation of fatty acids can occur in these forms of GSD, ketosis can develop rapidly and UCS doses are titrated to maintain a normal blood ketone concentration (<0.3 mmol/L). Protein dosing is aimed at normalization of total protein and prealbumin concentrations. Strict avoidance of fructose and sucrose is unnecessary; however, intake of simple sugars is still discouraged to avoid excessive storage of glycogen. This is particularly important in GSD III because excessive glycogen storage has been associated with worsening of the associated hypertrophic cardiomyopathy.[5]

Treatment of the muscle glycogenoses is shown in Table 207-1.

Prevention of Complications

Despite improvements in care, long-term complications are common in GSD I and III. There is increasing evidence, however, that complications can be delayed or even prevented with optimal metabolic control.[6] Hepatic adenomas can develop in patients with GSD I during adolescence or early adulthood and may gradually enlarge, undergo malignant transformation, or hemorrhage into the peritoneal cavity. Nephrocalcinosis and nephrolithiasis, caused by hypocitraturia, are also common and can be prevented by oral citrate supplementation. Maintenance of optimal metabolic control can prevent renal tubular dysfunction, focal segmental glomerulosclerosis, anemia, gout, and osteoporosis.[7,8]

In type III GSD, a hypertrophic cardiomyopathy can develop. The cardiac disease appears to be caused by overstorage of glycogen, and restriction of simple sugars and carbohydrates has resulted in normalization of cardiac function. Hepatic adenomas develop in 10% of patients, and hepatocellular cancer is rare. Most patients with type III do not have myopathic symptoms during childhood and early adulthood. Progressive myopathy can develop beginning in the teenage years, and it can become debilitating. A very high protein diet (3 or 4 g/kg) may slow the progression of the muscle disease.[2]

Short stature and osteoporosis are the only common complications in GSD O, VI, and IX, but these can be prevented by maintenance of optimal metabolic control and avoidance of ketosis.[9] Cirrhosis has been described as a complication in untreated patients with GSD IX, but scarring may be prevented with treatment.[10]

PROGNOSIS

The prognosis for all of the hepatic GSDs is now excellent. Almost all complications can be delayed or prevented with optimal metabolic control. Patients are doing well into adulthood, and pregnancies have now become routine.[11] Liver transplantation should be viewed as a treatment of last resort,[12] especially because gene therapy may be available for treatment of these disorders in the future.

The Association for Glycogen Storage Disease website (*http://www.agsdus.org*) provides basic information about the GSDs intended to be of use to people affected by one of the GSDs, their families, and other interested parties. Similar organizations exist in the United Kingdom, France, Spain, the Netherlands, Germany, Italy, Sweden, the Faroe Islands, Poland, and Russia.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Brown LM, Corrado MM, van der Ende RM, et al. Evaluation of glycogen storage disease as a cause of ketotic hypoglycemia in children. *J Inherit Metab Dis*. 2014; [Epub ahead of print].
2. Kishnani PS, Arn P, Austin S, et al. Glycogen storage disease type III: diagnosis and management guidelines. *Genet Med*. 2010;12:446-463.
3. Shah KK, O'Dell SD. Effect of dietary interventions in the maintenance of normoglycaemia in glycogen storage disease type 1a: a systematic review and meta-analysis. *J Hum Nutr Diet*. 2013;26:329-339.
4. Jun HS, Weinstein DA, Lee YM, et al. Molecular mechanisms of neutrophil dysfunction in glycogen storage disease type Ib. *Blood*. 2014;123:2843-2853.
5. Sentner CP, Caliskan K, Vletter WB, et al. Heart failure due to severe hypertrophic cardiomyopathy reversed by low calorie, high protein dietary adjustments in a glycogen storage disease type IIIa patient. *JIMD Rep*. 2012;5:13-16.
6. Wang DQ, Fiske LM, Carreras CT, et al. Natural history of hepatocellular adenoma formation in glycogen storage disease type I. *J Pediatr*. 2011;159:442-446.
7. Wang DQ, Carreras CT, Fiske LM, et al. Characterization and pathogenesis of anemia in glycogen storage disease type Ia and Ib. *Genet Med*. 2012;14:795-799.
8. Minarich LA, Kirpich A, Fiske LM, et al. Bone mineral density in glycogen storage disease type Ia and Ib. *Genet Metab*. 2012;14:737-741.
9. Bali DS, Goldstein JL, Fredrickson K, et al. Variability of disease spectrum in children with liver phosphorylase kinase deficiency caused by mutations in the PHKG2 gene. *Mol Genet Metab*. 2014;111:309-313.
10. Tsilianidis LA, Fiske LM, Siegel S, et al. Aggressive therapy improves cirrhosis in glycogen storage disease type IX. *Mol Genet Metab*. 2013;109:179-182.
11. Ferrecchia IA, Guenette G, Potocik EA, et al. Pregnancy in women with glycogen storage disease Ia and Ib. *J Perinat Neonatal Nurs*. 2014;28:26-31.
12. Boers SJB, Visser G, Smit GPA, et al. Liver transplantation in glycogen storage disease type I. *Orphanet J Rare Dis*. 2014;9:47.

REVIEW QUESTION

1. Which of the following symptoms and signs would *not* be a likely presenting manifestation of glycogen storage disease in an adult?

A. Unexplained hepatomegaly
B. Limb-girdle muscle weakness
C. Transient ischemic attacks
D. Symptoms of hypoglycemia
E. Muscle cramps

Answer: C Most cases of glycogen storage disease seen by internists will have been already diagnosed in childhood. However, individuals with much milder disease (presumably associated with less severe enzyme activity as determined by genotype) may present with symptoms and signs in adulthood. Hepatomegaly (A) and hypoglycemia (D) are the principal clinical manifestations of the hepatic glycogenoses; muscle cramps (E), exercise intolerance, and progressive fatigue and weakness are the major manifestations of the muscle glycogenoses (see Table 207-1). For example, whereas infants with Pompe's disease (type II glycogen storage disease) present at birth (floppy infant syndrome) or shortly thereafter and progress to respiratory failure, cardiomyopathy, and death by the age of 1 year unless treated, adult-onset disease commonly is manifested as limb-girdle muscle weakness (B), which can mimic that of other musculoskeletal disorders, especially muscular dystrophies. (Vissing J, Lukacs Z, Straub V. Diagnosis of Pompe disease. JAMA Neurol. 2013;70:923-927.) Cerebrovascular disease (C) may be seen in homocystinuria but is not known to be associated with glycogen storage diseases.

208

LYSOSOMAL STORAGE DISEASES

DONNA M. KRASNEWICH AND ELLEN SIDRANSKY

The lysosomal storage diseases encompass a group of more than 45 different inherited disorders, all sharing a defect in lysosomal function. Lysosomes are acidic, membrane-bound organelles present in the cytoplasm containing enzymes that degrade macromolecules. These disorders ensue when one or more of the hydrolytic enzymes are deficient or when essential lysosomal transporters, receptors, cofactors, or protective proteins are defective or lacking. Typically, complex macromolecules, including glycolipids, mucopolysaccharides, and glycoproteins are delivered to the lysosome, where they undergo sequential modification by a series of hydrolases. An enzymatic deficiency becomes clinically important when macromolecules accumulate owing to inadequate degradation. Different categories of defects resulting in lysosomal dysfunction are encountered in the lysosomal storage disorders; examples of each type are listed in Table 208-1.

Although most lysosomal storage disorders are rare, as a group their frequency is estimated to be 1 per 7000 to 8000 live births. This is an underestimate because milder or attenuated forms of these disorders are often not identified. All of the disorders have an autosomal recessive pattern of inheritance, with the exception of Fabry disease and Hunter's syndrome (mucopolysaccharidosis II), which are X-linked recessive, and Danon's disease, caused by mutations in the lysosome-associated membrane protein 2 (LAMP-2), which is inherited in an X-linked dominant manner. As a whole, these disorders are characterized by a vast spectrum of manifestations, often causing them to evade diagnosis. Many were traditionally classified into infantile, juvenile, and adult types on the basis of the patient's age at the onset of manifestations, but atypical presentations complicate these distinctions. Among the factors contributing to this phenotypic diversity are the amount of residual enzyme activity, the cellular localization of the enzyme, the genotype, and the genetic background of the affected individual as well as other genetic, environmental, and epigenetic influences.[1]

With the advent of new therapies for some of the lysosomal storage disorders, early establishment of the diagnosis is paramount. Suggestive clinical findings include coarse facial features; organomegaly; specific eye findings, including corneal clouding or a cherry red spot (of the macula); cytopenia; and skeletal abnormalities, notably dysostosis multiplex. Disorders associated with each of these findings are listed in Table 208-2. There should be a greater index of suspicion whenever these features occur in concert, the findings are progressive, there is developmental regression, or the patient appears dissimilar to other family members.

The diagnostic work-up includes a careful history, with analysis of the family pedigree and assessment of developmental milestones in childhood and adolescence. A family history of consanguinity, other affected siblings, multiple miscarriages, or early deaths can aid in making the diagnosis. Ethnicity can be a helpful clue because some of the lysosomal disorders occur with increased incidence in specific populations, such as Ashkenazi Jews (Gaucher disease type 1, Tay-Sachs' disease, mucolipidosis type IV) and Scandinavians (mannosidosis, aspartylglucosaminuria, Salla's disease, Gaucher disease type 3). On physical examination, special attention should be paid to head circumference, facial appearance, enlargement of the tongue, hepatosplenomegaly, skeletal manifestations (including kyphosis), broadening of the long bones, and stiffness of the joints. Skin evaluation may reveal angiokeratoma, especially around the umbilicus and in skin creases. The eye evaluation should include a funduscopic and slit-lamp examination as well as an assessment for atypical eye movements, which may be pathognomonic for disorders such as

TABLE 208-1 CLASSIFICATION OF LYSOSOMAL STORAGE DISORDERS BASED ON THE TYPE OF DEFECT*

SPHINGOLIPIDOSES
Fabry disease (α-galactosidase)
Farber disease (ceramidase)
GM_1 gangliosidosis/Landing's disease (β-galactosidase)
GM_2 gangliosidosis/Tay-Sachs' disease
Sandhoff's disease (α-hexosaminidases A and B)
Gaucher disease (glucocerebrosidase)
Niemann-Pick disease, types A and B (sphingomyelinase)
Metachromatic leukodystrophy (arylsulfatase A)
Krabbe's disease (β-galactocerebrosidase)
LIPID STORAGE DISORDERS
Wolman's disease (acid lipase)
Ceroid-lipofuscinosis, adult type, Kufs'/Parry's (CLN4, heterogeneous)
MUCOPOLYSACCHARIDOSES
Type I/Hurler's disease (α-L-iduronidase)
Type II/Hunter's disease (iduronate-2-sulfatase)
Type III/Sanfilippo's disease (four different enzymes in the degradation of heparan sulfate defining types A-D)
Type VI/Maroteaux-Lamy's disease (*N*-acetylgalactosamine-4-sulfatase)
Type VII/Sly's disease (α-glucuronidase)
OLIGOSACCHARIDOSES
Aspartylglucosaminuria (aspartylglucosaminidase)
Fucosidosis (α-fucosidase)
α-Mannosidosis (α-mannosidase)
Schindler's disease (α-*N*-acetylgalactosaminidase)
Sialidosis I (sialidase)
Sialidosis II/mucolipidosis I (sialidase)
MUCOLIPIDOSES
Mucolipidosis II/I-cell disease (*N*-acetylglucosaminylphosphotransferase)
Mucolipidosis III/pseudo-Hurler (*N*-acetylglucosaminylphosphotransferase)
Mucolipidosis IV (*MCOLN1* mutation)
LYSOSOMAL GLYCOGEN STORAGE
Glycogenosis type II/Pompe's disease (α-1,4-glucosidase)
LYSOSOMAL TRANSPORT DISORDERS
Sialic acid storage disease/Salla's disease (sialin/SLC17A5)
Cystinosis (cystine transporter)
Niemann-Pick disease, type C (intracellular cholesterol transport)
MULTIPLE ENZYME DEFICIENCY
Galactosialidosis (β-galactosidase and sialidase)
Multiple sulfatase deficiency/Austin's disease (sulfatases)

*Full chapters describing each of these disorders are available in Valle D, Beaudet AL, Vogelstein B, et al, eds. The Online Metabolic and Molecular Bases of Inherited Disease. http://www.ommbid.com/OMMBID.

neuronopathic Gaucher disease and Niemann-Pick disease type C. Unexplained cardiomyopathy and cryptogenic stroke may be the initial presentation of Fabry's disease. A careful neurologic and cognitive evaluation can be fruitful because some of the later presentations include dementia and psychiatric manifestations.[2] Moreover, regression of milestones can provide an early diagnostic clue. Preliminary diagnostic studies include urine for thin-layer chromatography, blood count with smear for vacuolated white blood cells, skeletal radiography, and ophthalmologic examination.

For the most part, the diagnosis of a specific lysosomal storage disorder is made by assaying enzymatic activity in a blood sample or fibroblast cell line. A lysosomal panel, evaluating the activity of multiple lysosomal enzymes from the same sample, should be the first-tier test. If a lysosomal storage disorder is still suspected, a tissue biopsy, most often of the bone marrow or liver, can be considered, although this is rarely indicated. Most of the genes encoding the lysosomal enzymes have been identified, and mutation analysis is available. However, in most cases, there is vast genotypic heterogeneity, and mutation screening is useful only when a mutation has already been identified in a specific family or when specific mutations are known to be common in a specific ethnic group.

Improved care and new therapeutic modalities have transformed the natural history of several of these disorders. With patients' increased longevity, diseases that were once encountered only by pediatricians have now made their way to the offices of internists. Moreover, many of the classic complications are now avoided by early therapeutic interventions, such as enzyme replacement therapy.[3] However, in some instances, prolonged longevity has unmasked unanticipated clinical manifestations. Physicians' increased awareness of the range of manifestations and presentations of lysosomal disorders may lessen the long delays in diagnosis that patients frequently describe. Lysosomal disorders encountered in adult patients are discussed here, with a focus on Gaucher disease and Fabry disease.[4] A brief discussion of specific disorders frequently diagnosed during adulthood as well as of childhood-onset lysosomal disorders that persist through adulthood is included.

GAUCHER DISEASE

PATHOBIOLOGY

Gaucher disease, the *autosomal* recessively inherited deficiency of the lysosomal enzyme glucocerebrosidase, is a disorder primarily of the reticuloendothelial system.[5] Lysosomes within macrophages become engorged with the substrate glucocerebroside, giving rise to the characteristic Gaucher cells, with a wrinkled-paper appearance resulting from intracytoplasmic substrate deposition. The accumulated glycolipid glucosylceramide is derived from the degradation of senescent leukocytes or erythrocyte membranes.

CLINICAL MANIFESTATIONS

Clinically, Gaucher disease has been divided into three types on the basis of the absence or presence and the rate of progression of neurologic involvement. Type 1, the non-neuronopathic form, is the most common type and can be manifested at any age. Type 2, the acute neuronopathic form, manifests before or shortly after birth and has a rapid and progressive course. Type 3 is the subacute neuronopathic form. The spectrum of manifestations encountered in this disorder ranges from asymptomatic octogenarians to infants who succumb in utero. Some patients defy classification into one of the three types. It is a pan-ethnic disorder, although type 1 Gaucher disease is more frequent among Ashkenazi Jews, in whom the carrier frequency is about 1 in 16; in contrast, the approximate carrier frequency in the general population is 1 in 100.

The gene encoding glucocerebrosidase (*GBA1*) is located on chromosome 1q21, and more than 300 different mutations have been found in patients. Several mutations are encountered with increased frequency in type 1 Gaucher disease; for example, among Ashkenazi Jews, mutation N370S is the most common allele. However, the mutations identified do not adequately account for the range of manifestations encountered.

In recent years, an association between Gaucher disease and parkinsonism (Chapter 409) has been reported.[6,7] Both patients with Gaucher disease and carriers of mutations in *GBA1* have a higher incidence of Parkinson's disease and Lewy body disorders.[8] Studies in cohorts of patients with Parkinson's disease around the world demonstrate that they have a more than five-fold increased frequency of *GBA1* mutations, rendering this the most common genetic risk factor for parkinsonism identified to date.

Commonly encountered symptoms in all types of Gaucher disease include easy bruisability, hepatomegaly, splenomegaly, chronic fatigue, and bone pain or pathologic fractures. Laboratory findings include anemia, thrombocytopenia, and elevations of ferritin, acid phosphatase, angiotensin-converting enzyme, and, at times, liver enzymes. Painless splenomegaly is the most common presentation in patients with type 1 Gaucher disease, and the spleen can be massively enlarged. Occasional patients have pulmonary involvement or pulmonary hypertension. Bone involvement is a significant cause of morbidity and can be manifested with extreme bone pain or pathologic fractures. Most patients have radiologic evidence of skeletal involvement, including the classic Erlenmeyer flask deformity of the distal end of the femur and osteopenia (Fig. 208-1A). Pathologic fractures (especially of the hip, ribs, or spine), lytic bone lesions, and osteoporosis may occur. Painful bone crises, episodes of bone infarcts, can last for weeks and may require aggressive pain management.

Type 2 disease, which is rare, is characterized by a rapid neurodegenerative course with extensive visceral involvement; death usually occurs within the first 2 years of life. The disease is diagnosed prenatally or in infancy and is associated with increased tone, strabismus, and organomegaly. Failure to thrive and compromised airway from laryngospasm are typical. Progressive psychomotor degeneration leads to death, usually secondary to an intercurrent respiratory infection and respiratory compromise.

Type 3 disease is clinically variable and is often noted in infancy or childhood. In addition to organomegaly and bone involvement, patients have abnormal horizontal eye movements, and some develop myoclonic epilepsy or neurodegenerative manifestations. A subgroup of patients has cardiac

TABLE 208-2 MANIFESTATIONS ENCOUNTERED IN DIFFERENT LYSOSOMAL STORAGE DISORDERS

FINDING	DISORDERS
Hepatosplenomegaly	GM_1 gangliosidosis, Niemann-Pick disease, Gaucher disease, Wolman's disease, fucosidosis, Pompe's disease, mannosidosis, multiple sulfatase deficiency, sialidosis, galactosialidosis, several mucopolysaccharidoses, cystinosis
Coarse facies	GM_1 gangliosidosis, fucosidosis, Pompe's disease, mannosidosis, multiple sulfatase deficiency, I-cell disease, several mucopolysaccharidoses, mucolipidosis II, sialic acid storage disease, aspartylglucosaminuria
Skeletal findings	GM_1 gangliosidosis, Gaucher disease, fucosidosis, mannosidosis, sialidosis, galactosialidosis, several mucopolysaccharidoses, I-cell disease, mucolipidosis III
Cherry red spot	Infantile forms of GM_1 gangliosidosis, Sandhoff's disease, Tay-Sachs' disease, Niemann-Pick disease, sialidosis, galactosialidosis, I-cell disease
Corneal clouding	GM_1 gangliosidosis, several mucopolysaccharidoses, mannosidosis, I-cell disease, mucolipidosis III and IV, multiple sulfatase deficiency, galactosialidosis
Cognitive impairment	GM_1 gangliosidosis, Sandhoff's disease, Tay-Sachs' disease, Niemann-Pick disease, Gaucher disease type 2, Wolman's disease, fucosidosis, mannosidosis, multiple sulfatase deficiency, sialidosis, galactosialidosis, several mucopolysaccharidoses, sialic acid storage disease, aspartylglucosaminuria, I-cell disease, mucolipidosis III and IV, Krabbe's disease, metachromatic leukodystrophy, neuronal ceroid lipofuscinosis
Hematologic	
Foam cells	GM_1 gangliosidosis, Niemann-Pick disease, Gaucher disease, acid lipase deficiency, fucosidosis
Granulated or vacuolated white blood cells	Several mucopolysaccharidoses, sialidosis, galactosialidosis. neuronal ceroid-lipofuscinosis, Niemann-Pick disease, Wolman's disease, fucosidosis, mannosidosis, aspartylglucosaminuria, I-cell disease, mucolipidosis III, multiple sulfatase deficiency
Psychiatric or behavioral manifestations	Several mucopolysaccharidoses (especially Sanfilippo's), sialidosis, galactosialidosis, Fabry disease, mannosidosis, neuronal ceroid-lipofuscinosis, metachromatic leukodystrophy, Tay-Sachs' disease, Niemann-Pick disease, type C
Newborn presentations	Gaucher disease, type 2; GM_1 gangliosidosis; Krabbe's disease; Niemann-Pick disease, types A and C; mucopolysaccharidosis I, IVA, VII; Pompe's disease; sialidosis, types I and II; mucolipidosis, types I and II; Schindler's disease; Wolman's disease; infantile sialic acid storage disease; sialuria; Salla's disease; galactosialidosis; multiple sulfatase deficiency; prosaposin deficiency

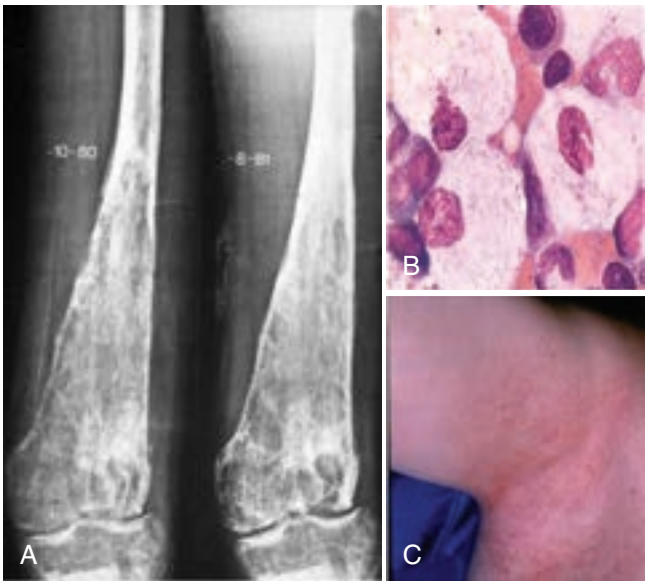

FIGURE 208-1. **A, Radiographic image showing the Erlenmeyer flask deformity in Gaucher disease. There is cortical thinning and widening of the medullary cavity of the metaphysis and adjacent diaphysis. B, Gaucher cells—reticuloendothelial cells storing abnormal amounts of lipid. C, Angiokeratomas, the nonblanching punctate skin lesion in Fabry disease.**

calcifications, hydrocephalus, and other atypical manifestations, and all carry the mutation D409H in the *GBA1* gene.

DIAGNOSIS

Gaucher disease should be considered in the differential diagnosis of patients of all ages with unexplained organomegaly, easy bruisability, or bone pain (Table 208-3). The diagnosis can be made by demonstration of deficient glucocerebrosidase activity in leukocytes or cultured cells. In some populations, particularly Ashkenazi Jews, mutation analysis can be diagnostic as mutation N370S accounts for about 70% of mutant alleles. However, the presence of a highly homologous pseudogene sequence nearby can complicate molecular analyses. Bone marrow and liver biopsies show pathologic changes (Fig. 208-1B) but are not indicated for diagnosis. Carrier identification is best achieved by DNA testing when the mutant allele is known. Prenatal diagnosis is possible by determining the enzymatic activity or specific mutations in chorionic villi or cultured amniotic fluid cells.

TABLE 208-3 SUGGESTIVE DIAGNOSTIC FEATURES IN ADULTHOOD FOR GAUCHER DISEASE AND FABRY DISEASE

GAUCHER DISEASE	FABRY DISEASE
Family member with Gaucher disease	Family history of Fabry disease
Hepatomegaly, splenomegaly (sometimes massive)	Cutaneous lesions of capillaries (angiokeratoma)
Frequent epistaxis	Hypohidrosis or heat intolerance
Easy bruising	Intermittent episodes of severe pain in the extremities (acroparesthesias)
Abnormal saccadic eye movements	Left ventricular hypertrophy of unknown etiology in young adulthood
Thrombocytopenia or anemia	Stroke of unknown etiology in young adulthood
Painful bone crisis	Chronic kidney disease of unknown etiology in young adulthood
Erlenmeyer flask deformity of the distal femur, aseptic necrosis of the femoral heads	Multiple renal sinus cysts discovered incidentally
Pathologic fractures, unexplained rib fracture	Female carriers may have more variable and less severe symptoms with later onset
Multiple myeloma	
Parkinsonism	
Elevated serum ferritin, angiotensin-converting enzyme, or tartrate-resistant acid phosphatase	

TREATMENT

Rx

Enzyme replacement with recombinant glucocerebrosidase is available for the treatment of symptomatic patients with type 1 disease (Table 208-4). Studies show that anemia, thrombocytopenia, and organomegaly are reversed within 12 to 36 months with enzyme doses between 30 and 120 IU/kg given monthly. The treatment is ongoing, administered intravenously, and extremely expensive. Asymptomatic and mildly symptomatic adults do not always require treatment. Several companies are now marketing different forms of recombinantly produced enzyme. The enzyme does not cross the blood-brain barrier and does not alter the neurologic progression of patients with neuronopathic forms of Gaucher disease, but it can still be useful in alleviating visceral manifestations. It does not prevent the development of parkinsonism. Efforts are also under way to develop alternative therapies, including substrate reduction therapy, chemical chaperones, and gene therapy. One oral form of substrate reduction therapy, an iminosugar derivative, has been approved and is efficacious for some of the systemic manifestations of type 1 Gaucher disease. Useful supportive therapies include bisphosphonates for osteoporosis, orthopedic surgery for bone fractures, and palliative therapy and hydration for bone crises. Total or partial splenectomy, once commonly performed in patients with Gaucher disease, is now rarely indicated. Bone marrow transplantation (Chapter 178) has improved systemic but not neurologic manifestations.

TABLE 208-4 ENZYME REPLACEMENT THERAPY (ERT) FOR GAUCHER DISEASE AND FABRY DISEASE

GAUCHER DISEASE	FABRY DISEASE
ERT is a costly but effective intravenous therapy generally administered every other week for life.	ERT is costly, and there are not uniform recommendations for its use.
Decreased splenic and hepatic volumes and increases in hemoglobin levels and platelet counts should be expected in the first year of treatment.	On the basis of some current trials, hemizygous males with a low or undetectable level of α-galactosidase A should be treated with ERT, whether or not clinical features are present.
Asymptomatic and mildly symptomatic adults do not always require treatment.	On the basis of current trials, female carriers and atypically affected males with clinical features of Fabry disease should be treated with ERT.
ERT does not cross the blood-brain barrier and does not correct neurologic features of neuronopathic forms of Gaucher disease.	Other trials suggest that ERT should not be started in patients with proteinuria or reduced glomerular filtration rate unless there are other findings of Fabry disease.
ERT will not prevent the development of parkinsonism.	

FABRY DISEASE

PATHOBIOLOGY

Fabry disease, an X-linked inherited deficiency of the lysosomal enzyme α-galactosidase A, has intermediate penetrance in females and is considered a systemic vascular disorder. The disease incidence is about 1 in 117,000 live male births. Evidence suggests that 50% of females with Fabry disease are either asymptomatic or are not identified. This defect in the hydrolytic cleavage of the terminal molecule of galactose from glycolipids causes lysosomal accumulation of globotriaosylceramide and galabiosylceramide in many cell types. Lysosomal inclusions or lipid deposits can be seen in vascular cells, including both endothelial and smooth muscle cells; cardiac cells, such as endocardial cells, cardiomyocytes, and cardiac valves; kidney epithelial cells, including tubular and glomerular cells and podocytes; and nerve cells, including dorsal root ganglia and some central nervous system neurons.

The gene encoding α-galactosidase A, *GLA*, is located on Xq22.1. More than 431 mutations have been described, including missense/nonsense mutations, small deletions, large deletions, splice defects, and complex rearrangements. Most affected individuals have 2 to 25% residual activity, but the most severe form of Fabry disease has been correlated with complete absence of α-galactosidase A activity.

CLINICAL MANIFESTATIONS

Clinically, angiokeratomas (nonblanching, punctate, blue-black skin lesions), debilitating pain, and corneal opacities can occur in early childhood and may lead to the diagnosis. If the diagnosis is missed, the disease can result in progressive renal and cardiac deterioration. Patients have a propensity for ischemic stroke, sometimes in their 20s but more commonly in the fourth and fifth decades of life. As with many metabolic disorders, there is a spectrum of presentations that can mimic more common disorders, and many patients are undiagnosed.

Fabry disease typically manifests in childhood in classically affected males with episodes of extremity pain. Angiokeratomas develop in adolescence, followed by advancing renal disease during adulthood. The progressive cardiac and cerebrovascular involvement accounts for a majority of the deaths in adulthood associated with Fabry disease. X-linked inactivation and the penetrance of this X-linked disorder are reflected in the fact that approximately 90% of females carrying the mutation have symptoms,[9] although affected males show more significant clinical manifestations at an earlier age than heterozygous females do.

The majority of patients experience acroparesthesia, or a "Fabry crisis," characterized by excruciating, burning pain that may be either continuous or episodic. The pain typically affects the feet first, followed by the hands, and may be triggered by exercise, stress, and extremes in environmental temperatures. Abdominal or flank pain, simulating renal colic, may occur.

Angiokeratomas (Fig. 208-1C) are often the first sign of Fabry disease and may be accompanied by hypohidrosis. These classic skin lesions increase in number and size over time and are typically most dense between the umbilicus and the knees; however, they may occur anywhere, including the buccal mucosa. Ophthalmologic findings include conjunctival and retinal tortuosity and corneal opacities (cornea verticillata). Characteristic lenticular lesions are observed during slit-lamp examination and are present in affected males and female heterozygotes. Progressive hearing loss may also occur.

Renal involvement is common and is first manifested as proteinuria, followed by progressive renal insufficiency, with birefringent "Maltese crosses" sometimes seen in the urinary sediment. Chronic kidney disease is part of the natural history, and end-stage renal disease may develop in the second to fourth decades. Fabry disease should be considered when multiple renal sinus cysts are seen on an imaging study. As affected men and women age, cardiovascular findings may include ventricular hypertrophy, conduction defects, coronary artery disease, aortic and mitral valve insufficiency, and aortic root dilation. Furthermore, atypical Fabry disease can be manifested with concentric left ventricular hypertrophy and no other disease findings.[10] Cerebrovascular involvement, leading to transient ischemic attacks and stroke, occur in approximately 25% of patients, with a mean age at onset of 40 years.

Female carriers of this X-linked disorder tend to have more variable and less severe symptoms, with a later age at onset. Affected women may not have proteinuria, even with pronounced renal impairment, and in almost 40%, stroke is the initial presentation. Whereas angiokeratomas are not typically seen in affected women, half have hypohidrosis and heat intolerance. The initial stages of cardiomyopathy in affected women is generally 10 years later than the classic presentation in men and may be the only manifestation of Fabry disease.

DIAGNOSIS

Fabry disease should be considered in individuals with angiokeratoma, acroparesthesia, and corneal lesions as well as in individuals with cryptogenic strokes, idiopathic cardiomyopathy, or renal disease (Table 208-3). Men and women with left ventricular hypertrophy without any other explanation or with a family history of renal, cardiovascular, cerebrovascular, or skin issues should be screened for Fabry disease. Angiokeratoma should be differentiated from Fordyce's disease, benign angiokeratomas of the scrotum, and angiokeratoma circumscriptum. Angiokeratomas are also seen in other lysosomal disorders including mannosidosis, fucosidosis, sialidosis, and β-galactosidase and β-hexosaminidase deficiency. Corneal abnormalities are similar to those seen secondary to the use of chloroquine or amiodarone; exposure to silicone dust can result in similar renal findings.

A presumed diagnosis can be confirmed by low α-galactosidase activity in peripheral white blood cells or cultured skin fibroblasts. Levels below 20% of normal are considered diagnostic, and levels up to 35% of normal should be considered suggestive. *GLA* mutation analysis is available and is critical for confirmation in atypically presenting males and heterozygote females because random X chromosome inactivation may lead to only slightly reduced or normal enzyme activity.

TREATMENT

Rx

Symptomatic treatment of the clinical manifestations of Fabry disease should follow standard medical care. Antiplatelet agents such as clopidogrel and aspirin or long-acting dipyridamole should be used for the prevention of strokes. Angiotensin-converting enzyme inhibitors and angiotensin receptor blockers are appropriate to manage hypertension and to preserve renal function. Kidney transplantation is effective in individuals with end-stage renal disease. Neuropathic pain can be treated with relatively low doses of antiepileptic medications, antidepressants, topical anesthetics, or pain relievers. Nonsteroidal anti-inflammatory agents should be avoided because of potential renal toxicity.

Enzyme replacement therapy has been available since 2001, and clinical trials suggest a modest benefit in modifying the natural course of the disease. Whereas there are no uniform recommendations, it is generally agreed that this therapy is appropriate for classically affected males, symptomatic females, and males with atypical disease (Table 208-4).[11]

OTHER LYSOSOMAL DISORDERS SEEN IN ADULTS

Metachromatic leukodystrophy, the deficiency of arylsulfatase A, results in the accumulation of sulfatides in the central and peripheral nervous systems, leading to demyelination of axons and peripheral nerves. It has a spectrum of manifestations, divided into childhood and adult variants. In both, gait disturbance and cognitive regression are seen. The adult form is associated with behavioral disturbances and dementia, often mistakenly resulting in a

diagnosis of psychosis. Metachromatic leukodystrophy is a recognized underlying cause of psychiatric illness in adults, and prominent features include auditory hallucinations, bizarre delusions, behavioral changes, personality changes, disinhibition and disorganization in daily life, and catatonic posturing. The diagnosis can be especially difficult in these individuals. Other neurologic signs, such as dysarthria and spasticity, manifest later as the disease progresses. The diagnosis is made initially by demonstration of low arylsulfatase A activity. However, because there can be a pseudodeficiency, it must be confirmed by molecular diagnosis or the demonstration of the excretion of sulfatides in urine.

Tay-Sachs' disease, caused by β-hexosaminidase A deficiency, is characterized by an excessive accumulation of the fatty acid derivative ganglioside GM_2 in neurons. There are three clinical variants based on the age at onset: type 1, infantile acute; type 2, subacute (2 to 18 years); and type 3, late onset. The main features of the disease are neurologic and cognitive deterioration as well as blindness, the macular cherry red spot, and deafness. In patients with late-onset GM_2 gangliosidosis, psychiatric signs may manifest years before the appearance of motor findings. The most common psychiatric signs include acute psychosis, mania, and depression without psychosis. Either recurrent progressive psychosis, consistent with schizophrenia-hebephrenia, or major depression followed by psychotic features may occur. Dysarthria and progressive speech loss are also common. The disorder is diagnosed by measurement of β-hexosaminidase A activity in the serum or white blood cells in the presence of normal or elevated activity of the β-hexosaminidase B isoenzyme. In pregnant women or those taking oral contraceptives, the test should be performed only in leukocytes.

Niemann-Pick disease type C has a spectrum of features ranging from a rapidly progressive, fatal neonatal phenotype to an adult-onset chronic, neurodegenerative course. Progressive ataxia, vertical supranuclear ophthalmoplegia, dystonia, and dementia are variable. Hepatosplenomegaly is frequently encountered. The disorder results from an error in cellular trafficking of exogenous cholesterol, leading to lysosomal accumulation of unesterified cholesterol, and has been linked to mutations in the *NPC1* (95% of cases) and *NPC2* (5% of cases) genes. The diagnosis is made by demonstration of impaired cholesterol esterification in cultured fibroblasts, termed the Filipin test, or by molecular genetic testing for mutations. Glycosphingolipids and cholesterol accumulate in different tissues such as the liver, spleen, bone marrow, and brain. In adults, psychiatric manifestations are observed that are consistent with acute psychosis, paranoid delusions or schizophrenia with delusions, hallucinations, disorganized behavior, and aggressiveness. Miglustat (*N*-butyldeoxynojirimycin) has been shown in clinical trials to stabilize but not to prevent or reverse key neurologic findings.[12] Decisions about starting this therapy typically involve the team of physicians, parents, and caregivers of the affected individual.

Aspartylglucosaminuria, a disorder more common in Finland than elsewhere in the world, is an autosomal recessive defect in glycoprotein degradation characterized by a slow or progressive delay in psychomotor development. Delayed speech and motor defects are often accompanied by repeated upper respiratory infections. Patients typically achieve the developmental competency of a 5- to 6-year-old child at around puberty; subsequently, they may experience progressive deterioration, resulting in the severe cognitive impairment seen in adulthood. The characteristic coarse facial features, thick calvaria, and osteoporosis result from mild connective tissue transformation. About 20% of patients experience seizures during the later stages of the disease, resulting primarily from abnormalities in the differentiation between gray and white matter and delayed myelination. The diagnosis is made by the detection of elevated urine oligosaccharides and deficient aspartylglucosaminidase activity in leukocytes.

The *neuronal ceroid-lipofuscinoses* are divided into four major groups—infantile, classic late infantile, juvenile, and adult—reflecting the age at symptom onset and the appearance of storage material on electron microscopy. Although generally inherited in an autosomal recessive fashion, the adult type can be inherited as a dominant allele. These diseases are characterized clinically by visual impairment leading to blindness, gait abnormalities, seizures, dementia, and early death. They are a genetically heterogeneous group of progressive, hereditary neurodegenerative diseases with variable onset of clinical manifestations. To date, approximately 160 mutations causing neuronal ceroid-lipofuscinoses have been found in eight human genes, complicating genetic analysis. Symptoms result from deficiencies in palmitoyl-protein thioesterase 1 and tripeptidyl-peptidase 1. The diagnosis is based on diminished enzyme activity and molecular genetic testing and, in some cases, clinical findings and electron microscopy of biopsied tissues. Accumulation of autofluorescent ceroid lipopigments is seen in the brain and other tissues. Adult patients may exhibit behavioral disturbances and cognitive decline, and parkinsonian features may be prominent.

Pompe's disease is discussed in Chapters 207 and 421.

The *mucopolysaccharidoses* (MPS) are a group of disorders resulting from defective lysosomal degradation of glycosaminoglycans. These chronic, progressive storage disorders show clinical manifestations that vary by type. Findings include coarse facial features, dysostosis multiplex, organomegaly, and neurologic manifestations with regression. All types are inherited in an autosomal recessive pattern except for MPS II/Hunter, which is X-linked. The mucopolysaccharidoses were once typically thought of as pediatric disorders, but with the advent of enzyme replacement therapy and the recognition of milder forms, more patients are reaching adulthood. Classically, patients with MPS type IS/Scheie, II/Hunter, III/Sanfilippo, IV/Morquio, and VI/Maroteaux-Lamy may have life expectancies beyond the pediatric years. Patients with MPS type IS may present with normal to short stature, normal intelligence, degenerative joint disease, corneal opacities, and cardiac valve lesions. Patients with MPS II/Hunter share the symptoms of MPS I/Hurler, except that airway involvement may be more significant in individuals with Hunter's syndrome, and the corneas are clear. Patients with MPS III/Sanfilippo have primarily central nervous system manifestations, with mild somatic involvement. They have progressive speech impairment, with the development of severe behavioral and sleep disturbances. Later they have an unrelenting loss of skills and deterioration into a vegetative state, with death in the third decade. Individuals with MPS IV have severe skeletal dysplasia, with normal intelligence in type A and a degenerative course in type B. Treatment is available in the form of enzyme replacement therapy for MPS IS, MPS II, and MPS IV, and enzyme therapies for other forms are in development. Medical management is type specific. In general, attention should focus on airway involvement resulting from progressive storage in the soft tissue of the upper airway and the need to optimize joint mobility and function by physical therapy. Patients with MPS I, II, and VI should be monitored for the development of cervical myelopathy due to dural thickening, which may lead to loss of endurance before ascending paralysis becomes apparent.[13]

Cystinosis is a lysosomal storage disorder with three clinical phenotypes: the most common infantile or nephropathic form, an intermediate or juvenile-onset form, and a benign form typically seen in adults and affecting primarily the eyes. Lysosomal cystine accumulation results from mutations in the gene *CTNS*, which codes for cystinosin, a lysosomal carrier protein. Individuals with the nephropathic form have lysosomal cystine accumulation leading to multiorgan system involvement, including progressive renal disease, corneal crystals, and effects on the thyroid, gonads, pancreas, muscle, and central nervous system. The adult form has only corneal crystals. Treatment is supportive and should include cysteamine, which is an oral medication that decreases cystine accumulation.[14] Cysteamine hydrochloride eyedrops dissolve corneal crystals and relieve photophobia.

Grade A Reference

A1. El Dib RP, Nascimento P, Pastores GM. Enzyme replacement therapy for Anderson-Fabry disease. *Cochrane Database Syst Rev*. 2013;2:CD006663.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Boustany RM. Lysosomal storage diseases—the horizon expands. *Nat Rev Neurol*. 2013;9:583-598.
2. Staretz-Chacham O, Choi JH, Wakabayashi K, et al. Psychiatric and behavioral manifestations of lysosomal storage disorders. *Am J Med Genet B Neuropsychiatr Genet*. 2010;153B:1253-1265.
3. Desnick RJ, Schuchman EH. Enzyme replacement therapy for lysosomal diseases: lessons from 20 years of experience and remaining challenges. *Annu Rev Genomics Hum Genet*. 2012;13:307-335.
4. Ferraz MJ, Kallemeijn WW, Mirzaian M, et al. Gaucher disease and Fabry disease: new markers and insights in pathophysiology for two distinct glycosphingolipidoses. *Biochim Biophys Acta*. 2014;1841:811-1825.
5. Grabowski GA. Gaucher disease and other storage disorders. *Hematology Am Soc Hematol Educ Program*. 2012;2012:13-18.
6. Siebert M, Sidransky E, Westbroek W. Glucocerebrosidase is shaking up the synucleinopathies. *Brain*. 2014;137:1304-1322.
7. Sidransky E, Lopez G. The link between the GBA gene and parkinsonism. *Lancet Neurol*. 2012;11:986-998.
8. Nalls MA, Duran R, Lopez G. A multicenter study of glucocerebrosidase mutations in dementia with Lewy bodies. *JAMA Neurol*. 2013;70:727-735.
9. Weidemann F, Niemann M, Sommer C, et al. Interdisciplinary approach towards female patients with Fabry disease. *Eur J Clin Invest*. 2012;42:455-462.
10. Palecek T, Honzikova J, Poupetova H, et al. Prevalence of Fabry disease in male patients with unexplained left ventricular hypertrophy in primary cardiology practice: prospective Fabry cardiomyopathy screening study (FACSS). *J Inherit Metab Dis*. 2014;37:455-460.
11. Pisani A, Visciano B, Roux GD, et al. Enzyme replacement therapy in patients with Fabry disease: state of the art and review of the literature. *Mol Genet Metab*. 2012;107:267-275.
12. Patterson MC, Hendriksz CJ, Walterfang M, et al. Recommendations for the diagnosis and management of Niemann-Pick disease type C: an update. *Mol Genet Metab*. 2012;106:330-344.
13. Muenzer J. Overview of the mucopolysaccharidoses. *Rheumatology (Oxford)*. 2011;50(suppl 5):v4-v12.
14. Nesterova G, Gahl WA. Cystinosis: the evolution of a treatable disease. *Pediatr Nephrol*. 2013;28:51-59.

REVIEW QUESTIONS

1. All except which of the following are different defects that can result in lysosomal storage disorders?

A. Sphingolipidosis
B. Lysosomal transport disorders
C. Phenylketonuria
D. Oligosaccharidoses
E. Lipid storage disorders

Answer: C Lysosomal storage disorders are a group of more than 45 different inherited disorders, all sharing a defect in lysosomal function. Lysosomes, acidic, membrane-bound organelles present in the cytoplasm, contain enzymes that degrade cellular macromolecules. These disorders occur when one or more of the hydrolytic enzymes are deficient or when essential lysosomal transporters, receptors, cofactors, or protective proteins are defective or lacking. Sphingolipidosis, the oligosaccharidoses, lipid storage disorders, and lysosomal transport disorders are all considered lysosomal storage disorders. Phenylketonuria is a disorder of amino acid metabolism, and abnormal lysosomal storage does not occur.

2. To make the diagnosis of a lysosomal storage disorder:

A. A tissue biopsy is often necessary.
B. DNA analyses are required.
C. Enzymatic assays can be performed on a blood or fibroblast sample.
D. Urine tests are the standard of care.
E. Clinical criteria alone are adequate.

Answer: C The diagnosis of most lysosomal storage disorders can be made by assaying the enzyme activity in a blood sample or fibroblasts pellet from a skin biopsy specimen. The first tier of diagnostics should be a lysosomal disease diagnostic panel, evaluating the enzyme activity of multiple lysosomal enzymes. Whereas the genes encoding the lysosomal enzymes have been identified and mutation analysis is available, there is vast genotypic heterogeneity. Thus, mutation screening is most useful when a mutation has already been identified in a specific family or when specific mutations are known to be common in an ethnic group.

3. A 24-year-old woman with no prior health issues presents with sudden onset of dysarthria and right hemiparesis. She also mentions that she "sweats less than her peers during exercise." She had a paternal grandmother and uncle who both died of renal disease in their fourth decade. Her work-up should include testing for:

A. Gaucher disease
B. Fabry disease
C. Glycogen storage disease
D. Wolman's disease
E. All of the above

Answer: B Female carriers of Fabry disease, which is an X-linked disorder, have less severe symptoms with a later age at onset. About 40% of affected women have stroke as the initial symptom. Whereas the typical skin finding in Fabry disease of angiokeratoma is not seen in women, about half complain of hypohidrosis and heat intolerance. Renal involvement is common, with progressive renal insufficiency and end-stage renal disease in the second to fourth decades. Affected men have proteinuria, which is less frequently seen in affected women, even those affected with renal disease. Gaucher disease, glycogen storage disease, and Wolman's disease are not manifested with stroke symptoms in men or women.

4. A 22-year-old Ashkenazi Jewish college student presents with thrombocytopenia and painless splenomegaly. He is otherwise healthy with no relevant family history. His work-up should include diagnostic testing for

A. Hurler's disease
B. Fucosidosis
C. Cystinosis
D. Gaucher disease
E. Hereditary hemorrhagic telangiectasia

Answer: D Type 1 Gaucher disease should be considered in an otherwise healthy young adult with thrombocytopenia and painless splenomegaly. The carrier frequency of mutations in the gene encoding glucocerebrosidase (*GBA1*), which underlies Gaucher disease, is about 1 in 16 in Ashkenazi Jews, in contrast to the carrier frequency in the general population of about 1 in 100. Hurler's disease is a mucopolysaccharidosis that manifests with hepatosplenomegaly, typically in childhood, although patients may live to adulthood. Cystinosis manifests in adults with corneal crystals. Hereditary hemorrhagic telangiectasia is an autosomal dominant disease that is manifested with abnormal blood vessel formations (telangiectasias) of the skin, mucous membranes, lungs, and gut. Fucosidosis manifests in childhood with developmental delay and a neurodegenerative course.

5. The following is true about the association between Gaucher disease and parkinsonism:

A. All patients with Gaucher disease ultimately develop parkinsonism.
B. The increased risk of Parkinson's disease is observed only in Ashkenazi Jews.
C. Mutations in *GBA*, the gene encoding glucocerebrosidase, are extremely rare in subjects with parkinsonism.
D. An increased frequency of parkinsonism is seen in both patients with Gaucher disease and Gaucher disease carriers.
E. Lewy body disorders other than Parkinson's disease are not associated with mutations in *GBA1*.

Answer: D Mutations in the glucocerebrosidase gene are one of the most common genetic risk factors for the development of parkinsonism. Both patients with Gaucher disease and Gaucher disease carriers have an increased risk of both Parkinson's disease and related synucleinopathies, such as dementia with Lewy bodies. The increased frequency of *GBA1* mutations is observed in both Jewish and non-Jewish patients with parkinsonism. However, the majority of patients with Gaucher disease never develop Parkinson's disease.

209

HOMOCYSTINURIA AND HYPERHOMOCYSTEINEMIA

MANUEL SCHIFF AND HENK BLOM

DEFINITION

Homocysteine is a nonprotein amino acid and is a key metabolic branch point metabolite between the trans-sulfuration and remethylation pathways of methionine metabolism. The many conditions associated with high

homocysteine levels encompass a wide range of clinical manifestations. The normal level of plasma total homocysteine (tHcy) is below 15 µM. However, the threshold of tHcy above which a disorder of homocysteine metabolism should be suspected and a specific therapy should be initiated is around 50 µM.

Homocystinuria and hyperhomocysteinemia primarily include inherited disorders of homocysteine metabolism but can also involve acquired nutritional cobalamin (Cbl, vitamin B_{12}) or folate deficiencies. Severely elevated plasma levels of homocysteine are accompanied by homocystinuria as rare autosomal recessively inherited disorders that are associated with vascular, neurologic, ocular, and skeletal abnormalities. Lesser elevations of plasma homocysteine, without homocystinuria, occur in about 5 to 7% of the population; these individuals with hyperhomocysteinemia only do not have the clinical stigmata of homocystinuria but are at increased risk of atherosclerotic cardiovascular disease (Chapter 52) and venous thromboembolism (Chapter 176).

Inherited disorders of homocysteine metabolism (Figs. 209-1 and 209-2) comprise *disorders of the trans-sulfuration pathway* with cystathionine β-synthase (CBS) deficiency (or classic homocystinuria) and *disorders of remethylation* of homocysteine to methionine. The latter include 5,10-methylenetetrahydrofolate reductase [MTHFR] deficiency and inherited disorders of cobalamin absorption, transport, and intracellular metabolism and the very rare congenital folate malabsorption disorder. Intracellular remethylation defects include disorders that all have defective methionine synthesis in common: MTHFR deficiency impairs methyltetrahydrofolate synthesis; defective lysosomal release of cobalamin (CblF and CblJ) and defects in cytosolic reduction and transport of hydroxocobalamin (CblC and CblD) impair the synthesis of both methylcobalamin and

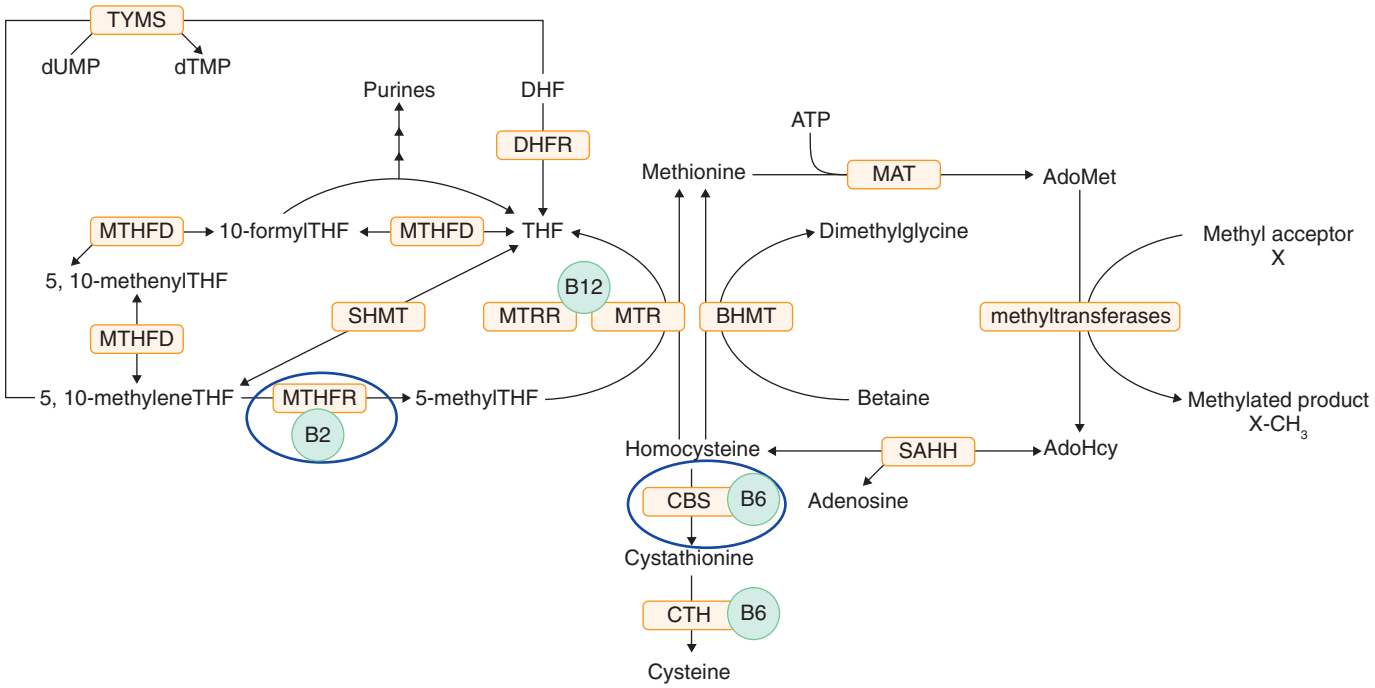

FIGURE 209-1. Homocysteine metabolism and the folate cycle. The two conditions leading to accumulation of homocysteine (CBS and MTHFR) are surrounded by a circle. AdoHcy = *S*-adenosylhomocysteine; AdoMet = *S*-adenosylmethionine; ATP = adenosine triphosphate; BHMT = betaine-homocysteine methyltransferase; CBS = cystathionine β-synthase; CTH = cystathionine γ-lyase; DHF = dihydrofolate; DHFR = dihydrofolate reductase; dTMP = deoxythymidine monophosphate; dUMP = deoxyuridine monophosphate; MAT = methionine-adenosyltransferase; MTHFD = methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase/formyltetrahydrofolate synthetase; MTHFR = methylenetetrahydrofolate reductase; MTR = methionine synthase; MTRR = methionine synthase reductase; SAHH = *S*-adenosylhomocysteine hydrolase; SHMT = serine-hydroxymethyltransferase; THF = tetrahydrofolate; TYMS = thymidylate synthase.

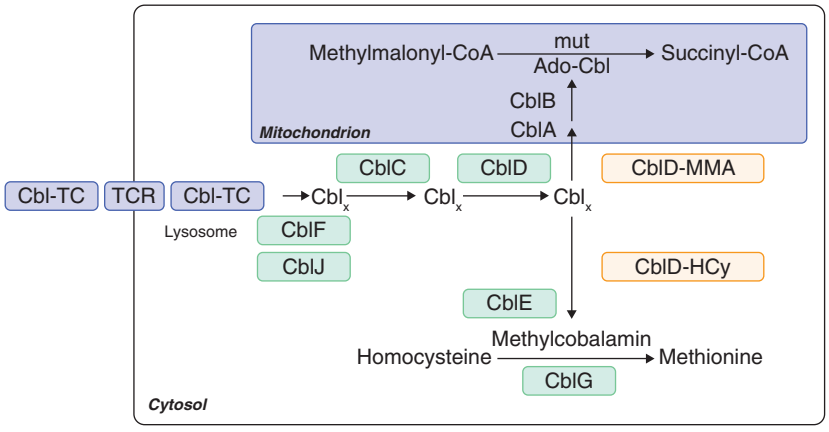

FIGURE 209-2. Intracellular cobalamin metabolism pathway and its defect. To date, 10 complementation group defects of the cobalamin pathway have been described. After binding to the transcobalamin receptor (TCR), cobalamin bound to TC enters the cell through lysosome-mediated endocytosis and is released through proteolysis. Export from the lysosome into the cytoplasm is defective in patients with the CblF and recently described CblJ defects. The steps in the cytosol after lysosomal release are defined by the complementation groups CblC and CblD. The exact form of cobalamin at this stage is unclear (as indicated by Cbl_x). In the cytosol, cobalamin is reductively methylated by methionine synthase reductase (CblE) to methylcobalamin, the cofactor for methionine synthase (CblG). After its transport into the mitochondrion, cobalamin is converted to adenosylcobalamin (Ado-Cbl), the cofactor for methylmalonyl–coenzyme A (CoA) mutase (mut), by cobalamin adenosyltransferase (CblB). The CblD defect can cause isolated methylmalonic aciduria (CblD-MMA complementation group), isolated homocystinuria (CblD-HCy), or both (CblD). In all these conditions with defective remethylation (TC and TC receptor deficiency, CblF, CblJ, CblC, CblD, CblD-HCy, CblE, and CblG defects), there is homocysteine accumulation due to dysfunction in methionine synthesis.

TABLE 209-1 GENETIC DEFECTS ASSOCIATED WITH HOMOCYSTINURIA

FUNCTIONAL DEFECT	COMMON NAME	ENZYME DEFECT	GENE	CHROMOSOME LOCUS
Trans-sulfuration	"Classic" homocystinuria	Cystathionine β-synthase	*CBS*	21q22.3
Remethylation	Folate-dependent homocystinuria	Methylenetetrahydrofolate reductase	*MTHFR*	1p36.3
	CblE	Methyltransferase reductase	*MTRR*	5p15.2-p15.3
	CblG	Methionine synthase	*MTR*	1q43
Cobalamin transport	TC-II	Transcobalamin	*TCN2*	22q11.2-qter
	CblF	Lysosomal B_{12} translocase	*LMBRD1*	6q13
Cobalamin processing	CblC	Intracellular cobalamin chaperone?	*MMACHC*	1p34.1
	CblD	Cobalamin reductase?	*MMADHC*	2q23.2
	CblJ	Cobalamin processing?	*ABCD4*	14q24.3

adenosylcobalamin; and isolated deficiencies of methionine synthase (CblE and CblG) are associated with defective methylcobalamin synthesis as well as the CblD-Hcy variant.

In CBS deficiency, in addition to severe elevations of plasma tHcy, plasma methionine level is also increased. In contrast, in remethylation disorders, the increased tHcy is accompanied by a low (or low-normal) level of methionine in plasma because of the ineffective conversion (remethylation) of homocysteine to methionine. Among the remethylation disorders, the cobalamin-processing (CblC) defect (the most frequent inherited disorder of intracellular cobalamin metabolism) and MTHFR deficiency are by far the most frequent, although in absolute numbers these disorders are still rare. Accordingly, this chapter focuses on CBS deficiency, CblC defect, and MTHFR deficiency.

EPIDEMIOLOGY

Prevalence and Incidence

The worldwide prevalence of CBS deficiency has been reported at 1 in 344,000. Minimum estimates of the incidence of CBS deficiency by newborn screening programs have ranged from 1 in 60,000 to 1 in 300,000 live births, varying with the population and the screening method. Estimates of its incidence in Europe have been in the area of 1 in 40,000, which corresponds to a carrier (heterozygote) frequency of about 1%, but studies screening for known mutations suggest that the prevalence may be more than twice that rate. Studies from the Middle Eastern countries have shown incidence rates as high as 1 in 1800. The incidence of severe homocysteine remethylation defects is probably less than 1 in 500,000 live births.

PATHOBIOLOGY

The homocysteine that accumulates (in both CBS deficiency and remethylation defects) is a multisystem toxic agent that exerts its effects either directly or indirectly through conversion to *S*-adenosylhomocysteine (see Fig. 209-1), which potentially inhibits many essential methyltransferases. Direct cellular homocysteine toxicity involves endothelial injury and neuronal cell death. The effects of homocysteine on vascular endothelium predispose to thrombosis that may occur at any age and affect arteries or veins of any size. Specifically in CBS deficiency, accumulation of homocysteine or cysteine deficiency induces modification of connective tissue proteins, possibly being the origin of the skeletal and ocular manifestations. These effects are probably related to fibrillin, which is a component of the matrix of periosteum and perichondrium, the major component of the zonular fibers of the ocular lens, and a protein singularly rich in cysteine. Fibrillin structure is affected by the linking of homocysteine to cysteine; as a result, some clinical features of homocystinuria overlap with fibrillin mutations (Marfan syndrome; Chapter 260).

Remethylation defects (CblC and MTHFR) result not only in severe elevations of homocysteine but also in a shortage of methionine, required for protein synthesis and *S*-adenosylmethionine generation. The latter causes, through this route, a further reduction of cellular methylation capacity in addition to the accumulation of *S*-adenosylhomocysteine, especially in the central nervous system. CblC defect is an inborn error of intracellular cobalamin metabolism due to a defect of the methylmalonic aciduria and homocystinuria type C protein (*MMACHC*). After normal dietary intake, intestinal absorption, blood transport, and cellular uptake, cobalamin is delivered in the cytosol, where it becomes bound to MMACHC. This protein has been shown to catalyze dealkylation of alkylcobalamins, such as adenosylcobalamin and methylcobalamin, and the decyanation of cyanocobalamin.[1] In the absence of normal MMACHC, neither methylcobalamin (the cofactor of methionine synthase) nor adenosylcobalamin (the cofactor of methylmalonyl-CoA mutase) is functional, leading to defects of the methionine synthase (remethylation defect) and methylmalonyl-CoA mutase (methylmalonic acid accumulation), respectively (see Fig. 209-2). In MTHFR deficiency, the methyl donor 5-methyltetrahydrofolate (5-methylTHF) cannot be produced, which secondarily impairs methionine synthase function and subsequent remethylation (see Fig. 209-2). The pathogenesis of MTHFR deficiency can be ascribed to homocysteine toxicity together with methionine cellular depletion. However, the pathophysiologic mechanism of the other remethylation defect, the CblC defect, remains incompletely understood. Toxic accumulation of methylmalonic acid in the CblC defect might play an additional role.

All these genetic disorders associated with homocystinuria are inherited in an autosomal recessive manner and are summarized in Table 209-1.

CLINICAL MANIFESTATIONS

CYSTATHIONINE β-SYNTHASE DEFICIENCY Affected individuals are normal at birth after an uneventful pregnancy and delivery. If untreated, they progressively develop the core clinical symptoms of CBS deficiency,[2] which involve four major organ systems.

CENTRAL NERVOUS SYSTEM Developmental delay and mental retardation affect about 60% of the patients to a variable degree of severity. Seizures, electroencephalographic abnormalities, and psychiatric disturbances have also been reported in approximately half of the cases. Focal neurologic signs may be a consequence of cerebrovascular accidents.

EYE Dislocation of the ocular lens (ectopia lentis), myopia, and glaucoma are frequent, severe, and characteristic complications. Retinal detachment and degeneration, optical atrophy, and cataracts may eventually appear. Myopia often precedes lens dislocation. Ectopia lentis is detected in most untreated patients from 5 to 10 years of age and in nearly all untreated patients by the end of the fourth decade. In children as well as in adults, it is often the clue to diagnosis. The dislocation is generally downward, whereas it is usually upward in Marfan syndrome. Once ectopia lentis has occurred, a peculiar trembling of the iris (iridodonesis) after eye or head movement may be evident.

SKELETON Osteoporosis is almost invariably detected, at least after childhood. Frequent consequences are scoliosis and a tendency toward pathologic fractures and vertebral collapse. Homocystinuric patients tend to be tall, with thinning and elongation (dolichostenomelia) of long bones near puberty, enlarged metaphyses and epiphyses (especially at the knees), and arachnodactyly, present in about half of the patients. Other bone deformities include genu valgum, pes cavus, and pectus carinatum or excavatum. Restricted joint mobility, particularly at the extremities, contrasts with the joint laxity observed in Marfan syndrome.

VASCULAR SYSTEM Thromboembolic complications, occurring in arteries and veins of all parts of the body, constitute the major cause of morbidity and mortality at any age. Because of the very low prevalence of arteriosclerosis and thrombosis in children or adolescents, homocystinurias should be excluded in any such case presenting with arterial or venous occlusive disease.

Two phenotypic variants are recognized, B_6-responsive homocystinuria and B_6-nonresponsive homocystinuria. B_6-responsive homocystinuria is usually milder than the nonresponsive variant.

Adult-Onset Clinical Presentation

The expression of all clinical symptoms is extremely variable and may be manifested during childhood but also during adulthood, even up to 60 years of age. Individuals are often tall and slender with a marfanoid habitus and are prone to osteoporosis. The main acute manifestation in adulthood is

cardiovascular disease. In all forms of inherited homocystinurias, isolated arterial and venous occlusive disease may be manifested at any age. Thromboembolism is the major cause of early death and morbidity. Adults can also present with isolated quick loss of diopter (refractive capacity), which is the most common symptom before diagnosis.

The IQ in individuals with untreated homocystinuria varies widely, from 10 to 138. In B_6-responsive untreated individuals, the mean IQ is 79 versus 57 for those who are B_6 nonresponsive. The neuropsychiatric symptoms (like schizophrenia or autism-like features) may remain apparently isolated, but a thorough clinical history and examination will often reveal associated features like marfanoid habitus, skeletal abnormalities, and lens dislocation.

Remethylation Disorders: CblC Defect and MTHFR Deficiency

Clinical signs of remethylation defects are mainly neurologic. Neonatal and early-onset patients exhibit acute neurologic distress. In childhood, patients exhibit nonspecific mental retardation often associated with acquired microcephaly. Without appropriate therapy, remethylation-defective patients may develop acute or rapidly progressive neurologic deterioration, sometimes leading to death. Adolescents and adults exhibit, after a period of normal development or mild developmental delay, a rapid mental or psychiatric deterioration. These patients may typically have signs of subacute combined degeneration of the cord. In addition, adults can be asymptomatic or present with isolated stroke.

In the CblC defect,[3-5] 5-methylTHF accumulates because of the block at methionine synthase (the 5-methylTHF trap), causing a functional cellular folate deficiency. This explains the hematologic manifestations (megaloblastic bone marrow leading to macrocytic anemia or pancytopenia) that are not present in MTHFR deficiency. In the CblC defect, severe (occasionally fatal) multisystem deterioration may occur. This includes hemolytic-uremic syndrome, cardiomyopathy, and interstitial pneumonia, which all share an identical pathologic hallmark (i.e., thrombotic microangiopathy). In addition, a peculiar and poorly understood retinopathy with nystagmus is often present.

Adult-Onset Clinical Presentation

Adult-onset CblC disease is less frequent than childhood onset and is dominated by neuropsychiatric manifestations such as ataxia, cognitive impairment, and psychosis. Hemolytic-uremic syndrome (even isolated) can also be present.

In MTHFR deficiency, the neuropsychiatric presentation is similar and can be subtle. Patients may remain asymptomatic or exhibit isolated arterial stroke.

In both remethylation defects, the neurologic disorder results in many signs similar to the late stage observed in childhood. Previously, most of these patients had either a normal or a mild developmental delay, and a striking feature is the rapid mental deterioration occurring in the second decade of life accompanied by bouts of unexplained lethargy and a progressive cerebral, myelopathic, and neuropathic disorder with variable results on neurophysiologic investigations.

In both groups of inherited homocystinurias, isolated arterial and venous occlusive disease may be manifested at any age. Because of the very low incidence of arteriosclerosis and thrombosis in children or adolescents, homocystinurias should be excluded in any case in this age presenting with these clinical manifestations.

DIAGNOSIS

If homocystinuria is suspected clinically, plasma tHcy should be determined. If tHcy (normal values, 5-15 µM) is higher than 50 µM, determination of serum vitamin B_{12} and folate levels in both serum and red blood cells, plasma amino acids, and urinary organic acids (or plasma methylmalonic acid) is warranted. If tHcy is below 50 µM, the probability of a homocystinuria is minute. If plasma methionine is elevated or high-normal with low plasma cysteine, the plasma abnormalities point to CBS deficiency, whereas decreased or low-normal methionine and elevated tHcy point to remethylation defects. In the CblC defect, hematologic abnormalities with megaloblastic bone marrow maturation are observed along with normal vitamin B_{12} and folate blood levels. Nutritional vitamin B_{12} or folate deficiencies and other genetic causes of cobalamin and folate metabolism should also be considered in the differential diagnosis of remethylation defects.[6] It has to be kept in mind that CBS deficiency can cause secondary vitamin B_{12} and folate deficiency and remethylation defects can cause secondary folate deficiency. In MTHFR deficiency, folate levels (red blood cells or plasma) are usually low. Blood cell count with blood smear (looking for characteristics of vitamin B_{12} and folate deficiency), blood vitamin status, and metabolic biomarkers (tHcy, methionine, and methylmalonic acid) are usually enough to differentiate between the different forms of homocystinurias (Table 209-2). Definitive diagnosis can be confirmed at the molecular level by sequence analysis of the putative gene (see Table 209-1); if the molecular studies are not conclusive, functional studies can be performed in fibroblasts or lymphocytes.

TREATMENT

Rx

In 2015, general treatment guidelines will be available through the European Network and Registry for Homocystinurias and Methylation Defects website (*http://www.e-hod.org*). None of the treatment options are evidence based because of the rarity of these disorders.

Therapeutic Goals

The general therapeutic goal is to reduce tHcy accumulation and, for remethylation disorders, to bypass the remethylation defect, thereby

TABLE 209-2 CLINICAL FEATURES OF HOMOCYSTINURIAS

	BIOCHEMICAL FEATURES			CLINICAL FEATURES	
CLASS	***tHcy***	***Meth***	***MMA***	***System***	***Signs***
CBS deficiency	↑	↑	–	Ocular	Ectopia lentis, myopia, glaucoma, optic atrophy, retinal detachment
				Skeletal	Elongated and thinned bones, arachnodactyly, genu valgum, pectus malformation, scoliosis
				Vascular	Thromboembolic events (arterial or venous)
				Neurologic	Mental retardation in untreated cases, cerebrovascular thromboses, seizures
					Psychiatric disorders, personality disorder
MTHFR deficiency	↑	↓	–	Ocular	Ectopia lentis
				Vascular	Thromboses
				Neurologic	Variable: psychiatric to severe neurologic
Transcobalamin	↑	–/↓	+	Hematologic	Early-onset pancytopenia, macrocytosis
CblF, CblJ	↑	–/↓	+	Pansystemic	MMA, macrocytosis, stomatitis, congenital heart defects
CblC, CblD	↑	↓	+	Hematologic	Pancytopenia
				Neuropsychiatric	
				Systemic (CblC)	Hemolytic-uremic syndrome
CblE, CblG	↑	↓	–	Vascular	Thromboses
				Hematologic	Pancytopenia
				Neurologic	

Cbl = cobalamin; CBS = cystathionine β-synthase; Meth = plasma methionine; MMA = methylmalonic acid (urine or plasma); MTHFR = methylenetetrahydrofolate reductase; tHcy = total plasma homocysteine.

maintaining normal methionine and folate concentrations.[7] This should correct the hematologic abnormalities and ensure normal neurologic development or prevent further neurologic deterioration. Normalization of plasma tHcy levels would be ideal, but in practice, it is difficult if not impossible to achieve. However, the experience with CBS deficiency patients has shown that treatment can prevent further thromboembolic events even when the tHcy levels remain clearly above the normal range. Decreasing tHcy to 50 to 70 µM would therefore be a more reasonable goal in many patients with CBS deficiency as well as in those with remethylation disorders.

Available Treatment Options

To date, all the remethylation disorders are similarly treated with the combined supplementation of vitamin B_{12}, vitamin B_9 (folate), vitamin B_6 (pyridoxine), betaine, and methionine, although the dosage and route of administration may vary with the type of defect.

In CBS deficiency, therapy is more standardized with a longer experience available. It is based on (1) increasing residual CBS activity with the use of vitamin B_6 (in B_6-responsive patients); (2) decreasing the load on the affected pathway and replacing the deficient products by a low-methionine diet, limitation of natural proteins, amino acid mixture, special low-protein foods, and supplementation of cystine to increase cysteine; and (3) increasing remethylation to methionine with folate, vitamin B_{12}, and betaine to reduce tHcy accumulation.

Vitamin B_{12}

Vitamin B_{12} is the cofactor of methionine synthase. Its natural form, hydroxocobalamin, is more effective than the synthetic form cyanocobalamin. In CBS deficiency, hydroxocobalamin may be given orally (1 mg/day to 1 mg/week) according to the serum B_{12} level to prevent cobalamin deficiency.

In remethylation disorders, initial treatment includes daily parenteral administration of hydroxocobalamin (1 mg/day). If MTHFR deficiency is confirmed, switching to oral hydroxocobalamin (1 mg/day to 1 mg/week) or even stopping hydroxocobalamin supplementation may deserve discussion. Conversely, in CblC defect, lifelong high-dose intramuscular hydroxocobalamin injections are needed. The optimal interval between intramuscular injections remains to be determined, but recent data support the need for long-term daily doses.

Folate (Vitamin B_9)

Folate (vitamin B_9) is available in three different forms: folic acid, a stable synthetic form of the vitamin used, for example, in food fortification; folinic acid (5-formylTHF), the most stable form of the reduced and active vitamin; and 5-methylTHF (CH3-THF), the main natural and circulating form of the vitamin. A folinic acid formulation for parenteral administration is available for emergency treatment, whereas the other forms are available only for oral use. Whatever the disorder, folinic acid is more appropriate because it is the most stable reduced form, and folic acid may exacerbate cerebral CH3-THF deficiency, especially in MTHFR deficiency. In CBS deficiency, folinic acid should be given orally (1 to 5 mg/day) to avoid folate depletion. Moreover, folate repletion may be necessary to permit a pyridoxine response, which means that pyridoxine responsiveness should always be tested after potential folate depletion correction. In the CblC defect, long-term high-dose oral folinic supplementation is added to compensate for the methyl folate trap and to correct the hematologic abnormalities. The daily dose varies from 5 to 30 mg. The same folinic acid doses should be used in MTHFR deficiency.

Vitamin B_6 (Pyridoxine)

Vitamin B_6 (pyridoxine), through its action as the cofactor for CBS, is given orally in pharmaceutical doses in CBS-deficient patients to treat possible B_6-responsive individuals. There is no consensus on the dose and duration of B_6, which is usually given from 100 to 500 to 1000 mg/day during a period of several weeks, after which tHcy is determined to evaluate whether B_6 has been effective in normalizing (or even lowering) tHcy. As discussed before, B_6 should always be combined with folinic acid. In B_6-responsive patients, long-term therapy with pyridoxine and folate prevents further deterioration. In remethylation disorders, B_6 might theoretically enhance homocysteine removal and may be given at a low dose (50 to 100 mg/day).

For pyridoxine-responsive CBS-deficient patients, pyridoxine should be kept at the lowest dosage able to achieve adequate metabolic control. Because of reported high-dose pyridoxine toxicity on the nervous system, daily dosages higher than 400 to 500 mg/day in children and adults should probably be avoided for a long-term treatment. In pyridoxine-nonresponsive patients, a daily dosage of 50 to 100 mg of pyridoxine can be added to the treatment.

Betaine

Betaine is derived from choline and is a substrate for the enzyme betaine-homocysteine methyltransferase and therefore acts as a methyl donor (see Fig. 209-1). Oral betaine supplementation decreases homocysteine levels. Despite widespread use, there is little consensus on betaine dosage and frequency of administration. Case studies and early literature have used doses of 150 to 250 mg/kg/day in children and 5 to 10 g/day in adults, usually given two or three times daily. These early data were confirmed by pharmacokinetic studies (performed in CBS-deficient patients), which showed that above 200 mg/kg/day in two to three divided doses, there was no obvious benefit in lowering tHcy.

In CBS deficiency, the use of betaine is usually followed by an increase in plasma methionine and a fall in plasma tHcy. No harmful effects from raised methionine have been documented in patients receiving betaine therapy except one case of cerebral edema in a CBS patient with a plasma methionine level above 1000 µM. In B_6-nonresponsive early-treated patients, a low-methionine diet alone can be highly successful with good long-term outcomes, provided lifelong compliance is good. The clinical benefits of betaine are therefore questionable in compliant patients on a low-methionine diet. However, in some patients (especially when dietary compliance is a problem), betaine has been of benefit and may allow an increase in natural protein intake, thus improving the nutritional status.

In remethylation disorders, betaine increases systemic methionine levels and probably also methionine availability to the central nervous system, especially in patients with MTHFR deficiency. Early betaine treatment prevents mortality and allows normal psychomotor development in patients with severe MTHFR deficiency, highlighting the importance of timely recognition.[8] In the CblC defect, there are data reporting a synergistic action of hydroxocobalamin with betaine, lowering the tHcy level.

Oral Methionine

Oral methionine may also hold promise as an additional therapeutic measure in remethylation defects for several reasons. Cerebral methionine depletion is a key pathogenic factor. Added methionine might act synergistically with betaine by supplying intracellular methionine, and acute methionine loading does not lead to further homocysteine accumulation (with the attendant risk of thromboembolism). As a whole, whatever the remethylation defect, methionine depletion is rarely corrected by betaine therapy alone, whereas the association of methionine and betaine usually corrects methionine depletion.

PREVENTION

It is prudent to adopt measures to prevent additional risk of thrombosis, such as using long-term, low-dose aspirin and avoiding smoking and oral contraceptives. Nitrous oxide may also be relatively contraindicated because it can inhibit methionine synthase. Surgery poses serious risks but can be performed safely as long as attention is paid to the patient's hydration and coagulation status.

PROGNOSIS

In CBS deficiency, pyridoxine responsiveness generally correlates with higher residual enzyme activity, and the prognosis is significantly better than that for unresponsive cases, with or without treatment. Skeletal, ocular, vascular, and neurologic complications are all reduced with successful treatment. Without the early institution of treatment, the median IQ in a large outcome study was 57 for unresponsive patients and 78 for responsive patients. With early treatment, pyridoxine-unresponsive patients have a nearly normal median IQ. For patients who are responsive to and compliant with treatment, the prognosis for intellectual development is good, especially with diet alone. However, in case of poor adherence, significant elevations in tHcy generally persist, and some increased risk for vascular complications probably remains.

In the few MTHFR-deficient patients treated promptly in their neonatal period, the outcome is good with respect to the first years of neurologic development despite suboptimal metabolic control. If undiagnosed or untreated or late treated, these patients have a very poor outcome with severe impairments. In spite of some pathogenic similarity (intracellular methionine depletion due to impaired remethylation), CblC-defective patients have in general a particularly poor long-term outcome with multisystem involvement, thrombotic microangiopathy, and retinopathy.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Quadros EV. Advances in the understanding of cobalamin assimilation and metabolism. *Br J Haematol*. 2010;148:195-204.
2. Skovby F, Gaustadnes M, Mudd SH. A revisit to the natural history of homocystinuria due to cystathionine β-synthase deficiency. *Mol Genet Metab*. 2010;99:1-3.
3. Carrillo-Carrasco N, Chandler RJ, Venditti CP. Combined methylmalonic acidemia and homocystinuria, cblC type. I. Clinical presentations, diagnosis and management. *J Inherit Metab Dis*. 2012;35:91-102.
4. Carrillo-Carrasco N, Venditti CP. Combined methylmalonic acidemia and homocystinuria, cblC type. II. Complications, pathophysiology, and outcomes. *J Inherit Metab Dis*. 2012;35:103-114.
5. Fischer S, Huemer M, Baumgartner M, et al. Clinical presentation and outcome in a series of 88 patients with the cblC defect. *J Inherit Metab Dis*. 2014;37:831-840.
6. Watkins D, Rosenblatt DS. Update and new concepts in vitamin responsive disorders of folate transport and metabolism. *J Inherit Metab Dis*. 2012;35:665-670.
7. Schiff M, Blom HJ. Treatment of inherited homocystinurias. *Neuropediatrics*. 2012;43:295-304.
8. Diekman EF, de Koning TJ, Verhoeven-Duif NM, et al. Survival and psychomotor development with early betaine treatment in patients with severe methylenetetrahydrofolate reductase deficiency. *JAMA Neurol*. 2014;71:188-194.

REVIEW QUESTIONS

1. A 56-year-old man presents with ectopia lentis and venous thrombosis. What would be the first step in ruling out homocystinuria due to cystathionine β-synthase deficiency?

A. Rule out Marfan syndrome
B. Measure cystathionine β-synthase in cultured fibroblasts
C. Measure total homocysteine and methionine in plasma
D. Try oral vitamin B_6
E. Sequence the cystathionine β-synthase gene

Answer: C Given the clinical picture, if plasma methionine and total homocysteine levels are elevated, the diagnosis of cystathionine β-synthase deficiency is certain and therapy is urgently needed.

2. A patient is diagnosed with cystathionine β-synthase deficiency. The mutation is not known. What is the best way to start therapy?

A. Oral aspirin
B. Oral vitamin B_6 along with oral folinic acid
C. Vitamin B_{12} injections
D. Oral vitamin B_6 alone
E. Cocktail of all the vitamins B_2, B_6, B_{12}, and folinic acid

Answer: B Determination of B_6 responsiveness is key as vitamin B_6 can be the sole and effective lifelong treatment. Folate repletion with folinic acid supplementation is necessary to permit a pyridoxine response.

3. A 27-year-old woman presents with renal failure, malignant hypertension, and microangiopathic hemolysis. She has early-onset nystagmus and had an unexplained stroke a few years earlier. What do you do?

A. Start immunosuppressive therapy
B. Start plasmapheresis
C. Measure plasma total homocysteine, methionine, and methylmalonic acid and start hydroxocobalamin parenteral injection
D. Give high-dose folinic acid and betaine
E. Give vitamin B_6

Answer: C Atypical hemolytic-uremic syndrome associated with stroke and nystagmus points toward CblC deficiency. Hydroxocobalamin IM is urgent to correct hemolytic-uremic syndrome.

4. If a cystathionine β-synthase–deficient patient does not respond to B vitamin therapy, what other therapeutic option remains?

A. Enzyme replacement
B. Adding cysteine as *N*-acetylcysteine
C. Dietary methionine restriction
D. Aspirin administration
E. Stem cell therapy

Answer: C Dietary methionine restriction is the first line of treatment in pyridoxine-unresponsive CBS deficiency.

5. One should determine plasma total homocysteine if a patient presents with

A. Neural tube defect
B. Hypotonia
C. Stroke before the age of 18 years
D. Brittle hair
E. Liver dysfunction

Answer: C Stroke or thrombosis, especially in adolescents and younger adults, should prompt measurement of total homocysteine in plasma because treatments aimed at lowering of total homocysteine levels and reducing (if not abolishing) the vascular risk are available.

210 THE PORPHYRIAS

RICHARD J. HIFT

DEFINITION

The porphyrias are a group of eight disorders arising from disturbances in heme biosynthesis. Each disorder corresponds to abnormal activity of one of the enzymes that catalyze the heme biosynthetic pathway. Most porphyrias are genetic and heritable, with the exception of the sporadic form of porphyria cutanea tarda and rare instances of porphyria arising from acquired somatic mutations.

CLASSIFICATION

Approximately 90% of heme biosynthesis occurs within the erythron (specifically within erythroid precursors), leading to the production of heme for incorporation into hemoglobin. The remainder is synthesized in all other nucleated cells of the body, producing heme for incorporation into a number of vital hemoproteins. The liver is the predominant site of nonerythroid heme synthesis, with much of the product being incorporated into enzymes of the cytochrome P-450 (CYP) system.

Thus, the porphyrias may be classified into two categories, the erythropoietic and hepatic porphyrias. Disturbances in erythroid heme biosynthesis give rise to three forms of erythropoietic porphyria: congenital erythropoietic porphyria (CEP), erythropoietic protoporphyria (EPP), and X-linked protoporphyria (XLPP). Disturbances in nonerythroid heme biosynthesis give rise to five forms of hepatic porphyria: acute intermittent porphyria (AIP), variegate porphyria (VP), porphyria cutanea tarda (PCT), hereditary coproporphyria (HCP), and ALA dehydratase porphyria (ALADP).

A clinically applicable classification divides the porphyrias into another two cross-cutting groups (Table 210-1). The acute porphyrias are those characterized by the potential to develop the potentially fatal acute attack (AIP, VP, HCP, and ALADP). The nonacute porphyrias are accompanied predominantly by cutaneous manifestations (PCT, CEP, EPP, and XLPP).

EPIDEMIOLOGY

Each porphyria has a prevalence that may vary between populations, depending on local gene frequencies for the mutations that give rise to it. Assessment of prevalence is complicated by incomplete penetrance, difficulties in case ascertainment, and unevenness in the accuracy of biochemical diagnosis performed by different laboratories. The most accurate figures have recently been reported from Europe, where diagnostic laboratories in each country are centralized and linked into an international network. The network has reported a prevalence of 9.2 per million for EPP, 5.9 per million for AIP, and 3.2 per million for VP.[1] It is likely that the figures for North America will be broadly similar. PCT, which is environmentally induced and treatable, has an annual incidence exceeding 6 per million in Norway and nearly 4 per million in Sweden; its prevalence in the United States has been estimated at 4 per 100,000. HCP has a lower prevalence; the two autosomal recessive disorders, CEP and ALADP, are very rare, and experience is restricted to small case series and case reports.

AIP and VP may become prevalent in specific populations as a result of a founder effect, whereby a particular mutation spreads widely in an isolated population and becomes locally common. In northern Sweden, AIP has a prevalence estimated at 23 per million. In South Africa, VP has become the most common monogenetic disorder among the Dutch-descended immigrant population of South Africa, with a prevalence estimated at 1.2 per thousand of the white population.

PATHOBIOLOGY

The pathobiology of the porphyrias is easily understood by reference to the heme biosynthetic pathway (Fig. 210-1). There are slight differences between erythroid and nonerythroid heme synthesis. The initial step in the pathway is the synthesis of 5-aminolevulinate (ALA), catalyzed by the enzyme 5-aminolevulinate synthase (ALAS). The ubiquitous or housekeeping enzyme ALAS1 is strongly expressed in the liver and is encoded by the *ALAS1* gene on chromosome 3. The erythroid form, ALAS2, is encoded by the *ALAS2* gene on the X chromosome. The two genes share 73% homology.

TABLE 210-1 SUMMARY OF THE PORPHYRIAS

PORPHYRIA	KEY CLINICAL FEATURES AND LONG-TERM COMPLICATIONS	INHERITANCE
ACUTE PORPHYRIAS		
More Common		
Acute intermittent porphyria	Acute attacks Hypertension Renal failure Hepatocellular carcinoma	Autosomal dominant
Variegate porphyria	Vesiculo-erosive skin disease Acute attacks Hepatocellular carcinoma	Autosomal dominant
Less Common		
Hereditary coproporphyria	Vesiculo-erosive skin disease Acute attacks	Autosomal dominant
Rare		
ALA dehydratase porphyria	Acute attacks Neuropathy	Autosomal recessive
NONACUTE PORPHYRIAS		
More Common		
Porphyria cutanea tarda	Vesiculo-erosive skin disease Associated with iron loading, alcoholic liver disease, hepatitis C, HIV infection, renal failure, and other disorders	Acquired, sometimes on a background of an inherited mutation
*Erythropoietic protoporphyria	Immediate photosensitivity Cholelithiasis, liver disease	Autosomal recessive; frequently in association with a common polymorphism in the ferrochelatase gene
Rare		
*X-linked protoporphyria	Immediate photosensitivity Cholelithiasis, liver disease	X-linked
*Congenital erythropoietic porphyria	Vesiculo-erosive skin disease Severe photomutilation	Autosomal recessive

*The three forms of erythropoietic porphyria. The remainder are classified as hepatic porphyrias.

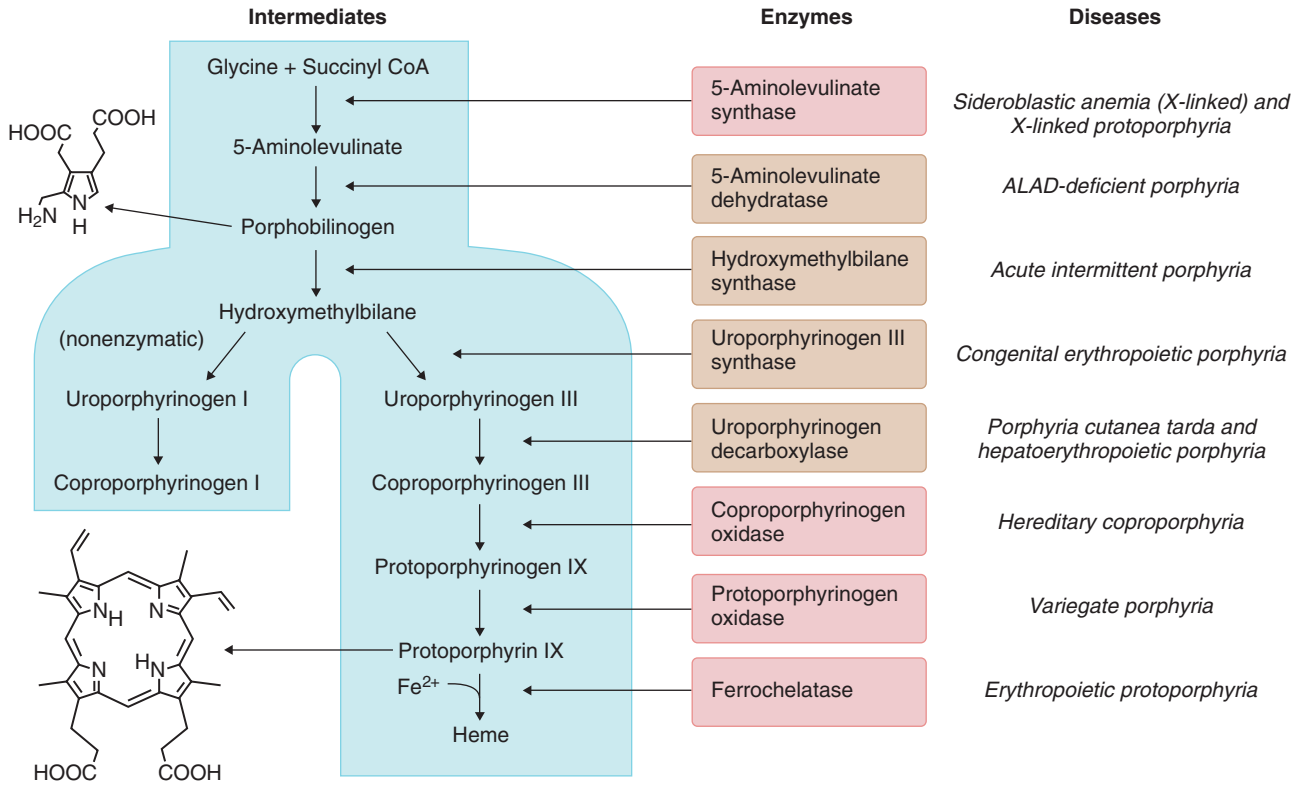

FIGURE 210-1. The heme biosynthetic pathway. Heme is synthesized through a series of porphyrin precursors and porphyrin intermediates, catalyzed by specific enzymes as indicated. The form of porphyria associated with abnormal enzyme function is shown on the right.

The remaining genes are common to both systems, although transcriptional variations lead to minor differences in size between the erythropoietic and hepatic forms of porphobilinogen synthase (previously known as ALA dehydratase), hydroxymethylbilane synthase, and uroporphyrinogen-III synthase.

Heme production is tightly coupled to heme requirement, and the flux of metabolites through the pathway is regulated by a process of negative feedback inhibition. The synthesis of ALA is the rate-limiting step in the pathway. In the liver, as heme use is increased (e.g., for incorporation into CYP), there is a reduction in a postulated hepatic regulatory free heme pool, the transcription of *ALAS1* is initiated, and ALAS1 activity increases. Conversely, when the requirement for heme drops, ALAS1 is repressed, leading to a reduction in porphyrin synthesis. Regulation in the erythropoietic system is somewhat different. Control is exercised through the binding of iron regulatory proteins to an iron response element in the 5′ region of the *ALAS2* mRNA. The extent of downregulation or upregulation of translation is controlled by the availability of iron.

The structure of the biochemical intermediates of heme biosynthesis as well as of the heme catabolic products biliverdin and bilirubin is shown in E-Figure 210-1. After ALAS-mediated synthesis, ALA is converted into porphobilinogen (PBG), which has a monopyrrolic ring structure. These two compounds, ALA and PBG, are classified as porphyrin precursors. Four PBG molecules are then combined to form the linear tetrapyrrole hydroxymethylbilane, which is then enzymatically cyclized by the enzyme uroporphyrinogen-III synthase, resulting in the tetrapyrrolic macrocycle uroporphyrinogen III. This is the first of a series of porphyrinogens. Spontaneous cyclization is possible but produces the series I isomer, which is not physiologic and is not further metabolized. Sequential modification of the porphyrinogen side chains results in the production of a sequence of porphyrinogens and their oxidized counterparts, the porphyrins. In the final step, iron is incorporated into the macrocycle, resulting in the functional heme molecule. The distinction between porphyrin precursors and porphyrins is important in diagnosis and in predicting the clinical presentation of each form of porphyria. The enzymes ALAS, coproporphyrinogen oxidase, protoporphyrinogen oxidase, and ferrochelatase are all mitochondrial, whereas the remaining enzymes are cytosolic. Heme synthesis therefore begins and ends in the mitochondrion, with intermediate metabolism occurring in the cytoplasm.

The porphyrin precursors and the initial porphyrins in the pathway are water soluble, circulate freely in the plasma, and are largely excreted in urine. The later porphyrins are hydrophobic, may be protein bound in the plasma, and undergo biliary excretion, eventually being excreted in the stool. The pattern of accumulation and excretion of precursors and porphyrins is exploited for the biochemical diagnosis of porphyrias and ultimately underlies the varying presentations of the acute and nonacute porphyrias. The acute attack is always associated with elevations in the precursors, and the acute porphyrias are therefore those in which such an accumulation occurs. Porphyrins are photosensitive molecules, and those forms of porphyria in which large amounts of porphyrin accumulate in plasma and skin have photosensitive skin disease as their major clinical presentation. In the nonacute porphyrias, the levels of precursors remain unchanged, accounting for the lack of acute attacks. Because porphyrins are typically elevated, they are manifested with skin disease alone. Two forms of porphyria, HCP and VP, may demonstrate both porphyrin elevation and a propensity to periodic elevation in the precursors; these porphyrias are therefore characterized by both skin disease and the risk of an acute attack.

Etiology

Acute Porphyrias

Acute Intermittent Porphyria

AIP is inherited as an autosomal dominant disorder. Because the enzyme block is early in the pathway, elevation of the precursors is characteristic and leads to a potential for development of the acute attack but not skin disease. Although an excess of nonphysiologic series I porphyrin isomers is typically observed in urine, this results from spontaneous cyclization of accumulated PBG, particularly in the bladder. Rare cases of homozygous AIP have been described.

Variegate Porphyria

VP is inherited as an autosomal dominant disorder. The *PPOX* gene is carried on chromosome 1. The number of known mutations approaches 200. Penetrance is estimated at approximately 37%. VP is associated with elevations in porphyrins, leading to photosensitivity, and also with acute attacks, during which phase of the illness the levels of precursors become markedly elevated. This has been ascribed to an allosteric inhibition of hydroxymethylbilane synthase by coproporphyrinogen and protoporphyrinogen, which accumulate in VP. Occasional cases of homozygous VP have been described. Such cases are either homozygous for mutations associated with some residual enzymatic activity or are compound heterozygotes, in which a mutation on one allele that results in complete loss of catalytic activity is accompanied on the other by a mutation associated with some residual catalytic activity. The homozygous state for mutations associated with complete loss of enzymatic activity is lethal.

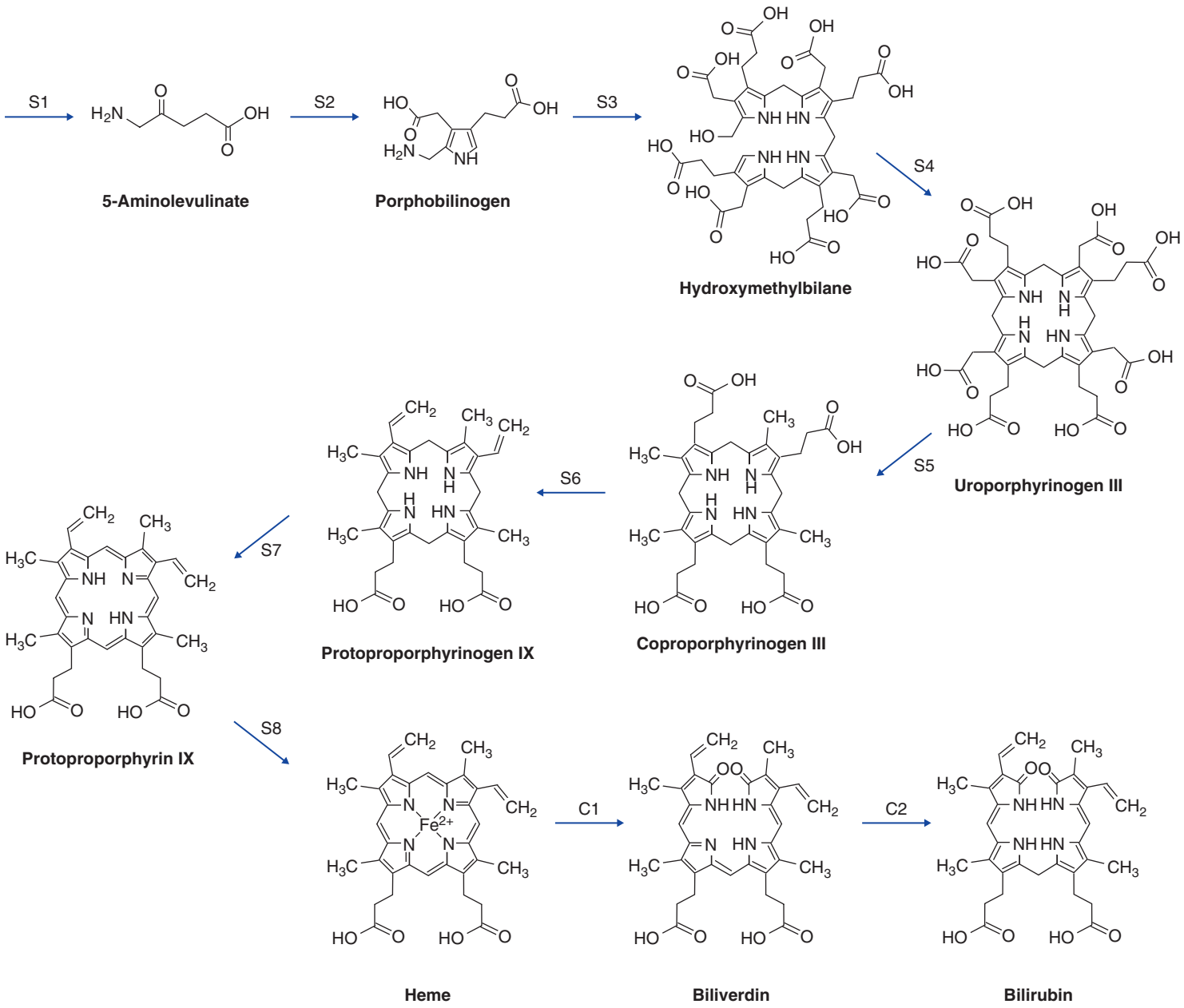

E-FIGURE 210-1. Heme biosynthesis and catabolism. S1-S3. Synthesis of the porphyrin precursors and the first tetrapyrrole. S1. 5-Aminolevulinate is synthesized from glycine and succinyl coenzyme A, catalyzed by the mitochondrial enzyme ALA synthase. S2. Two molecules of 5-aminolevulinate condense to form a cyclic pyrrole, porphobilinogen, catalyzed by the cytosolic enzyme ALA dehydratase. S3. Four molecules of porphobilinogen are combined to form an open tetrapyrrole, hydroxymethylbilane, catalyzed by the cytosolic enzyme hydroxymethylbilane synthase. **S4-S7. Sequential modification of tetrapyrrolic porphyrinogens, porphyrins, and heme.** S4. The cytosolic enzyme uroporphyrinogen-III synthase catalyzes ring closure and brings about a specific isomerization, forming a closed macrocycle. This is uroporphyrinogen III. In the absence of the enzyme, spontaneous cyclization occurs, giving rise to uroporphyrinogen I. This nonphysiologic isomer is not available for further enzymatic conversion. S5. Uroporphyrinogen III undergoes four sequential decarboxylation reactions all catalyzed by the cytosolic enzyme uroporphyrinogen decarboxylase, producing sequentially heptacarboxylic porphyrinogen, hexacarboxylic porphyrinogen, pentacarboxylic porphyrinogen (not shown), and coproporphyrinogen. S6. Coproporphyrinogen III re-enters the mitochondrion and is oxidized by coproporphyrinogen oxidase to protoporphyrinogen IX. S7. Protoporphyrinogen IX is oxidized to protoporphyrin IX by the mitochondrial enzyme protoporphyrinogen oxidase. S8. Elemental iron is inserted into the macrocycle by the mitochondrial enzyme ferrochelatase, forming heme. **Heme catabolism.** C1. Catalyzed by heme oxygenase, the tetrapyrrolic macrocycle is opened, releasing iron and carbon monoxide and forming biliverdin. C2. A reduction step catalyzed by biliverdin reductase results in bilirubin, which is bound to albumin and other plasma proteins and excreted.

Hereditary Coproporphyria

HCP is an autosomal dominant disorder. The *CPOX* gene is carried on chromosome 7, and more than 50 disease-associated mutations have been described. Both acute attacks and skin disease are encountered; the mechanism for the acute attack is similar to that of VP, with hydroxymethylbilane synthase being allosterically inhibited by accumulated coproporphyrinogen. Rare cases of homozygous HCP have been described.

ALA Dehydratase Porphyria

This is a rare autosomal recessive disorder, described in fewer than 10 cases. The *ALAD* gene is carried on chromosome 9.

Nonacute Porphyrias

Porphyria Cutanea Tarda

This disorder results from reduced activity of the enzyme uroporphyrinogen decarboxylase (UROD). PCT is unique among the porphyrias in that it is in most cases not inherited. Approximately 75% of patients present with sporadic PCT. In these cases, the *UROD* gene is normal, but patients will demonstrate a reduction in UROD activity. This is due to a chemically mediated inhibition of the enzyme. The inhibitor has been identified as uroporphomethene, an aberrant oxidative product of the normal substrate of the enzyme uroporphyrinogen. Sporadic PCT has a number of specific associations. First, nearly all cases are associated with some degree of hepatic iron loading. It is thought that the iron functions as an oxidant, facilitating the formation of the uroporphomethene inhibitor. Second, many cases will show evidence of liver dysfunction, commonly due to alcohol or hepatitis C viral infection. A number of other miscellaneous factors may also precipitate PCT, including HIV infection, estrogen exposure, renal failure, lymphoma, systemic lupus erythematosus, and toxins such as hexachlorobenzene. The relationship of viral infection and liver disease to PCT predisposition is not understood. It has been suggested that common factors may be downregulation of hepcidin, leading to iron overload, and increased oxidative stress, with these two factors potentiating the inhibitory mechanism.[2] An association with diabetes has also become evident. Reversal of the UROD inhibition results in biochemical and clinical remission.

In some cases, iron loading has been shown to relate directly to the inheritance of mutations in genes responsible for iron regulation. For example, the C282Y mutation in the *HFE* gene commonly found in white patients with hereditary hemochromatosis (Chapter 212) is over-represented in PCT. In other cases, the reason for the iron loading is less clear.

Approximately 25% of cases of PCT are classified as familial PCT. In these patients, an inherited mutation in a *UROD* allele can be shown. The resulting 50% reduction in UROD activity is not in itself sufficient to precipitate clinical expression. When disease does become clinically manifested, these patients commonly show evidence of the same precipitating factors described in sporadic PCT, but they may present at a younger age. Disease therefore results from a combination of an inherited mutation and environmental factors that result in inhibition of the remaining functional enzyme.

Rare cases of homozygous PCT have been described. These are known as hepatoerythropoietic porphyria and are manifested with severe photomutilation resembling that seen in CEP.

Erythropoietic Protoporphyria

EPP is inherited in a complex fashion. The *FECH* gene is carried on chromosome 18, and more than 100 disease-associated mutations have been reported. Molecular analysis demonstrates the presence of an inherited mutation on one allele resulting in reduced enzymatic activity. However, this is insufficient in its own right to result in clinical symptoms. A small proportion will be found to be homozygous, resulting in sufficiently reduced ferrochelatase activity for the disease to become clinically manifested. Approximately 94%, however, can be shown to have coinherited a *FECH* polymorphism, prevalent in white populations, that is associated with moderately reduced ferrochelatase activity. This *FECH**IVS3-48C low-expression allele is subject to aberrant splicing and decreased stability of the transcript, resulting in low expression. The gene frequency may reach 11% in European populations. The summative effect of the family-specific mutation and low-expression allele is sufficient to reduce ferrochelatase activities to a level below approximately 35%, at which stage the clinical syndrome may develop. In the strict sense, EPP is a recessive disorder, but the high prevalence of the low-expression allele in the population means that the compound heterozygous state occurs commonly in families carrying an EPP mutation. Thus, the prevalence is much higher than would typically be expected of an autosomal recessive disorder; its inheritance has been described as pseudodominant. A recent study in North America of 155 unrelated patients reported that 136 carried the combination of a loss-of-function *FECH* mutation and the low-expression allele, whereas only three carried two loss-of-function mutations. The remaining 15 patients were shown to have XLPP.[3]

X-Linked Protoporphyria

This is the most recent form of porphyria to be identified. The clinical presentation is nearly identical to that of EPP, and its existence as an independent entity was not suspected until a subgroup of families with EPP were found not to carry ferrochelatase mutations. XLPP is associated with a number of mutations occurring in a sharply restricted region at the C terminus of the *ALAS2* gene. These mutations appear to modify an important control region leading to stabilization of the mRNA transcript or, possibly, enhanced access of succinyl coenzyme A to ALAS, either of which will result in abnormally elevated ALAS2 activity. These are therefore gain-of-function mutations that result in an increased flux of porphyrins through the erythroid heme biosynthetic pathway. Both ferrochelatase and iron availability become rate limiting, such that large amounts of protoporphyrin cannot be further metabolized to heme but are diverted to the plasma. Because the pathophysiologic mechanism for EPP is also dependent on the overproduction of protoporphyrin, the two disorders are nearly identical clinically. The *ALAS2* gene is carried on the X chromosome, making XLPP an X-linked disorder. Loss-of-function mutations in the *ALAS2* gene are responsible for X-linked sideroblastic anemia (Chapter 159), a condition unrelated to XLPP, that is associated with gain-of-function mutations. Erythrocyte protoporphyrin levels tend to be higher in patients with XLPP than in those with EPP, with a higher proportion being zinc chelated. Penetrance appears to be near 100%, and liver complications are more prevalent. Transmission is as expected of an X-linked disorder, one manifestation of which is the absence of father to son transmission.

Somatic Mutations

Very rarely, non-germline tissue-specific mutations may result in manifest porphyria. Cases of CEP, EPP, and VP have been described in neoplastic and paraneoplastic settings, such as myelodysplasia, myeloproliferative disorders, and hepatocellular carcinoma.

Pathogenesis

Photosensitivity

In those porphyrias associated with skin disease, porphyrins are found in high concentrations in plasma, skin, and blister fluid. Porphyrins are fluorescent. Stimulation by light results in an excitation of the porphyrin molecules, the promotion of electrons to a higher energy state, and the production of singlet oxygen. Relaxation to the ground state is accompanied by loss of energy manifesting as a radiation of red light. In the skin, this energy may be transferred to biologic molecules, resulting in oxidation of membrane lipids, polypeptides, and nucleic acids, thus accounting for the skin disease. The most potent wavelengths for porphyrin excitation lie within the ultraviolet spectrum, in the Soret band between 400 and 410 nm. Four additional absorption bands are present in the range of 500 to 700 nm; thus even visible light, against which most sunscreens are ineffective, is harmful to the skin in subjects with a cutaneous porphyria.

Pathologic examination will reveal epidermal bullae, duplication of basement membranes, and deposition of hyaline material, which appears to be associated with fibrin, immunoglobulins, and complement in and around the blood vessels of the dermis, suggesting that these vessels may be the principal target for light-induced injury.

Acute Attack

Elevated concentrations of the heme precursors ALA and PBG are always present in patients during the acute attack, and remission is commonly accompanied by a reduction in these concentrations. A causal role for either or both of these molecules has therefore long been suspected but never unequivocally proven. ALA is structurally similar to known neurotransmitters such as glutamine and γ-aminobutyric acid. Given the central role of nervous system dysfunction in the acute attack, it is suspected that ALA may be directly involved. Alternatively, it may serve as a proxy marker for some other as yet unidentified neurotoxin. An earlier contrasting hypothesis, that the neuropathy associated with an acute attack is mediated by intraneuronal heme deficiency, appears less likely because recent clinical experience has shown that liver transplantation cures AIP and VP and has induced acute

attacks of porphyria in previously nonporphyric recipients who received an explanted porphyric liver as part of a domino transplant.

On histologic examination, nerve damage is characterized by axonal loss, although some degree of segmental demyelination may be present. There may be a reduction in intradermal nerve fiber density on skin biopsy. Skeletal muscle may show neurogenic atrophy. Nerve conduction studies, although not pathognomonic, tend to show a fairly characteristic pattern suggesting axonal neuropathy with relatively little evidence of demyelination. Upper limbs may be affected more than the lower limbs. Sensory nerves may be variably affected; electromyography initially shows a pattern of denervation, with widespread fibrillation, later replaced by a pattern of reinnervation marked by polyphasic motor unit potentials with increased amplitude and duration.

Induction of Porphyria

The block in porphyrin synthesis with consequent heme deficiency that characterizes all forms of porphyria does not appear to result in any direct clinical adverse effect. Although patients with CEP, EPP, and XLPP may be mildly anemic, this is in part secondary to hemolysis in CEP and iron deficiency in XLPP. In the acute porphyrias, hemoproteins such as CYP and tryptophan pyrrolase will reveal evidence of heme desaturation, again without obvious clinical consequences. The principal effect of the heme deficiency is the derepression of ALAS1, thus substantially increasing the flux of porphyrins through the pathway, accentuating the rate-limiting effect of the enzyme block and resulting in a significant overproduction of porphyrins and, possibly, precursors, which then result in the characteristic clinical syndromes. This mechanism is key to the understanding of the porphyrias.

The development of the acute attack is strongly associated with hyperinduction of ALAS1, typically by increased gene transcription in response to a reduction in free heme concentrations. Whereas the resultant increased activity of the heme biosynthetic pathway leads to an appropriate increase in heme levels in normal individuals, after which ALAS1 is suppressed, in patients with an acute porphyria, the defective enzyme is rate limiting; adequate heme concentrations are not reached, thereby leaving the pathway in a state of hyperinduction, and ALA and PBG accumulate in quantities sufficient to induce clinical symptoms. The most common precipitant of the acute attack is exposure to a number of drugs that share the ability to induce ALAS1. Such drugs are termed porphyrogenic.

This synthesis of CYP apoenzyme and heme is tightly correlated. The most powerfully porphyrogenic compounds include multifunctional inducers that induce multiple hepatic microsomal enzymes, those that induce the CYP3A and CYP2C9 subclasses, and those that are associated with irreversible mechanism-based inhibition of CYP. Such inhibition results in destruction of the enzyme, release of heme (which is then catabolized by heme oxygenase, leading to a reduction in the free heme pool), and consequently ALAS1 induction. These processes are mediated by nuclear receptors, particularly the constitutively active receptor and the pregnane xenobiotic receptor. Now that these mechanisms are understood, it is possible with high accuracy to predict which drugs are most likely to be porphyrogenic.

Calorie deprivation is known to induce porphyrin synthesis and even the acute attack, whereas glucose administration has a suppressive effect. This so-called glucose effect has been shown to be mediated by the transcriptional coactivator PGC-1α, which is induced when the liver shifts from the use of glucose as an energy substrate to β-fatty acid oxidation, activating the ALAS1 promoter and increasing heme synthesis.

Patients carrying a gene for an acute porphyria do not respond uniformly or predictably to drug exposure. Some patients will not respond at all, others may show some biochemical evidence of increased porphyrin production, and others will develop severe clinical symptoms. The reason for this variation in response or for the observation that acute attacks are extremely rare before puberty is not well understood. It is thought that the variability may result from the inheritance of polymorphisms in other genes responsible for drug metabolism, including cytochrome P-450, and possibly by metabolome-level variations, in other words, a complex interaction of genetic, biochemical, and hormonal interactions at a particular point in time. Similarly, the reason for the incomplete penetrance of many of the porphyrias is not yet understood; some 60% of patients carrying a gene for VP, for instance, fail to manifest biochemical or clinical evidence of disturbed porphyrin synthesis.

Women with AIP may show a pattern of regularly recurring attacks associated with the luteal phase of the menstrual cycle, although this is fortunately uncommon. It appears that endogenous hormone production is sufficient to stimulate ALAS and to cause an acute attack. Acute attacks have also been ascribed to calorie deprivation. Both stress and infection have been listed as possible inducers of the acute attack, although the evidence for this is weak.

CLINICAL MANIFESTATIONS

Patients with CEP, EPP, and the homozygous forms of the acute porphyrias usually present in childhood. This is an extremely rare occurrence in the three dominantly inherited acute porphyrias, in which both biochemical and clinical evidence of disease expression is typically delayed until after puberty. Although adult patients may present at any age, the first presentation is typically in the third decade.

Acute Porphyrias

The cardinal manifestation of the acute porphyrias is the acute attack. This metabolic crisis is characteristically manifested as severe, generalized abdominal pain, felt throughout the abdomen and sometimes in the lower back, buttocks, and thighs. The pain is severe and requires opioids for relief. It is not associated with peritonitis, and abdominal examination is typically unremarkable. There is autonomic overactivity manifested as hypertension, tachycardia, and gastrointestinal dysfunction, typically vomiting, constipation, and occasionally ileus.

Severe acute attacks are associated with a number of other features. Hyponatremia is common and when severe may lead to seizures and altered consciousness, particularly when hypotonic intravenous fluids are administered. Although often ascribed to the syndrome of inappropriate antidiuretic hormone (Chapter 116), the pattern of electrolyte excretion frequently suggests renal salt wasting, sometimes associated with marked urinary losses of potassium, calcium, and magnesium.

A severe, untreated acute attack may be complicated by a rapid-onset motor neuropathy, usually developing 24 hours or more after the onset of the abdominal pain. Very occasionally, a patient will present with neuropathy with a history of minimal or no abdominal pain; this may lead to a delay in suspecting porphyria as the cause. The neuropathy is typically symmetrical and may affect proximal muscles predominantly. Although motor signs predominate, there may be some sensory involvement in a central, so-called bathing suit distribution. The neuropathy may result in quadriparesis. Weakness of the respiratory muscles is common and may result in respiratory failure requiring ventilation. Once established, motor neuropathy typically requires months for recovery. Occasional patients may develop a prominent small-fiber neuropathy after attacks that is manifested with a generalized dysesthesia and hyperalgesia.

Cranial nerve or cerebellar involvement is occasionally noted, with facial and vagus nerve involvement in particular, although involvement of the trigeminal, hypoglossal, accessory, and oculomotor nerves has also been observed. The most extreme cases may develop the posterior reversible encephalopathy syndrome with radiologic evidence of reversible cerebral ischemia. On occasion, autonomic overactivity is so severe as to resemble a pheochromocytoma crisis (Chapter 228).

There is a widespread and unjustified misperception that psychiatric manifestations are a prominent part of the symptoms of the acute porphyrias. Claims that historical figures such as King George III of England and his relatives and Vincent van Gogh had porphyria all perpetuate the myth that AIP is associated with chronic insanity but are not supported by convincing evidence.[4] During the acute phase, the acute attack frequently is manifested with anxiety and sometimes with a short-lived psychotic episode, which reverses completely with remission of the attack. A statistical association with chronic anxiety and depression in patients with AIP has been shown. There is no association with chronic psychosis or a need for institutionalization.

The acute attack is more common in females than in males, is extremely rare before puberty, and becomes uncommon from the sixth decade onward. Pregnancy may precipitate acute attacks, but this is infrequent. Large studies suggest that there may be a slight increase in risk of perinatal death in women with the acute porphyrias; in a study of 136 deliveries, this did not reach statistical significance, but significance was shown when restricted to first pregnancies alone.[5]

The acute attack has become less common in recent decades as a result of earlier diagnosis and careful attention to its prevention. It is uncommon for a patient who has had an attack to suffer a recurrence once preventive measures have been instituted. Patients repeatedly exposed to porphyrogenic medication, including recreational drugs, may suffer recurrent acute attacks. A very small number of patients may show a course characterized by recurrent acute attacks of unknown etiology; this is particularly a feature of some young women with AIP. In some of these women, there is a clear relationship

to the menstrual cycle. In the remainder, such an association is not obvious, and a vicious circle of recurrent attacks and hospitalizations becomes established. Such patients become severely debilitated and cachectic, show evidence of accumulating neuronal damage, experience a poor quality of life, and may ultimately die.

Other Manifestations

Patients with AIP are prone to develop chronic hypertension and renal dysfunction, probably due to chronic activation of the sympathetic nervous system. A strong and unexplained association between both AIP and VP and noncirrhotic hepatocellular carcinoma (Chapter 196) has been reported from several centers. Swedish patients with AIP older than 50 years have been shown to be at 86-fold increased risk of hepatocellular carcinoma.[6] The risk is significantly higher in women. There is evidence of an increased risk in VP as well, but surprisingly, such an association has not been seen in South Africa despite the frequency of VP in that population.

Homozygous Acute Porphyrias

Homozygous AIP is a serious disorder that may be associated with severe neurodevelopmental abnormalities, including porencephaly, psychomotor and developmental retardation, ataxia, epilepsy, cataracts, and a number of other neurologic manifestations. It is usually fatal in childhood. Homozygous VP is associated with skeletal dysmorphism, severe skin disease, nystagmus, seizures, a sensory neuropathy, and cognitive impairment. An association with cerebral demyelination, detectable on magnetic resonance imaging, has recently been described. Despite the severity of the symptoms, there is no appreciable early mortality, and for reasons as yet unknown, acute attacks are not a feature of homozygous VP. Homozygous HCP takes two forms: the first resembles homozygous VP with small stature, photosensitivity, psychomotor retardation, and neurologic defects; the second is manifested at birth with hemolytic anemia and severe jaundice. This variety is associated with specific mutations in the *CPOX* gene that specifically block an intermediate stage in the oxidation of coproporphyrinogen and protoporphyrinogen, resulting in the accumulation of a harderoporphyrinogen intermediate, and is known as harderoporphyria.

ALA Dehydratase Deficiency Porphyria

This is an extremely rare recessive disorder that may be manifested in either childhood or adulthood, depending on the severity of the phenotype, typically with chronic neuropathy, other neurologic symptoms, and acute attacks.

Vesiculo-erosive Skin Disease

The characteristic skin disease of VP, HCP, PCT, and CEP is described as vesiculo-erosive. Patients present with blistering and erosions, typically in response to minor skin trauma, in sun-exposed areas, particularly the dorsal surface of the hands and forearms, the face, and, if sun exposed, the nape of the neck and the dorsal surfaces of the feet. There is no immediate photosensitivity, and the changes develop insidiously. Therefore, patients frequently fail to make the association between sun exposure and skin damage. The lesions heal slowly, leaving a residuum of scarring and small, localized areas of hypopigmentation or hyperpigmentation (Fig. 210-2). Milia may be present, particularly on the dorsal surfaces of the hands and in the digital clefts.

It is unusual for the skin disease of VP and HCP to develop beyond this. Patients with PCT, however, may show marked facial hypertrichosis, hyperpigmentation, and sometimes alopecia. They may also develop thickening of the skin of the fingers and hands; these pseudosclerodermoid changes may occasionally lead to a misdiagnosis of localized scleroderma.

The most severe skin disease is characterized by marked photomutilation, including loss of skin appendages such as the nose, ears, and lips. This is seen in CEP and in hepatoerythropoietic porphyria. The skin disease of homozygous VP is marked by an accentuation of the features characteristic of heterozygous VP, alopecia and resorption of digits, which may be due to photo-osteolysis (Fig. 210-3) or, alternatively, may represent part of the skeletal dysmorphism characteristic of the disorder.

Porphyria Cutanea Tarda

PCT typically presents in middle-aged and older subjects who develop characteristic skin lesions in sun-exposed areas. Acute attacks are not a feature. Given the strong association with iron loading and liver disease, clinical and biochemical assessment commonly reveals evidence of increased iron storage, liver dysfunction, alcohol abuse, or renal dysfunction. In the less common setting of HIV infection, systemic lupus erythematosus, or lymphoma, clinical evidence for these disorders will be present.

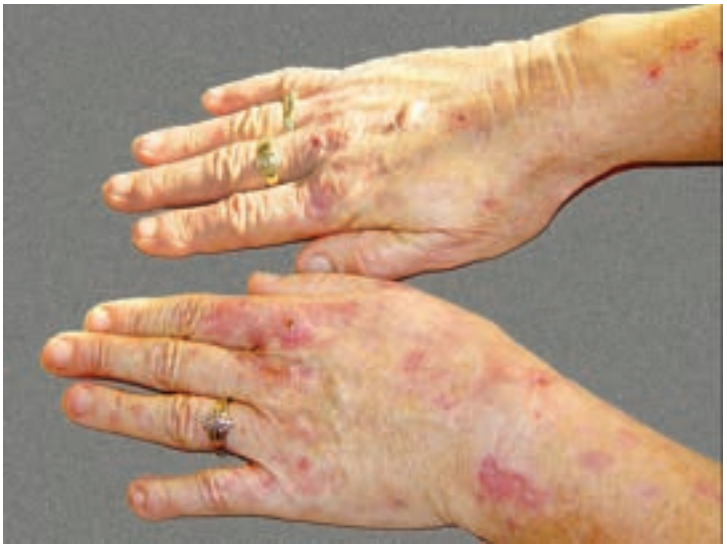

FIGURE 210-2. The hands in variegate porphyria. The characteristic lesions are bullae, shallow erosions that develop scabs and heal slowly, leaving areas of hypopigmentation and hyperpigmentation. The skin disease of porphyria cutanea tarda and hereditary coproporphyria is similar.

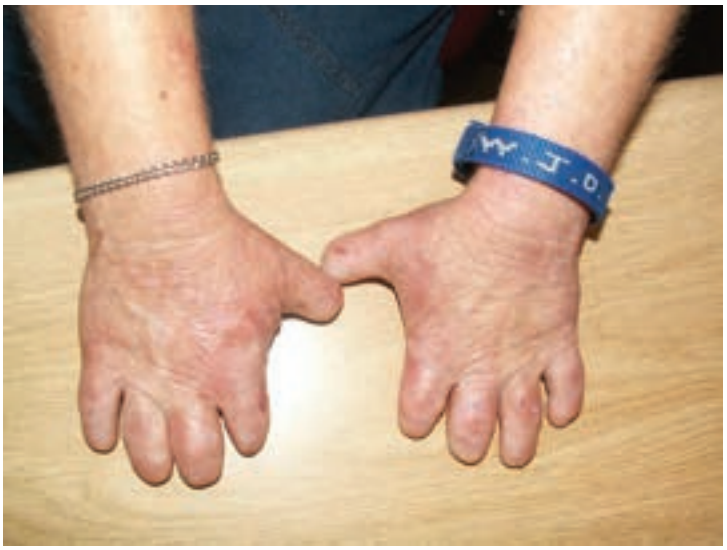

FIGURE 210-3. The hands of a patient with a homozygous form of variegate porphyria. In addition to the characteristic vesiculo-erosive skin lesions, there is marked brachydactyly, representing both photo-osteolysis and a skeletal developmental defect.

Congenital Erythropoietic Porphyria

This rare autosomal recessive disorder is associated with a spectrum of disease severity. Severely afflicted patients demonstrate photomutilation, with scarring, loss of skin appendages (such as ears, nose, lips, fingernails, and digits), ulcerative keratitis, and corneal scarring. A form of immediate photosensitivity after sunlight exposure and characteristic pink erythematous facial papules have been described.[7] Patients may show erythrodontia, osteodystrophy, hypercellular bone marrow, hemolytic anemia, and splenomegaly. Other patients are more mildly affected and may present later in life. The skin damage deteriorates progressively with increasing age. Impact on quality of life and psychosocial consequences may be severe. Prenatal cases presenting in utero with severe anemia associated with hydrops fetalis have been described. There is a variable genotype-phenotype correlation. Although some mutations tend to be associated with more severe disease, the severity may vary between patients even though they carry the same mutation. The most predictive features of a severe course are early age at onset and the presence of hematologic complications, particularly severe anemia and

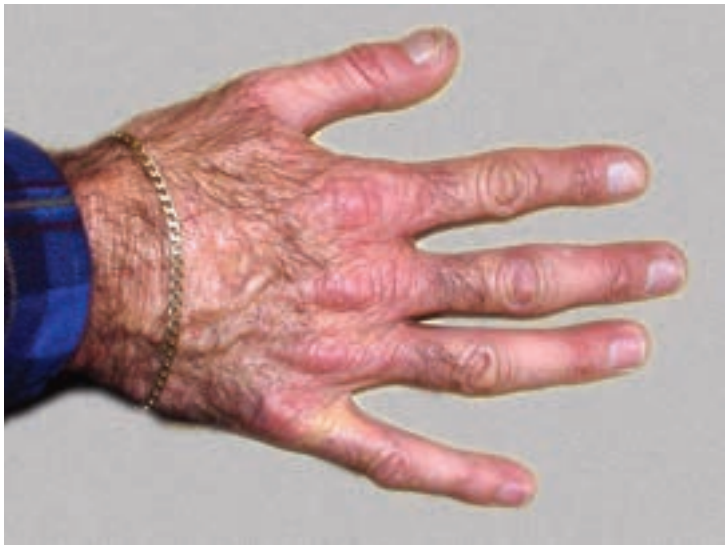

FIGURE 210-4. The hands in erythropoietic protoporphyria. There is thickening and grooving of the skin over the knuckles.

thrombocytopenia. One study has suggested that *ALAS2* acts as a modifier gene, with severity of phenotype being modulated by mutations in that gene.[8]

Skin Disease Associated with Immediate Photosensitivity

Erythropoietic Protoporphyria

Patients with EPP and XLPP do not manifest the typical vesiculo-erosive pattern of skin disease described before but develop a characteristic pattern of immediate photosensitivity. Patients will report that after a period of sun exposure, they develop severe discomfort and pain in sun-exposed areas. This may be associated with erythema and edema. The discomfort may take 24 to 48 hours to settle after cessation of sun exposure. The onset of the illness is often in childhood but diagnosis is frequently delayed, and the reason for a child's reluctance to remain outdoors is often not recognized immediately. Chronic skin changes are minimal and are usually limited to the development of a waxy thickening and grooving of the skin, typically over the bridge of the nose and over the knuckles (Fig. 210-4). A mild microcytic hypochromic anemia may be present. A subset of patients develop a condition known as seasonal palmar keratoderma; it has been shown that these patients carry homozygous or compound heterozygous ferrochelatase mutations.

Approximately 10% of patients with EPP will develop hepatobiliary disease secondary to the massive accumulation of protoporphyrin within the hepatocytes and biliary porphyrin excretion, typically with cholelithiasis secondary to elevated biliary porphyrin concentrations. Approximately 2% may present with severe and potentially life-threatening liver disease, including cirrhosis and liver failure.

X-Linked Protoporphyria

XLPP and EPP cannot for practical purposes be distinguished clinically. XLPP is likely to have a higher penetrance as well as a typical X-linked pattern of inheritance. Patients are more prone to liver disease than are those with EPP. Some patients will report a correlation between iron deficiency and severity of symptoms.

Other Settings

A transient immediate acute photosensitivity is occasionally observed in patients with VP as they emerge from an acute attack. This is sometimes associated with acute loss of fingernails. Some patients with CEP may manifest immediate photosensitivity in addition to the characteristic vesiculo-erosive skin response. Immediate photosensitivity is observed in some patients in whom the synthetic metalloporphyrin tin protoporphyrin was administered as a heme oxygenase inhibitor to prolong remission with repeated acute attacks of AIP. It is also noted in patients treated with ALA or synthetic porphyrin analogues as part of photodynamic therapy for cancer.

DIAGNOSIS

An accurate diagnosis is essential in the management of porphyria. Although the pattern of symptoms and skin manifestations may suggest the diagnosis, a clinical diagnosis alone is notably inaccurate, given the varying and often nonspecific manifestations of the disorders as well as the inexperience of most physicians in dealing with them. A failure to diagnose the porphyrias may have serious consequences, with unnecessarily impaired quality of life and, in some, the potential for a possibly fatal acute attack. Conversely, it is not unusual for patients, typically with a history of frequent, unexplained abdominal pain, to be erroneously labeled as having an acute porphyria.

Biochemical Diagnosis

Biochemical analysis is central to the diagnosis and evaluation of the porphyrias. For AIP, hydroxymethylbilane synthase assays have been widely used but are now discouraged owing to their inaccuracy. Recent reports have highlighted the gain in diagnostic accuracy that results from restricting diagnostic testing for porphyria to a small number of national reference laboratories that analyze sufficient cases to develop expertise and that cooperate in a quality enhancement network.

The pattern of porphyrin accumulation in urine, stool, plasma, and erythrocytes that characterizes a specific porphyria forms the basis of diagnosis. In practice, this is complicated by the varying water solubility of the precursors and porphyrins, leading to differential patterns of accumulation in urine, stool, and plasma. It must be stressed that examination of urine alone may lead to both misdiagnosis and misclassification of the porphyria. Given the differential excretion of porphyrins, it is essential that testing include urine, stool, and plasma. Where an erythropoietic porphyria is suspected, an erythrocyte porphyrin analysis must be performed as well.

Porphyrin analysis is typically performed with a high-performance liquid chromatographic separation technique with fluorometric detection of the porphyrins. The series I and III isomers can be distinguished and the concentration of each individual porphyrin species quantitated by reference to standard specimens. Differentiation of isomers is particularly important in the identification of HCP. The erythropoietic porphyrias are easily confirmed by assessing the biochemical profile of porphyrins in erythrocytes.

A useful screening assessment is plasma fluorescence scanning. The fluorescence emission maximum varies between different porphyrins and with the extent of their protein binding. When it is subjected to ultraviolet light, a plasma sample will typically demonstrate an emission peak at approximately 630 nm in EPP, 625 nm in VP, and 619 nm in AIP, HCP, and PCT.

The acute attack is always associated with an elevation of the porphyrin precursors ALA and PBG. The first step in the assessment of the confirmation of an acute attack is therefore to submit urine for determination of these levels, specifically PBG. In the appropriate clinical setting, elevated levels are highly confirmatory; and conversely, when these are normal, the diagnosis must be reconsidered. Some laboratories are able to measure ALA and PBG in plasma, which may have a slight advantage in accuracy. For immediate use in the emergency setting, semiquantitative test kits that allow the identification of elevated urinary PBG concentrations without the use of specialized equipment have been developed. It is recommended that every emergency department have ready access to such a kit.

Genetic Testing

Given that the various porphyrias result from mutations in different genes and that within each porphyria, numerous family-specific mutations may give rise to the characteristic phenotype, molecular diagnosis is not well suited to the primary evaluation of a patient. No mutation-specific diagnostic test will exclude any form of porphyria other than the one specifically associated with that gene and that mutation. It is for this reason that biochemical analysis is recommended as the first step. Furthermore, whereas biochemical testing, by quantitating precursor and porphyrin levels, provides information on the degree of activity of the porphyria at a particular moment, molecular techniques do not. Once a biochemical diagnosis of porphyria has been made, the underlying mutation should be identified. Screening of family members for this mutation will then determine carrier status. This is particularly important in the case of the acute porphyrias, such that as-yet unaffected family members can practice risk avoidance, particularly in terms of their exposure to potentially porphyrogenic medication. Where there is a very high prevalence as a result of a founder effect, screening for that mutation may prove useful in preliminary assessment. This is the case in South Africa, where a single *PPOX* mutation, the R59W mutation, accounts for more than 95% of all cases of VP.

Differential Diagnosis

The differential diagnosis of an acute porphyria will include any cause of severe abdominal pain, including surgical emergencies. Forty percent of

children with hereditary tyrosinemia type I may develop a syndrome closely resembling the acute attack; this is mediated by the accumulation of succinylacetone, which is a potent inhibitor of 5-aminolevulinate dehydratase.

When patients present with a motor neuropathy or quadriparesis, other causes of neuropathy may be considered. The initial diagnosis in such cases is often the Guillain-Barré syndrome. A careful history may reveal a history of abdominal pain (although very occasionally this is absent), and neurophysiologic testing will show a pattern of axonal necrosis rather than demyelination. A biochemical analysis for porphyria will confirm the correct diagnosis.

The differential diagnosis of a typical vesiculo-erosive porphyria is limited. Other chronic bullous diseases will require exclusion. Epidermolysis bullosa is similar but is not restricted to sun-exposed areas. A common differential is pseudoporphyria, which may be found in association with end-stage renal or liver disease, with tanning bed use, or as a class of drug-induced skin reactions, particularly in response to nonsteroidal anti-inflammatory drugs, nalidixic acid or tetracycline, sulfur-containing diuretics, systemic retinoids, cyclosporine, and dapsone. The skin manifestations closely resemble a cutaneous porphyria. Despite the clinical similarity, there is no underlying enzymatic or genetic defect, and plasma porphyrin profiles are normal.

It is not uncommon for patients with liver disease to excrete elevated amounts of coproporphyrin in the urine. This is termed coproporphyrinuria. It represents a slight diversion of coproporphyrin excretion from bile to urine, is not related to any disturbance in heme biosynthesis, and is essentially irrelevant.

TREATMENT

Rx

Given the rarity of the porphyrias, the evidence for efficacy of therapy seldom reaches grade A status. In most instances, treatment recommendations are based on experience in small case series or on expert opinion.

Acute Porphyrias

It is essential that the physician treating the acute attack of porphyria have a clear concept of the disorder, what treatment is appropriate, and when and how it should be applied. It is therefore important for patients to be referred to a physician with experience in porphyria or to be managed in close cooperation with such an expert.

Supportive therapy includes opioid analgesia in doses sufficient to relieve the abdominal pain. This is severe, and physicians with little experience in porphyria will frequently underdose and may even disbelieve the patient's complaints, being misled by the lack of physical signs in the abdomen, even though this is typical of the acute attack. Morphine and newer opioids should be selected in preference to meperidine, given its addictive potential. It is essential that all porphyrogenic drugs or other potential precipitants be stopped or corrected, and no medication may be administered to the patient unless its safety in porphyria has been checked.

Patients may require antiemetics. Although β-blockers may assist in slowing the pulse and reducing blood pressure, these are rarely sufficiently elevated to require treatment in their own right. Furthermore, they settle rapidly once specific therapy is administered.

Electrolyte balance requires careful monitoring. Hyponatremia may be a problem, and hypotonic fluids, such as glucose, should not be administered in large volumes. Although carbohydrate loading has been shown to have a suppressive effect on porphyrin synthesis, its effect is minimal in comparison with heme therapy.

Specific Therapy

Administration of exogenous heme results in negative feedback inhibition of ALAS, resulting in a rapid reduction in porphyrin synthesis. Such therapy is effective, has been confirmed in a controlled trial,[■] and is now the standard of care for the acute attack. The practical advantages have been three-fold: to shorten the period of symptoms for the patient; to shorten the course of the acute attack, allowing earlier discharge; and to prevent severe complications, such as encephalopathy and motor neuropathy. Although lyophilized hematin is effective, current practice is to administer heme arginate, a more stable compound. It is administered intravenously in a dose of 3 mg/kg daily for 4 days. The manufacturer recommends that the dose be reconstituted in 100 mL of a 0.9 % sodium chloride solution and infused into a large vein during at least 30 minutes. Some authorities suggest administration in human serum albumin as albumin has a buffering effect that may reduce the incidence of phlebitis at the site of infusion. There is also some evidence that administration in albumin may facilitate hepatic uptake. Typically, administration of heme arginate is followed within 24 hours by a reduction in symptoms, and after 72 hours the patient is usually symptom-free and may be discharged. Treatment during pregnancy has been shown to be safe.

It is essential that heme arginate be administered at an early stage of the acute attack. It will prevent but not reverse motor neuropathy. Once this has developed, the patient is committed to a lengthy period of hospitalization and rehabilitation.

Very rarely, patients present with a syndrome of severe, accelerated hypertension, tachycardia, and cerebral complications that may include coma, seizures, and the posterior reversible encephalopathy syndrome. This may resemble an uncontrolled pheochromocytoma. Administration of magnesium sulfate in association with combined α- and β-blockade is useful in controlling the autonomic overactivity while the attack is brought into remission with heme arginate. Recovery is usually complete.

Recurrent Attacks

Recurrent acute attacks are an uncommon manifestation of the acute porphyrias. A careful drug history is important in excluding exposure to porphyrogenic medication. Where a relationship between the attacks and the menstrual cycle is suspected, gonadal suppression with gonadotropin-releasing hormone agonists such as goserelin and buserelin may be attempted. In some patients, this is efficacious in aborting the pattern of recurrent attacks. In a small trial in 16 women, four responded with a complete cessation of symptoms, and 11 had some improvement. Because patients are at risk of osteoporosis after gonadal suppression, add-back hormonal therapy was attempted; estradiol and progesterone precipitated attacks in two and five of nine women, respectively.[9]

Some patients with a history of recurrent acute attacks have received prophylactic heme arginate at scheduled intervals, apparently with benefit in some. A permanent indwelling central venous catheter is often required for this. Currently, there is concern that frequent administration of heme arginate may induce heme oxygenase, leading to rapid catabolism of heme, thus initiating a vicious circle of reduction in hepatocyte free regulatory heme levels, ALAS1 induction, and increased porphyrin synthesis, promoting the development of another attack. Theoretically, this vicious circle might be ameliorated by the administration of heme oxygenase inhibitors, such as the metalloporphyrins like tin protoporphyrin, tin mesoporphyrin, and zinc mesoporphyrin. This has been attempted in a few patients and appeared to be of some short-term benefit, although it did not affect the overall outcome. Furthermore, exogenous hematin and heme arginate are iron rich, and patients receiving frequent courses of therapy may become iron overloaded. For these reasons, repetitive administration of heme arginate should be undertaken with extreme caution.

Orthotopic liver transplantation has now proved its value in the management of patients with recurrent acute attacks.[10] It is effectively curative in both AIP and VP, preventing further acute attacks. Transplantation should therefore be considered in any patient who develops a pattern of severe repetitive acute attacks, particularly when there is incomplete recovery between attacks, progressive disability, or severely impaired quality of life, and in patients for whom no causal factor amenable to removal or amelioration is identified.

Skin Disease

There is no specific therapy for vesiculo-erosive skin disease. Sun avoidance is central in management. This may require behavioral modification as well as careful attention to dress, wearing nontranslucent clothing to reduce the skin's ultraviolet exposure. Trauma to exposed areas should be minimized. Sunscreens may have a role but must prevent the transmission of both UVA and UVB wavelengths and preferably light at visible wavelengths as well. Zinc oxide is more effective than titanium dioxide. Although sunscreens containing micronized zinc oxide or titanium are translucent and cosmetically more acceptable, they reflect less light and therefore provide only partial protection. Nonmicronized pastes are more effective.

A number of photostable UVA filters are under development and may prove beneficial. Where skin disease is severe, as in CEP, consideration may be given to replacement of fluorescent lighting with other forms of illumination with lower short-wavelength light emission and the application of transparent film to windows, spectacles, and windshields to exclude the relevant wavelengths.

Established lesions should be carefully cleaned with nonastringent antiseptics. In our experience, aseptic lancing of bullae may hasten resolution. Where secondary infection is noted, topical or systemic antibiotics are indicated.

Porphyria Cutanea Tarda

PCT may be expected to remit once the precipitating factors have been removed. Therefore, alcohol use should be severely restricted. Hepatitis C, if present, should be treated. The reduction of hepatic iron stores is highly effective in the treatment of PCT. A common regimen is to carry out a 500-mL venesection (phlebotomy) fortnightly until iron parameters are in the low-normal range; this typically requires about 8 to 12 sessions. Intravenous and oral iron chelators are effective but are associated with more serious adverse effects than venesection. Patients with renal failure or conditions such as the myelodysplastic syndrome present a special problem as they are typically anemic. The drop in hemoglobin may be counteracted by administration of erythropoietin (erythropoiesis-stimulating agents), with the additional benefit

of its effect in mobilizing iron from the liver. However, in our experience, careful venesection without erythropoietin has proved safe and effective, the hemoglobin level returning to its usual set point after each session.

Chloroquine has been shown in a trial to be as effective as venesection in inducing remission in PCT, with an efficacy similar to that of venesection.[11] By disrupting lysosomal structure, it allows the release of porphyrins stored in the liver into the plasma, from which they are cleared by the kidneys. Thus, it is common to see a transient increase in plasma porphyrins and worsening of skin disease in the first few weeks of therapy. Chloroquine must be used in low doses, typically 125 mg twice weekly, because larger, daily doses may result in a severe transaminitis. Units with experience in treating PCT tend to have their own preference for chloroquine therapy or venesection. Our practice has been to combine phlebotomy with chloroquine, and this has proved extremely satisfactory. Remission of PCT, once attained, is usually maintained for many years, although re-treatment may occasionally be necessary. There is some evidence that α-tocopherol may provide additional benefit when it is prescribed in association with standard therapy for PCT.

Reports have suggested that direct treatment of hepatitis C and HIV with antiviral agents has resulted in an improvement in PCT even in the absence of venesection or chloroquine therapy. Given the deleterious effects of iron overload, however, it would appear prudent to combine such treatment with iron-lowering therapy.

Erythropoietic Protoporphyria

A reduction in sun exposure is central to the management of EPP and requires behavioral change, attention to dress, and use of broad-spectrum sunscreens. A small proportion of patients appear to respond positively to the administration of β-carotene in doses sufficiently large to induce carotenodermia; however, despite a number of trials of varying quality, efficacy has not been convincingly proved.[12] If it is effective at all, it is not known whether the benefit is due to the light-reflecting effect of the increased skin pigmentation, to its antioxidant effects, or to a combination. Narrow-band UVB phototherapy has in some cases resulted in increased phototolerance. Treatment with slow-release subcutaneous implants of afamelanotide, a synthetic melanocyte-stimulating hormone analogue, has yielded encouraging results in preliminary clinical studies.[13] Such analogues induce the synthesis of both melanin and eumelanin, which absorbs and reflects radiation over a wider light spectrum. In addition to inducing hyperpigmentation, afamelanotide also has antioxidant properties that may be of some benefit.

Claims have been made for a number of interventions in terms of their utility in reducing the acute pain of EPP, including lotions, steroids, local anesthetics, antihistamines, water immersion, and ice packs. Although individual patients may feel that they are helped by one or another, a study has not shown a consistent benefit for any single intervention.

The liver disease found in association with EPP in a small proportion of patients constitutes a serious problem. Such patients require the care of an experienced hepatologist, and there are no clear guidelines in treatment. Administration of oral sorbents such as activated charcoal and cholestyramine, which interrupt hepatic porphyrin recycling, have yielded inconsistent results. Hypertransfusion and administration of heme arginate may suppress porphyrin synthesis but are not suitable for long-term use. Severely affected patients are candidates for orthotopic liver transplantation. Given that the viscera are porphyrin laden and prone to severe light-induced necrosis, careful preparation of the patient is necessary, and surgery has to be carried out with appropriately filtered operating lights. Severe motor neuropathy has proved to be an unexpected but not unusual complication of liver transplantation in EPP. After transplantation, protoporphyrin will reaccumulate in the liver. Consideration may therefore be given to performing combined bone marrow and liver transplantation.

Congenital Erythropoietic Porphyria

Light avoidance, with protection for both skin and eye, is essential. Afamelanotide has shown benefit in a single case. Given the hemolysis and resultant jaundice, neonates with CEP may be subjected to phototherapy, which may seriously damage the skin. Some patients require chronic transfusion for anemia and may benefit from splenectomy. CEP is often a severe disease, and successful autologous stem cell transplantation will prevent further disfigurement and inevitable psychosocial consequences. There is emerging consensus that young subjects with severe CEP should be offered stem cell transplantation.[14] It should be borne in mind that the phenotype of CEP is extremely variable. Currently, stem cell transplantation is reserved for patients with a mutation known to be associated with a severe phenotype and for those presenting at a younger age or with progressive severe hemolytic anemia or thrombocytopenia because these factors are predictive of a poor outcome. This will require an expert decision.

Investigational Therapies for the Porphyrias

Intravenous administration of recombinant hydroxymethylbilane synthase to patients with AIP was studied in a small series. Although it reduced plasma ALA and PBG levels, it was ineffective in treating symptoms; this was presumably due to limited access to the hepatocyte. Gene replacement, with a variety of vectors, is under active development and has shown promise in laboratory studies, particularly for the management of CEP and AIP. Gene silencing by RNA interference directed at *ALAS1* mRNA to block ALAS1 induction in the acute porphyrias appears encouraging. Where the effect of a mutation is to reduce protein stability, interventions to improve stability may lead to improvement in enzyme concentration and activity. In CEP, laboratory studies suggest that administration of a proteosome inhibitor may improve UROS function where the enzyme deficiency is due to an unstable and rapidly degraded protein.[15]

Given the success of autologous stem cell transplantation in CEP and EPP and of liver transplantation in AIP and VP, experimental work on transplantation continues. A laboratory study suggesting that induced pluripotent stem cells are capable of correcting CEP was encouraging.[16] Hepatocyte transplantation has shown promise experimentally in the treatment of AIP and would have advantages over orthotopic liver transplantation.[17]

PREVENTION

Primary Prevention

Given that most of the porphyrias are genetically determined, there is currently no practical way in which these illnesses may be prevented. Diagnosis in utero by molecular methods for the identification of family-specific mutations is possible, and theoretically selective abortion would prevent transmission of the disease. In practice, however, this is not indicated. Only in a tiny minority of patients are the common acute porphyrias of sufficient severity to seriously impair the quality of the patient's life. Penetrance is incomplete, and even when the disease is clinically expressed, symptoms are not expected to develop until the end of the second decade. Termination of pregnancy is therefore inappropriate. Although patients with EPP may experience an impaired quality of life, few would argue that this is of a severity that would justify termination. The only situation in which it might be justified could be in CEP and in the homozygous acute porphyrias. However, these being recessive disorders, the incidence is extremely low, and these disorders are for practical purposes unforeseeable. The exception may be in cases of known consanguinity, in which the possibility of recessive inheritance is increased. Given that both familial and sporadic PCT tend to occur in association with disorders such as iron loading, hepatitis C, and alcoholic liver disease, treatment or avoidance of these conditions would be expected to reduce incidence.

Secondary Prevention

The most important aspect of prevention is anticipating and preventing or ameliorating the potential clinical effects and complications in the patient with known porphyria. In the acute porphyrias, interventions are directed at preventing the onset of the acute attacks. It is essential that the patient avoid exposure to any drug that might potentially be porphyrogenic. It is thus necessary that all gene carriers and the health professionals who care for them understand the absolute importance of consulting a drug safety database before taking or prescribing any medicinal agent. We recommend use of the European web-based database The Drug Database for Porphyria (*http://www.drugs-porphyria.org*) maintained by the Norwegian Porphyria Centre. Traditional drug safety lists are incomplete because the information on which they are based is typically obtained from clinical experience in porphyria and animal or tissue culture experiments and may therefore not be available for many drugs, particularly those that have come into use more recently. Second, the information derived from these sources is frequently poorly generalizable to the porphyric population at large, given the unreliability of some of the sources from which the information is derived, the extreme variability in response between individual patients, and the major differences in drug metabolism between species. The Drug Database for Porphyria, by contrast, is based on the prediction of porphyrogenicity on the grounds of metabolism and information is therefore available even before clinical experience has accumulated; preliminary analysis suggests that the predictions are highly reliable, and where a drug has been predicted to be safe, there have not as yet been any instances of clinical use resulting in an adverse effect.[18] The patient should wear a medical bracelet. Given the association of acute attacks with calorie deprivation, subjects with the acute porphyrias should avoid periods of low calorie intake.

Children with cutaneous porphyrias or those known to be a gene carrier for one of the adult-onset porphyrias with photocutaneous sensitivity, such as VP, should be encouraged to develop healthy habits of sun avoidance and

sun protection before the onset of symptoms. Patients with established disease need to modify their behaviors and attend to their dress to limit sun exposure. In cases of extreme photosensitivity, filtering of natural and artificial light may be beneficial.

Patients with PCT should be screened for hemochromatosis-associated mutations. These may be of prognostic significance in family members in detecting and therefore treating hemochromatosis early.

PROGNOSIS

With few exceptions, the prognosis of all the porphyrias is good. PCT is treatable, and remission is expected once the precipitating factors, including iron overload, have been corrected. Patients with the acute porphyrias may expect a normal lifespan, provided the appropriate precautions are taken to prevent the acute attack or, should such an attack develop, it is appropriately treated at an early stage. The skin disease of porphyria is not life-threatening.

The prognosis is poor in those patients who present with a pattern of frequent, repeated attacks and do not respond to interventions intended to break the cycle. The course is frequently one of slow deterioration during several years and may ultimately be fatal. Such patients should be assessed for orthotopic liver transplantation. Patients with CEP are subject to severe photomutilation with serious psychosocial consequences and should be assessed for allogeneic stem cell transplantation. Patients with homozygous AIP are prone to severe developmental abnormalities and may die in childhood. By contrast, there appears to be no early mortality in patients with homozygous VP, although the photomutilation and neurodevelopmental effects will have psychosocial and educational consequences.

The prognosis for the individual acute attack is excellent, provided the condition is recognized, confirmed, and appropriately treated at an early stage before neuropathy has developed. Where the patient has developed quadriparesis, the prognosis for ultimate recovery is good with excellent supportive care, including assisted ventilation. Recovery to the point of independence and nearly full power may require a year of support and rehabilitation and is not always complete. Experience has shown that it is important to avoid further acute attacks during the period of convalescence and to treat them immediately and effectively should they develop to prevent a relapse in neuropathy.

Grade A Reference

A1. Herrick AL, McColl KE, Moore MR, et al. Controlled trial of haem arginate in acute hepatic porphyria. *Lancet*. 1989;1:1295-1297.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Elder G, Harper P, Badminton M, et al. The incidence of inherited porphyrias in Europe. *J Inherit Metab Dis*. 2013;36:849-857.
2. Ryan Caballes F, Sendi H, Bonkovsky HL. Hepatitis C, porphyria cutanea tarda and liver iron: an update. *Liver*. 2012;32:880-893.
3. Balwani M, Doheny D, Bishop DF, et al. Loss-of-function ferrochelatase and gain-of-function erythroid-specific 5-aminolevulinate synthase mutations causing erythropoietic protoporphyria and X-linked protoporphyria in North American patients reveal novel mutations and a high prevalence of X-linked protoporphyria. *Mol Med*. 2013;19:26-35.
4. Hift RJ, Peters TJ, Meissner PN. A review of the clinical presentation, natural history and inheritance of variegate porphyria: its implausibility as the source of the 'Royal Malady'. *J Clin Pathol*. 2012;65:200-205.
5. Tollånes MC, Aarsand AK, Sandberg S. Excess risk of adverse pregnancy outcomes in women with porphyria: a population-based cohort study. *J Inherit Metab Dis*. 2011;34:217-223.
6. Sardh E, Wahlin S, Björnstedt M, et al. High risk of primary liver cancer in a cohort of 179 patients with acute hepatic porphyria. *J Inherit Metab Dis*. 2013;36:1063-1071.
7. Katugampola RP, Badminton MN, Finlay AY, et al. Congenital erythropoietic porphyria: a single-observer clinical study of 29 cases. *Br J Dermatol*. 2012;167:901-913.
8. To-Figueras J, Ducamp S, Clayton J, et al. ALAS2 acts as a modifier gene in patients with congenital erythropoietic porphyria. *Blood*. 2011;118:1443-1451.
9. Innala E, Backstrom T, Bixo M, et al. Evaluation of gonadotropin-releasing hormone agonist treatment for prevention of menstrual-related attacks in acute porphyria. *Acta Obstet Gynecol Scand*. 2010;89:95-100.
10. Singal AK, Parker C, Bowden C, et al. Liver transplantation in the management of porphyria. *Hepatology*. 2014;60:1082-1089.
11. Singal AK, Kormos-Hallberg C, Lee C, et al. Low-dose hydroxychloroquine is as effective as phlebotomy in treatment of patients with porphyria cutanea tarda. *Clin Gastroenterol Hepatol*. 2012;10:1402-1409.
12. Tintle S, Alikhan A, Horner ME, et al. Cutaneous porphyrias part II: treatment strategies. *Int J Dermatol*. 2014;53:3-24.
13. Minder EI. Afamelanotide, an agonistic analog of α-melanocyte-stimulating hormone, in dermal phototoxicity of erythropoietic protoporphyria. *Expert Opin Investig Drugs*. 2010;19:1591-1602.
14. Katugampola RP, Anstey AV, Finlay AY, et al. A management algorithm for congenital erythropoietic porphyria derived from a study of 29 cases. *Br J Dermatol*. 2012;167:888-900.
15. Fortian A, González E, Castaño D, et al. Intracellular rescue of the uroporphyrinogen III synthase activity in enzymes carrying the hotspot mutation C73R. *J Biol Chem*. 2011;286:13127-13133.
16. Bedel A, Taillepierre M, Guyonnet-Duperat V, et al. Metabolic correction of congenital erythropoietic porphyria with iPSCs free of reprogramming factors. *Am J Human Genet*. 2012;91:109-121.
17. Yin Z, Wahlin S, Ellis EC, et al. Hepatocyte transplantation ameliorates the metabolic abnormality in a mouse model of acute intermittent porphyria. *Cell Transplant*. 2014;23:1153-1162.
18. Hift RJ, Thunell S, Brun A. Drugs in porphyria: from observation to a modern algorithm-based system for the prediction of porphyrogenicity. *Pharmacol Ther*. 2011;132:158-169.

REVIEW QUESTIONS

1. Which is the most common form of acute hepatic porphyria?

A. Acute intermittent porphyria
B. ALA dehydratase deficiency
C. Hereditary coproporphyria
D. Porphyria cutanea tarda
E. Variegate porphyria

Answer: A All forms of acute porphyria are hepatic porphyrias, and acute intermittent porphyria is the most common everywhere except in the South African white population, in which variegate porphyria is locally prevalent as a result of a founder effect. Hereditary coproporphyria is considerably less common than both these forms of porphyria, and ALA dehydratase porphyria is excessively rare, having been described in fewer than 10 patients. Although porphyria cutanea tarda is a hepatic porphyria, it is not an acute porphyria because it presents clinically with skin disease alone and never with acute attacks.

2. The enzyme 5-aminolevulinate synthase 1 (ALAS1) is regulated by

A. Cytochrome P-450 concentration
B. Erythrocyte hemoglobin concentration
C. Free heme pool within the hepatocyte
D. Iron availability
E. Substrate availability

Answer: C ALAS1 is the hepatic isoenzyme. When the regulatory free heme pool in the hepatocyte is reduced, ALAS1 is induced, heme biosynthesis is upregulated, and heme levels rise. Conversely, when it is replete, heme synthesis stops. The erythroid isoenzyme ALAS2, by contrast, is principally regulated by iron availability. The availability of the substrates glycine and succinyl coenzyme A is not a limiting factor in heme synthesis. Hemoglobin concentrations are of no relevance to the regulation of hepatic heme synthesis.

3. What is the characteristic clinical presentation of erythropoietic protoporphyria?

A. Erosions and blisters
B. Hyperpigmentation
C. Hypertrichosis
D. Immediate photosensitivity
E. Photomutilation

Answer: D Patients with erythropoietic protoporphyria and X-linked protoporphyria complain of immediate photosensitivity, which is manifested as burning, itching, and pain in the skin after a short period of exposure to sunlight. This may be accompanied by erythema and edema. The other options on this list are not characteristic of this form of photosensitivity. Blisters and erosions are typical of the vesiculo-erosive form of cutaneous porphyria noted in variegate porphyria, porphyria cutanea tarda, hereditary coproporphyria, and congenital erythropoietic porphyria. Hypertrichosis may be noted in chronic cases, particularly in porphyria cutanea tarda, and photomutilation is encountered only in the most severe forms of vesiculo-erosive porphyria, such as congenital erythropoietic porphyria, hepatoerythropoietic porphyria, and the homozygous forms of variegate porphyria and hereditary coproporphyria.

4. Which of the following statements best characterizes the solubility and route of excretion of protoporphyrin?

A. It is of intermediate water solubility and may be identified in both urine and stool.
B. It is water insoluble and may be identified in urine.
C. It is water insoluble and may be identified in stool.
D. It is water soluble and may be identified in stool.
E. It is water soluble and may be identified in urine.

Answer: C The different porphyrins vary in their water solubility. The earlier porphyrins, such as uroporphyrin and the porphyrin precursors, are water soluble or hydrophilic and are readily detectable in urine, whereas very little emerges in the stool. Water solubility decreases with each step in the heme biosynthetic pathway. Protoporphyrin is hydrophobic and for practical purposes is undetectable in urine. It is excreted in bile and is therefore identified in stool. Some of the intermediate porphyrins, including coproporphyrin, are of intermediate solubility and detectable in both urine and stool. These properties are important because it is necessary to analyze both urine and stool (as well as plasma and, in the case of erythropoietic porphyrias, in erythrocytes) for an accurate diagnosis.

5. Which of the following treatment modalities is of most importance in modifying the course of the acute attack of porphyria?

A. Assisted ventilation
B. β-Blockade
C. Heme arginate
D. Liver transplantation
E. Meperidine

Answer: C Heme arginate is safe and highly effective in terminating the acute attack. When administered early, it will prevent serious complications such as quadriparesis and will reduce the hospital stay. Analgesia is an essential part of management but does not affect the natural history of the attack. β-Blockade may counteract the autonomic activity but is usually not necessary and will not alter the natural course. Assisted ventilation is relevant only in the patient who has had a serious attack with motor neuropathy, quadriparesis, and respiratory failure. Orthotopic liver transplantation is reserved for patients with a severe progressive course marked by recurrent acute attacks.

211

WILSON DISEASE

STEPHEN G. KALER AND MICHAEL L. SCHILSKY

DEFINITION

Wilson disease is an autosomal recessive disorder of copper transport. Affected individuals accumulate abnormal levels of copper in the liver and later in the brain as a consequence of mutations in both alleles of the Wilson disease gene (*ATP7B*). The gene encodes a copper-transporting ATPase expressed primarily in the liver, where its major function is excretion of hepatic copper into the biliary tract. The clinical condition of hepatolenticular degeneration with associated cirrhosis was first described in 1912 by S.A.K. Wilson. There are wide differences between patients in the age at onset and the spectrum of symptoms.

EPIDEMIOLOGY

The incidence of Wilson disease, defined as the occurrence of new cases, is approximately 1 in 30,000 to 40,000 live births. For special populations in which consanguinity is common, the risk of autosomal recessive traits such as Wilson disease is higher. In the general population, the prevalence of heterozygous gene carriers (defined as the ratio of all individuals with one mutant *ATP7B* allele to the population at risk of harboring one) is estimated to be 1 in 90. Some consider this figure an underestimate, however, on the basis of recent population data of the frequency of *ATP7B* mutation in the United Kingdom.[1]

PATHOBIOLOGY

Individuals normally consume 1 to 3 mg of dietary copper daily, of which approximately 50% is absorbed through the gastrointestinal tract. Most diets contain adequate amounts of copper, and certain foods (e.g., shellfish, liver, mushrooms, chocolate, nuts) contain higher quantities. In normal homeostasis, copper is absorbed from the stomach and duodenum, where absorption at the apical surface of the enterocyte is mediated by a specific copper transporter, hCTR1. The Menkes disease gene (*ATP7A*), which encodes a copper-transporting ATPase with high homology to *ATP7B*, transports copper from intestinal epithelial cells into the blood stream, where it is bound by albumin or amino acids, carried to the liver and other organs and tissues, or excreted by the kidney. This last pathway represents a minor pathway for copper excretion, and adults excrete up to 40 μg of copper per day in the urine.

Within the liver, copper may be (1) incorporated into ceruloplasmin, a multifunctional 132-kD α_2-glycoprotein enzyme containing six or seven copper atoms per molecule; (2) used in the synthesis of other copper-requiring enzymes; (3) bound by metallothionein, a low-molecular-weight, cysteine-rich protein that provides a storage and detoxification depot for copper and other trace metal elements; or (4) excreted into the bile. Copper excreted into bile does not undergo enterohepatic recirculation and thus represents a pathway for copper excretion.

In Wilson disease, metallation of ceruloplasmin is usually reduced, reflecting impaired transport of copper into the trans-Golgi compartment, where glycoprotein processing and the copper acquisition by apoceruloplasmin occurs. This results in low circulating levels of holoceruloplasmin (the protein with its full complement of copper). Because ceruloplasmin accounts for 90% of circulating copper, total serum copper is also low in most Wilson disease patients. However, free copper (non–ceruloplasmin-bound) is abnormally elevated in untreated Wilson disease patients. The flaw in copper incorporation into ceruloplasmin does not cause hepatic copper accumulation in Wilson disease, as evidenced by patients with aceruloplasminemia, in whom complete absence of this protein is associated with normal hepatic copper content. Rather, it is the effect of mutant ATP7B to reduce biliary copper excretion that produces massive hepatic copper overload when the condition is unrecognized. If the diagnosis goes unrecognized, copper overload subsequently involves other tissues, including the brain, which is particularly sensitive to perturbations in trace metal homeostasis.

The brain concentrates copper and other heavy metals for metabolic use. Copper is important for brain development and function, and copper excess (as well as deficiency) can seriously affect brain function.[2] In brain, astrocytes are considered important regulators of copper homeostasis, and in Wilson disease, the occurrence of abnormal astrocytes is a typical neuropathologic feature.

CLINICAL MANIFESTATIONS

Presenting clinical features of Wilson disease include nonspecific liver disease (Fig. 211-1), neurologic abnormalities, psychiatric illness, hemolytic anemia, renal tubular Fanconi syndrome, and various skeletal abnormalities.

There is considerable variation in clinical presentation and phenotype in Wilson disease.[3,4] Age influences the specific presentation. Most individuals who present with liver disease are younger than 30 years, sometimes in the first decade of life, whereas those presenting with neurologic or psychiatric signs range in age from the first to the eighth decade of life. This reflects the sequence of events in the pathogenesis of this disease (see earlier discussion). However, regardless of clinical presentation, some degree of liver disease is invariably present.[5] In one series of 400 adult patients with Wilson disease, approximately 50% presented with neurologic and psychiatric symptoms, 20% with neurologic and hepatic symptoms, and 20% with purely hepatic symptoms.

In patients with neurologic presentations, abnormalities include speech difficulty (dysarthria), dystonia, rigidity, tremor or choreiform movements, abnormal gait, uncoordinated handwriting, and (rarely) a combined motor

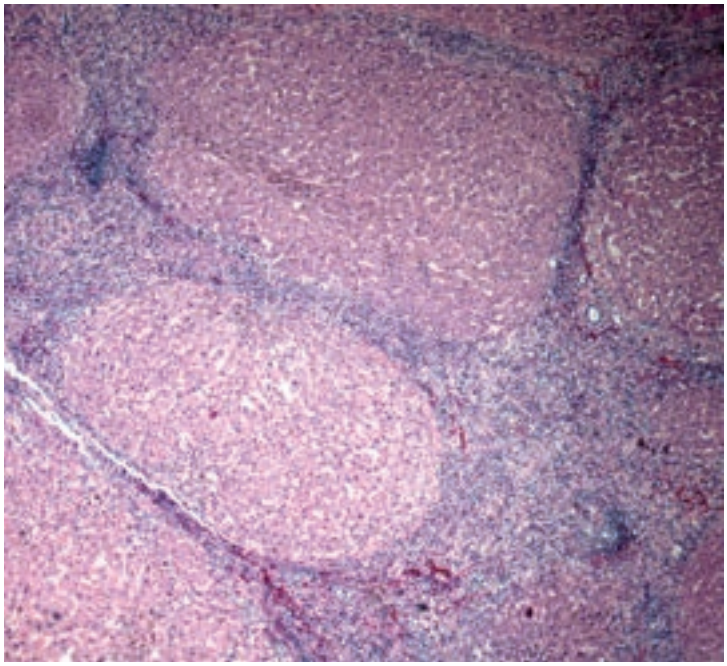

FIGURE 211-1. Hepatic cirrhosis in Wilson disease. (Image courtesy Kisha A. Mitchell, MD.)

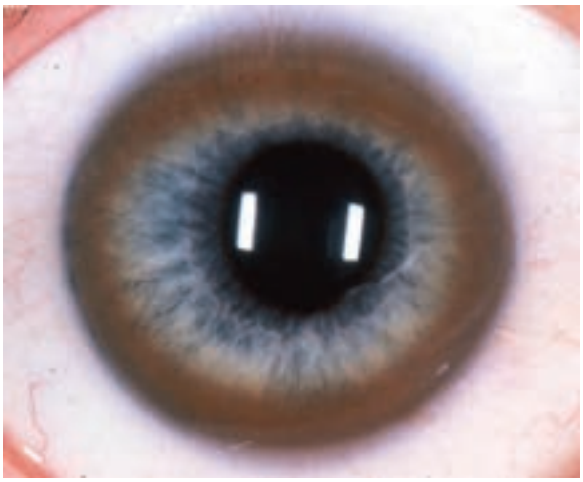

FIGURE 211-2. Kayser-Fleischer ring in a newly diagnosed patient with Wilson disease.

and sensory peripheral neuropathy.[6] Wilson disease may properly be classified as a movement disorder. The neurologic signs and symptoms reflect a predilection for involvement of the basal ganglia (e.g., caudate, putamen) in the brains of these individuals. Parkinson disease or other movement disorders may be mistakenly diagnosed.

In psychiatric presentations, changes in personality (irritability, anger, poor self-control), depression, and anxiety are common symptoms. Psychosis or bipolar disorder may also occur. Patients presenting with psychiatric symptoms are typically in their late teens or early 20s, a period during which substance abuse and schizophrenia are also prime diagnostic considerations. Wilson disease should be formally excluded in all teenagers and young adults with new-onset psychiatric signs, especially if results of liver function tests are abnormal or a family history of Wilson disease is noted.

In hepatic presentations, signs and symptoms include jaundice, hepatomegaly, edema, and ascites. Secondary endocrine effects of liver disease may include delayed puberty or amenorrhea. Viral hepatitis, autoimmune hepatitis, and cirrhosis are often initial diagnostic considerations in individuals with Wilson disease. Rare patients have Wilson disease concurrent with another liver disorder, and the diagnosis of Wilson disease is often delayed in these individuals as a consequence.

In addition to brain and liver, the eye is a primary site of copper deposition in Wilson disease, producing a benign but pathognomonic sign, the Kayser-Fleischer ring (Fig. 211-2). The Kayser-Fleischer ring is a golden to green-brown annular deposition of copper in the periphery of the cornea. This important diagnostic sign first appears as a superior crescent, then develops inferiorly and ultimately becomes circumferential. Slit-lamp examinations are required to detect rings in their early stage of formation. Copper can also accumulate in the lens and produce "sunflower" cataracts.

Approximately 95% of patients with neurologic signs manifest the Kayser-Fleischer ring, compared with approximately 50 to 65% of those with hepatic presentations. Copper chelation therapy, zinc treatment, and liver transplantation cause fading and even disappearance of corneal copper over time.

Hemolytic anemia resulting from the direct toxic effects of copper on red blood cell membranes has been observed in Wilson disease. This is usually associated with the release of massive quantities of hepatic copper into the circulation, a phenomenon that can be sudden and catastrophic due to the development of acute (fulminant) liver failure. On occasion, there may be bouts of hemolytic anemia unassociated with liver failure, but these individuals eventually develop progressive liver disease.

Renal dysfunction in Wilson disease is tubular in nature and leads to abnormal losses of amino acids, electrolytes, calcium, phosphorus, uric acid, and glucose. This effect is presumably related to direct copper toxicity or toxicity of copper complexes with metallothionein. High copper levels have been noted previously in the kidneys of patients with Wilson disease. Treatment with copper chelation often improves the renal disturbances.

There can be skeletal effects of Wilson disease, including osteoporosis and rickets; these may be attributable to renal losses of calcium and phosphorus. Osteoarthritis primarily affecting the knees and wrists also occurs in Wilson disease patients and may involve excess copper deposition in the bone and cartilage.

DIAGNOSIS

Wilson disease should be considered in patients with liver disease without a clear etiology; in patients presenting with acute liver failure with associated hemolysis; in patients with neurologic and psychiatric disease, especially if there is concomitant liver disease; and in first-degree relatives of identified patients.

Laboratory findings that support the diagnosis include low levels of serum copper and serum ceruloplasmin, elevated urine copper excretion (>100 μg/24 hours), elevated hepatic transaminase levels, low serum albumin, elevated prothrombin time (international normalized ratio), aminoaciduria, low uric acid levels, and direct antiglobulin test (Coombs)–negative hemolytic anemia. Analysis of liver biopsy specimens for histologic features and copper content also can assist diagnosis; most patients have copper contents above 250 μg/g dry weight liver, although some may have values as low as 75 μg (normal, <40 μg). Clinical signs of the disease include the stigmata of chronic liver disease, neurologic signs and symptoms, and Kayser-Fleischer rings (possibly requiring slit-lamp examination for detection). A scoring system was developed by experts who attended an international meeting on Wilson disease in Leipzig that was designed to help determine if the diagnosis should be pursued further. This scoring system was the first to use biochemical, clinical, and molecular genetic data to give a cumulative score that would suggest further evaluation is needed or achievement of the diagnosis is accomplished; it has been used in a diagnostic algorithm developed by the European Association for the Study of the Liver.[7]

Molecular diagnostics for *ATP7B* mutations has been extremely helpful, especially for difficult to diagnose cases and for family screening, in which it may be used as first-line testing if the mutations in the proband have been identified.[8] Cost and the large numbers of mutations for *ATP7B* (now more than 500) have hampered use of genetic testing for all patients being considered for Wilson disease; however, technical advances will likely increase future use of this test.

Incorporation of a stable radioisotope, ^{64}Cu, into serum ceruloplasmin is a highly specific diagnostic test; patients with Wilson disease incorporate very little ^{64}Cu into ceruloplasmin. This test is particularly useful in patients thought to have Wilson disease despite normal ceruloplasmin levels, and it distinguishes affected individuals from heterozygotes. Clinical availability of this test, however, is limited.

Increased urinary excretion of copper (>100 μg/24 hours) is another easily performed and important diagnostic test for this disorder. Copper-free collection containers should be used. A variation involving serial urine copper measurements is the penicillamine "challenge," in which 500 mg of

penicillamine is administered orally after collection of a baseline 24-hour urine specimen and repeated after 12 hours during the second 24-hour urine collection. A more than 10-fold increase in copper excretion is highly suggestive of Wilson disease.

Percutaneous needle liver biopsy for quantitative measurement of hepatic copper remains a useful test for the diagnosis of Wilson disease. As noted before, hepatic copper values higher than 250 μg/g of dry weight (normal, 20 to 50 μg/g dry weight) are characteristic of Wilson disease, although individuals with Wilson disease may have levels as low as 75 μg/g dry weight liver. Copper quantitation by inductively coupled plasma mass spectrometry or by atomic absorption spectrometry on dried and digested specimens is preferred to paraffin-embedded specimens, although paraffin-embedded specimens may be used when the diagnosis is considered retrospectively and adequate tissue was obtained. Histochemical staining of a liver biopsy specimen for copper by rhodanine may suggest Wilson disease but is less reliable.

In summary, in the absence of formal molecular evidence, the diagnosis of Wilson disease should be considered when at least two of the following are present: a positive family history, Kayser-Fleischer rings, Coombs-negative hemolytic anemia, low serum copper and ceruloplasmin levels, elevated hepatic copper content, increased 24-hour urine copper excretion, and positive penicillamine challenge result.[9,10]

TREATMENT

Penicillamine contains a free thiol that binds copper and greatly enhances urinary excretion, thereby preventing copper overload and its effects. Faithful compliance with oral penicillamine treatment has enabled the good health of thousands of patients with Wilson disease worldwide during the past 50 years.[11,12] Pyridoxine (vitamin B_6) should be prescribed concomitantly to counter the vitamin B_6 deficiency that tends to develop with long-term penicillamine administration.

Certain individuals, about 20%, are intolerant of penicillamine. Significant side effects include hypersensitivity; nephrotoxicity; hematologic abnormalities; and a distinctive rash, elastosis perforans serpiginosa, that often involves the neck and axilla. Furthermore, in some patients with neurologic presentations, penicillamine treatment induces paradoxical worsening of the neurologic disease.

Even though penicillamine is the therapy with the longest experience, other pharmaceutical agents are available and may be considered for use as first-line drugs. For example, zinc acetate and triethylene tetramine dihydrochloride (trientine) are suitable alternative agents with somewhat less significant side effect profiles.

Oral zinc acetate also has proved highly effective in Wilson disease.[13] The mechanism involves decreased copper absorption by the intestine into the blood stream by induction of the copper storage protein metallothionein in intestinal epithelial cells. Zinc monotherapy has particular value in young, presymptomatic patients; in patients who are pregnant, given the possible fetal teratogenic effects of other compounds; and as maintenance therapy for patients. Whereas most patients do well with zinc therapy, 10 to 20% of users note dyspepsia, and a higher incidence of hepatic decompensation has been observed with long-term zinc therapy compared with chelation therapy. Another drawback to zinc is the relatively long time (4 to 6 months) needed to restore proper copper balance if zinc monotherapy is used in the initial stage of treatment.

Tetrathiomolybdate forms stable tripartite complexes with proteins and copper. This drug both decreases copper absorption and reduces circulating free copper. It is fast-acting and can restore normal copper balance within several weeks, compared with the several months required with other copper chelators or with zinc. Tetrathiomolybdate is especially appropriate for the initial treatment of patients with neurologic presentations on the basis of a completed clinical trial.

Regardless of the specific regimen chosen, treatment of Wilson disease is lifelong because noncompliance eventually leads to symptomatic disease or liver failure.[14]

Liver transplantation is a rare consideration in Wilson disease because the condition is typically responsive to medical therapy. It should be considered for patients presenting with acute liver failure due to Wilson disease or those presenting with end-stage liver disease with irreversible hepatic damage who are unlikely to respond to medical therapy. Long-term outcomes after transplantation for Wilson disease are excellent, and the disease does not recur in the transplanted organ.[15]

Apart from pharmacologic treatment, there are several other important considerations in the treatment of Wilson disease. These include dietary restriction of copper-containing foods, especially shellfish and liver, both of which are copper rich. The major sources of patients' drinking water should be tested for copper concentration and avoided if levels approach 1.3 mg/L, which is the current maximum contaminant level goal (MCLG) established by the U.S. Environmental Protection Agency. The MCLG for copper in drinking water is set at a concentration at which no known or expected adverse health effects occur and for which there is an adequate margin of safety.

In newly diagnosed patients with neurologic manifestations, there is frequently a need for speech therapy and physical or occupational therapy and, for many others, psychological and genetic counseling.

Wilson Disease Heterozygotes

There is some debate about the risk of copper overload among individuals who are heterozygous carriers for Wilson disease. Even though Wilson disease is a classic autosomal recessive trait—that is, requiring two mutant alleles at the *ATP7B* locus for expression of the disease—a report from the U.S. National Academy of Sciences suggested that heterozygous carriers of Wilson disease may be a relatively sensitive population in terms of copper overload, particularly when dietary or drinking water copper exposure is higher than usual. Abnormally increased urinary copper excretion has been documented among some siblings of patients with Wilson disease, although genetic confirmation of the carrier or noncarrier status of these individuals was not available. Further patient and family studies are needed to formally address these questions.

PROGNOSIS

The prognosis in Wilson disease is generally favorable. Current therapeutic approaches can prevent, stabilize, or reverse most of the significant clinical signs and symptoms, including Kayser-Fleischer rings. However, if treatment is stopped, recurrence of symptoms and potentially fatal liver damage inevitably occurs.

FUTURE DIRECTIONS

Gene therapy for Wilson disease is a possibility. Because the Wilson copper transporter is expressed most prominently and functions most critically in the liver, this organ could be specifically targeted by the use of adenoviral or adeno-associated viral vectors (e.g., AAV8). Hepatocyte transplantation, an alternative to gene therapy, may also be applicable to the treatment of liver-specific metabolic disorders through therapeutic liver repopulation.

Grade A Reference

A1. Brewer GJ, Askari F, Lorincz MT, et al. Treatment of Wilson disease with ammonium tetrathiomolybdate: IV, comparison of tetrathiomolybdate and trientine in a double-blind study of treatment of the neurologic presentation of Wilson disease. *Arch Neurol*. 2006;63:521-527.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Coffey AJ, Durkie M, Hague S, et al. A genetic study of Wilson disease in the United Kingdom. *Brain*. 2013;136(Pt 5):1476-1487.
2. Scheiber IF, Mercer JF, Dringen R. Metabolism and functions of copper in brain. *Prog Neurobiol*. 2014;116:33-57.
3. Lutsenko S. Modifiying factors and phenotypic diversity in Wilson's disease. *Ann N Y Acad Sci*. 2014;1315:56-63.
4. Ferenci P. Phenotype-genotype correlations in patients with Wilson's diease. *Ann N Y Acad Sci*. 2014;1315:1-5.
5. Shah D. Wilson's disease: hepatic manifestations. *Dis Mon*. 2014;60:465-474.
6. Dalvi A. Wilson's disease: neurological and psychiatric manifestations. *Dis Mon*. 2014;60:460-464.
7. 2012 European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Wilson disease. *J Hepatol*. 2012;56:671-685.
8. Schilsky ML, Ala A. Genetic testing for Wilson disease: availability and utility. *Curr Gastroenterol Rep*. 2010;12:57-61.
9. Dalvi A, Padmanaban M. Wilson's disease: etiology, diagnosis, and treatment. *Dis Mon*. 2014;60:450-459.
10. Kanwar P, Kowdley KV. Metal storage disorders: Wilson disease and hemochromatosis. *Med Clin North Am*. 2014;98:87-102.
11. Lowette KF, Desmet K, Witters P, et al. Wilson disease: long-term follow-up of a cohort of 24 patients treated with D-penicillamine. *Eur J Gastroenterol Hepatol*. 2010;22:564-571.
12. Weiss KH, Gotthardt DN, Klemm D, et al. Zinc monotherapy is not as effective as chelating agents in treatment of Wilson disease. *Gastroenterology*. 2011;140:1189-1198.
13. Czlonkowska A, Litwin T, Karlinski M, et al. d-Penicillamine versus zinc sulfate as first-line therapy for Wilson's disease. *Eur J Neurol*. 2014;21:599-606.
14. Harada M. Pathogenesis and management of Wilson's disease. *Hepatol Res*. 2014;44:395-402.
15. Guillaud O, Dumortier J, Sobesky R, et al. Long term results of liver transplantation for Wilson's disease: experience in France. *J Hepatol*. 2014;60:579-589.

REVIEW QUESTIONS

1. The diagnosis of Wilson disease can be excluded in a patient

A. If the individual is younger than 21 years
B. If a Kayser-Fleischer ring is detected on slit-lamp examination
C. If the individual's penmanship is neat
D. If results of repeated 24-hour urine copper collections are normal
E. If serum ceruloplasmin and serum copper levels are normal

Answer: D Wilson disease may be manifested at any age, although appearance of symptoms before the age of 5 years is rare. The Kayser-Fleischer ring representing copper accumulation in the periphery of the cornea is a diagnostic hallmark of Wilson disease. Dysgraphia is sometimes reported by affected patients, but if it is not present, it should not exclude the diagnosis. The 24-hour urine copper excretion is reliably elevated in Wilson disease, and normal results exclude the diagnosis.

2. All except which of the following are true concerning the diagnosis of Wilson disease?

A. Schizophrenia in a 20-year-old can be a presenting manifestation.
B. Hepatic manifestations usually precede neurologic symptoms.
C. Serum copper and serum ceruloplasmin levels are invariably low.
D. *ATP7B* mutation analysis is useful for detection of presymptomatic affected individuals.
E. Neurologic symptoms often occur later than hepatic signs.

Answer: C An estimated 10% of individuals with Wilson disease have normal serum copper and ceruloplasmin levels, for reasons that remain unclear.

3. ATP7B is a copper-transporting molecule that

A. Mediates copper transport to the brain
B. Enables biliary excretion of copper
C. Is mutated in Menkes disease
D. Has its gene located on the X chromosome
E. Is responsible for intestinal copper absorption

Answer: B The Wilson disease gene product, ATP7B, normally mediates excretion of copper from the liver into the bile. The gene is located on chromosome 13. *ATP7A* is the gene mutated in X-linked recessive Menkes disease.

4. Low serum copper and ceruloplasmin levels may be found in all except which of the following?

A. ATP7A-related distal motor neuropathy
B. Wilson disease
C. Aceruloplasminemia
D. Menkes disease
E. Post–bariatric surgery

Answer: A Serum copper and ceruloplasmin levels are normal in patients with ATP7A-related isolated distal motor neuropathy. These can be low in all the other conditions listed.

5. Neurologic presentations of Wilson disease

A. Never coincide with hepatic symptoms
B. Do not include dysarthria, dystonia, rigidity, tremor, or choreiform movements
C. Usually involve stroke events affecting the basal ganglia
D. Never feature a sensory neuropathy
E. Are nearly always (95% of the time) associated with presence of Kayser-Fleischer rings

Answer: E Kayser-Fleischer rings are typically associated with neurologic presentation of Wilson disease. Basal ganglia copper deposition is a pathologic feature in Wilson disease; however, stroke is not.

212

IRON OVERLOAD (HEMOCHROMATOSIS)

BRUCE R. BACON

DEFINITION AND EPIDEMIOLOGY

Hereditary hemochromatosis (HH) is a common inherited disorder of iron metabolism. The genetic abnormality responsible in most patients with typical HH is found in homozygous form in about 1 in 250 persons of northern European descent. It is characterized by an increase in iron absorption from the upper gastrointestinal tract, with subsequent tissue iron deposition in parenchymal cells of the liver, heart, pancreas, joints, and endocrine organs. The autosomal recessive inheritance pattern of HH was clearly shown in the 1970s, and the gene responsible for most cases of HH was identified in 1996 by investigators using a positional cloning technique.[1] The gene is called *HFE*

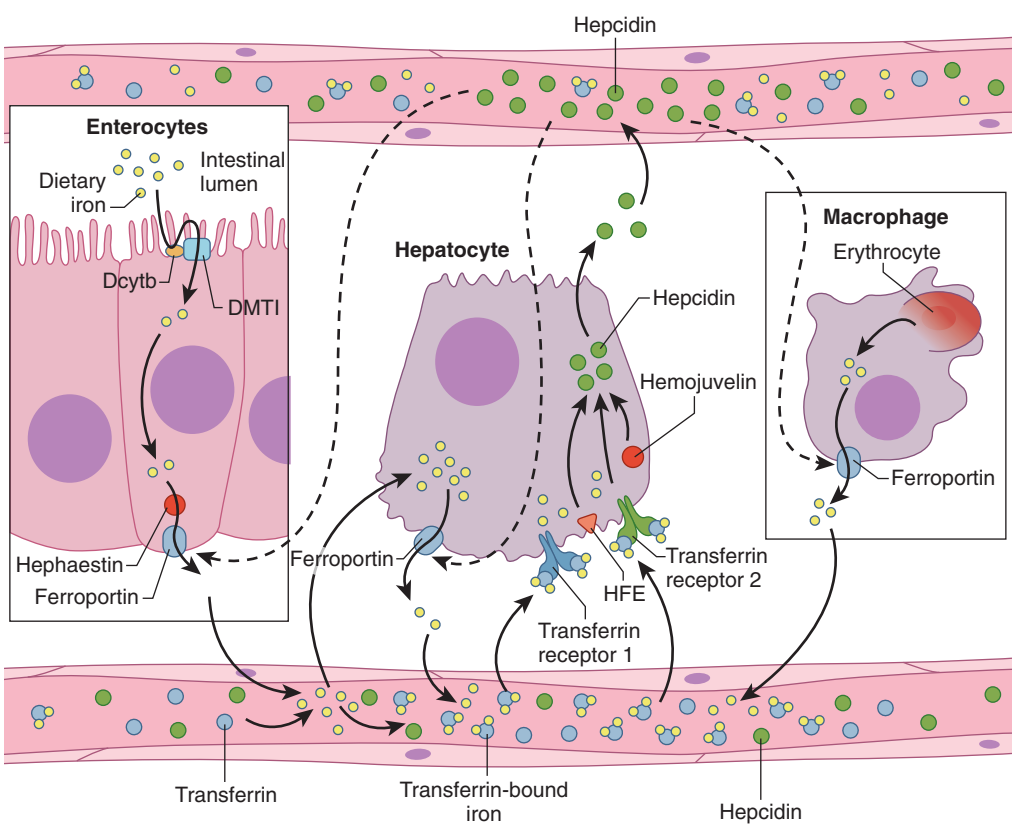

FIGURE 212-1. Orchestration of iron homeostasis. In the duodenal enterocyte, dietary iron is reduced to the ferrous state by duodenal ferric reductase (Dcytb), transported into the cell by divalent metal transporter 1 (DMT1), and released by way of ferroportin into the circulation. Hephaestin facilitates enterocyte iron release. Hepatocytes take up iron from the circulation either as free iron or as transferrin-bound iron (through transferrin receptors 1 and 2). Transferrin receptor 2 may serve as a sensor of circulating transferrin-bound iron, thereby influencing expression of the iron regulatory hormone hepcidin. The hepcidin response is also modulated by HFE and hemojuvelin. Hepcidin is secreted into the circulation, where it downregulates the ferroportin-mediated release of iron from enterocytes, macrophages, and hepatocytes (*dashed lines*). (From Fleming RE, Bacon BR. Orchestration of iron homeostasis. *N Engl J Med.* 2005;352:1741-1744.)

and encodes a novel major histocompatibility complex (MHC) class I–like molecule that binds with transferrin receptor (TfR) and affects hepcidin homeostasis (Fig. 212-1). Prospective population studies have demonstrated that only about 50 to 60% of patients who are homozygous for the major mutation found in *HFE* (called C282Y) have evidence of phenotypic expression of iron overload, and only a small percentage (<10%) go on to develop tissue damage from excess iron deposition. These findings of the highly variable penetrance of C282Y homozygosity have changed modern thinking about HH and must be considered when patients are evaluated for this disease in a physician's office and when national health policy for screening of this genetic disorder is developed.

The discovery of *HFE* has had a tremendous impact in a number of areas. The ability to accurately diagnose disorders of iron overload has been strengthened, family screening is improved, and the evaluation of patients with other forms of liver disease complicated by moderate to severe iron overload is possible. Furthermore, with the discovery of *HFE*, a considerable new body of knowledge about the mechanisms and regulation of iron absorption has been identified, both in the normal situation and in the pathologic condition seen when *HFE* mutations are present.

PATHOBIOLOGY

Classification of Iron Overload Syndromes

The term *hereditary hemochromatosis* should be reserved for inherited disorders of iron metabolism (see Fig. 212-1) that lead to tissue iron loading (Table 212-1). The most common form of this disease, *HFE*-related HH, is caused primarily by homozygosity for the C282Y mutation in the *HFE* gene. Other heritable forms of iron overload have also been recognized (non–*HFE*-related HH).[2,3] These include autosomal recessive forms of HH characterized by rapid iron accumulation and caused by mutations in the genes for hemojuvelin (*HJV*) and hepcidin (*HAMP*) (also called juvenile hemochromatosis); an autosomal dominant form of HH caused by mutations in the ferroportin gene; an autosomal recessive form of HH resulting from mutations in the gene for *TFR2* (*SLC40A1*); and rare forms of HH resulting from mutations in the *DMT1* gene or mutations in the portion of the ferritin gene encoding the iron-responsive regulatory element. Some other types of iron overload may have a familial or inherited component, but the genes involved have not yet been identified. For example, African iron overload is a familial disorder of iron loading prevalent in sub-Saharan Africa that is exacerbated by the ingestion of an iron-rich home-brewed beer. However, iron overload can also occur in individuals who do not drink this beverage. A similar form of iron overload has been suggested in African Americans, and further study is necessary to clarify this condition. The degree of iron loading can be similar to that seen in *HFE*-related HH, but the cellular and lobular distribution of iron is different. In addition, a rare disorder termed congenital alloimmune hepatitis is responsible for most cases of neonatal iron overload and is characterized by a modest increase in hepatic iron accompanied by severe liver injury present at birth.

It has been recognized during the last several years that many patients who have *HFE*-linked hemochromatosis have no evidence of iron overload. With this in mind, four stages of HH have been described[4]:

1. Genetic predisposition with no phenotypic abnormality
2. Iron overload (approximately 2 to 5 g total body iron) without symptoms
3. Iron overload with mild or early symptoms
4. Iron overload with organ damage, such as cirrhosis

The ability to establish a genetic diagnosis has led to a much greater understanding of genotype-phenotype correlations.

Genetics and Pathophysiology of Hemochromatosis

Since the classic linkage studies of Simon and colleagues in the mid-1970s, it has been known that the gene for hemochromatosis is located in the human leukocyte antigen (HLA) region on chromosome 6. In 1996, a team of molecular geneticists using a positional cloning technique identified a candidate gene for HH, which is now called *HFE*. *HFE* codes for a novel MHC class I–like molecule, which, like all MHC proteins, requires interaction with β_2-microglobulin for normal presentation on the cell surface. Three principal missense mutations have been identified in *HFE:* one results in a change of cysteine at position 282 to tyrosine (Cys282 → Tyr, C282Y), the second

TABLE 212-1 CLASSIFICATION OF IRON OVERLOAD SYNDROMES

HEREDITARY HEMOCHROMATOSIS
***HFE* Related**
C282Y/C282Y C282Y/H63D C282Y/S65C Other mutations
Non–*HFE* Related
Hemojuvelin *(HJV)* mutations (autosomal recessive) Hepcidin *(HAMP)* mutations (autosomal recessive) Ferroportin *(SLC40A1)* mutations (autosomal dominant) Transferrin receptor 2 *(TFR2)* mutations (autosomal recessive) Divalent metal transporter 1 *(SLC11A2)* mutations (rare) Ferritin regulatory mutations (rare)
Miscellaneous
African iron overload Neonatal iron overload (rare)
SECONDARY IRON OVERLOAD
Anemia Caused by Ineffective Erythropoiesis
Thalassemia major Sideroblastic anemias Congenital dyserythropoietic anemias Congenital atransferrinemia
Liver Disease
Alcoholic liver disease Chronic viral hepatitis B and C Porphyria cutanea tarda Nonalcoholic steatohepatitis After portacaval shunt
Miscellaneous
Transfusional iron overload Excessive parenteral iron administration

results in a change of histidine at position 63 to aspartate (His63 → Asp, H63D), and the third results in a change of serine at position 65 to cysteine (Ser65 → Cys, S65C). A few other mutations have been identified in *HFE*, but their frequency is low and their clinical impact is limited. In the original studies, 83% of typical phenotypic HH patients were found to be homozygous for the C282Y mutation. Several other studies from around the world in predominantly white populations demonstrated that among patients with typical hemochromatosis, about 85 to 90% were homozygous for C282Y. Thus, about 10 to 15% of patients with typical phenotypic HH have some reason other than C282Y homozygosity for their iron overload.

Nearly all absorption of dietary iron occurs in the duodenum, where iron may be taken up either as ionic iron or as heme. Ionic iron requires reduction to the ferrous state, which is accomplished by the ferric reductases (e.g., duodenal ferric reductase), which are expressed on the luminal surface of duodenal enterocytes (see Fig. 212-1). This ferrous iron crosses the apical membrane through divalent metal transporter 1, and iron taken up by the enterocyte is either stored as ferritin or transferred across the basolateral membrane to the plasma. This latter process occurs by the iron transporter ferroportin and requires oxidation of iron to the ferric state by the ferroxidase hephaestin.[5] Hypoxia-inducible factor-2 (HIF-2α) regulates the expression of key genes involved in iron absorption and may be involved in the hyperabsorption of iron in the context of hepcidin deficiency in HH.[6]

Hepcidin is a 25–amino acid, liver-derived peptide that influences systemic iron status such that it is now considered to be the principal iron regulatory hormone (see Fig. 212-1). Dysregulation of hepcidin expression is thought to play a role in the pathogenesis of HH. Patients with *HFE*-related HH have low hepatic expression of hepcidin, as do *HFE* knockout mice, despite excess hepatic iron stores. Conversely, overexpression of hepcidin in *HFE* knockout mice prevents the HH phenotype. Iron-induced regulation of hepcidin expression involves a bone morphogenetic protein (BMP)–dependent signaling pathway. BMPs bind to specific receptors on hepatocytes, thereby triggering SMAD protein–dependent activation of hepcidin expression. Selective inhibition of BMP signaling abrogates iron-induced upregulation of hepcidin. Hemojuvelin is a BMP coreceptor and facilitates the binding of BMP to its receptor; knockout of the hemojuvelin gene markedly decreases BMP signaling and hepcidin expression and causes iron overload. It has been hypothesized that *TFR2* in hepatocytes may act as an iron sensor. Mutations in *TFR2* cause a rare form of HH in humans, and *Tfr2* mutant mice have an HH phenotype. In HH, excess iron (both transferrin bound and non–transferrin bound) is avidly taken up by hepatocytes and stored. Iron stores increase to the point at which iron-induced oxidative damage occurs, resulting in cell injury and cell necrosis with phagocytosis by Kupffer cells. Iron-laden Kupffer cells become activated and produce profibrogenic cytokines (transforming growth factor-β, platelet-derived growth factor), which stimulate hepatic stellate cells to synthesize excess collagen and other matrix proteins. Increased fibrosis and then cirrhosis result.

CLINICAL MANIFESTATIONS

Several symptoms and clinical findings have been identified in patients with fully established HH, and all physicians should be aware of these symptoms and findings, which are summarized in Tables 212-2 and 212-3. Table 212-4

TABLE 212-2 SYMPTOMS IN PATIENTS WITH HEREDITARY HEMOCHROMATOSIS

ASYMPTOMATIC
Abnormalities of serum iron studies on routine screening chemistry panel Abnormal liver test results Identified by family screening Identified by population screening
NONSPECIFIC SYSTEMIC SYMPTOMS
Weakness Fatigue Lethargy Apathy Weight loss
SPECIFIC ORGAN-RELATED SYMPTOMS
Abdominal pain (hepatomegaly) Arthralgias (arthritis) Symptoms of diabetes mellitus (pancreas) Amenorrhea (cirrhosis) Loss of libido, impotence (pituitary, cirrhosis) Congestive heart failure symptoms (heart) Arrhythmias (heart)

TABLE 212-3 PHYSICAL FINDINGS IN PATIENTS WITH HEREDITARY HEMOCHROMATOSIS

ASYMPTOMATIC
No physical findings Hepatomegaly
SYMPTOMATIC
Liver
Hepatomegaly Cutaneous stigmata of chronic liver disease Splenomegaly Signs of liver failure: ascites, encephalopathy
Joints
Arthritis Joint swelling
Heart
Dilated cardiomyopathy Congestive heart failure
Skin
Increased pigmentation
Endocrine
Testicular atrophy Hypogonadism Hypothyroidism

TABLE 212-4 LABORATORY FINDINGS IN PATIENTS WITH HEREDITARY HEMOCHROMATOSIS

MEASUREMENTS	NORMAL SUBJECTS	PATIENTS WITH HEREDITARY HEMOCHROMATOSIS	
		Asymptomatic	*Symptomatic*
BLOOD (FASTING)			
Serum iron level (μg/dL)	60-180	150-280	180-300
Serum transferrin level (mg/dL)	220-410	200-280	200-300
Transferrin saturation (%)	20-45	45-100	80-100
Serum ferritin level (ng/mL)			
Men	20-200	150-1000	500-6000
Women	15-150	120-1000	500-6000
GENETIC (*HFE* MUTATION ANALYSIS)			
C282Y/C282Y	wt/wt‡	C282Y/C282Y	C282Y/C282Y
C282Y/H63D*	wt/wt	C282Y/H63D	C282Y/H63D
LIVER			
Hepatic iron concentration			
μg/g dry weight	300-1500	2000-10,000	8000-30,000
μmol/g dry weight	5-27	36-179	140-550
Hepatic iron index†	<1	1 to >1.9	>1.9
Liver histology			
Perls' Prussian blue stain	0, 1+	2+ to 4+	3+, 4+

*Compound heterozygote.
†Calculated by dividing the hepatic iron concentration (in μmol/g dry weight) by the age of the patient (in years). With the increased use of genetic testing in patients with iron overload, the specificity of the hepatic iron index has diminished.
‡wt/wt: wild type (normal).

summarizes the typical laboratory findings in symptomatic and asymptomatic patients with HH. Recent series have revealed that many asymptomatic patients who are C282Y homozygotes are now coming to medical attention because they are identified by family screening studies or population surveys or after abnormalities of iron studies are discovered on routine blood chemistry testing. It is ideal to identify patients who have some phenotypic expression with abnormal results of iron studies but no evidence of organ damage. Several large population screening studies have shown evidence of phenotypic expression with abnormal findings of iron studies in about 40 to 50% of C282Y homozygotes, but less than 10% of these individuals actually have signs and symptoms of the disease.

DIAGNOSIS

Because patients with genetic abnormalities can be identified before there is evidence of phenotypic expression, the method of diagnosing HH has undergone a change. The role of liver biopsy has lessened considerably with the advent of genetic testing. Nonetheless, some general principles should be acknowledged. If the diagnosis of HH is being considered, blood tests, including transferrin saturation (serum iron ÷ transferrin or total iron-binding capacity × 100%) and ferritin levels, should be obtained. Transferrin saturation does not need to be measured in the fasting state for reliable results to be obtained. In patients with symptoms (see Table 212-2), both these values are elevated; however, transferrin saturation is typically the earliest phenotypic marker of HH and may be elevated in young C282Y homozygotes with normal ferritin levels. Serum ferritin is sometimes elevated in other conditions in which there is no evidence of iron overload.[7] Confounding causes of high serum ferritin include alcohol consumption, metabolic syndrome, liver damage (acute or chronic), and more unusual disorders, such as Gaucher's disease and macrophage activation syndrome (hemophagocytic lymphohistiocytosis). Overall, about 90% of patients with hyperferritinemia do not have iron overload, which often remains unexplained. Thus, ferritin is relatively sensitive but not specific for iron overload.[8]

Magnetic resonance imaging is a fast and efficient noninvasive technique to assess and to monitor liver iron concentration. It is based on the accumulation of iron leading to signal loss in the liver, particularly with T2*-weighted sequences. Magnetic resonance sequences do lose accuracy, however, when the hepatic iron concentration is very high.

In the past, if an elevated transferrin saturation or ferritin level was identified, a liver biopsy would be performed to establish a diagnosis by histochemical iron stains and biochemical determination of the hepatic iron concentration with calculation of the hepatic iron index (HII). The HII is the patient's hepatic iron concentration (in μmol/g dry weight) divided by the patient's age in years. Previously, when the HII was higher than 1.9, the diagnosis of HH was established. Recent studies with genetic testing have shown that many (>50%) HH patients may have an HII of less than 1.9. Thus, the HII is no longer important in the diagnosis of HH.

Currently, when abnormalities of iron studies are identified, it is reasonable to proceed to genetic testing. Among individuals who are C282Y homozygotes or compound heterozygotes (C282Y/H63D), liver biopsy is reserved for those with elevated liver enzymes or ferritin levels above 1000 ng/mL (Fig. 212-2). Several studies have shown that advanced fibrosis or cirrhosis is not seen in HH patients when ferritin levels are below 1000 ng/mL or when liver enzymes are normal. Accordingly, as genetic testing has become more widely available, liver biopsy is less necessary.

When liver biopsy is performed, iron deposition is found preferentially in a periportal (acinar zone 1) region of the hepatic lobule, with a decrease in gradient in acinar zones 2 and 3. With significant iron loading, sinusoidal lining cell (Kupffer cell) iron deposition can be identified, and iron can be found in bile duct cells and in fibrous tissue in portal tracts or septa. In patients with secondary iron overload related to alcoholic liver disease or chronic viral hepatitis, iron deposition is typically in Kupffer cells as well as in hepatocytes, and it occurs in a panlobular (as opposed to a periportal) distribution. Histologic evaluation of iron-staining patterns provides information complementary to that obtained by traditional biochemical testing for iron overload along with genetic testing.

TREATMENT

Even though there have been advances in the molecular and cellular biologic understanding of HH, and although the impact of *HFE* mutation analysis on diagnosis has been significant, the treatment of HH remains simple, inexpensive, and safe. Patients should have therapeutic phlebotomy of 500 mL of whole blood (approximately 200 to 250 mg of iron, depending on the hemoglobin concentration) on a weekly basis, if tolerated. Therapeutic phlebotomy should be performed until iron-limited erythropoiesis develops, identified by failure of the hemoglobin level and hematocrit to recover before the next phlebotomy.[9] It is reasonable to monitor transferrin saturation and ferritin levels periodically (every 3 months) to predict the return of iron stores to normal and to provide encouragement to patients who are undergoing phlebotomy. Therapeutic phlebotomy should be continued until the ferritin level is into the normal range of 50 to 100 ng/mL. Transferrin saturation may still be elevated (> 50%) but should not dictate further phlebotomy.[10] It is not necessary for patients to become anemic or iron deficient—just depleted of excess iron stores. Some patients may require weekly phlebotomy for 1 year or longer. Others may be able to tolerate phlebotomy of only a half-unit of blood (250 mL) every other week. Once the initial *therapeutic* phlebotomy has been completed, most patients require *maintenance* phlebotomy, with 1 unit of blood removed every 2 to 3 months. This requirement is derived empirically, with the intent being to maintain a ferritin level between 50 and 100 ng/mL. Recognizing that HH subjects could constitute a safe source of blood for transfusion, the U.S. Food and Drug Administration now allows blood phlebotomized from HH donors to be used for transfusion as long as the phlebotomies are performed in authorized blood centers that comply with specific provisions for safeguard.[11]

With successful iron depletion, patients have an improved sense of well-being, right upper quadrant abdominal pain dissipates, liver test results improve, and diabetes may be easier to manage. Established cirrhosis, arthropathy, and testicular atrophy generally do not improve (Table 212-5).

Iron chelation therapy may be necessary for patients who are anemic and cannot tolerate phlebotomy. Parenteral infusions (administered either subcutaneously by an infusion pump or intravenously overnight through a port) of deferoxamine can be used. Ototoxicity and ocular toxicity are possible with deferoxamine, and appropriate monitoring should be performed. An orally administered iron chelator, deferasirox (Exjade), has been approved and is available for treatment of iron overload in patients with iron-loading anemias and HH.[12] This therapy is expensive and has potential for toxicity, but it is more convenient than phlebotomy. Minihepcidins that mimic hepcidin activity are being tested in laboratory animals as a possible future treatment of iron overload.[13]

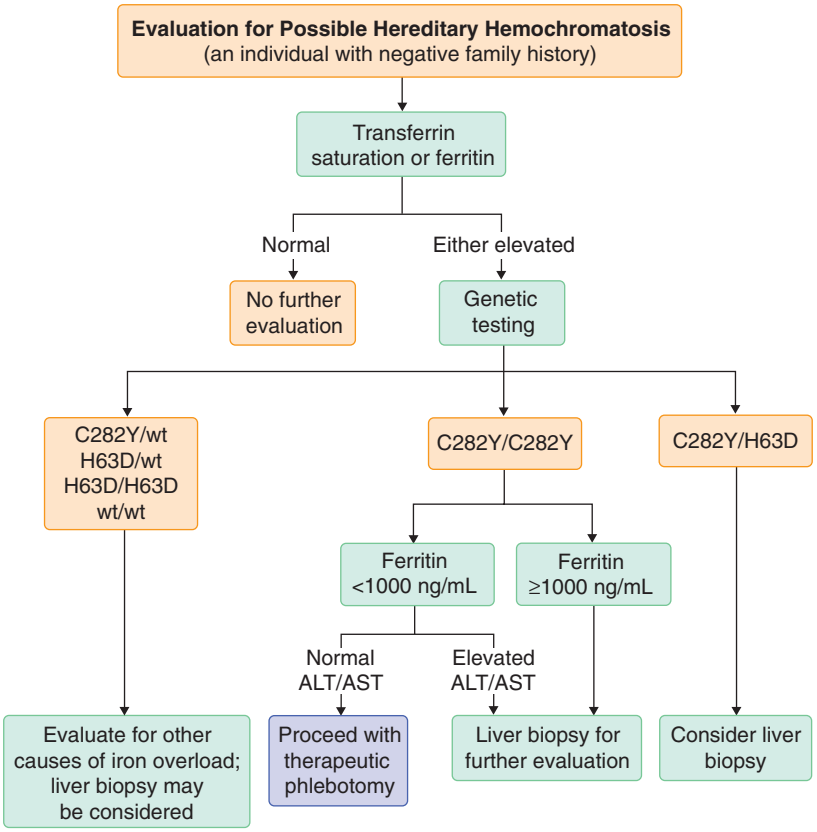

FIGURE 212-2. **Algorithm for evaluation of possible hereditary hemochromatosis in a person with a negative family history.** ALT = alanine transaminase; AST = aspartate transaminase.

TABLE 212-5 RESPONSE TO PHLEBOTOMY TREATMENT IN PATIENTS WITH HEREDITARY HEMOCHROMATOSIS

Reduction of tissue iron stores to normal
Improvement in survival if diagnosis is made and treatment is instituted before the development of cirrhosis and diabetes
Improvement in sense of well-being and energy level
Improvement in cardiac function
Improvement in controlling diabetes
Reduction in abdominal pain
Reduction in skin pigmentation
Normalization of elevated liver enzymes
Reversal of hepatic fibrosis
No reversal of established cirrhosis
Eliminates the risk of HH-related HCC if iron removal happens before the development of cirrhosis
Reduction in portal hypertension in patients with cirrhosis
No improvement in arthropathy
No reversal of testicular atrophy

HCC = hepatocellular cancer; HH = hereditary hemochromatosis.

PREVENTION

Family Screening

Once an HH proband is recognized, all first-degree relatives should be offered testing. In the past, HLA haplotyping was performed, but now *HFE* mutation analysis is recommended, along with the determination of transferrin saturation and ferritin levels. In probands with children, *HFE* mutation analysis is performed in the spouse to accurately predict the genotype in the child. If the spouse has either mutation, testing of the child is necessary, although the value and availability of genetic testing in children are debated. If an adult relative of a C282Y homozygote is identified and is either a C282Y homozygote or a compound heterozygote, and if results of blood iron studies are abnormal, a presumptive diagnosis can be made, and therapeutic phlebotomy can be initiated with the guidelines already discussed.

Population Screening

Because HH is a common disorder with a well-described treatment and a long latent period (i.e., time before disease occurs), some have suggested that it is an ideal candidate for population screening by genetic testing. However, studies have shown a less than expected phenotypic expression and a decreased number of patients with clinical manifestations of iron-mediated disease, raising questions about this recommendation. Initial results from a large National Institutes of Health–sponsored screening study in North America demonstrated a prevalence of C282Y homozygosity of 1 in 227 (Table 212-6). The C282Y homozygotes had higher ferritin levels than the general population, but 25% had normal ferritin levels. In screening for iron overload as opposed to screening for *HFE*-linked HH, transferrin saturation should be measured. In this situation, when abnormalities of iron studies are identified and the patient does not have a mutation in *HFE*, liver biopsy should be considered to clarify the situation relative to iron stores.

PROGNOSIS

If patients are identified, diagnosed, and treated[14] before the development of cirrhosis, their life expectancy is the same as that for an age- and sex-matched control population. If patients are not identified and treated before the development of cirrhosis, they are at risk for premature death from complications of diabetes, chronic liver disease, or hepatocellular cancer.[15] Treated cirrhotic patients are still at risk for hepatocellular cancer and should undergo surveillance abdominal imaging every 6 to 12 months.

Mutation Analysis in Patients with Liver Disease

Many patients with liver disease have abnormalities in serum parameters of iron metabolism. These abnormalities are more commonly seen in patients with hepatocellular liver diseases than in those with cholestatic liver diseases. Several clinical studies have shown that approximately 50% of patients with alcoholic liver disease, chronic viral hepatitis C, and nonalcoholic steatohepatitis (NASH) have abnormalities of serum iron studies. This abnormality is usually an elevation in serum ferritin, but elevated transferrin saturation is occasionally seen as well. When liver biopsy is performed, increased iron deposits can be seen, usually in a panlobular distribution, with iron in both hepatocytes and sinusoidal lining cells (Kupffer cells)

TABLE 212-6 PREVALENCE OF C282Y HOMOZYGOTES WITHOUT IRON OVERLOAD IN SCREENING STUDIES

POPULATION SAMPLE	COUNTRY	*n*	PREVALENCE OF HOMOZYGOTES	C282Y HOMOZYGOTES WITH A NORMAL FERRITIN LEVEL (%)
Electoral roll	New Zealand	1064	1 in 213	40
Primary care	United States	1653	1 in 276	50
Epidemiologic survey	Australia	3011	1 in 188	25
Blood donors	Canada	4211	1 in 327	81
General public	United States	41,038	1 in 270	33
Primary care	North America	44,082	1 in 227	25
General public	Australia	29,676	1 in 146	32
Total		124,636	1 in 240	41

(E-Figs. 212-1 to 212-5). Hepatic iron concentrations may be slightly increased or normal. When *HFE* mutations have been evaluated in patients with alcoholic liver disease, there has been no increased incidence of either C282Y or H63D (either heterozygote or homozygote) compared with control populations. Furthermore, there was no increase in *HFE* mutations in patients with alcoholic liver disease who had an increased amount of fibrosis. Thus, the abnormal iron studies frequently seen in patients with alcoholic liver disease are most likely due to mechanisms other than mutations in *HFE*.

In chronic hepatitis C, the relationship of abnormalities of iron studies and elevated hepatic iron concentration with a response to interferon monotherapy has been known for several years. Numerous studies have shown that patients who fail to respond to interferon monotherapy have a higher hepatic iron concentration than those who do respond. A corollary to this observation involves therapeutic phlebotomy to deplete iron stores in the hope of improving response to therapy. Reduction in iron stores by therapeutic phlebotomy does reduce elevated liver enzymes and has had some marginal beneficial effect on liver histologic features, but it does not have any virologic effects. When *HFE* mutation analysis has been investigated in patients with chronic hepatitis C, the frequency of C282Y and H63D has been equivalent to that in control populations. Most studies have shown that when *HFE* mutations are present, they correlate with increased iron stores seen histologically. Some studies have shown a synergistic effect with the development of fibrosis. At present, it is recommended that *HFE* mutation analysis be done when abnormalities of iron studies are seen in patients with chronic hepatitis C. Also, iron stains are typically performed on liver biopsy samples when biopsies are done to grade and to stage chronic hepatitis C. If iron stores are increased, it is reasonable to perform therapeutic phlebotomy to deplete excess iron stores before antiviral therapy is initiated.

In patients with NASH, several studies have provided conflicting results.[16] Some have shown an increase in *HFE* mutations in patients with NASH, and others have shown no difference from control populations. When there has been an increased prevalence of *HFE* mutations in NASH, there has been good correlation between abnormal serum parameters of iron and a correlation with hepatic iron concentration. Furthermore, some studies have shown an increase in fibrosis in NASH patients with *HFE* mutations. Finally, one study showed a reduction in elevated liver enzymes and improvement in parameters of insulin resistance in patients with NASH treated by phlebotomy to produce near iron deficiency. These observations suggest an interaction between the expression of fatty liver disease and iron metabolism.

Finally, in porphyria cutanea tarda (PCT), the relationship between abnormalities of iron metabolism and the role of therapeutic phlebotomy has been known for many years.[17] Also, it has recently been shown that as many as 70% of patients with PCT are infected with hepatitis C virus, and many patients with PCT drink excessive amounts of alcohol. An increased prevalence of *HFE* mutations has been shown in both European and American studies of PCT patients, and the use of phlebotomy to deplete excess iron stores is still recommended. Thus, *HFE* mutation analysis is of value in patients with PCT and may be of value in patients with chronic hepatitis C and NASH. It is probably not of value in patients with alcoholic liver disease.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

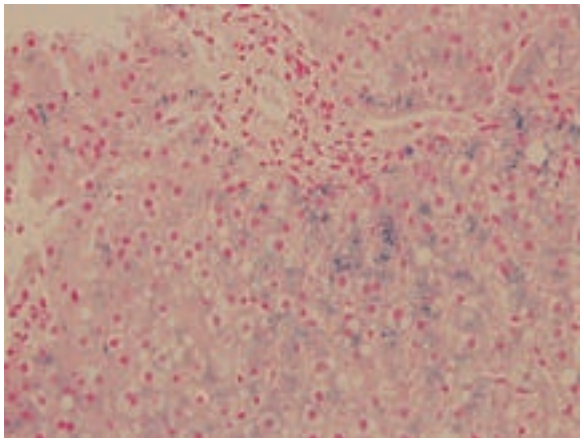

E-FIGURE 212-1. Mild iron deposition in a patient with chronic hepatitis C who is *HFE* negative. (Courtesy G. Chen, MD.)

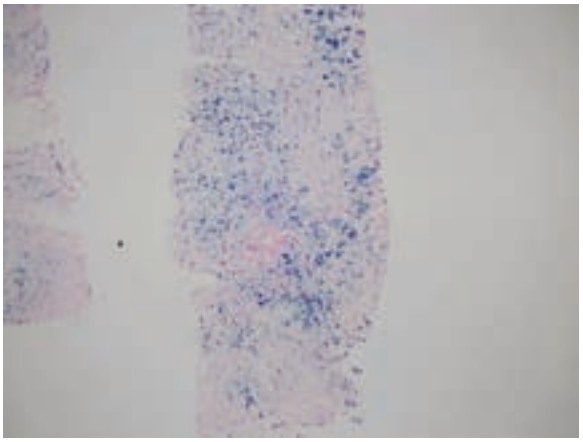

E-FIGURE 212-2. Moderate secondary iron overload in a patient with hepatitis C, with elevated ferritin and negative *HFE*. Note panlobular distribution with increased iron in sinusoidal lining cells. (Courtesy G. Chen, MD.)

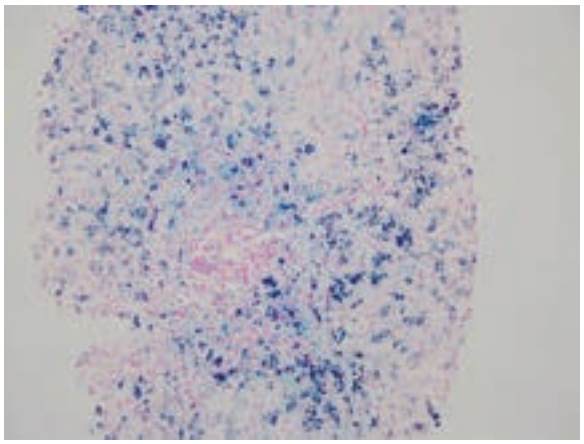

E-FIGURE 212-3. Higher power view of E-Figure 212-1. (Courtesy G. Chen, MD.)

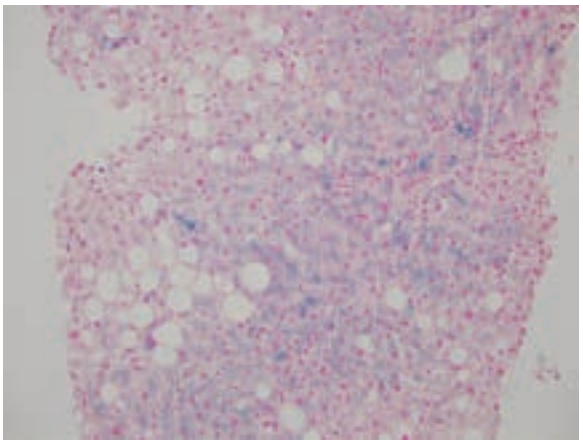

E-FIGURE 212-4. This patient with nonalcoholic steatohepatitis had an increase in ferritin with a negative *HFE* and grade 2 iron stores on liver biopsy. Iron is in both hepatocytes and sinusoidal lining cells. (Courtesy G. Chen, MD.)

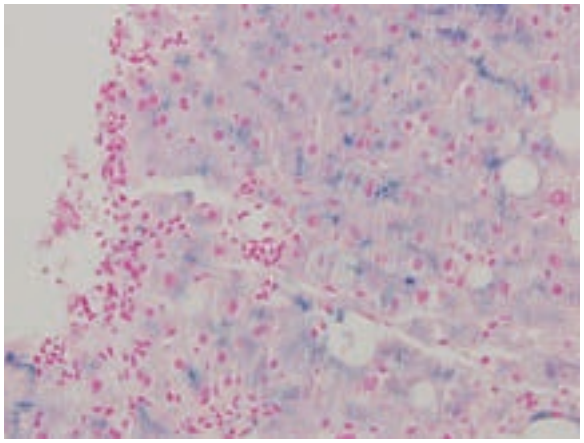

E-FIGURE 212-5. Higher power view of E-Figure 212-3.

GENERAL REFERENCES

1. Piperno A. Molecular diagnosis of hemochromatosis. *Expert Opin Med Diagn*. 2013;7:161-177.
2. Bardou-Jacquet E, Ben Ali Z, Beaumont-Epinette MP, et al. Non-HFE hemochromatosis: pathophysiological and diagnostic aspects. *Clin Res Hepatol Gastroenterol*. 2014;38:143-154.
3. Santos PC, Dinardo CL, Cancado RD, et al. Non-HFE hemochromatosis. *Rev Bras Hematol Hemoter*. 2012;34:311-316.
4. Pietrangelo A; European Association for the Study of the Liver. EASL clinical practice guidelines for HFE hemochromatosis. *J Hepatol*. 2010;53:3-22.
5. Ganz T. Systemic iron homeostasis. *Physiol Rev*. 2013;93:1721-1741.
6. Mastrogiannaki M, Matak P, Peyssonnaux C. The gut in iron homeostasis: role of HIF-2 under normal and pathological conditions. *Blood*. 2013;122:885-892.
7. VanWagner LB, Green RM. Elevated serum ferritin. *JAMA*. 2014;312:743-744.
8. McKinnon EJ, Rossi E, Beilby JP, et al. Factors that affect serum levels of ferritin in Australian adults and implications for follow-up. *Clin Gastroenterol Hepatol*. 2014;12:101-108.
9. Salgia RJ, Brown K. Diagnosis and management of hereditary hemochromatosis. *Clin Liver Dis*. 2015;19:187-198.
10. Bacon BR, Adams PC, Kowdley KV, et al. Diagnosis and management of hemochromatosis: 2011 practice guideline by the American Association for the Study of Liver Diseases. *Hepatology*. 2011;54:328-343.
11. Leitman SF. Hemochromatosis: the new blood donor. *Hematology Am Soc Hematol Educ Program*. 2013;645-650.
12. Hoffbrand AV, Taher A, Cappellini MD. How I treat transfusional iron overload. *Blood*. 2012;120:3657-3669.
13. Camaschella C. Treating iron overload. *N Engl J Med*. 2013;368:2325-2327.
14. Fleming RE, Ponka P. Iron overload in human disease. *N Engl J Med*. 2012;366:348-359.
15. Tirnitz-Parker JE, Glanfield A, Olynyk JK, et al. Iron and hepatic carcinogenesis. *Crit Rev Oncog*. 2013;18:391-407.
16. Nelson JE, Klintworth H, Kowdley KV. Iron metabolism in nonalcoholic fatty liver disease. *Curr Gastroenterol Rep*. 2012;14:8-16.
17. Caballes FR, Sendi H, Bonkovsky HL. Hepatitis C, porphyria cutanea tarda and liver iron: an update. *Liver Int*. 2012;32:880-893.

REVIEW QUESTIONS

1. How common is the C282Y/C282Y genotype in the United States?

A. 1 in 70
B. 1 in 250
C. 1 in 500
D. 1 in 1000

Answer: B Numerous large-scale prospective population surveys from North America, Australia, and Europe have determined that the prevalence of the C282Y/C282Y genotype is about 1 in 250 individuals (see Table 212-6).

2. Of individuals who are C282Y homozygotes, what percentage go on to develop phenotypic expression?

A. 100%
B. 75%
C. 50%
D. 25%

Answer: C Before the discovery of *HFE* in 1996, it was thought by experts in hereditary hemochromatosis (HH) that all patients with genotypic susceptibility would have phenotypic expression. After the discovery of *HFE*, several large population surveys from North America, Europe, and Australia have shown that about half of C282Y homozygotes do not have phenotypic expression. There is highly variable penetrance of C282Y homozygosity (see Definition and Epidemiology).

3. In patients with hemochromatosis who have symptoms, what is the most common presenting complaint?

A. Arthralgias
B. Heart failure
C. Impotence
D. Complications of cirrhosis

Answer: A Surveys of HH patients who present with symptoms show that arthralgias, usually in the hands and fingers, are the most common. Impotence, heart failure, and cirrhosis with complication are much less commonly seen.

4. Compared with the general population, what is the likelihood for development of hepatocellular cancer in cirrhotic patients with hemochromatosis?

A. 10-fold increase
B. 50-fold increase
C. 200-fold increase
D. 500-fold increase

Answer: C Cirrhosis from HH results in a major increase in risk for hepatocellular cancer above that in the general population. At one time, there was a concern that extrahepatic malignant disease was also increased, but this has not been confirmed.

5. What is the level of hepcidin in inherited syndromes of iron overload?

A. Increased
B. Same as normal
C. Decreased

Answer: C Inherited syndromes of iron overload should be remembered as hepcidin deficiency conditions. The mechanisms whereby hepcidin regulation results in reduced levels are still to be determined.

213 NUTRITION'S INTERFACE WITH HEALTH AND DISEASE

DOUGLAS C. HEIMBURGER

OLD AND NEW PARADIGMS IN THE SCIENCE OF NUTRITION

Nutrition science was characterized by two major phases in the 20th century. During the first phase, nutrition scientists discovered, characterized, and synthesized the essential nutrients and described their deficiency syndromes in detail. The dietary requirements for these nutrients were estimated and periodically updated as recommended dietary allowances (RDAs). Beginning in 1997, the RDAs were reformulated in a series of volumes containing dietary reference intakes (DRIs); in addition to the recommended intakes judged to be sufficient to meet the nutrient requirements of nearly all healthy individuals, the DRIs include estimates of tolerable upper intake levels—that is, the highest intake levels likely not to pose any adverse health risks.

More fundamentally, the DRIs focus on accumulating evidence related to the relationships of diet and nutritional status to the diseases that plague Western societies, such as coronary heart disease (CHD), cancer, diabetes, and the other leading causes of death. DRIs recommend intake levels that not only prevent deficiencies but also may promote long-term health and disease prevention.

NUTRITION'S INFLUENCE ON MORTALITY AND MORBIDITY

Evidence of a Connection between Diet and Disease

It has been estimated that between 300,000 and 800,000 deaths per year could be prevented in the United States if Americans followed evidence-based dietary recommendations. Substantial additional benefits from reduced morbidity and enhanced functional status would also accrue. However, the causal connections between diet and chronic diseases are difficult to tease out of the complex network of other risk factors, including social and behavioral variables, so a wide variety of studies must be relied on to establish these connections with reasonable certainty.

Epidemiologic studies are unable to infer causal relationships and may be confounded by variables that have not been examined. They are also challenged by the difficulty of accurately assessing the diets of free-living individuals.

Animal and in vitro studies can overcome some of these drawbacks but may be confounded by experimental conditions that differ from those encountered by humans. A large number of prospective, randomized human intervention trials have been undertaken to test the effects of dietary change on risk for disease. However, even these trials are not always conclusive because of pitfalls associated with selecting study populations and isolating individual dietary factors.

Taken together, epidemiologic, animal, in vitro, and intervention studies are proving that human dietary habits contribute importantly to the pathogenesis of most of the major causes of death in developed countries.[1] In the recently published Global Burden of Disease Study 2010, 15 dietary risk factors and physical inactivity collectively accounted for 10% of global deaths and disability-adjusted life years.[2]

Diseases Influenced by Nutrition

Table 213-1 lists 8 of the top 15 causes of death in the United States that are influenced by nutrition. Five are strongly linked to dietary habits, and three are associated with alcohol abuse. The table also outlines dietary contributions to obesity, atherosclerosis, osteoporosis, diverticular disease, and neural tube defects. A meta-analysis indicates that vitamin D_3 supplementation reduces mortality in adults.■ Table 213-2 summarizes the 2010 Dietary Guidelines for Americans, and Table 213-3 compares dietary recommendations promulgated by professional societies for risk reduction and/or management of the major chronic diseases. The close agreement among these recommendations enhances their credibility.

Coronary Heart Disease

Nutritional influences on the leading cause of death in the United States, CHD, have been the subject of a great deal of research. The overall U.S. mortality rate from CHD peaked in the 1960s and, in a trend that initially surprised medical science, has declined steadily since then. Changes in lifestyle, including diet, are responsible for a substantial proportion of this decline. Elevated plasma low-density lipoprotein (LDL) cholesterol levels are a major risk factor for CHD and peripheral atherosclerosis, and they correlate strongly with dietary saturated fat intake and less strongly with cholesterol intake. Intake of both these substances in the United States is derived largely from foods of animal origin, such as meats, dairy products, and eggs. Attempts to produce less atherogenic substitutes for some of these foods have not always proved beneficial. For instance, hydrogenation of vegetable oils to create margarine and shortening results in the formation of *trans*-fatty acids, which affect serum cholesterol levels in a manner similar to—and are perhaps even worse than—the saturated fatty acids found in butter and lard. LDL cholesterol levels can be lowered modestly by increasing the intake of soluble fiber from legumes, fruits, vegetables, and flax seed, as well as by consuming proteins and isoflavones from soy foods. LDL must be oxidized before it induces injury to the arterial wall. Although adequate dietary levels of the antioxidant vitamins C and E and β-carotene have been shown to inhibit LDL oxidation, pharmacologic doses of these vitamins have not reduced CHD events when tested in randomized trials.■ In fact, pharmacologic doses of vitamin E (>400 IU/day) and other antioxidants have no benefit and may increase all-cause mortality.

Epidemiologic evidence suggests that fish consumption may reduce CHD risk, perhaps through the action of omega-3 fatty acids, but randomized trials have found no benefit from omega-3 supplementation.■ Evidence also indicates that moderate consumption of alcohol, especially wine, is associated with a decreased risk for CHD, possibly by increasing high-density lipoprotein (HDL) cholesterol levels, preventing the oxidation of LDL, or both. The polyphenols in red wine are also apparently beneficial. Traditional Mediterranean lifestyle patterns—with diets high in vegetables, fruits, olive oil, fish, nuts, complex grains and carbohydrates, and red wine, along with physical activity—are associated with reduced risks for cardiovascular disease.■[3] A conservative estimate suggests that moderate dietary modification by the U.S. population, consisting mainly of replacing saturated fats with complex carbohydrates, fiber, monounsaturated fats, and fish, could easily lead to a 10% reduction in serum cholesterol levels and a 20% or greater reduction in CHD.[4] The actual risk reduction could be much greater.

Cancer

Nutrients, non-nutritive dietary constituents, and nutritional status can influence the risk for cancer in a variety of ways. Nutrition interacts with each step of carcinogenesis (carcinogen activation and tumor initiation, promotion, and progression). Excess energy intake may favor the generation of free radicals and reduce the body's ability to detoxify carcinogens. By contrast, antioxidant nutrients scavenge free radicals and other (pre)carcinogens and may thereby inhibit their activation, their ability to initiate mutations, or both. Folic acid may improve a cell's ability to preserve, repair, and methylate its DNA, either preventing or reversing the tendency toward mutation. Folic acid supplementation, however, may increase the cancer risk in some individuals, so supplementation beyond the levels present in multivitamins is not advised.■ Obesity has emerged as a major risk factor for many cancers, perhaps by inducing insulin resistance and elevating serum levels of insulin, insulin-like growth factor, and related tumor-promoting hormones. Excessive alcohol intake also promotes tumor growth.

Lung Cancer

Evidence indicates that the number-one cancer killer, lung cancer (Chapter 191), is influenced by diet. Although the most important causal factor is cigarette smoking, consumption of fruits and vegetables is inversely associated with lung cancer risk in both smokers and nonsmokers. It is probable that both nutrients and the non-nutritive phytochemicals in fruits and vegetables are responsible for the protective effects. However, in view of the disappointing results of randomized trials of supplementation with β-carotene, which increased mortality from lung cancer and other causes, antioxidant supplements should not be used to reduce disease risk, especially in tobacco smokers.

Breast Cancer

The number-two cause of cancer deaths in women, breast cancer (Chapter 198), is positively associated with obesity, especially when excess adiposity is located predominantly in the abdomen, and with physical inactivity. The

TABLE 213-1 DIETARY INFLUENCES ON MAJOR CAUSES OF DEATH AND MORBIDITY IN THE UNITED STATES

CAUSE OF DEATH OR MORBIDITY	FACTORS ASSOCIATED WITH DECREASED RISK	FACTORS ASSOCIATED WITH INCREASED RISK
DEATH		
Heart disease	Intake of complex carbohydrates, particular fatty acids (e.g., monounsaturated, polyunsaturated, and omega-3 fatty acids from fish), soluble fiber, polyphenols, soy proteins, antioxidants (vitamins E and C, β-carotene, selenium), folic acid, moderate alcohol	Intake of saturated fat, cholesterol, excess calories, sodium; abdominal distribution of body fat
Cancer	Intake of fruits and vegetables (for β-carotene; vitamins A, C, D, and E; folic acid; calcium; selenium; phytochemicals), fiber	Intake of excess calories, fat, alcohol, red meat, salt- and nitrite-preserved meats, possibly grilled meats; abdominal distribution of body fat
Cerebrovascular disease	Intake of potassium, calcium, omega-3 fatty acids	Intake of sodium, alcohol (as with hypertension)
Accidents		Excess alcohol consumption
Diabetes mellitus	Intake of fiber	Intake of excess calories, fat, alcohol; abdominal distribution of body fat
Suicide		Excess alcohol consumption
Chronic liver disease and cirrhosis		Excess alcohol consumption
Hypertension and hypertensive renal disease	Intake of fruits and vegetables, potassium, calcium, magnesium, omega-3 fatty acids	Intake of sodium, alcohol, excess calories, total and saturated fat; abdominal distribution of body fat
MORBIDITY		
Obesity		Intake of excess calories and fat
Osteoporosis	Intake of calcium, vitamin D, vitamin K	Intake of excess vitamin A, sodium, protein
Diverticular disease, constipation	Intake of fiber	
Neural tube defects	Intake of folic acid	

TABLE 213-2 DIETARY GUIDELINES FOR AMERICANS, 2010: KEY RECOMMENDATIONS

BALANCING CALORIES TO MANAGE WEIGHT

Prevent and/or reduce overweight and obesity through improved eating and physical activity behaviors.
Control total calorie intake to manage body weight. For people who are overweight or obese, this will mean consuming fewer calories from foods and beverages.
Increase physical activity and reduce time spent in sedentary behaviors.
Maintain appropriate calorie balance during each stage of life—childhood, adolescence, adulthood, pregnancy and breast-feeding, and older age.

FOODS AND FOOD COMPONENTS TO REDUCE

Reduce daily sodium intake to less than 2300 milligrams (mg) and further reduce intake to 1500 mg among persons who are 51 years and older and those of any age who are African American or have hypertension, diabetes, or chronic kidney disease. The 1500-mg recommendation applies to about half of the U.S. population, including children and most adults.
Consume less than 10% of calories from saturated fatty acids by replacing them with monounsaturated and polyunsaturated fatty acids.
Consume less than 300 mg per day of dietary cholesterol.
Keep *trans*-fatty acid consumption as low as possible by limiting foods that contain synthetic sources of *trans* fats, such as partially hydrogenated oils, and by limiting other solid fats.
Reduce the intake of calories from solid fats and added sugars.
Limit the consumption of foods that contain refined grains, especially refined grain foods that contain solid fats, added sugars, and sodium.
If alcohol is consumed, it should be consumed in moderation—up to one drink per day for women and two drinks per day for men—and only by adults of legal drinking age.
Individuals should meet the following recommendations as part of a healthy eating pattern while staying within their calorie needs:
- Increase vegetable and fruit intake.
- Eat a variety of vegetables, especially dark-green and red and orange vegetables and beans and peas.
- Consume at least half of all grains as whole grains. Increase whole-grain intake by replacing refined grains with whole grains.
- Increase intake of fat-free or low-fat milk and milk products, such as milk, yogurt, cheese, or fortified soy beverages.
- Choose a variety of protein foods, which include seafood, lean meat and poultry, eggs, beans and peas, soy products, and unsalted nuts and seeds.
- Increase the amount and variety of seafood consumed by choosing seafood in place of some meat and poultry.
- Replace protein foods that are higher in solid fats with choices that are lower in solid fats and calories and/or are sources of oils.
- Use oils to replace solid fats where possible.
- Choose foods that provide more potassium, dietary fiber, calcium, and vitamin D, which are nutrients of concern in American diets. These foods include vegetables, fruits, whole grains, and milk and milk products.

BUILDING HEALTHY EATING PATTERNS

Select an eating pattern that meets nutrient needs over time at an appropriate calorie level.
Account for all foods and beverages consumed and assess how they fit within a total healthy eating pattern.
Follow food safety recommendations when preparing and eating foods to reduce the risk for food-borne illnesses.

RECOMMENDATIONS FOR SPECIFIC POPULATION GROUPS

Women Capable of Becoming Pregnant

Choose foods that supply heme iron, which is more readily absorbed by the body, additional iron sources, and enhancers of iron absorption such as vitamin C–rich foods.
Consume 400 μg per day of synthetic folic acid (from fortified foods and/or supplements) in addition to food forms of folate from a varied diet.

Women Who Are Pregnant or Breast-Feeding

Consume 8 to 12 ounces of seafood per week from a variety of seafood types.
Because of their high methyl mercury content, limit white (albacore) tuna to 6 ounces per week and do not eat the following four types of fish: tilefish, shark, swordfish, and king mackerel.
If pregnant, take an iron supplement, as recommended by an obstetrician or other health care provider.

Individuals Aged 50 Years and Older

Consume foods fortified with vitamin B_{12}, such as fortified cereals, or dietary supplements.

TABLE 213-3 DIETARY GUIDELINES PROMULGATED BY NATIONAL ORGANIZATIONS*

	U.S. DEPARTMENT OF AGRICULTURE AND DEPARTMENT OF HEALTH AND HUMAN SERVICES: DIETARY GUIDELINES FOR AMERICANS (2010)	NATIONAL CHOLESTEROL EDUCATION PROGRAM ADULT TREATMENT PANEL III: THERAPEUTIC LIFESTYLE CHANGE DIET (2002)	NATIONAL HIGH BLOOD PRESSURE EDUCATION PROGRAM/JOINT NATIONAL COMMITTEE: 7 DIETARY APPROACHES TO STOP HYPERTENSION (DASH; 2006)	AMERICAN DIABETES ASSOCIATION (2004 AND 2012)	AMERICAN CANCER SOCIETY (2012)
Indication or Objective	***General Health Promotion and Disease Prevention***	***Elevated Cholesterol, Heart Disease Prevention***	***Prehypertension and Hypertension***	***Diabetes Prevention and Treatment***	***Cancer prevention***
NUTRIENT/FOOD GROUP					
Total energy	Prevent and/or reduce overweight and obesity through improved eating and physical activity behaviors	Manage weight, increase physical activity	Reduce energy intake to lose weight if overweight	Reduced energy intake and modest weight loss can improve glycemia and insulin resistance and reduce risk for type 2 diabetes	Achieve and maintain a healthy weight throughout life. Choose foods and beverages in amounts that help achieve and maintain a healthy weight
Fruits and vegetables	Increase vegetable and fruit intake. Eat a variety of vegetables, especially dark-green and red and orange vegetables and beans and peas		8-10 servings/day		Consume a healthy diet, with an emphasis on plant foods. Eat at least 2.5 cups of vegetables and fruits each day
Meat	Choose a variety of protein foods, which include seafood, lean meat, and poultry (for more, see Protein). Choose seafood in place of some meat and poultry		≤6 servings/day		Limit consumption of processed and red meats
Dairy	Increase intake of fat-free or low-fat milk and milk products, such as milk, yogurt, cheese, or fortified soy beverages		2-3 servings/day of low-fat dairy		
Grains, fiber	Consume at least half of all grains as whole grains. Replace refined grains with whole grains	Consider increased viscous (soluble) fiber (10-25 g/day) and plant stanols/sterols (2 g/day) to enhance LDL lowering	6-8 servings of whole grains and whole-grain products	Achieve U.S. Dietary Guidelines recommendation	Choose whole grains instead of refined grain products
Fat	Replace protein foods that are higher in solid fats with choices that are lower in solid fats and calories and/or are sources of oils. Use oils to replace solid fats where possible	25-35% of daily energy intake	<27% of daily energy intake		The totality of the evidence does not support a relationship between total fat intake and cancer risk
Saturated fats	<10% of daily calories, by replacing them with MUFA and PUFA	<7% of daily energy intake	6% of daily energy intake	<7% of daily energy intake	
Polyunsaturated fats	Replace solid fats with PUFA	Up to 10% of daily energy intake		Up to 10% of daily energy intake	
Monounsaturated fats	Replace solid fats with MUFA	Up to 20% of daily energy intake		15-20% of daily energy intake; combination of MUFA and carbohydrates should equal 60-70% of total energy intake	
Trans-fats	Keep as low as possible	Intake should be kept low		Intake should be minimized	
Cholesterol	<300 mg/day	<200 mg/day	150 mg/day	<300 mg/day; those with LDL cholesterol ≥100 may benefit from lowering cholesterol intake to 200 mg/day	
Carbohydrates		50-60% of daily energy intake	55% of daily energy intake	Total amount of carbohydrate is more important than the source or type. Individualize the mix of carbohydrate, protein, and fat	
Sugar	Reduce added sugars			Sucrose and sucrose-containing foods do not need specific restriction by diabetic persons. Persons at risk for type 2 diabetes should limit sugar-sweetened beverages	

TABLE 213-3 DIETARY GUIDELINES PROMULGATED BY NATIONAL ORGANIZATIONS—cont'd

	U.S. DEPARTMENT OF AGRICULTURE AND DEPARTMENT OF HEALTH AND HUMAN SERVICES: DIETARY GUIDELINES FOR AMERICANS (2010)	NATIONAL CHOLESTEROL EDUCATION PROGRAM ADULT TREATMENT PANEL III: THERAPEUTIC LIFESTYLE CHANGE DIET (2002)	NATIONAL HIGH BLOOD PRESSURE EDUCATION PROGRAM/JOINT NATIONAL COMMITTEE: 7 DIETARY APPROACHES TO STOP HYPERTENSION (DASH; 2006)	AMERICAN DIABETES ASSOCIATION (2004 AND 2012)	AMERICAN CANCER SOCIETY (2012)
Indication or Objective	***General Health Promotion and Disease Prevention***	***Elevated Cholesterol, Heart Disease Prevention***	***Prehypertension and Hypertension***	***Diabetes Prevention and Treatment***	***Cancer prevention***
Protein	Choose a variety of protein foods, which include eggs, beans and peas, soy products, and unsalted nuts and seeds (for more, see Meat)	Approximately 15% of daily energy intake	18% of daily energy intake	10-20% of daily energy intake if renal function is normal	
Alcohol	Up to 2 drinks/day for men and up to 1 drink/day for women; persons in special circumstances (e.g., pregnancy, history of alcoholism) should abstain		<2 drinks/day for men and <1 drink/day for women	If individuals choose to drink, limit to 2 drinks/day for men and 1 drink/day for women, and take extra precautions to prevent hypoglycemia	Drink no more than 1 drink/day for women or 2 drinks/day for men
Sodium	Reduce daily sodium intake to less than 2300 mg, and further reduce intake to 1500 mg among persons who are 51 years and older or are members of special groups		<2400 mg/day	<2400 mg/day	
Potassium			4700 mg/day		
Calcium			1250 mg/day	1000-1500 mg/day	1000-1200 mg/day
Magnesium			500 mg/day		

*For further details, see the websites listed in this chapter.
LDL = low-density lipoprotein; MUFA = monounsaturated fatty acid; PUFA = polyunsaturated fatty acid.
Adapted from Heimburger DC, Ard JD, eds. *Handbook of Clinical Nutrition*, 4th ed. Philadelphia: Elsevier; 2006.

positive association between obesity and postmenopausal breast cancer has been attributed in part to the synthesis of estrogen (a risk factor for breast cancer) in adipose tissue. Epidemiologic evidence suggests that alcohol intake may also be a risk factor for this disease, particularly in women with lower intakes of folic acid.

Colorectal Cancer

Colorectal cancer (Chapter 193) is the third leading cause of cancer mortality in men and women. The risk for colorectal cancer correlates positively with the intake of red meat (especially when it is overcooked) and dietary fat and with obesity, and the risk correlates inversely with the intake of calcium and folic acid. Evidence regarding the effect of dietary fiber is somewhat equivocal, but the preponderance of evidence points to a protective effect; there is no evidence of harm. Higher physical activity is associated with a 30 to 50% lower risk for colon cancer.

Summary

The interaction of all these influences is powerful enough to suggest that diet and physical inactivity contribute to well over 35% of cancer deaths in Western countries. Even though no independent benefit has been found from dietary supplementation with potentially protective nutrients such as carotenoids, vitamins C and E,[■■] folic acid, selenium,[■] and fiber, they are all present in vegetables and fruits, and a liberal intake of fruits and vegetables is strongly recommended. A very large randomized trial indicated that lowering dietary fat intake did not reduce rates of breast cancer[■] or colon cancer, but it is very likely that maintaining an appropriate body weight and being physically active can reduce cancer risks.

Hypertension

Elevated blood pressure (Chapter 67) is a major risk factor for stroke, CHD, heart failure, peripheral vascular disease, and renal disease. It is often associated with obesity, especially abdominal obesity, and weight reduction in obese hypertensive people generally leads to an improvement in blood pressure. Sodium restriction also usually reduces blood pressure levels. The Dietary Approaches to Stop Hypertension (DASH) diet, which is rich in fruits, vegetables, and low-fat dairy products and advocates a reduced saturated and total fat content, can also decrease blood pressure levels; reducing one's sodium intake provides an additional benefit when included as part of the DASH diet.[5] Because alcohol intake elevates blood pressure, its use should be limited in hypertensive patients.

Diabetes Mellitus

Type 2 diabetes mellitus (Chapter 229) is strongly associated with obesity, especially abdominal obesity, so maintenance of a desirable body weight throughout life is of major importance in both preventing and treating type 2 diabetes. Sugar consumption does not lead to diabetes, except to the extent that it may promote weight gain. A focus on dietary carbohydrate restriction is no longer recommended for persons with type 2 diabetes; rather, a variety of dietary patterns, including low-carbohydrate, low-glycemic index, and Mediterranean diets, are effective in improving glycemic control and markers of cardiovascular disease risk.[■] Because higher-fat diets tend to promote both obesity and CHD, for which diabetic patients are at high risk, dietary fat intake should be kept low. Alcohol can cause hypoglycemia, hyperglycemia, and increased triglyceride levels in diabetic patients, and its use should be limited. In both diabetic and nondiabetic people, excess alcohol intake is responsible for many deaths, particularly from accidents and liver disease, and it is a factor in some suicides.

Osteoporosis

Osteoporosis (Chapter 243) is influenced by several dietary factors. Inadequate calcium intake during adolescence can result in suboptimal peak bone mass in early adulthood, and during later life it can lead to accelerated bone loss, thereby increasing the risk for osteoporosis. Sodium and protein, which are consumed by Americans in greater quantities than required, may promote excess bone loss. Excessive supplementation with vitamin A reduces bone mass and increases fracture risk. Vitamin D, vitamin K, and magnesium assist in maintaining optimal bone mass.

Other Conditions

Obesity

The causes and health effects of obesity, the most prevalent nutritional disorder in the United States, are reviewed in Chapter 220. The metabolic

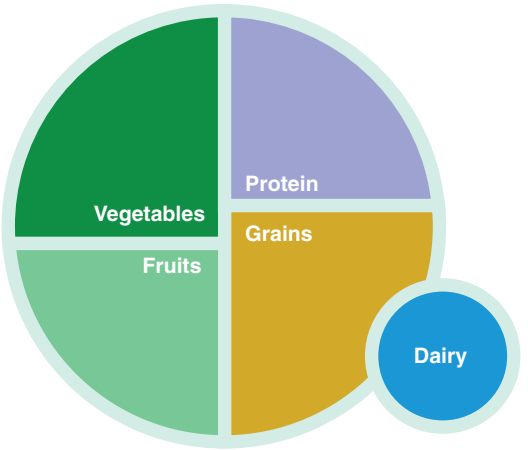

FIGURE 213-1. U.S. Department of Agriculture's MyPlate. MyPlate illustrates that a substantial proportion of dietary intake should be derived from vegetables, fruits, and grains, as contrasted with many Americans' expectation that protein sources should dominate. For more information, see www.ChooseMyPlate.gov.

syndrome, a constellation that includes obesity with an enlarged waist circumference; increased serum glucose, triglycerides, and blood pressure; and reduced HDL cholesterol, is strikingly prevalent in the United States and is a major risk factor for CHD, cancer, type 2 diabetes, and hypertension (Chapter 67).

Intestinal Diverticular Disease

Low dietary fiber intake causes constipation, and it is thought to be a cause of intestinal diverticular disease.

Congenital Neural Tube Defects

Inadequate maternal folic acid intake has been definitively proved to be a major risk factor for congenital neural tube defects such as spina bifida and myelomeningocele. For this reason, cereal and grain products have been fortified in the United States with folic acid since 1998.[6]

TRANSLATING EVIDENCE INTO DIETARY CHANGE

Thus, the evidence is strong that dietary habits can influence the incidence and severity of many incapacitating or lethal diseases in the United States. No justification exists for the belief that modification of the "usual" American diet is unnecessary or futile. The only questions are whether change is feasible and what is required to effect it. Various health agencies and the U.S. government have used public education, particularly the publication of dietary goals, as their primary means (see Table 213-3). The Department of Agriculture and the Department of Health and Human Services have developed and periodically revised the Dietary Guidelines for Americans (see Table 213-2) and a food guidance system, now called MyPlate (Fig. 213-1). MyPlate is part of a larger communications initiative based on the 2010 Dietary Guidelines for Americans to help consumers make better food choices. Although the practical application of nutritional genomics is not yet ready for routine clinical practice, it is an important, emerging science that may in the future be applied to nutrigenetic testing to provide dietary advice.[7]

Physicians can influence their patients' health by encouraging them to optimize their dietary habits and providing them with instructional materials and assistance from dietitians in making needed changes. A significant barrier to practical nutritional interventions could be removed if health insurers would reimburse dietitians' services.

Grade A References

A1. Bjelakovic G, Gluud LL, Nikolova D, et al. Vitamin D supplementation for prevention of mortality in adults. *Cochrane Database Syst Rev*. 2014;1:CD007470.
A2. Fortmann SP, Burda BU, Senger CA, et al. Vitamin and mineral supplements in the primary prevention of cardiovascular disease and cancer: An updated systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2013;159:824-834.
A3. Rees K, Hartley L, Flowers N, et al. "Mediterranean" dietary pattern for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev*. 2013;8:CD009825.
A4. Rizos EC, Ntzani EE, Bika E, et al. Association between omega-3 fatty acid supplementation and risk of major cardiovascular disease events: a systematic review and meta-analysis. *JAMA*. 2012;308:1024-1033.
A5. Ebbing M, Bønaa KH, Nygård O, et al. Cancer incidence and mortality after treatment with folic acid and vitamin B12. *JAMA*. 2009;302:2119-2126.
A6. Lin J, Cook NR, Albert C, et al. Vitamins C and E and beta carotene supplementation and cancer risk: a randomized controlled trial. *J Natl Cancer Inst*. 2009;101:14-23.
A7. Gaziano JM, Glynn RJ, Christen WG, et al. Vitamins E and C in the prevention of prostate and total cancer in men: the Physicians' Health Study II Randomized Controlled Trial. *JAMA*. 2009;301:52-62.
A8. Lippman SM, Klein EA, Goodman PJ, et al. Effect of selenium and vitamin E on risk of prostate cancer and other cancers: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *JAMA*. 2009;301:39-51.
A9. Thomson CA, Van Horn L, Caan BJ, et al. Cancer incidence and mortality during the intervention and postintervention periods of the Women's Health Initiative dietary modification trial. *Cancer Epidemiol Biomarkers Prev*. 2014;23:2924-2935.
A10. Ajala O, English P, Pinkney J. Systematic review and meta-analysis of different dietary approaches to the management of type 2 diabetes. *Am J Clin Nutr*. 2013;97:505-516.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Samieri C, Sun Q, Townsend MK, et al. The association between dietary patterns at midlife and health in aging: an observational study. *Ann Intern Med*. 2013;159:584-591.
2. Lim SS, Vos T, Flaxman AD, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study, 2010. *Lancet*. 2012;380:2224-2260.
3. Kastorini CM, Milionis HJ, Esposito K, et al. The effect of Mediterranean diet on metabolic syndrome and its components: a meta-analysis of 50 studies and 534,906 individuals. *J Am Coll Cardiol*. 2011;57:1299-1313.
4. Artinian NT, Fletcher GF, Mozaffarian D, et al. Interventions to promote physical activity and dietary lifestyle changes for cardiovascular risk factor reduction in adults: a scientific statement from the American Heart Association. *Circulation*. 2010;122:406-441.
5. Koliaki C, Katsilambros N. Dietary sodium, potassium, and alcohol: key players in the pathophysiology, prevention, and treatment of human hypertension. *Nutr Rev*. 2013;71:402-411.
6. Copp AJ, Stanier P, Greene ND. Neural tube defects: recent advances, unsolved questions, and controversies. *Lancet Neurol*. 2013;12:799-810.
7. Camp KM, Trujillo E. Position of the Academy of Nutrition and Dietetics: nutritional genomics. *J Acad Nutr Diet*. 2014;114:299-312.

SUGGESTED WEBSITES

US Dietary Guidelines 2010: http://www.health.gov/dietaryguidelines/2010.asp.
US Dietary Guidelines 2015: http://www.health.gov/dietaryguidelines/2015.asp.
American Cancer Society: www.cancer.org.
American Diabetes Association: www.diabetes.org.
American Heart Association: www.americanheart.org.
NIH National Heart, Lung, and Blood Institute Nutrition Tools and Resources: http://www.nhlbi.nih.gov/health/public/heart/obesity/wecan/tools-resources/nutrition.htm.
Nutrition Data.com: www.nutritiondata.com.
USDA MyPlate: www.ChooseMyPlate.gov.

REVIEW QUESTIONS

1. Which of the following, when included in the diet, is most likely to increase a person's risk for cardiovascular disease?

A. Monounsaturated fatty acids
B. *Trans*-fatty acids
C. Soy protein
D. Alcohol

Answer: B As noted in the Coronary Heart Disease section of the chapter, hydrogenation of vegetable oils to create margarine and shortening results in the formation of *trans*-fatty acids, which affect serum cholesterol levels in a manner similar to—and are perhaps even worse than—the saturated fatty acids found in butter and lard. The other three choices are all noted in the same section to have potentially protective effects against coronary heart disease.

2. The risk for hypertension is reduced by relatively high intakes of which of the following dietary sources?

A. Alcohol
B. Fat
C. Fruits and vegetables
D. Sodium

Answer: C As noted in the Hypertension section of the chapter, the Dietary Approaches to Stop Hypertension (DASH) diet, which is rich in fruits, vegetables, and low-fat dairy products and advocates a reduced saturated and total fat content, can decrease blood pressure levels; reducing one's sodium intake provides an additional benefit both alone and when included as part of the DASH diet.

3. Which of the following lifestyle habits is most likely to reduce the risk for chronic diseases?

A. Multivitamin supplementation
B. Vitamin A supplementation
C. Dietary fat restriction
D. General adherence to dietary recommendations

Answer: D There is no clear evidence that supplementation with multivitamins or vitamin A, or restriction of dietary fat alone, will reduce the risk for chronic diseases. As noted in the Osteoporosis section of the chapter, excessive supplementation with vitamin A reduces bone mass and increases fracture risk. Evidence supports general adherence to dietary recommendations as the best nutritional means to reduce the risk for chronic diseases.

4. Long-term control of the blood glucose level in an obese type 2 diabetic patient is likely to be improved most by which of the following?

A. Moderate caloric restriction
B. Fat restriction
C. Protein restriction
D. Carbohydrate restriction

Answer: A As noted in the Diabetes Mellitus section of the chapter, dietary carbohydrate restriction is no longer the focus of management of type 2 diabetes, and persons with obesity and type 2 diabetes are best served by weight reduction. This is best accomplished by moderate caloric restriction, not by focusing on restriction of carbohydrates, fats, or protein.

5. A 54-year-old white woman has a bone mineral density that indicates osteopenia in both the spine and hip. Which of the following nutrients may enhance the efficacy of calcium in preserving her bone mineral density?

A. Vitamin A
B. Vitamin K
C. Protein
D. Sodium

Answer: B As noted in the Osteoporosis section of the chapter, vitamin K assists in maintaining optimal bone mass, whereas excess intakes of vitamin A, protein, and sodium promote bone loss.

214

NUTRITIONAL ASSESSMENT

BRUCE R. BISTRIAN

GOALS AND IMPORTANCE OF NUTRITIONAL ASSESSMENT

Nutritional assessment in clinical medicine has three primary goals: to identify the presence and type of malnutrition, to define health-threatening obesity, and to devise suitable diets as prophylaxis against disease later in life. The focus of this chapter is on the diagnosis of protein-energy malnutrition because of its wide prevalence and major impact on disease outcome.[1,2] Other deficiency diseases are of much less relevance in that most occur in conjunction with protein-energy malnutrition or in specific disease states, such as thiamine deficiency in alcoholic liver disease and fat-soluble vitamin deficiency in malabsorptive states. The classic deficiency diseases, whether primary or secondary, are considered elsewhere in those chapters specifically dealing with the diseases mentioned here. The widespread availability of parenteral and enteral therapeutic measures since the mid-1980s that can provide adequate feeding regimens for virtually any disease condition makes a rudimentary knowledge of the pathophysiology of protein-energy malnutrition and its nutritional assessment essential for all primary care practitioners (Chapter 213).

CLINICAL NUTRITIONAL ASSESSMENT

Clinical assessment of protein nutritional status is based principally on the clinical history, physical examination including simple anthropometry, and measurement of the levels of several secretory proteins.

Clinical History

Although detailed dietary assessment can at times be helpful, in most circumstances physicians can safely limit their diet questions to whether patients have been following a prescribed diet, how much alcohol they drink, and whether they habitually take dietary supplements, including vitamins, minerals, and herbs. In ambulatory patients, the ability to maintain usual and adequate weight generally indicates that serious micronutrient deficiency is probably not the result of dietary inadequacy. Isolated vitamin deficiencies in the absence of weight loss or symptoms are rare, except perhaps for folate and vitamin B_{12} (Chapter 164). Although nutritional anemias do exist, as a consequence of strict vegetarian or vegan diets for vitamin B_{12}, or uncommonly today for folate except with extreme diets (as a result of widespread fortification of folate in food in developed countries), the role of dietary deficiency in anemias is limited in the absence of underlying disease, altered physiology (e.g., achlorhydria with aging), or weight loss. Only iron deficiency is a common cause of dietary anemia (Chapter 159). By contrast, full dietary assessment and diet prescriptions are likely to help patients with conditions such as fat malabsorption accompanied by weight loss, cramps, or diarrhea. Such evaluations are most effectively carried out by dietitians. Thus, detailed nutritional assessment of protein-energy malnutrition with secondary assessment of vitamin and mineral deficiencies is usually needed only when

protein-energy malnutrition or a specific disorder known to interfere with nutrient metabolism coexists, such as celiac disease, pernicious anemia, or nutrient-drug interactions. Even then, the assessment should emphasize the likely deficiencies. For fat malabsorption (Chapter 140), one should check levels of the fat-soluble vitamins A, D, E, and K; important divalent and trivalent cations (calcium [Ca^{2+}], zinc [Zn^{2+}], magnesium [Mg^{2+}], and iron [Fe^{3+}]); and phosphorus and alkaline phosphatase. When ileal resection has occurred, serum vitamin B_{12} levels should be measured, and the potential for bile salt depletion should be considered. Weight loss resulting from short-gut syndrome should prompt assessment of the fat-soluble vitamins, folic acid, vitamin B_{12}, calcium, magnesium, phosphorus, zinc, and iron. At least initially, levels of the water-soluble vitamins likely to have clinical impact such as thiamine and ascorbic acid should be checked, but provision of the full complement of vitamins should be routine in the management of this condition. Measurements of body water status (weight, blood urea nitrogen, serum creatinine, serum sodium) and acid-base balance (serum carbon dioxide combining power, chloride and potassium, and urine and arterial pH) should be obtained if the diarrhea is profuse (Chapter 140).

Protein-energy malnutrition, as defined by significant weight loss of more than 5% or hypoalbuminemia, affects at least 25% of patients hospitalized for acute care and are secondary to the underlying disease in most instances. Many of these patients can benefit from nutritional support and require a thorough clinical nutritional assessment, including a dietary history, physical examination, and laboratory tests that serve to confirm clinical impressions. The history should list information about the timing and amount of weight loss, medical illnesses, medications, gastrointestinal symptoms (abdominal pain, diarrhea, dysphagia), diet habits (eating fewer than two meals per day, alcohol consumption, dietary supplement intake, dental status), social habits (eating alone, needing assistance in self-care), economic status (having enough money for food), and mental status, particularly the presence of depressive symptoms. A special focus should be reserved for elderly people, in whom protein-energy malnutrition secondary to these last factors is more common (Chapter 24).

Nutritional Support

Four factors principally determine the timing, need, and appropriateness of nutritional support: (1) the presence and severity of protein-energy malnutrition, defined primarily by degree of weight loss and weight-to-height ratio as a percentage of standard or body mass index (BMI); (2) the presence and severity of the systemic inflammatory response, defined principally by the serum albumin level but also by the presence of fever, leukocytosis, and increased band forms; (3) the actual or expected duration of inadequate nutritional intake; and (4) the prognosis of the underlying condition. Well-nourished individuals have a 7- to 10-day reserve of energy and protein to withstand a moderate systemic inflammatory response without adverse nutritional consequences. Greater degrees of systemic inflammatory response and preexisting protein-energy malnutrition dramatically shorten the period that semistarvation, defined as consuming less than 50% of the energy and protein needs, can be tolerated. An important corollary is that when energy intake is limited, protein requirements for optimal outcomes are increased such that a reduction in energy intake to 50% of needs nearly doubles the desirable intake for protein.

Weight Loss

A recent unintended weight loss of 10 lb, or more than 5% of usual weight, should prompt efforts to diagnose the underlying disorder or social circumstance. Weight loss alone does not distinguish the composition of tissue loss, which can range from 25 to 30% lean tissue in semistarvation alone to 50% lean tissue loss following semistarvation plus injury and as much as 75% of the weight loss with the severest forms of injury, severe sepsis, major body burns, severe closed head injury, and multiple trauma. Therefore, unintentional weight loss of more than 10 lb or more than 5% of usual weight indicates a need for thorough nutritional assessment. Weight loss in excess of 10% of usual weight should be considered to represent protein-energy malnutrition that will impair physiologic function, particularly muscle strength and endurance. Weight loss in excess of 20% should be considered severe protein-energy malnutrition that will substantially impair the function of most organ systems. If major elective surgery is planned (Chapter 431), such individuals would benefit from adequate feeding preoperatively for up to 7 days or at least early nutritional intervention postoperatively. If palliative or curative radiation therapy or systemic chemotherapy is planned, adequate feeding during therapy with the use of supplemental formulas, tube feeding, or parenteral nutrition (in that order) is indicated, with enteral feeding preferred. However, if the weight loss represents end-stage systemic illness, commonly seen as a cachexia syndrome of loss of weight and muscle mass with symptoms of weakness (e.g., cancer; end-stage liver, renal, or lung disease; acquired immunodeficiency syndrome) for which no primary therapy is planned or is effective, invasive nutritional support is rarely indicated.

Physical Examination

Although the patient's external appearance and a check of the skin, eyes, mouth, hair, and nails often provide clues to the presence of nutritional abnormalities (Table 214-1), the physical findings of deficiency syndromes of vitamins, essential fatty acids, and trace metals are relatively insensitive and nonspecific. With respect to protein-energy malnutrition, only the marasmic form in which semistarvation is the principal mechanism and cachexia syndromes resulting from semistarvation and a mild systemic inflammatory response are evident at examination. Adult marasmus and cachexia, the more severe forms, can be defined as 15% or more of recent weight loss, a weight-to-height ratio lower than 85% of desirable weight, or a BMI of less than 17. Loss of subcutaneous fat and skeletal muscle is manifested by sunken temples, thin extremities, wasting of the muscles of the hand, and, rarely, edema. Although kwashiorkor in children is characterized by severe edema and a potbelly appearance from hepatomegaly and ascites, one rarely encounters these clinical signs in cases of hypoalbuminemic malnutrition that develops in the setting of systemic inflammation resulting from disease in industrialized societies. Significant hypoalbuminemia of less than 3.5 g/dL results from a systemic inflammatory response, often without the accompanying anthropometric changes for which nutritional support can help to improve outcome if the response is prolonged beyond a week. With severe injury and greater depression of serum albumin to the 2.4 g/dL level or lower, early feeding within the first 24 to 48 hours can improve outcome. A mixed picture of protein-energy malnutrition with weight loss and lean tissue loss with mild hypoalbuminemia (cachexia) is found with many chronic inflammatory conditions with end-stage systemic illness or the sarcopenia of aging.

Body Weight

The most useful element in the physical examination is body weight, which is expressed as a relative value to evaluate the patient in relation to the healthy population. Weight and height are easily obtained, and standards for comparison have been established (Table 214-2). Although newer standards are available, they reflect the increasing prevalence of obesity in the U.S. population. Use of the 1959 standards allows the same tables to be used to diagnose significant protein-energy malnutrition (<85% of desirable weight, which approximates the fifth percentile) and significant obesity, defined as obesity predisposing to excessive mortality risk (>130% of desirable weight or BMI of ≥30). Although severe protein-energy malnutrition often occurs at levels greater than 85% of desirable weight because of the greater likelihood of preexisting obesity, this condition is generally detected by percentage of weight loss or by upper arm anthropometry. Height can be measured in a reclining patient with a tape measure, and in certain situations the clinician may rely on the patient's history. The major confounding variable that limits the value of weight and height as an index of protein-energy malnutrition is the tendency for water retention with disease, and thus weight gain may not reflect an increase in lean body mass or protein content. Fluid retention is particularly a problem in patients with hypoalbuminemic malnutrition because of the effects of aldosterone, antidiuretic hormone, and insulin stimulated by the stress response, which causes sodium and fluid retention. Fluid retention, however, does not usually confound initial weight assessment in patients who are first seen at the physician's office, except in those patients with diseases such as cardiac failure, end-stage liver disease, and severe renal disease in whom the disturbance in water metabolism results from the underlying disease and not principally from the hormonal response to systemic inflammation.

Body Mass Index

The BMI, which is the weight in kilograms divided by the height in meters squared, has gained favor as a nutritional measure because of two valuable attributes. The measure is relatively independent of height, and the same standards apply to male and female patients. The following BMI values are used: normal nutrition, BMI of 20 to less than 30; significant protein-energy malnutrition, less than 18.5; overweight, from 25 to less than 30; and obesity, 30 or greater, with severe obesity defined as a BMI of 35 to less than 40, and morbid obesity as 40 and greater. Evidence from developing countries

TABLE 214-1 CLINICAL SIGNS AND SYMPTOMS OF NUTRITIONAL INADEQUACY IN ADULT PATIENTS

	CLINICAL SIGN OR SYMPTOM	NUTRIENT
General	Wasted, skinny appearance	Calorie
	Loss of appetite	Protein-energy, zinc
Skin	Psoriasiform rash, eczematous scaling	Zinc, vitamin A, essential fatty acids
	Pallor	Folate, iron, vitamin B_{12}, copper
	Follicular hyperkeratosis	Vitamin A, vitamin C
	Perifollicular petechiae	Vitamin C
	Flaking dermatitis	Protein-energy, niacin, riboflavin, zinc
	Bruising	Vitamin C, vitamin K
	Pigmentation changes	Niacin, protein-energy
	Scrotal dermatosis	Riboflavin
	Thickening and dryness of skin	Linoleic acid
Head	Temporal muscle wasting	Protein-energy
Hair	Sparse and thin, dyspigmented	Protein
	Easy to pull out	Protein
	Corkscrew hairs	Vitamin C
Eyes	History of night blindness (also impaired visual recovery after glare)	Vitamin A, zinc
	Photophobia, blurring, conjunctival inflammation	Riboflavin, vitamin A
	Corneal vascularization	Riboflavin
	Xerosis, Bitot's spots, keratomalacia	Vitamin A
Mouth	Glossitis	Riboflavin, niacin, folic acid, vitamin B_{12}, pyridoxine
	Bleeding gums	Vitamin C, riboflavin
	Cheilosis	Riboflavin, pyridoxine, niacin
	Angular stomatitis	Riboflavin, pyridoxine, niacin
	Hypogeusia	Zinc
	Tongue fissuring	Niacin
	Tongue atrophy	Riboflavin, niacin, iron
	Nasolabial seborrhea	Pyridoxine
Neck	Goiter	Iodine
	Parotid enlargement	Protein
Thorax	Thoracic rosary	Vitamin D
Abdomen	Diarrhea	Niacin, folate, vitamin B_{12}
	Distention	Protein-energy
	Hepatomegaly	Protein-energy
Extremities	Edema	Protein, thiamine
	Softening of bone	Vitamin D, calcium, phosphorus
	Bone tenderness	Vitamin D
	Bone ache, joint pain	Vitamin C
	Muscle wasting and weakness	Protein, calorie, vitamin D, selenium, sodium chloride
	Muscle tenderness, muscle pain	Thiamine
Nails	Spooning	Iron
	Transverse lines	Protein
Neurologic	Tetany	Calcium, magnesium
	Paresthesias	Thiamine, vitamin B_{12}
	Loss of reflexes, wristdrop, footdrop	Thiamine
	Loss of vibratory and position sense	Vitamin B_{12}
	Ataxia	Vitamin B_{12}
	Dementia, disorientation	Niacin
Blood	Anemia	Vitamin B_{12}, folate, iron, pyridoxine
	Hemolysis	Phosphorus, vitamin E

TABLE 214-2 DESIRABLE WEIGHT IN POUNDS IN RELATION TO HEIGHT FOR MEN AND WOMEN 25 YEARS OR OLDER*

MEN, MEDIUM FRAME				*WOMEN, MEDIUM FRAME*			
HEIGHT		**WEIGHT (lb)**		**HEIGHT**		**WEIGHT (lb)**	
Ft	***In***	***Range***	***Midpoint***	***Ft***	***In***	***Range***	***Midpoint***
				4	8	93-104	98.5
				4	9	95-107	101
				4	10	98-110	104
				4	11	101-113	107
				5	0	104-116	110
5	1	113-124	118.5	5	1	107-119	113
5	2	116-128	122	5	2	110-123	116.5
5	3	119-131	125	5	3	113-127	120
5	4	122-134	128	5	4	117-132	124.5
5	5	125-138	131.5	5	5	121-136	128.5
5	6	129-142	135.5	5	6	125-140	132.5
5	7	133-147	140	5	7	129-144	136.5
5	8	137-151	144	5	8	133-148	140.5
5	9	141-155	148	5	9	137-152	144.5
5	10	145-160	153	5	10	141-156	148.5
5	11	149-165	157				
6	0	153-170	161.5				
6	1	157-175	166				
6	2	162-180	171				
6	3	167-185	176				

*Corrected to nude weights and heights by assuming 1-inch heel for men, 2-inch heel for women, and indoor clothing weight of 5 and 3 lb for men and women, respectively.
Adapted from the *Metropolitan Life Insurance Company Statistical Bulletin,* 1959;4:1.

suggests that the BMI is better correlated with outcome than are weight and height.

Upper Arm Anthropometry

Approximately 50% of body fat is subcutaneous. The use of skinfold calipers to define the triceps skinfold thickness (TSF) is the most practical technique to estimate body fat. Standards for skinfold measurement are available from the National Health and Nutrition Examination Surveys I and II and were derived from a probability sample of the U.S. population. Generally, a value lower than the fifth percentile is used to define abnormality (Table 214-3).

The principal value of the TSF measurement is to determine the arm muscle circumference (AMC) or arm muscle area.

$$\text{AMC (cm)} = \text{arm circumference} - (\pi)(\text{TSF})[\text{mm}])/10$$

The AMC is a specific measure of protein-energy malnutrition if the fifth or tenth percentile is chosen as the cutoff point, and it is particularly valuable in patients with edematous states and in amputees, in whom weights are inaccurate or insensitive. The TSF and AMC measurements[3] are most useful in initially defining marasmic-type malnutrition or the mixed disorder. Many dietitians are skilled in upper arm anthropometry.

Serum Proteins

Despite many concerns, the serum albumin level remains the traditional standard for nutritional assessment by virtue of its extensive history and its continued use to separate the principal forms of protein-energy malnutrition. Hypoalbuminemia is a strong predictor of risk for morbidity and mortality[4] in both hospitalized and ambulatory patients. In almost all cases, except perhaps for hereditary analbuminemia, excessive loss secondary to nephrotic syndrome (Chapter 121), and, occasionally, protein-losing enteropathy, hypoalbuminemia identifies the recent or ongoing presence of a systemic inflammatory response. A value for serum albumin of less than 3.5 g/dL is considered to indicate a mild systemic inflammatory response, whereas a value of less than 2.4 g/dL represents a severe systemic inflammatory response, reflecting systemic inflammation that produces anorexia (limiting food intake) and increases protein catabolism and thus accelerates the development of protein-calorie malnutrition. With a half-life for albumin of 18 to 20 days and the fractional replacement rate of about 10% per day, the return of serum albumin to normal takes about 2 weeks of feeding when the stress response remits. Adequate feeding in the presence of systemic inflammation will not increase the serum albumin concentration, even though substantial nutritional benefit will occur in terms of wound healing and immune function. Levels of other proteins, such as transferrin, prealbumin, and retinol-binding protein, with respective half-lives of 7 days, 2 days, and half a day, also fall acutely with injury and respond more quickly when systemic inflammation remits. Serum transferrin also varies with iron status, however, and prealbumin and retinol-binding protein vary with dietary carbohydrate

TABLE 214-3 FIFTH, TENTH, AND FIFTIETH PERCENTILES FOR TRICEPS SKINFOLD AND MID-UPPER ARM MUSCLE CIRCUMFERENCE OF U.S. MEN AND WOMEN FROM THE FIRST NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY

AGE GROUP	MUAMC (cm) PERCENTILE			TSF (mm) PERCENTILE		
	5th	10th	50th	5th	10th	50th
MEN						
18-74	23.8	24.8	27.9	4.5	6	11
18-24	23.5	24.4	27.2	4	5	9.5
25-34	24.2	25.3	28	4.5	5.5	12
35-44	25	25.6	28.7	5	6	12
45-54	24	24.9	28.1	5	6	11
55-64	22.8	24.4	27.9	5	6	11
65-74	22.5	23.7	26.9	4.5	5.5	11
WOMEN						
18-74	18.4	19	21.8	11	13	22
18-24	17.7	18.5	20.6	9.4	11	18
25-34	18.3	18.9	21.4	10.5	12	21
35-44	18.5	19.2	22	12	14	23
45-54	18.8	19.5	22.2	13	15	25
55-64	18.6	19.5	22.6	11	14	25
65-74	18.6	19.5	22.5	11.5	14	23

MUAMC = mid-upper arm muscle circumference; TSF = triceps skinfold.
From Bishop CW, Bowen PE, Ritchey SJ. Norms for nutritional assessment of American adults by upper arm anthropometry. *Am J Clin Nutr*. 1981;34:2530-2539.

and renal function. As a result, these proteins do not reliably identify the presence and severity of the systemic inflammatory response any better than does albumin, but they reflect the nutritional response more quickly when inflammation lessens.

Composite Screening Tools

Investigators have made numerous attempts to combine the various components of nutritional assessment, including clinical history, physical examination, anthropometry, and serum proteins, into a single score.[5] Some of the more widely used tools include the following: the Subjective Global Assessment,[6] which classifies patients as A, B, or C or as having normal, mild, or moderate malnutrition; the Nutritional Risk Index, which is based on weight loss and serum albumin only; and the more extensive evaluations with the Mini-Nutritional Assessment and Malnutrition Universal Screening Tool. A clear advantage of one technique over another has not been established.

NUTRITIONAL THERAPY AND ITS ASSESSMENT

The same indices that are used in the baseline nutritional assessment can be used to assess response to therapy, provided certain points are kept in mind.

Assessing Lean Body Mass and Total Body Water

In a stressed, hospitalized patient receiving nutritional support, day-to-day weight changes generally reflect changes in fluid balance rather than energy balance. In an ambulatory setting, weight increases or decreases are most likely to reflect changes in protein nutritional status and body fat because the underlying illness is usually less severe. Even the most sensitive research methods for assessing changes in lean body mass, however, do not offer major improvements in diagnosis in the more seriously ill patients. Techniques that measure total body water, such as isotope dilution and underwater weighing, from which lean tissue is extrapolated, fail to account for the distortion in hydration of lean tissue with illness. Surrogate measures of total body protein to estimate lean tissues such as total body potassium measurement and dual x-ray absorptiometry do not adjust for differing body composition with disease. A newer method, single-frequency or multifrequency body impedance analysis, does show promise as a simple, accurate, noninvasive method[7] that may allow distinction between intracellular and extracellular water, with the former used to estimate lean tissue for an initial assessment. However, the inherent difficulties resulting from the greater disturbance of total body water in critically ill patients have not been overcome with this technique. Magnetic resonance imaging and total body nitrogen analysis are the most reliable tools for assessing the amount of lean tissue but are primarily useful for research purposes. The gold standard for assessing lean tissue is in vivo neutron activation analysis for nitrogen, but this is and will remain a research procedure. Combining clinically available, rapid techniques like dual x-ray absorptiometry scanning to assess body fat content with body impedance analysis[8] to measure total body water has shown some promise in estimating true lean tissue in critically ill patients.

Restoration of Lean Tissue

In an unstressed patient with the marasmic form of protein-calorie malnutrition, providing adequate energy and 1.2 to 1.5 g/kg protein should cause a positive nitrogen balance of 2 to 6 g/day (60 to 180 g lean tissue) and slow weight gain, depending on the extent of positive energy balance. For instance, a 300-kcal excess of intake over expenditure would provide approximately 120 g of lean tissue (100-kcal equivalent) in addition to 200 kcal (22 g) as fat, for a total of approximately 140 g, or approximately ⅓ pound of weight per day. Weight gains in excess of this number usually reflect sodium and thus water retention from the insulin stimulated by dietary carbohydrate. Such overhydration can be improved by reducing salt and limiting fluid intake. In patients with hypoalbuminemic malnutrition who are no longer stressed, a similar nutritional regimen will lead to a comparable gain of tissue, but weight change is often less as edema becomes mobilized, with normalization of serum albumin in 2 to 4 weeks and of retinol-binding protein, prealbumin, and transferrin more quickly. In stressed patients with hypoalbuminemic malnutrition, appropriate nutritional support often does not restore lean tissue but does improve other important functions, such as wound healing and immunocompetence. These are important treatment goals because they can improve the ultimate clinical outcome. Both the systemic inflammatory response and the limited activity level reduce the efficiency of skeletal muscle repletion, which represents 30% of body weight and 75% of actively metabolizing lean tissue. Functional testing of muscle strength and endurance, such as hand dynamometry, can be useful as a means of assessing this response but has not found wide clinical acceptance. Similarly, any reduction in other physiologic functions or impairment in the patient's ability to perform the usual activities of daily living will accentuate the consequences of protein-energy malnutrition. Cachexia syndromes reflect the loss of lean tissue with mild to moderate persistent inflammation due to an underlying chronic disease. Response to nutritional therapy is generally poor unless there is improvement or correction of the basic disease process.

Measures of Energy Expenditure and Caloric Need

Although caloric expenditure can now be reliably and easily measured with portable indirect calorimeters, estimated energy expenditure is sufficient in most clinical situations. The three components of total energy expenditure are basal energy expenditure (~55 to 65% of total energy expenditure), thermal effect of feeding (~10% of total energy expenditure), and activity energy expenditure (25 to 33%). An energy intake of 30 to 35 kcal/kg of body weight will maintain weight in most sedentary ambulatory patients, with adjustments upward or downward in 200- to 300-kcal increments as prompted by biweekly changes in weight. Although young, severely burned, or traumatized patients may require 35 to 40 kcal/kg in the acute phase to meet total energy expenditure, providing energy intakes principally as carbohydrates that exceed 35 kcal/kg substantially increases the likelihood of hyperglycemia. Evidence strongly implicates hyperglycemia in excess of 180 mg/dL as a major risk factor for nosocomial infection, thus emphasizing the importance of better glycemic control by the use of insulin infusions or by reducing the level of energy intake, or both. Most postoperative patients who require invasive nutritional support for mechanical or infectious complications usually require approximately 25 kcal/kg to meet energy needs and not more than 30 kcal/kg because of their older age and reduced activity and energy expenditure. Overfeeding should be avoided in such patients, and modest underfeeding for the first several weeks of acute illness may actually improve outcome in seriously ill patients.

The nutritional monitoring of the most critically ill patients[9] is primarily to assess how closely estimated nutritional needs are delivered on a daily basis, rather than any other presently defined nutritional marker beyond weight as an estimate of fluid status. Early parenteral feeding is no better than enteral feeding. Moreover, it is yet to be determined how early the full feeding should begin (immediately or within 7 days) and what constitutes best feeding practice (hypocaloric feeding at 50 to 70% of estimated energy expenditure with higher protein of at least 1.5 g/kg, versus 100% of both energy and protein needs at least for the first several weeks of critical illness). The disadvantage of full feeding in critically ill patients is the greater risk for hyperglycemia and its increased infectious complications and greater fluid

administration with its adverse consequences. Hypocaloric feeding appears to be as effective as full feeding for at least the first several weeks of critical illness. Measurement of energy expenditure by indirect calorimetry is generally reserved for the most severely marasmic patients, in whom estimates are often inaccurate, and for critically ill patients on prolonged mechanical ventilation or invasive nutritional support beyond several weeks, in whom provision of full energy needs may be beneficial.

Grade A References

A1. Casaer MP, Mesotten D, Hermans G, et al. Early versus late parentral nutrition in critically ill adults. *N Engl J Med*. 2011;365:506-517.

A2. Heidegger CP, Berger MM, Graf S, et al. Optimisation of energy provision with supplemental parenteral nutrition in critically ill patients: a randomized controlled clinical trial. *Lancet*. 2013;381:385-393.

A3. Harvey SE, Parrott F, Harrison DA, et al. Trial of the route of early nutritional support in critically ill adults. *N Engl J Med*. 2014;371:1673-1684.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Jensen GL, Mirtallo J, Compher C, et al. Adult starvation and disease-related malnutrition: a proposal for etiology-based diagnosis in the clinical practice setting from the International Consensus Guideline Committee. *JPEN J Parenter Enteral Nutr*. 2010;43:156-159.
2. van der Pols-Vijlbrief R, Wijnhoven HA, Schaap LA, et al. Determinants of protein-energy malnutrition in community-dwelling older adults: a systematic review of observational studies. *Ageing Res Rev*. 2014;18:112-131.
3. Almeida AI, Correia M, Camilo M, Ravasco P. Length of stay in surgical patients: nutritional predictive parameters revisited. *Br J Nutr*. 2013;109:322-328.
4. Van Stijn MF, Korkic-Halilovic I, Makker MS, et al. Preoperative nutrition status and postoperative outcome in elderly general surgery patients: a systematic review. *JPEN J Parenter Enteral Nutr*. 2013; 37:37-43.
5. Coltman A, Peterson S, Roehl K, et al. Use of 3 tools to assess nutrition risk in the intensive care unit. *JPEN J Par Ent Nutr*. 2015;39:28-33.
6. Steenson J, Vivanti A, Isenring E. Inter-rater reliability of the Subjective Global Assessment: a systematic literature review. *Nutrition*. 2013;29:350-352.
7. Böhm A, Heitmann BL. The use of bioelectrical impedance analysis for body composition in epidemiological studies. *Eur J Clin Nutr*. 2013;67(suppl 1):S79-S85.
8. Wilson JP, Strauss BJ, Fan B, et al. Improved 4-compartment body-composition model for a clinically accessible measure of total body protein. *Am J Clin Nutr*. 2013;97:497-504.
9. Hoffer LJ, Bistrian BR. Appropriate protein provision in critical illness: a systematic and narrative review. *Am J Clin Nutr*. 2012;96:591-600.

REVIEW QUESTIONS

1. Which of the following statements is correct regarding clinical nutritional assessment?

A. Normal values of body mass index (BMI) are higher in women than men in the second through fifth decades of life.
B. BMI values are useful to grade different levels of obesity or overweight, but not as an indicator of malnutrition.
C. The use of calipers to measure triceps skinfold is the easiest way to estimate body fat.
D. The level of hypoalbuminemia cannot be used to predict risk for morbidity and mortality in ambulatory patients.
E. Adequate feeding in the presence of systemic inflammation increases the serum albumin in malnourished patients.

Answer: C Approximately 50% of body fat is subcutaneous, so the use of skinfold calipers to define the triceps skinfold thickness (TSF) is the most practical technique to estimate body fat. (see Upper Arm Anthropometry section). The same standards of BMI (weight in kilograms divided by the height in meters squared) apply to males and females; and there are BMI standards to define malnutrition (see Body Mass Index section). Hypoalbuminemia is a strong predictor of risk for morbidity and mortality in both hospitalized and ambulatory patients (Van Stijn MF, Korkic-Halilovic I, Makker MS, et al. Preoperative nutrition status and postoperative outcome in elderly general surgery patients: a systematic review. *JPEN J Parenter Enteral Nutr.* 2013;37:37-43). Even though adequate feeding of malnourished patients who have coexisting inflammatory states benefits them in wound healing and immune function, it will not increase serum albumin levels under these conditions (see Serum Proteins section).

2. Which of the following clinical measures provides the most accurate indicator of changes in energy balance and lean body mass in seriously ill hospitalized patients?

A. Day-to-day weight changes
B. Isotope dilution
C. Dual x-ray absorptiometry
D. Total body nitrogen analysis
E. None of the above

Answer: E The gold standard for assessing lean tissue is in vivo neutron activation analysis for nitrogen, but this is and will remain a research procedure. Day-to-day weight changes in hospitalized patients generally reflect fluid balance rather than energy balance. Total body nitrogen analysis is likewise primarily a research tool. The other tests listed provide flawed estimates in acute illness (see Assessing Lean Body Mass and Total Body Water section).

3. Which of the following is *not* a risk of full feeding or overfeeding (as opposed to hypocaloric feeding) of critically ill patients?

A. Hyperviscosity
B. Hyperglycemia
C. Fluid overload
D. Nosocomial infections
E. Worse postoperative outcomes

Answer: A Hyperviscosity is not known to be associated with full feeding or overfeeding. Hyperglycemia and its attendant increase in risk for nosocomial infections are significant risks, as is fluid overload. Overfeeding should be avoided in postoperative patients who require invasive nutritional support, and modest underfeeding for the first several weeks of acute illness may actually improve outcomes in seriously ill patients (see Measures of Energy Expenditure and Caloric Need).

215

PROTEIN-ENERGY MALNUTRITION

MARK J. MANARY AND INDI TREHAN

DEFINITION

The term *protein-energy malnutrition* encompasses at least three distinct clinical syndromes. The first and most common, *stunting*, occurs throughout the developing world. It is a consequence of chronic macronutrient and micronutrient deficiency prenatally and during early childhood and is manifested as low birth weight and irreversible cognitive and physical stunting, including below normal weight and short stature in the first few years of life. In contrast to this chronic form of protein-energy malnutrition, a second manifestation takes the form of *acute malnutrition*, a primarily macronutrient deficiency. In its most severe forms, it includes kwashiorkor, marasmus (wasting), and marasmic kwashiorkor. The third syndrome is the wasting that occurs secondary to acute or chronic underlying illnesses such as renal or hepatic insufficiency, inflammatory bowel disease, malignancy, or a systemic infection such as HIV or tuberculosis.

EPIDEMIOLOGY

Global rates of stunting and acute malnutrition are difficult to quantify accurately, given the primarily rural populations where they mostly occur. Approximately 25% of children younger than 5 years are stunted worldwide, and another 8% suffer from marasmus. The number with kwashiorkor is unknown and underreported because of minimal high-quality surveillance data[1]; therapeutic feeding programs in southern Africa (the area with the highest prevalence) often report that 50 to 70% of cases of severe acute malnutrition have kwashiorkor rather than marasmus. It is estimated that some 15% of the total mortality for children younger than 5 years worldwide is attributable to stunting and that another 12% is attributable to wasting.[2]

Rates of secondary protein-energy malnutrition among those with medical and surgical illnesses vary widely and are a function of underlying disease processes, comorbidities, nutritional status before illness, and level of financial resources and social support. It is not unusual for malnutrition rates of 25 to 60% among hospitalized patients to be reported in the literature.

PATHOBIOLOGY

Most stunting in children occurs during the critical "1000 days" window between conception and 2 years of age, although there is some evidence that partial catch-up growth can occur in later childhood.[3] Maternal undernutrition contributes to low birth weight, which persists as underweight, short stature, and cognitive stunting. Stunting also places children at elevated risk for acute malnutrition when challenged by food shortages or acute infections. Even children without prenatal stunting are at high risk for stunting and acute malnutrition when raised in impoverished environments. HIV infection and exposure, diarrhea, pneumonia, measles, and malaria are common in the developing world and can lead to anorexia with decreased dietary intake, increased energy expenditure, and poor nutrient absorption, placing the child at risk for stunting and acute malnutrition. The end of exclusive breastfeeding (whether prematurely or at the recommended 6 months of age) and the introduction of complementary feeding is also a high-risk period as the child ingests a variety of environmental pathogens and often suffers a relative loss of high-quality protein, lipids, and micronutrients.

Two major pathobiologic factors have been recently appreciated to contribute significantly to the development of protein-energy malnutrition in children.[4] The first is environmental enteric dysfunction (abbreviated EED, formerly "environmental enteropathy" or "tropical enteropathy"), endemic in children and adults throughout the developing world. It is characterized by blunted and atrophied intestinal villi, hyperplasia of the crypts, and lymphocytic infiltration of the lamina propria, histologically similar in many ways to celiac disease. EED appears to be a T-cell-mediated response to repeated environmental insults, such as gastrointestinal infections with a fecal-oral transmission pattern. The net effect of EED is an increase in intestinal permeability with bacterial and toxin translocation due to loss of tight junction integrity, impaired immune functioning, and malabsorption. EED is generally subclinical, predisposing children to growth faltering, and leaves them more susceptible to developing acute malnutrition. A second, related, risk factor for the development of acute malnutrition is a disturbed configuration of the intestinal microbiome, which fails to develop in an age-appropriate manner compared with unaffected children.[5,6] The ability to clinically identify these abnormalities in the microbiome and provide specific therapy is not yet available.

Secondary protein-energy malnutrition that occurs in the context of an underlying illness often results from a triad of decreased energy intake, malabsorption, and catabolic stressors. Virtually any chronic and/or critical illness can precipitate protein-energy malnutrition, but among the most common are cancer, HIV/AIDS, tuberculosis, inflammatory bowel disease, chronic kidney disease, chronic liver disease, and rheumatologic illnesses. The patient is in a state of net negative energy balance manifested by decreased weight and metabolic rate, accompanied to varying degrees by muscle wasting, depletion of fat stores, reduced cardiorespiratory capacity, skin thinning with easy breakdown and ulceration, hypothermia, immunodeficiency with impaired wound healing, and apathy.

The specific pathobiologic etiologies and risk factors for the development of this secondary protein-energy malnutrition are numerous. Primary among them is poor dietary intake as a result of nausea, anorexia, depression, poor dentition, and oral-motor weakness accompanying the underlying illness. Even nutrients that are ingested may not be absorbed, for example, because of reduced bile salt secretion leading to steatorrhea, pancreatic insufficiency, and damage to the intestinal mucosa in Crohn's disease. Systemic inflammation and oxidative stress are common in critical illness, cirrhosis, patients with HIV/AIDS and other infections, and hemodialysis patients, contributing to a catabolic state. In patients with chronic renal disease, altered amino acid homeostasis by the kidney, resistance to growth hormone and insulin-like growth factor-1, low testosterone levels, insulin resistance, and altered insulin signaling are all important factors in protein-energy wasting.[7] Patients with chronic liver disease often suffer from protein-energy malnutrition due to a combination of altered gut motility, dyspepsia, cholestasis with poor absorption of fat-soluble vitamins, small intestine bacterial overgrowth, a hypermetabolic state, inadequate hepatic protein synthesis, lack of glycogen reserves, and blood loss from varices and the intestinal lumen.

CLINICAL MANIFESTATIONS

Stunting manifests quite simply as short stature and underweight for age. Brain development, and thus head circumference, may also be small for age, although this is relatively proportional for overall body size.

Acute malnutrition manifests in at least three different forms. Children with wasting are emaciated and weak, having suffered significant weight loss in a relatively short period. The wasting often first manifests in the axilla and groin, progressing to the thighs and buttocks, and eventually visible in the face, which may take on an "old man" appearance. These children may appear apathetic but in fact may be quite inconsolable when approached. Wasting may be mild, moderate, or severe ("marasmus") (Fig. 215-1).

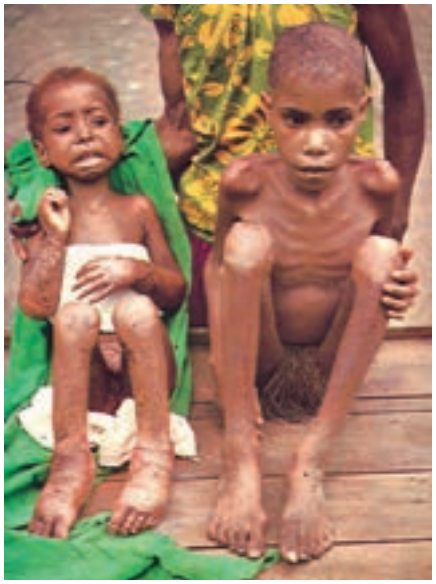

FIGURE 215-1. Kwashiorkor and marasmus in brothers. The younger brother, on the left, has kwashiorkor with generalized edema, skin changes, pale reddish yellow hair, and an unhappy expression. The older child, on the right, has marasmus, with generalized wasting, spindly arms and legs, and an apathetic expression. (From Peters W, Pasvol G, eds. *Tropical Medicine and Parasitology*, 5th ed. London: Mosby; 2002, Fig. 986.)

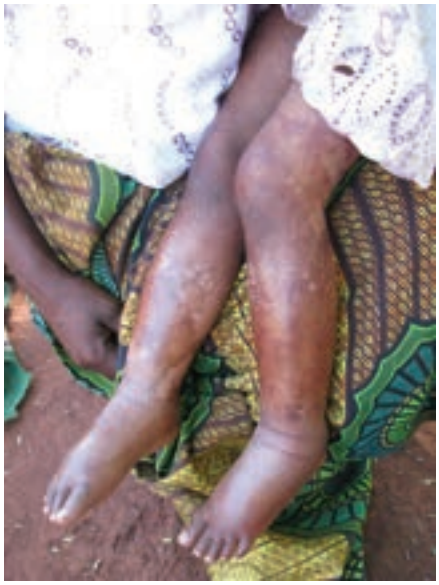

FIGURE 215-2. Edematous malnutrition or kwashiorkor.

The second form of acute malnutrition, edematous malnutrition or kwashiorkor, was classically described to occur when a child was weaned rapidly from the high-quality protein source that is breast milk, although this is not universal. Children with kwashiorkor have a weight that is generally low-to-average for their age. They present with bilateral peripheral edema that begins in the feet and progresses cephalad as it worsens. Despite what can be a relatively profound edema, they generally do not have ascites, nor is their illness a result of hepatic insufficiency or hypoalbuminemia. The skin may have patchy areas of "flaky paint" depigmentation, commonly with areas of breakdown and resultant infection (Fig. 215-2). The hair is often sparse, dry, brittle, and depigmented, appearing brown-yellow in children whose hair is normally black and thick.

Why some children in nearly identical living conditions develop marasmus while others develop kwashiorkor and most remain without acute malnutrition is not understood but does not appear to be a result of disparate protein intake; recent evidence suggests that alterations in the commensal intestinal microbiome may be responsible. Finally, children with marasmic kwashiorkor have both the severe weight loss of marasmus and the edema of kwashiorkor. Compared with children with either syndrome alone, these children are generally the most ill with the poorest prognosis.

Protein-energy malnutrition in patients with severe or chronic underlying disease generally presents as low body-mass index (BMI) or with progressive weight loss.

Virtually every organ system and tissue type is starved of energy in all forms of protein-energy malnutrition, resulting in a homeostatic drive to adapt to the decreased energy available. Fat is increasingly used as the body's primary fuel source within the first few days, having replaced glucose, and ketosis quickly develops. Overall glucose production and protein breakdown decrease markedly, decreasing urea production and urinary fluid losses. The basal metabolic rate decreases, accompanied by hypothermia and easy fatiguing. As starvation continues, nearly all of the body's fat stores are depleted, and lean muscle tissue may be cut in half. In addition to decreased muscle mass, hypokalemia leads to rapid muscle fatigue. Cardiac muscle is not spared, potentially leading to bradycardia, decreased stroke volume, hypotension, and poor tissue perfusion. Intravascular volume may be diminished at the same time that cellular and capillary leakage increases, leading to generalized edema, particularly in kwashiorkor.

Pulmonary capacity is adversely affected because of decreased respiratory muscle mass and electrolyte disturbances. The skin and hair often atrophy, depigment, and break down, leaving the patient susceptible to cutaneous infections. Insulin and thyroid hormone levels decrease, and cortisol concentrations increase. A state of immunodeficiency develops as lymphoid tissues atrophy and cell-mediated immunity is diminished, placing the malnourished patient at high risk for opportunistic infections. Pancytopenia can occur as a result of bone marrow suppression. Prolonged malnutrition leads to deterioration of all portions of the gastrointestinal system, including atrophy and blunting of the intestinal villi (thereby complicating therapeutic feeding), impaired exocrine pancreatic function, and decreased gastric and biliary secretions. Hepatomegaly and fatty liver infiltration are seen. Except in cases in which renal pathology is the inciting pathway, kidney function is relatively well preserved until late in the course. Although the brain is preserved longer than other organs, cerebral atrophy is seen in acute malnutrition, and (often permanently) delayed cognitive development is a profound complication with lifelong consequences among survivors.

DIAGNOSIS

Careful anthropometry must be conducted to accurately evaluate any individual for protein-energy malnutrition. In the case of children, it can be particularly challenging to gain their cooperation for anthropometry. Precise measurements of height (to the nearest 0.5 cm or less), weight (to the nearest 100 g or less), mid-upper arm circumference (MUAC; to the nearest 2 mm or less), and an assessment of peripheral pitting edema should be performed. Standardized World Health Organization growth charts should be used for diagnosis and classification.[8]

A child is considered to have *moderate stunting* when the height-for-age Z-score (HAZ) lies between 2 and 3 standard deviations (SD) below the mean. *Severe stunting* is diagnosed when the HAZ is more than 3 SD below the mean. Familial short stature, hypothyroidism, growth hormone deficiency, and micronutrient deficiency are all on the differential diagnosis for a stunted child, although these will be relatively rare compared with stunting in the relevant epidemiologic context.

Moderate wasting is based on having a weight-for-height Z-score (WHZ) between 2 and 3 SD below the mean. *Severe wasting*, or marasmus, is based on a WHZ more than 3 SD below the mean. For children aged 6 to 59 months, the MUAC can also be used to diagnose wasting because a healthy minimum MUAC remains relatively constant throughout this age. In these children, MUAC cutoffs of 115 to 125 mm and of less than 115 mm are generally considered diagnostic for moderate and severe wasting, respectively. Most children diagnosed as wasted by WHZ and MUAC criteria will be the same, but there is a sizeable population that will only be diagnosed by one or the other criterion. Given that the population identified by MUAC is generally younger and at higher risk for death than the population identified by WHZ,[9] MUAC is preferred for identifying severely malnourished children. Nevertheless, if resources exist to screen children by WHZ in addition to MUAC, that strategy is most likely to identify all acutely wasted children. Congenital heart disease, severe diarrhea and dehydration, malabsorptive syndromes due to intestinal parasites, malaria, HIV/AIDS, and tuberculosis should be considered in the differential diagnosis of a child presenting with wasting. Still, if a child has wasting based on anthropometric criteria,

nutritional rehabilitation needs to be provided while their underlying illness is addressed simultaneously.

The presence of edema in the appropriate clinical context, regardless of other anthropometric parameters, should prompt serious consideration of the diagnosis of kwashiorkor. Edema is most easily detected on the dorsum of the feet. The degree of edema is graded as 1+ if confined to the lower extremities, 2+ if additionally present on the upper extremities, and 3+ if extending to the face. The usual physiologic causes of edema, including underlying cardiac, hepatic, and renal diseases, should be considered in the differential diagnosis. In the rural impoverished populations where kwashiorkor is found, routine health care is also generally limited; thus, the possibilities of congenital or rheumatic heart disease, acute proliferative glomerulonephritis (postinfectious or post-streptococcal), profound anemia (from primary iron deficiency, severe malaria, or hookworm, among other causes), and tuberculosis should be considered. Nevertheless, for the overwhelming majority of children with edema presenting for care in these populations, kwashiorkor remains the leading diagnosis (E-Fig. 215-1).

Secondary malnutrition due to underlying illness also requires careful anthropometry, with a BMI of less than 18.5 kg/m^2 representing moderate malnutrition and a BMI of less than 15 kg/m^2 representing severe protein-energy malnutrition. Even without a BMI this low, any significant weight loss, especially if associated with lean muscle loss in addition to depletion of fat stores, should prompt consideration of protein-energy malnutrition and necessitates an investigation into an underlying illness if one had not been previously identified. No reliable diagnostic tests are available to identify those with protein-energy malnutrition because serum markers such as albumin, prealbumin, and C-reactive protein are themselves acute phase reactants and nonspecific for this purpose.

TREATMENT

Rx

Studies of probiotics and antibiotics to treat environmental enteric dysfunction have not shown benefit. However, albendazole, zinc, and micronutrient supplementation have shown some benefit.

There is relatively little that can be done for a child with stunting because the physical and cognitive growth faltering suffered in the first "1000 days" is not likely to be amenable to therapy. These children remain at high risk for further stunting during childhood, although some degree of catch-up growth is possible in later childhood and adolescence, even without specific nutritional interventions. Attention should be directed to an overall improvement in their dietary quality and diversity, particularly with respect to increased protein intake. Exclusive breast-feeding until 6 months of age and continued breast-feeding with appropriate supplementary foods until at least 2 years of age should be encouraged whenever possible, except in the case of an HIV-infected mother. Routine health care, including immunizations, vitamin A supplementation, and periodic deworming, should be ensured. In high-prevalence settings, HIV testing and treatment should be sought. Improvements in sanitary living conditions are likely to be the most beneficial.

Children with severe acute malnutrition have traditionally been managed as inpatients with fortified milk-based formulas as the key nutritional component of their overall care. A coordinated 10-step plan for the inpatient management of severe acute malnutrition has led to tremendous improvements in recovery and mortality rates (Fig. 215-3). Children are initially started on F-75 formula every 2 to 4 hours and then progress to F-100 as their appetite increases and they demonstrate that they are able to tolerate the solute load without severe diarrhea (Table 215-1).

However, the development and increasingly widespread availability of ready-to-use therapeutic food (RUTF), most often a fortified peanut paste, makes it preferable to treat children in community-based outpatient feeding programs instead, assuming that the child demonstrates a mental status and appetite conducive to feeding at home (see E-Fig. 215-1).[10] After ensuring that the child has an appetite and is not suffering from complications such as hypoglycemia, severe dehydration, respiratory distress, severe anemia, and high fevers, a test feeding of approximately 30 g RUTF is given under directly observed therapy. Most children will successfully complete this test feeding and can be discharged home with 1 to 2 weeks of 175 kcal/kg/day of RUTF. Anthropometry and clinical assessments are repeated at 1- to 2-week intervals; children are treated until their edema resolves and either their WHZ is no more than 2 SD below the mean or their MUAC reaches 125 mm (depending on whether they were diagnosed based on WHZ or MUAC criteria). Almost all children will recover within 8 to 12 weeks, with approximately half recovering by 4 to 6 weeks if the child is receiving all of the intended food and there are no interim complications. Any child not improving as expected should have a thorough social and medical assessment to evaluate whether they are being fed appropriately and whether any acute infectious complications have developed; inpatient care may be necessary in those situations.

Mortality and nutritional recovery outcomes from community-based programs are generally superior to those achieved by inpatient care, making this the current international standard of care[11] and relegating inpatient care only to those children with anorexia or medical complications or who are in an area where RUTF is not available. Outpatient care can be improved further with empirical antibiotic therapy for all patients because of the high risk for overwhelming sepsis and death. Integration of outpatient nutritional care into a complete package of routine health interventions, including linkage to HIV testing and treatment, will likely only improve outcomes further.

Patients with secondary protein-energy malnutrition should, first and foremost, have their underlying illness addressed because this is most likely to lead to long-term recovery of their nutritional status. Concomitantly, nutritional rehabilitation should be undertaken to prevent further energy losses and allow for recovery of damaged organic pathways. Fluid, electrolyte, and acid-base status should be corrected cautiously in the usual fashion. In general, enteral nutrition (Chapter 216) is preferred over parenteral, assuming the gastrointestinal tract is functioning adequately. Frequent small feedings or slow drip feedings may be necessary. Aggressive tube feeding or parenteral nutrition should be avoided in the early stages of rehabilitation because the

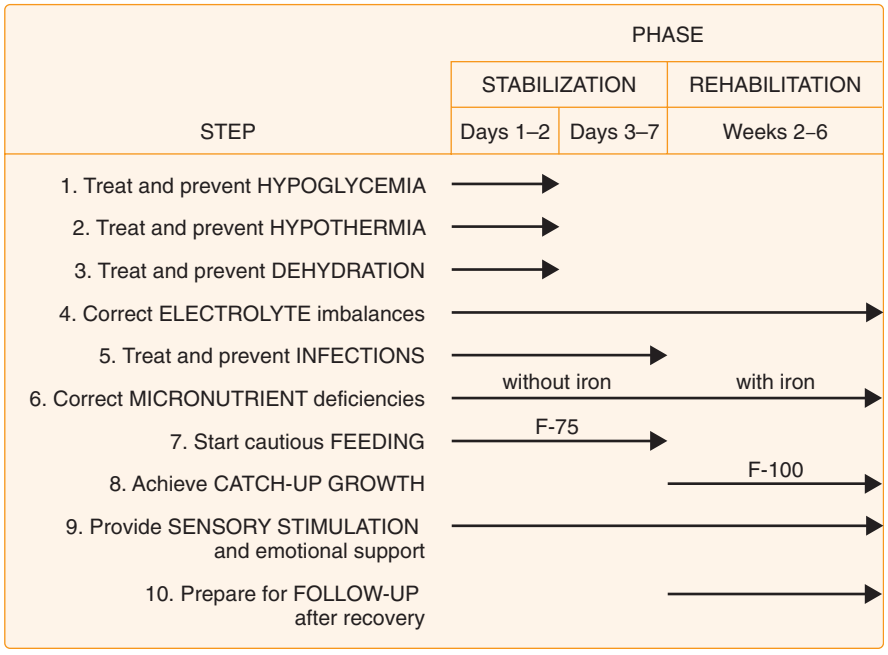

FIGURE 215-3. Ten-step inpatient management protocol for severe acute malnutrition. (Adapted from Ashworth A, Khanum S, Jackson A, Schofield C. *Guidelines for the Inpatient Treatment of Severely Malnourished Children.* Geneva: World Health Organization; 2003.)

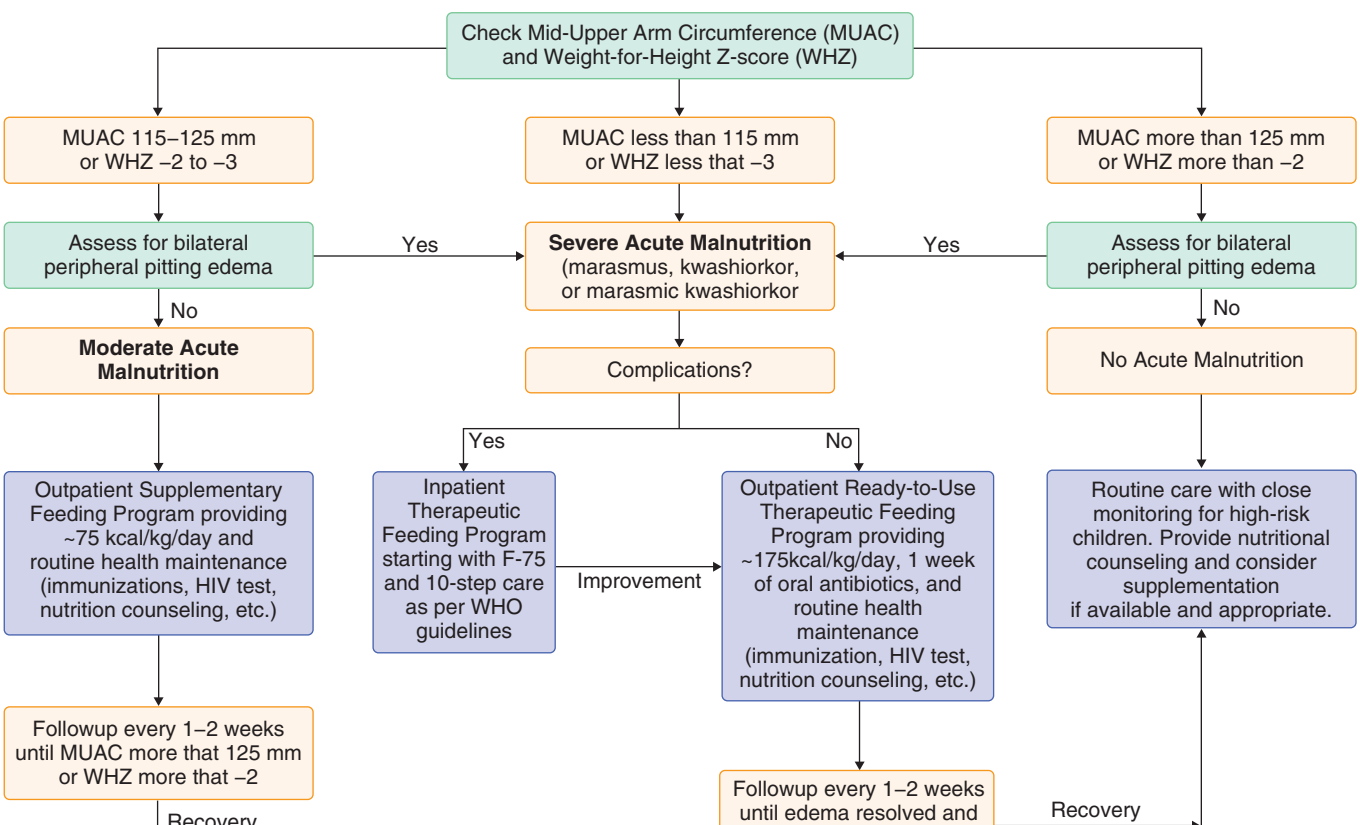

E-FIGURE 215-1. Diagnostic and treatment algorithm for acute malnutrition in children. MUAC only to be used for children 6 to 59 months of age. See World Health Organization (WHO). *WHO Child Growth Standards and the Identification of Severe Acute Malnutrition in Infants and Children.* Geneva: WHO, United Nations Children's Fund; 2009 for anthropometry details. See Ashworth A, Khanum S, Jackson A, Schofield C. *Guidelines for the Inpatient Treatment of Severely Malnourished Children.* Geneva: World Health Organization; 2003, Fig. 3, for 10-step inpatient therapy details. See World Health Organization (WHO). *Community-based Management of Severe Acute Malnutrition.* Geneva: World Health Organization, World Food Programme, United Nations System Standing Committee on Nutrition, United Nations Children's Fund; 2007, for outpatient therapy details.

TABLE 215-1 NUTRITIONAL COMPOSITION OF THERAPEUTIC FOODS FOR SEVERE ACUTE MALNUTRITION

	F-75 (100 mL)	F-100 (100 mL)	RUTF (100 mg)
Energy (kcal)	75	100	543
Protein (g)	0.9	2.9	13.6
Lactose (g)	1.3	4.2	
Potassium (mg)	156	246	1111
Sodium (mg)	14	44	189
Magnesium (mg)	10.5	17.7	92
Zinc (mg)	2	2.3	14
Copper (mg)	0.25	0.25	1.78
Osmolarity (mOsm/L)	413	419	
% of total energy from protein	5	12	10-12
% of total energy from fat	36	53	45-60

RUTF = ready-to-use therapeutic food.

refeeding syndrome is a real danger in patients who are malnourished. Hypophosphatemia, hypokalemia, hypomagnesemia, hyperglycemia, fluid overload, muscular weakness, cardiac arrhythmias, and diarrhea are all possible with aggressive refeeding. Fluid status, glycemic status, and electrolyte levels should be monitored closely throughout the refeeding process.[12]

PREVENTION

Prevention of stunting and acute malnutrition remains one of the most challenging and elusive goals in global health, especially given the intergenerational effects of maternal malnutrition. Providing pregnant women with nutritional supplementation, intermittent malaria treatment, timely care for sexually transmitted infections, and comprehensive prenatal care shows some success in improving birth weights and decreasing premature delivery. Adherence to recommendations for exclusive breast-feeding for the first 6 months and continued breast-feeding for at least the first 2 years also shows profound benefit.[13] Early diagnosis and treatment of HIV in mothers and their infants is also beneficial.[14] Nevertheless, children in the developing world remain at high risk for stunting and acute malnutrition because of the relative lack of sanitation, high rates of food insecurity, spotty vaccination coverage, and limited access to medical care. Addressing these underlying societal factors may provide the most benefit to decreasing environmental enteric dysfunction, stunting, and acute malnutrition.

Close attention to the nutritional status of medical and surgical patients, with risk stratification based on BMI and weight trajectory, is necessary to limit the risk for protein-energy malnutrition in these patients as they are treated for their underlying illnesses. Optimizing nutritional status will aid in recovery from their primary conditions, and similarly recovery from the primary condition will improve benefit to overall nutritional status. Efforts to empirically provide extra protein, calories, and micronutrients during times of particular oxidative stress and catabolism will help limit weight loss and its adverse effects.

PROGNOSIS

The prognosis for stunted children generally remains poor because the physical and cognitive deficits they suffer are mostly carried with them throughout life,[15] despite some catch-up growth in later childhood and adolescence. Children who recover from an episode of severe acute malnutrition are somewhat more susceptible to further episodes over the next several months than their peers, although those with underlying illnesses such as HIV or tuberculosis remain at significantly elevated risk. The highest recovery rates in severe acute malnutrition treatment programs are approximately 90%, with about 4 to 5% mortality. Untreated, episodes of severe wasting carry an estimated 10%-20% mortality risk per month. Although many children will recover spontaneously, most will never return to their baseline nutritional status and fully thrive. Early childhood growth faltering may be linked to an increased risk for the metabolic syndrome as adults, contributing to the double epidemic of childhood undernutrition and adult obesity being increasingly observed throughout the developing world.

Grade A References

A1. Trehan I, Shulman RJ, Ou CN, et al. A randomized, double-blind, placebo-controlled trial of rifaximin, a nonabsorbable antibiotic, in the treatment of tropical enteropathy. *Am J Gastroenterol*. 2009;104:2326-2333.

A2. Ryan KN, Stephenson KB, Trehan I, et al. Oral zinc or albendazole ameliorates environmental enteropathy in Malawian children: a randomized controlled trial. *Clin Gastroenterol Hepatol*. 2014;12:1507-1513.

A3. Smith HE, Ryan KN, Stephenson KB, et al. Multiple micronutrient supplementation transiently ameliorates environmental enteropathy in Malawian children aged 12-35 months in a randomized controlled clinical trial. *J Nutr*. 2014;144:2059-2065.

A4. Trehan I, Goldbach HS, LaGrone LN, et al. Antibiotics as part of the management of severe acute malnutrition. *N Engl J Med*. 2013;368:425-435.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Briend A, Collins S, Golden M, et al. Maternal and child nutrition. *Lancet*. 2013;382:1549.
2. Black RE, Victora CG, Walker SP, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. *Lancet*. 2013;382:427-451.
3. Prentice AM, Ward KA, Goldberg GR, et al. Critical windows for nutritional interventions against stunting. *Am J Clin Nutr*. 2013;97:911-918.
4. Jones KD, Thitiri J, Ngari M, et al. Childhood malnutrition: toward an understanding of infections, inflammation, and antimicrobials. *Food Nutr Bull*. 2014;35:S64-S70.
5. Smith MI, Yatsunenko T, Manary MJ, et al. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science*. 2013;339:548-554.
6. Subramanian S, Huq S, Yatsunenko T, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature*. 2014;510:417-421.
7. Ruperto M, Sanchez-Muniz FJ, Barril G. A clinical approach to the nutritional care process in protein-energy wasting hemodialysis patients. *Nutr Hosp*. 2014;29:735-750.
8. World Health Organization (WHO). WHO Child Growth Standards and the Identification of Severe Acute Malnutrition in Infants and Children. Geneva: WHO, United Nations Children's Fund; 2009.
9. Briend A, Maire B, Fontaine O, et al. Mid-upper arm circumference and weight-for-height to identify high-risk malnourished under-five children. *Matern Child Nutr*. 2012;8:130-133.
10. Park SE, Kim S, Ouma C, et al. Community management of acute malnutrition in the developing world. *Pediatr Gastroenterol Hepatol Nutr*. 2012;15:210-219.
11. World Health Organization (WHO). Guideline: Updates on the management of severe acute malnutrition in infants and children. Geneva: World Health Organization; 2013.
12. Ridley E, Gantner D, Pellegrino V, et al. Nutrition therapy in critically ill patients—a review of current evidence for clinicians. *Clin Nutr*. 2015;[Epub ahead of print].
13. Kamudoni P, Maleta K, Shi Z, et al. Exclusive breastfeeding duration during the first 6 months of life is positively associated with length-for-age among infants 6-12 months old, in Mangochi district, Malawi. *Eur J Clin Nutr*. 2015;69:96-101.
14. Trehan I, O'Hare BA, Phiri A, et al. Challenges in the management of HIV-infected malnourished children in sub-Saharan Africa. *AIDS Res Treat*. 2012;2012:790786.
15. Adair LS, Fall CH, Osmond C, et al. Associations of linear growth and relative weight gain during early life with adult health and human capital in countries of low and middle income: findings from five birth cohort studies. *Lancet*. 2013;382:525-534.

REVIEW QUESTIONS

1. Which one of the following is *not* considered a form of protein-energy malnutrition?

A. Stunting
B. Beriberi
C. Marasmus
E. Kwashiorkor
E. HIV wasting

Answer: B Beriberi occurs as a result of vitamin B_1 (thiamine) deficiency and may be due to inadequate intake, malabsorption in the gastrointestinal tract, alcoholism, dialysis, genetic disorders, and other causes (Chapter 218). Although beriberi may be present in patients with protein-energy malnutrition, it is a micronutrient deficiency and would be part of a more global malnutrition syndrome in these patients.

2. Which of the following is the most common underlying factor in childhood mortality worldwide?

A. Malaria
B. HIV
C. Intestinal parasites
D. Malnutrition
E. Premature birth

Answer: D Malnutrition, in all of its various forms, is an underlying factor in about 45% of all deaths in children younger than 5 years worldwide.[2] Although malaria, HIV, intestinal parasites, and prematurity remain extremely common and are a cause of significant mortality and morbidity, malnutrition continues to exceed all of these as a contributing factor to childhood mortality. Malaria accounts for approximately 7% of all childhood deaths, HIV 2%, and prematurity 14%, although each of these numbers would be lower if not for the malnutrition suffered by the mothers and children in question. Intestinal parasites rarely lead directly to mortality, but instead contribute greatly to morbidity, including malabsorptive syndromes and iron deficiency anemia.

3. Which of the following interventions is most likely to be effective in reducing the burden of environmental enteric dysfunction (EED)?

A. Ensuring all children receive their routine immunizations
B. Antibiotic prophylaxis
C. Routine probiotic administration
D. Zinc supplementation
E. Albendazole for deworming
F. Improved sanitation

Answer: F EED is most likely a result of repeated pernicious insults to the intestinal tract that are found most often in developing settings, influenced to a large degree by the lack of sanitation present there in impoverished areas (Korpe PS, Petri WA Jr. Environmental enteropathy: critical implications of a poorly understood condition. *Trends Mol Med.* 2012;18:328-336). Immunizations are unlikely to have an effect on EED because most do not target enteric pathogens, and none target intestinal bacterial or parasitic pathogens. Antibiotics and probiotics have not shown benefit in rigorous clinical prospective clinical trials (Trehan I, Shulman RJ, Ou CN, et al. A randomized, double-blind, placebo-controlled trial of rifaximin, a nonabsorbable antibiotic, in the treatment of tropical enteropathy. *Am J Gastroenterol.* 2009;104:2326-2333). Zinc and albendazole have recently shown modest benefits (Ryan KN, Stephenson KB, Trehan I, et al. Oral zinc or albendazole ameliorates environmental enteropathy in Malawian children: a randomized controlled trial. *Clin Gastroenterol Hepatol.* 2014;12:1507-1513), but these are relatively minimal compared with the degree of benefit that would be expected from modern sanitation systems.

4. Nutritional supply and function of which of the following organs are preserved the longest and most in protein-energy malnutrition?

A. Brain
B. Heart
C. Intestine
D. Liver
E. Lung
F. Skin

Answer: A Nutrient supply to the brain is preserved the longest in patients with protein-energy malnutrition, whereas the other organs must adapt to the decreased nutrients available to them. Nevertheless, some degree of cerebral atrophy is seen, with often permanent irreversible cognitive stunting and developmental delay for children who suffer protein-energy malnutrition.

5. Which of the following is the most useful anthropometric measure for diagnosing acute wasting in a 3-year-old child?

A. Height
B. Height-for-age Z-score
C. Mid-upper arm circumference
D. Weight
E. Weight-for-age Z-score
F. Weight-for-height Z-score

Answer: C The two parameters that are used for diagnosing wasting in all patients are mid-upper arm circumference (MUAC) and weight-for-height Z-score (WHZ). Although WHZ has traditionally been the preferred method for diagnosing acute wasting, MUAC has in recent years proved to be a better predictor of future mortality than WHZ and is now the preferred anthropometric criterion for diagnosing acute wasting. If resources are available to assess both criteria in any given health system, then both should be performed, but this may be tedious, time-consuming, and more expensive (Briend A, Maire B, Fontaine O, Garenne M. Mid-upper arm circumference and weight-for-height to identify high-risk malnourished under-five children. *Matern Child Nutr.* 2012;8:130-133). Height-for-age Z-score and weight-for-age Z-score are useful for diagnosing stunting and underweight, respectively, which are both markers of chronic malnutrition. Height and weight alone do not provide any specific information because there is wide genetic variability in each individual's potential height and weight, and because this is also a function of gender and age.

216 ENTERAL NUTRITION

STEPHEN A. MCCLAVE

The benefit of nutrition therapy in the hospitalized patient is clearly related to the provision of early enteral nutrition (EN).[1] Outside the setting of intestinal failure or short bowel syndrome, a true benefit over providing parenteral nutrition (PN) is not clear, and its use should be determined on a case-by-case basis. Providing early EN to the critically ill patient in the intensive care unit (ICU), on the other hand, exerts a beneficial physiologic effect that downregulates inflammation and improves patient outcome.[2] Because the gut is the largest immune organ in the body, the timing, volume, and content of nutrient that is infused into the lumen of the gut have a major impact on the level of oxidative stress, the tone of systemic immune responses, and the likelihood for complications. A window of opportunity exists shortly after admission to the ICU, during which placement of enteral access and initiation of feeding will improve patient outcome by reducing rates of infection, overall complications, organ failure, length of hospital stay, and, in some disease processes, even mortality.[3] Failure to utilize the gut in critical illness results in a different physiologic response, whereby the gut becomes a pro-inflammatory organ and outcome is worsened. Depending on the patient population, the window of opportunity for such modulation may be very narrow. Burns are considered "nutritional emergencies," in which the window of opportunity may be as short as 3 to 6 hours before increased permeability occurs. In other disease processes, such as pancreatitis or major elective operations, the window may be longer (i.e., 48 to 72 hours). At some point, the ability to modulate the immune system and contain gut barrier function begins to diminish. Preventing increases in permeability of the gut is easier than containment after barrier function is compromised. The main purpose of early EN in the hospitalized critically ill patient then is to maintain gut integrity and modulate systemic immunity. With time, provision of nutrients to prevent the deterioration of nutritional status and the development of malnutrition becomes increasingly important. Therefore, EN is indicated in any critically ill patient who is unable to eat and should be initiated as soon as the patient receives full volume resuscitation. If the patient is not critically ill, EN is indicated if the patient is anticipated to be unable to eat for more than 7 days. The only true contraindications to feeding are bowel ischemia, mechanical obstruction of the gut, and peritonitis. It is important for the clinician to realize that although ileus, pancreatitis, nausea, vomiting, and other such factors may make enteral feeding difficult, none is an absolute contraindication to providing EN.

PHYSIOLOGIC BENEFIT OF ENTERAL NUTRITION

The provision of EN maintains both the functional and structural integrity of the intestinal epithelium. Early EN stimulates intestinal contractility and

TABLE 216-1 COMPARISON OF PHYSIOLOGIC EFFECTS OF FEEDING VERSUS STARVATION

	ENTERAL FEEDING	STARVATION
Intestinal contractility	Increased	Decreased
Release of secretory IgA from gut	Increased	Decreased
Release of trophic agents (bile salts, gastrin)	Increased	Decreased
Splanchnic intestinal blood flow	Increased	Decreased
Luminal microbiota Organisms Population	 Commensal Normal	 Pathogenic Overgrowth
Gut integrity and epithelial cell tight junctions	Intact	Permeable
Mass of GALT	Maintained	Diminished
Support of MALT at distant sites	Maintained	Diminished
Population of CD_4 helper lymphocytes emerging from gut into systemic circulation	T_H2 (anti-inflammatory)	T_H1 (pro-inflammatory)
Expression of adhesion molecules	MAdCAM	E-selectin
Cell line affected (pass out of vascular space)	GALT cells	Neutrophils

GALT = gut-associated lymphoid tissue; IgA = immunoglobulin A; MALT = mucosal-associated lymphoid tissue; MAdCAM = mucosal addressin adhesion molecule.

causes release of trophic substances such as bile salts, gastrin, bombesin, and motilin. Good contractility controls the overall number of bacteria within the lumen of the gut by peristaltic movement of organisms downstream. EN stimulates the release of secretory immunoglobulin A (IgA), which helps coat luminal bacteria and prevents their adherence to the epithelial wall (Table 216-1). EN stimulates blood flow to the gut and supports the mass of gut-associated lymphoid tissue (GALT). EN stimulates the production of T_H2 anti-inflammatory CD4 helper lymphocytes, which enter the systemic circulation and support immune function. Secretory immunoglobulin A (IgA)-producing immunocytes form mucosal-associated lymphoid tissue (MALT), which generates further production of secretory IgA. EN promotes the role of commensal bacteria, which provide *direct protection* by degrading bacterial toxins and *indirect protection* by preventing colonization of pathogenic organisms (such as *Pseudomonas aeruginosa*). By fermenting prebiotic fiber to short-chain fatty acids, commensal bacteria stimulate butyrate receptors in the colon, which further downregulates inflammation. EN also simulates the immune process of oral tolerance, which further supports the anti-inflammatory functions of T_H1 and T_H3 lymphocytes and the production of the anti-inflammatory agent transforming growth factor-β (TGF-β). The T_H2 cytokines that are produced in this process suppress the adhesion molecule E-selectin, which functions to trap neutrophils within the vasculature. T_H2 cytokines also stimulate the release of mucosal addressin cell adhesion molecule-1 (MAdCAM), which allows GALT cells to mobilize from the splanchnic circulation and return to the lamina propria of the gut.

In specific patient populations, the addition of certain pharmaconutrient agents to an enteral formula exerts a synergistic effect to the benefit already induced by standard enteral formula alone. The addition of arginine has a direct stimulant effect on immune function and thus may further support the T_H2 response, proliferation of lymphocytes, and secretory IgA-producing immunocytes created by the delivery of enteral feeding. The addition of antioxidants, particularly selenium, serves to decrease oxidative stress. Providing fat in the form of fish oil reduces stimulation of the inflammatory ligand toll-like receptor-4 (TLR-4) on macrophages, neutrophils, and adipocytes, reducing the production and generation of NFκ-B and tumor necrosis factor-β. Provision of fish oil generates alternative prostaglandins (PGE_3), leukotrienes (LTB_5), and thromboxane (TXA_3), which have a downregulatory effect on inflammatory mediators compared with the corresponding products of omega-6 fatty acids (PGE_2, LTB_4, TXA_2). Glutamine added to an enteral formula serves as an antioxidant and helps maintain gut integrity. Zinc has a direct effect on the zona occludens, maintaining tight junctions between the gut epithelial cells.

CONSEQUENCES OF NOT PROVIDING ENTERAL NUTRITION

For the critically ill patient, failure to provide enteral feeding generates a different physiologic response that is pro-inflammatory and is associated with worsened outcome. Increases in gut permeability allow activation of macrophages and neutrophils and upregulation of the innate immune responses. Engagement of the luminal bacteria by adherence to the gut epithelium with loss of barrier function adversely affects the acquired immune response, with stimulation and proliferation of T_H1 lymphocytes. Reduced contractility of the gut promotes bacterial overgrowth. Bacteria express virulence genes, which allow them to adhere to the epithelium. A contact-dependent activation of the epithelial cells results in the release of cytokines into the lymphatic channels. A gut-lung axis of inflammation is initiated, in which cytokines pass through the lymphatic channels and the thoracic duct to the systemic circulation and the capillary system of the lungs, promoting acute respiratory distress syndrome (ARDS) (Chapter 104) and pneumonia. The production of T_H1 cytokines stimulates the adhesion molecule E-selectin, which allows extravasation of activated neutrophils out of the vascular space and into the pulmonary alveoli. The same cytokines suppress the release of MAdCAM, which effectively traps GALT cells in the vascular space, preventing their return to the intestinal lamina propria. With the increases in gut permeability and the upregulation of immune responses, the gut becomes a pro-inflammatory organ and contributes its own component to the systemic inflammatory response syndrome (SIRS).

INITIATING ENTERAL FEEDING

In most cases, enteral access may be achieved quickly and easily by placing a nasogastric tube and initiating EN immediately after full volume resuscitation and attainment of hemodynamic stability. Although orogastric feeding may reduce the incidence of sinusitis compared with nasogastric feeding, it may be more difficult to secure a tube adequately that is protruding from the patient's mouth. Surprisingly, more than 95% of critically ill patients will tolerate gastric feeding despite risk for ileus and aspiration. Compared with small bowel feeding, gastric feeding tends to be initiated almost one full day sooner. Small bowel (postpyloric enteral) feeding quickly "catches up" such that the overall mean provision of calories and time to advancement to goal calories are the same between the two levels of feeding. Small bowel feeding does reduce the risk for aspiration but may not significantly reduce the incidence of pneumonia. Small bowel feeds are perceived to be tolerated better because residual volumes are lower compared with gastric feeding. Percutaneous access is required only in those patients anticipated to require tube feeding for longer than 4 weeks.

With initiation of feeding, it often is important to advance to goal as quickly as possible (within 24 to 36 hours).[4] For most patients, the goal of feeding is set by protein and calorie requirements. Accurate determination of caloric requirements helps avoid the danger of overfeeding. The caloric goal can be determined by weight-based equations, such as 25 kcal/kg/day, or by measuring requirements with indirect calorimetry. Protein requirements are best estimated again by weight-based equations, such as 1.2 to 1.5 g/kg/day, or by calculating nitrogen balance using measurement of urinary urea nitrogen excreted in 24 hours (protein requirement = 1 g urine urea nitrogen × 6.25 g protein/1 g N).

There are three clinical scenarios in which it may be appropriate to "permissively" underfeed the hospitalized patient. In obese critically ill patients, meeting protein requirements (estimated at 2.0 to 2.5 g/kg ideal body weight per day) while providing only 60 to 70% of caloric requirements may help mobilize fat stores while maintaining lean body mass.[5] In any patient placed on PN, providing 80% of caloric requirements over the first week of hospitalization helps avoid overfeeding, improves insulin sensitivity, and may increase the chance for an outcome benefit from the PN. In patient with ARDS, trophic feeds providing only 25% of caloric requirements over the first 6 days of hospitalization has been shown to have similar outcomes as full feeds.

However, patients at high nutritional risk should receive feeds as close to goal as possible. Nutrition risk is determined both by evidence of malnutrition (low body mass index, weight loss, or reduced nutrient intake before admission) and disease severity. Nutrition risk may be determined objectively by assessment tools such as the Nutrition Risk Score (NRS) 2002 and the Nutric Score.[6] Patients at high nutritional risk (NRS 2002 score ≥5 or Nutric Score ≥6) should receive at least 80% of goal feeds to obtain optimal benefit (and lowest mortality) from nutrition therapy (E-Table 216-1).

E-TABLE 216-1 NUTRITION ASSESSMENT SCORING SYSTEMS USED TO DETERMINE NUTRITION RISK

NRS 2002: FACTORS USED TO DETERMINE SCORE

IMPAIRED NUTRITIONAL STATUS		SEVERITY OF DISEASE	
Absent score 0	Normal nutritional status	Absent Score 0	Normal nutritional requirements
Mild score 1	Wt loss > 5% in 3 mo *Or* Food intake below 50-75% of normal requirement in preceding week	Mild Score 1	Hip fracture Chronic patients in particular with acute complications: cirrhosis, COPD *Chronic hemodialysis, diabetes, oncology*
Moderate score 2	Wt loss >5% in 2 mo *Or* BMI 18.5-20.5 + impaired general condition *Or* Food intake 25-50% of normal requirement in preceding week	Moderate Score 2	Major abdominal surgery, stroke *Severe pneumonia, hematologic malignancy*
Severe score 3	Wt loss >5% in 1 mo (15% in 3 mo) *Or* BMI <18.5 + impaired general condition *Or* Food intake <25% of normal requirement in preceding week	Severe Score 3	Head injury Bone marrow transplantation *Intensive care patients (APACHE II ≥10)*

Note: If age ≥70 years, add 1 point.
Total score = (points for nutritional status) + (points for disease severity) + (points for age).

NUTRIC SCORE: FACTORS USED TO DETERMINE SCORE

	NUTRIC POINTS			
FACTORS	***0***	***1***	***2***	***3***
Age (yr)	<50	50-74	≥75	—
APACHE II score	<15	15-19	20-27	≥28
Baseline SOFA score	<6	6-9	≥10	—
No. of comorbidities	0-1	≥2	—	—
Days in hospital to ICU admit	0	≥1	—	—
Interleukin-6 (μ/mL)	0-399	≥400	—	—

Total score = (total from six separate factors).

APACHE = Acute Physiologic and Chronic Health Evaluation; BMI = body mass index; COPD = chronic obstructive pulmonary disease; LOS = length of stay; SOFA = simplified organ failure assessment; Wt = weight.

Formula selection has been simplified with the advent of pharmaconutrient formulas. The initial decision in any critically ill patient concerns whether the individual is a candidate for an immune-modulating formula. Patients expected to benefit from an arginine-containing pharmaconutrient formula (compared with a standard formula) are those who require major elective gastrointestinal surgery (esophagectomy, gastrectomy, or pancreatectomy), patients who have experienced major trauma (with an Abdominal Trauma Index Score >20, especially with head injury), burn patients (total body surface area >30%), and patients with head and neck cancer.[■] Critically ill patients who are septic or on mechanical ventilation have not been shown to benefit from use of an arginine-containing pharmaconutrient formula or from supplemental glutamine, omega-3 fatty acids, selenium, or antioxidants.[■] As a result, a standard enteral formula providing 1.0 to 1.5 kcal/mL should be selected for most patients. The only other type of clinical scenario in which a specialty formula is required is when the patient demonstrates evidence of malassimilation (either maldigestion from pancreatic disease or malabsorption from compromise of small bowel absorption). In these patients, a formula composed of small peptides and medium-chain triglyceride oil or a fiber-containing formula should be initiated. Specialty formulas designed for specific organ failure are rarely indicated, owing to their increased expense and the fact that their use has not been shown to alter patient outcome.

After feeds are initiated, the clinician must assess tolerance of EN (Table 216-2). Patients should be assessed to make sure they have achieved adequate volume resuscitation before initiation. Evaluating gut motility is important in selecting the proper tube, the level of feeding within the gastrointestinal tract, and whether there is need for simultaneous gastric decompression while feeding into the small bowel. Adequate gastric contractility is indicated by a nasogastric output of less than 1200 mL/day. Small bowel contractility may be assessed by bowel distention, bowel sounds on physical examination, and air-fluid levels on abdominal radiograph. Colonic contractility is evaluated by passage of stool and flatus. In the absence of risk for ischemia (such as a patient who is hypotensive on pressor agents), ileus is not a contraindication to feeding. In fact, feeding will stimulate promotility agents such as bombesin and motilin. Clinicians should be encouraged to "feed an ileus." Rapid ramp-up in the rate of feeding is tolerated better than slow ramp-up, and enteral feeding protocols should be in place to prevent inappropriate cessation of feeding. The head of the patient's bed should be elevated 35 to 45 degrees, and oral hygiene with chlorhexidine mouthwash twice a day should be used to decrease the bacterial count in oropharyngeal secretions.

TABLE 216-2 ENTERAL NUTRITION (PROTOCOL FOR NASOGASTRIC FEEDING

1. Elevate head of bed 30 to 45 degrees at all times.
2. Scrutinize and correct electrolyte abnormalities (esp. K^+, Ca^{2+}, Mg^{2+}, and phosphorus).
3. Initiate proton pump inhibitor intravenously every 8 hours.
4. Place nasal bridle.
   - Place 12-French nasogastric tube into stomach.
   - Secure tube to bridle.
   - Confirm position by abdominal radiograph.
5. Initiate enteral nutrition (EN) feeds with small peptide/medium-chain triglycerides (MCT) oil formula full strength at 25 mL/hr.
   - Advance by 25 mL/hr every 12 hours as tolerated to goal.
   - State goal feeds: _____ kcal/day, infused at final rate _____ mL/hr.
6. Administer chlorhexidine mouthwash with good oral hygiene nursing care twice daily.
7. Check gastric residual volume (GRV) every 4 hours.
   - Return all contents <500 mL to the patient.
8. If GRV > 400 mL, initiate the following:
   - Continue EN at the current rate.
   - Turn patient to right lateral decubitus position if possible for 30 minutes.
   - Begin metoclopramide 10 mg IV every 6 hours (if patient is receiving opioid narcotics).
   - Begin naloxone, 8 mg in 10 mL saline per tube every 6 hours.
   - Recheck GRV in 4 hours.
9. Only if second GRV 4 hours later is >400 mL, hold EN.
   - Recheck GRV every 2 hours and restart EN when GRV is <400 mL.
   - If no other signs of intolerance, restart at same rate.
   - If other evidence of intolerance is present, consider reducing rate by 25 mL/hr when GRV <400 mL (or to baseline 25 mL/hr).
10. If tube in small bowel and GRV > 50 mL, recheck position of tube by abdominal radiograph. Consider switching to aspirate/feed nasojejunal tube.

Physicians should have a low threshold for initiating prokinetic agents, should monitor laboratory tests to avoid the adverse effect of electrolyte abnormalities on motility, and in the future may use narcotic antagonists (a nonabsorbable methylnaltrexone or an intravenous μ-receptor antagonist, alvimopam) infused through the feeding tube to reverse the effects of any opioid narcotic in order to promote intestinal contractility. Promotility agents such as metoclopramide or erythromycin should be used with caution because they prolong QT intervals and thus may precipitate dysrhythmias.

Gastric residual volume (GRV) is a poor marker of gastric emptying and risk for aspiration. Raising the GRV cutoff level for cessation of feeds does not lead to increased aspiration, and lowering the number does not protect the patient from aspiration or pneumonia. Enteral feeding protocols should set the cutoff value for GRV somewhere between 400 and 500 mL, and feeds should not be stopped for the first GRV above this set point. After the first elevated GRV, the head of the bed should be elevated, the patient should be rolled over if possible into the right lateral decubitus position to promote gastric emptying, promotility agents should be considered, and the EN may be continued. In the absence of other signs of intolerance, the feedings should not be stopped for a single GRV above this cutoff point alone. Only after a second GRV is obtained above the cutoff level 4 hours later should feeds be stopped and the patient assessed further for evidence of intolerance. Eliminating use of GRV as a monitor has been shown surprisingly to increase delivery of EN without increases in aspiration, pneumonia, or any other adverse outcome.[■]

COMPLICATIONS OF ENTERAL FEEDING

Aspiration is the most feared complication arising from provision of EN, but the ability of the clinician to monitor such events is limited. Studies using a very specific and very sensitive laboratory marker for aspiration (tracheal pepsin levels) have shown that most critically ill patients (>75%) show signs of aspiration, occurring at a frequency of 22 to 36% of bedside assessments done every 4 hours. These aspiration events are unwitnessed and unmeasurable by clinical parameters, but an increase in their frequency does correlate with increased risk for pneumonia.[7] Ironically, aspiration of bacteria-laden oropharyngeal secretions is more likely than aspiration of bacteria-laden gastric contents to cause pneumonia. Likelihood for developing pneumonia may be reduced in the high-risk patient by elevating the head of the bed, switching from bolus to continuous infusion, displacing the levels of feeds lower in the gastrointestinal tract below the ligament of Treitz, adding a prokinetic agent, initiating oral hygiene twice daily with chlorhexidine mouthwash, and adding simultaneous gastric decompression.

Tube occlusion occurs when acid from the stomach comes in contact with the formula in the tube and forms a clot. The best declogging agent is a Viokase pancreatic enzyme preparation combined with sodium bicarbonate tablets mixed in warm water. This combination is more than twice as effective as soft drinks or papain meat tenderizer. Failure to declog the tube with these agents may necessitate mechanical declogging with some kind of cytology brush, stylette, or commercial corkscrew device.

Although diarrhea is a frequent complaint in the ICU for patients on tube feeding, most cases represent low-volume incontinence. Although diarrhea in the ICU certainly creates a nursing problem, feeding does not need to be stopped. The most frequent cause of diarrhea in the ICU is the addition of sorbitol as a mixing agent for drugs infused through the feeding tube. Pseudomembranous colitis from *Clostridium difficile* occurs in less than 20% of patients (Chapter 296). Rarely, diarrhea is related to the osmolarity of the formula. Switching to a small peptide formula with medium-chain triglyceride oil or adding fiber to a formula often corrects the problem.

Bowel ischemia (Chapter 143) is a rare and unpredictable complication of enteral feeding. This complication is more often described in patients undergoing surgical placement of a small bowel feeding tube, but cases nonetheless have been shown to occur with nasoenteric feeding tubes. In the patient on EN who becomes hypotensive, feeds should be held if pressor therapy is being initiated, the dose of pressors are being increased, or a second or third agent is being added to the first. Feeds may be restarted in the hypotensive patient on pressor agents if they have been stable for 24 to 36 hours or the doses have already been reduced, or both. It is important to confirm adequate volume resuscitation, fiber should be avoided in these situations, and it may be safer to feed into the stomach than into the small bowel. Enteral feeding in the patient on pressor agents should be held for any sign of intolerance, such as increases in nasogastric output, sudden abdominal distention, new abdominal pain, or cessation of flatus and stool, because these may be the first signs of intestinal ischemia.

One other complication that is seen even today in modern urban university-based hospitals is the *refeeding* syndrome,[8] a syndrome of sudden death associated with abrupt initiation of nutrition therapy. Potential candidates for refeeding syndrome are patients who have not been fed for more than 7 to 10 days, are already severely malnourished, require mechanical ventilation and are prone to hypercapnia, or have congestive heart failure. The mechanism of the refeeding syndrome is related to underlying cardiomyopathy and congestive heart failure, electrolyte shifts when the patient becomes anabolic, or volume overload from the feeding itself. Refeeding syndrome may be prevented by closely monitoring fluid volume, electrolytes, and caloric requirements, starting at a low rate of infusion with a mixed fuel substrate and advancing to goal slowly over 3 to 4 days.

CONCLUSION

The provision of early enteral feeding is one of the most important proactive therapeutic strategies that can favorably alter the outcome of critically ill patients through their course of hospitalization. A fairly narrow window of opportunity in time exists during which to initiate feeds in order to achieve attenuation of oxidative stress and modulate systemic immune responses. The timing, dose, and aspects of delivery of EN determine whether the patient receives the full benefit. The physiologic response seen from enteral feeding cannot be duplicated with parenteral nutrition or starvation. Instituting infusion protocols, improving clinician education (regarding issues such as ileus, GRVs, and perceived tolerance), and having the skills for deep jejunal placement of feeding tubes are all factors that promote the delivery of early EN in the ICU setting.

Grade A References

A1. Casaer MP, Mesotten D, Hermans G, et al. Early versus late parenteral nutrition in critically ill adults. *N Engl J Med*. 2011;365:506-517.

A2. Harvey SE, Parrott F, Harrison DA, et al. Trial of the route of early nutritional support in critically ill adults. *N Engl J Med*. 2014;371:1673-1684.

A3. Jiang H, Sun MW, Hefright B, et al. Efficacy of hypocaloric parenteral nutrition for surgical patients: a systematic review and meta-analysis. *Clin Nutr*. 2011;30:730-737.

A4. Rice TW, Wheeler AP, Thompson BT, et al. National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network. Initial trophic vs full enteral feeding in patients with acute lung injury: the EDEN randomized trial. *JAMA*. 2012;307:795-803.

A5. Drover JW, Dhaliwal R, Weitzel L, et al. Perioperative use of arginine-supplemented diets: a systematic review of the evidence. *J Am Coll Surg*. 2011;212:385-399.

A6. van Zanten AR, Sztark F, Kaisers UX, et al. High-protein enteral nutrition enriched with immune-modulating nutrients vs standard high-protein enteral nutrition and nosocomial infections in the ICU: a randomized clinical trial. *JAMA*. 2014;312:514-524.

A7. Reignier J, Mercier E, Le Gouge A, et al. Clinical Research in Intensive Care and Sepsis (CRICS) Group. Effect of not monitoring residual gastric volume on risk of ventilator-associated pneumonia in adults receiving mechanical ventilation and early enteral feeding: a randomized controlled trial. *JAMA*. 2013;309:249-256.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Martindale RG, McCarthy MS, McClave SA. Guidelines for nutrition therapy in critical illness: are they not all the same? *Minerva Anesthesiol*. 2011;77:463-467.
2. Desai V, McClave SA, Rice TW. Nutrition in the ICU: an evidence-based approach. *Chest*. 2014;145:1148-1157.
3. McClave SA, Martindale RG, Rice TW, et al. Feeding the critically ill patient. *Crit Care Med*. 2014;42:2600-2610.
4. Casaer MP, Van den Berghe G. Nutrition in the acute phase of critical illness. *N Engl J Med*. 2014;370:1227-1236.
5. Kushner RF, Drover JW. Current strategies of critical care assessment and therapy of the obese patient (hypocaloric feeding): what are we doing and what do we need to do? *JPEN*. 2011;35:36-43S.
6. Heyland DK, Dhaliwal R, Jiang X, et al. Identifying critically ill patients who benefit the most from nutrition therapy: the development and initial validation of a novel risk assessment tool. *Crit Care*. 2011;15:R268.
7. Metheny NA, Stewart BJ, McClave SA. Relationship between feeding tube site and respiratory outcomes. *JPEN*. 2011;35:346-355.
8. Rio A, Whelan K, Goff L, et al. Occurrence of refeeding syndrome in adults started on artificial nutrition support: prospective cohort study. *BMJ Open*. 2013;3:e002173.

REVIEW QUESTIONS

1. When performing nutritional assessment on a patient in the ICU, the least important factor in the critical care setting is which of the following?

A. Disease severity
B. Nutritional intake before admission
C. Weight as a percentage of ideal body weight
D. Body mass index
E. Anthropometric measures

Answer: E Nutritional risk is determined by both evidence of malnutrition and disease severity. Evidence of malnutrition is determined by evaluating body mass index, weight as a percentage of ideal body weight, and nutrition intake before admission. These markers of nutritional status and disease severity then can be combined through an assessment tool such as the NRS 2002 or the Nutric Score to determine overall nutrition risk (see E-Table 216-1). Anthropometric measures such as arm muscle circumference and skinfold thickness have little application in the hospitalized setting.

2. In a critically ill patient with acute sepsis, serum albumin levels are low due to which of the following?

A. Increased intestinal permeability
B. Increased vascular permeability
C. Increased hepatic production of albumin
D. Shift to greater protein synthesis of skeletal muscle
E. Increased mobilization of albumin from tissue stores

Answer: B Low albumin levels that follow an acute insult, injury, or admission to an intensive care unit occur because of increased vascular permeability, change in prioritization of hepatic protein synthesis (from proteins of homeostasis to acute phase proteins), and use of albumin as a source of cysteine by the body for antioxidant defense. Serum albumin levels thus represent a negative acute phase response and do not really reflect nutritional status.

3. Which of the following regarding use of the gut in nutrition therapy is true?

A. The gut contains less than 25% of the immune tissue in the body.
B. Antigen processing at the level of the gut supports immune tissue localized only within the gut wall.
C. Bacteria that translocate from the gut migrate out to distant organ sites through the systemic circulation and cause direct infection to that organ.
D. Maintaining gut integrity suppresses a gut-lung axis of inflammation that would otherwise lead to respiratory failure.
E. Opening of the paracellular channels between the epithelial cells stimulates blood flow to the mucosa.

Answer: D The gut contains 65% of overall immune tissue of the body and is responsible for 80% of immunoglobulin production. Antigen processing occurs at the level of the gut but supports and delivers immune tissue to distant sites of mucosal-associated lymphoid tissue (MALT), such as the lungs, genitourinary system, and lacrimal glands. Although bacteria and their products may translocate across the gut wall, they rarely make it past mesenteric lymph nodes or the liver. On the rare occasion that bacteria do make it to the systemic circulation, the inoculum is low, and clinical sequelae are minimal. Maintaining gut integrity through provision of enteral nutrition suppresses the gut-lung axis of inflammation, preventing inflammatory cytokines from passing from the gut up through lymphatics to the lung and causing subsequent respiratory failure. Opening the pericellular channels between the epithelial cells is what accounts for increased permeability and allows bacteria within the gut to engage the immune system.

4. Which of the following is true regarding gastric residual volumes as a monitor for aspiration?

A. Gastric residual volumes correlate well with gastric emptying and risk for aspiration pneumonia.
B. Decreasing gastric residual volume from 400 to 200 mL protects the patient from aspiration.
C. Gastric residual volumes in the range 200 to 500 mL indicate a major risk factor for aspiration and should prompt a change in strategy to reduce risk.
D. Eliminating gastric residual volumes as a monitor of enteral nutrition in the intensive care unit is associated with increased risk for aspiration pneumonia.
E. Raising gastric residual volumes from 200 to 400 mL will result in fewer occasions for cessation of formula infusion, but the incidence of aspiration may be expected to increase dramatically.

Answer: C Gastric residual volumes correlate poorly with gastric emptying and risk for aspiration pneumonia. Changing the cutoff value for residual volume (up or down) does not change the incidence of aspiration or gastroesophageal reflux. Lowering the cutoff value for gastric residual volumes leads simply to cessation of nutrition therapy. Elevated gastric residual volumes in the range of 200 to 500 mL have been identified as a major risk factor for aspiration and should prompt a change in strategy to reduce risk, such as elevating the head of the bed, initiating prokinetic therapy, and placing the patient in the right lateral decubitus position. Surprisingly, eliminating the use of gastric residual volume as a monitor of enteral nutrition in the intensive care unit is associated with increased delivery of enteral nutrition and no additional adverse sequelae.

5. Ileus during enteral nutrition represents which of the following?

A. Decreased splanchnic blood flow
B. Occult infection
C. Reduced intestinal contractility
D. Loss of GALT tissue
E. Reduced gastrointestinal absorptive capacity

Answer: C Ileus represents reduced intestinal contractility in response to acute injury, insult, or critical illness. Early provision of enteral nutrition helps preserve contractility and minimize ileus. Although failure to provide enteral nutrition in the critical care setting does result in reduced splanchnic blood flow, loss of GALT tissue, and reduced absorptive capacity with loss of intestinal villi, ileus relates solely to issues of contractility.

217

MALNUTRITION, NUTRITIONAL ASSESSMENT, AND NUTRITIONAL SUPPORT IN ADULT HOSPITALIZED PATIENTS

THOMAS R. ZIEGLER

MALNUTRITION IN HOSPITALIZED PATIENTS

Numerous surveys conducted in developed countries in the 21st century continue to demonstrate the frequent rate of protein-energy malnutrition or depletion of specific micronutrients in patients with chronic illnesses and in those requiring elective or emergent hospital admission. Hospitalized patients commonly receive inadequate amounts of calories, protein, vitamins, and minerals during their stay, and ad libitum intake of prescribed diets is typically inadequate. Studies have shown that worsening of malnutrition during the hospital admission is common. This is a problem because adequate intake of essential macronutrients (energy, carbohydrate, protein/amino acids, and fats) and micronutrients (vitamins, minerals, and electrolytes) is critical for optimal cellular and organ structure and function, muscle mass, tissue repair, immune function, ambulatory capacity, and recovery of the patient. Significant erosion of lean body mass (predominantly derived from skeletal muscle) or deficiency of specific vitamins and minerals is variously associated with weakness and fatigue, increased rates of infection, impaired wound healing, and delayed convalescence. This relationship is especially apparent in those with chronic protein-energy malnutrition and body weight loss associated with illness (Chapters 215 and 216).

Patients with acute and chronic illnesses typically have experienced several days to several weeks or months of continuous or intermittent decreased food intake because of anorexia, gastrointestinal symptoms, depression or anxiety, and other medical factors. They may also have had food intake restricted because of surgical operations or diagnostic or therapeutic procedures and recovery from these. Some patients have abnormal nutrient losses due to diarrhea (e.g., with chronic malabsorptive and maldigestive disorders or infectious diarrhea), vomiting, polyuria (as in uncontrolled diabetes mellitus), wound drainage, dialysis, and other causes. Certain drugs, including corticosteroids, chemotherapeutic agents, antirejection drugs, and diuretics, are associated with skeletal muscle breakdown, gastrointestinal injury, and electrolyte or water-soluble vitamin losses, respectively. Bedrest or markedly decreased ambulation is common in outpatient and inpatient settings and in turn associated with skeletal muscle wasting and impaired protein synthesis. Catabolic and critical illnesses are associated with concomitantly increased blood concentrations of "counter-regulatory" hormones derived from the adrenal glands and pancreas (e.g., cortisol, catecholamines, glucagon); release of proinflammatory cytokines from stimulated immune, endothelial, and epithelial cells (interleukins 1, 6, and 8 and tumor necrosis factor-α); and peripheral tissue resistance to anabolic hormones (insulin and insulin-like growth factor-I). These hormonal and cytokine alterations serve to increase the availability of endogenous metabolic substrates critical for cellular and organ function, wound healing, and host survival (e.g., glucose through glycogenolysis and gluconeogenesis, amino acids through skeletal muscle breakdown, and free fatty acids through lipolysis). This combination of decreased nutrient intake and increased tissue nutrient losses, coupled with increased energy (calorie), protein, and micronutrient needs due to inflammation, infection, and cytokinemia, is responsible for the wasting and micronutrient depletion common in medical patients with acute and chronic illness. Common causes of protein-energy malnutrition and micronutrient depletion in medical patients are shown in Table 217-1. Obesity has become a widespread medical problem and is a form of malnutrition but is considered in detail elsewhere (Chapter 220).

NUTRITIONAL ASSESSMENT

Serial assessment of nutritional status is a critically important component of routine medical care (Chapter 214). The major objectives are to detect preexisting depletion of body protein, energy reserves, and micronutrients; to identify risk factors for malnutrition (see Table 217-1); and to take steps to prevent nutrient deficiencies, depletion of lean body mass, and loss of skeletal muscle. Unfortunately, there are still no practical "gold standard" tests that can be used as an index of general nutritional status. Blood concentrations of specific micronutrients (e.g., copper, zinc, thiamine, 25-hydroxyvitamin D, vitamin B_6, folate, vitamin B_{12}) and electrolytes (e.g., magnesium, potassium, phosphorus) are important to guide needs and repletion responses. Nutritional assessment (Chapter 214) involves an integration of multiple factors: medical and surgical history; type and severity of the acute or chronic underlying illness and anticipated medical and surgical course; fluid drainage sites and amounts; physical examination findings; history of body weight change (degree and temporal aspects); dietary intake pattern; use of nutritional supplements, including prior administration of specialized enteral nutrition (EN) or parenteral nutrition (PN); evaluation of current organ function and fluid status; and determination of selected vitamin, mineral, and electrolyte concentrations in blood (E-Table 217-1). In the intensive care unit (ICU) setting, measured body weight typically reflects recent intravenous fluid administration and is typically much higher than recent "dry" or preoperative body weight, which is the best parameter to use.

E-TABLE 217-1 COMPREHENSIVE NUTRITIONAL ASSESSMENT OF MEDICAL AND SURGICAL PATIENTS

REVIEW PAST MEDICAL AND SURGICAL HISTORY AND CURRENT ILLNESS

Degree of catabolic stress (e.g., fever, infections, sepsis, surgeries, lung failure)
Organ function (e.g., liver, kidneys, lung, heart/vascular, gastrointestinal)
Use of medications that may decrease nutrient absorption (e.g., phenytoin, sulfasalazine, elixir-based medications), alter metabolism/utilization (e.g., warfarin, isoniazid, methotrexate), or increase excretion (e.g., gentamicin, loop diuretics)
Recent intravenous fluid and electrolyte therapy
Medical and surgical procedures that are likely in the near term
Hemodynamic status and requirements for pressor agents to maintain blood pressure

OBTAIN BODY WEIGHT HISTORY

Current body weight: dry weight, if available (e.g., preoperative, recent clinic visit), and usual body weight when healthy or clinically stable; calculate percentage body weight loss from usual body weight during the past several weeks and months
Calculate current weight as percentage of ideal body weight*
Determine body mass index (BMI = weight (kg)/height $(m)^2$; BMI < 18.5 is considered underweight)

DETERMINE DIETARY INTAKE PATTERN IN RELATION TO NUTRIENT NEEDS

General food and beverage intake pattern; percentage of usual dietary intake consumed in recent weeks and months (e.g., percentage of usual); unusual or excessive consumption of specific foods or beverages, including alcoholic beverage consumption
Previous use and type of enteral tube feedings or parenteral nutritional support
Previous use of liquid or solid nutritional supplements, multivitamin-mineral preparations, specific vitamins or minerals
Consult registered dietitian for more detailed nutrient intake assessment

PERFORM DETAILED PHYSICAL EXAMINATION

Skeletal muscle wasting (cannot be assessed accurately in overweight or obese patients)
Loss of body fat stores
Presence and qualities of wounds
Skin, hair, tongue, or conjunctival lesions suggestive of micronutrient deficiency (see Table 217-2)
Evidence of organ dysfunction (gastrointestinal, liver, renal, cardiopulmonary)
Fluid status (e.g., normal, dehydrated, fluid overload, capillary leak); fluid requirements are typically ≈ 30-40 mL/kg body weight

EVALUATE GASTROINTESTINAL TRACT FUNCTION

Swallowing or chewing difficulties, nausea, emesis, abdominal pain
Intestinal ileus, motility disorders, partial or complete obstruction
Diarrhea history (frequency, amount, other characteristics)
Acute or chronic gastrointestinal bleeding
Presence of fistulas; history of recent abdominal surgery
Drainage tube losses (e.g., gastric, biliary, intestinal, peritoneal)

DETERMINE FUNCTIONAL STATUS

Ability to perform daily activities, ambulatory capacity, bedrest, chemical paralysis
History or physical examination evidence of muscle weakness and fatigue
Mental capacity, history of psychiatric disorders that may preclude oral food intake

SERIAL EVALUATION OF SELECTED BIOCHEMICAL TESTS

Concentrations of standard blood measures of organ function
Electrolyte concentrations (e.g., calcium, magnesium, phosphorus, and potassium)
Blood pH (in ICU patients on mechanical ventilation)
Blood triglyceride concentrations (in patients receiving intravenous lipid emulsion)
Blood concentrations of selected vitamins and minerals if suggested by medical or dietary history, physical examination, or underlying illness (e.g., blood levels of zinc, selenium, copper, thiamine, vitamin B_6, vitamin B_{12}, folate, 25-hydroxyvitamin D)
Serum prealbumin (in stable outpatients)†

ESTIMATE CALORIE, PROTEIN, AND MICRONUTRIENT NEEDS

Calorie needs by Harris-Benedict equation,‡ guidelines on kilocalories per kilogram of body weight,§ or indirect calorimetry results (use serially in ICU patients and in initial assessment of very underweight or obese hospitalized patients requiring prolonged nutritional support)
Protein needs vary as a function of recognized clinical situations (see Table 217-3)
Vitamin and mineral needs are based on conventional requirements, serial blood levels, and clinical judgment regarding intake and estimated losses from comprehensive nutritional assessment
Nitrogen balance studies are not useful in non-research settings because of their variability and inaccuracy
Serial body composition measurements for lean body mass and body fat estimates may be useful in the outpatient setting (by BIA or DEXA) but are not practical or reliable in inpatient non-research settings because of fluid shifts and other factors

EVALUATE ENTERAL AND PARENTERAL ACCESS FOR NUTRIENT DELIVERY

Ability to take oral diet or liquid supplements
Central venous or PICC line access; peripheral line access
Nasogastric, nasoenteric, or percutaneous feeding tube availability or feasibility

CONSULTATION WITH MULTIDISCIPLINARY NUTRITION SUPPORT TEAM

BIA= bioelectrical impedance analysis; DEXA = dual-energy x-ray absorptiometry; ICU = intensive care unit; PICC = peripherally inserted central venous catheter.
*Ideal body weight can be estimated in men as 48 kg (106 pounds) per 5 feet of height + 2.7 kg (6 pounds) for each inch of height above 5 feet and in women as 45 kg (100 pounds) per 5 feet of height + 2.3 kg (5 pounds) for each inch of height above 5 feet.
†Blood concentrations of albumin and prealbumin in hospitalized patients, especially in the ICU setting, are markedly affected by non-nutritional factors (inflammation, infection, fluid status, capillary leak, decreased hepatic synthesis, and increased clearance from blood).
‡Harris-Benedict equation to estimate basal energy expenditure (BEE) for males and females:

$$\text{Males (kcal per 24 hours)} = 66.5 + (13.8 \times \text{kg body weight}) + (5.0 \times \text{height in cm}) - (6.8 \times \text{age in years})$$

$$\text{Females (kcal per 24 hours)} = 655 + (9.6 \times \text{kg body weight}) + (1.8 \times \text{height in cm}) - (4.7 \times \text{age in years})$$

§Calorie needs can also be estimated as kcal/kg/day (using dry weight or ideal body weight in ICU patients with fluid overload): ICU settings = 20 to 25 kcal/kg/day (some studies suggest that 15 to 20 kcal/kg/day or lower may be appropriate); non-ICU settings = 25 to 35 kcal/kg/day.

TABLE 217-1 COMMON CAUSES OF PROTEIN-ENERGY MALNUTRITION AND MICRONUTRIENT DEPLETION IN MEDICAL PATIENTS WITH ACUTE OR CHRONIC ILLNESSES

- Decreased spontaneous food intake due to anorexia from chronic or acute illness, gastrointestinal symptoms (e.g., nausea, vomiting, abdominal pain), depression/anxiety
- Restricted food intake required for surgical operations or diagnostic/therapeutic procedures and gastrointestinal dysfunction that follows these
- Abnormal macronutrient and micronutrient losses from the body due to malabsorption (e.g., celiac sprue, short-gut syndrome, inflammatory bowel disease, cystic fibrosis, diarrhea), maldigestion (e.g., pancreatitis), emesis, polyuria (e.g., in diabetes), wound drainage, or renal replacement therapy
- Periods of increased energy expenditure (calorie needs), protein requirements, and micronutrient needs (e.g., critical illness, increased inflammation)
- Catabolic effects of counter-regulatory hormones (e.g., cortisol, catecholamines, glucagon), release of proinflammatory cytokines from stimulated immune cells and endothelial and epithelial cells (interleukins 1, 6, and 8 and tumor necrosis factor-α), and peripheral tissue resistance to the anabolic hormones insulin and insulin-like growth factor-I
- Bedrest, decreased ambulation, and chemical paralysis during mechanical ventilation (skeletal muscle wasting due to impaired protein synthesis)
- Administration of drugs that induce skeletal muscle breakdown, gastrointestinal injury, or electrolyte and water-soluble vitamin losses (e.g., corticosteroids, chemotherapeutic agents, diuretics, and antirejection regimens)
- Socioeconomic deprivation, inadequate caregivers, ambulation difficulties in the home setting
- Inadequate provision of calories, protein, and essential micronutrients (vitamins, minerals, and trace elements) during hospitalization

TABLE 217-2 CLINICAL MANIFESTATIONS OF SPECIFIC NUTRIENT DEFICIENCIES

PHYSICAL EXAMINATION SIGN OR SYMPTOM OF NUTRIENT DEPLETION	SPECIFIC NUTRIENT DEPLETED
Muscle and fat wasting, weakness	Calories, protein, combined calories and protein
Anorexia	Calories, protein
Glossitis (discolored, smooth, painful tongue)	Folate, vitamin B_{12}, niacin, riboflavin, thiamine, iron
Cheilosis, angular stomatitis	Riboflavin, niacin, folate, vitamin B_{12}
Symmetrical motor or sensory dysfunction, ataxia, nystagmus, heart failure, mental status changes or confusion	Thiamine (beriberi)
Peripheral edema	Thiamine (heart failure), protein (low oncotic pressure)
Loss of vibratory or position sense, fatigue	Vitamin B_{12}
Dermatitis (sun-exposed skin), diarrhea, dementia	Niacin (pellagra)
Bleeding gums, petechiae, ecchymosis	Vitamins C and K
Poor wound healing	Calories, protein, calories and protein, vitamin C, vitamin A, zinc, others
Bone pain	Vitamin D (osteomalacia)
Follicular hyperkeratosis, night blindness, Bitot's spots	Vitamin A
Flaky, whitish dermatitis	Essential fatty acid (linoleic, α-linolenic)
Hair sparse or easily pluckable	Zinc, protein
Pale skin, nail spooning (koilonychia)	Iron
Loss of taste; reddish dermatitis around nose, mouth, groin; hair loss	Zinc
Peripheral neuropathies, gait abnormalities, weakness, fatigue	Copper
Muscle pain, heart failure	Selenium
Paresthesias, carpal pedal spasm	Calcium, magnesium, phosphorus, or potassium

Note: Typically severe deficiency of specific nutrients (with depletion of initially tissue and later blood concentrations) has occurred before physical manifestations of deficiency.

Integration of the factors outlined in E-Table 217-1 provides important information on whether patients are likely to be adequately nourished; to have mild, moderate, or severe protein-energy malnutrition; or to have depletion or deficiency of specific vitamins, minerals, or electrolytes. Patients with an involuntary body weight loss of 5 to 10% or more of their usual body weight in the previous few weeks or months, those weighing less than 90% of their ideal body weight, or those who have a body mass index less than 18.5 kg/m^2 should be carefully evaluated as these individuals are likely to be malnourished.

In hospitalized patients, especially those in the ICU, circulating concentrations of proteins (e.g., albumin, prealbumin) are generally low and not useful as protein nutritional status biomarkers, given their lack of specificity. Plasma concentrations of albumin and prealbumin typically fall during active inflammation or infection, in critical illness, and after traumatic injury (deceased synthesis by liver, catabolism of blood proteins) and are markedly affected by non-nutritional factors, including fluid status, capillary leak, decreased hepatic synthesis, and increased clearance from blood. Because of the long circulating half-life of albumin (18 to 21 days), concentrations in blood remain low despite adequate feeding and are slow to respond to nutritional repletion, irrespective of the other confounding factors noted before. Prealbumin has a much shorter circulating half-life than albumin (several days), and serial blood levels can be used as a general indicator of protein status in clinically stable outpatients. Table 217-2 illustrates physical examination findings that may be observed in association with depletion of specific nutrients.

Energy requirements can be estimated by standard equations, such as the Harris-Benedict equation, which incorporate the patient's age, gender, weight, and height to determine basal energy expenditure (BEE; see E-Table 217-1). Physical activity and the thermic effect of macronutrient administration can be added to the BEE to arrive at the energy prescription to maintain current body weight; this is estimated for most hospital patients and outpatients as 1.2 to 1.3 times BEE, unless the patient is sedated or at bedrest (common in the ICU), which decreases energy needs. The estimated maintenance energy requirement is approximately 1.3 times BEE in ambulatory subjects. Lower amounts of calories are now typically given in ICU patients (see later). Data obtained from a bedside metabolic cart machine (indirect calorimeter), which measures expired breath to determine oxygen consumption and carbon dioxide production, provide accurate actual energy expenditure in most settings and can be useful (see E-Table 217-1). A simple and relatively accurate method to estimate energy needs is simply to use 20 to 25 kcal/kg/day actual dry or ideal body weight in most patients. This assumes that the body weight used does not reflect intravenous fluid administration or capillary leak syndromes (see earlier). In ICU patients, even lower calorie doses (equivalent to 15 to 20 kcal/kg dry weight/day) have been advocated by some on the basis of known complications of overfeeding (see later) and limited clinical outcome data as a function of energy dose. In clinically stable, malnourished, non-ICU patients who require nutritional repletion, higher doses of calories (up to 35 kcal/kg/day) appear to be generally well tolerated, as long as refeeding syndrome is avoided (see later). In obese subjects (defined for these calculations as >20 to 25% above ideal body weight), adjusted body weight should be used in the calculation of energy and protein needs by the following equation:

$$\text{Adjusted body weight} = \text{current weight} - \text{ideal body weight}\ (\text{from standard tables or equations}) \times 0.25 + \text{ideal body weight}$$

Guidelines for protein/amino acid administration are given in Table 217-3. Studies in nonburned ICU patients indicate that protein loads of more than 2.0 g/kg/day are not efficiently used for protein synthesis and the excess may be oxidized and contribute to azotemia. In most catabolic patients requiring specialized feeding, a generally recommended protein dose is 1.5 g/kg/day in individuals with normal renal function. This is about twice the recommended dietary allowance for healthy adults of 0.8 g/kg/day. The administered protein dose should be adjusted downward as a function of the degree and tempo of azotemia (in the absence of dialysis therapy) and hyperbilirubinemia (see Table 217-3). This strategy takes into account the relative inability of catabolic patients to efficiently use exogenous nutrients and the knowledge that most protein and lean tissue repletion occurs in a period of several weeks to months during post-hospital convalescence. Adequate

TABLE 217-3 ESTIMATION OF PROTEIN/AMINO ACID REQUIREMENTS IN ADULT PATIENTS

CLINICAL CONDITION	PROTEIN/AMINO ACID DOSE (g/kg/day)*,†
Well nourished with acute illness	1.2-1.5
Malnourished or severe catabolic stress	1.5-2.0
Postoperative	1.2-1.5
Acute hepatic failure	0.6-1.2
Encephalopathy	0-0.6
Acute renal failure, not receiving renal replacement therapy	0.6-1.0
Renal failure, receiving renal replacement therapy	1.2-2.5

*Oral/enteral nutrient supplements and tube feedings contain either intact or partially hydrolyzed high-quality protein (typically casein, soy, or whey). Parenteral nutrition solutions for peripheral or central vein administration provide known essential L-amino acids combined with several nonessential amino acids. These may be limiting in certain conditionally essential amino acids (e.g., cysteine, taurine) in some clinical conditions.
†Limited data from randomized controlled trials on optimal protein/amino acid dosing in hospitalized patients are available.

TABLE 217-4 SOME CLINICAL INDICATIONS FOR SPECIALIZED ORAL/ENTERAL OR PARENTERAL NUTRITION SUPPORT

- Patient currently exhibits moderate to severe protein or protein-energy malnutrition or has evidence of specific deficiency of one or more essential micronutrients
- Patient with involuntary body weight loss of 5-10% or more of usual body weight in the previous few weeks or months, who weighs <90% of ideal body weight, or who has BMI <18.5 kg/m^2.
- Dietary food intake in hospital or outpatient setting likely to be <50% of needs for more than 5-10 days because of underlying illness
- Patient with severe catabolic stress (e.g., ICU care, serious infection) and adequate nutrient intake unlikely for >3-5 days
- After major gastrointestinal surgery or other major operations (e.g., hip replacement, partial organ resection)
- Medical illness associated with prolonged (>5-10 days) gastrointestinal dysfunction (diarrhea, nausea/vomiting, gastrointestinal bleeding, severe ileus, partial obstruction) or short-bowel syndrome, chronic/severe diarrhea, or other malabsorptive disorders
- Clinical settings in which adequate oral food intake may be contraindicated or otherwise significantly decreased, such as respiratory or other acute or severe organ failure, dementia, dysphagia, chemotherapy/irradiation, inflammatory bowel disease, pancreatitis, high-output enterocutaneous fistula, alcoholism, drug addiction
- Chronic obstructive lung disease, chronic infection, and other chronic inflammatory or catabolic disorders with documented poor nutrient intake or recent weight loss

BMI = body mass index; ICU = intensive care unit.

nonprotein energy is essential to allow amino acids to be effectively used for protein synthesis and to not be oxidized for energy (adenosine triphosphate) production. The nonprotein calorie-to-nitrogen ratio used in most centers now typically ranges from 75 : 1 to 125 : 1 (nitrogen = protein/6.25; thus, 75 to 125 nonprotein kilocalories for each 6.25 g of protein or amino acid administered). Highly catabolic patients in the ICU are typically given protein loads at the lower end of this range, assuming near-normal renal and hepatic function.

NUTRITIONAL SUPPORT

Table 217-4 lists common clinical scenarios in which specialized oral/EN or PN support may be indicated. In these settings, consultation with a multidisciplinary nutrition support team, if one is available, has been shown to reduce complications and costs and to increase the appropriate use of EN and PN in both academic and community medical centers.[1]

Oral nutrition supplementation includes provision of balanced oral diets of usual foods supplemented with complete liquid (or solid) nutrient products, protein supplements (e.g., hydrolyzed whey or casein powder that can be mixed with dietary beverages), high-potency multivitamin-mineral supplements, and specific micronutrients required to treat a diagnosed deficiency (e.g., zinc, copper, vitamin B_6, vitamin B_{12}, and vitamin D). Special supplements designed for patients with chronic renal failure featuring concentrated calories and low amounts of protein and electrolytes are available, as are a variety of formulations designed for other specific disease categories (see later). Several studies show that convalescence is enhanced with addition of one or two cans per day of complete liquid nutrient supplements to meals after stresses such as total hip replacement and gastrointestinal surgery. These provide calories, carbohydrate, high-quality protein, fat, and micronutrients; they are lactose and gluten free and may contain small peptides and medium-chain triglycerides to facilitate amino acid and fat absorption, respectively. Some formulations also contain soluble fiber or prebiotics (e.g., fructo-oligosaccharides) designed to decrease diarrhea. For patients who can tolerate oral medications, it is probably prudent to prescribe a potent oral multivitamin-mineral preparation, at least for several months, especially for those who either exhibit or are at risk for micronutrient depletion (see E-Table 217-1 and Table 217-2).

Administration of Enteral Tube Feeding

Patients with conditions outlined in Table 217-4 may have a functional gastrointestinal tract but may be unable to consume an adequate diet orally because of medical or surgical conditions (e.g., mechanically ventilated patients; those with pancreatitis, dementia, or dysphagia; and after trauma or burns). Although PN is commonly administered in these settings, this practice is not evidence based; academic guidelines strongly suggest that oral nutritional supplements or enteral tube feedings be used if specialized nutrition support is indicated in patients with a functional gastrointestinal tract ("if the gut works, use it"). On an individualized basis, aggressive nutrition support, including placement of feeding tubes, may not be desired by a competent patient or legally authorized representatives, such as in premorbid states or terminal illness. In these cases, full discussion with the patient and family or representatives is required with regard to the plan for EN.

Detailed discussion of EN and enteral tube feeding is provided in Chapter 216.

Administration of Parenteral Nutrition

PN support includes administration of standard complete nutrient mixtures, which contain dextrose, L-amino acids, lipid emulsion, electrolytes, vitamins, and minerals (and certain medications as indicated, such as insulin or octreotide), given through either a peripheral or central vein. PN technically also includes parenteral administration of specific micronutrients or micronutrient combinations to replete a deficiency (e.g., thiamine, copper, electrolytes). Administration of complete PN therapy in patients with gastrointestinal tract dysfunction has become a standard of care in most hospitals and ICUs throughout the world, although use in individual institutions varies widely.[2,3] PN is life-saving in patients with intestinal failure (e.g., short-bowel syndrome); unfortunately, in patient subgroups with lesser degrees of intestinal failure, few objective data from properly designed, large, randomized controlled studies are available to determine the true efficacy of and optimal indications for PN.[4] The use of PN in ICUs in the United States declined during the past decade, with EN increasing coincidentally during the same time.[5]

Existing data indicate that PN does benefit patients with preexisting moderate to severe malnutrition or with critical illness by decreasing overall morbidity and possibly mortality compared with patients receiving inadequate EN or hydration (intravenous dextrose) therapy alone. An earlier meta-analysis of well-designed, intent-to-treat, randomized controlled trials in adult ICU patients showed that early PN and early EN (each started within 24 hours of ICU admission) were equivalent in terms of mortality, but PN was associated with decreased mortality compared with ICU patients who received delayed enteral feeding (begun >24 hours after admission).■ In this and earlier trials, however, PN use was associated with a higher rate of infection compared with enterally fed patients. Randomized controlled trials suggest that patient subtype and ability to tolerate EN can influence the clinical effects of different strategies for the timing of initiation of PN in critical illness. A consensus based on recent rigorous studies in critical illness is emerging that early PN is no better than early EN.■ PN should probably not be initiated until day 3 or 4 after ICU admission in patients unable to tolerate adequate EN.■■ The provision of early PN to critically ill adults with relative contraindications to early EN, compared with standard care, was recently found not to result in a difference in mortality at 60 days, fewer days of ventilation, or significantly shorter ICU stay or hospital days.■

The basic principle in considering PN therapy is that the patient must be unable to achieve adequate nutrient intake by the enteral route. Compared with PN, EN is less expensive, probably maintains intestinal mucosal structure and function to a greater extent, is safer in terms of mechanical and metabolic complications (see later), and is associated with reduced rates of nosocomial infections. Thus, the enteral route of feeding should be used and advanced whenever possible and the amount of administered PN correspondingly reduced. Generally recognized indications for PN include the following:

- Patients with short-bowel syndrome or other conditions causing intestinal failure that prohibit adequate intake or absorption of enteral nutrients (e.g., motility disorders, obstruction, severe ileus, severe inflammatory bowel disease), especially in those with preexisting malnutrition
- Clinically stable patients in whom adequate enteral feeding (e.g., >50% of needs) is unlikely for 7 to 10 days because of any underlying illness
- Patients with severe catabolic stress requiring ICU care in whom adequate enteral nutrient intake is unlikely for more than 3 to 5 days

There is no reason to withhold PN in hospital patients for any time if they exhibit preexisting moderate to severe malnutrition and are deemed to be unlikely to meet their needs by the oral or enteral route.

Generally accepted contraindications to PN (that are largely not evidence based) include the following:

- The gastrointestinal tract is functional and access for enteral feeding is available
- PN is thought to be required for 5 days or less
- The patient cannot tolerate the extra intravenous fluid required for PN or has severe hyperglycemia or electrolyte abnormalities on the planned day of PN initiation
- The patient has uncontrolled blood stream infection or severe hemodynamic instability
- New placement of an intravenous line solely for PN poses undue risks on the basis of clinical judgment
- On an individualized basis when aggressive nutrition support is not desired by the competent patient or legally authorized representatives, such as in premorbid patients or those with terminal illness (full discussion with the patient and family or representatives is required)

PN can be delivered as either peripheral vein solutions or central vein solutions through percutaneous subclavian vein or internal jugular vein catheters for infusion into the superior vena cava (nontunneled in the hospital setting), subcutaneously tunneled central venous catheters (e.g., Hickman catheters) or central venous ports (for chronic home PN therapy), or peripherally inserted central venous catheters (PICC). Although data are limited, it is clearly preferable to manage patients requiring long-term central venous PN at home with a tunneled central venous catheter compared with a PICC line because of the higher rate of local complications (e.g., phlebitis, catheter breakage) and possibly catheter-associated infections with PICC lines.

A comparison of typical fluid, macronutrient, and micronutrient content of peripheral and central vein PN is shown in Table 217-5. To diminish the risk of phlebitis, typical peripheral vein PN solutions provide low concentrations of dextrose (5%; provides 3.4 kcal/g) and amino acids (<3.5 %; provides 4 kcal/g), with a large proportion of energy administration as fat emulsion (50 to 60% of total calories). Because fluid restriction or organ dysfunction often precludes use of large fluid volumes for PN, peripheral vein PN is generally not indicated in ICU patients or in patients with fluid overload or renal, hepatic, or cardiac failure. These solutions are most useful in stable patients who can tolerate the large fluid volumes required to meet amino acid and energy goals (usually 2.5 to 3 L/day) without providing excessive lipid.

Intravenous lipid emulsions (typically added to PN as a 20% soybean oil–based solution in the United States) provide both essential linoleic and α-linolenic fatty acids and energy (10 kcal/g); these are generally infused during a 24-hour period in the complete PN administration bag. The maximal recommended rate of fat emulsion infusion is approximately 1.0 g/kg/day. Most patients clear triglyceride from intravenous fat emulsion well from plasma. In some studies, larger doses of soybean oil–based fat emulsion were associated with proinflammatory and pro-oxidative effects and possibly immune suppression, presumably due to the high amount of omega-6 fatty acids derived from the linoleic fatty acid component. This has led to the approval and clinical availability of intravenous fish oil, olive oil/soybean oil, medium-chain triglyceride/soybean oil, and combinations of these formulations in Europe and other non-U.S. countries. An intravenous lipid emulsion of 80% olive oil/20% soybean oil was recently approved for use in adult PN in the United States. It is important to monitor blood triglyceride levels at baseline and then approximately weekly and as indicated to assess clearance of intravenous fat. Triglyceride levels should be maintained below 400 mg/dL to decrease the risk of pancreatitis or diminished pulmonary diffusion capacity in patients with severe chronic obstructive lung disease.

TABLE 217-5 COMPOSITION OF TYPICAL PARENTERAL NUTRITION SOLUTIONS

COMPONENT*	PERIPHERAL PN	CENTRAL PN
Volume (L/day)	2-3	1-1.5
Dextrose (%)	5	10-25
Amino acids (%)†	2.5-3.5	3-8
Lipid (%)‡	3.5-5.0	2.5-5.0
Sodium (mEq/L)	50-150	50-150
Potassium (mEq/L)	20-35	30-50
Phosphorus (mmol/L)	5-10	10-30
Magnesium (mEq/L)	8-10	10-20
Calcium (mEq/L)	2.5-5	2.5-5
Trace elements§		
Vitamins¶		

*Electrolytes in parenteral nutrition (PN) are adjusted as indicated to maintain serially measured serum levels within the normal range. The percentage of sodium and potassium salts as chloride is increased to correct metabolic alkalosis, and the percentage of salts as acetate is increased to correct metabolic acidosis. Regular insulin is added to PN as needed to achieve blood glucose goals (separate intravenous insulin infusions are commonly required with hyperglycemia in intensive care unit settings).
†Provides all essential amino acids and several nonessential amino acids. Dose of amino acids is adjusted downward or upward to goal as a function of the degree of azotemia or hyperbilirubinemia in patients with renal and hepatic failure, respectively.
‡Lipid is given as soybean oil– or olive oil/soybean oil–based fat emulsion in the United States, Europe, and other non-U.S. countries; intravenous fish oil, olive oil, medium-chain triglycerides, and combinations of these are available for use in PN. Lipid is typically mixed with dextrose and amino acids in the same PN infusion bag ("all-in-one" solution).
§Trace elements added on a daily basis to peripheral vein and central vein PN are mixtures of chromium, copper, manganese, selenium, and zinc (can also be supplemented individually).
¶Vitamins added on a daily basis to peripheral vein and central vein PN are mixtures of vitamins A, B_1 (thiamine), B_2 (riboflavin), B_3 (niacinamide), B_6 (pyridoxine), B_{12}, C, D, and E and biotin, folate, and pantothenic acid. Vitamin K is added on an individual basis (e.g., in patients with cirrhosis). Specific vitamins can also be supplemented individually.

Central venous administration of PN allows higher concentrations of dextrose (3.4 kcal/g) and amino acids (4 kcal/g) to be delivered as hypertonic solutions, and thus lower amounts of fat emulsion are needed to reach calorie goals (see Table 217-5). Requirements for potassium, magnesium, and phosphorus are typically higher with central vein PN compared with peripheral vein PN because of the increased dextrose provided by insulin-mediated intracellular electrolyte shifts, use in anabolic pathways, glucose metabolism, and adenosine triphosphate production. The higher concentrations of dextrose and amino acids possible allow most patients to achieve calorie and amino acid goals with only 1 to 1.5 L/day of PN. In central vein PN, initial orders typically provide 60 to 70% of non–amino acid calories as dextrose and 30 to 40% of non–amino acid calories as fat emulsion. These percentages are adjusted as indicated on the basis of blood glucose and triglyceride levels, respectively. The dextrose amount in central vein PN should be reduced or regular insulin added to the PN bag to maintain blood glucose concentration within the desired range. Separate intravenous insulin infusions should usually be used in the ICU when patients receiving central vein PN develop hyperglycemia.[6]

Specific requirements for intravenous trace elements and vitamins have not been rigorously defined for patient subgroups, and therapy is directed at meeting published recommended doses that maintain blood levels in the normal range in most stable patients with standardized intravenous preparations (see Table 217-5). Several studies have shown that a significant proportion of ICU patients have low zinc, selenium, vitamin C, vitamin E, and vitamin D levels despite receiving specialized PN (or EN). Depletion of these essential nutrients may in turn impair antioxidant capacity, immunity, wound healing, and other important body functions. For example, zinc is known to be important for immune function, wound healing, protein synthesis, and gastrointestinal mucosal regeneration. Zinc (and other micronutrients, such as copper) should probably be increased in the PN of patients with burns, large wounds, significant gastrointestinal fluid losses, and other conditions if

serum concentrations indicate low levels. This practical recommendation is not evidence based, however, because recent rigorous randomized trials in ICU patients administered large doses of selenium could not reproduce the positive clinical benefits observed in earlier smaller studies, and studies of zinc supplementation in the ICU setting have been inconclusive. Recent data suggest that thiamine depletion is not uncommon in patients receiving chronic diuretic therapy or in those with severe malabsorption.

Complications of Parenteral Nutrition

The most common complication of peripheral vein PN is local phlebitis due to the catheter. Alterations in blood electrolytes can be treated with adjustment of concentration in the peripheral PN prescription. Hypertriglyceridemia typically responds well to lowering of the total PN lipid dose. Central vein PN is associated with a much higher rate of mechanical, metabolic, and infectious complications than peripheral vein PN. Mechanical complications include those related to insertion of the central venous catheter (e.g., pneumothorax, hemothorax, malposition of the catheter, and thrombosis). Infectious complications include catheter-related blood stream infections and non–catheter-related infections by bacteria and fungal species that may in some cases be due to endogenous bacterial translocation from the gut lumen. The risk for these infections appears to be increased with use of non–subclavian vein central venous access (e.g., jugular, femoral veins) and multiple-use catheters with non-dedicated PN infusion ports used for additional purposes, such as blood drawing or medication administration. Poorly controlled blood glucose concentration (>140-180 mg/dL) is not uncommon in patients requiring central vein PN and is associated with an increased risk of nosocomial infection. Risk factors for hyperglycemia include poorly controlled blood glucose concentration at PN initiation; use of high dextrose concentrations (>10%) in the initial few days of PN administration or too rapid an increase in total dextrose load; insufficient exogenous insulin administration; inadequate monitoring of blood glucose responses to central vein PN administration; and administration of corticosteroids and vasopressor agents such as norepinephrine, which stimulate gluconeogenesis and cause insulin resistance.

Recent data also suggest that inadequate or no provision of the amino acid glutamine may increase infection risk in patients requiring PN. This amino acid appears to be conditionally essential in catabolic states and serves as an important fuel for immune cells and cells of the gut mucosa, among other potentially beneficial functions. A large number of animal studies and several human trials show that supplementation of glutamine in EN and PN enhances immunity, decreases hospital infections, and maintains indices of gut barrier function. Several expert panels now recommend that glutamine be routinely added to the PN in ICU patients, but this practice remains controversial because some studies show no benefit (or even harm) in certain patient subgroups, and an improvement in hospital mortality has not been documented.■

Studies of nutrient use efficiency and metabolic complications in severely catabolic patients suggest that lower amounts of total energy and amino acid/protein should be administered than were routinely given in the past, particularly in unstable and ICU patients. High calorie, carbohydrate, amino acid, and fat loads (hyperalimentation) are easily administered by central vein PN but, if ordered by physicians, can induce severe metabolic complications, including carbon dioxide overproduction, azotemia, hyperglycemia, electrolyte alterations, and hepatic steatosis and injury (Table 217-6). Dextrose and lipid doses in PN should be advanced during several days after initiation; blood glucose concentration, electrolyte values, triglyceride levels, organ function test results, intake and output measurements, and the clinical course must be closely monitored.

Refeeding syndrome[7] with central vein PN administration is relatively common in patients at risk, including those with preexisting malnutrition, electrolyte depletion, or alcoholism, and after prolonged periods of intravenous hydration therapy (e.g., 5% dextrose) without nutrition support, all of which are common in hospitalized patients. Refeeding syndrome is mediated by administration of excessive intravenous dextrose (>150-250 g, such as given in 1 L of PN with 15-25% dextrose). This, in turn, markedly stimulates insulin release, which rapidly lowers blood potassium, magnesium, and especially phosphorus concentrations because of intracellular shift and use in carbohydrate metabolic pathways. Administration of high doses of carbohydrate also consumes thiamine, which is required as a cofactor for carbohydrate metabolism and can precipitate symptoms of thiamine deficiency (see Table 217-2), especially in patients with poor thiamine nutrition at baseline. Hyperinsulinemia also tends to cause sodium and fluid retention at the level of the kidney. Together, fluid and sodium retention, the drop in blood electrolyte levels (which can cause arrhythmias), and hypermetabolism due to excessive calorie provision can result in heart failure, especially in patients with preexisting heart disease, as well as cardiac muscle atrophy due to prolonged protein-energy malnutrition. Prevention of refeeding syndrome requires vigilance to identify patients at risk; use of initially low PN dextrose concentrations; and empirical provision of higher doses of potassium, magnesium, and phosphorus based on current blood levels and renal function and supplemental thiamine (100 mg/day for 3 to 5 days).

For patients in whom home PN is indicated, primary physicians should consult with social service professionals to identify appropriate home care companies and nutrition support professionals to assess intravenous line access, metabolic status, and the home PN order and to arrange for follow-up care and monitoring of PN. It is important not to arrange for hasty hospital discharge in patients newly started on PN; obtaining appropriate venous access and monitoring of fluid and electrolyte status during a 2- to 3-day period are important aspects of care for most patients started on PN and imperative in those with severe malnutrition and those at risk for refeeding syndrome.

TABLE 217-6 SOME COMMON METABOLIC COMPLICATIONS OF PARENTERAL NUTRITION

PN ORDER PROBLEM	METABOLIC OR CLINICAL CONSEQUENCE
Excess kcal, CHO, fat	Abnormal liver function test results, hepatic steatosis
Excess CHO	Hypercapnia
Excess fluid, kcal, CHO, fat	Respiratory insufficiency
Excess amino acids	Azotemia
Excess sodium and fluid	Sodium and fluid retention
Excess CHO; inadequate insulin	Hyperglycemia-mediated immune cell dysfunction, infection
Inadequate or excessive electrolytes	Abnormal blood electrolyte levels
Excess fluid, kcal, sodium, CHO; inadequate electrolytes	Cardiac failure, arrhythmias
Excess CHO; inadequate electrolytes, thiamine	Refeeding syndrome

CHO = carbohydrate; kcal = calories; PN = parenteral nutrition.

Grade A References

A1. Simpson F, Doig GS. Parenteral versus enteral nutrition in the critically ill patient: a meta-analysis of trials using the intention to treat principle. *Intensive Care Med*. 2005;31:12-23.

A2. Harvey SE, Parrott F, Harrison DA, et al. Trial of the route of early nutritional support in critically ill adults. *N Engl J Med*. 2014;371:1673-1684.

A3. Heidegger CP, Berger MM, Graf S, et al. Optimisation of energy provision with supplemental parenteral nutrition in critically ill patients: a randomised controlled clinical trial. *Lancet*. 2013;381:385-393.

A4. Casaer MP, Mesotten D, Hermans G, et al. Early versus late parenteral nutrition in critically ill adults. *N Engl J Med*. 2011;365:506-517.

A5. Doig GS, Simpson F, Sweetman EA, et al. Early parenteral nutrition in critically ill patients with short-term relative contraindications to early enteral nutrition: a randomized controlled trial. *JAMA*. 2013;309:2130-2138.

A6. Tao KM, Li XQ, Yang LQ, et al. Glutamine supplementation for critically ill adults. *Cochrane Database Syst Rev*. 2014;9:CD010050.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Dhaliwal R, Cahill N, Lemieux M, et al. The Canadian critical care nutrition guidelines in 2013: an update on current recommendations and implementation strategies. *Nutr Clin Pract*. 2014;29:29-43.
2. Ziegler TR. Parenteral nutrition in the critically ill patient. *N Engl J Med*. 2009;361:1088-1097.
3. Casaer MP, Van den Berghe G. Nutrition in the acute phase of critical illness. *N Engl J Med*. 2014;370:1227-1236.
4. Ziegler TR. Nutrition support in critical illness—bridging the evidence gap. *N Engl J Med*. 2011;365:562-564.
5. Gershengorn HB, Kahn JM, Wunsch H. Temporal trends in the use of parenteral nutrition in critically ill patients. *Chest*. 2014;145:508-517.
6. McClave SA, Kozar R, Martindale RG, et al. Summary points and consensus recommendations from the North American Surgical Nutrition Summit. *JPEN J Parenter Enteral Nutr*. 2013;37:99S-105S.
7. Byrnes MC, Strangenes J. Refeeding in the ICU: an adult and pediatric problem. *Curr Opin Clin Nutr Metab Care*. 2011;14:186-192.

REVIEW QUESTIONS

1. A 73-year-old woman with type 2 diabetes mellitus is admitted to the medical intensive care unit (ICU) with abdominal pain on eating on occasion and a 10% weight loss from her usual weight during the past 2 months due to decreased food intake. The patient is obese on physical examination, and skeletal muscle mass cannot be assessed. Which of the following is the most appropriate regarding nutritional assessment?

A. Malnutrition cannot be assessed.
B. An involuntary weight loss of less than 15% during 2 months is not yet a signal for malnutrition.
C. The patient is likely to be malnourished.
D. Begin parenteral nutrition within 7 days.
E. Insert a feeding tube for tube feeds within 2 days of admission.

Answer: C An involuntary weight loss of 10% or more is a classic sign of nutritional inadequacy. Muscle mass cannot be assessed in obese patients accurately in most cases. Use of parenteral or enteral tube feeds for nutrition is indicated only after further evaluation of gut function and oral food tolerance. (Ziegler TR. Parenteral nutrition in the critically ill patient. *N Engl J Med*. 2009;361:1088-1097.)

2. A 23-year-old man with Crohn's disease and chronic, intermittent flares with bloody diarrhea is dependent on oral prednisone (10 mg/day) and is seen in the clinic. He exhibits evidence of skeletal muscle wasting on physical examination. Which of the following is likely to be true?

A. A combination of protein loss through stool and muscle breakdown from corticosteroids has probably contributed to muscle wasting.
B. Use of prednisone at doses of 10 mg/day is unlikely to influence catabolism.
C. Inflammation contributes to the patient's muscle loss.
D. Iron depletion is unlikely.
E. Parenteral nutrition should be started immediately.

Answer: A Protein loss in stool and steroid-induced muscle catabolism probably contributed to muscle loss in this patient. Inflammation also contributes to net body protein catabolism. Iron depletion from stool blood loss is likely. No evidence exists for immediate institution of parenteral nutrition in this setting. (Brown RO, Minard G, Ziegler TR. Parenteral nutrition. In: Ross AC, Caballero B, Cousins RJ, et al, eds. *Modern Nutrition in Health and Disease*. 11th ed. Philadelphia: Lippincott Williams & Wilkins; 2012:1136-1161.)

3. A 65-year-old man with a history of chronic obstructive pulmonary disease exhibits clinical evidence of sepsis and respiratory failure in the emergency department and is admitted to the medical ICU. He is placed on ventilator support. Gut dysfunction is present due to apparent ileus and emesis. The patient's body mass index is very low at 17 kg/m^2. Which of the following is true?

A. Parenteral nutrition should be held for at least 8 to 10 days.
B. Severely malnourished patients, as in this case, have been shown to be harmed by early parenteral nutrition begun after 4 days of ICU admission.
C. A large clinical trial suggests that some patients may have worse clinical outcomes with early (within 2 days) use of parenteral nutrition in ICU settings.
D. Tube feedings should be started, regardless of gut function.
E. Parenteral micronutrients should be initiated.

Answer: C No clinical practice guideline recommends withholding of parenteral nutrition in patients with gut dysfunction for more than 7 days. A large study showed possible harm from early parenteral nutrition in the ICU (but severely malnourished patients were excluded). Another recent trial showed no difference in mortality at 60 days or length of ICU stay or total hospital days when early parenteral nutrition is started in critically ill patients with relative contraindications to early enteral nutrition. Tube feeds should not be started with clinical evidence of gut dysfunction. Micronutrient requirements and timing for initiation in the ICU are unclear. (Casaer MP, Mesotten D, Hermans G, et al. Early versus late parenteral nutrition in critically ill adults. *N Engl J Med*. 2011;365:506-517. Doig GS, Simpson F, Sweetman EA, et al. Early parenteral nutrition in critically ill patients with short-term relative contraindications to early enteral nutrition: a randomized controlled trial. *JAMA*. 2013;309:2130-2138.)

4. A 55-year-old woman with partial small bowel obstruction precluding any food intake for 10 to 12 days presents to the emergency department and is admitted for failure to thrive. Parenteral nutrition is deemed to be needed for nutritional repletion. Which of the following is true?

A. Parenteral nutrition with a high dextrose content may cause hypophosphatemia due to insulin-induced movement of phosphorus intracellularly.
B. Enteral feeding should be instituted as it may prevent bacterial translocation from the gut.
C. Extra vitamin C should be added to parenteral nutrition as an antioxidant.
D. Parenteral nutrition should improve gut function in this patient.
E. Parenteral nutrition containing micronutrients can be initiated after 5 days.

Answer: A High dextrose parenteral nutrition can drive phosphorus intracellularly. Enteral feeds are contraindicated clinically, and there are no data that extra vitamin C or parenteral nutrition would be beneficial. Micronutrient needs in hospital patients are unknown, as is the timing of administration. (Brown RO, Minard G, Ziegler TR. Parenteral nutrition. In: Ross AC, Caballero B, Cousins RJ, et al, eds. *Modern Nutrition in Health and Disease*. 11th ed. Philadelphia: Lippincott Williams & Wilkins; 2012:1136-1161.)

5. A 20-year-old woman is admitted to the trauma ICU after a motor vehicle accident with multiple large bone fractures but no sign of gut dysfunction. She is mechanically ventilated to protect her airway and in anticipation of orthopedic surgery within several days. Which of the following is true?

A. Parenteral nutrition should be instituted.
B. Energy needs of the patient probably exceed 300% of normal.
C. An enteral feeding tube should be placed and enteral feeds initiated within the first 1 or 2 days of ICU admission.
D. Parenteral nutrition should be used to complement enteral feeds in this patient.
E. Enteral tube feeds containing adequate energy, protein, and micronutrients can be initiated after 6 days of parenteral nutrition.

Answer: A Most trauma patients can tolerate enteral nutrition, and there is no evidence of gut dysfunction in this patient. Energy needs go up with trauma, but not to 300% of normal. A feeding tube should be placed for institution of early enteral nutrition (within 2 days of admission), but use in such a non-malnourished patient early on is unclear. Optimal timing of enteral nutrition in the ICU remains unclear. (Doig GS, Simpson F, Sweetman EA, et al; Early PN Investigators of the ANZICS Clinical Trials Group. Early parenteral nutrition in critically ill patients with short-term relative contraindications to early enteral nutrition: a randomized controlled trial. *JAMA*. 2013;309:2130-2138.)

218 VITAMINS, TRACE MINERALS, AND OTHER MICRONUTRIENTS

JOEL B. MASON

MICRONUTRIENTS IN NUTRITIONAL SCIENCE

Dietary Requirements

Micronutrients are a diverse array of dietary components necessary to sustain health. The physiologic roles of micronutrients are as varied as their composition. Some micronutrients are used in enzymes as either coenzymes or prosthetic groups, others as biochemical substrates or hormones; in some instances, the functions are not well defined. Under normal circumstances, the average daily dietary intake for each micronutrient that is required to sustain normal physiologic functions is measured in milligrams or smaller quantities. In this manner, micronutrients are distinguished from macronutrients, which encompass carbohydrates, fats, and proteins as well as the macrominerals calcium, magnesium, and phosphorus.

Optimal Intake

For orderly homeostasis to proceed, most dietary nutrients must be ingested in quantities that are neither too small nor too great. Disorders may arise, therefore, when intake regularly falls outside of this physiologic window. The size of this physiologic window varies for each micronutrient and should be kept in mind, particularly in this era when the administration of large quantities of certain micronutrients is increasingly explored for possible therapeutic implications. The dietary requirement for a particular micronutrient is determined by many factors, only one of which is the amount needed to sustain those physiologic functions for which it is used (Table 218-1). The U.S. Institute of Medicine Food and Nutrition Board regularly updates dietary guidelines that define the quantity of each micronutrient that is "adequate to meet the known nutrient needs of practically all healthy persons." These *recommended dietary allowances* (RDAs) were most recently revised between 1998 and 2001, and the values for adults appear in Tables 218-2 and 218-3. Also established for the first time for each micronutrient were *tolerable upper limits* (TULs), which are the "maximal daily levels of oral intake likely to pose no adverse health risks." *Adequate intake*, the amount necessary to prevent a deficiency state, is not necessarily synonymous with *optimal intake*.

TYPES AND FUNCTION OF MICRONUTRIENTS

Vitamins

Vitamins are categorized as either fat soluble (A, D, E, K) or water soluble (all the others), as shown in Table 218-2. This categorization remains physiologically meaningful. None of the fat-soluble vitamins appears to serve as a coenzyme. Intestinal absorption of the fat-soluble vitamins is primarily through a micellar phase, and pathophysiologic conditions associated with fat malabsorption frequently are associated with selective deficiencies of the fat-soluble vitamins. In contrast, most of the functions of the water-soluble vitamins are as coenzymes, and they are not absorbed through the lipophilic phase in the intestine.

TABLE 218-1 FACTORS THAT DETERMINE DIETARY REQUIREMENT OF A MICRONUTRIENT

PHYSIOLOGIC FACTORS
Bioavailability: the proportion of a micronutrient that is ingested and is capable of being assimilated and used for physiologic purposes
Quantity required to fulfill physiologic roles
Extent to which the body can reuse the micronutrient
Distribution of nutrient in the body: storage compartments
Gender
Stage of life cycle: intrauterine development, childhood, adulthood, elder adulthood, pregnancy, lactation
PATHOPHYSIOLOGIC AND PHARMACOLOGIC FACTORS
Inborn errors of metabolism: variously affect assimilation, utilization, or excretion of micronutrients
Acquired disease states that alter the amounts required to sustain homeostasis (e.g., malabsorption, maldigestion, states that increase use)
Lifestyle habits: smoking, ethanol consumption
Drugs: may alter bioavailability or utilization

Trace Elements

Fifteen trace elements have been identified as essential for health: iron, zinc, copper, chromium, selenium, iodine, fluorine, manganese, molybdenum, cobalt, nickel, tin, silicon, vanadium, and arsenic (see Table 218-3), but only for the first 10 of these has compelling evidence indicated that they are essential nutrients in humans. Cobalt appears to be essential solely as a component of vitamin B_{12}, but an isolated deficiency state has never been described. Deficiency syndromes for several of the essential trace elements were not recognized until recently because of their exceedingly small requirements and because of the ubiquitous nature of these elements in foodstuffs. Only under exceptional circumstances, such as long-term reliance on total parenteral nutrition lacking these elements, have some of the deficiency syndromes been observed.

The biochemical functions of trace elements appear to be as components of prosthetic groups or as cofactors for enzymes. Determination of essential trace element status is problematic with the exception of iron. The low concentrations of these elements in body fluids and tissues, the finding that blood levels frequently do not correlate well with levels in the target tissues, and the fact that functional tests cannot be devised until their biochemical functions are better understood preclude an accurate laboratory method of assessing the adequacy of most trace elements.

Additional Compounds with Nutritional Relevance

Evidence indicates that humans also have an absolute requirement for the dietary component choline, which is a necessary precursor for acetylcholine and phospholipids and is needed to sustain normal levels of biologic methylation. To date, the most significant adverse effect of dietary inadequacy has been hepatic inflammation. Deficiency is nevertheless thought to be extremely rare, although pregnancy, and particularly lactation, increases the apparent requirement. Individuals whose long-term nutritional requirements are solely derived from total parenteral nutrition appear to be susceptible to choline deficiency. Both an RDA (425 mg, women; 550 mg, men) and a TUL (3.5 g) have now been established.

L-Carnitine is a dietary component that facilitates the transport of fatty acids into mitochondria, and a deficit therefore limits the fatty acid β-oxidation that occurs in those organelles.[1] Although no evidence exists for a dietary requirement in healthy children or adults, premature infants have very low stores of skeletal muscle carnitine. Therefore, preterm infants receiving parenteral nutrition without carnitine supplements appear to be a group at risk for deficiency. In clinical trials, parenteral supplementation of carnitine in such infants increases serum carnitine concentration, although it has not improved clinical end points in most studies. Similarly, it has been suggested that individuals on long-term hemodialysis and those dependent on total parenteral nutrition may also be susceptible to the clinical consequences of L-carnitine depletion, but convincing evidence is lacking.

CONDITIONS THAT INCREASE REQUIRED DIETARY INTAKE

Many physiologic, pathophysiologic, and pharmacologic factors increase the dietary requirements for micronutrients (see Table 218-1), thereby enhancing the risk for development of a deficiency state.

Physiologic Factors

Stages of the life cycle frequently have a significant impact on the requirements of nutrients. Phases of rapid growth and development, such as in utero development, infancy, adolescence, and pregnancy, are associated with increases in the utilization of certain micronutrients on a per-kilogram basis.

Pregnancy

Requirements for most micronutrients are increased in pregnancy, but, proportionately, the observed increases in the maternal requirements for iron and folate are particularly great and are related to the rapid proliferation of the placental and fetal tissues. Periods of lactation are similarly associated with remarkable increases in requirements; a lactating woman experiences disproportionately large increases in her requirements for zinc and vitamins

Text continued on p. 1452

TABLE 218-2 VITAMINS AND THEIR FUNCTIONS

	BIOCHEMISTRY AND PHYSIOLOGY	DEFICIENCY [RDA*]	TOXICITY [TUL†]	ASSESSMENT OF STATUS
FAT-SOLUBLE VITAMINS				
Vitamin A	A family of the retinoid compounds, each member having biologic activity qualitatively similar to retinol. Carotenoids are structurally related to retinoids. Some carotenoids, most notably β-carotene, are metabolized into compounds with vitamin A activity and are therefore considered to be provitamin A compounds. Vitamin A is an integral component of rhodopsin and iodopsins, light-sensitive proteins in rod and cone cells in the retina. *Additional functions:* induction and maintenance of cellular differentiation in certain tissues; signal for appropriate morphogenesis in the developing embryo; maintenance of cell-mediated immunity. 1 μg of retinol = 3.33 IU of vitamin A.	Follicular hyperkeratosis and night blindness are early indicators. Conjunctival xerosis, degeneration of the cornea (keratomalacia), and dedifferentiation of rapidly proliferating epithelia are later indications of deficiency. *Bitot spots* (focal areas of the conjunctiva or cornea with foamy appearance) are an indication of xerosis. Blindness, due to corneal destruction and retinal dysfunction, ensues if left uncorrected. Increased susceptibility to infection is also a consequence. [F: 700 μg; M: 900 μg]	In adults, >150,000 μg may cause *acute* toxicity: fatal intracranial hypertension, skin exfoliation, and hepatocellular necrosis. *Chronic* toxicity may occur with habitual daily intake of >10,000 μg: alopecia, ataxia, bone and muscle pain, dermatitis, cheilitis, conjunctivitis, pseudotumor cerebri, hepatocellular necrosis, hyperlipidemia, and hyperostosis are common. Single, large doses of vitamin A (30,000 μg) or habitual intake of >4500 μg/day in early pregnancy can be teratogenic. Excessive intake of carotenoids causes a benign condition characterized by yellowish discoloration of the skin. Habitually large doses of canthaxanthin, a carotenoid, have the additional capability of inducing a retinopathy. [3000 μg]	Retinol concentration in the plasma and vitamin A concentrations in the milk and tears are reasonably accurate measures of adequate status. Toxicity is best assessed by elevated levels of retinyl esters in plasma. A quantitative measure of dark adaptation for night vision and electroretinography are useful functional tests.
Vitamin D	A group of sterol compounds whose parent structure is cholecalciferol (vitamin D_3) Cholecalciferol is formed in the skin from 7-dehydrocholesterol (provitamin D_3) by exposure to UVB radiation. A plant sterol, ergocalciferol (provitamin D_2), can be similarly converted into vitamin D_2 and has similar vitamin D activity. The vitamin undergoes sequential hydroxylations in the liver and kidney at the 25 and 1 positions, respectively, producing the most bioactive form of the vitamin, 1,25-dihydroxy vitamin D. Vitamin D maintains intracellular and extracellular concentrations of calcium and phosphate by enhancing intestinal absorption of the two ions and, in conjunction with PTH, promoting their mobilization from bone mineral. It retards proliferation and promotes differentiation in certain epithelia. Purported actions of vitamin D as an anti-diabetes, anti-inflammatory, and cancer preventive agent remain controversial and are under investigation. 1 μg = 40 IU.	Deficiency results in decreased mineralization of newly formed bone called *rickets* in childhood and *osteomalacia* in adults. Expansion of the epiphyseal growth plates and replacement of normal bone with unmineralized bone matrix are the cardinal features of rickets; the latter feature also characterizes osteomalacia. Deformity of bone and pathologic fractures occur. Decreased serum concentrations of calcium and phosphate may occur. [15 μg, ages 19-70 yr; 20 μg, age >70 yr]	Excess amounts result in abnormally high concentrations of calcium and phosphate in the serum; metastatic calcifications, renal damage, and altered mentation may occur. [100 μg for ages ≥9 yr]	The serum concentration of the major circulating metabolite, 25-hydroxyvitamin D, is the best indicator of systemic status except in advanced kidney disease (stages 4 and 5), in which the impairment of renal 1-hydroxylation results in disassociation of the mono- and dihydroxyvitamin concentrations. Measurement of the serum concentration of 1,25-dihydroxyvitamin D is then necessary.
Vitamin E	A group of at least 8 naturally occurring compounds, some of which are tocopherols and some of which are tocotrienols At present, the only dietary form that is thought to be biologically active in humans is α-tocopherol. Vitamin E acts as an antioxidant and free radical scavenger in lipophilic environments, most notably in cell membranes. It acts in conjunction with other antioxidants, such as selenium.	Deficiency due to dietary inadequacy is rare. It is usually seen in premature infants, individuals with fat malabsorption, and individuals with abetalipoproteinemia. Red blood cell fragility occurs and can produce a hemolytic anemia. Neuronal degeneration produces peripheral neuropathies, ophthalmoplegia, and destruction of posterior columns of spinal cord. Neurologic disease is frequently irreversible if deficiency is not corrected early enough. May contribute to the hemolytic anemia and retrolental fibroplasia seen in premature infants. Reported to suppress cell-mediated immunity [15 mg]	Depressed levels of vitamin K–dependent procoagulants and potentiation of oral anticoagulants have been reported, as has impaired WBC function. Doses of 800 mg/day have been reported to increase slightly the incidence of hemorrhagic stroke. [1000 mg]	Plasma or serum concentration of α-tocopherol is most commonly used. Additional accuracy is obtained by expressing this value per milligram of total plasma lipid. RBC peroxide hemolysis test is not entirely specific but is a useful functional measure of the antioxidant potential of cell membranes.

TABLE 218-2 VITAMINS AND THEIR FUNCTIONS—cont'd

	BIOCHEMISTRY AND PHYSIOLOGY	DEFICIENCY [RDA*]	TOXICITY [TUL†]	ASSESSMENT OF STATUS
Vitamin K	A family of naphthoquinone compounds with similar biologic activity. Phylloquinone (vitamin K_1) is derived from plants; a variety of menaquinones (vitamin K_2) are derived from bacterial and animal sources. Vitamin K serves as an essential cofactor in the post-translational γ-carboxylation of glutamic acid residues in many proteins. These proteins include several circulating procoagulants and anticoagulants as well as proteins in a variety of tissues.	Deficiency syndrome is uncommon except in breast-fed newborns, in whom it may cause "hemorrhagic disease of the newborn"; in adults with fat malabsorption or who are taking drugs that interfere with vitamin K metabolism (e.g., coumarin, phenytoin, broad-spectrum antibiotics); and in individuals taking large doses of vitamin E and anticoagulant drugs. Excessive hemorrhage is the usual manifestation. [F: 90 µg; M: 120 µg]	Rapid intravenous infusion of K_1 has been rarely associated with dyspnea, flushing, and cardiovascular collapse; this is likely related to the dispersing agents in the solution. Supplementation may interfere with coumarin-based anticoagulation. Pregnant women taking large amounts of the provitamin menadione may deliver infants with hemolytic anemia, hyperbilirubinemia, and kernicterus. [no TUL established]	Prothrombin time is typically used as a measure of functional K status; it is neither sensitive nor specific for vitamin K deficiency. Determination of fasting plasma vitamin K is an accurate indicator of status. Undercarboxylated plasma prothrombin is also an accurate metric, but only for detecting the deficient state, and is less widely available than plasma vitamin K.
WATER-SOLUBLE VITAMINS				
Thiamin (vitamin B_1)	A water-soluble compound containing substituted pyrimidine and thiazole rings and a hydroxyethyl side chain. The coenzyme form is thiamin pyrophosphate (TPP). Thiamin serves as a coenzyme in many α-ketoacid decarboxylation and transketolation reactions. Inadequate thiamin availability leads to impairments of these reactions, resulting in inadequate adenosine triphosphate synthesis and abnormal carbohydrate metabolism, respectively. It may have an additional role in neuronal conduction independent of the aforementioned actions.	Classic deficiency syndrome (beriberi) is described in Asian populations consuming a polished rice diet. Alcoholism, chronic renal dialysis, and persistent nausea and vomiting after bariatric surgery are also common precipitants. High carbohydrate intake increases need for B_1. *Mild deficiency*: irritability, fatigue, and headaches *More severe deficiency*: combinations of peripheral neuropathy, cardiovascular dysfunction, and cerebral dysfunction. Cardiovascular involvement (wet beriberi): congestive heart failure and low peripheral vascular resistance. Cerebral disease: nystagmus, ophthalmoplegia, and ataxia (Wernicke's encephalopathy); hallucinations, impaired short-term memory, and confabulation (Korsakoff's psychosis) Deficiency syndrome responds within 24 hr to parenteral thiamin but is partially or wholly irreversible after a certain stage. [F: 1.1 mg; M: 1.2 mg]	Excess intake is largely excreted in the urine, although parenteral doses of >400 mg/day are reported to cause lethargy, ataxia, and reduced tone of the gastrointestinal tract. [TUL not established]	The most effective measure of B_1 status is the erythrocyte transketolase activity coefficient, which measures enzyme activity before and after addition of exogenous TPP; RBCs from a deficient individual express a substantial increase in enzyme activity with addition of TPP. Thiamin concentrations in blood or urine are also used.
Riboflavin (vitamin B_2)	Consists of a substituted isoalloxazine ring with a ribitol side chain Riboflavin serves as a coenzyme for a diverse array of biochemical reactions. The primary coenzymatic forms are flavin mononucleotide and flavin adenine dinucleotide. Riboflavin holoenzymes participate in oxidation-reduction reactions in myriad metabolic pathways.	Deficiency is usually seen in conjunction with deficiencies of other B vitamins. Isolated deficiency of riboflavin produces hyperemia and edema of nasopharyngeal mucosa, cheilosis, angular stomatitis, glossitis, seborrheic dermatitis, and a normochromic, normocytic anemia. [F: 1.1; M: 1.3]	Toxicity is not reported in humans. [TUL not established]	The most common method of assessment is to determine the activity coefficient of glutathione reductase in RBCs (the test is invalid for individuals with glucose-6-phosphate dehydrogenase deficiency). Measurements of blood and urine concentrations are less desirable methods.

TABLE 218-2 VITAMINS AND THEIR FUNCTIONS—cont'd

	BIOCHEMISTRY AND PHYSIOLOGY	DEFICIENCY [RDA*]	TOXICITY [TUL†]	ASSESSMENT OF STATUS
Niacin (vitamin B_3)	Refers to nicotinic acid and the corresponding amide, nicotinamide The active coenzymatic forms are composed of nicotinamide affixed to adenine dinucleotide, forming NAD or NADP. More than 200 apoenzymes use these compounds as electron acceptors or hydrogen donors, either as a coenzyme or as a co-substrate. The essential amino acid tryptophan is a precursor of niacin; 60 mg of dietary tryptophan yields approximately 1 mg of niacin. Dietary requirements thus depend partly on tryptophan intake. Requirement is often determined on basis of calorie intake (i.e., niacin equivalents/1000 kcal). Large doses of nicotinic acid (1.5-3 g/day) effectively lower low-density lipoprotein cholesterol and elevate high-density lipoprotein cholesterol.	*Pellagra* is the classic deficiency syndrome and is often seen in populations in which corn is the major source of energy; it is still endemic in parts of China, Africa, and India. Diarrhea, dementia (or associated symptoms of anxiety or insomnia), and a pigmented dermatitis that develops in sun-exposed areas are typical features. Glossitis, stomatitis, vaginitis, vertigo, and burning dysesthesias are early signs. It is reported occasionally to occur in carcinoid syndrome because tryptophan is diverted to other synthetic pathways. [F: 14 mg; M: 16 mg]	Human toxicity is known largely through studies examining hypolipidemic effects. Includes vasomotor phenomenon (flushing), hyperglycemia, parenchymal liver damage, and hyperuricemia. [35 mg]	Assessment of status is problematic; blood levels of the vitamin are not reliable. Measurement of urinary excretion of the niacin metabolites *N*-methylnicotinamide and 2-pyridone is thought to be the most effective means of assessment at present.
Vitamin B_6	Refers to several derivatives of pyridine, including pyridoxine, pyridoxal, and pyridoxamine, which are interconvertible in the body. The coenzymatic forms are pyridoxal-5-phosphate (PLP) and pyridoxamine-5-phosphate. As a coenzyme, B_6 is involved in many transamination reactions (and thereby in gluconeogenesis), in the synthesis of niacin from tryptophan, in the synthesis of several neurotransmitters, and in the synthesis of δ-aminolevulinic acid (and therefore in heme synthesis). It also has functions unrelated to coenzymatic activity: pyridoxal and PLP bind to hemoglobin and alter oxygen affinity; PLP also binds to steroid receptors, inhibiting receptor affinity to DNA and thereby modulating steroid activity.	Deficiency is usually seen in conjunction with other water-soluble vitamin deficiencies. Stomatitis, angular cheilosis, glossitis, irritability, depression, and confusion occur in moderate to severe depletion; normochromic, normocytic anemia has been reported in severe deficiency. Abnormalities on electroencephalography and, in infants, convulsions have also been observed. Some sideroblastic anemias respond to B_6 administration. Isoniazid, cycloserine, penicillamine, ethanol, and theophylline can inhibit B_6 metabolism. [Ages 19-50 yr: 1.3 mg; >50 yr: 1.5 mg for women, 1.7 mg for men]	Long-term use with doses exceeding 200 mg/day (in adults) may cause peripheral neuropathies and photosensitivity. [100 mg]	Many useful laboratory methods of assessment exist. The plasma or erythrocyte PLP levels are most common. Urinary excretion of xanthurenic acid after an oral tryptophan load and activity indices of RBC alanine or aspartate transaminase are functional measures of B_6-dependent enzyme activity.
Folate	A group of related pterin compounds More than 35 forms of the vitamin are found naturally. The fully oxidized form, folic acid, is not found in nature but is the pharmacologic form of the vitamin. All folate functions relate to its ability to transfer one-carbon groups. It is essential in the de novo synthesis of nucleotides and in the metabolism of several amino acids; it is an integral component for the regeneration of the "universal" methyl donor, *S*-adenosylmethionine. Inhibition of bacterial and cancer cell folate metabolism is the basis for the sulfonamide antibiotics and chemotherapeutic agents, such as methotrexate and 5-fluorouracil, respectively.	Women of childbearing age are most likely to be deficient. *Classic deficiency syndrome*: megaloblastic anemia, diarrhea. The hematopoietic cells in bone marrow become enlarged and have immature nuclei, reflecting ineffective DNA synthesis. The peripheral blood smear demonstrates macro-ovalocytes and polymorphonuclear leukocytes with an average of more than 3.5 nuclear lobes. Megaloblastic changes also occur in other epithelia that proliferate rapidly (e.g., oral mucosa and gastrointestinal tract, producing glossitis and diarrhea, respectively). Sulfasalazine and diphenytoin inhibit absorption and predispose to deficiency. [400 μg of dietary folate equivalents (DFE); 1 DFE = 1 μg food folate = 0.6 μg folic acid]	Doses >1000 μg/day may partially correct the anemia of B_{12} deficiency and may therefore mask (and perhaps exacerbate) the associated neuropathy. Large doses are also reported to lower seizure threshold in individuals prone to seizures. Parenteral administration is rarely reported to cause allergic phenomena, which is probably due to dispersion agents. [1000 μg]	Serum folate measures short-term folate balance, whereas RBC folate is a better reflection of tissue status. Serum homocysteine rises early in deficiency but is nonspecific because B_{12} or B_6 deficiency, renal insufficiency, and older age may also cause elevations.

TABLE 218-2 VITAMINS AND THEIR FUNCTIONS—cont'd

	BIOCHEMISTRY AND PHYSIOLOGY	DEFICIENCY [RDA*]	TOXICITY [TUL†]	ASSESSMENT OF STATUS
Vitamin C (ascorbic and dehydroascorbic acid)	Ascorbic acid readily oxidizes to dehydroascorbic acid in aqueous solution. Dehydroascorbic acid can be reduced in vivo, so it possesses vitamin C activity. Total vitamin C is therefore the sum of ascorbic and dehydroascorbic acid content. Vitamin C serves primarily as a biologic antioxidant in aqueous environments. Biosyntheses of collagen, carnitine, bile acids, and norepinephrine as well as proper functioning of the hepatic mixed-function oxygenase system depends on this property. Vitamin C in foodstuffs increases the intestinal absorption of nonheme iron.	Overt deficiency is uncommon in developed countries. The classic deficiency syndrome is *scurvy:* fatigue, depression, and widespread abnormalities in connective tissues, such as inflamed gingivae, petechiae, perifollicular hemorrhages, impaired wound healing, coiled hairs, hyperkeratosis, and bleeding into body cavities. In infants, defects in ossification and bone growth may occur. Tobacco smoking lowers plasma and leukocyte vitamin C levels. [F: 75 mg; M: 90 mg; increase requirement for cigarette smokers by 35 mg/day]	≥500 mg/day (in adults) may cause nausea and diarrhea. >1 g/day modestly increases risk for oxalate kidney stones. Supplementation may interfere with laboratory tests based on redox potential (e.g., fecal occult blood testing, serum cholesterol, and glucose). Withdrawal from chronic ingestion of high doses of vitamin C supplements should be done gradually because accommodation appears to occur, raising a concern of "rebound scurvy." [2 g]	Plasma ascorbic acid concentration reflects recent dietary intake, whereas WBC levels more closely reflect tissue stores. Women's plasma levels are approximately 20% higher than men's for any given dietary intake.
Vitamin B_{12}	A group of closely related cobalamin compounds composed of a corrin ring (with a cobalt atom in its center) connected to a ribonucleotide through an aminopropanol bridge. Microorganisms are the ultimate source of all naturally occurring B_{12}. The two active coenzyme forms are deoxyadenosylcobalamin and methylcobalamin. These coenzymes are needed for the synthesis of succinyl CoA, which is essential in lipid and carbohydrate metabolism, and for the synthesis of methionine. The synthesis of methionine is essential for amino acid metabolism, for purine and pyrimidine synthesis, for many methylation reactions, and for the intracellular retention of folates.	Dietary inadequacy is a rare cause of deficiency except in strict vegetarians. Most deficiencies arise from loss of intestinal absorption, which may occur with pernicious anemia, pancreatic insufficiency, atrophic gastritis, small bowel bacterial overgrowth, or ileal disease. Megaloblastic anemia and megaloblastic changes in other epithelia (see Folate) are the result of sustained depletion. Demyelination of peripheral nerves, posterior and lateral columns of spinal cord, and nerves within the brain may occur. Altered mentation, depression, and psychoses occur. Hematologic and neurologic complications may occur independently. Folate supplementation, in doses of 1000 μg/day, may partly correct the anemia, thereby masking (or perhaps exacerbating) the neuropathic complication. [2.4 μg]	A few allergic reactions have been reported to crystalline B_{12} preparations and are probably due to impurities, not the vitamin. [TUL not established]	Serum or plasma concentrations are generally accurate. Subtle deficiency with neurologic complications, as described in the Deficiency column, can best be established by concurrently measuring the concentration of plasma B_{12} and serum methylmalonic acid, which is a sensitive indicator of cellular deficiency.
Biotin	A bi-cyclic compound consisting of a ureido ring fused to a substituted tetrahydrothiophene ring. Endogenous synthesis by intestinal flora may contribute significantly to biotin nutriture. Most dietary biotin is linked to lysine, a compound called biotinyl lysine, or biocytin. The lysine must be hydrolyzed by an intestinal enzyme called biotinidase before intestinal absorption occurs. Biotin acts primarily as a coenzyme for several carboxylases; each holoenzyme catalyzes an adenosine triphosphate–dependent carbon dioxide transfer. The carboxylases are critical enzymes in carbohydrate and lipid metabolism.	Isolated deficiency is rare. Deficiency in humans has been produced by prolonged total parenteral nutrition lacking the vitamin and by ingestion of large quantities of raw egg white, which contains avidin, a protein that binds biotin with such high affinity that it renders it biounavailable. Alterations in mental status, myalgias, hyperesthesias, and anorexia occur. Later, a seborrheic dermatitis and alopecia develop. Deficiency is usually accompanied by lactic acidosis and organic aciduria. [30 μg]	Toxicity has not been reported in humans with doses as high as 60 mg/day in children. [TUL not established]	Plasma and urine concentrations of biotin are diminished in the deficient state. Elevated urine concentrations of methyl citrate, 3-methylcrotonylglycine, and 3-hydroxyisovalerate are also observed in deficiency.

TABLE 218-2 VITAMINS AND THEIR FUNCTIONS—cont'd

	BIOCHEMISTRY AND PHYSIOLOGY	DEFICIENCY [RDA*]	TOXICITY [TUL†]	ASSESSMENT OF STATUS
Pantothenic acid	Consists of pantoic acid linked to β-alanine through an amide bond Pantothenic acid is an essential component of CoA and phosphopantetheine, which are essential for synthesis and β-oxidation of fatty acids as well as for synthesis of cholesterol, steroid hormones, vitamins A and D, and other isoprenoid derivatives. CoA is also involved in the synthesis of several amino acids and δ-aminolevulinic acid, a precursor for the corrin ring of vitamin B_{12}, the porphyrin ring of heme, and of cytochromes. CoA is also necessary for the acetylation and fatty acid acylation of a variety of proteins.	Deficiency is rare; it has been reported only as a result of feeding of semisynthetic diets or an antagonist to the vitamin. Experimental, isolated deficiency in humans produces fatigue, abdominal pain, vomiting, insomnia, and paresthesias of the extremities. [5 mg]	In doses of 10 g/day, diarrhea is reported to occur. [TUL not established]	Whole blood and urine concentrations of pantothenate are indicators of status; serum levels are not thought to be accurate.

CoA = coenzyme A; NAD = nicotinamide adenine dinucleotide; NADP = nicotinamide adenine dinucleotide phosphate; PTH = parathyroid hormone; RBC = red blood cell; UVB = ultraviolet B; WBC = white blood cell.
*Recommended daily allowance (RDA) established for female (F) and male (M) adults by the U.S. Food and Nutrition Board, 1999-2001. In some instances, insufficient data exist to establish an RDA, in which case the adequate intake (AI) established by the board is listed.
†Tolerable upper limit (TUL) established for adults by the U.S. Food and Nutrition Board, 1999-2001.

TABLE 218-3 NUTRITIONAL TRACE ELEMENTS AND THEIR CLINICAL IMPLICATIONS

	BIOCHEMISTRY AND PHYSIOLOGY	DEFICIENCY [RDA*]	TOXICITY [TUL†]	ASSESSMENT OF STATUS
Chromium	Dietary chromium consists of both inorganic and organic forms. Its primary function in humans is to potentiate insulin action. It accomplishes this function as a circulating complex called *glucose tolerance factor*, thereby affecting carbohydrate, fat, and protein metabolism.	Deficiency in humans has been described only in long-term total parenteral nutrition (TPN) patients receiving insufficient chromium. Hyperglycemia or impaired glucose tolerance occurs. Elevated plasma free fatty acid concentrations, neuropathy, encephalopathy, and abnormalities in nitrogen metabolism are also reported. Whether supplemental chromium may improve glucose tolerance in glucose-intolerant individuals remains controversial. [F: 25 μg; M: 35 μg]	Toxicity after oral ingestion is uncommon and seems confined to gastric irritation. Airborne exposure may cause contact dermatitis, eczema, skin ulcers, and bronchogenic carcinoma. [no TUL established]	Plasma or serum concentration of chromium is a crude indicator of chromium status; it appears to be meaningful when the value is markedly above or below the normal range.
Copper	Copper is absorbed by a specific intestinal transport mechanism. It is carried to the liver, where it is bound to ceruloplasmin, which circulates systemically and delivers copper to target tissues in the body. Excretion of copper is largely through bile and then into the feces. Absorptive and excretory processes vary with the levels of dietary copper, providing a means of copper homeostasis. Copper serves as a component of many enzymes, including amine oxidases, ferroxidases, cytochrome *c* oxidase, dopamine β-hydroxylase, superoxide dismutase, and tyrosinase.	Dietary deficiency is rare; it has been observed in premature and low-birthweight infants fed exclusively a cow's milk diet and in individuals on long-term TPN lacking copper. It has also been described after gastric bypass surgery and with chronic zinc supplementation. Clinical manifestations include depigmentation of skin and hair, myelopathy and other neurologic lesions, leukopenia, anemia, and skeletal abnormalities. Anemia arises from impaired utilization of iron and therefore often is manifested as a sideroblastic anemia. The peripheral smear and bone marrow may mimic myelodysplasia. A deficiency syndrome is also observed in Menkes' disease, a rare inherited condition associated with impaired copper utilization. [900 μg]	Acute copper toxicity has been described after excessive oral intake and with absorption of copper salts applied to burned skin. Milder manifestations include nausea, vomiting, epigastric pain, and diarrhea; coma and hepatic necrosis may ensue in severe cases. Toxicity may be seen with doses as low as 70 μg/kg/day. Chronic toxicity is also described. Wilson's disease is a rare, inherited disease associated with abnormally low ceruloplasmin levels and accumulation of copper in the liver and brain, eventually leading to damage to these two organs. [10 mg]	Practical methods to detect marginal deficiency are not available. Marked deficiency is reliably detected by diminished serum copper and ceruloplasmin concentrations as well as by low red blood cell superoxide dismutase activity.

TABLE 218-3 NUTRITIONAL TRACE ELEMENTS AND THEIR CLINICAL IMPLICATIONS—cont'd

	BIOCHEMISTRY AND PHYSIOLOGY	DEFICIENCY [RDA*]	TOXICITY [TUL†]	ASSESSMENT OF STATUS
Fluorine	Known more commonly by its ionic form, fluoride. Fluorine is incorporated into the crystalline structure of bone, thereby altering its physical characteristics.	Intake of <0.1 mg/day in infants and <0.5 mg/day in children is associated with an increased incidence of dental caries. Optimal intake in adults is between 1.5 and 4 mg/day. [F: 3 mg; M: 4 mg]	Acute ingestion of >30 mg/kg body weight is likely to cause death. Excessive chronic intake (0.1 mg/kg/day) leads to mottling of teeth (dental fluorosis), calcification of tendons and ligaments, and exostoses and may increase the brittleness of bones. [10 mg]	Estimates of intake and clinical assessment are used because no good laboratory test exists.
Iodine	Iodine is readily absorbed from the diet, concentrated in the thyroid, and integrated into the thyroid hormones thyroxine and triiodothyronine. These hormones circulate largely bound to thyroxine-binding globulin. They modulate resting energy expenditure and, in the developing human, growth and development.	In the absence of supplementation, populations relying primarily on food from soils with low iodine content have endemic iodine deficiency. Maternal iodine deficiency leads to fetal deficiency, which produces spontaneous abortions, stillbirths, hypothyroidism, cretinism, and dwarfism. Permanent cognitive deficits may result from iodine deficiency during the first 2 years of life. In the adult, compensatory hypertrophy of the thyroid (goiter) occurs along with varying degrees of hypothyroidism. [150 μg]	Large doses (>2 mg/day in adults) may induce hypothyroidism by blocking thyroid hormone synthesis. Supplementation with >100 mg/day to an individual who was formerly deficient occasionally induces hyperthyroidism. [1.1 mg]	Iodine status of a population can be estimated by the prevalence of goiter. Urinary excretion of iodine is an effective laboratory means of assessment. Thyroid-stimulating hormone blood level is an indirect and therefore not entirely specific means of assessment.
Iron	Conveys the capacity to participate in redox reactions to a number of metalloproteins, such as hemoglobin, myoglobin, cytochrome enzymes, and many oxidases and oxygenases. The primary storage form of iron is ferritin and, to a lesser degree, hemosiderin. Intestinal absorption is 15-20% for "heme" iron and 1-8% for iron contained in vegetables. Absorption of the latter form is enhanced by the ascorbic acid in foodstuffs; by poultry, fish, or beef; and by an iron-deficient state. It is decreased by phytate and tannins.	Iron deficiency is the most common micronutrient deficiency in the world. Women of childbearing age are the group at highest risk because of menstrual blood losses, pregnancy, and lactation. The classic deficiency syndrome is hypochromic, microcytic anemia. Glossitis and koilonychia ("spoon" nails) are also observed. Easy fatigability often is an early symptom, before anemia appears. In children, mild deficiency of insufficient severity to cause anemia is associated with behavioral disturbances and poor school performance. [postmenopausal F and M: 8 mg; premenopausal F: 18 mg]	Iron overload typically occurs when habitual dietary intake is extremely high, intestinal absorption is excessive, repeated parenteral administration occurs, or a combination of these factors exists. Excessive iron stores usually accumulate in the reticuloendothelial tissues and cause little damage (hemosiderosis). If overload continues, iron eventually begins to accumulate in tissues such as the hepatic parenchyma, pancreas, heart, and synovium, causing hemochromatosis (Chapter 212) Hereditary hemochromatosis results from homozygosity of a common recessive trait. Excessive intestinal absorption of iron is seen in homozygotes. [45 mg]	Negative iron balance initially leads to depletion of iron stores in the bone marrow; a bone marrow biopsy and the concentration of serum ferritin are accurate indicators of early depletion. As the severity of deficiency proceeds, serum iron (SI) decreases and total iron-binding capacity (TIBC) increases; an iron saturation (SI/TIBC) of <16% suggests iron deficiency. Microcytosis, hypochromia, and anemia ensue. Elevated levels of serum ferritin or an iron saturation of >60% suggest iron overload, although systemic inflammation elevates serum ferritin regardless of iron status.
Manganese	A component of several metalloenzymes. Most manganese is in mitochondria, where it is a component of manganese superoxide dismutase.	Manganese deficiency in the human has not been conclusively demonstrated. It is said to cause hypocholesterolemia, weight loss, hair and nail changes, dermatitis, and impaired synthesis of vitamin K–dependent proteins. [F: 1.8 mg; M: 2.3 mg]	Toxicity by oral ingestion is unknown in humans. Toxic inhalation causes hallucinations, other alterations in mentation, and extrapyramidal movement disorders. [11 mg]	Until the deficiency syndrome is better defined, an appropriate measure of status will be difficult to develop.
Molybdenum	A cofactor in several enzymes, most prominently xanthine oxidase and sulfite oxidase	A probable case of human deficiency is described as being secondary to parenteral administration of sulfite and resulted in hyperoxypurinemia, hypouricemia, and low sulfate excretion. [45 μg]	Toxicity not well described in humans, although it may interfere with copper metabolism at high doses. [2 mg]	Laboratory means of assessment are not meaningful until the deficiency syndrome is better described.

TABLE 218-3 NUTRITIONAL TRACE ELEMENTS AND THEIR CLINICAL IMPLICATIONS—cont'd

	BIOCHEMISTRY AND PHYSIOLOGY	DEFICIENCY [RDA*]	TOXICITY [TUL†]	ASSESSMENT OF STATUS
Selenium	Most dietary selenium is in the form of an amino acid complex. Nearly complete absorption of such forms occurs. Homeostasis is largely performed by the kidney, which regulates urinary excretion as a function of selenium status. Selenium is a component of several enzymes, most notably glutathione peroxidase and superoxide dismutase. These enzymes protect against oxidative and free radical damage of various cell structures. The antioxidant protection conveyed by selenium apparently operates in conjunction with vitamin E because deficiency of one seems to potentiate damage induced by a deficiency of the other. Selenium also participates in the enzymatic conversion of thyroxine to its more active metabolite, triiodothyronine.	Deficiency is rare in North America but has been observed in individuals on long-term TPN lacking selenium. Such individuals have myalgias or cardiomyopathies. Populations in some regions of the world, most notably some parts of China, have marginal intake of selenium. In these regions *Keshan's disease*, a condition characterized by cardiomyopathy, is endemic; it can be prevented (but not treated) by selenium supplementation. [55 µg]	Toxicity is associated with nausea, diarrhea, alterations in mental status, peripheral neuropathy, and loss of hair and nails; such symptoms were observed in adults who inadvertently consumed 27-2400 mg. [400 µg]	Erythrocyte glutathione peroxidase activity and plasma or whole blood selenium concentrations are the most commonly used methods of assessment. They are moderately accurate indicators of status.
Zinc	Intestinal absorption occurs by a specific process that is enhanced by pregnancy and corticosteroids and diminished by coingestion of phytates, phosphates, iron, copper, lead, or calcium. Diminished intake of zinc leads to an increased efficiency of absorption and decreased fecal excretion, providing a means of zinc homeostasis. Zinc is a component of more than 100 enzymes, among which are DNA polymerase, RNA polymerase, and transfer RNA synthetase.	Zinc deficiency has its most profound effect on rapidly proliferating tissues. *Mild deficiency:* growth retardation in children. *More severe deficiency:* growth arrest, teratogenicity, hypogonadism and infertility, dysgeusia, poor wound healing, diarrhea, dermatitis on the extremities and around orifices, glossitis, alopecia, corneal clouding, loss of dark adaptation, and behavioral changes. Impaired cellular immunity is observed. Excessive loss of gastrointestinal secretions through chronic diarrhea and fistulas may precipitate deficiency. *Acrodermatitis enteropathica* is a rare, recessively inherited disease in which intestinal absorption of zinc is impaired. [F: 8 mg; M: 11 mg]	Acute zinc toxicity can usually be induced by ingestion of >200 mg of zinc in a single day (in adults). It is manifested by epigastric pain, nausea, vomiting, and diarrhea. Hyperpnea, diaphoresis, and weakness may follow inhalation of zinc fumes. Copper and zinc compete for intestinal absorption: long-term ingestion of >25 mg/day of zinc may lead to copper deficiency. Long-term ingestion of >150 mg/day has been reported to cause gastric erosions, low high-density lipoprotein cholesterol levels, and impaired cellular immunity. [40 mg]	No accurate indicators of zinc status exist for routine clinical use. Plasma, red blood cell, and hair zinc concentrations are often misleading. Acute illness, in particular, is known to diminish plasma zinc levels, in part by inducing a shift of zinc out of the plasma compartment and into the liver. Functional tests that determine dark adaptation, taste acuity, and rate of wound healing lack specificity.

*Recommended daily allowance (RDA) established for female (F) and male (M) adults by the U.S. Food and Nutrition Board, 1999-2001. In some instances, insufficient data exist to establish an RDA, in which case the adequate intake (AI) established by the board is listed.
†Tolerable upper limit (TUL) established for adults by the U.S. Food and Nutrition Board, 1999-2001.

A, E, and C to meet the metabolic demands incurred by milk production in addition to the aforementioned needs observed in pregnancy.

Aside from its general role in supporting the rapid proliferation of placental and fetal tissues, folate plays a specific role in the prevention of particular birth defects. A 20 to 85% reduction in births complicated by neural tube defects (NTDs, i.e., spina bifida and anencephaly) has been realized by providing women with a daily supplement of folic acid in the form of supplements or fortified foods. The optimal dose is not well defined, but 200 to 400 µg/day clearly affords a substantial degree of protection. Populations with a high background rate of NTD births attain the largest reductions in NTDs from supplemental folate. However, because the nascent neural tube closes about day 20 after conception, the additional folate must be provided before this time to be effective.

Infancy

Infancy carries particular vulnerabilities to specific micronutrient inadequacies. Healthy infants in the United States are typically supplemented with vitamin K at birth and with iron and vitamin D during the course of the first year because of their particular susceptibility to deficiencies of these nutrients.

Women of Childbearing Age

The ability to maintain adequate iron status from menarche through menopause is compromised in women by the additional losses incurred by menstruation, pregnancy, and lactation. Therefore, it is not surprising that the population subset that almost invariably displays the highest rate of iron deficiency is women of childbearing age.

Elderly Persons

Specific dietary recommendations for elderly people have been formally incorporated into the recommended dietary allowances (RDA) because aging has an impact on the need for certain micronutrients. Vitamin B_{12} status declines significantly with aging, in large part because of the high prevalence of atrophic gastritis and its associated impairment in protein-bound vitamin B_{12} absorption.[2] Estimates suggest that 10 to 20% of the elderly population is at risk for clinically significant vitamin B_{12} deficiency. Consequently, elderly persons should consume some of their vitamin B_{12} requirement in the crystalline form rather than solely from the naturally occurring protein-bound forms found in food because absorption of the crystalline form is not impaired by atrophic gastritis. Elderly people also require greater quantities of vitamins B_6 and D to maintain health compared with younger adults, as reflected in

the new RDAs (see Table 218-2). For instance, the RDA of vitamin D in persons older than 70 years is now set at 20 μg/day (800 IU), as opposed to adults who are 70 years of age or younger, whose RDA is 15 μg/day.[3,4] This increased need appears to result from diminished cutaneous synthesis of vitamin D by senile skin and from decreased sun exposure, which appears to be particularly important in elders residing in institutional facilities. The need for crystalline vitamin B_{12} and for a quantity of vitamin D that is difficult to achieve without resorting to a supplement suggests that universal use of a daily supplement pill containing these nutrients would benefit elderly people. Widespread use of a multivitamin that contains a broad spectrum of micronutrients is more controversial, in part because of concerns about subtle toxicity. For example, elders with chronic renal failure appear to have a vulnerability to vitamin A toxicity, suggesting that use of supplements containing this vitamin is contraindicated.

PATHOPHYSIOLOGIC AND PHARMACOLOGIC FACTORS

Diseases of the Gastrointestinal Tract

Malabsorption and maldigestion predispose to multiple micronutrient deficiencies. Both fat- and water-soluble micronutrients (except vitamin B_{12}) are absorbed predominantly in the proximal small intestine. Therefore, diffuse mucosal diseases affecting the proximal portion of the gastrointestinal tract are likely to result in deficiencies. Even in the absence of mucosal disease of the proximal small intestine, extensive ileal disease, small bowel bacterial overgrowth, and chronic cholestasis can each interfere with the maintenance of adequate intraluminal conjugated bile acid concentrations and thereby impair absorption of fat-soluble vitamins. Maldigestion is usually the result of chronic pancreatitis. Untreated, it frequently causes malabsorption and deficiencies of fat-soluble vitamins. Vitamin B_{12} malabsorption can often be demonstrated in this setting, a result of inadequate R-protein digestion, but clinical vitamin B_{12} deficiency is rarely reported.

Inborn Errors of Metabolism

Myriad rare inborn errors of metabolism have been described for vitamins and minerals that impair an individual's ability to assimilate, to use, or to retain a particular micronutrient (Chapter 205). Such defects are usually partial and can often be overcome, to a certain extent, by administering doses of the nutrient that are several orders of magnitude greater than usually required. Suspicion for such defects should be entertained if a known defect exists in the family, a deficiency syndrome arises at birth or during infancy, or the deficiency syndrome is present despite adequate dietary intake and the absence of any disease that would impair the ability to assimilate the nutrient.

Medications

Long-term administration of many drugs may adversely affect micronutrient status. The manner in which drug-nutrient interactions occur varies; some of the more common mechanisms are outlined in Table 218-4. Some drugs exert their therapeutic effects by specifically inhibiting the actions of a micronutrient. Examples include coumarin, which inhibits γ-carboxylation reactions mediated by vitamin K, and methotrexate, which binds tightly to dihydrofolate reductase, thereby inhibiting folate metabolism.

TABLE 218-4 DRUG-MEDIATED EFFECTS ON MICRONUTRIENT STATUS: EXAMPLES

DRUG	NUTRIENT	MECHANISM OF INTERACTION
Dextroamphetamine, fenfluramine, levodopa	Potentially all micronutrients	Induces anorexia
Cholestyramine	Vitamin D, folate	Adsorbs nutrient, decreases absorption
Omeprazole	Vitamin B_{12}	Modest bacterial overgrowth, decreases gastric acid, impairs absorption
Sulfasalazine	Folate	Impairs absorption and inhibits folate-dependent enzymes
Isoniazid	Pyridoxine	Impairs utilization of B_6
Nonsteroidal anti-inflammatory drugs	Iron	Gastrointestinal blood loss
Penicillamine	Zinc	Increases renal excretion

Toxins

Tobacco smoking alters the metabolism of several vitamins, including folate and vitamins C and E. In large surveys, diminished plasma levels of folate and ascorbic acid have been observed in chronic smokers. Smoking is also associated with diminished levels of folate in cells of the oral mucosa, diminished ascorbic acid levels in leukocytes, and decreased concentrations of vitamin E in the alveolar fluid, findings providing evidence that many tissues can be affected by smoking and that the effect does not simply represent a shift of these micronutrients out of the plasma compartment.

ADVANCES IN NUTRITIONAL SCIENCE

New Frontiers in Marginal Deficiency States of Micronutrients

Does Optimal Intake of Micronutrients Optimize Health?

Updating the definition of a micronutrient deficiency and establishing recommended daily intakes that are consistent with the most recent evidence have proved difficult for several reasons. In some instances, a novel biochemical or physiologic role for a nutrient has been identified but the question that arises is whether optimization of such functions translates into optimization of health. For example, providing supplemental vitamin E to elderly individuals whose vitamin E status falls within normative standards enhances T-lymphocyte responsiveness; nevertheless, it is unclear whether this translates into diminished infection rates. Another difficult problem pertains to the use of micronutrients in supraphysiologic quantities that exceed all conventional concepts of what is necessary for health. Some micronutrients, when they are taken in large quantities, have effects on physiologic functions that impart apparent health benefits. The ingestion of gram quantities of niacin to reduce low-density lipoprotein (LDL) cholesterol is an example. Such physiologic effects are not observed at more conventional levels of intake and are therefore usually considered pharmacologic effects of the nutrient. Thus, the determination of optimal nutrient intake is highly dependent on which physiologic effect is sought. Furthermore, if only a segment of the population will benefit from supraphysiologic quantities of a nutrient, should dietary guidelines for the remainder of the population be established according to this effect?

Determining an adequate level of intake implies the existence of a means of measuring nutrient status. In seeking an appropriate measure of nutrient status, the diversity of function often makes it difficult to decide which measurement is the most germane. Tobacco smoking, for example, diminishes vitamin E levels in alveolar fluid but not in the serum. Thus, the concepts of localized nutrient deficiencies and tissue-specific requirements add an additional level of complexity to the determination of nutrient status.

Redefinition of Nutritional Requirements

Folate

An example of the complexities that have arisen in redefining the criteria for vitamin deficiencies and vitamin requirements is the water-soluble vitamin folate. In the past, guidelines regarding its necessary intake were straightforward because they were based solely on the prevention of megaloblastic anemia. Measurement of serum and erythrocyte folate concentrations was the most common means of assessing status, and maintaining these levels within accepted normative ranges provided assurance that folate status was adequate to prevent anemia. However, degrees of deficiency that are insufficient to cause anemia may still disturb normal biochemical and physiologic homeostasis and, in some instances, cause clinical disease. Clinical trials have demonstrated that women taking folic acid supplements at the time of conception have a markedly lower chance of delivering a baby with an NTD compared with women who are not folate supplemented but whose folate status falls within a conventionally accepted range. This observation compelled the U.S. government to mandate the fortification of flour, beginning in 1998. Present recommendations are that women of childbearing age consume 400 μg/day of folic acid in the form of supplements or fortified foods, although the dose-response curve of this effect is ill-defined.

Less than optimal intake of folate is also evidenced by an increase in serum homocysteine, an amino acid that is normally metabolized by a folate-dependent pathway. Before the federally mandated fortification of flour, the median intake of folate among adults was half of the present RDA, and a substantial minority of Americans had significantly elevated serum homocysteine levels. Elevated homocysteine is associated with the development of

occlusive vascular disease and accelerated cognitive decline. In randomized clinical trials, however, supplementation with folate, vitamin B_{12}, and vitamin B_6 has shown no benefit against cardiovascular disease despite its ability to lower homocysteine levels. Such supplementation also has no clear benefit for cognitive function, except perhaps in patients with low baseline folate levels.

A compelling body of observations in both humans and animals has demonstrated that habitually low consumption of folate substantially increases the risk of colorectal cancer[5] and perhaps cancers of other organs, such as those of the breast and pancreas. This inverse relationship is observed even when folate status (or dietary intake) falls within the range of conventionally accepted normative values. This relationship has further complicated the determination of what constitutes an optimal intake of folate because the recent epidemiologic data suggest that about 500 μg constitutes the optimal daily intake for suppressing the risk of colon cancer. The issue is further confounded by observations, albeit controversial, suggesting that exceptionally high doses of supplemental folic acid among those who unknowingly harbor precancerous or cancerous lesions may paradoxically enhance the progression of these neoplasms,[6] thereby underscoring the potential for harm produced by taking a nutrient outside of its physiologic window.

The most recent update of the U.S. RDA for folate raised the value from 200 to 400 μg/day, citing both the prevention of anemia and optimization of serum homocysteine as criteria, and recommended that women capable of becoming pregnant consume an additional 400 μg/day in the form of supplements or fortified food. The issues surrounding the prevention of cardiovascular disease, cancer, and cognitive decline were not incorporated into that 1998 determination because the existing data at the time were inconclusive. However, future revisions of the RDAs may integrate some of this new knowledge. The potential for toxicity, the criterion for which was primarily linked to its ability to mask vitamin B_{12} deficiency, was dealt with by setting the TUL at 1000 μg/day of folic acid obtained from supplements and fortified foods in addition to that obtained from natural food sources (see Table 218-2).

Table 218-5 lists several examples of biochemical functions of vitamins that were not formerly recognized. As the clinical significance of each of these new roles is defined and as quantities of each vitamin needed to optimize such functions are determined, redefinition of the desirable range of vitamin status is likely to occur. Future efforts to refine appropriate dietary goals for each micronutrient will, however, need to take into consideration an important theme that is underscored by the previous discussion: the level of consumption of a particular micronutrient that conveys health benefits to one segment of the population is not necessarily beneficial, or even appropriate, for all segments of society.

Antioxidant and Free Radical Scavenging Vitamins and Provitamins

Vitamins A, C, and E as well as many of the carotenoids are effective antioxidants. In addition, vitamins C and E and some of the carotenoids can scavenge free radicals when these nutrients are taken in adequate quantities. Oxidation and free radical damage have been implicated as important contributors to common degenerative illnesses, such as atherosclerosis, cancer, cataracts, and retinal degeneration. Clinical trials to test the efficacy of antioxidant supplements have generally shown no benefit and in some instances harm,[7] although growing evidence indicates that health benefits of such supplements can be realized in populations with marginal antioxidant status. Two large-scale clinical intervention trials with β-carotene supplements conducted in the 1990s reported increased rates of lung cancer among the recipients of the carotenoid. Subsequent mechanistic studies indicated that the large doses administered (20 to 30 mg/day) result in asymmetrical cleavage of the carotenoid into unnatural products that antagonize normal signaling pathways in the lung epithelium, whereas lower supplemental doses undergo symmetrical cleavage into two molecules of vitamin A, thereby protecting against neoplastic transformation.

LDL oxidized in vivo is atherogenic. Prevention of LDL oxidation, at least in animal models, retards the process of atherogenesis. Supplementation of human subjects with several times the RDA of α-tocopherol, and perhaps some of the other antioxidant micronutrients, is an effective means of preventing LDL oxidation. Human intervention trials with vitamin E or other antioxidant nutrients, however, have generally been unable to demonstrate clinical benefits in the reduction of cardiovascular events. There nevertheless has been a sizable reduction in cardiovascular events observed with vitamin E supplementation among populations of patients who are under exceptional oxidative stress, such as those with chronic renal failure and certain classes of diabetics, suggesting that it is only among select groups of individuals that a clinical benefit may be realized.

Epidemiologic studies indicate that occurrence of cancers of the oral cavity, lung, esophagus, and stomach (and perhaps the colorectum) is inversely related to dietary intake of fresh vegetables and fruits. Careful dissection of dietary data suggests that β-carotene and vitamin E content are strongly predictive components of these foodstuffs. High doses of vitamin A and some of its synthetic analogues (e.g., 13-*cis*-retinoic acid) can effectively reduce the recurrence of head and neck cancers, although hepatic toxicity is sometimes a limiting factor in such cancer preventive therapy. Similarly, these agents, as well as β-carotene or vitamin E, taken in large doses have been shown significantly to promote the regression of oral leukoplakia, a premalignant lesion. Daily supplementation with one to three times the U.S. RDA of β-carotene, selenium, and vitamin E has been shown to reduce the incidence of adenocarcinoma of the stomach in a region of China where the disease as well as marginal vitamin status is particularly prevalent. However, as mentioned earlier, trials conducted in developed Western countries have observed no diminution of lung cancer among smokers with daily supplementation of β-carotene and vitamin E.

Epidemiologic associations also suggest an inverse relationship between lens cataract or macular degeneration and the intake of vitamin C, vitamin E, and β-carotene. These common degenerative conditions of the eye are caused, at least in part, by photo-oxidation. Some evidence in animal models indicates that they can be retarded by supraphysiologic supplementation with vitamin C or E. When tested under the conditions of a rigorously conducted multicenter, controlled trial, daily supplementation with a combination of vitamin C, vitamin E, and β-carotene (with or without zinc) had no effects compared with placebo on the likelihood for development of cataracts. However, the combination that included zinc produced an approximately 30% decline in the progression of early macular degeneration to an advanced stage and the likelihood of moderate visual acuity loss.

Further investigation is necessary to define the circumstances more clearly under which antioxidant nutrients can be used to prevent or to treat chronic degenerative diseases.

TABLE 218-5 NEWLY IDENTIFIED ROLES FOR VITAMINS

VITAMIN OR PROVITAMIN	CLASSIC ROLE	NEW ROLE
β-Carotene	Pro-vitamin A	Antioxidant, free radical
Niacin	NAD/NADP coenzyme	Reduction of LDL, elevation of HDL cholesterol
Folate	Hematopoietic factor	Diminishes homocysteinemia
Vitamin A	Transduction of visual input in retina	Induction and maintenance of epithelial differentiation, signal in embryogenesis
Vitamin D	Regulator of calcium	Retards epithelial proliferation; promotes differentiation
Vitamin B_6	Coenzyme for transamination	Modulation of steroid activity

HDL = high-density lipoprotein; LDL = low-density lipoprotein; NAD = nicotinamide adenine dinucleotide; NADP = nicotinamide adenine dinucleotide phosphate.

Vitamin B_{12} and Neuropsychiatric Disease

Plasma vitamin B_{12} concentrations are considered to be an accurate indication of vitamin B_{12} status. Values greater than 150 pg/mL were thought, until recently, to exclude vitamin B_{12} deficiency as a cause of neurologic or psychiatric syndromes.[8] Recent observations now indicate that 7 to 10% of individuals who have plasma vitamin B_{12} values between 150 and 400 pg/mL may develop neuropsychiatric complications of vitamin B_{12} deficiency in the absence of any indications of megaloblastic anemia. Such individuals can be identified by the demonstration of an elevated level of methylmalonic acid in the blood that decreases to normal levels with parenteral vitamin B_{12} administration. An elevation in serum methylmalonic acid is both a sensitive and a specific indication of cellular vitamin B_{12} deficiency. An alternative approach is to administer several parenteral injections of vitamin B_{12} to an individual who has an otherwise unexplained neuropsychiatric syndrome and whose plasma vitamin B_{12} level falls in the range of 150 to 400 pg/mL. Awareness of this phenomenon is particularly important because it has become clear that atrophic gastritis, an asymptomatic condition that affects approximately

30% of the elderly population, frequently produces a modest decrease in vitamin B_{12} status; similarly, long-term use of proton pump inhibitor drugs inhibits absorption and also increases the risk of clinically significant deficiency.[9]

Is Routine Multivitamin and Multimineral Supplementation Beneficial?

A common query by patients is whether regular use of a multivitamin or multimineral supplement is safe and efficacious in the maintenance of health. Although there is not a unanimous consensus about the "correct" answer to this question, the weight of available evidence indicates that for the general adult North American population, supplementation offers little or no benefit in regard to the prevention of the common chronic degenerative diseases, such as vascular disease, cancer, and dementia.[A1–A5] Although this apparent lack of efficacy has been notably contradicted by two clinical trials conducted in Western industrialized countries in which men taking multivitamins realized modest decreases in the incidence of cancer,[10] such benefits have not been substantiated by other investigations.

Although daily supplementation at the levels found in most multivitamin preparations probably presents no risk of harm, adverse health effects have been observed in several rigorously performed clinical trials in which long-term supplementation with micronutrients at levels that exceed the RDA (or conventional levels of dietary intake) by several-fold was examined. For example, an increased incidence of prostate cancer was observed in the SELECT trial, in which vitamin E was administered at a dose of 400 IU/day, and β-carotene supplementation resulted in an increased incidence of lung cancer among heavy smokers in the ATBC and CARET trials at doses of 20 to 30 mg/day.

This is not to say that health benefits cannot be realized from supplementation in select groups of individuals, although some thought needs to be exercised to determine which segments of the population should be targeted and what specific nutrients should be administered. Certainly, health benefits are likely in individuals whose dietary intake is chronically inadequate or in patients whose medical conditions are often complicated by micronutrient deficiencies, such as those on chronic renal dialysis or among individuals with marginally controlled intestinal malabsorption. The elderly frequently cannot achieve recommended intakes of vitamin D and calcium with diet alone, and therefore targeted supplementation with these nutrients is often indicated. Similarly, the high prevalence of atrophic gastritis among the elderly as well as the frequent use of proton pump inhibitor drugs each conspire to impair adequate vitamin B_{12} status.[11] Moreover, in many regions of the world, there continues to be a high prevalence of marginal micronutrient status among the general adult population, and in such areas widespread supplementation may be indicated; the Linxian trial in China, in which supplementation with a mixture of several antioxidant micronutrients led to a sizable decrease in gastric cancer, is one such example.

Grade A References

A1. Clarke R, Halsey J, Lewington S, et al. Effects of lowering homocysteine levels with B vitamins on cardiovascular disease, cancer, and cause-specific mortality: meta-analysis of 8 randomized trials involving 37,485 individuals. *Arch Intern Med*. 2010;170:1622-1631.

A2. Balk EM, Raman G, Tatsioni A, et al. Vitamin B_6, B_{12}, and folic acid supplementation and cognitive function: a systematic review of randomized trials. *Arch Intern Med*. 2007;167:21-30.

A3. Fortmann S, Burda B, Senger C, et al. Vitamin and mineral supplements in the primary prevention of cardiovascular disease and cancer: an updated systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2013;159:824-834.

A4. Grodstein F, O'Brien J, Kang J, et al. Long-term multivitamin supplementation and cognitive function in men: a randomized trial. *Ann Intern Med*. 2013;159:806-814.

A5. Lamas G, Roineau R, Goertz C, et al. Oral high-dose multivitamins and minerals after myocardial infarction. A randomized trial. *Ann Intern Med*. 2013;159:797.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ribas GS, Vargas CR, Wajner M. L-Carnitine supplementation as a potential antioxidant therapy for inherited neurometabolic disorders. *Gene*. 2014;533:469-476.
2. Johnson M, Hausman D, Davey A, et al. Vitamin B_{12} deficiency in African American and white octogenarians and centenarians in Georgia. *J Nutr Health Aging*. 2010;14:339-345.
3. Dietary reference intakes for calcium and vitamin D. www.iom.edu/Reports/2010/Dietary-Reference-Intakes-for-calcium-and-vitamin-D.aspx; Accessed March 23, 2015.
4. Chung M, Lee J, Terasawa T, et al. Vitamin D with or without calcium supplementation for prevention of cancer and fractures: an updated meta-analysis for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2011;155:827-838.
5. Gibson T, Weinstein S, Pfeiffer R, et al. Pre- and postfortification intake of folate and risk of colorectal cancer in a large prospective cohort study in the U.S. *Am J Clin Nutr*. 2011;94:1053-1062.
6. Figueiredo JC, Grau MV, Haile RW, et al. Folic acid and risk of prostate cancer: results from a randomized clinical trial. *J Natl Cancer Inst*. 2009;101:432-435.
7. Klein E, Thompson I Jr, Tangen C, et al. Vitamin E and the risk of prostate cancer: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *JAMA*. 2011;306:1549-1556.
8. Hunt A, Harrington D, Robinson S. Vitamin B12 deficiency. *BMJ*. 2014;349:g5226.
9. Lam JR, Schneider JL, Zhao W, et al. Proton pump inhibitor and histamine 2 receptor antagonist use and vitamin B_{12} deficiency. *JAMA*. 2013;310:2435-2442.
10. Gaziano J, Sesso H, Christen W, et al. Multivitamins in the prevention of cancer in men: the Physicians' Health Study II randomized controlled trial. *JAMA*. 2012;308:1871-1880.
11. Stabler SP. Clinical practice. Vitamin B12 deficiency. *N Engl J Med*. 2013;368:149-160.

219

EATING DISORDERS

MARIAN TANOFSKY-KRAFF

DEFINITION

Feeding and eating disorders are defined as syndromes "characterized by a persistent disturbance of eating or eating-related behavior that results in the altered consumption or absorption of food and that significantly impairs physical health or psychosocial functioning." The *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-5) defines anorexia nervosa (AN), bulimia nervosa (BN), and binge eating disorder (BED) as primary diagnoses in adolescents and adults. All other diagnoses are identified as Unspecified Feeding or Eating Disorder and represent presentations that do not meet the criteria for the primary eating disorders but nonetheless cause significant distress and impairment. The severity of each disorder is also specified as mild, moderate, severe, or extreme. Given the recent publication of the DSM-5, most empirical data available to date involve the criteria of the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition, text revision (DSM-IV-TR, published in 2000).

ANOREXIA NERVOSA

AN involves a restriction of "energy intake relative to requirements, leading to a significantly low body weight in the context of age, sex, developmental trajectory, and physical health."[1] Individuals with AN experience an intense fear of gaining weight or becoming fat, are overly concerned with weight or shape, and often may not recognize the seriousness of their low body weight. AN has two subtypes: restricting and binge-eating/purging. DSM-5 criteria for AN are listed in Table 219-1.

BULIMIA NERVOSA

A diagnosis of BN requires recurrent episodes of binge eating (i.e., the consumption of an unambiguously large amount of food given the context, accompanied by a sense of loss of control over eating). Episodes of binge eating co-occur with behaviors intended to compensate for energy consumed and to prevent weight gain, such as self-induced vomiting and fasting. Binge eating and compensatory behaviors must occur, on average, at least once a week for 3 months. The self-esteem of individuals with BN is excessively influenced by their body weight and shape. DSM-5 criteria for BN are outlined in Table 219-2.

BINGE EATING DISORDER

BED is characterized by recurrent episodes of binge eating in the absence of regular compensatory behaviors that are present in BN. The binge

TABLE 219-1 DSM-5 DIAGNOSTIC CRITERIA FOR ANOREXIA NERVOSA

A. Restriction of energy intake relative to requirements, leading to a significantly low body weight in the context of age, sex, developmental trajectory, and physical health. *Significantly low weight* is defined as a weight that is less than minimally normal or, for children and adolescents, less than minimally expected.
B. Intense fear of gaining weight or becoming fat, or persistent behavior that interferes with weight gain, even though at a significantly low weight.
C. Disturbance in the way in which one's body weight or shape is experienced, undue influence of body weight or shape on self-evaluation, or persistent lack of recognition of the seriousness of the current low body weight.

Specify whether:

Restricting type: During the last 3 months, the individual has not engaged in recurrent episodes of binge eating or purging behavior (i.e., self-induced vomiting or the misuse of laxatives, diuretics, or enemas). This subtype describes presentations in which weight loss is accomplished primarily through dieting, fasting, and/or excessive exercise.

Binge-eating/purging type: During the last 3 months, the individual has engaged in recurrent episodes of binge eating or purging behavior (i.e., self-induced vomiting or the misuse of laxatives, diuretics, or enemas).

From Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Washington, DC: American Psychiatric Association; 2013.

TABLE 219-2 DSM-5 DIAGNOSTIC CRITERIA FOR BULIMIA NERVOSA

A. Recurrent episodes of binge eating. An episode of binge eating is characterized by both of the following:
   1. Eating, in a discrete period of time (e.g., within any 2-hour period), an amount of food that is definitely larger than most people would eat during a similar period of time and under similar circumstances.
   2. A sense of lack of control over eating during the episodes (e.g., a feeling that one cannot stop eating or control what or how much one is eating).

B. Recurrent inappropriate compensatory behavior in order to prevent weight gain, such as self-induced vomiting; misuse of laxatives, diuretics, enemas, or other medications; fasting; or excessive exercise.

C. The binge eating and inappropriate compensatory behaviors both occur, on average, at least once a week for 3 months.

D. Self-evaluation is unduly influenced by body shape and weight.

E. The disturbance does not occur exclusively during episodes of anorexia nervosa.

From Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Washington, DC: American Psychiatric Association; 2013.

TABLE 219-3 DSM-5 DIAGNOSTIC CRITERIA FOR BINGE EATING DISORDER

A. Recurrent episodes of binge eating. An episode of binge eating is characterized by both of the following:
   1. Eating, in a discrete period of time (e.g., within any 2-hour period), an amount of food that is definitely larger than most people would eat during a similar period of time and under similar circumstances.
   2. A sense of lack of control over eating during the episodes (e.g., a feeling that one cannot stop eating or control what or how much one is eating).

B. The binge eating episodes are associated with three (or more) of the following:
   1. Eating much more rapidly than normal.
   2. Eating until feeling uncomfortably full.
   3. Eating large amounts of food when not feeling physically hungry.
   4. Eating alone because of feeling embarrassed by how much one is eating.
   5. Feeling disgusted with oneself, depressed, or very guilty afterward.

C. Marked distress regarding binge eating is present.

D. The binge eating occurs, on average, at least once a week for 3 months.

E. The binge eating is not associated with the recurrent use of inappropriate compensatory behavior as in bulimia nervosa and does not occur exclusively during the course of bulimia nervosa or anorexia nervosa.

From Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Washington, DC: American Psychiatric Association; 2013.

episodes are distinguished by at least three associated characteristics, such as eating rapidly, eating until feeling uncomfortably full, and feeling disgust and guilt regarding the episodes. Individuals experience marked distress surrounding the binge episodes, and the binge eating episodes must occur, on average, at least once a week for 3 months. DSM-5 criteria for BED are listed in Table 219-3.

EPIDEMIOLOGY

Data suggest a lifetime prevalence of AN of approximately 0.6%, with higher rates among women (0.9%) compared with men (0.3%).[2] The lifetime prevalence of BN appears to be about 1%, with higher rates among women (1.5%) than among men (0.5%). The lifetime prevalence of BED is estimated at 3.5% for women and 2.0% for men. Among adolescents, lifetime prevalence estimates of AN, BN, and BED have been reported at 0.3%, 0.9%, and 1.6%, respectively.[3] Contrary to the view that eating disorders afflict only non-Hispanic white, affluent women, individuals of all races, ethnicities, and cultures are affected by these diagnoses.[4]

PATHOBIOLOGY

Research regarding the neuropathology of eating disorders is in nascent stages. Data suggest that several brain regions may be involved in and potentially interact in the manifestation of all eating disorders. Individuals with eating disorders appear to have brain function alterations in emotional/limbic, reward, and cognitive control circuits.[5] Fear circuitry networks involving the amygdala, anterior cingulate cortex, hippocampus, insula, striatum, and prefrontal cortex have demonstrated differential activation among individuals with eating disorders (with the majority of research in AN) compared with controls. Specifically, there tends to be a hyper-responsiveness in the limbic circuitry in response to potentially threatening cues, such as food and body weight/shape.

There also appear to be alterations in reward function in patients with AN, but the direction is unclear. By contrast, individuals with BN and BED consistently demonstrate hyper-responsivity in reward and somatosensory regions on exposure to food images. Data also suggest that individuals with eating disorders may have dysregulated frontal cortical cognitive neural networks acting in concert with regional reward systems.[6] Individuals with eating disorders have demonstrated impaired cognitive flexibility. Specific to BN, impulsivity and poor inhibitory control have also been reported.

Given the brain regions implicated in eating disorders, current studies have focused on the role of dopamine and serotonin in the manifestation of eating disorders. Individuals with AN appear to have impaired dopaminergic signaling, particularly in striatal circuits, that might contribute to altered reward and affect, decision making, and executive control as well as compulsivity and decreased food ingestion. Moreover, emerging clinical research suggests that striatal dopamine abnormalities exist in individuals with BN and BED. Because serotonin (5-hydroxytryptamine) 1A and 2A receptors and the serotonin transporter may play a part in symptoms of eating disorders, such as impulse control and associated mood symptoms, it is likely that interactions between the serotonin and dopaminergic systems contribute to eating disorders.

Risk Factors

Eating disorders develop as the result of multiple biological, psychological, and sociocultural factors. AN, BN, and BED aggregate in families, with estimates from twin studies suggesting that 40 to 60% of vulnerability for eating disorders is genetic. Studies have reported links between eating disorders and polymorphisms in the serotonin transporter gene (*SLC6A4*), the dopamine D_2 receptor (*DRD2*) gene, the μ_1 opioid receptor (*OPRM1*) gene, the fat mass and obesity-associated (*FTO*) gene, and the brain-derived neurotrophic factor (*BDNF*) gene. Although genetic linkage and association studies have implicated several susceptibility loci for AN, BN, and BED, specific genes that consistently lend vulnerability to eating disorders are less conclusive.[7]

Female sex, pediatric overweight, elevated shape and weight concerns, sexual abuse, trauma, and mood disorders have been identified as risk factors for all eating disorders. Personality-related variables, such as impulsivity and perfectionism, appear to be linked to eating disorders.[8] Importantly, internalization to the "thin ideal" (a sociocultural emphasis on shape and weight and a marked preference for a thin body type) with resulting weight and shape concerns has been proposed to contribute to eating disorder development, particularly among adolescents who are under strong influence from their peer and family environments. For example, parental overconcern about eating, shape, and weight as well as weight-related teasing by family members confers risk for eating disorders. Specific to BED, maltreatment, including teasing and bullying, and perceived stress are risk factors for the disorder.[9]

CLINICAL MANIFESTATIONS

Symptoms and Signs

For AN, physical symptoms and signs may include amenorrhea, constipation, cold intolerance, anemia, and lanugo hair. Reduced bone density is believed to predict the onset of premature osteopenia and osteoporosis. Health problems associated with malnutrition affect cardiovascular, gastrointestinal, reproductive, and endocrine systems. Individuals with AN frequently present with comorbid psychiatric disorders, including mood and anxiety disorders (e.g., social phobia, specific phobia, post-traumatic stress disorder), and high rates of suicidal ideation and behavior.

Individuals with BN present with signs and symptoms most commonly associated with purging behavior. These include dental enamel erosion secondary to vomiting, gastrointestinal symptoms, salivary gland hypertrophy, and electrolyte disturbances. Electrolyte abnormalities can have deleterious effects on the renal and cardiovascular systems. BN patients are at risk for cardiometabolic conditions (e.g., diabetes, stroke) as well as chronic pain. Metabolic acidosis can also occur in patients who are abusing laxatives as a result of the loss of bicarbonate from the bowel. Noninflammatory swelling of the salivary glands is a common clinical manifestation of BN. The most common psychiatric comorbidities in BN are major depressive disorder, anxiety disorders, substance use disorders, and disruptive behavioral disorders.

Individuals with BED are frequently overweight or obese. However, adults with BED are likely to report the development of diagnoses of metabolic syndrome components (e.g., dyslipidemia, hypertension, type 2 diabetes) after accounting for the contribution of body weight. The presence of BED may affect bariatric surgery outcome, resulting in less weight loss or more weight regain, but this is not a consistent finding. However, the presence of "loss of control" eating after surgery consistently predicts less weight loss or greater weight regain. Compared with obese adults without BED, those with the disorder experience significant impairment in a number of domains of

psychosocial functioning, including a poorer quality of life and more impaired functioning in their home and social lives. Individuals with BED often have higher levels of disability, health problems, and work productivity impairment compared with obese and healthy controls without binge eating. With regard to comorbid psychiatric diagnoses, adults with BED experience Axis I psychiatric disorders at a rate comparable to (or higher than) that of individuals with AN or BN, including major depressive disorder, anxiety disorders, substance use disorders, and disruptive behavioral disorders.

Natural History

AN is typically manifested during adolescence, although the disorder can develop before puberty. BN frequently develops during later adolescence or early adulthood. BED is often manifested in adulthood, but adolescents also present with the disorder. Several retrospective and prospective studies report that binge and "out of control" eating occur as early as middle childhood.

Data on the natural course of eating disorders in the clear absence of treatment are limited. Eating disorders tend to exhibit a remitting and relapsing natural course across the lifespan, and there appear to be high rates of diagnostic crossover.[10] Treatment outcome data indicate that AN tends to transition to BN or an Unspecified Eating Disorder, and those with BN and BED tend to migrate from one to the other.

DIAGNOSIS

A number of structured, well-validated assessments for the diagnosis of eating disorders exist. These include but are not limited to the Structured Clinical Interview for the DSM and the Eating Disorder Examination. However, eating disorders are typically diagnosed by review of the patient's history, symptoms, and behaviors in an interview format. Evaluation of comorbid psychiatric problems, most notably mood, anxiety, substance use disorders, and disruptive behavioral disorders, is also required. Information should be gathered on interpersonal relationships, history of sexual and physical abuse, self-harm, and suicidal ideation or behavior. Family involvement is crucial, particularly for pediatric patients. A complete physical examination to assess body composition, vital signs, cardiovascular function, and hematologic and blood chemistry parameters is recommended for all patients.

TREATMENT Rx

Anorexia Nervosa

There is limited evidence on effective treatments for AN. For severely underweight patients, inpatient medical monitoring and supervised nutrition rehabilitation are required. The optimal setting (inpatient versus outpatient treatment) remains a subject of debate, and the evaluation of treatment costs in AN plays an important role in determining treatment. However, for pediatric patients, family-based psychotherapy, particularly during the early phases of the disorder, has demonstrated effectiveness.[11] Maudsley's family-based therapy involves both joint family sessions and simultaneous but independent patient/family intervention. Antidepressants (e.g., selective serotonin reuptake inhibitors) are associated with high rates of noncompliance, and compelling evidence of beneficial effects has not been found. The use of antipsychotic drugs has been explored, but results regarding their effectiveness remain nondefinitive.

Bulimia Nervosa

Cognitive-behavioral therapy (CBT) has been recognized as the treatment of choice for BN. Interpersonal psychotherapy (IPT) is also effective for the treatment of BN, particularly for those who are nonresponsive to CBT. There is growing support that pharmacotherapy may be helpful for some patients with BN. Antidepressants, especially selective serotonin reuptake inhibitors, are modestly effective for reducing binge eating in BN over the short and long term. Topiramate has consistently been shown to decrease binge eating in BN, but side effects may limit its usefulness. It is unclear whether combination therapy may be required for optimal outcomes.

Binge Eating Disorder

Psychological treatment for BED aims to reduce binge eating, weight and shape concerns, and prevent excess weight gain and/or induce modest weight loss. The psychotherapies most evaluated in clinical trials include CBT, IPT, behavioral weight loss, and CBT guided self-help (CBTgsh) approaches. CBT and IPT are first-line treatments. Given its cost-effectiveness, CBTgsh may be an optimal treatment option when specialist care is not available. With regard to pharmacologic treatment in BED, three medications or classes of medications have been studied in two or more placebo-controlled trials. Selective serotonin reuptake inhibitors, sibutramine, and topiramate all produce reductions in frequency of binge eating relative to placebo in short-term trials.[12] However, sibutramine has been withdrawn from the market, and topiramate is frequently associated with problematic cognitive effects, thus limiting its clinical utility.

PREVENTION

Whereas an increasing number of macro-level environmental public health initiatives have emerged (i.e., anti-dieting media campaigns and sanctions on advertising practices propagating an ideal of extreme thinness), few empirical data exist evaluating their efficacy. However, there are more data on individual, micro-level interventions aimed at reducing proximal eating disorder risk factors as well as current and distal eating pathology. Selected, interactive, multisession programs with adolescent girls may be more effective than universal, didactic, heterogeneous-sampled and single-session programs in reducing risk factors for eating disorder symptoms. For example, a dissonance-based program aimed at reducing eating disorder risk factors in adolescent girls has demonstrated effectiveness.

PROGNOSIS

Anorexia Nervosa

Remission rates vary widely for AN. Lower remission rates (29%) have been observed, particularly in studies with the shortest follow-up duration. However, most individuals with AN (approximately 76%) treated in outpatient settings will remit within 5 years after the initiation of treatment. Most individuals who do not achieve remission from AN during follow-up periods transition to a diagnosis of BN or an Unspecified Eating Disorder, which likely captures partial syndrome AN. Among psychiatric diagnoses, AN consistently has one of the highest mortality rates due to suicide, nutritional deficits, cardiac complications, and substance abuse.[13] The crude cumulative mortality rate is 2.8%, with longer duration of illness before receiving treatment and the need for inpatient treatment as negative prognostic indicators for AN.[14] Predictors of relapse include desiring a lower weight at the end of treatment and receiving treatment in a general (versus specialty) clinic.

Bulimia Nervosa

Similar to AN, most individuals with BN (70% or more) who receive treatment fully remit when assessed 5 to 20 years later, with remission rates being much lower (27 to 28%) at 1-year follow-up. If individuals with BN do not achieve remission within 5 years, however, they are likely to exhibit a chronic course of the illness. Mortality rates for BN range between 0 and 2%. Diagnostic crossover from BN to AN is relatively rare; yet, there is frequent diagnostic crossover between BN and BED, which may suggest a possible common psychological and/or biologic maintaining process. Negative prognostic indicators for BN include endorsement of greater psychiatric comorbidity, multiple impulsive behaviors (e.g., self-harm, substance use disorder), and a family history of alcohol abuse. Individuals who receive inpatient treatment or have a low motivation for engaging in treatment are more likely to relapse.

Binge Eating Disorder

A paucity of data exists on the long-term outcomes for BED patients. There are data to suggest that at 1 year after outpatient treatment, upwards of 80% of patients remit. In one clinical trial that examined 4-year outcomes, between 52 and 76% of individuals receiving psychological treatment for BED demonstrated remission from binge eating.[15] These preliminary data suggest that the prognostic trajectory may be similar to that of BN. Diagnostic crossover from BED to BN is high, whereas crossover to AN is relatively rare. Although examination of prognostic indicators for BED is in its early stages, patients reporting an undue influence of their body shape or weight on self-evaluation are less likely to have remission from binge eating at 12-month follow-up.[16] Rapid remission of binge eating has also been shown to be a positive prognostic indicator for binge remission.[17]

Grade A References

A1. Lock J, Le Grange D, Agras WS, et al. Randomized clinical trial comparing family-based treatment with adolescent-focused individual therapy for adolescents with anorexia nervosa. *Arch Gen Psychiatry*. 2010;67:1025-1032.

A2. Poulsen S, Lunn S, Daniel SI, et al. A randomized controlled trial of psychoanalytic psychotherapy or cognitive-behavioral therapy for bulimia nervosa. *Am J Psychiatry*. 2014;171:109-116.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Attia E, Becker AE, Bryant-Waugh R, et al. Feeding and eating disorders in DSM-5. *Am J Psychiatry*. 2013;170:1237-1239.
2. Smink FR, van Hoeken D, Hoek HW. Epidemiology, course, and outcome of eating disorders. *Curr Opin Psychiatry*. 2013;26:543-548.
3. Swanson SA, Crow SJ, Le Grange D, et al. Prevalence and correlates of eating disorders in adolescents. Results from the national comorbidity survey replication adolescent supplement. *Arch Gen Psychiatry*. 2011;68:714-723.
4. Pike KM, Hoek HW, Dunne PE. Cultural trends and eating disorders. *Curr Opin Psychiatry*. 2014;27:436-442.
5. von Hausswolff-Juhlin Y, Brooks SJ, Larsson M. The neurobiology of eating disorders-a clinical perspective. *Acta Psychiatr Scand*. 2014;[Epub ahead of print].
6. Frank GK, Kaye WH. Current status of functional imaging in eating disorders. *Int J Eat Disord*. 2012;45:723-736.
7. Stefano GB, Ptacek R, Kuzelova H, et al. Convergent dysregulation of frontal cortical cognitive and reward systems in eating disorders. *Med Sci Monit*. 2013;19:353-358.
8. Kessler RC, Berglund PA, Chiu WT, et al. The prevalence and correlates of binge eating disorder in the World Health Organization World Mental Health Surveys. *Biol Psychiatry*. 2013;73:904-914.
9. Tanofsky-Kraff M, Bulik CM, Marcus MD, et al. Binge eating disorder: the next generation of research. *Int J Eat Disord*. 2013;46:193-207.
10. Allen KL, Byrne SM, Oddy WH, et al. DSM-IV-TR and DSM-5 eating disorders in adolescents: prevalence, stability, and psychosocial correlates in a population-based sample of male and female adolescents. *J Abnorm Psychol*. 2013;122:720-732.
11. Focker M, Knoll S, Hebebrand J. Anorexia nervosa. *Eur Child Adolesc Psychiatry*. 2013;22 (suppl 1):S29-S35.
12. McElroy SL, Guerdjikova AI, Mori N, et al. Pharmacological management of binge eating disorder: current and emerging treatment options. *Ther Clin Risk Manag*. 2012;8:219-241.
13. Campbell K, Peebles R. Eating disorders in children and adolescents: state of the art review. *Pediatrics*. 2014;134:582-592.
14. Keel PK, Brown TA. Update on course and outcome in eating disorders. *Int J Eat Disord*. 2010;43:195-204.
15. Franko DL, Keshaviah A, Eddy KT, et al. A longitudinal investigation of mortality in anorexia nervosa and bulimia nervosa. *Am J Psychiatry*. 2013;170:917-925.
16. Hilbert A, Bishop ME, Stein RI, et al. Long-term efficacy of psychological treatments for binge eating disorder. *Br J Psychiatry*. 2012;200:232-237.
17. Grilo CM, White MA, Gueorguieva R, et al. Predictive significance of the overvaluation of shape/weight in obese patients with binge eating disorder: findings from a randomized controlled trial with 12-month follow-up. *Psychol Med*. 2013;43:1335-1344.

REVIEW QUESTIONS

1. A 21-year-old woman presents with scarred knuckles, faded wrist scarring, electrolyte imbalance, and low weight (body mass index [BMI] = 20). The menstrual cycle is reported as abnormal and infrequent. The patient denies current abnormal or restrictive eating patterns and self-injurious and compensatory behaviors. The patient expresses clear discomfort during collection of weight and avoids eye contact with clinical staff. No binge eating is reported. What is the most likely diagnosis for this individual?

A. Anorexia nervosa, general diagnosis: patient exhibits low body weight, potential self-harming behaviors, amenorrhea, and clear discomfort with body weight.
B. Bulimia nervosa, general diagnosis: low weight, knuckle scarring, and electrolyte imbalance are consistent with bulimia nervosa.
C. Anorexia nervosa, purging subtype: low weight, potential self-harming behaviors, and amenorrhea are consistent with anorexia nervosa; knuckle scarring and electrolyte imbalance are consistent with anorexia nervosa purging subtype.
D. No current eating disorder diagnosis: patient's BMI is not *significantly* low within the context of age and sex to indicate a clinical diagnosis of anorexia nervosa, and binge eating and compensatory behaviors are not present.
E. Bulimia nervosa, restricting subtype: presence of menstrual cycle, discomfort with body weight, knuckle scarring, and electrolyte imbalance are consistent with bulimia nervosa; low body weight is consistent with bulimia nervosa restricting subtype.

Answer: D The patient's body weight is within a normal range, and denial of abnormal or restrictive eating patterns excludes a diagnosis of anorexia nervosa. Although knuckle scarring and electrolyte imbalance are suggestive of self-induced vomiting, a diagnosis of bulimia nervosa cannot be made without clear indication of the regular use of compensatory behaviors. Although amenorrhea is often a symptom, it is currently neither a necessary nor conclusive criterion for the diagnosis of anorexia nervosa. Although the patient cannot be given a conclusive eating diagnosis, further psychological and physical examination is warranted.

2. A 19-year-old female college student presents with current bulimia nervosa. The patient reports feeling very dissatisfied with her body shape and weight and engaging in binge eating while watching reality television. What is the best *initial* course of treatment action for you to suggest to the patient?

A. Instruct the patient to modify television preferences to reduce episodes of binge eating while opening up a dialogue with friends about healthy eating habits.
B. Refer the patient to a therapist for cognitive-behavioral therapy.
C. Prescribe topiramate to help the patient reduce episodes of binge eating.
D. Instruct the patient to keep a journal documenting emotions surrounding binge episodes for personal reflection.
E. Refer the patient to a therapist for dialectical behavioral therapy.

Answer: B Cognitive-behavioral therapy remains the "gold standard" for the treatment of bulimia nervosa. Although other interventions, such as interpersonal psychotherapy and dialectical behavioral therapy, have shown efficacy in treatment of bulimia nervosa, cognitive-behavioral therapy shows the greatest efficacy across studies and subjects. Although discussing or expressing emotions regarding food and body shape and weight can be helpful for individuals with bulimia nervosa, constructive guidance is generally necessary for symptom reduction and cessation. Although topiramate has been shown to help individuals reduce binge eating, it has a number of cognitive side effects that may make its use impractical.

3. Which of the following is *not* a criterion for the diagnosis of binge eating disorder?

A. BMI ≥ 30 (body mass index obesity threshold)
B. Marked distress surrounding binge episodes
C. Binge eating
D. Lack of regular compensatory behaviors
E. Weekly episodes of binge eating for at least 3 months

Answer: A There is no weight requirement for a diagnosis of binge eating disorder. Although individuals with binge eating disorder are often overweight or obese, BMI is not a required criterion for the diagnosis.

4. Which of the following statements regarding the genetics of disordered eating is true?

A. Identical twins are 100% concordant for anorexia nervosa.
B. Several genes have been conclusively linked to increased vulnerability for eating disorders.
C. Of the neurotransmitters, dopamine and norepinephrine are the most likely candidates for eating disorder maintenance.
D. Media generally play a bigger role than genetics in the onset of eating disorders.
E. Children of individuals with binge eating disorder are more likely to experience out of control eating episodes than anorexia nervosa.

Answer: E Studies show that disordered eating behaviors, such as binge or loss of control eating, are often highly heritable. Offspring of mothers with binge eating have a higher genetic propensity to develop binge eating than do children of mothers who do not exhibit the behavior. Although monozygous twins share equal genetic predisposition for the development of disordered eating behaviors, no eating disorder shows 100% concordance between monozygous twins. Furthermore, although several gene loci have been indicated in the development and maintenance of eating disorders, evidence does not conclusively point to specific genes as responsible for disordered eating. Although studies have indicated the role of norepinephrine in the onset and maintenance of eating disorders, dopamine and serotonin are currently generally considered to be the most likely neurotransmitters to influence disordered eating behaviors. Findings regarding the effect of media on increasing risk for disordered eating are mixed; genetics, however, are strongly implicated.

5. Which of the following is *not* a finding regarding the neuropathology of eating disorders?

A. Individuals with bulimia nervosa and binge eating disorders are hyperresponsive to food images in reward and somatosensory regions.
B. Individuals with bulimia nervosa exhibit decreased impulsivity and enhanced inhibitory control relative to control subjects.
C. Individuals with anorexia nervosa exhibit impaired dopaminergic signaling in striatal circuits.
D. Individuals with anorexia nervosa exhibit differential activation of fear circuits relative to control subjects.
E. Findings indicate that dysregulation in the limbic system may be influential in the development of eating disorders.

Answer: B Data suggest that individuals with bulimia nervosa exhibit *increased* impulsivity and *decreased* inhibitory control relative to control subjects.

220 OBESITY

MICHAEL D. JENSEN

Obesity is the most common nutritional disorder in the United States and directly or indirectly accounts for a significant portion of health-related expenses. The safest treatment approaches (lifestyle change and behavior modification) are not those commonly employed by physicians and require considerable time to implement. The recently released Guideline for the Management of Overweight and Obesity in Adults provides direction for clinicians for the treatment of obesity.

DEFINITION

The Guideline for the Management of Overweight and Obesity in Adults produced by the National Institutes of Health and the National Heart, Lung, and Blood Institute (NHLBI) and disseminated by the American College of Cardiology (ACC), the American Heart Association (AHA), and The Obesity Society (TOS) provides evidence-informed, scientifically based recommendations on evaluation and management of overweight and obesity.

Body mass index (BMI) continues to be the recommended approach to categorize weight relative to height for adults. BMI is calculated as weight (in kilograms) divided by height squared (in meters):

$$BMI = \frac{weight\ (kg)}{height^2\ (m^2)}$$

To calculate BMI with pounds and inches, the formula is modified as follows:

$$BMI = \frac{weight\ (lb)}{height^2\ (in^2)} \times 703$$

The guideline suggested that no changes are indicated in the weight classifications by BMI, which are summarized in Table 220-1. Individuals who are overweight (BMI of 25.0 to 29.9) may or may not be overfat. Some adults may be overweight because of increased muscle mass, which is a straightforward clinical observation. Although, in general, the risk for development of adiposity-related health problems increases continuously as the BMI exceeds 25, the new guideline continues to recommend the use of waist circumference measurements to discriminate among patients who may require more testing. Overweight and class I obese patients with a waist circumference in the high-risk category deserve a discussion of lifestyle issues as they relate to health and weight loss. Some individuals with a BMI of 27 to 29.9 develop serious metabolic complications that improve with weight loss and are candidates for more aggressive treatment, including pharmacotherapy if it is needed. Asian populations, in particular, are at risk for the typical metabolic complications of obesity at lower BMI and waist circumferences than those for whites, Hispanics, blacks, and Polynesians; the guideline for at-risk BMI in Asian populations is 23 to 24.

TABLE 220-1 CLASSIFICATION OF OVERWEIGHT AND OBESITY BY BODY MASS INDEX (BMI)

	OBESITY CLASS	BMI (kg/m^2)
Underweight		<18.5
Normal		18.5-24.9
Overweight		25.0-29.9
Obesity	I	30.0-34.9
Obesity	II	35.0-39.9
Extreme obesity	III	≥40

Jensen MD, Ryan DH, Apovian CM, et al. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. *J Am Coll Cardiol.* 2014;63:2985-3023.

The prevalence of comorbidities and risk of future morbidities increase considerably at a BMI of more than 30, the cut point for obesity. Obesity is divided into three classes, also depending on BMI (see Table 220-1). Treatment approaches may differ for those who are overweight and for different classes of obesity. For example, current U.S. Food and Drug Administration guidelines indicate that pharmacotherapy can be adjunct treatment for any class of obesity, even if medical complications are not present. Familiarity with the guidelines is important. Supervisory agencies and third-party payers use them to determine who is eligible for treatment benefits. Extreme obesity (BMI > 40) is one of the key features that would prompt consideration of a patient for bariatric surgery when medical treatments have failed. Patients with class II obesity (BMI of 35.0 to 39.9) may be considered for bariatric surgery if medical treatments have failed and if severe, life-threatening complications are present.

As noted, the new NHLBI/ACC/AHA/TOS guidelines continue to recommend waist circumference as an office assessment tool to help with the treatment decision-making process. The new guidelines suggest that the previous waist circumference cut points of more than 102 cm (40 inches) for men and more than 88 cm (35 inches) for women are indicators of increased metabolic risk. However, the report stated that the relationships between disease risk and waist circumference are continuous and progressive, with no obvious cut points. The recommendation is to measure waist circumference in overweight and class I obesity adults. Those adults with waist circumferences above the cut points deserve further evaluation to detect other cardiovascular disease risk factors. Adults with class II or class III obesity are at sufficiently high risk that waist circumference information does not appear to add valuable information. These definitions of overweight and obesity and of high-risk waist circumference are generally applicable to those of European and African descent, but lower values are recommended for those of Asian descent. The risks of metabolic abnormalities occur at lower BMI and lower waist circumference in these populations.

EPIDEMIOLOGY

Prevalence of Obesity

Although the number of overweight and obese adults in the United States has increased dramatically during the past 30 years, the increase in the prevalence is now slowing or leveling off.[1] In 2009-2010, the prevalence of obesity was 35.5% among adult men and 35.8% among adult women, with no significant change compared with 2003-2008. Approximately 60% of U.S. men and 51% of U.S. women are overweight or obese, although a greater percentage of women than men are obese. There are substantial differences in the prevalence of obesity by age, race, and socioeconomic status. The prevalence of obesity in adults tends to rise steadily from the ages of 20 to 60 years, decreasing in later years. It has been estimated that almost 75% of men aged 60 to 69 years in the United States have a BMI of more than 25. The increase in mean BMI with age is not as much of a threat to population health as is a similar increase in the BMI of younger populations. The lowest mortality rates for young adults are for a BMI in the lower part of the normal range (20.0 to 24.9), whereas the BMI associated with the lowest mortality rates is somewhat above 25 kg/m^2 for those in the 60s and 70s. Physicians should base their weight recommendations for individual patients on whether adverse health consequences associated with obesity are present.

The differences in overweight and obesity among African Americans, Mexican Americans, and European Americans are not subtle. African American women and Mexican Americans of both sexes have the highest rates of overweight and obesity in the United States. In interpreting these data, however, it is important to keep in mind that there is an inverse relationship between socioeconomic status and obesity, especially among women (Chapter 5). Women in lower socioeconomic classes are much more likely than those in higher socioeconomic classes to be obese. This association reduces but does not eliminate the racial differences in the prevalence of obesity. Whether the remaining racial differences in the prevalence of obesity are due to genetic, constitutional, or social factors is not yet known.

PATHOBIOLOGY

Etiology

Genetic and constitutional susceptibility to obesity are heavily influenced by the environment. Evidence from studies of twins adopted into different families indicates that within a given environment, a significant portion of the variation in weight is genetic.[2] That said, the remarkable increase in the prevalence of obesity in the United States during the past 3 decades is unlikely to be due to wholesale changes in the genetic makeup of Americans.

Genetic Aspects of Human Obesity

Although obesity susceptibility is a classic polygenic condition, there are also a number of syndromic and monogenic obesity syndromes. The long-recognized genetic defects resulting in obesity include Prader-Willi and Laurence-Moon-Biedl syndromes. More recently, rare monogenic forms of human obesity due to mutations in the leptin gene, the leptin receptor gene, and the melanocortin signaling system genes have been described. These gene mutations are most often associated with increased appetite rather than with reduced energy expenditure. Genome-wide association studies have reported a number of genes associated with higher BMI. Those that appear to predict the greatest amount of variance in BMI include the fat mass and obesity-associated (*FTO*) gene and the melanocortin-4 receptor (*MC4R*) gene. Other genes that have been reliably associated with obesity include *TMEM18*, *KCTD15*, *GNPDA2*, *SH2B1*, *MTCH2*, and *NEGR1*. Together, however, the combined effects of all the identified genetic contributions account for less than 1% of the variance in BMI. This emphasizes both the huge environmental effects and the polygenic nature of susceptibility to obesity.

Constitutional Influences on Obesity

A number of environmental factors can result in long-term, epigenetic effects on body weight regulation and the susceptibility to obesity-related health problems. These epigenetic effects are ascribed to processes that include changes in DNA methylation, acetylation, and chromatin remodeling. The effect of the intrauterine environment and the perinatal period on subsequent weight and health is best studied. Undernutrition in the last trimester of pregnancy and in the early postnatal period decreases the risk of adult obesity, although the low birthweight associated with undernutrition (or smoking) in late pregnancy also increases the risk of adulthood hypertension, abnormal glucose tolerance, and cardiovascular disease. In contrast, undernutrition limited to the first two trimesters of pregnancy is associated with an increased probability of adult obesity. The infants of diabetic mothers tend to be fatter than those of nondiabetic mothers, and children of diabetic mothers have a greater prevalence of obesity when they are 5 to 19 years old, independent of whether their mother is obese. Finally, intrauterine exposure to the diabetic environment results in an increased risk of diabetes mellitus and obesity in the offspring. Thus, the issue of the genes versus the environment in regard to obesity and metabolic complications of obesity is blurred in the intrauterine and perinatal time intervals. One of the striking and worrisome aspects of these metabolic effects is not only the long-term effects on the individual's weight regulation and health but also the suggestion that these traits can be passed on to future generations.

Environmental Contributors to Human Obesity

Dramatic changes in the environment of Western countries have occurred during the past 50 years, including reduced demands for physical activity and alterations in the food supply. These food supply changes appear to have either increased or prevented the expected decrease in energy intake that would be needed to match the reduced energy expenditure from physical activity.[3]

Food

A number of environmental factors can influence food intake (Table 220-2). Consuming energy-dense foods results in greater energy intake because adults tend to respond to food volume rather than to the energy content. This factor likely accounts for the association between high-fat diets and excess body weight; many high-fat foods are also energy dense. When humans consume diets that are high in fat but low in energy density, energy intake is not greater than would be expected on the basis of the energy density of the foods. Larger food portion size has also been shown to increase food intake. Given the trend in the United States to serve larger portions of food and beverage, this could contribute to greater obesity risk. Food variety can also affect energy intake. An increased variety of entrees, sweets, snacks, and carbohydrates in the diet is associated with an increase in body fatness and food intake. In contrast, an increase in the variety of vegetables available does not appear to increase energy intake and is not associated with increased body fatness. Other factors that may have broad population effects in the United States include the reduced costs of food, increased availability, and palatability. Finally, there is evidence that consumption of sugar-sweetened beverages, such as soft drinks and fruit juices, is not accompanied by a decrease in food intake to offset the extra energy intake. The implication is that some types of beverages will add to the energy intake during the day and promote weight gain.

A number of psychological factors also influence how the properties of food affect energy intake. Individuals vary with respect to their dietary restraint (the tendency to consciously limit food intake to control weight), their feelings of hunger, or their disinhibition (the tendency to overeat opportunistically). It has been proposed that interindividual differences in these factors modify how food variety and portion size affect the eating profile. The social context in which food is consumed and the emotional state of the individual also modulate food intake.

TABLE 220-2 ENVIRONMENTAL FACTORS PROMOTING OBESITY

DIETARY	ACTIVITY
↑ Energy density of foods	↑ Sedentary behavior
↑ Portion size	↓ Activities of daily living
↑ Variety*	↓ Employment physical activity
↑ Palatability	
↑ Availability	
↓ Cost	
↑ Caloric beverages (sugar-sweetened beverages)	

*Variety of sweets, snacks, and entrees.

Physical Activity

Physical activity can be divided into three categories: (1) exercise (fitness- and sports-related activities); (2) work-related physical activity; and (3) non-exercise, nonemployment (spontaneous) activity. Tables are widely available that allow one to calculate energy expenditure on the basis of an individual's weight as well as the type and duration of exercise. Only about 20 to 30% of Americans engage in exercise at the recommended frequency, intensity, or duration that could be expected to have a protective effect on the development of obesity and other health problems, but this does not seem to have changed in recent decades. Recent data suggest that the amount of time spent in sedentary activities (e.g., watching television, using the computer) is an independent predictor of metabolic abnormalities associated with obesity over and above the effects of exercise. Thus, to the extent that reduced physical activity is contributing to the epidemic of obesity, it is likely that it is reduced employment-related and spontaneous physical activity that is changing.

Although it is becoming easier for individuals to measure the energy expended in nonexercise activity with step counters and electronic motion detection devices, there are insufficient longitudinal, population-based data to define the extent to which changes in this activity parameter have occurred. Certainly, employment-related physical activity has decreased with the advent of more automated systems in the workplace. One estimate suggests that between 1982 and 1992, energy expenditure at work decreased by about 50 kcal/day. The additional workplace changes since that time have probably reduced employment physical activity further.

The other component of nonexercise physical activity, the activities of daily living, has probably been reduced by the plethora of labor-saving conveniences (e.g., drive-through food and banking, escalators, remote controls, e-mail, online shopping) now available. Again, there are few hard data to assess how much of a change has actually occurred, although a reduction in daily walking trips and an increase in daily automobile trips have been documented.

There is a large amount of information on how differences in sedentary activity (television watching, video games, and computer use) relate to obesity and obesity complications. The evidence indicates that more time spent in sedentary pursuits is associated with an increased risk of overweight and obesity. The striking aspect to these studies is that the adverse effect of sedentary activities is independent of participation in traditional exercise activities.

Understanding the contributions of decreased work-related physical activity, decreases in activity of daily living, and increases in sedentary behavior can help the physician working with the patient to uncover patterns that may relate to weight gain. Physicians who are aware of these environmental factors are in a better position to help their obese patients identify which of these

environmental factors are contributing to the problem and to develop plans for intervention. In this regard, patients who regularly use step counters or other types of activity-monitoring devices will be better able to self-identify and modify their behavior to obtain sufficient physical activity.

Regulation of Body Weight and Energy Balance

The regulation of adult body weight is a well-balanced process. For example, the typical U.S. adult will take in and expend approximately 2000 to 3000 kcal/day. If there were a consistent error of even 1% in overconsumption of food, this would result in the gain of approximately 25 to 30 pounds of fat every 10 years, assuming no change in energy expenditure. It follows that most adults regulate their average energy balance with greater than 1% precision. There appears to be regulation of both energy intake and energy expenditure through conscious and unconscious processes.

The excess energy consumed by adults is generally stored as triglycerides in adipocytes. Humans continuously recruit new adipocytes from a large preadipocyte pool to replace dying adipocytes. Although the primary means by which abdominal adipose tissue mass expands is through increased fat cell size (adipocyte hypertrophy), this process can store only a limited amount of fat. Adults who gain leg fat accumulate more rather than larger adipocytes on average, resulting in a net increase in adipocyte number as more new adipocytes are created than needed to replace dying cells. Some adults recruit new adipocytes more readily than others do and thus gain weight more so from adipocyte hyperplasia (increased fat cell number) than from hypertrophy. Those who gain fat with large adipocytes, especially in association with an adipose tissue inflammatory response (greater numbers of classically activated macrophages and other immune cells), are more likely to be insulin resistant and to have signs of low-grade systemic inflammation (increased C-reactive protein, mildly elevated interleukin-6 and tumor necrosis factor).

Leptin, a cytokine family protein that is secreted almost exclusively by adipocytes, was the first identified adipose tissue hormone; it has been shown to have potent central nervous system effects on food intake in humans. Leptin also has other hypothalamic-pituitary functions and is proposed to have diverse peripheral physiologic actions. The leptin-deficient animal model of obesity, the *ob/ob* mouse, is severely obese, hyperphagic, hypometabolic, and sexually immature and has low levels of spontaneous activity. Administration of leptin to this animal corrects all of these defects. A few leptin-deficient humans (due to mutations in the leptin gene) have been identified. These children had very low plasma leptin concentrations, were hyperphagic and severely obese, and responded to exogenous leptin administration with dramatic weight loss, reduced food intake, and accelerated maturation of the pituitary-gonadal axis. Overwhelmingly, however, obese humans are not leptin deficient and in fact have high plasma leptin concentrations unless they are in a major negative energy balance circumstance. Because leptin is secreted as a function of percentage body fat, and because women have more body fat than men for any given BMI, they also have higher plasma leptin concentrations. Thus, screening for leptin deficiency is not warranted except in severe, hyperphagic obesity that begins in early childhood, is accompanied by sexual immaturity, and exists in the absence of other known causes (e.g., Prader-Willi syndrome).

Some animal models of genetic obesity (the *db/db* mouse and *fa/fa* rat) have defective leptin receptors, making them unresponsive to leptin. Although rare cases of obese humans with defective leptin receptor genes have been reported, again it appears that leptin resistance due to leptin receptor defects (or genetic post-receptor signaling abnormalities) is extremely uncommon. Clinical screening for leptin receptor mutations is not warranted, given that no treatment exists.

Energy Intake

Much of what has been learned about the biologic regulation of food intake has been from the study of animal models. These signals may affect different aspects of eating behavior. They can affect *hunger*, the compelling need or desire for food; *satiation*, the state of being satisfactorily full and unable to take on more; or *satiety*, the sense of no longer being hungry, a complex set of postprandial events that affect the interval to the next meal or the amount consumed at the next meal. Some of the signals that alter eating behavior affect one aspect and others affect multiple aspects. For example, ghrelin, a peptide produced by the stomach, increases hunger but does not appear to affect satiation or satiety. Cholecystokinin causes satiation but has no effect on satiety. Leptin appears to act on multiple pathways; leptin deficiency is associated with increased hunger and reduced satiation and satiety.

TABLE 220-3 SUGGESTED BIOLOGIC MODULATORS OF FOOD INTAKE

PERIPHERAL SIGNAL	PROPOSED EFFECT ON FOOD INTAKE
Vagal	–
Cholecystokinin	–
Apolipoprotein A-IV	–
Insulin	–
Peptide YY_{3-36}	–
Glucagon-like peptide 1	–
Other glucagon-related peptides	–
Leptin	+ when leptin ↓↓
Ghrelin	+
Tumor necrosis factor-α	–
Obestatin	–

TABLE 220-4 CENTRAL NERVOUS SYSTEM MODULATORS OF ENERGY BALANCE

CENTRAL ANABOLIC (↑ INTAKE)	CENTRAL CATABOLIC (↓ INTAKE)
Neuropeptide Y	α-Melanocyte-stimulating hormone
Agouti-related protein	Corticotropin-releasing hormone
Melanin-concentrating hormone	Thyrotropin-releasing hormone
Hypocretins and orexins	Cocaine- and amphetamine-regulated transcript
Galanin	Interleukin-1β
Norepinephrine	Urocortin
Endogenous endocannabinoids (anandamide and 2-arachidonoylglycerol)	Oxytocin
	Neurotensin
	Serotonin

Peripheral satiety signals act to inhibit further food intake at some point during meal consumption. Some of the signals reach the brain through the vagus nerve and some through the systemic circulation. Examples of the proposed factors modulating appetite are listed in Table 220-3. The compounds range from gut-derived (ghrelin, cholecystokinin, glucagon-like peptide 1) and pancreas-derived (insulin) hormones to peptides such as apolipoprotein A-IV, which is secreted with chylomicrons. The signals are thought to be triggered both by mechanical stimuli (e.g., the fullness of the stomach) and by the presence of nutrients in the jejunum and ileum.

The central nervous system regulation of food intake is becoming better understood. A number of neuropeptides, lipid derivatives, and monoamines have either anabolic (increased food intake with or without decreased energy expenditure) or catabolic (decreased food intake with or without increased energy expenditure) properties. A list of these molecules is provided in Table 220-4. Many of these compounds serve more than one function, such as regulation of hormone secretion (thyrotropin-releasing hormone and corticotropin-releasing hormone), wakefulness (norepinephrine), and behavior-reinforcing systems (endocannabinoids).

Energy Expenditure

There is a wide range of daily energy expenditure in adults, from less than 1400 kcal/day to more than 5000 kcal/day, with larger, more physically active individuals having the greatest energy needs. Typically, daily energy expenditure is divided into resting (or basal) metabolic rate, the thermic effect of food, and physical activity energy expenditure.

Basal Metabolic Rate

The basal metabolic rate (BMR) is the energy expenditure of lying still at rest, awake, in the overnight postabsorptive state. The resting metabolic rate (RMR) is similarly defined but is not necessarily measured before arising from bed. For most sedentary adult Americans, the RMR represents the

major portion of energy expended during the day and may range from less than 1200 to more than 3000 kcal/day. Most (~80%) of the BMR can be explained by the amount of lean tissue an individual has. There are a number of formulas that can be used to estimate BMR. The Harris-Benedict formula (available through numerous online calculators) predicts BMR on the basis of height, weight, age, and sex and is accurate to within 10% in approximately 90% of adults with BMIs of 18.5 to 45 kg/m^2.

Not all components of lean tissue consume oxygen at the same relative rates. Visceral or splanchnic bed tissues account for about 25% of RMR but a much smaller proportion of body weight. The brain, which is only a small percentage of body weight, accounts for almost 15% of RMR. Likewise, the heart (~7%) and kidneys (~5 to 10%) account for greater portions of resting energy needs than their relative contribution to body mass. In contrast, resting muscle makes up 40 to 50% of lean tissue mass but accounts for only 25% of RMR. This contribution changes dramatically with exercise, however, at which time muscle can account for 80 to 90% of energy expenditure. Adipose tissue is a minor contributor to daily energy expenditure, consuming only approximately 3 kcal/kg of body fat per day.

Brown fat is adipose tissue that expresses large amounts of uncoupling protein-1, a protein that allows a mitochondrial membrane proton leak, resulting in heat release as opposed to chemical work from adenosine triphosphate—"uncoupling" of substrate oxidation from chemical or mechanical work. This thermogenic tissue was thought to be present only in human infants but has recently been shown to exist in adults.[4] Methods used to detect brown fat largely rely on ^{18}F-fluorodeoxyglucose positron emission tomography scanning of humans exposed to cold. Lean adults are more likely than obese adults to have brown fat, and brown fat is more readily detectable after obese adults lose weight. Whether brown fat plays any meaningful role in thermogenesis is currently a matter of debate.

Although most of the RMR can be accounted for by the mass of lean tissue, there are also other, more subtle influences on RMR. Age, sex (women have slightly lower BMR even corrected for fat-free mass), and fat mass affect RMR. Small changes in BMR occur during the menstrual cycle (luteal phase > follicular phase). There is also evidence that heritable or family factors influence BMR, accounting for as much as 10% of the interindividual differences.

There are both obligatory and facultative components to RMR. With an energy-restricted diet, significant reductions in BMR relative to the amount of fat-free mass occur. Reductions in the production of triiodothyronine from thyroxine and the sympathetic nervous system drive are thought to contribute to this phenomenon. Likewise, during brief periods of overfeeding, RMR increases slightly above that which would be expected for the amount of lean tissue present.

It has been proposed that individuals with BMRs lower than predicted are at increased risk of future weight gain. Published data suggest that the relative risk is small, and clinical effort to identify such patients is not warranted. Measurement of BMR is sometimes helpful in the evaluation of patients who insist that they are unable to lose weight while following diets containing less than 1000 kcal/day. Almost without fail, if BMR is measured with a reliable instrument, it is substantially greater than the reported food intake. This underscores the fact that most adults are unreliable in assessing their own food intake.

The Thermic Effect of Food

An average of 10% of the energy content of food is expended in the process of digestion, absorption, and metabolism of nutrients. There is a significant interindividual variability in this value, however, ranging from a low of about 3% to a high of about 15% of meal calories that are "wasted" in the postprandial interval. The thermic effect of a meal is related to its carbohydrate and protein calorie content; the fat content has little stimulatory effect. Both obligatory (60 to 70%) and facultative (30 to 40%) components of the thermic effect of food have been identified. The obligatory components no doubt reflect the energy costs of digestion, absorption, and storage of nutrients. The two factors thought to play a role in the facultative component of the thermic effect of food are the postprandial insulin response and activation of the sympathetic nervous system. The thermic effect of food is somewhat lower in insulin-resistant and obese humans, but this has not been linked to future obesity.

Physical Activity Energy Expenditure

The energy expenditure of physical activity is a product of the amount of work done and the work efficiency of the individual. Tracking the total amount of physical activity that humans perform throughout the day is becoming easier with a variety of relatively inexpensive devices. By doing so, it is also possible to calculate the energy expended with published values for estimating the energy costs of work performed. Work units are expressed as metabolic equivalents (METs), a multiple of the RMR. If an individual's RMR is 1 kcal/minute (1440 kcal/day), a workload of 5 METs would be 5 kcal/minute. Although most sedentary individuals can work for only a limited amount of time at relatively low workloads, highly trained athletes can work at extremely high METs (>16) for extended periods. This is because athletes have both a greater peak work capacity (or maximal amount of calories or oxygen that can be consumed) and a higher lactate threshold. The lactate threshold is closely related to the level at which exercise begins to become so uncomfortable that it cannot be maintained much longer. The biochemical definition of lactate threshold describes the progressive rise in blood lactate concentrations observed during sufficiently high-intensity exercise. The lactate threshold may range from 50 to 90% of an individual's peak work capacity. Training raises the lactate threshold closer to the maximal workload and thus allows individuals to work at higher rates for longer periods. Thus, highly fit individuals can expend much greater amounts of energy per minute of exercise with less sense of discomfort than can obese, sedentary individuals who typically have low aerobic fitness and low lactate thresholds (sometimes on the order of 4 to 5 METs). The lactate threshold can be even lower in obese patients with type 2 diabetes, such that walking a mere 3 miles per hour can exceed their lactate threshold. Appreciation of the physical limitations of patients, which can usually be overcome with a carefully designed training program, is necessary to provide realistic activity recommendations.

Exercise (fitness- and sports-related activities) is commonly considered the main component of physical activity thermogenesis. Because most adults do not exercise at high levels or for a sufficient duration to expend a large amount of energy, focusing solely on "exercise" as the main component of physical activity will miss significant opportunities for improving energy balance. The benefits of and energy expended in nonexercise activity can be far greater than with exercise, given the limited amount of time and effort that most patients can commit to exercise.

Nonexercise activity thermogenesis (NEAT) is the calorie expense of performing all activities other than exercise- or employment-related and spontaneous activity. The range of observed NEAT under controlled (metabolic chamber) conditions has been less than 100 to more than 800 kcal/day. The energy expended from a physically demanding job or volitional exercise may or may not be offset by reductions in spontaneous (nonemployment) activity. For example, young adult men and women respond differently to 1 year of extra exercise; men lose weight and women do not, despite the absence of detectable change in food intake. Women must either reduce spontaneous activity in response to exercise or have subtle increases in food intake. It has also been shown that the variations in unconscious increases in NEAT relate strongly to the amount of fat gained in response to overeating. Low levels of NEAT have been reported to predict future weight gain in some populations, and there may be differences between lean and obese persons in the daily amount of NEAT, which could relate to differential tendencies to regulate weight.

Secondary Causes of Obesity

Medications

A number of medications cause weight gain in some or most of the patients for whom they are prescribed. Awareness of the medications that have this potential can facilitate weight loss treatment in some patients. Table 220-5 lists a number of medications that are associated with weight gain as well as alternative treatment approaches, if any, for the underlying condition.

Diseases

Less than 1% of obese patients have an underlying disease that can explain the development of their obesity. Endocrinopathies are the most common secondary cause of obesity. These include Cushing's syndrome (Chapter 227), hypothalamic damage resulting in overeating (most commonly after pituitary surgery), insulinoma (Chapter 230), and hypothyroidism (Chapter 226). A Cushing syndrome–like fat distribution is common; therefore, other physical or laboratory findings are the best clues to whether to test for this condition. These include the classic purple striae, thinning skin, easy bruising, proximal muscle weakness, and electrolyte abnormalities. Correction of Cushing's syndrome commonly results in substantial loss of excess body fat. Insulinoma is a rare tumor, and only a small portion of patients with

TABLE 220-5 PHARMACOLOGIC INFLUENCES IN WEIGHT GAIN AND ALTERNATIVE THERAPIES

DRUGS THAT MAY PROMOTE WEIGHT GAIN	ALTERNATIVE TREATMENTS: WEIGHT NEUTRAL OR WEIGHT LOSS
PSYCHIATRIC AND NEUROLOGIC MEDICATIONS Antipsychotics: olanzapine, clozapine, risperidone, quetiapine, aripiprazole Antidepressants Tricyclics: imipramine, amitriptyline Triazolopyridines: trazodone Serotonin reuptake inhibitors: paroxetine, fluoxetine, citalopram Tetracyclics: mirtazapine Monamine oxidase inhibitors Antiepileptic drugs: gabapentin (higher doses), valproic acid, carbamazepine, divalproex Mood stabilizers: lithium, carbamazepine, lamotrigine, gabapentin (higher doses)	**ALTERNATIVE PSYCHIATRIC AND NEUROLOGIC MEDICATIONS** Ziprasidone Nortriptyline, bupropion, nefazodone, fluvoxamine, sertraline, duloxetine Topiramate, zonisamide (weight loss), lamotrigine (less weight gain)
STEROID HORMONES Progestational steroids Corticosteroids Hormonal contraceptives	**ALTERNATIVES TO STEROID HORMONES** Barrier methods, intrauterine device Nonsteroidal anti-inflammatory drugs
ANTIDIABETES AGENTS Insulin (most forms) Sulfonylureas Thiazolidinediones	**ALTERNATIVE ANTIDIABETES AGENTS** Metformin Acarbose, miglitol Exenatide Dipeptidyl peptidase 4 inhibitors Liraglutide Sodium-glucose co-transporter 2 inhibitors
ANTIHISTAMINES Commonly reported with older agents; also oxatomide, loratadine, and azelastine	**ALTERNATIVE TO ANTIHISTAMINES** Decongestants, mast cell stabilizers, antagonists of endogenous mediators of inflammation
ANTIHYPERTENSIVE AGENTS α-Adrenergic and β-adrenergic receptor blockers Calcium channel blockers: nisoldipine	**ALTERNATIVE ANTIHYPERTENSIVE AGENTS** Angiotensin-converting enzyme inhibitors Calcium channel blockers: most other agents Angiotensin receptor blockers Diuretics
HIGHLY ACTIVE ANTIRETROVIRAL THERAPY	

insulinoma develop obesity. The weight gain associated with hypothyroidism is largely due to fluid retention and resolves with thyroid hormone replacement. Unfortunately, successful treatment is not available for hyperphagia due to hypothalamic damage. Adult patients with growth hormone deficiency, most commonly after hypophysectomy, may lose excess body fat with growth hormone replacement therapy.

Psychosocial Aspects of Obesity

Sexual, physical, and emotional abuse, especially in women, can result in long-term adverse consequences, including obesity. The effects of the abuse tend to be most profound if it occurs in childhood and adolescence. These women may be severely obese, suffer from chronic depression, and experience a number of psychosomatic symptoms, particularly chronic gastrointestinal distress. Identifying these issues before initiation of weight loss programs is important because successful weight loss may actually aggravate the distress experienced by these women. In addition, appropriate referral for psychiatric help may be needed before initiation of treatment for obesity.

PATHOPHYSIOLOGY

Metabolic Complications of Obesity

A central or upper body fat distribution is more predictive than total fat mass of the metabolic complications of obesity. Adipose tissue release of free fatty acids (FFAs) and glycerol into the circulation through lipolysis provides 50 to 100% of daily energy needs. Adipose tissue lipolysis is regulated primarily by insulin (inhibition) and catecholamines (stimulation), although growth hormone, cortisol, and atrial natriuretic peptide also stimulate lipolysis. Upper body obesity is associated with several abnormalities of adipose tissue lipolysis, most remarkably with higher postprandial FFA release and concentrations; this abnormality is particularly evident in type 2 diabetes mellitus. Abnormally high FFA concentrations can contribute to a number of the metabolic complications of obesity.

Insulin Resistance

The term *insulin resistance* is typically used in referring to the ability of insulin to promote glucose uptake and to inhibit the release of glucose into the circulation. The primary site of insulin-stimulated glucose uptake, oxidation, and storage is skeletal muscle. The principal site of glucose production is the liver. Insulin resistance initially leads to hyperinsulinemia and may eventually lead to the development of type 2 diabetes mellitus (Chapter 229).

The ability of insulin to promote glucose uptake, oxidation, and storage in muscle and to suppress plasma FFA concentrations is reduced in upper body obesity. High plasma FFA concentrations can induce a state of insulin resistance both in the muscle (glucose uptake) and in the liver (glucose release), independent of obesity. Thus, abnormal regulation of adipose tissue FFA export is a significant component of the development of insulin resistance. It is hypothesized that excess FFAs induce muscle insulin resistance by promoting increased synthesis of diacylglycerols and ceramides, both of which can interfere with the normal insulin signaling pathway.

Dysregulated production of a number of adipose-derived hormones, also called adipokines, is hypothesized to contribute to insulin resistance and the metabolic complications of obesity. Adiponectin, an adipocyte-derived hormone that improves insulin action, is secreted at reduced rates in obesity and diabetes. Increased production of resistin, interleukin-6, tumor necrosis factor, and retinol-binding protein-4 by adipose tissue has been linked to insulin resistance in animal models. We currently lack the experimental evidence from human studies to know what role adipokines play in the metabolic complications of obesity.

Islet Cell Failure and Type 2 Diabetes Mellitus

Type 2 diabetes usually results from defects in both insulin secretion and insulin action (Chapter 229). Many obese individuals are insulin resistant, yet only a subset will develop diabetes mellitus. It follows that those who develop type 2 diabetes develop pancreatic β-cell decompensation with subsequent hyperglycemia. Animal (rodent) studies have suggested that a process referred to as lipotoxicity is involved in pancreatic β-cell failure. In this model, increased FFAs are proposed to contribute to the insulin secretory abnormalities seen in obesity and ultimately to lead to β-cell failure. There is some evidence that elevated FFAs have adverse effects on islet β-cell function in humans. Another potential contributor to β-cell failure in obesity is the overproduction of islet amyloid polypeptide. This protein is co-secreted with insulin and, because of its tertiary structure, can form toxic amyloid deposits in β cells. Amyloid deposits have been found in the pancreatic islets obtained at autopsy from patients with type 2 diabetes mellitus.

Hypertension

Blood pressure can be increased by a number of mechanisms (Chapter 67). Increased circulating blood volume, abnormal vasoconstriction, decreased vascular relaxation, and increased cardiac output may all contribute to hypertension in obesity. The effect of hyperinsulinemia to increase renal sodium absorption may contribute to hypertension through increased circulating blood volume. Abnormalities of vascular resistance also contribute to the pathophysiologic process of obesity-related hypertension. Under some experimental conditions, elevated FFAs have been found to cause increased vasoconstriction and reduced nitric oxide–mediated vasorelaxation, similar to that seen in the metabolic syndrome. Some obese adults have increased sympathetic nervous system activity, which could contribute to obesity-associated hypertension. Finally, angiotensinogen (also produced by adipocytes) is a precursor of the vasoconstrictor angiotensin II and is proposed to contribute to elevated blood pressure.

Dyslipidemia

Upper body obesity and type 2 diabetes mellitus are associated with increased triglycerides, decreased high-density lipoprotein (HDL) cholesterol, and a high proportion of small, dense low-density lipoprotein (LDL) particles (Chapter 206). This dyslipidemia contributes to the increased cardiovascular risk observed in the metabolic syndrome. Fasting hypertriglyceridemia is caused by increased hepatic very low density lipoprotein (VLDL) secretion, which may be driven by increased delivery of FFAs to the liver coming from both visceral fat and upper body subcutaneous fat. The reduced HDL cholesterol concentrations and the increased small, dense LDL particle concentrations associated with upper body obesity are likely an indirect consequence of elevated triglyceride-rich VLDL. Increased cholesterol ester transfer protein activity and hepatic lipase activity can theoretically account for the atherogenic shifts in triglycerides and cholesterol between lipoproteins. Genetic influences play a significant role in the expression of these lipid abnormalities. Polymorphisms in the genes for apolipoprotein E, lipoprotein lipase, apolipoprotein B-100, and apolipoprotein A-II are correlated with increased triglycerides and decreased HDL.

Endocrine Manifestations of Obesity

Obesity is associated with abnormalities of the endocrine system, one of the most common being polycystic ovary syndrome. This syndrome (Chapter 236) is characterized by mild hirsutism and irregular menses or amenorrhea with anovulatory cycles. It is most commonly linked with obesity and often improves with weight loss and other treatments that improve insulin resistance. The insulin resistance associated with obesity may trigger the development of polycystic ovary syndrome in susceptible individuals. Whereas mild to moderate androgen overproduction is a feature of upper body obesity in women, obese men may suffer from mild to severe hypothalamic hypogonadism. This androgen deficiency improves with weight loss, and attempts to treat this condition with testosterone replacement offer little clinical benefit. There has been some concern that testosterone treatment of obese men may increase the risk of obstructive sleep apnea and perhaps even cardiovascular events. Although estrogens are not elevated in obese premenopausal women, they remain somewhat above postmenopausal levels in obese postmenopausal women. Serum growth hormone concentrations are commonly low in obese adults, but insulin-like growth factor-I concentrations are often normal, and growth hormone concentrations increase with weight loss. Treatment of these patients with growth hormone has been reported to worsen insulin resistance and glucose intolerance and cannot be justified, considering the costs and poor risk-to-benefit ratio.

Mechanical Complications of Obesity

The excess body weight associated with obesity is thought to be responsible for the increased prevalence of lower extremity degenerative joint disease. Extreme obesity can result in premature degenerative joint disease, and this may be especially difficult to treat surgically, given the greater stress on joint replacements. Severely obese individuals may also have problems with venous stasis, which is occasionally aggravated by right-sided heart failure (see later).

Obstructive Sleep Apnea and Sleep Restriction

Sleep apnea (Chapter 100) is common in severely obese patients, tending to be more prevalent in men and in women with an upper body/visceral obesity. Sleep apnea is most likely explained by enlargement of upper airway soft tissue, resulting in collapse of the upper airways with inspiration during sleep. The obstruction leads to apneas, with hypoxemia, hypercarbia, and high catecholamine and endothelin levels. The frequent arousals to restore breathing result in poor sleep quality. Sleep apnea is associated with an increased risk of hypertension, and if sleep apnea is severe, it can lead to right-sided heart failure and sudden death. A history of daytime hypersomnolence, loud snoring, restless sleep, or morning headaches is suggestive of obstructive sleep apnea. Treatment of sleep apnea is important to improve cardiovascular risk, and the failure to recognize and to treat this complication may make weight loss intervention strategies much less successful.

Epidemiologic studies have linked short sleep duration and disruptions of circadian rhythm with increased risk of metabolic syndrome and diabetes. Experimentally induced sleep restriction combined with circadian disruption in humans led to decreased RMR and increased postprandial plasma glucose levels due to inadequate insulin secretion.

Cancer

The risk of breast cancer and endometrial cancer is increased in obese women (Chapter 180). It is thought that this may be due to the increased estrogen levels associated with obesity in the postmenopausal woman. Obese men also have a higher mortality of cancers of the prostate and colon. The reasons for this association are unknown.

Gastrointestinal Disorders

Gastroesophageal reflux disease and gallstones are more prevalent in obese patients. Likewise, fatty liver and nonalcoholic steatohepatitis (Chapter 152) are more common in obesity. Nonalcoholic steatohepatitis can eventually progress to life-threatening hepatic cirrhosis. Weight loss and interventions that improve insulin sensitivity have been shown to improve fatty liver and nonalcoholic steatohepatitis.

DIAGNOSIS

Evaluation of Obesity

In the office practice, obtaining height and weight allows calculation of BMI. For patients with a BMI above 25 and below 35, a second piece of information—the waist circumference—provides an added indicator as to whether the patient is at greater risk for adverse consequences (see earlier). Measurement of blood pressure (which may require a large blood pressure cuff) then provides a third item of health information at almost no cost. The presence or absence of dyslipidemia (HDL cholesterol < 45 mg/dL for women, HDL cholesterol < 35 mg/dL for men, or triglycerides > 150 mg/dL), hypertension, glucose intolerance and diabetes, and hyperuricemia should be documented. A history suggestive of sleep apnea should prompt a referral for overnight oximetry or a sleep disorder evaluation.

A review of the patient's lifestyle, including an assessment of physical activity level and eating habits, may help provide information about why the patient is obese. A family history of obesity, or long-standing obesity, provides evidence against a secondary cause of obesity. A careful medication history and social history may help the clinician identify precipitating factors that can be modified. By emphasizing the role of modifiable lifestyle factors that predispose to disease risk, as opposed to focusing solely on the patient's weight, it may be possible to initiate a conversation about weight/disease management in a less threatening manner from the patient's perspective.

Before a patient enters a weight management program, it is helpful to ensure that the patient is interested and ready to make lifestyle changes and has realistic goals and expectations. Patients who expect to lose large amounts of weight in a short time are virtually doomed to disappointment. Medical treatment programs, even if they include pharmacotherapy, struggle to routinely achieve sustained weight loss of more than 10%. Although this amount of weight loss is sufficient to markedly reduce the medical complications of obesity, disappointment with "only" 10% weight loss may cause patients to abandon a medically successful program. Helping the patient to understand that lifestyle changes resulting in achievable (10%) weight loss is a reasonable, initial goal can be one of the more challenging aspects for a physician.

It is sometimes necessary to delay entry into treatment programs if a patient is not ready to make lifestyle changes. In this case, a reasonable strategy is to remind the patient periodically of the potential health benefits of improved activity and eating habits. Once a willingness to make changes is apparent, treatment is more likely to succeed.

TREATMENT

Obesity represents an individual's response to the environment based on genetics and learned behavior and is best viewed as a chronic disease. Therefore, treatment must be considered a long-term issue, much like diabetes, hypertension, or dyslipidemia. Substantial weight loss can be induced through severe calorie restriction, but without approaches to ensure behavioral changes, body fat is invariably regained. To the extent that environmental factors contribute to a patient's overweight status, and to the extent that the macroenvironment is unlikely to change, patients must learn how to make permanent lifestyle changes (eating and activity behavior) to hope for permanent weight loss. Behavior modification approaches,[5] which can help patients recognize and circumvent environmental cues for sedentary behavior and overeating, can increase the likelihood that patients will accomplish these lifestyle changes. A randomized study has shown that intensive lifestyle intervention (as compared to only support and education) is associated with fewer hospitalizations, fewer medications, and lower health care costs in overweight or obese adults with type 2 diabetes.

Reducing energy intake is the most efficient and effective means to lose weight. For example, creating a 500 kcal/day deficit by reduced food intake will theoretically result in the loss of 1 pound of fat per week. Although possible, it is much more difficult to increase energy expenditure by 500 kcal/week through exercise. Higher levels of physical activity can prevent weight gain (or weight regain after weight loss). Some patients are able to change eating and activity habits on their own, given the proper information, whereas others require formal or informal behavior modification interventions (see later) to help make these changes. In some instances, pharmacotherapy or surgery may be needed for treatment of obesity. A flow diagram on how to evaluate and to manage patients with overweight and obesity is presented in Figure 220-1.

Diet

Changes in eating habits must be permanent if weight loss is to be maintained. An experienced registered dietitian can be helpful in the evaluation of a patient's eating habits and will be able to provide the needed education. The diet history may identify eating behaviors that result in excess energy intake. Although it is important to address specific adverse eating behaviors, patients need to understand some general principles regarding diet. Reducing the energy density of food (most commonly accomplished by reducing dietary fat) can allow patients to feel satiated while consuming fewer calories. The NHLBI/AHA/ACC/TOS obesity guideline[6] recommends that providers prescribe 1200 to 1500 kcal/day for women and 1500 to 1800 kcal/day for men. Alternatively, diets that produce an energy deficit of 500 to 750 kcal/day can be recommended. Because there appears to be no clear superiority of one diet over another with regard to weight loss, it is recommended that providers prescribe one of the evidence-based diets that restricts selected food types (e.g.,

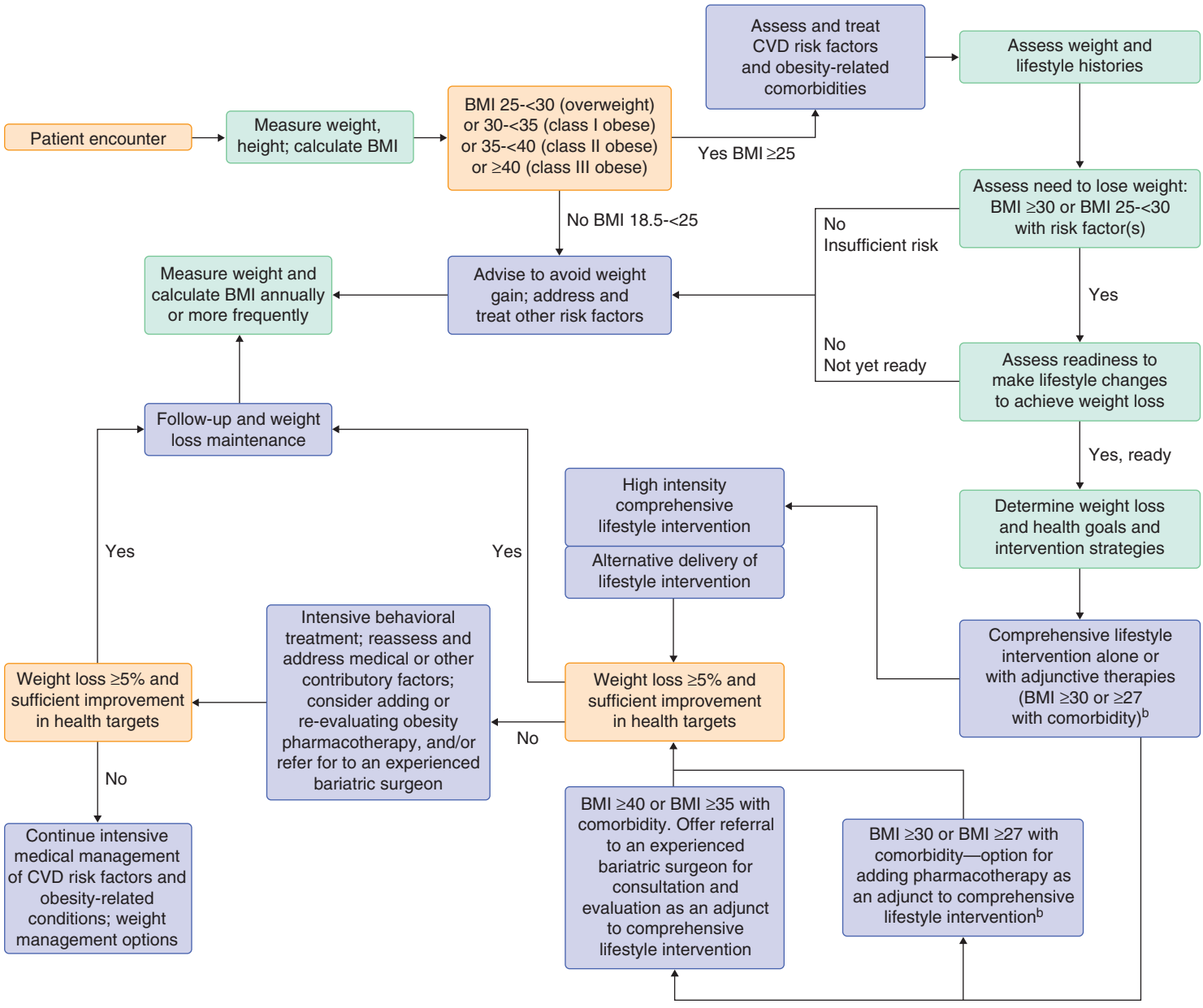

FIGURE 220-1. Flow diagram for the evaluation and management of overweight and obesity. BMI = body mass index; CVD = cardiovascular disease. [b]BMI cutpoint determined by the U.S. Food and Drug Administration (FDA) and listed on the package inserts of FDA-approved obesity medications. (Modified from Jensen MD, Ryan DH, Apovian CM, et al. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. *J Am Coll Cardiol*. 2014;63:2985-3023.)

high-carbohydrate foods, low-fiber foods, or high-fat foods) to create an energy deficit by reduced food intake as well as to address issues such as dyslipidemia, diabetes, and hypertension. Patients should be informed that consuming foods high in water and fiber (fruits, vegetables, legumes, and soups) can provide satiety without excess calories. Patients should be counseled to reduce the intake of beverages containing substantial calories, most often sugar-sweetened beverages. Finally, a regular pattern of eating should be encouraged.

New diets are continually being promoted with the promise of easy weight loss. A common feature of these diets is the claim that special properties of certain foods help people lose weight or are the cause of obesity. If followed, most of these diets result in weight loss because of a reduced energy intake. The reduced intake can be related to the monotony of the diet, and there have been no diets identified that cause persons to lose weight not in accordance with physiologic principles. Although a number of dietary approaches can be successful in promoting weight loss, if there is no peer-reviewed evidence for safety and success of new diets, a review by a dietitian for nutritional safety is warranted. The NHLBI/AHA/ACC/TOS obesity guideline concluded that many types of diets are able to help patients achieve long-term, medically significant weight loss. Thus, it may be less important what specific type of diet (DASH diet, Mediterranean diet, high-carbohydrate/low-fat or high-fat/low-carbohydrate diet) is recommended than for the patient to find dietary adherence to be relatively easy.[7] A comprehensive lifestyle intervention that includes a high-intensity, on-site behavioral intervention provides the best results.

Very low calorie diets (<800 calories per day) are still used to achieve accelerated weight loss. Because the long-term results of these diets are no better and sometimes worse than the results from the standard low-calorie diet combined with behavior modification, these diets are no longer commonly used. The expensive laboratory monitoring required for very low calorie diets without an improved long-term outcome raises questions about the costs versus benefits of this approach.

Physical Activity

A long-term increase in physical activity, either through the activities of daily living or through regular exercise, is key to preventing weight regain, thereby increasing the amount of successful, long-term weight loss. Unfortunately, many overweight and obese patients are unfit, being unable to walk even 1 mile continuously. It is not possible for most adults to expend a great deal of energy through exercise. For example, only about 100 kcal are expended by a 70-kg adult walking one mile. Losing weight solely by increasing exercise is impractical for most patients. However, increasing physical activity as a means of maintaining weight loss is an attainable goal for most patients.

Successful maintenance of weight loss requires that daily energy expenditure be an average of 80 to 90% above RMR. This is a considerable increase for most patients. For example, someone with an RMR of 1500 kcal/day would need to expend about 1000 kcal/day in physical activity to meet this target. Activities other than exercise are important means to achieve this goal. The most commonly applicable approach is by increasing the amount of walking done throughout the day.

The health benefits from regular physical activity over and above the effects on weight include lower cardiovascular and all-cause mortality as well as improved mood and cognition. The options for increasing physical activity include exercise (sports or fitness pursuits) and use of lifestyle approaches. Both methods can improve fitness and allow weight stability; however, persuading obese patients to become more active is not easy. Physicians can begin by asking patients about their current and past activity habits as well as what barriers they see to increasing physical activity. This accomplishes the goal of stimulating patients to think about the issue in a tactful manner. It can help to ask the patient what personal benefits are envisioned as a result of increasing the level of activity. If patients agree to begin an exercise or physical activity program, they will need to monitor their activity and set realistic goals for the amount of exercise they are going to achieve. The ready availability of step counters and electronic activity monitoring devices offers practical means for patients to track physical activity throughout the day and to assess the effects of changes in lifestyle on their activity level. Patients should be advised to acquire devices that have been shown to accurately count steps. Self-monitoring of how many steps are taken each day for 1 to 2 weeks can give patients a good sense of their baseline activity level. Many Americans take as few as 4000 to 5000 steps per day, whereas it may take as many as 15,000 to 17,000 steps per day to help those who have lost significant amounts of weight to maintain that lower weight. Gradually increasing the number of steps regularly taken during the day through a series of changes in habits (e.g., parking farther away, walking during work breaks) is more likely to result in long-term success for most persons than setting aside 2 hours or more for continuous walking.

Behavior Modification

Patients who are unable to make changes in eating activity habits on their own or with informal office counseling may benefit from referral to an interventionist trained in behavior therapy. The goals are to help patients modify their eating, activity, and thinking habits that predispose to obesity and focus on specific pathways to achieve the goals. These pathways may include identifying and removing barriers to development of better eating or activity habits. Small, incremental, and consistent changes in behavior are encouraged. Self-monitoring of food and activity is considered a key feature to success because most obese patients underestimate food intake and overestimate exercise. Cognitive restructuring has been introduced as a way to help overcome the thought processes that can lead to failure of a weight management program. Patients are taught to identify, to challenge, and to correct self-defeating thoughts.

The best weight loss results are provided by in-person, high-intensity (≥14 sessions in 6 months) comprehensive behavioral interventions, which average an 8-kg (5 to 10% of body weight) loss in 6 months. Approaches that provide electronically delivered counseling (telephone or Internet), including some commercial programs, can also achieve weight loss, but generally less than with in-person delivery approaches. Commercial programs that have published their results in peer-reviewed journals are preferred. Physicians who refer patients to programs that offer intensive, comprehensive lifestyle interventions are encouraged to obtain outcomes data from those programs.

TABLE 220-6 INDICATIONS FOR PHARMACOLOGIC TREATMENT OF OBESITY

Body mass index > 27 kg/m^2
One or more complications or conditions that are likely to improve with weight loss
Previous failure of conservative treatment with behavioral intervention, diet, and exercise
Agree to 2- to 4-wk trial of making initial changes in diet and exercise before starting pharmacotherapy
Agree to continued treatment with diet, exercise, and behavioral modification while receiving pharmacologic treatment
Agree to periodic follow-up
Premenopausal women (able to have children) must use some form of contraception
Consider a pregnancy test on initiation of treatment if there is any possibility of pregnancy
No contraindications to the specific drug used for pharmacologic treatment

Pharmacotherapy

A limited number of drugs are currently available to help patients with weight loss.[8] Optimally, any medication for obesity treatment will be prescribed in the context of a comprehensive lifestyle intervention by providers knowledgeable in its use. Not all overweight or obese patients are candidates for pharmacologic treatment of obesity. Table 220-6 provides criteria that should help select patients for pharmacologic treatment. Because pharmacologic treatment of obesity exposes patients to some risks and expense, it is reasonable to require an objective benefit. A rational argument can be made that prioritization should be given to those with one or more medical complications or conditions that are likely to improve with weight loss. In prescribing antiobesity medications, it is important to set clear goals with respect to both weight loss and health benefits.

Currently Available Medications

The medications currently available for long-term use act through either appetite reduction or inhibition of pancreatic lipase, which results in fat maldigestion. Phentermine is approved only for short-term (3 months) use. Recent additions to the medications approved by the Food and Drug Administration for chronic treatment of obesity include lorcaserin, a selective serotonin 2C receptor agonist,[9] the combination of topiramate and phentermine (see later), and a combination of bupropion with naltrexone.[10] Because weight that is lost with pharmacotherapy (especially when it is used without a comprehensive program) is regained once the medication is discontinued, agents that are approved for long-term use are better therapeutic choices.

Orlistat at the typical dose of 120 mg three times daily with meals causes about 30% of dietary fat to be malabsorbed. As expected, adverse gastrointestinal side effects, such as oily spotting, abdominal pain, excess flatus, fecal urgency, and fatty or oily stools, are not uncommon. These side effects decrease over time, and the concomitant use of bulk-forming laxatives (e.g., psyllium, methylcellulose) can reduce these symptoms. A daily multivitamin is recommended for those receiving long-term orlistat therapy. It is not necessary to take orlistat if a nonfat meal is being consumed. Orlistat is now available as an over-the-counter medication. Orlistat improves the results of medical treatment programs that include diet, exercise, and behavior modification, resulting in almost twice as many patients achieving goal weight loss (10% of body weight).

The combination of phentermine and topiramate in an extended-release capsule is approved for chronic treatment of obesity. The highest dose (15 mg phentermine/92 mg controlled-release topiramate) resulted in an average of about 10% weight loss at 2 years (8% more than with placebo), with slightly

more than half of patients achieving 10% weight loss and 15% of patients achieving 20% weight loss.[A] As both components of this medication were previously approved (topiramate for seizures and migraine prevention), the side effects were most commonly upper respiratory tract infection, constipation, paresthesia, sinusitis, and dry mouth; the incidence of individual adverse effects diminishes significantly after the first year.

Lorcaserin inhibits the serotonin pathway in a manner similar to fenfluramine, but without the cardiac valvulopathy effects. Patients taking lorcaserin 10 mg/day for 1 year lost an average of 5.8 kg compared with 2.2 kg with placebo. At 1 year, 47% of patients treated with lorcaserin lost 5% or more of their body weight compared with 20% of the placebo group.[A]

Combined bupropion and naltrexone addresses the issues of dopamine-induced gratification and addictive behavior. The drug results in an average weight loss comparable to other approved drugs, with few side effects and no abuse potential.

Success of Medical Therapy

It has been estimated that more than 95% of those embarking on self-diets or fad diets fail to maintain a significant weight loss for a time that would have meaningful health benefits. The published results from two commercial programs have shown better results. These commercial weight loss interventions provided a comprehensive intervention that was delivered in person and resulted in an average weight loss of 4.8 to 6.6 kg at 6 months in trials in which conventional foods were consumed and 6.6 to 10.1 kg at 12 months in trials in which prepared food was provided. Comprehensive weight management programs delivered at academic medical centers that employ behavior modification, dietary instruction, and physical activity can achieve equally or more impressive results. Average 1-year weight losses of about 10% can be achieved and maintained for 1 to 2 years, depending on the intensity of follow-up. The addition of medications, when indicated (see earlier), can produce even greater amounts of weight loss.

Bariatric Surgery

Surgical treatment can provide more weight loss than medications for class II and class III (see Table 220-1) obese patients with medical complications that could be expected to improve with successful weight loss, such as uncontrolled type 2 diabetes,[11] assuming past attempts at medical treatment have failed.[A] Patients with a BMI of 35 to 40 with life-threatening complications can be considered, but more typically patients with a BMI higher than 40 and several complications are candidates for surgery. Because the risks and costs of surgical treatment are greater than for medical treatment, selection of patients who stand to obtain more potential benefit from surgery should optimize the risk-to-benefit ratio. Contraindications to surgery include active substance abuse, defined noncompliance or inability to comply with medical care, and schizophrenia, borderline personality disorder, or uncontrolled depression.

A multidisciplinary team, including a physician, dietitian, psychologist or psychiatrist with expertise in this area, and surgeon experienced in bariatric procedures, is important for optimal outcome. Defining realistic expectations is an important part of the evaluation process. Patients undergoing bariatric surgery are not likely to be reduced to their ideal body weight. Successful weight loss is typically defined as losing an average of 50 to 60% of excess body weight, which is a difficult criterion to explain to patients. An easier explanation is that most successful patients will achieve weight losses of 25 to 35% of body weight. Follow-up to support the necessary changes in long-term behavior is recommended to optimize weight loss outcomes.

A variety of bariatric surgical procedures have been used. The jejunoileal bypass, long abandoned because of complications that include liver failure, renal failure, and arthropathy, is typically encountered only in those who underwent bariatric surgery more than 3 decades ago and who present with these complications. Procedures that modify the capacity of the stomach (laparoscopic gastric banding and sleeve gastrectomy) but do not create malabsorption of nutrients are employed, but these are generally less effective than the Roux-en-Y gastric bypass in terms of long-term weight loss and outright surgical success. The partial pancreaticobiliary bypass, the very long limb Roux-en-Y gastric bypass, and the duodenal switch procedures create malabsorption that results in greater weight loss than with the standard Roux-en-Y gastric bypass. Unfortunately, the incidence of severe, even fatal vitamin and mineral deficiencies (Chapter 218) is much higher with these procedures. Laparoscopic approaches are now routinely employed for bariatric surgery as they have allowed the time of hospitalization to be reduced considerably and with lesser rates of incisional hernias compared with open procedures.

After surgery, almost all of the weight loss that occurs will happen during the first 1 to 2 years. Long-term (>5 year) success rates are outstanding in good programs. Virtually all patients with successful weight loss will have a dramatic improvement in the medical complications of obesity. For these reasons, bariatric surgery has become an important tool in the treatment of severe, medically complicated obesity.

The results of the Roux-en-Y gastric bypass for treatment of morbid obesity have been favorable. Approximately 70% of patients achieve success as defined previously with this procedure. The mortality and morbidity (e.g., infection, anastomotic leak, wound dehiscence) of this procedure are low in centers with expertise, despite the high-risk population. Laparoscopic gastric banding, once popular because of the reduced short-term risk, is becoming less so as the inferior long-term weight loss results and late complications of band slippage, erosion, and weight regain become more apparent. Whether the sleeve gastrectomy results will more closely mimic gastric bypass or laparoscopic banding is not yet known. After any of the malabsorptive procedures, the patients require permanent follow-up specifically for adverse nutritional consequences.

The long-term follow-up of patients who have undergone gastric bypass surgery is needed to ensure adequate protein, calorie, vitamin, and mineral nutrition. Supplemental vitamin B_{12}, iron, and calcium are routinely added to standard multivitamins. The most common nutritional consequences of malabsorptive procedures are disorders of calcium and vitamin D metabolism, although many severely obese patients have low vitamin D levels even before surgery (Chapter 244). An increase in bone alkaline phosphatase may signal calcium or vitamin D deficiency. Low plasma vitamin D levels and low urinary calcium excretion should prompt aggressive replacement therapy. Iron deficiency, other fat-soluble vitamin deficiencies, and cases of copper deficiency occurring more than 5 to 10 years after surgery have been described. There have also been cases of pancreatogenous hypoglycemia that develop after bariatric surgical procedures. The symptoms are primarily postprandial and are occasionally severe enough to warrant partial pancreatectomy.

PREVENTION

The dramatic increase in the prevalence of obesity during the past few decades strongly suggests that preventive strategies are needed. Public health approaches that emphasize education have been almost uniformly unsuccessful at preventing weight gain or producing weight loss. Public health strategies that virtually impose behavior change are more successful in this regard. Unless widespread efforts are made to address the problem of obesity, it is likely that its prevalence and complications will become an ever-increasing health burden.

Grade A References

A1. Wadden TA, Butryn ML, Hong PS, et al. Behavioral treatment of obesity in patients encountered in primary care settings: a systematic review. *JAMA*. 2014;312:1779-1791.

A2. Espeland MA, Glick HA, Bertoni A, et al. Impact of an intensive lifestyle intervention on use and cost of medical services among overweight and obese adults with type 2 diabetes: the action for health in diabetes. *Diabetes Care*. 2014;37:2548-2556.

A3. Johnston BC, Kanters S, Bandayrel K, et al. Comparison of weight loss among named diet programs in overweight and obese adults: a meta-analysis. *JAMA*. 2014;312:923-933.

A4. Schwingshackl L, Dias S, Hoffmann G. Impact of long-term lifestyle programmes on weight loss and cardiovascular risk factors in overweight/obese participants: a systematic review and network meta-analysis. *Syst Rev*. 2014;3:130.

A5. Garvey WT, Ryan DH, Look M, et al. Two-year sustained weight loss and metabolic benefits with controlled-release phentermine/topiramate in obese and overweight adults (SEQUEL): a randomized, placebo-controlled, phase 3 extension study. *Am J Clin Nutr*. 2012;95:297-308.

A6. Smith SR, Weissman NJ, Anderson CM, et al. Behavioral Modification and Lorcaserin for Overweight and Obesity Management (BLOOM) Study Group. Multicenter, placebo-controlled trial of lorcaserin for weight management. *N Engl J Med*. 2010;363:245-256.

A7. Colquitt JL, Pickett K, Loveman E, et al. Surgery for weight loss in adults. *Cochrane Database Syst Rev*. 2014;8:CD003641.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ogden CL, Carroll MD, Kit BK, et al. Prevalence of childhood and adult obesity in the United States, 2011-2012. *JAMA*. 2014;311:806-814.
2. Jou C. The biology and genetics of obesity—a century of inquiries. *N Engl J Med*. 2014;370: 1874-1877.
3. Casazza K, Fontaine KR, Astrup A, et al. Myths, presumptions, and facts about obesity. *N Engl J Med*. 2013;368:446-454.
4. Harms M, Seale P. Brown and beige fat: development, function and therapeutic potential. *Nat Med*. 2013;19:1252-1263.
5. Lin JS, O'Connor E, Evans CV, et al. Behavioral counseling to promote a healthy lifestyle in persons with cardiovascular risk factors: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014;161:568-578.
6. Jensen MD, Ryan DH, Apovian CM, et al. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. *Circulation*. 2014;129:S102-S138.
7. Longitudinal Assessment of Bariatric Surgery (LABS) Consortium, Flum DR, Belle SH, et al. Perioperative safety in the longitudinal assessment of bariatric surgery. *N Engl J Med*. 2009;361: 445-454.
8. Yanovski SZ, Yanovski JA. Long-term drug treatment for obesity: a systematic and clinical review. *JAMA*. 2014;311:74-86.
9. Hainer V, Aldhoon-Hainerova I. Tolerability and safety of the new anti-obesity medications. *Drug Saf*. 2014;37:693-702.
10. Verpeut JL, Bello NT. Drug safety evaluation of naltrexone/bupropion for the treatment of obesity. *Expert Opin Drug Saf*. 2014;13:831-841.
11. Schauer PR, Bhatt DL, Kirwan JP, et al. Bariatric surgery versus intensive medical therapy for diabetes—3-year outcomes. *N Engl J Med*. 2014;370:2002-2013.

REVIEW QUESTIONS

1. A 52-year-old woman with type 2 diabetes, normal renal function, and a body mass index (BMI) of 36 asks for a referral to a surgeon for evaluation of possible bariatric surgery. Her glycemic control is adequate with a hemoglobin A_{1c} value of 7.2% on 500 mg/day of metformin, 30 U/day of intermediate-acting insulin, and 30 mg/day of pioglitazone. She also has well-managed depression treated with fluoxetine. She has seen a dietitian in the past and has been unable to lose weight. Which of the following is the most appropriate recommendation?

A. Advise the patient to begin a walking program and refer her to a surgeon with experience in laparoscopic banding procedures.
B. Advise the patient to begin a walking program and refer her to a surgeon with experience in laparoscopic gastric bypass procedures
C. Advise the patient to begin a walking program and refer her to a commercial weight loss program that has good outcome results based on peer-reviewed, published data and change the fluoxetine to bupropion.
D. Advise the patient to begin a walking program, gradually increase the metformin to 2500 mg/day, substitute liraglutide for pioglitazone, and request the dietitian to re-evaluate the patient.
E. Refer the patient to an academic weight management program that includes cognitive-behavioral therapy and dietary and exercise intervention components.

Answer: D Although the patient's BMI and medical condition technically meet the criteria for bariatric surgery, her previous weight loss attempts have not included a comprehensive program, and she is taking three medications that have weight gain as a side effect. Failure to address iatrogenic causes of weight gain will predispose to medical treatment failure. Surgery should be an option for those who fail to respond to optimal medical management. Pioglitazone and insulin promote weight gain to a greater extent than fluoxetine does, and the decision to change from fluoxetine to bupropion is best made by a clinician responsible for managing her depression. Increasing the metformin to maximum doses and substituting liraglutide for pioglitazone may reduce the need for insulin while maintaining glycemic control. Revisiting dietary strategies is likely to be more effective in the absence of medications that promote weight gain.

2. A 33-year-old woman presents for treatment of obesity. She has been overweight since her teens and experienced net weight gain with each of two pregnancies such that her BMI is now 32.5 kg/m^2. She had normal glucose tolerance during both pregnancies. Her waist circumference is 29 inches, and her blood pressure in the office is 110/70. Which of the following approaches best matches evaluation/treatment risks and costs with the potential health benefits of weight loss interventions?

A. Recommend over-the-counter orlistat, advise the patient to begin a walking program, and refer her to a dietitian for instruction in a 1200 kcal/day diet.
B. Measure thyroid-stimulating hormone, fasting triglycerides, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, aspartate transaminase, alanine transaminase, and alkaline phosphatase and perform an oral glucose tolerance test to look for hormonal causes of obesity and to screen for metabolic complications. Refer the patient to a dietitian for specific dietary advice based on test results.
C. Advise the patient to begin a walking program, prescribe phentermine/topiramate in an extended-release capsule, and refer her to a dietitian for a 1200 kcal/day diet.
D. Advise the patient to begin a walking program and refer her to a commercial weight loss program that has good outcome results based on peer-reviewed publications.

Answer: D This patient is at low risk for metabolic complications of obesity. She has a low waist circumference for a woman in her BMI range and had no pregnancy-associated glucose abnormalities. Weight gain with pregnancy is common, and in the absence of symptoms of hypothyroidism, laboratory screening is unlikely to be revealing. Costly laboratory testing looking for metabolic abnormalities is likewise a low-yield strategy in this patient. Although this patient is technically a candidate for pharmacotherapy, in practice, the out-of-pocket expense is such that the majority of patients will not continue with treatment because the amount of weight lost is insufficient. In addition, the side effects and risks of some medications make the risk-to-benefit ratio unappealing. Proven commercial weight loss programs can offer support and education not commonly available in a primary care provider's office. However, providers may wish to confirm that local programs maintain the integrity of the program's published results.

3. A 23-year-old man with a BMI of 56 kg/m^2 presents in the office after a work screening examination revealed a fasting blood glucose concentration of 180 mg/dL. The patient admits to nocturia for the past 6 months. He has been severely obese since early childhood and reports being frequently hungry and eating large amounts of food. All members of his immediate family are obese. On examination, he has acanthosis nigricans, large amounts of subcutaneous fat, no cushingoid striae, and normal chest and pubic hair. Which of the following further evaluations is most helpful in making disease management decisions?

A. Administer the Berlin Questionnaire, perform overnight oximetry, and measure hemoglobin A_{1c}, lipids, and liver function.
B. Measure serum leptin concentration to test for leptin deficiency and refer the patient to a center with a leptin treatment IND if indicated.
C. Perform genetic screening for leptin receptor and melanocortin-4 receptor gene mutations.
D. Measure serum testosterone, cortisol, thyroid-stimulating hormone, and growth hormone to screen for endocrine causes of obesity

Answer: A The patient has evidence of sexual maturation, which would not be expected with either leptin deficiency or leptin receptor mutations. Furthermore, there are currently no therapies for patients with leptin or melanocortin-4 receptor gene mutations. Endocrine causes of obesity that begin in childhood would prevent normal adult development (Cushing's syndrome, growth hormone deficiency, hypothyroidism), and testosterone may be low because of obesity. If the patient has sleep apnea, treatment with continuous positive airway pressure is urgently needed, and knowing whether the patient has dyslipidemia or fatty liver disease will inform the urgency of medical or surgical treatment of obesity.

4. A 42-year-old premenopausal woman with type 2 diabetes, dyslipidemia, and a BMI of 44 kg/m^2 failed to lose a clinically meaningful amount of weight during a comprehensive lifestyle program and is now scheduled for gastric bypass surgery in 1 week's time. In the process of evaluating microcytic anemia and elevated alkaline phosphatase, she is found to have severe iron and vitamin D deficiency. Gastrointestinal studies are negative for bleeding or malabsorption. Which of the following approaches best balances the risks and benefits of treatment?

A. Prescribe the patient oral vitamin D 50,000 U three times weekly and ferrous sulfate 325 mg, one tablet three times daily, in preparation for surgery.
B. Prescribe the patient oral vitamin D 2000 U daily and ferrous sulfate 325 mg, one tablet three times daily, in preparation for surgery. Inform the patient that she will likely need parenteral supplements after surgery.
C. Prescribe the patient oral vitamin D 50,000 U three times weekly and ferrous sulfate 325 mg, one tablet three times daily, and defer surgery until her blood levels are replete.
D. After surgery, prescribe the patient oral vitamin D 50,000 U three times weekly and ferrous sulfate 325 mg, one tablet three times daily, and monitor blood levels.

Answer: C Bariatric surgery is never an emergency, although many patients are impatient to have this procedure. Absorption of iron, vitamin D, and calcium is often reduced after gastric bypass, making correction of preexisting deficiencies even more difficult. It is highly unlikely that it will be possible to correct this patient's iron and vitamin D deficiencies in 1 week. Aggressive replacement therapy and deferment of surgery until her deficiencies have been corrected will make it easier to manage this patient after bariatric surgery and may improve the surgical risks.

5. A 48-year-old woman, height 175 cm, weight 104 kg, requests treatment for obesity. She states that she has not lost weight following a 1200 kcal/day diet prescribed by a dietitian despite walking 5 miles/day. Her serum thyroid-stimulating hormone level is normal with no thyroid hormone replacement, and her only medication is an angiotensin-converting enzyme inhibitor for hypertension. She requests instruction in a very low calorie (800 kcal/day) diet and a prescription for a medication to help her lose weight. This patient is most likely to benefit from which of the following approaches?

A. Refer the patient for measurement of a basal metabolic rate to assess whether she is hypometabolic.
B. Ask the patient to use a pedometer to ensure that she is walking 5 miles per day.
C. Prescribe an 800 kcal/day high-protein diet and monitor electrolytes carefully.
D. Request that the patient complete a real-time diet diary for 2 weeks and weigh the patient when she returns.

Answer: D The patient's basal metabolic rate calculated by the Harris-Benedict formula is 1750 kcal/day, which is likely to be within 10% of her actual metabolic rate if it were measured. It is virtually physiologically impossible for her metabolic rate, let alone total daily energy expenditure, to be less than 1200 kcal/day even if she is not walking at all. Even a very sedentary woman at this height, weight, and age would be expected to expend at least 2000 kcal/day. If she is walking 5 miles, her daily energy expenditure will exceed 2500 kcal/day. The patient must not be following a 1200 kcal/day diet. Self-monitoring of food intake is a proven approach to improve awareness of the types and amounts of food consumed. Many patients will lose weight while self-monitoring, and this approach can help identify problem eating habits. On occasion, patients will record perfect eating habits of unbelievably low energy intake and still not lose weight. Under these circumstances, measurement of a basal metabolic rate may help the physician and patient understand that inaccurate perception of food intake is the issue.

221

APPROACH TO THE PATIENT WITH ENDOCRINE DISEASE

DAVID R. CLEMMONS AND LYNNETTE K. NIEMAN

Most endocrine disorders are due to either an excess or a deficiency of a hormone that is transported in the systemic circulation and therefore result in multiorgan manifestations. It is unusual for patients to present with a single isolated set of symptoms that are referable to only one organ system. Additionally, generalized nonspecific symptoms such as weakness, difficulty concentrating, lack of energy, and change in appetite occur commonly in patients with endocrine disorders. Often, a constellation of symptoms is required to point the clinician in the direction of the correct diagnosis, and single symptoms evaluated in isolation are rarely helpful even if one considers an exhaustive differential diagnosis (Table 221-1). Another important feature of the evaluation of patients presenting with endocrinologic disease is obtaining a good longitudinal history. The duration of exposure to a hormone excess or deficiency often dictates the severity of symptoms, and characterizing the symptomatic changes that occur over time can be very helpful in both selecting the tests to confirm the diagnosis and either substantiating the need for treatment or selecting the correct mode of treatment. Physical examination is helpful for confirming the likelihood of a diagnosis; for example, the presence of a symmetrically enlarged thyroid gland indicates that Graves disease is the most likely cause of hyperthyroidism. However, because of the accuracy and precision of modern endocrinologic testing, endocrine disorders sometimes are diagnosed in the absence of symptoms or signs of overt disease (e.g., hyperparathyroidism). In such instances, a thorough history and physical examination are still important because they establish that the patient is truly in the asymptomatic phase of the disease, and this is often helpful in determining whether therapy or observation is required. A careful temporal history of a symptomatic change may also be helpful in the differential diagnosis—for example, in ascertaining whether a thyroid mass is likely due to a hemorrhagic cyst (i.e., occurring suddenly) or is a thyroid adenoma that might have evolved over an extended period. A thorough general history and physical examination may help to establish the diagnosis of diseases that are associated with endocrinologic abnormalities, such as cancers that ectopically secrete hormones, or may help to exclude the likelihood of an endocrinologic etiology of a specific disease manifestation.

Precise genetic testing to establish the etiology of endocrinologic syndromes is becoming more prevalent; therefore, an accurate family history can be useful in determining the need for genetic testing and familial screening. A thorough evaluation of medication use is also mandatory. Some medications can clearly mask the symptoms of overt endocrine disease, such as β-blockers in hyperthyroidism, and others can exacerbate the findings, such as hydrochlorothiazide use in hyperparathyroidism. Often, specific medications will confound laboratory evaluation, such as diuretics in patients with hyperaldosteronism or antihypertensive drugs in patients being screened for pheochromocytoma. Appropriate withdrawal of the medication may be required to establish a firm diagnosis. Medications may also directly alter a laboratory test, such as use of oral contraceptives in attempting to diagnose hyperthyroidism or hypothyroidism. Finally, an accurate history of the specific complaint may be challenging. For example, in evaluating male sexual dysfunction, a corroborating history from the patient's partner is often necessary for verification.

COMMON SYMPTOMS OF ENDOCRINOLOGIC DISEASE

Generalized symptoms such as weakness and fatigue are prominent features of adrenal insufficiency, hyperthyroidism and hypothyroidism, hypopituitarism, and poorly controlled diabetes mellitus. Pain is an uncommon complaint in endocrinologic disorders and is usually only seen with acute endocrinologic emergencies, such as diabetic ketoacidosis or acute adrenal insufficiency (abdominal pain). Chronic pain seen in primary or secondary hyperparathyroidism due to bone reabsorption and in osteomalacia (Chapter 244) can occur in the shafts of the long bones. Symptoms of menstrual dysfunction are common in women with adrenal insufficiency, hypopituitarism, Cushing's syndrome, hyperprolactinemia, hyperthyroidism, hypothyroidism, polycystic ovarian disease, and primary ovarian failure. Weakness and easy fatigability can be secondary to anemia, which is common in adrenal insufficiency, androgen deficiency, hypothyroidism, hyperparathyroidism, and panhypopituitarism.

The most common endocrine-related symptom of intestinal dysfunction is constipation. This occurs frequently in patients with diabetic autonomic neuropathy and in hypercalcemia, hypothyroidism, or pheochromocytoma. Diarrhea can be an early and prominent symptom in hyperthyroidism and metastatic carcinoid tumors. Fever, like abdominal pain, usually only occurs in endocrine emergencies, typically in severe adrenal insufficiency and very severe hyperthyroidism. Generalized hair loss not related to androgen excess occurs in hypothyroidism, hypopituitarism, and thyrotoxicosis. Typical male-pattern baldness can be a feature of hirsutism in women as well as Cushing's syndrome and acromegaly. Persistent recurrent headaches are a feature of expanding pituitary tumors, whereas episodic headaches occur frequently in patients with pheochromocytoma and hypoglycemia. Changes in libido are often present in patients with hyperthyroidism, hypopituitarism, hypogonadism, Cushing's syndrome, and poorly controlled diabetes mellitus. Polyuria and nocturia are features of both diabetes insipidus and diabetes mellitus and can also occur in severe hypercalcemia.

Weight gain is an early symptom of Cushing's syndrome and hypothyroidism. Weight loss accompanied by anorexia is common in adrenal insufficiency, metastatic hormone-producing tumors such as ectopic adrenocorticotropic (ACTH) hormone syndrome, type 1 diabetes mellitus, and panhypopituitarism. Weight loss with no change or an increase in appetite is common in hyperthyroidism (see Table 221-1). An important differential diagnostic issue in patients with weight loss is depression. Depression can occur concomitantly in patients with adrenal insufficiency, Cushing's syndrome, hypercalcemia, and hypothyroidism.

An important neuropsychiatric symptom is widening of mood swings—that is, an exaggeration of normal cyclothymic changes. This occurs in hyperthyroidism and Cushing's syndrome. A history of skin changes occurs with several endocrine disorders. Acanthosis nigricans occurs with severe obesity, polycystic ovarian disease, severe insulin resistance syndromes, Cushing's syndrome, and acromegaly. Acne is a symptom of androgen excess and occurs with androgen-producing tumors, polycystic ovarian disease, and Cushing's syndrome. Generalized hyperpigmentation occurs in Addison's disease and Nelson's syndrome. Dry skin is present in almost all patients with hypothyroidism. It also occurs in panhypopituitarism. In Cushing's syndrome, skin changes such as striae, plethora, and easy bruisability are common. Vitiligo occurs in association with several endocrine autoimmune diseases, most prominently with autoimmune thyroid disease and Addison's disease.

Assessment of any of these symptoms in isolation is unlikely to lead a clinician to the correct diagnosis. However, assessment of combinations of these symptoms (e.g., the combination of weight gain, constipation, cold intolerance, and dry skin in hypothyroidism) is likely to lead to the correct set of diagnostic decisions.

PHYSICAL EXAMINATION

Physical examination of a patient with a suspected endocrine disorder is extremely helpful in terms of confirming the significance of findings that were ascertained by a careful medical history (Table 221-2). Examination of the skin is important because endocrinopathies lead to skin changes that progress over time. These often occur early in the course of the illness and can be a major aid in making the correct diagnosis. Primary adrenal failure is usually accompanied by increased skin pigmentation, particularly over the creases of the palms and extensor surfaces. The oral mucosa is often hyperpigmented, which can be helpful in African Americans. Patients with Cushing's syndrome present with persistent facial plethora, acne, and characteristic striae, which are generally larger than 1 cm and exhibit a dark red to purplish hue. These are most common over the abdomen. The presence of axillary purpura is also suggestive of Cushing's syndrome and results from increased capillary fragility. Patients with acromegaly often present with excessive skin tags that increase in number over time and multiple types of benign skin tumors such as dermatofibromas and lipomas. The skin is also much thicker over the dorsum of the hand, and the amount of subcutaneous tissue in the palms is prominent. Patients with Cushing's syndrome have thinning of the skin over the forehead and around the eyes. Patients with hypothyroidism often present with evidence of hyperkeratosis, particularly over the extensor surfaces. Severe long-standing thyroid disease is also accompanied by myxedema of

TABLE 221-1 SYMPTOM CONSTELLATIONS SUGGESTING SPECIFIC ENDOCRINE DISORDERS

SYMPTOM CONSTELLATION	DIAGNOSIS
Weakness, fatigue, anorexia, loss of appetite, postural blood pressure changes	Adrenal insufficiency
Cold intolerance, dry skin, constipation, weight gain	Hypothyroidism
Fatigue, easy bruising, striae, proximal muscle weakness, central obesity, hypertension, acne	Cushing's syndrome
Weight loss, increased appetite, palpitations, tremor, emotional lability, diffuse hair thinning	Hyperthyroidism
Galactorrhea, amenorrhea, headaches	Prolactin-producing pituitary tumor
Weight loss, anorexia, loss of pubic and axillary hair	Hypopituitarism
Episodic palpitations, tremor, anxiety, headaches, sweating, weight loss	Pheochromocytoma
Episodic flushing, palpitations, abdominal cramping, and diarrhea	Carcinoid syndrome

TABLE 221-2 PHYSICAL SIGNS THAT SUGGEST SPECIFIC ENDOCRINE DISORDERS

Hyperpigmentation of palms, extensor surfaces, and buccal mucosa	Adrenal insufficiency
Facial plethora, moon facies, striae, purpura	Cushing's syndrome
Skin tags, acral enlargement, prognathism, orthodontia	Acromegaly
Proptosis, lid lag, symmetrically diffuse thyroid enlargement, hyperreflexia	Graves disease
Retinal microaneurysms, macular edema	Diabetic retinopathy
Short stature, web neck, loss of tears	Turner's syndrome
Shield-like chest, short fourth vertebra	Primary ovarian failure

the extremities, a specific skin finding of thickening of both the dermis and epidermis due to mucopolysaccharide accumulation. In contrast, patients with Graves disease present with thinning of the skin, but more prominently with skin that is moist and warm. Both adrenal insufficiency and hyperthyroidism or hypothyroidism can be associated with vitiligo. The nails in hyperthyroidism reveal onycholysis in cases of longer duration. Hair changes occur commonly with several endocrinopathies. In hypothyroidism, the hair becomes coarse, whereas in hyperthyroidism, diffuse thinning occurs. Long-standing hypothyroidism can be accompanied by loss of the lateral third of the eyebrow hair. Hirsutism with hair growth over the areolae and along the linea alba is present in patients with gonadal dysfunction or with hyperandrogenism due to polycystic ovarian disease or androgen-producing tumors. In contrast, patients with hypopituitarism or hypogonadism often present with loss of pubic and axillary hair.

The eye examination can be helpful in establishing the diagnosis of hyperthyroidism. Graves ophthalmopathy occurs in 40% of patients and presents with exophthalmos that can be unilateral. Extraocular movement assessment is important in patients with pituitary tumors because local expansion of these tumors can lead to third, fourth, or sixth nerve palsies, whereas compression of the optic tracts can cause the distinctive visual field defect of bitemporal hemianopsia. On neck examination, ascertainment of whether a goiter is smooth, symmetrical, and multinodular or a single nodule can be helpful in guiding further work-up in the evaluation of hyperthyroidism.

Cardiovascular evaluation is helpful for evaluating the severity and duration of endocrine illness. Patients with hyperthyroidism have evidence of increased sympathomimetic activity, including tachycardia, a widened pulse pressure, and prominent precordial activity. Flow murmurs may be heard as a result of increased cardiac output, and occasionally a bruit is easily heard on auscultation of the thyroid. In patients with hypothyroidism and Addison's disease, these findings are reversed. Specific cardiac abnormalities, such as coarctation of the aorta, may be present in patients with primary ovarian failure due to Turner's syndrome. Patients with pheochromocytoma often have a postural blood pressure change of greater than 20 mm Hg. Cardiomyopathy can be a feature of acromegaly. Hypertension is common in many endocrine disorders, including hypercalcemia, hyperparathyroidism, acromegaly, diabetes mellitus, obesity, Cushing's syndrome, primary aldosteronism, and pheochromocytoma.

Nearly all patients with acromegaly demonstrate evidence of hand and foot enlargement, a very unusual occurrence in adulthood. This often manifests as a history of changing ring size or shoe size and is easily demonstrable on physical examination. Patients with hyperthyroidism often have a significant hand tremor and evidence of peripheral bruits on auscultation. Evaluation of the extremities is also helpful in the differential diagnosis of metabolic bone disease. The presence of bowing of the legs can be found in either dietary or X-linked hypophosphatemic rickets. The presence of edema is usually an early sign of disorders of hormone-producing tumors that result in salt retention, including Cushing's syndrome and hyperaldosteronism. Measurement of upper to lower segment ratios and arm span is helpful in establishing the timing and onset of puberty in primary gonadal disorders.

Pelvic examination is a major aid to the differential diagnosis of ovarian disorders. Palpation may identify polycystic ovaries and suggest the absence of ovarian tissue. The absence of a uterus is important in the differential diagnosis of pseudohermaphroditism, and the evaluation of external genitalia can be important in establishing the presence of congenital adrenal hyperplasia. Vaginal dryness is a sign of severe estrogen deficiency, as is breast atrophy. Hyperprolactinemia often first manifests as the presence of expressible galactorrhea.

Neurologic changes can occur in a variety of diseases. Evaluation of the ability to detect monofilament or vibratory sensation is an important tool for evaluating the presence of diabetic neuropathy. The presence of peripheral motor nerve defects in diabetes is also an important presenting sign, as is the presence of a cranial nerve palsy, such as third nerve palsy. Funduscopic examination can reveal microaneurysms even in patients with undiagnosed type 2 diabetes, and this can be helpful for estimating antecedent duration of illness. Hyporeflexia with a delayed relaxation phase is one of the earliest physical changes in hypothyroidism and is prominent in most patients with clinically significant disease; hyperreflexia occurs in hyperthyroidism. Patients with Cushing's syndrome often have extreme proximal muscle weakness, and this occurs in severe long-standing hyperthyroidism. Midgut carcinoids frequently present with severe flushing and diarrhea. The skin characteristically has an unusual purplish hue, and this change typically lasts for the duration of the episode, about 20 to 30 minutes. Mental status changes occur frequently in patients with Cushing's syndrome, hyperthyroidism, and extreme hypercalcemia.

LABORATORY EVALUATION

Patients with endocrinologic diseases are frequently diagnosed in an asymptomatic state as a result of abnormal radiologic evaluation or hormonal testing abnormalities.

The most common radiologic conundrums occur in patients who are noted incidentally to have small pituitary, thyroid, or adrenal masses. Although most of these patients generally do not have a hormonal dysfunction syndrome, proper evaluation to exclude the presence of a functionally active tumor is usually necessary. For pituitary tumors, the minimal evaluation would include measurements of a baseline serum prolactin level and, in cases with suggestive symptoms, of 24-hour urine free cortisol and of growth hormone after glucose suppression. In patients with incidentally found adrenal masses ("incidentalomas"),[1] important findings are a large tumor mass (i.e., >4 cm), hypertension or hypokalemia, and symptoms and signs of Cushing's syndrome. If any of these are present, appropriate evaluation to exclude hyperaldosteronism, Cushing's syndrome, and pheochromocytoma should be undertaken.

The sensitivity and specificity of hormonal testing have reached levels that have significantly changed the initial evaluation of patients for endocrine disorders. For example, screening of asymptomatic patients for thyroid disease, disorders of lipoprotein metabolism, and disorders of gonadal dysfunction are common and widespread.

Hormones are often measured using immunologic detection methods. Most measurements use blood or urine samples to detect the active hormone.

Occasionally, measurement of a metabolite of the active hormone, such as 25-hydroxycholecalciferol (a vitamin D metabolite), is more reliable than that of the parent hormone. Some hormones (e.g., thyroxine) have a very long half-life in plasma, and therefore a measurement at any time of day reflects the ability of the gland to produce that hormone. However, other hormones (e.g., growth hormone) are secreted episodically, and therefore a static measurement may or may not be indicative of hormone excess or deficiency. In these cases, suppression or stimulation testing is used to confirm the diagnosis. Usually, an exogenous substance (e.g., ACTH) is administered either orally or intravenously, and the production of the hormone (e.g., cortisol) by the gland is either stimulated or suppressed.

Many hormones are present in the circulation bound to binding proteins. With the exception of peptide hormones that circulate in the free (unbound) form, this can cause problems in interpretation: medications or concomitant illnesses can result in a major change in the concentration of the binding protein, which in turn alters the total hormone concentration. This problem can be obviated either by measuring the binding protein itself or by direct measurement of the free hormone, such as measurement of free thyroxine.[2] Free hormone measurements, however, can be less reliable than total hormone measurements; therefore, a decision sometimes needs to be made wherein the reliability of the assay is compared with the quality of the information that will be provided.

Often, static hormone measurements are used for screening—for example, morning cortisol in screening for the presence of hypoadrenalism. Subsequently, stimulation or suppression testing is used for confirmation of the diagnosis.[3] In some cases, plasma measurements are much less reliable than urinary testing (e.g., a single morning cortisol level to rule in or out the presence of Cushing's syndrome). In these cases, 24-hour urinary measurement of hormone is often required to document overproduction. Urinary assays have the advantage of providing integrative assessment over 24 hours and thus are less likely to be susceptible to errors due to episodic hormonal secretion over time. Occasionally, measurements of other substances, such as electrolytes or metabolites, are also informative for confirming the diagnosis. For example, measurement of 24-hour urinary calcium can be important in the differential diagnosis of hypercalcemia. Metabolites in the urine may be extremely important in the evaluation of adrenal disorders and in documentation of pheochromocytoma and carcinoid syndrome. Simultaneous measurements of two substances is extremely helpful in the diagnosis of some disorders. Measurement of simultaneous serum calcium and parathyroid hormone (PTH) is important for confirming the presence of hyperparathyroidism. Likewise, simultaneous measurement of blood glucose and insulin can be important in screening for the presence of an insulin-producing tumor. Indirect measurement of hormonal status can also be important; for example, measurement of insulin-like growth factor (IGF)-I, which is inducible by growth hormone (GH), provides an integrative measure of GH secretion. Similarly, measurement of hemoglobin A_{1c} provides an integrative measurement of long-term blood sugar control in diabetes.

Imaging studies are commonly used in endocrine diagnosis. Magnetic resonance imaging and computed tomography are helpful in evaluating pituitary and adrenal masses. Scanning of the thyroid gland using radioactive iodine is useful for evaluation of the functional status of thyroid nodules and the functional activity of the thyroid gland as a whole.[4] Bone mineral density testing is now commonplace in screening of patients for osteoporosis and in the evaluation of patients with established fracture syndromes. Imaging can also be combined with hormonal measurements. Specifically, cannulation of the adrenal veins can be helpful in confirming the presence of functioning adrenal tumors such as aldosteronomas. Likewise, determining the location of tumors such as those producing PTH can often be confirmed by venous sampling procedures. Similarly, intraoperative measurements of hormones that change rapidly, such as intraoperative PTH, can help determine whether surgical removal of the hormone-secreting tumor has been adequate. The primary use of biopsy in endocrinologic diagnosis is fine-needle aspiration of the thyroid gland. This procedure can be done safely in the outpatient setting, often aided by ultrasound guidance. It is a highly accurate method for determining whether further diagnostic evaluation or therapeutic intervention is necessary.

GENETIC EVALUATION

Use of genetic testing in endocrinologic diagnosis has become much more commonplace. Polymerase chain reaction amplification of DNA obtained from peripheral blood cells is often used to determine the presence of a specific disorder. This has been extremely useful in differential diagnosis, for determining prognosis, and for deciding whether family screening is required (e.g., in the presence of multiple endocrine neoplasia). Whether genetic testing is required is often dictated by whether this information is necessary to solve a differential diagnostic problem (e.g., disorders of vitamin D metabolism) or to decide whether more extensive surgery should be required (e.g., in pheochromocytoma) or whether family screening is necessary (e.g., in multiple endocrine neoplasia 2 syndrome).

EVALUATION OF THE RESPONSE OF AN ENDOCRINE DISEASE TO TREATMENT

Most hormone excess syndromes are treated surgically by removal of the particular endocrine gland or tumor that is oversecreting the hormone.[5] However, follow-up of these patients mandates that (1) it be established that the disease has truly been cured by the resection or (2) if residual disease is present, and (3) determining whether repeat operation is likely to be of benefit to the patient or some other means of treating the patient should be undertaken. These decisions generally need to be made in consultation with surgeons and radiotherapists. If an endocrine deficiency is present, hormone replacement therapy is most often used to correct the disorder. In some cases, the efficacy of substitution therapy can be measured directly by laboratory testing, such as measurement of thyroid-stimulating hormone during thyroxine replacement. In other cases (e.g., measurement of ACTH suppression during substitution therapy for adrenal insufficiency), testing could lead to misleading information and overtreatment of patients; therefore, a combination of return of symptoms and signs to normal, laboratory testing, and indirect tests (e.g., potassium, blood urea nitrogen, and creatinine in the case of adrenal insufficiency) is required to determine the adequacy of replacement therapy. Understanding the pharmacology of the particular synthetic hormone used is important for proper replacement therapy. For example, synthetic glucocorticoids vary greatly in their half-life, and therefore dosage and timing of administration are important issues for patients receiving these hormones. Some patients, such as those with hypopituitarism, require substitution with multiple hormones, and often these need to be coordinated (e.g., glucocorticoid dosage in a patient with panhypopituitarism is dependent on the dosage of thyroid hormone that is administered). Because monitoring thyroid-stimulating hormone is not useful in such patients, empirical substitution with both hormones is required to establish that the therapeutic regimen results in return of functional activity to normal. At times, the signs of hormone excesses may have to be treated with ancillary medications; for example, β-blockade may be useful in patients with hyperthyroidism, and both α- and β-blockade are often necessary in patients with pheochromocytoma. The etiology of osteoporosis is often unknown. In most cases, this disease is generally treated by administration of agents that inhibit bone reabsorption by a variety of mechanisms, including the bisphosphonates, estrogen replacement therapy, and calcitonin.

Hormones are also used throughout medicine for treatment of other disorders, and sometimes these treatments result in a hormone excess syndrome. The most common example is administration of high-dose glucocorticoid therapy for immune suppression resulting in Cushing's syndrome. Similarly, growth hormone may be administered to children with short stature who do not have GH deficiency. PTH is the only agent available for stimulation of an anabolic effect in bone; therefore, it is administered to patients with osteoporosis even though they do not have hypoparathyroidism. Octreotide acetate, a long-acting derivative of somatostatin, can be useful for controlling gastrointestinal symptoms in patients with neuroendocrine tumors. Supraphysiologic doses of progesterone are commonly used as contraceptives. Hormone antagonists are also used in both endocrine and nonendocrine disorders. Estrogen receptor antagonists are commonly used in breast cancer therapy, as are androgen receptor antagonists in prostate cancer therapy. Similarly, gonadotropin-releasing hormone agonists have been found to be useful in patients with prostate cancer that is metastatic to bone. Prostaglandin antagonists are commonly used in acute and chronic inflammatory disorders, and angiotensin receptor antagonists as well as renin antagonists are used in the treatment of hypertension, whether it is due to an endocrine or nonendocrine etiology. A comprehensive knowledge of the actions of these hormones can intelligently guide the proper use of these agents in treating nonendocrine disorders.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Nieman LK. Approach to the patient with an adrenal incidentaloma. *J Clin Endocrinol Metab*. 2010;95:4106-4113.
2. Koulouri O, Moran C, Halsall D, et al. Pitfalls in the measurement and interpretation of thyroid function tests. *Best Pract Res Clin Endocrinol Metab*. 2013;27:745-762.
3. Petersenn S, Quabbe JH, Schöfl C, et al. The rational use of pituitary stimulation tests. *Dtsch Arztebl*. 2010;107:437-443.
4. Vazquez BJ, Richards ML. Imaging of the thyroid and parathyroid glands. *Surg Clin North Am*. 2011;91:15-32.
5. Lenders JW, Duh QY, Eisenhofer G, et al. Pheochromocytoma and paraganglioma: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2014;99:1915-1942.

222

PRINCIPLES OF ENDOCRINOLOGY

ALLEN M. SPIEGEL

INTRODUCTION

The principal manifestation of most endocrine diseases is over- or undersecretion of one or more hormones, but the causes of endocrine disease are not unique to endocrinology as a subspecialty of medicine. Benign or malignant proliferation of endocrine cells, destruction of endocrine cells by autoimmune, infectious, or other infiltrative processes, mutations in genes expressed by endocrine cells, and alterations in endocrine cell function caused by metabolic abnormalities or drugs are major causes of endocrine disease shared with diseases of other organ systems. If the causes of endocrine disease are not unique, there are nonetheless some general principles of endocrinology that define it as a medical subspecialty. These principles all derive from the study of hormones. Endocrinology was born with the recognition that certain cells secrete specific chemical entities—hormones—directly into the blood stream to act on specific distant targets. This immediately posed a series of questions: how are hormone synthesis and secretion regulated, how are hormones transported and metabolized, and how do hormones exert their actions on target tissues? This chapter will provide an overview of the answers to each of these questions and how they inform our current approach to the diagnosis and treatment of endocrine diseases.

WHAT IS A HORMONE?

The initial definition of a hormone was based on physiology rather than chemistry. Action on target cells reached via the blood stream was the operative principle. Secretin, now known to be a peptide hormone secreted by enteroendocrine cells in the gastrointestinal lining and acting on pancreatic exocrine cells, was the first example.[1] In contrast to enteroendocrine cells dispersed in the gut lining with other cell types, discrete collections of hormone-secreting cells, endocrine glands such as the adrenals, gonads, thyroid, and parathyroids, were soon recognized and their hormonal secretions chemically characterized. We now know that peptides, steroids, and many other chemical substances fit the definition of a hormone.

Endocrine action, a hormone secreted into the blood stream acting at a distance, has been contrasted with *paracrine action*, a growth factor or other signaling molecule secreted from one cell and acting on adjacent cells, and *autocrine action*, a cell secreting a signaling molecule that acts on the same cell. The distinction between endocrine, paracrine, and autocrine actions is not sharp. In some cases, a factor such as parathyroid hormone–related peptide (PTHrP) that acts physiologically in paracrine fashion during normal bone development may act as an endocrine factor in the syndrome of humoral hypercalcemia of malignancy. The definition of what constitutes an endocrine gland has also blurred. First came the discovery that specialized neurons could synthesize and secrete hormones directly into the blood stream, so-called neuroendocrine action, exemplified by vasopressin secretion by posterior pituitary cells. This contrasts with classic neuronal secretion of neurotransmitters into the synaptic cleft. With increasing recognition that many tissues secrete hormones (e.g., erythropoietin by the kidney and leptin and other adipokines by fat cells), the role of endocrine glands as the exclusive purveyors of hormonal secretions has diminished. This blurring of the boundaries between endocrinology and other medical specialties is a general phenomenon in which study of hormones has informed seemingly disparate fields. Radioimmunoassay, the concept of receptors, and other principles of signal transduction first elucidated in studying hormone action are now broadly applied in all fields of medicine.

REGULATION OF HORMONE SYNTHESIS AND SECRETION

There are two broad categories of hormone synthesis: (1) that responsible for synthesis of peptide hormones and (2) that responsible for synthesis of steroids, including the active form of vitamin D, thyroid hormones, catecholamines, and other nonpeptide hormones. In the former category, hormone structure is encoded genetically. mRNA translation yields a protein precursor (pre-pro-hormone) that is generally cleaved through successive steps to yield the mature secreted product. Some protein precursors such as proopiomelanocortin contain within them multiple hormonal products, adrenocorticotropic hormone (ACTH), melanocyte-stimulating hormone (MSH), and endorphins in that example. In certain pathologic conditions, inappropriate immature hormone secretion occurs (e.g., excessive proinsulin secretion by insulinomas). Post-translational modifications occur for some hormones, such as disulfide bond formation for vasopressin and insulin, C-peptide cleavage for insulin, and glycosylation of the pituitary glycoprotein hormones, thyroid-stimulating hormone (TSH), follicle-stimulating hormone (FSH), and luteinizing hormone (LH). Mutations in genes encoding peptide hormones can lead to disruption of normal hormone synthesis or secretion, a rare cause of hormone deficiency. Peptide hormones are typically stored in secretory granules, and they are secreted by exocytosis, a process regulated by Ca^{++} and other factors. For steroids and other nonpeptide hormones, hormone synthesis is accomplished by a series of enzymatic steps acting on precursors (cholesterol for steroid hormones; aromatic amino acids for thyroid hormones, catecholamines, and related compounds). Mutations in genes encoding enzymes responsible for one or more steps in hormone synthesis can lead to hormone deficiency.

Negative feedback regulation is the general principle that governs normal hormone synthesis and secretion. For endocrine glands whose growth and hormone secretion is stimulated by pituitary trophic hormones (gonads, adrenal cortex, thyroid), the hormone secreted by the gland acts directly on cognate pituitary trophic cells (e.g., cortisol acting on ACTH-secreting pituitary corticotrophs) to suppress hormone secretion (Fig. 222-1). Conversely, a physiologically meaningful reduction in target gland hormone secretion leads to increased pituitary trophic hormone secretion. In many other cases, negative feedback regulation operates without the pituitary as an intermediate (e.g., PTH secretion from the parathyroid glands regulates extracellular Ca^{++} homeostasis, and Ca^{++} feeds back directly on parathyroid cells to regulate PTH secretion). Chronic hormone deficiency with resultant loss of negative feedback can lead to hypersecretion of the cognate trophic hormone and even to neoplastic proliferation of trophic hormone–secreting cells. Examples include Nelson's syndrome in which corticotroph tumors form secondary to adrenalectomy, and tertiary hyperparathyroidism in which parathyroid adenomas occur in the setting of chronic hypocalcemia. In some adrenal cortical disorders, hormone deficiency leading to loss of negative feedback of trophic hormone secretion causes pathologic hypersecretion of alternative steroid hormones. The various forms of congenital adrenal hyperplasia are caused by mutations in one of the several enzymes in the cortisol biosynthetic pathway. 21-Hydroxylase deficiency, the most common, can lead to virilization in female infants, with excessive adrenal androgen secretion caused by ACTH stimulation in the face of inability to synthesize cortisol. Inhibition of enzymatic steps in hormone synthesis, such as aromatase inhibitors to decrease estrogen formation in estrogen receptor–positive forms of breast cancer, may be an important therapeutic target.

Hormone hypersecretion syndromes in which excessive hormone secretion occurs in the face of "normal" levels of the factor that ordinarily suppresses the cognate hormone are by definition caused by some intrinsic defect in negative feedback suppression. This may occur secondary to a neoplastic proliferation of hormone-secreting cells, so that "basal" hormone secretion from the increased mass of cells exceeds physiologic levels. This may also be due to alterations in the intrinsic "set-point" for negative feedback suppression of hormone secretion. In practice, it may be impossible to differentiate these two mechanisms, and they are not mutually exclusive.

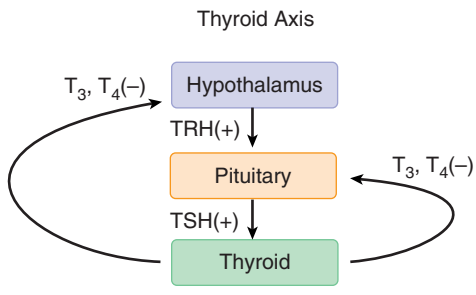

FIGURE 222-1 Hypothalamic-pituitary-thyroid axis illustrating negative feedback regulation. Following thyroid-stimulating hormone (TSH) secretion by the pituitary, the thyroid gland secretes T_3 and T_4, which feed back on the hypothalamus and pituitary to suppress further increases in thyrotropin-releasing hormone (TRH) and TSH secretion.

Hormone secretion is subject to many additional forms of regulation beyond simple negative feedback suppression. These include metabolic, neural, and other internal and environmental inputs. The temporal pattern of hormone secretion is often related to diurnal rhythms, as classically seen for cortisol, and also pulsatility. Changes in gonadotropin secretion during the menstrual cycle and during the course of puberty are striking examples of complex regulation of temporal patterns of hormone secretion.

HORMONE TRANSPORT AND METABOLISM

Most peptide hormones circulate as the free peptide, but insulin-like growth factor 1 (IGF-1) uniquely binds to a number of specialized binding proteins. Steroid and thyroid hormones are lipophilic molecules that circulate largely in protein-bound form. Specialized binding proteins for cortisol, androgens, estrogens, and thyroid hormones are selective for their cognate hormone. The free circulating hormone concentration is only a small fraction of the total hormone measured by routine analytic methods. Binding protein abnormalities can occur in liver disease, because most are synthesized by the liver. Thus, the determination of free as opposed to total plasma hormone concentration may be critical to accurate diagnosis in certain clinical conditions.

Hormone metabolism is another critical determinant of action for some hormones. For testosterone, thyroxine, and vitamin D, enzymatic conversion to more potent hormones—dihydrotestosterone formation in target tissues such as skin by 5-α reductase, thyroxine conversion to triiodothyronine by deiodinases, and 1,25 dihydroxyvitamin D formation by sequential hydroxylations in the liver and kidney—are all critical to normal hormone action. Defects in these metabolic steps leads to impaired hormone action. Even some peptide hormones such as angiotensin must undergo enzymatic conversion from secreted precursor form to generate the active hormone.

MECHANISM OF HORMONE ACTION

The central question in hormone action is how a hormone circulating in the blood stream in minute concentrations recognizes its specific target cells and regulates physiologic processes within them. Research addressing this question over the past four decades defined receptors, previously a purely theoretical concept, in molecular terms. Receptors are highly selective molecules that bind their cognate hormones with high affinity and specificity. Two broad classes of receptors were identified: (1) cell surface receptors that typically span the plasma membrane one or more times (Fig. 222-2) and (2) so-called nuclear receptors that reside either in the nucleus or in the cytoplasm, with subsequent translocation to the nucleus[2] (Fig. 222-3).

Hormones regulate cellular physiologic processes such as secretion of hormones, enzymes, and other compounds, muscle contraction, growth, and proliferation. *Signal transduction* is the general term for the biochemical steps between hormone binding to receptor and alterations in cell physiology. Most peptide and protein hormones (e.g., insulin, growth hormone, ACTH) bind to cell surface receptors that can be classified according to the

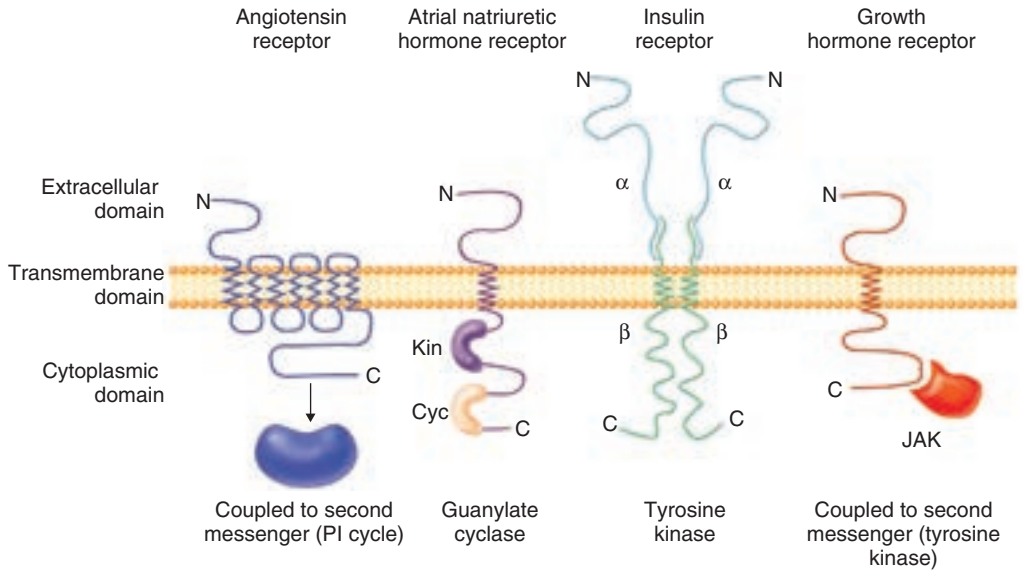

FIGURE 222-2. **Structures of different types of peptide hormone receptors.**

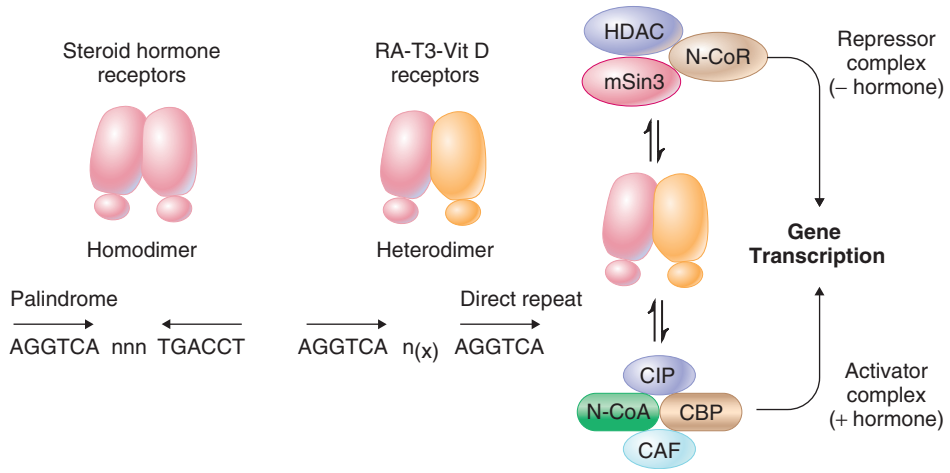

FIGURE 222-3. **How steroid hormone receptors function.** *Left,* Glucocorticoid receptor family members bind as homodimers to palindromic DNA sites. Thyroid hormone receptor family members bind primarily as heterodimers with retinoid X receptor to direct repeat DNA sites separated by varying numbers of base pairs. *Right,* As a result of hormone binding, repressor complexes dissociate and activator complexes bind to nuclear receptors. Repressor complexes contain histone deacetylase (HDAC), and activator complexes contain histone acetylase (CAF).

mechanism of signal transduction to which they are coupled (e.g., G protein–coupled[3] receptor tyrosine or serine kinase, JAK/STAT coupled; see Fig. 222-2). Cell surface receptors may generate "second messengers," which in turn regulate a kinase cascade. Receptor activation may have rapid effects such as exocytosis of secretory granules, but longer-term actions involving gene regulation may also be a consequence of second messenger and kinase cascade activation. Steroid and thyroid hormones and vitamin D bind to members of the nuclear receptor family. The latter act as "ligand-regulated" transcription factors to regulate gene expression (see Fig. 222-3).

Selectivity of hormone binding by receptors is not absolute. "Specificity spillover" is a clinically important phenomenon in which supraphysiologic concentrations of a hormone leads to binding and activation of a non-cognate receptor for a closely related hormone. Examples include: hypoglycemia seen with non–islet cell tumors secreting IGF-2, which binds to the insulin receptor[4]; skin hyperpigmentation in subjects with Addison's disease in whom excessive ACTH secretion activates the melanocortin receptor in melanocytes normally regulated by MSH; and hyperthyroidism in pregnant women with high human chorionic gonadotropin (HCG) concentrations activating the TSH receptor.

Genetic endocrine diseases include those caused by mutations in one component of a signal transduction pathway leading either to hormone "resistance" or hormone-independent activation (Tables 222-1 and 222-2). In the former, subjects present with apparent hormone deficiency, but direct hormone measurement reveals high concentrations of bioactive hormone that fails to act owing to target organ resistance. In the latter, patients present with apparent endocrine hyperfunction, but direct measurement reveals suppressed hormone concentration due to intact negative feedback. Loss-of-function mutations in receptors and in signaling intermediates such as G proteins have been identified in patients with hormone resistance. The converse, mutations that constitutively activate receptors or downstream signaling components,[5] have been identified in endocrine hyperactivity diseases such as familial male precocious puberty and familial nonautoimmune hyperthyroidism. Receptor inactivating mutations that lead to hormone-resistance diseases have been identified for both the cell surface and nuclear classes of receptors.

TABLE 222-1 DISEASES CAUSED BY G PROTEIN–COUPLED RECEPTOR LOSS-OF-FUNCTION MUTATIONS

RECEPTOR	DISEASE	INHERITANCE
V2 vasopressin	Nephrogenic diabetes insipidus	X-linked
ACTH	Familial ACTH resistance	Autosomal recessive
GHRH	Familial GH deficiency	Autosomal recessive
GnRH	Hypogonadotropic hypogonadism	Autosomal recessive
GPR54	Hypogonadotropic hypogonadism	Autosomal recessive
Prokineticin receptor 2	Hypogonadotropic hypogonadism	Autosomal dominant*
FSH	Hypergonadotropic ovarian dysgenesis	Autosomal recessive
LH	Male pseudohermaphroditism	Autosomal recessive
TSH	Familial hypothyroidism	Autosomal recessive
Ca^{2+} sensing	Familial hypocalciuric hypercalcemia, neonatal severe primary hyperparathyroidism	Autosomal dominant Autosomal recessive
Melanocortin 4	Obesity	Autosomal recessive
PTH/PTHrP	Blomstrand chondrodysplasia	Autosomal recessive

*With incomplete penetrance
ACTH = Adrenocorticotropic hormone; FSH = follicle-stimulating hormone; GH = growth hormone; GHRH = growth hormone–releasing hormone; GnRH = gonadotropin-releasing hormone; LH = luteinizing hormone; PTH = parathyroid hormone; PTHrP = parathyroid hormone–related protein; TSH = thyroid-stimulating hormone.

TABLE 222-2 DISEASES CAUSED BY G PROTEIN–COUPLED RECEPTOR GAIN-OF-FUNCTION MUTATIONS

RECEPTOR	DISEASE	INHERITANCE
LH	Familial male precocious puberty	Autosomal dominant
TSH	Sporadic hyperfunctional thyroid nodules	Noninherited (somatic)
TSH	Familial nonautoimmune hyperthyroidism	Autosomal dominant
Ca^{2+} sensing	Familial hypocalcemic hypercalciuria	Autosomal dominant
PTH/PTHrP	Jansen's metaphyseal chondrodysplasia	Autosomal dominant
V2 vasopressin	Nephrogenic inappropriate antidiuresis	Autosomal dominant

LH = Luteinizing hormone; PTH = parathyroid hormone; PTHrP = parathyroid hormone–related protein; TSH = thyroid-stimulating hormone.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Liu EH, Oberg K. The history and development of the gastroenteropancreatic endocrine axis. *Endocrinol Metab Clin North Am*. 2010;39:697-711.
2. Ahmadian M, Suh JM, Hah N, et al. PPARγ signaling and metabolism: the good, the bad and the future. *Nat Med*. 2013;19:557-566.
3. Spiegel AM. G proteins—the disease spectrum expands. *N Engl J Med*. 2013;368:2515-2516.
4. Dynkevich Y, Rother KI, Whitford I, et al. Tumors, IGF-2 and hypoglycemia: insights from the clinic, the laboratory and the historical archive. *Endocr Rev*. 2013;34:798-826.
5. Stratakis CA. E Pluribus Unum. The main protein kinase A catalytic subunit, a likely oncogene, and cortisol-producing tumors. *J Clin Endocrinol Metab*. 2014;99:3629-3633.

REVIEW QUESTIONS

1. Hormones are synthesized and secreted by:

A. Specialized collections of cells that form endocrine glands
B. Cells dispersed throughout the lining of the small intestine and within the pancreas
C. Neurons that comprise the posterior pituitary
D. The kidney, adipose tissue, and some other organs
E. All of the above

Answer: E Research has shown that not only classical endocrine glands such as the thyroid and adrenal, but also dispersed enteroendocrine cells in the gut lining and islets in the pancreas, neuroendocrine cells in the posterior pituitary, and non-classical hormone-secreting sites such as kidney, heart, and adipose are all sources of hormone secretion.

2. A genetic hormone deficiency disorder may be caused by:

A. A mutation in the gene encoding the hormone's signal sequence
B. Exposure of the pregnant mother to an endocrine "disruptor"
C. A mutation in a gene encoding an enzyme in the TCA cycle
D. Paternal advanced age at time of conception
E. All of the above

Answer: A A mutation in the hormone gene altering the signal sequence of a peptide hormone and thereby impairing its ability to be secreted may cause genetic forms of hormone deficiency such as hypoparathyroidism.

3. Negative feedback suppression is:

A. The basis for the diurnal rhythm in serum cortisol
B. The mechanism whereby an "output" regulated by a hormone feeds back to regulate the secretion of that hormone
C. The basis for menopause in women and lower testosterone with age in men
D. Always involves one hormone regulating the secretion of another hormone
E. None of the above

Answer: B B is the correct definition of negative feedback suppression. The latter is not the primary cause of either A or C. D is incorrect because outputs other than hormones can exert negative feedback suppression (e.g., Ca^{++} suppressing PTH secretion).

4. Hormone receptors are:

A. Always located on the cell surface
B. When activated, may exert their effects via a second messenger cascade
C. When activated, may exert their effects by directly regulating gene transcription
D. Irrespective of starting location, always translocate to the nucleus to exert their actions
E. Both B and C

Answer: E Because there are cell surface receptors that do not translocate to exert their action, and nuclear receptors for steroid hormones, A and D are incorrect. Both B and C are true for cell surface and nuclear receptors, respectively.

5. Genetic hormone resistance disorders:

A. Are due to gain-of-function mutations in hormone receptors
B. Are always due to loss-of-function mutations in nuclear receptors
C. Are due to overactivity of enzymes that degrade hormones
D. May be caused by loss-of-function mutations in genes encoding cell surface or nuclear receptors or other components of the signal transduction pathway
E. Are generally associated with autoimmune disease

Answer: D All the cited examples may be causes of inherited hormone resistance diseases, not just B. A, C, and E do not apply.

223

NEUROENDOCRINOLOGY AND THE NEUROENDOCRINE SYSTEM

MARK E. MOLITCH

NEUROENDOCRINE REGULATION

Neuroendocrinology refers to the area of endocrinology in which the nervous system interacts with the endocrine system to link neural activity with metabolic and hormonal homeostatic activity. The neurohypophysial neurons originate from the paraventricular and supraoptic nuclei, traverse the hypothalamic-pituitary stalk, and release vasopressin and oxytocin from nerve endings in the posterior pituitary. The hypophysiotropic neurons localized in specific hypothalamic nuclei project their axons to the median eminence to secrete their peptide and bioamine releasing and inhibiting hormones into the proximal end of the hypothalamic-pituitary portal vessels (Fig. 223-1). The median eminence receives its blood supply from the superior hypophysial artery, which arborizes into a rich capillary bed. The capillary loops extend into the median eminence and coalesce to form the long portal veins that traverse the pituitary stalk and end in the pituitary. The neuroendocrine system operates through a series of feedback loops that regulate pituitary and target organ hormone levels. Target organ hormones can feed back at both the hypothalamic and the pituitary levels to complete the loop. The feedback loops can be perturbed, resulting in alterations of set points by such factors as length of day (circadian periodicity), stress, nutritional status, and systemic illness.

Hypophysiotropic Hormones

Regulation of pituitary hormones by the hypophysiotropic hormones is quite complex, in part because of the multiplicity of substances present in the hypothalamus that can affect pituitary hormone secretion and in part because of the redundancy and overlapping nature of the feedback loops alluded to earlier. In addition, some hypophysiotropic hormones exert effects on more than one pituitary hormone.

Thyrotropin-Releasing Hormone

The primary neuroendocrine functions of thyrotropin-releasing hormone (TRH) are to stimulate the synthesis and release of thyroid-stimulating hormone (TSH) and prolactin. In cases of hypothyroidism, the increased TRH synthesis and binding to the pituitary results in increased TSH and prolactin levels. Correction of the hypothyroidism with thyroid hormones decreases the elevated TSH and prolactin levels. Conversely, in cases of hyperthyroidism, TSH levels are markedly suppressed. Although TRH is the major regulator of TSH synthesis and secretion, the role of TRH as a physiologic prolactin-releasing factor (PRF) is unclear.

Gonadotropin-Releasing Hormone

Gonadotropin-releasing hormone (GnRH) is a 10–amino acid peptide; its neurons originate in the epithelium of the medial part of the olfactory

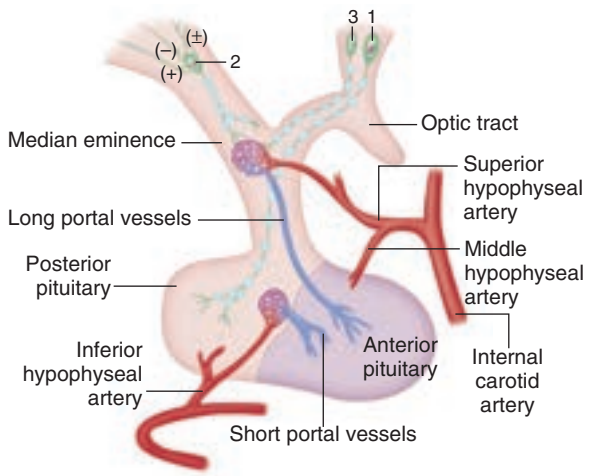

FIGURE 223-1. Neuroendocrine organization of the hypothalamus and pituitary gland. The posterior pituitary is fed by the inferior hypophyseal artery and the hypothalamus by the superior hypophyseal artery, both branches of the internal carotid artery. Most of the blood supply to the anterior pituitary is venous by way of the long portal vessels, which connect the portal capillary beds in the median eminence to the venous sinusoids in the anterior pituitary. Hypophysiotropic neuron 3 in the parvocellular division of the paraventricular nucleus and neuron 2 in the arcuate nucleus are shown to terminate in the median eminence on portal capillaries. These neurons of the tuberoinfundibular system secrete hypothalamic releasing and inhibiting hormones into the portal veins for conveyance to the anterior pituitary gland. Neuron 2 is innervated by monoaminergic neurons. Note that the multiple inputs to such neurons, using neuron 2 as an example, can be stimulatory, inhibitory, or neuromodulatory, in which another neuron may affect neurotransmitter release. Neuron 1 represents a peptidergic neuron originating in the magnocellular division of the paraventricular nucleus or supraoptic nucleus and projecting directly to the posterior pituitary by way of the hypothalamic-neurohypophyseal tract. (From Gay VL. The hypothalamus: physiology and clinical use of releasing factors. *Fertil Steril.* 1972;23:50-63, with permission of the American Society for Reproductive Medicine.)

placode. This origin of GnRH-producing neurons from olfactory epithelium is of clinical interest with respect to the entity of Kallmann's syndrome, in which GnRH deficiency is associated with agenesis of the olfactory bulbs. One form of Kallmann's syndrome is caused by loss of function of a protein (anosmin) that facilitates the embryologic migration of these GnRH-producing neurons. The primary function of GnRH is to stimulate the secretion of luteinizing hormone (LH) and follicle-stimulating hormone (FSH). Only one GnRH has been identified, and the differential secretion of LH and FSH is due to variations in sensitivity of the feedback effects of steroid and peptide hormones and variations in sensitivity to GnRH. GnRH pulsatile secretion directly increases its own receptor number, whereas continuous administration of GnRH is associated with a decrease. In women, positive and negative steroid hormone feedback regulation of the hypothalamic-pituitary-gonadal axis occurs at both the pituitary and hypothalamic levels, the hypothalamic effects being alteration of GnRH pulse amplitude and frequency, and the pituitary effects being modulation of the gonadotropin response to GnRH. In males, testosterone decreases GnRH pulsatile secretion, with a resultant decrease in gonadotropin pulse amplitude and frequency as well as a decreased gonadotropin response to exogenous GnRH.

The negative feedback effects of inhibin, a peptide produced by testicular Sertoli cells and ovarian granulosa cells, are predominantly on FSH at the pituitary, where inhibin causes a decrease in the sensitivity of gonadotrophs to GnRH. The related ovarian protein activin stimulates basal and GnRH-stimulated FSH synthesis and release from the pituitary, but its primary action is to facilitate the response of ovarian granulosa cells to FSH. Another gonadal peptide, follistatin, also inhibits the oophorectomy- and GnRH-induced rise in FSH selectively, primarily by binding to activin. These ovarian peptides are also found in the pituitary and may therefore have additional local effects on gonadotropin secretion.

GnRH has been administered with great success in pulsatile fashion to individuals with hypogonadotropic hypogonadism secondary to GnRH deficiency, leading to restoration of normal sexual function and fertility. Long-acting GnRH agonists have been used to downregulate GnRH receptors and gonadotropin secretion in a variety of conditions, including precocious puberty, prostate cancer, breast cancer, uterine fibroids, and endometriosis. Direct GnRH antagonists that compete for the GnRH receptor are used for similar conditions.

Somatostatin

Somatostatin (also known as somatotropin release inhibiting factor) inhibits GH secretion. The interaction of somatostatin and growth hormone–releasing hormone (GHRH) on GH secretion is complex. GH secretory episodes are associated with increased GHRH secretion, often accompanied by low somatostatin levels; the basal or trough GH levels are associated with low GHRH levels and more elevated somatostatin levels. Somatostatin also inhibits basal and stimulated TSH secretion. However, GH is about 10-fold more sensitive to inhibition by somatostatin than is TSH, suggesting that the physiologic role of somatostatin in inhibiting TSH secretion is limited. Somatostatin is also present in D cells of the pancreatic islets, the gut mucosa, and the myenteric neural plexus. Through paracrine and endocrine actions, it suppresses the secretion of insulin, glucagon, cholecystokinin, gastrin, secretin, vasoactive intestinal polypeptide (VIP), and other gastrointestinal hormones, as well as such functions as gastric acid secretion, gastric emptying, gallbladder contraction, and splanchnic blood flow. Analogues of somatostatin have been developed for the treatment of acromegaly, carcinoid tumors, VIP-secreting tumors, TSH-secreting pituitary tumors, and islet cell tumors.

Corticotropin-Releasing Hormone

Corticotropin-releasing hormone (CRH) releases adrenocorticotropic hormone (ACTH), β-endorphin, β-lipotropin, melanocyte-stimulating hormone (MSH), and other peptides generated from proopiomelanocortin (POMC) in equimolar amounts. CRH mediates 75% of the ACTH response to stress. The remaining 25% is due to vasopressin. CRH and vasopressin have synergistic effects on ACTH release. CRH and vasopressin are not always released coordinately, however, and stress selectively activates the vasopressin-containing subset of CRH neurons. Cortisol feeds back to decrease ACTH secretion at both the hypothalamic and the pituitary levels. ACTH and β-endorphin also feed back negatively to decrease CRH release by the hypothalamus. Central bioamines, opioids, and peptides influence CRH secretion. Monokines released by inflammatory tissue, such as interleukin (IL)-1, IL-6, and tumor necrosis factor (TNF)-α, stimulate the synthesis and release of CRH and vasopressin from the hypothalamus. The consequent increase in cortisol then reduces the intensity of the inflammatory response and release of these monokines, thus completing this inflammation-modulating feedback loop. CRH receptors are widely distributed in the brain, and increases in CRH are associated with activation of the sympathetic nervous system, suppression of the parasympathetic nervous system, stimulation of arousal, and increased learning performance. CRH may also be involved in the regulation of body weight; with overfeeding, increased leptin stimulates CRH, which causes decreased food intake and increased energy expenditure.

Biosynthetic human CRH has become available for clinical use. Its major utility is in the differential diagnosis of Cushing's disease versus ectopic ACTH syndrome, with the finding that patients with Cushing's disease respond with a greater than 35% increment, whereas those with ectopic ACTH secretion have a lesser response. If the results are equivocal, CRH testing during bilateral inferior petrosal sinus sampling for ACTH often provides additional discriminatory information.

Growth Hormone–Releasing Hormone

GHRH dose-dependently stimulates GH secretion. With repetitive administration, GHRH can cause the release of sufficient GH in children with GHRH deficiency to result in an increase in insulin-like growth factor (IGF)-I levels and an acceleration of growth. Both IGF-I and GH feed back negatively on GH secretion. This negative feedback results in both a decrease in GHRH and an increase in somatostatin. The feedback effect of IGF-I is clinically relevant, as documented by the high circulating GH levels that occur in IGF-I-deficient states such as renal insufficiency and cirrhosis. In children with mutations of the GH receptor resulting in their not being responsive to GH (GH insensitivity syndrome, also known as Laron-type dwarfism), IGF-I levels are very low and GH levels are correspondingly elevated.

Prolactin Inhibitory Factor

The inhibitory component of hypothalamic regulation of prolactin secretion predominates over the stimulatory component. Dopamine is the major physiologic prolactin inhibitory factor. It is likely that in most physiologic circumstances that cause a rise in prolactin (e.g., lactation), a simultaneous fall in dopamine occurs, along with a rise in a prolactin-releasing factor (PRF) such as VIP. Blockade of endogenous dopamine receptors by a variety of drugs, such as the antipsychotics, causes a rise in prolactin. Lesions that interrupt

A separate GH-stimulating system for GH secretion involves a distinct receptor, termed the *GH secretagogue* (GHS) receptor, that interacts with a 28–amino acid peptide called ghrelin that was initially isolated from the stomach. GHS receptor and ghrelin messenger RNA are both present in the pituitary and hypothalamus of humans. The interaction of ghrelin physiologically with GHRH and somatostatin is complex but appears to be of minor importance with regard to GH secretion. However, ghrelin may play a more important role in the regulation of appetite and food intake.

the basal hypothalamic neuronal pathways carrying dopamine to the median eminence or that interrupt portal blood flow, such as craniopharyngiomas or other large mass lesions, result in decreased dopamine reaching the pituitary and hyperprolactinemia.

Prolactin-Releasing Factor

A number of hypothalamic peptides other than TRH have also been shown to have PRF activity. VIP stimulates prolactin synthesis and release at concentrations found in hypothalamic-pituitary portal blood. Within the VIP precursor is another similarly sized peptide known as peptide histidine methionine, which also has PRF activity. The precise roles of VIP versus peptide histidine methionine are not clear. They appear to be of negligible physiologic importance in humans.

Endogenous Opioid Peptides

Most data now suggest a modest role for endogenous opioid peptides in neuroendocrine regulation. The endogenous opioid peptides have a common 5–amino acid sequence at their amino termini (Tyr-Gly-Gly-Phe-Met [or Leu]) that is important for their binding to endogenous opioid receptors and bioactivity. Three major opioid peptide receptors and three major groups of opioid peptides are recognized. The μ-receptor mediates most of the endocrine effects and analgesia. The primary peptide ligand for the μ-receptor is β-endorphin, which is derived from POMC. The δ-receptor mediates behavioral, analgesic, and some endocrine effects and has as its primary peptide ligands met- and leu-enkephalins, which are derived from proenkephalin A. It is much less well blocked by naloxone than is the μ-receptor. The κ-receptor mediates sedation and ataxia and binds primarily dynorphin and the neoendorphins. A fourth receptor has considerable sequence homology with the δ-receptor and binds to an endogenous 17–amino acid peptide called nociceptin (also known as orphanin FQ).

POMC is a 31-kD precursor peptide that harbors within it ACTH, β-lipotropin, and β-endorphin. POMC undergoes tissue-specific posttranslational processing. In the anterior pituitary, its major cleavage products are β-lipotropin and ACTH, with a significant proportion of β-lipotropin being further processed to β-endorphin. In the pituitary intermediate lobe, the major products are α-melanocyte stimulating hormone (α-MSH), corticotropin-like intermediate peptide, β-endorphin, and γ-lipotropin. Brain POMC, however, is processed primarily to β-endorphin, γ-lipotropin, and ACTH, with most of the ACTH being further processed to corticotropin-like intermediate peptide and α-MSH. The pentapeptide enkephalins are derived from the 28-kD precursor proenkephalin A. Neuronal perikarya containing the enkephalins are widely distributed throughout the brain. Dynorphin is a 17–amino acid peptide derived from a 28-kD precursor called proenkephalin B or prodynorphin. Shorter peptides called α- and β-neoendorphin, which have 10 and 9 amino acids, respectively, have also been isolated. These peptides react almost exclusively with the κ-receptor. Nociceptin is a 17–amino acid peptide derived from a κ precursor called pronociceptin. High concentrations of nociceptin and its receptor are present in the hypothalamus as well as in other areas of the brain that serve as the sources of monoamine neurotransmitters. In general, nociceptin appears to have an antiopioid or antinociceptive effect. The hypothalamus contains abundant opioid receptors. The effects of opioid peptides on anterior pituitary hormone secretion are produced through modulation of hypothalamic bioamines and hypophysiotropic factors.

The specific functions of the various opioid peptides and the opioid receptors are still not completely understood, although evidence links them to a number of body functions, including stress, mental illness, narcotic tolerance and dependence, eating, drinking, gastrointestinal function, learning, memory, reward, cardiovascular responses, respiration, thermoregulation, seizures, brain electrical activity, locomotor activity, pregnancy, and neuroimmune activity. Endogenous opioids have an inhibitory influence on gonadotropin secretion through action on GnRH secretion. Exogenous β-endorphin and enkephalin analogues increase serum GH and prolactin levels, but blockade of endogenous opioid pathways with naloxone does not alter basal or stimulated GH or prolactin levels. Opioids feed back negatively on ACTH and β-endorphin secretion, and naloxone can increase basal and stimulated ACTH levels. Overall, the effects of the endogenous opioids on normal physiologic regulation of the various pituitary hormones in humans is minimal. However, exogenous opioids in pharmacologic doses can impair GnRH and gonadotropin secretion, causing hypogonadism with reduced libido, sexual function, and fertility, and can also impair CRH and ACTH secretion, causing adrenal insufficiency.[1]

Central Nervous System Rhythms and Neuroendocrine Function

Pituitary hormones are secreted in a pulsatile fashion with a number of rhythms superimposed. The pulse amplitude of a pituitary hormone reflects the amount of releasing hormone and factors that alter sensitivity to that releasing hormone. Thus, the amplitude can be altered by the presence of inhibitory factors (e.g., GHRH versus somatostatin), nutritional factors, feedback effects of target organ hormones, and prior stimulation that depletes the releasable hormone pool. Pulse frequency is governed by the frequency of release of the hypophysiotropic factor, which is regulated by the hypothalamic pulse generator system.

The pituitary has an intrinsic rhythm of small amplitude with a frequency of every 2 to 10 minutes. Superimposed on this intrinsic rhythm is a rhythm caused by the pulsatile release of hypophysiotropic releasing factors, with or without the withdrawal of a corresponding inhibitory factor. Rhythms that are shorter than 1 day are referred to as ultradian rhythms. The next layer of rhythmicity is the circadian rhythm—that is, a rhythm with approximately 24-hour periodicity. These rhythms are usually synchronized with the 24-hour period by a periodic environmental cue such as the dark-light cycle. The suprachiasmatic nucleus functions as a circadian pacemaker and receives light-induced electrical impulses from the retina, transmitting those impulses to the pineal gland where they are converted to hormonal signals. Signals for a rhythm with a periodicity longer than 24 hours, an infradian rhythm, include the gravitational influence of the moon, which gives rise to the menstrual cycle.[2]

A number of factors may influence circadian and infradian rhythms. One of the most important is the sleep-wake cycle. GH, TSH, prolactin, ACTH, and pubertal LH secretion are all entrained more to the sleep-wake cycle than to the dark-light cycle (E-Fig. 223-1). Each has an increase and maximal level that occur following sleep onset. The profound diurnal variation in cortisol and ACTH is often used as an index of "normality" of the system. Loss of this diurnal rhythm occurs with disordered regulation by CRH, which may be due to endogenous depression or excessive alcohol intake, or autonomous secretion of ACTH in Cushing's disease (Chapter 224). Loss of the diurnal rhythm of cortisol has been used as a diagnostic test for Cushing's syndrome.

Interesting changes occur in gonadotropin secretion as a child passes through puberty into adulthood. Early in puberty, the amplitude of the pulses increases during sleep at night, especially for LH, but in adulthood, this nocturnal rise is lost. In patients with anorexia nervosa, the pattern of gonadotropin secretion often reverts to this pubertal pattern. This phenomenon suggests that body composition may in some way affect regulation of the pulsatile secretion of gonadotropins. The percentage of body fat has been proposed as being important in the timing of the onset of puberty. Recent studies implicate leptin as the signal indicating this change in body composition.

NEUROENDOCRINE DISEASE

Diseases of the Hypothalamus

Diseases may affect the hypothalamus by discrete localization in the hypothalamus, being part of a generalized central nervous system (CNS) disease such as neurosarcoidosis, or indirectly by a process such as hydrocephalus (Table 223-1). Furthermore, hormonal changes mediated by functional alterations in hypothalamic regulation may occur in systemic illnesses.

The axons projecting to the median eminence that contain the various hypophysiotropic factors are concentrated in the basal portion of the hypothalamus. Thus, lesions located within this final common pathway might be expected to cause significant decreases in secretion of some or all of the pituitary hormones except prolactin, which may increase because of the elimination of tonic inhibition by dopamine. Diabetes insipidus (Chapter 225) may also occur. Symptoms resulting from hypothalamic dysfunction are related to the size of the lesion and consequently to the area of the hypothalamus involved, as well as to the rapidity of the increase in lesion size. Slowly growing lesions tend to cause problems of hormone dysregulation rather than dramatic symptoms. Large, slowly growing lesions may cause more acute problems, such as when a slight increment in growth eliminates the remaining vestiges of vasopressin or ACTH secretion. The best way of discerning lesions affecting the hypothalamus is by magnetic resonance imaging (MRI) with gadolinium enhancement, although computed tomographic (CT) scanning with intravenous contrast is also effective. Formal visual field testing may discern impingement of the optic nerves and chiasm by hypothalamic lesions. Detailed testing of hypothalamic-pituitary function may reveal evidence of functional hypothalamic disruption with great sensitivity.

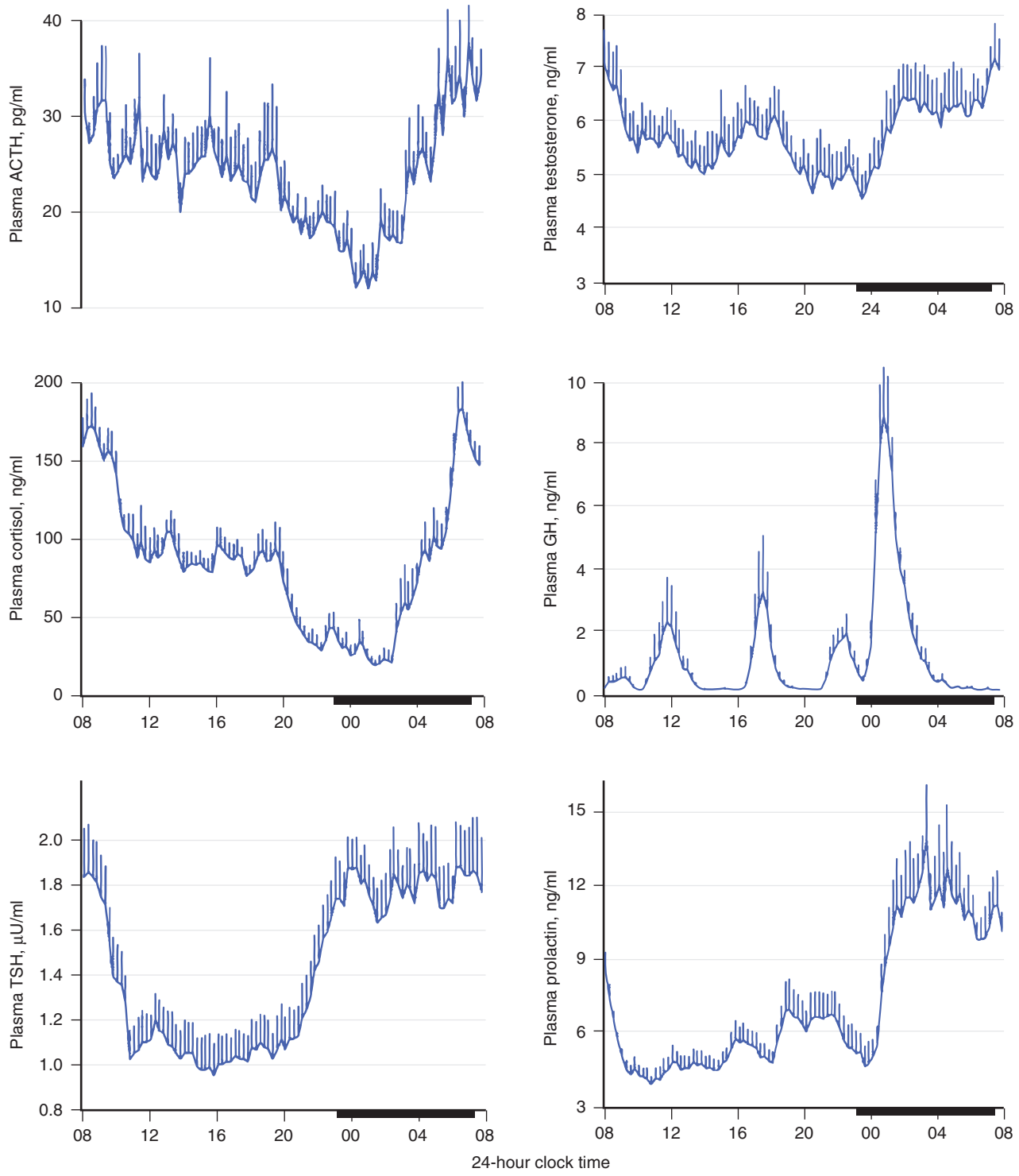

E-FIGURE 223-1. Hormonal circadian rhythms. Mean (+SEM) profiles of plasma levels of adrenocorticotropic hormone (ACTH), cortisol, thyroid-stimulating hormone (TSH), testosterone, growth hormone (GH), and prolactin from a group of 8 to 12 normal men studied at 15-minute intervals for 24 hours. The sleep times are indicated by black bars. (Redrawn with permission from Van Cauter E. Diurnal and ultradian rhythms in human endocrine function: a minireview. *Horm Res.* 1990;34:45-53.)

TABLE 223-1 ETIOLOGY OF HYPOTHALAMIC DISEASE

NEONATES
Congenital embryopathic disorders: agenesis of the corpus callosum, cleft palate *(HESX1)*
Congenital disorders: isolated hormone and receptor mutations, combined pituitary hormone deficiency *(PIT1, PROP1)*, Laurence-Moon-Bardet-Biedl syndrome, Prader-Labhart-Willi syndrome
Tumors: glioma, hemangioma
Trauma
Hydrocephalus, hydranencephaly, kernicterus
1 MONTH TO 2 YEARS
Tumors: glioma, especially optic glioma, hemangiomas
Infiltrative disease: Langerhans cell histiocytosis, meningitis
Hydrocephalus
2-10 YEARS
Tumors: craniopharyngioma, glioma, dysgerminoma, hamartoma, leukemia, ganglioneuroma, ependymoma, medulloblastoma
Infiltrative disease: Langerhans cell histiocytosis, meningitis, tuberculosis, encephalitis
Irradiation: for nasopharyngeal tumors, intracranial tumors, leukemia
Functional: psychosocial deprivation
10-25 YEARS
Congenital disorders: Kallmann's syndrome, gonadotropin-releasing hormone receptor defects
Tumors: craniopharyngioma, pituitary tumors, glioma, hamartoma, dysgerminoma, dermoid, lipoma, neuroblastoma
Trauma: subarachnoid hemorrhage, vascular aneurysm, arteriovenous malformation
Infiltrative diseases: Langerhans cell histiocytosis, sarcoidosis, tuberculosis, meningitis, encephalitis, leukemia
Chronic hydrocephalus or increased intracranial pressure
Functional: hypogonadotropic hypogonadism associated with weight loss, exercise
25-50 YEARS
Tumors: pituitary tumors, meningioma, craniopharyngioma, Rathke's cleft cyst, glioma, lymphoma, angioma, colloid cysts, ependymoma
Infiltrative diseases: sarcoidosis, Langerhans cell histiocytosis, tuberculosis, viral encephalitis
Subarachnoid hemorrhage, vascular aneurysms, arteriovenous malformation
Irradiation: for pituitary adenoma, nasopharyngeal tumors, intracranial tumors
Nutritional: Wernicke's disease
Functional: hypogonadotropic hypogonadism associated with weight loss, exercise
50 YEARS AND OLDER
Tumors: pituitary tumors, meningioma, craniopharyngioma, sarcoma, glioblastoma, lymphoma, colloid cysts, ependymoma
Vascular: infarct, subarachnoid hemorrhage, pituitary apoplexy, aneurysm
Irradiation: for pituitary adenoma, nasopharyngeal tumors, intracranial tumors
Infiltrative diseases: encephalitis, sarcoidosis, meningitis
Nutritional: Wernicke's disease

Modified from Plum F, Van Uitert R. Non-endocrine diseases of the hypothalamus. In: Reichlin S, Baldessarini RJ, Martin JB, eds. *The Hypothalamus.* New York: Raven Press; 1978:415.

CONGENITAL EMBRYOPATHIC DISORDERS

The most common embryopathic disorders to affect the hypothalamus are the midline cleft syndromes, which cause varying degrees of defects of midline structures, especially the optic and olfactory tracts, the septum pellucidum, the corpus callosum, the anterior commissure, the hypothalamus, and the pituitary. The clinical features of patients with midline cleft defects varies in severity from cyclopia to cleft lip and from isolated hypothalamic hormone defects to panhypopituitarism. The combination of absent septum pellucidum associated with optic nerve hypoplasia is referred to as septo-optic dysplasia and is associated with abnormalities of hypothalamic and other diencephalic structures. Some patients with septo-optic dysplasia and hypothalamic hypopituitarism have sexual precocity, presumably caused by a lack of inhibitory influences from other parts of the hypothalamus and intact GnRH-producing structures. Children with very mild midline cleft defects consisting of cleft lip, cleft palate, or both have an increased risk of GH and other pituitary hormone deficiencies. MRI studies of patients with "idiopathic" GH deficiency show absence of the infundibulum in nearly 50%.

Mutations responsible for these developmental defects are the subject of active investigation.[3] Mutations in the human *HESX1*, *SOX2*, *SOX3*, and *OTX2* transcription factor genes cause agenesis of the corpus callosum and panhypopituitarism. Case reports of mutations in other transcription factors, such as *Pitx2* (Rieger's syndrome) and *GLI2*, describe patients with varying brain and skull developmental abnormalities, along with varying degrees of hypopituitarism. Kallmann's syndrome is a condition characterized by anosmia or hyposmia and hypogonadotropic hypogonadism. The diagnosis is made by finding anosmia and low gonadotropin levels, and MRI will show absence or hypoplasia of the olfactory bulbs. The X-linked form of Kallmann's syndrome, representing about 85% of cases, is due to a gene defect *(KAL1)* resulting in loss of function of a protein called anosmin that facilitates the embryologic migration of GnRH-producing neurons from the olfactory placode to the hypothalamus and the olfactory nerves to the olfactory bulbs. Other genes implicated in Kallmann's syndrome include *PROK2*, *PROKR2*, *FGFR1*, and *FGF8*. The pituitary is usually intact in this condition, and treatment with pulsatile GnRH therapy or gonadotropins results in spermatogenesis and normal gonadal function. In some patients, other abnormalities may be present, including cerebellar ataxia, nerve deafness, color blindness, cleft lip and palate, mental retardation, disordered thirst, unilateral renal agenesis, synkinesia, dental agenesis, and skeletal anomalies such as syndactyly and polydactyly; these abnormalities tend to track with specific mutations of the genes outlined above.

TUMORS

The most common tumors affecting the hypothalamus are pituitary adenomas that have significant suprasellar extension. These tumors can cause varying degrees of hypopituitarism and hyperprolactinemia, either by compressing the normal pituitary or, more commonly, by affecting the pituitary stalk and mediobasal hypothalamus. Surprisingly, diabetes insipidus (Chapter 225) is a rare finding in patients with pituitary adenomas. In patients with normal or elevated prolactin levels, indicating a hypothalamic/stalk site of the lesion rather than pituitary destruction, pituitary function often returns following therapy.

Craniopharyngiomas are the next most common tumors affecting the hypothalamus.[4] Microscopically, craniopharyngiomas consist of cysts alternating with stratified squamous epithelium. The cyst fluid is usually thick and dark, and the material is often calcified. They arise from remnants of Rathke's pouch. A closely related lesion is Rathke's cleft cyst, which develops from the space between the anterior and rudimentary intermediate lobes. Rathke's cleft cysts are lined with cuboidal (as opposed to squamous) epithelium, and the cyst fluid is usually a white mucoid fluid. Craniopharyngiomas may be difficult to remove in their entirety, and postoperative radiation reduces recurrences. Rathke's cleft cysts less commonly recur. Craniopharyngiomas most commonly arise during childhood, but they may also occur in adults and even elderly people. These tumors come to attention because of mass effects, including headache, vomiting, visual disturbance, seizures, hypopituitarism, and polyuria. Some patients have galactorrhea, amenorrhea, and hyperprolactinemia, features suggestive of a prolactinoma. Careful endocrine testing reveals varying degrees of hypopituitarism in 50 to 75%, modest hyperprolactinemia in 25 to 50%, and often diabetes insipidus. Surgical extirpation of craniopharyngiomas commonly causes a worsening of pituitary function, often resulting in complete panhypopituitarism and diabetes insipidus because of stalk section, and may cause damage to the hypothalamic centers that regulate thirst, body temperature, and food intake, resulting in severe obesity. Recently, the technique of hypothalamic-sparing surgery for craniopharyngiomas has resulted in less obesity without increasing the recurrence rate.[5] Irradiation may also be helpful, especially in children.

Suprasellar dysgerminomas arise from primitive germ cells that have migrated to the CNS during fetal life and are structurally identical to germ cell tumors of the gonads. They most commonly occur in children, in whom they cause decreased growth because of hypopituitarism, as well as diabetes insipidus and visual problems. Hyperprolactinemia occurs in more than 50% of affected children, and 10% have precocious puberty from the production of human chorionic gonadotropin (HCG) by the tumor. The finding of an elevated HCG level in the spinal fluid may be diagnostic. As opposed to craniopharyngiomas, these tumors are very radiosensitive, and radiation therapy combined with chemotherapy is the preferred treatment.

A hypothalamic hamartoma is a nodule of growth of hypothalamic neurons, glia, and fiber bundles attached by a pedicle to the hypothalamus between the tuber cinereum and the mammillary bodies and extending into the basal cistern.[6] Asymptomatic hamartomas may be present in up to 20% of random autopsies; rarely, these lesions may enlarge and disrupt hypothalamic function because of compression of adjacent tissue. Less commonly, they may cause seizures, especially gelastic seizures. Some hamartomas can be associated with other congenital anomalies and mutations in the

transcription factor gene *GLI3*. A variant of hamartoma consisting of similar tissue present within the anterior pituitary but without a neural attachment to the hypothalamus is called a choristoma or gangliocytoma. These neuronal tumors are of particular endocrine interest because they can produce hypophysiotropic hormones. A number of cases associated with precocious puberty have been reported in which the hamartomas produced GnRH. Successful treatment has been reported with surgery and with the administration of a long-acting GnRH analogue that suppresses gonadotropin secretion but does not affect the tumor itself. If the hamartoma does not cause other problems from mass effects, medical therapy with the GnRH analogue may be the best choice because surgery can be noncurative or even fatal. Some gangliocytomas have been reported that produce GHRH and acromegaly or CRH and Cushing's syndrome.

Other tumors and space-occupying lesions occurring in the suprasellar area include arachnoid cysts, meningiomas, gliomas, astrocytomas, chordomas, infundibulomas, cholesteatomas, neurofibromas, lipomas, and metastatic cancer (particularly from the breast and lung). Any such lesion may manifest as varying degrees of hypopituitarism, diabetes insipidus, and hyperprolactinemia, and surgical therapy often worsens the hormonal deficit and may cause other hypothalamic damage.

INFLAMMATORY AND INFILTRATIVE DISORDERS

CNS involvement in cases of sarcoidosis (Chapter 95) occurs in 1 to 5% of patients as determined on clinical grounds, and in up to 16% of cases at autopsy. Isolated CNS sarcoidosis is uncommon. When sarcoidosis does involve the CNS, the hypothalamus is involved in 10 to 20% of cases. Sarcoid granulomas can involve the hypothalamus, stalk, or pituitary and may be infiltrative or occur as a mass lesion.[7] The most common endocrine findings are varying degrees of hypopituitarism, diabetes insipidus, and hyperprolactinemia. Obesity secondary to hypothalamic involvement by sarcoidosis has also been reported. In patients with isolated CNS sarcoidosis, the diagnosis may be extremely difficult. Examination of cerebrospinal fluid usually shows elevated protein levels, low glucose levels, pleocytosis, and variable elevations of angiotensin-converting enzyme. However, biopsy is often necessary. Although corticosteroid therapy has been reported to at least partially reverse the thirst disorders, anterior pituitary hormone deficits usually do not respond.

Langerhans cell histiocytosis infiltration of the hypothalamus may cause diabetes insipidus, varying degrees of hypopituitarism, and hyperprolactinemia.[8] It is the most common cause of diabetes insipidus in children. Usually, this infiltration appears as a thickening of the pituitary stalk, but it may also appear as a mass lesion of the hypothalamus or pituitary (Fig. 223-2). Osteolytic lesions may be present in the jaw or mastoid, so radiographs of the jaw are a worthwhile part of the diagnostic evaluation of an unknown suprasellar mass or diabetes insipidus. Therapy consists of local surgery, focal irradiation, or chemotherapy with alkylating agents and high-dose corticosteroids.

Other infiltrative diseases (e.g., tuberculosis, lymphomas, fungal diseases) can also cause a progressive alteration in hypothalamic regulation of pituitary hormone secretion (E-Fig. 223-2).

VASCULAR DISEASE

An enlarging aneurysm may manifest as a mass lesion of the hypothalamic-pituitary area and may cause hypopituitarism and visual field defects. Obviously, the distinction must be made before surgery. Tumors and aneurysms may also coexist, and careful radiologic evaluation with MRI is necessary to discern such association. Hypothalamic disease caused by vascular infarction is extremely rare. In the past several years, it has been found that subarachnoid hemorrhage may be associated with varying degrees of hypopituitarism in almost half of cases. On the other hand, diabetes insipidus is uncommon.

TRAUMA

Traumatic brain injury (TBI) (Chapter 399) can cause defects ranging from isolated ACTH deficiency to panhypopituitarism with diabetes insipidus. Within the first 72 hours of trauma, GH, LH, ACTH, TSH, and prolactin levels may actually be elevated in blood, perhaps because of acute release. These levels subsequently fall, and either pituitary function returns to normal or hypopituitarism develops. Overall, the frequency of hypopituitarism in TBI is less than in subarachnoid hemorrhage, occurring in about one fourth of surviving patients. In patients dying of head injury, damage to the hypothalamus, pituitary stalk, or anterior pituitary has been found in up to 86% of cases.[9] The paraventricular and supraoptic nuclei and median eminence are particularly involved with microhemorrhages, hence the high frequency of panhypopituitarism with diabetes insipidus. With frontal injuries, the brain travels backward, but the pituitary cannot move; consequently, the pituitary stalk becomes avulsed, with interruption of the portal vessels. Most patients with head injury are hyperprolactinemic, which clinically confirms that the hypothalamus or stalk is the primary site of injury. In the past, pituitary function in patients with TBI and subarachnoid hemorrhage has not been assessed, and the role of untreated hypopituitarism in the long-term disability of such patients is unknown. Acute deficiency of ACTH/cortisol may be life-threatening. Therefore, cortisol levels should be monitored carefully in the hours and days after such events, and hypocortisolism treated with stress doses of glucocorticoids when necessary.

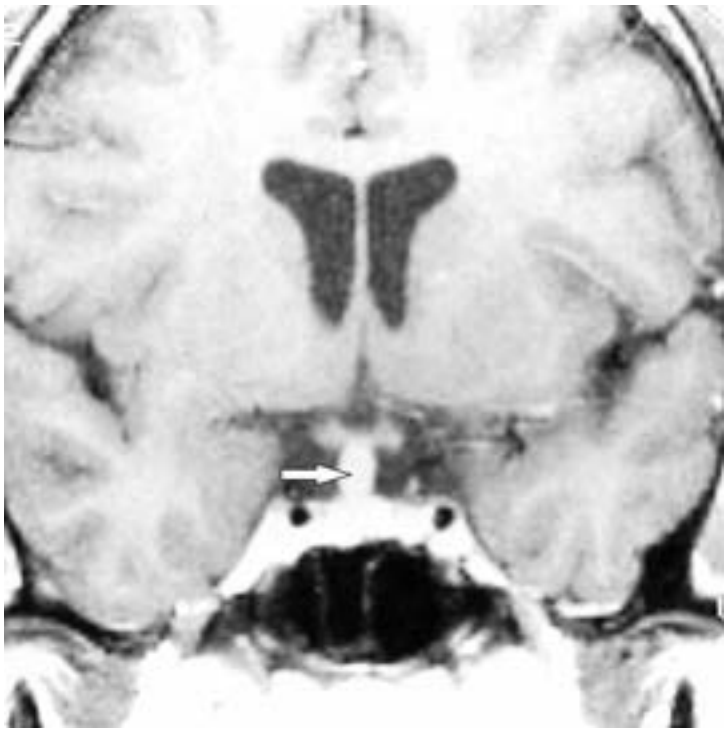

FIGURE 223-2. Thickened pituitary stalk in Langerhans cell histiocytosis. Magnetic resonance image of patient with Langerhans cell histiocytosis who manifested initially with amenorrhea, galactorrhea, and diabetes insipidus. Arrow points to the thickened pituitary stalk. (Reproduced with permission from Purdy LP, Molitch ME. Sudden onset of diabetes insipidus in an adolescent. *Endocr Trends.* 1998;5:1-7.)

IRRADIATION

Whole-brain irradiation for intracranial neoplasms frequently results in hypothalamic dysfunction, as evidenced by endocrine abnormalities and behavioral changes. The most common endocrine abnormality is hyperprolactinemia, but hypopituitarism also occurs. When the radiotherapy is targeted to the hypothalamic area, hypopituitarism occurs even more frequently.[10] The frequency of loss of pituitary function is so high that all patients who have had their pituitary and hypothalamic areas irradiated must be monitored periodically for the purpose of detecting these deficits. However, the development of such deficiencies may take many years, so that yearly testing is warranted for up to 20 years. It appears that stereotactic irradiation using the gamma knife apparatus or a linear accelerator for pituitary and other parasellar tumors causes a risk of hypopituitarism similar to that of conventional irradiation.

Effects of Hypothalamic Disease on Pituitary Function

Hypothalamic disease can cause both pituitary hyperfunction and hypofunction in varying degrees of severity. Although severe disease can cause absolute deficiencies of the various hormones, milder disease may cause a subtle alteration in feedback loops and timing such that, for example, the integration of signals necessary for menstrual cycling is lost, with subsequent "hypothalamic" amenorrhea (Chapter 236). Furthermore, the hypothalamic defects may be interrelated. The rather common finding of hyperprolactinemia occurring with hypothalamic dysfunction causes a hypogonadotropic hypogonadism that is reversible when the elevated prolactin levels are brought down to normal. In many cases, no structural lesion can be found on MRI,

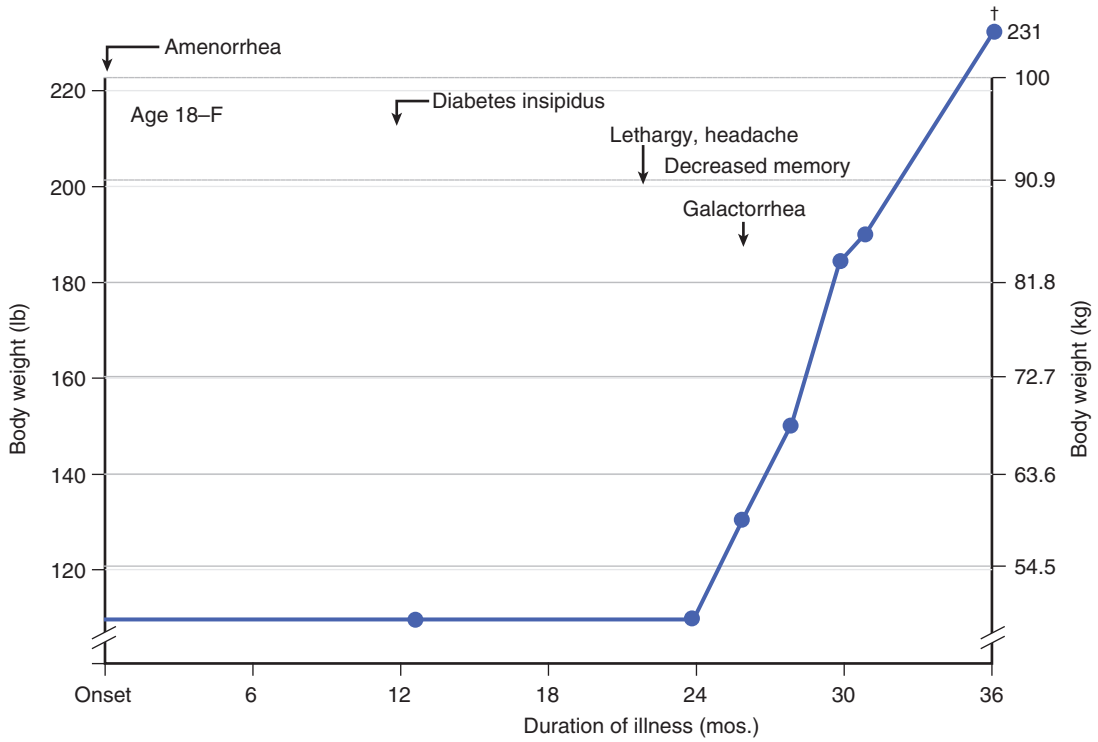

E-FIGURE 223-2. Progressive hypothalamic dysfunction from infiltrative disease of the hypothalamus. Clinical sequence in an 18-year-old woman with hypothalamic tuberculosis that progressed despite antituberculosis therapy. Her initial endocrine symptom was amenorrhea, which was followed over several months by the development of diabetes insipidus, lethargy, headache, decreased memory, galactorrhea, and culminating in a massive weight gain over the year prior to her demise. (Redrawn with permission from Bray GA, Gallagher TF Jr. Manifestations of hypothalamic obesity in man: a comprehensive investigation of eight patients and a review of the literature. *Medicine.* 1975;54:301-330.)

and a functional defect caused by altered neurotransmitter regulation is invoked.

GROWTH HORMONE

Loss of normal GH secretion is the most common hormonal defect occurring with structural hypothalamic disease. About three fourths of cases with congenital idiopathic GH deficiency have a normal GH response to exogenous GHRH, which implies that the defect is probably disordered hypothalamic regulation. Defects in the gene for GHRH have not been found, but a rare form of GH deficiency has been found to be caused by a mutation in the GHRH receptor. A reversible form of idiopathic GH deficiency caused by inadequate parental care and affection is referred to as the emotional deprivation syndrome or psychosocial dwarfism. Restoration of a proper social environment for such a child results in prompt normalization of GH secretion and growth. It has been hypothesized that the disordered GH regulation is due to psychogenic alteration of the neurotransmitter balance necessary for normal GHRH and somatostatin secretion. Other systemic illnesses such as inflammatory bowel disease may also cause decreased GH secretion and growth; treatment of the systemic illness will correct the growth abnormality. Treatment of children and adults with GH deficiency is discussed in Chapter 224.

GONADOTROPINS

Hypothalamic Hypogonadism

The primary defect in this group of disorders is thought to involve the secretion of GnRH, with resultant impairment in pituitary gonadotropin secretion and gonadal function. The disorders causing these conditions may be primary (i.e., congenital) defects or acquired. Depending on the time of onset, they are manifested as either delayed puberty, interruption of pubertal progression, or loss of adult gonadal function. The lesions causing these disorders may cause loss of other hormones or may be isolated to GnRH. Loss of gonadotropin secretion as the result of hypothalamic structural damage is the second most common defect after GH deficiency. However, a substantial portion of these defects are due to hyperprolactinemia and are reversible with correction of the hyperprolactinemia. In some cases, the defect is idiopathic. Defects in the gene for GnRH have not been found, but mutations in the GnRH receptor do occur.

In children, if the disorder is limited to GnRH and the gonadotropins, prior growth and development are normal, but the growth spurt occurring at puberty is lost. Undescended testes are present in 50% of patients with GnRH deficiency, probably secondary to the absence of gonadotropins during fetal development. The most common congenital lesion causing prepubertal GnRH deficiency is Kallmann's syndrome, which affects 50% of males and 37% of females seen with isolated gonadotropin deficiency (see earlier). In patients with GnRH deficiency, replacement of GnRH through subcutaneous administration every 2 hours with a portable pump causes a rapid rise in LH and FSH, a rise in testosterone to normal, and the development of normal spermatogenesis. Similar approaches in women result in ovulatory cycles in 80%. The success of such therapy confirms the original hypothesis of a primary defect of GnRH secretion. In men, comparable results can be obtained with exogenous gonadotropins given three times per week and is much more practical. GnRH therapy is not successful in those with GnRH receptor mutations. Replacement with testosterone alone causes adequate androgenization but does not result in an increase in testicular size or in spermatogenesis.

Loss of formerly normal GnRH secretion in adults may be due to structural hypothalamic damage such as a tumor, a functional change unassociated with a detectable lesion, or hyperprolactinemia.[11] Structural disease must be excluded in such patients by CT or MRI. Most cases of functional hypogonadotropic hypogonadism occur in women, the most common causes being weight loss, excessive exercise, psychogenic stress, or systemic illness, but idiopathic forms occur as well. In some patients, exercise results in a loss of body fat not detected with total body weight measures, and it is unclear whether the hypogonadism is directly due to the loss of body fat or to the exercise per se. Studies of pulsatile gonadotropin secretion in such patients reveal absent pulses. Usually, the gonadotropin response to injected GnRH is normal. Regain of weight and stopping of the exercise result in resumption of normal gonadal function. Furthermore, the administration of leptin to such women results in a resumption of normal gonadotropin pulsatile secretion and ovulation, confirming the key role leptin has in mediating the influence of body energy stores on reproductive function.[12] However, in the idiopathic form, spontaneous resolution does not occur. Hyperprolactinemia occurring postpubertally can also decrease GnRH and the pulsatile secretion of LH and FSH and thereby result in anovulation with oligomenorrhea and amenorrhea in women and impotence and infertility in men.

Two goals in the treatment of idiopathic functional hypogonadotropic amenorrhea are (1) restoration of a normal estrogen status to promote well-being and prevent osteoporosis and (2) facilitation of ovulation for fertility. The former can generally be achieved with cyclic estrogen and progesterone, whereas the latter may require clomiphene or GnRH or gonadotropin therapy. In men, similar goals may be achieved with testosterone or GnRH or gonadotropins.

Hypothalamic Hypergonadism (Precocious Puberty)

Precocious puberty is defined as the onset of puberty before the age of 8 years in girls or 9 years in boys. Pseudoprecocious puberty is that resulting from peripheral (gonadal or adrenal) causes. Central, "true," or GnRH-dependent precocious puberty is characterized by hormonal changes similar to those that occur at the time of normal puberty—that is, an increase in the pulsatile release of LH, an increase in the gonadotropin response to GnRH, and an increase in gonadal steroid secretion.[13] GnRH-dependent precocious puberty therefore represents premature activation of the GnRH pulse generator by a variety of lesions, or it may also be idiopathic. Only about 10% of cases of central precocious puberty occur in boys, but they tend to have more serious underlying disease. In boys with central GnRH-dependent precocious puberty, hypothalamic hamartomas account for 38% of cases, other CNS lesions represent 31%, familial disease accounts for 23%, and idiopathic disease accounts for only 8%. The picture is quite different in girls, however: hypothalamic hamartomas account for only 15% of cases, other CNS lesions represent 14%, the McCune-Albright syndrome (polyostotic fibrous dysplasia [Chapter 231]) accounts for 6%, and fully 65% are idiopathic. Dysgerminomas in the suprasellar or pineal region can produce HCG, which acts like LH in its stimulation of gonadal function. Usually, such tumors cause increased sex steroid formation but fail to cause ovulation.

Therapy for central GnRH-dependent precocious puberty consists of surgical removal of the tumor or medical therapy with a long-acting GnRH analogue, either in the form of monthly injections or yearly implants.[14] The latter can suppress gonadotropin and sex steroid hormone levels and cause a stabilization or even regression of secondary sex characteristics and a slowing of growth and bone maturation in most cases. When therapy is discontinued at the normal time of puberty, sex steroid levels increase, secondary sexual characteristics again develop, growth increases, and regular menses develop spontaneously.

PROLACTIN

Hypothalamic Hyperprolactinemia

Structural or infiltrative lesions of the hypothalamus, such as those discussed earlier, can decrease the amount of dopamine reaching the lactotrophs and thus cause modest hyperprolactinemia (usually < 100 ng/mL).[15] Because their therapy is quite different, it is very important to differentiate nonsecreting pituitary adenomas with extensive suprasellar extension causing prolactin elevations in this range from prolactin-secreting adenomas, which usually cause prolactin elevations 5 to 50 times higher. A peculiarity of some two-site immunoassays, referred to as the "hook effect," can sometimes cause a very high prolactin level to read falsely normal or just mildly elevated; a 1 : 100 dilution of the serum sample with saline will show the true level when the specimen is rerun. If there is any question about an assay being susceptible to the hook affect, in patients with very large tumors, prolactin should be measured undiluted and at 1 : 100 dilution to avoid this important spurious finding. A number of medications, antipsychotics in particular, can cause hyperprolactinemia, primarily by interfering with central catecholamines (Table 223-2).

Therapy is generally directed at the underlying cause. The hyperprolactinemia itself may impair gonadal function, so efforts may also be made to lower prolactin levels with dopamine agonists. Prolactin levels usually fall quite readily in such patients. Restoration of gonadal function is not automatic, however, because the primary hypothalamic lesion may also directly impair release of GnRH. In that circumstance, both dopamine agonists and sex steroid replacement may be necessary. When administration of psychotropic medications that cause the hyperprolactinemia cannot be stopped, dopamine agonists may be used but very rarely have been reported to exacerbate the psychosis. In such cases and in others in which fertility is not an issue, treatment with cyclic estrogen and progesterone replacement can be carried out safely.

TABLE 223-2 ETIOLOGIES OF HYPERPROLACTINEMIA

PITUITARY DISEASE
Prolactinomas
Acromegaly
Empty sella syndrome
Lymphocytic hypophysitis
Cushing's disease
Pituitary stalk section
HYPOTHALAMIC DISEASE
Craniopharyngiomas
Meningiomas
Dysgerminomas
Nonsecreting pituitary adenomas
Other tumors
Sarcoidosis
Langerhans cell histiocytosis
Neuraxis irradiation
Vascular
NEUROGENIC
Chest wall lesions
Spinal cord lesions
Breast stimulation
OTHER
Pregnancy
Hypothyroidism
Chronic renal failure
Cirrhosis
Pseudocyesis
Adrenal insufficiency
Idiopathic
MEDICATIONS
Antipsychotics
Atypical antipsychotics
Monoamine oxidase inhibitors
Tricyclic antidepressants
Reserpine
Methyldopa
Metoclopramide
Domperidone
Cocaine
Verapamil
Serotonin reuptake inhibitors

Modified from Molitch ME. Medication-induced hyperprolactinemia. *Mayo Clin Proc.* 2005;80:1050-1057.

Idiopathic Hyperprolactinemia

Idiopathic hyperprolactinemia is a diagnosis of exclusion. Prolactin levels in this condition are usually less than 100 ng/mL. In such cases, small pituitary or hypothalamic tumors could exist that are beyond the resolution of current imaging techniques, but when such patients are monitored for many years, it is very uncommon for tumors to later be visualized. Idiopathic hyperprolactinemia can cause amenorrhea, galactorrhea, impotence, infertility, and loss of libido, just as occurs with hyperprolactinemia of other causes, so the idiopathic hyperprolactinemia may need to be treated. Premature osteoporosis related to the estrogen deficiency may also occur. The only possible treatment is dopamine agonists, and these agents are successful in more than 90% of cases. Alternatively, cyclic estrogen and progesterone replacement may be given, but fertility will not be restored.

THYROID-STIMULATING HORMONE

Hypothalamic hypothyroidism is due to a central lesion that impairs the secretion of TRH, usually along with the loss of other hormones.[16] It occurs considerably less commonly than hypothalamic GH and gonadotropin deficiency. Defects in the gene for TRH have not been detected, but a case has been reported of a TRH receptor mutation causing hypothyroidism. TSH levels in this syndrome are generally normal or even slightly elevated. TSH in these patients is biologically less active than normal and binds to the TSH receptor less well because of altered glycosylation as a result of the TRH deficiency. Treatment is with L-thyroxine, and monitoring of therapy is done solely by measurement of free thyroxine (T_4) levels and not TSH levels.

ADRENOCORTICOTROPIC HORMONE

ACTH deficiency caused by hypothalamic lesions is uncommon.[17] It may occur with the loss of other hormones but may also appear as an isolated deficiency. The most common cause, of course, is prior suppression by exogenous or endogenous glucocorticoids. In the absence of CNS lesions or a history of trauma, many cases of isolated ACTH deficiency appear to be due to a pituitary autoimmune disorder. Treatment is with glucocorticoids; mineralocorticoids are not needed.

Effects of Hypothalamic Disease on Other Neurometabolic Functions

A number of functions that affect the internal milieu, in addition to anterior and posterior pituitary function, are regulated at least in part by the hypothalamus and include temperature control, behavior, consciousness, memory, sleep, food intake, and carbohydrate metabolism.

ALTERATIONS IN FOOD INTAKE

Hypothalamic Obesity

Destruction of the mediobasal hypothalamus will sometimes inhibit satiety and may result in hyperphagia and hypothalamic obesity.[18] The hyperphagia is due to destruction of noradrenergic fibers originating in the paraventricular nucleus and passing through the mediobasal hypothalamus. Because of their location, such lesions also usually produce hypopituitarism and diabetes insipidus. In a number of rare syndromes (Prader-Willi, Laurence-Moon-Biedl-Bardet) with obesity as a major characteristic, hypothalamic causes have been postulated but not proved.

Hypothalamic Anorexia

Lesions of the lateral hypothalamus, which destroy nigrostriatal dopaminergic fibers that pass through this area, produce hypophagia along with an increase in peripheral norepinephrine turnover and metabolic rate. This syndrome is very rare, probably owing to the requirement for bilateral lesions. The hormonal changes that occur in anorexia nervosa appear to all be secondary to the weight loss, and no evidence has been found for a primary hypothalamic disorder in this syndrome.

HYPERGLYCEMIA

Hypothalamic activation as part of the generalized response to stress can cause release of GH, prolactin, and ACTH, which serve as counter-regulatory hormones with respect to insulin. These hormones promote lipolysis, gluconeogenesis, and insulin resistance, resulting in glucose elevation. Of more importance in the acute response to stress, this hypothalamic response results in sympathetic activation with release of catecholamines that inhibit insulin secretion and stimulate glycogenolysis.

TEMPERATURE REGULATION

The anterior hypothalamus and preoptic area contain temperature-sensitive neurons that respond to internal temperature changes by initiating certain thermoregulatory responses necessary to restore a constant temperature. Measures that dissipate heat include cutaneous vasodilation, sweating, and panting, and measures that increase body heat include increasing metabolic heat production, shivering, and cutaneous vasoconstriction.

Rare patients have been reported with anterior hypothalamic lesions that cause paroxysmal or sustained hypothermia or hyperthermia from failure of these thermoregulatory activities. Some cases of paroxysmal hypothermia and hyperthermia respond to anticonvulsant medications, which suggests that the neuronal discharge effecting the temperature changes is seizure-like.

Poikilothermy results from an inability to dissipate or generate heat to keep the body temperature constant in the face of varying ambient temperatures. This condition results from bilateral lesions in the posterior hypothalamus and rostral mesencephalon, which are the areas responsible for the final integration of thermoregulatory neural efferents. Patients with this condition do not feel discomfort with temperature changes and are unaware of having a problem. Depending on the ambient temperature, they may experience life-threatening hypothermia or hyperthermia.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Brennan MJ. The effect of opioid therapy on endocrine function. *Am J Med*. 2013;126:S12-S18.
2. Morris CJ, Aeschbach D, Scheer FA. Circadian system, sleep and endocrinology. *Mol Cell Endocrinol*. 2012;349:91-104.
3. McCabe MJ, Dattani MT. Genetic aspects of hypothalamic and pituitary gland development. *Handb Clin Neurol*. 2014;124:3-15.
4. Larkin SJ, Ansorge O. Pathology and pathogenesis of craniopharyngiomas. *Pituitary*. 2013;16:9-17.
5. Elowe-Gruau E, Bertrand J, Brauner R, et al. Childhood craniopharyngiomas: hypothalamus-sparing surgery decreases the risk of obesity. *J Clin Endocrinol Metab*. 2013;98:2376-2382.
6. Mittal S, Mittal M, Montes JL, et al. Hypothalamic hamartomas. Part I. Clinical, neuroimaging, and neurophysiological characteristics. *Neurosurg Focus*. 2013;34:E1-E6.
7. Langrand C, Bihan H, Raverot G, et al. Hypothalamo-pituitary sarcoidosis: a multicenter study of 24 patients. *QJM*. 2012;105:981-995.
8. Kurtulmus N, Mert M, Tanakol R, et al. The pituitary gland in patients with Langerhans cell histiocytosis: a clinical and radiological evaluation. *Endocrine*. 2014;[Epub ahead of print].
9. Hannon MJ, Sherlock M, Thompson CJ. Pituitary dysfunction following traumatic brain injury or subarachnoid hemorrhage. *Best Pract Res Clin Endocrinol Metab*. 2011;25:783-798.
10. Sathyapalan T, Dixit S. Radiotherapy-induced hypopituitarism: a review. *Expert Rev Anticancer Ther*. 2012;12:669-633.
11. Silveira LFG, Latronico AC. Approach to the patient with hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*. 2013;98:1781-1788.
12. Michalakis K, Mintziori G, Kaprara A, et al. The complex interaction between obesity, metabolic syndrome and reproductive axis: a narrative review. *Metabolism*. 2013;62:457-478.
13. Fuqua JS. Treatment and outcomes of precocious puberty. An update. *J Clin Endocrinol Metab*. 2013;98:2198-2207.
14. Li P, Li Y, Yang CL. Gonadotropin releasing hormone agonist treatment to increase final stature in children with precocious puberty: a meta-analysis. *Medicine (Baltimore)*. 2014;93:e260.
15. Majumdar A, Mangal NS. Hyperprolactinemia. *J Hum Reprod Sci*. 2013;6:168-175.
16. Persani L. Clinical review: central hypothyroidism: pathogenic, diagnostic, and therapeutic challenges. *J Clin Endocrinol Metab*. 2012;97:3068-3078.
17. Crowley RK, Argese N, Tomlinson JW, et al. Central hypoadrenalism. *J Clin Endocrinol Metab*. 2014;99:4027-4036.
18. Bereket A, Kiess W, Lustig RH, et al. Hypothalamic obesity in children. *Obes Rev*. 2012;13:780-798.

224

ANTERIOR PITUITARY

MARK E. MOLITCH

ANATOMY AND EMBRYOLOGY

The pituitary is divided into anterior (adenohypophysis) and posterior (neurohypophysis) lobes. The optic chiasm, formed by the decussation of the optic nerves, is positioned directly above the pituitary gland. Specialized vascular structures located in the median eminence of the hypothalamus drain into portal veins that course down the pituitary stalk to join the sinusoidal capillaries of the anterior lobe. Hypothalamic hormones enter these capillaries and flow to the anterior pituitary (Fig. 224-1). Venous drainage from the anterior lobe is into the cavernous sinuses, which drain into the petrosal sinuses. The six major pituitary cell types include somatotrophs (growth hormone [GH] producing), lactotrophs (prolactin [PRL] producing), corticotrophs (adrenocorticotropic hormone [ACTH] producing), thyrotrophs (thyroid-stimulating hormone [TSH] producing), and gonadotrophs (follicle-stimulating hormone [FSH] and luteinizing hormone [LH] producing). Somatotrophs constitute 40 to 50% of anterior pituitary cells; lactotrophs, 15 to 25%; corticotrophs, 10 to 20%; and gonadotrophs 10%. Only 5% of pituitary cells are thyrotrophs.

The pituitary is formed from the fusion of Rathke's pouch (which gives rise to the anterior pituitary) and a portion of the ventral diencephalon (which gives rise to the posterior pituitary). Several transcription factors are important in the development of the various types of pituitary cells. LHX3 and LHX4 are present in somatotrophs, lactotrophs, gonadotrophs and thyrotrophs, and mutations in these genes result in deficits of GH, PRL, TSH, and the gonadotropins, although the deficits with LHX4 mutations are more variable.[1] Mutations in *HESX1* usually are associated with dysplasia of the septum pellucidum and optic tracts, in addition to multiple pituitary hormone deficits. The transcription factor Pit-1 is produced in somatotrophs, lactotrophs, and thyrotrophs. Mutations in the *PIT1 (POU1F1)* gene prevent the development of these cells and cause deficiencies of GH, PRL, and TSH. This lineage relationship explains why some GH-producing tumors also secrete PRL, and some TSH-producing tumors co-secrete GH.[2] PROP-1, another transcription factor, is critical for the development of somatotrophs, lactotrophs, and thyrotrophs. Mutations in the *PROP1* gene result in deficiencies of GH, PRL, and TSH, and in some affected individuals there is delayed puberty. Combined pituitary hormone deficiency (GH, PRL, TSH) has an incidence of about 1 in 8000 births, and 10% have an affected relative. Between 25 and 50% of these cases are due to *PIT1* or *PROP1* mutations. Tpit is specific to corticotroph cells, and TPIT gene mutations cause isolated ACTH deficiency. Rare mutations in several other genes (*SOX 3*, *OTX2*, *FGF8*, and *GLI2*) have also been shown to cause hypopituitarism, and the list continues to grow.

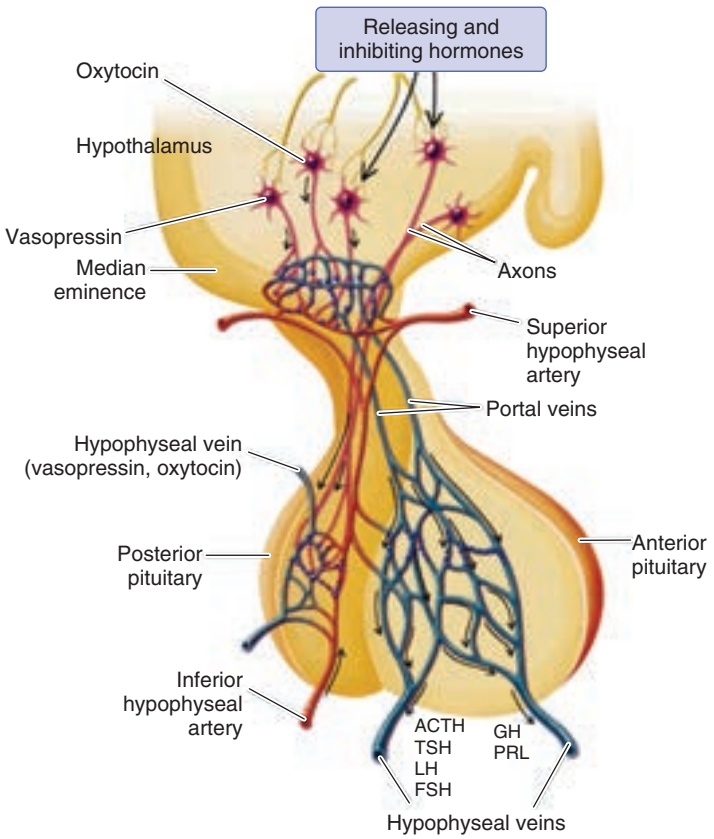

FIGURE 224-1. Structural-functional, humoral, endocrine, and neuroendocrine relationships within the hypothalamic-pituitary unit emphasize the unique and intimate interdependence of neural structures and hormone secretion with the circulation. Oxytocin and vasopressin neuron bodies located in the hypothalamus send axons through the pituitary stalk that terminate in the posterior pituitary, where they release oxytocin and vasopressin into blood vessels within the posterior pituitary. Hypothalamic neurons that produce growth hormone–releasing hormone, corticotropin-releasing hormone, thyrotropin-releasing hormone, and gonadotropin-releasing hormone send their axons through the median eminence to terminate and release their hormones into the hypophyseal-portal circulation. This network of blood vessels is located at the median eminence, which surrounds the pituitary stalk and penetrates into the anterior lobe of the pituitary. These hypothalamic neurohormones stimulate responsive anterior pituitary cells to secrete growth hormone (GH), adrenocorticotropic hormone (ACTH), thyroid-stimulating hormone (TSH), luteinizing hormone (LH), and follicle-stimulating hormone (FSH), respectively. Dopamine neurons reaching the median eminence are responsible for tonic inhibition of prolactin (PRL) secretion from the anterior pituitary, whereas somatostatin released from somatostatinergic neurons inhibit GH and TRH release. (From Melmed S. *The Pituitary.* 3rd ed. London: Elsevier; 2011.)

RADIOLOGY OF THE PITUITARY

Magnetic resonance imaging (MRI) provides excellent resolution of the pituitary and surrounding cerebrospinal fluid (CSF) and vascular and central nervous system structures (Fig. 224-2).[3] There is less radiation exposure with MRI than with computed tomography (CT), allowing repeated imaging. However, bone structures are not as well defined by MRI compared with CT. On MRI, the normal anterior pituitary appears isointense with brain white matter, whereas the posterior pituitary exhibits high signal intensity ("bright spot"). The optic chiasm is readily identified because it is surrounded by hypodense areas. Pituitary adenomas typically appear hypointense on T1-weighted images and show less enhancement with gadolinium than surrounding normal tissue (Figs. 224-3 and 224-4). Focal hypodense areas are also seen in about one fourth of normal individuals, which may correspond to cysts or nonfunctioning small adenomas, emphasizing the importance of endocrine evaluation.

REGULATION OF THE PITUITARY AXIS

The pituitary gland integrates the influences of an array of positive and negative signals to modulate hormone secretion. PRL is the only major pituitary

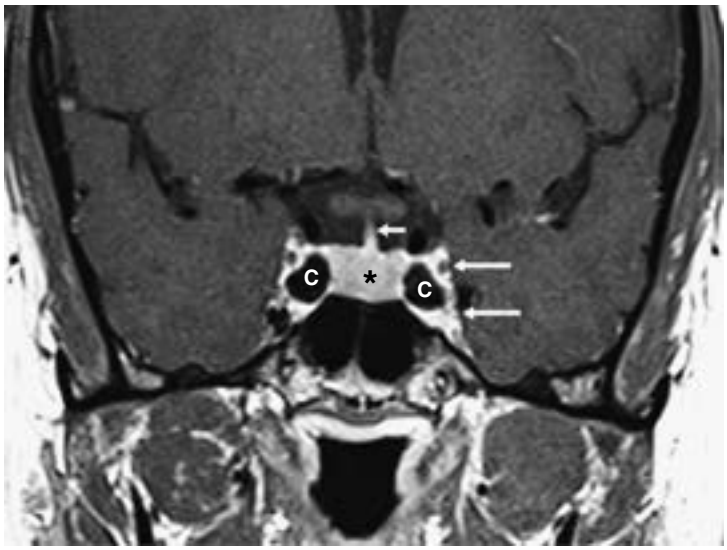

FIGURE 224-2. Magnetic resonance image of normal pituitary. Coronal postcontrast image shows homogeneously enhancing gland (*) and stalk (*short arrow*). Note the greater degree of contrast uptake in the cavernous sinuses, which contain the carotid arteries (C), easily depicted by their flow voids. Small hypointense dots (*long arrows*) are the cranial nerves within the cavernous sinuses. (From Melmed S. *The Pituitary.* 3rd ed. London: Elsevier; 2011.)

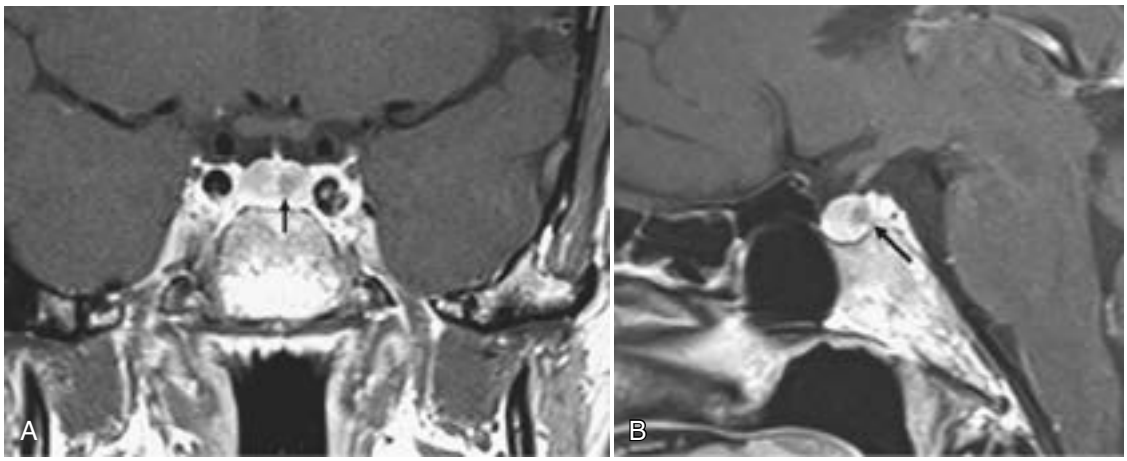

FIGURE 224-3. Magnetic resonance image showing a pituitary microadenoma. (**A**) Coronal and (**B**) sagittal T1-weighted images demonstrate a hypoenhancing lesion within the pituitary gland (*arrows*). (From Melmed S. *The Pituitary.* 3rd ed. London: Elsevier; 2011.)

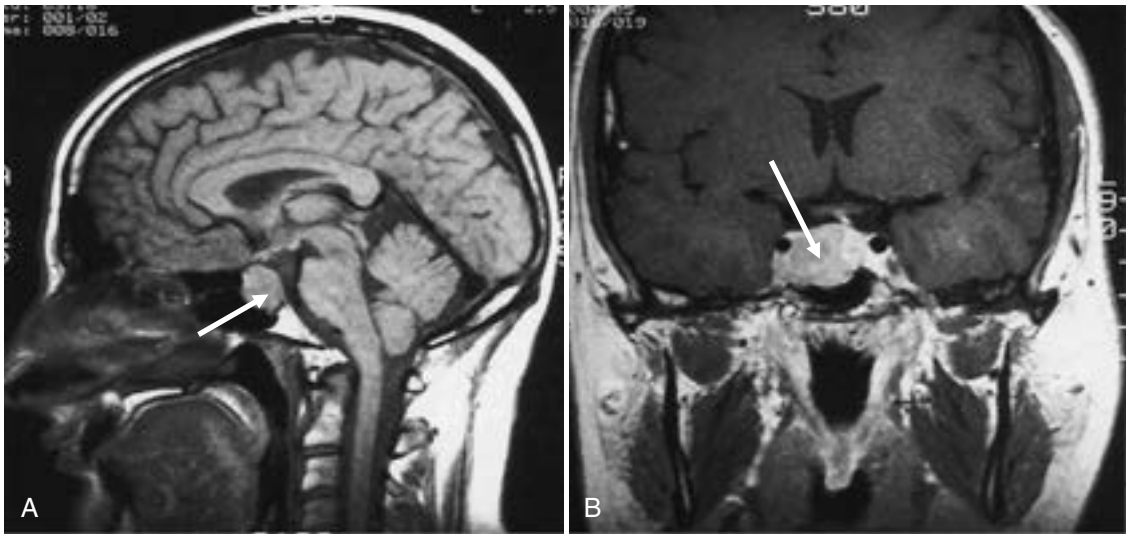

FIGURE 224-4. Magnetic resonance image showing a pituitary macroadenoma. Arrows point to the adenoma. (**A**) Sagittal view. (**B**) Coronal view.

hormone that is not subject to feedback inhibition by hormones produced in target tissues. It is controlled by positive and negative input from the hypothalamus, the latter being dominant.

The principles of feedback regulation are illustrated by the hypothalamic-pituitary-thyroid axis. Hypothalamic thyrotropin-releasing hormone (TRH) stimulates TSH secretion from the pituitary. TSH increases thyroid hormone secretion, which in turn suppresses hypothalamic TRH as well as pituitary TSH. A typical regulatory loop therefore has both positive (TRH, TSH) and negative (thyroxine [T_4], triiodothyronine [T_3]) components, allowing a high degree of control of hormone levels. The pituitary gland integrates positive TRH signals and the negative effects of thyroid hormone. The concept of feedback regulation is important not only for understanding pituitary physiology but also because it provides the basis for analyzing pituitary gland function using stimulation and suppression tests.

The feedback regulatory systems just described are superimposed on hormonal rhythms that are used for adaptation to the environment. Seasonal changes, the daily occurrence of the light-dark cycles, and stress have major impacts on the secretion of pituitary hormones (Chapter 2). Because many hormones are released in a pulsatile manner and in a rhythmic fashion, these characteristics of secretion should be considered when attempting to relate serum measurements to normal values. Although it is possible to characterize pulsatile patterns of hormone secretion using frequent blood sampling (every 10 minutes), this is not practical in a clinical setting. Alternative approaches include stimulation and suppression tests or the use of "integrated" measurements of hormone production, such as 24-hour urine free cortisol as an index of ACTH secretion, or insulin-like growth factor (IGF)-I as a biological marker of GH action.

HYPOPITUITARISM

Pituitary hormone deficiencies can be caused by loss of hypothalamic stimulation or by direct loss of pituitary function. When hypopituitarism is accompanied by diabetes insipidus or hyperprolactinemia, one should consider hypothalamic etiologies.

PATHOBIOLOGY

A variety of congenital and acquired causes of hypopituitarism have been described (Table 224-1). Congenital deficiencies of multiple pituitary hormones are often caused by mutations in the genes for the transcription factors mentioned previously *(HESX1, LHX3, PIT1, PROP1)*. Gene mutations have been found at several steps leading to pituitary hormone secretion, including those for the hypophysiotropic releasing factor receptors for gonadotropin-releasing hormone (GnRH), growth hormone–releasing hormone (GHRH), and TRH; those for the pituitary hormone structures for GH, ACTH, and the β-subunits of FSH, TSH, and LH; and those for the target organ receptors for GH, ACTH, TSH, and LH. Mutations of the GH gene have been shown to be heterogeneous. Some are large deletions that are inherited in an autosomal recessive manner and involve genetic recombination between related DNA sequences in the duplicated GH gene cluster. Point mutations have also been described, and some of these can be inherited in an autosomal dominant manner. Mutations of the other types described earlier generally cause autosomal recessive forms of selective hormone deficiencies.

Neoplastic lesions, particularly pituitary macroadenomas, are the most common cause of acquired hypopituitarism. Pituitary adenomas cause hypopituitarism in several different ways. In some cases, there is direct destruction

TABLE 224-1 CAUSES OF HYPOPITUITARISM

GENETIC DEFECTS
Hypophysiotropic hormone gene defects
Hypophysiotropic hormone receptor gene defects
GHRH receptor defect
GnRH receptor defect
TRH receptor defect
Pituitary hormone gene defects
Gonadotropins: LH β- and FSH β-subunit gene defects
GH: defects in GH gene
Thyrotropin: defects in TSH β-subunit gene
Multiple hormone (GH, PRL, TSH) defects due to mutation in *PIT1* and *PROP1* genes
Pituitary hormone receptor genetic defects
GH receptor defects: GH insensitivity syndrome (Laron-type dwarfism)
ACTH receptor defects: congenital insensitivity to ACTH
LH receptor defects
FSH receptor defects
TSH receptor defects
CONGENITAL EMBRYOPATHIC DEFECTS
Anencephaly
Midline cleft defects: septo-optic dysplasia, basal encephalocele, cleft lip and palate
Pituitary aplasia
Kallmann's syndrome (GnRH defect with anosmia)
ACQUIRED DEFECTS
Tumors: pituitary adenomas, craniopharyngiomas, dysgerminomas, meningiomas, gliomas, metastatic tumors, hamartomas, Rathke's cleft cysts
Irradiation
Trauma: surgery, external blunt trauma
Empty sella syndrome
Vascular
Pituitary apoplexy
Sheehan syndrome
Internal carotid aneurysm
Vasculitis
Subarachnoid hemorrhage
Inflammatory and infiltrative diseases
Sarcoidosis
Langerhans cell histiocytosis (histiocytosis X, eosinophilic granuloma)
Tuberculosis, syphilis
Meningitis
Lymphocytic hypophysitis, infundibulohypophysitis
Metabolic
Hemochromatosis
Amyloidosis
Critical illness
Malnutrition
Anorexia nervosa
Psychosocial deprivation
Idiopathic

ACTH = adrenocorticotropic hormone; FSH = follicle-stimulating hormone; GH = growth hormone; GHRH = growth hormone releasing hormone; GnRH = gonadotropin-releasing hormone; LH = luteinizing hormone; PRL = prolactin; TRH = thyrotropin-releasing hormone; TSH = thyroid-stimulating hormone.

or compression of the normal pituitary. Compression of the pituitary stalk can impair blood supply to the pituitary as well as decrease input from hypothalamic hormones. Hemorrhage into tumors can lead to pituitary infarction. A mild degree of hyperprolactinemia (usually < 200 ng/mL) is characteristic of disorders that cause stalk compression, and hyperprolactinemia further impairs gonadotropin secretion. A variety of other neoplasms that occur near the sella, such as craniopharyngiomas, can also cause hypopituitarism (see Table 224-1).

Radiation causes hypopituitarism primarily because of its effects on hypothalamic function, although high-dose radiation (e.g., proton beam) can also cause direct pituitary damage. The sellar region is subjected to radiation in the treatment of pituitary adenomas, craniopharyngiomas, optic gliomas, meningiomas, dysgerminomas, and neoplasms of the oropharynx. The effects of radiation can be delayed several years, and patients at high risk should be evaluated yearly. Although GH and gonadotropin deficiencies develop first in most patients, ACTH or TSH deficiencies occasionally occur first, emphasizing the need to evaluate each of the major axes.

Empty sella syndrome can occur as a primary or as an acquired condition. It is caused by defects in the diaphragma sella that allow herniation of the arachnoid membrane into the hypophyseal fossa. In long-standing cases, sellar enlargement occurs, probably because of persistent transmission of intracranial pressure. Imaging studies reveal the pituitary gland as a flattened rim of tissue along the floor of the sella. Primary empty sella occurs most commonly in women and may be associated with features of benign intracranial hypertension. Pituitary function is usually normal, but 10% have mild hyperprolactinemia, probably because of stretching of the pituitary stalk. Acquired forms may occur as a result of surgery, radiation, or pituitary infarction (usually of an adenoma).

Pituitary apoplexy is a syndrome caused by hemorrhage into a tumor, with associated infarction and leakage of the blood into the arachnoid space and resultant fever and stiff neck. When the hemorrhage is large, there can be mass effects, with headache and cranial nerve compression within the cavernous sinus. Asymptomatic infarctions are found in about 10 to 15% of pituitary adenomas. In the absence of a tumor, predispositions to apoplexy include trauma, pregnancy, anticoagulation, sickle cell anemia, and diabetes mellitus. Pituitary infarction in the peripartum period (Sheehan syndrome) is usually associated with significant obstetric hemorrhage and hypovolemia. Although Sheehan syndrome can manifest acutely with vascular collapse, it more commonly has a subacute manifestation consisting of postpartum inability to lactate, amenorrhea, and symptoms of adrenal insufficiency.

In lymphocytic hypophysitis, there is infiltration of the pituitary by lymphocytes and plasma cells, with destruction of the parenchyma; it is believed to have an autoimmune basis. The lesion is usually large, and patients present with either symptoms or signs of hypopituitarism or those of a mass lesion (i.e., visual field defects and headaches). Some patients may have mild hyperprolactinemia and diabetes insipidus. Almost all cases have been reported in women, and most present during or after pregnancy. Because of the presentation as a mass lesion during pregnancy, such lesions may be confused with prolactinomas. Mild PRL elevation points to a nonsecretory lesion rather than a prolactinoma. MRI cannot reliably differentiate pituitary adenoma from hypophysitis, although hypophysitis usually manifests with a diffuse enlargement of the pituitary that enhances, rather than as a focal lesion. Diagnosis is usually made by biopsy, but the lesion may be suspected clinically if it manifests during or just after pregnancy. Careful pituitary function testing is mandatory because many patients have died of adrenocortical insufficiency. Although the prognosis is not clear, a number of cases have resolved spontaneously. An entity with similar histologic findings involving the stalk and posterior pituitary, referred to as infundibuloneurohypophysitis, can cause diabetes insipidus. A third variant, panhypophysitis, involves both lobes of the pituitary. These last two forms occur in both sexes and are generally not associated with pregnancy. The causes and interrelationships between these entities remain unknown. In the past several years, secondary hypophysitis has been described in association with ipilimumab, a CTLA-4 blocking antibody used as an immunostimulant in some cancer chemotherapy regimens, and a plasmacytic infiltrative form in which the plasma cells produce antibodies of the IgG4 subclass.[4]

The pituitary may undergo damage because of iron deposition in patients with hemochromatosis (Chapter 212) and amyloid fibrils in patients with systemic amyloidosis (Chapter 188). Functional reversible hypopituitarism of varying degrees occurs in patients with severe systemic illness, severe psychosocial and emotional deprivation, and severe weight loss—particularly in those with anorexia nervosa.

DIAGNOSIS AND TREATMENT

Rx

The diagnosis of hypopituitarism rests on the stimulation tests that are summarized in Table 224-2. Therapy depends on the nature and severity of the hormone deficiencies as well as on the desired clinical end points. The goals are to replace hormones in a physiologic manner and to avoid the consequences of over-replacement. Examples of hormonal replacement paradigms are provided in Table 224-3. Adjustment of hormone doses is done primarily based on clinical findings; TSH levels are not helpful for adjusting thyroxine doses in patients with central hypothyroidism. Even when conventional hormone replacement (adrenal, thyroid, gonadal) is carried out appropriately, there is an approximately two-fold excess risk for death. Although untreated GH deficiency has been hypothesized to be the cause of this excess risk, this has not been proved. Other causes of death in this population include infections with inappropriate increase in steroid dose, and brain tumors related to prior irradiation.[5] The benefits of GH therapy are less clear than those of the other pituitary hormones and include improvements in body composition, bone, lipids, and quality of life; although there are few adverse effects, treatment involves daily injections.[6]

TABLE 224-2 TESTS OF PITUITARY INSUFFICIENCY

HORMONE	TEST	INTERPRETATION
Growth hormone (GH)	*Insulin tolerance test:* Regular insulin (0.05-0.15 U/kg) is given IV, and blood is drawn at −30, 0, 30, 45, 60, and 90 min for measurement of glucose and GH.	If hypoglycemia occurs (glucose < 40 mg/dL), GH should increase to > 5 μg/L.*
	Arginine-GHRH test: GHRH 1 μg/kg IV bolus followed by 30-min infusion of L-arginine (30 g)	Normal response is GH > 4.1 μg/L.
	Glucagon test: 1 mg IM with GH measurements at 0, 60, 90,120, 150, and 180 min	Normal response is GH > 3 μg/L.
Adrenocorticotropic hormone (ACTH)	*Insulin tolerance test:* Regular insulin (0.05-0.15 U/kg) is given IV, and blood is drawn at −30, 0, 30, 45, 60, and 90 min for measurement of glucose and cortisol.	If hypoglycemia occurs (glucose < 40 mg/dL), cortisol should increase by > 7 μg/dL or to > 20 μg/dL.
	CRH test: 1 μg/kg ovine CRH IV at 8 AM, with blood samples drawn at 0, 15, 30, 60, 90, and 120 min for measurement of ACTH and cortisol	In most normal individuals, the basal ACTH increases two- to four-fold and reaches a peak (20-100 pg/mL). ACTH responses may be delayed in cases of hypothalamic dysfunction. Cortisol levels usually reach 20-25 μg/dL.
	ACTH stimulation test: $ACTH_{1-24}$ (cosyntropin), 0.25 mg IM or IV. Cortisol is measured at 0, 30, and 60 min.	A normal response is cortisol > 18 μg/dL. In suspected hypothalamic-pituitary deficiency, a low-dose (1-μg) test may be more sensitive.
Thyroid-stimulating hormone (TSH)	*Basal thyroid function tests:* free T_4, free T_3, TSH	Low free thyroid hormone levels in the setting of TSH levels that are not appropriately increased
Luteinizing hormone (LH), follicle-stimulating hormone (FSH)	*Basal levels of LH, FSH, testosterone, estrogen*	Basal LH and FSH should be increased in postmenopausal women. Low testosterone levels in conjunction with low or low-normal LH and FSH are consistent with gonadotropin deficiency.

*Values are with polyclonal assays.
CRH = corticotropin-releasing hormone; GHRH = growth hormone–releasing hormone; IM = intramuscularly; IV = intravenously; T_3 = triiodothyronine; T_4 = thyroxine.

TABLE 224-3 HORMONAL REPLACEMENT THERAPY IN HYPOPITUITARISM*

PITUITARY AXIS	HORMONAL REPLACEMENTS
Growth hormone (GH)	In children, GH (0.25 mg/kg) SC daily. In adults, GH (0.3-1.2 mg) SC daily. Titrate dose to achieve IGF-I levels in middle to upper part of normal range. Women receiving oral estrogens require higher doses.
Prolactin	None
Adrenocorticotropic hormone–cortisol	Hydrocortisone (10-15 mg PO q AM; 5-10 mg PO q PM) or prednisone (2.5 mg PO q AM; 2.5 mg PO q PM). Dose adjusted on clinical basis. Stress dosing: 50-75 mg hydrocortisone IV q8h
Thyroid-stimulating hormone–thyroid	L-thyroxine (0.075-0.15 mg) PO daily
Gonadotropins–gonads	FSH and LH (or HCG) can be used to induce ovulation in women. HCG alone or with FSH can be used to induce spermatogenesis in men. In men, testosterone enanthate (100-300 mg) IM q1-3wk or testosterone cyclopentylpropionate (100-300 mg) IM q1-3wk. Testosterone transdermal patches can also be used (5 mg daily). Testosterone gel 5-10 g daily. In women, conjugated estrogens (0.625-1.25 mg) PO days 1-25 each month, cycled with medroxyprogesterone acetate (5-10 mg) PO days 15-25 each month. Low-dose contraceptive pills may also be used. Estrogen-containing transdermal patches are also available.
Posterior pituitary	Desmopressin, 0.05-0.2 mL (5-20 μg) intranasally once or twice daily, or tablets (0.1-0.4 mg q8-12h) or 0.5 mL (2 μg) SC

*Replacement therapy is dictated by the types of hormone deficiencies and by the clinical circumstances. In each case, the recommended preparations and doses are representative but need to be adjusted for individual patients. Other hormonal preparations are also available.
FSH = follicle-stimulating hormone; GnRH = gonadotropin-releasing hormone; HCG = human chorionic gonadotropin; IGF-I = insulin-like growth factor-I; IM = intramuscularly; LH = luteinizing hormone; PO, orally; SC = subcutaneously.

PITUITARY TUMORS

PATHOBIOLOGY

Pituitary tumors are classified according to the hormones they produce and their size: microadenomas, less than 10 mm in diameter; macroadenomas, more than 10 mm in diameter; and macroadenomas with extrasellar extension. In general, the levels of hormones produced by the tumors parallel the size of the tumors. The prevalence of the different types of pituitary adenomas, based on surgical data, is summarized in Table 224-4. Immunohistochemical studies using antibodies specific for each of the major pituitary hormones have been used to define tumor phenotype. Pituitary adenomas are rarely malignant but can be locally invasive.

Most pituitary tumors are monoclonal. This finding does not exclude a role for hormonal stimulation as a predisposing factor for somatic mutations, and the hormonal environment may also affect the rate of tumor growth (e.g., the growth of ACTH-secreting tumors following bilateral adrenalectomy). Supporting the concept that somatic mutations lead to pituitary tumorigenesis, a subset (35 to 40%) of somatotroph adenomas have activating mutations in the gene for the Gsα-subunit, resulting in two different amino acid (Arg201 and Glu227) substitutions.[7] Either mutation causes the Gsα-subunit to stimulate adenylyl cyclase in a constitutive manner. The elevated intracellular cyclic adenosine monophosphate levels lead to increased cell growth as well as GH production. Mutations in other oncogenes, such as *ras*, *Rb*, and *p53*, are uncommon in pituitary tumors. Thus, the nature of the somatic mutations causing most pituitary tumors remains unknown.

At least five types of inherited predispositions to pituitary tumors are recognized. Patients with McCune-Albright syndrome (Chapter 231) occasionally develop pituitary adenomas as well as characteristic abnormalities in other tissues, particularly the ovary, bone, and thyroid. Interestingly, the McCune-Albright syndrome is also caused by mutations in the gene for the Gsα-subunit. However, the somatic mutations in McCune-Albright occur early during development, so that multiple tissues are affected. In multiple endocrine neoplasia type 1 (MEN 1) (Chapter 231), the menin gene is mutated, so that the predisposition to pituitary tumors is inherited in an autosomal dominant manner and occurs in conjunction with tumors of the parathyroid and pancreas.[8] Familial isolated pituitary adenoma (FIPA) syndrome is autosomal dominant and has low or variable penetrance. Germline mutations have been found in the gene for the aryl hydrocarbon receptor–interacting protein *(AIP)*, which functions as a tumor suppressor. Such mutations have been found in about one third of FIPA families, most commonly

TABLE 224-4 PREVALENCE OF DIFFERENT TYPES OF PITUITARY ADENOMAS

TYPE OF PITUITARY ADENOMA	DISORDER	HORMONE PRODUCED	PREVALENCE (%)*
Somatotroph	Acromegaly and gigantism	Growth hormone	10-15
Lactotroph (prolactinoma)	Hypogonadism, galactorrhea	Prolactin	25-40
Corticotroph	Cushing's disease	Adrenocorticotropic hormone	10-15
Gonadotroph	Mass effects, hypopituitarism	Follicle-stimulating hormone and luteinizing hormone	15-20
Thyrotroph	Hyperthyroidism	Thyroid-stimulating hormone	<3
Nonfunctioning/ null cell	Mass effects, hypopituitarism	None	10-25

*The prevalence rates represent ranges described in several different large surgical series. Mixed tumors (e.g., growth hormone and prolactin) and plurihormonal adenomas are not shown. Rates vary depending on methods used to establish the diagnosis. Prolactinomas were underestimated in most recent pathologic series because they are largely managed medically. Most glycoprotein hormone–producing pituitary tumors were classified as nonfunctioning adenomas until the application of immunohistochemical studies.

in those with GH- and PRL-producing tumors. Carney's complex is an autosomal dominant condition consisting of pituitary adenomas, atrial myxomas, spotty skin pigmentation, and schwannomas. Primary pigmented nodular adrenocortical disease causing Cushing's syndrome occurs in about one third of patients with Carney's complex. The complex is caused by an inactivating mutation in the gene for the type 1A regulatory subunit of protein kinase A (PRKARIA).[7] The most recently discovered familial form involves mutations in the succinate dehydrogenase subunit genes, resulting in combinations of pituitary tumors and pheochromocytomas/paragangiomas.[9] In general, those tumors that arise as part of MEN 1 and FIPA tend to be more aggressive, occur at a younger age, and are less responsive to therapeutic interventions. It is now recommended that when a macroadenoma occurs in the context of a very young age or when there is a family history of pituitary or other endocrine tumors, that appropriate genetic screening be carried out.

CLINICAL MANIFESTATIONS

Many of the clinical manifestations of pituitary adenomas are related to the hypersecretion of hormones. However, the mass effects of the enlarging tumor can also lead to specific signs and symptoms (Chapter 189). Particularly in the case of nonfunctioning tumors or in those that produce gonadotropins, the primary clinical manifestations are related to effects of the tumor on surrounding structures.

Headaches are common in patients with macroadenomas and appear to be caused by expansion of the diaphragma sellae or by invasion of bone. The sudden onset of severe headache associated with nausea, vomiting, and altered consciousness can also be caused by hemorrhagic infarction with sudden enlargement of a pituitary adenoma. Severe cases require glucocorticoid treatment and possible surgical decompression.

Pituitary tumors that extend superiorly can affect visual fields. Expansion into the suprasellar region exerts pressure on the optic chiasm, usually in the central region where nerves emanating from the inferior and medial part of the retina (superior and temporal visual fields) cross, leading to bitemporal hemianopsia. Visual field loss is variable, however, and is affected by the location and flexibility of the chiasm as well as by the direction and extent of tumor growth.[10] Large tumors may invade the cavernous sinus or surround an optic nerve, leading to other patterns of visual field changes or loss of visual acuity. The tumor size and the direction and degree of extrasellar extension are best evaluated with MRI with gadolinium. If the tumor abuts the chiasm on MRI, formal visual field testing should be performed. Even long-standing visual loss may be reversible.

The normal pituitary is often compressed into a thin rim of tissue by large pituitary adenomas. Hypopituitarism probably results more from compression of the hypothalamic-pituitary stalk than from direct pressure on the normal pituitary. GH deficiency and hypogonadotropic hypogonadism are particularly common. Slightly elevated PRL levels (generally < 200 ng/mL) occur in cases of stalk compression because of diminished inhibition by dopamine. It is important not to mistake such tumors for prolactinomas, because they will only rarely decrease in size in response to medical therapy with dopamine agonists, unlike prolactinomas, which commonly do. A peculiarity of some two-site immunoassays, referred to as the "hook effect," can sometimes cause a very high PRL level to read falsely normal or just mildly elevated; a 1:100 dilution of the serum sample with saline will show the true PRL level.[11] Preoperative hypopituitarism caused by a large pituitary mass is reversible in up to half of patients after surgical decompression.[9] Diabetes insipidus (vasopressin deficiency) is rarely caused by pituitary tumors and should raise the suspicion of a craniopharyngioma or other hypothalamic disorders.[12]

TREATMENT Rx

Surgery

Except for prolactinomas, surgery is the primary mode of therapy for pituitary tumors that warrant intervention (Video 224-1). Indications for surgery include reduction in hormone levels and decompression to relieve mass effects or prevent further tumor expansion. Currently, the transsphenoidal route, usually with an endoscopic endonasal approach, is used almost exclusively for decompression or extirpation of pituitary tumors.[13] Because of substantially greater morbidity, subfrontal craniotomy is reserved for patients with tumors that require extensive exploration of the suprasellar region and surrounding structures. In experienced hands, transsphenoidal surgery is effective, and complications are uncommon (<5% complication rate) but include CSF leak, hemorrhage, optic nerve injury, hypopituitarism, and sinusitis. Transient diabetes insipidus occurs in about 5% of patients after surgery but rarely persists long term. Mortality rates are less than 1%. Complication rates increase with increasing size of the tumor and when a craniotomy is performed.

Surgical cure rates are largely a function of the size and location of the pituitary mass. When stringent hormonal criteria are used to assess surgical success rates, 30 to 60% of macroadenomas are cured by transsphenoidal surgery, although considerable improvements in hormone levels or mass effects can be achieved. On the other hand, hormone hypersecretion by microadenomas can be corrected completely in 80 to 90% of patients, although the cure rates vary considerably at different institutions. Even with apparent surgical cure, 10 to 20% of tumors may recur over several years, resulting in a redevelopment of the hormone oversecretion syndrome.

Radiation Therapy

Irradiation is usually used as adjunctive therapy after surgery or in combination with medical therapy. Radiation has typically been administered over 5 weeks at a dose of 45 Gy using cobalt-60 or a linear accelerator. Proton beam therapy delivers very high doses of radiation within a localized region, but it is limited to intrasellar lesions and is not widely available. More recently, a radiation therapy technique referred to as stereotactic radiotherapy has been employed for many patients with pituitary tumors. With this technique, approximately the same dosage of radiation is administered as a single dose through multiple ports, using a computerized matching of irradiation to tumor geometry.[14] Response rates are slow (several years, but somewhat more rapid with stereotactic therapy). Complete remission is rarely achieved with any of these types of irradiation. Because the time to recurrence for most nonfunctioning macroadenomas is 5 to 10 years, and not all tumors recur, it is often reasonable to follow patients with imaging techniques, reserving irradiation for those with evidence of residual tumor or recurrence if no tumor was visible postoperatively.

Complications of irradiation are dose related but can also be idiosyncratic. Partial or complete hypopituitarism occurs in 50 to 70% of patients and is primarily due to hypothalamic injury. Conventional irradiation is associated with an increased risk for stroke. Second tumors occur in the radiation field in about 2% of patients over a 20-year period. Less common complications include optic nerve damage, brain necrosis, and cognitive dysfunction. Whether stereotactic radiotherapy will have similar rates of complications is at present unknown. Early studies show that hypopituitarism is developing at a rate similar to that seen with conventional therapy.

Medical Therapy

The emergence of medical therapies for pituitary tumors has dramatically affected patient management. The dopamine agonists bromocriptine and cabergoline have a primary role in the management of prolactinomas. They induce a rapid fall in PRL levels and, importantly, decrease tumor size. Dopamine agonists are also used in the management of acromegaly, although the GH responses and effects on tumor size are much less pronounced than in prolactinomas. Somatostatin analogues such as octreotide, lanreotide, and pasireotide act to suppress the secretion of a number of hormones, including GH, TSH, and ACTH and have been used to treat acromegaly, TSH-producing tumors, and Cushing's disease. Other medical therapy for Cushing's disease has primarily been directed toward inhibition of steroid biosynthesis; these drugs include ketoconazole, metyrapone, etomidate, and mitotane. Because of

VIDEO 224-1. **Pituitary surgery.**

substantial side effects and because patients with Cushing's disease tend to escape from the cortisol-suppressing effects of these drugs, medical therapy is used primarily as an adjunctive treatment or to reduce cortisol levels preoperatively. Another new category of drugs to treat pituitary tumors includes the receptor blockers. Pegvisomant is a GH analogue with altered binding to the GH receptor that competitively inhibits GH binding to its receptor and is used in the treatment of acromegaly. Mifepristone can block the action of cortisol at its receptor and has been found to be effective for the treatment of Cushing's syndrome.

GROWTH HORMONE

The most important regulators of GH are the hypothalamic hormones: GHRH, which is stimulatory, and somatostatin, which is inhibitory. GH increases the production of IGF-I (formerly known as somatomedin C). The Gsα-subunit, which is coupled to the GHRH receptor, is one of the targets for activating mutations that lead to somatotroph adenomas.

IGF-I inhibits GH secretion, and it acts at both the pituitary and hypothalamic levels. In addition to reflecting GH action (primarily at the liver), serum IGF-I is also sensitive to nutritional and metabolic changes. In cases of starvation, anorexia nervosa, and poorly controlled diabetes, IGF-I levels are low, resulting in increased levels of GH. In cases of obesity, GH levels are low and GHRH responses are blunted, but IGF-I levels are generally normal and actually increase with increasing body mass index. Stress, exercise, and a variety of neurogenic stimuli also increase GH secretion. Estrogens stimulate GH secretion, but their effects are less pronounced than for PRL, and they inhibit the stimulatory effect of GH on IGF-I production.

Large bursts of GH secretion characteristically occur at night in association with slow-wave sleep. GH levels increase during puberty and decline gradually in adulthood. The amplitude of GH pulses is greater in women than in men, likely reflecting the effects of estrogens. Spontaneous GH pulses can reach 50 ng/mL; consequently, random GH levels can be quite variable. GH responses to GHRH are also variable even within an individual, owing to changes in endogenous somatostatin tone.

GH acts through a single transmembrane receptor that is structurally related to PRL and cytokine receptors. The GH molecule has two distinguishable receptor binding domains that allow it to contact two separate receptor molecules to induce receptor dimerization. Mutations in the gene for the GH receptor cause GH resistance and severe growth retardation, a condition referred to as the GH insensitivity syndrome (Laron-type dwarfism). In such patients, GH levels are elevated and IGF-I levels are low, reflecting the inability of the mutant receptor to transduce the GH signal. Mutations have also been found in signaling intermediates (e.g., the signal transducers of activators of transcription *STAT5b* gene) that mediate GH actions.

Many of the growth and metabolic effects of GH are transmitted indirectly through the actions of IGF-I. GH stimulates IGF-I production in most tissues, where it then exerts autocrine or paracrine effects. Circulating IGF-I is derived predominantly from the liver and is a useful marker of GH action because it has a longer half-life and integrates the effects of GH pulses. Although IGF-I levels are used in the diagnosis of acromegaly and to assess the integrity of the GH axis, factors other than GH (e.g., malnutrition) can alter IGF-I levels. IGF-I acts through widely distributed receptors that are structurally related to insulin receptors. In addition to its growth-promoting and anabolic effects, IGF-I also stimulates mitogenesis in many tissues. The bioactivity of IGF-I is itself modulated by six IGF-binding proteins (IGFBPs). These IGFBPs can inhibit or enhance IGF actions. IGFBP-3 is the major IGFBP in plasma; it is regulated by GH, and its levels generally parallel those of IGF-I.

GH has its major effects on linear growth but also influences a variety of metabolic pathways. Some of these effects are mediated by GH directly, whereas others are conferred by IGF-I. Although the relative roles of GH and IGF-I are debated, their actions are cooperative in many cases. The effects of GH on linear growth appear to be mediated largely by IGF-I, which has been used to stimulate growth in patients with the GH insensitivity syndrome. Linear growth in the fetus and neonate is not GH dependent, as illustrated by the fact that GH-deficient infants have normal birth lengths, although intrauterine IGF-I and IGF-II may be important for fetal growth independent of GH. In contrast, normal postnatal linear growth requires GH, as illustrated by the clinical manifestations of GH deficiency. GH and IGF-I act together to accelerate linear growth markedly, particularly at the time of puberty when sex steroids enhance GH and IGF-I levels.

GH also induces lipolysis and stimulates anabolic activity, including amino acid uptake and protein synthesis. As a result, it reduces body fat, increases lean body mass, and leads to positive nitrogen balance. These properties of GH are most strikingly seen in GH-deficient children who have undergone replacement. GH opposes many of the actions of insulin and can be considered diabetogenic. In diabetic individuals, nocturnal GH secretion accounts in large part for the so-called dawn phenomenon in which there is a decrease in glucose utilization, causing a tendency toward hyperglycemia.

Growth Hormone Deficiency

PATHOBIOLOGY

Causes of GH deficiency include hypothalamic-pituitary disorders, GHRH receptor mutations, GH gene mutations, combined pituitary hormone deficiencies, GH receptor mutations, IGF-I receptor mutations, radiation, and psychosocial deprivation (Chapter 223). Isolated idiopathic GH deficiency is the most common category, however, in children.

CLINICAL MANIFESTATIONS

The clinical manifestations of GH deficiency depend on the time of onset and the severity of hormone deficiency. Children with complete GH deficiency have slow linear growth rates (≈3 cm/year), and they rapidly fall below normal on standardized growth charts. GH-deficient children have normal skeletal proportions, and many have a pudgy, youthful appearance because of decreased lipolysis. Particularly in the setting of cortisol deficiency, there is a predisposition to hypoglycemia. Adults may acquire GH deficiency due to hypothalamic/pituitary disease such as tumors or infiltrative disease. Manifestations of adult GH deficiency include increased fat mass, decreased muscle mass, decreased bone mineral density, and decreased quality of life.

DIAGNOSIS

Basal GH does not provide a reliable measure of GH reserve, whereas low IGF-I is consistent with GH deficiency. However, not all patients documented to have GH deficiency by stimulation testing have IGF-I levels that are below the normal range. GH deficiency is assessed using insulin-induced hypoglycemia, which activates central nervous system pathways, leading to stimulation of both GH and ACTH secretion (see Table 224-2). The insulin tolerance test requires careful monitoring for symptoms of severe hypoglycemia, such as confusion or depressed consciousness. This test should be avoided in patients with seizure disorders or coronary artery disease. Insulin doses (0.1 to 0.15 U/kg) may need to be decreased if glucocorticoid deficiency is suspected or increased in conditions of insulin resistance (e.g., obesity). Alternatives to the insulin tolerance test for evaluation of GH are a combination of arginine and GHRH given intravenously or glucagon given intramuscularly.

TREATMENT

Rx

In children with well-documented GH deficiency, GH replacement is effective and is essential to increase final adult height. In a typical regimen, recombinant GH (0.025 mg/kg) is given daily as subcutaneous injections. The efficacy of GH treatment depends on when it is initiated as well as replacement of other hormone deficiencies, if they coexist. In the setting of multiple hormone deficiencies, replacement of thyroid hormone and cortisol is necessary for effective GH action. On the other hand, sex steroids lead to epiphyseal closure and limit linear growth. Consequently, GH is more effective before puberty; if exogenous sex steroids are given, low doses should be used. GH has also been shown to increase the final height of girls with Turner's syndrome (chromosomal XO state) (Chapter 236) and children with end-stage renal disease.

Only about one third of children with isolated idiopathic GH deficiency are found to be GH deficient when retested as adults. Thus, all such patients should be retested before GH therapy is continued or restarted unless they have proven molecular defects as the cause of their GH deficiency. Studies show that GH treatment can increase bone density and lean body mass, decrease fat mass, and improve the sense of well-being in adults with documented GH deficiency. In most studies, safety data for long-term GH administration show negligible adverse effects. Whether GH treatment in adults will affect the increased mortality rate associated with hypopituitarism remains to be seen. Adverse effects occur at lower doses in adults compared with children; a starting dose of 0.2 to 0.3 mg/day has been recommended, with gradual titration guided by clinical benefits, adverse effects, and IGF-I levels.

Growth Hormone Excess: Acromegaly and Gigantism

PATHOBIOLOGY

GH-producing pituitary tumors account for 10 to 15% of pituitary tumors (see Table 224-4). GH-producing tumors are frequently mixed tumors that secrete more than one hormone. PRL is produced in about 40% of somatotroph adenomas, and some patients may present because of symptoms due to the hyperprolactinemia (i.e., amenorrhea, galactorrhea, or both). Ectopic production of GHRH (usually carcinoid or pancreatic islets) is a well-documented but rare (<1%) cause of acromegaly that can result in somatotroph hyperplasia. Activating Gsα-subunit mutations occur in 35 to 40% of somatotroph adenomas. Molecular defects in the remaining 60 to 65% of somatotroph adenomas remain unidentified.

CLINICAL MANIFESTATIONS

Tumors that secrete GH cause acromegaly in adults and gigantism in children in whom GH excess occurs before epiphyseal closure. Acromegaly affects men and women with equal frequency and is most often recognized when patients are in their 30s or 40s, usually after a decade of GH excess. The most striking features of acromegaly usually involve the face, hands, and feet. The diagnosis is often suspected because of changes in facial appearance that include enlargement of the lower jaw (prognathism), the nose and lips, and the sinuses (causing frontal bossing) (Fig. 224-5). Oral cavity changes include malocclusion, increased spacing between the teeth, and enlargement of the tongue. A hollow, resonant voice is caused by changes in the vocal cords and the soft tissues of the hypopharynx. Sleep apnea may occur in patients with soft tissue obstruction of the pharynx but may also occur because of a central disorder. In addition to bony enlargement, there is a marked increase in the soft tissue of the hands and feet, leading to progressive increases in ring, glove, and shoe size. A moist, doughy, enveloping handshake is characteristic of acromegaly. Arthritis (hands, feet, hips, knees) is common (75%) and is caused by cartilage and synovial overgrowth. Some degree of carpal tunnel syndrome is seen in about half of patients. Skin changes include increased skinfolds, particularly over the brow and forehead. The skin is usually oily, owing to increased sebaceous activity and sweating. Skin tags are common, and their presence correlates with the presence of colonic polyps. Galactorrhea may be seen in women, and reproductive dysfunction occurs in both women and men when PRL levels are elevated. Headaches, visual field defects, and other neurologic symptoms depend on the location and extent of tumor growth.

Acromegaly causes a two- to three-fold increase in mortality rate.[15] Most of the increased mortality can be attributed to cardiovascular and cerebrovascular diseases and may be related to the increased prevalence of hypertension (25 to 35%) and diabetes mellitus (10 to 25%) in patients with acromegaly. There is evidence for cardiac hypertrophy in most patients, and symptomatic heart disease, consisting of coronary ischemia or congestive heart failure or both, occurs in 15 to 20% of patients. Sleep apnea may predispose patients to cardiac dysrhythmias. Some analyses have found an increased risk of premalignant polyps and colon cancer, and screening with colonoscopy is generally recommended. The disfigurement, metabolic complications, and increased mortality associated with acromegaly emphasize the importance of early diagnosis and implementation of appropriate therapy to lower the GH levels into the normal range.

DIAGNOSIS

Because GH is secreted in a pulsatile manner and because the amplitude of normal GH pulses can be large (>50 ng/mL), random GH level measurements are not very useful in making the diagnosis of acromegaly. IGF-I levels provide an integrated index of GH production and provide a better screening test for acromegaly. IGF-I levels decrease with age, so normal ranges must be age adjusted. IGF-I levels correlate well with 24-hour GH production rates and with disease activity. The most standardized test for acromegaly is the glucose tolerance test (Table 224-5). In acromegaly, increased glucose levels fail to suppress GH levels to below 1 ng/mL with polyclonal antibody immunoassays and to levels below 0.4 ng/mL using newer monoclonal antibody two-site assays. Co-secretion of PRL should be evaluated. After the diagnosis of acromegaly is made, radiologic studies, preferably using MRI, should be used to evaluate the extent of tumor growth. Unlike in Cushing's disease and prolactinomas, most patients with acromegaly have macroadenomas.

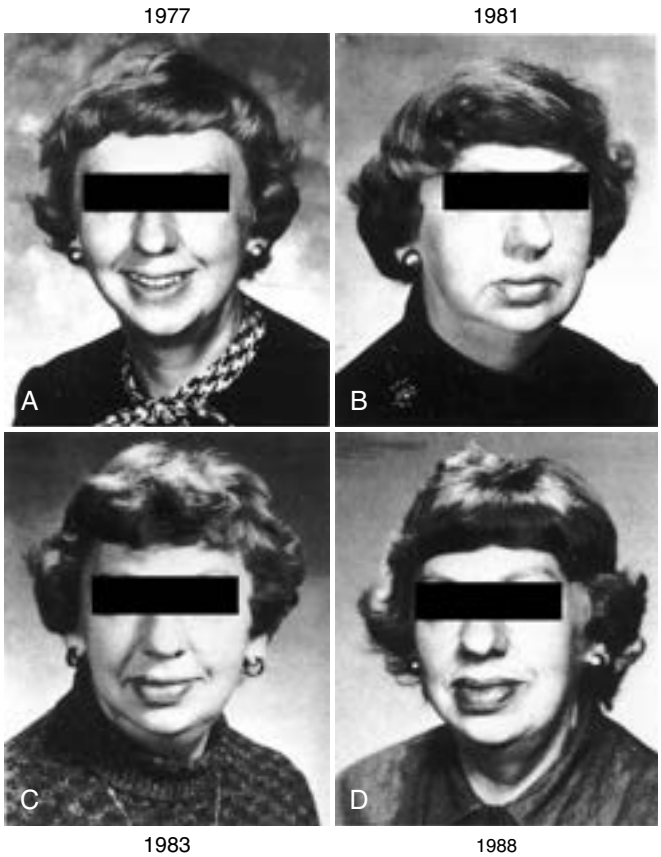

FIGURE 224-5. Clinical features of acromegaly. Serial photographs of a 64-year-old woman with acromegaly. Over an 11-year period, there is a progressive coarsening of facial features, including enlargement of the nose and lips and development of prognathism. She also experienced hypertension, arthropathy, and enlargement of the hands (not shown). (From Molitch ME. Clinical manifestations of acromegaly. *Endocrinol Metab Clin North Am.* 1992;21:597-614.)

TREATMENT

Rx

The goals of therapy are to reverse or prevent tumor mass effects and to reduce the long-term morbidity and mortality that result from excess GH production. Correction of the disorder prevents further physical disfigurement and can result in substantial resolution of soft tissue changes and improvements in metabolic derangements. Although reductions in GH levels are associated with improvements in symptoms, the ultimate goal is to achieve normal GH and IGF-I levels and to prevent tumor recurrence without incurring hypopituitarism.

Transsphenoidal surgery results in GH levels below 2.5 ng/mL in 80 to 90% of patients with microadenomas when performed by experienced neurosurgeons. This level, along with GH suppression below 1 ng/mL during an oral glucose tolerance test, and a normal IGF-I level have been associated with a normalization of the increased mortality of acromegaly. Patients with macroadenomas are less often cured by surgery (<30%) but usually have reductions in GH levels.

Medical therapies for acromegaly include: dopamine agonists such as cabergoline; somatostatin analogues such as octreotide, lanreotide, and pasireotide; and the GH receptor antagonist, pegvisomant. Although cabergoline can reduce GH and IGF-I levels in many patients, normal levels are achieved in only about one third. Long-acting preparations of octreotide, lanreotide, and especially pasireotide that can be given by intramuscular injection every 4 weeks reduce GH and IGF-I levels in almost all patients, with normal levels of IGF-I being achieved in about 50 to 60% of cases. A recent multicenter, randomized trial showed that the somatostatin analog pasireotide (long-acting release), at a dose of either 40 mg or 60 mg administered once every 28 days for 24 weeks, provides superior efficacy compared with continued treatment with octreotide or lanreotide, and could become the new standard pituitary-directed treatment in patients with acromegaly who are inadequately controlled with first-generation somatostatin analogs. Tumor size is reduced modestly in about one half of cases. Of those achieving normal levels of GH and IGF-I, about 10 to 20% can eventually be successfully withdrawn from treatment after several years. Somatostatin analogues are useful as adjunctive therapy in patients who are not cured by surgery or radiation and in some cases are used primarily when a surgical cure is not possible, such as in those with cavernous sinus invasion. Side effects of somatostatin analogues include diarrhea and increased risk of cholelithiasis, although cholecystitis and need for cholecystectomy are rare. Some patients experience additive beneficial effects from combining these two classes of medications while keeping the dose of each drug low enough to avoid adverse effects.

TABLE 224-5 SELECTED TESTS OF EXCESS PITUITARY FUNCTION

HORMONE	TEST	INTERPRETATION
Growth hormone (GH)	*Basal IGF-I*	Elevated IGF-I levels are consistent with acromegaly when interpreted in the context of age and nutritional status.
	Oral glucose suppression test: after 75-g glucose load, GH is measured at −30, 0, 30, 60, 90, 120 min	GH should be suppressed to < 1 μg/L in normal persons with polyclonal radioimmunoassays; < 0.4 μg/L with two-site monoclonal assays. GH may paradoxically increase in acromegaly.
Prolactin	*Basal prolactin levels*	Elevated prolactin (>200 μg/L) is consistent with a prolactinoma. When prolactin levels are between 20 and 200 μg/L, other causes of hyperprolactinemia should be considered.
Adrenocorticotropic hormone (ACTH)	*Measurement of 24-hr urine free cortisol*	Elevated level is suggestive of Cushing's syndrome, but it has several other causes as well.
	Midnight salivary cortisol: special tubes with cotton pledgets available to collect saliva at 11 PM to midnight	In normal persons, the midnight salivary cortisol is very low because of the normal diurnal variation. In patients with Cushing's syndrome, the salivary cortisol is elevated.
	Overnight dexamethasone suppression test: dexamethasone (1 mg) PO at midnight, followed by 8 AM plasma cortisol	In normal persons, AM cortisol should be suppressed to < 5 μg/dL. Normal test excludes Cushing's syndrome. Other disorders can cause failure to suppress normally.
	CRH test: ovine CRH (1 μg/kg) is administered IV, and ACTH and cortisol are drawn at −15, 0, 15, 30, 60, 90, and 120 min.	In Cushing's disease, there is usually a 50% increase in ACTH and a 20% increase in cortisol. Adrenal adenoma is associated with suppressed ACTH. Ectopic ACTH is associated with high basal ACTH and cortisol levels that are not affected by CRH.
	Petrosal sinus ACTH sampling: the inferior petrosal sinus is catheterized bilaterally, and plasma ACTH is compared with simultaneous peripheral samples. The sampling can be done in conjunction with CRH stimulation.	In Cushing's disease, the ratio of ACTH in the petrosal sinus to the periphery is at least 2 basally and at least 3 after CRH. In ectopic ACTH, the ratio of petrosal sinus to peripheral level is < 1.5.
Thyroid-stimulating hormone (TSH)	*Basal thyroid function tests*	An inappropriate normal or elevated TSH in the setting of increased free thyroid hormone levels is consistent with a TSH-producing tumor or other causes of inappropriate TSH secretion.
	Free α-subunit level	Elevated levels associated with inappropriately elevated TSH are suggestive of a TSH-producing tumor.
Follicle-stimulating hormone (FSH), luteinizing hormone (LH)	*Basal FSH, LH, testosterone*	Increased LH and testosterone levels in males are consistent with LH-secreting tumors. Elevated FSH and low-normal testosterone are suggestive of an FSH-producing tumor if primary gonadal failure is not present. In females, assessment of excess hormone secretion is difficult because of changes during the menstrual cycle and at menopause.

CRH = corticotropin-releasing hormone; IGF = insulin-like growth factor; TRH = thyrotropin-releasing hormone.

Pegvisomant is a biosynthetic GH analogue that prevents binding of GH to its receptor. It is capable of normalizing IGF-I levels in more than 90% of patients with corresponding clinical benefits, but it has no effects on the tumor itself. Pegvisomant is given by daily subcutaneous injection; although it generally has been held in reserve for patients not responding optimally to other treatment modalities, its high biochemical and clinical efficacy have led to increasing use now as initial medical therapy when tumors are not large. It has also been given in combination with somatostatin analogues. The most common adverse effect of pegvisomant is an increase in serum transaminase levels.

Radiation is not recommended as primary therapy for acromegaly because of the long time (5 to 10 years) required for reductions in GH levels and the high incidence of hypopituitarism and other complications. Adjunctive radiation therapy may be required for patients with macroadenomas when GH levels or mass effects persist after transsphenoidal surgery and medical therapy. Recent data suggest that stereotactic radiotherapy may be the most efficacious form of radiotherapy for acromegaly.

PROLACTIN

PRL and GH appear to be derived from a common ancestral gene, which accounts for the similarities in their structures and some overlap in their functional properties. Although large-molecular-weight forms of PRL (due to binding to immunoglobulin [Ig]G, termed *macroprolactin*) react in radioimmunoassays, they have diminished biologic potency. Estrogen stimulates lactotroph proliferation, and their number is consequently greater in females than in males and during pregnancy (≈70% of pituitary cells).

Secretion of PRL is controlled by tonic inhibition by dopamine, which acts through D_2-type receptors on lactotrophs. PRL biosynthesis and secretion are stimulated by the hypothalamic peptides TRH and vasoactive intestinal peptide (VIP). Hypothyroidism causes increased TRH output and increased sensitivity of the lactotrophs to TRH and can result in hyperprolactinemia. Dopamine inhibition is the dominant influence for PRL secretion; therefore, PRL is the one pituitary hormone that increases after pituitary stalk section. Secretion of PRL is pulsatile and increases with sleep, stress, chest wall stimulation, and pregnancy. PRL levels are usually less than 15 to 20 ng/mL in women and 10 to 15 ng/mL in men. The primary function of PRL is to induce and sustain lactation. During pregnancy, PRL levels increase, and in conjunction with other hormones (estrogens, progesterone, thyroid hormone, cortisol, and insulin), breast epithelium is stimulated to proliferate and milk synthesis is induced. High levels of estrogen and progesterone inhibit lactation during pregnancy, and their decline post partum permits lactation to occur. Neural pathways leading to the secretion of oxytocin provide the "let-down" reflex that induces lactation in response to suckling. Early in the postpartum period, PRL secretion is stimulated by suckling, but this response becomes damped with time as the frequency of suckling episodes decreases. PRL also suppresses gonadotropins. As a result, breast-feeding can suppress ovulation.

Prolactin Deficiency

PRL deficiency is rare and occurs primarily in the setting of combined hormone deficiencies. The only recognized consequence of PRL deficiency is the absence of postpartum lactation, and this scenario may be found with pituitary infarction occurring as a result of obstetric hemorrhage (Sheehan syndrome). No effects on breast development or other tissues have been described in PRL deficiency.

Hyperprolactinemia

PATHOBIOLOGY

Hyperprolactinemia can occur as a consequence of pharmacologic alterations in the pathways that control PRL secretion or of physiologic or metabolic effects on PRL production and clearance or as a neoplastic condition. Prolactinomas are neoplastic growths of lactotroph cells and are the most common type of pituitary adenoma (25 to 40%). Estrogen is a potent stimulus for lactotroph proliferation; however, there is no clear association between estrogens (e.g., oral contraceptive use) and the incidence of prolactinomas. However, the very high estrogen levels present during pregnancy may cause

about 30% of large prolactinomas to increase in size. Diminished dopamine tone, such as may occur with prolonged treatment with antipsychotic agents, results in increased PRL but has not been shown to cause prolactinomas.

Microprolactinomas constitute the great majority of tumors in premenopausal women. In contrast, macroadenomas are more commonly seen in men and postmenopausal women. The predominance of smaller tumors in premenopausal women may be accounted for by an ascertainment bias, because elevated PRL levels in this group lead to clinical manifestations (amenorrhea, galactorrhea, or infertility). Subclinical prolactinomas exist in men and many older women, and about 5% of apparently normal individuals have PRL-positive microadenomas in autopsy series. Prolactinomas in children tend to be macroadenomas, possibly owing to an increased prevalence of *AIP* mutations (see earlier).

CLINICAL MANIFESTATIONS

Hyperprolactinemia causes galactorrhea and oligomenorrhea or amenorrhea in premenopausal women. Estrogen facilitates PRL-induced galactorrhea, which explains why it is less common in postmenopausal women. Amenorrhea is primarily a consequence of PRL suppression of GnRH, although PRL may also have inhibitory effects at the level of the pituitary and the gonad. Amenorrhea is associated with infertility, and PRL levels should be a routine part of the hormonal evaluation of infertility. Estrogen deficiency can cause decreased libido, vaginal dryness, and dyspareunia. Long-standing estrogen deficiency also leads to osteopenia in many women. Oral contraceptives may mask PRL-induced oligomenorrhea or amenorrhea that becomes apparent on their discontinuation. In postmenopausal women, prolactinomas are often identified because of mass effects rather than because of their hormonal effects.

In men, hyperprolactinemia causes hypogonadism with suppressed LH and FSH levels and low testosterone levels. Hypogonadism causes diminished libido, impotence, infertility, and rarely, gynecomastia or galactorrhea. Diminished libido may also reflect suppression of GnRH, because testosterone replacement is not as effective as suppression of hyperprolactinemia. Hyperprolactinemia is found in 1 to 2% of men being evaluated for sexual dysfunction.

DIAGNOSIS

There are four primary categories of causes of hyperprolactinemia that must be distinguished if the correct therapy is to be instituted: (1) physiologic or metabolic hyperprolactinemia, (2) pharmacologic hyperprolactinemia, (3) hypothalamic or pituitary stalk compression, and (4) prolactinoma (see Table 224-4).[16] With the exception of pregnancy and renal failure, physiologic causes of increased PRL result in minor elevations in PRL, usually less than 50 ng/mL. Primary hypothyroidism should be excluded as a cause of mild hyperprolactinemia. A careful drug history should be obtained in all patients with hyperprolactinemia because of the large number of agents that can stimulate PRL secretion. Psychotropic medications in particular can increase PRL, either by reducing dopamine production or by blocking its action. In most cases, the degree of hyperprolactinemia caused by drugs is less than 150 ng/mL. A variety of suprasellar and parasellar mass lesions cause hyperprolactinemia (generally between 20 and 100 ng/mL) because of compression of the hypothalamus or pituitary stalk. Unless there is very good evidence for physiologic or drug-induced hyperprolactinemia, even patients with mild hyperprolactinemia should be evaluated with MRI to distinguish among idiopathic hyperprolactinemia, microprolactinomas, and other large mass lesions that cause stalk compression resulting in decreased dopamine reaching the lactotrophs. However, specific caution is needed when some two-site assays are used, because patients with very high PRL levels may appear to have PRL levels that are normal or only modestly elevated, owing to the hook effect, in which the very high PRL levels saturate the antibodies in the assay. To avoid this problem, PRL levels should be remeasured at 1 : 100 dilution in patients with large macroadenomas (>3 cm) and normal to modestly elevated PRL levels, because PRL levels in samples with the hook effect will then increase dramatically. When no pituitary lesions are seen by radiographic studies and physiologic and pharmacologic causes of hyperprolactinemia cannot be identified, the diagnosis of idiopathic hyperprolactinemia is made. Idiopathic hyperprolactinemia may represent microprolactinomas too small to be detected accurately by imaging or altered hypothalamic regulation of PRL secretion. Whether such patients should be treated depends on the clinical effects of hyperprolactinemia. When followed for several years, few of these patients develop large tumors, only 10 to 15% show MRI evidence of microadenomas, and in one third of cases, the hyperprolactinemia resolves.

TREATMENT Rx

The natural history of prolactinomas has been evaluated in several series. Although large prolactinomas evolve from smaller lesions, it is uncommon (≈7%) for microprolactinomas to progress to macroadenomas. Because of the slow rate of growth, it is reasonable to monitor patients with microprolactinomas without treatment by periodic measurement of PRL levels unless the hyperprolactinemia is causing symptoms that warrant therapy.

When hyperprolactinemia causes hypogonadism, osteopenia, or infertility, a dopamine agonist such as cabergoline or bromocriptine is the therapy of choice. Dopamine agonists normalize PRL levels and correct amenorrhea-galactorrhea in 80 to 90% of patients. Cabergoline is more effective, has fewer adverse effects than bromocriptine, and has the additional advantage in only having to be taken once or twice weekly. Bromocriptine must be started in low doses, with gradual increases to avoid side effects (nausea, dizziness, somnolence, and nasal stuffiness).

Cabergoline may cause a considerable reduction in tumor size in patients with macroprolactinomas (≈80 to 90% having > 50% reduction in tumor size), but such size reduction is seen in only about two thirds of patients treated with bromocriptine. Improvements in visual field defects can be seen in about 90% of patients with defects when treated with cabergoline. Thus, it is reasonable to use cabergoline as first-line therapy even in patients with visual field defects, so long as visual acuity is not threatened by rapid progression or recent tumor hemorrhage. Many patients treated with cabergoline whose tumors shrink to the point of nonvisualization on MRI and whose PRL levels are normal can maintain normal PRL levels and not experience tumor reexpansion after therapy has been tapered off. In some cases, prolactinomas appear to be resistant to a dopamine agonist. In these cases, switching from bromocriptine to cabergoline may be successful. Larger-than-standard doses (>2 mg/week) of cabergoline may be effective in normalizing PRL levels. The very high doses of cabergoline used in patients with Parkinson's disease have been associated with cardiac valvular abnormalities; such abnormalities have not been found with conventional doses of cabergoline used in patients with prolactinomas, but monitoring with echocardiography may be prudent in patients taking larger-than-standard doses. Alternatively, transsphenoidal surgery may be used. Although initial remission rates (80 to 90%) for transsphenoidal surgery of microprolactinomas are good, there is long-term recurrence in about 20% of patients. For macroprolactinomas, the initial remission rates with surgery are closer to 30%, with a similar recurrence rate. Radiation therapy, usually stereotactic, is reserved for patients with macroadenomas not responding to either medical or surgical treatment.

Dopamine agonist therapy for infertility, or when there is a possibility of pregnancy, deserves special consideration. These medications can induce ovulation in 80 to 90% of patients with hyperprolactinemia. Although neither bromocriptine nor cabergoline has been associated with congenital malformations, they should be stopped once pregnancy has been achieved. A form of barrier contraception is usually recommended until two to three regular menstrual cycles have occurred. Subsequently, pregnancy can be confirmed if a menstrual period is missed, allowing discontinuation of medication with exposure of the fetus to the drug for only 3 to 5 weeks. At present, the safety data for pregnancy outcome are more limited for cabergoline; therefore, some clinicians prefer bromocriptine when fertility is desired. Less than 3% of patients with microadenomas, but 23% of patients with macroadenomas, develop symptoms of tumor enlargement (headaches, visual field defects) during pregnancy (Fig. 224-6). If symptoms develop, MRI and formal visual field testing should be performed. If there is evidence of visual field compromise or tumor growth, dopamine agonist therapy should be restarted to shrink the tumor. PRL levels are not very useful because they are normally increased in pregnancy and an enlarging tumor may not cause PRL production to increase substantially. Because problems of tumor growth occur most often in patients with macroadenomas, consideration can also be given to the option of transsphenoidal decompression before pregnancy in women with large tumors, so long as fertility can be preserved. If the patient is far advanced in her gestation at the time tumor growth occurs, consideration could also be given to delivering the baby.

ADRENOCORTICOTROPIC HORMONE

ACTH is a 39–amino acid peptide that is derived from a precursor polypeptide, proopiomelanocortin (POMC; 241 amino acids), which encodes several peptides, including ACTH and β-lipotropin (Chapter 223). The biologically active portion of ACTH resides within the first 18 of its 39 amino acids. However, because a synthetic peptide (cosyntropin) that includes the first 24 amino acids has a longer half-life, it is used clinically to assess adrenocortical function. In cases with neoplastic ectopic production of ACTH, the levels of precursor peptides or their processed products may be elevated.

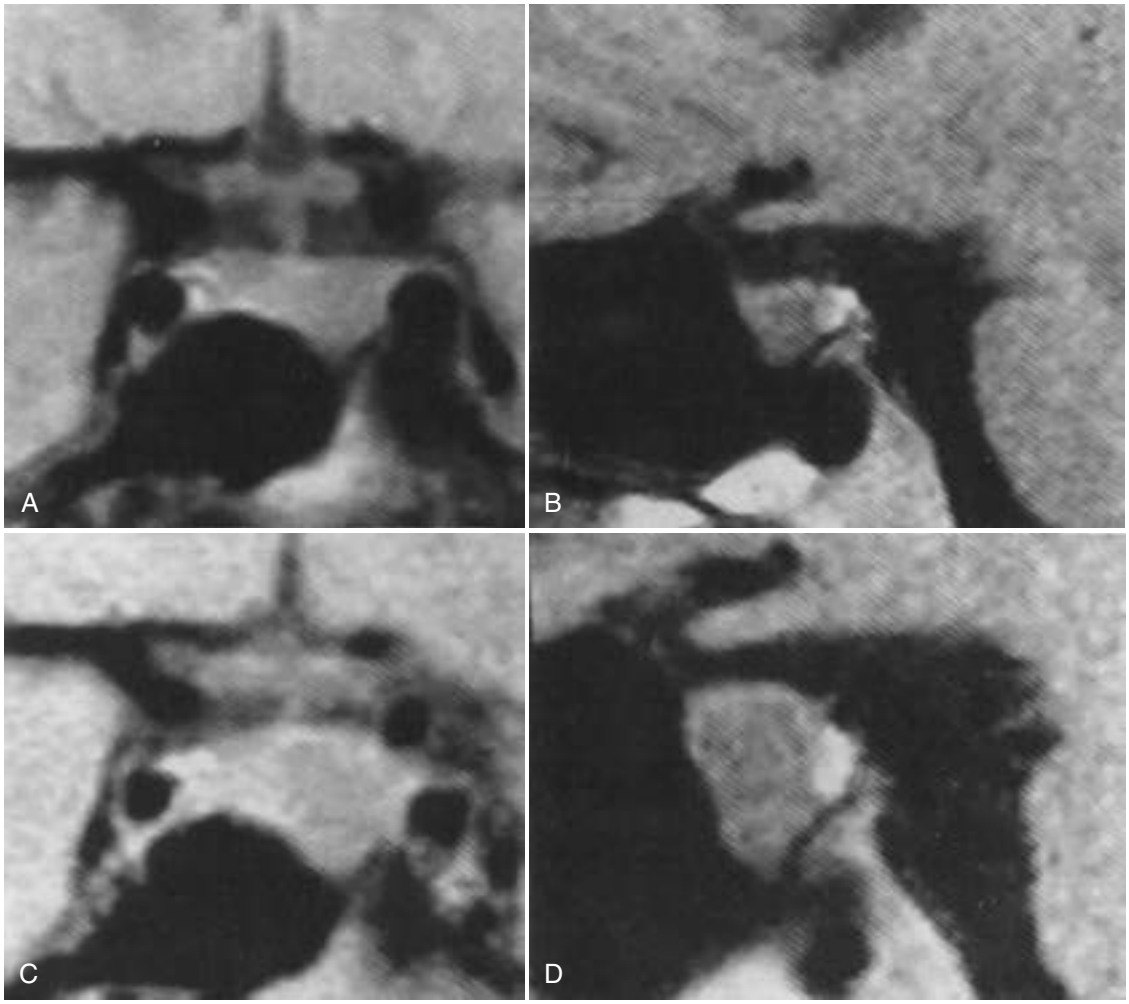

FIGURE 224-6. Magnetic resonance image showing enlargement of prolactinoma during pregnancy. ***Above,*** **Coronal (A) and sagittal (B) views show intrasellar prolactin-secreting macroadenoma prior to conception.** ***Below,*** **Coronal (C) and sagittal (D) views show enlargement of the prolactinoma at 7 months' gestation. (From Molitch ME. Medical treatment of prolactinomas.** ***Endocrinol Metab Clin North Am.*** **1999;28:143.)**

The primary effect of ACTH is to stimulate the adrenal gland to produce cortisol. It also stimulates secretion of adrenal androgens and mineralocorticoids, although production of mineralocorticoids is controlled primarily through non–ACTH-dependent mechanisms (i.e., the renin-angiotensin system) (Chapter 227). Consequently, mineralocorticoid function is preserved in ACTH deficiency, in contrast to primary adrenal insufficiency, which is characterized by loss of glucocorticoid and mineralocorticoid function. Long-term stimulation by ACTH causes adrenal hyperplasia and enlargement. On the other hand, ACTH deficiency leads to adrenal atrophy.

Hypothalamic corticotropin-releasing hormone (CRH) is the most important stimulator of ACTH secretion. Chronic stimulation by CRH causes corticotroph cell hyperplasia, which can be seen in cases of ectopic CRH production. Cortisol inhibits ACTH secretion, blunts the ACTH response to CRH, and inhibits CRH production. After prolonged glucocorticoid suppression of the hypothalamic-pituitary-adrenal axis, the amount of endogenous CRH secretion appears to be rate limiting and can require several months to recover.

Plasma ACTH is secreted in discrete pulses (10 to 80 pg/mL), so random measurements are of little value. Most clinical tests are therefore based on levels of cortisol or its metabolites, which tend to integrate the effects of ACTH. ACTH and cortisol secretion exhibit marked diurnal rhythms, being greatest at night several hours after the initiation of sleep. Cortisol levels are highest in the early morning and reach a nadir in the late afternoon and evening. Patients with Cushing's disease lose or exhibit a blunted diurnal rhythm of ACTH and cortisol secretion. ACTH secretion can be stimulated by a variety of different forms of stress, including psychological stimuli such as fright, anticipation of athletic competition, or surgery. Depression is associated with activation of the hypothalamic-pituitary-adrenal axis and impaired dexamethasone suppressibility. Hypoglycemia induces ACTH secretion through a central mechanism. The resulting increase in cortisol secretion represents one of several counter-regulatory mechanisms that increase glucose production. Insulin-induced hypoglycemia provides a mechanism for testing the integrity of the hypothalamic-pituitary-adrenal axis (see Table 224-2). Serious trauma and infection activate an array of cytokines that stimulate CRH and ACTH secretion. Because cortisol levels are often increased substantially in these circumstances, similar adjustments in cortisol replacement doses may be required in seriously ill patients with adrenal insufficiency.

Adrenocorticotropic Hormone Deficiency: Secondary Hypocortisolism

Secondary hypocortisolism causes symptoms of glucocorticoid deficiency, including nausea, vomiting, weakness, fatigue, fever, and hypotension. In addition to reduced levels of cortisol, abnormal laboratory test findings can include hyponatremia, hypoglycemia, and eosinophilia. Depending on its cause, the severity of cortisol deficiency in cases of secondary adrenal insufficiency is often not as marked as in primary adrenal insufficiency (Chapter 227). In addition, mineralocorticoid function is preserved in secondary adrenal deficiency. Consequently, the clinical manifestations of volume depletion are less pronounced, and hyperkalemia is not a feature of ACTH deficiency. Because ACTH levels are low, hyperpigmentation is not seen as in primary adrenal insufficiency. In women, reduced adrenal androgens can decrease libido and cause loss of axillary and pubic hair.

The most common cause of ACTH deficiency is treatment with exogenous glucocorticoids, which causes suppression of the hypothalamic-pituitary-adrenal axis. Sudden withdrawal of glucocorticoids or an increased requirement induced by the superimposition of severe illness can elicit symptoms of glucocorticoid deficiency. Congenital forms of ACTH deficiency are rare. When present, ACTH deficiency usually occurs in combination with the loss

of other pituitary hormones, although acquired, isolated ACTH deficiency does occur, particularly in women with lymphocytic hypophysitis.

ACTH reserve is most often evaluated using the insulin tolerance test. Caution should be exercised before inducing hypoglycemia in patients with suspected adrenal insufficiency. Insulin-induced hypoglycemia stimulates central responses to neuroglycopenia (Chapter 230) and mimics some stresses that activate ACTH secretion. ACTH stimulation tests using $ACTH_{1-24}$ (cosyntropin) can accurately evaluate primary adrenocortical insufficiency but may less accurately assess secondary adrenal insufficiency. A variation of the ACTH stimulation test using the low dose of 1 μg has been found to be useful for diagnosing secondary adrenal insufficiency in some studies.

Deficiency of ACTH is treated by replacement with glucocorticoids. Doses need to be individualized and are based largely on clinical criteria in which symptoms of glucocorticoid deficiency are balanced against features of glucocorticoid excess. Typical amounts of hydrocortisone are in the range of 15 to 20 mg/day in divided doses. Such doses are usually doubled in the event of mild to moderate illness. Patients should wear MedicAlert tags and be instructed in the warning signs of cortisol deficiency: nausea, vomiting, abdominal pain, low-grade fever, fatigue, and postural dizziness. Emergency injection kits of hydrocortisone are frequently provided for home use in the event that vomiting precludes taking oral steroids, or for severe sudden stress (e.g., a fracture). Stress doses of steroids should be used during times of illness. Current recommendations call for doses in the range of 50 to 75 mg every 8 hours for severe stress. Mineralocorticoid replacement is not required in patients with ACTH deficiency.

Cushing's Disease

PATHOBIOLOGY

Cushing's disease results from a pituitary adenoma that causes excess production of ACTH. It should be distinguished from a variety of other causes of Cushing's syndrome (glucocorticoid excess), which include adrenal causes (adenomas, carcinomas) of cortisol excess, ectopic production of ACTH and CRH, and physiologic states that result in overproduction of cortisol. Cushing's disease accounts for 60 to 70% of cases of Cushing's syndrome. Ten to 15% of pituitary tumors secrete ACTH. Cushing's disease occurs about eight times more often in women than in men.

Most ACTH-producing pituitary neoplasms, like other pituitary tumors, are monoclonal, implying a primary defect in corticotroph cells. In addition, there are rare cases of corticotroph hyperplasia causing Cushing's syndrome that are secondary to CRH production by either adjacent CRH-producing intrasellar gangliocytomas or ectopic CRH-producing cancers. Most (80 to 90%) of the ACTH-secreting tumors are microadenomas at the time of diagnosis. The clinical features of cortisol excess may allow detection of corticotroph adenomas before they have grown to a larger size. High levels of cortisol may also restrain tumor growth. ACTH-secreting macroadenomas may be locally invasive.

CLINICAL MANIFESTATIONS

The clinical features of Cushing's disease are caused by the effects of excess glucocorticoids and by the hypersecretion of ACTH and other POMC peptide products. The severity of the features of Cushing's disease varies greatly and appears to reflect not only the level of free cortisol but also the duration of the disease and perhaps the sensitivity to glucocorticoid action. In florid cases of Cushing's disease (Fig. 224-7), the constellation of symptoms and physical features is readily recognized. Early in the disease or in mild cases, it can be challenging to distinguish the clinical features of Cushing's disease from similar traits that are seen in the normal population. Clinical suspicion is of paramount importance. On the other hand, one must be discriminating and not formally evaluate everyone with obesity, hypertension, and glucose intolerance. Of the many features listed in Table 224-6, some are relatively specific for Cushing's disease. For example, the centripetal distribution of fat with the characteristic "buffalo hump," "moon facies," and deposition of fat in the supraclavicular area but not in the extremities is much more specific than generalized obesity. Striae that are wide (>1 cm) and purple reflect steroid-induced thinning of the dermis and can be distinguished from the more common "stretch marks." Numerous spontaneous ecchymoses also occur because of thinning of the skin and capillary fragility. Proximal muscle weakness represents another manifestation of glucocorticoid excess. Osteopenia and hypokalemia, when present, provide objective evidence consistent with ACTH excess. Hypokalemia results from the effects of ACTH on mineralocorticoid production but also from the ability of high levels of cortisol to saturate 11β-dehydrogenase, an enzyme in the kidney that inactivates cortisol. As a result, cortisol can "spill over" and act on mineralocorticoid receptors in the distal tubule. The hyperpigmentation associated with Cushing's disease is not as striking as that seen in Addison's disease or in ectopic ACTH syndrome, but in association with other findings, it should raise the suspicion of Cushing's disease and help distinguish it from adrenal causes of hypercortisolemia. Hirsutism and acne are caused by increased production of adrenal androgens and are more prominent in patients with Cushing's disease than in those with adrenal adenomas, in whom glucocorticoids tend to be the predominant product. Oligomenorrhea and amenorrhea probably have several causes, including androgen effects on the reproductive axis and glucocorticoid inhibition of GnRH, which may also account for diminished libido. Hypertension and glucose intolerance are caused by glucocorticoid excess. Immunosuppression, venous thrombo-

FIGURE 224-7. **(A) This 30 year old woman initially presented with a three year history of increasing facial hair, facial rounding, abdominal obesity, hypertension, diabetes, and oligomenorrhea. She had no muscle weakness or pigmented striae. (B) Following successful transsphenoidal resection of her ACTH-secreting microadenoma, she had a dramatic improvement in her clinical appearance with resolution of her diabetes and hypertension.**

TABLE 224-6 CLINICAL FEATURES OF CUSHING'S DISEASE

GENERAL
Obesity (centripetal distribution)
"Moon facies" and mild proptosis
Increased supraclavicular fat and "buffalo hump"
Hypertension
SKIN
Hyperpigmentation
Facial plethora
Hirsutism
Violaceous striae and thin skin
Capillary fragility and easy bruising
Acne
Edema
MUSCULOSKELETAL
Muscle weakness (proximal)
Osteoporosis and back pain
REPRODUCTIVE
Decreased libido
Oligomenorrhea and amenorrhea
NEUROPSYCHIATRIC
Depression
Irritability and emotional lability
Psychosis
METABOLIC
Hypokalemia and alkalosis
Hypercalciuria and renal stones
Glucose intolerance or diabetes mellitus
Impaired wound healing
Impaired resistance to infection
Granulocytosis and lymphopenia
TUMOR MASS EFFECTS
Headache
Visual field loss
Hypopituitarism

TABLE 224-7 TESTS USED IN THE DIFFERENTIAL DIAGNOSIS OF CUSHING'S SYNDROME*

ETIOLOGY	OVERNIGHT DEXAMETHASONE SUPPRESSION TEST	PLASMA ACTH	CORTICOTROPIN-RELEASING HORMONE STIMULATION OF ACTH	PETROSAL-TO-PERIPHERAL ACTH RATIO
Normal	Suppression	Normal	Normal	
Pituitary	No suppression	Normal or high	Normal or increased	>2
Ectopic	No suppression	High or normal	No response	<1.5
Adrenal	No suppression	Low	No response	

*Classic responses are indicated. Certain cases of ectopic adrenocorticotropic hormone (ACTH) production are suppressed by high-dose dexamethasone (not shown in this Table) or are stimulated by corticotropin-releasing hormone. In these cases, petrosal sinus sampling is the most reliable method for distinguishing pituitary and ectopic sources of ACTH.

embolism, opportunistic infections, and impaired wound healing can lead to considerable morbidity and mortality.[17] Neuropsychiatric symptoms, including depression, can be prominent effects of Cushing's disease. Suicide occurs with increased frequency in patients who receive no treatment for Cushing's disease.

DIAGNOSIS

The screening tests and differential diagnosis of Cushing's syndrome represent one of the greatest diagnostic challenges in endocrinology (Chapter 227). The first step is to determine whether a patient truly has cortisol excess. After confirmation of Cushing's syndrome, one must distinguish among (1) adrenal causes of cortisol excess, (2) pituitary causes of ACTH excess (Cushing's disease), (3) ectopic sources of ACTH, and (4) ectopic CRH (Table 224-7).

In screening for hypercortisolism, random cortisol levels are not useful because of diurnal variation of the hormone. The overnight dexamethasone test has been the most widely used screening test (see Table 224-5). A normal result of the dexamethasone test excludes Cushing's syndrome. It should be noted, however, that abnormal overnight dexamethasone suppression can be seen in up to 30% of hospitalized patients and in many patients with depression or during alcohol withdrawal. An elevated 24-hour urine free cortisol value provides an alternative or additional screening test for hypercortisolism. Often, two sequential specimens are collected because of day-to-day variations in hormone production. The sensitivity and specificity of urinary free cortisol measurements are greater than those of the overnight dexamethasone suppression test, particularly in hospitalized patients. A third test takes advantage of the observation that there is a loss of diurnal variation of cortisol levels in all forms of Cushing's syndrome. This test consists of finding elevation of a midnight cortisol level in the saliva. Kits are available for patients to obtain a late-night salivary cortisol sample. The sensitivity and specificity of late-night salivary cortisol measurements are very high, and this test now has become the one most commonly used by endocrinologists.

After demonstrating that cortisol excess is present, the next step is to determine the source of excess ACTH or cortisol. This is done by measuring an ACTH level along with a cortisol level, with primary adrenal disease causing suppressed ACTH levels. The classic approach of performing a low-dose, followed by a high-dose, dexamethasone suppression test has been largely abandoned (see Tables 224-5 and 224-7). The high-dose dexamethasone test is one of several means to to discriminate between pituitary and ectopic causes of ACTH-dependent Cushing's syndrome (see Table 224-7). Pituitary and ectopic causes of Cushing's disease are both ACTH dependent but respond differently to high-dose dexamethasone. Pituitary adenomas have an altered set point for glucocorticoid inhibition but retain a partial ability to respond to high-dose dexamethasone. The exact criteria for dexamethasone suppression in the high-dose test are debated. In most cases of ACTH-producing pituitary adenomas, urinary free cortisol is suppressed below 90% of baseline during the high-dose dexamethasone test.

The ectopic ACTH syndrome should be suspected in patients with known malignancies, particularly small-cell carcinoma of the lung; bronchial, thymic, or gastrointestinal carcinoids; islet cell tumors; and medullary carcinoma of the thyroid, among others. Plasma ACTH levels are often very high (>200 pg/mL) and can be associated with hyperpigmentation. Clinical features of Cushing's syndrome may be altered by the rapid onset of extreme hypercortisolemia coincident with elements of tumor cachexia. Pronounced weakness, fluid retention, glucose intolerance, hypokalemia, and poor skin integrity are often seen. Ectopic ACTH syndrome is readily recognized in its classic form. However, a subset of tumors, particularly carcinoids (Chapter 232), exhibit dexamethasone suppression that is similar to that seen with pituitary adenomas. When suspected, carcinoids can sometimes be detected by CT or MRI, but many are too small to be seen even with these techniques. Because of these exceptions to the high-dose dexamethasone test, a variety of procedures have been devised in an attempt to further distinguish ectopic and pituitary dependent sources of ACTH. CRH testing may also prove useful, with pituitary tumors exhibiting an increase in ACTH and tumors making ACTH ectopically having little or no response.

In recent years, inferior petrosal sinus sampling has been used to distinguish pituitary and ectopic sources of ACTH when the source of ACTH is not obvious based on the clinical circumstances, biochemical evaluation, and imaging studies. This test requires an experienced radiologist for safe and effective catheterization of the petrosal sinuses. Blood samples are taken simultaneously from the left and right petrosal sinuses and from the periphery before and after CRH stimulation. In the case of ACTH-producing pituitary adenomas, there is a gradient in ACTH levels between the central and peripheral blood specimens. When clinical and biochemical studies suggest the presence of a pituitary adenoma, pituitary imaging should be performed using CT or MRI. Most ACTH-secreting pituitary adenomas are small, and scans are normal in more than half of patients.

TREATMENT

Rx

The efficacy of transsphenoidal surgery for Cushing's disease is greatly aided by making the correct diagnosis preoperatively. In experienced hands, surgical cures of ACTH-producing microadenomas occur in 75 to 90% of patients undergoing a first operation. As in other pituitary tumors, complete remissions with macroadenomas are much less common. In the event of surgical remission or cure, postoperative hypocortisolism is to be expected because of suppression of the hypothalamic-pituitary axis. After coverage for steroid withdrawal in the postoperative period, cortisol replacement should gradually be decreased to allow recovery of the hypothalamic-pituitary-adrenal axis; recovery may take up to 1 year.

If transsphenoidal surgery is unsuccessful, reoperation may be indicated and can result in remission in up to 50% of patients[18]; in this circumstance, consideration should be given to performing a total hypophysectomy at reoperation. If transsphenoidal surgery cannot be performed or has failed, alternative forms of therapy should be used to prevent the long-term consequences of hypercortisolism.[19] Pituitary irradiation is often the second line of treatment for Cushing's disease. It is more efficacious in children and in younger patients, but even in older adults, remissions can be achieved in about 50% within 2 years. To prevent the continued ravages of hypercortisolism during this period, however, concomitant medical therapy (see later) is usually given. Bilateral adrenalectomy represents another alternative for patients with severe hypercortisolism after transsphenoidal surgery. It rapidly and effectively lowers cortisol levels but is associated with relatively high morbidity and mortality rates (as high as 5%) because of the associated metabolic and immune system alterations caused by hypercortisolism. The morbidity has been reduced in recent years by introduction of the laparoscopic approach. After adrenalectomy, patients must be maintained on glucocorticoids and mineralocorticoids and are at risk for the development of Nelson syndrome (see later).

Medical therapy for Cushing's disease has its primary role in preparation for surgery or control of hypercortisolism following unsuccessful surgery. It may also be used during the interval when radiation therapy is taking effect (Fig. 224-8). The antifungal agent ketoconazole is effective in decreasing glucocorticoid biosynthesis and also inhibits ACTH secretion, so it has been the most common medical therapy used; it has hepatotoxicity and is not actually approved for use for Cushing's syndrome. Other drugs that interfere with cortisol synthesis (e.g., mitotane, etomidate, metyrapone) have been used less commonly. In a few small studies, cabergoline has also been shown to cause a normalization of cortisol levels in about one third of patients with Cushing's disease, although it has not been approved for this indication. Recently, two new drugs have been approved for the treatment of Cushing's disease. Mifepristone is a progesterone receptor blocker that also is a glucocorticoid receptor blocker and is highly effective in improving clinical signs and symptoms of Cushing's syndrome; because of its mechanism of action, with

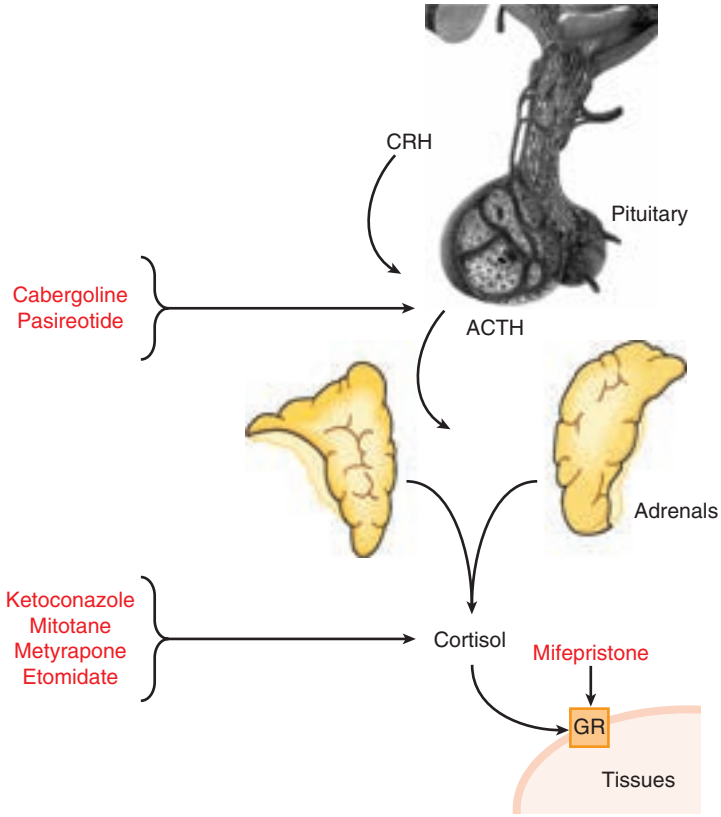

FIGURE 224-8. Medical therapies for Cushing's disease. Cabergoline and pasireotide act to decrease ACTH secretion from the corticotroph tumor. Several drugs (ketoconazole, metyrapone, mitotane, etomidate) work at the adrenal level to decrease cortisol synthesis. Mitotane works at the adrenal level by blocking cortisol action at the glucocorticoid receptor. ACTH = adrenocorticotropic hormone; CRH = corticotropin-releasing hormone; GR = glucocorticoid receptor. (Adapted from Petersenn S. Medical management of Cushing's disease. In: Swearingen B, Biller BMK (eds) Endocrine updates vol. 31: Cushing's disease. 2011, Springer, New York, pp 167–182, Figure 1. As adapted from Petersenn S, Endocrine Updates, 2011, Springer Science+Business Media B.V.)

treatment, cortisol and ACTH levels actually rise while the clinical signs improve. The higher levels of cortisol may "spill over" to the mineralocorticoid receptor, causing a blood pressure rise and hypokalemia. Symptoms of glucocorticoid withdrawal and even adrenal insufficiency may occur. Blockade of the progesterone receptor by mifepristone may cause menorrhagia. Pasireotide is a somatostatin receptor analog that has much greater activity on corticotroph adenomas than other somatostatin analogs because it has additional activity at the somatostatin-5 receptor. A large multicenter study has recently shown clinical and biochemical efficacy in patients with Cushing's disease[■]; the major adverse effect is hyperglycemia and a worsening of preexisting diabetes mellitus. This worsening of glucose levels is likely due to a reduction in insulin and glucagon-like peptide (GLP)-1 levels.

Nelson Syndrome

Nelson syndrome was initially described as the appearance of a pituitary adenoma after bilateral adrenalectomy. In addition to an enlarging pituitary mass, the syndrome is characterized by very high ACTH levels and hyperpigmentation. It is caused by a preexisting ACTH-producing tumor that grows in the absence of feedback inhibition by high levels of glucocorticoids. The incidence of clinically significant Nelson syndrome after adrenalectomy for Cushing's disease varies from 10 to 50% in different series.[20] Patients with Cushing's disease who have undergone adrenalectomy should be followed with imaging studies and plasma ACTH levels, because tumors that cause Nelson syndrome can be very aggressive. When there is evidence of mass effects or rapid growth, transsphenoidal surgery should be performed. Postoperative irradiation may provide additional benefit, although it appears to be less efficacious than in other ACTH-producing adenomas. In theory, Nelson syndrome could also occur in patients being treated with mifepristone, and such patients should also be monitored with periodic pituitary MRI scans, although short-term follow-up data are reassuring at this point.

GONADOTROPINS (FOLLICLE-STIMULATING HORMONE AND LUTEINIZING HORMONE)

The pituitary glycoprotein hormones include FSH, LH, and TSH. Chorionic gonadotropin, which is structurally very similar to LH, is made in the placenta. Each of the glycoprotein hormones has a specific β-subunit that forms a noncovalently bound dimer with the common α-subunit. The α- and β-subunits each undergo glycosylation, which is important for correct hormone folding, intracellular transport, and secretion. Glycosylation is also required for biologic activity, presumably because of effects on the tertiary structure of the hormones. The gonadotropins are involved in sexual differentiation, sex steroid production, and gametogenesis. The regulation and physiologic roles of gonadotropins are quite different in males and females.

In males, receptors for FSH are located on Sertoli cells and seminiferous tubules, whereas LH receptors are located on Leydig cells in the testis. LH stimulates androgen production by the Leydig cells. FSH is involved primarily in sperm maturation in the seminiferous tubules. Thus, FSH and LH act together to induce spermatogenesis (Chapter 234).

In females, ovarian FSH receptors are located on granulosa cells, where they induce enzymes involved in estrogen biosynthesis. LH receptors are located predominantly on thecal cells in the ovary and stimulate the production of ovarian androgens and steroid precursors that are transported to granulosa cells for aromatization to estrogens. The pattern of FSH and LH secretion during the menstrual cycle results in follicular recruitment and maturation (largely FSH mediated), followed by ovulation (largely LH mediated) and steroid production by the corpus luteum (Chapter 235).

Gonadotropin secretion is regulated primarily by the hypothalamic decapeptide GnRH. The gonadotroph cell is exquisitely sensitive to the pattern of GnRH stimulation. Continuous, rather than pulsatile, exposure to GnRH causes gonadotroph desensitization and suppression of LH and FSH. Gonadotroph sensitivity to GnRH is modulated by sex steroids and probably other hypothalamic peptides such as neuropeptide Y. Increased GnRH secretion, in combination with a higher density of GnRH receptors and rising estradiol concentrations, accounts in part for the dramatic release of gonadotropins that induces ovulation.

Hypogonadotropic Hypogonadism

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Clinical features of hypogonadotropic hypogonadism in women are primarily due to estrogen deficiency and include breast atrophy, vaginal dryness, and diminished libido. Hot flashes are uncommon, in contrast to postmenopausal estrogen deficiency. In premenopausal women, normal menstrual cycles provide evidence for an intact hypothalamic-pituitary-gonadal axis. LH and FSH levels should be increased in postmenopausal women, and normal levels may indicate deficiency. Hypogonadism in men causes decreased libido and sexual function. In men, low testosterone without elevation of LH and FSH is consistent with impaired hypothalamic-pituitary reserve. GnRH stimulation can distinguish hypothalamic and pituitary deficiency but may require multiple injections to prime the pituitary.

A congenital form of hypogonadotropic hypogonadism is caused by deficiency of GnRH, which in turn causes deficiencies of LH and FSH. When associated with anosmia (absent sense of smell), the condition is referred to as Kallmann's syndrome (Chapter 223).

Secondary hypogonadotropic hypogonadism is relatively common. In most cases, it is reversible and is caused by weight loss, anorexia nervosa, stress, heavy exercise, or severe illness. Reversible forms of secondary hypogonadotropic hypogonadism are caused by GnRH deficiency and are more common in women than men.

A variety of pathologic conditions can cause secondary hypogonadotropic hypogonadism, often in association with deficiencies of other pituitary hormones (see Table 224-1). These include hypothalamic lesions and central nervous system irradiation. Pituitary tumors can suppress gonadotropins because of stalk compression and disruption of pulsatile GnRH input, as well as by direct destruction of normal pituitary tissue. Hyperprolactinemia can suppress GnRH and lead to reduced gonadotropin levels.

In contrast to the aforementioned causes of hypogonadotropic hypogonadism that result from GnRH deficiency, primary deficiencies of LH and FSH are uncommon. An acquired form of isolated gonadotropin deficiency is rarely encountered and may have an autoimmune basis. Mutations in the *LHβ* or *FSHβ* genes have been described in case reports and cause selective loss of individual gonadotropins. Inactivating mutations in the GnRH

In women, the pattern of GnRH pulse frequency varies across the menstrual cycle (Chapter 235). The combination of GnRH stimulation in conjunction with ovarian feedback regulation results in a complex orchestration of positive and negative hormonal signals that converge at the gonadotroph to regulate LH and FSH secretion. The typical 28-day menstrual cycle is divided into follicular and luteal phases that are separated by ovulation on day 14. Unlike chronic exposure to low concentrations of estrogens, which exert negative feedback regulation and inhibit GnRH, the increasing concentration of estrogen before the LH surge exerts positive feedback regulation that results in increased GnRH pulse frequency. Increased GnRH in combination with increased gonadotroph sensitivity to GnRH results in the LH-FSH surge. During the luteal phase, the gonadotropin pulse frequency is reduced. In addition to feedback regulation by steroids, ovarian peptides such as inhibin also play a role in control of the reproductive axis. Inhibin causes selective suppression of FSH without affecting LH secretion. A homodimer of inhibin β-subunits, referred to as activin, has opposite actions and selectively stimulates FSH, but its predominant physiologic action is to increase ovarian granulosa cell responsivity to FSH. Circulating inhibin provides one of the negative feedback inputs that leads to FSH suppression as the follicle develops. The perimenopause is characterized by a gradual cessation of ovarian function. After several years of menstrual cycles that are sometimes anovulatory or irregular, menses cease, thereby defining the menopause (Chapter 240). At menopause, the decline in estrogen and progesterone causes loss of feedback inhibition and a marked increase in LH and FSH levels.

In males, the regulation of the hypothalamic-pituitary-gonadal axis is relatively constant. Testosterone inhibits the hypothalamic-pituitary axis, although its actions are mediated in part by aromatization to estrogens. Most of the inhibition by gonadal steroids occurs at the hypothalamic level. In contrast to menopause in women, there is no analogous abrupt change in hormone levels in men. There is, however, a gradual decline in testosterone levels associated with an increase in LH and FSH with aging.

receptor and the LH and FSH receptors causing hypogonadotropic hypogonadism have also been reported.

TREATMENT

In premenopausal women, preparations of estrogen and progestins should be used for hormonal replacement and to allow cyclical growth of the endometrium. Pulsatile GnRH (for GnRH-deficient patients) has been given to induce ovulation and fertility but is not commonly used at present. Gonadotropin injections are more commonly used when fertility is desired. Testosterone can be replaced in men, using intramuscular injections that are given at 2- to 4-week intervals. Doses and the intervals between injections should be adjusted on an individual basis using libido and testosterone levels before the next injection as a guide. Oral preparations of androgens should be avoided because of hepatotoxicity. Transdermal patch and gel preparations are also available and maintain more stable testosterone levels but are more expensive. With gels, care has to be used to prevent exposure of the partner or children. Although induction of spermatogenesis can be achieved using pulsatile GnRH (for GnRH-deficient patients), injections of gonadotropins are more commonly used.

Pulsatile GnRH has been used to induce puberty and fertility in both males and females with Kallmann's syndrome and other forms of GnRH deficiency, but more commonly, injections of gonadotropins are used.

Secondary hypogonadotropic hypogonadism is ideally treated by correcting the underlying cause.[21] Many women have a discrete threshold for weight or exercise level that will cause loss of menstrual periods. When it is not possible to correct the underlying abnormality, hormonal replacement can be used in women for protection against osteopenia and to cycle the endometrium. Permanent idiopathic hypogonadotropic hypogonadism can also occur in both sexes and will require hormone replacement.

Follicle-Stimulating Hormone– and Luteinizing Hormone–Producing Tumors

PATHOBIOLOGY

The majority (70 to 80%) of pituitary tumors classified previously as nonfunctioning adenomas can be shown to produce low levels of intact glycoprotein hormones or their uncombined α- or β-subunits. Biosynthetic defects in the tumor cells account for relatively inefficient hormone secretion as well as the propensity to produce uncombined subunits. FSH is produced more commonly than LH. Elevated levels of free α-subunits are noted more often than increased free β-subunits.

CLINICAL MANIFESTATIONS

Gonadotropin-producing tumors are somewhat more common in men than women and increase in prevalence with age. FSH- and LH-producing tumors do not usually cause a characteristic hormone excess syndrome. The tumors, typically large macroadenomas, present as clinically nonfunctioning tumors with symptoms and signs related to local mass effects. Visual field loss is found in more than 70% of patients. Many are detected incidentally by CT and MRI performed for unrelated indications. Symptoms of hypogonadism with loss of libido are also common. Men with predominantly FSH-secreting tumors may present with testicular enlargement from hypertrophy of the seminiferous tubules but may also be hypogonadal due to low levels of testosterone. These patients must be distinguished from those with primary hypogonadism due to testicular dysfunction. Tumors that primarily secrete LH are rare but can cause increased testosterone levels. Premenopausal women with gonadotropin-producing tumors may experience menstrual irregularity or secondary hypogonadism. Postmenopausal women often show reduced gonadotropin levels because the mass effects of the gonadotropin-producing tumors cause stalk compression, impairing GnRH stimulation of gonadotropins from normal pituitary cells.

DIAGNOSIS

Because of the absence of a clinical syndrome in most patients, almost all gonadotropin-producing pituitary tumors are diagnosed by postoperative immunohistochemistry, because they had presented with mass effects. There is no particular clinical benefit to distinguish whether a nonfunctioning adenoma is truly a gonadotroph adenoma. Some patients can have moderately elevated PRL levels that are caused by tumor mass effects. It is important to distinguish this group from patients with true prolactinomas. As noted earlier, many women, including those in the postmenopausal group, have paradoxically low gonadotropin levels. Thus, the absence of elevated gonadotropins does not exclude the diagnosis of a gonadotropin-producing tumor.

TREATMENT

Because the major symptoms of the gonadotropin-producing tumors are due to extrasellar extension and local mass effects, the main aim of treatment is reduction in tumor size. Complete or partial reversal of visual field defects and hypopituitarism can be accomplished by surgery unless these conditions have been of long standing. Transsphenoidal surgery is rarely curative, however, because of the large size of the tumors. Patients with significant residual tumor may benefit from radiation therapy. Because most tumors are slow growing, when no tumor is visible postoperatively by MRI, the patient may be followed with yearly monitoring for tumor recurrence, using visual fields and CT or MRI. If tumor markers such as free α- or β-subunit levels are available, they can also be used to monitor tumor function. When follow-up studies show tumor regrowth, repeat surgery, radiation therapy, or both are indicated. Medical therapy with dopamine agonists and somatostatin analogues has been successful in only a minority of patients.

THYROID-STIMULATING HORMONE

Like the other glycoprotein hormones, TSH is a heterodimer composed of the common α-subunit and the unique TSH β-subunit. Normal levels of TSH range from 0.4 to 4.0 μU/mL. The detection limit for current TSH assays is less than 0.01 μU/mL, allowing measurement of suppressed TSH levels in patients with hyperthyroidism. TSH controls thyroid hormone (T_4 and T_3) synthesis and secretion from the thyroid gland. Hypothalamic TRH stimulates TSH synthesis and secretion. Somatostatin and dopamine can inhibit TSH secretion, but their roles in normal physiology have not been clearly elucidated. Thyroid hormones have an inhibitory effect on the production of TRH and TSH and constitute a powerful negative feedback loop acting at both the hypothalamic and pituitary levels. Secretion of TSH is pulsatile, but the amplitude of the pulses is relatively small and does not create the difficulties in measurement of TSH that are encountered with measurements of other pituitary hormones. Because of the integrated nature of the hypothalamic-pituitary-thyroid axis, thyroid function tests are best interpreted when concentrations of TSH, free T_4, and free T_3 levels are known. Except in conditions of secondary hypothyroidism or TSH-secreting pituitary tumors (see later), TSH levels provide an excellent screening test for thyroid dysfunction.

Central Hypothyroidism

Central forms of hypothyroidism are due to loss of either TSH or TRH.[22] Three different types of congenital TSH deficiency are caused by genetic mutations. One type involves mutations in the *TSHβ* gene. A second involves mutations in *PIT1*, which causes combined deficiencies of GH, PRL, and TSH (see earlier). A third involves a mutation in the gene for TRH. Acquired central forms of hypothyroidism are usually associated with other pituitary hormone deficiencies, and usually there is no goiter because of low TSH levels.

Tests for TSH deficiency are best performed by analyzing free T_4 levels in combination with TSH. Low free T_4 without elevated TSH is consistent with central hypothyroidism. In some patients with hypothalamic disease, the TSH level is partially elevated in the presence of low levels of free T_4, but the bioactivity of the TSH is reduced. Central forms of hypothyroidism must be distinguished from the sick-euthyroid condition (Chapter 226). Laboratory tests in the sick-euthyroid syndrome progress through several phases but can include prolonged periods when both TSH and free thyroid hormone levels are low. It can be very difficult in these patients to exclude central hypothyroidism unequivocally. In addition to the clinical setting in which thyroid function tests are measured, the presence of normal thyroid function tests before the illness and the absence of known hypothalamic or pituitary disease make true central hypothyroidism unlikely. Increased levels of reverse T_3 are suggestive of sick-euthyroidism, and free T_4 and T_3 may be in the normal or low-normal range in sick-euthyroid patients. When TSH deficiency is documented, thyroid hormone is replaced using daily doses of L-thyroxine (0.05 to 0.15 mg/day). Because TSH cannot be used as an end point, one monitors the patient clinically and with serum levels of free T_4 and T_3.

Thyroid-Stimulating Hormone–Secreting Tumors

TSH-secreting tumors are rare and account for between 1 and 3% of pituitary tumors. A recent analysis from Sweden showed that the prevalence was only

2.8 per 1 million.[23] As many as 45% of TSH-producing tumors are plurihormonal. GH and PRL are co-secreted most often, perhaps reflecting the common cellular lineage for thyrotrophs, somatotrophs, and lactotrophs. Long-standing severe hypothyroidism can cause thyrotroph hyperplasia and pituitary enlargement. These hyperplastic masses regress with thyroid hormone replacement therapy, however. Most true TSH-producing tumors are relatively autonomous and respond weakly, if at all, to TRH stimulation or thyroid hormone suppression.

TSH-secreting tumors are usually macroadenomas by the time a diagnosis has been made. Consequently, many patients exhibit mass effects of the tumor, as well as hyperthyroidism. The clinical features of TSH-secreting tumors resemble those of Graves' disease, except that features of autoimmunity (e.g., ophthalmopathy) are absent. Circulating levels of T_4 and T_3 range widely but can be elevated as much as two- to three-fold. Diffuse goiter is present in most patients with TSH-producing tumors, and the 24-hour uptake of radioiodine is elevated.

Because feedback inhibition of TSH is impaired in TSH-producing tumors, TSH levels are inappropriately elevated in the presence of high levels of T_4 and T_3. TSH levels produced by tumors range from the low-normal range to as high as 500 μU/mL, but most levels are minimally elevated. Most TSH-producing tumors (>80%) secrete excess free α-subunit, and its assessment can be very useful in confirming the diagnosis. Thus, the diagnosis can usually be made by demonstrating that a hyperthyroid patient has a detectable serum TSH level associated with excess secretion of the free α-subunit. The finding of a mass lesion on CT or MRI confirms the diagnosis. Several other causes of inappropriate TSH secretion should be considered, including resistance to thyroid hormone and familial dysalbuminemic hyperthyroxinemia and other disorders that alter serum thyroid hormone binding proteins.

TREATMENT

Rx

The goals of therapy are to treat the underlying TSH-secreting tumor and to correct the hyperthyroidism. Transsphenoidal surgery alone is rarely curative because of the large size of most tumors, but it can alleviate mass effects and lower TSH levels. As in other large pituitary tumors, adjunctive irradiation may be required to control tumor growth. Somatostatin analogues have been used as adjunctive medical therapy, and they decrease TSH and α-subunit levels in about 80% of patients with TSH-secreting tumors, but consistent effects on tumor growth have not been demonstrated. Hyperthyroidism caused by TSH-secreting tumors can also be treated using antithyroid drugs or radioiodine.

CLINICALLY NONFUNCTIONING PITUITARY TUMORS

Most clinically nonfunctioning adenomas can be shown to produce low levels of the free α-subunit, free β-subunits of FSH and LH, and intact FSH and LH when analyzed by immunocytochemistry or messenger RNA expression. A smaller fraction can be shown to produce low levels of other pituitary hormones, particularly ACTH or GH, that escaped detection based on routine endocrine testing.[24] Even with detailed analyses of hormone production, a subset (10 to 20%) of nonfunctioning adenomas does not appear to produce any of the known pituitary hormones.

The clinical features and management of nonfunctioning tumors are similar to those for gonadotropin-producing tumors. The major signs and symptoms result from tumor mass effects that cause visual field defects, headache and other neurologic symptoms, and hypopituitarism. Transsphenoidal surgery is the primary mode of treatment, with a goal of debulking the tumor to relieve mass effects. Because there are no serum tumor markers, patients must be followed by CT or MRI in conjunction with visual field tests.

Some pituitary tumors are discovered as incidental findings on CT or MRI scans that were done for other reasons.[25] Such tumors should be screened for hormone oversecretion with measurement of PRL, IGF-I, and a midnight salivary cortisol or an overnight dexamethasone suppression test, but most will be found to be nonfunctioning. If the tumor abuts the optic chiasm, a formal visual field examination should be performed. Over several years, about 10% of incidental microadenomas and 20% of macroadenomas enlarge. Indications for surgery include compression of the optic chiasm, with or without visual field defects and significant tumor enlargement. Hypopituitarism is also a relative indication for surgery (Fig. e224-1). In the absence of these indications for surgery, it is reasonable to follow such patients with MRI scans to look for size change at yearly intervals initially and then at less frequent intervals.

Grade A References

A1. Barake M, Klibanski A, Tritos NA. Effects of recombinant human growth hormone therapy on bone mineral density in adults with growth hormone deficiency: a meta-analysis. *J Clin Endocrinol Metab*. 2014;99:852-860.

A2. Gadelha MR, Bronstein MD, Brue T, et al. Pasireotide versus continued treatment with octreotide or lanreotide in patients with inadequately controlled acromegaly (PAOLA): a randomised, phase 3 trial. *Lancet Diabetes Endocrinol*. 2014;2:875-884.

A3. Colao A, Petersenn S, Newell-Price J, et al. A 12-month phase 3 study of pasireotide in Cushing's disease. *N Engl J Med*. 2012;366:914-924.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

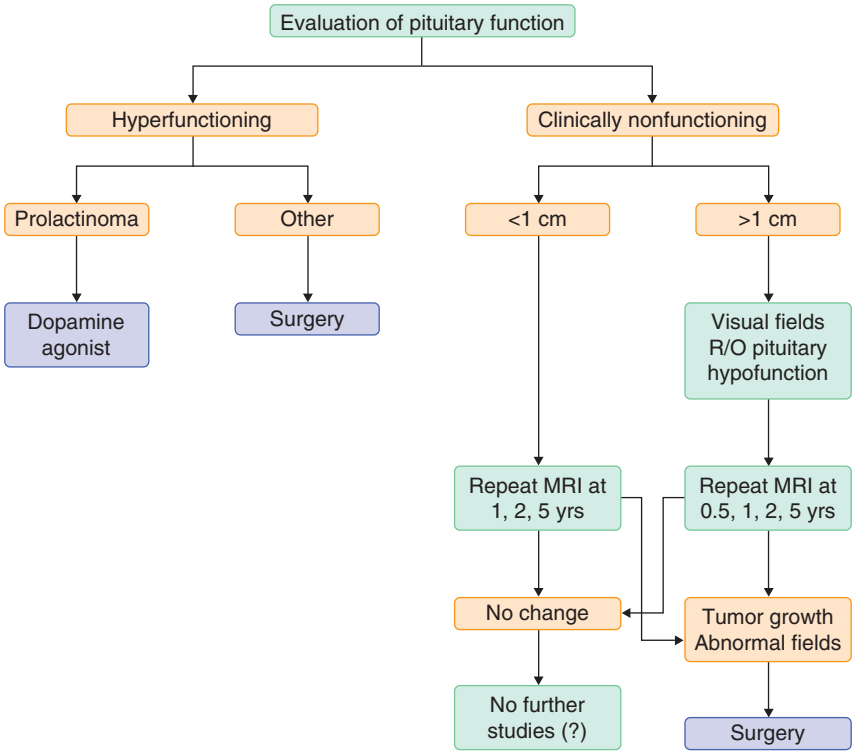

E-FIGURE 224-1. Flow diagram indicating the approach to the patient found to have a pituitary incidentaloma. The first step is to evaluate patients for pituitary hyperfunction and then treat those found to be hyperfunctioning. Of patients with tumors that are clinically nonfunctioning, those with macroadenomas are evaluated further for evidence of chiasmal compression and hypopituitarism. Scans are then repeated at progressively longer intervals to assess for enlargement of the tumors. (Reproduced from Molitch ME. Non-secreting tumors and pituitary incidentalomas. *Endocrinol Metab Clin North Am.* 2008;37:151-171.)

GENERAL REFERENCES

1. Cohen LE. Genetic disorders of the pituitary. *Curr Opin Endocrinol Diabetes Obes*. 2012;19:33-39.
2. Mete O, Asa SL. Clinicopathological correlations in pituitary adenomas. *Brain Pathol*. 2012;22:443-453.
3. Hess CP, Dillon WP. Imaging the pituitary and parasellar region. *Neurosurg Clin North Am*. 2012;23:529-542.
4. Caturegli P, Iwama S. From Japan with love: another tessera in the hypophysitis mosaic. *J Clin Endocrinol Metab*. 2013;98:1865-1868.
5. Burman P, Mattsson AF, Johannsson G, et al. Deaths among adult patients with hypopituitarism: hypocortisolism during acute stress, and de novo malignant brain tumors contribute to an increased mortality. *J Clin Endocrinol Metab*. 2013;98:1466-1473.
6. Molitch ME, Clemmons DR, Malozowski S, et al. Evaluation and treatment of adult growth hormone deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2011;96:1587-1609.
7. Melmed S. Pathogenesis of pituitary tumors. *Nat Rev Endocrinol*. 2011;7:257-266.
8. Vasilev V, Daly AF, Petrossians P, et al. Familial pituitary tumor syndromes. *Endocr Pract*. 2011;17(suppl 3):41-46.
9. Beckers A. Means, motives and opportunity: SDH mutations are suspects in pituitary tumors. *J Clin Endocrinol Metab*. 2013;98:2274-2276.
10. Fraser CL, Biousse V, Newman NJ. Visual outcomes after treatment of pituitary adenomas. *Neurosurg Clin North Am*. 2012;23:607-620.
11. Melmed S, Casaneuva FF, Hoffman AR, et al. Diagnosis and treatment of hyperprolactinemia. An Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2011;96:273-288.
12. Devin JK. Hypopituitarism and central diabetes insipidus. Perioperative diagnosis and management. *Neurosurg Clin North Am*. 2012;23:679-689.
13. Swearingen B. Update on pituitary surgery. *J Clin Endocrinol Metab*. 2012;97:1071-1081.
14. Sheehan JP, Xu Z, Lobo MJ. External beam radiation therapy and stereotactic radiosurgery for pituitary adenomas. *Neurosurg Clin North Am*. 2012;23:571-586.
15. Katznelson L, Laws E, Molitch M, et al. Diagnosis and treatment of acromegaly. An Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2014;99:3933-3951.
16. Rogers A, Karavitaki N, Wass JA. Diagnosis and management of prolactinomas and non-functioning pituitary adenomas. *BMJ*. 2014;349:g5390.
17. Dekkers OM, Horváth-Puhó E, Jørgensen JOL, et al. Multisystem morbidity and mortality in Cushing's syndrome: a cohort study. *J Clin Endocrinol Metab*. 2013;98:2277-2284.
18. Bertagna X, Guignat L. Approach to the Cushing's disease patient with persistent/recurrent hypercortisolism after pituitary surgery. *J Clin Endocrinol Metab*. 2013;98:1307-1318.
19. Feelders RA, Hofland LJ. Medical treatment of Cushing's disease. *J Clin Endocrinol Metab*. 2013;98:425-438.
20. Barber TM, Adams E, Wass JA. Nelson syndrome: definition and management. *Handb Clin Neurol*. 2014;124:327-337.
21. Silveira LFG, Latronico AC. Approach to the patient with hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*. 2013;98:1781-1788.
22. Persani L. Clinical Review: central hypothyroidism: pathogenic, diagnostic, and therapeutic challenges. *J Clin Endocrinol Metab*. 2012;97:3068-3078.
23. Önnestam L, Berinder K, Burman P, et al. National incidence and prevalence of TSH-secreting pituitary adenomas in Sweden. *J Clin Endocrinol Metab*. 2013;98:626-635.
24. Cooper O, Melmed S. Subclinical hyperfunctioning pituitary adenomas: the silent tumors. *Best Pract Clin Endocrinol Metab*. 2012;26:447-460.
25. Molitch ME. Management of incidentally found nonfunctional pituitary tumors. *Neurosurg Clin North Am*. 2012;23:53-554.

REVIEW QUESTIONS

1. A 25-year-old woman has had amenorrhea and galactorrhea for 2 years and is found to have hyperprolactinemia. Her PRL level is 513 ng/mL (normal 2.5 to 23.6 ng/mL) and magnetic resonance imaging showed a 1.3-cm macroadenoma. The rest of her evaluation is normal, with the exception of a serum calcium of 11.4 mg/dL. She recalls that one of her cousins also has a pituitary tumor. Which of the following genes should be analyzed for a mutation?

A. *PROP1*
B. *Pit1*
C. *Menin*
D. *Ret* proto-oncogene
E. *Ras*

Answer: C Menin is encoded by the gene mutated in multiple endocrine neoplasia *(MEN1)*, characterized by multiple endocrine tumors of the parathyroid glands, pancreatic islets, and anterior pituitary, especially prolactinomas. For incorrect answers, see introductory section on "Anatomy and Embryology."

2. Which of the following medications used to treat pituitary tumor syndromes acts by blocking a hormone receptor?

A. Pasireotide
B. Cabergoline
C. Ketoconazole
D. Mifepristone
E. Bromocriptine

Answer: D See Figure 224-8 and its legend.

3. A 67-year-old man is found to have enlarging hands and feet and has been referred by his dentist because of prognathism. To determine whether he has acromegaly, which of the following tests should be carried out?

A. Pituitary magnetic resonance imaging
B. Overnight 1-mg dexamethasone suppression test
C. Insulin-induced hypoglycemia stimulation test
D. Measurement of insulin-like growth factor (IGF)-I
E. Inferior petrosal sinus sampling with GHRH stimulation

Answer: D See section on "Diagnosis" under "Growth Hormone Excess: Acromegaly and Gigantism."

4. Magnetic resonance imaging (MRI) on a 73-year-old woman experiencing dizziness reveals an incidental 3-mm lesion in her pituitary that is compatible with a pituitary adenoma. She has otherwise been well. Testing for hormone oversecretion is negative. What should be the next step in management?

A. Refer her to an experienced neurosurgeon.
B. Repeat the MRI in 1 year to look for size change.
C. Refer her to ophthalmology for a visual field examination.
D. Begin bromocriptine treatment.
E. Refer her for a preoperative cardiac stress test.

Answer: B See section on "Clinically Nonfunctioning Pituitary Tumors" at the end of the chapter and also e-Figure 224-1.

225

POSTERIOR PITUITARY

JOSEPH G. VERBALIS

ANATOMY AND HORMONE SYNTHESIS

The hormones of the posterior pituitary, vasopressin and oxytocin, are synthesized in specialized neurons in the hypothalamus, the neurohypophysial neurons. These neurons, notable for their large size, are termed *magnocellular neurons*. In the hypothalamus, the magnocellular neurons are clustered in the paired paraventricular and supraoptic nuclei (Fig. 225-1). Vasopressin and oxytocin are also synthesized in parvicellular (i.e., small cell) neurons of the paraventricular nuclei, and vasopressin (but not oxytocin) is also synthesized in the suprachiasmatic nucleus.

Transcription of vasopressin and oxytocin messenger RNA and translation of the vasopressin and oxytocin prohormones occur entirely in the cell bodies of the neurohypophysial neurons. The prohormones provasopressin and prooxytocin are packaged along with processing enzymes into neurosecretory granules that are transported out of the perikaryon of the neurohypophysial neurons via microtubules and down the long axons that form the supraopticohypophysial tract, which terminates in the posterior pituitary. During transport, the processing enzymes cleave provasopressin into vasopressin (9 amino acids), vasopressin-neurophysin (95 amino acids), and vasopressin glycopeptide, or copeptin (39 amino acids). Pro-oxytocin is similarly cleaved to oxytocin (which differs from vasopressin by only two of nine amino acids) and oxytocin-neurophysin. The neurophysins form neurophysin-hormone complexes that stabilize the hormones. Stimulatory (e.g., glutamatergic, cholinergic, and angiotensin) neurotransmitter terminals and inhibitory (e.g., γ-aminobutyric acid and noradrenergic) neurotransmitter terminals control the release of vasopressin through the activity of synaptic contacts on the neurohypophyseal cell bodies. Physiologic release of vasopressin or oxytocin into the general circulation occurs at the level of the posterior pituitary, where, in response to an action potential, intracellular calcium is increased and causes the neurosecretory granules to fuse with the axon membrane, thereby releasing each hormone into the general circulation. Although each of the other prohormone fragments are released into the circulation, vasopressin and oxytocin are the only biologically active components of the prohormones. Factors that stimulate the release of neurohypophysial hormones also stimulate their synthesis. Because synthesis is delayed, maintenance of a large store of hormone in the posterior pituitary is essential to enable the instantaneous release of each hormone that is necessary following acute hemorrhage (vasopressin) or during parturition (oxytocin). In most species, sufficient vasopressin is stored in the posterior pituitary to support maximal antidiuresis for several days and to maintain baseline levels of antidiuresis for weeks.

Vasopressin

Vasopressin and Regulation of Osmolality

The primary physiologic action of vasopressin is its function as a water-retaining hormone. The central sensing system (osmostat) for controlling the release of vasopressin is anatomically discrete, located in a small area of the hypothalamus just anterior to the third ventricle (see Fig. 225-1). The

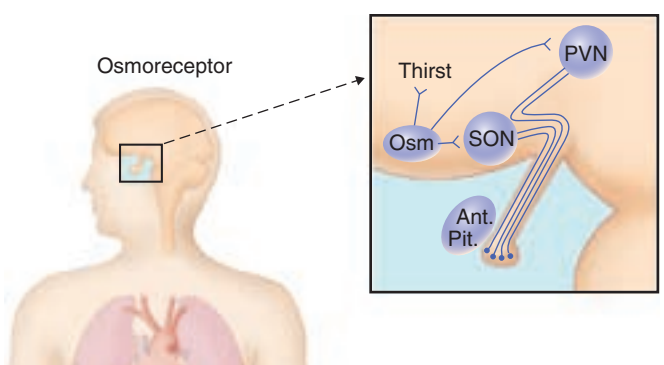

FIGURE 225-1. Sagittal view of the head, demonstrating the position of the neurohypophysis. The magnocellular neurons are clustered in two paraventricular nuclei (PVN) and two supraoptic nuclei (SON). Only one nucleus of each pair is illustrated. The supraoptic nuclei are lateral to the edge of the optic chiasm, whereas the paraventricular nuclei are central along the wall of the third ventricle. The axons of the four nuclei combine to form the supraopticohypophysial tract as they course through the pituitary stalk to their storage terminals in the posterior pituitary. The osmostat (Osm) is in the hypothalamus anterior to the third ventricle; the thirst center (Thirst) is distributed across different brain areas. Ant. Pit. = anterior pituitary. (From Buonocore CM, Robinson AG. Diagnosis and management of diabetes insipidus during medical emergencies. *Endocrinol Metab Clin North Am.* 1993;22:411-423.)

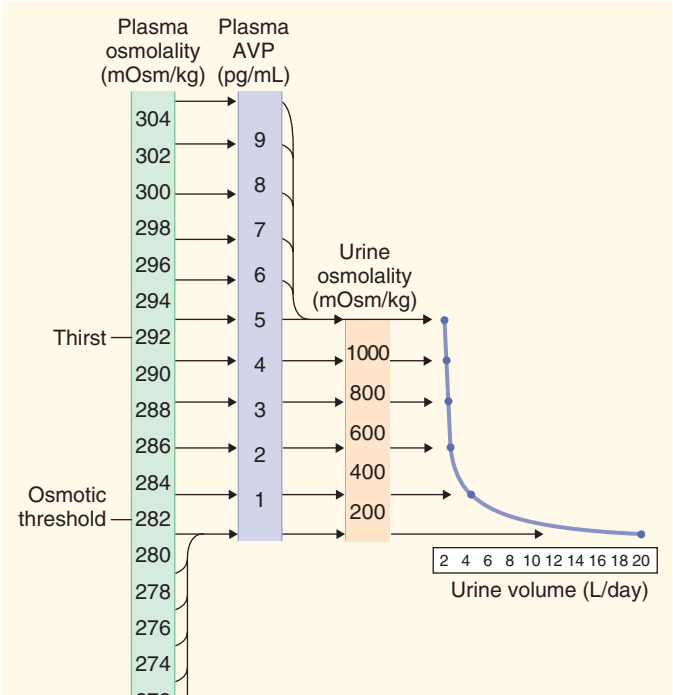

FIGURE 225-2. Idealized schematic of the normal physiologic relationships among plasma osmolality, plasma vasopressin (AVP), urine osmolality, and urine volume. The entire physiologic range of urine osmolality occurs with plasma vasopressin levels from 0 to 5 pg/mL. Increases in plasma osmolality above approximately 290 to 295 mOsm/kg H_2O result in increases in plasma vasopressin but no further concentration of the urine, which is limited by the maximal osmolality in the inner medulla. The relation of volume (calculated on the basis of a constant osmolar load) is inversely exponential to the other parameters. Because of this relationship, urine volume does not change substantially until there is nearly absent vasopressin secretion, after which urine volume increases dramatically. (Calculated from formulas presented in Robertson GL, Shelton RL, Athar S. The osmoregulation of vasopressin. *Kidney Int.* 1976;10:25-37. Figure drawn by J.G. Verbalis, Georgetown University, Washington, DC.)

osmostat controls the release of vasopressin to cause water retention and also stimulates thirst to cause water repletion.

Osmotic regulation of vasopressin release and osmotic regulation of thirst are usually tightly coupled, but they can be dissociated under pathologic conditions. The primary extracellular osmolyte to which the osmoreceptor responds is sodium. Under normal physiologic conditions, glucose and urea cross neuron cell membranes and do not stimulate the release of vasopressin. Although basal osmolality in normal subjects ranges between 280 and 295 mOsm/kg H_2O, extracellular fluid osmolality for each individual is maintained within narrow ranges. Increases in plasma osmolality as small as 1 to 2% are sufficient to stimulate vasopressin release. Basal plasma levels of vasopressin are generally 0.5 to 2 pg/mL, which maintains urine osmolality above plasma osmolality and urine volume in the range of 2 to 3 L/day. When vasopressin levels are suppressed below 0.5 pg/mL, maximal urine osmolality decreases to below 100 mOsm/kg H_2O, and a free water diuresis (or "aquaresis") ensues at levels that approach 800 to 1000 mL/hour (18 to 24 L/day). Increases in plasma osmolality cause a linear increase in plasma vasopressin and a corresponding linear increase in urine osmolality. At a plasma osmolality of approximately 295 mOsm/kg H_2O, urine osmolality is maximally concentrated to 1000 to 1200 mOsm/kg H_2O. Thus, the entire physiologic range of urine osmolality is accomplished by relatively small changes in plasma vasopressin levels of 0 to 5 pg/mL (Fig. 225-2).

To maintain fluid balance, water must be not only conserved but also consumed to replace insensible water losses and obligate urine output. Thirst is not stimulated until a somewhat higher plasma osmolality (5 to 10 mOsm/kg H_2O) than the threshold for release of vasopressin. Most humans derive sufficient water from habitual fluid intake and catabolism of food to maintain plasma osmolality below the threshold that activates thirst. Therefore, under normal physiologic conditions, water balance (and hence plasma osmolality) is regulated more by secretion of vasopressin than by thirst. However, with severe degrees of dehydration, thirst is essential to restore body water deficits.

Vasopressin acts on the V_2 subtype of vasopressin receptors in the collecting duct principal cells of the kidney to cause water retention, or antidiuresis. Vasopressin V_2 receptors are G protein–coupled receptors that activate adenylate cyclase, with subsequent increased intracellular cyclic adenosine monophosphate (cAMP) levels upon ligand activation of the receptor. The increased cAMP initiates the movement of aquaporin-2 (AQP2) water channels to the apical (luminal) membrane of the collecting duct cells. These channels allow facilitated rapid transport of water from the collecting duct lumen into the principal cell along osmotic gradients. The water then exits the cell through the basolateral membrane into the kidney medullary circulation through constitutively expressed aquaporin-3 and aquaporin-4 water channels.[1] This entire process is termed *antidiuresis*. In the absence of vasopressin, the AQP2 channels are reinternalized from the apical membrane into subapical vesicles. This prevents active reabsorption of water from the collecting duct lumen, resulting in diuresis. In addition to this rapid "shuttling" of the AQP2 channels to regulate water reabsorption on a minute-to-minute basis, vasopressin also acts through the V_2 receptors to regulate long-term stores of AQP2—that is, increased vasopressin stimulates AQP2 synthesis, and the absence of vasopressin results in decreased AQP2 synthesis. The hypertonic medullary interstitium is the determinant of the maximal concentration of the urine, which is in equilibrium with the osmolality of the inner medulla of the kidney under conditions of maximal antidiuresis (Chapter 115).

Vasopressin and Pressure and Volume Regulation

High-pressure baroreceptors are located in the aorta and carotid sinus, and low-pressure baroreceptors are located in the right and left atria. Decreases in blood pressure or intravascular volume stimulate vasopressin release, whereas situations that increase blood volume or left atrial pressure (e.g., negative-pressure breathing) decrease the secretion of vasopressin. The release of vasopressin in response to changes in volume or pressure is much less sensitive than the release in response to osmoreceptors; generally a 10 to 15% reduction in blood volume or pressure is needed to stimulate the release of vasopressin. However, once arterial pressure falls below this threshold, the stimulated response is exponential resulting in plasma levels of vasopressin that are markedly greater than those resulting from osmotic stimulation.

The pressor effects of vasopressin are mediated through a separate vasopressin receptor subtype, the V_{1a} receptors, located on vascular smooth muscle. The relatively insensitive regulation of vasopressin secretion by changes in volume and pressure and the modest role of vasopressin to regulate blood pressure are consistent with the notion that regulation of sodium homeostasis by the renin-angiotensin-aldosterone system (Chapter 227) is more important for controlling extracellular and blood volume than is the regulation of water homeostasis. However, the pressor effects of vasopressin to increase blood pressure can become prominent when other blood pressure

regulatory systems are deficient (e.g., autonomic neuropathy or renin-angiotensin-aldosterone system blockade) or in states of pathologic vasodilation (e.g., liver cirrhosis, septic shock).

Vasopressin and Adrenocorticotropic Hormone

Vasopressin stimulates adrenocorticotropic hormone (ACTH) secretion via stimulation of the vasopressin V_{1b} receptor subtypes that are located on anterior pituitary corticotroph cells. Although the major regulator of ACTH secretion is corticotropin-releasing hormone (Chapter 224), vasopressin activates a different signal transduction system in the corticotrophs, so these hormones have synergistic effects on ACTH secretion.

Interaction of Osmotic and Volume Regulation

The vasopressin system has evolved to optimize mammalian water homeostasis. Water is consumed as available in the absence of stimulated thirst, and vasopressin secretion then regulates water excretion to maintain plasma osmolality. Thirst serves as a backup mechanism if dehydration becomes excessive. Because pressure-volume regulation of vasopressin is less sensitive, modest changes in pressure or volume, which are exacerbated by upright posture, do not interfere with the regulation of osmolality. Yet the pressor effect of high vasopressin levels serves to maintain blood pressure if volume depletion or hypotension becomes excessive. Usually, the physiologic regulation of osmolality and pressure-volume are synergistic. Dehydration causes an increase in plasma osmolality and a decrease in blood volume, both of which stimulate the release of vasopressin. Conversely, excess fluid administration causes a decrease in plasma osmolality and an expansion of blood volume, both of which inhibit vasopressin secretion.

Other factors can also modulate osmotic release and action of vasopressin. With volume expansion, natriuretic factors such as atrial natriuretic peptide and brain natriuretic peptide are released from atrial myocytes and act at the kidney to induce a natriuresis. Brain natriuretic peptide is also synthesized in the hypothalamus, where it may act to decrease vasopressin secretion. During pregnancy, there is a decrease of plasma osmolality by approximately 10 mOsm/kg H_2O as a result of a resetting of the osmostat for vasopressin secretion, and the osmostat for thirst is reset downward in parallel. These effects appear to be mediated by the placental hormone relaxin.

Abnormalities in water and electrolyte balance are common in the elderly. This is due in part to age-related changes in body volume (as much as a 50% decrease in total body water occurs in those older than 75 years) and renal function. The elderly also have a decreased sense of thirst. Although there is a normal or even increased ability to secrete vasopressin with age, there is a decreased ability to achieve either maximal urine concentration to retain water or maximal dilution of urine to excrete water. Consequently, the elderly are particularly prone to both hypernatremia or hyponatremia with diseases that affect water balance or from the drugs used to treat various diseases.[2]

Oxytocin

Prolactin is the main hormone necessary for milk production, but oxytocin is essential for milk secretion. Suckling stimulates tactile receptors, producing an afferent signal to the hypothalamus that causes a synchronized release of oxytocin from the posterior pituitary. Oxytocin binds to oxytocin receptors in the breast and induces contraction of myoepithelial cells around the alveoli and ductules to eject milk. In addition, upregulation of uterine oxytocin receptors dramatically increases uterine smooth muscle contractions in response to oxytocin secretion at the end of pregnancy. The greatest release of oxytocin occurs with, not before, delivery of the infant, probably secondary to stretching of the vaginal wall. Because transgenic mice lacking either oxytocin or oxytocin receptors have normal parturition, oxytocin release may be more important to induce uterine contraction to inhibit blood loss after delivery than to initiate parturition. No pathologic syndromes of either increased or decreased secretion of oxytocin have yet been defined, but experimental studies have implicated oxytocin in maternal and affiliative behavior as well as bone formation.[3] However, because of the structural similarity between vasopressin and oxytocin, at high plasma levels oxytocin can activate vasopressin receptors, and vasopressin can activate oxytocin receptors, both of which can have pathologic consequences.

SYNDROME OF INAPPROPRIATE ANTIDIURETIC HORMONE SECRETION

Excess secretion of vasopressin can be caused by abnormally regulated secretion from the posterior pituitary, or by ectopic synthesis and secretion of vasopressin by tumors. Osmotically inappropriate secretion of vasopressin causes renal water retention and volume expansion of body fluids, with consequent dilutional hyponatremia. This disorder is called the syndrome of inappropriate antidiuretic hormone secretion (SIADH) and is discussed in Chapter 116.

DIABETES INSIPIDUS

DEFINITION

Diabetes insipidus is the excretion of a large volume of hypotonic insipid (tasteless) urine, usually manifested by polyuria (increased urination) and polydipsia (increased thirst).[4] The large urine volume, usually in excess of 50 to 60 mL/kg/day, must be distinguished from an increased frequency of small urine volumes and from large volumes of isotonic or hypertonic urine, both of which have different clinical significance.

PATHOBIOLOGY

Five pathophysiologic mechanisms must be considered in the differential diagnosis of diabetes insipidus.

1. Central diabetes insipidus is caused by the inability of the hypothalamus–posterior pituitary to secrete (and usually to synthesize) vasopressin in response to increased osmolality. No concentration of the dilute glomerular filtrate takes place in the renal collecting duct, and consequently, a large volume of hypotonic (i.e., dilute) urine is excreted. This produces a secondary increase in serum osmolality, with stimulation of thirst and secondary polydipsia. Levels of vasopressin in plasma are unmeasurable or inappropriately low for the plasma osmolality.
2. Nephrogenic diabetes insipidus is caused by the inability of an otherwise normal kidney to respond to vasopressin. As in hypothalamic (central) diabetes insipidus, the dilute glomerular filtrate entering the collecting duct is excreted as a large volume of hypotonic urine. The rise in plasma osmolality that occurs stimulates thirst and produces polydipsia. Unlike central diabetes insipidus, however, measured levels of vasopressin in plasma are high or appropriate for plasma osmolality.
3. Gestational diabetes insipidus is a rare condition produced by elevated levels or activity of placental cystine aminopeptidase (oxytocinase or vasopressinase) during pregnancy. The rapid destruction of vasopressin produces diabetes insipidus with polyuria and secondary stimulation of thirst with polydipsia. Because of the circulating vasopressinase, plasma vasopressin levels usually cannot be measured.
4. Primary polydipsia is a disorder of excess fluid ingestion rather than of vasopressin secretion or activity. Excessive ingested water produces a mild decrease in plasma osmolality that shuts off the secretion of vasopressin. In the absence of vasopressin action on the kidney, urine does not become concentrated, and a large volume of hypotonic urine is excreted. The amount of vasopressin in plasma is unmeasurable or low but is appropriate for the low plasma osmolality.
5. Osmoreceptor dysfunction is a variant of central diabetes insipidus in which the neurohypophysis is intact, but the osmoreceptive cells in the anterior hypothalamus have been damaged (see Fig. 225-1). Because the osmoreceptor cells are necessary for osmotically stimulated vasopressin secretion, the patient manifests polyuria. However, because the osmoreceptor cells also control thirst, these patients do not have polydipsia. As a result, they are characterized by elevated serum sodium levels and plasma osmolalities. For this reason, this disorder has also been called essential hypernatremia and adipsic hypernatremia, in recognition of the profound thirst deficits found in most of the affected patients.

Although the pathophysiologic mechanisms for each of these five disorders are distinct, patients in the first four categories usually manifest polyuria and polydipsia, and the serum sodium level is usually normal because an intact thirst mechanism is sufficiently sensitive to maintain water homeostasis in the first three disorders, and the normal kidney has sufficient capacity to excrete the excess water load in the fourth. The fifth category of osmoreceptor dysfunction is the exception, owing to a defective thirst mechanism leading to hypernatremia.

CLINICAL MANIFESTATIONS

Central Diabetes Insipidus

The sudden appearance of hypotonic polyuria[5] after transcranial surgery in the area of the hypothalamus or after head trauma with basal skull fracture and hypothalamic damage obviously suggests the diagnosis of central diabetes insipidus.[6] In these situations, if the patient is unconscious and unable to recognize thirst, hypernatremia is a common accompaniment. However, even in patients with more insidious progression of a specific disease or in patients

with idiopathic central diabetes insipidus, the onset of polyuria is often relatively abrupt and occurs over several days or weeks. Most patients do not notice polyuria until urine volume exceeds 4 L/day, and as illustrated in Figure 225-2, urine volume does not exceed 4 L/day until the ability to concentrate the urine is severely limited and plasma vasopressin is nearly absent. As few as 10 to 15% of the normal number of vasopressinergic neurons in the hypothalamus is sufficient to maintain an asymptomatic urine volume, but the further loss of just a small number of these neurons produces a rapid increase in urine volume and symptomatic polyuria. Urine volume seldom exceeds the amount of dilute fluid delivered to the collecting duct (≈18 to 24 L in humans); in many cases, urine volume is significantly less because patients voluntarily restrict fluid intake, which causes some mild volume contraction and increased proximal tubular reabsorption of fluid. Patients often express a preference for cold liquids, which are more effective in assuaging thirst. Both thirst and increased urine output persist through the night, impairing sleep. Patients with partial central diabetes insipidus have some ability to secrete vasopressin, but this secretion is markedly attenuated at normal levels of plasma osmolality. Therefore, these patients often have symptoms and urine volume similar to those of patients with complete central diabetes insipidus. Because most patients with central diabetes insipidus have sufficient thirst to drink fluid to match urine output, few laboratory abnormalities are present at the time of initial evaluation. The serum sodium level can be in the high-normal range, whereas the blood urea nitrogen level can be low secondary to the large urine volume. Uric acid is relatively high because of the modest volume contraction and lack of action of vasopressin on V_{1a} receptors in the kidney, which stimulate the clearance of uric acid. Uric acid levels greater than 5 mg/dL can distinguish diabetes insipidus from primary polydipsia.

Osmoreceptor Dysfunction

A variant of central diabetes insipidus is the syndrome of osmoreceptor dysfunction. Physiologic maneuvers demonstrate that when patients are euvolemic, an increase in plasma osmolality produces neither secretion of vasopressin nor a sensation of thirst. However, vasopressin is still synthesized by the hypothalamus and stored in the posterior pituitary, because stimulation of baroreceptors by hypovolemia or hypotension results in the prompt secretion of vasopressin; the kidney is responsive because vasopressin release by volume receptor stimulation causes urinary concentration. Because patients lack thirst, they are chronically dehydrated, often with markedly increased serum sodium levels. However, it is the dehydration-induced volume depletion, not the increased osmolality, that eventually stimulates the secretion of vasopressin. The volume of urine output depends on the degree of dehydration-induced secretion of vasopressin. If sufficient fluid replacement is given to return extracellular fluid volume to normal, these patients are unable to regulate vasopressin by osmolality and again become polyuric, thereby manifesting their underlying central diabetes insipidus. Lesions that cause osmoreceptor dysfunction are similar to lesions that can cause central diabetes insipidus, but in contrast to central diabetes insipidus these lesions usually occur more rostrally in the hypothalamus, consistent with the anterior hypothalamic location of the primary osmoreceptor cells (see Fig. 15-2). One lesion that is unique to this disorder is an anterior communicating cerebral artery aneurysm, particularly following resection of the aneurysm.

Central diabetes insipidus can be inherited as an autosomal dominant disease that is typically characterized by an asymptomatic infancy and an onset later in childhood. Most genetic defects are either in the signal peptide of the pre-prohormone or in the neurophysin portion of the prohormone.[7] Mutations involving the vasopressin sequence itself are few. Most cases are believed to result from disruption of cleavage from the signal peptide or abnormal folding of the neurophysin, which slows trafficking of the mutant prohormone through the endoplasmic reticulum, leading to neuronal cell dysfunction and/or death. Because this is a cumulative process, this explains the later onset of central diabetes insipidus with these types of mutations.

Myxedema and adrenal insufficiency both impair the ability to excrete free water by renal mechanisms. The simultaneous occurrence of either of these diseases with central diabetes insipidus (as can occur with a tumor of the hypothalamus or pituitary) can decrease an otherwise large urine output, thereby masking the symptoms of diabetes insipidus. Replacement treatment for the anterior pituitary deficiency, especially glucocorticoids, can then cause a sudden and massive excretion of dilute urine. Similarly, the onset of either hypothyroidism or adrenal insufficiency during the course of diabetes insipidus can decrease the need for vasopressin replacement and in some cases can even cause hyponatremia. Central diabetes insipidus occurs commonly in patients with severe brain ischemia, and is often indicative of brain death. Treatment of the diabetes insipidus along with any coexistent anterior pituitary hormone deficiencies can be used to preserve donor organs in such cases.

Nephrogenic Diabetes Insipidus

Nephrogenic diabetes insipidus is caused by mutations of the vasopressin V_2 receptor or the vasopressin-induced water channel AQP2, or by impairments in the signal transduction system linking V_2 receptor activation and AQP2 membrane insertion. Familial nephrogenic diabetes insipidus is a rare disease, most cases of which (>90%) are due to mutations of the V_2 receptor. More than 100 different V_2 receptor mutations have been described and can be classified into several different general categories based on differences in transport of the mutant receptor to the cell surface and vasopressin binding or stimulation of adenylate cyclase. Because the gene for the V_2 receptor is located on the X chromosome, this is an X-linked recessive disease. Symptoms are noted only in affected males, who often present with vomiting, constipation, failure to thrive, fever, and polyuria during the first week of life. Hypernatremia with a hypotonic urine is typically present. The phenotype is similar in the less than 10% of patients with mutations of the AQP2 water channel, but because the AQP2 gene is located on chromosome 12, mutations cause autosomal recessive disease; consequently, consanguinity and a family history of the disease in men and women is common, and this disorder should be suspected when the proband is female.[8]

Nephrogenic diabetes insipidus can also be acquired during treatment with certain drugs such as demeclocycline (which can be used to treat inappropriate secretion of vasopressin), lithium carbonate (used to treat bipolar disorders), and fluoride (previously used in fluorocarbon anesthetics), and from electrolyte abnormalities such as severe hypokalemia and hypercalcemia. All causes of acquired nephrogenic diabetes insipidus have in common the decreased synthesis and function of AQP2 due to impaired vasopressin signaling from V_2 receptor binding and activation. Other diseases of the kidney can produce polyuria and inability to concentrate the urine secondary to altered renal medullary blood flow or to other disorders that inhibit maintenance of the hyperosmolar concentrating gradient in the inner medulla. Renal manifestations of such disorders (e.g., sickle cell disease, sarcoidosis, pyelonephritis, multiple myeloma, analgesic nephropathy) are discussed in Chapter 121.

Gestational Diabetes Insipidus

In pregnancy, there is an increased metabolism of vasopressin due to cystine aminopeptidase (oxytocinase or vasopressinase), an enzyme that degrades oxytocin and prevents premature uterine contractions. Normally, this is compensated for by increased synthesis and secretion of vasopressin. Rarely, women with normal regulation of vasopressin develop diabetes insipidus because of markedly elevated levels of vasopressinase. Some of these patients have accompanying preeclampsia, acute fatty liver, and coagulopathies, but causal relations between diabetes insipidus and these abnormalities have not been identified. In general, diabetes insipidus does not persist after the pregnancy ends and does not recur in subsequent normal pregnancies.[9]

Polyuria can also manifest in patients who have limited vasopressin reserve (partial central diabetes insipidus) or who respond poorly to vasopressin action (compensated nephrogenic diabetes insipidus). Treatment may be required only during the pregnancy, and the patient often returns to her previous baseline function without the need for therapy when the pregnancy ends. Less commonly, central diabetes insipidus of another cause first becomes symptomatic during pregnancy and then persists afterward, following the usual course of diabetes insipidus.

Primary Polydipsia

Excessive fluid intake also causes hypotonic polyuria and, by definition, polydipsia. This disorder must be differentiated from the various causes of diabetes insipidus. Despite normal pituitary and kidney function, patients with this disorder share many characteristics of both central diabetes insipidus (vasopressin secretion is suppressed as a result of decreased plasma osmolality) and nephrogenic diabetes insipidus (kidney AQP2 expression is decreased as a result of suppressed plasma vasopressin levels). Many different names have been used for this excessive fluid intake, including dipsogenic diabetes insipidus, but primary polydipsia remains the best descriptor to avoid confusing this order with diabetes insipidus as classically defined.

Primary polydipsia is sometimes due to a severe mental illness such as schizophrenia, mania, or obsessive-compulsive disorder, in which case it is called psychogenic polydipsia. These patients usually deny true thirst and attribute their polydipsia to bizarre motives, such as a need to cleanse the

body of poisons. The incidence in psychiatric hospitals can be as high as 40%, and there is no obvious explanation for the polydipsia. Primary polydipsia can also be caused by an abnormality in the osmoregulatory control of thirst, in which case it is called dipsogenic diabetes insipidus. These patients have no overt psychiatric illness and invariably attribute their polydipsia to a nearly constant thirst. Dipsogenic diabetes insipidus is usually idiopathic, but it can also be secondary to organic structural lesions in the hypothalamus identical to those causing central diabetes insipidus, such as neurosarcoidosis of the hypothalamus, tuberculous meningitis, multiple sclerosis, or trauma. Consequently, all polydipsic patients should be evaluated with magnetic resonance imaging (MRI) of the brain before it is concluded that excessive water intake is due to an idiopathic or psychiatric cause. Primary polydipsia can also be produced by diseases or drugs that cause a dry mouth, or by any peripheral disorder causing marked elevations of renin or angiotensin.

Finally, primary polydipsia is sometimes caused by physicians, nurses, lay practitioners, or health writers who recommend a high fluid intake for valid (e.g., recurrent nephrolithiasis) or unsubstantiated health reasons. These patients lack overt signs of mental illness, but they also deny thirst and usually attribute their polydipsia to habits acquired from years of adherence to a drinking regimen. Laboratory studies in these patients are generally normal, although the serum sodium concentration is sometimes at the low end of the normal range, and the level of uric acid is lower than in patients with other forms of diabetes insipidus.

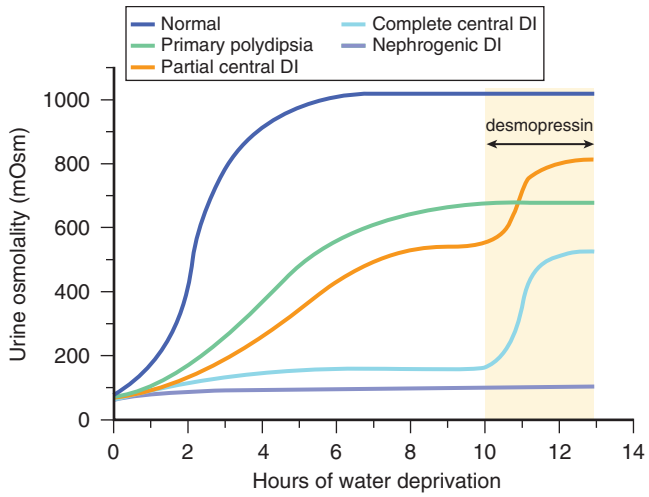

FIGURE 225-3. Responses to the water deprivation test to differentiate various types of diabetes insipidus (DI) and primary polydipsia (as described by Miller M, Dalakos T, Moses AM, et al. Recognition of partial defects in antidiuretic hormone secretion. *Ann Intern Med.* 1970;73:721-729). The response to dehydration reaches a plateau, and the subsequent change in urine osmolality in response to administered desmopressin is illustrated. See the discussion in the text.

DIAGNOSIS

Physiologic Diagnosis

Diabetes insipidus should be considered in all patients presenting with significant polyuria, defined as urine output greater than 50 mL/kg/day. Although osmotic diuresis secondary to hyperglycemia, intravenous contrast agents, or renal injury is a more common clinical cause of polyuria, the medical history, an isotonic urine osmolality, and routine clinical laboratory tests generally distinguish these disorders from diabetes insipidus. A diagnosis of diabetes insipidus can be made when urine osmolality is inappropriately low in the presence of an elevated plasma osmolality as a result of increased serum sodium concentration. These criteria are sometimes met at the initial examination, especially in cases of acute diabetes insipidus occurring after trauma or surgery with inadequate fluid replacement. In such patients with hypernatremia and hypotonic urine osmolality with normal renal function, one need only administer a vasopressin agonist to differentiate central diabetes insipidus, in which a renal response with decreased urine volume and increased urine osmolality occurs, from nephrogenic diabetes insipidus, in which a subnormal renal response is seen. Sometimes in the postoperative state, a water diuresis occurs as a result of water retention during the surgical procedure. Vasopressin is normally secreted in response to surgical stress, causing fluid administered intravenously during the procedure to be retained. During recovery, vasopressin levels fall, and a diuresis of the retained fluid occurs. In this case, the serum sodium level is almost always normal; however, if additional fluid is administered to match the urine output, persistent polyuria can be mistaken for diabetes insipidus. In this situation, the physician should decrease the rate of fluid administered and follow the urine output and serum sodium level. If the urine output decreases and the serum sodium level remains normal, no treatment is necessary; if serum sodium rises above the normal range and the urine remains hypotonic, diabetes insipidus is likely, and the response to a vasopressin agonist can ascertain the type (central versus nephrogenic).

Most outpatients with diabetes insipidus are not hypernatremic, because the polydipsia produced by a normal thirst response is generally sufficient to maintain water homeostasis. Instead, they present with polyuria, polydipsia, and a normal sodium level. In these patients, further testing is necessary to increase serum osmolality and then measure the plasma vasopressin level or the urinary response to an administered vasopressin agonist. The best described test is the water deprivation test (Fig. 225-3),[10] which should be carried out under controlled observation in the hospital or an appropriately equipped outpatient area. The exact timing of the test depends on the patient's symptoms. If the patient has marked polyuria during the night, it is best to begin the test during the day because the patient may become overly dehydrated overnight. However, if the patient has only two or three episodes of nocturia per night, it is best to begin the test in the evening so that the major part of the dehydration takes place when the patient is asleep. In either case, the patient is weighed at the beginning of the test, and all subsequent fluids are withheld. The volume and osmolality of all excreted urine are measured, and the patient is reweighed after each liter of urine output. When three consecutive urine samples have an osmolality differing by no more than 10% and the patient has lost at least 2% of body weight, a blood sample is obtained for the measurement of serum osmolality, sodium, and plasma vasopressin. The patient is then given 2 μg of desmopressin intravenously or intramuscularly and observed for an additional 2 hours.

Adults with normal vasopressin secretion concentrate their urine to greater than 800 mOsm/kg H_2O and have less than a 10% increase in urine osmolality in response to administered desmopressin. Patients with complete central diabetes insipidus have minimal concentration of the urine with dehydration, and a marked increase in urine osmolality (usually >50%) in response to administered desmopressin. Patients with nephrogenic diabetes insipidus usually have no increase in urine concentration in response to administered desmopressin, although in some cases of acquired nephrogenic diabetes insipidus, some increased urinary concentration (but generally < 10%) can occur. Nephrogenic diabetes insipidus is best distinguished from central diabetes insipidus by the measurement of vasopressin in plasma; plasma vasopressin levels are elevated in cases of nephrogenic diabetes insipidus, especially after dehydration.

In patients with partial central diabetes insipidus and patients with primary polydipsia, the urine is often somewhat concentrated in response to dehydration, but not to the maximum of a normal person. The chronically reduced level of vasopressin downregulates the synthesis of AQP2 water channels, and the large urine volume, regardless of cause, washes out the medullary osmotic gradient that is the determinant of maximal urine concentration. When desmopressin is administered, patients with partial central diabetes insipidus have a further increase (usually > 10% but < 50%) in urine osmolality, whereas most patients with primary polydipsia have no further increase (i.e., <10%). However, the reliability of distinguishing between these two disorders by the water deprivation test is suboptimal. Some patients with primary polydipsia may not become sufficiently dehydrated to secrete maximal vasopressin and hence have an increase in urine osmolality in response to administered desmopressin. Alternatively, some patients with partial central diabetes insipidus can become sufficiently dehydrated that their maximal concentration of urine is reached during the test, and no further concentration is seen with administered desmopressin. Plasma vasopressin levels at the end of dehydration are better at discriminating between these two disorders, but only at high serum sodium concentrations (i.e., >145 mmol/L). Consequently, some investigators recommend a limited infusion of hypertonic (3%) sodium chloride solution to achieve these elevated levels if they are not achieved by the water deprivation itself. Measurement of the C-terminal fragment of the vasopressin prohormone copeptin may be a better surrogate measure of vasopressin secretion.[11]

In some difficult cases, the response to treatment with a vasopressin agonist can be a useful aid to diagnosis. If a decrease in polyuria and thirst with maintenance of normal serum sodium concentration occurs, a diagnosis

of partial central diabetes insipidus is likely; however, if polydipsia persists and hyponatremia develops, a diagnosis of primary polydipsia is confirmed.

Etiologic Diagnosis

If the water deprivation test confirms that inadequate vasopressin secretion is responsible for the polyuria, the underlying cause must be determined. MRI of the hypothalamic-pituitary area is the most important diagnostic tool in these cases. The three areas of interest are the immediate suprasellar region of the hypothalamus, the pituitary stalk, and the posterior pituitary within the sella turcica (see the earlier discussion of anatomy). Most slow-growing tumors confined to the sella do not cause diabetes insipidus. To cause central diabetes insipidus, tumors in the hypothalamic area immediately above the sella must be either sufficiently large to destroy 80 to 90% of the vasopressin cells or located where the paths of the four nuclear groups converge at the origin of the pituitary stalk, just above the diaphragma sellae. Primary tumors, especially craniopharyngioma and suprasellar germinoma, metastatic tumors, and infiltrative diseases can also cause diabetes insipidus by infiltration of the pituitary stalk, which is then thickened (i.e., >2 mm) on MRI. On T1-weighted MRI, the vasopressin and oxytocin stored in neurosecretory granules in the posterior pituitary are visualized as a bright spot in the sella turcica. Most but not all normal subjects have this bright spot (it is absent more frequently in elderly and dehydrated patients); in most but not all patients with central diabetes insipidus, the bright spot is absent. Thickening of the stalk and absence of the bright spot are therefore especially suggestive of a hypothalamic disease process.[12]

Tumors that cause central diabetes insipidus are most often benign primary intracranial tumors such as craniopharyngioma, ependymoma (suprasellar germinoma), and pinealoma, which arise in the third ventricle. Primary tumors of the anterior pituitary (Chapter 224) cause diabetes insipidus only when substantial suprasellar extension is present. However, rapidly growing intrasellar lesions, such as metastases from carcinomas of the lung, breast, and melanoma or hemorrhage into pituitary adenomas, can cause diabetes insipidus because there is insufficient time for the vasopressin axons to adapt by releasing vasopressin from the hypothalamus. Metastases to the hypothalamus can also destroy the supraopticohypophysial tract and produce diabetes insipidus. Granulomatous diseases, such as Langerhans cell histiocytosis, sarcoidosis, tuberculosis, and leukemic infiltrates and lymphomas of the hypothalamus, can cause diabetes insipidus by destroying vasopressin cells. In such patients, the diagnosis is usually suspected on the basis of peripheral manifestations of the respective diseases. Lymphocytic infundibuloneurohypophysitis is an autoimmune disease similar to lymphocytic hypophysitis of the anterior pituitary (Chapter 224) in which lymphocytes infiltrate the neurohypophysis to produce diabetes insipidus. The hallmarks of this process are a thickened pituitary stalk and an absence of the pituitary bright spot in a patient with the abrupt onset of polyuria and polydipsia, particularly a postpartum female. The diagnosis was originally demonstrated by pituitary biopsy, but now is more commonly made by regression of the thickened stalk with continued MRI follow-up. When no specific cause is identified, the diagnosis of exclusion is idiopathic diabetes insipidus; but most such cases are probably caused by an autoimmune disease, and other autoimmune diseases, including anterior pituitary hypophysitis,[13] are often recognized in affected patients. When central nervous system disease is suspected but not diagnosed by MRI or general physical examination, cerebrospinal fluid obtained by lumbar puncture may be helpful in identifying tumor cells or markers of tumors or inflammatory processes (e.g., elevated angiotensin-converting enzyme levels with neurosarcoidosis, elevated β-HCG levels with germinomas). A family history suggestive of diabetes insipidus should be investigated with genetic testing for inherited mutations in the vasopressin or vasopressin receptor genes depending on the site of the defect.

TREATMENT

Rx

Because excess excretion of water is the primary manifestation of diabetes insipidus, water replacement in adequate quantities avoids the metabolic complications of all forms of this disease. However, oral or intravenous administration of the volume of fluid required to replace the often large urinary losses in diabetes insipidus is difficult and inconvenient. The goal of therapy is therefore to reduce the amount of polyuria and polydipsia to a tolerable level while avoiding overtreatment, which can produce water retention and hyponatremia.

Central Diabetes Insipidus

The best therapeutic agent for the treatment of central diabetes insipidus is the vasopressin agonist desmopressin.[14] Desmopressin is different from vasopressin in that the amino group of the N-terminal cystine residue has been removed to prolong the duration of action, and D-arginine has been substituted for L-arginine in position 8 to decrease the vasopressor effects. At therapeutic dosages, this agent acts primarily on V_2 or antidiuretic receptors, with minimal activity at V_{1a} or pressor receptors. Desmopressin is available as tablets of 0.1 or 0.2 mg for oral administration and in either a spray bottle that delivers a fixed dose of 10 µg in 100 µL or a bottle with a rhinal catheter that can deliver 50 to 200 µL (5 to 20 µg) for intranasal administration. When therapy is initiated, it is generally best to begin with a low dose (e.g., half of a 0.1-mg tablet, 5 µg by the rhinal tube, or a single 100 µL spray of 10 µg) at bedtime to allow the patient to sleep through the night, and then determine the duration of action by quantifying the polyuria the next day. The duration of action of a single dose varies from 6 to 24 hours, but in most patients, a good therapeutic response can be achieved on an every-12-hour schedule for the nasal spray or an 8- or 12-hour schedule for the tablets. Desmopressin is also available for parenteral use in 1-mL vials of 4 µg/mL. Parenteral administration is especially useful postoperatively or when a patient is unable to take the nasal preparation. In hospitalized patients, some physicians add vasopressin directly to a crystalloid solution to infuse doses in the range of 0.25 to 2.7 mIU/kg/hour to cause modest but persistent urinary concentration as a treatment of diabetes insipidus. With any form of desmopressin administration, serum sodium levels should be monitored regularly to prevent the development of hyponatremia.[15]

Osmoreceptor Dysfunction

Because the diabetes insipidus of patients with osmoreceptor dysfunction is central, they respond to desmopressin as do patients with central diabetes insipidus. However, because of their thirst defect, this is usually not sufficient to maintain normal plasma osmolality. Consequently, they must be given a "prescription" for amounts of fluids to be consumed each 24 hours in order to maintain normal serum sodium levels and plasma osmolalities. This must be individualized to each patient because overconsumption of fluid coupled with desmopressin administration can produce severe hyponatremia. Body weights using an accurate scale is useful as a guide to preventing under- or overhydration, but frequent monitoring of serum sodium levels is usually necessary as well.

Nephrogenic Diabetes Insipidus

Although most patients with nephrogenic diabetes insipidus do not respond to desmopressin, a small number have a partial response to higher doses (e.g., 10 to 20 µg subcutaneously or intranasally). For the majority of patients who have no response to desmopressin, some orally administered pharmacologic agents are also useful in treating nephrogenic diabetes insipidus. Thiazide diuretics cause sodium depletion and volume contraction and decrease urine volume by increasing proximal tubular reabsorption of glomerular filtrate. Prostaglandin synthase inhibitors (e.g., indomethacin) block the action of prostaglandin E to inhibit the action of vasopressin on the kidney. Chlorothiazide, amiloride, and prostaglandin synthase inhibitors are useful to reduce polyuria in nephrogenic diabetes insipidus. However, none of these agents has been approved by the U.S. Food and Drug Administration for the treatment of diabetes insipidus; therefore, the prescribing physician should be aware of potential toxicities and side effects. In cases of drug-induced nephrogenic diabetes insipidus, the most direct therapy is discontinuation of the offending agent, if possible. Symptomatic nephrogenic diabetes insipidus is usually treated with a thiazide diuretic, which is enhanced by coadministration of the potassium-sparing diuretic amiloride. Amiloride can be especially beneficial in cases of nephrogenic diabetes insipidus induced by lithium, because the drug decreases the entrance of lithium into cells in the distal tubule. When diuretics are used to treat nephrogenic diabetes insipidus, special attention should be paid to the possibility that the induced dehydration may increase the concentration of other drugs.

Gestational Diabetes Insipidus

During pregnancy, vasopressinase increases the metabolism of vasopressin but not of desmopressin, so desmopressin is the drug of choice for these patients. The vasopressinase activity subsides by a few weeks after delivery, and patients with the onset of partial diabetes insipidus during pregnancy may become asymptomatic after delivery. An additional advantage of desmopressin is that it has little action on the oxytocin receptors of the uterus. During pregnancy, normal plasma osmolality decreases by approximately 10 mOsm/kg H_2O because of changes in serum sodium, so pregnant patients with diabetes insipidus require only enough desmopressin to maintain the serum sodium at this lower level.

Correction of Hyperosmolality

Some situations require special attention during therapy. Rarely, if patients with diabetes insipidus are unable to drink or are given a hypertonic solution, severe hypernatremia can develop acutely. Osmotic equilibrium with the

intracellular water of neurons and glia produces shrinking of the brain. The brain is in a closed vault (i.e., the skull), and when the brain shrinks, traction on the vasculature of the central nervous system can cause the rupture of blood vessels and subarachnoid or intracerebral hemorrhage. If the hypernatremia persists for a longer time, the neurons accommodate by producing organic osmolytes (previously called idiogenic osmoles), which limit the amount of brain shrinkage. Once this adaptation has occurred, a too-rapid lowering of osmolality in the extracellular fluid will produce a shift of water into the brain and cause cerebral edema. In this situation, desmopressin can be administered to produce constant antidiuresis, and the amount of water given can be regulated to decrease osmolality by no more than approximately 12 mEq/L every 24 hours. Postoperatively or after head trauma, diabetes insipidus can be transient (see Prognosis), and the need for long-term maintenance therapy cannot be immediately established.

PROGNOSIS

The prognosis of properly treated diabetes insipidus is excellent. If nephrogenic diabetes insipidus is diagnosed and treated early, intracranial calcification and mental retardation do not occur. When the diabetes insipidus is secondary to a recognized disease process, that disease generally determines the ultimate prognosis. In some specific clinical situations, the course is different and characteristic. The development of diabetes insipidus after surgical or traumatic injury to the neurohypophysis can follow any of several well-defined patterns (Fig. 225-4). In some patients, polyuria develops 1 to 4 days after injury and resolves spontaneously. Less often, the diabetes insipidus is permanent and continues indefinitely. Most interestingly, one can see a "triphasic" response that has been well described after pituitary stalk transection. The first phase of diabetes insipidus is due to axon shock and lack of function of the damaged neurons. This phase lasts several hours to several days and is followed by a second, antidiuretic phase that is due to the uncontrolled release of vasopressin from the disconnected and degenerating posterior pituitary or from the remaining severed neurons. Overly aggressive administration of fluids during this second phase does not suppress the uncontrolled vasopressin release from the damaged neurohypophysis and can lead to hyponatremia. The antidiuresis can last 2 to 14 days, after which diabetes insipidus recurs after depletion of vasopressin from the degenerating posterior pituitary gland (third phase). Transient hyponatremia without preceding or subsequent diabetes insipidus has been reported after transsphenoidal surgery for pituitary microadenomas.

Once a deficiency of vasopressin secretion has been present for more than a few weeks, it rarely improves, even if the underlying cause of the neurohypophysial destruction is eliminated. The major exception to this is postoperative diabetes insipidus, in which spontaneous resolution is the rule. Although recovery from diabetes insipidus that persists more than several weeks postoperatively is less common, and is uncommon after 1 year of continued diabetes insipidus, well-documented cases of recovery as long as 10 years after the initiating event have been reported. Potential return of function is a reason to occasionally withhold therapy during long-term treatment. Diabetes insipidus should not be considered idiopathic until at least 4 years of follow-up. During this interval, annual computed tomography or MRI is indicated to search for a tumor or infiltrative process that may not have been detected at the initial examination.

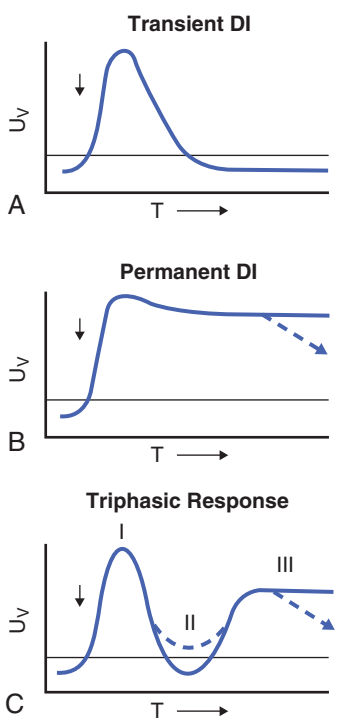

FIGURE 225-4. A to C, Diagrammatic summary of the major patterns of postoperative and post-traumatic diabetes insipidus (DI). The abscissa represents time (T) after the initial injury (*arrow*); the ordinate represents urinary volume (Uv) relative to a hypothetical "normal" urine output of 2 to 3 L/24 hours (*solid line*). See the discussion in the text. During the triphasic response (C), uncontrolled release of vasopressin from the disconnected or damaged posterior pituitary gland causes an antidiuresis that can lead to water retention and a dilutional hyponatremia. Diabetes insipidus returns as the third phase after the stored hormone in the posterior pituitary has been depleted. (From Verbalis JG, Robinson AG, Moses AM. Postoperative and post-traumatic diabetes insipidus. In: Czernichow AP, Robinson A, eds. *Diabetes Insipidus in Man: Frontiers of Hormone Research.* Basel: S Karger; 1985:247.)

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Kortenoeven ML, Fenton RA. Renal aquaporins and water balance disorders. *Biochim Biophys Acta*. 2014;1840:1533-1549.
2. Cowen LE, Hodak SP, Verbalis JG. Age-associated abnormalities of water homeostasis. *Endocrinol Metab Clin North Am*. 2013;42:349-370.
3. Colaianni G, Tamma R, Di BA, et al. The oxytocin-bone axis. *J Neuroendocrinol*. 2014;26:53-57.
4. Leroy C, Karrouz W, Douillard C, et al. Diabetes insipidus. *Ann Endocrinol (Paris)*. 2013;74:496-507.
5. Jakes AD, Bhandari S. Investigating polyuria. *BMJ*. 2013;347:f6772.
6. Schreckinger M, Szerlip N, Mittal S. Diabetes insipidus following resection of pituitary tumors. *Clin Neurol Neurosurg*. 2013;115:121-126.
7. Babey M, Kopp P, Robertson GL. Familial forms of diabetes insipidus: clinical and molecular characteristics. *Nat Rev Endocrinol*. 2011;7:701-714.
8. Bichet DG. Physiopathology of hereditary polyuric states: a molecular view of renal function. *Swiss Med Wkly*. 2012;142:w13613.
9. Aleksandrov N, Audibert F, Bedard MJ, et al. Gestational diabetes insipidus: a review of an underdiagnosed condition. *J Obstet Gynaecol Can*. 2010;32:225-231.
10. Fenske W, Allolio B. Current state and future perspectives in the diagnosis of diabetes insipidus: a clinical review. *J Clin Endocrinol Metab*. 2012;97:3426-3437.
11. Fenske W, Quinkler M, Lorenz D, et al. Copeptin in the differential diagnosis of the polydipsia-polyuria syndrome–revisiting the direct and indirect water deprivation tests. *J Clin Endocrinol Metab*. 2011;96:1506-1515.
12. Di IN, Napoli F, Allegri AE, et al. Diabetes insipidus–diagnosis and management. *Horm Res Paediatr*. 2012;77:69-84.
13. Bando H, Iguchi G, Fukuoka H, et al. The prevalence of IgG4-related hypophysitis in 170 consecutive patients with hypopituitarism and/or central diabetes insipidus and review of the literature. *Eur J Endocrinol*. 2014;170:161-172.
14. Oiso Y, Robertson GL, Norgaard JP, et al. Clinical review: treatment of neurohypophyseal diabetes insipidus. *J Clin Endocrinol Metab*. 2013;98:3958-3967.
15. Behan LA, Sherlock M, Moyles P, et al. Abnormal plasma sodium concentrations in patients treated with desmopressin for cranial diabetes insipidus: results of a long term retrospective study. *Eur J Endocrinol*. 2015;172:243-250.

REVIEW QUESTIONS

1. The hormone vasopressin (AVP) that regulates body water balance is synthesized where?

A. The anterior pituitary gland
B. The posterior pituitary gland
C. The supraoptic and paraventricular nuclei of the hypothalamus
D. The anterior hypothalamus near the osmoreceptor cells
E. The principal collecting duct cells of the kidney

Answer: C The supraoptic and paraventricular nuclei of the hypothalamus. The hormones of the posterior pituitary, vasopressin and oxytocin, are synthesized in specialized neurons in the hypothalamus, the neurohypophysial neurons. These neurons, notable for their large size, are termed *magnocellular neurons*. In the hypothalamus, magnocellular neurons are clustered in the paired paraventricular and supraoptic nuclei (see Fig. 232-1). The synthesized vasopressin prohormone is transported down the axons of the magnocellular neurons to the posterior pituitary gland, where vasopressin is released in response to specific stimuli.

2. Vasopressin secretion is stimulated at what levels of plasma osmolality?

A. Plasma osmolality above the thirst threshold of 295 mOsm/kg H_2O
B. Increases of plasma osmolality of 4 to 5%
C. Increases of plasma osmolality of 2 to 3%
D. Increases of plasma osmolality of 1 to 2%
E. Decreases in plasma osmolality of 1 to 2%

Answer: D Increases of plasma osmolality of 1 to 2%. Increases in plasma osmolality as small as 1 to 2% stimulate vasopressin release. Basal plasma levels of vasopressin are generally 0.5 to 2 pg/mL, which is sufficient to maintain urine osmolality above plasma osmolality and urine volume in the range of 2 to 3 L/day. When vasopressin levels are suppressed below 0.5 pg/mL, maximal urine osmolality decreases to less than 100 mOsm/kg H2O, and a free water diuresis (or "aquaresis") ensues at levels approaching 800 to 1000 mL/hour (18 to 24 L/day). Increases in plasma osmolality cause a linear increase in plasma vasopressin and a corresponding linear increase in urine osmolality. At a plasma osmolality of approximately 295 mOsm/kg H_2O, urine osmolality is maximally concentrated to 1000 to 1200 mOsm/kg H_2O. Thus, the entire physiologic range of urine osmolality is accomplished by relatively small changes in plasma vasopressin of 0 to 5 pg/mL (see Fig. 232-2).

3. Patients with diabetes insipidus and inability to concentrate their urine usually present with what manifestations?

A. Polyuria, polydipsia, and hyperosmolality
B. Polyuria, polydipsia, and dehydration
C. Polyuria, polydipsia, and hypernatremia
D. Polyuria, polydipsia, and elevated BUN and creatinine
E. Polyuria, polydipsia, and normal serum sodium, osmolality, BUN, and creatinine

Answer: E Polyuria, polydipsia, and normal serum sodium, osmolality, BUN, and creatinine. Most patients with diabetes insipidus are not hypernatremic, because the polydipsia produced by a normal thirst response is generally sufficient to maintain water homeostasis. Instead, they present with polyuria, polydipsia, and a normal sodium level and osmolality. Because their fluid intake is sufficient to maintain homeostasis, they are not dehydrated and do not have an elevated BUN or creatinine. In these patients, further testing is necessary to increase serum osmolality and then measure the plasma vasopressin level or the urinary response to an administered vasopressin agonist.

4. After a water deprivation test, a patient increases their urine osmolality from 350 to 375 mOsm/kg H_2O following administration of desmopressin (7% increase). What is the most likely diagnosis?

A. Primary polydipsia
B. Nephrogenic diabetes insipidus
C. Partial hypothalamic diabetes insipidus
D. Gestational diabetes insipidus
E. Osmoreceptor dysfunction

Answer: A Primary polydipsia. Adults with normal vasopressin secretion can concentrate their urine to greater than 800 mOsm/kg H_2O and have less than a 10% increase in urine osmolality in response to administered desmopressin. Patients with complete central diabetes insipidus have minimal concentration of the urine with dehydration, and a marked increase in urine osmolality (usually > 50%) in response to administered desmopressin. Patients with nephrogenic diabetes insipidus usually have no increase in urine concentration in response to administered desmopressin, although in some cases of acquired nephrogenic diabetes insipidus, some increased urinary concentration (generally < 10%) can occur. In patients with partial central diabetes insipidus and patients with primary polydipsia, the urine is often somewhat concentrated in response to dehydration, but not to the maximum of a normal person. The chronically reduced level of vasopressin downregulates the synthesis of aquaporin-2 water channels, and the large urine volume, regardless of cause, washes out the medullary osmotic gradient that determines the maximal urine concentration. When desmopressin is administered, patients with partial central diabetes insipidus have a further increase (usually > 10% but < 50%) in urine osmolality, whereas most patients with primary polydipsia have no further increase (i.e., <10%). Patients with gestational diabetes insipidus respond well to desmopressin because desmopressin is resistant to destruction by placental cystine aminopeptidase. Osmoreceptor dysfunction also responds well to desmopressin because it is a variant of central diabetes insipidus, with inadequate osmotic stimulation of vasopressin secretion.

5. What is the treatment of choice for patients with central diabetes insipidus?

A. Vasopressin
B. Desmopressin
C. Thiazide diuretics
D. Water
E. Insulin

Answer: B Desmopressin. The best therapeutic agent for the treatment of hypothalamic diabetes insipidus is the vasopressin agonist desmopressin. Desmopressin is different from vasopressin in that the amino group of the N-terminal cystine residue has been removed to prolong the duration of action, and D-arginine has been substituted for L-arginine in position 8 to decrease the vasopressor effects. At therapeutic dosages, this agent acts on V_2 or antidiuretic receptors, with minimal activity at V_{1a} or pressor receptors. Because excess diuresis of water is the primary manifestation of diabetes insipidus, water replacement in adequate quantities avoids the metabolic complications of this disease. However, oral or intravenous administration of the volume of fluid required to replace urinary losses in diabetes insipidus is difficult and inconvenient. Thiazide diuretics and amiloride are useful to reduce polyuria in nephrogenic diabetes insipidus by causing sodium depletion and volume contraction, thereby decreasing urine volume by increasing proximal tubular reabsorption of glomerular filtrate, but do not have as great an effect as desmopressin for hypothalamic diabetes insipidus. Insulin is used to treat diabetes mellitus, not diabetes insipidus.

226

THYROID

MATTHEW KIM AND PAUL W. LADENSON

The adult thyroid gland contains two lobes that wrap along the anterolateral aspects of the trachea, midway between the thyroid cartilage and the suprasternal notch. Each lobe is demarcated into upper, middle, and lower poles. The right and left lobes are connected by an isthmus on the anterior aspect of the trachea just below the cricoid cartilage. The normal terminus of the thyroglossal duct can persist as a pyramidal lobe, which is often palpably enlarged in diffuse thyroid disorders such as autoimmune thyroiditis and Graves disease.

With thyroid enlargement, the attachment of the sternothyroid muscle to the trachea limits upward expansion of the lobes, but further lateral, posterior, and downward growth may lead the gland to extend into the superior mediastinum, compressing the trachea and veins at the thoracic outlet. The parathyroid glands usually lie behind the superior and inferior poles of the thyroid lobes. The recurrent laryngeal nerves course upward along the tracheoesophageal groove, from which branches pass behind each thyroid lobe to innervate the larynx.

Thyroid tissue is composed of clustered spherical follicles, each containing a single layer of follicular epithelial cells known as thyrocytes that surround a lumen containing colloid. The principal component of colloid is thyroglobulin, a thyrocyte-specific protein. Parafollicular C cells, which are derived from neural crest tissue and produce calcitonin, are widely dispersed between follicles.

PHYSIOLOGY

Thyroid Hormone Synthesis and Secretion

Dietary iodine in the form of iodide (I^-) or iodate (IO_3^-) is absorbed by the gastrointestinal tract and distributed in the extracellular fluid. Daily iodine intake in the United States equals or exceeds the recommended daily intake of 150 μg because of the widespread use of iodized salt and iodate preservatives in baked goods. Circulating iodide is actively transported into the thyrocyte by the sodium-iodide symporter. Within the thyrocyte, iodide is rapidly oxidized by H_2O_2 in a reaction catalyzed by thyroid peroxidase. The reactive intermediate formed is covalently bound to tyrosyl residues present in thyroglobulin to generate monoiodotyrosine and diiodotyrosine residues through a process known as organification. Thyroid peroxidase also catalyzes the coupling of the monoiodotyrosine and diiodotyrosine residues to generate thyroxine (T_4) and triiodothyronine (T_3) residues in thyroglobulin,

which is secreted into the follicular lumen. Thyroglobulin is then pinocytosed at the apical membrane, and T_4 and T_3 are secreted after proteolysis of thyroglobulin. In the normal state, approximately 100 μg of T_4 and 5 μg of T_3 are directly released into the circulation each day. Pharmacologic amounts of iodine inhibit iodide trapping, organification, and release of the thyroid hormones. Lithium can also block thyroid hormone release.

Thyroid Hormone Transport and Metabolism

Circulating thyroid hormones are more than 99% bound to three classes of plasma proteins. Thyroxine-binding globulin (TBG) functions as the principal transport protein. Thyroxine-binding prealbumin (also known as transthyretin) and albumin make lesser contributions to T_4 and T_3 transport in blood. Pregnancy and exposure to pharmacologic doses of estrogens can increase the TBG level, as can hepatitis, familial TBG excess, and certain medications including 5-fluorouracil, tamoxifen, and methadone. Conversely, decreased TBG levels may occur with systemic illness, severe hepatic disease, nephrotic syndrome, and treatment with androgens, glucocorticoids, and slow-release nicotinic acid. Whereas total T_4 and T_3 levels rise and fall with changes in TBG, free T_4 and T_3 levels remain constant. Familial dysalbuminemic hyperthyroxinemia is an autosomal dominant disorder characterized by the production of an albumin fraction that binds T_4 with a higher-than-normal affinity. Affected individuals may present with high total T_4 levels with an inappropriately normal thyroid-stimulating hormone (TSH), and with normal free T_4 when this is measured by equilibrium dialysis.

The receptor binding and biologic activity of T_3 is eight-fold greater than that of T_4. More than 80% of the T_3 present in target tissues is derived from T_4 through the action of deiodinase enzymes that remove an outer-ring iodine, converting T_4 to T_3, generating the pool of T_3 in target tissues, and contributing to T_3 in the circulation. The type 2 deiodinase is present in the pituitary gland and brain, whereas type 1 deiodinase predominates in other peripheral tissues such as the liver and kidney. Activity of these deiodinases may be inhibited by systemic illness, iodide-containing compounds (e.g., amiodarone, radiocontrast agents), glucocorticoid therapy, and selenium deficiency. T_4 can also be converted by inner-ring deiodination to biologically inactive reverse T_3 (rT_3) by type 1 deiodinase and the type 3 deiodinase present in glial cells of the central nervous system. It deactivates thyroid hormone by inner-ring monodeiodination, a process that converts T_4 to inactive reverse T_3 (rT_3) and T_3 to inactive diiodothyronine (T_2). Type 3 deiodinase is also expressed in placenta, accounting for the increased thyroxine dose requirement in pregnant women.

Control of Thyroid Function

The growth of thyroid tissue and the production of thyroid hormones are controlled by the hypothalamus and the pituitary gland. Thyrotropin-releasing hormone (TRH) is a tripeptide synthesized by the hypothalamus and transported to the pituitary gland by the hypothalamic-pituitary portal system, where it binds to receptors on thyrotrophic cells, stimulating the synthesis and secretion of TSH (thyrotropin). TSH is a heterodimeric glycoprotein composed of a unique β-subunit coupled to an α-subunit identical to that present in follicle-stimulating hormone, luteinizing hormone, and chorionic gonadotropin (CG). It is transported in the circulation to the thyroid gland, where it binds to the TSH receptor on thyrocytes. The binding of TSH to the TSH receptor stimulates thyrocyte growth and all of the steps in synthesis and secretion of thyroid hormone. Circulating T_4 and T_3 exert negative feedback at the levels of both the hypothalamus and the pituitary gland, inhibiting the synthesis and secretion of TRH and TSH, respectively.

Thyroid Hormone Action

Thyroid hormone binds receptors that are members of the nuclear receptor superfamily, regulating expression of thyroid hormone–responsive genes. Isoforms of the thyroid hormone receptors (α_1, β_1, and β_2) bind to a specific hexameric oligonucleotide sequence in the transcriptional regulatory region of thyroid hormone–responsive genes. For example, in liver, T_3 increases expression of the low-density lipoprotein (LDL) receptor, resulting in accelerated LDL cholesterol clearance. In myocardium, T_3 increases myocyte contractility and relaxation by promoting expression of alpha myosin heavy chain and sarcoplasmic reticulum adenosine triphosphatase (ATPase), respectively. In the cardiac conducting system, T_3 increases the heart rate by accelerating sinoatrial node depolarization and repolarization. Other physiologic effects of thyroid hormone include increases in basal metabolic rate, mental alertness, ventilatory drive, gastrointestinal motility, and bone turnover. During fetal development, thyroid hormone plays critical roles in brain development and skeletal maturation.

DIAGNOSIS

Physical Examination

Examination of the thyroid begins with inspection of the lower anterior portion of the neck to check for diffuse or asymmetrical gland enlargement, tracheal deviation, lymphadenopathy, and jugular venous distention. Palpation can be performed by an anterior or posterior approach. Anterior palpation can be performed by using the thumb of one hand to locate the isthmus of the gland beneath the cricoid cartilage. The right lobe of the thyroid gland can be palpated by placing the left thumb along the left side of the trachea to displace the contralateral lobe, using the tips of the fingers of the right hand medial to the right sternocleidomastoid muscle at the level of the isthmus, and using the pads of the fingers of the right hand to define, while the patient swallows, the characteristics of that thyroid lobe (i.e., its size, firmness, contour, mobility, and potential tenderness). The maneuver is reversed to examine the left lobe.

Laboratory Findings

TSH and Thyroid Hormone Levels

The serum TSH level is a sensitive indicator of primary thyroid gland dysfunction. Contemporary TSH immunoassays permit accurate detection of all common causes of thyroid hormone deficiency and excess. Indeed, TSH levels even become abnormal when patients' thyroid hormone levels remain within broad reference ranges, conditions termed *subclinical hypothyroidism* and *subclinical thyrotoxicosis*. In most patients with primary thyroid gland dysfunction, measurement of a single TSH level permits an accurate classification of the thyroid status. Limitations of TSH testing occur when there is TSH-mediated secondary thyroid dysfunction, reduced biologic activity of T_3 or TSH itself, temporary disequilibrium of the hypothalamic-pituitary-thyroid axis, or analytic problems affecting the TSH immunoassay. A longitudinal community-based cohort study has documented that aging is associated with increased serum TSH concentrations with no change in free T_4 levels.[1] These findings suggest that the TSH increase in many elderly individuals arises from age-related alteration in the TSH set point or reduced TSH bioactivity, rather than occult thyroid disease.

Measurements of serum T_4 and T_3 levels confirm the significance of an abnormal TSH level, define the severity of thyroid dysfunction, and provide a clue to the underlying cause. Whereas assays that measure total T_4 and T_3 levels are accurate, their results do not distinguish between large plasma protein–bound and free fractions of each hormone. Consequently, congenital and acquired derangements of TBG (and, less commonly, transthyretin and albumin) can alter the total, but not the free T_4 and T_3 levels. These conditions can be misdiagnosed as abnormal thyroid function unless a discordance in TSH is noted or one of these underlying conditions is suspected (Tables 226-1 and 226-2). There are several methods of estimating unbound T_4 and T_3 levels. Free T_4 and free T_3 immunoassays are now widely employed for this purpose and yield reliable results in common conditions that alter plasma protein levels, such as estrogen-induced TBG excess. Free T_4 measurement after equilibrium dialysis of serum is the most accurate approach, but it is technically demanding and less readily available.

Other Laboratory Tests

Measurement of thyroid autoantibody titers can be useful in the evaluation of thyroid dysfunction. Antithyroid peroxidase and antithyroglobulin antibody titers can confirm the diagnosis of autoimmune thyroiditis. TSH receptor binding and stimulating immunoglobulin levels can be used to confirm the diagnosis of Graves disease. The erythrocyte sedimentation rate (ESR) can be helpful in the diagnosis of subacute thyroiditis. Serum thyroglobulin and calcitonin levels are used as tumor markers when observing patients treated for differentiated and medullary thyroid cancers, respectively.

Imaging

Anatomic Imaging

Ultrasonography provides images that characterize the thyroid gland's size, symmetry, texture, vascularity, and structural abnormalities including solid nodules, simple cysts, and partially cystic nodules. Diffuse heterogeneity suggests autoimmune thyroiditis. Certain characteristics of nodules—including their number, echogenicity, capsular regularity, vascularity, and patterns of calcification—alter the probability of malignancy but rarely confirm or

TABLE 226-1 CAUSES OF EUTHYROID HYPERTHYROXINEMIA (INCREASED TOTAL T_4, NORMAL TSH, NORMAL FREE T_4)

- Increased synthesis of thyroxine-binding globulin
  - Pregnancy
  - Hepatitis
  - Acute intermittent porphyria
  - Drugs
    - Estrogens
    - Tamoxifen
    - Raloxifene
    - Methadone
    - 5-Fluorouracil
- Increased binding of thyroid hormone to albumin
  - Familial dysalbuminemic hyperthyroxinemia
- Increased binding of thyroid hormone to transthyretin
  - Hereditary variants
  - Pancreatic neuroendocrine tumors

T_4 = thyroxine; TSH = thyroid-stimulating hormone.

TABLE 226-2 CAUSES OF EUTHYROID HYPOTHYROXINEMIA (DECREASED TOTAL T_4, NORMAL TSH, NORMAL FREE T_4)

- Increased metabolism of thyroid hormone
  - Drugs
    - Phenytoin
    - Phenobarbital
    - Carbamazepine
    - Rifampin
- Decreased synthesis of thyroxine-binding globulin
  - Severe liver disease
  - Malnutrition
  - Drugs
    - Androgens
    - Danazol
    - L-Asparaginase
- Increased clearance of thyroxine-binding globulin
  - Nephrotic syndrome
  - Protein-losing enteropathy
- Decreased binding of thyroid hormone to thyroxine-binding globulin
  - Drugs
    - Salicylates (high dose)
    - Phenytoin (high dose)
    - Furosemide (intravenous)

T_4 = thyroxine; TSH = thyroid-stimulating hormone.

exclude thyroid cancer with certainty. Imaging of surrounding structures may identify cervical lymphadenopathy not detectable on physical examination.

The value of computed tomography (CT) and magnetic resonance imaging (MRI) lies in their ability to delineate tracheal deviation, narrowing, and substernal extension of the thyroid into the mediastinum. Cervical CT scanning can also help define and localize regional lymphadenopathy. Positron emission tomography (PET) plays a role in the localization of metastatic thyroid cancer.

Functional Imaging

Radionuclide scanning takes advantage of the fact that gamma ray–emitting tracers transported into thyrocytes by the sodium-iodide symporter can generate images that reflect the regional activity of thyroid tissue. Technetium-99m (^{99m}Tc) and iodine-123 (^{123}I) are commonly used for this purpose. ^{99m}Tc-pertechnetate is rapidly trapped by thyrocytes, and scans obtained using this tracer can be acquired 20 to 30 minutes after injection. ^{123}I and ^{131}I thyroid scans generate images that more precisely reflect thyroid tissue function—for example, indicating whether a nodule is hypofunctioning (cold), hyperfunctioning (hot), or equivalent in function to extranodular tissue (warm). However, radionuclide imaging no longer plays a central role in the differential diagnosis of most thyroid nodules. When needed, ^{123}I is the preferred tracer because of its lower thyroidal and whole body radiation dose.

Thyroid Uptake

The fraction of an administered radioactive iodine or technetium dose taken up and retained by the thyroid gland during a defined period represents an index of the gland's activity. Typically, technetium pertechnetate uptake at 20 minutes ranges from 0.5 to 3%, whereas radioiodine uptake at 24 hours ranges from 8 to 28%. These fractional thyroid uptakes can be useful in the differential diagnosis of thyrotoxicosis. Radioiodine uptake values are also used to calculate effective ^{131}I doses to be administered for the treatment of hyperthyroidism and thyroid cancer.

TABLE 226-3 ETIOLOGIES OF HYPOTHYROIDISM

PRIMARY HYPOTHYROIDISM

- Insufficient functioning thyroid tissue
  - Congenital absence of thyroid tissue
  - Autoimmune destruction of thyroid tissue
    - Autoimmune thyroiditis (Hashimoto's thyroiditis)
  - Surgical removal of thyroid tissue
  - Radioablation of thyroid tissue by radioactive iodine or external beam radiation
  - Infiltrative destruction of thyroid tissue
    - Hemochromatosis
    - Scleroderma
    - Amyloidosis
- Impaired thyroid hormone synthesis
  - Iodine deficiency
  - Congenital enzymatic defects that disrupt thyroid hormone synthesis
  - Drug-mediated inhibition of thyroid hormone production and release
    - Thionamides
    - Amiodarone
    - Lithium
    - Aminoglutethimide
    - Certain tyrosine kinase inhibitors (e.g., sunitinib)

SECONDARY HYPOTHYROIDISM

- Insufficient secretion of TRH or TSH
  - Hypothalamic disorders
    - Tumor (lymphoma, germinoma, glioma)
    - Irradiation
    - Inflammation (sarcoidosis, vasculitis)
  - Hypopituitarism
    - Mass lesions
    - Pituitary surgery
    - Pituitary radiation
    - Hemorrhagic apoplexy (Sheehan syndrome)
    - Infiltration (hemochromatosis, tuberculosis, fungal infection)
    - Lymphocytic hypophysitis
- Thyroid hormone resistance syndrome

HYPOTHYROIDISM

DEFINITION

Primary hypothyroidism (termed *myxedema* when it is severe) refers to hormone deficiency caused by intrinsic thyroid gland dysfunction that disrupts the synthesis and secretion of T_4 and T_3 (Table 226-3). Overt primary hypothyroidism is characterized by an elevated TSH level (usually > 10 mIU/L) in conjunction with a free T_4 level below the lower limit of the reference range. In subclinical hypothyroidism, the TSH level is only modestly elevated; the free T_4 level remains in the low-normal to normal range.

Secondary or central hypothyroidism refers to deficient thyroid gland function that is the result of inadequate stimulation by TSH. This is due in turn to production of either insufficient or inactive TSH from a number of congenital or acquired pituitary and hypothalamic disorders (Chapter 224).

EPIDEMIOLOGY

Primary hypothyroidism is common, occurring in 5% of individuals. Mild hypothyroidism is present in as many as 15% of older adults. Hypothyroidism is more common in women. It is more prevalent among whites and Latin Americans. Secondary hypothyroidism is rare, representing less than 1% of cases.

PATHOBIOLOGY

Dietary iodine deficiency is a cause of primary hypothyroidism in regions where this micronutrient deficiency exists and is uncorrected by iodine supplementation. The most common cause of primary hypothyroidism in developed countries is autoimmune (or Hashimoto's) thyroiditis, a condition in which defective immune tolerance leads to inflammatory destruction of thyroid tissue and impaired gland function.[2] The condition is characterized by a lymphocytic infiltrate and fibrosis. Circulating antithyroid

antibodies directed against thyroid peroxidase and thyroglobulin are markers of the disease, but glandular inflammation is principally the result of altered T-cell-mediated immunity. There is a genetic predisposition to the condition, with linkage studies suggesting a polygenic basis. Patients with autoimmune thyroiditis may have other endocrine and nonendocrine autoimmune disorders. It may be a component of the type 2 polyglandular autoimmune syndrome associated with autoimmune adrenal insufficiency and type 1 diabetes mellitus. It is less commonly a component of the type 1 syndrome, which includes adrenal insufficiency, hypoparathyroidism, and chronic mucocutaneous candidiasis (Chapter 231). Other nonendocrine autoimmune conditions associated with autoimmune thyroiditis include atrophic gastritis, pernicious anemia, systemic sclerosis, Sjögren's syndrome, celiac disease, and vitiligo. Individuals treated with the immunomodulatory agent interferon-α may develop autoimmune thyroiditis with transient or permanent hypothyroidism.

Surgical resection of the thyroid gland predictably leads to hypothyroidism. Radioactive iodine therapy for treatment of hyperthyroidism commonly destroys sufficient thyroid tissue to cause postablative hypothyroidism. External beam radiation therapy for head and neck cancer can also cause thyroid gland failure. Exposure to pharmacologic and radiocontrast agents that contain large amounts of iodine (e.g., amiodarone, radiocontrast dyes, some expectorants, topical disinfectants) can disrupt thyroid hormone production. Lithium inhibits secretion of T_4 and T_3, leading to hypothyroidism in 10% of treated patients. Other pharmacologic agents reported to cause hypothyroidism include stavudine, thalidomide, lenalidomide, imatinib, sunitinib, sorafenib, motesanib, bexarotene, ipilimumab, and aminoglutethimide.

There are a number of other rare causes of primary hypothyroidism (see Table 226-3). Congenital hypothyroidism can be due to agenesis or dysgenesis of the thyroid gland or to mutations in genes encoding the enzymes catalyzing thyroid hormone synthesis. Infiltrative disorders that can disrupt thyroid function include hemochromatosis, amyloidosis, systemic sclerosis, and invasive fibrous thyroiditis (also known as Riedel's thyroiditis). The thyroid gland inflammation that occurs with subacute thyroiditis and painless (postpartum) thyroiditis causes transient hypothyroidism from which most patients recover. Consumptive hypothyroidism can occur in individuals with hemangiomas expressing the type 3 deiodinase, which converts T_4 to biologically inactive reverse T_3.

Secondary or central hypothyroidism may be caused by a number of disorders that impair normal hypothalamic or pituitary control of the thyroid gland (Chapter 224). Infiltrative disorders affecting the hypothalamus that can interfere with TRH secretion include sarcoidosis, hemochromatosis, and histiocytosis. Masses that impinge on the pituitary stalk can impede TRH delivery through the hypophyseal portal system. Compression of thyrotrophic cells by pituitary adenomas and other masses in the sella turcica can inhibit synthesis and secretion of TSH. Surgery and radiation therapy to treat pituitary adenomas can destroy thyrotrophic cells, leading to secondary hypothyroidism that develops as a component of panhypopituitarism. Other disorders associated with secondary hypothyroidism include lymphocytic hypophysitis, pituitary metastases from primary malignant neoplasms, apoplexy, infarction caused by hemorrhage at the time of delivery in women (also known as Sheehan syndrome [Chapter 224]), and head trauma.

CLINICAL MANIFESTATIONS

Symptoms and Signs

Symptoms of hypothyroidism include fatigue, lethargy, weight gain despite poor appetite, cold intolerance, hoarseness, constipation, weakness, myalgias, arthralgias, paresthesias, dry skin, and hair loss. Females may develop precocious puberty, menorrhagia, amenorrhea, and galactorrhea. Affected individuals may experience depressed mood with limited initiative and sociability. Cognitive deficits can range from mild lapses in memory to delirium, dementia, seizures, and coma. The nonspecific nature of most of these symptoms makes it difficult to determine which patients presenting with them have hypothyroidism rather than other causes. Furthermore, in most cases, hypothyroidism is insidious in onset, making its recognition difficult. Symptoms that are new, progressive, or present in combination are more likely to be due to hypothyroidism.

The physical findings associated with hypothyroidism vary according to the age at onset and disease severity. Children may present with delayed linear growth despite weight gain, precocious or delayed puberty, and pseudohypertrophy of muscle. Adults can present with bradycardia, diastolic hypertension, and mild hypothermia. The skin may be coarse, dry, yellow, and cool to the touch as a result of peripheral vasoconstriction. Diffuse thinning of scalp hair accompanied by thinning of the lateral eyebrows may occur. The nails may become brittle. Examination of the chest may reveal distant heart sounds. The extremities may reveal diffuse nonpitting edema caused by the deposition of glycosaminoglycans. Neurologic examination may reveal slow, dysarthric speech and diffuse slowing of deep tendon reflexes with a marked delay in the terminal relaxation phase.

Examination of the neck may reveal a range of findings. Healed cervical incisional scars in this region may indicate a history of surgical resection of thyroid tissue. In autoimmune thyroiditis, the thyroid gland may be normal in size, diffusely enlarged, or atrophic to the degree it may be difficult to palpate at all. It may be soft and smooth with a lobular texture, or firm and irregular with a variegated nodular texture.

Other Routine Test Abnormalities

Routine blood tests may reveal anemia (which is typically macrocytic), hyponatremia, hypoglycemia, and elevated creatine phosphokinase, prolactin, homocysteine, triglyceride, and total and LDL cholesterol levels. Electrocardiography may show sinus bradycardia with low voltage in the limb leads. Chest radiography may show a widened cardiac silhouette, and echocardiography may confirm a pericardial effusion.

DIAGNOSIS

Suspected primary hypothyroidism is confirmed by an elevated TSH level (Fig. 226-1). Established reference ranges for TSH levels typically extend from 0.5 to 4.5 mIU/L. However, the distribution of values within this range is skewed toward the lower half, such that the mean TSH level in adults is

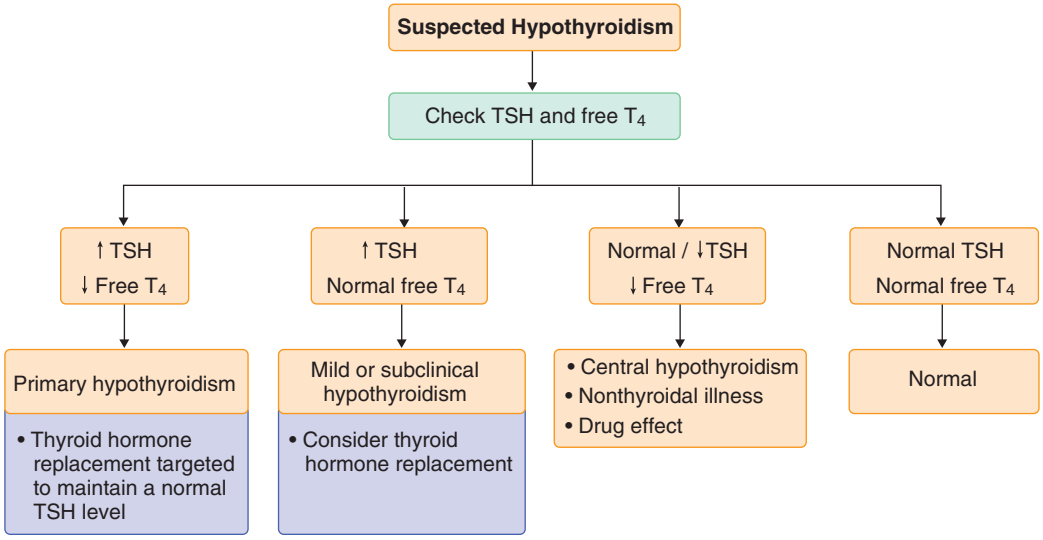

FIGURE 226-1. Laboratory assessment of suspected hypothyroidism. TSH = thyroid-stimulating hormone.

1.5 mIU/L. Measurement of the free T_4 level confirms the diagnosis of primary hypothyroidism and characterizes its severity. A low free T_4 level in conjunction with a persistently elevated TSH level represents overt primary hypothyroidism, whereas a low-normal free T_4 level with an elevated TSH level is termed *mild* or *subclinical primary hypothyroidism*. Other uncommon causes of isolated TSH elevation should be considered in appropriate settings, including recovery from severe systemic illness, renal failure, and adrenal insufficiency.

The underlying cause of primary hypothyroidism is usually clinically obvious, and laboratory testing is unnecessary in most cases. When confirmation is required (e.g., to convince a patient the condition is permanent), serum antithyroid antibodies may be assessed. Measurement of thyroid peroxidase antibody is a more sensitive test than thyroglobulin antibody for this purpose. However, 10% of patients with histologically documented autoimmune thyroiditis have no circulating antithyroid antibodies.

When clinical findings such as the presence of a sellar mass, previous pituitary surgery or irradiation, or other pituitary axis hormone deficiencies suggest the possibility of secondary hypothyroidism, the TSH level cannot be relied on to provide an accurate index of thyroid function. In these settings, the serum free T_4 level must be assessed, and a low or even low-normal free T_4 level can confirm the diagnosis. The TSH level in patients with secondary hypothyroidism can be low, normal, or even modestly elevated.

TREATMENT

Rx

The goals of thyroid hormone replacement therapy are straightforward: to replace endogenous thyroid hormone production, to avoid iatrogenic thyrotoxicosis, and (rarely) to treat systemic complications of severe hypothyroidism. Levothyroxine sodium (hereafter thyroxine) is the hormonal preparation of choice.[3] Thyroxine has a number of favorable pharmacokinetic characteristics. It is well absorbed, and its plasma protein binding gives it a 7-day half-life, permitting daily dosing. Thyroxine is physiologically deiodinated to the more biologically active T_3 in peripheral tissues. However, thyroxine has a narrow therapeutic index, and doses differing by as little as 12% can have clinical consequences. Tablets of multiple dose strengths ranging from 25 to 300 µg are available. Regulatory standards ensure pharmaceutical equivalence in terms of mass of thyroxine, but bioavailability may differ by as much as 12% among different preparations. Consequently, adherence to a single thyroxine formulation is advisable.

The optimal dose of thyroxine for replacement therapy is related to lean body weight, with most adults requiring a daily dose of approximately 1.8 µg/kg. The dose requirement for elderly adults is typically lower (e.g., 1 µg/kg/day) because of slower metabolic clearance. Patients with postsurgical or postablative hypothyroidism usually require a higher daily dose than patients with autoimmune thyroiditis, who may have some residual gland function. Patients with coexisting malabsorptive disorders may require higher and variable doses. Certain medications, mineral supplements, and foods can interfere with thyroxine absorption, including ferrous sulfate, calcium carbonate, aluminum hydroxide, sucralfate, cholestyramine, and soy-containing foods (Table 226-4). Thyroxine doses should be separated from these substances by 8 hours or longer.

Thyroxine dose requirements may increase as a result of accelerated metabolic clearance in several circumstances. Patients with nephrotic syndrome and other systemic illnesses that lead to rapid clearance of thyroid hormone require higher daily doses. Dose requirements increase by an average of 75% in most pregnant women as a result of placental deiodinative metabolism of thyroxine.[4] Simultaneous treatment with phenytoin, phenobarbital, carbamazepine, or rifampin also typically accelerates thyroxine metabolism.

Most adults without known or suspected coronary artery disease can be started on a full replacement dose of thyroxine. The initial dose can be calculated on the basis of the patient's weight and age, rounding down to the nearest available dose strength. For patients with primary hypothyroidism, adequacy of thyroxine therapy can be assessed by TSH measurement 4 to 6 weeks after therapy is started. The target TSH level for most treated individuals should be the lower half of the reference range (i.e., 0.5 to 2.0 mIU/L). Once an adequate dose has been established, the TSH level should be checked annually. In patients with secondary hypothyroidism, the serum free T_4 level should be monitored 2 to 4 weeks after the thyroxine dose is started or adjusted, with a target free T_4 level in the upper half of the reference range.

Management of Complications

Complications of thyroxine therapy are limited to iatrogenic thyrotoxicosis and, rarely, adverse effects of restoring euthyroidism. Typical symptoms and signs of thyrotoxicosis usually accompany significant degrees of overtreatment. However, even a modestly excessive thyroxine dose can induce bone mineral loss, especially in postmenopausal women, and it can increase the risk of atrial fibrillation in older individuals. In patients with underlying coronary artery disease, the positive chronotropic and inotropic effects of thyroxine may exacerbate myocardial ischemia.[5] Consequently, adults with known or suspected ischemic heart disease should be started on a low dose that is titrated upward in small increments once tolerance is ensured (e.g., starting with 25 µg daily, then increasing the dose by 12.5 to 25 µg every 4 to 6 weeks). In some cases, β-blocker therapy may need to be intensified to counter the induction of myocardial ischemia. However, deliberate suboptimal dosing of thyroxine should be avoided. If necessary, coronary revascularization may be required before euthyroidism can be fully restored. Coexisting adrenal insufficiency associated with hypopituitarism or the type 2 polyglandular autoimmune syndrome may be unmasked when cortisol clearance is accelerated by a return to the euthyroid state. Other adverse effects that infrequently occur with thyroxine therapy include transient hair loss, acute sympathomimetic symptoms that resolve with dose reduction and slow advancement, and pseudotumor cerebri in children.

A minority of patients with thyroxine-treated hypothyroidism continue to report bothersome symptoms despite biochemical evidence of adequate thyroid hormone replacement. Several randomized clinical trials have shown that combinations of T_3 and T_4—in the form of desiccated thyroid or synthetic thyroid hormone preparations—are not superior to T_4 alone.

Subclinical and Mild Hypothyroidism

Whether individuals diagnosed with subclinical hypothyroidism (i.e., an elevated or high-normal TSH level with a free T_4 level within the reference range) benefit from thyroxine therapy remains controversial.[6] In practice, many providers opt for a trial of therapy in mildly hypothyroid patients who are symptomatic, have underlying hypercholesterolemia, or have a high likelihood of progressing to overt hypothyroidism. Predictors of progressive thyroid failure include age older than 65 years, TSH level higher than 10 mIU/L, and the presence of circulating thyroid autoantibodies, indicating underlying autoimmune thyroiditis.

Myxedema Coma

Severe hypothyroidism can culminate in myxedema coma, a life-threatening condition characterized by hypothermia, bradycardia, hypotension, altered mental status, and multisystem organ failure. Risk factors include advanced age, poor access to health care, and other underlying major organ system diseases. Most patients have severe and long-standing thyroid hormone deficiency. Treatment should include thyroxine (1.8 µg/kg/day, with or without a 500-µg loading dose). Some experts advocate coadministration of triiodothyronine in divided doses to compensate for impaired conversion of T_4 to T_3. No controlled trials have been performed to evaluate the relative benefits and risks of these different approaches. Glucocorticoids should be administered in stress doses after a cosyntropin stimulation test has been performed to check for evidence of concomitant adrenal insufficiency (Chapter 227). Care should be taken to avoid exposure to potent sedative or analgesic agents that may exacerbate altered mental status. Hypothermia should be treated with external warming to reduce the risk of circulatory collapse.

TABLE 226-4 INTERFERENCE WITH THYROXINE REPLACEMENT THERAPY

FACTORS CONTRIBUTING TO UNDERREPLACEMENT
Inadequate prescribed dose
Limited compliance
Decreased absorption due to ingestion of agents that bind thyroxine
Ferrous sulfate
Calcium carbonate
Aluminum hydroxide
Sucralfate
Cholestyramine
Soy protein
Increased metabolism of thyroxine
Pregnancy
Drugs
Phenytoin
Phenobarbital
Carbamazepine
Rifampin
Diminishing residual thyroid function
Changing formulations
FACTORS CONTRIBUTING TO OVERREPLACEMENT
Excessive prescribed dose
Factitious ingestion of additional doses
Decreased metabolism of thyroxine due to aging
Increasing residual thyroid function
Changing formulations

Nonthyroidal Illness

In patients with severe nonthyroidal illness, a characteristic constellation of thyroid function test changes occurs that often appears to be consistent with hypothyroidism (see Fig. 226-1).[7] The T_3 level usually declines as a result of decreased extrathyroidal T_4-to-T_3 conversion. With increasingly severe disease, total T_4 and free T_4 levels also decline. TSH levels are usually low to low-normal. During the course of recovery, the TSH level can rise above the upper limit of the normal range, producing a profile that can be mistaken for primary hypothyroidism. Clinical correlation is essential to assess thyroid function in severely ill patients (e.g., a history of preexisting thyroid or pituitary disease, the presence of a goiter, or features suggesting other elements of hypopituitarism). Because no benefit of thyroid hormone treatment has been shown for these patients, observation with retesting 6 to 8 weeks after recovery is the preferred approach.

THYROTOXICOSIS

DEFINITION AND EPIDEMIOLOGY

Thyrotoxicosis is a systemic syndrome caused by exposure to excessive thyroid hormone (Table 226-5). Its prevalence is 1 in 2000 adults, affecting 1% of all individuals during the course of their lifetime.

PATHOBIOLOGY

Thyrotoxicosis is the result of excessive circulating and tissue effects of thyroid hormone. Strictly speaking, hyperthyroidism refers to those forms of thyrotoxicosis that are caused by excessive production of thyroid hormone by the thyroid gland due to a thyrotropic stimulus or autonomous thyroid tissue function (see Table 226-5). In Graves disease, the most common cause of hyperthyroidism, the thyroid gland is stimulated by autoantibodies that bind to and activate the TSH receptor. Excessive secretion of TSH causes hyperthyroidism in patients with rare TSH-secreting pituitary adenomas (Chapter 224). CG, a glycoprotein with high TSH homology, can cause transient gestational hyperthyroidism during pregnancy, when a choriocarcinoma or a germ cell tumor produces variant forms of HCG that are more active or when mutant TSH receptors bind HCG more avidly, as occurs in familial gestational thyrotoxicosis.

Autonomous production of thyroid hormone occurs when thyrocytes function independently of TSH receptor activation. This can occur as a result of growth of a benign functioning thyroid adenoma or growth of multiple autonomously functioning nodules forming a toxic multinodular goiter. In rare cases, it can occur when patients with well-differentiated thyroid cancer present with functioning metastases. In some toxic adenomas, somatic mutations in the TSH receptor gene lead to constitutive activation. In patients whose thyroid glands have the potential for autonomous function, exposure to excessive amounts of iodine in the form of amiodarone or iodinated contrast agents can provoke hyperthyroidism.

Transient thyrotoxicosis can also be caused by inflammatory conditions that release an excessive amount of thyroid hormone stored in the gland (see the section on thyroiditis). These include subacute thyroiditis, which is believed to be caused by a viral infection; acute or suppurative thyroiditis, caused by bacterial infection; radiation-induced thyroiditis; and pharmacologic thyroiditis (e.g., due to amiodarone). Autoimmunity can also provoke an inflammatory thyroiditis that causes transient thyrotoxicosis. This commonly occurs in the setting of lymphocytic thyroiditis (also known as silent, painless, or postpartum thyroiditis). It rarely occurs in the setting of autoimmune thyroiditis (also known as Hashimoto's thyroiditis).

In rare cases, excess thyroid hormone can be secreted by ectopic thyroid tissue located anywhere from the base of the tongue to the mediastinum, or by heterotopic thyroid tissue that develops as part of an ovarian teratoma (a condition known as struma ovarii).

Thyrotoxicosis can also be caused by ingestion of excessive amounts of thyroid hormone. This is most often the result of the prescription of excessive doses of pharmacologic preparations of thyroid hormone, but it can rarely be due to surreptitious or accidental ingestion.

TABLE 226-5 ETIOLOGIES OF THYROTOXICOSIS

HYPERTHYROIDISM
Antibody-mediated stimulation of thyroid tissue
Graves disease
Autonomously functioning thyroid tissue
Toxic multinodular goiter
Toxic adenoma
Iodine exposure
Autonomously functioning heterotopic thyroid tissue
Struma ovarii
Metastatic differentiated thyroid cancer
Excessive secretion of TSH
TSH-secreting pituitary adenoma
NONHYPERTHYROID THYROTOXICOSIS
Ingestion of exogenous thyroid hormone
Pharmacologic
Levothyroxine
Liothyronine
Combination preparations
Nonpharmacologic
Dietary supplements
Improperly processed meat products
Inflammation causing release of endogenous thyroid hormone
Subacute thyroiditis
Autoimmune thyroiditis

TSH = thyroid-stimulating hormone.

CLINICAL MANIFESTATIONS

Symptoms and Signs

The classic symptoms of thyrotoxicosis include weight loss despite a hearty appetite, heat intolerance, palpitations, tremor, and hyperdefecation (increased frequency of formed bowel movements). Thyrotoxicosis can escape early detection because of its presentation with common nonspecific symptoms such as fatigue, insomnia, anxiety, irritability, weakness, atypical chest pain, or dyspnea on exertion. Delayed recognition may also occur when atypical symptoms such as headache, weight loss, periodic paralysis, or nausea and vomiting dominate the clinical picture. Elderly patients may present with apathetic thyrotoxicosis typified by weight loss and the absence of sympathomimetic symptoms and signs.

Signs of thyrotoxicosis include resting tachycardia, systolic hypertension with a widened pulse pressure, warm moist skin with a velvety texture, onycholysis, and a staring gaze with lid lag (noted to be present when a rim of sclera is visible between the upper eyelid and the superior margin of the iris on downward gaze). Cardiac examination may reveal a prominent apical impulse and a systolic flow murmur. Neurologic findings may include a restless, impatient demeanor, pressured speech, proximal muscle weakness, distal hand tremor, and brisk deep-tendon reflexes.

Clinical findings often provide clues to the underlying cause.[8] In Graves disease, the gland is diffusely enlarged with a smooth or slightly lobulated contour, and may manifest an audible bruit or palpable thrill. Thyroid ophthalmopathy and dermopathy are also unique to Graves disease. In patients with toxic nodular goiter, one or more discrete nodules may be appreciated. In subacute thyroiditis, the gland is modestly enlarged, extremely tender, and firm. A history of recent pregnancy suggests possible painless thyroiditis. Recent exposure to amiodarone, other iodine-containing compounds, interferon-α, or pharmacologic preparations of thyroid hormone may suggest the characteristic forms of thyrotoxicosis associated with these agents.

Graves Disease

DEFINITION

Graves disease is an autoimmune disorder characterized by a variable combination of hyperthyroidism, ophthalmopathy (also known as thyroid eye disease), and dermopathy.

EPIDEMIOLOGY

Graves disease is more common among women, but it also affects men. It can develop at any time during life, but the onset most often occurs between 30 and 60 years of age.

PATHOBIOLOGY

The proximate cause of hyperthyroidism in Graves disease is the production of thyroid-stimulating immunoglobulins that bind to and activate the TSH receptor, promoting thyroid hormone secretion and gland growth. Thyrotropin (TSH) receptor antibodies of the stimulating variety are the hallmark of hyperthyroidism in Graves disease. Other thyroid autoantibodies commonly identified in the setting of Graves disease include thyroid peroxidase

antibodies, thyroglobulin antibodies, and TSH receptor antibodies. Although the fundamental cause of Graves disease remains unknown, a genetic predisposition is implicated by a higher incidence in monozygotic twins and first-degree relatives of affected individuals. Environmental factors implicated in triggering the onset of Graves disease include exposure to cigarette smoke, high dietary iodine intake, and perhaps stressful life events and certain antecedent infections.

CLINICAL MANIFESTATIONS

Affected individuals usually present with thyrotoxicosis and a thyroid gland that is diffusely enlarged with a rubbery consistency, smooth contour, definable pyramidal lobe, and audible bruit or palpable thrill due to increased blood flow. When it is clinically evident, thyroid eye disease usually presents within a few months of onset. In rare cases, it may develop long before, long after, or without any biochemical confirmation of hyperthyroidism.

PROGNOSIS

The hyperthyroidism associated with this condition often follows a persistent and progressive course, but one fourth of patients with Graves disease demonstrate spontaneous disease remission.

OPHTHALMOPATHY

DEFINITION

Thyroid eye disease is a distinctive disorder characterized by inflammation and swelling of the extraocular muscles and orbital fat, eyelid retraction, periorbital edema, episcleral vascular injection, conjunctival swelling (chemosis), and proptosis (also called exophthalmos).[9] Swelling of soft tissues within the confines of the orbits precipitated by fibroblast growth and inflammatory cell infiltrate can cause proptosis, entrapment of extraocular muscles, and compression of the optic nerve.

CLINICAL MANIFESTATIONS

Affected individuals typically complain of a change in eye appearance, ocular irritation, foreign body sensation, dryness, and ironically, excessive tearing. More severe involvement may cause exposure keratitis with corneal ulceration, diplopia, and blurred vision. On examination, patients may have a staring gaze, a rim of sclera visible between the upper eyelid and the superior margin of the iris during downward gaze (lid lag), signs of conjunctival inflammation, periorbital edema, and abnormalities of conjugate gaze, color vision, and visual acuity (Fig. 226-2). The precise degree of proptosis can be measured with an exophthalmometer. Orbital imaging with CT scanning or ultrasonography can confirm the diagnosis, which must sometimes be distinguished from other causes of bilateral and unilateral proptosis.

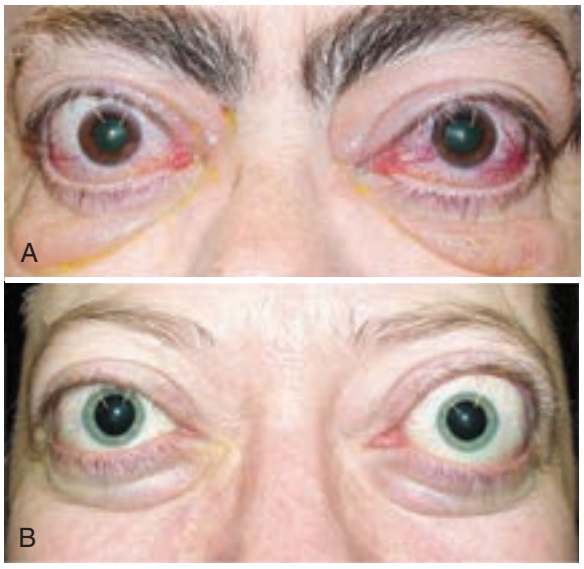

FIGURE 226-2, Graves ophthalmopathy. A, A 59-year-old woman with excess proptosis, moderate eyelid edema, and erythema with moderate eyelid retraction affecting all four eyelids. Conjunctival chemosis (edema) and erythema with bilateral edema of the caruncles, with prolapse of the right caruncle, are evident. **B,** A 40-year-old woman with excess proptosis, minimal bilateral injection, and chemosis with slight erythema of the eyelids. On slit lamp examination, she also had evidence of moderate superior limbic keratoconjunctivitis. (From Bahn RS. Graves' ophthalmopathy. *N Engl J Med.* 2010;362:726-738. Copyright 2010, Massachusetts Medical Society. All rights reserved.)

TREATMENT

Rx

Treatment of mild thyroid eye disease focuses on protecting the cornea from exposure and desiccation with moisturizing drops and ointment, glasses, and sometimes taping the eyelids closed at bedtime. Selenium supplementation may help to relieve some of the symptoms associated with active inflammation in mild to moderate cases. High-dose systemic glucocorticoid therapy can attenuate orbital inflammation in more severe cases. Orbital irradiation may be helpful in controlling inflammatory symptoms in some patients. Persistent corneal exposure, diplopia, altered vision due to optic nerve compression, and cosmetic issues may require surgery to decompress the orbits and readjust the extraocular muscles. Immunosuppressive agents and plasmapheresis have been used in severely affected patients, with anecdotal success.

DERMOPATHY

Infiltrative dermopathy, the least common aspect of Graves disease, is precipitated by the deposition of glycosaminoglycans in the dermis of the skin. Affected individuals usually present with mildly pruritic, orange peel–like thickening of the skin along the anterior aspects of the shins, known as pretibial myxedema. The dorsal aspects of the feet and fingers, the extensor surface of the elbows, and the face are more rarely affected.

The diagnosis can be confirmed by skin biopsy. Treatment of early infiltrative dermopathy with topical glucocorticoids under an occlusive wrap may limit its progression. Treatments involving the use of intradermal or systemic glucocorticoids, long-acting somatostatin analogues, and even surgical resection of soft tissue have demonstrated limited success.

Toxic Adenoma

A toxic adenoma is a solitary, autonomously functioning thyroid neoplasm that synthesizes and secretes excessive amounts of thyroid hormone independent of TSH stimulation. These neoplasms are almost always benign. Most grow large enough to be palpated by the time they present with thyrotoxicosis. Somatic gene mutations causing constitutive activation of the TSH receptor and the α-subunit of the stimulatory guanine nucleotide binding protein ($G_s\alpha$) have been identified in a subset of toxic adenomas. Hyperthyroidism caused by a toxic adenoma does not remit spontaneously, except in unusual cases complicated by hemorrhagic infarction of the neoplasm.

Toxic Multinodular Goiter

A toxic multinodular goiter is composed of multiple autonomously functioning thyroid nodules that synthesize and secrete excessive amounts of thyroid hormone. In some patients with nontoxic multinodular goiters, hyperthyroidism can be precipitated by exposure to excessive amounts of iodine. Most affected individuals have a goiter with multiple palpable thyroid nodules. Progressive enlargement may go undetected when there is substernal extension of nodular tissue. Toxic multinodular goiters are more common among older individuals.

TSH-Secreting Pituitary Adenoma

TSH-secreting pituitary adenomas represent less than 1% of all functioning pituitary tumors (Chapter 224). Patients may present with typical clinical manifestations of thyrotoxicosis, a diffuse goiter, symptoms and signs precipitated by an expanding sellar mass, syndromes associated with co-secretion of other anterior pituitary hormones (growth hormone, prolactin, or adrenocorticotropic hormone), or symptoms and signs of hypopituitarism. The key to suspecting the condition is usually recognition of an inappropriately nonsuppressed TSH level in a patient with thyrotoxicosis. The diagnosis is confirmed in most cases when laboratory testing reveals an elevated circulating level of the pituitary glycoprotein α-subunit in conjunction with a radiographically definable sellar mass.

DIAGNOSIS

Laboratory Findings

Abnormalities detected in routinely ordered laboratory tests are often the first clues to the presence of thyrotoxicosis. Thyrotoxic patients may have hypercalcemia or hypercalciuria, increased alkaline phosphatase levels, modestly elevated transaminase levels, and low or declining total and LDL

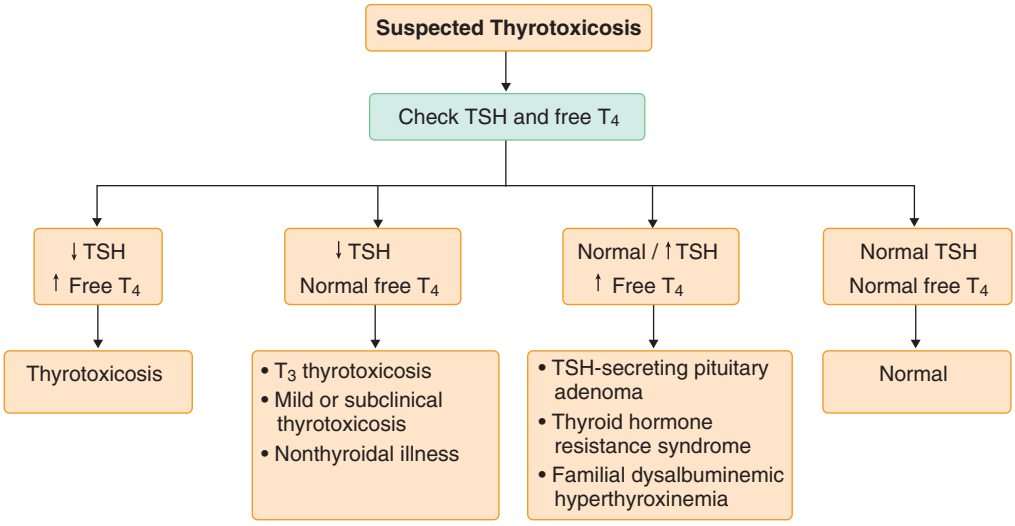

FIGURE 226-3. **Laboratory assessment of suspected thyrotoxicosis.** TSH = thyroid-stimulating hormone.

cholesterol levels. When they are measured, ferritin and angiotensin-converting enzyme levels are often increased. Electrocardiography typically reveals resting sinus tachycardia or atrial tachyarrhythmias, particularly atrial fibrillation with a rapid ventricular response. In severe cases, chest radiography may reveal cardiomegaly.

In most patients with suspected thyrotoxicosis, the diagnosis can be confirmed by measurement of a TSH level (Fig. 226-3). Sensitive TSH immunoassays with a detection limit of less than 0.02 mIU/L can accurately discriminate between clearly suppressed TSH levels, characteristic of all common forms of thyrotoxicosis, and mildly suppressed levels that fall just beneath the reference range, as may occur in otherwise sick individuals. Only the rare conditions associated with TSH-mediated hyperthyroidism (TSH-secreting pituitary tumors and isolated pituitary resistance to thyroid hormone) lack TSH suppression when testing for thyrotoxicosis. Measurement of serum free T_4 and T_3 levels confirms the diagnosis of thyrotoxicosis, defines its severity, and occasionally provides a clue to its underlying cause. Overt thyrotoxicosis is characterized by free T_4 or T_3 levels above the upper limit of the reference range, whereas mild or subclinical thyrotoxicosis is characterized by a suppressed TSH level with free T_4 and T_3 levels within the normal reference range. When only the free T_4 or T_3 concentrations are elevated, the terms *T_4 toxicosis* and *T_3 toxicosis* are applied, respectively.

Differential Diagnosis

Once thyrotoxicosis is confirmed, it is important to define its underlying cause to determine the most appropriate course of treatment. The relative degrees of T_4 and T_3 elevation sometimes can be helpful. Predominantly T_3 toxicosis is typical of Graves disease and can also occur with toxic nodular goiter. In contrast, predominantly T_4 toxicosis is more typical of subacute or painless thyroiditis. T_4 toxicosis is also more common in patients with iodine-induced hyperthyroidism.

Other laboratory tests are sometimes helpful in differential diagnosis. Antithyrotropin-receptor antibodies are pathognomonic of Graves disease. Levels of antithyrotropin-receptor antibodies are especially high in thyroid dermopathy and correlate positively with the clinical features and prognosis of Graves ophthalmopathy. An elevated ESR is typically seen in subacute thyroiditis.

Imaging studies can be helpful for the differential diagnosis. The fractional thyroidal uptake of radiotracer by the thyroid and its distribution in the gland on scintigraphic scanning often helps establish a definitive diagnosis (Table 226-6). Thyroid ultrasonography can confirm the presence of solitary or multiple thyroid nodules. Chest radiography and CT scanning may help delineate a substernal goiter.

TABLE 226-6 RADIOGRAPHIC EVALUATION OF SUSPECTED THYROTOXICOSIS

ETIOLOGY	FRACTIONAL 24-HOUR RADIOIODINE UPTAKE (%)	THYROID SCAN APPEARANCE
Graves disease	35-95	Diffuse increased homogeneous uptake; visible pyramidal lobe extending from isthmus
Toxic adenoma	20-60	Solitary focus of intense uptake; suppression of uptake in remainder of thyroid
Toxic multinodular goiter	20-60	Patchy heterogeneous foci of increased uptake interspersed with regions of diminished uptake
Subacute thyroiditis	0-2	Minimal to absent uptake
Autoimmune thyroiditis	0-2	Minimal to absent uptake; patchy heterogeneous uptake during recovery
Iodine-induced hyperthyroidism	0-2	Minimal to absent uptake
Exogenous thyroid hormone intoxication	0-2	Minimal to absent uptake
Metastatic differentiated thyroid cancer	0-5	Focal uptake in metastases
TSH-secreting pituitary adenoma	30-80	Diffuse increased homogeneous uptake

TSH = thyroid-stimulating hormone.

TREATMENT

Rx

Selection of the most effective treatment for a specific condition causing thyrotoxicosis requires an understanding of the underlying pathophysiologic process and natural history. For example, toxic multinodular goiter does not remit and requires definitive radioiodine treatment or surgery; subacute thyroiditis subsides spontaneously and requires only temporizing symptomatic therapy.

β-Blockers

β-Blockers help alleviate the sympathomimetic manifestations of thyrotoxicosis, regardless of the underlying cause. Palpitations, tremor, and anxiety can often be promptly controlled. However, other clinical features of thyrotoxicosis, including weight loss, heat intolerance, and fatigue, are not ameliorated by these agents. In thyrotoxic patients with marked sinus tachycardia or atrial fibrillation with a rapid ventricular response rate, β-blockers can be used as rate-controlling agents. Propranolol also partially inhibits extrathyroidal conversion of T_4 to T_3, which may be of added benefit in patients with severe thyrotoxicosis.

Propranolol can be started at a dose of 20 to 40 mg every 8 hours and titrated upward to a maximal daily dose of 240 mg on the basis of symptom control. Sustained-release propranolol or longer-acting β-blockers, such as

metoprolol and atenolol, can also be used. β-Blockers should be used with caution in thyrotoxic patients with a history of obstructive pulmonary disease, Raynaud's phenomenon, or heart failure. Esmolol can be used when a short-acting parenteral agent is required for heart rate control in patients with thyrotoxic heart failure.

For patients with transient forms of thyrotoxicosis (subacute thyroiditis, autoimmune thyroiditis, or exogenous thyroid hormone intoxication), a β-blocker may be the only treatment required. In patients with more sustained conditions, such as Graves disease or toxic nodular goiter, β-blockers provide prompt initial relief of symptoms while definitive treatment with antithyroid drugs, radioiodine, or surgery is implemented.

Antithyroid Drugs

The thionamides inhibit thyroid hormone biosynthesis by competitively inhibiting iodine organification and iodotyrosine coupling. These agents are used for the treatment of thyrotoxicosis caused by overproduction of thyroid hormones. Because the thionamides block only new thyroid hormone synthesis, glandular stores of preexisting thyroid hormone must be exhausted before they are fully effective. This may require 3 to 8 weeks in patients with Graves disease or toxic multinodular goiter. Although antithyroid drugs can provide long-term control of hyperthyroidism, they are most appropriately used when there is a possibility that the underlying condition will remit, as in Graves disease, or when thyrotoxicosis must be attenuated before radioiodine treatment or surgery.

Two thionamide agents are currently available: methimazole and propylthiouracil. Methimazole can be taken as a single daily dose because of its longer half-life and higher effective intrathyroidal concentration. This can bolster patients' adherence and drug effectiveness. Propylthiouracil also inhibits extrathyroidal conversion of T_4 to T_3, an effect that may be beneficial in patients with severe complicated thyrotoxicosis. Propylthiouracil is preferred for pregnant hyperthyroid women in the first trimester because methimazole has been rarely associated with the congenital anomalies of choanal atresia and cutis aplasia.[10] However, the shorter half-life of propylthiouracil necessitates its administration three or four times daily. Furthermore, risk of severe hepatotoxicity associated with the use of propylthiouracil has prompted the recommendation that methimazole be the first-line antithyroid drug to treat hyperthyroidism in children and adults, including women after the first trimester of pregnancy.

For patients with mild to moderate hyperthyroidism, methimazole is usually started at a dose of 10 to 30 mg once daily and increased to as much as 90 mg daily. For patients with more severe hyperthyroidism, thyrotoxicosis complicated by cardiac disease, or concomitant pregnancy, propylthiouracil can be started at a dose of 50 to 200 mg every 6 to 8 hours. Methimazole can be given rectally if necessary. The anticipated duration of treatment depends on the underlying cause. In patients with toxic multinodular goiter, antithyroid drugs are generally used only to restore euthyroidism in anticipation of definitive therapy. An effective dose can be continued for 6 to 24 months in a patient with Graves disease, before it is tapered off to determine whether there has been a remission of the patient's autoimmune thyroid disease. Patients most likely to respond are those who present with mild clinical and biochemical hyperthyroidism, a small thyroid gland, and no active ophthalmopathy.

Patients treated with antithyroid drugs should have thyroid function tests checked every 3 to 12 weeks during dose titration to monitor for iatrogenic hypothyroidism. Common side effects include rash, pruritus, fever, and arthralgias, which affect 5% of thionamide-treated patients. Agranulocytosis and hepatitis are rare but potentially fatal adverse reactions to thionamide medications. Their presentations are relatively sudden in onset and unpredictable. Monitoring of leukocyte counts and liver function test results is not useful as a preventive measure. Patients who are prescribed antithyroid drugs should be cautioned about manifestations of these adverse reactions and should be instructed to discontinue treatment and seek medical attention if they develop a high fever, pharyngitis, jaundice, or abdominal pain.

Radioactive Iodine

The selective uptake and concentration of iodide in thyrocytes permits the use of radioactive iodine to treat hyperthyroidism.[11] Once it is concentrated in the gland after oral administration, ^{131}I destroys thyroid tissue and controls hyperthyroidism, usually within 1 to 2 months. The dose of ^{131}I can be calculated on the basis of the fractional uptake of radioiodine, but the outcome of dosimetry is not superior to that achieved with the administration of empirical doses. Patients can be treated on an outpatient basis, with precautions taken to prevent exposure of others. Approximately three quarters of patients are cured with a single dose of radioiodine.

The principal side effect of radioactive iodine therapy is postablative hypothyroidism, which develops in most individuals receiving treatment for Graves disease and in a lesser proportion of patients treated for toxic nodular goiter. Lifelong monitoring of thyroid function is required because patients develop postablative hypothyroidism at a rate of 3% per year. Another less common complication is a transient exacerbation of thyrotoxicosis, which occurs in one quarter of patients during the first month after treatment as a result of radiation thyroiditis. Long-term follow-up studies have shown that radioiodine-treated patients with Graves disease do not have any greater risk of thyroid cancer or other malignant neoplasms. However, hyperthyroid children and adolescents treated with radioactive iodine are more likely to develop benign nodules. Among hyperthyroid women treated with radioiodine, the incidences of infertility, spontaneous abortion, and children with birth defects are not increased. Diagnostic or therapeutic radioactive iodine is contraindicated in women during pregnancy, and treated women should avoid pregnancy until euthyroidism has been confirmed 3 to 6 months after administration of a dose.

Other Drugs

Saturated solution of potassium iodide (SSKI) or Lugol's solution transiently inhibit the synthesis and release of thyroid hormone from the gland. They may be used to accelerate recovery after radioactive iodine treatment, to prepare patients for thyroidectomy, and to augment other treatments used to control severe thyrotoxicosis (see later). Iodinated radiocontrast agents inhibit the release of thyroid hormone while blocking peripheral conversion of T_4 to T_3. Lithium carbonate also inhibits the release of thyroid hormone. Rarely, these agents are used in combination with thionamides to treat patients with severe thyrotoxicosis. They may also help provide temporary control of hyperthyroidism when severe allergies preclude the continued use of thionamides. Cholestyramine can be used to bind thyroid hormone in the gut to interrupt enterohepatic circulation in cases of suspected exogenous thyroid hormone intoxication.

Surgery

Surgery has a limited role because of its potential to injure the adjacent recurrent laryngeal nerves and parathyroid glands. Resection of a toxic adenoma by lobectomy is curative and often preserves sufficient normal thyroid tissue for euthyroidism to be maintained. Consequently, it is often recommended in younger individuals. Toxic multinodular goiters causing compressive symptoms or cosmetic disfigurement may be appropriately managed with surgical resection. Although surgery is seldom recommended in the United States for the treatment of hyperthyroid Graves disease, it may be appropriate when other modalities are contraindicated, such as when there has been an adverse reaction to an antithyroid drug in pregnancy, when a thyroid nodule is thought to be malignant, or when hyperparathyroidism also requires surgical intervention.

Specific Treatment Scenarios

Pregnancy

Pregnant patients with hyperthyroidism present special challenges. Diagnosis requires a careful assessment of symptoms, especially heat intolerance, palpitations, and vomiting, which also occur during normal pregnancy. The serum total T_4 level is elevated because of increased TBG, and the TSH level can be suppressed in the first trimester as a result of HCG-mediated thyroid stimulation. Diagnostic radionuclide imaging studies are contraindicated. After diagnostic confirmation, hyperthyroidism must be treated because it is associated with an increased risk of spontaneous abortion, premature labor, low birth weight, and toxemia. β-Blockers should be used only transiently to control severe symptoms. Propylthiouracil is the preferred thionamide for treatment of Graves disease during the first trimester of pregnancy, because it crosses the placenta less readily than methimazole and because methimazole has been rarely linked to congenital malformations (i.e., choanal atresia and cutis aplasia). However, owing to the risk of very rare but potentially fatal propylthiouracil-related hepatitis, methimazole is preferred after the first trimester. Because Graves disease often remits later in pregnancy, antithyroid drug dose requirements often decline as gestation progresses. Measurement of maternal thyroid-stimulating immunoglobulin levels can help predict the risk of an infant developing neonatal Graves disease.

Subclinical and Mild Hyperthyroidism

Patients with subclinical or mild hyperthyroidism (i.e., a suppressed serum TSH with normal free T_4 and T_3 levels) may have symptoms that justify treatment. In patients with a serum TSH level suppressed to less than 0.1 mIU/L, bone mineral loss can lead to osteoporosis, particularly in postmenopausal women. Atrial fibrillation occurs more commonly in mildly hyperthyroid patients aged 60 years and older with TSH suppression below normal. It is less clear, however, whether younger asymptomatic patients with modestly suppressed TSH levels (e.g., 0.1 to 0.5 mIU/L) require anything more than periodic monitoring.

Thyrotoxic Crisis

Thyrotoxic crisis, also known as thyroid storm, is a potentially life-threatening syndrome that is usually the end result of severe and sustained thyrotoxicosis. It can affect patients with other medical conditions that render them vulnerable to the cardiovascular, neuropsychiatric, and gastrointestinal effects of exposure to excessive amounts of thyroid hormone. Thyrotoxic crisis typically develops in the setting of inadequately treated Graves disease and may be precipitated by intercurrent illness, surgery, or treatment with radioactive iodine. Affected individuals present with fever, atrial tachyarrhythmias,

congestive heart failure, nausea and vomiting, diarrhea, and seizures. Mental status changes can include agitation, delirium, psychosis, and coma. Prompt recognition and treatment in a monitored setting are crucial. A multifaceted treatment regimen should incorporate antipyretics, β-blockers, thionamides, iodinated contrast agents, and glucocorticoids, as well as aggressive evaluation and management of underlying medical problems.

THYROIDITIS

Subacute (de Quervain's) Thyroiditis

PATHOBIOLOGY

Transient thyrotoxicosis results from the uncontrolled release of thyroid hormone from the inflamed gland. After 2 to 8 weeks, when the supply of stored hormone is exhausted, thyrotoxicosis resolves spontaneously. Hypothyroidism ensues because the gland's biosynthetic capabilities remain impaired. This is also transient (lasting ≈ 1 month), with subsequent restoration of normal thyroid function in most patients.

CLINICAL MANIFESTATIONS

Subacute thyroiditis is characterized by painful enlargement of the thyroid, systemic inflammatory symptoms, and transient thyrotoxicosis that is often followed by transient hypothyroidism. The histologic pattern shows inflammatory cell infiltrates that are believed to be the result of a viral infection. Many patients with subacute thyroiditis report antecedent upper respiratory infections.

Patients usually present with pain localized to the thyroid or radiating to the throat, ears, or jaw. Constitutional symptoms, including fever, chills, sweats, and malaise, are often present. On occasion, these inflammatory features may dominate the presentation. Examination of the thyroid typically reveals an exquisitely tender, modestly enlarged, and woody, hard gland.

DIAGNOSIS

Differential Diagnosis

The differential diagnosis of thyroid pain must be considered in the evaluation of patients presenting with pain and tenderness localized to the lower anterior neck. In addition to subacute thyroiditis, potential causes of thyroid pain include acute (suppurative) thyroiditis, hemorrhage into an existing thyroid nodule, and rapid growth of anaplastic thyroid cancer, diffusely infiltrating thyroid cancer, or thyroid lymphoma.

Laboratory Findings

Laboratory testing in patients with subacute thyroiditis reveals a profile of overt thyrotoxicosis. Elevated T_4 levels are usually proportionately higher than T_3 levels. Patients typically have an elevated ESR during the acute phase. The fractional uptake of radioiodine is typically less than 2% at 24 hours (see Table 226-6).

TREATMENT

Rx

High-dose aspirin or naproxen sodium can be used to treat thyroid pain and systemic inflammatory symptoms. Patients who fail to respond may require glucocorticoid therapy, but it must be tapered over several weeks to prevent a relapse, prolonging the overall course of the illness. Symptoms ascribed to transient thyrotoxicosis may respond to treatment with a β-blocker continued for a limited course of 1 to 3 weeks. Patients who progress to symptomatic hypothyroidism may need short-term thyroxine replacement therapy, but most do not require long-term thyroid hormone replacement.

Lymphocytic (Postpartum, Painless, Silent) Thyroiditis

EPIDEMIOLOGY

Lymphocytic thyroiditis occurs most commonly in postpartum women, affecting as many as 6% of women 2 to 12 months after delivery or termination. Rarely, this condition occurs in non-postpartum women or in men. Predisposing factors include a history of previous episodes of postpartum thyroiditis, type 1 diabetes mellitus, and circulating antithyroid autoantibodies.

PATHOBIOLOGY

This painless inflammation of the thyroid gland can cause transient thyrotoxicosis followed by transient or persistent hypothyroidism. Each of these phases of thyroid dysfunction typically lasts 2 to 8 weeks. This condition is believed to reflect transient autoimmunity.

DIAGNOSIS

The diagnosis of lymphocytic thyroiditis is often overlooked when nonspecific symptoms of thyrotoxicosis (e.g., weight loss, insomnia, anxiety) or hypothyroidism (e.g., fatigue, depression) are misinterpreted as common postpartum complaints. The thyroid gland is nontender and either normal in size or modestly enlarged. Once it is considered, a diagnosis of lymphocytic thyroiditis can be readily confirmed or excluded by laboratory testing, which reveals a suppressed TSH level during phases of thyrotoxicosis and an elevated TSH level during phases of hypothyroidism. This condition must be distinguished from Graves disease, which can also present in the same time frame after delivery. Relative degrees of T_4 and T_3 elevation can sometimes provide a clue to which condition is present; lymphocytic thyroiditis is typically characterized by predominant increases in T_4 levels. Fractional uptake of radioiodine is either absent or very low in the setting of lymphocytic thyroiditis, whereas it is increased in active Graves disease (see Table 226-6).

TREATMENT

Rx

Lymphocytic thyroiditis can often be managed with reassurance and observation alone. Symptomatic thyrotoxicosis can be treated with a course of β-blocker therapy. Overt hypothyroidism may require short-term thyroxine replacement.

PROGNOSIS

Most patients with lymphocytic thyroiditis eventually return to a euthyroid state, but 25% develop persistent hypothyroidism due to classic autoimmune thyroiditis.

Acute (Suppurative) Thyroiditis

Infection of the thyroid gland is a rare condition that typically presents with severe thyroid pain, fever, and other systemic manifestations of infection. Bacterial infection of thyroid tissue can be the result of direct spread of gram-positive or gram-negative pathogens through fistulas communicating with the piriform sinus or the skin. Hematogenous spread of bacterial, mycobacterial, fungal, or parasitic organisms, especially *Pneumocystis carinii*, can occur in immunocompromised individuals. On examination, affected patients are typically febrile, with asymmetrical swelling of a thyroid that is tender, warm, and fluctuant to firm in consistency beneath erythematous skin. Ultrasonography may reveal an abscess that can be aspirated to identify a pathogen. Patients with suppurative thyroiditis require prompt treatment with appropriate antibiotics. Surgical drainage of abscesses may be required.

Other Forms of Thyroiditis

Certain drugs can cause thyroid gland inflammation. Amiodarone can produce a painless thyroiditis associated with thyrotoxicosis. Whenever possible, this should be distinguished from the iodine-induced form of thyrotoxicosis that can also be associated with amiodarone therapy. The former is optimally treated with glucocorticoid therapy, whereas the latter is managed with antithyroid drugs.[12] Interferon-α can provoke a painless thyroiditis associated with transient thyrotoxicosis. This must be differentiated from interferon-α–induced Graves disease; the former is managed with β-blockers and the latter with antithyroid drugs.

Riedel's thyroiditis or struma is characterized by fibrotic replacement of the thyroid, with adherence and infiltration of adjacent structures that causes local compressive symptoms. In this idiopathic condition, the thyroid is substantially enlarged, hardened, and fixed. Affected patients may also develop mediastinal and retroperitoneal fibrosis, sclerosing cholangitis, or orbital pseudotumor. Diagnosis requires open biopsy. Surgical excision is difficult or impossible. Glucocorticoid therapy and tamoxifen therapy have been anecdotally reported to be effective.

GOITER

DEFINITION

Goiters can be classified as diffuse or nodular, nontoxic or toxic (i.e., associated with thyroid hormone overproduction), and benign or malignant. Thyroid enlargement can be the result of thyrocyte proliferation stimulated by circulating factors (e.g., TSH and thyroid-stimulating autoantibodies), infiltration of the gland by inflammatory or malignant cells, or benign or malignant neoplastic changes within the gland itself. In a patient with a goiter, three clinical issues must be considered: enlargement causing local compressive or cosmetic concern, gland hyperfunction or hypofunction, and potential malignancy.

EPIDEMIOLOGY

Dietary iodine deficiency represents the most common cause of goiter worldwide. It is encountered in the United States only among immigrants from iodine-deficient regions. Younger patients present with diffuse or simple goiters that shrink in response to adequate iodine supplementation. In older individuals, iodine-deficient goiters become multinodular and do not decrease in size with iodine repletion. Excessive iodine exposure can provoke thyrotoxicosis in these patients.

PATHOBIOLOGY

Benign multinodular goiter or adenoma can be the result of genetic defects that lead to dyshormonogenesis, including mutations in the thyroglobulin, thyroid peroxidase, dual oxidase, and pendrin genes. Similarly, exposure to goitrogenic substances in foodstuffs, water, or drugs (e.g., lithium carbonate) that inhibit the normal steps in thyroid hormone synthesis can lead to goiter. In most patients, the underlying cause is unknown.

Autoimmune thyroiditis typically produces a modest goiter as a result of glandular infiltration with lymphocytes, inflammatory changes in thyrocytes, and fibrosis. The hypothyroid state caused by autoimmune thyroiditis results in increased TSH, which further stimulates thyroid enlargement. Graves disease is also characterized by diffuse thyroid enlargement due to the action of thyroid-stimulating immunoglobulins. Other forms of thyroiditis can present with goitrous enlargement of the thyroid gland, including subacute, lymphocytic, and acute (suppurative) thyroiditis (see earlier sections).

Malignant neoplasms that involve the gland diffusely, including thyroid lymphoma and infiltrative papillary, medullary, and anaplastic thyroid cancers, may present as rapidly enlarging goiters (see later sections). Affected patients often experience local pain and symptoms related to tumor expansion.

DIAGNOSIS

Clinical Examination

The first step in evaluating a suspected goiter is to confirm whether neck swelling represents enlargement of the thyroid. Redundant skin and subcutaneous fat in the lower anterior neck can be mistaken for an enlarged thyroid. These findings can usually be distinguished from true thyroid enlargement by palpating a normal thyroid beneath the misleading soft tissue and by observing that the fullness does not rise and fall with deglutition. Ultrasonography may help resolve uncertainty.

A patient's history can provide important clues to the underlying cause. A childhood social history may confirm previous iodine deficiency. Symptoms of hypothyroidism may suggest autoimmune thyroiditis, whereas clinical evidence of thyrotoxicosis may suggest Graves disease or toxic multinodular goiter. Clinical findings may lead to recognition of one of the various forms of thyroiditis (e.g., pain in subacute thyroiditis or postpartum status in lymphocytic thyroiditis). Symptoms suggesting the invasion of adjacent structures may raise concerns about malignant disease or Riedel's thyroiditis.

On examination, diffuse enlargement favors one of the forms of thyroiditis, Graves disease, or a diffusely infiltrating malignant neoplasm. Nodular enlargement is more likely to reflect a benign multinodular goiter or malignant neoplasm. The precise size of the gland should be documented. Dysphonia, tracheal deviation, cervical lymphadenopathy, and venous engorgement in the neck should be noted. Subtotal obstruction of the thoracic outlet may be revealed by having the patient touch his or her hands together above the head (Pemberton's maneuver) while checking for signs of facial plethora and cervical venous distention.

Laboratory Findings

A TSH level determines whether there is primary hypothyroidism or thyrotoxicosis. Elevated antithyroid peroxidase antibody titers can confirm suspected autoimmune thyroiditis. In asymptomatic patients with a modest diffuse goiter, no further evaluation may be indicated. Other blood tests (e.g., ESR for subacute thyroiditis or calcitonin for medullary thyroid cancer) can be useful when clinical clues suggest specific diagnoses.

Imaging

Cervical ultrasonography is the best imaging technique to define the character and extent of a goiter limited to the neck. It can help determine whether a goiter is diffuse or nodular, whether the thyroid is impinging on other cervical structures, and whether lymphadenopathy is present. Ultrasonography is also essential for guidance of fine-needle aspiration for cytologic differential diagnosis (see later). When a goiter extends posteriorly or beneath the sternal notch into the thorax, CT or MRI may be required. The administration of iodine-containing radiocontrast dye should generally be avoided in the evaluation of patients with goiters, because the stable iodide load may interfere with subsequent radioiodine imaging or therapy. Thyroid radionuclide uptake studies with ^{99m}Tc pertechnetate or ^{123}I can help characterize the functional status of the gland. Radionuclide scanning can help determine the cause of a goiter and whether a superior mediastinal mass is thyroid tissue. Barium swallow radiographs with fixed-diameter markers and pulmonary function testing with flow-volume loops can help determine whether symptoms are directly related to compression of the esophagus or trachea, respectively. Laryngoscopy is useful to evaluate vocal cord function in patients with potential recurrent laryngeal nerve involvement.

TREATMENT

Rx

Once thyroid dysfunction and malignant disease have been excluded, asymptomatic patients with goiters can be observed with periodic clinical assessment. Ultrasonography can be relied on as a reproducible technique for monitoring the size of an enlarged thyroid gland. Thyroxine therapy to suppress TSH levels is effective in shrinking goiters in only a minority of patients. Furthermore, chronic thyroid hormone treatment carries the risks of symptomatic thyrotoxicosis, atrial fibrillation, and bone mineral loss.

Patients with benign multinodular goiters causing local compressive symptoms or cosmetic concerns can be treated with surgery or radioactive iodine therapy. Surgery is often preferred when a patient has substantial gland enlargement causing compressive complications, especially when there is substernal extension of the goiter or acute obstructive symptoms. When surgery is contraindicated by the patient's health status, radioactive iodine therapy has been shown to reduce goiter size by an average of 50% over 12 to 24 months.

THYROID NODULES

EPIDEMIOLOGY

Thyroid nodules are common, being detected by palpation in 6% of women and 2% of men. Contemporary high-resolution ultrasonography identifies thyroid nodules in as many as 50% of all adults. Although the majority of these represent small, benign adenomatoid nodules or cysts, 5 to 10% of thyroid nodules are malignant. Less commonly, thyroid nodules are clinical problems by virtue of being hyperfunctioning or causing local compressive symptoms or cosmetic dissatisfaction.

DIAGNOSIS

Thyroid nodules can be noted by the patient or their physician in the absence of any other complaints.[13] It is also common for thyroid nodules to be detected incidentally on imaging procedures, such as carotid ultrasonography and cervical spine CT or MRI. Symptoms of compression or invasion of adjacent tissues suggest that a nodule may be malignant. These include pain in the lower anterior neck, cough or dyspnea due to tracheal compression, hemoptysis due to tracheal invasion, dysphonia due to recurrent laryngeal nerve encasement, and dysphagia or odynophagia due to esophageal compression. Certain other symptoms and signs lead to the consideration of specific underlying conditions. A toxic adenoma should be suspected in a patient with a thyroid nodule and the classic clinical manifestations of thyrotoxicosis. Hypothyroid symptoms and signs suggest autoimmune thyroiditis

with asymmetrical thyroid enlargement. Hypercalcitoninemia associated with the metastatic spread of medullary thyroid cancer can cause pruritus, flushing, and diarrhea. The clinical assessment should also include symptoms and signs related to common sites of thyroid cancer metastasis, such as chest pain, dyspnea, bone pain, and neurologic findings. Thyroid nodules rarely can be due to metastasis from other primary malignant neoplasms including kidney, colon, and breast cancers.

History

A special predisposition to thyroid cancer is suggested by a personal history of therapeutic neck irradiation in childhood. Family history can be informative if relatives have had medullary or papillary thyroid cancers, which are familial in 50% and 10% of cases, respectively. The possibility of medullary thyroid cancer should also be considered when there is a personal or family history of clinical problems associated with multiple endocrine neoplasia type 2 (MEN 2) syndromes, including hyperparathyroidism and pheochromocytoma (Chapter 231).

Physical Examination

Physical examination of a thyroid nodule should seek to define its size, consistency, surface texture, mobility, and tenderness. The presence of malignant disease is suggested by fixation and ipsilateral regional adenopathy or vocal cord paresis. Multinodularity of the gland may reflect benign nodular goiter, but it is not sufficiently reassuring to dispense with further diagnostic testing.[14] This is particularly true for a so-called dominant nodule that is larger, enlarging faster, or more symptomatic than others present in the thyroid.

Laboratory Findings

Routine laboratory testing includes measurement of TSH levels to identify patients with hyperthyroidism or hypothyroidism. When the TSH level is low or undetectable, the possibility of a benign autonomously functioning toxic adenoma can be pursued with radionuclide thyroid scanning (see Table 226-6). If an elevated TSH level indicates primary hypothyroidism, antithyroid peroxidase antibody titers can confirm whether the patient has autoimmune thyroiditis. Ultrasonography can distinguish asymmetrical enlargement caused by autoimmune thyroiditis from a discrete nodule. Calcitonin levels should be measured in patients with a known or suspected family history of MEN 2 or familial medullary thyroid cancer. Serum thyroglobulin measurement is not helpful in distinguishing benign from malignant thyroid abnormalities.

Imaging

Cervical ultrasonography helps confirm that a mass is within the thyroid, accurately defines its size, classifies it as cystic or solid, and determines whether additional nodules are present. Ultrasonography occasionally reveals other suspicious findings in nodules, such as fine calcifications, irregular nodule borders, and cervical adenopathy.

Radionuclide scanning with radioiodine or technetium pertechnetate is helpful only in selected cases. In patients with a thyroid nodule and a suppressed TSH level, scanning can confirm that the nodule is hyperfunctioning or "hot," in which case biopsy is usually not required.

Invasive Evaluation

Fine-Needle Aspiration Biopsy

Fine-needle aspiration biopsy is the most accurate test to exclude or confirm malignant disease in patients with a nodule and a normal TSH level (Fig. 226-4). Most solid nodules and complex cysts larger than 1.0 to 1.5 cm in diameter should be sampled. Although aspiration can be directed by palpation alone when a nodule is readily definable, ultrasonography provides more certain guidance for the sampling of poorly localized lesions, often revealing additional nodules that should be assessed.

The cytologic assessment of aspirated material must first confirm that there is adequate material for assessment (e.g., 6 clumps of 10 cells on 2 slides). Biopsies with inadequate specimens, which are more common in cystic lesions, must be repeated. Ultrasonographic guidance and on-site preliminary cytologic assessment can improve the yield of biopsy. In accordance with the Bethesda System for Reporting Thyroid Cytopathology, a sampled nodule can be categorized as benign, atypical, suspicious for a follicular neoplasm, suspicious for malignancy, or malignant (Table 226-7).[15]

Benign nodules typically yield samples containing clusters of normal-appearing follicular epithelial cells with colloid. Pure colloid cysts may have scant epithelium. This classification is highly accurate, with a false-negative rate of less than 3% in sonographically directed biopsy specimens, and surgical resection is not required. In most cases, conservative observation based on yearly clinical or sonographic reassessment can be recommended. Further enlargement during observation (i.e., >20% increase in two of three dimensions) should prompt a repeat biopsy. Surgical resection should be considered if a cytologically benign nodule continues to grow, causing compressive symptoms or cosmetic disfigurement.

Cytologic material classified as malignant typically contains abundant epithelial cells with atypical nuclear features, overlapping, and scant or absent colloid. This is also a highly reliable finding, with 98% of such lesions found to be thyroid cancers on subsequent resection. Consequently, bilateral thyroidectomy is indicated in patients without contraindications to operation. Samples that contain sparser quantities of epithelial cells with similar atypical nuclear features may be classified as suspicious for malignancy. Approximately 75% of nodules in this category represent thyroid cancers.

One in five biopsies yields adequate but diagnostically indeterminate cytologic material.[16] Specific findings that classify an indeterminate nodule as suspicious for a follicular neoplasm include abundant follicular or Hürthle cells in microfollicles with little or no colloid and minor degrees of nuclear atypia, potentially indicative of papillary cancer. Although the majority of such indeterminate nodules are benign follicular adenomas, 15 to 30% are thyroid carcinomas. Biopsy samples that reveal nuclear or architectural features considered to be abnormal but not clearly suspicious for malignancy or a follicular neoplasm are classified as demonstrating atypia of undetermined significance. Nodules that initially fall into this category have been estimated to harbor malignancy at rates ranging from 5 to 25%. Repeat sampling may provide a more specific diagnosis to guide further management in 75% of cases.

Definitive determination of whether a suspicious or atypical nodule represents a focus of malignancy requires surgery targeted to remove either the lobe of the thyroid containing the nodule or the entire gland for surgical pathologic examination. Unilateral thyroid lobectomy has the advantage of a lower incidence of surgical complications and postoperative hypothyroidism when the lesion is benign, but it necessitates a subsequent completion thyroidectomy for most patients who prove to have cancer.

Molecular diagnostic testing is available to reduce the number of surgeries performed in patients with cytologically indeterminate nodules, approximately 75% of which prove to be histopathologically benign. There are two general strategies: (1) testing aspirated material for oncogenic mutations associated with thyroid malignancies and (2) gene expression classifier microarrays designed to identify benign nodules. For a typical population of cytologically indeterminate nodule patients with a prevalence of thyroid cancer of 20 to 35%, the negative predictive value of oncogenic testing and gene expression classification have been shown to be approximately 85% and 95%, respectively. For patients with no clinical features of malignancy, particularly middle-aged or older women with multinodular glands in whom the prevalence of malignancy is 5% or less, vigilant observation with serial sonography is an alternative.

THYROID CANCER

Cancers of the thyroid gland have a spectrum of behavior that ranges from incidentally detected and clinically inconsequential microcarcinomas to aggressive and virtually untreatable anaplastic malignant neoplasms. When thyroid cancer is diagnosed early, treatment is effective for most types. Most thyroid cancers present as thyroid nodules that are either asymptomatic or associated with local cervical symptoms or adenopathy. Less often, thyroid cancers first present with manifestations of metastatic disease, such as a pulmonary mass or bone pain.

Papillary and Follicular (Epithelial) Thyroid Carcinomas

Papillary and follicular thyroid cancers arise from follicular epithelium and often retain responsiveness to TSH, produce thyroglobulin, and concentrate iodide. They are distinguished by their histopathologic appearances and characteristic patterns of progression. Hürthle cell carcinoma of the thyroid is composed of thyrocytes with abundant mitochondria-laden cytoplasm, and behaves like a follicular thyroid cancer, although it typically does not have iodine-concentrating ability.

EPIDEMIOLOGY

Approximately 60,000 new cases of thyroid cancer are diagnosed annually in the United States. Thyroid cancer is three times more common in women, in

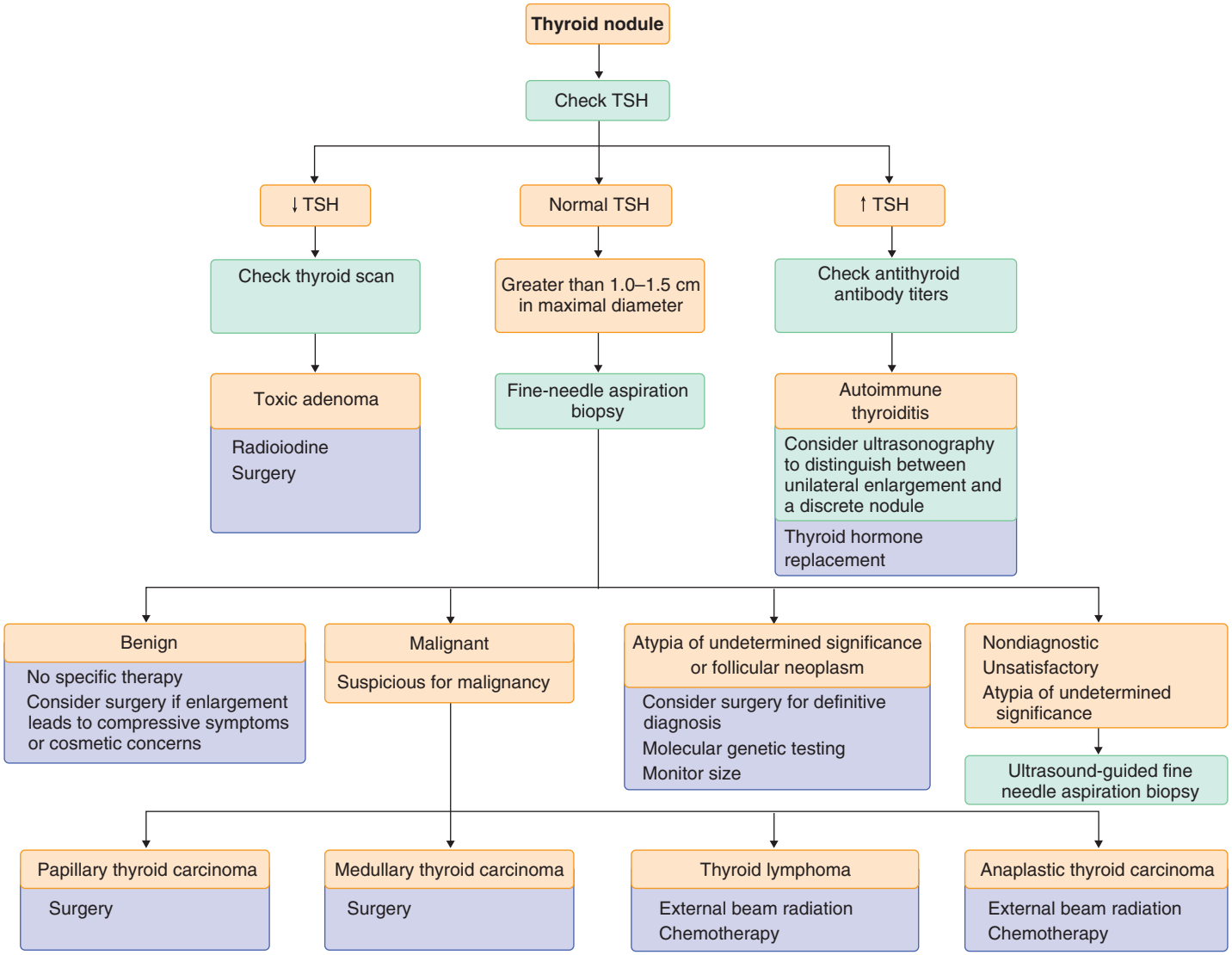

FIGURE 226-4. Evaluation of a thyroid nodule. TSH = thyroid-stimulating hormone.

TABLE 226-7 BETHESDA SYSTEM FOR REPORTING CYTOPATHOLOGY

CYTOLOGIC DIAGNOSIS	RISK OF MALIGNANCY
Benign	0-3%
Atypia of undetermined significance	20-25%
Suspicious for a follicular neoplasm	15-30%
Suspicious for malignancy	60-77%
Malignant	97-99%

Diagnostic categories associated with risk of malignancy.
Adapted from Bongiovanny M, Spitale A, Faquin WC, et al. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. *Acta Cytol.* 2012;56:333-339.

whom its incidence is currently rising faster than that of any other malignancy. There are estimated to be 450,000 U.S. thyroid cancer survivors who require lifelong follow-up for recurrence. Papillary thyroid carcinoma is the most common form of thyroid cancer, representing 90% of cases. The mean age at diagnosis is 45 years, but papillary thyroid carcinoma does occur in children and increases in incidence with age.

PATHOBIOLOGY

Irradiation of the thyroid gland in childhood is a risk factor, as evidenced by the epidemics of thyroid cancer that have followed both external beam radiation therapy for benign childhood conditions (e.g., tonsillitis and acne) and radioiodine exposure after nuclear incidents. A substantial body of evidence now implicates *RET/PTC* and *BRAF* gene mutations that activate the MAP kinase signaling pathway in the pathogenesis and progression of papillary thyroid cancer. Most papillary thyroid carcinomas are slow growing and either remain confined to the gland or metastasize to cervical lymph nodes. Papillary microcarcinomas are a common incidental pathologic finding in 5% of thyroid glands excised for other reasons. However, papillary thyroid carcinomas can be more aggressive, with extension into adjacent tissues, extensive nodal involvement, and distant metastatic spread, most commonly to the lungs. Such aggressive behavior is, in general, more common in older patients.

Follicular and Hürthle cell thyroid carcinomas account for 9% of all thyroid cancers. When these tumors show histologic evidence of invading only the tumor capsule, they are termed *minimally invasive* and generally behave like papillary thyroid carcinomas. However, follicular and Hürthle cell carcinomas with vascular invasion are more likely to be associated with distant metastatic disease, which most commonly involves the lungs and skeleton.

TREATMENT

Rx

Treatment of epithelial thyroid cancer entails surgery, often followed by radioiodine ablation of remnant thyroid tissue. Total or near-total thyroidectomy with selective central compartment lymph node resection is usually the appropriate initial surgical procedure. Thyroid surgery can be complicated by hypoparathyroidism or recurrent laryngeal nerve injury, which causes hoarseness if it is unilateral and airway obstruction if it is bilateral. The rationale for bilateral surgery is the frequent presence of bilateral disease in papillary

thyroid cancer and the lower risk of recurrence after bilateral gland removal. In addition, there is greater accuracy in detecting residual disease after the eradication of all remaining normal thyroid tissue.

A prospective cohort study of a national database of 30-day follow-up of patients undergoing thyroidectomy (for cancer or other indications) documented the increased risk of major pulmonary, cardiac, and infectious complications in the elderly. Elderly patients (65 to 79 years old) are twice as likely, and the most elderly (80 years or older) are 5 times as likely as young patients (16 to 64 years old) to have major systematic complications.[17]

Follow-up

Postoperatively, ^{131}I administration after TSH stimulation can be employed to ablate the small amount of normal thyroid tissue that usually remains after surgery. This tissue, if it is not destroyed, leaves patients with circulating thyroglobulin and iodine-concentrating tissue on whole body scanning, decreasing the accuracy of these tests to identify residual disease. In controlled but nonrandomized trials, radioiodine has been associated with a lower rate of tumor recurrence in patients with advanced disease (stages 3 and 4) at presentation (see later), but there is no demonstrated clinical benefit of adjunctive radioiodine therapy for patients with lower stages of disease. TSH stimulation of residual thyroid tissue, which is essential for effective radioiodine therapy, can be accomplished either by the temporary withdrawal of thyroid hormone therapy to promote endogenous TSH production or by the administration of recombinant thyrotropin, which avoids the morbidity of hypothyroidism.

Thyroxine therapy is appropriate for all patients with treated thyroid cancer, regardless of the extent of surgery and whether they received radioiodine ablative therapy. In addition to providing thyroid hormone replacement, thyroxine can be adjusted to suppress the patient's circulating TSH level to the low or low-normal range to reduce the likelihood of tumor recurrence. In determining the extent to which the TSH level should be suppressed, the patient's risk of cancer recurrence must be balanced against potential thyrotoxic complications such as bone mineral loss in postmenopausal women and atrial fibrillation in older patients.

Long-term monitoring of patients entails periodic clinical assessment, measurement of serum thyroglobulin levels, radioiodine imaging in the early postoperative phase, and occasional use of ultrasonography. Clinically, patients should be assessed for local neck symptoms or recurrent cervical masses, as well as for optimization of thyroid hormone therapy. For patients with treated epithelial thyroid cancers, thyroglobulin is a more specific tumor marker if all remaining normal thyroid tissue has been ablated. For patients with undetectable thyroglobulin levels on TSH-suppressive thyroid hormone therapy, thyroglobulin measurement after recombinant TSH stimulation can sometimes reveal residual disease. Radioiodine scanning after TSH stimulation can be helpful in patients who have previously undergone radioiodine ablation, but once radioiodine imaging is negative, it offers little or no advantage over measurement of stimulated thyroglobulin levels. This is particularly true in recurrent papillary thyroid cancers, which often lose the ability to concentrate iodine. Unfortunately, thyroglobulin testing is impossible in the 20% of patients who have circulating thyroglobulin autoantibodies that interfere with thyroglobulin immunoassays. Because most epithelial thyroid cancer recurrences are in cervical nodes or soft tissues, ultrasonography is useful for postoperative monitoring, particularly in patients who presented with extensive cervical disease or who have persistently detectable serum thyroglobulin. CT scanning of the chest should be employed to detect intrathoracic disease in patients whose findings suggest recurrence outside the neck. In patients with substantial detectable thyroglobulin levels (>10 ng/mL) and negative findings on standard imaging studies, PET scanning can identify sites of residual disease in more than 50% of patients.

Localization of recurrent cervical disease is usually an indication for comprehensive compartmental neck dissection. Distant and nonresectable metastases that are iodine avid, which occur more commonly in patients with invasive follicular thyroid cancer, can be treated with repeated doses of ^{131}I. Symptomatic hilar node and bone metastases can be treated palliatively with external beam radiation therapy. Surgery can be employed for isolated metastatic disease sites. Conventional chemotherapy has limited efficacy in the treatment of differentiated thyroid cancer, but newer biologic agents targeting the molecular pathways involved in the pathogenesis of thyroid cancer hold promise.[18] For example, the multikinase inhibitor sorafenib was shown in a phase 3 trial to double progression free survival to almost 11 months in patients with metastatic non–iodine avid epithelial thyroid cancers and shrink disease sites in 12% of patients.[A6] Another multikinase inhibitor, vandetanib, is also effective against locally advanced or metastatic differentiated thyroid cancer.[A6] However, multikinase inhibitors commonly have adverse effects, and because they are only tumoristatic must be used continuously.

PROGNOSIS

The TNM (tumor, node, metastasis) staging system is commonly used to stage epithelial thyroid cancers. In addition to tumor size, extent of node involvement, and presence of distant metastatic disease, the age of the patient at presentation is an important predictor of outcome. Patients younger than 45 years have a better prognosis than older individuals. The overall age-adjusted 10-year survival rates for patients with papillary and follicular thyroid cancer are 98% and 92%, respectively. However, disease recurrence is relatively common, occurring in approximately one third of patients with papillary thyroid cancer. Consequently, patients with treated thyroid cancer must be monitored for recurrent disease.

Medullary Thyroid Carcinoma

Patients with medullary thyroid cancer (Chapter 246) typically present with a thyroid nodule, cervical adenopathy, distant disease, or symptoms of flushing, diarrhea, and pruritus when the circulating calcitonin level is markedly elevated. Features of the other elements of MEN 2a (e.g., hypertension) or MEN 2b (e.g., marfanoid habitus, submucosal neuromas) should be sought.

Anaplastic Thyroid Carcinoma

Anaplastic thyroid carcinoma is a rare, histologically undifferentiated, clinically aggressive malignant neoplasm that typically arises in older patients, one fourth of whom present with evidence of a preceding differentiated thyroid cancer. Affected patients present with a rapidly enlarging mass in the anterior or lateral neck associated with pain, tenderness, and compressive symptoms including dysphagia, dysphonia, and stridorous dyspnea. Fine-needle aspiration biopsy of the mass usually yields large, pleomorphic, undifferentiated cells, but open surgical biopsy is sometimes required to confirm the diagnosis.

Most cases are unresectable at presentation because of invasion of cervical structures. Surgery is not curative and should aim to secure the patient's airway. A percutaneous gastrostomy tube is often placed to ensure adequate nutrition in the face of esophageal impingement. Conventional therapy consisting of combined external beam radiation therapy and chemotherapy with doxorubicin with or without cisplatin produces an initial response in 25% of patients. Rare patients with disease limited to the neck may have extended survival, but almost all patients relapse within a few months and succumb to their disease, with median survival ranging from 3 to 7 months. Current research is focused on the use of targeted antiangiogenic agents to treat unresponsive disease.

Thyroid Lymphoma

Lymphoma rarely arises in the thyroid gland, typically presenting in older persons as a rapidly enlarging and painful diffuse goiter. Patients often have a preceding history of autoimmune thyroiditis. The diagnosis is further suspected when fine-needle aspiration biopsy yields abundant lymphocytes without other cellular features of autoimmune thyroiditis. Immunohistochemical staining and flow cytometry of sampled material can characterize a monoclonal lymphocyte population. Surgical biopsy is sometimes required to establish the diagnosis. In 50% of cases, lymphoma is primary to the thyroid gland, and it is usually an intermediate-grade non-Hodgkin's–type lymphoma (Chapter 185).

Surgical resection of the thyroid is usually not indicated, but elective tracheostomy may be required if tracheal compression is imminent. Most patients respond to treatment with combined external beam radiation therapy and chemotherapy. Disease-free survival rates vary with the disease stage at diagnosis and the initial response to combination therapy.

Grade A References

A1. McDermott MT. Does combination T4 and T3 therapy make sense? *Endocr Pract*. 2012;18:750-757.

A2. Marcocci C, Kahaly GJ, Krassas GE, et al. Selenium and the course of mild Graves' orbitopathy. *N Engl J Med*. 2011;364:1920-1931.

A3. Bahn Chair RS, Burch HB, Cooper DS, et al. Hyperthyroidism and other causes of thyrotoxicosis: management guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. *Thyroid*. 2011;21:593-646.

A4. Stagnaro-Green A, Abalovich M, Alexander E, et al. Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. *Thyroid*. 2011;21:1081-1125.

A5. Smallridge RC, Ain KB, Asa SL, et al. American Thyroid Association guidelines for management of patients with anaplastic thyroid cancer. *Thyroid*. 2012;22:1104-1139.

A6. Brose MS, Nutting CM, Jarzab B, et al. Sorafenib in radioactive iodine-refractory, locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, phase 3 trial. *Lancet*. 2014;384:319-328.

A7. Leboulleux S, Bastholt L, Krause T, et al. Vandetanib in locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, phase 2 trial. *Lancet Oncol*. 2012;13: 897-905.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bremner AP, Feddema P, Leedman PJ, et al. Age-related changes in thyroid function: a longitudinal study of a community-based cohort. *J Clin Endocrinol Metab*. 2012;97:1554-1562.
2. Caturegli P, De Remigis A, Rose NR. Hashimoto thyroiditis: clinical and diagnostic criteria. *Autoimmun Rev*. 2014;13:391-397.
3. Garber JR, Cobin RH, Gharib H, et al. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. *Thyroid*. 2012;22:1200-1235.
4. Stagnaro-Green A, Pearce E. Thyroid disorders in pregnancy. *Nat Rev Endocrinol*. 2012;8:650-658.
5. Grais IM, Sowers JR. Thyroid and the heart. *Am J Med*. 2014;127:691-698.
6. Rugge JB, Bougatsos C, Chou R. Screening and treatment of thyroid dysfunction: an evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2015;162:35-45.
7. Pappa TA, Vagenakis AG, Alevizaki M. The nonthyroidal illness syndrome in the non-critically ill patient. *Eur J Clin Invest*. 2011;41:212-220.
8. Vaidya B, Pearce SH. Diagnosis and management of thyrotoxicosis. *BMJ*. 2014;349:g5128.
9. Bahn RS. Graves' ophthalmopathy. *N Engl J Med*. 2010;362:726-738.
10. Cooper DS, Laurberg P. Hyperthyroidism in pregnancy. *Lancet Diabetes Endocrinol*. 2013;1:238-249.
11. Lee SL. Radioactive iodine therapy. *Curr Opin Endocrinol Diabetes Obes*. 2012;19:420-428.
12. Bogazzi F, Tomisti L, Bartalena L, et al. Amiodarone and the thyroid: a 2012 update. *J Endocrinol Invest*. 2012;35:340-348.
13. Niedziela M. Thyroid nodules. *Best Pract Res Clin Endocrinol Metab*. 2014;28:245-277.
14. Brito JP, Yarur AJ, Prokop LJ, et al. Prevalence of thyroid cancer in multinodular goiter versus single nodule: a systematic review and meta-analysis. *Thyroid*. 2013;23:449-455.
15. Bongiovanni M, Spitale A, Faquin WC, et al. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. *Acta Cytol*. 2012;56:333-339.
16. Alexander EK, Kennedy GC, Baloch ZW, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med*. 2012;367:705-715.
17. Grogan RH, Mitmaker EJ, Hwang J, et al. A population-based prospective cohort study of complications after thyroidectomy in the elderly. *J Clin Endocrinol Metab*. 2012;97:1645-1653.
18. Marotta V, Sciammarella C, Vitale M, et al. The evolving field of kinase inhibitors in thyroid cancer. *Crit Rev Oncol Hematol*. 2015;93:60-73.

REVIEW QUESTIONS

1. A 74-year-old man with a history of hypertension and hypercholesterolemia is hospitalized after presenting with a 3-month history of progressive fatigue and dyspnea on exertion. His weight on admission is 164 lbs, his pulse is 44 bpm, and lab tests show CPK 528 U/L, TSH 65 mU/L, free T_4 0.2 ng/dL, and T_4 2.3 µg/dL. A pharmacologic nuclear stress test reveals findings consistent with diffuse myocardial ischemia. Subsequent coronary angiography reveals diffuse three-vessel disease that is not amenable to percutaneous stenting. A consulting cardiologist has recommended that he undergo coronary artery bypass surgery. What would you recommend?

A. Start levothyroxine at a dose of 125 µg daily.
B. Administer a 60-µg intravenous dose of levothyroxine daily.
C. Check antithyroid peroxidase and antithyroglobulin antibodies.
D. Start levothyroxine at a dose of 12.5 µg daily.

Answer: D Start levothyroxine at a dose of 12.5 µg daily. This patient requires treatment with levothyroxine, but it should be started at the lowest possible dose, with provisions to gradually increase it as tolerated in light of his coronary artery disease. A full replacement dose given orally or an adjusted dose given intravenously could exacerbate cardiac ischemia to the point of causing an infarction. Checking antithyroid peroxidase and antithyroglobulin antibodies would not provide any additional information. In the absence of any known history of thyroid surgery or external radiation treatment to the head and neck, it can be presumed that his severe hypothyroidism is due to autoimmune thyroiditis.

2. A 28-year-old woman with a 6-year history of hypothyroidism has been treated with levothyroxine at a dose of 125 µg daily. Lab tests checked 5 months ago right before she stopped taking cyclic oral contraceptives showed TSH 1.3 mU/L. Three weeks ago she checked a home pregnancy test that was positive. She had lab tests checked by her obstetrician that showed hemoglobin 10.2 g/dL, T_4 12.5 µg/dL, and T_3 185 ng/dL. She is estimated to be at 8 weeks' gestation and has started taking prenatal vitamins and iron sulfate at a dose of 325 mg twice daily. What should you do?

A. Have her continue levothyroxine at a dose of 125 µg daily, with instructions to take it regularly with other medications.
B. Increase her dose of levothyroxine to 150 µg daily.
C. Decrease her dose of levothyroxine to 112 µg daily.
D. Check a thyroid uptake and scan.
E. Check a TSH level.

Answer: E Check a TSH level. The only way to determine whether her current dose of levothyroxine is providing an adequate level of replacement is to check a TSH level. Total T_4 and T_3 levels measured while taking oral contraceptives or during pregnancy may be elevated or high-normal owing to increased production of thyroxine-binding globulin stimulated by increased estrogen levels. As such, they may not correlate with free thyroid hormone levels. Decreasing her dose of levothyroxine would be inappropriate because she may require a moderate to substantial increase in her dose during the course of a pregnancy. Empirically increasing a dose of levothyroxine may cause iatrogenic thyrotoxicosis characterized by symptoms that may be difficult to distinguish from normal physiologic changes of pregnancy. Exposure to radionuclide tracer is contraindicated during pregnancy, and in any event a thyroid uptake and scan would be unnecessary in a patient with known hypothyroidism. Doses of levothyroxine should always be separated from doses of iron sulfate by at least 4 hours to avoid interactions that can block absorption of both agents.

3. A 76-year-old woman presenting with a 1-week history of a cough, fever, and audible stridor is diagnosed with community-acquired pneumonia. A chest x-ray does not show evidence of an infiltrate but does reveal marked rightward tracheal deviation with a visible mediastinal soft tissue mass. A non-contrast chest computed tomography scan reveals multiple bilateral thyroid nodules, with an 8.5-cm left lower-pole nodule extending below the clavicle and sternum, with compression and narrowing of the trachea. Lab tests show TSH less than 0.001 mU/L, free T_4 2.8 ng/dL, and T_3 245 ng/dL. A thyroid uptake and scan reveals 24-hour uptake of 37%, with tracer accumulation localized to two right-sided thyroid nodules and the substernal left-sided thyroid nodule. What should you do next?

A. Check pulmonary function tests with flow-volume loops.
B. Prescribe a 12-month course of methimazole 10 mg daily.
C. Refer the patient to a thyroid surgeon.
D. Administer a 30-mCi dose of I-131.
E. Start levothyroxine at a dose of 137 µg daily.

Answer: C Refer the patient to a thyroid surgeon. A multinodular goiter that has extended substernally to the point of causing tracheal compression should be resected by an experienced thyroid surgeon, irrespective of its functional status. The presence of audible stridor and evidence of tracheal narrowing on radiographic images obviates the need for pulmonary function testing with flow-volume loops. Treatment with methimazole might help control hyperthyroidism caused by autonomously functioning thyroid nodules but will not be a permanent solution and would not shrink the dominant nodule to any extent. Treatment with I-131 might help shrink the dominant nodule over time but will require 12 to 24 months to be effective. Treatment with levothyroxine to try to suppress further enlargement of thyroid tissue may be marginally effective at best in euthyroid patients and would be completely ineffective—and potentially dangerous—in an elderly woman presenting with hyperthyroidism.

4. A 33-year-old man is noted to have palpable enlargement of the right side of his thyroid on a routine physical exam. A thyroid ultrasound reveals a solitary 3.1-cm right-sided nodule with smooth borders. Lab tests show TSH 0.1 mU/L and T_4 11.5 µg/dL. He reports a history of occasional symptomatic palpitations and weight loss of 5 lbs over the course of 3 months despite an increase in his appetite. He is not taking any medications and has not noted any problems with dysphagia or dysphonia. What should you do next?

A. Perform a fine-needle aspiration biopsy of the right-sided nodule.
B. Administer a 15-mCi dose of I-131.
C. Refer the patient to a thyroid surgeon.
D. Start methimazole at a dose of 5 mg daily.
E. Perform a 123-I thyroid scan.

Answer: E Perform a 123-I thyroid scan. When a patient presenting with a thyroid nodule who is not taking levothyroxine is noted to have a suppressed TSH level, a radionuclide thyroid scan should be checked to determine if the nodule is an autonomously functioning toxic adenoma. If a nodule is "hot" on the scan, it does not need to be biopsied. If it is "cold" on the scan, with evidence of increased tracer uptake in surrounding tissue consistent with Graves disease, fine-needle aspiration biopsy should be performed to confirm it is benign. Treatment of a toxic adenoma or Graves disease with methimazole or radioactive iodine may be indicated but should only be considered after it has been determined whether a biopsy is necessary. Referral for thyroid surgery would only be indicated if a biopsy of a cold nodule revealed suspicious or malignant cytopathology, or if there were contraindications to treatment of a toxic adenoma with radioactive iodine or methimazole.

5. A 55-year-old woman is noted to be tachycardic with a resting pulse of 104 during a routine physical exam. Lab tests checked to evaluate this show TSH 0.002 mU/L, with follow-up lab tests showing free T_4 2.3 ng/dL and T_3 289 ng/dL. She denies any history of anterior neck discomfort, weight loss, palpitations, anxiety, tremor, heat intolerance, or insomnia. Physical examination reveals tachycardia with a regular rhythm, a slightly enlarged thyroid without any discrete nodularity, and no evidence of proptosis or ocular irritation. She does have a history of osteopenia, with a lumbar spine T-score of −2.3 identified on a DEXA scan checked soon after the onset of menopause. What should you do next?

A. Check antithyroid peroxidase and antithyroglobulin antibodies.
B. Check a radionuclide thyroid uptake and scan.
C. Start methimazole at a dose of 10 mg daily.
D. Refer the patient to a thyroid surgeon.
E. Check an ESR.

Answer: B Check a radionuclide thyroid uptake and scan. This patient is presenting with thyrotoxicosis without any referable symptoms or clinical findings suggestive of a specific cause. Checking a thyroid uptake would help distinguish between a high-uptake state caused by hyperthyroidism driven by increased production of thyroid hormone, and a low-uptake state caused by inflammation with leakage of stored thyroid hormone. If a high-uptake state is identified, a scan will help distinguish whether hyperthyroidism is caused by Graves disease, a toxic adenoma, or a toxic multinodular goiter. Elevated antithyroid peroxidase and antithyroglobulin antibodies may identify underlying autoimmune thyroiditis but will not definitively determine the proximate cause of thyrotoxicosis. Treatment with methimazole or referral for thyroid surgery would only be considered after confirmation of a diagnosis. Checking an erythrocyte sedimentation rate (ESR) would not be informative, because subacute thyroiditis would be unlikely in the absence of localized discomfort.

227

ADRENAL CORTEX

LYNNETTE K. NIEMAN

The adrenal glands weigh 6 to 8 g in adults (Fig. 227-1). Each contains a cortex, which makes steroid hormones, and a medulla, which produces catecholamines. Diseases of the adrenal medulla are discussed in Chapter 228. In the adrenal cortex, production of the three major classes of steroids occurs in specific zones: the outermost layer, the glomerulosa, produces mineralocorticoids, primarily aldosterone; the middle layer, the fasciculata, produces glucocorticoids, primarily cortisol; the innermost layer, the reticularis, produces adrenal "androgens," primarily dehydroepiandrosterone (DHEA) and its sulfated conjugate (DHEA-S) (Fig. 227-2). This division reflects the fact that certain critical enzymes are restricted to specific zones, resulting in the ability or inability to synthesize specific end products.

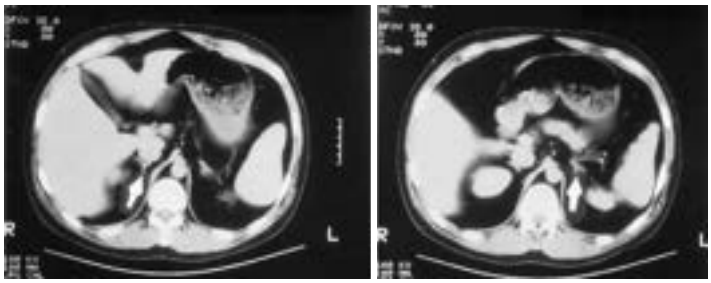

FIGURE 227-1. Magnetic resonance images of the abdomen showing the position and relative size of the normal adrenal glands.

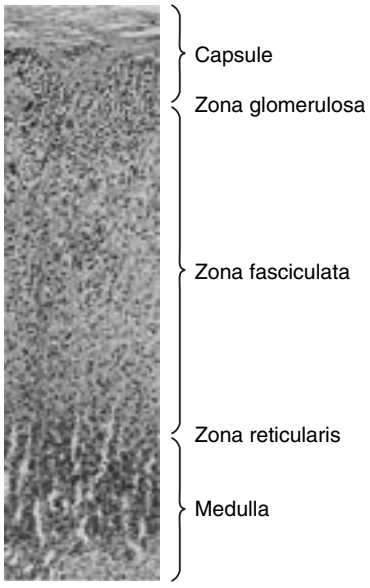

FIGURE 227-2. Histologic section through a normal adult adrenal gland showing the progression (from outside to inside) of the zona glomerulosa, zona fasciculata, zona reticularis, and medulla.

FUNCTION

The actions and regulation of these steroid classes differ. Mineralocorticoids act through the renal mineralocorticoid receptor to promote the reabsorption of sodium and the secretion of potassium. In addition to this classic action, mineralocorticoids have important action on the vasculature and may exacerbate the metabolic syndrome.[1] Aldosterone secretion is stimulated primarily by hyperkalemia and angiotensin II (which itself is stimulated by hypovolemia and excess renin). These agents increase the production of aldosterone synthase to restore homeostasis through this feedback loop. Aldosterone production is stimulated to a much smaller degree by adrenocorticotropic hormone (ACTH).

Cortisol and other glucocorticoids act through the glucocorticoid receptor type 2 and its isoforms. The actions of this class of steroids are much broader, including effects on carbohydrate handling, lipid and calcium metabolism, and the immune and nervous systems. Cortisol production is regulated primarily by ACTH, which is secreted in a circadian rhythm in response to corticotropin-releasing hormone (CRH) so that cortisol levels are highest in the morning and fall to a nadir around midnight. Cortisol coordinates ACTH production through negative feedback at the pituitary (ACTH) and hypothalamus (CRH). Vasopressin secretion also plays a role in stimulating ACTH release.

DHEA and DHEA-S are the most abundant products of the adrenal gland. They exert their estrogenic and androgenic effects as prohormones, being converted to estrogens and testosterone in the peripheral tissues and activating the androgen and estrogen receptors. There is no known regulator of DHEA synthesis, but its production declines with age.

DISORDERS OF ADRENAL FUNCTION

Most disorders of the adrenal cortex reflect overproduction or underproduction of the products of a single synthetic zone—cortisol, aldosterone, or testosterone or estrogen (Fig. 227-3). The congenital adrenal hyperplasias are an exception and are manifested with both overproduction and underproduction. Abnormal secretion is suggested by clinical features of each disorder and is reflected in plasma or urine levels of the relevant hormones or by the consequent increases or decreases in feedback systems, which form the basis of the biochemical diagnostic tests.

Glucocorticoid Excess: Cushing Syndrome

CLINICAL MANIFESTATIONS

Cushing syndrome is a symptom complex that reflects excessive tissue exposure to cortisol. Classic features of Cushing syndrome include weight gain, plethora, hypertension, and striae (Table 227-1). Not all patients have all features; the number and severity of features correlate roughly with the duration and severity of hypercortisolism. Because many of the signs and symptoms are nonspecific, the diagnosis may be confused with psychiatric disorders, polycystic ovary syndrome, the metabolic syndrome, simple obesity, fibromyalgia, or acute illness. However, because worsening hypercortisolism may precipitate hypertension, glucose intolerance, infections, psychiatric disturbances, impaired cognition, and hypercoagulability, it is important to identify this treatable disorder to prevent its associated morbidity and mortality.[2]

Changes in mood and cognition are useful markers of hypercortisolism. These include irritability, crying, and restlessness; depressed mood; decreased libido; insomnia; anxiety; and decreased concentration and impaired memory.

DIAGNOSIS

Clinical Examination

Cushing syndrome screening is most likely to be positive in the presence of signs that are typical of glucocorticoid excess, such as abnormal fat distribution in the supraclavicular and temporal fossae, proximal muscle weakness, wide (>1 cm) purple striae, and new irritability, decreased cognition, and decreased short-term memory. Testing is indicated when clinical features have progressed over time. For example, oligomenorrhea is more suggestive of Cushing syndrome if a woman previously had regular menses. Serial seven subtractions and recall of three cities (or objects) are useful bedside strategies to identify deficits in cognition and memory.

Laboratory Findings

Exogenous administration of glucocorticoid should be excluded before screening for endogenous Cushing syndrome. In the absence of pseudo-Cushing states (see later), at least two different screening test results should

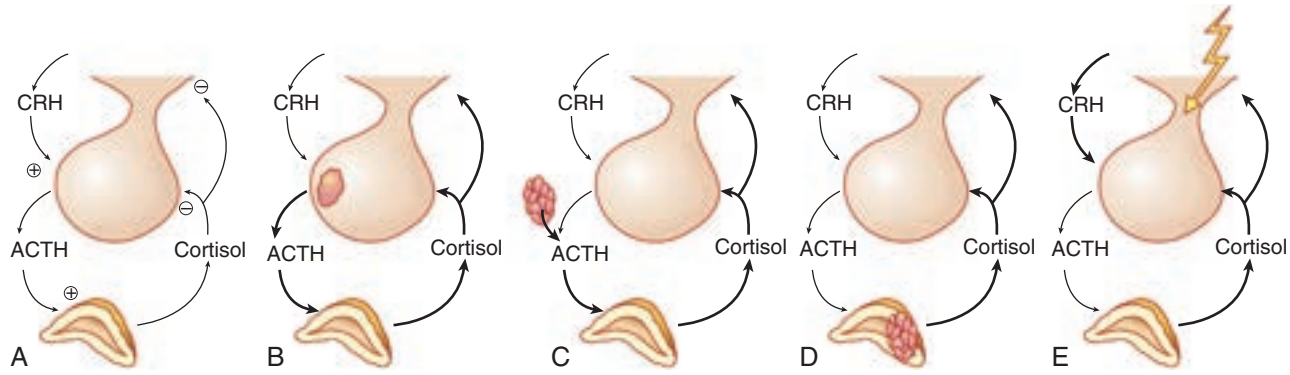

FIGURE 227-3. Physiology of the adrenal axis in health, Cushing syndrome, and pseudo-Cushing states. A, In healthy individuals, cortisol production is stimulated by the increased hypothalamic release of corticotropin-releasing hormone (CRH), which then travels down the pituitary stalk to stimulate adrenocorticotropic hormone (ACTH) secretion and release by corticotropes. Circulating ACTH stimulates adrenal gland production and secretion of cortisol. Cortisol then functions in a negative feedback mechanism to inhibit both CRH and ACTH. **B,** In Cushing disease, a pituitary tumor releases excessive amounts of ACTH, which results in increased cortisol secretion by the adrenal glands. **C,** In ectopic ACTH secretion, a nonpituitary ACTH-secreting tumor releases excessive amounts of ACTH, which results in increased cortisol secretion by the adrenal glands. **D,** In ACTH-independent adrenal forms of Cushing syndrome, the adrenal tumor autonomously releases excess amounts of cortisol. In all forms of Cushing syndrome, the negative feedback effects of excessive cortisol inhibit endogenous CRH and ACTH secretion, so that circulating ACTH levels reflect the underlying tumor (levels are normal or increased) or independent cortisol production (levels are suppressed). **E,** In pseudo-Cushing states, central stimulation increases CRH secretion, which in turn increases ACTH and hence cortisol production. In this setting, the negative feedback effects of excessive cortisol inhibit endogenous CRH and ACTH secretion, so that cortisol levels are ultimately constrained, albeit at an increased level.

TABLE 227-1 THE FREQUENCY OF CLINICAL SIGNS AND SYMPTOMS OF CUSHING SYNDROME

SIGN OR SYMPTOM	PERCENTAGE
Decreased libido in men and women	100
Obesity or weight gain	97
Plethora	94
Round face	88
Menstrual changes	84
Hirsutism	81
Hypertension	74
Ecchymoses	62
Lethargy, depression	62
Striae	56
Weakness	56
Electrocardiographic changes or atherosclerosis	55
Dorsal fat pad	54
Edema	50
Abnormal glucose tolerance	50
Osteopenia or fracture	50
Headache	47
Backache	43
Recurrent infections	25
Abdominal pain	21
Acne	21
Female balding	13

be abnormal to establish the diagnosis. Tests for the differential diagnosis of Cushing syndrome should not be used to make the diagnosis. Figure 227-4 is the Endocrine Society's recommended algorithm for testing of patients suspected of having Cushing syndrome.[3]

Urine, Saliva, and Serum Cortisol Measurements

Urine free cortisol (UFC) excretion during 24 hours is a good screening test. Specific, structurally based assay techniques, such as high-performance liquid chromatography and tandem mass spectrometry, are the "gold standard." The upper-normal limit of these tests is much lower and more specific than that of antibody-based assays, in which other steroids may cross-react. This cross-reactivity may be an advantage in screening for hypercortisolism.

UFC excretion also may be increased in the so-called *pseudo-Cushing states*, including psychiatric disorders (depression, anxiety disorder, obsessive-compulsive disorder), chronic pain, severe exercise, alcoholism, uncontrolled diabetes, and morbid obesity. Here, it is hypothesized that higher brain pathways stimulate CRH release and activation of the entire hypothalamic-pituitary-adrenal axis (see Fig. 227-3E). Cortisol negative feedback inhibition on CRH and pituitary ACTH release restrains the resulting hypercortisoluria to less than four-fold greater than normal. Thus, Cushing syndrome cannot be diagnosed with certainty unless values reach this threshold. Conversely, patients with Cushing syndrome may have normal UFC excretion because of mild or intermittent hypercortisolism or altered renal metabolism of cortisol. If UFC is only mildly elevated and clinical features are minimal, it is best to treat any pseudo-Cushing state and to remeasure UFC excretion with the expectation that it will normalize. Alternatively, if UFC values are normal but clinical suspicion is high, repeated measurement might disclose intermittent hypercortisolism.

Measurement of plasma cortisol at midnight distinguishes pseudo-Cushing states from Cushing syndrome with 95% diagnostic accuracy; a level greater than 7.5 μg/dL is required for the diagnosis of Cushing syndrome. Measurement of salivary cortisol at bedtime or at midnight works as well, is more convenient, and may be the best screening test in patients with mild or intermittent hypercortisolism.[4,5] However, the criteria for its interpretation differ, so each assay must be validated before it is used for this purpose.

Dexamethasone Suppression Tests

The dexamethasone suppression test is a simple screening test that takes advantage of the negative feedback effect of glucocorticoids to reduce ACTH (and hence serum cortisol). Dexamethasone 1 mg is given orally between 11:00 PM and midnight, and plasma cortisol is measured between 8:00 and 9:00 the next morning. The test has an 8% false-negative rate in patients with Cushing disease and a 30% false-positive rate in chronic illness, obesity, psychiatric disorders, and normal individuals. As a result, Cushing syndrome cannot be diagnosed by this test alone unless the result is extremely abnormal.

The 2-day, 2-mg dexamethasone suppression test discriminates patients with a pseudo-Cushing state if plasma cortisol end points of less than 1.4 or 2.2 μg/dL are used. Dexamethasone 500 μg is given orally every 6 hours for eight doses, and plasma cortisol is measured 2 hours after the last dose. The test has excellent sensitivity (90 to 100%) and specificity (97 to 100%) for discriminating Cushing syndrome, but it is costly and requires excellent compliance of the patient. The immediate subsequent administration of CRH (1 μg per kilogram of body weight intravenously) and the measurement of cortisol 15 minutes later increased the sensitivity and specificity to 100% in a small study of patients, with values above 1.4 μg/dL indicating Cushing syndrome. Although this combined dexamethasone-CRH test has high diagnostic accuracy, it has the same disadvantages as the 2-day dexamethasone suppression test and the added cost of CRH testing. Because of these drawbacks, these tests are usually reserved for patients with ambiguous or confusing results on other screening tests. CRH is available commercially (Acthrel), with Food and Drug Administration–approved labeling for the differential diagnosis of Cushing syndrome. Its use in the dexamethasone-CRH test is an off-label use.

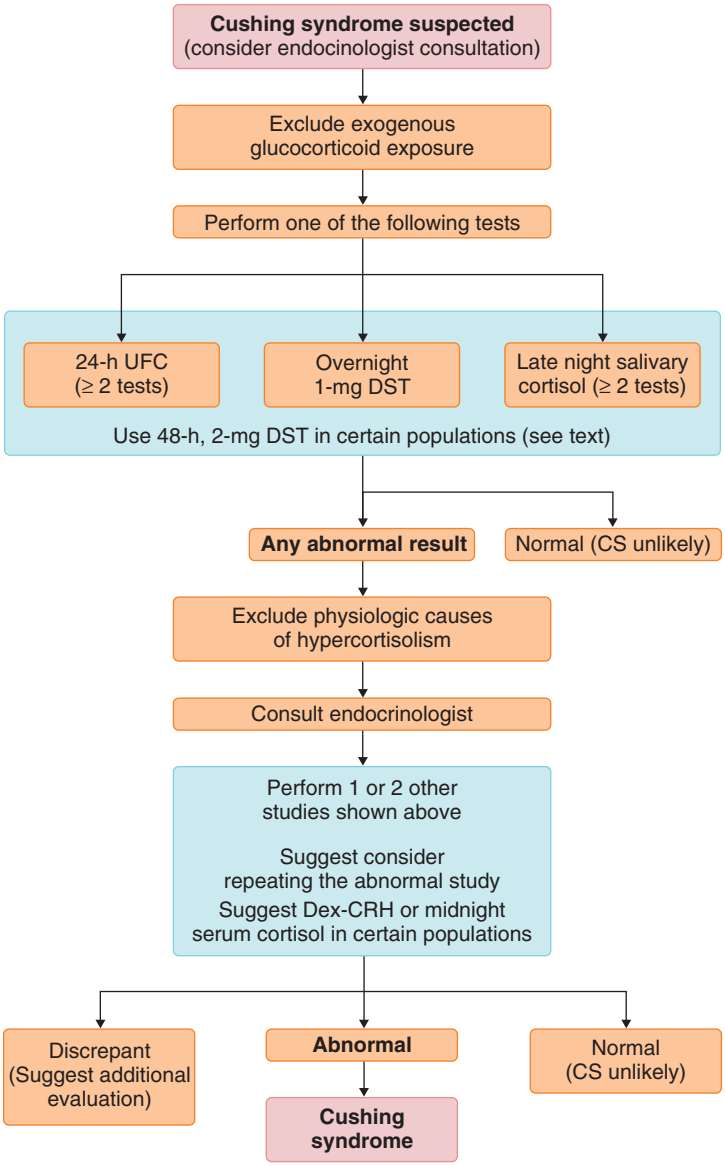

FIGURE 227-4. Algorithm for testing of patients suspected of having Cushing syndrome (CS). All statements are recommendations except for those prefaced by "suggest". Diagnostic criteria that suggest Cushing syndrome are urine free cortisol (UFC) greater than the normal range for the assay, serum cortisol greater than 1.8 µg/dL (50 nmol/liter) after 1 mg dexamethasone (1-mg DST), and late-night salivary cortisol greater than 145 ng/dL (4 nmol/liter). Dex-CRH = dexamethasone–corticotropin-releasing hormone test; DST = dexamethasone suppression test. (Reprinted with permission from Nieman LK, Biller BM, Findling JW, et al. The diagnosis of Cushing syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2008;93:1526-1540.)

Any dexamethasone test may give false results in patients with abnormal metabolic clearance of the drug. Agents that induce the cytochrome P-450 CYP3A4 enzymes (alcohol, rifampin, phenytoin, phenobarbital) increase dexamethasone clearance, whereas renal or hepatic failure decreases it. Measurement of a dexamethasone level can determine whether its clearance has been altered.

Differential Diagnosis

The causes of endogenous Cushing syndrome can be divided broadly into ACTH-dependent (80%) and ACTH-independent (20%) forms (Table 227-2). Hypercortisolism from autonomously functioning adrenal tumors suppresses ACTH, whereas in primary disorders of ACTH excess, the adrenal glands respond to tumor-derived ACTH. Plasma ACTH concentration distinguishes between these causes. ACTH is usually less than 10 pg/mL in primary adrenal disorders but is also suppressed by exogenous steroids, whether they are prescribed intentionally (iatrogenic Cushing syndrome) or taken factitiously. Patients in the latter group often have had multiple surgical procedures and do not reveal that they are self-administering steroids. As a result, patients must be queried closely about exogenous steroid administration, recognizing that parenteral, inhaled, and topical steroids can all cause glucocorticoid excess. Patients with endogenous Cushing syndrome and low ACTH concentrations should undergo adrenal imaging to identify the site of adrenal abnormality. Nonautonomous adrenal tissue atrophies when ACTH support is subnormal. Because of this, the common ACTH-independent forms of Cushing syndrome—adrenal adenoma and carcinoma—are manifested as a unilateral adrenal mass, with atrophy of the adjacent and contralateral tissue on magnetic resonance imaging or computed tomography.

TABLE 227-2 ETIOLOGY OF CUSHING SYNDROME

EXOGENOUS	ENDOGENOUS
Most common cause of Cushing syndrome: Glucocorticoid or ACTH driven May be factitious or iatrogenic	ACTH independent—autonomous adrenal activation (20% of all cases) Adrenal adenoma (40-50%) Adrenal carcinoma (40-50%) Primary pigmented nodular adrenal disease McCune-Albright syndrome Massive macronodular adrenal disease Gastric inhibitory polypeptide or food induced ACTH dependent—adrenal activation by excessive ACTH (80% of all cases) Corticotrope adenoma (80%) Ectopic ACTH secretion (20%) Ectopic CRH secretion (rare)

ACTH = adrenocorticotropic hormone; CRH = corticotropin-releasing hormone.

Bilateral forms of primary adrenal disease are rare and may be manifested with small or large adrenal nodules.[6] Primary pigmented nodular adrenal disease occurs primarily in children and young adults and is characterized by small to normal-sized adrenal glands containing small (<5 mm) black-brown cortical nodules. About half of these patients have additional features, termed *Carney complex*, which are often inherited in an autosomal dominant fashion. The clinical features of Carney complex include myxomas of the skin, breast, and heart; spotty pigmentation, such as lentigines and blue nevi; and other endocrine overactivity, such as acromegaly and testicular tumors. Some of these patients have mutations leading to a truncated form of protein kinase A regulatory 1α subunit. The resultant increase in protein kinase A activation by cyclic adenosine monophosphate presumably allows tumor formation. Bilateral nodular hyperplasia with Cushing syndrome can occur in the setting of *McCune-Albright syndrome*, mostly in infants or children. Massive macronodular adrenal disease generally is manifested after the age of 40 years with huge adrenal glands and aberrant expression of "illicit" receptors for various ligands (gastric inhibitory polypeptide, β-adrenergic, vasopressin), which presumably mediates autonomous cortisol production.

A normal or elevated plasma ACTH level (>15 pg/mL; 3.3 pmol/L) is consistent with an ACTH-producing tumor.[7] Intermediate ACTH concentrations between 5 and 15 pg/mL (1.1 to 3.3 pmol/L) in a two-site sandwich assay are not diagnostic. In these patients, suboptimal cortisol responses to CRH stimulation may identify the minority of cases of ACTH-independent Cushing syndrome with borderline basal ACTH values. In addition, a suppressed plasma DHEA-S value supports the diagnosis of an ACTH-independent disorder.

Cushing disease,[8] an ACTH-secreting pituitary adenoma, is the most common cause of Cushing syndrome. It is more common in women than in men (6 : 1 ratio), with a mean age at onset in the fourth decade. ACTH also may be secreted ectopically by a variety of neuroendocrine tumors, as shown in Table 227-3.

Pituitary magnetic resonance imaging shows a tumor in only about 40 to 50% of patients with Cushing disease, but it is obtained routinely in patients with ACTH-dependent disease to exclude a macroadenoma or abnormal anatomy before petrosal sinus sampling or surgery. A pituitary lesion less than 6 mm is seen in up to 10% of healthy individuals and so does not always indicate Cushing disease. Biochemical tests must be used to distinguish among the ACTH-dependent causes of Cushing syndrome, and they must be performed after a 6- to 8-week period of sustained hypercortisolism sufficient to suppress normal corticotrope function.

Inferior petrosal sinus sampling is the best test to distinguish between a pituitary and an ectopic source of excess ACTH; worldwide, the overall

TABLE 227-3 THE INCIDENCE AND TYPES OF TUMORS CAUSING THE SYNDROME OF ECTOPIC ACTH SECRETION

TUMOR TYPE	PERCENTAGE
Carcinoma of lung (small cell or oat cell)	19-50
Carcinoid of bronchus	2-37
Carcinoid of thymus	8-12
Pancreatic tumors, carcinoid and islet cell	4-12
Pheochromocytoma, neuroblastoma, ganglioma, paraganglioma	5-12
Medullary carcinoma of the thyroid	0-5
Miscellaneous*	<1

*Miscellaneous tumors reported to secrete ACTH in 1 to 10 cases include carcinoma of the ovary, prostate, breast, thyroid, kidney, salivary glands, testes, gallbladder, esophagus, and appendix; gastric carcinoid and renal carcinoid; acute myeloblastic leukemia; melanoma; and cloacogenic carcinoma of the anal canal.
ACTH = adrenocorticotropic hormone.

sensitivity and specificity are about 94%. The test involves catheterization of a peripheral vein and also the petrosal sinuses draining the pituitary gland; simultaneous measurement of ACTH levels at each site before and 3, 5, and 10 minutes after administration of CRH; and calculation of the central-to-peripheral ACTH ratio at each time point. Ratios of more than 2 before CRH administration or more than 3 after CRH administration are consistent with Cushing disease.

Although it is accurate in experienced hands, inferior petrosal sinus sampling carries a small risk of stroke, is expensive, and is not widely available. Other tests, such as the CRH test and the 8-mg dexamethasone suppression test, may be useful if both responses indicate Cushing disease. In this setting, the likelihood of ectopic ACTH secretion is low. However, the diagnosis is not clear if both responses are negative or if they are mixed.

If endocrine tests suggest ectopic ACTH secretion, imaging is obtained to localize the tumor. Computed tomography and magnetic resonance imaging of the chest are the best initial screens because these tumors are most often in the thoracic cavity. Octreotide scintigraphy is a useful adjunctive test. Measurement of serum calcitonin and gastrin and measurement of plasma or urine catecholamines may identify medullary carcinoma of the thyroid, gastrinoma, and pheochromocytoma. The process can be repeated every 6 to 12 months; tumors that make ACTH ectopically have a spectrum of malignant potential, and annual screening should continue, regardless of treatment for hypercortisolism.

TREATMENT

Rx

Surgical Therapy

The optimal treatment of Cushing syndrome is surgical resection of the lesion that is producing excessive ACTH or cortisol. In ACTH-dependent Cushing syndrome, if this is unsuccessful or cannot be done, bilateral adrenalectomy is an option.

Transsphenoidal resection of a microadenoma is the optimal therapy for a patient with Cushing disease, with up to a 90% chance of cure in the hands of an experienced neurosurgeon. A successful outcome is less likely if the initial surgery is not curative, in cases of recurrence, and for macroadenomas. Controversy exists about the criteria for remission; although a low postoperative cortisol level is encouraging, it does not preclude later recurrence. If recurrence develops, additional resection or alternative therapy should be considered.

Patients with ectopic ACTH secretion can be cured if the tumor can be removed and is not metastatic. Otherwise, adrenalectomy or medical therapy is chosen (see later). Adrenalectomy is appropriate when the patient cannot tolerate the medical toxicity, cost, or adverse psychological effects of long-term medical therapy and monitoring; when rapid correction of hypercortisolism is needed; or if maximal daily doses of ketoconazole (1600 mg) and metyrapone (2 g) given in combination do not render the patient eucortisolemic.

Nonmalignant primary adrenal causes of Cushing syndrome are cured by resection of the abnormal tissue. Laparoscopy is the preferred approach. Surgery is the mainstay in the treatment of adrenal cancer; multiple operations may be needed to resect primary lesions, local recurrences, and hepatic, thoracic, and intracranial metastases. Adjuvant adrenolytic therapy with mitotane may provide a chemotherapeutic benefit.

Radiation Therapy

Radiation therapy to the pituitary gland, with adjunctive therapy with steroidogenesis inhibitors to normalize cortisol levels, is a good option for patients with Cushing disease who cannot undergo surgery, for those in whom the risk of Nelson syndrome (Chapter 224) is deemed great, and for those with recurrent disease. This is usually delivered in 200-rad daily increments to a total dose of 4500 cGy. The disadvantage of radiation therapy is the length of time needed for a full response—up to 10 years—and the possibility of hypopituitarism. There is less experience with high-energy radiosurgery, such as the gamma knife, which has the advantage of requiring only one or two treatments. Adrenalectomy is preferable if rapid normalization of hypercortisolism is needed, and this option may be chosen by patients who have concerns about radiation-induced hypopituitarism and loss of reproductive function.

Medical Therapy

Medical therapy can also be used for patients with occult ectopic ACTH-secreting tumors or in combination with pituitary irradiation to treat Cushing disease.

Medical therapy with steroidogenesis inhibitors alone is rarely appropriate for Cushing disease because it requires close monitoring and adjustment of dose. Cabergoline or pasireotide may normalize UFC in 40% and 20% of patients, respectively.[8] There are limited data on the long-term efficacy of any medical therapy in Cushing disease.

For advanced adrenocortical carcinoma, rates of response and progression-free survival but not overall survival are significantly better with first-line therapy using a combination of etoposide (100 mg/m^2 on days 2 to 4), doxorubicin (40 mg/m^2 on day 1), and cisplatin (40 mg/m^2 on days 3 and 4) plus oral mitotane (to achieve a blood level of 14 to 20 mg/L) than with streptozocin plus mitotane as first-line therapy, with similar rates of toxic events.[8]

Mineralocorticoid Excess

DIAGNOSIS

Patients with mineralocorticoid excess often have few clinical symptoms apart from fatigue and muscle weakness or cramps related to hypokalemia. Most often the condition is suspected because of hypertension, especially if it occurs at an early age in association with spontaneous hypokalemia or is difficult to control.[9] Mineralocorticoid excess can result from primary adrenal disease, in which aldosterone (or another mineralocorticoid) is produced autonomously (and renin levels are low), or it may be due to nonadrenal causes as a result of elevated renin values, which stimulate aldosterone secretion. The latter situations include states of contracted arterial intravascular volume, such as congestive heart failure or cirrhosis with ascites, decreased renal arterial blood flow, and tumor production of renin (Table 227-4).

RENIN-INDEPENDENT MINERALOCORTICOID EXCESS

DIAGNOSIS

Although most of these conditions result from excessive aldosterone production by one or both adrenal glands, excessive production of other mineralocorticoids or constitutive activation of the renal sodium channel must be excluded. In these latter conditions, both aldosterone and renin values are low, resulting in the so-called syndrome of apparent mineralocorticoid excess. In this setting, diagnostic information is obtained by history (licorice ingestion) or measurement of other mineralocorticoids (see Table 227-4).

Primary hyperaldosteronism is diagnosed when there is an increased ratio (>20) of morning aldosterone to plasma renin activity (Fig. 227-5). One of four tests (usually salt loading) is used to confirm primary hyperaldosteronism by demonstrating a lack of aldosterone suppression.[10]

Differential Diagnosis

Having made the diagnosis of aldosterone-dependent mineralocorticoid excess, one must differentiate between the two most common adrenal causes—hyperplasia and adenoma—after excluding potential rare causes of hyperaldosteronism. Two rare autosomal dominant forms of familial hyperaldosteronism are type 1, a glucocorticoid-suppressible hyperaldosteronism, and type 2. Familial hyperaldosteronism type 1 is caused by a genetic swap of the promoter for *CYP11B1* (11β-hydroxylase) with that of *CYP11B2* (aldosterone synthase), forming a chimeric gene in which ACTH stimulates aldosterone synthase. It should be suspected in the setting of familial disease, particularly if there is a history of early-onset cardiovascular events, and is confirmed by gene testing (see http://www.brighamandwomens.org/Departments_and_Services/medicine/services/endocrine/Services/gra/default.aspx). The genetic abnormality in familial hyperaldosteronism

TABLE 227-4 CAUSES OF MINERALOCORTICOID EXCESS

PRIMARY HYPERALDOSTERONISM: HIGH ALDOSTERONE, LOW RENIN
Aldosterone-producing adenomas (30-50%)
Bilateral zona glomerulosa hyperplasia
Familial hyperaldosteronism
Type 1: glucocorticoid-remediable hyperaldosteronism—this results from formation of a chimeric gene containing the regulator portion of 11β-hydroxylase (normally regulated by ACTH) and the synthetic region of aldosterone synthase; as a result, ACTH stimulates aldosterone synthase and hence aldosterone production
Type 2: adrenal adenomas or hyperplasia expressed in a familial pattern
Type 3: caused by mutant KCNJ5, often younger and more severe than Type 2
Aldosterone-producing adrenal carcinoma
Ectopic aldosterone secretion (rare): kidney, ovary
SECONDARY HYPERALDOSTERONISM: HIGH ALDOSTERONE, HIGH RENIN
Renovascular hypertension and aortic stenosis
Diuretic use
Renin-secreting tumors
Severe cardiac failure
APPARENT MINERALOCORTICOID EXCESS: LOW ALDOSTERONE, LOW RENIN
Licorice ingestion: licorice (candy or flavored tobacco) containing glycyrrhetinic acid (or similar compounds such as carbenoxolone) inhibits renal 11β-hydroxysteroid dehydrogenase type 2, reducing cortisol conversion to cortisone and enabling cortisol to act as an endogenous mineralocorticoid
Severe hypercortisolism: similar in mechanism to licorice ingestion; very high cortisol levels are thought to overwhelm the ability of 11β-hydroxysteroid dehydrogenase type 2 to convert cortisol to cortisone in the kidney; cortisol itself then acts as a potent mineralocorticoid
Liddle's syndrome: mutation of the β or γ subunit of the collecting tubule sodium channel leads to a constitutive increase in sodium reabsorption and potassium excretion
11β-Hydroxylase deficiency form of congenital adrenal hyperplasia: 11-deoxycortisol accumulates because of an inability to convert it to cortisol
17-Hydroxylase deficiency form of congenital adrenal hyperplasia: deoxycorticosterone and corticosterone are increased

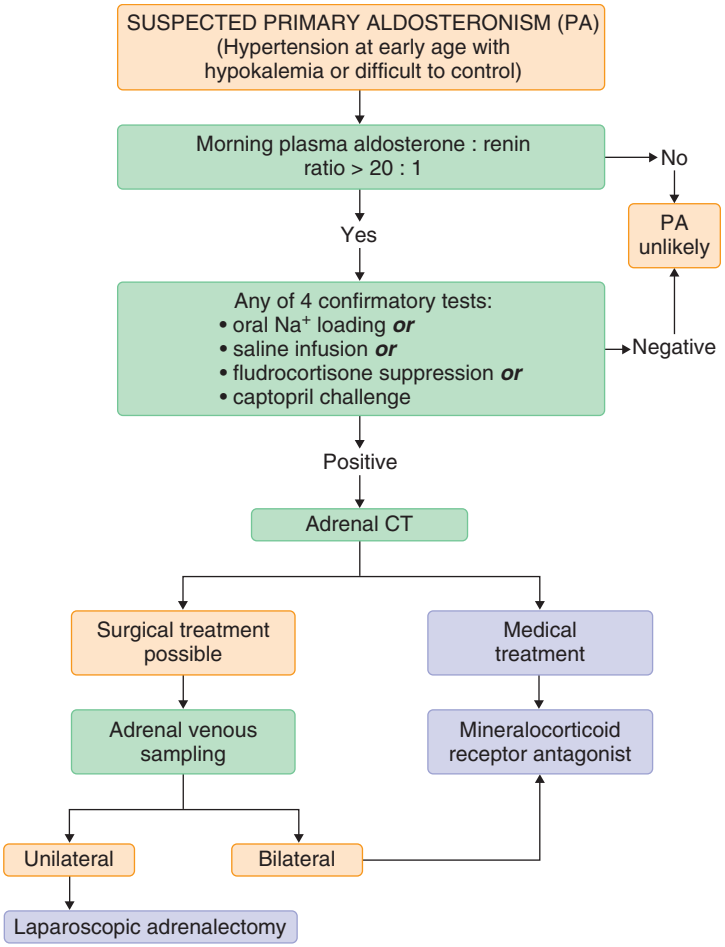

FIGURE 227-5. Algorithm for the detection, confirmation, subtype testing, and treatment of primary aldosteronism (PA). We recommend the case detection of PA in patient groups with relatively high prevalence of PA. These include patients with moderate, severe, or resistant hypertension; spontaneous or diuretic-induced hypokalemia; hypertension with adrenal incidentaloma; or a family history of early-onset hypertension or cerebrovascular accident at a young age (<40 years). We recommend use of the plasma aldosterone-to-renin ratio (ARR) to detect cases of PA in these patient groups. We recommend that patients with a positive ARR undergo testing, using any of four confirmatory tests, to definitively confirm or exclude the diagnosis. We recommend that all patients with PA undergo an adrenal computed tomography (CT) scan as the initial study in subtype testing and to exclude adrenocortical carcinoma. When surgical treatment is practicable and desired by the patient, the distinction between unilateral and bilateral adrenal disease should be made by adrenal venous sampling (AVS). We recommend that treatment by unilateral laparoscopic adrenalectomy be offered to patients with AVS-documented unilateral aldosterone-producing adenoma. If a patient is unable or unwilling to undergo surgery, we recommend medical treatment with a mineralocorticoid receptor (MR) antagonist. In patients with PA due to bilateral adrenal disease, we recommend medical treatment with an MR antagonist. *In patients with confirmed PA who have a family history of PA or of strokes at young age (<40 years), or with onset of hypertension earlier than at 20 years of age, we suggest genetic testing for glucocorticoid-remediable aldosteronism. In patients with glucocorticoid-remediable aldosteronism, we recommend the use of the lowest dose of glucocorticoid receptor agonist that can normalize blood pressure and serum potassium levels. (Reprinted, with slight modification of text, from Funder JW, Carey RM, Fardella C, et al. Case detection, diagnosis, and treatment of patients with primary aldosteronism: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab.* 2008;93:3266-3281.)

type 2 is not known; its clinical presentation is similar to sporadic hyperaldosteronism. Recent studies demonstrate rare germline mutations of a potassium channel in familial hyperalsosteronism, and somatic adrenal mutations in about 40% of patients.

For the more common conditions, adrenal computed tomography scans may show nonfunctioning nodules and falsely suggest an adenoma.[11] The responses to physiologic maneuvers, such as upright posture, and salt loading with oral or intravenous sodium tend to be preserved in patients with hyperplasia, but there is significant overlap among groups of patients. The best diagnostic test involves the measurement of cortisol and aldosterone in bilateral adrenal venous effluent and a peripheral vein before and during an ACTH infusion. Cortisol is used to evaluate catheter placement in the adrenal veins, as levels from the two sides should be similar. When an adenoma is present, the aldosterone-to-cortisol ratio on one side is usually at least five-fold greater than the other, which may be similar to the periphery, indicating suppression. Bilateral hyperplasia tends to produce similar values on each side.

TREATMENT

Rx

Treatment of primary hyperaldosteronism includes laparoscopic resection for adenomas.[12] Afterward, hypokalemia generally resolves, but hypertension persists in up to 65% of patients. A mineralocorticoid antagonist, spironolactone or eplerenone, is used to treat patients unable to undergo surgery or those with hyperplasia. Eplerenone is a more selective mineralocorticoid antagonist (with fewer side effects of sexual dysfunction and gynecomastia compared with spironolactone). A sodium channel blocker (e.g., amiloride) may be helpful, and antihypertensive agents are continued as needed.

Androgen Excess

DEFINITION

Women with excess circulating androgens or increased sensitivity to androgens present with complaints of hirsutism, acne, and anovulation or infertility. When testosterone is secreted in great excess, women may virilize and exhibit a deepened voice, clitorimegaly, masculinized body habitus, and alopecia.

DIAGNOSIS

The adrenal causes of hyperandrogenism—congenital adrenal hyperplasia, Cushing disease, adrenal cancer, and androgen-producing adrenal adenoma—are uncommon. Most women have no clear-cut cause (idiopathic hirsutism) or polycystic ovary syndrome. Rarely, androgen-secreting ovarian tumors, hyperprolactinemia, glucocorticoid resistance, or exogenous drugs cause hyperandrogenism. Patients with an adrenal source of hyperandrogenism usually have increased serum levels of DHEA, DHEA-S, or androstenedione, in contrast to the testosterone excess that is more typical of an ovarian source. DHEA and DHEA-S are weak androgens that can be converted locally to

testosterone in the hair follicles. Because DHEA and DHEA-S levels decline throughout adult life, these values must be interpreted within age-specific normal ranges. Although a tumor is more likely if DHEA-S is greater than 500 μg/dL or testosterone is greater than 200 ng/mL, it is not excluded at lower levels.

Imaging identifies nearly all adrenal tumors but may miss a small intraovarian one. UFC may be elevated in patients with virilizing adrenal carcinoma or Cushing disease (see earlier) and in those with glucocorticoid resistance. By contrast, androgen-secreting adrenal adenomas do not have glucocorticoid excess. In women suspected of having nonclassic forms of congenital adrenal hyperplasia, precursor and product hormones should be measured before and after ACTH to confirm the diagnosis.

TREATMENT Rx

Treatment of adrenal causes of hyperandrogenism varies according to the disorder. Classic congenital adrenal hyperplasia is treated by glucocorticoids to normalize ACTH and hence androgen levels (typically, dexamethasone 0.125 to 0.375 mg at bedtime). The nonclassic forms respond well to oral contraceptive or antiandrogen treatment, with dexamethasone reserved for ovulation induction. Surgery with adjunctive medical treatment may be used in adrenal carcinoma (see earlier).

Mixed Mineralocorticoid and Glucocorticoid Deficiency: Adrenal Insufficiency

PATHOBIOLOGY

Primary Adrenal Insufficiency

Autoimmune Destruction

Autoimmune destruction is the most common cause of primary adrenal insufficiency in industrialized countries and may occur alone or, rarely, in association with autoimmune polyglandular syndromes. These syndromes tend to be manifested either in childhood (type 1), in association with hypoparathyroidism and mucocutaneous candidiasis, or in adulthood (type 2), in association with insulin-dependent diabetes mellitus, autoimmune thyroid disease, alopecia areata, or vitiligo. The glands are small on imaging.

Adrenoleukodystrophy

Adrenoleukodystrophy, a rare (1 in 25,000) X-linked condition, is characterized by a deficiency of peroxisomal membrane adrenoleukodystrophy protein, which transports activated acyl-coenzyme A derivates into the peroxisomes, where they are shortened by β-oxidation. This deficiency results in the accumulation of very long chain fatty acids in the central nervous system and other tissues and increased plasma $C_{26:0}$ fatty acids. Incomplete penetrance of the genetic defect and variable accumulation of very long chain fatty acids in the adrenal gland, brain, testis, and liver account for the clinical phenotypes, which differ by age and presentation.[13]

Replacement of Adrenal Tissue

Infections cause about 15% of primary adrenal insufficiency. Typical infections include tuberculosis and systemic fungal diseases (histoplasmosis, coccidioidomycosis, blastomycosis), in which the adrenal tissue is replaced by caseating granulomas. End-stage AIDS-associated opportunistic infections, such as cytomegalovirus or *Mycobacterium avium-intracellulare*, may reduce adrenal function. Adrenal tissue may be replaced by bilateral metastases (most commonly primary carcinoma of the lung, breast, kidney, or gut) or primary lymphoma, although adrenal insufficiency is uncommon. Intraadrenal hemorrhage may also lead to insufficient steroidogenesis. Hemorrhage typically occurs in a stressed, hospitalized patient receiving long-term prophylactic anticoagulation and is often accompanied by back pain. The adrenal glands tend to be large on imaging.

Congenital Adrenal Hyperplasias

The congenital adrenal hyperplasias[14] are a disparate group of diseases caused by a genetic deficiency of one of the enzymes needed for adrenal steroidogenesis. Patients with nearly complete deficiency of an enzyme required for cortisol synthesis present in infancy with adrenal insufficiency and salt-wasting crisis. This is most problematic in patients with mutation of the 21-hydroxylase (*CYP21A2*) or 11β-hydroxylase (*CYP11B1*) gene. The increase in ACTH levels caused by cortisol deficiency drives the intact steroidogenic pathways so that there is excessive production of the steroids just proximal to the enzymatic block—17-hydroxyprogesterone and 11-deoxycortisol, respectively, in 21-hydroxylase and 11β-hydroxylase deficiency. The increased levels of precursor steroids enable increased adrenal androgen synthesis, so that severely affected girls may be virilized in utero. Girls and women with nonclassic congenital adrenal hyperplasia present later. They have greater enzyme activity, so that cortisol production is adequate, but increased ACTH levels cause hyperandrogenism.

Rare Causes

Other rare causes of primary adrenal insufficiency include ACTH resistance, congenital adrenal hypoplasia, Smith-Lemli-Opitz syndrome, and amyloidosis. Patients with primary adrenal insufficiency should undergo further evaluation to determine its cause (Table 227-5). Detection of antibodies to 21-hydroxylase identifies nearly all patients with idiopathic disease. In a male with negative results, measurement of plasma $C_{26:0}$ fatty acids will detect adrenoleukodystrophy. Taken together, this strategy identifies the cause in

TABLE 227-5 CAUSES OF ADRENAL INSUFFICIENCY AND ANCILLARY TESTS

SPECIFIC CAUSES	SUGGESTIVE CLINICAL FEATURES	USEFUL ANCILLARY TESTS
Primary adrenal insufficiency	Hyperpigmentation, orthostatic hypotension	Hyperkalemia, elevated ACTH
Idiopathic autoimmune destruction	Most common cause (80%) in developed countries; with or without other endocrinopathies, as below	Antibodies to 21-hydroxylase are present; on imaging, adrenal glands are small
Polyglandular failure type 1	Hypoparathyroidism, mucocutaneous candidiasis, vitiligo; age <20 years	
Polyglandular failure type 2	Insulin-dependent diabetes, autoimmune thyroid disease, alopecia areata, vitiligo; age >40 years	On imaging, adrenal glands are small
Infections: tuberculosis, systemic fungal diseases, AIDS-associated opportunistic infections (e.g., cytomegalovirus)	15% of patients in U.S. series	Adrenal glands tend to be large on CT and may be calcified
Space-occupying adrenal lesions	Metastases from carcinoma of lung, breast, kidney, gut; lymphoma or hemorrhage (heparin use)	Abnormal shape of adrenal glands on CT; evidence of hemorrhage
Bilateral adrenalectomy or treatment with steroidogenesis inhibitors		Ketoconazole, mitotane, aminoglutethimide, trilostane, and metyrapone reduce cortisol levels
Adrenoleukodystrophy	X-linked—screen males; in childhood, cognitive and gait disturbances; in adults, spastic paraparesis	Deficiency of peroxisomal very long chain acyl-coenzyme A synthetase leads to elevated plasma $C_{26:0}$ fatty acid levels
Secondary adrenal insufficiency		
Suppression of the adrenal axis by exogenous or endogenous glucocorticoids	Medication history; history of Cushing syndrome	Adrenal glands are small on imaging
Structural lesions of the hypothalamus or pituitary gland (tumors, destruction by infiltrating disorders, x-irradiation, and lymphocytic hypophysitis)	Other pituitary deficiencies	Adrenal glands are normal or small on imaging; MRI or CT may show pituitary or hypothalamic lesion
Isolated ACTH deficiency		
Head trauma		

ACTH = adrenocorticotropic hormone; AIDS = acquired immunodeficiency syndrome; CT = computed tomography; MRI = magnetic resonance imaging.

nearly all adult patients with idiopathic adrenal insufficiency. Patients with autoimmune disease should be tested for other endocrine deficiencies, and those with adrenoleukodystrophy require neurologic evaluation.

Secondary Adrenal Insufficiency

Suppression of the Pituitary Axis

Suppression of the hypothalamic-pituitary-adrenal axis by exogenous or endogenous glucocorticoids is the most common cause of secondary adrenal insufficiency. This phenomenon depends on the dose, duration, and schedule of glucocorticoid administration. Thus, adrenal suppression is unusual with "replacement" doses of glucocorticoid that are roughly equivalent to daily production (e.g., total daily doses of 20 mg hydrocortisone, 5 mg prednisone, or 0.3 to 0.5 mg dexamethasone). At higher doses, adrenal suppression is usually not seen until after 3 weeks of administration, and a single morning administration is less suppressive than are divided doses given during the day. When potentially suppressive doses of glucocorticoids are stopped, symptoms of adrenal insufficiency may occur within 48 hours, and the entire axis may not recover for up to 18 months. During this time, the patient should receive replacement glucocorticoid treatment or supplemental steroids at times of physiologic stress, depending on the degree of impairment (see later).

Lesions of the Hypothalamus or Pituitary

Secondary adrenal insufficiency also may result from structural lesions of the hypothalamus or pituitary gland that interfere with CRH production or transport or with corticotrope function. These causes include tumors, trauma, destruction by infiltrating disorders, x-irradiation, and lymphocytic hypophysitis. In general, these are not reversible conditions. Patients with secondary adrenal insufficiency not ascribed to glucocorticoid use should undergo imaging of the pituitary and hypothalamus to exclude a structural or infiltrating lesion as well as tests of other pituitary function to exclude additional deficiencies.

CLINICAL MANIFESTATIONS

The clinical presentation of adrenal insufficiency reflects the cause and duration of this uncommon condition. Primary adrenal insufficiency eventually destroys the entire adrenal cortex, with loss of both glucocorticoid and mineralocorticoid activity. By contrast, secondary adrenal insufficiency reflects an inability of the hypothalamic-pituitary unit to deliver CRH or ACTH, thus reducing trophic support to otherwise normal glands. As a result, only cortisol production decreases because mineralocorticoid production is not very ACTH dependent (Fig. 227-6).

The characteristic clinical presentation of acute primary adrenal insufficiency includes orthostatic hypotension, agitation, confusion, circulatory collapse, abdominal pain, and fever.[15] These features are most likely to be caused by hemorrhage, metastasis, or acute infection and can lead to death if not treated. In contrast, the typical history and clinical findings of chronic primary adrenal insufficiency include a longer history of malaise, fatigue, anorexia, weight loss, joint and back pain, and darkening of the skin (especially in the creases of the hands, extensor surfaces, recent scars, buccal and vaginal mucosa, and nipples). Patients may crave salt and may develop unusual food preferences, such as drinking the brine from pickles. Associated biochemical features for both acute and chronic presentations include hyponatremia, hypoglycemia, hyperkalemia, unexplained eosinophilia, and mild prerenal azotemia.

Chronic secondary adrenal insufficiency is manifested in a similar way, but without hyperpigmentation or mineralocorticoid abnormalities.

DIAGNOSIS

Biochemical testing confirms the diagnosis of adrenal insufficiency. A morning serum cortisol measurement is an inexpensive but relatively insensitive screening test for adrenal insufficiency in patients who are not acutely ill. The diagnosis is virtually excluded by values greater than 19 μg/dL (524 nmol/L) and is likely if the value is less than 3 μg/dL (83 nmol/L). However, both healthy individuals and patients with adrenal insufficiency may have indeterminate results (3 to 19 μg/dL) that require additional evaluation.

Patients with acute adrenal insufficiency should be evaluated for sepsis, adrenal metastases, and hemorrhage. Imaging of the glands and other testing may reveal an infectious cause. In acute adrenal insufficiency, a serum cortisol value is generally inappropriately normal or subnormal in the setting of hypotension, in which cortisol values are usually well above 18 μg/dL.

There is controversy about the best test to diagnose chronic adrenal insufficiency. Many use the cortisol response to exogenous ACTH as a gold standard test of adrenal steroidogenic ability. In the classic test, 250 μg of ACTH (1-24, cosyntropin) is given intravenously at any time of day. This dose of ACTH is a maximal stimulus to the adrenal gland, so that the serum cortisol level measured 30 to 60 minutes later is greater than 18 μg/dL. Lower values indicate adrenal insufficiency. Insulin-induced hypoglycemia, metyrapone stimulation, and lower doses of ACTH stimulation have been proposed as better tests for patients with mild or recent secondary adrenal insufficiency, who may respond to pharmacologic doses of ACTH. None of these is ideal. Also, because there is no commercial formulation of ACTH for lower-dose tests, the product must be diluted on site, leading to concerns about the accuracy of the administered dose and the validity of results. Metyrapone has limited availability in the United States.

Cerebral adrenoleukodystrophy, presenting in childhood, is characterized by cognitive and gait disturbances; the adult form, adrenomyeloneuropathy, is characterized by spinal cord and peripheral nerve demyelination. In both forms, the accumulation of very long chain fatty acids in the adrenal cortex alters membrane function and inhibits signal transduction by ACTH. Because a substantial minority of patients in both groups present first with adrenal insufficiency, boys and young men with adrenal insufficiency should be screened for adrenoleukodystrophy.

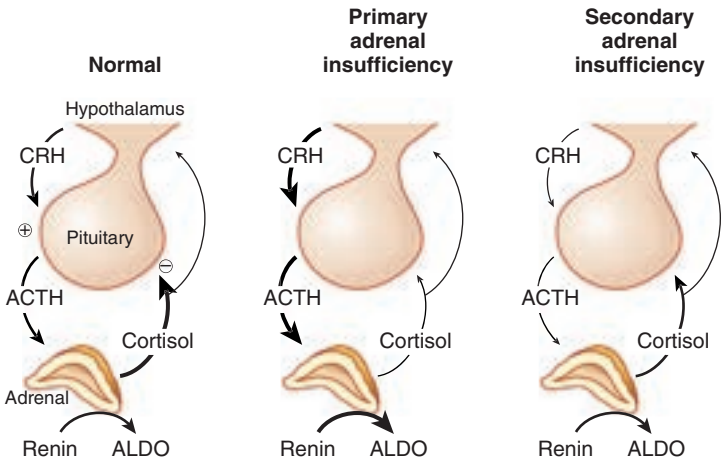

FIGURE 227-6. Physiology of the adrenal axis in health, primary adrenal insufficiency, and secondary adrenal insufficiency. ***In healthy individuals*** **(Normal), cortisol production is stimulated by the increased hypothalamic release of corticotropin-releasing hormone (CRH), which then travels down the pituitary stalk to stimulate adrenocorticotropic hormone (ACTH) secretion and release by corticotropes. Circulating ACTH stimulates adrenal gland production and secretion of cortisol. Cortisol then functions in a negative feedback mechanism to inhibit both CRH and ACTH.** ***In patients with primary adrenal insufficiency,*** **destruction or replacement of the entire adrenal cortex results in decreased cortisol, aldosterone, and dehydroepiandrosterone (DHEA, not shown) secretion by the adrenal glands. As a result of decreased cortisol negative feedback, the normal hypothalamus and pituitary gland increase CRH and ACTH secretion. The decreased aldosterone levels lead to an increase in renin levels.** ***In patients with secondary adrenal insufficiency,*** **ACTH or CRH secretion is reduced because of destruction or replacement of the hypothalamus or pituitary gland or because of disruption of the pituitary stalk. The decreased ACTH stimulation results in decreased cortisol (and, not shown, DHEA) secretion by the adrenal glands. Aldosterone production is only slightly affected by ACTH stimulation, and levels remain normal. The abnormal hypothalamus and pituitary gland do not increase CRH and ACTH secretion in response to the decreased cortisol negative feedback.**

Differential Diagnosis

Primary and secondary adrenal insufficiency can be distinguished by measurement of plasma ACTH. In primary adrenal insufficiency, ACTH levels are generally above the normal range and may exceed the normal range before the cortisol response to exogenous ACTH stimulation is subnormal. In addition, hyperkalemia and elevated renin values are characteristic of primary but not of secondary adrenal insufficiency, which is identified by a suppressed or inappropriately normal ACTH level.

TREATMENT Rx

Acute Adrenal Insufficiency

In suspected acute adrenal insufficiency, hydrocortisone is the treatment of choice because it has both glucocorticoid and mineralocorticoid activity.

Treatment with intravenous saline for volume expansion, glucose for hypoglycemia, and intravenous hydrocortisone (100 mg) is started immediately after placement of an intravenous line and withdrawal of blood for documentation of the cortisol value.

Chronic Adrenal Insufficiency

Therapy for chronic adrenal insufficiency[16] aims to provide the physiologic replacement of steroids. Glucocorticoid replacement is achieved by administration of 10 to 12 mg/m^2 of hydrocortisone daily in one to three oral doses, attempting to mimic the physiologic diurnal variation of cortisol concentrations. Hydrocortisone offers the advantage of multiple-dose tablets, which allows fine adjustment and splitting of the daily dose. Ideally, the morning dose is given as soon after waking as possible; for individuals who feel extremely fatigued in the morning before the agent is absorbed, a strategy of taking the medication 30 minutes before arising may be helpful. Although many patients do well with a single dose, others complain of pronounced fatigue in the afternoon and evening. For them, a split-dose regimen, in which about one third of the daily dose is given around 4:00 PM or two afternoon doses are given, may be useful.

Other glucocorticoids may be used for daily replacement therapy. Prednisone, 5 to 7.5 mg daily, has the advantage of a long half-life and may be particularly helpful in patients with afternoon or evening fatigue. Dexamethasone may be used, but because of variable interindividual metabolism, it is difficult to recommend a specific replacement dose; in addition, few options for fixed doses are available, so it is difficult to adjust the dose.

Patients with primary adrenal insufficiency should be encouraged to salt their food and not to limit salt intake. Nearly all patients also require a mineralocorticoid, such as fludrocortisone 50 to 300 µg/day. The dose is adjusted until plasma renin activity is normal. If a mineralocorticoid is not given, the dose of hydrocortisone or other steroid with mineralocorticoid activity is often mistakenly increased to reduce an "unwell" feeling or hyperkalemia or salt craving. However, if a supraphysiologic dose is given, the patient becomes cushingoid.

Patients with primary adrenal insufficiency also have decreased serum DHEA levels. Controversy exists about its replacement. A meta-analysis concluded that there is insufficient evidence to support its routine use in these patients.

Ensuring Proper Dosing

Education of the Patient

All patients receiving chronic glucocorticoid replacement therapy should be instructed that they must take the glucocorticoids as prescribed and that failure to take or to absorb the medication will lead to adrenal crisis and possibly death. They should wear medical information bracelets or necklaces that identify this requirement. It is important to educate patients and their families about glucocorticoid adjustment during physiologic stress conditions, including the emergency administration of intramuscular glucocorticoid by means of a kit containing prefilled syringes with injectable steroid.

Dosing for Stress

The daily oral glucocorticoid dose is usually doubled for "stressful" physiologic conditions, such as fever, nausea, and diarrhea, although there are few data to support this strategy. In addition, this practice may lead to chronic overmedication by the patient because of a liberal interpretation of what constitutes physical stress. Thus, education about when and how to change the dose of steroid should be reinforced periodically, preferably with written material, and the dangers of excessive steroid use should be emphasized. If the patient is vomiting, has severe diarrhea, or has collapsed, intramuscular glucocorticoids should be given before transport to a medical facility.

The glucocorticoid dose is increased in proportion to the amount of stress. Thus, during maximally stressful situations (e.g., adrenal crisis, major surgery, trauma, labor and delivery), the daily hydrocortisone dose is 100 to 300 mg. Few data support the need for this supraphysiologic dose, but the safety of not following this practice has not been established. The dose may be tapered by 50% per day if the patient is clinically stable. For more moderate stress, such as that of cholecystectomy, 75 to 100 mg of hydrocortisone is given on the day of surgery, and the dose is tapered more rapidly. Patients undergoing minimal stress, such as tooth extraction or short operative orthopedic procedures, may not require any additional supplementation.

Assessment to Ensure Proper Dosing

Clinical assessment is the best way to judge whether the glucocorticoid dose is correct. Symptoms of adrenal insufficiency improve with adequate therapy. The development of cushingoid features or osteopenia suggests frank or subtle overreplacement, respectively, and the presence of adrenal insufficiency symptoms (fatigue, anorexia, weight loss) suggests underreplacement. In women, DHEA replacement increases testosterone levels, so that hirsutism, acne, or other signs of androgen excess may suggest overreplacement. In primary adrenal insufficiency, adequate hormone replacement results in plasma ACTH levels that decrease but remain elevated, in the range of 100 to 200 pg/mL. Renin values, however, normalize completely and may be used to judge the adequacy of mineralocorticoid replacement. Although hydrocortisone is metabolized to cortisol, plasma cortisol values should not be used to monitor therapy because clearance from the blood stream is rapid, and circulating values are low for most of the day. UFC does not reflect adequate replacement; the increase in plasma cortisol levels after a single daily dose may exceed corticosteroid-binding globulin capacity, resulting in excessive urine levels and overestimation of integrated cortisol levels.

Mineralocorticoid Deficiency

Hypoaldosteronism may be classified as a low-normal or a high renin state on the basis of plasma renin activity after 4 hours of upright posture. Renin deficiency is the most common cause of hypoaldosteronism, occurring most often in older patients with mild, nonoliguric renal disease who often have insulin-dependent diabetes and potentially diabetic nephropathy. Indomethacin and other prostaglandin synthesis inhibitors as well as autonomic dysfunction associated with prolonged bedrest can also result in hyporeninemic hypoaldosteronism.

CLINICAL MANIFESTATIONS

There are few clinical features associated with mineralocorticoid deficiency; as a result, it is usually suspected when laboratory results reveal hyperkalemia, hyponatremia, and a mild metabolic alkalosis. If glucocorticoid deficiency is excluded, isolated hypoaldosteronism is established if the circulating level of aldosterone is inappropriately low.

High renin states of hypoaldosteronism include congenital adrenal hyperplasias with mineralocorticoid deficiency and primary adrenal insufficiency when it is treated with pure glucocorticoid replacement.

Treatment of these conditions involves sodium replacement with at least 10 mEq/kg/day, roughly equivalent to the 4 g of sodium chloride found in a typical diet in the United States. For individuals who do not maintain such a diet (often the elderly or the young), fludrocortisone can be given at the same doses used in primary adrenal insufficiency.

Grade A References

A1. Colao A, Petersenn S, Newell-Price J, et al. A 12-month phase 3 study of pasireotide in Cushing disease. *N Engl J Med*. 2012;366:914-924.

A2. Fassnacht M, Terzolo M, Allolio B, et al. Combination chemotherapy in advanced adrenocortical carcinoma. *N Engl J Med*. 2012;366:2189-2197.

A3. Alkatib AA, Cosma M, Elamin MB, et al. A systematic review and meta-analysis of randomized placebo-controlled trials of DHEA treatment effects on quality of life in women with adrenal insufficiency. *J Clin Endocrinol Metab*. 2009;94:3676-3681.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Briet M, Schiffrin EL. The role of aldosterone in the metabolic syndrome. *Curr Hypertens Rep*. 2011;13:163-172.
2. Dekkers OM, Horváth-Puhó E, Jørgensen JO, et al. Multisystem morbidity and mortality in Cushing syndrome: a cohort study. *J Clin Endocrinol Metab*. 2013;98:2277-2284.
3. Nieman LK, Biller BM, Findling JW, et al. The diagnosis of Cushing syndrome: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2008;93:1526-1540.
4. Manetti L, Rossi G, Grasso L, et al. Usefulness of salivary cortisol in the diagnosis of hypercortisolism: comparison with serum and urinary cortisol. *Eur J Endocrinol*. 2013;168:315-321.
5. Graham UM, Hunter SJ, McDonnell M, et al. A comparison of the use of urinary cortisol to creatinine ratios and nocturnal salivary cortisol in the evaluation of cyclicity in patients with Cushing syndrome. *J Clin Endocrinol Metab*. 2013;98:E72-E76.
6. Louiset E, Duparc C, Young J, et al. Intraadrenal corticotropin in bilateral macronodular adrenal hyperplasia. *N Engl J Med*. 2013;369:2115-2125.
7. Else T, Kim AC, Sabolch A, et al. Adrenocortical carcinoma. *Endocr Rev*. 2014;35:282-326.
8. Tritos NA, Biller BM. Cushing disease. *Handb Clin Neurol*. 2014;124:221-234.
9. Calhoun DA. Hyperaldosteronism as a common cause of resistant hypertension. *Annu Rev Med*. 2013;64:233-247.
10. Funder JW, Carey RM, Fardella C, et al. Case detection, diagnosis, and treatment of patients with primary aldosteronism: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2008;93:3266-3281.
11. Nieman LK. Approach to the patient with an adrenal incidentaloma. *J Clin Endocrinol Metab*. 2010;95:4106-4113.
12. Harvey AM. Hyperaldosteronism: diagnosis, lateralization, and treatment. *Surg Clin North Am*. 2014;94:643-656.
13. Horn MA, Retterstøl L, Abdelnoor M, et al. Adrenoleukodystrophy in Norway: high rate of de novo mutations and age-dependent penetrance. *Pediatr Neurol*. 2013;48:212-219.
14. Auschus RJ, Arlt W. Approach to the patient: the adult with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2013;98:2645-2655.
15. Charmandari E, Nicolaides NC, Chrousos GP. Adrenal insufficiency. *Lancet*. 2014;383:2152-2167.
16. Falorni A, Minarelli V, Morelli S. Therapy of adrenal insufficiency: an update. *Endocrine*. 2013;43:514-528.

REVIEW QUESTIONS

1. A 35-year-old woman has had weight gain, insomnia, decreased short-term memory, and depression for the last year. She states that her symptoms are worsening. Two recent urine free cortisol (UFC) test results are three-fold increased. She takes lisinopril for recent-onset hypertension. What is the next best diagnostic test?

A. A 1-mg dexamethasone suppression test
B. A pituitary magnetic resonance imaging study
C. An adrenocorticotropic hormone (ACTH) measurement
D. Another UFC determination
E. A dexamethasone–corticotropin-releasing hormone test

Answer: A Recent guidelines and expert opinion suggest that at least two different screening test results should be abnormal to confirm the clinical suspicion of Cushing syndrome (see Fig. 227-4). This patient has had two abnormal UFC results and should undergo another screening test, either measurement of late-night salivary cortisol or a 1-mg or 2-day 2-mg dexamethasone suppression test. Another UFC determination will not obviate the need for a different screening test. The dexamethasone–corticotropin-releasing hormone stimulation test is more cumbersome and expensive than the dexamethasone suppression test and would not be a better choice. Tests that are used for the differential diagnosis, such as ACTH level or magnetic resonance imaging, should not be used to establish the diagnosis of Cushing syndrome.

2. A 32-year-old white woman with known idiopathic primary adrenal insufficiency complains of recent fatigue and a feeling of being unwell. This has been worse lately when she has been more active with gardening and is worse in the afternoon and evening. She denies dizziness, anorexia, increase in pigmentation, or change in weight. On physical examination, she weighs 125 pounds, with a height of 63 inches. Blood pressure is 110/75, and pulse is 86 without orthostasis. She has pigmented palmar creases and some buccal pigmentation, unchanged from the last visit. There are no other skin lesions. She does not appear cushingoid, there is no abdominal tenderness, and the joint examination findings are normal. Her only medications are hydrocortisone, 15 mg in the morning and 5 mg in the afternoon, and fludrocortisone, 50 mg every other day. What is the next best step?

A. Measure a plasma renin level
B. Increase the afternoon hydrocortisone dose to 10 mg
C. Measure the urine free cortisol excretion
D. Add a late afternoon hydrocortisone dose of 5 mg
E. Measure plasma ACTH before the morning hydrocortisone dose

Answer: A This woman has somewhat nonspecific symptoms that are exacerbated by gardening (perhaps by heat). She is receiving adequate hydrocortisone replacement but taking a low fludrocortisone dose. Such nonspecific symptoms may occur in the setting of relative mineralocorticoid deficiency, which can be evaluated through renin measurement. If the renin were elevated, an increased fludrocortisone dose might eliminate her vague symptoms of being unwell. The history and the amount of prescribed hydrocortisone do not suggest that she needs additional glucocorticoid. If she is mineralocorticoid deficient, she might feel better with a higher hydrocortisone dose because of its mineralocorticoid activity; however, a slightly supraphysiologic dose would put her at risk for osteopenia and other features of glucocorticoid excess. (See the section on adrenal insufficiency, assessment to ensure proper dosing.) Measurement of ACTH and urine free cortisol usually is not helpful in assessing adequacy of dosage.

3. A 30-year-old woman presents to the emergency department with the acute onset of nausea, vomiting, diarrhea, and dizziness. Her only medication is levothyroxine, 88 mcg daily. On examination, she is hypotensive, tachycardic, and orthostatic. She has dark skin, and no hyperpigmentation is seen on cursory examination. She has abdominal pain and increased bowel sounds. She is extremely weak but able to move all extremities and to respond to questions. The sodium concentration is 129 mEq/L, potassium concentration is 5.3 mEq/L, and blood urea nitrogen level is 40 mg/dL. Apart from initiating fluid resuscitation and evaluation for sepsis, what treatment and testing should be ordered immediately?

A. Draw blood for cortisol measurement before giving hydrocortisone 100 mg IV.
B. Give dexamethasone 8 mg and perform an ACTH stimulation test.
C. Measure an ACTH level and give hydrocortisone 100 mg IV.
D. Give dexamethasone 8 mg and defer further testing.
E. Give hydrocortisone 100 mg IV and obtain an adrenal computed tomography scan to evaluate for hemorrhage.

Answer: A This woman presents with the signs and symptoms of acute primary adrenal insufficiency. The knowledge that she takes thyroid hormone suggests polyglandular failure. However, it is possible that this constellation of features is caused by sepsis, renal failure, or hypothyroidism. Thus, it is prudent to measure cortisol in blood obtained during intravenous line insertion to confirm the diagnosis of adrenal insufficiency. Other biochemical and stimulation testing can be deferred. She has no apparent reason to have adrenal hemorrhage. Dexamethasone 8 mg is more than what is needed for treatment of adrenal insufficiency; hydrocortisone 100 mg is the recommended initial dose. (See the section on adrenal insufficiency, replacement of adrenal tissue, and treatment.)

4. A 50-year-old man is evaluated for hyperaldosteronism because of difficult-to-control hypertension. He is otherwise healthy. Evaluation was done during treatment with agents that do not affect aldosterone or renin levels and include an aldosterone-to-renin ratio of more than 40 and lack of aldosterone suppression to oral salt loading. A computed tomography (CT) scan of the adrenal glands shows a 1-cm mass on the right with a density of 10 HU. What is the next best step?

A. Adrenal venous sampling
B. Right laparoscopic adrenalectomy
C. Treatment with eplerenone
D. Review of CT scan to exclude hyperplasia of the left adrenal gland
E. Ultrasound ablation of the adrenal gland

Answer: A Adrenal venous sampling should be performed to identify the source of excess aldosterone. It is possible that the visualized mass is a nonfunctioning tumor. This man is a good surgical candidate, and laparoscopic adrenalectomy would be recommended if venous sampling is consistent with unilateral disease. The CT scan is not helpful in identifying bilateral disease. Treatment with eplerenone and ultrasound ablation would be possibilities if he were not a surgical candidate; eplerenone would be an appropriate treatment for bilateral disease. (See section on mineralocorticoid excess, diagnosis and treatment.)

5. A 35-year-old woman presents with a 6-month history of increasing body hair growth. Before that time, she states that she was mildly "hairy" but now needs to shave her face daily. (She has stopped shaving for 3 days so that her hair growth would be visible.) She has had only two episodes of vaginal spotting during this time. Her menses started at the age of 13 years and are usually irregular, occurring every 30 to 45 days. The extra hair growth started in her teens and until 6 months ago was similar to the hair pattern of her mother and sister. She admits to some recent irritability and perhaps has gained 5 pounds. She has always had a low voice and does not notice any recent change. She has a 3-year-old daughter. She has been able to continue her 35 minutes of daily cardiovascular exercise. On physical examination, body mass index is 30, blood pressure 125/85 mm Hg, pulse 82. The Ferriman-Galway score is 17. Fat distribution is normal with some increased abdominal fat, and she appears "well developed" but not especially muscular. There is no temporal hair recession, but her scalp hair appears thin. Acne is present on the chest and back. The clitoral index is 100 mm. Laboratory results include low luteinizing hormone and follicle-stimulating hormone, estradiol 50 pg/mL, dehydroepiandrosterone sulfate (DHEA-S) 653 ng/dL (reference range, 4 to 332), and testosterone 235 ng/dL. These abnormalities are confirmed on a second set of tests. What is the next best diagnostic step?

A. Obtain a CT scan of the adrenal glands
B. Obtain an ovarian ultrasound
C. Perform a 1-mg overnight dexamethasone suppression test
D. Measure a morning17-hydroxyprogesterone level
E. Measure an androstenedione level.

Answer: A This woman has recent-onset hirsutism and virilization. Whereas worsening polycystic ovary syndrome is a possibility, the degree of DHEA-S and testosterone elevation suggests the possibility of an adrenal tumor. The slight weight gain and irritability might also point to possible cortisol excess. Although androstenedione would confirm an adrenal source, it is unlikely to point to another cause, and so proceeding to adrenal CT scan is reasonable, particularly because adrenal cancer is part of this differential diagnosis. An ovarian ultrasound is less likely to be abnormal (i.e., an ovarian tumor is less likely). Congenital adrenal hyperplasia is consistent with these signs and symptoms but not with the time course of rapid worsening. Cushing syndrome is a possibility, but it is less likely to be Cushing disease with these levels of hormones; adrenal cancer is a possibility that would be identified by the CT scan, making a 1-mg dexamethasone suppression test less urgent. (See the section on androgen excess.)

228

ADRENAL MEDULLA, CATECHOLAMINES, AND PHEOCHROMOCYTOMA

WILLIAM F. YOUNG, JR.

ADRENAL MEDULLA AND CATECHOLAMINES

The adrenal medulla occupies the central portion of the adrenal gland. Adrenomedullary cells are called chromaffin cells (they stain brown with chromium salts). Chromaffin cells differentiate in the center of the adrenal gland in response to cortisol; some chromaffin cells also migrate to form paraganglia. The largest cluster of chromaffin cells outside the adrenal medulla is near the level of the inferior mesenteric artery and is referred to as the organ of Zuckerkandl.

The term *catecholamine* refers to substances that contain catechol (*o*-dihydroxybenzene) and a side chain with an amino group—the catechol

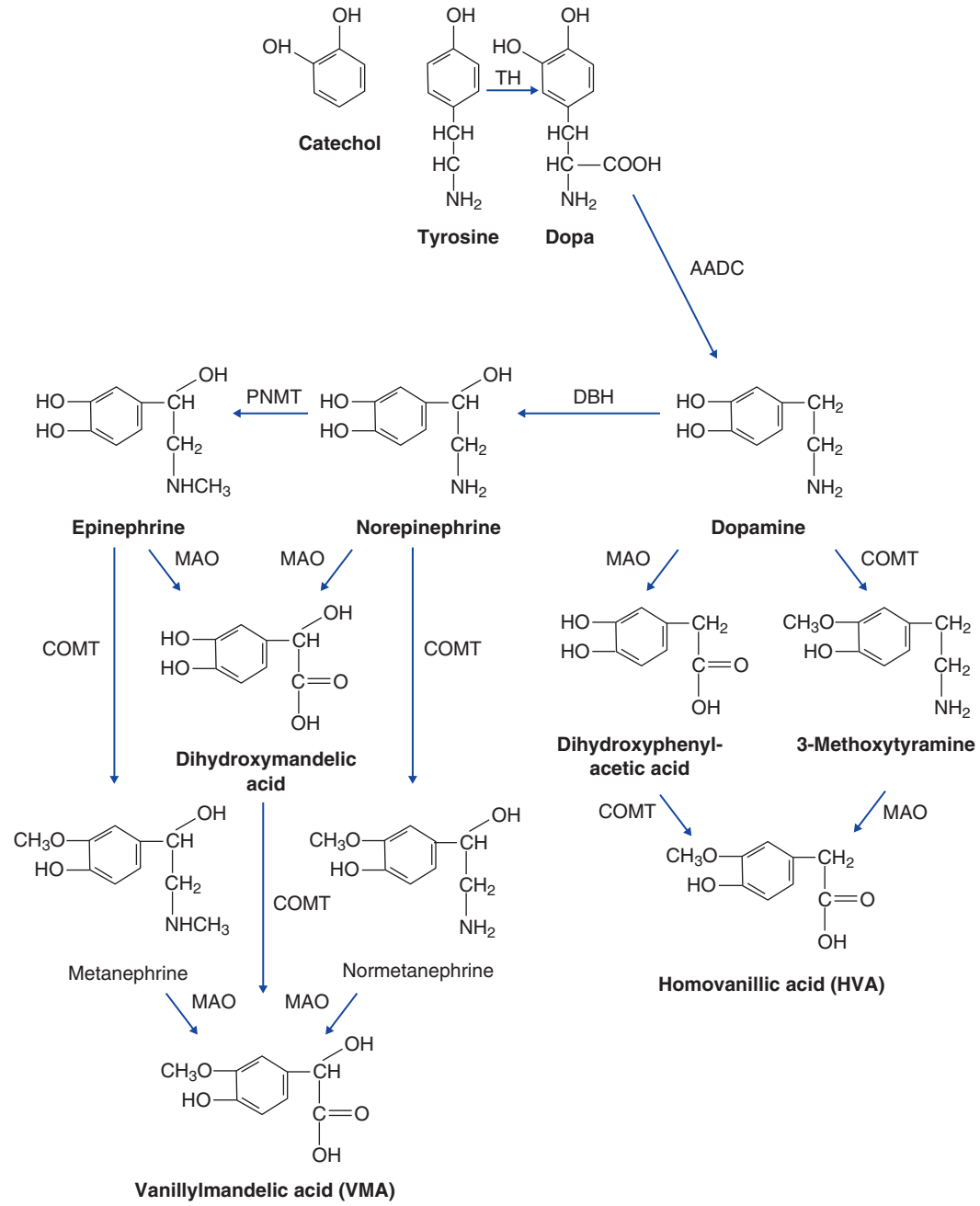

FIGURE 228-1. Biosynthetic and metabolic pathways for catecholamines. The term *catecholamine* comes from the catechol (*o*-dihydroxybenzene) structure and a side chain with an amino group—the catechol nucleus (*top left*). Tyrosine is converted to 3,4-dihydroxyphenylalanine (dopa) in the rate-limiting step by tyrosine hydroxylase (TH). Aromatic L-amino acid decarboxylase (AADC) converts dopa to dopamine. Dopamine is hydroxylated to norepinephrine by dopamine β-hydroxylase (DBH). Norepinephrine is converted to epinephrine by phenylethanolamine *N*-methyltransferase (PNMT); cortisol serves as a cofactor for PNMT, and this is why epinephrine-secreting pheochromocytomas are almost exclusively localized to the adrenal medulla. Metabolism of catecholamines occurs through two enzymatic pathways. Catechol-*O*-methyltransferase (COMT) converts epinephrine to metanephrine and norepinephrine to normetanephrine by meta-*O*-methylation. Metanephrine and normetanephrine are oxidized by monoamine oxidase (MAO) to vanillylmandelic acid by oxidative deamination. Monoamine oxidase also may oxidize epinephrine and norepinephrine to dihydroxymandelic acid, which is then converted by catechol-*O*-methyltransferase to vanillylmandelic acid. Dopamine is also metabolized by monoamine oxidase and catechol-*O*-methyltransferase, with the final metabolite homovanillic acid.

nucleus (Fig. 228-1). Epinephrine is synthesized and stored in the adrenal medulla and released into the systemic circulation. Norepinephrine is synthesized and stored not only in the adrenal medulla but also in the peripheral sympathetic nerves. Dopamine, the precursor of norepinephrine, is found in the adrenal medulla and peripheral sympathetic nerves.

Catecholamines affect many cardiovascular and metabolic processes, including increasing the heart rate, blood pressure, myocardial contractility, and cardiac conduction velocity. Specific receptors mediate the biologic actions. The three types of adrenergic receptors are α, β, and DA; their receptor subtypes are α_1, α_2, β_1, β_2, β_3, DA_1, and DA_2. The α_1 subtype is a postsynaptic receptor that mediates vascular and smooth muscle contraction; stimulation causes vasoconstriction and increased blood pressure. The α_2-receptors are located on presynaptic sympathetic nerve endings and, when activated, inhibit the release of norepinephrine; stimulation causes suppression in central sympathetic outflow and decreased blood pressure. Stimulation of the β_1-receptor causes positive inotropic and chronotropic effects on the heart, increased renin secretion in the kidney, and lipolysis in adipocytes as well as bronchodilation, vasodilation in skeletal muscle, glycogenolysis, and increased release of norepinephrine from sympathetic nerve terminals. The β_3-receptor regulates energy expenditure and lipolysis. DA_1 receptors are localized to the cerebral, renal, mesenteric, and coronary vasculature; stimulation causes vasodilation in these vascular beds. DA_2 receptors are presynaptic and localized to sympathetic nerve endings, sympathetic ganglia, and brain; stimulation inhibits the release of norepinephrine, inhibits ganglionic transmission, and inhibits prolactin release.

Catecholamines are synthesized from tyrosine by a process of hydroxylation and decarboxylation (see Fig. 228-1). Tyrosine is derived from ingested food or synthesized from phenylalanine in the liver, and it enters

neurons and chromaffin cells by active transport. Tyrosine is converted to 3,4-dihydroxyphenylalanine (dopa) by tyrosine hydroxylase, the rate-limiting step in catecholamine synthesis. α-Methyl-*p*-tyrosine (metyrosine) is a tyrosine hydroxylase inhibitor that may be used therapeutically in patients with catecholamine-secreting tumors. Aromatic L-amino acid decarboxylase catalyzes the decarboxylation of dopa to dopamine. Dopamine is actively transported into granulated vesicles to be hydroxylated to norepinephrine by the dopamine β-hydroxylase. These reactions occur in the synaptic vesicle of adrenergic neurons and the chromaffin cells of the adrenal medulla. In the adrenal medulla, norepinephrine is released from the granule into the cytoplasm, where phenylethanolamine *N*-methyltransferase converts it to epinephrine. Expression of phenylethanolamine *N*-methyltransferase is positively regulated by glucocorticoids. Thus, catecholamine-secreting tumors that secrete primarily epinephrine are localized to the adrenal medulla. In normal adrenal medullary tissue, approximately 80% of the catecholamine released is epinephrine.

The biologic half-life of circulating catecholamines is between 10 and 100 seconds. Thus, plasma concentrations of catecholamines fluctuate widely. Catecholamines are removed from the circulation by either reuptake by sympathetic nerve terminals or metabolism through two enzyme pathways (see Fig. 228-1), followed by sulfate conjugation and renal excretion. Almost 90% of catecholamines released at sympathetic synapses are taken up locally by the nerve endings (uptake 1). Uptake 1 can be blocked by cocaine, tricyclic antidepressants, and phenothiazine. Extraneuronal tissues also take up catecholamines (uptake 2). Most of these catecholamines are metabolized by catechol-*O*-methyltransferase. Metanephrine and normetanephrine are oxidized by monoamine oxidase to vanillylmandelic acid by oxidative deamination. Monoamine oxidase may also oxidize epinephrine and norepinephrine to 3,4-dihydroxymandelic acid, which is then converted by catechol-*O*-methyltransferase to vanillylmandelic acid. In the storage vesicle, norepinephrine is protected from metabolism by monoamine oxidase. Monoamine oxidase and catechol-*O*-methyltransferase metabolize dopamine to homovanillic acid (see Fig. 228-1).

PHEOCHROMOCYTOMA AND PARAGANGLIOMA

DEFINITION

Catecholamine-secreting tumors that arise from chromaffin cells of the adrenal medulla and the sympathetic ganglia are referred to as pheochromocytomas and extra-adrenal catecholamine-secreting paragangliomas, respectively. Because the tumors have similar clinical presentations and are treated with similar approaches, many clinicians use the term *pheochromocytoma* to refer to both entities. However, the distinction between pheochromocytoma and paraganglioma is an important one because there are differences in the risk for associated neoplasms, risk for malignant transformation, and type of genetic testing that should be considered.

EPIDEMIOLOGY

Catecholamine-secreting tumors are rare; the annual incidence is 2 to 8 cases per million people. Nevertheless, it is important to suspect, confirm, localize, and resect these tumors. The associated hypertension is curable with surgical removal of the tumor, a risk of lethal paroxysm exists, and at least 10% of the tumors are malignant. Approximately 30% of cases are familial, so detection of this tumor in the proband may result in early diagnosis in other family members.

PATHOBIOLOGY

Genetics

Approximately 30% of patients with catecholamine-secreting tumors have germline mutations (inherited mutations present in all cells of the body) in genes associated with the genetic disease.[1] Hereditary catecholamine-secreting tumors typically are manifested at a younger age than sporadic neoplasms are. Sporadic pheochromocytoma is typically diagnosed on the basis of symptoms or incidental discovery on computed imaging, whereas syndromic pheochromocytoma and paraganglioma are frequently diagnosed earlier in the course of disease because of biochemical surveillance or genetic testing.

Multiple Endocrine Neoplasia

Multiple endocrine neoplasia (MEN) type 2A is an autosomal dominant disorder (Chapters 231 and 246). The phenotype includes adrenal pheochromocytoma in 50% (usually bilateral and may be asynchronous), medullary carcinoma of the thyroid in 100%, hyperparathyroidism in 20 to 30%, and cutaneous lichen amyloidosis in 5%.[2] Medullary carcinoma of the thyroid is usually detected before pheochromocytoma. Numerous activating mutations in the *RET* proto-oncogene have been documented in persons with MEN type 2A (these are described in detail in Chapter 246).

MEN type 2B is also an autosomal dominant disorder, and it represents approximately 5% of all MEN type 2 cases. The phenotype includes pheochromocytoma in 50% (usually bilateral), aggressive medullary carcinoma of the thyroid in 100%, mucosal neuromas (typically involving the tongue, lips, and eyelids) in most patients, thickened corneal nerves, intestinal ganglioneuromatosis, and marfanoid body habitus. MEN 2B–associated tumors are caused by mutations in the RET protein's intracellular domain, as described in detail in Chapter 246.

Von Hippel–Lindau Disease

Von Hippel–Lindau (VHL) disease is an autosomal dominant disorder characterized by pheochromocytoma (frequently bilateral), paraganglioma (mediastinal, abdominal, pelvic), hemangioblastoma (involving the cerebellum, spinal cord, or brain stem), retinal angioma, clear cell renal cell carcinoma, pancreatic neuroendocrine tumor, endolymphatic sac tumor of the middle ear, serous cystadenoma of the pancreas, and papillary cystadenoma of the epididymis and broad ligament.[3] Pheochromocytoma occurs in about 10 to 20% of patients with VHL disease. Nearly 100% of patients have an identifiable gene mutation (VHL tumor suppressor gene). Certain missense mutations appear to be associated with a "pheochromocytoma only" presentation of VHL disease.

Neurofibromatosis

Neurofibromatosis type 1 (NF1) is an autosomal dominant disorder characterized by neurofibromas, multiple café au lait spots, axillary and inguinal freckling, iris hamartomas (Lisch nodules), bone abnormalities, central nervous system gliomas, pheochromocytoma and paraganglioma, macrocephaly, and cognitive deficits. The expression of these features is variable. Approximately 2% of patients with NF1 develop catecholamine-secreting tumors; in these patients, the tumor is usually a solitary benign adrenal pheochromocytoma, occasionally bilateral adrenal pheochromocytomas, and rarely an abdominal paraganglioma. Inactivating *NF1* mutations cause the disorder (NF1 tumor suppressor gene).

Familial Paraganglioma

Familial paraganglioma is an autosomal dominant disorder characterized by paragangliomas that are located in the skull base and neck, thorax, abdomen, and pelvis. Most cases of familial paraganglioma are caused by mutations in the succinate dehydrogenase (SDH; succinate:ubiquinone oxidoreductase) subunit genes (*SDHB*, *SDHC*, *SDHD*, *SDHAF2*, *SDHA*), which compose portions of mitochondrial complex II. Inactivating germline mutations in *SDHD* have been identified in multigenerational families with head and neck parasympathetic paragangliomas that are usually nonfunctional and occasionally in those with adrenal pheochromocytoma. In patients with *SDHD* mutations, penetrance depends on the mutation's parent of origin. Hence, the disease does not manifest when the mutation is inherited from the mother but is highly penetrant when it is inherited from the father. This phenomenon is known as maternal imprinting. Multiple cofactors are required for normal activity of the SDH complex, including flavin adenine dinucleotide (FAD) in the SDH1subunit. FAD is covalently attached to Sdh1, and deletion of *SDHAF2* causes a complete loss of FAD cofactor attachment (flavination) of Sdh1. Germline loss-of-function mutations in the *SDHAF2* gene, located on chromosome 11q13.1, have been associated with disease in a family with hereditary paraganglioma. Like families with mutations in *SDHD*, those with mutations in *SDHAF2* also exhibit maternal imprinting and parasympathetic paragangliomas that typically occur in the skull base and neck. Inactivating mutations in the tumor suppressor gene *SDHB*, located on chromosome 1p35-36, are associated with paragangliomas in the abdomen, pelvis, and mediastinum. Adrenal pheochromocytomas may also be found in patients with *SDHB* mutations. Patients with *SDHB* mutations have an increased risk for malignant paraganglioma.

Genetic Testing

Since 1990, 15 different pheochromocytoma and paraganglioma susceptibility genes have been reported: *NF1*, *RET*, *VHL*, *SDHD*, *SDHC*, *SDHB*, *EGLN1/PHD2*, *KIF1B*, *SDHAF2*, *IDH1*, *TMEM127*, *SDHA*, *MAX*, *HIF2A*, and *FH* gene encoding fumarate hydratase.[4,5]

Genetic testing should be considered if a patient has one or more of the following: paraganglioma, bilateral adrenal pheochromocytomas, unilateral adrenal pheochromocytoma and a family history of pheochromocytoma or paraganglioma, onset of unilateral adrenal pheochromocytoma at a young age (before 45 years), or other clinical findings suggestive of one of the previously discussed syndromic disorders.[6] Genetic testing can be complex; testing of one family member has implications for related individuals.[7] Genetic counseling is recommended to help families understand the implications of genetic test results; to coordinate the testing of at-risk individuals; and to help families work through the psychosocial issues that may arise before, during, or after the testing process. A list of clinically approved molecular genetic diagnostic laboratories is available at www.genetests.org.

CLINICAL MANIFESTATIONS

Catecholamine-secreting tumors occur with equal frequency in men and women, primarily in the third, fourth, and fifth decades of life. These tumors are rare in children; when discovered, they may be multifocal and associated with a hereditary syndrome. When symptoms are present, they are due to the pharmacologic effects of excess concentrations of circulating catecholamines (Table 228-1). The resulting hypertension may be sustained (in approximately half of patients) or paroxysmal (in approximately a third of patients). The remaining patients have normal blood pressure. Episodic symptoms may occur in spells, or paroxysms, that can be extremely variable in presentation but typically include forceful heartbeat, pallor, tremor, headache, and diaphoresis.[8] The spell may start with the sensation of a "rush" in the chest and a sense of shortness of breath, followed by a "pounding" heartbeat in the chest that typically progresses to a throbbing headache. Peripheral vasoconstriction with a spell results in cool or cold hands and feet and facial pallor. Increased sense of body heat and sweating are common symptoms that occur toward the end of the spell. Spells may be spontaneous or precipitated by postural change, anxiety, medications (e.g., metoclopramide, β-adrenergic inhibitors, anesthetic agents), exercise, or maneuvers that increase intra-abdominal pressure (e.g., change in position, lifting, defecation, exercise, colonoscopy, pregnancy, trauma). Although the types of spells experienced by patients are highly variable, spells tend to be stereotypical for each patient. Spells may occur multiple times a day or as infrequently as once a month. The typical duration of a pheochromocytoma spell is 15 to 20 minutes, but it may be much shorter or last several hours. The clinician must recognize that most patients with spells do not have a pheochromocytoma.

Additional clinical signs of catecholamine-secreting tumors include hypertensive retinopathy, orthostatic hypotension, angina, nausea, constipation (megacolon may be the presenting symptom), hyperglycemia, diabetes mellitus, hypercalcemia, Raynaud's phenomenon, livedo reticularis, erythrocytosis, and mass effects from the tumor. Fasting hyperglycemia and diabetes mellitus are caused in part by the α-adrenergic inhibition of insulin release. Painless hematuria and paroxysmal attacks induced by micturition and defecation are associated with urinary bladder paragangliomas. Some of the co-secreted hormones that may dominate the clinical presentation include corticotropin (Cushing's syndrome), parathyroid hormone–related peptide (hypercalcemia), vasopressin (syndrome of inappropriate antidiuretic hormone secretion), vasoactive intestinal peptide (watery diarrhea), and growth hormone–releasing hormone (acromegaly) (see Table 228-1). Cardiomyopathy and congestive heart failure are the symptomatic presentations of pheochromocytoma that are most frequently unrecognized by clinicians. Cardiomyopathy, whether dilated or hypertrophic, may be totally reversible with tumor resection. Some patients with pheochromocytoma may be asymptomatic despite high circulating levels of catecholamines, probably reflecting adrenergic receptor desensitization related to chronic stimulation.

Symptomatic pheochromocytomas are localized to the adrenal glands, with an average diameter of 4.5 cm (Fig. 228-2). Paragangliomas are found where there is chromaffin tissue: along the para-aortic sympathetic chain, within the organs of Zuckerkandl (at the origin of the inferior mesenteric artery), in the wall of the urinary bladder, and along the sympathetic chain in the neck or mediastinum. Paragangliomas in the head and neck region (e.g., carotid body tumors, glomus tumors, chemodectomas) usually arise from parasympathetic tissue and typically do not hypersecrete catecholamines and metanephrines, whereas paragangliomas in the mediastinum, abdomen, and pelvis usually arise from sympathetic chromaffin tissue and usually do hypersecrete catecholamines and metanephrines.

TABLE 228-1 SIGNS AND SYMPTOMS ASSOCIATED WITH CATECHOLAMINE-SECRETING TUMORS

SPELL RELATED
Anxiety and fear of impending death
Diaphoresis
Dyspnea
Epigastric and chest pain
Headache
Hypertension
Nausea and vomiting
Pallor
Palpitation (forceful heartbeat)
Tremor
CHRONIC
Anxiety and fear of impending death
Cold hands and feet
Congestive heart failure—dilated or hypertrophic cardiomyopathy
Constipation
Diaphoresis
Dyspnea
Ectopic hormone secretion–dependent symptoms (e.g., CRH/ACTH, GHRH, PTH-RP, VIP)
Epigastric and chest pain
Fatigue
Fever
General increase in sweating
Grade II to IV retinopathy
Headache
Hyperglycemia
Hypertension
Nausea and vomiting
Orthostatic hypotension
Painless hematuria (associated with urinary bladder paraganglioma)
Pallor
Palpitation (forceful heartbeat)
Tremor
Weight loss
NOT TYPICAL OF PHEOCHROMOCYTOMA
Flushing

ACTH = adrenocorticotropic hormone; CRH = corticotropin-releasing hormone; GHRH = growth hormone–releasing hormone; PTH-RP = parathyroid hormone–related peptide; VIP = vasoactive intestinal polypeptide.
Modified from Young WF Jr. Pheochromocytoma: 1926-1993. *Trends Endocrinol Metab.* 1993;4:122-127.

DIAGNOSIS

Differential Diagnosis

Numerous disorders can cause signs and symptoms that may lead the clinician to test for pheochromocytoma. These disorders span much of medicine and include endocrine disorders (e.g., primary hypogonadism), cardiovascular disorders (e.g., idiopathic orthostatic hypotension), psychological

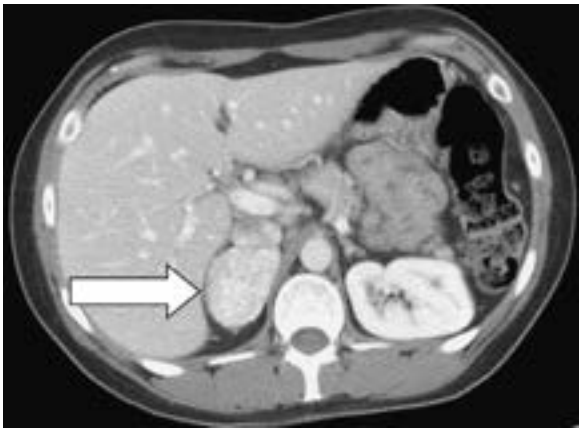

FIGURE 228-2. Contrast-enhanced computed tomography of the abdomen in a 32-year-old second-year medical student with the peripartum discovery of a pheochromocytoma. The plasma fractionated metanephrines were abnormal: metanephrine, 0.19 nmol/L (normal, <0.5 nmol/L); normetanephrine, 28.6 nmol/L (normal, <0.9 nmol/L). The 24-hour urine values were abnormal: norepinephrine, 781 µg (normal, <170 µg); epinephrine, 2.4 µg (normal, <35 µg); dopamine, 197 µg (normal, <700 µg); metanephrine, 117 µg (normal, <400 µg); normetanephrine, 8760 µg (normal, <900 µg). The axial image shows a typical 5-cm heterogeneously enhancing right adrenal mass, consistent with pheochromocytoma (*arrow*). After α- and β-adrenergic blockade, a 5.3 × 5.0 × 2.0-cm, 40-g pheochromocytoma was removed laparoscopically.

disorders (e.g., panic disorder), pharmacologic causes (e.g., withdrawal from an adrenergic inhibitor), neurologic disorders (e.g., postural orthostatic tachycardia syndrome), and miscellaneous disorders (e.g., mast cell disease). Indeed, most patients tested for pheochromocytoma do not have it. In addition, fractionated catecholamines and metanephrines may be elevated in several clinical scenarios: withdrawal from medications or drugs (e.g., clonidine, alcohol), any acute illness (e.g., subarachnoid hemorrhage, migraine headache, preeclampsia), and administration of many drugs and medications (e.g., tricyclic antidepressants, levodopa, cocaine, phencyclidine, lysergic acid diethylamide, amphetamines, ephedrine, pseudoephedrine, phenylpropanolamine, isoproterenol) (Table 228-2).

Pheochromocytoma should be suspected in patients who have one or more of the following: hyperadrenergic spells (e.g., self-limited episodes of nonexertional palpitations, diaphoresis, headache, tremor, or pallor); resistant hypertension; a familial syndrome that predisposes to catecholamine-secreting tumors (e.g., MEN type 2, NF1, VHL disease); a family history of pheochromocytoma or a history of a resected pheochromocytoma and a present history of recurrent hypertension or spells; an incidentally discovered adrenal mass; hypertension and diabetes; pressor response during anesthesia, surgery, or angiography; onset of hypertension at a young age (before 20 years); and idiopathic dilated cardiomyopathy.

Laboratory Findings

The diagnosis must be confirmed biochemically by increased concentrations of fractionated metanephrines in the plasma or fractionated catecholamines and metanephrines in a 24-hour urine collection (Fig. 228-3).[9] Most laboratories now measure fractionated catecholamines (dopamine, norepinephrine, and epinephrine) and metanephrines (metanephrine and normetanephrine) by high-performance liquid chromatography with electrochemical detection or tandem mass spectroscopy.[10] These techniques have overcome the problems with fluorometric analysis (e.g., false-positive results caused by α-methyldopa, labetalol, sotalol, and imaging contrast agents). One of the most reliable methods of identifying catecholamine-secreting tumors is measurement of fractionated metanephrines and catecholamines in a 24-hour urine collection (sensitivity, 98%; specificity, 98%).[11] If clinical suspicion is high, plasma fractionated metanephrines should also be measured. Some groups have advocated the measurement of plasma fractionated metanephrines as a first-line test for pheochromocytoma; the predictive value of a

TABLE 228-2 MEDICATIONS THAT MAY INCREASE MEASURED LEVELS OF CATECHOLAMINES AND METANEPHRINES

Tricyclic antidepressants (including cyclobenzaprine)
Levodopa
Drugs containing adrenergic receptor agonists (e.g., decongestants)
Amphetamines
Buspirone and antipsychotic agents
Prochlorperazine
Reserpine
Withdrawal from clonidine and other drugs
Ethanol

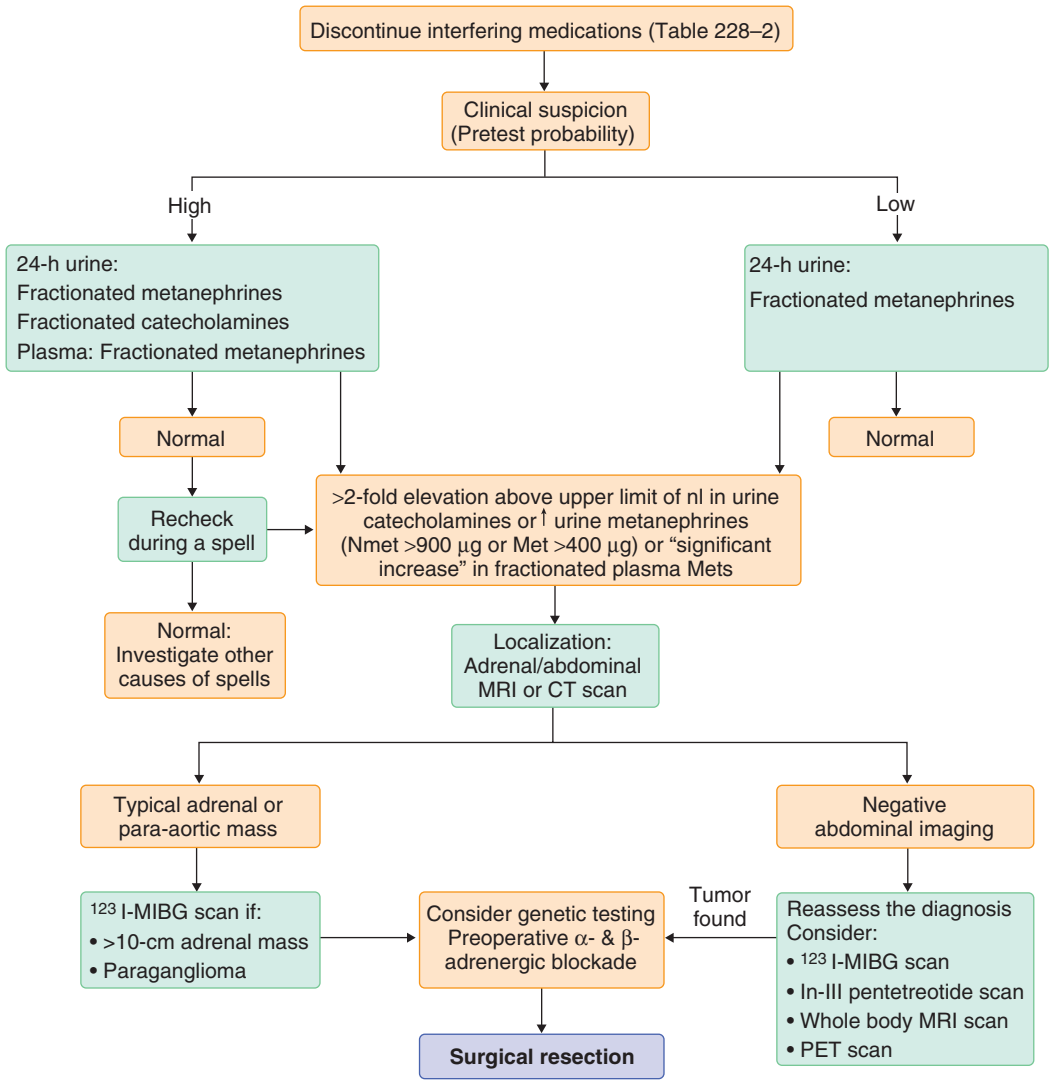

FIGURE 228-3. Evaluation and treatment of catecholamine-secreting tumors. Clinical suspicion is triggered by the following: paroxysmal symptoms (especially hypertension); hypertension that is intermittent, unusually labile, or resistant to treatment; family history of pheochromocytoma or associated conditions; or incidentally discovered adrenal mass. The details are discussed in the text. CT = computed tomography; ^{123}I-MIBG = ^{123}I-metaiodobenzylguanidine; Met = metanephrine; MRI = magnetic resonance imaging; nl = normal; Nmet = normetanephrine; PET = positron emission tomography. (Modified from Young WF Jr. Pheochromocytoma: 1926-1993. *Trends Endocrinol Metab.* 1993;4:122.)

negative test result is extremely high, and the finding of normal plasma fractionated metanephrines excludes pheochromocytoma, except in patients with early preclinical disease and those with strictly dopamine-secreting neoplasms.[12] A plasma test is also attractive because of its simplicity. Although measurement of plasma fractionated metanephrines has a sensitivity of 96 to 100%, the specificity is poor at only 85 to 89%; the specificity falls to 77% in patients older than 60 years. It has been estimated that 97% of patients with hypertension seen in a tertiary care clinic who have abnormal plasma fractionated metanephrine measurements do not have pheochromocytoma. This high false-positive rate results in excessive health care expenditures because of subsequent imaging as well as potentially inappropriate surgery. Thus, plasma fractionated metanephrines lack the necessary specificity to be recommended as a first-line test; this measurement is reserved for cases in which the index of suspicion is high. However, measurement of plasma fractionated metanephrines is a good first-line test for children, in whom it is difficult to obtain a complete 24-hour urine collection.

The index of suspicion for pheochromocytoma should be high (see Fig. 228-3) in patients with the clinical scenarios described earlier and in those with an incidentally discovered adrenal mass that has imaging characteristics consistent with pheochromocytoma. These include increased baseline computed tomography (CT) Hounsfield unit density (e.g., >20 HU), marked enhancement with intravenous contrast medium on CT, high signal intensity on T2-weighted magnetic resonance imaging (MRI), cystic and hemorrhagic changes, bilaterality, and large size (>4 cm) (see later).

Although it is preferable for patients not to receive any medications during the diagnostic evaluation, treatment with most medications can be continued. Tricyclic antidepressants interfere most frequently with the interpretation of fractionated catecholamines and metanephrines. For the effective detection of catecholamine-secreting tumors, treatment with tricyclic antidepressants and other psychoactive agents listed in Table 228-2 should be tapered and discontinued at least 2 weeks before any hormonal assessments. Furthermore, catecholamine secretion may be appropriately increased in situations of physical stress or illness (e.g., stroke, myocardial infarction, congestive heart failure, obstructive sleep apnea). Therefore, the clinical circumstances under which catecholamines and metanephrines are measured must be assessed in each case.

Imaging

Localization studies should not be initiated until biochemical studies have confirmed the diagnosis of a catecholamine-secreting tumor (see Fig. 228-3). Computer-assisted imaging of the adrenal glands and abdomen with CT or MRI should be the first localization test (sensitivity, >95%; specificity, >65%). Approximately 85% of these tumors are found in the adrenal glands, and 95% are found in the abdomen. If the results of abdominal imaging are normal, scintigraphic localization with ^{123}I-labeled metaiodobenzylguanidine (^{123}I-MIBG) is indicated (see Fig. 228-3). This radiopharmaceutical agent accumulates preferentially in catecholamine-producing tumors (sensitivity, 88%; specificity, 94%). If a typical (<10 cm) unilateral adrenal pheochromocytoma is found on CT or MRI, ^{123}I-MIBG scintigraphy is superfluous, and the results may even cause confusion. If the adrenal pheochromocytoma is larger than 10 cm in diameter or if a paraganglioma is identified on CT or MRI, ^{123}I-MIBG scintigraphy is indicated because the patient has an increased risk of malignant disease and additional paragangliomas.^{18}F-Fluorodeoxyglucose positron emission tomography is an excellent imaging modality to detect metastatic disease.[13,14]

Other localizing procedures that can be used but are rarely required include computer-assisted imaging of the chest and neck as well as somatostatin receptor imaging with [^{111}In]pentetreotide (see Fig. 228-3). Because of marked gradients between the adrenal glands in non-pheochromocytoma patients, adrenal venous sampling for catecholamines is not helpful in the investigation of adrenal pheochromocytoma.

TREATMENT

Medical Therapy

Some form of preoperative pharmacologic preparation is indicated for all patients with catecholamine-secreting neoplasms. However, no randomized controlled trials have compared the different approaches. Combined α- and β-adrenergic blockade is one approach to control blood pressure and to prevent intraoperative hypertensive crises. α-Adrenergic blockade should be started 7 to 10 days preoperatively to normalize blood pressure and to expand the contracted blood volume. A longer duration of preoperative α-adrenergic blockade is indicated in patients with recent myocardial infarction, catecholamine cardiomyopathy, and catecholamine-induced vasculitis. Blood pressure should be monitored with the patient in the seated and standing positions twice daily. Target blood pressure is less than 120/80 mm Hg (seated), with systolic blood pressure greater than 90 mm Hg (standing); both targets should be modified on the basis of the patient's age and comorbid disease. On the second or third day of α-adrenergic blockade, patients are encouraged to start a diet high in sodium content (≥5000 mg daily) because of the catecholamine-induced volume contraction and the orthostasis associated with α-adrenergic blockade. This degree of volume expansion may be contraindicated in patients with congestive heart failure or renal insufficiency. After adequate α-adrenergic blockade has been achieved, β-adrenergic blockade is initiated, which typically occurs 2 or 3 days preoperatively.

Phenoxybenzamine is the preferred drug to control blood pressure and arrhythmia preoperatively.[15] It is an irreversible, long-acting, nonspecific α-adrenergic blocking agent. The initial dosage is 10 mg once or twice daily, and the dose is increased by 10 to 20 mg in divided doses every 2 or 3 days as needed to control blood pressure and spells. The final dosage of phenoxybenzamine is typically between 20 and 100 mg daily.

The β-adrenergic antagonist should be administered only after α-adrenergic blockade is effective; with β-adrenergic blockade alone, hypertension may be more severe from the unopposed α-adrenergic stimulation. Preoperative β-adrenergic blockade is indicated to control the tachycardia associated with both the high concentrations of circulating catecholamines and the α-adrenergic blockade. The clinician should exercise caution if the patient is asthmatic or has congestive heart failure. Chronic catecholamine excess can produce a cardiomyopathy that may become evident with the initiation of β-adrenergic blockade, resulting in acute pulmonary edema. Therefore, the β-adrenergic blocker should be administered cautiously and at a low dose. Other agents that may be used to prepare the patient with pheochromocytoma for surgery include α-methyl-*p*-tyrosine (metyrosine) and calcium-channel blockers. Acute hypertensive crises may occur before or during an operation, and they should be treated with intravenous sodium nitroprusside, phentolamine, or nicardipine.

Surgical Therapy

The treatment of choice for pheochromocytoma is complete surgical resection. Surgical survival rates are 98 to 100% and are highly dependent on the skill of the endocrinologist–endocrine surgeon–anesthesiologist team. Careful preoperative pharmacologic preparation is crucial for successful treatment. Most catecholamine-secreting tumors are benign and can be totally excised. Tumor excision usually cures hypertension.

In the past, an anterior midline abdominal surgical approach was generally used to resect adrenal pheochromocytoma. However, the laparoscopic approach to the adrenal gland is currently the procedure of choice for a solitary intra-adrenal pheochromocytoma smaller than 8 cm in diameter.[16] Laparoscopic adrenalectomy for pheochromocytoma should be converted to open adrenalectomy in cases of difficult dissection, invasion, adhesions, or an inexperienced surgeon. An anterior midline abdominal surgical approach is indicated for abdominal paragangliomas. The midline abdomen should be inspected carefully. Paragangliomas of the neck, chest, and urinary bladder require specialized approaches.

Management of Complications

Hypotension may occur during and after surgical resection of the pheochromocytoma, and it should be treated with fluids and colloids and then intravenous pressor agents if necessary. Postoperative hypotension is less frequent in patients who have had adequate preoperative α-adrenergic blockade and volume expansion. If both adrenal glands were manipulated during surgery, adrenocortical insufficiency should be considered a potential cause of postoperative hypotension. Because hypoglycemia can occur in the immediate postoperative period, blood glucose levels should be monitored, and intravenous fluids should contain 5% dextrose.

Approximately 1 to 2 weeks after surgery, fractionated catecholamines and metanephrines should be measured by collection of a 24-hour urine specimen. If the levels are normal, resection of the pheochromocytoma should be considered complete. Increased levels of fractionated catecholamines and metanephrines detected postoperatively are consistent with residual tumor due to either a second primary lesion or occult metastases. Common sites of metastasis include lymph nodes, liver, lung, and bone.

Follow-up

The 24-hour urinary excretion of fractionated catecholamines and metanephrines or plasma fractionated metanephrines should be checked annually for life. Annual biochemical testing assesses for metastatic disease, tumor recurrence in the adrenal bed, or delayed appearance of multiple primary tumors. Follow-up CT or MRI is not needed unless the metanephrine or catecholamine levels become elevated or the original tumor was associated with minimal or no catecholamine or metanephrine excess.

MALIGNANT PHEOCHROMOCYTOMA AND PARAGANGLIOMA

Distinguishing between benign and malignant catecholamine-secreting tumors is difficult on the basis of clinical, biochemical, or histopathologic characteristics. Malignant disease is rare in patients with an adrenal familial syndrome but common in those with familial paraganglioma caused by mutations in *SDHB*. Although the 5-year survival rate for patients with malignant pheochromocytoma is less than 50%, the prognosis is variable[17]; approximately 50% of patients have an indolent form of the disease, with a life expectancy of more than 20 years, and the other 50% have rapidly progressive disease, with death occurring 1 to 5 years after diagnosis. The clinician should assess the pace of the malignant disease and base the level of therapy on the aggressiveness of the tumor's behavior. A multimodality, multidisciplinary, individualized approach is indicated to control catecholamine-dependent symptoms, local mass effects, and overall tumor burden.

PHEOCHROMOCYTOMA IN PREGNANCY

Pheochromocytoma in pregnancy can cause the death of both the fetus and the mother. The approach to the biochemical diagnosis is the same as for nonpregnant patients. MRI without gadolinium enhancement is the preferred imaging modality.[123]I-MIBG scintigraphy is contraindicated. The treatment of hypertensive crises is the same as for nonpregnant patients, except that nitroprusside should be avoided. Although the most appropriate management is debated, adrenal pheochromocytomas should be removed promptly after α- and β-adrenergic blockade if the diagnosis is made during the first two trimesters of pregnancy.[18] The preoperative preparation is the same as for nonpregnant patients. If the pregnancy is in the third trimester, one operation is recommended for cesarean delivery and removal of the adrenal pheochromocytoma at the same time. Spontaneous labor and delivery should be avoided. The management of catecholamine-secreting paragangliomas in pregnancy may require modification of these guidelines, depending on tumor location.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bausch B, Wellner U, Bausch D, et al. Long-term prognosis of patients with pediatric pheochromocytoma. *Endocr Relat Cancer*. 2013;21:17-25.
2. Thosani S, Ayala-Ramirez M, Palmer L, et al. The characterization of pheochromocytoma and its impact on overall survival in multiple endocrine neoplasia type 2. *J Clin Endocrinol Metab*. 2013;98:E1813-E1819.
3. Binderup ML, Bisgaard ML, Harbud V, et al. Von Hippel–Lindau disease (vHL). National clinical guideline for diagnosis and surveillance in Denmark. 3rd edition. *Dan Med J*. 2013;60:B4763.
4. Burnichon N, Cascón A, Schiavi F, et al. *MAX* mutations cause hereditary and sporadic pheochromocytoma and paraganglioma. *Clin Cancer Res*. 2012;18:2828-2837.
5. Castro-Vega LJ, Buffet A, De Cubas AA, et al. Germline mutations in FH confer predisposition to malignant pheochromocytomas and paragangliomas. *Hum Mol Genet*. 2014;23:2440-2446.
6. Rattenberry E, Vialard L, Yeung A, et al. A comprehensive next generation sequencing-based genetic testing strategy to improve diagnosis of inherited pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab*. 2013;98:E1248-E1256.
7. Brito JP, Asi N, Bancos I, et al. Testing for germline mutations in sporadic pheochromocytoma/paraganglioma: a systematic review. *Clin Endocrinol (Oxf)*. 2015;82:338-345.
8. Shah NH, Ruan DT. Pheochromocytoma: a devious opponent in a game of hide-and-seek. *Circulation*. 2014;130:1295-1298.
9. Lenders JW, Duh QY, Eisenhofer G, et al. Endocrine Society. Pheochromocytoma and paraganglioma: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2014;99:1915-1942.
10. van Berkel A, Lenders JW, Timmers HJ. Diagnosis of endocrine disease: biochemical diagnosis of phaeochromocytoma and paraganglioma. *Eur J Endocrinol*. 2014;170:R109-R119.
11. Eisenhofer G, Pacak K, Maher ER, et al. Pheochromocytoma. *Clin Chem*. 2013;59:466-472.
12. Därr R, Pamporaki C, Peitzsch M, et al. Biochemical diagnosis of phaeochromocytoma using plasma-free normetanephrine, metanephrine and methoxytyramine: importance of supine sampling under fasting conditions. *Clin Endocrinol (Oxf)*. 2014;80:478-486.
13. Rufini V, Treglia G, Castaldi P, et al. Comparison of metaiodobenzylguanidine scintigraphy with positron emission tomography in the diagnostic work-up of pheochromocytoma and paraganglioma: a systematic review. *Q J Nucl Med Mol Imaging*. 2013;57:122-133.
14. Timmers HJLM, Chen CC, Carrasquillo JA, et al. Staging and functional characterization of pheochromocytoma and paraganglioma by ^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography. *J Natl Cancer Inst*. 2012;104:700-708.
15. Agrawal R, Mishra SK, Bhatia E, et al. Prospective study to compare peri-operative hemodynamic alterations following preparation for pheochromocytoma surgery by phenoxybenzamine or prazosin. *World J Surg*. 2014;38:716-723.
16. Scholten A, Cisco RM, Vriens MR, et al. Variant adrenal venous anatomy in 546 laparoscopic adrenalectomies. *JAMA Surg*. 2013;148:378-383.
17. Hescot S, Leboulleux S, Amar L, et al. One-year progression-free survival of therapy-naive patients with malignant pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab*. 2013;98:4006-4012.
18. Dong D, Li H. Diagnosis and treatment of pheochromocytoma during pregnancy. *J Matern Fetal Neonatal Med*. 2014;27:1930-1934.

REVIEW QUESTIONS

1. An effective class of antihypertensive drugs could be based on which of the following?

A. Postsynaptic α_1-receptor agonists
B. Presynaptic α_2-receptor agonists
C. Tyrosine hydroxylase activators
D. β_1-Receptor agonists
E. β_2-Receptor agonists

Answer: B See introduction section of the chapter on catecholamines and adrenergic receptors. Agonism at the presynaptic α_2-receptor is the only option listed that would decrease catecholamine synthesis or release. Central α_2-receptor agonists are Food and Drug Administration–approved antihypertensive drugs (e.g., clonidine, guanfacine).

2. Your patient has a disease-causing mutation in succinate dehydrogenase subunit D (*SDHD*) and is currently pregnant. She asks you if her child inherits the *SDHD* mutation from her, what is the chance that her child will develop a paraganglioma in his or her lifetime? You advise her that if her child inherits the *SDHD* mutation, the risk for development of a paraganglioma is closest to

A. 0%
B. 25%
C. 50%
D. 75%
E. 100%

Answer: A See section on familial paraganglioma under Pathobiology and Genetics. In patients with *SDHD* mutations, penetrance depends on the mutation's parent of origin. Hence, the disease does not manifest when the mutation is inherited from the mother but is highly penetrant when it is inherited from the father. This phenomenon is known as maternal imprinting.

3. In a patient with an incidentally discovered adrenal mass, which of the following imaging characteristics is consistent with pheochromocytoma?

A. Lack of enhancement with intravenous contrast medium on computed tomography (CT)
B. Low signal intensity on T2-weighted magnetic resonance imaging (MRI)
C. Homogeneous mass with absent cystic and hemorrhagic changes
D. Precontrast CT Hounsfield unit density (e.g., >20 HU)
E. Rapid contrast washout with more than 50% washout at 10 minutes

Answer: D See section on diagnosis. Imaging characteristics of an incidentally discovered adrenal mass consistent with pheochromocytoma include increased baseline CT Hounsfield unit density (e.g., >20 HU); marked enhancement with intravenous contrast medium on CT; high signal intensity on T2-weighted MRI; cystic and hemorrhagic changes; bilaterality; and large size (>4 cm). In addition, pheochromocytomas are characterized by slow contrast washout (e.g., <50% at 10 minutes after administration of contrast material).

4. Medications and agents that frequently increase measured levels of catecholamines and metanephrines include

A. β-Adrenergic inhibitors
B. Presynaptic α_2-receptor agonists
C. Tyrosine hydroxylase inhibitors
D. Tricyclic antidepressants
E. Calcium-channel blockers

Answer: D See Table 228-2. In addition, β-adrenergic inhibitors do not affect circulating levels of catecholamines or metanephrines in a clinically important way. Central α-receptor agonists (e.g., clonidine) decrease the release of catecholamines. Tyrosine hydroxylase is the rate-limiting step in catecholamine synthesis, and inhibition of this step decreases total body catecholamine production. Use of tricyclic antidepressants is the most common cause of false-positive results of biochemical testing for pheochromocytoma.

5. In preparing the patient with pheochromocytoma for surgery, which of the following statements is correct with regard to β-adrenergic blockade?

A. The β-adrenergic antagonist should be administered before the α-adrenergic blocker is started.
B. It is indicated to control the tachycardia associated with both the high concentrations of circulating catecholamines and the α-adrenergic blockade.
C. The dosing of the β-adrenergic antagonist is not affected by asthma or congestive heart failure in this unique setting of catecholamine excess.
D. The dose of the β-adrenergic antagonist should be titrated for a heart rate of 50 beats per minute.
E. The dose of the β-adrenergic antagonist should be titrated for a systolic blood pressure of 100 mm Hg.

Answer: B See medical therapy section on treatment. The β-adrenergic antagonist should be administered only after α-adrenergic blockade is effective; with β-adrenergic blockade alone, hypertension may be more severe from the unopposed α-adrenergic stimulation. Preoperative β-adrenergic blockade is indicated to control the tachycardia associated with both the high concentrations of circulating catecholamines and the α-adrenergic blockade. The clinician should exercise caution if the patient is asthmatic or has congestive heart failure. Chronic catecholamine excess can produce a cardiomyopathy that may become evident with the initiation of β-adrenergic blockade, resulting in acute pulmonary edema. Therefore, the β-adrenergic blocker should be administered cautiously and at a low dose.

229

DIABETES MELLITUS

JILL CRANDALL AND HARRY SHAMOON

Diabetes mellitus is a chronic disorder characterized by abnormal metabolic regulation as well as by the potential for vascular and neuropathic complications. Diabetes comprises a cluster of heterogeneous disorders with elevated blood glucose levels as a common diagnostic feature; however, as genetic and molecular studies have suggested, it is likely that the cluster includes many subcategories, each of which requires tailored prevention, diagnosis, and treatment approaches. Depending on the context in which the patient presents, diabetes can be an acute life-threatening condition, a pregnancy-associated disorder, or a gradually evolving chronic disorder that carries with it secondary complications that may be ultimately more debilitating than hyperglycemia. Other factors make diabetes an unusual clinical challenge, including the need for active participation by patients in their treatment, the varying presentations across the age spectrum, and the unstable and evolving clinical presentation. Because the severity of the underlying metabolic defects does not remain static, diabetes management always requires changes in treatment according to the stage of the disease. These patterns of evolution are superimposed on the phenotypes at presentation and depend on a host of factors including age, sex, race, societal setting, and others.

It is now established that diabetes-related vascular and neuropathic complications stem from imperfect treatment of the metabolic disturbances, defined principally by hyperglycemia. There is also evidence that genetic factors may predispose or protect individual patients from the deleterious effects of hyperglycemia. Regardless of the specific subtype of diabetes, all have in common some degree of insulin deficiency; insulin deficiency may be absolute, as in type 1 diabetes, or a relative deficit with coexisting insulin resistance, as in type 2 diabetes. Deficient insulin is the primary driver of impaired fuel homeostasis, whereas hyperglycemia plays the dominant role in disease-related complications. Major strides in our understanding of diabetes have been made during the last 40 years, with accompanying additions to the diagnostic and treatment armamentarium.

TABLE 229-1 CLASSIFICATION OF DIABETES

	TYPE 1	TYPE 2
Age at onset	Childhood or early adulthood, but can be manifested at any age	Middle age or older, but can be manifested in obese children and adolescents
Family history/genetic factors	Genetic risk defined, but most cases are sporadic	Strong genetic component, polygenic in most cases
Environmental triggers	Largely unknown	Obesity, sedentary lifestyle
Requirement for insulin therapy	Universal	Variable
Frequency among people with diabetes	5-10%	~90%
Associated disorders	Autoimmunity, especially thyroid, other endocrine disorders	Hypertension, dyslipidemia, metabolic syndrome, polycystic ovary syndrome

DEFINITIONS

Despite the heterogeneity of phenotypes, it is possible to generally classify diabetes into two major subgroups, type 1 (previously referred to as juvenile-onset or insulin-dependent diabetes) and type 2 (previously referred to as adult-onset or non–insulin-dependent diabetes). The major clinical features of type 1 and type 2 are shown in Table 229-1 and are described in detail in the corresponding sections later.

In addition to these two large categories, diabetes may occur in association with other disorders, with use of certain medications, or, rarely, as a result of a specific genetic mutation, such as maturity-onset diabetes of youth (MODY).

Diabetes Associated with Other Disorders or Syndromes

Diabetes may occur as part of several inherited syndromes, including the Turner, Klinefelter, Prader-Willi, Down, and Wolfram syndromes, among others. The genetic and metabolic defects involved are heterogeneous but usually result in impaired β-cell function. The obesity (and resulting insulin resistance) associated with many of these syndromes also contributes. Diseases of the exocrine pancreas, such as pancreatitis, pancreatic cancer, hemochromatosis, and cystic fibrosis, can be accompanied by impaired pancreatic endocrine function, leading to insulin-deficient diabetes. Several endocrinopathies that are associated with insulin resistance, including acromegaly, Cushing syndrome, and pheochromocytoma, may result in impaired glucose tolerance or frank diabetes in predisposed individuals. Viral infections, such as congenital rubella and cytomegalovirus, may cause diabetes by β-cell destruction. Finally, hyperglycemia may be associated with the use of certain drugs, including those that worsen insulin resistance (e.g., glucocorticoids, nicotinic acid, thiazide diuretics) and those that impair β-cell function (e.g., pentamidine, diazoxide, interferon gamma).

Diagnostic Criteria for Diabetes

Diabetes is diagnosed on the basis of one of several criteria, including fasting plasma glucose concentration, plasma glucose concentration after a standard 75-g oral glucose challenge (oral glucose tolerance test), and percentage of glycosylated hemoglobin (HbA_{1c}) (Table 229-2). In most cases, abnormal results require a confirmatory test, but diabetes can be diagnosed in the presence of unequivocal hyperglycemia (casual plasma glucose concentration > 200 mg/dL) and typical symptoms of polyuria, polydipsia, and weight loss.

Because plasma glucose levels exist on a continuum, the selection of a specific diagnostic threshold is in some respects arbitrary. Current criteria are based on the plasma glucose or HbA_{1c} level above which the risk of diabetes-specific microvascular complications (e.g., retinopathy) is perceptibly

TABLE 229-2 DIAGNOSTIC CRITERIA FOR DIABETES

	NORMAL	IMPAIRED (PRE-DIABETES)	DIABETES
Fasting glucose concentration (mg/dL)	<100	100-125	≥126
OGTT 2-hour glucose concentration (mg/dL)	<140	140-199	≥200
HbA_{1c} (%)	<5.7	5.7-6.4	≥6.5

OGTT = oral glucose tolerance test.
Modified from American Diabetes Association. Standards of medical care in diabetes—2014. *Diabetes Care.* 2015;38(Suppl 1):S8-S16.

increased. In situations of altered red blood cell turnover or certain hemoglobinopathies, HbA_{1c} may not accurately reflect mean plasma glucose levels (see later section on glycosylated hemoglobin), and direct glucose measurement should be used. Separate glucose criteria exist for the diagnosis of gestational diabetes (see section on gestational diabetes under clinical manifestations of type 2 diabetes).

States of impaired glucose regulation, not meeting the criteria for diabetes, have also been defined (fasting glucose concentration of 100 to 125 mg/dL, 2-hour glucose concentration of 140 to 199 mg/dL, or HbA_{1c} level of 5.7 to 6.4%). Individuals in these categories are at increased risk for diabetes, although not all will progress and some may revert to normal glucose regulation. Impaired glucose tolerance (oral glucose tolerance test 2-hour glucose concentration of 140 to 199 mg/dL) has also been associated with increased risk of atherosclerotic cardiovascular disease (CVD), which may be independent of future development of diabetes.

Glycosylated Hemoglobin

Measurements of glycosylated hemoglobin have been in clinical use since the 1980s as a means of assessing glucose control in patients with diabetes and more recently for the diagnosis of diabetes and pre-diabetic states. Hemoglobin A_{1c} (HbA_{1c}) is formed by the nonenzymatic glycosylation of hemoglobin, and its percentage reflects the exposure of the hemoglobin A molecule to glucose during the lifespan of red blood cells (~120 days). Thus, HbA_{1c} has a predictable (but nonlinear) relationship with mean plasma glucose levels during the preceding 3 to 4 months, although more recent exposure (preceding 4 weeks) contributes relatively more to the percentage of glycosylation. The relationship between HbA_{1c} and mean glucose levels was initially based on data obtained from the Diabetes Control and Complications Trial (DCCT) and recently updated on the basis of data obtained from studies using continuous glucose monitoring in ambulatory individuals, including those with and without diabetes (Table 229-3).

Although several different types of assays (e.g., affinity chromatography, immunoassay) are used to measure HbA_{1c}, most methods have been harmonized to a common standard and generally allow results from different laboratories to be used interchangeably. HbA_{1c} results may be influenced by a number of factors, including conditions that alter red cell survival (e.g., hemolytic anemia) or cause interference with a specific assay. In these situations, measurement of fructosamine (glycosylated serum proteins) or glycated albumin, both of which reflect mean glucose levels during the preceding 2 to 3 weeks, may provide more accurate assessment of recent glucose levels. However, these assays have not been as well standardized, and the relationship with mean plasma glucose levels is less well established (Table 229-4).

PATHOBIOLOGY OF DIABETES

Figure 229-1 summarizes the effects of insulin deficiency on body fuel metabolism.

Given the dominant role of insulin in carbohydrate metabolism, it is not surprising that its availability and effectiveness play a role in every form of diabetes. However, because many other diabetogenic factors can be invoked and there is interdependence of many of these homeostatic mechanisms, teasing out their individual contributions is virtually impossible in any given patient.

Normal insulin physiology is orchestrated in a complex dynamic involving metabolic fuels, neurotransmitters, and other hormones. Insulin is synthesized as preproinsulin in the ribosomes of the rough endoplasmic reticulum of pancreatic islet β cells and is then converted to proinsulin, which in turn

TABLE 229-3 THE RELATIONSHIP BETWEEN HbA_{1c} AND ESTIMATED AVERAGE GLUCOSE LEVELS DURING THE PRECEDING 3 MONTHS

	ESTIMATED AVERAGE GLUCOSE LEVEL	
HbA_{1c} (%)	*mg/dL*	*mmol/L*
5	97	5.4
6	126	7.0
7	154	8.6
8	183	10.2
9	212	11.8
10	240	13.4
11	269	14.9
12	298	16.5

From Nathan DM, Kuenen J, Borg R, et al. Translating the A1c assay into estimated average glucose values. *Diabetes Care.* 2008;31:1473-1478.

TABLE 229-4 CONDITIONS THAT MAY AFFECT MEASUREMENT OR INTERPRETATION OF HbA_{1c}

MECHANISM OF HbA_{1c} INTERFERENCE	CONDITION OR DISEASE	EFFECT ON HbA_{1c}
Reduced red cell lifespan	Hemolytic anemia Acute blood loss Hypersplenism	Falsely low
Increased red cell lifespan	Iron deficiency anemia	Falsely high
Altered glycation	High-dose vitamin supplementation (vitamins A and C)	Falsely low
Assay interference	Hemoglobins S, G, D, C, E Hemoglobin F	Falsely high Falsely low
Miscellaneous	Chronic renal disease Chronic liver disease Red cell transfusion African ancestry	Falsely high Falsely low Falsely low or high Falsely high

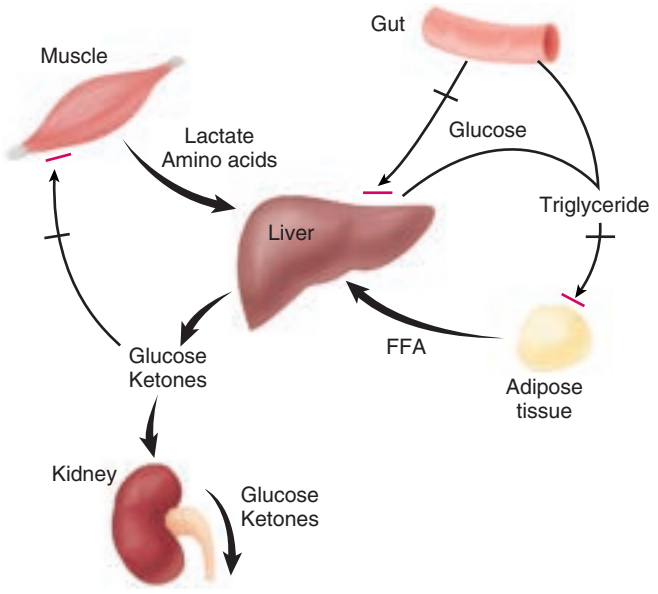

FIGURE 229-1. Effects of insulin deficiency on body fuel metabolism. Lack of insulin leads to mobilization of substrates for gluconeogenesis and ketogenesis from muscle and adipose tissue, accelerated production of glucose and ketones by the liver, and impaired removal of endogenous and exogenous fuels by insulin-responsive tissues. The net results are severe hyperglycemia and hyperketonemia that overwhelm renal removal mechanisms. FFA = free fatty acids.

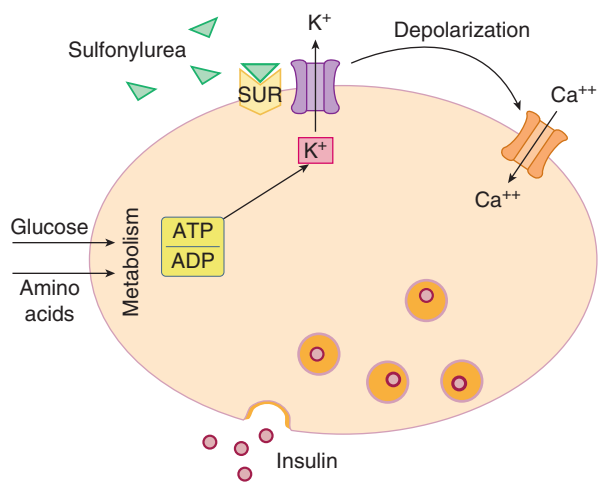

FIGURE 229-2. Nutrient regulation of insulin secretion. Glucose is taken up by the β cell through the GLUT2 glucose transporter and is metabolized (initially through phosphorylation by the glucokinase to glucose 6-phosphate). This leads to an increase in intracellular ATP (and an increase in the cytoplasmic ATP/ADP ratio), which causes closure of the ATP-dependent potassium channel, followed by membrane depolarization and the subsequent opening of voltage-gated calcium channels. The influx of calcium mobilizes the insulin secretory granules to fuse with the cell membrane and to release insulin into the extracellular fluid. The sulfonylurea 1 receptor (SUR1) is a component of the ATP-dependent potassium channel. ADP = adenosine diphosphate; ATP = adenosine triphosphate.

is transported to the Golgi apparatus, where it is packaged into secretory granules. Proinsulin is cleaved into equimolar amounts of insulin and a connecting segment (C-peptide) in the secretory granules. The stimulation of insulin secretion results in release of equimolar quantities of insulin and C-peptide (as well as a small amount of proinsulin) into the hepatic portal vein. Whereas a large proportion of insulin is bound to its hepatic receptor and metabolized in its "first pass" through the liver, C-peptide is much less prone to hepatic metabolism and is a better reflection of insulin secretion, although it is quantitatively of limited usefulness in the clinical diagnosis or treatment of diabetes.

The principal regulator of insulin secretion is glucose. The process of β-cell insulin secretion is shown schematically in Figure 229-2. Glucose is taken up by the β cells through the GLUT2 glucose transporter system and then phosphorylated to glucose 6-phosphate by an islet-specific glucokinase. Thus, glucokinase can be considered the "glucose sensor" of the β cell; mutations in this enzyme can lead to a specific diabetes syndrome (MODY2), and there is evidence of its role in common forms of type 2 diabetes. The conversion of glucose to glucose 6-phosphate results in a sequential increase in intracellular adenosine triphosphate (ATP), closing of the ATP-dependent potassium (K_{ATP}) channels in the β-cell membrane, membrane depolarization and influx of calcium, migration of the insulin secretory granules to the cell membrane and their fusion with the membrane, and finally release of insulin into the extracellular fluid. The K_{ATP} channel is made up of the sulfonylurea 1 receptor (SUR1) and an inward potassium channel subunit, Kir6.2. Mutations in either the SUR1 gene or the Kir6.2 gene lead to loss of K_{ATP} activity; as a result, the cell is depolarized, resulting in chronic release of insulin and a syndrome termed *persistent hyperinsulinemic hypoglycemia of infancy*. Mutations in Kir6.2 and SUR1 have also been identified in patients with permanent neonatal diabetes mellitus; treatment with sulfonylurea can normalize insulin secretion in these patients.

The magnitude of the insulin secretory response is determined by the level of blood glucose as well as by the mode of glucose entry. Compared with intravenous administration of glucose, higher insulin levels are produced when glucose is taken orally because of the simultaneous release of gut-derived incretins that include glucagon-like peptide 1 (GLP-1) and glucose-dependent insulinotropic peptide (GIP), both of which augment insulin secretion. In fact, drugs that mimic or enhance this incretin effect are useful in the treatment of type 2 diabetes.

Rapid increases in blood glucose concentration (e.g., after intravenous administration of glucose) cause a spike of insulin secretion that peaks within a few minutes and declines quickly (so-called first-phase insulin secretion). With more persistent elevations of plasma glucose concentration, insulin secretion is sustained (so-called second-phase insulin secretion). The earliest pathophysiologic indicator of defective β-cell function may be the loss of first-phase secretion of insulin, which precedes by years the decline in insulin secretory reserve sufficient to lead to overt glucose intolerance or diabetes.

TABLE 229-5 THE METABOLIC EFFECTS OF INSULIN

METABOLIC EFFECT	STIMULATED BY INSULIN	INHIBITED BY INSULIN
Carbohydrate metabolism	Glucose transport Glycolysis Glycogen synthesis	Glycogen breakdown Gluconeogenesis
Protein metabolism	Amino acid transport Protein synthesis	Protein breakdown
Lipid metabolism	Triglyceride uptake Lipogenesis	Lipolysis Fatty acid oxidation

Insulin Action

The actions of insulin on its principal target organs (i.e., muscle, fat, liver) have complex and coordinated effects on the metabolism of carbohydrates, proteins, and lipids and are mediated by its interaction with the insulin receptor. Insulin receptor signaling through insulin receptor substrate 1 and phosphatidylinositol 3-kinase is a major pathway in the mediation of insulin-stimulated glucose transport, notably by stimulating the translocation of the glucose transporter GLUT4 to the cell membrane. This pathway is also responsible for the vasodilator effects of insulin (through increased expression of endothelial nitric oxide synthase), which may contribute to glucose utilization by increasing nutrient delivery to tissues. Defects in these intracellular signaling pathways are an important cause of impaired insulin action, or "insulin resistance" (see section on impaired insulin action [insulin resistance] under pathobiology of type 2 diabetes).

The overall actions of insulin tend to promote uptake and storage of nutrients in the fed state and release of nutrients from body stores in the fasting state, as summarized in Table 229-5.

In the *postprandial period*, rising glucose levels simultaneously trigger insulin secretion and suppress glucagon release. The resulting rise in the insulin-to-glucagon ratio increases hepatic glycogen synthesis and inhibits release of glucose from the liver. Insulin stimulates glucose uptake into skeletal muscle and adipose tissue, promoting the synthesis of protein and triglycerides. In the *fasting state*, declining glucose levels inhibit insulin release, thereby increasing glycogenolysis and gluconeogenesis and the resulting delivery of glucose into the circulation. In states of absolute or relative insulin deficiency, inadequate basal insulin levels allow unrestrained hepatic glucose production, which results in fasting hyperglycemia. Inadequate insulin in the fed state impedes peripheral (predominantly skeletal muscle) glucose uptake, thereby contributing to postprandial hyperglycemia. Impaired suppression of hepatic glucose production also contributes to postprandial hyperglycemia in patients with diabetes (see also the section on type 2 diabetes).

TYPE 1 DIABETES

EPIDEMIOLOGY

Type 1 diabetes may be manifested at any age but most typically appears in childhood, especially around puberty. However, new cases of type 1 diabetes can appear at any time in life, and in the United States, approximately 30% of patients are diagnosed after young adulthood.[1]

Worldwide, the incidence of type 1 diabetes varies 50- to 100-fold, with the highest rates occurring in individuals of northern European descent. Both sexes are equally affected in childhood, but men are affected more commonly in early adult life. The incidence of childhood type 1 diabetes is rising rapidly in all populations, especially in the age group younger than 5 years, with a doubling time of less than 20 years in Europe. The increasing incidence of type 1 diabetes suggests a major environmental contribution, but the role of specific pathogenic factors remains largely unsettled. The distinction between type 1 and type 2 diabetes can become blurred in later life, and the true lifetime incidence of the condition is therefore unknown.

In Europe, the highest rates of childhood diabetes are found in Scandinavia, with an incidence for children from birth to 14 years of age ranging from 57/100,000 per year in Finland to 4/100,000 in Macedonia. In the United States, the overall annual incidence in youths is about 19/100,000. Prevalence rates are strikingly different among ethnic groups living in the same

geographic region, probably because of genetic differences in susceptibility to the disease. Early-onset diabetes carries a higher familial risk, and affected fathers are more likely to transmit type 1 diabetes to their offspring than affected mothers are, with risks being 6 to 9% and 1 to 3%, respectively.

Given that the United States does not have a systematic health registry and that its population is multiethnic, previous estimates of the prevalence and incidence of type 1 diabetes have been based on extrapolations from limited cohorts. The SEARCH for Diabetes in Youth multicenter study (funded by the Centers for Disease Control and Prevention and the National Institutes of Health) examined diabetes among children and adolescents in the United States. During 2008-2009, an estimated 18,436 people younger than 20 years in the United States were newly diagnosed with type 1 diabetes annually, and 5089 people younger than 20 years were newly diagnosed with type 2 diabetes annually. For those younger than 10 years, new cases of type 1 far outweighed type 2 (22.2/100,000 per year for type 1 diabetes vs. 0.8/100,000 for type 2 diabetes). Among youth aged 10 years or older, the rate of new cases of type 1 was about double that of type 2 (21.9/100,000 per year for type 1 diabetes vs. 11.0/100,000 for type 2 diabetes). Non-Hispanic white youth had the highest rate of new cases of type 1 diabetes in all age groups. Diabetes incidence rates by age and race/ethnicity are summarized in Figure 229-3.

Higher body mass index (BMI) is associated with younger age at diagnosis of type 1 diabetes, but this appears to be the case only in children with already compromised β-cell function. In addition, low birth weight may be a factor in accelerating the onset of type 1 diabetes, suggesting that the intrauterine environment may be an important determinant of age at onset for type 1 diabetes.

PATHOBIOLOGY

In type 1 diabetes, a complex interplay of genetic, environmental, and autoimmune factors selectively targets insulin-producing β cells and ultimately produces complete β-cell destruction. The role of genetic factors in type 1 diabetes has long been appreciated, emphasized by familial clustering with other autoimmune endocrine disorders and by concordance rates in identical twins of 30 to 40%. Because these concordance rates are not as high as in type 2 diabetes (i.e., >80%), environmental factors must clearly play a major role. Although the presence of an environmental trigger for type 1 diabetes is highly likely, even identical twins do not express identical T-cell receptor and immunoglobulin genes; as a result, total concordance might not be expected. Siblings who are HLA identical to the proband have a 12 to 15% risk for development of diabetes by the age of 20 years.

Although many of the genes linked to type 1 diabetes have yet to be identified, some are known. HLA genes, located on the short arm of chromosome 6, contribute about 50% of genetic susceptibility to type 1 diabetes. Two HLA class II haplotypes, DR4-DQ8 and DR3-DQ2, are present in about 90% of children with type 1 diabetes. The genotype containing both haplotypes carries the highest risk of diabetes (about 5%) and is most commonly seen in early-onset disease. In contrast, the DR15-DQ6 haplotype is highly protective, being found in only 1% of children with type 1 diabetes in contrast to 20% in the general population. HLA susceptibility haplotypes are overrepresented in adult-onset type 1, but at lower frequency than in classic type 1 diabetes in youth. Other genes likely contribute to the genetic susceptibility to type 1 diabetes. These include the insulin gene (on chromosome 11) and a number of other loci that are associated with other autoimmune conditions, suggesting the existence of common pathways predisposing to loss of self-tolerance. Another gene, *IFIH1*, located on chromosome 2, encodes a protein involved in innate immunity and plays a role in recognition of the RNA genomes of certain viruses. It is suggested that high IFIH1 levels might provoke exaggerated antiviral immune responses that predispose to autoimmunity. Many other genes have also been implicated, underscoring the polygenic nature of this disease.

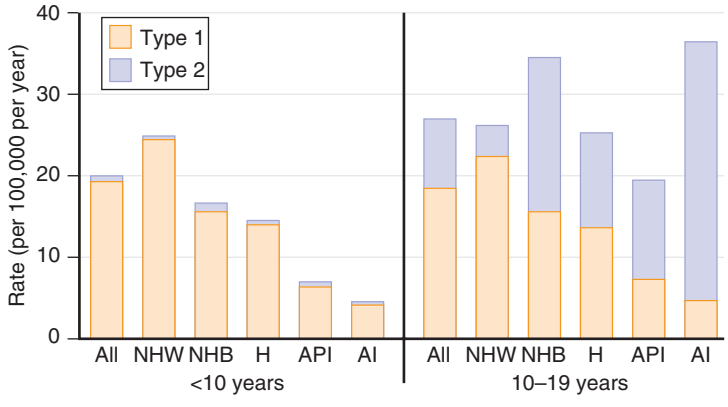

FIGURE 229-3. Rate of new cases of type 1 and type 2 diabetes among people younger than 20 years in the United States, by age and race/ethnicity, 2008-2009. AI = American Indians; API = Asian/Pacific Islander Americans; H = Hispanics/Latinos; NHB = non-Hispanic blacks; NHW = non-Hispanic whites. (From Centers for Disease Control and Prevention. *National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States, 2014.* Atlanta, GA: US Department of Health and Human Services; 2014. Source: SEARCH for Diabetes in Youth Study).

Historically, environmental causes of type 1 diabetes focused on viruses because of associations with seasonal pandemics of infections and rarely because of the isolation of a specific pathogen. Epidemics of mumps, rubella, and coxsackievirus infection have been associated with an increased frequency of type 1 diabetes. Moreover, specific and convincing rare examples of virus-induced diabetes have been reported. However, it is believed that virus-mediated β-cell damage is not responsible for the massive destruction of β cells but that it triggers an autoimmune response in genetically predisposed individuals. Thus, viruses may contain molecules that resemble a β-cell protein, and viral infection could thus nullify self-tolerance and trigger autoimmune responses.

It has long been recognized that about 80% of patients with new-onset type 1 diabetes have antibodies directed against various islet cell proteins, including insulin, glutamic acid decarboxylase (GAD65 and GAD67), and the secretory granule protein islet cell antigen 512 (IA-2). These antibody biomarkers have been important tools for studying the potential for early identification and prevention of total β-cell destruction in individuals susceptible to type 1 diabetes. Until the mid-1980s, it was mistakenly surmised that the autoimmune destruction of β-cells was mediated by these antibodies rather than their being epiphenomena, as is now understood. Rather, β-cell destruction is mediated by a variety of cytokines or by direct T-lymphocyte activity that causes apoptosis or cellular destruction. Both animal models and human pathologic studies have established that islet-targeted inflammatory cell infiltrates (termed insulitis) that are composed of $CD8^+$ and $CD4^+$ T cells, macrophages, and B cells are linked to the onset of diabetes. Over time, the islets become completely devoid of β cells and inflammatory infiltrates; α, δ, and pancreatic polypeptide cells are left intact, thus illustrating the specificity of the autoimmune attack on β cells.

A critical role for T cells is suggested by studies involving pancreatic transplantation in identical twins. Monozygotic twins with diabetes who received kidney and pancreas grafts from their nondiabetic, genetically identical siblings required little or no therapeutic immunosuppression. However, these patients eventually experienced a resumption of insulitis, with the subsequent recurrence of diabetes. Evidence implicating T cells in diabetes autoimmunity also derives from clinical trials using immunosuppressive drugs. Drugs such as cyclosporine or antibodies directed against a component of the T-cell receptor (anti-CD3) or that alter antigen presentation by B cells (anti-CD20) slow the progression of recent-onset diabetes, but this effect is not sustained if immunosuppression is withdrawn.

CLINICAL MANIFESTATIONS

It has been clearly established that type 1 diabetes has a long preclinical phase, best described in Figure 229-4. At the time of clinical diagnosis, about 10 to 20% of the original β-cell mass may still be functional. In most cases, overt hyperglycemia (and ketosis if it is present) may be precipitated by an unrelated medical illness or stress placed on an already-limited islet reserve, thus triggering the diagnosis. Typically, symptomatic hyperglycemia, manifested by polyuria, polydipsia, weight loss, and fatigue, occurs abruptly in an otherwise healthy child or young adult. For a minority of patients, the initial presentation may be diabetic ketoacidosis (DKA), which can occur if there is a delay in recognizing the symptoms of diabetes. Whereas the disease has an increased incidence in the winter months, classically attributed to respiratory viral infections, this seasonal pattern may be the result of illness-associated counter-regulatory hormones that drive hyperglycemia in individuals with already compromised β-cell function. Similarly, the coincidence of type 1diabetes with puberty has been attributed to insulin resistance associated with increases in sex and growth hormone secretion.

The diagnosis of diabetes is made according to glucose criteria (see Table 229-2). Measurement of anti–glutamic acid decarboxylase antibodies is sometimes performed, but the determination of type 1 etiology is generally

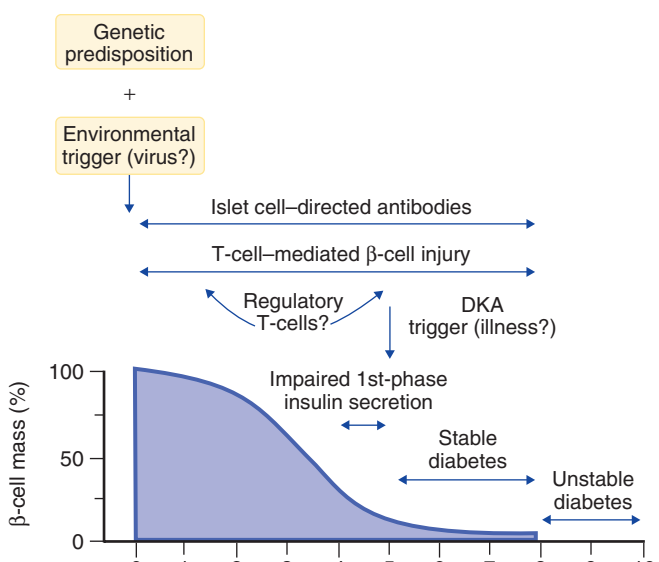

FIGURE 229-4 **Summary of the sequence of events that lead to pancreatic β-cell loss and ultimately to the clinical evolution of type 1 diabetes.** DKA = diabetic ketoacidosis.

made on clinical grounds. After initiation of insulin therapy and stabilization of plasma glucose levels, the patient may experience a period of weeks to months of relatively mild and easily controlled hyperglycemia. This so-called honeymoon phase of type 1 diabetes results from transient improvement in β-cell function and reflects the phenomenon of severe but not total β-cell destruction with ongoing (albeit reduced) insulin secretion. Ultimately, patients with type 1 diabetes experience progressive decline in insulin production, generally to undetectable levels after a few years. However, with highly sensitive C-peptide assays, low levels of insulin production have been detected in some patients with long-standing type 1 diabetes and are associated with more stable glycemic control and reduced risk of vascular complications.[2] In patients with onset of type 1 diabetes in adulthood, the clinical presentation may follow a more indolent course (termed latent autoimmune diabetes in adults), perhaps because β-cell mass declines at a slower pace. In fact, type 1 diabetes may be misdiagnosed as type 2 in many of these patients until the progression of insulin deficiency reveals the phenotype of permanent and complete insulin dependence.

TREATMENT

Rx

The key to successful treatment of type 1 diabetes is to achieve physiologic insulin replacement, that is, to replicate the normal and tightly regulated relationship between plasma glucose and insulin secretion. Although current technology can only mimic this normal physiology, substantial progress has been made to permit maintenance of relative euglycemia by many patients. Successful glucose management requires substantial commitment by the patient and health care practitioner.

Insulin Therapy

All patients with type 1 diabetes require insulin treatment to maintain life. The approach to insulin replacement in type 1 diabetes requires consideration of both basal insulin requirements (insulin required to maintain homeostasis in the fasting state) and insulin required for the influx of nutrients that occurs with meals. A variety of insulin preparations are available, which differ by their pattern of absorption after subcutaneous injection. Most currently used insulin preparations are analogues of human insulin that have been modified (usually by changing one or more amino acids) to alter pharmacokinetics to speed or to delay absorption (Table 229-6).

Patients with type 1 diabetes are treated with both a long-acting "basal" insulin and a shorter-acting "prandial" insulin at mealtime, by a multiple daily insulin injection regimen or a continuous subcutaneous insulin infusion pump. Typically, the daily insulin requirement for patients with type 1 diabetes is between 0.3 and 1.0 unit/kg/day, with half given as basal insulin and the remainder divided into pre-meal boluses. Prandial insulin doses are determined by meal carbohydrate content plus a "correction factor" if glucose is elevated before the meal. For example, a common approach is to use 1 unit for every 10 to 15 g of meal carbohydrate plus a correction factor of 1 unit to lower plasma glucose concentration by 20 to 50 mg/dL. However, insulin requirements are influenced by a number of factors (e.g., age, body size, insulin sensitivity) and vary substantially among patients; therefore, these algorithms need to be individualized. A number of mobile phone applications and computer programs are available to assist patients with dose calculation. Critical to the success of physiologic insulin replacement is the need for the patient to monitor blood glucose concentration, generally several times a day (see later).

A continuous subcutaneous insulin infusion pump using a short-acting insulin analogue can be programmed to deliver both a basal infusion and a preprandial bolus. Most insulin pumps contain an insulin reservoir attached by thin flexible tubing to a very small catheter that is inserted subcutaneously by the patient and changed every 2 or 3 days to avoid local inflammation and fibrosis, which can interfere with insulin absorption. The basal insulin delivery rate can be programmed to vary throughout the day and may be especially useful to prevent hyperglycemia associated with the "dawn phenomenon" (rising blood glucose levels in the early morning hours, thought largely to be due to increased growth hormone secretion). Most insulin pumps can be programmed to calculate prandial insulin doses, based on pre-meal glucose level and meal carbohydrate content, which is entered by the patient. However, in the event of pump malfunction, metabolic decompensation, including DKA, can develop within several hours because there is no subcutaneous reservoir of long-acting insulin. Successful use of an insulin pump requires a motivated and educated patient plus the support of a specialized diabetes team, including a certified diabetes educator. When it is used appropriately, continuous subcutaneous insulin infusion provides patients with maximal lifestyle flexibility and the best chance to achieve near-normal blood glucose levels.

Some patients who find adherence to a multiple injection or insulin pump regimen difficult can be treated with premixed "biphasic" insulin combinations, for example, a mixture of NPH and regular insulin given twice daily. This approach may be appropriate for patients with recent onset of type 1 diabetes who still maintain some endogenous insulin production. However, for most patients, this regimen is rarely optimal because it lacks flexibility and often increases the risk of hypoglycemia.

TABLE 229-6 INSULIN PREPARATIONS

TYPE OF INSULIN	ONSET OF ACTION	PEAK EFFECT	DURATION OF ACTION
BASAL INSULIN			
Glargine	~ 2 hours	None	~ 24 hours
Detemir	~ 2 hours	3-9 hours	6-24 hours
Degludec	~ 2 hours	None	~ 40 hours
NPH/NPL	~ 2 hours	6-12 hours	14-24 hours
PRANDIAL INSULIN			
Lispro, aspart, glulisine	5 to 15 minutes	45-75 minutes	2-4 hours
Regular	~ 30 minutes	2-4 hours	5-8 hours

NPH = neutral protamine Hagedorn; NPL = neutral protamine lispro.

Diet and Lifestyle Treatment

In type 1 diabetes, the focus of dietary planning is on accurate estimation of meal carbohydrate content to allow appropriate prandial insulin dosing. This can be approached by promotion of "carbohydrate consistency" from meal to meal and the use of relatively fixed pre-meal insulin dosing. A more flexible approach is for the patient to learn "carbohydrate counting," which specifies an insulin dose per amount of carbohydrate in the meal. With either approach, patients need to monitor the nutrient content of their meals. Avoidance of concentrated sweets and other high-carbohydrate meals, including those with a high "glycemic index," tends to facilitate accurate insulin dosing and to minimize postprandial glycemic excursions. In contrast to type 2 diabetes, most patients with type 1 diabetes are not overweight or obese, and calorie restriction is neither required nor helpful. A variety of eating patterns are considered acceptable, and recommendations for a "heart healthy" diet (low in saturated fat and cholesterol) are the same as for the general population.[3]

Glucose Self-monitoring

Successful management of type 1 diabetes requires consistent self-monitoring of blood glucose concentration by the patient or caregiver several times a day. Small portable meters with disposable strips are easy to use and reasonably accurate in most ambulatory care settings. Frequent testing (i.e., before meals and at bedtime) allows appropriate prandial insulin dosing and correction of unexpected hyperglycemia as well as detection or confirmation of hypoglycemia. Most current meters store a large number of readings, which

can be downloaded to a computer for analysis by the patient and health care team. Subcutaneous glucose monitors that provide continuous reading of interstitial glucose levels are available and are most commonly used in conjunction with an insulin pump. These monitors are most useful to determine glucose patterns and can be programmed to sound an alarm when the glucose level exceeds a preset range or rate of change. However, the accuracy of current monitors is such that they cannot replace conventional blood glucose monitoring for immediate treatment decisions. Research is progressing in the development of a "closed loop" system combining an insulin pump with a highly accurate continuous glucose monitor, allowing adjustments of insulin dosing without direct input of the patient. The first such device has received Food and Drug Administration approval for patients older than 16 years, but experience with this approach remains limited.

Patients with type 1 diabetes should be instructed to test urine ketones (with a reagent strip) in situations in which blood glucose concentration is unexpectedly and persistently elevated, especially if it is accompanied by symptoms suggestive of DKA (see section on diabetic ketoacidosis under hyperglycemic states in acute metabolic complications of diabetes). Small or trace amounts of urinary ketones are not cause for concern, but moderate or large amounts may indicate the onset of DKA and should prompt the patient to seek urgent medical attention.

Whole Pancreas and Islet Cell Transplantation

The ultimate goal of a "cure" of type 1 diabetes could most likely be achieved by successful transplantation of insulin-producing β cells. Whole pancreas transplantation has been performed for more than two decades and has a reasonable success rate, with 5-year graft survival rates of about 70%. However, the surgical procedure is complicated, and lifelong immunosuppression is required, as with any organ transplant. For these reasons, pancreas transplantation is generally reserved for patients who already have or are concurrently receiving a kidney transplant. In the absence of indications for kidney transplantation, pancreas-alone transplantation may be considered for patients who have a history of frequent, acute and severe metabolic complications (especially severe hypoglycemia) or severe and incapacitating psychosocial problems related to insulin therapy. Pancreatic islet cell transplants hold significant potential advantages over whole-gland transplants. However, at this time, islet cell transplantation is an experimental procedure, also requiring systemic immunosuppression, and is performed only within the setting of controlled research studies.

PREVENTION OF TYPE 1 DIABETES

Given that type 1 diabetes is an immunologically mediated disease, it has long been supposed that immune intervention should alter its natural history and perhaps even prevent it altogether. Furthermore, the significant heritability of type 1 diabetes suggests that treatment could target only susceptible individuals, and the existence of known biomarkers (antibodies that reflect disease activity as well as levels of insulin or C-peptide that reflect islet function) also lends credence to experimental immunologic treatments. Unfortunately, however, the major challenge for most immunologic interventions has been their lack of specificity for immune-mediated insulitis or the risks of spillover immune suppression in otherwise healthy persons. Given the experimental nature of all the tested therapies, we provide only a brief overview here.[4]

Prevention of type 1 diabetes can theoretically be undertaken at three stages: (1) in susceptible individuals before there is evidence of immune attack against islet cells (primary prevention); (2) in nondiabetic people who already have evidence of immune activation (antibodies, insulin defects) to prevent progression to actual diabetes (secondary prevention); and (3) in newly diagnosed patients in whom the goal is to slow the β-cell destructive process (tertiary prevention).

Avoidance of putative environmental triggers of islet autoimmunity (e.g., cow's milk) is one approach, and dietary supplementation with nutrients that may diminish islet autoimmunity (e.g., omega-3 fatty acids or vitamin D) has been attempted. On the basis of positive results of an earlier small cohort study, the first large trial of primary prevention by removal of cow's milk from the infant diet was initiated in 2002 and is expected to be completed in 2017. Secondary prevention trials have also been undertaken with oral, inhaled, or injected insulin and with nicotinamide, but the results have been equivocal. There are secondary prevention studies with teplizumab (an FcR-nonbinding anti-CD32 monoclonal antibody) and with abatacept (a costimulation modulator) under way. Several tertiary prevention studies (i.e., after the diagnosis of diabetes) have been published. Nonspecific immune interventions, such as cyclosporine, demonstrate that immunotherapy can indeed rescue β cells from ongoing destruction, but it is not an acceptable therapeutic alternative given that it preserves β-cell function only transiently and carries heightened risks of adverse effects, such as nephropathy. Anti-CD3 antibodies currently appear more promising and are being evaluated in clinical trials.

PROGNOSIS

Substantial progress has been made in recent decades in improving the prognosis for patients affected by type 1 diabetes. This is largely due to adoption of more intensive glucose control and more effective nonglycemic treatment of early stages of renal disease and retinopathy. Data from long-term follow-up of the intensively treated DCCT cohort showed that after 30 years' duration of diabetes, rates of serious diabetes complications were substantially lower than in historical controls, and less than 1% became blind, required renal replacement, or had an amputation due to diabetes. In the absence of renal disease, life expectancy for type 1 patients in the United States is comparable to that of the general population.[5] However, the mortality rate of all type 1 diabetic patients from age 35 onward is about twice as high as in nondiabetics even if the $HgbA_{1c}$ level is 6.9% or lower and becomes progressively higher for $HgbA_{1c}$ levels above 7.9%.[6] Analysis of nationally representative hospitalization and registry data has shown large reductions in the incidence of a broad spectrum of diabetes-related complications between 1990 and 2010 in the U.S. population of adults with diabetes; however, despite the substantial decline in the rates of diabetes-related complications in the past two decades, a large burden of disease persists because of the continued increase in the prevalence of diabetes.[7]

TYPE 2 DIABETES

EPIDEMIOLOGY

Type 2 diabetes is one of the most common chronic diseases, affecting more than 25 million people in the United States and an estimated 366 million worldwide. The prevalence of type 2 diabetes has been increasing in the United States, from approximately 3% of the population in 1995 to more than 9% in 2012. This increase is in part due to demographic shifts (i.e., the aging of the population), but incidence rates are also increasing and parallel the rise of overweight and obesity as well as increasingly sedentary lifestyles. A similar pattern is observed globally, with projections of 550 million (approximately half undiagnosed) to be affected by 2030. Although type 2 diabetes is being increasingly recognized in obese adolescents and young adults, older age remains a major risk factor for type 2 diabetes. More than one quarter of adults aged 65 years and older have diabetes, and another 50% have glucose or HbA_{1c} levels in the impaired or pre-diabetic range. Type 2 diabetes in the United States is more common among some racial and ethnic groups, with prevalence rates highest among non-Hispanic blacks (13%), Hispanics (12%), and American Indians (16%) and lowest among non-Hispanic whites (7%). Individuals from the Indian subcontinent (i.e., India, Pakistan, Bangladesh) and the Pacific Islands (e.g., Hawaii, Nauru, Samoa) also have high rates of type 2 diabetes. In general, men and women have about equal prevalence of type 2 diabetes.

PATHOBIOLOGY

Type 2 diabetes is characterized by variable defects in both insulin secretion and insulin action.[8] The underlying metabolic phenotype of type 2 diabetes is distinctly heterogeneous among individuals with the disease; some have a more pronounced defect in insulin secretion, and others have greater resistance to insulin action. The metabolic profile also varies within a given patient over time, as insulin secretion progressively declines with longer duration of disease. Although heterogeneous, type 2 diabetes is characterized in all cases by inadequate insulin secretion for the prevailing glucose level and degree of insulin sensitivity.

Impaired Insulin Secretion

The relative insulin deficiency characteristic of type 2 diabetes appears to be a consequence of both functional (i.e., reduced responsiveness to secretagogues) and quantitative (i.e., reduction in β-cell mass) factors. Insulin secretory capacity is difficult to directly measure in humans, but reductions of β-cell mass up to 60% are estimated to occur in type 2 diabetes. However, this alone is insufficient to explain insulin deficiency in type 2 diabetes, as evidenced by the observation that 50% surgical pancreatectomy does not lead to hyperglycemia in otherwise healthy individuals. Classic studies in diabetic patients have demonstrated failure of insulin secretion in response to glucose but a normal response to the amino acid arginine, providing further evidence

for the presence of a functional defect specific to glucose sensing. Abnormalities in the usual pulsatile and oscillatory patterns of insulin secretion and inefficient insulin biosynthesis have also been demonstrated in type 2 diabetes. For example, abnormal peptide processing results in increased secretion of intact proinsulin, which serves as a useful biomarker of future diabetes risk. Increased accumulation of amyloid also occurs within diabetic islets and may contribute to impaired secretory function. Ultimately, the β-cell defects in type 2 diabetes appear to be multifactorial, in part genetically determined (see later) but also influenced by environmental exposure, for example, to high levels of circulating glucose (glucotoxicity) and lipids (lipotoxicity). In addition, the β-cell defects are not static but worsen with increasing duration of diabetes.

Impaired Insulin Action (Insulin Resistance)

Resistance to the metabolic effects of insulin is also a characteristic although variable feature of type 2 diabetes. Hyperinsulinemia, thought to be a compensatory response to impaired insulin action, can be demonstrated in patients with pre-diabetes and in many patients with established type 2 diabetes, particularly early in its course. More precise techniques to measure insulin action (e.g., the euglycemic hyperinsulinemic clamp) have demonstrated resistance to insulin action primarily in peripheral tissues (reduced capacity to stimulate glucose uptake in muscle and fat) but also in the liver (reduced capacity of insulin to suppress hepatic glucose production). Insulin resistance is closely associated with obesity (see later) but also has genetic determinants, reflected by the observation that some obese patients do not have severe insulin resistance. Insulin resistance frequently occurs as part of a constellation of features, termed the metabolic syndrome, which include hypertension, abdominal obesity, dyslipidemia, glucose intolerance, and increased cardiovascular risk. Insulin resistance is also a common feature of the polycystic ovary syndrome.

There are multiple molecular mechanisms that can lead to resistance to physiologic insulin action, including pre-receptor defects (e.g., an abnormal insulin molecule) and abnormal insulin receptors (e.g., due to gene mutations). However, common forms of insulin resistance that occur in association with type 2 diabetes are generally due to post-receptor defects, that is, abnormalities in intracellular signaling. In insulin target tissues, signaling through the phosphatidylinositol 3-kinase pathway is responsible for translocation of the glucose transporter GLUT4, which is necessary for uptake of glucose into the cell. Several defects in this pathway have been described in humans with insulin resistance, including abnormalities in insulin receptor substrate 1 and protein kinase B/Akt2. Some specific gene mutations associated with insulin resistance have been identified, but insulin resistance may also be acquired as a consequence of obesity (see later), increases in circulating free fatty acids, certain medications (e.g., glucocorticoids, niacin), and inflammatory states.

Evidence from natural history and genetic association studies (see later) indicates that defects in either insulin action or insulin secretion can remain clinically silent. For example, insulin resistance can induce compensatory hyperinsulinemia, which early in the course of the disease is sufficient to maintain euglycemia. However, in individuals with inherited or acquired defects in β-cell function, this compensation ultimately fails and hyperglycemia ensues. Viewed another way, a subclinical β-cell defect may remain silent in the setting of normal insulin sensitivity but be manifested as hyperglycemia when acquired insulin resistance develops because of weight gain, aging, or some other factor. A unifying theory that explains the coexistence of defects in both insulin action and insulin secretion is appealing but so far elusive.

Genetics

The evidence for familial aggregation of type 2 diabetes is substantial and supports the presence of an important genetic influence. An individual with one parent with type 2 diabetes has a lifetime risk for development of type 2 diabetes of approximately 40%, with risk increasing to approximately 70% if both parents are affected. Further, the concordance rate among monozygotic twins is as high as 70%. The greater risk of type 2 diabetes among certain racial and ethnic groups also supports an important genetic component. The overall heritability of type 2 diabetes is estimated to be between 25 and 50%, although the specific gene or genes ultimately responsible for common forms of type 2 diabetes have not been established.[9]

Single-Gene Mutations Linked to Type 2 Diabetes Phenotype

A number of syndromes, termed maturity-onset diabetes of youth (MODY), characterized by impaired β-cell function have been recognized and linked to specific single-gene mutations. The phenotypes vary with the mutation but generally include early onset of relatively mild hyperglycemia in nonobese children or young adults and an autosomal dominant pattern of inheritance. Although not common (representing 1 to 3% of diabetes cases worldwide), their discovery has provided insight into the role of β-cell function in the more common forms of type 2 diabetes. MODY 2 is associated with a mutation of glucokinase, which acts as a glucose sensor within the β cell. In this case, higher levels of glucose are required to stimulate release of insulin from the β cell. MODY 3 is due to a mutation of the gene for hepatic nuclear factor 1α, which is involved in early pancreatic development and in the regulation of insulin gene expression. Other MODY forms (1, 4, 5, 6) are much less common, having been described in only a few families (Table 229-7).

TABLE 229-7 SINGLE-GENE MUTATIONS RESPONSIBLE FOR THE MORE COMMON FORMS OF MATURITY-ONSET DIABETES OF YOUTH (MODY)

	MUTATION	METABOLIC DEFECT	CLINICAL PHENOTYPE
MODY 2	Glucokinase	Decreased β-cell sensitivity to glucose	Mild nonprogressive hyperglycemia, may not require pharmacologic treatment; diabetes complications rare
MODY 3	Hepatic nuclear factor 1α	Abnormal regulation of β-cell gene transcription	Mild hyperglycemia, may be progressive; renal glycosuria; increased sensitivity to sulfonylurea drugs; susceptibility to microvascular complications

Other examples of single-gene mutations associated with specific diabetes syndromes include activating mutations of *KCNJ11* (encoding a portion of the β-cell sulfonylurea receptor), which causes severe neonatal diabetes, and *WFS1*, which encodes a protein that is defective in Wolfram syndrome (diabetes insipidus, diabetes mellitus, optic atrophy, and deafness).

Polygenic, Common Type 2 Diabetes

Common forms of type 2 diabetes are likely polygenic and multifactorial and represent a complex interaction between genes and environment. In the past several years, more than 50 genetic risk loci for common forms of type 2 diabetes have been identified by genome-wide association studies, but collectively these explain only about 15% of the heritability of the disorder. The gene with the strongest effect size to date (odds ratio for diabetes, 1.4), *TCF7L2*, is associated with reduced insulin secretion, as are the majority of other recognized gene variants. Others include a β-cell zinc transporter (ZnT-8; odds ratio, 1.15), the sulfonylurea receptor (*KCNJ11*; odds ratio, 1.1), and melatonin receptor 1B (*MTNR1B*; odds ratio, 1.10). A smaller number of gene variants associated with insulin resistance have been identified and include genes encoding peroxisome proliferator-activated receptor γ (odds ratio, 1.20) and insulin receptor substrate 1 (odds ratio, 1.10). Sequence variants in *SLC16A11*, a gene involved in intracellular lipid metabolism, were discovered to be a relatively common risk allele (odds ratio, 1.29) in Mexican populations. Current knowledge about genetic variants associated with type 2 diabetes risk is not useful for clinical disease prediction, offering no advantage to simple clinical tools based on traditional risk factors.

There is emerging evidence that epigenetic changes may play an important role in development of type 2 diabetes. Epidemiologic studies suggest that "metabolic programming" may occur in utero, with both fetal starvation and fetal overnutrition predisposing to diabetes in adult life. One example comes from experience with the Pima Indians, who have an extremely high prevalence of type 2 diabetes. Children born to mothers who were diabetic during their pregnancy had higher rates of adult diabetes than did children born to the same mothers before they became diabetic, suggesting that intrauterine exposure can have long-lasting metabolic effects. Studies of DNA methylation in animal models support this hypothesis, although human epigenome-wide studies are just now being conducted. Conversely, maternal undernutrition is associated with diabetes in offspring. Low birth weight has been linked to predisposition to CVD and diabetes in adulthood. According to the widely cited Barker hypothesis, nutrient deficiency in utero (e.g., due

to maternal starvation or placental insufficiency) impairs development of the endocrine pancreas, leading to inadequate insulin production later in life.[10] The available data suggest that maternal nutrition may play an important role in metabolic programming, resulting in increased diabetes susceptibility in adulthood.

Obesity

The presence of overweight or obesity (Chapter 220) substantially increases the risk for type 2 diabetes and likely accounts for the dramatic increase in diabetes prevalence during the past several decades. In fact, the presence of overweight or obesity is the single most important clinical predictor of type 2 diabetes, particularly for young or middle-aged individuals. The relationship between BMI and type 2 diabetes is linear, and increased risk can be observed even within the BMI range defined as normal (<25 kg/m^2). Related factors, such as sedentary lifestyle and diet (increased consumption of foods with high glycemic load, increased *trans* and saturated fat), may also contribute to diabetes risk, independent of BMI. Distribution of body fat also plays an important role, with visceral adiposity (as assessed by waist circumference or waist-to-hip ratio) being a particularly strong diabetes risk factor in Asian populations, who tend to develop type 2 diabetes at a lower BMI than some other racial or ethnic groups do. Ectopic accumulation of adipose tissue in the liver, often manifested as nonalcoholic fatty liver disease, is also strongly associated with increased diabetes risk.

The increase in adipose tissue mass impairs insulin action by a number of proposed mechanisms, including alterations in fatty acid metabolism, accumulation of triglycerides in the liver, and low-grade systemic inflammation. Adipose tissue macrophages produce proinflammatory cytokines, including tumor necrosis factor-α and interleukin-6, which can interfere with insulin signaling. Obesity is also associated with reduced levels of the fat-derived peptide adiponectin, which exhibits both anti-inflammatory and insulin-sensitizing properties. The increase of circulating free fatty acids that is characteristic of obese states can interfere with insulin action in skeletal muscle and liver, and increased intramyocellular lipid is also associated with insulin resistance. Further, increased lipid accumulation in pancreatic islets may lead to impaired insulin secretion. Interestingly, some obese individuals have apparently normal insulin sensitivity and glucose metabolism, sometimes referred to as the obesity paradox. The mechanisms that may protect someone from the diabetogenic effects of excess adiposity are not known, but intact cardiorespiratory fitness may play a role.

CLINICAL MANIFESTATIONS

Typical Type 2 Diabetes

The classic hyperglycemic symptoms of polyuria, polydipsia, and weight loss occur when the renal threshold for glucose reabsorption (~180 mg/dL) is exceeded and glycosuria with osmotic diuresis occurs. Therefore, patients may have plasma glucose concentration that is elevated but below this threshold, for years if not for decades, before specific symptoms appear. In the current era, many patients are found to have diabetes during routine screening or in the course of investigation for another disorder (typically, CVD). The initial presentation for some patients may be severe decompensated hyperglycemia, with profound dehydration, electrolyte imbalance, and plasma glucose levels of 400 mg/dL or higher, with the most striking examples being hyperosmolar hyperglycemic state (HHS) and DKA (see corresponding sections later).

A key feature of type 2 diabetes is that the metabolic defects are not static but tend to worsen over time. A patient early in the course of type 2 diabetes may maintain acceptable glucose control with simple dietary modification and modest weight loss. For many patients, these measures alone fail over time, and combinations of oral medications and often insulin therapy become necessary to control blood glucose levels.

Although the defining clinical feature of type 2 diabetes is hyperglycemia, it is actually the vascular complications of the disorder that cause the greatest morbidity and mortality. For a minority of patients, the initial clinical presentation of diabetes may be the presence of diabetic microvascular complications (retinopathy, neuropathy, nephropathy), which usually indicates many years of unrecognized hyperglycemia. More typical is the insidious onset of symptomatic microvascular complications after many years of diabetes, especially if it is poorly controlled.

Atypical Diabetes

DKA may be the initial clinical presentation for a minority of patients with type 2 diabetes, who subsequently recover β-cell function and do not require insulin treatment. This entity, referred to as ketosis-prone type 2 or Flatbush diabetes (named for the New York City neighborhood where it was first described), appears to be more common among African Americans and some other ethnic minorities. These patients typically lack markers of β-cell autoimmunity and have a strong family history of type 2 diabetes. Once the initial episode of DKA is treated and glucose levels stabilize, patients may have near-normoglycemic remissions lasting many years. The pathogenesis of this form of diabetes is not clear, but a unique β-cell predisposition to glucose desensitization ("glucose toxicity") has been proposed.

Gestational Diabetes

Diabetes that appears for the first time during pregnancy and typically regresses after delivery is termed gestational diabetes (Chapter 239). Women who develop gestational diabetes usually have risk factors including overweight or obesity, older age (>30 years), and a family history of type 2 diabetes. The majority will develop permanent type 2 diabetes during their lifetime. Hormonal changes (increases in placental lactogen, estrogen, progesterone) induce insulin resistance during pregnancy and may uncover latent β-cell defects in predisposed women. Babies born to mothers with diabetes mellitus are at risk for a number of adverse outcomes, especially macrosomia but also preterm birth, neonatal hypoglycemia, and hyperbilirubinemia. Routine screening, with an oral glucose tolerance test, of all pregnant women at 24 to 28 weeks of gestation is currently recommended. Because glucose levels tend to be lower than in the nonpregnant state, separate criteria have been developed for the diagnosis of diabetes in pregnancy. These include the presence of any of the following: fasting glucose concentration of 92 mg/dL or higher; glucose concentration of 180 mg/dL or higher at 1 hour or 153 mg/dL or higher at 2 hours after a 75-g oral glucose load. Aggressive glycemic control has been shown to reduce adverse pregnancy outcomes, including macrosomia and traumatic delivery, although its effects on long-term outcomes in offspring have not been established. Medical nutrition therapy is recommended for all women with gestational diabetes, with emphasis on moderate carbohydrate intake and avoidance of excessive weight gain. If diet modification is inadequate to maintain euglycemia, insulin has historically been considered first-line pharmacologic treatment for gestational diabetes. Oral diabetic medications, including glyburide and metformin, are increasingly being used to treat gestational diabetes, although they are not approved for this indication by the U.S. Food and Drug Administration. After delivery, women with gestational diabetes should continue to be observed for the development of type 2 diabetes.

TREATMENT

Effective treatment of type 2 diabetes is uniquely challenging because it encompasses management of lifestyle factors (including diet, exercise, and weight control), use of multiple oral or injectable medications, self-monitoring of blood glucose concentration, and surveillance and treatment for acute and chronic diabetic complications. The active participation of the patient in this complex program is critical for successful diabetes management, and many patients benefit from participation in a program of diabetes self-management education.

Goals of Therapy Including Glucose Targets

The primary goals of diabetes management are to prevent symptomatic hyperglycemia and hypoglycemia and to prevent the vascular complications associated with diabetes (see later section on chronic vascular complications). Intensive glycemic control (near-normoglycemia) has been shown to reduce microvascular and neuropathic complications of diabetes but not CVD or mortality.■ The current consensus view is that lowering of the HbA_{1c} level to 7% or below is an appropriate goal for most patients with diabetes. More stringent glycemic control (HbA_{1c} level close to the normal range) may be appropriate for some individuals (e.g., young patients with short duration of disease) if it can be achieved without excessive hypoglycemia. Conversely, less stringent goals may be suitable for patients with established vasculopathy, significant comorbidities, or reduced life expectancy. Generally accepted targets for fasting and postprandial glucose levels are 70 to 130 mg/dL and less than 180 mg/dL, respectively. Glycemic targets during pregnancy are different, in part because plasma glucose levels normally are lower during pregnancy and because of the risk of adverse fetal outcomes with even modest hyperglycemia (Table 229-8).

Diet and Lifestyle Management

Dietary recommendations for patients with type 2 diabetes have varied over the years and in the past included strict avoidance of sugars and the use

TABLE 229-8 RECOMMENDED GLYCEMIC TARGETS FOR ADULTS WITH DIABETES

	FASTING GLUCOSE LEVEL	POSTPRANDIAL GLUCOSE LEVEL	HbA_{1c}
Nonpregnant adults	70-130 mg/dL	<180 mg/dL	<7%
Gestational diabetes	≤95 mg/dL	<140 mg/dL (1 hour after meal) or <120 mg/dL (2 hours after meal)	—
Pre–gestational diabetes	60-99 mg/dL	100-129 mg/dL	<6%

Modified from American Diabetes Association. Standards of medical care in diabetes—2014. *Diabetes Care.* 2015;38(Suppl 1):S33-S40.

of specific diet plans (e.g., "exchange systems") that provided prescribed amounts of carbohydrate, fat, and protein. Current approaches for most patients focus on calorie restriction to achieve and to maintain modest (approximately 5 to 10% of body weight) weight loss, moderate carbohydrate intake, and avoidance of concentrated sweets and foods high in saturated fats and cholesterol. An optimal macronutrient distribution for patients with type 2 diabetes has not been established, and individualization of nutrition plans, depending on such factors as renal function, weight status, and the patient's preference, is recommended. Evidence suggests that a low-fat, low-carbohydrate (Atkins-type) diet and a Mediterranean-type diet can each be effective in promoting weight loss and improving glucose control in patients with diabetes. Moderate alcohol consumption is not prohibited, with the proper consideration of calorie intake (7 kcal/g) and hypoglycemia risk if alcohol is consumed without food, particularly in insulin-treated patients. Referral to a registered dietitian for medical nutrition therapy should be considered for patients newly diagnosed with type 2 diabetes and those who are not achieving glycemic or weight targets.

Regular exercise is an important but often overlooked component of diabetes management. Both aerobic exercise and resistance training can improve blood glucose control, even in the absence of significant weight change. Current recommendations are for a minimum of 150 minutes per week of moderate-intensity physical activity, such as brisk walking, biking, or swimming, and muscle-strengthening exercises two or three times per week. Assessment of cardiovascular status before beginning of an exercise program should be considered for selected patients, but routine screening (e.g., with an exercise stress test) of asymptomatic patients is not recommended. The presence of some diabetic complications may require restriction of certain activities. For example, in patients with proliferative retinopathy, vigorous aerobic or resistance exercise could precipitate retinal hemorrhage or detachment. The presence of significant sensory loss due to peripheral neuropathy can increase the risk of foot injury, including skin ulceration and Charcot joint destruction. Use of proper footwear and careful foot inspection are recommended, and avoidance of weight-bearing exercise may be required for high-risk patients. Finally, exercise-induced hypoglycemia can occur in patients treated with insulin or some secretagogues (i.e., sulfonylureas) and may require adjustment of the medication regimen or added carbohydrate before exercise.

Bariatric Surgery

Weight loss is considered the cornerstone for treatment of patients with type 2 diabetes, the majority of whom are overweight or obese. It has been clearly shown to improve glucose control. As expected, diabetic patients who undergo bariatric (weight reduction) surgery (Chapter 220) also show improvement in glucose control, which in some cases is dramatic. Improvements in glycemia often occur almost immediately after surgery, before significant weight loss has occurred, and appear to be related to changes in gut hormones (including GLP-1 and GIP) or bile acid metabolism. Many patients are able to discontinue diabetes medications, and diabetes remission rates above 50% have been reported. Among obese patients with uncontrolled type 2 diabetes, 3 years of intensive medical therapy plus bariatric surgery was reported to result in glycemic control in significantly more patients than did medical therapy alone; bodyweight, use of glucose-lowering medications, and quality of life also showed more favorable results at 3 years in the surgical groups.

Pharmacologic Therapy

Although weight control and nutrition form the foundation of effective management, most patients with type 2 diabetes will require use of pharmacologic agents, often multiple, to maintain recommended levels of glycemic control. During the last two decades, several new classes of drugs, targeting different metabolic pathways, have become available for treatment of type 2 diabetes. However, some of the most effective drugs are the oldest, and the long-term safety profile of newer agents remains to be established. Medications can be broadly categorized as those that enhance insulin availability (insulin and insulin secretagogues), those that enhance insulin action, or a miscellaneous group with other targets. Insulin therapy is also covered later and in the section on type 1 diabetes.

Insulin Sensitizers

Metformin

The biguanide drug metformin is the most widely used antidiabetic medication and is considered preferred initial therapy for patients with type 2 diabetes. The pleiotropic effects of metformin are thought to be mediated primarily through inhibition of mitochondrial complex 1 (i.e., effects on mitochondrial oxidative phosphorylation and cellular energy charge) and, in part, through regulation of the activity of 5′-adenosine monophosphate–activated protein kinase and the mammalian target of rapamycin. Metformin lowers glucose levels primarily through suppression of hepatic glucose production, but it also may enhance insulin sensitivity (improved insulin-mediated glucose uptake) and limit intestinal glucose absorption. Modest and sustained weight loss (~2 to 4 kg) is common with metformin.[11] Metformin is used orally twice a day, and extended-release forms are available for once-daily dosing. Hypoglycemia occurs rarely if at all with metformin monotherapy. The most common adverse effect is gastrointestinal intolerance (dyspepsia, diarrhea), which can be minimized by slow upward dose titration. Vitamin B_{12} malabsorption, leading to clinical B_{12} deficiency, has also been reported. The occurrence of lactic acidosis is the most serious although rare adverse effect, which occurs almost exclusively in patients with renal insufficiency and another precipitating factor, such as sepsis or shock. Renal function should be monitored periodically; metformin must be used with caution in those with an estimated glomerular filtration rate (GFR) of 45 mL/minute or lower and should be discontinued for an estimated GFR of 30 mL/minute or lower. Unique among available antidiabetic therapies, metformin was shown to reduce cardiovascular and all-cause mortality in the U.K. Prospective Diabetes Study (UKPDS), which adds to its appeal as a first-line agent. Metformin has also been used for diabetes prevention and for treatment of polycystic ovary syndrome.

Thiazolidinediones

The thiazolidinediones, which include rosiglitazone and pioglitazone, improve insulin-mediated glucose uptake and reduce hepatic glucose production. They bind to a nuclear receptor, peroxisome proliferator-activated receptor γ, and thus regulate the transcription of a variety of genes involved in carbohydrate and lipid metabolism. Thiazolidinedione therapy has pronounced effects on adipose tissue, reducing lipolysis, increasing fat mass, and causing redistribution of fat away from visceral to subcutaneous depots. Increases in circulating adiponectin, an adipokine with insulin-sensitizing and anti-inflammatory properties, may also play a role in the glucose-lowering effect of these drugs. Thiazolidinediones are given orally in once-a-day dosing. Common adverse effects include weight gain and fluid retention, including precipitation or worsening of congestive heart failure. Also reported have been an increase in fractures in postmenopausal women and increased risk of bladder cancer. The potential cardiovascular toxicity of rosiglitazone remains controversial, and its use has been restricted in many countries; these effects have not been observed for pioglitazone.

Insulin Secretagogues

Sulfonylureas

The sulfonylurea class of insulin secretagogues is among the oldest available oral antidiabetes drugs. Sulfonylureas currently in common use include glipizide, glyburide, and glimepiride; older sulfonylureas (chlorpropamide, tolbutamide) are still sometimes used outside of the United States. Their mechanism of action is to bind to the ATP-sensitive potassium channel in the β-cell membrane (at a site termed the sulfonylurea receptor), resulting in membrane depolarization and, ultimately, release of insulin from preformed secretory granules. Therefore, the presence of a sufficient mass of intact β cells is required for efficacy of these drugs. They can be used as monotherapy or in combination with other drugs. The major adverse effect of sulfonylureas is their potential to cause hypoglycemia because insulin secretion occurs regardless of ambient plasma glucose. Modest weight gain is also common. Results of a study conducted in the 1970s (University Group Diabetes Program) suggested that sulfonylurea drugs may increase the risk of cardiovascular events and mortality. These findings were not confirmed in other trials, but the issue remains controversial. Despite this, sulfonylureas are among the most widely used antidiabetic medications.

Glinides

Repaglinide and nateglinide are chemically distinct non-sulfonylurea insulin secretagogues that also bind to the ATP-sensitive potassium channel in the β-cell membrane. Their onset and duration of action are much shorter than those of sulfonylureas, and the frequency of fasting hypoglycemia may be less. They are administered orally before each meal, making them

somewhat less convenient than medications with a single daily dose but potentially providing an advantage for patients with inconsistent meal timing or content.

Incretin-Based Therapies/GLP-1 Agonists

Exenatide and liraglutide are analogues of the endogenous incretin hormone GLP-1 and stimulate insulin secretion by binding to GLP-1 receptors on β cells. These drugs augment glucose-stimulated insulin secretion and thus have less potential to cause hypoglycemia than sulfonylureas and glinides do. They also suppress hepatic glucose production (by reduction of glucagon secretion), delay gastric emptying, and suppress appetite, resulting in modest weight loss for many patients. GLP-1 agonists are given by injection once or twice a day, and a weekly long-acting formulation is also available. Major adverse effects include gastrointestinal intolerance (nausea and vomiting), which can be minimized by initiation with a low dose and gradual titration. An increased risk of acute pancreatitis has been reported with GLP-1 agonists (and DPP-4 inhibitors; see later), but the magnitude of the risk is uncertain and requires additional research. An increase in C-cell hyperplasia and medullary thyroid cancer was found in laboratory animals, although the relevance of this to humans is unclear.

Incretin-Based Therapies/DPP-4 Inhibitors

Inhibitors of dipeptidyl peptidase 4 (DPP-4), a ubiquitous serine protease, work by preventing the breakdown of endogenous GLP-1, thus prolonging its effects. DPP-4 inhibitors, including sitagliptin, saxagliptin, and linagliptin, are given orally in a single daily dose. Similar to GLP-1 agonists, they rarely cause hypoglycemia, but they are generally weight neutral and cause fewer gastrointestinal side effects. Concern about potential risk of pancreatitis and medullary thyroid cancer has also been raised but unconfirmed.

Other Pharmacologic Agents

SGLT2 Inhibitors

Canagliflozin and dapagliflozin are inhibitors of the sodium glucose cotransporter 2 (SGLT2) in the proximal renal tubule. This inhibition prevents the reabsorption of filtered glucose and results in glycosuria, which is accompanied by mild osmotic diuresis and modest weight loss. Experience to date with SGLT2 inhibitors is limited, and little is known about long-term efficacy and toxicity. The most common adverse effect is an increase in mycotic genital infections; hyperkalemia, urinary tract infections, and reductions in blood pressure have also been reported.

α-Glucosidase Inhibitors

Acarbose and miglitol are inhibitors of α-glucosidase enzymes in the intestinal lumen that are required for the breakdown and absorption of complex carbohydrates. Use of α-glucosidase inhibitors slows carbohydrate absorption in the small intestine, thus delaying the systemic delivery of glucose in the postprandial period and allowing better coordination with sluggish endogenous insulin secretion. α-Glucosidase inhibitors are given with meals, are weight neutral, and do not cause hypoglycemia when used as monotherapy. The major limiting factor for α-glucosidase inhibitor use is gastrointestinal side effects, predominantly flatulence and bloating due to the action of colonic bacteria on intestinal contents. Extremely slow dose titration may overcome this effect, but many patients cannot or will not tolerate these symptoms.

Pramlintide

Pramlintide is an analogue of the peptide amylin, which is co-secreted from the β cell along with insulin. The primary effects of pramlintide are to suppress hepatic glucose production in the postprandial period and to delay gastric emptying. Pramlintide is given by injection at mealtime and is approved for use in insulin-treated patients. Mild nausea and anorexia are common, which results in modest weight loss in some patients.

Bromocriptine

The dopamine agonist bromocriptine has been in use for many years as a therapy for Parkinson's disease and hyperprolactinemia. It was observed to have a glucose-lowering effect and has been approved as an antidiabetic drug. Its mechanism of action is thought to be through reduction in sympathetic nervous system activation, resulting in lower rates of hepatic glucose production and lipolysis. Its glucose-lowering effect is modest, and adverse effects (nausea, dizziness, weakness) are common.

Colesevelam

The bile acid–binding resin colesevelam was originally approved as a cholesterol-lowering drug and was also discovered to have a glucose-lowering effect. The mechanism of its antidiabetic action is not well understood but may involve enhanced availability of incretin hormones, including GLP-1. Colesevelam treatment requires two- or three-times-daily dosing and is accompanied by gastrointestinal side effects, notably constipation. The drug has a modest low-density lipoprotein (LDL)–lowering effect, but increases in triglycerides are common and may be limiting for some patients.

Insulin Therapy

Insulin treatment can be considered for patients with type 2 diabetes at any point in the course of the disorder, although typically it is used after "failure" of oral or other noninsulin therapies. Insulin may also be preferred therapy in specific situations, such as during hospitalizations (especially in the perioperative period) or in pregnancy. In contrast to patients with type 1 diabetes, patients with type 2 diabetes may be adequately controlled with basal insulin alone or in combination with other antidiabetic medications. Basal insulin is frequently used in combination with oral medications (e.g., metformin, DPP-4 inhibitors) or GLP-1 agonists (exenatide, liraglutide). However, reflecting the heterogeneity of type 2 diabetes, some patients may require physiologic insulin replacement similar to that used in type 1 diabetes, and insulin pump therapy is a safe and valuable option in patients who otherwise require multiple daily injections. Daily insulin requirements tend to be higher for patients with type 2 compared with type 1 diabetes, reflecting the existence of insulin resistance. Information about available insulin preparations and insulin regimens is provided in the section on insulin therapy under type 1 diabetes earlier and in Table 229-6.

Treatment Algorithms

Use of multidrug regimens is common in type 2 diabetes, and algorithms have been developed to guide therapy; however, the current evidence base to support these recommendations is limited. There is general agreement that metformin should be initial therapy for most patients and that subsequent drugs (when needed) are added to but do not replace metformin. The choice of a specific drug combination is driven by a number of factors, including efficacy, cost, side effect profile (e.g., hypoglycemia, weight gain), and preference of the patient (Fig. 229-5).

Metabolic Monitoring

Ongoing assessment of glycemic control is necessary to ensure optimal outcomes in patients with diabetes. Measurement of HbA_{1c}, which reflects mean glucose levels during the preceding 2- to 3-month period, should be performed routinely in all patients with diabetes, beginning at diagnosis and periodically thereafter. Quarterly tests should be done for patients whose therapy has been recently changed or who are not meeting glycemic goals. More stable patients can be tested twice a year. Self-monitoring of blood glucose levels is recommended for all patients using insulin and may be useful for any patient trying to achieve target glucose control. Monitoring for the development of vascular complications is addressed in the later section on chronic vascular complications.

Inpatient Management

Management of blood glucose levels during hospitalization is increasingly recognized as an important clinical issue, especially because 40 to 70% of hospitalized patients carry a concomitant diagnosis of diabetes. Frequently, diabetes is not the reason for admission, and attention to glucose management is secondary to other more critical medical problems. However, both hyperglycemia and hypoglycemia are associated with adverse outcomes in hospitalized patients, which has stimulated the development of algorithms and guidelines for inpatient glucose management, although the evidence base to support them is limited. Obviously, all patients with type 1 diabetes require continued insulin use during hospitalization.

Critically Ill Patients

After initial enthusiasm for intensive glucose control (maintenance of near-normoglycemia) for critically ill patients, more recent evidence suggests that it may be harmful, particularly when it is accompanied by hypoglycemia. Current guidelines recommend intravenous administration of insulin for critically ill patients in intensive care settings, with a goal of maintaining plasma glucose concentration between 140 and 180 mg/dL.[12] Application of standardized infusion protocols, which include frequent glucose monitoring, is recommended.

Non–Critically Ill Patients

The evidence base to support specific treatment guidelines for non–critically ill hospitalized patients is weak because this has not been systematically studied in randomized trials. However, there is agreement that subcutaneous administration of insulin is the preferred therapy to control glucose for most hospitalized (non–critically ill) patients with diabetes. Generally accepted targets are below 140 mg/dL for fasting glucose concentration and below 180 mg/dL for random or postprandial glucose concentration, if this can be achieved with minimal hypoglycemia risk. The patient's status needs to be reassessed frequently and insulin doses adjusted as needed to maintain target glucose levels. Use of basal insulin (see Table 229-6) is sufficient for many type 2 patients, but some may require the addition of prandial or corrective doses of short-acting insulin. However, prolonged dependence on insulin "sliding scales" to manage hyperglycemia should be avoided as this is rarely successful and carries increased risk of hypoglycemia. For stable patients who are eating consistent meals and those nearing hospital

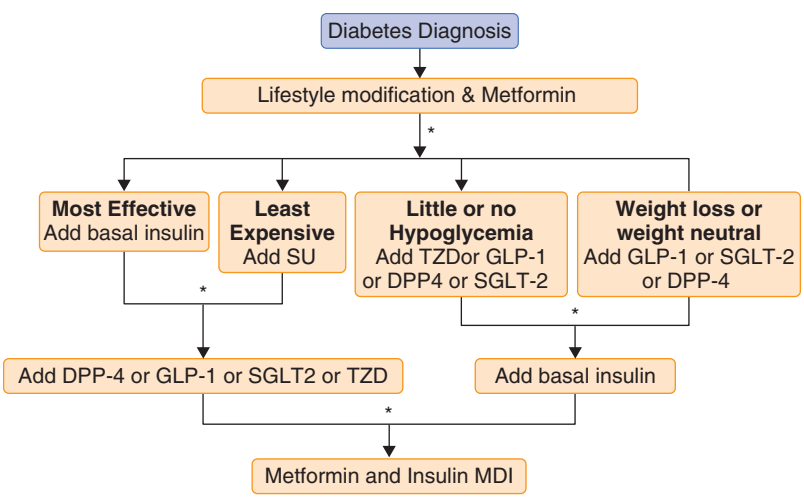

FIGURE 229-5. Algorithm for pharmacologic treatment of type 2 diabetes. DPP-4 = dipeptidyl peptidase 4; DPP-4-i = DPP-4 inhibitor; GLP-1 = glucagon-like peptide 1; GLP-1-RA = GLP-1 receptor agonist; HbA_{1c} = glycosylated hemoglobin; TZD = thiazolidinedione.

discharge, resumption of their usual oral or noninsulin injectable medications can be considered. Most patients with type 1 diabetes can be managed with their usual insulin injection regimen during hospitalization, but extra attention should be paid to hypoglycemia risk due to missed or delayed meals. Insulin pump therapy can be continued during hospitalization if the patient is able to direct its use and hospital personnel are sufficiently familiar with this form of treatment.

PREVENTION OF TYPE 2 DIABETES

The substantial burden, both human and societal, that accompanies type 2 diabetes and the difficulty in treating it effectively once it has developed make it an appropriate target for prevention. Further, the existence of a defined state of increased risk, pre-diabetes (i.e., impaired glucose tolerance and impaired fasting glucose), allows identification of patients who are most likely to benefit. Interventions that have been studied to date include lifestyle change (i.e., weight loss and exercise) and several antidiabetic medications.

Lifestyle Change

The largest and longest diabetes prevention study to date was the Diabetes Prevention Program, conducted in the United States beginning in the 1990s. Individuals at high risk for type 2 diabetes on the basis of the presence of overweight or obesity and pre-diabetic hyperglycemia (fasting glucose concentration of 95 to 125 mg/dL and 2-hour glucose concentration of 140 to 199 mg/dL) were randomly assigned to an intensive lifestyle program or a medication arm (metformin vs. placebo) and observed for a mean of 3 years. The lifestyle intervention stressed modest weight reduction (minimum 7% of body weight) with a reduced fat, hypocalorie diet and moderate-intensity physical activity for 150 minutes/week. Incident diabetes (determined by oral glucose tolerance test) was reduced by 58% compared with placebo, although the risk reduction was somewhat diminished (34%) with longer-term follow-up of the cohort. Successful weight loss was the major predictor of diabetes prevention, with every kilogram of weight loss reducing diabetes risk by 16%. Similar findings were reported from other studies, including the Finnish Diabetes Prevention Study. Even among individuals who did not lose weight, achieving the physical activity goal was associated with lower diabetes risk.

Medication

Several classes of antidiabetic drugs have been studied for diabetes prevention, including metformin, which reduced the risk of diabetes by 31% in the Diabetes Prevention Program. In smaller studies, the α-glucosidase inhibitor acarbose showed modest reduction in diabetes risk (~25%). The thiazolidinediones (e.g., troglitazone and rosiglitazone) have also shown diabetes prevention effects but are not widely used for this purpose because of concerns about their long-term safety. None of these drugs is approved by the Food and Drug Administration for diabetes prevention.

Recommendations

Lifestyle modification and metformin can both be recommended for individuals at high risk of diabetes. Candidates for prevention include those with defined glucose abnormalities (impaired glucose tolerance, impaired fasting glucose) and those with overweight or obesity plus an additional risk factor, such as family history of diabetes. The curriculum for the lifestyle intervention used in the Diabetes Prevention Program is available online (http://www.bsc.gwu.edu/dpp/lifestyle/dpp_part.html) and has been widely implemented in community settings, including the YMCA. Both lifestyle modification and metformin have shown positive effects on cardiovascular risk factors, but whether interventions to prevent diabetes will result in lower rates of microvascular or macrovascular complications remains to be determined.

Screening for Type 2 Diabetes

Individuals with risk factors for type 2 diabetes should be considered for screening for diabetes and impaired glucose regulation. This is especially important given that hyperglycemia can be present for years without specific symptoms and up to 30% of people with diabetes in the United States are undiagnosed. Screening will also allow identification of people with pre-diabetes, who may benefit from prevention interventions (Table 229-9).

Diabetes screening may be conducted with HbA_{1c} level, fasting glucose concentration, or oral glucose tolerance test, with the choice of test depending on the clinical setting and the preference of the patient. Screening should also be considered for asymptomatic children with BMI above the 85th percentile for age and sex plus any two of the following risk factors: family history of type 2 diabetes, high-risk race/ethnicity, or evidence of insulin resistance or features associated with insulin resistance (acanthosis nigricans, hypertension, dyslipidemia, polycystic ovary syndrome, or small-for-gestational-age birth weight). Pregnant women with risk factors for diabetes should be screened for undiagnosed diabetes at the first prenatal visit. Otherwise, a 75-g oral glucose tolerance test should be performed at 24 to 28 weeks of gestation to detect gestational diabetes.

PROGNOSIS

Type 2 diabetes is a chronic and, in most cases, progressive condition with potentially serious health consequences. However, it is also uniquely sensitive to modification of nutritional and lifestyle factors, which has been shown to be effective for both prevention and treatment of diabetes. Further, several classes of effective antihyperglycemic medications are available. There is substantial evidence that early intervention with a multifactorial approach to achieve and to maintain metabolic control, plus aggressive control of CVD

TABLE 229-9 CRITERIA FOR DIABETES SCREENING IN ASYMPTOMATIC ADULTS

Testing should be considered for any adult who is overweight (BMI >25 kg/m^2) and has at least one additional risk factor:
- Physical inactivity
- First-degree relative with type 2 diabetes
- High-risk race/ethnicity (e.g., African American, Latino, Native American, Asian American, Pacific Islander)
- Women with history of gestational diabetes or who delivered a baby weighing > 9 lb
- Hypertension
- HDL cholesterol < 35 mg/dL or triglyceride level > 250 mg/dL
- Women with polycystic ovary syndrome
- Clinical conditions associated with insulin resistance (e.g., severe obesity, acanthosis nigricans)

In the absence of the above risk factors, testing should begin at the age of 45 years. If results are normal, screening should be repeated at least at 3-year intervals.

BMI = body mass index; HDL = high-density lipoprotein.
Modified from American Diabetes Association. Standards of medical care in diabetes—2014. *Diabetes Care.* 2015;38(Suppl 1):S8-S16.

TABLE 229-10 CLASSIFICATION OF IATROGENIC HYPOGLYCEMIA IN TREATED DIABETIC PATIENTS

	CLINICAL FEATURES
Severe hypoglycemia	Episode with neurocognitive impairment that requires another person to administer treatment
Documented symptomatic hypoglycemia	Measured glucose concentration ≤70 mg/dL that coincides with sympathoadrenal or neurologic symptoms. Episode is self-managed.
Asymptomatic hypoglycemia	Measured glucose concentration ≤70 mg/dL, but without concomitant symptoms. Absence of symptoms may be due to hypoglycemia unawareness or hypoglycemia-associated autonomic failure.
Pseudo-hypoglycemia	Typical hypoglycemia symptoms, but with measured glucose concentration >70 mg/dL. Symptoms may be caused by resetting of counter-regulatory system in the setting of chronic poor glucose control.

risk factors, will substantially reduce the burden of diabetes complications and improve quality of life.

ACUTE METABOLIC COMPLICATIONS OF DIABETES

Hypoglycemia

Iatrogenic hypoglycemia in people with diabetes is the most frequent cause of low blood glucose concentration. Hypoglycemia (Chapter 230) affects the daily lives of persons with diabetes and can have a dramatic effect on quality of life. It can induce great fear, preclude comfortable engagement in routine activities (e.g., driving, uninterrupted sleep), and lead both patient and clinician to set higher glycemic targets and hence worse metabolic control. Thus, hypoglycemia continues to be a major limiting factor in the treatment of diabetes with most medications, particularly the use of insulin.[13]

Whereas insulin-stimulatory drugs (e.g., sulfonylureas) and parenteral insulin are the primary causes of drug-induced iatrogenic hypoglycemia, underlying defects in some parts of the counter-regulatory cascade contribute to the greater frequency and potential morbidity and mortality of hypoglycemia among patients with diabetes. The normal counter-regulatory response to hypoglycemia and the typical adrenergic and neuroglycopenic hypoglycemia symptoms are described in Chapter 230.

The threshold plasma glucose value that results in hypoglycemic symptoms is not constant; it is lower after recent antecedent hypoglycemia and higher in patients with poor glycemic control. However, there is general consensus that a self-monitored glucose level of 70 mg/dL or lower is a value that should alert the patient or caregiver, regardless of the presence of symptoms. A more detailed classification system to describe hypoglycemia has been established and widely adopted in research settings (Table 229-10).

However, these distinctions are not commonly used in clinical practice, and the severity of symptoms is often confused with severity of the actual prevailing physiologic state. Thus, a patient may feel intense symptoms at a glucose level of 50 to 60 mg/dL, for which there is no evidence of cognitive impairment or imminent danger, whereas potentially dangerous plasma glucose levels in the range of 20 to 40 mg/dL might go unappreciated owing to lack of classical symptoms. This also has implications for the epidemiology of hypoglycemia; most studies have reliably ascertained only the rates of severe hypoglycemia because other episodes are less likely to be documented. In type 1 diabetes, the DCCT reported 62 severe hypoglycemic episodes per 100 patient-years, although the actual risk may be higher in clinical settings. An episode of severe hypoglycemia can be the immediate cause of death in patients with type 1 diabetes, with recently reported mortality rates ranging from 4 to 10%. There remains uncertainty about the temporal relationship between hypoglycemia and death, and although prolonged episodes of very low circulating glucose (<15 mg/dL) can cause brain death, episodes of fatal hypoglycemia may be due to other mechanisms, such as ventricular arrhythmias. Episodes of severe hypoglycemia are much less common in patients with type 2 diabetes (see later).

In patients with treated diabetes, the initiation of the hypoglycemic event is due to mismatching of prevailing insulin levels to the underlying physiologic state of the individual. Thus, even absent overt insulin overdosage, factors such as missed meals, exercise, recent weight loss, alcohol, or insulin-sensitizing drugs create this mismatch and may set the plasma glucose concentration on a downward trajectory. In addition, the counter-regulatory systems that normally would counteract the decline of glucose to dangerous levels may be impaired. In patients with type 1 diabetes, glucagon release during hypoglycemia may become impaired shortly after the onset of diabetes, although glucagon is still secreted in response to other secretagogues, suggesting the presence of a functional defect. Epinephrine release during hypoglycemia also becomes progressively defective in type 1 diabetes; it is not triggered until the plasma glucose level is lower, and the maximal concentration of epinephrine released is significantly reduced. This decrease in epinephrine response during hypoglycemia is accompanied by an attenuated autonomic neural response, which results in the clinical syndrome of *impaired awareness of hypoglycemia*. Without autonomic symptoms, mild hypoglycemia may proceed unnoticed to more advanced and dangerous phases. Patients who have both impaired awareness of hypoglycemia and defective counter-regulation are at the greatest risk for development of severe hypoglycemia.

Hypoglycemia-associated autonomic failure in type 1 diabetes apparently results from antecedent episodes of mild hypoglycemia that further degrade the counter-regulatory response. In experiments in people without diabetes, recurrent or recent episodes of hypoglycemia are associated with reduced autonomic (epinephrine and norepinephrine), symptomatic, and cognitive functional responses to subsequent episodes of hypoglycemia, impairing the endogenous defense mechanisms and the clinical signs required for hypoglycemia detection. Because patients with type 1 diabetes already have a reduced counter-regulatory response, hypoglycemia-associated autonomic failure may play a role in the vicious circle of hypoglycemia begetting hypoglycemia. Meticulous avoidance of hypoglycemia is the only current approach proven to improve the epinephrine response and to reverse impaired awareness of hypoglycemia.

Compared with type 1 diabetes, type 2 diabetes is associated with a much lower risk of hypoglycemia. However, hypoglycemia remains a major clinical problem in this population. Episodes of severe hypoglycemia become progressively more common in patients with longer duration of type 2 diabetes, due in part to progressive β-cell failure and increased dependence on pharmacologic treatments. Use of sulfonylureas accounts for a substantial proportion of cases of drug-induced hypoglycemia, and severe episodes characterized by coma have been reported with all the agents in common use. The hypoglycemic potential of an agent is related to its potency, its plasma and biologic half-lives, its metabolism, and the concomitant use of other drugs. For example, liver disease prolongs the hypoglycemic actions of glyburide and glipizide because these drugs are partially metabolized in the liver. Similarly, renal disease may prolong the action of insulin (due to impaired insulin clearance) or potentiate the effects of hypoglycemic drugs by other mechanisms. Other antidiabetic agents, such as metformin, thiazolidinediones, and incretin-based drugs, have been associated with measureable albeit lower risks of hypoglycemia; however, symptomatic hypoglycemia is rare unless these drugs are used in combination with insulin. The elderly are at

particularly high risk for iatrogenic hypoglycemia because the intensity of adrenergic symptoms may be reduced and hypoglycemia-induced cognitive impairment greater.

CLINICAL APPROACH TO HYPOGLYCEMIA PREVENTION AND TREATMENT

Rx

Patients with diabetes need to be well informed about the symptoms of hypoglycemia and the factors that predispose to its occurrence: meal timing and content, exercise, and the expected time course of the drugs in use (especially insulin). Patients should also be made aware that the accuracy of some home glucose meters may be reduced in the hypoglycemia range and that the typical sympathoadrenal symptoms may wane during years of diabetes. A history of recurrent hypoglycemia should be carefully evaluated and attempts made to determine whether the patient had experienced events that went unrecognized. For example, reports of unexplained night sweats or a clouded mental state on arising in the morning may be due to nocturnal hypoglycemia and should be investigated. Patients should be encouraged to document episodes of hypoglycemia and to contact the care team if they have unexpected or more frequent episodes.

Table 229-11 lists several risk factors for severe hypoglycemia. Patients with these characteristics require greater vigilance, both in selection of treatment regimen and in the recognition and treatment of acute episodes.

Most mild or moderate episodes of hypoglycemia can be self-treated by ingestion of fast-acting carbohydrates such as glucose tablets, glucose gels, or food (juices, soft drinks, or a meal). The suggested amount of carbohydrate to be ingested is about 15 g, which will increase the plasma glucose concentration by about 15 mg/dL. Importantly, foods that are rich in fat delay glucose absorption and are thus less effective. If plasma glucose levels are still below 70 mg/dL and if symptoms have not abated after 15 minutes, the patient should take an additional 15 g of carbohydrate. Because the glycemic response to oral glucose is relatively transient, ingestion of a snack or a meal shortly after correction of hypoglycemia is recommended.

Parenteral treatment of hypoglycemia is recommended if the patient is unwilling or unable to ingest carbohydrates (e.g., due to impaired mental status) or if a patient has sulfonylurea- induced hypoglycemia (which may be prolonged). Intravenous administration of glucose (25 g) is the preferred treatment of hypoglycemia. Parenteral glucagon (1 mg subcutaneously) is an alternative, especially in patients with type 1 diabetes who may have to be treated by family members for severe hypoglycemia. Because glucagon stimulates secretion of insulin in addition to promoting glucose production, it is less effective in patients with type 2 diabetes.

Nocturnal hypoglycemia may be a particular problem for patients with type 1 diabetes. It may be asymptomatic and unsuspected because plasma glucose concentration is rarely measured during the night. Risk factors for nocturnal hypoglycemia include increased physical activity in the last 24 hours, certain insulin regimens (e.g., use of NPH or regular insulin), meal content (e.g., the amount of fat), and alcohol consumption. In addition, sleep is associated with a decrease in the autonomic response to hypoglycemia. Currently, the only practical approaches to detection of nocturnal hypoglycemia are regular nocturnal (3 AM) self-monitoring or the use of continuous glucose monitors with alarm features. Some patients with nocturnal hypoglycemia present with sleep disturbances, morning headache, chronic fatigue, or depression. Children in particular may present with seizures or enuresis. Strategies to prevent nocturnal hypoglycemia include eating "long-acting" bedtime snacks (slowly absorbed carbohydrate, such as uncooked cornstarch) and regular monitoring of blood glucose concentration at bedtime. The bedtime glucose level has been reported to be highly predictive of subsequent development of hypoglycemia during sleep.

TABLE 229-11 RISK FACTORS FOR SEVERE HYPOGLYCEMIA IN PATIENTS WITH DIABETES

Youth (children)
Elderly taking sulfonylurea drugs or insulin
Altered consciousness
Ethanol use
Strenuous exercise in the previous 24 hours
Recent antecedent hypoglycemia
Use of pentamidine, quinine, or nonselective β-blocker drugs
Concomitant illnesses, such as sepsis, or hepatic, renal, or cardiac failure
Type 1 diabetes with history of recurrent severe hypoglycemia
Recent rapid improvement in HbA_{1c} into the normal range

Hyperglycemic States

DKA and HHS are the most serious acute hyperglycemic complications of diabetes. DKA is typically associated with severe insulin-deficient states (i.e., type 1 diabetes). It may also occur rarely in type 2 diabetes under conditions of extreme stress, such as major infection or trauma or as a presentation of a variant of type 2 diabetes (ketosis-prone or Flatbush diabetes). On the other hand, HHS typically occurs in patients with type 2 diabetes. However, the distinction between the two clinical scenarios is sometimes blurred (e.g., patients with HHS may present with ketosis and acidosis), and these states may be considered as parts of the spectrum of severe metabolic decompensation. Despite aggressive treatment, mortality rates remain high for both conditions, approaching 5% for DKA and 15% for HHS. Mortality is associated with the extremes of age (i.e., the very young and the elderly) and comorbidities but, importantly, with the severity of the precipitating illness or event. Thus, in addition to correction of fluid and electrolyte imbalance and administration of insulin, treatment also includes prompt recognition of and therapy for any precipitating illness or event. A list of precipitating conditions commonly associated with DKA and HHS is shown in Table 229-12.

PATHOBIOLOGY

The pathogenesis of DKA and HHS mirrors the underlying respective forms of diabetes. The three fundamental biochemical features of DKA—hyperglycemia, ketosis, and acidosis—result from the combined effects of deficient circulating insulin and counter-regulatory hormone excess. This hormonal milieu promotes the delivery of substrates from muscle (amino acids, lactate, pyruvate) and adipose tissue (free fatty acids, glycerol) to the liver, where they are converted to glucose or to ketone bodies (β-hydroxybutyrate, acetoacetate, acetone). Glucose and ketones are thus released into the circulation at greater rates than their utilization, resulting in severe hyperglycemia (>250 mg/dL), ketoacidosis (arterial pH <7.30), and an osmotic diuresis that promotes dehydration and electrolyte loss. In HHS, despite comparable elevations of glucagon, the presence of some endogenous insulin modulates the ketosis even though the plasma glucose concentration in HHS typically exceeds 600 mg/dL, whereas in DKA it is usually more than 250 mg/dL.

In both states, fluid depletion plays a major role in causing dramatic elevations in circulating glucose. Indeed, the hyperosmolality accompanying DKA and HHS is best linked to the patient's level of neural and cognitive function, and treatment of both conditions depends on restoration of fluid balance. Finally, other factors have been invoked, including other hormones (such as epinephrine, growth hormone, and cortisol), proinflammatory cytokines (such as tumor necrosis factor-α, interleukin-1β, interleukin-6, and interleukin-8), and lipid peroxidation markers as well as plasminogen activator inhibitor 1 and C-reactive protein. Whether all these factors are simply "stress markers" reflecting the disordered metabolic state or true pathogenetic factors remains uncertain.

TABLE 229-12 PRECIPITANTS OF DIABETIC KETOACIDOSIS AND HYPEROSMOLAR HYPERGLYCEMIC STATE

MOST COMMON
Inadequate insulin treatment or noncompliance
New-onset diabetes
Infections
Myocardial infarction
OTHER PRECIPITATING FACTORS
Cerebrovascular accident
Acute pulmonary embolism
Acute pancreatitis
Intestinal or mesenteric thrombosis
Alcohol intoxication
Endocrinopathies: Cushing syndrome, thyrotoxicosis, acromegaly
Severe burns, hyperthermia, hypothermia
Drugs: clozapine, olanzapine, cocaine, lithium, sympathomimetics, corticosteroids, thiazide diuretics

DIABETIC KETOACIDOSIS

CLINICAL MANIFESTATIONS

DKA (Chapter 118) may signal the onset of type 1 diabetes, but changes in medical practice in the developed world during the past several decades have enhanced earlier diagnosis of type 1 diabetes, and now the majority of childhood cases are detected and treated before ketoacidosis occurs. Thus, DKA is more frequently seen in those with established diabetes, usually in the setting of coexisting illness or poor adherence. For example, a patient may be unable to maintain adequate hydration during an illness, such as a viral gastroenteritis, and may mistakenly omit insulin because of inability to eat. A key component of a diabetes treatment program is education in "sick-day" rules focused on home-based prevention of DKA (e.g., frequent blood glucose monitoring, serum or urine ketone testing, fluid intake, determination of insulin dosing or delivery problems). Behavioral factors may also be involved; some younger patients may omit insulin deliberately to promote weight loss or to call attention to a dysfunctional home situation. This should be suspected in cases of recurrent episodes of DKA.

The clinical history of DKA typically involves deterioration during several hours to days, with progressive polyuria, polydipsia, and other symptoms of hyperglycemia. Other common clinical features are weakness, lethargy, nausea, and anorexia. Nonlocalizing upper abdominal pain in the setting of DKA can mimic an acute abdomen. Reduced motility of the gastrointestinal tract or, in severe cases, paralytic ileus may further contribute to diagnostic confusion. Nausea and vomiting are symptoms that indicate the need for in-hospital treatment because they preclude oral fluid intake. Physical findings in DKA are mainly secondary to dehydration, hyperosmolality, and acidosis; these include dry skin and mucous membranes, reduced jugular venous pressure, tachycardia, orthostatic hypotension, depressed mental function, and deep, rapid respirations (Kussmaul breathing).

DIAGNOSIS

In DKA, glucose levels may vary from modestly elevated to more than 1000 mg/dL, serum bicarbonate concentration drops below 18 mEq/L, and there is an excess anion gap that is generally proportional to the decrease in serum bicarbonate (Table 229-13). Hyperchloremia may be superimposed if the patient maintains an adequate GFR and is able to exchange keto acids for chloride in the kidney. The degree of depression of arterial pH depends largely on respiratory compensation. In mild cases, the pH may range from 7.20 to 7.30; in severe cases, it can fall below 7.00. On occasion, a degree of superimposed metabolic alkalosis (e.g., caused by vomiting or diuretic use) may obscure the true severity of the ketoacidosis. An anion gap out of proportion to the fall of bicarbonate should suggest this possibility. Other laboratory abnormalities commonly seen in DKA include a reduced measured serum sodium concentration (due to hyperosmolarity and the resulting osmotic shift of intracellular water into the intravascular space), prerenal azotemia, and elevated serum amylase. The last is usually of nonpancreatic origin and can lead to an erroneous diagnosis of pancreatitis. Normal, elevated, or reduced concentrations of potassium, phosphate, and magnesium may exist when DKA is diagnosed; however, large deficits of these electrolytes invariably accompany the osmotic diuresis and become readily apparent during the course of treatment. The serum triglyceride concentration is frequently elevated, a reflection of deranged lipid metabolism in the setting of insulin deficiency. The white blood cell count is typically elevated; the hemoglobin and hematocrit may be elevated, reflecting volume contraction.

Special care should be taken in interpreting serum or urine ketone results. Because quantitative measurements of β-hydroxybutyrate and acetoacetate are not readily available, rapid diagnosis usually requires qualitative assessment of serum ketones by the use of serum dilutions and reagent strips (e.g., Ketostix) or tablets (e.g., Acetest), which depend on a nitroprusside reaction with acetoacetate. However, acetone reacts weakly with nitroprusside, and β-hydroxybutyrate does not react at all; thus, the results of qualitative testing for ketones can be misleadingly low. Furthermore, because of the presence of intracellular acidosis, β-hydroxybutyrate levels are often much higher than acetoacetate levels, which may further conceal the true degree of ketoacidosis. Conversely, after insulin therapy is begun, the nitroprusside reaction may give the "false" impression of sustained ketoacidosis for hours or even days. This results because nonacidic acetone is slowly cleared from the circulation and also because, as acidosis improves, β-hydroxybutyrate is converted to acetoacetate, giving the false impression that ketosis is worsening.

TREATMENT

Rx

An overview of the treatment of DKA and HHS is shown in Figure 229-6.

In the early hours of treatment, the primary considerations are to restore intravascular volume, to correct tissue hypoperfusion, and to restore insulin sensitivity. With DKA, large total body deficits of water (5 to 10 L), sodium (5 to 10 mEq/kg), and other electrolytes may exist (Chapter 118). These losses are even more profound in HHS, which typically develops during a longer time. Although water loss usually exceeds the loss of sodium, it is almost always preferable to begin fluid replacement with isotonic normal saline (0.9% NaCl solution) for efficient intravascular volume restoration. Fluid replacement regimens vary, but it is common to administer 1 L of normal saline within the first hour, followed by a continuous infusion with either 0.45% NaCl or 0.9% NaCl, depending on the corrected serum sodium concentration, the patient's hemodynamic status, and the clinical assessment of tissue perfusion. Likewise, the rate of infusion (commonly 250 to 500 mL/hour) should be adjusted according to both biochemical responses and the age and clinical status of the patient (e.g., oliguria or underlying CVD). In children, isotonic solutions are generally preferred because they are less likely than hypotonic solutions to accelerate water shifts into the intracellular space and contribute to cerebral edema. As the blood glucose concentration falls below 250 mg/dL, dextrose should be added to intravenous fluids to avoid later insulin-induced hypoglycemia because continued insulin delivery may be required to correct the persistent acidemia.

Although insulin resistance is present in both DKA and HHS, supraphysiologic doses of insulin are unnecessary and are more likely to provoke hypokalemia, hypophosphatemia, and delayed hypoglycemia. A typical insulin replacement regimen uses an intravenous bolus of 0.15 U/kg of rapid-acting (e.g., regular) insulin, followed by 0.1 U/kg/hour thereafter. Intravenous administration is the most predictable way to deliver insulin to target tissues, particularly in severely hypovolemic patients with reduced peripheral blood flow. If intravenous administration is not possible, intramuscular or subcutaneous routes of administration can be used. It is ideal if blood glucose levels fall at a steady and predictable rate (50 to 75 mg/dL/hour), so it is important to monitor blood glucose levels hourly during insulin therapy to ensure an appropriate rate of decline. Blood glucose levels should not fall too rapidly, especially in young children, in whom accelerated correction of plasma glucose concentrations has been associated with cerebral edema.

After a stable blood glucose level of 150 to 250 mg/dL is achieved, with resolution of the anion gap acidosis, subcutaneous administration of insulin can be started and the intravenous insulin infusion discontinued. With DKA, it is important to overlap the intravenous and subcutaneous routes by at least 1 to 2 hours to avoid the rebound ketoacidosis if insulin levels drop precipitously. After stabilization, and with resumption of oral food intake, long-term

TABLE 229-13 DIAGNOSTIC CRITERIA FOR DIABETIC KETOACIDOSIS (DKA) AND HYPEROSMOLAR HYPERGLYCEMIC STATE (HHS)

CRITERION	MILD DKA	MODERATE DKA	SEVERE DKA	HHS
Plasma glucose concentration (mg/dL)	≥250	≥250	≥250	≥600
Effective serum osmolality (mOsm/kg)	Variable	Variable	Variable	≥320
Urine or serum ketones (nitroprusside reaction)	Positive	Positive	Positive	Negative to small
Arterial pH	7.25-7.30	7.00-7.24	<7.00	>7.30
Serum bicarbonate (mEq/L)	15-18	10-15	<10	>15
Anion gap (mEq/L)	>10	>12	>12	Variable, usually <12
Typical mental status	Alert	Drowsy	Stupor or coma	Stupor or coma

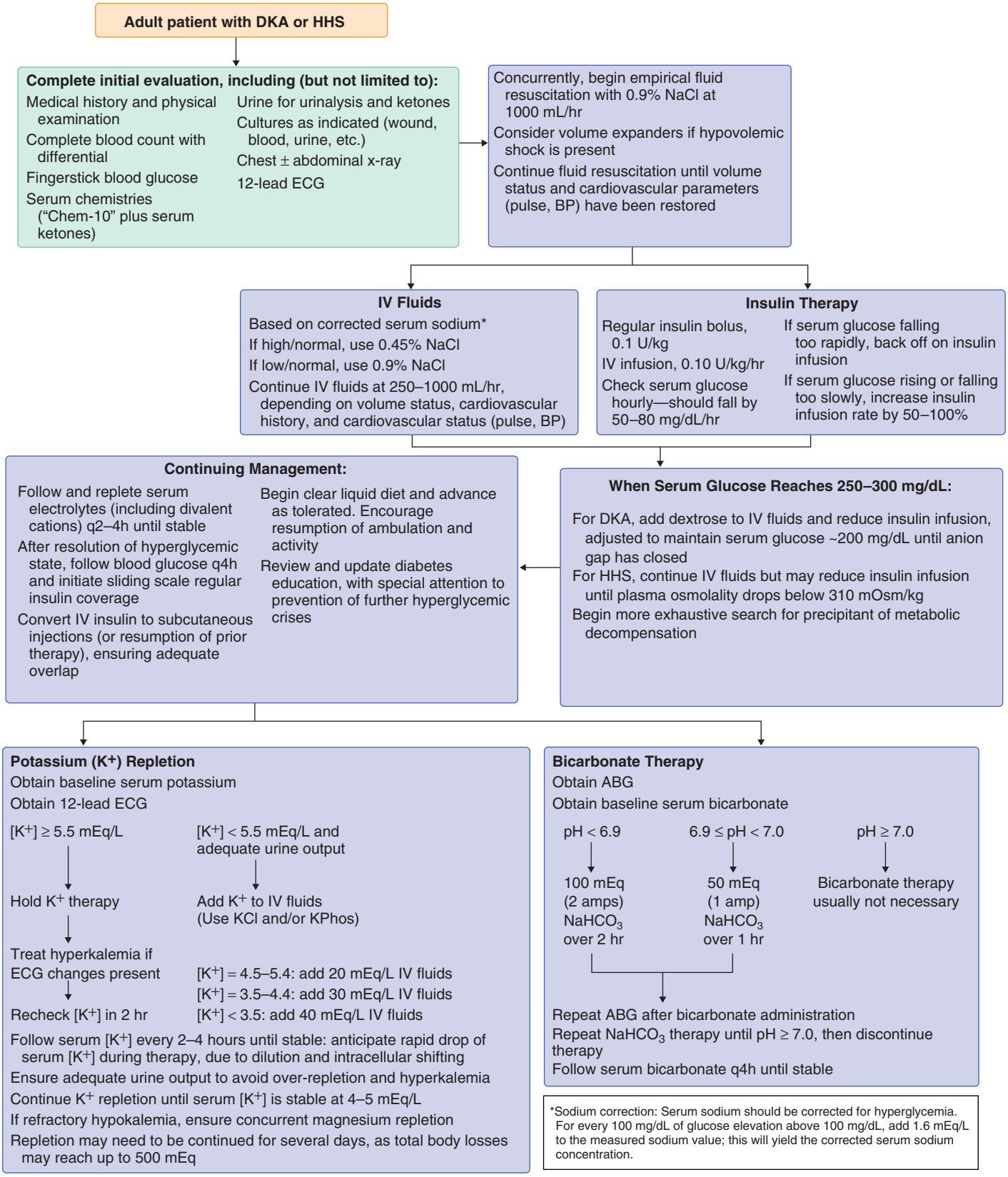

FIGURE 229-6. Management of diabetic ketoacidosis (DKA) and hyperosmolar hyperglycemic syndrome (HHS). ABG = arterial blood gas; BP = blood pressure; ECG = electrocardiogram; IV = intravenous.

medical management should be initiated (or resumed), with both long-acting and short-acting insulins, to approximate the desired outpatient regimen. A temporary "regular insulin sliding scale" should be avoided because such therapy is reactive to hyperglycemia and the swings in glycemia will not allow safe discharge of the patient. The eventual dosage and frequency of insulin depend on multiple factors, including body weight, comorbidity, insulin sensitivity, and effectiveness of prior therapeutic regimens.

Potassium replacement is usually required in DKA. Overt hypokalemia can result in muscle weakness, cramps, and nausea; both hyperkalemia and hypokalemia are associated with cardiac arrhythmias. Even absent severe hypokalemia, patients have a significant total body potassium deficit (about 3 to 7 mEq/kg), and measured serum potassium levels may be normal or high as acidosis and renal failure can mask the potassium deficiency. As insulin is infused, potassium will move into the intracellular space, further lowering

serum potassium to levels that may trigger life-threatening arrhythmias. In addition, fluid replacement causes extracellular dilution of potassium, leading to improved renal perfusion and increased urinary potassium excretion. Thus, potassium replacement should be initiated as soon as it is established that the patient is not in renal failure. A low potassium level (<3.5 mEq/L) requires prompt treatment with up to 40 mEq/hour, whereas "normal" serum levels (3.5 to 5.0 mEq/L) call for less aggressive repletion of potassium (10 to 30 mEq/hour), assuming adequate urine output. In patients who may have lost potassium for additional reasons, such as diuretic use or gastrointestinal loss, there will be need for greater potassium supplementation.

In the majority of patients with mild to moderate DKA, keto acids clear spontaneously with standard therapeutic measures, and correction of the pH with alkali (as bicarbonate) is unnecessary. Suppression of lipolysis by insulin reduces free fatty acid flux to the liver and blocks ketogenesis, and circulating keto acids are then cleared or oxidized, with subsequent regeneration of bicarbonate and restoration of arterial pH. However, in cases of severe acidosis (pH <6.9 to 7.0), bicarbonate administration may be indicated if the clinical picture dictates (e.g., hypotension that is unresponsive to fluids, cardiac dysfunction, respiratory exhaustion). Bicarbonate therapy should be used with caution and only at the minimal doses required to stabilize the patient because it can further provoke hypokalemia. In addition, by causing a sudden left shift of the dissociation curve for oxyhemoglobin, bicarbonate may impair oxygen delivery to the tissues. Therefore, if alkali therapy is given, small amounts should be administered slowly: 50 mEq of $NaHCO_3$ during 1 hour for arterial pH 6.9 to 7.0, and 100 mEq during 2 hours for pH below 6.9. After bicarbonate administration, arterial pH (and serum potassium levels) should be rechecked every 2 hours, and alkaline therapy should be discontinued when the pH rises above 7.0.

In the setting of DKA, phosphate losses average 3 to 7 mmol/kg; magnesium losses reach 1 to 2 mEq/kg. Phosphate is shifted extracellularly during hyperosmolar states, so initial serum levels may be falsely elevated and may drop rapidly during therapy. Complications of hypophosphatemia generally occur at serum levels below 1.0 mg/dL and include respiratory and skeletal muscle weakness, impaired cardiac systolic performance, and hemolytic anemia. Phosphate repletion should be used in patients with serum phosphate levels below 1.0 mg/dL and in patients with evidence of cardiac or respiratory compromise, hypoxia, or hemolytic anemia. An effective means of replacing phosphate is to replace one third to one half of the potassium losses (discussed previously) as potassium phosphate. In severe hypophosphatemia, cautious intravenous administration of additional small amounts of potassium phosphate may be necessary. Because of calcium binding, hypocalcemic tetany may complicate phosphate therapy unless magnesium supplements are also provided; for this reason, serum calcium, phosphate, and magnesium levels should be monitored during any phosphate infusion.

HYPEROSMOLAR HYPERGLYCEMIC SYNDROME

CLINICAL MANIFESTATIONS

The metabolic state formerly known as the hyperglycemic hyperosmolar nonketotic state or coma has been renamed the *hyperosmolar hyperglycemic syndrome* (HHS) to highlight two important points: (1) ketosis (and acidosis) may in fact be present to varying degrees in HHS, and (2) alterations in sensorium most commonly occur in the absence of coma. In fact, only 10% of HHS patients present with frank coma, and an equal percentage show no signs whatsoever of mental status change. Major risk factors for HHS include older age (most cases occur in patients aged 65 years and older) and impaired cognition (i.e., impaired ability to recognize thirst or to obtain access to water).

As shown in Table 229-13, the hallmarks of the HHS are severe hyperosmolarity (>320 mOsm/L) and hyperglycemia (>600 mg/ dL). Severe hyperglycemia occurs because patients cannot consume enough liquid to keep pace with a vigorous osmotic diuresis. The resulting impairment in renal function eventually further reduces glucose excretion through the kidney, leading to remarkable blood glucose elevations, sometimes exceeding 1000 mg/dL. In contrast to DKA, even though glucose concentrations are generally higher, severe acidosis and ketosis are usually absent in the HHS. This is probably explained by the presence of some residual insulin secretory capacity that is sufficient to suppress lipolysis and to avoid significant keto acid production. However, some type 2 patients with depressed endogenous insulin secretion may be unable to suppress ketone production fully in the face of elevated counter-regulatory hormones produced by physical illness. However, because HHS patients have higher portal vein insulin concentrations than do patients with DKA, keto acid production by the liver is quantitatively less, yielding only mild acidosis. In the HHS, in the absence of concurrent acid-base disturbances, arterial pH rarely drops below 7.30, and serum bicarbonate levels typically do not fall below 18 mEq/L.

In the HHS, clinical severity and levels of consciousness generally correlate with the severity and duration of hyperosmolarity. Clinical signs indicate profound dehydration; gastrointestinal symptoms are seen less frequently than in DKA. A variety of often reversible neurologic abnormalities may exist, including grand mal or focal seizures, extensor plantar reflexes, aphasia, hemisensory or motor deficits, and worsening of a preexisting organic mental syndrome. The laboratory picture is dominated by the effects of uncontrolled diabetes and dehydration; renal function is impaired, hemoglobin and hematocrit are elevated, and liver function test results may be abnormal because of baseline hepatic steatosis. Although severe hyperglycemia would be expected to lower measured serum sodium concentration, it is not uncommon to see normal or even elevated sodium levels because of the severity of dehydration. The serum osmolarity can be measured directly or estimated.

TREATMENT

Rx

The approach to treatment of HHS is similar to that of DKA and requires aggressive management of fluids and electrolytes (see Fig. 229-6). Importantly, patients with HHS tend to have more dramatic volume contraction, and by definition, acidosis is not present or is minimal in degree. It is important to volume resuscitate the patient adequately before insulin is administered because intracellular fluid shifts that occur as glucose levels are reduced may worsen systemic tissue perfusion. In fact, glucose levels usually drop substantially with hydration alone, in part because of improved renal perfusion, thus promoting glycosuria. Coadministration of dextrose along with insulin, as is recommended in patients with DKA to allow ketones to clear and the acidosis to resolve, is rarely required. Further, because recurrent acidosis is less of a concern, patients may be transitioned directly from insulin infusion to subcutaneous injections. Because altered mental status (and, in some cases, coma) is a frequent feature of HHS, attention should be paid to respiratory status and appropriate airway protection. A diligent search for underlying precipitating illness should be made, keeping in mind that the typical HHS patient is elderly and may well have overt or subclinical CVD. The presence of impaired cardiac function, also more common among the elderly, needs to be considered in the management of intravenous fluid resuscitation.

After resolution of the HHS episode, some patients may ultimately be able to be managed with oral agents alone. However, the development of HHS signifies a significant degree of insulin deficiency. As a consequence, it is always best to prescribe insulin injections before the patient is discharged and to reserve judgment about the appropriateness of using oral agents until the patient's progress can be monitored and reassessed in the outpatient setting.

CHRONIC VASCULAR COMPLICATIONS

EPIDEMIOLOGY

The major clinical burden associated with long-standing diabetes is the development of vascular disease, which includes characteristic microvascular complications (retinopathy, nephropathy, neuropathy) and accelerated medium- and large-vessel atherosclerosis. Diabetes is the leading cause of kidney failure, nontraumatic lower limb amputations, and new cases of blindness among adults in the United States. Diabetes is also a major cause of coronary heart disease, heart failure, and stroke and is the seventh leading cause of death in the United States. The microvascular complications are directly linked to hyperglycemia, with both the duration of diabetes and the degree of glucose elevation constituting the major risk factors. Other factors, including genetic susceptibility, smoking, and concomitant conditions like hypertension, also contribute to the risk of complications (Fig. 229-7). Diabetic microvascular complications occur in both type 1 and type 2 diabetes; given that most patients with type 1 diabetes develop it when younger, they may face greater lifetime risk of complications.

The central role for hyperglycemia in the development of diabetic complications was long suspected and ultimately confirmed by the landmark DCCT, which was reported in 1993.[14] In this study, 1441 adolescents and younger adults with type 1 diabetes were randomly assigned to conventional treatment designed to avoid symptomatic hypoglycemia or hyperglycemia (standard therapy at the time) or to an experimental treatment group designed to achieve near-normoglycemia. The experimental group received intensive management with multiple daily insulin injections or use of a continuous subcutaneous insulin pump; frequent self-monitored blood glucose determinations; and adoption of detailed algorithms to guide the patient in

determining insulin dosing in response to meals, glucose, and exercise. During the course of the study, mean HbA_{1c} levels were 7.2% in the intensive group compared with 9% for the conventional treatment group. The unequivocal DCCT results showed substantially lower rates of retinopathy, nephropathy, and neuropathy in the intensively treated group and led to major changes in the approach to diabetes treatment in the United States and worldwide. Results of the UKPDS, conducted in a cohort of recently diagnosed patients with type 2 diabetes, later confirmed the benefits of more intensive glucose control in the prevention of microvascular complications. These and other studies have provided convincing evidence that hyperglycemia is the driving force behind diabetic microvascular disease. Indeed, the long-term follow-up studies of the DCCT cohort showed that the benefits seen in the intensively treated group persisted for at least a decade after the study ended, even after HbA_{1c} levels between the two treatment groups converged, suggesting that the mechanisms underlying microvascular complications are conditioned by the prevailing metabolic milieu.

PATHOBIOLOGY

The cellular and molecular mechanisms that mediate hyperglycemic tissue damage are complex and still being elucidated. We now know that multiple interrelated pathways are involved, including four that have received the most attention as key mediators of vasculopathy (Fig. 229-8).

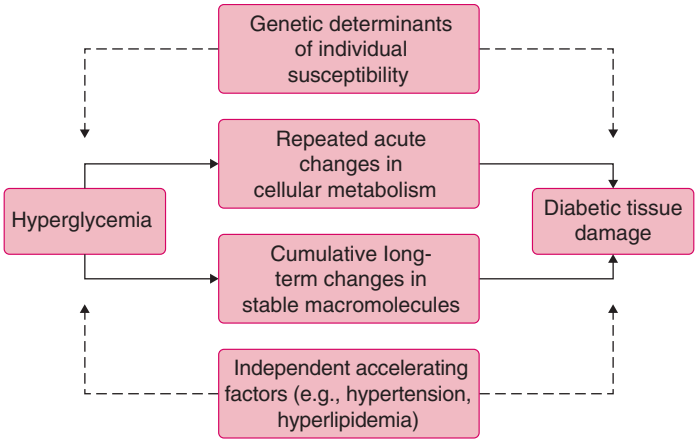

FIGURE 229-7. Factors related to the pathogenesis of diabetes complications. (From Brownlee M. The pathobiology of diabetic complications: a unifying mechanism. *Diabetes*. 2005;54:1615-1625.)

Advanced Glycation End Products

Advanced glycation end products (AGEs) are a heterogeneous group of compounds that form by the nonenzymatic interaction of glucose with amino groups on proteins. This process occurs continuously in vivo but is markedly accelerated in the presence of hyperglycemia. Indeed, the HbA_{1c} test to monitor the chronic level of glycemia was the result of observations of the glycosylation of subfractions of adult hemoglobin. Levels of AGEs in serum and tissues (e.g., skin collagen) correlate with diabetic vascular complications and mean glucose levels over time. AGEs can alter the properties and function of long-lived proteins, such as collagen and elastin, leading to vascular stiffness and increases in basement membrane thickness. AGE binding to specific cell surface receptors (e.g., receptors for AGE, RAGE), particularly on macrophages and endothelial cells, stimulates activation of signaling cascades that promote inflammation and oxidative stress. For example, AGE-RAGE interaction activates the transcription factor NF-κB, leading to multiple pathologic changes in gene expression. Further, AGEs formed intracellularly alter the function of many important cellular proteins. Studies in animal models provide strong evidence that AGE formation is a key process mediating hyperglycemic damage. However, to date, studies of anti-AGE compounds (e.g., aminoguanidine) have failed to demonstrate efficacy in preventing or ameliorating diabetic complications in humans.

Increased Polyol Pathway Flux

Metabolism of glucose through the aldose reductase pathway is generally minor because this enzyme has a low affinity for glucose. However, in the setting of intracellular hyperglycemia (most likely to occur in tissues that cannot downregulate glucose uptake, such as neurons and endothelial cells), there is increased flux through this pathway, leading to an accumulation of osmotically active sorbitol within the cell. Increased cellular osmolarity occurs, along with an increase in redox stress due to depletion of the reduced form of nicotinamide adenine dinucleotide phosphate and reduced glutathione. Inhibitors of aldose reductase have been proposed as a therapeutic strategy to reduce diabetic complications. Current evidence from clinical trials does not support their use, but this remains an active area of research.

Activation of Protein Kinase C

Intracellular hyperglycemia leads to increased de novo synthesis of diacylglycerol, which is a major activator of the protein kinase C family of enzymes. Activation of protein kinase C initiates a complex network of intracellular signaling that alters gene expression and results in enhanced angiogenesis, vasoconstriction, vascular permeability (by increases in vascular endothelial growth factor), cytokine activation, and extracellular matrix expansion. These alterations in cellular function have been linked to the development of

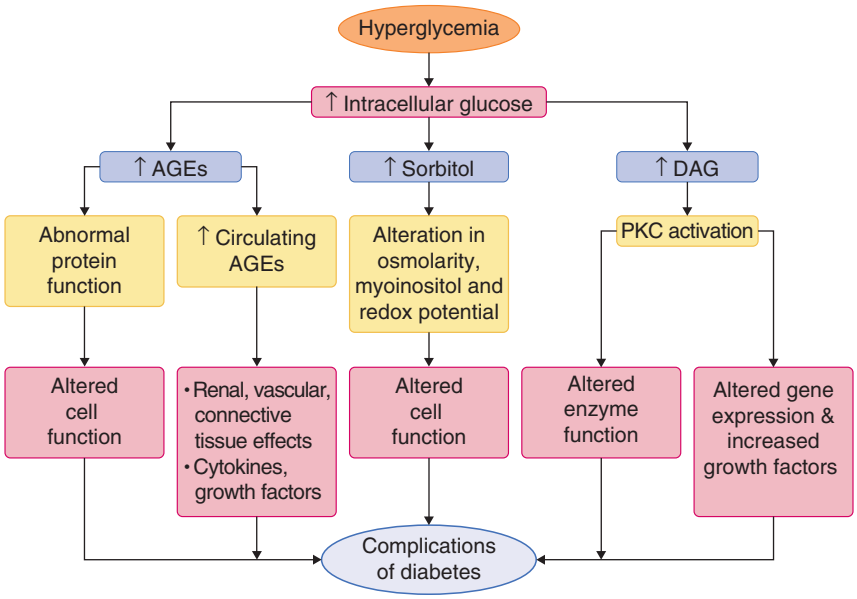

FIGURE 229-8. Proposed mechanisms of hyperglycemia-induced vascular complications. See text for discussion. AGEs = advanced glycation end products; DAG = diacylglycerol; PKC = protein kinase C.

microvascular complications (especially retinopathy) and atherosclerosis. Inhibitors of specific protein kinase C isoforms are being studied in clinical trials as agents specific for diabetic retinopathy and macular edema.

Increased Hexosamine Pathway Flux

In the setting of hyperglycemia and excess fatty acid oxidation, there is also increased flux of glucose through the hexosamine pathway, leading to increases in glucosamine 6-phosphate and ultimately post-translational modification of certain cytoplasmic and nuclear proteins. Associated with this are increases in expression of key genes, including those for transforming growth factor (α and β_1) and plasminogen activator inhibitor 1, and inhibition of endothelial nitric oxide synthase activity. Whereas the pathway has been linked to defective insulin action, its role in specific complications remains unclear.

These multiple and complex pathways are not mutually exclusive but are interconnected and may have a common antecedent process, which is overproduction of superoxide by the mitochondrial electron transport chain. Superoxide generates the production of other reactive oxygen species that can lead to cellular damage in a variety of ways. Data from animal models support the possibility that correction of diabetes-induced superoxide overproduction will have positive downstream effects on the various pathways leading to hyperglycemic tissue damage, but this remains to be confirmed in human studies.

Microvascular Complications

DIABETIC RETINOPATHY

Diabetic retinopathy (Chapter 423) is a highly prevalent, pathognomonic, microvascular complication, eventually affecting more than 50% of patients with long-term diabetes, although it causes vision impairment less frequently. The occurrence of vision loss due to diabetic retinopathy has declined during the past few decades as glucose and blood pressure control have improved in the population with diabetes. Nonetheless, it remains an important cause of preventable blindness, especially among patients with poor metabolic control. Both vascular and neural tissues in the retina are affected by chronic hyperglycemia. Early changes include the loss of retinal supporting cells (pericytes), basement membrane thickening, and retinal blood flow changes. Damaged retinal capillaries leak protein, red blood cells, and lipids, leading to retinal edema. Chronic retinal hypoxia (due to capillary occlusion) promotes neovascularization; these new vessels are abnormal and prone to rupture. Retinal hemorrhage, inflammation, and scarring can ultimately lead to traction retinal detachment and permanent vision loss (Table 229-14).

Diabetic retinopathy can be detected by dilated funduscopy, with early signs being the presence of microaneurysms, exudates, and intraretinal hemorrhages. Additional tests, including fluorescein angiography and ocular coherence tomography, are helpful to detect abnormal vessel permeability and macular edema, which can threaten vision. Regular screening by an eye care specialist (an ophthalmologist or optometrist) is recommended for all patients with diabetes because significant and potentially vision-threatening retinopathy can be present in the absence of any symptoms. Screening should begin at diabetes diagnosis for patients with type 2 diabetes because hyperglycemia has typically been present for years before it is recognized clinically. For patients with type 1 diabetes, screening can begin at 5 years after diagnosis or after puberty for childhood onset. Because retinopathy can progress rapidly during pregnancy, screening and follow-up should be more aggressive during this time (Table 229-15).

As with other diabetic complications, intensive glycemic control can prevent diabetic retinopathy and delay its progression but has limited effects on advanced retinal disease. Blood pressure control is also important to prevent worsening of retinopathy; there is some evidence that renin-angiotensin system (RAS) blockers may be especially beneficial.

Treatment of diabetic retinopathy (Chapter 423) includes laser photocoagulation, which can ablate abnormal vessels (thus reducing the risk of hemorrhage) and treat macular edema. Laser photocoagulation can be focal (to treat clinically significant macular edema or nonproliferative diabetic retinopathy) or panretinal (to treat severe nonproliferative diabetic retinopathy or proliferative diabetic retinopathy). Vitrectomy is a surgical procedure to remove hemorrhage and scar tissue that is obscuring vision. Nonsurgical therapies include intravitreal injection of glucocorticoids or anti–vascular endothelial growth factor monoclonal antibodies (e.g., ranibizumab) to treat macular edema. The established efficacy of retinopathy treatment, particularly photocoagulation, in preventing vision loss provides strong justification for routine retinopathy screening. There is evidence that treatment with fenofibrate reduces the progression of retinopathy, although the medication has not been approved for this indication in the United States. In addition to its well-known effects on lipid metabolism, fenofibrate appears to have significant anti-inflammatory, antiangiogenic, and antioxidant properties that are relevant to retinal disease. The presence of retinopathy is not considered a contraindication to the use of aspirin for CVD prevention.

Other eye conditions also affect patients with diabetes. Transient osmotically induced refractive error is common, especially at the time of diabetes diagnosis, but resolves with glucose control. Age-related eye conditions, such as cataracts and glaucoma, tend to occur at younger ages among diabetic patients. Diplopia and other gaze disorders due to acute mononeuropathy involving the cranial nerves (typically III or VI) are also more common in diabetes.

DIABETIC NEPHROPATHY

Diabetic nephropathy (Chapter 125) remains the most common single cause of end-stage renal failure, accounting for up to 50% of the cases in Western societies. Further, despite advances in the management of glucose and hypertension, the prevalence of chronic kidney disease among patients with diabetes has not declined in the past several decades. Overall, 20 to 30% of type 1 and type 2 diabetic patients develop evidence of nephropathy, although fewer type 2 patients progress to end-stage renal disease (ESRD). This may be because of competing mortality from CVD, with fewer surviving to ESRD. However, because of their much greater frequency in the population, the majority of diabetic patients presenting for treatment of ESRD (dialysis or transplantation) have type 2 diabetes. The major risk factor for the development of diabetic neuropathy is the duration and severity of hyperglycemia, but there is evidence for variation in genetic susceptibility. For example, African Americans and individuals with a family history of diabetic or nondiabetic renal disease are at higher risk for diabetic nephropathy. An insertion/deletion polymorphism in the gene encoding angiotensin-converting enzyme (ACE) has been widely reported to be associated with increased risk of diabetic nephropathy, but variants in genes involved in the polyol pathway, lipid metabolism, inflammatory cytokines, angiogenesis, and oxidative stress have also been identified.

Diabetic nephropathy develops during many years to decades, with a prolonged "silent" period before clinical detection, followed by more rapid progression to overt renal disease (Chapter 125). In the classic view, the hallmark of diabetic nephropathy is the development of proteinuria, which is due to alteration in glomerular basement membrane permeability and increases in intraglomerular pressure. The first clinical evidence of incipient nephropathy is the development of albuminuria, which is quantitatively minor at first

TABLE 229-14 CLASSIFICATION OF DIABETIC RETINOPATHY

	CLINICAL FEATURES
Mild NPDR	At least one microaneurysm
Moderate NPDR	Microaneurysms, intraretinal (blot) hemorrhage, soft exudates, venous beading, intraretinal microvascular abnormalities
Severe NPDR	More extensive intraretinal hemorrhages (>20 in each of four quadrants) *or* venous beading in at least two quadrants *or* prominent intraretinal microvascular abnormalities
PDR	Neovascularization and/or vitreous or pre-retinal hemorrhage; traction retinal detachment
Clinically significant macular edema	Retinal thickening or hard exudates approaching or involving the center of the macula

NPDR = nonproliferative diabetic retinopathy; PDR = proliferative diabetic retinopathy.

TABLE 229-15 RECOMMENDED INTERVALS FOR DIABETIC RETINOPATHY SCREENING

DIABETES TYPE	FIRST EXAMINATION	FOLLOW-UP
Type 1	5 years after diagnosis	Annual
Type 2	At time of diagnosis	Annual
Established diabetes during pregnancy	Before or soon after conception	At least every 3 months

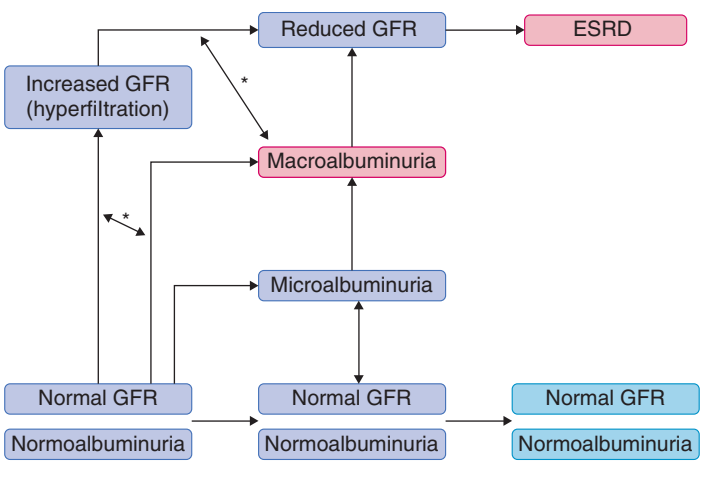

FIGURE 229-9. Development of diabetic nephropathy. See text for discussion. GFR = glomerular filtration rate; ESRD = end-stage renal disease. *GFR and albuminuria may progress independently of each other, i.e., patients may have micro- or macroalbuminuria even though their GFR is normal or even slightly elevated. However, macroalbuminuria is usually associated with reduced GFR, and is a strong risk for progressive ESRD. (From Boger CA, Sedor JR. GWAS of diabetic nephropathy: is the GENIE out of the bottle? *PLoS Genet*. 2012;8:e1002.)

(microalbuminuria, urine albumin-to-creatinine ratio of 30 to 300 mg/g) and then progresses to overt proteinuria, sometimes in the nephrotic range (>2 g/day). During the microalbuminuria phase, GFR is preserved but begins to decline in parallel with increasing proteinuria, leading to ESRD 5 to 15 years after the first detection of abnormal albumin excretion. However, recent evidence suggests that chronic kidney disease in diabetes is more heterogeneous than previously thought, with some patients progressing to advanced stages of chronic kidney disease in the absence of albuminuria (Fig. 229-9). Nonalbuminuric diabetic kidney disease appears more likely to occur in older patients with type 2 diabetes and may reflect, in part, the concurrence of multiple renal risk factors, including hypertension, obesity, and dyslipidemia. Further, microalbuminuria does not inevitably progress, with some patients regressing to normal or maintaining small but stable amounts of albuminuria. However, persistent and increasing albuminuria is a marker of high risk for progression to clinical nephropathy. Pathologic changes that are typical of diabetic nephropathy include an increase in glomerular basement membrane thickness and increased accumulation of extracellular matrix leading to mesangial expansion and the classic Kimmelstiel-Wilson nodular lesion.

Patients with diabetes should be screened annually for renal involvement (Chapter 125) by measurement of albumin on a spot urine sample with a sensitive immunoassay to detect microalbuminuria and by measurement of serum creatinine for calculation of estimated GFR. The finding of moderately increased urine albumin-to-creatinine ratio (30 to 300 mg albumin per gram of creatinine) should be confirmed on two of three repeated tests because transient increases are not uncommon but may not be clinically important. Data from the DCCT and other studies provide strong evidence that aggressive glycemic control can prevent the development of diabetic nephropathy and can retard the progression of microalbuminuria. However, there is little evidence that glycemic control can modulate the course once clinical albuminuria (>300 mg/day) and declining GFR occur. Central to the treatment of patients with albuminuria (micro or clinical) is intensive blood pressure control, preferentially by blockade of the RAS. Both ACE inhibitors and angiotensin receptor blockers have been shown to delay the progression of diabetic nephropathy and are recommended for patients with albuminuria even in the absence of hypertension. Despite initial enthusiasm, combined ACE inhibitor and angiotensin receptor blocker therapy is not recommended because of high rates of hyperkalemia and acute renal injury. In hypertensive patients, other drugs, such as calcium-channel blockers, diuretics, and β-blockers, can be used as additional therapy if needed to achieve adequate blood pressure control. There is little evidence to support use of RAS blockade in diabetic patients who are normotensive and normoalbuminuric, although there may be a therapeutic rationale for use of these agents in patients who cannot achieve adequate glycemic control. Dietary protein restriction has been recommended in the past for patients with nephropathy, but recent trials have been unable to demonstrate an effect of a low-protein diet on the rate of deterioration of GFR.[15]

DIABETIC NEUROPATHY

Diabetic neuropathy (Chapter 420) is a common complication of diabetes, with an estimated lifetime prevalence of about 50%. Diabetic neuropathy can be manifested in a variety of syndromes, including radiculoplexopathy and autonomic neuropathy, but the most common form is a characteristic distal symmetrical polyneuropathy (DSP), resulting from large-fiber nerve damage. Despite its high prevalence, there is no distinct neuropathic symptom or lesion that is specific to diabetes, and separation of diabetic neuropathy from other causes of nerve damage can be problematic. As for other microvascular complications, the etiology of DSP is attributed to hyperglycemic damage, as demonstrated by the dramatic 60% reduction in neuropathy in the intensive treatment group in the DCCT study. However, the possibility that the pathogenesis of DSP may differ in type 2 diabetes, with dyslipidemia and insulin resistance also contributing, has recently emerged. Support for this view comes from largely negative neuropathy results of clinical trials of intensive glucose control in type 2 diabetes (e.g., Action to Control Cardiovascular Risk in Diabetes [ACCORD] trial, VA Cooperative study) and the observation that the prevalence of DSP is already increased in the setting of prediabetes and the metabolic syndrome.

The clinical manifestations of DSP include symptoms of pain, paresthesias, and numbness that typically begin in the feet and progress more proximally in a "stocking and glove" distribution (Chapter 420). For some patients, neuropathic pain can be severe and disabling, resulting in a major reduction in quality of life. Loss of sensation, which may not be noticed by the patient, constitutes an important risk factor for falls due to gait instability. Ulceration, uncontrolled infection, and amputation can also occur from altered foot mechanics and inability to perceive repetitive trauma or other foot injury. DSP can be diagnosed by the presence of classic symptoms and by loss of ability to perceive pressure from a nylon (Semmes-Weinstein) monofilament. Additional tests, such as nerve conduction studies or electromyography, are occasionally indicated to distinguish DSP from radiculopathy. Current treatment options are mostly limited to control of metabolic risk factors (i.e., glucose, lipids) and symptoms, although some agents in clinical trials (e.g., aldose reductase inhibitors) show some promise. The chronic pain of DSP can be difficult to manage. Available therapies include tricyclic antidepressants, serotonin-norepinephrine reuptake inhibitors, anticonvulsants (such as gabapentin and pregabalin), and opioids.

Other forms of diabetic nerve damage (Chapter 420)[16] include small-fiber predominant neuropathy, radiculoplexopathy (diabetic amyotrophy), noncompressive radiculopathy, and mononeuritis multiplex. Autonomic neuropathy can be manifested as gastroparesis, urinary retention, erectile dysfunction, sudomotor dysfunction (typically anhidrosis of the extremities with or without hyperhidrosis of the trunk), cardiac arrhythmias, and disturbance of gut motility (diabetic diarrhea or constipation). Cardiac autonomic neuropathy is an especially ominous form of diabetic autonomic neuropathy. Typical clinical manifestations of cardiac autonomic neuropathy include resting tachycardia, diminished heart rate variability, and orthostatic blood pressure changes. Patients with cardiac autonomic neuropathy are at high risk for myocardial infarction, congestive heart failure, and sudden cardiac death.

DIABETIC FOOT

The combination of sensory impairment due to peripheral neuropathy and reduced tissue perfusion due to large-vessel atherosclerosis (peripheral arterial disease) or microvascular dysfunction can result in ulceration, infection, and ultimately lower extremity amputation. A typical case involves development of an ulceration (often surrounded by callus formation) on the plantar surface of the foot, often underneath the metatarsal heads. Ulceration can be slow to heal because of repetitive trauma from walking and impaired blood flow; hyperglycemia may also impair wound healing by effects on white blood cell migration and function. In the absence of protective sensation, an infection may fester for weeks and eventually invade the bone, leading to osteomyelitis. Altered foot mechanics can also lead to repeated (and usually undetected) fractures that destroy normal foot architecture and result in the classic Charcot foot deformity.

For many patients, foot amputation is the most feared diabetic complication; fortunately, it can be prevented in most cases but requires vigilance on the part of the patient and health care team. Routine foot examination,

especially for patients who have evidence of sensory loss, should be performed at every medical visit, and patients should be instructed to inspect their feet daily for cracks, fissures, ulcers, or inflammation. Patients should avoid walking barefoot (even at home) and should wear protective covering (avoid sandals) outside. Thermal injuries can be prevented by avoiding use of heating pads or hot-water bottles on the feet. Referral to a foot care specialist should be considered for patients with sensory loss, foot deformity, extensive callus formation, and nonhealing ulcers. Ulcers are treated with aggressive débridement of necrotic tissue and systemic antibiotics (guided by culture of infected tissue) if infection is present. Pressure "off-loading" by use of special shoes, orthotics, or application of total contact casts may be necessary to allow healing. Additional treatments include use of topical platelet-derived growth factor, bioengineered skin substitutes, hyperbaric oxygen, and negative-pressure wound therapy, although none of these has shown conclusive evidence of effectiveness in promoting wound healing.

OTHER ASSOCIATED CONDITIONS

Although not traditionally recognized as diabetic complications, there are a number of disorders that are increased in frequency or severity in patients with diabetes and that have a plausible or established relationship with hyperglycemia. These include periodontal disease, Alzheimer dementia, and musculoskeletal disorders, such as limited joint mobility, adhesive capsulitis, Dupuytren contracture, and trigger finger (flexor tenosynovitis). Patients with poorly controlled diabetes are widely thought to have increased susceptibility to infection, particularly with fungal pathogens. Defects in immune function (impaired neutrophil chemotaxis) have been described in diabetes, but whether this occurs in reasonably controlled diabetes or contributes to clinical infection is unclear. The incidence of osteoporotic fractures appears to be increased in women with diabetes, despite the presence of normal or even increased bone density. There is also emerging evidence that the frequency of some cancers (e.g., pancreatic, endometrial, colorectal, breast) is increased among people with diabetes.

Cardiovascular Disease in Diabetes

Atherosclerotic CVD is the major cause of morbidity and mortality for patients with diabetes and contributes substantially to its economic costs. The clinical and pathologic features of CVD in diabetes are generally not distinguishable from those occurring in nondiabetic individuals, but they are manifested at an earlier age, are more aggressive, and are associated with mortality rates that are two to four times higher in patients with diabetes (Chapter 52). This increased CVD risk is true for both type 1 and type 2 diabetes, with CVD in type 1 diabetes being strongly associated with concurrent presence of renal disease. Diabetes is also an important risk factor for peripheral vascular disease and stroke, which carries greater mortality risk than in nondiabetic patients.

PATHOBIOLOGY OF CARDIOVASCULAR DISEASE IN DIABETES

The pathogenesis of atherosclerotic CVD in diabetes is complex and multifactorial, with several mechanisms playing key roles. *Metabolic factors*, including hyperglycemia, insulin resistance, dyslipidemia, and increases in circulating free fatty acids, contribute to atherosclerotic plaque formation. Increases in *oxidation and glycoxidation* of lipoproteins increase their atherogenicity and enhance foam cell formation. *Endothelial dysfunction*, an early event in the development of atherosclerosis, has been described in association with several metabolic syndrome components, including hyperglycemia, insulin resistance, hypertension, and dyslipidemia. Systemic *inflammation*, which contributes to accelerated plaque formation, is increased in diabetes and obesity as a consequence of increased cytokine production by adipose tissue. Finally, diabetes is characterized by a *prothrombotic* state due to enhanced platelet reactivity and alterations in coagulation factors, including increased circulating levels of fibrinogen and plasminogen activator inhibitor 1.

DIABETIC CARDIOMYOPATHY AND HEART FAILURE

Diabetic cardiomyopathy is defined as alterations in cardiac structure and function that are not directly attributable to coronary artery disease or hypertension (Chapters 58 and 59). Characteristic features include cardiac hypertrophy, left ventricular dysfunction (diastolic may precede systolic), and altered myocardial metabolism. Diabetes is a recognized risk factor for the development of heart failure, even in the absence of atherosclerotic heart disease. For example, in the Framingham Heart Study, the frequency of heart failure was twice in diabetic men and five times in diabetic women compared with age-matched controls and persisted despite correction for hypertension, obesity, dyslipidemia, and coronary artery disease. Increased activation of the renin-angiotensin-aldosterone system and formation of AGEs are thought to contribute to myocardial fibrosis and stiffness, and altered substrate utilization (preferential use of free fatty acids) can promote myocyte dysfunction by enhanced production of reactive oxygen species and other mechanisms. Characteristic changes in myocardial function and structure were reported in type 1 diabetes in the DCCT and Epidemiology of Diabetes Interventions and Complications (EDIC) study and were related to long-term glycemic control.

PREVENTION OF CARDIOVASCULAR DISEASE IN DIABETES

Aggressive control of CVD risk factors is recommended for most patients with diabetes, keeping in mind that the presence of diabetes is considered the risk equivalent of a prior myocardial infarction by most risk assessment algorithms (e.g., Framingham risk score, Adult Treatment Panel III Report of the National Cholesterol Education Program). Assessment of blood pressure, lipid profile, and smoking status should be included as part of routine diabetes care. Determination of the optimal targets for risk factor control has been the subject of several large randomized trials, which have informed consensus guidelines.

GLUCOSE CONTROL

Hyperglycemia is a major risk for atherosclerotic CVD. In population-based studies including diabetic and nondiabetic cohorts, HbA_{1c} has been reported as an independent predictor of all-cause and CVD mortality, and among individuals with diabetes, every 1% rise in HbA_{1c} is associated with a 30% increase in all-cause mortality and a 40% increase in CVD mortality. Compelling evidence for the benefit of intensive glucose control in patients with type 1 diabetes was shown in the DCCT/EDIC study, in which CVD events were reduced by 58%. However, in type 2 diabetes, hyperglycemia occurs in the setting of multiple other CVD risk factors, including hypertension, dyslipidemia, and obesity, which also contribute to risk, so the contribution of glucose control is unclear. Several large clinical trials in patients with type 2 diabetes have failed to show that control of hyperglycemia has important effects on CVD outcomes (see later), highlighting the complex pathogenesis of vascular disease in diabetes. Similarly, an intensive lifestyle program designed to achieve weight loss and exercise goals also failed to demonstrate significant effects on CVD outcomes in type 2 diabetes patients.

The strongest evidence in favor of intensive glucose control comes from the long-term follow-up of the UKPDS, which demonstrated 15% reduction in myocardial infarction and 13% reduction in all-cause mortality in the intensive versus conventional treatment group. More recently, in the ACCORD trial, an intensive treatment arm, designed to maintain HbA_{1c} below 6%, was compared with conventional treatment with HbA_{1c} goal of 7.5% in a cohort of type 2 diabetes patients at high risk for CVD. This trial was stopped early because of unexpected increased mortality, largely CVD related, in the intensive treatment group. The reasons for increased mortality with intensive treatment are not known for certain, but increased frequency and severity of hypoglycemia or the toxicity of specific drugs or combinations has been proposed. Secondary analysis of ACCORD data showed a reduction in nonfatal myocardial infarction in the intensive treatment group, leading to speculation that some patients might still benefit. Other studies designed to address this, including the VA Cooperative Study and ADVANCE, also failed to show CVD benefit for intensive glucose control. These trials differed somewhat in patient characteristics, HbA_{1c} goal, and specific treatment regimens, and the largely negative results stimulated controversy. However, some consensus views have emerged[17]: (1) in the current era of effective treatment of other CVD risk factors (i.e., with statins, RAS blockers, antiplatelet therapy), the additional benefits of intensive glycemic control are modest at best; (2) patients with long-standing diabetes or established CVD are least likely to benefit from intensive glucose lowering; (3) the benefits of glucose lowering in the prevention of microvascular complications provide an independent rationale for strict glucose control for many patients; and (4) specific glycemic targets should be individualized according to the patient's characteristics (e.g., comorbidities, life expectancy, hypoglycemia risk) and preferences.

HYPERTENSION

Hypertension (Chapter 67) is a common comorbidity in diabetes, affecting the majority of patients with type 2 diabetes, and constitutes an important modifiable CVD risk factor. Further, in even the earliest stages of diabetic nephropathy (i.e., microalbuminuria), hypertension is further accelerated. In

type 1 diabetes, hypertension is generally the result of concurrent renal disease, with both contributing to CVD risk. The importance of blood pressure control in reducing CVD events as well as microvascular outcomes in patients with diabetes was established by several major trials, including the UKPDS, Systolic Hypertension in the Elderly Program (SHEP), Hypertension Optimal Treatment (HOT) study, and others. However, analysis of these and other trials failed to show evidence of improved outcomes (i.e., in myocardial infarction or mortality) with systolic blood pressure targets of 130 mm Hg or lower. An even more aggressive target systolic blood pressure of less than 120 mm Hg was shown to be of no additional benefit in reducing CVD events in the ACCORD trial. The current consensus is that blood pressure goals should be less stringent than previously recommended. The Eighth Joint National Committee (JNC8) guidelines recommend a blood pressure goal of less than 140/90 mm Hg for all patients younger than 60 years, regardless of diabetes or renal status, and less than 150/90 mm Hg for patients 60 years and older.[18] Other guidelines, including those from the American Diabetes Association (ADA), suggest a blood pressure target of less than 140/80 mm Hg for patients with diabetes but with the further recommendation that a lower target may be considered for younger patients if it can be achieved without excessive treatment burden. However, many of these recommendations are based on expert opinion rather than on evidence from randomized trials, and some uncertainty remains.

The choice of antihypertensive agent has also received considerable study, which is complicated by the fact that many patients will require treatment with two or more drugs to achieve target blood pressure. ACE inhibitors and angiotensin receptor blockers are generally considered first-line therapy for patients with diabetes, in part on the basis of their demonstrated renoprotective benefits. In addition, results from several randomized trials, including the Heart Outcomes Protection Study (HOPE), Fosinopril versus Amlodipine Cardiovascular Events Trial (FACET), and Appropriate Blood Pressure Control in Diabetes (ABCD), indicated improved cardiovascular outcomes with ACE inhibitors compared with other antihypertensive drugs, although this was not the case for the UKPDS, in which β-blockers were equally effective. Calcium-channel blockers and low-dose diuretics are also recommended as add-on therapy if needed to achieve blood pressure targets. Use of β-blockers should be considered in the setting of established CVD because of their proven benefits in patients with prior myocardial infarction and congestive heart failure. However, β-blockers should be used with caution in patients at high risk for hypoglycemia because they may blunt the autonomic warning symptoms associated with low glucose concentration. Both β-blockers and thiazide diuretics have been reported to increase the risk for development of diabetes, although there is little evidence for significant deterioration of glycemic control in patients with diabetes.

DYSLIPIDEMIA

The characteristic dyslipidemia of type 2 diabetes and insulin-resistant states, which includes low levels of high-density lipoprotein (HDL) cholesterol, elevated triglycerides, and small dense LDL particles, is highly atherogenic (Chapter 206). LDLs also are prone to oxidative modification in the setting of hyperglycemia, which enhances their atherogenicity. There is substantial clinical trial evidence to support lowering of LDL cholesterol levels with statin drugs in the majority of patients with diabetes older than 40 years. These findings come from trials limited to diabetes (Collaborative Atorvastatin Diabetes Study [CARDS]) and to diabetes subset analysis of larger trials (Heart Protection Study), all of which report similar CVD benefits of statin therapy among diabetics and nondiabetics. ADA recommendations are for target LDL levels of less than 100 mg/dL for most adult patients with diabetes and less than 70 mg/dL for diabetic patients with established CVD or multiple risk factors. Recent guidelines from the American Heart Association (AHA) and American College of Cardiology (ACC) have focused on CVD risk stratification to determine the need for and intensity of statin therapy.[19] With this approach, virtually all patients with diabetes (aged 40 to 75 years) would be candidates for statin therapy, regardless of baseline level of LDL cholesterol. Diabetic patients with established atherosclerotic CVD or estimated 10-year CVD risk of more than 7.5% would receive high-intensity statin treatment (regimens sufficient to lower LDL cholesterol >50% from untreated baseline); all others would be considered for moderate-intensity treatment (lowering of LDL cholesterol 30 to <50%). The evidence base to support these new recommendations is considered relatively strong. Although the ADA and the AHA/ACC guidelines differ in structure, ultimately the recommendations for most patients with diabetes will be similar with either approach.

Recent observations from several trials (e.g., JUPITER) and observational cohort studies of an increase in incident diabetes with statin therapy have generated concern, although the risk appears to be small in magnitude (hazard ratio, ~1.2) and is outweighed by the substantial benefits of CVD protection.[20] Clinically relevant effects of statins on glucose control among established diabetics have not been reported. In patients intolerant of statins, nicotinic acid (niacin) can be used, although CVD outcome trials have been disappointing despite substantial improvement in lipid parameters, including lowering of LDL cholesterol and increasing of HDL cholesterol levels. Further, nicotinic acid may worsen insulin resistance and glycemic control in some patients. Bile acid sequestrants, such as colesevelam or cholestyramine, can also be used but may exacerbate the hypertriglyceridemia characteristic of diabetic dyslipidemia.

In contrast to the definitive benefits of LDL lowering, there is less evidence that pharmacologic treatment of hypertriglyceridemia or of low HDL cholesterol levels reduces CVD risk. This may be in part due to the lesser efficacy of available drugs to alter these lipid subfractions. Trials with fibrate derivatives (gemfibrozil and fenofibrate) have yielded mixed results, and the addition of fenofibrate to a statin did not reduce the rate of major CVD events compared with statin alone in the ACCORD trial. Because most statins have some triglyceride-lowering effect, maximizing statin dose should be considered for patients with high triglyceride levels. Lifestyle factors are also effective, including weight loss and dietary modification (reduced fat diet, avoidance of alcohol). Omega-3 fatty acid supplementation is another option to lower triglyceride levels, although CVD outcome data are lacking. Pharmacologic treatment (i.e., with fibrates or fish oil supplements) of severe hypertriglyceridemia (triglyceride level >1000 mg/dL) is indicated to prevent acute pancreatitis.

ANTIPLATELET THERAPY

Prophylactic aspirin therapy is widely used for prevention of cardiovascular events in high-risk patients (i.e., those with prior myocardial infarction or stroke), with reported risk reductions of about 12% (Chapter 38). Results from clinical trials in patients with diabetes suggest that aspirin may be somewhat less effective for CVD prevention than in patients without diabetes, although this has not been a consistent finding. Current guidelines recommend aspirin therapy for diabetic patients with a prior CVD event (secondary prevention) or with increased CVD risk (10-year risk of >10%). This includes most men older than 50 years or women older than 60 years who also have one or more additional CVD risk factors: smoking, hypertension, albuminuria, dyslipidemia, or family history of CVD. For patients at lower CVD risk, the potential adverse effects from bleeding may outweigh the potential benefits, and routine use is not recommended. The optimal dose (balancing thrombosis prevention with the risk of bleeding) of aspirin has not been established and may differ according to patient characteristics, but 75 to 162 mg/day is commonly recommended. For high-risk patients who are unable to tolerate aspirin, clopidogrel is an effective alternative.

TREATMENT OF ESTABLISHED CARDIOVASCULAR DISEASE IN DIABETES

Rx

In general, treatment of clinically established CVD, including acute coronary syndromes and stable angina, is similar in diabetic and nondiabetic patients. There is some evidence that ischemic symptoms may be less intense, atypical, or absent in diabetic patients, leading to higher rates of "silent" myocardial infarction. However, a strategy of screening for ischemic heart disease, by exercise stress testing, in asymptomatic patients did not result in lower event rates or improved outcomes. Therefore, current recommendations are for coronary artery disease screening in patients with symptoms suggestive of ischemia.

The role of intravenous insulin (with or without potassium and glucose infusion) in the setting of acute myocardial infarction has been considered in a few studies. In the Diabetes and Insulin-Glucose Infusion in Acute Myocardial Infarction (DIGAMI) study, acute myocardial infarction patients with diabetes were treated with standard therapy or with insulin infusion during the first 48 hours, followed by continued insulin use after hospital discharge. Mortality after 1 year was reduced by 30% in the insulin-treated group. However, the implications of these results have been debated because factors other than insulin treatment differed between the two groups (i.e., sulfonylureas were routinely used in the standard therapy group but withdrawn from the insulin group). These findings subsequently were not confirmed in a follow-up study, and this approach has largely been abandoned.

Several studies have addressed the roles of medical therapy and revascularization in diabetic patients with coronary artery disease. Among them, the Bypass Angioplasty Revascularization Investigation 2 Diabetes (BARI 2D) study demonstrated that a policy of medical management (including aggressive risk factor modification) was as effective as early revascularization in diabetic patients with stable angina. In the Future Revascularization Evaluation in Patients with Diabetes Mellitus: Optimal Management of Multivessel Disease (FREEDOM) trial, diabetic patients with multivessel coronary disease had better outcome (reduced rates of death from any cause or nonfatal myocardial infarction) with coronary bypass surgery compared with percutaneous intervention with drug-eluting stents, although strokes were more frequent in the surgical group.

Grade A References

A1. Misso ML, Egberts KJ, Page M, et al. Continuous subcutaneous insulin infusion (CSII) versus multiple insulin injections for type 1 diabetes mellitus. *Cochrane Databse Syst Rev*. 2010;1: CD005103.

A2. Hemmingsen B, Lunc S, Gluud C, et al. Targeting intensive glycaemic control versus targeting conventional glycaemic control for type 2 diabetes mellitus. *Cochrane Database Syst Rev*. 2013;11:CD008143.

A3. Schauer PR, Bhatt DL, Kirwan JP, et al. Bariatric surgery versus intensive medical therapy for diabetes—3 year outcomes. *N Engl J Med*. 2014;370:2002-2013.

A4. Reznik Y, Cohen O, Aronson R, et al. Insulin pump treatment compared with multiple daily injections for treatment of type 2 diabetes (OpT2mise): a randomised open-label controlled trial. *Lancet*. 2014;384:1265-1272.

A5. Finfer S, Liu B, Chittock DR, et al. The NICE-SUGAR Study Investigators. Hypoglycemia and risk of death in critically ill patients. *N Engl J Med*. 2012;367:1108-1118.

A6. Diabetes Prevention Program Research Group. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet*. 2009;374:1677-1686.

A7. Gaede P, Lund-Andersen H, Parving HH, et al. Effect of a multifactorial intervention on mortality in type 2 diabetes. *N Engl J Med*. 2008;358:580-591.

A8. Nguyen Q, Brown D, Marcus D, et al. Ranibizumab for diabetic macular edema: results from 2 phase III randomized trials: RISE and RIDE. *Ophthalmology*. 2012;119:789-801.

A9. Keech AC, Mitchell P, Summanen PA, et al. Effect of fenofibrate on the need for laser treatment for diabetic retinopathy (FIELD study): a randomized controlled trial. *Lancet*. 2007;370: 1687-1697.

A10. Fried L, Emanuele N, Zhang J, et al. Combined angiotensin inhibition for the treatment of diabetic nephropathy. *N Engl J Med*. 2013;369:1892-1903.

A11. Wing RR, Bolin P, Brancati FL, et al. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *N Engl J Med*. 2013;369:145-154.

A12. Zoungas S, Chalmers J, Neal B, et al. Follow-up of blood-pressure lowering and glucose control in type 2 diabetes. *N Engl J Med*. 2014;371:1392-1406.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *Lancet*. 2014;383:69-82.
2. Keenan H, Sun JK, Levine J, et al. Residual insulin production and pancreatic B-cell turnover after 50 years of diabetes: Joslin Medalist Study. *Diabetes*. 2010;59:2846-2853.
3. Evert A, Boucher J, Cypress M, et al. Nutrition therapy recommendations for the management of adults with diabetes. *Diabetes Care*. 2014;37(suppl 1):S120-S143.
4. Skyler JS. Primary and secondary prevention of type 1 diabetes. *Diabet Med*. 2013;30:161-169.
5. Miller RG, Secrest AM, Sharma RK. Improvements in the life expectancy of type 1 diabetes: the Pittsburgh Epidemiology of Diabetes Complications study cohort. *Diabetes*. 2012;61:2987-2992.
6. Lind M, Svensson AM, Kosiborod M, et al. Glycemic control and excess mortality in type 1 diabetes. *N Engl J Med*. 2014;371:1972-1982.
7. Gregg EW, Li Y, Wang J, et al. Changes in diabetes-related complications in the United States, 1990-2010. *N Engl J Med*. 2014;370:1514-1523.
8. Kahn SE, Cooper ME, Del Prato S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. *Lancet*. 2014;383:1068-1083.
9. Ng HJ, Gloyn AL. Bridging the gap between genetic associations and molecular mechanisms for type 2 diabetes. *Curr Diab Rep*. 2013;13:778-785.
10. Hales CN, Barker DJ. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Int J Epidemiol*. 2013;42:1215-1222.
11. Bray G, Edelstein S, Crandall JP, Diabetes Prevention Program Research Group, et al. Long-term safety, tolerability, and weight loss associated with metformin in the Diabetes Prevention Program Outcomes Study. *Diabetes Care*. 2012;35:731-737.
12. Qaseem A, Humphrey LL, Chou R, et al. Use of intensive insulin therapy for the management of glycemic control in hospitalized patients: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. 2011;154:260-267.
13. Seaquist ER, Anderson J, Childs B, et al. Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. *Diabetes Care*. 2013;36:1384-1395.
14. Nathan DM, Bayless M, Cleary P, et al. Diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: advances and contributions. *Diabetes*. 2013;62:3976-3986.
15. Wheeler ML, Dunbar SA, Jaacks LM, et al. Macronutrients, food groups and eating patterns in the management of diabetes: a systematic review of the literature. *Diabetes Care*. 2012;35:434-445.
16. Griebeler ML, Morey-Vargas OL, Brito JP, et al. Pharmacologic interventions for painful diabetic neuropathy: An umbrella systematic review and comparative effectiveness network meta-analysis. *Ann Intern Med*. 2014;161:639-649.
17. Kishore P, Kim S, Crandall JP. Glycemic control and cardiovascular disease: what's a doctor to do? *Curr Diab Rep*. 2012;12:255-264.
18. James PA, Oparil S, Carter BL, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014;311:507-520.
19. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014;63:2889-2934.
20. Preiss D, Seshasai SR, Welsh P. Risk of incident diabetes with intensive-dose compared with moderate-dose statin therapy: a meta-analysis. *JAMA*. 2011;305:2556-2564.

REVIEW QUESTIONS

1. A 51-year-old man has a history of type 2 diabetes mellitus for 6 years. Past medical history is significant for chronic hepatitis C infection, chronic kidney disease stage 3, and a recent hospitalization for an upper gastrointestinal bleed. He takes a sulfonylurea for blood glucose control and rarely checks his blood glucose level. Fasting plasma glucose concentration in the office is 195 mg/dL, and his HbA_{1c} is 6.8%. What do you conclude about his glucose control?

A. His average blood glucose concentration during the past 3 months is approximately 140 mg/dL.
B. HbA_{1c} may be falsely high because of chronic kidney disease.
C. HbA_{1c} may be falsely low because of liver disease.
D. HbA_{1c} levels are increased after acute blood loss.
E. HbA_{1c} levels are more reflective of postprandial than of fasting glucose concentration.

Answer: C This patient has several reasons that his HbA_{1c} may not accurately reflect his mean plasma glucose concentration. HbA_{1c} results may be influenced by a number of factors, including conditions that alter red cell survival or cause interference with a specific assay. The HbA_{1c} may be falsely low in this patient because of cirrhosis (increased red cell turnover), recovery from recent acute blood loss (greater percentage of younger erythrocytes with shorter exposure to glucose), or transfusion (dilution of patient's blood with nondiabetic donor blood). In these instances, measurement of glycated serum proteins (fructosamine) or direct measurement of plasma glucose concentration will more accurately reflect glycemic control.

Additional information about HbA_{1c} assay methodology and interpretation of results can be obtained from the National Glycohemoglobin Standardization Program: http://www.ngsp.org.

2. A 38-year-old woman has had type 1 diabetes mellitus since the age of 12 years. She has maintained excellent control (HbA_{1c} 6.0%) with a basal/bolus injection regimen. She tests her glucose level four or five times a day, and review of her meter download shows many glucose levels in the 30s and 40s. However, the patient is unconcerned because she has no symptoms at these times. On questioning, she admits to recently "spacing out" while driving, which led to a minor traffic accident. Regarding the etiology and treatment of hypoglycemia in this patient:

A. She has adapted to low blood glucose concentration and no change in treatment is required.
B. She has developed hypoglycemia unawareness and her target HbA_{1c} should be increased.
C. Strict avoidance of hypoglycemia is of little benefit in reversing hypoglycemia-associated autonomic failure.
D. An excessive counter-regulatory hormone response to hypoglycemia may contribute to her lack of symptoms.
E. Treatment with β-blocker should be considered.

Answer: B In patients with long-standing diabetes, the counter-regulatory systems that normally would counteract the decline of glucose to dangerous levels may be impaired. This is especially true for patients with type 1 diabetes, who often have defects in glucagon and epinephrine response during hypoglycemia. This decrease in epinephrine response during hypoglycemia is accompanied by an attenuated autonomic neural response, which results in the clinical syndrome of *impaired awareness of hypoglycemia*. Without autonomic symptoms, mild hypoglycemia may proceed unnoticed to more advanced and dangerous phases. There is, however, evidence that hypoglycemia-associated autonomic failure can be reversed by strict avoidance of hypoglycemia, which can be facilitated by increasing target glucose levels.

Reno CM, Litvin M, Clark AL, Fisher SJ. Defective counterregulation and hypoglycemia unawareness in diabetes: mechanisms and emerging treatments. *Endocrinol Metab Clin North Am*. 2013;42:15-38.

3. A mother brings her 19-year-old son, who has type 1 diabetes and uses an insulin pump, to see you and to ask for referral to a dietitian. She complains that her son refuses to follow his "diabetic diet" and frequently eats junk food, including fast food (burgers, fries, pizza) and ice cream. She also worries that he sometimes skips meals, saying he is not hungry. His body mass index is 22, and his recent laboratory results show an HbA_{1c} of 7.8%, low-density lipoprotein cholesterol of 95, and normal triglycerides. Which of the following dietary recommendations is most appropriate for this patient?

A. An 1800-calorie/day American Diabetes Association diet
B. A low-carbohydrate diet
C. A low-protein diet
D. A low-fat, high-fiber diet
E. A flexible "heart healthy" meal plan that limits concentrated sweets and emphasizes fruits and vegetables

Answer: E Dietary recommendations for patients with diabetes have changed substantially over time, from the extremely low-carbohydrate, high-fat diets used before the discovery of insulin as a therapy, to "exchange diets," to more flexible meal plans. For patients with type 1 diabetes, the key element is for the patient to learn to match mealtime insulin doses to the carbohydrate content of the meal. Severely restricted diets (very low carbohydrate or low calorie) are neither required nor advisable, although avoidance of large carbohydrate loads will help minimize post-meal glycemic excursions. For most patients with type 2 diabetes who are typically overweight or obese, moderate carbohydrate intake and reduction in total calories are advised. Current recommendations allow a variety of eating styles and ethnic food preferences, with emphasis on fruits and vegetables, low-fat protein sources, and use of monounsaturated or polyunsaturated fats.

Evert A, Boucher J, Cypress M, et al. Nutrition therapy recommendations for the management of adults with diabetes. *Diabetes Care*. 2013;36:3821-3842.

Lasa A, Miranda J, Bullo M, et al. Comparative effect of two Mediterranean diets versus a low-fat diet on glycaemic control in individuals with type 2 diabetes. *Eur J Clin Nutr*. 2014;68:767-772.

4. A 54-year-old woman presents to her physician for treatment of hypertension. She had gestational diabetes during her last pregnancy 15 years ago, and there is a family history of type 2 diabetes (mother and older brother). Her body mass index is 36. Fasting glucose concentration is 110 mg/dL, and HbA_{1c} is 6.2%. Which of the following have been shown to reduce the progression to diabetes in high-risk patients?

A. Weight loss by reduced calorie diet
B. Treatment with metformin
C. Treatment with acarbose
D. Bariatric surgery
E. All of the above

Answer: E This patient has multiple risk factors for the development of type 2 diabetes, including family history, prior history of gestational diabetes, obesity, and hypertension. In addition, her glucose and HbA_{1c} levels are already elevated above normal, in the defined "pre-diabetes" range. All of the treatments listed have been shown to prevent or to delay the onset of diabetes in randomized clinical trials. The most consistent evidence comes from weight loss trials. Both the Finnish Diabetes Prevention Study and the U.S. Diabetes Prevention Program reported 58% reduction in diabetes with a hypocalorie, reduced fat diet combined with moderate-intensity physical activity. Weight loss achieved with bariatric surgery is also highly effective in preventing (or even reversing) diabetes. Medications, including metformin, the α-glucosidase inhibitor acarbose, and troglitazone (a thiazolidinedione), have also been shown to reduce diabetes in high-risk patients, although somewhat less effectively than by lifestyle modification. Lifestyle changes and metformin are reported to be cost-effective interventions, but whether delay or prevention of type 2 diabetes will result in lower rates of cardiovascular disease and diabetes microvascular complications will require longer-term follow-up studies.

Schwarz PE, Greaves CJ, Lindstrom J, et al. Nonpharmacologic interventions for the prevention of type 2 diabetes mellitus. *Nat Rev Endocrinol*. 2012;8:363-373.

5. A 28-year-old woman with type 1 diabetes since the age of 12 years is considering having a child. Currently, her blood glucose is reasonably well controlled, although she admits this was not the case during her teens and early 20s, when her HbA_{1c} was in the 9 to 11% range. She has mild background diabetic retinopathy, normal blood pressure, urine albumin-to-creatinine ratio of 25 mg/g, and normal findings on foot examination. Which of the following is true?

A. She should delay pregnancy until she has achieved optimal glucose control (HbA_{1c} ~6.5%).
B. She should be treated with an angiotensin-converting enzyme inhibitor to prevent progression of renal disease during pregnancy.
C. Progression of her retinopathy during pregnancy is likely to result in vision loss.
D. The risk of her child's developing type 1 diabetes is 25 to 50%.
E. She should be advised to avoid pregnancy because of the risk of both maternal and fetal complications.

Answer: A Although pregnancies in women with type 1 diabetes are generally considered "high risk," the outlook for patients with minimal complications and good metabolic control is good. Women with advanced renal disease (proteinuria, reduced glomerular filtration rate) or proliferative retinopathy may experience rapid progression during pregnancy because of the influence of hormonal and hemodynamic changes and should be monitored closely by specialists. Use of angiotensin-converting enzyme inhibitors is contraindicated during pregnancy because of the risk of fetal renal damage. The key to successful pregnancy outcomes is achieving optimal glucose control before conception because the developing fetus is most susceptible to the teratogenic effects of hyperglycemia in the first 6 to 8 weeks of pregnancy, before the time that most women are aware of being pregnant. Maintaining strict glucose control during the pregnancy will reduce the risk of fetal complications, such as macrosomia and hyperbilirubinemia. Many patients benefit from use of an insulin pump and continuous glucose monitoring during this time. Although a child of a mother with type 1 diabetes is at increased risk of diabetes compared with the general population, the risk is less than 10%.

230

HYPOGLYCEMIA AND PANCREATIC ISLET CELL DISORDERS

KHALID HUSSAIN

DEFINITIONS

Hypoglycemia is one of the most common biochemical abnormalities observed in clinical practice. It is a biochemical finding and not a diagnosis. Hypoglycemic disorders are more common in neonates, infants and children as compared to adults. Inappropriately treated hypoglycemia can have severe consequences, including seizures, permanent brain injury, or death. This is especially the case in neonates with persistent forms of hypoglycemia, who are at high risk of brain injury from delays in diagnosis and effective therapy.

Hypoglycemic disorders in neonates, infants, and children differ from adults in important aspects. First, they are most often due to congenital or genetic disorders, such as disorders of insulin secretion, as well as a range of metabolic and endocrine diseases. Second, during a transitional period of 1 to 3 days after birth, low plasma glucose concentrations are common in normal neonates, which makes it difficult to identify the minority who have a persistent hypoglycemia disorder or a genetic hypoglycemia disorder. Third, the importance of early recognition and treatment of such persistent hypoglycemia disorders in neonates is emphasized by reports that developmental handicap, which might have been avoidable by early recognition and treatment, occurs in 25 to 50% of cases with congenital hyperinsulinism.[1]

There is no absolute number that defines hypoglycemia in adults and in children. The current adult recommendations define clinical hypoglycemia as a plasma (or serum) glucose concentration low enough to cause symptoms and/or signs, including impairment of brain function. Because the clinical manifestations and symptoms of hypoglycemia are nonspecific, it is therefore not possible to state a single plasma glucose concentration that categorically defines hypoglycemia. The measured plasma or serum glucose concentration may be low owing to an artifact (e.g., when the blood sample is collected in a tube that does not contain an inhibitor of glycolysis and when separation of the plasma or serum from the formed elements is delayed).

For these reasons, guidelines in adults emphasize the value of Whipple triad for confirming hypoglycemia: (1) symptoms and/or signs compatible with hypoglycemia, (2) a low measured plasma glucose concentration, and (3) resolution of symptoms and signs when glucose concentrations are raised. Because circulating fuels such as ketone bodies can be used by the brain, lower plasma glucose concentrations can occur in healthy individuals, particularly in women and children, without symptoms or signs during extended fasting. Therefore, for all of these reasons, it is not possible to state a single plasma glucose concentration that categorically defines hypoglycemia.

The aim of this chapter is to outline the physiologic and biochemical changes associated with maintenance of a normal blood glucose level, describe the role of the counter-regulatory hormones, review the different hypoglycemia disorders observed in adults and children, and then finally discuss the various management strategies.

PATHOBIOLOGY

Physiologic and Biochemical Changes During Fasting and Feeding

Overview

Plasma glucose concentration is tightly controlled by a balance between glucose production and utilization. Glucose is derived from three sources: (1) intestinal absorption that follows digestion of dietary carbohydrates; (2) glycogenolysis, the breakdown of glycogen, which is the polymerized storage form of glucose; and (3) gluconeogenesis, the formation of glucose from precursors including lactate (and pyruvate), amino acids (especially alanine and glutamine), and to a lesser extent, glycerol. Normally, there is tight coordination between rates of endogenous glucose influx into the circulation and glucose efflux out of the circulation into insulin-dependent tissues (skeletal muscle, adipose tissue, and liver). This coordination, despite periods of feeding and fasting, maintains the plasma glucose concentration in a relatively narrow range between 70 and 110 mg/dL (3.8 to 6 mmol/L). Figure 230-1 shows an outline of glucose physiology.

Glucose is an obligate metabolic fuel for the brain under physiologic conditions. Unlike other body tissues, the brain cannot oxidize fatty acids, and neither can it synthesize/store glucose for later use. It is dependent on a continuous supply of glucose from the circulation. Given the vital importance of brain function and the above circumstances, it is not surprising that physiologic mechanisms have evolved for the maintenance of plasma glucose concentrations.

Changes During Fasting

During fasting, the basal rate of glucose output by the liver is precisely matched to glucose uptake by various body tissues. They average 2.2 mg/kg/minute in healthy adults after an overnight fast. In infants, these rates are much higher (≈6 mg/kg/minute) because of their greater brain mass relative to their body weight. The brain is responsible for nearly two thirds of basal glucose utilization. The remaining one third is used by red blood cells, renal medulla, and to some extent muscle and fat.

Hepatic glucose production results from a combination of glycogenolysis and gluconeogenesis. Endogenous glucose production is also contributed by gluconeogenesis in the kidneys. Breakdown of stored hepatic glycogen is a readily available source of free glucose. However, in an average adult, this process can only provide less than an 8-hour supply of free glucose. (In infants, this may provide only 4 hours of free glucose.) Considering this

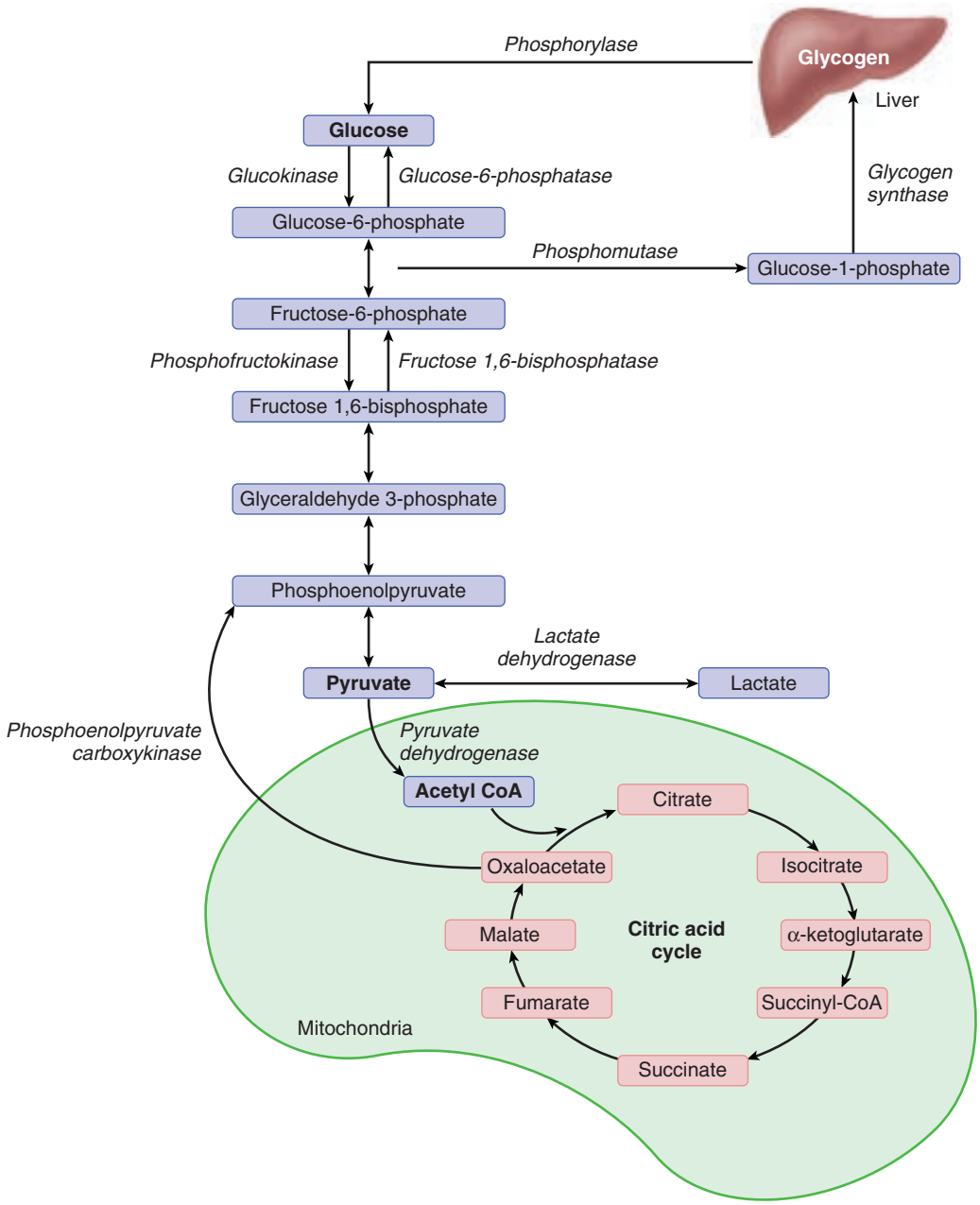

FIGURE 230-1. Outline of the biochemical pathways involved in glucose physiology.

limited capacity of glycogenolysis, gluconeogenesis is very important in supporting hepatic glycogen stores during an overnight fast.

Gluconeogenesis uses a number of key enzymes: pyruvate carboxylase, phosphoenolpyruvate carboxykinase (PEPCK), and fructose-1,6-bisphosphatase and its precursors, including lactate, alanine, glutamine, glycerol, and pyruvate. Muscle and adipose tissue, which utilize glucose in the fed state, respond to prolonged fasting by reducing their glucose uptake virtually to zero and satisfying their energy requirements by the β-oxidation of fatty acids. Additionally, through the process of proteolysis, muscle tissue provides amino acids to the liver to serve as gluconeogenic precursors for net glucose formation. Changes in the hormonal milieu during fasting (suppressed insulin and elevated counter-regulatory hormones) stimulate ketogenesis. Ketones become a major source of fuel for the brain when glucose utilization by the brain declines. This results in a decrease in the rate of gluconeogenesis required to maintain the plasma glucose concentration and hence in diminished protein wasting.

Changes During Feeding

After a meal, glucose absorption into the circulation increases glucose concentrations, which stimulates secretion of insulin from the pancreatic β-cells and suppresses secretion of glucagon from the pancreatic α-cells. This change in the hormonal milieu switches off endogenous hepatic glucose production and accelerates glucose utilization by liver, muscle, and adipose tissue. Glucose concentration then returns gradually to the postabsorptive level, at which endogenous glucose production is equal to the glucose uptake by peripheral tissues.

Counter-Regulatory Hormonal Responses to Hypoglycemia

The counter-regulatory hormones play a key role in the maintenance of normal blood glucose concentration.[2] If counter-regulation is intact, hypoglycemia (irrespective of the cause) will result in a decrease in insulin secretion and an increase in glucagon, epinephrine, norepinephrine, cortisol, and growth hormone (GH) secretion. Glucagon secretion increases rapidly in response to hypoglycemia, and studies have shown that the glucagon response is the primary essential defense mechanism against acute hypoglycemia. GH and cortisol have numerous effects on glucose metabolism, including increasing the rate of gluconeogenesis and antagonizing the effects of insulin. In adults, the glycemic thresholds for the activation of glucose counter-regulatory hormones such as GH and cortisol lie within or just below the physiologic blood glucose concentration and slightly higher than the threshold for symptoms. This suggests that GH and cortisol secretion increase in response to blood glucose concentrations within the normoglycemic range,

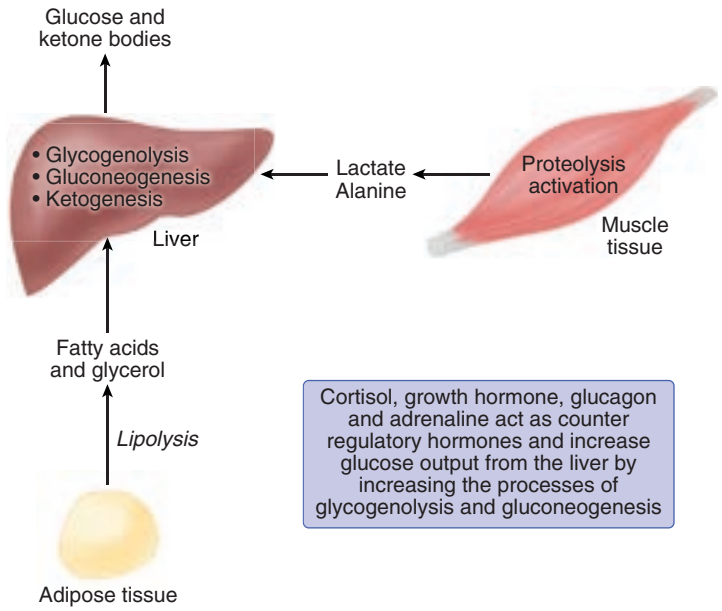

FIGURE 230-2. **The role of the counter-regulatory hormones, glycogenolysis, gluconeogenesis, and lipolysis in glucose physiology.**

and these increases are inversely proportional to the nadir in blood glucose. Figure 230-2 outlines the role of the counter-regulatory hormones.

Insulin secretion from β-cells of the pancreas in healthy individuals is inhibited as blood glucose concentration falls below 72 mg/dL (4.0 mmol/L). As insulin secretion is reduced, the repressive effect of insulin on pancreatic α-cell function is removed, thereby rapidly increasing glucagon secretion. Glucagon acts on the liver to increase hepatic glycogenolysis and gluconeogenesis. When the blood glucose concentration falls further (≈68 mg/dL [3.8 mmol/L]), epinephrine and norepinephrine are released both from the adrenals and directly into interstitial fluid from nerve terminals, further suppressing insulin secretion, increasing glucagon secretion, and decreasing peripheral glucose utilization in the muscle and increasing lipolysis in the adipose tissues.

Additional responses include GH and cortisol secretion, which occur below a blood glucose concentration of around 66 mg/dL (≈3.7 mmol/L) and are initiators of the adaptive response to hypoglycemia (e.g., during prolonged starvation); glucose-raising actions are much slower in onset (several hours). These hormone responses stimulate lipolysis, ketogenesis, and gluconeogenesis. Permissive amounts of cortisol and GH are required for a normal hepatic response to glucagon and epinephrine. In healthy individuals, this system ensures that hypoglycemia is rarely experienced and would only occur during starvation or ultra-endurance sports. Drugs or diseases that inhibit counter-regulatory secretion or action predispose patients to hypoglycemia.

Activation of counter-regulation depends on effective detection of falling blood glucose levels. This is achieved by the complex integration of various glucose-sensing systems in both the periphery and central nervous system.[3] Fluctuations in peripheral glucose levels are detected by glucose-sensing neurons in the oral cavity, gut, portal/mesenteric vein (PMV), and carotid body. PMV neurons detect changes in blood glucose prior to entry into the liver from the gut. This information is then relayed through the vagus nerve and spinal cord to the hindbrain and then to the hypothalamus. In addition, the hypothalamus, because of its location adjacent to the third ventricle and median eminence, may sample factors from peripheral circulation, including glucose, as well as hormones such as insulin and leptin. Although a complex network of glucose sensors has been described in the central nervous system and peripherally, the brain appears to have the dominant role during hypoglycemia and, specifically, the ventromedial region of the hypothalamus (VMH). VMH neurons contain the same glucose-sensing mechanisms (e.g., glucokinase, ATP-sensitive K^+ channels) as found in pancreatic β-cells.

CLINICAL MANIFESTATIONS

Symptoms of Hypoglycemia

The symptoms of hypoglycemia reflect the responses of the brain to a decrease in the blood glucose level; such symptoms may be nonspecific and vague, especially in the childhood period. Children may not be able to communicate their hypoglycemic symptoms. The symptoms of hypoglycemia may be categorized into two main groups: (1) those that arise as a result of the central nervous system being deprived of glucose (neuroglycopenic) and (2) symptoms arising from the perception of physiologic changes caused by the central nervous system–mediated sympatho-adrenal discharge triggered by hypoglycemia (neurogenic or autonomic).[4] The neurogenic symptoms of hypoglycemia are largely the result of sympathetic neural, rather than adrenomedullary, activation.

Neuroglycopenic symptoms (e.g., dizziness, confusion, tiredness, difficulty with speaking, headache, inability to concentrate, coma, and seizures) arise from the failure of brain function itself and are caused by deficient supply of glucose to the brain.[5] Neurogenic symptoms include both adrenergic responses (catecholamine-mediated symptoms such as palpitations, tremor, and anxiety) and cholinergic responses (acetylcholine-mediated symptoms such as sweating, hunger, paresthesias). Awareness of hypoglycemia chiefly depends on perception of the central and peripheral effects of neurogenic (as opposed to neuroglycopenic) responses to hypoglycemia.

In nondiabetic adults during acute insulin-induced hypoglycemia, autonomic symptoms become apparent at a threshold of approximately 60 mg/dL (3.3 mmol/L), and impairment of brain function manifested by neuroglycopenic symptoms occurs at a threshold of approximately 50 mg/dL (2.8 mmol/L) in arterialized venous blood (venous levels would be ≈ 3 mg/dL [0.16 mmol/L] less). However, in patients with recurrent hypoglycemia, the glycemic thresholds for responses to hypoglycemia are reset at a lower plasma glucose concentration. The glucose thresholds for the activation of neuroglycopenic and autonomic symptoms in children are not as clearly defined as in adults. The symptoms and signs of hypoglycemia are not influenced by the rate of blood glucose decline in nondiabetic individuals.

Clinical Approach to the Patient with Hypoglycemia

A careful clinical history, description of symptoms, physical examination, and a systematic step-by-step approach are the cornerstones of establishing a diagnosis. The symptoms of hypoglycemia may be very nonspecific, hence any symptomatic child or adult must have the blood glucose level measured and documented.

The relationship of a hypoglycemic episode to the most recent meal can be important diagnostically. Hypoglycemia occurring after a short fast (2 to 3 hours) may be suggestive of hyperinsulinism or glycogen storage disease. Hypoglycemia occurring after a long fast (12 to 14 hours) may suggest a disorder of gluconeogenesis. Postprandial hypoglycemia may indicate galactosemia, hereditary fructose intolerance, dumping syndrome, insulinoma, insulin autoimmune syndrome, and noninsulinoma pancreatogenous hypoglycemia syndrome. In both children and adults, a clear documentation of the medication history is important.

DIAGNOSIS

After the clinical history has been taken and the examination completed, a diagnostic cascade of appropriate tests is necessary. These may be guided in the context of the most common causes of hypoglycemia as listed in Table 230-1.

The current adult recommendations[6] state that evaluation and management of hypoglycemia should only be undertaken in patients in whom Whipple's triad—symptoms, signs, or both consistent with hypoglycemia, a low plasma glucose concentration, and resolution of those symptoms or signs after the plasma glucose concentration is raised—is documented. However, this does not apply to children for the reasons discussed earlier.

Causes of Hypoglycemia

Hypoglycemia is more common in the childhood period than in adults and can be due to a large number of causes. Table 230-1 summarizes the differential diagnosis of hypoglycemia.

Hypoglycemia Due to Excess Production of Hormones

Inappropriate and excess production of certain hormones can lead to hypoglycemia. The two most common conditions associated with excess production of a hormone are hyperinsulinemic hypoglycemia (HH) and non–islet cell tumor hypoglycemia (NICTH) or IGF-2-oma (insulin-like growth factor–secreting tumor). Inappropriate production of insulin can either lead to fasting hypoglycemia or postprandial hypoglycemia.

TABLE 230-1 DIFFERENTIAL DIAGNOSIS OF HYPOGLYCEMIA*

HYPERINSULINEMIC HYPOGLYCEMIA (INCLUDING POSTPRANDIAL)
Transient: infant of diabetic mother, perinatal asphyxia, Rhesus disease, intrauterine growth retardation, Beckwith-Wiedemann syndrome
Congenital: *ABCC8*, *KCNJ11*, ***GCK***, *GDH*, *HADH*, *HNF4A*, ***SLC16A1***
Dumping syndrome
Insulin receptor mutations and antibodies
Insulinoma
Noninsulinoma pancreatogenous hypoglycemia (adults)
Gastric bypass surgery for morbid obesity
Non–islet cell tumor hypoglycemia (NICTH) or IGF-2-oma
Insulin autoimmune syndrome
Insulin factitious hypoglycemia
HORMONAL DEFICIENCY/RESISTANCE
Adrenocorticotropic hormone
Cortisol
Growth hormone
Glucagon†
Adrenaline†
DEFECTS IN HEPATIC GLYCOGEN RELEASE/STORAGE
Glycogen storage diseases: **glucose-6-phosphatase**, **amylo-1,6-glucosidase deficiency**, liver phosphorylase deficiency, glycogen storage disease type 0
DEFECTS IN GLUCONEOGENESIS
Fructose-1,6-bisphosphatase deficiency, phosphoenolpyruvate carboxykinase deficiency, pyruvate carboxylase deficiency
CARNITINE METABOLISM
Carnitine deficiency (primary and secondary)
Carnitine palmitoyltransferase deficiency (CPT 1 and 2)
Carnitine transporter defects
FATTY ACID OXIDATION
Medium-chain acyl-CoA dehydrogenase (MCAD) deficiency
Very long-chain acyl-CoA dehydrogenase (VLCAD) deficiency
Short-chain acyl-CoA dehydrogenase (SCAD) deficiency
Long/short-chain L-3-hydroxyacyl-CoA (L/SCHAD) deficiency
DEFECTS IN KETONE BODY SYNTHESIS/UTILIZATION
HMG-CoA synthase deficiency, HMG-CoA lyase deficiency
Succinyl-CoA: 3-oxoacid-CoA transferase (SCOT) deficiency
METABOLIC CONDITIONS (COMMON ONES)
Organic acidemias (propionic, methylmalonic)
Maple syrup urine disease, galactosemia, fructosemia, tyrosinemia
Hereditary fructose intolerance
Mitochondrial respiratory chain complex deficiencies
Congenital disorders of glycosylation (CGD)
DRUG INDUCED
Sulfonylureas
Insulin
β-Blockers
Salicylates
Alcohol
Quinine
Haloperidol
Pentamidine
Levofloxacin
Methadone
Disopyramide
Indomethacin
Cibenzoline
Gatifloxacin
MISCELLANEOUS CAUSES (MECHANISM[S] NOT CLEAR)
Idiopathic ketotic hypoglycemia (diagnosis of exclusion)
Infections (sepsis, malaria), congenital heart disease

*Boldface indicates more common in adults.
†No human case yet reported with glucagon or adrenaline deficiency.
HMG = 3-hydroxy-3-methylglutaryl; IGF = insulin-like growth factor.

Hyperinsulinemic Hypoglycemia

HH is a heterogeneous group of disorders characterized by unregulated insulin secretion from pancreatic β-cells. In the face of hypoglycemia, patients have inappropriately detectable serum insulin levels, low ketone bodies, and low fatty acids and show a glycemic response to glucagon.[7]

Congenital Forms of Hyperinsulinemic Hypoglycemia

In patients with congenital forms of HH, mutations in the key genes (*ABCC8*, *KCNJ11*, *GLUD1*, *GCK*, *HADH*, *SLC16A1*, *HNF4A*, *HNF1A*, and *UCP2*) regulating insulin secretion have been identified.[8] Children with inactivating mutations in the genes *ABCC8* and *KCNJ11* present with the most severe forms of congenital HH, typically in the newborn period. Hyperinsulinism-hyperammonemia syndrome due to activating mutations in the *GLUD1* gene and activating mutations in the *GCK* gene, leading to HH, have both been described in adults as well as children. Exercise-induced HH due to activating mutations in the *SLC16A1* gene has also been recognized in adults.

Insulinoma

An insulinoma is the commonest cause of endogenous HH in adults. Insulinomas have the highest incidence in the fifth and sixth decades.[9] Insulinomas are insulin-secreting tumors of pancreatic origin, with an incidence of 1 to 4 per million. The majority (90%) of them are benign, solitary, intrapancreatic and less than 2 cm in diameter. Classically, symptoms become evident in the fasting state or following exercise. However, it is now known that insulinoma can also present with postprandial symptoms. Diagnosis is based on findings of abnormal serum levels of insulin and C-peptide (also proinsulin) at the time of fasting hypoglycemia. An insulinoma can occur either in isolation or in association with multiple endocrine neoplasia type 1 (MEN 1), with a lifetime prevalence of 10% among adults carrying mutations in *MEN1* (Chapter 231). Around 6% of insulinomas occur in patients with MEN 1, and most insulinomas are benign, but 5 to 10% are malignant.

Postprandial Hyperinsulinemic Hypoglycemia

Postprandial hyperinsulinemic hypoglycemia (PPHH) refers to hypoglycemia within a few hours of meal ingestion, secondary to inappropriate insulin secretion in response to a meal. If PPHH is clinically suspected, then an oral glucose tolerance test (OGTT) or a mixed-meal provocation test is performed. (See later in section "Investigations for Hypoglycemia.") A physiologic dip in the blood glucose level seen in OGTT might lead to misdiagnosis. However, corresponding biochemical evidence of endogenous HH and symptoms of neuroglycopenia during a hypoglycemic episode would help distinguish between pathologic PPHH and reactive hypoglycemia. A decrease of more than 108 mg/dL (6 mmol/L) between peak and nadir blood glucose during OGTT has been used as a diagnostic criterion for dumping syndrome in adults.

DUMPING SYNDROME. Dumping syndrome seen in infants after Nissen fundoplication is a classic example of PPHH. Precipitous emptying of hyperosmolar carbohydrate-containing solutions into the small bowel results in rapid glucose absorption, hyperglycemia, and reactive hypoglycemia. These children also tend to have abnormally exaggerated secretion of glucagon-like peptide-1 (GLP-1), which may contribute to the exaggerated insulin surge and resultant hypoglycemia.[10]

INSULIN AUTOIMMUNE SYNDROME. Insulin autoimmune syndrome, or Hirata's disease, is a rare condition characterized by HH associated with high titers of antibodies to endogenous insulin in the absence of pathologic abnormalities of pancreatic islets and prior exposure to exogenous insulin. The disease is extremely uncommon in Western countries. Insulin autoimmune syndrome affects men and women equally and is seen more frequently in patients older than 40 years. The binding kinetics of endogenous insulin by the antibodies are thought to lead to physiologically inappropriate levels of bioavailable insulin, causing either hyper- or hypoglycemia.

In this syndrome, the insulin levels are markedly elevated, usually above 100 mU/L. After a meal or glucose load, these patients often demonstrate initial hyperglycemia, followed by hypoglycemia a few hours later. The hyperglycemia is caused by the anti-insulin antibodies that bind the insulin secreted in response to rising blood glucose levels after a meal. This binding reduces the bioavailability of the secreted insulin to the receptors in the liver and peripheral tissues, resulting in hyperglycemia and further insulin secretion. As blood glucose concentrations begin to decrease and insulin secretion declines, the insulin bound to the antibodies is released, resulting in inappropriately high free insulin concentrations for the blood glucose, causing hypoglycemia.

PPHH IN PATIENTS WITH INSULIN-RECEPTOR MUTATIONS. PPHH has been described in patients who carried a heterozygote mutation (Arg-1174Gln) in the insulin-receptor gene. Hyperinsulinism seems to be associated with decreased degradation rather than increased secretion of insulin, as evidenced by increased fasting levels of serum insulin despite normal levels of serum C-peptide and reduced clearance of exogenous insulin during clamp studies.

PPHH AFTER GASTRIC BYPASS SURGERY. A consequence of the obesity epidemic is the increasing use of gastric bypass surgery for patients with severe, medically complicated obesity (Chapter 220), which has led to a number of reports of postprandial HH.[11] In a review of the Swedish Bariatric Surgery registry, the incidence of hospitalization for hypoglycemia in post–gastric bypass patients was reported as less than 1%.

A number of different explanations have been suggested to explain hypoglycemia post gastric bypass surgery. This can either be a manifestation of dumping syndrome or improved insulin sensitivity following weight loss unmasking an underlying hyperinsulinemia syndrome. The hypoglycemia could also be due to an effect on the enteroinsular axis induced by the diversion of nutrients into the small intestine.

The principal reason seems to be enhanced postprandial insulin secretion, thought to be due primarily to increased secretion of glucose-dependent insulinotropic polypeptide (GIP) and especially GLP-1. GLP-1 levels are now well documented to be increased two- to five-fold after gastric bypass. The elevations of incretins tend to be seen early, even as early as 2 days after gastric bypass, and levels may decline as substantial weight loss and normalization of insulin sensitivity occurs. In patients with PPHH, elevated levels of GIP and GLP-1 persist for years after surgery.

Increased postprandial insulin secretion by incretins is mediated by islet cell hypertrophy and hyperplasia. Both GIP and GLP-1 have been implicated in increasing pancreatic β-cell mass in rodent models. GLP-1 regulates islet growth by inducing the expression of the transcription factor pancreatic-duodenum homeobox-1 (PDX-1).

Overexpression of IGF-2 and IGF-1 receptor alpha (IGF1Rα) have been found in pancreatic tissue removed from patients with persistent PPHH after gastric bypass surgery as compared to controls. These findings are suggestive of the role of growth factors in islet hyperfunction seen in post–gastric bypass patients.

NONINSULINOMA PANCREATOGENOUS HYPOGLYCEMIA SYNDROME. Noninsulinoma pancreatogenous hypoglycemia syndrome (NIPHS) is characterized by postprandial neuroglycopenia in the presence of negative prolonged fasting tests and negative perioperative localization studies for insulinoma.[12] However, in some patients the selective arterial calcium stimulation test is positive, with the histology of the resected pancreas showing nesidioblastosis. The underlying genetic basis of NIPHS is not known.

These patients are negative for *ABCC8/KCNJ11* mutations and show islet hypertrophy histologically (as observed in diffuse congenital HH).

Immunohistologic studies of the resected pancreatic tissue have failed to show an increased rate of proliferation of β-cells or abnormal synthesis and/or processing of either proinsulin or amylin. Neither has there been any evidence of overexpression of pancreatic differentiation factors, PDX-1, and Nkx-6.1, nor the calcium-sensing receptor (CaSR).

Insulin Factitious Hypoglycemia

Hypoglycemia can also be induced pharmacologically, either intentionally as a diagnostic tool, accidentally as a complication of the treatment of diabetes mellitus, or as a consequence of poisoning either with insulin itself or with drugs (e.g., sulfonylureas) that stimulate insulin release. Whenever severe hypoglycemia occurs with documented hyperinsulinism, the possibility of Munchausen's syndrome by proxy should be considered in children. The possibility of malicious administration of insulin or an oral sulfonylurea should always be suspected in cases of sudden onset of hypoglycemia in a previously healthy individual. In the case of insulin administration, the clue in the biochemistry will be a raised insulin level accompanied by normal C-peptide.

Non–Islet Cell Tumor Hypoglycemia, or IGF-2-oma

NICTH, or IGF-2-oma, denotes the syndrome of hypoglycemia produced by or associated with any neoplasm other than an insulinoma. These are usually tumors of mesenchymal and epithelial origin (including hepatomas, fibromas, and fibrosarcomas). The underlying mechanism of hypoglycemia in nearly all patients with this syndrome is overproduction of IGF-2 by the tumor, which includes mature IGF-2 and incompletely processed forms of IGF-2, referred to collectively as "big" IGF-2.[13] The elevated IGF-2–related peptides mimic the fasting hypoglycemia characteristic of patients with insulin-producing islet cell tumors. Rarely, markedly elevated IGF-2 levels produce somatic changes suggestive of acromegaly. Typically, the elevated IGF-2 levels are associated with suppressed plasma levels of insulin, IGF-1, and GH.

Hypoglycemia Due to Hormone Deficiency

Deficiency of glucagon, adrenaline, GH, and cortisol can cause hypoglycemia. Glucagon and adrenaline deficiency is extremely rare, and so far no true human, genetically proven defects in glucagon and adrenaline deficiency have been described. Children and adults can present with hypoglycemia due to deficiency of various hormones. This might be either in isolation (e.g., isolated GH, adrenocorticotropic hormone [ACTH], or cortisol deficiency) or in combination with other hormones, such as in patients with hypopituitarism. The etiology of the hypoglycemia resulting from cortisol and GH deficiency is due to a combination of factors, including reduced gluconeogenic substrate availability (decreased mobilization of fats and proteins) and increased glucose utilization due to increased insulin sensitivity of tissues in the absence of these two hormones.

Acquired hypopituitarism may result from tumors (most commonly craniopharyngioma), radiation, infection, hydrocephalus, vascular anomalies, and trauma. Addison disease (AD) results from adrenal cortex hypofunction/dysfunction, with deficient production of glucocorticoids, mineralocorticoids, and androgens, and with high levels of both ACTH and plasma renin activity (Chapter 227). Autoimmune AD is the most frequent etiologic form in adult patients, accounting for about 80% of cases, followed by post-tuberculosis AD in 10 to 15%; the remaining 5% of cases are due to vascular, neoplastic, or rare genetic forms.

The markers of autoimmune AD are adrenal cortex (ACA) or 21-hydroxylase autoantibodies (21-OHAbs), and they are present at diagnosis in more than 90% of cases. In autoimmune AD, the adrenal cortex is infiltrated by lymphocytes and plasma cells, and the glands are sclerotic and reduced in volume. Autoimmune AD occurs mainly in middle-aged women, alone or associated with other (clinical, subclinical, or potential) autoimmune diseases, giving rise to various forms of autoimmune polyglandular syndrome. Replacement therapy with gluco- and mineralocorticoids is life-saving for patients with chronic adrenal insufficiency.

Hypoglycemia Due to Defects in Hepatic Glycogen Release/Storage

Glucose-6-phosphatase deficiency (glycogen storage disease [GSD] type I, Von Gierke disease) is the commonest of the glycogen storage diseases causing hypoglycemia (Chapter 207). Deficiency of this enzyme results in the inability to release free glucose from glucose-6-phosphate, with resultant hepatomegaly due to stored glycogen. These children and adults present with recurrent hypoglycemia associated with lactic acidosis, hyperuricemia, and hyperlipidemia.[14] The two other glycogen storage diseases causing hypoglycemia are due to deficiencies of the enzymes amylo-1,6-glucosidase (GSD type III) and liver phosphorylase (GSD type VI). The clinical and biochemical features of GSD-III subjects are quite heterogeneous.

Hypoglycemia Due to Defects in Gluconeogenesis

Gluconeogenesis, or the formation of glucose from mainly lactate/pyruvate, glycerol, glutamine, and alanine, plays an essential role in the maintenance of normoglycemia during fasting. Inborn deficiencies are known in each of the four enzymes of the glycolytic-gluconeogenic pathway that ensure a unidirectional flux from pyruvate to glucose: pyruvate carboxylase, phosphoenolpyruvate carboxykinase (PEPCK), fructose-1,6-bisphosphatase, and glucose-6-phosphatase. Gluconeogenesis can essentially be viewed as a reversal of glycolysis but with a few important differences. Patients with defects in gluconeogenesis present with fasting hypoglycemia and lactic acidosis. Pyruvate carboxylase deficiency may lead to a more widespread clinical presentation, with lactic acidosis, severe mental and developmental retardation, and proximal renal tubular acidosis.

Hypoglycemia Due to Disorders of Carnitine Metabolism and Defects of Fatty Acid Oxidation

Serious clinical consequences may occur if fatty acid oxidation (FAO) is impaired, including hypoglycemic seizures, muscle damage, cardiomyopathy, metabolic acidosis, and liver dysfunction. Fatty acids are taken up by

hepatocytes and muscle, where they are subsequently activated to their coenzyme A (CoA) esters. FAO disorders are individually rare, but they are collectively common because of the number of different enzymes affected. When defects occur in fatty acid degradation, excess acylcarnitine intermediates accumulate in the tissues, including heart, liver, and skeletal muscle, which can lead to organ dysfunction. The diversion of acyl-CoA intermediates into β-oxidation results in accumulation of toxic dicarboxylic acids. Acylcarnitines that spill into the blood provide a marker for diagnosis.

Primary carnitine deficiency is an autosomal recessive disorder of fatty acid oxidation that can present at different ages with hypoketotic hypoglycemia and cardiomyopathy and/or skeletal myopathy (Chapter 205). This disease is suspected based on reduced levels of carnitine in plasma and confirmed by measurement of carnitine transport in the patient's fibroblasts. Carnitine transport is markedly reduced (usually < 5% of normal) in fibroblasts from patients with primary carnitine deficiency. Patients with the hepatic isoform of carnitine palmityltransferase (CPT)-1 deficiency present with hypoketotic hypoglycemia in the neonatal period.

The commonest disorder of fatty acid β-oxidation is medium-chain acyl-CoA dehydrogenase (MCAD) deficiency, an autosomal recessive disease presenting in children who are typically asymptomatic except during times of fasting and metabolic stress, usually associated with a viral illness, when they present with fasting nonketotic hypoglycemia; if undiagnosed, 20 to 25% of affected patients will die during the first episode.

Metabolic Diseases

Hypoglycemia can also be due to a number of metabolic conditions (Chapter 205), including galactosemia, fructosemia, tyrosinemia, organic acidemias, maple syrup urine disease, glutaric aciduria type II, and in mitochondrial respiratory chain defects. Hereditary fructose intolerance, caused by catalytic deficiency of aldolase B (fructose-1,6-phosphate aldolase), is a recessively inherited condition in which affected homozygotes develop hypoglycemia and severe abdominal symptoms after taking foods containing fructose and cognate sugars. Continued ingestion of noxious sugars leads to hepatic and renal injury and growth retardation.

Noninsulinoma Islet Cell Tumors

Islet cell tumors present an important challenge to the clinician because of their protean manifestations and potential lethality. These tumors can be clinically silent or active (functioning). Early diagnosis is essential and depends on recognition of the classic and variant clinical syndromes followed by confirmation of elevated peptide levels by radioimmunoassay.[15] Glucagonoma, gastrinoma, VIPoma (VIP = vasoactive intestinal peptide), somatostatinoma, and ACTHoma are functioning tumors that may occur in isolation but can also be part of MEN 1 syndrome (Chapter 231) and von Hippel-Lindau disease (Chapter 417).

Tumor marker measurement gives useful information for the follow-up and management of patients with noninsulinoma islet cell tumors (neuroendocrine tumors). The currently used tumor markers are neuron-specific enolase (NSE) and chromogranin A (CgA). The clinical accuracy of these biomarkers depends on histotype and disease extent. CgA is thought to be the optimal marker for most neuroendocrine tumors, because it is independent of the biological characteristics of the tumor.

Glucagonoma

Glucagonomas are α-cell tumors that, when they are active, produce a syndrome characterized by necrolytic migratory erythema, diabetes mellitus, weight loss, anemia, glossitis, thromboembolism, neuropsychiatric disturbances, and hyperglucagonemia. Tumor characterization is made by computed tomography (CT) and/or pancreatic endoscopic ultrasonic and indium-labeled octreoscan. The diagnosis is established by documenting the presence of hyperglucagonemia, with diagnostic levels being generally above 500 pg/mL (normal, <120). It is important to remember that other diseases can also cause hyperglucagonemia, including cirrhosis, pancreatitis, diabetes mellitus, prolonged fasting, sepsis, burns, renal failure, acromegaly, and familial hyperglucagonemia. Surgery is the main component of treatment, in some cases in association with chemotherapy.

Gastrinoma

Gastrinomas are uncommon tumors of the endocrine system, occurring within the pancreas and duodenum. Overproduction of the hormone gastrin by these tumors produces a sustained increase in gastric acid secretion, leading to complications of peptic ulceration known as the Zollinger-Ellison syndrome (ZES). Gastrinomas can occur sporadically or in a familial pattern as a component of the MEN 1 syndrome. Gastrinomas have the potential to metastasize to regional lymph nodes, the liver, and other distant sites.

VIPoma

VIPoma is very rare, with 80% of these tumors originating from the pancreas, mostly in the tail. The majority of cases are sporadic. About 50 to 60% of cases have metastasized by the time the diagnosis is made. Most patients have secretory watery diarrhea, resulting in electrolyte disturbances, such as hypokalemia, hypophosphatemia, hypomagnesemia, and metabolic acidosis (Verner-Morrison syndrome, pancreatic cholera, WDHA syndrome). Hypochlorhydria or achlorhydria occurs in 75% of cases, owing to the inhibition of gastric acid production by VIP. Hyperchloremic acidosis can also occur as a result of low bicarbonate levels from severe intestinal loss. Occasionally, hypercalcemia, glucose intolerance, and hypotension may be present. The VIP level is elevated in almost all cases, but it can also be normal between episodes of diarrhea.

Somatostatinoma

Somatostatinomas are rare neuroendocrine tumors with an incidence of 1 in 40 million. These unusual tumors arise predominantly in the pancreas and peripancreatic duodenum, and patients often present with nonspecific symptoms. Rarely, patients present with somatostatinoma syndrome (diabetes, gallstones, and steatorrhea) when the tumor is secretory.

Investigations for Hypoglycemia

From the clinical history, description of symptoms, and physical examination, there might be important clues to the underlying cause of hypoglycemia, and the investigations can then be tailored to the particular cause. However, in some cases the clinical history and physical examination may not provide any clues, and in these cases the patient will need to be investigated more extensively.

Reagent strips in combination with a reflectance meter are the most common method of measuring bedside blood glucose levels. However, it is important to remember that these should be used only as a guide (they can be inaccurate), and the blood glucose concentration should always be checked in the laboratory. Whole-blood glucose is approximately 15% lower than serum glucose levels because of the lower glucose content and intracellular water content of the red cells. Glucose concentrations in venous blood are 10% lower than arterial blood. The blood sample for glucose measurement should be collected in a fluoride container to inhibit glycolysis. It should also be analyzed immediately because, even in the presence of fluoride, the blood glucose concentration will decrease over time.

In an ideal situation, the blood glucose level should be measured at the time of a spontaneous episode of hypoglycemia, and samples for plasma glucose, insulin, C-peptide, proinsulin, and β-hydroxybutyrate concentrations and toxicology screen for oral hypoglycemic agents taken. The blood glucose must be considered in the context of the whole fuel economy and in the light of concurrent hormone concentrations. However, this is not always possible, and patients may require further tests (e.g., fasting, mixed-meal, or provocation testing) to unravel the cause of the hypoglycemia. The various tests used for eliciting hypoglycemia in adults and children are described next. Table 230-2 shows the routine baseline investigations that should be performed in children and adults presenting with hypoglycemia.

TABLE 230-2 ROUTINE BASELINE INVESTIGATIONS IN PATIENTS WITH SUSPECTED HYPOGLYCEMIA

BLOOD	URINE
Glucose	Ketones
Insulin	Reducing substances
Cortisol	Organic acids
Lactate	
Growth hormone	
Nonesterified fatty acids	
3β-Hydroxybutyrate	
Carnitine (free and total)	
Blood spot acylcarnitine	
Ammonia	

TABLE 230-3 PROTOCOL FOR 72-HOUR FAST IN ADULTS

1. Start the fast from the last ingestion of a meal. Stop all medications that might interfere with test.
2. The patient can drink water during the test.
3. The patient must be active during waking hours.
4. Measure plasma glucose, insulin, C-peptide, and β-hydroxybutyrate (on the same venipuncture specimen) every 6 hours until plasma glucose reaches 60 mg/dL (3.3 mM). Then measure every 1 to 2 hours.
5. End the fast when the plasma glucose is 45 mg/dL (2.5 mM) and the patient has symptoms or signs of hypoglycemia, or plasma glucose is 55 mg/dL if Whipple triad had been demonstrated previously.
6. At the end of the fast, measure plasma glucose, insulin, C-peptide, β-hydroxybutyrate, and sulfonylurea (on the same venipuncture specimen). Then inject glucagon, 1 mg intravenously, and measure plasma glucose every 10 minutes three times. Once the fast is completed, allow the patient to eat normally.

Adapted from Cryer PE, Axelrod L, Grossman AB, et al. Evaluation and management of adult hypoglycemic disorders: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab.* 2009;94:709-728.

TABLE 230-4 SUGGESTED PROTOCOL FOR A MIXED-MEAL DIAGNOSTIC TEST IN ADULTS

1. Fast patient overnight. Stop all medications that might interfere with test.
2. Use a mixed meal similar to one that causes patient to experience symptoms.
3. Collect samples for plasma glucose, insulin, C-peptide, and proinsulin before ingestion and every 30 minutes through 300 minutes after meal ingestion.
4. Observe the patient for symptoms and/or signs of hypoglycemia, and ask the patient to keep a written log of all symptoms, timed from the start of meal ingestion.
5. The mixed-meal test should be interpreted on the basis of laboratory measured plasma glucose concentrations, not those estimated with a point-of-care glucose monitor. If it is judged necessary to treat before 300 minutes because of severe symptoms, obtain samples for all the following *before* administering carbohydrates: plasma insulin, C-peptide, and proinsulin (sent for analysis only in those samples in which plasma glucose is < 60 mg/dL [3.3 mmol/L]), and a measurement of oral hypoglycemic agents. If Whipple triad is demonstrated, antibodies to insulin should also be measured.

Adapted from Cryer PE, Axelrod L, Grossman AB, et al. Evaluation and management of adult hypoglycemic disorders: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab.* 2009;94:709-728.

Fasting Tests

Controlled fasting tests are important procedures for eliciting the cause of hypoglycemia in both children and adults. In adults it is recommended that a prolonged supervised (72-hour) fast be conducted in a standardized fashion (Table 230-3). During the fast, if patients have any signs or symptoms of hypoglycemia, with a documented low blood glucose level, the fast should be terminated. It is currently recommended that the fast not be prolonged beyond 72 hours if patients do not have any symptoms or signs of hypoglycemia and no documentation of low blood glucose. Monitoring for symptoms or signs of hypoglycemia during the fast is essential because patients may experience some symptoms but have serum glucose levels higher than the hypoglycemic range. In some healthy women (thin and lean) and men, blood glucose levels may drop to 40 mg/dL (2.2 mmol/L) during prolonged fasting. Some patients have lower glycemic thresholds without symptoms or signs of hypoglycemia.

In an adult patient where Whipple triad has been demonstrated, the 72-hour fast may be terminated if the plasma glucose concentration is 55 mg/dL (3 mmol/L) or less. The interpretation of serum insulin, C-peptide, and proinsulin during the 72-hour fast will depend on the concurrent plasma glucose concentration. Pancreatic β-cell insulin secretion becomes undetectable in healthy persons when the plasma glucose concentration is down to 55 mg/dL (3 mmol/L). Most patients with an insulinoma become hypoglycemic before 72 hours. However, continuation of the fast to 72 hours is necessary to rule out the likelihood of organic hypoglycemia.

HH in adults is characterized by plasma insulin concentrations of 3 μU/mL or greater (C-peptide of 200 pmol/L or more and proinsulin 5 pmol/L or more). Insulinoma patients have plasma insulin concentrations that rarely exceed 100 μU/mL, and plasma insulin levels greater than 1000 μU/mL suggest exogenous insulin administration or the presence of insulin antibodies. In the childhood period, any detectable plasma insulin in the presence of hypoglycemia is inappropriate and is highly suggestive of HH.

Measurement of the plasma β-hydroxybutyrate concentration is used as a surrogate marker of insulin action at the end of the 72-hour fast in healthy adult individuals and when Whipple triad is fulfilled in patients during a diagnostic fast. In patients with HH, because of the suppressive action of insulin on ketogenesis, plasma concentrations of β-hydroxybutyrate are typically less than 2.7 mmol/L. This is in contrast to healthy individuals, who will show a progressive rise in the concentration of β-hydroxybutyrate during the 72-hour test. If a value above 2.7 mmol/L is documented at any time point in the fast, the test can be terminated. In the childhood period, there is no clear-cut plasma level of β-hydroxybutyrate that can be used to confirm HH.

Another useful marker of insulin action in both adults and children is the glycemic increment in response to an intravenous/intramuscular injection of glucagon (1-mg dose in adults). Patients with HH will have increased glycogen stores, and giving glucagon will result in glycogenolysis. A positive response is defined as a maximal increment at least 25 mg/dL (1.3 mmol/L) greater than the terminal fasting serum glucose.

During the 72-hour prolonged fast, blood should also be collected for measurement of plasma sulfonylureas and meglitinides if the patient develops hypoglycemia. Sulfonylureas stimulate pancreatic β-cell insulin and C-peptide secretion, and the biochemical pattern is similar to that of an insulinoma.

Mixed-Meal Test

A mixed-meal test is performed in patients in whom there is a history suggestive of neuroglycopenic symptoms for up to 5 hours after food ingestion (Table 230-4). A positive test is defined as the onset of neuroglycopenic symptoms in association a documented low blood glucose level (e.g., ≤50 mg/dL). In postprandial HH, plasma insulin and C-peptide levels might be inappropriately elevated. Neuroglycopenic symptoms after a meal are reported in patients with insulinoma, patients with NIPHS, and patients who have undergone surgery for obesity.

The combination of a positive mixed-meal test and a negative 72-hour fast may occur in a patient with insulinoma or with NIPHS. The 5-hour oral glucose tolerance test should not be used as a diagnostic test for hypoglycemia, because a substantial percentage of healthy persons may have a serum glucose concentration of 50 mg/dL (2.7 mmol/L) or less.

Insulin Antibodies

Insulin autoimmune syndrome (IAS) is an uncommon cause of HH characterized by autoantibodies to endogenous insulin in individuals without previous exposure to exogenous insulin. IAS is the third leading cause of spontaneous hypoglycemia in Japan, and is increasingly being recognized worldwide in non-Asian populations. In patients with insulin autoimmune hypoglycemia from the spontaneous generation of insulin antibodies, such antibodies may be monoclonal or polyclonal and are present in very high titers, in contrast to the much lower titers in insulin-treated diabetes. It is important to test for the presence of insulin antibodies, because even low titers—which may have no diagnostic significance—may cause spurious results of the assay for insulin. Hypoglycemia due to IAS typically occurs in the fasting period but can occur postprandially as well.

Radiologic Investigations

Noninvasive imaging procedures such as CT and magnetic resonance imaging (MRI) are used when a diagnosis of insulinoma has been made to localize the source of pathologic insulin secretion. Invasive modalities, such as endoscopic ultrasonography (EUS) and arterial stimulation venous sampling (ASVS), are highly accurate in the preoperative localization of insulinomas and have frequently been shown to be superior to noninvasive localization techniques.[16] Intraoperative manual palpation of the pancreas by an experienced surgeon and intraoperative ultrasonography are both sensitive methods with which to finalize the location of insulinomas.

The sensitivity of transabdominal ultrasonography in the localization of insulinomas is poor (ranging from 9% to 64%). However, insulinomas demonstrate characteristic features when imaged with both CT and MRI, and the sensitivity of these techniques is 33 to 64% and 40 to 90%, respectively. The sensitivity and specificity of MRI is generally superior to that of CT, as is the detection of extrapancreatic extensions. Insulinomas generally demonstrate low signal intensity on T1-weighted images and high signal intensity on T2-weighted images.

Invasive modalities such as EUS and ASVS have been shown to be highly accurate in the preoperative localization of insulinomas and have frequently been shown to be superior to noninvasive localization techniques. EUS is

currently the test of choice in most Western centers, with reported detection rates of 86.6 to 92.3%.

ASVS has greatly facilitated the precise regionalization of insulinomas smaller than 2 cm, which noninvasive techniques like ultrasonography, CT, and MRI often fail to localize. This test requires access to intra-abdominal vessels, including the right hepatic vein, splenic artery, gastroduodenal artery, and superior mesenteric artery. A two- to three-fold increase in insulin concentration in the right hepatic vein in response to calcium injection into one or more of the arteries supplying the pancreas suggests that the region served by that artery may harbor abnormally functioning β-cells, whether from insulinoma or islet hypertrophy or nesidioblastosis. Calcium injection will stimulate a brisk response of insulin, C-peptide, and proinsulin simultaneously, and the magnitude of increase of both insulin and C-peptide appears to be correlated well with the degree of differentiation of the tumor cells.

Insulinomas have been shown to express GLP-1R in high density, and GLP-1R imaging has been used in a few patients for insulinoma localization. Fluorine-18-L-dihydroxyphenylalanine (18F-DOPA) positron emission tomography (PET) has also been used in localizing neuroendocrine tumors. In children with congenital HH, 18F-DOPA PET/CT is the gold standard for localizing focal lesions prior to surgery. 18F-DOPA PET has been found to be useful in some patients with insulinoma who had negative CT, MRI, and ultrasound results.

The positive responses to selective arterial calcium stimulation in some patients with NIPHS, despite negative radiologic localizing studies, establish that this technique should be performed in all adults with HH of unknown etiology.

Protein/Leucine Sensitivity Testing

Protein sensitivity is observed in some patients with congenital HH. Typically, mutations in the genes *ABCC8/KCNJ11*, *GDH*, and *HADH* lead to protein-induced hypoglycemia. These patients will demonstrate severe hypoglycemia in response to a protein or leucine load.

Exercise Test

In some patients, exercise can trigger unregulated insulin secretion. This is referred to as exercise-induced hyperinsulinism. Promoter-activating mutations induce expression of the *SLC16A1* gene in β-cells, where this gene is not usually transcribed, permitting pyruvate uptake and pyruvate-stimulated insulin release despite ensuing hypoglycemia. A physical exercise test will identify this group of patients.

Genetic Studies

Genetic testing should be undertaken in children who have been diagnosed with congenital HH and other causes of hypoglycemia that might have a genetic basis. Mutation in the genes *ABCC8/KCNJ11* is the commonest cause of medically unresponsive congenital HH. All patients with insulinoma and noninsulinoma islet cell tumors should be tested for mutations in the *MEN1* gene.

TREATMENT

The correct management of hypoglycemia will depend on the underlying cause. Therefore, establishing the correct diagnosis is fundamentally important in both children and adults.

Emergency Management

Acute management of hypoglycemia involves giving a bolus of intravenous glucose (in adults a bolus of 50% dextrose) to correct the blood glucose level. This will then need to be followed by an infusion of 10% dextrose to maintain normoglycemia. Glucagon can be used in the emergency management of hypoglycemia (in emergency, give 1 mg stat intramuscularly). Glucagon can cause rebound hypoglycemia, so the patient will need blood glucose monitoring after the administration of glucagon.

Management of Specific Causes of Hypoglycemia

The long-term management of hypoglycemia depends on the underlying cause. Below is a summary of the management of the different types of hypoglycemia.

Hyperinsulinemic Hypoglycemia

Diazoxide (5 to 20 mg/kg/day given orally three times daily) is the first-line medical therapy in children and adults with HH. Fluid retention is a major side effect. Diazoxide may be combined with a diuretic to reduce the side effect of fluid retention. Second-line therapies include the use of octreotide (5 to 35 μg/kg/day as an infusion or injection 3 or 4 times daily) and glucagon (1 to 10 μg/kg/hour given as a subcutaneous or intravenous infusion).

For adult patients with insulinoma, pancreatectomy is the treatment of choice.[17] However, there are reports of insulinoma in adults responding to therapy with diazoxide and octreotide (including long-acting octreotide). The mTOR inhibitor everolimus has been reported to be effective in controlling hypoglycemia in patients with malignant insulinomas or those who cannot undergo surgical resection. Adult patients with PPHH after gastric bypass surgery can be treated with diazoxide and octreotide, but some will require pancreatectomy.[18] The insulin autoimmune syndrome will respond to therapy with glucocorticoids, but some patients have responded to diazoxide and octreotide. NICTH or IGF-2-oma patients will require resection of the primary tumor.

Hypoglycemia Due to Hormonal Deficiencies

Adult patients and children with GH and cortisol deficiency will require replacement therapy with recombinant GH and hydrocortisone (prednisolone), respectively. Replacement therapy with gluco- and mineralocorticoids can be life-saving in patients with adrenal insufficiency.

Hypoglycemia Due to Glycogen Storage Diseases

Patients with hypoglycemia due to disorders of hepatic glycogen storage and release need to avoid prolonged periods of fasting. Children will require overnight continuous feeding. Raw, uncooked cornstarch is commonly used as a slow-release source of glucose and helps with prolonging the period of fasting.

Hypoglycemia Due to Defects in Fatty Acid Oxidation, Disorders of Gluconeogenesis, and Disorders of Ketone Body Metabolism

Principles similar to those already discussed also apply to these patients. However, patients with fatty acid oxidation disorders should have carnitine supplementation.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ludwig A, Ziegenhorn K, Empting S, et al. Diabetes Patienten-Verlaufsdokumentationssystem (DPV) Group, Mohnike K. Glucose metabolism and neurological outcome in congenital hyperinsulinism. *Semin Pediatr Surg*. 2011;20:45-49.
2. Tesfaye N, Seaquist ER. Neuroendocrine responses to hypoglycemia. *Ann N Y Acad Sci*. 2010;1212:12-28.
3. Watts AG, Donovan CM. Sweet talk in the brain: glucosensing, neural networks, and hypoglycemic counterregulation. *Front Neuroendocrinol*. 2010;31:32-43.
4. Chan O, Sherwin R. Influence of VMH fuel sensing on hypoglycemic responses. *Trends Endocrinol Metab*. 2013;24:616-624.
5. McCrimmon RJ. Update in the CNS response to hypoglycemia. *J Clin Endocrinol Metab*. 2012;97:1-8.
6. Cryer PE, Axelrod L, Grossman AB, et al. Endocrine Society. Evaluation and management of adult hypoglycemic disorders: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2009;94:709-728.
7. Douillard C, Mention K, Dobbelaere D, et al. Hypoglycaemia related to inherited metabolic diseases in adults. *Orphanet J Rare Dis*. 2012;7:26.
8. Senniappan S, Shanti B, James C, et al. Hyperinsulinaemic hypoglycemia: genetic mechanisms, diagnosis and management. *J Inherit Metab Dis*. 2012;35:589-601.
9. Okabayashi T, Shima Y, Sumiyoshi T, et al. Diagnosis and management of insulinoma. *World J Gastroenterol*. 2013;19:829-837.
10. Salehi M, Gastaldelli A, D'Alessio DA. Blockade of glucagon-like peptide 1 receptor corrects postprandial hypoglycemia after gastric bypass. *Gastroenterology*. 2014;146:669-680.
11. Milone M, Di Minno MN, Leongito M, et al. Bariatric surgery and diabetes remission: sleeve gastrectomy or mini-gastric bypass? *World J Gastroenterol*. 2013;19:6590-6597.
12. Galati SJ, Rayfield EJ. Approach to the patient with postprandial hypoglycemia. *Endocr Pract*. 2014;20:331-340.
13. Dynkevich Y, Rother KI, Whitford I, et al. Tumors, IGF-2, and hypoglycemia: insights from the clinic, the laboratory, and the historical archive. *Endocr Rev*. 2013;34:798-826.
14. Douillard C, Mention K, Dobbelaere D, et al. Hypoglycemia related to inherited metabolic diseases in adults. *Orphanet J Rare Dis*. 2012;7:26.
15. Ito T, Igarashi H, Jensen RT. Pancreatic neuroendocrine tumors: clinical features, diagnosis and medical treatment: advances. *Best Pract Res Clin Gastroenterol*. 2012;26:737-753.
16. Okabayashi T, Shima Y, Sumiyoshi T, et al. Diagnosis and management of insulinoma. *World J Gastroenterol*. 2013;19:829-837.
17. Partelli S, Maurizi AI, Tamburrino D, et al. GEP-NETS update: a review on surgery of gastro-entero-pancreatic neuroendocrine tumors. *Eur J Endocrinol*. 2014;171:R153-R162.
18. Iglesias P, Díez JJ. Management of endocrine disease: a clinical update on tumor-induced hypoglycemia. *Eur J Endocrinol*. 2014;170:R147-R157.

REVIEW QUESTIONS

1. Which of the following is true in relation to glucose physiology?

A. Insulin stimulates gluconeogenesis.
B. Glycogenolysis predominantly takes place in the kidney.
C. Lipolysis is dependent on ketogenesis.
D. Glucagon stimulates glycogenolysis.
E. Pancreatic beta cells stop producing insulin when the blood glucose level reaches 2.4 mmol/L.

Answer: D Glucagon stimulates glycogenolysis and releases stored glucose.

2. In the counter-regulatory hormonal response to hypoglycemia:

A. Growth hormone is the first hormone to increase in response to a falling blood glucose level.
B. Serum cortisol levels increase in parallel to glucagon.
C. Adrenaline and noradrenaline are the first line of defense against a falling blood glucose level.
D. Pancreatic delta cells produce glucagon in response to hypoglycemia.
E. The main glucosensors for hypoglycemia are located in the carotid body.

Answer: C Adrenaline and noradrenaline are the first line of defense against a falling blood glucose level. Adrenaline and noradrenaline with glucagon are the key counter-regulatory hormones produced in response to a falling blood glucose level.

3. Which of the following does *not* constitute Whipple triad?

A. A low measured plasma glucose concentration
B. Symptoms and/or signs compatible with hypoglycemia
C. Occurrence of any symptoms and/or signs during hypoglycemia
D. Resolution of symptoms and signs when glucose concentrations are corrected

Answer: C Whipple triad includes a low measured plasma glucose concentration, symptoms and/or signs compatible with hypoglycemia, and resolution of symptoms and signs when glucose concentrations are corrected.

231

POLYGLANDULAR DISORDERS

LYNNETTE K. NIEMAN AND ALLEN M. SPIEGEL

DEFINITION AND CLINICAL SIGNIFICANCE

Polyglandular syndromes are disorders in which there is dysfunction and pathology of more than one endocrine gland. These disorders can be classified into (a) neoplastic syndromes in which there is abnormal endocrine cell proliferation, and often, but not invariably, hormone hypersecretion, and (b) into autoimmune syndromes in which there is evidence of immune destruction of endocrine cells, often resulting in hypofunction and reduced hormone secretion. Both the neoplastic and autoimmune polyglandular syndromes often have nonendocrine manifestations that are relatively syndrome specific.

With few exceptions, the polyglandular syndromes are due to germline mutations of key growth regulatory genes (neoplastic syndromes) or immune regulatory genes (autoimmune syndromes).[1] It is important to recognize these disorders and differentiate them from sporadic single–endocrine gland diseases for several reasons. First, recognition of a specific polyglandular syndrome should alert the clinician to look for other endocrine and extra-endocrine manifestations of the syndrome. Although some patients will present with multiple endocrine gland manifestations, some will initially present with only a single endocrine gland affected. Careful family history and screening for other endocrine and characteristic extra-endocrine manifestations is needed in such cases. Second, treatment of polyglandular disease may differ from treatment of individual gland disease. Third, because of the

TABLE 231-1 POLYGLANDULAR NEOPLASIA SYNDROMES

SYNDROME	GENETIC BASIS*	ENDOCRINE TUMORS	NONENDOCRINE FEATURES
MEN 1	*MEN1*	Parathyroid Anterior pituitary Pancreatic islet	Subcutaneous lipomas Skin collagenomas
MEN 2 (A and B)	*RET*	Medullary thyroid cancer Pheochromocytoma Parathyroid (2A)	Mucosal neuromas (2B) Megacolon (2B)
MEN 4	*CDNK1B*	Anterior pituitary Parathyroid	Renal tumors
Carney complex	*PKAR1A*	Adrenal cortex Anterior pituitary Thyroid	Atrial myxomas Skin lentigines
von Hippel-Lindau disease	*VHL*	Pheochromocytoma Pancreatic islet	Renal cell cancer CNS hemangioblastoma
McCune-Albright syndrome	*GNAS* (mosaic)	Thyroid Anterior pituitary Adrenal cortex Gonads	Fibrous dysplasia Café au lait skin lesions

*With the exception of McCune-Albright, all syndromes are caused by heterozygous germline mutations of the gene listed and show autosomal dominant inheritance.
CA = cancer; CNS = central nervous system.

genetic basis of most of these syndromes, taking a careful family history and, in some cases, screening other family members to allow disease prevention in affected individuals is indicated.

This chapter discusses the best-characterized polyglandular disorders. Other chapters on the anterior pituitary (Chapter 224), thyroid (Chapter 226), adrenal cortex (Chapter 227), adrenal medulla (Chapter 228), pancreatic islets (Chapters 195 and 230), and parathyroids (Chapter 245) should be consulted for more detailed discussion of the diseases of individual glands.

NEOPLASTIC SYNDROMES

Six distinct neoplastic syndromes involve more than one endocrine gland. These include multiple endocrine neoplasia type 1 (MEN 1), multiple endocrine neoplasia types 2A (MEN 2) and 2B (sometimes referred to as MEN 3), multiple endocrine neoplasia type 4 (MEN 4), Carney complex,[2] von Hippel-Lindau disease (VHL), and McCune-Albright syndrome (MAS) (Table 231-1).[3] All but the latter are caused by heterozygous germline mutations and are inherited in autosomal dominant fashion. The genes responsible for MEN 1 and 4, Carney complex, and von Hippel-Lindau disease act as tumor suppressor genes. Germline loss-of-function mutations in one allele are followed by somatic mutations inactivating the second normal allele, leading to tumorigenesis. The basis for the tissue-specific expression of both the endocrine and extra-endocrine manifestations of these syndromes is not well understood, but the germline nature of the mutation and the expression of the affected gene in more than one endocrine gland explains the polyglandular aspect of the disorder. MEN 2A and B, in contrast, are caused by germline activating mutations of an oncogene, *RET*. The pattern of expression of this gene, involving chromaffin cells, helps explain the specific clinical manifestations. McCune-Albright syndrome is caused by a somatic rather than germline mutation that constitutively activates the ubiquitously expressed *GNAS* gene. This mutation, which may occur early in embryogenesis, leads to unregulated cyclic adenosine monophosphate (cAMP) formation in affected cells. The resultant mosaic distribution of the mutant gene helps explain the pleiotropic manifestations of the disease.

Patients with polyglandular neoplastic syndromes typically present with their respective endocrine tumors at a younger age than patients with single-gland sporadic endocrine tumors. Treatment of the polyglandular neoplastic syndromes, both in terms of the neoplastic component and the hormone hypersecretion, poses greater challenges than treatment of individual endocrine tumors.[4] For disorders with high risk of fatal cancer, such as medullary thyroid cancer in MEN 2, early genetic diagnosis and prophylactic surgical removal of the thyroid is indicated.[5] In other disorders such as MEN 1 and Carney complex, less aggressive approaches such as selective tumor resection and pharmacologic treatment to reduce hormone hypersecretion may be more appropriate. More detailed discussion of the clinical features, diagnosis, and treatment of these neoplastic syndromes may be found in other chapters (MEN 1 in Chapter 245; MEN 2 and 3 in Chapter 246; McCune-Albright syndrome in Chapters 233 and 235).

TABLE 231-2 CLINICAL FEATURES OF AUTOIMMUNE POLYGLANDULAR SYNDROMES

FEATURE	TYPE 1	TYPE 2
Mucocutaneous candidiasis	Very common	Not seen
Hypoparathyroidism	Common	Rare
Addison's disease	Common	Common
Primary hypogonadism	Common	Occurs
Autoimmune thyroid disease	Rare	Common
Autoimmune diabetes	Occurs	Common
Hypophysitis	Occurs	Occurs
Autoimmune hepatitis	Occurs	Not seen
Pernicious anemia	Occurs	Occurs
Vitiligo	Occurs	Occurs
Malabsorption syndrome	Occurs	Occurs as celiac disease
Alopecia	Common	Occurs
Myasthenia gravis	Not seen	Occurs
Keratopathy	Common	Not seen
Tympanic membrane calcification	Common	Not seen
Inheritance	Autosomal recessive	HLA association
Age at onset	Usually childhood	Usually adulthood

HLA = human leukocyte antigen.

AUTOIMMUNE SYNDROMES

Organ-specific autoimmune disease, characterized by lymphocytic infiltration and organ-specific autoantibodies, commonly results in endocrine hypofunction. Not uncommonly, however, disorders of more than one endocrine gland appear in families or individual patients. Characteristic patterns of disease presentation and genetic inheritance allow the definition of two syndromes with overlapping manifestations (Table 231-2).[6]

Autoimmune Polyglandular Syndrome Type 1

DEFINITION

Autoimmune polyglandular syndrome (APS) type 1 is a rare disease that is also known as autoimmune polyendocrinopathy, candidiasis, and ectodermal dystrophy syndrome. It typically manifests in early childhood.

PATHOGENESIS

APS type 1 is an autosomal recessive disorder caused by a variety of inactivating mutations in the gene encoding autoimmune regulator-1 (AIRE-1),[7] which controls the expression of autoantigens by medullary epithelial cells of the thymus. These antigens are also expressed in peripheral tissues. Their expression in the thymus is important for negative selection (elimination) of autoreactive T cells, which underlies the development of (self-) tolerance. These autoreactive T cells escape to the periphery in the absence of AIRE, and if activated, induce autoimmune destruction of the specific tissue. The appearance of organ-specific autoantibodies precedes disease presentation and predicts the development of specific end-organ damage. The role of these antibodies is unknown, however.

CLINICAL MANIFESTATIONS

Mucocutaneous candidiasis (Chapter 338) occurs in virtually all patients and is usually the first manifestation of disease. Hypoparathyroidism and Addison's disease are the most common endocrine manifestations; each of these diseases occurs in 70 to 80% of patients. Hypoparathyroidism usually precedes Addison's disease; both diseases typically manifest before age 15 years. Premature ovarian failure (in 60% of affected women) usually presents as secondary amenorrhea; testicular failure occurs less frequently. Insulin-dependent diabetes mellitus occurs in 12% of patients, usually in adulthood; hypothyroidism is uncommon.

Nonendocrine components of this syndrome, in addition to the mucocutaneous candidiasis, include alopecia, vitiligo, corneal opacities, autoimmune hepatitis, enamel hypoplasia of teeth, tympanic membrane calcification, nail dystrophy that correlates only loosely with obvious candidiasis, parietal cell atrophy and vitamin B_{12} malabsorption, and more general intestinal malabsorption with steatorrhea. Asplenism, with Howell-Jolly bodies on peripheral blood smears (Chapter 157), has been noted in several patients. Each of the disease components should be sought when any patient presents with hypoparathyroidism, primary adrenal insufficiency, or mucocutaneous candidiasis.[8]

TREATMENT

Rx

The hypoparathyroidism is treated, like the sporadic disease, with oral calcium and 1,25-dihydroxyvitamin D, although variable intestinal malabsorption can present a particular therapeutic challenge. The candidiasis can be satisfactorily controlled with ketoconazole. Primary adrenal insufficiency is treated with glucocorticoid and mineralocorticoid replacement.

PROGNOSIS

The prognosis of the variably expressed hormonal disorders is similar to that of their sporadic counterparts. When the diagnosis of APS type 1 is made, surveys for other components of the syndrome can allow earlier treatment than would otherwise occur.

Autoimmune Polyglandular Syndrome Type 2

PATHOGENESIS

APS type 2 is usually inherited in families with characteristic (normal) variants of genes that regulate the presentation of antigens to T cells and subsequent T-cell function. The most common genetic locus associated with this syndrome is the human leukocyte antigen (HLA) locus, particularly the B8, DR3, and D4 alleles.[9] The HLA associations do not predict disease absolutely, even in identical twins, so environmental factors must contribute to disease presentation. Abnormal expression of the gene encoding cytotoxic T-lymphocyte antigen-4 (CTLA-4) can also predispose to APS type 2. The protein tyrosine phosphatase nonreceptor type 22 gene that encodes the lymphoid tyrosine phosphatase opposes signaling from the activated T-cell receptor. A variant of this gene is enriched in families with both type 1 diabetes mellitus and autoimmune thyroid disease. Genome-wide association studies have identified a large number of genes associated with type 1 diabetes. Although this approach has not yet been used for APS type 2, it seems likely that many of the genes identified in these studies will prove relevant because a large fraction of them appear to influence immune responsiveness.

CLINICAL MANIFESTATIONS

APS type 2 is considerably more common than type 1 and typically manifests in the fourth decade of life. Insulin-dependent diabetes mellitus and thyroid dysfunction, either autoimmune hypothyroidism or Graves disease, are the most frequent manifestations. Addison's disease (Chapter 227) is the third major endocrine component of this disorder. Although most patients who present with autoimmune diabetes or thyroid disease have clinical involvement of only one gland, many patients with autoimmune Addison's disease develop clinically evident disease in other endocrine glands. Less common components of the type 2 APS include primary hypogonadism and hypophysitis. Pernicious anemia, vitiligo, celiac disease, alopecia, and myasthenia gravis are also associated with this syndrome.

Organ-specific antibodies appear before clinical disease and predict subsequent disease. The role of these antibodies in organ hypofunction has not been established, however.

TREATMENT

Rx

The treatment of each component of this syndrome is identical to the treatment of each disorder in isolation, although possible clustering of diseases must be kept in mind during the evaluation and follow-up of all patients with each individual component disorder. Thyroid hormone therapy can precipitate symptoms of adrenal insufficiency in patients with both disorders. Consequently, a careful history (including family history), physical examination, and a low threshold for specific laboratory testing for adrenal insufficiency should be part of the evaluation of every patient with autoimmune hypothyroidism. Further, combinations of hypothyroidism, adrenal insufficiency, and hypogonadism can mimic hypopituitarism, although specific hormonal testing (Chapter 224) can easily distinguish these disorders. Because multiple components of the syndrome can appear asynchronously, periodic evaluation for the early appearance of additional disease components is indicated.

PROGNOSIS

The prognosis of the individual components of APS type 2 is the same as for the sporadic versions of each component.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Marx SJ. Multiplicity of hormone-secreting tumors: common themes about cause, expression, and management. *J Clin Endocrinol Metab*. 2013;98:3139-3148.
2. Salpea P, Stratakis CA. Carney complex and McCune Albright syndrome: an overview of clinical manifestations and molecular genetics. *Mol Cell Endocrinol*. 2014;386:85-91.
3. Thakker RV. Multiple endocrine neoplasia type 1 (MEN1) and type 4 (MEN4). *Mol Cell Endocrinol*. 2014;386:2-15.
4. Thakker RV, Newey PJ, Walls GV, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab*. 2012;97:2990-3011.
5. Wells SA Jr, Pacini F, Robinson BG, et al. Multiple endocrine neoplasia type 2 and familial medullary thyroid carcinoma: an update. *J Clin Endocrinol Metab*. 2013;98:3149-3164.
6. Michels AW, Gottlieb PA. Autoimmune polyglandular syndromes. *Nat Rev Endocrinol*. 2010;6:270-277.
7. Akirav EM, Ruddle NH, Herold KC. The role of AIRE in human autoimmune disease. *Nat Rev Endocrinol*. 2011;7:25-33.
8. Husebye ES, Perheentupa J, Rautemaa R, et al. Clinical manifestations and management of patients with autoimmune polyendocrine syndrome type I. *J Intern Med*. 2009;265:514-529.
9. Weinstock C, Matheis N, Barkia S, et al. Autoimmune polyglandular syndrome type 2 shows the same HLA class II pattern as type 1 diabetes. *Tissue Antigens*. 2011;77:317-324.

REVIEW QUESTIONS

1. Which of the following polyglandular neoplastic syndromes is ***not*** caused by a germline mutation?

A. McCune-Albright syndrome
B. MEN 1
C. Carney complex
D. von Hippel-Lindau disease
E. MEN 2

Answer: A The polyglandular neoplastic syndromes are all due to germline mutations in either tumor suppressor genes (*MEN1*, *MEN4*, Carney complex, *VHL*) or in an oncogene (*MEN2*). Only McCune-Albright syndrome is due to a somatic mutation, but because the mutation may occur early in embryogenesis, multiple endocrine and nonendocrine organs may be affected.

2. A 25-year-old man presents with kidney stones and is found to be hypercalcemic with an elevated PTH. All the following steps are indicated ***except:***

A. Ultrasound localization of parathyroids and immediate parathyroidectomy
B. Take a careful family history and consider screening first-degree relatives for hypercalcemia.
C. Consider testing the patient for anterior pituitary and pancreatic islet lesions.
D. Test the patient's germline DNA for *MEN1* gene mutations.

Answer: A Age 25 is quite young to present with primary hyperparathyroidism on a sporadic basis and raises suspicion of familial disease, particularly MEN 1. Steps in B through D are indicated because a positive family history for parathyroid, anterior pituitary, and/or pancreatic islet tumors would bolster the likelihood of MEN 1, as would positive results of screening for pituitary and pancreatic islet disease, and of course if *MEN1* gene mutation is identified.

3. A 35-year-old woman with type 1 diabetes and Hashimoto's thyroiditis presents for routine follow-up. She has no specific complaints, and her physical examination is normal except for a small thyroid gland. Which additional screening test should be done to evaluate adrenal function?

A. Measurement of antiadrenal antibodies
B. Measurement of morning serum cortisol
C. No screening test is necessary; adrenal insufficiency is unlikely
D. A Cortrosyn stimulation test

Answer: A Antiadrenal antibodies predict the development of adrenal insufficiency, and in the absence of clinical features are a reasonable screening test in this woman with a high probability of having autoimmune polyglandular syndrome 2. One might argue that no test is needed in the absence of clinical symptoms; however, many would advocate measurement of the antibodies, which if positive, might increase the frequency of clinical surveillance.

4. A patient with autoimmune adrenal insufficiency and type 1 diabetes presents with complaints of abdominal discomfort and occasional diarrhea. This is not precipitated by food. She does not have orthostasis, nausea, vomiting, or joint aches and otherwise feels well, although she recently lost 5 pounds. What is the next best step?

A. Consider increasing her hydrocortisone dose to test whether this represents subtle glucocorticoid deficiency.
B. Measure tissue transglutaminase (TTG) antibody to exclude celiac disease.
C. Assess gastric motility to evaluate for gastroparesis.
D. Measure liver function tests to exclude hepatitis.

Answer: B This patient may have celiac disease, which is associated with autoimmune adrenal insufficiency. These symptoms do not suggest that the other conditions are more likely.

5. A 22-year-old woman with autoimmune polyglandular syndrome (APS) 1 mentions during a routine visit that she is thinking of becoming pregnant. She wants to know whether she will pass on the disorder to her child and what genetic testing should be done. She is not related to her partner. You reassure her that it is extremely unlikely that her child will have APS 1 because of its recessive genetic transmission. What else should you tell her regarding testing for the *AIRE* gene.

A. She could get the *AIRE* gene tested to see if her child will carry a mutation.
B. Her partner could get the *AIRE* gene tested to see if her child will carry a mutation.
C. The child should be tested for *AIRE* mutations at birth.
D. No gene testing is necessary.

Answer: D Her child will be an obligate carrier of an *AIRE* haplotype mutation; this is not sufficient to cause the disease and does not need to be documented by analysis of her mutations. Given that she is not related to her partner, it is unlikely that he will carry a mutation, thus testing of the partner is not indicated. There is no reason to test the child at birth.

232

NEUROENDOCRINE TUMORS AND THE CARCINOID SYNDROME

KENNETH R. HANDE

NEUROENDOCRINE TUMORS: DESCRIPTION, INCIDENCE, AND PRESENTATION

Neuroendocrine tumors are cancers that arise from enterochromaffin cells found throughout the body. A "carcinoid tumor" implies a well-differentiated neuroendocrine tumor and excludes high-grade or poorly differentiated neuroendocrine tumors. Although several classification systems for neuroendocrine tumors currently exist, there has been a shift away from the term *carcinoid tumor* in favor of the term *well-differentiated neuroendocrine tumor*.[1]

Well-differentiated neuroendocrine tumors, or carcinoids, occur most often in the lung (30% of all carcinoid tumors), small intestine (25%), rectum (15%), appendix (10%), and stomach (5%) but may be seen in many other organs. Typical well-differentiated neuroendocrine tumors demonstrate a histologic pattern of dense nests of cells of uniform size and nuclear appearance that contain secretory granules. The neurosecretory granules contain various amines such as 5-hydroxytryptamine (serotonin), peptides, tachykinins, and prostaglandins. Low-grade neuroendocrine

tumors have a low mutation rate compared to most neoplasms. Mutations in key oncogenic pathways, such as the PI3K/AKT/mTOR pathway, have been identified.[2] The incidence of carcinoid tumors (2 to 5 per 100,000 population) has been increasing over the past 20 years.[3] Because carcinoid tumors are relatively slow growing, their prevalence is actually greater than that of esophageal, gastric, or pancreatic cancers. Carcinoid tumors may be associated with multiple endocrine neoplasia type 1 (MEN 1) syndrome (Chapter 231). Thymic carcinoid tumors present in patients with MEN 1 portend a worse prognosis.[4]

Patients with neuroendocrine tumors present either with symptoms related to tumor mass or symptoms resulting from the release of biologically active peptides into blood. Abdominal pain from a tumor mass or bowel obstruction related to the desmoplastic reaction in the surrounding mesentery is a common presenting symptom. The desmoplastic reaction is believed to develop in response to the secretion of growth factors such as platelet-derived growth factor, insulin-like growth factor, epidermal growth factor, and transforming growth factor-β. Infrequently, primary tumors cause hemoptysis or gastrointestinal bleeding. Hepatomegaly may be noted at diagnosis. Other patients may present with symptoms related to the systemic release of peptides from tumor cells, referred to as the carcinoid syndrome.

THE CARCINOID SYNDROME

The term *carcinoid syndrome* refers to the systemic signs and symptoms resulting from the release of neuroendocrine mediators by some carcinoid tumors. Cutaneous flushing, diarrhea, and cardiac valvular lesions are the most common manifestations of the carcinoid syndrome. Only 8 to 10% of all neuroendocrine tumors are associated with the carcinoid syndrome, usually ileal carcinoids with hepatic metastases. Carcinoids from different primary sites possess unique clinical characteristics. Carcinoid tumors arising from organs of the embryonic foregut (e.g., bronchus, stomach, pancreas, and thyroid) are infrequently associated with the carcinoid syndrome; carcinoids from the distal large intestine may metastasize but do not exhibit endocrine effects.

PATHOPHYSIOLOGY OF THE CARCINOID SYNDROME

The carcinoid syndrome results from the production of a variety of biologically active substances by the neuroendocrine tumor cells, including serotonin, tachykinins, histamine, and prostaglandins. Most low-grade neuroendocrine tumors contain the enzyme tryptophan hydroxylase, which catalyzes the formation of 5-hydroxytryptophan (5-HTP) from tryptophan (Fig. 232-1). The typical ileal carcinoid tumor also contains aromatic L-amino-acid decarboxylase, which catalyzes the conversion of 5-HTP to 5-hydroxytryptamine (5-HT or serotonin). Following its release, serotonin is oxidized to 5-hydroxyindoleacetaldehyde and rapidly converted to 5-hydroxyindoleacetic acid (5-HIAA) by aldehyde dehydrogenase. This acid is excreted into the urine, and almost all circulating serotonin can be accounted for as urinary 5-HIAA. Tachykinins are also stored in neuroendocrine tumors. Of these, neuropeptide K, neurokinins A and B, and substance P have been identified in tumors and blood from patients with the carcinoid syndrome. Some carcinoid tumors, particularly those of gastric origin, release excessive amounts of histamine. Secretion of a variety of prostaglandins by carcinoids has also been demonstrated. Neuroendocrine tumors, particularly of the thymus and lung, have been associated with ectopic production of adrenocorticotropic hormone and growth hormone–releasing hormone.

Serotonin contributes to the intestinal hypermotility and diarrhea associated with the carcinoid syndrome. A secondary effect of serotonin overproduction occurs when a large fraction of dietary tryptophan is shunted into the hydroxylation pathway, leaving less tryptophan available for the formation of nicotinic acid and protein. When urinary excretion of 5-HIAA exceeds 100 mg/day, low levels of plasma tryptophan and evidence of nicotinic acid deficiency (pellagra) can be seen. The interaction of serotonin with platelets and the cardiac endothelium is considered the cause of carcinoid heart disease. This hypothesis is supported by the finding of valvular heart disease in patients who took appetite suppressants, such as fenfluramine, that release serotonin. The risk of valvular heart disease in patients with the carcinoid syndrome is correlated with the amount of 5-HIAA excreted in the urine.

Most evidence points to the tachykinins as mediators of the carcinoid flush. Tachykinins are known vasodilators. Tachykinin levels are increased during pentagastrin-induced flushing; when pentagastrin-induced flushing is inhibited by somatostatin, the rise in tachykinin levels is also blocked. Serotonin does not appear to be the mediator of flushing. Flushing attacks can be attributed to histamine in certain gastric carcinoids. Flushing can be triggered by catecholamines, and this probably accounts for the association of flushing with exercise and emotional stimuli. Injection of isoproterenol or pentagastrin can also trigger flushing, an action that may explain the provocation of flushes by eating in some patients. The carcinoid syndrome occurs when mediators produced by the tumor and normally metabolized by the liver escape into the systemic circulation. Thus, most patients with the carcinoid syndrome have hepatic metastasis.

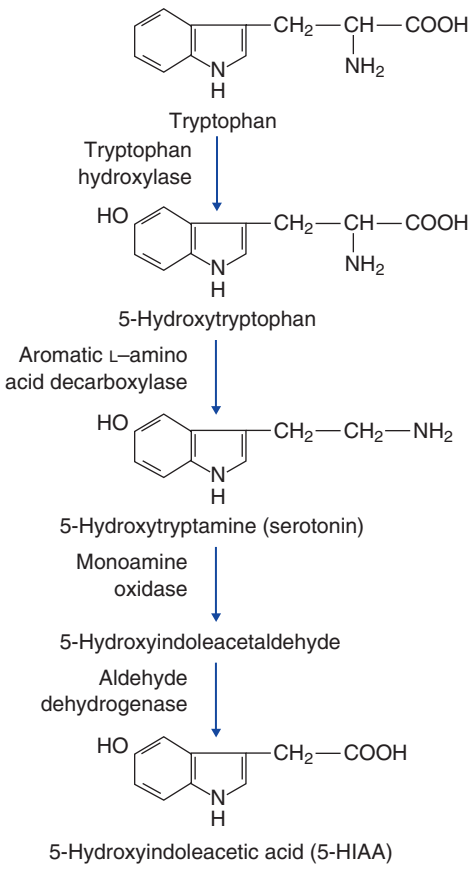

FIGURE 232-1. Synthesis and degradation of serotonin.

CLINICAL MANIFESTATIONS OF THE CARCINOID SYNDROME

VASODILATOR PAROXYSMS. Cutaneous flushing, which occurs in 80% of patients with the carcinoid syndrome, is the most common clinical feature. The typical flush is dark red to violaceous and involves the head, neck, and upper trunk (blush area). The flush usually lasts for 30 seconds to 3 minutes. Neuroendocrine tumors of the foregut produce a slightly different flush, characteristically bright salmon pink to red. Prolonged flushing attacks may be associated with lacrimation and periorbital edema. The flush may be accompanied by tachycardia. The blood pressure usually falls or does not change. A rise in blood pressure during flushing is rare, and the carcinoid syndrome is not a cause of sustained hypertension. Flushing may be provoked by excitement, exertion, eating, and ethanol ingestion. In patients with the bronchial carcinoid variant, flushing may last for hours. In addition to paroxysms of cutaneous vasodilatation, some patients develop telangiectasias, which are most marked in the malar area. These patients may have the characteristic features of rosacea.

GASTROINTESTINAL SYMPTOMS. Intestinal hypermotility with borborygmi and cramping occurs in 50 to 70% of patients with the carcinoid syndrome. Explosive secretory diarrhea may occur, although chronic diarrhea with a secretory component is more common. When diarrhea is severe, malabsorption may occur. Gastrointestinal transit times through the small and large bowel are two- to six-fold faster than in physiologically normal patients.

PELLAGRA. Tryptophan is normally used to form nicotinic acid. In patients with the carcinoid syndrome, up to 60% of dietary tryptophan may be used to form 5-HTP and 5-HT. Nicotinic acid levels are occasionally depleted, resulting in symptoms of pellagra (dermatitis, diarrhea, and dementia) (Chapter 218).

CARDIAC MANIFESTATIONS. Symptomatic valvular heart disease is present in 15 to 20% of patients with the carcinoid syndrome (Chapter 60). Up to 50% of patients have echocardiographic evidence of heart disease.[5] Plaque-like thickening of the endocardium of the valvular cusps and cardiac chambers occurs primarily on the right side of the heart but may rarely involve the left side (<10%). Lesions of the tricuspid valve (usually regurgitation) are present in 65% of patients with carcinoid heart disease, and pulmonic valvular disease (again, usually regurgitation) is seen in 20%.

OTHER SYMPTOMS. Generalized fatigue and debilitation are underappreciated features of the carcinoid syndrome. Bronchoconstriction, usually most pronounced during flushing attacks, is a less common feature of the syndrome, but when it occurs it may be severe. Attacks of severe and sustained flushing with life-threatening hemodynamic compromise and bronchoconstriction are referred to as carcinoid crisis. Precipitating factors include anesthesia, surgery, tumor necrosis, and catecholamine infusion. Cognitive impairment has also been associated with the carcinoid syndrome.[6]

DIAGNOSIS

When all its clinical features are present, the carcinoid syndrome is easily recognized. The diagnosis also must be considered when any one of its clinical manifestations is present. The diagnostic hallmark consists of overproduction of 5-hydroxyindoles accompanied by increased excretion of urinary 5-HIAA in a patient with a biopsy-proven carcinoid tumor. Normally, excretion of 5-HIAA does not exceed 10 mg/day. Ingestion of foods containing serotonin may complicate the biochemical diagnosis of the carcinoid syndrome; bananas, walnuts, and certain other foods contain enough serotonin to produce abnormally elevated urinary excretion of 5-HIAA after their ingestion. Selected drugs (e.g., guaifenesin, acetaminophen) may also falsely elevate urinary 5-HIAA measurements. When dietary 5-hydroxyindoles are excluded, urinary excretion of 25 mg/day of 5-HIAA is diagnostic of the carcinoid syndrome. Elevation in the range of 9 to 25 mg/day may be seen with the carcinoid syndrome, nontropical sprue, vomiting, or acute intestinal obstruction. Measurement of serotonin in blood or platelets is of interest but has less diagnostic value than assay of the major metabolite of serotonin in the urine. Plasma chromogranin A concentrations are often elevated in carcinoid patients, including those who do not have the carcinoid syndrome, and may serve as a marker of tumor mass. The diagnostic value of plasma chromogranin A is relatively low, however, because this substance is increased in patients with renal failure, atrophic gastritis, and patients taking proton pump inhibitors. Assessment of the extent and localization of both primary and metastatic tumor is aided by computed tomography of the abdomen and chest and by imaging with radionuclide-labeled somatostatin receptor ligands.

The typical carcinoid syndrome usually results from tumors of midgut origin, which almost invariably secrete serotonin. In contrast, tumors arising from the embryonic foregut have a lower serotonin content and may secrete 5-HTP. Patients with gastric carcinoids may exhibit unique flushing, beginning as bright, patchy erythema with sharply delineated serpentine borders that coalesce as the blush heightens. Food ingestion is especially likely to produce flushes. With carcinoid tumors arising from the bronchus, attacks of flushing tend to be prolonged and severe and may be associated with periorbital edema, excessive lacrimation and salivation, hypotension, tachycardia, anxiety, and tremulousness. This group is therapeutically unique in that severe flushes can sometimes be prevented by corticosteroids.

Differential Diagnosis of the Carcinoid Syndrome

Attacks of flushing in a patient with normal urinary excretion of 5-HIAA raise other diagnostic possibilities. Systemic mastocyte activation disorders, including systemic mastocytosis (Chapter 255) and idiopathic anaphylaxis (Chapter 253), produce flushing and diarrhea and should be considered when 5-HIAA excretion is not elevated. Flushing also occurs in genetically predisposed individuals following ethanol ingestion, in the postmenopausal state (Chapter 240), and in conjunction with other neuroendocrine tumors such as VIPomas and medullary carcinomas of the thyroid.

TREATMENT

Rx

Treatment of the carcinoid syndrome is directed toward pharmacologic therapy for humorally mediated symptoms and at measures designed to reduce the tumor mass.

Somatostatin Analogues

More than 80% of neuroendocrine tumors have somatostatin receptors on the cell surface. Somatostatin can bind to these receptors and prevent flushing and other endocrine symptoms. The development of analogues of somatostatin with longer biologic half-lives than the native hormone has made subcutaneous and intramuscular administration feasible and has been a major advance in the treatment of these patients.

Roughly 70% of patients with carcinoid syndrome have a 50% or greater reduction in the frequency of diarrhea and/or flushing with the use of octreotide, one of the somatostatin analogs. Therapy is usually associated with a decrease in urinary 5-HIAA excretion and in plasma tachykinin levels. With the improvement of endocrine symptoms and fatigue, a considerable improvement in quality of life may be achieved. Symptom control with octreotide can be durable, with 50 to 60% of patients continuing to have improvement in symptoms of diarrhea and flushing following 12 months of therapy.[7] Long-acting somatostatin analogues (octreotide LAR and lanreotide) have been developed, permitting once-monthly dosing. Two to 4 weeks may be required to achieve steady-state levels of octreotide following administration of octreotide LAR, during which time supplementation with subcutaneous octreotide may be needed. Octreotide is generally well tolerated. However, it may suppress pancreatic exocrine function, causing steatorrhea, abnormal glucose control, and inhibition of the release of cholecystokinin. Hyperglycemia, symptomatic cholelithiasis, steatorrhea, and hypoglycemia are seen in 9, 15, 3, and 2% of patients, respectively. Octreotide should be used to prevent carcinoid crises that accompany the massive release of mediators that may occur during operative procedures. In patients with histamine-secreting gastric neuroendocrine tumors, blockade of both histamine $(H)_1$ and H_2 receptors ameliorates flushing. In patients receiving octreotide, regression in tumor mass is uncommon. However, octreotide slows the growth rate of low-grade neuroendocrine tumors. With octreotide therapy, disease stability is seen in 65% of patients at 6-month follow-up, compared to only 35% of patients not receiving octreotide.

Surgery

Early diagnosis of the carcinoid syndrome leads to complete surgical cure in the few neuroendocrine tumors that release their humoral mediators directly into the systemic circulation (e.g., bronchial carcinoids). In contrast, tumors that release humoral substances into the portal circulation usually produce the syndrome only after hepatic metastases occur. Given the slow growth rate of low-grade neuroendocrine tumors, effective reduction in tumor mass can ameliorate morbidity and improve quality of life even after metastases have occurred. In selected patients, this goal can be achieved by surgical debulking of tumor, including partial hepatectomy for unilobar metastases, excision of large superficial hepatic metastases, and removal of the primary tumor and regional lymph nodes. In selected series,[8] over 90% of patients report symptom relief from pain and/or carcinoid syndrome following cytoreductive surgery (50 to 70%, complete relief). Median time to recurrence of symptoms is 3 years, and overall survival is greater than 5 years. Elective cholecystectomy during the surgical intervention prevents cholelithiasis that may result from octreotide treatment. Because the blood supply of hepatic metastases is largely arterial, percutaneous embolization of the hepatic arterial supply to the most involved hepatic lobe can shrink inoperable hepatic metastases; however, the procedure carries a risk of serious complications, with mortality of up to 6%.[9] In selected series, 90% of patients report improvement in symptoms for an average of 2 years' duration. Median survival in embolized patients has been 24 to 30 months. Valve replacement surgery is indicated in severe disease, in which it can usually improve life expectancy (Chapter 60).[10]

Chemotherapy

In contrast to high-grade, poorly differentiated neuroendocrine cancers, chemotherapy with single or combination cytotoxic agents rarely produces tumor regression in low-grade neuroendocrine tumors. However, recent studies suggest that several drugs, including the antiangiogenic agents bevacizumab and sunitinib and the mTOR inhibitor everolimus, may slow the growth rate of low-grade neuroendocrine tumors. In a randomized double-blind placebo-controlled study, 10 mg/day of everolimus increased progression-free survival from 11 months in control patients to 16 months (hazard ratio, 0.77). No overall survival advantage was noted, however, and stomatitis, rash, fatigue, and diarrhea are noted treatment side effects. For patients who exhibit tumor progression or whose clinical syndrome has failed to improve following cytoreduction and octreotide, everolimus may be considered for palliative therapy.

PROGNOSIS

The metastatic potential of localized carcinoid tumors correlates with tumor size, location, and histologic grade. Even when metastatic, typical neuroendocrine tumors generally have a slow rate of growth, and many patients survive for years after metastatic disease is recognized. Prior to development

of therapies for the carcinoid syndrome, morbidity resulted from the endocrine manifestations of the tumor. However, with current therapy, symptoms of the carcinoid syndrome can usually be controlled. Death from typical neuroendocrine tumors is usually caused by complications associated with tumor growth, such as bowel obstruction or hepatic failure. A concerted strategy consisting of removal of the primary tumor, reduction in tumor bulk, and administration of octreotide can lead to considerable amelioration of symptoms, improvement in the quality of life, reduction in the release of the humoral substances that engender cardiac lesions, and prolongation of survival.[11] The median survival of patients with metastatic neuroendocrine tumors and the carcinoid syndrome now exceeds 5 years in carefully managed patients, with patients having small bowel primaries living longer than those with lung or colon primaries.

Grade A References

A1. Rinke A, Muller HH, Schade-Brittinger C, et al. Placebo-controlled, double-blind, prospective, randomized study of the effect of octreotide LAR in the control of tumor growth in patients with metastatic neuroendocrine midgut tumors: a report from the PROMID study group. *J Clin Oncol*. 2009;27:4656-4663.

A2. Pavel ME, Hainsworth JD, Baudin E, et al. Everolimus plus octreotide long-acting repeatable for the treatment of advanced neuroendocrine tumors associated with carcinoid syndrome (RADIANT-2): a randomized, placebo-controlled trial. *Lancet*. 2011;378:2005-2012.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bergsland EK. The evolving landscape of neuroendocrine tumors. *Semin Oncol*. 2013;40:4-22.
2. Banck MS, Kanwar R, Kulkarni AA, et al. The genomic landscape of small intestine neuroendocrine tumors. *J Clin Invest*. 2013;123:2502-2508.
3. Fraenkel M, Kim MK, Faggiano A, et al. Incidence of gastroenteropancreatic neuroendocrine tumors: a systematic review of the literature. *Endocr Relat Cancer*. 2014;21:R153-R163.
4. Ito T, Igarashi H, Uehara H, et al. Causes of death and prognostic factors in multiple endocrine neoplasia type 1: a prospective study: comparison of 106 MEN1/Zollinger-Ellison syndrome patients with 1613 literature MEN1 patients with or without pancreatic endocrine tumors. *Medicine*. 2013;92:135-181.
5. Patel C, Mathur M, Escarcega RO, et al. Carcinoid heart disease: current understanding and future directions. *Am Heart J*. 2014;167:789-795.
6. Pasieka JL, Longman RS, Chambers AJ, et al. Cognitive impairment associated with carcinoid syndrome. *Ann Surg*. 2014;259:355-359.
7. Anthony L, Vinik AI. Evaluating the characteristics and the management of patients with neuroendocrine tumors receiving octreotide during a 6-year period. *Pancreas*. 2011;40:987-994.
8. Gu PI, Wu J, Newman E, et al. Treatment of liver metastases in patients with neuroendocrine tumors of gastroesophageal and pancreatic origin. *Int J Hepatol*. 2012;2012:131659.
9. Harring TR, Nguyen TN, Goss JA, et al. Treatment of liver metastasis in patients with neuroendocrine tumors: a comprehensive review. *Int J Hepatol*. 2011;2011:1-11.
10. Dobson R, Burgess MI, Pritchard DM, et al. The clinical presentation and management of carcinoid heart disease. *Int J Cardiol*. 2014;173:29-32.
11. Kunz PL, Reidy-Langunes D, Anthony LB, et al. Consensus guidelines for the management and treatment of neuroendocrine tumors. *Pancreas*. 2013;42:557-577.

REVIEW QUESTIONS

1. A 56-year-old woman presents with diarrhea, episodes of facial flushing, and fatigue. An enlarged liver is noted on exam, and a computed tomography scan shows metastatic nodules within the liver. A biopsy of one of these nodules shows a well-differentiated neuroendocrine tumor. The most likely primary site for this cancer is:

A. Thymus
B. Lung
C. Stomach
D. Small intestine
E. Rectum

Answer: D Small bowel carcinoid with hepatic metastases is the neuroendocrine tumor most commonly associated with carcinoid syndrome (General Reference 5).

2. Evaluation to identify a primary tumor site and corroborate the diagnosis in the patient in question 1 should include:

A. FDG-PET scan
B. Echocardiogram
C. Upper endoscopy
D. Bone scan
E. Radionucleotide-labeled somatostatin receptor ligand imaging

Answer: E Most (80%) low-grade neuroendocrine tumors will have somatostatin receptors (Grade A Reference A2).

3. Initial therapy for the patient in question 1 should be:

A. Octreotide
B. Chemoembolization of hepatic metastasis
C. Everolimus
D. Surgical tumor debulking
E. Benadryl

Answer: A Octreotide reduces symptoms of flushing and diarrhea in the majority of patients and should be used before any surgical intervention to prevent carcinoid crisis (General Reference 6).

4. A 24-hour urine for 5-HIAA is collected from the patient in question 1 and found to be elevated at 180 mg excretion per 24 hours (normal, <10 mg/day). The increased secretion of serotonin is associated with all of the following ***except***:

A. Flushing episodes
B. Diarrhea
C. Hepatic metastasis
D. Risk of carcinoid heart disease
E. Development of pellagra

Answer: A Histamine and catecholamines (the latter as the trigger) but not serotonin are the mediators of carcinoid flushing in many cases. Serotonin is most likely responsible for diarrhea, valvular heart disease, and niacin deficiency. In patients with carcinoid syndrome, up to 60% of dietary tryptophan may be used to form 5-HTP and 5-HT. Because tryptophan is used to form nicotinic acid, depletion of the latter in such patients can result in symptomatic pellagra.

5. The prognosis for the patient listed above is most likely:

A. Cure following surgical resection
B. 4- to 6-month survival
C. 1- to 2-year survival
D. 4- to 6-year survival
E. >10-year survival

Answer: D Metastatic neuroendocrine tumors are rarely curable, but because of their slow rate of growth, appropriately managed patients may live many years with good quality of life (General References 1 and 10).

233

DISORDERS OF SEXUAL DEVELOPMENT

PERRIN C. WHITE

DEFINITION

This chapter reviews the concepts underlying the initial evaluation and management of patients with disorders of sexual development (DSD). By definition, such an individual has lack of concordance of various aspects of gender. These include chromosomal sex (46,XX, 46,XY, or other), gonadal or reproductive sex (ovaries, fallopian tubes, and uterus vs. testes, seminal vesicles, prostate gland, and ejaculatory ducts), genital sex (vagina and clitoris vs. scrotum and penis), and gender-specific behavior. Depending on chromosomal sex, most patients can be classified as incompletely masculinized 46,XY males, virilized 46,XX females, and patients with abnormalities of sex chromosomes, such as those with mixed gonadal dysgenesis. (In the past, 46,XY and 46,XX DSD patients were referred to as male and female pseudohermaphrodites, respectively, but these terms are no longer preferred.) Many conditions can cause DSD (Table 233-1).

TABLE 233-1 DISORDERS OF SEXUAL DEVELOPMENT*

VIRILIZATION OR SEX REVERSAL IN XX FEMALES
Virilizing forms of congenital adrenal hyperplasia
21-Hydroxylase deficiency (1 : 16,000 births) [*CYP21*]: salt-wasting or simple virilizing forms
11β-Hydroxylase deficiency [*CYP11B1*]
3β-Hydroxysteroid dehydrogenase deficiency [*HSD3B2*]
Cytochrome P-450 oxidoreductase deficiency (also has a maternal effect) [*POR*]
Maternal or exogenous androgens
Drugs (danazol, progestins)
Luteoma
Aromatase deficiency [*CYP19*]
Transcription factor mutations
Mutations in genes affecting gonadal differentiation
SRY (translocation to X)
SOX9 (duplication)
WT1 (Denys-Drash syndrome)
Structural/idiopathic
UNDERVIRILIZATION OR SEX REVERSAL IN XY MALES
Biosynthetic defects
Lipoid adrenal hyperplasia [*STAR*]
17α-Hydroxylase/17,20 lyase [*CYP17*]
3β-Hydroxysteroid dehydrogenase [*HSD3B2*]
17-Ketosteroid reductase [*HSD17B3*]
5α-Reductase [*SRD5A2*]
Cytochrome P-450 oxidoreductase deficiency [*POR*]
Smith-Lemli-Opitz syndrome (1 : 20,000) [*DHCR7*]
Androgen insensitivity (1 : 20,000) [*AR*]: complete or partial
Luteinizing hormone insensitivity [*LHR*]
Mutations in genes affecting gonadal differentiation
SRY
SOX9 (campomelic dysplasia)
SF1 (sometimes associated with adrenal hypoplasia)
WT1 (WAGR and Denys-Drash syndromes)
DAX1 or WNT4 duplications
DHH (associated with peripheral neuropathy)
ATRX (X-linked α-thalassemia and mental retardation)
Exposure to 5α-reductase inhibitors, other endocrine disruptors
MICROPENIS
Panhypopituitarism [including *PROP1* mutations]
Septo-optic dysplasia [including *HESX* mutations]
Hypogonadotropic hypogonadism
Kallmann syndrome [*KAL1*]
Prader-Willi syndrome [paternal chromosome 15q11 deletion]
Adrenal hypoplasia congenita [*DAX1*]
Vanishing testes (may also cause ambiguous genitalia)
OTHER SYNDROMES AFFECTING REPRODUCTIVE SYSTEMS
Chromosomal aneuploidy
Turner syndrome (1 : 2500): 45,X; 45,X/46,XX mosaics; 46,XXr; 46,XXq−
Klinefelter's syndrome (1 : 1000): 47,XXY
Mixed gonadal dysgenesis (1 : 20,000): 45,X/46,XY; 45,X/47,XXY
Other: Trisomy 13, trisomy 18, triploidy, 4p−, 13q−
Persistent müllerian duct syndrome in XY males
Type 1 [*AMH*]
Type 2 [*AMHR2*]
Mayer-Rokitansky-Küster-Hauser syndrome (vaginal atresia) (1 : 6000)

*Frequencies of relatively common (at least 1:20,000) diseases are noted in parentheses. When causative genetic mutations have been identified, the affected locus is noted in square brackets. WAGR = Wilms' tumor, aniridia, genitourinary abnormalities or gonadoblastoma, and mental retardation.

Normal Sexual Differentiation

Gonadal Differentiation

At 5 to 6 weeks' gestation, the gonadal primordia (gonadal ridges) develop from the coelomic epithelium overlying the medial surface of the mesonephros (primitive kidneys; Fig. 233-1). These primitive gonads are identical in both sexes. Germ cells form at 3 to 4 weeks' gestation and migrate through the gut mesentery into the gonads at this early bipotential stage. Whether germ cells are directed toward male or female gametogenesis depends largely on the environment generated by surrounding somatic cells rather than on factors intrinsic to the germ cells.

During the seventh week, XY male gonads begin to differentiate under the influence of testis-determining genes.[1,2] The first to be expressed is *SRY*, the key gene on the Y chromosome controlling male differentiation, which initiates the development of Sertoli cells. Sertoli cells surround germ cells to form testis cords, which nourish primordial germ cells and direct them into the pathway for male gametogenesis. Recruitment of endothelial cells leads to development of a testis-specific vasculature that is required for normal organization of the testis.

Steroidogenic cells develop from the mesonephros and migrate into the developing adrenal cortex and testis at 8 weeks. In the testis, they become Leydig cells, which secrete the testosterone required for subsequent male reproductive development. In the first trimester, testosterone secretion is mainly under the control of human chorionic gonadotropin (HCG); it subsequently requires luteinizing hormone (LH) secreted by the fetal anterior pituitary.

Ovaries are recognizable at approximately 10 weeks. The signaling molecule WNT4 plays an active role in ovarian development, repressing expression of testis-specific genes and vascular development.[3] Germ cells in the ovary continue into the first meiotic prophase beginning at 12 weeks' gestation and continuing until 7 months' gestation.

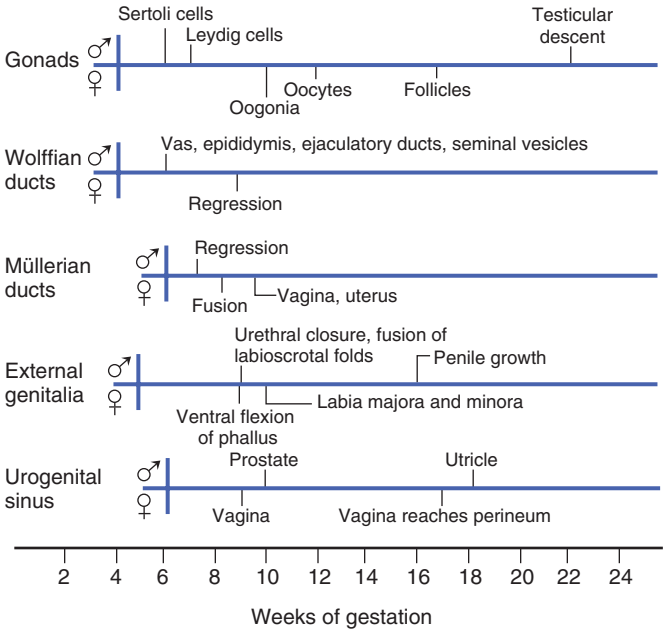

FIGURE 233-1. Time course of prenatal sexual differentiation in male and female fetuses. (Modified from Barthold JS, Gonzalez R. Intersex states. In: Gonzales ET, Bauer SB, eds. *Pediatric Urology Practice*. Philadelphia: Lippincott Williams & Wilkins; 1999.)

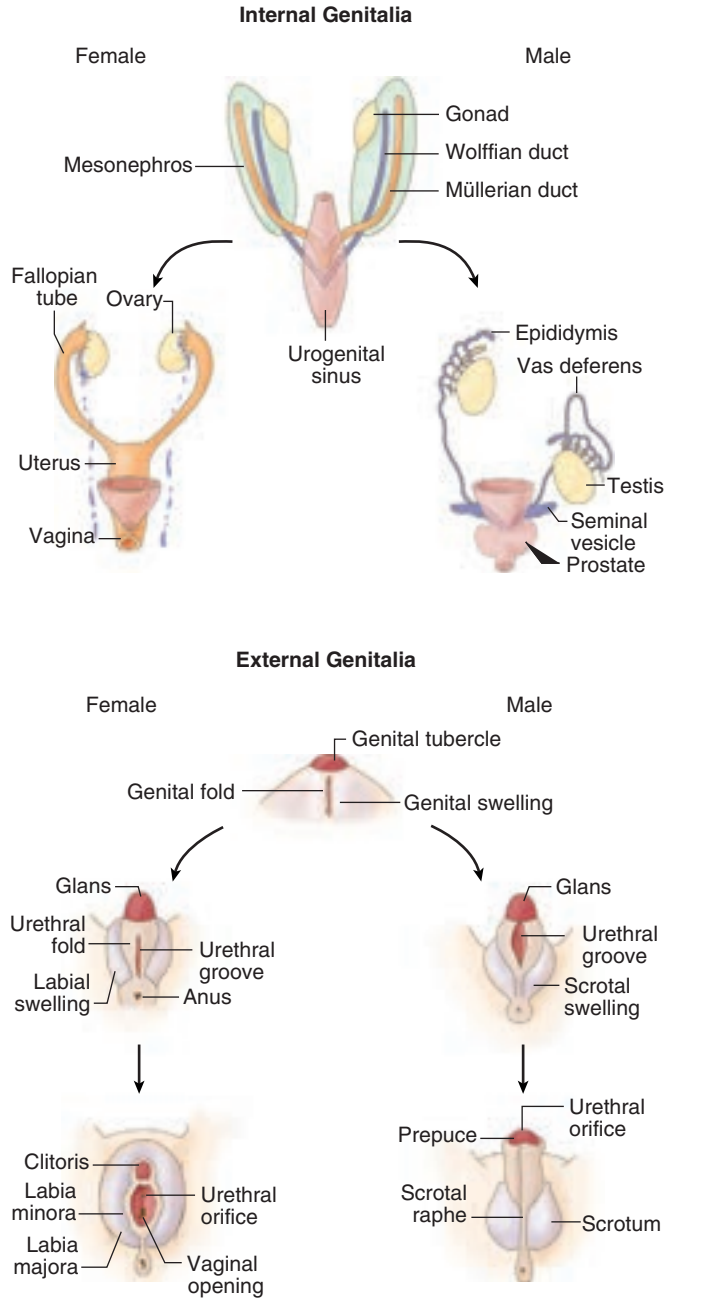

FIGURE 233-2. Differentiation of the internal and external genitalia of the human fetus. (Modified from Griffen JE, Ojeda SR, eds. *Textbook of Endocrine Physiology*. New York: Oxford University Press; 1996.)

Development of Male and Female Internal Reproductive Tracts

The reproductive tracts are derived from intermediate mesoderm. The male reproductive tract develops from the mesonephric (wolffian) ducts, and the female reproductive tract develops from the paramesonephric (müllerian) ducts (Fig. 233-2). Both sets of ducts are present in normal embryos.

Development of wolffian or müllerian structures depends on the presence or absence of normally functioning testes. The Sertoli cells secrete antimüllerian hormone (AMH, also termed müllerian inhibiting substance) starting when the testes differentiate.[4] Expression of AMH is controlled by several transcription factors: SF1 and WT1 (Wilms' tumor locus) synergize to promote transcription, whereas DAX1 antagonizes it. The overall effect of AMH is to induce regression of müllerian structures between 8 and 12 weeks' gestation. In its absence, development of the müllerian ducts proceeds, and the female internal structures (fallopian tubes, uterus, cervix, and upper vagina) are formed.

Development of the structures derived from the wolffian ducts, including the epididymis, ductus deferens, ejaculatory ducts, and seminiferous tubules, requires high local concentrations of testosterone secreted from Leydig cells of the testis beginning at approximately 7 weeks' gestation. In the absence of testosterone, wolffian ducts regress. Levels of testosterone in the circulation are insufficient to develop wolffian structures. Thus, in conditions in which the gonads develop asymmetrically (e.g., true hermaphroditism or mixed gonadal dysgenesis; see later), wolffian structures develop asymmetrically as well. Development of wolffian structures requires an intact androgen receptor.

Development of the External Genitalia

External genital structures are also bipotential in early gestation and consist of the genital tubercle, genital folds (later, urethral-labial folds), and genital swelling (later, labioscrotal folds) (see Fig. 233-2). Differentiation to male genitalia occurs from approximately 8 to 14 weeks' gestation under the influence of dihydrotestosterone, which must interact with an intact androgen receptor. The genital tubercle becomes the glans penis; the genital folds fuse to become the shaft of the penis and penile urethra, and the labioscrotal folds (derived from the genital swelling) fuse to become the scrotum. Without androgens, these structures become the clitoris, labia minora, and labia majora, respectively.

Normal Gonadal and Adrenal Steroidogenesis

Many forms of genital ambiguity result from defects in steroid biosynthesis in the testes or adrenal cortex or from defective steroid metabolism in the placenta or in target tissues[5] (Fig. 233-3 and E-Fig. 233-1).

Steroid biosynthesis in the testes and adrenals begins with the importation of cholesterol into mitochondria, a highly regulated process controlled largely by the steroidogenic acute regulatory (StAR) protein; levels of StAR are controlled within the adrenals by adrenocorticotropic hormone (ACTH) and within the testis by HCG during the first trimester and by LH later in pregnancy.

Within mitochondria, the side chain of cholesterol is cleaved between carbons 20 and 22 by the cholesterol side-chain cleavage enzyme (cholesterol desmolase, CYP11A), a cytochrome P-450 enzyme. The product is pregnenolone, which is transported to the endoplasmic reticulum. Some pregnenolone is converted by 17α-hydroxylase (CYP17) to 17-hydroxypregnenolone. Both 17-hydroxypregnenolone and the remaining pregnenolone are converted by 3β-hydroxysteroid dehydrogenase (HSD3B2) to 17-hydroxyprogesterone and progesterone, respectively. The side chain of 17-hydroxypregnenolone is cleaved by the 17,20-lyase activity of CYP17 to dehydroepiandrosterone (DHEA). DHEA may also be converted to androstenedione by HSD3B2.

All the preceding steps can occur in the adrenal cortex, in Leydig cells of the testis, and (after puberty) in theca cells of ovarian follicles. Subsequent

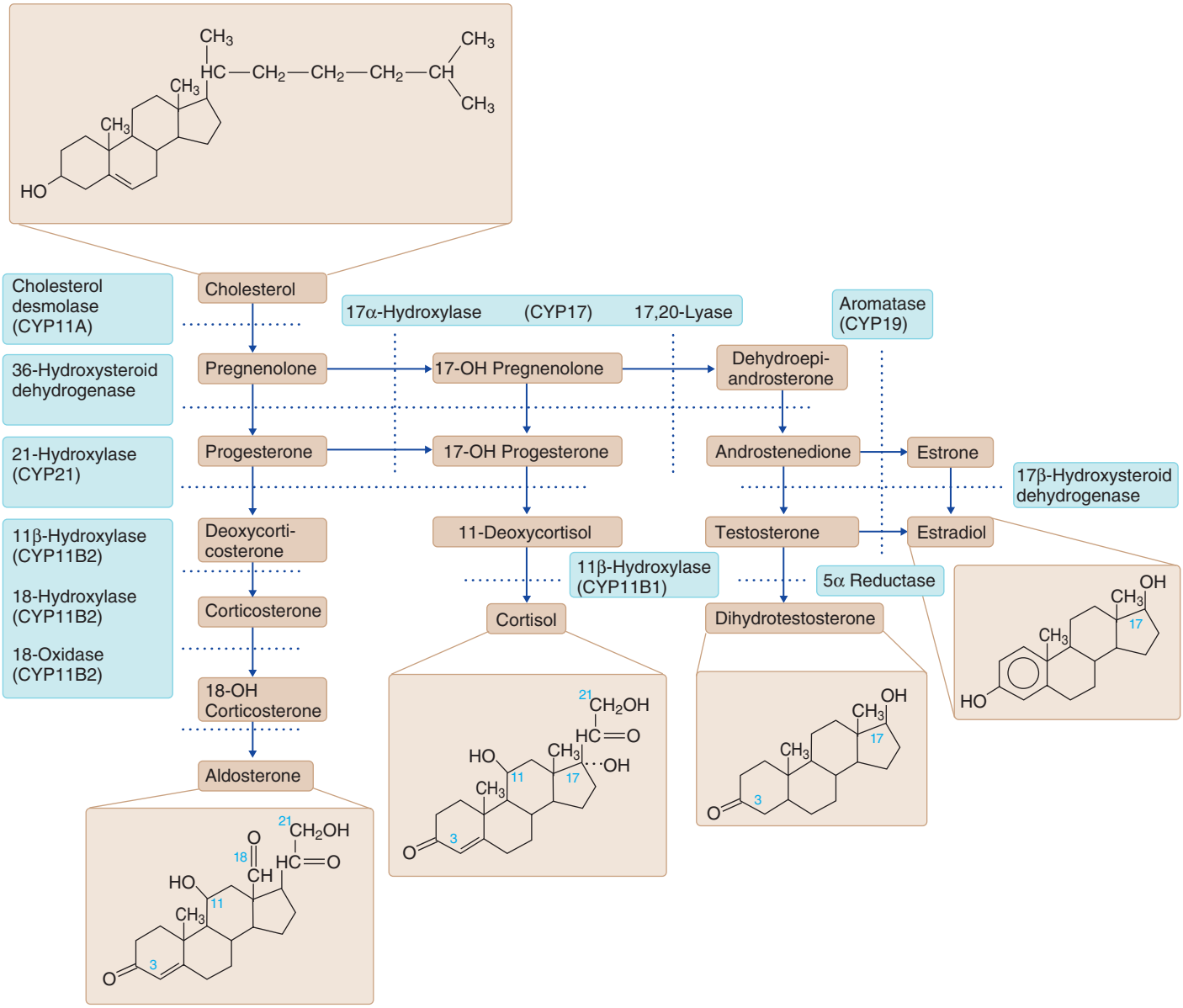

E-FIGURE 233-1. Steroidogenesis. The pathways for synthesis of progesterone and mineralocorticoids (aldosterone), glucocorticoids (cortisol), androgens (testosterone and dihydrotestosterone), and estrogens (estradiol) are arranged from left to right. Planar structures are shown for cholesterol and the end products of each pathway. The enzymatic activities catalyzing each bioconversion are written in boxes. For those activities mediated by specific cytochrome P-450 subsets, the systematic name of the enzyme (CYP followed by a number) is listed in parentheses. CYP11B2 and CYP17 have multiple activities.

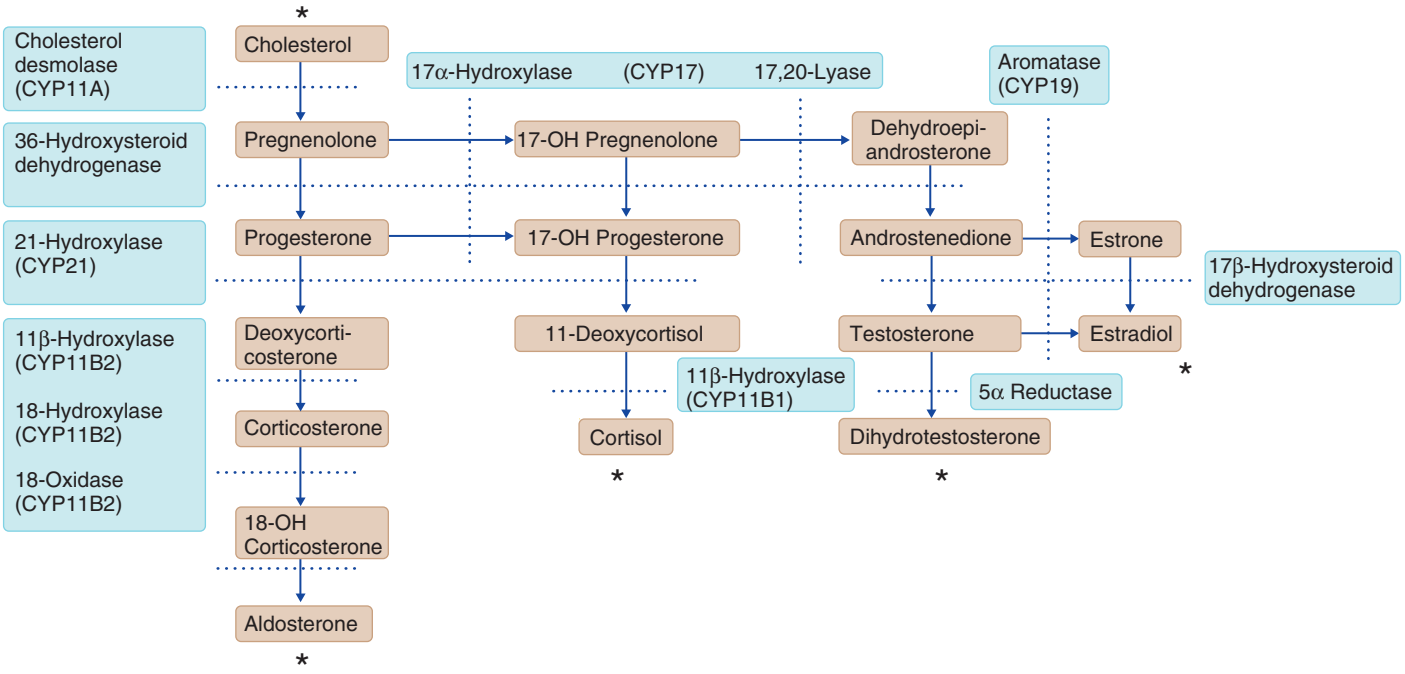

FIGURE 233-3. Steroidogenesis. Pathways for the synthesis of progesterone and mineralocorticoids (aldosterone), glucocorticoids (cortisol), androgens (testosterone and dihydrotestosterone), and estrogens (estradiol) are arranged from left to right. The enzymatic activities catalyzing each bioconversion are written in boxes. For those activities mediated by specific cytochrome P-450 subsets, the systematic name of the enzyme (CYP followed by a number) is listed in parentheses. CYP11B2 and CYP17 have multiple activities. *For an expanded version of this image, including planar structures for cholesterol and the end products of each pathway, see E-Figure 233-1.

biosynthetic steps are specific to different glands. In the adrenal cortex, 17-hydroxyprogesterone is converted by 21-hydroxylase (CYP21, a microsomal P-450) to 11-deoxycortisol, which is then converted in mitochondria to cortisol by 11β-hydroxylase (CYP11B1). Additionally, progesterone is converted to deoxycorticosterone by CYP21, which is then converted to aldosterone by aldosterone synthase (CYP11B2).

In Leydig cells of the testis, androstenedione is converted to testosterone by 17-ketosteroid reductase (17β-hydroxysteroid dehydrogenase type 3 [HSD17B3]); the same reaction occurs in theca cells of the ovary, catalyzed by 17β-hydroxysteroid dehydrogenase type 1 (HSD17B1). In granulosa cells of the ovary (after puberty), androstenedione and testosterone are converted by aromatase (CYP19) to estrone and estradiol, respectively. In skin of the developing male external genitalia, steroid 5α-reductase (SRD5A2) converts testosterone to a more potent androgen, dihydrotestosterone.

The placenta is a steroid-synthesizing and steroid-metabolizing tissue as well; it has high steroid sulfatase activity that converts DHEA sulfate from the fetal adrenal gland back to DHEA. This is then successively converted by 3β-hydroxysteroid dehydrogenase type 1 (HSD3B1) and aromatase (CYP19) to androstenedione and estrone, respectively, which are then converted to estradiol by HSD17B1.

DEFECTS OF SEX DIFFERENTIATION

Defects of Steroidogenesis

PATHOBIOLOGY

Genital ambiguity in genetic females is usually the result of exposure to excessive levels of androgens. Virilizing congenital adrenal hyperplasia (CAH), the most common cause of genital ambiguity in female infants, occurs in 1 in 16,000 births.[6]

Conversely, severe deficiencies of androgens, if present early in gestation, cause ambiguous or female-appearing external genitalia in male infants. Usually, müllerian structures such as the uterus, cervix, and upper vagina are not present because the testes are able to secrete müllerian inhibitory substance. Thus, individuals with these conditions have a short vagina ending in a blind pouch.

CONGENITAL ADRENAL HYPERPLASIA

The fundamental defect among patients with any form of CAH is inadequate synthesis of cortisol (see Fig. 233-3 and E-Fig. 233-1). Inefficient cortisol synthesis signals the hypothalamus and pituitary to increase corticotropin-releasing hormone and ACTH, respectively (Chapter 223). Consequently, the adrenal glands become hyperplastic, and steroid precursors accumulate proximal to the block in biosynthesis. In some conditions, these precursors can be converted to androgens.

Lipoid Hyperplasia

Lipoid hyperplasia results from mutations in the *STAR* gene. Cholesterol is not imported efficiently into mitochondria and thus accumulates in cells. Steroid biosynthesis is drastically reduced because of the lack of substrate, and the lipid accumulation quickly kills steroid-synthesizing cells in both the adrenals and the testes. Thus, affected male patients are born as phenotypically female because they cannot synthesize testosterone. Affected female patients may undergo transient spontaneous puberty because human ovarian granulosa cells do not synthesize steroid hormones (and thus do not accumulate cholesterol) until puberty. Both sexes have adrenal insufficiency and are unable to synthesize either cortisol or aldosterone.

17α-Hydroxylase/17,20 Lyase Deficiency

Severe mutations in the *CYP17* gene prevent the synthesis of any sex hormones. Affected male patients have female-appearing external genitalia but have no müllerian structures because the testes synthesize AMH. Affected female patients remain sexually infantile without hormone replacement. Milder mutations result in ambiguous genitalia in male patients. Although cortisol synthesis is also abolished, even severely affected individuals are able to synthesize corticosterone, an active glucocorticoid, as well as aldosterone. Thus, they do not develop adrenal insufficiency. On the contrary, they secrete excessive amounts of deoxycorticosterone, which has mineralocorticoid activity, and are therefore prone to develop hypertension.

Whereas most *CYP17* mutations affect both the hydroxylase and lyase activities, rare mutations can affect the lyase activity alone. Additionally, mutations in genes other than *CYP17* can have the same phenotype as 17,20, lyase deficiency (i.e., deficient androgen synthesis with normal cortisol synthesis). These include an accessory electron transfer protein, cytochrome b_5, and mutations in the genes for two aldo-keto reductases, AKR1C2 and AKR1C4. These AKR1C isozymes normally catalyze 3α-hydroxysteroid dehydrogenase activity, which allows synthesis of the potent androgen, dihydrotestosterone, through an alternative "backdoor" biosynthetic pathway that does not include testosterone as an intermediate.

3β-Hydroxysteroid Dehydrogenase Deficiency

Severe mutations in the *HSD3B2* gene prevent the synthesis of aldosterone, cortisol, testosterone, and estrogens. Because DHEA, a weak androgen, is synthesized and secreted at high levels, some degree of phallic growth is present. Thus, affected male patients have severely ambiguous genitalia, but affected female patients may have clitorimegaly. Both sexes develop adrenal insufficiency if they are untreated.

Because many children with premature adrenarche (early development of axillary and pubic hair), as well as many women with polycystic ovary syndrome, have elevated levels of DHEA, it was once thought that such individuals could have a mild form of HSD3B2 deficiency. However, mutations in *HSD3B2* are rarely if ever found in such individuals, who instead have an imbalance in the relative levels of HSD3B2 and CYP17 activity within the adrenal cortex.

21-Hydroxylase Deficiency

More than 90% of cases of CAH are caused by 21-hydroxylase deficiency resulting from mutations in the *CYP21* (or *CYP21A2)* gene. *CYP21* and a highly homologous pseudogene, *CYP21P (or CYP21A1P),* are located within the major histocompatibility complex on chromosome 6p21.3, a genomic region noteworthy for a high rate of recombination. More than 90% of all mutations are the result of intergenic recombination between *CYP21* and *CYP21P*. Most are transfers of deleterious mutations from *CYP21P* to *CYP21,* whereas 20% are net deletions of *CYP21* resulting from unequal meiotic crossover.

In patients with 21-hydroxylase deficiency, the adrenals produce excess 17-hydroxyprogesterone, 17-hydroxypregnenolone, and progesterone, which are further metabolized to DHEA and androstenedione. Once secreted, these substances are further metabolized to active androgens (testosterone and dihydrotestosterone) and, to a lesser extent, to estrogens (estrone and estradiol).

Adrenal secretion of excess androgen precursors does not significantly affect male sexual differentiation. In affected female patients, the urogenital sinus is in the process of septation when the fetal adrenal begins to produce excess androgens, which function to prevent the formation of separate vaginal and urethral canals. Further adrenal-derived androgens interact with androgen receptors in genital skin and induce clitoral enlargement, promote fusion of the labial folds, and cause rostral migration of the urethral/vaginal perineal orifice. However, internal wolffian structures such as the prostate gland and spermatic ducts are usually not virilized, presumably because development of the wolffian ducts requires markedly higher local concentrations of testosterone than the external genitalia. Nevertheless, severely affected female patients occasionally have some development of typically male internal genital structures.

Thus, the typical result in severely affected girls is ambiguous or male-appearing external genitalia with perineal hypospadias and chordee, but without palpable testes (Fig. 233-4). The severity of virilization is often quantitated using a five-point scale developed by Prader (Fig. 233-5). Not all female patients with classic CAH develop the same degree of genital ambiguity. One could speculate that the physical signs of androgen excess depend not only on direct adrenal secretion of androgen precursors but also on the efficiency with which such hormones are converted to more potent products, such as dihydrotestosterone, by peripheral enzymes such as 5α-reductase. Additionally, the concentration and transcriptional activity of androgen receptors may play a role in determining genital phenotype.

Most patients (75%) cannot synthesize sufficient aldosterone to maintain sodium balance and are termed *salt wasters.* These patients are predisposed to episodic and potentially life-threatening hyponatremic dehydration. Patients with sufficient aldosterone production to prevent salt wasting and who have signs of prenatal virilization and/or markedly increased production of hormonal substrates of 21-hydroxylase (e.g., 17-hydroxyprogesterone) are termed *simple virilizers.*

11β-Hydroxylase Deficiency

Patients with 11β-hydroxylase deficiency have mutations in the *CYP11B1* gene. They have elevated levels of deoxycorticosterone and 11-deoxycortisol, as well as earlier cortisol precursors such as 17-hydroxyprogesterone. These patients secrete excess adrenal androgens, with consequences similar to those

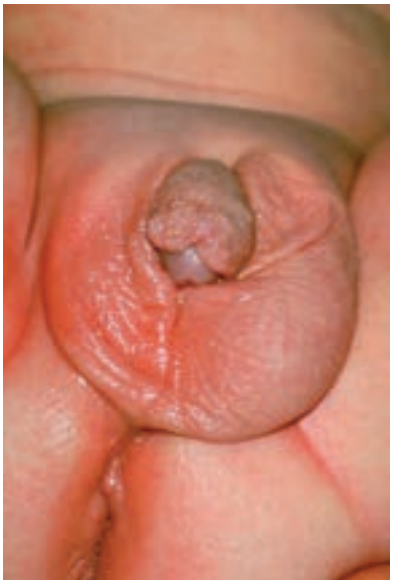

FIGURE 233-4. Virilized external genitalia in a female infant with congenital adrenal hyperplasia caused by 21-hydroxylase deficiency. No gonads are present in the scrotum.

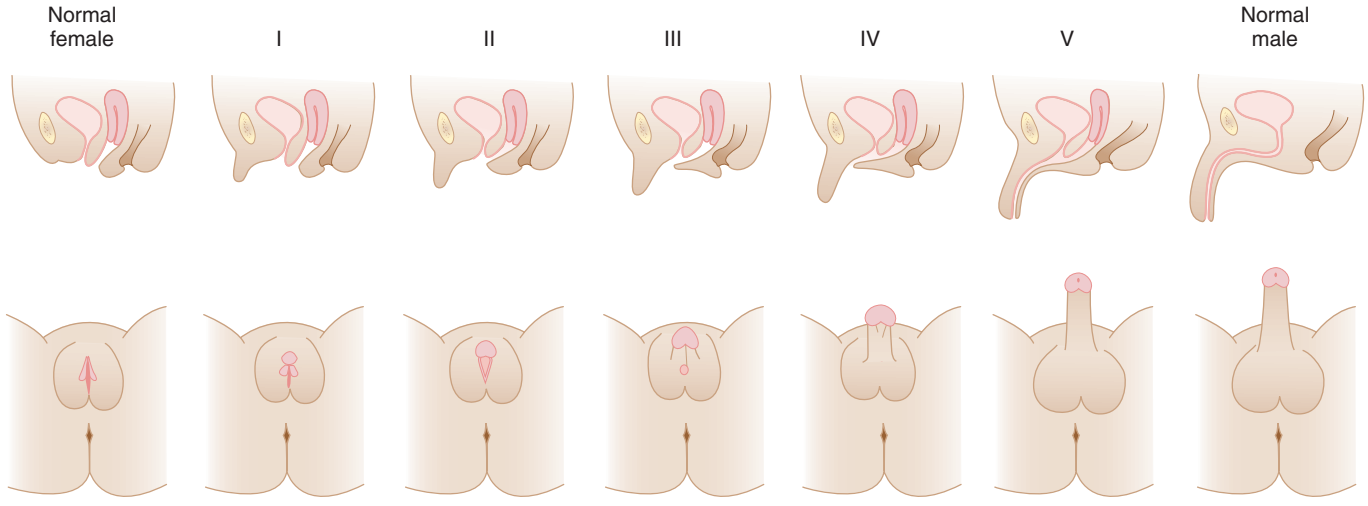

FIGURE 233-5. Abnormal differentiation of the urogenital sinus and external genitalia. Schematic representations of normal female and male anatomy flank a series of schematics illustrating different degrees of virilization of females, graded using the scale developed by Prader. The uterus (*shaded*) persists in virilized females even when the external genitalia have a completely masculine appearance (Prader grade V). (Modified from Prader A. Der Genitalbefund beim Pseudohermaphroditismus femininus der kengenitalen adrenogenitalen Syndroms. *Helv Paediatr Acta*. 1954;9:231-248.)

seen in 21-hydroxylase deficiency. However, patients with 11β-hydroxylase deficiency synthesize aldosterone normally and do not have problems with salt wasting. Instead, they are likely to become hypertensive as a result of elevated levels of deoxycorticosterone and its metabolites.

DEFECTS OF ANDROGEN BIOSYNTHESIS

Lipoid hyperplasia, 17-hydroxylase/17,20 lyase deficiency, and HSD3B2 deficiency affect the biosynthesis of both corticosteroids and sex hormones. In contrast, two enzymatic defects affect only androgen biosynthesis. They have similar phenotypes. Affected male patients are born with ambiguous genitalia, but they virilize at puberty and often reassign themselves to a male gender if they were raised as females. They have absent müllerian structures as a result of the secretion of AMH by the testes.

17-Ketosteroid Reductase (17-Hydroxysteroid Dehydrogenase 3) Deficiency

This disorder is caused by mutations in the *HSD17B3* gene. Although testosterone is not synthesized well, androstenedione, an active androgen, is synthesized. Because several other isozymes have 17-ketosteroid reductase activity in other tissues, some testosterone is invariably synthesized, especially at puberty, when circulating levels of androstenedione increase.

5α-Reductase Deficiency

Patients with 5α-reductase deficiency resulting from mutations in the *SRD5A2* gene synthesize entirely normal amounts of testosterone, but they cannot synthesize adequate amounts of dihydrotestosterone, the most potent naturally occurring androgen. This enzyme is not expressed at high levels in the testes (circulating levels of dihydrotestosterone are relatively low); instead, it is expressed in genital skin. Internal wolffian structures do not require this enzyme and are intact; high testosterone levels at puberty induce significant phallic growth without 5α-reductase activity.

OTHER DEFECTS OF STEROIDOGENESIS

Aromatase Deficiency

Mutations in *CYP19* cause aromatase deficiency in both the fetus and the placenta. The placenta can convert DHEA sulfate to androstenedione and testosterone normally, but it cannot convert these androgens to estrone and estradiol. These androgens accumulate in both the fetal and maternal circulations and virilize both the mother and the affected fetus if it is female. Affected male infants are phenotypically normal. Affected female patients virilize further at puberty if they are untreated. The lack of aromatase activity within bone leads to tall stature in both sexes (because estrogens are required to close the growth plates) and later to osteoporosis.

Cytochrome P-450 Oxidoreductase Deficiency

This deficiency is one form of Antley-Bixler syndrome, which is characterized by skeletal anomalies and craniosynostosis; most patients also have ambiguous genitalia associated with mutations in the *POR* gene. Antley-Bixler syndrome without genital abnormalities is caused by mutations in the FGF receptor (*FGFR2*) gene.

Because this disorder affects the activity of all microsomal cytochrome P-450 subsets, complete deficiency of cytochrome P-450 oxidoreductase (POR) is lethal, and identified mutations in humans yield POR with partial activity. *POR* mutations may cause skeletal anomalies by interfering with cholesterol synthesis. Genital ambiguity is caused by several mechanisms. Decreased activity of 17α-hydroxylase/17,20 lyase (CYP17) affects androgen synthesis and leads to undervirilization in males. Conversely, decreased activity of 21-hydroxylase (CYP21) can virilize affected females. These two deficiencies also can cause adrenal insufficiency. Finally, decreased activity of placental aromatase (CYP19) virilizes both the mother and the affected fetus (if female).

Smith-Lemli-Opitz Syndrome

This relatively frequent (1 in 20,000 northern Europeans) disorder of the final step of cholesterol biosynthesis (conversion from 7-dehydrocholesterol) is caused by mutations in the *DHCR7* gene encoding 7-dehydrocholesterol reductase. The syndrome is characterized by multiple congenital anomalies, including being small for gestational age, short stature, microcephaly, mental retardation, aggressive behavior, seizures, hypotonia, polydactyly, cleft palate, cardiac defects, lung hypoplasia, and renal anomalies. Male patients have ambiguous genitalia. The range of clinical severity is wide and depends on the nature of the mutations. Pathogenetic mechanisms for the ambiguous genitalia may include insufficient provision of cholesterol for steroid hormone biosynthesis or toxic effects of the precursor, 7-dehydrocholesterol, on steroidogenic cells.

MATERNAL CONDITIONS AFFECTING THE FETUS

LUTEOMA OF PREGNANCY

Luteomas are the most common causes of maternal virilization during pregnancy. They often occur bilaterally. Although many luteomas are discovered incidentally during cesarean sections or postpartum tubal ligations, one fourth of mothers virilize during the latter half of pregnancy, and half of female infants born to these mothers also exhibit signs of virilization, most typically clitorimegaly and labial fusion. Spontaneous regression of the luteoma generally begins within days after delivery.

DRUG EXPOSURE

Depending on the agent, maternal drug exposure may affect either male or female fetuses. Females may be virilized by androgens such as 19-nor-testosterone or progestins administered to prevent spontaneous abortion. Undervirilized males can be born to women exposed to 5α-reductase inhibitors such as finasteride. The antifungal agent fluconazole can inhibit many cytochrome P-450 enzymes and can lead to a condition closely resembling Antley-Bixler syndrome.

The synthetic estrogen diethylstilbestrol was used several decades ago to prevent spontaneous abortion (it was actually ineffective for this purpose). Males exposed to this agent in utero were born with testicular hypoplasia, cryptorchidism, hypospadias, and/or microphallus. Females had uterine, cervical, and vaginal abnormalities and an increased risk for clear cell adenocarcinoma of the vagina. Considering that many cases of genital ambiguity are idiopathic, it is likely that additional endocrine disruptors in the environment have not yet been identified.[7]

HORMONE INSENSITIVITY SYNDROMES AND OTHER HORMONE DEFICIENCIES

ANDROGEN INSENSITIVITY

Males normally carry a single copy of the X-linked androgen receptor (AR) gene.[8] Thus, a single mutation can completely inactivate the receptor in males and lead to complete androgen insensitivity (formerly termed *testicular feminization syndrome*). This is one of the most frequent forms of 46,XY DSD, occurring in approximately 1 in 20,000 male births.[9]

Patients with the complete form of androgen insensitivity have normal female external genitalia. Unless suspicion is raised by prior knowledge of the infant's karyotype, the condition is rarely discovered before puberty unless the testes are palpated in the groin or labia on routine examination. Because the testes secrete AMH, müllerian structures are absent, including the uterus, fallopian tubes, and cervix. Thus, the vagina is usually shallow and ends blindly. Wolffian structures are also absent. The testes may be located in the abdomen or in the labia majora and do not undergo spermatogenesis. AMH levels are elevated during the first year and (if the testes have not been removed) after puberty. Testosterone and LH levels in infancy and at puberty are elevated as a result of defective feedback regulation caused by androgen resistance at the level of the hypothalamus.

At puberty, pubic and axillary hair is scant or absent. Testosterone can be aromatized to estradiol by CYP19 in breast fat, and estrogen receptors are unaffected in this condition. Thus, breast development is that of a normal female.

Partial androgen insensitivity (Reifenstein's syndrome) is characterized by a variable degree of genital ambiguity, and both virilization and breast development occur at puberty. Mild androgen insensitivity also can occur with a male phenotype, with gynecomastia and infertility as the sole manifestations. Mutations in the androgen receptor are not detected in many mild cases, which may result from defects in other transcription factors affecting actions of the receptor.

LEYDIG CELL AGENESIS

Leydig cell agenesis or hypoplasia is a rare autosomal recessive syndrome caused by mutations in the *LHGCR* gene encoding the LH receptor. Without stimulation by LH (or by HCG early in gestation), Leydig cells do not differentiate normally and do not secrete testosterone. Thus, affected male infants are born with female-appearing or ambiguous external genitalia. Müllerian structures are absent because of unaffected secretion of AMH by Sertoli cells. LH levels are high in infancy and at puberty, and they respond normally to gonadotropin-releasing hormone, whereas testosterone levels are low and do not respond to stimulation by HCG. Affected females are

phenotypically normal but may have oligomenorrhea resulting from primary ovarian dysfunction.

PERSISTENT MÜLLERIAN DUCT SYNDROME

Persistent müllerian duct syndrome (PMDS) is a rare autosomal recessive condition that results from mutations in the genes for either AMH (PMDS type I) or the AMH receptor (*AMHR2* gene, PMDS type II). The two are distinguished clinically by low or absent AMH levels in patients with AMH mutations and by AMH levels in the high-normal range in those with AMH receptor mutations.

Affected male patients have unimpaired testosterone secretion and thus have normal external genitalia and wolffian structures. However, the lack of AMH action prevents regression of müllerian structures, so these patients also retain a uterus and fallopian tubes. These structures are often closely approximated to the vas deferens. The müllerian structures are usually dragged into the inguinal canal by the descending testes. However, these structures typically prevent the testes from descending into the scrotum and thus cause bilateral inguinal hernias (with the uterus on one side) and bilateral or occasionally unilateral cryptorchidism. The condition is usually discovered only at surgery. Fertility in affected patients may be normal or impaired, with an increased risk for malignant disease in undescended testes left in the abdomen.

HYPOGONADOTROPIC HYPOGONADISM

Milder or later appearing deficiencies of androgen biosynthesis (after 13 to 14 weeks) may allow complete fusion of the labioscrotal folds and normal positioning of the urethral meatus, but subsequent growth of the phallus is suboptimal. Such individuals have a micropenis. The most common cause is lack of gonadotropin (specifically, LH) secretion; even when LH is lacking, early male development is normal because testosterone secretion is controlled mostly by HCG during the first trimester.

Defective LH and follicle-stimulating hormone secretion can result when the neurons that normally secrete gonadotropin-releasing hormone fail to migrate into the hypothalamus.[10] This condition, *Kallmann syndrome,* is most often X-linked, resulting from mutations in the *KAL1* gene. It is often associated with anosmia. Other conditions that affect hypothalamic development and cause hypogonadotropic hypogonadism include *Prader-Willi syndrome,* which is a result of paternal deletions, methylation defects, and maternal uniparental disomy of imprinted loci on chromosome 15q12. Children with this syndrome have a characteristic appearance consisting of a narrow bitemporal diameter, almond-shaped eyes with an antimongoloid slant, and small hands and feet. They typically have marked hypotonia as infants, with subsequent moderate developmental delay and slow somatic growth. Hypothalamic obesity develops during childhood. Patients with adrenal hypoplasia congenita resulting from mutations in the DAX1 transcription factor have defective development of the ventromedial hypothalamus and consequent hypogonadotropic hypogonadism, associated with adrenal insufficiency that typically manifests with aldosterone deficiency and salt wasting. Steroidogenic factor-1 (SF-1), encoded by the *NR5A1* gene, is an orphan nuclear receptor that is critical for the development and function of the adrenal glands, gonads, pituitary gonadotropes, ventromedial nucleus of the hypothalamus, and male sexual differentiation. Heterozygous null mutations or homozygous milder mutations have been identified mainly among undervirilized 46,XY individuals and 46,XX women with premature ovarian failure. Only a minority have adrenal insufficiency.

Hypogonadotropic hypogonadism often results from failure of the entire anterior pituitary gland, or particular cellular populations therein, to develop.[11] Pituitary gland abnormalities can be associated with other midline defects, including hypoplasia of the optic nerves and the septum pellucidum, a condition termed *septo-optic dysplasia.* Associated pituitary hormone deficiencies may include growth hormone, ACTH, and thyroid-stimulating hormone. These deficiencies may manifest in the neonatal period as hypoglycemia or hypothyroidism (detected by newborn screening programs). Optic nerve dysfunction is difficult to detect by routine examination in the neonatal period, but it causes a characteristic wandering nystagmus after a few months of age.

Although panhypopituitarism is most often sporadic, mutations in transcription factors controlling pituitary development have been documented (Chapter 224), particularly PROP1, and septo-optic dysplasia has been associated with mutations in the *HESX* gene. Rarely, mutations in the gene encoding the β-subunit of LH may yield a phenotype similar to hypogonadotropic hypogonadism.

OTHER GENETIC CONDITIONS

ANEUPLOIDY OF SEX CHROMOSOMES

Turner Syndrome

Patients with Turner syndrome have normal female external genitalia and a normal uterus and fallopian tubes, but they have dysgenetic streak ovaries.[12] Most fetuses with Turner syndrome spontaneously abort, but the incidence in live births is approximately 1 in 2500. Classically, the karyotype is 45,X, but many patients retain an abnormal second X chromosome or even a fragment of a Y chromosome lacking *SRY*. Other patients are mosaic for 46,XX and 45,X cells and may have relatively mild phenotypes.

Untreated patients are short. Many have typical dysmorphic features, including lymphedema of the neck at birth, webbed neck, low posterior hairline, increased carrying angle of the arms, shield chest with widely spaced nipples, low-set ears, and micrognathia. Patients typically have primary amenorrhea and are infertile, but they occasionally have menarche followed by premature ovarian failure.

Klinefelter's Syndrome

In this condition, male patients have normal development of the penis and scrotum, but the testes are small and firm. Patients tend to be tall. At adolescence, gynecomastia is frequent. Signs of testosterone deficiency occur in most affected adults, and most have azoospermia. The usual karyotype is 47,XXY. Hormonal findings include elevated gonadotropin levels and a decreased serum testosterone concentration. Klinefelter's syndrome is a common disorder that occurs in 1 in 500 to 1000 men.

Mixed Gonadal Dysgenesis

Mixed gonadal dysgenesis, a frequent cause of sexual ambiguity, occurs in approximately 1 in 20,000 births.[13] The karyotype is usually mosaic 45,X/46,XY. Gonadal pathologic features can vary from fibrous streaks indistinguishable from those in Turner syndrome to normally developed testes and a normal male phenotype. Typically, patients have a testis on one side and a fibrous streak on the other. Some patients may have a Turner-like phenotype. A fallopian tube is usually present on the side of the streak gonad. Leydig cell function, evaluated by testosterone response to HCG, and Sertoli cell function, evaluated by serum AMH levels, vary from poor to normal.

XX MALE SYNDROME

Males with a 46,XX karyotype have normal external and internal male genitalia; however, they resemble patients with Klinefelter's syndrome in that they have small testes, azoospermia, and infertility. Translocation of the *SRY* gene to the X chromosome is detected in 75 to 90% of sporadic cases; this can occur because the gene is located very near the pseudoautosomal region, where the short arms of the X and Y chromosomes are homologous and meiotic recombination is possible. Duplication of the SOX9 transcription factor may be responsible for some familial cases of XX sex reversal.

XY FEMALE SYNDROMES

Patients with pure XY gonadal dysgenesis (Swyer syndrome) have a normal female phenotype, including a uterus and fallopian tubes, but they have streak gonads. These patients are free of Turner-like malformations and attain normal height. Mutations of the *SRY* gene have been identified in 15% of cases. Unlike 45,X patients with Turner syndrome, these patients have an increased risk for gonadoblastoma.

Similar phenotypes result from duplication of the region of the X chromosome containing the *DAX1* gene, from duplication of the *WNT4* gene, or from haploinsufficiency of the SF1 transcription factor (see earlier). XY sex reversal also can result from mutations in the SOX9 transcription factor associated with campomelic dysplasia, a form of dwarfism. Mutations of *DHH* cause XY gonadal dysgenesis, associated with peripheral neuropathy.

Some 46,XY patients with absent gonads have various degrees of sexual ambiguity and no müllerian derivatives. The implication that some testicular tissue was functional at least up to 10 weeks' gestation and subsequently regressed led to the name *fetal testicular regression syndrome.* Testicular regression may occur in late pregnancy or even postnatally; these fully virilized male patients have isolated anorchia.

VAGINAL ATRESIA

Mayer-Rokitansky-Küster-Hauser syndrome refers to aplasia of the uterus and upper vagina, occurring in approximately 1 in 5000 women. In approximately

one third of cases, it occurs along with other abnormalities including unilateral renal aplasia, and cervicothoracic somite dysplasia (MURCS association). The genetic basis is unknown in most cases. Rare affected individuals have heterozygous mutations of *WNT4;* such patients usually have clinical and biochemical signs of androgen excess.

TRUE HERMAPHRODITISM (OVOTESTICULAR DISORDER OF SEXUAL DEVELOPMENT)

True hermaphroditism (ovotesticular DSD), a rare and usually sporadic disorder, is defined as the coexistence of seminiferous tubules and ovarian follicles. Most patients have an ovotestis with either an ovary or a testis on the opposite side; a gonad in the scrotum is usually a testis but may be an ovotestis.

The genitalia are usually ambiguous, but they may appear completely masculine or feminine. The anatomy of the internal reproductive tract depends on the nature of the gonads, particularly whether they secrete AMH. A uterus or uterine horn is present in 90% of cases. Testosterone response to HCG is variable, and AMH levels are usually low. Most patients experience breast development, ovulation, and even menstruation at puberty; pregnancy and successful childbirth are possible if selective removal of testicular tissue is feasible. Unless sex of rearing has already been chosen, male gender assignment should be restricted to patients with no uterus and descended testicular tissue because the latter is usually dysgenetic and prone to malignant degeneration. Most patients with ovotesticular DSD have a 46,XX karyotype. Despite the presence of testicular tissue, they usually lack *SRY;* this suggests that the condition is the result of constitutive activation of a gene normally triggered by *SRY*.

MANAGEMENT OF INDIVIDUALS WITH DISORDERS OF SEXUAL DEVELOPMENT: GENDER ROLE AND IDENTITY

The influence of prenatal sex steroid exposure or epigenetic changes in the estrogen receptor[14] on personality is controversial.[15] In considering this question, it is important to distinguish among gender role, sexual orientation, and gender identity.

Gender Role

Gender role refers to gender-stereotyped behaviors, such as the choice of toys by young children. For example, parents of young girls with CAH often report that their daughters prefer to play with trucks rather than dolls and tend to be tomboyish later in childhood. Decreased interest in maternal behavior, beginning with infrequent doll play in early childhood and extending to lack of interest in child rearing in older girls and women, occurs frequently.

Sexual Orientation

Sexual orientation refers to homosexual versus heterosexual preferences. In many studies, a significant minority of women with CAH have been actively homosexual or bisexual or have had an increased tendency toward homoerotic fantasies. These characteristics occur more frequently in women with the salt-wasting form of 21-hydroxylase deficiency, suggesting that they are a consequence of prenatal exposure of the brain to androgens. However, the vast majority of both male and female homosexuals have no identifiable endocrinologic abnormality.

Gender Identity

Gender identity refers to self-identification as male or female. Gender self-reassignment back to male has been reported in cases of male patients with penile trauma or exstrophy of the bladder who were raised as girls. This may also occur in 46,XY patients with disorders of sexual development (DSD) raised as girls, especially in cases of 5α-reductase or 17-ketosteroid reductase deficiencies, in which the brain may be exposed to high circulating levels of androgens. Self-reassignment to the male gender is unusual in women with CAH. When it occurs, it may be related to delays in gender assignment or genital surgery or to inadequate suppression of adrenal androgens with glucocorticoid therapy.

Transgendered individuals rarely have identifiable hormonal abnormalities; nonhormonal mechanisms governing gender identity are poorly understood. Gender identity disorders are much more likely to occur in both identical twins than in fraternal twins, suggesting a high degree of heritability. Neuroanatomic studies suggest that the bed nucleus of the stria terminalis is larger in males and female-to-male transsexuals and smaller in females and male-to-female transsexuals. Similar findings involving other sexually dimorphic brain regions have been identified by MRI. Thus, gender identity disorder may be considered a brain-limited form of DSD.

DIAGNOSIS

Management of a child born with ambiguous genitalia presents a difficult challenge to medical personnel.[16] It is important to refrain from assigning sex until diagnostic information can be gathered. Usually, test results can be obtained within 24 to 48 hours and parents can be advised about the child's chromosomal and gonadal sex and the anatomy of internal sexual structures.

In addition, the physician must keep in mind that DSD may be associated with life-threatening biochemical or anatomic abnormalities. In particular, the most common cause of severely masculinized external genitalia in females, the salt-wasting form of CAH resulting from steroid 21-hydroxylase deficiency, may cause hyponatremia, hyperkalemia, hypovolemia, and shock. In contrast, male patients with ambiguous genitalia may have lipoid adrenal hyperplasia or a salt-wasting form of 3/3-hydroxysteroid dehydrogenase (HSD3B2) deficiency. Males with micropenis may have panhypopituitarism; in this case they are at risk for significant hypoglycemia and hyponatremia resulting from low cortisol (because of low ACTH) and low growth hormone levels, or they may have adrenal hypoplasia congenita, in which case they could have adrenal insufficiency. Finally, patients with ambiguous genitalia are at increased risk for renal anomalies, or they may have chromosomal syndromes with other associated anomalies.

History

The gestational history should concentrate on potential exposure to agents that could interfere with normal sexual differentiation. For a female infant with virilized genitalia, these include progestational agents, whereas the mother of a male with incompletely masculinized genitalia may have been exposed to a 5α-reductase inhibitor through her husband's use of such an agent for male pattern baldness or prostate enlargement. It should be determined whether amniocentesis and karyotyping have been performed. A family history should elicit similar cases of genital ambiguity or cases of sudden death, which could raise suspicion of undiagnosed salt-wasting CAH or adrenal hypoplasia congenita.

Physical Examination

The physical examination should document the size of the phallus (clitoris or penis), the degree of chordee (ventral bowing of the phallus), and the extent of fusion of the labioscrotal folds. The urethral meatus should be identified, and there must be careful palpation for gonads in the inguinal canals and labia or scrotum. Bilateral cryptorchidism, even if an isolated finding in a phenotypic male patient, should always lead to evaluation for a possible DSD.

Biochemical Evaluation of the Virilized Female

The minimal diagnostic tests should include measurement of basal serum 17-hydroxyprogesterone, androstenedione, and testosterone. Preferably, a complete profile of adrenocortical hormones is obtained before and 1 hour after stimulation of the adrenal cortex with 125 to 250 μg of cosyntropin ($ACTH_{1\text{-}24,}$). These assays should be deferred until after the first 24 hours of life. They will identify potential defects in adrenal steroidogenesis (i.e., CAH); 21-hydroxylase deficiency is identified by elevations in 17-hydroxyprogesterone, whereas 11-deoxycortisol and 11-deoxycorticosterone are high in 11β-hydroxylase deficiency.

Biochemical Evaluation of the Undervirilized Male

In 46,XY DSD patients, it is necessary to test adrenal and gonadal function as well as extragonadal androgen metabolism. With regard to adrenal defects, 11-deoxycorticosterone and the ratio of pregnenolone to 17-hydroxypregnenolone are high in 17α-hydroxylase deficiency, 17-hydroxypregnenolone and DHEA are high in HSD3B2 deficiency, and all steroids are low in lipoid hyperplasia.

Defects in gonadal steroidogenesis are best evaluated after stimulation with HCG (1500 IU intramuscularly on days 1, 3, and 5, with blood drawn on day 6). However, 17-hydroxylase and HSD3B2 deficiencies affect both the gonads and the adrenal cortex and thus are often diagnosed by cosyntropin stimulation testing. Low levels of all androgen precursors suggest lipoid hyperplasia, 17α-hydroxylase/17,20 lyase deficiency, or a generalized defect in testicular function, such as the vanishing testes syndrome (testicular regression-syndrome) or gonadotropin insensitivity. A high ratio of

androstenedione to testosterone is indicative of 17-ketosteroid reductase (HSD17B3) deficiency, and a high ratio of testosterone to dihydrotestosterone is diagnostic of 5α-reductase deficiency. The diagnosis of androgen insensitivity syndrome is suspected when a 46,XY patient has ambiguous or female-appearing external genitalia despite normal or high circulating levels of testosterone and dihydrotestosterone.

Gonadal Biopsies

Patients with mixed gonadal dysgenesis, true hermaphroditism, or unclear diagnoses should undergo bilateral gonadal biopsies (histology of the two gonads is often not identical). Dysgenetic gonads have a high potential for malignant transformation and usually need to be removed in childhood.

TREATMENT

Rx

Initial Medical Management

Patients with CAH resulting from 21-hydroxylase or HSD3B2 deficiencies or those with lipoid hyperplasia or adrenal hypoplasia congenita require replacement of both glucocorticoids and mineralocorticoids, usually with hydrocortisone (15 to 20 mg/m^2/day in divided doses) and fludrocortisone (usually 0.1 mg/day, but as much as 0.4 mg/day in neonates with salt-wasting crises). Neonates with severe salt losses may require sodium chloride supplementation (≤8 mEq/kg/day). Patients with 11β-hydroxylase or 17α-hydroxylase deficiencies have normal aldosterone biosynthesis and usually require only glucocorticoids. Patients with panhypopituitarism usually require treatment with hydrocortisone, thyroxine, and growth hormone.

All male infants with ambiguous genitalia or micropenis in whom rearing as a boy is contemplated should have a 3- or 4-month therapeutic trial of monthly depot testosterone injections (25 mg) to attempt to increase the size of the phallus during infancy. This treatment may improve social acceptability of the genitalia later in childhood and adolescence and may make reconstructive surgery easier. In cases of suspected partial androgen insensitivity, this treatment also documents the degree to which the patient is androgen responsive and thus may provide useful information about whether rearing as a boy is feasible. Higher doses of testosterone (75 mg every 4 weeks) may be used under these circumstances.

Considerations Related to Sex Assignment

In large medical centers, a team consisting of a neonatologist, a pediatric endocrinologist, a urologist, and preferably an experienced social worker and/or child psychiatrist or psychologist should promptly review the early diagnostic data and make a recommendation to the family as to the sex of rearing and any medical or surgical treatments. These recommendations should be based on both current knowledge of psychosexual development in DSD individuals and the feasibility of surgical treatment (see later).[17]

In general, the recommended sex assignment should be that of the genetic/gonadal sex, if for no other reason than to retain the possibility of reproductive function. This is especially true for female infants with CAH who have normal internal genital structures and the potential for childbearing. An exception may be considered in a genetically female infant with completely male-appearing genitalia, especially if the child has been raised as a boy for more than a few months. Such children need to be castrated at puberty to avoid feminization.

Conversely, genetic male infants with completely female-appearing external genitalia (usually resulting from complete androgen insensitivity syndrome, but also seen with severe testosterone biosynthetic defects) should be raised as female because the potential for reconstruction of male genitalia is poor. They, too, need to be castrated by early adulthood to avoid malignant transformation of the testes. However, male infants with 17-ketosteroid reductase or 5α-reductase deficiency should usually be reared as boys because they have normal levels of androstenedione or testosterone, respectively, and often virilize significantly at puberty. Indeed, many of these patients reassign themselves to the male gender when they are made aware of the diagnosis. The same considerations pertain to male patients with normal testosterone biosynthesis who have penile trauma or anatomic abnormalities such as bladder exstrophy.

Recommendations for sex assignment are to some extent culture specific. In cultures that value boys over girls, parents may strongly resist rearing a female infant with ambiguous genitalia as a girl, and many girls with severely virilized external genitalia will be raised as boys.

Surgical Management

Surgery for Ambiguous Genitalia

Whether, how, and when to intervene surgically in the correction of genital anomalies are the subject of continuing debate. Some adult patients with DSD conditions are unhappy with their gender assignments or surgical outcomes. Some physicians advocate postponing cosmetic genital surgery until the affected individual is able to provide informed consent, thus keeping all the options open if the adult patient wishes to function sexually with abnormal genitalia that nevertheless have sensation undiminished by surgery or chooses to reassign his or her gender. Declining or postponing surgery should not be confused with raising the child with an indeterminate gender, a concept currently well outside the mainstream. The option of deferring surgery should always be presented as part of the informed consent process. Nevertheless, most parents want their child to look as "normal" as possible, and they often resist suggestions to postpone genital surgery.

In addressing this question, it is best to consider the various general types of genital surgery separately. The greatest change in practice over the past few decades probably pertains to male infants with ambiguous (but not completely female) external genitalia. Physicians are far less likely to recommend that such patients be reared as female because it is now recognized that many of these patients reassign themselves as male at puberty. Thus, the ambiguous genitalia in such patients should rarely be "corrected" to female. On the contrary, surgical techniques for hypospadias repair have advanced significantly, and reconstruction of male genitalia is attempted more often, particularly if the infant responds to a course of testosterone with significant phallic growth.

Surgery for female infants with ambiguous genitalia may need to address an enlarged clitoris, the lack of a vaginal introitus, and the presence of a urogenital sinus. The clitoris is normally prominent in many infant girls. Even when enlarged in a girl with virilizing CAH, the clitoris can be prevented from growing larger with adequate suppression of adrenal androgens by glucocorticoids, and it will become less prominent as the patient grows. Thus, mild-to-moderate clitorimegaly is often best managed without surgery. When clitoroplasty is attempted, one must keep in mind the important role of clitoral sensation in the female sexual response. Such surgery must be performed only by experienced operators with scrupulous attention to the preservation of clitoral innervation.

Consensus is still lacking regarding the best age for vaginoplasty. Although many surgeons advocate a first procedure in infancy, it is difficult to maintain a functionally adequate introitus in the absence of estrogen exposure and mechanical dilation (with dilators or through sexual intercourse), and many patients require reoperation as young adults. Conversely, many women with atresia of the upper vagina (owing to complete androgen insensitivity or Mayer-Rokitansky-Küster-Hauser syndromes) can use dilators to lengthen the vagina without the need for surgery.

There is a dearth of large longitudinal studies comparing outcomes in patients who have had early genital surgery versus those who have had no surgery or surgery in adolescence. According to self-assessment surveys among sexually active women with CAH who have had genital surgery, most are able to have satisfactory sexual intercourse. As surgical and medical treatment regimens have improved in recent years, more women with CAH have successfully conceived spontaneously, completed pregnancies, and given birth. Most often, delivery is by cesarean section because of an inadequate introitus, but vaginal delivery is possible in some cases.

Hypospadias repair is usually begun in the first year of life, after testosterone treatment (if necessary to increase phallic size). Depending on the degree of hypospadias, more than one surgical procedure may be required.

Removal of Intra-abdominal Testes in 46,XY Patients with Disorders of Sexual Development

Intra-abdominal testes are at increasing risk for malignant transformation over time. In a boy with cryptorchidism who is being reared as male, orchiopexy should be performed as quickly as possible; this also maximizes the possibility of fertility when the underlying condition does not preclude it. Dysgenetic gonads that cannot be brought into the scrotum should be removed soon after diagnosis because the risk for malignant transformation in childhood is relatively high.

There is a lack of consensus regarding nondysgenetic testes in severely undervirilized genetic male infants in whom rearing as female is planned. In patients with complete androgen insensitivity or complete defects in testosterone biosynthesis, no possibility of fertility exists, and so there seems to be no reason to retain the testes. Conversely, the risk for malignant transformation in such gonads is low before puberty, and patients with complete androgen insensitivity can undergo spontaneous breast development at puberty. At that time, patients themselves can assent or consent to gonadectomy, which can usually be accomplished laparoscopically. This is of particular importance in genetic male patients with partial androgen insensitivity or incomplete defects of testosterone biosynthesis, because such patients may eventually desire a male gender role.

Patients with persistent müllerian duct syndrome have a reduced but still appreciable potential for fertility, and virilization is unaffected. Thus, the testes should be removed only if they cannot be brought into the scrotum. Because the müllerian and wolffian structures are closely approximated in these patients, surgical excision of the uterus and fallopian tubes may result in ischemic and/or traumatic damage to the vas deferens and testes; thus, salpingectomy and hysterectomy are indicated only in patients whose müllerian structures limit intrascrotal placement of the testes.

Space does not permit extensive discussion of surgical management of transgendered adults; options for male-to-female transsexuals include genitoplasty and, for those who did not have hormonal management during adolescence (see the next section), breast augmentation, body contouring, and facial and/or laryngeal surgery to produce a more feminine appearance. Female-to-male transsexuals often desire breast reduction surgery or complete mastectomies.

Treatment of Transgendered Individuals

Children and Adolescents

Transgendered individuals should be treated by multidisciplinary teams that can provide psychosocial evaluation and support.[18,19] Prepubertal children do not require medical management. The majority of such children do not persist in their identification with the opposite sex, although many will be homosexual as adults. Persistently transgendered children may develop significant gender dysphoria (distress at functioning in their natal gender) when puberty commences and are at increased risk for self-injury and suicide as adolescence progresses. If at all possible, such children should be allowed to function in the desired gender role. The current standard of care in many centers for children who have lived in a transgendered role for at least 6 months is to delay pubertal progression until mid-adolescence (~16 years old), with use of gonadotropin-releasing hormone (GnRH) agonists such as leuprolide depot injections or histrelin (Supprelin) implants. This will prevent the development of secondary sexual characteristics that may be distressing to the patient and may present cosmetic barriers to functioning in the desired, non-natal gender. These include breast enlargement, widened hips, and gynecoid adipose distribution in natal females or penile enlargement, facial and body hair, laryngeal enlargement and deep voice, and prominent jaw in natal males.

When the patient is confirmed in the desired gender role (e.g., by living completely in that role for at least 1 year), treatment may begin with the appropriate sex hormones. In high doses, such treatment will itself suppress gonadotropin secretion, and the GnRH agonist may be discontinued. Female-to-male transsexuals can be treated with depot testosterone whereas male-to-female transsexuals can be treated with parental forms of estradiol. Oral estrogen preparations should be avoided because they tend to increase production of clotting factors by the liver and may thus increase the risk for thromboembolism.

Adults

The medical treatment of transgendered adults follows the same principles as for adolescents, except that because secondary sexual characteristics have already developed, prolonged treatment with GnRH agonists is unnecessary. However, continuing such treatment in male-to-female transsexuals permits use of much lower estradiol doses with a concomitant reduction in the risks associated with high dose estrogen treatment. Nevertheless, cost represents a barrier to the long term use of GnRH agonists.

Prenatal Diagnosis and Treatment

Many conditions causing ambiguous genitalia can be detected by karyotyping of chorionic villus samples (for chromosomal abnormalities) or by direct molecular genetic testing. In most cases, this information is useful only for counseling purposes. In the case of virilizing forms of CAH (particularly 21-hydroxylase deficiency), the mother of an affected female fetus can take dexamethasone (20 μg/kg/day), which can cross the placenta and suppress the fetal adrenal gland, thus reducing the secretion of androgens and ameliorating virilization of the external genitalia. To be most effective, this treatment should be started by the sixth week of gestation, before the sex or genotype of the fetus is known. Thus, seven unaffected or male fetuses must be treated with dexamethasone to avoid genital ambiguity in one affected female, until chorionic villus sampling can be performed. Although effective in reducing prenatal virilization, this dose of dexamethasone can cause Cushing syndrome in the mother, and the long-term sequelae in the fetus are not known. Recent consensus statements suggest that this treatment be used only under approved research protocols that allow for case registries and long-term follow-up.

Psychosocial Support

Families of patients with DSD should be assessed for emotional health, initially by the pediatrician and/or pediatric endocrinologist. Parents should be offered psychological counseling soon after the diagnosis is made. Intermittent assessment of family functioning may be a useful tool in predicting future problems. As children mature, they should repeatedly be informed about their condition by parents and physicians in a sensitive and age-appropriate manner. When psychotherapy is undertaken, medical and psychiatric caregivers should communicate with each other so both are aware of the patient's and family's status. Unfortunately, many locales lack mental health professionals with experience in counseling patients with DSD and their families.

Although the psychosexual development of individuals with DSD cannot be predicted with confidence, patients' families should receive anticipatory counseling. For example, counseling of parents of girls affected with CAH should address the high likelihood that such girls will exhibit tomboyish behavior, masculine play preferences, and perhaps, when older, a preference for a career over domestic activities. The endocrinologist and/or mental health professional (depending on inclination and experience) caring for the adolescents with DSDs should address sexual orientation, both fantasized and actual. For example, some women with CAH are most comfortable as homosexuals; such individuals should receive appropriate psychosocial support. Adult patients also should be made aware of relevant patient advocacy groups.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Arnold AP, Chen X, Itoh Y. What a difference an X or Y makes: sex chromosomes, gene dose, and epigenetics in sexual differentiation. *Handb Exp Pharmacol*. 2012;214:67-88.
2. Jørgensen A, Rajpert-De Meyts E. Regulation of meiotic entry and gonadal sex differentiation in the human: normal and disrupted signaling. *Biomol Concepts*. 2014;5:331-341.
3. Biason-Lauber A. WNT4, RSPO1, and FOXL2 in sex development. *Semin Reprod Med*. 2012;30: 387-395.
4. Josso N, Rey R, Picard JY. Testicular anti-Müllerian hormone: clinical applications in DSD. *Semin Reprod Med*. 2012;30:364-373.
5. Flück CE, Pandey AV. Steroidogenesis of the testis—new genes and pathways. *Ann Endocrinol (Paris)*. 2014;75:40-47.
6. Auchus RJ, Arlt W. Approach to the patient: the adult with congenital adrenal hyperplasia. *J Clin Endocrinol Metab*. 2013;98:2645-2655.
7. Vandenberg LN, Colborn T, Hayes TB, et al. Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. *Endocr Rev*. 2012;33:378-455.
8. Tadokoro-Cuccaro R, Hughes IA. Androgen insensitivity syndrome. *Curr Opin Endocrinol Diabetes Obes*. 2014;21:499-503.
9. Hughes IA, Davies JD, Bunch TI, et al. Androgen insensitivity syndrome. *Lancet*. 2012;380: 1419-1428.
10. Silveira LF, Latronico AC. Approach to the patient with hypogonadotrapic hypogonadism. *J Clin Endocrinol Metab*. 2013;98:1781-1788.
11. King TF, Hayes FJ. Long-term outcome of idiopathic hypogonadotropic hypogonadism. *Curr Opin Endocrinol Diabetes Obes*. 2012;19:204-210.
12. Levitsky LL, Luria AH, Hayes FJ, et al. Turner syndrome: update on biology and management across the life span. *Curr Opin Endocrinol Diabetes Obes*. 2015;22:65-72.
13. Ocal G, Berberoglu M, Siklar Z, et al. The clinical and genetic heterogeneity of mixed gonadal dysgenesis: does "disorders of sexual development (DSD)" classification based on new Chicago consensus cover all sex chromosome DSD? *Eur J Pediatr*. 2012;171:1497-1502.
14. Matsuda KI. Epigenetic changes in the estrogen receptor alpha gene promoter: implications in sociosexual behaviors. *Front Neurosci*. 2014;8:344.
15. Ubuka T, Tsutsui K. Review: neuroestrogen regulation of socio-sexual behavior of males. *Front Neurosci*. 2014;8:323.
16. Faisal AS, Achermann JC, Arlt W, et al. UK guidance on the initial evaluation of an infant or an adolescent with a suspected disorder of sex development. *Clin Endocrinol (Oxf)*. 2011;75:12-26.
17. Furtado PS, Moraes F, Lago R, et al. Gender dysphoria associated with disorders of sex development. *Nat Rev Urol*. 2012;9:620-627.
18. Spack NP. Management of transgenderism. *JAMA*. 2013 6;309:478-484.
19. Byne W, Bradley SJ, Coleman E, et al. Report of the American Psychiatric Association Task Force on Treatment of Gender Identity Disorder. *Arch Sex Behav*. 2012;41:759-796.

REVIEW QUESTIONS

1. Bilateral inguinal masses are palpated on routine examination of a newborn girl. There is a patent vaginal orifice. A karyotype is 46,XY. Which of the following disorders is *not* a possibility in this child?

A. Complete androgen insensitivity
B. 17-Hydroxylase deficiency
C. 11 β-Hydroxylase deficiency
D. 5α-Reductase deficiency
E. 17-Ketosteroid reductase deficiency

Answer: C This is a genetic male who has testes, with female-appearing external genitalia. The failure of development of normal male external genitalia must be due to failure of biosynthesis (answers B, D, or E) or action (answer A) of the most potent androgen, dihydrotestosterone. Deficiency of 17-hydroxylase (B) prevents synthesis of all androgens; 5α-reductase deficiency (D) prevents conversion of testosterone to dihydrotestosterone, and 17-ketosteroid reductase deficiency (E) prevents conversion of androstenedione to testosterone. In contrast, 11β-hydroxylase deficiency (answer C) is associated with increased secretion of androgen precursors by the fetal adrenal gland, which causes virilization of affected 46,XX female fetuses but no genital abnormality in males.

2. An 8-year-old boy has bilateral inguinal hernias and bilateral cryptorchidism. The penis and scrotum are normal. At operation, a uterus and fallopian tube are noted in the left inguinal canal, following which the surgeon terminates the operation and pages you, the on-call endocrinologist. What gene is *most likely* affected in this patient?

A. AMHR2 (anti-müllerian hormone receptor)
B. CYP21A2 (steroid 21-hydroxylase)
C. SRY (sex-determining region Y)
D. SF1 (steroidogenic factor 1)
E. LHGCR (luteinizing hormone receptor)

Answer: A This boy has persistent müllerian duct syndrome, which can be caused by mutations in either anti-müllerian hormone or (equally likely) in its receptor. Persistence of müllerian structures (uterus and fallopian tubes) in this otherwise normal boy prevents descent of the testes. CYP21A2 mutations (answer B) do not cause any abnormalities of reproductive structures in 46,XY males. Affected 46,XX females might have virilization of the external genitalia that could be severe enough to be mistaken for males and would indeed have a normal uterus and fallopian tubes. However, such individuals almost invariably develop adrenal insufficiency shortly after birth, and all newborns in the United States are screened for this disorder. Moreover, the müllerian structures in females with CYP21A2 mutations are found in their normal positions because there are no testes and spermatic ducts to drag them into the inguinal canals. Mutations of SRY (answer C), SF1 (D), and LHGCR (E) are found in 46,XY phenotypic females.

3. A 15-year-old girl presents for evaluation of primary amenorrhea. She has Tanner 5 breasts and an adult female distribution of pubic and axillary hair. There are no inguinal or labial masses. An ultrasound examination shows ovaries in the normal position but no uterus. The patient's karyotype and diagnosis are *most likely* to be which of the following?

A. 46,XY; complete androgen insensitivity syndrome
B. 45,X/46,XX mosaic; Turner syndrome
C. 46,XY; 17,20-lyase deficiency
D. 46,XX; Mayer-Rokitansky-Küster-Hauser syndrome
E. 46,XX; Kallman's syndrome

Answer: D Patients with Mayer-Rokitansky-Küster-Hauser syndrome have normal ovaries and therefore have normal female secondary sexual characteristics. By definition, they have atresia of the uterus and upper vagina. This phenotypic female has axillary and pubic hair, meaning that she responds normally to androgens secreted by the adrenal glands and (if present) by the ovaries. This rules out complete androgen insensitivity syndrome (answer A); 46,XY males with that condition have absent or sparse axillary and pubic hair. Moreover, they have testes that are often palpable in the inguinal canals or labia. Girls with Turner syndrome (answer B) often have pubic and axillary hair (owing to adrenal androgen secretion) and may have some degree of breast development, owing to local aromatization of adrenal androgens in breast fat and (if mosaic) to some residual ovarian function. But the uterus is normally formed. 46,XY males with 17,20 lyase deficiency (answer C) cannot synthesize androgens and have a phenotype similar to patients with complete androgen insensitivity syndrome. Patients with Kallman's syndrome (answer E) have hypogonadotropic hypogonadism; affected females indeed have primary amenorrhea and may or may not have development of secondary sexual characteristics depending on degree of severity, but they have normal development of the uterus.

4. A 13-year-old boy is brought to your office by his parents because he insists that he is really a girl. He has had this feeling as long as he can remember, but it is causing him increasing distress now that he is in early adolescence. He has insisted on wearing dresses or gender-neutral clothing whenever possible. On examination, he is at the 50th percentile for height and weight for boys for his age. He has downy hair on his upper lip, axillary hair, a normally formed penis with the urethral meatus at the tip of the glans, bilaterally descended 4-mL testes, Tanner stage 2 pubic hair, and 1 cm of firm glandular tissue under each nipple. What would be the single most informative evaluation at this point?

A. Psychology consultation
B. Karyotype
C. Serum levels of dehydroepiandrosterone, androstenedione, and testosterone
D. FSH and LH
E. Estradiol

Answer: A Any patient with gender dysphoria requires psychological evaluation, preferably in the context of evaluation by an experienced multidisciplinary team, before considering any medical management. Transgendered patients rarely have identifiable medical etiologies for their condition unless the history and physical examination raise suspicion for a DSD. This phenotypic boy has palpable testes, strongly suggesting that he is 46,XY (answer B), and a normally formed penis, strongly suggesting intact ability to synthesize androgens (answer C). He has age-appropriate development of secondary sexual characteristics, suggesting that his ability to secrete gonadotropins is intact (answer D). Mild gynecomastia is typical in boys in early to mid-adolescence (owing to aromatization of androgens in breast tissue, in the context of low testosterone levels relative to adults) and is without clinical significance; estradiol levels will not be elevated.

5. You are paged to the intensive care nursery to evaluate a 2300-g term infant for ambiguous genitalia. The baby developed respiratory distress at delivery and is now intubated on a respirator. You are told that a cleft palate was noted on intubation. You hear a grade 3/6 harsh holosystolic murmur. The baby has accessory digits bilaterally. The penis is small and the urethral meatus is at the base of the shaft. There is a bifid scrotum; gonads are palpable bilaterally in the inguinal canals. What laboratory tests are *most likely* to establish the diagnosis?

A. 17-Hydroxyprogesterone and androstenedione
B. Testosterone and dihydrotestosterone
C. Androstenedione and testosterone
D. 11-Deoxycortisol and cortisol
E. 7-Dehydrocholesterol and cholesterol

Answer: E This male infant has Smith-Lemli-Opitz syndrome, owing to an inability to synthesize cholesterol from its immediate precursor. He is small for gestational age and, as a term infant, must have serious cardiorespiratory dysfunction (such as lung hypoplasia) to require ventilator support. The character of the murmur suggests a ventriculoseptal defect. The other physical findings are all typical of this syndrome. Boys with 5α-reductase deficiency are unable to synthesize dihydrotestosterone from testosterone (answer B); they may have female-appearing or ambiguous genitalia, but have no other medical problems. Boys with 17-ketosteroid reductase deficiency cannot synthesize testosterone from androstenedione (answer C); the phenotype is identical to 5α-reductase deficiency. Patients with 11-hydroxylase deficiency cannot synthesize cortisol from 11-deoxycortisol (answer D). Males are phenotypically normal; females may have ambiguous genitalia but no other congenital anomalies.

234

THE TESTIS AND MALE HYPOGONADISM, INFERTILITY, AND SEXUAL DYSFUNCTION

RONALD S. SWERDLOFF AND CHRISTINA WANG

PHYSIOLOGY

The testis is a bifunctional organ serving as the site of sex steroid (i.e., testosterone) synthesis and sperm production in the male. Androgens and their metabolites (including estrogens) also act on nonreproductive organs and serve essential roles in muscles, adipose tissues, bones, metabolism, and brain functions.

The male reproductive axis consists of six main components: (1) extrahypothalamic central nervous system (CNS), (2) hypothalamus, (3) pituitary, (4) testes, (5) sex steroid–sensitive end organs, and (6) sites of androgen transport and metabolism (Fig. 234-1). The components of this system function in an integrated fashion to control the concentrations of circulating gonadal steroids required for normal male sexual development

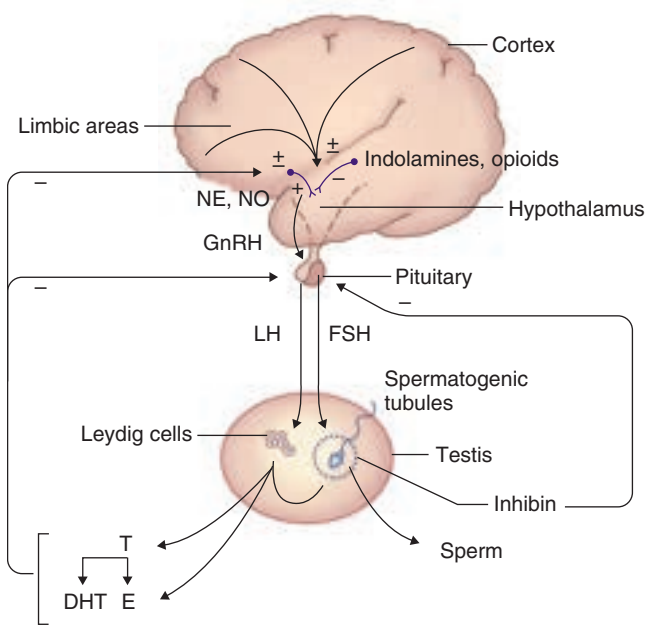

FIGURE 234-1. The hypothalamic-pituitary-gonadal axis in the male. DHT = dihydrotestosterone; E = estrogen; FSH = follicle-stimulating hormone; GnRH = gonadotropin-releasing hormone; LH = luteinizing hormone; NE = norepinephrine; NO = nitric oxide; T = testosterone.

and function and for androgen- and estrogen-mediated metabolic effects on critical end organs. The reproductive axis is also responsible for normal germ cell development and maturation. Accessory sexual organs, including the epididymides, seminal vesicles, and prostate gland, are important for sperm maturation (epididymis) and seminal fluid production. An anatomically functional sperm transport and ejaculatory system are necessary to ensure male fertility.

Hypothalamic-Pituitary Function

The hypothalamus is responsible for the normal pulsatile secretion of gonadotropin-releasing hormone (GnRH) (Chapter 224). The pulsatile release of GnRH provides the signals for the timing of the release of luteinizing hormone (LH) and follicle-stimulating hormone (FSH), which occurs every 60 to 90 minutes in men. The secretion of GnRH is regulated by neuronal input from higher cognitive and sensory centers and by circulating levels of sex steroids and peptide hormones such as prolactin and leptin (Chapter 223). Testosterone or its metabolic products (i.e., estradiol and dihydrotestosterone [DHT]) inhibit the secretion and release of GnRH, LH, and FSH. Prolactin is also a potent inhibitor of GnRH secretion.

Both LH and FSH consist of two subunits (α and β). Both subunits are required for biologic activity; the subunits can be detected in serum and may be increased in certain pathologic conditions (e.g., α-subunit elevations in gonadotropin-secreting pituitary adenomas). LH has a shorter half-life than FSH. Feedback regulation of LH and FSH secretion also occurs at the pituitary, with testosterone, DHT, and estrogens inhibiting the synthesis or release of both gonadotropins. The circulating testicular peptide inhibin, produced by Sertoli cells, also selectively inhibits FSH. LH and FSH act on the testes through specific cell surface receptors on the Leydig and Sertoli cells, respectively.

Testosterone

The testis is a complex organ consisting of (1) seminiferous tubules containing Sertoli cells and germ cells and (2) the interstitium, which contains the steroid-secreting (Leydig) cells. Leydig cells synthesize steroid hormones under the regulation of LH. The LH receptors on the cell surface of the Leydig cells lead to G protein–, adenyl cyclase–, and cyclic adenosine monophosphate–mediated activation of steroid biosynthesis.

Testosterone is the principal male hormone secreted by the testes; approximately 5 to 10 mg/day is produced in adult men. Testosterone synthesis occurs in the human testes through either the Δ^4 or the predominant Δ^5 pathway. The enzymatic rate-limiting steps in the process are the LH-inducible steroid acute regulatory (StAR) protein and the conversion of cholesterol to pregnenolone by the cholesterol side-chain cleavage enzyme P450SCC (Fig. 234-2).

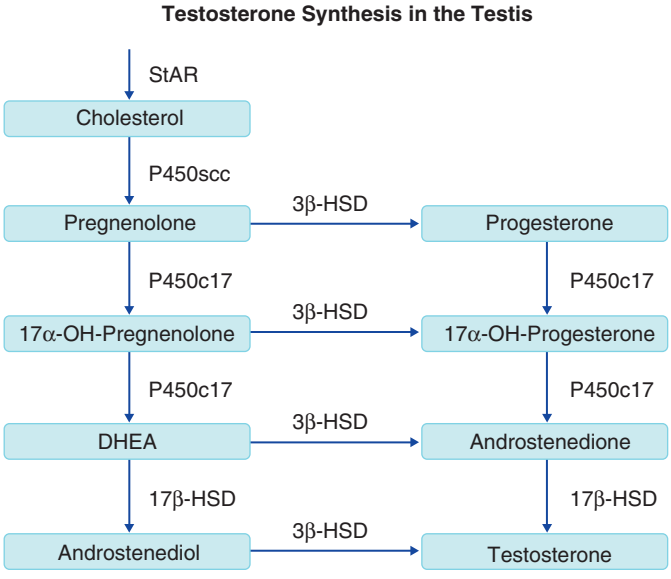

FIGURE 234-2. The steroid acute regulatory (*StAR*) protein mobilizes cholesterol from cellular stores to the mitochondria. Intratesticular steroidogenic pathways for the synthesis of testosterone. Although both the Δ^5 (*left*) and Δ^4 (*right*) pathways exist, the Δ^5 pathway predominates in the testis. DHEA = dehydroepiandrosterone; HSD = hydroxysteroid dehydrogenase.

Testosterone circulates mainly bound to two plasma proteins: sex hormone–binding globulin (SHBG) and albumin. In young adult men, approximately 54% of testosterone is bound to albumin, 44% is bound to SHBG, and 2 to 3% is unbound. Bioavailable testosterone refers to the sum of albumin-bound and free testosterone and is measured by separating SHBG-bound testosterone from total testosterone. Serum SHBG levels are increased in hyperestrogenemic states, hyperthyroidism, aging, phenytoin treatment, anorexia nervosa, and prolonged stress. SHBG levels are lowered in hyperandrogenic states (endogenous and exogenous as in androgen treatment), obesity, acromegaly, and hypothyroidism. In most instances, measurement of serum total testosterone provides biochemical support for the diagnosis of androgen deficiency. In conditions with abnormal SHBG levels, however, the total testosterone measurement may be misleading, and measurement of non–SHBG-bound testosterone may allow better interpretation of the active testosterone levels. This can be done by direct measurement of free testosterone by the equilibrium dialysis method, measurement of bioavailable testosterone, or calculation of the free testosterone by a formula requiring the serum testosterone and SHBG concentrations. Most guidelines recommend against measurement of free testosterone by a "direct" or analogue displacement method because of lack of accuracy traceable to a standard.

Testosterone exerts its effects either through direct action or after conversion to DHT by two separate 5α-reductase isozymes (1 and 2) or to estradiol by the aromatase enzyme (Fig. 234-3). Thus, testosterone can act directly on the androgen receptor or as a precursor for DHT, which also binds efficiently to the androgen receptor. Different tissues have coactivators or coinhibitors that modify the action of the androgen-receptor complex, providing tissue selectivity and amplification. Testosterone also can serve as a precursor for estradiol in some tissues, and after conversion, estrogen binds the estrogen receptors (α or β) to induce its effects. Various end organs differ in their 5α-reductase isoenzyme and aromatase concentrations and/or activity. Congenital and acquired defects in these two enzymes, as well as in the estrogen and androgen receptors, result in distinct syndromes with characteristic phenotypes that are experiments in nature and provide understanding of the actions of specific receptors and enzyme activities (Chapter 233).

Spermatogenesis

The spermatogenic compartment of the testis consists of the Sertoli and germ cells that are intimately interactive with the interstitial compartment. The Sertoli cells bridge the entire space between the basement membrane and the lumen of the tubules. They are the target of androgenic and FSH stimulation of spermatogenesis and also the source of a multitude of paracrine regulators of spermatogenesis (e.g., inhibin, activin, growth factors, cytokines).

Germ cell development and maturation depend on the proper hormonal (FSH) and paracrine (testosterone) milieu. Both testosterone and FSH stimulate progression of spermatogonia to mature spermatozoa, limit the amount of germ cell death (apoptosis), and regulate sperm release from the germinal epithelium.

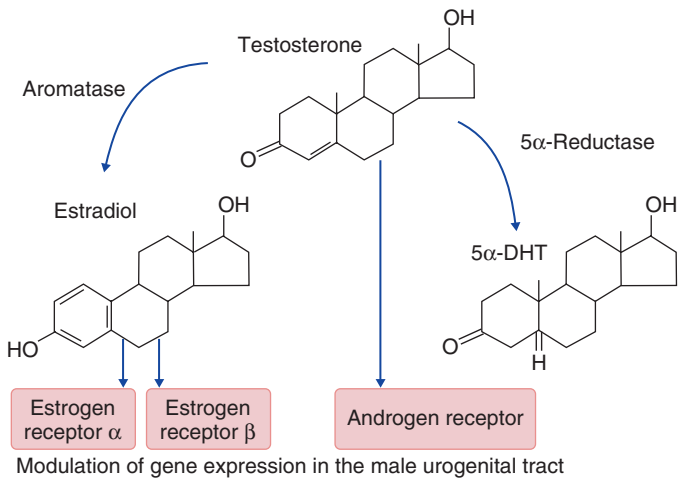

FIGURE 234-3. Testosterone action is mediated directly (androgen receptor), after conversion to estradiol (estrogen receptor α or β), or after conversion to dihydrotestosterone (DHT; androgen receptor). (From Kuiper GCJM, Carlquist M, Gustafsson JA. Estrogen is a male and female hormone. *Sci Med*. 1998;5:36-45.)

After spermatogenesis is completed, mature spermatozoa are released into the excretory system and travel through the rete testes and epididymis, where they become functionally mature and acquire fertilizing capacity before traversing the vas deferens. The seminal fluid gains constituents from the seminal vesicles, prostate, and bulbourethral glands before ejaculation.

Sexual Function and Erectile Physiology

Sexual function in men requires normal sexual desire (libido) and erectile, ejaculatory, and orgasmic capacity. The process is complex, involving cognitive, sensory, hormonal, autonomic neuronal, and penile vascular integrative actions for normal function. Defects can occur at multiple levels.

The brain is the integrative center of the sexual response system. It processes sensory input and hormonal signals to create the hypothalamic neuronal message that traverses the spinal cord to the T9-12 sympathetic and sacral parasympathetic outflow tracts. The nonadrenergic, noncholinergic autonomic plexus nerves initiate vasodilation of the cavernosal arterial and corpora cavernosal sinusoids of the penis through the release of local vasodilators (e.g., nitric oxide and vasoactive intestinal peptide) from the vascular endothelium and the sinusoidal smooth muscle cells of the sinusoids (Fig. 234-4). Nitric oxide produces smooth muscle dilation by the generation of cyclic guanosine monophosphate (cGMP) and the modification of calcium flux. The neurogenic mechanisms leading to vasodilation of the cavernosal arterioles and sinusoids lead to a rapid increase in penile blood flow and expansion of the vascular channels; this, in turn, inhibits venous return through compression of the venous channels against the tunica albuginea and limits venous drainage.

Testosterone's primary effect on erectile function is to enhance libido. Testosterone also increases penile nitric oxide synthase activity and enhances smooth muscle cell growth. Sexual desire and fantasy are highly sensitive to testosterone, explaining the preservation of erectile capacity in many men with partial androgen deficiency.[1]

Physiology in Development and Aging

Reproductive Axis Development during Childhood and Puberty

Adrenarche and Puberty

Adrenarche occurs at approximately 7 or 8 years of age when the zona reticularis of the adrenal gland undergoes maturation, leading to increased secretion of androstenedione, dehydroepiandrosterone (DHEA), and DHEA sulfate (DHEA-S). The process is under the control of adrenocorticotropic hormone, not LH or FSH. Androstenedione and DHEA are technically androgenic prehormones, and the prepubertal growth spurt, as well as the early development of pubic and axillary hair, are mediated to a great extent by the conversion of these precursors to testosterone and DHT in peripheral tissues.

Initiation of puberty is determined by an increase in the pulsatile pattern of hypothalamic GnRH secretion. This is marked by nocturnal bursts of LH secretion. As puberty progresses, feedback sensitivity of the hypothalamus and pituitary to circulating steroids lessens, thus increasing the secretion of gonadotropins. The increasing concentrations of intratesticular testosterone and circulating FSH stimulate the Sertoli cell to produce factors leading to the maturation of spermatogenesis. The majority of the extratesticular end-organ events of puberty are secondary to the increased testosterone and its metabolic products (DHT and estradiol) (Table 234-1). The penis and scrotum grow and become pigmented. As spermatogenesis advances, the testes increase in size from 1 to 2 mL at the outset of puberty to 15 to 35 mL in adulthood. There is a progressive increase in facial, axillary, chest, abdominal, thigh, and pubic hair; frontal scalp hair regresses, and the voice deepens (Fig. 234-5). Genital and sexual hair development and temporal scalp hair regression require DHT. The increased levels of sex steroids result in closure of the epiphysis and achievement of adult height.

Aberrations of Timing of Puberty

Delayed puberty, more common in boys than in girls, is usually defined as a temporary form of hypothalamic hypogonadotropic hypogonadism in which sexual development has not begun by age 13.5 years. Height age (the age

TABLE 234-1 PUBERTAL STAGES IN BOYS

STAGE	PUBIC HAIR	GENITAL
1	Absence of pubic hair	Childlike penis, testes, and scrotum (testes 2 mL)
2	Sparse, lightly pigmented hair mainly at base of penis	Scrotum enlarged with early rugation and pigmentation; testes begin to enlarge (3-5 mL)
3	Hair becomes coarse, darker, more curled, and more extensive	Penis has grown in length and diameter; testes now 8-10 mL; scrotum more rugated
4	Hair adult in quality, but distribution does not include medial aspect of thighs	Penis further enlarged, with development of glans; scrotum and testes (10-13 mL) further enlarged
5	Hair is adult and extends to thighs	Penis and scrotum fully adult; testes ≥ 15 mL

Modified from Marshall WA, Tanner JM. Variation in pattern of pubertal changes in boys. *Arch Dis Child.* 1970;45:13-23.

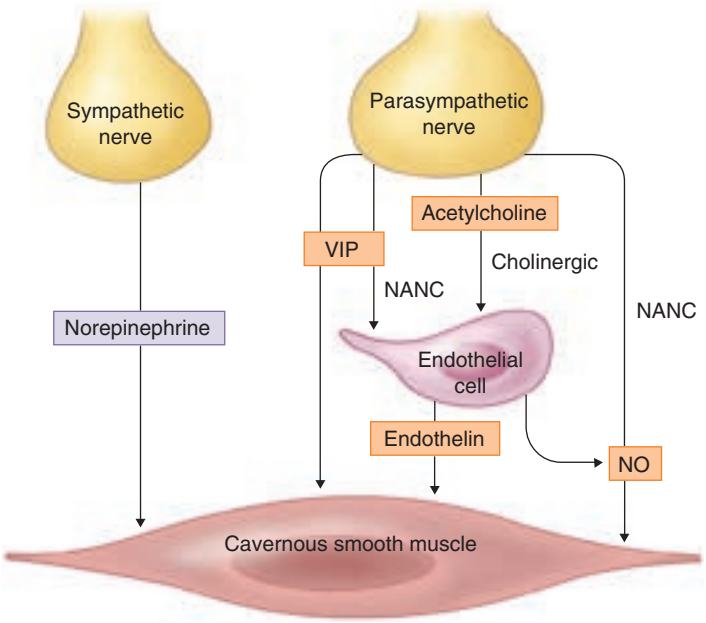

FIGURE 234-4. The interaction among cholinergic, adrenergic, and nonadrenergic, noncholinergic (*NANC*) neuronal pathways and their contribution to penile smooth muscle contraction and dilation (*arrows*). NO = nitric oxide; VIP = vasoactive intestinal polypeptide. (From Lue TF. Physiology of penile erection and pathophysiology of erectile dysfunction and priapism. In: Walsh P, Retick A, Vaughn E, Wein A, eds. *Campbell's Urology,* 7th ed. Philadelphia: WB Saunders; 1998:1164.)

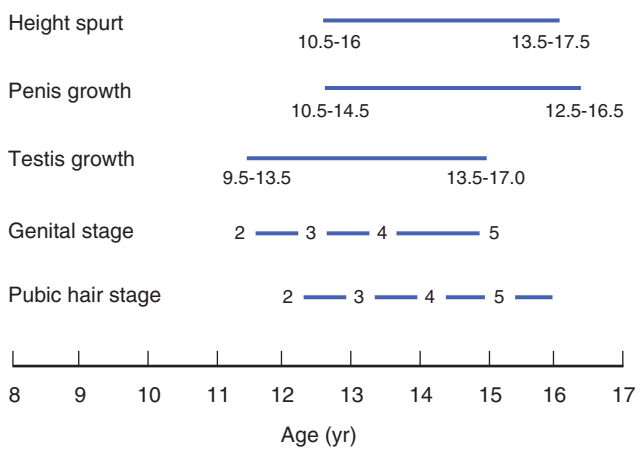

FIGURE 234-5. Diagram of the timing of the various components of puberty. The range of ages at which each parameter begins and is completed is shown for each bar. These data were obtained from European children 40 years ago. Since then, there may be a slight trend for an earlier onset of puberty. (From Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child*. 1970;45:13-23.)

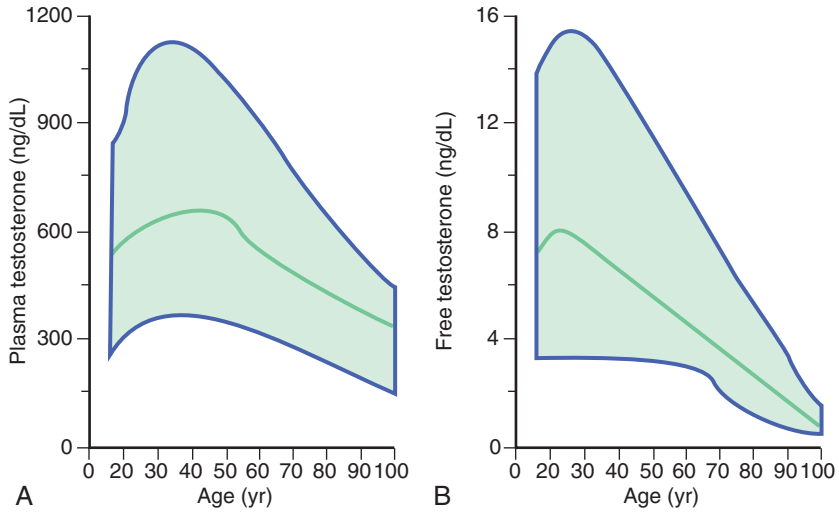

FIGURE 234-6. Relationship between plasma testosterone (A) and free testosterone (B) levels and age in normal males. (From Baker HWG, Berger HG, DeKretser DM, et al. Changes in the pituitary-testicular system with age. *Clin Endocrinol*. 1996;5:349-372.)

representative of 50% of normal children at the patient's height) is delayed with respect to chronologic age, but it is concordant with bone age. Once initiated, puberty is normally completed within 4.5 years. The majority of boys with delayed development eventually attain full sexual adulthood. There is often a family history of a parent or sibling being a "late bloomer." The *GPR54* gene encodes a kisspeptin-responsive G protein–coupled receptor whose absence results in hypogonadotropic hypogonadism resulting from impaired secretion of GnRH. Careful documentation of changing physical findings and measurement of serum LH, FSH, and testosterone concentrations may provide valuable clues to the beginning of puberty. An increase in testicular size to more than 3 mL usually heralds other signs of pubertal onset. Inquiring and testing for hyposmia or anosmia and other midline defects may indicate a common variant of congenital hypogonadotropic hypogonadism (Kallmann's syndrome). The decision to institute early treatment depends on the perceived degree of psychological stress associated with the maturational delay. The major concern is early fusion of the epiphyses induced by treatment with testosterone, which compromises optimal height; however, with proper dosing and monitoring of bone age, this is unusual. In adolescent boys with delayed puberty and low levels of gonadotropins, periodic withdrawal of treatment is used to determine whether spontaneous puberty has occurred. Many adult men diagnosed with and treated as adolescents for a presumed diagnosis of hypogonadotropic hypogonadism achieve normal reproductive function when they discontinue therapy.

Precocious puberty in boys is defined as the onset of pubertal (genital and secondary sexual) development before 9 years of age. Sexual precocity can be subcategorized as true (complete and incomplete) isosexual precocious puberty and pseudo–precocious puberty. The distinction is that true precocious puberty is associated with increases in GnRH-stimulated LH and FSH secretion (hypothalamic-pituitary origin), whereas pseudo–precocious puberty is independent of GnRH stimulation of LH and FSH secretion. True precocious puberty in boys is often associated with CNS disease (two thirds of boys), including hypothalamic tumors, cysts, inflammatory conditions, and seizure disorders. Diagnostic findings include sexual precocity, inappropriately elevated serum LH levels, and associated elevations of testosterone. Magnetic resonance imaging can localize most lesions. Another cause of central precocious puberty is human chorionic gonadotropin secretory germinomas (testicular, hepatic, hypothalamic, or pineal tumors). Pseudo–precocious puberty is characterized by increased testosterone with suppressed LH. Causes of pseudo–precocious puberty include congenital virilizing adrenal hyperplasia, testicular testosterone-secreting neoplasms, and constitutively active LH receptor mutations; the latter condition results in uncontrolled testosterone secretion (testotoxicosis). Treatment of true precocious puberty is removal or correction (with surgery or radiation therapy) of the CNS lesion, if possible, and treatment with GnRH analogues to temporarily suppress LH and FSH secretion. Treatment of pseudo–precocious puberty depends on the cause but includes glucocorticoids for congenital virilizing adrenal hyperplasia and ketoconazole (to suppress steroidogenesis), with or without antiandrogens (e.g., spironolactone, flutamide).

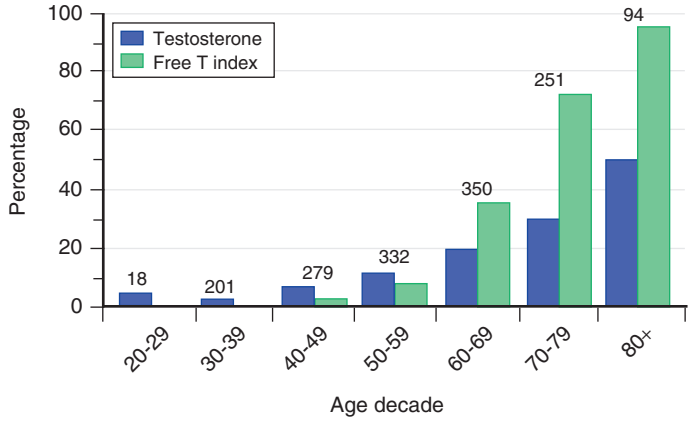

FIGURE 234-7. Hypogonadism in aging men. Bar height indicates the percentage of men in each 10-year interval, from the third to the ninth decades, with at least one testosterone value in the hypogonadal range. The criteria used for these determinations are total testosterone less than 11.3 nmol/L (325 ng/dL) and testosterone and sex hormone–binding globulin (free T index) less than 0.153 nmol/nmol. The numbers above each pair of bars indicate the number of men studied in the corresponding decade. The fraction of men who are hypogonadal increases progressively after age 50 years by either criterion. More men are hypogonadal by free T index than by total testosterone after 50 years, and there seems to be a progressively greater difference, with increasing age, between the two criteria.

Male Senescence: Decreased Testosterone and Other Anabolic Hormones

Testosterone Deficiency in the Elderly

Blood concentrations of testosterone, other anabolic hormones (e.g., growth hormone), and prehormones (e.g., DHEA, DHEA-S) are significantly lower in older men than in young adult men. Both total and bioavailable or free serum testosterone levels progressively decrease with aging (Fig. 234-6). The percentage decline in serum testosterone has been estimated as 1 to 2% per year. Serum SHBG levels also rise with age in men, resulting in a higher percentage of circulating testosterone that is tightly bound and less bioavailable. Between 20 and 80% of men older than 70 years have blood levels of bioavailable or free testosterone below the normal range for young adults (Fig. 234-7). Many older men with serum testosterone below the young adult reference range do not complain of symptoms attributable to low testosterone. In men between 30 and 70 years of age the crude prevalence rate of symptomatic testosterone deficiency has been estimated to be approximately 6%. Low testosterone levels are associated with comorbidities such as obesity, metabolic syndrome, and the commonest symptoms associated with low total or free testosterone in older men are sexual symptoms.[2]

Many of the effects of low testosterone levels in aging men are similar to those observed in younger hypogonadal men. These include decreases in libido, erectile function, muscle mass, muscle strength, bone mass, and impaired mood and sense of well-being. Older men have increased body fat, particularly visceral fat. The benefits of testosterone treatment for symptomatic older men with low serum testosterone levels remain controversial. Recommendations on the diagnosis, investigation, and treatment of late-onset hypogonadism are available but lack evidence from large-scale randomized, controlled national or international intervention studies. A randomized trial of testosterone versus placebo gel application in elderly men with a high prevalence of chronic disease was discontinued early because of an increased risk for cardiovascular events in those receiving testosterone, but the small size of the trial and the unique population prevented broader inferences about the safety of the testosterone therapy.[2] In a more recent study of men in the Veterans Administration health system who underwent coronary angiography and had low serum testosterone levels, the use of the testosterone therapy was found to be associated with increased risk for adverse outcomes.[3] Beneficial effects of testosterone replacement have been demonstrated in elderly men with relatively low serum testosterone levels. Testosterone replacement therapy (up to 3 years) decreases fat mass, increases lean body mass, improves strength, and increases bone mineral density in older men.[2] In larger, more recent studies of treatment with testosterone improved muscle strength and physical function were found in frail elderly men.[2,4] Erectile dysfunction in older men is usually multifactorial (see later), with impaired penile vasodilatory function a predominant factor in many cases. Thus testosterone replacement therapy in older men may enhance libido but often does not improve erectile dysfunction. At present, testosterone treatment is not recommended for men with or suspected of having prostate cancer, moderate-to-severe heart failure, severe and uncorrected sleep apnea, or high red blood cell mass. A digital rectal examination should be performed, prostate-specific antigen level determined, and symptoms of severe urinary tract obstruction evaluated before testosterone treatment is instituted.[5]

Deficiency of Adrenal Androgen in Older Men

A marked decline in the circulating levels of adrenal androgens, especially DHEA and DHEA-S, has been recognized in elderly men and women (Chapter 234). Serum levels of DHEA and DHEA-S peak at approximately the third decade of life and then decline at about 2% per year, resulting in levels 10 to 20% of baseline by 80 years of age. DHEA is a precursor to androgens such as testosterone and DHT. Studies have shown that the oral administration of 50 mg of DHEA to older men raises serum DHEA and DHEA-S concentrations to the levels found in young men without changing serum levels of testosterone, with no beneficial effects on quality of life, sexual function, mood, body composition, or exercise capacity. In the United States, DHEA is available without prescription as a health supplement and is widely used, making large-scale multicenter, prospective, placebo-controlled trials difficult to perform.

MALE HYPOGONADISM

DEFINITION

Hypogonadism (androgen deficiency) is diagnosed in men with consistent symptoms and signs and unequivocally low circulating levels of testosterone.[6] Most men with more severe androgen deficiency have very low intratesticular testosterone concentrations and are infertile. Primary hypogonadism indicates that the abnormality originates in the testis; it is characterized by increased serum LH and FSH levels. Secondary hypogonadism indicates a defect at the hypothalamus or pituitary, resulting in decreased gonadotropins (LH, FSH, or both). Combined primary and secondary hypogonadism occurs in aging and in a number of systemic diseases, such as alcoholism, liver disease, metabolic syndrome, type 2 diabetes mellitus, human immunodeficiency virus (HIV) infection, hemochromatosis, and sickle cell disease. Obesity leads to low total and free testosterone levels. Greater decreases are seen in the total testosterone level because obesity not only decreases testosterone secretion but also lowers SHBG levels. Decreased androgen action with normal or elevated testosterone levels, mimicking androgen deficiency, may occur in patients with androgen receptor defects (androgen resistance), postreceptor signaling abnormalities, and inability to convert testosterone to the active metabolite DHT (5α-reductase abnormalities).

Many of the causes of primary and secondary hypogonadism are listed in Tables 234-2 and 234-3 (see also Chapter 233).

TABLE 234-2 CAUSES OF PRIMARY TESTICULAR FAILURE AND END-ORGAN RESISTANCE

Congenital disorders
Chromosome disorders
Klinefelter's and related syndromes (e.g., XXY, XXY/XY, XYY, XX males)
Testosterone biosynthetic enzyme defects
Myotonic dystrophy
Developmental disorders
Prenatal diethylstilbestrol syndrome
Cryptorchidism
Acquired defects
Orchitis
Mumps and other viruses
Granulomatous disease (e.g., tuberculosis, leprosy)
Human immunodeficiency virus infection
Infiltrative disease (e.g., hemochromatosis, amyloidosis)
Surgical, traumatic injuries, torsion of testis
Irradiation
Toxins (e.g., alcohol, fungicides, insecticides, heavy metals, cottonseed oil, DDT, other environmental "endocrine disruptors")
Drugs
Cytotoxic agents
Inhibitors of testosterone synthesis and antiandrogens (e.g., ketoconazole, cimetidine, flutamide, cyproterone, spironolactone)
Ethanol, opioids, other recreational drugs
Autoimmune testicular failure
Isolated
Associated with other organ-specific disorders (e.g., Addison's disease, Hashimoto's thyroiditis, insulin-dependent diabetes)
Androgen resistance syndromes
5α-Reductase deficiency
Systemic diseases* (e.g., cirrhosis, chronic renal failure, sickle cell disease, acquired immunodeficiency syndrome, amyloidosis)
Aging*

*Systemic diseases and aging produce a mixed pattern of testicular and hypothalamic-pituitary dysfunction.

TABLE 234-3 CAUSES OF HYPOGONADOTROPIC HYPOGONADISM

IDIOPATHIC OR CONGENITAL
Isolated deficiency of gonadotropin-releasing hormone
With anosmia (Kallmann's syndrome)
With other abnormalities (Prader-Willi syndrome, Laurence-Moon-Biedl syndrome, basal encephalocele)
Partial deficiency of gonadotropin-releasing hormone (fertile eunuch syndrome)
Multiple hypothalamic and pituitary hormone deficiency
Pituitary hypoplasia or aplasia
ACQUIRED
Traumatic brain injury, after surgery or irradiation
Neoplastic
Pituitary adenoma (prolactinoma, other functional and nonfunctional tumors)
Craniopharyngioma, germinoma, glioma, leukemia, lymphoma
Pituitary infarction, carotid aneurysm
Infiltrative and infectious diseases of hypothalamus and pituitary (sarcoidosis, tuberculosis, coccidioidomycosis, histoplasmosis, syphilis, abscess, histiocytosis X, hemochromatosis)
Autoimmune hypophysitis
Aging and systemic diseases*
Obesity
Malnutrition
Anorexia nervosa, starvation, renal failure, liver failure
Exogenous hormones and drugs
Antiandrogens, estrogens and antiestrogens, progestogens, glucocorticoids, cimetidine, spironolactone, digoxin, drug-induced hyperprolactinemia (metoclopramide, tranquilizers, antihypertensives)

*Aging and systemic diseases produce a mixed pattern of central and testicular dysfunction.

CLINICAL MANIFESTATIONS

History

The medical history should focus on testicular descent, pubertal development, shaving frequency, changes in body hair, and present and past systemic illnesses. A complete sexual history includes changes in libido, erectile and

ejaculatory functions, frequency of masturbation, coital activity, and fertility (including that of present and previous partners). Information should be obtained on previous orchitis, sinopulmonary complaints, sexually transmitted diseases, human immunodeficiency virus (HIV) status, genitourinary infections, and previous surgical procedures that might affect the reproductive tract (e.g., vasectomy, hernia repair, prostatectomy, varicocele ligation). Social history includes tobacco and alcohol intake. Medication and self-prescribed drug history includes recreational drugs; opioids; anabolic steroids; 5α-reductase inhibitors; and psychiatric, antihypertensive, antiandrogenic, cytotoxic, and alternative medicine therapies; environmental toxins; and exposure to heat (including saunas and Jacuzzis) and irradiation.

Physical Examination

The general physical examination is supplemented by height and span measurements; characterization of facial, pubic, and body hair distribution; presence of acne and facial wrinkling; breast examination for gynecomastia; assessment of muscle mass and adiposity; measurement of penile length and urethral meatus localization; digital rectal prostate examination; and visual field assessment if secondary hypogonadism is suspected. The scrotal examination should include an assessment of midline fusion (e.g., bifid scrotum, hypospadias); testicular size and consistency; presence of intratesticular masses; abnormalities of the epididymis; bilateral presence of vas deferens; and varicoceles, hydroceles, or hernias. Normal testicular size ranges from 3.6 to 5.5 cm in length, 2.1 to 3.2 cm in width, and 15 to 35 mL in volume in white and black men. Asian men have a slightly smaller mean testicular size. A decrease in testicular volume usually implies decreased spermatogenic cells because the seminiferous tubules account for more than 80% of testicular volume.

Laboratory Studies

Because there is a strong diurnal rhythm in testosterone secretion in young men (highest in the morning), testosterone, LH, and FSH are routinely determined from morning blood samples. There is a broad range of reference values of these hormones due partly to measurement variability but also influenced by the selection criteria for the reference population. Most hospital laboratories used immunoassay methods to measure serum testosterone, which may lack precision at low serum testosterone levels. Recent studies suggest that methods using liquid or gas chromatography and mass spectrometry may give more accurate results even at very low serum testosterone levels. Elevated LH and FSH levels distinguish primary from secondary hypogonadism (both have low serum testosterone levels), but many older men with low serum testosterone levels have normal LH concentrations. Serum prolactin levels should be measured in all low testosterone, low LH cases (hypogonadotropic hypogonadism) and in men with known pituitary mass lesions, or galactorrhea. DHT is measured in cases of abnormal differentiation of the genitalia and when 5α-reductase deficiency is suspected. Serum estradiol should be measured in cases of gynecomastia. Assessment of other testosterone precursors and products may be required in special circumstances, including suspected congenital enzyme defects. The semen analysis is the "cornerstone" of the laboratory examination for male infertility.

Primary Testicular Hypogonadism

Primary hypogonadism refers to a condition of androgen deficiency with or without infertility in which the pathologic process lies at the testis level. A list of common causes is given in Table 234-2.

CONGENITAL DEFECTS

The commonest congenital defect is due to chromosomal abnormalities (Klinefelter's syndrome), and other causes are listed in Table 234-2 and described in Chapter 233.

ACQUIRED DEFECTS

Mumps, Orchitis, Leprosy, Human Immunodeficiency Virus Infection, and Hemochromatosis

After puberty, mumps (Chapter 369) is associated with clinical orchitis in 25% of cases, and 60% of those affected become infertile. During acute orchitis, the testes are inflamed, painful, and swollen. This is followed by a gradual decrease in size. The testes may return to normal size and function, or they may atrophy. Spermatogenic defects occur more often and earlier than Leydig cell dysfunction. Thus, patients with postorchitic infertility may have normal testosterone and LH levels with increased serum FSH levels. Over time, elevations in LH and lower serum testosterone levels may appear. Leprosy (Chapter 326) also may cause orchitis and gonadal insufficiency. HIV infection is often associated with hypogonadism, which can be either hypogonadotropic or hypergonadotropic. Hemochromatosis (Chapter 212) may affect the hypothalamus-pituitary, as well as act directly on the testis.

Trauma

The exposed position of the testes in the scrotum makes them particularly susceptible to injury. Surgical injury during scrotal surgery for hernia, varicocele, and vasectomy can result in permanent testicular damage.

Irradiation

Exposure of the testes to irradiation in the treatment of malignant diseases produces testicular germ cell, and less commonly, Leydig cell damage.

DRUGS AND TOXINS

Chemotherapy, in particular alkylating agents such as cyclophosphamide, busulfan, frequently leads to irreversible germ cell damage. Heavy metals (lead, cadmium) and cottonseed oil (gossypol) cause damage to the germ cells. Leydig cells are relatively less susceptible to most chemotherapeutic drugs than are Sertoli and germ cells. Some medications may interfere with testosterone biosynthesis (e.g., ketoconazole, spironolactone) or action (e.g., cyproterone, flutamide). Ethanol, independent of its role in causing liver disease, inhibits testosterone biosynthesis. Marijuana, heroin, methadone, medroxyprogesterone acetate, other progestins, and estrogens lower testosterone, mainly by decreasing LH. Medical treatment with androgens such as testosterone, DHT, and synthetic anabolic steroids or their illicit use (e.g., in athletes, bodybuilders) lowers serum LH and FSH and sperm counts. Serum testosterone levels are low after the use of DHT and synthetic anabolic agents. Environmental toxins such as fungicides and insecticides (e.g., DBCP, metabolites of DDT, vinclozolin) and byproducts of the plastics industry (e.g., phthalates, bisphenol A) are called "endocrine disruptors" because these chemicals may have either weak estrogenic or antiandrogenic effects and have been shown to cause testicular dysgenesis in male offspring when administered in large doses to pregnant female rodents. Data linking "endocrine disruptors" to male reproductive dysfunction in humans are principally associations studies and do not prove causality.[7]

AUTOIMMUNE TESTICULAR FAILURE

Antibodies against the microsomal fraction of the Leydig cells may occur either as an isolated disorder or as part of a multiglandular disorder (Chapter 231) involving, to variable degrees, the thyroid, pituitary, adrenals, pancreas, and other organs.

ANDROGEN RESISTANCE (ANDROGEN-SENSITIVE END-ORGAN DEFICIENCY)

Certain conditions have clinical phenotypes mimicking testosterone deficiency in the absence of lowered testosterone levels. These androgen-resistant states may be drug-induced (antiandrogens) or genetic sensitive defects in the androgen receptor, congenital or acquired post–androgen receptor signaling defects, or 5α-reductase deficiency (Chapter 233).

Hypogonadism Associated with Systemic Diseases

Abnormalities of the hypothalamic-pituitary-testicular axis occur in a number of systemic diseases, including liver failure, renal failure, severe malnutrition, sickle cell anemia, advanced malignant disease, severe obesity, metabolic syndrome, type 2 diabetes, cystic fibrosis, and amyloidosis, as well as in those on chronic hemodialysis. The effects of cirrhosis of the liver on testicular function are complex and may be either independent of or associated with the direct toxic effects of the continued use of alcohol. Gynecomastia, testicular atrophy, and impotence are concomitant signs of cirrhosis. Decreased spermatogenesis with peritubular fibrosis occurs in 50% of cases. Estradiol levels are usually elevated. This results in an increased ratio of serum estradiol to testosterone, often associated with gynecomastia. In sickle cell anemia and thalassemia major, boys may have impaired sexual maturation, and men are often infertile. Diabetes and obesity are two major factors in hypogonadism. Emerging data show that type 2 diabetes is associated with low blood testosterone levels mainly as a result of hypothalamic-pituitary dysfunction; the decrease in serum testosterone correlates with the degree of hyperglycemia.

Secondary Gonadal Insufficiency (Hypogonadotropic Hypogonadism)

CONGENITAL HYPOGONADOTROPIC HYPOGONADISM

Hypogonadotropic hypogonadism represents a deficiency in the secretion of gonadotropins (LH and FSH) because of an intrinsic or functional abnormality in the hypothalamus or pituitary glands (see earlier and Chapter 233). Such disorders result in secondary Leydig cell dysfunction (see Table 234-3). The clinical manifestations depend on the age of the patient at the onset of the disorder.

ACQUIRED HYPOGONADOTROPIC DISORDERS AND FUNCTIONAL DISORDERS

Anorexia Nervosa and Weight Loss

Anorexia nervosa (Chapter 219) and weight loss are examples of functional defects resulting in low serum testosterone levels. Men and women with anorexia nervosa present with manifestations of hypogonadotropic hypogonadism. Starvation also may reduce gonadotropic secretion. Strenuous exercise has minimal effects on testicular function in men.

Stress and Illness

Severe stress (e.g., surgery, trauma) and systemic illness also lower gonadotropin and testosterone levels. Organic hypothalamic-pituitary disorders include neoplastic, granulomatous, infiltrative, and post-traumatic lesions in the region of the hypothalamus and pituitary.

Pituitary Tumors

Prolactinomas manifest differently in men than in women (Chapter 224). In men, these tumors are usually large (>1 cm in diameter; macroadenomas) by the time they are detected. Male patients with prolactin-secreting macroadenomas usually present with hypogonadism, erectile dysfunction, and visual manifestations from suprasellar extension. In small tumors, hypogonadotropic hypogonadism may be due to suppressive effects on GnRH described earlier, but in large tumors, it also may be due to a mass effect damaging the non-neoplastic gonadotrophs.

Large non–prolactin-secreting pituitary tumors (growth hormone, adrenocorticotropic hormone, glycopeptide, and null cell) also may produce gonadotropin insufficiency from damage to the adjacent normal pituitary gland (Chapter 224), resulting in decreased serum LH and testosterone levels.

DIAGNOSIS

The diagnosis is based on clinical symptoms and signs and a reduced serum testosterone level. The normal range of serum total testosterone in a young adult male population varies across different laboratories but should be in the general range of 300 to 1000 ng/dL (10-35 nmol/L). Accurate measurements of testosterone in the severely hypogonadal range are best done by gas or liquid chromatography followed by tandem mass spectrometry. Total testosterone measurements may be misleading indicators of Leydig cell secretory status in conditions in which SHBG levels are abnormal (see earlier section). In these circumstances, a measurement of free testosterone (by an equilibrium dialysis method), bioavailable testosterone (consisting of free plus albumin bound), or calculated free testosterone (by total testosterone and SHBG measurements) may be useful.

The following rules on measurement of serum testosterone apply to most young and middle-aged men thought to have hypogonadism. If a morning serum total testosterone level is repeatedly below 230 ng/dL (8 nmol/L), and he has symptoms or signs compatible with low testosterone state, the patient is probably hypogonadal, and testosterone replacement is indicated. If the serum testosterone level is between 230 and 320 ng/dL with normal serum LH levels, the patient may or may not be clinically hypogonadal and androgen replacement may not improve the symptoms (e.g., sexual dysfunction). Thus, when serum total testosterone is borderline and LH is not increased, one of the measurements of bioactive testosterone is indicated (e.g., free testosterone). The guidelines for men older than 60 years are less certain. Because SHBG levels are often increased, total testosterone levels may overestimate the biologically active forms. A serum total testosterone level above 350 to 400 ng/dL indicates that hypogonadism is very unlikely to be the cause of the symptoms, and the clinician should look for other etiologies for the symptoms.

TREATMENT

Rx

The main medical indication for androgen replacement therapy is male hypogonadism (Table 234-4). In approximately 10% of men with idiopathic hypogonadism reversed by testosterone therapy, the reversal is sustained after therapy is stopped. This suggests that some patients with low serum testosterone may have a transient cause of the deficiency. Administration of testosterone to elderly men with low-normal testosterone concentrations increases lean body mass and decreases fat mass. There are insufficient data to judge whether testosterone treatment will improve functional status or cognition. Carefully designed studies of the efficacy of testosterone treatment in older men are ongoing.

Absolute contraindications to androgen replacement therapy include carcinoma of the prostate and the male breast. Androgens should be used with caution in older men with an enlarged prostate and urinary symptoms, elevated hematocrit, and sleep-related breathing disorders. The various methods of delivering testosterone treatment are shown in Table 234-5.

TABLE 234-4 INDICATIONS FOR ANDROGEN THERAPY

- Androgen deficiency (hypogonadism)
- Microphallus (neonatal)
- Delayed puberty in boys
- Elderly men with low total or bioavailable or free testosterone levels and symptoms
- Angioneurotic edema
- Other possible uses or under investigation
  - Hormonal male contraception
  - Sarcopenia associated with cancer, human immunodeficiency virus infection, chronic infection, frailty in older men and women
  - Hypoactive sexual disorder in postmenopausal women

TABLE 234-5 ANDROGEN PREPARATIONS

ROUTE	PREPARATION	DOSE AND FREQUENCY OF ADMINISTRATION
Oral*	Testosterone undecenoate (not available in United States; available in Canada, Mexico, Europe, Asia)	40-80 mg PO two or three times daily
Buccal	Transbuccal testosterone, mucoadhesive tablets (Striant)	30 mg two times daily
Injection	Testosterone enanthate and cypionate	100 mg/wk IM or 150-200 mg IM every 2-3 wk
	Testosterone undecanoate (not available in United States)	750-1000 mg IM every 10-12 wk
Implant	Testosterone implants	75-mg pellets (in United States), 6-10 inserted once every 4-6 months
Transdermal	Scrotal patch	One patch delivering testosterone 4 or 6/day
	Nonscrotal patch Androderm	Two patches, each delivering testosterone 2.5 mg/day; or one patch delivering testosterone 5 mg/day
	Testoderm TTS	One patch delivering testosterone 5 mg/day
Transdermal gels	AndroGel or Testogel; Testim; Axiron; Fortesta	1 to 2% gel applied once daily delivering 50-100 mg testosterone on skin and 5 to 10 mg to body

*Oral modified 17α-alkylated androgens such as methyltestosterone, fluoxymesterone, oxymetholone, stanozolol, and oxandrolone are not recommended for the treatment of androgen-deficient states because of potential hepatotoxicity and adverse effects on serum lipids.
IM = intramuscularly; PO = orally.

Testosterone esters, such as testosterone enanthate (or cypionate) injections, are widely used in the United States and throughout the world. The recommended dose is 150 to 200 mg administered intramuscularly once every 2 to 3 weeks. Testosterone undecenoate injections administered every 10 to 12 weeks are available in many parts of the world but not yet in the United States.

Modified 17α-alkylated androgens (methyltestosterone and many anabolic steroids), which are available in oral preparations, are not recommended as androgen replacement. These agents may lead to abnormalities in liver function, marked decreases in high-density lipoprotein cholesterol, and increases in total cholesterol levels compared with the testosterone esters. Oral testosterone undecanoate capsules have been available for over 20 years in many parts of the world but not in the United States. Transbuccal delivery of testosterone by mucoadhesive tablets (30 mg applied twice daily) results in physiologic-range testosterone levels through direct absorption into the systemic circulation, thus avoiding first-pass effects on the liver. The tablets may be dislodged from the buccal mucous membrane. Other oral formulations are in clinical trials in the United States.

Implants (pellets) of crystalline testosterone are available for chronic treatment of hypogonadism. Serum testosterone levels are maintained in the physiologic range for 4 to 6 months. Implants are not usually used but are gaining some popularity with urologists in the United States; they are widely used in Australia and the United Kingdom.

Transdermal testosterone delivery through skin patches and gels have been available in the United States for over 15 years. The nonscrotal patches deliver 5 mg/day of testosterone, which is the physiologic production rate. These patches deliver levels of testosterone within the normal range but have a high incidence of skin irritability (redness, swelling, and blisters). Hydroalcoholic and nonalcoholic testosterone gels have been developed for transdermal application and have become the most widely used testosterone formulations in the United States. The usual dosage is 50 to 100 mg of 1, 1.62, and 2% testosterone gel applied daily to the skin, delivering 5 to 10 mg of testosterone to the body. This results in a more consistent serum concentration and causes little skin irritation. Transfer from the user to others is possible during routine use and may be a concern if there is close skin contact with women and children. Protective clothing or a shower is necessary to avoid transferring testosterone through skin-to-skin contact.

Table 234-6 shows the benefits and potential side effects of androgen treatment. In hypogonadal men, androgen replacement leads to the development and maintenance of secondary sexual characteristics. Testosterone has important anabolic effects on muscle and bone and improves libido and sexual dysfunction. It has less effect on erectile dysfunction. It has no major short-term effects on prostate tissue.[8] Epidemiologic studies indicate that lower testosterone is a risk factor for cardiovascular disease. Many small studies have shown benefit or no effect of cardiovascular disease but more recent studies suggest testosterone replacement may increase the risk for cardiovascular events in elderly men, especially those who are frail and have multiple comorbidties.[9]

TABLE 234-6 ANDROGEN THERAPY: RISKS VERSUS BENEFITS

BENEFITS	RISKS
Development or maintenance of secondary sex characteristics	Fluid retention
Improved libido and sexual function	Gynecomastia
Increased muscle mass and strength	Acne, oily skin
Increased bone mineral density	Increased hematocrit
Decreased body and visceral fat	Decreased high-density lipoprotein cholesterol (oral 17α-alkylated agents produce the greatest effect)
Improved mood	Sleep apnea
Effect on cognition (?)	Aggressive behavior (?)
Effect on vitality and quality of life (?)	Prostate disease
Decreased cardiovascular disease risk (epidemiologic studies); clinical study no benefits/risk	Benign prostatic hyperplasia (?)
	Carcinoma of prostate (aggravate existing cancer)
	Increased cardiovascular adverse events in one study in frail elderly men with multiple comorbid conditions

MALE INFERTILITY

DEFINITION

Infertility is defined as the failure of a couple to achieve pregnancy after at least 1 year of frequent unprotected intercourse. If a pregnancy has not occurred after 3 years, infertility will most likely persist without medical treatment.

EPIDEMIOLOGY

Studies in the United States and Europe showed a 1-year prevalence of infertility in 15% of couples. The prevalence in developing countries is likely to be higher because of the higher prevalence of genital tract infection. Of subfertility cases, 30 to 35% can be attributed to predominantly female factors, 25 to 30% to male factors, and 25 to 30% to problems in both partners.

PATHOBIOLOGY

Hypothalamic-pituitary disorders are infrequent causes of male infertility and are discussed in the section on hypogonadism and androgen deficiency. Testicular disorders are the most frequent identifiable cause of infertility (see Table 234-2). Y chromosome microdeletions are increasingly recognized as a genetic cause of azoospermia and severe oligozoospermia. Up to 25% of infertile men have microdeletions in the long arm of the Y chromosome, many of which map to the Yq11 region of the chromosome, which is called the azoospermic factor (AZF). Mutations in the AZF a and b regions are associated with azoospermia, whereas mutations of AFZc region may be associated with oligozoospermia.[10]

DIAGNOSIS

The approach to the diagnosis of an infertile couple includes management of both the male and the female partner (Figs. 234-8 and 234-9).

Examination of the ejaculate is the cornerstone for the investigation of an infertile man (Table 234-7). Semen samples are collected at the physician's office or at home, preferably after 2 to 7 days of abstinence from ejaculatory activity. The generally accepted reference values for a semen analysis are given in Table 234-8. A normal sperm concentration is greater than 15 million/mL, with a total sperm number greater than 39 million per ejaculate; however, men with lower sperm counts can be fertile. More than 40% of the spermatozoa should be motile, and more than 32% should demonstrate a progressive motility pattern. Using strict criteria to assess sperm morphology, the percentage of morphologically normal forms should be above 4%. There is considerable overlap in the semen quality of fertile and subfertile men. Low sperm concentration and/or poor sperm morphology are associated with lower chances of natural conception in the female partner. In patients with abnormal semen analyses, measurements of serum FSH, LH, and testosterone are indicated (see Fig. 234-7). Elevated FSH levels usually indicate severe germinal epithelium damage. A decreased serum inhibin B level also reflects poor Sertoli cell function and may indicate spermatogenic dysfunction. Elevated serum LH and FSH concentrations together with a low serum testosterone level indicate pantesticular failure leading to hypogonadism and infertility. Low serum FSH, LH, and testosterone concentrations suggest hypothalamic-pituitary dysfunction; serum prolactin should be measured, and additional investigations may be required. A low sperm concentration and suppressed LH level with an increased, normal, or low serum testosterone level (without clinical manifestations of androgen deficiency) may suggest exogenous androgen use. The hormonal pattern in androgen insensitivity (an uncommon cause of male infertility) is elevated LH, normal FSH, and high-normal to increased serum testosterone levels. Normal hormonal parameters in azoospermic (no sperm in the ejaculate) men with normal-sized testes may suggest congenital or acquired obstruction in the epididymis or vas deferens.

TREATMENT

Rx

An algorithmic approach to the treatment of male infertility is illustrated in Figures 234-8 and 234-9. The principles of managing male infertility can be summarized as follows. (1) Men with mild-to-moderate oligozoospermia, with or without decreased sperm motility and some impairment of motility, are

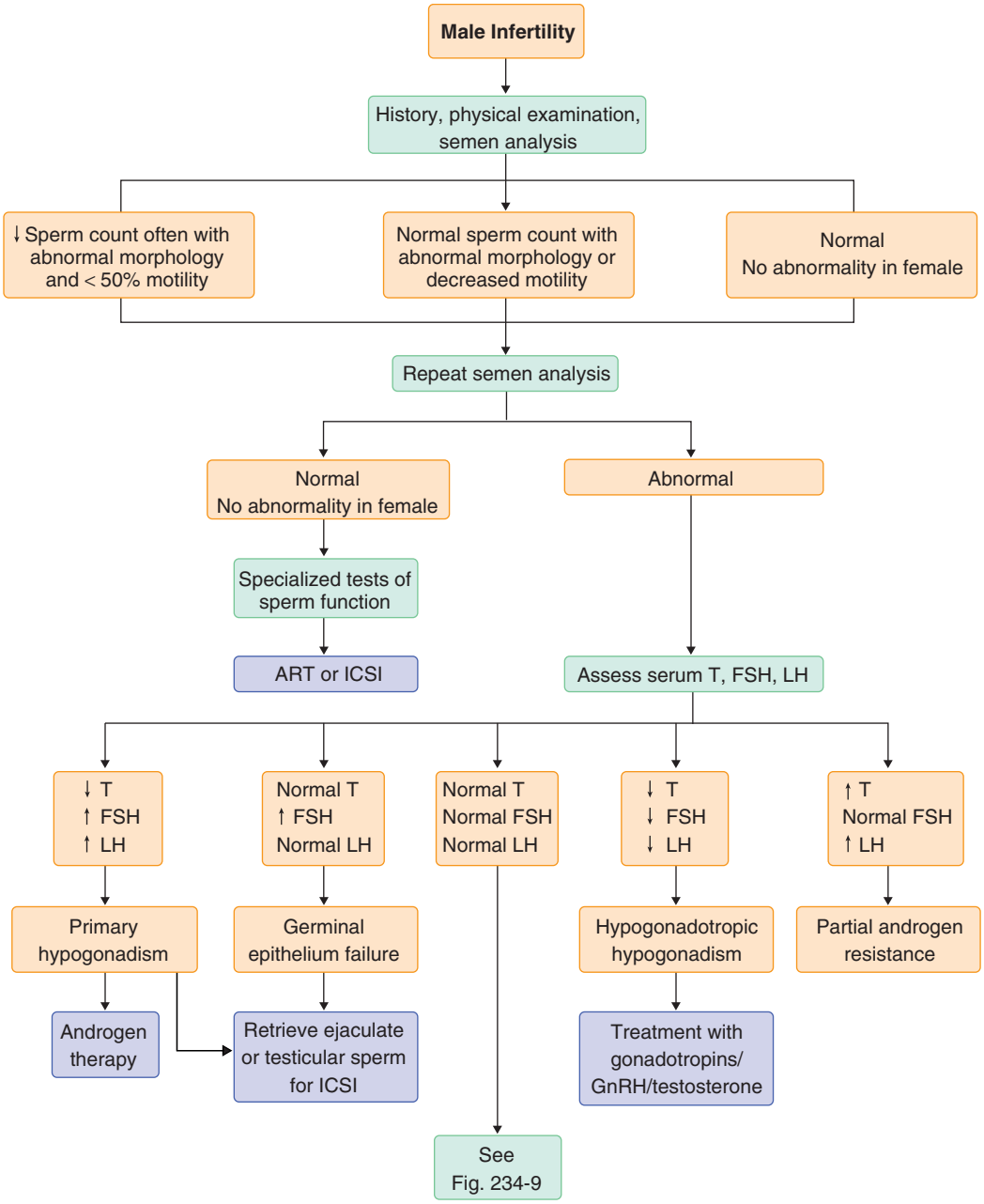

FIGURE 234-8. Algorithmic approach to the diagnosis and treatment of male infertility. ART = assisted reproductive technology; FSH = follicle-stimulating hormone; GnRH = gonadotropin-releasing hormone; ICSI = intracytoplasmic sperm injection; LH = serum luteinizing hormone; T = serum testosterone.

subfertile rather than infertile. Spontaneous pregnancies can occur in this group. (2) Reliable pharmaceutical treatment is limited to the 1 to 2% of infertile men with gonadotropin insufficiency. (3) Assisted reproductive technologies, including in vitro fertilization and intracytoplasmic sperm injection, have dramatically improved pregnancy rates. (4) In male factor infertility azoospermia (absence of sperm in the ejaculate) may occur in men with obstruction of the ejaculatory system. In these patients, in vitro fertilization and intracytoplasmic sperm injection after either percutaneous epididymal sperm extraction or microsurgical epididymal sperm extraction are highly successful. (5) Azoospermia resulting from impaired spermatogenesis may not be a sterile state, because sperm may be present within the testes. These sperm can be extracted, and intracytoplasmic sperm injection can be performed with good success, even in patients with Klinefelter's syndrome.

SEXUAL DYSFUNCTION

Sexual dysfunction can be divided into four main categories: (1) loss of desire (libido), (2) erectile dysfunction, (3) ejaculatory insufficiency, and (4) anorgasmic states.

Decreased Libido

Loss of libido refers to a reduction in sexual interest, initiative, and frequency and intensity of responses to internal or external erotic stimuli. Causal factors include psychogenic factors, CNS disease, androgen deficiency and resistance, and side effects from medications (e.g., antihypertensives, psychotropics, alcohol, narcotics, dopamine blockers, antiandrogens, and possibly 5α-reductase inhibitors). Treatment is directed toward the causal mechanism.

Ejaculatory Failure and Impaired Orgasm

Ejaculatory insufficiency refers to absent or reduced seminal emission or impaired ejaculatory contraction. It is usually associated with neurologic conditions and medication therapy. An anorgasmic state is a distressing but relatively uncommon condition in men in which the normal process of erection and ejaculation occurs in the absence of the subjective sensation of pleasure initiated at the time of emission and ejaculation. Premature ejaculation is the most common form of male sexual dysfunction. Estimates of prevalence vary, but 25 to 30% seems to be a reasonable estimate. The recently published *Diagnostic and Statistical Manual of Mental Disorders-5 (2013)* defines premature ejaculation as ejaculation occurring within approximately 1 minute of vaginal penetration before the person wishes it on 75% of occasions for at least 6 months and causing personal distress. The

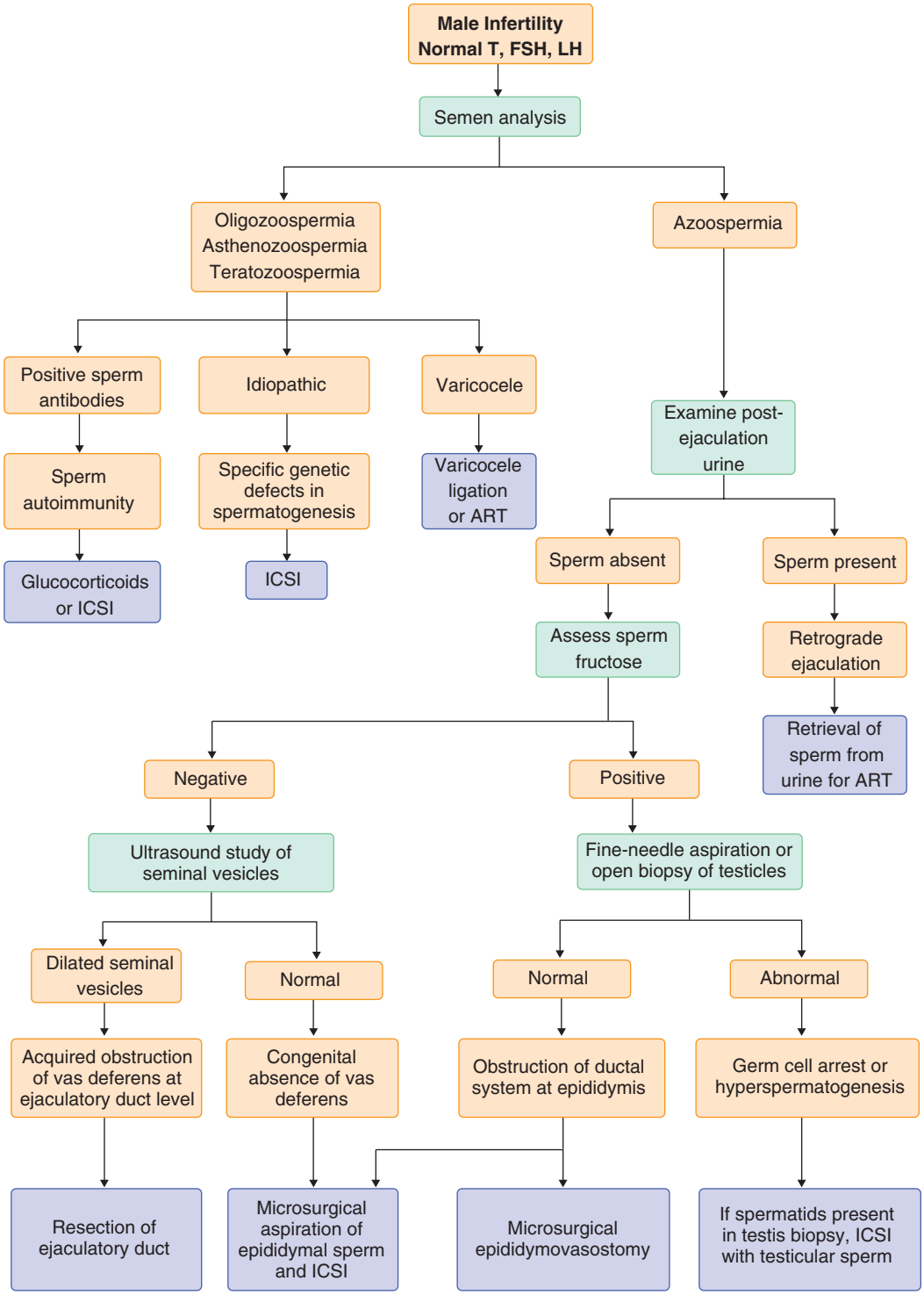

FIGURE 234-9. Algorithmic approach to the diagnosis and treatment of male infertility in patients with normal serum hormone concentrations. ART = assisted reproductive technology; FSH = follicle-stimulating hormone; ICSI = intracytoplasmic sperm injection; LH = serum luteinizing hormone; T = serum testosterone.

pathobiology of premature ejaculation is unknown. It may be associated with marked distress or interpersonal difficulty and is not a direct effect of substance abuse such as opiate withdrawal. The diagnosis is based mainly on sexual history and includes assessment of intravaginal ejaculatory latency time, perceived control, distress, and interpersonal difficulty. The first-line treatment is with selective serotonin reuptake inhibitors or a serotonin transporter inhibitor (e.g., dapoxetine 60 mg as on-demand therapy) together with behavioral therapy and relationship counseling. Topical anesthetic creams can be used as alternatives.[11]

Erectile Dysfunction

DEFINITION

Erectile dysfunction can be defined as a man's inability to obtain rigidity sufficient to permit coitus of adequate duration to satisfy himself and his partner.

EPIDEMIOLOGY

Current estimates suggest that 10 to 15% of all American men suffer from erectile dysfunction, with the incidence progressively increased as men become older. Data from the Massachusetts Aging Study report that 52% of men 40 to 70 years of age experience some degree of erectile dysfunction. The prevalence of erectile dysfunction is even higher in men with type 2 diabetes mellitus and after radical prostatectomy for prostate cancer. Recent epidemiologic studies in the United States and Europe in men between 50 to 80 years indicate that erectile dysfunction is associated with lower urinary tract obstructive symptoms/benign prostatic hyperplasia.

PATHOBIOLOGY

The causes of erectile dysfunction are many, but they can generally be categorized as follows: vasculogenic, psychological, endocrine, neurologic,

TABLE 234-7 MALE INFERTILITY: BASIC LABORATORY TESTS

SEMEN ANALYSES	HORMONE ANALYSES (IN PATIENTS WITH ABNORMAL SEMEN ANALYSES)
Volume, pH	Serum luteinizing hormone and follicle-stimulating hormone
Microscopy: Agglutination, debris	Serum testosterone
Sperm: Concentration, motility, morphology, vitality	If luteinizing hormone and testosterone levels are low, serum prolactin
Leukocytes	
Immature germ cells	
Sperm autoantibodies (sperm and semen biochemistry, sperm function tests)	

TABLE 234-8 SEMEN ANALYSIS: REFERENCE RANGE FROM FERTILE MEN*

PARAMETER	REFERENCE RANGE
Semen volume	>1.5 mL
Sperm	
Concentration	>15 million/mL
Total count	>39 million/ejaculate
Motility	>40% motile
	>32% progressively motile
Morphology	>4% normal†
Vitality (live)	>58%
Leukocytes	<1 million/mL

*Men whose partners had a time-to-pregnancy of ≤12 months were chosen to provide reference distributions for semen parameters.
†This value is based on the strict criteria for assessing sperm morphology in studies using in vitro fertilization as an end point.

iatrogenic (post–radical prostatectomy), drug related, systemic illness, and aging. Erectile dysfunction is common in older men, despite normal serum testosterone levels; this effect appears to be the result of impaired penile vasodilatory capacity as a result of endothelial dysfunction. Decreased nonadrenergic, noncholinergic nerve activity and reduced production of nitric oxide by endothelial cells result in decreased cavernous smooth muscle relaxation, decreased filling of the cavernous sinusoids, and reduced compression of the venous plexus against the tunica lead to failure of erection. Men presenting with erectile dysfunction share common risk factors with cardiovascular disease (smoking, obesity, metabolic syndrome, hyperlipidemia, and type 2 diabetes mellitus).[12] Recent evidence indicates that men presenting with mild erectile dysfunction should be assessed for cardiovascular disease, particularly when other risk factors are present.

DIAGNOSIS

The diagnosis of erectile dysfunction is based mainly on a detailed medical and sexual history of the patient and his partner when available. The history may reveal the underlying cause or other common disorders associated with erectile dysfunction. Physical examination should focus on genitourinary, cardiovascular, endocrine, and neurologic systems. Prostate examination is important because erectile dysfunction is commonly associated with symptomatic benign prostatic hyperplasia. Laboratory tests should include a morning serum testosterone, and, if indicated, prostate-specific antigen, fasting glucose (or hemoglobin A_{1C}), and cholesterol. Specific diagnostic tests are rarely required.

TREATMENT

Rx

The treatment of erectile dysfunction is to find the cause and treat the cause if found.[13] Symptoms can be effectively treated by the oral administration of penile-selective phosphodiesterase-5 inhibitors (sildenafil, vardenafil, tadalafil). A treatment algorithm for erectile dysfunction is given in Fig. 234-10. Lifestyle interventions reduce obesity and improve erectile function.[14] Combined androgen deficiency with decreased libido and decreased penile responsiveness resulting from impaired nitric oxide synthase activity may be common in elderly men. With the availability of effective penile vasodilatory medications to ensure erectile capacity, complaints of diminished libido can be effectively treated with androgen supplementation.

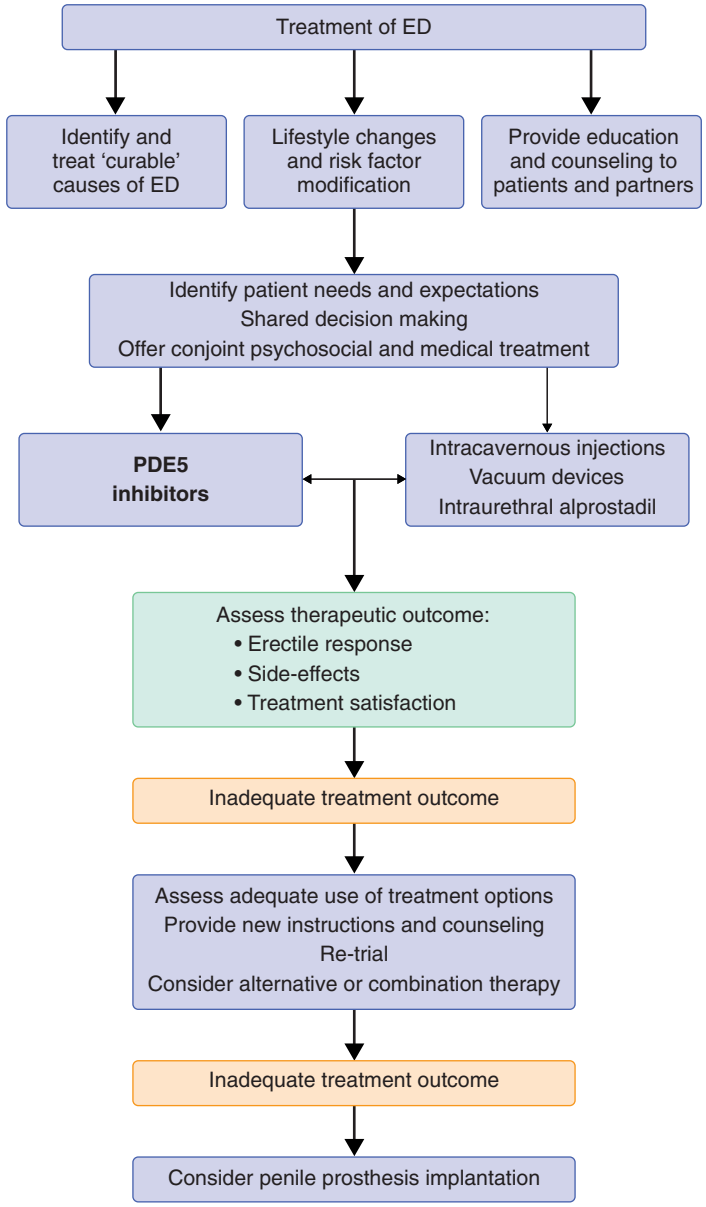

FIGURE 234-10. Treatment algorithm for erectile dysfunction (*ED*). PDE_5 = cyclic GMP phosphodiesterase-5. (Reprinted with permission from Hatzimouratidis K, Amar E, Eardley I, et al. Guidelines on male sexual dysfunction: erectile dysfunction and premature ejaculation. *Eur Urol.* 2010;57:804-814. 2015 update: http://uroweb.org/guideline/male-sexual-dysfunction/. Accessed March 26, 2015.)

Medical Therapy

Oral Medications

Oral and selective inhibitors of cGMP phosphodiesterase-5 (the primary phosphodiesterase in the penile cavernosal tissue) are effective for at least 60% of men. Inhibition of phosphodiesterase-5 causes persistence of normally (sexually) stimulated cGMP in the corpora cavernosa, resulting in protracted cavernosal tumescence and rigidity. Patients with diabetes mellitus, spinal cord injuries, prostatic surgery, and pelvic irradiation also benefit, but with a somewhat lower response rate. The usual starting dose of sildenafil is 50 mg, increasing in 25-mg increments up to 100 mg. Because of its mechanism of action, sildenafil is used on demand with recommended administration 20 to 60 minutes before intercourse. Two other potent phosphodiesterase-5 inhibitors (vardenafil and tadalafil) are widely used for the treatment of erectile dysfunction and appear to be equally effective. Vardenafil (5, 10, and 20 mg) has a relatively longer duration of action (4 to 6 hours), and tadalafil (10 or 20 mg) has an even longer duration of action (up to 36 hours). Randomized controlled trials showed that daily administration of tadalafil (5 mg) improved erectile function compared to on-demand treatment. Daily dosing of tadalafil is well tolerated and effective. Thus, daily dosing starting with 2.5 mg of tadalafil may

be an alternative to on-demand administration if intercourse is expected to be more frequent—for example, more than twice per week. Hypogonadal men with erectile dysfunction and low libido may benefit from combined treatment with testosterone and phosphodiesterase-5 inhibitors (PDE-5). However, the addition of testosterone to sildenafil does not further improve erectile dysfunction.[A6] PDE-5 inhibitors should not be administered with nitrates because the accumulation of cGMP may result in lowering of blood pressure and hypotension. PDE-5 inhibitors also may interact with antihypertensive agents, including α-blockers, resulting in orthostatic hypotension.

Intracavernosal Injection

Second-line treatment of erectile dysfunction involves intracavernosal injection with vasodilators such as prostaglandin E_1 (Alprostadil) alone or with other vasodilators (papaverine, phentolamine) These medications are injected into the cavernosal space with a 27- to 30-gauge needle and may be useful in men who are refractory to oral agents. The main side effects of penile injections are pain and cavernosal fibrosis which usually resolves after discontinuation of injections. Presence of tunica fibrosis may suggest early Peyronie's disease,[15] and injections should be stopped. The intraurethral prostaglandin E_1 suppository alprostadil is believed to work locally on the corpora cavernosa as a vasodilatory agent. The suppository is apparently successful in improving erectile function in 30 to 66% of cases.

Penile Prostheses

Surgical implantation of penile prostheses that include inflatable and malleable devices are the third line of treatment for men who prefer a permanent solution of their problem or for those who do not respond to other therapies.

Grade A References

A1. Basaria S, Coviello AD, Travison TG, et al. Adverse events associated with testosterone administration. *N Engl J Med*. 2010;363:109-122.

A2. Corona G, Rastrelli G, Giagulli VA, et al. Dehydroepiandrosterone supplementation in elderly men: a meta-analysis study of placebo-controlled trials. *J Clin Endocrinol Metab*. 2013;98:3615-3626.

A3. Srinivas-Shankar U, Roberts SA, Connolly MJ, et al. Effects of testosterone on muscle strength, physical function, body composition, and quality of life in intermediate-frail and frail elderly men: a randomized, double-blind, placebo-controlled study. *J Clin Endocrinol Metab*. 2010;95:639-650.

A4. De Hong C, Ren LL, Yu H, et al. The role of dapoxetine hydrochloride on-demand for the treatment of men with premature ejaculation. *Sci Rep*. 2014;4:7269.

A5. Yuan J, Zhang R, Yang Z, et al. Comparative effectiveness and safety of oral phosphodiesterase type 5 inhibitors for erectile dysfunction: a systematic review and network meta-analysis. *Eur Urol*. 2013;63:902-912.

A6. Spitzer M, Basaria S, Travison TG, et al. Effect of testosterone replacement on response to sildenafil citrate in men with erectile dysfunction: a parallel, randomized trial. *Ann Intern Med*. 2012;157:681-691.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Isidori AM, Buvat J, Corona G, et al. A critical analysis of the role of testosterone in erectile function: from pathophysiology to treatment-a systematic review. *Eur Urol*. 2014;65:99-112.
2. Wu FC, Tajar A, Beynon JM, et al. Identification of late-onset hypogonadism in middle-aged and elderly men. *N Engl J Med*. 2010;363:123-135.
3. Vigen R, O'Donnell CI, Baron AE, et al. Association of testosterone therapy with mortality, myocardial infarction, and stroke in men with low testosterone levels. *JAMA*. 2013;310:1829-1836.
4. Travison TG, Basaria S, Storer TW, et al. Clinical meaningfulness of the changes in muscle performance and physical function associated with testosterone administration in older men with mobility limitation. *J Gerontol A Biol Sci Med Sci*. 2011;66:1090-1099.
5. Surampudi P, Swerdloff RS, Wang C. An update on male hypogonadism therapy. *Expert Opin Pharmacother*. 2014;15:1247-1264.
6. Bhasin S, Cunningham GR, Hayes FJ, et al. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2010;95:2536-2559.
7. Knez J. Endocrine-disrupting chemicals and male reproductive health. *Reprod Biomed Online*. 2013;26:440-448.
8. Fernandez-Balsells MM, Murad MH, Lane M, et al. Clinical review 1: adverse effects of testosterone therapy in adult men: a systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2010;95:2560-2575.
9. Xu L, Freeman G, Cowling BJ, et al. Testosterone therapy and cardiovascular events among men: a systematic review and meta-analysis of placebo-controlled randomized trials. *BMC Med*. 2013;11:108.
10. Hotalin JM. Genetics of male infertility. *Urol Clin North Am*. 2014;41:1-17, 10.
11. Althof SE, McMahon CG, Waldinger MD, et al. An update of the International Society of Sexual Medicine's guidelines for the diagnosis and treatment of premature ejaculation (PE). *J Sex Med*. 2014;11:1392-1422.
12. Dong JY, Zhang YH, Qin LQ. Erectile dysfunction and risk of cardiovascular disease: meta-analysis of prospective cohort studies. *J Am Coll Cardiol*. 2011;58:1378-1385.
13. Shamloul R, Ghanem H. Erectile dysfunction. *Lancet*. 2013;381:153-165.
14. Gupta BP, Murad MH, Clifton MM, et al. The effect of lifestyle modification and cardiovascular risk factor reduction on erectile dysfunction: a systematic review and meta-analysis. *Arch Intern Med*. 2011;171:1797-1803.
15. Shaw EJ, Mitchell GC, Tan RB, et al. The non-surgical treatment of Peyronie disease: 2013 update. *World J Mens Health*. 2013;31:183-192.

REVIEW QUESTIONS

1. A 22-year-old man presents to his primary care physician with decreased libido and loss of morning erections. He had a recent low-impact fracture of his right tibia. His height was 185 cm and weight was 230 pounds. Laboratory tests included a serum testosterone of 96 ng/dL on a sample obtained at 4:00 PM and serum LH of 1.0 miU/mL. Which of the following information is *not* necessary in the evaluation of this patient?

A. DEXA scan of the hip and spine
B. Karyotype to evaluate for Klinefelter' syndrome
C. Serum prolactin level
D. Repeat serum testosterone obtained between 7:00 AM and 10:00 AM
E. MRI of the sella with and without gadolinium

Answer: B This patient has a low serum testosterone consistent with his clinical complaints. A DEXA scan is appropriate because he had a low-impact trauma and low serum T is a risk factor for osteoporosis and fractures. The serum testosterone should be repeated even though it is very low because it was collected in the afternoon and serum testosterone declines because of diurnal variation. An MRI and prolactin are appropriate because the low serum testosterone is coupled with a low serum LH, thus raising the possibility that the patient has hypogonadotropic hypogonadism, which could be due to a pituitary mass lesion that might be secreting prolactin (prolactinoma).

2. A 37-year-old man presented to his primary care physician complaining of inability of he and his wife to conceive. They have been married for 3 years. She is 27 years old, is not on contraceptives, and has regular menstrual periods. Neither has had a prior pregnancy. His semen specimen showed no spermatozoa in the ejaculate with a specimen volume of 3 mL. The man has a serum testosterone of 240 ng/dL and serum LH and FSH levels were 2 to 3 times above the upper reference range. A karyotype was XXY. The most appropriate recommendation would be:

A. Referral to a urologist for consideration of testicular sperm extraction and intracytoplasmic sperm injection into eggs obtained from his wife.
B. MRI of the sella.
C. Treatment with HCG and FSH.
D. Treatment with clomiphene.
E. Repeat karyotype.

Answer: A This patient has Klinefelter's syndrome as diagnosed by karyotype. Other than artificial insemination of the partner using donor spermatozoa partner or adoption of a child, the best approach to achieve fertility would be to attempt to extract spermatozoa by testicular sperm extraction and intracytoplasmic sperm injection into oocytes obtained from his female partner. The success rates depend on whether the urologist can find spermatozoa in the testis and the quality of the spermatozoa and the oocyte. There is no need for an MRI of the sella or treatment with HCG or clomiphene, because LH and FSH levels are elevated and a pituitary/hypothalamic lesion is not likely. A repeat karyotype could be done but is not necessary because everything fits the present diagnosis of Klinefelter's syndrome.

3. A 75-year-old man complains of erectile dysfunction (ED). His libido is mildly decreased. He has had two previous myocardial infarctions and is being treated with a calcium channel blocker for hypertension. His morning serum testosterone is 380 ng/dL, with serum LH and FSH in the normal range. The best first-line treatment is:

A. Transdermal testosterone to increase his serum testosterone to over 500 ng/dL.
B. An aromatase inhibitor.
C. A β-blocker to reduce the stress levels of the penis.
D. Phosphodiesterase-5 (PDE-5) inhibitor.
E. Referral to a psychologist to evaluate for psychogenic erectile dysfunction.

Answer: D This elderly man has ED without hypogonadism. The first-line treatment is a PDE-5 inhibitor to increase the vasodilatory functioning in the penis. The other treatments are unnecessary because they are directed toward treatment of primary or secondary hypogonadism (use of testosterone or aromatase inhibitors). A β-blocker is not an effective treatment for ED.

4. An 82-year-old man has decreased libido, decreased muscle strength, and osteopenia on DEXA scan. He has a strong family history of prostate cancer. A serum testosterone level was 300 ng/dL with a free testosterone by equilibrium dialysis that is below the reference range. Serum LH was in the high-normal range. A PSA was 7 ng/mL. The best strategy is to:

A. Treat with a 5α-reductase inhibitor.
B. Treat with testosterone to reverse the manifestation of hypogonadism.
C. Refer to a urologist for assessment of his prostate.
D. Treat with an aromatase inhibitor to reduce estradiol levels.
E. Ignore his complaints because it is normal and acceptable for age

Answer: C The appropriate first-line approach is to refer to a urologist to rule out prostate cancer. The patient is elderly and hypogonadism might be present, but his total testosterone is 300 ng/dL with a low free testosterone that is probably the result of an age-associated increase in SHBG. Even if he were hypogonadal, his PSA is elevated, thus requiring a prostate assessment. Although the clinical symptoms are important to the patient, all treatments (testosterone replacement, use of an aromatase inhibitor to increase endogenous testosterone levels) must be delayed until the prostate risk is assessed. Treatment with a 55α-reductase inhibitor is not appropriate until he has been assessed for possible prostate cancer.

5. A 50-year-old obese man complains of low libido. He had a serum testosterone checked by liquid chromatography tandem mass spectrometry; it was 190 ng/dL. A serum LH level was above the reference range for adult men. The physician decided to treat the patient for hypogonadism with testosterone replacement. Which one of the following adverse events is the most commonly seen?

A. Elevated LDL cholesterol
B. Anemia
C. Hypercalcemia
D. Priapism
E. Erythrocytosis

Answer: E Testosterone increases red blood cell mass by several mechanisms. An increase in both hemoglobin and hematocrit of 5 to 7% from baseline after testosterone is not uncommon. The increase in hemoglobin and hematocrit is dose dependent. Elevation of hematocrit to greater than 54% or hemoglobin greater than 17 g/dL requires either phlebotomy or the withdrawal of testosterone until the hematocrit and hemoglobin decrease into the reference range. Anemia may be seen in testosterone deficiency but not with testosterone treatment. With physiologic replacement of testosterone, LDL cholesterol levels are not elevated. Hypercalcemia does not occur as a result of testosterone therapy. Priapism may occur with PDE-5 inhibitors but not with testosterone treatment.

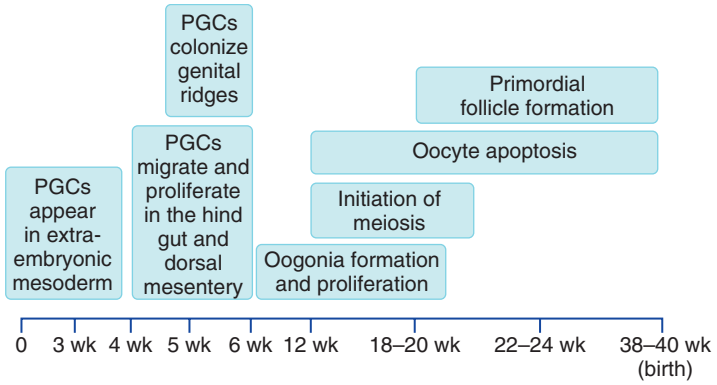

FIGURE 235-1. Diagram illustrating the developmental timetable of the major events that ultimately lead to the formation of primordial follicles during the process of human ovary organogenesis. PGC = primordial germ cell.

235

OVARIES AND PUBERTAL DEVELOPMENT

ROBERT W. REBAR AND WILLIAM H. CATHERINO

DEFINITION

The ovaries or female gonads episodically release female gametes (oocytes or eggs) and secrete sex steroid hormones, principally androstenedione, estradiol, and progesterone. Oocytes are released only during the adult reproductive years, when sex steroid secretion is also greatest, but the ovaries are physiologically active throughout life.

Sex steroids affect the growth, differentiation, and function of a variety of tissues and organs throughout the body; therefore, abnormalities of the ovaries and of sex steroid secretion should be recognized by all physicians. A rational approach to the diagnosis and treatment of reproductive disorders in women requires an understanding of the functions of the ovaries and of their most important unit, the follicle, throughout life.

PHYSIOLOGY

Embryology

Embryogenesis and Differentiation

Organogenesis of the ovaries occurs during fetal life. Ovarian cells are derived from two different sources: (1) primordial germ cells (PGCs) originate at a site outside the prospective gonads, and (2) somatic cells differentiate from the coelomic epithelium and gonadal mesenchyme. In females, PGCs become oocytes, whereas somatic cells differentiate into a variety of cell types, including granulosa, theca, and vascular cells.

PGCs in the human embryo can be distinguished at the gastrula stage (Fig. 235-1). Shortly after formation, PGCs migrate through the dorsal mesentery to the genital ridges. Chemotaxis plays a role in directing PGCs to the gonads. During migration, PGCs proliferate in response to growth factors, most notably Kit ligand. The importance of Kit is demonstrated by the finding that loss-of-function mutations result in a paucity of PGCs, which in turn results in premature ovarian failure.

The genital ridges are characterized by a thickening of the coelomic epithelium and underlying primary mesenchyme. Initially, the gonads are sexually indifferent. Male gonadal differentiation is triggered by the Y chromosome–encoded testis-determining factor SRY. SRY expression results in the differentiation of Sertoli cells and the secretion of müllerian-inhibiting substance, which induces regression of the müllerian ducts. Testicular interstitial cells differentiate into Leydig cells, which secrete testosterone, which in turn stimulates wolffian duct development. In the female, the absence of müllerian-inhibiting substance and testosterone and the activation of the WNT4 pathway leads to the degeneration of the wolffian ducts and the development of the müllerian ducts.[1,2] Thus, the development of the ovaries and female reproductive system is considered a "default" pathway.

When PGCs enter the genital ridges, they begin gametogenesis (see Fig. 235-1). In females, this process is termed *oogenesis* and involves the differentiation of PGCs into oogonia and oocytes. When sex-specific differentiation of the ovary commences, the inactive X chromosome in the PGCs becomes active. This denotes the formation of mitotically active oogonia. The importance of two functional X chromosomes during oogenesis is emphasized by the fact that 45,X females lack oocytes and undergo premature menopause.

After repeated mitosis, oogonia initiate meiosis and become oocytes (see Fig. 235-1). At approximately the same time, granulosa cells differentiate within the gonadal mesenchyme and establish intimate associations with oocytes. Oocytes that become surrounded by granulosa cells stop meiosis after diplotene, and the bivalents enter an interphase state known as dictyotene. If an oocyte is not surrounded by granulosa cells, meiosis continues to diakinesis, and the oocyte dies by apoptosis, although this may not always be true.[3] Granulosa cells, therefore, are critical for oocyte survival. The majority of oocytes die during fetal ovary development (Fig. 235-2), apparently from a lack of contact with granulosa cells.

With further development, the oocyte–granulosa cell complex becomes a primordial follicle. This occurs between the sixth and ninth months of gestation (Fig. 235-3). A primordial follicle consists of a single layer of squamous granulosa cells, a small (about 15 μm in diameter) dictyotene oocyte, and a thin basal lamina (see Fig. 235-3). In the human female, all potential future eggs have entered diplotene of meiosis at the time of birth.

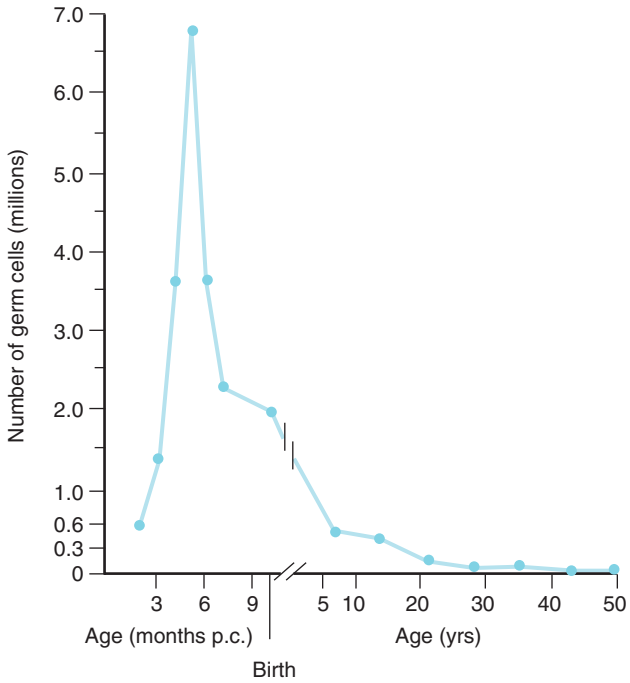

FIGURE 235-2. Changes in the total number of germ cells in the human ovaries during aging. At early to mid-gestation, the number of germ cells increases to almost 7 million; shortly thereafter, the number declines rapidly to about 2 million at birth. The number continues to decline until no oocytes are detected at 50 years of age. p.c. = Post conception. (From Baker TG. Radiosensitivity of mammalian oocytes with particular reference to the human female. *Am J Obstet Gynecol*. 1971;110:746-761.)

Anatomy

The Adult Ovary

The ovaries, which normally measure 2.5 to 5.0 cm long, 1.5 to 3.0 cm wide, and 0.6 to 1.5 cm thick in adults, are organized into two principal parts. A central zone, the medulla, is surrounded by a prominent peripheral zone, the cortex (Fig. 235-4). Growing follicles at different stages of development are present in the cortex. Typically, one follicle per cycle reaches maturity and ovulates its ovum. After ovulation, the follicle transforms into a corpus luteum. The corpus luteum of the cycle lasts for about 14 days, after which it dies and becomes a nodule of dense connective tissue, the corpus albicans (see Fig. 235-4).

The medulla is composed of loose connective tissue with numerous blood vessels and associated nerves. The arterial supply to the ovary originates from two principal sources: the ovarian artery and the uterine artery. These two vessels, which enter the medulla from opposite directions, form an anastomotic trunk and become a common vessel called the ramus ovaricus artery. This artery gives rise to a series of primary branches (spiral arteries) that enter the hilum. In the hilum, numerous secondary and tertiary branches are given off to supply the medulla and the follicles and luteal tissue in the cortex (see Fig. 235-4). The hilum also contains the hilus cells (see Fig. 235-4), which, like the testicular Leydig cells, contain Reinke crystals and secrete testosterone. The physiologic role of the hilus cells is still unknown.

Ovarian Function in Childhood and Puberty

Physical Changes at Puberty

Puberty extends from the earliest signs of sexual maturation until the attainment of physical, mental, and emotional maturity. Pubertal changes in girls result directly or indirectly from maturation of the hypothalamic-pituitary-ovarian (HPO) unit.[4] Human puberty is characterized hormonally by a resetting of the negative gonadal steroid feedback loop, the establishment of new circadian and ultradian (frequent) gonadotropin rhythms, and the acquisition in the female of a positive estrogen feedback loop controlling the menstrual cycle as interdependent expressions of the gonadotropins and ovarian steroids. In girls, pubertal development generally occurs between 7 and 14 years of age. The age at onset and the rate of progress through puberty are variable and depend on genetic, socioeconomic, nutritional, physical, and psychological factors. It appears that there are racial differences in the onset of pubertal development. In the United States, development begins earlier in African American than in white girls.

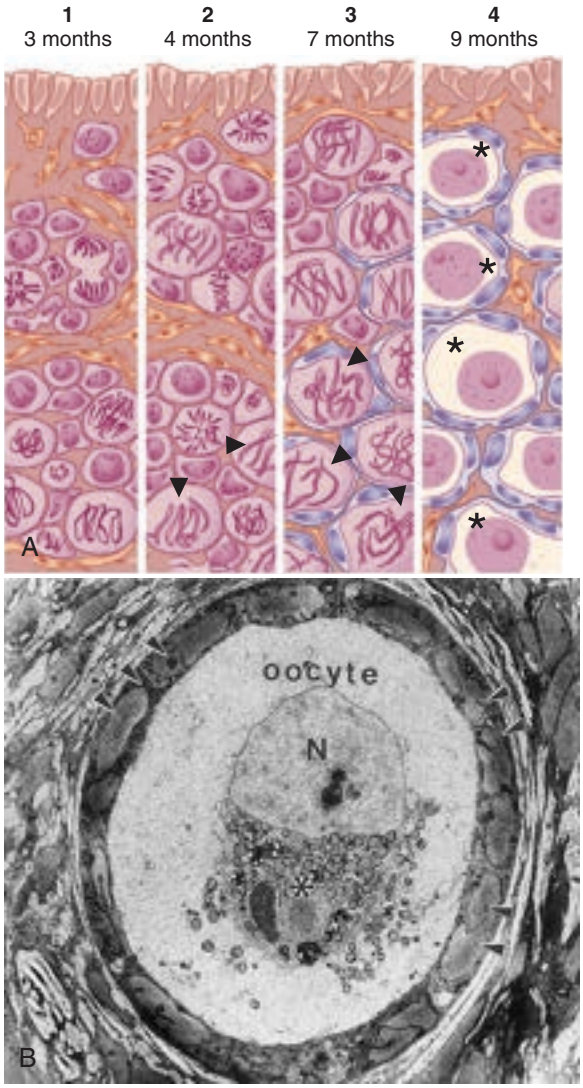

FIGURE 235-3. A, Drawing showing gametogenesis in the human fetal ovary leading to the formation of primordial follicles. At 3 months (1), oogonia divide mitotically. At 4 months (2), some oogonia deep within the cortical cords enter meiosis *(arrowheads)*. At 7 months (3), the cords are no longer distinct, and all germ cells are in meiotic prophase I. At 9 months (4), some oocytes become associated with granulosa cells and appear as primordial follicles *(asterisks)*. **B,** Electron micrograph of human primordial follicle. Granulosa cells *(arrowheads)*, oocyte nucleus (N), and Balbiani's body *(asterisk)* are shown. (From Erickson GF. The ovary: basic principles and concepts. In: Felig P, Baxter JD, Broadus AE, et al, eds. *Endocrinology and Metabolism*, 3rd ed. New York: McGraw-Hill; 1995.)

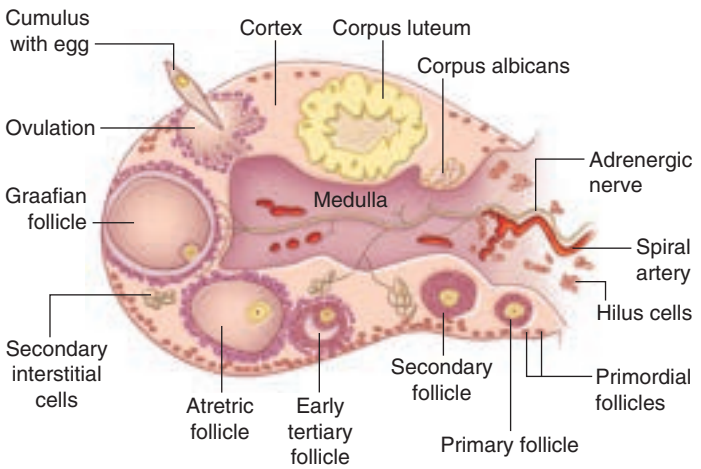

FIGURE 235-4. Diagram showing the anatomy of the human ovary during the reproductive years. Developing follicles and the corpus luteum are located in the cortex; the hilus cells, autonomic nerves, and spiral arteries are present in the medulla. (From Erickson GF. The ovary: basic principles and concepts. In: Felig P, Baxter JD, Broadus AE, et al, eds. *Endocrinology and Metabolism*, 3rd ed. New York: McGraw-Hill; 1995.)

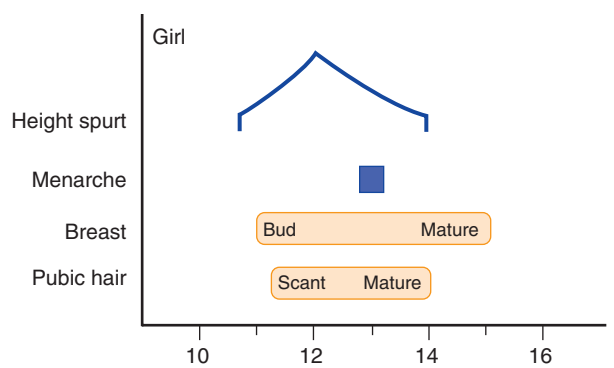

FIGURE 235-5. Temporal sequence of events for the "average" girl during puberty. (From Rebar RW. Practical evaluation of hormonal status. In: Yen SSC, Jaffe RB, Barbieri RL, eds. *Reproductive Endocrinology: Physiology, Pathophysiology, and Clinical Management*, 4th ed. Philadelphia: WB Saunders; 1999:710.)

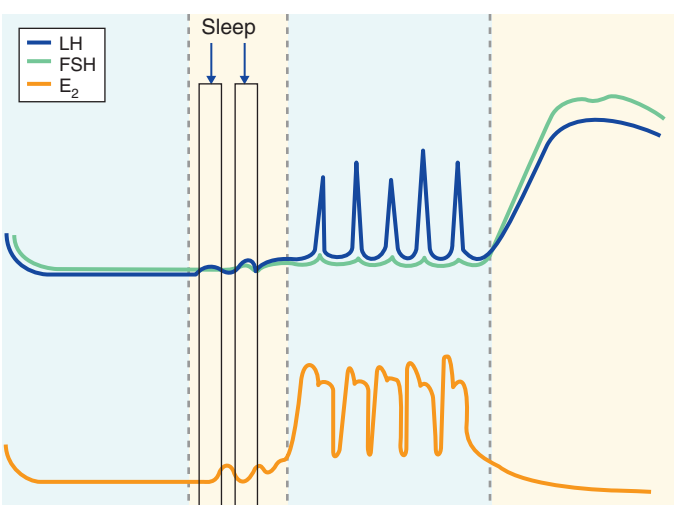

FIGURE 235-6. The changing patterns of luteinizing hormone (LH), follicle-stimulating hormone (FSH), and estradiol (E_2) concentrations in peripheral blood throughout the life of a woman. Not shown is the fact that both LH and FSH are secreted in a pulsatile fashion. The pubertal period has been expanded to illustrate the sleep-associated increases in LH and FSH followed by morning increases in E_2 that are observed during puberty. (Reprinted with permission from Endocrine and Metabolism Continuing Education Quality Control Program, 1982. Copyright American Association for Clinical Chemistry Inc.)

Physical changes occur in an orderly sequence during a definite time frame in puberty (Fig. 235-5).[5] Breast budding in girls is usually the first pubertal change, followed shortly by the appearance of pubic hair, with menarche occurring late in pubertal development. The time from breast budding (mean age of 10.0 years in white girls and 8.9 years in African Americans) to menarche is 2 years. Breast development results from increasing ovarian estrogen production, and pubic and axillary hair results from increasing androgen production. Estrogens are required for the growth of pubic hair as well.

The ovarian sex steroids join with growth hormone and adrenal androgens to produce the adolescent growth spurt. Peak growth velocity is achieved relatively early, with little growth observed after menarche. It has been estimated that more than 50 genes play roles in determining final adult height. It is now clear that estrogen, and not testosterone, is the primary hormone mediating pubertal bone growth in both males and females. Lean body mass, skeletal mass, and body fat are equal in prepubertal boys and girls, but by maturity, women have twice as much body fat as men and less lean body mass and skeletal mass as a result of differences in sex steroid secretion. Estrogens are necessary for the normal formation, mineralization, and maturation of bones. Well-established standards exist for determining whether bone age is appropriate for chronologic age, typically by examining radiographs of the bones of the wrist. Estrogen deficiencies retard, and excesses advance, bone age in relation to chronologic age.

Hormonal Changes

The ovaries function even in early childhood. Low levels of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) are normally present, and these levels increase if the ovaries are removed before puberty, just as they do later in life, indicating the exquisite sensitivity of the hypothalamic-pituitary unit to extremely low circulating sex steroid levels. As puberty nears, there is a progressive decrease in sensitivity of the hypothalamic-pituitary unit to sex steroids, leading to the increased secretion of pituitary gonadotropins, stimulation of sex steroid output, and development of secondary sex characteristics. Increased secretion of both LH and FSH initially occurs at night with sleep and is associated with increased estradiol secretion the next morning (Fig. 235-6). As is true for most hormones, LH and FSH are secreted in an episodic or pulsatile rather than a continuous fashion. Later in puberty, secretion of LH and FSH is increased throughout the 24-hour period, except during the early follicular phase, when nighttime increases still occur. Basal levels of estradiol, the major estrogen secreted by the ovaries, increase throughout puberty. A "critical body mass" may be required for positive estrogen feedback and ovulation. During the first 2 years after menarche, up to 90% of menstrual cycles may be anovulatory because of a delay in synchronization of the HPO axis.

ABERRATIONS IN PUBERTAL DEVELOPMENT

Abnormalities of pubertal development can be divided into four major categories (Table 235-1):

1. Precocious puberty represents any pubertal changes before age 9 years in white girls and before age 8 years in African American girls. This remains controversial. Some clinicians believe evaluation is warranted only if pubertal development begins before age 7 years in white girls and 6 years in African American girls. A compromise may involve careful screening by history and physical examination of girls with early-onset puberty, looking for central nervous system (CNS) symptoms, behavioral concerns, and any other abnormal findings that might warrant further evaluation. The nearer pubertal development begins to the mean age of puberty onset, the less likely it is to have a pathologic basis. Precocious development is isosexual when it is common to the phenotypic sex of the individual and heterosexual when the development is characteristic of the opposite sex. True or central precocious puberty is due to premature maturation of the hypothalamic-pituitary axis. In the absence of increased hypothalamic-pituitary activity, precocious pseudopuberty exists.
2. Delayed (or interrupted) puberty is defined as the absence of any secondary sex characteristics by age 13 years, the absence of menarche by age 16 years, or the passage of 5 years or more from breast budding to menarche.
3. Asynchronous pubertal development occurs when there is deviation from the normal pattern of pubertal development.
4. Heterosexual pubertal development occurs at the appropriate time but has some features characteristic of the opposite sex.

Precocious Puberty

DIAGNOSIS

The overall incidence of precocious puberty has been estimated at 1 in 5000 to 10,000 children. About 10 times as many girls as boys are affected.

Differential Diagnosis

The temporal sequence in which the signs and symptoms of sex steroid hormone excess appear is most important. *Incomplete isosexual precocious puberty* indicates premature development of only a single pubertal feature. If breast budding occurs before age 8 years in the absence of any other development, the diagnosis may be *premature thelarche.* Premature thelarche is believed to be due to transient increases in estrogen secretion or increased breast sensitivity to the small quantities of circulating estrogens present before puberty. Simple ovarian cysts may be present in some girls with this disorder and may be due in some cases to the same genetic abnormality found in girls with McCune-Albright syndrome (Chapters 231 and 248). If pubic or axillary hair develops alone and persists, *premature pubarche* and *adrenarche* must be considered. These abnormalities are associated with slight increases in adrenal androgen secretion but not with clitorimegaly or other signs of virilization. These syndromes require no treatment, and affected girls typically begin true puberty at the usual age. Careful follow-up is required to distinguish these disorders from true precocious puberty.

TABLE 235-1 ABERRATIONS OF PUBERTAL DEVELOPMENT

PRECOCIOUS DEVELOPMENT

- Isosexual precocity
  - Incomplete sexual precocity
    - Premature thelarche
    - Premature pubarche
    - Premature adrenarche
  - True (central) precocious puberty
    - Idiopathic (constitutional)
    - Due to central nervous system lesions
    - Primary hypothyroidism
    - Silver-Russell syndrome
  - Precocious pseudopuberty (of peripheral origin)
    - Ovarian neoplasms
    - Adrenal neoplasms
    - Iatrogenic (estrogen-containing preparations)
    - Human chorionic gonadotropin–secreting neoplasms distinct from central nervous system and ovarian tumors
    - McCune-Albright syndrome
- Heterosexual precocity
  - Ovarian neoplasms
  - Adrenal neoplasms
  - Congenital adrenal hyperplasia
  - Other rare disorders of sexual differentiation

DELAYED PUBERTAL DEVELOPMENT*

- Anatomic abnormalities
  - Mayer-Rokitansky-Küster-Hauser syndrome
  - Distal genital tract obstruction
    - Transverse vaginal septum
    - Imperforate hymen
    - Vaginal agenesis
- Hypergonadotropic hypogonadism (FSH >30-40 mIU/mL)
  - Gonadal dysgenesis
    - With stigmata of Turner syndrome
    - Pure (46,XX or 46,XY)
    - Mixed
  - Ovarian failure with normal ovarian development
    - Genetic disorders
    - Autoimmune disorders
    - Gonadotropin receptor or postreceptor defects (resistant ovary or Savage's syndrome?)
    - Enzymatic defects (17α-hydroxylase deficiency, galactosemia)
    - Physical causes
      - Irradiation
      - Chemotherapeutic agents
      - Viral agents
    - Idiopathic
- Hypogonadotropic or normogonadotropic hypogonadism (LH and FSH <10 mIU/mL, or LH and FSH 6-25 mIU/mL with at least one being <10 mIU/mL)
  - Isolated gonadotropin deficiency
  - In association with midline defects (Kallmann's syndrome)
    - Independent of associated disorders
  - Neoplasms of the hypothalamic-pituitary axis
    - Craniopharyngiomas
    - Pituitary tumors
    - Others
  - Infiltrative processes (Langerhans-type histiocytosis)
  - Idiopathic hypopituitarism
  - "Hypothalamic" forms of amenorrhea
    - Psychogenic
    - Exercise associated
    - Associated with malnutrition
    - Anorexia nervosa
  - Miscellaneous disorders
    - Prader-Labhart-Willi syndrome
    - Laurence-Moon-Bardet-Biedl syndrome
    - Primary hypothyroidism
  - Constitutional delayed puberty

ASYNCHRONOUS PUBERTAL DEVELOPMENT

- Incomplete forms of androgen insensitivity
- Complete forms of androgen insensitivity

HETEROSEXUAL PUBERTAL DEVELOPMENT

- Polycystic ovary syndrome
- Congenital adrenal hyperplasia (female pseudohermaphroditism)
  - 21-Hydroxylase deficiency
  - 11β-Hydroxylase deficiency
  - 3β-ol-Hydroxysteroid dehydrogenase deficiency
- Male pseudohermaphroditism due to 5α-reductase deficiency
- Male pseudohermaphroditism due to partial androgen insensitivity
- Mixed gonadal dysgenesis
- Androgen-producing neoplasms
  - Ovarian
  - Adrenal
- Cushing's syndrome

*No development by age 13 yr, absence of menarche by age 15 yr, or passage of ≥5 yr from breast budding without menarche.
FSH = follicle-stimulating hormone; LH = luteinizing hormone.

When precocious development is isosexual, the purpose of the evaluation is to determine whether the cause is central (true precocious puberty) or peripheral, in which case it is considered gonadotropin-releasing hormone (GnRH)–independent precocious puberty or precocious pseudopuberty. Careful questioning of the patient and her parents may indicate the inadvertent ingestion or absorption of sex steroids (iatrogenic or factitious). As many as 20% of individuals with true precocious puberty have one of several organic brain diseases, including any of several neoplasms, tuberous sclerosis, neurofibromatosis, encephalitis, meningitis, vascular malformations, and hydrocephalus. Because of the seriousness of intracranial lesions, girls with precocious puberty must have radiographic evaluation of the CNS, most effectively by magnetic resonance imaging (MRI). In at least 75% of girls with true precocious puberty, however, no cause is identified (idiopathic or constitutional).

The physical examination may also provide critical information about the cause of the precocious development. Cutaneous café au lait spots, facial asymmetry, polyostotic fibrous dysplasia and other skeletal abnormalities, cranial nerve deficits, and multiple ovarian follicular cysts suggest McCune-Albright syndrome (Chapters 231 and 248) in a girl with precocious development. It is now known that various clones of cells in the endocrine glands of girls with this disorder function autonomously with respect to cyclic adenosine monophosphate production as a consequence of a mutation within exon 8 of the *GNAS* gene, encoding the G_s protein α–subunit. This same mutation probably accounts for the bone lesions and café au lait hyperpigmentation. Other endocrine cells may be similarly affected and lead to pituitary adenomas (usually secreting growth hormone), hyperthyroidism, and, rarely, adrenal hyperplasia.

Studies on the etiologic causes for precocious puberty are ongoing. Heterozygous mutations in the paternal *MKRN3* gene are a diagnosable cause of familial central precocious puberty.[6] Another molecule, kisspeptin, is involved in pubertal development. Kisspeptin and its receptor GPR54 are essential regulators of GnRH-induced gonadotropin secretion and pubertal onset. Activation results in stimulation of the HPO axis, and elevated kisspeptin levels are associated with precocious puberty. Additionally, kisspeptin loss-of-function mutations result in normosmic idiopathic hypogonadotropic hypogonadism.

Abdominal and rectal examination may reveal a mass, suggesting an adrenal or ovarian tumor. Because palpable ovarian cysts may rarely develop before ovulation in true precocious puberty, the presence of a mass does not confirm the diagnosis of precocious pseudopuberty.

When vaginal bleeding is the only sign of development, the diagnosis of sexual precocity should be suspect. Common causes of bleeding in this age group include irritation from a vaginal infection or foreign body, sexual assault, prolapse of the urethral meatus, and ingestion of estrogen-containing medications (most commonly, oral contraceptive preparations). A vaginal or cervical neoplasm is also a rare possibility. Thus, vaginal bleeding requires a vaginal examination, which is often best performed with the patient under anesthesia, before further evaluation is undertaken.

Heterosexual precocity in an apparent prepubertal female is almost always due to congenital adrenal hyperplasia or to an androgen-secreting adrenal or ovarian neoplasm. Only rarely must another disorder of sexual differentiation be considered (Chapter 233). It is important to examine the external genitalia carefully because congenital adrenal hyperplasia is usually associated with some degree of sexual ambiguity.

Excessive androgens produced endogenously by abnormal fetal adrenal glands in utero or diffusing across the placenta to the fetus from the mother can virilize the external genitalia and result in female pseudohermaphroditism. The extent of virilization varies from only an enlarged clitoris to sexual ambiguity sufficient to make gender assignment difficult.

Excessive maternal androgen secretion, typically from an ovarian or adrenal neoplasm, can lead to virilization of a female fetus. This occurs very rarely because of the great capacity of the placenta to aromatize naturally occurring androgens to estrogens. Virilization of a female fetus is much more likely to occur if a pregnant woman has ingested a synthetic steroid preparation with androgenic properties because synthetic compounds generally cannot be aromatized.

Excessive androgen secretion beginning in utero is usually associated with defective cortisol synthesis. As a consequence, pituitary corticotropin secretion is increased, resulting in congenital adrenal hyperplasia and excessive androgen secretion. The three different enzyme defects in the steroidogenic pathway that can lead to virilization of the female fetus are described in Chapter 233. The most common form of congenital adrenal hyperplasia is 21-hydroxylase deficiency, accounting for the disorder in more than 90% of affected individuals. The defect may vary from partial to complete deficiency of the enzyme.

Diagnostic Tests

Measurement of Peptide and Steroid Hormones

Increased levels of immunoreactive human chorionic gonadotropin (HCG) may suggest an HCG-secreting neoplasm, most commonly an ovarian teratoma or dysgerminoma. In such cases, the HCG, which is antigenically and biologically similar to LH, stimulates ovarian steroid secretion and pseudopubertal development. Because even specific LH immunoassays show some cross-reactivity with HCG, values for serum LH may be elevated in individuals with HCG-secreting tumors. Immunoreactive HCG is always elevated in the presence of such tumors. Levels and ratios of FSH and LH typical of pubertal as opposed to prepubertal girls help in the diagnosis of true precocious puberty. Timed urine collections rather than blood samples can be used to measure gonadotropin secretion if necessary. The use of exogenous GnRH to stimulate endogenous LH and FSH secretion can help differentiate gonadotropin-dependent from gonadotropin-independent precocious puberty and is regarded as the "gold standard" in the diagnosis of central precocious puberty.[7] If GnRH is not available, a GnRH analogue can be substituted. Excessively high circulating levels of estrogen (>100 pg estradiol) suggest an estrogen-producing neoplasm or a functioning ovarian cyst. High levels of serum testosterone suggest an ovarian source of excess androgen in girls with heterosexual development, whereas increased levels of dehydroepiandrosterone or its sulfate (the principal precursors of 17-ketosteroids) suggest an adrenal source. High levels of serum 17-hydroxyprogesterone imply congenital adrenal hyperplasia secondary to 21-hydroxylase deficiency, whereas high levels of serum 11-deoxycortisol imply an 11β-hydroxylase deficiency (Chapter 233). In congenital adrenal hyperplasia, these hormone levels should decrease promptly after the oral administration of suppressive doses of dexamethasone. Suppression in response to exogenous corticoids occurs much less consistently in individuals with adrenal cortical adenomas and carcinomas (Chapter 227) and rarely in those with ovarian androgen-secreting neoplasms.

Additional Studies

Imaging of the CNS is the most important test if true precocious puberty is present or if there are any neurologic deficits. Ultrasonography of the adrenals and ovaries or computed tomography (CT) of the adrenals may be indicated to confirm clinical suspicions. In girls with ovarian or adrenal neoplasms, the tumor can almost always be localized radiographically. Catheterization of the ovarian and adrenal veins and measurements of the effluent steroids from each gland should be pursued only when CT, ultrasonography, or MRI fails to identify a suspected neoplasm. Radiographic estimation of bone age is also indicated and serves as a useful tool to follow the results of treatment.

TREATMENT

Rx

Treatment of precocious puberty should be initiated promptly so that the patient's ultimate height is not compromised as a result of sex steroid–induced premature epiphyseal closure and to prevent or attenuate emotional disturbances in the patient and her parents.[8]

GnRH analogues are now the preferred therapy for suppressing gonadotropin secretion, and they also may prevent early bone maturation. No randomized trials have been conducted, but there is universal acknowledgment that GnRH analogues increase ultimate adult height in girls presenting before 6 years of age. Two unresolved issues are whether to initiate treatment with GnRH analogues in girls between 6 and 8 years of age and at what age to stop treatment. In addition, some data suggest that metformin may prevent hirsutism and oligomenorrhea in the teenage years when started in 8- to 12-year-old girls with a history of low-normal birth weight and precocious puberty, although these findings have not been confirmed by other groups. The analogues are not effective in children with McCune-Albright syndrome, and ketoconazole and testolactone have been only marginally successful. Aqueous depot medroxyprogesterone acetate (100 to 200 mg intramuscularly every 2 to 4 weeks) also may be used to suppress gonadotropin secretion; however, it does not always prevent premature epiphyseal closure and the resultant short stature. Effective treatment appears to improve fertility in adult life.[9]

Individuals with CNS or steroid-secreting neoplasms must undergo therapy appropriate for the particular lesion. Girls with congenital adrenal hyperplasia are appropriately managed with glucocorticoids (plus mineralocorticoids when indicated), as outlined in Chapter 233.

Delayed Puberty

Girls who have no evidence of thelarche by age 13 years or who fail to undergo menarche by age 15 years have delayed puberty and should be evaluated.[10] Ovarian failure, congenital absence of the uterus and vagina, and constitutional delay constitute about two thirds of cases in large series. Because of the anxiety generated by delayed puberty, some evaluation is always indicated regardless of the age of the patient.

When pubertal development progresses normally but menstruation does not begin, an abnormality in the genital tract should be considered. Congenital malformations of the müllerian ducts are uncommon, occurring in 0.02% of all women. Most do not cause amenorrhea, and many do not impair reproduction. The anomalies associated with amenorrhea vary in severity from an imperforate hymen to complete aplasia of all müllerian duct derivatives, with vaginal atresia. Although aplasia generally involves all the müllerian duct derivatives, defects may involve only a single part of the distal genital tract. Family aggregates of the most common disorders of müllerian differentiation in females—müllerian aplasia and incomplete müllerian fusion—do occur and are best explained by polygenic or multifactorial inheritance. It is clear that the *HOX* genes, a family of regulatory genes that encode transcription factors, are essential for proper development of the müllerian tract.

A müllerian duct anomaly is suggested by (1) normal levels of serum gonadotropins and steroids, (2) an abnormal outflow tract, (3) a history of cyclic abdominal pain with or without a palpable mass, and (4) normal development of secondary sex characteristics. Normal ovarian function still induces endometrial growth and shedding after menarche if the uterus is normal. In the absence of a normal outflow tract, however, the menstrual effluent is retained and may or may not escape into the abdominal cavity. Free in the abdominal cavity, the effluent may cause endometriosis. Constrained to the uterine cavity, the effluent causes hematometra and a large abdominal mass. In the absence of a mass or cyclic pain, karyotyping is indicated in girls with evidence of an abnormal genital tract to rule out disorders of sexual differentiation (Chapter 233). Such disorders, however, almost never occur together with completely normal pubertal development. In girls with a normal karyotype and a genital tract anomaly, examination under anesthesia and diagnostic laparoscopy should be undertaken to delineate the extent of the defect. When the abnormality consists of an imperforate hymen or transverse vaginal septum only, surgical restoration can be accomplished relatively simply. Attempts to provide an outflow tract for the uterus should not be undertaken if there is no cervix because of the high risk for recurrent pelvic infection. Even with a functional cervix, the construction of an outflow tract that permits successful pregnancy is unlikely. A functional vagina can be constructed surgically or by the daily use of ever-larger dilators. To prevent shrinkage and scarring, surgery should be deferred

until the patient is willing to use dilators on a daily basis or she is about to become sexually active.

Other causes of delayed puberty and primary amenorrhea are the same as those that cause amenorrhea in older women (Chapter 236). When no apparent cause of delayed development is found, constitutional delayed puberty must be entertained as a diagnosis of exclusion.[11] A strong family history of delayed maturation supports this presumption. Small doses of estrogen can be administered to induce some pubertal development, but this may obscure a pathologic cause of the delay and may compromise linear growth and ultimate height.[12]

Asynchronous Pubertal Development

Asynchronous pubertal development is characteristic of male pseudohermaphroditism due to androgen insensitivity, especially complete testicular feminization. This syndrome of androgen insensitivity is inherited as either an X-linked recessive trait or a sex-limited autosomal dominant trait. Despite the presence of intra-abdominal or inguinal testes, there is complete failure of virilization. Affected individuals develop breasts (but only to Tanner stage 3) and a typical female habitus with unambiguous female external genitalia, but with the absence of internal female structures and generally only a foreshortened, blind-ending vagina. Little or no pubic and axillary hair develops. The karyotype is 46,XY in these individuals. Circulating testosterone levels are equivalent to or higher than those found in normal men, LH levels are elevated, and FSH levels are normal compared with those in menstruating women. For a more detailed description, see Chapter 233.

Heterosexual Pubertal Development

POLYCYSTIC OVARY SYNDROME

Polycystic ovary syndrome (PCOS), by far the most common cause of heterosexual pubertal development, is associated with the development of some secondary sex features characteristic of males at the normal age of puberty. Feminization occurs in affected girls, and they develop normal breasts and a typical female habitus, but masculinization also occurs (in contrast, girls with congenital adrenal hyperplasia generally show little if any female development at puberty). A heterogeneous syndrome, PCOS typically begins at or near puberty with hirsutism and irregular menses from the time of menarche. Many girls who develop PCOS are overweight in childhood, and obesity is clearly a risk factor. It now appears that many girls who develop PCOS have alterations in insulin signaling.[13] Menarche may be delayed in a few cases, so young women may present with primary amenorrhea. Basal LH levels tend to be somewhat elevated in perhaps two thirds of cases, and circulating levels of all androgens are moderately elevated. Some degree of insulin resistance is commonly present as well, and hypercholesterolemia may predispose to cardiovascular disease later in life. This is discussed more completely in Chapter 236.

CONGENITAL ADRENAL HYPERPLASIA

Congenital adrenal hyperplasia is generally diagnosed before puberty, and heterosexual precocious pseudopuberty is typical. However, if the defect is mild and changes to the external genitalia are minimal, masculinization may occur at the expected age of puberty. This attenuated or nonclassic form of 21-hydroxylase deficiency seems to occur in families with a strong history of hirsutism. Affected girls generally have some defeminization, with flattening of the breasts, severe hirsutism, relatively short stature, and obesity. For a more detailed description, see Chapter 233.

MIXED GONADAL DYSGENESIS

Mixed gonadal dysgenesis designates asymmetrical gonadal development, with a germ cell tumor or a testis on one side and an undifferentiated streak, rudimentary gonad, or no gonad on the other. The extent of genital virilization before puberty is variable in this rare disorder. Most individuals are reared as girls, in whom virilization occurs at puberty; some may note breast development as well. Affected individuals generally have a mosaic karyotype, with 45,X/46,XY being most common. Short stature and other stigmata associated with a 45,X karyotype in Turner syndrome are less common in patients with tumors than in patients with testes. Gonadectomy is indicated in all individuals with a Y chromosome to eliminate the increased neoplastic potential of such dysgenetic gonads and in all patients in whom virilization occurs at puberty to remove the source of androgen. Estrogen replacement therapy is warranted after gonadectomy. Other causes of male pseudohermaphroditism associated with heterosexual pubertal development are described in Chapter 233.

OTHER CAUSES

An androgen-producing adrenal neoplasm or Cushing's syndrome may occur rarely during the pubertal years and lead to heterosexual development (Chapter 227).

Grade A Reference

A1. Ibáñez L, López-Bermejo A, Díaz M, et al. Early metformin therapy (age 8-12 years) in girls with precocious pubarche to reduce hirsutism, androgen excess, and oligomenorrhea in adolescence. *J Clin Endocrinol Metab*. 2011;96:E1262-E1267.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Sarraj MA, Drummond AE. Mammalian foetal ovarian development: consequences for health and disease. *Reproduction*. 2012;143:151-163.
2. Biason-Lauber A. WNT4, RSPO1, and FOXL2 in sex development. *Semin Reprod Med*. 2012; 30:387-395.
3. Hayashi K, Ogushi S, Kurimoto K, et al. Offspring from oocytes derived from in vitro primordial germ cell-like cells in mice. *Science*. 2012;338:971-975.
4. Choi JH, Yoo HW. Control of puberty: genetics, endocrinology, and environment. *Curr Opin Endocrinol Diabetes Obes*. 2013;20:62-68.
5. Colvin CW, Abdullatif H. Anatomy of female puberty: the clinical relevance of developmental changes in the reproductive system. *Clin Anat*. 2013;26:115-129.
6. Abreu AP, Dauber A, Macedo DB, et al. Central precocious puberty caused by mutations in the imprinted gene *MKRN3*. *N Engl J Med*. 2013;368:2467-2475.
7. Sultan C, Gaspari L, Kalfa N, et al. Clinical expression of precocious puberty in girls. *Endocr Dev*. 2012;22:84-100.
8. Fugua JS. Treatment and outcomes of precocious puberty: an update. *J Clin Endocrinol Metab*. 2013;98:2198-2207.
9. Lazar L, Meyerovitch J, de Vries L, et al. Treated and untreated women with idiopathic precocious puberty: long-term follow-up and reproductive outcome between the third and fifth decades. *Clin Endocrinol (Oxf)*. 2014;80:570-576.
10. Palmert MR, Dunkel L. Clinical practice. Delayed puberty. *N Engl J Med*. 2012;366:443-453.
11. Harrington J, Palmert MR. Clinical review. Distinguishing constitutional delay of growth and puberty from isolated hypogonadotropic hypogonadism: critical appraisal of available diagnostic tests. *J Clin Endocrinol Metab*. 2012;97:3056-3067.
12. Dunkel L, Quinton R. Transition in endocrinology: induction of puberty. *Eur J Endocrinol*. 2014;170:R229-R239.
13. Diamanti-Kandarakis E, Dunaif A. Insulin resistance and the polycystic ovary syndrome revisited: an update on mechanisms and implications. *Endocr Rev*. 2012;33:981-1030.

REVIEW QUESTIONS

1. The testis-determining factor sex-determining region Y (SRY) directly induces differentiation of which one of the following?

A. Sertoli cells
B. Leydig cells
C. Spermatocytes
D. Spermatids
E. Spermatozoa

Answer: A SRY expression leads to the differentiation of Sertoli cells and expression of müllerian-inhibiting substance (or antimüllerian hormone) by these cells. It does not play a role in the differentiation of Leydig cells (which is induced by thyroid hormone and luteinizing hormone, or the various developmental groups of spermatozoa development (which develop in the presence of testosterone). (See Embryogenesis and Differentiation.)

2. In utero, oocyte meiosis arrests at which stage?

A. Prophase
B. Metaphase
C. Anaphase
D. Telophase
E. Interphase

Answer: A During oogenesis, oocytes enter the first phase of meiosis but halt at the diplotene stage of prophase 1. The process of maintaining the prolonged resting state of meiosis is called dictyotene and continues until puberty. Progression beyond the prophase (metaphase, anaphase, telophase, or interphase) involves chromatin condensation and could limit the molecular function of the oocyte DNA. (See Embryogenesis and Differentiation.)

3. Which one of the following is true of precocious puberty in girls?

A. Is diagnosed when pubertal changes occur before 10 years of age
B. Can occur earlier in white girls compared with African American girls
C. Can result in masculine pubertal development
D. Requires premature activation of the hypothalamic-pituitary axis
E. Is less likely to be pathologic if diagnosed at an earlier age

Answer: C Precocious puberty is diagnosed if puberty commences before 9 years of age in white girls and 8 years of age in African American girls. It can result in either isosexual or heterosexual pubertal development. If the cause is central, then the hypothalamic-pituitary-ovarian (HPO) axis is activated; however, noncentral causes (e.g., exogenous gonadal hormone exposure) do not require activation of the HPO axis. The earlier the age of the girl, the more likely it is to be pathologic. (See Precocious Puberty.)

4. Delayed menarche is caused by dysfunction in which one of the following?

A. Hypothalamus
B. Pituitary
C. Ovary
D. Uterus
E. All of the above

Answer: E Delayed menarche can result from the absence of gonadotropin-releasing hormone pulsatility, disruptions in gonadotropin release from the pituitary, loss of functional oocytes, or obstruction of the outflow tract, to name a few. Disruptions at any point along the female reproductive axis can result in disruptions in the orderly transition through puberty. (See Delayed Puberty.)

5. The diagnosis of polycystic ovarian syndrome (PCOS) absolutely requires which one of the following?

A. Evidence of oligo-ovulation or anovulation
B. Evidence of hyperandrogenism
C. Evidence of polycystic-appearing ovaries by imaging
D. Exclusion of other disorders that can cause menstrual irregularity and hyperandrogenism
E. All of the above

Answer: D The diagnosis of PCOS requires two of the following three: evidence of oligo-ovulation/anovulation, hyperandrogenism, and/or polycystic-appearing ovaries. No one of them is absolutely required. However, ruling out other disorders that can appear to be PCOS must be done before PCOS can be diagnosed. (See Polycystic Ovarian Syndrome.)

236

REPRODUCTIVE ENDOCRINOLOGY AND INFERTILITY

ROBERT W. REBAR AND WILLIAM H. CATHERINO

THE NORMAL MENSTRUAL CYCLE

Between menarche and menopause, the reproductive organs of normal women undergo a series of closely coordinated changes at monthly intervals that constitute the normal menstrual cycle. The menstrual cycle is the expression of the coordinated interactions of the hypothalamic-pituitary-ovarian axis, with associated changes in the target tissues (endometrium, cervix, vagina) of the reproductive tract.

A menstrual cycle begins with the first day of genital bleeding (day 1; menses) and ends just before the next menstrual period. The median menstrual cycle length is 28 days but normally ranges from 21 to 35 days. Menstrual cycles vary most in the years immediately after menarche and preceding menopause, partly because of an increase in anovulatory cycles. Irregularities in menstrual cycle length may be caused by changes in diet, exercise, emotional disturbances, parturition, or abortion. The menstrual cycle has three distinct phases: follicular, ovulatory, and luteal.

Follicular (Preovulatory) Phase

The follicular phase begins with the first day of bleeding and extends to the day before the preovulatory luteinizing hormone (LH) surge. A rise in serum follicle-stimulating hormone (FSH) begins in the late luteal phase of the previous menstrual cycle and continues into the early follicular phase. This initiates development of a group of follicles (Fig. 236-1). The preovulatory follicle destined for ovulation is selected from this cohort. After the early follicular phase, FSH levels fall, and LH levels rise slowly. About 7 days before the preovulatory LH surge, estradiol and estrone increase until the day before the LH surge. The divergence in LH and FSH levels may be related to secretion of inhibin B, which inhibits the release of FSH. Several days before the LH surge, plasma androgens begin to increase. They peak on the day of the LH surge. Progesterone does not increase until just before the LH surge onset.

Ovulatory Phase

During this phase, the ovum is released from the mature graafian follicle about 32 to 34 hours after the preovulatory LH surge. The ovulatory phase extends from 1 day before the LH surge to 1 day after the LH surge (see Fig. 236-1). Some women experience unilateral pelvic pain near the time of ovulation, termed *mittelschmerz*, which occurs before or after ovulation. During the ovulatory phase, a rapid rise in plasma LH in response to positive estrogen feedback leads to ovulation. As peak LH levels are reached, estradiol levels drop, but progesterone levels increase.

Luteal (Postovulatory) Phase

The luteal phase is about 14 days in length and ends with the onset of menses (see Fig. 236-1). This phase includes the functional lifespan of the corpus

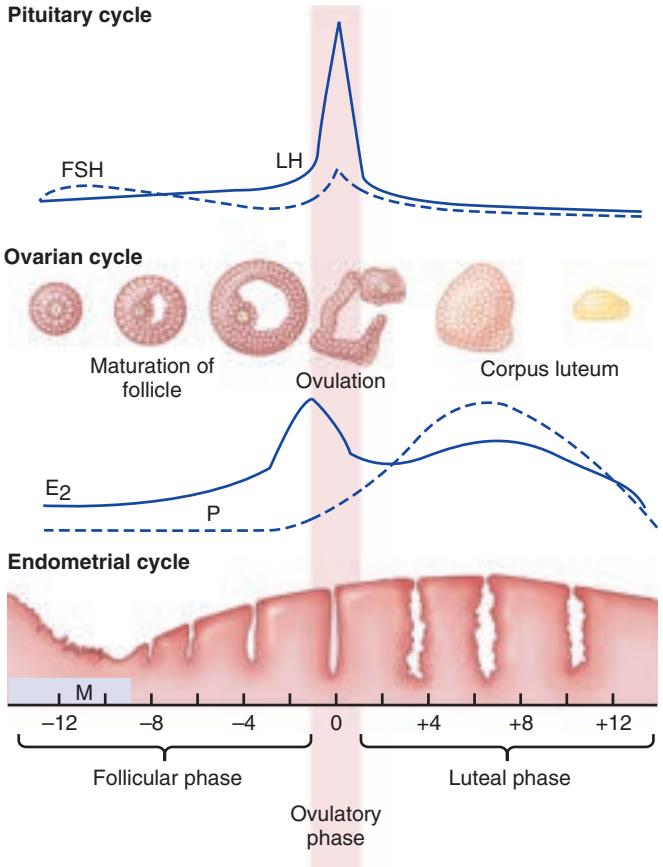

FIGURE 236-1. The idealized cyclic changes observed in gonadotropins, estradiol (E_2), progesterone (P), and uterine endometrium during the normal menstrual cycle. The data are centered on the day of the luteinizing hormone (LH) surge (day 0). Days of menstrual bleeding are indicated by M. FSH = follicle-stimulating hormone; LH = luteinizing hormone. (From Endocrine and Metabolism Continuing Education Quality Control Program, 1982. Copyright American Association for Clinical Chemistry, Inc.)

luteum, which supports the released ovum by secreting progesterone. Progesterone secretion increases up to 8 days after the LH surge. Smaller increases in 17α-hydroxyprogesterone, estradiol, and estrone occur. Progesterone decreases before menses unless the ovum is fertilized and pregnancy results. A serum progesterone level higher than 3 ng/mL 1 week before menses is probably diagnostic of ovulation.

CYCLIC CHANGES IN TARGET ORGANS

Endometrium

The endometrium undergoes histologic and cytologic changes that culminate with menstrual bleeding when the corpus luteum ceases to secrete progesterone (see Fig. 236-1). The basal layer of the endometrium then regenerates the superficial layer of compact epithelial cells lining the uterine cavity and an intermediate layer of spongiosa. Both superficial layers are shed at menstruation. Endometrial glands proliferate under the influence of estrogen, and the mucosa thickens. In the luteal phase, the glands become coiled and secretory, with increased vascularity and edema of the stroma. When estradiol and progesterone decline, the stroma becomes edematous, endometrial and blood vessel necrosis occurs, and bleeding ensues. Local release of prostaglandins may initiate vasospasm with ischemic necrosis and uterine contractions that accompany menstrual flow. Prostaglandin synthetase inhibitors can relieve menstrual cramping. The histologic changes are characteristic; therefore, endometrial biopsies can be used to characterize the stage of the cycle and to assess the tissue response to gonadal steroids.

Cervix and Cervical Mucus

During the follicular phase, cervical vascularity, congestion, and edema increase as a result of estrogen. Cervical mucus increases in quantity (10- to 30-fold) and in elasticity. So-called ferning becomes prominent. Progesterone stimulates cervical mucus thickening and loss of elasticity and ability to fern. These characteristics are useful in evaluating the stage of the cycle and the amount of estrogen present.

Vagina

Low estrogen is associated with pale, thin vaginal epithelium. As estrogens rise, the number of cornified epithelial cells increases. Subsequently, progesterone decreases the percentage of cornified cells and increases the number of precornified intermediate cells. There is also increased cellular debris and clumping of shed desquamated cells. Histologic changes in the vaginal epithelium are sensitive indicators of estrogen status.

Ovary

The ovaries produce a single dominant graafian follicle that grows and develops to the preovulatory stage during the follicular phase. This process is brought about by the combined actions of FSH and LH on the follicle wall to increase estradiol biosynthesis. The LH surge acts on the preovulatory follicle to cause the secretion of the mature fertilizable oocyte. After ovulation, the follicle wall transforms into the corpus luteum, which produces progesterone and estradiol. If implantation does not occur, the corpus luteum undergoes luteolysis and stops hormone production. In the late luteal phase, another dominant follicle develops, and a new menstrual cycle begins.

Chronology of Folliculogenesis

The preovulatory follicle begins its development when a primordial follicle is recruited into the pool of growing follicles. There are two major phases of folliculogenesis: the preantral (gonadotropin-independent) and the antral (gonadotropin-dependent) periods. The first phase is characterized by growth of the oocyte and granulosa proliferation. Preantral folliculogenesis proceeds slowly, requiring at least 300 days. During the second phase, granulosa and theca cells proliferate, and the antrum enlarges. The graafian follicle increases relatively rapidly as it develops. The mature graafian follicle that will ovulate requires 40 to 50 days to complete the antral phase.

Selection

The dominant follicle is selected from a cohort at the end of the luteal phase of the previous menstrual cycle. The selected follicle requires 20 days to develop to the ovulatory stage.

Shortly after the midluteal phase of the cycle, the granulosa cells show a sharp increase in the rate of mitosis. The first indication of selection is that the granulosa cells continue dividing at a high rate. As a consequence of the high and sustained mitotic rate and the progressive accumulation of follicular fluid, the dominant follicle undergoes remarkable growth. The increase in plasma FSH levels that begins at the end of luteal phase and continues through the early follicular phase evokes follicle selection. The concentration of FSH in the follicular fluid of the healthy (dominant) follicle increases but does not increase in the nondominant atretic follicles. The manner in which this selective increase in FSH is controlled is unknown. More than 99.9% of all follicles are not selected and undergo atresia.

Mechanism of Follicle-Stimulating Hormone Action

FSH exerts its influence on follicle growth and development by stimulating granulosa cell mitosis and cytodifferentiation. FSH activates high-affinity receptors on the granulosa cell. The binding event is transduced into an intracellular signal through the cyclic adenosine monophosphate (cAMP)-dependent protein kinase A signal transduction pathway. This leads to an increase in cell number and cytodifferentiation.

The granulosa cells in the chosen follicle continue to divide at a relatively rapid rate throughout the follicular phase of the cycle, increasing from 1×10^6 cells to 50×10^6 cells at ovulation. FSH induces granulosa cytodifferentiation in the dominant follicle. Expression of cytochrome P-450 aromatase (P450arom) is induced by FSH. The temporal pattern of expression of P450arom determines when and how much estradiol is produced. FSH also induces the expression of LH receptors. This is delayed until the dominant follicle is fully differentiated at day 12. The large number of LH receptors in the granulosa cells is essential for the LH surge to trigger ovulation. The terminal differentiation of the granulosa cells is characterized by the accumulation of other FSH-regulated gene products, including inhibin B, activin, and follistatin.

Mechanism of Luteinizing Hormone Action

The primary function of LH in follicle development is to stimulate androgen production by the theca interstitial cells. LH binds to LH-human chorionic gonadotropin (HCG) receptors located on theca cells, which interact with G

proteins and activate adenylate cyclase, leading to the synthesis of cAMP, which stimulates gene expression through protein kinase A. This leads to increased conversion of cholesterol to androstenedione. Theca cells express insulin receptors, and insulin stimulates theca androgen production. The crosstalk between the insulin and LH receptor signaling is clinically relevant because of the relationship between hyperinsulinemia and hyperandrogenism in women.

One of the most important consequences of FSH and LH action on the dominant follicle is the production of estradiol. This physiologically important process is called the *two-gonadotropin, two-cell concept* of follicular estrogen production.

Ovulation

At midpoint in the menstrual cycle, the preovulatory surges of LH and FSH act on the preovulatory follicle to initiate the events leading to ovulation. The LH surge induces meiotic maturation, a process that converts the oocyte into a fertilizable egg arrested at the second meiotic metaphase. During meiotic maturation, the granulosa cells next to the oocyte are stimulated by FSH to undergo cumulus expansion. This is a prerequisite for the oocyte's pickup and transport by the oviduct. The LH surge also stimulates production of proteolytic enzymes in the vicinity of the presumptive stigma. This process requires the LH stimulation of progesterone and prostaglandins, which are obligatory for stigma formation. After 36 hours, the fertilizable egg and surrounding cumulus cells are secreted through the stigma.

Luteogenesis

Ovulation leads to changes in the granulosa and theca cells of the ovulated follicle that result in increased production of progesterone and estradiol during the first week of the luteal phase. This event, termed *luteinization*, is important for the formation and development of a secretory endometrium. Three major physiologic mechanisms are responsible for luteinization: removal of luteinization inhibitors; secretion of LH by the pituitary; and delivery of high levels of cholesterol. The induction of StAR, P450c22, and 3β-hydroxysteroid dehydrogenase in the granulosa lutein cells leads to progesterone production by the corpus luteum. The two-cell, two-gonadotropin mechanism is responsible for estradiol production. If implantation does not occur, the corpus luteum initiates luteolysis, leading to decreases in progesterone, estradiol, and apoptosis. When luteolysis occurs, another dominant follicle is selected, and a new menstrual cycle begins.

Neuroendocrine Regulation of the Ovaries

Neurons containing various peptide hormones that can release or inhibit secretion of the gonadotropins are found in the hypothalamus (Chapter 223). Cells containing gonadotropin-releasing hormone (GnRH) occur in the arcuate nucleus, median eminence, and preoptic area. Axons from these neurons run in the tuberoinfundibular tract and terminate on capillaries within the median eminence, allowing delivery of their products to the anterior pituitary gland. Neurotransmitters, including norepinephrine, dopamine, and serotonin, as well as neuromodulators, such as endogenous opiates and prostaglandins, influence secretion of GnRH. Estrogens and androgens bind to cells in the hypothalamus and the anterior pituitary, and progestins bind to cells in the hypothalamus, to influence hypothalamic-pituitary regulation of ovarian function.

GnRH is secreted in a pulsatile fashion and is responsible for pulsatile release of gonadotropins. Pulsatile gonadotropin release accounts for the pulsatile secretion of sex steroids from the ovaries. The ovarian sex steroids feed back on the hypothalamic-pituitary unit to modulate both the frequency and amplitude of the gonadotropin pulse (Fig. 236-2). Gonadotropin pulses vary throughout the menstrual cycle. Pulses occur at approximately 60- to 90-minute intervals in the follicular phase and at intervals of 180 minutes in the luteal phase.

Gonadal steroids can exert both negative and positive feedback effects on gonadotropin secretion. 17β-Estradiol is the most potent inhibitor of gonadotropin secretion. For women to ovulate, estradiol must also elicit a positive feedback effect on gonadotropin release. The feedback effects are both time and dose dependent. In the normal menstrual cycle, the positive feedback action of estradiol leading to the LH surge is preceded by a period when lower estradiol levels are present.

It appears that the ovary is the "clock" for the timing of ovulation, with the hypothalamus stimulating pulsatile release of the gonadotropins. The follicle complex and corpus luteum develop in response to gonadotropin stimulation. For appropriate ovarian regulation of reproductive function in women, three biologic characteristics are necessary: an appropriate balance and sequence of negative and positive feedback actions; differential feedback effects on the release of LH and FSH; and local intraovarian controls on follicular growth and maturation.

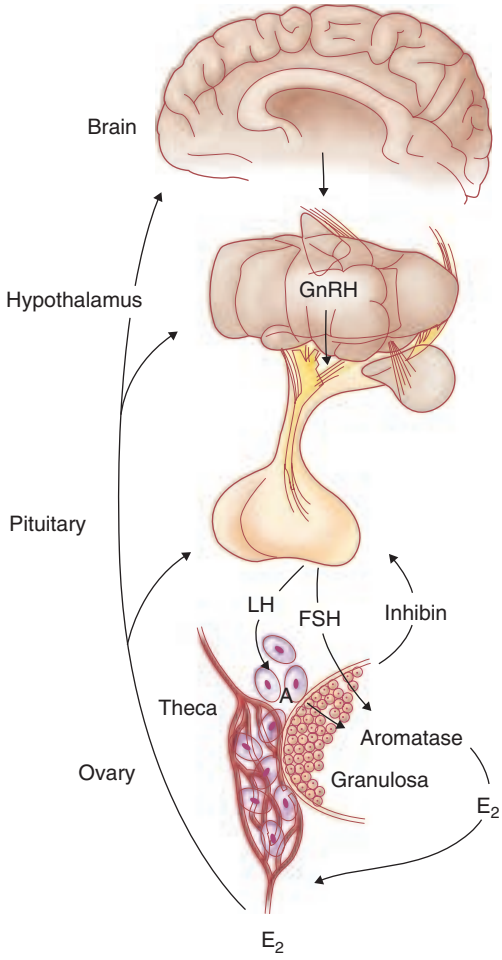

FIGURE 236-2. The hypothalamic-pituitary-ovarian axis in the regulation of follicular maturation and steroidogenesis. A = androgens; E_2 = estradiol; FSH = follicle-stimulating hormone; GnRH = gonadotropin-releasing hormone; LH = luteinizing hormone. (Modified from Endocrine and Metabolism Continuing Education Quality Control Program, 1982. Copyright American Association for Clinical Chemistry, Inc.)

ABNORMALITIES OF THE REPRODUCTIVE YEARS

Dysmenorrhea and Endometriosis

DEFINITION

Dysmenorrhea, defined as painful menstruation, affects about 50% of postpubertal women[1] and can be classified as primary or secondary. *Endometriosis,* which may result in dysmenorrhea, infertility, and dyspareunia (i.e., painful intercourse), is the ectopic occurrence of endometrial tissue, most commonly within the abdominal cavity but sometimes in surgical scars, on the vulva, in the umbilicus, and elsewhere.[2]

PATHOBIOLOGY

Primary dysmenorrhea occurs only in ovulatory cycles. Prostaglandins produce dysmenorrhea by initiating painful, exaggerated uterine contractions and myometrial ischemia. Associated systemic symptoms include nausea, diarrhea, headache, and emotional changes. In secondary dysmenorrhea, there is a pathologic cause, with endometriosis being the most common. Other causes include pelvic inflammatory disease; congenital abnormalities, such as atresia of a portion of the distal genital tract and cystic duplication of the paramesonephric ducts; and cervical stenosis. Recent studies suggest the possibility that the pain of endometriosis is caused by the presence of nerve fibers within ectopic endometrium.

TREATMENT Rx

Prostaglandin synthetase inhibitors (Chapter 37), such as naproxen, ibuprofen, mefenamic acid, and indomethacin, are used to treat primary dysmenorrhea.[3] If the dysmenorrhea persists, addition of an oral contraceptive to inhibit ovulation and limit prostaglandin release is generally effective. In cases in which the pelvic pain remains intractable, additional evaluation is warranted. If thorough evaluation of the gastrointestinal and urinary tracts fails to reveal a definitive cause, examination under anesthesia and diagnostic laparoscopy may be indicated.

If endometriosis is diagnosed at laparoscopy, treatment varies according to the severity of the disease and the goals of the patient regarding fertility.[4] It may be possible to fulgurate implants or to lyse adhesions. Available data do not demonstrate a benefit of excision over ablation at the time of initial diagnosis as the preferred approach to controlling symptoms. From this point onward, efforts should be directed toward treating endometriosis medically, with additional surgery deferred until infertility (if present) becomes manifest. Medical therapy can consist of continuous suppression with oral contraceptives, progestins (oral, injectable, or implantable), or GnRH analogues or danazol for 3 to 6 months. GnRH analogues are currently the most frequently used form of medical suppressive therapy. After a course of treatment, use of oral contraceptive agents should probably be continued until fertility is desired. Conservative surgical resection of endometriotic tissue should almost always be deferred until it is established as the cause of infertility. Surgery may be required, however, for continuing severe pain, severe endometriosis, or large ovarian cysts containing endometriosis (endometriomas). If symptoms continue despite adequate treatment or if psychological overlay is suspected, psychiatric evaluation may be indicated. Medical causes of dysmenorrhea, however, should be eliminated first.

Premenstrual Syndrome

DEFINITION

Premenstrual syndrome (PMS), also known as *premenstrual tension*, is a complex of physical and emotional symptoms that occur repetitively in a cyclic fashion before menstruation and that diminish or disappear with menstruation.[5]

DIAGNOSIS

The cyclic symptoms typically are sufficiently severe to interfere with some aspects of life. More than 150 different symptoms are now thought to vary with the menstrual cycle (Table 236-1). Estimates of the prevalence of PMS range from 25 to 100%. The *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) classifies severe PMS as premenstrual dysphoric disorder (PMDD). For most women, PMS is merely annoying; severe PMS (or PMDD) causes serious difficulties for 3 to 5% of women of reproductive age. The diagnosis of both PMS and PMDD is best established by requiring patients to keep prospective daily records of symptoms during a 2- to 3-month period. Less than 50% of women complaining of PMS are found to have the syndrome when such records are examined.

TABLE 236-1 COMMON SYMPTOMS OF CYCLIC PREMENSTRUAL SYNDROME

SOMATIC SYMPTOMS	
Abdominal bloating	Constipation or diarrhea
Acne	Headache
Alcohol intolerance	Peripheral edema
Breast engorgement and tenderness	Weight gain
Clumsiness	
EMOTIONAL AND MENTAL SYMPTOMS	
Anxiety	Insomnia
Change in libido	Irritability
Depression	Lethargy
Fatigue	Mood swings
Food cravings (especially salt and sugar)	Panic attacks
Hostility	Paranoia
Inability to concentrate	Violence toward self and others
Increased appetite	Withdrawal from others

Most women seek help for PMS in their 30s after 10 or more years of symptoms. Many report that their symptoms began at menarche; approximately half state that symptoms followed childbirth. Severity and duration of symptoms are often reported to increase after each successive pregnancy and to become more severe with advancing age. Women with severe long-standing PMS almost always describe associated psychological reactions, including social difficulties, such as marital discord, difficulty relating to their children, difficulty maintaining friendships, and withdrawal from social activities.

TREATMENT Rx

General Measures

The cause of PMS is unknown, and patients should be informed that no one therapy has been effective in all women. Women with mild premenstrual symptoms often benefit from simple changes in lifestyle, including daily mild aerobic exercise; reduction in intake of caffeine-containing beverages, salt, and refined sugar, particularly in the luteal phase; stress reduction; and adequate rest.

Medical Therapy

Women with more severe PMS may benefit from symptomatic treatment. Continuous oral contraceptives have inconsistent but generally positive therapeutic benefit.[6] Bromocriptine (generally 2.5 mg twice a day) or danazol (100 to 400 mg/day in two divided doses) may be given continuously for relief of mastalgia (breast pain), although this use is not listed in the manufacturer's directive or approved by the U.S. Food and Drug Administration (FDA), and efficacy has not been documented by rigorous randomized trials. Prostaglandin synthetase inhibitors may help reduce dysmenorrhea and may alleviate headaches. Mild sedatives and tranquilizers may help reduce insomnia and anxiety. Low doses of fluoxetine (20 mg) and other selective serotonin reuptake inhibitors, administered either daily or for the last 2 weeks of each menstrual cycle, are highly effective in reducing the emotional symptoms associated with PMS. Mild diuretics (especially spironolactone at doses up to 100 mg each morning) may benefit cyclic edema.

Natural progesterone, given in the form of vaginal suppositories, has been used, but results of double-blind placebo-controlled trials show no efficacy. Likewise, the value of large quantities of multiple vitamins or of oil of evening primrose, containing the essential fatty acid γ-linolenic acid, a precursor of prostaglandins, is unsubstantiated.

Surgical Therapy

Because PMS requires the occurrence of cyclic ovulation, oophorectomy is occasionally considered for patients with particularly intractable symptoms. However, oophorectomy may create new problems related to estrogen deficiency for women with PMS treated in this permanent fashion. Several trials employing a GnRH agonist together with exogenous steroids have been described as reducing PMS. Whether such therapy can be used in the long term remains to be determined.

Abnormal Uterine Bleeding

DIAGNOSIS

Differential Diagnosis

Because there is considerable confusion about terminology of abnormal uterine bleeding, it is important to determine just what is being included in any term used other than *abnormal uterine bleeding*.[7] Postmenarchal bleeding in adolescents secondary to immaturity of the hypothalamic-pituitary-ovarian axis (resulting in anovulation) accounts for about 20% of cases, and perimenopausal bleeding consequent to incipient ovarian failure constitutes more than half.

The causes of abnormal uterine bleeding in the reproductive years include complications from the use of oral contraceptives; complications of pregnancy, especially threatened, incomplete, or missed miscarriages and ectopic pregnancy; coagulation disorders, most commonly idiopathic thrombocytopenic purpura (Chapter 172) and von Willebrand disease (Chapter 173); and pelvic disease, such as intrauterine polyps, leiomyomas, and tumors of the vagina and cervix. Clear cell adenocarcinoma of the vagina or cervix (Chapter 199) may occur in women exposed to diethylstilbestrol during fetal life. Affected women may also have congenital abnormalities of the upper vagina, cervix, and uterus. Women with a history of diethylstilbestrol exposure should be reassured that the incidence of malignant change is extremely low. Trauma (coital or otherwise), foreign bodies, systemic illnesses including various endocrinopathies (e.g., diabetes mellitus, hypothyroidism and hyperthyroidism, Cushing's syndrome, and Addison's disease), leukemia, and

renal disease may also be associated with abnormal bleeding as the presenting manifestation.

Abnormal uterine bleeding with no demonstrable organic genital or extragenital cause (75% of cases) is most frequently associated with anovulation and is appropriately termed *anovulatory* (sometimes termed *dysfunctional*) *bleeding*. Most anovulatory bleeding is due to either estrogen withdrawal or estrogen breakthrough bleeding. In anovulatory women, estrogen stimulates the endometrium unopposed by progesterone. The endometrium proliferates, becomes thicker, and may shed irregularly. Anovulatory bleeding tends to occur at less frequent intervals, and organic lesions tend to cause bleeding more frequently than cyclic menses.

Clinical Evaluation

All cases of abnormal bleeding should be evaluated, beginning with a thorough history emphasizing the amount and duration of blood loss. Prospective charting of the days on which bleeding occurs may be required to evaluate the bleeding pattern. Complications of pregnancy or a bleeding diathesis must always be ruled out.

The findings on physical examination (including the Papanicolaou smear) are normal in anovulatory bleeding except for signs of anemia in the more severe cases. Laboratory tests should include a complete blood count, platelet count, coagulation studies, thyroid function tests, and fasting blood glucose concentration. Anovulatory bleeding must be a diagnosis of exclusion, with management depending on the age of the patient and the extent of the bleeding. A sample of the endometrium should be obtained by biopsy or by dilation and curettage from all women older than 35 years and from those at increased risk for endometrial carcinoma because of prolonged anovulatory bleeding.

TREATMENT

Even profuse bleeding in hemodynamically stable anovulatory women can almost always be successfully treated by the administration of one combination oral contraceptive pill every 6 hours for 5 to 7 days, although this use is not listed in the manufacturer's directive or approved by the FDA. Bleeding should cease within 24 hours, but patients should be warned to expect heavy bleeding 2 to 4 days after therapy is stopped. If anemia is profound, blood transfusion may be necessary. If the bleeding continues despite therapy, curettage can be carried out. Recurrence can be prevented by giving the patient combination oral contraceptive agents cyclically if pregnancy is not desired. If pregnancy is desired, ovulation can be induced.

Acute episodes of anovulatory bleeding can also be treated with conjugated estrogens administered intravenously (25 mg every 4 hours for up to three doses) until bleeding ceases, although this use is not listed in the manufacturer's directive or approved by the FDA. Progestin therapy (medroxyprogesterone acetate, 5 to 10 mg orally for 10 days) should be started simultaneously. Withdrawal bleeding occurs after cessation of therapy, and the patient can then be treated with oral contraceptive agents for at least three cycles.

For individuals with anovulatory bleeding without an episode of profuse bleeding, treatment with cyclic oral contraceptive agents or progestin can be provided unless pregnancy is desired, in which case ovulation must be induced.

Endometrial ablation by any of several methods is being used increasingly to treat persistent bleeding. However, ablation is not 100% effective, and medical management remains the first line of therapy for most women. Hysterectomy may be an appropriate choice for a small number of women.

Amenorrhea

DEFINITION

Amenorrhea is the absence of menstruation for 3 months or more in women with past menses (secondary amenorrhea) or the absence of menarche by the age of 15 years regardless of the absence or presence of secondary sex characteristics (primary amenorrhea).[8]

PATHOBIOLOGY

If an intact genital outflow tract exists and there is no primary disease of the uterus, amenorrhea is a sign of failure of the hypothalamic-pituitary-ovarian axis to produce cyclically the hormones necessary for menses. Amenorrhea is physiologic in the prepubertal girl, during pregnancy and early in lactation, and after menopause. At any other time, it is pathologic and demands evaluation. Use of the term *postpill amenorrhea* to refer to failure to resume menses within 3 months of discontinuation of oral contraceptives is inappropriate. Women so affected should be evaluated in the same manner as any woman with amenorrhea. Similarly, individuals with menses occurring at infrequent intervals of more than 40 days or having fewer than nine menses per year, termed *oligomenorrhea*, should be evaluated identically to women with amenorrhea.

DIAGNOSIS

Clinical Evaluation

In patients with amenorrhea, even subtle hormonal abnormalities may lead to signs and symptoms. Breast development indicates exposure to estrogens, and the presence of pubic and axillary hair indicates androgenic stimulation.

Patients should be questioned especially closely for evidence of psychological disturbances, dietary and exercise habits, lifestyle, environmental stresses, family history of genetic anomalies, abnormal growth and development, and signs of hyperandrogenism, including hirsutism, temporal balding, deepening of the voice, increased muscle mass, clitorimegaly, and increased libido, and signs of defeminization, including decreasing breast size and vaginal atrophy. Any history of galactorrhea should be determined. A history of symptoms related to thyroid and adrenal dysfunction should also be sought (Chapters 224, 226, and 227).

The physical examination should focus on the evaluation of body dimensions and habitus, extent and distribution of body hair, breast development and secretions, and genitalia. In normal adult women, the arm span is similar to the height; in hypogonadal women, the span is generally more than 5 cm greater than the height. The distribution and quantity of body hair should be considered in view of the family history. The extent of any hirsutism should be recorded, preferably by photographs. Other signs of virilization should be sought carefully. Breast development should be graded according to the method of Tanner (Table 236-2).[9] Breast secretion should be sought by applying pressure to the breasts while the patient is seated. Any secretion should be examined microscopically for the presence of perfectly round fat globules of varying size, which indicate galactorrhea. Finally, the female genitalia should be examined carefully because they are such sensitive indicators of the hormonal milieu. The Tanner stage of pubic hair development should be noted (see Table 236-2).

Because the sensitivity of the genitalia to androgens decreases onward from early in fetal development, the extent of any virilization is important. Fusion of the labia and enlargement of the clitoris with or without formation

TABLE 236-2 CRITERIA FOR DISTINGUISHING TANNER STAGES 1 TO 5 DURING PUBERTAL MATURATION

TANNER STAGE	BREAST	PUBIC HAIR
1 (Prepubertal)	No palpable glandular tissue or pigmentation of areola; elevation of areola only	No pubic hair; short, fine vellus hair only
2	Glandular tissue palpable with elevation of breast and areola together as a small mound; areolar diameter increased	Sparse, long, pigmented terminal hair chiefly along the labia majora
3	Further enlargement without separation of breast and areola; although more darkly pigmented, areola still pale and immature; nipple generally at or above midplane of breast tissue when individual is seated upright	Dark, coarse, curly hair, extending sparsely over mons
4	Secondary mound of areola and papilla above breast	Adult-type hair, abundant but limited to mons and labia
5 (Adult)	Recession of areola to contour of breast; development of Montgomery's glands and ducts on areola; further pigmentation of areola; nipple generally below midplane of breast tissue when individual is seated upright; maturation independent of breast size	Adult-type hair in quantity and distribution; spread to inner aspects of the thighs in most racial groups

Data from Ross GT. Disorders of the ovary and female reproductive tract. In: Wilson JD, Foster DW, eds. *Textbook of Endocrinology*, 7th ed. Philadelphia: WB Saunders; 1985:206; Speroff L, Glass RH, Kase N. *Clinical Gynecologic Endocrinology and Infertility*, 3rd ed. Baltimore: Williams & Wilkins; 1983:377; and Kustin J, Rebar RW. Menstrual disorders in the adolescent age group. *Primary Care*. 1987;14:139-166.

of a penile urethra are observed in women exposed to androgens during the first 3 months of fetal development (Chapter 233). Significant clitorimegaly in the absence of other signs of sexual ambiguity and in the presence of other signs of virilization requires marked androgenic stimulation and strongly implicates an androgen-secreting neoplasm. The development of the labia minora in postpubertal women indicates the influence of estrogens. Overt anomalies of the distal genital tract and any evidence of obstruction to the escape of menstrual blood should be sought. Under the influence of estrogen, the vaginal mucosa changes during sexual maturation from a tissue with a shiny, bright red appearance with sparse, thin secretions to a dull, gray-pink rugated surface with copious, thick secretions.

The history and physical examination quickly differentiate among several causes of amenorrhea (Table 236-3). The various disorders of sexual differentiation and the other anatomic causes are often apparent on inspection. Distal genital tract obstruction should be identified at the time of pelvic examination even if the specific abnormality is not obvious. The physical stigmata of Turner's syndrome, discussed subsequently, generally make the diagnosis simple. Any sexual ambiguity indicates the need for chromosomal analysis and the measurement of 17α-hydroxyprogesterone to rule out congenital adrenal hyperplasia. Pregnancy and gestational trophoblastic disease may be diagnosed by measurement of HCG. The possibility of intrauterine synechiae or adhesions (Asherman's syndrome) must be considered in individuals in whom amenorrhea develops after curettage or endometritis. Tuberculous endometritis, especially in younger women, may also lead to this disorder. Without hormonal measurements, it may be impossible to distinguish between individuals with chronic anovulation, in whom hypothalamic-pituitary-ovarian function is insufficiently coordinated to produce cyclic ovulation, and those with ovarian failure. However, it is generally possible to form a clinical impression about the cause of the amenorrhea. It can be noted whether the patient has absence of, incomplete, or complete development of secondary sex characteristics. The presence of excess body hair or galactorrhea may provide clinical evidence of the pathogenesis of the amenorrhea. Signs and symptoms of adrenal or thyroid dysfunction may be important as well. Administration of a progestin (typically medroxyprogesterone acetate, 5 to 10 mg given orally for 5 to 10 days, or progesterone in oil, 100 mg given intramuscularly) has been advocated to assess the level of endogenous estrogen. This test is of limited value, however, because almost half the young women with premature ovarian failure experience withdrawal bleeding in response to progestin.

To ascertain whether the outflow tract is intact, an orally active estrogen, such as 2.5 mg of conjugated estrogen daily for 21 days, with 5 to 10 mg of oral medroxyprogesterone acetate for the last 5 to 10 days, is administered. Withdrawal bleeding should occur if the endometrium is normal. Still, hysterosalpingography and hysteroscopy may be required for the diagnosis of Asherman's syndrome because some patients with a normal endometrium may not have a withdraw bleed due to obstruction of the cervical os by scar tissue.

TABLE 236-3 CAUSES OF AMENORRHEA

ANATOMIC CAUSES
Pregnancy
Various disorders of sexual differentiation
Distal genital tract obstruction (müllerian agenesis or dysgenesis)
Gonadal dysgenesis*
Ambiguity of external genitalia (male and female pseudohermaphroditism)
Intrauterine adhesions (Asherman's syndrome)
Gestational trophoblastic disease
CHRONIC ANOVULATION
Due to CNS-hypothalamic-pituitary dysfunction
With inappropriate steroid feedback (e.g., polycystic ovary syndrome)
Due to thyroid or adrenal disorders
OVARIAN "FAILURE"
Menopause
Genetic abnormalities
Physical and environmental causes (e.g., chemotherapeutic agents, irradiation)
Autoimmune disorders
Idiopathic

*Gonadal dysgenesis may be viewed as both a disorder of sexual differentiation and a form of gonadal "failure."
CNS = central nervous system.

Laboratory Findings

Basal levels of FSH, prolactin, and thyroid-stimulating hormone (TSH) should be measured in all amenorrheic and oligomenorrheic women to confirm the clinical impression (Fig. 236-3).[10]

Increased TSH levels with or without increased levels of prolactin imply primary hypothyroidism, and further evaluation for this disorder is indicated (Chapter 226). Although hypothyroidism commonly results in anovulation, amenorrhea occurs in only some hypothyroid women. Menorrhagia and oligomenorrhea may occur as well. The sensitive immunoassays for TSH permit identification of women with hyperthyroidism as well because TSH levels are suppressed in those individuals.

If the prolactin concentration is minimally increased and the TSH level is normal, measurement of the prolactin concentration should be repeated before more extensive evaluation is undertaken because prolactin levels are increased by nonspecific stressful stimuli, sleep, and food ingestion. Prolactin levels may be elevated in as many as one third of women with amenorrhea.

Increased FSH levels (generally >30 mIU/mL) imply ovarian failure and require further evaluation. Incipient ovarian failure should be considered in any woman with basal FSH levels of 15 mIU/mL or higher other than during the midcycle LH surge. Many clinicians believe that chromosomal evaluation is indicated in all individuals with elevated FSH levels before age 40 years, and it is certainly indicated if hypergonadotropic amenorrhea begins before age 30 years.

If FSH levels are low or normal, the measurement of total testosterone levels may be helpful whether or not there is any evidence of hirsutism or virilization. Hyperandrogenic women need not be hirsute because some have relative insensitivity of the hair follicles to androgens. Mildly increased levels of testosterone (and perhaps dehydroepiandrosterone sulfate as well) suggest polycystic ovary syndrome (PCOS). However, total circulating androgen levels need not be elevated because of the alterations in metabolic clearance rate and sex hormone–binding globulin that are present in PCOS. Consequently, some clinicians prefer to measure circulating free testosterone levels.

Circulating levels of LH and FSH may aid in differentiation of PCOS from hypothalamic-pituitary dysfunction. LH levels are often elevated in PCOS so that the ratio of LH to FSH is increased; however, LH levels may be identical to those observed in normal women in the follicular phase. In contrast, levels of LH and FSH are normal or slightly reduced in hypothalamic-pituitary dysfunction. There is some overlap between women with "polycystic ovarian–like" disorders and those with hypothalamic-pituitary dysfunction. Radiographic assessment of the sella turcica is indicated in all amenorrheic women in whom both LH and FSH levels are consistently low (both <10 mIU/mL) to exclude a pituitary or parapituitary neoplasm (Chapter 224). Other pituitary functions should be evaluated in any individual with significantly impaired LH and FSH secretion. Both total testosterone and dehydroepiandrosterone sulfate levels should be measured in hirsute or virilized women. Testosterone levels greater than 200 ng/dL should lead to investigation for an androgen-producing neoplasm, most likely of ovarian origin. Dehydroepiandrosterone sulfate levels greater than 7.0 μg/mL should lead to evaluation for an adrenal neoplasm, and levels between 5.0 and 7.0 μg/mL should lead to evaluation for adult-onset congenital adrenal hyperplasia.

Hypergonadotropic Amenorrhea (Presumptive Ovarian Failure, Primary Hypogonadism, Primary Ovarian Insufficiency)

DIAGNOSIS

Differential Diagnosis

Gonadal failure may begin at any time during embryonic or postnatal development and may result from many causes. Normally, the ovaries fail at menopause, when virtually no functioning follicles remain. However, premature loss of oocytes before the age of 40 years may occur and lead to premature ovarian failure. Circulating gonadotropin levels increase whenever ovarian failure occurs because of decreased negative estrogen feedback to the hypothalamic-pituitary unit.

PATHOBIOLOGY

There are several causes for premature ovarian failure, including genetic causes (a growing list that includes karyotypic abnormalities, single gene

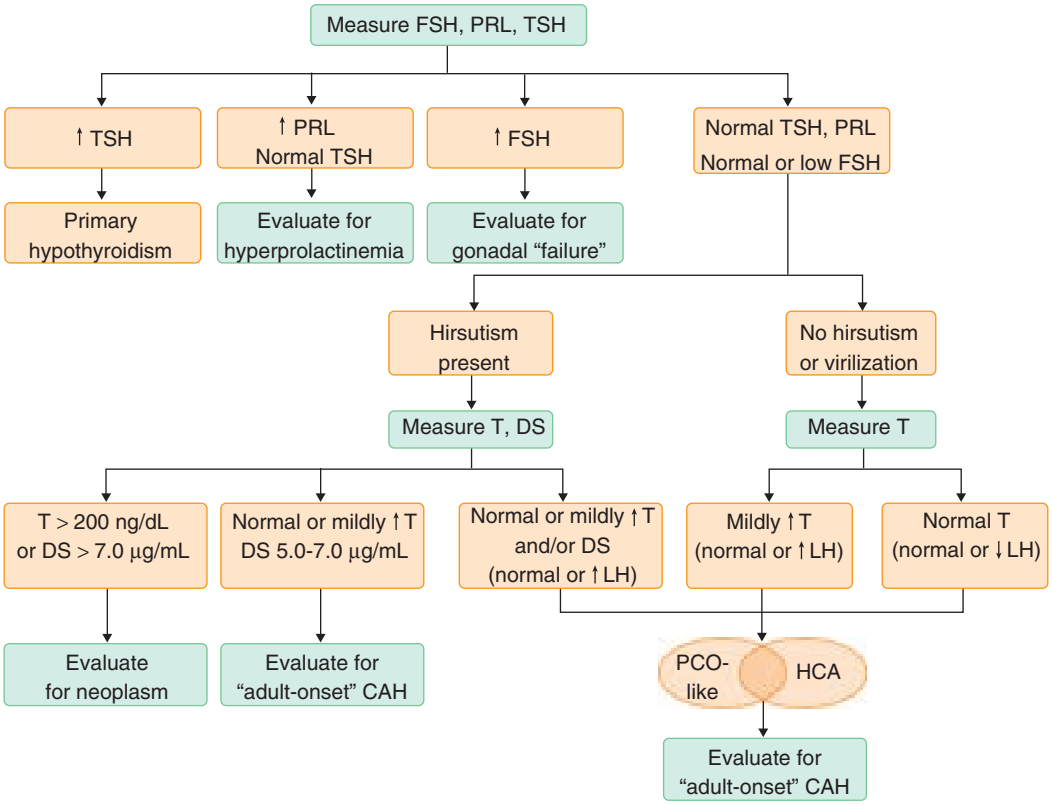

FIGURE 236-3. Biochemical evaluation of amenorrhea. This schema must be considered an adjunct to the clinical evaluation of the patient. See text for details. CAH = congenital adrenal hyperplasia; DS = dehydroepiandrosterone sulfate; FSH = follicle-stimulating hormone; HCA = hypothalamic chronic anovulation; LH = luteinizing hormone; PCO-like = polycystic ovarian syndrome–like; PRL = prolactin; T = testosterone; TSH = thyroid-stimulating hormone.

mutations, and complex multifactorial polygenic inheritance), physical and environmental causes, and autoimmune disturbances. In addition, there may be families in which menopause begins earlier than the expected age without any further pathologic cause.

Genetic Abnormalities

Several pathologic conditions with dysgenetic gonads involve elevated gonadotropin levels and amenorrhea as well as abnormalities of the X chromosome. The term *gonadal dysgenesis* refers to individuals with undifferentiated streak gonads without any association with either extragonadal stigmata or sex chromosome aberrations. Because individuals with gonadal dysgenesis have the normal complement of oocytes at 20 weeks of fetal age but virtually none by birth, this disorder is a form of premature ovarian failure.

Turner's Syndrome

Turner's syndrome (also see Chapter 233) describes patients with streak gonads composed of fibrous stroma and four cardinal features: a female phenotype; sexual infantilism; short stature; and several physical abnormalities, sometimes including webbed neck, low-set ears, multiple pigmented nevi, double eyelashes, micrognathia, epicanthal folds, shieldlike chest with microthelia, short fourth metacarpals, increased carrying angle of the arms, and certain renal and cardiovascular defects (most commonly coarctation of the aorta and aortic stenosis) (Fig. 236-4).[11] The diagnosis can sometimes be made at birth because of unexplained lymphedema of the hands and feet. The syndrome is associated with an abnormality of sex chromosome number, morphology, or both. Most commonly, the second sex chromosome is absent (45,X). Turner's syndrome is the single most common chromosome disorder in humans, but more than 95% of such fetuses are aborted, and the incidence in newborns is approximately 1 in 3000 to 5000. Chromosome breakage and mosaicism occur as well. In mosaic individuals with a normal 46,XX cell line, sufficient follicles may persist postnatally to initiate pubertal changes and to cause ovulation so that pregnancy is possible. Deletions of the X-chromosome-linked *SHOX* gene explain many of the dysmorphic skeletal features that are present, including the short stature. It is believed that the number of phenotypic findings may be related to the percentage of cells that are 45,X. There also may be an effect of imprinting with the variation in phenotype partly explained by the parental origin of the one remaining X chromosome.

Pure Gonadal Dysgenesis

Pure gonadal dysgenesis is the term given to phenotypically female individuals with streak gonads who are of normal stature and have none of the physical stigmata associated with Turner's syndrome. Such individuals have either a 46,XX or 46,XY karyotype. The 46,XX defect may be inherited as an autosomal recessive, with 10% having associated nerve deafness. The 46,XY defect may be inherited as an X-linked recessive, with clitorimegaly occurring in 10 to 15% and gonadal tumors developing in 25% if the gonads are not removed.

Mutations in the X Chromosome Associated with Premature Ovarian Failure

Several regions of the X chromosome are now recognized to contain mutations in genes that may result in premature ovarian failure. Of particular note is the fragile X mental retardation *(FMR1)* gene. More than 5% of women with 46,XX spontaneous premature ovarian failure have mutations of the *FMR1* gene. This risk is increased if there is a family history of premature ovarian failure. A family history of fragile X syndrome, unexplained mental retardation, dementia, developmental delay of a child, or tremor-ataxia syndrome is reason for genetic counseling. Mutations in the *FMR1* gene are known to be associated with a neurodegenerative disorder. Women with mutations in the *FMR1* gene are at risk for having a child with mental retardation, should they be one of the 6 to 8% of women with premature ovarian failure who conceive spontaneously. For *FMR1*, a CGG repeat sequence occurs, with up to 60 repeats being normal. Expansion to more than 200 repeats leads to the fragile X syndrome, with the high level of repeats causing hypermethylation of the gene promoter and silencing of the gene. Female carriers of the permutation have an unstable intermediate number of repeats (i.e., 60 to 199) and the predisposition for premature ovarian failure.

Trisomy X

Trisomy X (46,XXX karyotype) is also associated with premature menopause, although many such individuals have normal reproductive lives. Premature menopause can also occur in mosaic individuals with cell lines with excess X chromosomes. When gonadal abnormalities occur in women with excess X chromosomes, they seem to occur after ovarian differentiation so that some ovarian function is possible. Only later in life do such women develop secondary amenorrhea and premature ovarian failure.

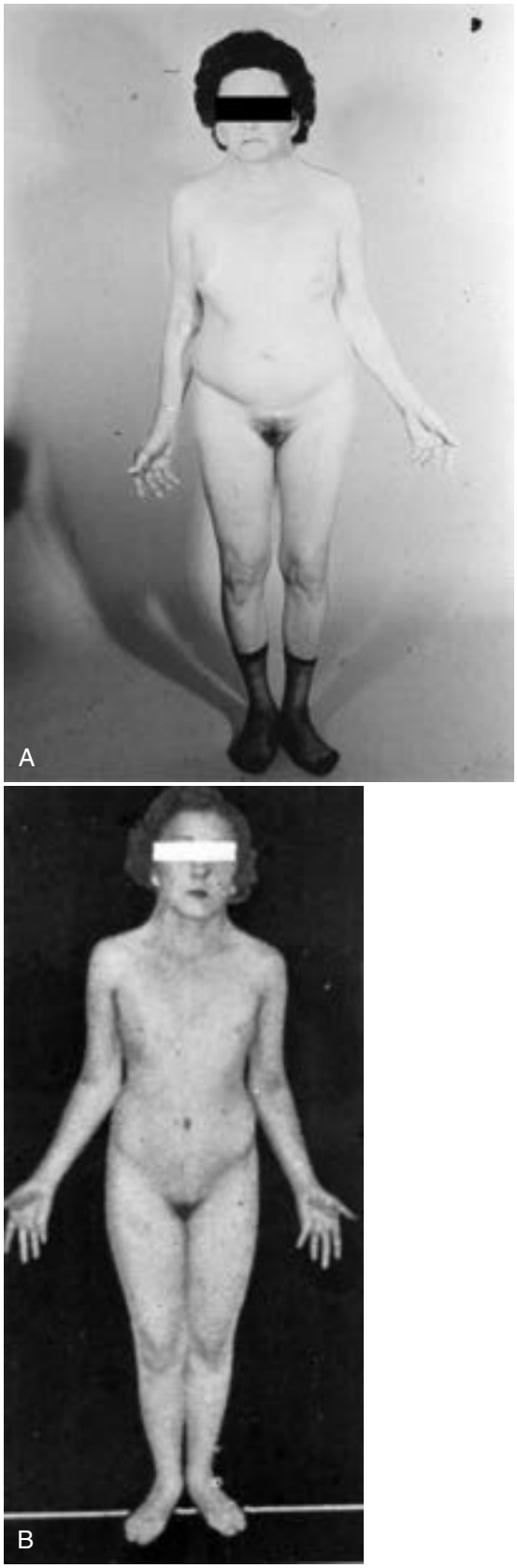

FIGURE 236-4. Adult with Turner's syndrome. This woman was seen at age 56 years by the author of the chapter (Dr. Rebar), and was case number 2, an adolescent at that time, in the original publication of Dr. Henry Turner describing the syndrome.

Known Genetic Alterations of Specific Genes

In girls with the rare syndrome of 17α-hydroxylase deficiency involving *p450c17* who survive until the expected age of puberty, sexual infantilism and primary amenorrhea occur together with elevated levels of gonadotropins (also see Chapter 233). Defects in the 20,22-lyase *(p450scc)* or aromatase *(p450arom)* enzymes may also lead to ovarian failure. Women with galactosemia also experience ovarian failure early in life, even when a galactose-restricted diet is introduced early in infancy.

Mutations of several autosomal genes result in premature ovarian failure. Included in this growing list are mutations involving *FSHR* (the FSH receptor gene), *FOXL2* (a forkhead transcription factor associated with the blepharophimosis-ptosis-epicanthus inversus syndrome), *INHA* (the inhibin-α gene), *E1F2B* (a family of genes associated with central nervous system leukodystrophy and ovarian failure), *PMM2* (the gene for phosphomannomutase), *GALT* (the gene for galactose-1-phosphate uridyltransferase), and *AIRE* (leading to the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy syndrome). Myotonic dystrophy (Chapter 421) is caused by an autosomal triple repeat mutation, like the fragile X syndrome, that is similarly associated with premature loss of germ cells from the ovary. The list of mutations associated with early ovarian failure continues to increase as the function of more genes is determined.

Mutations Involving Reproductive Hormones, Their Receptors, and Action

The resistant ovary (Savage's) syndrome occurs in young amenorrheic women who have elevated peripheral gonadotropin concentrations, normal (although immature) follicles present on ovarian biopsy, 46,XX karyotype with no evidence of mosaicism, fully developed secondary sex characteristics, and ovarian resistance to stimulation with human menopausal or pituitary gonadotropins. At least some of these women have mutations in the FSH receptor. It is probably inappropriate to use the term "resistant ovary syndrome" because it is likely that this is a heterogeneous disorder due to various genetic mutations.

Other Causes

Physical and Environmental

Irradiation and chemotherapeutic agents used to treat various malignant diseases may also cause premature ovarian failure. Ovulation and cyclic menses return in some of these patients even after prolonged intervals of hypergonadotropic amenorrhea associated with signs and symptoms of profound hypoestrogenism. In general, the younger the individual at the time of treatment, the less likely is she to have permanent ovarian failure after the completion of therapy. Rarely, mumps affects the ovaries and causes ovarian failure.

Autoimmune Disorders

Premature ovarian failure may occur in conjunction with a variety of autoimmune disorders. The most well-known syndrome (autoimmune polyglandular syndrome type 1) involves hypoadrenalism, hypoparathyroidism, and mucocutaneous candidiasis together with ovarian failure (Chapter 231). Testing for adrenal antibodies by indirect immunofluorescence will identify the 4% of women with spontaneous premature ovarian failure who have steroidogenic cell autoimmunity and are at risk for adrenal insufficiency. Thyroiditis is the most commonly associated abnormality. Antibodies to the FSH receptor have been identified in a few cases. These associations make it mandatory to rule out other potentially life-threatening endocrinopathies in young women with hypergonadotropic amenorrhea.

TREATMENT

Rx

Women with hypergonadotropic amenorrhea and ovarian failure should be treated identically whether or not they have signs of hypoestrogenism or desire pregnancy. Counseling and psychological support are indicated in women in whom the diagnosis of premature ovarian failure is made. Ovarian biopsy is not indicated to document the existence of follicles because only a small portion of each ovary can be sampled and because pregnancies have resulted in patients who had biopsy samples devoid of follicles. Estrogen replacement is warranted to prevent the accelerated bone loss known to occur in affected women (Chapter 243). The estrogen should be given sequentially with a progestin to prevent endometrial hyperplasia. Young women with ovarian failure may require twice as much estrogen as postmenopausal women for relief of signs and symptoms of hypoestrogenism. Inexplicably, women with premature ovarian failure may conceive while taking exogenous estrogen, even in the form of oral contraceptive agents, at the same rate as those not taking estrogen, so barrier contraception should be discussed if pregnancy is not desired.

Women with hypergonadotropic amenorrhea are rarely able to become pregnant. It is not clear why pregnancy may rarely occur in such women, but

the pregnancy and delivery rate is 6 to 8%. Infertility treatment of young women with hypergonadotropic amenorrhea involves hormone replacement to mimic the normal menstrual cycle and embryo transfer by use of donor oocytes. Whether women with gonadal dysgenesis should be offered pregnancy by use of donor oocytes is now the subject of debate because a markedly increased incidence of aortic rupture during pregnancy secondary to medial necrosis has been documented.

Women with Turner's syndrome contemplating pregnancy should be counseled regarding the risks. The coordination of health care of adult women with Turner's syndrome often falls to the endocrinologist because many of the complications of the disease are endocrinologic: hypothyroidism, diabetes, hypertension, obesity, osteoporosis, and hypogonadism. However, guidelines have been published about the surveillance of other multisystem conditions for which Turner's syndrome patients are at risk, including significant psychosocial problems, congenital heart disease, deafness, and gastrointestinal and hepatic disorders.

CHRONIC ANOVULATION

Chronic anovulation, the most frequent form of amenorrhea encountered in women of reproductive age, implies that functional ovarian follicles remain and that cyclic ovulation can be induced with appropriate therapy (Table 236-4). The cause of the anovulation should be determined. The pathophysiologic bases for several forms of anovulation are unknown, but the anovulation can be interrupted transiently by nonspecific induction of ovulation in most affected women. Anovulation can result in either amenorrhea or irregular (generally less frequent) menses.

Hypothalamic Chronic Anovulation

DEFINITION

Hypothalamic chronic anovulation (HCA) represents a heterogeneous group of disorders with similar manifestations. Emotional and physical stress, excessive exercise, nutritional deficiencies, weight loss, reduced body fat, and other unrecognized factors may contribute in varying proportions to the anovulation. Women with HCA have normal neuroanatomic findings.

TABLE 236-4 CAUSES OF CHRONIC ANOVULATION

- Chronic anovulation of hypothalamic-pituitary origin
  - Hypothalamic chronic anovulation
    - Psychogenic
      - Exercise associated
      - Associated with diet, weight loss, or malnutrition
      - Anorexia nervosa and bulimia
      - Pseudocyesis
  - Forms of isolated (idiopathic) hypogonadotropic hypogonadism (including Kallmann's syndrome)
  - Due to hypothalamic-pituitary damage
    - Pituitary and parapituitary tumors
    - Empty sella syndrome
    - Following surgery
    - Following irradiation
    - Following trauma
    - Following infection
    - Following infarction
  - Idiopathic hypopituitarism
  - Hypothalamic-pituitary dysfunction or failure with hyperprolactinemia (multiple causes)
  - Due to systemic diseases
- Chronic anovulation due to inappropriate feedback (i.e., polycystic ovary syndrome)
  - Excessive extraglandular estrogen production (i.e., obesity)
  - Abnormal buffering involving sex hormone–binding globulin (including liver disease)
  - Functional androgen excess (adrenal or ovarian)
  - Neoplasms producing androgens or estrogens
  - Neoplasms producing chorionic gonadotropin
- Chronic anovulation due to other endocrine and metabolic disorders
  - Adrenal hyperfunction
    - Cushing's syndrome
    - Congenital adrenal hyperplasia (female pseudohermaphroditism)
  - Thyroid dysfunction
    - Hyperthyroidism
    - Hypothyroidism
  - Prolactin or growth hormone excess
    - Hypothalamic dysfunction
    - Pituitary dysfunction (microadenomas and macroadenomas)
    - Drug induced
  - Malnutrition

ANOREXIA NERVOSA

Individuals with amenorrhea and significant weight loss should be examined for the possibility of anorexia nervosa (Chapter 219).

ISOLATED HYPOGONADOTROPIC HYPOGONADISM

Affected individuals have absence of spontaneous pubertal development. Most have functional GnRH deficiency, but some have abnormalities of gonadotropin deficiency localized to the pituitary gland.

Kallmann's syndrome is a familial disorder consisting of gonadotropin deficiency, anosmia or hyposmia, and color blindness in men or, more rarely, in women (Chapter 223). Partial or complete agenesis of the olfactory bulb is present on autopsy, accounting for use of the term *olfactogenital dysplasia.* Isolated gonadotropin deficiency in the absence of anosmia occurs as well. Sexual infantilism with a eunuchoid habitus is the clinical hallmark of this disorder, but moderate breast development may occur. Circulating LH and FSH levels are low but almost always detectable. Mutations in *KAL1*, *FGFR1*, *FGF8*, *PROK2*, *ROKR2*, *HS6ST1*, *WDR11*, or *CHD7* have been identified in a minority of patients with Kallmann's syndrome. Ovulation induction requires use of exogenous gonadotropins and HCG or pulsatile GnRH. Estrogen replacement therapy is indicated in these women until pregnancy is desired. It may not be possible to distinguish between partial isolated gonadotropin deficiency and functional HCA in all cases.

HYPOPITUITARISM

Hypopituitarism may be obvious on cursory inspection or sufficiently subtle to require endocrine testing (Chapter 224). The clinical presentation depends on the age at onset, the cause, and the nutritional status of the individual. Failure of development of secondary sex characteristics must always raise the question of hypopituitarism. Ovulation can be induced successfully with exogenous gonadotropins when pregnancy is desired and after the hypopituitarism is treated appropriately. Replacement therapy with estrogen is indicated.

HYPERPROLACTINEMIA

Galactorrhea associated with hyperprolactinemia, whatever the cause, almost always occurs together with amenorrhea caused by hypothalamic-pituitary dysfunction or failure. Many conditions can cause excess prolactin secretion (Chapter 224). A prolactinoma must be excluded. Hirsutism may be observed occasionally in association with amenorrhea-galactorrhea and hyperprolactinemia. Elevated levels of the adrenal androgens dehydroepiandrosterone and dehydroepiandrosterone sulfate may be observed and may account for the polycystic-type ovaries present in some hyperprolactinemic women.

FAILURE OF THE HYPOTHALAMIC-PITUITARY UNIT

The hypothalamic-pituitary unit may also fail to function normally in a number of stressful, debilitating, systemic illnesses that interfere with somatic growth and development. Chronic renal failure, liver disease, and diabetes mellitus are the most prominent examples.

DIAGNOSIS

Abrupt cessation of menses in women younger than 30 years who have no anatomic abnormalities of the hypothalamic-pituitary-ovarian axis and no other endocrine disturbances suggests a diagnosis of HCA. Affected individuals tend to be bright, educated, and engaged in intellectual occupations and may well give a history of psychosexual problems and socioenvironmental trauma. HCA is characterized by low to normal levels of gonadotropins and relative hypoestrogenism. Rarely, however, do affected women present with signs and symptoms of estrogen deficiency. It is important to rule out a central lesion as the cause of the hypogonadotropic hypogonadism in women who appear to have HCA.

TREATMENT

Rx

Psychological counseling or a change in lifestyle, especially for women engaged in strenuous exercise programs, may be effective in inducing cyclic ovulation and menses in women with functional HCA. Cognitive behavior therapy is effective in a proportion of women with functional HCA. For women

desiring pregnancy, ovulation can also be induced with clomiphene citrate (50 to 100 mg/day for 5 days beginning on the third to fifth day of withdrawal bleeding). Treatment with exogenous gonadotropins to induce follicular maturation followed by HCG to induce follicular rupture may be effective in women who do not ovulate in response to clomiphene. Because women with HCA have low circulating levels of leptin, investigators have given recombinant leptin and documented that ovulation may resume in some affected women. Given the heterogeneous nature of the disorder, it is not surprising that exogenous leptin is not effective in all women.

Most physicians advocate the use of exogenous gonadal steroids to prevent osteoporosis. A regimen can consist of daily oral conjugated or esterified estrogens (0.625 to 1.25 mg), ethinyl estradiol (20 μg), or micronized estradiol-17β (1 to 2 mg) or transdermal estradiol-17β (0.05 to 0.10 mg) daily, with oral medroxyprogesterone acetate (5 to 10 mg) added for the first 12 to 14 days of each month. Sexually active women can be given oral contraceptive agents as an alternative. If steroid therapy is administered, patients must be informed that the amenorrhea will probably be present when therapy is discontinued. Other physicians believe that only periodic observation is indicated, with barrier methods of contraception recommended for fertility control. Adequate ingestion of calcium should be ensured regardless of therapy. Contraception is needed for sexually active women with HCA because the functional defect is mild in these disorders and may resolve spontaneously at any time, with ovulation occurring before any episode of menstruation.

Chronic Anovulation Related to Inappropriate Feedback

DEFINITION

PCOS is a heterogeneous disorder in which there is considerable clinical and biochemical variability among affected individuals. (See PCOS as the most common cause of heterosexual pubertal development in Chapter 235.) PCOS is currently considered to exist in women with any two of the following: (1) oligo-ovulation or anovulation, (2) hyperandrogenism, or (3) polycystic ovaries on ultrasound, and in whom other etiologies have been eliminated. PCOS is the classic disorder in which the amenorrhea or oligomenorrhea results from inappropriate feedback of gonadal steroids from the ovaries.

PATHOBIOLOGY

Current evidence suggests that the hypothalamic-pituitary unit is intact and that a functional derangement, perhaps involving insulin-like growth factors (IGFs) such as IGF-I within the ovary, results in abnormal gonadotropin secretion. PCOS is characterized by insulin resistance and compensatory hyperinsulinemia. (See association between PCOS, insulin resistance, and obesity in Chapter 220.) The insulin resistance has been found in affected women of many racial and ethnic groups, implying that it is a universal characteristic and that a common defect may be present. There is increasing evidence of specific genetic abnormalities in some women with PCOS.

CLINICAL MANIFESTATIONS

Although patients usually present with amenorrhea, hirsutism, and obesity, affected women may instead complain of irregular and profuse uterine bleeding, may not have hirsutism, and may be of normal weight (Fig. 236-5). Excess androgen from any source or increased extraglandular conversion of androgens to estrogens can lead to the typical findings of PCOS. Included are such diverse disorders as Cushing's syndrome, mild congenital adrenal hyperplasia, virilizing tumors of adrenal or ovarian origin, hyperthyroidism and hypothyroidism, obesity, and primary PCOS with no other recognizable cause.

In the primary syndrome, the irregular menses, mild obesity, and hirsutism begin during puberty and typically become more severe with time, although there is increasing evidence of improvement in the years just before menopause. Obesity alone can lead to a polycystic ovarian–like syndrome, with the degree of obesity required to cause anovulation varying widely. The increase in the prevalence of obesity is leading to an increased prevalence of PCOS. All such patients are well estrogenized regardless of whether they present with primary or secondary amenorrhea or dysfunctional bleeding. LH concentrations tend to be elevated, with relatively low and constant FSH levels, but both may be in the normal range for the follicular phase of the menstrual cycle. Levels of most circulating androgens, especially testosterone, tend to be mildly elevated.

DIAGNOSIS

After exclusion of other etiologies, two of the following three are required for diagnosis of PCOS: (1) hyperandrogenism (clinical or biochemical); (2) oligo-ovulation or anovulation; (3) polycystic ovaries on ultrasound examination or at surgery.[12]

This definition is confusing to clinicians because it implies that hirsute women with polycystic ovaries on ultrasound examination who ovulate regularly should be considered to have PCOS. Moreover, it is clear that

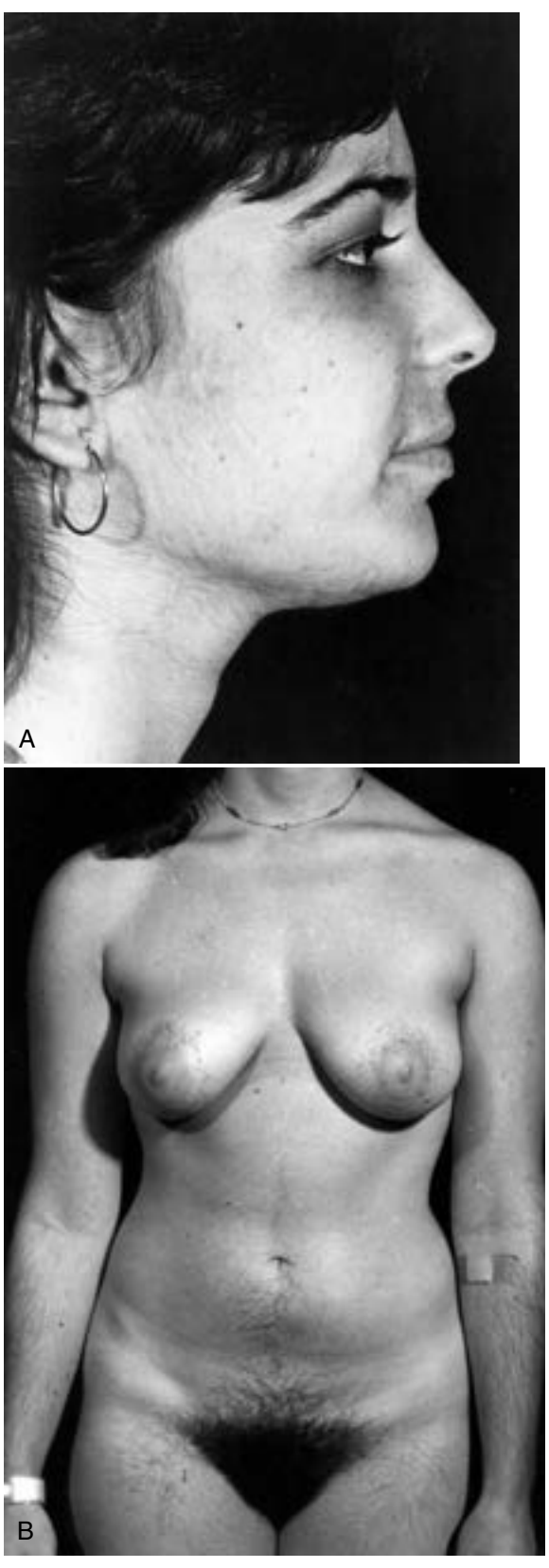

FIGURE 236-5. Adult with polycystic ovary syndrome (PCOS). This 28-year-old woman with documented PCOS had elevated luteinizing hormone levels, irregular menses, and hirsutism since puberty. Note the increased hair in the midline extending up to and above the umbilicus. Other findings (which are not necessarily abnormal) are periareolar hair and hypertrichosis of the arms.

polycystic ovaries may be identified on ultrasound examination in normal women. In any case, the aim of the diagnostic evaluation is to rule out any causes (such as neoplasms) that require definitive therapy. Hirsutism should be evaluated as detailed in Chapter 442.

A particularly severely affected subset of women present with marked obesity, anovulation, mild glucose intolerance with high levels of circulating insulin, acanthosis nigricans, hyperuricemia, severe hirsutism, and elevated circulating androgen levels. These women have hyperthecosis of the ovaries, in which the androgen-producing cells in the stromal, hilar, and thecal regions are increased greatly in number. Hyperthecosis should probably be viewed as a part of the spectrum of disorders constituting PCOS.

TREATMENT

Rx

Patients generally require therapy for hirsutism, for induction of ovulation if pregnancy is desired, and for prevention of estrogen-induced endometrial hyperplasia and cancer. No ideal therapy exists; the therapeutic approach must be individualized. The risks for metabolic syndrome, cardiovascular disease, and diabetes mellitus are increased in women with PCOS, at least in part because of the increased androgens and insulin resistance. Moreover, many women have elevated cholesterol levels.

Medical Therapy

In the anovulatory woman not desiring pregnancy who is not hirsute, therapy with intermittent progestin administration (e.g., medroxyprogesterone acetate, 5 to 10 mg orally for 10 to 14 days each month) or oral contraceptives can be provided to reduce the increased risk for endometrial carcinoma that is present in such a woman with unopposed estrogen. All women using intermittent progestin administration should be cautioned about the need for effective contraception if they are sexually active because these agents do not inhibit ovulation when they are administered intermittently.

Improvements in insulin sensitivity in women with polycystic ovaries, either through lifestyle changes (i.e., exercise and diet) or through pharmacologic intervention, consistently result in improvements in the reproductive and metabolic abnormalities. Resumption of ovulation may occur in up to 60 to 70% of affected women.

The longest and largest published experiences with any agent that improves insulin sensitivity in PCOS is with metformin, a biguanide that functions primarily by suppressing hepatic gluconeogenesis and also improves insulin sensitivity. Its use in PCOS leads to reductions in insulin and androgen levels and resumption of menses in some women. Divided doses of 1500 to 2000 mg/day have proved effective.

Some clinicians advocate giving metformin to all women with polycystic ovaries, whereas others would administer such an agent only to those with documented insulin resistance. Some clinicians also advocate giving metformin first to women who desire pregnancy and then adding an agent to induce ovulation if the metformin proves ineffective. These agents are not approved for use in pregnant women or for the induction of ovulation.

Treatment Considering Pregnancy

Oral contraceptive agents are the first line of therapy for hirsute anovulatory woman not desiring pregnancy and offer protection from endometrial hyperplasia. In women with PCOS desiring pregnancy, clomiphene citrate or letrozole can be used to induce ovulation. Letrozole is not approved for this use by the FDA, but a large multicenter randomized trial has demonstrated its superiority to clomiphene in obese women with PCOS. About 75 to 80% conceive with such therapy. In addition to insulin-sensitizing agents, other possible methods of inducing ovulation include use of exogenous gonadotropins and HCG, and laparoscopic ovarian surgery with multiple punctures of the ovary by diathermy or laser. A large clinical trial documented that clomiphene citrate is more effective than metformin in inducing ovulation and resulting in pregnancy; there was no further improvement when the two agents were used concurrently.

Surgical Treatment

Laparoscopic ovarian surgery can achieve unifollicular ovulation or make it easier for medical ovulation induction but increases the risk for development of ovarian adhesions (themselves leading to infertility). It may be successful in a small subset of women with PCOS who are geographically removed from good medical care.

Chronic Anovulation Related to Other Endocrine and Metabolic Disorders

Adrenal hyperfunction appears to cause chronic anovulation by inducing a polycystic ovarian-like syndrome secondary to increased adrenal androgen secretion. Both hyperthyroidism and hypothyroidism are associated with a variety of menstrual disturbances, including dysfunctional uterine bleeding and amenorrhea as a result of alterations in the metabolism of androgens and estrogens. These metabolic changes in turn result in inappropriate steroid feedback and chronic anovulation.

Luteinized Unruptured Follicle Syndrome

The luteinized unruptured follicle syndrome refers to the development of a dominant follicle without its subsequent disruption and release of the ovum. The abnormality can be diagnosed by ultrasonography or by the absence of evidence of ovulation when the ovary is viewed at laparoscopy. The disorder occurs infrequently and is not a significant cause of infertility. Menstrual cycles in which no ovum is released are characterized by presumptive evidence of ovulation, including biphasic basal body temperatures, secretory endometrium, normal LH surge, and normal progesterone production in the luteal phase.

Luteal Phase Dysfunction

PATHOBIOLOGY

Progesterone secretion in the luteal phase may be reduced in duration (termed *luteal phase insufficiency*) or in amount (termed *luteal phase inadequacy*). More rarely, the endometrium may be unable to respond to secreted progesterone because of the absence of progesterone receptors. These disorders represent causes of infertility (because of inability of fertilized ova to implant) in less than 5% of infertile couples. Abnormalities of the follicular phase, especially in the frequency of gonadotropin pulses, may account for most luteal phase defects. Luteal phase defects may also occur sporadically in normally ovulating women.

DIAGNOSIS

Luteal phase dysfunction may be associated with several clinical entities, including mild or intermittent hyperprolactinemia, strenuous physical exercise, inadequately treated 21-hydroxylase deficiency, and recurrent miscarriage. Luteal dysfunction occurs more commonly at the extremes of reproductive life and in the first menstrual cycles after full-term delivery, abortion, or discontinuation of oral contraceptives. It may also occur during ovulatory cycles induced with clomiphene citrate or exogenous gonadotropins and HCG.

TREATMENT

Rx

Treatment of luteal dysfunction is controversial. Any underlying defect should be treated. If subsequent luteal function depends on prior follicular development, modification of follicular development with either clomiphene citrate (25 to 100 mg daily by mouth for 5 days beginning on cycle day 3 to 5) or FSH (75 to 300 IU intramuscularly for 3 to 5 days beginning on cycle day 3 to 5) can be used; HCG (2500 to 5000 IU intramuscularly at 2- to 3-day intervals beginning with the shift in basal body temperature) or progesterone (12.5 mg intramuscularly in oil daily or 25 mg twice a day as rectal or vaginal suppositories) can be used as well. Bromocriptine may correct the abnormality in individuals with hyperprolactinemia. Synthetic progestational agents should not be used to treat luteal phase defects because of their possible association with congenital anomalies. Furthermore, the synthetic progestins produce an abnormal endometrium. None of these agents has been shown to increase the pregnancy rate.

INFERTILITY

DEFINITION

The World Health Organization (WHO) has defined *infertility* as "a disease of the reproductive system defined by the failure to achieve a clinical pregnancy after 12 months or more of regular unprotected sexual intercourse." *Sterility* is total inability to reproduce. More than 10% of couples in the United States seek medical assistance for infertility.

The requirements for pregnancy to occur are several:

- The male must produce adequate numbers of normal, motile spermatozoa.
- The male must be capable of ejaculating the sperm through a patent ductal system.
- The sperm must be able to traverse an unobstructed female reproductive tract.
- The female must ovulate and release an ovum.

TABLE 236-5 CAUSES OF INFERTILITY AND THEIR APPROXIMATE INCIDENCE (WHERE AVAILABLE)*

Male factors (40%)
Decreased production of spermatozoa
Varicocele
Testicular failure
Endocrine disorders
Cryptorchidism
Stress, smoking, caffeine, nicotine, recreational drugs
Ductal obstruction
Epididymal (after infection)
Congenital absence of vas deferens
Ejaculatory duct (after infection)
After vasectomy
Inability to deliver sperm into vagina
Ejaculatory disturbances
Hypospadias
Sexual problems (i.e., impotence), medical or psychological
Abnormal semen
Infection
Abnormal volume
Abnormal viscosity
Immunologic factors
Sperm-immobilizing antibodies
Sperm-agglutinating antibodies
Female factors
Fallopian tube disease (20-30%)
Pelvic inflammatory disease or puerperal infection
Congenital anomalies
Endometriosis
Secondary to past peritonitis of nongenital origin
Amenorrhea and anovulation (15%)
Minor ovulatory disturbances (<5%)
Cervical and uterine factors (10%)
Leiomyomas and polyps
Uterine anomalies
Intrauterine synechiae (Asherman's syndrome)
Destroyed endocervical glands (after surgery or after infection)
Vaginal factors (<5%)
Congenital absence of vagina
Imperforate hymen
Vaginismus
Vaginitis
Immunologic factors (<5%)
Sperm-immobilizing antibodies
Sperm-agglutinating antibodies
Nutritional and metabolic factors (5%)
Thyroid disorders
Diabetes mellitus
Severe nutritional disturbances
Idiopathic or unexplained (<10%)

*World Health Organization definition of infertility: ≥12 months of regular unprotected sexual intercourse. In about one third of couples, more than one cause contributes to the infertility.

- The sperm must be able to fertilize the ovum.
- The fertilized ovum must be capable of developing and implanting in appropriately prepared endometrium.

In approximately 40% of cases, infertility is caused by the male (Table 236-5). In one third of couples, more than one cause contributes to the infertility.

Peak age for fertility in the female is 25 years. For nulliparous women of this age, the average time during which unprotected intercourse occurs until conception is 5.3 months. For parous women, the average duration of intercourse until conception is 2.7 months. The reproductive performance of couples is influenced by the ages of the female and male partners, the frequency of intercourse, and the length of time the couple has been attempting to conceive. There is a decline in both female and male reproductive performance after the age of 25 years.

DIAGNOSIS

Couples who complain of infertility merit evaluation regardless of the length of infertility. Evaluation is warranted in all women after 12 months and in women 35 years of age or older after 6 months of regular unprotected intercourse.

The evaluation begins with a detailed history obtained from both partners and physical examinations of both individuals. If possible, the couple should be seen together. Each couple should be questioned together and separately because separate interviews may uncover information that would not be imparted in the presence of the partner.

Initial evaluation for infertility includes assessment of semen; documentation of ovulation by basal body temperature, serum progesterone determination 6 to 8 days before menses, serum thyroid hormone, or (rarely) endometrial biopsy less than 3 days before onset of menses; and evaluation of the female genital tract by hysterosalpingography or sonohysterography. Diagnostic laparoscopy with tubal dye instillation may be performed if results of all previous tests are normal because 30 to 50% of women are found to have endometriosis or tubal disease on surgical evaluation; alternatively, patients with initial normal findings may be merely treated as having idiopathic infertility.

TREATMENT Rx

Treatment must be predicated on the findings of the infertility evaluation. Abnormalities of sperm are difficult disorders to treat. Low sperm count or poor motility is best treated either by donor insemination or in vitro fertilization with intracytoplasmic injection of a single viable sperm into each oocyte. Obstruction of the fallopian tubes may be amenable to surgical intervention, but success rates are often greater with in vitro fertilization. Endometriosis causing infertility may be treated by surgery or various suppressive drugs as indicated; however, here, too, in vitro fertilization may be indicated.

Induction of ovulation is one of the most successful therapies when used in anovulatory women.

Induction of ovulation should never be attempted until serious disorders precluding pregnancy are ruled out or treated. Furthermore, ovulation induction should not be used in women with ovarian failure because they are unresponsive to any form of ovulation induction.

Clomiphene citrate is the agent that usually induces ovulation most easily. Clomiphene should be used in individuals without hyperprolactinemia who have the ability to release LH and FSH. A typical course of clomiphene therapy is begun on the third to fifth days after either spontaneous or induced uterine bleeding. The initial dosage is 50 mg daily for 5 days. Clomiphene appears to act as an antiestrogen and stimulates gonadotropin secretion by the pituitary gland to initiate follicular development. If ovulation is not achieved in the first cycle of treatment, the daily dosage is increased to 100 mg. If ovulation is still not achieved, dosage is increased in a stepwise fashion in 50-mg increments to a maximum of 200 to 250 mg daily for 5 days. The highest dose should be continued for 3 to 6 months before the patient is regarded as unresponsive to clomiphene. The quantity of drug and the length of time that it can be used, as suggested here, are greater than those recommended by the manufacturers and the FDA but conform to published series. Despite absence of FDA approval, letrozole is being used increasingly in place of clomiphene.

The ovulatory surge of LH may occur 5 to 12 days (average, 7 days) after the completion of the last day of clomiphene treatment. Couples are advised to have intercourse every other day during this interval. Ovulation can be documented by monitoring changes in basal body temperature or preferably by measuring serum progesterone 14 days after the last clomiphene dose. Menses should occur after 3 weeks. Withdrawal bleeding with progestin can be induced if the patient fails to bleed within 4 weeks of therapy and if a serum HCG level documents that the patient is not pregnant. Testing the urine for an LH surge may also be useful in timing ovulation.

Some clinicians give 5000 to 10,000 IU of HCG intramuscularly 7 days after the last day of clomiphene therapy to trigger ovulation, but this approach has not been established to increase effectiveness. The administration of HCG, however, does serve to time ovulation and may be helpful in selected couples. Ovulation can be expected to occur approximately 36 hours after HCG administration.

Of appropriately selected patients, 75 to 80% ovulate, and 40 to 50% can be expected to become pregnant. About 15% of pregnancies can be expected with each ovulatory cycle. The multiple pregnancy rate is about 8%, with almost all being twins. The incidence of congenital anomalies is not increased.

Side effects of clomiphene are uncommon and rarely serious. The most serious ones include vasomotor flushes (10%), abdominal discomfort (5%), breast tenderness (2%), nausea and vomiting (2%), visual symptoms (1.5%), and headache (1%). Ovarian enlargement may occur but is rare (5%). Concern has been raised about the potential for clomiphene to increase the risk for epithelial ovarian cancer. The bulk of the evidence now indicates that clomiphene does not increase this risk.

The addition of dexamethasone, 0.5 mg orally at bedtime, to blunt the nighttime secretion of adrenocorticotropic hormone may be useful in hyperandrogenic women who fail to ovulate in response to clomiphene. Other individuals who do not respond to clomiphene typically require exogenous gonadotropins and HCG or perhaps pulsatile GnRH to induce ovulation.

Both bromocriptine and cabergoline are effective in inducing ovulation in hyperprolactinemic women. The drug should be stopped when pregnancy is confirmed. Ovulatory menses and pregnancy are achieved in about 80% of patients with galactorrhea and hyperprolactinemia. Most women with prolactin-secreting pituitary tumors remain asymptomatic during pregnancy. It is rare for a patient with either a microadenoma or a macroadenoma to develop a problem related to the tumor that affects either the mother or the fetus during pregnancy. Monitoring during pregnancy need consist only of questioning the patient about the development of visual symptoms and headaches. Formal assessment of visual fields and computed tomography or magnetic resonance imaging should be carried out in any patient experiencing suggestive symptoms. Symptoms generally abate with institution of therapy with a dopamine agonist. No adverse effects of dopamine agonists on fetuses or pregnancies have been reported. Concerns have been raised that ergot-derived dopamine agonists, in the large doses used in the treatment of Parkinson disease, may increase the risk for cardiac valve regurgitation. Although there is no evidence of risk in women treated with much lower doses for hyperprolactinemia, they should be counseled about this potential side effect.

Several preparations of purified and synthetic biochemically engineered gonadotropins for use for induction of ovulation now exist. Synthetic preparations consist entirely of FSH, whereas most purified preparations contain some LH as well. Each vial typically contains 75 IU of gonadotropin. Individuals with gonadotropin deficiency require a preparation containing some LH. Exogenous gonadotropins are typically administered at doses of two to four vials (intramuscularly or subcutaneously, depending on the preparation) for 5 to 12 days to achieve follicular development as monitored by ultrasonography and serum or urinary estradiol concentrations; HCG, 5000 to 10,000 IU, is administered as a single intramuscular dose when follicular maturation is apparent. The HCG should be withheld if more than three follicles mature together. GnRH analogues are now being used to suppress endogenous follicular activity before initiation of therapy with exogenous gonadotropins and continued until HCG is given in older women and those with poor responses to exogenous gonadotropins. Use of the analogues necessitates administration of larger doses of exogenous gonadotropins. Success rates, however, appear to be somewhat improved with this combined therapy. Because of the expense and the complication rate, thorough evaluation should be carried out to exclude other causes of infertility before exogenous gonadotropins and HCG are used. Ovulation can be induced in almost 100% of patients, but pregnancy occurs in only 50 to 70%. There is no increased risk for congenital anomalies with exogenous gonadotropins and HCG. The rate of multiple pregnancies with exogenous gonadotropins and HCG may approach 30%, with 5% being triplets or more.

Ovarian hyperstimulation (*ovarian hyperstimulation syndrome*, or OHSS) is the major side effect and may be life-threatening. The ovaries enlarge remarkably, and multiple follicle cysts, stromal edema, and multiple corpora lutea are present. There is a shift of fluid from the intravascular space into the abdominal cavity with resultant hypovolemia and hemoconcentration. The cause of the ascites is unknown. The most serious complications of OHSS may include thromboembolism, renal failure, adult respiratory distress syndrome, and hemorrhage from ovarian rupture. Treatment is conservative, with monitoring of fluid and electrolyte status. Pelvic examinations should not be performed for fear of rupturing the ovaries. The hyperstimulation generally resolves slowly during about 7 days but lasts longer if the cycle results in pregnancy.

Clomiphene citrate or exogenous gonadotropins together with intrauterine insemination of spermatozoa may be used in women with unexplained infertility as so-called controlled ovarian hyperstimulation (COH). The intent is to stimulate several oocytes to be ovulated, but multiple (sometimes high-order) gestations are a significant risk. A randomized trial has noted that the risk for multiple gestation and the costs are reduced if COH with gonadotropins is not used and patients are advanced immediately to treatment by in vitro fertilization.

Assisted Reproductive Technologies

The assisted reproductive technologies, in which by definition both eggs and sperm are handled outside of the body, are being used commonly to treat infertile couples with tubal disease, endometriosis, oligospermia and azoospermia, sperm antibodies, and unexplained infertility. The procedure consists of in vitro fertilization and several variants. In vitro fertilization involves ovarian hyperstimulation, oocyte retrieval, fertilization, embryo culture, and embryo transfer. Ovarian hyperstimulation with clomiphene citrate and exogenous gonadotropins, gonadotropins alone, or a GnRH agonist or antagonist plus gonadotropins typically causes 1 to 20 oocytes to mature, depending on the patient's age and ovarian "reserve." After follicular growth is judged sufficient by ultrasound examination, HCG is given to induce final follicular maturation. About 34 hours after HCG administration, the oocytes are retrieved by direct needle puncture of each follicle, usually transvaginally with ultrasound guidance. The oocytes are then inseminated in vitro with washed sperm, or a single sperm is injected directly into a single egg (so-called intracytoplasmic sperm injection). The embryos are cultured for about 40 to 120 hours, after which one or more embryos are transferred to the uterine cavity. Embryos may be cultured to the blastocyst stage (at 120 hours) before transfer. Additional embryos can be frozen in liquid nitrogen for transfer in a subsequent natural cycle. The success rate is most dependent on the age of the woman. In the United States, the percentage of cycles resulting in live births ranges from 40.1% in women younger than 35 years to 12.2% in women aged 41 to 42 years. Approximately 30% are twins and 1% are triplets or higher-order multiples.

It is now possible to test the early embryo for genetic abnormalities by removal of either a single cell (i.e., blastomere) or a polar body from the embryo in vitro and testing it with probes by fluorescent in situ hybridization, with the assistance of polymerase chain reaction, or most recently by comparative genomic hybridization. Identification of normal and abnormal embryos allows only normal embryos to be transferred in families with recognized and testable genetic abnormalities.

SEXUAL FUNCTION AND DYSFUNCTION

Sexual Function

DEFINITION

Sexual responses historically have been divided into four phases: excitement, plateau, orgasm, and resolution. With sexual arousal and excitement, vasocongestion and muscle tension increase progressively, primarily in the genital region, manifested by vaginal lubrication in the female. The lubrication is due to formation of a transudate in the vagina. Sexual excitement is initiated by any of a variety of psychogenic or somatogenic sexual stimuli and must be reinforced to result in orgasm. With continued stimulation, the excitement phase increases in intensity into a plateau phase during which a high state of sexual interest is maintained. The plateau phase may be short or long, and it is from this phase that an individual can shift to orgasm. The orgasmic phase tends to be brief and is characterized by rapid release from the developed vasocongestion and muscle tension. The orgasmic release is also known as the climax because peak psychological and physical intensity is achieved and there is an attendant feeling of satisfaction. Copious secretions and transudate may flow during orgasm in women. Characteristic genital and extragenital responses occur during these phases. Estrogens magnify the sexual responses, but responses may occur in estrogen-deficient women. For women, these changes occur in the breasts and in the pudendal region and are variable from one response cycle to another. For some women, excitement proceeds quickly through plateau to orgasm, and orgasm is explosive and accompanied by vocalization and involuntary contractions of the pelvic skeletal muscles. For other women, the responses are slow in building, controlled in amplitude, and long lasting. For a few women, orgasm never occurs; for many, it is intermittently absent.

The somatic sensate focus enabling orgasmic release is variable and may include stimulation of the breasts, vagina, or clitoris. The psychological aspect of coitus may involve concentration on the current partner or act or fantasies about other times and persons. Although orgasms may vary in physiologic intensity, what is important is psychological satisfaction. Satisfaction for both men and women may be had without orgasm.

Many clinicians have noted several limitations of this traditional human sex response cycle. Many clinicians and researchers see the cycle as circular with stimuli of different types leading to arousal. Clinicians in this field now have extended this theory to include desire and arousal. Women seek sexual experiences for intimacy as well as for sexual gratification. Women may be receptive to or seek out sexual stimuli to enhance intimacy. Biologic and psychological factors contribute to the processing of these stimuli and can enhance arousal and desire simultaneously.

Sexual Dysfunction

Women may seek consultation because of disturbances in normal sexual arousal or orgasm.[13] Such sexual dysfunction may be due to either organic or functional disturbances.

A variety of diseases affecting neurologic function, including diabetes mellitus and multiple sclerosis, may prevent sexual arousal. So, too, may local pelvic disorders, such as endometriosis and vaginitis, which cause dyspareunia and lead to sexual avoidance. Estrogen deficiency causing vaginal atrophy and dyspareunia is a relatively common cause of sexual dysfunction. Debilitating systemic diseases such as malignant disease may also affect sexual function indirectly.

In many cases, the cause of sexual dysfunction is psychological.[14] For instance, vaginismus involves involuntary contractions of the muscles

surrounding the introitus and leads to dyspareunia. It is a conditioned response engendered by a previous real or imagined traumatic sexual experience. Feelings of guilt (caused by incest or rape, as examples), of inadequacy (caused by hysterectomy or mastectomy), or of depression or anxiety may lead to failure to be aroused. Failure to achieve orgasm may be viewed as a dysfunction if the woman is frustrated or dissatisfied.

TREATMENT

Rx

Treatment of sexual dysfunction should eliminate functional causes and provide the patient, often together with her partner, with appropriate psychological counseling.[15] Behavioral modification is effective in treating many women with psychological sexual dysfunction. In one randomized trial, self-reported sexual satisfaction was increased in women treated with testosterone.[16] However, dosing guidelines are not clear, and the therapy cannot be considered standard care based on current evidence.

Grade A References

A1. Healey M, Ang WC, Cheng C. Surgical treatment of endometriosis: a prospective randomized double-blind trial comparing excision and ablation. *Fertil Steril*. 2010;94:2536-2540.

A2. Alkatout I, Mettler L, Beteta C, et al. Combined surgical and hormonal therapy for endometriosis is the most effective treatment: prospective, randomized, controlled trial. *J Minim Invasive Gynecol*. 2013;20:473-481.

A3. Guzick DS, Huang LS, Broadman BA, et al. Randomized trial of leuprolide versus continuous oral contraceptives in the treatment of endometriosis-associated pelvic pain. *Fertil Steril*. 2011;95:1568-1573.

A4. Marjoribanks J, Brown J, O'Brien PM, et al. Selective serotonin reuptake inhibitors for premenstrual syndrome. *Cochrane Database Syst Rev*. 2013;6:CD001396.

A5. Ford O, Lethaby A, Roberst H, et al. Progesterone for premenstrual syndrome. *Cochrane Database Syst Rev*. 2012;3:CD003415.

A6. Kjøtrød SB, Carlsen SM, Rsmussen PE, et al. Use of metformin before and during assisted reproductive technology in non-obese young infertile women with polycystic ovary syndrome: a prospective, randomized, double-blind, multi-centre study. *Hum Reprod*. 2011;26:2045-2053.

A7. Roy KK, Baruah J, Singla S, et al. A prospective randomized trial comparing the efficacy of letrozole and clomiphene citrate in induction of ovulation in polycystic ovarian syndrome. *J Hum Reprod Sci*. 2012;5:20-25.

A8. Badawy A, Gibreal A. Clomiphene citrate versus tamoxifen for ovulation induction in women with PCOS: a prospective randomized trial. *Eur J Obstet Gynecol Reprod Biol*. 2011;159:151-154.

A9. Legro RS, Brzyski RG, Diamond MP, et al. Letrozole versus clomiphene for infertility in the polycystic ovary syndrome. *N Engl J Med*. 2014;371:119-129.

A10. Legro RS, Barnhart HX, Schlaff WD, et al. Clomiphene, metformin, or both for infertility in the polycystic ovary syndrome. *N Engl J Med*. 2007;356:551-566.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ju H, Jones M, Mishra G. The prevalence and risk factors of dysmenorrhea. *Epidemiol Rev*. 2014;36:104-113.
2. Vercellini P, Viganò P, Somigliana E, et al. Endometriosis: pathogenesis and treatment. *Nat Rev Endocrinol*. 2014;10:261-275.
3. Harel Z. Dysmenorrhea in adolescents and young adults: an update on pharmacological treatments and management strategies. *Expert Opin Pharmacother*. 2012;13:2157-2170.
4. Dunselman GA, Vermeulen N, Becker C, et al. ESHRE guideline: management of women with endometriosis. *Hum Reprod*. 2014;29:400-412.
5. Rapkin AJ, Akopians AL. Pathophysiology of premenstrual syndrome and premenstrual dysphoric disorder. *Menopause Int*. 2012;18:52-59.
6. Freeman EW, Halbreich U, Grubb GS, et al. An overview of four studies of a continuous oral contraceptive (levonorgestrel 90 mcg/ethinyl estradiol 20 mcg) on premenstrual dysphoric disorder and premenstrual syndrome. *Contraception*. 2012;85:437-445.
7. Deligeoroglou E, Karountzos V, Creatsas G. Abnormal uterine bleeding and dysfunctional uterine bleeding in pediatric and adolescent gynecology. *Gynecol Endocrinol*. 2013;29:74-78.
8. Roberts-Wilson TK, Spencer JB, Fantz CR. Using an algorithmic approach to secondary amenorrhea: avoiding diagnostic error. *Clin Chim Acta*. 2013;423:56-61.
9. Biro FM, Greenspan LC, Galvez MP, et al. Onset of breast development in a longitudinal cohort. *Pediatrics*. 2013;132:1019-1027.
10. Azurah AG, Zainuddin AA, Jayasinghe Y. Diagnostic pitfalls in the evaluation and management of amenorrhea in adolescents. *J Reprod Med*. 2013;58:324-336.
11. Chacko E, Graber E, Regelmann MO, et al. Update on Turner and Noonan syndromes. *Endocrinol Metab Clin North Am*. 2012;41:713-734.
12. Conway G, Dewailly D, Diamanti-Kandarakis E, et al. The polycystic ovary syndrome: a position statement from the European Society of Endocrinology. *Eur J Endocrinol*. 2014;171:P1-P29.
13. Chen CH, Lin YC, Chiu LH, et al. Female sexual dysfunction: definition, classification, and debates. *Taiwan J Obstet Gynecol*. 2013;52:3-7.
14. Bradford A, Meston CM. Behavior and symptom change among women treated with placebo for sexual dysfunction. *J Sex Med*. 2011;8:191-201.
15. Frühauf S, Gerger H, Schmidt HM, et al. Efficacy of psychological interventions for sexual dysfunction: a systematic review and meta-analysis. *Arch Sex Behav*. 2013;42:915-933.
16. Folladi E, Davis SR. An update on the pharmacological management of female sexual dysfunction. *Expert Opin Pharmacother*. 2012;13:2131-2142.

REVIEW QUESTIONS

1. Which of the following is the main biologic mediator that causes menstrual cramping?

A. Follicle-stimulating hormone (FSH)
B. Prostaglandin
C. Estradiol
D. Inhibin
E. Aromatase

Answer: B Prostaglandins increase in concentration during the luteal phase, and their release results in greater and/or prolonged uterine contraction. This excessive contraction results in ischemic pain. Prostaglandin synthesis inhibitors decrease prostaglandin concentration and thereby alleviate dysmenorrhea. FSH stimulates granulosa cell production of estradiol and is predominant before ovulation. Estradiol is a gonadal hormone involved in the development of female secondary sexual characteristics and reproduction but does not induce menstruation or menstrual pain. Inhibin is released from the follicle to downregulate FSH production by the pituitary. Aromatase converts testosterone to estradiol. (See Cyclic Changes in Target Organs: Endometrium; and Harel Z. Dysmenorrhea in adolescents and young adults: an update on pharmacological treatments and management strategies. *Expert Opin Pharmacother.* 2012;13:2157-2170.)

2. Which of the following directly regulates the secretion of gonadotropin-releasing hormone (GnRH)?

A. Endogenous opiates
B. Follicle-stimulating hormone (FSH)
C. Luteinizing hormone (LH)
D. Follistatin
E. Antimüllerian hormone

Answer: A GnRH secretion is regulated by neurotransmitters and neuromodulators, including endogenous opiates (see Neuroendocrine Regulation of the Ovaries). FSH and LH release are regulated by GnRH, but there is no direct feedback mechanism with these molecules. Instead, FSH and LH act on the ovary to produce gonadal hormones, which can impact GnRH release. Follistatin binds to activin and regulates activin's activity on gonadotropin production but does not itself directly regulate GnRH. Antimüllerian hormone is produced by the follicle and can be used as a marker for ovarian reserve but does not regulate GnRH secretion.

3. Endometriosis is most likely to cause which of the following?

A. Menorrhagia
B. Premenstrual syndrome
C. Dyspareunia
D. Ovarian hyperstimulation
E. Amenorrhea

Answer: C Rationale: Endometriosis is defined as the presence of endometrial glands and stroma outside of the uterine cavity, and typical clinical characteristics are infertility and dysmenorrhea. Endometriosis-related peritoneal scarring can also result in dyspareunia and dyschezia. (See Abnormalities of the Reproductive Years: Dysmenorrhea and Endometriosis: Definition.) Endometriosis does not significantly influence menstrual flow and therefore does not cause menorrhagia or amenorrhea. Premenstrual syndrome occurs before menstruation and has not been demonstrated to by caused by endometrium outside of the uterine cavity. Ovarian hyperstimulation occurs typically with exogenous gonadotropin use, resulting in massive enlargement of the ovaries and ascites. It is not caused by endometriosis.

4. Among women with hypergonadotropic amenorrhea (premature ovarian failure or insufficiency), what is the lifetime likelihood of delivering a live-born progeny using their own oocytes?

A. Same as in normal population
B. 1 in 2
C. 1 in 5
D. 1 in 15
E. Less than 1 in 10,000

Answer: D The lifetime likelihood of delivering a live-born child using their own (e.g., autologous) oocytes is 6 to 8% in this patient population. (See Hypergonadotropic Amenorrhea: Treatment.) As a result, women with this condition should not be managed expectantly, regardless of age. Standard therapies directed at inducing ovulation (clomiphene citrate, letrozole, or gonadotropins) are also not indicated because these women are already exposed to high gonadotropins and are unlikely to respond to exogenous sources. Should such patients desire pregnancy, the most effective intervention requires the use of donor oocytes.

5. Which of the following is the infertility treatment most likely to cause triplet pregnancy?

A. Clomiphene citrate
B. Letrozole
C. Bromocriptine
D. Controlled ovarian stimulation with intrauterine insemination
E. Controlled ovarian stimulation with in vitro fertilization

Answer: D Clomiphene citrate induces ovulation by inhibiting estradiol-mediated feedback, thereby increasing endogenous gonadotropins. Although twinning is common (8%), triplets or more are rare. Letrozole is an aromatase inhibitor that decreases estradiol production and thereby decreases estradiol-mediated negative feedback on gonadotropins in a manner similar to clomiphene. Again, there is minimal evidence of high-order multiple pregnancies. Bromocriptine is used in hyperprolactinemic women to induce normal ovulation, and multiple pregnancies are rare. When exogenous gonadotropins are used without oocyte retrieval, there is a risk for fertilization of multiple mature oocytes, resulting in a twin rate of 30% and a higher order multiple rate of 5%. (See Infertility: Treatment.) When oocytes are retrieved, and the number of embryos transferred to the uterus is limited, the risk for higher order multiples drops to 1%.

237

APPROACH TO WOMEN'S HEALTH

KAREN FREUND

An approach to the care of women must go beyond an understanding of differences in the incidence of disease between men and women. Providers need to consider the impact of sex differences (those based on genetic and hormonal differences) and gender differences (those attributable to the roles men and women are ascribed in society). Therapeutic decisions should take into account both genetic and environmental differences in the presentation of disease and the effectiveness of therapeutic options, the patient's reproductive life stage, comorbidities, and social and cultural contexts of care. Empirical evidence about care for women has expanded since 1994, when the National Institutes of Health required the inclusion of women as research subjects. The increased number of women participating in clinical research, coupled with initiatives through the Office on Research and Women's Health, has led to a broad expansion of knowledge; however, there is a continued need for gender-specific analyses of data to evaluate the impact of any therapeutic intervention.

LIFESPAN GROUPS

Many important women's health issues are linked to the social, psychological, and biologic context at certain ages and stages of life. When considering both preventive care and common causes of mortality and morbidity, these lifespan stages provide a context for organizing care (Table 237-1).

For most women, young adulthood (15 to 44 years) is marked by social transitions in family structure, such as forming one's own family and parenting, and entering the work force. Mortality rates are low, and health visits can focus on behavioral decisions that will influence the risk for future disease, such as those pertaining to sexual behavior, smoking, alcohol and drug use, diet, and exercise. Major sources of morbidity include intentional and unintentional injury, including interpersonal violence (Chapter 241) and motor vehicle crashes. HIV is a leading cause of morbidity and mortality in this age group. Depression and anxiety are common in this and all life stages. Reproductive issues are considered in most therapeutic decisions.

The middle years (ages 45 to 65 years) continue to be influenced by behavioral decisions, especially diet, exercise, and alcohol and substance use. The social context includes role changes as children reach adulthood; caregiving responsibilities are for dependent children and possibly grandchildren, as well as aging parents. The menopausal transition may be accompanied by new symptomatic concerns. Common causes of morbidity, including diabetes and obesity, now reflect earlier behavioral decisions. Cancer is the leading cause of mortality.

Health issues in older women (65+ years) may occur in the context of loss of function and independence, and because women commonly survive their male partners, they are more likely at this stage of life to be single and possibly more isolated. Cardiovascular disease is the major cause of mortality, followed by cancer, cerebrovascular disease, chronic obstructive lung disease, and pneumonia. Loss of independence is related to cognitive decline, osteoarthritis, osteoporotic fractures, and incontinence.

HEALTH DISPARITIES AMONG WOMEN

It is critical to consider the racial and ethnic differences in outcomes for most common causes of mortality and morbidity when addressing the health status of women. American women with a minority racial or ethnic affiliation, including those of African descent (whether born in the United States or abroad), women from many Asian and Pacific Island nations, Native Americans, and women of Latino background, share poorer outcomes for a wide variety of conditions. This broad finding of poorer outcomes across many diverse ethnic and racial groups points away from specific genetic differences in these populations and toward social determinants of health. Minority racial and ethnic affiliation is correlated with lower educational attainment, lower income, residence in neighborhoods with higher crime and more environmental health hazards, less access to comprehensive health insurance, and less access to care even when insured. Common recommendations for health promotion, including a diet low in fats and processed sugars and high in whole grains, fruits, and vegetables, along with regular exercise, such as walking, may be difficult to follow in high-crime neighborhoods or in those without markets offering a variety of affordable, nutritious food options. Barriers to health care access and poor adherence to medical therapy are more common in low-income women. For instance, they may have difficulty scheduling appointments that do not interfere with their work schedules or unpaid time off from clerical or service jobs. The need to care for dependent children or elders may interfere with women's ability to address their own health care needs. Lack of health literacy and cultural barriers can create additional barriers (Chapter 5).

COMMON CAUSES OF MORTALITY IN WOMEN

Cardiovascular Disease

Cardiovascular disease (CVD; Chapter 52) is the overall leading cause of death in women. However, because CVD-related death occurs most often in women older than 65 years, its impact is often under-recognized or underestimated. Racial ethnicity plays a large role in the risk for CVD; African American women have much higher rates of CVD and death from CVD than any other racial or ethnic group, and Latinas and Asian women also have higher rates of CVD and CVD death than white populations.

The incidence rate of CVD in women lags 10 years behind that of men from ages 40 through 70 years; that is, a 65-year-old woman has a similar risk to a 55-year-old man. There is no abrupt increase in CVD risk at menopause in women, suggesting that menopausal changes in estrogen or progesterone do not account for this sex difference. Furthermore, randomized clinical trials have confirmed that hormone therapy in women does not prevent CVD and is not indicated for CVD prevention. Re-analysis of the Women's Health Initiative confirmed the increased risk for CVD *with* hormone therapy, even when treatment is initiated within 10 years of menopause.[1] There has been some controversy about the impact of calcium supplementation for bone and health on increased risk for myocardial infarction, but not CVD deaths. However, most observational studies do not find an association in women, and calcium and vitamin D supplementation are still recommended for women regardless of CVD risk (Chapter 243).[2-4] Other risk factors for CVD in women are the same as those in men: elevated lipids, lack of physical activity, obesity, smoking, hypertension, and diabetes. Women with diabetes have the same risk for CVD as men of the same age with diabetes. Because smoking rates in current cohorts of young women continue to increase, cigarette smoking continues to be an important behavioral risk factor for CVD, especially in younger women, American Indian women, and women with low incomes and low educational attainment. The incidence of many known risk factors, including obesity, hypertension, and hypercholesterolemia, is greater in African American women than in white women. Furthermore, data suggest that women are less likely to receive risk factor reduction therapy, reflecting an inertia caused by misperception by clinicians and patients of heart disease risk.

The presentation of coronary disease in women differs from that in men. Although chest pain is the most common presentation in women, atypical and noncardiac pain is a more common presentation in women than in men, and fewer women have typical chest tightness or pressure as their presenting complaint. For this and other reasons, there are delays in seeking and provision of emergency care at all steps from home to arrival to hospital for women compared with men, which can limit some therapeutic options and increase the severity of disease complications, including congestive heart failure.

The guidelines for the management of lipid disorders (Chapter 206) do not differ per se by gender, but do account for gender differences in the risk models used to guide therapy. The decision to begin statin therapy is based on the absolute low-density lipoprotein (LDL) level as well as age and the presence of diabetes. High intensity statin therapy is recommended for women with clinical vascular disease under age 75 years, for primary prevention when LDL is greater than 190 mg/dL, or for women with a 10 year calculated risk of atherosclerotic cardiovascular disease greater than 7.5%.

Medical management is similar for men and women with acute coronary syndromes, including unstable angina and acute myocardial infarction (Chapter 72). The use of aspirin, β-blockers, angiotensin-converting enzyme (ACE) inhibitors, heparin, and thrombolytic therapy and recommendations for noninvasive testing are the same for women and men. Randomized controlled trial data show that low-risk women with unstable angina or non–ST segment elevation myocardial infarction do not benefit from early revascularization and that medical management is indicated.

TABLE 237-1 IMPORTANT HEALTH ISSUES FOR WOMEN THROUGH THE LIFESPAN

	AGES		
ISSUE	***15-44 yr***	***45-65 yr***	***65+ yr***
Behavioral issues	Risk behaviors Sexual behavior Smoking Alcohol and drug use Exercise Diet	Risk behaviors Smoking Alcohol and drug use Exercise Diet	Risk behaviors Smoking Exercise
Social roles	Entering work force Relationship transitions Parenting	Caregiving to several generations Transitions in family and work environments	Losses and social isolation
Reproductive issues	Reproductive health issues	Menopause transition	
Injury	Intentional and unintentional interpersonal violence Motor vehicle crashes		Falls
Common causes of mortality and morbidity	Depression Anxiety HIV/AIDS	Cancer Obesity Diabetes Depression Anxiety	Cardiovascular disease Cancer Cognitive decline Osteoporosis Osteoarthritis Incontinence Depression Anxiety

Type 2 Diabetes

The rates of type 2 diabetes (Chapter 229) continue to rise, with a greater risk in women who are overweight or obese and physically inactive. Rates are higher in many racial and ethnic groups than among white women, including African Americans, Asians, Native Americans, and Latinas. No randomized clinical trials to date on low- to moderate-risk populations have shown a mortality benefit of screening, and most guidelines do not currently recommend universal screening. Screening women for known risk factors (e.g., hypertension [blood pressure >130/85 mm Hg], obesity, family history) is recommended by some. For women who develop gestational diabetes (Chapter 239), the risk of developing type 2 diabetes later in life increases five-fold; therefore, screening with hemoglobin A_{1C} 6 to 12 weeks postpartum and every 3 years thereafter is recommended by some.

The incidence of diabetes is rising in women at younger ages, with significant perinatal implications. Ongoing discussion of fertility control and family planning is critical in diabetic women with childbearing potential. Tight glycemic control before conception and through the first trimester is critical to reduce the risk for birth anomalies, and it is important later in pregnancy to reduce the risk for adverse fetal events, including macrosomia (large for gestational age) and preterm birth (Chapter 239). Ideally, women planning a pregnancy should switch from oral agents to insulin, with home glucose monitoring for tight control (Table 237-2).

Cancer

Cancer is the leading cause of death in women 40 through 65 years, and it is the leading cause of years of life lost in women younger than 65. Breast cancer (Chapter 198) is the most common type of cancer and the second leading cause of cancer death in women, although most women who develop breast cancer do survive. Studies indicate that many women significantly overestimate their personal risk for the disease. The lack of understanding of the cause of most breast cancers and the limitations of the imaging modalities used as screening tests have hampered efforts at breast cancer control. Recent data suggesting overdiagnosis of breast cancer, especially in women 40 to 50 years of age, has prompted growing interest in joint decision making regarding screening in this age group.[5] After decades of an increasing incidence of breast cancer in the United States, incidence rates have fallen in the past 12 years. The cause of this decline is not understood; reduced use of postmenopausal hormone therapy and recent declines in screening rates have both been implicated.

Lung cancer (Chapter 191) is now the leading cause of cancer death in women, and rates continue to rise, commensurate with the number of women who took up smoking in the 1940s and 1950s and the lower rates of smoking cessation in women compared with men. Most lung cancers are smoking related, although women are more likely than men to have lung cancers that are not tobacco related. Low-dose chest tomography has been shown to reduce death from lung cancer in those with greater than a 30 pack-year smoking history who are current smokers or quit within the previous 15 years; the harms of are false-positive scans and potential overdiagnosis.■ The National Lung Screening Trial enrolled 41% women but has not at this time provided any gender-specific analyses. An analysis based on quintiles of lung cancer risk, including sex, showed female sex associated with a lower risk for cancer. This subanalysis suggests that women may benefit from screening less than men, independent of pack years of smoking. Because of the harms of screening, smoking cessation efforts continue to be recommended as the most effective strategy to prevent lung cancer.

Colorectal cancer (Chapter 193) has a similar incidence in both women and men. Testing for occult blood in the stool using fecal immunochemical test, sigmoidoscopy, and colonoscopy is effective at detecting precancerous lesions and thereby reducing rates of both new cancers and later stage disease.

Ovarian cancer (Chapter 199) is relatively rare, with 21,500 cases in the United States annually, but mortality is high owing to its presentation at late stages in most cases. Lower abdominal or pelvic pain, urinary symptoms, and changes in bowel habits are nonspecific and common in many women, limiting the ability to detect this cancer at an early stage. There is no effective screening test for women at average risk, and screening pelvic examination is no longer recommended in asymptomatic women.[6] CA-125 has both low sensitivity and low specificity, with normal levels in early-stage disease and elevations with nonmalignant ovarian pathology. One randomized clinical trial of average-risk women showed no benefit of mortality and increased harms in terms of false-positive studies and complications due to operative procedures.■ While awaiting the results of another large randomized controlled trial, current recommendations are to consider genetic counseling and testing for women with a family history of ovarian cancer.

The greatest recent advance in cancer prevention is the development of vaccines against the most carcinogenic subtypes of human papillomavirus (HPV) associated with cervical cancer (Chapters 199 and 373) and anal cancer. The vaccine has the potential to reduce significant morbidity caused by the management of premalignant lesions, including the risk for cervical incompetence and preterm labor. Vaccine rates in the United States are low, especially in minority communities. Despite this, there is evidence of reduced HPV prevalence since the vaccine became available. Although concern of tacit approval resulting in increased sexual activity among young adolescents has been cited by parents as a reason to delaying vaccination, evidence does not demonstrate this concern to be founded. Guidelines now suggest delaying Pap test screening in women until age 21 years and increasing the screening interval to every 2 years in women aged 20 to 29 years and to every 3 to 5 years in women aged 30 years and older if previous screening has demonstrated no high-risk subtypes of the HPV and negative HPV serology.[7] Pap tests are not recommended for women after hysterectomy for nonmalignant indications, or for women older than 65 years with adequate recent screening and no high-risk factors.

Osteoporosis

Hip fracture due to osteoporosis (Chapter 243) is one of the major causes of disability, loss of independence, and mortality in older women. Osteoporosis prevention through calcium and vitamin D intake and weight-bearing exercise begins at puberty and extends throughout adulthood. Women require a calcium intake of 1000 mg/day, increasing to 1300 mg in puberty and breastfeeding and 1200 mg after menopause, in order to maintain the structural strength of bone. Vitamin D is a necessary element for the absorption of calcium (Chapter 218), and it is available through direct sun exposure or vitamin D supplementation of milk (not most other dairy products) and some juices. Epidemiologic data suggest widespread vitamin D deficiency in women in the United States. This is attributed to low levels of dietary replacement, decreased sun exposure with the use of sunscreens, and lack of vitamin D production in the skin (even with sun exposure) in northern climates during the winter months. Screening for a single baseline serum 25-hydroxyvitamin D level can provide women with useful information about the adequacy of their diets, with levels above 20 ng/mL considered

TABLE 237-2 PREFERRED MEDICATIONS FOR WOMEN OF REPRODUCTIVE POTENTIAL

COMMON CONDITIONS IN WOMEN OF REPRODUCTIVE POTENTIAL	ISSUES TO BE AWARE OF	PREFERRED MEDICATIONS OR GROUPS OF MEDICATIONS
Depression	Must consider effects of untreated depression on mother and infant	Avoid newer medications when older drugs with more information are available SSRIs are generally considered safe; fluoxetine has most safety data
Anxiety	Commonly associated with depression in women	Benzodiazepines generally considered safe
GERD	No data on harmful effects of either H_2-blockers or PPIs Misoprostol is contraindicated; can cause miscarriage, fetal death, congenital anomalies	Calcium-containing antacids are first-line therapy H_2-blockers preferred over PPIs
Acne	Oral isotretinoin and topical tazarotene are contraindicated owing to congenital defects	Most other topical agents are considered classes B and C
Asthma	Extensive data on safety of common drug categories: benefit ratios far exceed risk in treating women	Cortisone inhalers Systemic prednisone for flares Short- and long-acting bronchodilators
Seizure disorders	Difficult to separate effects of medication from effects of seizure on fetal development All agents associated with some increased risk for fetal abnormality (4-8%, compared with 1-2% in general population). Valproate and phenobarbital with highest risks	Monotherapy at lowest doses recommended Avoid medication changes in first trimester Folate supplementation in preconception period
Hypertension (not preeclampsia)	ACE inhibitors and ARBs are contraindicated in pregnancy; possibly associated with cardiovascular and neurologic anomalies in first trimester; may cause abnormalities in renal hemodynamics in third trimester Most data suggest thiazide diuretics are safe if stable use before pregnancy	β-blockers, especially labetalol, and methyldopa are first-lines choices Calcium-channel blockers are also considered safe
Lipid disorders	Circulating lipid levels are elevated with pregnancy and breast-feeding.	No medications during pregnancy; ideally, stop statins before conception
Analgesics for fever and pain management	Controversy about whether NSAIDs slightly increase risk for miscarriage in first trimester	Acetaminophen preferred; aspirin also considered safe
Headache	Controversy about whether NSAIDs slightly increase risk for miscarriage in first trimester	Acetaminophen preferred; aspirin also considered safe
Diabetes	First-generation sulfonylureas contraindicated in pregnancy; may cause fetal hyperinsulinemia and birth defects	Conversion to insulin during planned preconception period Insulin, metformin, and glyburide preferred in women of childbearing potential
Autoimmune conditions	Must weigh benefits of immunosuppressive use against potential risks to both mother and infant Methotrexate is contraindicated in pregnancy because it induces miscarriage No data on safety or risk for TNF and IL-1 inhibitors	Prednisone generally considered safe in pregnancy Azathioprine, sulfasalazine, cyclosporine, and hydroxychloroquine generally preferred if required
Tobacco control	Cigarette smoking has known harmful effects on fetus In one small randomized controlled trial, nicotine replacement was associated with reduced levels of nicotine and improved birth outcomes Bupropion is linked with reports of some fetal anomalies	Try nonpharmacologic approaches to cessation first Short course of nicotine replacement likely better for fetus than smoking
Polycystic ovarian disease	Metformin can restore ovulatory cycles; use with contraception	
Bacterial infections	Tetracyclines accumulate in fetal bone and teeth Sulfa drugs may increase risk for neural tube defects with use in first trimester and kernicterus with use in third trimester Trimethoprim interferes with folic acid metabolism Streptomycin and kanamycin associated with bilateral deafness	Penicillins, cephalosporins, erythromycin, azithromycin

ACE = angiotensin-converting enzyme; ARB = angiotensin receptor blocker; GERD = gastroesophageal reflux disease; IL-1 = interleukin-1; NSAID = nonsteroidal anti-inflammatory drug; PPI = proton pump inhibitor; SSRI = selective serotonin re-uptake inhibitor; TNF = tumor necrosis factor.

sufficient or ideal. Daily replacement of vitamin D of 600 IU is recommended for women, with 800 IU recommended after age 70 years.

Although no long-term outcome data exist on the benefits of dual-energy x-ray absorptiometry (DXA) screening (DEXA bone densitometry), most guidelines recommend DXA screening at age 65 years in women, and in women ages 50 to 65 year who have one risk factor (smoking, family history, body mass index < 22, or alcohol use). An algorithm to assess risk for bone fracture based on DXA and individual risk factors is available through the World Health Organization and can guide decision making on preventive therapy.[8] Long-term estrogen therapy is not recommended for osteoporosis prevention. Complications from bisphosphonate therapy include gastroesophageal erosions, preventable in most women with weekly doses on an empty stomach while sitting upright for 30 minutes. Rare but debilitating jaw osteonecrosis (Chapter 248) has been seen in women with and without risk factors, such as dental disease. Raloxifene, a selective estrogen receptor modulator, has the benefits of preventing osteoporosis while reducing breast cancer risk without increasing the risk for CVD; it has been underused as a preventive agent.

COMMON CAUSES OF MORBIDITY IN WOMEN

Obesity

Obesity rates continue to rise, and the prevalence of obesity is higher in women than men, especially among minority and low-income women (Chapter 220). The rates of increase in obesity in women have doubled from 1976-80, when it was at 17%, to the rate of 35% in 2007-10. Dietary reduction in calories and increased caloric output with aerobic exercise are the short- and long-term strategies for care. Studies suggest that no single diet is superior to others, and many common diet strategies can reduce weight. Increasing activity during the course of one's daily routine is as effective as shorter intervals of more strenuous activity. The most effective programs include a

combination of behavioral therapy, either individually or in groups, to address behavioral patterns in food intake, and with diet and exercise. Weight loss targets of 1 to 2 lb/week and a total loss of up to 5 to 10% of body weight are realistic goals. Even modest changes in weight and modest increases in physical activity reduce morbidity and mortality on a population basis.

Exercise and diet changes are difficult for many people to achieve, for a wide variety of reasons. The physical environment may not be conducive to physical activity owing to the sedentary nature of most workplaces, the lack of safe areas to exercise, and architectural design features, such as the lack of easy access to stairs instead of elevators. The use of prepared foods and the consumption of fast foods, high in fat and calories and low in nutritional content, increase the risk for obesity.

There are currently no evidence-based interventions to support weight reduction in the setting of a brief office visit. Most guidelines focus on assessing body mass index (BMI) in all women and recommending weight reduction with diet restriction and exercise, including behavioral therapy to support these behavioral changes. Orlistat and sibutramine are U.S. Food and Drug Administration (FDA)-approved medications with moderate weight reduction efficacy that should be used in conjunction with an exercise and diet program, and not as sole therapy. For obese women with a BMI greater than 40, or those with a BMI greater than 35 and major comorbidities such as diabetes, sleep apnea, or osteoarthritis, bariatric surgery (Chapter 220) is indicated after other weight loss methods have failed. Bariatric surgery leads to short- and long-term benefits in multiple comorbidities and reduced mortality because of improvements in comorbidities. Critical to the success of this intervention is a team approach, with psychological assessment and diet and exercise programs beginning before and continuing after surgery. Management of obesity in general is discussed in detail in Chapter 220.

Depression

Major depression and other related disorders, including dysthymia, predominate in women (Chapter 434). Multiple short screening tools have been developed to identify depression. Women with depression are highly likely to come to the attention of the health care system. Depression is a significant comorbidity in many chronic medical conditions. Somatic complaints are a common presentation of depressive disorders. All specialties are likely to see patients with unexplained symptoms, such as chest pain, headache, abdominal pain, and other complaints. It is critical to include depression as either a primary or a secondary diagnosis and to treat depression as part of the overall management plan.

Anxiety Disorders

Anxiety disorders also predominate in women. They commonly coexist with depressive disorders. Anxiety disorders, including post-traumatic stress disorder, may be a consequence of violence against women, which often remains unidentified (Chapter 241). Benzodiazepines have been demonstrated to be safe and effective in trials for short-term use. Selective serotonin re-uptake inhibitors (SSRIs) and behavioral therapies are effective for the long-term management of anxiety disorders.

Osteoarthritis

Osteoarthritis (Chapter 262) is one of the most common causes of morbidity and functional status limitation, especially as women age. Assessment of pain and functional status is at the core of management. Physical and occupational therapy to restore functional status is a critical component of care. Joint replacement (Chapter 276) should be considered and recommended when functional status interferes with activities of daily living and supportive care and other symptom management strategies are ineffective.

Smoking

Tobacco use, most commonly cigarette smoking, continues to be one of the major preventable causes of morbidity and mortality in women (Chapter 32). Although there has been much progress in smoking cessation efforts, lower income women and younger women continue to start smoking and fail to quit smoking at high rates. Health care providers can achieve 1 to 2% smoking cessation rates by asking all patients about their smoking status and making a simple statement encouraging smokers to quit. Further gains in smoking cessation are possible with targeted counseling, which involves assessing the patient's stage of readiness to change and providing counseling relevant to that stage.

Nicotine replacement is equally effective in women and men. Smokers with a physiologic addiction to nicotine (those who smoke more than one pack daily or smoke within 20 minutes of awakening) benefit the most from nicotine replacement. Many women report weight gain as a major barrier to smoking cessation. Chapter 32 describes approaches to smoking cessation in detail.

Alcohol Use and Substance Use

Alcohol dependence is estimated in 5% of women; however, this is underrecognized in clinical practice (Chapter 33). It is well established that lower amounts of alcohol cause alcohol-related liver and other disease in women compared with men and increases the risk for breast cancer. Also worrisome is the frequency of binge drinking (defined as four drinks or more at a sitting) by women, with reports that one in eight women drink with this pattern, with risk for poor judgment in personal safety.

Although women in general have lower rates of most substance abuse than men, they have similar rates of nonmedical use of narcotic medications (about 2% of population), a problem that has more than doubled in the past decade, and narcotic overdose deaths have increased five-fold in women in the same timeframe. Current recommendations include specific training for all prescribing physicians; in addition, all nononcology patients who are prescribed more than 30 days of narcotics should be part of a narcotics program, with signed consent on risks and benefits, agreement to obtain medications from a single practice source, and agreement to monitoring, including random pill counts and drug testing, for the presence of the prescribed medication and the absence of other medications.[9]

Alcohol abuse and dependence and drug abuse and dependence are described and discussed in detail in Chapters 33 and 34, respectively.

Incontinence

Urinary incontinence (Chapter 26) is a frequently overlooked cause of major functional status limitations in middle-aged and older women. As many as half of affected women underreport this problem to their physicians and alter their lifestyles to adapt to the problem, including reducing fluid intake, avoiding activities that exacerbate the problem, and restricting travel where access to facilities is uncertain. There are two broad categories of incontinence—stress incontinence and urge incontinence—although women commonly have aspects of both types. Stress incontinence is defined as leaking with increases in intra-abdominal pressure, such as occurs with sneezing or coughing as well as running or walking. The most common reasons are pelvic floor laxity, often from childbirth. Kegel exercises are frequently recommended but are of limited value. A number of surgical procedures are available to address this condition. Less invasive procedures may be tried first, including the fitting of a vaginal pessary or periurethral injections with biodegradable materials such as collagen or with nonbiodegradable materials. Urge incontinence, described as detrusor muscle instability, results in the urge to void with low volumes. Anticholinergic medications and bladder training are both effective. Urodynamic evaluation should be considered if the history does not clearly identify a cause. The problem of incontinence is discussed in more detail in Chapter 26.

HIV Disease

The risk factors for HIV disease in women are heterosexual contact in 80% of new cases and injection drug use in 20%. HIV continues to affect minority women disproportionately, with 61% of incident cases affecting African American women. Incidence rates remain lower in women than in men, with only 27% of incident cases affecting women; in large part, this is due to the fact that male-to-male sexual contact continues to account for 72% of new cases in men. However, the absolute number of new infections from heterosexual contact is twice as high in women as it is in men. Women with HIV/AIDS continue to have poorer survival than men, despite the availability of antiretroviral therapies. There are no data of differential effectiveness of therapy by gender, and there are no gender-specific recommendations regarding timing and type of antiretroviral therapy. Some data suggest that women are less likely than men to adhere to an antiretroviral therapy regimen. The gender difference in therapy adherence was associated with caring for dependent children in one study, suggesting that women's caregiving roles may be a barrier to their own care. Cervical and anal dysplasia and cancer are more common among women with HIV than those without. Therefore, annual Pap tests are recommended for all women with HIV infection, even in the absence of positive HPV testing. There is insufficient evidence to support anal screening for dysplasia, but providers should be aware of the increased risk and perform an external examination for evidence of lesions.

Prophylaxis and management of complications of HIV/AIDS are discussed in detail in Chapter 389, and other aspects of HIV/AIDS are covered in individual chapters in Section XXIV.

REPRODUCTIVE HEALTH ISSUES

All providers caring for women with reproductive potential should consider the reproductive implications of preventive and therapeutic decisions. With half of all pregnancies in the United States unplanned, providers should routinely inquire about contraceptive practices and consider these in their care plans.

All primary care providers should be comfortable counseling patients about contraceptive choices and prescribing oral contraceptives. The absolute contraindications to oral contraceptives are a personal history of CVD, thromboembolic disease, migraine headache with aura, and gynecologic or breast cancer. A family history of cancer is not considered a contraindication; in fact, data suggest that the use of oral contraceptives decreases the risk for both endometrial and ovarian cancer. Smoking while using oral contraceptives increases the risk for thromboembolic events in all women, but especially in those older than 30 years. Oral contraceptive use is considered safe in nonsmoking women until menopause. Although factor V Leiden and other thrombophilias (Chapter 176) have been associated with an increased risk for deep vein thrombosis in those taking oral contraceptives, the absolute risk to any woman is still very low; therefore, screening for this and other genetic thrombophilias is not indicated. Preexisting hypertension is a relative contraindication to oral contraceptive use. Some women develop elevated blood pressures on oral contraceptives; therefore, blood pressure should be monitored at 3 months after starting the drug and then annually.

Providers should consider the reproductive implications of all chronic medications in women of reproductive potential (Chapter 239). Given that the teratogenic effects of medications may occur during the first trimester and before an initial obstetric assessment, the principle when choosing chronic medications for women during their reproductive years is to select those with the greatest safety profile during the first trimester of pregnancy.[10] Table 237-2 outlines common drug categories and recommendations for use in pregnancy.

Antidepressant medications deserve particular attention because of the conflicting data regarding their use in pregnancy.[11] Some initial reports suggested that antidepressants, especially paroxetine, were associated with congenital defects, preterm birth, and neonatology mortality. However, in studies that are able to assess whether prescriptions were filled or not and account for severity of depression and other risk factors of birth defects, specifically smoking, the risks from antidepressants are no greater than in the overall population. For women who wish to take no medication during pregnancy, the recommendation is to gradually reduce the dosage over the course of several weeks and not to abruptly stop taking the medication. The risk for untreated depression during pregnancy and the risk for postpartum depression to the woman and her infant are substantial. Therefore, treatment goals should be to provide adequate and even increased dosing to prevent the worsening of depression during this period and to provide close surveillance of women, whether or not they stop antidepressant medications.

Grade A References

A1. Rossouw JE, Anderson GL, Prentice RL, et al; Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288:321-333.

A2. O'Donoghue M, Boden WE, Braunwald E, et al. Early invasive vs conservative treatment strategies in women and men with unstable angina and non–ST-segment elevation myocardial infarction: a meta-analysis. *JAMA*. 2008;300:71-80.

A3. Kovalchik SA, Tammemagi M, Berg CD, et al. Targeting of low-dose CT screening according to the risk of lung-cancer death. *N Engl J Med*. 2013;369:245-254.

A4. Buys SS, Partridge E, Black A, et al. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *JAMA*. 2011;305:2295-2303.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Mosca L, Benjamin EJ, Berra K, et al. Effectiveness-based guidelines for the prevention of cardiovascular disease in women—2011 update: a guideline from the American Heart Association. *Circulation*. 2011;123:1243-1262.
2. Li K, Kaaks R, Linseisen J, Rohrmann S. Associations of dietary calcium intake and calcium supplementation with myocardial infarction and stroke risk and overall cardiovascular mortality in the Heidelberg cohort of the European Prospective Investigation into Cancer and Nutrition study (EPIC-Heidelberg). *Heart*. 2012;98:920-925.
3. Van Hemelrijck M, Michaelsson K, Linseisen J, et al. Calcium intake and serum concentration in relation to risk of cardiovascular death in NHANES III. *PLoS ONE*. 2013;8:e61037.
4. Xiao Q, Murphy RA, Houston DK, et al. Dietary and supplemental calcium intake and cardiovascular disease mortality: the National Institutes of Health-AARP diet and health study. *JAMA Intern Med*. 2013;173:639-646.
5. Bleyer A, Welch HG. Effect of screening mammography on breast cancer incidence. *N Engl J Med*. 2012;367:1998-2005.
6. Qaseem A, Humphrey LL, Harris R, et al. Screening pelvic examination in adult women: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. 2014;161:67-72.
7. Saslow D, Solomon D, Lawson HW, et al; ACS-ASCCP-ASCP Cervical Cancer Guideline Committee. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *CA Cancer J Clin*. 2012;62:147-172.
8. FRAX WHO Risk Assessment Tool. http://www.shef.ac.uk/FRAX/; Accessed February 7, 2015.
9. Vital signs: overdoses of prescription opioid pain relievers and other drugs among women—United States, 1999–2010. *MMWR Morb Mortal Wkly Rep*. 2013;62:537-542.
10. Hernández-Díaz S, Smith CR, Shen A, et al; North American AED Pregnancy Registry; North American AED Pregnancy Registry. Comparative safety of antiepileptic drugs during pregnancy. *Neurology*. 2012;78:1692-1699.
11. Stephansson O, Kieler H, Haglund B, et al. Selective serotonin reuptake inhibitors during pregnancy and risk of stillbirth and infant mortality. *JAMA*. 2013;309:48.

REVIEW QUESTIONS

1. Which of the following is the reason to consider teratogenicity of medication in 30-year-old women who are not currently pregnant?

A. Half of all pregnancies in the United States are unplanned, putting the fetus at risk.
B. Many medications have long half-lives and can affect the fetus even if stopped at the time of conception.
C. Almost all medications are known to be unsafe in the first trimester.
D. Women in their 30s are less likely than younger women to identify that they are pregnant in the first weeks of pregnancy.

Answer: A Half of all pregnancies in the United States are unplanned, including those in women in their 30s and 40s. Most women will not begin prenatal care until 12 weeks' gestation, after which major organogenesis has occurred. It is therefore incumbent on the prescribing physician to consider the impact of the medications they prescribe in any patient of childbearing potential.

2. The decision to undergo screening mammography in women aged 40 to 49 years should be shared decision making because of all except which of the following?

A. Breast cancer is less common than lung cancer is this age group.
B. Overdiagnosis of breast cancer may result in detecting cancers that would not require treatment.
C. Women in their 40s have a higher false-positive rate of screening studies than women older than 50 years.
D. Women tend to overestimate their breast cancer risk and benefit from education and discussion of risks and benefits.

Answer: A U.S. Preventive Services Task Force guidelines recommend shared decision making for women of average risk aged 40 to 49 years for screening mammography, given the potential harms of both a higher false-positive rate and overdiagnosis. Several studies indicate that normal-risk women in this age group tend to overestimate their personal breast cancer risk.

3. Management of a non–ST elevation acute coronary syndrome (ACS) in women should differ from men in which of the following ways?

A. Medical management is preferred over revascularization in women but not in men.
B. Noninvasive testing is not recommended in women.
C. Aspirin is not indicated for emergent treatment of ACS because of bleeding risk.
D. β-Blockers should be avoided because of increased risk for congestive heart failure.
E. Lipid management goals are low-density lipoprotein cholesterol levels of less than 120 mg/dL.

Answer: A Medical management of ACS and non–ST elevation myocardial infarction is the same for men and women, with aspirin, β-blocker, thrombolytic therapy, noninvasive testing, and cardiovascular rehabilitation. Randomized trials suggest that medical management has better outcomes than early revascularization in women with non–ST elevation myocardial infarction (MI) only, but not in men, one place where gender differences in guidelines exist. Early revascularization is recommended for women and men with ST elevation MI.

4. A woman taking fluoxetine for depression is now planning pregnancy. In counseling her, you would advise which of the following?

A. If possible, she should switch to sertraline.
B. She should wait until the pregnancy test is positive, then stop the medication.
C. An increase in dose should be considered if depressive symptoms increase in pregnancy.
D. To avoid fetal abnormalities, she should stop the medication before attempting conception.

Answer: C Although several studies have suggested a possible association between selective serotonin reuptake inhibitor (SSRI) use in pregnancy and fetal anomalies, these associations usually disappear in studies that are able to control for other causes of fetal abnormalities such as smoking, and in those that are able to ascertain that the SSRI is the only prescription used. Fluoxetine appears to have fewer concerns than sertraline in the available studies. Because of drug distribution in pregnancy, increased doses in pregnancy may be required. The benefits of treated depression and the risks for depression to the woman and her child, including evidence of some developmental delays, must be weighed against any small increased risk for fetal anomalies.

5. Screening for problem alcohol use in women should consider all except which of the following?

A. Binge alcohol drinking is common in women.
B. Younger women have an alcohol threshold (amount and duration) for complications that is lower than men.
C. Women develop complications from alcohol abuse with lower intake for shorter time periods.
D. Rates of alcohol dependence in women are about 5%.

Answer: B Although women have lower rates of alcohol dependence (5%) compared with men (10%), they develop complications with lower amounts of alcohol and with short length of abuse. Binge drinking is common not only among young adults but also among those older than 30 years, and it puts women at risk for violence and poor choices, including sexually transmitted diseases and sexual assault.

238

CONTRACEPTION

BEVERLY WINIKOFF AND DANIEL GROSSMAN

CONTRACEPTIVE USE

Contraception enables men and women to avoid unwanted fertility by preventing pregnancy. Methods can be classified in many different ways. Some classification schemes distinguish among mechanisms (e.g., barriers to the encounter of sperm and ovum versus methods that prevent ovulation); other categories emphasize the timing of use (at the time of intercourse versus ongoing); still other classifications focus on the permanence of the method (sterilization, which is intended as a permanent method, long-acting methods that last for years, and short-term methods that depend on the behavior of the user periodically, every day or at every exposure to pregnancy). There are advantages and disadvantages of each contraceptive method. These advantages and disadvantages should be thoroughly explained so that the individual or couple will choose the most acceptable method that suits their lifestyles and will be used most effectively. Because medical contraindications to individual methods are uncommon among young women, in most cases the choice of contraceptive method depends most on the user's preferences.

In the United States, about 62 million women are in the reproductive age group (15 to 44 years), and about 38 million (62%) are using a method of contraception.[1] Of the remainder, most were either noncontraceptively sterile (about 2%), pregnant or trying to conceive (8%), or never sexually active or had no recent sexual activity (19%). About 8% of women were sexually active in the prior 3 months but were not using a method of contraception. About 50% of U.S. pregnancies are unintended,[2] and more than half of such pregnancies occur in women who are not practicing contraception.[3]

In the United States in 2006 to 2010, the most common methods of fertility prevention were oral contraceptives (OCs) and female sterilization, used by 17.1% and 16.5% of women aged 15 to 44 years, respectively.[4] Next in frequency of use was the male condom (10.2%), followed by male sterilization (6.2%). The injectable progestin was used by 2.4%. However, the intrauterine device (IUD), the most effective method of reversible contraception, was used by only 3.5%, although this is about three times higher than the proportion using an IUD in 2002. Between 1982 and 2006 to 2010, there was a marked decrease in diaphragm use and an increase in condom use.

Contraceptive Effectiveness

Despite an increased use of contraceptive methods by U.S. women since 1982, about half of pregnancies are unintended, meaning they are either mistimed or unwanted. Of the 6.6 million pregnancies that occurred in the United States in 2006 (the most recent data available), 49% were unintended, and 43% of these pregnancies were terminated by elective abortion. Of the women with an unintended pregnancy, 50% stated they were using a method of contraception in the month they conceived. In recent years, teen pregnancy rates have declined, and women aged 18 to 24 years have the highest unintended pregnancy rates. Poor women and those who are unmarried and living with a partner have significantly higher rates of unintended pregnancy compared with women with higher incomes and those who are married or not living with a partner. Research has identified a variety of factors associated with nonuse of contraception or gaps in use, including side effects (both experienced and feared), not liking a method, personal or religious reasons, and barriers to access, including difficulty obtaining a prescription or the method itself, and high cost of a method.

The most pressing need worldwide is that the highly effective contraceptive methods already available should be affordable to most of the population and also that these methods should fulfill the needs of women of different ages and with different reproductive requirements.[5]

Perfect Use versus Typical Use

The terms *perfect use* and *typical use* describe different aspects of the effectiveness of the various contraceptive methods. Perfect use refers to use of the method as intended and covering all acts of exposure to pregnancy. So, for user-dependent methods such as oral pills or condoms, for example, this measure can only be applied to situations in which the user reliably uses the

method every day or at every act of intercourse. Perfect use is a measure of the maximum possible efficacy of the method. *Typical use,* on the other hand, is a measure of how effective methods are when used as a group of people under study actually use them. These rates may be considerably lower than perfect use rates, especially if there is the possibility of not using the method at every intercourse or in other ways not using the method as intended. Methods used at the time of coitus have higher failure rates than OCs, implants, injections, IUDs, and sterilization. IUDs, implants, and sterilization have lower failure rates than pills, patches, and injections because they act over a long period of time, and there is nothing a user needs to do to keep using them. Table 238-1 illustrates the differences in failure rates among the contraception methods under conditions of perfect and typical use.

Cumulative failure rates for use of long-acting methods are low. The effectiveness of long-acting reversible contraception is superior to that of contraceptive pills, patch, or ring and is not altered in adolescents and young women.[6] The cumulative failure rates of all types of tubal sterilization are 1.31% during the first 5 years after the procedure and 1.85% after 10 years; rates are higher for tubal fulguration and lower for segmental resection. The cumulative pregnancy rate for 5 years' use of the levonorgestrel-releasing IUD is 0.5%, and for 10 years' use of the copper T380 IUD, it is 1.7%.

The World Health Organization has developed a chart that nicely represents the actual use (typical) failure rates of most contraceptives, dividing them into three main classes of effectiveness (Figure 238-1).

Contraindications, Risks, and Benefits

Most contraceptives can be used safely by most people, but some conditions or concomitant medications are considered contraindications to use. The U.S. Centers for Disease Control and Prevention have developed evidence-based Medical Eligibility Criteria (MEC) for contraceptive use, based on a similar document developed by the World Health Organization.[7,8] The MEC categorizes conditions or medications into four groups for each contraceptive method: 1—no restriction on use; 2—advantages of use generally outweigh theoretical or proven risks; 3—theoretical or proven risks generally outweigh advantages of using the method; and 4—the condition represents an

TABLE 238-1 PERCENTAGE OF WOMEN EXPERIENCING AN UNINTENDED PREGNANCY DURING THE FIRST YEAR OF TYPICAL USE AND THE FIRST YEAR OF PERFECT USE OF CONTRACEPTION AND THE PERCENTAGE CONTINUING USE AT THE END OF THE FIRST YEAR—UNITED STATES

	WOMEN EXPERIENCING AN UNINTENDED PREGNANCY WITHIN THE FIRST YEAR OF USE (%)		
METHOD	**TYPICAL USE***	**PERFECT USE†**	**WOMEN CONTINUING USE AT 1 YEAR (%)‡**
No method§	85	85	—
Spermicides‖	28	18	42
Fertility awareness–based methods¶	24	—	47
Standard-days method	—	5	—
Two-day method	—	4	—
Ovulation method	—	3	—
Symptothermal method	—	0.4	—
Withdrawal	22	4	46
Sponge			
Parous women	24	20	36
Nulliparous women	12	9	—
Condom**			
Female	21	5	41
Male	18	2	43
Diaphragm††	12	6	57
Combined pill and progestin-only pill	9	0.3	67
Evra patch	9	0.3	67
NuvaRing	9	0.3	67
Depo-Provera	6	0.2	56
Intrauterine devices			
Paragard (copper containing)	0.8	0.6	78
Mirena (levonorgestrel releasing)	0.2	0.2	80
Implanon	0.05	0.05	84
Female sterilization	0.5	0.5	100
Male sterilization	0.15	0.10	100
Lactational amenorrhea method‡‡	—	—	—

*Among typical couples who initiate use of a method (not necessarily for the first time), the percentage who experience an accidental pregnancy during the first year if they do not stop use for any other reason. Estimates of the probability of pregnancy during the first year of typical use for spermicides and the diaphragm are taken from the 1995 National Survey Growth (NSFG) and corrected for underreporting of abortion; estimates for fertility awareness-based methods, withdrawal, the male condom, the pill, and Depo-Provera are taken from the 1995-2002 NSFG corrected for underreporting of abortion.

†Among couples who initiate use of a method (not necessarily for the first time) and who use it perfectly (both consistently and correctly), the percentage who experience an accidental pregnancy during the first year if they do not stop use for any other reason.

‡Among couples attempting to avoid pregnancy, the percentage who continues to use a method for 1 year.

§The percentage who become pregnant in the second and third columns are based on data from populations in which contraception is not used and from women who cease using contraception to become pregnant. Among such populations, approximately 89% become pregnant within 1 year. This estimate was lowered slightly (to 85%) to represent the percentage who would become pregnant within 1 year among women not relying on reversible methods of contraception if they abandoned contraception altogether.

‖Foams, creams, gels, vaginal suppositories, and vaginal film.

¶The ovulation and 2-day methods are based on evaluation of cervical mucus. The standard-days method avoids intercourse on cycle days 8 to 19. The symptothermal method is a double-check method based on evaluation of cervical mucus to determine the first fertile day and evaluation of cervical mucus and temperature to determine the last fertile day.

**Without spermicides.

††With spermicidal cream or jelly.

‡‡This is a highly effective, temporary method of contraception. However, to maintain effective protection against pregnancy, another method of contraception must be used as soon as menstruation resumes, the frequency of duration of breast-feeds is reduced, bottle-feeds are introduced, or the baby reaches age 6 months.

Adapted from Trussell J. Contraceptive failure in the United States. *Contraception.* 2011;83:397-404.

Effectiveness of Family Planning Methods

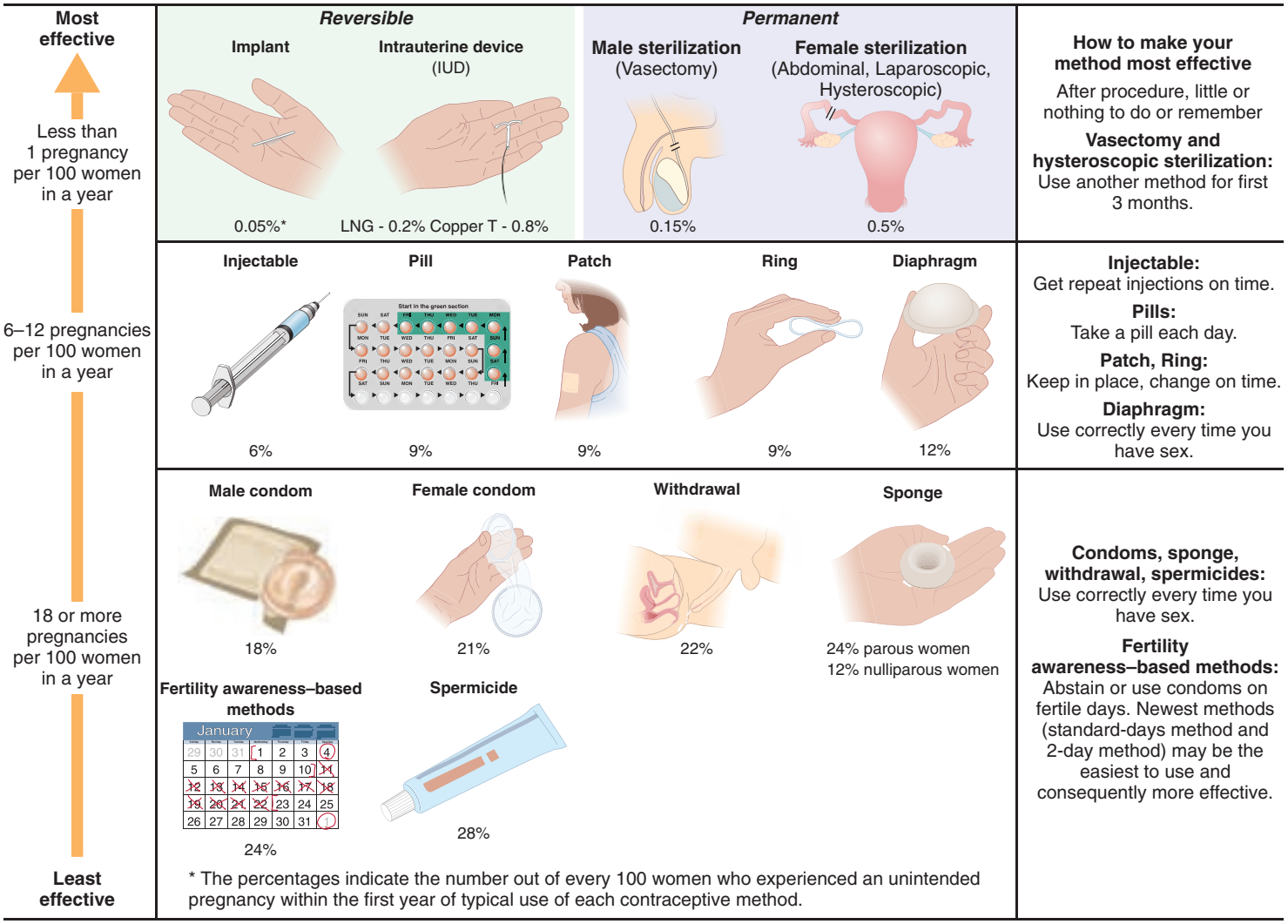

Condoms should always be used to reduce the risk for sexually transmitted infections.

Other Methods of Contraceptions

Lactational Amenorrhea Method: LAM is a highly effective, temporary method of contraception.

Emergency contraception: Emergency contraceptive pills or a copper IUD after unprotected intercourse substantially reduces risk for pregnancy.

FIGURE 238-1. Detailed chart of the effectiveness of family planning methods with instructions for subjects. (Adapted from World Health Organization (WHO) Department of Reproductive Health and Research, Johns Hopkins Bloomberg School of Public Health/Center for Communication Programs (CCP), Knowledge for Health Project. *Family Planning: A Global Handbook for Providers.* 2011 updates. Baltimore, Geneva: CCP and WHO; 2011, and Trussell J. Contraceptive failure in the United States. *Contraception.* 2011;83:397-404.)

unacceptable health risk if the method is used. For conditions that represent relative contraindications (MEC category 3), it is important to recognize that pregnancy may also be risky, and if the method is the best choice to avoid an unintended pregnancy, it may be worth the risk.

In addition to preventing unintended pregnancy, many contraceptive methods have additional benefits. Some of the most significant noncontraceptive benefits include the reduced risk for transmission of HIV and other sexually transmitted infections associated with the use of male and female condoms, as well as the reduction in dysmenorrhea and menorrhagia associated with the use of combined hormonal contraception.

TYPES OF CONTRACEPTIVES

Natural Methods

Methods that rely on the natural infertility of different times in the menstrual or life cycle are often called "natural" methods. These methods are not really more natural than other methods, because they involve disruptions in the "natural" desire for sexual intimacy; however, they do not rely on any specific external technologies to create a state of low fertility potential. Because male sperm can live only 5 days in the female genital tract and female ova have a lifespan of only about 24 hours, the window for fertilization is only 5 to 6 days each month. In theory, if couples avoid unprotected intercourse on those 5 to 6 days, the potential for pregnancy is markedly reduced. The calendar method counts the days of the cycle to predict fertile and infertile days, and the symptothermal method relies on the calendar plus the biologic signals of impending ovulation (changes in vaginal mucus) and of ovulation itself (rise in basal body temperature) to enhance prediction of "safe" days for intercourse. Breast-feeding in the postpartum period also lowers fertility and is considered another natural method of contraception. Exclusive breast-feeding during the first 6 months after a birth provides a good level of protection against pregnancy. However, once a baby is older than 6 months or if foods other than breast milk become part of the infant's diet, then a woman is at much higher risk for ovulation and possible pregnancy.

Withdrawal, or coitus interruptus, when the penis is removed from the vagina before ejaculation, is a commonly used method, with up to 60% of women reporting ever using it. Although the method can be effective with perfect use, with a pregnancy rate of only 4% in the first year of use, failure is much more common in typical use. A recent study reported that among U.S. women aged 15 to 24 years using withdrawal as their primary method, 21% experienced an unintended pregnancy, which was a significantly higher pregnancy rate than that among users of other contraceptive methods.

Barriers

Barrier methods are so called because their mechanism of action is to impose a chemical or physical barrier between ovum and sperm so that fertilization

is not possible. Barrier methods include spermicides, diaphragm, and male and female condoms, among others.

Spermicides

All spermicidal agents contain a surfactant (in U.S. products this chemical is nonoxynol-9) that immobilizes or kills sperm on contact. The spermicidal products are available in foams, creams, and vaginal suppositories that need to be placed into the vagina before each coital act—and reapplied even if coital acts follow immediately after each other. The typical-use effectiveness of spermicides as a sole contraceptive is among the lowest of modern methods (about 28% of women using the method report an unintended pregnancy during 1 year of use.) There is no increased risk for birth defects in the offspring of women who conceive while using spermicides.

Diaphragm

A diaphragm is a dome-shaped latex or silicone device with a flexible rim that seals off the upper genital tract from contact with deposited semen. It is usually used with a spermicide applied inside and around the rim. The device must be fitted by a health care provider using the largest size that does not cause discomfort or undue pressure on the vagina. The woman will need to insert the diaphragm before every act of intercourse. It should be left in place for 6 hours after intercourse, but should not be left in place for more than 24 hours because it may cause ulceration of the vaginal epithelium. In actual use, the diaphragm is more effective than other barrier methods (12% unintended pregnancy rate in first year of use) and provides contraceptive protection almost as well as hormonal pills in actual use. Diaphragm size for any one woman may change after a birth, miscarriage or abortion after 14 weeks, abdominal/pelvic surgery, or a change in weight of more than 20%. The diaphragm may provide some protection against gonorrhea and *Chlamydia* infection, but not against HIV and herpes. A condom is recommended if infection is a concern. Diaphragm users have an increased risk for urinary tract infection.

Male Condom

The male condom is one of the oldest contraceptives known. It is safe, easy to use, and widely available. The modern version is a stretchy latex or plastic sheath that fits over the erect penis and captures the man's ejaculate during intercourse. It is the most effective way to prevent transmission of infections (including HIV) during sex and can be used during vaginal, oral, or anal intercourse. However, natural membrane condoms (made from sheep intestine) do not prevent sexually transmitted infections. Condoms can be used for protection against infection even if another method is used for protection against pregnancy. If the condom is used alone for contraception, in typical use about 18% of women will experience an unintended pregnancy in 1 year. For optimum protection, the condom must be used at every sex act and requires the active participation of the man. Lubricants containing oil-based products may weaken latex condoms and should not be used with them; water-based lubricants (K-Y jelly is one) are safe to use. The male condom causes no side effects except possible irritation or allergy.

Female Condom

The female condom is a soft, loose-fitting prelubricated pouch with two flexible polyurethane rings, one at each end. The smaller ring at the closed end is inserted well into the vagina, creating a barrier to sperm. The larger ring remains outside the vagina, covering the vulva and providing additional protection. The female condom can be inserted before beginning sexual activity and left in place for a longer time than the male condom after ejaculation occurs. Because polyurethane is stronger than the latex used in most male condoms, the female condom is less likely to rupture. Both polyurethane and latex prevent virus transmission and should reduce the risk for acquiring HIV infection.

Hormonal (Steroidal) Contraception

Contraception using steroid hormones has been available since the 1960s. The use of hormones (or derivatives/analogues of hormones) that occur naturally in the female reproductive cycle can alter the reproductive system so that ovulation does not occur or so that physical factors (such as mucus production, tubal motility, and endometrial thickness) that enhance the probability of fertilization or implantation are altered. All modern hormonal formulations are made from synthetic steroids. The hormones are either a combination of an estrogen and a progestin or, in some formulations, a progestin alone. There are two major types of synthetic progestins: derivatives of 19-nortestosterone (which are used in OCs) and derivatives of 17α-acetoxyprogesterone (pregnanes). Pregnanes are structurally related to progesterone and are used in injectable contraceptives, but are not used in pills.

After the discontinuation of hormonal contraceptives, the rate of return of fertility is slightly lower for users of OCs than for users of barrier methods—but faster than for users of Depo-Provera. OCs do not cause permanent infertility or adversely affect pregnancies that occur after their discontinuation. OCs are not teratogenic if they are accidentally ingested during pregnancy.

From the user's point of view, the main differences are route of administration, length of action, how much attention the user needs to pay to the administration of the drug, and side effects. All these methods are very effective, and if used consistently they have very low pregnancy rates. Even with typical use, they are among the more effective methods, although, with the exception of the implants, they are less effective than sterilization or the IUD. The most commonly used of these methods is the OC ("the pill"), which was also the first hormonal contraceptive and the one most widely used globally.

Oral Contraceptives

There are three major types of OC formulations: fixed-dose combination, combination phasic, and daily progestin. The combination formulations are the most widely used and most effective. They consist of tablets containing both an estrogen and a progestin, usually given continuously for 3 weeks. Generally, no steroids are given for the fourth week. Three types of pills provide active tablets for 24 days, with 4 days of inactive tablets. Other types provide active tablets for 84 days followed by 7 days without active tablets or with a low dose of estrogen to allow withdrawal bleeding. The endometrium usually begins to slough 1 to 3 days after steroid ingestion is stopped, causing withdrawal bleeding, which usually lasts 3 to 4 days (and which users interpret as menstrual bleeding). The uterine blood loss with OC use averages about 25 mL per cycle, less than the 35 mL average for ovulatory cycles.

Three estrogens (ethinyl estradiol and its 3-methyl ether, mestranol, as well as one agent with estradiol valerate) are used in combined OCs. They are combined with one of two major types of 19-nortestosterone progestins—estranes and gonanes—both of which have androgenic activity. The estranes currently used in several OCs are norethindrone and its acetates, norethindrone acetate, and ethynodiol diacetate. Gonanes have greater progestational activity per unit weight than estranes, and thus a smaller amount of these progestins is used in OC formulations. One other progestin that is structurally related to spironolactone has been formulated in an OC. This progestin is called drospirenone and has antimineralocorticoid and antiandrogenic actions as well as progestational activity without androgenic activity. There are also daily progestin-only formulations that include norethindrone, levonorgestrel, or desogestrel.

Combined OCs, which contain both estrogen and progestin, consistently inhibit the midcycle gonadotropin surge and thus prevent ovulation. The progestin-only formulation has a lower dose of progestin than the combined agents and does not consistently inhibit ovulation, even though it is ingested every day. Progestin-only pills containing desogestrel appear to more consistently inhibit ovulation than other progestin-only formulations. Both combined OCs and progestin-only formulations also act on the cervical mucus and tubal motility to interfere with sperm transport. Progestins also alter the endometrium so that if fertilization occurs, implantation may be prevented. For contraceptive effectiveness to be maintained with the combination formulations, it is important that the pill-free interval be limited to no more than 7 days. This is made easier to remember by inclusion of placebo pills in the packet for the 7 hormone-free days. Continuous or extended cycle combined OCs seem to be an equally safe option for women who prefer it.

Side Effects

The synthetic steroids in OC formulations have many metabolic effects in addition to their contraceptive actions. These effects can cause the more common, less serious side effects as well as the rare, serious complications. The magnitude of these effects is directly related to the dosage and potency of the steroids in the formulations.

The most frequent symptoms produced by the estrogen component include nausea, breast tenderness, and fluid retention (bloating). The progestins can produce certain androgenic effects, such as weight gain, acne, and depression. But because estrogens decrease sebum production, women who have acne may experience improvement in their symptoms. Insufficient estrogen, too much progestin, or a combination of both may result in unscheduled (breakthrough) bleeding. This problem is more common with

formulations containing 20 μg of estrogen than with those containing 30 to 35 μg and is increased in women who also smoke cigarettes. Shortening the pill-free interval to 3 or 4 days may decrease the incidence of unscheduled bleeding with low-estrogen formulations.

The synthetic estrogens used in OCs cause an increase in the hepatic production of several proteins. Some of the proteins that are increased by ethinyl estradiol, such as factors V, VIII, and X and fibrinogen, have the potential to enhance thrombosis (see later), and an increase in angiotensinogen levels may elevate blood pressure in some users. The incidence of both venous and arterial thrombosis is higher with 50-μg estrogen formulations than with those with 20 to 35 μg of estrogen. Blood pressure should be followed in all users of combined OCs and the drug discontinued if there is a clinically significant increase. The progestins do not affect protein synthesis except to reduce levels of sex hormone–binding globulin.

High-progestin formulations have an adverse effect on the lipid profile. However, estrogen has a beneficial effect on the arterial wall and on serum lipids, so users of these agents do not have an increased risk for cardiovascular disease. The newer combination formulations with less androgenic progestins have a more favorable effect on the lipid profile. The effect of OCs on glucose metabolism is directly related to the dose, potency, and type of progestin. Although high-progestin formulations caused peripheral insulin resistance, the low-progestin formulations in current use do not significantly alter levels of glucose, insulin, or glucagon after a glucose load.

Complications and Risk Factors

Thrombosis

The background rate of venous thrombosis and embolism in women of reproductive age is about 3 per 10,000 woman years. Women of reproductive age who are not pregnant or using OCs experience thrombosis at a rate of 1.9 to 3.7 per 10,000 woman years. Among users of OCs, the relative risk is 3.5 (95% confidence interval, 2.9 to 4.3) compared with nonusers, but less than the rate of 5 to 20 per 10,000 woman years that occurs in association with pregnancy. The risk for venous thrombosis and embolism is higher for women using OCs with 50 μg of ethinyl estradiol than for those using 30 to 35 μg. In the presence of an inherited hypercoagulable state (Chapter 176), the risk for venous thrombosis is increased several-fold. Screening for coagulation deficiencies before women are started on OCs is not recommended unless the individual has a personal or significant family history of thrombotic events. Women with known inherited or acquired thrombogenic conditions should not use estrogen-containing steroid contraceptives in pills, rings, or patches because each of these agents has thrombogenic effects. Some epidemiologic studies have found that the risk for venous thromboembolism is greater in individuals ingesting OCs with the newer, less androgenic progestins than those containing levonorgestrel with the same amount of estrogen. However, other studies have reported that the risk is similar with formulations containing these two types of progestins. These studies are all observational and thus subject to bias.

Myocardial Infarction and Stroke

Myocardial infarction is rare among women of reproductive age, with a rate of 10.1 per 100,000 person years in a recent Danish cohort. Although the absolute risks for myocardial infarction and thrombotic stroke associated with the use of hormonal contraception were found to be low, the risk was increased by a factor of 0.9 to 1.7 with OCs that included ethinyl estradiol at a dose of 20 μg and by a factor of 1.3 to 2.3 with those that included ethinyl estradiol at a dose of 30 to 40 μg, with relatively small differences in risk according to progestin type. The use of high-dose OCs by women who smoke cigarettes increases the risk for myocardial infarction by about 10-fold. Therefore, combination OCs should not be prescribed to women older than 35 years who smoke cigarettes or use alternate forms of nicotine. Epidemiologic studies indicate that use of low-dose OCs by nonsmoking women without hypertension is not associated with a significantly increased incidence of either myocardial infarction or either hemorrhagic or thrombotic stroke.

Cancers of the Reproductive System

An analysis of worldwide epidemiologic data in 1988 showed that the risk for breast cancer diagnosis was increased by about 25% in young women who were currently using OCs, but this increased risk was no longer present 10 years or more after they stopped using OCs. Several studies have reported that use of OCs by women with a family history of breast cancer does not increase their risk for developing breast cancer. A large study of women aged 35 to 64 years in the United States reported that there was no significantly increased risk for breast cancer among current and former OC users compared with women who had not used OCs. A very large cohort study in Great Britain of OC users and aged-matched nonusers was initiated in 1968. Data accumulated until 2004 showed a similar incidence of breast cancer in both groups.

The epidemiologic data are conflicting regarding OC use and the risk for invasive cervical cancer or cervical intraepithelial neoplasia. Most well-controlled studies indicate that there is no change in risk for cervical intraepithelial neoplasia with OC use. The single most comprehensive study now indicates a significant increase in the risk for cervical cancer among OC users that increases with duration of use and declines with interval since last use.

Several studies have shown that the use of OCs has a protective effect against endometrial cancer. Moreover, the decrease in risk persists for many years after OCs are stopped. This protective effect is related to duration of use, increasing from a 20% reduction with 1 year of use to a 60% reduction with 4 years of use. The level of protection declines with time after use is stopped.

In addition, OCs reduce the risk for development of epithelial ovarian cancer as well as cancers with low malignant potential. The magnitude of the decrease in risk is directly related to the duration of OC use, increasing from about a 40% reduction with 4 years of use to a 60% reduction with 12 years of use. The protective effect continues for at least 20 years after the use of OCs ends. As with endometrial cancer, the protective effect occurs only in women of low parity (fewer than four), who are at greatest risk for this type of cancer.

Studies have reported that OCs significantly reduce the risk for development of colorectal cancer by about 20%.

Benign Hepatocellular Adenoma

The development of a benign hepatocellular adenoma was a rare occurrence in long-term users of high-dose OCs containing mestranol, but it is not increased by use of ethinyl estradiol OCs. There is no increased risk for liver cancer associated with OC use.

Contraindications

OCs can be prescribed for most women of reproductive age. According to the MEC, several conditions are considered absolute contraindications to use of combined hormonal contraceptives (category 4), including smoking 15 cigarettes or more per day, at age 35 or older, and severe hypertension, among others. There is no evidence that individuals with asymptomatic mitral valve prolapse should avoid using OCs. The presence of migraine headaches without aura is also not a contraindication to OC use, but if aura is present, combination OCs should not be prescribed because of a possible increased risk for stroke. OC use does not increase the risk for development of malignant melanoma or prolactin-secreting pituitary adenomas.

Management of OC Therapy

If a healthy woman has no contraindications to OC use, it is unnecessary to perform any laboratory tests, including cervical cytology, before she uses them. A pelvic examination is not required. Starting pills on the day of the visit is associated with better long-term use of the method. There is no reason to discontinue OC use unless pregnancy is desired. Intermittent discontinuation is unnecessary and puts women at risk for an unwanted pregnancy.

Although synthetic sex steroids can retard the biotransformation of certain drugs (e.g., phenazone and meperidine) as a result of substrate competition, such interference is usually not important clinically. However, some drugs can interfere clinically with the action of OCs by inducing liver enzymes that convert the steroids to more polar and less biologically active metabolites. These drugs include barbiturates, sulfonamides, cyclophosphamide, griseofulvin, and rifampin. There is a high incidence of OC failure in women ingesting rifampin as well as systemic griseofulvin, and neither should be given concurrently with OCs. Products containing St. John's wort reduce contraceptive effectiveness and cause breakthrough bleeding. Women taking certain medications for epilepsy should be treated with 50-μg estrogen formulations because many antiepileptic medications lower ethinyl estradiol levels and cause breakthrough bleeding, which may cause premature discontinuation of use.

Because of their many health benefits, including reduction in risk for endometrial and ovarian cancer and induction of regular cyclic uterine bleeding, OC use can be continued until menopause in normotensive, nonsmoking women without contraindications.

A common clinical question is what to do if a pill is missed. The standard advice for combined OCs is to take the first missed pill as soon as possible

and take the remaining pills at the usual time, even if it means taking two pills in one day (discarding any additional missed pills). If two or more pills are missed, take the most recent missed pill as soon as possible, continue taking the remaining pills at the usual time, even if it means taking two or more pills on the same day, and use back up contraception (e.g., condoms) or avoid sexual intercourse until pills have been taken for at least 7 consecutive days. If pills were missed in the last week of hormonal pills (third week of cycle), omit the hormone-free interval and start a new pack the next day. Emergency contraception should be considered. Vomiting and diarrhea for up to 48 hours should be considered as one missed pill; vomiting and diarrhea for more than 48 hours should be treated as two or more missed pills. For norethindrone-containing progestin-only pills, a pill is considered "missed" if it is more than 3 hours late.

Emergency Contraception

There is now a way that a women can avoid pregnancy even after unprotected sex acts. The method is termed *emergency contraception* because it should be used as early as possible after the unprotected sex. A formulation of 1500 µg (1.5 mg) levonorgestrel in a single tablet prevents about 85% of expected pregnancies, if used within 72 hours after coitus.[9,10] Another recently approved agent for emergency contraception is the selective progesterone receptor modulator ulipristal acetate given as a single 30-mg dose. This agent is as effective as levonorgestrel and is effective for 5 days after intercourse.

Transdermal and Intravaginal Steroid Contraceptives

Transdermal Patch

In the United States, there is one transdermal contraceptive patch that contains both estrogen and progestin (Ortho Evra). The patch has an area of 20 cm^2 and delivers 150 µg of the progestin norelgestromin, the active metabolite of norgestimate, and 20 µg of ethinyl estradiol daily. It may be applied to the buttocks, lower abdomen, upper arm, or upper torso (but not the breasts). The patch should be removed after 7 days and a new patch applied to a different area of skin. A woman using this method uses three patches sequentially, each for 7 days. After the third patch is removed, she waits 7 days before starting her next patch, thus mimicking the 28-day combined OC cycle (21 hormone days, followed by 7 hormone-free days, during which withdrawal bleeding occurs). Because the patch does not require daily attention, adherence with the patch is somewhat higher than with OCs. Contraceptive efficacy, bleeding patterns, and side effects are similar to those associated with OCs, and the contraindications are similar. Although the effectiveness of the patch may be decreased among women weighing more than 90 kg, there does not appear to be an association between pregnancy risk and body mass index (BMI). For all combined hormonal contraceptives, a BMI of 30 kg/m^2 or greater is considered category 2 (benefits of use outweigh potential risks) by the MEC.

Intravaginal Ring

Another option for nonoral hormonal contraception is the vaginal ring (NuvaRing in the United States). This soft, flexible ring measures 58 mm in diameter and is 4 mm thick. The ring is composed of ethinyl vinyl acetate and contains the progestin etonogestrel, a major metabolite of desogestrel, and ethinyl estradiol. The ring is inserted and removed by the woman herself. There is no "wrong" position or placement of the ring as long as it is inside the vagina. Each ring is left in place for 3 weeks, after which time it is removed for 1 week to allow withdrawal bleeding. Each day, 120 µg of etonogestrel and 15 µg of ethinyl estradiol are released from the ring, and bleeding with the ring in place is uncommon. Contraceptive efficacy and side effects are similar to those of combined OCs, as are contraindications. Women may keep the ring in place during intercourse, or it can be safely removed for up to 3 hours and then reinserted. Tampons may also be used concurrently with the ring without affecting efficacy.

Injectable Steroid Contraceptives

Constituents and Use

Although several types of injectable steroid formulations are in use for contraception throughout the world, currently the only injectable available in the United States is depot medroxyprogesterone acetate (DMPA). The initial formulation of this contraceptive was administered as an intramuscular injection of 1 mL of an aqueous suspension containing 150 mg of crystalline medroxyprogesterone acetate once every 3 months. A recently developed formulation that is administered subcutaneously (DMPA-SC) contains 104 mg of DMPA in 0.65 mL of solution. This lower dose formulation has a lower peak medroxyprogesterone acetate concentration than DMPA and a long duration of action that suppresses ovulation for at least 13 weeks and is not affected by body mass. The formulation for subcutaneous administration allows the possibility for women to self-inject the medication. Other injectable contraceptives include norethindrone enanthate, given in a dose of 200 mg every 2 months, and several once-a-month injections of combinations of different progestins and estrogens.

DMPA has a low failure rate, 0.1% at 1 year and 0.4% at 2 years. The major contraceptive action of DMPA is inhibition of ovulation, and it also impedes sperm transport by thickening cervical mucus. With DMPA and DMPA-SC, serum medroxyprogesterone levels rapidly increase to contraceptively effective blood levels (>0.5 ng/mL) within 24 hours after the injection. With DMPA, medroxyprogesterone levels plateau for about 3 months, after which there is a gradual decline until levels become undetectable 7 to 9 months after the injection. With DMPA-SC, medroxyprogesterone levels steadily decline after the initial peak and reach 0.2 ng/mL 3 to 4 months after the injection.

Side Effects

With both formulations, mean endogenous estradiol levels remain above the postmenopausal range (40 to 60 pg/mL), and symptoms of estrogen deficiency do not occur. Although DMPA may decrease bone mineral density during use, it is unnecessary to measure bone mineral density or to administer bone antiresorptive agents in DMPA users because the bone loss is temporary and reversible after stopping DMPA.

Because of the lag time it takes to clear DMPA from the circulation, resumption of ovulation is delayed for a variable time after the last injection. It may take as long as 1 year for ovulatory cycles to return. After this initial delay, fecundity resumes at a rate similar to that found after discontinuation of a barrier contraceptive.

The major side effect of DMPA is complete disruption of the menstrual cycle. Because this formulation contains only a progestin, without an estrogen, endometrial integrity is not maintained, and usually light uterine bleeding occurs at irregular and unpredictable intervals. As duration of therapy increases, the incidence of frequent bleeding steadily declines and the incidence of amenorrhea steadily increases so that at the end of 2 years, about 70% of users are amenorrheic. Because the major reason for discontinuance of all progestin-injectable contraceptives is menstrual irregularity, several combined progestin-estrogen injectables that are given once monthly and produce regular withdrawal bleeding have been developed, but these are not available in the United States.

Most DMPA users gain between 1.5 and 4 kg in their first year of use and continue to gain weight thereafter. If weight gain occurs, calorie intake should be decreased. Because there is no estrogen in DMPA, its use does not cause hypertension or thromboembolism. DMPA use is associated with a reduction in seizures among women with epilepsy, as well as a reduction in painful crises among women with sickle cell disease.

Subdermal Implants

Constituents and Use

The only subdermal implant currently available in the United States is a single 4-cm by 2-mm ethylene vinyl acetate rod containing 68 µg of etonogestrel, the active metabolite of desogestrel (Implanon or Nexplanon, which is radiopaque). It provides effective contraception for 3 years. The rod is packaged in a disposable metal trocar inserter and does not require a skin incision for insertion, only for removal. Ovulation is inhibited by the circulating etonogestrel levels, and no pregnancies were reported in three large clinical trials. As with other progestin-only implants, irregular bleeding is the most common clinical complaint. Because implants are not user dependent, the typical-use and perfect-use failure rates are identical and very low, making this method essentially as effective as IUDs and sterilization. Another subdermal implant, Jadelle, consists of two 4.3-cm rods each containing 75 mg of levonorgestrel and is approved for 5 years of contraception; it is not yet available in the United States.

Intrauterine Devices

Two options for intrauterine contraception are available in the United States: the copper-containing IUD and the levonorgestrel intrauterine system (LNG-IUS). Both methods are exceedingly effective with perfect- and typical-use failure rates of less than 1%.

The copper T380A IUD is approved for use in the United States for 10 years and maintains its high levels of effectiveness for at least 12 years. The

LNG-IUS is approved for 5 years of use, and it releases a dose of 20 μg of levonorgestrel from the device into the endometrial cavity each day. This causes atrophy of the endometrial lining, which markedly reduces the amount of uterine bleeding, and it is approved to treat menorrhagia. A newer LNG-containing IUD recently became available in the United States that is smaller and designed for use by nulliparous women.

The main mechanism of action for copper IUD is spermicidal. This effect is caused by a local sterile leukocytic response produced by the copper as well as the plastic IUD. The levonorgestrel-releasing IUD acts mainly by preventing transport of spermatozoa through the cervical mucus and thus preventing fertilization of the ovum. In addition, some women do not ovulate because of the systemic absorption of levonorgestrel. After removal of each type of IUD, the inflammatory reaction rapidly disappears, and resumption of fertility is prompt.

The main difference between the two IUDs is the menstrual bleeding pattern. With the copper IUD, women generally continue to have a regular menstrual period, which may be associated with more pain and heavier bleeding. With the LNG-IUS, irregular bleeding is common in the first 4 to 6 months of use, but after that time, most women develop amenorrhea.

Both IUDs can be easily inserted by any clinician who has been trained. No special tests are needed routinely before insertion, and if it is reasonably certain that the woman is not pregnant, the IUD may be inserted on the same day she presents requesting it. It is not necessary to wait for the next menstrual period. Almost all women, including nulliparous and young women, are considered good candidates for the IUD. Uterine perforation is a rare complication of IUD insertion, occurring in less than 0.1% of cases. Spontaneous expulsion of the IUD after insertion is also rare and happens in less than 5% of users. An IUD may be safely inserted immediately after a delivery or abortion, although the expulsion rate may be slightly higher.

Development of acute salpingitis more than 1 month after insertion of the IUD is due to infection with a sexually transmitted pathogen and is unrelated to the presence of the device. All IUD-related upper genital tract infections occur only during the insertion process. If there is clinical suspicion that cervicitis is present, an endocervical test for chlamydia and gonorrhea should be performed and the insertion delayed until negative results are obtained. It is not recommended to administer antibiotics routinely with IUD insertion.

Sterilization

Considering both tubal ligations for women and vasectomy for men, sterilization is the most common contraceptive method used by couples in the United States. Female sterilization may be performed transabdominally, such as at the time of cesarean delivery; through a minilaparotomy incision immediately postpartum; laparoscopically; or hysteroscopically. Both laparoscopic tubal ligation and hysteroscopic sterilization may be performed as outpatient procedures. Hysteroscopic tubal occlusion using the Essure device requires evaluation with a hysterosalpingogram 3 months after the procedure to confirm tubal occlusion.

Vasectomy is a simple outpatient procedure that can be performed under local anesthesia. Although many men are concerned about the possibility, sexual function is not affected by vasectomy. There are often programs that support contraceptive services for low-income women, but it is often more difficult for low-income men to access vasectomy.

Grade A References

A1. Edelman A, Micks E, Gallo MF, et al. Continuous or extended cycle vs. cyclic use of combined hormonal contraceptives for contraception. *Cochrane Database Syst Rev*. 2014;7:CD004695.

A2. Stegeman BH, de Bastos M, Rosendaal FR, et al. Different combined oral contraceptives and the risk of venous thrombosis: systematic review and network meta-analysis. *BMJ*. 2013;347:f5298.

A3. Vessey M, Yeates D. Oral contraceptive use and cancer: final report from the Oxford-Family Planning Association contraceptive study. *Contraception*. 2013;88:678-683.

A4. Piaggio G, Kapp N, von Hertzen H. Effect on pregnancy rates of the delay in the administration of levonorgestrel for emergency contraception: a combined analysis of four WHO trails. *Contraception*. 2011;84:35-39.

A5. Hohmann HL, Reeves MF, Chen BA, et al. Immediate versus delayed insertion of the levonorgestrel-releasing intrauterine device following dilation and evacuation: a randomized controlled trial. *Contraception*. 2012;85:240-245.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Finer LB, Zolna MR. Unintended pregnancy in the United States: incidence and disparities, 2006. *Contraception*. 2011;84:478-485.
2. Finer LB, Zolna MR. Shifts in intended and unintended pregnancies in the United States, 2001-2008. *Am J Public Health*. 2014;104:S43-S48.
3. Fact sheet. Unintended pregnancy in the United States. *Guttmacher Inst*. 2013; www.guttmacher.org/pubs/FB-Unintended-Pregnancy-US.html. Accessed February 7, 2015.
4. Jones J, Mosher W, Daniels K. Current contraceptive use in the United States, 2006-2010, and changes in patterns of use since 1995. *Natl Health Stat Report*. 2012;1-25.
5. Bahamondes L, Bahamondes MV. New and emerging contraceptives: a state-of-the-art review. *Int. J Women's Health*. 2014;6:221-234.
6. Winner B, Peipert JK, Zhao Q, et al. Effectiveness of long-acting reversible contraception. *N Engl J Med*. 2012;366:1998-2007.
7. Division of Reproductive Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention (CDC). U.S. Selected Practice Recommendations for Contraceptive Use, 2013: adapted from the World Health Organization selected practice recommendations for contraceptive use, 2nd edition. *MMWR Recomm Rep*. 2013;62:1-60.
8. Division of Reproductive Health, National Center for Chronic Disease Prevention and Health Promotion. U.S. Selected Practice Recommendations for Contraceptive Use, 2013. Adapted from the World Health Organization selected practice recommendations for contraceptive use, 2nd ed. *MMWR*. 2013;62:1-46.
9. American College of Obstetricians and Gynecologists. Practice bulletin no. 112: emergency contraception. *Obstet Gynecol*. 2010;115:1100-1109.
10. Koyama A, Hagopian L, Linden J. Emerging options for emergency contraception. *Clin Med Insights Reprod Health*. 2013;7:23-35.

REVIEW QUESTIONS

1. A 34-year-old nulliparous woman requesting contraception is found to have a blood pressure of 165/90 mm Hg. Otherwise, her screening for medical contraindications is negative. Which contraceptive methods would *not* be appropriate for her?

A. Progestin-only pills (POPs)
B. Combined oral contraceptives (COCs)
C. Copper-containing IUD
D. Etonogestrel subdermal implant
E. Levonorgestrel-IUS

Answer: B According to the Medical Eligibility Criteria, this level of hypertension is considered a category 4 contraindication for COCs. Hypertension is considered category 2 for POPs and implants, and it is category 1 for the copper IUD. Nulliparity is considered a category 2 condition for the copper IUD.

2. What would you advise a woman who comes seeking protection from one act of unprotected intercourse 60 hours before her visit?

A. Have a copper IUD inserted.
B. Find a regular method of contraception.
C. Fill a prescription for ulipristal emergency contraception (EC).
D. Buy a package of levonorgestrel emergency contraception if other alternatives are not readily available to her.
E. All of the above

Answer: E The clinician should give the woman the choice of how to proceed after explaining the options and their advantages and disadvantages. The IUD works as emergency contraception in this time frame, but there is no large series of data on use of the progestin-containing device as opposed to the copper IUD, largely because it was not available when the earlier studies were done. The time period after exposure is associated with a diminution of efficacy for LNG EC pills, but there is probably still some effect, and they are easy to obtain. Ulipristal works in this time frame, but it must be supplied by prescription and is costly in the United States unless it is covered by insurance. Women who find themselves in need of EC should review their usual contraceptive practices and try to find an effective method that suits them for regular use.

3. A woman calls your office saying that she took her last combined oral contraceptive (COC) pill 3 days ago. She was traveling and lost her pill pack, and she has now returned home. She had sex yesterday and the day before, both times using a male condom, but she thinks the condom broke yesterday. What would you advise her?

A. Start taking a new pack of COC pills immediately.
B. Take two pills today and tomorrow.
C. Use condoms for the next 7 days.
D. Take levonorgestrel emergency contraception.
E. A, C, and D

Answer: E For any scenario with missed pills, women should take a pill as soon as it noticed to be late, and then continue taking one at the usual time. At the most, this would require taking two pills once; repeated days of double-dosing of pills is not recommended. Condoms must be used until the woman is back on COCs for 7 days. Because there is the possibility that the condom broke and she had unprotected sex within the last 24 hours, she should consider taking emergency contraception.

4. What contraceptive method would you recommend for a woman seeking protection from sexually transmitted infection (STI) and pregnancy who wishes to use only one method?

A. Combined oral contraceptives
B. Progestin-containing IUD
C. Female condom
D. Male condom
E. Spermicides

Answer: D Hormonal contraception does not protect against the acquisition of infection. Spermicides, although toxic to infectious organisms in vitro, do not seem to provide protection against STIs in vivo, and the irritation of the genital tract caused by these compounds may create a portal of entry for pathogens. The female condom is also not a proven method to protect against HIV and other sexually transmitted pathogens, but there is a suggestion that it may reduce risk. The only well-studied method that clearly prevents both pregnancy and STI transmission is the male condom.

5. A woman had a copper IUD inserted 3 weeks ago and complains that she cannot feel the strings in her vagina. On speculum examination, the strings are not visible. On pelvic ultrasound, a bright echogenic linear structure is noted within the uterine cavity. Her vital signs and examination are otherwise normal. What is the appropriate diagnosis and next step?

A. The IUD is correctly placed, but the strings have withdrawn into the uterine cavity. Reassure her that the IUD will be effective; no further action is necessary.
B. The IUD has perforated the uterus. The patient must be scheduled for laparoscopic removal of the misplaced IUD.
C. The IUD is misplaced. It must be removed and replaced in order to be effective.

Answer: A The ultrasound confirms that the IUD is correctly placed within the uterus. Even though the strings are not visible, the IUD will be effective. At the time of removal, small forceps may be used to grasp the strings, which are likely in the endocervical canal or lower portion of the uterus. There is no evidence of perforation in this case.

239

COMMON MEDICAL PROBLEMS IN PREGNANCY

KAREN ROSENE-MONTELLA

There are 62 million women of childbearing age in the United States, 85% of whom will give birth by the age of 44 years. The majority of these women will not have obtained preventive health services in any given year, and more than half the pregnancies will be unplanned or unintended. At least 25% will enter pregnancy with a chronic medical illness, and more than half will be overweight or obese, making the role of the internist paramount in maternal health. In the most recent Confidential Enquiry into maternal mortality in the United Kingdom, more than half of all women who died of direct or indirect causes were overweight or obese, and more than 15% of all deaths were in morbidly obese women. Sixteen percent of pregnant women have depression in the perinatal period, and depression rates are even higher in those with chronic illnesses, such as diabetes and asthma.

By the time pregnant patients are seen by their obstetricians, most major teratogenic abnormalities have already occurred (Fig. 239-1), and the window of opportunity to enter pregnancy in a quiescent disease state, on the safest possible medication profile, may have passed. For this reason, internists caring for women of childbearing age have a unique responsibility to provide preconception care at a time when interventions will be of maximum benefit to both the fetus and the mother. Table 239-1 describes preconception interventions for women with chronic medical illnesses.

The basic principles involved in the care of pregnant patients with medical disorders are reviewed here, followed by a more detailed discussion of certain medical conditions, selected because of their contribution to maternal mortality or because of the frequency with which they occur.

BASIC PRINCIPLES

Pregnancy is associated with significant but normal physiologic changes that have an impact on the diagnosis and management of disease states and the pharmacokinetics of most drugs (Table 239-2).[1] The physiologic changes required during pregnancy may stress the woman's ability to adapt, particularly in the presence of an underlying disease. The mother's response to pregnancy often unmasks diseases or predicts future risk, so pregnancy is an opportunity to identify women at risk for other non–pregnancy-related illnesses. For example, gestational diabetes is predictive of an increased risk for type 2 diabetes; preeclampsia is predictive of increased risk for ischemic heart disease and stroke[2]; and thrombosis, late fetal loss, or preeclampsia may unmask an underlying thrombophilia.

Fetal well-being depends on maternal well-being. Although there is often thought to be a dichotomy between maternal and fetal needs, they are usually one and the same. The fetus is dependent on maternal perfusion, oxygenation, and nutrition. Thus, more harm may be done by withholding necessary treatments and investigations from pregnant women than by providing them. Uninvestigated symptoms lead to the progression of untreated disease, and untreated maternal disease compromises fetal safety, growth, and development. The major cause of asthma exacerbations and seizures during pregnancy is abrupt discontinuation of medications, exposing the fetus to hypoxemia and acidosis in an effort to save the fetus from drug exposure. A population analysis of prescriptions for asthma medications in The Netherlands showed that prescriptions for controller medications decreased by 30% during the first months of pregnancy.[3] In the U.K. Confidential Enquiry, in more than half the cases of maternal death from pulmonary embolism, failure to make the diagnosis was due to the unfounded fear that diagnostic testing would be harmful to the fetus. Most diagnostic imaging can be used safely in pregnancy. The effects of radiation in utero depend on both the gestational age at exposure and the level of exposure. Recommendations on fetal exposure from the National Commission on Radiation Protection are summarized in Table 239-3. Radiation exposure from specific diagnostic tests is provided in Table 239-4.

The effect of contrast agents is related to the bioavailability of iodine, and there is concern about the impact on the fetal thyroid. Iodine availability is extremely low, and single-dose exposures, even if they are high, are unlikely

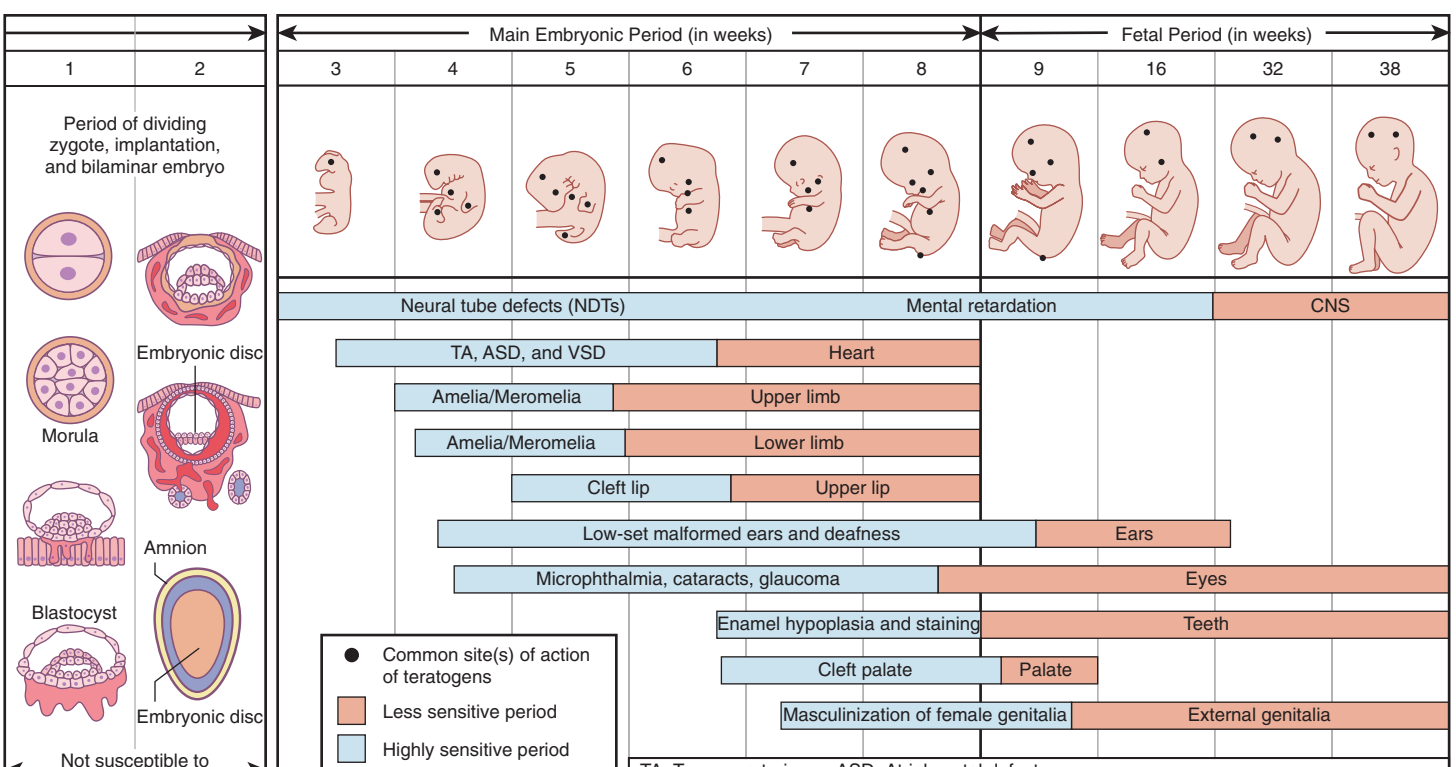

FIGURE 239-1. The developing fetus. CNS = central nervous system. **(From Moore K. The Developing Human: Clinically Oriented Embryology. Philadelphia: WB Saunders; 1982, with permission from Annals of Internal Medicine.)**

to be harmful. Therefore, contrast agents may be used when necessary. There are limited data on gadolinium, so the current recommendation is to avoid gadolinium exposure if possible.

The use of medications to treat pregnant women requires a rational risk-benefit analysis and a good understanding of the maternal indications. It is helpful to view treatment as justifiable or not justifiable rather than as safe or not safe. It is important to consider whether the condition is self-limited or harmless, what the maternal and fetal consequences of discontinuing a medication will be, and the safety data for the drug. U.S. Food and Drug Administration categories may be misleading and often do not include adequate data for a proper risk-benefit analysis. Resources such as the Teratology Information Service, found at http://depts.washington.edu/terisweb/teris, offer more complete information.

The list of known human teratogens is small and includes warfarin, cyclophosphamide, diethylstilbestrol, lithium, thalidomide, penicillamine, isotretinoin, methotrexate, acetazolamide, and the antiepileptic drugs phenytoin, carbamazepine, phenobarbital, and valproic acid. Of the antiepileptic drugs, valproate has the most significant data, and it is the only antiepileptic drug for which discontinuation during pregnancy is recommended if there is an effective alternative.[4] Angiotensin-converting enzyme (ACE) inhibitors and angiotensin II receptor blockers (ARBs) should be added to this list on the basis of data confirming that first-trimester exposure is associated with fetal renal agenesis and renal failure. Tetracyclines should be avoided because of later effects on fetal teeth and bone.

HYPERTENSIVE DISORDERS OF PREGNANCY

DEFINITION

Hypertension in pregnancy is defined as a blood pressure (BP) of 140/90 mm Hg or higher. It is defined as chronic hypertension when it predates pregnancy, is diagnosed before 20 weeks' gestation, or persists post partum. Transient late or gestational hypertension occurs toward term and resolves post partum in the absence of any other signs or symptoms of preeclampsia.

EPIDEMIOLOGY

Chronic hypertension is the most common medical condition encountered in women of childbearing age. The incidence is increasing parallel to the increase in obesity, insulin resistance, and pregnancies in women older than 30 years. Hypertension complicates 5 to 8% of pregnancies and is associated with a 20% risk for the development of preeclampsia.

PATHOBIOLOGY

Systemic arterial BP decreases by 10 to 15 mm Hg during normal pregnancy, with a greater fall in diastolic than in systolic pressures, probably because of the decreased sensitivity to angiotensin II that has been demonstrated in pregnant women. BP begins to fall in the first trimester, reaching a nadir toward the end of the second trimester and returning toward baseline at term. This decrease may be exaggerated in women with chronic hypertension, making the diagnosis of chronic hypertension difficult during pregnancy and affecting both diagnostic and therapeutic considerations.

DIAGNOSIS

The diagnosis of hypertension relies simply on a BP measurement, in the sitting position at the level of the heart, of 140/90 mm Hg or higher on two occasions 6 hours apart. Later in pregnancy, inferior vena cava compression by the gravid uterus may lower BP substantially in the supine position, so it is critical to measure maternal BP in the sitting position. The initial evaluation should document target organ damage (such as left ventricular hypertrophy), renal disease (creatinine concentration, urinalysis, and potassium concentration), and retinopathy so that a baseline is established. Consideration of secondary causes is necessary in this young population (Chapter 67), but the diagnosis of secondary causes of hypertension is complicated by normal pregnancy-related changes.[5] The diagnosis of Cushing's syndrome is complicated by increased levels of cortisol and the placental production of adrenocorticotropic hormone and corticotropin-releasing hormone, so the best test is a 24-hour urine free cortisol measurement with higher pregnancy-specific reference ranges. Primary hyperaldosteronism (Chapter 227) may also be

TABLE 239-1 PRECONCEPTION INTERVENTIONS FOR WOMEN WITH MEDICAL ILLNESSES

TYPE 1 AND TYPE 2 DIABETES

Discuss importance of a normal hemoglobin A_{1c} before conception and importance of using contraception until that is achieved
Evaluate for microvascular complications
  Obtain remission for proliferative retinopathy
  Emphasize need to discontinue ACE inhibitor after first missed period
Discontinue thiazolidinediones and statins
Consider change to insulin therapy for type 2 diabetic patients on oral agents unless using metformin for ovulation induction in PCOS
Discuss probable need to reduce insulin dose in first trimester

THYROID DISEASE

Screen for hypothyroidism in women at risk
Normalize TSH and free T_4 before pregnancy
Counsel women taking levothyroxine on probable need to increase dose soon after conception
Diagnose cause of hyperthyroidism and consider ablative therapy for women with Graves' disease requiring high doses of PTU

CHRONIC HYPERTENSION/RENAL DISEASE

Rule out secondary causes of hypertension if appropriate
Evaluate extent of end-organ disease
Quantify GFR and proteinuria
Discuss drugs of choice for hypertension and replace ACE inhibitor
Discuss risk of superimposed preeclampsia and use of low-dose aspirin for women at significant risk of preeclampsia

THROMBOEMBOLIC DISEASE

Consider evaluation for congenital or acquired thrombophilias in women with previous VTE, previous poor obstetric outcome, or family history
Discuss risks of warfarin in pregnancy, need to discontinue warfarin by 4-6 weeks' gestation, and conversion to unfractionated or low-molecular-weight heparin
Discuss options to combined oral contraceptives

EPILEPSY

Determine whether patient is a candidate for withdrawal of antiepileptic drugs
Consider monotherapy with most effective agent at lowest dose possible
Prescribe folate at 1 to 4 mg/day
Discuss possible ineffectiveness of low-dose contraceptives with phenobarbital, phenytoin, and carbamazepine
Consider discontinuing valproate

CARDIAC DISEASE

Obtain baseline echocardiography if congenital disease, stenotic lesion, or pulmonary hypertension suspected
Evaluate for coronary artery disease in women with multiple risk factors

ASTHMA

Verify patient's asthma action plan and peak flowmeter use
Discuss relative safety of all asthma medications except leukotriene modifiers

SYSTEMIC LUPUS ERYTHEMATOSUS AND AUTOIMMUNE DISEASE

Evaluate for renal and cardiopulmonary disease and antiphospholipid, anti-Ro, anti-La antibodies
Avoid pregnancy if disease is active
Discuss relative safety of most immunosuppressants

ACE = angiotensin-converting enzyme; GFR = glomerular filtration rate; PCOS = polycystic ovary syndrome; PTU = propylthiouracil; T_4 = thyroxine; TSH = thyroid-stimulating hormone; VTE = venous thromboembolism.
From Rosene-Montella K, Keely EJ, Lee RV, Barbour LA, eds. *Medical Care of the Pregnant Patient.* 2nd ed. Philadelphia: ACP Press/American College of Physicians; 2008.

difficult to diagnose in the face of normal pregnancy-related elevations in plasma renin activity and aldosterone and because progesterone ameliorates both the hypertensive and the kaliuretic effects of aldosterone. Primary hyperaldosteronism should be strongly considered in any patient with chronic hypertension in whom there is a marked increase in BP toward term or post partum. Pheochromocytoma (Chapter 228) is associated with a high maternal and fetal mortality rate, in large part owing to a delay in diagnosis. Both magnetic resonance imaging and magnetic resonance angiography (which do not require gadolinium) can be used safely in pregnancy to evaluate the adrenal glands and renal arteries.

TABLE 239-2 NORMAL PHYSIOLOGIC CHANGES IN PREGNANCY

CARDIAC

Cardiac output increased 40%
Blood volume increased 30-50%
Heart rate increased 10-20 beats/min
Blood pressure decreased 10-15 mm Hg
ECG changes related to widened thorax, dextrorotation of heart, elevation of diaphragm

PULMONARY

Upper airway hyperemia and glandular hyperactivity leading to increased edema and friability
Nasal congestion, gestational rhinitis, snoring
Difficult airway management and failed intubation
Minute ventilation increased (owing to an increase in tidal volume, *not* respiratory rate, which remains unchanged), which leads to relative respiratory alkalosis (pH 7.4-7.45)
  Normal Pao_2 100-105 mm Hg
  Normal $PaCO_2$ 28-32 mm Hg

RENAL

Increased GFR to 150-180 mL/min/1.73 m^2
Normal serum creatinine concentration <0.8 mg/dL
Increased renal excretion of bicarbonate, limiting buffering capacity in patients who become acidotic
Decreased oncotic pressure

ALTERED PHARMACOKINETICS

Increased renal and hepatic clearance of drugs
Altered absorption
Altered protein binding
Increased volume of distribution

ECG = electrocardiogram; GFR = glomerular filtration rate.
These physiologic changes generally progress throughout gestation.

TABLE 239-3 NATIONAL COMMISSION ON RADIATION PROTECTION (NCRP) RECOMMENDED PREGNANCY EXPOSURES

TOTAL EXPOSURE DURING PREGNANCY (rad)	NCRP RECOMMENDATIONS
≤5	Acceptable; low likelihood of problems
5-10	Low risk for problems
10-15 (at ≤8 wk gestation)	Higher risk; consideration of termination
>15	Termination of pregnancy recommended

From Rosene-Montella K, Keely EJ, Lee RV, Barbour LA, eds. *Medical Care of the Pregnant Patient.* 2nd ed. Philadelphia: ACP Press/American College of Physicians; 2008.

TABLE 239-4 RADIATION EXPOSURE

STUDY	RADIATION EXPOSURE (rad)
Chest radiography	<0.001
Lung scan	0.01-0.02 ventilation 0.01-0.03 perfusion
Pulmonary angiography	<0.050 by brachial route 0.2-0.3 by femoral route
CT angiography	0.2-0.3
Ultrasound	None
MRI, MRA, MRV	None
Upper gastrointestinal series	0.1
Lumbar spine series	0.9
Barium enema	1
Complete IVP	0.5
Head CT	<0.01
CT of abdomen	2.0-3.0

CT = computed tomography; IVP = intravenous pyelography; MRA = magnetic resonance angiography; MRI = magnetic resonance imaging; MRV = magnetic resonance venography.

TREATMENT

Rx

Patients previously receiving drug therapy can often discontinue antihypertensives and restart them when BP gradually rises to prepregnant values toward term. It is difficult to determine whether rising BP represents a normal physiologic return to earlier pressure or the development of preeclampsia. Baseline preeclampsia laboratory tests (see later), very close follow-up, and comanagement with the patient's obstetrician are required. Patients with good BP control may prefer to continue safe medications or switch to another regimen. The drugs for which there is grade A evidence of efficacy and safety are methyldopa and labetalol (Table 67-5). Nifedipine, hydralazine, and other β-blockers, especially those with intrinsic sympathomimetic activity, have also been studied and are acceptable second- and third-line agents.

Dosing should take into consideration the increase in renal and hepatic clearance and the increased volume of distribution, which may require higher doses or narrowed dosing intervals during pregnancy. ACE inhibitors and ARBs should be discontinued at the diagnosis of pregnancy because of their teratogenicity and their association with fetal and neonatal renal agenesis and renal failure even when they are used later in gestation.

The goal of antihypertensive therapy in pregnancy is not clear. Most consensus recommendations, which address fetal concerns and short-term maternal safety only, recommend keeping BP below 160/100 mm Hg. Given the long-term maternal data, most centers prefer to keep maternal BP, particularly in patients with diabetes or renal disease, below 140/90 mm Hg. Consensus recommendations agree that maintaining BP above 120/80 mm Hg is necessary to preserve placental perfusion. There is no evidence that salt restriction or dietary changes improve BP control in pregnancy, and weight loss is not recommended. Likewise, there is no evidence that BP control decreases the risk of preeclampsia. It is important to obtain baseline preeclampsia laboratory tests (complete blood count, platelet count, creatinine concentration, uric acid level, aspartate transaminase level, urinalysis) in all patients with hypertension, given the 20% risk of preeclampsia, and low-dose aspirin and calcium supplementation should be considered to prevent preeclampsia (see later). Fetal monitoring with serial ultrasound for growth and amniotic fluid volume, nonstress testing (fetal heart rate acceleration in response to movement) once or twice a week after 32 weeks, and consideration of Doppler flow velocimetry are recommended.

Most antihypertensives are safe for breast-feeding, which should be encouraged. Hydrochlorothiazides, α-methyldopa, nifedipine, acebutolol, and metoprolol are all approved by the American Academy of Pediatricians. There is no evidence that hydrochlorothiazides affect milk volume. There is evidence that propranolol and atenolol are concentrated in breast milk, so they should be avoided. Enalapril and captopril are the preferred ACE inhibitors in breast-feeding women, but it may be prudent to delay ACE inhibitors for the first few weeks of the baby's life and for mothers of premature babies, given the adverse pregnancy data.

PROGNOSIS

Hypertension may increase the risk of placental abruption, intrauterine growth restriction (IUGR), and low-birthweight babies. The major risk, however, is its contribution to the risk for preeclampsia and the associated increase in perinatal morbidity and mortality. In addition, chronic hypertension in patients with other comorbidities significantly increases the risk for maternal and fetal complications[6] (Table 239-5).

Women who develop hypertension during pregnancy are at increased lifetime risk for the development of chronic hypertension, even if the BP normalizes post partum.[7]

PREECLAMPSIA

DEFINITION

Preeclampsia is a multisystem disorder defined as BP of 140/90 mm Hg or higher, accompanied by proteinuria of more than 300 mg/24 hours after the 20th week of gestation in a previously normotensive patient. When it is diagnosed in a patient with preexisting chronic hypertension, it is referred to as chronic hypertension with superimposed preeclampsia.

A urine protein-creatinine ratio of at least 0.3 may become a criterion for proteinuria, but the American College of Obstetricians and Gynecologists has not yet added this to the definition. Edema and hyperreflexia are no longer considered diagnostic criteria, and the 30 mm Hg increase in systolic pressure or 15 mm Hg increase in diastolic pressure has been dropped from the criteria for hypertension. Severe preeclampsia is defined as the presence of one of the following symptoms or signs with preeclampsia: systolic BP of 160 mm Hg or higher, or diastolic BP of 110 mm Hg or higher, on two occasions at least 6 hours apart; proteinuria of more than 5 g in a 24-hour period; pulmonary edema; oliguria (<400 mL in 24 hours); persistent headaches; epigastric pain or impaired liver function; thrombocytopenia; and IUGR.

EPIDEMIOLOGY

Preeclampsia complicates 6 to 8% of pregnancies worldwide. Preeclampsia/eclampsia is a leading cause of maternal mortality in the developing world and continues to contribute to maternal mortality in the United States despite the availability of antihypertensives and antiseizure medications. In the United States, preeclampsia is believed to be responsible for 15% of premature deliveries and 17.6% of maternal deaths. Worldwide, preeclampsia and eclampsia are estimated to be responsible for approximately 14% of maternal deaths per year (50,000 to 75,000).

Primigravida and multigravida women with new partners are at increased risk, suggesting a role for paternal antigens. Additional risk factors include prior history of preeclampsia, black race, diabetes or insulin resistance, obesity, systemic lupus erythematosus (SLE), renal disease, hypertension, thrombophilia, obesity, molar pregnancy, multiple gestation, and extremes of age (younger than 20 years or older than 40 years).

PATHOBIOLOGY

Preeclampsia is a disorder of abnormal placentation that begins early in gestation, well before its manifestations are clinically apparent. In normal pregnancy, uterine spiral arteries undergo remodeling when they are invaded by fetal cytotrophoblastic cells, resulting in an adhesion receptor switch from cells with characteristics of epithelial cells to cells with the phenotype of endothelial cells. This leads to the transformation of previously narrow, high-resistance maternal uterine blood vessels into dilated, high-capacitance blood vessels. The proximal portions of the spiral arteries are further dilated by the hormonal effects of estrogen and progesterone, resulting in an overall increase in uterine blood flow from 45 mL/minute during menstruation to 750 mL/minute at term. In preeclampsia, this cell switching does not occur, and the fetal cells' only superficial invasion into maternal vasculature results in limited placental perfusion. As the pregnancy progresses, abnormal placentation produces relative hypoxia and ischemia as this compromised uterine blood flow cannot keep up with the growing demands of the fetus and the placenta. The result is diffuse endothelial dysfunction that is manifested as the clinical syndrome of preeclampsia (Fig. 239-2).

Recent work on the potential mechanism of the endothelial dysfunction that underlies this disease state has focused on the imbalance between pro-angiogenic and anti-angiogenic factors, arising from studies showing elevated concentrations of placental soluble film-like tyrosine kinase 1 in the plasma of women with preeclampsia. This protein prevents the interaction of placental growth factor and vascular endothelial growth factor with endothelial receptors, inducing endothelial dysfunction. A placenta-derived soluble transforming growth factor-β coreceptor, soluble endoglin, is elevated in the sera of patients with preeclampsia, inducing vascular permeability and hypertension; the degree of elevation correlates with disease severity, and levels fall after delivery. Additional factors currently being studied include ADAM12 trophoblast and angiogenesis markers, pregnancy-associated plasma protein A, placental protein 13, and placental growth factor.[8] To date, no definitive predictive model has been established. Current studies to address both the cause of and the mechanism linking preeclampsia and the risk of cardiovascular disease are ongoing and focus on endothelial dysfunction.[9]

DIAGNOSIS

The diagnosis of preeclampsia depends on a BP of 140/90 mm Hg or higher and proteinuria of more than 300 mg/24 hours after 20 weeks' gestation. Eclampsia is diagnosed when a patient with preeclampsia has a seizure. Additional abnormalities that contribute to the diagnosis include hyperuricemia, hemoconcentration, elevated creatinine or liver function test results, and thrombocytopenia. The HELLP syndrome (hemolysis, elevated liver enzymes, and low platelets) (Chapter 172) is likely to be a more severe form of preeclampsia. The diagnosis of severe preeclampsia depends on the criteria previously listed.

Diagnostic evaluation should include a careful history, asking about headache, visual complaints, epigastric pain, weight gain, and edema and reviewing the presence of risk factors. The physical examination should include a careful neurologic examination, looking for funduscopic changes (retinal vasospasm, edema, or hemorrhage) or hyperreflexia, and examination for any

TABLE 239-5 ODDS RATIOS FOR FETAL AND MATERNAL COMPLICATIONS: 1995-2008

VARIABLE	PREGESTATIONAL DIABETES		CHRONIC RENAL DISEASE		COLLAGEN VASCULAR DISEASE		THYROID DISORDERS	
	With Chronic Hypertension	*Without Chronic Hypertension*	*With Chronic Hypertension*	*Without Chronic Hypertension*	*With Chronic Hypertension*	*Without Chronic Hypertension*	*With Chronic Hypertension*	*Without Chronic Hypertension*
FETAL OUTCOMES								
Stillbirth[a]	4.30 (3.81-4.85)	3.05 (2.88-3.23)	7.29 (5.59-9.52)	1.74 (1.51-2.02)	7.42 (5.37-10.25)	2.74 (2.35-3.20)	1.86 (1.48-2.33)	0.98 (0.92-1.05)
Poor fetal growth[a]	2.66 (2.40-2.94)	1.20 (1.14-1.27)	7.94 (6.67-9.44)	2.29 (2.12-2.49)	7.99 (6.44-9.91)	3.87 (3.55-4.22)	3.59 (3.20-4.02)	1.29 (1.25-1.34)
Spontaneous delivery <37 wk gestation[a]	4.88 (4.63-5.15)	2.90 (2.83-2.98)	8.60 (7.64-9.67)	2.25 (2.15-2.35)	7.19 (6.22-8.30)	3.15 (2.98-3.33)	3.24 (3.02-3.48)	1.24 (1.21-1.27)
MATERNAL OUTCOMES								
Preeclampsia[a]	13.96 (13.29-14.66)	3.80 (3.69-3.91)	27.87 (24.85-31.25)	3.28 (3.10-3.47)	17.41 (15.09-20.09)	2.96 (2.76-3.18)	9.74 (9.15-10.35)	1.38 (1.35-1.42)
Stroke/cerebrovascular complications[a]	7.14 (4.90-10.40)	1.85 (1.41-2.44)	13.73 (6.63-28.44)	3.52 (2.34-5.31)	23.00 (11.47-46.14)	7.60 (5.26-10.97)	3.87 (2.07-7.23)	1.58 (1.29-1.94)
Acute renal failure[a]	35.41 (28.39-44.16)	4.43 (3.57-5.48)	253.4 (199.5-321.9)	62.40 (54.37-71.63)	191.5 (141.4-259.4)	12.60 (8.88-17.88)	14.17 (9.65-20.82)	1.27 (0.97-1.65)
Pulmonary edema[a]	11.97 (7.86-18.24)	4.01 (3.07-5.25)	23.29 (10.32-52.56)	9.06 (5.84-14.06)	15.52 (4.92-48.95)	6.08 (3.46-10.69)	9.85 (5.64-17.19)	1.54 (1.16-2.05)
Ventilation[a]	11.87 (9.22-15.26)	3.34 (2.80-4.00)	19.29 (11.36-32.76)	8.25 (6.43-10.60)	26.20 (15.04-45.63)	11.09 (8.46-14.52)	5.71 (3.69-8.86)	1.84 (1.55-2.18)
Cesarean delivery[b]	5.75 (5.46-6.05)	3.33 (3.26-3.41)	5.73 (5.03-6.53)	1.74 (1.68-1.81)	4.38 (3.74-5.12)	1.89 (1.80-1.98)	3.16 (2.97-3.36)	1.27 (1.25-1.29)
Length of stay >6 days[c]	14.74 (13.68-15.89)	5.34 (5.09-5.60)	42.16 (36.78-48.32)	6.52 (6.12-6.95)	30.29 (25.45-36.04)	6.18 (5.69-6.71)	8.40 (7.60-9.28)	1.77 (1.71-1.84)
In-hospital mortality[a]	6.02 (2.71-13.40)	2.58 (1.59-4.17)	27.02 (8.72-83.73)	6.88 (3.56-13.29)	88.81 (41.90-188.2)	23.81 (14.67-38.66)	1.74 (0.24-12.40)	1.72 (1.06-2.77)

For each analysis, reference group was delivery admissions without chronic hypertension and without comorbidity of interest. Admissions with chronic hypertension but without comorbidity of interest were included as a group in each analysis. Because of similarity of estimates of association in these groups to those obtained when analyzing effect of overall chronic hypertension, results are not shown.

[a]Adjusted for multiple birth, year of study, insurance status, region, and age.

[b]Adjusted for previous cesarean delivery, multiple birth, year of study, insurance status, region, and age.

[c]Adjusted for disposition status, admission status, multiple birth, year of study, insurance status, region, and age.

From Bateman BT, Bansil P, Hernandez-Diaz S, et al. Prevalence, trends, and outcomes of chronic hypertension: a nationwide sample of delivery admissions. *Am J Obstet Gynecol* 2012;206:134.e1-134.e8, 2012.

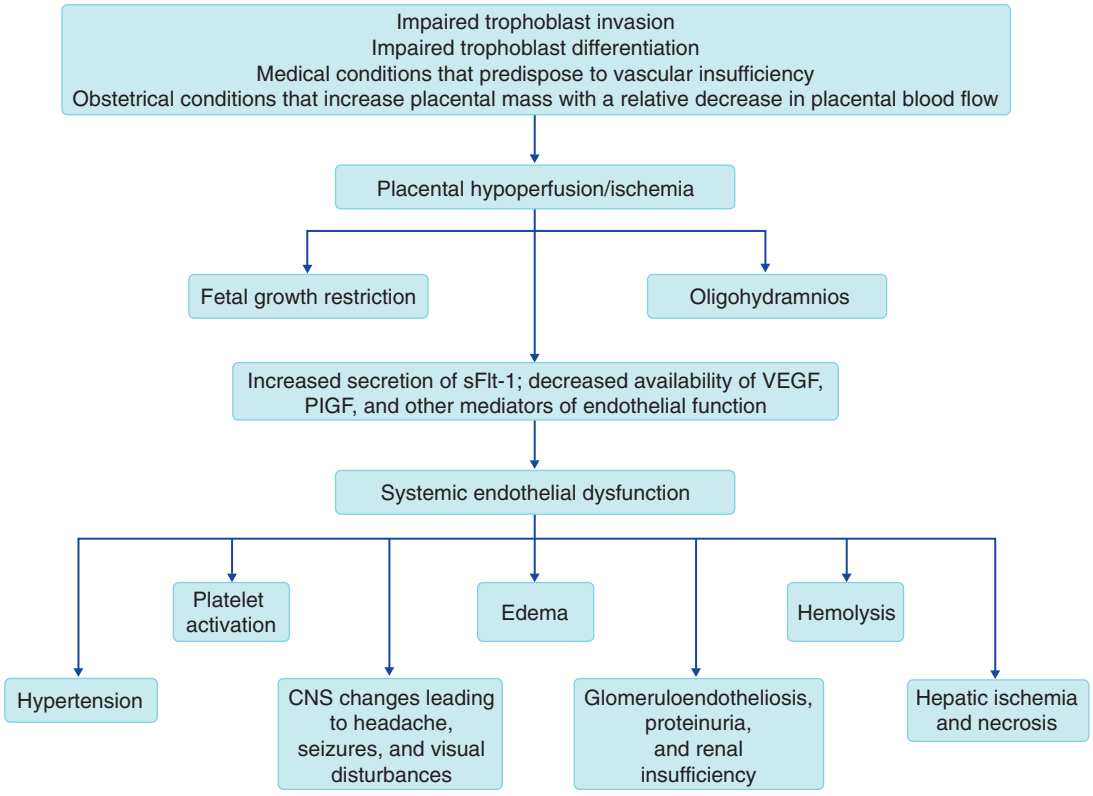

FIGURE 239-2. Model for the pathogenesis of preeclampsia. CNS = central nervous system; PlGF = placental growth factor; sFlt-1= soluble film-like tyrosine kinase; VEGF = vascular endothelial growth factor. (From Rosene-Montella K, Keely EJ, Lee RV, Barbour LA, eds. *Medical Care of the Pregnant Patient.* 2nd ed. Philadelphia: ACP Press/American College of Physicians; 2008.)

focal findings suggestive of mass effect, hepatic tenderness, and edema. Laboratory testing for preeclampsia includes a complete blood count, platelet count, urine protein-creatinine ratio or 24-hour urine protein, liver function tests, creatinine concentration, and uric acid level. Additional evaluation includes fetal testing and close maternal monitoring for life-threatening consequences, such as severe hypertension, seizures, pulmonary edema, cerebral hemorrhage, hepatic infarction or rupture, disseminated intravascular coagulation, and renal failure.

Currently, diagnosis depends on development of the full clinical syndrome, but earlier diagnosis may be possible on the basis of biomarkers. It is likely that a combined model that looks at soluble endoglin, soluble film-like tyrosine kinase 1, pregnancy-associated plasma protein A, ADAM12, and placental growth factor concentrations will be a better prediction instrument before 20 weeks' gestation than any individual marker.

The differential diagnosis of each individual manifestation of preeclampsia is broad, so the diagnosis centers on the constellation of signs and symptoms that suggest preeclampsia. The clinical conditions that can mimic preeclampsia include SLE with nephritis (Chapter 266), thrombotic thrombocytopenic purpura, and hemolytic-uremic syndrome (Chapter 172). Differentiation of preeclampsia from a flare of SLE with nephritis (Chapter 266) is difficult because both can cause hypertension, proteinuria, thrombocytopenia, and rises in serum creatinine concentration. The differential diagnostic features that favor SLE include falling serum complement levels, rising anti-DNA antibodies, and extrarenal manifestations of SLE such as rash and arthralgias. The proteinuria and hypertension in preeclampsia are more likely to be of sudden onset.

TREATMENT

Rx

Once it occurs, the only known treatment of preeclampsia is delivery as soon as it is obstetrically feasible. Nevertheless, preeclampsia can be manifested post partum, and both preeclampsia and eclampsia have been reported up to 21 days after delivery. Management of preeclampsia includes treatment of hypertension, seizure prophylaxis, and limitation of fluids due to the risk of pulmonary edema. Treatment of severe hypertension in preeclampsia is reviewed in Table 239-6. Magnesium sulfate is recommended as first-line treatment of eclampsia as well as for prophylaxis against eclampsia in women with

TABLE 239-6 TREATMENT OF SEVERE HYPERTENSION IN PREECLAMPTIC PATIENTS

MEDICATION	ONSET AND DURATION OF ACTION	ACUTE DOSING FOR SEVERE HYPERTENSION	MAINTENANCE DOSE
Labetalol	Begins to work in 5-10 min Lasts 3-6 hr	Given as a series of boluses until BP reaches the desired level: 10 mg IV push; then in 10 min, 20 mg IV push; then in 10 min, 40 mg IV push; then in 10 min, 80 mg IV push; then in 10 min, 80 mg IV push, up to a total dose of no more than 300 mg Follow with PO labetalol or labetalol drip	100-200 mg PO bid-tid (100-600 mg bid-tid; maximum 2400 mg/day) IV infusion 0.5-2.0 mg/min (labetalol comes in vials of 100 mg/20 mL) Put 5 vials (100 mL) labetalol into 150 mL IV fluid (D_5W, LR, or NS) to get a solution of 2 mg/mL; start at 15 mL/hr (0.5 mg/min); titrate up to as high as 60 mL/hr (2 mg/min)
Nifedipine	Begins to work in <30 min Lasts 4-5 hr	10-20 mg PO q30min to a maximum of 50 mg	10-20 mg PO tid of short-acting nifedipine or 30-120 mg once daily of long-acting formulation
Hydralazine	Begins to work in 10-20 min Lasts for 3-6 hr	2.5-10 mg IV q30min	Start at 10 mg PO qid; can be gradually increased to 50 mg PO qid

BP = blood pressure; D_5W= dextrose 5% in water (solution); LR = lactated Ringer's (solution); NS = normal saline.
From Rosene-Montella K, Keely EJ, Lee RV, Barbour LA, eds. *Medical Care of the Pregnant Patient.* 2nd ed. Philadelphia: ACP Press/American College of Physicians; 2008.

severe and nonsevere preeclampsia. Phenytoin and benzodiazepines should not be used for eclampsia prophylaxis or treatment unless there is a contraindication to magnesium sulfate or it is ineffective. There is evidence from two randomized controlled trials that magnesium is superior to phenytoin for the prevention of both primary seizures and recurrent seizures in eclampsia.

Treatment of acute seizures in eclampsia includes airway protection, fetal monitoring, magnesium, BP control, and benzodiazepines as needed to stop seizures acutely. Treatment of severe hypertension is as outlined in Table 239-6.

Severe maternal manifestations of preeclampsia that may warrant early delivery include seizure, renal failure, severe hypertension, severe thrombocytopenia or hemolysis, aspartate transaminase or alanine transaminase elevation of more than two to three times normal, pulmonary edema, retinal hemorrhage, and other symptoms suggestive of end-organ damage (headache, visual disturbance, epigastric or right upper quadrant pain). Fetal indications for delivery may include significant IUGR, oligohydramnios, and nonreassuring fetal testing. Women with preeclampsia before 34 weeks' gestation should receive a corticosteroid that crosses the placenta, such as betamethasone or dexamethasone, to accelerate fetal lung maturation.

PREVENTION

Multiple trials of antihypertensives, antioxidant supplementation with vitamins C and E, magnesium, protein or salt restriction, fish oil, and other dietary changes have failed to prevent preeclampsia. Low-dose aspirin in high-risk populations is the only intervention with data to support a positive effect. Initial trials showed limited effects, but a subsequent meta-analysis found that low-dose aspirin (<100 mg/day) decreases both the risk of preeclampsia and fetal and neonatal deaths. We recommend aspirin 81 mg/day in all patients with risk factors for preeclampsia. Trials of calcium supplementation have shown conflicting results, but given the inverse relationship between dietary calcium intake and BP in the general population, calcium supplementation of at least 1 g/day is recommended for women with a low dietary intake of calcium (<600 mg/day). An alternative to supplementation may be to increase dietary calcium by eating three or four servings per day of dairy products (assuming 250 to 300 mg of calcium per serving).

PROGNOSIS

Women who have had preeclampsia are at increased risk of heart disease, stroke, and cardiovascular death. Preeclampsia is also a marker for increased risk of end-stage renal disease. One year after delivery, preeclampsia patients observed longitudinally by Smith and colleagues had evidence of increased insulin resistance, BP, cholesterol, and triglycerides, which may be the first manifestations of the metabolic syndrome. The same group recently reported on the calculated 10-year, 30-year, and lifetime risk of cardiovascular disease in this cohort.[10] The 10-year, 30-year, and lifetime risk for development of cardiovascular disease was 18.2% versus 1.7%, 31.3% versus 5.1%, and 41.4% versus 17.8% in matched controls who did not have preeclampsia. It is unclear whether there is a shared pathogenesis, an unmasking of already established disease, or a contribution to the development of disease. It is possible that preexisting abnormal endothelial function predisposes to renal and vascular disease later in life and is, in fact, the same abnormality that disturbs implantation, resulting in preeclampsia and fetal loss. It is also possible that preeclampsia itself contributes to the later development of disease.

Continuing care beyond 6 weeks post partum is strongly recommended. Women with a history of severe preeclampsia should be screened for preexisting hypertension, underlying renal disease, thrombophilia, and possibly secondary causes of hypertension. They should also be informed of the risk for preeclampsia in subsequent pregnancies, particularly if the birth interval is less than 2 years or more than 10 years. Women who are overweight should be advised to normalize their body mass index before another pregnancy and to reduce long-term risk. Both women with preexisting hypertension and those whose BP normalizes are likely to benefit from an overall assessment of cardiovascular risk that includes a lipid profile, smoking cessation, and early interventions to reduce risk.

DEEP VENOUS THROMBOSIS, PULMONARY EMBOLISM, AND THROMBOPHILIA
(Chapters 81, 98, and 176)

Pulmonary embolism (PE) is the leading medical cause of maternal mortality in the developed world. It was responsible for 30% of direct maternal deaths in the most recent U.K. Confidential Enquiry. Despite our best efforts, mortality rates from PE in pregnancy have not changed in more than 2 decades, and the incidence of PE in the United States is rising, likely owing to the increase in obesity and cesarean deliveries. Current strategies to reduce risk from PE must address the widespread use of appropriate prophylaxis, early detection of venous thromboembolism (VTE), and prompt, safe, and effective therapy.[11]

EPIDEMIOLOGY

More than half of the VTE events in women younger than 40 years occur in association with pregnancy. VTE is 10 times more common in pregnant than in nonpregnant women of comparable age. It occurs in 5 to 12 of 10,000 pregnancies ante partum and in 3 to 7 of 10,000 post partum. The risk of VTE with pregnancy is increased by the presence of additional risk factors, including prolonged bedrest, cesarean section, preeclampsia, three or more children, smoking, obesity, previous superficial thrombophlebitis, previous VTE, thrombophilia, and family history of VTE.

PATHOBIOLOGY

Pregnancy is a hypercoagulable state (Chapter 176) characterized by venous stasis, maternal prothrombotic imbalance in which activation of the coagulation system exceeds the fibrinolytic response progressively through the course of pregnancy, and endothelial disruption. Venous stasis occurs as a result of progesterone-induced venodilation early in pregnancy and is later increased by the compressive effects of the gravid uterus. Compression of the left iliac vein by the right iliac artery further increases venous stasis on the left, which may explain the finding that more than 90% of cases of deep venous thrombosis (DVT) in pregnancy occur in the left leg. Endothelial damage occurs with preeclampsia and with both vaginal and operative delivery, further contributing to VTE risk.

Genetic and Acquired Thrombophilias

A positive family history for VTE (possibly a marker for thrombophilia) or a known thrombophilia significantly increases the risk for VTE in pregnancy (see Chapter 176). The best-described genetic thrombophilias include deficiencies in protein C, protein S, and antithrombin III, all of which appear to have autosomal dominant inheritance with variable penetrance, and the presence of the single-gene mutations factor V Leiden and prothrombin G202010. Of these, antithrombin-deficient homozygotes (rare) and compound heterozygotes have the highest risk of VTE in pregnancy. The thrombophilias have also been associated with obstetric complications, including IUGR, abruption, both early and late pregnancy loss, and preeclampsia (early, severe, or recurrent). The antiphospholipid antibody syndrome is the major acquired thrombophilia for which there are compelling pregnancy data supporting a link with both thrombosis risk and obstetric complications as well as the use of thromboprophylaxis to prevent poor obstetric outcome (Chapter 176). A large international, multicenter trial, the Thrombophilia in Pregnancy Prophylaxis Study, is currently under way, randomizing patients with thrombophilias and adverse pregnancy outcomes to surveillance versus thromboprophylaxis. Preliminary data analysis has failed to demonstrate efficacy of thromboprophylaxis for the prevention of the adverse pregnancy outcomes described before, and the most recent American College of Obstetricians and Gynecologists technical bulletin recommends against prophylaxis for prevention of adverse pregnancy outcome in patients with thrombophilias other than the antiphospholipid antibody syndrome.[12]

DIAGNOSIS

The diagnosis of VTE during pregnancy is complicated by both normal pregnancy-related physiologic changes and the reluctance to use diagnostic imaging in pregnancy. Clinical signs are unreliable, and leg swelling and complaints of dyspnea are common during pregnancy, making it difficult to decide when to investigate for VTE. The finding that 90% of DVT occurs in the left leg led to the observation that the combination of symptoms in the left leg, calf circumference difference of 2 cm or more, and first-trimester presentation (when leg swelling is less likely) is highly predictive of DVT. Most DVTs occur ante partum, and events are evenly distributed throughout gestation. The majority of fatal PEs in most studies occurred in the postpartum period, so vigilance is required for a prolonged period after delivery. Diagnosis of DVT requires compression ultrasonography that includes the iliac veins and the inferior vena cava at the level of the liver (Chapter 81). It also requires repeated compression ultrasonography if study findings are normal but there is high pretest probability and continuing symptoms. In

patients with suspected iliac or pelvic vein thrombosis and normal findings on ultrasound studies, magnetic resonance imaging or magnetic resonance venography is recommended.

The diagnosis of PE is even more of a problem, given the frequency of dyspnea, the likelihood of normal oxygenation in young patients with no underlying cardiopulmonary disease, and the more invasive nature of diagnostic testing. Arterial blood gas analysis is not helpful; the A-a gradient was normal in 60% of pregnant patients with documented PE in a retrospective review done at two centers.

The radiation exposure from imaging required for the diagnosis of PE is well below that allowed by the National Commission on Radiation Protection (see Table 239-3), so testing should never be withheld out of concern for fetal exposure. Ventilation-perfusion (V/Q) scanning is still the diagnostic test of choice in most centers outside the United States. It is better validated in pregnancy, involves no administration of contrast material, and has good negative predictive value in normal scans and in low-probability scans when it is paired with leg studies. If V/Q is used, it must be understood that there is still a significant risk of PE in patients with scans interpreted to be "intermediate" and "indeterminate," so additional testing is required in those cases. Computed tomography angiography has replaced V/Q scanning in most U.S. centers on the basis of its use in the nonpregnant population. The technique is dependent on cardiac output and plasma volume, both of which are increased during pregnancy. This may lead to poor opacification of the vessels, causing artifacts to be read as filling defects or the failure to visualize clots, so the technique must be adjusted for pregnancy. It is a sensitive, cost-effective test that offers an alternative diagnosis in 25 to 40% of cases, and it is preferred if there is an abnormality on the chest radiograph. It is well tolerated and has a shorter breath-holding time than V/Q scans, so it is also preferred in unstable patients, especially if an alternative diagnosis is suspected. Computed tomography angiography exposes the maternal breast to 2 to 3.5 rad, and exposure of the breast to 1 rad increases the lifetime risk of breast cancer by 13%. The use of breast shields decreases this exposure by about 50% without compromising the integrity of the test, so breast shields are strongly recommended.

The role of D-dimer testing in pregnancy has not yet been elucidated because D-dimer is elevated during normal pregnancy. It may have some use for its negative predictive value, but studies are inadequate to recommend its use at this time.

TREATMENT

Rx

The safety of unfractionated heparin (UFH) and low-molecular-weight heparin (LMWH) (Chapter 38) for the fetus is well established, so they are the drugs of choice for both the treatment and the prevention of VTE. Warfarin is a teratogen that crosses the placenta and has been associated with fetal bleeding and central nervous system abnormalities later in gestation, so it is not used for this indication in pregnancy. Initial treatment of VTE in the pregnant patient is either intravenous UFH, followed by subcutaneous UFH or LMWH, or an initial adjusted dose of LMWH that is then continued; both are acceptable.

LMWH causes a lower incidence of heparin-induced thrombocytopenia (Chapter 172) and less osteoporosis, so, given the prolonged exposure during pregnancy, it is the preferred agent. The 2012 American College of Chest Physicians consensus guidelines now recommend LMWH as the preferred agent in pregnancy.[13] The same consensus conference suggests limiting the use of fondaparinux and parenteral direct thrombin inhibitors to patients with severe allergic reactions to heparin (e.g., heparin-induced thrombocytopenia) who cannot receive danaparoid. The use of oral direct thrombin and anti-Xa inhibitors is not recommended (Table 239-7). LMWH has increased bioavailability, but the ease of administration in pregnancy is mitigated by the need for twice-daily dosing and frequent monitoring. Dosing requirements increase with increasing gestation, so it is necessary to follow anti-Xa levels. Because LMWH has limited reversibility with protamine and because it has been associated with epidural hematomas in nonpregnant patients given spinal or epidural anesthesia, most centers recommend a switch to UFH by 34 to 36 weeks. This allows patients the option of epidural anesthesia for delivery, and in the event of an emergent delivery before holding of anticoagulation, UFH can be reversed with protamine. Specific treatment recommendations are outlined in Table 239-8.

PREVENTION

The overall recurrence risk of VTE during pregnancy ranges from 5 to 20%, depending in part on the circumstances of the index clot. Patients with a previous idiopathic VTE (while not pregnant) or a secondary VTE that occurred during a previous pregnancy or while taking oral contraceptives and patients with a positive family history of or identified thrombophilia are at the highest risk. The thrombophilias with the highest recurrence risk are the antiphospholipid antibody syndrome and homozygosity or compound heterozygosity with more than one mutation and antithrombin deficiency. Thromboprophylaxis is required in all these groups, and its intensity is related to the severity of the risk (see Table 239-8). The lowest risk of recurrence, based on retrospective data, appears to be in patients without a family history of thrombophilia in whom the previous VTE occurred in association with a transient risk factor other than pregnancy or oral contraceptive use. The American College of Chest Physicians consensus guidelines now suggest surveillance with postpartum thromboprophylaxis in this group, but many U.S. centers would also offer antepartum thromboprophylaxis. Prophylaxis should be continued for at least 6 to 8 weeks post partum, when the hemostatic changes of pregnancy return to prepregnant values. Additional groups that should be considered for thromboprophylaxis are patients who have undergone cesarean section, particularly if they have an additional risk factor for VTE, and patients on prolonged bedrest. Patients receiving continued prophylactic or treatment doses post partum have the option of switching to warfarin, which is safe in breast-feeding women.

PROGNOSIS

VTE during pregnancy may be the first manifestation of a hypercoagulable state, as pregnancy acts as a "stress test" for thrombophilia. Fifty percent of initial episodes of VTE in women younger than 40 years are manifested in association with pregnancy. A thrombophilia evaluation is indicated in all patients who present with VTE during pregnancy to assess long-term maternal and family risk and to guide future secondary prophylaxis recommendations. Patients with an identified thrombophilia and an adverse pregnancy outcome may be at risk for a similar outcome in a subsequent pregnancy and should be counseled about this risk and considered for thromboprophylaxis.

Patients who have had DVT during pregnancy have a high risk of postphlebitic syndrome and venous insufficiency. Two randomized controlled trials demonstrated a 50% risk reduction in symptoms of post-thrombotic syndrome when compression stockings were used within 1 month of diagnosis and continued for a minimum of 1 year after diagnosis.

ASTHMA (Chapter 87)

Maintaining adequate control of asthma during pregnancy is important for both maternal and fetal outcome. Asthma may be associated with increased perinatal mortality, preterm birth, IUGR, gestational diabetes, and preeclampsia. Well-controlled asthma reduces the likelihood of these adverse outcomes to baseline, so it is safer for both mother and fetus to treat maternal asthma than to allow exacerbations to occur.

EPIDEMIOLOGY

Asthma is the most common respiratory disease in pregnancy. It affects 3.7 to 8.4% of pregnancies in the United States and 12 to 13% of pregnancies in Australia and the United Kingdom. Approximately 10% of U.S. women of reproductive age have asthma, and the rates of asthma reported during labor and delivery have doubled during the last decade.

Effect of Pregnancy on Asthma

The course of asthma in pregnancy is unpredictable, and most studies have found that a third of patients improve, a third worsen, and a third stay the same. The most likely predictor in any individual patient is her course during a previous pregnancy. In most studies, the majority of exacerbations occurred between 17 and 32 weeks, with some improvement reported by 36 weeks' gestation. Patients with mild asthma do well during labor and delivery, but almost 50% of patients with severe asthma worsen during labor and delivery. Risk factors for exacerbations include severe asthma, poor compliance with medications (especially inhaled corticosteroids), obesity, viral infections, rhinitis, gastroesophageal reflux, and poor prenatal care. The highest morbidity and mortality are reported in African American patients.

Effect of Asthma on Pregnancy

Pregnancy and perinatal outcome are improved when asthma is well controlled. Poorly controlled asthma increases the risk of spontaneous abortion,

TABLE 239-7 SAFETY AND PHARMACOKINETICS OF ANTICOAGULANTS IN PREGNANCY

	UFH	LMWH	SEMISYNTHETIC HEPARINOIDS (DANAPAROID)	SYNTHETIC HEPARINS AND FACTOR Xa INHIBITOR (FONDAPARINUX, RIVAROXABAN)	THROMBIN INHIBITORS (RECOMBINANT HIRUDINS)	THROMBIN INHIBITORS (ARGATROBAN, DABIGATRAN)	WARFARIN (COUMADIN)
Monitoring	aPTT	Anti-Xa level	Anti-Xa level	Anti-Xa level	aPTT	aPTT	INR
Half-life	1.5 hr	Enoxaparin: 4.5-7 hr Tinzaparin: 3-4 hr Dalteparin: 3-5 hr All prolonged in renal impairment	24 hr Prolonged in severe renal impairment	17-21 hr Prolonged in severe renal impairment	Lepirudin: 1.3 hr Bivalirudin: 25 min Desirudin: 2 hr All prolonged in renal impairment	Argatroban: 39-51 min Dabigatran: 12-17 hr	20-60 hr
Clearance	Liver, reticuloendothelial system	Liver 40% urine excretion	Plasma Urine excretion	Metabolism unknown Urine excretion	Lepirudin: metabolism unknown; 48% urine excretion Bivalirudin: plasma (80%); urine (20%) excretion Desirudin: kidney	Argatroban: liver; urine and fecal excretion Dabigatran: liver; urine excretion	Liver 92% urine excretion
Safety	Does not cross the placenta No known risk of teratogenicity	Enoxaparin, tinzaparin, and dalteparin: do not appear to cross the placenta and are not believed to increase risk of birth defects on the basis of animal studies and some human studies outcomes.	No longer available in the United States Many case reports of use of danaparoid in pregnancy in various doses and duration have shown successful pregnancy outcomes.	Fondaparinux: on the basis of experimental animal studies, use of fondaparinux in pregnancy is not expected to increase the risk of malformations. Small amounts cross the placenta, but clinical significance of such is unknown. Rivaroxaban: postimplantation pregnancy loss, increased fetal toxicity, and maternal hemorrhagic complications have been observed in animal studies. There are no adequate and well-controlled human studies.	Lepirudin: on the basis of experimental animal studies, it is not expected to increase the risk of congenital malformations, although it is known to cross rat placenta. Case reports of its use during various times of pregnancy did not show adverse events in exposed neonates. Bivalirudin: no epidemiologic studies of congenital anomalies among infants born to women treated with bivalirudin during pregnancy have been reported. Desirudin: teratogenic effects were observed in some animal reproductive studies.	Argatroban: did not produce malformations in rats and rabbits, but dosing was low compared with human therapeutic dose levels. Few case reports describing its use during pregnancy with no adverse newborn outcomes. Dabigatran: adverse events were observed in some animal reproductive studies. There are no adequate and well-controlled studies in pregnant women.	Warfarin crosses the placental barrier. Risk of birth defect with early exposure; potential for fatal hemorrhage to the fetus in utero.
Lactation	Safe	Enoxaparin: excretions in milk unknown Tinzaparin, dalteparin: no data available	Little or no danaparoid appears in breast milk and would likely be inactivated in infant's stomach	Fondaparinux: appears in rat milk. Possible adverse effects of exposure through milk have not been described. Rivaroxaban: no data available	Lepirudin: in one case report, it was used during lactation without adverse events. It was not detectable in milk. Bivalirudin and desirudin: no data available	Argatroban, dabigatran: no data available	Safe
Administration	SC and IV	SC and IV	SC and IV	Fondaparinux: IV Rivaroxaban: PO	IV	Argatroban: IV Dabigatran: PO	PO

aPTT = activated partial thromboplastin time; INR = international normalized ratio; LMWH = low-molecular-weight heparin; UFH = unfractionated heparin.
From Mazer J, Zouein J, Bourjeily G. Treatment of pulmonary embolism in pregnancy. *US Respir Dis.* 2012;8:30-35.

TABLE 239-8 TREATMENT OF VENOUS THROMBOEMBOLISM IN PREGNANCY: ANTEPARTUM ANTICOAGULATION

	PROPHYLAXIS		AGGRESSIVE PROPHYLAXIS		
DRUG	***First 20 Weeks***	***20-37 Weeks***	***First 20 Weeks***	***20 Weeks to Term***	**FULL TREATMENT**
Dalteparin	5000 U/day	5000 U q12h	100 U/kg/day	100 U/kg/day	100 U/kg q12h with anti-Xa monitoring
Enoxaparin	30 mg/day	30 mg q12h	1 mg/kg/day	1 mg/kg/day	1 mg/kg q12h with anti-Xa monitoring
Tinzaparin	4500 U/day	4500 U q12h	88 U/kg/day	88 U/kg/day	88 U/kg q12h with anti-Xa monitoring
Heparin	Alternative if LMWH unaffordable: 750 U bid first 20 wk; 10,000 U bid wk 20-37		Alternative if LMWH unaffordable: 10,000 U bid to achieve anti-Xa level of 0.1-0.3 U/mL		Adjusted to mid-interval anti-Xa of 0.35-0.67 with q12h SC injections

LMWH = low-molecular-weight heparin.
Modified from Bourjeily G, Rosene-Montella K, eds. Venous thromboembolism in pregnancy. In: *Pulmonary Problems in Pregnancy, Respiratory Medicine.* New York: Humana Press; 2009.

low birthweight, IUGR, and cesarean section. Preterm delivery, gestational diabetes, and preeclampsia have also been associated with poorly controlled asthma, but it is unclear how systemic steroids contribute to these complications. Systemic steroids have been associated with an increased risk of premature rupture of membranes, preeclampsia, prematurity and low birthweight, and gestational diabetes. A retrospective study suggested that some complications may increase even in patients with mild asthma or asthma in good control.

PATHOBIOLOGY

The normal physiologic changes of pregnancy may contribute to variations in asthma severity. Factors contributing to the worsening of asthma include gastroesophageal reflux disease and rhinitis or sinusitis, triggers for asthma that are common during pregnancy. Gastroesophageal reflux may be manifested initially during pregnancy or worsen in patients with preexisting reflux owing to both hormonal and mechanical effects. Progesterone acts as a smooth muscle dilator that reduces lower esophageal sphincter pressure and contributes to delayed gastric emptying. Later in gestation, uterine enlargement further contributes to gastric displacement and increased reflux. Rhinitis and sinusitis clearly contribute to asthma exacerbations in nonpregnant patients. Gestational rhinitis related to hormonal effects is present in most pregnant women, and its behavior seems to parallel that of asthma. Bacterial sinusitis is five to six times more common in pregnancy and should be treated aggressively.

Hormonal effects on the airway may also contribute to asthma status. There is a progressive increase in serum cortisol and estradiol, which affects the quality of mucus production, and in progesterone, which decreases smooth muscle contractility and thereby causes airway dilation and improves minute ventilation. Immunologic factors during normal pregnancy may also contribute to the course of asthma. There is a suppression of cell-mediated immunity, with a predominant T_H2 environment and high interleukin-5 and tumor necrosis factor messenger RNA. In pregnant women with asthma (not receiving inhaled corticosteroid therapy), the T_H2/T_H1 ratio is even higher, possibly contributing to exacerbations.

The mechanism by which asthma exacerbations affect perinatal outcome is probably related to chronic maternal hypoxia, with consequent placental dysfunction and decreased uteroplacental flow, which contributes to decreased fetal growth. Poorly controlled asthma increases low birthweight 2.5-fold. Relative placental ischemia in asthma, particularly in disease that was poorly controlled before conception, is likely the link to an increased risk for preeclampsia. Placentas from women with asthma show a change in response to vasodilators and constrictors in vitro, similar to that seen in preeclampsia.

DIAGNOSIS

The diagnosis of asthma during pregnancy is the same as in the nonpregnant state (Chapter 87): normal forced expiratory volume in 1 second (FEV_1)/forced vital capacity on baseline pulmonary function tests with an obstructive physiology during exacerbations that is reversible either spontaneously or with medications. Airway hyperresponsiveness, as demonstrated by a methacholine challenge causing a 20% drop in FEV_1 from baseline, is also useful for diagnosis in pregnancy. Asthma severity in pregnancy is classified the same as in nonpregnant patients by the new classification of asthma severity that incorporates short-acting β-agonist use. The new classification includes both the level of impairment (daytime and nighttime frequency, quality of life and interference with normal activities, lung function) and the risk of exacerbations based on frequency and severity of prior exacerbations.

Differential Diagnosis

Dyspnea of pregnancy is a benign condition that often occurs later in pregnancy and is characterized by an increased awareness of the work of breathing that is disturbing for many patients. It is not likely to be acute, occurs less with rest, and should not interfere with normal daily activities. Dyspnea of pregnancy should not be accompanied by an increase in respiratory rate, wheezing, or hypoxia. It is important to consider pulmonary edema in any pregnant patient complaining of shortness of breath (Chapter 69). Pregnancy-related causes of pulmonary edema and acute respiratory distress syndrome include tocolytics (drugs that slow contractions), preeclampsia, gastric aspiration, amniotic fluid embolism, sepsis (related to pyelonephritis, chorioamnionitis, endometritis, septic abortion), abruption, and obstetric hemorrhage. Cardiac causes should be suspected when pulmonary edema is manifested at the peak of blood volume (28 to 32 weeks), when occult valvular disease (Chapter 75) is most likely to be unmasked. Additional cardiac considerations are peripartum cardiomyopathy, preeclampsia, and ischemic heart disease, which in pregnancy may also be caused by coronary dissection.

TREATMENT

Rx

Management of asthma during pregnancy does not differ greatly from that of the nonpregnant patient (Chapter 87).[15] However, normal arterial carbon dioxide pressure ($PaCO_2$) in pregnancy is 28 to 32 mm Hg, so a tachypneic pregnant patient with a $PaCO_2$ above this range may be in impending respiratory failure. Minute ventilation in pregnancy increases by an increase in tidal volume, but respiratory rate is unchanged by pregnancy, so tachypnea is always an abnormal finding.

The goal of asthma therapy during pregnancy is to maintain adequate control to ensure maternal and fetal health. It is always safer for pregnant women with asthma to be treated with asthma medications than to experience symptoms and exacerbations. Careful monitoring during all prenatal visits, preferably with spirometry, and stepped-up therapy are required both for maternal asthma control and to ensure appropriate oxygenation of the fetus. Maternal arterial oxygen saturation should be maintained at 95% or more, or the arterial oxygen pressure (PaO_2) should be maintained at 80 mm Hg or more, to maintain fetal oxygenation. E-Figure 239-1 outlines the classification of asthma and care for pregnant patients with asthma. More detailed recommendations for the home management of exacerbations and for hospitalization and emergency care for pregnant patients can be found in the *National Asthma Education and Prevention Program Working Group Report for Managing Asthma During Pregnancy.*

Albuterol is the preferred short-acting β-agonist because it has an excellent safety profile and the most data related to safety during human pregnancy. Inhaled corticosteroids are the preferred medication for long-term control. Budesonide is the preferred inhaled corticosteroid solely because of the amount of reassuring data on its use in pregnant patients. There are, however, no adverse data on the other inhaled corticosteroids. Data on the effectiveness and safety of long-acting β-agonists during pregnancy are limited, although it is reasonable to assume that they have a safety profile similar to that of albuterol. Salmeterol is the preferred agent, based only on its longer availability and lack of reports of adverse outcomes in exposed pregnancies. Cromolyn has an excellent safety profile but has limited effectiveness compared with inhaled corticosteroids.

Minimal published reports are available on the use of leukotriene receptor antagonists during pregnancy; however, animal safety data are reassuring. Current guidelines do not recommend leukotriene receptor antagonists because of the limited data, unless a patient's asthma was well controlled with this type of drug before pregnancy.

Classify Severity: Clinical Features Before Treatment or Adequate Control			Medications Required to Maintain Long-Term Control
	Symptoms/Day / Symptoms/Night	PEF or FEV_1 / PEF Variability	**Daily medications**
Step 4 Severe Persistent	Continual / Frequent	≤60% / >30%	• Preferred treatment: –High-dose inhaled corticosteroid AND –Long-acting inhaled β_2-agonist AND, if needed, –Corticosteroid tablets or syrup long term (2 mg/kg per day, generally not to exceed 60 mg per day). (Make repeat attempts to reduce systemic corticosteroid and maintain control with high-dose inhaled corticosteroid.)
			• Alternative treatment: –High-dose inhaled corticosteroid AND –Sustained-release theophylline to serum concentration of 5–12 μg/mL.
Step 3 Moderate Persistent	Daily / >1 night/week	>60%–<80% / >30%	• Preferred treatment: EITHER –Low-dose inhaled corticosteroid and long-acting inhaled β_2-agonist OR –Medium-dose inhaled corticosteroid If needed (particularly in patients with recurring severe exacerbations): –Medium-dose inhaled corticosteroid and long-acting inhaled β_2-agonist
			• Alternative treatment: –Low-dose inhaled corticosteroid and either theophylline or leukotriene receptor antagonist. If needed: –Medium-dose inhaled corticosteroid and either theophylline or leukotriene receptor antagonist.
Step 2 Mild Persistent	>2 days/week but <daily / >2 nights/month	≥80% / 20%–30%	• Preferred treatment: –Low-dose inhaled corticosteroid
			• Alternative treatment (listed alphabetically): cromolyn, leukotriene receptor antagonist OR sustained-release theophylline to serum concentration of 5–12 μg/mL.
Step 1 Mild Intermittent	≤2 days/week / ≤2 nights/month	≥80% / <20%	No daily medication needed.
			• Severe exacerbations may occur, separated by long periods of normal lung function and no symptoms. A course of systemic corticosteroid is recommended.
Quick Relief All Patients	• Short-acting bronchodilator: 2–4 puffs short-acting inhaled β_2-agonist as needed for symptoms. • Intensity of treatment will depend on severity of exacerbation; up to 3 treatments at 20-minute intervals or a single nebulizer treatment as needed. Course of systemic corticosteroid may be needed. • Use of short-acting inhaled β_2-agonist >2 times a week in intermittent asthma (daily, or increasing use in persistent asthma) may indicate the need to initiate (increase) long-term control therapy.		

E-FIGURE 239-1. Stepwise approach for management of asthma during pregnancy and lactation. FEV_1 = forced expiratory volume in 1 second; PEF = peak expiratory flow. (From National Asthma Education and Prevention Program Working Group Report on Managing Asthma During Pregnancy. *Recommendations for Pharmacologic Treatment—Update 2004. NIH Publication No. 05-5236.* U.S. Department of Health and Human Services; National Institutes of Health; National Heart, Lung, and Blood Institute; March 2005.)

Intranasal corticosteroids are recommended for the treatment of allergic rhinitis (Chapter 251), given the limited systemic effect. The current nonsedating antihistamines of choice are loratadine and cetirizine.

Patients at risk for fatal asthma are those with a large bronchodilator response, overreliance on short-acting bronchodilators, marked circadian variation in lung function, history of hospitalization or intubation, and frequent systemic steroid use. There are specific considerations based on pregnancy physiology in pregnant patients who may require airway intubation. Pregnant patients have a low functional residual capacity and oxygen reserve, a more profound response to sedatives, airway edema, and larger airways. Intubation failure is much higher in pregnant women, so intubation should be performed by the most experienced professional available.

Breast-feeding should be encouraged in all patients with asthma as there is some evidence that it decreases atopy in offspring. Data are conflicting regarding the development of asthma in offspring.

DIABETES (Chapter 229)

Diabetes affects 1.85 million women of reproductive age, and it is estimated that preconception management could reduce the risk for 113,000 births per year. All women of reproductive age with diabetes should be counseled about the relationship between glucose control and congenital anomalies. Hyperglycemia is a teratogen, and the incidence of congenital anomalies is directly related to the hemoglobin A_{1c} level at conception (Fig. 239-3). The anomaly rate was as high as 11% in women without preconception care, including cardiac anomalies, neural tube defects, and sacral agenesis. The single most important contribution an internist can make to the prevention of congenital anomalies is to address pregnancy risk with all women of childbearing age with diabetes. The responsibility to normalize hemoglobin A_{1c} before conception falls to the medical care provider; once pregnancy is diagnosed and the patient is seen by her obstetrician, the teratogenic effects of glucose have already occurred.

DEFINITION

Gestational diabetes mellitus (GDM) is defined as glucose intolerance that first occurs or is first identified during pregnancy. Either type 1 or type 2 diabetes in a pregnant patient is referred to as preexisting or pregestational diabetes.

EPIDEMIOLOGY

The frequency of GDM is rising in the United States; it now occurs in 4 to 14% of all pregnancies, depending on the patient's characteristics. The epidemic of type 2 diabetes has resulted in a higher prevalence at a younger age; in the United States, there has been a 70% increase in the prevalence of diabetes in the 30- to 39-year-old age group versus 33% overall. The proportion of women with type 2 versus type 1 pregestational diabetes has also increased, from 26% in 1980 to 65% in 2000, and it is still increasing. The perinatal morbidity and mortality associated with type 2 diabetes is at least as great as that associated with type 1 during pregnancy.

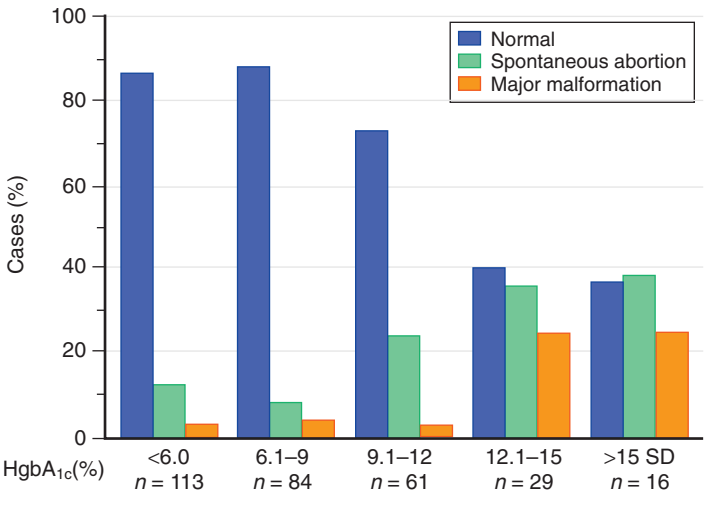

FIGURE 239-3. Relationship of hemoglobin A_{1c} ($HgbA_{1c}$), congenital anomalies, and spontaneous abortion.

PATHOBIOLOGY

Type 1 diabetes is caused by autoimmune destruction of pancreatic beta cells, resulting in an absolute insulin deficiency. Ninety percent of cases are diagnosed before the age of 25 years and are often associated with other autoimmune illnesses or a family history of autoimmune illness, including thyroid disorders, Addison's disease, and celiac disease. Type 2 diabetes is part of the metabolic syndrome, which includes insulin resistance, hyperinsulinemia, dyslipidemia, abdominal obesity, and hypertension with premature atherosclerosis; it probably has a genetic component. GDM may also be a manifestation of the metabolic syndrome, unmasked by the insulin-resistant state of pregnancy. Patients with GDM have a 50% chance for development of type 2 diabetes in the ensuing 5 to 10 years and a long-term risk of approximately 70%.

Pregnancy is a state of accelerated starvation and marked insulin resistance. Lower fasting glucose levels are seen early in the first trimester, and nocturnal hypoglycemia is common. There is blunted hypoglycemic awareness due to decreased epinephrine and norepinephrine release, with falls in blood glucose concentration and increased ketogenesis, resulting in an increased risk for diabetic ketoacidosis. Insulin requirements may decrease 20% in weeks 7 to 12 but then gradually rise later in gestation, such that insulin doses need to be increased by 16 weeks' gestation. Marked insulin resistance is related to the presence of elevated levels of cortisol, prolactin, human placental lactogen, and human placental growth hormone. Insulin sensitivity is decreased by about 50% in the third trimester, resulting in increased serum insulin and postprandial glucose levels, explaining the timing of the onset of GDM.

DIAGNOSIS

The diagnosis of pregestational diabetes is based on the finding of a fasting blood glucose level of more than 125 mg/dL or a 2-hour or random blood glucose level of 200 mg/dL or more. The American Diabetes Association recently added a hemoglobin A_{1c} level of 6.5% or higher as an acceptable alternative diagnostic criterion. GDM is based on the results of a blood glucose screen after a 50-g glucose challenge test (≥140 mg/dL), followed by a 3-hour confirmatory 100-g oral glucose tolerance test. Positive results are any two of the following: fasting, 95 mg/dL or more; 1 hour, 180 mg/dL; 2 hours, 155 mg/dL; and 3 hours, 140 mg/dL. It may be difficult to distinguish between GDM and type 2 diabetes that was not diagnosed before pregnancy. Elevated fasting glucose levels before 24 weeks' gestation and elevated hemoglobin A_{1c} are both suggestive of type 2 diabetes. Any diabetes diagnosed during pregnancy is referred to as GDM; if it persists post partum, it is reclassified as type 2.

TREATMENT

Rx

Dietary recommendations are for 30 kcal/kg, with 40 to 50% carbohydrates, divided into three meals and three snacks. Aerobic exercise may decrease insulin resistance and reduce maternal glucose levels and may be an effective adjunct to diet in patients with GDM. Adjustment of medications should include the discontinuation of ACE inhibitors, ARBs, and statins and the institution of folate and prenatal vitamins. An assessment of baseline disease status in all patients with diabetes should include hemoglobin A_{1c}, ophthalmologic examination, electrocardiogram, assessment of urine protein excretion, serum creatinine concentration, and thyroid-stimulating hormone level. Baseline preeclampsia laboratory studies are also recommended.

The fetal assessment plan includes ultrasound to confirm dates and viability; this is done by the patient's obstetric care provider. There will also be a quad screen, looking for serum markers that suggest neural tube defects or Down syndrome; an ultrasound nuchal translucency and a level 2 ultrasound to assess for congenital anomalies; and a fetal echocardiogram.

Control of Blood Glucose

The goal of treatment for pregestational diabetes is the best hemoglobin A_{1c} level possible, without excessive hypoglycemia. This is achieved by frequent insulin adjustments and self-monitoring of blood glucose level at least four times a day. The specific goals for both GDM and pregestational diabetes are fasting glucose concentration of 65 to 95 mg/dL, 1-hour postprandial glucose concentration of less than 140 mg/dL, and 2-hour glucose concentration of less than 120 mg/dL. This is best achieved by multiple doses of insulin, including basal, intermediate, and long-acting insulin, with preprandial bolus rapid-acting insulin to cover the anticipated carbohydrate load. This regimen requires self-monitoring of blood glucose levels six or seven times a day, so it may be difficult to comply with. Further, the risk of serious

hypoglycemia may limit these glycemic targets, particularly in women with type 1 diabetes. Extreme vigilance is required to avoid serious hypoglycemia, particularly in weeks 7 to 12, when insulin requirements are the lowest during pregnancy.[16] In patients with decreased glycemic awareness, the risk of nighttime hypoglycemia is significant, and patients' partners should be counseled about this risk.

Insulin analogues are being used with increasing frequency. Of the rapid-acting analogues, there are data that lispro does not cross the placenta; there are no data yet on aspart. Although long-acting glargine is being used in pregnancy, its placental transfer is unknown, and there are theoretical concerns about its binding to the insulin-like growth factor receptor and its mitogenic potential.

Of the oral agents for which there are data, glyburide is safe and efficacious in women with GDM, and it is more efficacious than metformin in this setting.[■] Oral agents are less useful with significant insulin resistance and type 2 diabetes, so insulin continues to be the "gold standard" in this group. Metformin does not appear to increase the risk of congenital anomalies or spontaneous abortion, but it does cross the placenta. Women with polycystic ovary syndrome treated with metformin may regain their fertility and should be advised to use contraception. Recent trials support the safety of metformin in the second and third trimesters. In the United States, metformin is not yet recommended for the treatment of type 2 diabetes in pregnancy or for GDM. There are inadequate pregnancy data for the meglitinides and glitazones.

Maternal and Fetal Monitoring

During pregnancy, increased vigilance and continual assessment for the development of complications, including hypertension, preeclampsia, worsening nephropathy, and retinopathy, are required. The incidence of retinopathy in pregnant women with type 1 and type 2 diabetes is 34 to 50% and 3 to 5%, respectively. Nephropathy is found in 4% of diabetic pregnancies and is associated with increased maternal and perinatal morbidity. Most studies agree that pregnancy may accelerate the progression of nephropathy, but the reversibility of this complication is unclear. Most studies show a worsening of retinopathy in pregnant patients similar to that occurring during the same period in nonpregnant patients. The level of severity of retinopathy before pregnancy is most predictive of worsening during pregnancy, and treatment is recommended before conception.

Labor and Delivery

During labor and delivery, tight glucose control is necessary to avoid neonatal hypoglycemia due to hyperinsulinemia at birth. The goal is to maintain serum glucose levels of 72 to 144 mg/dL. The use of an insulin drip with dextrose infusion in active labor is recommended. Immediately after delivery, insulin requirements decrease to prepregnancy levels. Insulin needs should be one half to two thirds prepregnancy requirements, with a further decreased need found in breast-feeding patients.

Postpartum Considerations

Fifteen percent to 25% of type 1 diabetics develop postpartum thyroiditis, so all patients require postpartum measurements of thyroid-stimulating hormone and follow-up for 6 months. Other postpartum recommendations are to restart ACE inhibitors and to monitor closely for infection. It is important to discuss diabetes prevention in the offspring and to address contraception. It is most important to recommend an effective contraceptive method that is acceptable to the patient. Oral estrogens can increase triglycerides, and both oral and injectable progesterone-only contraceptives may increase insulin resistance. Low-dose combined oral contraceptives and the progesterone-releasing intrauterine device appear to have little effect on glucose.

Insulin is acceptable in breast-feeding women, and limited data suggest that glyburide and metformin are safe as well. One small study found that glyburide is not excreted in breast milk and that metformin is excreted in a small amount that is probably not clinically significant.

PROGNOSIS AND COMPLICATIONS

The maternal complications of diabetes may be affected by pregnancy and may affect the course of the pregnancy. Patients with nephropathy experience an increase in proteinuria and a risk for progression of renal disease, especially if the serum creatinine concentration is more than 1.4 mg/dL. There is an increased risk for hypertension, which is seen in 30% of patients during the first trimester and 75% of patients by the third trimester. Autonomic neuropathy may worsen, as manifested by increasing gastroparesis, orthostatic hypotension, and decreased hypoglycemic awareness. Patients with long-standing diabetes may need to be evaluated for ischemic heart disease, which may impair the heart's ability to meet the cardiovascular demands of pregnancy. Pregnant patients are also at risk for hyperlipidemia and for preeclampsia. Diabetes increases the risk for operative delivery and for infections, the most common of which are wound, urinary tract, and respiratory.

Diabetic ketoacidosis may be precipitated by steroid use for fetal lung maturity, hyperemesis, infection, and noncompliance with insulin regimens. Acidosis may occur more quickly and at lower glucose levels in pregnant than in nonpregnant patients. There is a high fetal mortality associated with diabetic ketoacidosis (9 to 10%), and patients should be monitored in an intensive care setting.

Fetal and Neonatal Effects (Fig. 239-4)

There is an increased risk of spontaneous abortion, fetal loss, congenital anomalies, preeclampsia, and preterm delivery in patients with diabetes. Poor

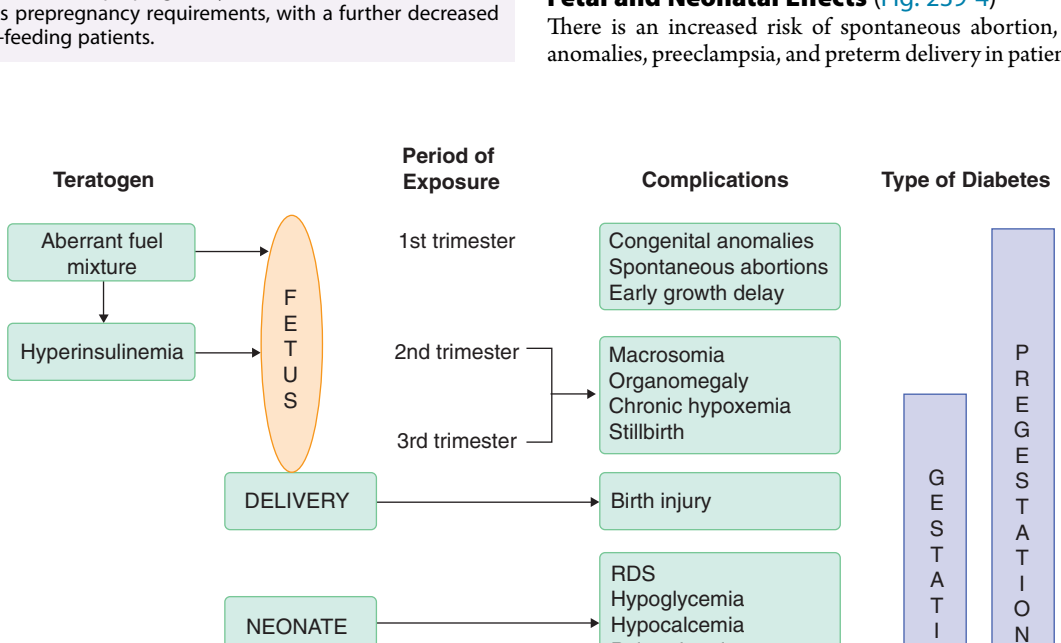

FIGURE 239-4. Fetal, neonatal, and childhood effects of exposure to hyperglycemia. CNS = central nervous system; RDS = respiratory distress syndrome.

glycemic control during pregnancy, especially in type 2 diabetes and GDM, is also associated with macrosomia (baby weighing >4000 g) and fetal intraventricular septal hypertrophy. Poor control is also associated with the effects of maternal vascular and renal disease and ketoacidosis, which include fetal loss, preeclampsia, and low birthweight. A study looking at differences in causes of pregnancy loss in type 1 and type 2 diabetic mothers compared the placental histology of patients with type 1 and type 2 diabetes and found an increase in histologic infarcts in type 2, suggesting a vascular rather than a glycemic cause of pregnancy complications and signs of abnormal development of placentas from patients with type 1 disease.[17]

Neonatal complications include respiratory distress syndrome, hypoglycemia, hypocalcemia, cardiac hypertrophy, hyperbilirubinemia, and polycythemia. The risk of hypoglycemia may be ameliorated by careful control of maternal glucose concentration during labor and delivery. Normalization of maternal glucose concentration prevents hyperinsulinemia in the fetus and mitigates the risk of neonatal hypoglycemia.

Maternal Effects

GDM is a marker for type 2 diabetes; 50% of patients will develop type 2 diabetes within 7 to 10 years, and overall, 70% will develop the disease. Patients who have had GDM need directed testing at their 6-week postpartum visit, annual screening, and recommendations for lifestyle modification and cardiovascular risk reduction. Offspring of patients with GDM and type 2 diabetes are at increased risk for obesity and glucose intolerance.

LIVER DISEASE IN PREGNANCY

Liver disease found during pregnancy may be unique to pregnancy, represent underlying liver disease unmasked during pregnancy, or develop during pregnancy. Most liver function test results are unchanged by pregnancy with the exception of an increase in alkaline phosphatase (which is produced by the placenta), an increase in fibrinogen, and a decrease in serum albumin.

Table 239-9 outlines pregnancy-related liver disease. This discussion focuses on diseases unique to pregnancy and those diseases for which there are specific management considerations during pregnancy.

Liver Diseases Unique to Pregnancy

INTRAHEPATIC CHOLESTASIS OF PREGNANCY (OBSTETRIC CHOLESTASIS)

DIAGNOSIS

Intrahepatic cholestasis of pregnancy (ICP) or obstetric cholestasis affects 0.5 to 2% of pregnant women, although in Bolivia and Chile, rates of 4 to 28% have been observed. ICP is manifested most commonly in the late second or third trimester with intense pruritus accompanied by an increase in serum bile acids and often transaminases and prothrombin time. ICP is likely to be a metabolic disease of multifactorial etiology characterized in most cases by a genetic variation in the biliary transporters and receptors that govern bile acid homeostasis.[18] It is more common in patients with underlying hepatitis C, so it is important to screen all patients with hepatitis C serology. ICP is associated with preterm labor, meconium staining, fetal hypoxia, and sudden fetal demise.

TABLE 239-9 LIVER DISEASE AND PREGNANCY

UNIQUE TO PREGNANCY
Acute fatty liver of pregnancy
HELLP syndrome (hemolysis, elevated liver enzymes, and low platelets)
Hyperemesis gravidarum
Intrahepatic cholestasis of pregnancy
Preeclampsia and eclampsia
INCREASED INCIDENCE DURING PREGNANCY
Budd-Chiari syndrome
Drug-induced hepatotoxicity
Gallstones
Liver transplantation
Sepsis
Viral hepatitis
UNDERLYING CONDITION THAT MAY BE REVEALED
Autoimmune hepatitis
Cirrhosis
Hepatitis B and C
Primary biliary cirrhosis
Primary sclerosing cholangitis
Wilson's disease

Modified from Mufti AR, Reau N. Liver disease in pregnancy. *Clin Liver Dis.* 2012;16:247-269.

TREATMENT Rx

Because no antenatal testing has been able to predict those patients at risk for fetal demise, consensus guidelines recommend delivery at 37 to 38 weeks of gestation in patients with a significant elevation in serum bile acids (≥40 µmol/L). The treatment of choice is ursodeoxycholic acid at 10 to 15 mg/kg, which leads to both symptomatic and biochemical improvement. There are no studies demonstrating a beneficial effect on pregnancy outcome.

PREECLAMPSIA AND ECLAMPSIA

As discussed before, preeclampsia can be associated with liver abnormalities including liver edema and infarction, subcapsular hematoma, liver laceration, and HELLP syndrome (hemolysis, elevated liver enzymes, and low platelets).

HYPEREMESIS GRAVIDARUM

Hyperemesis gravidarum, defined as severe persistent nausea and vomiting of pregnancy with weight loss, ketosis, or dehydration, can be associated with transaminase elevations in 50 to 60% of cases.

TREATMENT Rx

Treatment is supportive with rehydration, antiemetics, and vitamin replacement,[19] often requiring hospitalization. Oral fluid repletion can often be accomplished as indwelling intravenous catheters have been associated with a significant risk of thrombosis and of infections. Hyperemesis gravidarum is usually a reversible condition with no permanent liver damage.

ACUTE FATTY LIVER OF PREGNANCY

DIAGNOSIS

Acute fatty liver of pregnancy is a rare condition that is estimated to occur in 5 per 100,000 pregnancies. It most commonly is manifested in the third trimester and post partum; in its most severe form, it can be associated with fulminant liver failure and the need for liver transplantation. Maternal mortality rates were earlier thought to be as high as 20%, but a more recent U.K. study found a 2% maternal and 11% perinatal mortality rate. The Swansea criteria have recently been validated as a diagnostic tool for the diagnosis of acute fatty liver of pregnancy (Table 239-10).

PROGNOSIS

The maternal mortality rate has been estimated at 18%. Acute fatty liver of pregnancy is associated with an inherited defect in mitochondrial fatty acid B oxidation. The defect results in accumulation of toxic metabolites produced by the fetus and placenta that, after entering the maternal circulation, are deposited in maternal liver. Diagnosis is clinical, based on the constellations of findings described in Table 239-10. Liver function abnormalities can be severe, and hypoglycemia is a poor prognostic sign. Anyone with evidence of liver failure should be seen at a transplant center as early in the course as possible.

TABLE 239-10 SWANSEA DIAGNOSTIC CRITERIA FOR ACUTE FATTY LIVER OF PREGNANCY

Six or more of the following features in the absence of another explanation:

- Vomiting
- Abdominal pain
- Polydipsia/polyuria
- Encephalopathy
- Elevated bilirubin
- Hypoglycemia
- Elevated urate
- Leukocytosis
- Ascites or bright liver on ultrasound scan
- Elevated transaminases
- Elevated ammonia
- Renal impairment
- Coagulopathy
- Microvesicular steatosis on liver biopsy

From Ch'ng CL, Morgan M, Hainsworth I, et al. Prospective study of liver dysfunction in pregnancy in Southwest Wales. *Gut.* 2002;51:876-880.

Preexisting or New-Onset Liver Disease during Pregnancy

The major importance of recognition of underlying liver disease in pregnancy is both for maternal health and for those diseases in which lack of treatment results in a high rate of vertical transmission to the fetus or neonate.

VIRAL HEPATITIS

Hepatitis C virus (HCV) infection (Chapter 148) has become an increasingly important and prevalent issue in pregnancy; maternal to child HCV vertical transmission rates are 5 to 10% and as high as 22% with HIV coinfection. New screening guidelines will lead to more universal screening in this age group. The mode of delivery does not have an impact on vertical transmission rates. The higher the viral load and the longer the duration of ruptured membranes, the higher the risk of transmission. Infection with hepatitis E virus and herpes simplex virus is much more likely to be severe in pregnant women. Particularly in the third trimester, it can be associated with fulminant disease and high maternal and perinatal mortality. All pregnant patients with new-onset hepatitis should also be screened for cytomegalovirus and Epstein-Barr virus.

TREATMENT Rx

All pregnant women in the United States are screened for hepatitis B virus (HBV) with hepatitis B surface antigen (Chapter 148). All neonates born to positive mothers are treated with hepatitis B immune globulin within 12 hours of birth and given their first dose of HBV vaccine at birth. This regimen is less effective in mothers with a high viral load or in the presence of hepatitis B e antigen positivity.

Fetal risks involved in the use of interferon during pregnancy outweigh its benefits. All current oral anti-HBV drugs (including lamivudine, entecavir, and adefovir) are categorized as Food and Drug Administration pregnancy category C, except telbivudine and tenofovir, which are pregnancy category B drugs.[20] Vertical transmission of HCV occurs, but data supporting recommendations for prevention are limited. Both ribavirin and interferon are contraindicated during pregnancy.

CHRONIC LIVER DISEASE

Chronic liver disease may be associated with anovulation, amenorrhea, and infertility. Thus, it is unusual to see pregnant patients with significant liver decompensation and cirrhosis. Patients with portal hypertension, autoimmune hepatitis, Wilson's disease, hepatic masses, and successful liver transplantation will be seen during pregnancy.

TREATMENT Rx

The management of these patients requires an understanding of their course in pregnancy and a recognition of the importance of continuing prepregnant treatment.

Portal hypertension (Chapter 153) of any cause will be affected by the increase in blood volume during pregnancy, requiring careful follow-up and management of esophageal varices and splenic artery aneurysm. β-Blockers should be continued, and baseline endoscopy should be done early to consider banding of larger varices.

Autoimmune hepatitis (Chapter 149) improves dramatically with immunosuppression, so many women regain fertility with treatment.[21] Immunosuppression with steroids and azathioprine should be continued to avoid relapse and progression of disease. The cholestasis associated with primary biliary cirrhosis can be treated with ursodeoxycholic acid as outlined for ICP.

Similar to women with autoimmune hepatitis, patients with treated Wilson's disease (Chapter 211) regain fertility. Chelation therapy should be continued as discontinuation is associated with marked rises in copper levels and can lead to fulminant liver failure.

Hepatic masses are most commonly benign adenomas, focal nodular hyperplasia, or hemangiomas in women of childbearing age. Careful follow-up of these estrogen-sensitive masses is important because enlargement and hemorrhage may be complications of pregnancy.

SUMMARY

Women of childbearing age with chronic medical conditions benefit greatly from preconception counseling and interventions that address control of their disease and safety of their medications. Pregnant patients with acute or chronic medical illnesses require a multidisciplinary team that understands the maternal and fetal risks related to both the underlying illness and untreated disease. Pregnancy is a window of opportunity to address maternal health, and the maternal response to pregnancy may be predictive of future risk. The 6-week postpartum visit, rather than being the end of pregnancy care, should represent the beginning of a woman's long-term health care.

Grade A References

A1. Koopmans CM, Bijlenga D, Groen H, et al. Induction of labour versus expectant monitoring for gestational hypertension or mild pre-eclampsia after 36 weeks' gestation (HYPITAT): a multicentre, open-label randomised controlled trial. *Lancet*. 2009;374:979-988.

A2. Belfort MA, Anthony J, Saade GR, et al. A comparison of magnesium sulfate and nimodipine for the prevention of eclampsia. *N Engl J Med*. 2003;348:304-341.

A3. Roberts JM, Myatt L, Spong CY, et al. Vitamins C and E to prevent complications of pregnancy-associated hypertension. *N Engl J Med*. 2010;362:1282-1291.

A4. Henderson JT, Whitlock EP, O'Connor E, et al. Low-dose aspirin for prevention of morbidity and mortality from preeclampsia: a systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014;160:695-703.

A5. Bain E, Pierides KL, Clifton VL, et al. Interventions for managing asthma in pregnancy. *Cochrane Database Syst Rev*. 2014;10:CD010660.

A6. Moore LE, Clokey D, Rappaport VJ, et al. Metformin compared with glyburide in gestational diabetes: a randomized controlled trial. *Obstet Gynecol*. 2010;115:55-59.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Sanghavi M, Rutherford JD. Cardiovascular physiology of pregnancy. *Circulation*. 2014;130: 1003-1008.
2. Mannisto T, Mendola P, Vaarasmaki M, et al. Elevated blood pressure in pregnancy and subsequent chronic disease risk. *Circulation*. 2013;127:681-690.
3. Zetstra-van der Woude PA, Vroegop JS, Bos HJ, et al. A population analysis of prescriptions for asthma medications during pregnancy. *J Allergy Clin Immunol*. 2013;131:711-717.
4. Hernandez-Diaz S, Smith CR, Shen A, et al. Comparative safety of antiepileptic drugs during pregnancy. *Neurology*. 2012;78:1692-1699.
5. American College of Obstetricians and Gynecologists. ACOG Practice Bulletin No. 125: Chronic hypertension in pregnancy. *Obstet Gynecol*. 2012;119:396-407.
6. Bateman BT, Bansil P, Hernandez-Diaz S, et al. Prevalence, trends, and outcomes of chronic hypertension: a nationwide sample of delivery admissions. *Am J Obstet Gynecol*. 2012;206:134.e1-134.e8.
7. Ahmed R, Dunford J, Mehran R, et al. Pre-eclampsia and future cardiovascular risk among women: a review. *J Am Coll Cardiol*. 2014;63:1815-1822.
8. Myatt L, Clifton RG, Roberts JM, et al. First-trimester prediction of preeclampsia in nulliparous women at low risk. *Obstet Gynecol*. 2012;119:1234-1242.
9. Myatt L, Clifton RG, Roberts JM, et al. The utility of uterine artery Doppler velocimetry in prediction of preeclampsia in a low-risk population. *Obstet Gynecol*. 2012;120:815-822.
10. Smith GN, Pudwell J, Walker M, et al. Ten-year, thirty-year, and lifetime cardiovascular disease risk estimates following a pregnancy complicated by preeclampsia. *J Obstet Gynaecol Can*. 2012;34: 830-835.
11. Bourjeily G, Paidas M, Khalil H, et al. Pulmonary embolism in pregnancy. *Lancet*. 2010;375: 500-512.
12. Practice Bulletin No. 138: Inherited thrombophilias in pregnancy. *Obstet Gynecol*. 2013;122: 706-716.
13. Bates SM, Greer IA, Middeldorp S, et al. VTE, thrombophilia, antithrombotic therapy, and pregnancy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e691S-e736S.
14. Mihălţan FD, Antoniu SA, Ulmeanu R. Asthma and pregnancy: therapeutic challenges. *Arch Gynecol Obstet*. 2014;290:621-627.
15. Murphy VE, Gibson PG. Asthma in pregnancy. *Clin Chest Med*. 2011;32:93-110.
16. Mathiesen ER, Ringholm L, Damm P. Therapeutic management of type 1 diabetes before and during pregnancy. *Expert Opin Pharmacother*. 2011;12:779-786.
17. Beauharnais CC, Roberts DJ, Wexler DJ. High rate of placental infarcts in type 2 compared with type 1 diabetes. *J Clin Endocrinol Metab*. 2012;97:E1160-E1164.
18. Dixon PH, Wadsworth CA, Chambers J, et al. A comprehensive analysis of common genetic variation around six candidate loci for intrahepatic cholestasis of pregnancy. *Am J Gastroenterol*. 2014;109:76-84.
19. van der Woude CJ, Metselaar HJ, Danese S. Management of gastrointestinal and liver diseases during pregnancy. *Gut*. 2014;63:1014-1023.
20. Piratvisuth T. Optimal management of HBV infection during pregnancy. *Liver Int*. 2013;33: 188-194.
21. Mufti AR, Reau N. Liver disease in pregnancy. *Clin Liver Dis*. 2012;16:247-269.

REVIEW QUESTIONS

1. A 30-year-old G1P0 at 36 weeks' gestation presents with a complaint of right upper quadrant pain. She denies nausea, vomiting, change in color of urine or stool, or change in the pain with eating. She does state that she has had a dull headache and "floaters" in her vision. Laboratory testing reveals an aspartate aminotransferase level of 82 (upper reference limit = 40), hemoglobin level of 14.3, creatinine concentration of 1.0, and platelet count of 100,000. What is her most likely diagnosis?

A. Cholelithiasis
B. Acute hepatitis
C. Cholestasis of pregnancy
D. Preeclampsia
E. Migraine syndrome

Answer: D This patient is a primigravida who presents in her third trimester with symptoms of preeclampsia, hemoconcentration, elevated creatinine for pregnancy, and thrombocytopenia, all most suggestive of preeclampsia.

2. You are caring for a 36-year-old patient with a 6-year history of type 2 diabetes. She has been erratically controlled with metformin, 500 mg daily, and basal insulin. She has never been pregnant but is considering starting a family. What is the single most important advice you can give her?

A. She is really too old to begin a family.
B. She needs to achieve normoglycemia and normalize her hemoglobin A_{1c} level before conception.
C. She will need to be tightly controlled during pregnancy, so she needs to call you as soon as she has a positive pregnancy test result.
D. She needs an eye examination before pregnancy.
E. She should see a nephrologist to screen her kidney function before pregnancy.

Answer: B This patient is in an ideal circumstance for preconception counseling. Normalization of her glucose concentration and hemoglobin A_{1c} level before conception will lower her risk of fetal congenital anomalies and spontaneous abortion to background risk levels.

3. A 24-year-old G2P1 woman whom you observe for asthma comes in at 14 weeks' gestation. She has had increasing problems with her asthma, needing a rescue inhaler two or three times a day, and has had nocturnal awakening. She denies chest pain, recent upper respiratory infection, gastroesophageal reflux, or new exposures. What is the most likely cause of her asthma exacerbation?

A. The normal effects of pregnancy
B. The increased work load and oxygen consumption associated with pregnancy
C. She has discontinued her steroid inhaler and controller medication because of concern about the drug's effect on the fetus.
D. She has decreased her exercise.
E. The stress of work, pregnancy, and sleep deprivation

Answer: C The most frequent cause of asthma exacerbation in pregnancy is the patient's or physician's discontinuation of controller medication because of the concern about fetal effects. Both maternal and fetal outcome are optimized by good control of her asthma, and there are no data of adverse effects on the fetus associated with her controller medications.

4. A 28-year-old woman currently at 8 weeks' gestation in a first pregnancy presents with left leg pain. She was seen in a local emergency department, having recently returned from a car trip from Boston to Florida. She was advised to use nonsteroidal anti-inflammatory drugs and to elevate her leg. She has had extensive nausea and vomiting during her pregnancy. The findings of your examination are benign, with minimal swelling in the left calf. What should your recommendation be?

A. Radiograph of left knee to ankle
B. Doppler ultrasound of left leg to rule out deep venous thrombosis (DVT)
C. Decrease exercise, especially aerobic classes
D. Examine her for probable restless leg syndrome associated with pregnancy
E. Encourage oral rehydration for her dehydration

Answer: B Pregnant patients are at an increased risk of thrombosis, beginning early in gestation. The hormonal effects leading to venodilation and stasis as well as pregnancy-related increase in coagulation activation are already present in this patient. Her dehydration further increases her hypercoagulability. In addition, more than 90% of DVT during pregnancy and the postpartum period occurs on the left side, so her pretest probability of this being DVT is higher than at baseline.

240

MENOPAUSE

DEBORAH GRADY AND ELIZABETH BARRETT-CONNOR

DEFINITION

All healthy women transition from a reproductive or premenopausal period marked by regular ovulation and cyclic menstrual bleeding to a postmenopausal period marked by infertility and amenorrhea (Fig. 240-1). The onset of the menopausal transition is generally marked by subtle shortening in the length of the menstrual cycle and changes in the duration or amount of menstrual flow. As the menopausal transition progresses, menstrual cycles are missed until complete amenorrhea occurs, but the pattern of missed cycles is not predictable. Amenorrhea for a few months is not a good indicator of menopause because one half to three fourths of middle-aged women who are amenorrheic for 6 months resume cycles. Thus, menopause is typically defined retrospectively after 12 months of amenorrhea.

EPIDEMIOLOGY

The menopausal transition usually begins in the middle to late 40s and lasts approximately 4 years, with menopause occurring at a median age of 51 years and ranging from approximately 45 to 57 years. Age at menopause has not changed significantly during the past century. However, a gradual increase in life expectancy to the low 80s now means that the average woman is postmenopausal for more than one third of her life. Age at menopause does not vary significantly by race or ethnicity, but on average, cigarette smokers experience menopause approximately 2 years earlier than nonsmokers do.

BIOLOGY

During the early menopausal transition, estrogen levels are generally normal (50 to 200 pg/mL, depending on the stage of the menstrual cycle) or even slightly elevated, whereas the levels of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) begin to increase (see Fig. 240-1). As the menopause transition progresses, estrogen levels fall markedly, and FSH continues to increase. After menopause, women do not ovulate, and their ovaries do not produce estradiol or progesterone. However, a small amount of estrogen may be produced by metabolism of adrenal steroids to estradiol in peripheral fat tissue. In the early postmenopausal period, mean estradiol

Final menstrual period

	Reproductive			Menopause Transition		Postmenopause	
Reproductive stage	Early	Peak	Late	Early	Late	Early	Late
				Perimenopause			
Menstrual cycle	Variable or regular	Regular		Cycle length variable, 1 or 2 missed cycles per year	3 or more missed cycles	None	
Age (duration)	Puberty to mid-40s			Mid-40s to mid-50s (4 yr)		Mean of 51 years to death	
Steroid hormones	Estradiol 50 to 200 pg/mL Testosterone 400 pg/mL			Same or slightly higher Same		40 pg/mL same	0-15 pg/mL same
Pituitary hormones	FSH 10 mIU/mL day 2-4 LH 10 mIU/mL day 2-4			Same or higher Same or higher		>100 mIU/mL >100 mIU/mL	

FIGURE 240-1. Stages of the menopause transition. FSH = follicle-stimulating hormone; LH = luteinizing hormone.

levels average approximately 40 pg/mL, and they fall to less than 15 pg/mL in the late postmenopausal period. Depending on the measurement method, estradiol is unmeasurable in approximately 15 to 30% of older postmenopausal women. After menopause, testosterone levels may fall slightly but are generally similar to premenopausal levels.

It is not clear what causes menopause, but two leading theories have been proposed.[1] Age-related depletion of ovarian follicles may lead to decreased production of estrogen and inhibin and may thus cause altered hypothalamic-pituitary feedback that results in menopause. Alternatively, age-related changes in hypothalamic production of gonadotropin-releasing hormone and subsequent effects on FSH and LH may be responsible for the increased rate of loss of ovarian follicles, declining ovarian function, and menopause.

MENOPAUSAL SYMPTOMS

Menopause is a positive occurrence in the life of many women. It marks the end of cyclic bleeding and the need for birth control. It occurs at an age when children generally have become independent adults, thereby reducing family and child care responsibilities. Conversely, menopause is a notable sign of aging in cultures that value youth. In addition, it often occurs with other stresses, such as caring for elderly or ill parents. Women in the menopausal transition commonly report a wide variety of symptoms, including hot flushes, night sweats, vaginal dryness, trouble sleeping, sexual dysfunction, depression, anxiety, labile mood, memory loss, fatigue, headache, joint pains, weight gain, and urinary incontinence.

Only vasomotor symptoms, vaginal dryness, and sleep disturbance are consistently associated with the menopausal transition. Other reported symptoms may result from aging or stress associated with menopause. Some symptoms, such as depression, anxiety, memory loss, and fatigue, may be the consequence of frequent hot flushes or poor sleep.

Vasomotor Symptoms

DEFINITION

Vasomotor symptoms include hot flushes, chills, and sweats.[2] A hot flush is a sudden feeling of warmth, generally most intense over the face, neck, and chest. The duration is variable, but it averages approximately 4 minutes. It is often accompanied by sweating that can be profuse and followed by a chill.

EPIDEMIOLOGY

The prevalence of hot flushes is maximal in the late menopausal transition, occurring in approximately 50% of women (Fig. 240-2).[3] However, prevalence varies markedly, depending on the definition of flushing (any flushing, daily flushing, troublesome flushing) and the population studied. Lower prevalence is reported among women in China, Japan, and other Asian countries. The reason for this variation is not clear, but investigators have suggested that it may result from differences in biology, cultural influences on

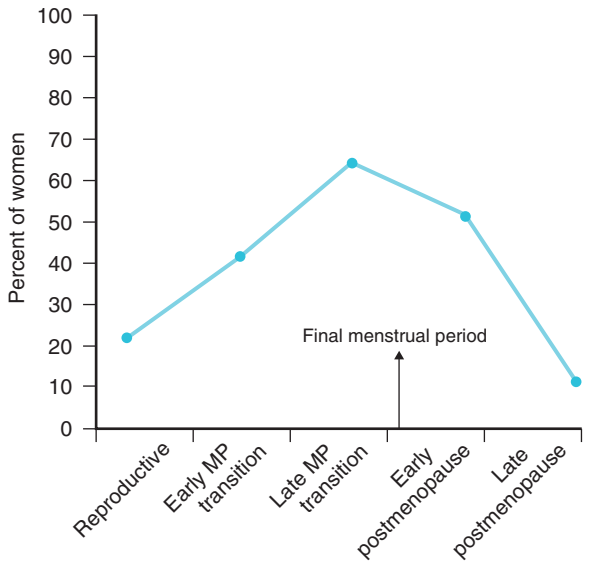

FIGURE 240-2. Prevalence of hot flushes during the menopause (MP) transition.

experiencing or reporting flushes, or diet and lifestyle. In the United States, flushes are more common in African American and Latina women and less common in Chinese and Japanese women compared with white women. Approximately 15% of women with menopausal symptoms consult a physician.

Cigarette smoking increases the likelihood of flushing, but other potential risk factors, including surgical menopause, physical activity, body mass index, alcohol consumption, and socioeconomic status, have been inconsistently associated with hot flushes. Currently, there is no way to predict whether an individual woman will suffer from hot flushes.

In most women, hot flushes are transient. Approximately 50% of women report resolution of symptoms within a few years, and symptoms resolve in about 90% within 8 years. However, some women continue to have frequent and severe flushes many years after menopause. Approximately 10% of women in their middle to late 60s report significant flushing. It is not clear why flushes persist for many years in some women and resolve in others.

PATHOBIOLOGY

Thermoregulation is abnormal in menopausal women with hot flushes. Body temperature is regulated by inducing vasodilation and sweating to release heat and vasoconstriction and shivering to conserve heat. Thermoregulation

is complex and depends on central stimuli from the anterior hypothalamus and local changes in cutaneous vasoconstriction or dilation (Chapters 223 and 418). A hot flush is similar to a heat dissipation response because both result in vasodilation, sweating, and reduction in core body temperature. The mechanism of altered thermoregulation is not clear. The acute vasodilation associated with hot flushes is preceded by a marked increase in skin sympathetic activity, which mediates vasodilation by a number of substances including nitric oxide, vasoactive intestinal peptide, prostaglandins, and substance P. Local blockade of skin sympathetic activity prevents vasodilation, as does local blockade of nitric oxide.

One theory suggests that abnormalities in central nervous system adrenergic neurotransmission cause hot flushes. This theory is supported by studies showing that systemic administration of yohimbine, an α_2-adrenergic antagonist that increases norepinephrine release, provokes hot flushes, whereas administration of clonidine, an α_2-adrenergic agonist that decreases norepinephrine release, reduces the frequency of hot flushes. Alternatively, some evidence indicates that changes in serotonergic neurotransmission could cause hot flushes. Lower estrogen levels are associated with lower levels of serotonin (5-hydroxytryptamine) in blood, resulting in increased sensitivity of 5-hydroxytryptamine type 2A receptors in the hypothalamus. Stimulation of these receptors can alter the thermoregulatory set point in animals. Mild stressors, such as heat and anxiety, cause a brief release of 5-hydroxytryptamine that may stimulate central 5-hydroxytryptamine type 2A receptors, lower the thermoregulatory set point, and cause flushing. This hypothesis is supported by the finding that drugs that increase central serotonin levels are modestly effective in the treatment of hot flushes.

Estrogen treatment effectively relieves hot flushes, but the exact role of estrogen in flushing is not clear. Fluctuations in estrogen levels in an individual woman do not correlate with the onset of flushes. Prepubertal girls with very low levels of endogenous estrogen, premenopausal women with marked fluctuations in estrogen during the menstrual cycle, and most postmenopausal women with low, constant levels of estradiol do not experience flushing. However, women with gonadal dysgenesis (Chapter 233) who are treated with estrogen for several months experience flushing when treatment is discontinued. Thus, withdrawal of estrogen, rather than the absolute estrogen level, appears to play a key role in the etiology of hot flushes.

In addition to changes in estradiol, menopause is associated with multiple other hormonal changes. In the Study of Women's Health Across the Nation, a large cohort study in the United States, lower estradiol was associated with flushing in middle-aged women in univariate models. However, higher FSH was the only measure independently associated with flushing after multivariate adjustment for other hormone levels. Hot flushes correlate with pulsatile increases in LH, but suppression of LH with gonadotropin-releasing hormone agonists does not eliminate flushing. Androgens may also play a role because men who are treated with androgen deprivation therapy for prostate cancer frequently report flushing.

DIAGNOSIS

Vasomotor symptoms are classic manifestations of the menopause transition, and the diagnosis is generally obvious from a woman's age and description of the symptoms. No abnormal physical findings are associated with hot flushes. Estradiol, FSH, and LH levels may be in the normal premenopausal range during the menopausal transition (see Fig. 240-1). A woman in her middle 40s to middle 50s who complains of classic hot flushes does not require any specific physical or laboratory evaluation unless there is good reason to suspect another cause of flushing (Table 240-1). However, an FSH determination may be helpful in assessing the risk for pregnancy. A woman in her middle 40s to middle 50s with an FSH level obtained on the third day after menses that is higher than 20 IU/L is at very low risk of becoming pregnant.

TABLE 240-1 DIFFERENTIAL DIAGNOSIS OF HOT FLUSHES

Alcohol consumption
Carcinoid syndrome
Dumping syndrome
Hyperthyroidism
Narcotic withdrawal
Pheochromocytoma
Medications
Aromatase inhibitors
Gonadotropin-releasing hormone agonists or antagonists
Nicotinic acid
Nitrates
Selective estrogen receptor modulators (tamoxifen and raloxifene)

TREATMENT

Rx

Because self-reported frequency and severity of hot flushes improve markedly with placebo, conclusive evidence of efficacy of treatments requires randomized, blinded trials.

General Measures

Behavioral and Alternative Therapies

Many women have mild flushes and obtain adequate relief with simple measures such as lowering ambient temperature and wearing lighter clothing. Weight loss has been shown to improve hot flashes, but moderate exercise does not alleviate flushing.

There is no convincing evidence that acupuncture, yoga, Chinese herbs, dong quai, evening primrose oil, ginseng, kava, omega-3 fatty acids, or red clover extract improves hot flushes, although a more recent trial found that, among health sedentary menopausal women, yoga appears to improve menopausal quality of lie. Evidence regarding black cohosh is mixed but primarily negative. Multiple trials have been performed with different phytoestrogen preparations. Although some of these studies have reported benefit, the weight of evidence, especially from good-quality trials with blinded comparisons, suggests little benefit. One trial of vitamin E supplementation found an improvement in flushes, but the decrease was only one hot flush per day. Many women prefer alternative medications because they believe that these treatments are harmless, but phytoestrogens and possibly black cohosh bind estrogen receptors and theoretically could cause adverse outcomes similar to those observed with estrogen. No studies of these preparations have been of adequate size or duration to document safety.

Medical Therapy

Estrogens

Multiple randomized trials have demonstrated that estrogen markedly improves the frequency and severity of hot flushes.[4,5] All types, preparations, and routes of administration of estrogen are effective, reducing the frequency of hot flushes 60 to 95%, depending on the dose. Higher doses of estrogen may control symptoms more rapidly but are also associated with a higher rate of side effects, including uterine bleeding, breast tenderness, and headache.

At similar biologically active doses, oral and transdermal estrogens are approximately equally effective for treatment of vasomotor symptoms. Oral estrogens undergo first-pass metabolism in the liver that results in changes in hepatic proteins and enzymes. Hepatic effects are responsible for the beneficial effects of estrogen on lipoproteins (reduced low-density lipoprotein cholesterol and increased high-density lipoprotein cholesterol) but may also cause adverse effects, such as increases in clotting factors. The transdermal route may be safer because it minimizes these changes.

Many estrogen preparations are approved for treatment of vasomotor symptoms (Table 240-2).[6] To individualize treatment, physicians should become familiar with several of these preparations.

Treatment with estrogen alone markedly increases the risk for uterine hyperplasia and cancer. The risk for endometrial abnormalities appears not to be increased with the use of vaginal estrogens that deliver low systemic doses, especially if they are used only a few times per week, as is generally recommended (Table 240-3).

Adding a progestin to the estrogen regimen prevents the increased risk for uterine cancer. For this reason, a woman with a uterus who takes estrogen should also be given a progestin.[7] There is no reason to add progestins to the hormone regimen in women who have had a hysterectomy. Several progestins are approved by the U.S. Food and Drug Administration (FDA) for this purpose and are available either to add to estrogen or in preparations combined with estrogen (see Table 240-2). Two general approaches are used in prescribing progestins to protect the endometrium. Sequential therapy (estrogen given daily with a progestin added on the last 10 to 14 days of a 28-day cycle) results in endometrial shedding and cyclic bleeding resembling a menstrual period in approximately 80% of women. To avoid monthly menstruation-like bleeding, the progestin may be added to estrogen every 3 to 6 months. However, data regarding the safety of this "long cycle" approach to prevent endometrial hyperplasia are mixed, and women using this regimen should undergo vaginal ultrasound or endometrial biopsy if they experience abnormal vaginal bleeding.

Alternatively, the progestin can be added to the estrogen every day. This continuous regimen results in endometrial atrophy and unpredictable uterine spotting or bleeding that can be difficult for the woman to anticipate and manage. Bleeding occurs in approximately 80% of women in the first 6 months of continuous treatment. Amenorrhea becomes common with prolonged use, but some women continue to bleed or spot for many years.

The most commonly used progestins in sequential regimens in women using standard doses of estrogens (0.625 mg oral conjugated estrogens, 1 mg

TABLE 240-2 ESTROGEN AND PROGESTIN PREPARATIONS FOR TREATMENT OF MENOPAUSAL VASOMOTOR SYMPTOMS

HORMONES	GENERIC NAME	BRAND NAME	DOSE (mg/day)
ESTROGENS*			
Oral	Conjugated estrogens	Premarin	0.3, 0.45, **0.625**, 0.9, 1.25
	17β-Estradiol	Estrace/ generics	0.5, **1.0,** 2.0
Transdermal	17β-Estradiol	Alora patch†	0.025, **0.05,** 0.075, 0.1
		Climara patch‡	0.025, 0.0375, **0.05,** 0.075, 0.1
		Menostar patch†	0.014
		Evamist spray	1.53/spray
		Elestrin gel	0.025/pump
Vaginal	Estradiol acetate	Femring vaginal ring§	0.05, 0.10
	0.05, 0.1		
PROGESTINS			
Oral	Medroxyprogesterone acetate	Provera/ generics	2.5, 5.0, 10.0
	Micronized progesterone	Prometrium	100, 200 (in peanut oil)
Vaginal	Progesterone	Prochieve 4%	45 every other day
COMBINATION PREPARATIONS			
Oral sequential‖	Conjugated estrogens and medroxyprogesterone acetate	Premphase	0.625 + 5.0
Oral continuous¶	Conjugated estrogens and medroxyprogesterone acetate	Prempro	0.625 + 2.5 or 5.0; 0.45 + 2.5 or 1.5; 0.3 + 1.5
	17β-Estradiol and norethindrone acetate	Activella	1.0 + 0.5; 0.5 + 0.1
	Conjugated estrogens and bazedoxifene	Duavee	0.45 + 20
Transdermal continuous¶	17β-Estradiol and levonorgestrel	Climara Pro‡	0.045 + 0.015
	17β-Estradiol and norethindrone acetate	CombiPatch‡	0.05 + 0.14 or 0.25

*Approximately equivalent doses of estrogens are shown in **bold**.
†Patch applied twice per week.
‡Patch applied once per week.
§Vaginal ring inserted every 90 days; note that Femring, as opposed to the vaginal preparations listed in Table 240-3, delivers systemic levels of estrogen and should be opposed by a progestin in women with a uterus.
‖Each pill contains estrogen days 1 to 14 and estrogen with progestin days 15 to 28.
¶Each pill or patch contains estrogen and progestin.

TABLE 240-3 ESTROGEN VAGINAL PREPARATIONS FOR TREATMENT OF VAGINAL DRYNESS*

PREPARATION	GENERIC NAME	BRAND NAME	DOSE
Vaginal cream	Conjugated estrogens	Premarin	0.625 mg/2 g cream: 2 g/day for 2 wk, then 1-2 g 2 to 3 times/wk
	17β-Estradiol	Estrace	0.1 mg/2 g cream: 2 g/day for 2 wk, then 1-2 g 2 to 3 times/wk
Vaginal tablet	Estradiol hemihydrate	Vagifem	0.025 mg tablet: 1 tablet/day for 2 wk, then 1 tablet twice/wk
Vaginal ring	17β-Estradiol	Estring†	0.0075 mg/day

*Most oral, transdermal, and vaginal products listed in Table 240-2 for treatment of vasomotor symptoms are also approved for treatment of vaginal dryness.
†Vaginal ring is inserted every 90 days.

TABLE 240-4 RESULTS OF THE WOMEN'S HEALTH INITIATIVE RANDOMIZED TRIALS OF THE EFFECTS OF POSTMENOPAUSAL HORMONE THERAPY ON DISEASE OUTCOMES

	RELATIVE RISK AND 95% CONFIDENCE INTERVAL	
OUTCOMES	*Estrogen and Progestin**	*Estrogen†*
Coronary heart disease events	1.29 (1.02-1.63)	0.91 (0.75-1.12)
Stroke	1.41 (1.07-1.85)	1.39 (1.10-1.77)
Pulmonary embolism	2.13 (1.39-3.25)	1.34 (0.87-2.06)
Breast cancer	1.26 (1.00-1.59)	0.77 (0.59-1.01)
Colon cancer	0.63 (0.43-0.92)	1.08 (0.75-1.55)
Hip fracture	0.66 (0.45-0.98)	0.61 (0.41-0.91)
Dementia	2.05 (1.21-3.48)	1.49 (0.83-2.66)
Death	0.98 (0.82-1.18)	1.04 (0.88-1.22)

*Results of the Women's Health Initiative estrogen plus progestin randomized trial: 16,608 postmenopausal women without hysterectomy randomized to 0.625 mg conjugated estrogen plus 2.5 mg medroxyprogesterone acetate per day or identical placebo and observed for 5.2 years.
†Results of the Women's Health Initiative estrogen-only randomized trial: 10,739 postmenopausal women with hysterectomy randomized to 0.625 mg conjugated estrogen per day or identical placebo and observed for 6.8 years.

oral estradiol, or 0.05 mg transdermal estradiol) are medroxyprogesterone acetate 5 mg and micronized progesterone 200 mg for 10 to 14 days per month. Continuous regimens generally include about half these progestin doses given daily. In some situations, it may be slightly less costly to prescribe estrogen and progestins separately, but the convenience of taking a single pill and the assurance that the estrogen is adequately opposed by progestin make combination preparations preferable.

A combination of conjugated equine estrogens (0.45 mg) and bazedoxifene (20 mg) has been approved by the FDA for treatment of hot flashes in women with a uterus. Bazedoxifene is a selective estrogen receptor modulator that, like progestin, appears to block the carcinogenic effects of estrogen on the uterus. In clinical trials, conjugated estrogens plus bazedoxifene reduced the frequency of hot flashes about 80%, improved dyspareunia, and increased bone density with few side effects. Studies of up to 2 years have not identified other important adverse effects, such as endometrial cancer, breast cancer, venous thromboembolic events, or stroke, but experience with this preparation is limited.

The use of "bioidentical hormones" is based on the concept that estrogens (estradiol, estrone, and estriol) and progestins (progesterone) made from plant products are identical to women's endogenous hormones and therefore more natural, safe, and effective than FDA-approved hormone preparations. Bioidentical hormone therapy often uses doses and combinations of steroid hormones guided by the patient's symptoms or serum hormone levels, and prescriptions are generally filled by Internet-based compounding pharmacies. There is little scientific rationale for the mixtures and ratios of hormones employed, and there are no adequate clinical trial data to support the safety or efficacy of these regimens, some of which include very large doses of estradiol. Bioidentical hormones are not approved by the FDA for treatment of menopausal symptoms.

Side Effects and Risks of Postmenopausal Hormone Therapy

Estrogen is generally well tolerated, but it may cause headache (especially in women with a history of migraine) and breast tenderness. Added progestins tend to make these side effects more severe and also cause uterine bleeding.

The effects of hormone therapy on disease outcomes have been evaluated among postmenopausal women in the Women's Health Initiative (WHI) randomized trials (Table 240-4). Both estrogen and estrogen in combination with progestin reduced the risk for hip fracture by 35 to 40%. Neither estrogen alone nor estrogen with a progestin reduced the risk for coronary events, and both increased the risk for stroke by approximately 40%.

Compared with estrogen alone, added progestin appears to increase the risk for coronary events, pulmonary embolism, breast cancer, and dementia (Table 240-4). This finding suggests that adding a progestin should be avoided, but treatment with unopposed estrogen in women with a uterus markedly increases the risk for uterine hyperplasia and cancer as well as the rate of gynecologic procedures and hysterectomy. Replacing progestins with the selective estrogen receptor modulator bazedoxifene is a novel approach to blocking the carcinogenic effects of estrogen on the uterus, but at this time bazedoxifene has not been proved to be safer than progestins.

The excess risk of any one of the adverse events listed in Table 240-4 in the WHI trials was about 2 per 1000 women treated for 1 year with estrogen in combination with progestin (about 4 per 1000 per year in women older than

65 years if dementia is included) and 1 stroke per 1000 women treated with unopposed estrogen. These risks are relatively small, but they cumulate such that treatment for 5 years is associated with an excess risk of 1 event per 100 women treated with estrogen in combination with progestin and 3 strokes per 1000 women treated with unopposed estrogen. Given these potential harms and the availability of other effective and safe drugs for prevention of osteoporotic fractures, postmenopausal hormone therapy currently has no role for prevention of disease, especially in older women.

The average age of women enrolled in the WHI trials was 63 years. In contrast, most women who take hormone therapy for treatment of hot flashes are generally in their 50s. Subgroup analyses by age from the WHI trials that include women who took hormone therapy for 5 to 7 years show that the *relative* risk for any adverse event among women who took estrogen plus progestin was similar for younger (50 to 59 years) and older (70 to 79 years) women; but given the lower rate of events in younger women, the *absolute* increase in the rate of any major adverse event in the younger women was about 1 per 1000 per year, compared with 4 per 1000 per year for older women. Among women who took estrogen alone, the *relative* risk was lower in younger women than in older women, and the *absolute* rate of adverse events was decreased by 2 per 1000 per year in the younger women compared with an increase of 5 per 1000 per year in the older women. These data suggest that the risk for women in their 50s taking estrogen plus progestin for a few years for the treatment of hot flashes is very low, and there may be no risk associated with taking estrogen alone.

Despite these findings, hormone therapy is relatively contraindicated in women with a history of stroke, breast cancer, or venous thromboembolic events and should be avoided in women at high risk for these conditions.

Stopping Hormone Therapy

Given the possible adverse effects of hormone therapy, current guidelines recommend that women use the lowest effective dose for the shortest time necessary. Vasomotor symptoms improve or resolve spontaneously within a few years of onset in most women, suggesting that most should be able to discontinue hormone therapy within a few years of starting. Women using hormone therapy for treatment of symptoms should stop every 6 to 12 months to determine whether symptoms have improved to the point that treatment is no longer needed. A small percentage of women are unable to stop hormone therapy because vasomotor symptoms persist for many years.

Women experiencing intolerable symptoms after stopping hormone treatment can be told to resume therapy and either to begin a slow taper or to wait 6 months before trying again to stop. Tapering can be accomplished by decreasing the dose of hormone therapy, but it may be easier to decrease the number of days per week that hormone therapy is used. The *dose taper* involves progressively reducing the dose of hormone therapy, for example, by reducing the dose of conjugated estrogens from 0.625 mg/day to 0.45 to 0.3 mg/day and then discontinuing therapy. If changing to a lower dose is associated with tolerable symptoms, the next reduction in dose should not occur until symptoms improve, which may require 3 to 6 months in some women. The *day taper* involves decreasing the number of days per week of hormone therapy use and effectively decreasing the weekly dose. For example, therapy with the same dose may be continued, but only Monday through Friday. If this reduced weekly dose is tolerated, therapy may be discontinued on Friday, and so on. As with the dose taper, if symptoms develop, the weekly dose should be maintained until symptoms improve. Both these approaches to tapering can require many months or even years until therapy is discontinued. The day taper has the advantages of allowing smaller reductions in weekly dose, ensuring that the estrogen dose is appropriately opposed by a progestin in women using continuous therapy, and does not require multiple new prescriptions for different doses of hormone therapy. For women who cannot tolerate even a slow taper, the value of symptom relief likely outweighs the risks of hormone therapy.

Other Prescription Drugs

The progestins megestrol and medroxyprogesterone acetate are effective for the treatment of hot flushes, but they have frequent side effects, and progestin use was associated with increased risk for adverse effects in the WHI trials. Several selective serotonin and serotonin-norepinephrine reuptake inhibitors and gabapentin have been studied (Table 240-5). Paroxetine and escitalopram reduce the frequency of hot flashes by about 0.5 to 1 hot flash per day more than placebo; these drugs have side effects typical of selective serotonin reuptake inhibitors. Low-dose paroxetine mesylate (7.5 mg/day) reduces hot flash frequency by about 1 hot flash per day more than placebo, and side effects were uncommon. This is the only nonhormonal drug currently approved by the FDA for the treatment of hot flashes. In a randomized, double-blinded trial, the efficacy of low-dose oral estradiol (0.5 mg daily) was found to be only slightly superior to that of the serotonin-norepinephrine reuptake inhibitor venlafaxine (extended release, 75 mg daily) in alleviating vasomotor symptoms. Gabapentin also has only modest efficacy (reducing the frequency of hot flashes by 0.5 to 1 hot flash per day more than placebo) but is associated with dizziness, somnolence, headache, and nausea. Extended-release gabapentin has been studied in three phase III randomized trials, but results have not been published and the drug was not approved by the FDA for treatment of hot flashes. Clinical trials of these prescription drugs have generally lasted only 3 months and have not been adequately large or prolonged to detect uncommon adverse effects.

Summary of Approach to Treatment of Vasomotor Symptoms

Women with mild vasomotor symptoms may find adequate relief by wearing layered clothing and keeping the home and bedroom cool. Women with moderate symptoms may choose a low dose of estrogen or a nonestrogen therapy. Low-dose paroxetine is modestly effective, has the best side effect profile of the nonhormonal drugs, and is approved by the FDA for this indication. For women with severe symptoms, hormone therapy is the most effective treatment.

Vaginal Symptoms

EPIDEMIOLOGY

The prevalence of vaginal dryness, discomfort, itching, and dyspareunia increases as women transition through the menopause. Up to 30% of perimenopausal and early postmenopausal women and a higher proportion of older menopausal women express these complaints. Urologic symptoms, including urgency, frequency, dysuria, and incontinence, are not clearly correlated with the menopause transition.

PATHOBIOLOGY

Vaginal symptoms generally correlate with findings (often called vaginal atrophy) including pallor, dryness, friability, and decreased rugosity of the vaginal mucosa. Vaginal fluid in premenopausal women is acidic, ranging from a pH of approximately 4.5 to 5.5 with mild alkalinization to approximately 6.0 before ovulation. Acidity is produced by proton excretion from the vaginal epithelial cells and by metabolism of glycogen stored in vaginal epithelial cells by *Lactobacillus* species, the normal vaginal flora. The acid environment of the vagina inhibits growth of *Escherichia coli* and other enteric gram-negative bacteria that are a potential cause of urinary tract infections. Vaginal pH can easily be measured from lateral vaginal wall fluid.

In postmenopausal women, vaginal pH is generally neutral, and the predominant flora are often *E. coli* and other gram-negative bacteria. This appears to occur because estrogen deficiency associated with menopause causes vaginal epithelial cell dysfunction, including decreased storage of glycogen, less ability to acidify the vaginal fluid, and lowered production of vaginal lubrication. Vaginal epithelial cells, which are primarily superficial and intermediate cells in premenopausal women, shift to predominantly immature parabasal cells in postmenopausal women. Treatment with estrogen improves or relieves vaginal dryness, lowers vaginal pH, and increases the proportion of superficial cells in the vaginal epithelium.

DIAGNOSIS

Diagnosis is primarily based on typical complaints of vaginal dryness, discomfort, itching, or dyspareunia in women undergoing the menopause transition or older postmenopausal women. Pelvic examination should be performed to exclude other causes of symptoms, including infections, lesions, and trauma. Physical findings of vaginal dryness, pallor, friability, and vaginal pH above 5.5 support the diagnosis. Cytologic examination of the proportion of superficial, intermediate, and parabasal cells from a scraping of the lateral vaginal wall (vaginal maturation index) showing primarily parabasal cells also supports the diagnosis. In clinical practice, measurement of pH and vaginal maturation index are not necessary to make the diagnosis.

TREATMENT

Rx

Some women with mild dyspareunia may obtain adequate relief with over-the-counter vaginal lubricants used as needed for sexual intercourse.[8] Over-the-counter vaginal moisturizers, such as Replens (a bioadhesive polycarbophil vaginal gel used daily or three times per week), have been shown to improve vaginal symptoms and findings. Estrogen therapy is highly effective.[9] Topical therapy is efficacious[10] and is preferred because it generally results in smaller increases in systemic estrogen levels than with oral or transdermal therapy. Estrogen vaginal creams, tablets, and rings approved for treatment of vaginal dryness are listed in Table 240-3.

Most clinicians do not add a progestin to protect the uterus in women treated with vaginal estrogen, but evidence to support the uterine safety of vaginal estrogen is limited to short-term studies. Low-dose, intermittent

TABLE 240-5 EVIDENCE FROM RANDOMIZED, CONTROLLED CLINICAL TRIALS OF THE EFFICACY OF NONESTROGEN DRUGS FOR TREATMENT OF MENOPAUSAL HOT FLUSHES

TREATMENT	EVIDENCE OF BENEFIT	COMMENTS	REFERENCES	SIDE EFFECTS*
ANTIDEPRESSANTS				
Citalopram	No	No benefit of 30 mg citalopram compared with placebo	Suvanto-Luukkonen E, Koivunen R, Sundstrom H, et al. *Menopause.* 2005;12:18-26.	Selective serotonin reuptake inhibitors citalopram, escitalopram, fluoxetine, paroxetine, and sertraline: nausea, vomiting, diarrhea, insomnia or somnolence, anxiety, decreased libido, dry mouth, worsening depression, mania, suicidality, serotonin syndrome, withdrawal syndrome, and possible decreased tamoxifen effectiveness
Escitalopram	Yes	Among generally healthy women, frequency of hot flushes reduced 47% with escitalopram 10 to 20 mg/day compared with 33% with placebo	Freeman EW, Guthrie KA, Caan B, et al. *JAMA.* 2011;305:267-274.	
Fluoxetine	Mixed	Among breast cancer survivors, frequency of hot flushes reduced 50% with fluoxetine 20 mg/day compared with 36% with placebo No benefit among women with breast cancer treated with 30 mg fluoxetine compared with placebo	Loprinzi CL, Sloan JA, Perez EA, et al. *J Clin Oncol.* 2002:20:1578-1583. Suvanto-Luukkonen E, Koivunen R, Sundstrom H, et al. *Menopause.* 2005;12:18-26.	
Paroxetine	Yes	Among generally healthy women, frequency of hot flushes reduced 62% with 12.5 mg and 65% with 25 mg paroxetine CR compared with 38% in placebo In a crossover trial in which 81% of participants had a history of breast cancer, paroxetine 10 mg reduced hot flush frequency by 41% compared with 14% with placebo, and paroxetine 20 mg reduced hot flush frequency by 52% compared with 27% placebo In 2 similarly designed trials of 7.5 mg of paroxetine mesylate daily compared with placebo, paroxetine reduced the frequency of hot flashes by about 1 hot flash per day more than placebo	Steans V, Beebe KL, Lyengar M, et al. *JAMA.* 2003;289:2827-2834. Stearns V, Slack R, Greep N, et al. *J Clin Oncol.* 2005;23:6919-6930. Simon JA, Portman DJ, Kaunitz AM, et al. *Menopause.* 2013;20:1027-1035.	
Sertraline	No	Among women with a history of breast cancer, no benefit of treatment with 50 mg of sertraline compared with placebo Among generally healthy women, no benefit of 100 mg of sertraline compared with placebo	Kimmick GG, Lovato J, McQuellon R, et al. *Breast J.* 2006;12:114-122. Grady D, Cohen B, Tice J, et al. *Obstet Gynecol.* 2007;109:823-830.	
Venlafaxine	Mixed	Among breast cancer survivors, frequency of hot flushes reduced 61% with 75 or 150 mg venlafaxine compared with 27% placebo Among women without breast cancer, no effect on frequency of flushes with 75 mg venlafaxine, but women treated with venlafaxine were more likely to report that flushes improved compared with placebo	Loprinzi CL, Kugler JW, Sloan JA, et al. *Lancet.* 2003;356:2059-2063. Evans ML, Pritts E, Vittinghoff E, et al. *Obstet Gynecol.* 2005;105:161-166.	In addition to the side effects noted above, the selective serotonin-norepinephrine reuptake inhibitors venlafaxine and desvenlafaxine can also cause hypertension.
Desvenlafaxine	Yes	Meta-analysis of the results of 6 randomized trials showed that treatment with 100 to 150 mg of desvenlafaxine per day reduces the frequency of hot flashes by 0.3 to 0.5 more than with placebo	Umland EM, Falconieri L. *Int J Womens Health.* 2012;4:305-319.	
ANTIHYPERTENSIVES				
Clonidine	Mixed	Small trials suggest little or no benefit	Nelson HD, Haney E, Humphrey L, et al. AHRQ Publication No. 05-E016-2. Rockville, MD: Agency for Healthcare Research and Quality; 2005. Goldberg RM, Loprinzi CL, O'Fallon JR, et al. *J Clin Oncol.* 1994;12:155-158. Pandya KJ, Raubertas RF, Flynn PJ, et al. *Ann Intern Med.* 2000;132:78-93.	α-Adrenergic antagonists clonidine and methyldopa: dry mouth, drowsiness, dizziness, hypotension, rebound hypertension
Methyldopa	No		Nelson HD, Haney E, Humphrey L, et al. AHRQ Publication No. 05-E016-2. Rockville, MD: Agency for Healthcare Research and Quality; 2005.	

TABLE 240-5 EVIDENCE FROM RANDOMIZED, CONTROLLED CLINICAL TRIALS OF THE EFFICACY OF NONESTROGEN DRUGS FOR TREATMENT OF MENOPAUSAL HOT FLUSHES—cont'd

TREATMENT	EVIDENCE OF BENEFIT	COMMENTS	REFERENCES	SIDE EFFECTS*
HORMONES				
Medroxyprogesterone acetate	Yes	Frequency of hot flushes was reduced 74% with 20 mg medroxyprogesterone acetate compared with 26% with placebo	Schiff I, Tulchinsky D, Cramer D, et al. *JAMA*. 1980;244:1443-1445.	Progestins medroxyprogesterone and megestrol: nausea, vomiting, constipation, somnolence, depression, breast tenderness, uterine bleeding; possible increased risk for venous thromboembolism, cardiovascular events, and breast cancer
Megestrol	Yes	Among breast cancer survivors, frequency of hot flushes reduced 74% with 20 mg megestrol twice a day compared with 27% with placebo	Loprinzi CL, Michalak JC, Quella SK, et al. *N Engl J Med*. 1994;331:347-352.	
OTHER DRUGS				
Gabapentin	Yes	Among healthy women, frequency of hot flushes was reduced 45% with gabapentin 300 mg three times/day compared with 29% with placebo Frequency of hot flushes was reduced 31% more than with placebo among breast cancer survivors	Guttuso TJ, Kurlan R, McDermott MP, et al. *Obstet Gynecol*. 2003;101:337-345. Pandya JK, Morrow GR, Rosco JA, et al. *Lancet*. 2005;366:818-824.	Nausea, vomiting, somnolence, dizziness, rash, ataxia, fatigue, leukopenia

*Side effects were reported in clinical trials of the therapy or from Epocrates RX drug reference (available at http://www.epocrates.com).

treatment (e.g., 1 to 2 g conjugated estrogen cream or 0.025 mg estradiol tablet twice a week) results in small increases in systemic estrogen levels that appear not to cause endometrial stimulation. However, full-dose daily treatment has been shown to increase estradiol levels to 50 pg/mL or higher in approximately half of treated women and has been associated with uterine bleeding and hyperplasia.

Ospemifene, an oral selective estrogen receptor modulator, is approved by the FDA for the treatment of menopausal vaginal dryness and dyspareunia. Treatment with 60 mg once daily reduces the bothersomeness of vaginal symptoms about 10 to 15% more than placebo does, but it is associated with hot flashes, urinary tract infection, and vaginal infections. Small studies of up to 1 year have not shown increased risk of endometrial cancer, breast cancer, venous thromboembolic events, or stroke, but because ospemifene is a selective estrogen agonist, these potential adverse effects are of concern.

Sleep Disturbance

The prevalence of self-reported sleep disturbance increases from about 40% of premenopausal women to approximately 60% of postmenopausal women. Sleep disturbances, including trouble falling asleep and early awakening, are reported by menopausal women, but awakening during the night appears to be most bothersome.

The etiology of sleep disturbance associated with menopause is unclear. Postmenopausal women with hot flushes are more likely to report sleep disturbance than are those without flushes, and women commonly report that they are awakened by hot flushes. However, studies using polysomnography find that nocturnal hot flushes do not consistently occur at the same time as sleep disturbance. Thus, disturbed sleep appears to be part of a menopausal syndrome, but it may not be caused by flushing.

Menopause-related sleep disturbance can be treated by standard approaches to sleep hygiene and prescription medications. Both oral and transdermal estrogen preparations improve sleep in perimenopausal and postmenopausal women with hot flushes.

Grade A References

A1. Reed SD, Guthrie KA, Newton KM, et al. Menopausal quality of life: RCT of yoga, exercise, and omega-3 supplements. *Am J Obstet Gynecol*. 2014;210:244.e1-244.e11.

A2. Pinkerton JV, Utian WH, Constantine GD, et al. Relief of vasomotor symptoms with the tissue-selective estrogen complex containing bazedoxifene/conjugated estrogens: a randomized, controlled trial. *Menopause*. 2009;16:1116-1124.

A3. Manson JE, Chlebowski RT, Stefanick ML, et al. Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials. *JAMA*. 2013;310:1353-1368.

A4. Simon JA, Portman DJ, Kaunitz AM, et al. Low-dose paroxetine 7.5 mg for menopausal vasomotor symptoms: two randomized controlled trials. *Menopause*. 2013;20:1027-1035.

A5. Freeman EW, Guthrie KA, Caan B, et al. Efficacy of escitalopram for hot flashes in healthy menopausal women: a randomized controlled trial. *JAMA*. 2011;305:267-274.

A6. Hayes LP, Carroll DG, Kelley KW. Use of gabapentin for the management of natural or surgical menopausal hot flashes. *Ann Pharmacother*. 2011;45:388-394.

A7. Joffe H, Guthrie KA, LaCroix AZ, et al. Low-dose estradiol and the serotonin-norepinephrine reuptake inhibitor venlafaxine for vasomotor symptoms: a randomized clinical trial. *JAMA Intern Med*. 2014;174:1058-1066.

A8. Suckling J, Lethaby A, Kennedy R. Local oestrogen for vaginal atrophy in postmenopausal women. *Cochrane Database Syst Rev*. 2006;4:CD001500.

A9. Portman DJ, Bachmann GA, Simon JA. Ospemifene, a novel selective estrogen receptor modulator for treating dyspareunia associated with postmenopausal vulvar and vaginal atrophy. *Menopause*. 2013;20:623-630.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Pollycove R, Naftolin F, Simin JA. The evolutionary origin and significance of menopause. *Menopause*. 2011;18:336-342.
2. Goodman NF, Cobin RH, Ginzburg SB, et al. American Association of Clinical Endocrinologists Medical Guidelines for Clinical Practice for the diagnosis and treatment of menopause: executive summary of recommendations. *Endocr Pract*. 2011;17:949-954.
3. Freeman EW, Sammel MD, Sanders RJ. Risk of long-term hot flashes after natural menopause: evidence from the Penn Ovarian Aging Study cohort. *Menopause*. 2014;21:924-932.
4. North American Menopause Society. The 2012 hormone therapy position statement of the North American Menopause Society. *Menopause*. 2012;19:257-271.
5. U.S. Preventive Services Task Force. Menopausal hormone therapy for the primary prevention of chronic conditions. January 2013. http://www.uspreventiveservicestaskforce.org/uspstf/uspspmho.htm; Accessed October 17, 2013.
6. Nelson HD, Walker M, Zakher B, et al. Menopausal hormone therapy for the primary prevention of chronic conditions: a systematic review to update the U.S. Preventive Services Task Force recommendations. *Ann Intern Med*. 2012;157:104-113.
7. Reid R, Abramson BL, Blake J, et al. Managing menopause. *J Obstet Gynaecol Can*. 2014;36:830-838.
8. North American Menopause Society. Position Statement. Management of symptomatic vulvovaginal atrophy: 2013 position statement of the North American Menopause Society. *Menopause*. 2013;20:888-902.
9. Singh S, van Herwijnen I, Phillips C. The management of lower urogenital changes in the menopause. *Menopause Int*. 2013;19:7-81.
10. Management of symptomatic vulvovaginal atrophy: 2013 position statement of The North American Menopause Society. *Menopause*. 2013;20:888-902.

REVIEW QUESTIONS

1. Which of the following case histories documents that the woman is in the early menopause transition?

A. A 50-year-old woman complains of labile mood, forgetfulness, and loss of libido and reports that she missed a menstrual period last month.
B. A 54-year-old woman reports that she has not had a menstrual period for a year and is having hot flashes.
C. A 48-year-old smoker complains that she is having hot flashes and trouble sleeping and reports that she has missed three menstrual periods in the past year.
D. A 48-year-old woman who is taking birth control pills reports gaining 10 pounds in the past year.
E. A 60-year-old woman complains that she has been having hot flashes for 10 years since her menstrual periods stopped, and the hot flashes recur every time she tries to stop her estrogen treatment.

Answer: C A woman in her mid-40s to mid-50s who complains of classic hot flashes is likely to be in the menopause transition. The fact that the woman in C has missed three menstrual periods in the last year suggests that she is in the early menopause transition. The symptoms reported by the women described in A and D are not associated with the menopause transition. The women described in B and E have undergone menopause or are in the late menopause transition because they have not had a menstrual period for a year.

2. What evaluation is required for a 52-year-old woman who has missed six menstrual periods in the previous year and complains of hot flashes?

A. Lateral vaginal wall pH
B. A careful pelvic examination to evaluate for vaginal atrophy
C. Serum follicle-stimulating hormone (FSH) level
D. Review of medications and prior medical history
E. Serum thyroid-stimulating hormone level

Answer: D A clear history of appropriate age (mid-40s to mid-50s), missing menstrual periods, and classic description of hot flashes with no other medical problems that might cause flushing (see Table 240-1) is diagnostic of menopause-related hot flashes, and no further evaluation is needed. There are no associated physical findings, and FSH levels may be normal in the early menopause transition. If the woman is not complaining of vaginal symptoms, a pelvic examination or vaginal wall pH measurement is not helpful.

3. What treatment would you recommend for a 50-year-old woman with mild hot flashes?

A. Reassure her that the symptoms are likely to resolve in a few years.
B. Suggest that she dress in layers and keep her bedroom cool.
C. If there are no contraindications, offer low-dose (7.5 mg/day) paroxetine.
D. If there are no contraindications, offer low-dose hormone therapy.
E. All of the above

Answer: E Many women with mild hot flashes want to be reassured that they are indeed in the menopause transition and that this is a normal stage of reproductive life. These women are often not bothered enough by hot flashes to want treatment. However, some women are bothered and would like to try treatment to alleviate the hot flashes. In this situation, a nonhormonal drug such as paroxetine, 7.5 mg daily, or a low-dose estrogen, such as 0.025 mg estradiol (combined with a progestin if she has a uterus), is likely to provide adequate relief.

4. What treatment would you recommend for a 50-year-old woman with severe hot flashes that are interfering with her ability to work? She has not had a hysterectomy, and her only other medical problem is osteoarthritis, for which she takes a nonsteroidal anti-inflammatory drug.

A. Reassurance that the symptoms are likely to resolve in a few years
B. A nonhormonal drug, such as low-dose paroxetine 7.5 mg daily
C. Bioidentical hormone therapy
D. Standard-dose estrogen given daily, combined with a progestin given on the last 10 to 14 days of the month or half-dose progestin given daily
E. Estradiol 0.5 mg given daily

Answer: D A woman with severe hot flashes is unlikely to obtain adequate relief with nonhormonal treatments. However, if she wishes to try these treatments before taking hormone therapy or there is a contraindication to hormone therapy, it would appropriate to try a nonhormonal treatment such as paroxetine or escitalopram. A woman with a uterus should have a progestin added to estrogen therapy to avoid the increased risk of endometrial hyperplasia and cancer associated with unopposed estrogen therapy. This can be done by adding standard-dose progestin on days 10 to 14 of the month (or by using a sequential combination estrogen plus progestin pill) or by adding half-dose progestin daily (or using a combination estrogen plus progestin pill or transdermal patch). Hormone therapy options are provided in Table 240-2. The various types of estrogen and progestins are equally effective at equivalent doses. Some women prefer a patch, whereas others prefer a pill; both the oral and transdermal routes are effective.

5. What treatment would you recommend for a 65-year-old woman with severe vaginal dryness, itching, and dyspareunia?

A. Reassurance that the symptoms are likely to resolve in a few years
B. Use of a vaginal lubricant as needed for sexual intercourse
C. Use of a vaginal moisturizer daily or three times per week
D. Estrogen vaginal cream
E. Estrogen vaginal ring

Answer: D or E The best answers are D and E. Reassurance is inappropriate, as unlike hot flashes, menopause-related vaginal symptoms generally do not resolve over time. The woman could try vaginal lubricants for sexual intercourse and a vaginal moisturizer such as Replens, but these are unlikely to provide adequate relief of symptoms. Low-dose topical estrogen applied to the vagina is the most effective treatment. Estrogen can be provided as a cream, vaginal tablet, or slow-release ring. Treatment options are provided in Table 240-3. If topical estrogens are used at low doses (two or three times per week as described in Table 240-3), added progestins are not required for women with a uterus.

241

INTIMATE PARTNER VIOLENCE

GENE FEDER AND HARRIET L. MACMILLAN

DEFINITION

Intimate partner violence (IPV) is defined as any behavior within an intimate relationship or ex-relationship that causes physical, psychological, or sexual harm. This includes physical aggression, such as hitting, kicking, and beating; psychological violence, such as intimidation and constant humiliation; various controlling behaviors, such as isolation from family and friends, monitoring of movements, financial control, and restricting access to services; and sexual violence, including forced intercourse and other sexual coercion. Lifetime prevalence of isolated violent acts within relationships is comparable for men and women, but repeated coercive, sexual, or severe physical violence is perpetrated largely against women by men. Although IPV also occurs in same-sex relationships, research evidence on the health consequences of IPV and the care of survivors is largely confined to women in heterosexual relationships.

Historically, there has been the stereotype of a male batterer as one who uses severe, repeated, and unilateral violence against a nonviolent female

victim. It is now recognized that *bilateral violence* is a common form of IPV, even though the overwhelming burden of morbidity and mortality related to IPV is experienced by women. Bilateral violence, sometimes referred to as *common couple violence,* is considered less severe than the pattern of abuse known as *battering* or *intimate terrorism,* a severe and escalating form of IPV characterized by threats, terrorization, multiple forms of abuse, and controlling behavior on the part of the abuser. Current research suggests that women rarely subject men to battering.

IPV is a risk factor for a wide range of medical and psychiatric conditions and therefore can be understood as an epidemiologic exposure. Yet violence perpetrated by an intimate partner or ex-partner is essentially a violation of human rights and a preventable psychosocial issue that needs to be addressed through social and educational policies.

EPIDEMIOLOGY

The prevalence of IPV against women varies internationally but is universally high, comparable to that of chronic conditions like diabetes and asthma. The most robust comparative study (24,097 women from 15 sites in 10 countries), conducted by the World Health Organization, found a lifetime prevalence of physical violence, sexual violence, or both ranging from 15 to 71%.[1] Prevalence of physical or sexual violence during the past year ranged from 15 to 54%. In all sites but one, women were more at risk for violence from a partner or ex-partner than from violence by other people. A systematic review that included data from 66 countries found that one in seven homicides globally are committed by an intimate partner; this figure is six times higher for female homicides compared with male homicides.[2]

Causation

Several theories about the causes of IPV have been proposed over the years. Social learning theory suggests that IPV is a learned behavior. The fact that male perpetrators and female victims are more likely to report histories of exposure to violence in childhood supports this theory. However, most individuals exposed to violence in childhood do not go on to commit violence as adults, and not all abusers have violent upbringings. Furthermore, the link between poor parenting generally, including neglect, and subsequent IPV in adulthood suggests that the effect is not simply one of modeling abusive behavior. Exposure to rejecting or neglectful parenting is associated with adverse effects on intrapersonal (e.g., poor self-worth) and interpersonal development that are associated with IPV.

A feminist perspective views IPV against women as a form of social control that results from society's patriarchal structure leading to inequality in power relationships between men and women. Lending support to this perspective is the finding that IPV appears to be less common in more democratic and less economically polarized societies. Although IPV occurs more often in contexts in which there is support for male authority in the family and women have less access to economic security, it is not clear why some individuals are more likely than others to be violent under such conditions.

With regard to psychological theory, there are conflicting views about the association between IPV and psychopathology. Some researchers argue that abusive males have deficits in one or more coping mechanisms, anger control, and communication skills, whereas others suggest that IPV results from dysfunctional interactional patterns between partners. Because types of IPV are not the same for all couples, there are likely to be multiple causes for its occurrence. Most of the research has focused on factors associated with increased risk of men abusing women (Table 241-1)[3]; however, whether these factors are "causal" is unknown.

TABLE 241-1 FACTORS ASSOCIATED WITH A MAN'S RISK FOR ABUSING HIS PARTNER

INDIVIDUAL	RELATIONSHIP	COMMUNITY	SOCIETAL
Young age	Poor family functioning	Weak community sanctions against intimate partner violence	Traditional gender norms
Heavy drinking	Marital instability	Poverty	Social norms supportive of violence
Depression	Marital conflict	Economic inequality	
Personality disorders	Male dominance	Low social capital	
Low academic achievement	Economic stress		
Low income			
Exposure to violence in childhood			

Modified from World Health Organization: World Report on Violence and Health. Geneva: World Health Organization; 2002.

CLINICAL MANIFESTATIONS

The information in this section pertains to female patients because most of the research examining clinical manifestations associated with IPV exposure has focused on women. However, studies of male victims suggest that they also experience increased risk for poor health as well as chronic physical and emotional health problems and injuries.[4]

Patients seldom present with a chief complaint of IPV. Injuries are the most obvious manifestation; a clinician should have increased suspicion for IPV if there are multiple injuries, the presenting history of injuries is not consistent with the physical examination, and there is a delay in seeking medical care for injuries. Patients exposed to physical violence may present with injuries that vary from minor abrasions to life-threatening trauma. Although there can be overlap between injuries resulting from IPV and injuries from other causes, the former typically involve trauma to the head, face, and neck, whereas the latter are more typically injuries of the extremities.[5] Multiple facial injuries are suggestive of IPV rather than of other causes, and those that are more specific for IPV include zygomatic complex fractures, orbital blowout fractures, and perforated tympanic membrane. Although facial injuries are the most common injuries associated with IPV, they have low specificity. Musculoskeletal injuries are considered the second most common type of injuries, including sprains, fractures, and dislocations.[6] Blunt-force trauma to the forearms should raise suspicion of IPV because these can occur when trying to block being struck.

Victims of IPV often experience multiple mechanisms of injury; being struck by a hand is the most common, followed by use of a household object. Injuries from weapons such as knives and guns are far less common (<1%) but are associated with higher risk for mortality. Strangulation also occurs frequently,[7] but less is known about the types of clinical manifestations that result from this form of IPV. Other injuries that raise suspicion of IPV include fractures of the spine or trunk, bites, hair pulling, and open wounds. Those exposed to sexual abuse may show signs of trauma to the genital area, but sexual assault is associated with signs of injury in less than one third of cases.

Most victims of IPV presenting to health care settings do not have signs of obvious trauma but rather have a constellation of overlapping physical and mental health problems. A patient presenting with vague signs and symptoms or chronic somatic complaints, including pain, suggests the possibility of IPV. Other behaviors that suggest IPV include delay in seeking medical care, multiple cancellations of medical appointments, and noncompliance with vital medications (e.g., insulin).

There are no systematic reviews of studies on the overall physical health consequences of IPV, but an overview of studies reported increased rates of chronic physical conditions, particularly gynecologic, gastrointestinal, and nervous system disorders, as well as increased cardiovascular risk, although most of the studies were small and poorly adjusted for other risk factors. The overview also found that women with a history of abuse, particularly physical and sexual violence, were more likely to experience chronic pain and nonspecific symptoms, although an association between abuse and number of physical symptoms is also found in women who experience emotional abuse without any physical abuse. The World Health Organization study reported significant associations between women's lifetime experiences of partner violence and self-reported poor health and specific health problems in the previous 4 weeks, such as difficulty in walking, difficulty with daily activities, pain, memory loss, dizziness, and vaginal discharge. Other physical conditions that should raise suspicion of IPV include chronic gynecologic or gastrointestinal symptoms, such as irritable bowel syndrome and chronic pelvic pain. It should not be assumed, however, that there is a specific association of functional disorders, such as irritable bowel syndrome and fibromyalgia, over and above the greater reporting of physical syndromes in general. IPV exposure is associated with an increased risk for sexually transmitted infections, including human papillomavirus.

Exposure to any type of IPV can be associated with a wide range of emotional and behavioral symptoms[8]; depression and post-traumatic stress disorder (PTSD) are the two most commonly associated emotional conditions, but other anxiety disorders and substance abuse are also associated with IPV exposure. In the World Health Organization study, women who reported IPV at least once in their life reported three to four times more emotional distress, suicidal thoughts, and suicide attempts than nonabused women. There is strong evidence of increased risk for depression, anxiety, substance abuse, and PTSD, with longitudinal studies reporting increased morbidity after

violence. A recent meta-analysis examining the association between IPV against adult women and depressive conditions found a 2- to 3-fold increase in risk of major depressive disorder and a 1.5- to 2-fold increased risk of postpartum depression and elevated depressive symptoms.[9] The cross-sectional design of most studies examining associated impairment precludes conclusions about the causal role of IPV in these conditions, but the few published longitudinal studies show the onset or worsening of depression, PTSD, and substance abuse *after* exposure to IPV.

Pregnant women deserve special mention because IPV can threaten the health of both mother and fetus. Injury patterns during pregnancy are more likely to be central, including blunt trauma to the head, torso, abdomen, breasts, and genitalia. Abuse directed to the abdomen may lead to poor pregnancy outcomes and perinatal death. Although the evidence regarding a direct association between IPV in pregnancy and low birthweight has been conflicting, the increased risk for preterm birth as well as other factors, such as psychosocial stress, may play a role in adverse outcomes for infants, but more investigation is required.

Although it is beyond the scope of this chapter, there is increasing recognition that children's exposure to IPV shows a significant association with children's internalizing and externalizing problems, including trauma symptoms, developmental delay, educational problems, and long-term mental health conditions.[10]

Identification

Despite some guidelines recommending universal screening for IPV, randomized controlled trials have shown that although such screening increases the identification of women with IPV, it has not been shown to improve women's health outcomes or to reduce the occurrence of IPV. Consistent with the World Health Organization guidelines referred to before, IPV screening is not recommended for women of any age group.[11] However, it is important to be alert to the signs and symptoms associated with IPV, including those associated with the broad range of physical and mental health conditions referred to previously, and for clinicians to have a low threshold for asking about abuse. Indicators that suggest a higher likelihood of IPV include symptoms of depression, somatization, and PTSD in the female patient and a history of alcohol or drug abuse and unemployment in the male partner (or ex-partner). It is important, when asking about exposure to IPV, to do so privately, with no one else present, including a child (beyond infancy) or partner. If the inquiry or response is overheard, it could put the patient at risk for further IPV. A meta-analysis of qualitative studies of women's expectations and experiences reported that when the topic of IPV is raised, patients want questioning that is nonjudgmental, compassionate, and caring.[11] Women want to be asked about IPV with confidentiality ensured but do not want to be pressured to disclose. In some jurisdictions, however, disclosure of IPV when a patient has children in the home can lead to mandatory reporting to child protection services. It is important that patients be advised about the limits of confidentiality before being asked about IPV exposure.

Possible questions to ask if IPV is suspected include the following:

Sometimes partners or ex-partners use physical force. Has this ever happened to you?

Have you felt humiliated or emotionally harmed by your partner or ex-partner?

Are you now or have you ever been afraid of your partner or ex-partner?

Have you ever been physically threatened or hurt by your partner or ex-partner?

Have you been forced to have any kind of sexual activity by your partner or ex-partner?

Has your partner or ex-partner ever tried to control your behavior, for example, control where you go or whom you see?

The initial clinical response when IPV is identified should include validation of the experience (e.g., everyone deserves to feel safe at home), affirmation that violence is unacceptable, and expression of support. The clinician needs to acknowledge the complexity of IPV and respect the patient's individual concerns and decisions. The assessment should include an evaluation of safety; the patient should be asked if it is safe for her (or him or any children) to return home. The following are examples of safety considerations:

Has the frequency or severity of the violence increased?

Is the partner or ex-partner obsessed with the patient?

How safe does she (he) feel?

Does the partner or ex-partner have a weapon or access to one?

Has she (he) been threatened with a weapon?

Although a general discussion of gun violence is beyond the scope of this chapter, having firearms in the home is associated with an increased risk for homicide associated with IPV. Another predictor of domestic homicide is threats of deadly violence.

TREATMENT

Inquiry and Disclosure

The initial response of clinicians to the disclosure of IPV by female patients, whether the disclosure is spontaneous or the result of clinical inquiry, is crucial in gaining trust and is the basis of further management. IPV is a highly stigmatized condition, akin to sexually transmitted infection or substance abuse, with the added dimension of risk for further harm from breach of confidentiality. A number of qualitative studies of women's expectations and experiences (847 informants) reported consistent messages about how clinicians can respond appropriately to disclosure. *Before disclosure or questioning,* they should understand the problem, including knowing about the available community services and appropriate referral systems; ensure that the clinical environment is supportive, welcoming, and nonthreatening; place brochures and posters in the clinical setting; try to ensure continuity of care; assure abused women about matters of privacy, safety, and confidentiality; be alert to the signs of abuse and raise the matter when indicated; use verbal and nonverbal communication skills to develop trust; and be compassionate, supportive, and respectful toward abused women. *When the topic of IPV is raised,* they should be nonjudgmental, compassionate, and caring when questioning about abuse; be confident and comfortable asking about domestic violence; not pressure women to disclose abuse because simply raising the topic may be helpful to women; ask about abuse during the course of several interviews because a woman may disclose abuse at a later date; ensure that the environment is private and confidential; and provide time. *Immediate response to disclosure* should be nonjudgmental, with compassion, support, and belief of experiences; acknowledge the complexity of the problem and respect the woman's unique concerns and decisions; put the needs identified by the woman first and help ensure that social and psychological needs are met; take time to listen, to provide information, and to offer referrals to specialist help; validate her experiences, challenge assumptions, and provide encouragement; and respond to any concerns about safety. *Response in later interactions* should be patient and supportive, allowing her to progress at her own therapeutic pace; understand the chronicity of the problem and provide follow-up and continued support; respect the woman's wishes and not pressure her into making any decisions; be nonjudgmental if a woman does not follow up with referrals immediately; and give abused women an opportunity to disclose abuse at a later date.

Referral Services and Advocacy

Beyond their initial response and managing the medical sequelae of abuse, most generalists have neither the expertise nor the capacity to meet the specific needs of women experiencing IPV, which include legal, financial, housing, and safety needs. A key step, particularly in the context of current or recent violence, is an offer of referral to some sort of specialist support. Two main types of services have been evaluated: advocacy programs and psychological interventions (individual or group based). In general, advocates engage with individual clients who are being abused, aiming to empower them and linking them to community services. Core activities of advocacy include provision of legal, housing, and financial advice; facilitation of access to and use of community resources, such as refuges (shelters, safe houses) and emergency housing; and provision of safety planning advice. Advocates can also provide ongoing support and informal counseling. A Cochrane review of 10 randomized controlled trials (1527 participants) of domestic violence advocacy concluded that there was equivocal evidence that advocacy for women recruited in domestic violence shelters (refuges) had a beneficial effect on their physical and psychosocial well-being and was unable to draw any conclusions for women receiving advocacy in or referred from health care settings. A broader systematic review that included all controlled studies of domestic violence advocacy concluded that most showed a reduction in abuse, increased social support, and improved quality of life as well as increased use of safety behaviors and accessing of community resources. Only two of the studies had participants referred from health care settings. Clinicians should be able to refer patients to specialist IPV advocacy and are more likely to ask about abuse if they have the support of these services. If such services are not immediately available, shelters and refuges often provide these kinds of services for women both in residences and on an outreach basis.

Individual and Group Psychological Interventions

A systematic review of controlled studies of psychological interventions for survivors of IPV identified 17 studies, 7 of individual and 10 of group interventions. There was a wide range of individual psychological interventions that

TABLE 241-2 SUMMARY OF SELECT INTIMATE PARTNER VIOLENCE RECOMMENDATIONS FROM THE WORLD HEALTH ORGANIZATION

CATEGORY	RECOMMENDATION	QUALITY OF EVIDENCE	STRENGTH OF RECOMMENDATION
Woman-centered care	Women who disclose any form of violence by an intimate partner (or other family member) should be offered immediate support by clinicians, at a minimum. If clinicians are unable to provide this first-line support, they should ensure that someone else (within their health care setting or another that is easily accessible) is immediately available to do so.	Indirect	Strong
Identification of survivors	Universal screening is not recommended.	Low-moderate	Conditional
	Ask about exposure to IPV when assessing conditions that may be caused or complicated by abuse.	Indirect	Strong
	Written information about IPV should be available in all health care settings.	No relevant evidence	Conditional
Care for survivors	Women with preexisting diagnosed or IPV-related mental disorders should receive mental health delivered by health care professionals with a good understanding of violence against women.	Indirect	Strong
	Cognitive-behavioral therapy or eye movement desensitization and reprocessing interventions, delivered by health care professionals with a good understanding of violence against women, should be offered to women with post-traumatic stress disorder who are no longer experiencing violence.	Low-moderate	Strong
	Women who have spent at least 1 night in a shelter, refuge, or safe house should be offered a structured program of advocacy, support, and/or empowerment.	Low	Conditional
	For children who are exposed to IPV at home, a psychotherapeutic intervention should be offered.	Moderate	Conditional
Training of clinicians	Training at prequalification level in first-line support for women who have experienced IPV should be given to clinicians (in particular physicians, nurses, and midwives).	Very low	Strong
	Clinicians offering care to women should receive in-service training integrated with training on managing sexual assault.	Low-moderate	Strong
Health care policy	Care for women experiencing IPV should be integrated into existing health services rather than as a stand-alone service.	Very low	Strong
Mandatory reporting	Mandatory reporting to the police by clinicians is not recommended; clinicians should offer to report the incident to the appropriate authorities (including the police) if the woman wants this and is aware of her rights.	Very low	Strong

IPV = intimate partner violence. (From Feder G, Wathen CN, MacMillan HL. An evidence-based response to intimate partner violence: WHO guidelines. *JAMA* 2013;310:479-480.)

demonstrated improvements in psychological outcomes, including depression, PTSD, and self-esteem. Well-executed trials of individual cognitive therapy–based interventions for women with PTSD who were no longer experiencing violence provided reasonable evidence for this intervention, but this cannot be extrapolated to women still in an abusive relationship.[■■] All the studies of group psychological interventions showed improvement in one or more psychological or mental health outcomes, but with the exception of one study, they were poorly conducted. Consequently, the effectiveness of this type of intervention remains uncertain, particularly for women who are still experiencing IPV.

Treatment of Abuser and Couple Therapy

Although the assessment and treatment of the abuser should be carried out by mental health professionals with expertise in this area, it can be helpful for general clinicians to have some awareness of the effects of treatment. The evidence for batterer treatment is mixed, with the better-designed studies generally indicating little or no benefit or potential harm (i.e., increased recidivism).[■] To date, there is insufficient evidence to recommend specific treatment for those committing IPV. The evidence for couple therapy is mixed, with trial-level evidence indicating no benefit in a military sample.[■] Most authors caution that these couple therapy programs are not safe for many abused women, particularly those experiencing "intimate terrorism." Furthermore, when abusers are enrolled in treatment programs, it is important that women be provided with concurrent advocacy and support. There is some evidence to suggest that permanent, but not temporary, civil protection orders may be effective in reducing future violence.

PREVENTION

From a public health perspective, primary prevention of IPV is desirable, although most of the available research focuses on the health care response to the survivors of IPV, both while a woman is still exposed to abuse (secondary prevention) and when she is experiencing the long-term health problems associated with IPV (tertiary prevention).

Efforts aimed at primary prevention of IPV through educational programs have generally focused on changes in attitude, knowledge, skills, or self-reports of dating (relationship) violence. No studies to date have measured physical or emotional health outcomes. A meta-analysis assessed the efficacy of interventions aimed at preventing dating or relationship violence in adolescents and young adults; such violence is often considered a precursor to IPV in adulthood.[■] The authors concluded that there was no evidence that the interventions were effective in improving attitudes, behaviors, or skills related to relationship violence or in reducing episodes of relationship violence; there was a small improvement in knowledge about relationships. However, given that there was substantial heterogeneity between the studies, and that when those studies considered at high risk of selection bias were excluded, the only remaining study showed no effect on knowledge, the overall conclusion was no evidence of effect for the interventions.

There is no clinical trial evidence for the effectiveness of interventions provided in general medical settings with the aim of secondary prevention. A systematic review concluded that there is insufficient evidence to determine the effectiveness of interventions in preventing IPV against pregnant women.[■] However, an advocacy and empowerment program in antenatal clinics reduced psychological and minor physical violence,[■] and a program for pregnant African American women based on individual counseling sessions reduced violence and improved pregnancy outcomes.[■] Within an Australian family medicine setting, there was equivocal benefit in terms of mental health and safety of a brief counseling intervention delivered by physicians.[■] Outside of health care settings, intensive advocacy (12 hours or more duration) may reduce physical abuse among women leaving shelters or refuges after 12 to 24 months of follow-up, but not for shorter or longer follow-up. There is evidence that a training and support program for primary care clinicians improves identification of women experiencing abuse and referral to advocacy services.[■] The World Health Organization has published guidelines for the health care response to IPV with recommendations linked to the current evidence base. Table 241-2 outlines the key recommendations related to prevention of IPV and treatment of conditions associated with exposure to IPV. Particularly noteworthy are the recommendation on training for clinicians in first-line support to women who have disclosed IPV and the recommendations against screening and mandatory reporting.

PROGNOSIS

The prognosis of IPV with and without intervention is uncertain. Trials of interventions have small samples and short follow-up, and most have substantial attrition of participants. As far as the "natural history" of the condition is concerned, cohort studies are rare, and cross-sectional studies are potentially misleading. In a 3-year follow-up of participants who received an advocacy intervention after leaving a shelter, 36% had been assaulted by their original partner or a new partner in the 6 months before the interview. The difference

in re-victimization at 2 years between intervention and control arms did not persist, but there was still a significant difference in quality of life and social support among women receiving advocacy. In a U.S. cohort study, 44% of participants were still being abused after 3½ years.

Grade A References

A1. MacMillan HL, Wathen CN, Jamieson E, et al. Screening for intimate partner violence in health care settings: a randomized trial. *JAMA*. 2009;302:493-501.

A2. Klevens J, Kee R, Trick W, et al. Effect of screening for partner violence on women's quality of life: a randomized controlled trial. *JAMA*. 2012;308:681-689.

A3. O'Doherty LJ, Taft A, Hegarty K, et al. Screening women for intimate partner violence in healthcare settings: abridged Cochrane systematic review and meta-analysis. *BMJ*. 2014;348:g2913.

A4. Ramsay J, Carter Y, Davidson L, et al. Advocacy interventions to reduce or eliminate violence and promote the physical and psychosocial well-being of women who experience intimate partner abuse. *Cochrane Database Syst Rev*. 2009;3:CD005043.

A5. Feder G, Ramsay J, Dunne D, et al. How far does screening women for domestic (partner) violence in different health-care settings meet criteria for a screening programme? Systematic reviews of nine UK National Screening Committee criteria. *Health Technol Assess*. 2009;13:iii-iv, xi-xiii, 1-113, 137-347.

A6. Kubany ES, Hill EE, Owens JA, et al. Cognitive trauma therapy for battered women with PTSD (CTT-BW). *J Consult Clin Psychol*. 2004;72:3-18.

A7. Tirado-Munoz J, Gilchrist G, Farre M, et al. The efficacy of cognitive behavioural therapy and advocacy interventions for women who have experienced intimate partner violence: a systematic review and meta-analysis. *Ann Med*. 2014;46:567-586.

A8. Babcock JC, Green CE, Robie C. Does batterers' treatment work? A meta-analytic review of domestic violence treatment. *Clin Psychol Rev*. 2004;23:1023-1053.

A9. Dunford FW. The San Diego Navy experiment: an assessment of interventions for men who assault their wives. *J Consult Clin Psychol*. 2000;68:468-476.

A10. Fellmeth GLT, Heffernan C, Nurse J, et al. Educational and skills-based interventions for preventing relationship and dating violence in adolescents and young adults. *Cochrane Database Syst Rev*. 2013;6:CD004534.

A11. Jahanfar S, Janssen PA, Howard LM, et al. Interventions for preventing or reducing domestic violence against pregnant women. *Cochrane Database Syst Rev*. 2013;2:CD009414.

A12. Tiwari A, Leung WC, Leung TW, et al. A randomised controlled trial of empowerment training for Chinese abused pregnant women in Hong Kong. *Br J Obstet Gynaecol*. 2005;112:1249-1256.

A13. Kiely M, El-Mohandes AA, El-Khorazaty MN, et al. An integrated intervention to reduce intimate partner violence in pregnancy: a randomized controlled trial. *Obstet Gynecol*. 2010;115:273-283.

A14. Hegarty K, O'Doherty L, Taft A, et al. Screening and counselling in the primary care setting for women who have experienced intimate partner violence (WEAVE): a cluster randomised controlled trial. *Lancet*. 2013;382:249-258.

A15. Feder G, Davies RA, Baird K, et al. Identification and Referral to Improve Safety (IRIS) of women experiencing domestic violence with a primary care training and support programme: a cluster randomised controlled trial. *Lancet*. 2011;378:1788-1795.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. World Health Organization. Responding to intimate partner violence and sexual violence against women: WHO clinical and policy guidelines. http://apps.who.int/iris/bitstream/10665/85240/1/9789241548595_eng.pdf; Accessed January 20, 2015.
2. Stöckl H, Devries K, Rotstein A, et al. The global prevalence of intimate partner homicide: a systematic review. *Lancet*. 2013;382:859-865.
3. Abramsky T, Watts CH, Garcia-Moreno C, et al. What factors are associated with recent intimate partner violence? Findings from the WHO multi-country study on women's health and domestic violence. *BMC Public Health*. 2011;11:109.
4. Jonas S, Khalifeh H, Bebbington PE, et al. Gender differences in intimate partner violence and psychiatric disorders in England: results from the 2007 adult psychiatric morbidity survey. *Epidemiol Psychiatr Sci*. 2014;23:189-199.
5. Wu V, Huff H, Bhandari M. Pattern of physical injury associated with intimate partner violence in women presenting to the emergency department: a systematic review and meta-analysis. *Trauma Violence Abuse*. 2010;11:71-82.
6. Sprague S, Bhandari M, Della Rocca GJ, et al. Prevalence of abuse and intimate partner violence surgical evaluation (PRAISE) in orthopaedic fracture clinics: a multinational prevalence study. *Lancet*. 2013;382:866-876.
7. Sorenson SB, Joshi M, Sivitz E. A systematic review of the epidemiology of nonfatal strangulation, a human rights and health concern. *Am J Public Health*. 2014;104:e54-e61.
8. Trevillion K, Oram S, Feder G, et al. Experiences of domestic violence and mental disorders: a systematic review and meta-analysis. *PLoS ONE*. 2012;7:e51740.
9. Beydoun HA, Beydoun MA, Kaufman JS, et al. Intimate partner violence against adult women and its association with major depressive disorder, depressive symptoms and postpartum depression: a systematic review and meta-analysis. *Soc Sci Med*. 2012;75:959-975.
10. Stewart DE, MacMillan H, Wathen N. Intimate partner violence. *Can J Psychiatry*. 2013;58:Insert 1-15, Encart 1-17.
11. Feder G, Wathen CN, MacMillan HL. An evidence-based response to intimate partner violence: WHO guidelines. *JAMA*. 2013;310:479-480.

REVIEW QUESTIONS

1. A 25-year-old woman comes to the emergency department with a bruise to her left eye. She is accompanied by a woman whom she refers to as a friend. The patient explains that she was hit in the eye while playing baseball. The friend insists on remaining with the patient while the initial history is taken by the nurse. Which of the following is the most appropriate recommendation?

A. The friend should be asked politely to leave during the history taking, which should include exploration about the possibility of intimate partner violence (IPV).
B. The patient should first be asked if she wants her friend to stay, and if she agrees, the patient can be asked about IPV.
C. The patient should be asked if the friend is a partner; if not, it is acceptable to proceed with exploration of IPV in the presence of the friend.
D. Because the patient is accompanied by a friend, it is important not to inquire about exposure to IPV because of confidentiality.
E. Because the patient has given a plausible explanation for the bruise to her eye and there is no evidence for universal screening, she should not be asked about exposure to IPV.

Answer: A The patient has an injury that is suggestive of exposure to IPV. It is important that she be asked privately about the nature of the injury. Such inquiry is a form of case finding, not universal screening.

2. A 17-year-old pregnant woman comes to the family physician's office to discuss her plans for the pregnancy. During the visit, she discloses that her 20-year-old boyfriend has twice threatened to hit her before she became pregnant. Which of the following is the most appropriate approach by the family physician?

A. Ask the patient whether she has ever experienced any physical injury from her boyfriend; if not, advise her that they should attend couple counseling together.
B. Advise the patient that the risk will likely escalate during the pregnancy and she should leave the partner.
C. Give the patient the opportunity to discuss her concerns at a pace that is comfortable for her.
D. Advise the patient that this creates significant risk for her infant in the future, and it is not safe for her to remain in the relationship.
E. Explain to the patient that it is best to return another day when the appointment can be set aside to discuss the issue in more detail.

Answer: C The patient needs to be given the opportunity to discuss her concerns at a pace that is comfortable for her. The other responses could put the patient at risk (A) or are too directive (B and D) or not sufficiently responsive to the patient's needs (E).

3. A 56-year-old woman is a new patient in a family medicine practice. In her first consultation, she presents with a 16-year history of recurrent depression. In taking a comprehensive history, the family physician asks about any experience of abuse. She discloses that she has been frightened of her husband for years. He is extremely controlling and occasionally physically violent. What is the most appropriate initial response from the physician to this disclosure?

A. Have you ever thought about packing your bags and leaving him?
B. Have you ever thought about going to couple counseling?
C. Is there something you could do differently so that he becomes less angry?
D. That must be very difficult for you; I can help you find support.
E. That sounds frightening; do you want me to have a word with him?

Answer: D The priority must be initial validation, empathizing with the patient and signaling that support is available. The other responses are too directive (A) or could increase risk to the patient (B and F) or are victim blaming (C).

4. A clinical team is developing a policy on IPV. It should include which of the following?

A. A protocol for annual screening of all women patients for current abuse
B. A mandatory reporting requirement ensuring that the police are informed in all cases of serious assault
C. Recording of abuse in the medical record of the victim and the perpetrator
D. A training program on IPV for all clinical staff that includes a referral pathway to IPV advocacy support

Answer: D Training of clinicians is a prerequisite for an appropriate response to patients experiencing IPV, and a referral pathway to advocacy removes one of the barriers to the clinician's asking about abuse. The other responses are not based on evidence of effectiveness (A) or could increase the risk to the patient (B and C), also reducing the likelihood of disclosure to health care providers.

242

APPROACH TO THE PATIENT WITH METABOLIC BONE DISEASE

THOMAS J. WEBER

DIAGNOSIS

History

Patients with metabolic bone disease may present to the clinician in a number of ways, ranging from no symptoms to disabling musculoskeletal pain, depending on the nature of the underlying disorder. The most common conditions, osteoporosis and primary hyperparathyroidism, encompass a clinical spectrum that ranges from asymptomatic (diagnosed by low bone density and an elevated serum calcium, respectively)[1] to severe disease (fractures and bone pain). Less common conditions, such as osteomalacia, have more predictable presentations. In osteoporosis, fractures of the long bones (humerus, distal forearm, femur, and tibia) are clearly evident, whereas fractures at other sites (vertebrae, ribs, pelvis) may not be (see later and Chapter 243).[2] Patients with osteomalacia may complain of deep bone pain or aches, although often it is difficult for them to distinguish such pain from muscular pain. They may also report proximal muscle weakness that impairs their ability to ascend stairs. Vitamin D deficiency, the most common cause of osteomalacia, may confer similar bone and muscle complaints as well. In hyperparathyroidism, either self-reported or elicited fatigue may be a common complaint, along with mildly impaired cognition and memory. A history of recurrent nephrolithiasis is a hallmark of symptomatic primary hyperparathyroidism.

Physical Examination

Patients may exhibit physical clues to their skeletal condition. Height loss of more than 2 inches from a self-reported maximum, measured accurately with a calibrated stadiometer, may suggest the presence of vertebral compression fractures, which are clinically silent in up to three fourths of patients. Corresponding thoracic kyphosis may be present, and tenderness to palpation or percussion over the spinous process may suggest a recent vertebral fracture. Thoracic kyphosis will also impart quantifiable physical characteristics, including decreased and increased rib-pelvis and wall-occiput distances, respectively, which may be followed clinically in patients and significantly predict the presence of vertebral fractures. The method for the rib-pelvis distance measurement is illustrated in Figure 242-1. The wall-occiput distance is the space between the wall and the occiput of the head when a patient stands straight with the heel, buttocks, and back against the wall: it reflects the degree of kyphosis. Patients may also exhibit signs of secondary causes of osteoporosis (e.g., blue sclerae with osteogenesis imperfecta, goiter and proptosis with hyperthyroidism, facial plethora and purple striae with Cushing's syndrome). Patients with Paget's disease may have a skeletal deformity and warmth over the affected sites. Patients with osteomalacia often have tenderness to palpation over the tibia or other long bones due to expansion of the subperiosteal space by undermineralized osteoid with nerve irritation. These patients may also have a wide-based, "waddling" gait due to pain. Hyperparathyroid patients may have flank tenderness if active nephrolithiasis is present and can rarely have corneal calcification if hypercalcemia is severe and longstanding. Such patients rarely have a palpable parathyroid adenoma. If present, however, a diagnosis of parathyroid carcinoma should be entertained.

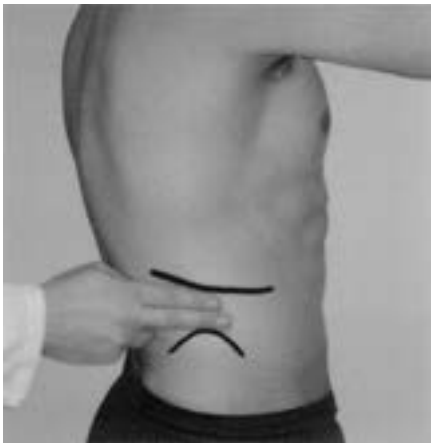

FIGURE 242-1. Method of assessing rib-pelvis distance. While standing behind the subject, the examiner holds his or her hands vertically and places them into the space between the inferior margin of the ribs and the superior surface of the pelvis in the mid-axillary line. The vertical distance is then measured in fingerbreadths.

Laboratory and Radiologic Investigations

Laboratory studies are a useful adjunct in the evaluation of metabolic bone disease, although their specificity depends somewhat on the disease in question. In particular, studies performed in the work-up of osteoporosis are generally not solely diagnostic but rather are supportive of secondary etiologies that may contribute to bone loss. Examples include thyroid-stimulating hormone, 25(OH)D, and testosterone levels (Table 242-1). In addition to a cursory work-up as detailed, additional investigations may also be indicated as clinically indicated and in individuals with a greater dual-energy x-ray absorptiometry (DXA) bone density deficit than expected for age (see later and Chapter 243).[3] In contrast, the diagnosis of other conditions is more securely based on abnormal studies, such as alkaline phosphatase (elevated in Pagets disease and depressed in hypophosphatasia) and parathyroid hormone (elevated in hyperparathyroidism). More sophisticated studies that target a specific diagnosis should be based on the history and examination (i.e., genetic testing for osteogenesis imperfecta). Bone turnover markers, which are cellular products of bone formation and resorption that can be measured in the blood and urine of patients, may provide noninvasive information on skeletal turnover (i.e., high or low) but cannot be used for diagnosis. They also have unacceptable biologic and measurement variability that precludes their clinical usefulness at this time.

Radiologic studies are critical to the diagnosis and management of these patients. Because the most common clinical event is fracture, plain radiographs of the involved skeletal sites are often indicated.[4] It is important to note that radiographs are relatively insensitive in identifying stress fractures and may also lag behind a frank fracture by hours or days. As such, additional, more sensitive modalities may be employed, including computed tomography and magnetic resonance imaging, to confirm a fracture. These studies also reveal characteristic patterns of skeletal involvement in certain conditions (i.e., Paget's disease). Whole body bone scintigraphy with the radioisotope techneticum-99m is the most sensitive tool to identify an active skeletal process but is nonspecific as to the nature of the underlying process (e.g., fracture, infection, malignancy). Perhaps most widely used and critical to management of osteoporosis is bone mineral density (BMD) testing, generally by DXA. As detailed in Chapter 243, DXA is a low-radiation, noninvasive examination of the spine, proximal femur, and distal forearm that may be used to identify and subsequently follow the treatment response to pharmacologic or conservative therapies. Finally, although rarely needed, tetracycline-labeled, transcortical bone biopsy of the iliac crest, with subsequent histomorphometic analysis, may be useful in the management of patients. Bone biopsy, which is generally performed by an orthopedic surgeon under conscious sedation, may be indicated to best guide management in patients with excessive bone fragility that cannot be adequately characterized by noninvasive means (e.g., patients with renal osteodystrophy, suspected osteomalacia, fractures with normal BMD by DXA).

TABLE 242-1 LABORATORY WORK-UP OF OSTEOPOROSIS

ALL OSTEOPOROSIS/OSTEOPENIA PATIENTS	AS CLINICALLY INDICATED
Serum creatinine, calcium, total protein, albumin, phosphorus, alkaline phosphatase, liver function tests	Serum and urine protein electrophoresis (SPEP and UPEP) (if total protein-to-albumin ratio is >2.0)
Complete blood count	Intact parathyroid hormone
Thyroid-stimulating hormone	24-Hour urine cortisol
24-Hour urine calcium and creatinine	Celiac panel (antigliadin/antiendomyseal antibodies)
Serum 25(OH)D	Fasting morning testosterone (men)

TREATMENT

Rx

Management of patients with metabolic bone disease is generally directed by the disease process, although there are some unifying aspects of treatment. Adequate intake of calcium and vitamin D, usually through a combination of diet and supplements, is recommended for patients with osteoporosis. High-dose vitamin D is indicated for low-vitamin D–related osteomalacia, and phosphorus in combination with vitamin D analogues (i.e., calcitriol) is necessary to heal osteomalacia and facilitate normal longitudinal growth in children and adolescents with certain osteomalacic conditions (e.g., X-linked hypophosphatemic rickets or X-linked hypophosphatemia). Weight bearing and resistive exercise are also advisable for osteoporotic patients, although physical therapy consultation may be indicated in patients at high risk for fracture (previous fractures, frequent falls). Pharmacotherapy with oral and parenteral bisphosphonates is frontline therapy for patients with both osteoporosis and Paget's disease because of the anticatabolic effect on excessive osteoclastic bone resorption that underlies these diseases. In addition, other antiresorptive drugs, such as raloxifene and denosumab, are approved for the treatment of osteoporosis. Parenteral bisphosphonates and denosumab also are effective in combating malignancy-related bone disease (e.g., metastatic breast cancer, multiple myeloma). Finally, the anabolic bone agent teriparatide, a recombinant parathyroid hormone analogue, is useful in "building" bone density and reducing fracture risk in patients with severe osteoporosis, defined as very low BMD, high fracture risk, and/or multiple fractures. Newer drugs in development that target more recently identified aspects of bone physiology (e.g., the Wnt pathway in bone formation), as well as specific derangements in less common diseases (e.g. FGF-23 antibody in X-linked hypophosphatemia), will necessarily further improve the management of patients with metabolic bone disease.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Leslie WD, Morin SN. Osteoporosis epidemiology 2013: implications for diagnosis, risk assessment, and treatment. *Curr Opin Rheumatol*. 2014;26:440-446.
2. McKenna MJ, Heffernan E, Hurson C, et al. Clinician approach to diagnosis of stress fractures including bisphosphonate-associated fractures. *QJM*. 2014;107:99-105.
3. Kling JM, Clarke BL, Sandhu NP. Osteoporosis prevention, screening, and treatment: a review. *J Womens Health (Larchmt)*. 2014;23:563-572.
4. McConnell CT Jr, Wippold FJ 2nd, Ray CE Jr, et al. ACR appropriateness criteria management of vertebral compression fractures. *J Am Coll Radiol*. 2014;11:757-763.

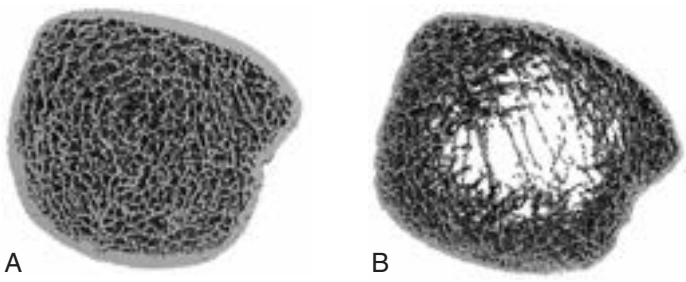

FIGURE 243-1. High-resolution peripheral computed tomography (Xtreme CT) of distal tibia in subject with normal bone mineral density (A) and severe osteoporosis (B). The deterioration in trabecular architecture with reduced trabecular number, trabecular thinning, increased trabecular spacing, generalized cortical thinning, and increased cortical porosity is readily appreciable in the osteoporotic subject. (From Griffith JF, Genant HK. New advances in imaging osteoporosis and its complications. *Endocrine.* 2012;42:39-51.)

243

OSTEOPOROSIS

THOMAS J. WEBER

DEFINITION

Osteoporosis is defined as a skeletal disorder characterized by compromised bone strength predisposing to an increased risk for fracture. The pertinent clinical outcomes of this disease include fractures, bone pain, height loss, and physical deformity. This definition was developed by the National Institutes of Health in 2000 to help clinicians better diagnose and treat patients with the disease. The concept of bone strength is central to understanding the disorder because patients who suffer an osteoporotic or fragility fracture may or may not have osteoporosis by bone mineral density (BMD) criteria. The World Health Organization defines osteoporosis as a BMD that is equal to or greater than 2 standard deviations (SD) below that of an average individual at peak bone mass (generally age 20 to 30 years, depending on the measured skeletal site). However, it is well established that most fragility fractures, which are defined as fractures occurring from the energy imparted from a fall from a standing height or less, occur in individuals who have low (osteopenia) or even normal BMD. (Osteopenia, or low bone density, is defined as a BMD between −1.0 and −2.5 SD below young average normal.) This observation is consistent with the lack of a specific BMD threshold for fracture. Given these observations, it is clear that other factors must also significantly influence fracture risk. Certainly, falls and traumatic injuries are a significant, independent risk factor for fractures. Excluding falls and trauma, however, studies to date have also identified qualitative factors that are integral to bone strength, including skeletal microarchitecture, bone turnover, damage accumulation (e.g., microfractures), and pattern or degree of mineralization. Newer technologies are currently in development to improve our understanding of how these qualitative changes in bone compromise skeletal strength, including high-resolution peripheral quantitative computed tomography (HR-pQCT) and high-resolution magnetic resonance imagine (MRI) (Fig. 243-1).[1] Although promising, these newer techniques are not widely available. More important, they have not enhanced fracture prediction to date over known risk factors and BMD. Indeed, the availability and widespread application of Internet-based fracture prediction tools such as FRAX, which incorporate independent, additive clinical risk factors for fracture with or without hip BMD, are currently the best approach for identifying individuals at risk for and most likely to benefit from treatment to prevent fragility fractures.

EPIDEMIOLOGY

Approximately half of white women will develop an osteoporosis-related fracture in their lifetime, which is greater than their risk for breast cancer, heart attack, and stroke combined. In addition, one in five men will also fracture.[2] More than 2 million fractures occur annually in the United States, at an estimated total direct cost of $17 billion. Nearly three fourths of them occur in women, with most occurring in white women. Nonetheless, there are no ethnic exclusions to developing the disorder. The most common site of osteoporotic fracture is the spine, accounting for more than 750,000 fractures annually. Fractures of the proximal femur, which disproportionately confer a greater cost than other osteoporotic fractures, account for 14% of incident fractures but nearly three fourths of costs. Additional sites of fracture include the distal forearm, proximal humerus, and pelvis, with the latter two more commonly occurring in elderly people. The risk for fracture increases markedly with age, although the pattern of fracture risk does differ by skeletal site. The risk of Colles fractures increases until the mid-60s and then plateaus, whereas the risk for hip fractures increases exponentially in woman after the age of 65 years. Vertebral fracture risk rises earlier than that of hip, although many spine fractures are not clinically apparent and only identified through radiographic assessment. Although clinically silent, such fractures do confer a significant independent risk for future fractures, particularly if they are of recent occurrence.[3] Men also ascribe to an age-independent increase in fracture risk, although the increase in incidence generally lags at least 5 to 10 years behind that of women (Fig. 243-2). The reason for the gender-based difference in fracture incidence is likely related to anatomic differences because although men and woman have similar volumetric bone density at a given skeletal site, bone size is larger in men than in women and confers an independent mechanical protection against fracture. Nonetheless, the independent contribution of age to fracture risk necessarily predicts a higher morbidity and cost related to osteoporosis in men as well as women, with an increase in the United States alone by the year 2025 to greater than 3 million fractures with an associated annual cost of $25.3 billion.

There are also significant ethnic and geographic differences in the rate of osteoporotic fractures. African Americans have a lower lifetime risk for osteoporotic fracture, approximately roughly half that of whites. Differences in bone size, bone microarchitecture (thicker trabeculae in blacks), body composition, calcium absorption in youth, and life expectancy are potential reasons for this observation. Asian Americans and Hispanics have a fracture risk that is intermediate between that of whites and blacks, despite the fact that the former have a BMD that generally approximates that of whites.

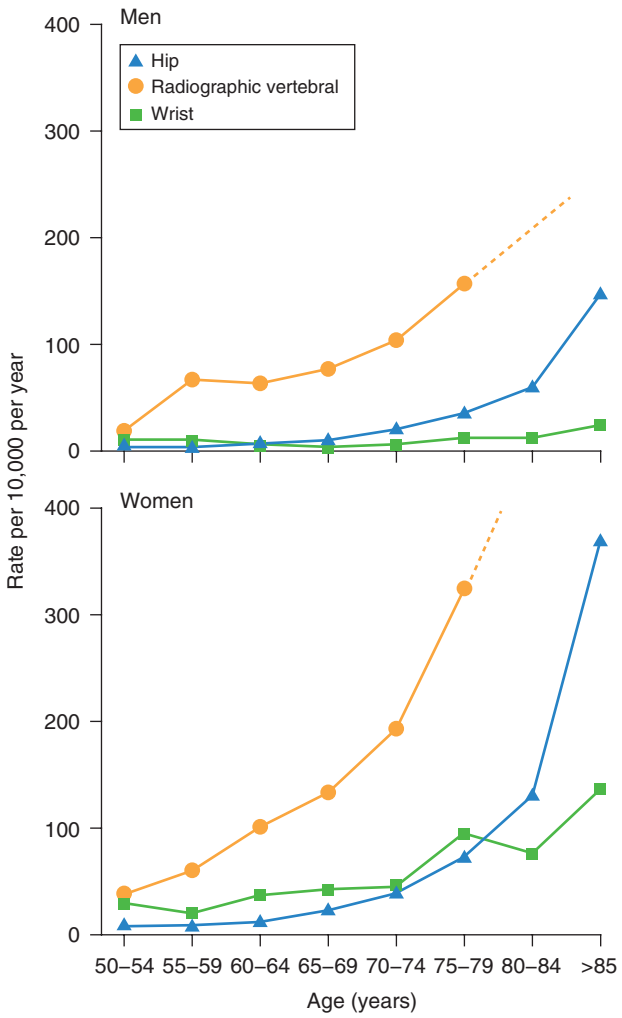

FIGURE 243-2. Age-specific and sex-specific incidence of radiographic vertebral, hip, and distal forearm fractures. (Data derived from European Prospective Osteoporosis Study and General Practice Research Database; from Sambrook P, Cooper C. Osteoporosis. *Lancet.* 2006;367:2010-2018.)

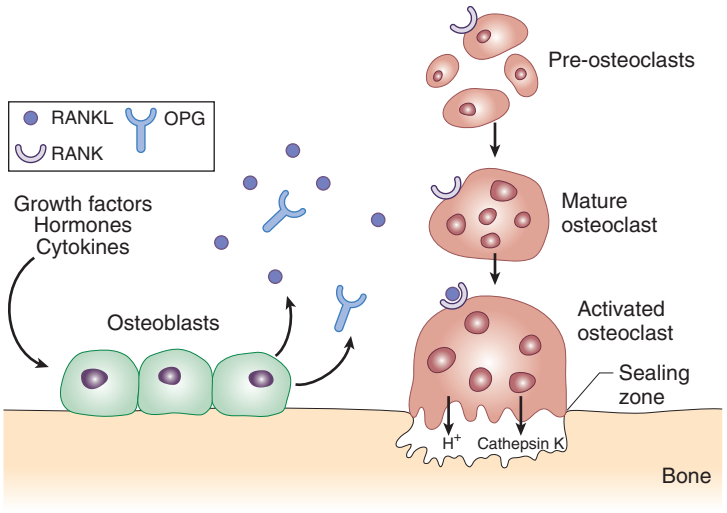

FIGURE 243-3. RANKL expressed by osteoblast lineage cells binds to RANK on the surface of pre-osteoclasts and mature osteoclasts, resulting in increased bone resorption through an increase in osteoclast differentiation, activity, and survival. Osteoprotogerin (OPG) is a "decoy receptor," also produced by osteoblasts, that binds to RANKL, preventing RANKL binding to RANK and thereby inhibiting osteoclastic bone resorption. It is the balance of RANKL and OPG that determines the ultimate rate of bone resorption. (Adapted from Lewiecki EM. New targets for intervention in the treatment of postmenopausal osteoporosis. *Nat Rev Rheumatol.* 2011;7:631-638.)

Indeed, the risk for hip fracture in U.S. Asian men and women is actually equal to or lower than that of U.S. blacks. The U.S. data are consistent with the global experience as well because hip fracture rates in China are lower than those observed in the United States, despite similar bone density at the hip. Differences in hip geometry, physical activity, and diet have been proposed as possible explanations. Despite this, rates of hip fracture have actually been increasing in the Far East, while declining in the United States for unclear reasons. Changing patterns of nutrition and physical activity could be responsible for the former, although the latter observation remains heretofore unexplained.

PATHOBIOLOGY

Normal Bone Biology

BMD in adults is determined by the magnitude of bone acquisition during adolescence and young adulthood, and the rate of bone loss that ensues thereafter. These processes are generally referred to as bone modeling and remodeling. Heritable factors, including gender and ethnicity, account for 60 to 80% of the variability in skeletal development, including peak bone mass, bone size, and bone geometry, although nutrition, lifestyle, and other factors also have a significant impact. Peak bone mass is achieved in most individuals by the early to late 20s and differs in timing by skeletal site (age 18 to 20 years for proximal femur, 25 to 30 years for spine). Modeling of the skeleton occurs during this time and represents a true increase in bone mass and bone size through endochondral ossification of the axial skeleton and periosteal apposition of the appendicular skeleton.

To best understand the underlying pathophysiology of osteoporosis, one must first appreciate the concept of bone remodeling. Skeletal remodeling is a finely orchestrated process of bone resorption and subsequent formation. It is a necessary physiologic function that results in repair of damaged bone and redistribution of the skeleton to adapt to changes in mechanical stress and to provide calcium to the systemic circulation for critical cellular processes. Recent evidence suggests that the osteocyte, which accounts for 90 to 95% of all bone cells, is the critical cell that regulates both resorption and formation. Osteocytes are derived from osteoblasts that are embedded within the bone matrix. During this maturation phase, this "osteoid-osteocyte" cell actively secretes and calcifies bone matrix material. In addition, mature osteocytes within bone contain dendritic processes that may directly regulate osteoblast recruitment and bone formation. Recent studies also suggest that mature osteocytes may form new bone within their lacunae. In addition, osteocytes produce proteins that regulate mineralization, including positive (PHEX, DMP-1) and negative (FGF-23) factors.

Osteocytes also regulate bone resorption, both directly through apoptosis proximate to skeletal microcracks or fatigue damage in need of repair and indirectly through enhancement of pre-osteoblast or mesenchymal stromal cell (MSC) development. (MSCs may also develop into adipocytes, chondrocytes, and muscle cells, depending on developmental stimuli.) These events result in the production, expression, and release of cytokines critical to osteoclast recruitment and development, including interleukin-1 (IL-1), interleukin-6 (IL-6), osteoprotogerin (OPG), and receptor activator of nuclear factor-κB ligand (RANKL). The cognate receptor for RANKL, RANK, is expressed on the surface of the developing and mature osteoclast, which itself is a derivative of cells of the monocyte-macrophage lineage. RANKL is a critical determinant of osteoclast recruitment, development, and survival, such that disruption of RANKL signaling results in high bone density fragility disorders (e.g., osteopetrosis). OPG, which is also produced by pre-osteoblasts, is a decoy receptor for RANK that binds to RANKL, preventing RANKL binding to RANK and thereby serving as an endogenous suppresser of osteoclast function. In essence, RANKL/OPG represent a "yin/yang" paradigm in which osteoclast biology and bone resorption are intricately regulated and controlled (Fig. 243-3).

Factors Affecting Peak Bone Mass and Remodeling

As mentioned, peak bone mass is achieved by the third decade. Longitudinal studies in children and adolescents suggest that hormonal, physical activity, nutritional, and genetic factors are all important in this process. Hereditary influence is the most important determinant and to date the least well understood. Growth hormone and sex hormones play critical roles in growth of the appendicular (long bones) and axial (vertebrae) skeleton, with the former maturing earlier than the latter (end of puberty vs. young adulthood). Males

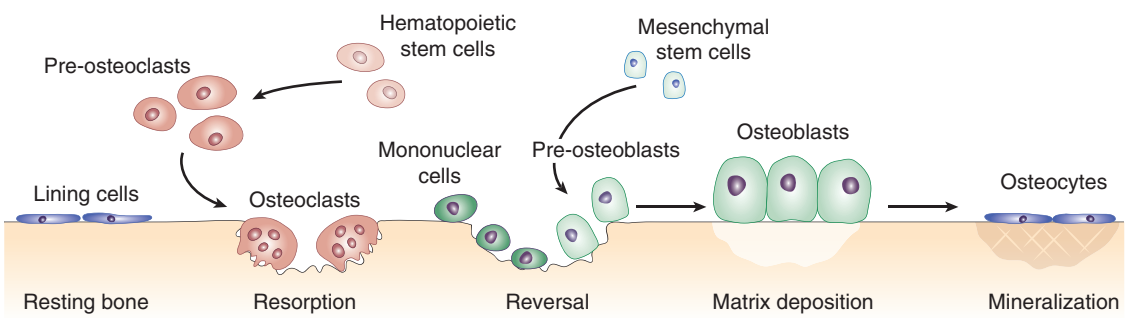

FIGURE 243-4. The basic multicellular unit (BMU) moving across the bone tissue. First, 10 to 20 osteoclasts resorb old or damaged tissue; then they recruit 1000 to 2000 osteoblasts, which produce new bone matrix. The BMU moves at a speed of 20 to 40 µm/day and survives for up to 6 months. (From Kapinas K, Delany AM. MicroRNA biogenesis and regulation of bone remodeling. *Arthritis Res Ther.* 2011;13:220.)

have larger bones as a result of greater periosteal (outer surface of bone) expansion of bone than occurs in females. Current evidence supports a positive effect of exercise and loading on bone size and mineral density, although a subsequent potential antifracture benefit later in adult life is not proved. Poor nutrition and concomitant disease may also affect bone accrual, primarily through delayed pubertal onset and progression, although studies do support a potential for "catch-up" growth, which depends on the degree of insult and timing of resolution.

Although these factors are all important, genetic factors appear to account for 60 to 80% of the variance in peak BMD. Genome-wide association studies have identified 62 distinct loci that are significantly associated with BMD.[4] Three known obligatory pathways of bone metabolism were identified by genome-wide association studies, namely Wnt, RANK-RANKL-OPG, and endochondral ossification. The RANK pathway was discussed previously. The Wnt signaling pathway is obligatory for bone formation, owing to its role in osteoblast proliferation and differentiation. It was initially discovered through study on kindreds with high bone mass and the osteoporosis pseudoganglioma syndrome, who have, respectively, activating and inactivating mutations in the Wnt pathway receptor LRP5. Endochondral ossification, a process that involves the cartilage growth plate and subsequent ossification of the cartilaginous skeleton, is dependent on transcription factors (SOX6, RUNX2) and proteins (parathyroid hormone–related peptide, bone sialoprotein 2, and osteopontin) that are necessary for the development of the cartilage growth plate, bone matrix mineralization, and osteoblast differentiation. Indeed, the first two pathways are already the target of existing and developing therapies for osteoporosis. Despite these advances, our understanding of the relationship between genes and bone mass accrual remains poor, as evidenced by the fact that less than 6% of the variance in femoral neck BMD is explained by the loci identified in genome-wide association studies.

In adults, bone remodeling is a physiologic process through which skeletal repair and adaption to changes in biomechanical stress occur. The basic multicellular unit (BMU), which is composed of bone-resorbing osteoclasts, bone-forming osteoblasts, bone lining cells, and embedded osteocytes, is the cellular apparatus that facilitates remodeling (Fig. 243-4). After the age of 30 years, this process is reasonably matched, resulting in rates of bone loss of only 0.3 to 0.5% per year from the third through fifth decades of life. As individuals age, mismatches in bone remodeling, either due to reduced formation, increased resorption, or a combination of both, result in greater rates of bone loss. Biologic changes, such as menopause and aging, systemic disease, personal vices, and medications (most notably glucocorticoids), may contribute to this imbalance. In addition, there is evidence that genetic factors also influence rates of bone loss and fracture. Although not as robust an influence as on peak BMD, twin and family-based studies suggest that genetic factors contribute roughly 30 to 50% of the variance in rates of bone loss at the spine, hip, and forearm. However, studies attempting to identify specific genes that are associated with accelerated rates of bone loss have been inconsistent to date, including ones that have identified the collagen gene (*COL1A1*), vitamin D receptor gene (*VDR*), estrogen receptor-1 gene (*ESR1*), interleukin-6 (IL-6), and the apoprotein AOE*4 allele as potential candidates. This inconsistency may be due to insufficient data sets that do not have repeated measures of bone density. Interestingly, there is a fairly robust heritability of menarche and menopausal onset, perhaps paralleling the observations on *ESR1* and IL-6 (whose production is suppressed by estrogen) and supporting hormonal pathways as a potential target of genetically mediated bone loss.

Although a better understanding of the genetic influences on bone mass accrual and subsequent loss is critically needed, it is perhaps more important to extend these analyses to the pertinent clinical outcome in this disease—fracture. This is further supported given the previously described significant unexplained variance between BMD and fracture risk. Unfortunately, such work to date has yielded disappointing results, likely in part because the heritability of fracture risk diminishes significantly with aging. In addition, the contributions of individual genetic variants to fracture risk are quite small and potentially falsely negative when studied. Nonetheless, there is preliminary evidence that simulated genetic profiling using a large number of variants could increase the discriminatory value of existing fracture prediction models and significantly re-stratify individuals in need of or not requiring antifracture therapies.[5]

Mechanisms of Bone Loss

"Natural" and Aging-Related Bone Loss

As noted previously, net bone loss during adult life is expected, although the rate of bone loss accelerates during the latter decades because of a number of factors. Perhaps most important for women, rates of bone loss increase substantially in the perimenopausal and early postmenopausal years, amounting to losses of 1 to 5% per year. Unfortunately, there are currently no biologic markers that help determine which women a priori are "rapid losers." Bone loss is slower in obese women, likely because of higher estrogen levels afforded by production through aromatization in adipose tissue. A decline in circulating estrogen is primarily responsible for bone loss following both natural and surgical menopause, which is mediated primarily through upregulation of cytokines (most notably RANKL) and a resultant increase in the number, activity, and depth of osteoclast-mediated bone resorption sites. In addition, OPG production is diminished, further amplifying bone resorption, although estrogen replacement can restore OPG production while reducing RANKL expression and thereby help mitigate bone loss during this period. Although bone resorption and formation do occur sequentially during this period, resorption outpaces formation because of potentiation of the former aided by the release of soluble cytokines, resulting in a significant uncoupling of bone remodeling and accelerated bone loss. Fortunately, this phase of rapid bone loss is typically limited to 5 to 7 years in most women.

Current evidence supports a key role in bone loss for reduced estrogen levels in men, perhaps owing to reduced aromatase activity in fat tissue. Decline in circulating insulin-like growth factor-I levels also occurs with aging, which may inhibit the production of MSCs and pre-osteoblasts, in addition to limiting periosteal bone expansion. The latter is a normal adaptive process to aging that increases bone size, more so in men than in women, somewhat attenuating the observed increase in fracture risk with aging. Bone loss in later decades also is caused by lower 25-hydroxy vitamin D (25[OH]D) levels, related to restricted sun exposure and a reduced capacity to generate pre–vitamin D in the skin with aging, as well as limited intake of vitamin D and calcium. Indeed, more than 75% in individuals in the United States are either vitamin D deficient or insufficient (25[OH]D < 30 ng/mL). Low vitamin D results in secondary hyperparathyroidism, which accelerates cortical bone loss, as well as likely increasing the risk for falls. Increase in bone

marrow adiposity is also inversely and significantly correlated with aging-related bone loss. This process likely reduces bone formation through the production of fewer osteoblast precursors, which share a common mesenchymal stem cell lineage with adipocytes. It is also likely the result of increased activation of the nuclear receptor peroxisome proliferator activator receptor-γ (PPARG), which commits stem cells toward adipocyte and away from osteoblast differentiation. Additionally, fat-derived adipokines (adiponectin and leptin) may negatively affect bone remodeling through a "bone-fat" connection by reducing bone formation, further exacerbating bone loss.[6] Finally, loss of weight and muscle mass (sarcopenia) is associated with bone loss, with evidence of a recently identified common factor (myostatin) that may provide insights on this muscle-bone interaction.[7]

Secondary Causes and Clinical Impact of Bone Loss

Bone loss also occurs as a result of secondary processes, such as diseases and medications. In fact, such processes are identifiable in more than one fourth of individuals with osteoporosis and may be more likely with greater degrees of skeletal deficit. A list of secondary causes of osteoporosis, based on a systems-based approach, is detailed in Table 243-1. As expected, endocrine disorders, including hypogonadism and Cushing's syndrome, predominate. Nonetheless, bone loss in several conditions is mediated indirectly through attendant effects on vitamin D metabolism (e.g., malabsorption, chronic liver disease) and is the case with certain medications as well (antiepileptic drugs phenytoin and phenobarbital). In the case of vitamin D deficiency, undermineralization of bone may confound the clinical picture of bone loss and should be considered first before initiation of bone active medications (see later). Rarely, malignancies can be the principal cause of osteoporosis, with the best example being multiple myeloma (Chapter 187). Myeloma causes bone loss through uncoupling of bone resorption and formation, the latter occurring through Wnt inhibition by the protein Dickopf. Medications also cause bone loss through both osteoclastic activation and osteoblastic inhibition. Reduced bone formation is the mechanism by which glucocorticoids cause bone loss, both through exogenous administration and exogenous overproduction. Previous work has confirmed reduced osteoblast and osteocyte function and survival as the principal mechanism of glucocorticoid-induced osteoporosis. Anorexia nervosa likely interferes with the anabolic effect of insulin-like growth factor-I on bone. Alcohol excess appears to suppress osteoblast function, perhaps both directly and indirectly through associated malnutrition, but it may also possibly cause bone loss through hypogonadism, which results in accelerated turnover. Finally, relatively rare connective tissue disorders, such as osteogenesis imperfecta and Marfan syndrome, increase skeletal fragility by means of disturbances in the skeletal matrix and mechanical integrity rather than altered bone remodeling.

TABLE 243-1 SECONDARY CAUSES OF OSTEOPOROSIS

ENDOCRINE DISORDERS
Hypogonadism: female and male
Hypercortisolism: endogenous and exogenous
Hyperthyroidism
Hyperparathyroidism
Idiopathic hypercalciuria
Diabetes mellitus: type 1 and type 2
NUTRITIONAL AND GASTROINTESTINAL DISORDERS
Malabsorption: celiac disease, gastrointestinal bypass
Vitamin D deficiency
Cirrhosis (including primary biliary cirrhosis)
Pancreatic insufficiency
Inflammatory bowel disease
Cystic fibrosis
Anorexia nervosa, bulimia
HEMATOLOGIC AND ONCOLOGIC DISORDERS
Multiple myeloma
Hemolytic anemia
Hemoglobinopathies: thalassemia, sickle cell
Myeloproliferative neoplasms
Skeletal metastases
Pompe's disease
Mastocytosis
CONNECTIVE TISSUE AND METABOLIC DISORDERS
Osteogenesis imperfecta
Ehlers-Danlos syndrome
Marfan syndrome
Homocystinuria
Gaucher's disease
Pompe's disease
MEDICATIONS
Glucocorticoids
Thyroxine (excessive)
Antiepileptics
Heparin
Gonadotropin-releasing hormone agonists
Depo-Provera
Immunosuppressants: tacrolimus, cyclosporine
Chemotherapy
Selective serotonin reuptake inhibitors*
Proton pump inhibitors*
Thiazolidinediones*
Alcohol
MISCELLANEOUS
Rheumatoid arthritis
Immobilization
Juvenile osteoporosis
Pregnancy-associated osteoporosis

*Association based.

The resulting clinical impact of bone loss and the persistence of effect depend on a number of factors, both potentially modifiable and nonmodifiable. The age at which bone loss occurs (e.g., young adult vs. 80-year-old) affects the ability to recover BMD with resolution or treatment of the condition (younger people have more robust recovery than elderly people) and also appears to hold for conditions of accelerated bone resorption (e.g., immobilization) or reduced bone formation (e.g., glucocorticoids). Predicated on and paralleling this observation, BMD response to treatment is generally more pronounced in individuals with higher rates of bone turnover, likely reflecting to some extent a "filling in" of bone remodeling space. Menopausal stage in women also influences rate of bone loss as described previously, with greater rates of decline occurring within the first 5 to 7 years of cessation of menses (Chapter 240). Concomitant vitamin D deficiency and resulting secondary hyperparathyroidism also potentiate rates of BMD loss, particularly in elderly people; it is thought to be a major contributing factor to the increase in hip fracture incidence in this group. Tobacco and alcohol overuse may also accelerate bone loss, owing to global effects that include reduced sex hormones, altered calcium metabolism, and weight loss and frailty, resulting in uncoupled bone remodeling. Therefore, a distinct appreciation of the clinical context is critical to defining an appropriate diagnostic and treatment approach to patients with osteoporosis.

CLINICAL MANIFESTATIONS

Historically, osteoporosis was diagnosed in an individual presenting with a low trauma or fragility fracture, typically of the vertebrae or hips. Classic representations of women with the so-called dowager's hump or kyphotic deformity were common depictions of the disease. Currently, however, the disease is appreciated both for its clinical and subclinical manifestations, because of both the advent of bone density testing and the appreciation that many vertebral fractures are clinically silent. This approach is akin to paradigms that identify a surrogate marker for both diagnosis and risk stratification, such as hypertension for stroke and hyperlipidemia for myocardial infarction.

A history of fragility fracture is strongly suggestive of osteoporosis, although there is evidence that a history of high trauma fractures also identifies persons with low BMD and those at higher risk for low-trauma fractures. The National Osteoporosis Foundation considers fractures of the spine, proximal femur, distal forearm, and proximal humerus "major" osteoporotic fractures, although other skeletal sites are also prone to fragility fractures. These include the pelvis, ribs, and proximal tibia, although there is controversy about whether ankle fractures should be considered as such. Fractures of the spine are generally from the mid-thoracic region through the lower lumbar region, with the greatest frequency at T11 through L2.[8] Patient often present after a fall or a spinal flexion-loading event in which they may hear a "pop" and complain of sharp midline back pain that may radiate to the flanks. Patients may also present with complaints of back "tiredness," which is improved with sitting or lying down. This symptom is likely related to paraspinal weakness or spasm from abnormal spinal curvature that occurs with

chronic vertebral compression. Back pain may commonly be related to other pathology, such as degenerative disc and spine disease that is concomitantly present. This is important to note because low bone mass in and of itself does not cause pain, unless it is due to osteomalacia (see later). Vertebral fractures may occur without acute symptoms as well, as noted later. In contrast, nonvertebral fracture events are always clinically evident. Hip fractures generally occur with falls, although they may rarely occur with limited force such as twisting.

Physical examination may also indicate the presence of osteoporosis and associated fractures, as well as potentially identifying underlying secondary processes contributing to the disease. Measured height loss, best confirmed using a calibrated device such as a stadiometer, of greater than 4 cm since young adult maximum height is suggestive of prior vertebral fractures. Height loss also occurs with scoliosis and aging (approximately ⅓ inch of height is lost per decade after age 50 years). A kyphotic deformity of the upper thoracic spine may be present, although it is important to distinguish it from accentuated cervical lordosis with associated prominence of T1. Spinal tenderness to palpation and percussion can occur with an acute vertebral compression fracture. Palpable tenderness of the long bones may suggest underlying osteomalacia instead, due to periosteal expansion and nerve irritation. Reduced rib-pelvis and increased wall-occiput distances are correlated with vertebral fractures as well.

In addition to the history and physical examination, radiologic findings may identify the presence of osteoporosis, sometimes somewhat surreptitiously. Plain films can detect bone loss by means of accentuation of vertical striations on spine radiographs that represent loss of horizontal trabeculae, although this generally indicates BMD loss of at least 25% or more. Kyphosis and compression fractures may be present, and patients often are unaware of the deformities because nearly three fourths of such fractures occur without acute pain. Furthermore, radiologic reporting of these fractures is inconsistent, suggesting that, if possible, the clinician review available digitized lateral chest radiographs and even lateral scout films often available from computed tomography (CT) to identify such fractures (Fig. 243-5). This is critical for optimal management, given the aforementioned risk conferred by previous fractures on future fracture events. The degree of compression fracture is also important because more severe fractures (>25% vertebral height loss) appear to better predict future fractures, as do nonvertebral fractures. When fractures are suspected, CT and MRI may be used, given that plain radiographs have a lower sensitivity acutely and with stress fractures. MRI also can be used to define a vertebral fracture with persistent swelling and edema based on T2 characteristics, potentially identifying patients who could benefit from vertebroplasty or kyphoplasty (see later). Finally, whole body bone scintigraphy is the most sensitive test for fracture but can be falsely positive because of inflammation, infection, or tumor and usually is positive for 6 to 12 months after a fracture event.

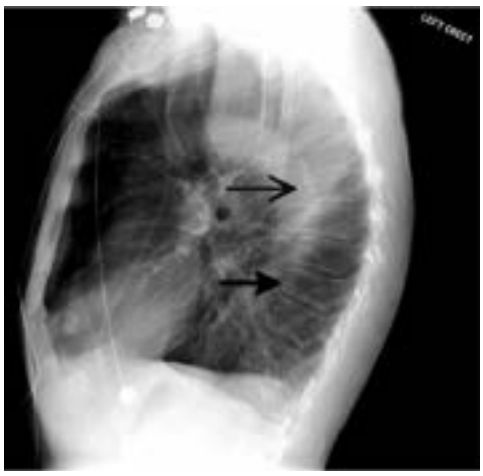

FIGURE 243-5. Incidental vertebral compression fractures on chest radiograph. Lateral radiograph of the chest of a 74-year-old man studied for cough. No relevant pulmonary abnormality was noted on the frontal radiograph (not shown). Examination of the thoracic spine shows the presence of a mild anterior wedge compression fracture of T9 (*thick arrow*) and moderate anterior wedge fracture of T6 (*thin arrow*). Neither fracture was reported on the radiographic report.

DIAGNOSIS

Although a recent fragility fracture is a reasonable basis for diagnosis of osteoporosis, other skeletal conditions should also be entertained, including inherited and acquired osteomalacias and pathologic fracture due to malignancy. These disorders can often be distinguished by history and physical examination, although additional investigations may be required. This distinction is critical because therapies may differ greatly between disorders. Most patients with osteoporosis are diagnosed on the basis of BMD measurement, generally by dual-energy x-ray absorptiometry (DXA). DXA is a low-radiation-based radiologic measurement of the areal bone density (g/cm^2) of the lumbar spine, proximal femur, and distal radius. Osteoporosis can be diagnosed if the BMD of a postmenopausal woman or man older than 50 years is more than 2.5 SD below young average normal (T score ≤ −2.5). A T score between −1.0 and −2.5 is considered low bone density or osteopenia, and a Z score (age-matched BMD) in premenopausal women and men younger than 50 years that is more than 2 SD below that of an a average age-matched individual is considered low bone density for age. BMD is an independent predictor of fracture risk, such that the relative risk for fracture increases by 1.5- to 2-fold for each 1 SD decrease in T score. In addition, fracture risk increases exponentially below a T score of −2.5. Furthermore, BMD of the femoral neck may be used in fracture prediction models such as FRAX to better define an individual's risk for subsequent fracture (see later). In addition to DXA, other modalities are also used to diagnose osteoporosis, including quantitative CT of the spine (QCT) and wrist and tibia (pQCT), finger DXA, and ultrasound of the calcaneus or wrist. Measurement of BMD by all of these techniques has been shown to globally predict fractures, akin to DXA. QCT and pQCT provide additional information on cortical and trabecular bone compartments but are accompanied by higher radiation exposure and poorer reproducibility compared with DXA. Ultrasound is radiation free and easy to operate but is less sensitive in diagnosing osteoporosis and does not measure change in a reliable fashion in response to age or treatment, making it useful as a screening modality but not for longitudinal care.

Although DXA is an effective diagnostic tool, several potential limitations and caveats need to be considered by the clinician. First, DXA cannot distinguish between low bone density and undermineralized bone matrix, the latter of which occurs in osteomalacia (Chapter 244). BMD may also be quite disparate between regions, perhaps in more than one third of individuals. This inconsistency results from a number of factors, including differences in bone composition (predominantly trabecular bone in the spine and cortical bone in the one-third radius) with resultant variations in rates of bone loss due to aging and disease (vertebral bone loss with menopause and glucocorticoid use versus cortical bone loss in hyperparathyroidism). Degenerative changes due to aging, such as facet osteoarthritis and aortic calcification, may artifactually raise spine BMD value. Given these considerations, the lowest skeletal site should be used for diagnosis. Finally, BMD should be measured longitudinally on the same DXA machine if possible, because of intermachine and intermanufacturer differences that may confound the ability to validly measure change over time. Despite these caveats, DXA remains the best method to diagnose and manage osteoporosis by bone density testing.

Despite its utility, bone density has been limited historically in optimally predicting fracture risk in individual patients. In addition, BMD does not take into account clinical factors that independently predict fracture. Under this premise, fracture prediction models have been developed that combine BMD and risk factors to better stratify fracture risk. The best known and most widely used of these prediction models is FRAX. FRAX was developed by the World Health Organization in collaboration with national and international osteoporosis foundations as an Internet-based computer algorithm that defines a person's 10-year risk for hip and major osteoporotic fracture (hip, clinical spine, forearm, and proximal humerus all combined). The model uses country-specific data on clinical risk factors and femoral neck BMD to calculate fracture probability and is available as a web-based tool that can be used by clinicians with their patients to assist in making informed decisions on osteoporosis management (http://www.shef.ac.uk/FRAX/). FRAX can be also be used to define country-specific recommended diagnostic and treatment thresholds. An example of this is the National Osteoporosis Foundation guidance that 10-year risks equal to or exceeding 3 and 20% for hip and major fracture risk, respectively, warrant consideration of pharmacologic treatment, which is based on cost-effective analyses in the United States. Furthermore, the number needed to treat (NNT) can be determined to best

inform patients of their expected risk and benefit of treatment (e.g., bisphosphonate use roughly reduces hip fracture risk by half, or from 10 to 5%, with NNT of 1/0.05, or 20 patients treated to prevent one hip fracture). Despite its utility and ease of use, FRAX does have limitations. These include inability to use patients who are not treatment naïve, underestimation of fracture risk in those with disparately lower BMD in the spine than hip, absence of fall history/fall risk in the model, and the use of fixed clinical risk factors. Modifications have been suggested for the latest guidelines, including adjusting FRAX score up or down based on glucocorticoid dose. In addition, although fracture risk calculators that incorporate fall risk are available (e.g., from the Garvan Institute), they do not include the competing risk of mortality as FRAX does. As such, FRAX should be viewed, not as a perfect, but rather as a complementary tool to BMD in best defining a person's risk for fracture and candidacy for pharmacologic intervention.

Finally, all patients presenting with osteoporosis require an assessment for secondary causes of bone loss, given that 20 to 25% of women and perhaps an even greater portion of men will have identifiable additional etiologies that may contribute to bone loss (see Table 243-1). Most patients will have had routine chemistry, hematology, and thyroid studies as part of their annual examination. The 25(OH)D level should be measured in all patients for multiple reasons, as previously discussed. Additional investigations may also be considered, as directed by the clinical history and physical examination. In addition, a greater degree of BMD deficit (i.e., lower Z score) indicates a need for more extensive testing, given the greater likelihood of secondary causes being present. Bone turnover markers (BTMs) are serum and urinary products of bone formation or resorption that can also be used to assist in management. Available tests include bone-specific alkaline phosphatase, osteocalcin, type I procollagen amino-terminal propeptide, and type I procollagen carboxy-terminal propeptide as formation markers, and serum and urine C- and N-terminal peptides of type I collagen as resorption markers, among others. Their use is predicated on studies showing that high bone turnover increases fracture risk independent of BMD. In addition, fracture risk reduction correlates well with reduction in bone turnover based on clinical trials with anticatabolic agents. Nonetheless, their clinical utility has been tempered to date by several issues. First, there is significant biologic variability due to nonmodifiable (e.g., age, gender, underlying comorbid disease, medications) and modifiable (e.g., time of day, food intake, presence of fracture) factors that limit the ability to detect meaningful change over time in an individual patient. Second, optimal specimen processing is required for valid results and interpretation. Finally, and perhaps in part secondary to these issues and others, evidence to date does not demonstrate significant benefit of BTMs in individual patients in securely predicting bone density increase, fracture risk reduction, or cost effectiveness through patient feedback and improved adherence. Therefore, at present, BTMs should not be used in routine clinical practice, although they could help inform management in more complicated cases of metabolic bone disease.[9]

PREVENTION AND TREATMENT

Calcium

Adequate intake of calcium is critical to the optimal accumulation and maintenance of bone mineral density. Calcium supplementation has meaningful impact on BMD and fracture risk reduction, although the latter is much less certain. BMD is modestly improved by 1 to 2%, although evidence of a definitive reduction in hip and nonvertebral fracture risk when given without vitamin D is lacking at present, although previous studies have suggested a trend toward vertebral fracture risk reduction. Given the established increase in rate of nephrolithiasis and a possible, albeit unproven, potential increase in nonfatal cardiac events with higher dose calcium supplementation, it would seem prudent to recommend that adults with osteoporosis obtain 1200 to 1500 mg of calcium from a combination of supplements and dietary sources.

Vitamin D

Appropriate circulating levels of 25(OH)D are necessary for optimal intestinal absorption of calcium and skeletal accrual and maintenance. Despite this, a significant proportion of children and adults have vitamin D levels that would be deemed insufficient (i.e., 25[OH]D < 20 ng/mL). Data in adults with osteoporosis confirm a benefit of vitamin D supplementation for fracture risk reduction, although the effect is dependent on the patient population and the amount of supplementation. Doses of 400 to 800 IU of vitamin D combined with 1000 mg of calcium reduce the risk for hip fracture in postmenopausal women and men aged 65 years and older, although the benefit is less certain for community-dwelling individuals than for those in assisted living centers. In addition, it appears that a 25(OH)D level of at least 30 ng/mL is needed to reduce the risk for hip fracture, although there is some controversy over this recommendation based on the results of other meta-analyses with likely different methodologies. In addition, a daily vitamin D intake of at least 800 IU also reduces the risk for falls, likely by improving muscle strength and reducing body sway. It should be noted that the recent U.S. Preventive Services Task Force recommendation against the use of calcium and vitamin D[10] was based on the general U.S. population and does not pertain to patients with osteoporosis. Finally, although activated vitamin D analogues such as calcitriol and α-calcidiol have been shown to reduce fracture risk, they are generally not indicated based on unacceptable risk for hypercalcemia. The exception to use of vitamin D analogues is possibly patients with stage 3 and 4 chronic kidney disease, wherein treatment of secondary hyperparathyroidism could provide skeletal benefit.

Exercise and Lifestyle

Physical activity is also a critical element of osteoporosis management, which can be indirectly inferred based on the known profound effects of decreased gravitational force (i.e., immobilization, paraplegia, weightlessness in space) on inducing bone loss. Physical activity likely also confers additional benefits through enhanced muscle strength, improved cardiovascular status, and reduction in fall risk. Meta-analysis confirmed a modest benefit of exercise on lumbar spine (mean difference = 0.85%) and trochanteric BMD (mean difference = 1.03%) in postmenopausal women compared with placebo, although it did not show significant changes in femoral neck or total hip BMD. However, studies to date have not confirmed an improvement in bone strength with exercise in this patient group. Studies concerning middle-aged and older men are much more limited in number and quality, although preliminary evidence suggests that resistance training with or without impact-loading activities has the greatest BMD benefit. Importantly, although none of the aforementioned studies have demonstrated a clear antifracture benefit from exercise, there are abundant data that multiple targeted exercise interventions do reduce either the risk for falling (Tai Chi) or both the rate and risk for falling (group and home-based exercise programs), which is most likely an inciting event in older patients incurring an osteoporotic fracture. Finally, modification of aberrant lifestyles is also indicated in patients with osteoporosis, especially tobacco cessation and moderation of caffeine, carbonated beverage, and alcohol intake. Data are lacking, however, on whether these reduce overall fracture risk.

Medications

There is robust evidence that pharmacologic therapy significantly reduces the risk for osteoporotic fracture in a clinically meaningful and cost-effective manner.[11] Medications approved for osteoporosis can be classified based on their mechanism of action: anticatabolic (i.e., antiresorptive) and anabolic (i.e., bone building).

Anticatabolic Agents

Anticatabolic medications, or antiresorptive as they were more commonly known, inhibit osteoclast recruitment, function, and/or survival, resulting in reductions in skeletal turnover and bone loss. These agents, depending on the potency and persistence of bone effect, reduce the number of new activation sites (BMUs) and the bone remodeling space, thereby improving BMD while strengthening the skeletal microstructure and reducing fracture risk.

Bisphosphonates

Bisphosphonates (BPs) are the most widely prescribed and used medications for the treatment of osteoporosis, owing in large part to good tolerability and an ability to dose them infrequently (from once weekly to once yearly, depending on the drug). BPs are chemically engineered analogues of the naturally occurring molecule pyrophosphate in which a carbon is substituted for an oxygen. As a result, BPs have an extremely high affinity for hydroxyapatite crystals within bone. After incorporation into bone, BPs are taken up by osteoclasts and thereafter inhibit cellular attachment, function, and survival. The carbon side-chain molecules largely determine skeletal affinity and potency of BP effect. The first-generation BP etidronate, which is not approved in the United States for treatment of osteoporosis, is the least potent agent of the class. It must also be given in an interrupted fashion for 2 weeks every 3 months owing to the potential to cause focal osteomalacia, and it may cause lower gastrointestinal symptoms (i.e., abdominal pain and diarrhea). Nonetheless, it is has been shown to reduce the risk for vertebral but not nonvertebral nor hip fractures.

Three oral bisphosphonates are approved by the U.S. Food and Drug Administration (FDA) and currently available in the United States: alendronate, risedronate, and ibandronate, in order of time since initial FDA approval. All three drugs are also available as generic preparations, although some differences do exist between the brand name and generic drugs in regard to the

TABLE 243-2 STRENGTH OF EVIDENCE FOR THE REDUCTION OF RISK FOR FRACTURE TYPES WITH PHARMACOTHERAPY IN WOMEN WITH POSTMENOPAUSAL OSTEOPOROSIS

	FRACTURE SKELETAL SITES			
	VERTEBRAL	**NONVERTEBRAL**	**HIP**	**WRIST**
Alendronate	•••	•••	•••	•
Ibandronate	•••	••	•	ı
Risedronate	•••	•••	•••	•
Zoledronate	•••	•••	•••	ı
Denosumab	•••	•••	•••	ı
Teriparatide	•••	••	•	ı
Raloxifene	•••	ı	ı	ı

Strength of evidence symbol legend: ı = insufficient strength of evidence; • = low strength of evidence; •• = moderate strength of evidence; ••• = high strength of evidence.
(Adapted with permission from Levis S, Theodore G. Summary of AHRQ's comparative effectiveness review of treatment to prevent fractures in men and women with low bone density or osteoporosis: update of the 2007 report. *J Manag Care Pharm.* 2012;18[4 Suppl B]:S1-S15, discussion S13.)

inactive excipients. The oral BPS may be administered once weekly (alendronate and risedronate) or once monthly (risedronate and ibandronate), fasting in the morning with water only, and the patient must remain fasting in a sitting or standing position for 30 to 60 minutes after the dose. Recently, a delayed-release formulation of risedronate (Atelvia) was approved that may be taken immediately after breakfast. The most common side effect is precipitation or aggravation of gastroesophageal reflux, although most patients tolerate the drugs without difficulty. In light of this side effect and a potential risk for esophageal irritation and ulceration, these drugs are contraindicated in patients with functional or anatomic disorders of esophageal transit (i.e., esophageal stricture, achalasia). All three drugs significantly reduce the risk for vertebral fractures, although high-strength evidence for hip and nonvertebral fracture risk reduction exists for alendronate and risedronate but not ibandronate (Table 243-2). In addition, studies confirm a persistent BMD and likely antifracture benefit after 5 years of therapy.

Parenteral BPs are also approved and available for osteoporosis treatment, although they should be considered second line to oral BPs based on overall risk-benefit assessment in most osteoporotic patients. They may be considered for use in patients with contraindications to oral BPs (e.g., esophageal disease, inability to sit upright and/or fast after dose), documented or expected poor adherence to oral BPs, or failure to respond to oral BPs or other FDA-approved therapies (recurrent fractures, declining BMD). Zoledronic acid, 5 mg once yearly, and ibandronate, 3 mg quarterly, may be given, although high-strength evidence would favor the use of zoledronic acid, given its unequivocal effect on spine, hip, and nonvertebral fracture risk reduction. Intriguingly, zoledronic acid has also been shown to reduce mortality in women and men following a low-trauma hip fracture, although the mechanism of the mortality benefit is unknown. Finally, BMD remains stable and the antifracture effect likely persists for 3 years after three annual doses of zoledronic acid. Both drugs are associated with an approximately 15 to 20% likelihood of a flu-like reaction, typically consisting of fever, arthralgias, and myalgias, usually limited to the first infusion, and generally lasting 24 to 48 hours, although symptoms lasting weeks to months have rarely been reported to the FDA. Both drugs confer a higher risk as well for delayed healing of exposed bone in the oral cavity compared with oral BPs (see later).

Rare, considerably more serious adverse effects have been associated with both oral and intravenous BPs. Osteonecrosis of the jaw, which is defined as exposed bone within the oral cavity for more than 8 weeks following an invasive dental procedure (e.g., tooth extraction, dental implant) or spontaneous tooth loss, occurs in roughly 1 in 10,000 to 100,000 patients treated with oral BPs, although it likely occurs in 1 in 1000 to 10,000 in intravenous BP–treated patients with osteoporosis. Current evidence suggests that microbial biofilm formation on an acellular bone surface, perhaps facilitated by BPs and the non-BP drug denosumab (see later), may be operative in the development of this disorder. As such, patients on intravenous BPs should maintain optimal oral hygiene and consider a BP holiday or delay in dose if invasive oral procedures are planned. Atypical femoral fractures have also been recently described in patients on long-term bisphosphonate therapy, generally after 5 years or more of treatment. Patients will typically have prodromal thigh or groin pain, which is referable to a stress fracture of a thickened lateral femoral cortex, inferior to the greater trochanter. These fractures can be bilateral in nature and may be identified radiographically with plain films, MRI, or CT. These patients are at risk for low-trauma, severe, oblique, "chalk-stick" fractures, which often represent orthopedic repair and healing challenges. Fortunately, the estimated prevalence of atypical femoral fractures is low (~1 in 5,000 to 10,000). Nonetheless, the severe manifestations of osteonecrosis of the jaw and atypical femoral fractures make it prudent for clinicians to consider a BP drug holiday, particularly given strong evidence of continued benefit on discontinuation.

Selective Estrogen Receptor Modulators

Selective estrogen receptor modulators (SERMs) are compounds that bind to the estrogen receptor and thereby influence bone and reproductive biology. As with estrogen (see later), SERMs are anticatabolic agents in bone, acting through a reduction in cytokines (RANKL, tumor necrosis factor-α) that engender osteoclast activation and function. Raloxifene is the only FDA-approved drug for prevention and treatment of osteoporosis in menopausal women, although the breast cancer drug tamoxifen likely has skeletal benefits as well. Both drugs have antiestrogenic effects in the breast and are FDA approved for the prevention of breast cancer in high-risk patients. Raloxifene does reduce the risk for vertebral fractures by approximately 30 to 50% but does not reduce the risk for hip and nonvertebral fractures. This antifracture profile positions it as an alternative to bisphosphonates in postmenopausal women with osteopenia and a relatively low risk for hip and other nonspine fractures. The most common side effects include hot flushes and leg cramps in about 10 to 15% and about 5% of patients, respectively. SERMs also increase the risk for deep vein thrombosis, with an absolute risk of roughly 1 in 400, akin to that seen with oral estrogen replacement therapy (ERT). Raloxifene has also been associated with an increased risk for fatal stroke in women at higher baseline risk for stroke, likely precluding its general consideration in women older than 65 years.

Estrogen

ERT, either alone or in combination with a progestin in women with an intact uterus, had historically been a frontline agent in the management of osteoporosis in postmenopausal women (Chapter 240). ERT prevents bone loss if administered to women at menopause and significantly increases BMD by approximately 3 to 5% in woman who are well into their menopausal years. Although lower doses of estrogen may have skeletal benefits, more standard doses of estrogen (0.625 mg of conjugated equine estrogen and 1.0 mg of ethinyl estradiol) have been proved efficacious. Long-term estrogen therapy reduces the risk for all clinical fractures by about 27%, based on the available moderate-quality evidence. ERT is also the most efficacious agent available for treatment of vasomotor symptoms. These data notwithstanding, ERT is associated with an increased risk for stroke (34% increase), and the use of continuous combined hormone replacement therapy (HRT) confers an unacceptable greater global risk than benefit in woman initiating HRT, based on the results of the Woman's Health Initiative. These results, however, may not be applicable to the younger postmenopausal population, based on differences in cardiovascular risk, although data confirming this are currently lacking. Both ERT and HRT are also associated with a two- to three-fold increase in the risk for venous thromboembolic disease. Therefore, ERT/HRT is recommended only for postmenopausal women at significant risk for fracture for whom other antifracture therapies are unsuitable.

Denosumab

As detailed previously, increased osteoclast activation through the RANKL pathway is a key mechanism through which bone loss occurs in menopause and other osteoporotic conditions. Intuitively, a therapy that targets this process directly would be desirable. Denosumab is a fully human monoclonal antibody to RANKL that is FDA approved for the treatment of osteoporosis in postmenopausal women and in men, as well as for individuals with breast and prostate cancer to reduce bone loss associated with hormonal deprivation therapy. It is administered twice yearly as a subcutaneous injection in the clinic and clearly reduces the risk for spine, hip, and nonvertebral fractures. Denosumab does not undergo hepatic or renal metabolism and thus can potentially be used in patients with more advanced renal dysfunction, unlike BPs. In contrast to BPs, it is reversible, such that robust bone loss ensues once the medication is stopped. Denosumab is well tolerated in clinical studies, although a higher incidence of skin conditions (eczema and erysipelas) and infections, including serious infections that required hospitalization, were observed in drug- versus placebo-treated subjects. Therefore, the drug is likely not suitable for patients on immunosuppressant therapy who are at higher baseline risk for infection.

Anabolic Agents

Although anticatabolic drugs are effective at retarding bone loss and reducing fracture risk, anabolic or "bone-building" drugs would be preferred. Teriparatide (TPTD) is a recombinant human parathyroid hormone analogue that encompasses amino acids 1 to 34 and was approved by the FDA in 2002. Given

as a self-administered once-daily subcutaneous injection, TPTD is truly anabolic based on robust increases in bone density (~10% over 2 years in the lumbar spine) and bone formation as determined by bone biopsies and other sophisticated imaging studies. More important, TPTD significantly reduces the risk for vertebral and nonvertebral fractures by approximately two thirds and one half, respectively. Because bone resorption increases along with bone formation, bone loss generally ensues on cessation of therapy, necessitating the initiation of an anticatabolic bone drug to preserve the increase in BMD facilitated by TPTD. Finally, although it is plausible to consider that a combination of TPTD and an anticatabolic drug is more beneficial than either drug alone, evidence from randomized controlled trials to date has failed to confirm this. Recent studies, however, suggest that the combination of TPTD and denosumab may have a truly synergistic effect on BMD.

TPTD is more expensive than other treatments for osteoporosis, although it is generally covered by insurance in patients who have severe osteoporosis (based on BMD and/or fracture risk) and who cannot tolerate or have contraindications to other antifracture agents. The drug is generally well tolerated, with the most common adverse effects being dizziness and leg cramps. TPTD has a black box warning, based on the fact that toxicology studies in rats revealed an increase in risk for osteosarcoma in animals treated with supratherapeutic doses of the drug, particularly in growing animals. Given this, the drug is contra-indicated for patients who are at a higher baseline risk for osteosarcoma, including patients with Paget disease and previous therapeutic radiotherapy, as well as younger individuals with open epiphyses. Fortunately, there has not been an observed increase in the rate of osteosarcoma in teriparatide treated patients above that expected in the general population to date.

Other Therapies and Treatment Considerations

Currently Available and Emerging Therapies

Nasal calcitonin is FDA approved and available at the time of this writing for treatment of postmenopausal osteoporosis, although it is widely considered the weakest antifracture agent based on marginal vertebral fracture benefit. In addition, recent human studies have suggested a possible link to cancer, potentially further limiting its clinical utility and future availability in the United States. Strontium ranelate is approved in Europe for the treatment of osteoporosis and may have a dual proformation-anticatabolic effect on bone. It has been shown to reduce the risk for vertebral and nonvertebral fractures, as well as clinical osteoporotic fractures. It is not available for use in the United States, and alternative forms of strontium salts cannot be assumed to be effective as well. In addition, BMD by DXA cannot be followed in patients on strontium because of artifactual increases in BMD related to the incorporation into bone of the strontium salt.

Emerging therapies on the horizon will likely provide additional tools to treat this debilitating disease, including new anticatabolic agents (e.g., cathepsin K inhibitors) and new anabolic therapies (e.g., sclerostin antibody). Odanacatib is an oral, small molecule that reversibly inhibits cathepsin K, which is produced by activated osteoclasts and primarily is responsible for the breakdown of type 1 collagen. Odanacatib also does not significantly suppress bone formation, perhaps "uncoupling" bone turnover in a favorable fashion to potentiate improvements in BMD. Sclerostin is a naturally occurring inhibitor of the Wnt pathway and bone formation, and preliminary clinical studies do confirm a significant anabolic effect and BMD increase with intermittent administration of a monoclonal antibody to sclerostin. Interestingly, unlike TPTD, inhibition of sclerostin does not appear to stimulate bone resorption, potentially affording greater and more persistent gains in BMD. Ongoing and future clinical studies are needed to confirm an antifracture benefit of this compound.

Glucocorticoid-Induced and Male Osteoporosis

As detailed previously, glucocorticoids are a major cause of and the most common etiology of medication-related secondary osteoporosis. Glucocorticoids are prescribed for a number of common inflammatory conditions, often in a chronic, long-term manner. They are potent suppressors of bone formation and at higher doses likely increase bone resorption, principally through central suppression of sex steroid production. This resultant "uncoupling" of bone turnover can result in dramatic declines in BMD within the first 6 months of starting therapy. In addition to bone loss, there is good evidence to support that individuals on glucocorticoids may fracture at a higher level on BMD compared with non-glucocorticoid-treated patients. Fracture rates are increased as well with doses of prednisone as low as 2.5 mg per day, although the increase in risk appears to attenuate with glucocorticoid discontinuation. The treatment approach to glucocorticoid-induced osteoporosis is similar to osteoporosis in general, with the exception that attempts should be made to reduce the steroid dose to as low as the underlying treated disease will permit.[12] Calcium and vitamin D are important adjuncts but are insufficient to prevent bone loss or fractures. Although not clearly evidence-based, replacement of deficient sex steroids is a reasonable strategy in younger individuals who are at lower risk for fracture. The BPs alendronate, risedronate, and zoledronic acid are FDA approved for glucocorticoid-induced osteoporosis in women and men, although the established benefit is based primarily on BMD improvement. A more logical and indeed superior treatment of glucocorticoid-induced osteoporosis is TPTD, which as an anabolic drug more directly addresses the primary mechanism of bone loss in glucocorticoid-induced osteoporosis: osteoblast inhibition. TPTD is FDA approved for treatment of glucocorticoid-induced osteoporosis in women and men and is superior to alendronate in improving BMD and vertebral fracture risk reduction. Although the drug was used for 36 months in this head-to-head trial, treatment is advised for no more than 24 months based on previously mentioned safety considerations.

Male osteoporosis historically has been under-recognized and underappreciated by primary care clinicians and patients alike, although the current data support a significantly more prevalent and clinically significant disorder. More than 2 million men in the United States have osteoporosis, and one in four men older than 50 years will suffer a fragility fracture in their remaining lifetime. Roughly 30% of vertebral and hip fractures combined occur in men, and these are the more common fractures in older men. In addition, men have a substantially higher mortality after hip fracture compared with women. As in women, aging, low body weight, and prior fragility fractures are independent predictors of fracture. In some contradistinction to women, however, osteoporosis in men is more commonly multifactorial in etiology, with the most common secondary causes being excess glucocorticoids, hypogonadism, and alcohol overuse. Despite these associations and others (current smoking, history of falls), there is not at present sufficient evidence to warrant use of a specific testing or screening strategy to identify men at higher risk for fracture. The laboratory work-up of male osteoporosis is similar to that for women, with the exception of a morning fasting testosterone level. Idiopathic osteoporosis may also occur, particularly in younger men with no discernable cause. Genetic factors may well be important in these men, with studies suggesting an association with lower production and circulating levels of estrogen. As in women, primary treatment of male osteoporosis is targeted at lifestyle changes, adequate nutrition (calcium and vitamin D), and exercise. Bisphosphonates (oral and intravenous), denosumab, and TPTD are all effective at improving BMD in men. Although more limited in scope, antifracture efficacy is evident for denosumab in men with prostate cancer on androgen deprivation therapy. True antifracture efficacy for the other agents is either less convincing or absent, based on the paucity of randomized controlled trial data, although this should not be construed as a reason not to treat. Testosterone replacement of men with significant biochemical hypogonadism (total T score < 200 ng/dL) does improve bone density, although data on fracture risk reduction are lacking. In older men (>50 years) at a substantial risk for fracture based on history and risk factors, androgen replacement should be considered second line behind the aforementioned other therapies, based on overall risk-benefit and lack-of-fracture data.

Vertebroplasty and Kyphoplasty and Low-Intensity Vibration

Although often clinically silent, vertebral fractures may cause acute and severe back pain. In addition, up to one third of vertebral fractures remain chronically painful, perhaps related to incomplete healing or instability of the fracture. Over the past decade, vertebroplasty and kyphoplasty have been developed and advanced to reduce the morbidity associated with acute spine fractures. These invasive procedures introduce, through the spinal pedicles, a cement-like substance (polymethylmethacrylate) to the compressed vertebral body, with (kyphoplasty) or without (vertebroplasty) use of saline-infused balloon tamps that permit a few millimeters of elevation of the vertebral end plates. Initial randomized trials suggested a benefit of vertebroplasty over conservative management in patients with acute vertebral fractures, although a recent meta-analysis of patient-level data from two randomized controlled trials did not confirm this finding. Additionally, there may be a concern about fracture of adjacent vertebrae following the procedure, reinforcing the need for further, adequately powered and designed clinical trials.

Low-intensity vibration is also under active investigation as an anticatabolic and possibly anabolic intervention for osteoporosis. Animal studies using low-intensity vibration appears to show enhanced osteoblast and hindered osteoclast development, thereby "coupling" bone remodeling. Clinical studies suggest a modest but significant BMD benefit in postmenopausal women and other groups (children with cerebral palsy, adults on prolonged bed rest), although further studies are needed to confirm a true clinical and ideally antifracture benefit of this intervention.

PROGNOSIS

It stands to reason that, based on the information and data discussed previously, the burden incurred by individual patients and society as a whole can be significantly lessened through a combination of diagnostic, preventive, and therapeutic interventions. Although there is no true "cure" for osteoporosis, current pharmacotherapies reduce the risk for fracture roughly by half.

This reduction is critical because there is robust evidence to suggest an independent increase in mortality after an osteoporotic fracture, including fractures of the spine, humerus, tibia, and pelvis, as well as the proximal femur. Moreover, available data, primarily from randomized controlled trials with bisphosphonates, confirm a statistically significant reduction in death with pharmacologic treatment of osteoporosis, although the mechanism of this effect is not known.[13] These data further underscore the importance of identifying and treating patients with osteoporosis.

Grade A References

A1. Avenell A, Gillespie WJ, Gillespie LD, et al. Vitamin D and vitamin D analogues for preventing fractures associated with involutional and post-menopausal osteoporosis. *Cochrane Database Syst Rev*. 2014;14:CD000227.

A2. Bischoff-Ferrari HA, Willett WC, Wong JB, et al. Prevention of nonvertebral fractures with oral vitamin D and dose dependency: a meta-analysis of randomized controlled trials. *Arch Intern Med*. 2009;169:551-561.

A3. Bischoff-Ferrari HA, Dawson-Hughes B, Staehelin HB, et al. Fall prevention with supplemental and active forms of vitamin D: a meta-analysis of randomised controlled trials. *BMJ*. 2009;339:b3692.

A4. Howe TE, Shea B, Dawson LJ, et al. Exercise for preventing and treating osteoporosis in postmenopausal women. *Cochrane Database Syst Rev*. 2011;7:CD000333.

A5. Gillespie LD, Robertson MC, Gillespie WJ, et al. Interventions for preventing falls in older people living in the community. *Cochrane Database Syst Rev*. 2012;9:CD007146.

A6. Levis S, Theodore G. Summary of AHRQ's comparative effectiveness review of treatment to prevent fractures in men and women with low bone density or osteoporosis: update of the 2007 report. *J Manag Care Pharm*. 2012;18:S1-S15.

A7. Lyles KW, Colon-Emeric CS, Magaziner JS, et al. Zoledronic acid and clinical fractures and mortality after hip fracture. *N Engl J Med*. 2007;357:1799-1809.

A8. Black DM, Rein IR, Boonen S, et al. The effect of 3 versus 6 years of zoledronic acid treatment of osteoporosis: a randomized extension to the HORIZON-Pivotal Fracture Trial (PFT). *J Bone Miner Res*. 2012;27:243-254.

A9. Marjoribanks J, Farguhar C, Roberts H, et al. Long term hormone therapy for perimenopausal and postmenopausal women. *Cochrane Database Syst Rev*. 2012;7:CD004143.

A10. Cummings SR, Ensrud K, Delmas PD, et al. Denosumab for prevention of fractures in postmenopausal women with osteoporosis. *N Engl J Med*. 2009;361:756-765.

A11. Kanis JA, Johansson H, Oden A, et al. A meta-analysis of the effect of strontium ranelate on the risk of vertebral and non-vertebral fracture in postmenopausal osteoporosis and the interaction with FRAX(®). *Osteoporos Int*. 2011;22:2347-2355.

A12. McClung MR, Grauer A, Boonen S, et al. Romosozumab in postmenopausal women with low bone mineral density. *N Engl J Med*. 2014;370:412-420.

A13. Saag KG, Zanchetta JR, Devogelaer JP, et al. Effects of teriparatide versus alendronate for treating glucocorticoid-induced osteoporosis: thirty-six-month results of a randomized, double-blind, controlled trial. *Arthritis Rheum*. 2009;60:3346-3355.

A14. Staples MP, Kallmes DF, Comstock BA, et al. Effectiveness of vertebroplasty using individual patient data from two randomised placebo controlled trials: meta-analysis. *BMJ*. 2011;343:d3952.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Griffith JF, Genant HK. New advances in imaging osteoporosis and its complications. *Endocrine*. 2012;42:39-51.
2. Leslie WD, Morin SN. Osteoporosis epidemiology 2013: implications for diagnosis, risk assessment, and treatment. *Curr Opin Rheumatol*. 2014;26:440-446.
3. van den Bergh JP, van Geel TA, Geusens PP. Osteoporosis, frailty and fracture: implications for case finding and therapy. *Nat Rev Rheumatol*. 2012;8:163-172.
4. Richards JB, Zheng HF, Spector TD. Genetics of osteoporosis from genome-wide association studies: advances and challenges. *Nat Rev Genet*. 2012;13:576-588.
5. Nguyen TV, Eisman JA. Genetic profiling and individualized assessment of fracture risk. *Nat Rev Endocrinol*. 2013;9:153-161.
6. Kawai M, de Paula FJ, Rosen CJ. New insights into osteoporosis: the bone-fat connection. *J Intern Med*. 2012;272:317-329.
7. Bonewald LF, Kiel DP, Clemens TL, et al. Forum on bone and skeletal muscle interactions: summary of the proceedings of an ASBMR workshop. *J Bone Miner Res*. 2013;28:1857-1865.
8. Gerdhem P. Osteoporosis and fragility fractures: vertebral fractures. *Best Pract Res Clin Rheumatol*. 2013;27:43-55.
9. Naylor K, Eastell R. Bone turnover markers: use in osteoporosis. *Nat Rev Rheumatol*. 2012;8:379-389.
10. Moyer VA. Vitamin D and calcium supplementation to prevent fractures in adults: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2013;158:691-696.
11. Andreopoulou P, Bockman RS. Management of postmenopausal osteoporosis. *Annu Rev Med*. 2014;66:329-342.
12. Rizzoli R, Biver E. Glucocorticoid-induced osteoporosis: who to treat with what agent? *Nat Rev Rheumatol*. 2015;11:98-109.
13. Grey A, Bolland MJ. The effect of treatments for osteoporosis on mortality. *Osteoporos Int*. 2013;24:1-6.

REVIEW QUESTIONS

1. A 60-year-old woman presents to her physician to discuss her recent bone density results and management options. She has treated hypertension and a history of atrial fibrillation, as well as rheumatoid arthritis, for which she takes infliximab. Family history is notable for a hip fracture in her mother after a fall at age 65 years. She drinks socially and does not smoke. Her physical examination is unremarkable, including only 1 inch of height loss from her young adult maximum. Her spine examination reveals normal curvature, no kyphosis, normal rib-pelvis distance of 3 fingerbreadths, and 0 fingerbreadth wall-occiput distance. Laboratory studies are normal, including 25(OH)D level. Duel-energy x-ray absorptiometry (DXA) bone density reveals lumbar spine, femoral neck, and total hip T scores of −2.5, −3.0, and −2.7, respectively. Her FRAX 10-year estimates of hip and major osteoporotic fracture are 5.6 and 29%, respectively. What is the best management recommendation for this woman?

A. Calcium and vitamin D supplementation alone
B. Hormone replacement therapy
C. Raloxifene
D. Oral bisphosphonate
E. Denosumab

Answer: D This patient has osteoporosis based on T score at the lumbar spine and proximal femur. In addition, she has an absolute fracture risk that supports pharmacologic intervention on a cost-effectiveness basis. Oral bisphosphonates are effective and generally frontline therapy. Active use of immunosuppressive therapy increases her risk for infection, which was seen more frequently in denosumab-treated patients in randomized controlled trials. Her cardiovascular history increases her risk for stroke, which has been observed more frequently in raloxifene- and estrogen-treated patients. Finally, calcium and vitamin D are important adjuncts to her management but should not be considered adequate alone for fracture risk reduction in this woman.

2. A 65-year-old woman presents with 6 months of progressive lower extremity pain and describes difficulty ascending stairs because of pain and weakness. She also brings a recent outside DXA bone density test, which shows total hip T and Z scores of −4.5 and −2.5, respectively. She does not have a history of fragility fractures. Past medical history is notable for hypertension and long-standing irritable bowel syndrome. She takes 600 mg of calcium and 400 IU of vitamin D daily, but no other medications. Family history is negative for osteoporosis or parental hip fracture. She does not smoke or drink alcohol. Physical examination is notable for tenderness to palpation over the mid-tibia bilaterally. She also has a wide-based, nonantalgic gait. What is the next best choice for her skeletal management?

A. Double her calcium and vitamin D daily intake.
B. Measure serum 25(OH)D level.
C. Start an oral bisphosphonate.
D. Measure fasting serum C telopeptide level (CTx)
E. Start teriparatide therapy.

Answer: B This patient has lower bone density than expected for age based on Z score of −2.0 or less. This suggests a secondary cause for low bone mineral density (BMD) besides menopause. Her clinical presentation is consistent with osteomalacia, which cannot be distinguished from osteoporosis based solely on BMD. In addition, the prevalence of vitamin D deficiency or insufficiency and established hip fracture efficacy warrant its identification and treatment as the next best step in patient management. The recommended increase in calcium and vitamin D would be insufficient to treat vitamin D deficiency–related osteomalacia in this woman, and bisphosphonate and teriparatide therapy would be inappropriate until the vitamin D deficiency is corrected. Finally, bone turnover markers cannot be used independently for diagnosis in patients with metabolic bone disease.

3. An 81-year-old man is admitted after a fall and low-impact fracture of the proximal femur. His history is notable for Parkinson's disease treated with dopamine agonist therapy. On physical examination, patient has a mild resting tremor while lying in bed. Laboratory investigations do not reveal secondary causes of osteoporosis, including a 25(OH)D level of 30 ng/mL and testosterone of 300 ng/dL. The patient has normal renal function and calcium. The patient undergoes successful operative repair. After discharge, the patient undergoes 4 weeks of inpatient rehabilitation. What is the most definitive choice for this patient's metabolic bone management?

A. Start calcium and vitamin D supplementation.
B. Increase dopamine agonist therapy for Parkinson's disease.
C. Give zoledronic acid, 5 mg intravenously.
D. Start testosterone replacement therapy.
E. Prescribe ergocalciferol 50,000 IU once weekly.

Answer: C This gentleman has incurred a low trauma femur fracture, which greatly increases his risk for subsequent fractures. Furthermore, the fracture confers a mortality rate up to 30% within 1 year of the event. Zoledronic acid has been proved not only to reduce the risk for subsequent fracture but also to reduce mortality by 27% Calcium and vitamin D are important adjuncts to treatment but not definitive. Given his level of 25(OH)D, additional antifracture and/or fall risk reduction would not be anticipated from pharmacologic vitamin D treatment. There is no evidence that testosterone treatment of either eugonadal or hypogonadal men reduces fracture risk. Finally, optimal management of his Parkinson's disease, which does not appear undertreated in this patient, might reduce his risk for falls but not his risk for fracture.

4. A 55-year-old white woman presents to her physician with intense midback pain after a fall onto her backside while walking. Her history is notable for a recent diagnosis of polymyalgia rheumatica, for which she has taken prednisone 10 mg daily for the last 6 months. She does have known osteopenia, with lumbar spine, femoral neck, and total hip T scores of −1.5, −1.2, and −0.8, respectively. Physical examination is notable for tenderness to palpation and percussion over the lower thoracic spine. Plain films of the thoracolumbar spine reveal a new moderate (35%) anterior compression fracture at T11. Laboratory work-up, including 25(OH), is within normal limits. Other than continuing the calcium and vitamin D that the patient is currently taking, what is the most appropriate pharmacologic choice for treatment at this time?

A. Vitamin D 50,000 IU once weekly
B. Alendronate 70 mg once weekly
C. Risedronate 150 mg once monthly
D. Teriparatide 20 mcg subcutaneously daily
E. Denosumab 60 mg subcutaneously every 6 months

Answer: D This woman presents with an acute vertebral fracture following a low trauma event. Her level of BMD is not osteoporotic, underscoring the independent contribution of glucocorticoids to fracture risk such that fractures may occur at a higher (i.e., better) level of BMD. The recent nature of her fracture further underscores the need for initiation of treatment now. There is no evidence that increasing this patient's vitamin D level will further reduce her risk for fracture. Denosumab will likely reduce her risk for subsequent fracture, although the association with infection and use of concurrent prednisone in this woman preclude its consideration at present. Alendronate and risedronate are approved for the prevention and treatment of glucocorticoid-induced osteoporosis (GIO). Despite this, teriparatide has been proved more effective that bisphosphonates in patients with GIO and is indicated for this woman, assuming she has no contraindications to treatment.

5. A 70-year-old man presents with acute, mid-back pain after lifting a bag of topsoil while gardening 1 week ago. His medical history is notable for known male hypogonadism due to mumps orchitis in his 20s, although he does not take testosterone because of severe benign prostatic hypertrophy. He does take 600 mg of calcium twice daily and 1000 IU of vitamin D once daily. On physical examination, he has tenderness to palpation and percussion over his mid-thoracic spine. Laboratory investigations are unrevealing, including a 25(OH) level of 45 ng/mL. DXA bone density study confirms osteoporosis with lumbar spine T score of −3.0, although femoral neck and total hip T scores are normal. Radiographs of the thoracic spine reveal a new severe (50%) biconcave compression fracture at T8 without apparent widening of the pedicles. In addition to prescribing alendronate, which of the following is best adjunct management for this patient?

A. Analgesic therapy and referral to physical therapy for post-fracture consultation
B. Prescription for a spinal brace that patient should wear fro 6 months
C. Percutaneous vertebroplasty
D. Nasal calcitonin use for 1 year
E. No additional interventions are required at this time.

Answer: A This gentleman has osteoporosis secondary to long-standing hypogonadism, exhibited by predominantly great bone density loss and deficit in cancellous bone (i.e., spine). He has incurred by definition a low-trauma fracture and is in need of a pharmacologic antifracture therapy (i.e., alendronate). Although he has significant pain that must be addressed, there is no evidence that conservative therapy with analgesics is inferior to vertebroplasty in regard to pain management. In addition, his radiographic findings do not suggest fracture instability that might benefit from vertebroplasty. Physical therapy is also helpful both short term as an adjunct to pain management and long term regarding modification of lifestyle and exercise to maximally reduce future fracture risk. A supportive spine brace may be used, although its use should be limited to only 4 to 6 weeks because of the necessary induction of paraspinal muscle weakness associated with long-term use. Similarly, nasal calcitonin may have some analgesic benefit in the acute, post-fracture period based on limited data, although there is no evidence of a long-term analgesic benefit with continued use of the drug.

TABLE 244-1 CAUSES OF OSTEOMALACIA

VITAMIN D DEFICIENCY
Dietary deprivation and lack of sunlight exposure
VITAMIN D MALABSORPTION
Postgastrectomy Gastric bypass for obesity Gluten enteropathy Small bowel disease or resection Pancreatic insufficiency Cholestyramine therapy for cholestatic liver disease Laxative abuse
IMPAIRED 1-HYDROXYLATION OF 25-HYDROXYVITAMIN D
Vitamin D–dependent rickets type I X-linked hypophosphatemia Autosomal dominant hypophosphatemic rickets/osteomalacia Oncogenic osteomalacia
IMPAIRED TARGET-ORGAN RESPONSE TO 1,25-DIHYDROXYVITAMIN D
Vitamin D–dependent rickets type II
HYPOPHOSPHATEMIA
X-linked hypophosphatemia Autosomal dominant hypophosphatemic rickets/osteomalacia Sporadic hypophosphatemia Fibrous dysplasia Oncogenic osteomalacia Antacid-induced osteomalacia Chronic metabolic acidosis Paraproteinemia Saccharated ferric oxide Tenofovir Cadmium
INHIBITORS OF MINERALIZATION
Etidronate Fluoride Aluminum Iron
MISCELLANEOUS
Hypophosphatasia Axial osteomalacia Fibrogenesis imperfecta ossium

244

OSTEOMALACIA AND RICKETS

ROBERT S. WEINSTEIN

DEFINITION

Rickets refers to impaired mineralization of the cartilaginous growth plate and abnormal endochondral bone formation and therefore cannot occur in adults after epiphyseal closure.[1] Osteomalacia, literally meaning softening of bone, refers to defective or delayed mineralization of the organic matrix of bone, or osteoid, at the interface between calcified bone and osteoid, and may occur at any age. Both rickets and osteomalacia may be present in a growing child, but defective mineralization can cause only osteomalacia in adults; therefore, this chapter will focus on osteomalacia. Despite advances in our understanding of vitamin D metabolism and the increased sensitivity of measurements of serum 25-hydroxyvitamin D, osteomalacia remains a common and frequently overlooked disorder in the world. Optimal therapy requires precise identification of the etiology of the abnormal mineralization, which may present a problem because there are numerous causes (Table 244-1). However, after a correct diagnosis is made, therapy is usually gratifying and often spectacular. Early recognition of osteomalacia depends on familiarity with the typical clinical manifestations and settings. It is helpful to appreciate that the bone disease almost always manifests in the same manner regardless of the cause of the osteomalacia.

EPIDEMIOLOGY

About 20% of North American women receiving treatment for osteoporosis have 25-hydroxyvitamin D levels below 20 ng/mL (adequate values are greater than 30 ng/mL), and 8% have levels below 15 ng/mL. This indicates that, at the least, impaired bone mineralization could be a confounding factor in their osteoporosis treatment and, at worst, osteomalacia is the correct diagnosis (a defect in mineralization) rather than osteoporosis (a reduced amount of normally mineralized bone). An inadequate response to the bisphosphonate treatment commonly used for postmenopausal osteoporosis is four times more likely when 25-hydroxyvitamin D levels are subnormal than when the levels are above 30 ng/mL.[2] Vitamin D deficiency is more common in elderly people, especially in nonaffluent people during the winter, at higher latitudes, and with low sun exposure. Vitamin D deficiency is also commonly found in medical inpatients, institutionalized patients, and postmenopausal women with acute hip fracture.[3]

The prevalence of osteomalacia due to vitamin D deficiency varies with the referral source. The disorder is far more frequent when patients are referred from geriatricians, gastroenterologists (osteomalacia may be found in up to 30% of patients with gastric surgery or bypass for obesity), nursing homes, or orthopedists concerned about symmetrical or nonhealing fractures. The most common hypophosphatemic osteomalacia is the inherited disease X-linked hypophosphatemia (XLH), but affected adults infrequently present to internists and then only when troubled by severe bone pain or nonunion of fractures.

PATHOBIOLOGY

A review of normal bone remodeling and the mineralization of osteoid (bone matrix) serves as a background to understand the abnormal mineralization characteristic of osteomalacia. Bone remodeling or turnover is carried out by teams of juxtaposed osteoclasts and osteoblasts, comprising temporary anatomical structures known as basic multicellular units (BMUs). In cortical bone, the BMUs drill tunnels or "cutting cones" through the compact tissue; whereas in spongy, cancellous bone, they gouge across the trabecular surface, forming serpiginous trenches. Bone turnover begins by conversion of a quiescent skeletal surface to a remodeling site, a process referred to as *activation.* Activation involves proliferation of new blood vessels needed to bring recruited osteoclast progenitors to the remodeling site and retraction of the flat, pavement-like bone-lining cells that cover the quiescent surfaces to expose the mineralized bone surface. The recruited cells become multinucleated osteoclasts, which attach to the newly exposed bone surface with a ring of contractile proteins sealing off a subosteoclastic resorption compartment. Lysosomal enzymes, hydrogen ions, and collagenase are secreted through the microvilli of the ruffled underside border of the osteoclasts, and these chemicals begin to excavate a resorption cavity. The osteoclasts remove both the bone mineral and matrix. It is a misunderstanding to attribute to these cells or to metabolic acidosis the ability to remove only the mineral, leaving behind demineralized osteoid. Demineralized bone in vivo is a misnomer. Demineralized or decalcified bone only occurs when bones are placed in acid (1N HCl) or chelating solutions (ethylenediaminetetraacetate, or EDTA). Osteoclasts are motile cells, capable of resorbing more than just the cavity within which they are identified. After an osteoclast digs a cavity, it may detach from bone and move on to a new resorption site or die by apoptosis and be quickly removed by phagocytes. When the osteoclasts have moved on, osteoblasts assemble to reconstitute the previously resorbed cavity with new bone. In any established BMU, both events are happening at the same time; bone formation begins to occur while bone resorption advances.

Between the end of bone resorption and the beginning of bone formation is the reversal phase, when mononuclear phagocytes smooth out the jagged erosion bays. During this phase, the old bone is coated by a thin layer of cement substance, a collagen- and mineral-poor matrix rich in glycosaminoglycans, glycoproteins, and acid phosphatase, to which the new osteoblasts attach. In adults, new osteoblasts assemble only at sites where osteoclasts have recently been eroding bone; a phenomenon referred to as *coupling.* The arrival of the osteoblasts in the right place at the right time and in sufficient numbers to reconstitute the cavity is referred to as *remodeling balance* and is likely due to proportional production of osteoblasts and osteoclasts in the bone marrow, release of osteoblast-recruiting substances from the resorbed bone, and chemotaxis by the cement substances. As osteoblasts complete their bone matrix synthesis and move away from the cement line, they gradually flatten. Some osteoblasts become bone-lining cells, and some become osteocytes, but as many as 65% of the osteoblasts that originally assembled at the remodeling site die by apoptosis. It is the balance between cell proliferation and apoptosis that determines the amount of work performed by these cells.

Normally, up to 70% of the mineralization of the osteoid deposited by the osteoblasts starts within 4 to 12 days and proceeds at about 1 µm per day; but in osteomalacia, mineral deposition in the osteoid slows or stops completely, while the osteoblasts continue to make osteoid, which then accumulates in excessive amounts. Therefore, normal osteoid width is about 4 to 12 µm, but in osteomalacia, the osteoid width may become dramatically augmented. Depending on the extent of the delay in mineralization, overt osteomalacia may take many years to develop. In normal subjects, further mineralization proceeds slowly over months to years and at the cost of displacement of the water in the hydroxyapatite crystals, resulting in a modest increase in brittleness and the eventual need for another round of remodeling. Even though 1 million BMUs are undergoing remodeling every day, bone mass in a healthy adult is preserved thanks to a remarkably tight balance between the amount of bone resorbed and the amount formed during each cycle of remodeling. By this means, the adult skeleton is almost completely regenerated every 10 years.

Mineralization requires the availability of sufficient calcium and phosphorus at the remodeling site, the presence of a normal bone collagen matrix, the absence of inhibitors of mineralization, and an adequate amount of skeletal alkaline phosphatase activity. Defects in these requirements are the cause of most forms of osteomalacia. Deficiency of vitamin D per se has traditionally been incriminated as the cause of the osteomalacia, but today, considerable evidence indicates that the abnormal mineralization associated with vitamin D deficiency depends more on the deficiency of calcium and phosphorus than the absence of a direct effect of vitamin D on bone cells. The primary function of vitamin D is to provide adequate levels of calcium and phosphorus by increasing their intestinal absorption. Chronic metabolic acidosis has also been identified as a cause of osteomalacia, but evidence suggests that the bone disease associated with chronic metabolic acidosis is primarily due to the associated hypophosphatemia.

CLINICAL MANIFESTATIONS

The clinical presentation of osteomalacia depends on three overlapping manifestations: those due to the underlying disorder, such as gastrointestinal disease or surgery (especially troublesome are gastric resection, stapling or bypass for obesity, and intestinal malabsorption); those due to hypocalcemia or hypophosphatemia; and those directly due to the bone disease. The most common symptoms and signs are bone pain, muscle weakness, and bone tenderness. The bone pain is usually nonspecific and poorly localized. Because of the paucity of findings, the pain is often attributed to rheumatism or neurosis. It may be worse at night and after sudden movements such as turning in bed or the change from sitting to standing. Most often, the pain is in the lower back, pelvis and legs and is worse on weight bearing, resulting in a characteristic flat-footed, springless, waddling gait made worse by proximal muscle weakness. The gait has been referred to as "mother penguin's walk." Patients may complain that they can only climb stairs by pulling themselves up with the hand rail or rise from sitting in a chair or on the toilet by using their hands to push off. The decrease in strength is usually far greater than the degree of muscle wasting. Fasciculations are absent, and both reflexes and sensation remain normal. The bulbar, facial, and ocular muscles are always spared. However, muscle weakness is conspicuously absent when the osteomalacia is due to X-linked hypophosphatemia (see Table 244-1). Often, bone tenderness can be elicited by rib cage compression or pressing on the tibiae, wrists, pubic rami, or iliac crests. Hypocalcemia is usually mild to moderate but, rarely, can be severe enough to present with paresthesias, muscle cramps, a positive Chvostek's sign, or seizures. If the osteomalacia is mistaken for osteoporosis and treatment is started with a bisphosphonate, the patient may experience new-onset paresthesias, muscle cramps, and palpitations. This not uncommon scenario occurs because the antiresorptive treatment interferes with the compensatory secondary hyperparathyroidism and aggravates the hypocalcemia.

DIAGNOSIS

Biochemical changes depend on the stage of the disease and its etiology. In vitamin D deficiency, hypophosphatemia precedes and is more severe than the hypocalcemia because of the secondary or compensatory hyperparathyroidism (Chapter 245) that almost invariably accompanies the disorder by the time that osteomalacia has occurred. In malabsorption, hypomagnesemia may contribute to the hypocalcemia, and hypoalbuminemia may lead to a spurious diagnosis of hypocalcemia. Increased serum alkaline phosphatase activity is classically associated with osteomalacia due to vitamin D deficiency but is not an early or reliable clue because some patients may have normal or borderline levels. The serum 25-hydroxyvitamin D levels are often less than 10 to 15 ng/mL. In contrast, serum 1,25-dihydroxyvitamin D levels are usually elevated because of the concomitant hyperparathyroidism and do not contribute to the diagnosis of osteomalacia except in the rare abnormalities of vitamin D resistance (when 1,25-dihydroxyvitamin D levels may be extraordinarily high) or when 1-hydroxylation is defective (and 1,25-dihydroxyvitamin D levels are low). Quite a different pattern occurs with the inherited disease hypophosphatasia: serum 25-hydroxyvitamin D and calcium are normal, phosphorus and vitamin B_6 levels are high normal or frankly elevated, and alkaline phosphatase activity is below the normal range.[4]

Radiographic findings may be absent with early osteomalacia, and only blurred margins of the cancellous bone with thin cortices may be noted. The presence of bilateral, thin (2 to 3 mm), radiolucent bands known as *pseudofractures* (Fig. 244-1) found perpendicular to the periosteal surface in ribs, pubic and ischial rami, the neck of the femur, and metatarsals and below the glenoid fossa on the outer border of the scapulae are generally considered to be pathognomonic of osteomalacia, but this classical radiographic sign is infrequent today. Rarely, it may be seen in disorders lacking excessive osteoid. These pseudofractures (sometimes called *Looser's zones* or *Milkman's fractures*) show increased uptake on bone scans (Fig. 244-2) and may lead to an inappropriate search for a primary malignancy. Bone mineral density T scores are often −3 or −4, with the radial diaphyseal density lower than that of the lumbar spine or total proximal femur.

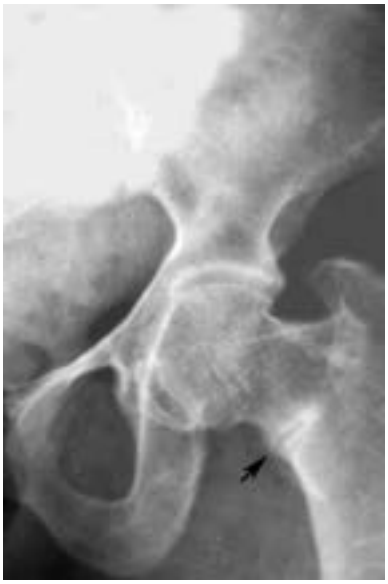

FIGURE 244-1. Radiographic evidence of a pseudofracture of the femoral neck is suspicious for osteomalacia *(arrow)*.

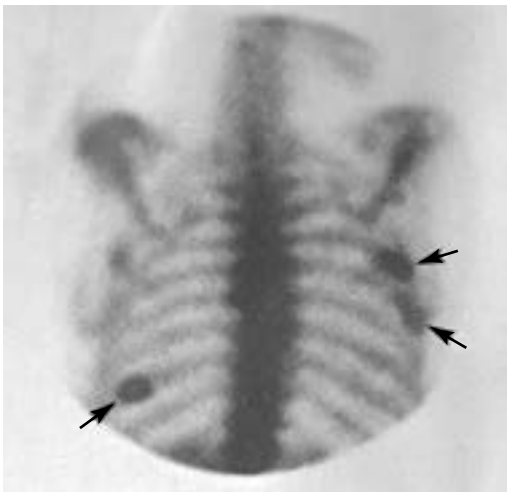

FIGURE 244-2. In osteomalacia, focal increased uptake of radionuclide on a bone scan may erroneously suggest metastatic disease *(arrows)*.

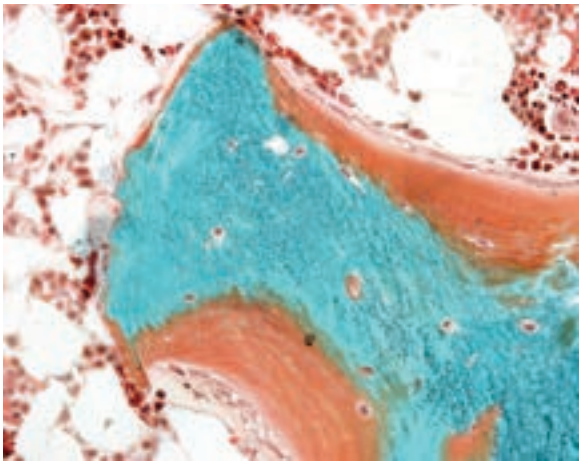

FIGURE 244-3. An undecalcified bone biopsy specimen shows the characteristic abundant osteoid and flattened osteoblasts of osteomalacia (normally mineralized bone is *blue* and osteoid is *red*).

Although characteristic clinical, radiographic, and biochemical findings may suggest osteomalacia, the absence of these findings cannot exclude the diagnosis. Quantitative histologic examination of undecalcified bone is, therefore, required to establish the unequivocal presence of osteomalacia (Fig. 244-3). Rigorous kinetic criteria for the histologic recognition of osteomalacia are necessary to preserve the traditional clinical, biochemical, and therapeutic connotations of the term. Therefore, a review of the quantitative bone histologic findings or histomorphometry in osteomalacia is useful.[5] The histomorphometric diagnosis of osteomalacia requires the simultaneous presence of three findings: (1) excessive osteoid (osteoid area >10% of the cancellous bone area; normal is <4%), (2) augmentation of the osteoid width (>15 μm; normal is 4 to 12 μm), (3) and prolongation of the mineralization lag time (>100 days; normal is 9 to 20 days), as determined by the osteoid width divided by the distance between and linear extent of double tetracycline labels observed in the bone after the patient receives two time-spaced courses of oral tetracycline. Tetracycline is deposited early in the course of hydroxyapatite crystal formation and generates bright stripes at the interface of mineralized bone and osteoid when viewed with fluorescent microscopy. If the two time-spaced courses of tetracycline (1 g/day for 3 days) are separated by a 14-day interval, the rate of mineralization (μm/day) can be calculated by measuring the average distance between the double labels divided by the number of days between the two courses. When the double labels are numerous and widely spaced, mineralization is intact and excess osteoid must be due to increased bone turnover. A paucity of tetracycline labels that are narrowly spaced indicates that if excessive osteoid is present, it must be due to the delayed or ceased mineralization of osteomalacia (Fig. 244-4).

Therefore, it follows that excessive osteoid can occur from two distinct mechanisms. Osteomalacia is the consequence of defective mineralization, while osteoid production continues. However, osteoid will also accumulate with accelerated bone formation if the rate of osteoid deposition exceeds the rate of mineralization, as occurs in states of greatly increased bone turnover, such as hyperparathyroidism (Chapter 245), Paget's disease (Chapter 247), or thyrotoxicosis (Chapter 226). Even though osteoblasts in osteomalacia are usually sparse and flattened, whereas they are numerous, plump, and cuboidal with high bone turnover, these two groups of disorders can only be reliably distinguished with the use of tetracycline markers. The treatment of increased bone turnover and of defective mineralization is completely different, which is why the three histomorphometric criteria are necessary. Additionally, evaluation of each of the criteria in isolation has limitations. Regarding the first requirement, a small increase in the osteoid area relative to the total bone area may occur in osteoporosis, with a decrease in the amount of mineralized bone. In the second requirement, wide osteoid seams may be seen in some specimens obtained from patients with severe secondary hyperparathyroidism, such as those on maintenance hemodialysis therapy (Chapter 131). In the third requirement, reduced mineral appositional rate and increased mineralization lag time are nonspecific indices of impaired matrix synthesis by osteoblasts, as is often found in patients with involutional osteoporosis. Only when all three requirements are fulfilled is the diagnosis of osteomalacia irrefutable.

Several presumed causes of osteomalacia (anticonvulsant drugs, metabolic acidosis without hypophosphatemia, pseudohypoparathyroidism, and chronic renal failure) have not fulfilled all of these requirements and primarily represent secondary hyperparathyroidism. Patients with the nephrotic syndrome lose albumin and vitamin D metabolites in the urine, but evidence indicates that serum ionized calcium and parathyroid hormone levels are normal and metabolic bone disease in adults with the nephrotic syndrome is absent. Muscle weakness and bone pain are significantly more common in patients in whom the rigorous histologic diagnosis of osteomalacia has been proved. However, bone biopsy is not always necessary to be reasonably certain of the diagnosis. When biopsy is necessary, the local pathologist must be familiar with the processing of undecalcified bone specimens and plastic embedding; otherwise, the best solution is to refer the patient to a histomorphometry center for biopsy. This ensures satisfactory communication between the clinician, operator, and pathologist and is the best insurance against the incomplete, broken, fragmented, or accidentally decalcified bone specimens. Such referral may be indispensable in the evaluation of a patient with unusually painful disease or progressive loss of bone mineral density, particularly when the results of the physical examination, radiographs, and biochemical findings are ambiguous. Biopsy may also be indicated in patients with unexplained chronic hypophosphatemia.

The best approach is to avoid overlooking the diagnosis of osteomalacia by maintaining a high degree of suspicion in the typical clinical settings.[6] This is especially important because osteomalacia can usually be successfully

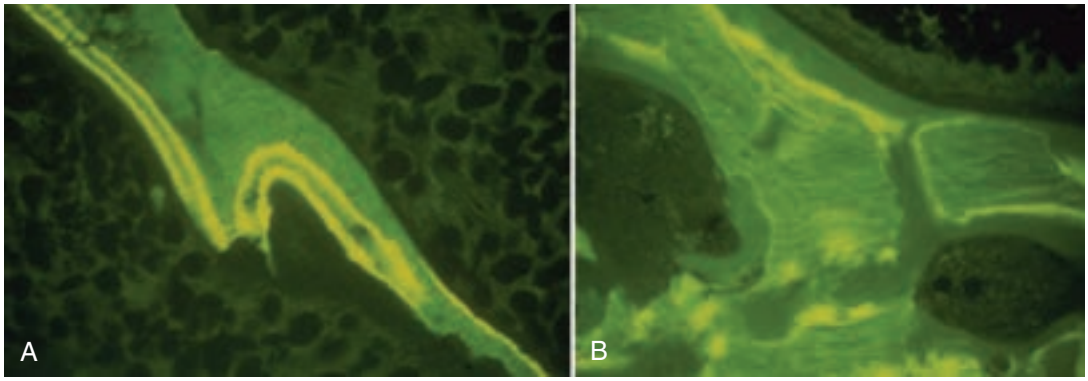

FIGURE 244-4. Histomorphometric diagnosis of osteomalacia by fluorescence imaging of double tetracycline labeling. **A,** Tetracycline double labels are numerous, discrete, and widely spaced, as is typical of intact mineralization. **B,** The tetracycline labels are mostly single despite the administration of two time-spaced doses of oral tetracycline, indicating that mineralization must be delayed or ceased, as is typical of osteomalacia.

treated. An investigation for osteomalacia is indicated in elderly patients with bone pain and muscle weakness, in patients with gastric surgery and low bone mineral density or bone pain, and in patients with persistent hypophosphatemia. Unexplained elevations of the serum alkaline phosphatase activity are usually due to drugs (e.g., anticonvulsants, anabolic steroids, phenothiazines, or antibiotics) or Paget's disease of bone (Chapter 247) but rarely may be the only biochemical clue to osteomalacia in a patient with variable skeletal discomfort. Bilateral or slowly healing fractures also warrant an investigation for osteomalacia.

TREATMENT

Rx

Understanding of the treatment of osteomalacia is facilitated by dividing the disease into four subgroups. The *first subgroup* is osteomalacia due to disorders of vitamin D absorption or metabolism; the *second* is osteomalacia due to chronic hypophosphatemia. Most patients with osteomalacia will be in these first two subgroups. Treatment of osteomalacia caused by these two subgroups is discussed in detail in the next two sections. The *third subgroup* includes osteomalacia caused by inhibitors of mineralization, such as etidronate (the first oral bisphosphonate, now rarely used in North America); high doses of fluoride; accumulation of a skeletal burden of aluminum from water used for dialysis or as a contaminate in solutions used for parenteral nutrition (now rarely seen); iron overload as in thalassemia; and cadmium, which induces the proximal tubular lesion of Fanconi's syndrome and causes osteomalacia due to the resultant hypophosphatemia. The *fourth subgroup* includes miscellaneous causes of osteomalacia that lack specific therapy but are fortunately quite rare. This last subgroup includes the variable forms of the heritable disorder hypophosphatasia, caused by a deficiency of the tissue-nonspecific (liver, bone, kidney) isoenzyme of alkaline phosphatase (although, therapeutic trials have shown that enzyme replacement is effective); axial osteomalacia, a sporadic osteosclerotic disorder primarily affecting middle-aged men and presenting with mild to moderate pain in the spine and pelvis (but without fractures), apparently due to the production of an abnormal and poorly mineralized bone matrix by osteoblasts; and fibrogenesis imperfecta ossium, another sporadic disorder presenting with intractable bone pain and fractures, mainly in middle-aged men and women and apparently also due to production of an abnormal bone matrix lacking the normal collagen birefringence by osteoblasts. In axial osteomalacia and fibrogenesis imperfecta ossium, serum calcium, phosphorus, and vitamin D levels are normal, but serum alkaline phosphates activity may be increased. General measures for this last subgroup include routine nutritional advice and avoidance of further bone loss due to postmenopausal or involutional osteoporosis. High-dose vitamin D therapy in these disorders has caused nephrocalcinosis, nephrolithiasis, and renal insufficiency and must be avoided.

Osteomalacia Due to Vitamin D Disorders

Iron deficiency anemia, hypocalcemia, weight loss, glossitis, or pruritic rash and bone discomfort in a patient with low bone mineral density point to celiac disease even without gastrointestinal symptoms. These signs suggest the need to test for immunoglobulin A antiendomysial and antitissue transglutaminase antibodies. Cholestyramine therapy for cholestatic liver disease may increase malabsorption of vitamin D by binding bile salts. Laxative abuse may cause osteomalacia and severe resistance to vitamin D supplementation, including treatment with calcitriol. Advice on nutrition and sun exposure, discontinuation of offending drugs, adherence to a gluten-free diet, and pancreatic enzyme replacement may cure the mineralization defect in mild cases without the need for additional treatment.

Patients with severe disease will usually require vitamin D and calcium supplementation. Replacement doses depend on the serum 25-hydroxyvitamin D level, as shown in Table 244-2. Because pharmacologic doses of any vitamin D preparation carry the risk for vitamin D intoxication, increases in the dose must be made carefully. The interval between increments in dosage should be at least the time required to reach maximal effects plus about 50%. However, experience with the doses given in Table 244-2 indicates that serum 25-hydroxyvitamin D levels rarely reach 80 to 100 ng/mL. Vitamin D intoxication is unlikely even with of levels of 200 to 250 ng/mL. The goal is to raise the serum 25-hydoxyvitamin D level well above 30 ng/mL and restore the elevated parathyroid hormone concentration to normal without hypercalcemia or hypercalciuria. Urinary calcium excretion should be monitored when treatment has normalized the serum calcium level. The urinary calcium-to-creatinine ratio (mg/mg) should be kept below 0.22. Approximately 1 to 1.5 g/day of oral elemental calcium is a reasonable initial dose. Frequent small doses (three times a day) are more effective and tolerable than fewer larger ones, and the absorbability of calcium supplements is enhanced with meals. Most patient do well using the calcium preparations used for osteoporosis, such as calcium carbonate (40% calcium as in Os-Cal or the equivalent) or calcium citrate (21% calcium as in Caltrate or the equivalent). Some patients who cannot tolerate calcium carbonate or citrate experience fewer adverse gastrointestinal symptoms with the use of the chocolate or coffee-flavored formulations known as Viactiv (500 mg calcium per tablet). The vitamin D content of these calcium supplements is trivial in the treatment of osteomalacia.

TABLE 244-2 VITAMIN D PREPARATIONS FOR TREATMENT OF OSTEOMALACIA

	VITAMIN D_3 (CHOLECALCIFEROL)	CALCITRIOL ($1,25[OH]_2D_3$)
Trade names	Calciferol, BIOTECH†	Calcitriol
Dosage form	Caps: 50,000 units = 1.25 mg	Caps: 0.25 and 0.50 µg
Dosage:		0.50-2.0 µg/day
If serum 25(OH)D = 20 to 30 ng/mL	50,000 units once a week × 10 weeks* and once a month thereafter	
If serum 25(OH)D = 10 to 20 ng/mL	50,000 units twice a week × 10 weeks* and twice a month thereafter	
If serum 25(OH)D = less than 10 ng/mL	50,000 units three times a week × 10 weeks* and three times a month thereafter	
Dosage in resistant cases	Up to 50,000 units per day	5-20 µg/day
Time to reach maximum effects	4-10 weeks	3-7 days
Persistence of effects after cessation	6-30 weeks	3-7 days
Cost in U.S.	$40/100 capsules of 50,000 units	$130/100 capsules of 0.25 µg $150/100 capsules of 0.5 µg

*If not >30 ng/mL after 10 weeks, exclude malabsorption, gluten enteropathy, and noncompliance. High-quality cholecalciferol, free from gluten, dairy, egg, fish nuts, soy, or artificial colors can be obtained from †BIOTECH at 1-800-345-1199. Weekly tanning bed treatments may be used if oral vitamin therapy fails or a switch to the more costly calcitriol may be necessary.
25(OH)D = 25-hydroxyvitamin D.

In patients with malabsorption, vitamin D requirements may increase during periods of increased diarrhea, and calcitriol may be easier for these patients to absorb. Its rapid onset of action and disappearance after cessation add to the safety of treatment, albeit at far greater cost. Calcium, phosphorus, potassium, magnesium, multivitamins, and gonadal steroids may also be beneficial in patients with malabsorption. Some patients do not tolerate any form of oral vitamin D, and the parenteral ergocalciferol preparations in North America are ineffective. These patients can be improved, although not restored to normal, by the use of weekly tanning bed treatments to areas of their bodies not normally exposed to the sun, an attempt to minimize solar-induced skin cancer. Calcitriol is the drug of choice in patients with the autosomal recessive disease, vitamin D–dependent rickets type I, in which the 1α-hydoxylase enzyme necessary to convert 25-hydroxyvitamin D to 1,25-dihydroxyvitamin D is deficient. In vitamin D–dependent rickets type II, another rare autosomal recessive disease presenting with alopecia, diminished target sensitivity to 1,25-dihydroxyvitamin D may require extraordinarily high doses of calcitriol. If oral treatment fails, nocturnal infusions of calcium and phosphorus have been successful, providing additional evidence that the osteomalacia is due to inadequate calcium and phosphorus rather than the defect in vitamin D metabolism.

An increase in the serum alkaline phosphatase activity (the healing "flare") and a small increase in the serum and urine calcium levels are the earliest signs of effective treatment. Thereafter, the serum alkaline phosphatase activity level falls progressively as healing occurs. At the start of therapy, serum calcium levels should be measured at weekly intervals. If hypoalbuminemia is present, serum ionized calcium determinations are more useful. When therapy appears stabilized, biweekly or monthly intervals are usually sufficient for the first 3 or 4 months, but even with long-term therapy, measurements should be at least two to three times a year. In some patients with severe osteomalacia, bone pain and paresthesias may increase and the serum calcium levels decrease during the first few weeks of therapy. This is due to the increased skeletal avidity for mineral during healing and indicates the need for additional calcium supplementation.

Osteomalacia Due to Hypophosphatemia

Therapy of chronic hypophosphatemia is aimed at maintaining normal concentrations of serum phosphorus without inducing secondary hyperparathyroidism or nephrocalcinosis. This considerably difficult task requires divided doses of phosphorus supplements (1 to 3 g/day) and calcitriol (1 to 4 μg/day) to increase the absorption of phosphorus and try to prevent the increase in parathyroid hormone (Tables 244-2 and 244-3). If phosphorus-induced secondary hyperparathyroidism develops, the phosphorus supplements are rapidly excreted, and therapy thus is not only futile but also causes the additional bone disease of hyperparathyroidism. Baseline and yearly renal ultrasound examinations are necessary to recognize early nephrocalcinosis or nephrolithiasis.

X-linked hypophosphatemia (XLH) is the most common cause of chronic hypophosphatemia, and the presence of a positive family history, pediatric onset, and bowed legs usually substantiates the diagnosis. Treatment with an anti-FG23 antibody can raise serum phosphorus and 1,25 vitamin D levels,[A1] although the long-term benefits are not yet known. However, some hypophosphatemic patients have an autosomal dominant family history and present in adulthood with osteomalacia but without lower extremity deformities. Like patients affected with XLH, these patients may present with bone pain, pseudofractures, and high-normal or frankly elevated levels of fibroblast growth factor 23 (FGF23), a phosphaturic protein that interferes with 1-hydroxylation of 25-hydroxyvitamin D.[7] Patients with autosomal dominant hypophosphatemic osteomalacia (ADHR) appear to acquire the renal phosphate losses in adolescence or adulthood, whereas other patients with ADHR may lose the defect as they age. A diagnostic problem arises when a patient with chronic hypophosphatemia presents without a positive family history or bowed legs because this presentation resembles that of patients with oncogenic osteomalacia. This disorder is associated with a variety of small, hard to find, benign mesenchymal tumors that secrete FGF23.[8,9] The muscle pain, weakness, fractures, and osteomalacia characteristic of this syndrome are due to hypophosphatemia made worse by inappropriately low levels of 1,25-dihydroxy vitamin D. Oncogenic osteomalacia is also treated with phosphorus supplementation and calcitriol until the offending tumor can be located and resected. Improved tumor localization has been reported with positron emission tomography and computed tomography.[9-11] Surgical correction of deformities should be postponed until medical management achieves persistently normal levels of calcium, phosphorus, and alkaline phosphatase activity. An exception to this rule is an acute fracture of the femoral neck. Prompt surgical repair may be essential to avoid osteonecrosis. Oncogenic osteomalacia rarely occurs with malignant tumors that secrete FGF23, but unless surgical resection is complete, the osteomalacia will persist. Recent evidence suggests that tumors in surgically difficult locations may be treated by radiofrequency ablation.[12] Antacid-induced hypophosphatemia due to the ingestion of large quantities of phosphate-binding antacids has become rare with the increased availability of proton pump inhibitors but still occurs occasionally.

TABLE 244-3 PHOSPHATE PREPARATIONS FOR TREATMENT OF OSTEOMALACIA

PREPARATION	TABLET MARKINGS AND SHAPE	SODIUM CONTENT (mEq)	POTASSIUM CONTENT (mEq)	AMOUNT THAT CONTAINS 1 GRAM OF ELEMENTAL PHOSPHORUS
Neutra-Phos	0	28.5	28.5	4 unit dose caps*
Neutra-Phos-K	0	0	57.0	4 unit dose caps*
K-Phos Neutral	"Beach 11-25" oblong	50.4	4.6	4 tabs
K-Phos Original	"Beach 1111" round	0	33.0	9 tabs

*Each unit dose cap is reconstituted with 75 mL of water, fruit juice, or cola, and this formulation is preferred by children. The unit dose cap contains the powder concentrate and is not to be swallowed undiluted. Adults prefer the K-Phos Neutral tablets.

PREVENTION

Advice about vitamin D supplementation should help to prevent osteomalacia caused by vitamin D deficiency, but this has proved to be difficult because routine supplements may be inadequate and compliance with nutritional supplements is poor. The optimal vitamin D supplementation dosage is not clear, but most bone and mineral problems are avoided by 50,000 units of ergocalciferol given once monthly. Notable exceptions occur in patients with celiac disease, gastric surgery, or bypass for obesity, who often require much larger amounts (see Table 244-2). In patients with osteomalacia due to hypophosphatemia, the need for phosphorus supplementation may be lifelong. Rare exceptions occur in oncogenic osteomalacia if complete surgical removal or destruction of the tumor is accomplished.

PROGNOSIS

The response to appropriate treatment in most forms of osteomalacia is usually excellent. Improvements in bone pain and muscle weakness usually occur within 2 or 3 months, and healing of skeletal lesions occurs within 6 to 18 months. Depending on the quantity of excess osteoid, repeat bone mineral density determinations may show as much as 20% gains at the lumbar spine and total proximal femur. However, bone density at the radial diaphysis may not improve because of the irreversible loss of cortical bone resulting from prolonged secondary hyperparathyroidism. Furthermore, if decreased bone volume is present in addition to excess osteoid, skeletal recovery may be incomplete, and the risk for fractures may remain increased.

Grade A Reference

A1. Carpenter TO, Imel EA, Ruppe MD, et al. Randomized trial of the anti-FGF23 antibody KRN23 in X-linked hypophosphatemia. *J Clin Invest*. 2014;124:1587-1597.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Elder CJ, Bishop NJ. Rickets. *Lancet*. 2014;383:1665-1676.
2. Peris P, Martinez-Ferrer A, Monegal A, et al. 25 Hydroxyvitamin D serum levels influence adequate response to bisphosphonate treatment in postmenopausal osteoporosis. *Bone*. 2012;51: 54-58.
3. van Schoor NM, Lips P. Worldwide vitamin D status. *Best Pract Res Clin Endocrinol Metab*. 2011;25:671-680.
4. Berkseth KE, Tebben PJ, Drake MT. Clinical spectrum of hypophosphatasia diagnosed in adults. *Bone*. 2013;54:21-27.
5. Kulak CA, Dempster DW. Bone histomorphometry: a concise review for endocrinologists and clinicians. *Arq Bras Endoocrinol Metab*. 2010;54:87-98.
6. Priemel M, von Domarus C, Klatte TO, et al. Bone mineralization defects and vitamin D deficiency: histomorphometric analysis of iliac crest bone biopsies and circulating 25-hydroxyvitamin D in 675 patients. *J Bone Miner Res*. 2010;25:305-312.
7. Quarles LD. "Dem bones" are made for more than walking. *Nat Med*. 2011;17:428-430.
8. Manger B, Schett G. Paraneoplastic syndromes in heumatology. *Nat Rev Rheumatol*. 2014;10: 663-670.
9. Chong WH, Molinolo AA, Chen CC, et al. Tumor-induced osteomalacia. *Endocr Relat Cancer*. 2011;18:53-77.
10. Chong WH, Andreopoulou P, Chen CC, et al. Tumor localization and biochemical response to cure in tumor-induced osteomalacia. *J Bone Miner Res*. 2013;28:1386-1398.
11. Clifton-Bligh RJ, Hofman MS, Duncan E, et al. Improving diagnosis of tumor-induced osteomalacia with gallium-68 DOTATATE PET/CT. *J Clin Endocrinol Metab*. 2013;98:687-694.
12. Jadhav S, Kasaliwal R, Shetty NS, et al. Radiofrequency ablation, an effective modality of treatment in tumor-induced osteomalacia: a case series of three patients. *J Clin Endocrinol Metab*. 2014;99: 3049-3054.

REVIEW QUESTIONS

1. Defective mineralization in osteomalacia is due to lack of one or more of the following except which one?

A. Adequate calcium and phosphorus at the remodeling site
B. Adequate amount of skeletal alkaline phosphatase
C. Normal pH at the site of calcification
D. Presence of a normal bone collagen matrix
E. Adequate level of fluoride

Answer: E Fluoride is an inhibitor of mineralization. See Pathobiology.

2. The correct sequence of steps involving bone remodeling is represented by which one of the following?

A. Reversal, activation, bone resorption by osteoclasts, osteoblast assembly and new bone formation
B. Activation, bone resorption by osteoclasts, reversal phase, osteoblast assembly and new bone formation
C. Bone resorption by osteoclasts, activation, osteoblast assembly and new bone formation, reversal
D. Activation, osteoblast assembly and new bone formation, bone resorption by osteoclasts, reversal
E. None of the above

Answer: B See Pathobiology.

3. All of the following may be signs or symptoms of osteomalacia except which one?

A. Nonspecific and poorly localized bone pain
B. Bone pain after sudden movements
C. Fasciculations and absent reflexes
D. Flat-footed waddling gait
E. Muscle weakness

Answer: C See Clinical Manifestations.

4. Excessive osteoid due to osteomalacia can be reliably distinguished from that caused by increased bone turnover by which one of the following?

A. Serum alkaline phosphatase
B. Presence of pseudofractures
C. Increased radionuclide on a bone scan
D. Phosphate level
E. Bone histomorphometry using tetracycline labels

Answer: E See Diagnosis.

5. A 30-year-old woman presents to her primary care physician with fatigue, generalized bone pains, weight loss, and a pruritic rash on her elbows and back. Laboratory studies show: hemoglobin 10 (male, 14-17 g/dL; female, 12-16 g/dL); serum iron 20 (60-160 µg/dL); serum calcium 8 (9-10.5 mg/dL). What is the next appropriate diagnostic test?

A. Bone mineral density scan
B. Hemoglobin electrophoresis
C. Colonoscopy
D. Antiendomysial immunoglobulin A antibodies
E. Skin biopsy

Answer: D The patient has gluten enteropathy, which may occur even without gastrointestinal symptoms. See Treatment.

245

THE PARATHYROID GLANDS, HYPERCALCEMIA AND HYPOCALCEMIA

RAJESH V. THAKKER

CALCIUM METABOLISM

A healthy adult body has a total of 1 kg of calcium; about 99% of this is present within the crystal structure of bone mineral, and less than 1% is in soluble form in the extracellular and intracellular fluid compartments. In the extracellular fluid compartment (ECF), about half of the total calcium is

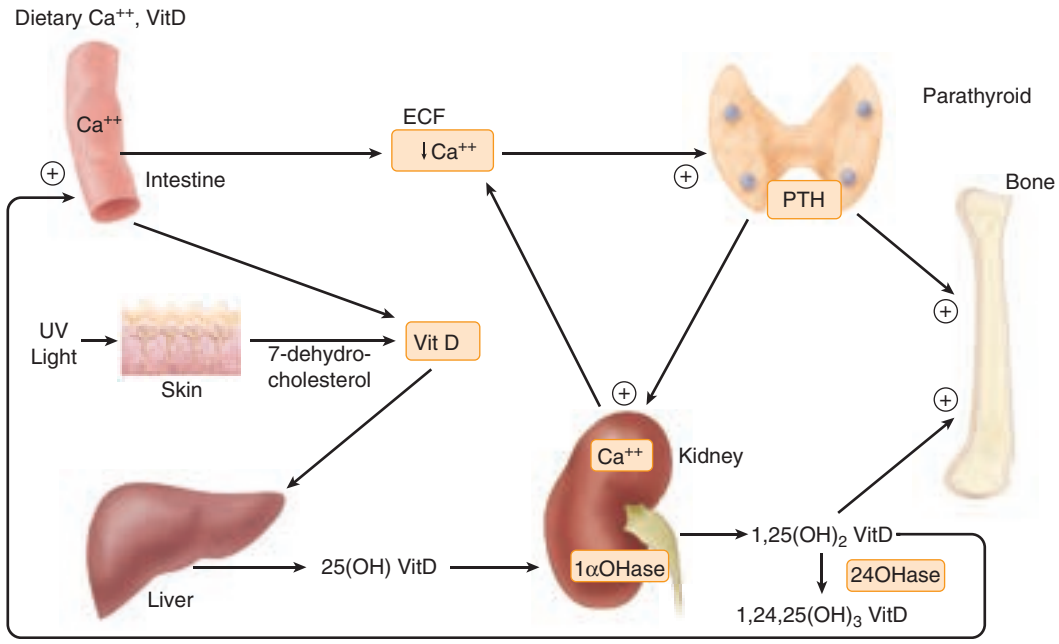

FIGURE 245-1. Regulation of extracellular fluid (ECF) calcium (Ca^{2+}) by parathyroid hormone (PTH) action on kidney, bone, and intestine. A decrease in ECF Ca^{2+} is sensed by the calcium-sensing receptor (see Fig. 245-2), and this leads to an increase in PTH secretion and a reduction in PTH degradation. The increased circulating PTH predominantly acts directly on kidney and bone that possess the PTH receptor (PTHR, Fig. 245-2). The skeletal effects of PTH are to increase (+) osteoclastic bone reabsorption. However, because osteoclasts do not have PTHRs, this action is mediated by the osteoblasts, which do have PTHRs and in response release cytokines and factors in turn that activate osteoclasts. In the kidney, PTH stimulates (+) the 1α-hydroxylase (1αOHase) to increase the conversion of 25-hydroxyvitamin D [25(OH)VitD] to the active metabolite 1,25-dihydroxyvitamin D [1,25$(OH)_2$VitD]. In addition, PTH increases (+) the reabsorption of Ca^{2+} from the renal distal tubule and inhibits the reabsorption of phosphate from the proximal tubule, thereby leading to hypercalcemia and hypophosphatemia. PTH also inhibits Na^+, H^+ antiporter activity and bicarbonate reabsorption, thereby causing a mild hyperchloremic acidosis. The elevated 1,25$(OH)_2$VitD acts on the intestine to increase (+) absorption of dietary calcium and phosphate. It is important to note that PTH does not appear to have a direct action on the gut. Thus, in response to hypocalcemia and the increase in PTH secretion, all of these direct and indirect actions of PTH on the kidney, bone, and intestine will help to increase ECF Ca^{2+}, which in turn will act through the calcium-sensing receptor to decrease PTH secretion. (From Thakker RV, Bringhurst FR, Jüppner HH. Regulation of calcium homeostasis and genetic disorders that affect calcium metabolism. In: Jameson JL, De Groot LJ, giudice LC, et al., eds. *Endocrinology: Adult & Pediatric.* 7th ed. Philadelphia: Saunders; 2016.)

ionized, and the rest is principally bound to albumin or complexed with counter-ions. Ionized calcium in the ECF plays an important role in many physiologic pathways, including muscle contraction, secretion of neurotransmitters and hormones, and coagulation pathways. Ionized serum calcium concentrations range from 4.65 to 5.25 mg/dL (1.16 to 1.31 mmol/L), and the total serum calcium concentration ranges from 8.5 to 10.5 mg/dL (2.12 to 2.62 mmol/L).[1] However, the usual 2 : 1 ratio of total to ionized calcium may be disturbed by disorders such as metabolic acidosis, which reduces calcium binding by proteins, or by changes in protein concentration, caused by cirrhosis, dehydration, venous stasis, or multiple myeloma1. In view of this, total serum calcium concentrations are adjusted, or "corrected," to a reference albumin concentration: the actual total serum calcium value is adjusted by adding or subtracting 0.8 mg/dL (0.016 mmol/L) for every 1 g/dL (1 g/L) of albumin below or above a reference albumin concentration of 4 g/dL (40 g/L), respectively.

The control of body calcium involves a balance between the amounts that are absorbed from the gut, deposited into bone and into cells, and excreted from the kidney (Fig. 245-1).[2] This fine balance, involving three organs, is chiefly under the control of parathyroid hormone (PTH), which is synthesized and secreted by the parathyroid glands. Hypocalcemia leads to an increased secretion of PTH, whereas hypercalcemia results in diminished PTH secretion. Regulation of extracellular calcium takes place through complex interactions (Fig. 245-2) at the target organs of the major calcium-regulating hormone, PTH, and vitamin D and its active metabolites, 1,25-dihydroxyvitamin D (1,25$[OH]_2$D).

PARATHYROID GLANDS, PARATHYROID HORMONE, *PTH* GENE, AND PARATHYROID HORMONE ACTIONS

Parathyroid Glands

There are usually four parathyroid glands, which are located in close proximity to the superior and inferior poles of the lobes of the thyroid gland. The superior parathyroids are derived from the endoderm of the embryonic fourth pharyngeal pouches, and the inferior parathyroids are derived with the thymus from the endoderm of the third pharyngeal pouches. Extra parathyroid glands are commonly found in aberrant locations along this migrating path and also within the thymus and thyroid. Parathyroid cells express a G protein–coupled receptor (GPCR), referred to as the *calcium-sensing receptor* (CaSR), that detects changes in extracellular calcium and leads to alterations in PTH secretions.[3] For example, activation of the CaSR, which is also expressed in renal tubular cells, as a result of elevated extracellular calcium concentrations causes G protein–dependent stimulation of phospholipase C activity through Gαq and Gα11, which leads to accumulation of inositol 1,4,5-trisphosphate and an increase in intracellular calcium concentrations.[4] These changes, in turn, lead to reduced circulating PTH concentrations and increased urinary calcium excretion. Disorders of the parathyroid glands may cause hypercalcemia or hypocalcemia, and these can be classified according to whether they arise from an excess of PTH, its deficiency, or insensitivity to its effects (Table 245-1; see Fig. 245-2).

Parathyroid Hormone and *PTH* Gene

The mature PTH peptide is encoded by the *PTH* gene and secreted from the parathyroid chief cells as an 84–amino acid peptide; however, when the *PTH* mRNA is first translated, it is as pre-proPTH peptide. The "pre" sequence consists of a 25–amino acid signal peptide (leader sequence) that is responsible for directing the nascent peptide into the endoplasmic reticulum to be packaged for secretion from the cell. The "pro" sequence is 6 amino acids in length and, although its function is less well defined than that of the "pre" sequence, is also essential for correct PTH processing and secretion. After the 84–amino acid mature PTH peptide is secreted from the parathyroid cell, it is cleared from the circulation with a short half-life of about 2 minutes, by nonsaturable hepatic uptake and renal excretion.

Parathyroid Hormone Actions

PTH shares a receptor with PTH-related peptide (PTHrP); this PTH/PTHrP receptor (see Fig. 245-2) is a member of a subgroup of the G protein–coupled receptor family.[5] PTH/PTHrP receptors are expressed in kidney and bone, where PTH is its predominant agonist, and thus PTH acts directly on kidney and bone cells and indirectly on intestinal cells (see Fig. 245-1) to enhance renal calcium reabsorption, release stored calcium in bones into the ECF, and increase gut calcium absorption, respectively. Expression of the PTH/PTHrP receptor also occurs in the brain, heart, skin, lung, liver, and testis, where it mediates the actions of PTHrP. Mutations

PTH is an 84–amino acid peptide encoded by the *PTH* gene, which is located on chromosome 11p15 and consists of three exons (transcribed regions) that are separated by two introns. Exon 1 of the *PTH* gene is 85 base pairs (bp) in length and is untranslated, whereas exons 2 and 3 code for the 115–amino acid pre-proPTH peptide. Exon 2 is 90 bp in length and encodes the initiation (ATG) codon, the prehormone sequence, and part of the prohormone sequence. Exon 3 is 612 bp and encodes the remainder of the prohormone sequence, the mature PTH peptide, and the 3′ untranslated region. The 5′ regulatory sequence of the human *PTH* gene contains a vitamin D response element 125 bp upstream of the transcription start site, which downregulates *PTH* messenger RNA (mRNA) transcription in response to vitamin D receptor binding. *PTH* gene transcription (as well as PTH peptide secretion) is also dependent on the extracellular calcium concentration, although the presence of a specific upstream "calcium response element" has not yet been demonstrated.

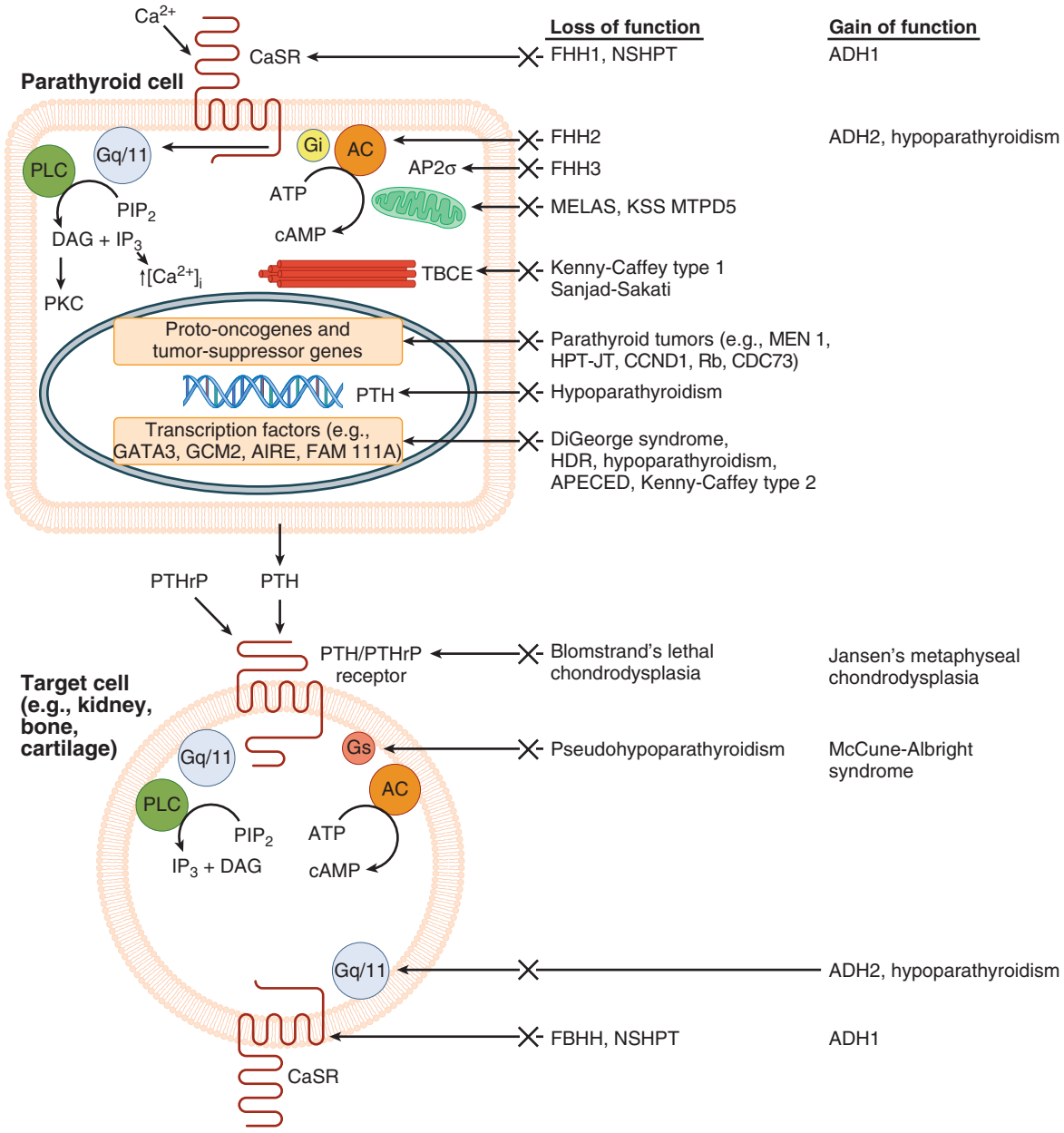

FIGURE 245-2. Schematic representation of some of the components involved in calcium homeostasis. Alterations in extracellular calcium are detected by the calcium-sensing receptor (CaSR), which is a 1078–amino acid G protein–coupled receptor. The PTH/PTHrP receptor, which mediates the actions of PTH and PTHrP, is also a G protein–coupled receptor. Thus, Ca^{2+}, PTH, and PTHrP involve G protein–coupled signaling pathways, and interaction with their specific receptors can lead to activation of Gs, Gi, and Gq, respectively. Gs stimulates adenylcyclase (AC), which catalyzes the formation of cyclic adenosine monophosphate (cAMP) from adenosine triphosphate (ATP). Gi inhibits AC activity. cAMP stimulates protein kinase A (PKA), which phosphorylates cell-specific substrates. Activation of Gq stimulates phospholipase C (PLC), which catalyzes the hydrolysis of the phosphoinositide (PIP_2) to inositol triphosphate (IP_3), which then increases intracellular calcium, and diacylglycerol (DAG), activating protein kinase C (PKC). These proximal signals modulate downstream pathways, which result in specific physiologic effects. Loss of function in several genes, shown with their respective sites of action on the right, has been identified in specific disorders of calcium homeostasis (also see Table 245-1). (From Thakker RV, Bringhurst FR, Jüppner H. Regulation of calcium homeostasis and genetic disorders that affect calcium metabolism. In: Jameson JL, De Groot LJ, giudice LC, et al., eds. *Endocrinology: Adult & Pediatric.* 7th ed. Philadelphia: Saunders; 2016.)

involving the genes that encode these proteins and receptors in this calcium-regulating pathway (see Fig. 245-2) are associated with hypercalcemic and hypocalcemic disorders (see Table 245-1).

Renal Actions

Calcium is absorbed at multiple sites and by different mechanisms, which include passive paracellular or active transcellular transport, along the renal tubule.[2] The renal actions of PTH are to (1) stimulate activity of the proximal tubular cell 1α-hydroxylase; (2) increase reabsorption of calcium by the cells of the distal tubule, connecting tubules and the thick ascending loop of Henle (TAL); and (3) inhibit phosphate reabsorption by proximal tubular cells (see Fig. 245-1). PTH increases the formation of biologically active $1,25(OH)_2D$ from its precursor 25-OH-D by stimulating the activity of the renal 1α-hydroxylase and inhibiting the 24-hydroxylase, which metabolizes $1,25(OH)_2D$ to the inactive $24,25(OH)_2D$ form (see Fig. 245-1). PTH regulates calcium reabsorption by distal tubular cells by upregulating expression of the transient receptor potential vanilloid 5 (TRPV5), thereby promoting calcium entry into the cell, and increasing calbindin-D28K expression to enhance transcellular calcium reabsorption by increased buffering of subapical Ca^{2+} ions. In the TAL, PTH may increase active transcellular transport of calcium, as well as paracellular calcium transport, by augmenting the transepithelial voltage gradient. Phosphate transport in proximal tubular cells is mediated by the luminal membrane sodium-phosphate cotransporters 2a and 2c (NPT2a and NPT2c), and PTH actions lead to internalization and degradation of NPT2a and NPT2c, thereby resulting in decreased reabsorption of phosphate.

Skeletal Actions

PTH acts directly on osteoblasts and indirectly on osteoclasts to increase their numbers and activity, thereby enhancing bone turnover and release of stored calcium. Thus, PTH increases the size of the osteoblast precursor pool, increases the bone-forming activity of mature osteoblasts, and stimulates

TABLE 245-1 PARATHYROID DISEASES AND THEIR CHROMOSOMAL LOCATIONS

METABOLIC ABNORMALITY	DISEASE	INHERITANCE	GENE/GENE PRODUCT	CHROMOSOMAL LOCATION
HYPERCALCEMIA				
	Multiple endocrine neoplasia type 1	Autosomal dominant	MENIN	11q13
	Multiple endocrine neoplasia type 2	Autosomal dominant	*RET*	10q11.2
	Hereditary hyperparathyroidism and jaw tumors (HPT-JT)	Autosomal dominant	PARAFIBROMIN	1q31.2
	Sporadic hyperparathyroidism	Sporadic	*PRAD1/CCND1*	11q13
			Retinoblastoma	13q14
			Unknown	1p32-pter
	Parathyroid carcinoma	Autosomal dominant or sporadic	PARAFIBROMIN	1q31.2
			Retinoblastoma	13q14
	Familial benign hypercalcemia (FBH)			
	FBH1	Autosomal dominant	*CaSR*	3q 21.1
	FBH2	Autosomal dominant	Gα11	19p13
	FBH3	Autosomal dominant	*AP2S1*	19q13
	Neonatal severe hyperparathyroidism (NSHPT)	Autosomal recessive or autosomal dominant	*CaSR*	3q21.1
	Jansen's disease	Autosomal dominant	PTHR/PTHrP receptor	3p21.3
	Williams syndrome	Autosomal dominant	*Elastin, LIMK* (and other genes)	7q11.23
	Infantile hypercalcemia	Autosomal recessive	*CYP24A*	20q13.2-q13.3
	McCune-Albright syndrome	Mutations during early embryonic development?	Gsα	20q13.3
HYPOCALCEMIA				
	Isolated hypoparathyroidism	Autosomal dominant	*PTH, GCMB*	11p15*
		Autosomal recessive	*PTH, GCMB*	11p15*, 6p24.2
		X-linked recessive	*SOX3*	Xq26–27
	Autosomal dominant hypocalcemia type 1 (ADH1)	Autosomal dominant	*CaSR*	3q21.1
	Autosomal dominant hypocalcemia type 2 (ADH2)	Autosomal dominant	Gα11	19p13
	Hypoparathyroidism associated with polyglandular autoimmune syndrome (APECED)	Autosomal recessive	*AIRE-1*	21q22.3
	Hypoparathyroidism associated with Kearns-Sayre and MELAS	Maternal	Mitochondrial genome	
	Hypoparathyroidism associated with complex congenital syndromes			
	DiGeorge syndrome	Autosomal dominant	*TBX1*	22q11.2/10p
	HDR syndrome	Autosomal dominant	*GATA3*	10p15
	Blomstrand's lethal chondrodysplasia	Autosomal recessive	PTHR/PTHrP receptor	3p21.3
	Kenney-Caffey syndrome type 1, Sanjad-Sakati syndrome	Autosomal dominant	*TBCE*	1q42.3
	Kenney-Caffey syndrome type 2	Autosomal recessive	*FAMIIIA*	11q12.1
	Barakat syndrome	Autosomal recessive†	Unknown	?
	Lymphedema	Autosomal recessive	Unknown	?
	Nephropathy, nerve deafness	Autosomal dominant†	Unknown	?
	Nerve deafness without renal dysplasia	Autosomal dominant	Unknown?	?
	Pseudohypoparathyroidism (type 1a)	Autosomal dominant parentally imprinted	*GNAS exons 1-3*	20q13.3
	Pseudohypoparathyroidism (type 1b)	Autosomal dominant parentally imprinted	*GNAS* Upstream deletion	20q13.3

HDR = hypoparathyroidism, deafness, and renal dysplasia; MELAS = mitochondrial encephalopathy, stroke-like episodes, and lactic acidosis; ? = location not known.
*Mutations of PTH gene are identified only in some families.
†Most likely inheritance.

osteoblasts to release cytokines such as colony-stimulating factor 1 and receptor activator of nuclear factor-κB (NF-κB) ligand (RANKL), which stimulate the formation of new osteoclasts and activate mature osteoclasts. PTH also inhibits osteoblast production of osteoprotegrin (OPG), which is a soluble decoy receptor for RANKL that inhibits osteoclast development. Calcium transport involves TRPV4 and TRPV5 in bone cells; TRPV4 regulates intracellular calcium concentrations in osteoblasts and osteoclasts, whereas TRPV5, expressed in osteoclasts, participates to remove the mineral bone matrix.[2] The net result of persistent elevations of PTH is linked to an increase in osteoclast activity more than osteoblast activity, hence liberating the stores of calcium to the ECF (see Fig. 245-1).

Intestinal Actions

Calcium is absorbed throughout the intestine by passive paracellular routes and active transcellular routes, which involve TRPV6 and calbindin D9K. PTH exerts indirect actions on intestinal calcium absorption by increasing the circulating $1,25(OH)_2D$ concentrations (see Fig. 245-1). The increased $1,25(OH)_2D$ concentrations increase TRPV6 expression, which facilitates enhanced calcium entry into the cell from the lumen, and cytosolic calbindin D9K expression, which facilitates transcellular transport of calcium.

HYPERCALCEMIA

DEFINITION

Hypercalcemia is defined as a serum calcium concentration greater than 2 standard deviations above the normal mean, and this is usually a total serum calcium above 10.5 mg/dL (2.62 mmol/L) and an ionized serum calcium of above 5.25 mg/dL (1.31 mmol/L). There is no formal grading system for defining the severity of hypercalcemia, but mild, moderate, and severe hypercalcemia is generally considered for total serum calcium concentrations less than 12 mg/dL (3 mmol/L), between 12 and 14 mg/dL (3 to 3.5 mmol/L), and greater than 14 mg/dL (3.50 mmol/L), respectively.

PATHOBIOLOGY

Hypercalcemia may arise through one of three mechanisms: increased bone resorption, increased gastrointestinal absorption of calcium, and decreased renal calcium excretion (see Fig. 245-1). For example, lytic bone metastases cause increased bone resorption; thiazide diuretics lead to a decrease in calcium excretion; and excessive PTH will either directly or indirectly, by increasing $1,25(OH)_2D$ production, stimulate bone resorption and calcium

absorption from the gut and renal tubules.[6] The causes of hypercalcemia may be classified according to whether serum PTH concentrations are elevated (i.e., primary or tertiary hyperparathyroidism due to parathyroid tumors) or reduced (i.e., not due to parathyroid tumors but instead to an excessive production of PTHrP by a cancer; a defect in the PTH receptor, for example, the PTH/PTHrP receptor; an excess production of downstream mediators, for example, $1,25(OH)_2D$; or an altered set point in the calcium-sensing receptor) (Table 245-2; see Fig. 245-2). Primary hyperparathyroidism and malignancy are the most common causes and account for more than 90% of patients with hypercalcemia. Detailed clinical history and examination will usually help to differentiate between these two diagnoses. In primary hyperparathyroidism, the hypercalcemia is often less than 12 mg/dL (3 mmol/L), asymptomatic, and may have been present for months or years. If symptoms, such as nephrolithiasis, are present, then they have usually been present for several months. However, in malignancy, the patients are usually acutely ill, often with neurologic symptoms; the hypercalcemia is more than 12 mg/dL (3 mmol/L); and the cancer (e.g., lung, breast, or myeloma) is often readily apparent. Hypercalcemia from causes other than primary hyperparathyroidism or malignancy may also occur (see Table 245-2), and a careful history (e.g., for vitamin D ingestion, drugs, renal disease) and examination (e.g., for thyrotoxicosis, adrenal disease, granulomatous diseases), together with appropriate investigations (Table 245-3; Fig. 245-3), are essential for establishing the diagnosis.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The clinical presentation of hypercalcemia varies from a mild, asymptomatic, biochemical abnormality detected during routine screening to a life-threatening medical emergency. In general, the presence or absence of symptoms correlates with the severity and rapidity of onset of the hypercalcemia. Thus, symptoms do not usually develop when serum calcium is below 12 mg/dL (3 mmol/L) and are invariably present when the hypercalcemia exceeds 14 mg/dL (3.5 mmol/L). However, there is a considerable variability, and some patients may be symptomatic with mild hypercalcemia. Although there are many causes of hypercalcemia (see Table 245-2), the signs and symptoms of hypercalcemia are similar, regardless of etiology. Indeed, the clinical manifestations of hypercalcemia involve several organ systems that include the renal, musculoskeletal, gastrointestinal, neurologic, and cardiac systems (Table 245-4), and many of these have been referred to as "moans, groans, pains, and stones." Investigations should be directed at confirming the presence of hypercalcemia and establishing the cause (Table 245-5; see Table 245-3).

TABLE 245-2 CAUSES OF HYPERCALCEMIA

HIGH PARATHYROID HORMONE LEVELS

Primary hyperparathyroidism* (adenoma, hyperplasia, or carcinoma): nonfamilial or familial, e.g., MEN 1, MEN 2, HPT-JT, FIHP
Tertiary hyperparathyroidism (hyperplasia or adenoma in chronic renal failure)

LOW PARATHYROID HORMONE LEVELS

Malignancy*
- Primary
  - Parathyroid hormone–related peptide (PTHrP): carcinoma of lung, esophagus, renal cell, ovary, and bladder
  - Excess production of $1,25(OH)_2D$ (lymphoma)
- Secondary
  - Lytic bone metastases* (multiple myeloma* and breast carcinoma*)
  - Other location, ectopic factors (e.g., cytokines)

Excess vitamin D
- Exogenous vitamin D toxicity by parent D compound, 25(OH) vitamin D_3, or $1,25(OH)_2$ vitamin D_3 in vitamin preparations, cod liver oil, herbal medicines
- Endogenous production of 25(OH) vitamin D_3—Williams syndrome
- Endogenous production of $1,25(OH)_2$ vitamin D_3, e.g., granulomatous disorders (sarcoidosis, HIV, TB, histoplasmosis, coccidioidomycosis, leprosy), lymphoma, and infantile hypercalcemia

Drugs
- Thiazide diuretics
- Lithium
- Total parenteral nutrition
- Estrogens/antiestrogens, testosterone
- Milk-alkali syndrome
- Vitamin A toxicity
- Aluminum intoxication (in chronic renal failure)
- Aminophylline

Nonparathyroid endocrine disorders
- Thyrotoxicosis
- Pheochromocytoma
- Acute adrenal insufficiency
- Vasoactive intestinal polypeptide hormone producing tumor (VIPoma)
- Immobilization

INAPPROPRIATE PARATHYROID HORMONE LEVELS DUE TO ALTERED SET POINT

Familial benign hypocalciuric hypercalcemia (FBH or FHH)

*Most common causes.
FIHP = familial isolated hyperparathyroidism; HIV = human immunodeficiency virus; HPT-JT = hyperparathyroidism with jaw tumors; MEN = multiple endocrine neoplasia; TB = tuberculosis.

TABLE 245-3 PRELIMINARY INVESTIGATIONS FOR HYPERCALCEMIA

BLOOD

× 2-3 estimations of serum calcium, phosphate, albumin, urea and electrolytes, creatinine, alkaline phosphatase, liver function tests
Parathyroid hormone
Complete blood count
Electrophoretic protein strip or serum protein electrophoresis
25-OH-D_3 (and if indicated, $1,25[OH]_2D_3$)
Thyroid function tests
Magnesium
Parathyroid hormone–related peptide (if malignancy suspected)

URINE

× 2-3 estimations of 24-hr urinary calcium and creatinine clearance, and clearance ratios

Imaging
- Chest radiograph
- Radiograph of hands
- Ultrasound of kidneys

TABLE 245-4 CLINICAL FEATURES OF HYPERCALCEMIA

Renal
- Stones (nephrolithiasis) and nephrocalcinosis, polyuria, polydipsia

Musculoskeletal
- Bone pain, osteopenia, fractures, muscular weakness, especially proximal myopathy

Gastrointestinal
- Nausea, vomiting, lack of appetite, constipation, peptic ulcers, and pancreatitis

Neurologic
- Tiredness, lethargy, inability to concentrate, increased sleepiness, depression, confusion, coma

Cardiac
- Bradycardia, first-degree atrioventricular block, arrhythmias, shortened QT interval

TABLE 245-5 SUMMARY OF GUIDELINES FOR PARATHYROID SURGERY IN PRIMARY HYPERPARATHYROIDISM PATIENTS

Surgery* recommended if patient meets any one of the following criteria:
- Serum calcium >1 mg/dL (0.25 mmol/L) above upper limit of normal
- Any complication of primary hyperparathyroidism (e.g., nephrolithiasis† or bone erosions of osteitis fibrosa cystica)
- An episode of acute primary hyperparathyroidism with life-threatening hypercalcemia
- Significant reduction in creatinine clearance (i.e., <60 mL/min)
- Reduction in bone mineral density (i.e., T score <–2.5, and/or previous fracture fragility)
- Age <50 years

*Surgery is also indicated in patients for whom medical surveillance is neither desired nor possible.
†Some physicians still regard marked hypercalciuria (>9 mmol/L per 24 hr or >400 mg/24 hr) as an indication for surgery.
Adapted from Bilezikian JP, Khan AA, Potts JT Jr. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the third international workshop. *J Clin Endocrinol Metab.* 2009;94:335-339.

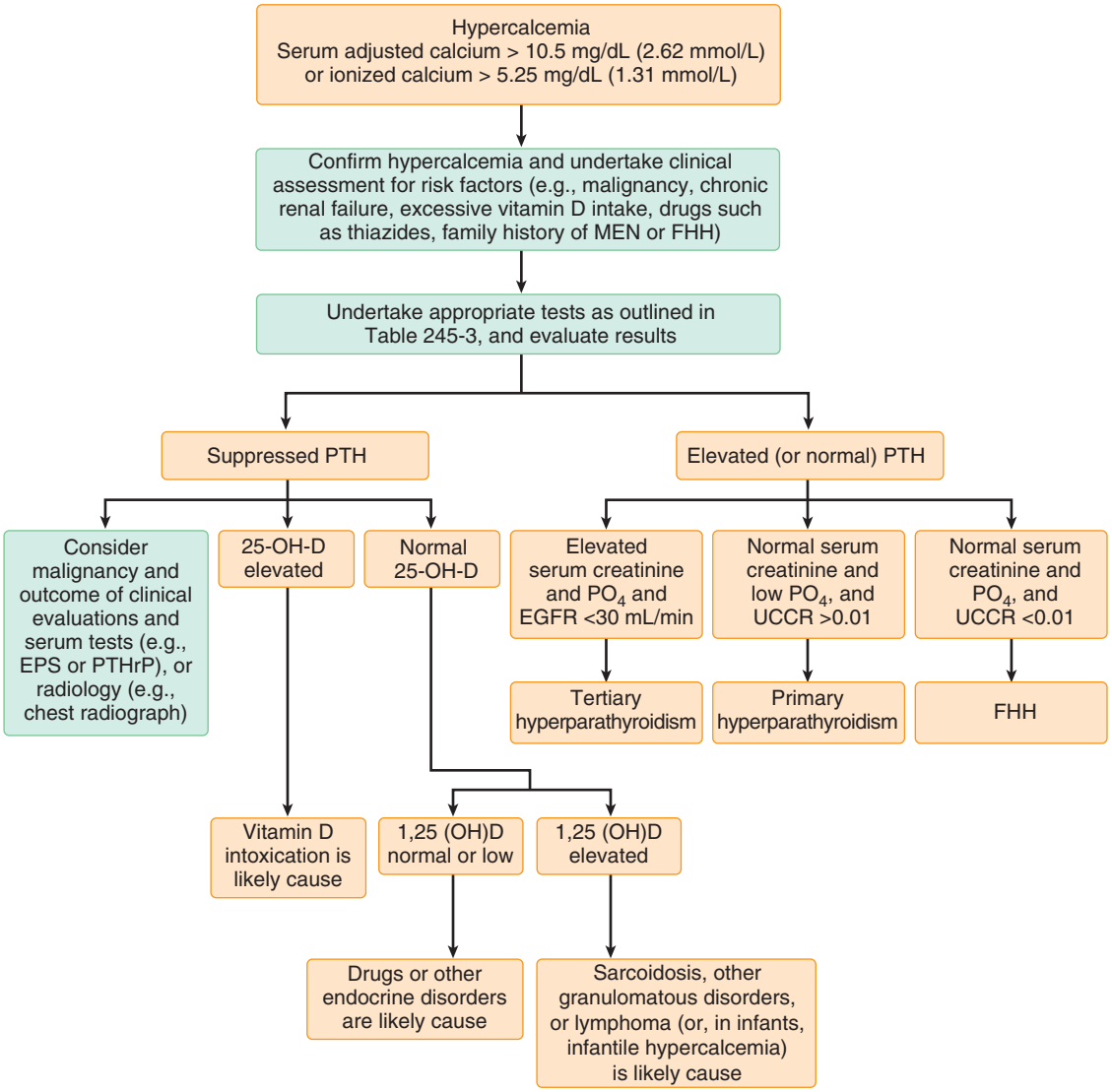

FIGURE 245-3. **Clinical approach to the investigation of causes of hypercalcemia. $1,25(OH)_2D$ = 1,25-dihyroxyvitamin D; 25-OH-D = 25-hydroxyvitamin D; EGFR = estimated glomerular filtration rate; EPS = electrophoretic strip (serum protein electrophoresis); FHH = familial hypocalciuric hypercalcemia; MEN = multiple endocrine neoplasia; PTH = parathyroid hormone; PTHrP = parathyroid hormone–related peptide; UCCR = 24-hour urinary calcium clearance–to–creatinine clearance ratio.**

TREATMENT

Rx

The treatment of hypercalcemia depends on the severity of the hypercalcemia and the presence of symptoms. Thus, asymptomatic patients with mild hypercalcemia do not usually need urgent treatment, whereas patients with severe hypercalcemia would require treatment regardless of symptoms, and patients with moderate hypercalcemia would require urgent treatment if symptomatic. Before instituting treatment, it is always important to consider the underlying causes (see Table 245-2) and to initiate investigations (see Table 245-3). In addition, drugs such as thiazides and vitamin D compounds, which cause hypercalcemia, should be discontinued and, if appropriate, dietary calcium restricted.

The acute management of hypercalcemia involves general measures to enhance hydration and diuresis and specific measures using drugs to lower serum calcium. Dehydration due to hypercalcemic symptoms, such as anorexia, nausea, vomiting, and polyuria because of defective urinary concentration, is very common, and patients may require 5 to 10 liters of 0.9% sodium chloride over a 24- to 48-hour period. This vigorous hydration with normal saline may lower serum calcium by 1 to 3 mg/dL (0.25 to 0.75 mmol/L); it enhances urinary calcium excretion by increasing glomerular filtration and reducing proximal and distal renal tubular reabsorption of calcium and sodium. This saline diuresis may need adjuvant therapy with a loop diuretic (e.g., furosemide, 10 to 20 mg), as necessary to control complications due to volume overload, especially in elderly patients and those with impaired cardiovascular and renal function. Note that excessive use of furosemide before intravascular volume has been restored may worsen the hypercalcemia by exacerbating volume depletion. Saline diuresis may lead to hypokalemia, hypomagnesemia, and electrolyte imbalance, which will need correction.

If saline diuresis is not successful, particularly if the hypercalcemia is very severe, then more specific measures, such dialysis and/or drugs, will be required. The drugs of choice are pamidronate and zoledronic acid, which are potent bisphosphonates, but these should not be used if the hypercalcemia is due to primary or tertiary hyperparathyroidism. Recommended treatments are to administer pamidronate (15-60 mg, depending on serum calcium concentration, in a single IV infusion or in divided doses, depending upon renal function and responses, over 2-4 days; maximum of 90 mg per treatment course) or zoledronic acid (4 mg as single IV infusion). Other bisphosphonates (e.g., etidronate and clodronate) and other agents, such as mithramycin, calcitonin, and gallium nitrate, have also been used in the past. Glucocorticoid therapy (e.g., hydrocortisone, 120 mg/day in three divided doses, in adults) is particularly effective when the hypercalcemia is mediated by the actions of $1,25(OH)_2D$, for example in granulomatous disease or lymphoma, or myeloma. Dialysis using a low or zero calcium dialysate should be considered if these treatments are not effective or if the patient has renal failure. When the acute management of hypercalcemia has been completed, appropriate treatment for the underlying cause needs to be undertaken.

HYPERPARATHYROIDISM

DEFINITION

Hyperparathyroidism is characterized by high concentrations of serum immunoreactive PTH, and three types, referred to as primary, secondary, and tertiary, are recognized. Primary and tertiary hyperparathyroidism are associated with hypercalcemia (see Table 245-2), whereas secondary

hyperparathyroidism is associated with hypocalcemia (see later). Primary hyperparathyroidism usually occurs as an isolated nonsyndromic endocrinopathy and less commonly as part of complex syndromic disorders such as the multiple endocrine neoplasia (MEN)[7] and hyperparathyroidism with jaw tumors (HPT-JT). Syndromic and nonsyndromic forms of primary hyperparathyroidism may also occur as hereditary (i.e., familial), usually autosomal dominant disorders, or they may occur as nonfamilial (i.e., sporadic) diseases. Tertiary hyperparathyroidism usually arises in association with chronic renal failure.

Primary Hyperparathyroidism

EPIDEMIOLOGY

Primary hyperparathyroidism, which affects 3 in 1000 adults, is one of the two most common causes of hypercalcemia and is due to an excessive secretion of PTH from one or more parathyroid tumors. Studies have estimated that the global prevalence of parathyroid tumors is 4 million. Primary hyperparathyroidism usually occurs as a nonsyndromic isolated endocrinopathy, between the ages of 40 and 65 years, and is three times more common in females than males.

PATHOBIOLOGY

Eighty percent of patients with primary hyperparathyroidism will have a solitary parathyroid adenoma, and 15 to 20% of patients will have hyperplasia involving all four parathyroid glands. Parathyroid carcinoma occurs in less than 0.5% of patients with primary hyperparathyroidism. The underlying causes of primary hyperparathyroidism are largely unknown. However, more than 10% of patients with clinically nonfamilial primary hyperparathyroidism occurring before 45 years of age have a germline mutation in 1 of 11 genes, including those of MEN 1 (*MEN1*), cell division cycle 73 (*CDC73*), and *CaSR*. In addition, studies of nonfamilial sporadic parathyroid adenomas have shown that 35% to 50% have somatic mutation of the *MEN1* gene; 15% have overexpression of cyclin D1; and more than 85% have an abnormality of the Wnt/β-catenin pathway.[8,9]

CLINICAL MANIFESTATIONS

Many patients with primary hyperparathyroidism are asymptomatic, and the hypercalcemia, which is usually mild, is detected by chance at the time of biochemical screening for other reasons.[10,11] However, it is important to note that nearly half the patients have subtle neuromuscular symptoms such as fatigue and weakness, and this becomes apparent only in retrospect after a successful parathyroidectomy.

Symptomatic hypercalcemia (see Table 245-4) predominantly affects the skeletal, renal, and gastrointestinal systems; peptic ulcers and pancreatitis may develop. The skeletal changes of osteitis fibrosa cystica due to subperiosteal resorption of the distal phalanges, tapering of the distal clavicles, a salt-and-pepper appearance of the skull, bone cysts, and brown tumors of the long bones are now identified in less than 5% of patients. However, osteopenia, as assessed by bone mineral density, occurs in 25% of patients. Renal stone disease (nephrolithiasis and nephrocalcinosis) occurs in 20% of patients, and hypercalciuria occurs in 30% of patients; renal impairment may complicate this disease.

DIAGNOSIS

In the presence of hypercalcemia, the finding of elevated circulating PTH concentrations establishes the diagnosis because PTH is elevated in approximately 90% of patients with primary hyperparathyroidism, who invariably have hypercalcemia (see Fig. 245-3). However, it is important to make sure that the immunoradiometric (IRMA) and immunochemiluminometric (ICMA) assays for PTH are being used to measure the intact molecule, rather than the older radioimmunoassays, which were not as reliable. The only other hypercalcemic disorders in which PTH may occasionally be elevated are those related to familial benign hypocalciuric hypercalcemia (FBH or FHH), immobilization, or lithium or thiazide use (see Table 245-2), and a careful history and a cessation of drug use helps to exclude these possibilities.[12,13] The hypercalcemia of primary hyperparathyroidism, unlike that of malignancy or granulomatous disease, is usually not suppressible by a 10-day course of oral hydrocortisone (120 mg/day given in three divided doses). This test, referred to as the *steroid suppression test,* was previously used to differentiate primary hyperparathyroidism from other causes of hypercalcemia; however, with the advent of more reliable PTH assays, this test is rarely used now. About one third of patients with primary hyperparathyroidism have a low serum phosphate level (see Fig. 245-3), and in the others, it is in the lower range of normal. In addition, some patients have a small increase in serum chloride concentration and a concomitant decrease in bicarbonate concentration. Serum alkaline phosphatase activity may be elevated in some patients, and urinary calcium excretion is increased in 30% of patients. The circulating $1,25(OH)_2D$ concentration is elevated in some patients with primary hyperparathyroidism, although it is not of diagnostic value because it is also elevated in other hypercalcemic disorders such as sarcoidosis and lymphomas (see Fig. 245-3). The serum 25-OH-D concentration is within the normal range. Densitometric scanning is of use in detecting early skeletal changes. Patients with primary hyperparathyroidism develop reduced bone mineral densities (osteopenia) primarily of the cortical bone (e.g., distal third of forearm) rather than the cancellous bone (e.g., lumbar spine). The hip bones, which are an equal mixture of cortical and cancellous bone, show intermediate reductions in bone mineral density. Overall, the risk for bone fractures in patients with mild primary hyperparathyroidism is similar to those in matched, normal controls. However, successful parathyoidectomy does lead to an increase in bone mineral density over a 6- to 12-month period, and this continues for up to 10 years. Indeed, bone mineral density measurements are used in the evaluation of patients with primary hyperparathyroidism and in deciding on conservative as opposed to surgical management (see Table 245-5).

Preoperative localization to define the sites of the parathyroid tumors may be undertaken.[14] The noninvasive tests consist of ultrasonography, computed tomography (CT), magnetic resonance imaging (MRI), and scintigraphy with technetium-99m sestamibi. Sestamibi scintigraphy has now become established as the best and most convenient localization test; this can be performed with CT techniques (e.g., single-photon emission computed tomography [SPECT]) to give a three-dimensional image with greater anatomic resolution. It is important to note that there is an appreciable incidence of false-positive rates with all the noninvasive localization procedures, so a confirmation using two methods is preferable. Invasive localization tests consist of arteriography and selective venous sampling for PTH in the veins draining the thyroidal region. These tests are time consuming, expensive, difficult, and dependent on the skill of the radiologist. It is generally accepted that these preoperative localization tests are indicated in those patients who have had previous neck surgery. However, their role in patients who have not had prior surgery remains to be established, and at present, the preferences and expertise of the local medical, radiology, and surgery teams usually determine the use of venous sampling procedures.

TREATMENT

Rx

Parathyroidectomy, which is the definitive cure, is a generally successful and safe procedure if undertaken by an experienced surgeon. There have also been major advances in surgery that have facilitated a surgical approach to be undertaken under local, as opposed to general, anesthesia. An example of this is the use of minimally invasive parathyroidectomy (MIP) in the patient with single gland disease that has been successfully localized by the combined use of sestamibi scintigraphy and ultrasonography. Surgery is recommended for symptomatic patients and for those who have skeletal and renal complications (see Table 245-5). Complications of parathyroid surgery include damage to the recurrent laryngeal nerve and permanent hypoparathyroidism. However, the decision to recommend surgery may be difficult in asymptomatic patients, who may constitute more than 50% of patients with primary hyperparathyroidism. The natural history of primary hyperparathyroidism in most patients is to progress slowly or not at all. For example, among asymptomatic patients, only 25% have progressive disease, which is usually manifested as a decrease in bone mineral density during a 10-year period. This has led to a controversy regarding the indications for surgery, and guidelines have been provided by the Third International Workshop (2008) on the Management of Asymptomatic Primary Hyperparathyroidism (see Table 245-5). However, these guidelines may not exclusively influence the decision for or against surgery, and a careful evaluation and assessment of the risks and benefits is considered by most medical and surgical teams in conjunction with the patient. Clearly, some patients will not wish to continue living with a curable disease and will prefer surgery despite the guidelines (see Table 245-5), whereas other patients will decline surgery, despite having guideline indications for surgery, because they may have coexisting medical conditions that make them feel that the risks for surgery are too great.

Patients who do not undergo parathyroidectomy should be evaluated clinically and also monitored for serum calcium, creatinine, and PTH at 12-month intervals, and for bone mineral density at 12- to 24-month intervals. In addition, the following medical guidelines are recommended. First, they should avoid dehydration and remain ambulant. Second, the dietary intake of calcium should be moderate, that is, at or below 1000 mg/day, and thiazide diuretics

should be avoided. Finally, they should avoid herbal and tonic remedies that may contain vitamin D or vitamin A. Drugs that have been used for the treatment of primary hyperparathyroidism include oral phosphate, estrogens, or selective estrogen receptor modulators (SERMs) in postmenopausal women; bisphosphonates; and the calcimimetic, cinacalcet. Phosphate is not used because of concerns related to soft tissue ectopic calcification. Estrogens and SERMs (e.g., raloxifene) do increase bone density in postmenopausal women with primary hyperparathyroidism, but they have only small effects on the serum calcium and PTH concentrations. The bisphosphonates (e.g., alendronate and zoledronic acid) inhibit bone resorption and reduce serum calcium. However, these effects are not sustained. In a randomized, double-blind, placebo-controlled clinical trial, cinacalcet was effective in lowering serum calcium concentrations to normal values and reducing PTH levels in patients with primary hyperparathyroidism.[E] These effects were maintained with long-term treatment without major adverse effects. Bone mineral density in the treated patients remained unchanged, but there was a reduction in biochemical markers for bone resorption and formation. Cinacalcet may therefore represent an effective nonsurgical treatment for the management of primary hyperparathyroidism. Daily high dose (70 mg or 2800U) cholecalciferol also can decrease PTH and improve bone density when used for 6 months before and after parathyroidectomy.[E]

FAMILIAL PRIMARY HYPERPARATHYROIDISM

Primary hyperparathyroidism is most frequently encountered as a nonfamilial (sporadic) disorder. However, approximately 10% of patients with primary hyperparathyroidism have a hereditary form that may either be part of the MEN 1, MEN 2, MEN 3, and MEN 4 syndromes or part of the HPT-JT syndrome. In addition, hereditary primary hyperparathyroidism may develop as a solitary endocrinopathy, and this has also been referred to as familial isolated hyperparathyroidism (FIHP). Patients with these familial forms of primary hyperparathyroidism, including the MEN syndromes, have important differences from those developing nonfamilial forms; these include an earlier age of onset (20 to 25 years versus 55 years) and an equal male-to-female ratio (1 : 1 versus 1 : 3). In addition, MEN syndromes are associated with the occurrence of multiple parathyroid tumors, rather than the solitary parathyroid adenomas typically found in the sporadic form; and HPT-JT is associated with occurrence of parathyroid carcinoma in 15% of patients. This has implications for the treatment of parathyroid tumors in patients with these disorders. Thus, minimally invasive parathyroidectomy is an unsuitable approach in MEN patients because of multigland disease, and patients with HPT-JT are likely to require earlier surgery because of the higher risk for parathyroid carcinoma. Investigations of the hereditary and sporadic forms of primary HPT have helped to identify some of the genes and chromosomal regions that are involved in the etiology of parathyroid tumors (see Table 245-1 and Fig. 245-2). FIHP has been reported in several kindreds, and some have been shown to harbor mutations of the *MEN1*, *CDC73*, or *CaSR* genes. The familial syndromes associated with parathyroid tumors include MEN 1, MEN 2, MEN 3, and MEN 4. They are reviewed in detail in Chapter 231.

HYPERPARATHYROIDISM–JAW TUMOR SYNDROME

The HPT-JT syndrome is an autosomal dominant disorder characterized by the occurrence of parathyroid tumors, which may be carcinomas in approximately 15% of patients, and ossifying fibromas that usually affect the maxilla or mandible. In addition, some patients may also develop Wilms' tumors, renal cysts, renal hamartomas, renal cortical adenomas, papillary renal cell carcinomas, uterine tumors that may be malignant, pancreatic adenocarcinomas, testicular mixed germ cell tumors with a major seminoma component, and Hurthle cell thyroid adenomas. Mutations of the *CDC73* gene, which is located on chromosome 1q31.2 and encodes a 531–amino acid protein, parafibromin, cause HPT-JT. Parafibromin has been shown to be associated with the human homologue of the Paf1 protein complex, which interacts with RNA polymerase II, and as part of this protein complex, parafibromin may regulate post-transcriptional events and histone modification. Patients with nonfamilial parathyroid carcinomas frequently harbor germline *CDC73* mutations.

UREMIC HYPERPARATHYROIDISM

PATHOBIOLOGY

Serum PTH levels rise in response to hypocalcemia, and this secondary hyperparathyroidism usually resolves with treatment of the underlying cause of hypocalcemia (Table 245-6). However, in chronic renal failure (Chapter 130), the secondary hyperparathyroidism may persist for a longer time, and eventually the parathyroid cells gain an autonomous function, secreting excessive PTH despite hypercalcemia; this state is referred to as tertiary hyperparathyroidism (see Table 245-2). The cause of progression from the early, presumably polyclonal, secondary hyperplasia of the parathyroids to the later, presumably monoclonal, tumors is not understood and appears to involve genes other than those involved in the etiologies of the sporadic and familial forms of primary hyperparathyroidism (see Table 245-2).

TABLE 245-6 CAUSES OF HYPOCALCEMIA

LOW PARATHYROID HORMONE LEVELS (HYPOPARATHYROIDISM)
Parathyroid agenesis
Isolated or part of complex developmental anomaly (e.g., DiGeorge syndrome)
Parathyroid destruction
Surgery*
Radiation
Infiltration by metastases or systemic disease (e.g., hemochromatosis, amyloidosis, sarcoidosis, Wilson's disease, thalassemia)
Autoimmune
Isolated
Polyglandular (type 1)
Reduced parathyroid function (i.e., parathyroid hormone secretion)
Parathyroid hormone gene defects
Hypomagnesemia*
Neonatal hypocalcemia (may be associated with maternal hypercalcemia)
Hungry bone disease (postparathyroidectomy)
Calcium-sensing receptor or Gα11 mutations
HIGH PARATHYROID HORMONE LEVELS (SECONDARY HYPERPARATHYROIDISM)
Vitamin D deficiency*
As a result of nutritional lack,* malabsorption,* liver disease, or vitamin D receptor defects
Inadequate production of active vitamin D ($1,25[OH]_2D$) as a result of chronic renal failure*
Vitamin D resistance (rickets)
As a result of renal tubular dysfunction (Fanconi's syndrome) or vitamin D receptor defects
Parathyroid hormone resistance
(e.g., pseudohypoparathyroidism, hypomagnesemia)
Drugs
Calcium chelators (e.g., citrated blood transfusions, phosphate—cow's milk is rich in phosphate)
Inhibitors of bone resorption (e.g., bisphosphonates, calcitonin, plicamycin)
Altered vitamin D metabolism (e.g., phenytoin, ketoconazole)
Foscarnet
Miscellaneous
Acute pancreatitis
Acute rhabdomyolysis
Massive tumor lysis
Osteoblastic metastases (e.g., from prostate or breast carcinoma)
Toxic shock syndrome
Hyperventilation

*Most common causes.

CLINICAL MANIFESTATIONS AND TREATMENT

Rx

In chronic renal failure, the ensuing phosphate retention and decreased production of $1,25(OH)_2D$ result in hypocalcemia and secondary hyperparathyroidism. This combination of biochemical abnormalities results in a severe bone disease that shows combined features of hyperparathyroidism and vitamin D deficiency (i.e., osteomalacia). Thus, in renal osteodystrophy, bone erosions and osteomalacia are simultaneously observed. Treatment is based on correcting the hypocalcemia, for example, with oral administration of calcium salts, which also ameliorates the hyperphosphatemia by chelating phosphate in the intestines, and with calcitriol ($1,25[OH]_2D$). The use of the most appropriate phosphate binder is not well established, but it is clear that aluminum-containing compounds are to be avoided. Aluminum in these preparations and as a contaminant of dialysis solutions contributed in the recent past to the osteomalacic osseous disease and other aspects of metal toxicity in patients with renal failure (e.g., hypochromic anemia and encephalopathy). Early treatment of the metabolic disturbance will prevent or delay

the onset of severe secondary hyperparathyroidism and tertiary hyperparathyroidism, which requires parathyroidectomy. For patients who have end-stage renal failure and are on dialysis, cinacalcet, the allosteric activator of the CaSR, can be used to treat the severe secondary hyperparathyroidism. Cinacalcet will reduce the PTH concentrations and may also have an antiproliferative effect.

DISORDERS AND SYNDROMES ASSOCIATED WITH HYPERCALCEMIA

Endocrine Causes of Hypercalcemia Other than Hyperparathyroidism

Several nonparathyroid disorders (see Table 245-2) are associated with hypercalcemia, and these include thyrotoxicosis, pheochromocytoma, Addison's disease, vasoactive intestinal polypeptide hormone producing tumor (VIPomas), familial benign hypocalciuric hypercalcemia, Jansen's disease, and Williams syndrome.

THYROTOXICOSIS

Mild hypercalcemia (<12 mg/dL, or 3 mmol/L)) frequently accompanies thyrotoxicosis, which leads to increased bone turnover and resorption. The hypercalcemia may respond to treatment with β-adrenergic blockers.

FAMILIAL BENIGN HYPOCALCIURIC HYPERCALCEMIA

Familial benign hypercalcemia (FBH), which is also referred to as familial hypocalciuric hypercalcemia (FHH), is an autosomal dominant disorder characterized by lifelong asymptomatic hypercalcemia in association with an inappropriately low urinary calcium excretion (i.e., calcium clearance–to–creatinine clearance ratio (CCR) <0.01), and normal circulating PTH concentrations in 80% of patients. Hypermagnesemia is also typically present. Although most patients with FBH are asymptomatic, chondrocalcinosis and acute pancreatitis have occasionally been observed. Patients with FHH have been misdiagnosed as having primary hyperparathyroidism because 20% of FHH patients may have elevated plasma PTH concentrations. In addition, 20% of FHH patients may have a CCR greater than 0.01 and therefore be indistinguishable from patients with primary hyperparathyroidism. Moreover, low CCRs are observed in patients with primary hyperparathyroidism who have vitamin D deficiency or renal insufficiency or are of African American origin. It is important to distinguish FHH patients from those with primary hyperparathyroidism because the hypercalcemia in FHH is generally benign and does not result in sequelae (see Table 245-4). Moreover, parathyroidectomy does not correct the hypercalcemia in FHH. Mutational analysis may help in identifying FHH patients from those with primary hyperparathyroidism.

FHH is genetically heterogenous, with three reported variants—FHH1, FHH2 and FHH3—whose loci are on chromosomes 3q21.1, 19p, and 19q13, respectively (see Table 245-1). FHH1 is due to heterozygous loss-of-function mutations of the CaSR, which is a GPCR that signals through Gαq and Gα11. The human CaSR, a 1078–amino acid cell surface GPCR encoded by the *CaSR* gene located on chromosome 3q21.1, is expressed in parathyroid cells, thyroid cells, and kidney (see Fig. 245-2). Approximately two thirds of FHH kindreds have unique heterozygous mutations of the CaSR, and expression studies of these mutations have demonstrated a loss of CaSR function whereby there is an increase in the calcium ion–dependent set point for PTH release from the parathyroid cell. FHH2 is due to loss-of-function mutations in the G-protein subunit α11 (Gα11) (see Fig. 245-2), which decrease the sensitivity of cells expressing the CaSR, probably by impairing the release of guanosine diphosphate. Such Gα11 loss-of-function mutations may occur in less than 5% of FHH patients.

FHH3 is due to loss-of-function mutations of the adaptor protein 2 (AP2) sigma subunit (AP2σ) (see Fig. 245-2). AP2 is a central component of clathrin-coated vesicles (CCVs) and is pivotal in clathrin-mediated endocytosis, which internalizes plasma membrane constituents such as GPCRs. AP2 is a heterotetramer of α, β, μ, and σ subunits and links clathrin to vesicle membranes and binds to tyrosine- and dileucine-based motifs of membrane-associated cargo proteins. The FHH3-associated AP2σ mutations, which all involve an Arg15 residue that forms key contacts with the dileucine-based motifs of CCV cargo proteins, result in a decreased sensitivity of CaSR-expressing cells to extracellular calcium and reduced CaSR endocytosis, probably through loss of interaction with a C-terminal CaSR dileucine-based motif, whose disruption also decreases intracellular signaling. Such AP2σ loss-of-function mutations occur in more than 5% of FHH patients. FHH1, FHH2, and FHH3 have similar clinical features, and thus genetic analysis is required to identify the relevant mutations.

AUTOIMMUNE HYPOCALCIURIC HYPERCALCEMIA

Some patients, who have the clinical features of FHH1, but not CaSR mutations, may have autoimmune hypocalciuric hypercalcemia (AHH). Such patients may have multiple clinical autoimmune manifestations, including antithyroid, antigliadin, or antiendomyseal antibodies. These patients were shown to have circulating antibodies to the extracellular domain of the CaSR, and these antibodies stimulated PTH release from dispersed human parathyroid cells in vitro, probably by inhibiting the activation of the CaSR by extracellular calcium. The effects of treatment with glucocorticoids have been variable, with the hypercalcemia responding in one patient but not in another. Thus, AHH is a condition of extracellular calcium-sensing that should be considered in FHH1 patients who do not have CaSR mutations.

NEONATAL SEVERE PRIMARY HYPERPARATHYROIDISM

Neonatal severe primary hyperparathyroidism (NSHPT) is defined as symptomatic hypercalcemia with skeletal manifestations of hyperparathyroidism in the first 6 months of life.[5] NSHPT children often present in the first few days or weeks of life with failure to thrive, dehydration, hypotonia, constipation, rib cage deformities, and multiple fractures due to bony undermineralization. Children with NSHPT often have life-threatening hypercalcemia and require urgent parathyroidectomy, which corrects the PTH-dependent hypercalcemia and bone demineralization. FBH or FHH is due to heterozygous inactivating mutations of the CaSR, and NSHPT is often associated with inactivating homozygous CaSR mutations when the children are from consanguineous parents with FHH1 (see Fig. 245-2). However, NSHPT has also been observed in children for whom one parent had clinically apparent FBH, and many other NSHPT patients appear to be sporadic; that is, both parents have normal serum calcium concentrations. In such NSHPT patients with heterozygous CaSR mutations, the mutant CaSR may exert a dominant negative action on the normal CaSR.

JANSEN'S DISEASE

Jansen's disease is an autosomal dominant disease that is characterized by short-limbed dwarfism due to metaphyseal chondrodysplasia and severe hypercalcemia and hypophosphatemia despite normal or undetectable serum levels of PTH. These abnormalities are associated with activating mutations of the PTH receptor (see Fig. 245-2), and thus this represents a PTH-independent activation of the PTH receptor (PTHR).

WILLIAMS SYNDROME

Williams syndrome is an autosomal dominant disorder characterized by supravalvular aortic stenosis, elfin-like facies, psychomotor retardation, and infantile hypercalcemia. The underlying abnormality causing hypercalcemia, which affects 5 to 50% of patients, remains unknown, but abnormal $1,25(OH)_2D_3$ metabolism and decreased calcitonin production have been implicated, although no abnormality has been consistently demonstrated. Hemizygosity due to a microdeletion of chromosome 7q11.23 involving the *ELASTIN* and *LIM-KINASE* genes, which may explain the respective cardiovascular and neurologic features, have been reported in Williams syndrome patients. However, the calcitonin receptor gene, located on chromosome 7q21 and close to the region deleted in Williams syndrome, was not involved in the deletion found in four patients with Williams syndrome, indicating that it is unlikely to be implicated in the hypercalcemia of such children. Another, as yet uncharacterized gene that is within this contiguously deleted region is likely to be involved to explain the abnormalities of calcium metabolism.

INFANTILE HYPERCALCEMIA

Infantile hypercalcemia is associated with failure to thrive and is characterized by severe hypercalcemia, hypercalciuria, and nephrocalcinosis and elevated circulating $1,25(OH)_2D$ concentrations. Some infants with this disorder have homozygous or compound heterozygous mutations of the gene encoding the 24-hydroxylase (CYP24A1) enzyme, which metabolizes the active $1,25(OH)_2D$ to the inactive $1,24,25(OH)_3D$ form (see Fig. 245-1).

Malignancy

Hypercalcemia may occur in 20 to 30% of patients with a malignancy, and this is usually due to increased bone resorption, which may either be directly

due to skeletal metastases or indirectly due to tumor production of a humoral factor that stimulates osteoclastic bone resorption (Chapter 179). The cancers that typically metastasize to produce lytic bone lesions are from the breast, lymphomas, or multiple myeloma (see Table 245-2). The associated osteolysis, mediated by recruitment and activation of osteoclasts, involves cytokines. Denosumab, a humanized neutralizing monoclonal antibody to RANKL, may be used to prevent the recruitment and activation of osteoclasts. The cancers that are typically associated with the humoral hypercalcemia of malignancy (HHM) are squamous carcinomas of the lung, esophagus, cervix, vulva, skin, head, or neck, but other types from the kidney, bladder, ovary, and breast may also occur. HHM accounts for up to 80% of patients with malignancy-associated hypercalcemia. The most common factor causing HHM is PTHrP, which can be measured in the serum by immunoassay. However, these assays are relatively insensitive, and the failure to detect serum PTHrP does not exclude the diagnosis of HHM. Patients with HHM generally have hypercalcemia associated with lower or undetectable serum PTH levels, marked hypercalcemia, and a reduced plasma $1,25(OH)_2D$ level. Therapy of HHM is aimed at (1) reducing the tumor load by surgery, radiotherapy, and/or chemotherapy; (2) reducing osteoclastic bone resorption by use of bisphosphonates (e.g., zoledronate) calcitonin or denosumab; and (3) increasing renal calcium clearance by a saline diuresis.

Granulomatous Disorders

Several granulomatous disorders are associated with hypercalcemia (see Table 245-2), and this is invariably associated with elevated circulating concentrations of $1,25(OH)_2D$, which is due to extrarenal synthesis. Sarcoidosis is the most frequently encountered granulomatous disorder associated with hypercalcemia; 10% of patients with sarcoidosis have hypercalcemia, and about half become hypercalciuric. The finding of raised serum angiotensin-converting enzyme (ACE) activity may help confirm the diagnosis. Glucocorticoids (e.g., 40 to 60 mg of prednisolone) decrease $1,25(OH)_2D$ production and restore the calcium concentration to normal. Failure to achieve normal serum calcium concentrations within 10 days of glucocorticoid therapy (e.g., hydrocortisone, 40 mg three times per day), the steroid suppression test, should suggest the coexistence of another cause for the hypercalcemia, such as primary hyperparathyroidism or malignancy.

Drugs

Several drugs (see Table 245-2) can cause hypercalcemia by different mechanisms. Compounds containing vitamins D and A are common and frequently associated with hypercalcemia. The use of thiazide diuretics is often associated with hypercalcemia. The hypercalcemia appears to be largely renal in origin because thiazides enhance distal renal tubular calcium reabsorption. Hypercalcemia reverses rapidly with discontinuation of the drug.

The milk-alkali syndrome was first described in the 1930s, generally in the context of ulcer treatment with large quantities of milk together with sodium bicarbonate. Today, the responsible agent is usually calcium carbonate, although consumption of large quantities of dairy products (milk, cheese, and yogurt) may still contribute. Classic features include moderate to severe hypercalcemia with alkalosis and renal impairment. The amount of calcium ingested by patients with this syndrome is usually 5 to 15 g/day. Treatment consists of (1) discontinuing the ingestion of the calcium containing compounds and antacids; (2) rehydration; and (3) saline diuresis.

HYPOCALCEMIA

DEFINITION

Hypocalcemia is defined as a serum calcium concentration below the lower limit of normal range, and this is usually an ionized serum calcium below 4.65 mg/dL (1.16 mmol/L) and a total serum calcium below 8.5 mg/dL (2.12 mmol/L). Mild hypocalcemia is defined as a total serum calcium of 8 to 8.5 mg/dL (2 to 2.12 mmol/L) and severe hypocalcemia as a total serum calcium below 7.6 mg/dL (1.9 mmol/L).

PATHOBIOLOGY

Hypocalcemia (see Table 245-6) can be classified by cause, according to whether serum PTH concentrations are low (i.e., hypoparathyroid disorders) or high (i.e., disorders associated with secondary hyperparathyroidism). Hypocalcemia is most commonly caused by hypoparathyroidism, a deficiency or abnormal metabolism of vitamin D, acute or chronic renal failure, or hypomagnesemia. Hypocalcemic diseases (see Table 245-6) may arise because of a destruction of the parathyroid glands, failure of parathyroid gland development, or reduced PTH secretion or PTH-mediated actions in target tissues. Thus, these diseases may be classified as being due to a deficiency of PTH, a defect in the PTH receptor (i.e., the PTH/PTHrP receptor), or insensitivity to PTH caused by defects downstream of the PTH/PTHrP receptor (see Fig. 245-2). The diseases may also be classified as being part of the hypoparathyroid disorders, of the CaSR abnormalities, or of the pseudohypoparathyroid disorders.

TABLE 245-7 HYPOCALCEMIC CLINICAL FEATURES OF NEUROMUSCULAR IRRITABILITY

Paresthesia, usually of fingers, toes, and circumoral regions
Tetany, carpopedal spasm, muscle cramps
Chvostek's sign*
Trousseau's sign†
Seizures of all types (i.e., focal or petit mal, grand mal, or syncope)
Prolonged QT interval on electrocardiogram
Laryngospasm
Bronchospasm

*Chvostek's sign is twitching of the circumoral muscles in response to gentle tapping of the facial nerve just anterior to the ear; it may be present in 10% of normal individuals.
†Trousseau's sign is carpal spasm elicited by inflation of a blood pressure cuff to 20 mm Hg above the patient's systolic blood pressure for 3 minute.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The clinical presentation of hypocalcemia ranges from an asymptomatic biochemical abnormality to a severe, life-threatening condition. In mild hypocalcemia, patients may be asymptomatic. Those with more severe and long-term hypocalcemia may develop acute symptoms of neuromuscular irritability (Table 245-7), ectopic calcification (e.g., in the basal ganglia, which may be associated with extrapyramidal neurological symptoms), subcapsular cataract, papilledema, and abnormal dentition. Investigations should be directed at confirming the presence of hypocalcemia and establishing the cause (Fig. 245-4).

In hypoparathyroidism, serum calcium is low, phosphate is high, and PTH is undetectable; renal function and concentrations of the 25-hydroxy and 1,25-dihydroxy metabolites of vitamin D are usually normal (see Fig. 245-4). The features of pseudohypoparathyroidism are similar to those of hypoparathyroidism except for PTH, which is markedly increased.[15] In chronic renal failure, which is the most common cause of hypocalcemia, phosphate is high, and alkaline phosphatase, creatinine, and PTH are elevated; 25-OH-D_3 is normal, and $1,25(OH)_2D_3$ is low (see Fig. 245-1). In vitamin D deficiency osteomalacia, serum calcium and phosphate are low, alkaline phosphatase and PTH are elevated, renal function is normal, and 25-OH-D_3 is low (see Fig. 245-4). The most frequent artifactual cause of hypocalcemia is hypoalbuminemia, such as occurs in liver disease or the nephrotic syndrome.

TREATMENT

Rx

Acute Hypocalcemia

The management of acute hypocalcemia depends on the severity of the hypocalcemia, the rapidity with which it developed, and the degree of neuromuscular irritability (see Table 245-7). Treatment should be given to symptomatic patients (e.g., with seizures or tetany) and asymptomatic patients with a serum calcium of less than 7.6 mg/dL (1.90 mmol/L) who are at high risk for developing complications. The preferred treatment for acute symptomatic hypocalcemia is calcium gluconate, 10 mL 10% w/v (2.20 mmol of calcium) intravenous, diluted in 50 mL of 5% dextrose or 0.9% sodium chloride and given by slow injection (>5 minutes); this can be repeated as required to control symptoms. Serum calcium concentrations should be assessed regularly. Continuing hypocalcemia may be managed acutely by administration of a calcium gluconate infusion; for example, dilute 10 ampules of calcium gluconate, 10 mL 10% w/v (22 mmol of calcium), in 1 liter of 5% dextrose or 0.9% sodium chloride, start infusion at 50 mL/hour, and titrate to maintain serum calcium concentrations in the normal range. Generally, 1.2 to 1.6 mg/kg (0.3 to 0.4 mmol/kg) of elemental calcium infused over 4 to 6 hours increases serum calcium by 2 to 3 mg/dL (0.5 to 0.75 mmol/L). If hypocalcemia is likely to persist, oral vitamin D therapy (see later) should also be administered. In hypocalcemic patients who are also hypomagnesemic, the hypomagnesemia must be corrected before the hypocalcemia will resolve. This may occur in the postparathyroidectomy period or in patients with severe malabsorption, for example, those with established celiac disease (Chapter 140).

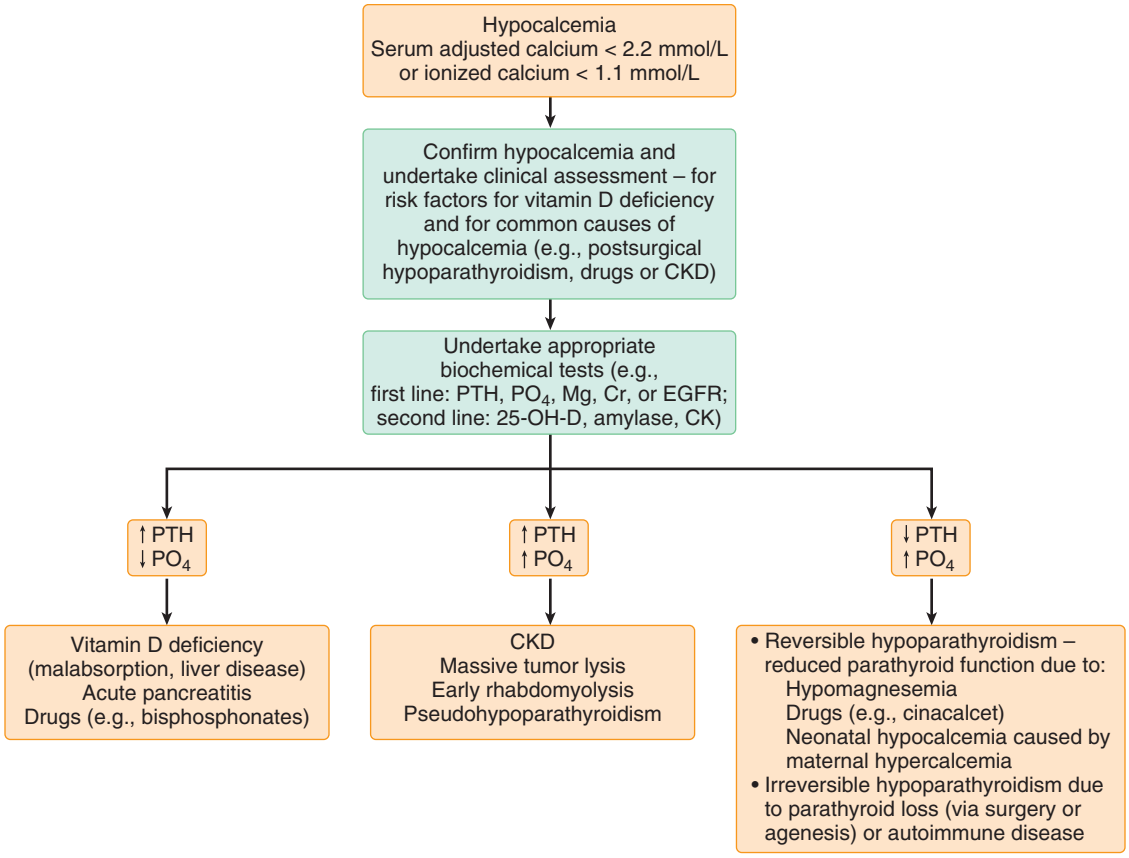

FIGURE 245-4. Clinical approach to investigation of causes of hypocalcemia. 25-OH-D = 25-hydroxyvitamin D; CK = creatinine kinase ; CKD = chronic kidney disease; Cr = creatinine; EGFR = estimated glomerular filtration rate; PTH = parathyroid hormone.

Chronic Hypocalcemia

The two main agents available for the treatment of chronic (long-term) hypocalcemia are supplemental calcium, about 10 to 20 mmol calcium every 6 to 12 hours, and vitamin D preparations. Patients with hypoparathyroidism seldom require calcium supplements after the early stages of stabilization with vitamin D. A variety of vitamin D preparations have been used. These include vitamin D_3 (cholecalciferol) or vitamin D_2 (ergocalciferol), 10,000 to 50,000 units (0.25 to 1.25 mg/day); dihydrotachysterol (now seldom used), 0.25 to 1.25 mg/day; alfacalcidol (1α-hydroxycholecalciferol), 0.25 to 1 μg/day; and calcitriol (1,25-dihydroxy cholecalciferol), 0.25 to 2 μg/day. In children, these preparations are prescribed in doses based on body weight. Cholecalciferol and ergocalciferol are the least expensive preparations but have the longest durations of action and may result in prolonged toxicity; however, they are the preparations of choice for treating hypocalcemia associated with vitamin D deficiency (see Table 245-6). The other preparations, which do not require renal 1α-hydroxylation, have the advantage of shorter half-lives and thereby minimize the risk for prolonged toxicity. For treatment of hypocalcemia due to hypoparathyroidism or chronic renal failure, calcitriol is the drug of choice because it is the active metabolite and, unlike alfacalcidol, does not require hepatic 25-hydroxylation. Close monitoring (at about 1- to 2-week intervals) of the patient's serum and urine calcium concentrations are required initially, and 3- to 6-month intervals are appropriate once stabilization is achieved. The aim is to avoid hypercalcemia, hypercalciuria, nephrolithiasis, and renal failure. It should be noted that hypercalciuria may occur in the absence of hypercalcemia. The use of PTH (1-84) in hypoparathyroid patients has been reported to be associated with improvement in the biochemical and skeletal indices, as well as in mental and physical health.[16]

HYPOPARATHYROIDISM

DEFINITION

Hypoparathyroidism is characterized by hypocalcemia and hyperphosphatemia, which are the result of a deficiency in PTH secretion or action.[17] Serum concentrations of immunoreactive PTH are low or undetectable, and the concentrations of $1,25(OH)_2D_3$ are usually in the low-normal to low range, but alkaline phosphatase activity is unchanged (see Fig. 245-4). The daily urinary excretion of calcium is reduced, although the fractional excretion of calcium is increased. Nephrogenous cyclic adenosine monophosphate (cAMP) excretion is low, and renal tubular reabsorption of phosphate is elevated. Urinary cAMP, plasma cAMP, and urinary phosphate excretion increase markedly after administration of exogenous bioactive PTH (Chase-Aurbach and Ellsworth-Howard tests).

PATHOBIOLOGY

Hypoparathyroidism may result from agenesis (e.g., DiGeorge syndrome) or destruction of the parathyroid glands (e.g., following neck surgery, in autoimmune diseases), from reduced secretion of PTH (e.g., neonatal hypocalcemia or hypomagnesemia), or from resistance to PTH (which may occur as a primary disorder (e.g., pseudohypoparathyroidism or secondary to hypomagnesemia) (see Table 245-6). In addition, hypoparathyroidism may occur as an inherited syndromic disorder (see Table 245-1) that may either be part of a complex congenital defect (e.g., DiGeorge syndrome) or part of a polyglandular autoimmune disorder (see Table 245-6 and Fig. 245-2). Hypoparathyroidism may also occur as a nonsyndromic solitary endocrinopathy, which has been referred to as isolated or idiopathic hypoparathyroidism. Familial occurrences of isolated hypoparathyroidism with autosomal dominant, autosomal recessive, and X-linked recessive inheritances have been established.

Isolated Hypoparathyroidism

Isolated hypoparathyroidism may either be inherited or acquired by damage to the parathyroids at surgery, by infiltrating metastases, or by systemic disease (see Table 245-6).

ACQUIRED FORMS OF HYPOPARATHYROIDISM

Hypoparathyroidism may occur after neck surgery, after irradiation, or because of infiltration by metastases or systemic disease, for example, hemochromatosis, amyloidosis, sarcoidosis, Wilson's disease, or thalassemia (see Table 245-6). Surgical damage to the parathyroids occurs most commonly after a radical neck dissection, such as for laryngeal or esophageal carcinoma treatment, after a total thyroid resection, or after repeated parathyroidectomies for polyglandular disease (e.g., in MEN 1 or MEN 2, discussed

previously). Hypocalcemic symptoms begin 12 to 24 hours after surgery and may need treatment with oral or intravenous calcium. Parathyroid function often returns, but persistent hypocalcemia requires treatment with vitamin D preparations.

Neonatal hypoparathyroidism resulting in hypocalcemia may occur in the infant of a mother with hypercalcemia caused by primary hyperparathyroidism (see Table 245-6). Maternal hypercalcemia results in increased calcium delivery to the fetus, and this fetal hypercalcemia suppresses fetal PTH secretion. Postpartum, the infant's suppressed parathyroids are unable to maintain normocalcemia. The disorder is usually self-limited, but occasionally therapy may be required. In addition, the feeding of cow's milk, which has a high phosphate content, to infants may result in hypocalcemia in some children.

Functional hypoparathyroidism may result from severe hypomagnesemia (<0.4 mmol/L), which may be due to a severe intestinal malabsorption disorder (e.g., Crohn's disease) or a renal tubular disorder (see Table 245-6). It is associated with hypoparathyroidism because magnesium is required for the release of PTH from the parathyroid gland and also for PTH action through adenyl cyclase. Magnesium chloride, 35 to 50 mmol intravenously in 1 liter of 5% glucose or other isotonic solution given over 12 to 24 hours, may be repeatedly required to restore normomagnesemia.

INHERITED HYPOPARATHYROIDISM

Patients with inherited forms of hypoparathyroidism may develop hypocalcemic seizures in the neonatal or infantile periods and require lifelong treatment with oral vitamin D preparations, such as calcitriol. Autosomal dominant, autosomal recessive, and X-linked recessive inheritances for hypoparathyroidism have been observed (see Table 245-1). Some of the autosomal forms are due to mutations of the *PTH* gene, the CaSR (see later), the Gα11 subunit (see later), and the transcriptional factor *GCMB* (glial cells missing B). The X-linked forms likely alter regulation of SOX3 (see Fig. 245-4).

Autosomal Dominant Hypocalcemia Type 1 and Type 2

Autosomal dominant hypocalcemia type 1 (ADH1) is characterized by lifelong mild or severe hypocalcemia in association with normal serum PTH concentrations in about 40% of patients or low serum PTH concentrations in about 60% of patients. Serum phosphate and magnesium concentrations may be elevated or low, respectively. Approximately 50% of ADH1 patients have asymptomatic hypocalcemia, and the remaining 50% may experience paresthesia, muscle cramps, carpopedal spasms, and seizures, which may be associated with a febrile illness. In addition, about 10% of ADH1 patients may have absolute hypercalciuria, which may be associated with nephrocalcinosis and kidney stones in 35% of patients. Vitamin D preparations and calcium supplementation to correct the hypocalcemia may worsen the hypercalciuria and lead to renal impairment. Basal ganglia or ectopic calcification may be found in more than 35% of patients. About 20% of ADH1 patients do not have a previously reported family history because they have de novo mutations. ADH1 is due to gain-of-function mutations of the CaSR (see Table 245-1 and Fig. 245-2). ADH2 is due to gain-of-function mutations of the Gα11 subunit, and ADH2 patients appear to have clinical features that are similar to those in ADH1 patients.

Complex Syndromes Associated with Hypoparathyroidism

Hypoparathyroidism may occur as part of a complex syndrome that may be associated with either a congenital developmental anomaly or an autoimmune syndrome. The congenital developmental anomalies associated with hypoparathyroidism, which occurs in 1 in 4000 live births, include DiGeorge syndrome, hypoparathyroidism, deafness, and renal anomalies (HDR) syndrome, Kenney-Caffey and Barakat syndromes, and also syndromes associated with either lymphedema or dysmorphic features and growth failure (see Table 245-1 and Fig. 245-2).

POLYGLANDULAR AUTOIMMUNE HYPOPARATHYROIDISM

Polyglandular autoimmune hypoparathyroidism comprises hypoparathyroidism, Addison's disease, candidiasis, and two or three of the following: type 1 diabetes mellitus, primary hypogonadism, autoimmune thyroid disease, pernicious anemia, chronic active hepatitis, steatorrhoea (malabsorption), alopecia (totalis or areata), and vitiligo. The disorder has also been referred to as either the autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED) syndrome or the polyglandular autoimmune type 1 syndrome (see Table 245-1). Antibodies directed against the adrenal, thyroid, and parathyroid glands are detected in the sera of some patients. The polyglandular autoimmune type 2 syndrome is characterized by adrenal insufficiency, type 1 diabetes mellitus, and thyroid disease and does not involve hypoparathyroidism. APECED, which has an autosomal recessive inheritance, has a high incidence in Finland and among Iranian Jews. The *APECED* gene, which has been located to chromosome 21q22.3, encodes a 545–amino acid protein that contains motifs associated with a transcriptional factor and includes two zinc-finger motifs, a proline-rich region, and three LXXLL motifs. The gene is referred to as *AIRE* (autoimmune regulator) (see Fig. 245-2). Four *AIRE* mutations are commonly found in APECED families, and these likely abolish the E3 ubiquitin ligase activity of the AIRE1 protein. AIRE1 has been shown to regulate the elimination of organ-specific T cells in the thymus, and APECED is likely to be caused by a failure of this specialized mechanism for deleting forbidden T cells and establishing immunologic tolerance.

AUTOIMMUNE ACQUIRED HYPOPARATHYROIDISM

Twenty percent of patients who had acquired hypoparathyroidism (AH) in association with autoimmune hypothyroidism were found to have autoantibodies to the extracellular domain of the CaSR (see Table 245-1 and Fig. 245-2). The CaSR autoantibodies did not persist for long; 72% of patients who had AH for less than 5 years had detectable CaSR autoantibodies, whereas only 14% of patients with AH for more than 5 years had such autoantibodies. The majority of the patients who had CaSR autoantibodies were female, a finding that is similar to that found in other autoantibody-mediated diseases. Indeed, a few acquired hypoparathyroidism patients have also had features of autoimmune polyglandular syndrome type 1. The epitopes for the anti-CaSR antibodies were localized to the N terminal of the extracellular domain of the receptor. These findings establish that the CaSR is an autoantigen in acquired hypoparathyroidism.

DIGEORGE SYNDROME

Patients with the DiGeorge syndrome suffer from neonatal hypoparathyroidism, T-cell immunodeficiency, congenital heart defects, and deformities of the ear, nose, and mouth (e.g., cleft lip and/or palate). Children with DiGeorge syndrome often die from infections related to the immunodeficiency. The disorder arises from a congenital failure in the development of the derivatives of the third and fourth pharyngeal pouches with resulting absence or hypoplasia of the parathyroids and thymus. Most cases are sporadic, but an autosomal dominant inheritance of DiGeorge syndrome has been observed, and an association between the syndrome and an unbalanced translocation and deletions involving chromosome 22q11.2 has also been reported (see Table 245-1). In some patients, deletions of another locus on chromosome 10p13-p14 have been observed in association with DiGeorge syndrome, and this is referred to as DGS2, whereas patients with the 22q11.2 deletions are referred to as having DGS1. Studies of the DGS1 deleted region on chromosome 22q11.2 have revealed four genes (*RNEX40*, *NEX2.2-NEX3*, *UDFIL*, and *TBX1*) to be involved. However, point mutations in DGS1 patients have been detected only in the *TBX1* gene, and *TBX1* is now considered to be the gene causing DGS1 (see Table 245-1 and Fig. 245-2). *TBX1* encodes a DNA-binding transcriptional factor, of the T-BOX family, that is known to have an important role in vertebrate and invertebrate organogenesis and pattern formation. The *TBX1* gene is deleted in approximately 96% of all DGS1 patients, and some of those without deletions have been shown to harbor mutations of *TBX1*.

HYPOPARATHYROIDISM, DEAFNESS, AND RENAL ANOMALIES SYNDROME

HDR syndrome is an autosomal dominant disorder in which patients often have asymptomatic hypocalcemia with undetectable or inappropriately normal serum concentrations of PTH. Bilateral, symmetrical, sensorineural deafness involving all frequencies occurs, and the renal abnormalities consist mainly of bilateral cysts that compress the glomeruli and tubules and lead to renal impairment. Cytogenetic abnormalities involving chromosome 10p14-10pter were identified in HDR patients, and HDR patients have deletions or mutations of the zinc-finger transcription factor *GATA3* (see Table 245-1 and Fig. 245-2).

MITOCHONDRIAL DISORDERS ASSOCIATED WITH HYPOPARATHYROIDISM

Hypoparathyroidism has been reported to occur in three disorders associated with mitochondrial dysfunction: Kearns-Sayre syndrome, MELAS syndrome, and a mitochondrial trifunctional protein deficiency syndrome (MTPDS) (see Table 245-1 and Fig. 245-2). Kearns-Sayre syndrome is characterized by progressive external ophthalmoplegia and pigmentary

TABLE 245-8 CLINICAL, BIOCHEMICAL, AND GENETIC FEATURES OF HYPOPARATHYROID AND PSEUDOHYPOPARATHYROID DISORDERS

		PSEUDOHYPOPARATHYROIDISM				
	HYPOPARATHYROIDISM	*PHP 1a*	*PPHP*	*PHP 1b*	*PHP 1c*	*PHP 2*
AHO manifestations	No	Yes	Yes	No	Yes	No
Serum calcium	↓	↓	N	↓	↓	↓
Serum PO_4	↑	↑	N	↑	↑	↑
Serum PTH	↓	↑	N	↑	↑	↑
Response to PTH:						
Urinary cAMP* (Chase-Aurbach test)	↑	↓	↑	↓	↓	↑
Urinary PO_4 (Ellsworth-Howard test)	↑	↓	↑	↓	↓	↓
Gsα activity	N	↓	↓	N	N	N
Inheritance	AD, AR, X	AD	AD	AD	AD	Sporadic
Molecular defect	PTH, CaSR, GATA3, Gcm2, others	GNAS1	GNAS1	GNAS1†	?Adenyl cyclase	?cAMP targets
Other hormonal resistance	No	Yes	No	No	Yes	No

↓ = decreased; ↑ = increased; ? = presumed, but not proved; AD = autosomal dominant; AHO = Albright's hereditary osteodystrophy; AR = autosomal recessive; N = normal; PHP = pseudoparathyroidism; PPHP = pseudopseudoparathyroidism; PTH = parathyroid hormone; X = X-linked.
*Plasma cyclic adenosine monophosphate (cAMP) responses are similar to those of urinary cAMP.
†Involves deletions that are located upstream of GNAS1.

retinopathy before the age of 20 years and is often associated with heart block or cardiomyopathy. MELAS syndrome consists of a childhood onset of mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes. In addition, varying degrees of proximal myopathy can be seen in both conditions. Both Kearns-Sayre and MELAS syndromes have been reported to occur with type 1 diabetes mellitus and hypoparathyroidism, and mitochondrial gene abnormalities have been identified in some patients. Mitochondrial trifunctional protein deficiency is a disorder of fatty acid oxidation that is associated with peripheral neuropathy, pigmentary retinopathy, and acute fatty liver degeneration in pregnant women who carry an affected fetus. Hypoparathyroidism has been observed in one patient with trifunctional protein deficiency.

KENNEY-CAFFEY, SANJAD-SAKATI, AND KIRK-RICHARDSON SYNDROMES

Hypoparathyroidism has been reported to occur in more than 50% of patients with Kenney-Caffey syndrome, which is associated with short stature, osteosclerosis, cortical thickening of the long bones, delayed closure of the anterior fontanel, basal ganglia calcification, nanophthalmos, and hyperopia. Parathyroid tissue could not be found in a detailed postmortem examination of one patient, and this suggests that hypoparathyroidism may be due to an embryologic defect of parathyroid development. In Kirk-Richardson and Sanjad-Sakati syndromes, which are similar, hypoparathyroidism is associated with severe growth failure and dysmorphic features. This has been reported in patients of Middle Eastern origin whose parents were consanguineous, indicating that these are autosomal recessive disorders. Molecular genetic investigations have identified mutations of the tubulin-specific chaperone (*TBCE*), located on chromosome 1q42-q43, to be associated with the Kenney-Caffey type 1 and Sanjad-Sakati syndromes (see Table 245-1 and Fig. 245-2). *TBCE* encodes one of several chaperone proteins required for the proper folding of α-tubulin subunits and the formation of αβ-tubulin heterodimers. Kenney-Caffey type 2 syndrome is due to mutations of a member of the family with sequence similarity 111 of gene (*FAM111A*), located on chromosome 11q12.1.

PSEUDOHYPOPARATHYROIDISM

Patients with pseudohypoparathyroidism (PHP), which may be inherited as an autosomal dominant disorder, are characterized by hypocalcemia and hyperphosphatemia due to PTH resistance rather than PTH deficiency (see Table 245-6). Five variants are recognized on the basis of biochemical and somatic features (Table 245-8), and three of these—PHP type 1a (PHP 1a), PHP type 1b (PHP 1b), and pseudopseudohypoparathyroidism (PPHP)—will be reviewed in further detail. Patients with PHP 1a exhibit PTH resistance (hypocalcemia, hyperphosphatemia, elevated serum PTH, and an absence of an increase in serum and urinary cAMP and urinary phosphate following intravenous human PTH infusion), together with the features of Albright's hereditary osteodystrophy (AHO), which includes short stature, obesity, subcutaneous calcification, mental retardation, round faces, dental hypoplasia, and brachydactyly (i.e., shortening of the metacarpals, particularly the third, fourth, and fifth). In addition to brachydactyly, other skeletal abnormalities of the long bones and shortening of the metatarsals may occur. Patients with PHP 1b exhibit PTH resistance only and do not have the somatic features of AHO, whereas patients with PPHP exhibit the somatic features of AHO in the absence of PTH resistance. The absence of a normal rise in urinary excretion of cAMP after an infusion of PTH in PHP 1a indicated a defect at some site of the PTH receptor–adenyl cyclase system (see Fig. 245-2). This receptor system is regulated by at least two G proteins, one of which stimulates (Gsα) and another of which inhibits (Giα) the activity of the membrane-bound enzyme that catalyses the formation of the intracellular second messenger cAMP. Patients with PHP 1a may also show resistance to other hormones, such as thyroid-stimulating hormone, follicle-stimulating hormone, and luteinizing hormone, that act through GPCRs. Inactivating mutations of the Gsα gene (referred to as *GNAS1*), which is located on chromosome 20q13.2, have been identified in PHP 1a and PPHP patients (see Table 245-1 and Fig. 245-2). However, *GNAS1* mutations do not fully explain the PHP 1a or PPHP phenotypes, and studies of PHP 1a and PPHP that occurred within the same kindred revealed that the hormonal resistance is parentally imprinted. Thus, PHP 1a occurs in a child only when the mutation is inherited from a mother affected with either PHP 1a or PPHP; and PPHP occurs in a child only when the mutation is inherited from a father affected with either PHP 1a or PPHP. PHP 1b is due to deletions that are located upstream of the *GNAS1* gene. Moreover, in affected individuals, the deletion involved the maternal allele, whereas its occurrence on the paternal allele resulted in unaffected healthy carriers. This is consistent with parental imprinting of the *GNAS1* abnormality causing PHP 1b.

Grade A References

A1. Peacock M, Bilezikian JP, Bolognese MA, et al. Cinacalcet HCl reduces hypercalcemia in primary hyperparathyroidism across a wide spectrum of disease severity. *J Clin Endocrinol Metab*. 2011;96:E9-E18.

A2. Rolighed L, Rejnmark L, Sikjaer T, et al. Vitamin D treatment in primary hyperparathyroidism: a randomized placebo controlled trial. *J Clin Endocrinol Metab*. 2014;99:1072-1080.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

Additional Familial Syndromes

Single familial syndromes in which hypoparathyroidism is a component have been reported (see Table 245-1). Thus, an association of hypoparathyroidism, renal insufficiency, and developmental delay has been reported in an Asian family in whom autosomal recessive inheritance of the disorder was established. The occurrence of hypoparathyroidism, nerve deafness, and a steroid-resistant nephrosis leading to renal failure, which has been referred to as Barakat syndrome, has been reported in four brothers from one family, and an association of hypoparathyroidism with congenital lymphedema, nephropathy, mitral valve prolapse, and brachytelephalangy has been observed in two brothers from another family. Molecular genetic studies have not been reported from these two families as of this time.

BLOMSTRAND'S DISEASE

Blomstrand's chondrodysplasia is an autosomal recessive disorder characterized by early lethality, dramatically advanced bone maturation, and accelerated chondrocyte differentiation. Affected infants, who usually have consanguineous unaffected parents, develop pronounced hyperdensity of the entire skeleton with markedly advanced ossification, which results in extremely short and poorly modeled long bones. Mutations of the PTH/PTHrP receptor that impair its function are associated with Blomstrand's disease (see Table 245-1 and Fig. 245-2). Thus, it seems likely that affected infants will, in addition to the skeletal defects, have abnormalities in other organs, including secondary hyperplasia of the parathyroid glands, presumably due to hypocalcemia.

GENERAL REFERENCES

1. Hannan FM, Thakker RV. Investigating hypocalcaemia. *BMJ*. 2013;9:f2213.
2. Ferrè S, Hoenderop JG, Bindels RJ. Sensing mechanisms involved in Ca^{2+} and Mg^{2+} homeostasis. *Kidney Int*. 2012;82:1157-1166.
3. Nesbit MA, Hannan FM, Howles SA, et al. Mutations affecting G-protein subunit alpha11 in hypercalcemia and hypocalcemia. *N Engl J Med*. 2013;368:2476-2486.
4. Nesbit MA, Hannan FM, Howles SA, et al. Mutations in AP2S1 cause familial hypocalciuric hypercalcemia type 3. *Nat Genet*. 2013;45:93-97.
5. Wysolmerski JJ. Parathyroid hormone-related protein: an update. *J Clin Endocrinol Metab*. 2012;97:2947-2956.
6. Marcocci C, Cetani F. Primary hyperparathyroidism. *N Engl J Med*. 2011;365:2389-2397.
7. Thakker RV, Newey PJ, Walls GV, et al. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab*. 2012;97:2990-3011.
8. Starker LF, Akerstrom T, Long WD, et al. Frequent germ-line mutations of the MEN1, CASR, and HRPT2/CDC73 genes in young patients with clinically non-familial primary hyperparathyroidism. *Horm Cancer*. 2012;3:44-51.
9. Costa-Guda J, Soong CP, Parekh VI, et al. Germline and somatic mutations in cyclin-dependent kinase inhibitor genes CDKN1A, CDKN2B, and CDKN2C in sporadic parathyroid adenomas. *Horm Cancer*. 2013;4:301-307.
10. Pallan S, Rahman MO, Khan AA. Diagnosis and management of primary hyperparathyroidism. *BMJ*. 2012;344:e1013.
11. Eastell R, Brandi ML, Costa AG, et al. Diagnosis of asymptomatic primary hyperparathyroidism: proceedings of the fourth international workshop. *J Clin Endocrinol Metab*. 2014;99:3570-3579.
12. Eldeiry LS, Ruan DT, Brown EM, et al. Primary hyperparathyroidism and familial hypocalciuric hypercalcemia: relationships and clinical implications. *Endocr Pract*. 2012;18:412-417.
13. Henry D, Vadhan-Raj S, Hirsh V, et al. Delaying skeletal-related events in a randomized phase 3 study of denosumab versus zoledronic acid in patients with advanced cancer: an analysis of data from patients with solid tumors. *Support Care Cancer*. 2014;22:679-687.
14. Augustine MM, Bravo PE, Zeiger MA. Treatment of primary hyperparathyroidism. *Endocr Pract*. 2011;17:75-82.
15. Mantovani G. Clinical review: pseudohypoparathyroidism: diagnosis and treatment. *J Clin Endocrinol Metab*. 2011;96:3020-3030.
16. Cusano NE, Rubin MR, McMahon DJ, et al. PTH(1-84) is associated with improved quality of life in hypoparathyroidism through 5 years of therapy. *J Clin Endocrinol Metab*. 2014;99:3694-3699.
17. De Sanctis V, Soliman A, Fiscina B. Hypoparathyroidism: from diagnosis to treatment. *Curr Opin Endocrinol Diabetes Obes*. 2012;19:435-442.

REVIEW QUESTIONS

1. Which one of the following statements is true for the actions of parathyroid hormone (PTH)?

A. Acts directly on renal, bone, and intestinal cells
B. Acts directly on renal cells and indirectly on bone and intestinal cells
C. Acts directly on renal and bone cells and indirectly on intestinal cells
D. Has no direct actions on renal, bone, and intestinal cells
E. Acts via a tyrosine kinase receptor

Answer: C PTH shares a receptor with PTH-related peptide (PTHrP), and this receptor is a member of the G protein–coupled receptor (GPCR) family. PTH/PTHrP receptors are expressed by kidney and bone cells, and PTH acts directly on renal and bone cells. PTH acts indirectly on intestinal cells to enhance calcium absorption as follows: PTH stimulates renal 1α-hydroxylase activity to promote synthesis of the active form of vitamin D (1,25-dihydroxyvitamin D), which then acts on the intestine to enhance calcium absorption.

2. Which of the following is not a cause of hypercalcemia?

A. Granulomatous disorders
B. PTH receptor loss-of-function mutation
C. Autoimmunity
D. Renal 1α-hydroxylase overactivity
E. Myeloma

Answer: B PTH receptor loss-of-function mutations result in Blomstrand's chondrodysplasia, which is not associated with hypercalcemia. Instead, PTH receptor gain-of-function mutations that lead to Jansen's metaphyseal chondroplasia are associated with severe hypercalcemia. Granulomatous disorders, autoimmunity (e.g., autoantibodies to the calcium-sensing receptor), renal 1α-hydroxylase overactivity, and myeloma may all be associated with hypercalcemia.

3. Which of the following statements is not included as a recommendation for parathyroid surgery in asymptomatic primary hyperparathyroid patients?

A. Serum calcium >10.mg/dL (0.25 mmol/L) above upper limit of normal
B. Reduction in bone mineral density by T score <–2.5
C. Reduction in creatinine clearance below 60 mL/min
D. Marked hypercalciuria (>400 mg per 24 hr, >9 mmol/L per 24 hr))
E. Age <50 years

Answer: D Marked hypercalciuria (>400 mg per 24 hr, >9 mmol/L per 24 hr) was included as a recommendation for parathyroid surgery in asymptomatic primary hyperparathyroidism by the Second International Conference (2002), but not by the Third International Conference (2008). However, some physicians still regard marked hypercalciuria as an indication for parathyroid surgery.

4. Which of the following statements regarding parathyroid tumors in multiple endocrine neoplasia (MEN) syndrome is *false*?

A. In MEN 1, parathyroid tumors occur with equal frequency in men and women.
B. In MEN 2b and MEN 3, parathyroid tumors are a rare occurrence.
C. In MEN 1, minimally invasive surgery is not a suitable approach because of the high occurrence of multiple parathyroid tumors.
D. Patients with MEN 2a have a higher risk for developing parathyroid carcinomas.
E. Parathyroid tumors occur in more than 95% of MEN 1 patients.

Answer: D MEN 1 is an autosomal dominant disorder, and this generally affects men and women equally. Parathyroid tumors are found in more than 95% of patients with MEN 1, who usually have multiple parathyroid tumors, and hence minimally invasive surgery is not recommended. Parathyroid tumors rarely occur in patients with MEN 2b or MEN 3, and about 20% of MEN 2a patients will usually have parathyroid hyperplasia and not parathyroid carcinoma. However, patients with hyperparathyroidism with jaw tumors are at high risk for developing parathyroid carcinomas.

5. Which of the following is not a cause of hypocalcemia?

A. Vitamin D–resistance disorders
B. Secondary hyperparathyroidism
C. Adrenal insufficiency
D. Autoimmunity
E. Acute pancreatitis

Answer: C Acute adrenal insufficiency is associated with hypercalcemia, not hypocalcemia. Hypocalcemia typically occurs in patients with secondary hyperparathyroidism and in those with vitamin D–resistance disorders, and it may be found in patients with acute pancreatitis. In addition, autoimmune destruction of the parathyroids, resulting in hypoparathyroidism, or autoantibodies to the calcium-sensing receptor may be associated with hypocalcemia.

246 MEDULLARY THYROID CARCINOMA

SAMUEL A. WELLS, JR.

MEDULLARY THYROID CARCINOMA

DEFINITION

Medullary thyroid carcinoma (MTC) is an uncommon cancer that arises from the neural crest—derived C cells of the thyroid gland.

EPIDEMIOLOGY

MTC accounts for 5% of thyroid cancers, and there will be approximately 3000 new cases in the United States in 2015. MTC occurs either sporadically (75% of cases) or as part of the multiple endocrine neoplasia (MEN) syndromes, MEN 2A (Online Mendelian Inheritance in Man [OMIM] #171400) or MEN 2B (OMIM #162300), or the related syndrome familial medullary thyroid carcinoma (FMTC) (OMIM #155240).[1] The incidence of MEN 2A and FMTC combined is approximately 1 in 100,000 live births, whereas the incidence of MEN 2B is approximately 1 in 2,000,000 live births.

PATHOBIOLOGY

As the lateral thyroid complex closes during embryogenesis, the C cells are incorporated within the middle and upper portions of the thyroid lobes. Because of its anatomic location, MTC is classified as a thyroid tumor; however, considering its origin from the neural crest rather than the thyroid follicular cells, it is a neuroendocrine tumor. Sporadic MTC occurs as a solitary tumor in one thyroid lobe, whereas hereditary MTC develops in both thyroid lobes and is multicentric. In patients with hereditary MTC, the first manifestation of a C-cell disorder is C-cell hyperplasia (CCH) that progresses over time to microinvasive MTC and then to invasive MTC.[2]

The C-cell mass is much greater in the thyroid glands of men than in women, which accounts for the higher serum calcitonin levels seen in men compared with women.

The C cells have diverse biosynthetic activity and secrete calcitonin (CTN) and carcinoembryonic antigen (CEA), which are excellent serum markers for the presence of a C-cell disorder. CTN was once thought to be important in calcium homeostasis; however, its physiologic importance has been called into question. The *RET* protooncogene encodes a single-pass transmembrane receptor of the tyrosine kinase family of proteins. At several stages of development, it is expressed in cells derived from the branchial arches (parathyroids), the neural crest (brain, parasympathetic and sympathetic ganglia, thyroid C cells, adrenal medulla, and enteric ganglia), and the urogenital system. Activating, germline point mutations in *RET* are present in virtually all hereditary MTCs, and somatic *RET* mutations are present in approximately half of sporadic MTCs.[3,4] Recently, it was discovered that somatic mutations in *HRAS*, *KRAS*, and rarely *NRAS* are present in sporadic MTCs and are almost always mutually exclusive with the presence of somatic *RET* mutations.[5]

Approximately 75 *RET* mutations have been reported in association with MEN 2A, MEN 2B, and FMTC. The mutations for MEN 2A and FMTC are located in exons 5, 8, 10, 11, 13, 14, 15, and 16. The mutations for MEN 2A are mostly located in the extracellular, cysteine-rich region of exon 10 (including codons 609, 611, 618, and 620) and exon 11 (including codons 630 and 634). Approximately 85% of the mutations associated with MEN 2A involve *RET* codon 634, about half of which are *C634R RET* mutations. The *RET* mutations in MEN 2B cause constitutive activation, which alters substrate specificity, presumably owing to a conformational change in the binding pocket of the kinase. Approximately 95% of mutations causing MEN 2B are in codon M918T, and 5% are in codon A883F. Rare cases of MEN 2B are caused by double somatic *RET* mutations involving codon V804M and either codon Y806C, S904C, or E805K.

In 50% of patients with MEN 2B and 10% of patients with MEN 2A and FMTC, the disease arises de novo. In such founder cases, the de novo mutation almost always derives from the paternal allele.[6]

CLINICAL MANIFESTATIONS

The peak incidence of sporadic MTC is in the fifth decade of life, and most patients present with a solitary thyroid nodule and lymph node metastases. Clinically, the tumors are more aggressive than papillary thyroid carcinoma and follicular thyroid carcinoma but less aggressive than anaplastic thyroid carcinoma (Chapter 226). The 10-year survival rate is 75%. In the expectation of detecting MTC at any early stage, clinicians in Europe evaluate serum CTN levels in patients with thyroid nodules who have no history of hereditary MTC. The detection rate of MTC is less than 0.5%, however, and clinicians in the United States have not adopted this practice. The clinical manifestations of MEN 2A, MEN 2B, and FMTC are listed in Table 246-1.

TABLE 246-1 CLINICAL MANIFESTATIONS OF MULTIPLE ENDOCRINE NEOPLASIA 2A, 2B, AND FAMILIAL MEDULLARY THYROID CARCINOMA

MULTIPLE ENDOCRINE NEOPLASIA (MEN) 2A
Medullary thyroid carcinoma (~100%) Pheochromocytoma (incidence of 50% in families with a *RET* codon 634 germline mutation but less in families with other *RET* codon mutations) Hyperparathyroidism (incidence of 30% in families with a *RET* codon 634 germline mutation but less in families with other *RET* codon mutations.)
VARIANTS OF MEN 2A
MEN 2A with cutaneous lichen amyloidosis (almost always associated with a *RET* codon 634 germline mutation.) MEN 2A with Hirschsprung disease (most common in families with *RET* germline mutation most commonly involving codon 620)
FAMILIAL MEDULLARY THYROID CARCINOMA (FMTC)
Since the original description of this syndrome, there has been confusion about the designation FMTC. Most clinicians now consider it a variant of MEN 2A.
MEN 2B
Medullary thyroid carcinoma (~100%) Pheochromocytoma (50%) Mucosal neuroma, ganglioneuromatosis, marfanoid habitus, colonic abnormalities, characteristic physical appearance (~100%)

MEN 2A (80% of cases), MEN 2B (5% of cases), and FMTC (15% of cases) are inherited as autosomal dominant traits with near-complete penetrance and, in the cases of MEN 2A and MEN 2B, variable expressivity. Approximately 50% of patients with MEN 2A (and a codon 634 mutation) develop pheochromocytomas (Chapters 228 and 231), the frequency being much lower in association with mutations in codons 609, 611, 618, and 620.[7] Before the availability of biochemical and genetic screening in families with MEN 2A, the most frequent cause of death was pheochromocytoma, not MTC. The deaths occurred most often in patients during childbirth or interventional procedures. Thus, pheochromocytoma must be excluded in patients with a confirmed or presumptive diagnosis of hereditary MTC (Chapter 228). With rare exceptions, the pheochromocytoma should be excised first in patients who also have MTC.

Parathyroid hyperplasia occurs in up to 30% of patients with MEN 2A and is usually associated with a *RET* codon 634 mutation. The disease is frequently asymptomatic, with the only abnormality being an elevated serum calcium concentration.[8]

Patients with MEN 2A may also develop cutaneous lichen amyloidosis (CLA) or Hirschsprung disease (HD).[9] CLA occurs in about 25% of patients and involves the interscapular region of the back, corresponding to dermatomes T2 through T6. Pruritus, the dominant symptom, leads to repetitive scratching and secondary skin changes characterized by the deposition of amyloid. The lesion may be evident in infancy, thus serving as a precursor marker of MEN 2A. Cutaneous lichen amyloidosis is almost always associated with a *RET* codon 634 mutation.

HD, manifested by the absence of intrinsic ganglion cells in the distal gastrointestinal tract, has been reported in 30 or more families with MEN 2A or FMTC and is associated with mutations in *RET* exon 10 involving codons 609 (15%), 611 (4%), 618 (30%), and 620 (50%). In functional studies, the cell surface expression of *RET* with these codon mutations is lower than that found with a codon 634 mutation. This suggests a novel mechanism whereby the specified *RET* mutations have low transforming activity, which is sufficient to trigger the development of MTC and pheochromocytoma, yet is insufficient to stimulate differentiation of intestinal ganglion cells. It is also of interest that 50% of patients with familial HD and 30% of patients with sporadic HD have germline *RET* mutations.

Patients with MEN 2B develop mucosal neuromas, ganglioneuromatosis throughout the aerodigestive tract, hypotonia, skeletal malformations, and medullated corneal nerves. They also develop colonic dysfunction manifested by abdominal pain and occasionally intestinal obstruction. Patients

have a characteristic physical appearance, which may not be evident early in life. The failure to diagnose MEN 2B at a young age can be catastrophic because MTC is often evident soon after birth, and regional or distant metastases occur soon thereafter. The MTC associated with MEN 2B is much more aggressive than that occurring with MEN 2A or FMTC. The primary basis for the difference is that MEN 2B mutations are associated with significantly higher basal kinase activity compared with mutations in MEN 2A and FMTC. Patients with FMTC develop only MTC, which, relative to the tumors in patients with MEN 2A and MEN 2B, is slow growing. Many clinicians consider FMTC a variant of MEN 2A.

DIAGNOSIS

The measurement of serum levels of CTN, either in the basal state or following the intravenous administration of the secretagogues calcium, pentagastrin, or both, was initially the primary method of establishing the diagnosis of a C-cell disorder. With the discovery that MEN 2A, MEN 2B, and FMTC are caused by mutations in the *RET* protooncogene, direct DNA analysis became the method of choice for identifying affected family members who had inherited a mutated *RET* allele. At present, the determination of CTN is primarily used to detect persistent or recurrent MTC following thyroidectomy or to evaluate response to therapy in patients with regional or metastatic disease. As we have learned more about the variable clinical expression of MEN 2A in families with the identical *RET* mutation, however, the measurement of basal and stimulated serum CTN levels has assumed importance in timing early thyroidectomy in young family members with *RET* mutations.[10] The two-site, two-step, chemiluminescent, immunometric assay that is highly specific for monomeric CTN is the preferred method for quantitating serum CTN levels.

At present, direct DNA analysis of RET has become the preferred method of detecting *RET* mutations in families with hereditary MTC. The Gene Tests directory currently lists 63 laboratories that perform DNA analysis for *RET* mutations (http://www.genetests.org). Almost all laboratories use direct sequence analysis to evaluate mutations in exons 10, 11, 13, 14, 15, and 16, and some laboratories include exon 8. If no mutations are found in these exons, the entire coding region of *RET* can be sequenced. It is important to perform direct DNA analysis for *RET* mutations in all patients with presumed sporadic MTC because approximately 7% of them will have hereditary MTC. A diagnosis of hereditary MTC in this setting mandates a different treatment strategy for the patient, as well as his family members, who should be offered the opportunity for clinical evaluation and genetic testing.

TREATMENT

Rx

The primary treatment for patients with MTC, whether sporadic or hereditary, is total thyroidectomy. Resection of lymph nodes in the central compartment is included in all adults and in children with MEN 2B but is excluded in outwardly normal youngsters with MEN 2A and FMTC who are undergoing early thyroidectomy based on directed DNA analysis. If enlarged cervical lymph nodes are evident on preoperative ultrasound examination or at the time of thyroidectomy, the involved anatomic nodal compartment should also be resected. During the thyroidectomy, great care must be taken to preserve the parathyroid glands, the recurrent laryngeal nerves, and the external branch of the superior laryngeal nerve. Postoperatively, serum calcitonin is normal in only 10% of patients with node-positive disease compared with 60% of patients with node-negative disease. Many patients with regional lymph node metastases have a good prognosis, however, with 5- and 10-year survival rates of 80 and 70%, respectively.

Repeat neck operation following initial thyroidectomy is indicated in patients with complications from recurrent tumor compressing or invading vital structures, such as the spinal cord, airway, or esophagus. Also, patients who have intractable diarrhea due to markedly elevated tumor hormone secretions, presumably CTN, may obtain symptom relief by tumor debulking. Patients who develop persistent or recurrent MTC following thyroidectomy, as indicated by elevated serum levels of CTN or CEA, are also candidates for reoperation; however, the benefit of such surgical procedures is open to question because there are no long-term data on quality of life and survival. Rarely, patients with MTC develop Cushing syndrome (Chapter 227) due to the inappropriate secretion of adrenocorticotrophic hormone (ACTH) or corticotropin-releasing hormone. Such patients have advanced disease, and bilateral adrenalectomy may be required if steroidogenesis inhibitors are ineffective. Inappropriate secretion of ACTH is a poor prognostic sign, associated with an average survival of 2 years.

The treatment of pheochromocytoma is adrenalectomy, as described in Chapter 228. Hyperparathyroidism (Chapter 245) is managed by either subtotal parathyroidectomy or total parathyroidectomy with heterotopic autotransplantation.

For patients with locally advanced or metastatic MTC, single-agent or combined chemotherapeutic regimens have been minimally effective, being characterized by low response rates of short duration. External beam radiotherapy is indicated primarily for the treatment of localized metastases, primarily of the central nervous system or bone.

With the demonstration that the tyrosine kinase inhibitor imatinib induced remissions in patients with chronic myelogenous leukemia and gastrointestinal stromal tumors, there was hope that similar molecular targeted therapeutics (MTTs) would be developed for other solid tumors, including MTC. In a recent prospective, randomized, placebo-controlled, double-blind, phase III trial, patients treated with the MTT vandetanib had a significantly prolonged progression-free survival compared with placebo.[A1] On the basis of this study, the U.S. Food and Drug Administration (FDA) approved vandetanib for the treatment of patients with advanced MTC.[11] The FDA also recently approved a second MTT, cabozantinib,[12] based on similar results of a phase III trial.[A2] Thus, effective systemic therapies are available for patients with advanced MTC, and additional studies of other MTTs have recently been initiated.

PREVENTION

In patients with a hereditary cancer syndrome, the removal of an organ destined to become malignant should be considered in the light of five factors. There should be (1) near-complete penetrance of the mutated gene, (2) a reliable method of detecting family members who have inherited a mutated allele, (3) minimal morbidity associated with removal of the organ at risk, (4) excellent replacement therapy for the function of the removed organ, and (5) a reliable method for determining whether the operative procedure has been curative. Few hereditary malignancies meet all of these criteria; fortunately, MEN 2A, MEN 2B, and FMTC meet each of them. Young members of kindred with hereditary MTC who are found to have a mutated *RET* allele on genetic screening have the greatest likelihood of being cured by early thyroidectomy. Surgeons in several countries have reported success with this operative procedure, and the question is no longer should it be done but at what age.

The Consensus Committee of the 7th International Workshop on MEN, the National Comprehensive Cancer Network, and the American Thyroid Association have all proposed guidelines for the timing of prophylactic thyroidectomy in patients with MEN 2A, MEN 2B, and FMTC[13]. The recommendations of the three groups are similar, in that children with MEN 2B (or with mutations in codons 918 or 882) should have thyroidectomy at the time of diagnosis, even during the first months of life. Children with MEN 2A and mutations in codons 611, 618, 620, or 634 should have the thyroid removed at or before 5 years of age. In children with mutations in other *RET* codons, the recommended timing of thyroidectomy is less clear but is generally between 5 and 10 years of age.

PROGNOSIS

Several factors portend an adverse outcome in patients with MTC.[14] Poor prognosis is associated with older age, advanced disease at the time of diagnosis of a large primary tumor, lymph node metastases, markedly elevated serum levels of CTN and CEA preoperatively, extrathyroidal invasion of the trachea or soft tissues, and distant metastases. Patients with MEN 2B and patients with MEN 2A who have *RET* mutations in codon 634 have a poorer prognosis than those with *RET* mutations in other codons. Also, in patients with sporadic MTC, the presence of a *RET M918T* mutation, compared with other codon mutations, is associated with a more aggressive tumor and a poor prognosis. Patients apparently cured by thyroidectomy are followed at 6-month intervals with measurement of serum levels of CTN and CEA. The doubling times of serum CTN are especially useful in predicting the course of the disease. CTN doubling times of less than 6 months (compared with those greater than 24 months) are associated with a very poor prognosis.

Grade A References

A1. Wells SA Jr, Robinson BG, Gagel RF, et al. Vandetanib in patients with locally advanced or metastatic medullary thyroid cancer: a randomized, double-blind phase III trial. *J Clin Oncol.* 2012;30:134-141.

A2. Elisei R, Schlumberger MJ, Müller SP, et al. Cabozantinib in progressive medullary thyroid cancer. *J Clin Oncol.* 2013;31:3639-3646.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Wells SA Jr, Pacini F, Robinson BG, et al. Multiple endocrine neoplasia type 2 and familial medullary thyroid carcinoma: an update. *J Clin Endocrinol Metab*. 2013;98:3149-3164.
2. Mete O, Asa SL. Precursor lesions of endocrine system neoplasms. *Pathology*. 2013;45:316-330.
3. Frank-Raue K, Rondot S, Raue F. Molecular genetics and phenomics of *RET* mutations: impact on prognosis of MTC. *Mol Cell Endocrinol*. 2010;322:2-7.
4. Agrawal N, Jiao Y, Sausen M, et al. Exomic sequencing of medullary thyroid cancer reveals dominant and mutually exclusive oncogenic mutations in *RET* and *RAS*. *J Clin Endocrinol Metab*. 2013;98:E364-E369.
5. Moura MM, Cavaco BM, Pinto AE, et al. High prevalence of *RAS* mutations in *RET*-negative sporadic medullary thyroid carcinomas. *J Clin Endocrinol Metab*. 2011;96:E863-E868.
6. Choi SK, Yoon SR, Calabrese P, et al. Positive selection for new disease mutations in the human germline: evidence from the heritable cancer syndrome multiple endocrine neoplasia type 2B. *PLoS Genet*. 2012;8:e1002420.
7. Welander J, Soderkvist P, Gimm O. Genetics and clinical characteristics of hereditary pheochromocytomas and paragangliomas. *Endocr Relat Cancer*. 2011;18:R253-R276.
8. Scholten A, Schreinemakers JM, Pieterman CR, et al. Evolution of surgical treatment of primary hyperparathyroidism in patients with multiple endocrine neoplasia type 2A. *Endocr Pract*. 2011;17:7-15.
9. Moore SW, Zaahl M. The Hirschsprung's-multiple endocrine neoplasia connection. *Clinics*. 2012;67:63-67.
10. Rowland KJ, Moley JF. Hereditary thyroid cancer syndromes and genetic testing. *J Surg Oncol*. 2015;111:51-60.
11. Karras S, Anagnostis P, Krassas GE. Vandetanib for the treatment of thyroid cancer: an update. *Expert Opin Drug Metab Toxicol*. 2014;10:469-481.
12. Nix NM, Braun K. Cabozantinib for the treatment of metastatic medullary thyroid carcinoma. *J Adv Pract Oncol*. 2014;5:47-50.
13. Tuttle RM, Ball DW, Byrd D, et al. Medullary carcinoma. *J Natl Compr Cancer Netw*. 2010;8:512-530.
14. Ho AS, Wang L, Palmer FL. Postoperative nomogram for predicting cancer-specific mortality in medullary thyroid cancer. *Ann Surg Oncol*. 2014; [Epub ahead of print].

REVIEW QUESTIONS

1. Sporadic medullary thyroid carcinomas (MTCs) without somatic *RET* mutations most commonly have which of the following?

A. Somatic *RAS* mutations
B. Germline *RET* mutations
C. Somatic *BRAF* mutations
D. No additional genetic mutations
E. Gene fusions involving the *RET* protooncogene

Answer: A About half of patients with sporadic MTC have somatic *RET* mutations, and about 70% of sporadic MTCs with no somatic *RET* mutations have *HRAS, KRAS*, or rarely *NRAS* mutations. (Moura MM, Cavaco BM, Pinto AE, Leite V. High prevalence of RAS mutations in RET-negative sporadic medullary thyroid carcinomas. *J Clin Endocrinol Metab.* 2011;96: E863-868.)

2. Patients with MEN 2A who develop pheochromocytomas and hyperparathyroidism, in addition to MTC, most commonly have mutations in which codon?

A. *RET* codon 620.
B. *RET* codon 918
C. *RET* codon 634
E. *RET* codon 791
E. *RET* codon 804

Answer: C There is a clear relationship between genotype and phenotype in patients with MEN 2A. Virtually all patients with this syndrome develop MTC, but the development of pheochromocytoma and hyperparathyroidism is highly variable, occurring much more often in patients with *RET* 634 codon mutations and much less frequently in patients with other *RET* codons mutations. (Wells SA Jr, Pacini F, Robinson BG, Santoro M. Multiple endocrine neoplasia type 2 and familial medullary thyroid carcinoma: an update. *J Clin Endocrinol Metab.* 2013;98:3149-3164.)

3. After thyroidectomy, the most reliable indicator of prognosis is which of the following?

A. The specific *RET* or *RAS* mutation associated with the MTC
B. The serum levels of calcitonin (CTN) and carcinoembryonic antigen (CEA) immediately postoperatively
C. The rate at which serum CTN and CEA double
D. The completeness of the thyroidectomy
E. The gender of the patient

Answer: C The presence of an elevated serum CTN or CEA in the immediate postoperative period or soon thereafter is diagnostic of persistent or recurrent MTC. Although this determination is valuable, it is not as significant an indicator of prognosis as the rate at which serum levels of CTN and CEA double. Doubling times of less than 6 months for either CTN or CEA are a very poor prognostic sign. (Meijer JA, le Cessie S, van den Hout WB, et al. Calcitonin and carcinoembryonic antigen doubling times as prognostic factors in medullary thyroid carcinoma: a structured meta-analysis. *Clin Endocrinol.* 2010;72:534-542.)

4. In patients with MEN 2B where the *RET* mutation arises de novo,the mutated allele comes from which parent(s)?

A. The mother
B. The father
C. Either parent
D. Neither parent

Answer: B In patients with de novo MEN 2B, where neither parent expresses the characteristic phenotype associated with the disease, the mutated *RET* allele is virtually always inherited from the father. The average age of males who transmit the disease is significantly greater than the average age of all fathers. This appears to be due to a selective advantage acquired by the testis stem cells, which increases the mutation's frequency in the testis. (Choi SK, Yoon SR, Calabrese P, Arnheim N. Positive selection for new disease mutations in the human germline: evidence from the heritable cancer syndrome multiple endocrine neoplasia type 2B. *PLoS Genet.* 2012;8:e1002420.)

5. In patients with hereditary MTC, thyroidectomy should be performed within the first 5 years of life in which cases?

A. All patients with a germline *RET* codon mutation
B. Only patients with MEN 2B
C. Patients with MEN 2A who have a germline *RET* codon 634 mutation and patients with MEN 2B
D. Patients with a germline *RAS* mutation
E. Patients who only have an elevated serum calcitonin level

Answer: C The published guidelines for managing patients with hereditary MTC all discuss the timing of thyroidectomy in children who have inherited a mutated *RET* allele. There is uniform agreement that thyroidectomy should be performed before 5 years of age in patients with *RET* mutations in codon 634, and as soon as the diagnosis is made in patients with MEN 2B, even in the first months of life. (Kloos RT, Eng C, Evans DB, et al. Medullary thyroid cancer: management guidelines of the American Thyroid Association. *Thyroid.* 2009;19:565-612.)

247

PAGET DISEASE OF BONE

STUART H. RALSTON

DEFINITION

Paget disease of bone is a disorder of skeleton characterized by increased and disorganized bone remodeling affecting one or more skeletal sites. Affected bones enlarge, become deformed, and are at increased risk for pathologic fractures

EPIDEMIOLOGY

The population prevalence of Paget disease is about 1% in the United States and 2% in the United Kingdom. It is also common in Western Europe and in people of European descent who have migrated to other parts of the world. Paget disease is rare in Scandinavians, Africans, and Asians. These differences are thought to have a genetic basis and to be caused by founder mutations that occurred in Europeans many centuries ago with subsequent spread to the rest of the world through emigration. Paget disease is strongly related to age; in the United Kingdom, the incidence is 0.3 to 0.5 per 10,000 person years in those aged 55 to 59 years but doubles in frequency each decade thereafter to reach an incidence of 5.4 per 10,000 person years in women and 7.6 per 10,000 person years in men in those 85 years and older. The prevalence and severity of Paget disease have diminished in most countries over the past 25 years.[1] The causes are unclear, but suggested explanations include influx of migrants from low prevalence areas in some populations, improved nutrition, a more sedentary lifestyle with a reduction in skeletal injuries, and reduced exposure to infections.[2]

PATHOBIOLOGY

Susceptibility to Paget disease seems to be genetically determined, but environmental factors also play a key role in regulating onset and severity of the disease. The importance of genetics is emphasized by the fact that between 15 and 40% of patients have a positive family history and that the risk for developing Paget disease in a first-degree relative of a patient is about sevenfold higher than in the general population.[3] In many families, the disease is transmitted in an autosomal dominant manner, although penetrance is incomplete. The most important susceptibility gene for classical Paget disease is *SQSTM1*. Mutations of *SQSTM1* are present in up to 40% of patients with a family history and 5 to 10% of people with sporadic disease. The *SQSTM1* gene encodes a protein called p62 that is involved in regulating signal transduction downstream of the receptor activator of nuclear factor κB (RANK), which plays a critical role in regulating osteoclastogenesis when activated by RANK ligand (RANKL) (Fig. 247-1). The disease-causing mutations cluster in the ubiquitin-associated domain and have the effect of upregulating nuclear factor κB (NFκB) signaling and stimulating osteoclastogenesis by complex mechanisms that are reviewed in detail elsewhere. Genome-wide association

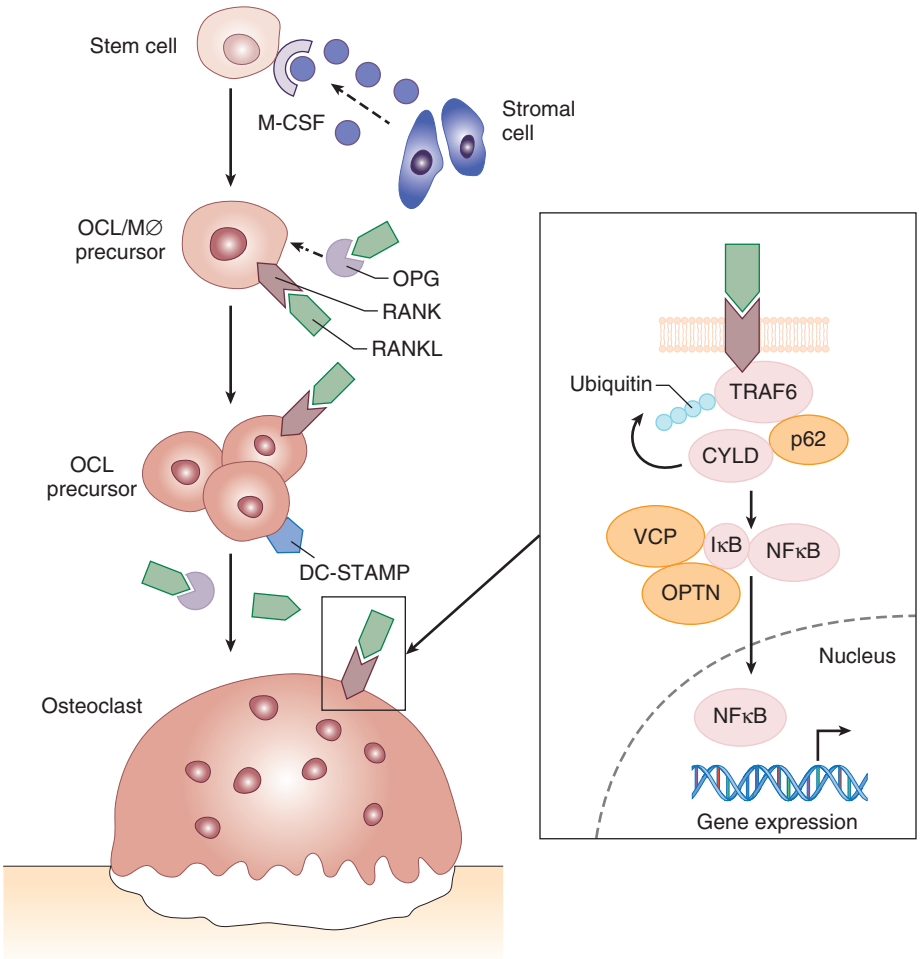

FIGURE 247-1. Regulators of osteoclast dysfunction in Paget disease. Some of the key molecules that have been implicated in the pathogenesis of Paget disease are illustrated. Macrophage colony-stimulating factor (M-CSF) encoded by *CSF1* is required for differentiation of stem cells to the osteoclast/macrophage (OCL/MØ) lineage. Osteoclast differentiation and activity are enhanced when RANK (encoded by *TNFRSF11A*) is activated by RANKL but inhibited by OPG (encoded by *TNFRSF11B*). Fusion of osteoclast precursors to form mature osteoclasts requires DC-STAMP (encoded by *TM7SF4*). Within the cell *(inset),* p62 (encoded by *SQSTM1*) is required for signal transduction downstream of the RANK receptor and is also involved in regulating autophagy. Both VCP (encoded by *VCP*) and OPTN (encoded by *OPTN*) also play a role in regulating NFκB signaling and autophagy.

TABLE 247-1 GENES THAT PREDISPOSE TO PAGET DISEASE–LIKE SYNDROMES

SYNDROME	GENE (PROTEIN)	GENE FUNCTION	INHERITANCE/MUTATION	CLINICAL FEATURES
Familial expansile osteolysis Early-onset familial Paget disease of bone Expansile skeletal hyperplasia	*TNFRSF11A* (RANK)	Enhances osteoclast differentiation and bone resorption	Autosomal dominant Insertion mutations exon 1	Onset during adolescence, extensive bone lesions, deafness, tooth loss.
Juvenile Paget disease	*TNFRSF11B* (OPG)	Inhibits osteoclast differentiation and bone resorption	Autosomal recessive Various loss of function mutations	Onset during childhood, extensive bone lesions, deafness, fractures, deformity, premature cardiovascular disease
Inclusion body myopathy, Paget disease, and frontotemporal dementia	*VCP* (p97)	Multiple cellular functions, including roles in NFκB signaling and autophagy	Autosomal dominant Loss of function mutations in UBA domain	Onset during 3rd-4th decades, with myopathy, Paget disease, and dementia occurring during 5th-6th decades

studies have identified seven other loci that predispose to Paget disease, which individually increase the risk between 1.4 and 1.7 fold.[4,5] These loci have additive effects such that individuals who carry several predisposing alleles have a substantially increased risk for developing Paget disease. Many of these loci lie close to genes that play key roles in osteoclast function, including *CSF1*, which encodes macrophage colony-stimulating factor (M-CSF); *TNFRSF11A*, which encodes RANK; *TM7SF4*, which encodes DC-STAMP; and *OPTN*, which encodes optineurin (see Fig. 247-1). Several inherited diseases with clinical features overlapping with those of Paget disease have also been described (Table 247-1). These syndromes are also caused by mutations in genes that regulate osteoclast function.

Under normal circumstances, bone is renewed and repaired in an orderly and tightly regulated fashion through the process of bone remodeling. The bone remodeling process is highly abnormal in Paget disease. Osteoclasts are increased in number, larger than normal, and hypernucleated. Some contain nuclear inclusion bodies. These were originally thought to be paramyxovirus nucleocapsids, but it has been suggested more recently that they may be aggregates of un-degraded proteins caused by defects in the autophagy pathway. Bone formation is also markedly increased, but the amount of new bone that is formed greatly exceeds that which has been removed by osteoclast activity, leading to enlargement and deformity of affected bones (Fig. 247-2). The bone that is formed is laid down in a disorganized fashion (woven bone) and has impaired mechanical strength. Other features include increased vascularity and marrow fibrosis. The focal nature of Paget disease remains a puzzle. Suggested explanations include the occurrence of somatic mutations in affected bones, which locally increase osteoclast activity, or excessive mechanical loading or skeletal injuries early in life, which by causing microdamage act as a focus for localized increases in bone remodeling.

CLINICAL MANIFESTATIONS

It has been estimated that between 7 and 16% of patients with Paget disease come to medical attention, and the presentation is highly variable.[6] Many patients are asymptomatic, and Paget disease is detected as the result of a raised serum alkaline phosphatase (ALP) or an abnormal radiograph in patients who are being investigated for another reason. In those that do present clinically, symptoms that are attributable to Paget disease are observed in about 75% of cases. The most common is pain, which can be due to either increased bone turnover or a complication such as osteoarthritis, spinal stenosis, pseudofractures, or nerve compression syndromes. Deafness may occur in patients with skull involvement, but this is usually conductive rather than due to auditory nerve compression. Osteosarcoma occurs in less than 0.5% of cases but should be suspected in patients who experience a sudden increase in bone pain or swelling of an affected site. Other, rare complications include obstructive hydrocephalus, high-output cardiac failure, and hypercalcemia in patients who are immobilized. The risk for cardiovascular disease is increased in patients with Paget disease compared with age- and gender-matched controls, probably owing to an increased prevalence of vascular calcification. Most patients have no clinical signs but some present with bone deformity (see Fig. 247-2) or warmth of the skin overlying an affected bone.

DIAGNOSIS

The diagnosis can usually be made by radiograph, which shows the typical features of focal osteolysis with coarsening of the trabecular pattern, bone expansion, and cortical thickening (see Fig. 247-2). Occasionally, the disease may be predominantly lytic in nature (see Fig. 247-2). The most sensitive way of defining the extent of Paget disease is a radionuclide bone scan in which tracer uptake is intensely increased at affected sites (see Fig. 247-2). Imaging with magnetic resonance imaging and computed tomography is not usually required unless complications such as spinal stenosis or osteosarcoma are suspected. Laboratory testing should include assessment of renal function, calcium, albumin, alkaline phosphatase (ALP), and 25(OH)D levels; liver function should be assessed to rule out the possibility that elevations in ALP are of hepatic origin. Typically, Paget disease presents with an elevation in ALP with otherwise normal biochemistries, but normal levels of ALP do not exclude the diagnosis. Vitamin D deficiency is a common finding but most likely reflects the fact that Paget disease predominantly affects older people in whom vitamin D deficiency is prevalent. Specialized markers such as bone-specific ALP or procollagen type 1 N-terminal propeptide can be useful in patients with coexisting liver disease but otherwise offer no advantage over total ALP in diagnosis and assessing treatment response. Susceptibility to Paget disease can be assessed in relatives of affected patients by genetic testing for *SQSTM1* mutations, although this is not commonly performed in routine clinical practice.

The differential diagnosis includes hyperostosis frontalis interna (a benign condition characterized by osteosclerosis of the frontal bones of the skull), fibrous dysplasia, pustulotic arthro-osteitis (which can present with mixed osteosclerotic and osteolytic lesions of the clavicle and ribs),[7] and osteosclerotic metastases, particularly from carcinoma of the prostate. Usually, Paget disease can be distinguished from these conditions biochemically and through imaging, but occasionally, biopsy of an affected site may be required.

Treatment

The most common indication for medical treatment of Paget disease is bone pain localized to an affected site.[8] Although such pain may be caused by increased metabolic activity, other causes may also be operative, including nerve compression syndromes, pseudofractures, secondary osteoarthritis, and other musculoskeletal conditions. Careful assessment of the patient is therefore necessary to decide on the most appropriate treatment. Bone pain caused by increased metabolic activity is localized to the affected site and is usually accompanied by a raised ALP level. It is common to encounter patients in whom pain occurs in the presence of coexisting osteoarthritis, bone deformity, or other musculoskeletal conditions. In such cases, it can be difficult to be sure about the origin of the pain, and many clinicians give a therapeutic trial of bisphosphonates. If the pain responds, then one can assume it was due to increased metabolic activity; if it does not, further evaluation should be undertaken to identify the cause and treat the patient appropriately. Pseudofractures represent a distinct management problem. These are areas of focal osteolysis that traverse the lateral cortex of weight-bearing bones of the lower limbs. Some remain stable for prolonged periods without causing symptoms; others regress spontaneously; and others progress to pathologic fracture, often in association with a localized increase in pain at the affected site.[9]

Bisphosphonates

Bisphosphonates are the drugs of first choice for the treatment of pain that is thought to be due to increased metabolic activity. Nowadays, nitrogen-containing bisphosphonates (aminobisphosphonates) are used in preference to older bisphosphonates because of their greater potency (Table 247-2).

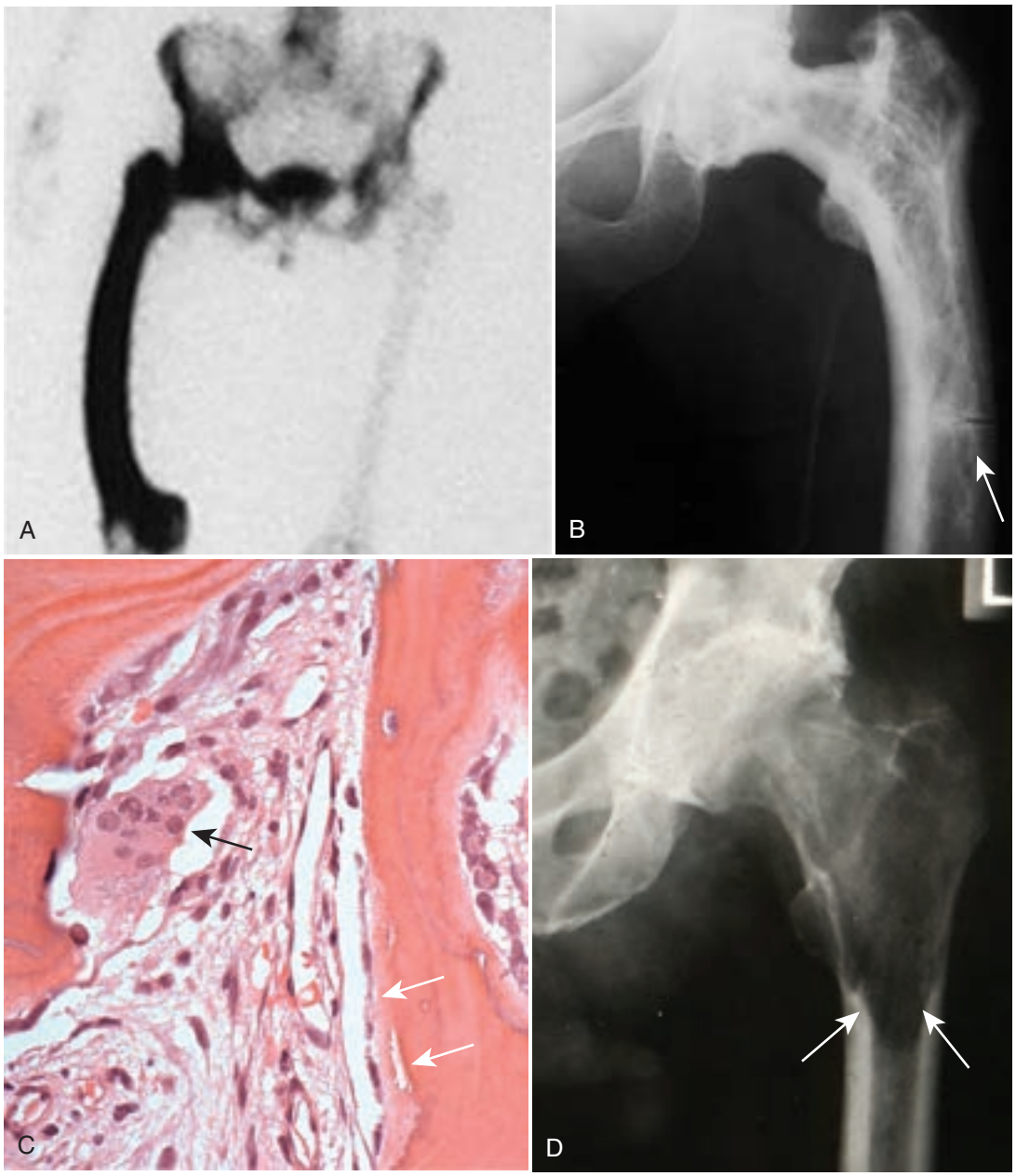

FIGURE 247-2. Radiographic and histologic features of Paget disease. A, Radionuclide bone scan image showing intense tracer uptake, typical of Paget disease of bone (PDB) in the right femur. **B,** Radiograph of an affected left femur showing bone expansion with mixed osteolytic/osteosclerotic areas and loss of normal trabecular pattern. A pseudofracture is visible in the lateral cortex *(arrow).* **C,** Histologic features from a hematoxylin and eosin stained section. A large osteoclast is visible *(black arrow)* close to an area of new bone formation *(white arrows).* There is extensive marrow fibrosis. Irregular cement lines typical of woven bone are apparent to the right of section. **D,** Predominantly lytic Paget disease of the left femur. The lytic area involves the intertrochanteric region and extends down the femoral shaft *(white arrows).*

Placebo controlled trials have shown that most bisphosphonates are effective at improving bone pain in Paget disease. Superiority of aminobisphosphonates over etidronate and tiludronate at suppressing ALP levels has been demonstrated, but with little difference in the response of pain. There are limited data comparing different aminobisphosphonates. In an open-label study comparing pamidronate 180 mg intravenously in unit doses of 30 mg weekly or 60 mg alternate weeks to oral alendronate 40 mg daily given in 3-monthly blocks over a 2-year period, there were no significant differences in the proportion of patients who achieved normal levels of ALP (86 and 91% respectively) or in symptomatic response.[■] Another study that compared a single infusion of 5 mg zoledronic acid with oral risedronate 30 mg daily for 2 months showed significant superiority of zoledronic acid in lowering ALP. Those randomized to zoledronic acid had greater improvement in some domains of health-related quality of life, but the differences between groups were small (1 to 2 points) and below the threshold of 5 points, which is considered clinically significant.[■] Another randomized trial compared the effects of giving repeated course of bisphosphonates (mainly risedronate) with the aim of normalizing ALP (intensive treatment), with therapy primarily aimed at controlling symptoms (symptomatic therapy) in Paget disease.[■] This showed no difference in response of pain, quality of life, or complications between the groups, indicating that trying to restore ALP to normal confers no clinical advantage in most patients with Paget disease.

After initiation of bisphosphonate therapy, levels of ALP start to fall within about 10 days and reach a nadir between 3 and 6 months. Levels of ALP can remain suppressed for many months or years thereafter, particularly with zoledronic acid. Symptoms can improve while ALP levels are still falling, and good clinical responses are often observed in patients whose ALP levels are not restored to normal.

Intravenous bisphosphonates can cause transient bone pain, myalgia, headache, nausea, pyrexia, and fatigue within 1 to 3 days of the infusion in about 25% of cases (acute phase response). These symptoms can be ameliorated by acetaminophen given before and for a few days after the infusion, but they almost always subside within 7 days even without treatment. The acute phase response is much less common after second and subsequent infusions. Hypocalcemia may occur, particularly in patients with substantial elevations in bone turnover and vitamin D deficiency. The risk can be

TABLE 247-2 BISPHOSPHONATES USED IN THE TREATMENT OF PAGET DISEASE

DRUG	DOSE	COMMON ADVERSE EFFECTS
Oral		
Etidronate*	400 mg/day orally for 3-6 mo	Diarrhea, nausea, abdominal pain
Tiludronate	400 mg/day orally for 3 mo	Diarrhea, nausea, dyspepsia
Risedronate	30 mg/day orally for 2 mo	Dyspepsia, esophagitis
Alendronic acid†	40 mg/day orally for 6 mo	Dyspepsia, esophagitis
Intravenous		
Pamidronate	180 mg IV in unit doses of 30 mg weekly or 60 mg alternate weeks	Acute phase response, hypocalcemia
Zoledronic acid	5 mg IV	Acute phase response, hypocalcemia

*Now seldom used.
†Not licensed in the United Kingdom or Europe for Paget disease. Etidronate, pamidronate, tiludronate, and risedronate should be avoided if estimated glomerular filtration rate (eGFR) <30; zoledronic acid and alendronic acid should be avoided if eGFR <35.

minimized by correcting vitamin D deficiency before treatment and providing calcium and vitamin D supplements for the first 1 or 2 weeks after the infusion.

Patients taking oral bisphosphonates must fast for 30 minutes (risedronate, alendronate) or 120 minutes (etidronate, tiludronate) before and after dosing to achieve adequate absorption. The most common adverse effects are dyspepsia (risedronate and alendronic acid) and diarrhea (tiludronate and etidronate). Other rare side effects of bisphosphonates include uveitis, skin rashes, atrial fibrillation, and osteonecrosis of the jaw, as well as atypical subtrochanteric fractures. Bisphosphonates can cause kidney injury and are contraindicated in patients with significant renal impairment.

Other Drug Treatments

Analgesics, anti-inflammatory drugs, and antineuropathic agents are often required in patients with Paget disease, particularly when there is coexisting osteoarthritis or a nerve compression syndrome. Calcitonin can improve bone pain due to metabolic activity in Paget disease but is seldom used except in patients for whom bisphosphonates are contraindicated. Adverse effects such as nausea and flushing can be problematic, and resistance may develop owing to the formation of neutralizing antibodies. Anecdotal reports suggest that the osteoclast inhibitor denosumab may also be effective at reducing ALP levels in Paget disease of bone, but it is not licensed for this indication.

Nonpharmacologic Treatments

Nonpharmacologic approaches (acupuncture, physiotherapy, hydrotherapy, and transcutaneous electrical nerve stimulation) are often used to control pain, but their effectiveness has not been specifically investigated in controlled trials. Clinical experience suggests that specific problems such as limb shortening and deformity can be helped by aids and devices such as walking sticks and shoe raises.

Monitoring Disease Activity and the Effects of Treatment

Metabolic activity and the response to treatment is typically assessed by measuring ALP, although levels can be normal in patients with localized disease that is metabolically active. Further courses of treatment should be considered in patients with recurrent or persistent pain in whom ALP levels remain or become elevated.

Surgery

Orthopedic surgery may be required for the management of coexisting osteoarthritis, pseudofractures, fractures, bone deformity, and spinal stenosis. Osteotomy is performed infrequently, but analyses of small cases series have reported good results in about 60% of patients. Surgery is much more frequently required to repair fractures and to replace joints that are affected by osteoarthritis. Surgical treatment of Paget disease can be technically challenging because of deformity, osteosclerosis, and increased vascularity, but evidence from case series indicate that fractures through pagetic bone heal normally and that joint replacement surgery has a good outcome. It has been suggested that a bisphosphonate should be given before orthopedic and spinal surgery with the aim of reducing operative blood loss, but the effectiveness of this has not been studied. There is a theoretical concern that previous bisphosphonate therapy might impair fracture union and bone repair, but there is little evidence to suggest that this is a problem in clinical practice.[10] Orthopedic surgery may also be required in patients who develop osteosarcoma, but the prognosis is poor even with aggressive operative treatment.

PROGNOSIS

The prognosis of Paget disease is highly variable. Some patients remain completely asymptomatic throughout life, but those that present clinically frequently have complications and have a significant reduction in quality of life. Although modern bisphosphonates are highly effective at suppressing bone turnover in Paget disease,[11] they have not as yet been shown to alter the natural history of the disease or prevent complications. Disease severity and extent can be predicted by genotyping for *SQSTM1* mutations and other risk alleles, and studies are currently in progress to determine whether genetic testing can be combined with prophylactic bisphosphonate therapy to prevent or delay the onset of disease.

Grade A References

A1. Walsh JP, Ward LC, Stewart GO, et al. A randomized clinical trial comparing oral alendronate and intravenous pamidronate for the treatment of Paget disease of bone. *Bone*. 2004;34:747-754.
A2. Reid IR, Miller P, Lyles K. Comparison of a single infusion of zoledronic acid with risedronate for Paget disease. *N Engl J Med*. 2005;353:898-908.
A3. Langston AL, Campbell MK, Fraser WD, et al. Randomised trial of intensive bisphosphonate treatment versus symptomatic management in Paget disease of bone. *J Bone Miner Res*. 2010;25:20-31.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Corral-Gudino L, Garcia-Aparicio J, Sanchez-Gonzalez MD. Secular changes in Paget disease: contrasting changes in the number of new referrals and in disease severity in two neighboring regions of Spain. *Osteoporos Int*. 2013;24:443-450.
2. Ralston SH. Clinical practice: Paget disease of bone. *N Engl J Med*. 2013;368:644-650.
3. Ralston SH, Layfield R. Pathogenesis of Paget disease of bone. *Calcif Tissue Int*. 2013;91:97-113.
4. Albagha OME, Wani S, Visconti MR, et al. Genome-wide association identifies three new susceptibility loci for Paget disease of bone. *Nat Genet*. 2011;43:685-689.
5. Albagha OM, Visconti MR, Alonso N. Genome-wide association study identifies variants at CSF1, OPTN and TNFRSF11A as genetic risk factors for Paget disease of bone. *Nat Genet*. 2010;42:520-524.
6. Tan A, Ralston SH. Clinical presentation of Paget disease: evaluation of a contemporary cohort and systematic review. *Calcif Tissue Int*. 2014;95:385-392.
7. Nguyen MT, Borchers A, Selmi C, et al. The SAPHO syndrome. *Semin Arthritis Rheum*. 2012;42:254-265.
8. Gruener G, Camacho P. Paget disease of bone. *Handb Clin Neurol*. 2014;119:529-540.
9. Tan A, Ralston SH. Paget disease of bone. *QJM*. 2014;107:865-869.
10. Goldhahn J, Feron JM, Kanis J, et al. Implications for fracture healing of current and new osteoporosis treatments: an ESCEO consensus paper. *Calcif Tissue Int*. 2012;90:343-353.
11. Devogelaer JP, Geusens P, Daci E, et al. Remission over 3 years in patients with Paget disease of bone treated with a single intravenous infusion of 5 mg zoledronic acid. *Calcif Tissue Int*. 2014;94:311-318.

REVIEW QUESTIONS

1. A previously healthy 75-year-old man presents to his primary care physician with pain in the left hip that is worse on walking. Systematic enquiry is unremarkable, but examination reveals pain and a reduction in range of movement of the left hip. Radiography shows mixed osteosclerotic-osteolytic lesions in the right hemipelvis with expansion of the iliac bone. Osteophytes and joint space narrowing are observed in the left hip. Routine hematology, urea and electrolytes, calcium and phosphate, and liver function tests are normal, but alkaline phosphatase is raised at 200 μ/L (reference range, 40 to 125). Serum 25(OH)D is 35 nmol/L, parathyroid hormone is 7ng/L (normal, 2 to 12). What is the most likely cause of the hip pain?

A. Paget disease of the pelvis
B. Osteoarthritis of the hip
C. Osteomalacia
D. Osteoporosis
E. Prostate cancer

Answer: B Osteoarthritis of the hip. Although the blood test results are consistent with Paget disease, the pain is localized to the noninvolved side of the pelvis. Although serum 25(OH)D is slightly reduced, osteomalacia is unlikely because this does not usually present with localized pain, and parathyroid is normal. Prostate cancer with bone metastases can cause osteosclerotic metastases, which can look similar to Paget disease on radiography, but the fact that he is clinically well makes this less likely.

2. A 47-year-old man with known Paget disease of the left tibia presents with pain localized to the anterior aspect of the affected tibia. On examination, there is mild deformity of the tibia on the left side, with warmth of the skin overlying the tibia. Hematology and biochemistry are normal, including alkaline phosphatase, which is 105 (reference range, 40 to 125). What is the most appropriate treatment?

A. Paracetamol
B. Surgical osteotomy of the tibia to correct the deformity
C. Amitriptyline
D. Intravenous zoledronic acid, 5 mg IV annually for 3 years
E. Intravenous zoledronic acid, 5 mg IV

Answer: E A single infusion of zoledronic acid would be the most appropriate option. The pain is clearly localized to an affected site and is likely due to Paget disease, even though the alkaline phosphatase is normal, because there is warmth of the overlying skin. To give three infusions would be unnecessary and would represent overtreatment because the clinical effects of zoledronic acid usually last for several years in Paget disease. Osteotomy can be used to correct significant deformity but would not be indicated to treat pain in a patient with mild deformity. Both paracetamol and amitriptyline can be used to help pain control in Paget disease, but in this case it would be appropriate to first assess the response to bisphosphonate before progressing to other treatments.

3. A 62-year-old women presents with a 5-year history of musculoskeletal pain affecting the low back region, both shoulders, and both hips, neck, and head. She reports that her father and older brother have a history of Paget disease. Routine hematology, urea and electrolytes, calcium and phosphate, and liver function tests are normal, but alkaline phosphatase is raised at 2350 μ/L (reference range, 40 to 125). A radiograph of the pelvis shows typical changes of Paget disease. What would be the most appropriate investigation to determine the extent of the disease?

A. Magnetic resonance imaging (MRI) of the spine
B. Computed tomography (CT) of the abdomen, chest, and pelvis
C. Serum bone specific alkaline phosphatase
D. Radionuclide bone scan.
E. Radiograph of the skull, shoulders, and cervical spine

Answer: D Radionuclide bone scan. Bones that are affected by Paget disease typically show intense tracer uptake, allowing one to delineate the extent of involvement in a single test. Imaging with MRI or CT can be useful if complications of Paget disease are suspected, such as spinal stenosis or osteosarcoma, but are not as useful in determining the extent of the disease. Bone-specific alkaline phosphatase will provide information on whether the elevated alkaline phosphatase is of bony origin but will not provide information on disease extent. Although radiographs can detect Paget disease with reasonable sensitivity, one would need to perform a whole skeletal survey to evaluate the extent of the disease.

4. A 75-year-old man develops pain in the low back region that is gradually worsening and has started to radiate into the buttocks on walking. A lateral radiograph of the lumbar spine shows mixed osteosclerotic-osteolytic lesions in lumbar vertebrae 2, 3, and 4, with expansion of all three vertebrae. Routine hematology, urea and electrolytes, calcium and phosphate, and liver function tests are normal, but alkaline phosphatase is raised at 350 μ/L (reference range, 40 to 125). What further investigations would be useful in determining the cause of the pain?

A. MRI of the lumbar spine
B. Anteroposterior radiograph of the spine
C. Radionuclide bone scan
D. Serum procollagen 1N peptide (PINP)
E. CT scan of the lumbar spine

Answer: A MRI of the lumbar spine would be the most appropriate test because the symptoms are suspicious of spinal stenosis, which is a known complication of Paget disease of the lumbar spine and caused by expansion of the affected bones with encroachment on the cauda equina. The anteroposterior radiograph would not be useful in identifying spinal stenosis. Although the CT scan could give information on narrowing of the lumbar canal, it does not provide detail on the soft tissues, which would be necessary to determine whether the cauda equina is being compromised. The bone scan may show evidence of Paget disease in the spine but would not provide information on cauda equina compression. Measurement of PINP would provide information on bone formation but would not be helpful in defining the cause of the pain.

5. A 54-year-old man falls and sustains an injury to his right ankle. A radiograph is taken to exclude a fracture. No evidence of fracture is found, but changes consistent with early Paget disease are noted affecting the upper part of the right tibia. On clinic review 6 weeks after the injury, the patient has fully recovered and is asymptomatic. Routine biochemistry and hematology are normal. What would be the most appropriate evidence-based treatment option for this patient?

A. Prophylactic therapy with intravenous zoledronic acid every 3 years for the next 10 years to prevent progression of the disease.
B. Prophylactic therapy with oral alendronate 70 mg once a week on a long-term basis to prevent progression of the disease.
C. Close observation with repeat radiographs on an annual basis followed by zoledronic acid treatment, 5 mg IV, if the radiographs shows signs of progression.
D. Surgical removal of the upper part of the tibia with total knee replacement.
E. Observation at periodic intervals with initiation of bisphosphonate treatment should symptoms that are suggestive of Paget disease develop.

Answer: E There is currently no evidence that prophylactic therapy with any bisphosphonate can alter the natural history of the disease. Similarly, there is no evidence that prophylactic surgery would be helpful in this situation. The most appropriate option would be to keep the patient under review to look for the development of symptoms that might be attributable to the Paget disease and give treatment if these occur.

248

OSTEONECROSIS, OSTEOSCLEROSIS/HYPEROSTOSIS, AND OTHER DISORDERS OF BONE

MICHAEL P. WHYTE

OSTEONECROSIS

DEFINITION

Osteonecrosis (aseptic, avascular, or ischemic necrosis of bone) refers to skeletal infarction. Bone infarcts may be asymptomatic, cause self-limited discomfort, or engender painful collapse of subarticular bone that leads to joint destruction.

PATHOBIOLOGY AND PATHOGENESIS

Many conditions are associated with osteonecrosis (Table 248-1). In adults, the most common causes are ethanol abuse and long-term glucocorticoid therapy, both of which have dose-dependent effects.

Skeletal infarction may result from blood vessel destruction (e.g., joint dislocation, fracture), obstruction (e.g., thromboemboli, sickle cell disease, fat emboli, caisson disease), or, hypothetically, compression from local expansion of fatty tissue (e.g., ethanol abuse, glucocorticoid treatment, diabetes mellitus). However, symptoms may not occur unless, weeks later, resorption of dead bone during skeletal repair leads to pathologic fracture. Certain skeletal sites (often subarticular) are predisposed to osteonecrosis,

TABLE 248-1 CAUSES OF ISCHEMIC NECROSIS OF CARTILAGE AND BONE

Endocrine/metabolic
Ethanol abuse
Glucocorticoid therapy
Cushing disease
Diabetes mellitus
Hyperuricemia
Osteomalacia
Hyperlipidemia
Bone antiresorptive therapy (osteonecrosis of the jaw)
Storage diseases (e.g., Gaucher disease)
Hemoglobinopathies (e.g., sickle cell disease)
Trauma (e.g., dislocation, fracture)
Human immunodeficiency virus (HIV) infection
Dysbaric conditions (e.g., caisson disease)
Collagen vascular disorders
Irradiation
Pancreatitis
Organ transplantation
Hemodialysis
Burns
Intravascular coagulation
Idiopathic, familial
Pregnancy

but the locations differ for traumatic and nontraumatic processes and for children and adults. *Osteochondrosis* refers to necrosis of ossification centers; more than 50 eponymic types have been recorded. The susceptibility of children to osteochondrosis and its pathogenesis are poorly understood. At all ages, however, the femoral head is especially prone to infarction. Nontraumatic osteonecrosis commonly affects the humeral head, femoral condyles, distal end of the tibia, and talus. Although the pathogenesis is uncertain, administration of potent bone antiresorptive agents, especially to patients with malignant disease, has been associated with osteonecrosis of the jaw (Fig. 248-1).[1]

CLINICAL MANIFESTATIONS

Pain occurs acutely if there is skeletal collapse. Chronic arthralgia results from desquamated necrotic tissue and articular destruction.

DIAGNOSIS

Magnetic resonance imaging that demonstrates bone marrow edema is especially sensitive for detection of early osteonecrosis. Bone scintigraphy discloses skeletal reconstitution with or without fracture. Relatively late in the pathologic process, radiographs first show patchy areas of osteopenia and osteosclerosis that reflect skeletal repair. A linear subchondral radiolucency (crescent sign) indicates bone collapse.

TREATMENT Rx

Non–weight bearing is advisable for the affected limb. Decompression by trephine insertion is used at some sites. Arthrotomy to remove debris, transpositional osteotomy, arthroplasty, or joint replacement may be necessary.

OSTEOSCLEROSIS/HYPEROSTOSIS

Many conditions are associated with radiographic evidence of increased bone density. Skeletal dysplasias, metabolic disturbances, and various other disorders can cause generalized or focal increases in bone mass (Table 248-2). Aberrations in skeletal growth, modeling (shaping), or remodeling (turnover) may be at fault. *Osteosclerosis* refers to thickening of trabecular (spongy, cancellous) bone. *Hyperostosis* describes widening of cortical (compact) bone. Increases in trabecular bone, cortical bone, or both may augment skeletal density.

Osteosclerosis

Neoplastic, hematologic, and metabolic disorders may preferentially cause sclerosis in trabecular bone because it houses marrow and remodels more rapidly than cortical bone.

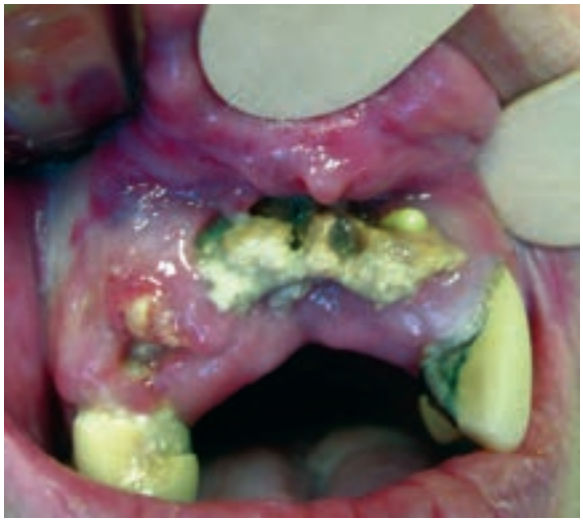

FIGURE 248-1. Exposed dead bone characterizes osteonecrosis of the jaw.

TABLE 248-2 DISORDERS THAT CAUSE DENSE BONES

DYSPLASIAS
Central osteosclerosis with ectodermal dysplasia
Craniodiaphyseal dysplasia
Craniometaphyseal dysplasia
Dysosteosclerosis
Endosteal hyperostosis
van Buchem disease
Sclerosteosis
Worth type (LRP5 activation)
Frontometaphyseal dysplasia
Infantile cortical hyperostosis (Caffey disease)
Juvenile Paget disease
Lenz-Majewski syndrome
Melorheostosis
Metaphyseal dysplasia (Pyle disease)
Mixed sclerosing bone dystrophy
Oculodento-osseous dysplasia
Osteodysplasia of Melnick and Needles
Osteopathia striata
Osteopetrosis (several types)
Osteopoikilosis
Progressive diaphyseal dysplasia (Engelmann disease)
Pyknodysostosis
METABOLIC CONDITIONS
Carbonic anhydrase II deficiency
Fluorosis
Heavy metal poisoning
Hepatitis C–associated osteosclerosis
Hyperparathyroidism, hypoparathyroidism, pseudohypoparathyroidism
Hypervitaminosis A, D
Hypophosphatemic rickets or osteomalacia
Milk-alkali syndrome
Renal osteodystrophy
OTHER DISORDERS
Axial osteomalacia
Fibrogenesis imperfecta ossium
Ionizing radiation
Lymphoma
Mastocytosis
Multiple myeloma
Myelofibrosis
Osteomyelitis
Osteonecrosis
Paget bone disease
Sarcoidosis
Skeletal metastases
Tuberous sclerosis

From Whyte MP. Skeletal disorders characterized by osteosclerosis or hyperostosis. In: Avioli LV, Krane SM, eds. *Metabolic Bone Disease.* 3rd ed. San Diego: Academic Press; 1998.

FIBROGENESIS IMPERFECTA OSSIUM

DEFINITION

This rare, sporadic condition features generalized osteopenia, but coarsening of the remaining trabeculae places it among disorders that manifest osteosclerosis.

PATHOBIOLOGY

The cause is unknown. Subperiosteal bone formation and collagen synthesis in nonosseous tissues seem to be normal.

CLINICAL MANIFESTATIONS

Intractable skeletal pain typically begins gradually during middle age or later and then rapidly increases with a debilitating course and eventual immobility. Spontaneous fractures are a prominent complication. Physical examination reveals marked bone tenderness.

DIAGNOSIS

On radiography, only the skull is spared. Initially, osteopenia and a slightly abnormal appearance of trabecular bone are noted. Subsequently, the changes suggest osteomalacia. Corticomedullary junctions become indistinct as compact bone is replaced by an abnormal cancellous pattern. Generalized osteopenia causes the remaining spongy bone to appear coarse and dense, in a fishnet pattern of mixed lytic and sclerotic areas. Alkaline phosphatase activity in serum is increased.

The skeletal lesion is a localized form of osteomalacia that varies considerably in severity from area to area.

Hyperostosis

PROGRESSIVE DIAPHYSEAL DYSPLASIA (CAMURATI-ENGELMANN DISEASE)

PATHOBIOLOGY

Progressive diaphyseal dysplasia (Camurati-Engelmann disease) affects all races and is inherited as an autosomal dominant trait with variable expressivity. New bone formation gradually envelops both the periosteal and endosteal surfaces of long bone diaphyses. In patients with severe disease, osteosclerosis also occurs in the axial skeleton.

Mutations alter the gene that encodes transforming growth factor-β1. Osteoblast differentiation may be deranged.

CLINICAL MANIFESTATIONS

During childhood, limping or a broad-based, waddling gait is noted. Muscular dystrophy can be diagnosed erroneously. Severely affected individuals may have a characteristic body habitus featuring an enlarged head with a prominent forehead, proptosis, and thin limbs with little subcutaneous fat or muscle mass and tender, thickened bones. Cranial nerve palsies and raised intracranial pressure can occur. Some patients have hepatosplenomegaly and Raynaud phenomenon. Symptoms may remit after puberty.

DIAGNOSIS

Irregular hyperostosis of the diaphyses of the major long bones slowly develops as a result of periosteal and endosteal new bone formation. The femur and tibia are most commonly affected. Metaphyses may be involved. The age at onset, rate of progression, and severity are variable. Clinical, radiographic, and bone scan findings are generally concordant. Serum alkaline phosphatase activity, biochemical markers of skeletal turnover, and erythrocyte sedimentation rate may be elevated. Histopathologic study reveals newly formed woven bone that matures and becomes incorporated into cortical bone. Electron microscopy of muscle may show myopathic changes and vascular abnormalities.

TREATMENT Rx

Glucocorticoid therapy (typically a low dose of prednisone on alternate days) can relieve bone pain and may normalize skeletal histology. Bisphosphonates are sometimes useful.

ENDOSTEAL HYPEROSTOSIS

PATHOBIOLOGY

Sclerosteosis and van Buchem disease, autosomal recessive disorders, are the most severe types of endosteal hyperostosis. Sclerosteosis is caused by deactivating mutations in a gene called *SOST*. Van Buchem disease involves a deletion downstream of *SOST*. Enhanced osteoblast activity from sclerostin deficiency, with failure of osteoclasts to compensate for the increased bone formation, leads to the skeletal changes.

CLINICAL MANIFESTATIONS

Sclerosteosis (cortical hyperostosis with syndactyly) occurs primarily in individuals of Dutch ancestry. The gender distribution appears to be equal. Patients are tall and heavy beginning in childhood; have a prominent, square mandible; and are deaf and experience facial nerve palsy from cranial nerve entrapment. Raised intracranial pressure and headache may reflect a small cranial cavity that can shorten life expectancy. Van Buchem disease causes progressive asymmetrical enlargement of the jaw during puberty. Patients may be symptom free, or, beginning as early as infancy, they may have recurrent facial nerve palsy, deafness, and optic atrophy from narrowing of cranial foramina. Long bones may hurt with applied pressure but are strong.

DIAGNOSIS

In sclerosteosis, the skeleton is radiographically normal in early childhood except for syndactyly, which is common and most often involves the index and third fingers. Progressive bone thickening widens the skull and causes prognathism. Osteosclerosis involves the skull base, facial bones, vertebrae, pelvis, and ribs. Endosteal thickening homogeneously widens diaphyseal cortices and narrows medullary canals. Computed tomography has shown fusion of ossicles and narrowing of the internal auditory canals and cochlear aqueducts. Serum alkaline phosphatase activity can be increased from enhanced bone formation.

TREATMENT Rx

Surgical decompression of narrowed foramina may alleviate cranial nerve palsies.

PACHYDERMOPERIOSTOSIS

PATHOBIOLOGY

Pachydermoperiostosis (hypertrophic osteoarthropathy, primary or idiopathic) is an autosomal dominant disorder that features clubbing of the digits, hyperhidrosis with thickening of the skin (especially of the face), and periosteal new bone formation, most prominently in the distal ends of the limbs. Not all patients manifest all three principal features.[2] A loss-of-function mutation is found in the gene that encodes 15-hydroxyprostaglandin dehydrogenase. Autosomal recessive inheritance also seems to occur.

CLINICAL MANIFESTATIONS

Men seem to be more severely affected than women, and blacks are affected more commonly than whites. Symptoms typically begin during adolescence, intensify during the next decade, but then become quiescent. Arthralgia and fatigue are common. Stiffness and limited mobility occur in both the appendicular and the axial skeleton. Clubbing, with slowly progressive enlargement of the hands and feet, results in a pawlike appearance. Cutaneous changes include thickening, furrowing, pitting, and oiliness, especially of the scalp and face.

DIAGNOSIS

Periostitis thickens the distal portions of the tibia, fibula, radius, and ulna. Clubbing is obvious, and acro-osteolysis can occur. Periosteal proliferation is exuberant, with an irregular texture, and it often involves the epiphyses, whereas secondary hypertrophic osteoarthropathy (pulmonary or otherwise) typically causes a smooth and undulating periosteal reaction. Ankylosis of the joints, especially in the hands and feet, may trouble older patients. Bone scanning reveals symmetrical, diffuse, regular uptake along the cortical margins of long bones, especially in the legs—the double stripe sign.

TREATMENT Rx

Patients with painful synovial effusions may respond to nonsteroidal anti-inflammatory drugs. Contractures or neurovascular compression by osteosclerotic lesions may require surgical intervention.

OSTEOSCLEROSIS WITH HYPEROSTOSIS

Osteopetrosis

DEFINITION

Osteopetrosis (marble bone disease) is a group of rare disorders characterized by an increase in bone mass due to dysfunction in or lack of osteoclasts during growth. There are two major clinical categories: the autosomal recessive or "malignant" type, which often results in death by early childhood if untreated; and the autosomal dominant or "benign" type, which causes lesser complications. Autosomal recessive types can also feature intermediate severity, neuronal storage disease, or renal tubular acidosis with cerebral calcification due to carbonic anhydrase II deficiency. Bisphosphonate-induced osteopetrosis has been reported.

PATHOGENESIS

The defective gene causing autosomal dominant osteopetrosis encodes a chloride channel important for osteoclast activity. Bi-allelic mutations in this gene, or ones that encode components of a vacuolar proton pump, result in malignant osteopetrosis. Carbonic anhydrase II deficiency is caused by deactivating mutations in the gene that encodes this isoenzyme. Especially rare autosomal recessive cases involve deficient osteoclastogenesis from loss-of-function mutations within the genes for either receptor activator of nuclear factor κB (RANK) or its ligand (RANKL).

Histopathologic studies show that all true forms of osteopetrosis feature profound deficiency of osteoclast action. Bone embedded primary spongiosa (calcified cartilage deposited during endochondral bone formation) persists away from growth plates and constitutes the pathognomonic finding. Defective endosteal bone resorption impairs the formation of marrow space. Quiescent skeletal remodeling leads to bone fragility from the diminished interconnection of osteons and the delayed conversion of immature (woven) bone to mature (compact) bone. Neuronal storage disease (ceroid-lipofuscin) may reflect a lysosomal defect. Deficient superoxide production (necessary for bone resorption) has been considered a pathogenetic factor.

CLINICAL MANIFESTATIONS

Malignant osteopetrosis can be manifested during infancy as nasal "stuffiness" from underdeveloped mastoid and paranasal sinuses. Small cranial foramina may cause optic, oculomotor, or facial nerve palsy. Failure to thrive, delayed dentition, and fractures are common. Hypersplenism and recurrent infection, bruising, and bleeding reflect myelophthisis. Short stature, large head, frontal bossing, nystagmus, hepatosplenomegaly, and genu valgum are characteristic physical features. Untreated children usually die during the first decade of life of hemorrhage, pneumonia, severe anemia, or sepsis. Benign osteopetrosis can cause fracture, facial palsy, deafness, mandibular osteomyelitis, bone marrow failure, impaired vision, psychomotor delay, carpal tunnel syndrome, or osteoarthritis. Carbonic anhydrase II deficiency can cause failure to thrive, fracture, developmental delay, mental subnormality, and short stature. Cerebral calcification develops during childhood, but defective skeletal modeling and osteosclerosis may resolve. Both proximal and distal renal tubular acidosis have been described.

DIAGNOSIS

A generalized increase in bone density is the radiographic hallmark of osteopetrosis. In severe disease, modeling defects in long bones produce an "Erlenmeyer flask" deformity (Fig. 248-2). Alternating dense and lucent bands commonly occur in the metaphyses and pelvis. The cranium is usually thickened and dense, especially at the base, and the paranasal and mastoid sinuses are underpneumatized. Vertebrae may show, on a lateral view, a "bone-in-bone" (endobone) configuration or end-plate sclerosis causing a "rugger jersey" appearance. Skeletal scintigraphy can disclose fractures and osteomyelitis. Magnetic resonance imaging helps monitor the response to bone marrow transplantation.

Serum levels of acid phosphatase and creatine kinase (brain isoenzyme) are often increased. In malignant osteopetrosis, hypocalcemia with secondary hyperparathyroidism and elevated serum concentrations of calcitriol can accompany radiographic changes that resemble rickets. In benign osteopetrosis, biochemical indices of mineral homeostasis are typically unremarkable, although serum parathyroid hormone levels may be elevated.

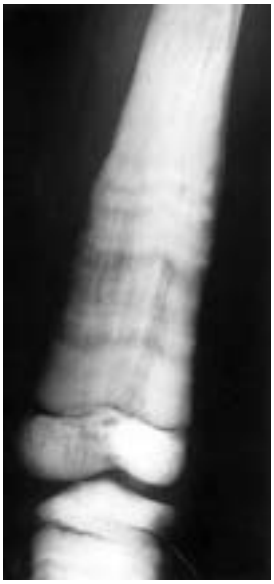

FIGURE 248-2. Osteopetrosis. Anteroposterior radiograph of the distal end of the femur shows a widened metadiaphyseal region, with characteristic alternating dense and lucent bands. (From Whyte MP, Murphy WA. Osteopetrosis and other sclerosing bone disorders. In: Avioli LV, Krane SM, eds. *Metabolic Bone Disease.* 2nd ed. Philadelphia: WB Saunders; 1990.)

TREATMENT

Rx

Because the origin, molecular pathogenesis, prognosis, and treatment of the more than 10 types of osteopetrosis can differ, precise diagnosis is crucial.[3] Now, commercially available mutation analysis can delineate most patients. For the malignant form, prompt use of human leukocyte antigen–identical bone marrow transplantation to supply functional osteoclasts has remarkably benefited some children. Calcium-deficient diets have been tried but may be limited by hypocalcemia and rickets and have uncertain efficacy. Pharmacologic doses of calcitriol (1,25-dihydroxyvitamin D_3) administered orally, together with dietary calcium restriction (to prevent hypercalciuria and hypercalcemia) or human interferon-γ, believed to enhance superoxide production, have reportedly stimulated osteoclast activity. Prednisone alone or with a low-calcium, high-phosphate diet can sometimes be effective. Glucocorticoid therapy stabilizes pancytopenia and hepatosplenomegaly. Hyperbaric oxygenation helps treat osteomyelitis. Surgical decompression of optic and facial nerves can be beneficial.

Pyknodysostosis

EPIDEMIOLOGY

Pyknodysostosis is believed to have affected French impressionist painter Henri de Toulouse-Lautrec (1864-1901). Most descriptions have come from Europe and the United States, but the disorder seems to be especially common in Japan.

PATHOBIOLOGY

This autosomal recessive condition is caused by loss-of-function mutations in *CTSK*, the gene that encodes cathepsin K.[4] Consequently, bone collagen degradation and skeletal turnover are diminished. In chondrocytes and osteoblasts, abnormal inclusions have been described.

CLINICAL MANIFESTATIONS

Characteristic features of pyknodysostosis seen during infancy or early childhood are relatively large cranium, fronto-occipital prominence, proptosis, bluish sclerae, beaked and pointed nose, small facies and chin, obtuse mandibular angle, high-arched palate, dental malocclusion with retention of primary teeth, and disproportionate short stature. Cranial sutures remain open. Fingers are short and clubbed from acro-osteolysis or aplasia of the terminal phalanges, and the hands are small and square. Repeated fractures cause knock-knee deformity. Mental retardation is noted in approximately 10% of patients. Adult height ranges from 4 feet 3 inches to 4 feet 11 inches. Life expectancy can be shortened by recurrent respiratory infections and

right-sided heart failure from chronic upper airway obstruction secondary to micrognathia.

DIAGNOSIS

Osteosclerosis is uniform, first becoming apparent in childhood and increasing with age. Skeletal modeling defects do not occur, although long bones appear to have thick cortices because of narrow medullary canals. Clavicles are gracile and hypoplastic at their lateral segments. The calvarum and base of the skull are sclerotic, orbital ridges are dense, and wormian bones are present.

TREATMENT Rx

No effective medical therapy has been documented. Fractures of the long bones usually mend satisfactorily. Internal fixation of long bones is formidable because of their narrow medullary space and bone hardness. Tooth extraction is difficult. Osteomyelitis of the mandible may require antibiotic, surgical, or hyperbaric therapy.

Hepatitis C–Associated Osteosclerosis

Rarely, achy and tender limbs develop periodically in individuals who are infected with hepatitis C virus. Radiographic studies reveal a marked generalized increase in bone mass from osteosclerosis and hyperostosis. Disturbances in the insulin-like growth factor system may explain the enhanced bone formation. Calcitonin or bisphosphonate therapy to slow bone turnover or antiviral treatment has benefited some patients.

FOCAL OSTEOSCLEROSIS/HYPEROSTOSIS

Osteopoikilosis

Osteopoikilosis ("spotted bones") is generally a radiographic curiosity due to a deactivating mutation of the *LEMD3* gene transmitted as a highly penetrant autosomal dominant trait. The bone lesions are usually asymptomatic. However, incorrect diagnosis may lead to confusion with serious conditions, including metastatic disease.[5] Some patients have connective tissue nevi called *dermatofibrosis lenticularis disseminata* (or Buschke-Ollendorff syndrome). On radiologic examination, numerous small, round or oval foci of bone sclerosis appear in cancellous bone in the tarsal, carpal, pelvic, and metaepiphyseal regions of tubular bones.

Osteopathia Striata

This finding is usually an autosomal dominant curiosity of asymptomatic linear striations in the metaphyseal regions of long bones and in the ilium. However, two clinically important X-linked dominant disorders with osteopathia striata affect predominantly females: osteopathia striata with cranial sclerosis due to mutation of the *WTX* gene[6]; and osteopathia striata with widespread linear areas of dermal hypoplasia and various bone defects in the limbs due to mutation of the *PORCN* gene (Goltz's syndrome).

Melorheostosis

DEFINITION

Melorheostosis is a sporadic disorder that features bone changes with the appearance of wax that has dripped down a candle. No mendelian basis has been established. The anatomic distribution suggests a postzygotic segmentary defect.

CLINICAL MANIFESTATIONS

Monomelic involvement is usually noted; bilateral disease is generally asymmetrical. Cutaneous changes over affected bones are common (e.g., linear scleroderma-like areas and hypertrichosis) and often appear before the hyperostosis. Symptoms typically begin during childhood, with pain and stiffness as the major complaints. Joints may become contracted and deformed from ectopic bone. Leg length inequality results from soft tissue contractures and premature fusion of epiphyses. Skeletal changes seem to progress most rapidly throughout childhood. During adult life, melorheostosis may or may not gradually spread, but pain is especially common.

DIAGNOSIS

As seen radiographically, irregular, dense, eccentric periosteal and endosteal hyperostosis affects a single bone or several adjacent bones. The lower limbs are most commonly involved. Endosteal thickening predominates during infancy and childhood, and periosteal new bone formation is prominent during adulthood. Ectopic bone formation may occur, particularly near joints.

TREATMENT Rx

Surgical correction of contractures is difficult. Recurrent deformity is common.

Mixed Sclerosing Bone Dystrophy

This typically sporadic disorder features combinations of osteopoikilosis, osteopathia striata, melorheostosis, cranial sclerosis, and other skeletal aberrations in one individual. Complications derive from the specific types of osteosclerosis or hyperostosis, such as nerve palsy with cranial sclerosis and bone pain with melorheostosis.

OTHER DISORDERS OF BONE

Fibrous Dysplasia

This sporadic developmental disorder features one or more expansile fibrous lesions within bone. Polyostotic disease is typically seen before the age of 10 years; monostotic disease begins in adolescence or early adulthood. *McCune-Albright syndrome* refers to polyostotic fibrous dysplasia, cafe au lait spots (Fig. 248-3), and endocrine hyperfunction.[7]

PATHOBIOLOGY

Fibrous dysplasia and McCune-Albright syndrome are caused by postzygotic mosaicism for an activating mutation in the gene that encodes the α subunit of the receptor subunit/adenylate cyclase–coupling G protein, *GNAS*. Imperfect bone forms because mesenchymal cells do not fully differentiate to osteoblasts.

CLINICAL MANIFESTATIONS

Monostotic fibrous dysplasia is more common than polyostotic disease. The skull and long bones are affected most often. The skeletal lesions can deform bones, cause fractures, and occasionally entrap nerves. Sarcomatous degeneration is rare (<1%) but typically occurs within the facial bones or femur and is more frequent with polyostotic disease. Pregnancy may reactivate quiescent lesions. McCune-Albright syndrome usually causes pseudo–precocious puberty in girls. Less commonly, one sees pseudo–precocious puberty in boys. There can also be thyrotoxicosis, Cushing's disease,

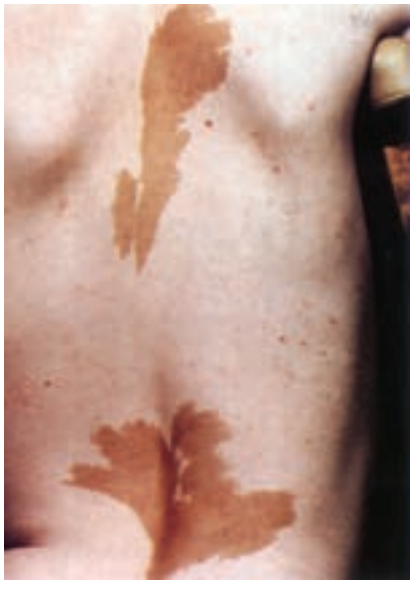

FIGURE 248-3. McCune-Albright syndrome. Typical rough-border ("coast of Maine"), pigmented cafe au lait spots. (From Whyte MP. Metabolic and dysplastic disorders. In: Coe FL, Favus MJ, eds. *Disorders of Bone and Mineral Metabolism*. New York: Raven Press; 1992.)

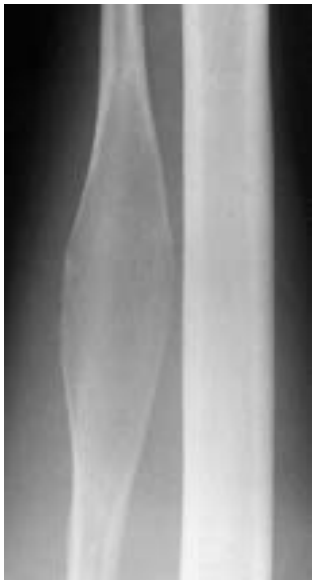

FIGURE 248-4. Fibrous dysplasia. A characteristic expansile lesion with a ground-glass appearance has caused thinning of the cortex in the mid-diaphysis of the fibula. (From Whyte MP. Fibrous dysplasia. In: Favus MJ, ed. *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism.* 3rd ed. Philadelphia: Lippincott-Raven; 1996.)

acromegaly, hyperprolactinemia, or hyperparathyroidism. In some patients, acquired renal phosphate wasting causes hypophosphatemic rickets or osteomalacia.

DIAGNOSIS

The skeletal lesions have a characteristic radiographic appearance. In the long bones, they are found in either the metaphysis or diaphysis, typically are well defined with thin cortices, and have a ground-glass appearance (Fig. 248-4). With aging, the defects can become lobulated, with trabeculated areas of radiolucency.

TREATMENT

Rx

In patients with mild disease, bone lesions may not expand. In severe cases, individual defects can progress, and new ones may appear, during childhood. Spontaneous healing does not occur, but pathologic fractures generally mend well. Stress fractures, however, can be difficult to detect and to treat. When the skull is involved, nerve compression may require surgical intervention. In McCune-Albright syndrome, search for and pharmacologic control of associated endocrinopathies are often important. Bone antiresorptive treatment has helped some patients.

Hereditary Multiple Exostoses

This relatively common, highly penetrant, autosomal dominant disorder features irregular bone excrescences that protrude from expanded metaphyses. Mutations have been identified in the *EXT1* and *EXT2* genes. Osteocartilaginous exostoses arise from growth plates and increase in size until linear growth ceases. Lesions may or may not become detached from the parent bone. Their structure is relatively unremarkable, with an outer cortex and an inner spongiosa. Disability results primarily from limb length discrepancies when linear bone growth suffers at the expense of transverse expansion. Compression of nerves, the spinal cord, or the vascular system occurs occasionally. Sarcomatous degeneration (0.5 to 2% of patients) should be suspected when an exostosis enlarges rapidly, especially in an adult.

Enchondromatosis (Dyschondroplasia, Ollier Disease)

This sporadic disorder features cartilaginous masses within the trabecular bone that arise from growth plates. The condition begins in childhood with localized swelling and interferes with linear bone growth. At puberty, expansion of cartilage masses ceases, and these lesions can be replaced by mature bone. Enchondromas appear radiographically as lucent defects in flat bones or in metaphyses of tubular bones, often with central calcific stippling. When enchondromatosis occurs together with multiple hemangiomas (Maffucci syndrome), the enchondromas or hemangiomas undergo malignant transformation in 15% of cases. Ollier disease and Maffucci syndrome are caused by somatic mosaic mutations in the *IDH1* and *IDH2* genes.

Achondroplasia

Chondrodystrophies are disorders of cartilage growth that result in disproportionate short stature. Achondroplasia is the most common. A defect occurs in the gene that encodes fibroblast growth factor receptor type 3; 80% of cases are the result of new autosomal dominant mutations, which are more prevalent with increasing paternal age. Short tubular bones form because of abnormal endochondral ossification in the limbs. In the chondrocranium, membranous ossification is undisturbed; hence the skull vault is normal. However, the cranial base and foramen magnum are small. The head is large, with frontal bossing and midface hypoplasia. Lumbar lordosis is greatly exaggerated, and the spinal canal narrows from the upper to lower segments of the vertebral column. This disturbance is revealed radiographically by a decreasing interpediculate distance. The trunk length is relatively normal, but the limbs show rhizomelic shortening, and the hands have a trident configuration. The long bones appear massive because of their disproportionately normal width. Growth plates are not grossly disorganized, and chondrocytes appear normal. Complications can include hydrocephalus and compression of the brain stem, spinal cord, or nerve roots. Minimal impingement by a disc or osteophyte on the small spinal canal can cause neurologic disturbances.[8] Despite these problems, achondroplasia is compatible with good health and a normal lifespan.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Sigua-Rodriguez EA, da Costa Ribeiro R, de Brito AC, et al. Bisphosphonate-related osteonecrosis of the jaw: a review of the literature. *Int J Dent*. 2014;2014:192320.
2. Pineda C, Martinez-Lavin M. Hypertrophic osteoarthropathy: what a rheumatologist should know about this uncommon condition. *Rheum Dis Clin North Am*. 2013;39:383-400.
3. Sobacchi C, Schulz A, Coxon FP, et al. Osteopetrosis: genetics, treatment and new insights into osteoclast function. *Nat Rev Endocrinol*. 2013;9:522-536.
4. Arman A, Bereket A, Coker A, et al. Cathepsin K analysis in a pycnodysostosis cohort: demographic, genotypic and phenotypic features. *Orphanet J Rare Dis*. 2014;9:60.
5. Ng C, Schwartzman L, Moadel R, et al. Osteopoikilosis: a benign condition with the appearance of metastatic bone disease. *J Clin Oncol*. 2014;[Epub ahead of print].
6. Fujita A, Ochi N, Fujimaki H, et al. A novel WTX mutation in a female patient with osteopathia striata with cranial sclerosis and hepatoblastoma. *Am J Med Genet A*. 2014;164A:998-1002.
7. Lietman SA, Levine MA. Fibrous dysplasia. *Pediatr Endocrinol Rev*. 2013;10(suppl 2):389-396.
8. Hecht JT, Bodensteiner JB, Butler IJ. Neurologic manifestations of achondroplasia. *Handb Clin Neurol*. 2014;119:551-563.

249 APPROACH TO THE PATIENT WITH ALLERGIC OR IMMUNOLOGIC DISEASE

STEPHEN I. WASSERMAN

Allergic diseases and disorders of the immune system affect multiple organ systems and may arise in a variety of manners. The reader is directed to Section VII for a detailed discussion of the immune system and specific autoimmune and acquired immune disorders. This chapter addresses allergic disorders, the most common manifestation of immune system dysfunction, and primary immune deficiencies, which are uncommon manifestations of immune dysfunction. For clarity, these two issues are treated separately.

ALLERGIC DISEASE

DEFINITION

Allergic disorders are common, and their prevalence is increasing, particularly in urbanized, Western societies. It is said that allergic diseases are the most common disorders seen by primary care physicians. Moreover, even in nonallergic patients, consideration of allergy frequently enters the differential diagnosis of a problem under consideration. Therefore, an appreciation of how to approach the diagnosis and treatment of allergic patients is of major importance to the practice of internal medicine. Allergic disorders are those caused by the interaction of a sensitized host (one who has made immunoglobulin E [IgE] antibody recognizing a specific antigen) with a specific allergen. Not all patients possessing specific IgE antibody react adversely on interaction with the allergen, and such individuals are sensitized but not allergic. The primary allergic conditions are seasonal allergic rhinoconjunctivitis (hay fever), perennial allergic rhinitis/sinusitis, asthma, anaphylaxis (especially secondary to foods, medications, and hymenopteran stings), urticaria or angioedema, atopic dermatitis (eczema), and food allergy.

EPIDEMIOLOGY

It is currently estimated that more than 50% of the population is atopic (i.e., able to mount an IgE immune response and to exhibit an immediate positive prick-puncture hypersensitivity response to common aeroallergens). Clinically, 10 to 20% of the general population will develop allergic rhinoconjunctivitis, 5 to 7% will have active asthma, and 20% will experience urticaria at some time. The incidence of allergic rhinitis and asthma is increasing worldwide and most rapidly in those areas with prior low incidence of these disorders.

The increase in allergic diseases noted in the past two decades is thought to result from better hygienic conditions, decreases in infant and childhood infections, and increasingly sedentary and indoor lifestyle with its attendant exposures to indoor allergens and risk for obesity. These changes appear to be associated with a less effective activation of the innate immune system, thereby altering the protective maturation of the acquired immune system. The immune bias in utero and in infancy is toward a type 2 helper T-lymphocyte (T_H2)–directed immune response, which is the immune pathway required for the expression of allergic disease. Ineffective generation of regulatory T lymphocytes underlies the genesis and persistence of allergy. It is therefore postulated that without sufficient early childhood immunologic exposure (e.g., with infection) to induce a switch to an effective and protective T_H1 immune response, allergic disease is more likely to emerge during childhood. Substantial epidemiologic evidence has been gathered to support this concept, now termed the hygiene hypothesis. Thus, allergy is more prevalent in individuals of higher socioeconomic status, among those living in urban areas, in less polluted communities (e.g., western Germany), in first-born children compared with later siblings, in multiply immunized individuals, and in those free of mycobacterial disease. Conversely, children living on farms, in rural communities, and in more highly polluted areas as well as children with mycobacterial infection and those who have experienced multiple early childhood infections are less likely to develop allergic disorders. A concentration-effect relationship appears to exist between exposure to endotoxin (as a marker for hygiene) and the incidence of allergic sensitization. Low and very high levels of exposure to endotoxin are associated with abnormal immune maturation and allergic expression, whereas moderate levels of exposure predispose to a nonallergic phenotype.

PATHOBIOLOGY

The persistence or aberrant activation of T_H2 lymphocytes leads to the generation of cytokines (e.g., interleukins-4, -5, -13), which stimulate B-lymphocyte synthesis of IgE antibody and the production of eosinophilic polymorphonuclear leukocytes. The expression of allergic disorders results from the interaction of a specific allergen with allergen-reactive IgE bound to high-affinity receptors on mast cells and basophils. This interaction leads to the activation of these target cells and to their release of preformed, granule-associated mediators (exemplified by histamine); the synthesis of lipid mediators from membrane lipids (sulfidopeptide leukotrienes); and the transcription and secretion of cytokines, including tumor necrosis factor-α and interleukins-4, -5, and -13. These mediators directly induce smooth muscle contraction, vascular dilation, and endothelial leakage; they also cause vascular adhesion molecule expression, and they attract and activate inflammatory leukocytes, particularly $CD4^+$ T lymphocytes, basophils, and eosinophils. These and other IgE-dependent mediators are thought to be responsible for stimulating smooth muscle proliferation and tissue remodeling.

DIAGNOSIS

Allergy is a systemic immune disorder, so its expression can be multifocal. It is essential to remember this fact when examining a patient with suspected allergic problems because a focus on only the major presenting symptom may be insufficient to identify all the pertinent medical issues in a given patient.

History

Allergic disease has a high degree of heritability, with a great degree of concordance in identical twins. The risk of expressing allergic disease is highest if both parents are atopic. The inheritance of specific manifestations of allergy and of the specific allergen to which a patient is sensitized is less simple. Often, the diagnosis of allergic disorders is straightforward and can be made by asking about the nature of the patient's complaints, when and where reactions occur, and what exposures the patient believes are relevant to symptom induction or exacerbation (Table 249-1).

Seasonal and Perennial Rhinitis

Patients with seasonal and perennial rhinitis (Chapter 251) commonly present with complaints of itchy nose and palate; sneezing; watery rhinorrhea; itching, watery, and burning eyes; and nasal obstruction, which, when severe, may cause anosmia. In the evaluation of possible causes of seasonal rhinoconjunctivitis or sinusitis, the time of the year when symptoms occur is pertinent. Symptoms may be associated with the pollination of trees (early spring), grasses (late spring and summer), or weeds (fall). Important geographic differences exist in allergen concentration and exposure; for example, there is little ragweed in the western United States, and the grasses present in New England and Florida are quite different. In some patients with perennial symptoms, the multiple overlapping pollen seasons are responsible for their symptoms. However, indoor exposures at home, school, work, or recreational sites to furred animals, dust mites or insects, and mold should be addressed in the search for additional causes of perennial symptoms. Mold and mites are to be expected in humid environments, and mites are nearly ubiquitous in bedding and in homes with pets, carpeting, and overstuffed furnishings. Additional occupational or recreational exposures may be pertinent in selected situations (e.g., bakers, health care workers, food handlers, horse fanciers, laboratory animal handlers) in which specific inciting allergens may be identified. Because many patients with rhinitis have concomitant asthma, it is important to obtain information about the presence of this disease in patients with rhinitis. The concurrence of asthma and rhinitis has led to the concept of a "single airway" in which insults/responses in one part of the airway (i.e., upper) may be reflected in the other (i.e., lower).

Asthma

Patients with asthma (Chapter 87) may present with cough or wheeze with dyspnea, which is reversible spontaneously or with treatment. In addition to the association with rhinitis, the influence of exercise, exposure to tobacco smoke, effect of respiratory infection (particularly viral), occupational exposures (e.g., ≤30% of atopic animal handlers develop asthma), and medication use (e.g., β-adrenergic blocking drugs) are pertinent. Because most patients with asthma have concurrent rhinitis, it is essential that the physician evaluate this issue in all asthmatic patients. Wheezing may accompany other

TABLE 249-1 SYMPTOMS, SIGNS, AND TREATMENT OF ALLERGIC DISEASE

SYMPTOMS AND SIGNS	APPROACH TO TREATMENT
SYMPTOMS	
Cutaneous: itch, rash	H_1-antihistamine
Ocular: gritty sensation, itch	Topical H_1-antihistamine or mast cell stabilizing agent
Upper respiratory: palatal pruritus, clear rhinorrhea, sneeze, nasal obstruction	Topical corticosteroid, oral H_1-antihistamine, leukotriene receptor antagonist, topical nasal H_1-antihistamine
Lower respiratory: wheeze, cough, dyspnea	β_2-Agonist, inhaled corticosteroid, inhaled β_2-agonist, leukotriene receptor antagonist, oral methylxanthine, parenteral corticosteroid, parenteral anti-IgE
Gastrointestinal: nausea, vomiting, cramping pain	Epinephrine (if caused by anaphylaxis), oral corticosteroid, oral cromolyn
SIGNS	
Cutaneous: flushing, urticaria, angioedema, eczema	
Ocular: conjunctival erythema, chemosis	
Upper respiratory: pallor, edema, clear rhinorrhea, polyps	
Lower respiratory: wheeze	

disorders, including pulmonary edema in congestive heart failure and pulmonary embolic disease.

Urticaria and Angioedema

Patients with urticaria (Chapter 252) describe pruritic, erythematous cutaneous lesions with regular or irregular borders occurring anywhere on the body; they may vary in size from small (1 × 1 mm) to extremely large. Skin lesions are often preceded by intense intertriginous pruritus and erythema. Individual urticarial lesions generally persist for a few hours and rarely last for more than 24 hours.[1] However, many disorders can cause a sensation of itching; for more on skin and systemic diseases associated with pruritus (see Chapter 436). Angioedema (Chapter 252) is most frequently appreciated in the face, hands, and other soft tissues and is generally accompanied by symptoms of stretching, tingling, and tightness of the skin rather than by pruritus. Lesions, especially in the face, typically last 24 to 36 hours. Although most cases of urticaria or angioedema are not IgE-allergen mediated (particularly in chronic urticaria), it is important to identify foods and medications used by patients with acute urticaria or angioedema, particularly those substances ingested within 2 to 4 hours of the development of lesions, and to inquire about insect stings. Chronic urticaria is less often IgE mediated; questions about medications, especially nonsteroidal anti-inflammatory drugs, recent infection (especially with Epstein-Barr virus), and the presence of autoantibodies to the IgE receptor must be addressed. Approximately one half to two thirds of patients with such autoantibodies also have antibodies to thyroid antigens. In angioedema, the use of angiotensin-converting enzyme inhibitors must be sought, with special attention to those of African American heritage. Atopic dermatitis is another allergic cutaneous disorder in which patients complain of intense pruritus, especially in flexural surfaces. In adults, foods (IgE mediated) and cutaneous infection with *Staphylococcus aureus* (superantigen mediated) are the most commonly identified precipitating events for atopic dermatitis.

Anaphylaxis

Anaphylaxis (Chapter 253) is the most important allergic emergency and is potentially fatal. It is an acute allergic response associated with cutaneous (urticaria, angioedema, flushing), respiratory (laryngeal edema, asthma), cardiovascular (arrhythmia, hypotension, extravascular fluid loss), gastrointestinal (nausea, vomiting, abdominal pain, diarrhea), and nonspecific symptoms (metallic taste, sense of impending doom) that may occur singly or together. Historical information of note includes all medications, foods, and other encounters occurring within 2 hours of the reaction. Epidemiologic data suggest that foods (especially peanuts, tree nuts, shellfish, milk, and eggs), hymenopteran stings, and medications (antibiotics, muscle relaxants, radiocontrast media) are the most frequently identified causes of this important problem. Patients with elevated basal tryptase levels are at increased risk for anaphylaxis, especially with hymenopteran stings, and all patients suffering anaphylaxis after such an exposure should have baseline tryptase determined; if it is elevated, appropriate follow-up evaluation for occult mast cell disorders should be undertaken.

Food Allergy

Patients presenting with food allergy[2] often complain of oral pruritus and nausea, vomiting, diarrhea, and abdominal pain. Eczema, urticaria, and anaphylaxis, as noted previously, may also be consequences of food allergy. In general, allergic symptoms consequent to foods occur within minutes to 2 hours of ingestion of the causative food; delayed symptoms are unlikely to be mediated by an IgE-allergen interaction. Eosinophilic esophagitis has recently been added to the list of disorders in which food allergy is thought to play a role. Other symptoms attributable to foods are less easily explained by allergic mechanisms and are termed food intolerance.

Physical Examination

The physical examination of a patient with suspected allergic disease should emphasize the organ systems pertinent to the patient's complaints. The skin should be examined for the presence of urticarial or angioedematous lesions and for signs of atopic dermatitis, including flexural papules, excoriations, and lichenification. Keratosis pilaris, particularly on the outer aspect of the upper arm, commonly accompanies atopic dermatitis. Urticaria typically consists of small, pink, irregular lesions that blanch on pressure and then clear, leaving normal skin. In a patient with urticaria, a simple test for dermatographism should be undertaken. Angioedematous lesions are larger, more diffuse, and pale, and they are found most often on the face and in acral areas.

The eyes, ears, nose, and throat should be examined in all patients thought to have allergic disease, particularly those whose symptoms suggest seasonal or perennial allergic rhinoconjunctivitis-sinusitis or asthma. In allergic disease, the conjunctivae are often injected and may be edematous. "Cobblestoning" of the epithelium may be present. The periorbital tissues may be swollen and darkened. Examination of the nares may show pale and edematous nasal mucous membranes and swollen turbinates, and polyps may be seen. Secretions, generally clear, may be seen in the nasal passages or in the posterior pharynx. Such secretions generally contain copious numbers of eosinophils (see later), and their absence is a point against an allergic cause. Fever and discolored secretions, particularly those that are thick and yellow or green, in the presence of neutrophilic polymorphonuclear leukocytes suggest infection. Percussion over the maxillary or frontal sinuses may elicit tenderness in acute sinusitis. In chronic sinusitis, the physical examination may be unrevealing. In acute otitis media, patients may have erythema and bulging or perforation of the tympanic membrane, with fluid in the canal; in chronic cases, the drum may be scarred and retracted. Alteration in airborne conduction may be noted as well.

Patients with acute asthma may display tachypnea and auditory wheezes, and they may be unable to speak in full sentences because of shortness of breath. Use of accessory muscles of respiration and evidence of cyanosis should be sought. Examination of the chest includes inspection for evidence of chronic hyperinflation and auscultation for wheezing (which, if unilateral, may suggest a foreign body or tumor). In mild asthma, the examination findings may be normal, or the only physical finding may be wheezing on forced expiration and a slight prolongation of the expiratory phase.

Patients experiencing acute anaphylaxis usually demonstrate flushing, and concomitant urticaria and angioedema are often present. Assessment of vital signs may disclose hypotension and tachycardia. In some situations, hoarseness or stridor related to laryngeal edema or wheezing secondary to asthma can be identified. Hyperactive bowel sounds may be noted. Progressive hypoxia and cyanosis may ensue. In severe anaphylaxis, cardiovascular collapse secondary to hypoxia and hypotension may result in death.

Laboratory Evaluation

In the evaluation of patients with allergic disorders, the laboratory may be of assistance in both the identification and the quantification of specific organ dysfunction as well as in the assessment of the presence and specificity of IgE antibody.[3]

Assessment of Total and Allergen-Specific Immunoglobulin E

Essentially all (>95%) IgE antibody is bound to specific high-affinity receptors on tissue mast cells and circulating peripheral blood basophils. The small amount of free serum IgE antibody circulates in nanogram quantities and can be identified only with techniques of sufficient sensitivity. A large proportion

TABLE 249-2 ADVANTAGES AND DISADVANTAGES OF DIFFERENT ALLERGY TESTING METHODS

METHOD	PATIENT SELECTION	CLINICAL ADVANTAGES	CLINICAL DISADVANTAGES
Skin testing	Clinical indication suggesting allergic disease	Rapid (15-30 min) turnaround Sensitive and specific; prick-puncture for aeroallergens; prick-puncture followed by intradermal testing for drugs, sera, and venoms	Patient must not be taking H_1-antihistamine agents for 5-7 days Not interpretable in the presence of dermatographism Requires sufficient normal skin to enable testing
In vitro testing	Clinical indication suggesting allergic disease	Antihistamine therapy not contraindicated Dermatographism not a problem Sensitive and specific; equal to prick-puncture skin testing	Requires blood to be drawn Slow turnaround (7-14 days)

of IgE in a given individual may be directed toward a single antigen. Therefore, total IgE levels may be normal in the presence of allergic disease, and the measurement of total serum IgE is rarely of help in making a diagnosis. In a few situations, such as adult atopic dermatitis or allergic bronchopulmonary aspergillosis, measurement of total serum IgE levels may provide insight into disease severity or risk of disease exacerbation.

Of more importance is the identification of allergen-specific IgE in a patient with suspected allergic disease (Table 249-2).[4] Such specific IgE may be identified in vitro or in vivo. A search for allergen-specific IgE is particularly useful in the evaluation of patients with suspected allergic rhinitis, asthma, eczema, food reactions, and anaphylaxis. In vitro assessment is similar to the quantification of total IgE, except the initial capture reagent bound to a solid phase is a specific allergen from pollen, mold, insect, venom, food, or other material. Development of the assay is identical to that used to quantify total IgE, and results are generally reported in a semiquantitative manner. The magnitude of the reaction is weakly correlated with the expression of allergy, although for certain foods, more precise correlative data exist on the risk of allergy and the amount of allergen-specific IgE detected. To assess allergen-specific IgE in vivo, a minute quantity of the allergen in question is introduced into the skin by a prick-puncture technique, and the cutaneous response is assessed 15 to 30 minutes thereafter. A positive response is one in which a wheal and flare at least 2 mm larger than that caused by a saline control occur at the injection site. In vivo tests are rapid and inexpensive; their use requires the absence of dermatographism, that patients not be taking antihistamine medications, and that patients exhibit a positive response to a control with histamine. In some situations (e.g., penicillin allergy or hymenopteran sting), a more diluted allergen is injected intradermally, and the wheal and flare response is assessed similarly. The presence of allergen-specific IgE antibody and a clear temporal correlation between exposure to allergen and genesis of symptoms are required to conclude that a patient is allergic to a specific allergen. In the absence of symptoms, a patient with allergen-specific IgE is termed sensitized but not allergic.

Specific in vivo challenge tests can also be used to identify allergen responsiveness. Such tests in the presence of specific IgE antibody may be useful in research settings, or they may be used clinically to clarify the exact relationship between exposure and symptoms. The tests can be dangerous, however, because they introduce the allergen to which the patient is presumed allergic. In the case of food allergy, such challenges are best done in a double-blind and placebo-controlled manner; they may be useful to distinguish allergy from sensitization or to eliminate a suspect food from consideration. However, food challenge tests are unnecessary in the case of anaphylaxis and a positive test response for IgE antibody to the putative allergen. Because many patients falsely believe that foods are responsible for their symptoms, such double-blind challenges may be useful in directing patients' concerns to more productive areas. Inhalation tests employing specific allergens or chemicals have been helpful in elucidating some cases of occupational allergy or asthma.

Other Laboratory Aids in Allergic Disease

In a patient with acute asthma, chest radiographs generally demonstrate hyperinflation. In some instances, evidence of bronchiectasis may be present, a finding that raises the specter of allergic bronchopulmonary aspergillosis. The presence of a tumor or radiopaque foreign body may be noted on a chest radiograph and should be sought in a patient with unilateral localized wheezing. In the examination of a patient with asthma, assessment of both airflow and volumes can provide a clear picture of the severity of asthma and its response to treatment. Flow-volume loops can also identify the presence of vocal cord dysfunction or tracheomalacia. When patients with airway obstruction are evaluated, their response to an inhaled β_2-adrenergic agonist medication (or a short-acting anticholinergic agent) can be helpful in elucidating the reversible nature of their disorder. A large majority of asthmatic patients exhibiting bronchoconstriction display a bronchodilatory response to the inhalation of such agents. In suspected cases of asthma in which pulmonary function is normal, a histamine or methacholine challenge can be performed. These agents take advantage of the nonspecific bronchial hyperresponsiveness that is characteristic of patients with asthma. Failure to develop bronchoconstriction on inhalation of either of these agents strongly argues against the diagnosis of asthma.

Other laboratory tools may be of clinical benefit in the identification and classification of allergic disorders. Audiometry may clarify the degree of hearing loss caused by otitis media in a patient with allergic rhinitis. When sinusitis is suspected, computed tomography of the sinuses provides the most complete image and has the highest degree of sensitivity for the identification of mucosal thickening, opacification of air spaces, and presence of polyps and bone erosions. Computed tomography is particularly useful in the examination of the ethmoid and sphenoid sinuses, which are often affected in chronic allergic disease and are difficult to assess on physical examination or with plain radiographs. Imaging studies are not indicated, however, in most cases of acute sinusitis.

The quantification of blood, sputum, nasal mucus, or tissue eosinophilia and the response to corticosteroid therapy are useful correlates in the identification and management of allergic disease. The quantification of tryptase, a mast cell–specific protease with a serum half-life of 2 hours, can assist in the diagnosis of anaphylaxis if it is performed on serum or plasma obtained within hours of a systemic response with associated hypotension. Quantification of exhaled nitric oxide has been used to assess airways inflammation in asthma and primarily reflects eosinophilic inflammation.

IMMUNOLOGIC DISEASE

EPIDEMIOLOGY

Diseases related to disordered immune function (immunodeficiency) are far less common than allergic disorders.[5] The most frequent is IgA deficiency, which occurs in approximately 1 in 1000 individuals and is often asymptomatic. Next most frequent are disorders of B and T lymphocytes, such as common variable hypogammaglobulinemia, and other disorders, including DiGeorge syndrome and severe combined immunodeficiency (Chapter 250). Much less common are defects in neutrophil function or complement.

DIAGNOSIS

The clinical expression of immunodeficiency disorders is primarily infection, related to impaired host defense. Thus, the diagnosis of suspected immunodeficiency involves the evaluation of recurrent, persistent, severe, unusual, and otherwise unexplained infections. Many but not all immune disorders arise in early childhood, and with improved management, many patients presenting in childhood live into adulthood. It is important for the general internist and internal medicine subspecialist to be cognizant of the presentation of these disorders.

History

The most important historical information includes the following: age at onset of the problem in question; family history of frequent infection or death at an early age from infection; number, site, and type of infections; and presence of other physical abnormalities (Table 249-3). The earlier the onset of infections, the more severe the immune defect is likely to be. T-lymphocyte defects, with or without B-cell deficiencies, usually are manifested in the first 3 to 5 months of life, whereas B-cell function is supported by maternal antibody until after 6 months of age. Many of the immune disorders are X-linked,

TABLE 249-3 KEY POINTS REGARDING IMMUNOLOGIC DISORDERS

ANTIBODY DEFICIENCY DISORDERS
Onset after 6 months of age
Recurrent respiratory infection
Infection with bacteria, especially encapsulated organisms
Absence of isohemagglutinins
Evaluation of B-cell function, not numbers
CELLULAR IMMUNE DEFECTS
Onset before 6 months of age
Recurrent viral, fungal, or parasitic (opportunistic) infection
Defective delayed hypersensitivity skin responses
Malabsorption or diarrhea
COMPLEMENT DEFICIENCIES
Recurrent bacterial infection
Recurrent neisserial infection (deficiency of late components)
Associated rheumatic disorder (especially systemic lupus erythematosus)
FACTORS SUGGESTING NEUTROPHIL DYSFUNCTION
Late separation of umbilical cord
Persistent neutrophilic leukocytosis
Recurrent or persistent gingivitis or periodontitis
Recurrent bacterial infection with granuloma formation

and a careful family history is critical in such situations. Infection-related death of a patient's male sibling or the patient's mother should lead to the search for such an X-linked disorder.

In a patient with a T-cell disorder, viral, fungal, mycobacterial, and other opportunistic infections (*Pneumocystis jiroveci, Toxoplasma gondii*) are most commonly noted, and live virus vaccination may be associated with disseminated and progressive viral disease. Persistent thrush, diarrhea, malabsorption, and failure to thrive occurring in early childhood may suggest the presence of T-cell abnormalities.

In B-cell or antibody deficiency, pyogenic bacterial infections predominate, particularly infections involving encapsulated microorganisms. Such infections usually affect the upper and lower respiratory tract and the skin and are severe, recurrent, and often persistent. Infections with unusual organisms, with unexpected complications, or involving multiple sites (lung, sinus, joint, bone, or meninges, with abscess formation or sepsis) should raise the index of suspicion. In adults, the most common disorder in this class is termed common variable immunodeficiency.

As in any patient with infection, information should be sought about exposure to ill individuals or to irritants such as tobacco smoke, the hygiene of the environment to which the patient has been exposed, and the presence of an anatomic abnormality or allergy that could predispose to infection.

Physical Examination

Physical examination beyond that necessary to assess the extent and severity of a particular infection should focus on immune organs. Assessment of tonsillar tissue and determination of the presence and size of lymph nodes, spleen, and liver are important. Patients with common variable immunodeficiency often present with hepatosplenomegaly and lymph node hyperplasia, whereas in X-linked hypogammaglobulinemia, lymph tissue is absent. Telangiectasia (ataxia-telangiectasia), cardiac defects (DiGeorge syndrome), chronic eczema (Wiskott-Aldrich syndrome), and chronic periodontitis (neutrophil defects) all suggest immunodeficiency syndromes.

Laboratory Evaluation

The proper use of the laboratory is essential to elucidate a suspected immunodeficiency disorder.[6] Screening tests appropriate to the generalist's initial approach include complete blood count, total neutrophil and lymphocyte enumeration, quantitative immunoglobulin levels, and assessment of isohemagglutinins (especially when common variable immunodeficiency is suspected). In some situations, quantification of IgG subclasses may be warranted to identify a specific subclass deficiency. In considering T-lymphocyte defects, it is important to enumerate total T cells and specific T-cell subsets. Delayed hypersensitivity skin testing to recall antigens is also helpful in assessing cellular immunity. When neutrophil defects are suspected, a nitroblue tetrazolium test or measurement of phagocytic potency can be performed. Complement defects are best addressed by obtaining a CH_{50} level. CH_{50} is the amount of the patient's serum required to cause lysis of 50% of test erythrocytes. It is compared with the amount of pooled normal serum required to cause the same degree of lysis. Tests for specific individual components of complement, or of complement regulatory proteins, can also be obtained under special circumstances (e.g., late complement components in patients with recurrent *Neisseria* species infection).

Additional tests of antibody production in response to defined stimuli, including vaccinations, may be helpful when selective antibody deficiency is suspected or when borderline immunoglobulin levels are encountered in the presence of frequent infection. In some situations, assessment of T-cell proliferation to mitogens or antigen may be of benefit. Further testing might include the assessment of natural killer-cell function and the production of cytokines by activated lymphocytes. In general, such additional laboratory tests should be performed in consultation with an expert in immune disorders.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Zuberbier T, Aberer W, Asero R, et al. The EAACI/GA(2) LEN/EDF/WAO Guideline for the definition, classification, diagnosis, and management of urticaria: the 2013 revision and update. *Allergy*. 2014;69:868-887.
2. Boyce JA, Assa'ad A, Burks AW, et al. Guidelines for the diagnosis and management of food allergy in the United States: summary and the NIAID-sponsored expert panel report. *Nutr Res*. 2011;31: 61-75.
3. Sastre J. Molecular diagnosis in allergy. *Clin Exp Allergy*. 2010;40:1442-1460.
4. National Clinical Guideline Centre (UK). *Drug Allergy: Diagnosis and Management of Drug Allergy in Adults, Children and Young People*. London: National Institute for Health and Care Excellence (UK);2014.
5. Al-Herz W, Natarangelo LD. Classification of primary immunodeficiency disorders: one size fits all does not help anymore. *Clin Immunol*. 2012;144:21-25.
6. Ochs HD, Hagin D. Primary immunodeficiency disorders: general classification, new molecular insights, and practical approach to diagnosis and treatment. *Ann Allergy Asthma Immunol*. 2014;112: 489-495.

REVIEW QUESTIONS

1. Which of the following young adults is most likely to suffer from hay fever, seasonal rhinoconjunctivitis, atopic eczema, and dermatitis?

A. Raised in a wealthy family with access to excellent preventive health care
B. Born, raised, and living on a farm
C. Grew up as a "sickly child" with multiple ear infections, upper respiratory tract infections, and other infections
D. During childhood, had a mother with active TB that was being treated at home "for a long time with pills"
E. Youngest of five children whose parents did not allow him to have childhood immunizations

Answer: A This question relates to the concept known as the hygiene hypothesis, which postulates that decreased microbial stimulation early in life may lead to delayed maturation of the immune system, which in turn results in an increased risk of allergic diseases later in life. Although the situation with asthma specifically appears to be more complex, other allergic diseases are more prevalent in individuals of high socioeconomic status, in first-born children compared with later siblings (especially in large families), in multiply immunized individuals, and in those free of mycobacterial disease. In contrast, children raised on farms are less likely to develop allergic disorders later in life, possibly because of their early exposure to diverse microbial environments in animal sheds, haylofts, harvesting, and other farm activities. See Epidemiology.

2. Which of the following symptoms is characteristic of an anaphylactic reaction?

A. Angioedema
B. Metallic taste
C. Abdominal pain
D. Sense of impending doom
E. All of the above

Answer: E Anaphylaxis, the most important and a potentially fatal allergic emergency, can be manifested with any single one or a combination of a wide variety of symptoms, including cutaneous (urticarial, angioedema, flushing), respiratory (stridor from laryngeal edema, wheezing), cardiovascular (hypotension, arrhythmia, extravascular fluid loss), gastrointestinal (vomiting, abdominal pain, diarrhea), or nonspecific symptoms (metallic taste in mouth, sense of impending doom). See Anaphylaxis under the History section of Diagnosis.

3. Diagnosis of a specific allergic disorder requires

A. No laboratory testing if the clinical history is characteristic
B. Increased serum level of total IgE
C. Increase in serum allergen-specific IgE
D. Increase in serum allergen-specific IgE coinciding with symptoms
E. Demonstration of dermatographism

Answer: D The presence of allergen-specific IgE antibody and a clear temporal correlation between exposure to allergen and the genesis of symptoms are required to conclude that a patient is allergic to a specific allergen. In the absence of symptoms, a positive test response for allergen-specific IgE alone is termed sensitization but not allergy. Total levels of IgE in serum may be normal in the presence of allergic disease, and therefore its measurement is rarely helpful in making a diagnosis. See Assessment of Total and Allergen-Specific Immunoglobulin IgE under Laboratory Evaluation.

4. A young woman is found unresponsive, lying on a sidewalk. She is brought to the emergency department, where she is resuscitated but found to be in respiratory failure, requiring intubation, and in shock. No previous history is known, and the cause of her shock is not clear. Which of the following tests can diagnose anaphylactic shock in this case?

A. Increased serum IgE level
B. Eosinophilia
C. Increased serum tryptase level
D. Histamine or methacholine challenge
E. None of the above is useful without a clinical history

Answer: C The quantification of tryptase, a mast cell–specific protease with a half-life of 2 hours, can be helpful to diagnose anaphylaxis if it is performed on serum or plasma obtained within hours of a systemic reaction with associated hypotension. Histamine or methacholine challenge is not even feasible under these circumstances. The other blood tests are not sufficiently sensitive or specific.

5. In the evaluation of a patient with frequent infections for a possible immunodeficiency disorder, her history reveals late separation of the umbilical cord at birth, recurrent and persistent gingivitis and periodontitis, and recurrent bacterial infections with granuloma formation. Which of the following categories of immunologic disorders is she most likely to have?

A. Cellular immune defect
B. Neutrophil dysfunction
C. Complement deficiency
D. IgA deficiency
E. Severe variable immunodeficiency

Answer: B These findings along with persistent neutrophilic leukocytosis are characteristic of disorders of neutrophil dysfunction. Infections in individuals with cellular immune defects are more likely to be viral, fungal, or parasitic (opportunistic) in etiology, and they may also have malabsorption and diarrhea. Complement deficiencies are typically associated with recurrent bacterial infections, especially neisserial infections, and rheumatic diseases like systemic lupus erythematosus. IgA deficiency and other antibody deficiency disorders are characterized by recurrent respiratory infections with encapsulated organisms.

*Note: Dr. Wasserman had no role in the development of these questions.

250

PRIMARY IMMUNODEFICIENCY DISEASES

CHARLOTTE CUNNINGHAM-RUNDLES

Since the descriptions of the first genetic immune defects, severe combined immunodeficiency (SCID) and X-linked agammaglobulinemia (XLA), in the 1940s, the number of known primary immune defects (175+) has expanded exponentially.[1] To keep pace, every few years the International Union of Immunological Societies has compiled these defects into eight general categories: T- and B-cell combined defects, B-cell/antibody defects, complement disorders, phagocyte defects, defects with syndromic features, diseases of immune dysregulation, autoinflammatory syndromes, and defects of innate immunity (Table 250-1).[2] Autoinflammatory syndromes are covered in Chapter 261. In this chapter, our current understanding of these other primary immune defects is considered with emphasis on immune defects found in adults. Complement and phagocyte disorders are discussed in more detail in Chapters 50 and 169, respectively.

AN APPROACH TO EVALUATION OF THE IMMUNE SYSTEM

Because of the numbers and types of immune deficiency, recognition of the many clinical phenotypes can be difficult, leading in many cases to a delayed diagnosis. The spectrum of immune defects found in populations varies with the age of the patient and with the combined T- and B-cell immune defects, defects of phagocyte function, defects with syndromic features, defects of immune dysregulation, and innate defects; selected infections are more commonly recognized in childhood, whereas defects of complement and antibody production are more characteristic of adults. However, there are many exceptions to this general observation; in addition, even if an immune defect has been diagnosed in childhood, adequate treatment has allowed these patients to increasingly appear in the offices of internists and adult specialists.

For most patients, the first symptom of an immune defect is a series of relatively common infections, particularly involving the respiratory tract. These may include chronic sinusitis, otitis, and bacterial pneumonia. For adults with immune defects, infections are likely to last longer, are likely to require additional courses of antibiotics, and tend to recur. Infections may lead to additional complications or procedures, such as empyema after bacterial pneumonia or the need for myringotomy tubes in an adult with chronic otitis. For infants and children, chronic infections lead to poor appetite and an obvious failure to grow normally; for adults, some weight loss may occur, but it is less apparent. Because of lack of immunity, shingles (Chapter 375)

is relatively common in patients with T-cell defects or lack of antibody. Other common clinical presentations include acute gastrointestinal infections with characteristic organisms such as *Giardia* (Chapter 351) and chronic intestinal inflammatory diseases with weight loss and malabsorption (Chapter 140). General guidelines to approach the laboratory evaluation of the main immune defects outlined in this chapter, based on clinical presentations,[3] are provided in Table 250-2. Flow charts of the work-up of immune defects in Table 250-1, based on clinical phenotypes, have been published.[4]

TABLE 250-1 CATEGORIES OF PRIMARY IMMUNODEFICIENCY DISEASES

T- and B-cell combined deficiencies
Antibody deficiencies
Complement disorders
Phagocyte defects
Well-defined defects with syndromic features
Immune dysregulation syndromes
Autoinflammatory defects
Defects of innate immunity

T- AND B-CELL COMBINED DEFECTS

DEFINITION

Combined immune defects are diseases in which both the T- and B-cell compartments are greatly impaired. With the very early onset and severe nature of these defects, this group contains all forms of SCID and other syndromes in which both T- and B-cell limbs of the immune system are markedly abnormal.

EPIDEMIOLOGY

The incidence of SCID has undergone a downward revision from the estimated 1 : 100,00 a few years ago because of newborn screening, currently performed for about half of all newborns in the United States.[5] The aggregate incidence of these forms, listed in Table 250-3, appears to be about 1 : 54,000. When the larger spectrum of combined immune defects is also included, the incidence is likely to be higher.

PATHOBIOLOGY AND GENETICS

The hallmark of combined defects is that they eliminate or greatly impair T-cell development, in most cases leading to profound lymphopenia. Infants

TABLE 250-2 CLINICAL PRESENTATION AND EVALUATION OF THE IMMUNE SYSTEM

CLINICAL PRESENTATION	DEFECTS	IMMUNE DEFECTS	CONDITIONS	LABORATORY TESTING
Recurrent or chronic bacterial, viral, or fungal infections Opportunistic infections	Cell-mediated immunity	Impaired killing of intracellular organisms Impaired viral immunity Hypogammaglobulinemia	SCID	Absolute lymphocyte count Enumeration of T cells and T-cell subsets Proliferative tests for T-cell function
Bacterial infections Viral infections Autoimmunity Inflammatory diseases Enteropathy Giardiasis	B cells	Hypogammaglobulinemia Impaired bacterial killing Impaired clearance of virus or toxins	Hypogammaglobulinemia Agammaglobulinemia IgA deficiency CVID IgG subclass defects Antibody deficiency	Enumeration of B cells Serum IgG, IgA, and IgM Antibody testing (i.e., tetanus, diphtheria) Vaccine challenge and antibody testing (pneumococcal vaccine)
Bacterial infections Susceptibility to meningococcal disease Autoimmunity Angioedema	Complement	Impaired opsonization Impaired bacterial killing Lack of clearance of immune complexes	Complement C2 deficiency Other complement defects HAE	CH_{50} AH_{50} Measuring individual components C1 inhibitor protein and function
Bacterial infections Poor skin healing Fungal infections Stomatitis Periodontal disease	Phagocytic cells	Impaired neutrophil mobilization Impaired opsonization Impaired bacterial killing	Chronic neutropenia Cyclic neutropenia Autoimmune neutropenia LAD CGD	Absolute neutrophil counts Neutrophil oxidative burst examined by dihydrorhodamine test by flow cytometry Examination of the blood smear Antineutrophil antibodies

CGD = chronic granulomatous disease; CVID = combined variable immunodeficiency; HAE = hereditary angioedema; LAD = leukocyte adhesion deficiency; SCID = severe combined immunodeficiency.

TABLE 250-3 EXAMPLES OF COMBINED IMMUNE DEFECTS

SCID TYPE	GENES	INHERITANCE	LABORATORY FEATURES	DISEASE AND COMPLICATIONS
Defects of V(D)J recombination	*RAG1*, *RAG2* ARTEMIS DNA-PKcs	AR	Very low lymphocyte numbers with loss of T and B cells; hypogammaglobulinemia	Severe infections, failure to thrive; leaky versions may have autoreactive T cells (Omenn syndrome)
Adenosine deaminase deficiency	Adenosine deaminase	AR	Variably low lymphocyte numbers with loss of T and B cells; also decreased NK cells; hypogammaglobulinemia	Severe infections, failure to thrive; often with costochondral junction flaring, neurologic features, hearing impairment, lung and liver manifestations
X-linked SCID	Common γ chain of cytokine receptors	XL	Low lymphocyte numbers with loss of T cells; B cells present; markedly decreased NK cells; hypogammaglobulinemia	Severe infections, failure to thrive; leaky cases may present with low T or NK cells or Omenn syndrome
JAK3 deficiency	*JAK3*	AR	Low lymphocyte numbers with loss of T cells; B cells present; hypogammaglobulinemia	Severe infections, failure to thrive; leaky cases may present with variable T or NK cells
IL-7 deficiency	α chain of the IL-7 receptor	AR	Low lymphocyte numbers with loss of T cells; B cells present; normal NK cells; hypogammaglobulinemia	Severe infections, failure to thrive; leaky cases may present with low T or NK cells or Omenn syndrome
T-cell receptor chain defects	γ-, ε-, and ζ-chain mutations	AR	Low lymphocyte numbers due to loss of T cells; normal B and NK cells; hypogammaglobulinemia	Severe infections, failure to thrive; leaky cases may present with low T or NK cells or Omenn syndrome

AR = autosomal recessive; DNA-PKcs = DNA-dependent protein kinase catalytic subunits; IL-7 = interleukin-7; JAK3 = Janus kinase 3; NK = natural killer; SCID = severe combined immunodeficiency; XL = X-linked.

with defects that affect the formation of T- and B-cell receptors, such as defects of the recombinase activating genes *RAG1* and *RAG2*, which impair VDJ recombination, have few if any T and B cells. Similarly, other defects of DNA recombination or repair genes (ARTEMIS, DNA-PKcs) will have a similar phenotype. When T-cell immunity is absent, B cells may be present, but they will have no function. This is the case for the most common form of SCID, the X-linked form, due to mutations in the cytokine γ chain, the signaling component of six cytokine receptors: interleukin (IL)–2, IL-4, IL-7, IL-9, IL-15, and IL-21. Similarly, defects of the *JAK3* gene, downstream from the cytokine γ chain, or of the IL-7 receptor itself lead to a similar immune profile.

CLINICAL MANIFESTATIONS

With loss of both essential limbs of the adaptive immune system, infants with combined immune defects have severe and recurrent infections due to bacteria, viruses, and fungi. Other common features include diarrhea, dermatitis, and failure to thrive. Clinically, most patients present before the age of 3 months, but there is a significant number of exceptions as infants may present later, although still usually in the first year of life. Without intervention, SCID usually results in severe infections and death by the age of 2 years. In some cases, the immune defect is such that a few T cells can develop, but these are often self-reactive; these cases are often termed leaky SCID. When the presentation of these cases includes rashes and evidence of autoimmunity, infants are said to have Omenn syndrome.

DIAGNOSIS

Newborns normally have an absolute lymphocyte count of 4000/mm^3 or higher. Thus, the first measure in suspected SCID is the complete blood count; most infants have significant lymphopenia. When the clinical manifestations suggest SCID, a flow cytometer panel is used to enumerate T, B, and NK cells; this will also suggest genes that may be responsible. Further genetic testing can be done, but only after stem cell transplant approaches have been launched.[6] A number of states have recently introduced newborn screening for SCID, based on standard blood spot (Guthrie) cards to determine if the DNA signature of new T cells emigrating from the infant thymus can be detected in normal numbers. This method has been shown to be both highly sensitive and specific and is likely to be universally adopted.

TREATMENT AND PROGNOSIS

Rx

Without immune reconstitution, infants with SCID will die, and prompt recognition is essential. Early reconstitution with stem cells from human leukocyte antigen (HLA)–matched bone marrow or mobilized peripheral blood is mandatory. When the diagnosis is made early and no severe infections have occurred, hematopoietic stem cell transplantation (HSCT; Chapter 178) is likely to be curative in 90% of cases. Gene therapy in some cases is gaining a new role and will likely lead to additional options in the future.

ANTIBODY DEFECTS

DEFINITION

The antibody defects are due to loss of B-cell development, loss of production of one or more of the immunoglobulin (Ig) isotypes, or loss of functional antibody production.

EPIDEMIOLOGY

As a group, antibody defects are the most prevalent immune defects and are found in patients of all ages. Selective IgA deficiency is most common in patients of white background, but the incidence varies with the population studied from 1 : 400 to more than 1 : 10,000. IgA deficiency is found in 1 : 400 in Finland but much less commonly in African Americans or Asians (1 : 14,000 or fewer). Common variable immune deficiency (CVID) has an estimated incidence of 1 in 50,000; IgG subclass or selective antibody defects are also common, but the incidence is unknown.

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

Whereas the genetic causes have been elucidated for many of the combined forms of immune deficiency, the genes are not yet known for many of the more common B-cell defects (Table 250-4). Antibody defects can be considered in three main forms: B cells are absent; B cells are present but one or more immunoglobulin isotypes is not made; and B cells and immunoglobulin levels are normal but the produced immunoglobulins have no function.

Lack of B Cells Leading to Agammaglobulinemia

The first severe antibody defect described was the X-linked form of agammaglobulinemia (XLA). The gene affected, a tyrosine kinase (BTK) located on the X chromosome, is essential for downstream signaling from the B-cell receptor. Without these signals, B cells do not survive, leading to profound hypogammaglobulinemia. The incidence of this disease is approximately 1 in 100,000. The main clinical manifestations appear in the first year of life, but males may come to clinical attention later, in some cases not until the second decade. Whereas X-linked inheritance is a central feature, the family history may or may not be positive because of de novo mutations. Infections are usually bacterial, generally with encapsulated organisms like *Streptococcus pneumoniae, Haemophilus influenzae, Staphylococcus aureus,* and *Pseudomonas* species. A particular propensity for *Mycoplasma* infections in XLA has been long noted; these may occur in joints or the urinary tract and can be difficult to diagnose as appropriate culture techniques are not widely available.

In addition to XLA, there are other genetic forms of agammaglobulinemia. These are gene defects of the B-cell receptor itself, such as the μ heavy chain, the surrogate light chain λ5, Igα, and Igβ. Similarly, mutations in signaling proteins immediately downstream from the B-cell receptor lead to the same outcome, with loss of all B cells. As these genes are not on the X chromosome, these defects, although rare, are found in both sexes.

Good Syndrome

A special case of agammaglobulinemia with loss of B cells in adults is a poorly understood immune defect associated with thymomas (Good syndrome). This appears to be a secondary immune defect, but it is important to include it here as the loss of B cells, with either agammaglobulinemia or hypogammaglobulinemia, leads to many of the same infectious manifestations as with the other profound antibody defects. Quite rare, Good syndrome occurs in adults, most often after the age of 40 years, and includes an increased incidence of opportunistic infections, such as *Pneumocystis jiroveci, Candida* infections with nail or other cutaneous involvement, viral infections, and inflammatory complications such as lichen planus. The connection between thymoma, loss of B-cell function, and the unusual infections remains unclear.

Hypogammaglobulinemia with B cells

Common Variable Immune Deficiency

Patients with CVID have varying degrees of hypogammaglobulinemia ranging from almost total loss of immunoglobulins to more modest reductions of IgG and IgA or IgM. From the clinical point of view, CVID is a noteworthy disorder as it is relatively common (1 : 25,000 to 1 : 50,000), has a later onset than other immune defects (generally between the ages of 20 and 40 years), and has a highly heterogeneous clinical presentation. Delays in diagnosis are common. Before diagnosis, about 80% of subjects with CVID will have had one or more episodes of pneumonia, sometimes leading to empyema. Over time, bronchiectasis may develop. The bacterial species most commonly found include *S. pneumoniae, H. influenzae, S. aureus,* and *Mycoplasma* species. The gastrointestinal tract is not uncommonly involved; this may be infectious (e.g., *Giardia, Campylobacter,* norovirus) or inflammatory, including lymphoid hyperplasia and forms of inflammatory bowel disease leading to malabsorption. However, a biopsy will show loss of plasma cells in the gastrointestinal mucosa. About one quarter of subjects with CVID have autoimmune conditions such as thrombocytopenia, hemolytic anemia, achlorhydria, pernicious anemia, and granulomatous disease suggesting sarcoidosis.[7] Clinically, lymphadenopathy is common, and splenomegaly is noted in 28%. Malignant disease is also increased, usually B-cell lymphomas, but other cancers appear more common as well. In the past decade, some of the gene mutations responsible for a minority of these cases have been discovered and are included in Table 250-4. However, at this time, most subjects with hypogammaglobulinemia do not have a known gene defect.

Hyperimmunoglobulin M Syndromes

The hyperimmunoglobulin M (hyper-IgM) syndromes are defects in which there is a loss of isotype switch; that is, whereas B cells do produce IgM, they do not secrete IgG or IgA. The prototypic form is the X-linked version in which an essential T-cell activation receptor, the CD40 ligand encoded on the X chromosome, is missing or nonfunctional. Mutations in the gene for its receptor partner, the CD40 on B cells, lead to a similar defect. Several

TABLE 250-4 EXAMPLES OF ANTIBODY DEFECTS

TYPE	GENES	INHERITANCE	LABORATORY FEATURES	DISEASE AND COMPLICATIONS
ABSENT B CELLS: SEVERE REDUCTIONS IN IgG, IgA, AND IgM				
X-linked agammaglobulinemia	Bruton tyrosine kinase	XL	IgG, IgA, and IgM are very low or absent	Severe bacterial infections
Autosomal forms of agammaglobulinemia	Defects of the B-cell receptor or its signaling pathways; λ5, Igα, Igβ	AR	IgG, IgA, and IgM are very low or absent	Severe bacterial infections
Good syndrome	Unknown	Unknown	Variable hypogammaglobulinemia	Associated with thymoma; may have opportunistic infections
B CELLS PRESENT BUT LOW SERUM IgG, IgA, AND/OR IgM				
Common variable immune deficiency	Unknown	Unknown	Low IgG, IgA and/or IgM	Bacterial infections and other inflammatory complications
ICOS deficiency	ICOS	AR	Low IgG, IgA and/or IgM	Recurrent infections; autoimmunity
Defects of the B-cell receptor	CD19, CD81, CD20, CD21	AR	Low IgG, IgA and/or IgM	Recurrent infections
Defects of other B-cell receptors	TACI, BAFFr, TWEAK	AD and sporadic	Variably low IgG, IgA and/or IgM; antibody defects	Recurrent infections and autoimmunity; variable clinical expression
B CELLS PRESENT: SEVERE REDUCTION IN SERUM IgG AND IgA BUT NORMAL OR ELEVATED IgM				
X-linked hyper-IgM syndrome	CD40 ligand	XL	IgG and IgA decreased; IgM may be normal or increased; B-cell numbers may be normal or increased	Bacterial and opportunistic infections, neutropenia, autoimmune disease
CD40 deficiency	CD40	AR	Low IgG and IgA; normal or increased IgM	Bacterial and opportunistic infections, neutropenia, autoimmune disease
Defects of DNA recombination	AID and UNG	AR	IgG and IgA decreased; IgM increased	Bacterial infections; enlarged lymph nodes and germinal centers
B CELLS PRESENT: ISOTYPE DEFICIENCIES				
Selective IgA deficiency	Unknown	Unknown	IgA absent	Usually asymptomatic; allergies and autoimmunity may be more common
IgA with IgG subclass deficiency	Unknown	Unknown	Reduced IgA with decrease in one or more IgG subclass	Infections in some with loss of antibody
IgG subclass deficiency	Unknown	Unknown	Reduction in one or more IgG subclass	Asymptomatic in many; infections in some with loss of antibody
B CELLS PRESENT: NORMAL IgG, IgA, AND IgM				
Antibody deficiency	Unknown	Unknown	Normal serum immunoglobulins but no vaccine responses to protein and carbohydrate antigens or vaccines	May lead to recurrent infections

AD = autosomal dominant; AR = autosomal recessive; ICOS = inducible T-cell costimulator; XL = X-linked.

other genetic defects lead to a similar immunologic phenotype, including defects of the gene for enzyme activation-induced cytidine deaminase (*AICDA*) and uracil-DNA glycosylase, both important for DNA recombination. The complications of the hyper-IgM syndromes include bacterial infections, autoimmunity, and enteropathy similar to CVID but also *P. jiroveci* pneumonia, neutropenia, and unusual cancers. A predilection to infections with *Cryptosporidium* has been noted in hyper-IgM syndromes, which unfortunately can lead to irreversible liver disease.

IgA Deficiency

Selective IgA deficiency (IgA <7 mg/dL with other isotypes normal) is the most common of the primary immunodeficiency disorders, but most subjects are asymptomatic. Lack of infections in most subjects is generally ascribed to the overlapping and compensatory role of other immune functions, but this is not clearly understood. However, allergies, autoimmunity, increased serum IgE, asthma, rheumatoid arthritis, gluten intolerance, and inflammatory bowel disease are found more commonly in selective IgA-deficient subjects than in other populations.[8] Presumably owing to the loss of secretory IgA, *Giardia* infections may occur (Chapter 351). The treatments used in subjects with IgA deficiency are based on the clinical conditions found. Some subjects with IgA deficiency are IgG2 and IgG4 deficient, with loss of antibacterial antibody leading to severe infections and, in some cases, chronic lung disease.

IgG Subclass Defects

Another variable immune deficiency is represented by the IgG subclass defects. The incidence of these is difficult to determine, partly because laboratory normal ranges vary. The clinical consequences depend on how much antibody function is lost. Whereas there are structural differences in IgG isotypes, their functional roles have considerable overlap; thus, the importance of isotype defects can be controversial, especially if loss of antibody is not demonstrable. In adults, IgG3 deficiency appears to be the most common but is likely to have no significance. However, low IgG2 or IgG4, most often found in subjects with selective IgA deficiency, may lead to a profound deficiency of antibody production, especially to carbohydrate antigens such as contained in the pneumococcal vaccine.

Antibody Deficiency with Normal Immunoglobulins

More complex and heterogeneous is the loosely described defect termed antibody deficiency with normal serum immunoglobulins. The incidence is unknown; all ages are affected, but in general, children younger than 5 years are not included to allow transient forms of physiologic immune deficiency to resolve. Although B cells are present and there are normal levels of IgG, IgA, and IgM, these subjects do not form protective levels of serum antibodies after having an infection exposure or vaccination with protein or carbohydrate vaccines. In the most severe cases, even strong immunogens such as herpes zoster or tetanus vaccines are ineffective; in milder cases, the pneumococcal vaccine does not result in titers of antibody considered sufficient for protection.

DIAGNOSIS

The diagnosis of antibody defects is based on the laboratory tests of numbers of B cells, the serum immunoglobulin levels (IgG, IgA, and IgM), and an evaluation of a panel of vaccine responses to determine the levels of functional

antibody.[9] If B cells are absent and the levels of immunoglobulins are very low, further antibody testing is not required. For a young male with a family history of males with immune deficiency, the diagnosis of XLA or hyper-IgM can be investigated by flow cytometer (to determine numbers of B cells for XLA) or by genetic tests (hyper-IgM). For older subjects (generally older than 45 years), a thymoma may be sought by chest computed tomography, which may show a mass in the mediastinum. Most subjects with hypogammaglobulinemia will be found to have B cells in peripheral blood and some amount of serum IgG, IgA, or IgM. In these cases, the loss of functional antibody should be tested by commercial laboratories to determine if protective titers of antibody to common vaccine antigens (i.e., tetanus, diphtheria, *H. influenzae*, and pneumococci) can be detected. In some cases, revaccination may be needed to determine if a response occurs (tested again in 4 to 6 weeks). Most authorities use the laboratory-stipulated protective ranges for protein vaccines and for pneumococcal vaccination, usually 1.3 µg/mL for individual serotypes. When high levels of B cells are found in adults, the possibility of a clonal B-cell expansion should be considered (e.g., chronic lymphocytic leukemia). For subjects with IgG subclass defects or normal immunoglobulins levels, the use of a panel of antibody titers is also recommended to have a clear understanding of immune competence or immune defect.

TREATMENT

Rx

The essential treatment of significant IgG antibody defects is intravenous or subcutaneous immune globulin, usually given in doses of 400 to 600 mg/kg body weight per month. The intravenous forms are usually given every 3 or 4 weeks, the subcutaneous forms weekly or biweekly, depending on body weight. Indwelling ports are not required and are discouraged. Most patients also require occasional courses of antibiotics, chosen on the basis of culture results, at intervals dictated by clinical events. Referring to Table 250-4, the defects that require IgG replacement are those in which B cells are absent (XLA, other agammaglobulinemias, the hyper-IgM syndromes, IgG subclass defects with demonstrable loss of antibody function, and some cases of loss of antibody with normal immunoglobulins). Subjects with IgA deficiency do not require immune globulin replacement unless there is clear loss of antibody. For unclear reasons, some of the antibody defects have an increased incidence of autoimmune or inflammatory complications. These require treatments commonly prescribed for immunocompetent subjects but with a view to minimizing courses of immune suppressants. Immune cytopenias may be treated with rituximab with some success; splenectomy is to be avoided. As many subjects with antibody defects have experienced one or more bouts of pneumonia, lung functions may be abnormal and intermittent or prophylactic antibiotics required, but there is no consensus on the medications, dose, or intervals to use.

PROGNOSIS

The prognosis for subjects with antibody defects is variable and depends on the degree of the defect, the response to treatment, whether organ damage has occurred, and whether other complications develop. Subjects with loss of B cells have a pure B-cell defect; when they are diagnosed and treated early with sufficient immune globulin, the prognosis appears excellent. Subjects with selective IgA deficiency can be indistinguishable from age-matched healthy peers. CVID subjects with varying degrees of hypogammaglobulinemia ranging from almost total loss of immunoglobulins to more modest reductions of IgG and IgA or IgM often have additional complications, in some cases because the diagnosis has been delayed and pulmonary or other damage has occurred. Chronic lung disease, lymphoid hyperplasia, and gastrointestinal enteropathy may be difficult to treat, leading to increased morbidity. Improved survival to prior years is likely overall in CVID, but the inflammatory complications still present additional challenges.[10] For subjects with IgG subclass defects or antibody deficiency, with immune reconstitution if required, no increased morbidity or mortality is expected.

COMPLEMENT DISORDERS

DEFINITION

The complement system is a network of proteins that both amplify and control many actions of the immune system. It is generally considered to have three main branches, the classical, alternative, and lectin pathways; deficiencies of individual components lead to increased susceptibility to infections, autoimmunity, and inflammatory diseases (Table 250-5). For further details of these disorders, see Chapter 50.

EPIDEMIOLOGY

Complement C2 deficiency is found in 1 : 10,000 white subjects and usually in subjects with a conserved major histocompatibility complex haplotype due to a founder defect; more than 95% of C2-deficient individuals are homozygous for the same *C2* mutation. The other complement component defects are rare but found in unequal distribution in selected populations; C6 deficiency is more common in persons of African descent and C9 deficiency in Asians, with an estimated incidence of 0.036 to 0.095%. Disorders of members of these pathways are discussed here; deficiency of C1 inhibitor is discussed separately.

PATHOBIOLOGY

The classical pathway is triggered by interaction of the Fc portion of an IgG1, IgG2, IgG3, or IgM antibody with C1q, which subsequently engages C1r,

TABLE 250-5 EXAMPLES OF COMPLEMENT DEFECTS

TYPE	GENES	INHERITANCE	LABORATORY FEATURES	ALTERED FUNCTIONS	DISEASE AND COMPLICATIONS
C1q, C1r, C1s deficiency	*C1qA, C1qB, C1qC; C1r, C1s*	AR	Absent CH_{50} hemolytic activity	Loss of early complement activation; impaired dissolution of immune complexes; impaired clearance of apoptotic cells	Bacterial infections; SLE-like syndrome, rheumatoid disease, multiple autoimmune diseases, infections
C4 deficiency	*C4A* and *C4B*	AR	Absent CH_{50} hemolytic activity	Loss of early complement activation	Bacterial infections
C2 deficiency	*C2*	AR	Absent CH_{50} hemolytic activity	Loss of early complement activation	Bacterial infections; SLE-like syndrome, vasculitis, early atherosclerosis, polymyositis, glomerulonephritis
C3 deficiency	*C3*	AR	Absent CH_{50} hemolytic activity	Loss of classical and alternative pathways of complement activation	Life-threatening pyogenic infections; SLE-like disease; glomerulonephritis; atypical hemolytic-uremic syndrome
C5, C6, C7, C8 deficiency	*C5*	AR	Absent CH_{50} hemolytic activity	Loss of complement activation	Neisserial infections, SLE
C9 deficiency	*C9*	AR	Reduced CH_{50} and AP_{50} hemolytic activity	Partial loss of complement activation	Some *Neisseria* infections
C1 inhibitor deficiency	C1 inhibitor	AD	Activation of complement; low levels of C4 and C2	Loss of regulation of activities of complement C1	Angioedema

AD = autosomal dominant; AR = autosomal recessive; SLE = systemic lupus erythematosus.

C1s, C2, and C3, leading to activation of C4, C5, C6, C7, C8, C9, resulting in lysis of bacteria (discussed in Chapter 50). As opsonization of bacteria is essential for antibody function, patients with these defects have infections similar to those of subjects with loss of immunoglobulin. The alternative pathway is activated in an antibody-independent manner and involves opsonization of bacteria with subsequent involvement of C3 and the alternative pathway. The lectin pathway includes other serum binding proteins that coat bacteria or fungi, leading to downstream complement activation and the assembly of the membrane attack complex, the C7, C8, C9 components responsible for microbial lysis. The genes of the complement system are located on many chromosomes, and in general, the defects are autosomal recessive in inheritance, with one exception, defects of X-linked properdin. In addition to the three pathways of activation, the complement system also includes an even larger number of control proteins that, when genetically defective, also lead to severe infections, hemolytic-uremic syndrome, severe eclampsia, glomerulonephritis, thrombosis, and macular degeneration, which are outside the scope of this chapter.

CLINICAL MANIFESTATIONS

With genetic loss of the classical and alternative pathways, severe bacterial infections are likely; this is particularly true of subjects with defects of C3, which lies at the convergence of the three pathways. For unclear reasons, with loss of C6, C7, C8, and C9 or properdin, *Neisseria gonorrhoeae* or *Neisseria meningitidis* infections are more common. More complex but equally potent is the role that the complement proteins play in immune regulation. With loss of the early components of the classical system, C1q, C1r , C1s, C2, and C4, autoimmunity, especially systemic lupus erythematosus, is common; this complication is estimated at 93% for subjects with defects of C1q and 75% for defects of C4. Complement is important for clearing immune complexes and possibly apoptotic cells, potentially explaining this observation.

DIAGNOSIS

Deficiencies of complement are diagnosed by testing total serum hemolytic complement (CH_{50}) and the alternative hemolytic complement (AP_{50}). The CH_{50} tests for deficiencies in the classical pathway by determining whether the patient's serum can lyse antibody-coated sheep erythrocytes; this will be zero if the proteins of the classical pathway are defective. The AP_{50} tests for alternative pathway activity. Further testing usually includes measurement and function of individual serum complement proteins to determine the most applicable diagnosis. (Note that the most common reason for low levels in CH_{50} and AP_{50} is improper blood handling.)

TREATMENT Rx

There are no treatments for complement deficiencies. Whereas loss of these classical components may lead to severe clinical consequences, for C2 in particular but also for C4 and C5-C9, there may be no history of illness. Prompt antibiotic therapy for acute infections and control of autoimmunity are the important therapeutics. However, periodic immunizations with pneumococcal, *H. influenzae*, and meningococcal vaccines may be helpful to boost antibody titers to enhance bacterial clearance.

PROGNOSIS

The prognosis of complement defects is highly variable because of the clinical complications; also, most of these defects have been found in healthy subjects. However, for defects of the classical pathway, prompt recognition and treatment of bacterial infections and possibly preemptive vaccination with appropriate vaccines would be important. The prognosis for subjects with autoimmunity will depend on disease manifestation and response to treatment. Whereas C2 deficiency is commonly viewed as usually asymptomatic, some data suggest a higher incidence of premature arteriosclerotic heart disease.

C1 INHIBITOR DEFICIENCY

DEFINITION

C1inhibitor (C1 INH) deficiency is discussed in detail in Chapter 252. It is classified as either genetic or acquired. The genetic form is called hereditary angioedema (HAE), due to mutations of C1 INH, inherited as an autosomal dominant trait. However, approximately 25% of patients with HAE have a spontaneous mutation in C1 INH, with no family history. C1 INH defects are classified as type I, due to loss of C1 INH protein, or type II, in which the protein is present but nonfunctional. Another inherited, estrogen-dependent form of HAE due to mutations in factor XII has also been recognized. The acquired form of C1 INH deficiency may be due to interfering autoantibodies and often clinically resembles the genetic forms, except that it usually is manifested later in life.

PATHOBIOLOGY AND GENETICS

Defects of C1 INH are inherited as autosomal dominant traits, but the outcome in families may vary considerably, from no angioedema in some members to frequent episodes in others. The C1 INH protein is the main regulator of the early activation steps of the classical complement pathway. It binds to and inactivates C1r and C1s in the C1 complex of the classical pathway, and when insufficient amounts are present, the classical pathway can be activated. C1 INH protein also inhibits proteases in the bradykinin pathways, increasing bradykinin production and leading to angioedema. With loss of C1 INH activity, fluid leaks from the vascular endothelium, leading to localized edema.

CLINICAL MANIFESTATIONS

A diagnosis of C1 INH deficiency is suggested by a history of recurrent attacks of angioedema or in some cases (25%) episodes of recurrent abdominal pain due to edema. The edema is nonpitting, nonpruritic, and localized to one regional area, such as the lips and mouth, hands, genitals, or abdomen, leading to concern for acute colitis or appendicitis. The presence of urticaria with angioedema suggests an IgE-mediated process and is not typical. As these defects are inherited as autosomal dominant defects, a family history is likely but not always identified. The hereditary forms usually become clinically apparent in early adolescence, but angioedema may occur before or after, with the first episode appearing first in adult life. Attacks may be preceded by "prodromes" of erythema, nausea, or a tingling sensation. Although the cause of an attack onset may often be unpredictable, there are certain triggers of HAE, including dental work, accidents, surgery, emotional stress, and infection. Unlike allergic angioedema, angioedema from C1 INH deficiency does not respond to treatment with antihistamines or corticosteroids and resolves slowly during 3 to 5 days. Selected medications (estrogen and angiotensin-converting enzyme inhibitors) can exacerbate or cause attacks and are to be avoided.

DIAGNOSIS

The diagnosis is suggested by a low serum C4 level. In HAE I, there is low C1 INH protein; in HAE II, there is normal C1 INH protein but a low function. The acquired form is suggested by a decreased C4 level, low C1 INH protein and function, and decreased C1q. A diagnosis of this form should prompt further evaluation of an underlying malignant or autoimmune disease. Blood tests for either HAE or the acquired form should be repeated to confirm the diagnosis because falsely low C1 INH function levels may result from improper handling of blood.

TREATMENT Rx

The treatment of hereditary C1 INH deficiency includes management of the attacks and prophylaxis against angioedema. These are described in Chapter 252.

PHAGOCYTE DEFECTS

DEFINITION

Abnormalities of the phagocytic system are presented in detail in Chapters 167 and 169. They are categorized as neutropenia, abnormal neutrophil morphology, defective cell adhesion and migration, or defective microbial killing (E-Table 250-1). Neutropenia (Chapter 167) is defined as counts of less than 1000×10^9/L.

EPIDEMIOLOGY

Whereas neutropenia is a common laboratory finding (due to conditions described in Chapter 167), genetic defects impairing neutrophil development, adhesion, locomotion, or intracellular killing are rare. The most

E-TABLE 250-1 EXAMPLES OF PHAGOCYTE DEFECTS

TYPE	GENES	INHERITANCE	LABORATORY FEATURES	ALTERED FUNCTIONS	DISEASE AND COMPLICATIONS
DEFECTS OF NEUTROPHIL DIFFERENTIATION					
Severe congenital neutropenia	*ELANE, GFI1, HAX1*	AD or AR	Neutropenia	Impaired myeloid differentiation	Susceptible to infections; myelodysplasia
Cyclic neutropenia	*ELANE*	AR	Neutropenia at 21-day intervals	Impaired and cyclic myeloid differentiation	Susceptible to infections at intervals, periodontitis
DEFECTS OF MOTILITY					
Leukocyte adhesion deficiency (types 1-3)	*ITGB2* Adhesion protein (CD18)	AR	Leukocytosis	Impaired neutrophil adherence and chemotaxis	Delayed cord separation for type 1; skin ulcers, poor wound healing, periodontitis
Shwachman-Diamond syndrome	*SBDS*	AR	Pancytopenia	Impaired chemotaxis	Exocrine pancreatic insufficiency, chondrodysplasia; risk of leukemia
Chédiak-Higashi syndrome	*LYST*	AR	Neutrophils with giant inclusions; hair: pigment clumps	Impaired chemotaxis	Partial albinism, recurrent infections, late-onset primary encephalopathy, increased lymphoma risk
DEFECTS OF RESPIRATORY BURST					
Chronic granulomatous disease	gp9 ^{phox} p22^{phox} p47^{phox} p67^{phox} p40^{phox}	XL, AR	Normal or increased neutrophil counts but loss of respiratory burst	Impaired bacterial killing	Susceptible to bacterial and fungal infections

AD = autosomal dominant; AR = autosomal recessive; XL = X-linked.

common is chronic granulomatous disease (CGD), with an estimated incidence of 1 : 100,000 to 1 : 200,000.

PATHOBIOLOGY AND GENETICS

Circulating neutrophils are attracted to sites of inflammation by complement components C5a, chemokines, and bacterial byproducts, but traveling to these sites requires migration through capillaries and into tissues. The best known diseases in which neutrophil adhesion is impaired are the leukocyte adhesion defects (LAD types 1, 2, and 3), discussed in Chapter 169. Other defects of neutrophil motility include juvenile periodontitis, Shwachman-Diamond syndrome, and the Chédiak-Higashi syndrome.

About two thirds of patients with CGD are males as they have defects in an X-linked gene encoding $gp91^{phox}$. Autosomal defects in $p47^{phox}$ are the next most common form, occurring in 20% of patients and often due to the same deletion. Other autosomal forms are due to defects in the gene encoding the $p22^{phox}$ or $p67^{phox}$ subunits (about 5% each).

CLINICAL MANIFESTATIONS

The genetic neutrophil disorders have specific clinical associations: delayed separation of the umbilical cord and poor wound healing in LAD-1; growth delay, mental retardation, and Bombay blood group in LAD-2; peripheral nerve conduction defects, pigmentary dilution with partial oculocutaneous albinism, easy bruising, and risk of hemophagocytic disease in Chédiak-Higashi syndrome; and pancreatic insufficiency (fat malabsorption), growth failure, and skeletal abnormalities in Shwachman-Diamond syndrome. For both Shwachman-Diamond syndrome and the severe congenital neutropenias, there is a risk for development of myelodysplastic disease and leukemia.

The clinical manifestations of CGD usually include bacterial or fungal infections. Males with the X-linked form will generally present in the first decade of life, whereas subjects with autosomal forms may have a later onset of symptoms (into the second decade). Regardless of the genetic cause, most patients with CGD will have one or more episodes of pneumonia; the most common causes of infection are *Staphylococcus, Burkholderia cepacia, Klebsiella, Aspergillus, Serratia,* and *Nocardia* species. Common clinical manifestations include acute or chronic lymphadenitis, colitis leading to recurrent diarrhea, *Staphylococcus* liver abscess, osteomyelitis, and rectal abscess. Patients with CGD are also prone to infections with unusual organisms, for example, *Chromobacterium violaceum, Trichosporon inkin, Francisella philomiragia,* and *Granulibacter bethesdensis*. For this reason, exposure to contaminated water or decaying plant material (compost, mulch) presents significant risk to subjects with CGD.

DIAGNOSIS

The differential diagnosis of neutropenia is presented in Table 167-4; the genetic diagnosis of congenital neutropenia syndromes, in Table 167-5; and a diagnostic approach to suspected phagocyte defects, in Figure 169-4.

TREATMENT Rx

The management of patients with neutropenia is discussed in Chapter 167. The treatment of CGD is summarized in Chapter 169.

WELL-DEFINED DEFECTS WITH SYNDROMIC FEATURES

Another group of primary immune defects are those that have distinctive systemic characteristics, aside from the obvious abnormities in the immune system (Table 250-6). The best known of these are the Wiskott-Aldrich syndrome, ataxia-telangiectasia, DiGeorge syndrome, hyperimmunoglobulin E (Buckley-Job) syndrome, and cartilage-hair hypoplasia. As these are distinct from each other, they are discussed separately.

Wiskott-Aldrich Syndrome

DEFINITION AND EPIDEMIOLOGY

Wiskott-Aldrich syndrome (WAS) is a rare X-linked recessive disease characterized by eczema, thrombocytopenia, and immune deficiency. WAS is rare, estimated as 1 and 10 cases per million males. Ethnic differences are not known.

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

WAS is inherited as an X-linked disease, and the main manifestations in early childhood include eczema, chronic thrombocytopenia sometimes leading to bloody diarrhea, and immune deficiency with recurrent infections.[11] Not uncommonly, autoimmunity or inflammatory disease including autoimmune hemolytic anemia, splenomegaly, arthritis, inflammatory bowel disease, and vasculitis appear. There is a clear increase in the incidence of lymphoma in WAS. The syndrome is caused by mutations in the *WAS* gene, which codes for the protein called WASP, an intracellular cytoplasmic scaffold protein important for the activation and mobility of all blood cells. WASP is involved in actin polymerization and in establishing an interface between immune cells (the immune synapse). Partly depending on the location of the

TABLE 250-6 WELL-DEFINED DEFECTS OF IMMUNITY WITH SYNDROMIC FEATURES

TYPE	GENES	INHERITANCE	LABORATORY FEATURES	ALTERED FUNCTIONS	DISEASE AND COMPLICATIONS
Wiskott-Aldrich syndrome	*WAS*	XL	Thrombocytopenia, small platelets	Impaired cell activation, mobility	Eczema; lymphoma; autoimmune disease; bacterial and viral infections
Ataxia-telangiectasia	*ATM*	AR	Some have IgA deficiency; IgG defects, lymphopenia in some	Impaired DNA double-stranded break repair	Ataxia; telangiectasia; pulmonary infections; lymphoreticular and other malignant neoplasms; increased α-fetoprotein; x-ray sensitivity
DiGeorge anomaly	22q11.2 deletion; rarely a deletion in 10p	De novo (majority) or AD	Lymphopenia; low T-cell numbers; large deletion in chromosome 22 on fluorescence in situ hybridization	Impaired T-cell immunity	Cardiac abnormalities; hypoparathyroidism, abnormal facies
Hyper-IgE syndrome (Buckley-Job syndrome)	*STAT3*	AD, de novo	Eosinophilia, high IgE	Loss of normal cytokine activation, defective IL-17	Bacterial infections; eczema, distinctive facial features, osteoporosis , fractures, scoliosis, delay of shedding primary teeth, hyperextensible joints, candidiasis
Cartilage-hair hypoplasia	*RMRP*	AR	Lymphopenia, low T-cell numbers	Impaired processing of mitochondrial RNA	Short-limbed dwarfism, sparse hair, celiac disease, Hirschsprung disease, bone marrow failure, autoimmunity, susceptibility to lymphoma

AD = autosomal dominant; AR = autosomal recessive; IL = interleukin; XL = X-linked.

mutation in the *WAS* gene, milder versions are known, leading to X-linked thrombocytopenia in some cohorts. Another, much rarer version leads to X-linked neutropenia.

DIAGNOSIS

The diagnosis is commonly made in the first few years of life in males with the characteristic symptoms of eczema with thrombocytopenia leading to petechiae. Typically, IgM levels are low, whereas IgA (and sometimes IgE) levels are increased. Platelet sizes are smaller than normal, and clot retraction is poor. Family history may include male relatives with WAS or thrombocytopenia. The diagnosis can be suggested by lack of the WAS protein as detected by flow cytometry in reference laboratories, but definitive diagnosis requires gene testing.

TREATMENT Rx

Treatment strategies in WAS are diverse and usually considered on a case-by-case basis. Conservative management includes prophylactic antibiotics, immunization with conjugated polysaccharide vaccines, and intravenous or subcutaneous immune goblin for subjects with repeated infections. For eczema, standard measures are used (Chapter 438). For significant thrombocytopenia (Chapter 172), splenectomy has been performed but is discouraged as lifelong post-splenectomy sepsis poses a significant risk. For these subjects, lifelong antibiotic prophylaxis is mandatory. Platelet transfusions should be reserved for active bleeding that cannot be managed with usual methods (e.g., aminocaproic acid) and avoided for subjects for whom transplantation is considered. Autoimmunity can be difficult to control, and immune suppression should be used with caution. Treatment of lymphomas is by standard regimens (Chapter 185). Definitive treatment is by HSCT, which has proved successful in many subjects. Trials with gene therapy are also ongoing.

PROGNOSIS

The prognosis in WAS is highly heterogeneous. Some have mild thrombocytopenia leading to occasional nose bleeds, whereas other subjects have inflammatory disease or other complications that require additional, sometimes intensive medical management. HSCT (Chapter 178) offers a cure, but this is best done early and requires careful matching and standard protocols.

Ataxia-Telangiectasia

DEFINITION AND EPIDEMIOLOGY

Ataxia-telangiectasia (AT) is a rare neurodegenerative disease that leads to cerebellar atrophy, skin telangiectasia, and immune defects. AT is estimated to occur in 1 in 40,000 to 100,000 but is more common in selected isolated populations. Sexes are affected equally.

PATHOBIOLOGY

AT is due to recessive mutations in the gene that encodes the ATM protein, important in both cell division and DNA repair. With the loss of the ATM protein, DNA breakage cannot be repaired, leading to cell death.

CLINICAL MANIFESTATIONS

The clinical manifestations include progressive difficulty in walking, with ataxia beginning around the age of 5 years. Skin telangiectasias develop on the bulbar conjunctiva and behind the ears. The immune defects include IgA deficiency, IgG subclass defects, and cellular defects leading to recurring pulmonary infections and lung damage in some. Subjects with AT have radiosensitivity, and the development of lymphomas is common with increasing age.

DIAGNOSIS

The diagnosis can usually be made by the characteristic clinical phenotype, coupled with an increase in α-fetoprotein in the blood. Radiosensitivity can be assessed in vitro in fibroblast cell lines. Definitive diagnosis is by *ATM* gene sequencing.

TREATMENT AND PROGNOSIS Rx

Treatment for AT includes supportive measures and physical therapy as needed. The life expectancy for subjects with AT varies greatly, but most live into early adulthood.

DiGeorge Syndrome

DEFINITION AND EPIDEMIOLOGY

DiGeorge syndrome is an autosomal dominant defect and one of the members of the 22q11.2 deletion syndrome that includes velocardiofacial syndrome, conotruncal anomaly face syndrome, congenital thymic aplasia, and thymic hypoplasia. DiGeorge syndrome is one of the most common of the immune defects, estimated at 1 : 4000. Sexes are affected equally.

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

Although it is classified as an immune defect because of thymic hypoplasia or aplasia, patients with DiGeorge syndrome are likely also to have congenital heart disease, cleft palate or pharyngeal closure defects, characteristic facies, hypocalcemia due to parathyroid insufficiency, and learning disability. The more common cardiac defects include tetralogy of Fallot, interrupted aortic arch, ventricular septal defects, vascular rings, and anomalous return of brachial arteries. The mnemonic CATCH-22 has been applied: cardiac issues, abnormal facies, thymic aplasia, cleft palate, and hypocalcemia. With loss of thymic tissue, cellular immunity is mildly to moderately impaired, leading to recurrent infections. Hypogammaglobulinemia is not uncommon and may be associated with autoimmune cytopenias, especially thrombocytopenia.[12]

DIAGNOSIS

The diagnosis of DiGeorge syndrome for most patients is based on the genetic test fluorescence in situ hybridization, which detects the loss of the 22q11.2 gene segment or, more rarely, a loss of 10p14-p13. However, about 10% do not have a gene defect but have the syndrome due to maternal diabetes, fetal alcohol syndrome, or prenatal exposure to isotretinoin (Accutane).

TREATMENT

Treatment of DiGeorge syndrome is based on individual need and may require cardiac surgery, cleft palate repair, and calcium and vitamin D supplementation if hypocalcemia is found. Some subjects are hypothyroid, requiring thyroid supplementation. The immune defect in DiGeorge syndrome varies widely, from complete loss of thymic development with no circulating T cells to normal T-cell numbers. For most, the thymus is hypoplastic, and whereas the level of T cells may be subnormal for age, sufficient T-cell function remains and no specific treatment is needed. Withholding of live viral vaccines may not be necessary as reports of ill effects are rare and the protection afforded is likely to outweigh any risk . For complete loss of the thymus, thymic transplantation, HSCT, or, for some, infusion of HLA-matched peripheral T cells can supply sufficient reconstitution.

PROGNOSIS

For most subjects, the prognosis depends on the concomitant medical issues, such as results of cardiac surgery, surgical repair of cleft palate, management of swallowing difficulties, and resources to enhance muscle strength and to overcome speech impediments and learning disabilities. For most, the T-cell defect is a minor component and a normal lifespan is likely; but with age, autoimmunity may become more prominent.

Hyperimmunoglobulin E Syndrome

DEFINITION

Hyperimmunoglobulin E syndrome (HIES), also called Buckley-Job syndrome, is an immune deficiency syndrome characterized by eczema, skin and lung abscesses, hyperextensible joints and recurrent bone fractures, distinctive coarse facies, eosinophilia, and high levels of serum IgE.

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

Most but not all patients with HIES have an autosomal dominant defect in the gene *STAT3* encoding a transcription factor, signal transducer and activator of transcription 3. After activation by selected cytokines and growth factors, STAT3 protein is phosphorylated and translocated to the cell nucleus. Whereas loss of STAT3 signaling affects many cellular processes, the syndrome itself is often clinically recognizable from the characteristic clinical findings of eczema, recurrent skin boils, unusual facies with tubular nose, cyst-forming pneumonias often due to *S. aureus*, and increased serum IgE. Other common manifestations include a rash in the newborn period; mucocutaneous candidiasis; and skeletal abnormalities, such as scoliosis, osteoporosis, fractures with minimal trauma, and delayed shedding of primary teeth.

Features of the HIES syndrome have been found in a few patients with rare autosomal recessive defects in genes encoding tyrosine kinase 2 (Tyk2) and dedicator of cytokinesis 8 (DOCK8), but in both of these, viral infections are also prominent.

DIAGNOSIS

The diagnosis can be strongly suspected on clinical grounds, but a useful scoring system composed of a composite of laboratory and clinical characteristics has been shown to aid in the dissection of hyperimmunoglobulin E subjects from other subjects with high serum IgE levels, for example, severely atopic subjects. IgE levels may range from 1000 to 40,000 IU or more; some references note that IgE levels may normalize in older subjects. Eosinophilia is common. However, a definitive diagnosis is best confirmed by identifying a *STAT*3 mutation.

TREATMENT

Rx

There is no definitive treatment for HIES syndrome. Because of the propensity for staphylococcal infections, prophylaxis with appropriate antibiotics (trimethoprim-sulfamethoxazole, 5 mg/kg/day trimethoprim divided twice daily) is commonly used, along with oral antifungals such as itraconazole, 100 mg daily for patients aged <13 years or weighing <50 kg; 200 mg daily for those aged >13 years or weighing >50 kg. Surgical drainage of abscesses is also important, but wound healing may be poor. Skin care for eczema may include bleach baths to reduce bacterial burden and antihistamines to control pruritus. Optimization of calcium and vitamin D levels may be useful to strengthen bones.

PROGNOSIS

HIES is a lifelong disease, and infections or complications require individual care. With increasing age, increased respiratory dysfunction is likely. Recent reports note vascular, especially arterial, abnormalities, which are important to define.

Cartilage-Hair Hypoplasia

DEFINITION AND EPIDEMIOLOGY

Cartilage-hair hypoplasia (CHH) is a rare autosomal recessive form of short-limbed dwarfism associated with a variable cellular immune deficiency. In the Old Order Amish population, CHH affects 1 in 1300 newborns; for those of Finnish descent, the incidence is about 1 in 20,000. A common point mutation in the gene is prevalent in the particularly affected populations.

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

The genetic defect in CHH is in the *RMRP* gene, which encodes the RNA in a mitochondrial RNA-processing endoribonuclease that helps copy mitochondrial DNA and process ribosomal RNA. In CHH, the encoded RNA is unstable, leading to skeletal dysplasia, sparse hair, and the predominantly T-cell deficiency disease. For unclear reasons, the cellular defect is varied, ranging from mildly impaired immunity to severe defects requiring HSCT. Other clinical features of CHH include short stature, anemia, autoimmunity, celiac disease, Hirschsprung disease, and several cancers including lymphoma.

DIAGNOSIS

The diagnosis can be suspected by inheritance patterns and clinical phenotype, but definitive diagnosis is made by genetic sequencing of the *RMRP* gene.

TREATMENT AND PROGNOSIS

Rx

For patients with severe T-cell defects and infections (essentially SCID phenotype) suggesting a significantly impaired immune system, HSCT is required. Aside from the cellular defect, treatment is directed at the other presenting clinical issues. The prognosis of CHH is varied, and the prognosis depends on the depth of the cellular defect, the therapies required, and the associated clinical complications. As the defect is so variable, normal lifespans may be achieved.

IMMUNE DYSREGULATION SYNDROMES

Clustered into a separate set of immunodeficiency diseases are syndromes that have immune dysregulation as a main theme. These mostly monogenic diseases have in common lymphoid proliferation, immune activation, and inflammatory or autoimmune complications. These include the hemophagocytic lymphohistiocytosis (HLH) diseases (Chapter 169), the lymphoproliferative syndromes linked to Epstein-Barr virus (EBV) infection, autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy (APECED), autoimmune lymphoproliferative syndromes (ALPS), and defects of T-regulatory cells (Table 250-7).

EPIDEMIOLOGY

The estimated incidence of the genetic HLH syndromes is 1 : 50,000; the incidence of X-linked lymphoproliferative disease is 1 to 3 in 1,000,000; and the incidence of APECED is high in Finland (1 in 25,000) and in Sardinians and Iranian Jews (1 in 9000) but otherwise much rarer. The incidence of ALPS and defects of T-regulatory cells is unknown.

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

HLH is a form of extreme and potentially life-threatening immune activation. It is further discussed under the section Hemophagocytic Lymphohistiocytosis and Macrophage Activation Syndrome in Chapter 169. There are two forms: genetic, due to mutations in genes that control cellular cytotoxicity; and secondary, due to acute viral illnesses, autoimmune activation, or underlying malignant disease.[13] The familial form is a heterogeneous autosomal recessive disorder due to mutations in one of five genes essential for control of T-cell cytotoxicity. Immune activation leads to expansion of poorly controlled cytotoxic T cells and macrophages, leading to the release of interferon (IFN)-γ, IL-1, IL-6, and IL-10. Patients have high fevers, cytopenias, liver dysfunction, coagulopathy, and sometimes neurologic symptoms. HLH may be fatal unless it is treated with aggressive measures and may require HSCT. Life-threatening accelerated immune activation syndromes are also characteristic of other genetic defects that impair cytotoxicity, such as the Chédiak-Higashi syndrome.

Monogenic defects leading to impaired immunity to EBV produce another group of immune dysregulation syndromes. The first described (and the most common, 70 to 80%) is the X-linked proliferative disorder (XLP) due to mutations of the X-linked gene *SH2D1A*, which encodes the gene *SAP*, a signaling lymphocytic activation molecule (SLAM)–associated protein. Other genetic causes for loss of control of EBV are due to mutations in *XIAP* (20 to 30%) or, rarely, *ITK* and *CD27*. *XLP* and *XIAP* are on the X chromosome; the other defects are inherited as autosomal recessive traits. In each case, infection with EBV leads to an acute illness with lymphoproliferation, progressive but variable hypogammaglobulinemia, and lymphoma in *XLP*, *ITP*, and *CD27* defects. A unifying theme of these syndromes is loss of function of NK-T cells, a subset of T cells important in viral immunity.

A unique member of the genetic immune dysregulation diseases is APECED or autoimmune polyglandular syndrome type 1 (Chapter 231). Whereas the clinical presentation is usually due to endocrine disease (hypoparathyroidism, Addison disease, hypogonadism, and secondary amenorrhea), the disease is caused by loss of thymic recognition of self-antigens due to mutations in the autoimmune regulator gene (*AIRE*).[14] Chronic mucocutaneous candidiasis is common, probably due to circulating anticytokine antibodies (interferon and IL-17). The transcription factor encoded by the *AIRE* gene, found in thymic epithelial cells, is involved in the early negative selection of cells with autoimmune potential. Clinically, cutaneous candidiasis or the endocrine defect may be the first sign of the syndrome; for unclear reasons, chronic diarrhea with malabsorption is also common. Other autoimmune complications may include hepatitis, alopecia, vitiligo, diabetes mellitus, anemia, and pernicious anemia.

Defects in lymphocyte apoptosis lead to another form of immune dysregulation; in these subjects, because of impaired death of lymphocytes, lymph nodes and spleen enlarge and autoimmunity, especially thrombocytopenia and hemolytic anemia, occur. Together, these are commonly referred to as ALPS. The most common of these defects are due to autosomal dominant mutations in the *FAS* gene, which encodes the important FAS death receptor, and less commonly in FAS ligand. Both are dominant but have variable penetrance. Much less common forms of autoimmune lymphoproliferation are due to mutations in caspase 10 or even more rarely the oncogenes *KRAS* and *NRAS* or protein kinase Cδ.[15]

TABLE 250-7 DISEASES OF IMMUNE DYSREGULATION

TYPE	GENES	INHERITANCE	LABORATORY FEATURES	ALTERED FUNCTION	DISEASE AND COMPLICATIONS
Familial hemophagocytic lymphohistiocytosis syndromes	*PRF1, UNC13D, STX11,* and *STXBP2*	AR	Anemia, neutropenia, thrombocytopenia, abnormal liver functions, high ferritin and serum IL-2 receptor, hemophagocytosis in bone marrow and liver	Decreased to absent NK cells and cytotoxic activities	Fever, hepatosplenomegaly, cytopenias, hemophagocytic lymphohistiocytosis, neurologic disease in some
Chédiak-Higashi syndrome	*LYST*	AR	Neutrophils with giant inclusions; hair: pigment clumps	Impaired chemotaxis	Partial albinism, recurrent infections, late-onset primary encephalopathy, increased lymphoma risk
Lymphoproliferative syndromes	*SAP, XIAP* (*ITK, CD27*)	XL, AR	Epstein-Barr virus infection; decreased NK cells and $CD8^+$ CTL activation; deficient NK-T cells; anemia; hypogammaglobulinemia in some	Loss of function of NK-T cells leading to impaired viral control	Clinical and immunologic features triggered by Epstein-Barr virus infection; lymphoproliferation, lymphoma
Autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy	*AIRE*	AR	Endocrine dysfunction; hepatitis	Loss of thymic self-tolerance	Autoimmunity leading to hypoparathyroidism, hypothyroidism, diabetes, adrenal and gonadal dysfunction; cutaneous candidiasis; hepatitis
Autoimmune lymphoproliferative syndrome	*FAS*, FAS ligand; (caspase 10; *KRAS; NRAS;* protein kinase Cδ)	AD	Increased double-negative T cells ($CD4^-/CD8^-$), increased serum B_{12}	Defects in lymphocyte apoptosis	Splenomegaly, lymphadenopathy, autoimmune cytopenias; increased risk of lymphoma
Genetic defects of T-regulatory cells	*FOXP3* (CD25, *STAT5B* and *ITCH, STAT1*)	XL, AR	Diabetes, anemia, eosinophilia, high serum IgE, hyperglobulinemia, loss of *FOXP3* expression in the XL form	Lack of (or impaired function of) $CD4^+$, $CD25^+$, $FOXP3^+$ regulatory T cells (Tregs)	Enteropathy, dermatitis, eczema, early-onset diabetes, thyroiditis, hemolytic anemia, thrombocytopenia, elevated IgE and IgA

AD = autosomal dominant; AR = autosomal recessive; CTL = cytotoxic T lymphocyte; IL = interleukin; XL = X-linked.

The defects of T-regulatory cells are the final member of this set of genetic defects leading to loss of regulation.[16] The first to be described was the immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome, a generally lethal disease in males, characterized by early-onset insulin-dependent diabetes mellitus, enteropathy with severe diarrhea, and eczema-like dermatitis. Other manifestations include anemia, thrombocytopenia, and neutropenia as well as liver or kidney autoimmune disease. The defect is generally due to mutations of the X chromosome forkhead box protein 3 (*FOXP3*) gene, a gene essential for the development of regulatory T cells. However, other genetic defects may lead to a similar clinical syndrome.[16]

DIAGNOSIS

The diagnosis of these syndromes may be suspected from the clinical manifestations, family history, and laboratory grounds, but genetic validation is required for definitive diagnosis.

TREATMENT

Rx

For the HLH diseases and related syndromes, prompt immune suppression by established protocols and intense supportive care are necessary. For the genetic forms, HSCT is often required. The lymphoproliferative syndromes associated with EBV are similar in that prompt supportive care is required and transplantation is potentially curative. Rituximab has been used in these defects to reduce B-cell numbers and EBV burden if infection occurs. Treatment of the cytopenias in ALPS includes corticosteroids, rapamycin, and other agents. For unclear reasons, rituximab may lead to permanent hypogammaglobulinemia in ALPS, and splenectomy is to be avoided. Patients with APECED usually require endocrine and possibly nutritional management as well as treatment for cutaneous candidiasis. For the defects of T-regulatory cells due to mutations in IPEX, HSCT is the only curative measure.

PROGNOSIS

The diseases of immune dysregulation have a varied prognosis. For the genetic HLH syndromes, EBV-related lymphoproliferative diseases, and IPEX, immune reconstitution is required. Because of the broad spectrum of manifestations for ALPS and APECED, management of the clinical issues may be sufficient.

DEFECTS OF INNATE IMMUNITY THAT LEAD TO SELECTED INFECTIONS

DEFINITION

As opposed to the adaptive immune system (in which a previous exposure is required to form immune memory; Chapter 46), many components of the immune system function quickly with no pre-exposure. These components of the innate immune system (Chapter 45) include, for example, complement, phagocytic cells, and natural killer cells. Screening of large populations for selected microbial diseases has revealed a number of novel defects of innate immunity. Some of these defects are discussed here (Table 250-8).

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

Innate defects appear to be rare and the incidence is not known. One of the first recognized was anhidrotic ectodermal dysplasia with immunodeficiency, a syndrome due to mutations in the *IKBKG* gene that encodes nuclear factor κB (NF-κB) essential modulator (NEMO). This is an X-linked disease and was first assigned to the category of hyper-IgM syndromes. However, the actual phenotype is broad due to impairment of NEMO, which is essential for both cytokine and toll-like receptor signaling pathways. Impairment in this gene leads to severe bacterial infections and mycobacterial disease as well as to the characteristics of ectodermal dysplasia: sparse hair, abnormal tooth development, and lack of sweat glands. After NEMO defects were recognized, a series of other genetic defects in toll-like receptors and their signaling

TABLE 250-8 EXAMPLES OF DISEASES OF INNATE IMMUNITY

DISEASE	GENES	INHERITANCE	LABORATORY FEATURES	ALTERED FUNCTION	ASSOCIATED FEATURES
Anhidrotic ectodermal dysplasia with immunodeficiency	*NEMO* (*IKBKG*), *IKBA*	XL, AD	Variable hypogammaglobulinemia for NEMO with increased IgM in some; lack of antibody response to polysaccharides	Defective NF-κB signaling pathway	Bacterial and mycobacterial infections Ectodermal dysplasia, hair loss, heat intolerance due to loss of sweat glands Tooth abnormalities
IRAK4 deficiency	*IRAK4*	AR	Impaired cytokine responses to toll receptor activators	Defective TIR-IRAK signaling pathway	Bacterial infections, especially *Staphylococcus* and *S. pneumoniae*
MyD88 deficiency	*MYD88*	AR	Impaired cytokine responses to toll receptor activators	Defective TIR-MyD88 signaling pathway	Bacterial infections, especially *Staphylococcus* and *S. pneumoniae*
Herpes simplex encephalitis	*TLR3* *UNC93B1* *TRAF3*	AD AR	Impaired cytokine responses to TLR3 activators	Defective IFN-α, IFN-β, and IFN-γ induction	Herpes simplex virus 1 encephalitis
Predisposition to fungal diseases	*CARD9*	AR	Fungal cultures positive	Defective CARD9 signaling pathway	Invasive candidiasis and other fungal diseases
Chronic mucocutaneous candidiasis	*IL17RA* *IL17F* *STAT1*	AR, AD	Fungal cultures positive	Defective IL-17R signaling pathways	Mucocutaneous candidiasis
IL-12, IL-23 receptor deficiency	*IL12RB* *IL12* *IL23*	AR	Mycobacterial cultures positive	Defective cytokine receptor binding and signaling	Mycobacterial and salmonella infections
IFN-γ receptors 1 and 2 deficiency	*IFNGR1* *IFNGR2*	AR	Mycobacterial cultures positive	Defective IFN-γ binding and signaling	Mycobacterial and salmonella infections
GATA2 deficiency	*GATA2*	AR, AD	Multilineage cytopenias; very low monocyte numbers		Infections with mycobacteria, papillomaviruses, histoplasmosis, alveolar proteinosis, but also myelodysplasia and leukemias

AD = autosomal dominant; AR = autosomal recessive; IFN = interferon; IL = interleukin; IL-17R = interleukin-17 receptor; NF-κB = nuclear factor-κB; TIR = intracytoplasmic Toll and IL-1 receptor; TLR = toll-like receptor; XL = X-linked.

pathways have been recognized, for example, autosomal recessive defects in IRAK4 and MyD88, both of which lead to severe pneumococcal and staphylococcal infections. In contrast, effects of the TLR3 pathway lead to early herpes simplex encephalitis. Much more clinically heterogeneous are the genetic defects that lead to chronic mucocutaneous candidiasis. These defects may be autosomal dominant or recessive and lead to simple onychomycosis in some to invasive fungal infections in others. Patients of any age may have defects in these pathways. The pathogenesis of some of these includes genes that disrupt the dectin-1 pathway. Dectin-1 is a surface lectin receptor that recognizes the β1-3 glucan of fungi; downstream mutations in *CARD9* impair the secretion of IL-17A, IL-17F, and IL-22, cytokines that are essential in fungal clearance.[17]

A separate and unique category of innate defects are the cytokine/receptor mutations that impair the functions of cytokines IL-12, IL-23, and IFN-γ, which are needed for control of mycobacterial and other intracellular infections, such as salmonella. Chronic mycobacterial infections may also occur in patients with autosomal recessive mutations in the gene for signal transducer and activator of transcription 1 (*STAT1*), a gene downstream of both IFN-γ and IFN-α receptors. However, as the functions of both cytokines are impaired, these patients may also have severe viral or fungal infections. On the other hand, dominant (activating) mutations in *STAT1* may lead to simple cutaneous candidiasis or more complex clinical outcomes in others. More complex is the syndrome of GATA2 deficiency, in which mycobacterial disease may also develop, but other organisms (papillomaviruses, fungi) and serious complications (cytopenias, myelodysplasia, pulmonary alveolar proteinosis, peripheral edema) may be foremost. Whereas *GATA2* defects are dominantly inherited, members of the same family with the same mutations may have very different clinical manifestations.

DIAGNOSIS

The diagnosis of innate defects is first based on exclusion of other causes and then confirmed by genetic testing. The family history may be helpful, but for patients with mutations in *STAT1* or *GATA2*, although dominant inheritance is likely, the extreme range of clinical phenotypes may obscure easy recognition.

TREATMENT AND PROGNOSIS

Rx

The treatment of innate defects includes antimicrobial therapy to clear active infections and, probably, relevant prophylactic therapy on an ongoing basis. For the more severe defects, HSCT is required.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Milner JD, Holland SM. The cup runneth over: lessons from the ever-expanding pool of primary immunodeficiency diseases. *Nat Rev Immunol*. 2013;13:635-648.
2. Al-Herz W, Bousfiha A, Casanova JL, et al. Primary immunodeficiency diseases: an update on the classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency. *Front Immunol*. 2014;5:162.
3. Ochs HD, Hagin D. Primary immunodeficiency disorders: general classification, new molecular insights, and practical approach to diagnosis and treatment. *Ann Allergy Asthma Immunol*. 2014;112:489-495.
4. Bousfiha AA, Jeddane L, Ailal F, et al. A phenotypic approach for IUIS PID classification and diagnosis: guidelines for clinicians at the bedside. *J Clin Immunol*. 2013;33:1078-1087.
5. Kwan A, Abraham RS, Currier R, et al. Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA*. 2014;312:729-738.
6. Shearer WT, Dunn E, Notarangelo LD, et al. Establishing diagnostic criteria for severe combined immunodeficiency disease (SCID), leaky SCID, and Omenn syndrome: the Primary Immune Deficiency Treatment Consortium experience. *J Allergy Clin Immunol*. 2014;133:1092-1098.
7. Tam JS, Routes JM. Common variable immunodeficiency. *Am J Rhinol Allergy*. 2013;27:260-265.
8. Singh K, Chang C, Gershwin ME. IgA deficiency and autoimmunity. *Autoimmun Rev*. 2014;13:163-177.
9. Orange JS, Ballow M, Stiehm ER, et al. Use and interpretation of diagnostic vaccination in primary immunodeficiency: a working group report of the Basic and Clinical Immunology Interest Section of the American Academy of Allergy, Asthma & Immunology. *J Allergy Clin Immunol*. 2012;130:S1-S24.
10. Resnick ES, Moshier EL, Godbold JH, et al. Morbidity and mortality in common variable immune deficiency over 4 decades. *Blood*. 2012;119:1650-1657.
11. Massaad MJ, Ramesh N, Geha RS. Wiskott-Aldrich syndrome: a comprehensive review. *Ann N Y Acad Sci*. 2013;1285:26-43.
12. Gennery AR. Immunological features of 22q11 deletion syndrome. *Curr Opin Pediatr*. 2013;25:730-735.
13. Chandrakasan S, Filipovich AH. Hemophagocytic lymphohistiocytosis: advances in pathophysiology, diagnosis, and treatment. *J Pediatr*. 2013;163:1253-1259.
14. Arstila TP, Jarva H. Human APECED; a sick thymus syndrome? *Front Immunol*. 2013;4:313.
15. Oliveira JB. The expanding spectrum of the autoimmune lymphoproliferative syndromes. *Curr Opin Pediatr*. 2013;25:722-729.
16. Verbsky JW, Chatila TA. Immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) and IPEX-related disorders: an evolving web of heritable autoimmune diseases. *Curr Opin Pediatr*. 2013;25:708-714.
17. Lanternier F, Cypowyj S, Picard C, et al. Primary immunodeficiencies underlying fungal infections. *Curr Opin Pediatr*. 2013;25:736-747.

REVIEW QUESTIONS

1. A 30-year-old man has had episodes of chronic diarrhea for several years. He is thin and has had a 10-pound weight loss during the previous year. He has had an extensive gastrointestinal evaluation for Crohn disease, but the biopsies were inconclusive; however, villous blunting and a lack of plasma cells were noted in the intestinal mucosa. Which of the following might be appropriate?

A. Test for anti-endomysial antibodies.
B. Treat with steroid.
C. Test for serum IgG, IgA, and Ig.
D. Test for stool parasite.
E. Order computed tomography of the abdomen with contrast enhancement.

Answer: C Loss of plasma cells in the gastrointestinal mucosa is a signal that there is a profound lack of B-cell development. This is a characteristic in common variable immune deficiency in general in those with or without gastrointestinal disease and would be expected to be associated with low serum levels of IgG, IgA, and IgM. Whereas celiac disease, for which testing for anti-endomysial antibodies would be sensitive and specific, is also characterized by villous atrophy, intestinal biopsy in this disease demonstrates increased lymphocyte and plasma cell infiltration, not absence of plasma cells.

2. A 64-year-old man goes to an allergist with chronic sinusitis for 3 years but no other significant illnesses aside from onychomycosis of several fingernails. Skin test results for allergy to aeroallergens are negative, and the allergist tests quantitative immunoglobulins. These show an IgG level of 160 mg/dL (normal, 700 to 1600 mg/dL), IgA level of 7 mg/dL (normal, 40 to 250 mg/dL), and IgM level of 12 mg/dL (normal, 50 to 300 mg/dL). Which of the following is the next step?

A. Give an antifungal cream.
B. Test serum Ig.
C. Culture the sinuses.
D. Order computed tomography of the chest.
E. Request a bone marrow examination.

Answer: D This is likely to be Good syndrome, thymoma with hypogammaglobulinemia (or agammaglobulinemia). The main characteristics are the later onset and the observation of nail fungus. For diagnosis of this syndrome, computed tomography (or magnetic resonance imaging) of the chest is needed. The loss of antibody will require immunoglobulin replacement.

3. A 21-year-old woman comes to the emergency department with 4 days of chest pain, some blood-stained sputum, and a low-grade fever for 2 days. Her medical history includes severe eczema since infancy, an abscess on her buttock that required in-hospital surgical drainage, and a wrist fracture on closing a file cabinet. Aside from the fairly severe eczema, she does not appear ill; her white blood cell count is 9400 with 70% neutrophils and 10% eosinophils. Which of the following is the next step?

A. It is likely a viral syndrome and she can be discharged.
B. Give her a prescription for an antibiotic for 5 days
C. Get a chest radiograph.
D. Treat the eczema more aggressively.
E. Evaluate her for asthma with pulmonary function tests.

Answer: C With the clinical history, this may be hyperimmunoglobulin E syndrome (HIES). Pneumonias in HIES often lead to fewer clinical signals of toxicity; her past history of eczema since infancy, a significant abscess requiring drainage, and a fracture with minimal trauma suggest this diagnosis. A white blood cell count may not show neutrophilia, but eosinophilia is common. She needs a chest radiograph and, if it is abnormal, computed tomography to demonstrate the infection. At least 80% of HIES patients have had one or more pneumonias, often with mild symptoms; these lead to cysts and lung cavities.

4. A 29-year-old software designer was referred for evaluation of hypogammaglobulinemia. Recent evaluation for an enlarged spleen had revealed low serum levels of all immunoglobulins, including an IgG level of 251 mg/dL, IgA level 10 mg/dL, and IgM level 39 mg/dL. At the age of 15 years, he was febrile, anemic, and thrombocytopenic and developed splenomegaly and lymphadenopathy. He had a bone marrow that showed normal hematopoiesis, but the stain for Epstein-Barr virus (EBV)–encoded RNA (EBER) was positive. He was diagnosed with mononucleosis and recovered. His complete blood count is normal. Which of the following should be considered?

A. HIV test
B. Bone marrow examination
C. Monospot
D. Protein loss in the intestinal tract
E. X-linked lymphoproliferative disease

Answer: E The X-linked proliferative disorder (XLP), due to mutations of the X-linked gene *SAP*, leads to a significant infection with EBV with lymphoproliferation, progressive but variable hypogammaglobulinemia, and, in some, lymphoma. This is the most common genetic disease leading to loss of control of EBV.

5. JF is a 27-year-old man who is referred because of the possibility of food allergy. For the past 5 years, he has had recurrent episodes of abdominal pain with nausea and vomiting and has concluded that he must be allergic to a common food ingredient. On one occasion, he had such severe abdominal pain that he went to the emergency department, where acute food poisoning and appendicitis were excluded. His father had a long history of "colitis" with intermittent abdominal pain and diarrhea. Which of the following tests might be helpful?

A. Serum tryptase
B. Serum C4
C. Total hemolytic complement
D. Computed tomography of abdomen with contrast enhancement
E. Serum Ig

Answer: B With the severity of the pain and family history of a first-degree relative with an unexplained but somewhat similar abdominal pain, a C1 inhibitor defect is possible. The abdominal pain may be due to angioedema of the intestine. The best screening test would be serum C4, and if the level is low, one would proceed to test the C1 inhibitor protein and function.

251 ALLERGIC RHINITIS AND CHRONIC SINUSITIS

LARRY BORISH

ALLERGIC RHINITIS

Allergic rhinitis (AR) refers to the nasal and ocular symptoms that result from an inflammatory hypersensitivity reaction to aeroallergens deposited on the nasal mucosa and conjunctiva.

EPIDEMIOLOGY

AR, the most common chronic disease in the United States, affects between 10 and 30% of adults and up to 40% of children. Each year, nearly 80 million people in the United States experience 7 days or more of nasal or ocular

symptoms as a result of AR. Although it is not a severe disorder, the socioeconomic costs of AR are substantial. AR is one of the chief reasons for visiting a primary care physician; it adversely affects work productivity and school performance, and it limits socialization. The impact of AR also reflects its association with a variety of comorbid conditions, including asthma (Chapter 87), acute and chronic sinusitis, nasal polyposis, secretory otitis media, and sleep disorders. Adequately addressing AR requires a thorough understanding of its pathophysiology, its relation to these comorbid conditions, and the effects of various therapeutic options on AR and its associated comorbidities.

PATHOBIOLOGY

Airborne Allergens

Allergic respiratory diseases result from a hypersensitivity immune reaction to airborne allergens. These include the pollens and molds that are responsible for seasonal AR (SAR) and the indoor allergens, such as dust mites, indoor molds, and animal proteins, that are responsible for perennial AR (PAR) (Table 251-1).

In any area, the specific pollens that are likely to cause symptoms can be predicted from the number of days a pollen is airborne in large numbers. All these pollens use a wind-borne mechanism to achieve fertilization. Insect-borne pollens, specifically those produced by flowers, are not significantly airborne and therefore are not inhaled in sufficient concentrations to generate immune responses. In the United States, grass pollens (May to June) and ragweed (mid-August to October) are the most important causes of SAR. Tree pollens vary locally but typically start in late February and continue through April. The major trees implicated in allergy include birch in the North, oak in the mid-Atlantic region, live oak in the South, and mountain cedar in the Southwest. In addition to pollens, outdoor molds, particularly *Alternaria* and *Cladosporium*, can produce symptoms. These molds have variable seasons, depending on the weather; high levels of airborne fungi may be common at any time between March and October. In recent years, climate change has influenced both the length and the region of pollen seasons.

TABLE 251-1 ALLERGENS CAUSING ALLERGIC RHINITIS

COMMON NAME	SEASON
SEASONAL ALLERGENS	
Trees	
Birch	March–May
Cottonwood	April–May
Elm	February–May
Cedar	March–May
Oak	May–June
Maple	March–May
Grasses	
Kentucky blue	Mid-May–June
Timothy	Mid-May–June
Orchard	Mid-May–June
Sweet vernal	Mid-May–June
Fescue	Mid-May–June
Bermuda	Mid-May–June
Weeds	
Ragweed	August–September
Kochia	July–September
Russian thistle	July–September
Sage	July–September
Marsh elder	July–September
English plantain	July–September
Outdoor Molds	
Alternaria	Spring–fall
Cladosporium	Spring–fall
PERENNIAL ALLERGENS	
Household Allergens	
Cockroaches (German and American)	
Dust mites: *Dermatophagoides farinae, D. pteronyssinus, Blomia tropicalis*	More active in summer and humid months
Other insects (spiders, ladybugs)	
Animals	
Cats	
Dogs	
Other pets (guinea pigs, ferrets, hamsters, horses)	
Rodents	
Indoor Molds	
Aspergillus	
Cladosporium	
Penicillium	

PAR may occur year-round, but the term is applied to any rhinitis that does not have a clearly defined seasonal association. The most common causes include (1) indoor fungi, which are related to periods of high indoor humidity; (2) animal danders, particularly cats, but rodents (mice, rats, guinea pigs, ferrets, hamsters), rabbits, dogs, and birds may also be significant; (3) dust mites of the genus *Dermatophagoides*, which grow in bedding and pillows and are semiseasonal, with maximal levels from August to December; and (4) other insects (the best studied is the cockroach, but gypsy moths, crickets, ladybugs, spiders, and beetles may be locally important). Dust mites and cats produce the most important indoor allergens. Dust mites grow well only with a relative humidity higher than 50%. Dust mite allergy is probably relevant in all areas with more than 6 humid months in the year.

Immunoglobulin E, Mast Cell, and Basophil Activation

The traditional view is that AR is caused by the triggering of mast cell degranulation resulting from the cross-linking of surface-bound immunoglobulin (Ig) E molecules by the aeroallergen. As with all antibody-mediated immune responses, the initial exposure to the antigen results in B-lymphocyte secretion of low-affinity IgM antibodies. Subsequent exposure to the allergen in genetically predisposed subjects leads to a secondary immune response characterized by the isotype switch to IgE. The specific mechanisms underlying isotype switching to IgE are controversial, including the exact location where this occurs and the extent to which this requires an IgG intermediate. Similarly, the relative production of IgE-producing long-lived B cells or plasma cells is very poorly understood, although experience with bee stings and drug allergy suggests ongoing IgE synthesis for up to 10 years after an isolated allergenic exposure.[1] The resulting release of IgE antibodies into the circulation does not cause allergic symptoms; only after these IgE antibodies bind to their high-affinity receptors on basophils and mast cells do symptoms develop with subsequent allergen exposure. It takes the cross-linking of approximately 300 IgE receptors and cells to stimulate degranulation; therefore, and with the large number of mast cells scattered through virtually all tissues, it often requires several allergy seasons before sufficient numbers of allergen-specific IgE molecules are present on the surface of a given mast cell to drive its degranulation. This means that the development of symptomatic AR is a protracted process, generally requiring at least three or four exposures. As a result, SAR is generally not observed in children until they are approximately 4 years of age. Similarly, in adults, symptomatic responses to local allergens may not develop until approximately 4 years after moving to a region. PAR, however, can develop much faster.

Within minutes of allergen exposure, IgE-sensitized mast cells degranulate and release preformed and newly synthesized mediators, including histamine, proteases (tryptase and chymase), cysteinyl leukotrienes, prostaglandins, platelet-activating factor, and cytokines. Some of these mediators produce the characteristic early-phase symptoms of AR, namely, sneezing, pruritus, rhinorrhea, and, to some extent, congestion. Other mediators stimulate infiltration of the nasal mucosa with inflammatory cells, including basophils, eosinophils, neutrophils, additional mast cells, and mononuclear cells. This infiltration of inflammatory cells and their subsequent release of a secondary wave of mediators sustain the inflammatory reaction and produce the late-phase response of AR. This slowly developing inflammatory response is characterized primarily by nasal congestion. The inflammation that develops during the course of an allergy season is associated with an approximately 10-fold increase in the number of mast cells present in nasal epithelial and submucosal tissue. This reflects the migration of preexisting mast cells into the epithelium and the differentiation and influx of newly synthesized mast cells under the influence of cytokine growth factors. During the course of chronic allergen stimulation, these mast cells also display increased *priming*, which reflects increases in the number of IgE receptors and surface-bound IgE as well as enhancement of signal transduction pathways. As a result, as the allergen season progresses, less and less allergen (engagement of fewer and fewer IgE receptors) is required to trigger mast cell degranulation.

Antigen-Presenting Cell and Helper T-Lymphocyte Activation

In addition to their interaction with mast cells, allergens behave like any other foreign antigen and are processed and presented by antigen-presenting cells

to helper T (T_H) lymphocytes. Activation of these antigen-presenting cells, including mononuclear phagocytic cells, B lymphocytes, and especially dendritic cells, is an important source of cytokines, especially those associated with innate immunity, such as interleukin (IL)–1, IL-6, and tumor necrosis factor-α (TNF-α). Recent understanding of the molecular mimicry of allergens has increasingly expanded our understanding of the role of pathogen-associated molecular patterns receptors in dendritic cell activation, such as the engagement of toll-like receptor 4 by *Dermatophagoides* and similarly engagement of the lectin receptor dectin-2 by *Aspergillus*.[2] Engagement of these receptors promotes an inflammatory response to these otherwise innocuous inhaled particles. With the ensuing development of allergic inflammation, there is an increased presence of B cells expressing allergen-specific surface immunoglobulin and dendritic cells expressing surface-bound IgE. These antibodies can function as receptors that "capture" allergen and increase these cells' effectiveness in antigen processing. The newly activated T lymphocytes tend to resemble T_H2 cells, characterized by the production of IL-4, IL-5, IL-9, IL-13, and granulocyte-macrophage colony-stimulating factor (GM-CSF). These cytokines are also major components of the inflammatory response in AR and contribute to the increased production, recruitment, and activation of eosinophils, mast cells, and basophils. A milieu rich in IL-4 and IL-13 drives the IgE isotype switch and contributes to the further production of allergen-specific IgE. IL-13 also drives the metaplastic differentiation of nasal epithelial cells into goblet cells that are characterized not only by production of mucus but, importantly, also by chemokines such as CCL5, CCL11, CCL24, and CCL26 that further mediate basophil and eosinophil chemotaxis.

Innate Lymphoid Type 2 Cells as Additional Sources of Type 2 Signature Cytokines

An additional cell type, the type 2 innate lymphoid cell (ILC2), having features of NK cells and sharing with the NK cell the absence of a T-cell receptor, has recently been defined and distinguished by its secretion of preformed IL-5 and IL-13.[3] ILC2s can be activated through several pathways, but it is increasingly appreciated that nasal epithelial cells themselves respond to allergens to secrete two cytokines, IL-25 and IL-33, that are central to this process. This cellular mechanism defines a pathway linking allergen engagement of receptors on epithelium that drives their secretion of IL-25 and IL-33, leading in turn to an allergic inflammatory milieu driven by ILC2–derived IL-5 and IL-13. This provides a parallel pathway for development of allergen-induced inflammation that does not require engagement of the adaptive immune system (i.e., allergen-specific IgE or allergen-targeting T effector cells).

Through these mechanisms, during the seasonal course of allergen exposure, rhinitis evolves and becomes more dependent on mediators associated with the infiltration of eosinophils, basophils, neutrophils, mononuclear cells, and T_H lymphocytes as well as the increasingly primed mast cells. The symptoms of acute rhinitis, such as sneezing, itching, and rhinorrhea, largely reflect vasoactive mediator release, especially histamine. As SAR or PAR persists, however, these infiltrating cells continue to produce cytokines and other inflammatory mediators, leading to mucus hypersecretion, tissue edema, goblet cell hyperplasia, and tissue damage that become the primary sources of allergy patients' symptoms. As the role of histamine diminishes with AR progression (discussed later), antihistamines become less and less effective.

Eosinophils represent an important component of the inflammation that develops in AR. Eosinophils release a wide variety of proinflammatory mediators, including cysteinyl leukotrienes (leukotrienes C4, D4, and E4), eosinophil cationic protein, eosinophil peroxidase, major basic protein, IL-3, IL-4, interferon-γ, GM-CSF, and platelet-activating factor. Eosinophil-derived mediators are major components of the chronic allergic response and produce many of the symptoms of AR, especially nasal congestion. The natural history of AR is for symptoms to worsen inexorably during several weeks in the presence of ongoing allergen exposure. Symptoms often do not peak until well after the peak in pollen counts, and then they persist after pollen counts have declined. These observations reflect the time frame of the onset of nasal inflammation and tissue damage. The influx of eosinophils into the nasal mucosa correlates with the development and progression of symptoms. In summary, the natural history of AR represents an evolution from an acute, primarily mast cell–mediated process that is responsive to antihistamines to a chronic inflammatory process that is primarily eosinophil mediated and is much less responsive to antihistamines.

CLINICAL MANIFESTATIONS

The diagnosis of AR is based on a history of sneezing, which is often paroxysmal; rhinorrhea with clear, watery secretions; nasal congestion; and itching in the nares and palate. These symptoms are generally associated with allergic conjunctivitis manifested by ocular itching, lacrimation, and conjunctival injection. Severe conjunctivitis is less common in PAR than in SAR. The best explanation for this difference is that pollen grains affect the eyes when they are blown into them. With the exception of cat-derived allergens, which have a strong tendency to remain airborne, indoor allergens including those derived from dog and dust mites are less likely to be blown into the eyes because these allergens are on heavy particles that do not remain airborne. Instead, these indoor aeroallergens are drawn into the nose by breathing.

What is less well appreciated—in the literature but certainly not in AR sufferers—is that AR is a systemic disease associated with circulating activated T lymphocytes and mononuclear phagocytic cells. The activation of these cells is demonstrated by their production of cytokines associated with innate immunity, such as IL-1, TNF-α, and IL-6. These cytokines are responsible for the lethargy, fatigue, arthralgias, myalgias, and cognitive impairment that frequently accompany AR. These systemic symptoms, which are often the chief complaints of allergy sufferers, contribute to a diminished quality of life and are often severe enough to make normal activities difficult, including work or school.[4] Although fever is not a feature of AR, it is intriguing that the lay term for this condition is *hay fever*, a designation reflecting the pronounced influenza-like nature of this disease.

DIAGNOSIS

AR is primarily a clinical diagnosis based on symptoms and exposure history. It is a complex genetic disorder, and affected patients generally give a positive family history. Vasoactive mediators induce glandular secretions and vascular leakage along with engorgement of capacitance nasal venous sinusoids. Physical examination therefore reveals the nasal mucosa to be cyanotic and swollen, with clear secretions. Smears of nasal secretions (Hansel stains) are seldom required, but when they are performed, they typically reveal eosinophils.

The diagnosis of AR is confirmed by the demonstration of specific IgE antibodies reactive to the relevant allergen through either positive allergy skin test responses or IgE immunoassays (Video 251-1). Identification of the specific triggering allergen is essential for the institution of appropriate environmental controls. Prick skin testing is safe, specific, and rapid, and it is the diagnostic test of choice for identifying relevant allergens. A positive intradermal test response in the presence of a negative prick skin test response is often a false-positive result. Intradermal testing is therefore rarely indicated and is associated with potentially life-threatening systemic reactions. If prick skin testing is not available or if the test cannot be performed (e.g., in patients with eczema or dermatographism or in those using antihistamines or antipsychotic agents), IgE immunoassays are an appropriate alternative. As with prick skin tests, positive IgE immunoassays correlate with symptoms caused by natural exposure, establish the diagnosis of AR, and can form the basis for environmental therapy; these assays should therefore be used extensively by primary care physicians who manage patients with AR. However, a negative IgE immunoassay with a strong clinical suspicion suggests the need for referral.

Differential Diagnosis

Other causes of rhinitis are shown in Table 251-2. The approach to patients with sneezing and rhinorrhea is illustrated in Figure 251-1.

Viral rhinitis may be difficult to distinguish from SAR. Viral rhinitis is not associated with the release of mast cell mediators. The main mediators present in nasal secretions from patients with the common cold are kinins; histamine and cysteinyl leukotrienes are less prevalent. In contrast to the pervasive eosinophilia in AR, infectious rhinitis is defined by the presence of polymorphonuclear leukocytes, reflecting in large part the secretion of IL-8 by infected epithelium. The presence of these different mediators is in keeping with the observation that pruritus, paroxysmal sneezing, and clear secretions help distinguish SAR from viral rhinitis, along with the distinct recurrent, seasonal nature of SAR. Viral rhinitis produces thicker, purulent secretions, with neutrophils present on the nasal smear. Conjunctival symptoms are less pronounced, and on physical examination, the nasal mucosa is erythematous and swollen.

Hormonal influences that may produce chronic nasal congestion and rhinorrhea include hypothyroidism, birth control pill use, pregnancy, and menopause. Abuse of topical nasal decongestants (e.g., oxymetazoline), with chronic reflex vasodilation, has historically been the most common cause of rhinitis medicamentosa; however, cocaine and nasal narcotic abuse is another frequent cause of this condition. Chronic unilateral nasal blockage suggests an anatomic defect, typically a deviated or fractured septum, but such

blockage can also result from polyps, foreign bodies, and tumors, which can be malignant. This history necessitates referral for rhinoscopy (Video 251-2) and possibly computed tomography (CT) of the nose and sinuses. Nasal septum deviation is an unlikely cause of bilateral nasal congestion, and surgical therapy has little role in the treatment of rhinitis that is producing symptomatic congestion.

TABLE 251-2 DIFFERENTIAL DIAGNOSIS OF RHINITIS

ALLERGIC
Seasonal allergic rhinitis (SAR)
Perennial allergic rhinitis (PAR)
INFLAMMATORY
Infectious rhinitis (viral)
Nonallergic rhinitis with eosinophilia syndrome (NARES)
Chronic sinusitis with or without nasal polyposis
Laryngopharyngeal reflux
HORMONAL
Pregnancy, oral contraceptives, perimenopause
Hypothyroidism
Hyperthyroidism
RHINITIS MEDICAMENTOSA
Topical or, less commonly, oral decongestants
Antihypertensives
Antidepressants
Cocaine
VASOMOTOR
Irritant induced (pollution, cigarette smoke)
Cold air induced
Gustatory (food induced)
ANATOMIC
Nasal septal deviation
Tumor, neoplasm
Foreign body
Cerebrospinal fluid leak
Atrophic (postsurgical or trauma)

An abnormal neurogenic response to irritants (e.g., cold air, pollutants, cigarette smoke, strong odors, alcohol, foods) is the predominant feature of vasomotor rhinitis. This disorder is characterized by nasal autonomic nerve dysfunction. Patients with vasomotor rhinitis typically have chronic nasal congestion and posterior pharyngeal drainage, but they lack the paroxysmal sneezing, rhinorrhea, pruritus, conjunctivitis, and systemic complaints typical of patients with AR. In addition, eosinophils are absent in their nasal mucus. Topical antihistamines (nasal azelastine, olopatadine) are often effective in vasomotor rhinitis. Patients with this condition also respond to therapy with topical corticosteroids but also to atropine (nasal ipratropium).

An increasingly recognized cause of perennial nasal congestion and posterior pharyngeal drainage that often occurs in association with cough and hoarseness can be ascribed to laryngopharyngeal reflux. These patients are often otherwise asymptomatic from ("silent") gastroesophageal reflux, making recognition challenging. Intra-arytenoid erythema on laryngoscopy is traditionally ascribed to reflux; however, the specificity of this finding has never been determined, and other causes of chronic cough may produce identical findings. Because symptoms may not resolve until after several months of aggressive reflux treatment and because proton pump inhibitors by themselves may not prevent reflux of digestive enzymes (e.g., pepsin) or other stomach-derived irritants, proper diagnosis can be daunting. More recent techniques, including esophageal impedance testing and laryngeal smears for pepsin, are promising diagnostic advances.

Chronic sinusitis (see next section) with or without nasal polyposis can produce a spectrum of symptoms, including rhinorrhea, mucopurulent posterior pharyngeal drainage, and nasal congestion, that can be mistaken for PAR. However, chronic sinusitis, particularly when it is eosinophilic (discussed later), is often asymptomatic, and a reduced or absent sense of smell may be the only complaint. In contrast to acute (infectious) sinusitis, headaches are an unusual manifestation of either perennial rhinitis (allergic or nonallergic) or chronic sinusitis, and virtually all patients who complain of "sinus headaches" suffer from atypical migraines or other headache syndromes.[5] Atypical migraines routinely produce headaches that occur in a bilateral distribution involving the maxillary or ophthalmic branches of the

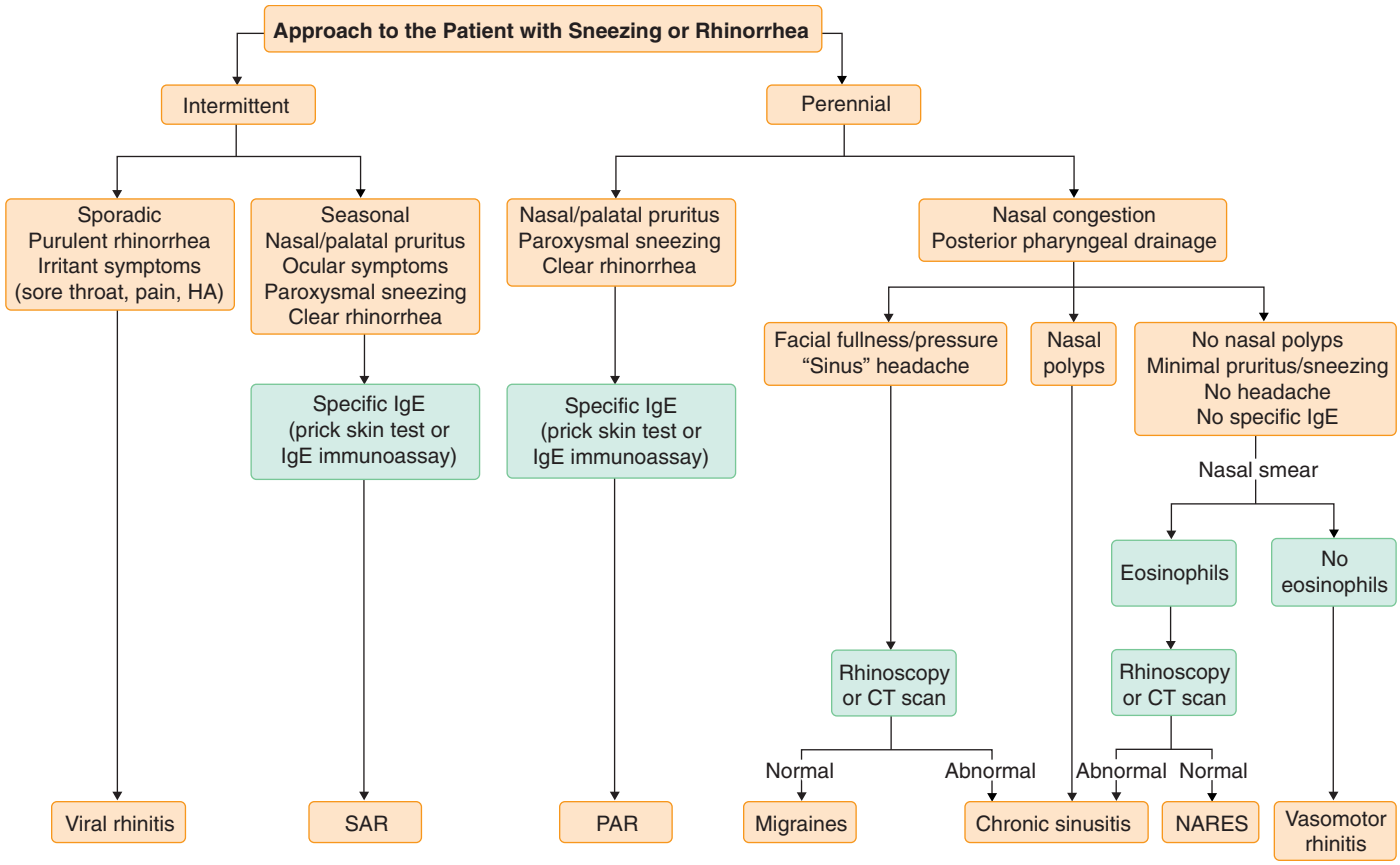

FIGURE 251-1. Approach to the patient with rhinitis symptoms. CT = computed tomography; HA = headache; IgE = immunoglobulin E; NARES = nonallergic rhinitis with eosinophilia syndrome; PAR = perennial allergic rhinitis; SAR = seasonal allergic rhinitis.

trigeminal nerve; this distribution, especially when it is combined with vasomotor symptoms such as nasal congestion, rhinorrhea, and conjunctival injection, often leads to misdiagnosis as chronic sinusitis or rhinitis. This complex becomes even more confusing with the recognition that vasoactive mediators released in association with allergic (or nonallergic) rhinitis (e.g., histamine) can trigger migraines. Because of this overlap in symptoms among chronic sinusitis, perennial rhinitis, and atypical migraines and their synergistic influences on each other, objective evaluation with either CT or rhinoscopy is usually required to establish the diagnosis of chronic sinusitis.

Atrophic rhinitis is characterized by atrophy of the nasal epithelium and is associated with complaints of nasal congestion and a perceived bad odor. It is observed in elderly patients, but the most common cause is devascularization secondary to nasal surgery or trauma.

Finally, a nonallergic nasal disease characterized by prominent eosinophilic inflammation has been described and termed non-AR with eosinophilia syndrome (NARES). On further analysis, many of these patients prove to have chronic sinusitis and nasal polyps. Patients with NARES present with symptoms similar to those of vasomotor rhinitis. NARES is diagnosed by performing a nasal smear (Hansel stain) for eosinophils; however, a CT scan or rhinoscopy is required to rule out chronic sinusitis or nasal polyposis because these conditions may be minimally symptomatic. In contrast to vasomotor rhinitis, NARES is more responsive to intranasal cromolyn and intranasal corticosteroids.

CHRONIC SINUSITIS

Chronic sinusitis represents many disease processes, including those caused by chronic bacterial infections, cystic fibrosis, immotile cilia syndrome, immune deficiencies, nonspecific inflammation, hypersensitivity to colonized fungi (allergic fungal sinusitis), aspirin-exacerbated respiratory disease (Samter's triad), and a disorder termed chronic hyperplastic eosinophilic sinusitis (CHES) (Table 251-3).[6] In the absence of immunodeficiency or cystic fibrosis, infection is a *very* unusual cause of chronic sinusitis. Noneosinophilic forms of chronic sinusitis most often reflect congenital anatomic variants or acquired nasal disorders (e.g., allergic rhinitis) that lead to ostial occlusion, with the secondary development of frequent, protracted acute infections. This ultimately produces tissue remodeling within the sinuses (with or, more often, without nasal polyps). These patients typically respond to surgical interventions that address the underlying anatomic precipitant.

TABLE 251-3 HETEROGENEITY OF CHRONIC SINUSITIS

PHENOTYPE	CHARACTERISTIC
Infectious	Very unusual as a cause of chronic sinusitis Occurs in association with cystic fibrosis, immune deficiency, ciliary dyskinesis
Inflammatory	Reflects remodeling secondary to frequent, recurrent acute sinusitis Develops secondary to anatomic obstruction of the sinus ostia; often resolves after surgical correction Pathologic examination demonstrates dense collagen fibrils, goblet cell metaplasia, and glandular hypertrophy and a chronic mononuclear infiltrate, with or without neutrophils Bacterial biofilms contribute to presence and severity Less often associated with nasal polyps
Hyperplastic eosinophilic	Prominent eosinophilic infiltrate with systemic eosinophilia Pathologic examination demonstrates edema, subbasement thickening Frequent association with allergies and asthma Usually associated with nasal polyps
Allergic fungal	Often unilateral; presents as expansive, dense infiltrate on CT scan Associated with elevated total IgE and specific IgE to colonizing fungi
Aspirin-exacerbated respiratory disease (Samter's triad)	Intense eosinophilic infiltrate Usually associated with nasal polyps and asthma Exacerbations of upper and lower respiratory disease (asthma) after ingestion of aspirin and nonselective COX1 inhibitors

COX1 = cyclooxygenase 1; CT = computed tomography; IgE = immunoglobulin E.

CHES, which more often occurs in association with nasal polyposis, is an inflammatory disorder characterized by the accumulation of eosinophils, fibroblasts, mast cells, other stromal cells, and T_H2-like lymphocytes along with goblet cell metaplasia and mucous gland hypertrophy. The prominent accumulation of eosinophils, however, is the diagnostic feature of this condition. The inflamed tissue expresses cytokines responsible for eosinophil hematopoiesis (IL-5), survival (IL-3, IL-5, and GM-CSF), recruitment (CCL11 [eotaxin]), and activation (CCL11, CCL5 [RANTES], IL-3, IL-4, IL-5, GM-CSF, and TNF-α). This histologic appearance and immune profile of CHES bears striking similarity to asthma. Indeed, approximately half of patients with CHES have asthma, and virtually all asthmatics have some component of CHES. Together, these observations suggest that CHES and asthma are similar diseases affecting the upper and lower portions of the airway (the "unified" airway concept). Whereas surgery may be a useful therapeutic adjuvant, as with asthma, treatment requires lifelong anti-inflammatory therapy. Therapies useful in asthma, such as topical corticosteroids and biotherapeutics that target IgE and IL-5, also appear useful in CHES.[7,8]

PREVENTION

Avoidance and Environmental Control

When it is feasible, avoidance or elimination of the source of the allergen is the treatment of choice for patients with AR. Avoidance studies in AR are limited, and the amount of allergen reduction needed to alleviate symptoms is unknown. Avoidance studies in asthma (Chapter 87) provide compelling evidence of a beneficial effect on bronchial hyperreactivity, symptom severity, and need for β-agonist rescue therapy.

Dust mite avoidance involves four principles: (1) remove reservoirs for mite growth (i.e., use allergen-impermeable mattress and pillow covers), (2) keep the relative humidity lower than 50%, (3) wash bedding in hot water (130° F) to kill mites, and (4) wear a simple mask when dust is being disturbed. Many of the measures suggested for mites are also helpful for fungi, especially dehumidification. Windows, shower curtains, and indoor plants are important sites for fungal growth and can be treated with mild fungicides (dilute household bleach).

In some houses, and particularly urban apartment blocks, large numbers of cockroaches are present, and IgE sensitivity is common. Although it may be difficult to kill cockroaches in an apartment, it is usually possible to keep a house clear of cockroaches by using chemical sprays and traps. Care must be taken in using chemical sprays because they can be an irritant to asthmatic patients.

Air-conditioning with closed windows is useful for reducing seasonal allergens, and the dehumidification provided by air-conditioning also mitigates the mite and indoor mold load.

Pets, especially cats, are the most preventable source of allergic diseases. Animal dander accumulates in houses during a prolonged period and takes many months to eliminate after the pet is removed. Although it is difficult to persuade patients to get rid of their animals, it may be possible to move the pet outside into the garage or to restrict the pet's access to certain parts of the house. Dogs kept outside and allowed into the house only occasionally do not appear to be an important cause of symptoms. Cat allergy is a much more serious problem because a single cat can deposit a huge concentration of allergen, and cat allergen remains airborne for prolonged periods. Cat owners, in turn, deposit sufficient concentrations of allergen in classrooms and work environments to induce symptoms in their allergic colleagues. The dominant rodent allergen is a urinary protein, and rodents, like cats, can deposit large quantities of allergen in a house.

TREATMENT

Rx

Although avoidance interventions can reduce allergen levels, they often fail to produce clinically significant improvement. As a result, pharmacotherapy is frequently required.

Antihistamines

Antihistamines are the oldest drugs used to treat AR and are considered first-line therapy. Antihistamines compete with histamine for the H_1-receptor sites that contribute to sneezing, itching, rhinorrhea, and conjunctivitis. Oral antihistamines ameliorate these symptoms of AR but generally do not improve nasal congestion. They also inhibit mast cell activation as manifested by diminished histamine, cysteinyl leukotrienes, and mast cell tryptase secretion. First-generation antihistamines cross the blood-brain barrier and have significant

sedative and anticholinergic effects. In addition to causing sleepiness, they interfere with school, work, driving, or use of machinery, and as such, the use of these drugs is no longer recommended. Second-generation antihistamines have a longer duration of action, do not cross the blood-brain barrier, and are nonsedating. These agents include cetirizine, levocetirizine, fexofenadine, descarboxyloratadine, and loratadine. Although less sedating than their parent compound hydroxyzine, cetirizine and levocetirizine may occasionally produce sedation. The intranasal antihistamines azelastine and olopatadine have a more rapid onset of action than oral antihistamines, and reflecting the high local concentrations they are able to achieve, they do have decongestant efficacy. They are also distinguished from oral antihistamines in being effective for nonallergic forms of rhinitis. No studies have convincingly demonstrated the superiority of one antihistamine over another.

As discussed earlier, the role of histamine diminishes during the course of an allergy season or with PAR, making antihistamines less effective. Antihistamines are effective for acute allergic reactions that are mediated predominantly by mast cell–derived histamine; as such, they are most beneficial in patients with intermittent allergen exposures, such as occasional outdoor exposure during pollen season. In patients with continuous allergen exposures, however, such as PAR caused by indoor allergens or after several days of continuous exposure to seasonal allergens, these drugs often prove to be little better than placebo.

Decongestants

Decongestants such as pseudoephedrine treat nasal stuffiness but are mild stimulants and even in oral formulations may produce rebound congestion and headaches. These drugs are usually used in combination with antihistamines to control the full spectrum of AR symptoms. Antihistamines and decongestants alone generally do not provide satisfactory relief in patients with moderate to severe AR.

Leukotriene Modifiers

Leukotriene modifiers (zileuton, zafirlukast, montelukast) have a confirmed efficacy in AR that is comparable to that of antihistamines but do significantly improve sneezing, rhinorrhea, nasal congestion, ocular symptoms, and quality of life in patients with both SAR and PAR. This efficacy reflects the presence and importance of these proinflammatory vasoactive mediators in AR.

Nasal Cromolyn

Nasal cromolyn stabilizes mast cells and mediates additional anti-inflammatory activities. Although it is not as effective as intranasal corticosteroids, cromolyn provides relief in patients with mild to moderate symptoms. The value of cromolyn is mitigated by the need for frequent doses (four times/day), a lack of efficacy in approximately 30 to 40% of recipients, and the superior efficacy of intranasal corticosteroids in controlled studies. Cromolyn may be especially useful preventively (e.g., immediately before cat exposure). Ocular cromolyn has been especially useful in the treatment of allergic conjunctivitis. No significant side effects are associated with its use.

Intranasal Corticosteroids

Intranasal corticosteroids (fluticasone [Flonase, Veramyst], beclomethasone [Qnasl], triamcinolone [Nasacort], flunisolide [Nasarel], budesonide [Rhinocort], mometasone [Nasonex], and ciclesonide [Omnaris]) are the most effective AR treatments and are considered the treatments of choice for patients with moderate to severe SAR or PAR.[A1] Comparative studies of antihistamines and intranasal corticosteroids consistently favor the corticosteroids.[A2] In well-performed placebo-controlled studies, intranasal corticosteroids provided a 50 to 90% reduction in symptoms (compared with 20 to 30% for antihistamines). In contrast to oral antihistamines, topical corticosteroids reduce nasal congestion in addition to relieving itching, rhinorrhea, sneezing, *and* allergic conjunctivitis.[9] Few studies have directly addressed the influence of corticosteroids on the systemic effects of AR—including missed work and school, poor productivity, reduced cognition, and fatigue—that are often the dominant complaints of patients with AR; however, intranasal corticosteroids significantly improve quality of life, reflecting relief from these complaints. Although they are slower in onset than antihistamines, intranasal corticosteroids produce some clinical improvement as quickly as 6 to 8 hours and achieve full effectiveness after several days.

Topical corticosteroid therapy does not inhibit IgE synthesis or mast cell degranulation, traditionally considered the two determinants for the development of AR. However, corticosteroids do inhibit T-lymphocyte proliferation, chemokine and cytokine production, recruitment of eosinophils and basophils, mucus secretion, vascular permeability, and mast cell proliferation. Intranasal corticosteroid use is therefore associated with diminished nasal eosinophilia, mast cell numbers, and cytokine expression. The efficacy of intranasal corticosteroids emphasizes the importance of these nonhistamine mechanisms to the pathophysiologic process of AR.

Several intranasal corticosteroid preparations are currently available and differ according to dose, approval age, and propellant (Table 251-4). No studies have demonstrated superior efficacy for any of the nasal corticosteroid preparations. Clinical experience with asthma (Chapter 87) suggests that patients refractory to one intranasal corticosteroid may be switched to a higher-potency corticosteroid; however, the best evidence is that all these agents are comparably valuable when patients are willing to comply with their use. Choices should be based primarily on the patient's preference.

TABLE 251-4 INTRANASAL CORTICOSTEROIDS*

GENERIC NAME	DOSE (PER ACTUATION)	MINIMUM APPROVED AGE	USUAL DOSING
Beclomethasone AQ	42 µg	6 yr	Twice daily
Beclomethasone HFA	80 µg	12 yr	Once daily
Flunisolide	25 µg	6 yr	Twice daily
Triamcinolone	55 µg	2 yr	Once daily
Budesonide	32 µg	6 yr	Twice daily
Fluticasone furoate	27.5 µg	2 yr	Once daily
Fluticasone propionate	50 µg	4 yr	Once daily
Mometasone	50 µg	2 yr	Once daily
Ciclesonide	50 µg	6 yr	Once daily
Ciclesonide HFA	37 µg	12 yr	Once daily

*Intranasal corticosteroids are generally administered at 2 sprays per nostril.

There is no convincing evidence of clinically significant systemic absorption or systemic side effects from intranasal corticosteroids. Given these drugs' hydrophobicity, local metabolism, and lack of absorption from lung tissue, clinically meaningful systemic absorption from the nasal passages is unlikely. Intranasal corticosteroids, even at greater than recommended doses, do not significantly suppress serum or urinary cortisol levels. There is a small but statistically significant effect on short-term growth velocity in children with PAR but minimal or no impact on ultimate adult height. Nasal corticosteroids cause minimal topical side effects, including local irritation, dryness, and epistaxis. Nasal perforation has been reported, but primarily in the setting of underlying devascularization (previous trauma, surgery, or cocaine abuse).

Immunotherapy

The clinical efficacy of subcutaneous immunotherapy for AR caused by grass, ragweed, many other pollens, cat and dog dander, and dust mites has been categorically established in innumerable well-designed, controlled studies.[10] Immunotherapy decreases the severity of AR, reduces the need for pharmacotherapy, and significantly improves quality of life. In patients with severe AR and conjunctivitis poorly controlled by antihistamines and intranasal corticosteroids, immunotherapy can reduce allergen sensitivity by more than 10-fold as well as significantly decrease total symptoms and reduce total antiallergic drug use. Efficacy depends on delivery of the correct antigen, regular injections for 3 to 5 years, and administration of an adequate dose of the allergen (10 to 15 µg, a dose much higher than that used historically).

Immunotherapy is indicated primarily in patients with refractory rhinitis or in those experiencing unacceptable side effects from standard medications. Because intranasal corticosteroids are not universally effective and do not provide complete relief in all patients, consideration of immunotherapy is necessary. In addition, despite the excellent safety profile of intranasal corticosteroids, many patients are reluctant to use them. Patients should normally go through at least one full pollen season before considering immunotherapy. Immunotherapy is the only treatment that produces long-term immune modulation. Both avoidance and pharmacotherapy are effective only if they are sustained. The effects of immunotherapy, in contrast, persist for many years after a 3- to 5-year course of treatment has been discontinued, and they could be lifelong. As such, a 5-year course of immunotherapy has cost advantages over lifelong pharmacotherapy. Many patients are attracted to immunotherapy by this potential for long-term immune modulation, remission of symptoms, and ability to discontinue daily pharmacotherapy.

Immunotherapy is associated with a small risk of fatal anaphylaxis (about 3 fatalities/year in the United States, of 2 million people receiving this form of treatment). Because of this risk, immunotherapy must be administered only in a facility where resuscitation equipment and trained personnel are available. Asthmatic patients are uniquely at risk for fatal anaphylaxis, so immunotherapy should be recommended with caution in these patients.

Because of the risk for anaphylaxis and the inconvenience of having to receive subcutaneous immunotherapy in a facility where evaluation and treatment of anaphylaxis are available, there is increasing enthusiasm for alternative forms of immunotherapy, specifically, sublingual immunotherapy. Sublingual immunotherapy has been extensively studied and proved to provide significant clinical benefit along with reduced need for pharmacologic therapy for numerous allergens including grass, ragweed, and dust mites.[A3-A6] In comparison to subcutaneous administration, sublingual immunotherapy is sufficiently

safe to permit home administration. As with subcutaneous immunotherapy, long-term clinical benefits are observed after it is discontinued, although recurrences are common after 5 to 10 years. Whereas few direct comparative studies have been performed, the preponderance of evidence suggests that subcutaneous immunotherapy does provide greater and longer lasting benefits.[11]

Summary

Many patients have multiple antigen sensitivities, and specific immunotherapy at effective doses may not be practical. Furthermore, immunotherapy has poor efficacy for many antigens, such as molds. These issues have led to a search for new immune-based therapies capable of attenuating allergic inflammation. Many experimental approaches being developed for asthma, including various anticytokine therapies, may have efficacy for AR.

In general, however, currently available therapies provide adequate relief for virtually all AR sufferers. The persistence of nasal symptoms in the face of adequate AR-directed pharmacotherapies should suggest the diagnosis of an alternative form of rhinitis (see Fig. 251-1) or the presence of sinusitis or—especially if headaches are present—atypical migraines and rebound headaches. Given the high frequency of AR in the population (and the even higher prevalence of allergic sensitization), treatment becomes particularly challenging when a patient presents with nonallergic nasal symptoms and either concomitant AR or clinically innocuous allergic sensitization (positive skin test responses or IgE immunoassays) occurring in the absence of actual AR.

Grade A References

A1. Zalmanovici Trestioreanu A, Yaphe J. Intranasal steroids for acute sinusitis. *Cochrane Database Syst Rev*. 2013;12:CD005149.

A2. Glacy J, Putnam K, Godfrey S, et al. Treatments for Seasonal Allergic Rhinitis [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013. Report No.: 13-EHC098-EF.

A3. Aasbjerg K, Backer V, Lund G, et al. Immunological comparison of allergen immunotherapy tablet treatment and subcutaneous immunotherapy against grass allergy. *Clin Exp Allergy*. 2014;44:417-428.

A4. Creticos PS, Maloney J, Bernstein DI, et al. Randomized controlled trial of a ragweed allergy immunotherapy tablet in North American and European adults. *J Allergy Clin Immunol*. 2013;131:1342-1349.

A5. Wang DH, Chen L, Cheng L, et al. Fast onset of action of sublingual immunotherapy in house dust mite–induced allergic rhinitis: a multicenter, randomized, double-blind, placebo-controlled trial. *Laryngoscope*. 2013;123:1334-1340.

A6. Lin SY, Erekosima N, Kim JM, et al. Sublingual immunotherapy for the treatment of allergic rhinoconjunctivitis and asthma: a systematic review. *JAMA*. 2013;309:1278-1288.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Davies JM, Platts-Mills TA, Aalberse RC. The enigma of IgE^+ B-cell memory in human subjects. *J Allergy Clin Immunol*. 2013;131:972-976.
2. Tengroth L, Millrud CR, Kvarnhammar AM, et al. Functional effects of Toll-like receptor (TLR)3, 7, 9, RIG-I and MDA-5 stimulation in nasal epithelial cells. *PLoS ONE*. 2014;9:e98239.
3. Walker JA, McKenzie AN. Development and function of group 2 innate lymphoid cells. *Curr Opin Immunol*. 2013;25:148-155.
4. Patterson AM, Yildiz VO, Klatt MD, et al. Perceived stress predicts allergy flares. *Ann Allergy Asthma Immunol*. 2014;112:317-321.
5. Marmura MJ, Silberstein SD. Headaches caused by nasal and paranasal sinus disease. *Neurol Clin*. 2014;32:507-523.
6. Sarber KM, Dion GR, Weitzel EK, et al. Approaching chronic sinusitis. *South Med J*. 2013; 106:642-648.
7. Gevaert P, Van Bruaene N, Cattaert T, et al. Mepolizumab, a humanized anti-IL-5 mAb, as a treatment option for severe nasal polyposis. *J Allergy Clin Immunol*. 2011;128:989-995.
8. Gevaert P, Calus L, Van Zele T, et al. Omalizumab is effective in allergic and nonallergic patients with nasal polyps and asthma. *J Allergy Clin Immunol*. 2013;131:110-116.
9. Baroody FM, Naclerio RM. Nasal-ocular reflexes and their role in the management of allergic rhinoconjunctivitis with intranasal steroids. *World Allergy Organ J*. 2011;4(suppl):S1-S5.
10. Roche AM, Wise SK. Subcutaneous immunotherapy. *Int Forum Allergy Rhinol*. 2014;4(Suppl 2):S51-S54.
11. Dretzke J, Meadows A, Novielli N, et al. Subcutaneous and sublingual immunotherapy for seasonal allergic rhinitis: a systematic review and indirect comparison. *J Allergy Clin Immunol*. 2013;131: 1361-1366.

REVIEW QUESTIONS

1. A patient presents with a chief complaint of recurrent, protracted headaches that occur bilaterally in a maxillary distribution and are associated with nasal congestion and ocular tearing. These episodes occur spontaneously throughout the year. The most common cause of this condition is

A. Allergic rhinitis
B. Chronic sinusitis
C. Migraine
D. Nasal polyps
E. Rebound headache

Answer: C The most common cause of "sinus" headache is migraine. Whereas acute sinusitis is often painful, chronic sinusitis is virtually never a source of pain. Migraines routinely occur in a maxillary distribution, are often bilateral, and often are inappropriately ascribed to sinusitis when they are associated with vasomotor symptoms such as nasal congestion and ocular tearing. Allergic rhinitis may trigger migraines and acute sinusitis but by itself would be an unusual cause of headaches. Nasal polyps do not produce pain. Rebound headaches most commonly occur in the setting of caffeine or short-acting nonsteroidal anti-inflammatory drugs.

2. A 55-year-old women presents with a chief complaint of continuous nasal congestion and posterior pharyngeal drainage. Symptoms are present year-round and interfere with her sleep and quality of life. Blood tests demonstrate specific IgE to cottonwood trees, *Alternaria*, and ragweed. The most likely explanation for this syndrome is

A. Allergic rhinitis
B. Chronic sinusitis
C. Atypical migraine
D. Nonallergic rhinitis
E. Nasal polyps

Answer: D Nonallergic rhinitis most often is manifested with the combination of nasal congestion and posterior pharyngeal drainage. The absence of typical allergy symptoms of paroxysmal sneezing, pruritus, clear anterior secretions, and concomitant allergic conjunctivitis argues against allergic rhinitis as a cause. Allergic sensitization (presence of specific IgE by blood or skin testing) is common, is often not associated with symptoms on natural exposure, and therefore should not by itself be used to diagnose an allergic disease. In addition, the absence of sensitization to perennial allergens that can explain a year-round syndrome rules out an allergic etiology in this case. Whereas chronic sinusitis (with or without associated nasal polyps) can produce similar symptoms, these patients typically have additional complaints, such as "pressure" or "fullness" in a sinus distribution and reduced sense of smell. Migraines are episodic and are usually associated with headaches.

3. A 21-year-old patient presents with allergic rhinitis each spring, lasting continuously from early March through the end of June, that is associated with numerous pollen sensitizations. The most effective therapeutic intervention for this patient would be

A. Oral antihistamines
B. Intranasal cromolyn
C. Intranasal corticosteroids
D. Leukotriene modifiers
E. Immunotherapy

Answer: C Intranasal corticosteroids are the most effective treatment for allergic rhinitis and are compellingly superior to antihistamines, leukotriene modifiers, and cromolyn in well-performed trials. Antihistamines are effective in mild intermittent disease, but once symptoms persist for more than a few days or weeks, they may prove little better than a placebo. Cromolyn and leukotriene modifiers are not as effective as corticosteroids. Immunotherapy may be effective, especially for grass pollen. It may be difficult to treat the patient for all of the pollinating trees present in her environment. In general, however, immunotherapy is reserved for patients who fail to respond to pharmacologic treatments.

4. A college student presents with a chief complaint of seasonal allergies that occur every year in late spring (May-June). Of the following, which is the most likely allergen explaining this history?

A. Birch pollen
B. Timothy grass pollen
C. Ragweed pollen
D. Dust mite (*Dermatophagoides pteronyssinus*)
E. *Alternaria*

Answer: B Early spring allergic rhinitis is triggered by trees. Grasses pollinate in late spring and are what is driving this patient's symptoms. Ragweed pollinates in late summer and autumn. *Alternaria* is an outdoor mold that is also particularly prevalent in the autumn. Whereas dust mites are often considered a perennial allergen, they are particularly active during periods of high humidity and as such often present as "seasonal" allergens with worse symptoms in the summer.

5. Left untreated, allergic rhinitis may contribute to the development of which of the following medical conditions?

A. Benign nasal tumors
B. Atypical migraines
C. Eosinophilic granuloma (Langerhans cell histiocytosis)
D. Eosinophilic esophagitis
E. Chronic inflammatory (noneosinophilic) sinusitis

Answer: E Chronic rhinitis of any cause, including allergic rhinitis, occludes the sinus ostia. This can lead to frequent, persistent sinus infections and ultimately the remodeling of sinus tissue that characterizes chronic inflammatory sinusitis. Chronic sinusitis may be associated with extrusion of sinus tissue into the nasal air space, a condition referred to as nasal polyps. However, neither chronic sinusitis nor persistent allergic rhinitis produces nasal tumors. Similarly, allergic rhinitis produces a local (nasal) eosinophilic inflammatory process but not either eosinophilic esophagitis or eosinophilic granuloma. Although allergic rhinitis may trigger a migraine, it is not a cause of headaches, atypical or otherwise.

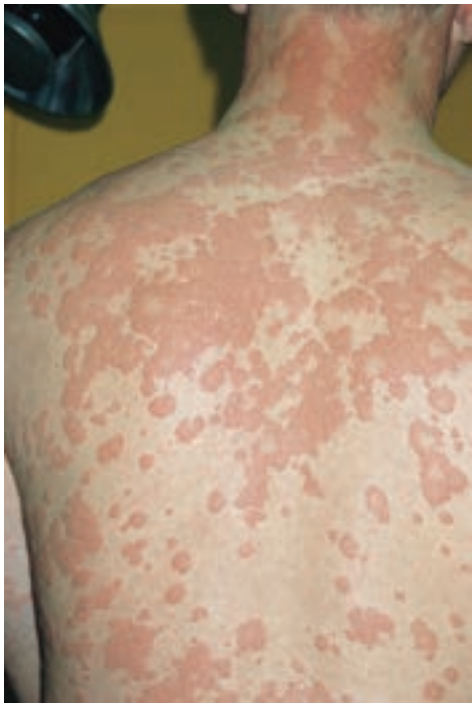

FIGURE 252-1. Extensive urticaria. Many presentations are more subtle. (From Roitt I, Brostoff J, Male D, eds. Immunology. 6th ed. London: Mosby; 2001.)

252

URTICARIA AND ANGIOEDEMA

STEPHEN C. DRESKIN

URTICARIA

Urticaria (hives) consists of pruritic, edematous, erythematous, blanching papules that are round or oval; have pale, raised centers (wheals); are several millimeters to a few centimeters in size; and are transient, lasting minutes to days (Fig. 252-1). Angioedema appears as a brawny, nonpitting edema, typically without well-defined margins and without erythema. Angioedema can be accompanied by a sense of burning, pressure, or aching but not pruritus; it is distinguished from other edematous states by its frequent involvement of the lips, tongue, eyelids, hands, feet, or genitalia and its rare occurrence in dependent areas of the body. Episodes (daily or almost daily symptoms) of recurrent hives or angioedema during a period of less than 6 weeks are considered acute, and those lasting longer are said to be chronic. Patients typically present with urticaria alone or urticaria with angioedema. Rarely, patients present with angioedema alone, and this becomes a diagnostic dilemma because angioedema as an isolated finding may be due to activation of the kinin system or may be due to activation of mast cells (see later). The terms *urticaria* and *urticaria/angioedema* are used interchangeably here to refer to illnesses characterized by urticaria or angioedema in which mast cells are activated.

EPIDEMIOLOGY

Urticaria/angioedema occurs in 15 to 25% of individuals at some time during their lives and can affect both genders and all races. Acute urticaria is more common in young adults and children. Chronic urticaria is more common in adults, affecting women (75% of cases) more often than men.[1]

PATHOBIOLOGY

Mast cells, the primary effector cells in urticaria/angioedema, are found in high numbers throughout the body, particularly within the subcutaneous tissue. After activation of mast cells, there is a rapid (<10 minutes) release of histamine, leukotriene C_4, and prostaglandin D_2, leading to vasodilation, subcutaneous and intradermal leakage of plasma from postcapillary venules, and pruritus. In addition, there is the delayed (4 to 8 hours) production and secretion of inflammatory cytokines such as tumor necrosis factor-α, interleukin-4, and interleukin-5, leading to an inflammatory infiltrate and perpetuation of longer-lived lesions. Angioedema is formed by a similar extravasation of fluid, not superficially in the skin but in deeper dermal and subdermal sites.

Lesions of acute urticaria typically show subcutaneous edema with widened dermal papillae, swollen collagen fibers, and rare inflammatory cells. Most episodes of acute urticaria/angioedema are caused by immediate hypersensitivity reactions to drugs or foods or result from inflammatory processes initiated by viral illnesses. The most common drugs that cause acute urticaria/angioedema are penicillins, sulfonamides, muscle relaxants, diuretics, and nonsteroidal anti-inflammatory drugs (NSAIDs), although any drug acting as a hapten can generate an allergic response. The predominant allergenic foods are milk, eggs, and peanuts for children and peanuts, tree nuts, fish, and shellfish for adults, although sensitization can occur to many other foods as well. These allergens cross-link immunoglobulin (Ig) E bound to the high-affinity receptor for IgE (FcεRI), leading to activation of mast cells. Some drugs (e.g., opioids, vancomycin, NSAIDs) and radiocontrast dye can activate mast cells by an IgE-independent (pseudoallergic) mechanism. Ingestion of fish contaminated with bacteria that produce histamine leads to hives as part of a toxic reaction to the histamine (scombroid food poisoning).

Lesions of chronic urticaria are characterized by similar edematous findings, with the addition of a dense perivascular inflammatory infiltrate

consisting of $CD4^-$ and $CD8^+$ T lymphocytes, eosinophils, basophils, and neutrophils. A minority of patients exhibit urticarial vasculitis, with lesions characterized by vascular destruction with leukocytoclasis.

The largest subgroup of chronic urticaria/angioedema is idiopathic urticaria, accounting for approximately 70 to 80% of cases. Recently, an effort has been made to replace the term *chronic idiopathic urticaria* with the more descriptive term *chronic spontaneous urticaria*. These patients have symptoms in the absence of a specific physical trigger, allergen exposure, or coexistent disease. Half the patients with idiopathic urticaria have evidence of autoimmunity based on the presence of IgG antibodies that can cross-link FcεRI or antithyroid antibodies. Some experts consider these patients to have a separate entity called autoimmune urticaria, whereas others consider these patients to have idiopathic urticaria with evidence of autoimmunity.[2]

Physical stimuli activate mast cells by unknown mechanisms and account for about 20 to 30% of cases of chronic urticaria. The most common of the physical urticarias is dermographism (also called dermatographism), in which wheals can be "written on the skin" by simple stroking or scratching. Cholinergic urticaria is a physical urticaria in which the trigger leading to mast cell activation is related to cholinergic stimuli occurring after exposure to heat or after exercise. Other physical stimuli can cause urticaria, including cold, solar radiation, pressure, vibration, and water. Cold-induced urticaria is rarely caused by cryoglobulins. Cold-induced urticaria needs to be distinguished from the cryopyrin-associated periodic fever syndromes that include familial cold autoinflammatory syndrome and the Muckle-Wells syndrome (see later and Chapter 261).

In approximately 1-2% of patients with chronic urticaria/angioedema, symptoms appear to be caused by ingestants (e.g., foods, medications, dietary supplements), contactants (e.g., soaps, detergents, cosmetics, hair or nail products, latex), concomitant infections, hormonal changes, or systemic illnesses. A food must be consumed regularly to cause chronic urticaria. Wheat is rarely found to be a trigger. Multicellular parasites (e.g., those causing strongyloidiasis or filariasis) elicit strong IgE responses and are important causes of chronic urticaria in endemic areas. Rare patients report that their urticaria occurs only during menses or is worsened by menses. Although hormonal variation often is associated with worsening of urticaria, a detailed history may reveal ingestion of NSAIDs taken for uterine cramping. Chronic urticaria/angioedema can be associated with flares of rheumatic conditions, other autoimmune conditions (including Hashimoto thyroiditis), or neoplastic conditions. Occult neoplasia is exceedingly unlikely to be the cause of chronic urticaria.

TABLE 252-1 CLASSIFICATION OF URTICARIA AND ANGIOEDEMA

I. Acute urticaria/angioedema
  A. Hypersensitivity reactions
    1. Drug allergy
    2. Food allergy
    3. Insect allergy
  B. Idiopathic
  C. Pseudoallergic reactions
    1. Drugs
    2. Radiocontrast dye
  D. Toxic reactions
  E. Immune complex
    1. Serum sickness
    2. Transfusion related
    3. Postviral
II. Chronic urticaria/angioedema
  A. Idiopathic
    1. Autoantibody associated
      a. Anti-IgE receptor (FcεRI)
      b. Anti-IgE
      c. Antithyroid
      d. Other
    2. Not associated with autoantibodies
  B. Physical
    1. Dermographism
    2. Cholinergic
    3. Delayed pressure
    4. Solar
    5. Cold
    6. Vibratory
    7. Aquagenic
  C. Immune complex
    1. Urticarial vasculitis
    2. Collagen vascular disease associated
III. Urticaria pigmentosa and systemic mastocytosis
IV. Complement-related and kinin-mediated angioedema
  A. Hereditary angioedema
  B. Acquired angioedema
  C. Angiotensin-converting enzyme inhibitor–induced angioedema
  D. Renin inhibitor–induced angioedema
V. Urticaria pigmentosa and systemic mastocytosis

CLINICAL MANIFESTATIONS

Patients often report that the first sensation of urticaria is poorly localized pruritus that quickly develops into the typical lesions of urticaria. The intensity of the pruritus varies from a minor inconvenience to an unbearable sensation that can lead to self-inflicted abrasion of the skin. Groups of hives often appear together during a short period, and episodes of hives can come in waves starting several times a day. Patients with cholinergic urticaria usually have a distinctive clinical presentation of diffuse urticarial lesions measuring a few millimeters in diameter on exertion sufficient to cause sweating. The pruritus is particularly intense, and all the symptoms are generally limited to the skin. A self-rated quality-of-life survey of patients with chronic urticaria revealed dramatic impairment in terms of loss of sleep, fatigue, and emotional discomfort. Angioedema can originate near a wheal or independently in other parts of the body. Symptoms vary from minor discomfort to an intense sense of pressure and may lead to other symptoms, such as severe shortness of breath if there is compromise of the airway. Rarely, patients report angioedema beginning 4 to 6 hours after application of local pressure, and this is called delayed pressure urticaria. This is a debilitating condition that is often difficult to treat.

DIAGNOSIS

The first episode of acute urticaria/angioedema may occur in the absence of an identifiable stimulus. If hives occur 5 to 30 minutes after ingestion of a drug or a food, the patient often can identify the association. If a physician is consulted, the best approach is to take a careful history, with attention to ingestants and intercurrent illnesses. Unnecessary drugs and food supplements should be discontinued, and any recently added medication should be changed to a structurally different agent. Most often, no causative agent is identified, and the hives are treated symptomatically (see the later discussion) for days or weeks before they resolve spontaneously.

Differential Diagnosis

The differential diagnosis of chronic urticaria/angioedema includes the subgroups of urticaria discussed earlier: idiopathic, autoimmune, physical, ingestant mediated, and associated with a variety of systemic illnesses.[3] Other conditions that can be confused with chronic urticaria/angioedema include diffuse pruritus complicated by dermographism, flushing disorders, urticarial vasculitis, urticaria pigmentosa, systemic mastocytosis, exercise-induced anaphylaxis, exercise-induced food-associated anaphylaxis, idiopathic anaphylaxis, hereditary angioedema, acquired angioedema, and angioedema associated with angiotensin-converting enzyme (ACE) inhibitors (Table 252-1).

Approximately 95% of patients with urticaria/angioedema are not reacting to an ingestant and do not have another illness that is causing their hives. However, it is sometimes difficult for patients (and some physicians) to accept this fact, prompting an extensive, invasive, expensive, and unnecessary investigation. The best "test" to identify patients with a specific underlying cause (i.e., physical trigger, autoimmune condition, allergen, or systemic disease) is a careful and detailed history and physical examination by a specialist knowledgeable in urticarial disease.

A good place to begin is by excluding possible physical triggers. Specific tests are available to establish the diagnosis of most physical urticarias, including scratching the skin and exposing the skin to heat, ice, vibration, pressure, ultraviolet radiation, or water. Cold urticaria must be distinguished from cryopyrin-associated periodic fever syndromes that are characterized by a cold-induced papular rash (not urticaria) and are now classified in the family of hereditary periodic fever syndromes. Solar urticaria must be distinguished from other types of light sensitivity, including metabolic abnormalities (e.g., erythrogenic porphyria) and photosensitivity due to drugs.

Even though foods and drugs are infrequent causes of chronic urticaria, many patients focus on ingestants and are not satisfied until these causes are ruled out. As in the evaluation of acute urticaria, the patient should discontinue all food supplements and medications that are not absolutely necessary and, if possible, change essential medications to structurally unrelated compounds. The patient then keeps a food diary to identify suspect foods that can be eliminated. Some allergists use skin tests with foods to identify "suspects" (Chapter 249), but this approach is unproven. For highly motivated patients, 2 weeks of a severely restricted diet, often based on lamb and rice, is recommended. Antihistamines and other medications used to control the urticaria must be discontinued. If the urticaria resolves, it is critical to reintroduce foods in a controlled fashion to identify the specific food causing the urticaria and to reinstate a healthy diet.

Chronic infections, including sinus infection, dental abscess, *Helicobacter pylori* gastric infection, cholecystitis, onychomycosis, and tinea pedis, have been associated with urticaria. Case reports indicate the resolution of urticaria after treatment of these infections, although rigorous proof of an association is lacking. The natural history of chronic urticaria probably accounts for coincidental spontaneous improvement after treatment of these conditions, at least in some cases.

Laboratory evaluation in a patient with typical urticaria should always include a complete blood count with differential, basic metabolic panel, liver enzymes, and urinalysis. Specialists are not in full agreement about the necessity of additional laboratory testing. Levels of thyroid-stimulating hormone and antithyroid antibodies may be measured in otherwise euthyroid-appearing patients to screen for subclinical Hashimoto thyroiditis. Skin tests for immediate hypersensitivity to foods may be ordered for patients with a suggestive history. Some specialists order no screening tests at all. As in vitro tests for anti-FcεRI autoantibodies have become more widely available, some specialists will perform this test. A positive test response for anti-FcεRI autoantibodies is useful because this reassures the patient that the urticaria is being driven by an internal process and is not caused by an ingestant or occult illness. Other tests should be ordered only as a result of positive findings in the history and physical examination.[4]

Although it is not routinely indicated in every case of chronic urticaria, a skin biopsy can provide useful information. The most common indication for this procedure is to rule out urticarial vasculitis when the hives are more painful than pruritic, last longer than 24 hours, or leave discolored skin. The presence of vascular destruction, fibrinoid necrosis, and immune complex deposition on microscopic examination (including immunofluorescence) should lead to a consideration of the specific causes of urticarial vasculitis (e.g., systemic lupus erythematosus) and the rapid initiation of more aggressive treatment.

Primary mast cell disorders rarely manifest as chronic urticaria (Chapter 255). Systemic mastocytosis is a very rare condition characterized by increased numbers of atypical mast cells in the bone marrow, skin, and other organs. Levels of tryptase (an enzyme specific for mast cells) are usually elevated in the serum. This condition is frequently accompanied by episodic flushing, urticaria pigmentosa, prominent gastrointestinal symptoms, neuropsychiatric symptoms, or recurrent anaphylaxis. Urticaria pigmentosa is characterized by distinctive pigmented cutaneous lesions containing nests of mast cells and is not easily confused with urticaria/angioedema.

Hereditary angioedema, acquired angioedema, and angioedema associated with ACE inhibitors are discussed later in this chapter. Briefly, these syndromes are characterized by episodic swelling without urticaria and are best identified by a careful history, physical examination, and focused laboratory evaluation.[5] An approach to the evaluation and treatment of patients with urticaria or angioedema is summarized in Figure 252-2.

TREATMENT

Rx

Acute urticaria is usually self-limited and responds well to histamine$_1$ (H_1)–type antihistamines. Antihistamines work better if they are taken prophylactically rather than after histamine has been released and is bound to the receptor. Patients often self-medicate with or are prescribed diphenhydramine (25 to 50 mg every 6 hours) or hydroxyzine (25 to 50 mg every 6 hours), but they may experience significant sedation. Second-generation antihistamines such as cetirizine (10 mg nightly at bedtime), fexofenadine (180 mg/day), and loratadine (10 mg/day) are much better tolerated and can be effective, although doses up to four times the standard doses are sometimes necessary.

On occasion, a brief course of corticosteroids is warranted to control severe symptoms. Epinephrine (0.3 mL of 1:1000 intramuscularly) quickly (but transiently) reverses the signs and symptoms of urticaria and angioedema. Patients who have experienced potentially life-threatening angioedema or anaphylaxis should have ready access to self-injectable epinephrine and be knowledgeable about its indications, administration, and brief duration of action. β-Blockers not only can aggravate urticaria but also can interfere with the action of epinephrine. NSAIDs and codeine can lead to IgE-independent mast cell activation. These medications should be discontinued if it is clinically safe to do so.

H_1 antihistamines are also the cornerstone of therapy for chronic urticaria/angioedema but are frequently inadequate to control symptoms. Certain H_1 antihistamines have been proposed as "preferred" for particular subtypes of chronic urticaria, such as hydroxyzine (25 to 50 mg three to four times daily) for cholinergic urticaria or cyproheptadine (2 to 4 mg every 6 hours) for cold-induced urticaria. Some have advocated using multiple H_1 antihistamines, changing or "rotating" agents, or using them in dosages well above those approved by U.S. Food and Drug Administration (FDA) labeling procedures.[6]

Approximately 15% of histamine receptors in the skin are of the H_2 subtype; therefore, the addition of an H_2 antihistamine, such as ranitidine (150 mg twice daily) or famotidine (20 mg twice daily), is a logical adjunct to H_1 antihistamine therapy, providing additional clinical benefit. The tricyclic antidepressant doxepin (10 to 100 mg nightly at bedtime) has highly potent H_1 and H_2 antihistamine activity, with an H_1-receptor affinity almost 800 times that of diphenhydramine and an H_2-receptor affinity 6 times that of cimetidine, but its use can be limited by significant sedation and its tendency to stimulate the appetite, leading to significant gain of weight.

Symptoms often persist despite the use of maximal or supramaximal doses of antihistamines. This is not surprising, considering the number of vasoactive and pruritogenic mediators released by mast cells, of which histamine is only one. Antileukotriene medications, such as montelukast (10 mg/day) or zafirlukast (20 mg twice daily), can be added to antihistamines, with some success. Especially severe symptoms may require systemic corticosteroids (prednisone 10 to 60 mg/day) to achieve symptomatic control, but strong concerns about side effects limit their usefulness.

Refractory symptoms have been treated with a wide variety of other medications. Some of these medications (adrenergic agents, calcium-channel blockers) are thought to decrease the ability of mast cells to release mediators. Others drugs are anti-inflammatory (hydroxychloroquine, sulfasalazine, dapsone, colchicine), immunomodulatory (cyclosporine, tacrolimus, mycophenolate), or antimetabolic (azathioprine, cyclophosphamide, methotrexate). Other treatments of refractory autoimmune chronic urticaria include intravenous immune globulin, plasmapheresis, and omalizumab (anti-IgE).

Evidence-Based Treatments

Multiple randomized, placebo-controlled studies of chronic urticaria/angioedema have shown the efficacy of both sedating and nonsedating antihistamines.[7] If sedating antihistamines must be used, doxepin is more effective than diphenhydramine, but it must be titrated carefully to avoid significant sedation. H_2-receptor antagonists alone are ineffective, but a meta-analysis of four studies with a total of 144 subjects demonstrated that they are effective when combined with H_1-receptor antagonists. Prednisone is generally accepted as a mainstay of therapy in difficult cases but has not been formally studied. Cyclosporine (4 mg/kg/day) was shown to be effective in a randomized, placebo-controlled, parallel study of 30 patients with autoimmune urticaria.[8] Leukotriene C_4 receptor antagonists (montelukast, zafirlukast) do not provide any incremental benefit when added to an antihistamine. Omalizumab (a humanized monoclonal antibody that binds and inactivates IgE; 300 mg SQ every 4 weeks) has been shown to be effective in a large randomized placebo-controlled trial. Although well characterized by described in case studies and used by experts, sulfasalazine, hydroxychloroquine, dapsone, colchicine, methotrexate, azathioprine, and intravenous gamma globulin are immunomodulatory agents that appear to be effective but have not been formally studied.

PREVENTION

It is essential to encourage patients with chronic urticaria to accept the chronicity of their illness and to focus on achieving reasonable symptomatic control with effective treatments that cause the fewest side effects. Many patients with physical urticaria can learn to avoid or to minimize triggers. The few patients for whom chronic urticaria is a feature of systemic illness may find relief if the underlying condition is appropriately treated. An excellent example is that chronic urticaria in patients with clinically apparent thyroid disease often resolves once the thyroid disease is treated. For many patients, other factors that exacerbate their specific symptoms can be identified, including stress or anxiety, hormonal fluctuations, aspirin and other NSAIDs,

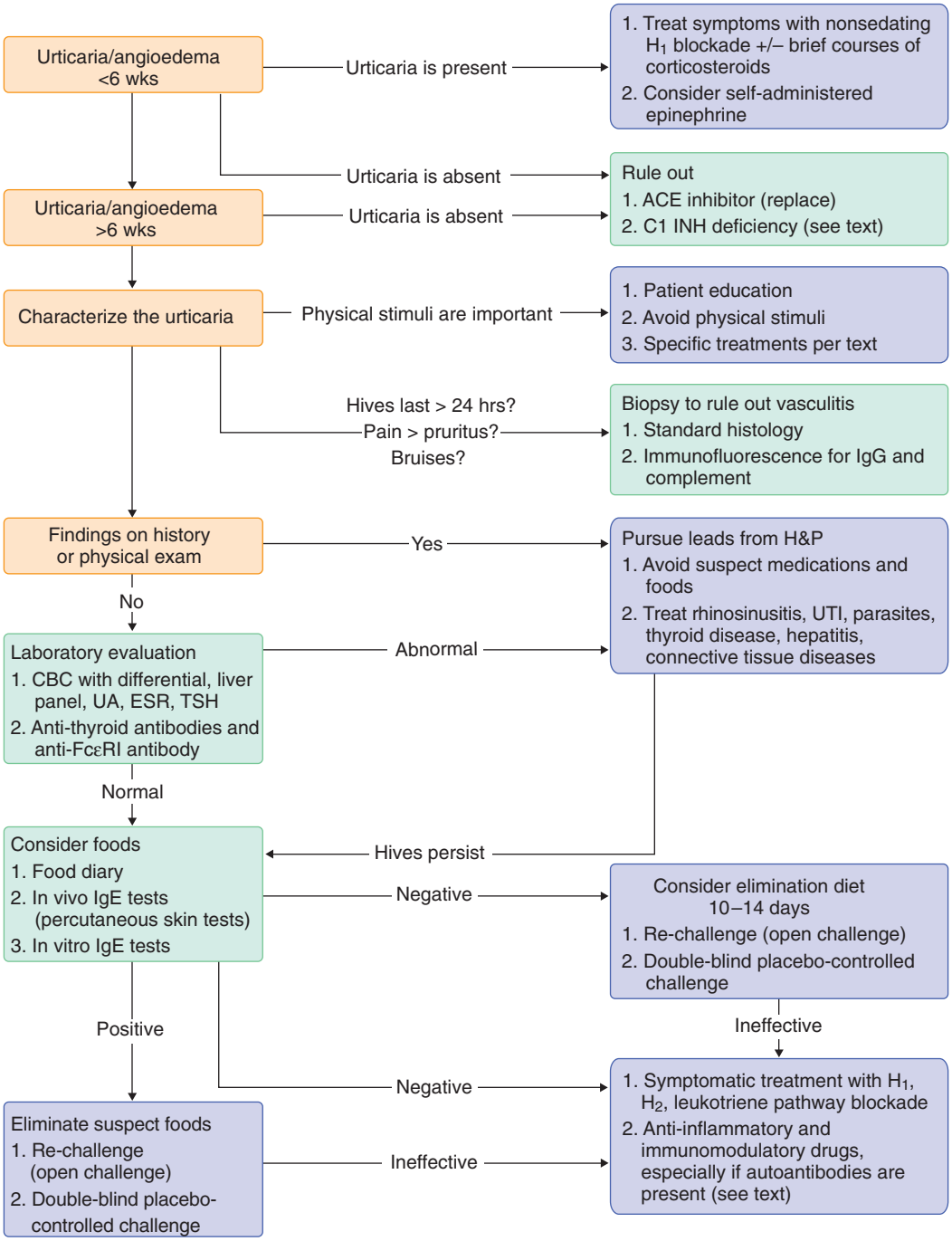

FIGURE 252-2. Evaluation and treatment of urticaria/angioedema. Treatment of urticaria with or without angioedema (AE) can be similar. However, treatment of AE without urticaria depends on the cause. If the AE is caused by an angiotensin-converting enzyme (ACE) inhibitor, discontinuation of the medication is required. Treatment of AE caused by a deficiency or dysfunction of C1 inhibitor (C1 INH) is discussed in the text. Idiopathic AE often responds to treatments described for urticaria/angioedema. CBC = complete blood count; ESR = erythrocyte sedimentation rate; FcεRI = high-affinity receptor for IgE; H & P = history and physical examination; H_1 = histamine$_1$-receptor antagonist; H_2 = histamine$_2$-receptor antagonist; IgE = immunoglobulin E; IgG = immunoglobulin G; TSH = thyroid-stimulating hormone; UA = urinalysis; UTI = urinary tract infection.

and agents that cause cutaneous vasodilation (e.g., alcohol, hot baths or showers, exercise, heated waterbeds). Psychosocial stress is a commonly reported trigger of worsening symptoms. A plausible biochemical mechanism may include increased release of cutaneous neuropeptides known to lower the threshold for mast cell degranulation.

PROGNOSIS

The prognosis for most patients with chronic urticaria/angioedema is excellent. Spontaneous resolution occurs within 12 months in 50% of patients and within 5 years in an additional 20%. However, 10 to 20% of patients, particularly those with physical or autoimmune urticaria, continue to have symptoms for as long as 20 years. Patients who had one episode of chronic urticaria that lasted for months or years and then resolved may experience one or more similar recurrences later in life.

FUTURE DIRECTIONS

The current trend in the treatment of urticaria/angioedema is to use multiple antihistamines and other agents that block the actions of the mediators produced by mast cells. In the near future, it is likely that patients will be treated earlier with anti-inflammatory and immunomodulatory drugs. Some agents under development for asthma and rhinitis may be useful for the treatment of urticaria/angioedema, including 5-lipoxygenase inhibitors, prostaglandin D_2-receptor antagonists, and more potent nonsedating antihistamines. Agents that decrease the sensitivity of mast cells to degranulation, such as phosphodiesterase 4 inhibitors and spleen tyrosine kinase (Syk) inhibitors, may also find a role in the treatment of this condition. In spite of the fact that chronic urticaria/angioedema is not thought to be an IgE-mediated disease, as mentioned before, omalizumab (anti-IgE) has been shown to be very

effective. This may be due to unexpected effects of IgE on mast cell activation. This finding could have a significant impact on future therapies.[9]

HEREDITARY ANGIOEDEMA AND RELATED DISEASES

DEFINITION

Hereditary angioedema (HAE) and related illnesses are characterized by recurrent attacks of angioedema mediated by vasoactive peptides such as bradykinin.[10]

EPIDEMIOLOGY

HAE affects approximately 1 in 50,000 people. It is an autosomal dominant disease and therefore affects 50% of offspring of both genders. Frequently, a history of several generations with this disease is obtained, but new mutations do occur, and a negative family history is not uncommon. Acquired angioedema (AAE) is more rare, affecting older persons who often have a monoclonal gammopathy or a malignant disease such as lymphoma. Angioedema associated with ACE inhibitors occurs in 0.1 to 0.2% of treated patients.[11]

PATHOBIOLOGY

HAE and AAE are caused by either low levels or abnormal function of a regulatory protein in the plasma, C1 inhibitor (C1 INH deficiency), which exerts control of the complement, fibrinolytic, and kinin-generating pathways. Because there is one normal gene, levels of C1 INH are detectable but, because of the abnormal gene, are not sufficient to control the generation of kinins. The C1 esterase enzyme, when activated, cleaves two complement products, C4 and C2. Without proper inhibition, this leads to low levels of circulating C4 and C2. C1 INH is also a critical modulator of the bradykinin pathway, and decreased C1 INH function leads to increased levels of bradykinin. Increased generation of bradykinin, not mediators from mast cells or activation of complement, leads to capillary leakage and angioedema. Changes in levels of C4 and C2, although not important in the pathophysiologic mechanism of the disease, are useful diagnostically.[12]

In HAE type I (85% of patients), the abnormal gene does not produce C1 INH. In HAE type II (15%), an antigenically detectable C1 INH protein is produced, but it is not functional. In HAE type III (very rare), C1 INH is present and functional, but there is a yet-to-be-defined abnormality in the generation of vasoactive compounds. In AAE, unknown factors activate C1 and deplete the C1 INH activity in plasma, or there is an autoantibody to C1 INH that interferes with its function. ACE inhibitor–associated angioedema is due to unintended inhibition of the enzyme that inactivates bradykinin; the complement pathway is unaffected.

CLINICAL MANIFESTATIONS

Children with HAE can have attacks shortly after birth, but these tend to be mild. For most patients, the severity of the attacks worsens at puberty, with episodes of swelling that can affect any external body surface, including the genitalia. Mucosal surfaces are also affected, and patients can have life-threatening swelling of the uvula and posterior pharynx, leading to asphyxiation. Swelling of the submucosa of the gastrointestinal tract can cause symptoms of an "acute abdomen," leading to unnecessary exploratory laparotomy. About half of patients report that trauma, particularly trauma associated with local pressure, precipitates an attack, and about half note an increased frequency of attacks during times of emotional stress. Attacks in patients with AAE are clinically similar to those in patients with HAE. In patients taking ACE inhibitors, angioedema may be manifested as severe swelling or simply as a chronic cough beginning days to months after ACE inhibitor therapy is initiated.

DIAGNOSIS

The best tests to support the diagnosis of HAE or AAE are measurements of C1 INH levels, C1 INH function, and C4 levels, particularly during an attack. The distinguishing features of AAE are onset later in life and the presence of a malignant disease or paraproteinemia. However, in addition to having low levels of C2 and C4, patients with AAE can have profound depressions in the level of C1, a protein that is commonly normal in HAE. Patients with ACE inhibitor–associated angioedema can present within hours after initiation of therapy or after many months and even years. The angioedema seen in urticaria/angioedema is distinctive in that it is usually associated with a pruritic urticarial rash, laboratory evaluation is normal, there is no history of treatment with an ACE inhibitor, and it responds to antihistamines, steroids, and epinephrine.

TREATMENT

Rx

Acute Attacks of Hereditary Angioedema

C1 INH concentrate purified from human plasma (Berinert; 20 units/kg intravenously), icatibant, a bradykinin receptor 2 antagonist (Firazyr; 30 mg SC), and ecallantide, a kallikrein inhibitor (Kalbitor; 30 mg SQ), are all FDA approved for acute attacks of HAE. If these agents are not available, treatment of angioedema of the airway should include racemic epinephrine (1:1000) delivered in the airway by nebulization and by intramuscular injections (0.2 to 0.3 mL of 1:1000 at intervals of 20 to 30 minutes). The addition of antihistamine for sedation may be helpful. Treating physicians must be prepared to perform nasotracheal intubation, preferably in the operating room under conditions in which tracheostomy can be performed if needed. Acute attacks can be terminated by administration of 2 units of fresh-frozen plasma (FFP) to supply the missing C1 INH; but in rare instances, patients may become more edematous, presumably reflecting the increased availability of substrates for the generation of kinins. Therefore, although FFP can be useful for treatment of non–life-threatening acute attacks, it is not recommended for life-threatening laryngeal edema.[13]

Long-term Treatment of Hereditary Angioedema

Attenuated androgens, such as danazol (50 to 200 mg up to twice daily), increase the production of C1 INH and lead to a marked amelioration of symptoms in patients with HAE. Masculinizing side effects are usually mild but can be problematic. These drugs are absolutely contraindicated in pregnancy. C1 INH concentrate (Cinryze; 1000 units IV every 3 to 4 days) is approved for long-term treatment. Many patients with relatively mild disease or infrequent attacks are treated with "on-demand" therapy with either C1 INH, icatibant, or ecallantide.[14]

Prophylaxis

Patients should be treated prophylactically before dental work or other procedures that involve trauma to tissue. Those treated with attenuated androgens, antifibrinolytic agents, FFP (2 units intravenously), or C1 INH concentrate (500 units subcutaneously) have fewer attacks.

Acquired Angioedema

Treatment of AAE is similar to that of HAE, but definitive treatment requires amelioration of the underlying disease.

ACE Inhibitor–Associated Angioedema

Treatment of angioedema associated with the use of an ACE inhibitor includes antihistamines, epinephrine, or both, as appropriate, and discontinuation of the ACE inhibitor. The direct renin inhibitor aliskiren is also associated with a significant risk of angioedema. Rare patients develop angioedema when taking angiotension receptor blockers. These patients are more likely to have idiopathic angioedema than angioedema due to the angiotension receptor blocker.

Evidence-Based Treatments

Current therapy for HAE in the United States includes both prophylactic treatment and on-demand, patient-centered treatment of attacks. Prophylactic administration of anabolic steroids and of C1 INH concentrate has been shown in a double-blind, placebo-controlled trial to significantly reduce the number of acute attacks. On-demand treatment for acute attacks has been shown to be effective with C1 INH, icatibant, and ecallantide. Ecallantide provides only marginal benefit for treating acute ACE inhibitor-induced angioedema in the emergency department setting.

PROGNOSIS

The long-term outlook for patients with HAE is largely dependent on the phenotype of the illness (frequency of laryngeal attacks), the ability of the patient to tolerate attenuated androgens, and the patient's access to C1 INH concentrate, icatibant, or ecallantide. Repeated use of these medications for recurrent acute episodes appears to be safe and effective.[15] For most patients, life expectancy should be normal. AAE usually resolves with treatment of the underlying condition, but the ultimate prognosis depends on the nature of that illness. Angioedema associated with the use of an ACE inhibitor can be fatal but usually resolves after the medication is removed.

FUTURE DIRECTIONS

In the last several years, there has been dramatic progress in the availability of medication for HAE.[16] In the near future, the focus will be on tailoring therapy to individual patients and controlling costs.

Grade A References

A1. Fedorowicz Z, van Zuren EJ, Hu N. Histamine H_2-receptor antagonists for urticaria. *Cochrane Database Syst Rev*. 2012;3:CD008596.

A2. Maurer M, Rosen K, Hsieh HJ, et al. Omalizumab for the treatment of chronic idiopathic or spontaneous urticaria. *N Engl J Med*. 2013;368:924-935.

A3. Zuraw BL, Busse PJ, White M, et al. Nanofiltered C1 inhibitor concentrate for treatment of hereditary angioedema. *N Engl J Med*. 2010;363:513-522.

A4. Riedl MA, Bernstein JA, Li H, et al. Recombinant human C1-esterase inhibitor relieves symptoms of hereditary angioedema attacks: phase 3, randomized, placebo-controlled trial. *Ann Allergy Asthma Immunol*. 2014;112:163-169.

A5. Cicardi M, Banerji A, Bracho F, et al. Icatibant, a new bradykinin-receptor antagonist, in hereditary angioedema. *N Engl J Med*. 2010;363:532-541.

A6. Lumry WR, Li HH, Levy RJ, et al. Randomized placebo-controlled trial of the bradykinin B_2 receptor antagonist icatibant for the treatment of acute attacks of hereditary angioedema: the FAST-3 trial. *Ann Allergy Asthma Immunol*. 2011;107:529-537.

A7. Cicardi M, Levy RJ, McNeil DL, et al. Ecallantide for the treatment of acute attacks in hereditary angioedema. *N Engl J Med*. 2010;363:523-531.

A8. Lewis LM, Graffeo C, Crosley P, et al. Ecallantide for the acute treatment of angiotensin-converting enzyme inhibitor-induced angioedema: a multicenter, randomized, controlled trial. *Ann Emerg Med*. 2014;65:204-213.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Zuberbier T. Chronic urticaria. *Curr Allergy Asthma Rep*. 2012;12:267-272.
2. Konstantinou GN, Asero R, Ferrer M, et al. EAACI taskforce position paper: evidence for autoimmune urticaria and proposal for defining diagnostic criteria. *Allergy*. 2013;68:27-36.
3. Lang DM. Evidence-based diagnosis and treatment of chronic urticaria/angioedema. *Allergy Asthma Proc*. 2014;35:10-16.
4. Sanchez-Borges M, Asero R, Ansotegui IJ, et al. Diagnosis and treatment of urticaria and angioedema: a worldwide perspective. *World Allergy Organ J*. 2012;5:125-147.
5. Tarbox JA, Gutta RC, Radojicic C, et al. Utility of routine laboratory testing in management of chronic urticaria/angioedema. *Ann Allergy Asthma Immunol*. 2011;107:239-243.
6. Zuberbier T. Pharmacological rationale for the treatment of chronic urticaria with second-generation non-sedating antihistamines at higher-than-standard doses. *J Eur Acad Dermatol Venereol*. 2012;26:9-18.
7. Bernstein JA, Lang DM, Khan DA, et al. The diagnosis and management of acute and chronic urticaria: 2014 update. *J Allergy Clin Immunol*. 2014;133:1270-1277.
8. Trojan TD, Khan DA. Calcineurin inhibitors in chronic urticaria. *Curr Opin Allergy Clin Immunol*. 2012;12:412-420.
9. Lang DM. A critical appraisal of omalizumab as a therapeutic option for chronic refractory urticaria/angioedema. *Ann Allergy Asthma Immunol*. 2014;112:276-279.
10. Altman KA, Naimi DR. Hereditary angioedema: a brief review of new developments. *Curr Med Res Opin*. 2014;30:923-930.
11. Vasekar M, Craig TJ. ACE inhibitor-induced angioedema. *Curr Allergy Asthma Rep*. 2012;12:72-78.
12. Caballero T, Baeza ML, Cabanas R, et al. Consensus statement on the diagnosis, management, and treatment of angioedema mediated by bradykinin. Part I. Classification, epidemiology, pathophysiology, genetics, clinical symptoms, and diagnosis. *J Investig Allergol Clin Immunol*. 2011;21:333-347.
13. Bork K. Current management options for hereditary angioedema. *Curr Allergy Asthma Rep*. 2012;12:273-280.
14. Betschel S, Badiou J, Binkley K, et al. Canadian hereditary angioedema guideline. *Allergy Asthma Clin Immunol*. 2014;10:50.
15. Malbrán A, Riedl M, Ritchie B, et al. Repeat treatment of acute hereditary angioedema attacks with open-label icatibant in the FAST-1 trial. *Clin Exp Immunol*. 2014;177:544-553.
16. Altman KA, Naimi DR. Hereditary angioedema: a brief review of new developments. *Curr Med Res Opin*. 2014;30:923-930.

REVIEW QUESTIONS

1. A 27-year-old woman wakes up with hives and appears in your office the same day for an urgent visit. On examination, she is not in acute distress but has multiple raised pruritic papules that blanch. The first medication you should prescribe is

A. First-generation type 1 antihistamine
B. Second-generation type 1 antihistamine
C. Type 2 antihistamine
D. Leukotriene pathway inhibitor
E. Corticosteroid

Answer: B Second-generation type 1 antihistamines (cetirizine, fexofenadine, and loratadine) are better tolerated than the first-generation type 1 antihistamines. Often, the first generation antihistamines (e.g., diphenhydramine) are prescribed for breakthrough symptoms.

2. A 45-year-old man has had hives daily for 8 weeks. A combination of antihistamines (types 1 and 2) at high doses plus prednisone 20 mg daily controls the hives. What is the next step?

A. Perform extensive testing to discover the offending food.
B. Continue prednisone after documenting a discussion of possible side effects.
C. Explain the importance of adherence to prescribed medications.
D. Discuss adding an immunomodulatory drug.
E. Initiate a work-up for systemic lupus erythematosus.

Answer: D This patient has failed to respond to conservative medical management. Chronic urticaria is a life-altering disease, and chronic steroid use is often associated with significant side effects. Current literature suggests that some immunomodulatory drugs, such as cyclosporine, can result in remission of the disease.

3. A 20-year-old man appears in your emergency department with acute abdominal pain. He has presented in this way in the past, without a definitive diagnosis. Examination now reveals a tender, swollen abdomen without bowel sounds. What is your next step?

A. Laparotomy
B. Appendectomy
C. Narcotics
D. Discharge secondary to malingering
E. Take a detailed personal and family history

Answer: E Attacks of abdominal angioedema in patients with hereditary angioedema can mimic an acute abdomen. Also, these patients often seek pain relief and are incorrectly labeled "drug seeking." A detailed personal and family history may reveal a pattern of swelling of the bowels or other parts of the body in either the patient or his immediate relatives. This will lead to appropriate laboratory studies and medical management.

4. A 50-year-old man carries the diagnosis of hereditary angioedema. He has been intubated twice and has minor episodes of swelling monthly. A trial of an attenuated androgen resulted in severe acne. Therapeutic options at this time include

A. C1 INH twice weekly
B. On-demand C1 inhibitor
C. On-demand icatibant
D. On-demand ecallantide
E. All of the above

Answer: E All of these options are possible. Patients now have multiple options for management of hereditary angioedema, and these should be instituted on the basis of the needs of each patient.

5. A 10-year-old girl presents with a history of a hive-like rash that occurs when she is exposed to cold weather. Further questioning reveals that this is sometimes associated with myalgias and arthralgias and that several first-degree relatives have similar complaints. What is the most likely diagnosis?

A. Chronic spontaneous urticaria
B. Cholinergic urticaria
C. Familial cold urticaria
D. Familial cold autoinflammatory syndrome
E. Systemic lupus erythematosus

Answer: D This uncommon disorder is often mistaken for familial cold urticaria, another unusual disease, and must be differentiated from common conditions such as chronic urticaria and systemic lupus erythematosus.

253

SYSTEMIC ANAPHYLAXIS, FOOD ALLERGY, AND INSECT STING ALLERGY

LAWRENCE B. SCHWARTZ

DEFINITION

Systemic anaphylaxis, a form of immediate hypersensitivity, arises when mast cells and possibly basophils are provoked to secrete mediators with potent vasoactive and smooth muscle contractile activities that evoke a systemic response. Although mast cells in any organ system may be involved, depending on the distribution of the instigating stimulus, the principal targets are the cardiovascular, cutaneous, respiratory, and gastrointestinal systems, sites where mast cells are most abundant. Systemic anaphylaxis can occur when these cells are activated by allergen that binds immunoglobulin E (IgE), or classic immediate hypersensitivity, and by alternative pathways.

EPIDEMIOLOGY

Assessments of the annual incidence of systemic anaphylaxis and the prevalence of those at risk for systemic anaphylaxis are compromised by imprecise diagnostic measures. Approximately 1500 to 2000 deaths in the United States per year are attributed to systemic anaphylaxis. The incidence of nonfatal cases has been estimated to be between 10 and 100 cases per 100,000 person-years. A random public telephone survey of 1000 adults conducted in 2011, designed to estimate the lifetime prevalence of systemic anaphylaxis, elicited a positive personal history in 7.7% of the participants, which was lowered to 1.6% using stringent criteria based on severity (hospitalization and "felt their life was in danger") and at least two organ systems being involved, including either respiratory or cardiovascular or both.[1] In children and adolescents,[2] because food allergy is more common, the incidence of anaphylaxis is likely to be higher. A separate survey of about 1000 at-risk patients also was conducted, focusing on subjects who had a history of some type of generalized allergic reaction, identifying systemic anaphylaxis in about one third. In both public and patient surveys, respiratory or cutaneous symptoms each occurred in more than 50%, whereas cardiovascular, neurologic, or gastrointestinal symptoms were recognized in less than 50%. Among patients, medications were the most common trigger, followed by insect stings, foods, environmental allergens, and latex, but the list of offending agents was lengthy and in some cases was unknown. About half of the reactions occurred at home, 14% at a hospital or clinic, and 6 to 7% at a family member's or friend's home, at work, or at a restaurant. Among adolescents, it has been estimated that one in four first-time reactions occurs outside the home, and therefore training of school and college staff also is essential.[2] Antibiotics and radiocontrast media are the most common triggers in hospitals. In the perioperative setting, systemic anaphylactic reactions occur with a frequency of about 1 in 3500, muscle relaxants being the most common, but antibiotics, latex, induction drugs, and other drugs can also be the culprit.[3]

Anaphylaxis to foods and insect stings each account for about 100 deaths per year. Most fatal anaphylactic reactions to injected venom proteins begin within 30 minutes after the sting. Most fatal food and insect sting reactions and many drug reactions are preceded by a mild immediate hypersensitivity reaction to the same allergen. Recognition of these earlier events as an important risk factor for future fatal anaphylaxis should lead to implementation of an action plan to prevent and deal with such reactions.

Food allergy is found in about 6% of children younger than 3 years of age and in half that percentage of adults, and these individuals are at risk for food-induced anaphylaxis. Most children lose their allergic sensitivities to cow's milk, egg, wheat, or soy by 5 years of age, whereas sensitivities to peanut, tree nuts, or seafood are typically long lasting. About 20% of children lose peanut sensitivity by school age, but a small portion of these regain peanut sensitivity later in life, particularly if they continue to avoid this food.

Latex provokes anaphylaxis in a small but significant group of individuals, particularly patients who have undergone multiple surgical procedures early in life, such as those with spina bifida or congenital urinary tract disorders, and those with frequent exposure later in life, such as medical personnel. Estimates of the prevalence of latex hypersensitivity range from 1 to 6% in the general population and about 10% among regularly exposed health care workers. Over a 5-year period, the U.S. Food and Drug Administration (FDA) collected approximately 1100 reports of latex-induced anaphylaxis, including 15 deaths. Elimination of powder latex gloves and availability of nonlatex gloves has diminished the prevalence of this problem. Contact hypersensitivity is diagnosed by patch testing, and immediate hypersensitivity by latex-specific IgE tests performed in vitro. Latex allergen skin test reagents have not yet received FDA approval.

PATHOBIOLOGY

Etiology

The mediators produced by activated mast cells and basophils initiate many of the signs and symptoms of anaphylaxis. These cells constitutively express the high-affinity receptor for IgE, FcεRI, on their cell surfaces. Consequently, these cells will always be armed with antigen-specific IgE, which is produced by sensitive individuals and enables cells to respond to antigens that aggregate IgE-FcεRI complexes on their surfaces. Therapeutic interventions aim to prevent the activation of these cells and to block the production or actions of their mediators. Cells other than mast cells and basophils also undoubtedly participate in systemic anaphylaxis, particularly those expressing FcεRI. Eosinophils, monocytes, antigen-presenting cells, and epithelial cells may be induced to express this receptor and thereby affect the intensity, duration, or character of anaphylactic reactions.

Most IgE-dependent mast cell activation events occur at local sites and result in local disease. For example, allergic conjunctivitis, allergic rhinitis, or allergic asthma typically occurs when allergen lands on the corresponding mucosal surface of a sensitive individual and diffuses into the tissue where mast cells reside. Systemic anaphylaxis presumably requires the allergen (or nonallergen agonist) to distribute systemically to activate mast cells at remote sites. This is more likely to occur when allergen is administered parenterally and is less likely after oral ingestion, inhalation, or cutaneous or ocular topical contact. Activation of mast cells in perivascular locations should have the greatest effect on systemic vascular responses. Additionally, the responsiveness of various organ systems to mast cell mediators may be influenced by local factors. Although mediators released at one tissue site could, in theory, spill into the circulation and affect remote sites, most vasoactive mediators are rapidly metabolized.

Allergens

The most common allergens causing systemic anaphylactic reactions include drugs, insect venoms, foods, radiocontrast media, allergen immunotherapy injections, and latex (Table 253-1). Most allergens are typically proteins or glycoproteins that serve as complete antigens, having at least two epitopes recognized by different IgE antibodies, and thereby capable of eliciting immediate hypersensitivity reactions in a sensitized subject without further processing. The protease activity of some allergens, such as house dust mite Der

TABLE 253-1 CAUSES OF SYSTEMIC ANAPHYLAXIS

IgE-MEDIATED	NON–IgE-MEDIATED
Insect stings	Aspirin
Foods	Radiocontrast media
Drugs	Exercise
Latex	Narcotics, vancomycin
Allergen extracts	Autoimmune
	Idiopathic

IgE = immunoglobulin E.

p1, may facilitate their penetration at mucosal sites. Others have lipid-binding domains, such as Der p2, that increase their antigenic potency. Anaphylactic reactions to a humanized monoclonal antibody, cetuximab, can occur on first exposure, owing to IgE against a nonhuman carbohydrate moiety, alpha-gal, which was made by the animal cells and conjugated to the recombinant humanized antibody being expressed. This IgE antibody seems to form against alpha-gal in tick secretions, also causing delayed (3 to 6 hours) anaphylactic reactions after ingestion of red meats, which contain alpha-gal.

In contrast to complete antigens, most drugs act as haptens. They become covalently linked to self-proteins in the circulation, in tissues or on cells, emerging as multivalent allergens. Multivalency is important for immediate hypersensitivity because cross-linking of at least two IgE molecules on the surface of cells aggregates FcεRI molecules, which then transmit an activating signal into the cell. Monovalent antigens fail to elicit mediator release because they bind IgE molecules without cross-linking them.

An allergen exposure must lead to sensitization before an immediate hypersensitivity reaction can occur. This process, which takes at least 1 week, involves antigen processing by antigen-presenting cells, which then present peptide antigens to T_H2 cells (helper T lymphocytes), which in turn select, nurture, and instruct allergen-specific B cells to switch from production of allergen-specific IgM or IgG to IgE. Production of interleukin-4 (IL-4) or IL-13 by T_H2 cells and binding of T_H2 CD40 ligand to B-cell CD40 are essential for this antibody class switch. Consequently, anaphylaxis does not typically occur on first exposure to an allergen (sensitization phase) because the antigen is likely gone by the time antigen-specific IgE is made, but it may occur after subsequent exposures.

Food

Most cases of *food-induced anaphylaxis* in children occur in response to egg, peanut, cow's milk, wheat, or soy, whereas peanuts, tree nuts, and seafood account for most reactions in adults. Reactions to seeds such as sesame seem to be growing in importance, and a variety of different foods have proved to be important allergens in specific individuals. Some patients have the *oral allergy syndrome*, which typically occurs in subjects sensitive to pollen allergens, their pollen-specific IgE cross-reacting to certain food allergens, such as ragweed with melon or birch with peach or apple. Also, the food epitopes involved are typically conformational (rather than linear), are therefore more easily destroyed by heating (cooking), by acid in the stomach, or by proteases in the intestines, and thus rarely progress to systemic reactions.

Food allergy-associated exercise-dependent anaphylaxis occurs when a sensitive subject exercises within several hours after eating the food to which they are sensitive, but not when eating the food without exercise. Shrimp and wheat are most commonly implicated. Exercise appears to increase intestinal permeability to food antigens, which then enter into the systemic circulation. Aspirin and nonsteroidal anti-inflammatory drugs (NSAIDs) also act to increase intestinal permeability. Avoiding the implicated food for 4 to 6 hours before exercise is recommended.

Insect Sting

Hymenoptera families primarily responsible for anaphylactic reactions include the Apidae (honey bees and bumble bees), Vespidae (hornets, yellow jackets, and paper wasps), and Formicidae (fire ants). Major allergens of honeybees include phospholipase A_2 (Api m 1), hyaluronidase (Api m 2), and melitin (Api m 4). Bumblebee venom proteins exhibit immunologic cross-reactivity with those of the honeybee, even though melitin is lacking. Vespid venoms cross-react among themselves and include proteins named antigen 5, phospholipase, and hyaluronidase, the latter allergen cross-reacting with bee hyaluronidase. Fire ant venom toxicity is caused principally by various alkaloids, which are not allergenic. Immediate hypersensitivity reactions to fire ant venom target a phospholipase that cross-reacts with the comparable vespid enzyme and various other proteins that are unique antigens. Allergens in fire ant venom cross-react with those in scorpion venom. A person may exhibit an anaphylactic reaction on first exposure to an insect's sting if previously sensitized to cross-reactive venom from a different insect. In contrast to stinging insects, allergens from biting insects of the Diptera order (mosquitoes, gnats, midges, true flies) are salivary in origin and do not cross-react with Hymenoptera venom allergens. Anaphylaxis to these salivary proteins appears to be uncommon, but precise epidemiologic data are problematic because people are often unaware of an ongoing mosquito bite, and commercial diagnostic reagents of high quality are not yet available.

Latex

Latex allergens are derived from the rubber tree, *Hevea brasiliensis.* Irritant dermatitis is the most frequent contact reaction and does not involve acquired immunity. Contact hypersensitivity, which results from cell-mediated immunity to haptenic chemicals added to latex during processing, produces a poison ivy–like local reaction that may appear the day after a sensitive subject is exposed. In contrast, immediate hypersensitivity occurs when IgE is made against proteins naturally found in this plant-derived product. Cutaneous (elastic materials), mucosal or intravascular (catheters), oral (balloon), and inhaled (powdered latex gloves) routes of exposure have been well documented and generally elicit signs and symptoms within minutes of exposure. IgE-mediated cross-reactivities between latex proteins and allergens in certain fresh foods such as banana, chestnut, avocado, kiwi, peach, bell pepper, and tomato have been reported and may necessitate avoidance of these foods.

Non–IgE-Dependent Agonists

Most foreign agents that trigger non–IgE-dependent anaphylaxis do not require antigen processing and can elicit a mast cell activation response on first exposure. These include radiocontrast dyes, narcotics such as codeine and morphine, and vancomycin (see Table 253-1). The dose and rate of administration and individual variations in sensitivity are determinants of severity. For radiocontrast dyes, media of low ionic strength and iso-osmolarity are less likely than those of high ionic strength and hyperosmolarity to elicit a systemic reaction. Vancomycin produces a non–IgE-dependent mast cell activation event known as *red man syndrome*, typically involving flushing and sometimes urticaria, but without cardiovascular compromise unless infused too rapidly; and these reactions usually can be avoided by reducing the rate of administration of the antibiotic, thereby reducing peak levels.

Endogenous mast cell activators include neuropeptides such as substance P, neurokinin A, calcitonin gene–related peptide, and the complement anaphylatoxins C3a and C5a. Whether a magnitude of mast cell activation sufficient to cause systemic anaphylaxis can result from endogenous secretion or generation of these peptides by themselves is unproved. For example, an anaphylactic shock–like syndrome occurred in hemodialysis patients exposed to a contaminated hemodialysis membrane that was associated with complement activation without detectable mast cell activation, and infusion of heparin contaminated with oversulfated chondroitin sulfate caused shock by activating the contact pathway and presumably generating bradykinin.

Aspirin and NSAIDs

Aspirin hypersensitivity typically manifests as either a respiratory or a cardiovascular reaction, although sometimes overlap is observed.[4] Respiratory reactions include bronchospasm, nasal congestion, and rhinorrhea and may extend beyond the respiratory tract to include abdominal cramping, watery diarrhea, and urticaria. Cardiovascular reactions that are identical clinically to allergen-induced systemic anaphylaxis and shock also can occur. In most cases, such reactions appear to be pharmacologically (not IgE) mediated, and in sensitive subjects they can occur in response to any of the cyclooxygenase 1 (COX1) inhibitors. Although cyclooxygenase inhibitors may shunt arachidonic acid metabolism to the lipoxygenase pathway, a mechanism to explain mast cell activation has not yet emerged. COX2-selective inhibitors appear to be relatively safe in aspirin-intolerant asthmatic patients but may still cause cardiovascular reactions in those who present with this manifestation. Less commonly, sensitivity occurs to only one of the drugs within this class and is caused by IgE against an associated unique chemical moiety on that particular drug. Human mast cells also express the low-affinity IgG receptor, FcγRIIa, which when aggregated by IgG immune complexes is capable of activating mast cells, and, at least in theory, may contribute to some episodes of anaphylaxis.[5]

Physical Stimuli

Physical stimuli may precipitate systemic anaphylaxis in certain individuals. Episodes can occur in response to exercise, heat, solar radiation, vibration, pressure, or cold. Exercise-dependent anaphylaxis is sometimes associated with ingestion of any food, regardless if sensitivity to the food can be documented, occurring within several hours of ingestion, and it might be avoided by delaying exercise until several hours after eating.

Autoimmunity and Activating Kit Mutations

Some patients experience spontaneous bouts of anaphylaxis without an obvious stimulus. Those with systemic mastocytosis (Chapter 255) are particularly prone to systemic anaphylaxis, perhaps because they have too many mast cells and because the mast cells they do have harbor a somatically acquired activating mutation of Kit tyrosine kinase that primes their activation status.[6] A corollary of this is that systemic anaphylaxis to an insect sting may be a presenting manifestation of mastocytosis, particularly if a baseline serum tryptase level is elevated (see later).[7] A related disorder, *mast cell activation syndrome*, includes patients with clonal mast cell disease, reflected by these same Kit mutations, who have recurrent bouts of anaphylaxis, but who either do or do not meet diagnostic criteria for systemic mastocytosis.[8] A cohort of patients, described in 2014, had elevated baseline serum tryptase levels, atopy and connective tissue disorders, and either spontaneous anaphylaxis or anaphylaxis triggered by heat, exercise, vibration, emotional stress, nonspecific foods, or minor physical trauma, inheriting this phenotype in an autosomal dominant pattern, but having no discernible mutations in their *c-Kit* gene, suggesting that mutation in other genes might increase the risk for anaphylaxis.

Some cases of chronic urticaria are known to be associated with IgG and IgM antibodies against FcεRI or IgE.[9] In such cases, complement activation leading to the generation of complement anaphylatoxins at the surface of mast cells has been postulated to synergize with FcεRI-mediated activation. These reactions may occur preferentially in the skin because of the expression of anaphylatoxin receptors on the type of mast cell that predominates in the skin but not on the type that predominates in lung. An analogous, albeit speculative, autoimmune process might activate mast cells localized in blood vessel walls, the result being anaphylaxis.

Autoimmune progesterone-mediated anaphylaxis, *catamenial anaphylaxis*, which tends to occur just before menses, is uncommon but well documented and may respond to medical or surgical interventions that prevent menses.[10]

Pathophysiology

Mast cells participate in both acquired and innate forms of immunity (Chapter 255). They develop in peripheral tissues from bone marrow progenitors, primarily under the influence of stem cell factor, the ligand for the tyrosine kinase receptor called Kit. Armed with allergen-specific IgE, mast cells are activated by multivalent allergens that cross-link IgE and aggregate FcεRI molecules on the cell surface. This may be important in the defense against certain parasites that elicit a strong IgE response. Experiments performed in rodents suggest that mast cells can be directly activated by microbial products, leading to the secretion of mediators that recruit neutrophils. This innate immune response may restrain bacterial dissemination until a more potent acquired immune response develops. Activation of mast cells by endogenous peptides such as substance P or calcitonin gene–related peptide may influence basic biologic processes such as wound healing and angiogenesis. Whether human mast cells have a critical, nonredundant role in these biologic and immunologic processes remains controversial. However, their central role in immediate hypersensitivity is clear.

Mediators released by mast cells include preformed mediators stored in secretory granules, some of which are preferentially or exclusively made by mast cells, newly generated lipid products, which are not precise biomarkers for mast cells and include metabolites of arachidonic acid, and an array of cytokines and chemokines. Histamine, formed from histidine by histidine decarboxylase, is the sole biogenic amine stored in all granules of human mast cells and basophils. Histamine released by mast cells or basophils diffuses freely and interacts with H_1, H_2, H_3, and H_4 receptors. H_1 receptors are found on endothelial cells, smooth muscle cells, and sensory nerves; when stimulated, bronchial and gastrointestinal smooth muscle contraction, vascular smooth muscle relaxation, increased permeability of postcapillary venules, coronary artery vasoconstriction, and pruritus can occur—signs and symptoms often associated with systemic anaphylaxis. In the central nervous system (CNS), blockade of H_1 receptors appears to cause drowsiness. H_2 receptors reside on gastric parietal cells and at lower levels on inflammatory cells, bronchial epithelium, and endothelium and in the CNS. H_2-receptor–mediated increased acid production in the stomach may occur transiently during systemic anaphylaxis, but it is more likely to become clinically significant if histamine levels are chronically elevated, as occurs with systemic mastocytosis. H_3 receptors are found primarily on cells in the CNS. H_4 receptors are found on hematopoietic cells, such as mast cells, basophils, eosinophils and lymphocytes, and may modulate certain aspects of inflammation, such as eosinophil recruitment, as well as pruritus. Histamine, after its secretion from mast cells and basophils, is rapidly metabolized to inactive methylhistamine and methylimidazole acetic acid.

Prostaglandin D_2 (PGD_2) is the principal COX-catalyzed product of arachidonic acid secreted by activated mast cells, but it is not made by basophils. It binds to the G protein–coupled receptors, CRTH2 and DP. Both COX1 and COX2 are involved in PGD_2 production by mast cells. Consequently, a COX inhibitor that is bipotent might be better than one that is selective at blocking PGD_2-mediated responses during anaphylaxis, which may include hypotension, bronchospasm, inhibition of platelet aggregation, and prolonged asymptomatic cutaneous erythema.

Leukotriene C_4 (LTC_4) is released by both mast cells and basophils after its formation from arachidonic acid and glutathione; its formation is sequentially catalyzed first by 5-lipoxygenase together with 5-lipoxygenase–activating protein and then by LTC synthase. Conversion to LTD_4 and LTE_4, which also are bioactive, occurs in the extracellular space. These sulfidopeptide leukotrienes bind to the G protein–coupled receptors cysteinyl leukotriene 1 ($CysLT_1$), on bronchial smooth muscle, epithelial and endothelial cells, and leukocytes, and $CysLT_2$, on vascular smooth muscle, endothelial and epithelial cells, leukocytes, and heart muscle. LTE_4 may have selective affinity for the P2Y receptor, gpr99. Sulfidopeptide leukotrienes cause bronchoconstriction, mucus secretion, eosinophil recruitment, vasopermeability, diminished cardiac contractility, vasoconstriction of coronary and peripheral arteries, vasodilation of venules, and a burning cutaneous wheal-and-flare response. Antagonists of $CysLT_1$ (montelukast, zafirlukast), but not of $CysLT_2$, as well as a 5-lipoxygenase inhibitor (zileuton), are currently available to patients.

Platelet-activating factor (PAF) is generated from 2-lyso-glycero-3-phosphorylcholine when an acetyl group is placed on the sn-2 carbon of glycerol by acetyltransferase. PAF activates platelets but also is a potent vasoactive mediator that enhances vasodilation and vasopermeability and a smooth muscle constrictor that is capable of inducing bronchospasm. PAF is generated by both mast cells and basophils, as well as by other cell types. Elevated levels of PAF and low levels of circulating PAF acetylhydrolase, which converts PAF back to inactive lyso-PAF by removing the sn-2 acetyl moiety, have been associated with more severe food-induced systemic anaphylaxis. Sphingosine-1-phosphate (S1P) is generated from sphingosine by sphingosine kinase in activated mast cells, as well as in other cell types, and may enhance the vascular response during systemic anaphylaxis.

Mast cells also are the sole or principal source of heparin proteoglycan and certain proteases. All express β-tryptase, and a subset also expresses chymase, mast cell carboxypeptidase, and cathepsin G (like neutrophils and monocytes). Mast cells that express only tryptase are called MC_T cells; those that also express the other proteases are called MC_{TC} cells. Mature tryptase is stored in the secretory granules of all mast cells and is released during degranulation of activated cells; acute levels in serum serve as a clinical biomarker for mast cell activation. In contrast, precursor forms of tryptase (protryptase) are spontaneously secreted by mast cells at rest; baseline levels in serum reflect the total body burden of mast cells, serve as a minor diagnostic criterion for systemic mastocytosis, and serve as a biomarker for anaphylactic risk in allergen-sensitized subjects. MC_{TC} but not MC_T cells express CD88, the C5a receptor, and therefore are activated by complement C5a. Basophils are relatively deficient in these proteases but likewise express CD88.

Cytokines (tumor necrosis factor-α [TNF-α]; IL-4, -5, -6, -8, -13, and 16; granulocyte-macrophage colony-stimulating factor [GM-CSF]; basic fibroblast growth factor [bFGF]; vascular endothelial growth factor [VEGF]) and chemokines (IL-8, monocyte chemotactic protein-1, monocyte inflammatory protein-1α) represent another dimension of the mediators released by mast cells and basophils. Although these mediators are not selectively produced by these cell types, their vasoactive and inflammatory potential could affect the severity and duration of anaphylaxis. As selective antagonists of the relevant cytokines and chemokines become available and are tested for therapeutic benefits, the roles of these mediators in the pathogenesis of anaphylaxis will be better understood.

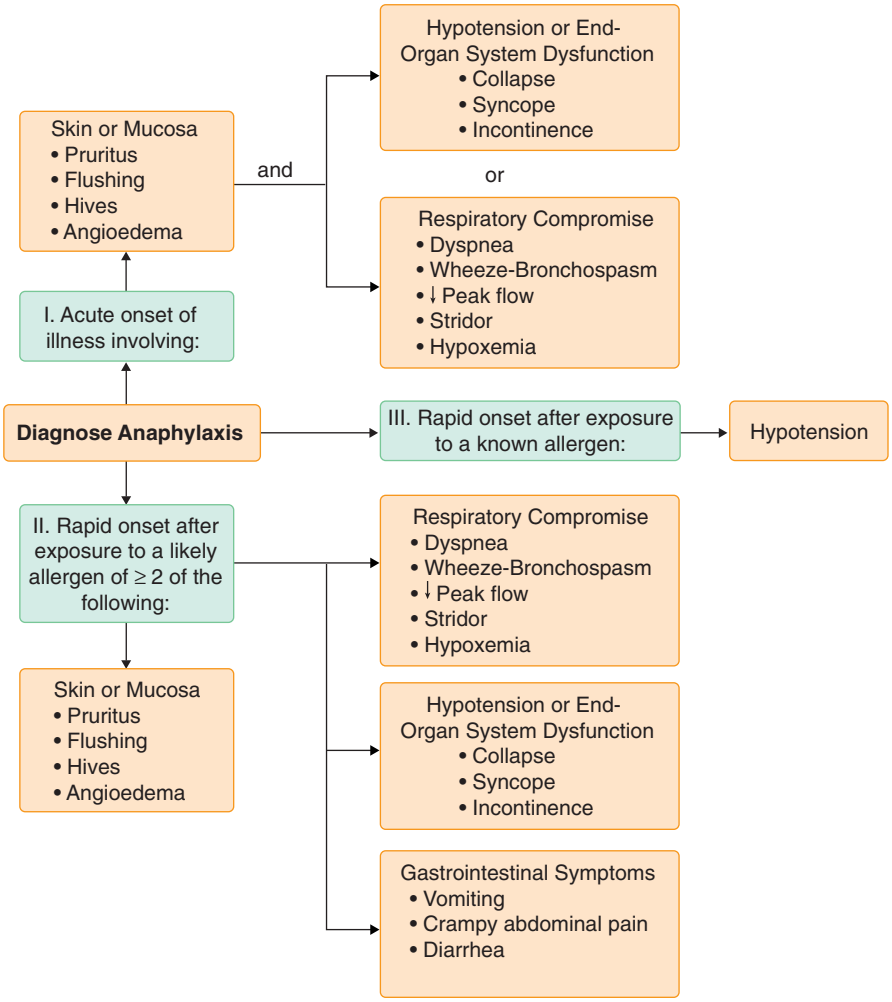

FIGURE 253-1. Clinical diagnosis of systemic anaphylaxis. Acute onset of systemic anaphylaxis in the apparent absence of allergen exposure means that the signs and symptoms develop over minutes to several hours, while rapid onset after exposure to a likely or known allergen means that these signs and symptoms begin to occur within minutes to several hours after that exposure. This is based on the Second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network Symposium (From Sampson HA, Muñoz-Furlong A, Campbell RL, et al. Second symposium on the definition and management of anaphylaxis: summary report—second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium. *J Allergy Clin Immunol*. 2006;117:391-397).

DIAGNOSIS

Systemic anaphylaxis can be diagnosed clinically in real time by consensus criteria outlined in Figure 253-1.[11-13] Acute onset of cutaneous signs of immediate hypersensitivity along with either hypotension or respiratory compromise in the apparent absence of allergen exposure, rapid onset of hypersensitivity signs involving at least two organs from among cutaneous, gastrointestinal, respiratory and cardiovascular systems after exposure to a likely allergen, or rapid onset of hypotension after exposure to a known allergen can be used to diagnose systemic anaphylaxis—which sometimes can be precisely confirmed in the laboratory by demonstration of antigen-specific IgE (sensitization) and an acute serum level of total (mature plus pro) tryptase (mast cell activation) that is greater than a baseline level obtained at least 24 hours after all signs and symptoms have resolved. Skin testing or in vitro measurements of antigen-specific IgE should be delayed for at least 2 weeks after the precipitating event to prevent false-negative results. Insect venom allergies also have been assessed by experimental sting challenges, but these are not recommended for routine evaluations. For food allergies, larger wheal-and-flare responses to prick skin tests and higher IgE titers to specific allergens are associated with more severe reactions. Oral food challenges are performed under certain circumstances, taking care to minimize the risk for systemic anaphylaxis. These food-allergic reactions involve IgE sensitization and IgE-dependent mechanisms and should be distinguished from a variety of other types of adverse food reactions, including lactose intolerance due to a deficiency in lactase, food-induced enterocolitis in infants (in reaction to cow's milk, soy, or grains), and celiac disease associated with ingestion of gluten in wheat and other grains.

An increased level of total tryptase in acute (over baseline) serum, which peaks 15 to 60 minutes after the onset of the signs or symptoms of anaphylaxis and then declines with a half-life of about 2 hours (normal baseline levels ranging from 1 to 11 ng/mL), indicates that mast cell activation has occurred. During a study of experimental insect sting–induced anaphylaxis, the increased level of tryptase correlated closely with the drop in mean arterial pressure, indicating that the magnitude of mast cell activation is a primary determinant of clinical severity. Although an increased serum total tryptase level during putative systemic anaphylaxis may be useful for distinguishing anaphylaxis from other conditions in the differential diagnosis, elevations may not be detected after anaphylaxis triggered by food ingestion, or, in general, if anaphylactic severity is either modest (no hypotension) or local (laryngeal edema), or if the acute sample was collected outside of the optimal time. Whether there are anaphylactic IgE-dependent pathways that do not require mast cell activation, but instead involve basophil activation, is unknown but has been considered for anaphylaxis triggered by food allergen ingestion. Plasma histamine, because it is rapidly metabolized, is not as practical as serum or plasma tryptase for detecting anaphylaxis. However, urinary histamine or methylhistamine levels also may reflect overall levels of released histamine, accumulating in urine during anaphylaxis and stored in the bladder until micturition; but levels are affected by ingested histamine-containing foods, histamine-producing mucosal bacteria, and variability in histamine metabolism. PGD_2, which is made by several cell types, including activated mast cells, is rapidly metabolized to $PGF_{2\alpha}$, and urinary levels of this metabolite should be elevated in urine formed during anaphylaxis.

An elevated baseline level of serum total tryptase also appears to be a risk factor for increasing severity of insect venom-mediated systemic anaphylaxis,

perhaps in part because of an underlying clonal mast cell disorder with an activating c-kit mutation. Future studies will determine whether baseline serum tryptase levels should guide therapy for venom-sensitive subjects as well as those with other allergic sensitivities. Other risk factors for severe systemic anaphylaxis include a prior allergic event to that allergen and having high blood pressure, particularly if being treated with nonspecific β-blockers or angiotensin-converting enzyme (ACE) inhibitors.

Low serum levels of PAF acetyl hydrolase, which metabolizes PAF, and of ACE, which metabolizes bradykinin, have been associated with more severe food-induced systemic anaphylaxis. Whether slow metabolism of PAF and bradykinin might allow these mediators to play a role in such reactions and whether mediator-specific therapies would be clinically useful in such reactions remain to be determined.

Differential Diagnosis

Anaphylaxis should be distinguished from a variety of disorders with overlapping presentations. Vasovagal syncope causes diaphoresis, nausea, hypotension, and bradycardia, but without urticaria and tachycardia. Flushing disorders may be benign and unrelated to anaphylaxis, or they could be a manifestation of pathologic conditions such as the carcinoid syndrome (Chapter 232), with which urticaria and profound hypotension are not typically associated, or pheochromocytoma (Chapter 228), which causes episodic hypertension. Precise detection of these latter conditions is beyond the scope of this chapter. Panic attacks and vocal cord dysfunction can be a challenge to distinguish from anaphylaxis, especially by history alone, but nevertheless must be considered. Acute attacks of hereditary and acquired angioedema (Chapter 252) caused by C1 esterase inhibitor deficiency are not associated with pruritus or urticaria and persist longer than attacks of anaphylaxis. Shock due to complement activation by contaminated hemodialysis tubing, leading to the generation of C3a and C5a anaphylatoxins, or to activation of the contact system by an oversulfated chondroitin sulfate contaminant in heparin preparations, leading to the production of bradykinin, can occur without involving mast cell activation. Scombroidosis (histamine fish poisoning) occurs 5 to 90 minutes after ingestion of histamine, typically in poorly stored fish, and manifests with flushing, palpitations, headache, and gastrointestinal symptoms. The condition lasts several hours, both duration and severity depending on the amount of histamine ingested, and usually responds to H_1-receptor and H_2-receptor antihistamines, but occasionally requires epinephrine and intravenous fluids. Acute serum sickness, various cell activation syndromes, endotoxin-mediated septic shock, and superantigen-mediated toxic shock syndromes manifest with fever, which is not characteristic of anaphylaxis by itself. Also, hypoglycemia, seizure, and primary pulmonary or cardiac events should be considered.

In some cases, systemic anaphylaxis occurs together with another disorder. For example, a 65-year-old man, after being stung by a wasp, complained of dizziness and shortness of breath, was hypotensive with urticaria, and responded to treatment with intramuscular epinephrine, but also complained of chest pressure; electrocardiography indicated an inferior wall infarction. Acute serum levels of both tryptase and cardiac enzymes were elevated, indicating that both anaphylaxis and myocardial infarction had occurred.

Systemic mastocytosis (Chapter 255) is an important condition to consider in the setting of anaphylaxis. In adults, a somatic activating mutation in the gene for Kit in mast cell progenitors results in an excessive body burden of mast cells. The disease may regress spontaneously in children with this disorder, but this is uncommon in adults. Patients with too many mast cells are at increased risk for anaphylaxis, and anaphylaxis may be a presenting manifestation of systemic mastocytosis. For example, anaphylaxis in response to an insect sting, particularly in the absence of venom-specific IgE (due to direct mast cell agonists), should raise the possibility of systemic mastocytosis. Diagnostic tests for systemic mastocytosis might include a lesional skin biopsy when urticaria pigmentosa is suspected and a bone marrow biopsy, each stained for mast cells by antitryptase and anti-Kit (CD117) immunohistochemistry. Findings of mast cell granulomas (major criterion) and numerous spindle-shaped mast cells (minor criterion) in the bone marrow biopsy, along with mast cells that express CD2 or CD25, a D816V Kit mutation, and an elevated baseline level of serum total tryptase (>20 ng/mL) during a clinically quiescent interval (minor criteria) are used to diagnose systemic mastocytosis. One major and one minor, or three minor, criteria are recommended by the World Health Organization for diagnosis. Mast cell activation syndrome can be diagnosed when spontaneous anaphylaxis occurs in association with only two of these minor criteria.

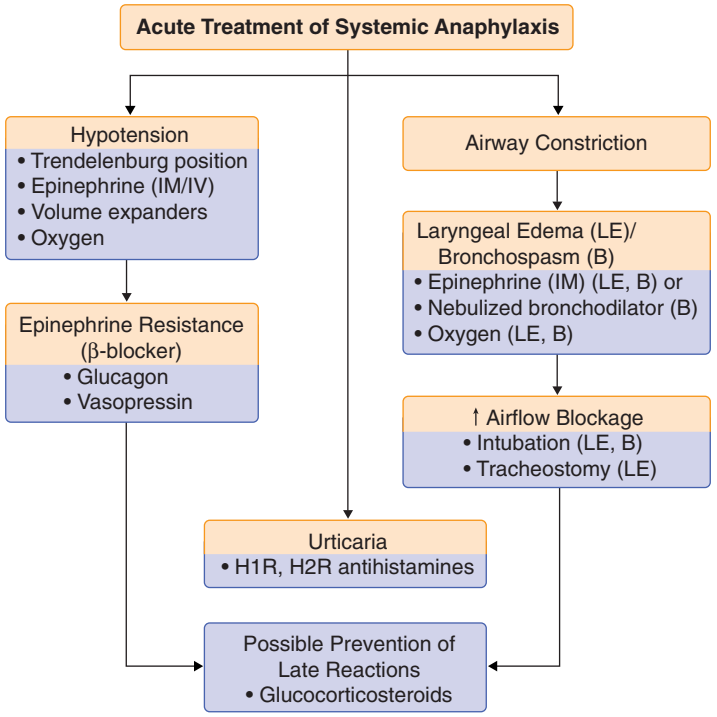

FIGURE 253-2. Acute treatment of systemic anaphylaxis. B = bronchospasm; H1R = histamine-1 receptor; H2R = histamine-2 receptor; IM = intramuscular; LE = laryngeal edema.

TREATMENT

Rx

Fatal outcomes in anaphylaxis are principally the result of either airway constriction or hypotension. Accordingly, the acute treatment of systemic anaphylaxis requires that airway patency, blood pressure, and cardiac status be addressed (Fig. 253-2).[14] Epinephrine, intubation, tracheostomy, volume expanders, and vasopressors may be needed. Patients exhibiting any signs or symptoms of hypotension should immediately assume the Trendelenburg position, which may prevent progression to anaphylactic shock or what has been called in postmortem examinations the *empty ventricle syndrome*—because almost all hypotensive anaphylactic deaths are preceded by syncope occurring in a sitting or upright posture. Epinephrine injected intramuscularly into the thigh (0.2 to 0.5 mg for adults, 0.01 mg/kg up to 0.3 mg for children, repeated every 5 to 30 minutes as indicated) is the most critical drug to administer, the earlier during the course of an anaphylactic event the better. Alternatively, intravenous administration of a solution of epinephrine (1 mg/100 mL solution starting at 30 to 100 mL/hour) and titrated to the lowest effective rate of infusion can be considered. Epinephrine relaxes bronchial smooth muscle and improves vasomotor tone and vasopermeability, thereby counteracting bronchospasm, hypotension, and tissue edema. However, the benefits of epinephrine need to be weighed against its disadvantages in elderly subjects and in those with cerebrovascular or coronary artery disease, hypertension, diabetes, hyperthyroidism, cardiomyopathy, or narrow-angle glaucoma, in whom adverse events such as myocardial infarction, stroke, or pulmonary edema can be precipitated. Also, patients taking a β-blocker may be resistant to epinephrine; in such a case, glucagon (1 to 5 mg/hour intravenously [IV] in adults or 20-30 mcg/kg in children, in each case administered over 5 min or vasopressin (2 to 40 IU IV in adults) may be used. Oxygen should be administered by nasal cannula. Inhaled bronchodilators can relieve bronchospasm. Parenteral administration of H_1-receptor (diphenhydramine, 1 to 2 mg/kg up to 50 mg) and H_2-receptor (ranitidine, 50 mg in adults and 1 mg/kg in children, in each case administered IV over 5 minutes) antihistamines may prevent progression of some of the signs and symptoms, particularly urticaria and pruritus, but is not likely to reverse hypotension or tissue edema. Prednisone (20 mg orally) or Solu-Medrol (40 mg IV) may reduce the risk for a protracted reaction or the late phase of biphasic anaphylaxis but is unlikely to be of benefit acutely.

PREVENTION

Patients who have experienced an anaphylactic reaction are at greatest risk for another episode. Such individuals should wear a Medic-Alert bracelet and be instructed in the use of epinephrine (e.g., EpiPen), which they should carry. Avoidance of nonspecific β-blockers and ACE inhibitors is

recommended because either may worsen the severity of an anaphylactic episode, and β-blockers might interfere with β-agonist treatment. In subjects with recurrent anaphylaxis, prophylactic use of H_1- and H_2-receptor antihistamines is beneficial. A leukotriene antagonist and cyclooxygenase inhibitor theoretically would provide additional benefit, but this has not been well studied. Finally, cyclosporine (3 to 5mg/kg per day) might be considered in difficult cases of recurrent anaphylaxis because of its ability to inhibit mast cell activation in vitro and in vivo in chronic urticaria. Omalizumab neutralizes free IgE, and anecdotal reports show benefit in controlling both urticaria and spontaneous episodes of anaphylaxis in mastocytosis patients, but it is not currently approved by the FDA for this indication. Glucocorticosteroids do not inhibit mast cell activation in vitro or immediate skin test responses to allergens in vivo but nevertheless may be beneficial in selected patients with recurrent anaphylaxis.

Specific anaphylactic syndromes have unique considerations. Anti-IgE therapy in peanut-allergic subjects can increase the threshold of sensitivity, on average, from the equivalent of half a peanut to almost nine peanuts.[A1] Insect venom sensitivity can be selectively treated by venom immunotherapy,[A2] dramatically decreasing the risk for anaphylaxis in response to future stings.[15] Reactions to radiocontrast media can be prevented or attenuated by prior administration of prednisone and H_1- and H_2-receptor antihistamines. Patients who are hypersensitive to penicillin should avoid β-lactam antibiotics in general but can be desensitized if an antibiotic in this class is critically needed (e.g., penicillin for neurosyphilis). However, desensitization is temporary; after the drug has cleared, sensitivity is likely to return. Catamenial anaphylaxis may respond to the luteinizing hormone–releasing hormone analog, Lupron, to oophorectomy, or to conjugated estrogens. Patients with systemic mastocytosis, in addition to prophylactic pharmacologic measures, should avoid using direct mast cell agonists such as codeine, morphine, and vancomycin. Aspirin-intolerant subjects can be desensitized but then must continue to ingest a daily dose of aspirin to maintain their desensitization status, thereby benefiting from better asthma control and regression of any nasal polyps that might be present. Food- and latex-sensitive subjects must practice avoidance of the provocative agent, although preliminary data with anti-IgE neutralization therapy and oral immunotherapy indicate that these measures might provide some protection against small inadvertent food allergen exposures. Future research should yield more effective and long-lasting interventions that reduce anaphylactic risk (including better desensitization regimens) and that more effectively reverse the signs and symptoms of this potentially fatal disorder.

Grade A References

A1. Anagnostou K, Islam S, King Y, et al. Assessing the efficacy of oral immunotherapy for the desensitisation of peanut allergy in children (STOP II): a phase 2 randomised controlled trial. *Lancet*. 2014;383:1297-1304.

A2. Boyle RJ, Elremeli M, Hockenhull J, et al. Venom immunotherapy for preventing allergic reactions to insect stings. *Cochrane Database Syst Rev*. 2012;10:CD008838.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Wood RA, Camargo CA Jr, Lieberman P, et al. Anaphylaxis in America: the prevalence and characteristics of anaphylaxis in the United States. *J Allergy Clin Immunol*. 2014;133:461-467.
2. Gupta RS. Anaphylaxis in the young adult population. *Am J Med*. 2014;127(1 suppl):S17-S24.
3. Mertes PM, Tajima K, Regnier-Kimmoun MA, et al. Perioperative anaphylaxis. *Med Clin North Am*. 2010;94:761-789.
4. Kowalski ML, Stevenson DD. Classification of reactions to nonsteroidal antiinflammatory drugs. *Immunol Allergy Clin North Am*. 2013;33:135-145.
5. Jonsson F, Mancardi DA, Zhao W, et al. Human FcgammaRIIA induces anaphylactic and allergic reactions. *Blood*. 2012;119:2533-2544.
6. Soucie E, Brenet F, Dubreuil P. Molecular basis of mast cell disease. *Mol Immunol*. 2015;63:55-60.
7. Alvarez-Twose I, Bonadonna P, Matito A, et al. Systemic mastocytosis as a risk factor for severe Hymenoptera sting-induced anaphylaxis. *J Allergy Clin Immunol*. 2013;131:614-615.
8. Valent P, Akin C, Arock M, et al. Definitions, criteria and global classification of mast cell disorders with special reference to mast cell activation syndromes: a consensus proposal. *Int Arch Allergy Immunol*. 2012;157:215-225.
9. Viegas LP, Ferreira MB, Kaplan AP. The maddening itch: an approach to chronic urticaria. *J Investig Allergol Clin Immunol*. 2014;24:1-5.
10. Bauer CS, Kampitak T, Messieh ML, et al. Heterogeneity in presentation and treatment of catamenial anaphylaxis. *Ann Allergy Asthma Immunol*. 2013;111:107-111.
11. Lieberman P, Nicklas RA, Oppenheimer J, et al. The diagnosis and management of anaphylaxis practice parameter: 2010 update. *J Allergy Clin Immunol*. 2010;126:477-480.
12. Simons FE, Ardusso LR, Dimov V, et al. World Allergy Organization anaphylaxis guidelines: 2013 update of the evidence base. *Int Arch Allergy Immunol*. 2013;162:193-204.
13. Lieberman PL. Recognition and first-line treatment of anaphylaxis. *Am J Med*. 2014; 127(1 suppl):S6-S11.
14. Casale TB, Burks AW. Hymenoptera-sting hypersensitivity. *N Engl J Med*. 2014;370:1432-1439.
15. Golden DBK. Advances in diagnosis and management of insect sting allergy. *Ann Allergy Asthma Immunol*. 2013;111:84-89.

REVIEW QUESTIONS

1. In adults, which of the following triggers of anaphylaxis accounts for the most deaths?

A. Peanuts
B. Latex
C. Penicillin
D. Hymenoptera venom
E. Aspirin and nonsteroidal anti-inflammatory drugs (NSAIDs)

Answer: D The most common cause of anaphylactic death reported in adults is due to stings from insects, whereas in children, food allergic reactions may be more common. Adults who have experienced anaphylaxis to an insect sting in the past should be evaluated by an allergist for current sensitivity and risk factors for severe anaphylaxis and encouraged to undergo venom immunotherapy as clinically appropriate, which can reduce the risk for severe anaphylaxis to future stings by more than 95%. The prevalence of latex allergy in general and latex-triggered anaphylaxis in particular seem to be declining now that powdered latex gloves have been removed from medical centers and latex-free products are available for almost any procedure. Medical centers and patients generally do a good job of avoiding penicillin exposure in penicillin-allergic patients, and alternatives to penicillins are available for most infections. Cardiovascular reactions associated with aspirin and NSAIDs that can mimic allergen-induced systemic anaphylaxis are, in most cases, pharmacologically (not immunoglobulin E [IgE]) mediated.

2. An anaphylactic reaction in a 45 year-old living in Virginia that occurs 3 to 6 hours after eating dinner is most likely due to which one of the following allergens?

A. Bovine serum albumin in beef
B. Tropomyosin protein in shrimp
C. Alpha-gal carbohydrate in veal
D. Penicillin in pork
E. Melon in salad

Answer: C Particularly in the Southeastern and Mid-Atlantic states, perhaps due to exposure and sensitization to alpha-gal during Lone Star tick bites, IgE antibodies form against this antigen. Sensitized individuals can exhibit delayed urticarial or anaphylactic responses 3 to 6 hours after eating red meats, presumably because they do not react to monovalent alpha-gal-conjugated proteins in the meat soon after ingestion, but instead require several hours for this alpha-gal to be processed and form polyvalent, haptenized antigens. Interestingly, skin testing to a beef allergen extract is typically positive with sensitivity to protein allergens but negative with sensitivity to only alpha-gal, whereas in vitro testing of serum for allergen-specific IgE is positive in both cases. Anaphylactic reactions to protein allergens such as bovine serum albumin or tropomyosin occur acutely, generally within an hour of ingestion. Penicillin in beef or pork from penicillin-fed animals can cause anaphylaxis in penicillin-sensitive patients, but is uncommon and does so in less than an hour after ingestion. Individuals with pollen allergies may have pollen-specific IgE cross-reacting to certain food allergens, such as ragweed with melon, typically causing immediate-onset pruritus, irritation, and swelling around the mouth ("oral allergy syndrome").

3. Which of the following drugs is likely to activate mast cells on first exposure?

A. Vancomycin
B. Penicillin
C. Phenytoin
D. Bactrim

Answer: A Vancomycin directly activates mast cells on first and subsequent exposures, without involving the IgE-FcεRI pathway, but instead activating a G protein–coupled receptor. When peak levels of vancomycin are modest, red man syndrome results; hypotensive anaphylaxis can occur if the drug is infused too rapidly. Penicillin, phenytoin, and Bactrim (sulfamethoxazole and trimethoprim) can each haptenize proteins, elicit an IgE antibody response during a sensitization period, and then cause anaphylaxis with a subsequent exposure.

4. Mastocytosis, a clonal mast cell disorder that increases the risk for anaphylaxis, should be suspected if which of the following is elevated in serum taken from a patient at a time when their medical problem is clinically quiescent?

A. Bradykinin
B. Tryptase
C. Serotonin
D. Complement C5a
E. Catecholamines and metanephrines

Answer: B Serotonin is not abundantly produced by mast cells. Bradykinin, generated extracellularly from kininogen precursors, also is not generated by mastocytosis patients during nonacute time periods. C5a is generated from C5 by proteases, typically by C5 convertases generated during activation of the complement pathways, and also by certain noncomplement pathway proteases, but is not generated during times when mastocytosis is clinically quiescent. Flushing syndromes like pheochromocytoma, which produces catecholamines and metanephrines, are only in the differential diagnosis to be distinguished from anaphylaxis. On the other hand, α- and β-protryptases are spontaneously released by mast cells at rest. Consequently, tryptase levels in serum, detecting both mature and protryptases and reflecting the increased mast cell burden and abnormal mast cell phenotype of mastocytosis mast cells, are typically elevated during nonacute periods of disease.

5. In the absence of exposure to a known allergen, the diagnosis and treatment of anaphylaxis should be most strongly considered in a patient who presents with the acute onset of which of the following signs or symptoms?

A. Vomiting and diarrhea
B. Abdominal pain
C. Flushing and diarrhea
D. Hypotension
E. Urticaria and wheezing

Answer: E Acute onset of cutaneous signs of immediate hypersensitivity along with either hypotension or respiratory compromise in the apparent absence of allergen exposure; rapid onset of hypersensitivity signs involving at least two organs from among cutaneous, gastrointestinal, respiratory, and cardiovascular systems after exposure to a likely allergen; or rapid onset of hypotension after exposure to a known allergen can be used to diagnose systemic anaphylaxis in real time. In the scenario presented, lacking exposure to a known allergen, answers A and B only involve one organ system, C does not exhibit hypotension or respiratory signs, and D lacks cutaneous signs.

254 DRUG ALLERGY

LESLIE C. GRAMMER

DEFINITION

Adverse drug reactions (ADRs) are recognized as an important public health problem as they result in both morbidity and mortality. An ADR is defined by the World Health Organization as an unintended, noxious response to a drug that occurs at a dose usually prescribed for human patients. The classic pharmacologic definition of ADRs by Rawlins and Thompson separates these into two major types: type A reactions, which are predictable and dose dependent; and type B reactions, which are unpredictable and not dose dependent. Type B reactions account for 10 to 25% of all ADRs and include drug allergy. The World Health Organization Nomenclature Review Committee defines *drug allergy* as a hypersensitivity reaction for which a definite immunologic mechanism, either a B-cell–mediated (antibody) or a T-cell–mediated process, is documented. Most published epidemiologic studies refer to ADRs in general and not to drug allergy specifically because the demonstration of drug-specific B-cell–mediated or T-cell–mediated mechanisms is often difficult, and the immunologic culprit may be a drug metabolite.

EPIDEMIOLOGY

Drug allergy is responsible for significant mortality, morbidity, and socioeconomic costs that are probably underestimated. Current data must be evaluated carefully because they involve different populations, different definitions of ADRs and drug allergy, and different methodologies, especially in terms of data analysis. The Boston Collaborative Drug Surveillance Program collected information on all ADRs in 4031 hospitalized patients during a period of 6 months. An incidence of 6.1% was reported, of which 42% were severe; 1% of the severe reactions resulted in the patient's death. Using an automatic detection system in a Salt Lake City hospital, Claussen and coinvestigators identified 731 ADRs among 36,653 hospitalized patients. Of note, only 12.3% of these were reported by physicians in the hospital. In a meta-analysis of 33 U.S. prospective studies from 1966 to 1996, Lazarou reported that 15% of hospitalized patients experienced ADRs and that the frequency of drug-related hospital admissions varied from 3 to 6%. Most other subsequent studies reported similar data.[1] Epidemiologic information on drug allergy in nonhospitalized people and in the general population is even more limited and is confined mainly to studies of antibiotics.

Risk Factors

Some risk factors have been identified for the development of drug allergy.[2] Certain drugs more commonly cause adverse reactions, and some drugs lead to more severe reactions (Table 254-1). The dosage and route of administration of a drug can also be risk factors; intermittent, repeated administrations of a drug can be more sensitizing than uninterrupted therapy. Drug allergy is more commonly reported in women and in patients with HIV infection[3] or reactivation of some herpes viruses. Some ethnic groups appear to be more prone to certain ADRs. For example, white Americans are at a higher risk than other ethnic groups for hypersensitivity reactions to abacavir, a reverse transcriptase inhibitor. For drug allergy caused by angiotensin-converting enzyme inhibitors, the more vulnerable population is African American.

In the United States, approximately 10% of individuals who seek health care have a history of penicillin allergy. However, if tested with an appropriate panel of skin tests, less than 10% of those individuals would be deemed to have a penicillin allergy. Individuals with a positive history and negative skin test results tolerate penicillin-type antibiotics at the same rate as the general population with a negative history; in addition, there is a very low rate of resensitization.

TABLE 254-1 DRUGS FREQUENTLY IMPLICATED IN ALLERGIC DRUG REACTIONS

Allopurinol
Amiodarone
Antiarrhythmic drugs (procainamide, quinidine)
Antibiotics (β-lactams, sulfas, nitrofurans)
Anticonvulsants (hydantoin, phenobarbital, carbamazepine)
Antihypertensive agents (angiotensin-converting enzyme inhibitors)
Antipsychotic tranquilizers
Antisera (antitoxins, antivirals)
Antituberculous drugs (isoniazid, rifampicin)
Aspirin and nonsteroidal anti-inflammatory drugs
Biologics (monoclonal antibodies such as anti–tumor necrosis factor and other recombinant DNA protein products)
Chemotherapy agents (cisplatin, doxorubicin, taxanes)
Enzymes (L-asparaginase, streptokinase, chymopapain)
Heavy metals (gold salts)
Muscle relaxants (rocuronium, succinylcholine)
Radiocontrast media
Vaccines (egg protein, gelatin)

PATHOBIOLOGY

Hypersensitivity reactions to drugs can be classified according to the type of immunologic reaction, as originally described by Gell and Coombs with later modifications by Janeway, Kay, and Pichler. An immunologic response to any antigen may be diverse and the resulting reaction complex; drugs are no exception. Drugs that are more frequent perpetrators of significant allergy are listed in Table 254-1.

Most pharmacologic agents are simple structures with a molecular mass of less than 1000 D. Alone, they are unable to induce hypersensitivity-type immunologic responses. However, most of these agents have the ability to covalently bind to proteins and form hapten-carrier complexes, with the low-molecular-weight agent acting as the hapten and the protein being the carrier. Hapten-carrier complexes can induce immunologic responses, with most responses being directed at the hapten. In addition to low-molecular-weight drugs acting as haptens, there is evidence that they may activate immune receptors by binding to them directly. This is known as the pharmacologic interaction with immune receptors (or the p-I) concept.[4]

A well-known example of a low-molecular-weight agent is penicillin. Benzylpenicillin has a molecular mass of approximately 300 D and is metabolized into a penicilloyl hapten moiety. The penicilloyl moiety, which constitutes about 95% of all penicillin metabolites, is referred to as the major determinant because it is the major metabolite in terms of quantity. It has been conjugated to poly-D-lysine to form penicilloyl-polylysine, which is now commercially available as Pre-Pen (ALK-Abelló, Round Rock, TX) for skin testing. The other 5% of penicillin metabolites are referred to as the minor determinants. Although they are minor in quantity, these determinants actually cause most of the immediate-type anaphylactic reactions, whereas the major determinant is associated with later and less severe reactions. Minor determinant reagents have never been commercially available in the United States. Penicillin skin testing[5] is not widely used by U.S. physicians; annually, only 40,000 doses of the major determinant are sold.

In contrast to simple low-molecular-weight drugs, therapeutic agents that are proteins with a molecular mass exceeding 5000 D can be recognized by the human immune system and can result in sensitization and hypersensitivity reactions on subsequent exposure. Because these proteins are complete antigens, they can be used as skin testing reagents or as antigens or allergens in in vitro assays. Included among therapeutic protein reagents that reportedly cause hypersensitivity are antithymocyte globulin (rabbit or equine), streptokinase, latex, and vaccines such as tetanus toxoid. Biologics, including monoclonal antibodies, are increasingly recognized causes of drug hypersensitivity. As anticipated, murine antibodies are most immunogenic, followed by chimeric and then humanized monoclonals. Unexpectedly, a variety of human recombinant proteins, including insulin and fully human monoclonal antibodies, can cause hypersensitivity reactions. In addition to hypersensitivity reactions, biologics such as monoclonal antibodies can cause other immunologic reactions (Chapter 36). One such reaction is the *cytokine release syndrome*, in which high cytokine levels result in systemic symptoms, including fever, arthralgia, and capillary leak; interleukin-2 is the original biologic agent in which this was described.[6] Immune imbalance is another immunologic reaction, exemplified by anti–tumor necrosis factor therapy that results in immune dysregulation consisting of increased susceptibility to infection or autoimmunity.

CLINICAL MANIFESTATIONS

The clinical manifestations of drug allergy often include a dermatologic component (Chapter 440). It is estimated that 80 to 90% of drug allergies result in one of the following cutaneous manifestations: exanthematous or morbilliform eruption; urticaria, angioedema, or both; contact dermatitis; fixed drug eruption; erythema multiforme–like eruption; or photosensitivity.[7] Severe cutaneous adverse reactions (SCARs) are generally induced by drugs and encompass the conditions of Stevens-Johnson syndrome and toxic epidermal necrolysis; drug-induced eosinophilia and systemic syndrome, also known as drug-induced hypersensitivity syndrome; and acute generalized exanthematous pustulosis. These conditions, although rare, cause significant morbidity and even mortality, which is why it is important for the treating physician to promptly recognize SCARs and to discontinue implicated drugs. Some features of SCARs that distinguish them from nonserious cutaneous reactions include involvement of other organs (e.g., liver, kidneys); fever; eosinophilia; mucosal involvement; and lesions that are painful, blistering, or pustular.

DIAGNOSIS

The diagnosis of drug allergy may be simple if a patient has recently started therapy with a single agent known to cause hypersensitivity, such as a β-lactam antibiotic. In contrast, in a hospitalized patient in whom multiple drugs have been started and stopped, identifying the offending drug may be difficult, requiring a complete and exhaustive history along with a physical examination. It also requires compatible clinical manifestations and temporal relationships. In vitro tests are rarely useful clinically. In vivo testing, such as cutaneous tests and provocative test dosing, may be indicated in some situations.

Differential Diagnosis

To distinguish drug allergy from other ADRs, several criteria are helpful.[8] Allergic reactions occur in a tiny fraction of individuals who receive the drug, and they cannot be predicted. The observed clinical effects do not resemble known pharmacologic actions of the drug. In the absence of prior exposure to the drug, allergic or hypersensitivity symptoms rarely appear before 1 week of continuous therapy. In general, drugs used consistently for several months or longer are rarely responsible.

Drug allergy often resembles other allergic or hypersensitivity reactions, such as anaphylaxis, urticaria, and serum sickness–like illness. Although most drug reactions include cutaneous manifestations, some involve only other organ systems, for example, pulmonary infiltrates with eosinophilia, hepatitis, and acute interstitial nephritis. A list of drugs that cause organ-specific reactions is provided in Table 254-2. Drug-specific antibodies or T-cell receptors have been identified that react with the suspected drugs or relevant drug metabolites. As with ADRs in general, the reaction often subsides after the drug is discontinued. However, a hypersensitivity reaction may persist or

TABLE 254-2 ORGAN-SPECIFIC REACTIONS AND IMPLICATED DRUGS

REACTION	IMPLICATED DRUG
PULMONARY MANIFESTATIONS	
Pulmonary infiltrates with eosinophilia	Minocycline, nitrofurantoin
Pneumonitis and fibrosis	Bleomycin, amiodarone
Noncardiogenic pulmonary edema	Hydrochlorothiazide, cocaine, heroin, methadone
AUTOIMMUNE MANIFESTATIONS	
Drug-induced lupus	Hydralazine, procainamide
DRUG-INDUCED IMMUNE CYTOPENIAS	
Thrombocytopenia	Quinidine, gold salts, sulfonamides, heparin
Hemolytic anemia	Penicillin, methyldopa
Agranulocytosis	Sulfonamides, propylthiouracil, quinidine, procainamide, phenytoin
HEPATIC MANIFESTATIONS	
Cholestasis	Aminosalicylic acid, dapsone
Hepatocellular damage	Phenothiazines, erythromycin
Mixed pattern	Halothane, isoniazid, diclofenac
	Phenytoin, sulfonamides
RENAL MANIFESTATIONS	
Nephrotic syndrome	Gold salts, captopril, NSAIDs, penicillamine
Acute interstitial nephritis	β-Lactam antibiotics, NSAIDs, sulfonamides
LYMPHOID SYSTEM MANIFESTATIONS	
Pseudolymphoma	Phenytoin
Infectious mononucleosis–like syndrome	Aminosalicylic acid, dapsone
CARDIAC MANIFESTATIONS	Sulfonamides, β-lactam antibiotics
NEUROLOGIC MANIFESTATIONS	
Peripheral neuritis	Colchicine, nitrofurantoin, sulfonamides

NSAIDs = nonsteroidal anti-inflammatory drugs.

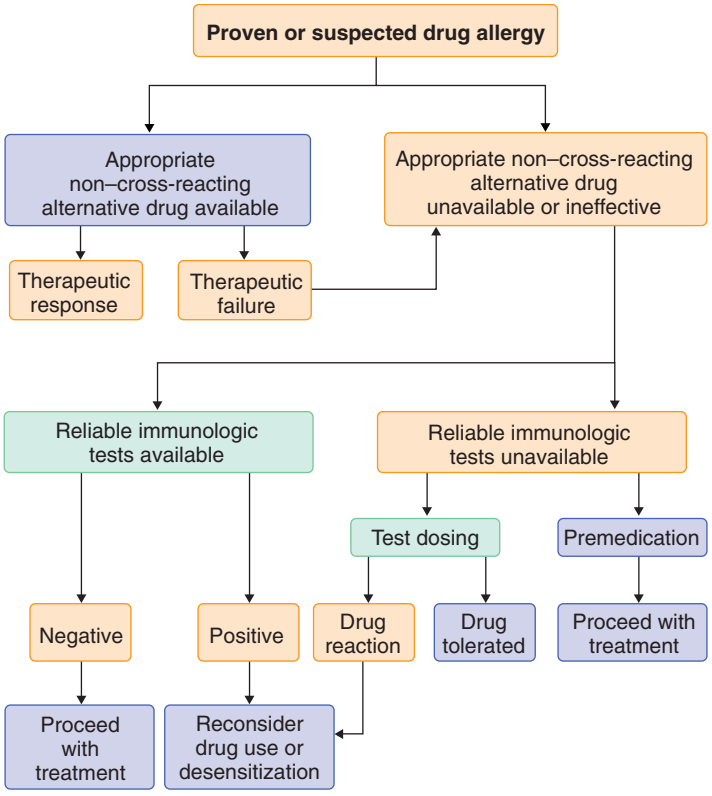

FIGURE 254-1. Guidelines for the treatment of patients with a history of drug allergy. In patients with a suspected or known drug allergy, the first choice is to use an appropriate non–cross-reacting drug. If such a drug is not available, or if the patient does not respond to it, further evaluation is based on the availability of a reliable immunologic test to detect drug hypersensitivity.

even intensify because of the formation of drug metabolites, which act as haptens and bind to carrier proteins such as human serum albumin.

TREATMENT

Evidence-Based Treatments

There is a paucity of evidence-based information regarding drug allergy, a disease that is generally iatrogenic. One study evaluated HIV patients who previously had an adverse reaction to cotrimoxazole; it was concluded that desensitization resulted in fewer adverse reactions and fewer treatment discontinuations in patients with a previous history of mild or moderate hypersensitivity.[A1] A second study, evaluating treatment for toxic epidermal necrolysis, concluded that there are no randomized controlled trials of the most commonly used therapies (i.e., systemic steroids, cyclosporine, intravenous immune globulins).

There are published clinical guidelines for the management of infusion-related hypersensitivity reactions caused by the administration of chemotherapeutic or biologic therapy. These guidelines were developed as part of a performance improvement initiative and resulted in a standardized approach to the management and reporting of ADRs.

PREVENTION

Although the outcome of ADRs is generally favorable, prevention is the obvious goal. The physician should prescribe medications only if they are clinically appropriate and, if possible, should avoid drugs that are known to produce significant hypersensitivity reactions (see Table 254-1). Before starting a medication, the patient should be asked about prior ADRs to the medication or to other pharmacologically related medications. If appropriate, oral administration is probably preferable to parenteral administration; anaphylaxis is less likely, as is sensitization. Protocols for skin testing to foreign antisera and for management of medication hypersensitivity reactions (e.g., premedication, test dosing, desensitization) are available.[9,10] Therapeutic guidelines regarding treatment of the most important and common ADRs are also reviewed in those references.[11] A general algorithm is provided in Figure 254-1.

The risk of an anaphylactic reaction to a drug such as penicillin is a function of the history of onset, severity, and proximity (Table 254-3). If an individual experienced an immediate-type reaction that was rapid in onset, involved life-threatening symptoms or signs, and occurred relatively recently, that individual is at high risk for a severe anaphylactic reaction on subsequent exposure.

TABLE 254-3 RISK OF ANAPHYLACTIC REACTION TO PENICILLIN OR OTHER PHARMACOTHERAPEUTIC AGENTS

FACTOR	LOW RISK	HIGH RISK
Onset of previous reaction	>24 hr	<30 min
Signs and symptoms of previous reaction	Morbilliform eruption Urticaria alone	Life-threatening symptoms: hypotension, upper airway angioedema, bronchospasm
Time elapsed since previous reaction	>20 yr	<1 yr

Even with a negative Pre-Pen skin test result, a patient could have reactivity against minor determinants; therefore, the approach to a patient who needs a β-lactam antibiotic depends on the risk as listed in Table 254-3. Risks and benefits should be thoroughly discussed and documented. In a high-risk individual, cautious test dosing can be performed. If a reaction occurs, desensitization can be considered if the clinical risks and benefits so warrant.

PROGNOSIS

Most drug allergies involve cutaneous eruptions that are self-limited and resolve shortly after the offending agent has been discontinued. However, severe, life-threatening reactions occur in approximately 1 in every 1000 hospitalized patients. SCARs are especially likely to cause morbidity and mortality. In 1998, the death rate for hospitalized Medicare patients was 20% higher in those who experienced an ADR. The proportion of ADRs that were allergic reactions was not determined in this study, but it can be estimated at about one fifth. The incidence of adverse cutaneous reactions to drugs is higher in women than in men. There is also an increased incidence of ADRs in the elderly.

One of the most severe reactions associated with drug allergy is anaphylactic shock (Chapter 253). It is usually immunoglobulin E (IgE) mediated, but it may occur with non–IgE-mediated reactions to drugs such as nonsteroidal anti-inflammatory drugs or radiocontrast media. It is estimated that approximately 1500 people die annually in the United States owing to anaphylaxis from medications. In the United Kingdom, drugs are the leading cause of anaphylactic fatalities.

FUTURE DIRECTIONS

Pharmacogenomics will be an important method of identifying those individuals at risk for a significant allergic reaction to a given drug.[12] Human leukocyte antigen (HLA) genotyping can reportedly identify individuals who are at increased risk for drug hypersensitivity. For example, individuals with HLA-B*5701 are at greater risk for a drug hypersensitivity reaction to abacavir, an HIV transcriptase inhibitor. Severe cutaneous adverse reactions to allopurinol are highly associated with the genetic marker HLA-B*5801. In patients of Asian ancestry, HLA-B*1508 is highly associated with development of Stevens-Johnson syndrome if carbamazepine is prescribed. Why this genetic variant is not a risk factor in patients of African or European ancestry is unclear.[13] Other avenues by which susceptible individuals may be identified include polymorphisms in genes for immune recognition molecules, drug-metabolizing enzymes, and macromolecular adduct repair systems.

Grade A Reference

A1. Lin D, Li WK, Rieder MJ. Cotrimoxazole for prophylaxis or treatment of opportunistic infections of HIV/AIDS in patients with previous history of hypersensitivity to cotrimoxazole. *Cochrane Database Syst Rev.* 2007;2:CD005646.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Classen DC, Jaser L, Budnitz DS. Adverse drug events among hospitalized Medicare patients: epidemiology and national estimates from a new approach to surveillance. *Jt Comm J Qual Patient Saf*. 2010;36:12-21.
2. Thong BYH, Tan TC. Epidemiology and risk factors for drug allergy. *Br J Clin Pharmacol*. 2011;71:684-699.
3. Yunihastuti E, Widhani A, Karjadi TH. Drug hypersensitivity in human immunodeficiency virus–infected patient: challenging diagnosis and management. *Asia Pac Allergy*. 2014;4:54-67.
4. Warrington R. Drug allergy: causes and desensitization. *Hum Vaccin Immunother*. 2012;8:1513-1524.
5. Fox S, Park MA. Penicillin skin testing in the evaluation and management of penicillin allergy. *Ann Allergy Asthma Immunol*. 2011;106:1-7.
6. Sathish JG, Sethu S, Bielsky MC, et al. Challenges and approaches for the development of safer immunomodulatory biologics. *Nat Rev Drug Discov*. 2013;12:306-324.
7. Stern RS. Clinical practice. Exanthematous drug eruptions. *N Engl J Med*. 2012;366:2492-2501.
8. Dworzynski K, Ardern-Jones M, Nasser S, et al. Diagnosis and management of drug allergy in adults, children and young people: summary of NICE guidance. *BMJ*. 2014;349:g4852.
9. Scherer K, Brockow K, Aberer W, et al. Desensitization in delayed drug hypersensitivity reactions—an EAACI position paper of the Drug Allergy Interest Group. *Allergy*. 2013;68:844-852.
10. Joint Task Force on Practice Parameters; American Academy of Allergy, Asthma and Immunology; American College of Allergy, Asthma and Immunology; Joint Council of Allergy, Asthma and Immunology. Drug allergy: an updated practice parameter. *Ann Allergy Asthma Immunol*. 2010;105:259-273.
11. Gangemi S, Guameri C, Romeo P, et al. Safety and reliability of the drug tolerance test: our experience in 739 patients. *Pharmacology*. 2011;87:90-95.
12. Karlin E, Phillips E. Genotyping for severe drug hypersensitivity. *Cur Allergy Asthma Rep*. 2014;14:418.
13. Yip VL, Alfirevic A, Pirmohamed M. Genetics of immune-mediated adverse drug reactions: a comprehensive and clinical review. *Clin Rev Allergy Immunol*. 2014; [Epub ahead of print].

REVIEW QUESTIONS

1. Which of the following drugs is *most* likely to cause drug hypersensitivity?

A. Murine monoclonal anti-CD3 antibody
B. Human insulin
C. Chimeric monoclonal anti-CD20 antibody
D. Humanized tumor necrosis factor antibody
E. Human tumor necrosis factor antibody

Answer: A Biologic proteins from most to least allergenic are murine, chimeric, humanized, and finally fully human protein. Even fully human proteins like human insulin or fully human anti–tumor necrosis factor can cause hypersensitivity. However, the most likely protein to cause drug allergy is the most foreign protein, that is, a murine protein. (See Pathobiology.)

2. Which of the following clinical characteristics is *least* concerning for a serious cutaneous adverse reaction (SCAR) to a drug?

A. Blistering lesions
B. Painful lesions
C. Pruritic lesions
D. Pustular lesions

Answer: C Some features of SCARs that distinguish them from nonserious cutaneous reactions include involvement of other organs (e.g., liver, kidneys); fever; eosinophilia; mucosal involvement; and lesions that are painful, blistering, or pustular. Pruritus is not a distinguishing feature.

3. A patient taking multiple medications develops acute interstitial nephritis with normal findings on skin examination. Which of the listed drugs is the most likely culprit?

A. Propranolol
B. Phenytoin
C. Colchicine
D. Ampicillin

Answer: D Although most drug reactions include cutaneous manifestations, some only involve other organ systems, for example, pulmonary infiltrates with eosinophilia, hepatitis, and acute interstitial nephritis. Drugs that have been described to cause acute interstitial nephritis in the absence of cutaneous involvement include sulfonamides, nonsteroidal anti-inflammatory drugs, and penicillins like ampicillin and methicillin.

4. HLA genotyping has been useful in identifying individuals at increased risk of hypersensitivity to which of the following drugs?

A. Amoxicillin
B. Minocycline
C. Ganciclovir
D. Abacavir

Answer: D Human leukocyte antigen (HLA) genotyping can reportedly identify individuals who are at increased risk for drug hypersensitivity. For example, individuals with HLA-B*5701 are at greater risk for a drug hypersensitivity reaction to abacavir, an HIV transcriptase inhibitor. At many centers, abacavir is prescribed only after the HLA genotype is determined not to be HLA-B*5701.

5. Among the following, what is the *most* significant risk factor for drug allergy?

A. Male sex
B. HIV infection
C. Atopy
D. HLA-B27 genotype

Answer: B Drug-related rashes have been estimated to be 100 times more common in HIV-positive patients than in the general population. The reasons for this are not clear but are likely to be multifactorial and include changes in drug metabolism, oxidative stress, cytokine profiles, and immune hyperactivation. (See section Risk Factors under Epidemiology.)

255

MASTOCYTOSIS

CEM AKIN

DEFINITION

Mastocytosis is a heterogeneous group of disorders characterized by pathologic accumulation of mast cells in tissues such as skin and bone marrow. According to the classification of the World Health Organization (WHO), based on clinical presentation and pathologic findings, there are seven distinct categories of mastocytosis (Table 255-1). The term *cutaneous mastocytosis* describes skin disease alone without any evidence of internal organ involvement, whereas the term *systemic mastocytosis* describes the disorder when it involves internal organs (most commonly the bone marrow) with or without skin disease.

EPIDEMIOLOGY

Mastocytosis can be diagnosed at any age.[1] Pediatric-onset and adult-onset forms are distinguished on the basis of the age of the patient at initial diagnosis. These forms display differences in their clinical course, molecular pathology, and prognosis. The most common clinical scenario leading to diagnosis in the pediatric population is a child presenting with skin lesions of cutaneous mastocytosis within the first year of life. Patients with a later onset of skin lesions are more likely to have systemic mastocytosis, as are most patients with adult-onset mastocytosis. The disease has been diagnosed in all ethnic populations. Estimates of the prevalence of patients with cutaneous mastocytosis range from 1 in 500 to 1 in 8000 patients presenting in dermatology clinics. The prevalence of systemic mastocytosis is more difficult to estimate because the diagnosis requires biopsy of an involved tissue and a high degree of clinical suspicion, especially if skin lesions are absent. Systemic mastocytosis is likely to be underdiagnosed, considering the fact that there are neither physical examination findings nor routine hematologic or chemistry laboratory abnormalities specifically associated with the disease. Consequently, it is not unusual to encounter several years' delay after the onset of symptoms in many patients before a diagnosis of mastocytosis is reached. The disease is sporadic, although rare cases of familial occurrence have been described.

PATHOBIOLOGY

Pathogenesis

The pathogenesis of mastocytosis involves the accumulation of mast cells in tissues, with mediators released by activated mast cells. The primary reason for the increased mast cell numbers in tissues appears to be defective apoptosis rather than uncontrolled proliferation. It is unusual to see increased mitotic activity in biopsy specimens from patients with mastocytosis, and, in most patients, the disease follows an indolent course. Tissue microenvironment and altered chemotaxis may also contribute to the final level of tissue mast cell burden.

Genetics

Mast cells are derived from hematopoietic progenitors (Chapter 156). Systemic mastocytosis is associated with somatic gain-of-function point mutations in the *KIT* (formerly *c-kit*) gene of the mast cell progenitor, leading to a clonal neoplastic expansion of mast cells. *KIT* encodes a transmembrane receptor (Kit) whose intracellular portion functions as a tyrosine kinase enzyme. The extracellular portion of Kit binds the cytokine stem cell factor (SCF or Kit ligand). The interaction between SCF and Kit provides the single most important growth and differentiation stimulus for mast cells from their progenitors. Under physiologic conditions, homodimeric SCF binds and cross-links two Kit receptor molecules, which leads to autophosphorylation of the tyrosine amino acids of the intracellular portion of the Kit molecule. Phosphorylated tyrosine residues in turn act as docking sites for downstream adaptor and signal transduction molecules that regulate the differentiation, proliferation, chemotaxis, and functional activation of mast cells.

The most common mutation reported in mastocytosis[2] involves codon 816 in *KIT* (located in exon 17), resulting in the replacement of an aspartic acid by a valine residue (D816V) in the Kit protein, leading to ligand-independent autophosphorylation. The D816V mutation has been shown in lesional mast cells from the skin or bone marrow tissue of more than 90% of adults and approximately 40% of pediatric patients with mastocytosis. Another 40% of pediatric patients carry *KIT* mutations in other exons, most commonly in exons 8 and 9. *KIT* mutations can be demonstrated in non–mast cell hematopoietic lineages in advanced variants of systemic mastocytosis, similar to the multilineage involvement observed in myeloproliferative neoplasms (Chapter 166). The sensitivity of detecting the mutation is much higher when a lesional tissue such as bone marrow or skin is analyzed compared with peripheral blood. Other pathogenetic factors, some yet to be determined, appear to be responsible for the final disease phenotype because the presence of the D816V *KIT* mutation alone does not explain the remarkable heterogeneity in the clinical presentation and prognosis of the disease. Molecular aberrations in *TET2*, *SRSF2*, *ASXL1*, *CBL*, *RUNX1*, and *DNMT3A* were the other most frequently identified mutations in advanced forms of systemic mastocytosis. In advanced systemic mastocytosis, most patients carry three or more mutations.

TABLE 255-1 WORLD HEALTH ORGANIZATION CLASSIFICATION OF MASTOCYTOSIS

Cutaneous mastocytosis
Indolent systemic mastocytosis
Systemic mastocytosis with associated clonal hematologic non–mast cell lineage disease
Aggressive systemic mastocytosis
Mast cell leukemia
Mast cell sarcoma
Extracutaneous mastocytoma

From Horny HP, Metcalfe DD, Bennett JM, et al. Mastocytosis. In: Swerdlow SH, Campo E, Harris NL, et al, eds. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon, France: IARC Press; 2008:54-63.

CLINICAL MANIFESTATIONS

Symptoms

The symptoms of mastocytosis are primarily related to the release of mast cell mediators and rarely by the destructive infiltration of mast cells into tissues. Mast cell activation results in the release of various preformed mediators stored in mast cell granules, de novo synthesis of sulfidopeptide leukotrienes such as LTC_4 and prostaglandins (mostly PGD_2) from membrane lipids, and cytokine synthesis. Preformed mediators stored in mast cell granules include histamine; proteases such as tryptase, chymase, and carboxypeptidase A; and proteoglycans such as heparin and chondroitin sulfate. Vasoactive mediators such as histamine, LTC_4, and PGD_2 at local or distant tissues cause vasodilation, which may lead to flushing, tachycardia, hypotension, presyncope, and syncope. Histamine also causes pruritus and stimulates gastric acid hypersecretion from parietal cells. Mast cells are rich sources of cytokines. Elevated serum levels of tumor necrosis factor-α and interleukin-6 have been found in patients with mastocytosis and may contribute to the pathophysiologic process of fatigue and accelerated osteoporosis observed in some patients. Rare aggressive categories of mastocytosis may be associated with an extensive destructive infiltration of mast cells into tissues such as the gastrointestinal tract, which may result in malabsorption, and the liver, which may cause portal fibrosis with associated portal hypertension.

Mast cell activation and mediator release may occur after triggers, such as temperature changes (e.g., hot showers), exercise, ingestion of alcohol or spicy foods, emotional stress, insect stings, and exposure to certain drugs (such as opioid analgesics, nonsteroidal anti-inflammatory drugs, or muscle relaxants), and sometimes spontaneously without an obvious trigger. The prevalence of atopic disease in patients with mastocytosis is similar to that in the general population, and the serum immunoglobulin E (IgE) level is often found to be low. However, patients with anaphylactic sensitivity to hymenoptera venoms appear to have a disproportionately high incidence of mastocytosis.[3]

Mastocytosis is a disease with protean clinical manifestations. Although in some patients the only complaint is the cosmetic appearance of urticaria pigmentosa lesions, others suffer from frequent episodes of vascular instability or have life-threatening hematologic disease. In general, patients with mastocytosis belong to one of two broad categories, according to the site of tissue involvement: those with cutaneous disease alone, or those with systemic disease with or without skin involvement. Cutaneous mastocytosis (i.e., disease limited to the skin in the absence of internal organ involvement)

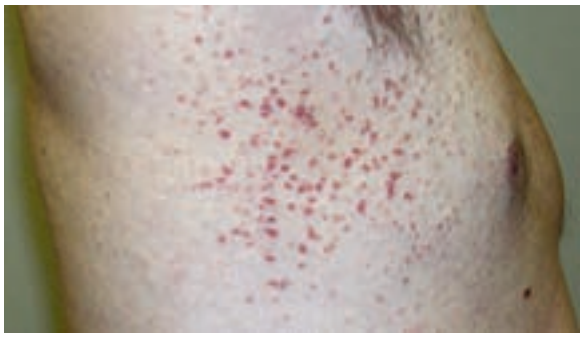

FIGURE 255-1. Urticaria pigmentosa.

is commonly diagnosed in children within the first year of life, whereas systemic mastocytosis is mostly diagnosed in adults by a bone marrow biopsy and aspirate.

Cutaneous Manifestations

Maculopapular skin lesions of urticaria pigmentosa are the most common presentation of cutaneous mastocytosis (Fig. 255-1). They are also present in 50 to 90% of patients with systemic mastocytosis, depending on the disease category. Remarkably different in appearance from urticaria or hives, lesions of urticaria pigmentosa are fixed, tan- to salmon-colored lesions varying in size from a few millimeters to a few centimeters. They are most prominently observed on the trunk and extremities and tend to spare the face and the sun-exposed areas of the skin, although facial involvement may be seen in children. Blistering of the lesions may occur in children in the first 3 years of life. The lesions are generally not pruritic at rest but may urticate after exposure to a number of triggers (see Pathobiology). Many patients note that the skin lesions become more prominent after exposure to heat or after physical irritation such as rubbing. The lesions may be found concentrated in skin areas that are prone to irritation, such as the axillae and groin.

Uncommon presentations of cutaneous mastocytosis include mastocytomas, diffuse cutaneous mastocytosis, and telangiectasia macularis eruptiva perstans (TMEP). Mastocytomas are benign and generally solitary mast cell tumors, although they have been known to precede urticaria pigmentosa lesions in some cases. They occur almost exclusively in children, and physical irritation of the lesion may result in generalized flushing and other symptoms of mast cell mediator release. Diffuse cutaneous mastocytosis is another form of skin involvement seen exclusively in children. It is characterized by diffuse thickening of the skin and appendages with a peau d'orange appearance without individual urticaria pigmentosa lesions. TMEP is a rare form of cutaneous mastocytosis characterized by the presence of diffuse telangiectatic macules. Because TMEP lesions are generally seen in the presence of urticaria pigmentosa, there is debate about whether TMEP represents a distinct form of cutaneous mastocytosis.

Patients with cutaneous mastocytosis may manifest systemic symptoms, such as abdominal pain, diarrhea, and flushing.

Systemic Manifestations

Symptoms caused by mast cell degranulation may be experienced as brief, recurrent, and self-limited episodes with multiorgan manifestations or as chronic complaints during a prolonged time course. A typical mast cell degranulation episode may variably involve flushing, conjunctival hyperemia, nausea, vomiting, abdominal cramping, diarrhea, tachycardia, and lightheadedness. Hypotension may develop, and the episode may progress to full loss of consciousness in some patients. Therefore, mastocytosis should be ruled out in all patients with recurrent anaphylaxis before a diagnosis of idiopathic anaphylaxis can be made.[4] Tryptase, a protease stored in mast cell granules, may be elevated above the patient's baseline level in the serum or plasma if it is measured within 3 hours after the onset of the episode in patients with suspected mast cell degranulation or anaphylaxis, regardless of the cause. Angioedema, hives, and wheezing are uncommon in mastocytosis, in contrast to idiopathic anaphylaxis. Flushing usually involves the face and upper chest area. A consistent trigger can be identified in only a small number of patients (see Pathobiology). The episodes usually last 30 minutes to a few hours. Hypotensive episodes can be life-threatening, particularly in the presence of comorbidities, such as cardiac or pulmonary disease. Systemic mastocytosis should be suspected in all patients with systemic reactions to hymenoptera stings, especially those involving hypotensive syncope or near-syncope.

Gastrointestinal Symptoms

Gastrointestinal symptoms are observed in more than 50% of patients with mastocytosis. Epigastric pain, lower abdominal cramping, nausea, vomiting, or diarrhea can occur episodically in the context of an acute mast cell degranulation episode or on a chronic basis. Gastric acid hypersecretion induced by mast cell–derived histamine may lead to esophagitis, gastritis, and peptic ulcer disease, although measurements of basal acid output have shown great variability in different studies, ranging from hypersecretion in the range of Zollinger-Ellison syndrome to achlorhydria. Mucosal edema, thickened gastric or duodenal mucosal folds, or nodular lesions may be observed in radiographic or endoscopic evaluations. Diarrhea alternating with constipation may be seen. Severe persistent diarrhea may be complicated by clinically significant malabsorption in patients with aggressive systemic mastocytosis. Hematochezia, hematemesis, and melena are uncommon symptoms and should prompt endoscopic evaluation to rule out coexisting disease. Mast cells are constituents of the normal lamina propria in gastrointestinal mucosa, and their numbers may be increased in inflammatory states affecting the gastrointestinal tract. However, quantitation of mast cell numbers in gastrointestinal biopsy specimens is generally not helpful, and diagnosis of mast cell disease by a gastrointestinal biopsy, solely based on increased mast cell numbers without evidence for other WHO criteria, should be avoided. Mild to moderate hepatomegaly with or without abnormalities in serum transaminases may be observed, although portal hypertension and ascites are rare and indicate the presence of advanced categories of mastocytosis. Jaundice and findings on cholangiography resembling those of primary sclerosing cholangitis have been reported in some patients.

Musculoskeletal Symptoms

Musculoskeletal pain is common in patients with mastocytosis and is mostly caused by soft tissue pain resembling fibromyalgia. Accelerated osteoporosis may be seen in a subgroup of patients, particularly those with other risk factors, such as postmenopausal women, and those receiving glucocorticoid therapy. Pathologic compression fractures may be the initial finding in some patients.[5] A bone densitometry measurement should be recommended as part of the standard evaluation of women with mastocytosis and of any patient with a history of pathologic fractures. Radiographic abnormalities have been reported in up to 75% of patients with mastocytosis. In addition to generalized osteoporosis, bone surveys may show a mixture of sclerotic or lytic lesions, and skeletal scintigraphy may reveal focal or diffuse radiotracer uptake.

Hematologic Manifestations

Peripheral blood abnormalities have been noted in up to 50% of patients with systemic mastocytosis. Mild normochromic normocytic anemia is the most common abnormality, followed by thrombocytopenia, eosinophilia, monocytosis, and leukopenia. Eosinophilia in mastocytosis rarely causes organ damage, as is observed in chronic eosinophilic leukemia or idiopathic hypereosinophilic syndrome (Chapter 170). It is important to differentiate a primary eosinophilic disorder from mastocytosis with eosinophilia. Some cases of chronic eosinophilic leukemia are associated with the *FIP1L1-PDGFRA* fusion gene and respond to the drug imatinib, whereas systemic mastocytosis is associated with codon 816 point mutations of the *KIT* gene, which confers resistance to this drug.

Approximately 20% of patients with systemic mastocytosis have been reported to display evidence of another clonal non–mast cell hematologic disease. Non–mast cell clonal hematologic neoplasms associated with mastocytosis are commonly myeloid in nature (myeloproliferative neoplasms, myelodysplastic syndromes, or myeloid leukemias) but may also involve lymphoproliferative disorders, such as lymphomas, myelomas, and lymphocytic leukemias.

DIAGNOSIS

The diagnosis and classification of mastocytosis are carried out according to the guidelines published by the WHO. A suggested algorithm for the diagnosis of mastocytosis is shown in Figure 255-2.

Cutaneous Mastocytosis

Diagnosis of cutaneous mastocytosis is made by observing the typical hyperpigmented maculopapular lesions of urticaria pigmentosa and is confirmed

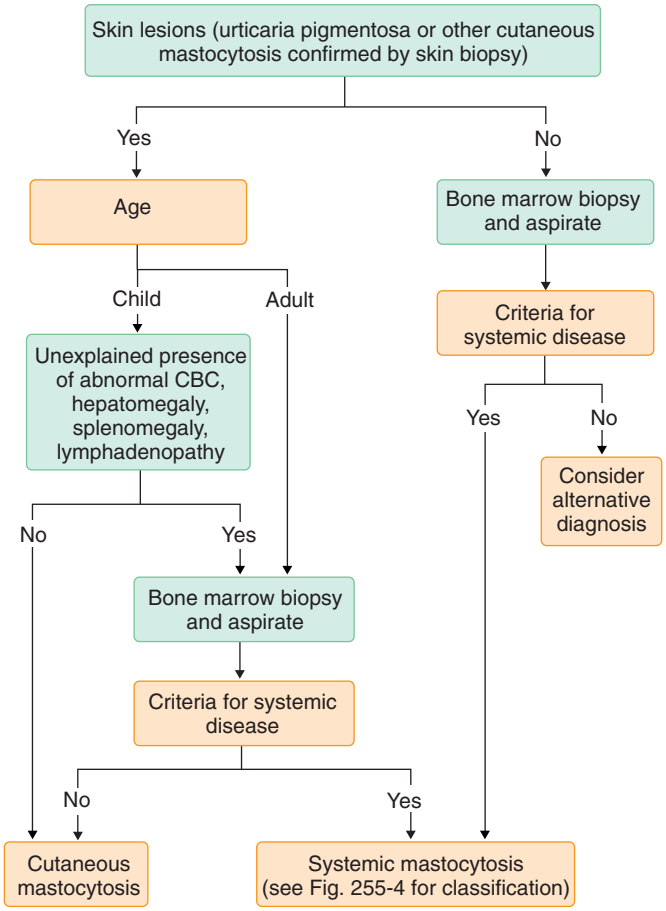

FIGURE 255-2. Suggested diagnostic algorithm for mastocytosis. CBC = complete blood count.

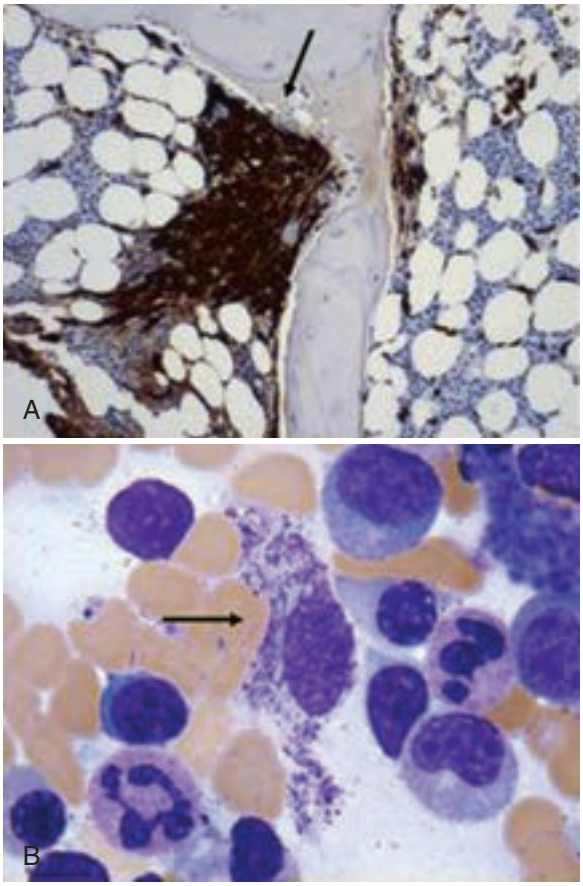

FIGURE 255-3. Diagnostic findings in the bone marrow biopsy specimen and aspirate smear. A, Characteristic mast cell aggregates on tryptase staining (major criterion) in biopsy section (*arrow*). **B,** Mast cells with atypical spindle shapes in aspirate smear *(arrow)*.

by skin biopsy, which shows infiltration of mast cells in the upper dermis, particularly in perivascular locations. Mild increases in mast cell numbers can be observed in inflammatory and neoplastic skin diseases, and establishing a diagnosis of cutaneous mastocytosis by a blind skin biopsy or biopsy of a lesion that does not have the typical appearance of urticaria pigmentosa should be avoided. A localized wheal-and-flare reaction limited to the lesional skin within a few minutes after rubbing or scratching of the skin is known as Darier's sign. Diagnosis of urticaria pigmentosa in adults should always prompt investigation of possible systemic mastocytosis.

Systemic Mastocytosis

Biopsy

The recommended diagnostic procedure to evaluate the presence of WHO diagnostic criteria for systemic disease (discussed later) is a bone marrow biopsy and aspiration. This procedure is recommended for all patients with adult-onset urticaria pigmentosa, patients with recurrent symptoms suggestive of mast cell degranulation (such as flushing and hypotension accompanied by abdominal complaints), patients with unexplained osteoporosis, and patients with suspected hematologic disease (see Clinical Manifestations). Children with an onset of lesions within the first year of life usually do not require a bone marrow biopsy unless they have abnormal blood counts, lymphadenopathy, hepatomegaly, or splenomegaly. Children with late-onset skin lesions and those who experience persistence of urticaria pigmentosa into adulthood should be considered for diagnostic evaluation for systemic disease.

World Health Organization Diagnostic Criteria

WHO guidelines for diagnosis of systemic mastocytosis consist of one major and four minor criteria (Table 255-2). Presence of the major criterion with at least one minor criterion, or demonstration of three minor criteria in the absence of the major criterion, is needed to establish a diagnosis of systemic mastocytosis and to distinguish it from reactive mast cell hyperplasia. The major diagnostic criterion is the presence of multifocal, dense aggregates of 15 or more mast cells in bone marrow or other extracutaneous tissue biopsy sections (Fig. 255-3A). Such clusters are frequently observed around blood vessels and next to bone trabeculae in bone marrow biopsy sections. Immunohistochemical staining for tryptase is the recommended method for visualization of mast cells. Routine hematoxylin and eosin or metachromatic stains such as toluidine blue are not sufficiently sensitive to demonstrate subtle mast cell infiltrates or abnormal morphologic features of mast cells within the infiltrates in decalcified bone marrow biopsy sections.

Mast cell morphology in bone marrow provides important clues to the diagnosis of systemic mastocytosis. Bone marrow mast cells in systemic mastocytosis often display atypical morphology, such as an elongated (spindle) shape, hypogranularity, and an eccentric or lobulated nucleus (Fig. 255-3B). These atypical mast cells are usually observed in close association with bone

TABLE 255-2 WORLD HEALTH ORGANIZATION DIAGNOSTIC CRITERIA FOR SYSTEMIC MASTOCYTOSIS

MAJOR
Multifocal, dense infiltrates of mast cells consisting of 15 or more mast cells in aggregates detected in sections of bone marrow and/or other extracutaneous organs, confirmed by tryptase immunohistochemistry or other special stains
MINOR
More than 25% of mast cells in biopsy sections or bone marrow aspirate smears showing spindle shape or atypical morphology
Detection of a *KIT* codon 816 point mutation in bone marrow, blood, or other extracutaneous organs
Expression of CD2 and/or CD25 by mast cells in bone marrow, blood, or extracutaneous organs
Persistent elevation of serum total tryptase >20 ng/mL*

*Criterion not valid if there is an associated clonal myeloid disorder.
From Horny HP, Metcalfe DD, Bennett JM, et al. Mastocytosis. In: Swerdlow SH, Campo E, Harris NL, et al, eds. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon, France: IARC Press; 2008:54-63.

marrow spicules in the aspirate smear. Mast cells in mast cell leukemia (MCL) may be very sparsely granulated.

Flow cytometric analysis of the mast cells in a bone marrow aspirate, when it is performed appropriately, is a sensitive diagnostic aid. The mean percentage of mast cells in a healthy bone marrow aspirate is approximately 0.02%, and it does not exceed 1% in most patients with mastocytosis. Therefore, to visualize the mast cell population correctly, the total cell numbers analyzed by flow cytometry should be significantly higher than those in other, more routine evaluations (e.g., leukemia phenotyping). The characteristic flow cytometric finding of systemic mastocytosis is the aberrant expression of CD25 or CD2 on $CD117^+$ mast cells. CD25 is more sensitive than CD2 as CD2 may be absent or weakly expressed in some cases of advanced mastocytosis. Aberrant CD25 expression can also be demonstrated by immunohistochemical staining of bone marrow biopsy specimens.[6] Serum tryptase level may be elevated in patients with mastocytosis.[7] Currently available commercial tryptase immunoassays measure levels of total tryptase, the sum of mature tryptase, and tryptase precursors. Mature tryptase enzyme is a serine protease stored in mast cell granules and is transiently elevated in serum or plasma after mast cell degranulation episodes, such as anaphylaxis. In contrast, tryptase precursor proenzymes (α and β protryptases) are constitutively secreted outside the cell, and their serum levels at baseline correlate with mast cell burden. The median serum tryptase level in a healthy population is approximately 5 ng/mL. A serum tryptase level higher than 20 ng/mL raises suspicion for systemic mastocytosis in the appropriate clinical setting. A normal tryptase level does not rule out a diagnosis of mastocytosis, and increased tryptase levels can be seen in other conditions, such as myelodysplastic syndromes, acute myeloid leukemias, chronic eosinophilic leukemia, and chronic renal insufficiency. Metabolites of histamine, such as *N*-methylhistamine, and prostaglandin D_2 can be elevated in a 24-hour urine specimen but are neither more sensitive nor more specific than the baseline serum tryptase measurement in mastocytosis.

Demonstration of a codon 816 *KIT* mutation (D816V) may be necessary to fulfill the diagnostic criteria in patients lacking the major criterion (see Pathobiology).[8] Examination of lesional tissues, such as skin and bone marrow, affords the highest sensitivity. Codon 816 *KIT* mutations have been detected in a variety of other neoplastic diseases, such as core binding factor acute myeloid leukemias, sinonasal lymphomas, and seminomas, in addition to mastocytosis.

A rare histologic variant with clustering of mature round mast cells without CD25 expression termed well-differentiated systemic mastocytosis has been described. These patients generally have a history of childhood-onset mastocytosis without the D816V *KIT* mutation and therefore may respond to imatinib as opposed to those with typical systemic mastocytosis carrying the D816V mutation (see Treatment).

World Health Organization Disease Categories

Each patient diagnosed with mastocytosis should be assigned a category of disease according to the WHO classification (see Table 255-1). *Cutaneous mastocytosis* in the absence of bone marrow and internal organ involvement is the most common category in patients with pediatric-onset disease.

Systemic mastocytosis is divided into the categories of indolent systemic mastocytosis, systemic mastocytosis with associated clonal hematologic non–mast cell lineage disease (SM-AHNMD), aggressive systemic mastocytosis, and MCL. An algorithm for classification of systemic mastocytosis is presented in Figure 255-4. *Indolent systemic mastocytosis* is the most common category in adults. Patients in this category usually have a normal life expectancy compared with age-matched general populations, although they experience symptoms related to release of mast cell mediators.[9] Indolent systemic mastocytosis follows a persistent course, and progression to a more advanced category is unusual (<5% of cases). *SM-AHNMD* is the second most common category in adults, and a non–mast cell hematologic disease is usually diagnosed at the time that the diagnosis of mastocytosis is made. Therefore, bone marrow biopsy and aspirate specimens should be carefully evaluated for the presence of other hematologic disease in every patient with newly diagnosed systemic mastocytosis. *Aggressive systemic mastocytosis* is a rare category characterized by the presence of organ dysfunction resulting from destructive mast cell infiltration. Aggressive systemic mastocytosis may involve the hematopoietic, gastrointestinal, and skeletal systems in the form of cytopenias, hypersplenism, malabsorption with weight loss, hepatomegaly with portal hypertension and ascites, and large osteolytic lesions with pathologic fractures; these constitute so-called C-findings as defined by the WHO criteria. MCL is characterized by 10% or more mast cells in the peripheral circulation or 20% or more mast cells in bone marrow aspirate smears, or both. To diagnose MCL, the mast cell percentage in bone marrow aspirate smears should be assessed in an area of the slide that is sufficiently distant from the spicules. *Mast cell sarcoma* and *extracutaneous mastocytoma* are rare diagnoses characterized by malignant and benign solid mast cell collections, respectively.

There is a subset of patients with recurrent idiopathic or hymenoptera venom–induced anaphylaxis who have evidence of clonal mast cells carrying the D816V *KIT* mutation or aberrantly expressing surface CD25, without fully meeting the WHO diagnostic criteria and without displaying urticaria pigmentosa skin lesions. Such patients are provisionally referred to as having a monoclonal mast cell activation syndrome.

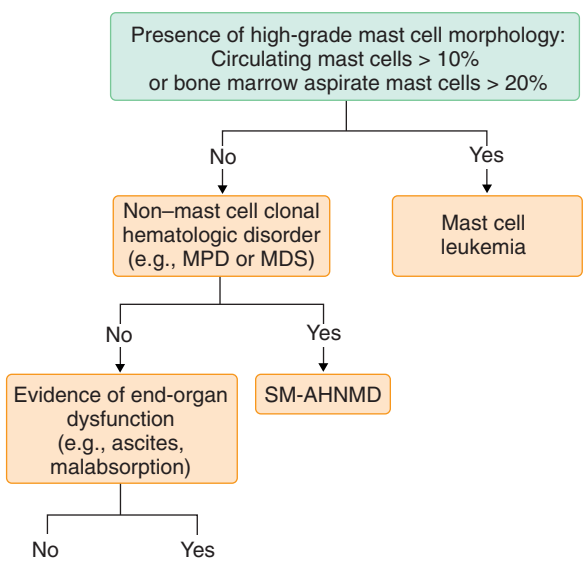

FIGURE 255-4. **An algorithm for classification of systemic mastocytosis.** ASM = aggressive systemic mastocytosis; ISM = indolent systemic mastocytosis; MDS = myelodysplastic syndromes; MPD = myeloproliferative disorders (neoplasms); SM-AHNMD = systemic mastocytosis with associated clonal hematologic non–mast cell lineage disease.

TREATMENT

Rx

The major goal of treatment for all categories of mastocytosis is symptom control. A reduction in mast cell numbers is considered only in disease categories with a poor prognosis (i.e., SM-AHNMD, aggressive systemic mastocytosis, MCL, and mast cell sarcoma).[10,11] Current treatment modalities have not been shown to change the natural course of the disease.[12]

Medical Therapy

Patients with cutaneous and indolent systemic mastocytosis are treated symptomatically. Pruritus in mastocytosis usually responds to scheduled doses of $histamine_1$-receptor blocker antihistamines, such as fexofenadine or cetirizine. Sedating antihistamines, such as hydroxyzine or diphenhydramine, may be used before bedtime. Photochemotherapy (oral psoralen plus ultraviolet A) or phototherapy may be helpful in patients with refractory pruritus; it results in symptomatic improvement and temporary fading of the pigmented skin lesions in up to 50% of patients. The side effects of phototherapy, including increased risk of skin cancer, should be taken into account when this treatment is considered.

$Histamine_2$-receptor blocker antihistamines, such as ranitidine or famotidine, are usually prescribed as a first-line treatment for patients with gastrointestinal complaints, such as heartburn, nausea, and abdominal pain. Proton pump inhibitors may be added in patients whose abdominal symptoms are refractory to $histamine_2$-receptor blockers. Oral cromolyn sodium (adult dose, 200 mg four times daily) has been effective in reducing abdominal pain, diarrhea, nausea, vomiting, and pruritus in various studies, although the beneficial effects are variable among patients. Finally, low to moderate doses of systemic glucocorticoids can be beneficial in unusual cases of aggressive mastocytosis presenting with recalcitrant diarrhea associated with malabsorption or hepatomegaly with ascites.

Cysteinyl leukotrienes, such as LTC_4, that are produced after mast cell activation are thought to contribute to symptoms in mastocytosis. Therefore, drugs targeting the synthesis or receptor binding of leukotrienes are usually added to the treatment regimens of patients who derive suboptimal relief of itching and abdominal pain from histamine receptor–blocking therapy. However, there have been no controlled studies evaluating the clinical efficacy of this class of drugs in patients with mastocytosis.

Self-administered epinephrine should be considered for all patients even if they do not have any history of hypotensive or anaphylactic episodes resulting in presyncope or syncope from acute mast cell degranulation. These episodes should be treated like systemic anaphylaxis (Chapter 253).

Cytoreductive therapy, considered in aggressive disease variants associated with poor prognosis, has yielded disappointing results thus far. Some patients with recurrent life-threatening episodes of mast cell mediator release unresponsive to conventional therapy may also be candidates for cytoreductive therapy after careful consideration of risks and benefits. Approaches to cytoreductive treatment of mastocytosis have included interferon alfa-2b and the nucleoside analogue 2-chlorodeoxyadenosine. Interferon alfa-2b (0.5 to 5 million units, three to five times per week), alone or with prednisone, has been reported to partially improve clinical and laboratory abnormalities in approximately 50% of patients with aggressive systemic mastocytosis, patients with osteoporosis and pathologic fractures, and patients with recalcitrant recurrent anaphylaxis, although complete histopathologic and molecular remissions appear to be rare. Interferon alfa is difficult to tolerate because of its many side effects, including influenza-like symptoms, bone pain, and depression. A regimen of 2-chlorodeoxyadenosine (0.10 to 0.17 mg/kg/day for 5 days, repeated at intervals of 4 to 8 weeks) has been reported to result in partial and transient responses in patients with advanced categories of disease in case reports and small series. MCL usually is treated with polychemotherapy as acute myeloid leukemia (Chapter 183), although a successful treatment regimen has not yet been identified.

Imatinib, a tyrosine kinase inhibitor with activity against wild-type *KIT*, *PDGFR*, and *abl*, has been effective in a small number of patients without D816V *KIT* mutation or with the *FIP1L1-PDGFRA* fusion gene, who present with chronic eosinophilic leukemia (Chapter 170) with a modest increase in bone marrow mast cells. However, most patients with mastocytosis have the D816V *KIT* mutation, which confers resistance to imatinib, and therefore are not appropriate candidates for this therapy.[13]

Ancillary and Other Therapies

Avoidance of the triggers of mast cell degranulation is an important adjunct to the pharmacologic treatment of symptoms. These show remarkable individual variation among patients (see Pathobiology), and the individual medical history can be helpful in identifying such triggers. General anesthesia and surgery impose an additional risk to patients with mastocytosis because several agents that are used perioperatively, such as muscle relaxants, opioid analgesics, and nonsteroidal anti-inflammatory drugs, can induce acute mast cell degranulation. Prior surgical and anesthesia records should be obtained if available, and an appropriate strategy for the anesthetic management of the patient should be determined, with close communication involving the patient, anesthesiologist, surgeon, and an allergist.

Non–mast cell clonal hematologic disorders associated with mastocytosis should be treated according to the standard-of-care guidelines for those disorders, regardless of the presence of mastocytosis. Bone marrow transplantation (Chapter 178) has yielded variable results for the treatment of mast cell disease, and occasional cases resulting in complete remission have been reported.

Venom immunotherapy is recommended for those with a history of systemic reactions to hymenoptera who have evidence of IgE-mediated sensitization (by blood or skin allergy testing). Most experts recommend the duration of the therapy to be indefinite as fatalities have been reported after discontinuation of immunotherapy.

Because of the high prevalence of osteoporosis and pathologic bone fractures in mastocytosis, bone densitometry should be considered a standard diagnostic procedure in adult patients with mastocytosis. If osteoporosis is detected, it should be treated per standard recommendations (Chapter 243).

PROGNOSIS

The prognosis for mastocytosis varies by the category of disease. At least 50% of patients with pediatric-onset cutaneous mastocytosis have complete resolution of the disease by adolescence, and the great majority of the rest of those patients experience improvement or fading of the skin lesions. Indolent systemic mastocytosis is a persistent disease but has a good prognosis without a decrease in life expectancy, and progression to a more aggressive disease category is rare.[14] Factors associated with poorer prognosis have been reported as the absence of urticaria pigmentosa, older age at onset of symptoms, elevated serum lactate dehydrogenase or alkaline phosphatase, thrombocytopenia, anemia, peripheral blood smear abnormalities, and detectability of the D816V *KIT* mutation in peripheral blood. The prognosis for SM-AHNMD is determined by the prognosis for the associated hematologic disorder. Aggressive systemic mastocytosis and MCL have poor prognoses, with median survival times of less than 3 years and less than 1 year, respectively.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Carter MC, Metcalfe DD, Komarow HD. Mastocytosis. *Immunol Allergy Clin North Am*. 2014;34:181-196.
2. Molderings GJ. The genetic basis of mast cell activation disease—looking through a glass darkly. *Crit Rev Oncol Hematol*. 2015;93:75-89.
3. Alvarez-Twose I, Bonadonna P, Matito A, et al. Systemic mastocytosis as a risk factor for severe Hymenoptera sting–induced anaphylaxis. *J Allergy Clin Immunol*. 2013;131:614-615.
4. Akin C. Anaphylaxis and mast cell disease. What is the risk? *Curr Allergy Asthma Rep*. 2010;10:34-38.
5. van der Veer E, van der Goot W, de Monchy JG, et al. High prevalence of fractures and osteoporosis in patients with indolent systemic mastocytosis. *Allergy*. 2012;67:431-438.
6. Morgado JM, Sánchez-Muñoz L, Teodósio CG, et al. Immunophenotyping in systemic mastocytosis diagnosis: "CD25 positive" alone is more informative than the "CD25 and/or CD2" WHO criterion. *Mod Pathol*. 2012;25:516-521.
7. Valent P, Sperr WR, Sotlar K, et al. The serum tryptase test: an emerging robust biomarker in clinical hematology. *Expert Rev Hematol*. 2014;7:683-690.
8. Valent P, Escribano L, Broesby-Olsen S, et al. Proposed diagnostic algorithm for patients with suspected mastocytosis: a proposal of the European Competence Network on Mastocytosis. *Allergy*. 2014;69:1267-1274.
9. Rabenhorst A, Christopeit B, Leja S, et al. Serum levels of bone cytokines are increased in indolent systemic mastocytosis associated with osteopenia or osteoporosis. *J Allergy Clin Immunol*. 2013;132:1234-1237.
10. Siebenhaar F, Akin C, Bindslev-Jensen C, et al. Treatment strategies in mastocytosis. *Immunol Allergy Clin North Am*. 2014;34:433-447.
11. Pardanani A, Tefferi A. Systemic mastocytosis in adults: a review on prognosis and treatment based on 342 Mayo Clinic patients and current literature. *Curr Opin Hematol*. 2010;17:125-132.
12. Cardet JC, Akin C, Lee MJ. Mastocytosis: update on pharmacotherapy and future directions. *Expert Opin Pharmacother*. 2013;14:2033-2045.
13. Ustun C, DeRemer DL, Akin C. Tyrosine kinase inhibitors in the treatment of systemic mastocytosis. *Leuk Res*. 2011;35:1143-1152.
14. Brockow K. Epidemiology, prognosis, and risk factors in mastocytosis. *Immunol Allergy Clin North Am*. 2014;34:283-295.

REVIEW QUESTIONS

1. Which of the following markers is aberrantly expressed by mast cells in patients with systemic mastocytosis?

A. CD25
B. CD117 (Kit)
C. High-affinity IgE receptor
D. CD45

Answer: A CD25 (alpha chain of the interleukin-2 receptor) is aberrantly expressed and is a sensitive diagnostic marker in systemic mastocytosis. All others are expressed by both normal and pathologic mast cells. (Valent P, Horny HP, Escribano L, et al. Diagnostic criteria and classification of mastocytosis: a consensus proposal. *Leuk Res.* 2001;25:603-625.)

2. A 25-year-old woman has recently been diagnosed with systemic mastocytosis by a bone marrow biopsy. There was no other hematologic disease noted. She experiences pruritus of the skin in hot weather and occasional abdominal cramping but is otherwise asymptomatic. *KIT* mutational analysis reveals the D816V mutation in bone marrow. Which of the following is the most appropriate treatment?

A. Imatinib
B. H_1 and H_2 antihistamines
C. Prednisone
D. Interferon alfa
E. Bone marrow transplantation

Answer: B This patient has indolent systemic mastocytosis and should be treated symptomatically. Cytoreductive therapies such as interferon alfa are indicated in patients with aggressive mastocytosis. The patient is not a candidate for imatinib because of indolent disease and presence of D816V *KIT* mutation. The patient's symptoms do not warrant bone marrow transplantation or use of systemic glucocorticoids, which are associated with long-term adverse effects. (Valent P, Akin C, Escribano L, et al. Standards and standardization in mastocytosis: consensus statements on diagnostics, treatment recommendations, and response criteria. *Eur J Clin Invest*. 2007;37:435-453.)

3. A 35-year-old farmer with a 10-year history of mastocytosis who is otherwise healthy has recently experienced anaphylaxis with brief loss of consciousness after being stung by a honeybee. Allergy skin testing confirmed honeybee allergy. What is the most appropriate next step in management of honeybee allergy in this patient?

A. Venom immunotherapy is not recommended because of increased risk of reactions to immunotherapy due to mastocytosis.
B. Recommend venom immunotherapy for 5 years.
C. Recommend venom immunotherapy indefinitely.
D. Confirm skin test results by blood testing before recommending immunotherapy.

Answer: C Patients with mastocytosis are at increased risk for life-threatening reactions to hymenoptera stings. Fatalities have been reported in patients after discontinuation of venom immunotherapy. Therefore, most experts recommend immunotherapy when IgE-mediated sensitization is found by either skin or blood testing in these patients. Whereas the patients are also at increased risk of having a systemic reaction during immunotherapy, the risk-benefit ratio is often considered favorable to initiate immunotherapy. The patients should always continue to carry a self-injectable epinephrine as the protection is lower than for those without mastocytosis. (Niedoszytko M, de Monchy J, van Doormaal JJ, et al. Mastocytosis and insect venom allergy: diagnosis, safety and efficacy of venom immunotherapy. *Allergy.* 2009;64:1237-1245; and Alvarez-Twose I, Bonadonna P, Matito A, et al. Systemic mastocytosis as a risk factor for severe Hymenoptera sting–induced anaphylaxis. *J Allergy Clin Immunol.* 2013;131:614-615.)

4. Which of the following patients is a candidate for therapy with imatinib?

A. 30-year-old with adult-onset indolent systemic mastocytosis with D816V *KIT* mutation who remains symptomatic despite symptomatic therapy
B. 30-year-old with adult-onset systemic mastocytosis with occasional symptoms on symptomatic therapy; peripheral blood analysis negative for D816V *KIT* mutation
C. 60-year-old with systemic mastocytosis, splenomegaly, and pancytopenias; peripheral blood and bone marrow positive for D816V *KIT* mutation
D. 40-year-old with childhood-onset cutaneous mastocytosis that persisted into adulthood; patient recently developed progressive splenomegaly and anemia; mutational analysis of bone marrow shows an exon 8 *KIT* mutation

Answer: D Patients with D816V *KIT* mutation regardless of the category of mastocytosis are resistant to therapy with imatinib. Patients with cutaneous or indolent systemic mastocytosis with symptoms manageable by symptomatic therapy are not candidates for cytoreductive therapy regardless of the mutational status. Patients with well-differentiated systemic mastocytosis without codon 816 (exon 17) *KIT* mutations have a high likelihood of response to imatinib. (Akin C, Fumo G, Yavuz AS, et al. A novel form of mastocytosis associated with a transmembrane c-kit mutation and response to imatinib. *Blood.* 2004;103:3222-3225.)

5. Which of the following mast cell mediator measurements raises the greatest suspicion for mastocytosis?

A. Tryptase level of 50 ng/mL in a patient with perioperative anaphylaxis, obtained 1 hour after the event
B. Baseline tryptase level of 90 ng/mL in a patient who experiences recurrent anaphylactic episodes of unclear etiology
C. Elevated urinary *N*-methylhistamine in a patient with recurrent unexplained flushing
D. Elevated urinary prostaglandin D_2 in a patient with chronic diarrhea and osteoporosis

Answer: B Urinary metabolites of mast cell mediators are neither more sensitive nor more specific than measurement of baseline tryptase in diagnosis of systemic mastocytosis. Tryptase levels increase (usually within 4 hours) after an anaphylactic event regardless of whether it is associated with mastocytosis. Some patients with recurrent unexplained anaphylaxis who may have been previously diagnosed with idiopathic anaphylaxis may have mastocytosis as an underlying disorder (especially if the anaphylactic episodes involve hypotension and syncope rather than urticaria or angioedema). The diagnosis should be confirmed by a bone marrow biopsy in the patient in option D. (Akin C, Scott LM, Kocabas CN, et al. Demonstration of an aberrant mast-cell population with clonal markers in a subset of patients with "idiopathic" anaphylaxis. *Blood.* 2007;110:2331-2333; and Akin C, Metcalfe DD. Surrogate markers of disease in mastocytosis. *Int Arch Allergy Immunol.* 2002;127:133-136.)

256 APPROACH TO THE PATIENT WITH RHEUMATIC DISEASE

VIVIAN P. BYKERK AND MARY K. CROW

Rheumatic diseases are common and an important cause of reduced quality of life, increased comorbidity, and reduced life expectancy. They incur a significant socioeconomic burden and warrant expertise on the part of all physicians who treat patients. This chapter provides a framework to approach the evaluation of patients who present with signs and symptoms suggesting a rheumatic disease. An algorithmic approach is provided that allows the physician to incorporate presenting features, patient characteristics, anatomic structures, along with diagnostic tests, to facilitate a diagnosis and treatment plan.

DEFINITION AND CATEGORIZATION

Rheumatic diseases are disorders of connective tissue in which general or localized inflammation frequently manifests as pain attributable to peripheral joints, the spine, or muscles. Systemic features such as stiffness, fever, or weight loss and a multitude of extramusculoskeletal features, ranging from skin rashes to renal dysfunction, often accompany rheumatic diseases. In most cases, the basic underlying pathology is understood, although this is less true of pain disorders appearing alone or accompanying a rheumatic disease (Chapter 30). For most rheumatic disorders, the underlying molecular mechanisms through which environmental triggers and genetic susceptibility factors collaborate to result in a particular rheumatic disease in an individual remain to be fully elucidated. These diseases can be broadly considered as those that are primarily degenerative, with inflammation occurring secondarily, and those in which inflammation is the primary mediator of the disease. In the latter, the pathogenesis can be mediated through aberrant immune responses or as a result of metabolic abnormalities.

Histopathology

Rheumatic diseases are often also termed *connective tissue diseases,* understandably, because connective tissue is the most abundant tissue in the body supporting and connecting other tissues and organs. Loose and dense connective tissues include cellular components and extracellular matrix. Loose connective tissue fills spaces between muscle sheaths, encases blood and lymphatic vessels, and holds fibroblasts that synthesize collagen fibers. It includes reticular fibers that provide the skeleton of muscle cells, nerves, and capillaries. Dense connective tissue supports the body's soft tissues and includes more collagen fibers and fewer cells. It is found in the dermis, joint capsules, cartilage, bone, and fascia of muscles, and it forms tendons, ligaments, and points of connection where these insert into bone (aponeurosis). Cells included in connective tissue may be wandering, such as mast cells or macrophages, or resident cells, such as fibroblasts, fibrocytes, and reticular cells. Fibroblasts are responsible for synthesizing collagen, elastic reticular fibers, and ground substance of extracellular matrix, including tissue fluids and collagen fibers. Importantly, connective tissue is integrated with cells associated with the body's defense system: lymphocytes, plasma cells, macrophages, dendritic cells, and eosinophils. The close proximity of connective tissue to blood vessels and cells of the immune system provides the setting for a group of disorders that are mediated by impaired immune system regulation and disruptions of the vascular system.

Classification of Rheumatic Diseases

More than 100 types of rheumatic diseases have been described. Although these can be considered to be based primarily on one of two degenerative or inflammatory overarching processes, one can further subdivide rheumatic diseases as follows (and also outlined in Table 256-1): (1) those associated with degeneration of connective tissues attributable to (a) trauma, (b) structural/mechanical imbalances, or (3) inherent early demise of cellular components; (2) those associated with systemic autoimmunity,[1] often linked with measurable autoantibodies that can manifest primarily with (a) synovitis, (b) widespread organ involvement, (c) inflamed blood vessels, or (d) inflammation of muscle; (3) other inflammatory connective tissue diseases involving more dense tissues, not associated with the formation of autoantibodies and hence termed *seronegative rheumatic diseases* or *spondyloarthropathies;* (4) diseases in which inflammation of the vasculature, particularly small, medium, or large arteries, is the predominant feature; (5) autoinflammatory diseases that can be associated with crystal deposition or genetic mutations involving cytokine pathways; and (6) pain syndromes that must often be considered in the context of these diseases, in which some appear to be comorbid and closely linked to the underlying rheumatic disease, such as diffuse pain associated with Sjögren's syndrome, hypermobility of connective tissue, or those regional pain syndromes that are anatomically linked to mechanical disruption. Patients presenting with generalized pain syndromes require investigation to exclude a connective tissue disease. Increasingly, genotypes have been identified that are associated with diseases that fall into each of these categories, and in some cases specific immunologic pathways have allowed grouping of a set of rheumatic diseases previously considered more distinct.

Mimics of rheumatic diseases exist, and clinicians need to be mindful of considering these when evaluating a patient for a rheumatic disease. For instance, arthropathies and syndromes resembling a rheumatic disease can occur in the settings of both infection and malignancy. Autoimmune phenomena are increasingly being recognized in the setting of malignancy.[2] Red flags for each need to be considered in the assessment of a patient for possible rheumatic disease.

No classification of rheumatic disease can completely explain its genesis. However, considering these in a classification schema can aid in the approach to a patient in whom these disorders are being considered (see Table 256-1).

EPIDEMIOLOGY

Although connective tissue diseases can generally be categorized as noted in Table 256-1, in adults there are six prototypical rheumatic diseases most often assessed and managed by rheumatologists: rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), systemic sclerosis (SSc), spondyloarthropathies (SpA) (primarily ankylosing spondylitis [AS]), Sjögren's syndrome (SS), and vasculitis. These diseases are ubiquitous throughout the world (Table 256-2) and for the most part have a similar incidence and prevalence throughout the globe. Each is associated with differing immune aberrations and mechanisms of inflammatory damage, although the cause and reasons for chronicity still remain unknown. Autoimmune rheumatic diseases are also among the leading causes of death and morbidity in the industrial world, in part related to associated comorbid diseases, particularly comorbid cardiovascular disease. They contribute to a significant socioeconomic burden. Increasing evidence points to risks for their genesis relating to environmental factors, socioeconomic factors, and exposure to infectious agents, ultraviolet radiation, and pollutants. Smoking, in particular, has been associated with an increased risk for SLE and RA in genetically susceptible individuals in Western cultures. Effects of migration elucidate some of these risks. For instance, Africans who migrate far away from their native environmental and cultural origins appear to have an increased susceptibility to SLE. Also, reports have linked occupational exposures, such as silica dust, mercury, pesticides, solvents, and metals, to an increased risk for SLE and RA.

In some cases, geographic clusters of rare autoimmune disease argue for specific genetic determinants. For example, with SSc, higher incidence, prevalence, and mortality rates have been reported in African American populations compared with white populations, and the prevalence has been reported as higher in southern Europe, particularly Italy (prevalence of 7 to 33 per 100,000). Additionally, social and demographic factors may contribute to the epidemiology of rheumatic diseases. For example, the prevalence of SLE is reported as very high in Georgia, United States, whereas the prevalence of AS is rare in malaria endemic regions where HLA-B27 genotypes are rare. Inflammatory arthropathies, including RA and AS, have a higher prevalence in North American Native populations.

CLINICAL MANIFESTATIONS

It is important to understand the potential clinical manifestations and natural history of rheumatic diseases. Primary care and hospital-based health care providers are often the first to evaluate a patient with an evolving rheumatic disease, and they need to be attuned to the presenting features to make a timely diagnosis. In many cases, the presentation could signal a life- or organ-threatening condition. Evaluation of constitutional, systemic and joint symptoms should always include rheumatic disease in the differential diagnosis.

TABLE 256-1 COMMON RHEUMATIC DISEASES BROADLY CLASSIFIED ACCORDING TO PATHOGENESIS

DEGENERATIVE DISEASES OF BONES AND JOINTS	SYSTEMIC AUTOIMMUNE DISEASES	SERONEGATIVE SPONDYLOARTHROPATHIES	VASCULAR RHEUMATIC DISEASES	AUTOINFLAMMATORY DISEASES	PAIN DISORDERS
Osteoarthritis	Rheumatoid arthritis	Ankylosing spondylitis	ANCA- associated vasculitis	Adult-onset Still's disease	Regional myofascial pain syndromes
DISH	Systemic lupus erythematosus	Psoriatic arthritis	Temporal artery vasculitis	Crystal diseases	Tendonitis/bursitis
Degenerative disc disease	Sjögren's syndrome	Reactive arthritis	Polymyalgia rheumatica	Pediatric periodic fever syndromes	Adhesive capsulitis
Spinal stenosis	Inflammatory myopathies (polymyositis, dermatomyositis)	Enteropathic arthritis	Behçet's disease		Reflex sympathetic dystrophy
Osteoporosis	Systemic sclerosis				Pain with hypermobility syndromes
					Fibromyalgia*

*The only pain disorder that has not been associated primarily with inflammation.
ANCA = antineutrophil cytoplasmic antibody; DISH = diffuse idiopathic sclerosing hyperostosis (also linked to metabolic factors, including elevated growth hormone).

TABLE 256-2 WORLDWIDE PREVALENCE* AND INCIDENCE OF RHEUMATIC DISEASES ASSOCIATED WITH AUTOIMMUNITY

DISEASE	NORTH AMERICA	CENTRAL AMERICA	SOUTH AMERICA	EUROPE	MIDDLE EAST	ASIA	SUB SAHARAN AFRICA	AUSTRALIA
RA	600-1000 (40)	400-2000	100-500	200-900 (2-7)	200-1500	100-800 (40-90)	Rare - 900	2000
SLE	20-60 (2-7)	50-60 (5)	N/A	20-70 (2-7)	N/A	20-70 (3)	Rare	20-80 (11)
SSc	13-28	N/A	N/A	<10-15 (<2)	N/A	<10	N/A	23 (2)
SpA	50-130 (7)	N/A	N/A	100-850 (2-9)	500	10-240	Rare	N/A
SS	320 (4)	N/A	N/A	200-600 (4-5)	N/A	330-700 (China)	N/A	N/A

*Prevalence (annual incidence) per 100,000 by world regions.
RA = rheumatoid arthritis; SLE = systemic lupus erythematosus; SpA = spondyloarthropathy (primarily ankylosing spondylitis); SS = Sjögren's syndrome; SSc = systemic sclerosis.
Data from Shapira Y, Agmon-Levin N, Shoenfeld Y. Geoepidemiology of autoimmune rheumatic diseases. *Nat Rev Rheumatol.* 2010;6(8):468-476; and Chaaya M, Slim ZN, Habib RR, et al. High burden of rheumatic diseases in Lebanon: a COPCORD study. *Int J Rheum Dis.* 2012;15(2):136-143.

Joint Symptoms as a Common Presenting Feature

Almost all rheumatic diseases can manifest with joint-related symptoms as a significant and frequently presenting feature. This can include symptoms of pain, stiffness, swelling, and erythema, as in the case of autoinflammatory diseases such as gout or pseudogout. The pattern of joint involvement, particularly duration and timing of maximal symptoms, can help the health care provider identify patients who present with spondyloarthropathy or any inflammatory arthritis, regardless of pathogenetic classification. For instance, joint pain that is worse in the morning, is associated with prolonged stiffness, and improves with activity is a classic presentation of inflammatory pain. On the contrary, pain that is worse with activity, better with rest, and associated with a very short period of stiffness signals that the category is degenerative. Location provides a clue to broad classification. Patients describing "pain all over" may have a primary pain syndrome. However pain with an inflammatory pattern localized to the spine or an enthesis (site of ligament insertion) is more likely to indicate a seronegative spondyloarthropathy. A patient noting an inflammatory pattern of symptoms in small joints of the hands and feet suggests the presence of one of the autoimmune rheumatic diseases associated with either a positive antinuclear antibody (ANA) or rheumatoid factor (RF). Thus, joint symptom patterns become central to the evaluation of any rheumatic disease.

Nonspecific Clinical Presentations

All of the rheumatic diseases can be associated with joint involvement. However, joint symptoms are not always present in many of these diseases. Thus, a working knowledge of other patterns of presentation is important. Fever or cutaneous manifestations, including rashes, are common in vasculitis and in the presentation of SLE. Sicca symptoms are pathognomonic of Sjögren's syndrome. Both Sjögren's syndrome and inflammation of blood vessels can occur concurrently in patients with SLE. Systemic features such as myalgias or fatigue are common to almost all rheumatic diseases regardless of their classification, whereas true weakness may be the only presenting complaint of an inflammatory myopathy. Renal involvement is common to seropositive systemic autoimmune diseases and vasculitis and can present with anasarca if proteinuria is severe or prolonged. Consequently, specific and nonspecific features associated with various connective tissue diseases must be identified to correctly develop a differential diagnosis that fits within the classification described in Table 256-1. Features may present sequentially over time; thus, rheumatic diseases from more than one category must often be considered in a patient whose disease has not yet been diagnosed, leaving the patient with a label of a nonspecific or undifferentiated connective tissue disease. Most rheumatic diseases have specific classification criteria. When these are not yet met, features are considered in the context of the broad classifications, and terms such as *undifferentiated inflammatory polyarthritis* or *undifferentiated spondyloarthropathy* may be used in the interim to aid in diagnosis and management. When a patient's disease is not yet diagnosed, investigations and monitoring over time, as indicated in Figure 256-1, can help to ultimately identify an emerging rheumatic disease.

Cutaneous Manifestations

Although cutaneous manifestations are frequent in patients with seropositive autoimmune diseases, particularly SLE, they are important manifestations of all autoimmune diseases. A nonblanching purpuric rash can indicate a vasculitis, and rashes involving specific extensor regions are common to dermatomyositis. In cases of SLE or dermatomyositis, rashes are triggered or worsened by exposure to ultraviolet light and occur in a light-exposed distribution. Rashes of vasculitic origin can indicate either the presence of an autoimmune disease such as SLE or an inflammatory vascular disease such as antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis. These tend to be present for days and are often palpable, and a tissue biopsy of the lesions will be very helpful in diagnosis. Rashes can be transient. In adult-onset Still's disease (an adult form of systemic inflammatory arthritis classified as an autoinflammatory disease), patients present with daily spiking fevers, peaking late in the day, associated with a salmon-colored evanescent blanching rash lasting only 1 to 2 hours. The specific location of a rash aids in diagnosis. A facial rash that spares the nasolabial folds is classic for lupus. A rash that does not spare the nasolabial folds suggests rosacea. Psoriasis is almost always present in psoriatic arthritis (a spondyloarthropathy variant). Although psoriasis is often widespread, involving extensor surfaces, it can be missed if there

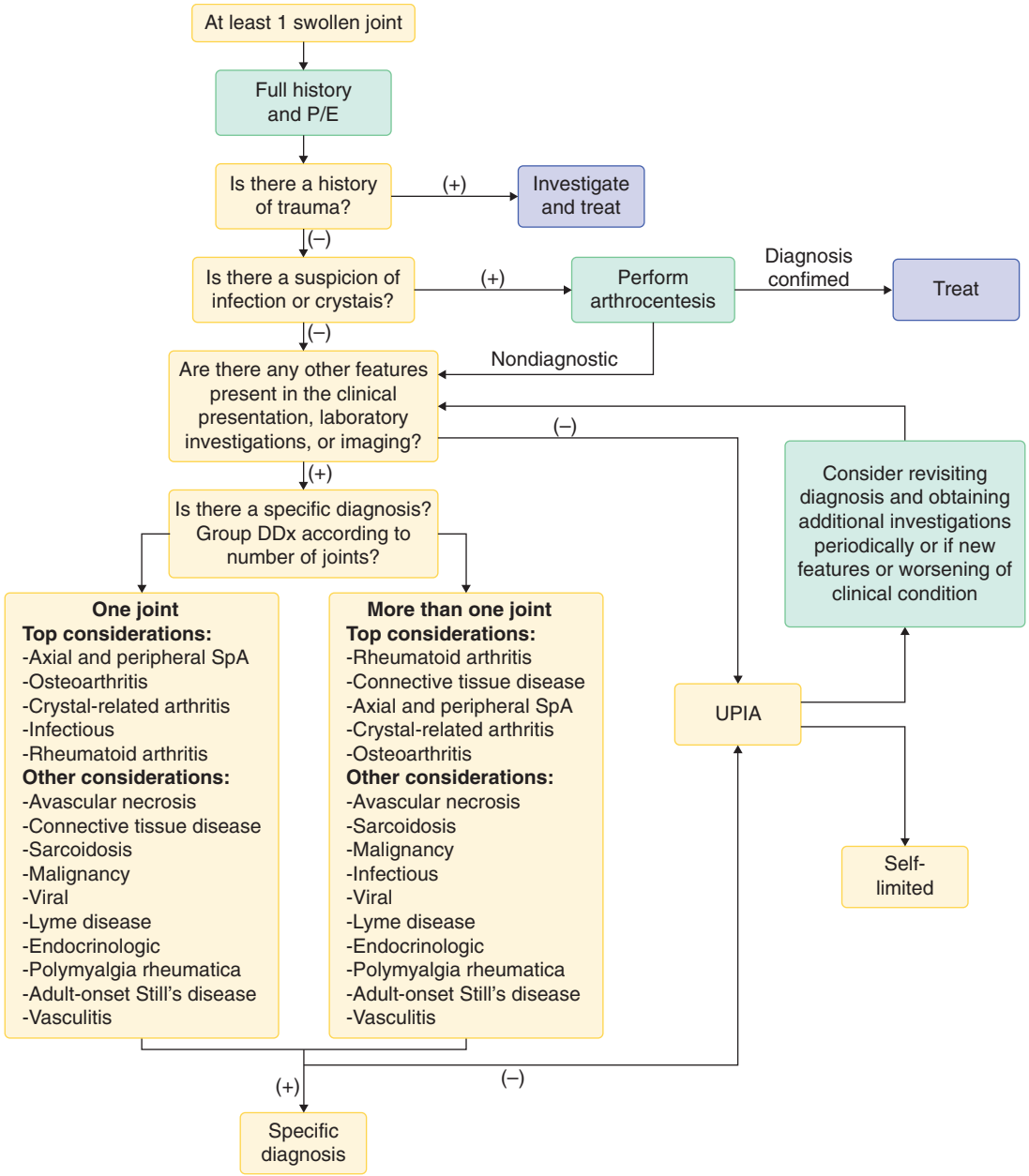

FIGURE 256-1. Algorithm for identification of undifferentiated peripheral inflammatory arthritis. Recommended minimum investigations in all patients: rheumatoid factor and/or anti–citrullinated peptide antibodies, erythrocyte sedimentation rate and/or C-reactive protein, complete blood count, and radiographs of affected joints. DDx = differential diagnosis; P/E = physical examination; SpA = spondyloarthritis; UPIA = undifferentiated peripheral inflammatory arthritis. (From: Hazlewood G, Alethaha D, Carmona L, et al. Algorithm for identification of undifferentiated peripheral inflammatory arthritis: a multinational collaboration through the 3e initiative. *J Rheumatol Suppl.* 2011;87:54-58.)

are very few lesions or if it is located in areas that are not easily seen (intertriginous regions, such as in the ear, umbilicus, buttock creases, or scalp) or if it only involves the nails. The rash of psoriasis can be seen in any of the variants of the spondyloarthropathies. Occasionally RA will manifest with nodules over the extensor surfaces or rashes in the same areas, but this is increasingly rare. However, the presence of nodules over extensor surfaces may also indicate gouty tophi. In SSc, distal tightening of the skin, presence of telangiectasias, and digital ulcers can be seen. These should also be sought in the context of Raynaud's syndrome, characterized by vascular spasm in the hands.

Pattern of Onset of Rheumatic Diseases by Category

Most rheumatic diseases appear spontaneously and often insidiously or with a subacute onset. Not all will manifest with all typical features at the time of presentation, and a diagnosis can take time to make as defining features present themselves. Almost all rheumatic diseases have had classification criteria published that account for the typical and specific features of the disease (these are detailed in other chapters). Although classification criteria are generally published to ensure homogenous recruitment of participants in research studies, they can also aid in diagnosis. However, a diagnosis can be made without meeting all classification criteria because some features of a rheumatic disease are highly specific and are not associated with other diseases. For example, SLE is the only disease that presents with a classic malar rash and a particular autoantibody, anti-Sm, in the serum.

The natural history of each rheumatic disease is related to the severity of presentation, the specific additional organs that become involved, and the development of comorbidities. RA may present with one or two swollen joints or pain in the forefoot. However, when the patient has a high titer of anti–cyclic citrullinated peptide antibody (ACPA), an autoantibody associated with development of erosive joint disease, the diagnosis can be made quickly and treatment initiated early. Autoimmune and vascular rheumatic diseases are associated with high morbidity and mortality when left untreated, and efforts to investigate for all manifestations early on are warranted.

Degenerative Rheumatic Diseases

These refer to rheumatic diseases commonly associated with advancing age. Degenerative joint disease (DJD) is usually thought to include osteoarthritis (OA) and degenerative disc disease (DDD). DJD is heralded by breakdown of articular collagenous structures (cartilage or intervertebral discs) and development of bony hypertrophy. Controversy remains as to which occurs

first. As collagenous structures degrade, associated inflammation commonly occurs. Resultant pain from varying causes contributes to immobility, secondary comorbidities, and disability. DJD represents, by far, the most common of the rheumatic diseases and is described in detail in Chapter 262. In general, these diseases are not associated with rashes or nonspecific constitutional symptoms.

Autoimmune Rheumatic Diseases

These diseases include SLE, RA, SSc, primary Sjögren's syndrome, idiopathic inflammatory myositis, and the systemic vasculitides. They involve multiple organ systems, are heterogeneous in clinical manifestations,[3] and are associated with and partially characterized by autoantibodies that can be measured in the serum (Tables 256-3 and 256-4). Many of these diseases affect females more than males. Most striking is SLE (Chapter 266), which affects 8 to 10 women for every 1 man and typically has its onset in the childbearing years. As noted, cutaneous manifestations, sicca symptoms (of dry eyes or mouth), mucosal ulceration, fevers, alopecia, and Raynaud's syndrome are common and often described by patients. These occur alone or in addition to symptoms affecting the joints and muscles. RA (Chapter 264) has a female-to-male ratio of approximately 3 : 1 and typically has its onset in the later adult decades, with symmetrical arthritis the classic presenting feature. Specific to SSc (Chapter 267), often referred to as scleroderma, are symptoms of skin tightness in the extremities. This is often associated initially with swelling of the digits, followed by tightness of the hands and ultimately tightness of the skin. In addition, skin tightness can involve the face, arms, and in the case of diffuse SSc, trunk, back, and legs. Although these diseases can present classically, facilitating rapid diagnosis, features can overlap. For instance, Sjögren's syndrome (Chapter 268), with prototypical sicca symptoms and autoantibodies, can occur concurrently in SLE and RA. Autoimmune vasculitides and SLE can share similar organ system manifestations, including those involving the lungs, kidneys, skin, and nervous systems. Most can have constitutional disturbance, arthralgia and arthritis, and myalgias. Overlap of symptoms expands the differential diagnosis, and a clear distinction is often not readily apparent, even after serologic testing. Tissue biopsy may help to provide a definitive answer.

Spondyloarthropathies

AS (Chapter 265) is more common in males than females. Localization of symptoms to the back, sacroiliac joints, and large lower extremity joints can be used to differentiate AS from RA. The typical presence of psoriatic skin lesions in patients with psoriatic arthritis can be a distinguishing feature from RA.

Forms of Vasculitis

There are many vasculitis syndromes, and these are generally grouped based on vessel size (Chapter 270). Some are classified as such though not always proved to be of vascular origin. Polymyalgia rheumatica (PMR) (Chapter 271) is a common inflammatory rheumatic disease of the elderly and shares many pathogenetic and epidemiologic features with giant cell arteritis (GCA), a form of older-onset vasculitis.[4] Patients complain of aching around the neck and bilateral involvement of the shoulder and hip girdles, along with significant stiffness that is most problematic in the morning. Symptom onset can be abrupt or insidious over weeks to months. The diagnosis of PMR is primarily clinical, although diagnostic criteria have been suggested, based on the typical clinical presentation and laboratory evidence of acute-phase reactants. Mimics of PMR can include elderly onset RA; thus, tests to exclude this may be indicated. More important, when considering PMR, always consider GCA, often involving inflammation of the temporal arteries, because the consequences of this disease can be damaging and severe, sometimes leading to blindness.

Autoinflammatory Diseases

Rare autoinflammatory diseases that are based on mutations in genes involved in inflammatory pathways are typically diagnosed in children and are covered in detail in Chapter 261. Gout (Chapter 273) is common in middle-aged and older men and may be increasing in prevalence. Exquisitely painful joint erythema and swelling are presenting features. Tophi that might be confused with rheumatoid nodules can be seen.

Pain and Pain Syndromes

Pain (also see Chapter 30) is a common and nonspecific, but very important, symptom central to nearly all rheumatic diseases. Pain is the key presenting feature of joint disease as perceived by the patient. In regional pain syndromes, the distribution of pain is the key clue to the diagnosis. Diffuse pain without evidence of underlying pathology associated with inordinate levels of fatigue, difficulty coping, and intricately detailed descriptions of pain using colorful analogies herald fibromyalgia (Chapter 274). Fibromyalgia is defined as widespread pain involving right and left sides and upper and lower extremities, as well as the neck and back. Most pain syndromes are regional pain syndromes. For instance, a regional pain syndrome relating to a mechanical neck and shoulder syndrome will result in the patient having pain in the involved neck and shoulder but also the trapezius, upper chest, and lower arm and hand. A large proportion of patients presenting with musculoskeletal pain will have regional pain relating to muscular imbalances with or without underlying degenerative arthritis, tendinopathy, or enthesopathy. As an

TABLE 256-3 CLUES TO DIAGNOSIS FOR EACH CATEGORY OF RHEUMATIC DISEASE

DEGENERATIVE DISEASES OF BONES AND JOINTS	SYSTEMIC AUTOIMMUNE DISEASES	SERONEGATIVE SPONDYLOARTHROPATHIES	VASCULAR RHEUMATIC DISEASES	AUTOINFLAMMATORY DISEASES	PAIN DISORDERS
Investigate if persistent symptoms >6 wk or failure of conservative measures (physical therapy, acetaminophen, NSAIDs). If at risk for bone loss and possible fragility fracture, consider osteoporosis.	Frequently associated with inflammatory joint pain and/or swelling with or without constitutional symptoms and other organ involvement. Perform ANA, RF, ESR, CRP.	Consider if psoriasis is present, or if back pain has inflammatory features; consider when known associated nonarticular features are present (e.g., uveitis, inflammatory bowel disease, urethritis, enthesitis, dactylitis).	Consider in all situations in which infarction of tissue has occurred or vasculitic rashes, pulmonary hemorrhage, or acute or subacute renal syndromes are present.	All can present with fever. In children, these are usually genetically determined syndromes mediated through interleukin-1 or TNF-α. In adults, consider crystal diseases.	Consider in situations in which pain is in excess of findings, with history of resolved trauma or repetitive strain, referred pain symptoms, diffuse pain, or colorful descriptions of pain.
INITIAL INVESTIGATIVE APPROACH TO CONFIRM SUSPECTED RHEUMATIC DISEASE IN EACH CATEGORY					
Image specific region giving rise to persistent pain (consider pain may be referred). If osteoporosis is a concern, perform BMD and investigate for metabolic bone diseases.	Investigations should specifically target suspected diseases, e.g., perform CPK if weakness or myalgia; anti-CCP (ACPA) if possible RA; ANA, dsDNA, C3, C4, ENA if possible SLE.	Imaging of sacroiliac (SI) joints (radiographs if long-standing symptoms, MRI if more recent onset). HLAB27 testing if inflammatory back pain or SI pain most predominant, clinical suspicion high for inflammatory back pain, but imaging is negative.	ANCA Anti-PR3 Anti-MPO Acute phase reactants (ESR, CRP) Tissue sampling of involved organs to facilitate pathophysiologic classification.	For suspected crystal disease, perform aspiration of synovial fluid or tophus and examine under polarized light microscopy.	Usually a diagnosis of exclusion. In reflex sympathetic dystrophy, tendonitis or enthesitis, specific physical findings and imaging can facilitate diagnosis.

ACPA = anti–citrullinated peptide antibodies; ANA = antinuclear antibodies; ANCA = antinuclear cytoplasmic antibodies; anti-CCP = anti–cyclic citrullinated peptide; anti-MPO = anti-myeloperoxidase; anti-PR3 = anti-proteinase-3; BMD = bone mineral density scan; CPK = creatinine phosphokinase; CRP = C-reactive protein; dsDNA = anti-double-stranded DNA; ENA = antibodies to extractable nuclear antigens (e.g., Ro, La, Sm, RNP, Scl70, Jo-1); ESR = erythrocyte sedimentation rate; MRI = magnetic resonance imaging; NSAIDs = nonsteroidal anti-inflammatory drugs; RF = rheumatoid factor; SLE = systemic lupus erythematosus; TNF-α = tumor necrosis factor-α.

TABLE 256-4 AUTOANTIBODIES ASSOCIATED WITH SPECIFIC FEATURES OF SEROPOSITIVE SYSTEMIC AUTOIMMUNE DISEASES

	ANTIBODY PREVALENCE	MAIN CLINICAL MANIFESTATIONS AND ASSOCIATIONS
SLE		
Double-stranded DNA	70-80	Kidney disease, skin disease
Nucleosomes	60-90	Kidney disease, skin disease
Smith	10-30	Kidney disease
Small nuclear ribonucleoproteins (spliceosomes, U1-RNP, 70kD, A, C)	15-25	Raynaud's syndrome, puffy fingers, myositis, and hypergammaglobulinemia
N-methyl-D-aspartate receptor	33-50	CNS lupus
Phospholipids (cardiolipin, β2 GP1, prothrombin)	20-30	Thrombosis, pregnancy loss, thickened heart valve disease, and livedo reticularis
α-Actinin	20	Kidney disease
Ribosomes P0, P1, P2	4-12	Hepatic, CNS manifestations (psychosis)
C1q	40-50	Kidney disease, associated with disease activity
SLE AND SJÖGREN'S SYNDROME		
Ro/SSA	30-40	Kidney disease in SLE in the absence of anti-La/SSB, skin disease in SLE and photosensitivity; congenital heart block and neonatal lupus erythematosus, sicca symptoms; subacute cutaneous lupus, hypergammaglobulinemia, leukopenia; interstitial nephritis and increased risk for non-Hodgkin's lymphoma in patients with Sjögren's syndrome
La/SSB	15-20	Congenital heart block and neonatal lupus erythematosus, sicca symptoms; photosensitivity, subacute cutaneous lupus erythematosus, hypergammaglobulinemia, leukopenia; increased risk for increased non-Hodgkin's lymphoma in patients with Sjögren's syndrome
α-Fodrine	46-100 in Sjögren's syndrome and 30 in SLE	Sicca symptoms
IDIOPATHIC INFLAMMATORY MYOSITIS		
Jo-1, Pl-7, Pl-12, OJ, EJ	20-30	Antisynthetase syndrome
Signal recognition particle	2-8	Necrotizing myopathy
Mi-2	8-12 in idiopathic inflammatory myositis, 15-20 dermatomyositis	Dermatomyositis and idiopathic inflammatory myositis
TRIM33	10-30 of dermatomyositis	Dermatomyositis, malignancy
U1-RNP/U2-RNP	8-15	Mixed connective tissue more commonly than system lupus erythematosus, systemic sclerosis, and undifferentiated connective tissue disorder
PM/Scl	12-16	Dermatomyositis and systemic sclerosis overlap, systemic sclerosis, and dermatomyositis
Ku	1-7	Myositis overlap, SLE, idiopathic inflammatory myositis, and systemic sclerosis
CADM-140/anti-MDA-5 antibody (clinically amyopathic dermatomyositis/antimelanoma-differentiation-associated gene 5)	Infrequent	Amyopathic dermatomyositis (53%) with interstitial disease
SYSTEMIC SCLEROSIS		
Centromere	15-40	Limited systemic sclerosis, pulmonary hypertension
Scl-70/topoisomerase	10-40	Diffuse cutaneous systemic sclerosis, pulmonary fibrosis
RNA polymerase III	5-25	Diffuse cutaneous systemic sclerosis, renal crisis, pulmonary hypertension
AUTOANTIBODIES WITHOUT DISEASE SPECIFICITY		
Rheumatoid factor	30-40, 90-95, and 10-20	SLE, Sjögren's syndrome, and myositis
Antibodies against proteasome subunits	40-60, 50-60, and 40	Idiopathic inflammatory myositis, SLE, and Sjögren's syndrome

CNS = central nervous system; SLE = systemic lupus erythematosus.
From Goldblatt F, O'Neill SG. Clinical aspects of autoimmune rheumatic diseases. *Lancet*. 2013;382(9894):797-808.

example, pain syndromes in the trapezial region, referring down the arm to the deltoid and even forearm, can be multifactorial and associated with a combination of muscular spasm, underlying degenerative arthritis in the cervical spine or rotator cuff impingement, not infrequently relating to repetitive activities. Pain in an extremity after trauma or surgery associated with a cold extremity is suggestive of reflex sympathetic dystrophy. The etiology of these will become apparent with a careful history of the pain, along with a medical history, physical examination, and exclusion of "red flags" or factors that indicate an underlying organic pathology. Pain in the setting of a history of malignancy should suggest the possibility of metastases. A tick bite may indicate prior Lyme disease. Most pain syndromes warrant a full medical evaluation before making a definitive diagnosis.

DIFFERENTIAL DIAGNOSIS AND DIAGNOSTIC EVALUATION:

A comprehensive history is needed to complete an evaluation of a patient with a rheumatic disease. In addition to considering age and gender, a patient's personal history, including marital status, occupation, and psychosocial factors, helps to elucidate a patient's diagnosis, prognosis, and treatment options. Clues helpful in making a diagnosis of one of the rheumatic diseases are summarized in Table 256-3.

Assessing clinical signs and symptoms is the cornerstone of diagnosis. Most rheumatic disorders will present with symptoms that involve, or seem to involve, joints. This can be limited to pain involving a specific joint or group of joints or periarticular structures. Querying the patient to determine the pattern of symptoms—whether pain, swelling, or stiffness associated with

joints—is key to narrowing the differential diagnosis of a rheumatic disease. Joint symptoms may have inflammatory features such as prolonged stiffness, pain at rest, or noninflammatory and mechanical features, such as instability or giving way, locking, or increased symptoms with use. Patterns of joint involvement—whether primarily small joints of the hands, wrist, and feet; large joints of the elbows, knees, ankles, or "root" (shoulders or hips); or spinal involvement—will help to narrow down the possible diagnoses. Moreover, questions regarding recent illnesses, exposure to possible infectious pathogens, and presence or absence of systemic features such as fever, fatigue, or weight loss will provide important clues. Appreciation of signs and symptoms indicating extra-articular features, particularly cutaneous, pulmonary, renal, neurologic, or vascular manifestations, will not only help to facilitate a diagnosis but aid in prognosis and an understanding of the needed intensity of therapy.

The presence of one rheumatic disease can be associated with comorbid manifestations unrelated to connective tissue. For example, myocardial infarction is more common in many patients with rheumatic diseases. Immobility or treatment-related factors leading to obesity can increase risks for diabetes and lower joint degeneration.

Factors in the Medical History that Contribute to Diagnosis and Prognosis

Age and Gender

Certain rheumatic diseases typically manifest in childhood. These include genetically based disorders such as hemophilia, associated with arthritis, and a number of autoinflammatory conditions that are by definition childhood diseases. Juvenile idiopathic arthritis refers to forms of arthritis in which the onset occurs before the age of 16 years. Autoimmune rheumatic diseases, and inflammatory rheumatic diseases such as spondyloarthropathies, RA, and SLE, can begin in young adulthood, whereas degenerative conditions such as OA rarely do and more often begin to manifest in the middle and late middle years. The peak onset of RA occurs in the late middle years, although onset can occur at almost any time in life. Elderly people are more prone to OA and PMR, but the latter has a wide differential diagnosis and should be considered at all ages. Autoimmune diseases are more common in women, whereas spondyloarthropathies can be equally common in men and women. Gouty arthritis is more common in men and rarely attacks women before menopause.

Occupation and Recreation

Occupation and recreational activities may give rise to physical and psychological stresses. The demands of a patient's occupation need to be understood, particularly when repetitive activities may contribute to the development of DJD or to regional pain syndromes. Similarly, trauma from sports, including prior injuries, can be significant contributors to DJD.

Family History

It is important to obtain a complete family history because autoimmune diseases, spondyloarthropathies, and gout occur with an increased incidence in families. It is common to see family pedigrees in which different forms of autoimmunity occur throughout a family. This does not mean that any one autoimmune disease has specific heritability. Also, generalized OA that involves the hands and other joints commonly runs in families.

Concomitant Medication Use

Concomitant medications may contribute to the genesis of a rheumatic disease. For example, diuretics can increase hyperuricemia and risk for gouty arthritis. Minocycline can be associated with lupus-like presentations. Antibiotics in the fluoroquinolone class have been associated with enthesopathies. A full medication history needs to be considered in the assessment of patients with rheumatic diseases.

Habits and Social Circumstances

Smoking has increasingly been associated with RA and SLE. Also, poor socioeconomic circumstances and psychosocial or physical stress may contribute to the severity of symptoms and should be considered in planning management strategies. Similarly, patients of different ethnic and cultural origins may have differences in their ability to describe symptoms and in preferences around treatment choices.

Onset and Evolution of Symptoms

Knowledge of the pattern of onset, location, and evolution of symptoms is essential to make an accurate diagnosis of a rheumatic disease. Symptoms that develop over hours to days typically suggest an inflammatory, or possibly an infectious or traumatic, process. When they persist for more than 6 weeks, symptom onset is considered subacute, and the disease chronic. Early in the presentation of some rheumatic diseases, the symptoms can be intermittent or palindromic before becoming constant. Sudden onset of joint pain and swelling, particularly involving one or a few joints, should be considered to be due to an infectious or crystalline etiology during the course of investigation.

Pain and Stiffness

Pain assessment should include a description of its onset, constancy/chronicity, severity, quality, factors that trigger or improve it, and location and radiation of the pain. Stiffness, often described as tightness or linked to difficulty with movement or function, should be determined in terms of location (e.g., is it in a particular joint or more diffuse) and timing and duration (e.g., occurring after a period of rest). Stiffness that resolves in 10 to 15 minutes is more characteristic of OA. In inflammatory disease, stiffness typically lasts longer, often at least 1 hour and even all day.

Joint Involvement

The distribution of joint involvement is key to making a diagnosis of a rheumatic disease. Monoarthritis describes symptoms in a single joint; oligoarthritis (or pauciarthritis) refers to symptoms in two to four joints; and polyarthritis indicates involvement of at least five joints. Peripheral arthritis involves an extremity, whereas spinal involvement is termed *axial disease*. Symmetrical as opposed to asymmetrical peripheral joint disease is more commonly associated with autoimmune rheumatic disorders, whereas asymmetrical arthritis can be associated with spondyloarthropathies or OA. Similarly, predominant small joint involvement is more typical in RA or SLE, whereas large joint involvement is classic for spondyloarthropathies. In addition the presence of associated enthesitis and axial symptoms herald spondyloarthropathy. Joint or spine symptoms associated with inflammatory causes often include predominance of symptoms in the morning, associated with stiffness for more than 60 minutes, worsening with rest, and improvement over the day and with activity. Joint or spine symptoms associated with degenerative joint disease typically worsen with activity, are often worse later in the day, and associated with stiffness, and typically resolve quickly over 15 to 30 minutes. Joint pain is usually felt at the joints (exceptions include shoulder pain, felt over the deltoid, and hip pain, felt in the groin). Joint pain from degenerative or inflammatory causes can vary in severity. Most patients will describe joint pain as aching and rarely rate the pain higher than 8/10 on an ascending severity scale. Pain relating to localized myofascial pain syndromes, including tendinopathies and enthesopathies, may be described as being close to joints, and being worse with specific movements. In people suffering from generalized pain syndromes, pain is often rated very highly (10/10), is poorly localized, involving upper and lower body regions, with descriptions including qualifiers to impress the severity of the pain ("like a truck ran over me").

It is also important to distinguish between arthralgia (subjective joint pain without objective signs) and arthritis, where pain and tenderness are associated with objective signs of joint swelling and warmth (synovitis), deformity, or limitation of movement. Objective findings on physical examination must be identified for a diagnosis of arthritis to be made.

Function

Function is commonly compromised in patients with rheumatic disease. Although this can be related to fatigue or muscular weakness, in the case of rheumatic disorders in which there is no articular involvement, most commonly functional impairment is related to joint involvement. Function should be assessed in terms of a patient's ability to perform activities of daily living, work, and participation. Validated questionnaires of function are available to identify functional limitations.

Physical Examination

Essential Concepts

A complete examination by a physician is required to identify and classify a rheumatic disease. This should include an assessment of temperature, body mass index, affect, pain behaviors, gait, and posture, as well as examination of the scalp, skin, eyes, lymph nodes, cardiovascular system, lungs, abdomen, joints, spine, and skeletal muscles. A systematic joint examination is key to the rheumatic disease examination and should include all regions, with comparisons of right and left sides. The pattern of joint involvement, including symmetry, and axial versus peripheral involvement, should be recorded. Use

of a joint diagram (homunculus) helps to track joint involvement. The joint examination should include documentation of the presence or absence of tenderness, periarticular wasting, erythema, swelling, limitation in range of motion (ROM), sites of prior surgery and trauma, and joint deformity, allowing for comparison over time and between different examiners. Using a four-step systematic approach to joint examination facilitates a thorough examination. This should include (1) inspection (looking for asymmetry, erythema, swelling, and deformity), (2) palpation (feeling for tenderness, specifically joint line tenderness, warmth, synovial thickening and effusion, bony hypertrophy, and crepitus), (3) ROM (both actively and passively for each joint), and (4) special tests specific to each joint or region. A complete examination should also consider relevant possible extra-articular manifestations.

Examples of Musculoskeletal Findings that Help to Classify a Rheumatic Disease

When considering the joint examination, a finding of redness (erythema) can indicate more acute and/or severe inflammation. The presence of erythema is more typically seen in the case of infection or crystalline arthritis. Joint warmth also signifies underlying inflammation. Joint swelling, a definitive sign of joint inflammation or arthritis, may indicate the presence of a joint effusion (excess synovial fluid) representing inflammation of the synovial membrane (synovitis). All rheumatic diseases short of specific pain syndromes can present with synovitis. Palpable bony thickening around a joint indicates swelling related to osteophytes, which is characteristic of OA. Crepitus feels like a grinding sensation under the examiner's hand during active or passive joint motion. Fine or velvety crepitus may indicate chronic proliferative synovitis, whereas coarse crepitus may indicate either roughening of the cartilage surface or complete loss of hyaline cartilage. Joint damage may manifest as loss of cartilage or bony hypertrophy. In some diseases such as RA, psoriatic arthritis, or Jaccoud's arthropathy (a form of SLE-related deforming arthropathy) deformity occurs as a result of joint subluxation or contracture related to nature's forces on joints and tendons and where chronic synovitis has caused joint capsule distension, ligamentous laxity, tendon rupture, or contracture.

Assessing Range of Motion

Both active and passive ROM should be assessed to appreciate joint function. Generally, active ROM is assessed first by asking patients to demonstrate full ROM of a joint; active ROM requires intact strength, innervation, muscle and tendon function, and joint mobility. Passive ROM is assessed by the examiner and for the most part assesses joint mobility or in some cases ligament or tendon impingement. First, assessing active ROM enables the examiner to appreciate potential areas of pain and where to examine carefully. If mobility is full on passive ROM, other causes can be considered for loss of mobility.

Establishing a Diagnosis

The findings from a thorough history and physical examination can be used to guide an appropriate set of investigations. Table 256-3 and Figure 256-1 provide clinical clues that are helpful in establishing a diagnosis based on the broad classifications in Table 256-1. By using the algorithm and Table 256-3, clinicians can develop a more parsimonious differential diagnosis and use investigations that will help to confirm the suspected diagnosis.

Laboratory and Imaging in Rheumatic Diseases

Identifying the presence of specific laboratory or imaging features can support a diagnosis and aid in specific classification of a rheumatic disease.[5] For example, in vasculitis, testing ANCAs or imaging the vascular tree of an involved area is key to establishing a diagnosis. Similarly, in seronegative spondyloarthropathies, imaging of the sacroiliac joint using radiographs when symptoms are sustained, or MRI if the disease is of more recent onset, is key. Radiographs and MRIs of specific joint areas in DJD not only help establish diagnosis but also aid in staging the disease. In the case of RA, testing of rheumatoid factor and ACPA is critical. When considering other prototypical seropositive systemic rheumatic diseases, the list of potential autoantibodies is much longer, and depending on the degree of difficulty in making the diagnosis, or appreciating the extent of disease and associated organ involvement, many of these can be considered for testing. These are listed in Table 256-4. It should be noted that no test alone should be used to diagnose a rheumatic disease, but rather that the tests should support the diagnosis.

TREATMENT

Rx

The treatment approaches to each rheumatic disease are highlighted in detail in the following chapters. Therapeutic approaches to degenerative rheumatic diseases continue to involve control of symptoms with either nonsteroidal anti-inflammatory drugs or analgesics. Additionally, physical modalities and injections of glucocorticoids (GCs) or other agents are employed to manage symptoms. When these conservative approaches fail, orthopedic surgery is often indicated. However, the approach to systemic and inflammatory rheumatic diseases usually requires more intense or immune-modulatory therapies. GCs are a major component of treatment regimens, particularly when organs are at risk for damage, when other agents take time to become fully effective, and in situations in which there are no alternative options for therapy. However, GCs are not without risk, and these should be discussed with each patient. Depending on the disease, GC-sparing therapies are usually initiated, the potency of which is tailored to the severity and risk for the illness itself. For example, SLE patients presenting with only a rash or synovitis may receive hydroxychloroquine, whereas patients with renal disease may receive mycophenolate mofetil, cyclophosphamide, or other therapies to most effectively treat that disease manifestation. Similarly in RA, patients presenting with high-titer ACPA levels, a high swollen joint count, and erosions on baseline radiographs of the hands and feet may receive rapidly escalating doses of methotrexate with or without other disease-modifying antirheumatic drugs and earlier use of tumor necrosis factor inhibitors. Specific approaches and GC-sparing therapies will be outlined in chapters addressing each rheumatic disease. In almost all rheumatic diseases, GCs are used as transitional therapy with a view to tapering these and using them again only for disease exacerbations.

SUMMARY

A broad-based set of rheumatic disease classifications can provide an overall construct for consideration of a multitude of possible rheumatic diseases. When the classifications are based on pathogenic mechanisms as well as clinical features, this facilitates identification of specific symptoms and signs and guides a further line of investigation. Although all rheumatic diseases will not fit within this classification, a systematic and directed analysis will accelerate achieving the correct diagnosis of the patient.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Wahren-Herlenius M, Dörner T. Immunopathogenic mechanisms of systemic autoimmune disease. *Lancet*. 2013;382:819-831.
2. Shah AA, Casciola-Rosen L, Rosen A. Cancer-induced autoimmunity in the rheumatic diseases. *Arthritis Rheumatol*. 2015;67:317-326.
3. Goldblatt F, O'Neill SG. Clinical aspects of autoimmune rheumatic diseases. *Lancet*. 2013;382:797-808.
4. Nesher G. Polymyalgia rheumatic: diagnosis and classification. *J Autoimmun*. 2014;48-49:76-78.
5. Binder A, Ellis S. When to order an antinuclear antibody test. *BMJ*. 2013;347:f5060.

REVIEW QUESTIONS

1. On physical examination, which of the following categories of rheumatic diseases, when active, is *not* typically associated with inflammatory signs of synovitis?

A. Systemic autoimmune diseases
B. Degenerative diseases of bones and joints
C. Pain syndromes
D. Seronegative spondyloarthropathies
E. By definition, they are all associated with synovitis.

Answer: C Signs of synovial inflammation (synovitis), including erythema, warmth, and swelling, are typically not found in the pain disorders like fibromyalgia, reflex sympathetic dystrophy, adhesive capsulitis, and regional myofascial pain syndromes. It is important to distinguish between *arthralgia,* which is subjective joint pain without objective signs, and *arthritis*, in which pain is associated with objective signs of tenderness, swelling, and warmth (synovitis); the latter is sometimes also associated with deformity or limitation of movement.

2. A 65-year-old woman presents with complaints of increasing pain in her fingers and hands. She states that she has morning stiffness in those regions that tends to resolve within 10 to 15 minutes after start of activity. There are no other rheumatic or systemic symptoms and physical examination of the hands is completely normal. Which of the following is the most likely clinical diagnosis?

A. Osteoarthritis
B. Rheumatoid arthritis
C. Fibromyalgia
D. Scleroderma/systemic sclerosis
E. Systemic lupus erythematosus (SLE)

Answer: A Morning stiffness that resolves in 10 to 15 minutes is most characteristic of osteoarthritis. The morning stiffness of rheumatoid arthritis and other inflammatory rheumatic diseases typically lasts longer: at least 1 hour and even for much of the day. Hand involvement with scleroderma/SSc is characteristically associated with initial swelling of the digits, then progressing to skin tightness, which was not observed in this patient. Fibromyalgia involves widespread body pain of both sides, including upper and lower extremities, back and neck. This would be an unusual age for presentation with SLE, which typically begins in women in the age range of 15 to 45 years.

3. Which of the following would *not* be considered an important epidemiologic factor in the differential diagnosis of a patient with rheumatic disease?

A. Occupation and recreational activities
B. Age and gender
C. Concomitant medication use
D. Family history
E. Global geographic origin

Answer: E Connective tissue diseases are ubiquitous around the world and for the most part have a similar incidence and prevalence throughout the globe (see Table 256-1). In contrast, the clinical onset of various rheumatic diseases is often characteristic of certain age groups (e.g., ranging from autoinflammatory diseases typically presenting in childhood to polymyalgia rheumatica or osteoarthritis presenting in older age groups) and gender (with increased prevalence of autoimmune diseases in women and gout in men). A variety of medications taken chronically can be linked to specific rheumatic syndromes. Positive family histories are not uncommon in patients with autoimmune and other rheumatic disorders.

257

LABORATORY TESTING IN THE RHEUMATIC DISEASES

DAVID S. PISETSKY

The rheumatic diseases are a heterogeneous group of conditions that result from diverse pathophysiologic mechanisms and involve the musculoskeletal system as well as other organs. These conditions range from mild, diffuse joint and muscle pain to severe life-threatening kidney failure and stroke. Although the rheumatic diseases have many origins, immune disturbances resulting in local and systemic inflammation are frequently the underlying basis for disease manifestations. The approach to diagnosis therefore entails a wide array of laboratory tests to assess functional disturbances of individual organs and their relationship to inflammation and autoimmunity.

Laboratory testing in patients with rheumatic disease involves determination of biomarkers of the following kinds: antecedent (risk for disease); screening (subclinical disease); diagnostic (overt disease); staging (disease severity or activity); and prognostic (disease course, response to therapy, monitoring therapy). Some tests are useful in all contexts, although others have more specific uses related to their performance characteristics, specificity, and pattern of expression during disease. In view of the increasing efficacy

of treatment for diseases such as rheumatoid arthritis (RA) and the availability of more specific serologic assays for early diagnosis, laboratory screening may be important to improve outcomes further by identifying individuals who have symptoms (e.g., arthralgias) that could represent the first manifestations of disease (preclinical autoimmunity); serologic screening could also be useful in identifying individuals at risk for disease (e.g., siblings or first-degree relatives). In general, the assessment of damage relies on specific tests of end-organ function or structure rather than process markers. Distinguishing activity from damage is important in patient management, especially with respect to the use of therapies associated with toxicity. In many diseases, prognosis can reflect ongoing disease activity as well as damage from past disease activity and effects of treatment, with laboratory testing needed to assess these various processes.

MARKERS OF INFLAMMATION

For many patients, the initial goal of evaluation is to determine the presence of inflammation. Inflammation is the body's response to injury and is characterized by a cascade of cellular and molecular events that arise irrespective of stimulus or locale (Chapter 48). The immediate response to inflammatory stimuli is termed the *acute phase response* and includes a set of proteins produced primarily in the liver in response to cytokines such as interleukin-6 (IL-6), tumor necrosis factor-α (TNF-α), and IL-1. These cytokines are produced by macrophages and dendritic cells after stimulation of pattern recognition receptors (PRRs) that include both toll-like receptors (TLRs) and other non-TLR sensing systems that trigger a system called the *inflammasome*; PRRs recognize intracellular as well as extracellular bacterial and viral products as well as large and small (e.g., adenosine triphosphate, uric acid) molecules released from damaged cells. The result is stimulation of innate immunity (Chapters 46 and 48). Many proteins in the acute phase response show large increases in serum levels, although some show a reduction. Because the levels of these proteins can increase dramatically in magnitude, they provide a sensitive and powerful set of markers for inflammation, whether induced by infection, trauma, or autoimmunity.

Of the proteins stimulated during the acute phase response, C-reactive protein (CRP) has received the most attention as a marker of inflammation in both rheumatic and nonrheumatic diseases. CRP is a member of the pentraxin family; although its function is not fully known, its ability to bind to phosphocholine suggests a scavenger function to eliminate bacterial products or damaged cells and to attenuate the consequences of infection or tissue injury.[1] Other molecules, such as serum amyloid protein (SAP), fibrinogen, and complement, also show marked elevations in levels during the acute phase response, signifying a broad-based effort at host defense.

The CRP level provides a very useful measure of inflammation and can convey information for categorization of a clinical process (e.g., inflammatory versus noninflammatory arthritis) as well as assessment of disease activity or prognosis (e.g., activity of RA or likelihood of joint erosion). The advantage of measuring CRP in the blood, rather than cytokines, is that the protein levels are much higher. Furthermore, CRP levels remain elevated for a longer period (days) than do cytokines; the latter may appear only transiently in the blood and thereby evade detection. The levels of CRP are in part determined by genetic factors, with baseline values important in determining the significance of any increases that can be associated with disease activity. Although CRP testing is commonly performed to assess the risk for atherosclerosis, a frequent complication of systemic rheumatic disease, the application of this screening in a patient with an inflammatory condition must take into account the various determinants of this marker, especially during active disease.

Another simple laboratory test reflecting the acute phase response is the erythrocyte sedimentation rate (ESR). In this test, commonly called the *sed rate*, anticoagulated blood is drawn into a long, thin tube and allowed to settle under the influence of gravity for 1 hour. The distance the blood falls depends on a number of factors, including the concentration of serum proteins such as immunoglobulins and fibrinogen, an acute phase reactant. The sedimentation rate is nonspecific with respect to disease association and also depends on the age and gender of the person. The upper limits of normal vary between women and men. Other simple laboratory tests point to an acute phase response. For example, patients with inflammation frequently exhibit a leukocytosis or thrombocytosis, most likely reflecting the action of cytokines and other mediators, including glucocorticoids, during this process. With chronic inflammation, anemia of chronic disease can also occur, with the hematocrit, in conjunction with the white blood cell and platelet counts, pointing to the presence of an inflammatory process. In this regard, in systemic lupus erythematosus (SLE), lymphopenia, thrombocytopenia, and low CRP values often characterize active disease, with the discordance between laboratory and clinical findings a clue to diagnosis.

LABORATORY EVALUATION OF MUSCULOSKELETAL DISEASE

The most common presentation of musculoskeletal disease is pain in and around the joints in association with functional impairment. Collectively, diseases causing joint symptoms are called arthritis, implying inflammation. The extent of inflammation in these diseases varies markedly, however, with some forms such as osteoarthritis (Chapter 262) showing only limited evidence of inflammation either locally or systemically.

Arthritis results from many different diseases and occurs in various patterns defined by the number and size of joints affected, symmetry, and involvement of the axial as well as peripheral joints. For each pattern (e.g., chronic polyarthritis), a key issue in diagnosis concerns its place in the spectrum of inflammatory versus noninflammatory arthritis. Furthermore, although many diseases can cause arthritis, their prevalence varies enormously, with osteoarthritis or degenerative joint disease being the most common form of noninflammatory arthritis and RA the most common form of inflammatory arthritis.

The differential diagnosis of arthritis is based on a comprehensive history and physical examination to assess symptoms suggesting inflammation (e.g., morning stiffness and fatigue), the presence of synovitis, and results of laboratory tests indicative of inflammation. Of these tests, the ESR and CRP are nonspecific indicators of inflammation. Depending on the stage of disease and prior therapy of the patient, however, both the CRP and ESR may not be elevated at the time of an initial evaluation because many treatments, especially those directed against cytokines, can effectively reduce the acute phase response; multiplex assays of various other mediators may allow assessment of disease activity in this circumstance. Two autoantibody tests, rheumatoid factor (RF) and antibodies to citrullinated proteins, provide more specific diagnostic information. Given the demographics of inflammatory arthritis, testing for antinuclear antibodies is often part of this evaluation as well.

Rheumatoid Factor

RF comprises a family of specificities that bind to the immunoglobulin G (IgG) molecule. These RFs target primarily the constant region or Fc portion of IgG, reacting with antigenic determinants that are most likely conformational in origin. IgM RFs are the most abundant of these antibodies and have been easiest to measure, using agglutination assays with red blood cells or latex beads coated with IgG. More recently, enzyme-linked immunosorbent assay (ELISA) and nephelometry have been used to detect RFs.

RFs occur in approximately 80% of patients with RA (Chapter 264) and represent one criterion for the classification or diagnosis of this disease. Furthermore, high levels of RFs are often associated with a worse prognosis, the occurrence of joint erosion as measured by radiographs, and deformity. Despite these associations, RFs occur in the sera of patients with a wide range of autoimmune and inflammatory diseases as well as in normal individuals, especially with age (Table 257-1). The frequent occurrence of RFs may reflect their etiology and role in innate immune responses to promote the binding of IgG to antigen by Fc cross-linking. Although RFs may occur in many settings other than RA, depending on the pretest probability, the test nevertheless remains useful in the evaluation of patients with inflammatory arthritis.

Antibodies to Citrullinated Proteins

Antibodies to citrullinated proteins are other autoantibody specificities important in the diagnosis of RA (Chapter 264). Citrulline is a post-translational modification of the amino acid arginine that results from deimidation. This chemical reaction is catalyzed by the enzyme peptidylarginine deiminase (PAD) and may occur in the setting of inflammation; the function of this modification is unknown. Citrullination can affect many different proteins, creating antigenic sites on proteins that include vimentin, enolase, and filaggrin.[2]

Although antibodies are directed to citrullinated residues on intact proteins, they can be conveniently measured using synthetic peptides containing citrulline. Among these synthetic antigens, a citrulline-containing protein with a cyclic structure provides sensitive and specific assays in an ELISA format. Antibodies directed to this type of antigen are known as anti-CCP (cyclic citrullinated peptide) and can be formally distinguished from antibodies to the citrullinated proteins themselves (ACPA, or anti–citrullinated protein antibodies); The term *anti-CCP* is commonly used for these

TABLE 257-1 RHEUMATIC DISEASES AND NONRHEUMATIC CONDITIONS ASSOCIATED WITH A POSITIVE RHEUMATOID FACTOR

DISEASES	FREQUENCY
Rheumatoid arthritis	50-90%
Systemic lupus erythematosus	15-35%
Sjögren's syndrome	75-95%
Systemic sclerosis	20-30%
Polymyositis/dermatomyositis	5-10%
Cryoglobulinemia	40-100%
Mixed connective tissue disease	50-60%
Aging (>70 yr)	10-25%
Infection	
Bacterial endocarditis	25-50%
Liver disease	15-40%
Tuberculosis	8%
Syphilis	Up to 13%
Parasitic diseases	20-90%
Leprosy	5-58%
Viral infection	15-65%
Pulmonary disease	
Sarcoidosis	3-33%
Interstitial pulmonary fibrosis	10-50%
Silicosis	30-50%
Asbestosis	30%
Miscellaneous diseases	
Primary biliary cirrhosis	45-70%
Malignancy	5-25%

Modified from Shmerling RH, Delbanco TL. The rheumatoid factor: an analysis of clinical utility. *Am J Med.* 1991;91:530.

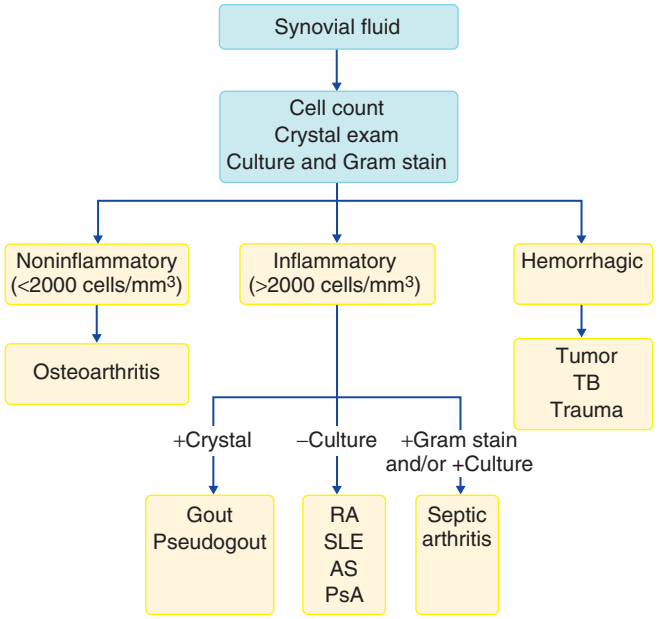

FIGURE 257-1. Algorithm for analysis of joint fluid. Examples of inflammatory arthritis are indicated, although many conditions can produce these findings. AS = ankylosing spondylitis; PsA = psoriatic arthritis; RA = rheumatoid arthritis; SLE = systemic lupus erythematosus; TB = tuberculosis.

specificities, although it is not formally synonymous with ACPA. ACPA can be assessed by a variety of analytic techniques using as antigens both modified proteins as well as arrays of peptides. For detection of anti-CCP antibodies, the formulation of peptides has changed over the years as designated by assay generation. Furthermore, among commercially available assays, results of assays can vary, making it important to know the performance characteristics of assays in interpreting the results of testing.

Anti-CCP antibodies are highly associated with RA and represent a criterion in the classification of patients with this disease.[3] Depending on the assay, these antibodies occur in 60 to 70% of patients with RA and uncommonly in those with other forms of inflammatory arthritis, making their presence important in diagnosis. Significantly, anti-CCP antibodies can occur before the onset of other signs and symptoms of RA, suggesting utility for screening of at-risk patients. In addition, in patients with arthalgias without clinical evidence of synovitis by examination, the presence of anti-CCP may predict the development of subsequent arthritis. Thus, because of the specificity of anti-CCP for RA, the presence of these antibodies in patients with early signs and symptoms of disease may indicate the diagnosis of RA and allow more prompt initiation of therapy before disease is fully manifest. In this regard, although RA can occur in the absence of anti-CCP antibodies, the presence of these antibodies may define disease subsets that differ in etiology, clinical course, and response to therapy.

Joint Fluid Analysis

Analysis of joint fluid can provide decisive data in the evaluation of arthritis and, in some instances, a definitive diagnosis. This analysis is essential in the setting of acute monoarthritis to investigate the possibility of infection; for chronic forms of arthritis, joint fluid should be analyzed if there is uncertainty about the diagnosis and involvement of one joint out of proportion to others. Joint aspiration is a sterile procedure performed with a local anesthetic. Although fluid can be analyzed by tests to assess viscosity and mucin content, the cell count, examination of crystals, and stains and cultures to evaluate infection are the most informative.

On the basis of cell counts, joint fluids can be categorized into four main types: noninflammatory, inflammatory, septic, and hemorrhagic. A noninflammatory fluid has fewer than 2000 cells/mm^3 with mononuclear cell predominance. An inflammatory fluid has more than 2000 cells/mm^3, with 50,000 cells/mm^3 frequently used as the upper limit for this type of fluid. In an inflammatory fluid, polymorphonuclear cells predominate. A septic fluid is an inflammatory fluid in which culture or staining for microorganisms demonstrates infection. Suspicion of infection is especially high for fluids with cell counts greater than 50,000/mm^3. However, crystal-induced arthritis can produce cell counts of this magnitude, and an infected fluid can have counts below this level. Hemorrhagic fluids have red cell predominance that can approximate that of blood.

In the setting of an acute monoarthritis, crystal-induced disease is much more common than infection, with the presence of crystals demonstrated by polarization microscopy. With this technique, monosodium urate crystals in gout appear needle shaped and are negatively birefringent. In contrast, calcium pyrophosphate dihydrate crystals in pseudogout are rhomboidal in shape and are weakly positively birefringent. Infection can coexist with crystal-induced disease, necessitating microbiologic evaluation even when crystals are found. Hemorrhagic fluids can also result from infection, although their presence suggests malignancy or trauma. Figure 257-1 provides an algorithm for the analysis of joint fluid.

Depending on the clinical findings and the results of initial laboratory testing, other studies may be performed to investigate less common diagnostic possibilities such as metabolic disease or malignancy. The laboratory evaluation of inflammatory arthritis may also include serologic tests for infections such as Lyme disease, HIV infection, or hepatitis.

LABORATORY EVALUATION OF SYSTEMIC INFLAMMATORY DISEASE

Among rheumatic diseases, some are characterized by severe systemic inflammation that can cause organ-threatening and life-threatening manifestations. These diseases can have arthritis as a component and presenting complaint, although the prominence of extra-articular manifestations, especially as they develop over time and involve organs such as the kidney, points to their systemic nature. These diseases can be categorized on the basis of clinical, serologic, and pathologic findings, with the presence of vasculitis, irrespective of blood vessel size, providing a unifying feature in disease classification.

The terms *connective tissue disease* (CTD) and *collagen vascular disease* are both used to denote a group of diseases that includes RA, SLE, Sjögren syndrome, polymyositis, dermatomyositis, and progressive systemic sclerosis. Diseases in this group can share common or overlapping clinical features, especially early in their course, when their presentations may be similar. In this stage of disease, the condition may be called *undifferentiated CTD*, with serologic markers sometimes predictive of the eventual diagnosis.

Antinuclear Antibodies

The expression of antibodies to components of the cell nucleus (antinuclear antibodies, or ANAs) is characteristic of CTD and is essentially invariable in patients with SLE (Chapter 266). These antibodies target a host of nuclear macromolecules, including DNA, RNA, and proteins as well as complexes of proteins with nucleic acid. These antigens are ubiquitously expressed in cells and subserve critical processes related to chromosomal structure, cell division, transcription, and translation. The basis for the antigenicity of these molecules is unknown, although DNA and RNA both have intrinsic immunologic activities and can stimulate cytokine production through action on both TLR and non-TLR PRRs, especially when in the form of immune complexes. Furthermore, these antigens may undergo post-translation modification as well as enzymatic cleavage reactions during cell death, perhaps increasing their immunogenicity.

ANAs are commonly measured by immunofluorescence (IF) assays in which sera are incubated with tissue culture cells (e.g., Hep2 cells) fixed to a glass slide. Antibody binding is revealed by fluorescence microscopy after incubation of the slide with a fluoresceinated anti-immunoglobulin reagent. Results are reported in terms of the pattern of fluorescence as well as the end-point titer of sera at which fluorescence can be observed. The patterns of binding differ depending on the location of the particular macromolecular target, although a few patterns predominate. These patterns include homogeneous, rim, nucleolar, and speckled; in addition, ANA tests can detect antibodies to cytoplasmic antigens. Despite some disease associations, these patterns do not have diagnostic significance. Table 257-2 presents a list of major ANAs with their pattern and disease associations.

A major limitation in the assays of ANAs concerns the frequency of positive reactivity in the sera of otherwise normal individuals who lack evidence of a CTD. Depending on the titer for screening, the sera of as many as 20% of normal individuals express reactivity in the IF ANA test.[4] The basis of this reactivity, which occurs more commonly in women than men, is not well understood, although it may reflect a predisposition to autoimmunity that is manifest in ANA production in the absence of other immunopathologic disturbances for the complete development of a CTD. Because ANA testing is often performed to evaluate nonspecific complaints such as arthralgias, fatigue, and fever, a positive test must be interpreted with caution and not used as proof of a CTD in the absence of correlative clinical or laboratory findings.

Because of detailed biochemical studies, the molecular identity of many ANAs is now known, allowing for the development of specific immunochemical assays using technologies that can eliminate the need for visual inspection of a slide stained for fluorescence. These tests (e.g., ELISA) can be performed individually, although multiplex assays provide simultaneous assessment of multiple specificities. With multiplex assays, a version of the ANA test can be generated by measuring antibodies to a select set of more commonly targeted autoantigens, with positive binding to any one antigen considered indicative of a positive value for an ANA. Because sera of patients with SLE and other systemic inflammatory diseases contain numerous ANAs, focus on a limited set of specificities may fail to detect some positive sera. In addition, because immunochemical assays may detect low levels of antibody binding, the significance of this reactivity is not certain. Thus, although multiplex assays are operationally easier than conventional ANA testing by immunofluorescent staining, in clinical situations, the IF assay remains an important laboratory test for patient evaluation.

Among many ANA specificities now identified, only a few are performed routinely because of their value for diagnosis and prognosis. For certain CTDs, diagnosis can be readily determined from clinical findings or other laboratory tests. In these instances, the ANA determination provides confirmatory information as well as clues to the occurrence of certain clinical manifestations.

Antibodies to DNA

Antibodies to DNA (anti-DNA) are serologic markers of SLE and represent a criterion in the classification of patients with this disease (Chapter 266). These antibodies bind sites on both single-stranded (ss) and double-stranded (ds) DNA, although anti-dsDNA antibodies are more specific for SLE and therefore routinely measured. Although these antibodies can bind free DNA, DNA in the cell occurs in association with histones to form a structure called the *nucleosome*, with DNA wrapped around a histone core. Anti-DNA may therefore be considered a subset of antibodies to nucleosomes, with nucleosomes probably serving as the driving antigen for this response.[5]

In clinical practice, the measurement of anti-DNA antibodies is an important element in the evaluation of patients with a broad array of clinical complaints, given the heterogeneity and multisystemic nature of SLE. Anti-DNA determinations, in addition to their value in diagnosis, also convey prognostic information and serve as an index of disease activity. The association with disease activity appears strongest with glomerulonephritis, most likely because of the role of DNA–anti-DNA immune complexes in immunopathogenesis. The association of anti-DNA antibodies with other disease manifestations is less certain, limiting the use of this marker as a measure of overall disease activity. The presence of anti-DNA may nevertheless be important in assessing likelihood of response to therapies such as belimumab (anti-BLyS or anti-BAFF), an agent indicated for treatment of patients with active disease as evidenced by the presence of either anti-DNA antibodies or a positive test for ANA.

Several immunochemical approaches can be used to detect anti-DNA antibodies, although solid-phase ELISA assays are convenient and sensitive and eliminate the need for radioactivity. The assays vary in regard to the spectrum of anti-DNA antibodies detected, and results between assays may not

TABLE 257-2 SELECTED ANTINUCLEAR ANTIBODIES AND RHEUMATIC DISEASES

DISEASE PATTERN	ANTIBODY	ANTIGEN	ASSOCIATION
Homogeneous	Antihistone	Histones H1, H2A, H2B, H3, H4	Drug-induced lupus (>95%)
Rim	Anti-double-stranded DNA	Double-stranded DNA	SLE (50%)
Speckled	Anti-Sm	snRNP proteins	SLE (30%)
	Anti-U1-RNP	U1 snRNP proteins	SLE (30%); MCTD (>95%)
	Anti-Ro (SS-A)	Two proteins complexed to small RNAs Y1-Y5	SLE (30%); Sjögren syndrome (70-80%)
	Anti-La (SS-B)	Single protein plus RNA polymerase III transcript	SLE (15%); Sjögren syndrome (50-70%)
	Anti-Ku	DNA binding protein	SLE (10%)
	Anti-SCL-70	DNA topoisomerase I	PSS (40-70%); CREST (10-20%)
Nucleolar	Anti-PM-Scl	Nucleolar protein complex	PSS (3%); PM (8%)
	Anti-Mi-2	Nuclear protein complex	DM (15-20%)
	Anti-RNA polymerase	Subunits of RNA polymerase I	PSS (4%)
Dividing cell	Anticentromere	Centromere/kinetochore protein	CREST (80%); PSS (30%)
	Antiproliferating cell nuclear antigen	Auxiliary protein of DNA polymerase δ	SLE (3%)
Cytoplasmic	Anti-Jo-1	Histidyl tRNA synthetase	ILD in PM/DM (18-25%)
	Anti-PL-7	Threonyl tRNA synthetase	PM/DM (3%)
	Anti-PL-12	Alanyl tRNA synthetase	PM (4%)
	Anti-SRP	Signal recognition particle	SLE (10%)
	Anti-ribosomal P	Large ribosomal subunit	PM/DM (3%)

CREST = calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia; DM = dermatomyositis; ILD = interstitial lung disease; MCTD = mixed connective tissue disease; PM = polymyositis; PSS = progressive systemic sclerosis (diffuse scleroderma); SLE = systemic lupus erythematosus; snRNP = small nuclear ribonucleoprotein; tRNA = transfer RNA.

correlate.[6] Nevertheless, for each assay, the dynamic range for testing is large. With treatment and disease quiescence, anti-DNA antibodies may essentially disappear; with flare, levels may increase dramatically. This property distinguishes anti-DNA antibodies from other ANAs in SLE, levels of which tend to be more consistent over time.

As is the case for other ANAs, the appearance of anti-DNA antibodies in the serum may precede other manifestations of SLE, suggesting vigilance if these antibodies are present in patients who have symptoms that suggest a CTD but lack other evidence to establish a firm diagnosis.

Other Antinuclear Antibodies

Anti-Sm and anti-RNP antibodies are related specificities that commonly occur together in the sera of patients with SLE, a phenomenon called *linkage*. These antibodies bind proteins on subcellular particles called *snRNPs* (small nuclear ribonucleoproteins) that are composed of a set of proteins and uridine-rich RNAs. Anti-Sm and anti-RNP antibodies differ in protein specificity and in the ability to cause immunoprecipitation of the bound RNA molecules. Anti-Sm antibodies occur only in patients with SLE and represent a serologic marker in disease classification. In contrast, anti-RNP antibodies can appear in the sera of patients with other clinical presentations and, in the absence of anti-Sm, may characterize patients with overlapping CTD features, so-called mixed CTD or MCTD. In SLE, the frequencies of anti-Sm and anti-RNP antibodies vary among racial and ethnic groups, although a clear association with particular clinical manifestations has not been established.

Anti-Ro and anti-La antibodies (or anti-SS-A and anti-SS-B), another set of linked ANAs, are directed to protein-RNA complexes that are involved in cellular metabolism of RNA. These antibodies are expressed more widely in patients with CTD and appear in the sera of patients with SLE, RA, and Sjögren's syndrome, among others. Assessment of these antibodies is important because of their association with the neonatal lupus syndrome, which results from the transplacental passage of antibodies and causes congenital heart block as well as rash in the neonate. Although both Sm/RNP and Ro/La are complexes of proteins and RNA, these antibodies appear to be expressed by different patient subsets, suggesting distinct mechanisms of induction and clinical associations.[7]

Although ANAs are directed to ubiquitous antigens, they nevertheless are expressed in disease-specific patterns and may show association with particular organ-specific manifestations. These associations include anti–ribosomal P antibodies with central nervous systemic involvement in SLE, antibodies to DNA topoisomerase 1 (anti-SCL-70) with progressive systemic sclerosis (diffuse scleroderma), antibodies to centromeres with CREST syndrome (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia), and antibodies to histidyl transfer RNA synthetase (anti-Jo-1) with interstitial lung disease in scleroderma (Chapter 267). In inflammatory myopathies, the presence of certain autoantibodies may be associated with particular patterns of disease, with antibodies to the enzyme 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase present in a syndrome of necrotizing myositis; the syndrome can occur in patients treated with statins, which can inhibit the enzyme.[8]

In addition to their association with specific disease manifestations, antibodies to both DNA- and RNA-binding proteins such as Sm and RNP may contribute to overall immune dysregulation in patients with autoimmune disease because of their formation of immune complexes containing DNA or RNA. These complexes can stimulate the production of type 1 interferon by triggering both TLR and non-TLR nucleic acid sensors as well as other cellular receptors (e.g., Fc receptors). Because immunoassays of interferon with patient sera are limited, the presence of interferon is observed more clearly in the pattern of gene expression known as the interferon signature in peripheral blood cells. This signature can be assessed by both microarray assays and measurement of more limited sets of messenger RNA molecules that are induced by interferon. Because antibodies to RNA-binding proteins in particular may promote this pattern, the serologic assay of these ANAs may allow assessment of the likelihood of both nonspecific and specific immunologic disturbances.[9]

Antibodies to Phospholipids

Originally defined by their effects on in vitro clotting tests, antibodies to phospholipids (APLs) are associated with in vivo thrombosis and have been termed *lupus anticoagulants* (LACs). Patients with these antibodies display a clinical condition, termed the *antiphospholipid antibody syndrome*, which is characterized by arterial or venous thrombosis, thrombocytopenia, and first-trimester spontaneous abortions (Chapter 176). This syndrome may occur by itself or in the context of SLE, where it may contribute to the acceleration of atherosclerosis, premature stroke, and myocardial infarction. The laboratory evaluation of this condition involves specific assays of antibodies to phospholipids and related proteins as well as functional assays of clotting. Because expression of these antibodies may vary over time, testing must be performed on more than one occasion at least 6 weeks apart. Furthermore, the results of immunochemical and functional assays may not be congruent because they are likely related to the heterogeneity of antibodies.

The serology of APLs is complicated because it is related to the nature of the antigenic targets as well as heterogeneity among patients.[10,11] These antigens include phospholipids such as cardiolipin. Cardiolipin, however, can bind to the protein β_2-glycoprotein 1, which is also a target for antibodies in this condition. Serologic evaluation thus involves assays with a complex of cardiolipin and β2-glycoprotein 1 as well as β2-glcycoprotein in an ELISA format using reagents to measure IgG and IgM. The association of antibodies with thrombosis appears strongest with IgG antibodies; determination of the IgA isotype may also be informative depending on the results of IgG and IgM assays. ELISAs for these antibodies are not yet standardized, making it important to specify assay features related to quantitation such as cutoff values used to define positivity and values that are considered significant.

Functional assays for LACs involve tests directed at inhibition of in vitro clotting (e.g., activated partial thromboplastin time, dilute Russell viper venom time), recognizing the discordance between in vivo thrombosis and in vitro anticoagulation. Functional assays to detect lupus anticoagulants involve a mixing step in which patient plasma is mixed with normal plasma to determine the presence of an inhibitor (i.e., an antibody) as opposed to a deficiency state. The mechanisms by which antibodies to phospholipids and related proteins may cause thrombosis in vivo are unknown, although these antibodies may interact with the surface of cells (e.g., endothelium) to promote a prothrombotic state. Assessing the likelihood of the syndrome is best accomplished by considering results of both the immunoassays and functional assays in the context of the individual patient because both factors may promote thrombotic events.

Complement

Assessment of the complement system can provide valuable information on the activity of diseases in which immune complex deposition may promote inflammation and tissue injury (Chapter 50). This system involves a large number of proteins that function in enzyme cascades to generate degradation products that amplify immunologic reactions and promote the destruction or removal of foreign organisms as well as damaged cells. In the setting of SLE and in certain forms of vasculitis and glomerulonephritis, immune complexes activate complement to promote local inflammation. This activation can be measured in terms of the total complement level in the blood by functional assays of hemolytic activity; by measurement of individual complement components such as C3 and C4, whose levels are reduced by cleavage during activation; by measurement of split products of cleaved complement components; and by measurement of complement fragments bound to red blood cells during complement activation. Proteins of the complement system are acute phase reactants and can increase with inflammation, including active disease. Correspondingly, low levels may reflect inherited complement deficiency rather than consumption; genetic deficiency of C1q, for example, is highly associated with SLE.

Antineutrophil Cytoplasmic Antibodies

Antineutrophil cytoplasmic antibodies (ANCAs) are autoantibodies that react to determinants in the neutrophil and occur prominently in patients with certain forms of necrotizing vasculitis or rapidly progressive glomerulonephritis. Reflecting the serology, conditions have been called ANCA-associated vasculitis (AAV). Two main forms of ANCA have been distinguished on the basis of the target antigens and pattern of immunofluorescence staining of fixed neutrophils: PR3-ANCA (C-ANCA), which reacts with proteinase-3 (PR3), and MPO-ANCA (P-ANCA), which reacts with myeloperoxidase (MPO). By immunofluorescence, PR3-ANCA shows staining in the cytoplasm; staining by MPO-ANCA localizes in the perinuclear area. ANCAs to other proteins have also been identified, but these may also occur in conditions other than vasculitis.

In the evaluation of severe, multisystem inflammatory disease, ANCA testing is important to evaluate diagnostic possibilities.[12] ANCAs occur in association with varying clinical manifestations in patients with AAV and help define patterns of clinical involvement in terms of organ system involvement as well as histopathology (e.g., presence of granulomatous inflammation). PR3-ANCA occurs commonly in patients with granulomatosis with polyangiitis (GPA, formerly called Wegener granulomatosis) as well

eosinophilic granulomatosis with polyangiitis (EGPA, formerly called Churg-Strauss disease); MPO-ANCA marks the course of vasculitis caused by microscopic polyangiitis.[13] Although there is overlap between serology and clinical features, PR3-ANCA occurs commonly in patients with upper airway disease, whereas MPO-ANA occurs commonly patients with patients with rapidly progressive renal disease (Chapter 270).

In patients with ANCA-associated glomerulonephritis, the kidney lacks evidence of immune deposits, as indicated by the lack of staining for immunoglobulins or complement. Kidney disease of this kind is termed *pauci-immune glomerulonephritis*. Although ANCA testing is useful in initial diagnosis, its role for assessing disease activity is less certain. Occasionally, in patients who are desperately ill and cannot tolerate a lung or kidney biopsy, the presence of an ANCA can be used as preliminary evidence for diagnosis to allow the initiation of immunosuppressive therapy. ANCA testing is also useful for assessing the likelihood for relapse because patients who express PR3-ANCA appear at risk for recurrent disease.

Cryoglobulins

Cryoglobulins are serum immunoglobulins that precipitate in the cold and promote the pathogenesis of systemic inflammatory disease through tissue deposition. The presence of a cryoglobulin is detected by allowing blood, collected warm, to remain cool at 2° to 4° C for 1 or more days. After centrifugation, the amount of cryoprecipitate is measured and expressed as a cryocrit. In the preanalytical phase, it is important that the blood remain at a temperature of 37° C during all steps from drawing the blood from the patient to separation of the serum fraction after coagulation. Thermos flasks with preheated sand or water and other special devices are available to keep the blood tubes at 37° C during transport. If these steps are not taken, cryoprecipitation at even room temperature may already occur before separation of the serum from the blood cells, possibly resulting in false-negative results.[14]

Subsequent analysis of the cryoprecipitate by immunochemical assays allows determination of its components. Cryoglobulins can be classified into three main types on the basis of their composition: (1) single, or type I; (2) mixed, type II; and (3) mixed, type III. A type I cryoglobulin consists of only a monoclonal immunoglobulin that precipitates in the cold. A mixed-type cryoglobulin contains RFs bound to polyclonal IgG to form an immune complex. In type II cryoglobulins, the IgM RF is monoclonal, and in type III, the IgM RF is polyclonal.

Type I cryoglobulins occur in patients with lymphoproliferative disorders such as Waldenström macroglobulinemia, multiple myeloma, or chronic lymphocytic lymphoma (Chapters 184 and 187). In contrast, patients with mixed cryoglobulins can present with a wide range of signs and symptoms resulting from vasculitis. These manifestations include purpura (a sign of leukocytoclastic vasculitis), weakness, arthritis, and neuropathy, representing a syndrome known as *essential mixed cryoglobulinemia*. Most patients with this condition have infection with hepatitis C virus, with viral components present in the complexes. These patients have serologic evidence of this infection as well as manifestations attributable to the underlying liver disease. As in the case of other CTDs and systemic inflammatory diseases, the evaluation of patients with essential mixed cryoglobulinemia demands attention to the entire patient and the impact of disease on multiple organs.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ansar W, Ghosh S. C-reactive protein and the biology of disease. *Immunol Res*. 2013;56:131-142.
2. Van Venrooij WJ, Van Beers JBC, Pruijn GJM. Anti-CCP antibodies: the past, the present and the future. *Nat Rev Rheumatol*. 2011;7:391-398.
3. Aleteha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League against rheumatism collaborative initiative. *Arthritis Rheum*. 2010;62:2569-2681.
4. Satoh M, Chan EK, Ho LA, et al. Prevalence and sociodemographic correlates of antinuclear antibodies in the United States. *Arthritis Rheum*. 2012;64:2319-2327.
5. Bizzaro N, Villata D, Giavarina D, et al. Are anti-nucleosome antibodies a better diagnostic marker than anti-dsDNA antibodies for systemic lupus erythematosus? A systematic review and a study of metanalysis. *Autoimmun Rev*. 2012;12:97-106.
6. Pisetsky DS. Standardization of anti-DNA antibody assays. *Immunol Res*. 2013;56:420-424.
7. Ching KH, Burbelo PD, Tipton C, et al. Two major autoantibody clusters in systemic lupus erythematosus. *PLoS ONE*. 2012;7:e32001.
8. Hamann PD, Cooper RG, McHugh NJ, et al. Statin-induced necrotizing myositis—a discrete autoimmune entity within the "statin-induced myopathy spectrum.". *Autoimmun Rev*. 2013;12:1177-1181.
9. Rönnblom L, Eloranta M-L. The interferon signature in autoimmune disease. *Rheumatology*. 2013;25:248-253.
10. Favaloro EJ. Variability and diagnostic utility of antiphospholipid antibodies including lupus anticoagulants. *Int J Lab Hematol*. 2013;35:269-274.
11. Lakos G, Favaloro EJ, Harris EN. International consensus guidelines on anticardiolipin and anti-β2-glycoprotein I testing. *Arthritis Rheum*. 2012;64:1-10.
12. Lionaki S, Blyth ER, Hogan SL. Classification of antineutrophil cytoplasmic autoantibody vasculitides. *Arthritis Rheum*. 2012;64:3452-3462.
13. Lally L, Spiera R. Current landscape of antineutrophil cytoplasmic antibody-associated vasculitis: classification, diagnosis, and treatment. *Rheum Dis Clin North Am*. 2015;41:1-19.
14. Damoiseaux J. The diagnosis and classification of the cryoglobulinemic syndrome. *Autoimmun Rev*. 2014;13:359-362.

258

IMAGING STUDIES IN THE RHEUMATIC DISEASES

RONALD S. ADLER

Historically, rheumatic disorders have been well characterized by conventional imaging. In as much as these disorders often manifest in characteristic distributions and present with specific alterations in the appendicular or axial skeleton and adjacent soft tissues, radiographic evaluation has been sufficient to both characterize the abnormalities as well as provide a relatively small number of differential possibilities as to the specific disease. The most well-studied example is rheumatoid arthritis (RA) in which symmetric involvement of the metacarpophalangeal joints, uniform joint space narrowing, periarticular osteopenia, and juxta-articular erosions along the "bare areas" are pathognomonic.

The development of new therapeutic alternatives for the inflammatory arthritides, so-called disease-modifying antirheumatic drugs (DMARDs), and chondroprotective strategies in the case of osteoarthrosis, require methods to diagnose these diseases at an earlier stage, characterize the degree of inflammation, and provide a useful metric to assess therapeutic response.[1] Indeed, it has become necessary to assess for possible joint and soft tissue abnormalities before irreversible tissue damage, the latter often being the case when the radiographic findings are abnormal. Fortunately, the requirement to achieve earlier diagnosis has paralleled advances in imaging. Ultrasonography and magnetic resonance imaging (MRI) have largely supplanted conventional radiographic evaluation in the imaging work-up of patients with suspected rheumatologic disorders and negative radiographs.[2,3] The term *molecular imaging* has been applied, particularly in the case of MRI and positron emission tomography (PET), in as much as these modalities reflect local tissue environment or metabolic activity.

RADIOGRAPHIC EVALUATION

Radiographic evaluation is among the first studies ordered in patients with a suspected rheumatologic disorder. In the current digital era, conventional analog-based radiographs have been largely replaced by computed radiography. Images are usually displayed on workstations with high-resolution monitors within the context of a picture archiving system (PACS). Digital radiographs are of high spatial resolution but relatively poor soft tissue contrast. These images are amenable to a variety of image processing schemes, resulting in enhanced definition of the cortical surfaces and cancellous bone, which may be of value in displaying subtle erosions.

It is important to recognize that radiographs are projection images. To detect an abnormality, it may be necessary to view a joint or other structure at a specific angle. For instance, subtle erosions may only be apparent when viewed tangentially as opposed to en face. It is therefore necessary to have specific image protocols in order to optimally display the joint, cortical surface, or soft tissue structure. Most radiographic evaluations contain at least two orthogonal projections. The addition of an oblique view or other specialized projection may be necessary to address a specific clinical question.

The nature and distribution of joint space narrowing, presence of osteopenia, new bone formation, soft tissue swelling, soft tissue calcification, chondrocalcinosis, presence and nature of erosions, and assessment for joint malalignment may allow a specific diagnosis and help determine the severity of disease (Fig. 258-1). For instance, the presence of a juxta-articular erosion extending over an adjacent area of slightly hyperdense soft tissue swelling in the setting of normal bone mineralization with maintenance of the adjacent joint space is diagnostic of gout, in contrast to RA noted earlier. The seronegative arthritides, such as psoriatic arthritis, have a characteristic appearance in the small joints of the hand and feet, including a predilection for distal joints, asymmetry, and appositional new bone formation.

Table 258-1 summarizes some of the features of several of the more common diseases that may be encountered in clinical practice.

Finally, radiographs provide a direct means for needle localization during percutaneous procedures, predominantly joint injections, aspirations, and some biopsies. These are generally performed while imaging in real time (fluoroscopy) using short bursts of low-intensity x-rays enhanced through an image intensifier. Injection of joints under fluoroscopic guidance provides a convenient means to ensure intra-articular deposition of therapeutic agent or for diagnostic aspiration. Intra-articular location is verified by injection of a small amount of a standard iodinated contrast material. Arthrography using fluoroscopic guidance can be used as a primary diagnostic tool, but this application has largely been replaced by intra-articular injection of contrast followed by computed tomography (CT) or MRI.

For some procedures, CT may be preferable, depending on the location of the abnormality. The principal disadvantages of fluoroscopy relate to the use of ionizing radiation and poor soft tissue contrast. The latter becomes important with needle placements near neurovascular structures that may be potentially compromised by poor position. CT allows greater control over needle placement at the cost of greater levels of radiation exposure. Ultrasonography has replaced fluoroscopy and CT for a large number of percutaneous procedures. MRI provides another method to perform a variety of procedures without the necessity of ionizing radiation. These options will be discussed in greater detail below.

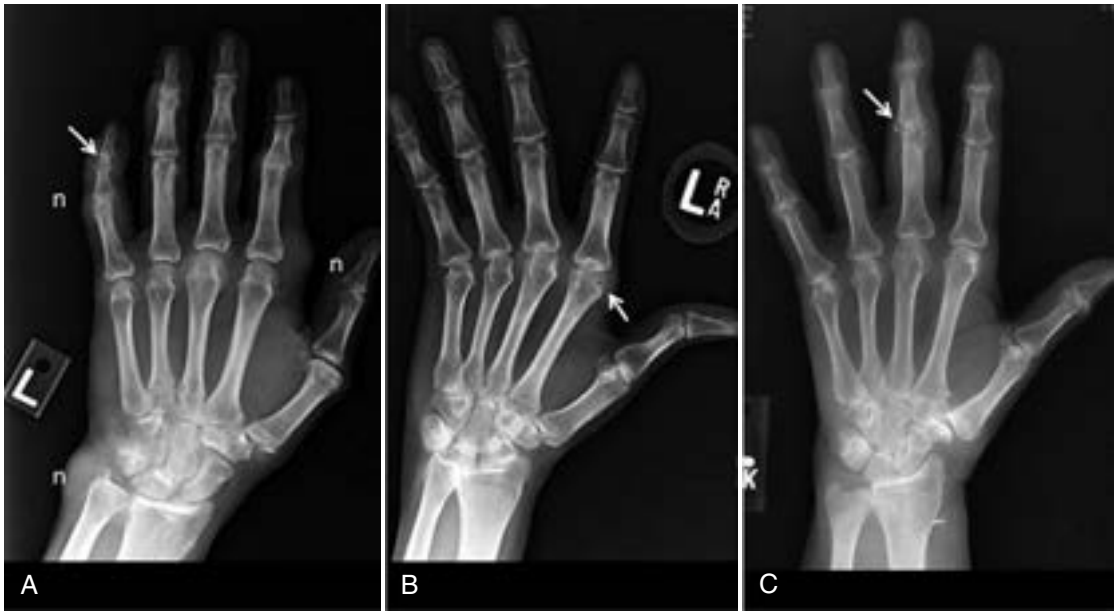

FIGURE 258-1. Three hands with different diagnoses. A, Gout. Radiograph of the left hand showing multiple dense soft tissue nodules (n) with multiple small erosions affecting the ulnar styloid, triquetrum and fifth ray. A large erosion (arrow) at the fifth distal interphalangeal (DIP) joint demonstrates bone formation extending circumferentially about the adjacent tophaceous deposit typical of an overhanging edge. Bone mineralization and joint spaces are preserved. **B, Rheumatoid arthritis.** There is ulnar deviation of the second through fifth metacarpophalangeal (MCP) joints with uniform joint space loss involving the MCP joints and the carpus. The DIP joints are spared. Periarticular demineralization is present with small erosions along the radiovolar aspect of the second (*arrow*) MCP joint. **C, Osteoarthrosis.** Soft tissue swelling affecting the third digit with joint space narrowing and bone production affecting the DIP joints, third and fifth proximal interphalangeal (PIP) joints, basal joint of the thumb, and scaphotrapeziotrapezoid joint. There are subchondral cystic changes at the third PIP joint having an erosive character (*arrow*). Mineralization is preserved, as are the radiocarpal and MCP joint spaces.

TABLE 258-1 DISTINGUISHING RADIOGRAPHIC FEATURES OF SEVERAL COMMON RHEUMATIC DISEASES

CONDITION	COMMON SITES	DISTRIBUTION	RADIOGRAPHIC FEATURES
Rheumatoid arthritis	Hands: MCP, PIP; wrists: intercarpal, DRUJ, ulnar styloid; feet: fifth MTP, cervical spine (atlantoaxial, apophyseal)	Bilateral, symmetric, polyarticular	Periarticular osteopenia, periarticular swelling, subluxations (e.g., ulnar, volar), uniform joint space loss, erosions (bare areas)
Osteoarthritis (primary)	Hands (DIP), wrists (basal joint, STT),feet (first MTP), hips (superolateral), knees (medial), spine (discs, facet, apophyseal, uncovertebral)	Symmetric, weight-bearing joints	Normal or increased density, nonuniform joint space loss, subchondral sclerosis, cysts, bone formation (osteophytes) Spine–disc space narrowing, end plate sclerosis and bone formation
Psoriatic arthritis	Hands (DIP, terminal tufts), feet (IP joints), entheses (calcaneus: plantar, posterior), spine, sacroiliac joints	Asymmetric (single ray), polyarticular, segmental (intervertebral, apophyseal)	Normal or increased density, periosteal bone formation, soft tissue swelling, ankylosis (SI joints), thick hyperostosis spine (nonmarginal syndesmophytes), juxta- and periarticular erosions
Ankylosing spondylitis	Spine, SI joints, fibrous joints (pubic symphysis), entheses (adductor origin), rhizomelic joints (hips, shoulders)	Symmetric, continuous (may affect entire spine-bamboo spine)	Normal or increased density, erosions (spine squaring, shining corner) with superimposed bone formation (ankylosis: SI), thin (marginal) syndesmophytes)
Gout	Feet (first MTP), other damaged joints, elbow, knee, hindfoot	Asymmetric, extensor surfaces (elbow), abnormal joints (e.g., osteoarthritic joints)	Normal joint space, normal or increased density, dense soft tissue nodules (tophi), para-articular and subchondral erosions with bone formation along tophi (overhanging edge)
Calcium pyrophosphate dihydrate crystal deposition disease	Hands (second, third MCP), wrists (radiocarpal), TFC, knees (lateral compartment and patella-femoral, menisci)	Symmetric, fibrocartilaginous joints	Normal or increased density, hypertrophic bone formation, subchondral or periarticular cysts, chondrocalcinosis (hyaline, fibrocartilage), periarticular, peritendinous, periligamentous calcification
Infection	Any joint, pyogenic, TB	Monoarticular (mostly), any joint	Pyogenic (osteopenia; 8-10 days), joint space widened (early), joint space loss (rapid development), soft tissue swelling, erosions (both sides of joint), sequestra, periostitis, TB (joint space and mineralization may be preserved), juxta-articular erosions, spine–disc space loss and end plate erosion

DIP = distal interphalangeal; DRUJ = distal radial ulnar joint; IP = interphalangeal; MCP = metacarpophalangeal; MTP = metatarsophalangeal; PIP = proximal interphalangeal; SI = sacroiliac; STT = scaphotrapezotraphezoid; TB = tuberculosis; TFC = triangular fibrocartilage.

COMPUTED TOMOGRAPHY

Computed tomography provides a two-dimensional map of tissue attenuation obtained from external x-ray source(s) located on a rotating gantry, whose radiation is detected by a series of detectors opposite the source. The current generation of CT scanners uses multiple detectors (16, 32, 64, and so on), allowing rapid image acquisition that can be displayed in a single plane in real time (CT-fluoroscopy) or as extremely thin section contiguous or overlapping acquisitions in the axial plane. The acquired images can be reconstructed in multiple planes with equivalent (isotropic) resolution elements (voxels) or as a three-dimensional rendering. Some scanners use dual energy sources, taking advantage of differences in the attenuation characteristics of various tissues at different energies. This has received greatest attention in the setting of gout, enabling a definitive diagnosis as well as depicting tophaceous deposits in anatomic locations not conducive to radiographs or ultrasound.[4]

Computed tomography allows the best assessment of trabecular and cortical bone, providing an excellent means to assess fractures and erosions, the presence of new bone formation (e.g., fracture callus), and degenerative or inflammatory arthritis. Soft tissue mineralization can likewise be well characterized, providing important information as to its etiology. Joints that are difficult to assess on radiographs, including the sacroiliac, temporomandibular, wrist, and sternoclavicular joints, are well seen on CT (Fig. 258-2).

Computed tomography generally has poor soft tissue contrast. Nevertheless, it is still very useful in performing a number of guided procedures because of its tomographic nature and rapid image acquisition capability. Improved soft tissue contrast can be obtained with use of iodinated contrast material. A number of soft tissue tumors, inflammatory synovitis, and infectious processes display pathologic enhancement after contrast administration. CT can likewise be used to produce angiographic displays (CTA) when used in combination with contrast, providing exquisite detail of central and peripheral vascular disease, including in patients with suspected vasculitis. These agents are typically administered intravenously following well-defined enhancement characteristics. CTA has become the method of choice in evaluating patients with suspected pulmonary embolism. Likewise, contrast agents may be used to improve intra-articular contrast (CT arthrography), currently the method of choice in assessing internal derangement in the postoperative shoulder, knee, and so on and in patients who are unable to undergo MRI (e.g., those with claustrophobia, aneurysm clips, or cardiac pacemakers). Imaging of cartilage and soft tissue abnormalities usually depends on pathologic imbibition of contrast material, indicative of degeneration or tearing. A limitation of this approach resides in the fact that some abnormalities may remain occult. An example is the inability to detect a bursal-sided rotator cuff tear after shoulder CT arthrography.

The radiation dose from CT can be high, especially when using the newer scanners. This is most significant when one is looking to minimize exposure, such as in children, requiring protocols specifically designed for the pediatric population. Intravenous (IV) use of iodinated contrast agents are contraindicated in patients with impaired renal function or history of allergic reaction. Nonionic agents can diminish the associated risks but still should be used with caution.

ULTRASONOGRAPHY

Ultrasound imaging takes advantage of the near uniform speed of sound and predictable attenuation characteristics of sound propagation in soft tissue. In general, anatomic images derive from specular surfaces whose dimensions exceed the ultrasound wavelength; inherent noise (speckle) within the image derives from small scatterers, smaller than the resolution element of the transducer. Modern ultrasound equipment contains various methods to reduce speckle in the image, resulting in a more anatomic rendition of the soft tissues. Rapid image acquisition and processing enables ultrasonography to be performed in real time (≈30 frames per second). Ultrasonography is also conducive to evaluation of blood flow from which estimates of flow velocity can be obtained through the Doppler equation. Doppler information is typically reported by either continuously estimating velocity at a specific depth (spectral Doppler) or through a color encoded two-dimensional map (color or power Doppler).

There is great appeal for using ultrasonography in patients with rheumatic disorders. There is no ionizing radiation, and it is real time, inexpensive, relatively portable, and well tolerated. Historically, however, ultrasonography has played only a limited role in the diagnostic assessment and treatment of patients with suspected musculoskeletal abnormalities, being used to differentiate fluid-filled from solid masses. The detection of a Baker cyst in the knee or the presence of a joint effusion constituted two major applications. There has also been limited application of ultrasonography to perform image-guided aspirations and biopsies. Within the United States, in particular, the development of MRI further limited the musculoskeletal applications of ultrasonography.

With the development of linear high-frequency small parts transducers; new imaging capabilities of ultrasound scanners; and the evolution of a new class of compact, portable (laptop) ultrasound units that have excellent image quality, the role of ultrasonography has dramatically changed in recent years.[5,6] These new applications have paralleled the development of new classes of DMARDs for which diagnosis of inflammatory synovitis prior to joint destruction is optimal.

The current generation of ultrasound scanners enables examination of the small joints of the hands and feet, allowing early detection of synovitis (Fig. 258-3). Typically, a 10-MHz or higher frequency linear transducer is used. The displacement of the joint capsule by hypoechoic (dark) soft tissue that displays vascularity on Doppler imaging or is incompressible with direct pressure by the transducer is characteristic, allowing differentiation of synovitis from an effusion. In addition to the detection of synovitis, ultrasonography has been shown to be more sensitive than conventional radiographs in the detection of erosions. Erosions appear as discrete irregular discontinuities

FIGURE 258-2. Infectious sacroiliitis in a 12-year-old boy with a 2-week history of back and left hip pain. A, Axial computed tomography (CT) image of the pelvis at the level of the sacroiliac joints (SIs) photographed using window settings optimized for bone detail. There is clear asymmetry in the two SI joints, with the left appearing more irregular. The cortical margins of the left sacral ala are less distinct and there is an isolated bone fragment (*arrow*) surrounded by soft tissue suspicious for a sequestrum. **B,** CT-guided aspiration of the left SI joint confirmed an infectious origin.

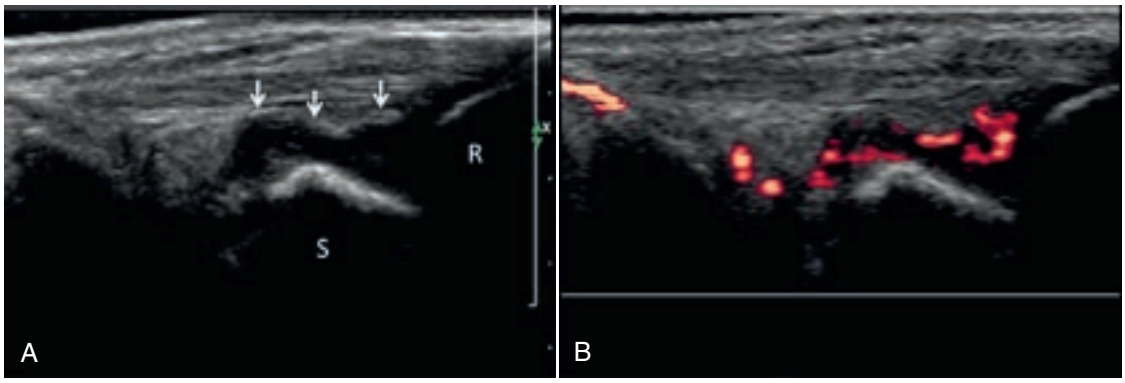

FIGURE 258-3. Synovitis on ultrasonography in a female patient with normal hand radiographs. A, Gray-scale ultrasound image obtained along the dorsal aspect of the radioscaphoid joint shows hypoechoic soft tissue (*arrows*) distending the dorsal recess. The cortical margins of the scaphoid (S) and radius (R) appear as bright reflectors on ultrasonography. **B,** Power Doppler image depicts the marked vascularity (*red color hues*) of the soft tissue illustrating the level of disease activity.

in the normally smooth hyperechoic (bright), reflecting cortical surfaces, often seen in continuity with adjacent inflammatory soft tissue. There is some variation in the appearance of synovitis among various arthritides. The distribution, presence, or lack of symmetry and other concomitant findings may be necessary to obtain a specific diagnosis.

The level of vascularity on color-flow imaging can reflect active inflammation, correlating with clinical and biochemical parameters. A parametric image encoding either mean Doppler shift (color Doppler) or amplitude (power Doppler) is typically used as a standard Doppler map. Both maps can be used to detect abnormal levels of vascularity. Whereas power Doppler provides an indirect measure of the number of moving scatterers within the region being scanned, color Doppler provides a velocity map, and therefore is more subject to artifact (angle dependence and sampling errors). When combined with color-flow imaging, the activity of the synovitis can be estimated. Ultrasound contrast agents can depict capillary flow, resulting in significantly improved detection sensitivity of synovial inflammation, and are used extensively in Europe. They constitute microbubble agents encased in a lipid or polysaccharide shell that can be instilled as either bolus or constant infusion, with the shell being metabolized in the liver and the gas exhaled in the lungs. These agents have biological half-lives on the order of minutes and are best suited to examining target joints. Contrast agents have received Food and Drug Administration approval for cardiovascular applications and therefore can only be used off label for the assessment of synovitis.

Articular and fibrocartilage have characteristic appearances on ultrasonography. Whereas the former appears as a thin hypoechoic band paralleling the articular surface, fibrocartilage appears hyperechoic. Chondrocalcinosis appears as discrete hyperechoic foci with the substance of the cartilage, in which case its presence is suggestive of calcium pyrophosphate deposition disease. Calcification along the margin of the articular cartilage gives rise to the double-line sign seen in gout.

Tendons and muscles have characteristic appearances on ultrasonography. The presence of tendinosis, tendon tears, muscle edema or inflammation, atrophy, and tears can be diagnosed. Ultrasonography is very sensitive, although not specific, for the detection of small amounts of calcification or ossification. It is an excellent method to assess for calcific peritendinitis and to provide guidance for treatment. Abnormal fluid distention of synovial lined structures can be assessed and treated under ultrasound guidance. Ultrasonography is an excellent modality to provide image guidance for therapeutic aspiration and injection of small and large joints, tendon sheaths, and cysts (e.g., bursae, ganglion, paralabral cysts, hematomas, abscesses,) (Fig. 258-4). The real-time capability of ultrasonography is useful to demonstrate the presence of subluxations or painful snapping, to document the distribution of injected material, and to assess adhesions. Ultrasonography is considered the method of choice to detect foreign bodies.

Nerves also have a characteristic appearance on ultrasonography. In cross-section, a nerve often has a "cluster of grapes" appearance, with nerve fascicles appearing hypoechoic and surrounded by hyperechoic endo- and epineural fat. In long axis, nerves display a characteristic tram-track appearance. Ultrasonography has been shown to be useful in the diagnosis and treatment of carpal tunnel syndrome and cubital tunnel syndrome. It is an excellent modality to assess for the presence of posttraumatic or postsurgical and interdigital neuromas and to provide image guidance for treatment, including therapeutic injections and ablative therapy.

Although ultrasonography is well suited to the evaluation of superficial structures, it is less well suited to deep structures. Frequency and penetration are reciprocally related: the higher the frequency, the better the axial resolution but poorer the degree of penetration. A 15-MHz linear transducer would work well in the hand but not in the hip. Examination of a hip might require a 5-MHz transducer and curved transducer geometry with reduced image quality. Excessive abdominal fat can further limit acoustic penetration and distort the ultrasound beam, limiting image quality. Diagnostic ultrasonography does not penetrate bone, resulting in limited acoustic access to joint structures. In some instances, soft tissue contrast can be poor. An inexperienced scanner may find ligaments and tendinous insertions difficult to differentiate from adjacent fibrofatty structures.

MAGNETIC RESONANCE IMAGING

The natural abundance of hydrogen in biological systems and an inherent property of hydrogen, called *spin,* form the basis of conventional MRI. When placed in a strong magnetic field, protons tend to align themselves along the direction of the field. Magnetic field strengths are specified as Tesla and can be variable between clinical scanners. The majority of scanners in clinical usage vary between 1 and 3 Tesla. Application of a radiofrequency (RF) pulse to the system of protons induces the spins to rotate away from the direction of the field, during which time they precess about the direction of the magnetic field at a characteristic frequency, called the Larmor frequency. When the RF pulse is turned off, the spins relax toward their initial state determined by two tissue-dependent relaxation times, T1 and T2, which vary with field strength. T1 (also known as the spin-lattice relaxation) and T2 (or the transverse relaxation time) along with proton density are the principal determinants of signal intensity. The image can emphasize either the T1 or T2 characteristics of the tissue, impacting tissue contrast. Different tissues have varying appearance often based of levels of fat and water content, reflected by their inherent T1 and T2 relaxation times. Tissue morphology is often characterized by their appearance on T1-weighted or proton density images: tendon, muscle, fat, marrow, cortical bone, articular, and fibrocartilage have characteristic appearances. Many pathologic states, alternatively, are characterized by increased mobile water or effective T2 lengthening. Examples include soft tissue edema, inflammatory infiltrates, and neoplasm (Fig. 258-5). Images that emphasize T2 contrast are therefore helpful to display most pathologic states. Selective maps of T2 have been used to characterize the state of articular cartilage in early degenerative disease.[7] Other cartilage specific properties that relate to water content, glycosaminoglycan (GAG) content, and integrity of collagen architecture can be assessed using T2 and other parametric maps that can be derived from the MR data (Fig. 258-6).

The widely used contrast for MRI studies is a neutral, hydrophilic salt of the gadolinium chelate, gadolinium diethylenetriamine-penta-acetic acid (Gd-DTPA). Gadolinium can be injected intravenously or directly into the joint. IV injection (indirect magnetic resonance arthrography) carries the contrast in the vascular system to areas of hyperemia and inflammation (Fig. 258-7). It can be used for assessment of synovial activity in inflammatory joint diseases.[8] Gadolinium is taken up in inflamed synovium and is able to demonstrate thickened pannus. The slope of the early time-signal intensity curve provides a measure of tissue perfusion and can quantify inflammatory activity. Contrast material excreted into the synovial fluid provides excellent depiction of intra-articular structures and can be used in lieu of arthrographic

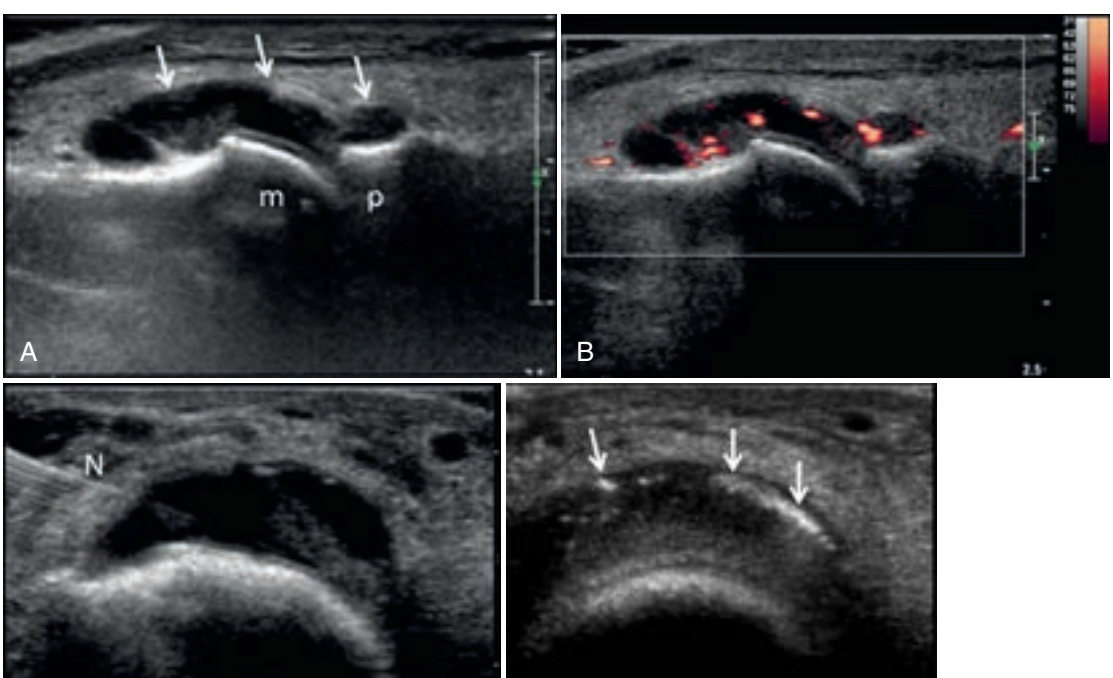

FIGURE 258-4. Ultrasound-guided therapy in the first metatarsophalangeal (MTP) joint of a patient with pain and swelling. A, Longitudinal gray scale image of the dorsal recess of the first MTP joint. Fluid and soft tissue distend the joint capsule (*arrows*). The metatarsal (m) and proximal phalanx (p) are labeled. Note that a thin hypoechoic (dark) band parallels the surface of the metatarsal head, corresponding to the overlying articular cartilage. **B,** Increased vascularity (*red color hues*) demonstrated on power Doppler imaging within the dorsal recess reflects the level of disease activity. **C,** Transverse gray scale ultrasound image shows a needle (N) within the distended dorsal recess from which several drops of synovial fluid were aspirated followed by therapeutic injection. **D,** Postinjection transverse ultrasonography depicts low level echoes (small echogenic foci within dorsal recess) and microbubbles (*arrows*) within the distended joint capsule from injected material. Whereas microbubbles aggregate along the nondependent portion of the distended joint capsule, injected material tends to settle to the deep portion of the recess.

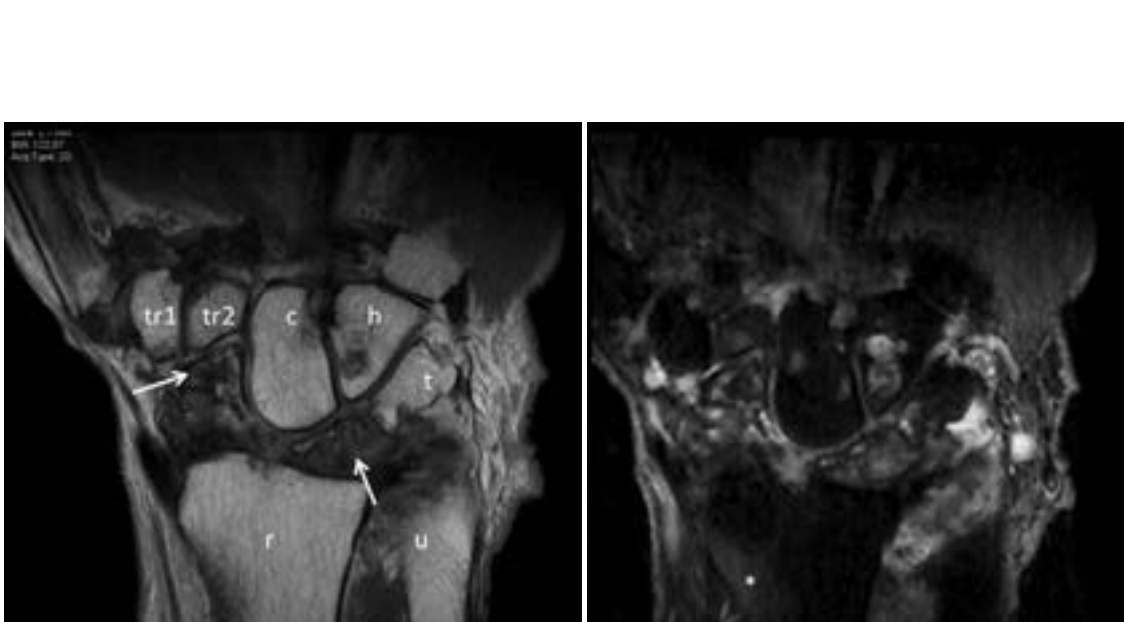

FIGURE 258-5. Magnetic resonance image of the right wrist in a female patient with advanced rheumatoid arthritis. A, Proton density coronal image shows loss of normally bright marrow signal within the scaphoid and lunate bones (*arrows*). The proximal scaphoid is eroded, and the lunate appears deformed and translocated and volarly tilted (not shown), giving rise to its triangular appearance. The distal ulna (u) is poorly visualized because of a large erosion. Intermediate-intensity material (appears dark gray) within the carpus and distal radioulnar joint is difficult to separate from the distal ulna, lunate, and scaphoid. The triquetrum (t), hamate (h), trapezium (tr1) and trapezoid (tr2), capitate (c), and radius (r) are labeled. **B,** Fluid-sensitive coronal image emphasizing T2 relaxation demonstrates increased signal intensity (bright) within the inflammatory pannus, compatible with increase in mobile water associated with inflammation. Increased signal intensity is evident within the lunate, scaphoid, and distal ulna, including focal areas within the distal row of carpal bones, corresponding to small erosions. Diffuse increased signal within the distal radius likely reflects reactive marrow edema (*asterisk*).

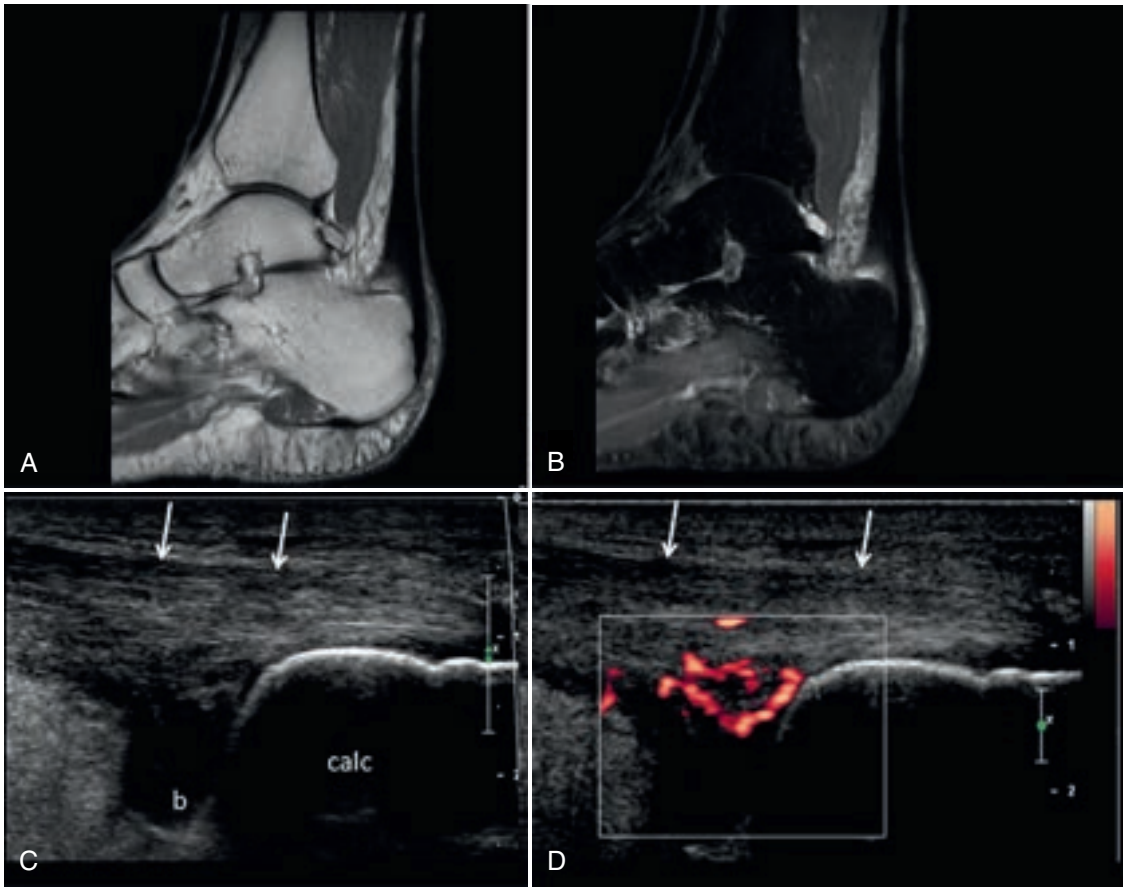

FIGURE 258-6. **Imaging of the soft tissues in a patient with retrocalcaneal pain demonstrating complementary nature of magnetic resonance imaging and ultrasonography. Whereas the proton-density sagittal image (A) emphasizes anatomic detail, the fluid-sensitive image (B) depicts thickening and increased signal intensity within the distal Achilles tendon, reflecting tendinosis, retrocalcaneal bursitis, or a tear of the deep surface of the tendon. Surrounding increased signal intensity (bright areas) within the adjacent soft tissue reflects adjacent soft tissue edema. Long-axis gray scale (C) and power Doppler (D) images of the same patient obtained when the patient presented for ultrasound-guided therapeutic injection. The tendon (*arrows*) is inhomogeneous. A prominent hypoechoic collection deep to the tendon is compatible with retrocalcaneal bursitis (b). There is prominent increased vascularity on power Doppler imaging at the margin of the bursa and tendon. The calcaneus (calc) is labeled.**

direct techniques. In glycosaminoglycan (GAG) -depleted cartilage, there can be delayed uptake of contrast into the cartilage, which would normally be inhibited by the negatively charged GAG molecules.

Patients with renal disease who receive IV injection of gadolinium can develop nephrogenic systemic fibrosis (NSF). (Also see the section Nephrogenic Systemic Fibrosis in Chapter 267.) When the kidney cannot sufficiently clear out the gadolinium, it produces fibrosis of many tissues, including the skin, muscle, heart, nerves, and pleura. To date, NSF has been seen only in patients who have been given IV gadolinium with acute or chronic renal insufficiency. The changes in the skin with NSF are usually bilateral and symmetrical, primarily involving the extremities and the trunk. These changes can mimic systemic sclerosis but, unlike that disease, the face is usually spared. If renal function improves, the skin lesions may stabilize or get better, although in some patients, the process progresses, affecting mobility and causing severe pain.

Injection of dilute gadolinium into the joint (direct magnetic resonance arthrography) is helpful for outlining structures to determine whether there is morphologic damage. Injection is usually performed either under fluoroscopic or ultrasound guidance. This technique is particularly effective for visualization of small structures such as the labrum of the hip or shoulder if there is no joint effusion. It is also helpful for demonstrating breakdown of soft tissue structures that normally prevent communication between joint compartments such as the rotator cuff, triangular fibrocartilage of the wrist, and ligaments in the various joints. Newer techniques that enable image acquisition in near real time as well as the development of MR-compatible needles now permit a variety of percutaneous procedures to be performed directly under MR guidance.

SCINTIGRAPHY

Scintigraphy by its nature represents physiologic imaging because it derives from labeling physiologically occurring substances with a gamma-emitting radionuclide and uses detectors in the form of gamma cameras arranged in a planar or circumferential configuration to determine the distribution of radionuclide within the tissue. Scintigraphy can provide a global assessment of abnormal tracer update or can be performed using a targeted approach (Fig. 258-8). Images often provide high tissue contrast but are of relatively poor spatial resolution. Commonly used agents vary from tagged red blood cells to assess blood flow; agents that reflect bone metabolism (technetium-99m methylene diphosphate [Tc-MDP]); agents that reflect glucose metabolism (18-fluorine deoxy-glucose [18-FDG]), in the case of PET ; and agents that concentrate at sites of inflammation, such as autologous white blood cells labeled with ^{111}In (Indium) and ^{67}Ga-citrate (gallium). Clinical applications include detection of a variety of malignancies, osteomyelitis, vascular graft infection, multifocal infectious disease, inflammatory diseases such as RA, vasculitis, inflammatory bowel disease, sarcoidosis, fever of unknown origin, and infection of joint prostheses.

Traditional nuclear medicine involves use of single gamma photon emissions as a product of nuclear decay. The information can be displayed using planar imaging through a single (or multiple) pinhole camera or displayed tomographically in a manner similar to CT (single-photon emission computed tomography [SPECT]). Bone scintigraphy uses Tc-MDP as the radioactive tracer. The isotope goes to areas of high bone turnover and vascular flow as well as areas of calcium or bone deposition. Three-phase bone scans are obtained at different intervals after injection, reflecting the early vascular phase, the intermediate blood pool phase, and the late phase. Each phase allows for further characterization of the disease process. Abnormal tracer uptake is seen in areas of inflammation, infection, neoplasm, osteonecrosis, and fracture. The scan is most useful to identify the location of lesions within the skeleton but is nonspecific.

Positron emission tomography scans use the appearance of two simultaneously produced 511-KEV gamma rays after annihilation of a positron and electron pair to localize the distribution of radionuclide. The

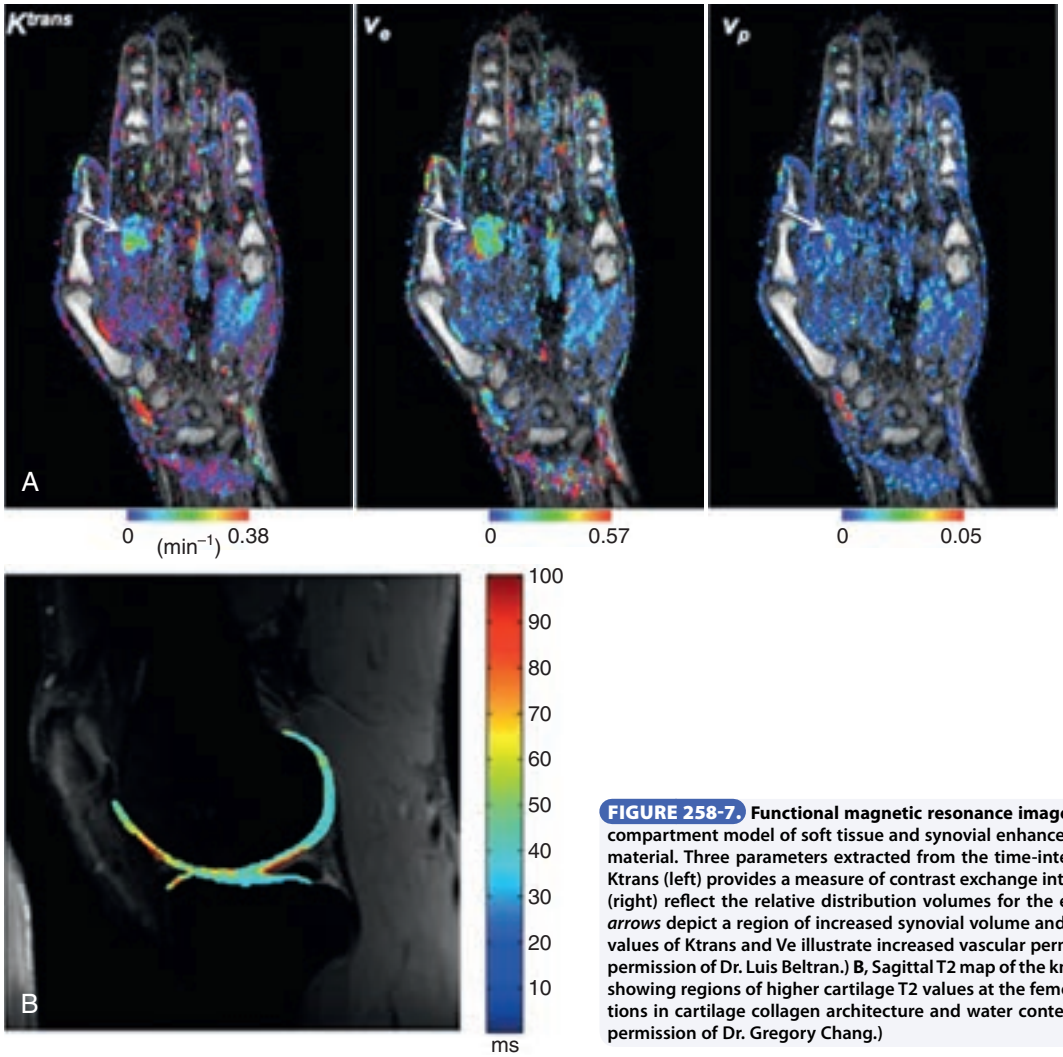

FIGURE 258-7. Functional magnetic resonance image. A, Parametric image derived from fitting a two-compartment model of soft tissue and synovial enhancement after intravenous administration of contrast material. Three parameters extracted from the time-intensity curves are displayed as parametric images: Ktrans (left) provides a measure of contrast exchange into the extravascular soft tissues; Ve (center) and Vp (right) reflect the relative distribution volumes for the extravascular space and plasma, respectively. The *arrows* depict a region of increased synovial volume and enhancement at the second MCP joint. Increased values of Ktrans and Ve illustrate increased vascular permeability at the site of inflammation. (Printed with permission of Dr. Luis Beltran.) **B,** Sagittal T2 map of the knee in which relative T2 relaxation is color encoded, showing regions of higher cartilage T2 values at the femoral condyle and tibial plateau. This reflects alterations in cartilage collagen architecture and water content and possibly early osteoarthritis. (Printed with permission of Dr. Gregory Chang.)

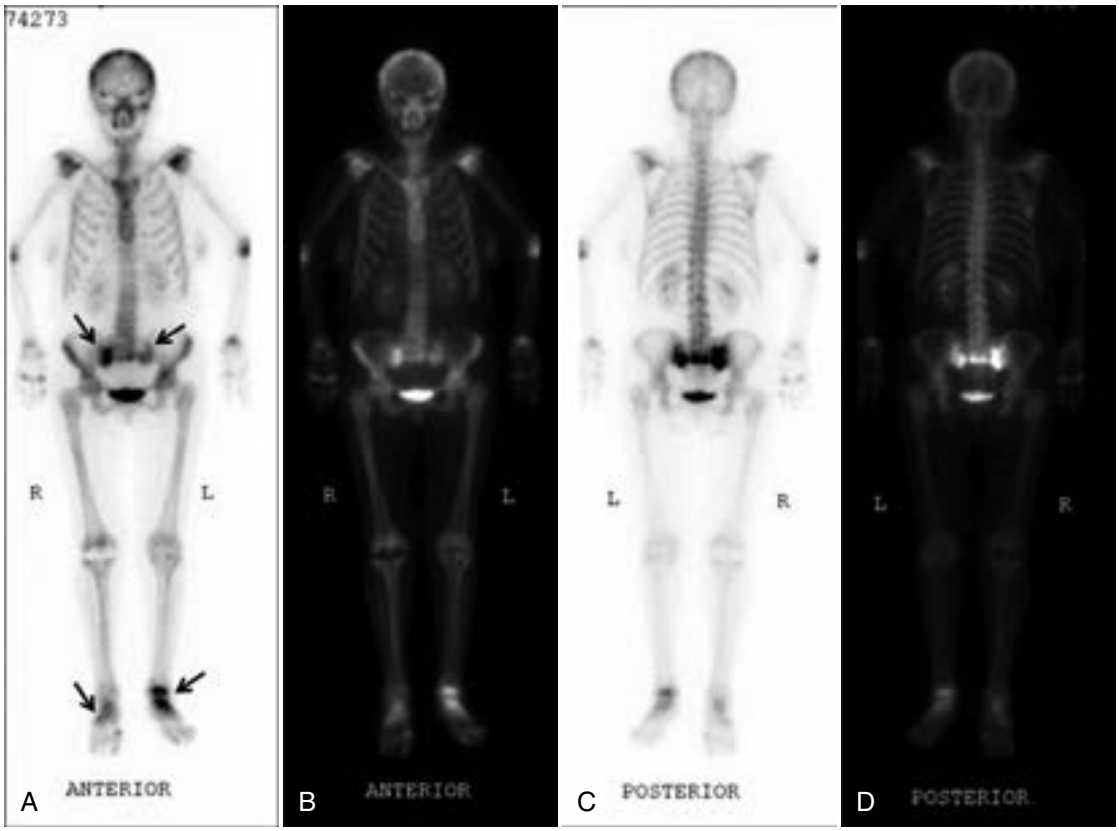

FIGURE 258-8. Rectilinear bone scan in a patient with back pain. Anterior (**A**) and posterior (**B**) delayed images of the axial and appendicular skeleton demonstrate increased tracer uptake in the region of the sacral ala, left ankle, and right midfoot (*arrows*). Follow-up radiographs confirmed the presence of bilateral sacral ala fractures. Note that the central pooling of tracer in the expected location of the urinary bladder is normal. Bone scans provide a sensitive but nonspecific method to evaluate the appendicular and axial skeletal. Increased uptake in the feet in this patient was attributed to degenerative change.

near-simultaneous detection of the photons (coincidence counting) provides an estimate of source tracer concentration. Newer PET scanners are often used in combination with either CT or MRI to achieve improved spatial registration, allow accurate estimates of soft tissue attenuation, provide high-quality anatomic images, and quantify metabolic activity.[9] Combined PET-CT or PET-MRI provides high-resolution images of abnormal metabolic activity and may ultimately provide the most definitive maps of inflammatory activity in patients with rheumatic disease.[10] Early results to date have been promising and are expected to provide sensitive evaluation of the response to DMARDs in patients with inflammatory arthritis.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Haavardsholm EA, Lie E, Lillegraven S. Should modern imaging be part of remission criteria in rheumatoid arthritis? *Best Pract Res Clin Rheumatol*. 2012;26:767-785.
2. Kang T, Horton L, Emery P, et al. Value of ultrasound in rheumatologic diseases. *J Korean Med Sci*. 2013;28:497-507.
3. Sofka CM. Tracking rheumatic disease through imaging. *Rheum Dis Clin North Am*. 2013;39: 633-644.
4. Desai MA, Peterson JJ, Garner HW, et al. Clinical utility of dual energy CT for evaluation of tophaceous gout. *Radiographics*. 2011;31:1365-1375.
5. Mandl P, Kurucz R, Niedermayer D, et al. Contributions of ultrasound beyond clinical data in assessing inflammatory disease activity in rheumatoid arthritis: current insights and future prospects. *Rheumatology (Oxford)*. 2014;53:2136-2142.
6. Rowbotham EL, Wakefield RJ, Granger AJ. The technique and application of ultrasound in the diagnosis and management of inflammatory arthritis. *Semin Musculoskel Radiol*. 2012;16:360-366.
7. Palmer AJR, Brown CP, McNally EG, et al. Non-invasive imaging of cartilage in early osteoarthritis. *Bone Joint J*. 2013;95-B:738-746.
8. McQueen FM. MRI in rheumatoid arthritis: a useful tool for the clinician? *Postgrad Med J*. 2014;90:332-339.
9. Yamashita H, Takahashi H, Kubota K, et al. Utility of fluorodeoxyglucose positron emission tomography/computed tomography for early diagnosis and evaluation of disease activity of relapsing polychondritis: a case series and literature review. *Rheumatology (Oxford)*. 2014;53: 1482-1490.
10. Gotthardt M, Bleeker-Rovers CP, Boerman OC, et al. Imaging 0f inflammation by PET, conventional scintigraphy and other imaging techniques. *J Nucl Med Technol*. 2013;41:157-169.

REVIEW QUESTIONS

1. With regard to imaging of gout, the following statement(s) is (are) true:

A. The presence of periarticular articular erosions and soft tissue nodules are pathognomonic.
B. Osteopenia is a hallmark of the disease.
C. Dual-energy computed tomography can be of value to assess disease extent.
D. It has a characteristic appearance on ultrasound in these patients.
E. C and D are correct.

Answer: E Dual-energy CT is well suited to assess the distribution of tophaceous deposits in patients with gout because of differences of the absorption characteristics of sodium urate crystals compared with other types of deposition diseases. On ultrasonography, calcification along the superficial margin of the cartilage is characteristic of gout, giving rise to a double-line sign.

2. A patient presents with radial-sided wrist pain that extends proximally into the forearm. What imaging test(s) would assess possible etiologies?

A. Radiographs alone should be adequate
B. MRI
C. Ultrasonography
D. B and C are correct
E. A and B are correct

Answer: D The clinical history suggests DeQuervain's tenosynovitis. Radiographs would allow assessment of the osseus structures but not the adjacent tendons. Although MRI would provide the most complete examination of the radial-sided structures, ultrasonography is well suited to evaluating the soft tissues, including the first dorsal compartment tendons. One also could perform ultrasound-guided therapy at the time of diagnosis.

3. A patient is suspected of having sacroiliitis. What imaging study would be appropriate to initially evaluate the patient?

A. Radiographs of the SI joints
B. Computed tomography
C. Nuclear scintigraphy (bone scan)
D. MRI
E. Ultrasonography

Answer: A Dedicated radiographic views of the sacroiliac joints should be the first study ordered to assess for possible erosions and may be sufficient to establish the diagnosis. Computed tomography would provide a more sensitive evaluation for subtle erosions as well as the adjacent soft tissues, particularly with the addition of intravenous contrast. Computed tomography provides an ideal method to perform guided therapy or aspiration. A bone scan would provide a sensitive evaluation of the SI joints but findings would be likewise abnormal for trauma, inflammatory, or degenerative etiologies. A bone scan would allow a global assessment of the axial and appendicular skeleton to determine whether other sites are potentially affected. MRI allows the best assessment of the bone, adjacent marrow space, and soft tissues.

4. A patient has shoulder pain and gives a history of prior dislocation. There is an equivocal abnormality on the humeral head on radiographic evaluation. Which additional study should be considered?

A. Ultrasonography to rule out a rotator cuff tear
B. Noncontrast computed tomography
C. Additional specialized radiographs to evaluate the scapula
D. Direct MR arthrography
E. Nuclear scintigraphy

Answer: D Although ultrasonography could evaluate the rotator cuff, it does not provide adequate assessment of the capsular labral complex. Cross-sectional imaging with intra-articular contrast would best accomplish this. MR arthrography would be optimal. Direct arthrography has been the method of choice in assessing the glenoid labrum, surrounding ligaments, and capsule. CT arthrography is of value, particularly in the postoperative shoulder, and provides indirect imaging of internal structures by coating them with contrast material. Noncontrast CT would be very limited in assessing the labroligamentous complex, even in the presence of a joint effusion caused by poor soft tissue contrast.

5. A patient with early rheumatoid arthritis is being considered for placement on a DMARD. All of the imaging studies below could assess the level of disease activity and response to therapy *except*

A. ^{18}FDG scan.
B. radiographs of the hands and wrists.
C. gray-scale ultrasonography with power Doppler.
D. MRI with gadolinium.
E. parametric MR imaging of distribution volumes of contrast in the extravascular space.

Answer: B Radiographic findings in rheumatoid arthritis usually occur when there has already been irreversible joint damage. Ideally, therapy would be instituted on a radiographically negative patient. The remaining examinations can provide sensitive evaluation of disease activity before the development of either bone or cartilage erosion.

259

CONNECTIVE TISSUE STRUCTURE AND FUNCTION

SUNEEL S. APTE

Connective tissues are those having a primary mechanical or structural role. They are typically skeletal or osteoarticular tissues such as bone, cartilage, tendon, ligament, fascia, intervertebral disc, and joint-associated structures such as synovium and fibrocartilaginous menisci, although adipose tissue is also included (Table 259-1). However, all viscera, glands, and vascular and nervous tissue contain varying amounts of connective tissue elements that maintain proper spatial relationships between their cells, organize them into functional units, and provide an internal fibrous skeleton or external protective capsule. These connective tissue elements primarily comprise extracellular matrix (ECM), which is organized as a basement membrane or an interstitial matrix. Examples include elastic fibers in the aorta and lung that mediate stretch and recoil, glomerular basement membrane that participates in filtration, and the fibrous endoskeleton that connects the heart valves and transmits myocardial contraction. Indeed, heart valves and chordae tendinae have regional structural similarities to cartilage and tendons.

In addition to their structural role, connective tissues are involved in storage and activation of growth factors, cytokines, and morphogens. Some connective tissue molecules mediate inflammation when released as intact molecules, or may acquire new properties as proteolytically derived fragments (matrikines).[1] Connective tissues such as bone, cartilage, and tendon harbor mesenchymal stem-like cells (MSC) that undergo expansion and differentiation during growth, repair, and regeneration and are therapeutically valuable. Bone contributes to calcium and phosphate homeostasis. A broad perspective on connective tissue is thus crucial for understanding the complex clinical presentation of inherited connective tissue disorders, pathways of tissue repair and regeneration, and pathogenesis of degenerative and autoimmune conditions.

CONNECTIVE TISSUE CELLS, THEIR EXTRACELLULAR MATRIX AND TISSUE-SPECIFIC BIOMECHANICS

Connective tissues comprise cells, typically derived from mesoderm, or in the craniofacial region, derived from the neural crest, and the ECM they secrete and assemble around them. Unique morphologic and biosynthetic properties of the differentiated cells specify osteoblasts, osteoclasts, and osteocytes (in bone); chondrocytes (in cartilage); and tenocytes (in tendons), as well as mixed phenotypes such as cells of fibrocartilage (in the intervertebral disc, some tendon and ligament insertions, and joint menisci). Chondrocytes alter their biosynthetic profile as they terminally differentiate into hypertrophic chondrocytes in growth plate cartilage, downregulating aggrecan and collagen II expression and synthesizing collagen X as a specialized product.[2] As the growth plate transitions into bone, the hypertrophic cells die, blood vessels invade the cartilage, and the cartilage matrix is gradually replaced by bone. Fibroblasts, which secrete collagens as a major product, are the dominant cells in tendons, fascia, ligaments, and dermis. The abundant ECM of connective tissues is their major defining characteristic, but in other tissues and organs, fibroblasts are relatively quiescent, and the interstitial ECM is less organized and not as abundant. However, under appropriate stimuli, fibroblasts transition to highly contractile, biosynthetically active myofibroblasts, which are associated with hypertrophic scars and fibrosis.

During embryonic development, undifferentiated connective tissue cells (mesenchymal cells) produce a hydrated, loose, provisional ECM optimal for migration, branching, and folding of individual cells and cell collectives, such as epithelial sheets. As embryogenesis progresses, provisional ECM is remodeled by matrix-degrading proteinases and replaced by specialized connective tissues whose mechanical properties[3] are better suited to weight bearing, locomotion, and increased circulatory stress. This requirement for increased mechanical strength is not met in the severest forms of inherited connective

TABLE 259-1 DIVERSITY OF CONNECTIVE TISSUES

STRUCTURE	FUNCTION	CELLS	KEY MATRIX COMPONENTS
Adipose tissue	Energy metabolism Physical protection	Adipocytes, fibroblasts, endothelial cells	Collagen I, collagen III, microfibrils, collagen IV, laminin
Basement membranes	Epithelial support Cell polarity Filtration barrier Cell barrier Transparency (lens)	Epithelial and endothelial cells, myotubes, adipocytes, lens fibers	Collagen IV, laminin, nidogens, perlecan
Bone	Structural support Hematopoiesis Mineral storage	Osteoblasts, osteoclasts, osteocytes	Collagen I, osteocalcin, bone sialoprotein, hydroxyapatite
Cartilage	Bone growth Joint motion Load transmission	Chondrocytes	Collagen II, aggrecan
Dermis	Elasticity and resiliency	Fibroblasts	Collagen I, elastin
Ligament	Connects bone to bone	Fibroblasts	Collagen I, SLRPs
Tendon	Connects muscle to bone	Fibroblasts	Collagen I, SLRPs, fibrillins
Visceral stroma	Internal scaffold and capsule	Fibroblasts	Collagen I, collagen III, versican, fibronectin
Synovium	Produces synovial fluid	Fibroblasts, macrophages	Collagen I, collagen III
Vessel wall	Barrier, elastic recoil	Endothelium, vascular smooth muscle cells	Collagen III, collagen IV, elastin, fibrillins, fibulins, versican

SLRP = small leucine repeat-rich protein.

tissue disorders affecting the vasculature, bone, or cartilage, with neonatal, juvenile, or early adult mortality resulting.

Connective tissues continually sense and adapt to their mechanical environment. The anabolic response of many connective tissues to optimal levels of mechanical stress is now recognized as a crucial, remediable determinant of health. This concept is embodied in Wolff's law of bone remodeling, which states that bone remodels in response to the mechanical loads imposed on it. The composition of specific connective tissues reflects adaptations in response to their mechanical environment or requirements for other specialized functions. For example, articular cartilage and other hyaline cartilages comprise proteoglycan aggregates that exert a swelling pressure and a network of collagen II fibrils that restrain them, contributing to compressive strength and shock absorption.[4] The superficial zone of articular cartilage, in contrast, in enriched in collagenous fibrils that are arranged parallel to the surface to resist shear forces. A mucinous glycoprotein, lubricin, present on the joint surfaces, ensures the low coefficient of friction of synovial joints. In tendons, the major ECM component is collagen I, which has high tensile strength, but in bone, a composite of collagen I and calcium hydroxyapatite provides tensile and compressive strength. The contractile apparatus of skeletal muscle is connected by the dystrophin–glycoprotein complex to the muscle basement membrane and via hierarchical assemblies of interstitial matrix (endomysium, perimysium, and epimysium) to tendons and ultimately to bone, thus efficiently transmitting muscle contraction forces.[5] In the intervertebral disc, the nucleus pulposus is rich in aggrecan, which exerts a swelling pressure that is constrained by the surrounding concentric lamellae of fibrillar collagen in the annulus fibrosus. This composite structure absorbs vertical loads on the spine while limiting deformation of the nucleus pulposus. The interfaces between different connective tissues show structured transitions, highlighted by fibrocartilage present at tendon insertion sites and by Sharpey fibers, which seamlessly integrate perimysial collagen of muscle fibers with collagen in bone. In knee menisci, the cells and ECM of the inner third are cartilaginous, with abundant proteoglycans. Whereas the cells of the outer third, which is vascular and connected to ligaments, are fibroblastic, with corresponding predominance of fibrillar collagens, the cells and ECM in the middle third have an intermediate fibrochondrocyte phenotype.

MAJOR EXTRACELLULAR MATRIX COMPONENTS

Extracellular matrix comprises collagens; the glycosaminoglycan (GAG) hyaluronan (HA); proteoglycans; a variety of glycoproteins; phosphoproteins (especially in bones and teeth); and matricellular proteins such as thrombospondin, tenascin, and periostin, which regulate cellular functions via ECM but do not have a structural role. The distribution of ECM molecules and macromolecules is tightly regulated to achieve specific microenvironments that control proliferation, differentiation, polarity, and migration of cells and provide niches for postnatal stem cells. Collagens, the most abundant proteins of the body, comprise several distinct molecules, each containing at least one triple-helical domain composed of αchains having a repeating Gly-X-Y sequence. Twenty-eight different collagen types are formed from the products of more than 40 genes encoding collagen α chains. The presence of glycine, the smallest amino acid, at every third position, permits triple-helix formation; the amino acid proline, which is frequently modified to form hydroxyproline at the Y position, ensures stability of the triple helix. Fibrillar collagens (e.g., types I-III, V, XI) have long triple-helical (collagenous) regions, so they form rodlike structures. Whereas collagen I, the most widely distributed and abundant, is a heterotrimer comprising two α1(I) chains and an α2(I) chain, collagen II and III are homotrimers of α1(II) and α1(III) chains, respectively. Mutations of either collagen I chain cause the majority of cases of osteogenesis imperfect (Chapter 260) and infrequently can also cause rare subtypes of Ehlers-Danlos syndrome (EDS) (Chapter 260). The triple helices of nonfibrillar collagens are shorter and interspersed with interruptions of the Gly-X-Y sequence or noncollagenous domains that introduce flexible regions. Collagen XIII, XVII, XXIII, and XXV are transmembrane proteins. Fibrillar collagens are synthesized as procollagens having bulky terminal propeptides. The folding of the triple helix occurs intracellularly and is propagated from the C- to the N-terminus to form homo- or heterotrimeric triple helices. The propeptides are subsequently excised by specific proteinases, resulting in rodlike tropocollagen that can be assembled into tightly packed, quarter-staggered fibrils. Failure to remove the N-propeptide of collagen I impairs fibril assembly and leads to a specific type of EDS with severe skin fragility (type VIIc or dermatosparactic type). Whereas collagen I is the major component of bone, dermis, tendons, ligaments, and the sclera of the eye, collagen III is abundant in the skin, lung, and vasculature, explaining the association of genes encoding these collagens with osteogenesis imperfecta and vascular EDS (also known as type IV). Collagen II, together with minor amounts of collagen IX and collagen XI, predominates in cartilage, ocular vitreous and the nucleus pulposus of the intervertebral disc, and mutations of either of these collagen genes can cause the Stickler syndrome (triad of cartilage, eye, and hearing anomalies). Collagen fibrils can consist of more than one collagen type; for instance, collagen V is found in heterotypic fibrils with collagen I and is required for nucleation of fibrillogenesis to form large collagen I fibrils.

Basement membranes are formed from collagen IV, with significant content of laminins, nidogens, and the heparin sulfate proteoglycan perlecan, which is also an important constituent of cartilage. Basement membranes have crucial roles in regulating molecular transport, such as in glomerular filtration, and in establishment and maintenance of epithelial polarity. Collagen VI is widely distributed as beaded microfibrils that form pericellular matrices in fibroblasts and, alongside collagen VII anchoring fibrils, connect basement membranes to interstitial ECM. Hyaluronan, a polymer with repeating disaccharide units, can achieve a molecular weight in the millions of Daltons. It is a key component of cartilage matrix, pericellular matrix of many cell types, and a mediator of inflammation. Proteoglycans are characterized by covalent attachment of glycosaminoglycans, such as chondroitin, heparan, and keratan sulfate, to a core protein. Aggrecan and versican are large chondroitin sulfate proteoglycans that form giant aggregates with hyaluronan in chondrocytes or nucleus pulposus cells and fibroblasts, respectively. On the other hand, some small leucine repeat-rich proteins (SLRPs) such as decorin, biglycan, and lumican are proteoglycans that may have only one or two glycosaminoglycan chains. This class of molecules, which also includes fibromodulin, interacts with collagen fibrils to cross-link them and regulate fibril diameter. They have been shown to bind transforming growth factor β (TGF-β) and, when released from connective tissue, to provide danger signals to the immune system. During connective tissue healing, fibrin and fibronectin provide a transitional ECM permissive for cell migration, differentiation, and other aspects of the repair process.

Elastic fibers are formed by coacervation of a soluble precursor named tropoelastin. Elastic fiber assembly is guided by tissue microfibrils that are formed from three large glycoproteins named fibrillins.[6] Fibrillin-1 is abundant in the aorta, ocular zonule, and perichondrium of bone and is mutated in Marfan syndrome (Chapter 260), resulting in aneurysms, ectopia lentis, and skeletal overgrowth. Fibrillin-2 mutations cause Beals syndrome, with skeletal overgrowth and limb contractures as major features but typically, not severe cardiac or eye problems. The elastic fiber-microfibril network additionally contains versican, microfibril-associated glycoproteins, latent TGF-β–binding proteins, and fibulins. Cutis laxa, which primarily affects the skin but can also involve vasculature and internal organs, can be caused by mutations affecting elastin and fibulin-5.

Most ECM molecules undergo one or more posttranslational modifications. These include enzymatic formation of intra- and interchain disulfide bonds, phosphorylation, and several kinds of glycosylation, including addition of N- and O-linked sugars or glycosaminoglycans such as chondroitin, keratan, or heparan sulfates. Whereas N-linked sugars on some proteins are essential for protein folding and secretion, glycosaminoglycan (GAG) chains provide crucial biophysical properties (such as of aggrecan) and mediate intermolecular interactions. During its biosynthesis, collagen undergoes lysyl and prolyl hydroxylation. Some lysyl and hydroxylysl residues are modified extracellularly by lysyl oxidase and form stable cross-links between adjacent collagen molecules, which strengthens bone, tendons, and skin.[7] Gene defects affecting collagen-modifying enzymes or subunits of the molecular complexes they operate in lead to various recessive forms of osteogenesis imperfecta or EDS. L-Ascorbic acid is a cofactor for lysyl hydroxylase and prolyl hydroxylase and stimulates procollagen synthesis. Its nutritional deficiency leads to scurvy by reducing collagen synthesis, triple-helical stability, and tropocollagen cross-linking. Lathyrism (a now rarely seen neurotoxic disease) is caused by excess ingestion of β-aminopropionitrile, which inhibits lysyl oxidase–mediated formation of lysine aldehydes, which are the precursors of the major collagen and elastin cross-links, and leads to connective tissue fragility. Copper is a required cofactor for lysyl oxidase, and its deficiency can also lead to lathyrism.

CELL-MATRIX INTERACTIONS IN CONNECTIVE TISSUE REGULATION

Cellular ECM receptors mediate cell-matrix interactions that ensure force transmission from cells to ECM[8]; the dynamic reciprocity between cells and

ECM is crucial for environmental sensing and adaptive responses. Receptors provide feedback to the cells regarding the quality, content, or mechanical properties of the matrix (outside-in signaling) to generate appropriate responses, including cell proliferation or migration, or lead to altered ECM synthesis or degradation. Alternatively, cytoplasmic signals may alter cell-matrix interactions (inside-out signaling). Integrins are a large group of heterodimeric receptors with distinct binding preferences that comprise α and β subunits having short cytoplasmic domains. Integrins are crucial for fibronectin assembly, which in turn substantially influences assembly of collagen, fibrillins, and TGF-β activation. Intracellular signals activate integrins to promote high-affinity binding to ECM proteins, and binding of ECM proteins via specific integrin-binding motifs such as Arg-Gly-Asp initiates clustering of cell adhesion complexes and intracellular signaling that, among diverse effects, may elicit the production of inflammatory cytokines. Integrin αvβ6 transmits cellular traction to matrix-bound latent TGF-β, exposing the active growth factor. Because osteoclast attachment to and spreading on bone surfaces is dependent on $\alpha_v\beta_3$, it is being targeted for treatment of osteoporosis (Chapter 243) using chemical antagonists or blocking antibodies. Syndecans are transmembrane heparan sulfate proteoglycans that work as coreceptors alongside high-affinity ECM receptors such as integrins. Discoidin domain receptors are receptor tyrosine kinases activated by binding to native triple-helical collagen.[9] DDR1 and DDR2 bind to fibrillar collagens I-III and collagen V. Collagen IV activates DDR1 but not DDR2, whereas collagen X activates DDR2. With their extended structure, extensive post-translational modification, and exposure of numerous binding sites, ECM molecules such as collagens, fibrillins, and the classical cell-binding protein fibronectin bind to multiple receptor types. Among several hyaluronan-binding molecules on the cell surface, including hyaluronan synthases, CD44 is a major hyaluronan receptor with a role in assembling the pericellular matrix and in inflammation.

EXTRACELLULAR MATRIX NETWORKS AND CONNECTIVE TISSUE DISORDERS

Extracellular matrix molecules and macromolecules are assembled to form higher order (supramolecular) complexes and networks whose varied composition and geometry further diversify connective tissues. These are exemplified by parallel arrays of extensively cross-linked collagen fibrils conferring high tensile strength to tendons, ligaments, and bone, by the multidirectional or "basket-weave" arrangement of collagen in dermis, both permitting and constraining multiaxial mobility, and by the concentric elastic lamellae and crimped collagen in the aorta that regulate hemodynamics. In the eye, orthogonally oriented collagen fibrils in corneal stroma and the lattice-like structure of collagen VIII in the Descemet membrane allow transparency and are crucial for normal vision.

Cartilage and nucleus pulposus ECM comprises large aggregates of hyaluronan and aggrecan, with hundreds of aggrecan molecules attached to each hyaluronan polymer via their N-terminal domain, leaving the C-terminal domain free for interactions with fibrillins, fibulins, and other ECM networks. Each aggrecan molecule has approximately 100 chondroitin sulfate chains and several keratan sulfate chains. Their high fixed-charge density creates an osmotic environment that favors water retention and restricts water flux in cartilage. Through the attachment of hyaluronan to cell-surface receptors, the aggrecan–hyaluronan network is retained by the chondrocyte as a conspicuous pericellular matrix (or glycocalyx). In connective tissue cells other than chondrocytes, pericellular matrices use versican or heparin sulfate proteoglycans instead of aggrecan. The pericellular matrix occupies the crucial interface between cells and their environment and influences cell behavior. It is a provisional ECM in microcosm because it is metabolically more active than further removed matrix, which is more stable. Some interstitial collagen and elastin can be stable for decades.

Connective tissue disorders arise from inborn defects affecting cells or ECM, as well as cellular dysfunction caused by repetitive injury, inflammation, or metabolic or aging processes. Because ECM is such a crucial component of connective tissues, most inherited and acquired connective tissue disorders result from mutant ECM molecules, insufficient but essentially normal matrix (e.g., osteoporosis), or excess or inappropriately deposited matrix (e.g., adhesions, fibrosis, and scleroderma). Osteoarthritis (OA) involves all structures that form the joint, including the synovium and subchondral bone, and appears to result from a combination of genetic and nongenetic factors (Chapter 262). OA of the hip and hands has a stronger genetic component than OA of the knee. Because the function of cartilage is to absorb impact and distribute it to bone without aberrant loading of joint structures, variations in genes that affect joint congruity and alignment or cartilage or meniscal integrity could predispose to OA. For instance, variations in genes encoding cartilage ECM components mutated in chondrodysplasias (Chapter 205), such as collagen II, IX, cartilage oligomeric protein, or matrilins, have been associated with early-onset OA, suggesting a genetic link and a spectrum extending from OA predisposition at the milder end to severe chondrodysplasia. Another genetic link is with the mechanisms that influence the formation of joints, correlating with the association between joint malalignment and OA. Variations in genes associated with the TGF-β superfamily pathway, including *GDF5* (encoding a morphogen required for joint development), *ASPN* (encoding an ECM protein that binds TGFβ), and SMAD3 (encoding a cytoplasm-nucleus signaling intermediary), are associated with OA.[10] Factors that affect load distribution across joints, such as ligament and meniscal tears or hip dysplasia, predispose to OA. Cartilage breakdown in arthritis is not caused by wear and tear but by altered cell-matrix interactions that lead to enzymatic digestion. Loss of aggrecan, followed by loss of SLRPs and other molecules from the surface of collagen fibrils, are initial changes that render collagen fibrils susceptible to destruction by collagenases. The loss of articular cartilage collagen II is thought to constitute an irreversible change in joint disease. Whatever the original insult, the response of chondrocytes or synovium can lead to a vicious cycle of joint destruction because the cells may respond by producing excess ECM and ECM-degrading proteases, with release of both intact and fragmented molecules that may further potentiate inflammation. Rodent immune arthritis models have revealed a role for the alternative complement pathway in arthritis, and cartilage proteins and their fragments released by proteolytic breakdown can have complex effects on the complement pathways that lead to both activation and suppression.

Cell-matrix interactions are profoundly affected in muscular dystrophies (Chapter 421), revealing a continuum that is essential for force transmission from the cytoskeleton to interstitial matrix. Mutations affecting components of the dystrophin–glycoprotein complex including dystroglycan, a receptor that connects muscle cytoskeleton to laminin in the muscle basement membrane, muscle basement membrane (laminin and collagen IV), and collagen VI filaments in ECM, cause a variety of muscular dystrophies. Specific adhesion complexes such as hemidesmosomes in the epidermis mediate anchorage to basement membranes and underlying interstitial ECM. Autoantibodies against collagen XVII and VII, or laminin 5 mutations, affect major components of these complexes and lead to blistering skin diseases (Chapter 439).

PROTEASES AND CONNECTIVE TISSUE TURNOVER

Reflecting continuous adaptation to their environment, connective tissues have intrinsic mechanisms that ensure ECM renewal (i.e., through coupled synthesis and degradation). ECM molecules are substrates for several proteinase classes, chiefly matrix metalloproteinases (MMPs), astacin MMPs such as BMP1 and tolloids, cathepsins, and A disintegrin-like and metalloproteinase domain with thrombospondin type 1 repeats (ADAMTS).[11] Although these proteases are typically considered to be catabolic, some are also essential for the maturation of precursor proteins, such as ADAMTS2, which excises the amino-propeptide of procollagen I, II, and III and is defective in EDS, dermatosparactic type. BMP1, which excises the C-propeptides of procollagen I, II, and III and cleaves lysyl oxidase and several other proteins, is deficient in a type of recessive osteogenesis imperfecta. Although most MMPs are secreted, a class of membrane-type MMPs is cell-surface bound. Most MMPs require proteolytic activation by other MMPs or serine proteinases such as plasmin and furin, and are inhibited by tissue inhibitors of MMPs (TIMPs) and α_2-macroglobulin. Whereas MT1-MMP is crucial for collagen I proteolysis in bone, MMP-13 has been implicated as a major collagen II degrading enzyme in arthritis. Cell surface MMPs such as MT1-MMP and A disintegrin-like and MMPs (ADAM) are responsible for ectodomain shedding of ECM receptors, cell-surface cytokines and cytokine receptors, together regulating a variety of inflammatory and oncogenic situations. ADAMTS4 and ADAMTS5, also known as aggrecanase-1 and -2, respectively, are principal aggrecan-degrading proteases implicated in OA and, together with MMP-13, are potential drug targets in this disorder. However, MMPs, ADAMs, and ADAMTSs have structurally similar catalytic domains and zinc and calcium-dependent proteolytic mechanisms, which renders selective inhibition of any single proteinase from these classes challenging. Indeed, a major side effect of MMP inhibitors used in clinical trials for cancer was connective tissue stiffness and inflammation, presumably resulting from reduced physiological turnover of collagen and other

ECM proteins. Among other proteinases, cathepsin K produced by osteoclasts is active at acidic pH, unlike MMPs, and therefore efficiently digests bone collagen in the acidic osteoclast–bone interface. Neutrophil elastase (a serine proteinase), MMP-9, and MMP-12 (metalloelastase) can degrade elastin. Proteolysis of ECM proteins can release bioactive fragments (matrikines) such as endostatin (from collagen XVIII) or endorepellin (from perlecan), which are antiangiogenic; some fibronectin fragments released in OA are proinflammatory. Hydroxyproline antibodies that recognize ECM-fragments with specific cleaved ends (neoepitope antibodies), as well as cross-linked collagen fragments released by proteolytic activity, are useful biomarkers of bone and cartilage turnover.

AGING OF CONNECTIVE TISSUE

Some visible hallmarks of aging result not only from cellular senescence but also reduced ECM synthesis and increased catabolism, as well as greater connective tissue fragility. This leads, for example, to reduced bone mass or osteoporosis; thinning and loss of elastic properties of dermal ECM, which are visible as wrinkles and sagging skin; reduced volume of intervertebral discs; and increased capillary fragility. Loss of HA and GAGs reduces tissue hydration and of collagen and elastin reduces tensile strength and elasticity, respectively. Products of ECM proteolysis stimulate cells to release free radicals and cytokines that further accelerate ECM breakdown or lead to cell death. Extrinsic factors such as sunburn and ultraviolet irradiation induce cytokines that accelerate this process. In addition, aging collagen can be cross-linked by the Maillard reaction, especially when glucose levels are high, resulting in its modification by advanced glycation end products (AGE), which renders it inflexible, alters the rate of turnover, and affects binding to matrix receptors.[12] AGE binding to a cellular receptor has been shown to lead to cellular dysfunction, oxidative stress, and inflammation.

In summary, connective tissue provides the framework of the musculoskeletal system and a structural scaffold for internal organs while serving as a reservoir for growth factors, cytokines, and stem-like precursor cells. Inherited connective tissue disorders have the potential to disturb these functions in specific ways. Acquired disorders such as OA and fibrosis reflect perturbation of diverse physiological networks and pathways, with altered cell-matrix interactions at the center.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Frey H, Schroeder N, Manon-Jensen T, et al. Biological interplay between proteoglycans and their innate immune receptors in inflammation. *FEBS J*. 2013;280:2165-2179.
2. Fosang AJ, Beier F. Emerging frontiers in cartilage and chondrocyte biology. *Best Pract Res Clin Rheumatol*. 2011;25:751-766.
3. Nandadasa S, Foulcer S, Apte SS. The multiple complex roles of versican and its proteolytic turnover by ADAMTS proteases during embryogenesis. *Matrix Biol*. 2014;35:34-41.
4. Heinegard D, Saxne T. The role of the cartilage matrix in osteoarthritis. *Nat Rev Rheumatol*. 2011;7:50-56.
5. Carmignac V, Durbeej M. Cell-matrix interactions in muscle disease. *J Pathol*. 2012;226: 200-218.
6. Baldwin AK, Simpson A, Steer R, et al. Elastic fibres in health and disease. *Expert Rev Mol Med*. 2013;15:e8.
7. Byers PH, Pyott SM. Recessively inherited forms of osteogenesis imperfecta. *Annu Rev Genet*. 2012;46:475-497.
8. Ross TD, Coon BG, Yun S, et al. Integrins in mechanotransduction. *Curr Opin Cell Biol*. 2013; 25:613-618.
9. Fu HL, Valiathan RR, Arkwright R, et al. Discoidin domain receptors: unique receptor tyrosine kinases in collagen-mediated signaling. *The J Biol Chem*. 2013;288:7430-7437.
10. Sandell LJ. Etiology of osteoarthritis: genetics and synovial joint development. *Nat Rev Rheumatol*. 2012;8:77-89.
11. Gargiulo S, Gamba P, Poli G, et al. Metalloproteinases and metalloproteinase inhibitors in age-related diseases. *Curr Pharm Des*. 2014;20:2993-3018.
12. Semba RD, Nicklett EJ, Ferrucci L. Does accumulation of advanced glycation end products contribute to the aging phenotype? *J Gerontol A Biol Sci Med Sci*. 2010;65:963-975.

260

INHERITED DISEASES OF CONNECTIVE TISSUE

REED E. PYERITZ

MUCOPOLYSACCHARIDOSES

DEFINITION

Proteoglycans are ubiquitous components of the extracellular matrix (ECM) and the surfaces of cells, and they are among the largest and most complex of human molecules. Proteoglycans consist of a protein core to which are covalently bound glycosaminoglycans (GAGs; formerly called mucopolysaccharides) of several types: dermatan sulfate, heparan sulfate, keratan sulfate, and chondroitin sulfate. These four polymeric molecules are cleaved from their protein core in lysosomes; then they, plus hyaluronan (a GAG lacking a protein core), are catabolized further in lysosomes in a stepwise fashion by more than a dozen enzymes. Genetic defects in any one of these enzymes lead to the accumulation of GAG metabolites in lysosomes, with profound disruption of cellular physiology. The phenotypes resulting from deficiencies of these catabolic enzymes are termed *mucopolysaccharidoses* (MPSs) and are classified into seven types (Table 260-1). Several additional storage disorders, termed *mucolipidoses* (MLs), are caused by a genetic defect in posttranslational modification of lysosomal enzymes and share features with the MPSs.

EPIDEMIOLOGY

All MPS disorders are rare, each with an incidence of one or fewer cases per 100,000 births, and are without ethnic predilection.

PATHOBIOLOGY

With the exception of MPS II (Hunter syndrome), which is X-linked, each of these disorders is autosomal recessive. All MPSs are caused by deficiency of a single lysosomal enzyme responsible for a specific step in GAG metabolism. Catabolism of GAG proceeds normally until the step requiring the defective enzyme, when further normal metabolism halts. Although a minor degree of nonspecific breakdown occurs, resulting in urinary excretion of cleaved GAG that can be useful diagnostically, the accumulation of GAG within lysosomes of cells of mesenchymal origin; endothelium; and, in most cases, neurons causes widespread, progressive cellular dysfunction and clinical effects. Lysosomal enzymes are targeted to lysosomes by posttranslational addition of mannose 6-phosphate. Deficiency of the phosphotransferase that catalyzes the first step in this reaction results in an inability to catabolize any GAG molecules. The catabolic enzymes, which normally would be transported into lysosomes, instead are secreted from the cell and are found in unusually high concentrations in plasma, providing one diagnostic test for MLs.[1]

Pathology

All pathologic manifestations of MPS and ML disorders worsen with age, and some are present from early developmental stages. Gross anatomic hallmarks are hepatosplenomegaly, marked skeletal alterations (termed *dysostosis multiplex*)[2] that result in short stature and thoracic cage deformity, thickening and narrowing of airways and arteries, and coarsening of facial features. Although mental retardation is a prominent feature of some of these conditions, the brain may show only ventriculomegaly secondary to communicating hydrocephalus. On microscopy, mesenchymal cells show a cytoplasm full of apparently empty vacuoles; these are lysosomes from which GAG has been removed by fixation. Cells cultured from patients show greatly enlarged lysosomes filled with granular material. In the severe form of ML, dense inclusions are present, which gave rise to the common name, *I-cell disease.*

CLINICAL MANIFESTATIONS

Each of the disorders in Table 260-1 shows a wide spectrum of clinical severity. This wide spectrum has led to a classification that gives the impression of separate disorders within some of the MPS and ML types, but these represent the apparent ends of the continuum. Some of the disorders result in death by adolescence (Hurler syndrome, severe Hunter syndrome, ML II), but others are commonly compatible with survival to adulthood. The latter group of disorders is emphasized here.

The milder end of the MPS I spectrum, Scheie syndrome, may not be diagnosed until adulthood; patients present with stiffened joints, corneal clouding and glaucoma, carpal tunnel syndrome, and aortic valvular disease. Stature and intelligence are not affected. The main health risks are valvular involvement, thickening of meninges that can produce a myelopathy, and thickening of the upper airways that can produce obstructive symptoms and sleep apnea.

The milder form of MPS II, Hunter syndrome, is distinctive because it is X-linked (affecting males almost exclusively), and the cornea shows little overt clouding. Cervical myelopathy, obstructive airway disease, and cor pulmonale are important concerns. A combined conductive and neurosensory hearing loss is common.

Neither MPS IV (Morquio syndrome) nor MPS VI (Maroteaux-Lamy syndrome) affects intelligence. Both syndromes often are associated with severe skeletal changes, which are distinct radiographically but produce similar problems of kyphoscoliosis, pectus carinatum, restrictive lung disease, severe short stature, and joint degeneration. Cervical myelopathy resulting from a thickened dura is common to both disorders and is accentuated by odontoid hypoplasia in MPS IV. Thickening of the aortic and mitral valves may produce severe dysfunction necessitating their replacement. General anesthesia is especially hazardous because of the narrow upper and middle airways and cervical instability.

Patients with ML III (pseudo-Hurler polydystrophy) resemble patients with MPS VI but often have mild to moderate mental retardation. Aortic regurgitation is common.

TABLE 260-1 MUCOPOLYSACCHARIDOSES AND MUCOLIPIDOSES

TYPE	EPONYM OR COMMON NAME	CLINICAL FEATURES	INHERITANCE	OMIM*	ENZYMATIC DEFECT
MPS IH	Hurler syndrome	DM and short stature; MR; corneal clouding; HS; heart disease; death in childhood	AR	252800	α-L-iduronidase
MPS IS	Scheie syndrome	Coarse facies; stiff joints, corneal clouding; aortic valve disease; normal intelligence and lifespan	AR	252800	α-L-iduronidase
MPS II	Hunter syndrome	Severe form: coarse facies, DM and short stature, HS; MR; no corneal clouding; death by late adolescence Mild form: coarse facies, short stature; normal intelligence; survival to adulthood	XL	309900	Iduronate sulfatase
MPS IIIA	Sanfilippo A	Severe MR and hyperactivity; mild somatic changes	AR	252900	Heparan *N*-sulfatase
MPS IIIB	Sanfilippo B	Same as MPS IIIA	AR	252920	α-*N*-acetylglucosaminidase
MPS IIIC	Sanfilippo C	Same as MPS IIIA	AR	252930	Acetyl-coenzyme A: α-glucosaminide acetyltransferase
MPS IIID	Sanfilippo D	Same as MPS IIIA	AR	252940	*N*-acetylglucosamine 6-sulfatase
MPS IVA	Morquio A	Short stature and distinct skeletal dysplasia with odontoid hypoplasia and myelopathy; corneal clouding; normal intelligence; valvular heart disease	AR	253000	Galactose 6-sulfatase
MPS IVB	Morquio B	Same as MPS IVA	AR	253010	β-Galactosidase
MPS VI	Maroteaux-Lamy	DM and short stature; corneal clouding; normal intelligence; aortic stenosis; leukocyte inclusions; hydrocephalus in severe form	AR	253200	*N*-acetylgalactosamine
MPS VII	Sly syndrome	DM; HS; widely variable, including MR	AR	253220	β-Glucuronidase
MPS IX	—	Short stature; periarticular soft tissue masses	AR	601492	Hyaluronidase
ML II	I-cell disease	Similar to but more severe than MPS IH but with cellular inclusions; no mucopolysaccharidurria	AR	252500	UDP-*N*-acetylglucosamine: lysosomal enzyme *N*-acetylglucosaminyl-1- phosphotransferase
ML III	Pseudo-Hurler polydystrophy	Short stature and mild DM; stiff joints, mild MR; survival to adulthood	AR	252500	Same as ML II arthropathy, coarse facies; variable but milder

*Entries in Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute of Genetic Medicine. Baltimore: Johns Hopkins University. http://omim.org.
AR = autosomal recessive; DM = dysostosis multiplex; HS = hepatosplenomegaly; MR= mental retardation; XL = X-linked.

DIAGNOSIS

Differential Diagnosis

Diagnosis of these conditions is difficult in young children, before most of the clinical features have progressed, but should be considered in any person with hepatosplenomegaly and coarsening of the facial features. Evaluation requires a pedigree analysis, ophthalmologic examination, skeletal radiographic survey, echocardiography, and analysis of the urine for excretion of GAGs. Often the specific MPS is evident from radiographs, the presence or absence of corneal clouding, and the pattern of mucopolysaccharidurria. Enzymatic analysis of leukocytes confirms the diagnosis. Patients with MLs do not show mucopolysaccharidurria but have marked elevation of all the GAG catabolic lysosomal enzymes in plasma.

TREATMENT Rx

Ventriculoperitoneal shunting is necessary if intracranial pressure is elevated. Close attention to hearing and visual problems is essential throughout life. Many adults with MPS or ML require surgery for carpal tunnel syndrome. Cardiovascular surgery for valvular or coronary disease may be necessary. All use of anesthesia is high risk because of the narrow airways and, in the case of MPS IV, atlantoaxial instability. For patients who remain ambulatory, selective joint replacement can be beneficial. Because of the morbidity associated with thoracic cage deformity, consideration should be given to stabilizing the spinal deformity before it becomes severe.

Replacement of the deficient enzyme via intravenous infusion is being studied for most of the MPS disorders. Laronidase (Aldurazyme) has been approved in the United States for treatment of MPS I. An infusion every 2 weeks for 1 year in adolescent and adult patients resulted in substantial reduction in hepatosplenomegaly and modest improvement in pulmonary function, sleep apnea, and joint mobility. Whether early institution of therapy in young children modulates mental retardation in the Hurler variant of MPS I is uncertain. Galsulfase (Naglazyme) has been approved for the treatment of MPS VI, in which somatic rather than neurologic problems predominate. Bone marrow transplantation has been attempted in many of the MPS disorders, with mixed success. The earlier transplantation occurs, the better the outcome in terms of somatic problems, but prevention of mental retardation has not occurred. Current recommendations based on consensus in Europe calls for hematopoietic stem cell transplantation for patients with Hurler syndrome before the age of 2.5 years. Enzyme replacement should be started in all patients when diagnosed.[3]

MARFAN SYNDROME

DEFINITION

Marfan syndrome is an autosomal dominant, pleiotropic disorder caused by defects in the principal component of the extracellular microfibril, the large glycoprotein fibrillin-1. The disease manifestations occur in multiple systems, especially the eye, skeleton, heart, aorta, lung, and integument. Notable features include dislocation of the ocular lens, tall stature with particularly long limbs and digits, deformity of the thoracic cage from pectus carinatum or excavatum with abnormal curvature of the spine, mitral and tricuspid valve prolapse, dilation of the sinuses of Valsalva and predisposition to aortic dissection, spontaneous pneumothorax, abnormal skin stretch marks, hernias, and dural ectasia. If untreated, patients often die before 30 or 40 years of age from aortic dissection or congestive heart failure.

EPIDEMIOLOGY

Marfan syndrome is a common Mendelian disorder, with an estimated incidence of about one per 5000 births. Marfan syndrome is found throughout the world, without ethnic or geographic predilection.

PATHOBIOLOGY

Pathogenesis

Mutations in *FBN1,* which maps to human chromosome 15q21.1 and encodes fibrillin-1, cause Marfan syndrome and related connective tissue disorders. More than 1000 distinct mutations have been found, and few occur in more than one family. Patients are heterozygous for mutations in *FBN1,* leading to autosomal dominant inheritance. Extracellular microfibrils are

polymers of many fibrillin-1 molecules and are ubiquitous in the ECM of most tissues. Latent transforming growth factor β (TGF-β) binding protein, which keeps the cytokine inactive, bears striking homology to regions of fibrillin. Abnormalities of either the quality or the quantity of microfibrils disrupt normal signaling by TGF-β, especially during embryonic development and postnatal growth. Studies in mice engineered to harbor human mutations in *FBN1* showed that excessive TGF-β signaling causes abnormal lung septation (the precursor to pneumothorax), mitral valve prolapse, muscular hypoplasia, and aortic dilatation. This fundamental shift in understanding of the pathogenesis of Marfan syndrome has suggested novel therapies, such as with small molecules that affect the activity of TGF-β or its downstream signaling.

The features of Marfan syndrome are highly variable, even among relatives who share the same mutation in *FBN1*. This variability persists after accounting for the effects of age. Men tend to be affected more severely, for unclear reasons.

Pathology

The features of Marfan syndrome are age dependent. Some severely affected infants have flagrant features and often die of mitral regurgitation and heart failure despite aggressive management. At the other end of the clinical spectrum, Marfan syndrome merges with several related disorders, and patients may not come to medical attention, let alone receive a definitive diagnosis, until adulthood.

None of the gross or microscopic pathologic changes is specific for Marfan syndrome. The medial degeneration of the aortic wall, characterized by disarray and fragmentation of the elastic fibers and increased proteoglycan (often inappropriately termed *cystic medial necrosis*) also can be seen in other disorders and in older people with hypertension. Aortic dissection (Chapter 78) usually begins just superior to the aortic valve (type A) and often progresses to the bifurcation. Death usually results from retrograde dissection and hemopericardium. About 10% of dissections begin in the descending thoracic aorta (type B).

CLINICAL MANIFESTATIONS

The lens tends to be displaced superiorly, and usually the zonules remain intact. The retina is at increased risk of detachment, especially in patients who are highly myopic. Tubular bones overgrow, accounting for the disproportionate tall stature (dolichostenomelia), long digits (arachnodactyly), and sternal deformity. Ligaments may be lax, causing scoliosis and joint hypermobility. Alternatively, congenital contractures are common, especially of the elbows. The palate typically is highly arched, and the dentition can be crowded and malocclused. Mitral valve prolapse occurs in about 80% of cases, and the valve leaflets become progressively thickened (myxomatous on histopathology) (Chapter 75). The mitral annulus may dilate and calcify. Aortic root dilation begins in the sinuses of Valsalva and progresses with age, albeit at highly variable rates (Chapter 78).[4] Most males with Marfan syndrome have an aortic root dimension above the upper limit of normal for their body surface area by adolescence. Some females show a slower progression and may have a root diameter near the upper limit of normal well into adulthood. The dilation usually does not involve the distal ascending aorta. Spontaneous pneumothorax, resulting from rupture of apical blebs, occurs in about 5% of patients. Stretch marks (striae atrophicae) occur over areas of flexural stress, such as the shoulders, breasts, and lower back. The neural canal in the lumbosacral region is enlarged in most people with Marfan syndrome; this may be visible on plain radiographs, especially if the neuroforamina are widened. Imaging by computed tomography or magnetic resonance imaging is diagnostic and should be used in patients with back pain and radicular symptoms. Dural ectasia progresses with age; large anterior meningoceles in the pelvis are a severe manifestation.[5] Simple cysts in the liver and kidneys are common, increase with age, and seldom cause clinical problems. Sleep apnea is of increased frequency in adults.[6]

DIAGNOSIS

Differential Diagnosis

The conditions that overlap clinically and genetically with Marfan syndrome include familial aortic aneurysm, familial ectopia lentis, mitral valve prolapse, mild aortic dilation, striae, skeleton (MASS) phenotype (which includes many families with mitral valve prolapse syndrome), and Loeys-Dietz syndrome. Most of these conditions are diagnosed clinically, so differentiating among them is arbitrary. A careful family history is essential to this process. Molecular genetic testing has a limited role. However, if the mutation in *FBN1* is known in a family, analysis of DNA can be used effectively for presymptomatic or prenatal diagnosis. Loeys-Dietz syndrome, which is associated with generalized arterial tortuosity and susceptibility to dissection, is caused by mutation in either of two receptors for TGF-β, *TGFBR1* and *TGFBR2*, and molecular analysis is clinically available.

A question of Marfan syndrome arises most commonly in tall, lanky adolescents who have several minor skeletal features, nearsightedness, and athletic desires. A detailed ophthalmologic examination with full pupillary dilation and a transthoracic echocardiogram are essential components in the evaluation. If these test results are negative and no one in the family has a history of Marfan syndrome or aortic dissection, the patient probably can be reassured.[7]

TREATMENT

Rx

Life expectancy for those with Marfan syndrome has improved markedly, to the point that many patients can expect survival to advanced years. All patients should be seen at least annually by a physician who manages the overall care. Most patients require annual ophthalmologic and cardiologic consultation and orthopedic consultation as required by specific problems. Lens subluxation often requires surgical correction.[8] A number of studies, but only one randomized clinical trial, support the prophylactic use of β-adrenergic blockade from an early age to slow the rate of aortic root dilation and protect against aortic dissection. Based on studies of the Marfan mouse, therapies that interfere with excess signaling through pathways mediated by TGF-β are being studied in human clinical trials. One large European trial suggested a benefit of the angiotensin receptor blocker losartan on aortic root dilatation rate,[9] but a large international trial found no benefit of losartan compared with atenolol.[10] Prophylactic surgical repair of the aortic root has had the greatest beneficial impact. The composite graft, involving a prosthetic valve in a Dacron tube and implantation of the coronary ostia into the graft, was the first approach to produce markedly improved survival in these patients. More recently, replacement of the aneurysm and preservation of the native aortic valve have shown promise and should be considered first.[9,10] For adults, aortic root surgery should be strongly considered when the maximal aortic diameter reaches 45 mm, and a family history of aortic dissection should prompt earlier repair (Chapter 78).

EHLERS-DANLOS SYNDROMES

DEFINITION

The Ehlers-Danlos syndromes (EDSs) are clinically variable and genetically heterogeneous. Diagnoses still are based largely on the bedside examination. The unifying themes among these disorders are fragility of tissues, joint hypermobility, and skin hyperextensibility.[11]

EPIDEMIOLOGY

No accurate data exist, but an incidence of about one in 5000 births is a reasonable estimate of how many individuals qualify for one of the EDS diagnoses. Each type represents something of a clinical spectrum, with the mild end merging with what might be considered normal variation. Just as the diagnostic criteria are arbitrary, so would be any determination of prevalence based on phenotypic criteria. The extent to which normal variation in joint hypermobility, skin elasticity, and tissue fragility represents genetic variation at loci that encode collagen or other ECM genes requires considerable research.

PATHOBIOLOGY

Pathogenesis

Defects in collagen and other proteins in the ECM of various tissues underlie all forms of EDS that have been elucidated so far. The specific mutations occur in a variety of genes, with the effect of altering the structure, synthesis, posttranslational modifications, or stability of the collagens involved. The known molecular defects are listed in Table 260-2.

Pathology

Few findings in the routine pathologic evaluation distinguish among the various types of EDS or even distinguish individual types from normal. Thickness of the dermis is decreased in some forms, especially the vascular type, and the walls of arteries are reduced in thickness in this type. By electron microscopy, the classic, hypermobile, and kyphoscoliotic types have abnormal collagen fibers, especially when viewed in cross section (variable and often increased fiber diameter with an irregular outline). In the vascular type,

TABLE 260-2 EHLERS-DANLOS SYNDROMES

TYPE	FORMER NAME	CLINICAL FEATURES*	INHERITANCE	OMIM†	MOLECULAR DEFECT
Classic	EDS I and II	Joint hypermobility; skin hyperextensibility; atrophic scars; smooth, velvety skin; subcutaneous spheroids	AD	130000 130010	Structure of type V collagen caused by mutations in *COL5A1* or *COL5A2*
Hypermobility	EDS III	Joint hypermobility; some skin hyperextensibility, with or without a smooth, velvety texture	AD AR	130020 225320	? Tenascin-X *(TNX)*
Vascular	EDS IV	Thin skin; easy bruising; pinched nose; acrogeria; rupture of large- and medium-caliber arteries, uterus, and large bowel	AD	130050 (225350) (225360)	Deficient type III collagen *(COL3A1)*
Kyphoscoliotic	EDS VI	Joint hypermobility; congenital, progressive rupture; scoliosis; scleral fragility with globe rupture; tissue fragility, aortic dilation, MVP	AR	225400	Deficiency of lysyl hydroxylase
Arthrochalasis	EDS VII A	Joint hypermobility, severe, with subluxations, congenital hip dislocation; and skin hyperextensibility; tissue fragility	AD	130060	No cleavage of amino terminus of type I procollagen caused by mutations in *COL1A1* or *COL1A2*
Dermatosparaxis	EDS VII C	Severe skin fragility; decreased skin elasticity, easy bruising; hernias; premature rupture of fetal membranes	AR	225410	No cleavage of amino terminus of type I procollagen caused by deficiency of peptidase
Unclassified types	EDS V	Classic features	XL	305200	?
	EDS VIII	Classic features and periodontal disease	AD	130080	?
	EDS X	Mild classic features, MVP	?	225310	?
	EDS XI	Joint instability	AD	147900	?
	EDS IX	Classic features; occipital horns	XL	309400	Allelic to Menkes syndrome
	EDS, progeroid form	Classic features and premature aging	AR	130700	Deficiency of galactosyltransferase I

*Listed in order of diagnostic importance.
†Entries in Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute of Genetic Medicine. Baltimore: Johns Hopkins University. http://omim.org.
AD = autosomal dominant; AR = autosomal recessive; EDS = Ehlers-Danlos syndrome; MVP = mitral valve prolapse; XL = X-linked.

some patients have dilated endoplasmic reticulum consistent with aberrant secretion of type III collagen molecules.

CLINICAL MANIFESTATIONS

The major and minor features of each EDS are detailed in Table 260-2. Infants with classic EDS often are born prematurely by 4 to 8 weeks because of rupture of fetal membranes. Diagnosis of the vascular and kyphoscoliotic types is important because of their cardiovascular features. The vascular type, previously termed *EDS IV,* is characterized by spontaneous rupture of large arteries and hollow organs, especially the colon and uterus, and pneumothorax. Because these events carry considerable morbidity, life expectancy is reduced, on average, by more than half. During pregnancy, women with this form of EDS are especially vulnerable to rupture of major arteries and the uterus. In the kyphoscoliotic type, aortic root dilation and aortic regurgitation can develop. Patients with most forms of EDS are prone to develop mitral valve prolapse, and progression to mitral regurgitation (Chapter 75) occurs more often than in the common form of mitral valve prolapse.

DIAGNOSIS

Differential Diagnosis

By carefully adherence to the clinical features shown in Table 260-2 and judicious use of laboratory tests, the various defined types of EDS can be differentiated. Many specific non-EDS syndromes need to be excluded. The kyphoscoliotic type of EDS in infants shares some features with severe Marfan syndrome. Patients with Larsen syndrome may resemble patients with the arthrochalasis type of EDS. The skin redundancy and loss of elasticity of the dermatosparaxis type of EDS is reminiscent of autosomal dominant cutis laxa, which is not associated with easy bruising or tissue fragility.

The most difficult decision is whether any diagnosis of EDS is warranted. Patients who have only joint hypermobility without skin changes should not be labeled with EDS; a diagnosis of familial joint hypermobility might be more appropriate. Familial joint instability involves a predisposition to dislocations of major joints that is rare in most types of EDS except for arthrochalasis.

TREATMENT

Rx

Management of most skin and joint problems should be conservative and preventive. Sutures need to be placed with careful attention to approximating the margins and avoiding tension; removable sutures should be left in place for twice the usual time. Most instances of joint hypermobility and pain in EDS do not require surgical treatment. Benefit often is derived from physical therapy designed to strengthen the muscles that provide support for the loose ligaments. All patients should receive genetic counseling about the mode of inheritance and their risk of having children affected with EDS. The possibility of prenatal diagnosis exists for all of the EDS types with defined molecular or biochemical defects.

The vascular type of EDS requires particular surgical care; the ruptured arteries are difficult to repair because of the pronounced vascular fragility. Experienced vascular surgeons are having some success with prophylactic repair of vessels deemed to be at risk of dissection or rupture.[12] One clinical trial suggested improved outcomes with prophylactic β-adrenergic blockade. Rupture of the bowel is a surgical emergency. Because the risk of uterine and vascular rupture is especially high during pregnancy in women with the vascular form, affected women should be advised that there is a substantial risk of death related to pregnancy and delivery. Patients should be advised to avoid contact sports and to treat blood pressure elevations aggressively. Arteriography and arterial lines should be avoided if possible. Biochemical and genetic screening holds the potential for reassuring relatives at risk that they do not have a defect in type III collagen.

The kyphoscoliotic type of EDS may improve with large doses of vitamin C (1-4 g/day) because ascorbate is a cofactor for the enzyme that is deficient. No other metabolic or genetic therapy is effective in other forms of EDS.

OSTEOGENESIS IMPERFECTA SYNDROMES

DEFINITION

The heterogeneous group of disorders called *osteogenesis imperfecta* (OI) includes, at one end of the severity spectrum, a type that is lethal prenatally or in the neonatal period and, at the other, such mild features that distinguishing affected individuals from the general population is difficult. The unifying feature is hereditary osteopenia (insufficient bone), with primary defects in the protein matrix in bone and other tissues. The clinical syndromes all involve osteoporosis with liability to fracture (Chapter 243).

EPIDEMIOLOGY

No careful epidemiologic study has been performed, and the milder forms of type I OI merge with the phenotypes of familial osteoporosis, fracture

TABLE 260-3 OSTEOGENESIS IMPERFECTA

TYPE	CLINICAL FEATURES	INHERITANCE	OMIM*	BASIC DEFECTS
I	Fractures variable in number; little deformity; stature normal or nearly so; blue sclerae; hearing loss common but not always present; DI uncommon	AD	166200	Typically, one nonfunctional *COL1A1* allele
II	Lethal in utero or shortly after birth; many fractures at birth typically involving ribs (may appear "beaded") and other long bones; little calvaria; pulmonary hypertension	AD	166210	*COL1A1* or *COL1A2:* substitution of glycyl residues; occasionally deletions of a portion of the triple-helical domain
		AR	259400	Deletion in *COL1A2* plus a nonfunctional allele
III	Fractures common, but long bones progressively deform starting in utero; stature markedly reduced; sclerae often blue but become lighter with age; DI and hearing loss common	AD	259420	One single amino acid substitution
		AR (rare)	259440	Two mutations in *COL1A1* and/or *COL1A2* (rarely)
IV	Fractures common; stature usually reduced; bone deformity common but rarely severe; scleral hue normal to grayish; hearing loss variable; DI common	AD	166220	Point mutations in *COL1A1* or *COL1A2*
			166240	Exon skipping mutations in *COL1A2*
V	Similar to type IV without DI or blue sclerae; fractures develop hyperplastic callus; calcification of the interosseous membrane between the radius and ulna	AD	610967	?
VI	Similar to type IV without DI, blue sclerae or Wormian bones; excess osteoid present in bone	?	610968	?
VII	Similar to types II or III with fractures at birth, blue sclerae, no DI; presence of rhizomelic limb shortening and coxa vara	AR	610682	Mutations in *CRTAP*
VIII	Similar to types II or III with fractures at birth	AR	610915	Mutations in *LEPRE1*
IX	Similar to types II or III with fractures at birth	AR	259440	Mutations in *PPIB*

*Entries in Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute of Genetic Medicine. Baltimore: Johns Hopkins University. http://omim.org.
AD = autosomal dominant; AR = autosomal recessive; DI = dentinogenesis imperfecta.

susceptibility, and joint hypermobility found in the general population. A crude estimate of the overall prevalence of OI is one to two per 20,000 births. The neonatal lethal form (type II), which is almost always caused by a new mutation in a parental gamete, has an incidence of about one in 50,000 births.

PATHOBIOLOGY

Pathogenesis

Most patients in whom mutations have been found usually have defects in the two genes that encode the procollagen chains of type I collagen, *COL1A1* and *COL1A2*. Type I collagen is composed of two α1(I) and one α2(I) procollagen chains; the mature fiber requires considerable posttranslational modification, which occurs appropriately only if the three procollagen chains have intertwined to form a triple helix that is perfect and completed at the right speed. A mutation that affects formation of the triple helix, such as substitution of one of the mandatory glycine residues that occurs at every third position, also has adverse effects on the modifications that render the molecule capable of forming effective mature fibers. As a result, a single nucleotide change resulting in a missense mutation can have profound effects on the ECM and produce a severe condition.[13] Alternatively, and at first glance paradoxically, a mutation that eliminates an entire allele, or at least production of any product capable of intertwining with normal procollagen chains, has a much milder effect on the ECM and on the severity of OI. Examples of the most common classes of mutations are shown in Table 260-3. Hundreds of mutations have been described. Patients with mutations in *COL1A1* or *COL1A2* are heterozygous, and thus the most common forms of OI are inherited as autosomal dominant traits. Several autosomal recessive forms of OI occur because of mutations in genes that encode enzymes that process type I collagen into mature fibrils.[14]

Pathology

Other than the gross pathology associated with the clinical manifestations, the most characteristic pathology is a primary reduction in bone matrix with secondary undermineralization.

CLINICAL MANIFESTATIONS

The major phenotypic features of OI are shown in Table 260-3. Among the most common forms, the most severe type is type II, followed in decreasing order by types III, IV, and I. In type II, infants either are stillborn or die soon after birth of pulmonary failure secondary to the small thorax, which usually is compromised further by myriad rib fractures. A few infants have survived for at least a few years but require enormous attention to their medical needs.

Type III OI may be confused with type II at birth, but survival alone helps make the distinction. Bony deformity is pronounced and not necessarily caused by fractures. Mobility is impaired, and most patients require a wheelchair at an early age. Stature may be severely compromised. Because of progressive vertebral column deformity and rib fractures, restrictive lung disease is a common problem as patients age; many die of pulmonary complications. Basilar impression causing compression of the brain stem and the craniocervical junction can produce central sleep apnea, headache, and upper motor neuron signs.

Patients with type IV OI generally have reduced stature, some bony deformity, and abnormal teeth that are opalescent and wear easily (dentinogenesis imperfecta). As in type I OI, the tendency to fracture is highest in childhood and lessens with adolescence. A distinguishing characteristic of type IV OI is a normal scleral hue.

Type I OI is probably the most common form and is associated with a bluish or blue-gray scleral hue. People with type I OI who also have dentinogenesis imperfecta tend to have more severe skeletal problems. The risk of fracture diminishes during adulthood but reemerges as a major concern for women after menopause. Hearing impairment in all forms of OI is common and age related, being rare before adolescence. The deficits are of either a mixed or a predominantly conductive form.

The recessive forms of OI (types VI-IX) range in severity from type IV to type II and may have distinctive radiologic or histopathologic findings.

DIAGNOSIS

Differential Diagnosis

The range of diagnostic possibilities in a person with multiple fractures largely depends on age. In infancy, the genetic conditions hypophosphatasia, severe osteochondrodysplasias (e.g., achondrogenesis and forms of spondyloepiphyseal dysplasia), and Menkes syndrome need to be excluded when a diagnosis of type II or type III OI is considered. The radiographic features eventually become entirely diagnostic, but often the neonatologist has to arrive at a definitive answer in short order. Analysis of serum alkaline phosphatase and copper can be helpful. In childhood, the most common situation leading to consideration of a mild form of OI is child abuse. In this situation, the pattern of fracture is usually distinct, and bone mineralization should be normal if the child is the object of nonaccidental or repeated accidental trauma. Abnormal scleral hue, dentinogenesis imperfecta, and wormian bones (microfractures along the cranial sutures) all support the diagnosis of

OI. The legal and child-protective systems often request exclusion of OI by analysis of collagen production from cultured skin fibroblasts or analysis of DNA for a mutation.

In older children, the disorder idiopathic juvenile osteoporosis should be considered in any patient seen initially with repeated fractures. Many osteochondrodysplasias are associated with short stature, skeletal deformity, and a tendency to fracture. Pyknodysostosis and osteopetrosis are associated with sclerotic bones rather than osteoporotic ones. In adulthood, early-onset osteoporosis may be confused with OI (Chapter 243). Mutations in type I collagen also cause familial osteoporosis, and the skeletal phenotypes merge; patients with true OI may have scleral, hearing, or dental abnormalities and a positive family history.

Analysis of the specific enzymes defective in the recessive forms of OI is useful for establishing the diagnosis and enabling reproductive counseling and prenatal diagnosis if desired.

TREATMENT

Rx

Management of the skeletal complications largely depends on orthopedic, physical, and occupational therapy approaches. Risedronate (2.5 or 5 mg daily) increases bone mineral density and reduces both first and recurrent fractures in children with OI.[A1] The long-term goals are for the patient to maintain function and independence as an individual. These goals can be advanced in some patients by judicious use of intramedullary rods in the long bones of the legs; if mobility and especially ambulation can be maintained, the demineralization associated with inactivity can be avoided.

Unaffected parents of a child with OI and all affected individuals should have genetic counseling. For the parents of a child with type II OI, the possibility of germinal mosaicism (which has been well documented in this condition) should not be overlooked. If one parent has a "new" mutation in one of the type I procollagen genes and multiple gonadal cells carry this mutation, the risk of recurrence in future children is not negligible. If the mutation in the affected child can be defined, the risk of recurrence can be quantified (through molecular analysis of sperm) if the mutation arose in the father.

PSEUDOXANTHOMA ELASTICUM

DEFINITION

Pseudoxanthoma elasticum (PXE) is a heritable disorder of connective tissue with pleiotropic manifestations wherever elastic fibers are found but primarily in the skin, eye, and vasculature.[15] Life expectancy is reduced, on average, because of a predisposition to myocardial infarction and gastrointestinal hemorrhage.

EPIDEMIOLOGY

The exact frequency of PXE is unknown, but it is probably underdiagnosed. Rough approximations suggest a prevalence of one in 25,000 to 100,000 births. Males and females are equally affected, although women are more likely to seek medical attention out of concern for the skin changes.

PATHOBIOLOGY

Pathogenesis

In most families, PXE occurs as an autosomal recessive trait, which means, given relatively small sibships, that many patients will have no affected relatives. Apparent autosomal dominant inheritance may reflect expression in occasional heterozygotes. The gene for PXE maps to human chromosome 16 and encodes one of the adenosine triphosphate (ATP)-binding cassette transporters *(ABCC6)*. Because of the prominent histopathologic feature of calcification of elastic tissue, this gene may be important in calcium homeostasis. It is unclear, however, whether calcification is a primary or a secondary phenomenon in PXE.

Pathology

The hallmark of PXE, and an important diagnostic clue, is the histopathologic finding of hyperproliferated elastic fibers in the mid-dermis; these fibers become fragmented, clumped, and calcified. An arteriolar sclerosis develops in the media of muscular arteries and arterioles; the lumen may become progressively and concentrically narrowed. Alternatively, microaneurysms can form. Thickening of the endocardium, especially in the atria, develops in some patients. In the eye, Bruch membrane becomes calcified and fragmented.

CLINICAL MANIFESTATIONS

Because of the pleiotropic nature of PXE, the diagnosis initially may be suspected by any of a variety of clinicians, especially dermatologists, ophthalmologists, cardiologists, and gastroenterologists. The condition gains its name from the dermatologic feature of yellowish papules that appear at areas of flexural stress, especially the neck, groin, and popliteal and cubital fossae; in periumbilical regions; and on the buccal mucosa. The appearance of affected skin has been likened to that of a "plucked chicken." Over time, affected areas coalesce and become thickened.

Changes in the eye begin as a generalized, subtle, mottled pattern in the retina (peau d'orange) and progress to the characteristic angioid streaks. The latter changes are not specific for PXE and can be seen in diabetes mellitus, sickle cell disease, and a variety of other conditions. Streaks represent breaks in Bruch's membrane, an elastic lamina that lies between the retinal vasculature and the choroid. Spontaneous hemorrhages, especially those involving the macula, lead to progressive visual loss.[16]

Involvement of arteries of various calibers produces problems because of occlusion and hemorrhage.[17] The lifetime risk of serious gastrointestinal hemorrhage from any site, but especially the stomach, is about 10%. Hypertension is relatively common, in part because of involvement of the renal vasculature. Progressive occlusion of peripheral arteries leads to absence of pulses; acral ischemia is rare because of the development of collaterals. The risk for stroke, myocardial infarction, abdominal angina, and intermittent claudication is increased independent of other risk factors. Impaired left ventricular function is common in adults.

DIAGNOSIS

Differential Diagnosis

Whole exome sequencing is an efficient and sensitive way to make the diagnosis.[18] An acquired form of PXE has been reported and is also of unclear etiology. This form is difficult to differentiate from a sporadic case in a family because of heterozygosity in the parents, but it tends to affect only the skin. As suggested by the name, the cutaneous features of PXE need to be differentiated from those of true xanthoma, which results from a disorder of lipid metabolism (Chapter 206). The dermatologic manifestations need to be differentiated from those of Miescher elastoma, elastic tissue nevi (Buschke-Ollendorff syndrome), and solar elastosis.

TREATMENT

Rx

No cure for or means of preventing PXE is known. In many instances, careful attention to the ocular features by a retinal specialist experienced in PXE can delay but not prevent loss of vision. The risk of gastrointestinal hemorrhage suggests that patients should avoid gastric irritants such as aspirin, nonsteroidal anti-inflammatory drugs, and excessive alcohol. Stool should be checked regularly for occult blood, and angiography may be necessary to detect the source of bleeding. All standard risk factors for atherosclerosis should be managed aggressively. Complaints of chest pain should prompt a rigorous investigation for coronary artery disease. Angioplasty has not been reported to be effective, and the coronary lesions tend to be diffuse. Coronary artery bypass graft surgery has been performed, but long-term results have not been reported. It may be theoretically advantageous to use vein grafts rather than the internal mammary artery for bypass. The excessive wrinkling and pseudoxanthoma in exposed areas can be ameliorated by plastic surgery.

FUTURE DIRECTIONS

Each of these disorders poses special considerations in clinical diagnosis, utility of molecular testing, genetic counseling, and management. For the storage disorders, the clinical utility of enzyme replacement therapy is actively being pursued by several pharmaceutical companies. For several of the other conditions, somatic stem cell therapy offers some promise but is years away from routine clinical use. In Marfan syndrome, clinical trials of drugs that modulate activity of TGF-β are underway. Additionally, close medical management for individuals detected as being at heightened risk for cardiovascular, skeletal, and ocular complications will remain a mainstay.

Grade A References

A1. Groenink M, den Hartog AW, Franken R, et al. Losartan reduces aortic dilatation rate in adults with Marfan syndrome: a randomized controlled trial. *Eur Heart J*. 2013;34:3491-3500.

A2. Lacro RV, Dietz HC, Sleeper LA, et al. Atenolol versus losartan in children and young adults with Marfan's syndrome. *N Engl J Med*. 2014;371:2061-2071.
A3. Ong KT, Perdu J, De Backer J, et al. Effect of celiprolol on prevention of cardiovascular events in vascular Ehlers-Danlos syndrome: a prospective, randomized, open, blinded-endpoints trial. *Lancet*. 2010;376:1476-1484.
A4. Bishop N, Adami S, Ahmed SF, et al. Risedronate in children with osteogenesis imperfecta: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2013;382:1424-1432.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Campos D, Monaga M. Mucopolysaccharidosis type I: current knowledge on it pathophysiological mechanisms. *Metab Brain Dis*. 2012;27:121-129.
2. Stevenson DA, Steiner RD. Skeletal abnormalities in lysosomal storage diseases. *Pediatr Endocrinol Rev*. 2013;10(suppl 2):406-416.
3. De Ru MH, Boelens JJ, Das AM, et al. Enzyme replacement therapy and/or hematopoietic stem cell transplantation at diagnosis in patients with mucopolysaccharidosis type I: results of a European consensus procedure. *Orphanet J Rare Dis*. 2011;6:55.
4. Jondeau G, Detaint D, Tubach F, et al. Aortic event rate in the Marfan population. *Circulation*. 2012;125:226-232.
5. Sheikhzadeh S, Sondermann C, Rybczynski M, et al. Comprehensive analysis of dural ectasia in 150 patients with a causative *FBN1* mutation. *Clin Genet*. 2014;86:238-245.
6. Rybczynski M, Koschyk D, Karmeier A, et al. Frequency of sleep apnea in adults with the Marfan syndrome. *Am J Cardiol*. 2010;105:1836-1841.
7. Pyeritz RE. Evaluation of the tall adolescent with some features of Marfan syndrome. *Genet Med*. 2012;14:171-177.
8. Miraldi Utz V, Coussa RG, Traboulsi EI. Surgical management of lens subluxation in Marfan syndrome. *J AAPOS*. 2014;18:140-146.
9. Song HK, Preiss LR, Maslen CL, et al. Valve-sparing aortic root replacement in patients with Marfan syndrome enrolled in the National Registry of Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions. *J Heart Valve Dis*. 2014;23:292-298.
10. Svensson LG, Blackstone EH, Alsalihi M, et al. Midterm results of David reimplantation in patients with connective tissue disorders. *Ann Thorac Surg*. 2013;95:555-562.
11. De Paepe A, Malfait F. The Ehlers-Danlos syndrome, a disorder with many faces. *Clin Genet*. 2012;82:1-11.
12. Lum YW, Brooke BS, Black JH III. Contemporary management of vascular Ehlers-Danlos syndrome. *Curr Opin Cardiol*. 2011;26:494-501.
13. Li Q, Jiang Q, Uitto J. Ectopic mineralization disorders of the extracellular matrix of connective tissue: molecular genetics and pathomechanisms of aberrant calcification. *Matrix Biol*. 2014;33:23-28.
14. Byers PH, Pyott SM. Recessively inherited forms of osteogenesis imperfect. *Annu Rev Genet*. 2012;46:475-497.
15. Uitto J, Jiang Q, Varadi A, et al. Pseudoxanthoma elasticum: diagnostic features, classification, and treatment options. *Expert Opin Orphan Drugs*. 2014;2:567-577.
16. Ebran JM, Milea D, Trelohan A, et al. New insights into the visual prognosis of pseudoxanthoma elasticum. *Br J Ophthalmol*. 2014;98:142-143.
17. Campens L, Vanakker OM, Trachet B, et al. Characterization of cardiovascular involvement in pseudoxanthoma elasticum families. *Arterioscler Thromb Vasc Biol*. 2013;33:2646-2652.
18. Hosen MJ, Van Nieuwerburgh F, Steyaert W, et al. Efficiency of exome sequencing for the molecular diagnosis of pseudoxanthoma elasticum. *J Invest Dermatol*. 2014;[Epub ahead of print].

REVIEW QUESTIONS

1. Osteogenesis imperfect syndromes demonstrate

A. intergenic heterogeneity.
B. intragenic heterogeneity.
C. variable expression.
D. pleiotropy.
E. all of the above.

Answer: E All four of the choices are correct. Mutations in multiple genes can cause OI (intergenic heterogeneity). With a single locus, many different mutant alleles have been discovered (intragenic heterogeneity). Within a given type of OI, relatives with the same mutation can demonstrate different features of varying severity (variable expression). Multiple organ systems are affected in OI (pleiotropy).

2. Which of the following is not a common feature of Marfan syndrome?

A. Pulmonary arteriovenous malformation
B. Aortic root dilatation
C. Mitral valve prolapse
D. Pulmonic artery dilatation
E. Aortic dissection

Answer: A All four of the latter choices are common features. Arteriovenous malformations are common in hereditary hemorrhagic telangiectasia but not in Marfan syndrome.

3. Hypertension is a relatively common feature in which syndrome?

A. Osteogenesis imperfect
B. Hurler syndrome
C. Marfan syndrome
D. Vascular Ehlers-Danlos syndrome
E. Pseudoxanthoma elasticum

Answer: E Because of partial occlusion of the renal arteries, elevated blood pressure can occur in PXE. None of the other syndromes has a predisposition to hypertension.

4. Which of the following is not a currently accepted approach to management of Marfan syndrome?

A. Prophylactic surgery of the aortic root
B. Exercise restriction
C. Annual echocardiography
D. Gene therapy
E. Chronic β-adrenergic blockade

Answer: D At the present time, no approach to correcting the specific mutation in *FBN1* is feasible. All of the other approaches are thought to be beneficial.

5. If treatment is not offered, which of the following conditions is associated with the best prognosis?

A. Ehlers-Danlos syndrome, hypermobility type
B. Osteogenesis imperfect, type II
C. Marfan syndrome
D. Hurler syndrome
E. Ehlers-Danlos syndrome, vascular type

Answer: A The hypermobility form of Ehlers-Danlos syndrome has little to no added mortality. Osteogenesis imperfect type II is usually lethal in infancy. Marfan syndrome has reduced life expectancy because of aortic dissection. Patients with Hurler syndrome rarely survive to their third decade. Patients with the vascular form of Ehlers-Danlos syndrome are at high risk of death from arterial or bowel rupture.

261

THE SYSTEMIC AUTOINFLAMMATORY DISEASES

RICHARD M. SIEGEL AND DANIEL L. KASTNER

DEFINITION

The systemic autoinflammatory diseases (Table 261-1) are a group of illnesses characterized by seemingly unprovoked inflammation, without evidence of high-titer autoantibodies or antigen-specific T cells, thus distinguishing them from the more classic autoimmune diseases. The first conditions recognized as autoinflammatory were the hereditary recurrent fevers, a group of mendelian disorders characterized by episodic or fluctuating degrees of fever and localized inflammation. The scope of autoinflammatory disease has been broadened to include other heritable illnesses, including disorders in which purulent or granulomatous inflammation predominates, as well as inherited disorders of the complement system (Chapter 50).[1-6] In addition, in numerous autoinflammatory conditions, some of which manifest in childhood and others that occur later in life, there is a complex interaction of genetic susceptibilities and environmental factors. These illnesses include systemic-onset juvenile idiopathic arthritis (Still's disease), Behçet's disease, and even the crystalline arthritides. Recent advances in the genetics and pathophysiology of the inherited autoinflammatory diseases suggest that these conditions are inborn errors of innate immunity, the phylogenetically more primitive part of the immune system that uses germline membrane and intracellular receptors expressed in granulocytes and macrophages to mount the body's first line of defense against pathogens (Chapters 45 and 48).

HEREDITARY RECURRENT FEVER SYNDROMES

Familial Mediterranean Fever

DEFINITION

Familial Mediterranean fever (FMF) is a recessively inherited illness that typically manifests with 12- to 72-hour episodes of fever and localized serosal, synovial, or cutaneous inflammation. Between attacks, patients usually feel completely well, although biochemical evidence of inflammation may remain, and some patients eventually develop systemic amyloidosis. Before the identification of the causative gene, FMF was defined purely clinically; clinical features remain an important part of the diagnosis, because some patients with typical disease have only one, or sometimes no, demonstrable mutation in *MEFV,* the only known causative gene.

EPIDEMIOLOGY

FMF is most common in individuals of Jewish, Arab, Armenian, Turkish, and Italian ancestry. The frequency of asymptomatic carriers of a single *MEFV* mutation in these populations is as high as 1 in 5, a finding that suggests a selective advantage for heterozygotes. With genetic testing, FMF is now frequently recognized in both Ashkenazi (eastern European) and non-Ashkenazi Jewish populations, as well as in Mediterranean populations previously thought not to be at risk. Mutation-positive individuals with typical symptoms have been documented worldwide. FMF usually manifests in childhood, sometimes even in infancy, although approximately 10% of patients experience their first attack as adults; infrequently, FMF first occurs in persons older than 40 years.

PATHOBIOLOGY

MEFV, the gene for FMF, was identified by positional cloning in 1997. It encodes a 781–amino acid protein denoted pyrin (or marenostrin) that is expressed in granulocytes, monocytes, and dendritic cells, as well as in peritoneal, synovial, and dermal fibroblasts. The N-terminal 92 amino acids of pyrin are the prototype for a motif, the PYRIN domain, that is involved in protein-protein interactions; this domain defines a family of more than 20 human proteins, including pyrin itself, involved in the regulation of cytokine production (particularly the interleukin-1 [IL-1] family), nuclear factor kappa B (NF-κB) activation, and apoptosis. More than 50 FMF-associated mutations in pyrin have been identified, many of which reside in the C-terminal domain encoded by exon 10 of *MEFV.* An even larger number of variants of unknown significance have been described in individual patients with a spectrum of inflammatory phenotypes.

CLINICAL MANIFESTATIONS

Episodes of FMF are more properly termed recurrent than periodic, and some patients associate attacks with psychological stress or physical exertion. Women of childbearing age sometimes experience their attacks with menses, with remissions during pregnancy. Some patients are unaware of fever during the attacks, but it is almost always observed when sought.

Serosal involvement in FMF is usually peritoneal or pleural. Abdominal attacks are the most frequent, and they may vary from mild discomfort to frank peritonitis, with boardlike rigidity, direct and rebound tenderness, and air-fluid levels on upright films of the abdomen. Regardless of the severity of the abdominal attack, constipation is much more common than diarrhea. When a laparotomy or laparoscopy is performed during an attack, a small amount of sterile exudate rich in polymorphonuclear leukocytes is found. Except for serosal inflammation, the appendix is normal. Repeated abdominal attacks may cause peritoneal adhesions, but ascites is rare. Pleurisy, usually unilateral, may accompany abdominal pain, or it may occur independently. Physical findings, if present, may include diminished breath sounds and a pleural friction rub, whereas x-ray films may show a small effusion or atelectasis. With multiple attacks, pleural thickening may develop. Symptomatic nonuremic pericardial involvement in FMF has been reported but is unusual.

In adults, the arthritis of FMF typically manifests as monoarticular involvement of the knee, hip, or ankle, and attacks of arthritis may persist for up to 1 week at a time. In children, oligoarticular or polyarticular joint involvement may occur. Large joint effusions are sometimes present, and the synovial fluid may have as many as 100,000 leukocytes/mm^3. In approximately 5% of patients who are not treated with prophylactic colchicine, chronic arthritis (usually of the hip or knee) may develop, often necessitating joint replacement surgery. Regardless of colchicine treatment or a particular human leukocyte antigen (HLA-B27) status, some patients with FMF develop sacroiliitis. Arthralgia without frank arthritis is common in FMF.

Cutaneous manifestations of FMF tend to be less common than serosal or synovial involvement. The characteristic skin lesion of FMF is erysipeloid erythema, a painful, demarcated erythematous area most often seen on the lower leg, ankle, or dorsum of the foot. This rash may occur independently, or it may accompany an episode of arthritis. Histologically, a mixed perivascular cellular infiltrate is seen. Other acute manifestations of FMF include unilateral scrotal inflammation (the tunica vaginalis is an embryologic remnant of the peritoneal membrane) and myalgia, either with fever or, especially in children, without fever and induced by vigorous exercise. Various forms of vasculitis also have been associated with FMF; Henoch-Schönlein purpura may occur in children with FMF; less frequently, polyarteritis nodosa is seen.

COMPLICATIONS

Before the widespread use of colchicine prophylaxis, systemic AA amyloidosis (Chapter 188) was a frequent complication of FMF, caused by the ectopic deposition of a misfolded fragment of serum amyloid A (SAA), an acute phase reactant, in the gastrointestinal tract, kidneys, spleen, lung, testes, and adrenals. Malabsorption and nephrotic proteinuria leading to renal failure are the most common manifestations of AA amyloidosis. Cardiomyopathy is less common, and neuropathy and arthropathy are rare. Several risk factors for amyloidosis development in FMF have been identified, including late diagnosis of FMF, colchicine noncompliance, male gender, and specific

TABLE 261-1 SYSTEMIC AUTOINFLAMMATORY DISEASES: A PARTIAL LISTING

INHERITED AUTOINFLAMMATORY DISEASES	INHERITANCE/ETIOLOGY	GENES OR RISK FACTORS	OMIM*
HEREDITARY RECURRENT FEVER SYNDROMES			
Familial Mediterranean fever (FMF)	Autosomal recessive	*MEFV*†	249100
Tumor necrosis factor receptor–associated periodic syndrome (TRAPS)	Autosomal dominant	*TNFRSF1A*†	142680
Hyperimmunoglobulinemia D with periodic fever syndrome (HIDS)	Autosomal recessive	*MVK*†	260920
Familial cold autoinflammatory syndrome (FCAS)	Autosomal dominant	*NLRP3* (formerly *CIAS1*)†	120100
Muckle-Wells syndrome (MWS)	Autosomal dominant	*NLRP3* (formerly *CIAS1*)†	191900
Neonatal-onset multisystem inflammatory disease (NOMID)/chronic infantile neurologic cutaneous and articular (CINCA) syndrome	Sporadic, autosomal dominant	*NLRP3* (formerly *CIAS1*)†	607115
GRANULOMATOUS DISORDERS			
Granulomatous inflammatory arthritis, dermatitis, and uveitis (Blau's syndrome)	Autosomal dominant	*NOD2/CARD15*†	186580
Early-onset sarcoidosis	Sporadic, autosomal dominant	*NOD2/CARD15*†	609464
Crohn's disease	Complex inheritance	*NOD2/CARD15*†	266600
PYOGENIC DISORDERS			
Syndrome of pyogenic arthritis with pyoderma gangrenosum and acne (PAPA)	Autosomal dominant	*PSTPIP1*†	604416
AUTOINFLAMMATORY DISORDERS OF SKIN AND BONE			
Deficiency of interleukin-1 receptor antagonist (DIRA)	Autosomal recessive	*IL1RN*	612852
Chronic recurrent multifocal osteomyelitis (CRMO)	Sporadic, autosomal recessive	*LPIN2*,† when associated with congenital dyserythropoietic anemia (Majeed syndrome)	259680
Synovitis acne pustulosis hyperostosis osteitis syndrome (SAPHO)	Idiopathic	—	—
COMPLEMENT DISORDERS			
Hereditary angioedema	Autosomal dominant	*C1NH*	106100
Hemolytic-uremic syndrome	Autosomal dominant, sporadic	HF1 (complement factor H)	235400
Age-related macular degeneration	Complex inheritance	HF1 (complement factor H)	603075
OTHER AUTOINFLAMMATORY SYNDROMES			
Syndrome of periodic fever with aphthous stomatitis, pharyngitis, and cervical adenopathy (PFAPA)	Idiopathic	—	—
Autoinflammation, lipodystrophy, and dermatosis syndrome (Nakajo-Nishimura syndrome, JMP syndrome, CANDLE syndrome)	Autosomal recessive	PSMB8	256040
Systemic-onset juvenile idiopathic arthritis (SOJIA)	Complex inheritance	IL-6, *MIF* polymorphisms	604302
Adult-onset Still's disease	Idiopathic	—	—
Schnitzler's syndrome	Idiopathic	—	—
Behçet's disease	Complex inheritance	HLA-B51, polymorphisms in *IL10, IL23R, CCR1, STAT4, KLRC4, ERAP1, MEFV, TLR4*	109650
Crystalline arthropathies	Complex inheritance	*SLC2A9/GLUT9, ABCG2*	—

*Online Mendelian Inheritance in Man, an online catalogue of genetic disorders, available at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM. Accessed September 29, 2014.
†An updated list of disease-associated mutations is available online at http://fmf.igh.cnrs.fr/infevers. Accessed September 29, 2014.

genotypes of the *MEFV* and *SAA* genes. Amyloidosis in FMF is less common in the United States than in the Middle East. Abdominal fat aspirates are much less sensitive than rectal or renal biopsy in detecting the amyloidosis of FMF. The latter procedure may be preferred, because of the increasing recognition of nonamyloid glomerular disease in FMF. With early diagnosis, aggressive suppression of the acute phase response with colchicine or adjunctive agents may lead to improvement, but for patients with renal failure, early renal transplantation is preferred.

DIAGNOSIS

Based on a simple recessive model of inheritance, two mutations in *MEFV,* in *trans,* should be identified to establish the genetic diagnosis of FMF. Nevertheless, the interpretation of genetic testing is complicated by complex alleles consisting of various combinations of mutations in *cis,* as well as by the observations that as many as one third of patients with clinically typical FMF have only one demonstrable mutation in *MEFV,* and a few patients with typical disease have no identifiable *MEFV* mutations. These latter two findings suggest that, under some circumstances, one *MEFV* mutation may be sufficient for symptoms or that additional genes for FMF exist.

For these reasons, clinical data remain an essential part of the diagnosis of FMF, and genetic testing plays an adjunctive role in settings in which clinical experience is limited.[7] Clinical criteria emphasize attack duration (12 to 72 hours); recurrence of symptoms (three or more episodes); documented fever (rectal temperature > 38° C); painful manifestations in the abdomen, chest, joints, or skin; and the absence of other causative factors. The differential diagnosis includes the other hereditary recurrent fever syndromes (Table 261-2), as well as other conditions specific to the clinical setting. For patients with recurrent abdominal pain, considerations include gynecologic disorders, porphyria (Chapter 210; which can be distinguished by hypertension during attacks, dominant inheritance, and urine porphyrins), and hereditary angioedema (Chapter 252; which usually does not cause fever). The syndrome of periodic fever with aphthous stomatitis, pharyngitis, and cervical adenopathy is probably the most common cause of unexplained recurrent fever in children and is also included in the differential diagnosis. In patients presenting primarily with recurrent monoarthritis, joint aspiration for cultures and crystals may aid in excluding bacterial and crystalline arthritis, respectively.

Still's disease in children (systemic-onset juvenile idiopathic arthritis) and adults (adult-onset Still's disease) is also considered in the differential diagnosis. *Adult-onset Still's disease*[8] (see Table 261-1) is an uncommon autoinflammatory condition of unknown cause that is not considered to be hereditary. It is characterized by spiking fever, an evanescent salmon-pink maculopapular rash, arthritis, and neutrophilic leukocytosis. It can be clinically distinguished from FMF by the pattern of fever (intermittent quotidian in Still's disease vs. discrete episodes in FMF), the pattern of arthritis (chronic polyarthritis vs. intermittent monoarthritis), the characteristic skin involvement (evanescent rash vs. erysipeloid erythema), and the presence of lymphadenopathy (more common in Still's disease).

TABLE 261-2 CLINICAL FEATURES OF SELECTED HEREDITARY RECURRENT FEVER SYNDROMES

CLINICAL FEATURE	FMF	TRAPS	HIDS	FCAS	MWS	NOMID/CINCA
Typical ethnicity	Arab, Armenian, Italian, Jewish, Turkish	Any ethnicity	Dutch, other North European	European	European	Any
Attack duration	12-72 hr	Days to weeks	3-7 days	12-24 hr	1-2 days	Continuous, with flares
Abdominal attacks	Sterile peritonitis, constipation more often than diarrhea	Severe pain, vomiting, peritonitis	Sterile peritonitis, diarrhea, rarely constipation	Nausea	Abdominal pain	Not common
Pleural attacks	Common	Common	Rare	Not seen	Rare	Rare
Joint/bone involvement	Monoarthritis, rarely protracted arthritis in knee or hip	Arthritis in large joints, arthralgia	Arthralgia, symmetrical polyarthritis	Polyarthralgia	Polyarthralgia, oligoarthritis, clubbing	Epiphyseal overgrowth, contractures, intermittent or chronic arthritis, clubbing
Skin rash	Erysipeloid erythema on lower leg, ankle, foot	Migratory rash, underlying myalgia	Diffuse maculopapular rash, urticaria	Cold-induced urticaria-like rash	Urticaria-like rash	Urticaria-like rash
Lymphatic involvement	Splenomegaly, occasional lymphadenopathy	Splenomegaly, occasional lymphadenopathy	Cervical adenopathy in children	Not seen	Rare	Hepatosplenomegaly, adenopathy
Neurologic involvement	Aseptic meningitis?	Controversial	Headache	Headache	Sensorineural deafness	Sensorineural deafness, chronic aseptic meningitis, intellectual disability, headache
Ophthalmologic involvement	Rare	Conjunctivitis, periorbital edema, rarely uveitis	Uncommon	Conjunctivitis	Conjunctivitis, episcleritis	Uveitis, conjunctivitis, progressive vision loss
Vasculitis	Henoch-Schönlein purpura (HSP), polyarteritis nodosa	HSP, lymphocytic vasculitis	Cutaneous vasculitis common, rarely HSP	Not seen	Not seen	Occasional
Systemic amyloidosis	Risk depends on *MEFV* and *SAA* genotypes; more common in Middle East	Occurs in ~10%; risk increased with cysteine mutations	Rare	Rare	Occurs in ~25%	May develop in some, usually in adulthood

FCAS = familial cold autoinflammatory syndrome; FMF = familial Mediterranean fever; HIDS = hyperimmunoglobulinemia D with periodic fever syndrome; MWS = Muckle-Wells syndrome; NOMID/CINCA = neonatal-onset multisystem inflammatory disease/chronic infantile neurologic cutaneous and articular syndrome; TRAPS = tumor necrosis factor receptor–associated periodic syndrome.

TREATMENT

Rx

The mainstay of therapy for FMF is daily oral colchicine, which can prevent both acute attacks of FMF and the development of systemic amyloidosis. Colchicine probably works by several mechanisms, including its effects on inhibiting leukocyte adhesion and modulating cytokine production. In adults, the therapeutic dose is 1.2 to 1.8 mg/day, and nearly 90% of patients note significant improvement at this dose. The major side effects are gastrointestinal, and they can usually be minimized by gradually increasing the dosage and avoiding milk products in patients who develop lactose intolerance. Most experts continue to prescribe colchicine to patients during pregnancy, with the recommendation that amniocentesis be performed to exclude trisomy 21, for which there may be a slightly increased risk. Use of colchicine in lactating women is considered safe. Intravenous colchicine should be used with extreme caution, if at all, in FMF, because fatal toxicity has been reported in patients already receiving oral colchicine who are given the drug intravenously. IL-1 inhibitors[9] may be effective in patients who are poorly responsive to colchicine or who cannot tolerate therapeutic doses.

Tumor Necrosis Factor Receptor–Associated Periodic Syndrome

DEFINITION

Worldwide, the tumor necrosis factor (TNF) receptor–associated periodic syndrome (TRAPS) is the second most frequently diagnosed hereditary recurrent fever syndrome, behind FMF. TRAPS is defined by recurrent episodes of fever and localized inflammation, in many ways resembling FMF, but differing in key details (noted later) and caused by mutations in the 55-kD receptor for TNF (TNFRSF1A, TNFR1, p55, CD120a). Whereas a positive genetic test is not necessary to diagnose FMF, the diagnosis of TRAPS requires the identification of a TNF receptor mutation. One of the first well-characterized families with what was later defined as TRAPS was of Irish ancestry, and the condition was termed *familial Hibernian fever* to emphasize the ethnic background and clinical differences from FMF. However, with the discovery of TNF receptor mutations in families of other ancestries, the ethnically neutral TRAPS nomenclature was proposed.

PATHOBIOLOGY

The p55 TNF receptor is composed of four cysteine-rich extracellular domains, a transmembrane region, and an intracellular death domain. To date, nearly all of the more than 90 mutations described are in the extracellular domains and approximately one third are missense substitutions of cysteine residues that abolish highly conserved disulfide bonds. The initial description of TRAPS documented a defect in activation-induced ectodomain cleavage of the p55 receptor in patients with the C52F *TNFRSF1A* mutation, possibly leading to a defect in homeostasis by impaired downregulation of membrane receptors and diminished shedding of potentially antagonistic soluble receptor molecules. More recent studies indicate a more complex pathogenetic picture, because not all mutant receptors exhibit this shedding defect. Additional mechanisms by which p55 mutations may lead to autoinflammation include impaired leukocyte apoptosis and impaired intracellular receptor trafficking, with possible constitutive activation of mitogen-activated protein (MAP) kinases by intracellular aggregates of mutant receptors.[10]

DIAGNOSIS

Although genetic testing is necessary for the diagnosis of TRAPS, certain clinical clues can help distinguish TRAPS from FMF. These include ethnicity (FMF is seen predominantly in Mediterranean and Middle Eastern populations, whereas TRAPS has a more widespread distribution), mode of inheritance (autosomal recessive in FMF, dominant in TRAPS), and duration of attacks, which tends to be longer in TRAPS and sometimes approaches continuous symptoms. The rash of FMF is typically erysipeloid erythema on the lower extremity, whereas patients with TRAPS often have a distinctive erythematous rash, often with underlying myalgia, which may migrate on the trunk or centrifugally on the extremities. Ocular involvement, with periorbital edema, conjunctivitis, and occasionally even uveitis, is observed in

TRAPS but not in FMF. Finally, whereas colchicine is much more effective than corticosteroids in FMF, the opposite is true in TRAPS. Nevertheless, aside from the difference in duration and susceptibility to pharmacologic intervention, the abdominal, pleural, synovial, and even scrotal manifestations of the two diseases are rather similar. The usual age of onset for TRAPS is also in childhood, and systemic AA amyloidosis is seen in approximately 10% of untreated patients with TRAPS. As in FMF, life expectancy in TRAPS is normal in patients whose disease is not complicated by amyloidosis.

As noted earlier, the diagnosis of TRAPS is established by the identification of *TNFRSF1A* mutations in the appropriate clinical setting. One variant, the substitution of glutamine for arginine at residue 92 (R92Q), is present in more than 1% of whites and may be associated with a broader spectrum of symptoms than is typically seen in TRAPS, including early inflammatory arthritis or, in some cases, no symptoms at all. The substitution of lysine for proline at residue 46 (P46L) is common among African American patients with TRAPS and is associated with a receptor shedding defect, but it is also seen among healthy African American controls. These findings establish a "gray zone" for the diagnosis of TRAPS and emphasize the potential role of polymorphisms in the recurrent fever genes in other more common phenotypes.

TREATMENT

Rx

The treatment of TRAPS depends on the frequency and severity of attacks. Patients with relatively infrequent, mild episodes may respond to nonsteroidal anti-inflammatory drugs (NSAIDs). Patients with more severe attacks that occur infrequently may be treated with corticosteroids, although increasing doses may be required as the episodes become more frequent and toxicities may become limiting. For patients with severe attacks occurring once a month or more frequently, treatment with etanercept, the soluble p75 TNF receptor:Fc fusion protein, may be warranted. This may be a unique effect of etanercept, because there is anecdotal evidence that monoclonal antibodies against TNF may actually exacerbate TRAPS. Consistent with a model implicating multiple cytokines in the pathogenesis of TRAPS, IL-1 inhibitors have also been found effective in TRAPS.[11]

INTERLEUKIN-1–ASSOCIATED AUTOINFLAMMATORY DISEASES

The IL-1–associated autoinflammatory diseases are linked by markedly increased expression or cellular responsiveness to this cytokine, and dramatic resolution of symptoms with IL-1 blockade. Interleukin-1α and Interleukin-1β (IL-1α and IL-1β) are structurally related cytokines released from cells triggered by a number of inflammatory stimuli, such as lipopolysaccharide. They mediate inflammatory responses by binding to a common receptor that is present on the surface of a wide variety of cell types and signals to activate inflammatory genes through the nuclear factor kappa B (NF-KB) transcription factor complex. IL-1 is part of a larger family of cytokines including IL-18, IL-33, and IL-36, which bind to related receptors and share the property of not having a characteristic signal peptide that normally targets cytokines to secretory vesicles. Because of this, IL-1 family cytokines may be secreted only by dead or dying cells, functioning as molecular markers of cellular stress, which can trigger beneficial inflammatory responses to infection and injury. IL-1β and IL-18 are unique in that they are not biologically active until cleaved by the protease caspase-1. Caspase-1 is itself activated in cytoplasmic protein complexes containing various sensor proteins such as NLRP3, and the adapter protein ASC. These complexes are referred to as *inflammasomes* because of their ability to trigger IL-1–mediated inflammation. Autoinflammatory diseases caused by mutations in genes encoding proteins that process or sense IL-1 are described later.

Cryopyrin-Associated Periodic Syndromes: The Cryopyrinopathies

Three rare, recurrent febrile disorders usually beginning early in life have been associated with mutations in *NLRP3* (formerly *CIAS1*), the gene encoding a protein variously named cryopyrin, NLRP3, NALP3, PYPAF1, or CATERPILLER 1.1, a key component of the NLRP3 inflammasome that activates caspase-1. These disorders are referred to as cryopyrinopathies or cryopyrin-associated periodic syndromes (CAPS). The least severe clinical phenotype is familial cold autoinflammatory syndrome (FCAS; formerly called familial cold urticaria), which is dominantly inherited and is notable for day-long attacks of chills, fever, headache, diffuse urticarial skin rash, arthralgia, and conjunctivitis, precipitated by generalized cold exposure. Amyloidosis is rare in FCAS. Of intermediate severity is Muckle-Wells syndrome (MWS), also dominantly inherited, in which 1- to 2-day episodes of chills, fever, urticarial rash, limb pain, and arthritis occur independently of cold exposure. Sensorineural hearing loss is common in MWS, and systemic amyloidosis also may occur. The most severe *NLRP3*-associated phenotype is neonatal-onset multisystem inflammatory disease (NOMID), known in Europe as chronic infantile neurologic cutaneous and articular (CINCA) syndrome. It is usually sporadic owing to the reduced reproductive fitness of most affected individuals. Fever and constitutional symptoms occur almost daily, often from birth, with generalized urticarial skin rash, a peculiar arthropathy characterized by epiphyseal overgrowth of the long bones, and central nervous system (CNS) involvement that includes chronic aseptic meningitis, uveitis, and cochlear inflammation, which may lead to intellectual disability, blindness, and deafness. In all three cryopyrinopathies, the rash is not true urticaria because there is a neutrophilic rather than a mast cell infiltrate and serum histamine levels are normal.

The protein mutated in all three disorders is NLRP3, a critical component of the eponymous NLRP3 inflammasome, which serves as an intracellular scaffold for the processing of IL-1β. The alternative name of this protein, cryopyrin, refers to its aminoterminal PYRIN domain, the basis for a structural and functional relationship to the protein mutated in FMF. Disease-associated cryopyrin mutations are thought to decrease the threshold for inflammasome activation, thereby increasing IL-1β production. The discovery that the NLRP3 inflammasome is also necessary for IL-1β production in response to crystalline forms of monosodium urate and calcium pyrophosphate connected the pathophysiology of these rare autoinflammatory diseases to crystal-induced arthritis (Chapter 273), which shares some clinical features, such as episodic, self-limited attacks, with autoinflammatory diseases.

Because there are patients with FCAS, MWS, and NOMID/CINCA without demonstrable *NLRP3* mutations, these diagnoses remain clinical, although genetic testing serves as a valuable adjunct and has greatly increased the recognition of all three conditions. Deep sequencing has identified somatic *NLRP3* mutations in some patients with symptoms consistent with CAPS who are negative for mutations by standard genetic testing. In addition, overlap syndromes that are intermediate between FCAS and MWS and between MWS and NOMID/CINCA have been reported.

TREATMENT

Rx

Blockade with anakinra, a recombinant IL-1 receptor antagonist, is effective in controlling fever and acute phase reactants in all three cryopyrinopathies, and longitudinal analysis of a large series of patients at the National Institutes of Health showed that long-term treatment with anakinra markedly decreased CNS inflammation and end-organ damage in NOMID/CINCA, which led to the regulatory approval of anakinra for the treatment of this condition in the United States and Europe. Recent studies have also documented the efficacy of rilonacept, another soluble IL-1 blocker, and canakinumab, a monoclonal antibody against IL-1β, in FCAS and MWS, although these agents may be less effective against NOMID/CINCA because of reduced penetration into the CNS. The efficacy of canakinumab suggests that the major biologic effect of cryopyrin in humans is mediated through IL-1β rather than by IL-1α or other distinct inflammatory pathways.

Deficiency of Interleukin-1 Receptor Antagonist

Deficiency of the IL-1 receptor antagonist (DIRA) is characterized by the neonatal onset of a pustular skin rash, multifocal osteomyelitis, periostitis, and, rarely, vasculitis.[12] Fever is not a prominent finding, although acute phase reactants are markedly elevated. DIRA is caused by recessively inherited loss-of-function mutations in *IL1RN*, which encodes the IL-1 receptor antagonist (IL-1Ra). Patients usually present within the first 2 weeks of life with skin lesions ranging from discrete crops of pustules to generalized severe pustulosis or ichthyosiform lesions. Histologic examination demonstrates extensive neutrophilic infiltrates in the dermis and epidermis. Typical radiographic findings include multifocal osteolytic lesions, periosteal elevation of the long bones, heterotopic ossification of the proximal femurs, and widening of the anterior rib ends. Bone biopsies demonstrate sterile purulent osteomyelitis, fibrosis, and sclerosis. To date, five different *IL1RN* mutations have been identified, three of which are truncating point mutations that drastically reduce IL-1Ra messenger RNA and protein levels. The fourth is a 15bp in-frame deletion, and the fifth is a 175-kilobase genomic deletion in

chromosome 2q that subsumes *IL1RN* and five other genes in the IL-1 family. In DIRA, the lack of IL-1Ra leads to unopposed IL-1β and IL-1α signaling, whereas in the cryopyrinopathies, *NLRP3* mutations lead to inflammasome activation and increased IL-1β production. DIRA patients respond dramatically to anakinra, a recombinant form of the protein they lack.

OTHER INHERITED SYSTEMIC AUTOINFLAMMATORY DISEASES

Syndrome of Pyogenic Arthritis with Pyoderma Gangrenosum and Acne

The syndrome of pyogenic arthritis with pyoderma gangrenosum and acne (PAPA) is a rare, dominantly inherited autoinflammatory disease characterized by intermittent episodes of sterile pyogenic arthritis, pyoderma gangrenosum, and severe cystic acne. It is caused by mutations in proline-serine-threonine phosphatase–interacting protein 1 (PSTPIP1), also known as CD2BP1. PSTPIP1 is a cytoskeletal protein that interacts with certain other proteins involved in the immune response, including CD2; the Wiskott-Aldrich syndrome protein (WASP); a phosphatase denoted PTP-PEST; and pyrin, the FMF protein. PAPA mutations abrogate the binding of PSTPIP1 to PTP-PEST, leading to hyperphosphorylation of PSTPIP1 and increased binding to pyrin. Both in patients and in cell lines, this finding is associated with markedly increased IL-1β production. Early in life, PAPA tends to present with monoarticular or pauciarticular pyogenic arthritis, sometimes induced by trauma. In the absence of treatment, arthritis may progress to severe joint damage and ankylosis. As patients reach puberty, skin manifestations begin to predominate, including disfiguring cystic acne. Pathergy also may develop, and extensive pyoderma gangrenosum may require opiates for pain control. The diagnosis of PAPA syndrome is made by documenting *PSTPIP1* mutations in the appropriate clinical setting. High doses of corticosteroids have been used in PAPA, with varying success, and patients with arthritis sometimes require aspiration, intra-articular corticosteroids, or open drainage. Newer investigational approaches for PAPA syndrome focus on the use of targeted cytokine inhibitors. Anecdotal evidence supports the use of anakinra for the arthritis and monoclonal anti-TNF antibodies for the pyoderma gangrenosum of PAPA.

Granulomatous Inflammatory Arthritis, Dermatitis, and Uveitis (Blau's Syndrome)

Blau's syndrome is a rare, dominantly inherited illness characterized by the following features: early-onset granulomatous synovitis often complicated by cyst formation and camptodactyly (flexion contractures of the fingers and toes); granulomatous anterior and posterior uveitis, sometimes causing retinal detachment, glaucoma, cataracts, and blindness; and an intermittent papular rash with noncaseating granulomas. Lung or other visceral involvement is generally not present. However, visceral involvement of the liver and spleen is observed in early-onset sarcoidosis (Chapter 95), which is phenotypically quite similar to Blau's syndrome. Both Blau's syndrome and some cases of early-onset sarcoidosis are caused by mutations in *NOD2/CARD15*. Distinct variants of *NOD2/CARD15* have been associated with susceptibility to Crohn's disease, which manifests as granulomatous inflammation of the gastrointestinal tract (Chapter 141). The protein encoded by this gene is thought to be an intracellular sensor of bacterial products. Crohn's disease—associated mutations in the ligand-binding, leucine-rich repeat region of the protein may alter responses to bacterial products in the gastrointestinal tract to cause inflammation, whereas Blau's syndrome mutations in the nucleotide binding domain may lead to constitutive extraintestinal inflammation. Topical and systemic corticosteroids are currently the mainstay of treatment of Blau's syndrome. There are case reports of the efficacy of TNF and IL-1 inhibitors in this disease.

Hyperimmunoglobulinemia D with Periodic Fever Syndrome

Hyperimmunoglobulinemia D with periodic fever syndrome (HIDS) was first described in 1984 as an FMF-like illness seen in six patients of Dutch ancestry. Besides the difference in ethnicity, a key distinction was the observation of extremely high levels of immunoglobulin D (IgD) in the serum of these patients, thus prompting the HIDS nomenclature. HIDS is now recognized in a broader ethnic distribution, although northern Europeans still predominate. Overall, HIDS is still quite rare. Family studies documented autosomal recessive inheritance. In 1999, patients with HIDS were found to have mutations in *MVK,* which encodes the mevalonate kinase enzyme involved in the biosynthesis of cholesterol and nonsterol isoprenes. Enzyme activity in patients is markedly reduced, but not absent. The elevated immunoglobulin D (IgD) levels seen in HIDS appear to be an epiphenomenon and do not correlate with disease severity either among patients or in a given patient over time, although IgD may contribute to the release of proinflammatory cytokines in vitro. Moreover, modest elevations of IgD are seen in several inflammatory conditions, including chronic infections, and can be observed in other hereditary recurrent fever syndromes. Up to 20% of patients (particularly young children) with typical recurrent fevers and *MVK* mutations have normal serum IgD levels. Current data suggest that isoprenoid deficiency may play a more important pathogenic role in the pathophysiology of HIDS. In vitro studies suggest that isoprenoid deficiency may lead to excessive IL-1β production, and increased body temperature can further decrease mevalonate kinase enzymatic activity, thereby creating a vicious circle in which infection or immunization can precipitate HIDS attacks. One of the well-recognized clinical characteristics of HIDS is the provocation of attacks by immunizations. Other distinguishing clinical features include a very early age of onset (average age, 6 months), a duration of attacks intermediate between FMF and TRAPS (3 to 7 days), prominent cervical lymphadenopathy during attacks, polyarticular joint involvement, a diffuse maculopapular rash, the predominance of diarrhea over constipation with abdominal attacks, and the infrequency of pleuritic attacks or systemic amyloidosis.

The diagnosis of HIDS can be established in a patient with recurrent episodes of fever and typical associated findings by documenting either two mutations in *MVK* or elevated levels of mevalonic acid, the substrate for mevalonate kinase, in the urine during attacks. Approximately 10% of patients with otherwise typical disease have only a single identifiable *MVK* mutation. The significance of elevated IgD without genetic or biochemical findings remains unknown. NSAIDs or corticosteroids are sometimes useful in the treatment of the arthritic manifestations of HIDS. Colchicine is generally not effective. Numerous agents are investigational in HIDS, including the statins, TNF inhibitors, and IL-1 inhibitors. Patients with HIDS have a normal lifespan, and attacks may become somewhat less frequent in adulthood.

Proteasome-Associated Systemic Inflammatory Diseases

Recently, a constellation of diseases have been described linked to recessive loss of function mutations in *PSMB8,* which encodes the β5i subunit of the proteasome, also known as LMP7. An autosomal recessive syndrome in adults characterized by recurrent fevers, progressive lipodystrophy, joint contractures, and cardiac manifestations was linked to homozygous missense mutations in *PSMB8.*[13] Patients with a pediatric syndrome termed CANDLE (chronic atypical neutrophilic dermatosis with lipodystrophy and elevated temperature) were found to have homozygous missense and nonsense mutations in *PSMB8,* with some patients having only one known *PSMB8* mutation. It is not yet clear whether these syndromes represent identical diseases related to loss of function of β5i. The β5i proteasome subunit is one of the subunits that that are induced in immune cells through immune stimuli such as interferons, altering the proteasome so that it more efficiently processes peptides for antigen presentation to T cells. However, there is no indication of a T cell component to this disease, and studies have shown that the β5i proteasome subunit can be expressed in nonimmune cells such as adipocytes. A striking interferon transcriptional signature, similar to that seen in systemic lupus erythematosus, was observed in circulating blood cells from patients with CANDLE. Defective degradation of proteins in cells lacking β5i may result in buildup of ubiquitinated proteins, which somehow triggers interferon production, or PSMB8 deficiency may enhance interferon signaling by stabilizing components of the interferon signal transduction machinery that are negatively regulated by ubiquitin-proteasome degradation. Whichever the mechanisms, the link to interferon hyperactivity suggests that blocking interferons with antibodies or inhibitors of interferon signal transcution may be effective in the therapy of CANDLE and possibly other PSMB8-associated syndromes.

New Autoinflammatory Syndromes and the Promise of Whole-Exome Sequencing

Recent years have seen a dramatic acceleration in the pace of discovery of new mendelian inflammatory diseases as a result of the availability of whole-exome sequencing, which allows unbiased identification of disease-causing mutations in protein coding sequences, although it should be noted that accurate clinical description of these syndromes is as important as the genetic tools for identification of new syndromes. These discoveries have confirmed

the role of gene products in human inflammation that were identified in other animal model systems and identified new genes and proteins not previously thought to be involved in the regulation of inflammation. Early-onset, apparently sporadic, cases of inflammatory syndromes have often turned out to be due to de novo mutations in a child when screened against parental DNA. For example, inherited gain-of-function mutations in *CARD14,* encoding an adapter protein in innate immune sensing, cause dominantly inherited familial psoriasis, and a more severe gain-of-function de novo mutation in the same gene caused infantile-onset severe pustular psoriasis. Recessive mutations causing systemic autoinflammatory disease also have been identified by whole-exome sequencing from just a few families. Recent examples of novel diseases discovered through these methods include a syndrome characterized by fevers, early-onset strokes, and vasculopathy or frank vasculitis caused by autosomal recessive mutations in *CECR1,* encoding adenosine deaminase 2 (ADA2), a serum protein with newly recognized effects on macrophage differentiation and vascular development.[14] Recessive mutations in *HOIL1* that impair the addition of linear ubiquitin chains to receptor signaling complexes cause a complex syndrome marked by autoinflammation and immunodeficiency and intramuscular glycogen deposition. Gain-of-function mutations in *PLCG2,* encoding phospholipase Cγ2, an enzyme with essential functions in B-cell receptor and Fc Rceptor signaling, cause a dominantly inherited autoinflammatory syndrome characterized by blistering skin lesions, bronchiolitis, arthralgia, ocular inflammation, and enterocolitis in the absence of autoantibodies.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Broderick L, De Nardo D, Franklin BS, et al. The inflammasomes and autoinflammatory syndromes. *Annu Rev Pathol*. 2015;10:395-424.
2. Henderson C, Goldbach-Mansky R. Monogenic autoinflammatory diseases: new insights into clinical aspects and pathogenesis. *Curr Opin Rheumatol*. 2010;22:567-578.
3. Jacobs Z, Ciaccio CE. Periodic fever syndromes. *Curr Allergy Asthma Rep*. 2010;10:393-404.
4. Kastner DL, Aksentijevich I, Goldbach-Mansky R. Autoinflammatory disease reloaded: a clinical perspective. *Cell*. 2010;140:784-790.
5. Wurster VM, Carlucci JG, Edwards KM. Periodic fever syndromes. *Pediatr Ann*. 2011;40:48-54.
6. Infevers database. A compendium of mutations associated with known autoinflammatory syndromes with links to relevant publications. http://fmf.igh.cnrs.fr/ISSAID/infevers/. Accessed January 20, 2015.
7. Berkun Y, Eisenstein EM. Diagnostic criteria of familial Mediterranean fever. *Autoimmun Rev*. 2014;13:388-390.
8. Gerfaud-Valentin M, Jamilloux Y, Iwaz J, et al. Adult-onset Still's disease. *Autoimmun Rev*. 2014;13:708-722.
9. Jesus AA, Goldbach-Mansky R. IL-1 blockade in autoimflammatory syndromes. *Annu Rev Med*. 2014;65:223-244.
10. Bachetti T, Ceccherini I. Tumor necrosis factor receptor-associated periodic syndrome as a model linking autophagy and inflammation in protein aggregation diseases. *J Mol Med (Berl)*. 2014;92:583-594.
11. Ter Haar NM, Frenkel J. Treatment of hereditary autoinflammatory diseases. *Curr Opin Rheumatol*. 2014;26:252-258.
12. Aksentijevich I, Masters SL, Ferguson PJ, et al. An autoinflammatory disease with deficiency of the interleukin-1-receptor antagonist. *N Engl J Med*. 2009;360:2426-2437.
13. Liu Y, Ramot Y, Torrelo A, et al. Mutations in proteasome subunit beta type 8 cause chronic atypical neutrophilic dermatosis with lipodystrophy and elevated temperature with evidence of genetic and phenotypic heterogeneity. *Arthritis Rheum*. 2012;64:895-907.
14. Zhou Q, Yang D, Ombrello A, et al. Early-onset stroke and vasculopathy associated with mutations in ADA2. *N Engl J Med*. 2014;370:907-916.

REVIEW QUESTIONS

1. Which of the following is not commonly associated with the development of AA amyloidosis?

A. Hyper-IgD syndrome (HIDS)
B. Familial Mediterranean fever (FMF)
C. Tumor necrosis factor receptor–associated periodic syndrome (TRAPS)
D. Muckle-Wells syndrome

Answer: A All of the listed periodic fever syndromes are commonly associated with the development of AA amyloidosis with the exception of hyper-IgD syndrome, which generally has a benign course.

2. Patients with which monogenic autoinflammatory disease often have a history of flaring after routine immunizations?

A. FMF
B. HIDS
C. TRAPS
D. FCAS

Answer: B Of these syndromes, only HIDS has a substantiated association of flares with routine immunizations. Psychological stress is often cited as a trigger for flares in FMF and TRAPS, and cold exposure triggers a systemic inflammatory flare and hivelike skin lesions in FCAS.

3. A 30-year-old woman of Armenian and Jewish ancestry presents with a life-long history of unexplained febrile episodes. These attacks are variable in length, with the shortest lasting a few days and the longest lasting over 1 month. The fevers are accompanied by severe abdominal pain or pleuritic chest pain, periorbital edema, arthralgia, and a painful migratory erythematous rash. Corticosteroids ameliorate her symptoms. During pregnancy 7 years ago she was totally free of fevers, but she developed a severe attack in the postpartum period. She is currently not experiencing an attack, but has an erythrocyte sedimentation rate of 85 (Westergren), C-reactive protein of 100 mg/L, urine protein-to-creatinine ratio of 5.3, and serum creatinine of 2.5. Which of the following is most likely true?

A. The patient probably has familial Mediterranean fever, and should undergo genetic testing for *MEFV* mutations, have a rectal biopsy to rule out amyloidosis, and commence treatment with intravenous colchicine.
B. The patient has chronic atypical neutrophilic dermatosis with lipodystropy and elevated temperature (CANDLE syndrome), with the T75M mutation in *PSMB8,* encoding a component of the immunoproteasome.
C. The patient has a mutation at a cysteine residue in the extracellular domain of the p55 TNF receptor and possible amyloidosis. She should undergo a rectal biopsy and should be treated with etanercept or anakinra in an effort to normalize her acute phase reactants.
D. The patient has a mutation in the *NLRP3 (CIAS1)* gene, and could be treated with anakinra, rilonacept, or canakinumab for neonatal-onset multisystem inflammatory disease (NOMID).

Answer: C Despite the demographic information that may point toward FMF, the presence of longer disease flares of longer than 7 days, periorbital edema, painful migratory rash, and remission of symptoms with pregnancy are all characteristic of the TNFR1-associated periodic syndrome (TRAPS). Patients with TRAPS and structure-disrupting mutations in TNFR1 have an elevated risk for amyloidosis. TRAPS does not have a predilection for a specific ethnic group or geographic location.

4. A 3-year-old boy of northern European ancestry presents to the autoinflammatory disease clinic for an initial evaluation. Per his mother's report, he began experiencing febrile attacks at the age of 3 months that last 4 to 5 days and occur every 1 to 2 months. Associated symptoms include a nonpruritic macular rash and oral ulcers, and on three occasions he has had genital ulcers. He tends to have flares approximately 2 days after immunizations and after routine viral illnesses. There is no family history of febrile illnesses. He is treated with ibuprofen and acetaminophen and has been given a couple of courses of prednisolone without complete resolution of symptoms. He had been well before his visit, but his mother thinks he is starting to have a flare. On physical examination, he has conjunctival injection, cervical lymphadenopathy but no rashes, oral or genital ulcers, or arthritis. Laboratory studies reveal an ESR of 57 mm/hour, CRP of 55 mg/L, and IgD level at the top of the normal range. Which of the following is most likely true?

A. The patient probably has periodic fever with aphthous stomatitis, pharyngitis, and cervical adenitis (PFAPA). He should be given either 1 mg/kg of prednisolone at the onset of symptoms or anakinra 2 mg/kg SC for 1 to 2 days at the onset of symptoms.
B. The patient has Behçet's disease and should consider colchicine for his oral ulcers. He should undergo HLA typing to assess for the presence of HLA-B51 or other Behçet's-associated MHC findings.
C. The patient has hyperimmunoglobulinemia D syndrome and should initiate therapy with anakinra at the time of flares or weekly etanercept. He should be tested for mutations in the *MVK* gene.
D. The patient has cyclic neutropenia and should be tested for mutations in the *ELA-2* gene. He should also have weekly CBC drawn to assess for periodic neutropenia.

Answer: C Despite the oral and genital ulcers that might suggest Behçet's disease, the history of flares after vaccinations or viral illness, the pattern and age of onset, and the ethnic background more strongly suggest that the patient has hyperimmunoglobulinemia D syndrome. IgD levels can be normal in this disease, and a normal IgD should not rule out this diagnosis, which can be confirmed by finding a heterozygous mutation in the mevalonate kinase gene.

5. A 16-year-old high school student is evaluated in the emergency department for 18 hours of abdominal pain, pleurisy, and fever. At that time, he had diffuse rebound tenderness, temperature of 102.5° F (39.2°degrees C), and WBC of 17,000/mm^3. He was admitted to general surgery and after 12 hours of persistent symptoms, he was taken to the operating room for exploratory surgery. A normal-appearing appendix was laparascopically removed. Postoperatively, the fever persisted and he was noted to have a left pleural effusion. Additional history reveals that he has had similar episodes of febrile illness in the past. On physical examination, he has no rash but has moderate abdominal tenderness with guarding and a left pleural friction rub. Laboratory data reveal an ESR of 80 mm/hour and a negative ANA and RF. While waiting for the results of genetic testing, which of the following medications is *most* appropriate to start at this time?

A. Indomethacin, 50 mg orally three times daily
B. Prednisone, 40 mg orally daily
C. Colchicine, 0.6 mg orally twice daily
D. Anakinra, 100 mg subcutaneously daily
E. Acetylsalicylic acid, 1000 mg orally daily

Answer: C The duration of fevers, association with pleural and peritoneal inflammation, and lack of other localizing symptoms are most consistent in this case with familial Mediterranean fever. Colchicine is effective in the prophylaxis of future attacks in FMF. Indomethacin and prednisone can be of some value in terminating attacks but should be reserved for refractory cases. Anakinra can be effective in the treatment of refractory FMF, but there is no need to use this agent in patients who respond to colchicine.

262

OSTEOARTHRITIS

JOEL A. BLOCK AND CARLA SCANZELLO

DEFINITION

Osteoarthritis (OA) is a heterogeneous disease that has many names, including degenerative joint disease and osteoarthrosis. It is a joint disease characterized clinically by pain and functional loss. Although local inflammation of involved joints is common, OA is not associated with a systemic inflammatory process, in contrast to other arthritides. It is the most common form of arthritis and accounts for the overwhelming majority of arthritis cases, and its prevalence is expected to rise dramatically during the next 20 years as global populations age. OA is often neglected either because it is not a fatal disease or because many physicians assume it is a normal part of aging and is not inherently treatable. Yet it results in vast direct medical costs and significant loss of work; it is the leading indication for total joint replacement and is a leading cause of work disability. Formal definitions of OA have evolved as our understanding of pathophysiology has progressed. Whereas it conventionally had been considered primarily a degenerative process of cartilage, it is now clear that OA involves the entire joint. Thus, a modern definition of OA is a *painful* degenerative process involving progressive deterioration of all joint structures and remodeling of subchondral bone that is not primarily inflammatory. It is important to distinguish true OA from asymptomatic structural degeneration of joints that is virtually universal during normal aging.

EPIDEMIOLOGY

As an age-related disease, OA prevalence has risen substantially with the aging population of the developed world; an estimated 27 million people had physician-diagnosed OA in the United States in 2005, increased from 21 million a decade earlier, and this number is expected to reach 67 million patients with clinically significant OA by 2030. For epidemiologic studies, OA is often defined radiographically by the presence of osteophytes, joint space narrowing, and subchondral sclerosis. However, a substantial number of individuals have these x-ray changes but remain clinically asymptomatic. Thus, estimates of OA prevalence vary widely depending on whether one is assessing *radiographic* or *symptomatic* OA. In either case, the lifetime risk is exceedingly high. There is general concordance among epidemiologic studies across North American, Asian, and European populations that the prevalence of radiographic OA in the knees, hips, and hands is quite low before age 45 and increases dramatically with aging, with most people having x-ray evidence of OA in at least one joint by the seventh decade. Symptomatic knee OA affects between 7% and 17% of those older than 44 years, with rates increasing with age; women have higher prevalence than men, and African Americans have higher prevalence than white Americans. Symptomatic hip OA is less prevalent than knee OA, with overall rates between 6.7% and 9.7% among those over age 44; as with knee OA, prevalence is higher among the elderly, women, and African Americans.[1] Symptomatic hand OA affects at least 6.8% of those older than 25, occurring in women more than two-fold more frequently than in men. Hand OA may be less common in African Americans than in white Americans.

Risk factors for the development of OA and for specific joint involvement have been extensively studied. Among nonmodifiable risk factors, the strongest is aging. This is true both for radiographic changes of OA and for symptomatic involvement. In addition, female sex is a risk for prevalence and severity of OA, especially after menopause. There is a significant heritable component, particularly for hand OA and hip OA. This component is estimated at 48 to 65% for so-called generalized OA characterized by osteophytes of the distal interphalangeal joints (Heberden nodes) or the proximal interphalangeal joints (Bouchard's nodes).

Modifiable OA risk factors may provide clues for preventive strategies. The most important of these is obesity, which alone confers an approximately three-fold increased risk for incident OA. Occupational and lifestyle activities that involve repeated trauma or excessive loading may be associated with increased risk for OA. These include chronic squatting, bending, and lifting such as by warehouse workers and laborers, who have increased knee involvement, and, classically, pneumatic drill operators who develop OA of the wrist and elbow. Significant trauma, such as major knee or ankle injury, is strongly associated with subsequent development of OA in the injured joint. Aberrant loading of joints is an important risk factor for the development and progression of OA. For example, excessive loading of the knee has been observed to result in significantly increased risk for progression to advanced knee OA. In addition, joint alignment is an important parameter in OA, and malalignment at the knee is among the strongest predictors of OA progression.[2]

PATHOBIOLOGY

Although degeneration of articular cartilage is a central common pathway in OA, multiple joint and periarticular tissues are compromised and contribute to clinical manifestations. Pathologic changes in synovium, ligaments, supporting musculature, and fibrocartilagenous structures such as the menisci in the knee are common.[3] Unlike autoimmune arthritides, OA does not affect extra-articular organs. However, the chronic pain of OA involves both the peripheral and central nervous system (CNS) and OA-related disability degrades the general physical and mental health of the patient. Appreciation of the global effects of OA has important implications for current and future treatment approaches.

Tissues Central to the Osteoarthritis Process

Cartilage

The hallmark of osteoarthritis is progressive deterioration of articular cartilage. Normal articular cartilage distributes loads across joint surfaces and allows for almost frictionless joint motion. These functions are furnished by the extracellular matrix, which accounts for more than 90% of the tissue volume and is organized by a network of collagen type II fibers, which provides tensile strength, entrapping aggrecan complexes, a proteoglycan that confers compressive stiffness and resilience (Fig. 262-1A and C). There is one major cell type, the chondrocyte, that synthesizes these matrix components. In mature cartilage, turnover of extracellular matrix molecules, particularly collagen type II, is slow.

Cartilage change in OA begins with swelling of the matrix, then progresses through stages of surface roughening, fibrillation, fissuring, and eventually full-thickness erosion. These are accompanied by activation of chondrocytes to increase synthesis of proteolytic enzymes that degrade matrix.[4] Matrix metalloproteinase-13 (MMP-13; collagenase-3) plays a central role in collagen type II degradation, while ADAMTS-4 and -5 (a disintegrin and metalloproteinase with thrombospondin motifs-4 and -5) proteases are important

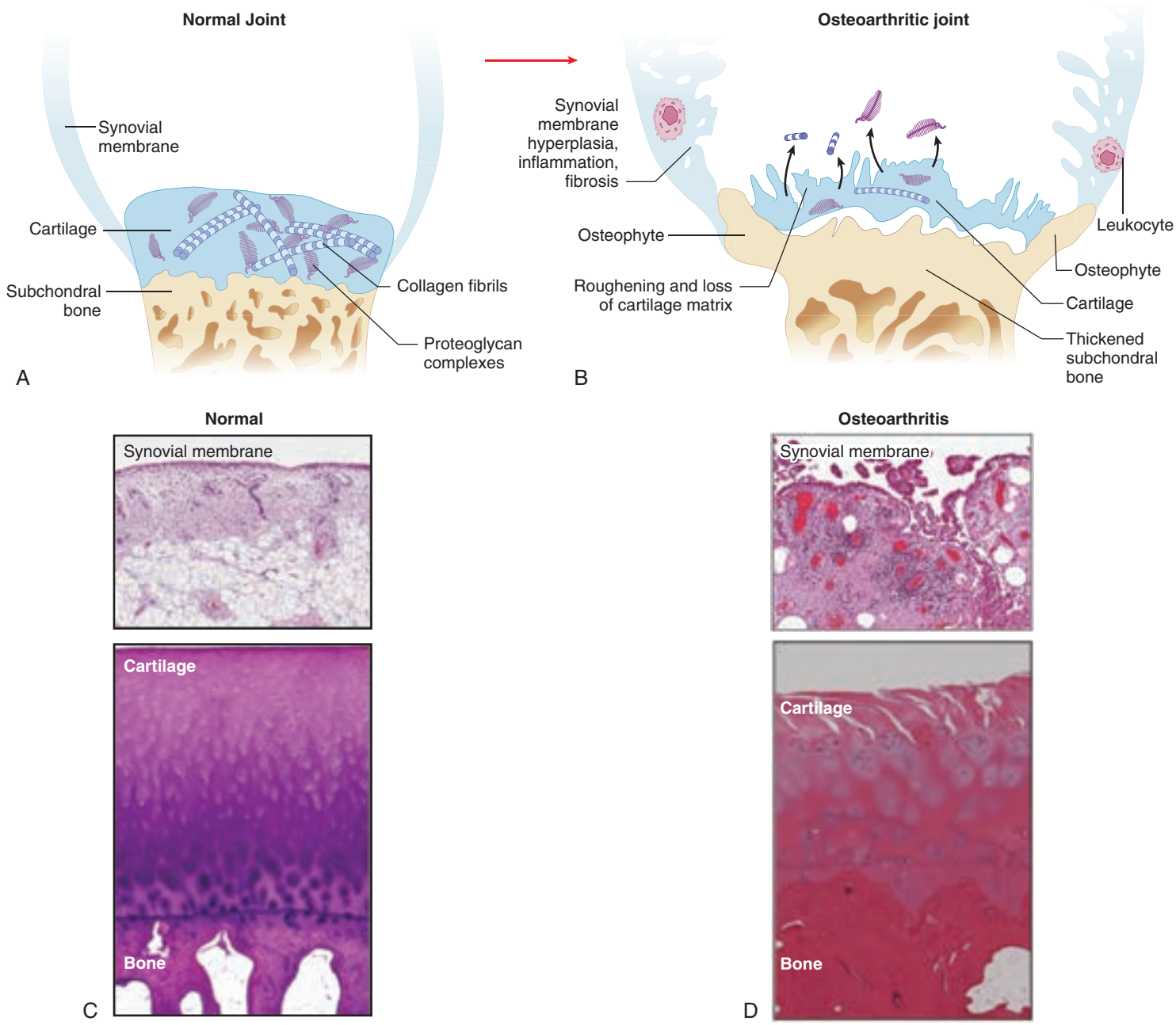

FIGURE 262-1. Pathologic features of osteoarthritic joint tissues. A, Features of a normal adult synovial joint. Healthy adult articular cartilage is characterized by a smooth surface and extracellular matrix (ECM) composed of a collagen type II fibrillar network and large proteoglycan complexes. The ECM is produced and maintained by the cellular components of cartilage, chondrocytes. The subchondral bone consists of a thin cortical layer and underlying trabecular bone. The synovial membrane lines the joint capsule and attaches at the cartilage-bone interface. In the normal state, it consists of a lining layer 1 or 2 cells thick, with underlying vascularized loose connective tissue. **B,** Typical changes to tissues seen in osteoarthritis (OA). Enzymatic activities (ADAMTS-4,5 and MMP-13 in particular) cleave proteoglycan and collagen components of the ECM, leading to loss of these molecules from the matrix. As the process advances, the articular cartilage thins and fibrillates and eventually fissures down to the underlying bone are seen. Simultaneously, a remodeling response in the bone is observed. Thickening of the cortical subchondral bone layer occurs, and new bone growth at the margins appears as osteophytes. The synovial membrane changes observed in OA patients include lining layer hyperplasia, inflammation in the form of leukocyte infiltration, and fibrosis which can be seen to varying degrees. Photomicrographs of human joint tissues showing these features are depicted in **C** (normal tissues) and **D** (OA tissues). (C and D courtesy of Edward F. DiCarlo, MD. Hospital for Special Surgery, New York, NY).

for loss of aggrecan, but other enzymes participate (see Fig. 262-1B and D). Concomitantly, the activated chondrocytes proliferate to form clonal clusters and produce inflammatory mediators, including interleukin-1α and β, (IL-1α and IL-1β), IL-6, tumor necrosis factor-α (TNF-α) and nitric oxide (NO), which accelerate the degradative cycle and stimulate chondrocyte apoptosis. Thus, both cellular *and* molecular components of cartilage are lost as the process progresses.[5]

Bone

The cortical bone underlying articular cartilage (subchondral bone) supports load-bearing and transmits mechanical signals to articular chondrocytes. In OA, there is increased remodeling, likely in response to abnormal biomechanical loading. This may result in thinning (attrition) and reduced bone density, leading to subchondral cyst formation in early disease, but progresses to subchondral sclerosis as bone formation outpaces resorption (see Fig. 262-1B and D).[6] Remodeling at joint margins and entheses results in osteophytes (bone spurs). An important role for the growth factors transforming growth factor-β (TGF-β) and bone morphogenetic protein-2 (BMP-2) in driving osteophyte formation has been demonstrated in animal models.

Synovium

Synovial involvement in OA is more variable than in rheumatoid arthritis, but low-grade synovitis, characterized by infiltration of macrophages and lymphocytes, increased vascularity, synovial lining hyperplasia, and fibrosis,

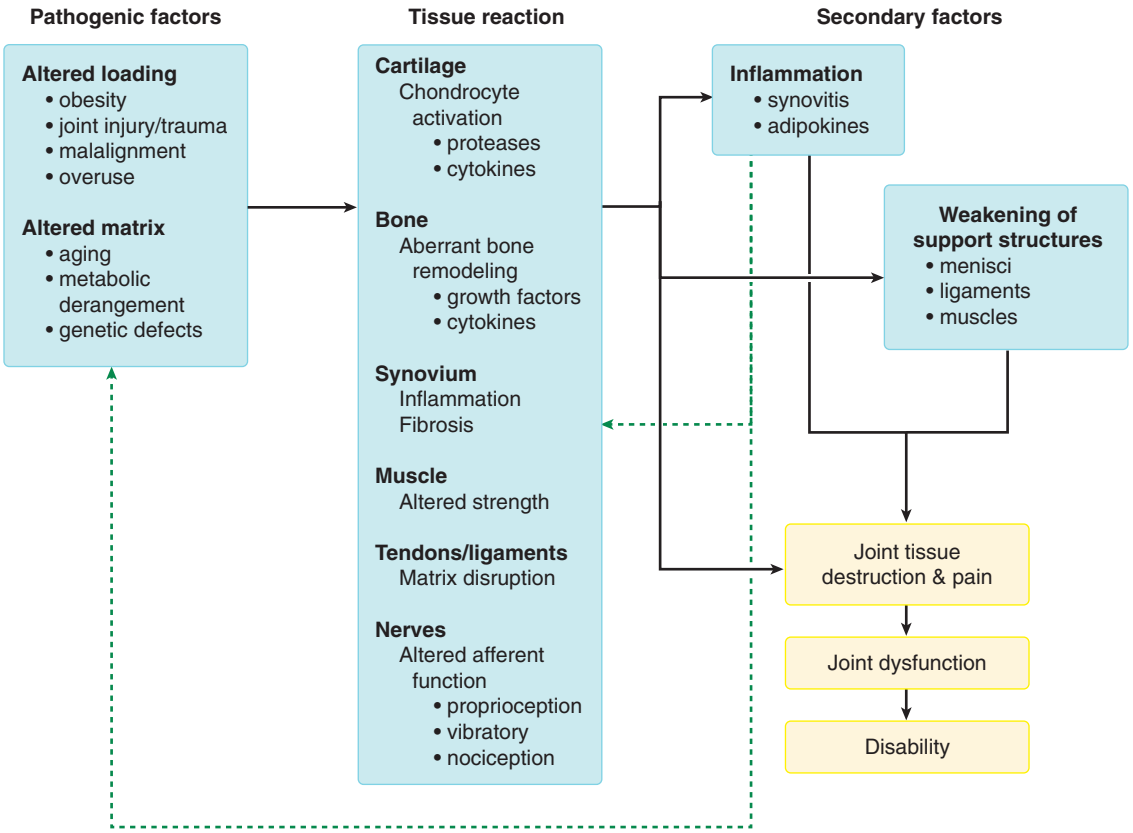

FIGURE 262-2. Schematic of the pathophysiology of osteoarthritis. A variety of stimuli can alter biomechanical loading patterns of joint tissues or exert biologic pressure on the extracellular matrix. These result in specific responses in the various joint tissues. Aided by secondary factors, these lead to tissue degeneration, pain, and ultimately joint dysfunction.

is common and occurs in up to 75% of patients (see Fig. 262-1B and D). Although this synovial reaction is likely a secondary response to molecular breakdown products present in the OA joint, it results in release of cytokines (e.g., IL-6, TNF-α), chemokines (IL-8, monocyte chemoattractant protein 1 [MCP-1], C-C motif chemokine 19 [CCL19]), enzymatic mediators of cartilage catabolism, and growth factors involved in bone remodeling, and when present it is associated with more severe symptoms and possibly more rapid progression of cartilage degradation.[7]

Menisci and Ligaments

Traumatic and athletic injuries to the menisci and ligaments are known risk factors for incident OA. However, meniscal and ligamentous abnormalities are common in patients with OA even without injury and are detectable in 80% and 30% (respectively) of patients undergoing magnetic resonance imaging (MRI). These structures are sensitive to the same inflammatory and enzymatic mediators that promote cartilage deterioration in OA, and their compromise promotes degenerative injuries, accelerated cartilage erosion, joint instability, and mechanical symptoms (i.e., locking or catching).

Muscles

Periarticular muscle weakness and atrophy are characteristic of hip and knee OA. Such weakness is associated with altered gait kinematics and pathologic joint loading, but it remains unclear whether this is a cause or an effect of OA.

Peripheral and Central Nervous System

Somatosensory deficits associated with neuromuscular function, including proprioceptive and widespread vibratory deficits, have been described.[8] Also, patients with OA exhibit signs of peripheral and CNS sensitization in the form of hyperalgesia and allodynia. Inflammatory mediators (e.g., MCP-1) and growth factors (e.g., nerve growth factor [NGF]) have been implicated in pain pathways in OA.[9] Additional work is necessary to fully understand the mechanisms leading to symptomatic OA, but pain in OA is clearly multifactorial, involving the peripheral and the central nervous systems in addition to joint tissues and inflammation.

Pathogenic Factors (Fig. 262-2)

Biomechanics

Aberrant loading of joints mediates evolution of structural joint degeneration and may contribute to initiation of the process. This is evident in the high risk that malalignment of the knee carries for knee OA progression. Weight-bearing regions of OA joints bear greater loads during ambulation than normal, and these loading patterns have been shown to be predictive of subsequent progression of lower extremity joint disease longitudinally. Mechanistically, abnormal biomechanical loading promotes pathologic activation of chondrocytes and bone remodeling, and may be precipitated by joint injury, congenital dysplasias, malalignment, joint instability (e.g., in the setting of aging or heritable connective tissue diseases), and abnormal neuromuscular control (e.g., neuropathic diseases), each of which is a risk factor for OA.

Metabolic Factors

Several metabolic conditions, including alkaptonuria, acromegaly, hemochromotosis, and the metabolic syndrome, among others, predispose individuals to early OA, though they have different mechanisms and disease patterns. In the metabolic syndrome, obesity increases loads on weight-bearing joints, but its association with OA of non–weight-bearing joints in the hands suggests a nonmechanical etiology as well. A possible mechanism may be related to the overproduction of adipokines, IL-6, and leptin leading to chronic low-grade systemic inflammation that may potentiate molecular deterioration of joint tissues and contribute to joint symptoms.[10]

Genetic Factors

Rare single-gene defects may lead to early or aggressive OA. Heritable defects in the collagen II gene have occasionally been described in families with chondrodysplasias, as have lubricin defects. However, variations in multiple genes are likely implicated in the majority of OA cases. Genome-wide association studies have identified polymorphisms in genes involved in proliferation, skeletal development (e.g., GDF5), and regulation of body weight associated with risk for radiographic knee and hip OA.[11] Additional polymorphisms have been associated with pain in OA (e.g., *PACE4* and *TRPV1*).

Aging

Aging-related molecular and cellular changes likely contribute to OA pathogenesis.[12] Modifications of extracellular matrix components, such as accumulation of advanced-glycation end products, and carboxylation associated with oxidative-stress, occur with advancing age. These can alter protein folding, weaken tissues, and increase susceptibility to proteolytic cleavage. Age-related changes to chondrocytes and other cells include the "senescence-associated secretory phenotype" associated with decreased proliferative capacity but increased secretory activity that may promote abnormal chondrocyte responses to injury and aberrant loading. Nonetheless, the link between aging-related changes and symptoms remains poorly understood.

CLINICAL MANIFESTATIONS OF OSTEOARTHRITIS

Symptoms

Pain is the most prominent symptom of OA. Although often limited to affected joints, it can become widespread over time.[13] The quality of pain experienced is variable, ranging from "aching" joint pain to less localized periarticular or radiating pain. OA pain is typically worsened by joint use while stiffness is exacerbated by prolonged inactivity. Morning stiffness may occur but is brief, lasting less than 30 minutes. Other symptoms include joint instability, limitation of motion, locking, and a grinding feeling with motion. Symptom severity varies over time, but with advanced disease, pain becomes persistent and can disturb sleep.

Patterns of Joint Involvement

The joint pattern may help distinguish OA from other forms of arthritis (Fig. 262-3).

Lower Extremities

The large weight-bearing joints (knees and hips) are most commonly affected. Knee OA encompasses any of the three compartments. The medial compartment is involved in the majority of patients and may lead to varus (bow-legged) deformity. Lateral compartment OA may result in valgus (knock-kneed) deformity, and patellofemoral OA typically causes pain exacerbated by descending stairs. Meniscal or ligamentous degeneration often accompanies OA and exacerbates knee instability. Effusions, synovial (Baker's) cysts, and anserine or prepatellar bursitis may be present and cause additional pain. Hip OA often begins with restricted internal rotation and progresses to limited motion in all directions and limb length discrepancy. Pain is felt in the groin area and may radiate to the anterior thigh and knee and be confused with knee pain. It also must be differentiated from lateral thigh pain, which more likely originates from other structures (e.g., trochanteric bursa, iliotibial band, lumbar spine). In the foot, OA typically involves the first metatarsophalangeal joint, resulting in bunion deformity.

Spine

Spinal OA typically involves the lumbar and cervical regions, affecting the apophyseal (facet) and uncovertebral (joints of Luschka) joints. Low-grade inflammation and bone remodeling can lead to local pain, and nerve root compression by osteophytes causes radicular, radiating pain. Degenerative disc disease often coexists with spinal OA, and together they contribute to spinal stenosis causing muscle weakness, paresthesias, and numbness. In severe cases, cord impingement with myelopathy may result.

Upper Extremities

The small joints of the hands are most commonly affected, specifically the distal and proximal interphalangeal joints (DIPs and PIPs) and first carpometacarpal (CMC) joints. Osteophytosis of DIPs and PIPs leads to bony, palpable Heberden and Bouchard nodes, respectively. This pattern is more common in white women and is termed *primary generalized OA*. Patients may experience difficulty grasping, opening jars, buttoning clothes, and turning doorknobs. Erosive or inflammatory OA is a less common but distinct subset in which erosions develop in the DIPs and PIPs, and the patient experiences repeated episodes of acute inflammatory symptoms.

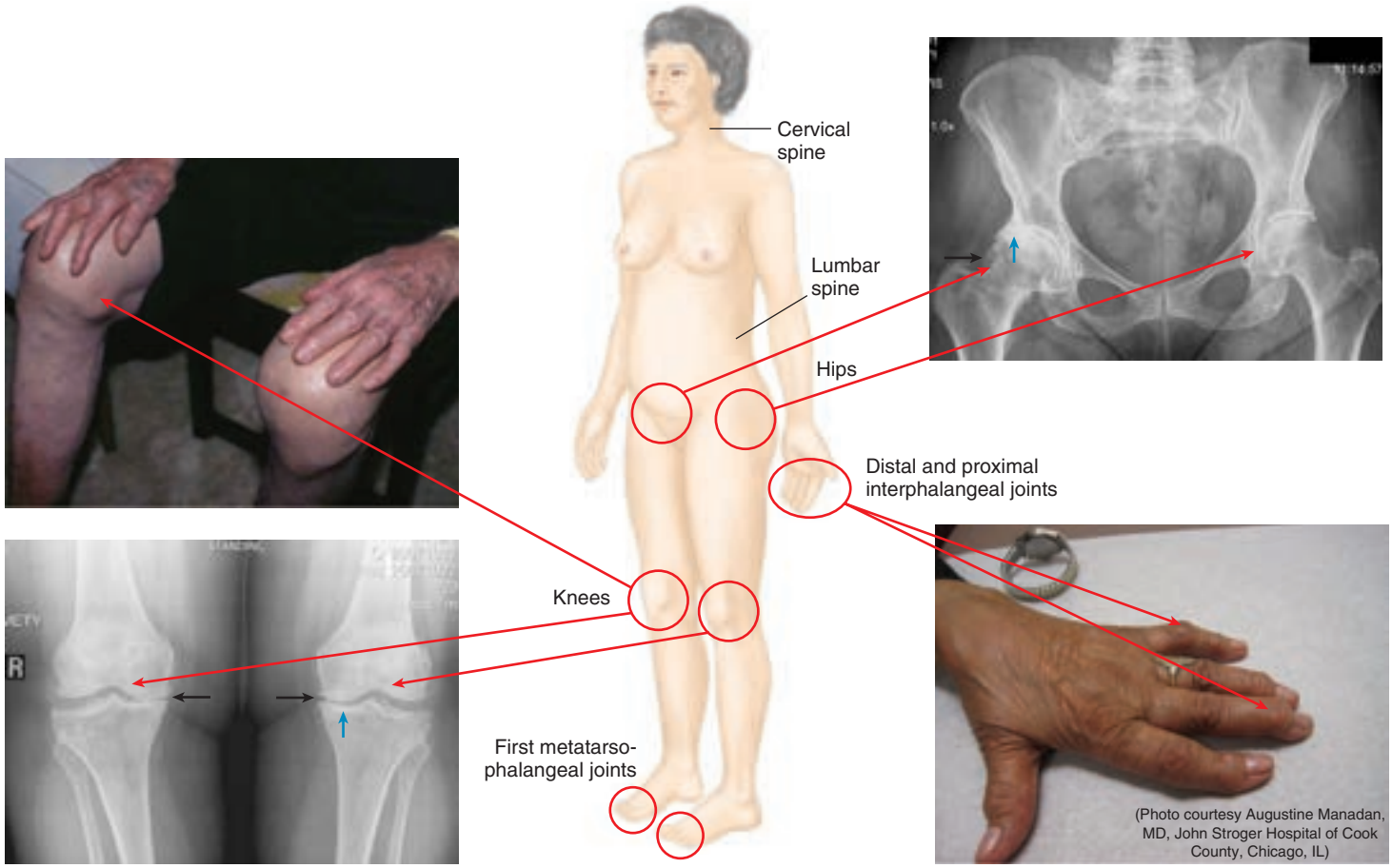

FIGURE 262-3. Commonly affected joints in osteoarthritis. Joints in which symptomatic osteoarthritis typically develops are depicted in the whole-body drawing in the center. The distal and proximal interphalangeal joints of the hands characteristically exhibit palpable bony bumps (*right lower panel*). Synovial effusions and bony outgrowths, or osteophytes, may be visualized in the knee (*left upper panel*), and radiographic involvement of the knee (*left lower panel*) or of the hip (*right upper panel*) shows characteristic features of joint space narrowing, osteophytes (*black arrows*), and subchondral sclerosis (*blue arrows*).

DIAGNOSIS

The diagnosis of OA depends primarily on the clinical presentation; unlike other rheumatic diseases, imaging and laboratory analysis have relatively small roles in diagnosis. Because OA primarily affects tissues of the joint and does not involve systemic inflammation, laboratory tests are more useful for excluding competing diagnoses than for establishing an OA diagnosis.

Physical Examination

When cartilage surfaces become roughened, crepitus (palpable or audible crackling) may be detected by physical examination. As cartilage is further compromised, small fragments can dislodge and, if sufficiently large, may restrict joint motion and cause locking.

Osteophytes can be palpated in superficial joints as bony projections and may result in deformities. Their growth is sometimes inflammatory, with erythema, tenderness, and swelling. Heberden and Bouchard nodes of the DIP and PIP joints, respectively, are typical, but squaring of the first CMC joint and palpable knee osteophytes are also common. These deformities may eventually restrict joint range of motion. Classically, severe joint inflammation is not appreciated; however, effusions are common and may cause mild joint warmth.

Laboratory Evaluation

Blood Tests

OA is limited to joints; thus, tests of systemic inflammation and those that assess critical organ function are generally normal. These include the sedimentation rate, conventional C-reactive protein, and other acute phase reactants, as well as the blood count and the comprehensive metabolic panel. Nonetheless, OA is prevalent among the elderly and concomitant diseases may confound the interpretation of laboratory testing. Tests for autoantibodies are unremarkable, though low-titer detection of nonspecific rheumatoid factors and antinuclear antibody (ANA) may be seen in the normal aging population.

Synovial Fluid

When sampled, synovial fluid total leukocyte counts are typically less than 1500 to 2000 cells/mm^3 with a predominance of lymphocytes rather than neutrophils. Incidental findings of crystals, such as calcium pyrophosphate dihydrate, or of cartilage fragments may sometimes be observed.

Biomarkers

There has been an aggressive search for macromolecules measurable in blood, synovial fluid, or urine that provide prognostic or diagnostic value in OA. A variety of macromolecules, typically breakdown or cleavage products of cartilage or bone matrix, have been identified that have a statistical correlation with OA progression or pain. The OA Initiative is an ongoing longitudinal study of several thousand individuals, spearheaded by the National Institutes of Health, to identify novel body fluid and imaging biomarkers that may provide predictive information about OA onset and progression. But no biomarkers have yet been demonstrated to be useful for clinical evaluation.

Imaging

Radiography

Conventional radiography is the mainstay for imaging OA. Characteristic features include narrowed joint space, osteophytes, and subchondral bone sclerosis (see Fig. 262-3). Not all patients have all three features, and there is an imperfect relationship between radiographic appearance and clinical symptoms. Disease progression can be monitored by longitudinal imaging, both by qualitative grading and by quantitative assessment of joint space narrowing.

Magnetic Resonance Imaging

MRI provides the most detailed images of joint structures and, in contrast to radiography, detects subtle defects of articular cartilage.[14] In addition, MRI detects subchondral bone marrow lesions that are associated with symptomatic disease. MRI remains predominantly a research tool in OA, where it has been helpful in identifying soft tissue pathologic conditions and measuring early cartilage pathology; clinically, it has little role in routine OA evaluation and management. MRI is often used in OA to exclude other potential sources of pain, such as degenerative menisci, ligamentous tears, and other intraarticular pathology.

Ultrasonography

Although operator dependent, careful ultrasonography reveals outstanding details of joint structure and does not expose the patient to ionizing radiation. It may identify osteophytes undetectable by standard radiography, and articular cartilage lesions can often be observed. Its primary use, though, may be to detect local synovitis that is prevalent in OA; it remains unclear whether it can play a significant role in diagnosis.

Radionuclide Imaging

Whereas radioisotope scanning formerly had utility in OA, it has been largely supplanted by more sensitive and specific modalities for detecting articular pathology.

TREATMENT Rx

Many physicians persist in counseling their symptomatic patients that there is nothing to be done, and that OA is an inevitable sign of "getting old." Although no interventions have been demonstrated to alter the natural history of structural joint degeneration, symptomatic OA is not inevitable, and many strategies provide relief and maintain function among those with even advanced OA.[15] Patient education and support are critical; self-help programs alone have been shown to improve outcomes in OA. Patients should be provided thorough information about the natural course of OA, their role in disease management, and appropriate expectations. In general, the goals of OA therapy are similar to the treatment of any disease—to maintain or restore function, relieve symptoms, and prevent disease progression.

Maintenance of Function

Strategies to retain function and independence in patients with OA include ambulatory assistive devices, such as canes and walkers, which provide stability in addition to reducing loading across arthritic joints. Motorized carts can assist individuals with severe knee or hip OA to retain independence in the community. Physical therapy can help to retain strength and range of motion, and occupational therapy can provide customized assist devices and braces. Whenever possible, counseling on weight loss for overweight or obese patients should be provided as weight loss can improve both pain and function.

Symptom Palliation

A variety of strategies can provide effective pain palliation in OA. These include physical measures, medical therapy, and surgical interventions.[16]

Physical Measures

Exercise is especially important in OA, both to improve function and to palliate pain.[17] Exercises aimed at strengthening muscles surrounding affected joints are of value physiologically and have been consistently demonstrated to provide significant pain relief. Patients should be encouraged to exercise regularly and may benefit from physical therapy for instruction in appropriate strength training and to improve and maintain range of motion. During painful flares and immediately after exertion, application of heat or ice to affected joints can be useful.

Pain may be mediated by aberrant biomechanical loading. Therefore, canes and walkers, which significantly reduce loads across the knee during gait, can reduce pain and improve stability. Similarly, unloading knee braces, when tolerated, may provide palliation for knee OA.

Medication

Most patients will need more pain relief than is provided by physical measures. Among the elderly, choice of medication is often influenced by comorbidities. Topical agents may reduce risk for of systemic adverse effects and are appropriate when only a few joints are symptomatic. Topical capsaicin has been approved for knee OA; its use requires frequent application and careful handwashing after contact. Topical nonsteroidal anti-inflammatory drugs (NSAIDs) are available, including salicylic acid and diclofenac.

When physical measures and topical agents are insufficient, oral analgesics are used. Many specialty organizations, including the American College of Rheumatology,[18] suggest that acetaminophen may be beneficial, especially among patients with contraindications to NSAIDs such as renal dysfunction or cardiac disease. Nonetheless, acetaminophen may not be effective for long-term analgesia and chronic use carries its own potential toxicities, including liver damage and hypertension. Therefore, acetaminophen may be most properly used for short-term flares of OA pain, typically lasting no more than a few weeks. Approved alternatives to acetaminophen include tramadol and opiates, which though demonstrated to relieve OA pain, also substantially increase morbidity among the elderly, especially the risk for traumatic falls. Finally, neuroactive agents are widely used for OA pain. One such medication, duloxetine hydrochloride, a serotonin and norepinephrine reuptake inhibitor (SNRI), received U.S. Food and Drug Administration approval in 2010 for the treatment of musculoskeletal pain, including OA, based on positive results in clinical trials.

NSAIDs have been demonstrated to be effective for OA pain and may maintain efficacy for years.[A1] These medications remain the mainstay of OA medical therapy, but are not an option for many patients with renal, cardiac, or gastrointestinal conditions. Proton pump inhibitors or misoprostol can provide gastric protection in middle-aged and elderly patients and those at risk for gastrointestinal bleeding. Cyclooxygenase-2 inhibitors also may be used; in the United States, celecoxib is the only representative of this class available. Six weeks of low-dose oral prednisolone (7.5 mg daily) is also effective, and its benefits can persist for at least another six weeks.[A4]

Intra-articular Therapy

Intra-articular glucocorticoids may provide short-term relief of OA pain, often lasting months.[A5] Their use is limited to three or four times per year in any single joint because of theoretical concerns of toxicity to articular cartilage. A variety of hyaluronan derivatives are available for injection to relieve OA pain, but controversy exists as to whether they are more effective than placebo. Originally developed to supplement viscosity of synovial fluid in an attempt to improve articular lubrication, residence time in the joint is too brief to have this effect.

Surgical Approaches

Joint replacement surgery restores function and relives pain in the majority of patients and is the most important therapeutic advance in OA treatment to date (Chapter 276). It should be reserved for those in whom pain or joint dysfunction significantly limits normal life activities despite optimal medical and physical management. The presence of advanced structural degeneration of the joints alone, without severe symptoms, should not be an indication for arthroplasty. Knees and hips are most frequently replaced, but good results are now obtained in other joints as well. The durability of joint prostheses is limited, so joint replacement surgery should be delayed in younger patients when practical. Aside from total joint replacement, there are a variety of temporizing strategies that may be used in joints that have less severe structural degeneration, including realignment osteotomy in the knee and hemiarthroplasty.

Delay of Disease Progression

There has been extensive effort to identify disease-modifying OA drugs (DMOADS) that would retard disease progression and affect OA morbidity. To date, no true DMOADs have been identified, although investigation continues into agents targeting cartilage and joint tissue metabolism and inflammation. There remains optimism that as the mechanism of joint degeneration becomes more fully elucidated, rational drug discovery may identify effective DMOADs. In addition, tissue engineering approaches and mesenchymal stem cell technology may permit the development of functional joint tissue replacement in the future. Current cartilage replacement techniques are not indicated for OA treatment, but are restricted to patients with isolated chondral defects.

Biomechanically Active Approaches

OA progression is at least partly mediated by aberrant loading of joints, so improving loading patterns should affect structural progression. At present, no approaches to alter loading have yet been shown to substantially affect OA progression, but specialized footwear, gait modifications, and mechanically derived exercise regimens are under active investigation.

Complementary and Alternative Approaches

Similar to what occurs with other chronic pain conditions, the overwhelming majority of patients with OA try complementary approaches (Chapter 39). Among the more popular are glucosamine, chondroitin, and acupuncture. None has been clearly demonstrated to substantially retard joint degeneration and independently funded trials have been negative, but many patients feel pain improvement. In any blinded study of OA pain, a substantial placebo response is typically observed. Many complementary approaches have been systematically studied, and controversy remains regarding the incremental pain relief provided by these modalities over that obtained with placebo. Regardless, many of these approaches can be safely used by individual patients who derive relief from them.

PROGNOSIS

Once structural joint damage is present, it is likely to progress, although at variable rates. Slowly progressive structural disease in the absence of severe symptoms may never require surgical intervention, whereas rapidly progressive symptomatic disease might prompt early intervention. The causes of this variability remain unclear, but several factors may contribute to an individual's prognosis. Female gender, obesity, and pain severity are associated with both the incidence and progression of radiographic knee OA, and joint malalignment (varus or valgus) is a strong predictor of progression. Symptomatically, MRI or sonographically detected synovial thickening or effusion may predict progressive cartilage defects and pain severity. In the hip, joint shape is predictive of radiographic progression, as is meniscal pathology (tears, extrusion, and maceration). MRI-identified subchondral bone marrow lesions are associated with pain severity and progressive cartilage loss.

PREVENTION

No current treatments substantially alter OA progression, but some strategies reduce the risk for development of OA and ameliorate symptomatic progression. Obesity may be the most modifiable of the strong OA risk factors. Weight loss in adulthood reduces the risk for incident radiographic and symptomatic OA and reduces pain severity in patients who already have OA.[A1] Exercise is an important component of weight strategies and ameliorates pain, but specific types of exercise have not yet shown consistent preventive effects. Strategies to decrease joint injuries in young athletes are critical to reducing post-traumatic OA,[19] and proper conditioning has been shown to reduce knee injuries among female soccer players.[A7] New insights into the biology and biomechanics of OA in the coming years may be expected to yield novel strategies to prevent the onset and progression of the disease.

Grade A References

A1. Bliddal H, Leeds AR, Stigsgaard L, et al. Weight loss as treatment for knee osteoarthritis symptoms in obese patients: 1-year results from a randomised controlled trial. *Ann Rheum Dis*. 2011;70:1798-1803.

A2. Chappell AS, Ossanna MJ, Liu-Seifert H, et al. Duloxetine, a centrally acting analgesic, in the treatment of patients with osteoarthritis knee pain: a 13-week, randomized, placebo-controlled trial. *Pain*. 2009;146:253-260.

A3. Chou R, McDonagh MS, Nakamoto E, et al. Analgesics for osteoarthritis: an update of the 2006 Comparative Effectiveness Review. Agency for Healthcare Research and Quality (US); 2011. Report No.: 11(12)-EHC076-EF.

A4. Abou-Raya A, Abou-Raya S, Khadrawi T, et al. Effect of low-dose oral prednisolone on symptoms and systemic inflammation in older adults with moderate to severe knee osteoarthritis: a randomized placebo-controlled trial. *J Rheumatol*. 2014;41:53-59.

A5. Yavuz U, Sokucu S, Albayrak A, et al. Efficacy comparisons of the intraarticular steroidal agents in the patients with knee osteoarthritis. *Rheumatol Int*. 2012;32:3391-3396.

A6. Messier SP, Mihalko SL, Legault C, et al. Effects of intensive diet and exercise on knee joint loads, inflammation, and clinical outcomes among overweight and obese adults with knee osteoarthritis: the IDEA randomized clinical trial. *JAMA*. 2013;310:1263-1273.

A7. Steffen K, Emery CA, Romiti M, et al. High adherence to a neuromuscular injury prevention programme (FIFA 11+) improves functional balance and reduces injury risk in Canadian youth female football players: a cluster randomised trial. *Br J Sports Med*. 2013;47:794-802.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Neogi T, Zhang Y. Epidemiology of osteoarthritis. *Rheum Dis Clin North Am*. 2013;39:1-19.
2. Chapple CM, Nicholson H, Baxter GD, et al. Patient characteristics that predict progression of knee osteoarthritis: a systematic review of prognostic studies. *Arthritis Care Res*. 2011;63:1115-1125.
3. Loeser RF, Goldring SR, Scanzello CR, et al. Osteoarthritis: a disease of the joint as an organ. *Arthritis Rheum*. 2012;64:1697-1707.
4. Xia B, Di C, Zhang J, et al. Osteoarthritis pathogenesis: a review of molecular mechanisms. *Calcif Tissue Int*. 2014;95:495-505.
5. Liu-Bryan R, Terkeltaub R. Emerging regulators of the inflammatory process in osteoarthritis. *Nat Rev Rheumatol*. 2015;11:35-44.
6. Burr DB, Gallant MA. Bone remodelling in osteoarthritis. *Nat Rev Rheumatol*. 2012;8:665-673.
7. Scanzello CR, Goldring SR. The role of synovitis in osteoarthritis pathogenesis. *Bone*. 2012;51: 249-257.
8. Roos EM, Herzog W, Block JA, et al. Muscle weakness, afferent sensory dysfunction and exercise in knee osteoarthritis. *Nat Rev Rheumatol*. 2011;7:57-63.
9. Malfait AM, Schnitzer TJ. Towards a mechanism-based approach to pain management in osteoarthritis. *Nat Rev Rheumatol*. 2013;9:654-664.
10. Zhuo Q, Yang W, Chen J, et al. Metabolic syndrome meets osteoarthritis. *Nat Rev Rheumatol*. 2012;8:729-737.
11. Reynard LN, Loughlin J. The genetics and functional analysis of primary osteoarthritis susceptibility. *Expert Rev Mol Med*. 2013;15:e2.
12. Loeser RF. Aging processes and the development of osteoarthritis. *Curr Opin Rheumatol*. 2013; 25:108-113.
13. Hunter DJ, Guermazi A, Roemer F, et al. Structural correlates of pain in joints with osteoarthritis. *Osteoarthritis Cartilage*. 2013;21:1170-1178.
14. Braun HJ, Gold GE. Diagnosis of osteoarthritis: imaging. [Review]. *Bone*. 2012;51:278-288.
15. Osteoarthritis: Care and Management in Adults. NICE Clinical Guidelines, No. 177. London: National Clinical Guideline Centre; 2014.
16. McAlindon TE, Bannuru RR, Sullivan MC, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthritis Cartilage*. 2014;22:363-388.
17. Golightly YM, Allen KD, Caine DJ. A comprehensive review of the effectiveness of different exercise programs for patients with osteoarthritis. *Phys Sportsmed*. 2012;40:52-65.
18. Hochberg MC, Altman RD, April KT, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care Res (Hoboken)*. 2012;64:465-474.
19. Bennell K, Hunter DJ, Vicenzino B. Long-term effects of sport: preventing and managing OA in the athlete. *Nat Rev Rheumatol*. 2012;8:747-752.

263

BURSITIS, TENDINITIS, AND OTHER PERIARTICULAR DISORDERS AND SPORTS MEDICINE

JOSEPH J. BIUNDO

DEFINITION

An array of painful and sometimes disabling musculoskeletal syndromes exist that are not articular in origin but arise from tendons and bursae. These conditions are referred to by various names, in addition to *tendinitis* and *bursitis,* including the terms *nonarticular rheumatism, soft tissue diseases, regional rheumatic pain syndromes, overuse syndromes*, and *repetitive use syndromes* (Tables 263-1 and 263-2). These entities are often ignored, misdiagnosed as arthritis, or attributed to the aging process; awareness of the existence of these conditions and knowledge of basic musculoskeletal anatomy (Figs. 263-1 and 263-2) are the fundamental requirements for diagnosis. This knowledge is coupled with brief but specific physical diagnosis techniques. The accurate diagnosis and successful treatment of these conditions is gratifying to the clinician because many people can be relieved of their chronic painful syndromes.

Various terms regarding tendon injuries are used and may be confusing. The main term used is *tendinitis. Tendinosis* has been proposed as the correct terminology because there are degenerative changes in the tendon but very few inflammatory cells. In addition, fatty mucoid degeneration and hyaline

TABLE 263-1 MUSCULOSKELETAL CONDITIONS BY ETIOLOGY

TENDINITIS	TENDON RUPTURE	BURSITIS
Rotator cuff	Rotator cuff	Subacromial
Bicipital	Bicipital	Olecranon
Volar flexor	Quadriceps	Trochanteric
de Quervain	Patellar	Ischial
Patellar	Posterior tibialis	Iliopsoas
Posterior tibialis	Achilles	Pes anserine
Achilles		Prepatellar
Epicondylitis		Retrocalcaneal

TABLE 263-2 TENDINITIS AND BURSITIS CONDITIONS BY REGION

SHOULDER
Rotator cuff tendinitis
Rotator cuff tear
Bicipital tendinitis
Subacromial bursitis
Adhesive capsulitis
ELBOW
Olecranon bursitis
Medial epicondylitis
Lateral epicondylitis
WRIST AND HAND
de Quervain's tenosynovitis
Volar flexor tenosynovitis
Ganglion
HIP
Trochanteric bursitis
Iliopsoas bursitis
Ischial bursitis
Coccydynia
KNEE
Prepatellar bursitis
Pes anserine bursitis
Popliteal cyst (Baker's cyst)
Patellar tendinitis
Patellar/quadriceps tendon tear
ANKLE AND FOOT
Achilles tendinitis
Achilles tendon tear
Posterior tibial tendinitis
Posterior tibial tendon tear
Retrocalcaneal bursitis
Plantar fasciitis

features occur in these tendon syndromes. These tendon conditions are described by some as a *tendinopathy* because use of this term avoids the need to decide whether inflammation is a factor. Also, tendons may *rupture* or *tear,* partially or completely. The term *tendon insufficiency* is used when the tendon is stretched or is partially or even completely torn. The terms *tenosynovitis* and *peritendinitis* refer to an inflammatory response of the tenosynovium or peritendon, respectively.

Tendon syndromes are basically "overuse" injuries. Tendinitis may occur when the tendon repeatedly bears more load than it can withstand. This may result from excessively high loads across normal tendons or from normal loads across degenerated tendons. In addition to load and repetitiveness, tendon changes resulting from immobility and from aging may play a role, as may the use of certain medications such as fluoroquinolones and corticosteroids.[1]

Bursae are closed sacs lined by a synovial membrane and serve as a cushion. They are located between tendon and bone, tendon and tendon, or bone and skin and allow smooth gliding between these structures. A bursa, which normally has a small amount of bursal fluid, can become inflamed from trauma or overuse, or become infected, producing a bursitis. When this occurs, some swelling and pain of the bursa may be present.

EPIDEMIOLOGY

The incidence of the nonarticular syndromes of bursitis and tendinitis is high. They are more common than both rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE). For example, the incidence of shoulder pain, largely a result of rotator cuff tendinitis and rotator cuff tear, was approximately 20% in a population older than 70 years of age.

DIAGNOSIS

A precise history is needed to identify the conditions present, and more than one syndrome can occur concomitantly. A working knowledge of regional anatomy and an approach that uses a regional differential diagnosis will help in obtaining a specific diagnosis. A complete neuromusculoskeletal examination should be performed, emphasizing careful palpation, passive range of motion (ROM), and active ROM alone or sometimes with resistance. Systemic and infectious causes must be considered. Diagnostic ultrasonography and magnetic resonance imaging (MRI) are sometimes useful in confirming a diagnosis.

TREATMENT

Rx

Treatment of tendinitis and bursitis includes use of nonsteroidal anti-inflammatory drugs (NSAIDs), relative rest of the injured site, stretching and strengthening exercises, friction massage, use of modalities (heat, ice, and ultrasound), splinting, corticosteroid injections,[2] and surgery. A comprehensive management of these regional syndromes should be undertaken, rather than relying on oral medications alone. The causative aspects should be evaluated, and activity modification should be advised as needed. The goals of therapeutic exercise are to increase flexibility by stretching, increase muscle strength by resistive exercises, and improve muscle endurance by some repetitive regimen. Caution should be exerted in performing corticosteroid injections; the injections should not be placed into the tendon proper, but rather into the peritendinous sheath. The injected solution should be placed beneath the subcutaneous tissue, to avoid skin and subcutaneous fat atrophy, and

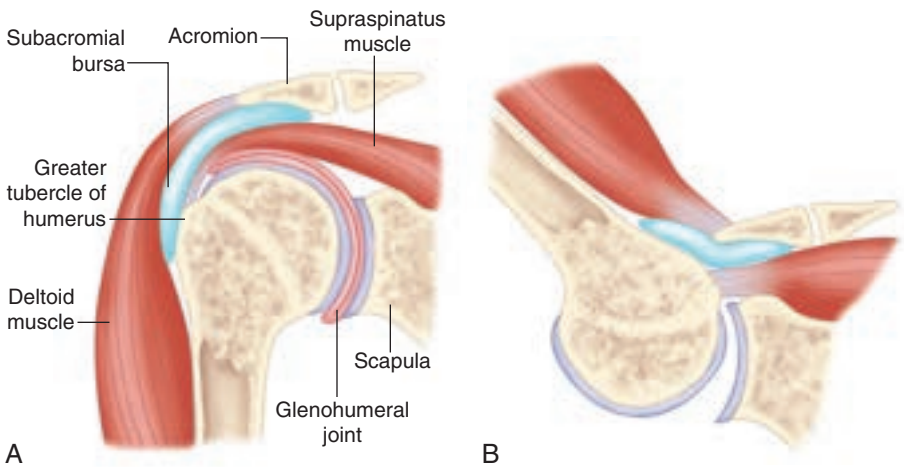

FIGURE 263-1. Relationship of subacromial bursa (shown in *blue*) to supraspinatus muscle and acromion process. A, In the position of adduction of the humerus. To show this bursa more clearly, the synovial membrane of the glenohumeral joint is not shown in *blue*. **B,** In the position of abduction of the humerus, the acromion impinges on the subacromial bursa and the insertion of the supraspinatus tendon. (From Polley HF, Hunder GG, eds. *Rheumatologic Interviewing and Physical Examination of the Joints,* 2nd ed. Philadelphia: WB Saunders; 1978:65.)

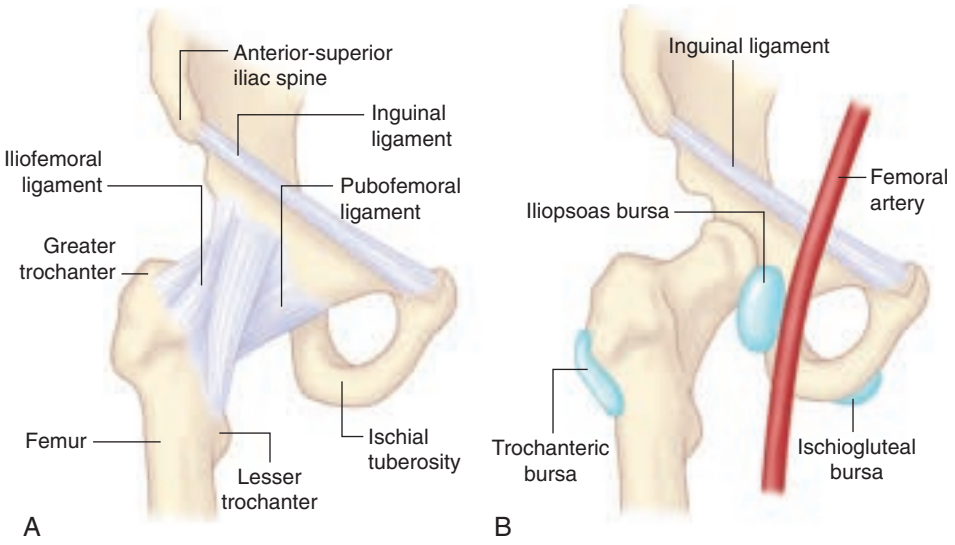

FIGURE 263-2. **Musculoskeletal anatomy of the hip. A, Anterior aspect of the hip joint and bony structures. B, Relationship of the distended iliopsoas, trochanteric, and ischiogluteal bursae (shown in *blue*) to the hip joint and adjacent structures. (From Polley HF, Hunder GG, eds. *Rheumatologic Interviewing and Physical Examination of the Joints,* 2nd ed. Philadelphia: WB Saunders; 1978:183.)**

injections should not be given too frequently, to avoid the possibility of weakening and rupture of the tendon. The accuracy of injections may be improved with concomitant use of diagnostic ultrasonography to assist with determining the correct needle site. Also, fluoroscopically guided injections can be used to increase accuracy.

SPORTS MEDICINE INJURIES

An overlap exists between commonly occurring conditions of tendinitis and bursitis and those attributed to sports injuries (Table 263-3). For example, lateral epicondylitis, frequently referred to as *tennis elbow,* occurs more commonly secondary to non–sports-related causes. In contradistinction, iliotibial band syndrome is usually related to sports. Other entities that occur more often in relation to sports include ligamentous knee injuries, patellar tendinitis, ankle sprains, turf toe, and acromioclavicular separations. It is important in both categories to know the anatomy and biomechanics of the condition, so as to better diagnose and treat the problem. Sports-related injuries are helped with the classic RICE treatment, consisting of *r*est, *i*ce, *c*ompression, and *e*levation. Often, anti-inflammatory and analgesic drugs are used. However, there is less use of corticosteroid injections in athletic injuries than in routine cases of tendinitis and bursitis.

TABLE 263-3 ADDITIONAL SPORTS-RELATED CONDITIONS

SHOULDER
Acromioclavicular separation
Glenoid labial tear (SLAP lesion)
Glenohumeral instability with dislocation
ELBOW
Triceps tendinopathy
Little League elbow (apophysitis)
Distal biceps tendinitis
WRIST AND HAND
Gamekeeper's thumb (skier's thumb)
Mallet finger (baseball finger)
Extensor carpi ulnaris tendinitis
Rupture of flexor digitorum profundus tendon
Injury to triangular fibrocartilage
HIP
Adductor strain (groin pull)
Hip pointer
Hamstring strain
KNEE
Anterior cruciate tear
Posterior cruciate tear
Medial collateral ligament tear/strain
Lateral collateral ligament tear/strain
Popliteal tendinitis
Medial and lateral meniscal tears
Patellar tendinitis
Iliotibial band syndrome
ANKLE AND FOOT
Ankle sprain
Turf toe
Stress fracture

DISORDERS OF THE SHOULDER REGION

Shoulder pain is one of the most common musculoskeletal complaints in people older than 40 years of age. In younger people, athletic injuries are a frequent source of such pain.

Rotator cuff tendinitis, or impingement syndrome, is the most common cause of shoulder pain.[2] Tendinitis (and not bursitis) is the primary cause of pain, but secondary involvement of the subacromial bursa occurs in some cases. The condition may be acute or chronic and may or may not be associated with calcific deposits within the tendon. The key finding is pain in the rotator cuff on active abduction, especially between 60 and 120 degrees, and sometimes when lowering the arm. In more severe cases, pain may begin on initial abduction and continue throughout the ROM. Typically, chronic rotator cuff tendinitis manifests as an ache in the shoulder, usually over the lateral deltoid, and occurs with various movements, especially abduction and internal rotation. Other symptoms include difficulty in dressing oneself and night pain because of difficulty in positioning the shoulders. The physical findings include pain and loss of active abduction and internal rotation, less pain on passive motion, tenderness of the area of supraspinatus insertion, and a positive impingement sign (Fig. 263-3, Neer's sign, which is pain occurring in forced flexion.[3] The causes of rotator cuff tendinitis are multifactorial, but relative overuse, especially from overhead activity causing impingement of the rotator cuff, is commonly implicated. Treatment consists of rest and modalities such as hot packs, ultrasound, or cold applications, with specific ROM exercises as soon as tolerated. NSAIDs are often beneficial, but the most frequent treatment is injection of a depot corticosteroid into the subacromial bursa, the floor of which is contiguous with the rotator cuff. Disodium ethylenediaminetetraacetic acid (EDTA) administered by phonophoresis and mesotherapy to patients with calcific tendinitis of the shoulder has been found to be effective in pain reduction, improvement in shoulder function, and disappearance of calcifications. Extracorporeal shock wave therapy can also be beneficial for chronic calcific tendinitis.

In a rotator cuff tear, an acute tear after trauma is usually easily recognized. The trauma may be superimposed on an already degenerative and possibly even partially torn cuff. In cases of trauma resulting in a ruptured cuff, fracture of the humeral head and dislocation of the joint also should be considered. However, most patients with a tear recall no trauma. In these cases, degeneration of the rotator cuff occurs gradually, resulting ultimately in a complete tear. Rotator cuff tears are classified as small (≤1 cm), medium (1 to 3 cm), large (3 to 5 cm), or massive (>5 cm). Shoulder pain, weakness on abduction, and loss of motion occur in varying degrees, ranging from severe pain and mild weakness to no pain and marked weakness. A positive drop-arm sign with inability to maintain actively 90 degrees of passive shoulder abduction may be present in patients with large or massive tears. Small complete tears

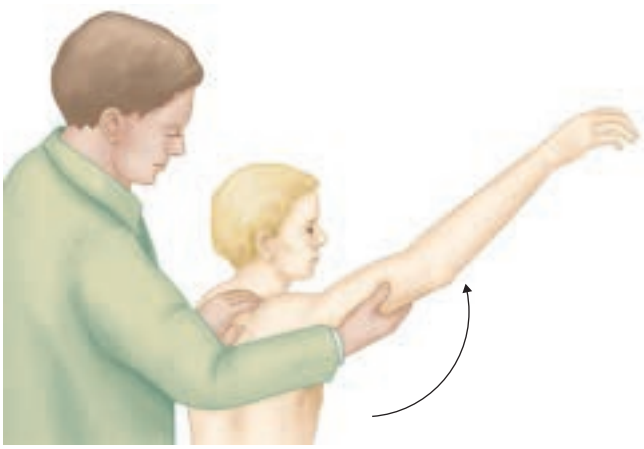

FIGURE 263-3. **The impingement sign is elicited by forced forward elevation of the arm.** Pain results as the greater tuberosity impinges on the acromion. The examiner's hand prevents scapular rotation. This maneuver may be positive in other periarticular disorders. (From Neer CS II. Impingement lesions. *Clin Orthop*. 1983;173:70-77.)

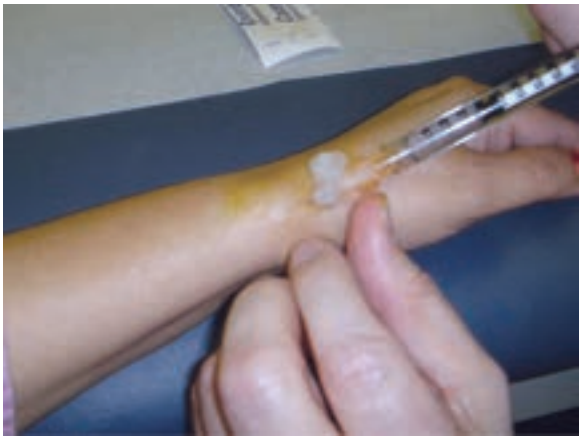

FIGURE 263-4. Injection of de Quervain tenosynovitis.

and incomplete tears of the rotator cuff are treated conservatively with rest, physical therapy, and NSAIDs. Although its role has not yet been established by careful studies, a subacromial injection of a corticosteroid may relieve pain. Surgical repair may be indicated in younger patients.

Bicipital tendinitis is manifested by pain, most often in the anterior region of the shoulder and occasionally more diffusely. The pain may be acute but is usually chronic and is related to impingement of the biceps tendon by the acromion. Tenosynovitis of the long head of the biceps is present, and the tendon may be frayed and fibrotic. Palpation over the bicipital groove reveals localized tenderness. The patient's response should be compared with the response to palpation of the opposite side (i.e., tendon with normal tenderness). Pain may be reproduced over the bicipital tendon in some cases by supination of the forearm against resistance (Yergason's sign), shoulder flexion against resistance (Speed's test), or extension of the shoulder. Treatment of bicipital tendinitis consists of rest, hot packs, ultrasound, and, as pain subsides, passive and then active ROM exercises. NSAIDs may be helpful, and occasionally a small amount of corticosteroid carefully injected into the tendon sheath may be of benefit. Rupture of the biceps tendon can occur at the superior edge of the bicipital groove, producing a characteristic bulbous enlargement of the lateral half of the muscle belly.

Adhesive capsulitis (frozen shoulder) is associated with generalized pain and tenderness and severe loss of active and passive motion in all planes. It is rare before 40 years of age but may occur secondary to any type of shoulder problem. However, not every stiff and painful shoulder is necessarily adhesive capsulitis. Inflammatory arthritis and diabetes can cause adhesive capsulitis. Additional factors such as immobility, low pain threshold, depression, and neglect or improper initial treatment also favor the development of a frozen shoulder. Many cases, however, are idiopathic. The joint capsule adheres to the anatomic neck, and the axillary fold binds to itself, causing restricted motion. The capsule becomes thickened and contracted. Arthrography can help confirm this diagnosis by showing a decrease in volume of the shoulder joint capsule. Oral steroids improve pain and range of motion in the short term, but a frozen shoulder is probably best treated with a comprehensive program involving NSAIDs and corticosteroid injections into the glenohumeral joint and the subacromial bursa. Physical therapy consists of ice packs, ultrasound, transcutaneous electrical nerve stimulation, and gentle ROM exercises, beginning with pendulum exercises and wall climbing with the fingers and progressing to active ROM and strengthening exercises.

DISORDERS OF THE ELBOW REGION

Olecranon bursitis occurs frequently and involves the subcutaneous olecranon bursa, either secondary to trauma or as an idiopathic condition. The bursa is characteristically swollen and tender on pressure, but pain may be minimal and usually no motion is lost. Aspiration may yield clear or blood-tinged fluid with a low viscosity or grossly hemorrhagic fluid. Inflammatory olecranon bursitis may be caused by gout, RA, or calcium pyrophosphate deposition disease, and infection can also cause a bursitis. Aspiration alone and protection from trauma are usually sufficient to resolve the condition. A small dose of corticosteroid may be injected into the bursa. With septic olecranon bursitis, localized erythema is the major clue. Heat, pain, and a positive culture are also frequently present.

Lateral epicondylitis, or tennis elbow, is a common condition in those who overuse their arms.[4] Localized tenderness directly over or slightly anterior to the lateral epicondyle is the hallmark of this disorder. Pain may occur during handshakes, while lifting a briefcase, or with other similar activities. Probably less than 10% of patients actually acquire lateral epicondylitis through playing tennis. Job and recreational activities, including gardening and athletics, are the usual causes. Pathologically, the condition consists of degeneration of the common extensor tendon, particularly of the extensor carpi radialis brevis tendon.[5] Treatment is aimed at altering activities and preventing overuse of the forearm musculature. Ice packs, heat, and NSAIDs are of some benefit. A forearm brace also can be used. A local corticosteroid injection with a 25-gauge needle over the lateral epicondyle often produces satisfactory initial relief. Isometric strengthening is important as the initial part of a rehabilitation program.

Medial epicondylitis, or golfer's elbow, which mainly involves the flexor carpi radialis, is less common and less disabling than lateral epicondylitis. Local pain and tenderness over the medial epicondyle are present, and resistance to wrist flexion exacerbates the pain.

DISORDERS OF THE WRIST AND HAND

A ganglion is a cystic swelling that arises from a joint or tendon sheath and occurs most commonly over the dorsum of the wrist. It is synovial lined and contains thick, jelly-like fluid. Ganglia apparently develop secondary to trauma or prolonged wrist extension. Usually, the only symptom is swelling, but occasionally a large ganglion produces discomfort on wrist extension.

De Quervain's tenosynovitis may result from repetitive activity that involves pinching with the thumb while moving the wrist. The symptoms are pain, tenderness, and occasionally swelling over the radial styloid. Pathologic findings include inflammation and narrowing of the tendon sheath around the abductor pollicis longus and extensor pollicis brevis. A positive Finkelstein test result is usually seen; pain increases when the thumb is folded across the palm and the fingers are flexed over the thumb as the examiner passively deviates the wrist toward the ulnar side. However, this test also may be positive in patients with osteoarthritis (OA) of the first carpometacarpal joint and must be differentiated from this common condition. Treatment involves splinting, local corticosteroid injection (Fig. 263-4), and NSAIDs as indicated. Rarely, surgical removal of the inflamed tenosynovium is needed.

Volar flexor tenosynovitis consists of inflammation of the tendon sheaths of the flexor digitorum superficialis and flexor digitorum profundus tendons in the palm. It is extremely common but often unrecognized. Pain in the palm is felt on finger flexion, but in some cases the pain radiates to the proximal interphalangeal (PIP) and metacarpophalangeal (MCP) joints on the dorsal side, misleading the examiner. The diagnosis is made by palpation and identification of localized tenderness and swelling of the volar tendon sheaths. The middle and index fingers are most commonly involved, but the ring and little fingers also can be affected. Often a nodule composed of fibrous tissue can be palpated in the palm just proximal to the MCP joint on the volar side. The nodule interferes with the normal tendon gliding and can cause a triggering or locking, which may be intermittent and may produce an uncomfortable sensation. Similar involvement can occur at the flexor tendon of the thumb. The most common cause is overuse trauma of the hands from gripping with increased pull on the flexor tendons. It may be part of inflammatory conditions, such as RA, psoriatic arthritis, or apatite crystal deposition disease. It is seen frequently in conjunction with OA of the hands. Injection

of a long-acting steroid into the tendon sheath usually relieves the problem, although surgery on the tendon sheath may be needed in unremitting cases.

Gamekeeper's thumb (skier's thumb) is caused by trauma to the thumb resulting in instability of the first MCP joint. This instability is due to laxity or rupture of the ulnar collateral ligament. It is treated by immobilization, but surgical repair may be necessary.

Avulsion of flexor digitorum profundus (jersey finger) may result from trauma, usually in football, when a player grabs onto a jersey. The distal phalanx, usually the fourth, is hyperextended while the digitorum profundus is contracting maximally. The avulsion of the tendon results in an inability to flex the distal phalanx of that digit. Surgery is required to correct the problem.

DISORDERS OF THE HIP REGION

Although trochanteric bursitis is common, it frequently goes undiagnosed. It occurs predominantly in middle-aged to elderly people, and somewhat more often in women. The main symptom is aching over the trochanteric area and lateral thigh. Walking, various hip movements, and lying on the involved hip may intensify the pain. Onset may be acute, but more often it is gradual, with symptoms lasting for months. In chronic cases, the patient may fail to locate or describe the pain adequately, or the physician may fail to note the symptoms or interpret them correctly. Occasionally, the pain has a pseudoradiculopathic quality, radiating down the lateral aspect of the thigh. In a few cases, the pain is so severe that the patient cannot walk and complains of diffuse pain of the entire thigh. The best way to diagnose trochanteric bursitis is to palpate over the trochanteric area and elicit point tenderness. In addition to specific pain on deep pressure over the trochanter, other tender points may be noted throughout the lateral aspect of the thigh muscle. Pain may be worse with external rotation and abduction against resistance. Although bursitis has historically been described as the principal problem, the condition may actually arise at the insertions of the gluteus medius and gluteus minimus tendons.[6] Local trauma and degeneration play a role in the pathogenesis, leading to tendinosis and/or tendon tears. Conditions that may contribute to trochanteric bursitis, apparently by adding stress to the area, include OA of the lumbar spine or of the hip, leg-length discrepancy, and scoliosis. Treatment consists of local injection of depot corticosteroid using a 22-gauge, 3.5-inch needle to ensure that the bursal area is reached (Fig. 263-5). NSAIDs, weight loss, and strengthening and stretching of the gluteus medius muscle and iliotibial band help in management.

Coccydynia is manifested by pain in the coccyx area when pressure is applied to the area. This most notably occurs on sitting. The patient squirms from buttock to buttock to relieve the pressure and consequent pain and often chooses to sit on a cushion. The symptoms may be chronic and severe. The condition may relate to a fall on the coccyx, dropping to a hard chair when sitting, or some related trauma to the coccyx. However, at times no obvious cause can be detected. Women are much more frequently affected, perhaps because the lordosis that often occurs in women exposes the coccyx to more trauma. The diagnosis is confirmed by finding localized tenderness over the coccyx on palpation. A plain x-ray film can be obtained to exclude a fracture or dislocation of the coccyx. Treatment with a local injection of 1 mL of a long-acting corticosteroid and 2 mL of a 2% lidocaine solution is usually very effective. The exact nature of the pathology of coccydynia has not been studied, but it is presumed to be a bone bruise.

In iliopsoas bursitis, groin and anterior thigh pain are present and worsen on passive hip hyperextension and sometimes on flexion, especially with resistance. Tenderness is palpable over an involved bursa. The patient may hold the hip in flexion and external rotation to eliminate pain and may limp to prevent hyperextension of the hip. The iliopsoas bursa lies behind the iliopsoas muscle, anterior to the hip joint and lateral to the femoral vessels. It communicates with the hip in 15% of cases. The diagnosis is more apparent if a cystic mass is seen (~30% of cases); however, other causes of cystic swelling in the femoral area must first be excluded. A bursal mass can cause femoral venous obstruction or femoral nerve compression. As with most cases of bursitis, acute or recurrent trauma and inflammatory conditions such as RA may lead to iliopsoas bursitis (also called *iliopectineal bursitis*). Iliopsoas tendinitis may overlap with the bursitis or occur independently in a similar clinical picture. The diagnosis is confirmed by plain x-ray with injection of a contrast medium into the bursa, or by ultrasonography, computed tomography, or MRI. Iliopsoas bursitis/tendinitis usually responds to conservative treatment including physical therapy and corticosteroid injections. With recurrent involvement, excision of the bursa may be necessary.

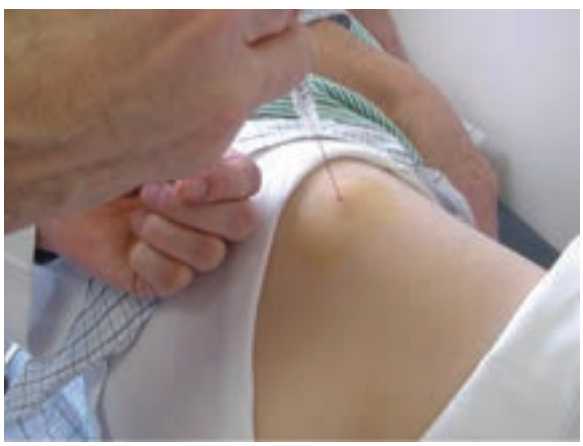

FIGURE 263-5. Injection of trochanteric bursitis.

Ischial or ischiogluteal bursitis is caused by trauma or by prolonged sitting on hard surfaces, as evidenced by the name *weaver's bottom*. Pain is often exquisite when sitting or lying down. The hamstring muscles originate from the ischial tuberosity, and the ischiogluteal bursa is superficial to the tuberosity. Because the bursa is superficial to the tuberosity, separating the gluteus maximus from the tuberosity, the pain may radiate down the back of the thigh. Point tenderness over the ischial tuberosity is present. Use of cushions, hamstring stretching, and local injection of a corticosteroid are helpful.

DISORDERS OF THE KNEE REGION

Anserine bursitis is seen predominantly in overweight, middle-aged to elderly women with large legs and OA of the knees. The symptoms are pain and tenderness over the medial aspect of the knee approximately 2 inches below the joint margin, with the pain worsened by climbing stairs. The pes anserinus (Latin for "goose foot") is composed of the conjoined tendons of the sartorius, gracilis, and semitendinosus muscles. The bursa extends between the above tendons and the tibial collateral ligament. Tendinitis of these tendons, rather than bursitis, is the predominate cause of the syndrome. The diagnosis is made by eliciting exquisite tenderness over the bursal area. Anserine bursitis is often overlooked because it frequently occurs concomitantly with OA of the knee, which, when present, is the assumed cause of pain; however, in some cases of dual involvement, anserine bursitis is the principal source of pain. The treatment is rest, stretching of the adductor and quadriceps muscles, and a corticosteroid injection into the bursa and tendon insertion site.

Prepatellar bursitis manifests as a swelling superficial to the kneecap and results from trauma such as frequent kneeling, leading to the name *housemaid's knee*. The prepatellar bursa lies anterior to the lower half of the patella and the upper half of the patellar ligament. The pain is generally slight unless pressure is applied directly over the bursa. The infrapatellar bursa, which lies between the patellar ligament and the tibia, is also subject to trauma and swelling. Chronic prepatellar bursitis can be treated by protecting the knee from the irritating trauma.

Patellar tendinitis (jumper's knee) is seen predominantly in athletes engaging in activities such as repetitive running, jumping, or kicking. Pain and tenderness are present over the patellar tendon.

Iliotibial band syndrome manifests by lateral knee pain caused by friction between the iliotibial band and the lateral femoral condyle. It is an overuse injury and is seen in runners, cyclists, and other athletes performing repetitive knee flexion activities.

Popliteal cysts,[7] also known as Baker's cysts, are not uncommon, and the clinician should be well aware of the possibility of their dissection or rupture. A cystic swelling behind the knee with mild or no discomfort can be the only initial finding. With further distention of the cyst, however, a greater awareness and discomfort are experienced, particularly on full flexion or extension. The cyst is best seen when the patient is standing and examined from behind. Any knee disease having a synovial effusion can develop into a popliteal cyst. Popliteal cysts are most common secondary to RA, OA, or internal derangements of the knee. There are a few reported cases secondary to gout and Reiter's syndrome. A syndrome of pseudothrombophlebitis may occur as a result of cyst dissection into the calf or actual rupture of the cyst. Findings include diffuse swelling of the calf, pain, and sometimes erythema and edema of the ankle. An ultrasound or arthrogram of the knee confirms both the cyst and the possible dissection or rupture. A cyst related to an inflammatory arthritis is treated by injection of a depot corticosteroid into the knee joint, and possibly into the cyst itself, which usually resolves the problem. If the cyst results from OA or an internal derangement of the knee, surgical repair of the underlying joint lesion is usually necessary to prevent a recurrence of the cyst.

In the knee area, tendon ruptures may occur, and quadriceps tendon rupture is involved approximately 50% of the time; otherwise, patellar tendon rupture occurs. Quadriceps tendon rupture is generally caused by sudden violent contractions of the quadriceps muscle when the knee is flexed. A hemarthrosis of the knee joint may follow. Patients with chronic renal failure, RA, hyperparathyroidism, or gout and patients with SLE taking steroids have been reported to have spontaneous ruptures of the quadriceps tendon. The patient experiences a sudden sharp pain and cannot extend the leg. X-ray studies may show a high-riding patella. The tendon is usually found to be degenerated, and surgical repair is often indicated. Rupture of the patellar tendon has been associated with a specific episode of trauma, repetitive trauma from sporting activities, and systemic diseases.

Meniscal tears are common causes of knee "locking" and pain. Physical examination may show pain, with or without clicking, when the hip and knee are bent to 90°. Magnetic resonance imaging is the diagnostic test of choice. Physical therapy is often as effective as surgery.[A5]

DISORDERS OF THE ANKLE AND FOOT REGION

Achilles tendinitis usually results from trauma, athletic overactivity, or improperly fitting shoes with a stiff heel counter, but it also can be caused by inflammatory conditions such as ankylosing spondylitis, Reiter's syndrome, gout, RA, and calcium pyrophosphate dihydrate crystal deposition disease.[9] Pain, swelling, and tenderness occur over the Achilles tendon at its attachment and in the area proximal to the attachment. Crepitus on motion and pain on dorsiflexion may be present. Management includes NSAIDs, rest, shoe corrections, heel lift, gentle stretching, and sometimes a splint with slight plantar flexion. Local injection of platelet-rich plasma (PRP) has become an increasingly used treatment for releasing growth factors into degenerative tendons; however, more recent randomized, placebo-controlled trials for treatment of chronic Achilles and other tendinopathies have found PRP injections to be ineffective in improving pain and activity.[A1] The Achilles tendon is vulnerable to rupture when involved with tendinitis, and treatment with a corticosteroid injection could increase this possibility.

Achilles tendon rupture is well known and occurs with a sudden onset of pain during forced dorsiflexion. An audible snap may be heard, followed by difficulty in walking and standing on toes. Swelling and edema over the area usually develop. Diagnosis can be made with the Thompson test, in which the patient kneels on a chair with the feet extending over the edge and the examiner squeezes the calf and pushes toward the knee. Normally this produces plantar flexion, but in a ruptured tendon, no plantar flexion occurs. Achilles tendon rupture usually occurs during athletic events or with trauma from jumps or falls. The tendon is more prone to tear in people with preexisting Achilles tendon disease and in those taking corticosteroids. Orthopedic consultation should be obtained, and immobilization or surgery may be selected, depending on the situation.

For acute, severe ankle sprain, a below-knee cast or Aircast produces a faster recovery than a tubular compression bandage, but there is no difference in outcomes at 9 months.[A8] Plantar fasciitis, which is seen primarily in persons between 40 and 60 years of age, is characterized by pain in the plantar area of the heel. The onset may be gradual, or it may occur with trauma or overuse from some activity, such as athletics, prolonged walking, using improper shoes, or striking the heel with some force. Plantar fasciitis may be idiopathic; it also is likely to be present in younger patients with spondyloarthritis (Chapter 265). The pain characteristically occurs in the morning on arising and is most severe for the first few steps. After an initial improvement, the pain may worsen later in the day, especially after prolonged standing or walking. The pain is burning, aching, and occasionally lancinating. Palpation typically reveals tenderness anteromedially on the medial calcaneal tubercle at the origin of the plantar fascia. Treatment includes relative rest with a reduction in stressful activities, NSAIDs, use of heel pad or heel cup orthosis, arch support, and stretching of the heel cord and plantar fascia. A local corticosteroid injection, using a 25-gauge needle, is often of help.

In posterior tibial tendinitis, pain and tenderness occur just posterior to the medial malleolus; it can be caused by trauma, excessive pronation, RA, or spondyloarthropathy. Extension and flexion may be normal, but pain is present on resisted inversion or passive eversion. The discomfort is usually worse after athletic activity, and swelling and localized tenderness may be present. Treatment usually includes rest, NSAIDs, and possibly a local injection of corticosteroid. Immobilization with a splint is sometimes needed.

Posterior tibialis tendon rupture, which is not commonly recognized, is a cause of progressive flat foot. It can result from trauma, chronic tendon degeneration, or RA. An insidious onset of pain and tenderness may be noted along the course of the tendon just distal to the medial malleolus, along with swelling medial to the hind foot. The unilateral deformity of hind foot valgus and forefoot abduction is an important finding. The forefoot abduction can best be seen from behind; more toes are seen from this position than would be seen normally. The result of the single heel rise test is positive when the patient is unable to rise onto the ball of the affected foot while the contralateral foot is off the floor. Treatment usually includes rest, NSAIDs, and possibly an orthosis. Surgical repair of the tendon is sometimes indicated. Manifestations of retrocalcaneal bursitis include pain at the back of the heel, tenderness of the area anterior to the Achilles tendon, and pain on dorsiflexion. Local swelling is present, with bulging on the medial and lateral aspects of the tendon. Retrocalcaneal bursitis, also called sub-Achilles bursitis, may coexist with Achilles tendinitis, and distinguishing the two is sometimes difficult. This condition may be secondary to RA, spondylitis, a reactive arthritis, gout, or trauma.

Turf toe is an injury of the big toe originally described during play on artificial turf. It results from hyperextension of the first metatarsophalangeal (MTP) joint when a fixed, dorsiflexed foot is forced into the ground. The plantar capsular ligament may be sprained or torn.

Stress fracture is also known as march fracture or fatigue fracture because it was first associated with spontaneous fracture after long marches in army recruits. Pain, swelling, tenderness, and occasionally erythema develop over the metatarsal area, usually without any clear history of trauma. On questioning, however, the episode of spontaneous pain related to onset of the fracture can be identified in some cases. The neck of the second metatarsal bone is most frequently involved, but the third metatarsal is also a site of fracture. Aside from prolonged marching, other athletic events with overactivity, including jogging, are common causes. Stress fractures may be seen in patients with RA and in elderly people. The difficulty in diagnosing stress fractures is that the initial x-ray films usually show no abnormalities or, at most, only a faint fracture line. A repeat x-ray examination several weeks later shows healing with callus formation. Bone scans aid the early diagnosis of stress fractures by showing an increase in uptake over the fracture site. Usually these fractures heal spontaneously, and rest and strapping of the foot are helpful. Occasionally, a cast is needed.

Grade A References

A1. Coombes BK, Bisset L, Vicenzino B. Efficacy and safety of corticosteroid injections and other injections for management of tendinopathy: a systematic review of randomised controlled trials. *Lancet*. 2010;376:1751-1767.

A2. Rhon DI, Boyles RB, Cleland JA. One-year outcome of subacromial corticosteroid injection compared with manual physical therapy for the management of the unilateral shoulder impingement syndrome: a pragmatic randomized trial. *Ann Intern Med*. 2014;161:161-169.

A3. Cacchio A, De Blasis E, Desiati P, et al. Effectiveness of treatment of calcific tendinitis of the shoulder by disodium EDTA. *Arthritis Rheum*. 2009;61:84-91.

A4. Bannuru RR, Flavin NE, Vaysbrot E, et al. High-energy extracorporeal shock-wave therapy for treating chronic calcific tendinitis of the shoulder: a systematic review. *Ann Intern Med*. 2014;160:542-549.

A5. Katz JN, Brophy RH, Chaisson CE, et al. Surgery versus physical therapy for a meniscal tear and osteoarthritis. *N Engl J Med*. 2013;368:1675-1684.

A6. Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. *N Engl J Med*. 2013;369:2515-2524.

A7. Moraes VY, Lenza M, Tamaoki MJ, et al. Platelet-rich therapies for musculoskeletal soft tissue injuries. *Cochrane Database Syst Rev*. 2014;4:CD010071.

A8. Lamb SE, Marsh JL, Hutton JL, et al. Mechanical supports for acute, severe ankle sprain: a pragmatic multicentre, randomised controlled trial. *Lancet*. 2009;373:575-581.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hall MM, Finnoff JT, Smith J. Musculoskeletal complications of fluoroquinolones: guidelines and precautions for usage in the athletic population. *PM R*. 2011;3:132-142.
2. Huegel J, Williams AA, Soslowsky LJ. Rotator cuff biology and biomechanics: a review of normal and pathological conditions. *Curr Rheumatol Rep*. 2015;17:476.
3. Hermans J, Luime JJ, Meuffels DE, et al. Does this patient with shoulder pain have rotator cuff disease?: The Rational Clinical Examination systematic review. *JAMA*. 2013;310:837-847.
4. Brummel J, Baker CL 3rd, Hopkins R, et al. Epicondylitis: lateral. *Sports Med Arthrosc*. 2014;22: e1-e6.
5. Frick MA, Murthy NS. Imaging of the elbow: muscle and tendon injuries. *Semin Musculoskelet Radiol*. 2010;14:430-437.
6. Klauser AS, Martinoli C, Tagliafico A, et al. Greater trochanteric pain syndrome. *Semin Musculoskelet Radiol*. 2013;17:43-48.
7. Herman AM, Marzo JM. Popliteal cysts: a current review. *Orthopedics*. 2014;37:e678-e684.
8. Nguyen JC, De Smet AA, Graf BK, et al. MR imaging-based diagnosis and classification of meniscal tears. *Radiographics*. 2014;34:981-999.
9. Calleja M, Connell DA. The Achilles tendon. *Semin Musculoskelet Radiol*. 2010;14:307-322.

REVIEW QUESTIONS

1. A 78-year-old woman who has had periodic discomfort in her neck and radiating to both shoulders for several years had a fall, landing on her outstretched right arm. She had pain in the right shoulder and weakness on use. At the urgent care center she was found to have some pain on passive motion and was unable to actively abduct the right shoulder. She had a positive drop arm sign of the right shoulder but not of the left one. All upper extremity deep tendon reflexes were normal. Which of the following is the *most likely* diagnosis?

A. Rotator cuff tendinitis is the main diagnosis.
B. X-ray examination of the humerus of the injured side is not indicated in this case because tendon injuries are not visualized on plain x-ray film of the shoulder.
C. Bicipital tendinitis is the likely diagnosis.
D. A tear of the rotator cuff best fits as the diagnosis.
E. Cervical spine disc herniation is the most likely diagnosis based on the history of neck pain and arm weakness.

Answer: D Acute trauma is often the cause of rotator cuff tears, although most tears occur gradually through chronic tendinitis. The positive drop arm sign is typical of a significant rotator cuff tear. Sometimes with trauma, a humeral fracture can occur along with the rotator cuff tear. Pain on abduction and especially on resisted abduction are seen in rotator cuff tendinitis; also, the drop arm sign is typically negative. Bicipital tendinitis usually does not cause shoulder weakness. The clinical picture does not fit a cervical disc herniation, and the reflexes are normal.

Biundo, JJ: Regional Rheumatic Pain Syndromes, Primer on the Rheumatic Diseases, Thirteenth Edit. 2006

2. A 51-year-old male insurance agent, whose hobby is wood working, complains of right hand pain of approximately 4 months' duration. He had been doing more wood working during that period until the hand pain began to limit this activity. On examination there were no Heberden's nodes of the DIP joints, and the PIP and MCP were not tender or swollen. Moderate-to-severe tenderness on palpation was present over the palm side of the hand at the region of the second and third metacarpal bones. Some swelling was detected at these sites. Crepitus was noted on palpation on flexion of these two digits. Which of the following is the *most likely* diagnosis?

A. Because rheumatoid arthritis is a likely diagnosis, you should obtain RF, ANA, and hand x-ray films.
B. Because of his age, he most likely has osteoarthritis of the hand.
C. He fits the picture of volar flexor tenosynovitis
D. A C7 radiculopathy would best explain pain in this part of his hand.
E. Overuse of his hands has caused carpal tunnel syndrome.

Answer: C Volar flexor tenosynovitis is one of the most common musculoskeletal syndromes and is frequently overlooked. It is manifest by pain in the palm of the hand. The volar flexor tendon sheath is tender on palpation and thickened. Triggering or snapping of the digit sometimes occurs, especially when more chronic. Overuse of the hand is the usual cause, but is also seen in association with OA of the hands and diabetes. Stretching of the involved digit and an injection of a small amount of a corticosteroid into the tendon sheath is usually helpful.

Biundo, JJ: Regional Rheumatic Pain Syndromes, Primer on the Rheumatic Diseases, Thirteenth Edit. 2006

3. A 48-year-old male professional cellist, whose hobby is gardening, has had a 4-month history of right shoulder pain. Pain occurs on various movements of the shoulder, and there is some limitation of movement. Night pain in bed is also reported. On examination, pain limits active abduction to 90 degrees. Less pain is noted on passive abduction, and 180 degrees of abduction is present. The drop arm sign is negative, but pain occurs on active internal rotation of the shoulder. There is no crepitus. Which of the following is the *most likely* diagnosis?

A. Rotator cuff tendinitis is the most likely diagnosis.
B. Rotator cuff tear best fits this clinical picture.
C. Rupture of the biceps tendon is the diagnosis because the drop arm sign is negative.
D. Adhesive capsulitis (frozen shoulder) is the diagnosis.
E. OA of the shoulder (glenohumeral joint) is the likely diagnosis, and a plain x-ray film of the shoulder is needed to confirm the diagnosis.

Answer: A The typical physical examination findings of rotator cuff tendinitis are pain on active abduction, less pain on passive abduction, and more pain on resistive abduction. Loss of motion on active movements can occur when more severe. Also, pain may occur on active internal and on external rotation. In addition to the physical examination, ultrasonography and MRI can detect tendinitis, but not plain x-ray examination. In adhesive capsulitis restriction, the range of motion in all directions is usually seen. The full range on abduction and other motions in this case are not typical of a frozen shoulder. OA of the shoulder would not have had a pain history of only 4 months, and it is a much less frequent entity than rotator cuff tendinitis. Crepitus on range of motion is very common in OA of the glenohumeral joint.

Biundo, JJ: Regional Rheumatic Pain Syndromes, Primer on the Rheumatic Diseases, Thirteenth Edit. 2006

4. A 76-year-old woman with a known history of moderate OA of both knees and a partially torn right medial meniscus experienced pain in her right knee after stooping down to clean something off the floor. The next day she had discomfort and fullness behind her right knee and swelling of her lower leg and ankle. She was able to walk but had a limp. She was seen that day by her primary care physician for evaluation. Which of the following is the *best* course of action?

A. Begin treatment with heparin or enoxaparin for a deep vein thrombosis.
B. Give oral prednisone for 5 days as a tapering dose for an acute flare of the OA of her knee.
C. Obtain a new x-ray film of the knees, with a standing AP and lateral view of the involved leg.
D. Because of the acute flare of knee symptoms, give colchicines for a possible acute attack of gout.
E. Obtain a Doppler ultrasound of the right leg for a deep vein thrombosis and an ultrasound image of the right knee to identify a Baker's cyst that has dissected, causing a pseudothrombophlebitis picture.

Answer: E Popliteal cysts, also known as Baker's cysts, are often asymptomatic. However, at times they may enlarge as a result of increasing pressure from a synovial effusion of the knee, flowing through a one-way valve from the knee to the popliteal cyst. The cyst can at times dissect downward, causing swelling of the leg and simulating thrombophlebitis. Thus, the name given to the syndrome is pseudothrombophlebitis. The swelling of the leg may resolve on its own after several days. Often, a corticosteroid injection into the knee joint may help by decreasing the inflammation and synovial effusion of the knee. OA of the knee should not produce swelling of the lower leg, nor should gout. Treatment with anticoagulants should not be given unless the diagnosis of DVT were confirmed.

Biundo, JJ: Regional Rheumatic Pain Syndromes, Primer on the Rheumatic Diseases, Thirteenth Edit. 2006

5. A 52-year-old moderately obese man complains of pain in his right foot that has been present for approximately 4 months. The pain has become worse recently, and he has had to stop the walks he was taking for exercise. The pain is described to be more in the rear foot and is particularly bad when he first gets out of bed in the morning. He has no history of gout or other arthritis except for some back pain. On examination there is no joint swelling. He does have pes cavus, and tenderness of the plantar surface of the rear foot. His calf muscles are also tight. What is the *most likely* diagnosis and/or the *best* course of evaluation and treatment?

A. Order an MRI of the foot or technetium bone scan for possible stress fracture.
B. Inject the calcaneus with a steroid, because that is the best initial treatment for heel pain.
C. Make a tentative diagnosis of plantar fasciitis and treat with analgesics, stretching of calf and plantar fascia, and use of a heel cushion as the initial treatment.
D. Obtain uric acid level and start treatment with colchicine.
E. Obtain orthopedic consultation for possible Achilles tendon tear.

Answer: C The complaint of morning foot pain, which is called "first step pain," is typical of plantar fasciitis. Tenderness often elicited on pressure exerted on the plantar surface of the calcaneus. Trauma to the foot from prolonged walking, tight calf muscles, high arched foot with tight plantar fascia, and obesity are factors related to acquiring plantar fasciitis. A steroid injection of the cacaneal area is often helpful, but is not the first line of treatment. With Achilles tendinitis/tear the pain is in the posterior heel and not the plantar surface. A stress fracture is more likely to occur in the metatarsal area and only very rarely occur in the calcaneus.

Biundo, JJ: Regional Rheumatic Pain Syndromes, Primer on the Rheumatic Diseases, Thirteenth Edit. 2006

264

RHEUMATOID ARTHRITIS

JAMES R. O'DELL

DEFINITION

Rheumatoid arthritis (RA) is a chronic systemic inflammatory disease of unknown etiology that primarily targets synovial tissues. It is relatively common, with a prevalence of slightly less than 1% in adults all over the

world. RA shortens survival and significantly affects quality of life in many patients. Essentially all patients exhibit some systemic features such as fatigue, low-grade fevers, anemia, and elevations of acute phase reactants (erythrocyte sedimentation rate [ESR] or C-reactive protein [CRP]). This systemic inflammation is believed to be responsible for vascular endothelial damage and a marked increased risk for coronary artery disease and congestive heart failure in patients with RA.[1] However, the primary target of RA is the synovium and it is responsible for most of the protean clinical features. Synovial tissues proliferate in an uncontrolled fashion, resulting in excess fluid production, destruction of cartilage, erosion of marginal bone, and stretching and damage of the tendons and ligaments.

In the past two decades, the treatment of RA has changed dramatically. Current therapeutic strategies should result in over 50% of patients achieving clinical remissions with treatment with appropriate disease-modifying antirheumatic drugs (DMARD) or combinations of DMARDs.

EPIDEMIOLOGY

RA is present all over the world, with a prevalence of 0.5 to 1% of adults and with some differences in certain population groups. For reasons that are still unclear, the prevalence in women is two or three times greater than that in men. RA can occur at any age, but onset before the age of 45 years in men is uncommon. The relatively few well-done inception cohorts suggest that the yearly incidence of RA is approximately 40 per 100,000 for women and about half that for men. These figures vary significantly based on the age of the cohort. The best available data suggest that the incidence of RA in women increases with age until approximately 60 years of age and then plateaus. The incidence rate is much lower in young men, approximately one third that in women, but increases steadily with age and approaches that of women older than 65 years. Because the incidence of RA increases or is stable with age and RA is a lifelong disease, the prevalence of RA increases with each decade. Recent data suggest that the incidence of RA, particularly rheumatoid factor (RF)-negative RA, may be decreasing. The reasons for this are unclear, but, if elucidated, they could provide valuable insights into the etiology and pathogenesis of RA and might allow the implementation of strategies to prevent clinical disease.

RA has a significant genetic component; therefore, it is not surprising that RA is reportedly very unusual in certain populations and more common in others. Most notably, cohorts have been described in rural Nigeria in which no individuals are affected with RA; in contrast, a prevalence of RA of 5% has been found in some studies of Chippewa, Yakima, and Inuit Native American tribes.

PATHOBIOLOGY

Genetics

Genetics play a significant role in determining both the risk for developing RA and the severity of the disease.[2] Twin studies reveal a concordance rate for RA that averages 15% for monozygotic twins and approximately 5% for dizygotic twins. These data in monozygotic twins simultaneously reveal both the significance of genetic factors and the fact that they are clearly not the only important factor, or else the concordance rate would approach unity.

It has been clearly shown that RA is a multigene disease with important contributions from both human leukocyte antigen (HLA) and non-HLA genes. The association of certain HLA alleles, specifically HLA-DR4, with an increased risk for developing RA and of having more severe disease has long been recognized. This association is explained by a particular amino acid sequence in the third hypervariable region on the DRβ1 chain. HLA-DR molecules are present on the surface of antigen-presenting cells and allow T cells to recognize antigen in the context of DR. Hypervariable regions on the DR molecule are particularly important for antigen recognition. E-Table 264-1 details the amino acid sequence of several DRβ1 chains that are associated with RA and some that are not. The amino acid sequence associated with RA has been called the *shared epitope* or the *at-risk allele*. It has been shown by a number of investigators that patients with the shared epitope have more severe RA and more extra-articular manifestations than those who are negative. Furthermore, individuals with two copies of the shared epitope, particularly those with HLA-DR4, have a further increased risk for the development of severe RA. This association with a particular antigen recognition site may ultimately aid understanding of the antigen or antigens that are important for triggering RA. Proteins in which arginine has been converted to citrulline are bound with greater avidity by the shared epitope. The importance of certain DRβ1 alleles in RA supports the concept that T cells are integrally involved in the pathogenesis.

Population-based studies have suggested that only 30 to 50% of the genetic risk for RA is explained by genes located in the HLA region. A functional polymorphism for the gene that encodes intracellular protein tyrosine phosphatase nonreceptor 22 *(PTPN22)* has been reproducibly associated with RA and a number of other autoimmune diseases, including type 1 diabetes, systemic lupus erythematosus, Graves' disease, and Hashimoto's thyroiditis. Genome-wide association studies have identified at least 80 other candidate genes associated with RA, including polymorphisms for signal transducer and activator of transcription (STAT4), tumor necrosis factor receptor–associated factor 1 (TRAF-1), and CD40. To complicate things further, HLA-DRB1 03 is associated with lower titers of anti–citrullinated peptide antigen (ACPA) antibodies but is associated with increased risk for cyclic citrullinated peptide–negative RA.

The shared epitope is present in approximately 25 to 35% of the white population, but the chance of developing RA among individuals who carry this allele is only approximately 1 in 25. Therefore, despite identifying the most important genetic risk factor for RA, this test has little or no clinical utility. The role of epigenetics in RA is currently receiving attention and may provide important insights.

Etiology

Clearly, other factors, in addition to genetics, are active in precipitating or triggering RA. RA appears to require the complex interaction of genetic and environmental factors with the immune system and ultimately in the synovial tissues throughout the body (Fig. 264-1).[3] Sera collected before the development of clinical RA show that immunologic changes predate clinical manifestations by years. Autoantibodies, particularly ACPA antibodies and rheumatoid factor (RF), are present in the sera of many individuals 5 to 10 years before the clinical onset of disease. By following cohorts of people at high risk for RA, investigators are learning much about these early immunologic changes and ultimately it is hoped about the triggers for the disease.

The use of oral contraceptives has been associated with a decrease in the incidence of RA; because the effect appears to be strongest for oral contraceptives that have high estrogen content, it is postulated that estrogen is responsible for this protective effect. Studies that have tried to address the question of postmenopausal estrogen use and its effect on RA have yielded conflicting results.

Smoking has long been associated with a significant increase in the risk for developing RA, but more recently it has been shown that this is true only for ACPA-positive patients and is not associated with ACPA-negative disease. Furthermore, smoking appears to be a risk factor for RA only in those patients who are positive for shared epitope. Purported triggers for RA in addition to smoking have included bacteria *(Mycobacteria, Streptococcus, Mycoplasma, Escherichia coli, Helicobacter pylori)*, viruses (rubella, Epstein-Barr virus, parvovirus), and periodontal disease.

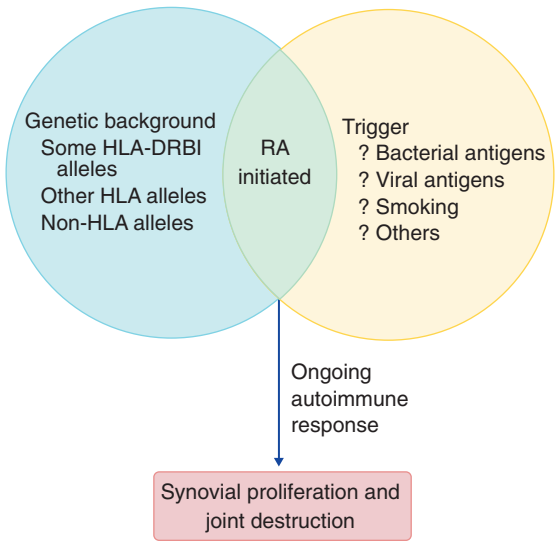

FIGURE 264-1 Initiation of rheumatoid arthritis (RA). HLA = human leukocyte antigen.

E-TABLE 264-1 HUMAN LEUKOCYTE ANTIGEN ASSOCIATIONS WITH RHEUMATOID ARTHRITIS

HLA TYPES (ALLELES) AND METHODS OF DETECTION			*THIRD HYPERVARIABLE REGION AMINO ACID SEQUENCES*					
ALLOANTISERA (DR)	**MLC (DW)**	**DNA (DRβ1)**	**70**	**71**	**72**	**73**	**74**	**MOST COMMON ETHNIC GROUPS**
ASSOCIATED WITH RA								
DR4	Dw4	*0401	Q	K	R	A	A	Whites (Western Europe)
DR4	Dw14	*0404		R				Whites (Western Europe)
DR4	Dw15	*0405		R				Japanese, Chinese
DR1	Dw1	*0101		R				Asian Indians, Israelis
DR6 (14)	Dw16	*1402		R				Yakima Native Americans
DR10	—	*1001	R	R				Spanish, Greeks, Israelis
NOT ASSOCIATED WITH RA								
DR4	Dw10	*0402	D	E				Whites (Eastern Europe)
DR4	Dw13	*0403		R			E	Polynesians
DR2	Dw2	*1501	D	A				Whites
DR3	Dw3	*0301				G	R	Whites

A = alanine; D = aspartic acid; E = glutamic acid; HLA = human leukocyte antigen; K = lysine; MLC = mixed leukocyte cultures; Q = glutamine; R = arginine; * = the same amino acid in that position as for DRB1*0401.

Pathogenesis

The pathogenesis of RA is complex, and there are almost certainly multiple triggering mechanisms, including but not limited to smoking, infection, molecular mimicry, immune complexes, altered T-cell repertoire, and T-cell reactivity. Furthermore, it is likely that the triggers may be different based on the genetic background. As mentioned previously, smoking is a well-known trigger for some individuals but appears to be a risk factor only in those patients who possess the shared epitope.

Rheumatic fever, reactive arthritis (formerly known as *Reiter's syndrome*), and, more recently, Lyme arthritis are examples of arthritic syndromes for which infectious triggers have clearly been demonstrated, but these triggering agents are often difficult or impossible to isolate at the time when the arthritic syndromes occur. Many other examples exist in animal models of arthritis, including syndromes induced by mycobacteria and streptococci. Reactive arthritis (Chapter 265) has clearly been shown to occur when any one of a myriad of different but specific infectious triggers is presented to a specific location in the body (the gastrointestinal or genitourinary tract) of individuals with a certain genetic background, in most cases HLA-B27. Additionally, in this syndrome, the age and gender of the individual and hence the maturity of the immune system may be critical in the development of clinical disease, which occurs primarily between the ages of 15 and 40 years in males. Once unraveled, the pathophysiology of RA is likely to be similarly complex.

Despite the absence of clear evidence linking any infectious agent to RA, it is widely believed that ultimately an important triggering role will be elucidated for infectious or other environmental agents. Once triggers for RA are identified, strategies for prevention can be addressed, but this information may not help individuals with established disease. Possibly infections involving the innate immune system are causative in an early subclinical phase of the rheumatoid disease process, with the agents being absent once clinical disease develops.

The relative roles of the cellular versus the humoral immune system in the initiation and perpetuation of RA are much debated; both appear to be important. Most likely, the mechanisms of initiation of the disease process are different from those that perpetuate the chronic disease. T cells, particularly of the activated T_H1 and T_H17 types, appear to predominate in synovial tissues. These T cells, presumably activated by some as yet unknown antigen presented by macrophages, B cells, or synoviocytes in the context of HLA-DR, secrete cytokines that drive further synovial proliferation. It is believed by many that, although RA may initially be triggered by exogenous antigens, the process, once initiated, may be perpetuated by autoantigens. Macrophage-derived cytokines, particularly interleukin-1 (IL-1) and tumor necrosis factor-α (TNF-α), play central roles in this ongoing inflammatory process. As definitive proof, biologic products directed against these cytokines have shown significant efficacy in the treatment of RA.

The humoral immune system also plays a role. RF has long been a serologic marker of RA and is well known to correlate with more severe disease, including erosions of bone, and with the presence of extra-articular features. The reason that RF is produced in excess and the exact role that it plays remain elusive. RF production may increase complement activation and result in the release of lysosomal enzymes, kinins, and oxygen free radicals. ACPA antibodies exhibit a high specificity (95 to 99%) for RA, although their sensitivity for RA with currently available assays is only approximately 70%. Even though both RF and ACPA antibodies also correlate with more aggressive erosive disease, this link is strongest for ACPA antibodies.

Pathology

The synovial tissues are the primary target of the autoimmune inflammatory process that is RA; the reason for this remains elusive. However, the generalized inflammation of RA also involves the vascular endothelium and results in significant premature atherosclerosis. Once RA is initiated, the synovial tissues throughout the body become the site of a complex interaction of T cells, B cells, macrophages, and synovial cells (Fig. 264-2). The resultant proliferation of the synovial tissues (synovitis) causes the production of excessive amounts of synovial fluid and the infiltration of pannus into adjacent bone and cartilage. Synovitis results in the destruction of cartilage and marginal bone and in the stretching or rupture of the joint capsule or tendons and ligaments. In patients, these effects are manifested by the deformities (Fig. 264-3 and E-Fig. 264-1) and disabilities that make up the clinical picture of RA.

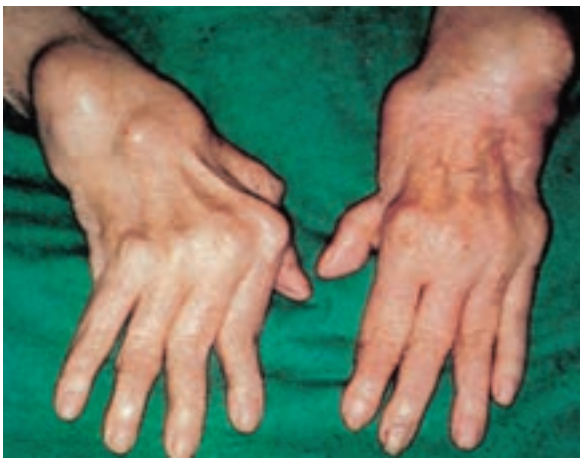

FIGURE 264-3. Severe advanced rheumatoid arthritis of the hands. There is massive tendon swelling over the dorsal surface of both wrists, severe muscle wasting, ulnar deviation of the metacarpophalangeal joints, and swan-neck deformity of the fingers. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine,* 3rd ed. London: Mosby; 2003.)

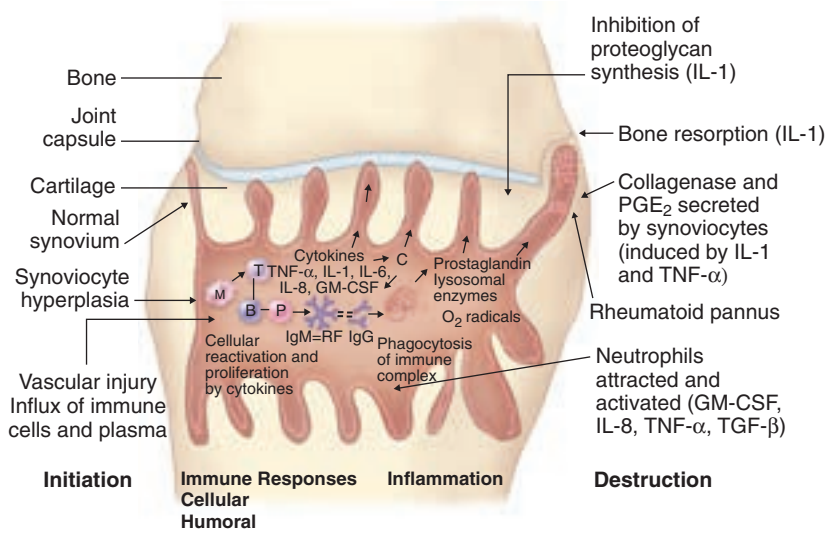

FIGURE 264-2. Events involved in the pathogenesis of rheumatoid synovitis (progressing from *left* to *right*). B = B lymphocyte; C = complement; GM-CSF = granulocyte-macrophage colony-stimulating factor; IgG, IgM = immunoglobulin G, M; IL = interleukin; M = macrophage; P = plasma cell; PGE_2 = prostaglandin E_2; RF = rheumatoid factor; T = T lymphocyte; TGF-β = transforming growth factor-β; TNF-α = tumor necrosis factor-α.

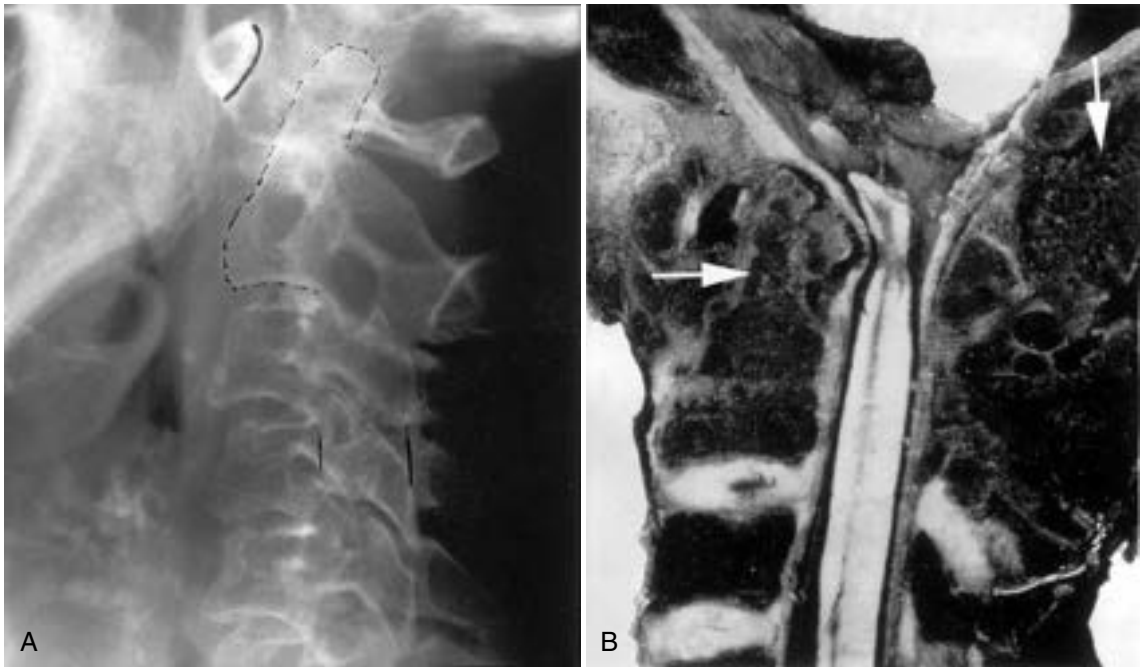

E-FIGURE 264-1. Subluxation of the cervical spine in patients with rheumatoid arthritis. A, In a lateral radiograph of the cervical spine, the body of C2 and its odontoid process are outlined by the *broken lines* and the posterior aspect of the anterior segment of C1 is indicated by a *solid line*. Normally, a space of only 2 to 3 mm separates C1 from C2. The space between C1 and the odontoid of C2 is markedly increased, indicative of subluxation of C1 and C2. **B,** Lateral view of a pathologic specimen from a patient who died of C1-C2 subluxation. The *horizontal arrow* shows the odontoid process that subluxed posteriorly, severely compressing and almost severing the cord. The *vertical arrow* shows a bone graft that had been put in place posteriorly in an attempt to prevent subluxation. Below the *arrow,* a nonhealing area is present through the bone graft, and inferior to that a wire fixation suture is still in place.

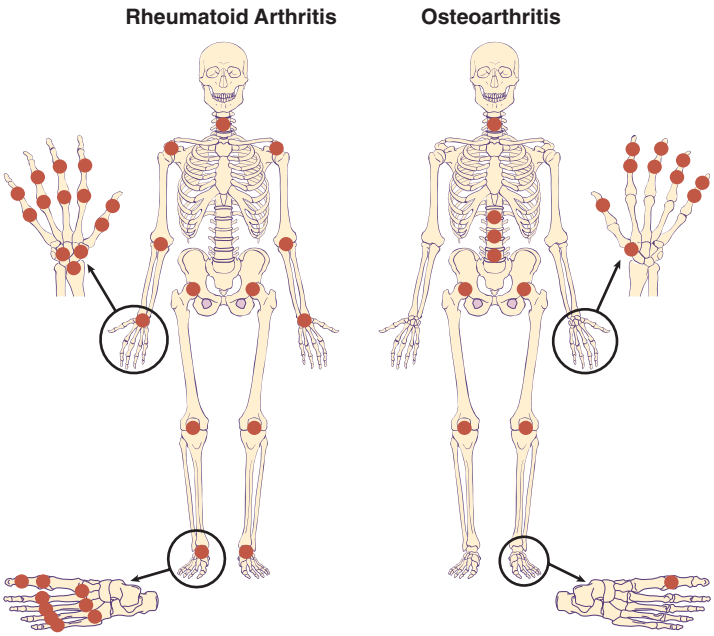

FIGURE 264-4. Distribution of involved joints in the two most common forms of arthritis: rheumatoid arthritis and osteoarthritis. *Shaded circles* are shown over the involved joint areas.

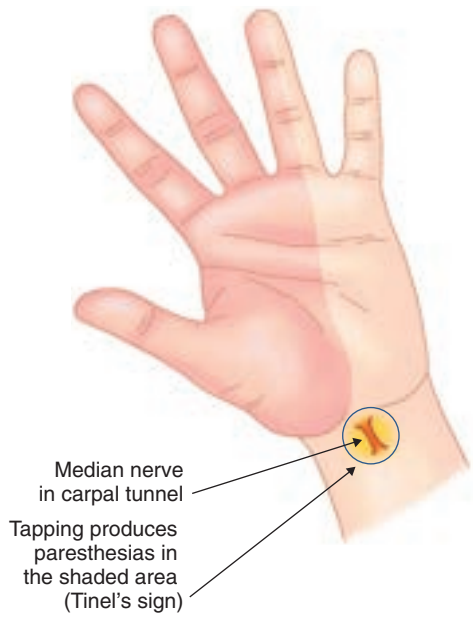

FIGURE 264-5. Carpal tunnel syndrome. Distribution of pain and/or paresthesias (*shaded area*) when the median nerve is compressed by swelling in the wrist (carpal tunnel).

CLINICAL MANIFESTATIONS

Articular Manifestations

RA can affect any of the synovial (diarthrodial) joints (Fig. 264-4). Most commonly, the disease starts in the metacarpophalangeal (MCP), proximal interphalangeal (PIP), and metatarsophalangeal (MTP) joints, followed by the wrists, knees, elbows, ankles, hips, and shoulders, in roughly that order. Early treatment limits the joints involved. Less commonly, and usually later, RA may involve the temporomandibular, cricoarytenoid, and sternoclavicular joints. RA may involve the upper part of the cervical spine, particularly the C1-C2 articulation, but, unlike the spondyloarthropathies (Chapter 265), it does not involve the rest of the spine. RA patients are at an increased risk for osteoporosis (Chapter 243), and this risk should be considered and dealt with early.

Hands

The hands are a major site of involvement, and a significant portion of the disability that RA causes is because of damage and dysfunction of the hands. Typically disease starts with swelling of the PIPs and MCPs. The distal interphalangeal (DIP) joints are almost never involved; significant involvement of the DIP joints should suggest the possibility of a different diagnosis (i.e., osteoarthritis or psoriatic arthritis). Figure 264-3 illustrates the classic ulnar deviation of the MCP joints and swan-neck deformities (hyperextension of the PIP joints) that are commonly seen in late disease. Boutonnière (or buttonhole) deformities also occur as a result of hyperflexion of the PIP joints. If the clinical disease remains active, hand function deteriorates. Sudden loss of function of individual fingers may occur as a result of tendon rupture, which requires the expertise of a hand surgeon to repair.

Feet

Feet, particularly the MTP joints, are involved early in most patients with RA. Radiographic erosions occur at least as early in the feet as in the hands. Subluxation of the toes is common and leads to the dual problem of breakdown of the skin and ulcers on the top of the toes and malalignment of the MTP heads. Painful ambulation develops owing to loss of the cushioning pads that usually protect the heads of the MTP joints.

Wrists

The wrist joints are involved in most patients with RA; radial deviation is the rule, and patients with severe involvement may progress to volar subluxation. Even early in the course of the disease, synovial proliferation in and around the wrists may compress the median nerve, causing carpal tunnel syndrome (Fig. 264-5). Later, this synovial proliferation may invade tendons and lead to rupture of extensor tendons.

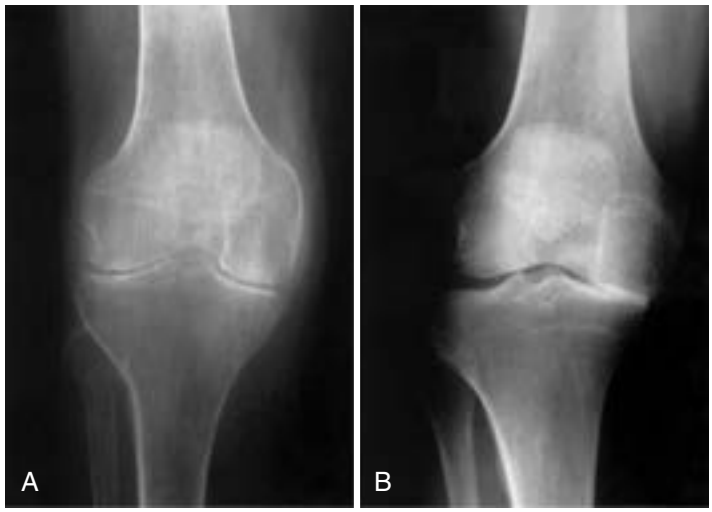

FIGURE 264-6. Radiographs of the knees in the two most common forms of arthritis: rheumatoid arthritis and osteoarthritis. A, Severe involvement in rheumatoid arthritis, with almost complete symmetrical loss of joint space in both the medial and the lateral compartments, but with little subchondral sclerosis or osteophyte formation. **B,** Typical osteoarthritis, with severe, near-total loss of joint space of one compartment and a normal or actually increased joint space of the other compartment. Note also the significant subchondral sclerosis in the involved area, typical of osteoarthritis.

Large Joints

Involvement of knees, ankles, elbows, hips, and shoulders is common. Characteristically, the whole joint surface is involved in a symmetrical fashion. Therefore, RA is symmetrical not only from one side of the body to the other but also within the individual joint. In the case of the knee (Fig. 264-6A), the medial and lateral compartments are both severely narrowed in RA; in contrast, in patients with osteoarthritis (see Fig. 264-6B) typically only one compartment of the knee is involved.

Synovial cysts may occur around any of the joints (large or small), and they occasionally manifest as soft, fluctuant masses that present diagnostic challenges. When the knee produces excess synovial fluid, it may accumulate in the popliteal space (popliteal or Baker's cyst) (E-Fig. 264-2). These cysts can cause problems by pressing on the popliteal nerve, artery, or veins. Baker's cysts may dissect into the tissues of the calf (usually posteriorly), or they may rupture into the upper calf. Dissection may produce only minor symptoms, such as a feeling of fullness; rupture of the cyst with extravasation of the

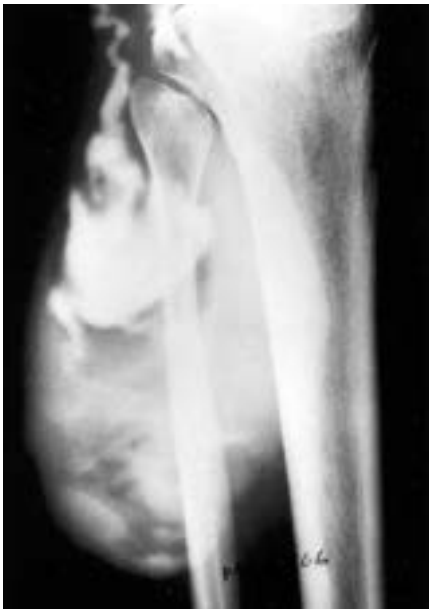

E-FIGURE 264-2. Arthrogram with a radiocontrast agent injected into the knee. The dye flows into the popliteal space and through a narrow channel into a large synovial cyst (Baker's cyst) that has dissected into the soft tissue of the calf.

inflammatory content produces significant pain and swelling and may be confused with thrombophlebitis, the so-called pseudothrombophlebitis syndrome. Ultrasonography of the popliteal fossa and calf is useful to establish the correct diagnosis and rule out thrombophlebitis, which may be precipitated by popliteal cysts. Treatment of popliteal or any other cyst should be directed at interrupting the inflammatory process through an intra-articular injection of corticosteroid into the associated joint.

Neck

Although most of the axial skeleton is spared in RA, the cervical spine is commonly involved, particularly the C1-C2 articulation. Bony erosions and ligament damage can occur in this area and may lead to subluxation (see E-Fig. 264-1). Most often, subluxation at C1-C2 is minor and without accompanying symptoms; patients and caregivers need only be cautious and avoid actively forcing the neck into positions of flexion. Occasionally, subluxation at C1-C2 is severe and leads to compromise of the cervical cord with symptoms and in some cases death.

Other Joints

Wherever synovial tissue exists, RA can cause problems. The temporomandibular, cricoarytenoid, and sternoclavicular joints are examples of other joints that may be involved in RA. The cricoarytenoid joint is responsible for abduction and adduction of the vocal cords. Involvement of this joint may lead to a feeling of fullness in the throat, to hoarseness, and, rarely, when the cords are essentially fused in a closed position, to a syndrome of acute respiratory distress with or without stridor. In this latter situation, emergent tracheotomy may be life-saving.

Extra-Articular Manifestations

Systemic features of RA such as fatigue, weight loss, and low-grade fevers occur frequently. As with all the other extra-articular features, they are more common in those patients who possess RF or ACPA antibodies or both (Table 264-1) and respond to treatment of the RA.

Skin

Subcutaneous nodules are seen in approximately one fifth of patients with RA, almost exclusively in those who are RF positive. Patients with nodules who are RF negative should be carefully scrutinized for a different diagnosis, such as chronic tophaceous gout. Nodules may occur almost anywhere (e.g., lungs, heart, eye), but most commonly they occur subcutaneously on extensor surfaces (particularly the forearms) (Fig. 264-7), over joints, or over pressure points. Rheumatoid nodules are firm on examination, usually are not tender, have a characteristic histologic picture, and are thought to be initiated by small vessel vasculitis. A syndrome of increased nodulosis, despite good control of the joint disease, has been described with methotrexate therapy (Fig. 264-8).

Small vessel vasculitis,[4] manifested as digital infarcts or leukocytoclastic vasculitis, may occur in RA (Fig. 264-9) and should prompt more aggressive DMARD treatment. A vasculitis of small and medium arteries that is indistinguishable from polyarteritis nodosa also can be seen with RA and requires aggressive systemic therapy. Finally, pyoderma gangrenosum occurs with increased frequency in association with RA.

TABLE 264-1 EXTRA-ARTICULAR MANIFESTATIONS OF RHEUMATOID ARTHRITIS

Skin	Nodules, fragility, vasculitis, pyoderma gangrenosum
Heart	Pericarditis, premature atherosclerosis, vasculitis, valve disease, and valve ring nodules
Lung	Pleural effusions, interstitial lung disease, bronchiolitis obliterans, rheumatoid nodules, vasculitis
Eye	Keratoconjunctivitis sicca, episcleritis, scleritis, scleromalacia perforans, peripheral ulcerative keratopathy
Neurologic	Entrapment neuropathy, cervical myelopathy, mononeuritis multiplex (vasculitis), peripheral neuropathy
Hematopoietic	Anemia, thrombocytosis, lymphadenopathy, Felty's syndrome
Kidney	Amyloidosis, vasculitis
Bone	Osteopenia

Cardiovascular Involvement

Cardiac involvement directly related to RA is uncommon; however, patients with RA have a significantly increased morbidity and mortality from coronary artery disease and congestive heart failure. A meta-analysis of observational studies has shown that the risk for incident cardiovascular disease is increased by 48% in patients with RA compared with that in the general population. The reasons are not clear, but chronic inflammation appears to be the major cause. Some of the medications used to treat RA and a sedentary lifestyle may be additional risk factors for the development of coronary artery disease. Pericardial effusions are common in RA (50% by echocardiography) but usually are asymptomatic. Rarely, long-standing pericardial disease may result in a fibrinous pericarditis, and patients may present clinically with constrictive pericarditis (Chapter 77). A population-based inception cohort of patients with RA in Olmstead County, Minnesota, has shown an increased incidence of venous thromboembolism compared with subjects without RA.

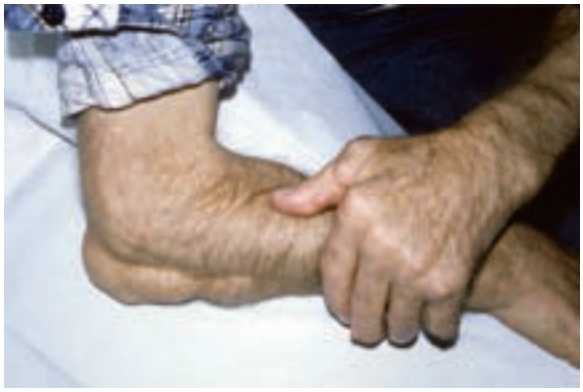

FIGURE 264-7. Rheumatoid nodules. Large rheumatoid nodules are seen in a classic location along the extensor surface of the forearm and in the olecranon bursa.

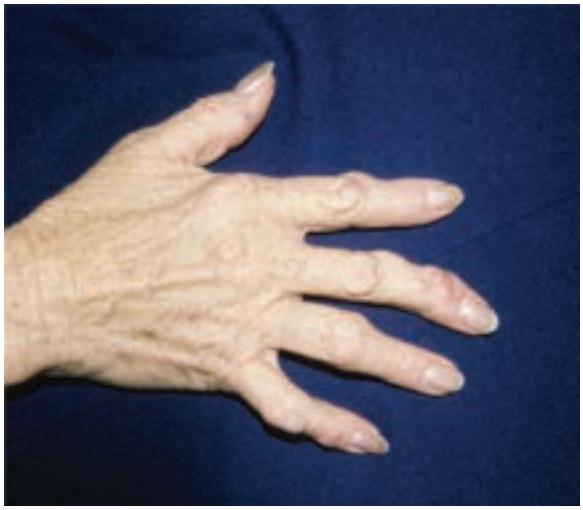

FIGURE 264-8. Rheumatoid nodulosis. In this patient, multiple rheumatoid nodules are present over joints. In some cases, nodules may dominate the clinical picture. Rarely, this may be seen as a side effect of methotrexate therapy.

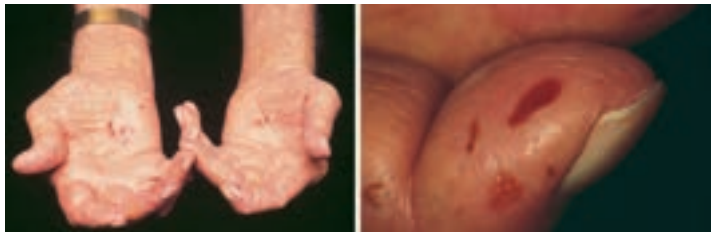

FIGURE 264-9. Small vessel vasculitis. A and B, Rheumatoid vasculitis with small brown infarcts of palms and fingers in chronic rheumatoid arthritis. (Courtesy Dr. Martin Lidsky, Houston, TX.)

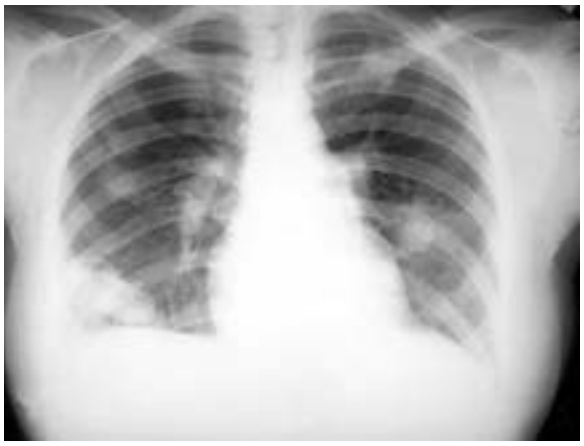

FIGURE 264-10. Rheumatoid nodules in the lung. Chest radiograph demonstrates discrete rheumatoid nodules in both right and left lower lobes. (Courtesy Dr. Martin Lidsky, Houston, TX.)

Pulmonary Manifestations

Pulmonary manifestations of RA include pleural effusions, rheumatoid nodules, and parenchymal lung disease (Chapters 84 and 92). Pleural effusions occur more commonly in men and are usually small and asymptomatic. Of interest, pleural fluid in RA is characterized by low levels of glucose and low pH and, therefore, may at times be confused with empyema. Rheumatoid nodules may occur in the lung, especially in men (Fig. 264-10); these are usually solid but may calcify, cavitate, or become infected. Rarely, pulmonary nodules rupture and produce a pneumothorax. If patients with RA are exposed to coal or silica dust, diffuse nodular densities may occur (Caplan's syndrome). Differentiating rheumatoid nodules from lung cancer can be problematic, particularly if the lesion is solitary. Therefore, the presence of pulmonary nodules in a patient with RA should precipitate an aggressive diagnostic evaluation.

Diffuse interstitial fibrosis occurs in RA and may progress to a honeycomb appearance on radiography with increasing dyspnea. Rarely, bronchiolitis obliterans can be seen, with or without organizing pneumonia.

Ophthalmologic Manifestations

The most common manifestation of RA in the eye is keratoconjunctivitis sicca (dry eyes) from secondary Sjögren's syndrome (Chapter 268). Patients may have associated xerostomia (dry mouth), parotid gland swelling, or, occasionally, lymphadenopathy. Scleritis also can occur and may be painful, with progression to thinning of the sclera, with deep pigment showing through on physical examination, and may progress to perforation (scleromalacia perforans). Rarely, tendonitis of the superior oblique muscles can result in double vision (Brown's syndrome).

Neurologic Manifestations

Peripheral nerve entrapment syndromes, including carpal tunnel syndrome (median nerve at the wrist), and tarsal tunnel syndrome (anterior tibial nerve at the ankle), are common in RA. Vasculitis can lead to a stocking and glove neuropathy or mononeuritis multiplex, both of which may require aggressive therapy. Subluxations at C1-C2 may produce myelopathy (see E-Fig. 264-1). Rheumatoid nodules in the central nervous system have been described but are rare and usually asymptomatic.

Felty's Syndrome

Felty's syndrome is the triad of RA, splenomegaly, and neutropenia. This complication is seen in patients with severe, RF/ACPA-positive disease and may be accompanied by hepatomegaly, thrombocytopenia, lymphadenopathy, and fevers. Most patients with Felty's syndrome do not require special therapy; instead, treatment should be directed toward their severe RA. If severe neutropenia (Chapter 167) exists (<500 cells/μL) and is accompanied by recurrent bacterial infections or chronic, nonhealing leg ulcers, splenectomy may rarely be indicated.

Some patients with RA, who were previously thought to have Felty's syndrome, have peripheral white blood cell counts dominated by large granular lymphocytes with almost complete absence of neutrophils. This condition is known as the *large granular lymphocyte syndrome* and is thought to be a variant of T-cell leukemia. In the setting of RA, this syndrome has a good prognosis, with the neutropenia often responding dramatically to methotrexate therapy.

TABLE 264-2 CLASSIFICATION CRITERIA FOR RHEUMATOID ARTHRITIS*

Morning stiffness (≥1 hr)
Swelling (soft tissue) of three or more joints
Swelling (soft tissue) of hand joints (PIP, MCP, or wrist)
Symmetrical swelling (soft tissue)
Subcutaneous nodules
Serum rheumatoid factor
Erosions and/or periarticular osteopenia in hand or wrist joints seen on radiograph

*Criteria 1 through 4 must have been continuously present for 6 wk or longer, and criteria 2 through 5 must be observed by a physician. A classification of rheumatoid arthritis requires that 4 of the 7 criteria be fulfilled.
MCP = metacarpophalangeal; PIP = proximal interphalangeal.

Clinical Course

Although the presentation is variable, most patients with RA have insidious onset of pain, stiffness, or swelling in multiple small joints over the course of weeks to months. Systemic features such as fatigue, low-grade fevers, and weight loss also may be present. Less commonly, the onset can be fulminant, occurring almost overnight, or patients may exhibit persistent monoarthritis or oligoarthritis for prolonged periods before manifesting the more typical pattern of joint involvement. Rarely, particularly men, develop extra-articular features of RA before the joint problems appear.

The distribution of involved joints is a critical clue to the underlying diagnosis. The joints that are involved in RA at presentation are variable; typically, the symptoms start in the small joints of the hands (PIP and MCP joints) and in the toes (MTP joints). Importantly, RA usually spares the DIP joints and the small joints of the toes (see Fig. 264-4). Later, RA moves, or some would say "metastasizes," to larger joints: wrists, knees, elbows, ankles, hips, and shoulders (roughly in that order). Although the patient's history of joint symptoms (arthralgia) is important, the diagnosis of RA requires the presence of inflammation (swelling, warmth, or both) on examination of the joints.

Morning stiffness is a hallmark of inflammatory arthritis and is a prominent feature of RA. Patients with RA are characteristically at their worst in the morning or after prolonged periods of rest. This stiffness in and around joints often lasts for hours, and quantifying it is one way to measure improvement. Stiffness is relieved by warmth and activity, and reducing or eliminating joint stiffness is a clear goal of therapy.

DIAGNOSIS

All current treatment paradigms for RA stress the early and aggressive use of DMARDs. Therefore, the importance of accurate early diagnosis of RA cannot be overemphasized. There is no one single finding on physical examination or laboratory testing that is pathognomonic of RA. Instead, the diagnosis of RA requires a collection of historical and physical features, as well as an alert and informed clinician.

Classification

There are currently two classification systems for RA: one designed for clinical use and one designed for studies. Table 264-2 lists the current clinical classification for RA; although designed for classification, these criteria are widely used as a diagnostic aid. The first five criteria are all clinical; in other words, they are established by physical examination or by talking with the patient. Only the last two criteria require laboratory tests or radiographs. The first four criteria must be present for at least 6 weeks before a diagnosis of RA should be made. This caveat is important, because a host of conditions, including many virus-related syndromes, can cause self-limited polyarthritis syndromes that look identical to RA, including at times the presence of RF. Such conditions usually last only 2 to 3 weeks. New classification criteria for RA have been developed for use in clinical trials and, although less specific, improve early classification.[5] These American College of Rheumatology/European League Against Rheumatism criteria do not require 6 weeks of disease and give significant weight to the presence of high-titer RF or ACPA positivity. The presence of ACPA antibodies, even in the first few weeks of an inflammatory arthritis, is strongly suggestive of ongoing aggressive RA.[6] The 2010 American College of Rheumatology/European League Against

TABLE 264-3 DIFFERENTIAL DIAGNOSIS OF RHEUMATOID ARTHRITIS

DISORDER	SUBCUTANEOUS NODULES	RHEUMATOID FACTOR
Viral arthritis (hepatitis B and C, parvovirus, rubella, others)	–	±
Bacterial endocarditis	±	+
Rheumatic fever	+	–
Sarcoidosis	+	+
Reactive arthritis	–	–
Psoriatic arthritis	–	–
Systemic lupus erythematosus	±	+
Primary Sjögren's syndrome	–	+
Chronic tophus gout	+	–
Calcium pyrophosphate disease	–	–
Polymyalgia rheumatica	–	–
Osteoarthritis (erosive)	–	–

– = Not present; + = frequently present; ± = occasionally present.

TABLE 264-4 KEYS TO OPTIMIZE OUTCOME OF TREATMENT OF RHEUMATOID ARTHRITIS

Early, accurate diagnosis
Early DMARD therapy
Strive for remission in all patients
Monitor carefully for treatment toxicities
Consider and treat comorbid conditions*

*Important comorbid conditions include cardiovascular disease, increased susceptibility to infections, and osteoporosis.
DMARD = disease-modifying antirheumatic drug.

Rheumatism collaborative initiative classification criteria for RA are shown in E-Figure 264-3.

Laboratory Findings

Historically, the most characteristic laboratory abnormality in RA is the presence of RF, which is found in approximately 80% of patients. RF was first described in the 1930s and is an antibody that recognizes the Fc portion of immunoglobulin G as its antigen. The presence of RF is strongly associated with more severe articular disease, as well as with essentially all the extra-articular features previously discussed. Importantly, RF is seen in association with many diseases other than RA, particularly in disease processes that provide chronic stimulation of the immune system (Table 264-3). ACPA antibodies, found in approximately 75% of patients with RA, have a high specificity (93 to 98%), are often present before clinical disease is diagnosed, and are associated with aggressive erosive disease. Approximately 15% of RA patients are negative for both RF and ACPA (seronegative). RA is associated with many other autoantibodies, including antinuclear antibodies (~30%) and antineutrophil cytoplasmic antibodies, particularly of the perinuclear type (~30%) (Chapter 257).

Most patients with RA have an anemia of chronic disease, and the degree is proportional to the activity of the disease. Therapy that controls the disease will normalize the hemoglobin levels. Other causes of anemia should also be considered in RA, particularly iron deficiency anemia from gastrointestinal blood loss. Thrombocytosis is common, with platelet counts returning to normal as the inflammation is controlled. Acute phase reactants such as ESR and CRP levels parallel the activity of the disease, and their persistent elevation portends a poor prognosis in terms of both joint destruction and mortality. White blood cell counts may be elevated, normal, or, in the case of Felty's syndrome, profoundly depressed. Eosinophilia is present in some patients with RA.

Synovial fluid in RA is characterized by white blood cell counts in the range of 5000 to 100,000/mm^3, with approximately two thirds of the cells being polymorphonuclear leukocytes. There are no synovial fluid findings that are pathognomonic of RA.

Differential Diagnosis

The accurate diagnosis of RA early in its course, although challenging, is critical if patients are to benefit maximally from therapeutic intervention. Once disease has been present and active for years and the characteristic deformities and radiographic changes have occurred, the diagnosis is all too obvious. Once RA has progressed to that point, deformities may no longer be amenable to medical therapy.

Many diseases can mimic RA (see Table 264-3). Early in the course of disease, self-limited viral syndromes need to be considered, especially hepatitis B and C, parvovirus, rubella (infection or vaccination), and Epstein-Barr virus. At any time, systemic lupus erythematosus, psoriatic arthritis, and reactive arthritis may present differential diagnostic challenges. In the case of these three mimics, a targeted history and examination to elucidate their associated clinical features, such as rashes, oral ulcers, nail changes, dactylitis, urethritis, and renal, pulmonary, gastrointestinal, or ophthalmologic involvement, is critical. Especially in elderly patients with fulminant-onset RA, remitting RF-negative symmetrical synovitis with pitting edema (the so-called RS3PE syndrome) and paraneoplastic syndromes should be considered. Chronic tophaceous gout also may mimic severe nodular RA. Hypothyroidism not only causes many rheumatic manifestations but also occurs commonly in conjunction with RA and, therefore, should be kept in mind.

TREATMENT

General Measures

RA is a lifelong disease process that has no known cure; the diagnosis is made based on clinical criteria, and many different options exist for treatment. These factors magnify the importance of the patient-physician relationship and place a premium on the art rather than the science of medicine. Optimal care for patients with RA requires effective ongoing interactions between primary care physicians and rheumatologists, and, in some cases, physical therapists, occupational therapists, and orthopedic surgeons.[7] Because of the serious nature of the disease, the rapid introduction of new treatments, and the need for expertise in monitoring these therapies, all patients with RA should be evaluated early and followed closely by a rheumatologist.

The goal of therapy is disease remission (Table 264-4) or very low disease activity.[8] When RA is treated early, remission is possible in over 50% of patients. However, remissions require the ongoing use of DMARDs and are not always durable. Essentially all patients with RA should be treated with DMARDs.[9,10] Some combination of nonsteroidal anti-inflammatory drugs (NSAIDs), steroids, and DMARDs is necessary in many patients. In many patients with RA, combinations of different DMARDs (conventional and biologic) are necessary for optimal control. Therapy should be escalated rapidly to ensure maximal suppression of disease while minimizing toxicity and expense. Patients with RA should be educated about their disease and its treatment. Patients should have an opportunity to spend time with physical therapists and occupational therapists to learn about range-of-motion exercises, joint protection, and assistive devices.

Medical Therapy

In the treatment of RA, three types of medical therapies are used: NSAIDs, glucocorticoids, and DMARDs (both conventional and biologic). Initial therapy should always include a DMARD.

Nonsteroidal Anti-inflammatory Drugs

NSAIDs are important for the symptomatic relief they provide to patients with RA; however, they play only a minor role in altering the underlying disease process (Chapter 37). Therefore, NSAIDs should rarely, if ever, be used to treat RA without the concomitant use of DMARDs. Many clinicians waste valuable time switching from one NSAID to another before starting DMARD therapy.

Much has been written about the gastrointestinal toxicity of NSAIDs, and these concerns are particularly relevant to patients with RA, who often have significant risk factors, including age and concomitant steroid use. Therefore, cyclooxygenase-2 (COX2)-selective agents have been a popular choice for patients with RA. The evidence linking these agents to increased cardiovascular toxicity has been particularly troubling for patients with RA, who are already at high risk for myocardial infarction. Therefore, if COX2-selective agents are used, they should be kept at a low dose. Consideration should be given to low-dose aspirin prophylaxis in RA, but this may increase the gastrointestinal toxicity of NSAIDs. The use of concomitant misoprostol or proton pump inhibitors should be considered in all patients with RA who are taking NSAIDs. Additionally, the potential for NSAIDs to decrease renal blood flow and to increase blood pressure should be kept in mind.

Glucocorticoids

Glucocorticoids have had a significant role in the treatment of RA for more than half a century (Chapter 35). Indeed, RA was chosen as the first disease to be treated with this new therapy, partly because it was thought that RA was

ACR / EULAR RA Classification Criteria

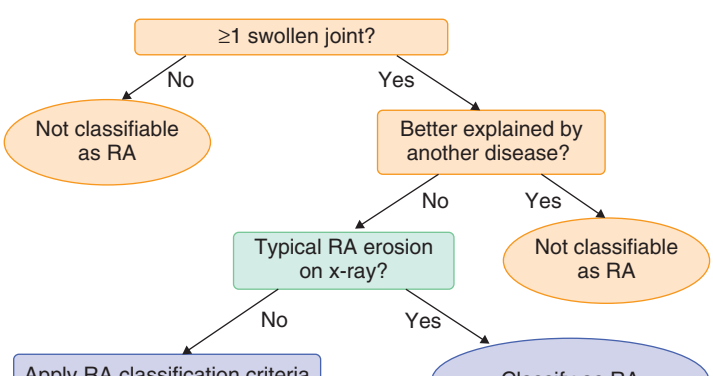

Joint Involvement (0-5)	
1 medium-large joint	0
2-10 medium-large joints	1
1-3 small joints	2
4-10 small joints	3
>10 joints (at least one small joint)	5
Serology (0-3)	
Neither RF or ACPA positive	0
At least one test low positive titre	2
At least one test high positive titre	3
Duration of Synovitis (0-1)	
<6 weeks	0
≥6 weeks	1
Acute Phase Reactants (0-1)	
Neither CRP or ESR abnormal	0
Abnormal CRP or abnormal ESR	1

Score ≥ 6 indicates 'Definite RA'

Low +: >1 & ≤3 × ULN
High +: >3 × ULN

Medium/Large joints
Knee, hip, shoulders, elbow and ankle

Small joints
PIPs, MCPs, wrists and MTPs

E-FIGURE 264-3. American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) classification criteria for rheumatoid arthritis (RA). The classification criteria are shown as a flow diagram for initial decision making and also in tabular American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) classification criteria for rheumatoid arthritis (RA). Form to calculate. ACPA = anti-citrullinated protein antibody; CRP = C-reactive protein; ESR = erythrocyte sedimentation rate; MCP = metacarpal phalangeal joint; MTP = metatarsal phalangeal joint; PIP = proximal interphalangeal joint; RA = rheumatoid arthritis; RF = rheumatoid factor; ULN = upper limit of normal. (From Aletaha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria. An American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum.* 2010;62:2569-2581.

TABLE 264-5 GUIDELINES FOR USE OF GLUCOCORTICOIDS

Avoid use of glucocorticoids without DMARDs
Prednisone, >10 mg/day, is rarely indicated for articular disease
Taper to the lowest effective dose
Use as "bridge therapy" until DMARD therapy is effective
Remember prophylaxis against osteoporosis

DMARD = disease-modifying antirheumatic drug.

a disease of glucocorticoid deficiency (an issue that remains unresolved). As was the case with the first patient treated in 1948, glucocorticoids are dramatically and rapidly effective in patients with RA. Not only are glucocorticoids useful for symptomatic improvement but they also significantly decrease the radiographic progression of RA. However, the toxicities of long-term therapy are extensive and potentially devastating. Therefore, the optimal use of these drugs requires an understanding of several principles (Table 264-5).

Glucocorticoids remain among the most potent anti-inflammatory treatments available; for this reason, and because of their rapid onset of action, they are ideally suited to help control the inflammation in RA while the much slower-acting DMARDs are starting to work. Prednisone, the most commonly used glucocorticoid, should rarely be used in doses higher than 10 mg/day to treat the stiffness and articular manifestations of RA. At this dose at the start of methotrexate-based treatment, the addition of prednisone reduces erosive joint damage, disease activity, physical disability, and the use of biologic treatment at 2 years. The dose should be slowly tapered to the lowest effective dose, and the concomitant DMARD therapy should be adjusted to make this possible. Glucocorticoids should rarely, if ever, be used to treat RA without concomitant DMARD therapy. The paradigm is to shut off inflammation rapidly with glucocorticoids and then to taper them as the DMARD is taking effect ("bridge therapy"). In all patients receiving glucocorticoids, strong measures should be taken to prevent osteoporosis. Bisphosphonates have been shown to be particularly effective in this regard but are contraindicated in women of childbearing age. Higher doses of glucocorticoids may be necessary to treat extra-articular manifestations, especially vasculitis and scleritis.

Disease-Modifying Antirheumatic Drugs

DMARDs are a group of medications that have the ability to halt the disease process in the synovium and to modify or change the disabling potential of RA. These drugs have the ability to halt or slow the radiographic progression of RA.

Conventional Disease-Modifying Antirheumatic Drugs

Included in this group of medications are methotrexate, sulfasalazine (Azulfidine), gold, antimalarials (hydroxychloroquine [Plaquenil] and others), leflunomide (Arava), azathioprine (Imuran), minocycline, and the newly approved tofacitinib (Xeljanz). It is critically important that clinicians and patients understand that conventional DMARDs take 2 to 6 months to exert their maximal effect, and all require some monitoring (Table 264-6). Therefore, other measures, such as glucocorticoid therapy, may be needed to control the disease while these medications are starting to work.

These DMARDs have been shown to be effective in treating both early and more advanced RA. Until additional research elucidates factors that allow selection of the best initial therapy for each patient, the choice will depend on patient and physician concerns about toxicity and monitoring issues, as well as the activity of disease and presence of comorbid conditions. The critical issue is not which DMARD to start first but rather getting the DMARD therapy started early in the disease process.

Methotrexate

Methotrexate is the preferred initial DMARD of most rheumatologists, in part because patients have a more durable response, and because, with correct monitoring, serious toxicities are rare. Methotrexate is dramatically effective in slowing radiographic progression and is usually given orally in doses ranging from 5 to 30 mg/week as a single dose. This once-per-week administration is worthy of emphasis; prior experience with daily therapy in psoriasis has demonstrated the importance of allowing the liver time to recover between doses. Oral absorption of methotrexate is variable; subcutaneous injections of methotrexate may be effective if oral treatment is not. Side effects of methotrexate include oral ulcers, nausea, hepatotoxicity, bone marrow suppression, and pneumonitis. With the exception of pneumonitis, these toxicities respond to dose adjustments. Monitoring of blood counts and liver blood tests (albumin and aspartate aminotransaminase [SGOT] or alanine aminotransferase [SGPT]) should be done every 4 to 8 weeks initially and, when stable, every 3 months thereafter, with adjustments in the dose of methotrexate as needed. Renal function is critical for clearance of methotrexate; previously stable patients may experience severe toxicities if renal function deteriorates. Pneumonitis, although rare, is less predictable and can be fatal, particularly if the methotrexate is not stopped or is restarted. Folic acid, 1 to 4 mg/day, can significantly decrease most methotrexate toxicities without interfering with efficacy. If methotrexate alone does not sufficiently control

TABLE 264-6 CAVEATS FOR MONITORING DISEASE-MODIFYING ANTIRHEUMATIC DRUG THERAPIES*

MEDICATION	CAVEATS
Prednisone	Use as bridge to effective DMARD therapy. Prophylaxis for osteoporosis? (see Table 264-5)
Hydroxychloroquine	Keep dosage lower than 6.5 mg/kg/day. Yearly eye checkup by ophthalmologist
Sulfasalazine	CBC for neutropenia, initially every month, then every 6 mo
Methotrexate	CBC and SGOT/SGPT every 8-12 wk when dose is stable. Many toxicities respond to folic acid or small dose reduction. If pneumonitis, stop and do not restart. Decreasing renal function may precipitate toxicities. Absolute contraindication in pregnancy
Leflunomide	CBC and SGOT/SGPT every 4-8 wk; long half-life may require cholestyramine washout; absolute contraindication in pregnancy
TNF inhibitors	If fevers or infectious symptoms of any kind, stop until symptoms resolve; aggressively work up and treat possible infections. May precipitate congestive heart failure, demyelinating syndromes, or lupus-like syndromes

*Patients receiving DMARDs, both conventional and biologic, should be monitored by a rheumatologist.
CBC = compete blood count; DMARD = disease-modifying antirheumatic drug; SGOT = serum glutamate oxaloacetate transaminase (aspartate aminotransferase); SGPT = serum glutamate pyruvate transaminase (alanine aminotransferase); TNF = tumor necrosis factor.

disease, it is combined with other DMARDs. Methotrexate in combination with virtually any of the other DMARDs (conventional or biologic) has been shown to be more effective than either drug alone.

Leflunomide

Leflunomide, a pyrimidine antagonist, has a very long half-life and is most commonly started at 10 to 20 mg/day orally. Diarrhea is the most common toxicity and responds to dose reduction, and doses of leflunomide of 10 to 20 mg three to five times per week are frequently used. Also, because of the long half-life and teratogenic potential of leflunomide, women wishing to become pregnant who have previously received leflunomide, even if therapy was stopped years ago, should have blood levels drawn. If toxicity occurs or if pregnancy is being considered, leflunomide can be rapidly eliminated from the body by treatment with cholestyramine. Laboratory monitoring for hematologic and hepatic toxicity should be done during treatment with leflunomide, as recommended for methotrexate.

Antimalarial Drugs

The antimalarial drugs hydroxychloroquine (Plaquenil) and chloroquine are frequently used for the treatment of RA. They have the least toxicity of any of the DMARDs and do not require monitoring of blood tests. Yearly monitoring by an ophthalmologist after 5 years of therapy is recommended to detect any signs of retinal toxicity (rare). Hydroxychloroquine is the most commonly used preparation and is given orally at 200 to 400 mg/day. These drugs are frequently used in combination with other DMARDs, particularly methotrexate. Hydroxychloroquine decreases cholesterol levels and has recently been shown to decrease the incidence of diabetes in patients with RA.

Sulfasalazine

Sulfasalazine is an effective treatment when given in doses of 1 to 3 g/day. Monitoring of blood counts, particularly white blood cell counts, in the first 6 months is recommended. Sulfasalazine and hydroxychloroquine are often combined with methotrexate, a regimen referred to as triple therapy.

Minocycline

Minocycline 100 mg twice daily has been shown to be an effective treatment for RA, particularly when used in early, RF-positive disease. Chronic therapy (>2 years) with minocycline may lead to cutaneous hyperpigmentation. Minocycline has been associated with drug-induced lupus.

Gold

Gold, the oldest DMARD, is effective but cumbersome and is currently rarely used.

Tofacitinib

Tofacitinib (Xeljanz) has been recently approved for the treatment of RA in the United States. It is the first Jak kinase inhibitor to be approved for RA. It is given orally at a dose of 5 mg twice daily, and complete blood count and liver function tests should be monitored. Additional toxicity concerns include infections, including tuberculosis, and malignancies. Tofactinib has been

shown to be effective as initial DMARD therapy,[A] when combined with methotrexate in patients who have had incomplete responses to methotrexate, and in patients who have failed TNF inhibitors. Tofacitinib has not yet been approved in Europe.

Biologic Disease-Modifying Antirheumatic Drugs

Recent research has continued to elucidate the central role that cytokines, most notably TNF-α, IL-1 and IL-6, play in the pathophysiology of RA (Chapter 36). This led directly to the development and clinical use of biologic agents directed against TNF-α[11] (etanercept [Enbrel], infliximab [Remicade], adalimumab [Humira], golimumab [Simponi], and certolizumab [Cimzia]), IL-1 (anakinra [Kineret]), and IL-6 (tocilizumab [Actemra). Additionally, monoclonal antibodies that deplete B cells (anti-CD20, rituximab [Rituxan]) and that block the second signal for T cell activation (abatacept [Orencia]) are effective agents in the treatment of RA. All patients with RA receiving biologic therapies should be monitored by a rheumatologist, and their physicians should be aware of the risk for infections that are often atypical.[12] All the biologics, when combined with methotrexate, have been shown to decrease disease activity and slow radiographic progression in patients with RA with active disease despite methotrexate.[A] Early treatment with abatacept plus methotrexate was shown to result in greater sustainable clinical, functional, and radiographic benefits than methotrexate alone, with acceptable safety and tolerability, in early erosive RA. Currently, biologic agents should not be used in combination with each other because all studies to date have shown a significant increase in infections. See Chapter 36 for further details on the use of biologic agents in the treatment of RA.

The Order of Therapy in Rheumatoid Arthritis

Several randomized double-blind trials have elucidated the order of therapy in RA. The Treatment of Early Aggressive Rheumatoid Arthritis (TEAR) trial nicely showed that initial therapy with methotrexate in patients with poor-prognosis RA was not inferior at 2 years to initial combinations of either conventional DMARDs or the combination of methotrexate and etanercept.[A] The Rheumatoid Arthritis: Comparison of Active Therapies (RACAT) trial has also shown that in those patients who are not controlled on methotrexate alone, the strategy of initially adding sulfasalazine and hydroxychloroquine to methotrexate (triple therapy) was not inferior to the addition of etanercept to methotrexate.[A] Therefore, because of the huge economic advantages, the typical RA patient should be started on methotrexate monotherapy, and, if not controlled after 3 months on maximum methotrexate, advanced to triple therapy. If the patient does not achieve adequate control after 3 to 6 months on triple therapy, either a TNF inhibitor or abatacept should be added to methotrexate.

Treatment of Underlying Conditions

Optimal care of patients with RA requires recognition of the associated comorbid conditions, including an increased risk for cardiovascular death, osteoporosis, infections (especially pneumonia), and certain cancers.

Cardiovascular Disease

Increasingly, cardiovascular disease is being recognized as the cause of much of the excess mortality in RA. Various factors contribute to this mortality, including sedentary lifestyle and glucocorticoid therapy. However, a strong association between chronic inflammation and cardiovascular disease has been identified, and it is likely that this may be the most significant factor. Therapies that control RA earlier and better can be expected to decrease cardiovascular morbidity and mortality. Both methotrexate and TNF inhibitors have been shown to decrease cardiovascular mortality in RA. Clinicians should consider RA a risk factor for cardiovascular disease and should aggressively address other cardiovascular risk factors (Chapter 52) in their patients with RA.[13]

Other Associated Diseases

Osteoporosis is common in patients with RA, and early treatment results in long-term dividends. Patients with RA are at an increased risk for infections, and some forms of treatment further increase this risk. Patients should be cautioned to seek medical attention early for even minor symptoms suggestive of infection, especially if receiving biologic therapy. All patients with RA should receive a pneumococcal vaccine at appropriate intervals and yearly influenza vaccinations. Finally, patients with RA have an increased risk for lymphoma. Occasionally, B-cell lymphomas are associated with immunosuppression and regress after immunosuppression is discontinued. Patients with RA have significantly decreased risk of developing colon cancer. This is thought to be secondary to chronic inhibition of COX by NSAIDs.

PROGNOSIS

RA is not a benign disease and is not limited to the joints. Once established, RA is a lifelong progressive disease that produces significant morbidity in most patients and premature mortality in many. Long-term studies have found that 50% of patients with RA have had to stop working after 10 years (~10 times the average rate). Patients who are RF or ACPA positive and those who are positive for the shared epitope have a worse prognosis, with more erosions and more extra-articular disease (see Table 264-1). Once deformities are found on examination or erosions on radiography, the damage is largely irreversible. It has been clearly shown that erosions occur in most patients in the first 1 to 2 years and that the rate of radiographic damage can be affected by early therapy. A recent cohort study has shown that the aggressive use of disease-modifying agents and the introduction of biologic agents have been associated with substantial reductions in disability. Therefore, early DMARD therapy is critical. Although limited long-term data are available, the current information strongly suggests that patients have the opportunity to benefit greatly if the newer principles of therapy are practiced.

FUTURE DIRECTIONS

Significant advances in the effective treatment of RA have come from an understanding of the cytokine imbalance that accompanies this disease. Much research is focused on the further development of biologic products to modulate this balance. There remains a critical need for a cytokine thermostat that would allow titration of the desired cytokine balance to control disease without altering critical immune functions.

Even with existing therapies, there are many different effective options for patients with RA. The challenge for the clinician is to pick the right option for each patient. Few data are currently available to aid in this choice, and the establishment of parameters, genetic or otherwise, that would allow selection of the best initial option for each patient would be a major breakthrough. Finally, elucidation of the trigger or triggers for RA may allow the development of strategies to prevent the onset of clinical disease.

Grade A References

A1. Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, et al. Comparison of treatment strategies in early rheumatoid arthritis: a randomized trial. *Ann Intern Med*. 2007;146:406-415.

A2. Bakker MF, Jacobs JW, Welsing PM, et al. Low-dose prednisone inclusion in a methotrexate-based, tight control strategy for early rheumatoid arthritis: a randomized trial. *Ann Intern Med*. 2012;156:329-339.

A3. Lopez-Olivo MA, Siddhanamatha HR, Shea B, et al. Methotrexate for treating rheumatoid arthritis. *Cochrane Database Syst Rev*. 2014;6:CD000957.

A4. Moreland LW, O'Dell JR, Paulus HE, et al. TEAR Investigators. A randomized comparative effectiveness study of oral triple therapy versus etanercept plus methotrexate in early aggressive rheumatoid arthritis. *Arthritis Rheum*. 2012;64:2824-2835.

A5. Lee EB, Fleischmann R, Hall S, et al. Tofacitinib versus methotrexate in rheumatoid arthritis. *N Engl J Med*. 2014;370:2377-2386.

A6. Nam JL, Ramiro S, Gaujoux-Viala C, et al. Efficacy of biological disease-modifying antirheumatic drugs: a systematic literature review informing the 2013 update of the EULAR recommendations for the management of rheumatoid arthritis. *Ann Rheum Dis*. 2014;73:516-528.

A7. O'Dell JR, Curtis JR, Mikuls TR, et al. Validation of the methotrexate-first strategy in patients with early, poor-prognosis rheumatoid arthritis: results from a two-year randomized, double-blind trial. *Arthritis Rheum*. 2013;65:1985-1994.

A8. O'Dell JR, Mikuls TR, Taylor TH, et al. Therapies for active rheumatoid arthritis after methotrexate failure. *N Engl J Med*. 2013;369:307-318.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Avina-Zubieta JA, Thomas J, Sadatsafavi M, et al. Risk of incident cardiovascular events in patients with rheumatoid arthritis: a meta-analysis of observational studies. *Ann Rheum Dis*. 2012;71:1524-1529.
2. Yarwood A, Huizinga TW, Worthington J. The genetics of rheumatoid arthritis: risk and protection in different stages of the evolution of RA. *Rheumatology (Oxford)*. 2014;[Epub ahead of print].
3. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. *N Engl J Med*. 2011;365:2205-2219.
4. Makol A, Matteson EL, Warrington KJ. Rheumatoid vasculitis: an update. *Curr Opin Rheumatol*. 2015;27:63-70.
5. Mjaavatten MD, Bykerk VP. Early rheumatoid arthritis: the performance of the 2010 ACR/EULAR criteria for diagnosing RA. *Best Pract Res Clin Rheumatol*. 2013;27:451-466.
6. Whiting PF, Smidt N, Sterne JA, et al. Systematic review: accuracy of anti-citrullinated peptide antibodies for diagnosing rheumatoid arthritis. *Ann Intern Med*. 2010;152:456-464.
7. Vliet Vlieland TP, van den Ende CH. Nonpharmacological treatment of rheumatoid arthritis. *Curr Opin Rheumatol*. 2011;23:259-264.
8. Gaujoux-Viala C, Gossec L, Cantagrel A, et al. Recommendations of the French Society for Rheumatology for managing rheumatoid arthritis. *Joint Bone Spine*. 2014;81:287-297.
9. Gramling A, O'Dell JR. Initial management of rheumatoid arthritis. *Rheum Dis Clin North Am*. 2012;38:311-325.
10. Singh JA, Furst DE, Bharat A, et al. 2012 Update of the 2008 American College of Rheumatology recommendations for the use of disease-modifying antirheumatic drugs and biologic agents in the treatment of rheumatoid arthritis. *Arthritis Care Res*. 2012;64:625-639.
11. Krishnan E, Lingala B, Bruce B, et al. Disability in rheumatoid arthritis in the era of biological treatments. *Ann Rheum Dis*. 2012;71:213-218.
12. Singh JA, Wells GA, Christensen R, et al. Adverse effects of biologics: a network meta-analysis and Cochrane overview. *Cochrane Database Syst Rev*. 2011;2:CD008794.
13. Martin-Martinez MA, Gonzalez-Juanatey C, Castaneda S, et al. Recommendations for the management of cardiovascular risk in patients with rheumatoid arthritis: scientific evidence and expert opinion. *Semin Arthritis Rheum*. 2014;44:1-8.

REVIEW QUESTIONS

1. A 32-year-old woman presents with 3 weeks of symmetrical swelling and stiffness in her PIPs and MCPs. The most helpful laboratory test result in establishing a diagnosis of RA in this patient would be:

A. ESR of 65 mm/hour.
B. RF of 120 IU (normal <40).
C. Anti-CCP of 87 (normal <20).
D. Positive ANA.

Answer: C The presence of anti-CCP antibodies in a patient with an inflammatory arthritis is 99% specific for the diagnosis of RA even when it is only of short duration as in this patient with only 3 weeks of disease. This is critically important because it allows very early institution of DMARD therapy. The presence of a positive ANA, an elevated RF, or elevated ESR would not be particularly helpful—all are nonspecific.

2. A 40-year-old attorney is diagnosed with RF- and CCP-positive rheumatoid arthritis. His mortality is increased, and he is most likely to die from which of the following?

A. Lymphoma
B. Pulmonary fibrosis from RA
C. Infections related to RA and its therapies
D. Coronary artery disease

Answer: D Heart disease is by far the biggest killer of patients with RA. Rheumatologists should do everything possible to get disease activity under control, and PCPs should do everything possible to control other risk factors, including ideal management of BP and cholesterol. RA patients have at least a two-fold increased risk for lymphoma, have increased risk for infections (particularly when treated with steroids or biologic agents), and can get pulmonary fibrosis, but these all have a small impact compared to that of heart disease.

3. A 28-year-old accountant with RA is considering another pregnancy. Reasons for concerns for conception and the health of her fetus during the first trimester include which of the following?

A. She is currently on 5 mg of prednisone.
B. She is currently taking sulfasalazine.
C. She has a history of treatment with leflunomide.
D. She is taking Naprosyn 500 mg.

Answer: C Leflunomide is a significant teratogen and has a half-life that is extremely prolonged. Any potential mother who has ever taken it needs blood levels drawn before conception. Leflunomide can be rapidly eliminated from the body by treatment with cholestyramine. Prednisone is considered safe in pregnancy; significant data suggest that sulfasalazine is well tolerated by both mother and child and NSAIDs are safe until the last trimester.

4. A 52-year-old woman with RA treated with methotrexate and adalimumab presents to the ER with fevers, cough, and increasing shortness of breath. A chest x-ray film shows diffuse patchy infiltrates. You would:

A. Admit, stop her adalimumab, and give her broad-spectrum antibiotics.
B. Admit, stop both her adalimumab and methotrexate, and aggressively pursue workup for infectious and inflammatory etiologies.
C. Treat her for community-acquired pneumonia.
D. Admit, stop adalimumab, and aggressively pursue workup for infectious etiology.

Answer: B All TNF inhibitors have a black box warning for fatal opportunistic infections. Tuberculosis, histoplasmosis, coccidioidomycosis, pneumocystis, legionellosis, and others have all been reported. The key to patient survival is aggressive diagnosis so appropriate treatment can be started rapidly. Methotrexate can be associated with unusual infections rarely, but the biggest concern here is that the patient may have methotrexate pneumonitis. This can be fatal if the methotrexate is not stopped. It usually presents with low-grade fevers, nonproductive cough, increasing dyspnea, and patchy infiltrates on chest x-ray films. Treatment of methotrexate pneumonitis after stopping methotrexate is usually supportive, but occasionally high-dose steroids may be indicated.

5. A 55-year-old carpenter is newly diagnosed with RA. His medical history includes type II diabetes for 15 years. CBC shows low-grade anemia with normal red cell indices, Hgb A_{1c} is 7, creatinine is 2.1, and urinalysis shows 3+ proteinuria. Laboratory studies are otherwise unremarkable. Appropriate initial DMARD therapy would be:

A. Cyclosporine.
B. Hydroxychloroquine.
C. Methotrexate.
D. Sulfasalazine.

Answer: B Although methotrexate at the doses used to treat RA rarely affects the kidney, well-functioning kidneys are important for the clearance of methotrexate. Methotrexate, if used, should be used with extreme caution and close monitoring in this situation. Cyclosporine, although effective in some cases of RA, also has renal toxicity and is not a good choice. Hydroxychloroquine is the best choice here—in addition to being effective against RA, it has been shown to lower blood sugar, particularly in type II diabetes. Additionally, patients with RA who are on hydroxychloroquine have a 60% reduction in their risk for incident diabetes. Finally, sulfasalazine would be a possible choice but not as good as hydroxychloroquine.

265

THE SPONDYLOARTHROPATHIES

ROBERT D. INMAN

COMMON FEATURES OF SPONDYLOARTHRITIS

DEFINITION

Spondyloarthritis (SpA) encompasses a group of clinical syndromes that are linked in terms of disease manifestations and genetic susceptibility. The

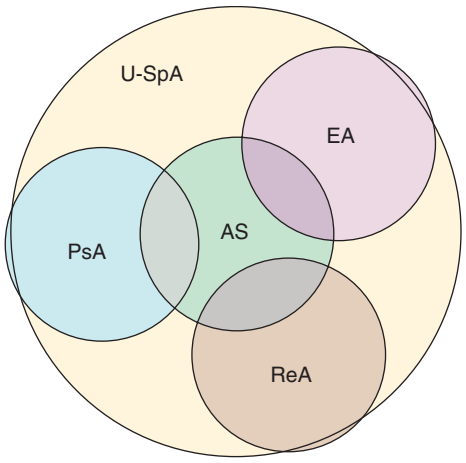

FIGURE 265-1. Schematic relationships among the different spondyloarthritis (SpA) subsets. Ankylosing spondylitis (AS), considered the classic SpA, encompasses the essential features of this family of diseases. AS may overlap with psoriatic arthritis (PsA), enteropathic arthritis (EA), or reactive arthritis (ReA). Many patients have clinical features of SpA that do not meet the diagnostic criteria for any of the four defined subsets. Such cases are termed *undifferentiated SpA* (U-SpA).

TABLE 265-1 ASAS CLASSIFICATION CRITERIA FOR AXIAL SPONDYLOARTHRITIS

Sacroiliitis* *Plus* ≥SpA feature†	OR	HLA-B27 *Plus* ≥2 other SpA features†
*Sacroiliitis (x-rays or MRI)		†SpA features:
• Definite radiographic sacroiliitis according to modified New York criteria (see Table 265-3) *Or* • Active (acute) inflammation on MRI highly suggestive of sacroiliitis associated with SpA		• IBP • Arthritis • Enthesitis (heel) • Dactylitis • Psoriasis • Crohn's disease/ulcerative colitis • Good response to NSAIDs • Family history of SpA • Elevated CRP • HLA-827

ASAS = Assessment in Spondyloarthritis International Society; CRP = C-reactive protein; IBP = inflammatory back pain; MRI = magnetic resonance imaging; NSAIDs = nonsteroidal anti-inflammatory drugs; SpA = spondyloarthritis.

clinical subsets most commonly recognized are ankylosing spondylitis (AS), reactive arthritis (ReA), psoriatic arthritis (PsA), and enteropathic arthritis (EA) (Fig. 265-1). In addition, a sizable number of patients do not fit into one of these distinct diagnostic categories but share some of the common clinical features described in this chapter. This syndrome is termed *undifferentiated spondyloarthritis* (USpA); it may evolve over time into a classic pattern such as AS, or it may retain an undifferentiated pattern in long-term follow-up studies.

PATHOBIOLOGY

Family studies involving multiple individuals with SpA have emphasized some of the common features among the four distinct subsets. The impression from such studies is that there is a shared common path of immunogenetic susceptibility, with further genetic and environmental influences that lead to characteristic clinical subsets. Thus, EA may occur in one such family, but in another family the disease may be PsA. In this sense, the subsets of SpA seem to "breed true." It should be recognized, however, that some distinct clinical features can be very similar in their manifestations (e.g., guttate psoriasis and keratoderma blennorrhagicum), making simple discrimination difficult.

CLINICAL MANIFESTATIONS

There are several common features in the clinical subsets of SpA that serve to both link them with and distinguish them from the other major contributor to chronic polyarthritis—rheumatoid arthritis (RA) (Chapter 264). SpA has a strong predilection for the spine, in particular the sacroiliac joints. There is a shared tendency for new bone formation at sites of chronic inflammation, with joint ankylosis as a consequence. When peripheral arthritis occurs, it is commonly in the lower extremity and asymmetrical. There is a predilection for involvement at sites of tendon insertion into bone (entheses), so enthesitis is one of the most specific clinical manifestations of SpA. Theories postulating the basis for this target organ involvement have invoked biomechanical factors, innervation, local vascularity, and bone marrow–derived inflammatory mediators, but the precise mechanism remains incompletely defined. Whatever the reason, inflammation in the enthesis and contiguous subchondral bone is a characteristic feature of this form of arthritis, and the appearance of this inflammation on magnetic resonance imaging (MRI) is distinct enough to be increasingly used for diagnostic purposes, particularly when the x-rays are not diagnostic.

A predilection for ocular inflammation, particularly acute anterior uveitis, is a common feature of SpA. Indeed, some investigators consider anterior uveitis to be a feature of SpA in its own right because it may occur in the same susceptible population of patients even in the absence of joint involvement, and it may have a unique genetic predisposition. Finally, all SpA subsets have an association with the class I human leukocyte antigen (HLA) allele B27, with the strength of the association varying somewhat among them. Newer genetic risk associations such as *IL23R*, which are shared between SpA, psoriatic arthritis, and inflammatory bowel disease (IBD), further link the clinical subsets of SpA.

DIAGNOSIS

Increasingly, diagnostic criteria (Table 265-1 [Assessment in Spondyloarthritis International Society criteria]) are emphasizing the common clinical features—namely, inflammatory spinal pain or asymmetrical lower extremity synovitis. Several distinct features differentiate SpA from RA, the other main contributor to the differential diagnosis of chronic polyarthritis (Table 265-2). These features include sex predilection, HLA association, pattern of joint involvement, and presence of rheumatoid factor, which becomes the serologic distinction between seropositive disease (RA) and seronegative disease (SpA).

At the level of joint histopathology, sites of chronic inflammation in RA are associated with erosions, but in SpA such sites are associated with new bone formation. This distinction suggests a fundamental difference in the cytokine profile in the microenvironment of the joint, but this issue has not been resolved, and the mediators of neo-ossification await identification. Dysregulation of the wnt/β-catenin pathway may play a key role in the ankylosing process. Synovial histopathology in SpA is characterized by abundant neutrophils, macrophages, and hypervascularity, whereas in RA the prominent features are lymphoid aggregates, dendritic cells, lining cell hyperplasia, and citrullinated proteins. These differences suggest that SpA reflects a fundamental alteration in innate immunity, whereas RA reflects dysregulation of adaptive immunity.

TREATMENT

General Measures

SpA necessitates a global approach to management in which education of patients is the cornerstone.[1-3] Because the typical onset is during young adulthood, these patients may experience significant frustration or depression if their acute arthritis evolves into a chronic disease that significantly impairs their functional capabilities and quality of life. A clinician managing patients with SpA should be aware that these psychosocial aspects are an important part of the burden of illness. Similarly, there may be important implications for the workplace, particularly if a job demands significant bending or twisting. It is important to include the mechanical demands of the workplace in the global assessment of patients with SpA.

Exercise is an important part of the treatment plan for patients with AS.[4] Generally, high-impact sports should be avoided, whereas swimming is an ideal exercise. Stretching to maintain mobility and maintenance of posture should be emphasized, and an experienced physiotherapist can greatly assist in instructing patients in daily exercises. Long car trips and air travel should include periodic stretching. Sleep position should emphasize a straight back position rather than one curled on the side. Deep breathing exercises and avoidance of smoking should be stressed.

One key area of concern for patients is prognosis, because SpAs, particularly ReA, often occur in young, active individuals for whom athletic activity is a

TABLE 265-2 DIFFERENTIAL DIAGNOSIS OF CHRONIC POLYARTHRITIS

FEATURE	RHEUMATOID ARTHRITIS	ANKYLOSING SPONDYLITIS	ENTEROPATHIC ARTHRITIS	PSORIATIC ARTHRITIS	REACTIVE ARTHRITIS
Male-female ratio	1 : 3	3 : 1	1 : 1	1 : 1	10 : 1
HLA association	DR4	B27	B27 (axial)	B27 (axial)	B27
Joint pattern	Symmetrical, peripheral	Axial	Axial and peripheral	Axial and asymmetrical, peripheral	Axial and asymmetrical, peripheral
Sacroiliac	0	Symmetrical	Symmetrical	Asymmetrical	Asymmetrical
Syndesmophyte	0	Smooth, marginal	Smooth, marginal	Coarse, nonmarginal	Coarse, nonmarginal
Eye	Scleritis	Iritis	+/–	0	Iritis and conjunctivitis
Skin	Vasculitis	0	0	Psoriasis	Keratoderma
Rheumatoid factor	>80%	0	0	0	0

HLA = human leukocyte antigen.

priority. There is general recognition that ReA has a greater propensity for chronicity than was previously appreciated, and this should temper an overly optimistic projection of the disease's natural history. At the 5-year follow-up of a cohort of patients with *Salmonella*-induced ReA, two thirds continued to have subjective complaints, and one third demonstrated objective changes in their joints. The variability in prognosis for the large group of patients falling into the USpA group is perplexing. At present, there is a lack of reliable predictors of progression in patients with this heterogeneous cluster of articular and extra-articular features.

Medical Therapy

Nonsteroidal Anti-inflammatory Drugs

In general, joint inflammation in SpA improves significantly after the introduction of nonsteroidal anti-inflammatory drugs (NSAIDs). Indomethacin and diclofenac (up to 150 mg/day in divided doses) or naproxen (up to 1000 mg/day in divided doses) are generally well tolerated in this population. These agents have to be used with caution in EA because of concern about exacerbating possible underlying IBD. In the case of AS, the goal with anti-inflammatory treatment is to achieve sufficient relief of pain and stiffness to allow an active, sustained program of exercise and physical activity that maintains posture and improves quality of life. Some studies have suggested that NSAIDs have disease-modifying capability, but this effect appears to be restricted to those patients with elevated C-reactive protein (CRP) or erythrocyte sedimentation rate (ESR).

Corticosteroids

The response to the intra-articular injection of steroids in the peripheral joints of patients with SpA is often neither as dramatic nor as sustained as in those with RA. Corticosteroid injection into the sacroiliac joints is usually performed under imaging guidance (fluoroscopy or computed tomography [CT]). One study found that such injections resulted in a good response in 79% of patients and that the improvement could persist for many months. Systemic corticosteroids (either orally or via an intravenous bolus protocol) have been used for severe symptomatic flares, but there are few controlled trials to validate their effectiveness. The goal should be prompt tapering of the dose when symptomatic control is achieved. The recognition that osteoporosis (Chapter 243) is a significant problem in AS provides further impetus to use corticosteroids sparingly. Topical steroids are usually effective for the treatment of the mucous membrane and skin manifestations of ReA. For uveitis, topical corticosteroid eye drops are an integral component of management, and treatment should be monitored jointly with an ophthalmologist.

Sulfasalazine

Randomized placebo-controlled trials have provided some support for the use of sulfasalazine (SSZ), particularly in PsA. Three 36-week randomized double-blind multicenter studies of patients with AS, PsA, and ReA, respectively, were undertaken to compare SSZ (2 g/day) with placebo in each case. An analysis of these studies stratified the patients into those having axial disease and those having peripheral disease. In patients with only axial disease, response criteria were met equally in the SSZ group and the placebo group. In patients with peripheral arthritis, significantly superior responses were seen with SSZ: 59% of the SSZ group and 43% of the placebo group responded ($P < .0005$). These findings are useful in guiding the selection of patients for SSZ treatment. A recent study comparing SSZ with etanercept in AS demonstrated the superiority of tumor necrosis factor (TNF) inhibitor therapy with respect to symptomatic improvement as well as MRI evidence of inflammation.

Methotrexate

Concurrent with the widespread use of methotrexate (MTX) in patients with RA, there has been increasing use of MTX in patients with SpA, but responses have been good only for peripheral joint disease. There is no evidence that MTX is effective for the spinal inflammation characteristic of AS, nor is there evidence that MTX changes the course of axial involvement in AS. Experience with long-term MTX therapy in patients with PsA has increased, although there has been little in the way of randomized controlled trials. Long-term follow-up may be required to resolve whether MTX has a joint-sparing effect in PsA.

Disease-Modifying Agents

There have been additional therapeutic approaches to the control of PsA. The clinical impression of leflunomide therapy has been positive, in general, although there are few formal trials to validate this impression. The mechanism of action of disease-modifying antirheumatic drugs (DMARDs) in SpA has not been resolved. The response of SpA patients to SSZ may be attributable to the antibiotic moiety of this compound (sulfapyridine) or to the anti-inflammatory moiety (5-aminosalicylic acid).

Antibiotic Therapy

The current concept of the pathogenesis of ReA postulates that a bacterial infection, usually gastrointestinal (GI) or genitourinary (GU), is the triggering event in an immunogenetically susceptible host. For the other subsets of SpA, there is less compelling evidence that infection plays a causal role. It is sound clinical practice to treat any culture-proven chlamydial urethritis in conjunction with treatment of the sexual partner. For this indication, a single 1-g dose of azithromycin is as effective as doxycycline 100 mg twice a day for 7 days. The role of antibiotics in the management of ReA has been controversial. A meta-analysis concluded that there is little evidence for efficacy of antibiotic therapy in ReA. But a recent report compared rifampin/azithromycin, rifampin/doxycycline, and placebo for chronic *Chlamydia*-induced ReA and observed that tender and swollen joint counts responded more significantly to the combination antibiotics than to placebo. The precise role of antibiotics in chronic post-*Chlamydia* ReA has not yet been determined.

Anti–Tumor Necrosis Factor Therapy

The pathogenic role of immunomodulatory cytokines in the pathogenesis of SpA has remained unresolved, but the advent of biologic agents has changed the landscape for SpA.[5,6] Biologic agents such as monoclonal antibodies to TNF-α (infliximab, adalimumab, golimumab) or the soluble TNF receptor (etanercept) have been used in the treatment of SpA. So far, these four anti-TNF agents have been comparably effective in trials of AS and PsA. These studies have generally reported a prompt response in clinical outcome measures and in laboratory indicators of inflammation, and MRI evaluations have shown improvement in local inflammation in the sacroiliac joints and spine. The anti-TNF treatments have been well tolerated, with no significant incidence of serious adverse events, but patients appear to relapse when treatment is discontinued. Experience with longer-term treatment with anti-TNF agents has been encouraging with regard to the persistence of the therapeutic effect and the infrequency of late adverse events. These biologic agents have been shown to retard radiographic progression in PsA, and a recent report suggests a similar effect in AS.[7] Additional data using MRI may be helpful in assessing the capacity for these agents to alter the long-term outcomes of AS structurally as well as functionally.

GENETIC SUSCEPTIBILITY

Recent genome-wide association studies in AS have identified additional genetic markers of susceptibility for AS. Polymorphisms in the *IL-23R* gene are associated with AS, and these particular variants are the same as those seen in IBD and psoriasis.[8] Thus the clinical convergence of these different diseases, well known to clinicians, now appears to have a common genetic

The major histocompatibility complex (MHC), on the short arm of chromosome 6 in humans, is one of the most polymorphic regions of the human genome. This is particularly so for the B locus, which constitutes part of the class I MHC genes in this complex. More than 200 different alleles have been detected at this locus, of which B27 is just one. As with all HLA alleles, there is codominant expression of B locus genes, such that most individuals who are B27 positive are heterozygous for the B locus. There appears to be little clinical or prognostic significance associated with the less common homozygous B27 state. The conventional role of class I HLAs is to present a processed peptide to the T-cell receptor of a specific $CD8^+$ cytotoxic T cell, thereby initiating an immune response against the pathogen from which that peptide was derived by intracellular proteolysis and processing. This function gives the HLAs a critical role in host defense against pathogens, and the heterogeneity of cellular immune responses is alleged to be an advantage for a species (such as humans) with extensive polymorphism in a region of the genome. This idea has led to the concept that infectious diseases have driven allelic polymorphisms in the MHC. Such a hypothesis postulates a selective advantage in the extensive peptide-binding capabilities conferred by different alleles of the B locus, as appears to be the case for HLA-B27's association with a more robust immune response to hepatitis C virus infection.

element. Polymorphisms in the endoplasmic reticulum aminopeptidase (ERAP) gene constitute the strongest genetic risk factor for AS after HLA-B27,[9] and the association with AS is restricted to HLA-B27$^+$ AS patients, suggesting a gene-gene interaction. ERAP plays a key role in trimming peptides in the endoplasmic reticulum before loading these peptide complexes onto a nascent class I MHC molecule. This finding continues to attribute a central role to MHC class I peptide presentation in the pathogenesis of AS. With larger numbers studied, the list of candidate genes conferring susceptibility to AS has now extended to more than 20, but the odds ratio for any one gene is modest, with the notable exception of HLA-B27.

It is believed that the prevalence of AS in various parts of the world closely parallels the prevalence of B27 in that population, and in general, this pattern is valid. What introduces complexity into this concept is the recognition that there are more than 30 subtypes of B27. HLA-B2705 is regarded as the primordial subtype, with variability developing over time on the basis of alterations in genomic DNA. Some subtypes, notably B2706 and B2709, do not seem to confer increased susceptibility to the development of AS. This observation has led to a search for "arthritogenic peptides" that are presented by the disease-associated subtypes such as B2705 and B2704, but not by the non–disease-associated subtypes. To date, no simple peptide-susceptibility relationship has been demonstrated, but this is an important clue to the pathogenic role of B27, and studies are ongoing to explore this relationship. Recent studies have suggested that certain B27 subtypes have specific interactions with ERAP, which might fundamentally alter MHC structure and function.

Genome-wide screening studies of multiplex families with SpAs, particularly AS, are ongoing in several countries to identify other genes involved in the predisposition to these diseases. The strongest association of SpA to date remains with the HLA complex, so at least in familial AS, B27 may to a certain extent be necessary (but not sufficient) to confer disease susceptibility. MRI studies in asymptomatic B27-positive individuals indicate that there is a much higher prevalence of sacroiliitis than previously recognized, and studies are continuing to define that prevalence and indeed the prevalence of SpAs in the general population. Some investigators have concluded that SpA is as common as RA.

CLINICAL SUBSETS OF THE SPONDYLOARTHROPATHIES

Ankylosing Spondylitis

EPIDEMIOLOGY

AS is the most common inflammatory disorder of the axial skeleton.[10] The following is a useful rule of thumb: AS occurs in 0.2% of the general population, in 2% of the B27-positive population, and in 20% of B27-positive individuals with an affected family member. There is a male preponderance in the disease, with the male-to-female ratio ranging from 2.5 : 1 to 5 : 1; however, recent epidemiologic studies have found more female involvement than these earlier estimates indicate. The basis for the gender bias has not been resolved. It is held, however, that AS is underrecognized in women, perhaps because of milder axial disease and a more delayed disease onset, but alternative diagnoses of pelvic and low back pain in women may hinder clinician awareness of the disease in female patients.

CLINICAL MANIFESTATIONS

AS typically begins in young adulthood, but symptoms may arise in adolescence or earlier. Up to 15% of children with juvenile idiopathic arthritis are classified as having juvenile AS. Such children may have a pauciarticular pattern, with a predilection for the tarsal joints and frequently minimal spinal complaints. During the adolescent years there is an increasing prevalence of radiographic sacroiliitis, with a significant proportion of patients manifesting this feature by the end of the teenage years. At the other end of the age spectrum, a small number of patients with late-onset AS may have sacroiliitis and oligoarthritis. The axial involvement and asymmetrical lower extremity involvement may serve to differentiate such patients from those with late-onset RA, although there may be overlapping clinical features. Recent studies indicate that the rate of radiographic progression may be less in juvenile-onset AS than in adult-onset AS.

The classic manifestation of AS is the onset of low back pain that persists for more than 3 months, is accompanied by early-morning stiffness, and is typically improved by exercise but not by rest (Table 265-3). Some studies would include a response to NSAID therapy as an additional feature differentiating AS from mechanical low back pain. Back pain that awakens the patient from sleep is often a clue to inflammatory back pain that may have been misdiagnosed as the pain of degenerative disc disease, the latter being a much more common cause of low back pain in the population at large. The pain typically occurs in the region of the sacroiliac joints, with or without slight radiation to the buttock area. Midthoracic pain and cervical pain, particularly at night, are less common but strongly suggest inflammatory back pain when they occur. Fatigue is also a suggestive symptom and is often a major concern for the typical young male patient who has a high functional target in terms of sports and recreation. If the inflammation is inadequately controlled, there is increasing stiffness that may persist most of the day, as well as progressive loss of mobility and flexibility.

TABLE 265-3 MODIFIED NEW YORK CRITERIA FOR ANKYLOSING SPONDYLITIS (1984)

CLINICAL CRITERIA
Low back pain and stiffness for > 3 mo that improve with exercise but are not relieved by rest
Limitation of motion of the lumbar spine in both sagittal and frontal planes
Limitation of chest expansion
RADIOLOGIC CRITERIA
Sacroiliitis: grade ≥ 2 bilateral or grade 3 or 4 unilateral
GRADING
Definite AS if the radiologic criterion is associated with at least one clinical variable
Probable AS if:
The three clinical criteria are present
The radiologic criterion is present without the clinical criteria

AS = ankylosing spondylitis.

Peripheral oligoarthritis is seen in up to 30% of patients with AS. Typically, it is an asymmetrical oligoarthritis with a predilection for the lower extremities. It is important to ask about concurrent or previous tendinitis (e.g., Achilles tendinitis) or heel pain (e.g., plantar fasciitis), because either may reflect an enthesitis that is part of the clinical picture. Involvement of the hip can occur at any point in the course of AS and can follow a course to joint destruction. A hip flexion contracture on this basis may contribute to increasing stoop on standing and walking, which may otherwise be attributed to spinal involvement in the disease.

Extra-articular features most commonly involve the eye. Ocular involvement may occur in up to 40% of AS patients, most typically acute anterior uveitis (iritis). The uveitis often manifests as a slight impairment in visual acuity, with accompanying photophobia and eye pain. Typically, it is unilateral and recurrent. IBD and psoriasis occur in approximately 10% of AS cohorts. Less common manifestations include aortic insufficiency, cardiac conduction defects, and pulmonary fibrosis.

DIAGNOSIS

Physical Examination

Physical examination of the spine characteristically indicates restricted movement, which in the early stages may reflect paraspinal muscle spasm in part; late in the course it reflects ankylosis of the zygapophyseal joints and syndesmophyte bridging of the vertebral bodies. Forward flexion is restricted and can be monitored by Schober's test. This test is used to measure mobility in the lower part of the back: with the patient standing upright, a 10-cm span is marked from the fifth lumbar vertebra upward. On maximal forward flexion, the distance between the marks is remeasured. With normal spinal mobility, the flexed distance should register as 15 cm or an increment of 5 cm. Thoracic involvement is measured in chest expansion, with the chest circumference at maximal inspiration being more than 5 cm greater than the circumference at maximal expiration. Changes in cervical mobility can be measured as the occiput-to-wall distance, with the patient's heels against the wall as the patient attempts to touch the back of the head to the wall. Restricted spinal mobility early in the course of the disease may best be detected by lateral spinal flexion, measured as the difference in the finger-to-floor distance when standing erect compared with maximal bending to the side. Inflammation in the sacroiliac joint may be reflected by joint line tenderness to direct pressure or by the FABERE test (for Flexion, ABduction, External Rotation, and Extension) or Gaenslen maneuver. In the former, the patient lies supine while the examiner flexes and externally rotates the hip. In the latter, the examiner extends the hip by letting the leg dangle off the side of the examining table. In both cases,

It is clear that HLA is strongly associated with SpA, yet the prevalence of HLA-B27 varies widely in different racial and ethnic clusters around the world. It is virtually absent in aboriginal populations in Australia, occurs in 1% of the population in Japan, in 7% in northern European countries, and in 50% in some of the native tribes in western Canada. The environmental-genetic interaction that may account for the expansion or restriction of this gene in human populations is unknown, but some evidence indicates that B27 may confer a more effective host response to some viruses, such as hepatitis C. This variability has a practical impact for the clinician. Because the relative risk conferred by a gene reflects the prevalence of the gene in affected individuals versus its prevalence in the normal population, the relative risk for SpA is higher in a population in which the gene is uncommon (e.g., Japan) than in a population in which B27 is more common (e.g., Scandinavia). In the North American white population, the prevalence of the gene is approximately 7%. Thus, there is a 7% "false-positive" rate if one is attempting to use the gene as a diagnostic marker to decipher the cause of chronic back pain in an unselected population of patients. In contrast, 90% of patients with AS are B27 positive, so there is a 10% "false-negative" rate when using the test diagnostically. The key factor is pretest probability. In a patient with chronic back pain that is clearly inflammatory in character, the addition of B27 positivity combines to strengthen the likelihood of AS accounting for the back problem. The presence of distinctive extra-articular features (e.g., uveitis) further increases this likelihood.

Other steps have also been taken to define the mechanism whereby B27 confers disease susceptibility, in addition to that of uniquely presenting an arthritogenic peptide to T cells. According to the theory of molecular mimicry, an autoimmune response can ensue after an infection if the immune response against the pathogen cross-reacts with host antigens. There is a degree of sequence homology between B27 and several candidate gram-negative enteric bacteria, and there is evidence for cross-reacting monoclonal antibodies, but the significance of such homology for disease pathogenesis remains unresolved. It has also been argued that B27 is distinctive in its propensity to misfold in the endoplasmic reticulum, which may induce a pro-inflammatory cascade called the unfolded protein response. Furthermore, B27 may have a distinct tendency to form heavy-chain homodimers at the cell surface, and the possible consequences of this change for the immune response are under investigation. There has also been investigation into the alteration of primary host-pathogen interactions, such as modulation of pathogen invasion, intracellular replication, and clearance. However, no definitive allele-specific relationships have been demonstrated in these studies.

In B27 transgenic rats, the spontaneous development of pathology that is strikingly similar to human SpAs has supported the notion that B27 itself is the critical genetic factor in disease pathogenesis. These animals demonstrate pathology similar to that of Crohn's disease in the GI tract, spondylitis, peripheral arthritis, uveitis, and psoriasiform skin and nail changes. Of interest, if such animals are raised in a germ-free environment, there is a marked reduction in joint and gut disease, implying a dynamic interrelationship between microbial triggers and background host genes that seems to recapitulate the situation seen clinically.

stress is placed on the sacroiliac joint and may reproduce the back pain if it derives from this site.

Laboratory Findings

Laboratory tests in the evaluation of inflammatory back pain are relatively nonspecific. The ESR and CRP are typically elevated, but normal levels do not exclude inflammatory back pain, and the degree of elevation is typically less than would be seen in acute RA. Anemia of chronic disease may be observed if the condition is long-standing. HLA-B27 is rarely the definitive factor for diagnosis, and the false-positive and false-negative rates have already been discussed; however, in the setting of characteristic back symptoms, the test has reasonably high sensitivity and specificity.

Imaging

Radiographic assessment is important for confirmation of disease, but early in the course there may be no radiographic changes in the sacroiliac joints. If the clinician has a high index of suspicion in such cases, MRI may improve the sensitivity of the plain radiograph because inflammatory changes on MRI predate radiographic changes.[11,12] When ordering x-rays, specific views of the sacroiliac joints can be requested. A routine anteroposterior pelvic radiograph is generally the standard diagnostic x-ray. The classic findings are bilateral changes in the sacroiliac joints (Fig. 265-2). Abnormalities include erosions in the joint line, pseudowidening, subchondral sclerosis, and, finally, ankylosis, reflecting complete bony replacement of the sacroiliac joints.

Radiographs of the spine may reveal squaring of the vertebral bodies (loss of the normal anterior concavity of the lumbar vertebra) and "shiny corners" (subchondral sclerosis at the upper edge of the vertebral body), both of which are manifestations of enthesitis. Syndesmophytes, which represent marginal bridging of the vertebrae (Figs. 265-3 and 265-4), eventually develop and make the diagnosis clear. Because ankylosis of the apophyseal joints may occur without syndesmophyte formation, it is important to assess the posterior joints on the lateral lumbosacral spine views, as well as the anterior margin of the vertebrae. Eventually, the changes may result in a "bamboo spine," so called because the bridging syndesmophytes can mimic the appearance of bamboo. It is now appreciated that osteoporosis (Chapter 243) is a significant feature of AS, probably reflecting both the local chronic inflammation and the abnormal biomechanical loading of the vertebrae as the disease progresses.

Differential Diagnosis

The differential diagnosis of AS includes the following: osteitis condensans ilii; diffuse idiopathic skeletal hyperostosis (DISH); the syndrome of synovitis, acne, pustulosis, hyperostosis, and osteomyelitis (SAPHO); and some induced hyperostotic states (vitamin A intoxication, fluorosis). New bone formation occurs in degenerative disc disease, but the bulky horizontal appearance of osteophytes is usually easily distinguished from that of syndesmophytes, and narrowing of the disc space is not a feature of AS. Osteoarthritis of the sacroiliac joint has recently been recognized as having a higher prevalence than previously appreciated.

The clinical course and severity of AS are highly variable. Inflammatory back pain and stiffness dominate the picture in the early stages, whereas chronic pain and deformity may develop over time. In both early and late phases of the disease, there may be a significant impact on work disability and quality of life. In only a minority of patients does the full-blown picture of a bamboo spine eventually develop, but there are few variables that can reliably aid in prognosticating the course. At present, the strongest predictor of new syndesmophyte formation is the presence of syndesmophytes at baseline. In AS patients in whom new, refractory spinal pain develops, an intervertebral fracture should be considered, which can occur after only minimal trauma.

Additional late complications may include cauda equina syndrome, osteoporotic compression fractures, spondylodiscitis, and restrictive lung disease.

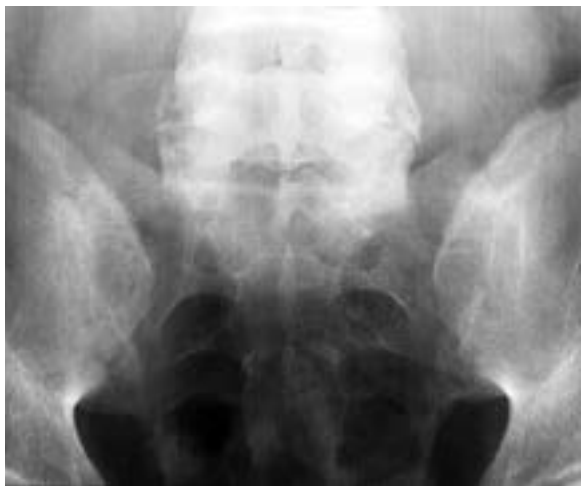

FIGURE 265-2. Bilaterally symmetrical sacroiliitis in ankylosing spondylitis.

Reactive Arthritis

DEFINITION

ReA is an aseptic arthritis that occurs subsequent to an extra-articular infection, most typically of the GI or GU tract. In the GI tract, the key pathogens

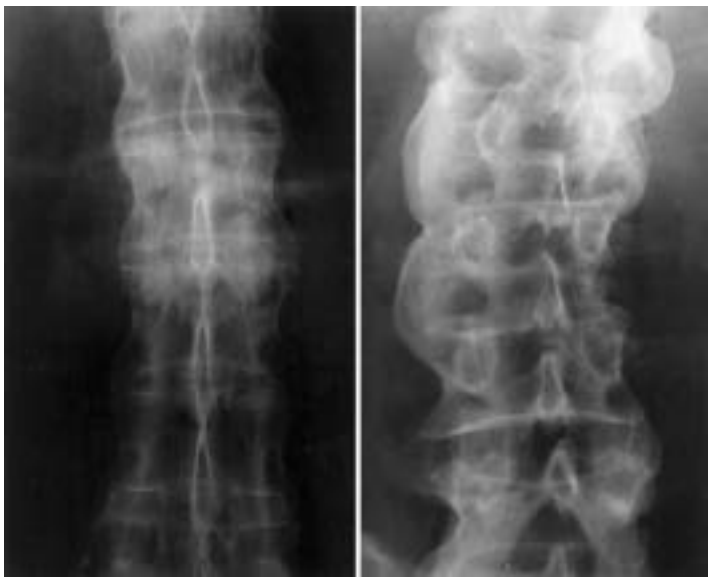

FIGURE 265-3. *Left,* Lumbar spondylitis in ankylosing spondylitis, with symmetrical marginal bridging syndesmophytes and calcification of the spinal ligament. *Right,* The bulky, nonmarginal, asymmetrical syndesmophytes of reactive arthritis with lumbar spondylitis.

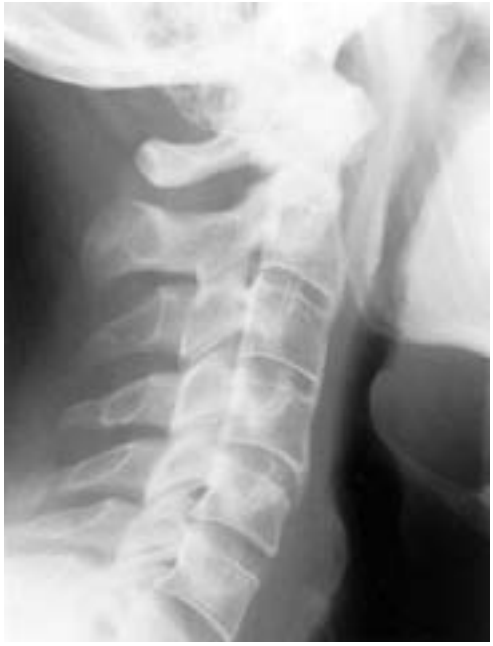

FIGURE 265-4. A 34-year-old man who has had ankylosing spondylitis for 9 years and neck pain. Radiographs demonstrate narrowing of the C2-C3 apophyseal joints posteriorly and anterior bridging marginal syndesmophytes extending from C2 to C5.

are *Salmonella typhimurium*, *Yersinia enterocolitica*, *Shigella flexneri*, and *Campylobacter jejuni*. In the GU tract, *Chlamydia trachomatis* is the most common offender.

EPIDEMIOLOGY

The true incidence and prevalence of ReA are not well defined. In epidemics involving *Salmonella* (Chapter 308) or *Yersinia* (Chapter 312), it is estimated that ReA develops in 2 to 7% of infected individuals but in as many as 20% of B27-positive infected individuals. In such epidemic studies, B27 confers risk not only for the onset of arthritis but also for axial involvement and chronicity. Genetic variants in toll-like receptor 2 (TLR-2) are associated with acute ReA, thus implicating host innate immunity as central in ReA. The variability in the rate of ReA is determined by the heterogeneity of the cohorts reported, which introduces confounding variables of different genetic backgrounds in the population and different species of pathogens. Even in the setting of an epidemic point source outbreak, the inoculum varies widely among the exposed individuals, and the genetic makeup of the population at risk (e.g., the prevalence of B27) may differ greatly among different studies. Case ascertainment and relative risk are even more difficult to determine for post-*Chlamydia* ReA. Young adults in the United States have a high prevalence of asymptomatic *Chlamydia* carriage in the GU tract, and establishing a causal link between *Chlamydia* and synovitis can be difficult. Nevertheless, it is with *Chlamydia* that ReA has been most intensively studied.

PATHOBIOLOGY

Although immunofluorescence studies have identified bacterial antigens in the joints of patients with ReA after both GI and GU infections, it is primarily in post-*Chlamydia* ReA that results of polymerase chain reaction studies on synovial tissues have most consistently been positive, suggesting that viable *Chlamydia* may persist in the joints of such patients, albeit in a metabolically altered state.

Typically, the onset of arthritis occurs 1 to 3 weeks after the GI or GU infection, but the temporal details are often difficult to define precisely.

Although the definition of aseptic arthritis after an extra-articular infection may include a broader range of pathogens (e.g., *Chlamydia pneumoniae*), sites of infection (e.g., streptococcal pharyngitis), and types of infections (e.g., *Giardia* infections of the GI tract), these clinical scenarios have not generally been included in the category of ReA. They lack the other associated clinical features of the SpA group of diseases, and they lack an association with B27.

DIAGNOSIS

The pattern of joint involvement in ReA is one of asymmetrical oligoarthritis with a predilection for the lower extremity, a pattern shared by most SpA syndromes. Enthesitis may present as Achilles tendinitis or plantar fasciitis. Dactylitis, appearing as a sausage digit, may also be seen. Dactylitis is the net result of inflammatory changes affecting the joint capsule, entheses, periarticular structures, and periosteal bone. Sacroiliitis may be seen in the acute phase, but radiographic changes are seen largely in patients with a more chronic course.[13]

When ReA is accompanied by certain extra-articular features such as urethritis, conjunctivitis, or mucocutaneous lesions, the term *Reiter's syndrome* has been applied historically, but it is no longer in common use. The urethritis may manifest as dysuria or discharge, and the rash as circinate balanitis, which appears as vesicles or shallow ulcerations on the glans penis. Painless lingual or oral ulcerations may also be seen. The fact that the cervicitis may be less symptomatic could partially account for the underdiagnosis in women. The classic skin manifestation of ReA is keratoderma blennorrhagicum, a painless papulosquamous eruption on the palms or soles (Fig. 265-5). Occasionally, nail dystrophy with pitting and onycholysis or subungual keratosis can be seen. The conjunctivitis can be bilateral and painful; in contrast, the acute anterior uveitis that can also be seen in this setting tends to be less painful and unilateral.

Radiographic changes of ReA can be seen in the involved peripheral joints, with early findings consisting of soft tissue swelling and juxta-articular osteopenia. Areas of periostitis and new bone formation may develop in peripheral joints. When changes in the sacroiliac joints are seen, they are typically asymmetrical (Fig. 265-6), in contrast to the symmetrical pattern seen in AS. In the chronic phase, syndesmophytes may develop, but they are described as bulky, nonmarginal, often asymmetrical formations that differ from the classic syndesmophytes of AS. The frequency with which ReA evolves into bona fide AS has not been determined definitively.

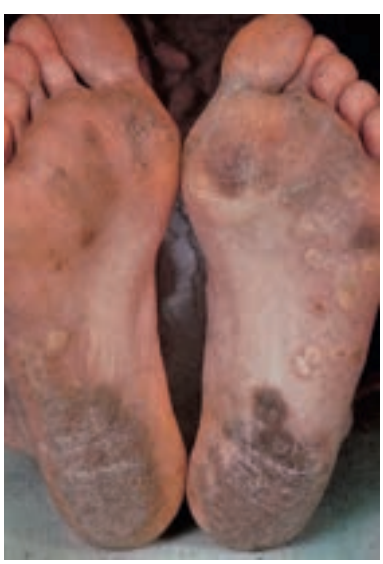

FIGURE 265-5. Keratoderma blennorrhagicum of the feet in reactive arthritis.

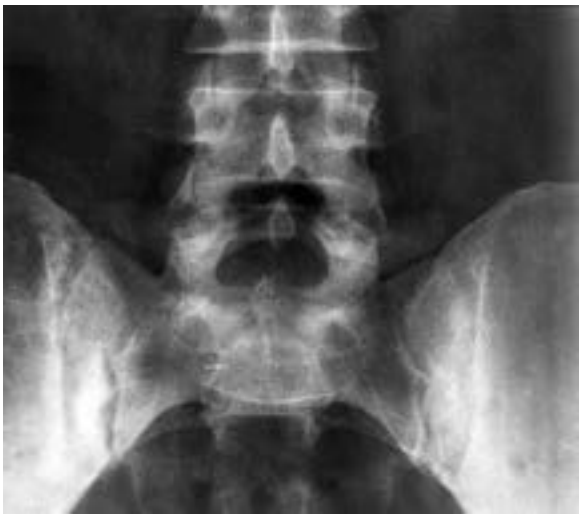

FIGURE 265-6. Bilaterally asymmetrical sacroiliitis in reactive arthritis. Erosions, pseudowidening, and ileal sclerosis are present.

Differential Diagnosis

The most important differential diagnosis for such reactive arthropathies is septic arthritis. Both *Yersinia* and *Salmonella* can cause septic arthritis, so an appropriate culture of synovial fluid should precede the diagnosis of ReA whenever possible. The course of ReA is variable, and few prognostic markers are available for the clinician to predict the course in an individual case. The majority of patients have an initial episode lasting 2 to 3 months, but synovitis may persist for a year or longer. In one 5-year follow-up of a point source cohort of post-*Salmonella* ReA, 20% of patients had ongoing inflammatory joint disease, and some degree of functional disability was observed in 30% of patients 5 years after the onset of disease.

REACTIVE ARTHRITIS AND HUMAN IMMUNODEFICIENCY VIRUS

An aggressive form of SpA may be seen in patients who are concomitantly infected with HIV.[14] There is no increased frequency of ReA in patients with HIV, but HIV may alter the course of these arthropathies, with a tendency for a more aggressive and more refractory joint disease. Aggressive skin and joint disease may be seen in patients in whom PsA develops in the setting of HIV infection. Most North American patients with the HIV-ReA constellation are B27 positive, but studies of comparable patients in Africa have found a sizable B27-negative component in such patients. The arthritis in these patients falls into two clinical patterns: (1) an additive, asymmetrical polyarthritis or (2) an intermittent oligoarthritis that most commonly affects the lower extremities. Enthesitis, fasciitis, conjunctivitis, and urethritis can all be

seen in such patients. Sacroiliitis can occur, although extensive spinal syndesmophyte formation is not common.

Psoriatic Arthritis

EPIDEMIOLOGY

PsA develops in 5 to 7% of patients with psoriasis. Although most cases arise in patients with established cutaneous disease, some patients (particularly children) have arthritis that antedates the appearance of the skin lesions.[15] Although the extent of psoriatic skin disease correlates poorly with the development of arthritis, the risk for PsA increases with a family history of SpA. The age at onset can range from 30 to 55 years, with an equal predilection for PsA in women and men. Psoriatic spondylitis has a slight male preponderance. Two large prospective studies suggest that obesity is a significant risk factor for psoriatic arthritis.[16,17]

PATHOBIOLOGY

The genetic associations with PsA are complex. Psoriasis itself is associated with several HLA loci; some B alleles have been reported, but the dominant element is HLA-Cw6. HLA-B39 and HLA-B27 have been associated with sacroiliitis and axial involvement. No etiologic agent has been proved in PsA, although some investigators have proposed that the disease process represents ReA in response to cutaneous bacteria. The histopathology of the synovitis of PsA is comparable to that of the other forms of SpA, with the absence of the local production of immunoglobulin and rheumatoid factor differentiating this disease from RA. There is the potential for aggressive osteolysis, fibrous ankylosis, and heterotopic new bone formation to occur in PsA. As mentioned earlier, the coexistence of HIV and PsA seems to set the stage for an aggressive course of joint destruction in some patients.

DIAGNOSIS

PsA has a variable manifestation and disease course, but several clinical patterns have been identified in prospectively monitored cohorts of patients. The clinical subsets are not mutually exclusive, nor are they static over time. The most common form, which affects 30 to 50% of patients, is an asymmetrical oligoarthritis that may involve both large and small joints. Dactylitis, arising as sausage digits, can be seen in fingers and toes and actually represents an enthesitis. In the second subset there is selective targeting of the distal interphalangeal joints, seen in 10 to 15% of patients. These changes are strongly associated with nail dystrophy, of which the features are onycholysis, subungual keratosis, pitting, and oil drop–like staining (Fig. 265-7). The third subset (15 to 30% of patients) has a symmetrical polyarthritis that mimics RA in many ways, except for the absence of rheumatoid nodules and rheumatoid factor. The fourth clinical variant is psoriatic spondylitis, which occurs in 20% of patients; 50% of such patients are B27 positive. Finally, arthritis mutilans (5% of patients) is a destructive, erosive arthritis that affects large and small joints and can be associated with marked deformities and significant disability.

Radiographic changes in PsA involve soft tissue swelling (particularly in the case of dactylitis), erosions, and periostitis. Axial involvement may lead to the appearance of asymmetrical sacroiliitis with syndesmophytes that are bulky, asymmetrical, and nonmarginal. The classic "pencil-in-cup" deformity may be seen in patients with distal interphalangeal joint disease or arthritis mutilans. Acro-osteolysis is noted in a minority of patients and reflects an aggressive erosive process.

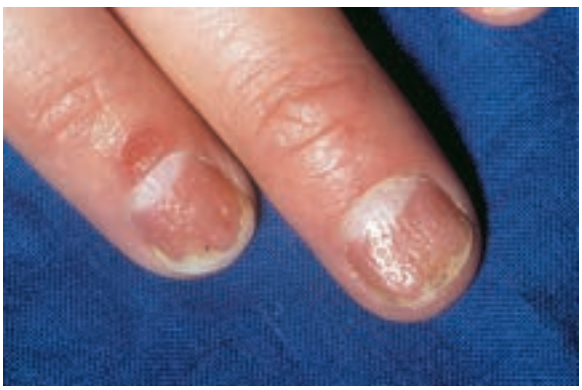

FIGURE 265-7 Nail pitting in psoriasis. The pits are more discrete and regular compared with pits affecting the nail plate in dermatitis.

Differential Diagnosis

The diagnosis of PsA depends on finding the typical skin or nail changes in association with one of the articular variants described previously. The differential diagnosis for the skin lesions can include seborrheic dermatitis, dyshidrotic eczema, fungal infection, keratoderma blennorrhagicum, and palmoplantar pustulosis.

TREATMENT Rx

Patients typically receive aggressive treatment for psoriasis (Chapter 438). The advent of biologic agents has had a major impact on the treatment of PsA. The anti-TNF agents have been studied most extensively, indicating the efficacy of infliximab, etanercept, adalimumab, and golimumab. In a phase III randomized placebo-controlled trial, treatment of psoriatic arthritis with subcutaneous golimumab (50 mg to 100 mg every 4 weeks) inhibited the progression of structural damage and demonstrated continued clinical efficacy and safety through 1 year.[■] In a study of patients with PsA, ustekinumab, a monoclonal antibody against interleukin-12/23, was well tolerated, reduced the extent and severity of psoriasis, and was safe.[■] Brodalumab, a human monoclonal antibody against interleukin-17 receptor A (IL17RA), significantly improved response rates among patients with PsA in a phase 2 randomized, double-blind, placebo-controlled study.[■] The European League Against Rheumatism in 2012 published recommendations for the management of PsA with systemic and local (nontopical) symptomatic and disease-modifying antirheumatic drugs. They suggest nonsteroidal anti-inflammatory drugs to relieve musculoskeletal signs and symptoms; treatment with disease-modifying drugs such as methotrexate, sulfasalazine, or leflunomide in patients with swollen joints, structural damage in the presence of inflammation, or clinically relevant extraarticular manifestations; and anti-TNF agents in patients with active enthesitis and/or dactylitis and insufficient response to other medications. Patients should be switched to another anti-TNF agent if the first is not successful.

Enteropathic Arthritis

DEFINITION

EA refers to the arthritis associated with Crohn's disease (CD) or ulcerative colitis (UC) (Chapter 141; Table 265-4).

PATHOBIOLOGY

All extraenteric manifestations, including arthritis, occur more commonly in CD than in UC. Peripheral arthritis occurs in 10 to 20% of CD patients and in 2 to 7% of UC patients. This pattern of arthritis occurs more commonly in patients with other extraenteric features (e.g., erythema nodosum, iritis). It is typically an inflammatory nonerosive polyarthritis, predominantly of large joints. In general, the clinical activity of the peripheral arthritis parallels the

TABLE 265-4 ENTEROPATHIC ARTHRITIS

FEATURE	PERIPHERAL ARTHRITIS	SACROILIITIS, SPONDYLITIS
CROHN'S DISEASE (CD)		
Frequency in CD	10-20%	2-7%
HLA-B27 associated	No	Yes
Pattern	Transient, symmetrical	Chronic
Course	Related to activity of CD	Unrelated to activity of CD
Effect of surgery	Remission of arthritis uncommon	No effect
Effect of anti-TNF therapy	Effective	Effective
ULCERATIVE COLITIS (UC)		
Frequency in UC	5-10%	2-7%
HLA-B27 associated	No	Yes
Pattern	Transient	Chronic
Course	More common in pancolitis than proctitis; related to activity of UC	Unrelated
Effect of surgery	Remission of arthritis	No effect

HLA = human leukocyte antigen; TNF = tumor necrosis factor.

activity of the gut inflammation, and measures that control the GI disease usually control the joint disease as well. The peripheral arthritis of EA is not associated with B27. As mentioned earlier, IBD and AS share genetic susceptibility associated with numerous genes.

In contrast, the sacroiliitis or spondylitis of EA follows a pattern in which the joint inflammation waxes and wanes independently of the bowel inflammation. Axial disease occurs in 2 to 7% of both CD and UC patients. HLA-B27 is found in 50% of patients with axial arthritis. The course tends to be chronic, as opposed to the transient course of peripheral arthritis.

The association of bowel inflammation and arthritis is supported by ileocolonoscopic studies in which subclinical inflammation of the bowel has been demonstrated in diseases covering the entire spectrum of SpAs. Histologic evaluation demonstrates that changes of acute ileitis are seen in postdysenteric ReA, whereas chronic inflammatory changes are more likely to be seen in patients with AS. As mentioned earlier, the abnormalities in the bowel of B27 transgenic rats have strong similarity to the lesions of CD, and a germ-free environment minimizes inflammatory changes in both the gut and the joints. This finding argues that altered bowel permeability, with enhanced bacteremia or antigenemia, may provide the link in both cases.

DIAGNOSIS

It is important to recognize that the musculoskeletal features of EA may precede any GI symptoms or signs. Conversely, the diarrhea preceding the onset of peripheral or axial arthritis in a young patient could just as likely represent a food-borne pathogen (e.g., *Salmonella*, *Yersinia*), with secondary ReA as IBD and accompanying EA. In the initial assessment of such a patient, it is important to carry out careful and complete stool cultures. If the GI symptoms persist, diagnostic colonoscopy is often required to resolve the issue.

Undifferentiated Spondyloarthritis

Despite careful clinical and radiographic assessment, there are still a substantial number of patients who do not fall into one of the classic diagnostic subsets of SpA outlined previously. These patients are often defined as having USpA with peripheral enthesitis, asymmetrical arthritis or sacroiliitis, or iritis in the absence of identifiable antecedent infection or concurrent IBD or psoriasis. The natural history of USpA has not been well defined, and case heterogeneity and diagnostic dilemmas plague a systematic or multicenter approach to the problem. When the clinical course is examined, a number of patients may finally meet the diagnostic criteria for AS, but many retain a distinct USpA pattern for prolonged periods.[18]

Grade A References

A1. Braun J, van der Horst-Bruinsma F, Huang F, et al. Clinical efficacy and safety of etanercept versus sulfasalazine in ankylosing spondylitis patients: a randomized, double-blind study (ASCEND Trial). *Arthritis Rheum*. 2011;63:1543-1551.

A2. Barber CE, Kim J, Inman RD, et al. Antibiotics for treatment of reactive arthritis: a systematic review and metaanalysis. *J Rheumatol*. 2013;40:916-928.

A3. Carter JD, Espinoza LR, Inman RD, et al. Combination antibiotics as a treatment for chronic *Chlamydia*-induced reactive arthritis: a double-blind, placebo-controlled, prospective trial. *Arthritis Rheum*. 2010;62:1298-1307.

A4. Inman RD, Davis JC, van der Heijde D, et al. Efficacy and safety of golimumab in patients with ankylosing spondylitis: results of the randomized, double-blind, placebo-controlled GO-RAISE trial. *Arthritis Rheum*. 2008;58:3402-3412.

A5. Sieper J, Lenaerts J, Wollenhaupt J, et al. Efficacy and safety of infliximab plus naproxen versus naproxen alone in patients with early, active axial spondyloarthritis: results from the double-blind, placebo-controlled INFAST study, Part 1. *Ann Rheum Dis*. 2014;73:101-107.

A6. Dougados M, van der Heijde D, Sieper J, et al. Symptomatic efficacy of etanercept and its effects on objective signs of inflammation in early nonradiographic axial spondyloarthritis: a multicenter, randomized, double-blind, placebo-controlled trial. *Arthritis Rheum*. 2014;66:2091-2102.

A7. Song IH, Hermann KG, Haibel H, et al. Consistently good clinical response in patients with early axial spondyloarthritis after 3 years of continuous treatment with etanercept: longterm data of the ESTHER trial. *J Rheumatol*. 2014;41:2034-2040.

A8. Kavanaugh A, van der Heijde D, McInnes IB, et al. Golimumab in psoriatic arthritis. One-year clinical efficacy, radiographic, and safety results from a phase III, randomized, placebo-controlled trial. *Arthritis Rheum*. 2012;64:2504-2517.

A9. Gottlieb A, Menter A, Mendelsohn A, et al. Ustekinumab, a human interleukin 12/23 monoclonal antibody, for psoriatic arthritis: randomised, double-blind, placebo-controlled, crossover trial. *Lancet*. 2009;373:633-640.

A10. Mease PJ, Genovese MC, Greenwald MW, et al. Brodalumab, an anti-IL17RA monoclonal antibody, in psoriatic arthritis. *N Engl J Med*. 2014;370:2295-2306.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Braun J, van den Berg R, Baraliakos X, et al. 2010 update of the ASAS/EULAR recommendations for the management of ankylosing spondylitis. *Ann Rheum Dis*. 2011;70:896-904.
2. Cantini F, Niccoli L, Nannini C, et al. Psoriatic arthritis: a systematic review. *Int J Rheum Dis*. 2010;13:300-317.
3. Gossec L, Smolen JS, Gaujoux-Viala C, et al. European League Against Rheumatism recommendations for the management of psoriatic arthritis with pharmacological therapies. *Ann Rheum Dis*. 2012;71:4-12.
4. Passalent LA. Physiotherapy for ankylosing spondylitis: evidence and application. *Curr Opin Rheumatol*. 2011;23:142-147.
5. Heldmann F, Dybowski F, Saracbasi-Zender E, et al. Update on biologic therapy in the management of axial spondyloarthritis. *Curr Rheumatol Rep*. 2010;12:325-331.
6. Haroon N, Inman RD, Weisman MH, et al. The impact of TNF-inhibitors on radiographic progression in ankylosing spondylitis. *Arthritis Rheum*. 2013;65:2645-2654.
7. Eder L, Thavaneswaran A, Chandran V, Gladman DD. Tumour necrosis factor α blockers are more effective than methotrexate in the inhibition of radiographic joint damage progression among patients with psoriatic arthritis. *Ann Rheum Dis*. 2014;73:1007-1011.
8. International Genetics of Ankylosing Spondylitis Consortium (IGAS). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet*. 2013;45:730-738.
9. Haroon N, Tsui F, Uchanska-Ziegler B, et al. Endoplasmic reticulum aminopeptidase 1 (ERAP1) exhibits functionally significant interaction with HLA B27 and relates to subtype specificity in ankylosing spondylitis. *Ann Rheum Dis*. 2012;7:589-595.
10. Dean LE, Jones GT, MacDonald AG, et al. Global prevalence of ankylosing spondylitis. *Rheumatology (Oxford)*. 2014;53:650-657.
11. Jang JH, Ward MM, Rucker AN, et al. Ankylosing spondylitis: patterns of radiographic involvement—a re-examination of accepted principles in a cohort of 769 patients. *Radiology*. 2011;258:192-198.
12. Paparo F, Revelli M, Semprini A, et al. Seronegative spondyloarthropathies: what radiologists should know. *Radiol Med*. 2014;119:156-163.
13. Selmi C, Gershwin ME. Diagnosis and classification of reactive arthritis. *Autoimmun Rev*. 2014;13:546-549.
14. Tikly M, Njobvu P, McGill P. Spondyloarthritis in sub-Saharan Africa. *Curr Rheumatol Rep*. 2014;16:421.
15. Ficco HM, Citera G, Cocco JA. Prevalence of psoriatic arthritis in psoriasis patients according to newer classification criteria. *Clin Rheumatol*. 2014;33:243-246.
16. Li W, Han J, Quereshi AA. Obesity and risk of incident psoriatic arthritis in US women. *Ann Rheum Dis*. 2012;71:1267-1272.
17. Love TJ, Zhu Y, Zhang Y, et al. Obesity and the risk of psoriatic arthritis: a population-based study. *Ann Rheum Dis*. 2012;71:1273-1277.
18. Sieper J, van der Heijde D. Review: Nonradiographic axial spondyloarthritis: new definition of an old disease? *Arthritis Rheum*. 2013;65:543-551.

REVIEW QUESTIONS

1. A 28-year-old man presents with a 2-year history of low back pain and intermittent arthritis in the lower extremity. Physical examination demonstrates bilateral Achilles tendonitis and tenderness to direct pressure over the sacroiliac joints. Which of the following would ***not*** support the diagnosis of spondyloarthritis (SpA)?

A. Presence of HLA-B27
B. History of anterior uveitis
C. Positive rheumatoid factor
D. Elevated CRP
E. History of inflammatory bowel disease

Answer: C The history and the other optional features would be entirely consistent with a diagnosis of SpA. HLA-B27 is strongly associated with SpA in general, and its strongest association is with ankylosing spondylitis (AS). Anterior uveitis occurs in approximately 30% of AS cases, and is itself associated with HLA-B27 even when AS is not present. Positive rheumatoid factor is characteristic of rheumatoid arthritis but not of SpA, so it would not be expected to be positive in such patients. Elevated C-reactive protein (CRP) is a nonspecific marker of inflammation and would be commonly seen in patients with active SpA. Inflammatory bowel disease occurs in approximately 10% of AS cases, and its presence strongly suggests that concurrent musculoskeletal inflammation that the underlying arthritis is part of the SpA spectrum.

2. A 32-year-old man presents with pain and swelling of one ankle and one knee. The history reveals that he had just returned from a trip to South America and developed diarrhea that was now subsiding. He denied any eye inflammation or skin lesions. Physical examination confirmed tenderness and swelling of an ankle and a knee. The next appropriate diagnostic test would be which of the following?

A. Aspiration of the knee swelling and culture of the synovial fluid
B. Initiate a course of broad-spectrum antibiotics.
C. X-rays of the lumbar and cervical spine
D. Check a blood test for the presence of antinuclear antibodies.
E. Initiate an empirical trial of methotrexate.

Answer: A A patient with new-onset swelling in the joints, particularly when one or two joints are affected, should always raise the possibility of septic arthritis. When there has been a recent infection elsewhere in the body, it is possible the inciting infection could spread by a hematogenous route to seed the joint. Although the history is suggestive of reactive arthritis (ReA), it is important to rule out septic arthritis with certainty before entertaining that diagnosis. Culture of synovial fluid is the definitive test to rule out septic arthritis.

Empirical use of broad-spectrum antibiotics rarely has a place in the management of new-onset arthritis. The first step is to make a diagnosis. X-rays of the spine are important in the evaluation of inflammatory back pain, but in this case there is only peripheral joint involvement. If the course of the arthritis becomes chronic or back pain develops over time, x-rays at that point would be appropriate. Antinuclear antibodies in the serum are the hallmark of many autoimmune rheumatic diseases—in particular, lupus. But with the clinical presentation in this case, there are no features to implicate lupus or related conditions, and it would not be appropriate to search for antinuclear antibodies. The immediate priority is to rule out septic arthritis. It would be inappropriate to institute methotrexate before that is resolved.

3. A 30-year-old man presents with worsening back pain. The history reveals that this has been present for the past 3 years and is increasing in severity. It is characterized by early-morning stiffness, nocturnal pain, and modest improvement with exercise. He initially noted an improvement with an NSAID, but subsequently that effect was lost, and he has tried two other NSAIDs without effect. Physical examination reveals limitation in forward flexion of the lumbar spine. An AP x-ray of the pelvis reveals bilateral erosive changes in the sacroiliac joints. A lateral x-ray of the lumbar spine reveals squaring of several vertebral bodies. The next step in pharmacologic treatment of this condition is which of the following?

A. Methotrexate
B. Leflunomide
C. Sulfasalazine
D. TNF inhibitor biologic agents
E. Azathioprine

Answer: D The patient meets the diagnostic criteria for ankylosing spondylitis (AS). He has long-standing back pain that has features of inflammatory back pain and has limitation in spinal mobility. The presence of bilateral erosive sacroiliitis provides imaging support for the diagnosis of AS. He has had an adequate trial of NSAIDs, and his back pain has not responded adequately to these agents. The classical DMARDs used in rheumatoid arthritis—methotrexate, leflunomide, sulfasalazine, and azathioprine—have proved ineffective in controlling the spinal inflammation seen in AS. On the other hand, the TNF inhibitors (infliximab, etanercept, adalimumab, golimumab) have established efficacy in controlling the signs and symptoms of AS.

4. A 35-year-old woman is under your care for ankylosing spondylitis (AS). She is HLA-B27 positive and has no extra-articular features accompanying her back pain. Her symptoms have been controlled with NSAIDs in combination with physiotherapy and exercises. She has an 8-year-old son, and she is inquiring about his current and future health prospects. Which of the following would be advisable in this circumstance?

A. The son should be checked for HLA-B27 status.
B. The mother should be advised that the likelihood of her son developing AS is less than 10%.
C. The son should be advised to avoid sports and physical activities.
D. Preventive low-dose NSAIDs should be instituted for the son.
E. The son should be advised to avoid foreign travel in order to avoid food-borne pathogens.

Answer: B There is a 50% chance that any particular child of a B27-positive patient will be B27 positive, but there is only a 20% chance of a B27-positive individual with a positive family history developing AS. So the likelihood of a child developing AS when he has a B27-positive parent with AS is maximally 10%. It would not be advisable to test an asymptomatic child for HLA-B27, because it has low predictive value (as discussed earlier) and could introduce chronic anxiety about future disease, which is unwarranted. There is no reason to limit the sports and recreation of the child. Maintaining general fitness and muscle strength and posture is sound advice even if there is a greater chance of later arthritis than in the general population. There is no evidence that sports-related injuries trigger AS. There is no evidence that NSAIDs prevent the subsequent development of SpA in individuals who are genetically predisposed to SpA. Chronic NSAID therapy carries its own risks and would not be justified in an asymptomatic individual. There is a theoretical concern that being B27 positive could increase the risk of reactive arthritis after a food-borne pathogen illness. This might lead to advising a patient with ***prior*** diagnosis of ReA to minimize exposure to food-borne pathogens, but it would be inappropriate to restrict travel for anyone who has never had an episode of ReA.

5. A 19-year-old female patient has developed a painful swollen knee over the course of 7 days. There is no history of trauma and no associated skin lesions or ocular complaints. The patient relates a history that she has just returned from a trip abroad with the family and had a transient diarrheal episode while traveling. Further questioning reveals that she has had two prior episodes of diarrhea in the past year, each associated with crampy abdominal pain, but on each occasion the diarrhea was self-limited and she did not seek medical attention. Which of the following would ***not*** be an appropriate next step in the management of this patient?

A. Referral to a gastroenterologist for ileocolonoscopy
B. Culture of a stool sample
C. Aspiration of the knee and culture of synovial fluid
D. Check hemoglobin and ESR.
E. Institute a 2-week course of ciprofloxacin.

Answer: E The patient has developed arthritis in the setting of a recurrent diarrheal course. The key differential is between reactive arthritis (ReA) and enteropathic arthritis (EA) as part of inflammatory bowel disease (IBD). In either case, it would be inappropriate to institute an empirical course of antibiotics before a diagnosis is made. In neither postdysenteric ReA nor in IBD-associated arthritis is there evidence that antibiotics alter the course of the arthritis. It would be appropriate in this case to determine whether the recurrent diarrhea could be a manifestation of Crohn's disease or ulcerative colitis, and ileocolonoscopy would be indicated to address that question. Recurrent diarrhea might reflect a infectious agent such as *Salmonella*. Stool cultures would be an appropriate test to address that. Culture of synovial fluid is the definitive test to rule out septic arthritis. Both *Salmonella* and *Yersinia* can cause a septic arthritis, so it is appropriate to exclude an infection in the joint before proceeding to other management decisions. Anemia of chronic disease might raise the suspicion that a chronic process, such as IBD, has been present for some time. Elevation in CRP or ESR is appropriate in the work-up of this patient, although this would not differentiate ReA from IBD-related arthritis. These acute phase reactants can serve as useful surrogate markers to monitor response to treatment.

266

SYSTEMIC LUPUS ERYTHEMATOSUS

MARY K. CROW

DEFINITION

Systemic lupus erythematosus (SLE) is a multisystemic autoimmune disease that results from immune system–mediated tissue damage. Manifestations of SLE can involve the skin, joints, kidney, central nervous system (CNS), cardiovascular system, serosal membranes, and the hematologic and immune systems. The disease is highly heterogeneous, with individual patients manifesting variable combinations of clinical features. In most patients with SLE, the disease is characterized by a waxing and waning clinical course, although some demonstrate a pattern of chronic activity. The molecular triggers of the disease are not known, but the pathogenesis is understood to involve the production of autoantibodies exhibiting multiple specificities, with reactivity with nucleic acid–binding proteins being a common feature. Immune complexes, along with immune system cells and soluble mediators, generate inflammation and tissue damage. Therapeutic approaches generally involve immunosuppression.

EPIDEMIOLOGY

A notable feature of SLE is that it occurs much more frequently in females than in males. Like Hashimoto's thyroiditis and Sjögren's syndrome, the female-to-male ratio is approximately 8 : 1 to 10 : 1 in adults, and most cases are diagnosed between the ages of 15 and 44 years. In children and women older than 55, the ratio is closer to 2 : 1. The prevalence of SLE in the United States is estimated to be approximately 73 per 100,000, and the incidence of new cases is 5.5 per 100,000 per year. The prevalence, severity, and characteristics of disease differ in different ethnic groups, with SLE being 2.3-fold more frequent in African Americans than in the white population. The severity of disease is also greater in Hispanic individuals than in whites, although data for Hispanic populations are less abundant. Asians may also have a higher prevalence of disease than whites. Recent studies of lupus in minority populations indicate that socioeconomic factors are major contributors to the increased prevalence and severity of disease in African Americans and Hispanic Americans.

PATHOBIOLOGY

Current understanding of lupus pathogenesis incorporates roles for genetic susceptibility based on a threshold model involving multiple genes[1]; environmental triggers, including microbial infection, sunlight, and certain drugs; and altered immune system function. Recent advances in immunology have focused attention on the mechanisms that account for innate immune system activation.[2] At least some of the genetic and environmental contributions to lupus are likely to promote innate immune system activation and subsequent autoimmunity. Others may contribute to inflammation and tissue damage. Induction of cellular stress responses, including oxidative modification of cell proteins, is of current interest as a mechanism that links environmental triggers to altered immune function.

Murine models have proved useful in identifying genes that could contribute to lupus susceptibility or define patterns of disease. Production of autoantibodies characteristic of SLE and development of nephritis and accelerated death have been demonstrated in numerous murine strains in which immune system genes have been modified. In most cases, no alterations have been noted in the homologous human genes. The ease of induction of lupus-like disease in murine models suggests that there are numerous possible pathogenic paths that might lead to the clinical manifestations of lupus. It is not known which of these molecular pathways is responsible for human SLE, although mediators of the immune response to viral infection, particularly components of the type I interferon response, are associated with lupus in both murine and human systems and are likely to be important in disease pathogenesis.[3]

Genetics

An important role for a genetic contribution to lupus susceptibility is suggested by the high concordance of disease in monozygotic twins (14 to 57%).

Genes that might account for increased lupus susceptibility or severity include those encoding components of the complement pathway, including C1q, C2, and C4A (E-Table 266-1). Impaired production of these early complement components may decrease the clearance of apoptotic cells, thereby augmenting the pool of available autoantigens, or decrease the solubility of immune complexes. Polymorphic variants in components of the toll-like receptor (TLR) pathways that regulate type I interferon production, including interferon regulatory factor 5 *(IRF5)*, have been associated with a diagnosis of SLE and increased plasma interferon activity in some populations. Association of SLE with the major histocompatibility complex (MHC) class II alleles human leukocyte antigen (HLA)-DR2 and HLA-DR3 has been documented in many studies and is most striking in patients expressing particular autoantibody specificities. Polymorphisms in the Fc receptor genes *FCGR2A* and *FCGR3A* have been associated with SLE nephritis, possibly based on altered clearance of immune complexes. Variants of the *PTPN22* gene, which encodes a phosphatase that regulates T-cell activation, are also associated with SLE. Genome-wide association studies have expanded the list of lupus-associated gene variants to include regulators of innate immune system activation *(TNFAIP3, ITGAM, IRAK1)* and signaling molecules important in lymphocyte activation *(STAT4, BANK1, BLK,* and *LYN)*. Rare mutations in genes encoding proteins that regulate nucleic acid integrity and degradation, including *TREX1*, encoding a DNase; *SAMHD1*, a triphosphohydrolase; *RNASEH2A, B* and *C*; and *ADAR*, an RNA-specific adenosine deaminase, have been documented in some patients with a lupus-like syndrome called Aicardi-Goutieres syndrome, characterized by skin lesions, CNS disease, autoantibodies, and high levels of interferon. Mutations in these genes have also been documented in rare patients with SLE and have provided new insights into the likely contribution of endogenous nucleic acids to innate immune system activation and lupus pathogenesis.[4] The available data suggest a common theme: the genes that have been associated with lupus confer either increased activation or impaired regulation of the innate or adaptive immune responses, with increased type I interferon often observed in association with the risk genotype.

Environmental Triggers

Several classes of potential environmental triggers for lupus have been studied.[5] Although the female preponderance of SLE implies a role for hormonal factors in the disease, recent concepts describe a possible contribution of epigenetic modification or dosage effects of the X chromosome as accounting for at least some of the sex skewing. A role for microbial triggers—particularly viral infection—has been postulated, consistent with the constitutional symptoms that often characterize the earliest stage of the disease. Epstein-Barr virus has garnered particular interest among investigators, because the frequency of previous infection in SLE patients is significantly higher than in the general population (99 vs. 94%). Evidence of exposure to other viruses, including cytomegalovirus, is equivalent between SLE patients and healthy control subjects. Ultraviolet light exposure is a well-described trigger of lupus flares. Possible mechanisms include DNA damage, induction of cellular stress responses, and induction of apoptosis of skin cells, which result in concentration of nucleic acids and associated proteins in cell membrane blebs and increased processing by antigen-presenting cells. Data also support an association between current tobacco use and anti–double-stranded DNA antibodies and lupus disease activity. Certain drugs, including procainamide and hydralazine, can induce a lupus-like syndrome, but the symptoms usually abate after discontinuing use of the drug. These agents may promote demethylation of DNA, thereby altering gene expression and potentially increasing the availability of immunostimulatory DNA. Sulfa antibiotics have been reported to induce lupus flare in some patients. Administration of recombinant interferon-α to patients with hematologic malignancies or hepatitis C infection has been associated with induction of a lupus-like syndrome. In addition, anti–tumor necrosis factor agents have induced lupus autoantibodies and occasionally clinical lupus in patients with rheumatoid arthritis.

Immunologic Triggers

Genetic and environmental factors that increase the probability of development of SLE are likely to act on the immune system to induce autoimmunity and consequent tissue inflammation and damage. In addition to mechanisms that increase the availability of self-antigens (such as ultraviolet light), altered expression of gene products that mediate or regulate apoptosis, or impaired clearance of apoptotic debris, results in generalized activation of the immune system and contributes to autoimmunity in lupus. In parallel with the events that account for effective immune responses directed at exogenous microbes, the autoimmunity that occurs in SLE patients is likely to require activation of both innate and adaptive immune responses. The innate immune response is first activated by common molecular patterns expressed on the microbe and results in augmented antigen-presenting cell capacity and successful generation of an antigen-specific adaptive immune response. The characterization of the TLR family of pattern recognition receptors has provided new understanding of the mechanisms through which the innate immune system is activated by exogenous and endogenous stimuli, including nucleic acid–containing immune complexes, and promotes induction of a self-directed adaptive immune response.[6]

Type I Interferon

Recent studies of gene expression in peripheral blood mononuclear cells of SLE patients using microarray technology have demonstrated that activation of genes regulated by type I interferon is a common feature in patients with active disease and may represent innate immune system activation. Interferon (IFN)-α may be responsible for many of the immunologic alterations that have been observed in SLE and is identified as a potential therapeutic target. Immune complexes containing DNA or RNA are postulated to contribute to the production of type I interferon in SLE. Demethylated CpG-rich DNA or RNA associated with nucleic acid–binding proteins can activate plasmacytoid dendritic cells and other immune system cells through TLRs and thereby result in the production of type I interferon (IFN-α or IFN-β) and other proinflammatory cytokines (E-Fig. 266-1). Diverse effects of type I interferon on immune system function are consistent with the altered immune responses observed in SLE patients, including maturation of dendritic cells, increased immunoglobulin class switching to mature immunoglobulin isotypes (immunoglobulin [Ig]G and IgA), and induction of soluble mediators that increase B-cell differentiation and inflammatory responses, such as B-lymphocyte stimulator (BLyS) and IFN-γ.[7] Induction of an immunostimulatory microenvironment by IFN-α may support the development of a humoral immune response directed at self-antigens, particularly intracellular particles that contain nucleic acids and nucleic acid–binding proteins. It is not known why some individuals initiate immune system activation directed at self-antigens and others do not. In addition to its effects on immune system function, type I interferon has been associated with altered endothelial cell function and may contribute to the development of atherosclerotic vascular pathology in patients with lupus.[8]

Autoantibodies

The most characteristic lupus autoantibodies target intracellular particles containing both nucleic acid and nucleic acid–binding proteins. Understanding the significance of induction of these particular autoantibody specificities may provide clues to the etiology of SLE. A recent analysis of the spectrum of autoantibodies present in the sera of individuals in whom SLE is later diagnosed has suggested that autoantibodies reactive with certain RNA-binding proteins, including the Ro protein, occur early in the preclinical stage of the disease, along with a positive antinuclear antibody (ANA) test. These are often followed by anti-DNA antibodies and, finally, by the development of antibodies specific for the spliceosomal proteins Smith (Sm) and ribonucleoprotein (RNP) at approximately the time of diagnosis (Fig. 266-1). These observations suggest that individuals who demonstrate progression from humoral immunity targeting proteins associated with RNA to antibodies that bind DNA and other specificities are those in whom sufficient autoimmunity develops to manifest clinical symptoms. Approximately one third of SLE patients have autoantibodies reactive with phospholipids or the proteins associated with them, particularly β_2-glycoprotein I (β_2GPI). These autoantibody specificities can also be present independently of SLE in primary antiphospholipid antibody syndrome (Chapter 174).

Immune Complexes and Complement

Tissue and organ damage in SLE is mediated by the deposition or in situ formation of immune complexes and subsequent complement activation and inflammation. The complement system (Chapter 50), composed of more than 30 proteins that act in concert to protect the host against invading organisms, initiates inflammation and tissue injury. Complement activation promotes chemotaxis of inflammatory cells and generates proteolytic fragments that enhance phagocytosis by neutrophils and monocytes. The classic pathway is activated when antibodies bind to antigen and generate potent effectors. Alternative pathway activation mechanisms differ in that they are initiated by the binding of spontaneously activated complement components

E-TABLE 266-1 GENES ASSOCIATED WITH SYSTEMIC LUPUS ERYTHEMATOSUS

GENE*	PROTEIN
BANK1	B-cell scaffold protein with ankyrin repeats 1
BLK	B-lymphocyte–specific tyrosine kinase
C1QA, *B*, and *C*	Complement component C1q
C2	Complement component 2
C4A and *C4B*	Complement component C4
CRP	C-reactive protein
DRB11501	MHC class II (DR2)
DRB10301	MHC class II (DR3)
FCGR2A	Activating FcγRIIA
FCGR3A	Activating FcγRIIIA
IRF5	Interferon regulatory factor 5
ITGAM	Mac1/complement receptor 3
IRAK1	Interleukin-1 receptor–associated kinase 1
LYN	Lyn tyrosine kinase
PTPN22	Protein tyrosine phosphatase nonreceptor type 22
STAT4	Signal transducer and activator of transcription 4
TNFAIP3	A20
TNFSF4	Ox40 ligand
TREX1	DNase III

*Genes listed have an odds ratio of ≥ 1.3.
MHC = major histocompatibility complex.

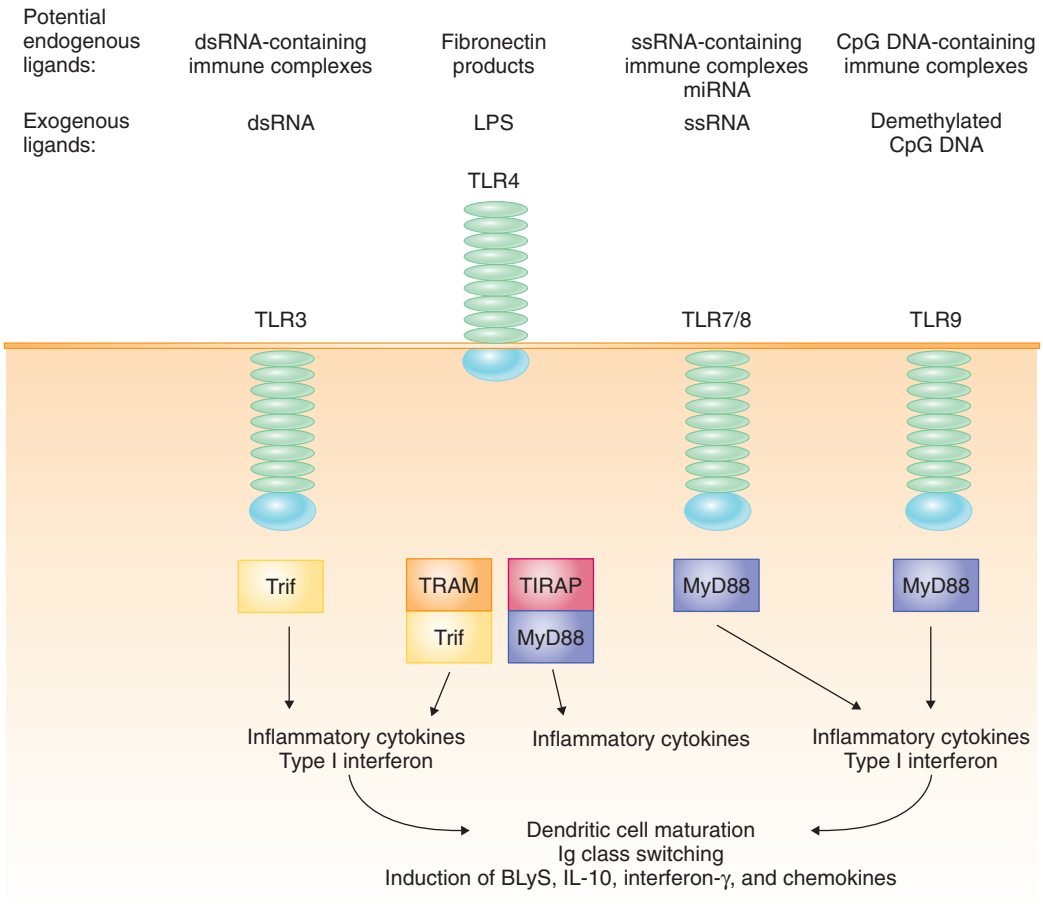

E-FIGURE 266-1. Model for induction of innate immune system activation in systemic lupus erythematosus. Both exogenous and endogenous stimuli can induce toll-like receptor (TLR) activation and thereby result in new gene transcription. Among potential endogenous ligands are immune complexes containing DNA or RNA or matrix-derived components. TLR ligands trigger the activation of intracellular adaptors, including TIR domain-containing adapter-inducing interferon-β (Trif), Trif-related adaptor molecule (TRAM), TIR domain-containing adapter protein (TIRAP), or myeloid differentiation primary response protein 88 (MyD88), and induce transcription of type I interferons or inflammatory cytokines. Type I interferons mediate diverse effects on immune system cells, including maturation of dendritic cells, increased immunoglobulin (Ig) class switching, and induction of cytokines that promote autoimmunity and inflammation, including B-lymphocyte stimulator (BLyS), interleukin-10 (IL-10), interferon-γ, and chemokines. LPS = lipopolysaccharide.

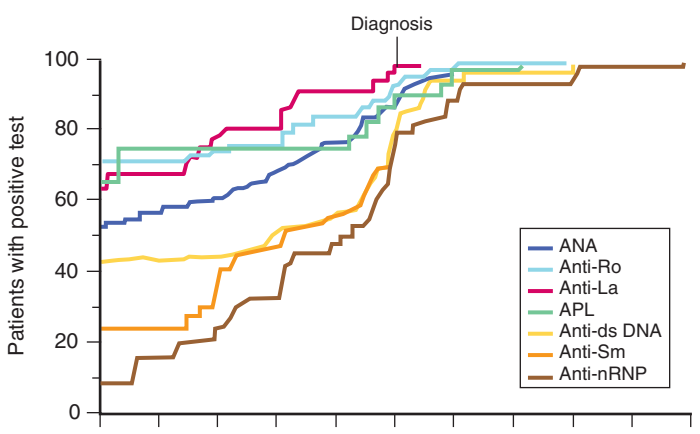

FIGURE 266-1. Proportion of patients with positive antibody tests relative to the time of diagnosis or appearance of the first clinical manifestation of systemic lupus erythematosus (SLE). For each autoantibody, the proportion of patients testing positive relative to the time of diagnosis or to the time of appearance of the first clinical criterion was assessed. In analyses of the time from antibody development to the diagnosis of SLE, antinuclear antibodies (ANAs) appeared significantly earlier than anti-Sm antibodies and antinuclear ribonucleoprotein (anti-nRNP) antibodies, but not significantly earlier than anti-Ro, anti-La, antiphospholipid (APL), or anti–double-stranded DNA antibodies (anti-dsDNA). (From Arbuckle MR, McClain MT, Rubertone MV, et al. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N Engl J Med.* 2003;349:16.)

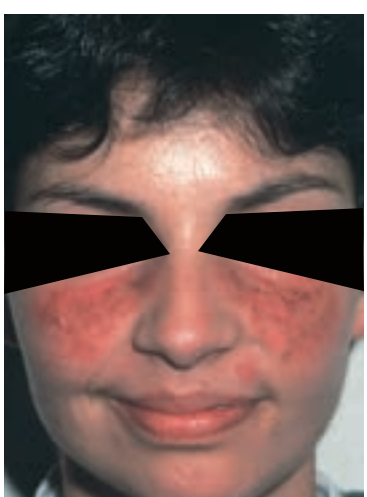

FIGURE 266-2. Malar rash in a patient with systemic lupus erythematosus. Note that the rash does not cross the nasolabial fold. (From Gladman DD, Urowitz MB. Systemic lupus erythematosus: clinical features. In: Klippel JH, Dieppe PA, eds. *Rheumatology.* 2nd ed. London: Mosby; 1998.)

TABLE 266-1 CLINICAL MANIFESTATIONS OF SYSTEMIC LUPUS ERYTHEMATOSUS

MANIFESTATION	APPROXIMATE FREQUENCY (%)
Cutaneous	88
Arthritis/arthralgias	76
Neuropsychiatric	66
Pleurisy/pericarditis	63
Anemia	57
Raynaud's phenomenon	44
Vasculitis	43
Atherosclerosis	37
Nephritis	31
Thrombocytopenia	30
Sensorimotor neuropathy	28
Cardiac valvar disease	18
Pulmonary alveolar hemorrhage	12
Pancreatitis	10
Myositis	5
Myocarditis	5

to the surfaces of pathogens or self-tissues. C3a, an anaphylatoxin that binds to receptors on leukocytes and other cells, causes activation and release of inflammatory mediators. C5a is a potent soluble inflammatory, anaphylatoxic, and chemotactic molecule that promotes recruitment and activation of neutrophils and monocytes and mediates endothelial cell activation through its receptor. The release of reactive oxygen and nitrogen intermediates is an additional mechanism that contributes to tissue damage.

Tissues targeted by immune system activity in lupus include the skin, where immune complexes and complement are deposited in a linear pattern (as demonstrated in the lupus band test, in which deposited antibodies are identified by a fluorescent tag), the glomeruli, and heart valves. Antibodies reactive with hippocampal neurons in the brain can mediate excitotoxic death. Immune and inflammatory mechanisms responsible for the vasculopathy of lupus are multifactorial and not clearly defined. Microvascular damage is observed in splenic arteries and is characterized by the typical onion-skin pattern of concentric connective tissue deposition. In addition to vascular damage mediated by inflammation, thrombosis, including microthrombi, contributes to ischemia and cell necrosis in the brain and other organs.

CLINICAL MANIFESTATIONS

Symptoms and Signs

Constitutional Symptoms

SLE is a disease that involves virtually all components of the immune system and can be accompanied by constitutional symptoms similar to those seen in the setting of microbial infection. Fatigue, headaches, weight loss, and fevers are common, along with generalized arthralgias, myalgias, and lymphadenopathy. The level of activity of lupus typically follows a pattern of flares and remissions, although some patients sustain active disease for prolonged periods. Careful monitoring for the development of major organ system disease is important to ensure timely adjustments in medical therapy.

Cutaneous and Mucous Membranes

The skin and mucous membranes are affected in most lupus patients (Table 266-1). The erythematous facial rash with a butterfly distribution across the malar and nasal prominences and sparing of the nasolabial folds is the classic rash of SLE and is seen in 30 to 60% of patients (Fig. 266-2). The butterfly rash is often triggered by sun exposure, but photosensitivity can also be demonstrated diffusely in other areas of the body.

The discoid skin lesions are erythematous plaques with central scarring and may be covered with scale. These lesions are seen in about 25% of patients, involve the scalp or the face and ears, and may be associated with alopecia. Discoid lesions can be present in the absence of systemic manifestations of SLE (discoid lupus). In addition to the scarring alopecia of discoid lupus, more transient alopecia may be a clinical sign of increased disease activity and is associated with apoptosis of cells in the hair follicle.

Inflammation of the deep dermis and subcutaneous fat can result in lupus panniculitis, with firm painful nodules that sometimes adhere to the epidermis, causing irregularities in the superficial skin. Subacute cutaneous lupus erythematosus is seen in sun-exposed areas and can involve erythematous plaques or psoriasiform lesions. It is associated with autoantibodies to the Ro (SSA) RNA-binding protein. Mucosal ulcerations, especially of the buccal mucosa and upper palate, result from mucositis and are typical of SLE. Manifestations of vasculopathy are also common in SLE, including arteriolar spasm or infarcts in the nail folds, a diffuse lacey pattern over the skin described as livedo reticularis, and petechial-purpuric or urticarial lesions on the extremities. Vasculopathy in SLE is often associated with the presence of antiphospholipid antibodies.

Musculoskeletal System

Arthralgias and nonerosive arthritis are among the most common clinical features of SLE and are experienced by more than 85% of patients. The

proximal interphalangeal and metacarpophalangeal joints of the hand are most commonly symptomatic, along with the knees and wrists. In some patients (≈10%), deformities resulting from damage to periarticular tissue can occur, a condition termed *Jaccoud arthropathy*. The heavy use of corticosteroids in many lupus patients can be accompanied by the development of osteoporosis, including osteoporotic fractures or osteonecrosis, most commonly of the hips, although the underlying vasculopathy can also contribute to joint damage.

Inflammation of the muscles with elevated creatine phosphokinase can occur rarely in SLE, and myopathy may be observed as a consequence of corticosteroid therapy. Fibromyalgia, characterized by painful trigger points at characteristic locations, commonly accompanies SLE and can contribute to fatigue and depression.

Renal System

Kidney involvement in SLE is common, with 74% of patients being affected at some time in the course of disease, and is a poor prognostic indicator. Renal pathology is generally attributed to the deposition of circulating immune complexes or in situ formation of these complexes in glomeruli and results in the activation of complement and subsequent recruitment of inflammatory cells. In addition to glomerular inflammation, necrosis, and scarring, renal pathology is characterized by vascular lesions, including thrombotic microangiopathy and extraglomerular vasculitis. Tubulointerstitial disease, including infiltration of the interstitium with mononuclear cells, tubular atrophy, and interstitial fibrosis, is increasingly recognized as associated with a poor prognosis for persistent nephritis and renal survival. Hypertension may be a consequence of significant renal involvement.

Most cases of lupus nephritis present a complex immunopathologic picture, but in general, the pattern of renal disease reflects the site of deposition of immunoglobulins and the quality of the effector mechanisms they induce. Mesangial deposition of immunoglobulin induces mesangial cell proliferation and is associated with microscopic hematuria and mild proteinuria (Fig. 266-3). Subendothelial deposition of immune complexes results in proliferative and exudative inflammation, together with hematuria, mild to moderate proteinuria, and reduced glomerular filtration rate. Subepithelial deposition of immune complexes adjacent to podocytes and along the glomerular basement membrane can result in membranous nephritis with nephrotic-range proteinuria. In addition, antiphospholipid antibodies may support the development of thrombotic or inflammatory vascular lesions within or external to glomeruli.

A World Health Organization classification of lupus nephritis lesions was first published in 1975, with subsequent revisions. These classifications were reviewed and rigorously reexamined in the revised International Society of Nephrology and Renal Pathology Society classification criteria for lupus glomerulonephritis (GN) (Table 121-7 in Chapter 121, and also E-Table 266-2). Class I and II GN involves mesangial deposition of immune complexes (class I without and class II with mesangial hypercellularity); class III describes focal GN involving less than 50% of total glomeruli; class IV includes diffuse GN involving 50% or more of glomeruli; class V designates membranous lupus nephritis; and class VI is characterized by advanced sclerotic lesions. Classes III and IV have subdivisions for active and sclerotic lesions, and class IV also has subdivisions for segmental and global involvement. Pathologic diagnosis should include descriptions of tubulointerstitial and vascular disease as well as glomerular involvement.

The prognosis of class I and class II disease is usually good, whereas class IV, the most common form of lupus nephritis, has the worst prognosis, particularly when the serum creatinine level is elevated at the time of diagnosis. Class V nephritis occurs in 10 to 20% of patients, and the implication for long-term outcome depends on the degree of proteinuria, with mild proteinuria having a good prognosis and nephrotic syndrome with chronic edema having a more negative prognosis. It should be noted that renal veins can occasionally become involved with thrombosis, which then also contributes to nephrotic syndrome. This complication can be evaluated by renal ultrasound (Chapter 121).

Cardiovascular System

Pericarditis and valve nodules were among the first clinical manifestations described in SLE. It is only recently that the extent of premature atherosclerotic disease has been well documented. Pericarditis (Chapter 77) is the most common cardiac manifestation, but it is sometimes recognized only on imaging studies or at autopsy. It is a component of the generalized serositis that is often a feature of SLE and is associated with local autoantibodies and

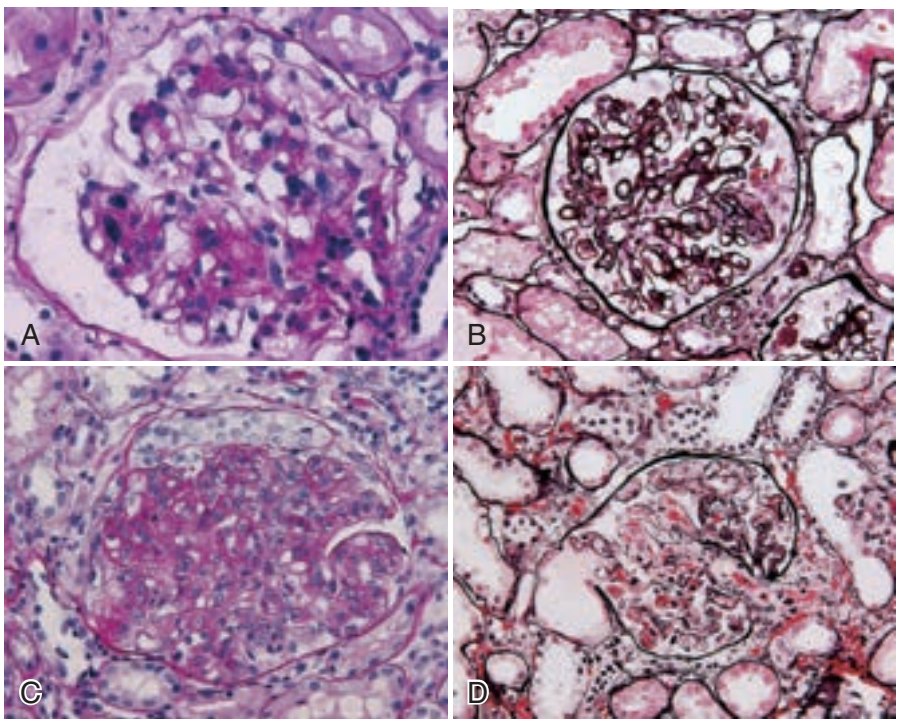

FIGURE 266-3. Histopathology of lupus nephritis. A, Lupus nephritis class II. A light micrograph of a glomerulus shows mild mesangial hypercellularity (periodic acid–Schiff). **B**, Lupus nephritis class III (A). Light micrograph showing a glomerulus with segmental endocapillary hypercellularity, mesangial hypercellularity, capillary wall thickening, and early segmental capillary necrosis (methenamine silver). **C**, Lupus nephritis class IV-G (A/C). A glomerulus manifests global endocapillary proliferation, leukocyte influx and apoptotic bodies, double contours, crescent formation with tubular transformation, early sclerosis, and disruption of Bowman's capsule (periodic acid–Schiff). **D**, Thrombotic microangiopathy in a patient with systemic lupus erythematosus and circulating anticoagulant. A glomerulus shows severe capillary and arteriolar thrombosis, endothelial cell swelling and necrosis, neutrophil influx, and stasis of erythrocytes. No signs of immune deposits were found (methenamine silver). (From Wenning JJ, D'Agati VD, Schwartz MM, et al. The classification of glomerulonephritis in systemic lupus erythematosus revisited. *J Am Soc Nephrol.* 2004;15:241.)

E-TABLE 266-2 INTERNATIONAL SOCIETY OF NEPHROLOGY/RENAL PATHOLOGY SOCIETY 2003 CLASSIFICATION OF LUPUS NEPHRITIS

Class I	**Minimal mesangial lupus nephritis** Normal glomeruli by light microscopy, but mesangial immune deposits by immunofluorescence
Class II	**Mesangial proliferative lupus nephritis** Purely mesangial hypercellularity of any degree or mesangial matrix expansion by light microscopy, with mesangial immune deposits. May be a few isolated subepithelial or subendothelial deposits visible by immunofluorescence or electron microscopy, but not by light microscopy
Class III	**Focal lupus nephritis*** Active or inactive focal, segmental, or global endocapillary or extracapillary glomerulonephritis involving < 50% of all glomeruli, typically with focal subendothelial immune deposits, with or without mesangial alterations
Class III (A)	Active lesions: focal proliferative lupus nephritis
Class III (A/C)	Active and chronic lesions: focal proliferative and sclerosing lupus nephritis
Class III (C)	Chronic inactive lesions with glomerular scars: focal sclerosing lupus nephritis
Class IV	**Diffuse lupus nephritis**† Active or inactive diffuse, segmental or global endocapillary or extracapillary glomerulonephritis Involving ≥ 50% of all glomeruli, typically with diffuse subendothelial immune deposits, with or without mesangial alterations. This class is divided into diffuse segmental (IV-S) lupus nephritis when ≥ 50% of the involved glomeruli have segmental lesions, and into diffuse global (IV-G) lupus nephritis when ≥ 50% of the involved glomeruli have global lesions. Segmental is defined as a glomerular lesion that involves less than half of the glomerular tuft. This class includes cases with diffuse wire loop deposits but with little or no glomerular proliferation.
Class IV-S (A)	Active lesions: diffuse segmental proliferative lupus nephritis
Class IV-G (A)	Active lesions: diffuse global proliferative lupus nephritis
Class IV-S (A/C)	Active and chronic lesions: diffuse segmental proliferative and sclerosing lupus nephritis Active and chronic lesions: diffuse global proliferative and sclerosing lupus nephritis
Class IV-S (C)	Chronic inactive lesions with scars: diffuse segmental sclerosing lupus nephritis
Class IV-G (C)	Chronic inactive lesions with scars: diffuse global sclerosing lupus nephritis
Class V	**Membranous lupus nephritis** Global or segmental subepithelial immune deposits or their morphologic sequelae by light microscopy and by immunofluorescence or electron microscopy, with or without mesangial alterations Class V lupus nephritis may occur in combination with class III or IV, in which case both will be diagnosed Class V lupus nephritis shows advanced sclerosis
Class VI	**Advanced sclerosis lupus nephritis** ≥90% of glomeruli globally sclerosed without residual activity

*Indicate the proportion of glomeruli with active and with sclerotic lesions.
†Indicate the proportion of glomeruli with fibrinoid necrosis and/or cellular crescents.
From Weening JJ, D'Agati VD, Schwartz MM, et al. The classification of glomerulonephritis in systemic lupus erythematosus revisited. *J Am Soc Nephrol.* 2004;15:241.

immune complexes. Pericarditis is usually manifested as substernal chest pain that is improved by bending forward and can be exacerbated by inspiration or coughing. The symptoms and effusions associated with pericarditis are quite responsive to moderate-dose (20 to 30 mg/day of prednisone) corticosteroid treatment.

Structural valve abnormalities in SLE range from sterile nodules (originally described by Libman and Sacks) to nonspecific valve thickening (Chapter 75). The nodules are immobile and usually located on the atrial side of the mitral valve and sometimes on the arterial side of the aortic valve. Right-sided lesions are rare. These structural changes may in some cases result in valvular regurgitation. Although valve nodules are detected in most patients with SLE at autopsy, clinically significant valvular heart disease is much less common (1 to 18%). The verrucous valvular lesions of Libman and Sacks are most likely inflammatory in nature and may be associated with the presence of antiphospholipid antibodies.

Premature and accelerated atherosclerosis is prevalent in lupus patients, and preclinical atherosclerotic carotid plaque has been documented in 37% of SLE patients as opposed to 15% of age- and sex-matched controls. Traditional cardiovascular risk factors apply, but the diagnosis of SLE is itself a significant risk factor for premature atherosclerosis (Chapter 70). Among the lupus-related mechanisms that confer additional risk for atherosclerosis, IFN-α and oxidative modification of lipid-associated proteins contribute to the accumulation of vascular damage.[9] Mortality from atherosclerosis may be up to 10 times greater in patients with SLE than in age- and sex-matched controls.

Although not specific to SLE, Raynaud's phenomenon (Chapter 80), characterized by episodic vasospasm and occlusion of the digital arteries in response to cold and emotional stress, is a feature in up to 60% of SLE patients and contributes to pain and sometimes necrosis of the distal ends of extremities. The character of the digits classically changes from pallor to cyanosis and then to rubor as vascular perfusion becomes impaired and then reperfusion ensues. In addition, small arteries, arterioles, and capillaries can be affected by vasculitis and fibrinoid necrosis, with clinical manifestations that include periungual telangiectases, abdominal pain, and neuropsychiatric symptoms.

Pulmonary System

Pleuritis is the most frequent manifestation of pulmonary involvement in SLE and occurs in about 30% of patients at some point in their disease course. Pleuritis is characterized by pain on respiration and exudative effusions (Chapter 99). Parenchymal disease is less common but may be based on several distinct mechanisms, including pneumonitis in the absence of documented infection and sometimes involving alveolar hemorrhage (in up to 12% of patients), pulmonary embolism secondary to venous thrombosis, or pulmonary hypertension with increased pulmonary resistance and impaired diffusing capacity.

Neuropsychiatric Involvement

Clinical features of SLE that involve the nervous system include both neurologic and psychiatric manifestations. The central and peripheral nervous systems can be affected by the disease. The American College of Rheumatology has identified 19 neuropsychiatric syndromes that can be associated with SLE, and validation of these neuropsychiatric findings has been substantiated in several independent studies. The most common manifestations that are probably attributable to SLE cerebritis include cognitive dysfunction, present in 17 to 66% of SLE patients; psychosis or mood disorder, the former reported in up to 8% of patients; cerebrovascular disease in 5 to 18% of patients; and seizures, present in 6 to 51% of patients. Headaches are also common. Because none of these CNS manifestations is found exclusively in SLE, it can be difficult to be certain that a neuropsychiatric complaint or symptom should be attributed to SLE.

Evaluation of neuropsychiatric lupus depends on a careful clinical history and physical and laboratory examinations and, in some cases, imaging studies and analysis of cerebrospinal fluid to rule out infection. Magnetic resonance imaging is useful for detecting intracranial abnormalities, which are seen in 19 to 70% of patients and include white matter lesions, cerebral infarction, venous sinus thrombosis, and sometimes atrophy. More sophisticated imaging techniques such as magnetic resonance angiography and magnetic resonance spectroscopy can be used to assess cerebral blood flow or neuronal metabolism.

Cranial nerve and ocular involvement, most likely based on vasculopathy and focal ischemia, can sometimes affect vision. Ocular examination of the retina can reveal cotton-wool spots as a result of retinal ischemia or necrosis. Although rare, transverse myelopathy, frequently associated with antiphospholipid antibodies, can have devastating consequences, including paraplegia. Sensorimotor neuropathies, often asymmetrical, are more common (up to 28%) and are based on damage to small nerve fibers with vasculopathy in the small arteries that supply the nerve fibers.

As is the case with lupus nephritis, the pathophysiologic mechanisms that account for the neuropsychiatric manifestations of SLE are diverse and complex. Recent data suggest that autoantibodies cross-reactive with neuronal cell surface glutamate receptors and DNA may mediate excitotoxic death of neurons and are proposed to contribute to cognitive dysfunction. Antibodies directed against ribosomal P protein have also been associated with neuropsychiatric lupus, and antiphospholipid antibodies can contribute to a procoagulant state, vascular thrombosis, and cerebral ischemia. Cerebral vasculopathy has been clearly demonstrated by angiographic and pathologic studies. Noninflammatory small vessel vasculopathy is the most common lesion and can be associated with microinfarcts. Inflammatory mediators, including the cytokines interleukin (IL)-6 and IFN-α, and matrix metalloproteinases may also contribute to the neuropsychiatric manifestations of SLE.

Gastrointestinal System

Although uncommon, vasculitis of the gastrointestinal tract or mesentery can result in pain and bowel necrosis. Less common than pleuritis and pericarditis, peritonitis can manifest as peritoneal effusion and abdominal pain. Pancreatitis occurs in less than 10% of patients but may also be due to vascular pathology. Lupoid hepatitis, a syndrome that was named for the presence of positive ANAs in patients with chronic active hepatitis, is a misnomer because elevated transaminases are only rarely seen in lupus patients.

Lymphadenopathy

About one third of SLE patients demonstrate diffuse lymphadenopathy at some time during the course of their disease. The nodes are often nontender, and lymphoma is sometimes considered in the differential diagnosis. Biopsy usually reveals follicular hyperplasia, although some histopathologic findings appear similar to the histiocytic necrotizing lymphadenitis that is a feature of Kikuchi's disease, a self-limited syndrome characterized by fever and lymphadenopathy. Recent multicenter studies have determined the frequency of malignancies in patients with SLE and have found a significant increase in hematologic malignancies, particularly non-Hodgkin's lymphoma. Splenomegaly is sometimes seen in SLE, and spleen pathology is characterized by a classic onion-skin histology that appears as concentric circles of collagen matrix surrounding splenic arteries and arterioles.

Hematologic System

In addition to autoantibody specificities that are fairly specific for SLE (anti-DNA, anti-Sm), antibodies that target each of the cellular blood elements are also common. Anemia is present in about 50% of patients and is multifactorial. It can be associated with a positive Coombs test or microangiopathic hemolysis (Chapter 160) or reflect chronic disease (normochromic, normocytic) (Chapter 158). Leukopenia, particularly lymphopenia, is observed, with the lymphocyte count decreasing in the setting of increased disease activity. Antibodies that bind to lymphocytes and neutrophils have been described, and an increased tendency for lymphocytes to undergo spontaneous apoptosis may contribute to lymphopenia. Idiopathic thrombocytopenic purpura (Chapter 172) can be an early manifestation of SLE, and thrombocytopenia induced by antiplatelet autoantibodies can sometimes lead to a life-threatening risk for hemorrhage. Autoantibodies to clotting factors can also occur and contribute to impaired clot formation and hemorrhage.

Lupus Pregnancy and Neonatal Lupus

Whether pregnancy increases the likelihood of lupus exacerbation has been debated, with differences on this point presented by different investigators. However, abundant data indicate that patients with SLE have worse fetal outcomes than healthy individuals.[10] Gestational hypertension, fetal growth restriction, and fetal distress are increased in patients with SLE and may lead to fetal loss or premature delivery. Preeclampsia can contribute to a poor outcome in both the mother and fetus and can be difficult to distinguish from a lupus flare associated with lupus nephritis.

Neonatal lupus is a distinct entity that can occur in infants of mothers with or without a diagnosis of SLE.[11] The syndrome is characterized by cutaneous lesions and congenital heart block in the infant and the presence of antibodies

TABLE 266-2 UPDATE OF THE 1982 REVISED CRITERIA FOR CLASSIFICATION OF SYSTEMIC LUPUS ERYTHEMATOSUS

CRITERION*	DEFINITION
1. Malar rash	Fixed erythema, flat or raised, over the malar eminences that tends to spare the nasolabial folds
2. Discoid rash	Erythematous raised patches with adherent keratotic scaling and follicular plugging; atrophic scarring may occur in older lesions
3. Photosensitivity	Rash as a result of unusual reaction to sunlight, by history or physician observation
4. Oral ulcers	Oral or nasopharyngeal ulceration, usually painless, observed by a physician
5. Arthritis	Nonerosive arthritis involving 2 or more peripheral joints and characterized by tenderness, swelling, or effusion
6. Serositis	A. Pleuritis—convincing history of pleuritic pain or rubbing heard by a physician or evidence of pleural effusion *or* B. Pericarditis—documented by electrocardiography, a rub, or evidence of pericardial effusion
7. Renal disorder	A. Persistent proteinuria > 0.5 g/day or > 3+ if quantitation is not performed *or* B. Cellular casts—may be red cell, hemoglobin, granular, tubular, or mixed
8. Neurologic disorder	A. Seizures—in the absence of offending drugs or known metabolic derangements (e.g., uremia, ketoacidosis, electrolyte imbalance) *or* B. Psychosis—in the absence of offending drugs or known metabolic derangements (e.g., uremia, ketoacidosis, electrolyte imbalance)
9. Hematologic disorder	A. Hemolytic anemia—with reticulocytosis *or* B. Leukopenia—<4000/mm^3 total on 2 or more occasions *or* C. Lymphopenia—<1500/mm^3 on 2 or more occasions *or* D. Thrombocytopenia—<100,000/mm^3 in the absence of offending drugs
10. Immunologic disorder	A. Deleted in 1997 update B. Anti-DNA: antibody to native DNA in abnormal titer *or* C. Anti-Sm: presence of antibody to Sm nuclear antigen *or* D. Positive finding of antiphospholipid antibodies based on (1) an abnormal serum level of immunoglobulin (Ig)G or IgM anticardiolipin antibodies, (2) a positive test for lupus anticoagulant using a standard method, or (3) a false-positive serologic test for syphilis known to be positive for at least 6 months and confirmed by *Treponema pallidum* immobilization or the fluorescent treponemal antibody absorption test (modified in the 1997 update)
11. Antinuclear antibody	An abnormal titer of antinuclear antibody by immunofluorescence or an equivalent assay at any point in time and in the absence of drugs known to be associated with "drug-induced lupus" syndrome

***The proposed classification is based on 11 criteria. For the purpose of identifying patients in clinical studies, a person shall be said to have systemic lupus erythematosus if any 4 or more of the 11 criteria are present, serially or simultaneously, during any interval of observation.**

Modified from Tan EM, Cohen AS, Fries JF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* 1982;25:1271.

to the Ro (SSA) or La (SSB) RNA-binding proteins (or both) in the mother. Mortality in babies with a congenital heart block is 15 to 31%. Deposition of anti-Ro IgG in the fetal heart, indicative of transplacental transfer of maternal autoantibody, and dense connective tissue encompassing the conduction system have been demonstrated in autopsy specimens. Prenatal testing of lupus mothers for the presence of anti-Ro and anti-La antibodies is appropriate, and careful monitoring with fetal echocardiography starting at week 16 of pregnancy can detect conduction defects. Fluorinated corticosteroids such as dexamethasone have been effective in reversing heart block in some cases. The role of hydroxychloroquine in prevention of neonatal lupus manifestations is under investigation.

Antiphospholipid Antibody Syndrome

Antiphospholipid antibodies represent a distinct class of autoantibodies that are seen in about one third of SLE patients but can also be present in individuals who do not carry a diagnosis of SLE (Chapter 171). Although these antibodies were initially thought to be specific for phospholipids exposed in cell membranes, particularly after "flipping" of the membranes of apoptotic cells, extensive data support their primary reactivity with phospholipid-binding proteins, particularly β_2GPI. Whether in primary antiphospholipid syndrome or in SLE, antiphospholipid antibodies have been associated with venous and arterial thromboses.[12] In addition to vascular thromboses, clinical manifestations of antiphospholipid syndrome include thrombotic microangiopathic glomerular disease, cardiac valve lesions, livedo reticularis, thrombocytopenia, hemolytic anemia, and CNS disease. Recent data indicate that these autoantibodies can contribute to fetal loss and growth restriction by binding to the placenta, activating the complement system, and inducing inflammation. Catastrophic antiphospholipid syndrome, triggered by the acute onset of multisystemic (three or more organs) thrombosis, is resistant to anticoagulation treatment and is fatal in approximately 50% of cases.[13]

DIAGNOSIS

Classification

Criteria for the classification of patients with SLE for the purpose of clinical studies were developed by the American College of Rheumatology, with the most recent full revision published in 1982 and an update published in 1997 (Table 266-2). The criteria include 11 features that encompass manifestations of skin and mucosal involvement, arthritis, serositis, renal disorder, neurologic disorder, hematologic disorder, immunologic disorder, and an abnormal titer in the ANA test, with at least four criteria required for classification as SLE. ANA has low specificity but strengthens the sensitivity of the criteria because it is positive in virtually all lupus patients. These criteria are not intended for use as diagnostic criteria, because more than 50% of patients with SLE do not meet four criteria at any point in time, although all do meet these criteria at some point during the course of the disease. The criteria are useful in reminding the clinician of the most characteristic features of SLE, but a careful history with detailed review of systems and triggering factors, as well as a family history, is essential in raising suspicion for a diagnosis of SLE. Because drugs can trigger a lupus-like syndrome, a careful drug history should be taken. Procainamide and hydralazine present the greatest risk for development of lupus, with quinidine, isoniazid, minocycline, and recombinant interferon-α presenting a lower risk. At the onset of clinical symptoms, the diagnosis of SLE can be uncertain because many of the systemic manifestations of lupus can mimic other conditions, particularly viral infections or malignancy, and only some of the typical clinical symptoms may be expressed at any one point in time. Important features of SLE are its multisystemic nature and characteristic serology. The differential diagnosis of SLE includes other rheumatic disorders, such as rheumatoid arthritis and vasculitis; infections, including gonococcal arthritis, parvovirus B19, and mononucleosis; inflammatory bowel disease; thrombotic thrombocytopenic purpura; drug reactions; and malignancies, particularly lymphoma. It should be recognized that the clinical manifestations of lupus can demonstrate overlap with those of other autoimmune rheumatic diseases and can evolve over time. Many genetic, environmental, and immunologic factors associated with lupus are also associated with other systemic autoimmune diseases, often contributing to a complex clinical picture.

Laboratory Findings

Laboratory tests can be very helpful in supporting the diagnosis of SLE. All cellular elements of blood can be affected in lupus, so the complete blood

count is an essential test that aids in diagnosis and management. A prolonged activated partial thromboplastin time (aPTT) can indicate the presence of pathogenic antiphospholipid antibodies (Chapter 171). These antibodies are also associated with a false-positive result in the serologic test for syphilis, an observation that is mainly of historical interest.

Evaluation of renal disease in SLE includes urinalysis with microscopic analysis of urine sediment, serum blood urea nitrogen and creatinine, and 24-hour urine collection (or alternatively, spot urine protein-to-creatinine ratio) for estimation of protein and creatinine clearance. Low serum albumin would be consistent with persistent proteinuria and membranous GN, whereas red and white blood cell casts in the urinary sediment suggest proliferative GN. Although a renal biopsy is usually performed only when the result may influence therapeutic decisions, pathologic classification of the features of renal disease can provide prognostic information.

The erythrocyte sedimentation rate (ESR), although a very nonspecific indicator of systemic inflammation, is often monitored and in many patients can provide an indication of disease activity. Interestingly, C-reactive protein, an acute phase reactant, is relatively uninformative in SLE because it is often low in comparison to an ESR performed on the same occasion.

Assaying and monitoring characteristic lupus serologic tests can strongly support the diagnosis of SLE and, in some cases, can assist in the assessment of disease activity. The ANA test is positive in virtually all patients and does not need to be repeated once it has been documented to be positive (Table 266-3). Anti–double-stranded DNA antibodies are common in SLE, and some studies have found that monitoring their titer can be useful in assessing the activity of lupus nephritis. Autoantibodies specific for proteins that associate with nucleic acids in intracellular particles are present in many patients and can provide support for a diagnosis of SLE. Anti-Sm antibodies are highly specific for SLE and, along with anti-RNP antibodies, react with the spliceosome particle. Anti-Ro (SSA) and anti-La (SSB) antibodies are specific for proteins in an RNA-containing particle and are common in patients with Sjögren's syndrome and in mothers of babies with neonatal lupus, as well as being a feature of SLE. It is useful to document the presence of anti-Sm, anti-RNP, anti-Ro, and anti-La antibodies when the diagnosis of SLE is being made, but the titers of these autoantibodies are not helpful in monitoring disease activity.

Proteins of the complement system are activated by immune complexes, such as those that form in patients with SLE. The activation products that result from enzymatic cleavage of the complement components promote inflammation directly by binding to cell surface receptors on mononuclear phagocytes, and indirectly by acting as chemotactic agents to recruit inflammatory cells. Decreased levels of two of the more stable complement components, C3 and C4, are typically measured in serum, and decreased C3 and C4 levels are often indicators of enhanced consumption and increased disease activity. Some laboratories also use a functional measure of total hemolytic complement activity (CH_{50}).

It is the global picture provided by a careful history, physical examination, and blood, urine, and serologic data that supports a diagnosis of SLE. It should be recognized that there is considerable heterogeneity among patients and that different combinations of clinical features will characterize any one individual. As is the case with many systemic diseases, infection and some malignancies may have a similar picture and should be included in the differential until the diagnosis of SLE is secure.

TABLE 266-3 AUTOANTIBODIES ASSOCIATED WITH SYSTEMIC LUPUS ERYTHEMATOSUS

TARGET ANTIGEN	APPROXIMATE FREQUENCY
Nuclear antigens	99
dsDNA	70
Sm	38
RNP (U1-RNP)	33
Ro (SSA)	49
La (SSB)	35
Phospholipids	21
Ribosomal P	10

TREATMENT

Although current knowledge of genetic risk factors for SLE is not sufficient to predict those in whom the disease will develop, once the diagnosis has been made, regular counseling and education are fundamental to the treatment of SLE patients. Patients should be advised to avoid known triggers of disease exacerbation, such as ultraviolet light, and should be instructed regarding the need for adequate rest. Pregnancy should be undertaken with caution and with careful monitoring. Lupus patients can be informed that data indicate that oral contraceptive agents and estrogen replacement therapy do not contribute to disease exacerbations.

Conventional Medical Therapy

Clinical manifestations of lupus that do not involve major organ systems can often be managed with nonsteroidal anti-inflammatory drugs, low-dose corticosteroids, and antimalarials. Corticosteroids (Chapter 35) are immunosuppressive agents that modulate many functions of lymphocytes and monocytes, including the production of proinflammatory cytokines. Oral prednisone in doses ranging from 5 to 30 mg daily is effective in treating constitutional symptoms, arthralgias, pericarditis and pleuritis, and skin disease. Topical corticosteroids are sometimes applied to cutaneous lesions. Although effective, corticosteroids also have toxicities that add to the morbidity associated with lupus. The broad immunosuppression mediated by these drugs contributes to the susceptibility to infection that is an inherent feature of SLE. Osteonecrosis, osteoporotic fractures, posterior subcapsular cataracts, diabetes, myopathy, hypertension, hypoadrenalism, and emotional disturbance are additional deleterious effects of corticosteroids.

Antimalarial agents, most commonly hydroxychloroquine administered at 200 to 400 mg/day, have long been used to control skin involvement and arthralgias and are now routinely used in most lupus patients. An important Canadian study demonstrating an increased frequency of disease flare in patients who discontinued hydroxychloroquine contributed to its recent use in lupus for a broader range of clinical manifestations. Hydroxychloroquine has been associated with a decreased incidence of thrombosis, a mechanism that could affect vasculopathy and end-organ damage. An additional potential mechanism of action implicates the TLR pathway, which is responsible for activation of the innate immune response.[14] The effects of antimalarial agents on acidification of the intracellular vesicles where TLRs associate with their ligands may inhibit immune cell activation mediated by stimulatory nucleic acids. Antimalarials are well tolerated. Because they can rarely cause eye toxicity, ophthalmologic examinations should precede initiation of therapy and take place every 6 to 12 months.

Management of Serious Organ-System Disease

For more serious disease, particularly active nephritis, CNS disease, or systemic vasculitis, prednisone at 60 mg daily or 1 g of intravenous methylprednisolone administered daily for 3 days can sometimes gain control of disease activity. In many situations, additional immunosuppressive, cytotoxic, or biologic therapies are required. Because lupus nephritis is the most common severe clinical organ-system manifestation of SLE, an algorithm describing guidelines for medical management of lupus nephritis developed by a Task Force Panel assembled by the American College of Rheumatology is useful (Fig. 266-4).[15] A similar approach can be applied to other serious clinical disease flares. Lupus flares involving rapid decompensation of renal function, CNS disease (including seizures, strokes, or psychosis), or widespread vasculitis or vasculopathy can be life-threatening and must be recognized and treated early and aggressively.[16] Careful attention to monitoring for concurrent or superimposed infection is an important priority during management of a severe lupus flare, and distinguishing sepsis from active lupus or catastrophic antiphospholipid syndrome can be a particular challenge. In general, cyclophosphamide is added to high-dose corticosteroid therapy in the setting of severe flare, although mycophenolate mofetil has gained increased use in situations where the physician or patient wishes to avoid the potential toxicities associated with cyclophosphamide. Randomized controlled studies have investigated options for maintaining improvement in those patients who respond to induction therapy in the setting of lupus nephritis flare, with current data favoring mycophenolate mofetil over azathioprine. One biologic agent, belimumab, a monoclonal antibody reactive with BLyS, has been approved by the U.S. Food and Drug Administration (FDA) for treatment of active autoantibody-positive SLE in conjunction with standard therapies, but belimumab has not yet been tested in patients with lupus nephritis or other severe manifestations of lupus. Agents used in management of lupus are reviewed below.

Immunosuppressive Agents

Alkylating Agents

Approximately 33% of lupus patients receive cytotoxic therapy during the course of their disease. Cyclophosphamide is a cytotoxic agent that has been one of the more reliable and studied treatments for severe organ system manifestations of lupus, particularly lupus nephritis and CNS involvement.

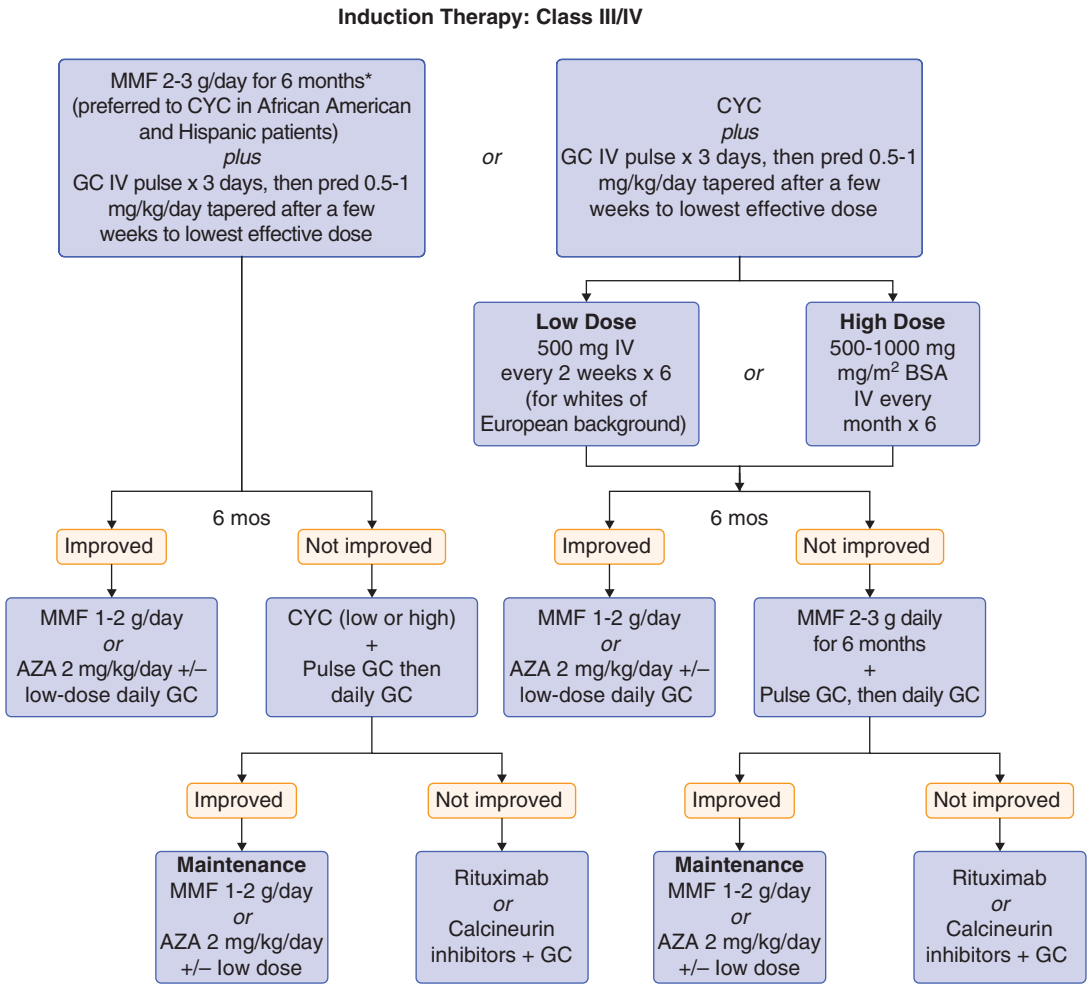

FIGURE 266-4. Algorithm for induction therapy for lupus nephritis. Guidelines developed by the Task Force Panel of the American College of Rheumatology for management of class III/IV lupus nephritis. Refer to reference 15 for guidelines for management of class III/IV lupus nephritis with crescents and class V membranous lupus nephritis without proliferative changes and nephrotic range proteinuria. *The Task Force Panel preferred MMF over CYC in patients who desire to preserve fertility. AZA = azathioprine; BSA = body surface area; CYC = cyclophosphamide; GC = glucocorticoid; IV = intravenous; MMF = mycophenolate mofetil; pred = prednisone. (From Hahn BH, McMahon MA, Wilkinson A, et al. American College of Rheumatology guidelines for screening, treatment, and management of lupus nephritis. *Arthritis Care Res (Hoboken)*. 2012;64:797-808.)

Studies performed at the National Institutes of Health in the 1980s led to recommendations of a standard regimen of cyclophosphamide, 0.5 to 1 g/m^2 body surface area administered intravenously monthly for 6 months, followed by quarterly doses through 2 years. Cyclophosphamide is usually given with oral prednisone in tapering doses or sometimes with pulse methylprednisolone. Although this regimen is often effective in controlling GN, overall patient survival has not been demonstrated to be increased, and cyclophosphamide is associated with significant toxicity, including cytopenia, infection, gonadal failure, and malignancy. Recent clinical studies have included modified immunosuppressive regimens, such as a 6-month induction followed by maintenance with less toxic immunosuppressive agents, including azathioprine or, even better, mycophenolate mofetil (MMF). Cyclophosphamide is relatively contraindicated in pregnant women.

Purine Synthesis Inhibitors

Azathioprine has been used for the treatment of lupus nephritis and as a steroid-sparing agent in SLE for many years. Azathioprine inhibits DNA synthesis and inhibits key signaling pathways in T lymphocytes. Azathioprine is commonly dosed at 2 to 3 mg/kg/day administered as a tablet. Toxicities of azathioprine target the bone marrow (and result in cytopenias) as well as the liver, occasionally resulting in transaminitis. Its use is rarely associated with non-Hodgkin's lymphoma (Chapter 185). It has been used safely in pregnant women.

MMF is an inhibitor that binds to the isoform of inosine monophosphate dehydrogenase that mediates purine synthesis in activated lymphocytes. It has a good track record of utility in inhibiting allograft rejection. Recent clinical trials have compared MMF with intravenous cyclophosphamide for induction therapy in lupus nephritis; the results demonstrated equivalence of MMF and cyclophosphamide in patients with lupus nephritis.[■] In a randomized trial, MMF (1 g twice daily) was more effective than azathioprine for the maintenance treatment of lupus nephritis, with treatment failure rates reduced from 32 to 16%.[■]

Methotrexate

Methotrexate is a folate antagonist that is commonly used in rheumatoid arthritis. A double-blind randomized placebo-controlled trial of oral methotrexate (15 to 25 mg/week for 6 months) in SLE controlled disease and allowed tapering of prednisone. The most responsive clinical manifestations were cutaneous and articular.

Ancillary and Other Therapies

Intravenous Gamma Globulin

Although positive data from controlled trials of intravenous gamma globulin are not available, case reports and clinical experience indicate that administration of pooled IgG fractions can sometimes be efficacious in gaining control of lupus disease activity that is refractory to other therapies. A common regimen is 2 g/kg in divided doses over a 3- to 5-day period. Several mechanisms have been proposed for this therapy, including blockade of Fc receptors, modulation of lymphocyte function through Fc receptors, increased catabolism of pathogenic immunoglobulin, and actions of the anti-idiotype antibody that is a component of the administered IgG.

Plasmapheresis

Removal of pathogenic antibodies and immune complexes is the goal of plasmapheresis, but there are scant data supporting the utility of this therapy. Nonetheless, plasmapheresis has been occasionally useful in lupus patients with life-threatening complications in which the clinical manifestations can be clearly attributed to pathogenic autoantibodies. In particular, plasmapheresis has been effective in cases of thrombotic thrombocytopenic purpura associated with SLE (Chapter 172).

Biologic Therapies

Biologic therapies (Chapter 36) are being actively investigated in clinical trials, but only one agent has been successful as yet in phase III studies. Belimumab, a monoclonal antibody, blocks a B-cell survival and differentiation

signal. At a dose of 1 mg/kg or 10 mg/kg intravenously on days 1 and 28 and then every 28 days for 48 weeks, belimumab reduced disease activity by both validated measures and physician global assessment and decreased autoantibody levels.[A9] In a second study, belimumab at a dose of 10 mg/kg administered intravenously on days 0, 14, and 28 and then every 28 days, the currently recommended regimen, for 72 weeks, added to standard therapy, significantly improved response rate, disease activity, and severe flares, and was generally well tolerated in SLE.[A10] The clinical settings that are most appropriate for the use of belimumab, which was approved by the FDA, will be determined by future clinical trials and clinical experience. Additional agents under investigation target B cells, T-cell activation, and cytokines.[17] Rituximab, a monoclonal antibody specific for the cell surface B-cell molecule CD20, is approved for use in B-cell lymphomas and has been used in some patients with SLE who are poorly responsive to other therapies. Rituximab depletes B cells, often for many months, and may limit T-cell activation by eliminating activated B cells that can serve as antigen-presenting cells. However, controlled clinical trials of rituximab in lupus have not shown efficacy. In a randomized double-blind placebo-controlled phase III trial in patients with active proliferative lupus nephritis, rituximab therapy led to more responders and greater reductions in anti-dsDNA and C3/C4 levels, but did not improve clinical outcomes after 1 year of treatment.[A11] Other agents that block B-cell survival and differentiation factors or target B-cell surface molecules are under study. T-cell targets include CD28 and inducible costimulator (ICOS), a T-cell surface molecule in the same molecular family as CD28 and CTLA4. Blockade of T-cell activation by CTLA4-Ig, a soluble inhibitor of CD28 ligation, or inhibition of ICOS-ICOS ligand interaction might inhibit B-cell differentiation in the germinal center. Monoclonal antibodies specific for most isoforms of IFN-α are under study and represent a rational approach in view of the strong genetic and immunologic role of type I interferon in many SLE patients.[18] IL-6 and IFN-γ represent additional cytokine targets under study.

Adjunctive Therapies

In addition to controlling autoimmunity and inflammation in SLE, it is essential to control hypertension adequately when it occurs. In those with lupus nephritis, treatment with an angiotensin inhibitor or an angiotensin receptor blocker reduces intraglomerular pressure, thus reducing proteinuria. In patients with a history of thrombosis, who will usually have antiphospholipid antibodies, long-term warfarin is recommended. The potential use of statins in lupus is of interest because those agents have anti-inflammatory as well as lipid lowering effects, but statins have not yet been shown efficacious in controlling lupus disease activity in controlled trials. The role of vitamin D supplementation in lupus is under study.

Future Directions

Recent advances in basic immunology, together with detailed molecular and clinical characterization of cohorts of SLE patients, have directed attention to the role adjuvant-like factors play in activating the innate immune response through TLRs. The primary triggers of that response are not known, but abundant data support production of type I interferon as an important consequence of immune activation that has an impact on many aspects of lymphocyte function, probably including induction of self-antigen–specific immune responses. In addition to the biologic therapies currently under study that target T and B lymphocytes, future therapies may be designed that can inhibit TLR activation, signaling components downstream from TLRs, cytoplasmic sensors of stimulatory nucleic acids, or type I interferon itself. Recognition that the complement system is an essential contributor to inflammation triggered by antiphospholipid antibodies, as well as by immune complexes, provides additional targets that might be inhibited therapeutically and limit tissue damage. Continued investigation of the genetic and environmental factors that contribute to disease susceptibility may permit identification of individuals at risk for the development of SLE and may elucidate the primary stimuli that lead to autoimmunity. Identification of informative biomarkers that reflect or even predict disease flares would improve medical management of lupus patients.

PROGNOSIS

Although survival of patients with a diagnosis of SLE is good, lupus remains a disease that is potentially fatal. SLE demonstrates a bimodal pattern of death, with deaths within the first year attributable to active lupus and infection and late deaths attributable to atherosclerotic cardiovascular disease. Recent cohort studies have estimated 5-year survival rates greater than 90%, with improvement in medical management probably contributing to improved outcomes, as opposed to earlier studies, and 85% survival rates at 10 years. However, once a diagnosis of SLE has been made, prolonged remission is rare. Of 702 patients registered in a lupus clinic in Canada, 6.5% achieved complete remission (score of 0 on the SLE Disease Activity Index), and only 1.7% maintained remission for at least 5 years with no treatment. The presence of any permanent organ damage within the first year after a diagnosis of SLE is associated with poorer survival at 10 years (a 75 vs. 95% rate in those without permanent organ damage). Regarding renal outcome, an elevated level of serum creatinine at the time of diagnosis has been correlated with an adverse outcome.

Recent studies of minority populations in the United States have indicated that predictors of high lupus disease activity include Hispanic Texan and African American ethnicities, lack of health insurance, and poor social support. African admixture and anti–double-stranded DNA antibodies also predicted high levels of disease activity, as did previous disease activity.

Data from a multicenter study of nearly 10,000 patients has supported an increased risk for hematologic malignancies in SLE patients, particularly non-Hodgkin's lymphoma. Prognostic factors for an adverse fetal outcome in pregnant lupus mothers are maternal renal disease and hypertension.

Grade A References

A1. Petri M, Kim MY, Kalunian KC, et al. Combined oral contraceptives in women with systemic lupus erythematosus. *N Engl J Med*. 2005;353:2550-2558.

A2. Sanchez-Guerrero J, Gonzalez-Perez M, Durand-Carbajal M, et al. Menopause hormonal therapy in women with systemic lupus erythematosus. *Arthritis Rheum*. 2007;56:3070-3079.

A3. Houssiau FA, Vasconcelos C, D'Crus D, et al. The 10-year follow-up data of the Euro-Lupus Nephritis Trial comparing low-dose and high-dose intravenous cyclophosphamide. *Ann Rheum Dis*. 2010;69:61-64.

A4. Arends S, Grootscholten C, Derksen RH, et al., Dutch Working Party on systemic lupus erythematosus. Long-term follow-up of a randomised controlled trial of azathioprine/methylprednisolone versus cyclophosphamide in patients with proliferative lupus nephritis. *Ann Rheum Dis*. 2012;71:966-973.

A5. Houssiau FA, D'Cruz D, Sangle S, et al. MAINTAIN Nephritis Trial Group. Azathioprine versus mycophenolate mofetil for long-term immunosuppression in lupus nephritis: results from the MAINTAIN Nephritis Trial. *Ann Rheum Dis*. 2010;69:2083-2089.

A6. Feng L, Deng J, Huo DM, et al. Mycophenolate mofetil versus azathioprine as maintenance therapy for lupus nephritis: a meta-analysis. *Nephrology (Carlton)*. 2013;18:104-110.

A7. Appel GB, Contreras G, Dooley MA, et al. Mycophenolate mofetil versus cyclophosphamide for induction treatment of lupus nephritis. *J Am Soc Nephrol*. 2009;20:1103-1112.

A8. Dooley MA, Jayne D, Ginzler EM, et al. Mycophenolate versus azathioprine as maintenance therapy for lupus nephritis. *N Engl J Med*. 2011;365:1886-1895.

A9. Navarra SV, Guzmán RM, Gallacher AE, et al. Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. *Lancet*. 2011;377:721-731.

A10. Furie R, Petri M, Zamani O, et al. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. *Arthritis Rheum*. 2011;63:3918-3930.

A11. Rovin BH, Furie R, Latinis K, et al. Efficacy and safety of rituximab in patients with active proliferative lupus nephritis. *Arthritis Rheum*. 2012;64:1215-1226.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Rullo OJ, Tsao BP. Recent insights into the genetic basis of systemic lupus erythematosus. *Ann Rheum Dis*. 2013;72(suppl 2):ii56-ii61.
2. Crow MK. Type I interferon in the pathogenesis of lupus. *J Immunol*. 2014;192:5459-5468.
3. Crow MK. Advances in understanding the role of type I interferons in systemic lupus erythematosus. *Curr Opin Rheumatol*. 2014;26:467-474.
4. Lee-Kirsch MA, Wolf C, Gunther C. Aicardi-Goutieres syndrome: a model disease for systemic autoimmunity. *Clin Exp Immunol*. 2014;175:17-24.
5. Somers EC, Richardson BC. Environmental exposures, epigenetic changes and the risk of lupus. *Lupus*. 2014;23:568-576.
6. Koh YT, Scatizzi JC, Gahan JD, et al. Role of nucleic acid-sensing TLRs in diverse autoantibody specificities and anti-nuclear antibody-producing B cells. *J Immunol*. 2013;190:4982-4990.
7. Gonzalez-Navajas JM, Lee MJ, David M, et al. Immunomodulatory functions of type I interferons. *Nat Rev Immunol*. 2012;12:125-135.
8. Somers EC, Zhao W, Lewis EE, et al. Type I interferons are associated with subclinical markers of cardiovascular disease in a cohort of systemic lupus erythematosus patients. *PLoS ONE*. 2012;7:e37000.
9. Skaggs BJ, Hahn BH, McMahon M. Accelerated atherosclerosis in patients with SLE–mechanisms and management. *Nat Rev Rheumatol*. 2012;8:214-223.
10. Peart E, Clowse ME. Systemic lupus erythematosus and pregnancy outcomes: an update and review of the literature. *Curr Opin Rheumatol*. 2014;26:118-123.
11. Izmirly PM, Buyon JP, Saxena A. Neonatal lupus: advances in understanding pathogenesis and identifying treatments of cardiac disease. *Curr Opin Rheumatol*. 2012;24:466-472.
12. Sciascia S, Khamashta MA, D'Cruz DP. Targeted therapy in antiphospholipid syndrome. *Curr Opin Rheumatol*. 2014;26:269-275.
13. Nayer A, Ortega LM. Catastrophic antiphospholipid syndrome: a clinical review. *J Nephropathol*. 2014;3:9-17.
14. Sacre KL, Criswell A, McCune JM. Hydroxychloroquine is associated with impaired interferon-alpha and tumor necrosis factor-alpha production by plasmacytoid dendritic cells in systemic lupus erythematosus. *Arthritis Res Ther*. 2012;14:R155.
15. Hahn BH, McMahon MA, Wilkinson A, et al., American College of Rheumatology. American College of Rheumatology guidelines for screening, treatment, and management of lupus nephritis. *Arthritis Care Res (Hoboken)*. 2012;64:797-808.
16. Lisnevskaia L, Murphy G, Isenberg D. Systemic lupus erythematosus. *Lancet*. 2014;384:1878-1888.
17. Stohl W. Future prospects in biologic therapy for systemic lupus erythematosus. *Nat Rev Rheumatol*. 2013;9:705-720.
18. Kirou KA, Gkrouzman E. Anti-interferon alpha treatment in SLE. *Clin Immunol*. 2013;148:303-312.

REVIEW QUESTIONS

1. Data from genome-wide association studies identify gene variants encoding the following proteins as associated with a diagnosis of systemic lupus erythematosus (SLE):

A. Major histocompatibility complex class I
B. Interferon regulatory factor 5 (IRF5)
C. Interferon-γ
D. Interleukin-17
E. Adiponectin

Answer: B Polymorphisms in IRF5 are associated with SLE. IRF5 is a transcription factor that is activated following cell triggering through endosomal toll-like receptors and induces transcription of type I interferon and other proinflammatory mediators. Although MHC class II alleles are associated with a diagnosis of SLE, associations with class I alleles have not been reported. Interferon-γ and interleukin-17 are T-cell products that may be implicated in lupus pathogenesis, but significant genetic polymorphisms in their genes have not been associated with SLE. Adiponectin is a product of adipocytes and has not been described as having a role in lupus susceptibility.

2. Environmental factors can trigger initiation or exacerbation of SLE. One mechanism by which drug-induced lupus can be triggered is through:

A. Induction of T regulatory cells
B. Damage to renal podocytes
C. DNA demethylation
D. Increased proteinuria
E. Hemolysis

Answer: C Procainamide is an antihypertensive agent that has been associated with induction of lupus. This drug impacts gene expression and cell function through demethylation of DNA, resulting in self-reactive T cells. T regulatory cells contribute to protection from autoimmune disease rather than induction of autoimmune disease. Damage to podocytes might be one mechanism of renal damage but has not been associated with drug-induced lupus. Increased proteinuria and hemolysis are clinical manifestations of lupus but do not represent mechanisms that account for drug induced lupus.

3. Clinical manifestations of lupus can include:

A. Seizures
B. Renal failure
C. Myocardial infarction
D. Pericarditis
E. All of the above

Answer: E The clinical manifestations of lupus are protean and can effect any organ system. It should be noted that premature atherosclerotic cardiovascular disease is frequent among lupus patients and can contribute to increased mortality.

4. Autoantibodies specific for the following antigen are specific for SLE:

A. Ro
B. Centromere
C. RNP
D. Scl70
E. Sm

Answer: E Autoantibodies associated with a diagnosis of SLE include those targeting protein components of ribonucleoprotein particles, such as Ro, which is also associated with Sjögren's syndrome and rheumatoid arthritis. Anti-RNP antibodies can be seen in patients with SLE or mixed connective tissue disease. Anti-centromere and anti-Scl70 antibodies are associated with systemic sclerosis. Anti-Smith (Sm) antibodies are nearly exclusive to patients with SLE.

5. Which therapeutic agent is appropriate for both induction and maintenance therapy for class III or IV lupus nephritis?

A. Intravenous pulse glucocorticoid therapy
B. Azathioprine
C. Cyclophosphamide
D. Mycophenolate mofetil
E. Rituximab

Answer: D Induction therapy for lupus nephritis is often conducted with cyclophosphamide, along with IV pulse glucocorticoid therapy. However, recent data indicate that mycophenolate mofetil (MMF) can be as effective as cyclophosphamide. In addition, MMF has shown efficacy as a maintenance therapy. IV pulse glucocorticoid therapy and cyclophosphamide are appropriate for induction therapy but should not be used as a chronic maintenance therapy for lupus nephritis. Azathioprine is appropriate as a maintenance therapy but is inferior to other options as induction therapy. Rituximab has not shown efficacy in randomized controlled clinical trials for lupus nephritis, but nonetheless is sometimes used in that clinical setting if other therapies are not successful.

267 SYSTEMIC SCLEROSIS (SCLERODERMA)

JOHN VARGA

DEFINITION

Systemic sclerosis (SSc) is a chronic autoimmune connective tissue disease of unknown cause. Although middle-aged women are most commonly affected, SSc can occur at any age and is associated with considerable morbidity and increased mortality. The disease shows marked clinical heterogeneity, has protean clinical manifestations, and commonly follows a progressive course.[1] The hallmark of SSc is thickening and hardening of the skin (scleroderma), but in most patients the lungs, gastrointestinal (GI) tract, kidneys, and heart are also affected. In its earliest stages, SSc is associated with prominent inflammation and autoimmunity and altered microvascular function. Over time, progressive structural alterations in small blood vessels and fibrosis in multiple organs cause organ failure. Although there is no approved disease-modifying therapy for SSc, current treatment strategies can control symptoms, slow disease progression, improve quality of life, and prolong

survival. The presence of scleroderma (hard skin) distinguishes SSc from other autoimmune and rheumatic diseases, but skin induration also features prominently in localized forms of scleroderma and multiple unrelated conditions (Table 267-1).

Classification

Systemic Sclerosis

SSc is clinically classified into two subsets: diffuse cutaneous SSc (dcSSc) and limited cutaneous SSc (lcSSc), defined by the pattern of skin involvement and associated with distinct clinical and laboratory manifestations and natural history (Table 267-2). In lcSSc, skin involvement is restricted to the fingers, toes, distal extremities, and face; proximal extremities and the trunk are spared. Diffuse cutaneous SSc is characterized by involvement of the skin proximal to the elbows and knees, including the trunk, along with the distal extremities. In patients with lcSSc, Raynaud phenomenon commonly precedes other disease manifestations. In contrast to lcSSc, dcSSc is generally rapidly progressive and may be complicated by early pulmonary fibrosis, accelerated hypertension, and acute renal failure. The constellation of calcinosis cutis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly (scleroderma of the fingers), and telangiectasia in a subset of patients with lcSSc is termed *CREST syndrome*. Patients with CREST generally follow an indolent course and have a relatively good prognosis. Raynaud phenomenon and other characteristic clinical and laboratory findings of SSc in the absence of obvious skin thickening is the hallmark of SSc sine scleroderma.

TABLE 267-1 CONDITIONS WITH SCLERODERMA-LIKE SKIN INDURATION

Systemic sclerosis (SSc)
- Limited cutaneous SSc
- Diffuse cutaneous SSc

Localized scleroderma
- Morphea (plaque, guttate, generalized)
- Pansclerotic morphea
- Linear scleroderma, "coup de sabre"

Scleredema and diabetic scleredema
Scleromyxedema (papular mucinosis)
Nephrogenic fibrosing syndrome (nephrogenic systemic fibrosis)
Chronic graft-versus-host disease
Diffuse fasciitis with eosinophilia (Shulman disease, eosinophilic fasciitis)
Eosinophilia-myalgia syndrome
Chemically induced scleroderma-like conditions
- Vinyl chloride–induced disease, other solvents
- Pentazocine-induced skin fibrosis
- Other drug associations

Paraneoplastic syndrome

TABLE 267-2 CLASSIFICATION OF SYSTEMIC SCLEROSIS

CHARACTERISTIC FEATURES	LIMITED CUTANEOUS SYSTEMIC SCLEROSIS	DIFFUSE CUTANEOUS SYSTEMIC SCLEROSIS
Skin induration	Limited to fingers, distal to elbows, face; progression slow	Diffuse: fingers, extremities, face, trunk; progression rapid; tendon friction rubs
Raynaud phenomenon	Precedes skin involvement; often severe; associated with critical ischemia	Onset occurs coincident with or subsequent to skin involvement
Pulmonary fibrosis	Occasional, moderately severe	Frequent, early, can be progressive and severe
Pulmonary arterial hypertension	Frequent, late, may be isolated	Occasional, commonly in association with pulmonary fibrosis
Scleroderma renal crisis	Very rare	Occurs in up to 15%; early
Calcinosis cutis	Frequent, prominent	Infrequent
Characteristic autoantibodies	Anticentromere	Antitopoisomerase I (Scl-70), anti-RNA polymerase III

Mixed Connective Tissue Disorder

Mixed connective tissue disorder (MCTD) is an overlap syndrome first described in 1971 that is characterized by features of systemic lupus erythematosus (SLE), SSc, and myositis, all occurring in the same patient. In the early phase, most patients have Raynaud phenomenon in association with edema of the hands and evidence of inflammatory muscle disease. Over time, these patients sequentially manifest other features of connective tissue diseases, including pericarditis, esophageal dysmotility, sclerodactyly, neuropathy, and pulmonary arterial hypertension (PAH). Erosive arthritis does not occur. On the other hand, some patients develop acute renal involvement similar to scleroderma renal crisis. In the early stage of this disorder, it is often difficult to predict whether the patient will progress to develop a distinct connective tissue disease such as SSc or SLE or will be ultimately diagnosed with MCTD. A diagnostic hallmark of MCTD is the presence in the serum of a highly characteristic autoantibody specificity against U1-ribonuclear protein (U1-RNP). Most MCTD patients have a very high titer (often > 1 : 1000) of anti-U1-RNP autoantibody. Some of the features of MCTD might respond to corticosteroid therapy. Overall, patients with MCTD generally have a better prognosis than those with SSc.

Localized Scleroderma

Localized scleroderma refers to a family of generally benign skin conditions primarily affecting children. These conditions are characterized by discreet areas of skin induration in the absence of Raynaud phenomenon or systemic involvement. Lesional skin is discolored and indurated and histologically may be indistinguishable from SSc. Localized scleroderma has multiple distinct forms. When it occurs as single or multiple solitary patches of induration, it is called morphea. When these patches coalesce, the condition is called generalized morphea. The lesions are generally asymmetrical in distribution and spare the digits. Some individuals have extensive and disabling induration (pansclerotic morphea). Induration may follow in a linear distribution, most commonly on the lower extremities (linear scleroderma). In children, linear scleroderma can be complicated by growth retardation and joint contractures.

EPIDEMIOLOGY

SSc is a sporadic disease with a worldwide distribution. The incidence is 9 to 19 cases per million per year, with an estimated 100,000 cases in the United States. Studies from several countries suggest that the incidence of SSc may be increasing. Like other connective tissue diseases, SSc shows a marked female predominance, particularly in the childbearing years. The peak age of onset is 30 to 50 years for both the limited and diffuse cutaneous forms. African Americans have a higher incidence and an earlier age of disease onset compared to whites, and are more likely to have the diffuse cutaneous form of SSc associated with interstitial lung involvement and a worse prognosis.

Etiology and Environmental and Occupational Exposures

Although the cause of SSc is unknown, the onset is commonly ascribed to an interplay between environmental factors and genetic susceptibility. Suspected environmental triggers include occupational, dietary, medical, and lifestyle exposures, and possibly certain infectious agents. Because some SSc autoantibodies cross-react with certain virus-associated epitopes, molecular mimicry has been proposed as a possible pathogenetic link between viral infection and SSc.

Epidemic outbreaks of SSc-like syndromes have been linked with toxic exposures. A Spanish outbreak of toxic oil syndrome in the 1980s affected more than 20,000 individuals. The syndrome was characterized by chronic skin induration and neuropathy, and the outbreak was linked to ingestion of contaminated rapeseed oils used for cooking. An outbreak of eosinophilia-myalgia syndrome (EMS) a decade later was linked to ingestion of over-the-counter L-tryptophan dietary supplements used for insomnia, weight loss, and other indications. The syndrome was characterized by peripheral blood eosinophilia and severe myalgia in the acute stage, followed by intractable scleroderma-like diffuse skin induration. Neither toxic oil syndrome nor EMS was associated with Raynaud phenomenon or SSc-specific autoantibodies.

Occupational exposures tentatively linked with SSc include silica (in miners), polyvinyl chloride, epoxy resins, and aromatic hydrocarbons including toluene and trichloroethylene. Certain drugs, including bleomycin, pentazocine, hormone replacement therapy, cocaine, and appetite suppressants, have been linked with SSc or PAH. Although earlier studies implied a

possible association of SSc with silicone breast implants, large-scale epidemiologic investigations failed to establish an increased risk.

Genetic Factors

A genetic contribution to SSc susceptibility is indicated by the fact that 1.6% of patients have a first-degree relative with SSc, a prevalence rate substantially higher than in the general population (0.026). Indeed, a family history is the strongest identified risk factor for SSc. Moreover, patients with SSc are more likely to have first-degree relatives with Raynaud phenomenon and interstitial lung disease, as well as other autoimmune diseases including multiple sclerosis, rheumatoid arthritis, and thyroiditis. Genomewide association and candidate-gene studies in SSc have identified association with multiple HLA loci, as well as non-HLA loci including *STAT4*, *IRF4*, *PTPN22*, *TNIP-1*, *IRAK1*, *CD247*, and *BANK1*, each of which encodes genes involved in immune regulation or autoimmunity. Moreover, most of these SSc-associated risk alleles are also linked with other autoimmune diseases, especially SLE.

PATHOBIOLOGY

The protean clinical and pathologic manifestations of SSc reflect a complex underlying pathobiology encompassing three interrelated cardinal processes: autoimmunity and inflammation, vascular injury and obliteration, and fibrosis and excessive matrix accumulation in multiple tissues and organs (Fig. 267-1).[2] This canonical triad is operative to a greater or lesser extent in all patients with SSc, and their variable relative contribution of each process to the individual phenotype account for the observed disease heterogeneity.

Pathology

The distinguishing pathologic hallmark of SSc is the constellation of capillary loss (rarefaction) and obliterative vasculopathy coexisting with fibrosis in the skin and internal organs. In early-stage disease, perivascular inflammation can be detected in multiple organs before the appearance of fibrosis. The vascular lesion is characterized by intimal proliferation in the small and medium-sized arteries, resulting in luminal narrowing and obliteration.[3] In later-stage SSc, fibrosis is prominent in the skin, lungs, GI tract, heart, tendon sheath, perifascicular tissue surrounding skeletal muscle, and some endocrine organs.[4] Accumulation of connective tissue rich in collagens, fibronectin, cartilage oligomeric matrix protein (COMP), and proteoglycans disrupts normal architecture, resulting in functional impairment of affected organs.

In the skin, dermal collagen deposition causes obliteration of the hair follicles, sweat glands, and other adnexae, as well as invasion of the subjacent adipose layer with entrapment of fat cells. The epidermis is atrophic, and the rete pegs are effaced. In late-stage disease, there is a paucity of vascular and lymphatic endothelium. In the lungs, the interstitium and alveolar spaces are infiltrated with inflammatory cells in early disease. With progression, interstitial fibrosis and vascular damage, often coexisting within the same lesions, dominate the pathologic picture. The most common histologic pattern in SSc-associated lung disease is nonspecific interstitial pneumonitis (NSIP). Progressive thickening of the alveolar septae results in obliteration of the air spaces, honeycombing, and loss of pulmonary blood vessels.

Intimal thickening of the pulmonary arteries (Fig. 267-2), best seen with elastin stain, underlies PAH (Chapter 68). These vascular lesions resemble those of, but are distinct from, idiopathic PAH, but the hallmark plexiform lesions are uncommon in SSc. In the GI tract, pathologic changes can be found at any level from the mouth to the rectum. Fibrosis of the lamina propria and submucosa with atrophy of the muscular layers are prominent in the lower esophagus, whereas striated muscle in the upper third of the esophagus is generally spared (Chapter 138). Replacement of the normal gut architecture leads to disordered peristaltic activity with gastroesophageal reflux and dysmotility, gastroparesis, and small bowel obstruction. Chronic reflux

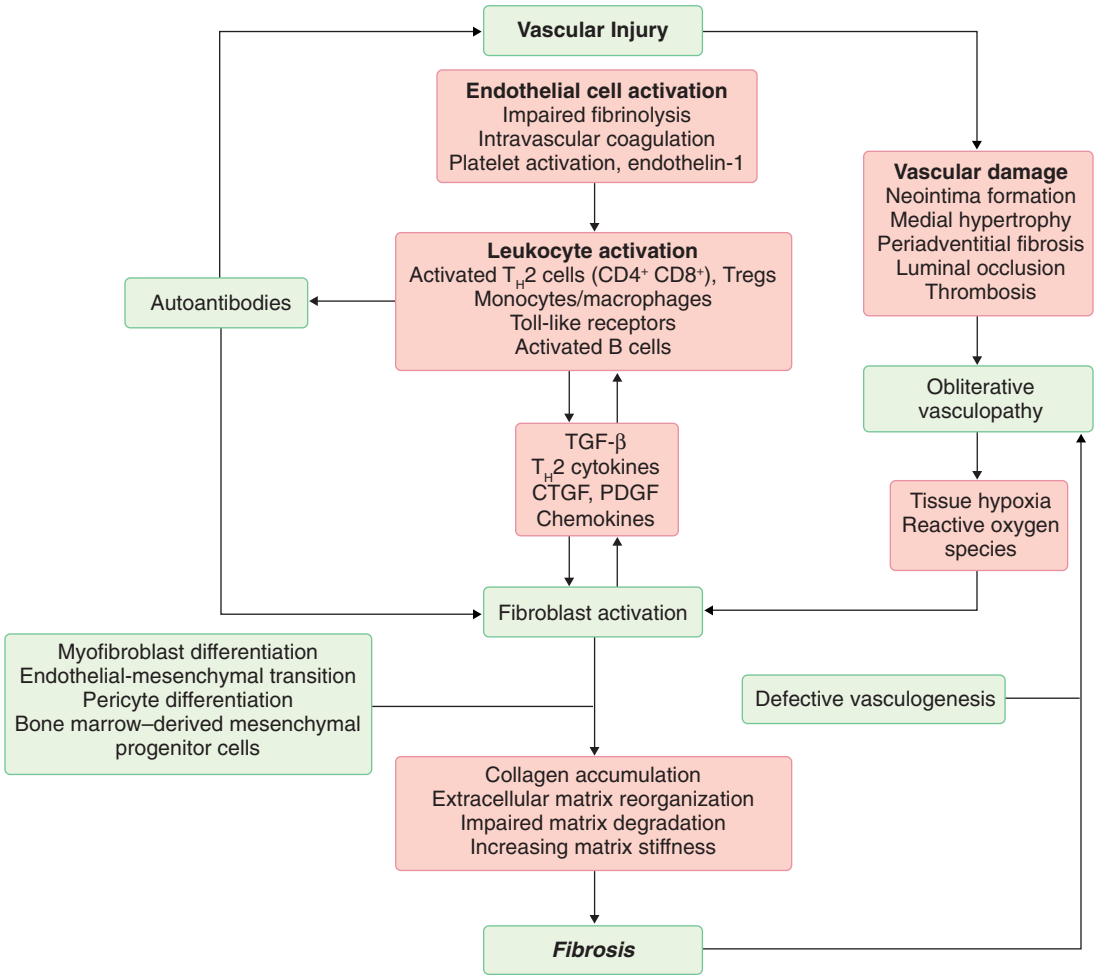

FIGURE 267-1. The pathogenetic triad of systemic sclerosis: vasculopathy, autoimmunity, fibrosis. Initial endothelial injury in a genetically susceptible individual leads to vascular damage, inflammation, and autoimmunity. The inflammatory and immune responses initiate fibroblast activation, resulting in intractable fibrosis. Vasculopathy, loss of microvasculature, and reduced blood flow result in ischemia and generation of reactive oxygen species that contribute to and further aggravate vascular damage, tissue fibrosis, and atrophy. CTGF = connective tissue growth factor; PDGF = platelet-derived growth factor; TGF-β = transforming growth factor-β.

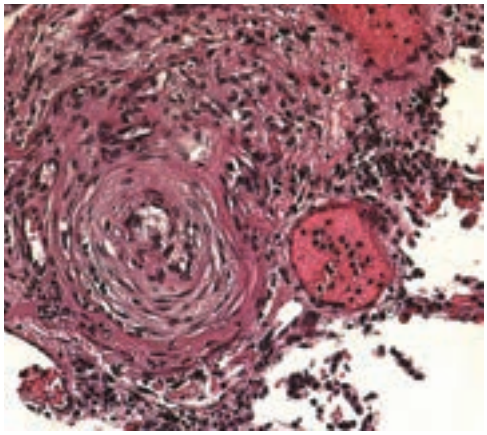

FIGURE 267-2. Pulmonary artery obliterative vasculopathy. Striking intimal hyperplasia and narrowing of the lumen of a small pulmonary artery, coexisting with interstitial pulmonary fibrosis, in a patient with diffuse cutaneous systemic sclerosis.

is associated with esophageal inflammation, ulcerations, stricture formation, and Barrett metaplasia.

Pathologic changes in the heart are common in SSc, with involvement of the myocardium and pericardium. Characteristic microvascular lesions include concentric intimal hypertrophy and luminal narrowing. Contraction band necrosis reflecting ischemia-reperfusion injury in the myocardium is prominent and may be accompanied by patchy myocardial fibrosis. In the kidneys, noninflammatory lesions occur in the interlobular arteries. Scleroderma renal crisis (Chapter 125) is associated with striking changes in small renal arteries, with reduplication of elastic lamina, marked intimal proliferation, and concentric narrowing of the lumen giving rise to the onion-skin appearance, frequently accompanied by thrombosis and microangiopathic hemolysis.[5]

Pathophysiologic Triad: Vasculopathy, Immune Dysregulation, and Fibrosis

Vasculopathy

Vascular injury is an early and presumably primary event in pathogenesis of SSc. It effects primarily the small and medium-sized arteries and arterioles in multiple vascular beds and accounts for major clinical complications. The initial vascular endothelial injury might be caused by viruses and other infectious agents, oxygen radicals, circulating cytotoxic factors, complement activation, or autoantibodies. Endothelial cell injury and apoptosis result in altered production of endothelium-derived vasodilatory (nitric oxide and prostacyclin) and vasoconstricting (endothelin-1) molecules. Endothelial dysfunction causes increased vascular permeability associated with upregulation of adhesion molecules with transendothelial leukocyte diapedesis, as well as platelet aggregation, activation of intravascular coagulation, defective fibrinolysis, and thrombosis. Small blood vessels show intimal hyperplasia with thickening and reduplication of the basement membrane. In the vascular media, myointimal cells proliferate, whereas the adventitial layers develop fibrosis; the net result is progressive narrowing and obliteration of capillaries, arterioles, and even large vessels. Impaired blood flow results in widespread tissue ischemia, which is a potent stimulus for fibrogenesis. Recurrent ischemia-reperfusion is associated with the generation of reactive oxygen species that further damage the endothelium. Paradoxically, despite the presence of tissue hypoxia and sometimes dramatically elevated levels of angiogenic factors, compensatory vasculogenesis is impaired owing to reduced production, mobilization, or maturation of endothelial progenitor cells. The combination of widespread capillary loss, obliterative vasculopathy of small and medium-sized arteries, and failure to regenerate damaged blood vessels are hallmarks of SSc that underlie Raynaud phenomenon, ischemic digital ulcers, cutaneous telangiectasia, PAH, and other major vascular manifestations.

Raynaud Phenomenon

The earliest and most common vascular complication of SSc, Raynaud phenomenon, reflects abnormal thermal regulation of blood flow and can precede other disease manifestations by years.[6] Raynaud phenomenon in SSc is characterized by autonomic and peripheral nervous system changes leading to impaired production of calcitonin gene–related peptide from sensory afferent nerves and heightened sensitivity of α_2-adrenergic receptors on vascular smooth muscle cells. In contrast to primary Raynaud phenomenon, a common and relatively benign condition, Raynaud phenomenon in SSc is generally progressive and complicated by vascular remodeling with irreversible structural changes that result in tissue damage.

TABLE 267-3 CHARACTERISTIC AUTOANTIBODIES IN SYSTEMIC SCLEROSIS

AUTOANTIBODY (FREQUENCY IN SSc)	SYSTEMIC SCLEROSIS SUBSET	CLINICAL ASSOCIATION
Topoisomerase-I (10-40%)	Diffuse cutaneous (less commonly limited)	Tendon friction rubs, ILD, cardiac involvement, scleroderma renal crisis; isolated PAH rare
Centromere (15-40%)	Limited cutaneous	Digital ischemia, calcinosis cutis, isolated PAH, PBC; severe ILD and scleroderma renal crisis rare
RNA polymerase III (4-25%)	Diffuse cutaneous	Extensive skin involvement; tendon friction rubs, scleroderma renal crisis, increased cancer risk
U3-RNP/fibrillarin (1-5%)	Diffuse cutaneous	PAH, ILD, myositis
Th/To (1-7%)	Limited cutaneous	ILD, isolated PAH
PM/Scl (0-6%)	Limited cutaneous	Calcinosis, myositis, arthritis
U1-RNP (5-35%)	MCTD	Severe PAH, myositis

ILD = interstitial lung disease; MCTD = mixed connective tissue disease; PAH = pulmonary arterial hypertension, PBC = primary biliary cirrhosis.

Inflammation and Autoimmunity

Cellular Immunity

Immune dysregulation is a hallmark SSc shares with other autoimmune diseases. In early SSc, activated T cells, B cells, and monocyte-macrophages accumulate in lesional tissues. Recent studies employing genomewide transcriptional profiling indicate that a subset of SSc patients demonstrate a strong inflammatory gene signature in their lesional skin, with evidence of activated innate and adaptive immune signaling and elevated expression of many inflammatory chemokines and cytokines.

Infiltrating $CD4^+$ T cells display restricted TcR receptor signatures indicative of their oligoclonal expansion in response to unknown antigens. Moreover, these T cells show T_H2 polarization with secretion of interleukin (IL)-4, IL-13, and IL-21, and low levels of interferon (IFN)-γ. T_H2 cytokines induce TGF-β and promote the synthesis of collagen and other extracellular matrix molecules. Increased levels of IL-17 detected in the serum suggest a role for T_H17 cells in SSc. The frequency of circulating regulatory T cells is elevated, but their immunosuppressive function appears to be defective. Myeloid dendritic cells show abnormally high secretion of inflammatory cytokines and chemokines such as CXCL.[7] Moreover, in response to thymic stromal lymphopoietin (TSLP), which is elevated in SSc, dendritic cells contribute to the persistence of the T_H2-polarized profibrotic immune response. Macrophages show evidence of alternative activation associated with pathologic fibrogenesis. In SSc, aberrant innate immune responses in dendritic cells and tissue fibroblasts might be triggered by damage-associated nucleic acids and matrix macromolecules. These responses are mediated through toll-like receptors (TLRs) and are likely to contribute to autoimmunity and progressive fibrosis.[8] The elevated expression of type I interferon-regulated genes (IFN signature) in SSc is consistent with innate immune activation and may contribute to vascular injury.

Autoantibodies and B Cells

Although circulating autoantibodies have well-established clinical utility in SSc as diagnostic and prognostic markers, their role in driving pathogenesis and tissue damage remains speculative. In addition to antinuclear antibodies (ANAs) that are detected in virtually all patients with SSc, a number of highly disease-specific and mutually exclusive autoantibodies occur (Table 267-3).

Animal Models of Disease

Although no single animal model fully recapitulates the tripartite immune-vascular-fibrotic pathobiology of SSc, certain genetically engineered mice and mice subjected to chemical or immunologic challenges exhibit particular disease features and are useful in investigating pathogenesis and therapy. Tight skin mice (Tsk1) spontaneously develop diffuse skin induration and fibrosis, with prominent thickening of the hypodermis. The Tsk1 phenotype is due to a duplication mutation in the gene for fibrillin-1, a component of extracellular microfibrils. Microfibrils provide a scaffold for elastin fibers but also control the storage and activity of transforming growth factor (TGF)-β and other regulatory molecules. Mutations in the fibrillin-1 gene also underlie Marfan syndrome. Scleroderma-like skin changes and organ fibrosis can be induced in mice by subcutaneous injection of bleomycin, TGF-β angiotensin II, or HOCl, or by transplantation of HLA-mismatched bone marrow or spleen cells. Increasingly, targeted genetic modifications such as deletion or transgenesis are used to create novel animal models for dissecting the molecular and cellular pathways in fibrosis and for the discovery and preclinical evaluation of novel therapies. For instance, transgenic mice with constitutive TGF-β, connective tissue growth factor [CTGF], platelet-derived growth factor [PDGF], Fra-2 or Wnt10 signaling spontaneously develop scleroderma-like changes in multiple organs and serve as valuable experimental tools.

Most SSc-specific autoantibodies are directed against intracellular proteins, such as topoisomerase-I, centromere, and RNA polymerases I and III. Recent studies identified autoantibodies in SSc that are directed against endothelial cell or myenteric neuron epitopes, or recognize cell surface receptors (platelet-derived growth factor receptor [PDGFR], angiotensin II receptor, and muscarinic-3-acetylcholine receptor), and appear to directly contribute to vascular injury, intestinal wall atrophy and dysmotility, or tissue fibrosis. B cells are implicated in mediating both the autoimmune and fibrotic components of SSc. In addition to their role in antibody production, B cells also present antigen, produce IL-6 and other profibrotic cytokines, and modulate the function of T cells and dendritic cells. Gene expression profiling of SSc skin biopsies has identified messenger RNA expression signatures characteristic of B-cell activation.

Fibrosis

Fibrosis of the skin and multiple internal organs is the distinguishing feature of SSc. It characteristically follows, and is thought to be a consequence of, inflammation and vascular injury. Fibrosis is characterized by replacement of normal tissue architecture with a collagen-rich extracellular matrix that is secreted by resident fibroblasts and myofibroblasts.[9] Under physiologic conditions, these mesenchymal cells undergo controlled activation triggered by TGF-β, IL-6, PDGF, hypoxia, oxygen radicals, and other factors. Fibroblasts proliferate, migrate, synthesize and secrete collagens and extracellular matrix, and transdifferentiate into contractile myofibroblasts, enabling them to repair damaged tissue with full regeneration. When this tightly regulated wound healing program becomes sustained and amplified, excessive scar tissue formation ensues, leading to intractable fibrosis.

Effector Cells in Fibrosis

Myofibroblasts are mesenchymal cells with both smooth muscle cell–like contractile and biosynthetic properties. Myofibroblasts appear transiently in wounds, where they contribute to healing through production of collagen and TGF-β and contraction of the surrounding extracellular matrix. In SSc, activated myofibroblasts accumulate in lesional tissue and persist there because of one of three pathways: (1) in situ activation of quiescent resident fibroblasts, (2) through transdifferentiation from damaged epithelial cells, endothelial cells, or pericytes, or (3) by migration and terminal differentiation of bone marrow–derived mesenchymal progenitor cells.

CLINICAL MANIFESTATIONS

Overview

Multiple organs are affected in SSc, but the frequency, tempo, and severity of their involvement show substantial patient-to-patient variability. Patients with dcSSc characteristically develop extensive skin induration associated with early and progressive internal organ involvement. In contrast, patients with lcSSc commonly present with long-standing Raynaud phenomenon, skin changes limited to the distal extremities and face, and indolent progression of internal organ disease. However, patients with SSc frequently defy easy subclassification, or show an overlap of typical SSc features coexisting with clinical and laboratory evidence of another autoimmune disease such as polymyositis, Sjögren syndrome, polyarthritis, or SLE.

Initial Clinical Presentation

Diffuse Cutaneous Systemic Sclerosis

Patients with dcSSc typically present with soft tissue swelling, erythema, and pruritus, often accompanied by fatigue, stiffness, and malaise. Although arthralgia, muscle weakness, and carpal tunnel syndrome are common, Raynaud phenomenon may not be present until later in the disease. In the ensuing weeks to months, the inflammatory edematous phase evolves into a chronic "fibrotic" phase with skin induration accompanied by hyperpigmentation, loss of body hair, and impaired sweating. The wrists, elbows, shoulders, knees, and ankles become stiff owing to fibrosis of the joint structures. Advancing skin changes are commonly accompanied by onset of internal organ involvement that is most rapidly progressive during the initial 4 years from disease onset. The risk for new organ involvement declines thereafter.

Limited Cutaneous Systemic Sclerosis

In lcSSc, the diagnosis is generally made at a more advanced stage of the disease. These patients give a history of long-standing Raynaud phenomenon, sometimes complicated by ischemic ulcerations at the fingertips. The course of disease is indolent, with delayed onset and slow progression of gastroesophageal reflux, telangiectasia, or cutaneous calcinosis. Vascular manifestations of lcSSc tend to be more pronounced compared with dcSSc, and digital ischemia, cutaneous telangiectasia, and progressive PAH are frequent late manifestations. In contrast, scleroderma renal crisis is uncommon in lcSSc.

Organ Involvement

Skin

Skin thickening, the distinguishing hallmark of SSc, typically starts in the fingers and advances in a centripetal pattern from distal to proximal extremities. The skin becomes hyperpigmented, but dark-skinned individuals may develop vitiligo-like hypopigmentation or "salt-and-pepper" changes, most prominently on the upper back and chest. Obliteration of eccrine sweat glands and sebaceous glands results in decreased sweating and oil secretion, causing dry and itchy skin. The face may assume a characteristic appearance with a beaklike nose, thinning and retraction of the lips, fine wrinkles (radial furrowing) around the mouth, and occasionally a masklike facies due to reduced mobility of the eyelids, cheeks, and mouth (Fig. 267-3). Decreased oral aperture (microstomia) interferes with eating and oral hygiene.

In patients with long-standing SSc, the skin is atrophic and tethered to the subcutaneous tissue. Telangiectasias are prominent on the face, hands, lips, and oral mucosa. They resemble the skin lesions of hereditary hemorrhagic telangiectasia and are due to dilatation of postcapillary venules in the upper dermis. Breakdown of atrophic skin leads to painful ulcerations at the extensor surfaces of the interphalangeal joints, fingertips, and bony prominences such as the elbows and malleoli. Ulcers may become secondarily infected, resulting in osteomyelitis. Ischemic fingertip ulcerations heal slowly and give rise to characteristic digital tip "pits." Ischemic soft tissue loss at the fingertips is associated with resorption of the terminal phalanges (acro-osteolysis) (Fig. 267-4).

Calcium deposits composed of calcium hydroxyapatite crystals develop in the skin and soft tissues. These deposits, varying in size from tiny punctate lesions to large conglomerate masses, can be readily visualized on plain radiographs. Frequent locations include the finger pads, extensor surfaces of the forearms, and olecranon and prepatellar bursae. Calcific deposits can ulcerate through the overlying skin, producing drainage of chalky white material, pain, and local inflammation.

Raynaud Phenomenon

Raynaud phenomenon (Chapter 80) is an episodic vasoconstriction in the digits that occurs in virtually all patients with SSc.[8] Typical attacks start with

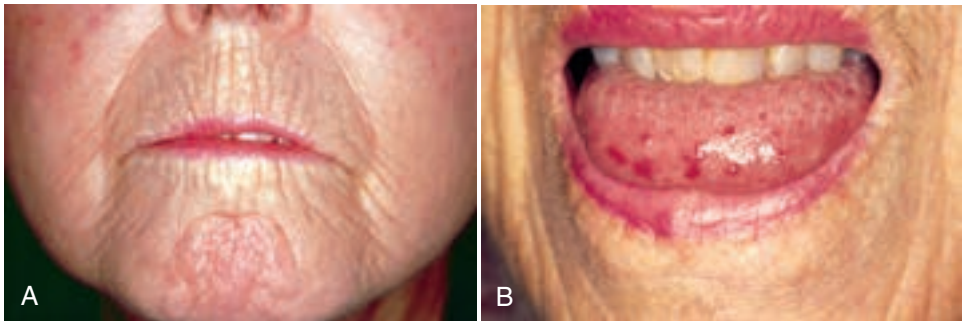

FIGURE 267-3. Facial features in systemic sclerosis. A, Perioral furrowing. Note vertical lines of furrowing around the mouth in a patient with diffuse cutaneous systemic sclerosis. **B,** Telangiectasia on the lips and tongue in a patient with long-standing limited cutaneous systemic sclerosis.

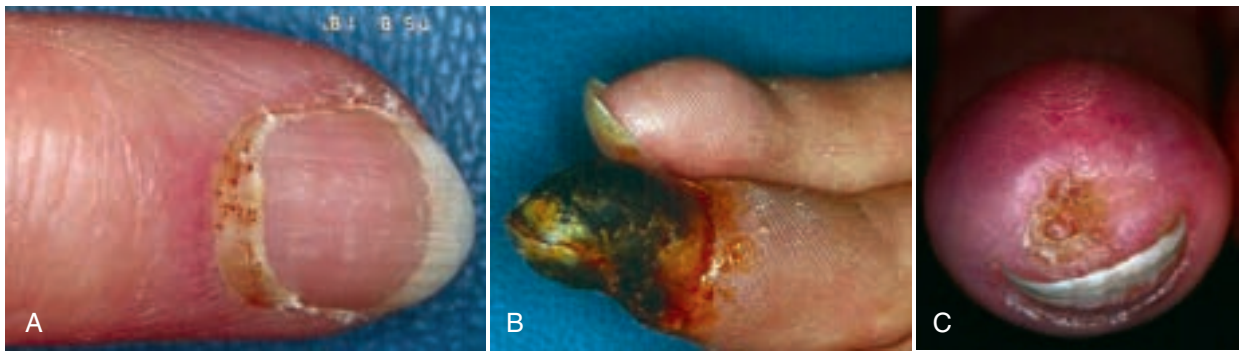

FIGURE 267-4. Vascular complications of systemic sclerosis in the fingers. **A,** Nailfold microvascular changes. **B,** Digital infarction. Sharply demarcated necrosis of the fingertip in a patient with limited cutaneous systemic sclerosis associated with severe Raynaud phenomenon. **C,** Digital tip ulceration and pitting.

pallor (vasoconstriction) followed by cyanosis (ischemia) and erythema (reperfusion), commonly triggered by exposure to cold or emotional stress. Primary Raynaud phenomenon, a benign condition representing an exaggerated physiologic response to cold, occurs in 3 to 5% of the population and is more frequent in women. Secondary Raynaud phenomenon occurs in SSc but also other connective tissue diseases, hematologic and endocrine conditions, and occupational disorders, and with the use of β-blockers and anticancer drugs such as cisplatin and bleomycin.

Distinguishing primary from secondary Raynaud phenomenon can present a challenge. Secondary Raynaud phenomenon typically develops at an older age (>30 years), tends to be more severe, and is frequently complicated by critical ischemia. Nailfold capillaroscopy allows cutaneous capillaries to be viewed under a drop of immersion oil using an ophthalmoscope. Patients with primary Raynaud phenomenon have normal nailfold capillaries that appear as regularly spaced parallel vascular loops, whereas in SSc, capillaries are distorted with widened and irregular loops, dilated lumen, and areas of vascular "dropout."

Gastrointestinal Involvement

GI tract involvement is very common in both lcSSc and dcSSc and may be the presenting manifestation of the disease. A pathologic picture of smooth muscle atrophy and obliterative small vessel vasculopathy with or without fibrosis is seen throughout the length of the GI tract, causing altered peristaltic activity and consequent complications. Severe intestinal involvement and malnutrition are associated with high mortality.

Upper Gastrointestinal Tract

Oropharyngeal manifestations of SSc include xerostomia, reduced oral aperture, periodontal disease, and resorption of the mandibular condyles. The frenulum of the tongue may be shortened. The most frequently affected GI organ is the esophagus. Gastroesophageal reflux is associated with heartburn, regurgitation, and dysphagia but can also be asymptomatic (Chapter 138). Reduced lower esophageal sphincter pressure resulting in gastroesophageal reflux frequently coexists with impaired esophageal clearance of refluxed gastric contents due to diminished motility in the distal two thirds of the esophagus. Delayed gastric emptying further aggravates the problem. On high-resolution computed tomography (HRCT) of the chest, the esophagus is dilated and shows intraluminal air. Endoscopy may show severe erosive esophagitis in patients with minimal reflux symptoms. Esophageal strictures and Barrett esophagus (Chapter 138) can complicate long-standing reflux. Because Barrett esophagus is associated with an increased risk for adenocarcinoma, SSc patients require periodic endoscopy with mucosal biopsy. Hoarseness and chronic cough may be extraesophageal manifestations of gastroesophageal reflux disease (GERD). Chronic microaspiration of gastric contents may aggravate underlying interstitial lung disease. A distinct pattern of interstitial lung disease called centrilobular fibrosis, associated with esophageal dilatation and chronic gastroesophageal reflux, is occasionally seen in SSc.

Stomach

Gastroparesis contributes to delayed gastric emptying with early satiety, abdominal distention, and aggravated reflux symptoms. Gastric vascular ectasia (GAVE) develops in 5% of patients and is equally prevalent in limited and diffuse cutaneous disease. On endoscopy, parallel longitudinal mucosal folds resembling the stripes of a watermelon are seen in the antrum. The histologic features of dilated thrombosed mucosal capillaries and fibromuscular dysplasia of the lamina propria reflect the diffuse small vessel vasculopathy of SSc. Patients with GAVE may develop recurrent GI bleeding and typically present with unexplained iron deficiency anemia.

Lower Gastrointestinal Tract

Impaired small bowel motility in SSc can cause chronic diarrhea due to bacterial overgrowth. Fat and protein malabsorption, vitamin B_{12} and D deficiency, and malnutrition may ensue and are associated with high mortality. Malabsorption is diagnosed by hydrogen breath test or 14C-D-xylose test, and serum prealbumin (transthyretin) is useful to monitor malnutrition (Chapter 140). Disturbed intestinal motor function can also cause recurrent episodes of intestinal pseudo-obstruction with acute abdominal pain, nausea, and vomiting. Differentiating pseudo-obstruction, which responds to supportive care and intravenous nutritional supplementation, from mechanical bowel obstruction is a difficult diagnostic challenge. Colonic and anorectal involvement causing constipation, rectal prolapse, and fecal incontinence is frequent and is the source of much distress. In late-stage SSc, wide-mouth colonic sacculations can occur and cause perforation and bleeding. An occasional radiologic finding is pneumatosis cystoides intestinalis due to air trapping in the bowel wall. These lesions can rupture and cause pneumoperitoneum. The liver is rarely affected in SSc. However, primary biliary cirrhosis associated with antimitochondrial antibodies may occur.

Lung Involvement

The two major forms of lung involvement in SSc are interstitial lung disease and PAH, with many patients developing both. Less frequent pulmonary manifestations include aspiration pneumonitis complicating gastroesophageal reflux, pulmonary hemorrhage, obliterative bronchiolitis, pleural reactions, restrictive ventilatory disease due to chest wall fibrosis, spontaneous pneumothorax, and drug-induced lung toxicity. The incidence of lung cancer, particularly bronchoalveolar carcinoma (Chapter 191), is increased.

Interstitial Lung Disease

Interstitial lung disease (Chapter 92) in SSc can remain asymptomatic until quite advanced.[10] The most frequent presenting symptoms are exertional dyspnea, fatigue, and reduced exercise tolerance. A chronic dry cough may be present. Physical examination may reveal "Velcro" crackles at the lung bases. Pulmonary function testing (Chapter 85) is a sensitive method for detecting early interstitial lung disease. The most common abnormalities are reductions in forced vital capacity (FVC) or single breath diffusing capacity (D_{LCO}). However, a reduction in D_{LCO} that is significantly out of proportion to the reduction in FVC (FVC/D_{LCO} ratio > 1.6) suggests pulmonary vascular disease.

Evidence of interstitial lung disease can be found in almost all patients with SSc and is clinically significant in up to 50%. Risk factors include male sex, African American race, diffuse skin involvement, severe gastroesophageal reflux, and the presence of topoisomerase-I autoantibodies. The most rapid progression in interstitial lung disease occurs within the first 3 years of the disease.

Chest radiography is useful for ruling out infection and other causes of pulmonary involvement but is relatively insensitive for detection of early interstitial lung disease. In contrast, HRCT is highly sensitive (Chapter 84). Prominent HRCT findings include reticular linear interstitial opacities,

predominantly in the lower lobe periphery, occurring in isolation or in combination with ground-glass opacification. Additional findings include mediastinal lymphadenopathy and, rarely, honeycombing. The extent of lung disease on initial HRCT correlates with progression and prognosis of interstitial lung disease and may provide useful information regarding the need for initiating therapy. Bronchoalveolar lavage (Chapter 85) may be indicated for ruling out occult infection. Lung biopsy is rarely useful.

Pulmonary Arterial Hypertension

Approximately 15% of SSc patients develop PAH, defined as a mean pulmonary arterial pressure of 25 mm Hg or greater, with a pulmonary capillary wedge pressure of 15 mm Hg or less (Chapter 68). In the setting of SSc, PAH can occur as an isolated abnormality (World Health Organization [WHO] group I), or coexist with interstitial lung disease (WHO group III). Although the natural history of SSc-associated PAH is variable, patients generally follow a progressive downhill course, with development of right heart failure and increased mortality. Risk factors for SSc-associated PAH include limited cutaneous disease, older age of disease onset, severe Raynaud phenomenon, large number of cutaneous telangiectasias, and anticentromere, U1-RNP, U3-RNP (fibrillarin), Th/To, B23, and β_2-glycoprotein I autoantibodies.

The initial symptoms of PAH are exertional dyspnea and reduced exercise capacity, but early-stage disease is often clinically silent. With progression, angina, syncope, and symptoms and signs of right-sided heart failure develop. Physical examination shows tachypnea, a prominent pulmonic S_2 heart sound, palpable right ventricular heave, elevated jugular venous pressure, and dependent edema. Pulmonary arterial systolic pressures above 40 mm Hg (determined by Doppler echocardiography) suggest PAH, as does an isolated low D_{LCO} or a FVC/D_{LCO} ratio over 1.6. Right heart catheterization is required for confirming the diagnosis of PAH, assessing its severity, and evaluating ventricular function. The serum levels of N-terminal brain natriuretic peptide (NT-pro-BNP) are elevated in PAH and correlate with severity and survival.

Kidney Involvement

Scleroderma renal crisis is an uncommon but life-threatening acute complication of SSc, but chronic and indolent kidney disease also occurs.

Scleroderma Renal Crisis

Scleroderma renal crisis, the most dreaded complication of SSc, develops in 10 to 15% of patients, almost always within 4 years of disease onset.[11] Prior to the advent of angiotensin-converting enzyme (ACE)-inhibiting drugs in the 1980s, scleroderma renal crisis was invariably fatal, often within weeks. The pathogenesis involves obliterative vasculopathy and luminal narrowing of the renal arcuate arteries. Progressive reduction in renal blood flow, aggravated by vasospasm, leads to juxtaglomerular hyperplasia and increased renin secretion, with further renal vasoconstriction resulting in a vicious cycle that culminates in malignant hypertension (Chapters 67 and 125).

Scleroderma renal crisis is a medical emergency. Although most patients present with abrupt onset of hypertension and progressive renal insufficiency, in some cases the blood pressure remains normal or only modestly elevated. Normotensive renal crisis is associated with a poor outcome. Hypertensive encephalopathy and retinopathy, pericarditis, and arrhythmias may complicate scleroderma renal crisis. Urinalysis shows mild proteinuria, granular casts, and microscopic hematuria. When thrombocytopenia and microangiopathic hemolysis with fragmented red blood cells are detected, the diagnosis of thrombotic thrombocytopenic purpura (Chapter 172) is sometimes entertained. In many patients, oliguric renal failure develops over a period of weeks. Kidney biopsy can be useful for diagnosis and prognosis, but the characteristic lesions of intimal and medial proliferation and luminal narrowing are indistinguishable from the changes of accelerated hypertension.

Risk factors for scleroderma renal crisis include early-stage disease, rapidly progressive skin involvement and the presence of tendon friction rubs, African American race, male sex, and autoantibodies to RNA polymerases I and III. In contrast, the presence of anticentromere antibodies is associated with a low risk for scleroderma renal crisis. Pericardial effusion, new-onset anemia, and thrombocytopenia may be harbingers of impending scleroderma renal crisis, and a history of recent corticosteroid use is associated with a more than 10-fold increased risk. Accordingly, SSc patients with early and progressive cutaneous disease should be counseled to determine their blood pressure daily. In these patients, corticosteroids should be used only when absolutely required, and at low doses.

Once scleroderma renal crisis sets in, hospitalization and prompt initiation of short-acting ACE inhibitors is essential. The goal is adequate blood pressure control before the onset of renal failure. Despite appropriate timely intervention, more than half of patients with scleroderma renal crisis require hemodialysis, although some of these ultimately recover sufficient renal function to be able to discontinue hemodialysis. Oliguria or a serum creatinine level higher than 3 mg/dL at presentation predict poor outcome. The "prophylactic" use of ACE inhibitors to prevent scleroderma renal crisis is associated with a worse outcome and is not recommended.

Chronic Kidney Disease

Kidney biopsies in patients with SSc commonly show chronic changes including reduplication of elastic fibers, sclerosed glomeruli, tubular atrophy, and interstitial fibrosis. In one study, abnormal renal function or proteinuria was detected in more than one third of patients, none of whom progressed to end-stage renal disease. Rarely, glomerulonephritis associated with lupus serologies or antineutrophil cytoplasmic antibody–positive renal vasculitis occurs.

Cardiac Involvement

Cardiac involvement is frequently detected using sensitive diagnostic tools but is commonly clinically silent. Clinical cardiac involvement is more frequently seen in patients with dcSSc; it generally develops early in the course of the disease and is a poor prognostic factor. The endocardium, myocardium, and pericardium may be affected separately or together. Clinical manifestations include tachyarrhythmias, conduction abnormalities, valvular regurgitation, diastolic heart failure, and pericardial effusion. Systemic and pulmonary arterial hypertension, as well as lung and renal involvement, also affect the heart. Conventional echocardiography has a low sensitivity for detecting SSc heart involvement. Tissue Doppler echocardiography, single-photon emission computed tomography, and especially cardiac magnetic resonance imaging (cMRI) reveal a high prevalence of myocardial abnormalities such as abnormal ventricular relaxation and reversible perfusion defects. An elevated level of serum NT-pro-BNP is a sensitive marker for increased pulmonary artery pressure but may also indicate primary cardiac involvement. Myocarditis can develop in association with muscle inflammation. Pericardial effusion develops in more than 15% of SSc patients but is rarely significant.

Musculoskeletal Complications

Carpal tunnel syndrome (Chapter 420) occurs frequently and may be a presenting manifestation of SSc. Joint mobility is progressively impaired, especially in the hands. Large joint contractures can be accompanied by audible or palpable tendon friction rubs that are caused by extensive fibrosis and adhesion of the tendon sheaths and fascial planes at the affected joint. The presence of tendon friction rubs often signals aggressive disease. Frank joint inflammation is uncommon in SSc; however, erosive polyarthritis in the hands can occur. Muscle weakness may be a sign of deconditioning, disuse atrophy, and malnutrition. Less commonly, inflammatory myositis indistinguishable from idiopathic polymyositis (Chapter 269) occurs in early disease. A noninflammatory myopathy characterized by atrophy and fibrosis in the absence of elevated muscle enzyme levels may occur in late disease. Bone resorption affects the distal tufts of terminal phalanges (acro-osteolysis), mandibular condyles, ribs, and distal clavicles.

Other Clinical Manifestations

In addition to microangiopathy, involvement of larger blood vessels (>100 μm) is common in SSc. Manifestations of macrovascular disease include occlusion of the digital and ulnar arteries, leading to ischemic ulcerations and even loss of digits or limbs. Epidemiologic studies indicate increased risk of coronary artery disease in patients with SSc. Dry eyes and dry mouth are common in SSc, but in contrast to Sjögren syndrome (Chapter 268), salivary gland biopsy in such cases shows fibrosis rather than focal lymphocytic infiltration. Hypothyroidism due to thyroid fibrosis is common and may be associated with antithyroid autoantibodies. Although the brain and central nervous system are generally spared in SSc, autonomic neuropathy, as well as a primarily sensory neuropathy of the trigeminal nerve due to fibrosis or vasculopathy, can occur. Pregnancy in women with active SSc is associated with an increased rate of adverse fetal outcomes. Furthermore, cardiopulmonary involvement might worsen during pregnancy, and scleroderma renal crisis can occur. Inability to attain or maintain penile erection is due to vascular insufficiency and fibrosis; the problem is frequent and may be the presenting disease manifestation in males with SSc.

Systemic Sclerosis and Cancer

Patients with SSc have an increased risk of cancer. In these patients, lung cancer and esophageal adenocarcinoma typically occur in the setting of long-standing interstitial lung disease or GERD, and chronic inflammation and tissue repair may be contributing factors. In contrast, breast, lung, and ovarian carcinoma and lymphoma in SSc tend to occur in close temporal association with the clinical onset of SSc and are often associated with anti-RNA polymerase III antibodies. In these cases, SSc might be a paraneoplastic syndrome that is triggered by the antitumor immune response.

DIAGNOSIS

Skin induration in the fingers or proximally (associated with Raynaud phenomenon) and characteristic visceral organ manifestations are sufficient to establish the diagnosis of SSc. Occasionally, diagnostic full-thickness skin biopsy may be required for ruling out scleroderma mimics such as scleredema, scleromyxedema, or nephrogenic systemic fibrosis (see Table 267-1). Primary Raynaud phenomenon is differentiated from SSc by normal-appearing nailfold capillaries and absence of autoantibodies. Diagnosing SSc can be difficult in the early stages of the disease because initial symptoms and findings are often nonspecific and can be mistaken for rheumatoid arthritis, SLE, myositis, or undifferentiated connective tissue disease. Rarely, patients with SSc first present with accelerated hypertension or GI bleeding caused by watermelon stomach as the initial manifestation.

Laboratory Features

Anemia is common and may reflect chronic inflammation, GI bleeding from gastric vascular ectasia, erosive gastritis or chronic esophagitis, or folate and vitamin B_{12} deficiency due to small bowel bacterial overgrowth and malabsorption. Microangiopathic hemolytic anemia (Chapter 160) caused by mechanical trauma and red blood cell fragmentation is a hallmark of scleroderma renal crisis. In contrast to other connective tissue diseases, the erythrocyte sedimentation rate and C-reactive protein generally show only modest elevation. Monitoring serum levels of prealbumin and vitamin K is useful in patients with small bowel bacterial overgrowth and malabsorption.

Antinuclear autoantibodies (ANAs) are present in virtually all patients with SSc and can be detected at, or even prior to, disease onset. Autoantibodies specific for SSc are described in Table 267-3. Anticentromere antibodies are associated with PAH, but cardiac involvement, significant pulmonary fibrosis, or scleroderma renal crisis occurs only rarely in these patients. Topoisomerase-I antibody positivity is associated with reduced survival, whereas anticentromere antibody–positive patients have improved survival compared with those without this antibody. Antibodies to RNA polymerase III (recognized based on its speckled immunofluorescence pattern) are associated with increased risk for scleroderma renal crisis. Antibodies to β_2-glycoprotein I are not specific but in SSc identify increased risk for critical ischemia.

TREATMENT AND PREVENTION Rx

With the exception of ACE inhibitors for scleroderma renal crisis, no therapy to date has been shown to significantly alter the natural history of SSc. In contrast, organ-based treatments are effective in alleviating symptoms and slowing progression of the cumulative organ damage. A significant reduction in disease-related mortality has occurred during the past 25 years. Treatment must be tailored to each patient's unique needs. Because of the marked heterogeneity in clinical presentation, a thorough and individualized baseline evaluation is paramount. Optimal management incorporates the following principles: prompt diagnosis, accurate classification and risk stratification, early recognition and assessment of organ-based complications, and monitoring progression, disease activity, and response to therapy. Management of complications should be proactive, with regular screening and initiation of appropriate intervention at the earliest possible opportunity. Given the multisystemic nature of SSc, an integrated team-based management approach, typically at specialized medical centers, is desirable. The team should incorporate appropriate medical specialists. Patients are empowered by learning about potential complications, therapeutic options, and the natural history of their disease.

Disease-Modifying Therapy

Immunosuppressive Agents

Immunosuppressive agents that are highly effective in the treatment of other connective tissue diseases have generally shown modest or no benefit in SSc.[12] Corticosteroids may alleviate stiffness and aching in early-stage dcSSc but do not slow the progression of skin or internal organ involvement and are associated with an increased risk for scleroderma renal crisis. Therefore, corticosteroids should be avoided if possible; when absolutely necessary, they should be given at the lowest dose possible and for brief periods only.

Cyclophosphamide was shown to reduce the progression of symptomatic interstitial lung disease in early SSc.[■■] Compared with placebo, patients treated with oral cyclophosphamide showed stabilization and, rarely, modest improvement in respiratory symptoms, pulmonary function, and abnormalities on chest HRCT after 1 year of treatment, but these benefits were short-lived. The use of cyclophosphamide in SSc needs to be balanced against its potential for side effects, including bone marrow suppression, opportunistic infections, hemorrhagic cystitis, bladder cancer, and premature ovarian failure.

In small clinical trials, methotrexate was associated with a modest improvement in skin involvement. Mycophenolate mofetil treatment was shown to improve skin involvement and stabilize lung disease. There is some support in the literature for the use of immunomodulatory agents and interventions including rituximab, intravenous immunoglobulin, and extracorporeal photopheresis for the treatment of SSc. Recent reports suggest that rituximab might be effective in ameliorating skin and lung involvement. In patients with severe SSc who fail to respond to other treatments (Chapter 178), autologous hematopoietic stem cell transplantation (HSCT) improves long term, event-free survival despite an increased treatment-related mortality in the first year.[■] Because of this potential morbidity and mortality and its substantial cost, HSCT is presently considered an investigational therapy for SSc.

Antifibrotic Therapy

Because tissue fibrosis causes progressive and irreversible organ damage, drugs that block or slow the fibrotic process represent a rational approach to therapy. D-Penicillamine has been extensively used as an antifibrotic agent. In retrospective studies, D-penicillamine stabilized and improved skin induration, prevented new internal organ involvement, and improved survival. However, in a randomized controlled clinical trial, there was no difference in the extent of skin involvement between patients treated with standard-dose (750 mg/day) or very low-dose (125 mg every other day) D-penicillamine. Minocycline, bosentan, recombinant relaxin, interferon-γ, and inhibitors of tumor necrosis factor are putative antifibrotic agents that have failed to show meaningful benefit in SSc clinical trials. Small-molecule inhibitors of protein tyrosine kinases used in malignancies (e.g., imatinib, nilotinib, and dasatinib) block signaling by TGF-β and PDGF and thereby prevent fibrotic responses in vitro and in vivo. These agents are currently in clinical trials for SSc.

Treatment of Organ-Specific Complications

Gastrointestinal Complications

Because significant gastroesophageal reflux may occur in the absence of symptoms, all patients with SSc should be treated for this complication. Proton pump inhibitors may need to be given in relatively high doses and for prolonged periods, and patients should be instructed to elevate the head of the bed and eat frequent small meals. Recurrent GI bleeding due to GAVE can be treated with laser or argon plasma photocoagulation. Bacterial overgrowth due to small bowel hypomotility causes bloating and diarrhea and may lead to malabsorption, weight loss, and malnutrition. Treatment with short courses of rotating broad-spectrum antibiotics such as metronidazole, erythromycin, and tetracycline can eradicate bacterial overgrowth, but many patients relapse when antibiotics are stopped. In patients with malnutrition but intact small bowel function, enteral nutrition via a jejunostomy can be effective. In others, total parenteral nutrition may be indicated. Refractory hypomotility of the small bowel may respond to subcutaneous octreotide injections. Anorectal complications may respond to sacral neuromodulation.

Vascular Therapy and Raynaud Phenomenon

The goal of therapy is to reduce the frequency and duration of vasospastic episodes, prevent ischemic complications and enhance their healing, and slow the progression of obliterative vasculopathy. Patients should dress warmly, minimize cold exposure, and avoid drugs that could precipitate or exacerbate vasospastic episodes. Calcium-channel blockers such as nifedipine and diltiazem are used commonly for Raynaud's phenomenon but show only moderate benefit, and their use is often limited by side effects (palpitations, dependent edema, light-headedness). ACE inhibitors do not reduce the frequency or severity of episodes, but angiotensin II receptor blockers such as losartan are effective and generally well tolerated. Patients with severe Raynaud phenomenon require α_1-adrenergic receptor blockers (e.g., prazosin), 5-phosphodiesterase inhibitors (e.g., sildenafil), topical nitroglycerine, or intravenous prostaglandins. Low-dose aspirin and dipyridamole prevent platelet aggregation and may have a role as adjunctive agents but must be used with caution in light of the risk of bleeding from GAVE lesions. The endothelin-1 receptor antagonist bosentan reduces development of new ischemic ulcers. Intravenous prostacyclin infusion, local injections of botulinum toxin, and digital sympathectomy are options for some patients with critical digital ischemia. Patients with ischemic digital ulcerations may require

surgical débridement, especially if necrotic tissue is present. Empirical long-term therapy with statins and antioxidants may slow the progression of vascular damage.

Pulmonary Arterial Hypertension

All patients with SSc should be screened for PAH at initial evaluation, and those at high risk on a yearly basis. Treatment for symptomatic PAH should be started with an endothelin-1 receptor antagonist or a 5-phosphodiesterase inhibitor. Diuretics, oral anticoagulation, and digoxin may be used when appropriate. If hypoxemia is documented, supplemental oxygen should be given. If clinical response is inadequate, 5-phosphodiesterase inhibitors may be used in combination with endothelin-1 receptor antagonists. Prostacyclin analogues can be administered intravenously, by continuous subcutaneous infusion, or by frequent inhalations. Lung transplantation remains an option for selected patients with SSc-associated PAH who fail medical therapy.

Treatment and Prevention of Scleroderma Renal Crisis

Prompt recognition of impending or early scleroderma renal crisis is essential. Because patients with early-stage SSc and progressive skin involvement are at highest risk, they should monitor their blood pressure daily and report significant alterations immediately. Corticosteroids should be used only when absolutely necessary and at the lowest possible doses. When scleroderma renal crisis occurs, patients should be hospitalized and treatment with short-acting ACE inhibitors started immediately to achieve prompt blood pressure normalization. There is no evidence that "prophylactic" use of ACE inhibitors can prevent the development of scleroderma renal crisis or ameliorate its severity. Although up to two thirds of patients who develop renal crisis require dialysis, delayed recovery of renal function can occur. Kidney transplantation is appropriate for patients unable to discontinue dialysis after 2 years. Survival with renal transplantation in SSc is comparable to that in other connective tissue diseases, and recurrence of scleroderma renal crisis is rare.

Skin Care

Skin involvement in early SSc is inflammatory and can be controlled with systemic antihistamines or short-term low-dose corticosteroids. Because of the increased risk for scleroderma renal crisis, blood pressure should be carefully monitored. Cyclophosphamide, methotrexate, D-penicillamine, and mycophenolate have been associated with modest improvement in skin induration in early-stage SSc. Skin dryness can be managed with the use of hydrophilic ointments and emollient bath oils. Fingertip ulcerations should be protected by occlusive dressing to promote healing and prevent infection. Infected skin ulcers are treated with topical or oral antibiotics and may necessitate surgical débridement. No medical therapy has been shown to be effective in preventing soft tissue calcification or in promoting its dissolution, and surgical therapy is only occasionally effective.

PROGNOSIS AND NATURAL HISTORY

Patients with dcSSc have a more rapidly progressive course, greater internal organ involvement, and worse prognosis compared to those with lcSSc. However, the outcome of the disease is difficult to predict.

Early inflammatory symptoms of dcSSc such as fatigue, edema, arthralgia, and pruritus commonly subside after 2 to 4 years, and skin thickening reaches a plateau followed by slow regression, which characteristically occurs in an order that is the reverse of initial involvement, with softening on the trunk followed by proximal and finally the distal extremities. Sclerodactyly and finger contractures generally persist. Relapse or recurrence of skin thickening is rare. Visceral organ involvement develops and progresses most rapidly during the initial 2 to 4 years of the disease. New organ involvement rarely occurs once the skin involvement has plateaued. Similarly, scleroderma renal crisis almost invariably occurs within the first 4 years of disease. In patients with lcSSc, Raynaud phenomenon may precede other disease manifestations by years or even decades, and visceral organ complications such as PAH and primary biliary cirrhosis generally occur late in the course of the disease.

Age- and gender-adjusted mortality rates in patients with SSc are more than five-fold higher than in the general population. The 10-year survival rate is 55% for patients with dcSSc and 75% for patients with lcSSc. Survival correlates with the extent of skin involvement, which represents a surrogate for visceral organ involvement. The leading causes of death are pulmonary fibrosis, PAH, severe GI involvement, and cardiac disease. Markers of poor prognosis include male sex, African American race, older age of disease onset, low body mass index, extensive skin thickening with truncal involvement, and evidence of significant or progressive visceral organ involvement. Autoantibodies to topoisomerase-I or absence of anticentromere antibodies are markers of poor prognosis. In one study, SSc patients who had extensive skin involvement, vital capacity less than 55% of predicted, significant GI involvement, and clinically evident cardiac involvement or scleroderma renal crisis had a less than 40% 10-year survival. The severity of PAH is correlated with mortality, and SSc patients with a mean pulmonary arterial pressure of 45 mm Hg or higher had a 33% 3-year survival rate. In scleroderma renal crisis, therapy with ACE inhibitors has had a dramatic effect on survival, increasing from less than 10% at 1 year in the pre-ACE inhibitor era to better than 70% 3-year survival at the present time.

NEPHROGENIC SYSTEMIC FIBROSIS

Nephrogenic systemic fibrosis (NSF) is a novel complication of renal insufficiency with certain clinical features resembling SSc.[13] The condition was initially described in 2000 and is now recognized as an emerging problem in patients with chronic renal failure. It is estimated that 2% of patients on long-term hemodialysis might develop NSF. Originally considered a purely dermatologic scleromyxedema-like condition and termed *nephrogenic fibrosing dermatopathy*, NSF is now recognized to be associated with visceral organ involvement and is therefore more accurately termed *nephrogenic systemic fibrosis*. The cutaneous manifestations of NSF share histopathologic and clinical features with other scleroderma-spectrum disorders, notably fasciitis and scleromyxedema. In most patients with NSF, the condition develops while undergoing long-term dialysis. However, no association with a particular route or type of renal replacement therapy has been demonstrated. Furthermore, NSF has also been described in patients who have never received dialysis. Histologic hallmarks include cutaneous fibrosis with mucin deposition and accumulation of spindle-shaped cells, including numerous CD34-positive cells, in the lesional skin.

The clinical hallmark of NSF is thickening and "woody" tightness of skin over the lower and, less commonly, upper extremities and contractures at large joints. A link between NSF and exposure to gadolinium-containing MRI contrast agents was suggested in 2006, leading to a warning by the U.S. Food and Drug Administration regarding the use of these agents in patients with renal insufficiency. This was followed by a substantial decline in the incidence of NSF. The course of NSF is generally progressive, and the prognosis is poor. Some patients show improvement with adjustment to renal replacement therapy, and others respond to renal transplantation. Anecdotal reports describe treatment with phototherapy, imatinib mesylate, and immunosuppressive agents. However, in most patients with NSF, the induration is resistant to therapy and leads to progressive induration, joint contractures, and reduced mobility.

The topic of immunoglobulin (Ig)G4-related disease is discussed in Chapter 275.

Grade A References

A1. Tashkin DP, Celli B, Senn S, et al. Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med*. 2006;354:2655-2666.

A2. Hoyles RK, Ellis RW, Wellsbury J, et al. A multicenter, prospective, randomized, double-blind, placebo-controlled trial of corticosteroids and intravenous cyclophosphamide followed by oral azathioprine for the treatment of pulmonary fibrosis in scleroderma. *Arthritis Rheum*. 2006;54:3962-3970.

A3. van Laar JM, Farge D, Sont JK, et al. Autologous hematopoietic stem cell transplantation vs intravenous pulse cyclophosphamide in diffuse cutaneous systemic sclerosis: a randomized clinical trial. *JAMA*. 2014;311:2490-2498.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Arthritis Rheum*. 2013;65:2737-2747.
2. Katsumoto TR, Whitfield ML, Connolly MK. The pathogenesis of systemic sclerosis. *Annu Rev Pathol*. 2011;6:509-537.
3. Trojanowska M. Cellular and molecular aspects of vascular dysfunction in systemic sclerosis. *Nat Rev Rheumatol*. 2010;6:453-460.
4. Bhattacharyya S, Wei J, Varga J. Understanding fibrosis in systemic sclerosis: shifting paradigms, emerging opportunities. *Nat Rev Rheumatol*. 2011;8:42-54.
5. Bussone G, Bérezné A, Pestre V, et al. The scleroderma kidney: progress in risk factors, therapy, and prevention. *Curr Rheumatol Rep*. 2011;13:37-43. *Review emphasizing use of ACE inhibitors and avoidance of corticosteroids*.
6. Goundry B, Bell L, Langtree M, et al. Diagnosis and management of Raynaud's phenomenon. *BMJ*. 2012;344:e289. *Review*.
7. van Bon L, Affandi AJ, Broen J, et al. Proteome-wide analysis and CXCL4 as a biomarker in systemic sclerosis. *N Engl J Med*. 2014;433-443.
8. O'Reilly S. Innate immunity in systemic sclerosis pathogenesis. *Clin Sci (Lond)*. 2014;126:329-337.
9. Bhattacharyya S, Tamaki Z, Wang W, et al. FibronectinEDA promotes chronic cutaneous fibrosis through Toll-like receptor signaling. *Sci Transl Med*. 2014;6:232ra50.
10. Akter T, Silver RM, Bogatkevich GS. Recent advances in understanding the pathogenesis of scleroderma-interstitial lung disease. *Curr Rheumatol Rep*. 2014;16:411.
11. Mouthon L, Bérezné A, Bussone G, et al. Scleroderma renal crisis: a rare but severe complication of systemic sclerosis. *Clin Rev Allergy Immunol*. 2011;40:84-91. *Treatment relies on aggressive control of blood pressure with ACE inhibitors, but dialysis is frequently needed, and 5-year survival is 65%*.
12. Denton CP, Ong VH. Targeted therapies for systemic sclerosis. *Nat Rev Rheumatol*. 2013;9:451-464.
13. Daftari Besheli L, Aran S, Shaqdan K, et al. Current status of nephrogenic systemic fibrosis. *Clin Radiol*. 2014;69:661-668.

REVIEW QUESTIONS

1. Which of the following does not contribute to the pathogenesis of systemic sclerosis (SSc)?

A. Vascular wall remodeling
B. Tissue fibrosis
C. Vasculitis
D. T cells
E. Hypoxia

Answer: C Vasculitis does not typically occur in SSc. Vascular remodeling leading to hypoxia, inflammation, and fibrosis are the hallmarks of the disease.

2. Which is a characteristic autoantibody seen in SSc patients?

A. Anticentromere
B. Anti-CCP
C. Antihistone
D. Antiphospholipid
E. Anti-Smith

Answer: A Anticentromere is a hallmark SSc-associated autoantibody, whereas the other autoantibodies are seen in rheumatoid arthritis (B), lupus (C, E), and the antiphospholipid syndrome (D).

3. Which is a major factor in SSc morbidity?

A. Glomerulonephritis
B. Episcleritis
C. Amyloidosis
D. Pulmonary hypertension
E. Bowel infarction

Answer: D Pulmonary hypertension develops in up to 15% of SSc patients and is a major cause of morbidity and mortality. The other clinical features are uncommon in SSc and indicate other autoimmune or rheumatic diseases.

4. In patients with SSc-associated lung disease, lung biopsy may show:

A. Honeycombing
B. Interstitial fibrosis
C. Plasma cell accumulation
D. Intimal proliferation in the small vessels
E. All of the above

Answer: E All of the listed pathologic features can be noted in lung biopsies from patients with SSc-associated interstitial lung disease.

5. Which of the following drugs may contribute to scleroderma renal crisis?

A. NSAIDs
B. Glucocorticoids
C. Mycophenolate
D. Rituximab
E. Penicillamine

Answer: B The use of glucocorticoids has been shown to increase the risk of new-onset scleroderma renal crisis in patients with SSc.

268

SJÖGREN SYNDROME

XAVIER MARIETTE

DEFINITION

Sjögren syndrome (SS) is a systemic autoimmune disease characterized by lymphocytic infiltrates of salivary and tear glands, leading to oral and ocular dryness, and by autoantibody secretion. It can be encountered either alone

(primary Sjögren syndrome [pSS]) or in the presence of other systemic autoimmune diseases (secondary Sjögren syndrome [sSS]) like rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), inflammatory myositis, and systemic sclerosis. SS in the setting of RA usually follows RA diagnosis by many years and is mainly manifested by keratoconjunctivitis sicca, with systemic features being rather uncommon. Associated with other systemic autoimmune disease, the presentation of sSS is very close to pSS. Of note, pSS may be also associated with organ-specific systemic autoimmune disease, such as autoimmune thyroiditis and primary biliary cirrhosis.

EPIDEMIOLOGY

Primary SS is a common disease that affects 0.1 to 0.6% of the general adult female population.[1] A higher prevalence of the disease has been reported (0.5 to 2%), but this must be considered with caution because the reported prevalence of SS depends on the classification criteria used in the various studies, and the prevalence of sicca symptoms in the general population is high. Conversely, in recent studies with strict criteria, a lower prevalence has been found: 1.02 per 10,000 adults.[2] Primary SS has a female preponderance (female-to-male ratio at least 9 : 1). The age peak of the disease occurs after menopause in the mid-50s.

PATHOPHYSIOLOGY

Recent years have witnessed major advances in the pathophysiologic mechanisms of the disease. Several studies have confirmed the role of innate immunity, genetics, and B-cell activation and the relation between abnormalities in them.

The presence of an interferon (IFN) signature has been shown both in salivary glands and blood.[3] Plasmacytoid dendritic cells, the professional cells secreting type 1 IFN, are present within the glands. Type 2 IFN-dependent genes can be overexpressed in salivary glands. Natural killer (NK) cells, another actor of innate immunity able to secrete type 2 IFN are present in salivary glands of patients and play a role in the disease.[4]

In line with this IFN signature, multiple viral agents have been incriminated as etiologic factors for either the development or the modulation of SS; these include Epstein-Barr virus, retroviruses, and coxsackieviruses, but in all cases the data remain controversial.[5]

The genetics of pSS is now better understood with the reports of two genome-wide association studies (GWAS).[6] Like in other systemic autoimmune diseases, HLA is the most important region associated with the disease, and especially HLA-DR3-DQ1 in patients with autoantibodies. Interestingly, other genes associated with the disease are involved in the IFN response. These include IFN regulatory factor 5 *(IRF-5)*, a pivotal transcription factor in the type 1 IFN pathway, and signal transducer and activator of transcription 4 *(STAT-4)*, and *IL-12A*, involved in the type 2 IFN pathway. Other genes found to be associated with the disease are *TNIP1*, playing a role in control of nuclear factor (NF)-κB activation, and *CXCR5*, involved in germinal center formation.

The presence of ectopic salivary gland germinal centers demonstrates the importance of B-cell activation in pSS. Different cytokines may explain this B-cell activation. Several studies have focused on the role of BAFF (B-cell activating factor of the tumor necrosis factor [TNF] family), a cytokine that promotes B-cell maturation, proliferation, and survival. It has been shown that BAFF is enhanced in sera and in salivary glands from pSS patients. Interestingly, BAFF can be secreted by salivary gland epithelial cells, the target of autoimmunity, after stimulation by the innate immune system (type 1 or type 2 IFN, or viral infections). Thus, this cytokine is likely to be a link between innate immunity and autoimmunity.

The current hypothetical scenario for the development of pSS is based on the successive activation of innate and adaptive immune systems (Fig. 268-1). Environmental factors such as viral infections or hormonal imbalance may act at the initial stage of the disease by activating epithelial cells. This epithelial cell activation is promoted in patients who carry susceptibility factors in the genes for IFN pathway proteins. These patients experience a greater degree of IFN pathway activation, which leads to BAFF overproduction,

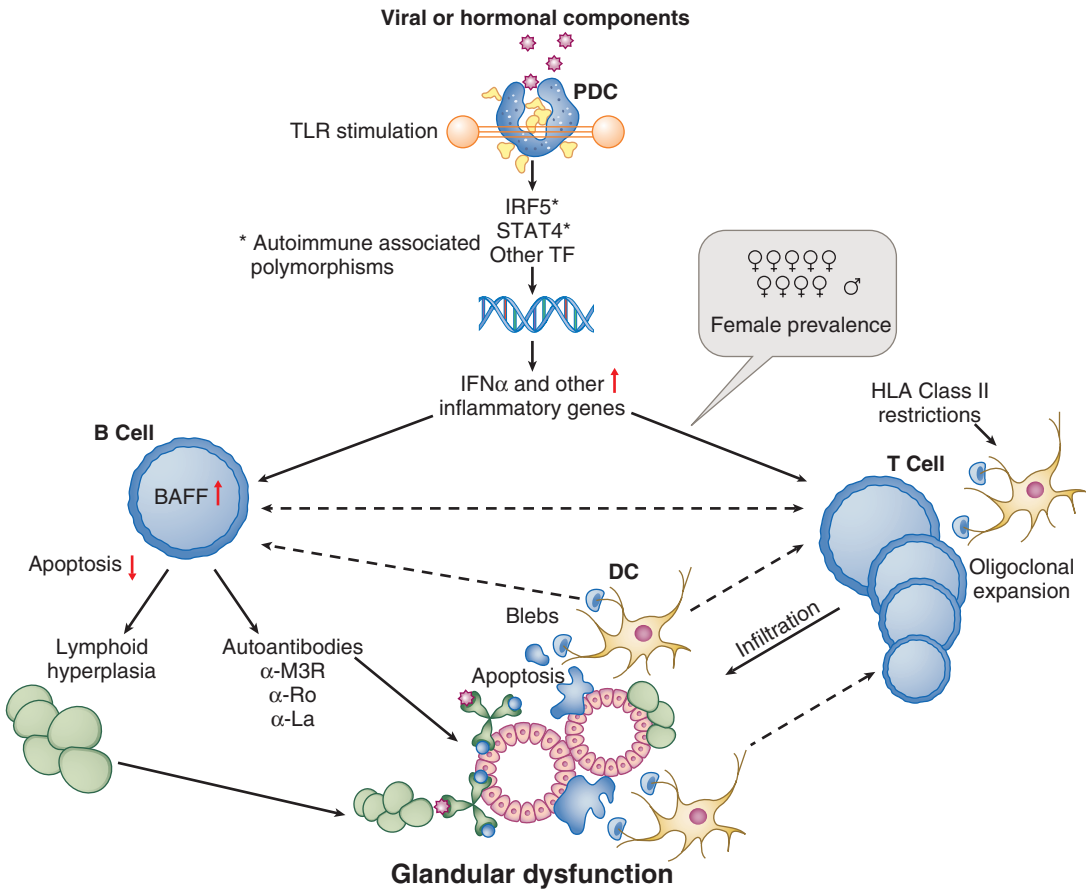

FIGURE 268-1. Hypothetical scenario for development of primary Sjögren syndrome. An environmental factor (e.g., virus) causes epithelial cell and dendritic cell (DC) activation. Plasmacytoid DCs are also activated by immune complexes, promoting interferon (IFN) pathway activation, which leads to BAFF overproduction and to B- and T-cell activation. B-cell activation leads to autoantibody production within germinal center–like structures. Interleukin-12 secreted by myeloid DCs leads to natural killer cell and T-helper 1 activation, which promotes tissue damage and IFN-γ production. IFN-α and IFN-γ enhance BAFF secretion. Epithelial cells release autoantigens that participate in immune complex formation and perpetuate the vicious cycle of immune system overactivation. BAFF = B-cell activating factor of the tumor necrosis factor family; IRF5 = interferon regulatory factor 5; PDC = plasmacytoid dendritic cell; STAT4 = signal transducer and activator of transcription 4; TF = transcription factors; TLR = toll-like receptors.

B- and T-cell activation, and secretion of autoantibodies, especially in predisposed patients. These autoantibodies constitute immune complexes that participate in the maintenance of IFN-α production. Altogether, these steps promote a vicious cycle of immune system activation leading to tissue damage.

CLINICAL MANIFESTATIONS

Glandular

Decreased salivary secretion results in mouth dryness and increased incidence of oral infections, mucosal friability, and dental caries due to loss of the lubricating, buffering, and antimicrobial capacities of saliva.[7] Fungal infections (primarily candidiasis) are also common. Parotid salivary gland or other major salivary gland enlargement can also occur. Persistent enlargement should be carefully followed, however, to exclude bacterial superinfection and, more importantly, the development of lymphoma.

Decreased lacrimal flow and impaired lacrimal composition lead to damage of the corneal and conjunctival epithelia, a condition known as keratoconjunctivitis sicca. As a result of keratoconjunctivitis sicca, SS patients might experience foreign-body sensation, grittiness, irritation, photosensitivity, and thick rope-like secretions at the inner canthus, all leading to increased discomfort and possibly visual impairment, with considerable functional disability. Furthermore, ocular complications include corneal ulceration and scarring, bacterial keratitis, and eyelid infections that require continuous ophthalmologic care and treatment.

Systemic

In addition to the sicca features, systemic manifestations occur in approximately 20 to 30% of pSS patients. Of note, it has been increasingly appreciated that the extraglandular manifestations in SS can be divided into two major types according to the underlying pathophysiologic mechanism. Thus, lymphocytic infiltration of the epithelia of organs beyond the exocrine glands (e.g., renal, liver, and bronchial epithelial cells) results in interstitial nephritis, autoimmune cholangitis, and obstructive bronchiolitis, respectively. These clinical features seem to appear early and usually have a benign course. On the other hand, immune complex deposition as a result of the ongoing B-cell hyperreactivity can give rise to the extraepithelial manifestations—palpable purpura, glomerulonephritis, interstitial pneumonitis, and peripheral neuropathy—that are linked to increased morbidity and risk for lymphoma development. The main systemic manifestations are listed in Table 268-1. Peripheral neuropathy may occur through various mechanisms. Vasculitis may be present with cryoglobulinemia, leading to both sensory and motor symptoms. More frequently, pure sensory neuropathy is present, sometimes purely ataxic and sometimes in the form of small-fiber neuropathy. This latter entity is difficult to diagnose because clinical and electromyographic examinations are normal. The diagnosis may be made by skin biopsy showing rarefaction of sensory small fibers.

Sjögren Syndrome and Non-Hodgkin Lymphomas

Chronic polyclonal B-cell activation is commonly present in pSS, which may explain why this autoimmune disease has the strongest association with the development of B-cell lymphoma (relative risk, 15 to 20). More recent studies have estimated this risk at a lower level: 6 in Denmark and Sweden, 7 in Taiwan, and 9 in Norway.

Lymphomas complicating pSS have specific features (Chapter 185). They are mostly B-cell non-Hodgkin lymphomas with a predominance of low-grade, marginal-zone histologic type. Mucosal localization is predominant, notably as mucosa-associated lymphoid tissue (MALT) lymphomas. Interestingly, lymphomas often develop in organs where pSS is active, such as salivary glands.

In the setting of SS, chronic autoimmune B-cell activation plays the major role in the lymphomagenesis process, and the identified predictors of lymphoma development in pSS are in line with this phenomenon. The main clinical predictors are permanent swelling of salivary glands, splenomegaly, lymphadenopathy, and palpable purpura. The main biological predictors are positivity of rheumatoid factor (RF), cryoglobulinemia, lymphopenia (especially CD4 lymphopenia), low complement levels, and a monoclonal component in serum or urine. Three novel predictive factors for lymphoma development have been recently described: (1) the presence of ectopic germinal centers associated with the occurrence of lymphoma in pSS patients[8]; (2) demonstration that BAFF levels are increased in pSS patients with current or previous lymphoma compared with patients without lymphoma[9]; (3) abnormalities of the gene *TNFAIP3* coding for the A20 protein that regulates NF-κB activation, found in up to 77% of MALT lymphomas complicating pSS.[10] In half of the cases, *TNFAIP3* mutations or deletions occur within lymphoma cells; in the other 50%, they involve germline *TNFAIP3* mutations with functional consequences.

TABLE 268-1 EXTRAGLANDULAR MANIFESTATIONS OF PRIMARY SJÖGREN SYNDROME

CONSTITUTIONAL SYMPTOMS
Fatigue
Low-grade fever
SKIN AND VASCULAR
Small vessel vasculitis
Raynaud phenomenon
Photosensitivity reactions similar to subacute cutaneous systemic lupus erythematosus
Xerosis
UPPER AND LOWER AIRWAYS
Pyogenic sialoadenitis or parotitis
Interstitial pneumonitis or fibrosis
Chronic bronchitis
Bronchiectasis
Bronchiolitis obliterans with organizing pneumonia
Chronic obstructive pulmonary disease
MUSCULOSKELETAL
Polyarthralgia, polyarthritis
Myopathy, polymyositis
RENAL
Type I renal tubular acidosis
Tubular interstitial nephritis
Glomerulonephritis
NEUROLOGIC
Peripheral motor sensory neuropathy
Pure sensory neuropathy (including pure ataxic neuropathy)
Small fiber sensitive neuropathy
Multiple sclerosis–like focal lesions
Spinal cord dysfunction, including transverse myelitis
NEOPLASIA
Lymphadenopathy, MALT (mucosa-associated lymphoid tissue) lymphoma

Laboratory Findings

The most common serologic finding in pSS is hypergammaglobulinemia. The elevated γ-globulins contain several autoantibodies directed against non–organ-specific antigens, such as RF and antinuclear antibody (ANA). Specific ANA, anti-SSA/Ro, and anti-SSB/La antibodies are present in 60 to 80% and 30 to 40% of patients, respectively, and anti-SSB/La is never present without anti-SSA/Ro. Of note, the presence of anti-SSA/Ro, possibly with anti-SSB/La, may mediate complete heart block of newborns owing to cross-mimicry between specific fetal myocardial antigens and epitopes of the SSA/Ro-SSB/La complex.

Anemia of chronic inflammation and high erythrocyte sedimentation rates (due to hypergammaglobulinemia) are frequently encountered, whereas C-reactive protein levels are usually within normal limits. Cytopenias (most frequently lymphopenia and neutropenia) can also occur. In the setting of interstitial nephritis, the presence of hypokalemic, hyperchloremic acidosis might reveal distal renal tubular acidosis.

A monoclonal immunoglobulin can be detected in 10 to 15% of patients with SS, depending on the technique used. Approximately 20% of patients with SS have cryoglobulins in their sera. Complement levels may be decreased, especially C4. This low C4 level may be either genetically determined or secondary to consumption (in immune complexes or cryoglobulinemia).

DIAGNOSIS

Differential Diagnosis

The definition of pSS had suffered for a long time from the absence of accurate and consensus-driven diagnostic criteria. This is important because the patients' main symptoms (dryness, fatigue, and pain) are frequent in the general population. They can be caused by numerous drugs (Table 268-2), anxiety and/or depression, other comorbidities, or aging (Table 268-3). Sarcoidosis can mimic the clinical picture of SS. However, in sarcoidosis minor salivary gland biopsy reveals noncaseating granulomas, and autoantibodies

TABLE 268-2 DRUGS AND TOXINS THAT MIGHT DECREASE LACRIMAL AND SALIVARY SECRETION

STRONG EFFECT	MODERATE EFFECT
Atropine, atropinic antiparkinsonian drugs, anticholinergic antihistaminic drugs	β-Adrenergic blockers
Antidepressants: imipraminic (amitriptyline) and inhibitors of monoamine oxidase	α-Adrenergic blockers
Neuroleptics	Calcium channel blockers
Morphine, codeine, tramadol	Benzodiazepines
A-type botulinum toxin	Inhibitors of serotonin reuptake (very slight effect)
Class IA antiarrhythmic (disopyramide)	H_1 antihistaminic drugs
Isotretinoin	Diuretics
Toxins and psychotropic drugs: tobacco, ecstasy, cannabis, cocaine	Some antiretroviral drugs

TABLE 268-3 THE DIFFERENT CAUSES OF SICCA SYMPTOMS

Drugs, particularly psychotropic drugs (see Table 268-2)
Aging, postmenopausal estrogen deficiency
Prolonged use of contact lenses
Fibromyalgia and chronic fatigue syndrome
Anxiodepressive syndromes
Head and neck radiotherapy
Diabetes (uncontrolled)
Severe hyperlipidemia
Amyloidosis
Sarcoidosis
Lymphoma
Graft-versus-host disease
Some viral infections (HIV, HCV, HTVL-1)
IgG4-related sialoadenitis
Sjögren syndrome

HCV = hepatitis C virus; HIV = human immunodeficiency virus; HTVL-1 = human T-lymphocytic virus-1.

are typically absent. Other SS mimickers include chronic graft-versus-host disease, amyloidosis, infection with viruses such as HIV, human T-lymphocytic virus-I (HTLV-I), and hepatitis C virus (HCV), and IgG4-related disease (Chapter 275). The latter disease is important in the differential diagnosis of SS. It more often involves men with salivary or lacrimal gland enlargement (previously called Mickuliz disease) with previous organ-specific autoimmune disease (like autoimmune pancreatitis) without anti-SSA/SSB antibodies. Sicca symptoms without salivary lymphoid infiltrate and without anti-SSA/SSB antibodies may be part of the fibromyalgia syndrome (Chapter 274), and several acronyms have been proposed for designating these patients: sicca asthenia polyalgia syndrome (SAPS) or dry eyes and mouth syndrome (DEMS).

Diagnostic Criteria

International agreement has established a definition of SS based on the American-European Consensus Group (AECG) criteria, which require the presence of either focal lymphocytic infiltrates in minor salivary glands with a focus score of 1 or more, or anti-SSA/SSB autoantibodies (Table 268-4). A new set of preliminary criteria for SS classification was proposed by an expert consensus panel (American College of Rheumatology [ACR]-Sjögren International Collaborative Clinical Alliance [SICCA]). According to these criteria, classification of an individual as a pSS patient requires the presence of two out of three of the following objective items: (1) a positive serum test for anti-Ro/SSA and/or anti-La/SSB antibodies, or positive rheumatoid factor (RF) and antinuclear antibody (ANA) (titer > 1: 320); (2) presence of keratoconjunctivitis sicca, defined by an ocular staining score over 3; and (3) presence of focal lymphocytic sialoadenitis, defined by a focus score of 1 focus/4 mm^2 or above in a labial salivary gland biopsy.[11]

Assessment of Activity of the Disease

An international expert group recently set up an SS activity score under the umbrella of the European League Against Rheumatism (EULAR). Two indices have been developed: (1) a patient-administered questionnaire to assess subjective features, the EULAR Sjögren Syndrome Patient Reported Index (ESSPRI), based on three different visual analogic scores: dryness, fatigue, and limb pain[12]; and (2) a systemic activity index to assess systemic complications, the EULAR Sjögren Syndrome Disease Activity Index (ESSDAI).[13] The latter index comprises 12 domains with 3 or 4 levels of activity for each domain. Determination of the threshold of moderate activity as well as the minimal clinically important improvement is in progress, with the objective to base inclusion criteria and primary end-points of future clinical studies on ESSDAI levels.

TABLE 268-4 CLASSIFICATION CRITERIA FOR SJÖGREN SYNDROME

I. OCULAR SYMPTOMS
Positive response to at least one of these three questions: 1. Have you had daily, persistent, troublesome dry eyes for more than 3 months? 2. Do you have a recurrent sensation of sand or gravel in the eyes? 3. Do you use tear substitutes more than three times a day?
II. ORAL SYMPTOMS
Positive response to at least one of these three questions: 1. Have you had a daily feeling of dry mouth for more than 3 months? 2. Have you had recurrent or persistently swollen salivary glands as an adult? 3. Do you frequently drink liquids to aid in swallowing dry food?
III. OCULAR SIGNS
Objective evidence of ocular involvement, defined as a positive result in at least one of the following two tests: 1. Schirmer test (≤5 mm in 5 min) 2. Rose bengal score (≥4 according to van Bijsterveld scoring system)
IV. HISTOPATHOLOGY
Focus score ≥1 in a minor salivary gland biopsy specimen (a focus is defined as an agglomerate of at least 50 mononuclear cells; the focus score is defined by the number of foci in 4 mm^2 of glandular tissue)
V. SALIVARY GLAND INVOLVEMENT
Objective evidence of salivary gland involvement, defined by a positive result in at least one of the following three diagnostic tests: 1. Salivary scintigraphy 2. Parotid sialography 3. Unstimulated salivary flow (≤1.5 mL in 15 min)
VI. AUTOANTIBODIES
Presence in the serum of the following autoantibodies: antibodies to Ro (SSA) or La (SSB) antigens, or both
RULES FOR CLASSIFICATION
In patients without any potentially associated disease, primary Sjögren syndrome is diagnosed if: Four of six criteria are met, including IV or VI; *or* Three of four criteria from III, IV, V, and VI are met For secondary Sjögren syndrome, criteria I or II plus any two from criteria III, IV, and V should be met
EXCLUSION CRITERIA
Preexisting lymphoma, AIDS, sarcoidosis, graft-versus-host disease, past head and neck radiation treatment, use of anticholinergic drugs, and hepatitis C

From Vitali C, Bombardieri S, Jonsson R, et al. Classification criteria for Sjögren syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis.* 2002;61:554-558.

TREATMENT

Symptomatic Treatment

A recent systematic review of the literature confirms benefits for muscarinic agonists (pilocarpine hydrochloride and more recently cevimeline hydrochloride) for sicca features (oral dryness and, to a lesser extent, ocular dryness). Topical cyclosporine collyrium (0.05%) also was effective for moderate or severe ocular dryness and inflammation in a randomized controlled trial versus placebo, as were 0.1% clobetasone butyrate eyedrops. Environmental measures (avoidance of hot air heating systems or excessive air conditioning, use of a humidifier, appropriate glasses to protect the eye from evaporating air flow) and "little means" (sugar-free chewing gums, regular water drinking, salivary substitutes) might be useful. Regular dental examinations and oral hygiene are crucial for reducing subsequent oral health issues (i.e., caries and periodontal disease associated with xerostomia). To treat pain, simple analgesics should be used first, particularly acetaminophen/paracetamol, which does not cause dryness.

Immunomodulatory Drugs

To date, no immunomodulatory drug has proved efficacious in pSS. Severe organ manifestations of pSS have to be treated in accordance with treatment modalities used in SLE or other connective tissue diseases. Randomized trials have assessed hydroxychloroquine in pSS and failed to demonstrate any clinical efficacy.[A3] In spite of these negative results on clinical outcomes, hydroxychloroquine is frequently used in pSS, especially to treat arthralgia with or without synovitis or purpura. Controlled studies are needed to assess the use of methotrexate, leflunomide, mycophenolate sodium, azathioprine, and cyclosporine. Intravenous gamma globulin (IVIG) has been used in the treatment of SS-associated sensorimotor neuropathies or non-ataxic sensory neuropathy without any necrotizing vasculitis.

Biologics

Two randomized controlled trials (RCT) of infliximab and etanercept did not show any efficacy of TNF-blocker agents in pSS on a composite primary outcome including limb pain, fatigue, and dryness visual analogue scales (VAS).[A4] B-cell targeting appears to be a promising strategy in pSS. Three randomized controlled trials assessed efficacy of the monoclonal anti-CD20 antibody (rituximab). In the first one, a significant improvement from baseline in a fatigue VAS was observed in the rituximab group but not in the placebo group.[A5] In the second one, rituximab demonstrated significant efficacy compared to placebo in improving stimulated salivary flow, the primary end-point, but also oral and ocular dryness, fatigue VAS, and systemic complications.[A6] In the third study, the composite primary end-point, using 4 VAS, was achieved at 6 weeks but not at 6 months.[A7] Lastly, recent data derived from the French Autoimmune and Rituximab (AIR) registry, including 78 pSS patients with mainly systemic manifestations, suggested the efficacy of rituximab on systemic manifestations in approximately two thirds of the patients.[14] Overall, rituximab seems to be useful in cases of persistent parotid swelling or systemic complications, especially in cryoglobulinemia-induced vasculitis. Inhibitors of BAFF, especially the anti-BAFF monoclonal antibody belimumab (which is approved in SLE), has been used in pSS in a first open phase 2 study, with promising results.[15]

FUTURE DIRECTIONS

SS is a model of autoimmune disease, because it can be primary or associated with other autoimmune diseases; it represents autoimmunity where the risk of lymphoma is most important. SS is the autoimmune disease for which the target tissue of autoimmunity is the most easily available, with the lip biopsy being necessary for diagnosis. Recent progress in pathophysiology has emphasized a number of similarities with SLE that support consideration of SS as a sort of lupus of the mucosa. Even if the pathogenetic mechanisms of the disease remain largely unknown, improved knowledge of the effector mechanisms will allow identification of new targets for future therapy. Moreover, with the recently validated composite activity scores of ESSPRI and ESSDAI, the tools are now available to begin new clinical trials with novel drugs for this disease that will improve the poor quality of life currently associated with it.

Grade A References

A1. Ramos-Casals M, Tzioufas AG, Stone JH, et al. Treatment of primary Sjögren syndrome: a systematic review. *JAMA*. 2010;304:452-460.

A2. Aragona P, Spinella R, Rania L, et al. Safety and efficacy of 0.1% clobetasone butyrate eyedrops in the treatment of dry eye in Sjogren syndrome. *Eur J Ophthalmol*. 2013;23:368-376.

A3. Gottenberg JE, Ravaud P, Puechal X, et al. Effects of hydroxychloroquine on symptomatic improvement in primary Sjogren syndrome: the JOQUER randomized clinical trial. *JAMA*. 2014;312:249-258.

A4. Mariette X, Ravaud P, Steinfeld S, et al. Inefficacy of infliximab in primary Sjogren's syndrome: results of the randomized, controlled Trial of Remicade in Primary Sjogren's Syndrome (TRIPSS). *Arthritis Rheum*. 2004;50:1270-1276.

A5. Dass S, Bowman SJ, Vital EM, et al. Reduction of fatigue in Sjögren syndrome with rituximab: results of a randomised, double-blind, placebo-controlled pilot study. *Ann Rheum Dis*. 2008;67:1541-1544.

A6. Meijer JM, Meiners PM, Vissink A, et al. Effectiveness of rituximab treatment in primary Sjögren syndrome: a randomized, double-blind, placebo-controlled trial. *Arthritis Rheum*. 2010;62:960-968.

A7. Devauchelle-Pensec V, Mariette X, Jousse-Jolin S, et al. Treatment of primary Sjögren syndrome with rituximab: a randomized trial. *Ann Intern Med*. 2014;160:233-242.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Qin B, Wang J, Yang Z, et al. Epidemiology of primary Sjogren's syndrome: a systematic review and meta-analysis. *Ann Rheum Dis*. 2014; [Epub ahead of print].
2. Maldini C, Seror R, Fain O, et al. Epidemiology of primary Sjögren syndrome in a French multiracial/ethnic area. *Arthritis Care Res (Hoboken)*. 2013;66:454-463.
3. Nocturne G, Mariette X. Advances in understanding the pathogenesis of primary Sjögren syndrome. *Nat Rev Rheumatol*. 2013;9:544-556.
4. Rusakiewicz S, Nocturne G, Lazure T, et al. NCR3/NKp30 contributes to pathogenesis in primary Sjogren's syndrome. *Sci Transl Med*. 2013;5:195-196.
5. Kivity S, Arango MT, Ehrenfeld M, et al. Infection and autoimmunity in Sjogren's syndrome: a clinical study and comprehensive review. *J Autoimmun*. 2014;51:17-22.
6. Lessard CJ, Li H, Adrianto I, et al. Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjögren syndrome. *Nat Genet*. 2013;45:1284-1292.
7. Mavragani CP, Moutsopoulos HM. Sjogren syndrome. *CMAJ*. 2014;186:E579-E586.
8. Theander E, Vasaitis L, Baecklund E, et al. Lymphoid organisation in labial salivary gland biopsies is a possible predictor for the development of malignant lymphoma in primary Sjögren syndrome. *Ann Rheum Dis*. 2011;70:1363-1368.
9. Gottenberg JE, Seror R, Miceli-Richard C, et al. Serum levels of beta2-microglobulin and free light chains of immunoglobulins are associated with systemic disease activity in primary Sjögren syndrome. Data at enrollment in the prospective ASSESS cohort. *PLoS ONE*. 2013;8:e59868.
10. Nocturne G, Boudaoud S, Miceli-Richard C, et al. Germline and somatic genetic variations of TNFAIP3 in lymphoma complicating primary Sjogren's syndrome. *Blood*. 2013;122:4068-4076.
11. Shiboski SC, Shiboski CH, Criswell L, et al. American College of Rheumatology classification criteria for Sjögren syndrome: a data-driven, expert consensus approach in the Sjögren International Collaborative Clinical Alliance cohort. *Arthritis Care Res (Hoboken)*. 2012;64:475-487.
12. Seror R, Ravaud P, Mariette X, et al. EULAR Sjogren's Syndrome Patient Reported Index (ESSPRI): development of a consensus patient index for primary Sjogren's syndrome. *Ann Rheum Dis*. 2011;70:968-972.
13. Seror R, Ravaud P, Bowman SJ, et al. EULAR Sjogren's syndrome disease activity index: development of a consensus systemic disease activity index for primary Sjogren's syndrome. *Ann Rheum Dis*. 2010;69:1103-1109.
14. Gottenberg JE, Cinquetti G, Larroche C, et al. Efficacy of rituximab in systemic manifestations of primary Sjogren's syndrome: results in 78 patients of the AutoImmune and Rituximab registry. *Ann Rheum Dis*. 2013;72:1026-1031.
15. Mariette X, Seror R, Quartuccio L, et al. Efficacy and safety of belimumab in primary Sjögren syndrome: results of the BELISS open-label phase II study. *Ann Rheum Dis*. 2015;74:526-531.

REVIEW QUESTIONS

1. A 59-year-old woman is referred to your office because of dry mouth, grittiness of eyes, and a rash on both legs. Physical exam discloses unilateral parotid enlargement and a purpuric rash of the lower extremities. Past medical history is unremarkable to date. Among the following diagnostic tests, which one is the most useful to establish the diagnosis of Sjögren syndrome?

A. Complement levels
B. Cryoglobulins
C. Serum protein electrophoresis
D. Salivary flow
E. Anti-Ro/SSA antibodies

Answer: E Anti-Ro/SSA positivity is one of the two required 2002 European/American classification criteria for Sjögren syndrome (SS) and one out of the three 2012 preliminary ACR/SICCA 2012 criteria. Low complement levels, especially C4, and cryoglobulinemia have been designated as adverse prognostic factors for lymphoma development and mortality among SS patients. However, they are not included in the classification criteria. Hypergammaglobulinemia or monoclonal gammopathy are likewise not included in the classification criteria of SS. Salivary flow rate of less than 1.5 mL/15 minutes is an objective criterion of oral dryness, included in the 2002 European/American classification SS criteria but not in the 2012 preliminary ACR/SICCA criteria. It is not as specific as anti-Ro/SSA antibodies.

2. A 55-year-old woman presents with mild pain affecting her hands, as well as ocular and mouth dryness for 10 years. Her examination reveals positive ocular staining for both eyes, revealing keratoconjunctivitis sicca, positive anti-Ro/SSA antibodies, and negative HCV and HIV serology. Additional testing included complete blood count, revealing lymphopenia with normal hemoglobin levels and platelet counts. Mixed monoclonal cryoglobulinemia and low C4 levels were also detected. Which of the following statements is true?

A. The patient fulfills both the the 2002 American/European Consensus Group (AECG) criteria and 2012 preliminary ACR/SICCA criteria for primary SS (pSS).
B. The patient fulfills only the 2002 AECG criteria for pSS.
C. A minor salivary gland biopsy is mandatory to classify the patient as having SS according to 2002 AECG group criteria for pSS.
D. A minor salivary gland biopsy is mandatory to classify the patient as having SS according to 2012 preliminary ACR/SICCA criteria for pSS.
E. The patient suffers from sicca asthenia polyalgia syndrome.

Answer: A Two out of three 2012 preliminary ACR/SICCA criteria are fulfilled, as well as four out of six 2002 AECG criteria. Since anti-Ro/SSA antibodies are present, no histopathologic confirmation is required for the classification of pSS according to 2002 AECG criteria in the presence of subjective and objective symptoms and signs of salivary and lacrimal gland involvement. Sicca asthenia polyalgia syndrome is a diagnosis of exclusion and refers to the presence of fibromyalgia-like features in association with sicca symptomatology.

3. Which of the following is the best option for improving the patient's ocular dryness?

A. Methotrexate
B. Pilocarpine
C. Hydroxychloroquine
D. Ocular cyclosporine drops
E. Infliximab

Answer: D Ocular cyclosporine drops have proved useful in the treatment of dry eye. The efficacy of methotrexate on ocular dryness has never been demonstrated. Ocular dryness is less frequently improved than salivary dryness with secretagogues (i.e., pilocarpine). The efficacy of hydroxychloroquine on dryness has not been demonstrated. No efficacy of infliximab has been demonstrated in pSS.

4. Which of the following features is ***not*** considered an adverse predictor for lymphoma development in Sjögren syndrome?

A. Presence of splenomegaly
B. Presence of purpura
C. Salivary flow rate less than 1 mL/15 minutes
D. Presence of persistent parotid gland enlargement
E. Presence of CD4 lymphopenia

Answer: C All the other items are predictive factors of lymphoma. An association has not been reported between salivary flow rate and risk of lymphoma development.

5. A 55-year-old woman is referred to your department for polysynovitis and palpable purpura. She fulfills the AECG 2002 criteria for Sjögren syndrome. Which of the following treatments would you recommend to treat these systemic manifestations?

A. Infliximab
B. Intravenous immunoglobulin
C. Pilocarpine
D. Rituximab
E. Cyclophosphamide

Answer: D Three randomized controlled trials (RCTs) and one registry study suggested possible efficacy of rituximab in pSS, at least in some systemic manifestations. Two RCTs did not demonstrate any efficacy of tumor necrosis factor blockers (i.e., infliximab) in pSS. IVIG has been proposed in open studies in the treatment of SS-associated neuropathies without vasculitis, but never has demonstrated any efficacy in other systemic manifestations. The secretagogue pilocarpine has no effect on systemic manifestations. The systemic manifestation presented by the patient is not severe enough to propose cyclophosphamide.

269

INFLAMMATORY MYOPATHIES

STEVEN A. GREENBERG

OVERVIEW

The inflammatory myopathies are a heterogeneous group of acquired disorders in which the immune system is thought to play a major pathogenic role. Though some genetic disorders affecting muscle also have significant involvement of the immune system and are treated with immunosuppressive therapy as standard of care (e.g., treatment of Duchenne's muscular dystrophy with corticosteroids), these genetic disorders are not classified as inflammatory myopathies. The four major subtypes of inflammatory myopathy are: dermatomyositis (DM), polymyositis (PM), immune-mediated necrotizing myopathy (IMNM), and inclusion body myositis (IBM; also called sporadic inclusion body myositis [sIBM]). These disorders have distinct clinical and pathologic features and pathophysiologies (Table 269-1). Whereas DM and PM have been described in the medical literature for over 100 years, IMNM and IBM have only become defined as syndromes distinct from PM within the last few decades.

EPIDEMIOLOGY

The prevalence of DM has been estimated at 100[1] to 210 per million. The estimated prevalence of PM is confounded by frequent misdiagnosis of IBM and muscular dystrophies as PM. Traditionally, PM has been considered more prevalent (70 per million[2]), but comparative studies with attention to IBM have found a prevalence of PM of 35 per million, approximately half the prevalence of IBM of 70 per million. The prevalence of IMNM is unknown.

DM peaks in prevalence in childhood (7 to 15 years) and in midlife (30 to 50 years), whereas PM peaks in prevalence in midlife. IBM is rarely diagnosed before the age of 40 and is most common after the age of 50. DM and PM have female predominance; IBM has male predominance. Ethnicity and worldwide distribution influence the development of various inflammatory myopathies.

PATHOBIOLOGY

The pathophysiologies of various forms of inflammatory myopathy are poorly understood. These disorders do share in common injury to muscle by the immune system. Much of the theory of pathophysiology of these disorders comes from microscopic examination of muscle biopsies and the distinct pathologies of these disorders (Fig. 269-1).

The muscle pathology of DM involves loss of muscle blood vessels and injury to myofibers at the edges of muscle fascicles (i.e., perifascicular atrophy; see Fig. 269-1). The relationship of these two features to each other is uncertain but has been postulated to be due to a primary injury to muscle capillaries, followed by ischemic injury to myofibers. An alternative view is that a common factor injures both myofibers and capillaries.[3] Skin pathology shows features analogous to that of muscle, with an interface dermatitis consisting of injury to the basal layer of keratinocytes.

Much evidence points toward DM as mediated by the type 1 interferon cytokine family, consisting mainly of interferon (IFN)-α and IFN-β.[4,5] Numerous studies of DM skin and muscle samples show marked upregulation of type 1 IFN-inducible transcripts and proteins uniquely in DM among muscle diseases, and similarly to systemic lupus erythematosus among skin diseases. The presence of autoantibodies in some patients with DM, such as antibodies to the type 1 IFN-inducible protein MDA5, is of uncertain significance but seems likely due to an immune reaction to proteins that are not normally expressed at high levels or exposed to the immune system. The paraneoplastic associations of DM suggest that in such patients, an immune reaction against an underlying malignancy results in bystander injury to muscle and skin.

Because PM is a diverse group of disorders, the mechanisms involved are likely to be varied. Pathologically, there is an appearance of invasion of muscle fibers by adaptive immune system cells (T cells) that appears to be antigen driven, so that cytotoxic T cell–mediated autoimmunity directed against an unknown target has been a favored hypothesis. The antigens targeted by this process and the fundamental cause are unknown.

TABLE 269-1 CLASSIFICATION OF INFLAMMATORY MYOPATHIES

DISORDER	AGE RANGE	CLINICAL FEATURES	MUSCLE PATHOLOGY
Dermatomyositis	Juvenile and adult forms	Proximal weakness plus skin	Perimysial and perivascular inflammation, perifascicular atrophy
Polymyositis	Adult (rare in childhood)	Proximal weakness	Endomysial inflammation with invasion of non-necrotic muscle fibers
Immune-mediated necrotizing myopathy	Adult	Proximal weakness	Multifocal necrotic muscle fibers
Inclusion body myositis	Adult > 40 years old	Prominent quadriceps and finger flexor weakness; treatment refractory	Endomysial inflammation with invasion of non-necrotic muscle fibers plus rimmed vacuoles
Overlap syndromes	Adult	Myositis plus defined connective tissue disease	Nonspecific inflammation
Other (granulomatous myositis, eosinophilic myositis)	All ages	Proximal or distal weakness	Specific to type (e.g., granulomas present with granulomatous myositis)

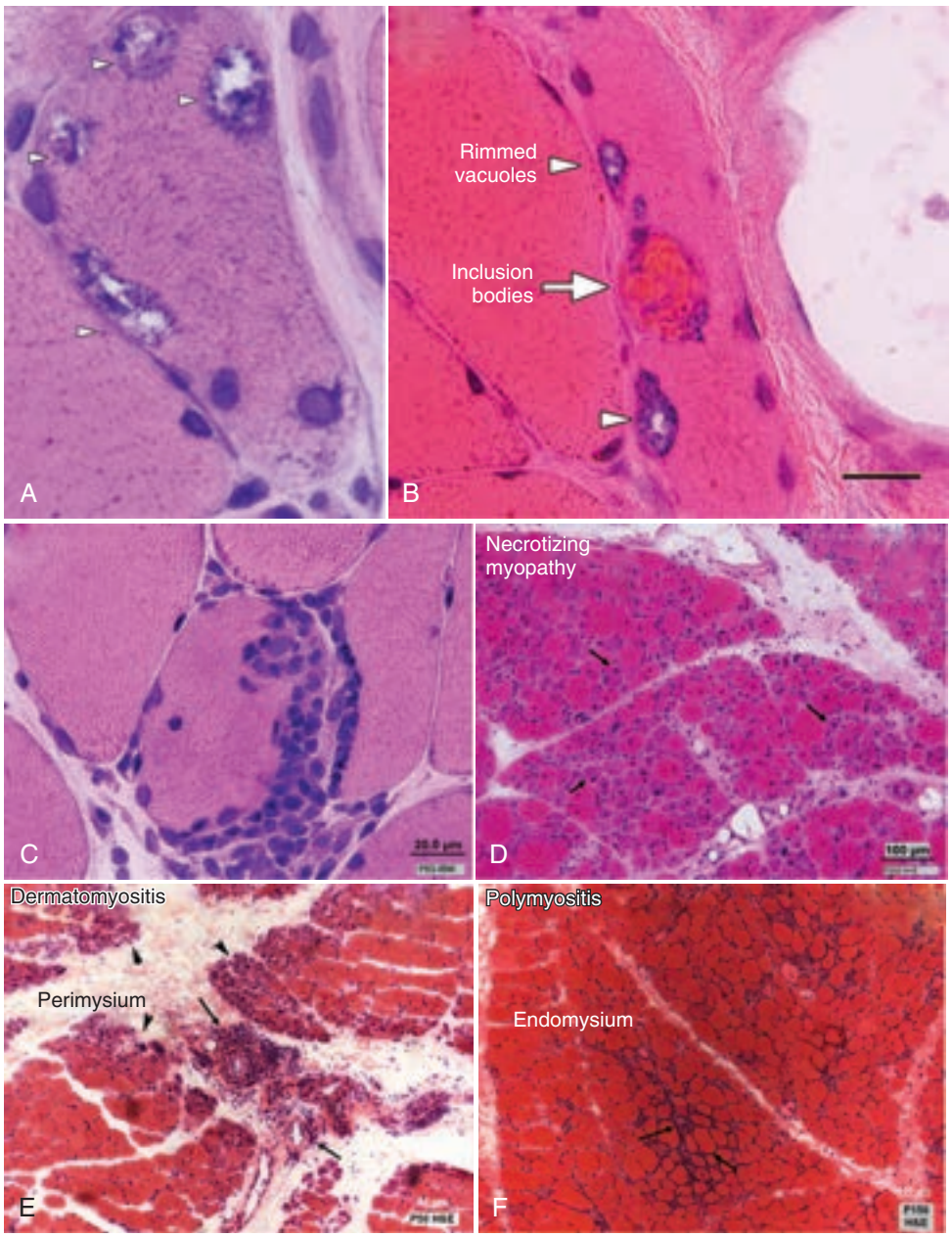

FIGURE 269-1. Pathologies of inflammatory myopathies. A and **B**, Rimmed vacuoles (*arrowheads*) of inclusion body myositis (IBM). **C**, Invasion of non-necrotic muscle fiber in IBM. **D**, Scattered necrotic and regenerating myofibers in immune-mediated necrotizing myopathy. **E**, Perivascular and perimysial inflammation (*arrows*), with perifascicular atrophy (*arrowheads*), in dermatomyositis. **F**, Endomysial inflammation in polymyositis. With permission from the Inclusion Body Myositis Foundation, Inc.

IMNM is also a poorly understood disorder.[6] It can also be paraneoplastic, suggesting cross-reactions by the immune system with the underlying malignancy and with muscle antigens. More commonly, IMNM occurs in association with treatment with statin drugs. The identification of autoantibodies against the target of statins, 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGCR), in the majority of patients who develop IMNM in association with statin use suggests that the upregulation of HMGCR in muscle is directly toxic to muscle and triggers an immune reaction against it.

The pathogenesis of IBM is complex. Two dual pathologies have been noted: degeneration of myofibers and of myonuclei in particular, evident as formation of rimmed vacuoles (see Fig. 269-1A and B), and involvement of the immune system.[7] The accumulation of more than 75 different proteins into sarcoplasmic aggregates in a small percentage of IBM myofibers has been reported, and has given rise to a number of molecular toxicity hypotheses in which certain specific protein aggregates are theorized as injurious to myofibers.

The immune system involvement in IBM is notable in that whereas most other forms of inflammatory myopathy are generally responsive to immunomodulatory treatments, IBM is refractory to treatment. This is particularly remarkable in that IBM has the greatest evidence of all the inflammatory myopathies of a highly refined antigen-driven adaptive immune system involvement. Pathology shows very chronic and often marked but variable inflammatory infiltrates of T cells, myeloid dendritic cells, and plasma cells in muscle. Studies of the T-cell receptors have strongly suggested that T-cell autoimmunity is driven by one or more specific antigens, though the identity of any of these antigens is unknown.

Studies of a B-cell pathway in IBM have led to identification of an autoantibody that is highly specific to IBM among muscle diseases. Circulating autoantibodies against a 43-kD muscle protein were reported in 2011, and the identity of this 43-kD protein as cytoplasmic 5′ nucleotidase 1A (cN1A; NT5C1A) was reported in 2013.[8,9] cN1A is a nucleotidase that is most abundant in skeletal muscle and involved in the metabolism of nucleic acids. Serum anti-cN1A autoantibodies are present in 50 to 70% of patients with IBM, depending on which assays and what cutoffs are used, and highly specific to IBM (>90 to 95%) among muscle diseases. The role of blood testing for anti-cN1A autoantibodies in the diagnosis and management of patients with suspected IBM is currently being defined, potentially shortening the time to diagnosis, reducing the misdiagnosis rate, and avoiding more invasive muscle biopsy in some patients.

TABLE 269-2 CLINICAL DIAGNOSTIC CRITERIA FOR INFLAMMATORY MYOPATHIES

DISORDER	DIAGNOSIS
Dermatomyositis	1. Diagnostic skin involvement (heliotrope rash, Gottron's papules) *OR* diagnostic muscle biopsy finding of perifascicular atrophy *OR* 2. All of the following: • Suggestive skin involvement • Subacute or chronic proximal or distal weakness • Muscle biopsy showing perimysial or perivascular inflammation without features suggesting another disorder (e.g., endomysial inflammation, rimmed vacuoles) *OR* skin biopsy showing interface dermatitis along with clinical exclusion of lupus erythematosus
Polymyositis	All of the following: 1. Subacute or chronic proximal weakness 2. Elevated serum creatine kinase (CK) 3. Muscle biopsy showing invasion of endomysial inflammation 4. Response to immunotherapy *OR* appropriate consideration and exclusion of limb-girdle muscular dystrophies and inclusion body myositis
Immune-mediated necrotizing myopathy	Both of the following: 1. Subacute or chronic proximal weakness 2. Muscle biopsy showing necrotizing myopathy, with scattered necrotic or regenerating myofibers and a lack of inflammation other than macrophage invasion of necrotic muscle fiber
Inclusion body myositis	All of the following: 1. Adult > 40 years old 2. Finger flexion or quadriceps weakness 3. Muscle biopsy showing endomysial inflammation *OR* the presence of serum anti-cN1A autoantibodies 4. Muscle biopsy showing rimmed vacuoles *OR* invasion of non-necrotic muscle fibers *OR* the presence of serum anti-cN1A autoantibodies

cNIA = cytoplasmic 5′ nucleotidase 1A.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

A diagnosis of inflammatory myopathy is considered when a patient presents with proximal or distal weakness without sensory symptoms, or in patients with the characteristic skin lesions of DM. Less frequently, asymptomatic elevated creatine kinase (CK) levels lead to a diagnosis of inflammatory myopathy. Most patients with DM, PM, or IMNM present with subacute proximal weakness of the arms and legs progressing over months, though these diseases may present acutely. Patients with IBM present later in life, usually symptomatic from slowly progressive weakness of knee extensors and finger flexors. More specific diagnostic considerations for these disorders are considered individually (Table 269-2). Most patients undergo muscle biopsy, or skin biopsy in the case of suspected DM, as part of the diagnostic evaluation.

Dermatomyositis

Patients with DM typically present with characteristic skin lesions or muscle weakness. Virtually pathognomonic skin features are a heliotrope rash, a violaceous periorbital macular erythema, sometimes with edema, and Gottron's papules, violaceous papules over dorsal metacarpophalangeal and interphalangeal joints of the hands (Fig. 269-2).[10] Periungual telangiectasias and thrombosed capillaries, poikiloderma over photoexposed areas such as the upper back ("shawl sign"), non-scarring alopecia, and subcutaneous calcification are other suggestive signs. Prominent pruritus is also a common feature of DM. Muscle weakness in DM is less specific, occurring in a pattern indistinguishable from many other muscle diseases.

Useful laboratory studies for the evaluation of suspected DM include serum CK (though CK can be normal or even below typical laboratory lower limits of normal in patients with highly active disease) and DM-associated autoantibody studies (e.g., anti-Jo-1, anti-Mi2, and anti-MDA5). Occasional patients have abnormal serum aldolase but normal serum CK. Skin biopsy showing a cell-poor interface dermatitis supports the diagnosis of DM. Muscle biopsy showing perimysial and perivascular inflammation also supports a diagnosis of DM, whereas the presence of perifascicular atrophy in a muscle biopsy is pathognomonic for DM. Because DM is associated with malignancy, appropriate laboratory and radiologic studies should be performed to search for underlying malignancy in all newly diagnosed patients. The most common DM-associated malignancies tend to reflect the overall age and gender cancer rates within the individual patient's population (i.e., breast, lung, and colorectal cancer in Western countries; nasopharyngeal cancer in Asian populations). This observation supports the notion of DM as a paraneoplastic process that can develop in virtually any kind of cancer.

Research diagnostic criteria for DM have been defined.[11] Clinical diagnostic criteria are outlined in Table 269-2. The clinical features of muscle weakness in DM are entirely nonspecific, with no particular pattern indicative of DM rather than other muscle diseases. In practice, certain dermatologic clinical findings (heliotrope rash, Gottron's papules) or muscle biopsy findings (perifascicular atrophy) are considered nearly pathognomonic for DM.

Polymyositis

The diagnosis of PM is often problematic, with historically many patients with genetically defined limb-girdle muscular dystrophies and IBM being misdiagnosed as PM. The 1975 criteria for PM that are frequently cited allow for a diagnosis of "definite" PM without a muscle biopsy. Other research criteria require muscle biopsy. In clinical practice, the core criteria for the diagnosis of PM are subacute proximal weakness, elevated serum CK, and muscle biopsy showing endomysial inflammation without features suggestive of another diagnosis such as IBM (see Table 269-2). Patients with defined connective tissue disorders such as Sjögren's syndrome or mixed connective tissue disease have "overlap syndromes," often also classified as PM. Patients with IMNM have historically been classified as PM but are increasingly classified separately. Patients with IBM are frequently misdiagnosed as PM because of a lack of appreciation of characteristic IBM finger flexor weakness, and because muscle biopsies show endomysial inflammation. The presence of autoantibodies such as anti-Jo-1 argue more for PM than IBM, though these may be seen in DM as well.

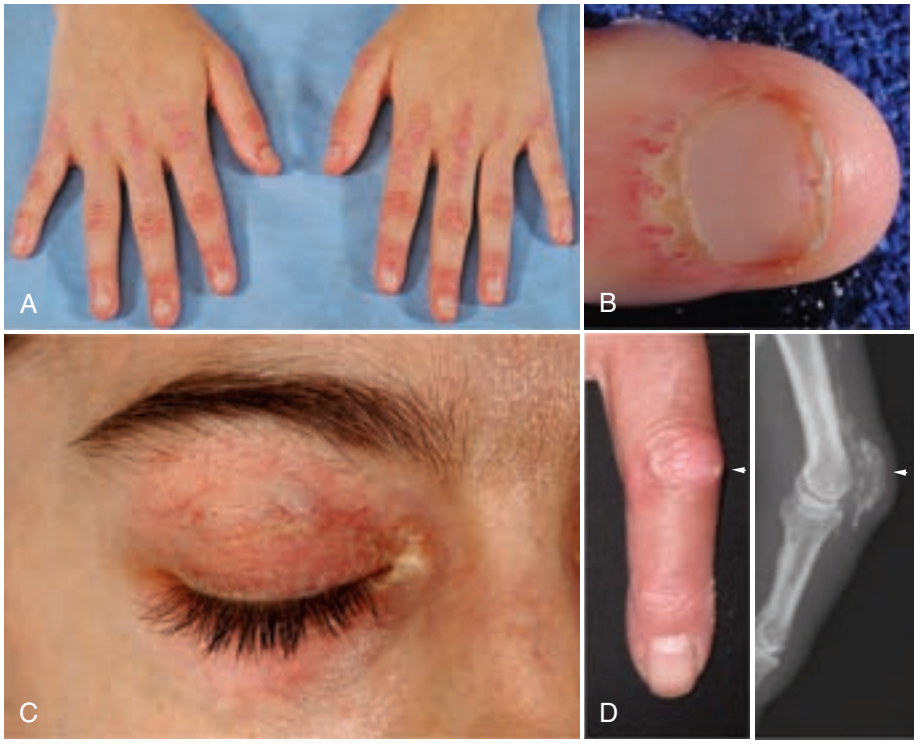

FIGURE 269-2. Clinical findings in dermatomyositis. A, Erythematous to violaceous raised papules overlying the metacarpal and interphalangeal joints, known as Gottron's papules. These are considered the hallmark finding in dermatomyositis. **B,** Cuticular overgrowth and periungual capillary changes, which include dilated and tortuous blood vessels with areas of atrophy, telangiectasia, vessel dropout, and bushy loop formation along the fingernail bed. **C,** Erythema and minimal edema involving the upper eyelids, with occasional telangiectasia, known as the heliotrope rash. **D,** Subcutaneous calcification erupting through skin (*arrowhead*), seen clinically and by x-ray.

Immune-Mediated Necrotizing Myopathy

IMNM has increasingly been separated from the PM category. Acute or subacute proximal weakness indistinguishable from that of PM or DM and an elevated CK are nonspecific, but muscle biopsy showing scattered necrotic or regenerating myofibers without inflammation other than macrophages invading these necrotic myofibers is typical of IMNM. The presence of anti-HMGCR (3-hydroxy-3-methylglutaryl-coenzyme A reductase) or anti-SRP (signal recognition particle) antibodies both suggest IMNM. IMNM, particularly when associated with anti-SRP antibodies, may be paraneoplastic, and laboratory and radiologic evaluation for malignancy should be considered.

Inclusion Body Myositis

IBM has a clinical presentation distinct from other inflammatory myopathies.[12] IBM weakness is always slowly progressive rather than the acute or subacute weakness more typically seen in other forms of inflammatory myopathies. Clinical diagnostic criteria are shown in Table 269-2. IBM has a high misdiagnosis rate, estimated at approximately 50% of patients. Symptoms of IBM rarely are present before the age of 40 years and most commonly occur after the age of 50. The distribution of weakness is usually in finger flexors or quadriceps rather than proximal arms (shoulder abduction) or proximal legs (hip flexion), more typical of PM or DM. IBM is a highly atrophying muscle disease, and loss of bulk in medial and lateral anterior thighs and ventral forearms is characteristic. Patients present with difficulty walking, buckling of knees, or weakness of grip. The diagnosis of IBM can be highly suspected in such patients of appropriate age and findings on examination of quadriceps atrophy and weakness of finger flexors, especially flexor digitorum profundi, responsible for flexion of distal fingertips. Examination of the strength in these distal fingertips, which needs to be done one finger at a time, is often the single most helpful approach to the diagnosis of IBM.

Serum CK is either normal or modestly elevated (typically < 5 times the upper limit of normal). A serum autoantibody, anti-cN1A (also called anti-NT5C1A), appears highly specific to IBM among muscle diseases and may be of diagnostic value. Most patients undergo muscle biopsy, with characteristic features being the presence of rimmed vacuoles seen on hematoxylin and eosin (H&E) and Gomori trichrome staining, along with endomysial inflammation or invasion of non-necrotic muscle fibers. Immunohistochemical stains detecting p62 or TDP-43 are of additional highly specific diagnostic value.

TREATMENT

Rx

Generally, most patients with DM, PM, and IMNM respond to immunomodulatory therapies, whereas patients with IBM are almost universally refractory. A general approach to treatment is shown in Figure 269-3.

Treatment of Dermatomyositis and Polymyositis

Most patients with DM and PM are treated with and respond to corticosteroids.[13,14] Dosing is typically prednisone at 1 mg/kg/day orally until significant improvement occurs (typically 1 to 3 months), followed by gradual taper of 10 mg/day/month. Second-line agents include methotrexate, azathioprine, cyclosporine, and intravenous immunoglobulin. Second-line agents are used for two reasons: they may have a better side-effect profile than chronic higher doses of corticosteroids, and they may be necessary for patients whose responses are insufficient to corticosteroids alone. An important decision is whether to start second-line agents concurrently with initial corticosteroid treatment or wait and see how low a dose of corticosteroids offers satisfactory control, and then add agents only if the corticosteroid dose cannot be lowered sufficiently. Thus, in the former approach, prednisone 60 mg/day and methotrexate 7.5 mg PO weekly might be started concurrently, and the methotrexate dose increased weekly to 15 to 20 mg PO weekly. Once improvement is substantial, the dose of prednisone may be tapered over 3 to 6 months. Stability on methotrexate alone would then be followed by gradual reduction in its dose.

For patients with severe initial presentations, the combination of corticosteroids and periodic intravenous immunoglobulin (1 g/kg every 2 weeks) may offer a better chance for more rapid improvement.

Approximately seven randomized placebo controlled trials have been reported to date in DM or PM. These studies have almost always used the Bohan and Peter criteria for the diagnosis, which may result in inclusion of patients with limb-girdle muscular dystrophies and IBM misdiagnosed as having PM. The largest trial, rituximab in myositis (RIM), enrolling 200 subjects, used a trial design in which all subjects received active drug, but comparator groups were treated "early" or "late" (8 weeks later), and efficacy was based on time to improvement. This study found no significant differences between these early and late treatment groups' time to improvement, as defined by a specific definition of improvement (DOI) used in that study.

Treatment of Inclusion Body Myositis

A number of controlled and uncontrolled trials of IBM have been published. None of these demonstrated efficacies of the therapeutic

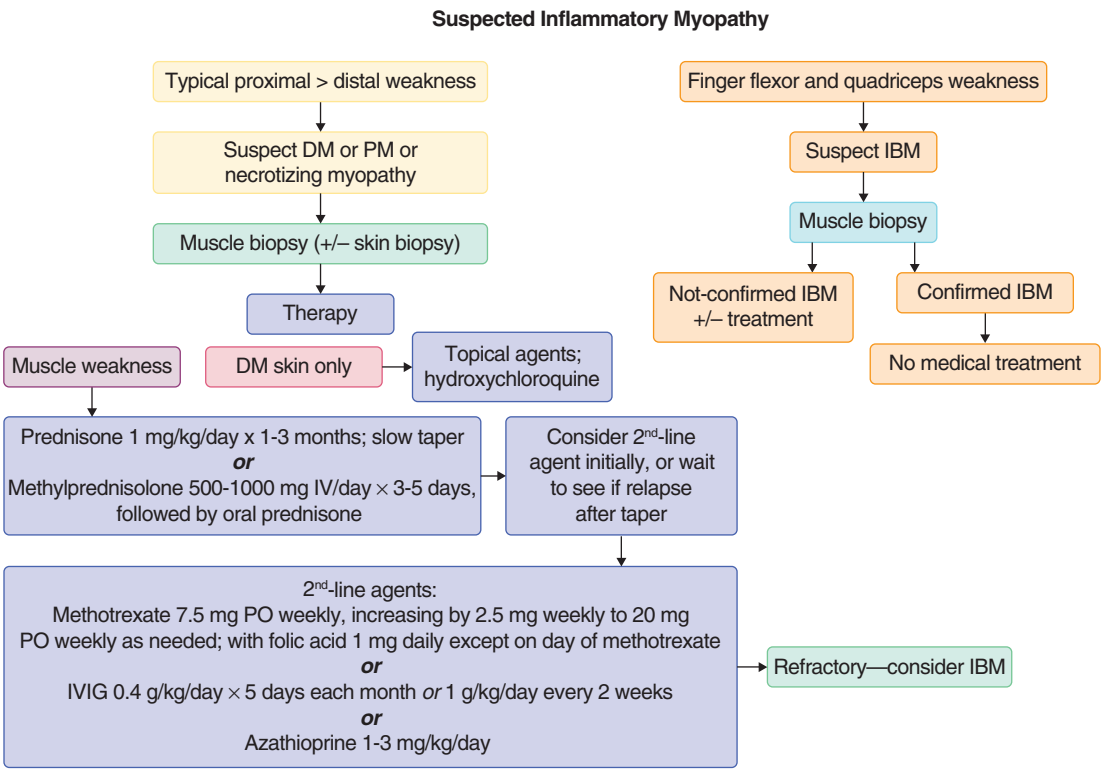

FIGURE 269-3. Approach to treatment of suspected inflammatory myopathy. DM = dermatomyositis; IBM = inclusion body myositis; IVIG = intravenous immunoglobulin; PM = polymyositis. With permission from the Inclusion Body Myositis Foundation, Inc.

interventions, which have included prednisone, intravenous immunoglobulin, methotrexate, antithymocyte globulin, oxandrolone, interferon-β, and alemtuzumab.[15] Current management of patients with IBM is supportive, involving avoidance of falls and the use of ankle supports and gait assistive devices. Tendon transfer to improve hand function has been used.

PROGNOSIS

Most patients with adult DM, PM, and statin-associated IMNM have a good prognosis but require long-standing immunomodulatory therapy. Many patients with juvenile DM may go into long-standing remission or cure with aggressive initial treatment. Patients with anti-SRP–associated IMNM may have severe and difficult-to-treat disease. Patients with IBM generally have a slowly progressive course, with one series showing a mean time to loss of ambulation of 12 years.

Grade A References

A1. Gordon PA, Winer JB, Hoogendijk JE, et al. Immunosuppressant and immunomodulatory treatment for dermatomyositis and polymyositis. *Cochrane Database Syst Rev*. 2012;8:CD003643.

A2. Oddis CV, Reed AM, Aggarwal R, et al. Rituximab in the treatment of refractory adult and juvenile dermatomyositis and adult polymyositis: a randomized, placebo-phase trial. *Arthritis Rheum*. 2013;65:314-324.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Rosa J, Garrot LF, Navarta DA, et al. Incidence and prevalence of polymyositis and dermatomyositis in a health management organization in Buenos Aires. *J Clin Rheumatol*. 2013;19:303-307.
2. Meyer A, Meyer N, Schaeffer M, et al. Incidence and prevalence of inflammatory myopathies: a systematic review. *Rheumatology (Oxford)*. 2015;54:50-63.
3. Greenberg SA. Sustained autoimmune mechanisms in dermatomyositis. *J Pathol*. 2014;233: 215-216.
4. Suárez-Calvet X, Gallardo E, Nogales-Gadea G, et al. Altered RIG-I/DDX58-mediated innate immunity in dermatomyositis. *J Pathol*. 2014;233:258-268.
5. Greenberg SA. Dermatomyositis and type 1 interferons. *Curr Rheum Rep*. 2010;12:198-203.
6. Mohassel P, Mammen AL. Statin-associated autoimmune myopathy and anti-HMGCR autoantibodies. *Muscle Nerve*. 2013;48:477-483.
7. Greenberg SA. Inclusion body myositis. *Curr Opin Rheumatol*. 2011;23:574-578.
8. Larman BH, Salajegheh M, Nazareno R, et al. Cytosolic 5′-nucleotidase 1A autoimmunity in sporadic inclusion body myositis. *Ann Neurol*. 2013;73:408-418.
9. Pluk H, van Hoeve BJ, van Dooren SH, et al. Autoantibodies to cytosolic 5′-nucleotidase 1A in inclusion body myositis. *Ann Neurol*. 2013;73:397-407.
10. Bailey EE, Fiorentino DF. Amyopathic dermatomyositis: definitions, diagnosis, and management. *Curr Rheumatol Rep*. 2014;16:465.
11. Hornung T, Wenzel J. Innate immune-response mechanisms in dermatomyositis: an update on pathogenesis, diagnosis and treatment. *Drugs*. 2014;74:981-998.
12. Dimachkie MM, Barohn RJ. Inclusion body myositis. *Neurol Clin*. 2014;32:629-646.
13. Lam C, Vleugels RA. Management of cutaneous dermatomyositis. *Dermatol Ther*. 2012;25: 112-134.
14. Lundberg I, Vencovsky J, Alexanderson H. Therapy of myositis: biological and physical. *Curr Opin Rheumatol*. 2014;26:704-711.
15. Greenberg SA. Pathogenesis and therapy of inclusion body myositis. *Curr Opin Neurol*. 2012;25: 630-639.

REVIEW QUESTIONS

1. A 65-year-old man developed slowly progressive difficulty arising from a chair and experienced buckling of the knees while walking, resulting in several falls. Muscle biopsy showed endomysial inflammation and rimmed vacuoles. The correct diagnosis is:

A. Polymyositis
B. Dermatomyositis
C. Inclusion body myositis
D. Immune-mediated necrotizing myopathy
E. Overlap syndrome

Answer: C Inclusion body myositis (IBM). IBM is a later-onset, slowly progressive disease presenting with prominent quadriceps or finger flexion weakness. In this case, the knee buckling indicates quadriceps weakness. Endomysial inflammation can be seen in either polymyositis or IBM, but rimmed vacuoles confirm IBM as the correct diagnosis.

2. A 35-year-old woman developed proximal weakness and a purplish papular rash over the dorsum of the hands. A skin biopsy might be expected to show what feature and lead to which diagnosis:

A. Lymphocytoclastic vasculitis AND polymyositis
B. Interface dermatitis AND dermatomyositis
C. Granulomas AND granulomatous myositis
D. Necrotic cells AND necrotizing myopathy
E. Necrotic cells AND dermatomyositis

Answer: B This patient has Gottron's papules (see Fig. 296-2A), virtually pathognomonic for dermatomyositis. Skin biopsy in dermatomyositis shows an interface dermatitis, with pathology of the basal layer of keratinocytes lying at the border (interface) between the epidermis and the dermis.

3. A 60-year-old man with proximal weakness and mildly elevated serum creatine kinase (CK) is diagnosed with polymyositis and treated with high-dose prednisone 80 mg/day for several months, before slow taper of prednisone to 30 mg/day by month 6. His serum CK improves, but no improvement in strength is present and he appears mildly weaker. Which statement is most likely true?

A. He has refractory polymyositis and needs more aggressive therapy.
B. He has developed steroid myopathy in addition to his polymyositis.
C. The diagnosis of polymyositis should be reconsidered, and he could undergo a second muscle biopsy or blood diagnostic testing for anti-cN1A autoantibodies as the next step.
D. The diagnosis of polymyositis should be reconsidered; he may have inclusion body myositis or a limb-girdle muscular dystrophy.
E. C and D

Answer: E Patients with IBM in particular, and sometimes limb-girdle muscular dystrophy, are often misdiagnosed as having polymyositis. Serum CK, modestly elevated in IBM, may lower with prednisone treatment, but patients rarely have improved strength.

4. Which syndrome can be paraneoplastic and should prompt thorough investigation for an underlying malignancy?

A. Dermatomyositis
B. Polymyositis
C. Inclusion body myositis
D. Granulomatous myositis
E. All of the above

Answer: A Dermatomyositis. A new diagnosis of dermatomyositis should prompt a thorough investigation for an underlying malignancy, with reported rates estimated at 15 to 23%.[12] The types of malignancy found in patients with dermatomyositis generally reflect those found in age-, sex-, and population-matched persons without dermatomyositis. Therefore, breast, lung, and colorectal cancer are the three most common cancers from Western country cohorts, while nasopharyngeal carcinoma is the most common dermatomyositis-associated cancer in Asian studies. This finding further supports the role of dermatomyositis as a paraneoplastic process that can occur with virtually any kind of cancer. (Ungprasert P, Bethina NK, Jones CH. Malignancy and idiopathic inflammatory myopathies. *N Am J Med Sci.* 2013;5:569-572.)

5. Normal serum CK in a patient with proximal muscle weakness excludes the diagnosis of:

A. Dermatomyositis
B. Inclusion body myositis
C. Muscular dystrophy
D. Granulomatous myositis
E. None of the above

Answer: E None of the above. Serum CK is a helpful indicator of muscle disease when elevated, but normal CK does not exclude muscle disease. In particular, patients with very active dermatomyositis frequently have normal serum CK.

270 THE SYSTEMIC VASCULITIDES

JOHN H. STONE

DEFINITION

The vasculitides are a heterogeneous group of disorders linked by the common finding of destructive inflammation within blood vessel walls. The most current nomenclature scheme[1] identifies at least 27 different forms of primary vasculitis (Table 270-1). The major forms of vasculitis are discussed in this chapter.

CLASSIFICATION

Classification by Vessel Size

The etiology of most forms of vasculitis remains unknown, and major gaps exist in our understanding of the pathophysiologic processes. The most valid basis for classification of the vasculitides is the size of the predominant blood vessels involved. The vasculitides are categorized initially by whether the vessels affected are primarily large, medium, or small (Table 270-2). Large vessels are considered the aorta, its primary branches, and any vessel that is not located within an organ such as a muscle, kidney, nerve, or the skin. Medium-sized vessels, in contrast, consist of the main visceral arteries and their branches. (Thus, the renal artery is considered a large vessel, but its intrarenal branches—the interlobar and arcuate arteries—are medium-sized vessels). Finally, small vessels include smaller intraparenchymal arteries as well as arterioles, capillaries, and veins.

Medium-vessel vasculitis and even large-vessel vasculitis can also affect small arteries. However, Large-vessel vasculitis affects large arteries more often than medium or small-vessel vasculitis, medium-vessel vasculitis affects predominantly medium arteries, and small-vessel vasculitis affects predominantly small arteries and other small vessels.

Additional Considerations in Classification

Several considerations other than blood vessel size are relevant to the classification of vasculitis (see Table 270-2). These are (1) age, sex, and ethnic

background of the patient; (2) tropism for particular organs; (3) presence or absence of granulomatous inflammation; (4) participation of immune complexes in the pathophysiologic process; and (5) detection of characteristic autoantibodies in the patients' serum, such as antineutrophil cytoplasmic antibodies (ANCAs).

Age, sex differences, and ethnic variation are discussed later in the section on Epidemiology. The organ tropisms of these disorders are illustrated by the following examples. Whereas immunoglobulin (Ig)A vasculitis (IgAV, also known as Henoch-Schönlein purpura) typically affects the skin, joints, kidneys, and gastrointestinal (GI) tract, granulomatosis with polyangiitis (GPA; formerly Wegener granulomatosis) classically involves the upper airways, lungs, and kidneys. In contrast to both IgAV and GPA, Cogan syndrome involves the eyes, the audiovestibular apparatus of the inner ear, and (in 10 to 15% of cases) the large arteries.

The presence or absence of granulomatous inflammation is a crucial element of vasculitis diagnosis and classification. Granulomatous inflammation implicates a small number of vasculitides that bear this hallmark, including GPA, giant cell arteritis, Takayasu arteritis, and eosinophilic granulomatosis with polyangiitis (EGPA; Churg-Strauss syndrome).

Immune complexes are essential to the pathophysiologic mechanism of some forms of small- and medium-vessel vasculitis. Complexes of IgA1, for example, are found in IgAV. Immune complexes consisting of IgG, IgM, complement components, and the hepatitis C virion characterize most cases of mixed cryoglobulinemia. In contrast, "pauci-immune" types of small- and medium-vessel vasculitis, such as GPA and microscopic polyangiitis, have little immunoglobulin or complement deposition within diseased tissues. Many but not all patients with pauci-immune forms of vasculitis are ANCA positive.

TABLE 270-1 NAMES FOR VASCULITIDES ADOPTED BY THE 2012 INTERNATIONAL CHAPEL HILL CONSENSUS CONFERENCE ON THE NOMENCLATURE OF VASCULITIDES.[1]

LARGE-VESSEL VASCULITIS
Takayasu arteritis
Giant cell arteritis
MEDIUM-VESSEL VASCULITIS
Polyarteritis nodosa
Kawasaki disease
Buerger disease*
SMALL-VESSEL VASCULITIS
Antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis
Microscopic polyangiitis
Granulomatosis with polyangiitis (formerly Wegener granulomatosis)
Eosinophilic granulomatosis with polyangiitis (Churg-Strauss syndrome)
Immune complex small-vessel vasculitis
Antiglomerular basement membrane disease
Cryoglobulinemic vasculitis
Immunoglobulin (Ig)A vasculitis (Henoch-Schönlein purpura)
Hypocomplementemic urticarial vasculitis
VARIABLE-VESSEL VASCULITIS
Behçet syndrome
Cogan syndrome
SINGLE-ORGAN VASCULITIS
Cutaneous leukocytoclastic angiitis
Cutaneous arteritis
Primary central nervous system vasculitis
Isolated aortitis
VASCULITIS ASSOCIATED WITH SYSTEMIC DISEASE
Lupus vasculitis
Rheumatoid vasculitis
Sarcoid vasculitis
Others (e.g., IgG4-related aortitis)
VASCULITIS ASSOCIATED WITH PROBABLE ETIOLOGY
Hepatitis C virus–associated cryoglobulinemic vasculitis
Hepatitis B virus–associated vasculitis
Syphilis-associated aortitis
Drug-associated immune complex vasculitis
Drug-associated ANCA-associated vasculitis
Cancer-associated vasculitis
Others

*Buerger disease (thromboangiitis obliterans) is not always considered to be a primary form of vasculitis and was not included in this consensus statement on nomenclature.[1]

EPIDEMIOLOGY

The epidemiologic features of individual forms of systemic vasculitis vary tremendously by geography (Table 270-3). This may reflect genetic influences, variation in environmental exposures, and other unknown disease risk factors. For example, whereas Behçet syndrome is rare in North Americans, affecting only 1 person in approximately 300,000, this condition is several hundred times more common among inhabitants of countries bordering the ancient Silk Route. Similarly, although Takayasu arteritis is rare in the United States—on the order of 3 new cases per million people per year—this disease is reportedly the most common cause of renal artery stenosis in India, where the incidence may be as high as 200 to 300 per million per year.

Age is an important consideration in the epidemiology of vasculitis. Eighty percent of patients with Kawasaki disease are younger than 5 years. In contrast, giant cell arteritis virtually never occurs in patients younger than 50 years, and the mean age of patients with this disease is 72. Age may also have an impact on disease severity and outcome. In IgAV, the overwhelming majority of cases in children (who represent 90% of all cases) have self-limited courses, resolving within several weeks. In adults, however, IgAV has a higher likelihood of chronicity and a poor renal outcome.

The distribution of sex varies across many forms of vasculitis. Buerger disease is the only form of vasculitis with a striking male predominance. The greater prevalence of smoking among males in most societies probably explains this predilection. In contrast, Takayasu arteritis has an overwhelming

TABLE 270-2 CONSIDERATIONS IN THE CLASSIFICATION OF SYSTEMIC VASCULITIS

Size of predominant blood vessels affected
Epidemiologic features:
Age
Sex
Ethnic background
Pattern of organ involvement
Pathologic features:
Granulomatous inflammation
Immune complex deposition vs. "pauci-immune" histopathology
Presence of ANCA in serum

ANCA = antineutrophil cytoplasmic antibody.

TABLE 270-3 EPIDEMIOLOGY OF SELECTED VASCULITIDES

DISEASE	UNITED STATES	ELSEWHERE	AGE, SEX, AND ETHNIC PREDISPOSITIONS
Giant cell arteritis	Incidence: 240/million (Olmsted County, MN)	220-270/million (Scandinavian countries)	Age > 50 yr, mean age 72 yr; females 3 : 1; northern European ancestry
Takayasu arteritis	Incidence: 3/million	200-300/million (India)	Age < 40 yr; females 9 : 1; Asian
Behçet syndrome	Prevalence: 3/million	3000/million (Turkey)	Silk Route countries
Polyarteritis nodosa	Incidence: 7/million	7/million (Spain)	Slight male predominance
Kawasaki disease	Incidence: 100/million*	900/million (Japan)	Children of Asian ancestry
Wegener granulomatosis	Incidence: 4/million	8.5/million (United Kingdom)	Whites ≫ blacks

*Among children younger than 5 years. From Gonzalez-Gay MA, Garica-Porrua C. Epidemiology of the vasculitides. *Rheum Clin North Am.* 2001;27:729-749.

tendency to occur in females (a 9 : 1 female-to-male ratio). The pauci-immune forms of vasculitis, such as GPA, EGPA, and microscopic polyangiitis, occur in males and females with approximately equal frequencies, but the phenotypic expression of these conditions may be affected by both age and sex.

The strongest link between any single gene and vasculitis is the association of HLA-B51 with Behçet syndrome. In Behçet syndrome, 80% of Asian patients have the HLA-B51 gene. The prevalence of HLA-B51 is significantly higher among patients with Behçet syndrome in Japan than among non-disease control subjects (55% versus < 15%). Among the sporadic cases of Behçet syndrome involving whites in the United States, however, HLA-B51 occurs in fewer than 15% of cases.

With the exception of Buerger disease and smoking, no definitive associations have been confirmed between disease and environmental or occupational exposures. Associations have been reported but not confirmed between exposures to silica and some types of pauci-immune vasculitis. Studies of potential associations between exposures of any type and vasculitis, however, are complicated frequently by difficulties in obtaining reliable measurements of the levels of the relevant exposure, the likelihood of recall bias among patients who are diagnosed with vasculitis, and the choice of appropriate control groups.

PATHOBIOLOGY

Table 270-4 illustrates the pathologic characteristics of selected forms of vasculitis. Specific pathologic features are discussed in the subsections on each disease. The type of inflammatory cell infiltrate in vasculitis is independent of the size of blood vessels involved. Mixed cell infiltrates in vasculitis are the rule rather than the exception, and histopathologic patterns of vasculitis may include leukocytoclasis (degranulation and destruction of neutrophils within blood vessel walls), granulomatous findings (with or without giant cells), lymphoplasmacytic infiltrates, varying degrees of eosinophilic infiltration, necrosis, and combinations of all these findings.

PATHOPHYSIOLOGY

Some pathophysiologic mechanisms are common to many different forms of vasculitis, regardless of the size of the predominant blood vessels involved. Immune complex deposition, for example, is present in several types of vasculitis that involve both medium-sized and small blood vessels. In this section, the general concepts related to the pathogenesis of large-vessel vasculitides are discussed separately from those of medium- and small-vessel vasculitides.

Large-Vessel Vasculitides

The pathologic process in large-vessel vasculitis appears to begin in the adventitia. In both Takayasu arteritis and giant cell arteritis, abundant numbers of activated T lymphocytes are found within inflamed arterial walls, centering on the adventitia. In Takayasu arteritis, most of these T cells appear to be of the $CD8^+$ subtype. Current evidence suggests that the cytotoxic functions of these cells, mediated by perforin and granzyme B, contribute to smooth muscle cell damage in this disease. $CD4^+$ T-cell responses in Takayasu arteritis have not been well defined.

In giant cell arteritis (Chapter 271), much evidence now suggests an antigen-driven disease, with the site of immunologic recognition events being the adventitia. $CD4^+$ T cells that secrete interferon (IFN)-γ appear to be recruited to the adventitia by a specific antigen(s), the identity of which remains unknown. Both the T cells that orchestrate the transmural inflammation and the inciting antigens are theorized to reach the adventitia through the vasa vasorum. Subsequently, T-cell signals from the adventitia stimulate macrophages and multinucleated giant cells to elaborate an array of downstream mediators, including metalloproteinases and platelet-derived growth factor. Interleukin (IL)-6, known to be a crucial cytokine in giant cell arteritis and probably Takayasu arteritis as well, is produced by macrophages residing in the blood vessel wall. The results of this inflammatory cascade are granulomatous inflammation, destruction of the internal elastic lamina, arterial wall hyperplasia, smooth muscle cell proliferation, intimal thickening, vascular occlusion, and in some cases, weakening of the vessel wall, leading to dilation and aneurysm formation.

Medium- and Small-Vessel Vasculitides

Several different pathophysiologic mechanisms are operative among the medium- and small-vessel vasculitides. In many cases, the mechanisms outlined in the following sections overlap.

Immune Complex–Mediated Vascular Injury

Immune complex–mediated tissue injury does not produce a single clinical syndrome but rather applies to many forms of vasculitis and overlaps with injuries caused by other immune mechanisms. Numerous variables influence immune complex–mediated injury, including the physical properties of the immune complexes (e.g., their size), the ability of the immune complexes to activate complement, the antigen-to-antibody ratio, and the hemodynamic features of specific vascular beds. Immune complexes participate in the pathophysiologic process of some forms of both medium- and small-vessel vasculitis, including polyarteritis nodosa, cryoglobulinemia, IgAV, cutaneous leukocytoclastic angiitis, and rheumatoid vasculitis.

Role of Antineutrophil Cytoplasmic Antibodies

ANCAs are directed against antigens that reside within the primary granules of neutrophils and monocytes. Two types of ANCA are relevant to vasculitis: (1) those directed against proteinase 3 (PR3), known as PR3-ANCA; and (2) those directed against myeloperoxidase (MPO), termed MPO-ANCA. ANCA interact with cytokines, neutrophils, monocytes, and other elements of the immune system to amplify ongoing inflammation in certain forms of vasculitis. A striking and still unexplained feature of ANCA-associated vasculitis (AAV) is that patients with primary forms of these conditions virtually never have antibodies to both PR3 and MPO. Despite the specificity of these antibodies, however, evidence for a primary role of ANCA in the etiology of human disease is still absent.

In GPA, abnormal cytokine regulation interacts with the production of ANCA to fuel the inflammatory response. T_H1 cytokines such as IFN-γ, IL-12, and tumor necrosis factor (TNF) appear to play important roles. Under the direction of IL-12, $CD4^+$ T cells from patients with GPA produce elevated levels of TNF, and peripheral blood mononuclear cells secrete increased amounts of IFN-γ. Serum levels of soluble receptors for TNF are elevated in patients with active GPA and normalize with the induction of remission. In vitro priming of activated neutrophils with TNF markedly

TABLE 270-4 PATHOLOGIC CHARACTERISTICS OF SELECTED FORMS OF VASCULITIS

	TAKAYASU ARTERITIS	POLYARTERITIS NODOSA	GRANULOMATOSIS WITH POLYANGIITIS (WEGENER GRANULOMATOSIS)	CHURG-STRAUSS SYNDROME	HENOCH-SCHÖNLEIN PURPURA	CUTANEOUS LEUKOCYTOCLASTIC ANGIITIS
Vessels involved	Elastic (large) or muscle (medium-sized) arteries	Medium-sized and small muscle arteries	Small arteries and veins; sometimes medium-sized vessels	Small arteries and veins; sometimes medium-sized vessels	Capillaries, venules, and arterioles	Capillaries, venules, and arterioles
Organ involvement	Aorta, aortic arch and major branches, and pulmonary arteries	Skin, peripheral nerves, gastrointestinal tract, and other viscera	Upper respiratory tract, lungs, kidneys, skin, eyes	Upper respiratory tract, lungs, heart, peripheral nerves	Skin, joints, gastrointestinal tract, kidneys	Skin, joints
Type of vasculitis and inflammatory cells	Granulomatous with some giant cells; fibrosis in chronic stages	Necrotizing, with mixed cellular infiltrate	Necrotizing or granulomatous (or both); mixed cellular infiltrate plus occasional eosinophils	Necrotizing or granulomatous (or both); prominent eosinophils and other mixed infiltrate	Leukocytoclastic, with some lymphocytes and variable eosinophils; IgA deposits in affected tissues	Leukocytoclastic, with occasional eosinophils

enhances the ability of ANCA to stimulate neutrophil degranulation. Despite the strong rationale for anti-TNF strategies in GPA, however, a randomized trial of etanercept showed no efficacy in the maintenance of disease remissions.

B-cell depletion is a more effective approach to the treatment of AAV. The efficacy of this treatment strategy probably relates to the removal of several B-cell functions beyond their evolution into plasma cells and the production of ANCA. These other B-cell functions include cytokine production, antigen presentation, and B cell–T cell crosstalk.

Superantigen Model

The degree of immune activation in Kawasaki disease and the acute but generally self-limited nature of this illness imply a potential role for superantigens. Superantigens are proteins produced by microbial pathogens (e.g., *Staphylococcus aureus* or *Streptococcus* species) that are capable of stimulating large populations of T cells in a manner unrestricted by the class II major histocompatibility complex (MHC). Superantigens bind directly to conserved amino acid residues outside the antigen-binding groove on class II MHC molecules, thereby selectively stimulating T cells that express particular β-chain variable gene segments. Through the binding of this MHC-superantigen complex to its cognate T-cell receptors, as many as 20% of circulating lymphocytes may become activated, leading to a potentially enormous outpouring of cytokines. With regard to the etiology of Kawasaki disease, substantial attention has focused on toxic shock syndrome toxin 1, an exotoxin produced by *S. aureus*.[2] Superantigens have also been postulated to play roles in the susceptibility to disease flares in GPA. Nasal carriage of *S. aureus* and superantigens associated with these organisms has been linked to a greater likelihood of disease flares in some studies.

Anti–Endothelial Cell Antibodies

Anti–endothelial cell antibodies can induce endothelial cell injury and lysis through either complement-mediated cytotoxicity or antibody-dependent cellular cytotoxicity. Both of these mechanisms have been demonstrated to cause endothelial injury in in vitro assays employing sera from patients with systemic vasculitis. The ability of these antibodies to damage endothelial cells is an appealing argument for their potential role in forms of vasculitis in which the endothelium is the focus of the inflammation (as opposed to the more external vessel wall layers). However, the true relevance of anti–endothelial cell antibodies to human disease and their importance within the larger context of other disease mechanisms remain unclear.

CLINICAL MANIFESTATIONS

Large-Vessel Vasculitides

Takayasu Arteritis

Takayasu arteritis (Chapter 78) affects the aorta and its major branches. In contrast to atherosclerosis, which is characterized by focal irregular lesions, the lesions of Takayasu arteritis are long, smooth, tapered stenoses (E-Fig. 270-1). The most commonly involved arteries are the subclavian and innominate arteries. Takayasu arteritis has been termed "pulseless disease" because of its ability to obliterate peripheral pulses (particularly in the upper extremities). Exuberant collateral circulation develops over time in response to the gradual narrowing of major arteries, making the loss of digits or limbs from ischemia extremely rare. The pulmonary circulation is involved in approximately 50% of cases of Takayasu arteritis.

Patients with severe narrowing of the aortic arch vessels supplying the head may develop Takayasu retinopathy, a hypotensive retinopathy leading to neovascularization. In contrast, patients with prolonged hypertension associated with renal artery stenosis demonstrate the classic ocular features of hypertension: "copper wiring" and multiple retinal infarctions. This complication is particularly difficult to diagnose and dangerous because vascular narrowings of large arteries to the arms and legs can cause underestimations of the true central aortic pressure. Takayasu arteritis involvement of the ascending aorta may lead to aortic dilation, aortic regurgitation, aneurysm formation, and aortic rupture.

TREATMENT Rx

The cornerstone of treatment of Takayasu arteritis is glucocorticoids. For patients with marked symptoms and signs of an inflammatory phase, prednisone (1 mg/kg/day) is usually effective in controlling the disease. This dose should be tapered within 8 to 12 weeks to less than 20 mg/day and ultimately to less than 10 mg/day as a maintenance dose. Emerging data support a role for IL-6 inhibition in patients with Takayasu arteritis whose prednisone doses cannot be tapered to reasonable levels. Patients have been treated with tocilizumab 8 mg/kg administered intravenously each month or with a corresponding subcutaneous preparation.

Giant Cell Arteritis

Giant cell arteritis is the other primary form of vasculitis that involves arteries far larger than vasculitides of any other category. This disease is discussed in detail elsewhere (Chapter 271).

Medium-Vessel Vasculitides

Polyarteritis Nodosa

Polyarteritis nodosa has a striking predilection for certain organs, particularly the skin, peripheral nerves, GI tract, and kidneys.[3] This disease usually begins with nonspecific symptoms such as malaise, fatigue, fever, myalgias, and arthralgias. Overt signs of vasculitis may not occur until weeks or months after onset of the first symptoms. Skin lesions of polyarteritis nodosa include livedo reticularis, subcutaneous nodules, ulcers, and digital gangrene. A majority of patients with polyarteritis nodosa (>80% in some series) have vasculitic neuropathy, typically in the pattern of a mononeuritis multiplex.

The classic GI manifestation of polyarteritis nodosa is "intestinal angina," the occurrence of postprandial abdominal pain. Polyarteritis nodosa can also affect individual GI tract organs such as the gallbladder or appendix, presenting as cholecystitis or appendicitis. The typical renal manifestation of polyarteritis nodosa is vasculitic involvement of the medium-sized intrarenal arteries, leading to renin-mediated hypertension and renal infarctions. Cardiac lesions, which usually remain subclinical, may lead to myocardial infarction or congestive heart failure. Polyarteritis nodosa usually spares the lungs.

The diagnosis of polyarteritis nodosa requires either a tissue biopsy or an angiogram that demonstrates microaneurysms (Fig. 270-1). Simultaneous nerve and muscle biopsies (e.g., sural nerve and gastrocnemius muscle) are of high yield if there is a clinical suspicion of vasculitic neuropathy. Symptoms suggestive of a neuropathy can be confirmed by electrodiagnostic studies that demonstrate a sensorimotor axonal neuropathy, often in a mononeuritis multiplex pattern. The pathologic changes in polyarteritis nodosa are limited to the arterial circulation, and the lesions are segmental, favoring the branch points of arteries. In gross pathologic specimens, aneurysmal bulges of the arterial wall may be visible. Histologic sections reveal infiltration and destruction of the blood vessel wall by inflammatory cells, accompanied by fibrinoid necrosis. Granulomatous inflammation is absent.

TREATMENT Rx

Approximately half of patients with polyarteritis nodosa achieve remissions or cures with high doses of glucocorticoids alone. Cyclophosphamide (2 mg/kg/day, adjusted for renal dysfunction) is indicated for patients whose disease is refractory to glucocorticoids or who have serious involvement of major organs. In recent years, therapeutic regimens involving lamivudine or entecavir and plasma exchange have substantially improved the treatment of hepatitis B virus (HBV)-associated polyarteritis nodosa. Because of increasing use of the HBV vaccine, fewer than 10% of polyarteritis nodosa cases now are associated with HBV infections.

Kawasaki Disease

Kawasaki disease occurs exclusively in young children. Because of its striking mucocutaneous findings and lymphadenopathy, Kawasaki disease is also known as mucocutaneous lymph node syndrome. Features of Kawasaki disease include high fevers, cervical adenopathy, conjunctival congestion, buccal erythema, prominence of the tongue papillae ("strawberry tongue"), a polymorphous truncal rash, erythema of the palms and soles, and desquamation of skin from the fingertips occurring days to weeks into the illness.[4] In its acuity and severity, Kawasaki disease resembles toxic shock syndrome and scarlet fever, both of which are mediated by superantigens (see Pathophysiology).

In a small number of patients with Kawasaki disease, panvasculitis in the coronary vessels leads to acute cardiac complications. Coronary arteritis leads to narrowing of the vessel lumen by the migration of myointimal cells

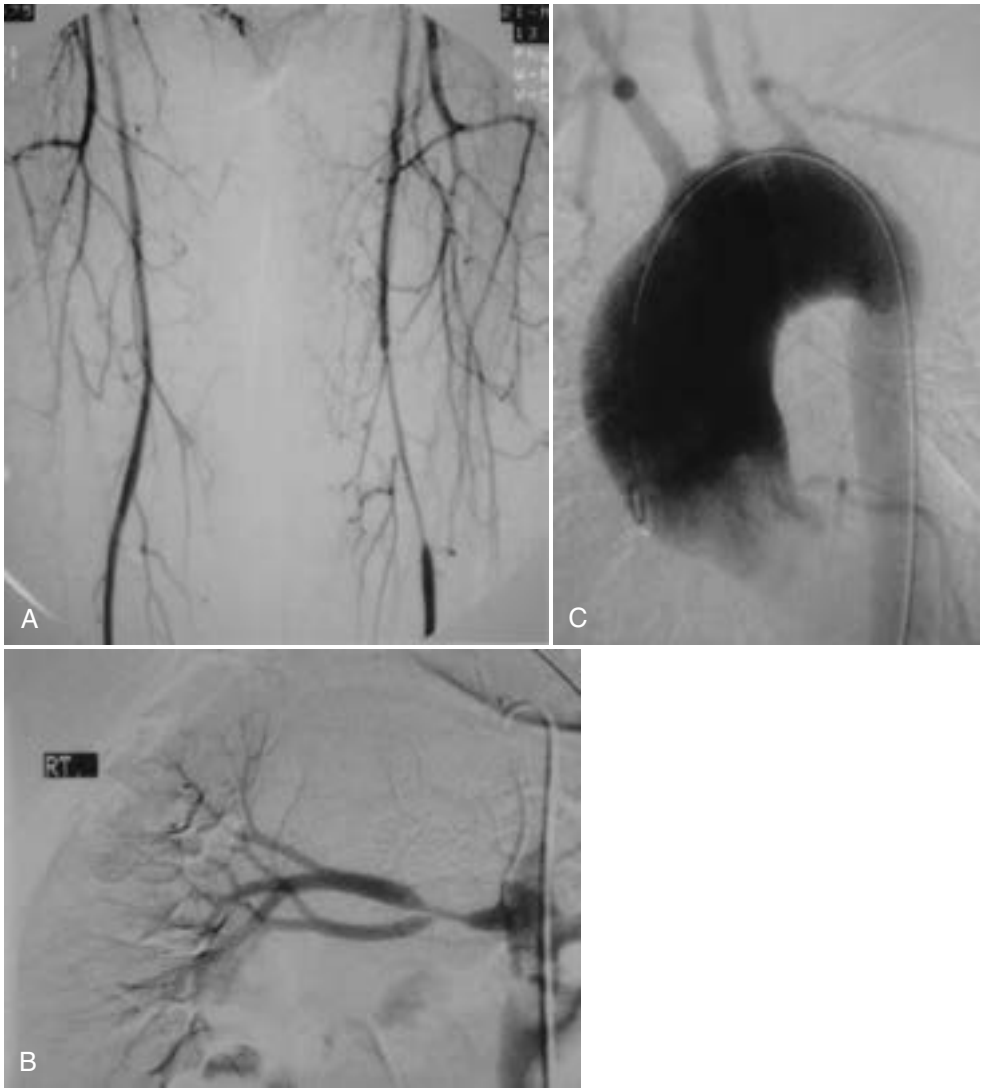

E-FIGURE 270-1. Large-vessel disease in Takayasu arteritis. Long, smooth tapering in the left common femoral artery (A) and the right renal artery (B). Dilation of the ascending aorta (C). Aortic regurgitation necessitated an aortic valve replacement and replacement of the ascending aorta with a Gore-Tex graft.

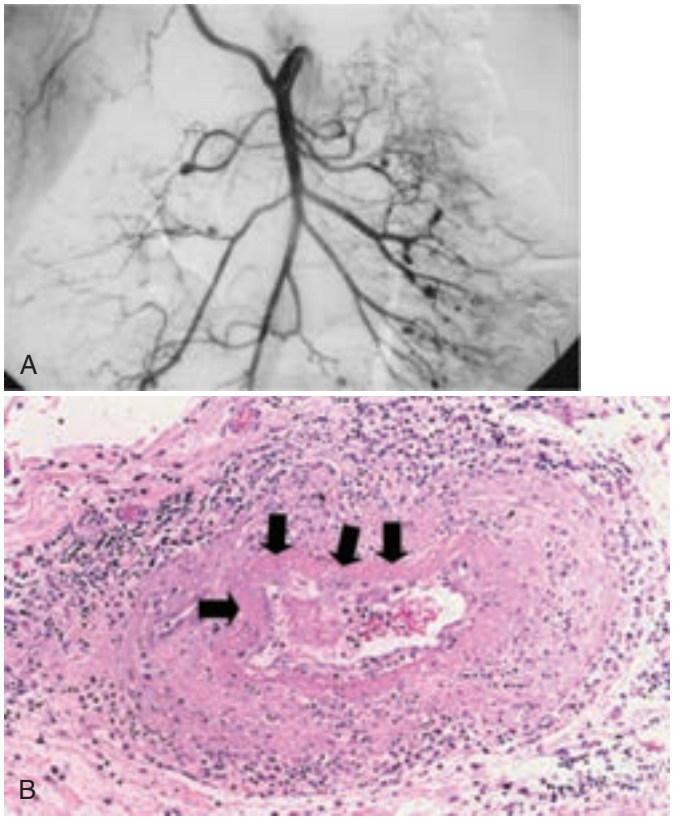

FIGURE 270-1. Vasculitis of medium-sized arteries in polyarteritis nodosa. A, Mesenteric angiogram showing numerous aneurysms in medium-sized arteries. B, Fibrinoid necrosis (*arrows*) in a jejunal artery from a patient who required surgical resection of necrotic bowel.

from the media through the fragmented internal elastic lamina. Direct complications include aneurysmal dilation and thrombosis of the coronary arteries, leading to myocardial infarction and possibly to death (in 1 to 2% of patients with Kawasaki disease during the acute illness). Late mortality from myocardial infarction may occur from the thrombosis of coronary artery aneurysms formed during the initial inflammatory stage. Such myocardial infarctions have been reported in middle-aged individuals who had febrile illnesses consistent with Kawasaki disease in childhood.

TREATMENT Rx

The recommended therapeutic regimen in Kawasaki disease is the combination of intravenous immune globulin (IVIG; 400 mg/kg/day on 4 consecutive days) and acetylsalicylic acid (100 mg/kg/day, lowered to 3 to 5 mg/kg/day after resolution of the fever). IVIG prevents the formation of coronary aneurysms in most cases. Glucocorticoids are reserved for salvage therapy in patients whose treatment with IVIG and acetylsalicylic acid has failed.

Buerger Disease

Buerger disease, also known as thromboangiitis obliterans (Chapter 80), is not always considered to be a primary form of vasculitis and was not included in the most recent consensus statement on nomenclature. Buerger disease has a remarkably strong yet poorly understood association with cigarette smoking and simply does not occur in the absence of exposure to tobacco. The vessels affected by Buerger disease are the distal medium-sized arteries and veins, particularly vessels at the levels of the ankles and wrists. The disease is characterized by thrombotic obliterations that begin distally and proceed proximally. Buerger disease tends to be segmental in nature, involving 5- to 10-cm lengths of blood vessels. Arterial obliteration leads to the development of collateral vessels with a "corkscrew" appearance on angiography. Vascular occlusion in Buerger disease often leads to the loss of digits and, if smoking persists, to loss of larger amounts of tissue (e.g., hands or feet). Despite the intense involvement of the extremities in Buerger disease, internal organ disease almost never occurs.

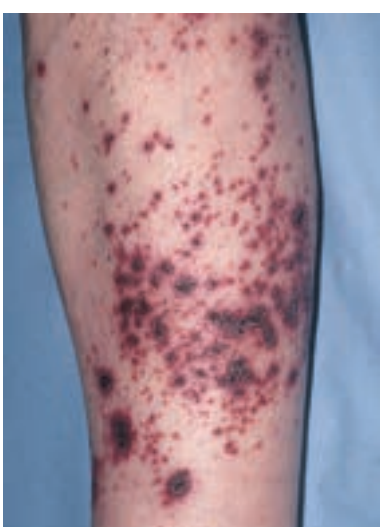

FIGURE 270-2. Cutaneous small-vessel vasculitis showing palpable purpuric lesions with necrosis and crusting.

TREATMENT Rx

Complete abstinence from tobacco is essential to the treatment of Buerger disease. Failure to stop smoking is associated with a dramatic increase in the risk of limb loss by amputation. No other therapeutic interventions, including glucocorticoids and anticoagulation, have dramatic effects on Buerger disease.

Small-Vessel Vasculitides

Antineutrophil Cytoplasmic Antibody–Associated Vasculitides

Granulomatosis with Polyangiitis

Classic GPA (formerly Wegener granulomatosis) involves the upper respiratory tract, lungs, and kidneys.[5] Distinctive features may also occur in the eyes, ears, and other organs. The three pathology hallmarks of GPA are (1) granulomatous inflammation in the upper or lower respiratory tract, (2) necrotizing vasculitis affecting arteries or veins, and (3) segmental glomerulonephritis associated with necrosis and thrombosis of capillary loops, with or without granulomatous lesions.

Approximately 90% of patients with GPA have nasal involvement, including crusting, bleeding, and obstruction. Cartilaginous inflammation may lead to nasal septal perforation and collapse of the nasal bridge ("saddle nose" deformity). Erosive sinus disease and subglottic stenosis (narrowing of the trachea just below the vocal cords) are highly characteristic of GPA.

Both conductive and sensorineural hearing loss can occur in GPA, though conductive lesions caused by middle ear disease are more common. Orbital masses ("pseudotumors" that develop behind the eye), scleritis, and peripheral ulcerative keratitis are the most dangerous ocular lesions. Episcleritis and conjunctivitis also occur. Anterior uveitis is rare. The clinical manifestations of GPA in the lung range from asymptomatic nodules to fulminant alveolar hemorrhage. The most common radiographic findings are pulmonary infiltrates, nodules, and cavitary lesions. Large-airway disease leading to bronchial narrowing is a challenging diagnosis to make because patients present with few symptoms until advanced disease is present.

The clinical presentation of renal disease in GPA is usually rapidly progressive glomerulonephritis: hematuria, red blood cell casts, and proteinuria (usually non-nephrotic). Without appropriate therapy, end-stage renal disease may ensue within weeks.

Sixty percent of patients with GPA have musculoskeletal symptoms during their disease course. The presenting complaint is frequently arthralgias or an oligoarthritis that is migratory in nature. Skin lesions in GPA include the full panoply of lesions associated with cutaneous vasculitis, including purpura (Fig. 270-2). Cutaneous nodules over the extensor surfaces of joints, particularly the elbow, may mimic rheumatoid nodules. These lesions are known as cutaneous extravascular necrotizing granulomata or Churg-Strauss lesions. Meningeal inflammation, presenting with headaches, cranial neuropathies, and a clinical picture compatible with chronic meningitis, is perhaps the most

common central nervous system (CNS) manifestation of GPA. Mononeuritis multiplex may affect the peripheral nervous system.

GPA is the prototype of conditions associated with ANCA. Positive results of immunofluorescence tests for ANCA in either the cytoplasmic (C-ANCA) or perinuclear (P-ANCA) pattern should be confirmed by enzyme immunoassays for antibodies to either proteinase 3 (PR3) or myeloperoxidase (MPO). An ANCA-negative assay sample does not exclude GPA, because a substantial minority of patients (between 15 and 40% overall) lack these antibodies. Furthermore, ANCA titers do not correlate reliably with disease activity.

TREATMENT Rx

Manifestations of GPA that constitute immediate threats either to the function of a vital organ or to the patient's life require treatment urgently. From the late 1960s until 2010, the combination of cyclophosphamide (2 mg/kg orally daily) and high doses of glucocorticoids (prednisone 1 mg/kg orally daily, tapered during 6 to 12 months) was the standard of care for GPA. Intermittent administration of cyclophosphamide by IV infusion is also effective in remission induction. However, a multicenter clinical trial that compared rituximab to cyclophosphamide in patients with either GPA or microscopic polyangiitis demonstrated that rituximab (375 mg/m^2 weekly times four) is at least as effective as the conventional regimen. Rituximab appears to be more effective for AAV patients who present with disease flares. An alternative dosing regimen of rituximab, 1 g times two separated by 2 weeks, may also be effective. Limited forms of GPA may respond to the combination of methotrexate (up to 25 mg/week) and glucocorticoids, but rituximab is now often employed in this setting as well. Rituximab (e.g., 500 mg every 6 months) is more effective than azathioprine for maintaining remission in patients who demonstrate a tendency to flare.

Microscopic Polyangiitis

Microscopic polyangiitis is characterized by (1) nongranulomatous necrotizing vasculitis with few or no immune deposits, (2) involvement of small (and possibly medium-sized) blood vessels in the arterial or venous circulation, and (3) tropism for the kidneys and lungs. Many cases of small-vessel vasculitis once regarded as polyarteritis nodosa are now classified more properly as microscopic polyangiitis. In contrast to polyarteritis nodosa, an ANCA-negative disorder, 70% of microscopic polyangiitis patients are ANCA positive.[6] Thus, microscopic polyangiitis is considered to be a form of AAV. The ANCAs in microscopic polyangiitis are usually directed against myeloperoxidase, leading to a perinuclear pattern of staining on immunofluorescence testing (P-ANCA). Microscopic polyangiitis is not characterized by granulomatous inflammation, and upper respiratory tract symptoms, if present at all, are much milder than those associated with GPA.

TREATMENT Rx

The approach to the treatment of microscopic polyangiitis is similar to the treatment of GPA. The combination of rituximab and glucocorticoids is the treatment regimen of choice for most patients with microscopic polyangiitis.

Eosinophilic Granulomatosis with Polyangiitis

EGPA is an eosinophil-rich form of granulomatous inflammation that involves the respiratory tract and other organs. The disease is associated with necrotizing vasculitis of small to medium-sized vessels. Two hallmarks of EGPA are asthma and eosinophilia. Several phases of EGPA are described:

- A prodromal phase characterized by the presence of allergic disease (typically asthma or allergic rhinitis), which may last months to many years
- An eosinophilia–tissue infiltration phase in which remarkably high peripheral eosinophilia may occur and tissue infiltration by eosinophils is observed in the lung, GI tract, and other tissues
- A vasculitic phase in which systemic necrotizing vasculitis afflicts a wide range of organs, ranging from the heart and lungs to peripheral nerves and skin

TREATMENT Rx

Patients with mild disease may be treated with prednisone. Those with evidence of neurologic, cardiac, renal, or GI involvement should be treated with cyclophosphamide in addition to glucocorticoids. Although clinical remissions are obtained in more than 90% of patients with EGPA, disease recurrences are seen in 25%. In most cases, relapses are heralded by the return of eosinophilia. Approximately 50% of cases of EGPA are associated with ANCA, usually directed against myeloperoxidase, but the percentage may be higher among untreated patients.

Immune Complex–Mediated Vasculitides

Anti–Glomerular Basement Membrane Disease

Anti–glomerular basement membrane (anti-GBM) disease is vasculitis affecting glomerular capillaries, pulmonary capillaries, or both, accompanied by the deposition of anti–basement membrane autoantibodies within basement membranes. Anti-GBM disease is discussed in detail elsewhere (Chapter 121).

Immunoglobulin A Vasculitis/Henoch-Schönlein Purpura

IgA vasculitis (IgAV) is characterized by non-thrombocytopenic purpura, arthritis, abdominal pain, and glomerulonephritis. The histopathologic findings are those of a leukocytoclastic vasculitis with IgA deposition. IgAV can develop at any age, but 80 to 90% of the cases occur in children. Although the cause is unknown, the disease's seasonal variation and the fact that two thirds of patients with IgAV experience antecedent acute upper respiratory illnesses suggest an infectious trigger. Medications such as antibiotics can also trigger IgAV, and environmental triggers are also likely. The diagnosis of IgAV can be confirmed only by demonstration of IgA deposition within and around blood vessel walls.

The classic IgAV patient presents with the acute onset of fever, palpable purpura on the lower extremities and buttocks, abdominal pain, arthritis, and hematuria. The clinician must be alert to the possibility of IgAV even when only parts of the syndrome are present. Most patients with IgAV, especially children, have a self-limited disease that lasts an average of 4 weeks.

TREATMENT Rx

Glucocorticoids ameliorate the GI, joint, and skin symptoms in many cases, but some patients respond surprisingly poorly to conventional doses of glucocorticoids, even in doses on the order of 40 to 60 mg/day. Anecdotal evidence suggests that pulse glucocorticoids (e.g., methylprednisolone 500 to 1000 mg/day times three doses) may abort persistent bouts of IgAV. The efficacy of glucocorticoids in the glomerulonephritis associated with this condition is controversial. Uncontrolled studies suggest that methylprednisolone pulses (1 g/day for three doses), followed by oral prednisone combined with azathioprine or mycophenolate mofetil may be useful in severe glomerulonephritis associated with IgAV.

Hypocomplementemic Urticarial Vasculitis

At least three subtypes of urticarial vasculitis are known: (1) normocomplementemic, a form that is generally idiopathic and benign (which may be viewed as a manifestation of cutaneous leukocytoclastic angiitis); (2) hypocomplementemic, a form that is often associated with a systemic inflammatory disease; and (3) hypocomplementemic urticarial vasculitis syndrome (HUVS), a potentially severe condition usually associated with autoantibodies to the collagen-like region of C1q. Most patients with the hypocomplementemic subtype have an underlying systemic disorder, such as systemic lupus erythematosus (Chapter 266) or Sjögren syndrome (Chapter 268). Many HUVS patients have C1q "precipitins," IgG autoantibodies to the collagen-like region of C1q that trigger the classical pathway of complement activation. The role of anti-C1q antibodies in disease pathogenesis remains unclear.

The lesions of urticarial vasculitis must be distinguished from the far more common chronic idiopathic urticaria (Chapters 252 and 440). Unlike idiopathic urticaria, the lesions of urticarial vasculitis last more than 48 hours, often have a purpuric component (i.e., they do not blanch), and resolve with postinflammatory hyperpigmentation. In urticarial vasculitis, lesions associated with vasculitis are often accompanied by stinging or burning. Urticarial vasculitis affects the capillaries and postcapillary venules, showing leukocytoclastic vasculitis on light microscopy. Direct immunofluorescence studies reveal both immunoglobulin and complement deposition in or around blood vessels of the upper dermis or the dermoepidermal junction.

TREATMENT Rx

Patients with urticarial vasculitis whose serum complement levels remain normal during attacks often have self-limited disease and require little therapy.

Other cases, especially HUVS, may cause life-threatening involvement of the lungs or other organs and require periods of intensive immunosuppression. Treatment decisions in HUVS must be individualized according to the patient's clinical status.

Cryoglobulinemia

Cryoglobulins are antibodies that precipitate from serum under conditions of cold and resolubilize on rewarming. Cryoglobulins are classified into types I, II, and III on the basis of whether monoclonality and rheumatoid factor activity (the ability to bind to the Fc portion of IgG) are present.[7] Type I cryoglobulins, which are monoclonal but lack rheumatoid factor activity, are associated with certain hematopoietic malignant neoplasms (e.g., multiple myeloma) and often lead to hyperviscosity rather than to vasculitis (Chapter 187). In contrast, type II and type III cryoglobulins may be associated with systemic vasculitis involving small (and often medium-sized) blood vessels. Vasculitis results from the deposition of cryoglobulin-containing immune complexes within blood vessel walls and the activation of complement.

Cryoglobulin types II and III are termed *mixed cryoglobulins* because they consist of complexes of both IgG and IgM antibodies. The IgM components in both type II and type III cryoglobulinemia possess rheumatoid factor activity (i.e., assays for rheumatoid factor are positive, indicating binding of the IgM antibody to the Fc portion of IgG). Whereas the IgM component in type II cryoglobulin is monoclonal, the IgM in type III cryoglobulin is polyclonal. Ninety percent of patients with vasculitis secondary to mixed cryoglobulins are hypocomplementemic, with C4 levels characteristically more depressed than C3. Infection with hepatitis C virus (HCV) accounts for at least 80% of the vasculitis cases associated with mixed cryoglobulins.[8]

TREATMENT Rx

Rapidly progressive sensorineural hearing loss requires early and requires aggressive therapy with high doses of systemic glucocorticoids. Some otolaryngologists also perform intratympanic injections of glucocorticoids. Cytotoxic agents can be considered for patients with suboptimal responses to glucocorticoids who still have salvageable hearing. Many Cogan syndrome patients become candidates for cochlear implants.

Variable-Vessel Vasculitides

The variable-vessel vasculitides have no predominant type of vessel involved but rather can affect vessels of any size (small, medium, and large) and any type (arteries, veins, and capillaries).

Cogan Syndrome

The combination of inflammatory eye disease and vestibuloauditory dysfunction is the sine qua non of Cogan syndrome.[9] In addition to inflammatory disease of the eyes and ears, up to 15% of patients with Cogan syndrome have vasculitis involving medium-sized to large blood vessels. Although the ocular manifestations vary, the classic presentation is the combination of interstitial keratitis and sensorineural hearing loss. Cogan syndrome may appear first in either the eyes or the ears. Although intervals as long as 1 to 2 years have been described between the start of disease in one organ and the appearance of disease in the other, the time between disease manifestations in these organs is usually only a matter of months. Patients usually present with photophobia and blurry vision, sometimes accompanied simultaneously by auditory or vestibular dysfunction. The vascular disease in Cogan syndrome resembles that of Takayasu arteritis.

TREATMENT Rx

The optimal therapy for most cases of cryoglobulinemic vasculitis is successful treatment of the underlying HCV infection. For cryoglobulinemic patients with relatively mild disease (e.g., frequent purpuric lesions, shallow cutaneous ulcers), short courses of prednisone followed by the institution of effective therapy for HCV may be sufficient. For patients with severe cutaneous ulcers, mononeuritis multiplex, glomerulonephritis, or other manifestations of severe disease, glucocorticoids, rituximab, and possibly a short course of plasma exchange may be indicated.

Behçet Syndrome

Behçet syndrome may affect small, medium, and large vessels in either the venous or the arterial circulation.[10] The most typical lesions in Behçet syndrome are mucocutaneous, reflecting the involvement of small blood vessels. The triad of recurrent mouth ulcers, genital ulcers, and eye inflammation is the classic presentation. The criteria for diagnosis of the International Study Group for Behçet Syndrome consist of one required manifestation—recurrent oral ulceration—plus at least two of the following: recurrent genital ulceration, characteristic eye or skin lesions, or a pathergy reaction (see later). However, the spectrum of Behçet syndrome encompasses many manifestations not included in these criteria.

Large-vessel complications of Behçet syndrome may include aneurysms in the pulmonary and systemic arterial systems. Venous complications include thromboses of the deep venous system, vena cava, portohepatic vein, and cerebral sinus. Pathergy—the development of pustules at the sites of sterile needle pricks—is a distinctive feature in many patients with Behçet syndrome, particularly those of Turkish origin. The arthritis of Behçet syndrome is a nondeforming, oligoarticular, asymmetrical arthritis of large joints. GI lesions in Behçet syndrome typically consist of ulcerations of the distal ileum or cecum. Crohn disease, which can cause genital ulcers as well as GI tract disease, may be particularly difficult to distinguish from Behçet syndrome.

TREATMENT Rx

Low-dose glucocorticoids are effective for intransigent mucocutaneous disease and may have a better side-effect profile than other medications used for this purpose (e.g., thalidomide). Intermittent courses of glucocorticoids during periods of particular mucocutaneous disease activity may be sufficient for patients with mild disease.

Severe disease in any organ system almost always requires high doses of prednisone (e.g., 1 mg/kg/day). Azathioprine (2 mg/kg/day), cyclosporine (3 to 5 mg/kg/day in two divided doses), methotrexate (up to 25 mg/week), and interferon alpha (3 million to 5 million units three times a week) are appropriate therapies for many complications of Behçet's syndrome. TNF inhibition with infliximab (5 mg/kg IV every 4 to 6 weeks) or adalimumab (40 mg every other week) is the treatment of choice for patients with the most severe forms of uveitis or meningoencephalitis.

Selected Single-Organ Vasculitides

Single-organ vasculitis is defined as vasculitis within the vessels of any type or size of a single organ, in the absence of any features (e.g., ANCA) suggesting one of the systemic forms of vasculitis.

Cutaneous Leukocytoclastic Angiitis

Cutaneous leukocytoclastic angiitis has also been termed *hypersensitivity vasculitis*. Cutaneous leukocytoclastic angiitis is the preferred name because no hypersensitivity or allergy is evident in many cases. Histories of exposure to new medications or to infections may be elicited. An immune complex deposition is central to the pathophysiologic process. Although it is occasionally associated with synovitis, other signs of systemic involvement are absent.

The skin lesions in cutaneous leukocytoclastic angiitis occur in "crops," coinciding with some period of elapsed time following exposure to the inciting antigen. The usual time between the exposure and the onset of clinically evident vasculitis is 10 to 14 days. The lesions typically occur first in dependent regions, such as on the lower extremities or buttocks. The rash may be asymptomatic but is usually accompanied by burning or tingling sensations.

TREATMENT Rx

Keys to the management of cutaneous leukocytoclastic angiitis include (1) exclusion of any underlying form of vasculitis that may cause subclinical involvement of other organs and (2) removal of any agent (e.g., a medication) that may have triggered the vasculitis. For patients in whom a precipitant can be identified, removal of the offending agent usually leads to resolution of the vasculitis within days to weeks. The type, intensity, and duration of therapy for cutaneous leukocytoclastic angiitis are based on the degree of disease severity. Mild cases may be treated simply with leg elevation, H_1 antihistamines, or low-dose prednisone. For persistent disease not associated with cutaneous gangrene, colchicine, hydroxychloroquine, or dapsone may be tried. For severe cases, high doses of glucocorticoids are indicated to suppress inflammation quickly and prevent skin ulceration.

Vasculitis of the Central Nervous System

CNS vasculitis includes two major categories of disease, one of which is not a true vasculitis. These conditions are primary angiitis of the CNS (PACNS)

TABLE 270-5 PRIMARY ANGIITIS OF THE CENTRAL NERVOUS SYSTEM (PACNS) VERSUS REVERSIBLE CEREBRAL VASOCONSTRICTION SYNDROME (RCVS)

	PACNS	RCVS
Female-to-male ratio	1 : 1	2-3 : 1
Onset	Subacute (weeks to months)	Sudden (seconds to minutes)
Headache	Insidious, dull	Thunderclap
Typical lumbar puncture findings	Abnormal in 50-80%: lymphocytic pleocytosis; elevated protein	Normal
Typical MRI findings	Multifocal subacute infarctions	Normal Watershed infarcts in minority
Typical angiogram findings	Normal in up to 40% of cases. Abnormal angiographic features when present cannot be distinguished from RCVS	Multifocal stenoses/dilatations
Utility of brain biopsy	Reasonable sensitivity in appropriately selected patients Important for excluding disease mimickers	Little to no role Helpful if confusing clinical situation confounds differentiation from PACNS or PACNS mimickers

MRI = magnetic resonance imaging.

and reversible cerebral vasoconstriction syndrome (RCVS). The diagnosis and management of these two conditions differ dramatically. The clinical, radiologic, and pathologic characteristics of PACNS and RCVS are shown in Table 270-5.

Primary Angiitis of the Central Nervous System

PACNS[11] typically develops in a subacute fashion, with the evolution of multifocal strokes, encephalopathy, headache, and other clinical features over months. Headache is often the first symptom. As the condition progresses, most patients develop lethargy, confusion, and memory loss. Some patients develop multifocal strokes, seizures, evidence of increased intracranial pressure, or myelopathy. The results of routine laboratory tests (e.g., erythrocyte sedimentation rate) are often normal in PACNS. Lumbar puncture demonstrates abnormalities of the cerebrospinal fluid in approximately 80% of cases, usually a modest monocytosis and elevated protein. Lumbar punctures should be performed in all patients in whom the diagnosis of PACNS is considered seriously. Although the findings on lumbar puncture in PACNS patients are nonspecific, a normal lumbar puncture argues against PACNS, and the procedure frequently identifies important PACNS mimickers such as infection or malignancy.

Magnetic resonance imaging (MRI) is the critical imaging modality in PACNS. Because of the subacute nature of the disorder, MRI studies reveal multifocal CNS infarctions in most cases. Strokes, hemorrhagic lesions, and mass lesions typically occur in more than one vascular territory. A normal brain MRI argues strongly against the diagnosis of PACNS. Angiography is less helpful in the evaluation of patients with PACNS for two main reasons. First, the sizes of blood vessels involved in PACNS are often too small to be resolved adequately, even by conventional angiography. The false-negative rate of angiography in PACNS is on the order of 35%. Second, the "classic" string-of-beads abnormality on angiography, produced by segmental arterial narrowing alternating with dilations, is nonspecific and can be mimicked perfectly by a host of nonvasculitic conditions (the most common of which is RCVS). No angiographic pattern is pathognomonic for PACNS, and there is a significant tendency to overdiagnose "vasculitis" on angiographic grounds alone. A normal brain MRI in the setting of an abnormal angiogram suggests RCVS, not PACNS.

TREATMENT Rx

When employed in appropriately selected patients whose history and radiologic studies suggest PACNS, brain biopsy is associated with reasonable positive and negative predictive values and frequently identifies important PACNS mimickers. Prednisone and cyclophosphamide are appropriate for treatment of patients who have abnormal findings on brain biopsy. Treatment courses of 6 to 12 months are recommended.

Reversible Cerebral Vasoconstriction Syndrome

Eighty percent of patients with RCVS[12] are women. RCVS is probably far more common than PACNS. Overtreatment of patients with RCVS who are misdiagnosed as having PACNS leads to substantial morbidity.

A careful history is the most important part of the evaluation. In contrast to the subacute course that typifies PACNS, RCVS usually begins in a more dramatic fashion with a "thunderclap" headache. Compared with PACNS, the neurologic signs are less severe in RCVS (e.g., encephalopathy is less common). RCVS frequently occurs in the setting of precipitants associated with vasospasm, such as in the postpartum setting or following the use of vasoactive agents such as nasal decongestants and recreational drugs.

The lumbar puncture is usually normal in RCVS, and brain MRI usually does not show multifocal CNS infarctions, with the exception of watershed infarctions mentioned earlier. The typical angiographic findings in RCVS—vascular narrowing and beading—are generally indistinguishable from those of PACNS and conditions that mimic PACNS. Multifocal vascular narrowing is particularly characteristic of RCVS. The most distinctive angiographic feature of RCVS is that the abnormalities are completely reversible, usually within 4 to 8 weeks. These abnormalities in RCVS are caused by vasospasm rather than true vasculitis. In the evaluation of patients with potential RCVS, a diagnostic strategy that can clinch the diagnosis is a follow-up angiogram 4 to 8 weeks after the first. Angiographic abnormalities due to RCVS will resolve in this interval.

TREATMENT Rx

Several approaches to the treatment of RCVS are reasonable. First, one may opt for watchful waiting. It is not clear that immunosuppression is either necessary or helpful. Moreover, attempts to treat vasospasm with calcium-channel blockers may lead to a vascular steal phenomenon, potentially causing harm. Second, because it is frequently difficult to do nothing for a patient with possibly serious CNS disease, calcium-channel blockers (e.g., nifedipine 30 mg three times daily) may be tried. Third, because of the frequent diagnostic uncertainty at the time of presentation, some clinicians opt to treat empirically with glucocorticoids (prednisone 1 mg/kg/day) for 1 month, followed by a taper over several weeks. Fourth, combinations of calcium-channel blockers and glucocorticoids are also reasonable, but cytotoxic therapy is not indicated in RCVS.

DIAGNOSIS

Differential Diagnosis

The major categories of diseases that can mimic vasculitis are displayed in Table 270-6. Certain features of a patient's case should raise the diagnostic suspicion for vasculitis. First, most cases of vasculitis do not begin suddenly but rather unfold subacutely during weeks or months. Second, pain is usually a prominent feature of vasculitis, resulting from arthritis or arthralgias, myalgias, headaches, neuropathy, testicular infarction, digital ischemia, sinusitis, otalgia, back pain (caused by aortic inflammation), postprandial abdominal pain (caused by mesenteric vasculitis), or other disease manifestations. Third, signs of inflammation such as fever, rash, weight loss, and elevated acute phase reactants are highly characteristic. Finally, multiorgan system involvement is the rule in vasculitis.

Ideally, the diagnosis of vasculitis is established through biopsy of an involved organ. Diagnoses based on angiography alone have many potential pitfalls, as discussed in the sections on PACNS and RCVS. Angiographic findings that are "consistent with vasculitis" must be interpreted in the proper context. A diverse array of other diseases, ranging from atherosclerosis to vasospasm to pheochromocytoma, may mimic the angiographic appearance of vasculitis. Systemic vasculitis can also be mimicked by two or more common medical problems or treatment complications occurring simultaneously in the same patient. Finally, high on the differential diagnosis of any individual form of vasculitis are other forms of vasculitis. For example, digital ischemia and splinter hemorrhages may be secondary to idiopathic polyarteritis nodosa. They may also be caused by polyarteritis nodosa associated with HBV infection, GPA, EGPA, microscopic polyangiitis, cryoglobulinemia, Buerger disease, or some other form of vasculitis. Because the

TABLE 270-6 MAJOR DISEASE CATEGORIES IN THE DIFFERENTIAL DIAGNOSIS OF VASCULITIDES

Other forms of vasculitis
Simultaneous occurrence of common medical problems in the same patient
Infections
Bacterial, viral, mycobacterial, fungal
Occlusive processes
Hypercoagulable states
Livedoid vasculopathy (atrophie blanche)
Atheroembolic disease
Malignant neoplasms
Lymphoma (including lymphomatoid granulomatosis)
Castleman disease
Amyloidosis
Paraproteinemias
Connective tissue disorders
Systemic lupus erythematosus, mixed connective tissue disease
Systemic sclerosis
Rheumatoid arthritis
Miscellaneous
Atrial myxoma
Calciphylaxis
Fibromuscular dysplasia
Neutrophilic dermatoses
Pyoderma gangrenosum
Sarcoidosis
Reversible cerebral vasoconstriction syndrome

appropriate interventions for these conditions vary widely, careful distinction among these potential etiologies is essential.

TREATMENT

Rx

The intensity of treatment in patients with vasculitis must be guided by the degree of disease activity. Specifically, the treatment of vasculitis should be predicated not only on abnormal laboratory test results but also on clear evidence of active disease. In addition, the intensity of treatment must be adapted to the type of vasculitis. Whereas giant cell arteritis responds to high doses of glucocorticoids in essentially all cases, for example, GPA nearly always requires an additional agent (rituximab, cyclophosphamide, or methotrexate) for disease control. In contrast, despite the dramatic fashion in which they sometimes present, most cases of IgA vasculitis and cutaneous leukocytoclastic angiitis require no immunosuppressive treatment at all.

Conventional therapies such as glucocorticoids, immunomodulating agents, and cytotoxic drugs induce remissions and control vasculitis in most cases. Moreover, in some cases—a variable percentage, depending on the type of vasculitis—the disease is curable. Unfortunately, the treatments of vasculitis have enormous potential for toxicity. Regular monitoring of patients' bone marrow, renal, and hepatic function is essential to avoid treatment-induced toxicity. Prophylaxis against opportunistic infections, particularly *Pneumocystis* pneumonia (Chapter 341), is an important part of many vasculitis treatment regimens. During the tapering of immunosuppressive medications, disease flares are common in many forms of vasculitis.

A common error is treating patients with high doses of immunosuppressive agents for too long. The most appropriate use of medications such as cyclophosphamide and glucocorticoids is to induce remission as quickly as possible with early, aggressive treatment regimens, and then to convert patients to safer treatments for the maintenance of remission. Rituximab is replacing cyclophosphamide as the drug of choice for some forms of vasculitis, particularly AAV. Patients with AAV who demonstrate a tendency to flare are often retreated with rituximab (500 mg or 1 g) every 4 to 6 months, at least until lengthy periods of disease control are established. Current treatment approaches to specific vasculitides are described under their "Clinical Manifestations."

PROGNOSIS

Assuming that the diagnosis is made before the patient has become catastrophically ill, the prognosis in systemic vasculitis is determined largely by the answers to four questions:

1. Was the diagnosis established before the occurrence of major irreversible organ damage?
2. Was aggressive (but appropriately dosed) treatment begun in a timely fashion?
3. Was there careful monitoring during treatment, and were specific steps taken to avoid drug-induced toxicity (e.g., opportunistic infection)?
4. Were the potentially toxic medications that induced remission stopped at an appropriate juncture and replaced with less dangerous medications (or was treatment stopped altogether)?

For most forms of vasculitis, the factors that determine long-term drug-free remissions remain poorly understood. The likelihood of achieving sustained remissions after discontinuation of all medications (or cures) varies according to the specific type of vasculitis.

FUTURE DIRECTIONS

Compelling laboratory and naturally occurring animal models of disease, combined with the known associations among HBV, HCV, and vasculitis in humans, suggest that additional links between infection and systemic vasculitis may be established in the future. Important strides have been made in the description of cytokine and chemokine pathways that are operative in vascular inflammation, but relevant anticytokine interventions remain to be defined for clinical therapies. B-cell depletion is emerging rapidly as the treatment of choice for some forms of severe vasculitis. IL-6 inhibition strategies are also likely to play important roles soon in the large-vessel vasculitides. Additional studies are required to define the full spectrum of clinical utility of these and other biologic agents.

Grade A References

A1. Newburger JW, Sleeper LA, McCrindle BW, et al. Randomized trial of pulsed corticosteroid therapy for primary treatment of Kawasaki disease. *N Engl J Med*. 2007;356:663-675.

A2. Stone JH, Merkel PA, Spiera R, et al. Rituximab versus cyclophosphamide for remission induction in ANCA-associated vasculitis. *N Engl J Med*. 2010;363:221-232.

A3. Specks U, Merkel PA, Seo P, et al. Efficacy of remission-induction regimens for ANCA-associated vasculitis. *N Engl J Med*. 2013;369:417-427.

A4. Guillevin L, Pagnoux C, Karras A, et al. Rituximab versus azathioprine for maintenance in ANCA-associated vasculitis. *N Engl J Med*. 2014;371:1771-1780.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Jennette JC, Falk RJ, Bacon PA, et al. 2012 Revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides. *Arthritis Rheum*. 2013;65:1-11.
2. Salgado-Pabon W, Case-Cook LC, Schlievert PM. Molecular analysis of staphylococcal superantigens. *Methods Mol Biol*. 2014;1085:169-185.
3. Forbess L, Bannykh S. Polyarteritis nodosa. *Rheum Dis Clin North Am*. 2015;41:33-46.
4. Sundel RP. Kawasaki disease. *Rheum Dis Clin North Am*. 2015;41:63-73.
5. Comarmond C, Cacoub P. Granulomatosis with polyangiitis (Wegener): clinical aspects and treatment. *Autoimmun Rev*. 2014;13:1121-1125.
6. Lyons PA, Rayner TF, Trivedi S, et al. Genetically distinct subsets within ANCA-associated vasculitis. *N Engl J Med*. 2012;367:214-223.
7. Damoiseaux J. The diagnosis and classification of the cryoglobulinemic syndrome. *Autoimmun Rev*. 2014;13:359-362.
8. Dammacco F, Sansonno D. Therapy for hepatitis C virus-related cryoglobulinemic vasculitis. *N Engl J Med*. 2013;369:1035-1045.
9. Kessel A, Vadasz Z, Toubi E. Cogan syndrome-pathogenesis, clinical variants and treatment approaches. *Autoimmun Rev*. 2014;13:351-354.
10. Hatemi G, Yazici Y, Yazici H. Behcet's syndrome. *Rheum Dis Clin North Am*. 2013;39:245-261.
11. Adams HP Jr. Cerebral vasculitis. *Handb Clin Neurol*. 2014;119:475-494.
12. Ducros A. Reversible cerebral vasoconstriction syndrome. *Handb Clin Neurol*. 2014;121:1725-1741.

REVIEW QUESTIONS

1. Immunosuppressive therapy is known to be futile for which disease associated with vascular inflammation?

A. Buerger disease (thromboangiitis obliterans)
B. Granulomatosis with polyangiitis (formerly Wegener granulomatosis)
C. IgA vasculitis (Henoch-Schönlein purpura)
D. Giant cell arteritis
E. Mixed cryoglobulinemia

Answer: A The only effective intervention known for Buerger disease is smoking cessation. Neither immunosuppression nor anticoagulation is of any demonstrable benefit in relieving the severe digital ischemia associated with this disease.

2. Which form of systemic vasculitis is most likely to present with an orbital pseudotumor?

A. Takayasu arteritis
B. Granulomatosis with polyangiitis (formerly Wegener granulomatosis)
C. Microscopic polyangiitis
D. Polyarteritis nodosa
E. Behçet syndrome

Answer: B Two common ocular manifestations of granulomatosis with polyangiitis are scleritis and orbital pseudotumor. The latter can lead to proptosis and vision loss if blood circulation to the eye is compromised by the retrobulbar mass.

3. A 24-year-old man from Lebanon presents with multiple aphthous ulcers, bilateral ocular erythema, tender erythematous nodules over his anterior legs, and a deep venous thrombosis. An ophthalmologic examination reveals severe bilateral uveitis affecting both the anterior and posterior compartments of the eye. In addition to anticoagulation and the institution of glucocorticoid therapy, what is the most appropriate approach to treating his disease at this time?

A. Intravenous cyclophosphamide
B. Azathioprine
C. Intravenous immunoglobulin
D. Tumor necrosis factor inhibition
E. B-cell depletion

Answer: D Tumor necrosis factor inhibition is highly effective in Behçet disease and is employed in the presence of severe disease manifestations such as posterior uveitis or central nervous system (CNS) disease.

4. A 32-year-old woman presents with the rapid onset of the worst headache of her life 1 day after giving birth to a healthy baby boy. The delivery followed an uneventful pregnancy. The patient has no history of migraine headaches, and her past medical history is unremarkable. The patient's blood pressure is 140/90 mm Hg, and although her neurologic examination is nonfocal, the severity of the headache prompts a computed tomographic (CT) scan of the brain. This is negative for an intracranial bleed, and the headache resolves after approximately 24 hours, during which time it was profoundly disabling.

Approximately 18 hours later, the headache returns, again rising to a crescendo within several minutes of onset. A repeat head CT is negative and a lumbar puncture is also performed, revealing a normal opening pressure, 1 white blood cell/mL (100% lymphocytes), and a cerebrospinal fluid protein of 40 mg/dL (normal 30 to 50 mg/dL). The headache improves following round-the-clock narcotic treatment but then worsens again on the third postpartum day. A magnetic resonance imaging study is negative for hemorrhage, cerebral infarctions, and masses, but a four-vessel cerebral angiogram shows alternating areas of vascular narrowing and dilatation (beading) in multiple vascular distributions.

What is the most appropriate intervention for this patient now?

A. Tighter blood pressure control, with transfer to the intensive care unit for sodium nitroprusside therapy
B. Addition of a baby aspirin to her anticoagulation with low-molecular-weight heparin
C. Pulse methylprednisolone (1000 mg/day for 3 days) followed by prednisone 60 mg/day and cyclophosphamide 100 mg/day
D. Nifedipine 30 mg three times daily
E. Proceed to brain biopsy

Answer: D Nifedipine. This patient has a history that is classic for reversible cerebral vasoconstriction syndrome (RCVS), an entity that mimics CNS vasculitis but is not an inflammatory condition. The features suggesting RCVS in this patient are the presence of a "thunderclap" headache, the nonfocal neurological examination, the normal cross-sectional imaging and lumbar puncture, and the diffuse beading on angiography. This beading represents vasospasm, which is the essential pathophysiology of RCVS. Calcium-channel blockers are often used in therapy, but prolonged courses of immunosuppression are not indicated. Brain biopsy can be diagnostic in primary angiitis of the CNS, but would not be indicated in a patient like this who has typical clinical and radiographic features of RCVS.

5. A 4-year-old boy presents with 3 days of a febrile illness with temperatures to nearly 105° Fahrenheit. Physical examination reveal an erythematous "strawberry" tongue, swollen lips, and a pink macular rash on the trunk. There is also tender cervical and axillary lymphadenopathy and diffuse swelling of the distal upper and lower extremities. Which treatment is indicated immediately now to prevent the occurrence of coronary aneurysms?

A. Intravenous immunoglobulin
B. High-dose aspirin therapy
C. Pulse methylprednisolone
D. Plasma exchange
E. Anticoagulation with heparin

Answer: A This child has Kawasaki disease, and the most feared complication of this disorder is formation of coronary aneurysms, which can lead to myocardial infarction. If administered in a timely manner, intravenous immunoglobulin prevents the occurrence of such aneurysms.

271

POLYMYALGIA RHEUMATICA AND TEMPORAL ARTERITIS

ROBERT F. SPIERA

DEFINITION

Polymyalgia rheumatica (PMR) and temporal arteritis, also called giant cell arteritis (GCA), are companion systemic inflammatory disorders of unknown etiology that represent a spectrum from severe proximal aches and pains and constitutional symptoms to an occlusive granulomatous vasculitis of medium and large vessels that can lead to permanent blindness or other organ and tissue damage. These disorders occur primarily in patients older than 50 years, in women more than in men; they are propagated by antigen-driven, cell-mediated (T_H1) immune mechanisms that may be associated with specific genetic markers and are highly responsive to corticosteroids.

EPIDEMIOLOGY

In the United States, the average annual incidence of PMR is 52.5 per 100,000 patients aged 50 years and older and increases with age. The prevalence is about 0.5 to 0.7%. Internationally, the frequency varies according to country, with the highest rates occurring in the Scandinavian countries.[1] The incidence and prevalence of GCA are approximately one third those of PMR.

PATHOBIOLOGY

The etiology of PMR and GCA is unknown, but both demonstrate familial aggregation and have a genetic association with human leukocyte antigen (HLA)-DR4 and a demonstrated sequence polymorphism encoded within the hypervariable region of the *HLA-DRβ1*04* gene.[2] Other genetic

associations have been suggested, including polymorphisms that may be seen in increased frequency in patients with the disease. Disease in genetically predisposed patients may be triggered by environmental factors such as viruses or endogenous antigens such as elastin, and their inflammatory manifestations are directed by specific patterns of cell-mediated, T_H1-associated cytokines. The cytokine production by the mononuclear cells in the involved tissues appears to influence the clinical phenotype. Cytokine profiles characterized in temporal artery biopsy specimens obtained from patients with PMR and GCA differ. GCA tissue contains the T-lymphocyte products interferon-γ and interleukin (IL)-2 and the macrophage products IL-1β, IL-6, and transforming growth factor-β. In PMR vascular tissue, transcripts are found for transforming growth factor-β, IL-1, and IL-2 but not for interferon-γ.[3] Patients with GCA who present with fever of unknown origin and who do not have ischemic symptoms, such as visual loss, have low interferon-γ levels. Arteries that express high interferon-γ levels typically have multinucleated giant cells present; these cells remove debris and secrete cytokines that stimulate intimal hyperplasia and lead to angiogenesis. IL-17-producing T_H17 cells have also been found in involved vascular tissue and peripheral blood of patients with untreated GCA but disappear rapidly with institution of corticosteroid therapy, whereas T_H1 cells are more persistent, speaking to the possibility that more than one antigenic trigger may be involved.[4]

The adventitia is considered the immunologic center in the pathogenesis of GCA. Macrophages and T lymphocytes enter the vessel wall through the vasa vasorum with the aid of adhesion molecules and come into contact with an inciting antigen. Here, it is likely that clonal proliferation of $CD4^+$ T cells is triggered by the presentation of unknown antigens by antigen-presenting cells. The activated CD4 cells produce interferon-γ that attracts macrophages to the arterial wall. Some of these macrophages fuse at the intima-media to form multinucleated giant cells. These cells produce vascular endothelial growth factor, which triggers neovascularization, both at the intima-media junction and at the level of the vasa vasorum, sprouting from the adventitia to the media. The subsequent immunologic events lead to a characteristic topography of mononuclear cells throughout the vessel wall. Products of the giant cells and macrophages at the intima-media junction include collagenase and nitric oxide, both of which probably contribute to tissue damage. The pathologic impact of cytokines leads not only to the characteristic medial damage but also to significant intimal hyperplasia that eventually, if it is not treated, may cause luminal narrowing and tissue ischemia.

In GCA, a transmural (involving all layers of the vessel) inflammatory infiltrate, comprising predominantly mononuclear cells and commonly with giant cells, is found in the superficial temporal arteries as well as in other large and medium-sized arteries. In elderly patients, fragmentation of the internal elastica is characteristic and helps differentiate this vascular lesion from that of atherosclerosis. Often, macrophages containing fragments of elastic tissue are found at the intima-media junction, the histologic center of the inflammatory process. As mentioned earlier, immunochemical techniques demonstrate differing patterns of cells and their proinflammatory and profibrotic products in the adventitia, media, and intima. Intimal proliferation may be prominent and lead to luminal narrowing. Fibrinoid necrosis, a common histologic feature in polyarteritis nodosa, is not seen in GCA.

In PMR, mononuclear cell inflammation can be found not only in the proximal joints, such as the shoulders, but also in the surrounding tendons, bursae, and soft tissues consistent with enthesitis. Although muscle pains may be present, no muscle inflammation is found.

CLINICAL MANIFESTATIONS

PMR and GCA are systemic inflammatory disorders that occur primarily in patients older than 50 years, in women more than in men (2:1), and in whites. PMR and GCA are particularly uncommon in African Americans. Shared characteristics of the two disorders include significant cytokine-driven constitutional symptoms, such as fever, fatigue, and weight loss, as well as a markedly elevated erythrocyte sedimentation rate (ESR), anemia, and thrombocytosis. The musculoskeletal hallmark of PMR is proximal, severe, and symmetrical morning and even day-long stiffness, soreness, and pain in the shoulder, neck, and pelvic girdles.[5] Fifty percent of patients with GCA share this characteristic proximal pain syndrome. Carpal tunnel syndrome and hand and knee synovitis may be seen in patients with PMR, but the overall presentation remains predominantly proximal, as opposed to rheumatoid arthritis, in which distal synovitis dominates. Whereas patients with PMR may appear to have proximal muscle weakness, this is invariably due to pain and not muscle inflammation (Table 271-1). Magnetic resonance imaging (MRI)[6] and ultrasound studies[7] in patients with PMR have confirmed the presence of inflammation of extra-articular synovial structures, in particular subacromial and subdeltoid bursae in the shoulders.

TABLE 271-1 GIANT CELL ARTERITIS: CLINICAL FEATURES

INFLAMMATORY
Polymyalgia rheumatica: constitutional symptoms
Fever
Weight loss
Fatigue
Laboratory abnormalities
Hematologic: anemia, thrombocytosis
Elevated sedimentation rate, C-reactive protein
ISCHEMIC
Ocular
Diplopia
Amaurosis fugax
Fixed vision loss
Complete blindness
Cranial symptoms
Headache
Jaw claudication
Scalp tenderness
Scalp or lingual necrosis (rare)
Cerebrovascular accidents
Large vessel disease
Leg or arm claudication
Diminished pulses, blood pressure asymmetry
Aortic aneurysms
LATE COMPLICATIONS
Aortic aneurysms
Thoracic aorta
Abdominal aorta
Corticosteroid complications
Osteoporosis
Fractures
Cataracts

Specific signs and symptoms of GCA are best appreciated in their anatomic and physiologic contexts. GCA preferentially affects certain blood vessels, including the branches of the external carotid artery, the ophthalmic artery and particularly its posterior ciliary branches, and the large arteries that arise from the aortic arch and abdominal aorta. Headache and scalp pain are probably the most frequent symptoms, occurring in 50 to 75% of patients. Headache is often the first manifestation of GCA and is described as boring, severe, and constant, unresponsive to simple pain medications and persisting through the night. Classically, patients complain of persistent and prominent temporal headaches, but occipital pains can also occur. Ear, pinna, or parotid gland pain may occur secondary to involvement of the posterior auricular artery. Jaw claudication and pain due to masseter muscle ischemia on chewing occur in 50% of patients. Lingual and maxillary artery involvement can lead to jaw or tongue pain on chewing or talking. The superficial temporal artery may become tortuous, prominent, nodular, or tender, but these findings are not invariable, and an abnormal temporal artery may be found on biopsy in vessels that appear normal. It is important to note that a dry, nonproductive cough can be a feature of the disease because this often may direct the clinician away from considering GCA and more toward consideration of an infectious or neoplastic respiratory cause of the symptoms. Rarely, mononeuritis multiplex and/or sensorineural hearing loss can occur but should lead the clinician to consider other possible vasculitides such as antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis or polyarteritis nodosa.

Fixed or intermittent symptoms related to vasculitic involvement of the ophthalmic arteries and its branches are the most dreaded in this illness and demand immediate therapeutic intervention. These symptoms are related to vascular narrowing due to both active inflammation and endothelial injury–mediated vasospasm. Decreased vision secondary to arteritis is the most common serious consequence of GCA, occurring in 20 to 50% of patients who present to ophthalmologists. It is the presenting symptom in 60% of patients with GCA who develop visual loss. A careful history of most patients who present with "sudden" visual loss reveals that preceding headache, constitutional symptoms, and PMR occurred in approximately 40% of patients. Even the evolution of the visual loss was often staggered, with amaurosis fugax in 10% and a partial field defect progressing to complete blindness over days. If GCA remains untreated, the second eye may become involved within

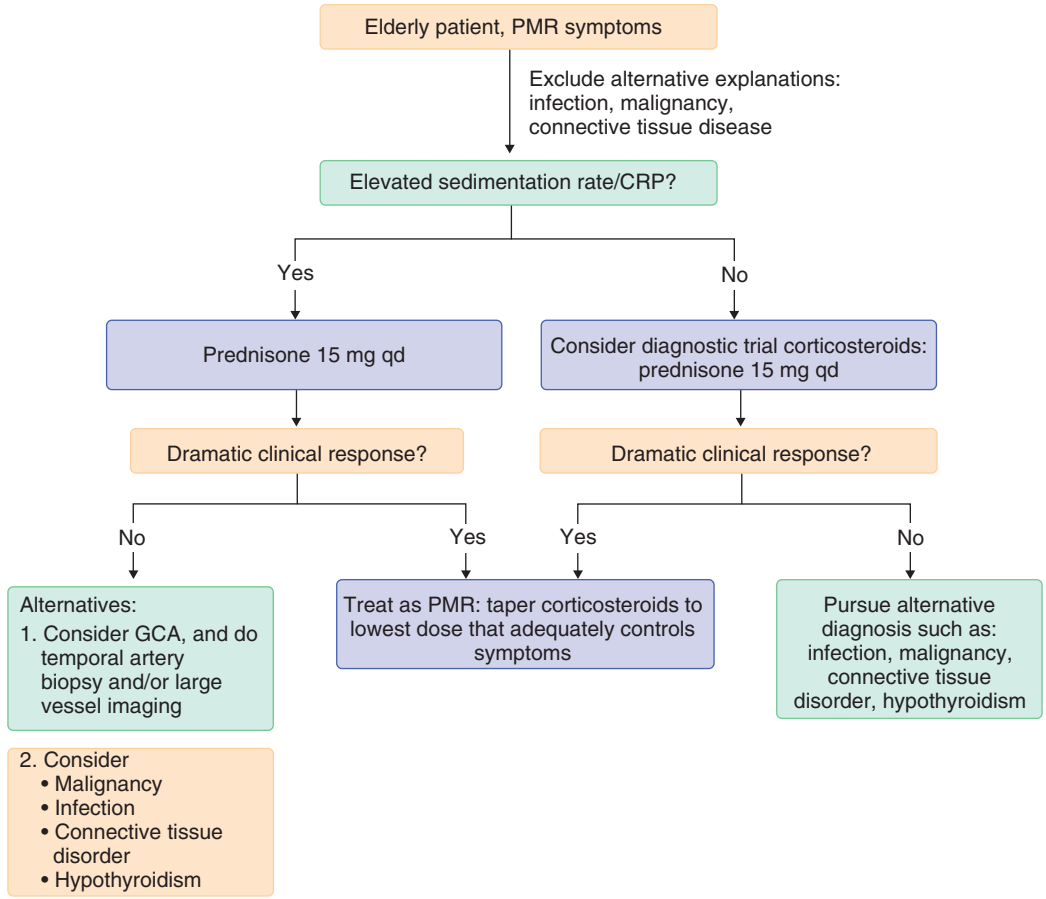

FIGURE 271-1. Diagnostic algorithm for polymyalgia rheumatica (PMR). CRP = C-reactive protein; GCA = giant cell arteritis.

1 to 2 weeks. The posterior ciliary arteries are the most frequently involved; thus, anterior ischemic optic neuropathy is the most common lesion, which can be easily defined by an ophthalmologist. Occlusion of the central retinal artery and its branches is uncommon; thus, exudates, hemorrhages, and frank vasculitis are infrequent. Five percent of patients with GCA may present with diplopia or ptosis, which may precede visual loss. The final visual abnormality can be a composite of many ischemic events occurring together in the optic nerve, the extraocular muscles, the chiasm, and the brain itself. Because GCA primarily involves arteries that contain elastica and the elastic lamina is lost as vessels pierce the dura, intracerebral lesions such as strokes are uncommon but not unheard of.

Large artery involvement most commonly presents as arm or leg claudication; rarer manifestations are stroke, subclavian steal syndrome, intestinal infarction, and symptomatic aortic aneurysm. Thus, a subclinical arteritis can exist and demands long-term monitoring. There is an emerging appreciation that some older patients classified as having GCA can present with large vessel disease resembling Takayasu's arteritis clinically, with a paucity of cranial ischemic symptoms but often the presence of PMR-like symptoms. Conversely, in patients presenting with typical GCA with cranial symptoms and a positive temporal artery biopsy, large vessel disease with aortic wall thickening is markedly more frequent than in matched controls without GCA, even early in the disease course. Steroid-treated PMR and GCA are self-limited illnesses lasting 1 to 2 years in most patients. However, a subgroup of patients with both disorders can have active inflammatory disease as manifested by persistent symptoms and blood test signs of active inflammation for 7 to 10 years. Of note is the fact that thoracic aneurysms with giant cells in the tissue can develop as long as 15 years after the diagnosis, successful treatment, and discontinuation of steroids. Indeed, the incidence of thoracic and aortic aneurysms is markedly higher in patients with prior history of presumably successfully treated GCA than in age-matched control subjects. Conversely, in studies of repaired aortic aneurysms, pathologic findings consistent with GCA have been found in approximately 2 to 4% of specimens from individuals without previously recognized or suspected arteritis. In most studies, survival rates for patients with PMR and GCA are similar to those of unaffected persons of the same age. However, one study did show that survival was decreased in a group of patients with GCA who had permanent visual loss and required more than 10 mg of prednisone per day at 6 months. This probably supports the experience that the morbidity and mortality are caused by steroid-related treatment complications in this high-risk, elderly group of patients possessing many comorbid conditions.

DIAGNOSIS

The diagnoses of PMR and GCA are based on clinical facts, with supporting but not diagnostic aid obtained from laboratory tests and temporal artery biopsy (Fig. 271-1).[8] No physician should await an abnormal finding on temporal artery biopsy or demand the presence of an elevated ESR before making the definitive diagnosis of GCA in the setting of a characteristic clinical picture. That said, the laboratory hallmark of PMR and GCA is an elevation in IL-6-stimulated acute phase reactants such as the ESR and C-reactive protein. The ESR is usually in excess of 50 mm/hour and may exceed 100 mm/hour. An ESR in the low 20s or 30s, however, does not exclude a diagnosis of PMR or GCA if other characteristic clinical features are present and especially if the patient is already taking steroids.

Normocytic, normochromic anemia and thrombocytosis occur in approximately 50% of patients with both disorders and are excellent guides to the state of inflammation. In both PMR and GCA, the frequency of rheumatoid factor, antinuclear antibody, ANCA, monoclonal proteins, and cryoglobulins is not higher than in age-matched control subjects, and complement is not reduced. Alkaline phosphatase activity may be elevated in one third of patients, primarily those with GCA. Although these tests are not indicated in PMR and GCA, muscle enzymes and electromyography are normal, and muscle biopsy shows type II fiber atrophy but no inflammation.

Superficial Temporal Artery Assessment

Temporal artery tenderness, nodularity, and diminished pulsation are typical findings on physical examination in a patient with GCA. Color duplex ultrasonography has been used as an adjunctive noninvasive diagnostic tool in GCA. A hypoechoic halo around the superficial temporal artery has been reported in 73% of patients with biopsy-proven GCA. The halo, representing edema in the arterial wall, was observed bilaterally in a significant subset of patients and disappeared in a mean of 16 days after the initiation of steroids in one study. The presence of the halo in this study had a sensitivity of 73% and was 100% specific for GCA. Other groups have been unable to replicate this experience, however, finding Doppler ultrasonography to be no more

sensitive or specific than physical examination in patients thought to have the disease. Findings of stenosis or occlusion of temporal arteries by Doppler ultrasound have also been recognized as being modestly sensitive and specific for the diagnosis of GCA in some studies. Operator dependency remains a challenge to the more widespread use of this modality diagnostically. [^{18}F] Fluorodeoxyglucose–positron emission tomography may be helpful in identifying large vessel inflammation suggestive of GCA, but it is not helpful in assessing the temporal arteries themselves, given their relatively small size and high background uptake in that area. Conventional angiography is rarely used in the diagnosis of GCA. Some studies have suggested that MRI/magnetic resonance angiography (MRA) may be a helpful noninvasive diagnostic modality. Superficial cranial arteries can be visualized, and mural inflammatory changes and luminal narrowing can be identified. Large vessel involvement can also be assessed. Studies have suggested sensitivities and specificities of MRI/MRA similar to those of biopsy in the diagnosis of GCA. Nevertheless, temporal artery biopsy remains the diagnostic "gold standard" in GCA, and given the relatively easy accessibility of the artery and potentially significant morbidity of therapy in GCA, histologic confirmation is favored in most cases.

Although temporal artery biopsy continues to be an important diagnostic test for the presence of GCA, a few caveats must be stated. First, in a patient in whom the clinical diagnosis is likely, treatment with steroids should be instituted immediately without waiting for the biopsy results. Second, because of the skipped nature of the pathologic inflammatory lesions in the vessel wall, as many as 20 to 30% of biopsy specimens may be normal despite an overwhelming diagnostic likelihood of GCA. However, because the biopsy is helpful in confirming the diagnosis of GCA, in which high doses of steroids are used, the following guidelines are given. Patients with pure PMR and no GCA signs or symptoms do not need a biopsy. However, because 10% of these patients may develop such clinical manifestations of GCA within the next year, they should be told to report such symptoms immediately. When GCA is likely, an outpatient biopsy should be performed on the symptomatic side of the head, preferably including inflamed areas with tenderness or nodularity and incorporating 2 to 3 cm of vessel. Multiple sections should be requested because of the segmental nature of the disease process. Some rheumatologists routinely request bilateral biopsies, which may increase the likelihood of obtaining an abnormal finding by up to 5%, whereas others perform a contralateral biopsy if the first specimen is normal. Diagnostic biopsy findings continue to be present for as long as 2 to 4 weeks after the clinical diagnosis is made and steroid treatment instituted.[9]

Differential Diagnosis

The systemic nature of these disorders and the fact that they occur in elderly people demand careful diagnostic scrutiny to avoid missing a malignant neoplasm or major infection and possibly treating patients inappropriately with high-dose steroids. This is true in PMR because there is no diagnostic test and in GCA because the GCA biopsy finding may be normal in the face of active, vision-threatening vasculitis. Infections that must be considered and ruled out if clinically appropriate include tuberculosis, endocarditis, and hepatitis B and C. Malignant neoplasms such as lymphoma and multiple myeloma may mimic PMR, and an age-appropriate cancer evaluation is always indicated in this age group. Autoimmune disorders such as elderly-onset rheumatoid arthritis and systemic lupus erythematosus, as well as dermatomyositis and other types of vasculitis, must be considered in the differential diagnosis and sorted out by employing clinical information and serologic testing. There is support for the concept that elderly-onset rheumatoid arthritis is the same disorder as PMR with negative rheumatoid factor, a more proximal focus of joint inflammation, and a good response to low-dose prednisone. The distinction may be semantic because neither disorder tends to evolve into an erosive arthritis. A more protracted clinical course, however, is often seen in patients in whom distal synovitis is a prominent feature, and those patients are classified as having elderly-onset rheumatoid arthritis. PMR and GCA should always be thought of in the setting of a fever of unknown origin because symptoms and signs can be occult or the history incomplete.

TREATMENT

Rx

Both PMR and GCA are highly responsive to corticosteroids, which are the preferred treatment choice.[10] This response is so characteristic that an immediate and dramatic improvement in PMR and GCA symptoms within 1 to 3 days after steroid institution supports the diagnosis. Conversely, a lack of rapid and significant improvement in signs, symptoms, and function within 5 to 7 days should lead the clinician to suspect the initial impression and consider an alternative diagnosis (e.g., tumor or infection) or the presence of GCA in PMR patients that might require a higher steroid dose. Because the inflammatory burden of the two disorders is different, different doses of steroids are employed at the onset of treatment. Whereas PMR usually responds to 15 mg of prednisone daily, GCA usually requires 40 to 60 mg of prednisone per day in divided doses or higher doses if organ or tissue damage is present or threatened.[11] In GCA, if visual symptoms are present as a fixed loss or amaurosis fugax, the patient often should be treated with high-dose intravenous methylprednisolone with doses ranging from 40 mg every 8 hours to 1 g/day for 3 days, followed by high-dose oral steroids in divided doses.

Within 2 to 3 days after the institution of steroids, most symptoms of PMR or GCA clear rapidly, and patients describe a miraculous improvement. The steroid dose is then maintained for 2 to 3 weeks, during which the ESR, C-reactive protein, hemoglobin, and platelet counts normalize. Steroid taper is then instituted and guided by the clinical response. In PMR, taper is commonly by 1 mg every 7 to 10 days; in GCA, taper is by 5 to 10 mg every 7 to 10 days. In GCA, the use of alternate-day corticosteroid regimens to minimize steroid side effects is generally not recommended because randomized controlled trials have demonstrated higher rates of treatment failure with alternate-day dosing schedules. It is important that the taper be guided primarily by clinical findings (e.g., PMR stiffness, headache, fatigue) and that the level of ESR elevation be considered within that clinical context. One should never "chase the ESR" because the elderly patient would be subjected inappropriately to a dangerously high cumulative dose of steroids with their attendant side effects. An increased dose of prednisone should be based on a change in symptoms, not solely on an increase in the ESR. One possible exception is in a patient with a history of GCA and prior abrupt vision loss in one eye, in whom any further compromise of vision would be catastrophic. The effective dose demanded for a flare often can be as low as 5 to 10 mg of prednisone, and uncommonly up to 60 mg/day to control symptoms (e.g., visual abnormalities). A persistently elevated ESR (>50 mm/hour) without PMR or GCA symptoms should alert the physician to look for alternative causes, such as infection. Treatment is a careful balancing act between disease control and avoidance of steroid-related toxicity. The overall goal of the patient and the physician is to attain the best disease control with the lowest dose of steroids. In most patients, prednisone can be tapered safely in 1 to 2 years. However, other patients may need to take low doses of steroids for 2 years or more. The higher the initial dose and cumulative dose, the greater the likelihood that the patient will develop a major steroid side effect such as sepsis, osteoporosis, osteonecrosis, diabetes, emotional lability, or myopathy. Appropriate immunizations, osteoporosis regimens (calcium, vitamin D, and bisphosphonates), and metabolic monitoring are mandatory in all patients prescribed chronic steroid therapy.

The major feared outcome in GCA is ischemic complications of the disease, most often vision loss or, less frequently, cerebrovascular accident. Vision loss is usually irreversible, and although it is uncommon after the diagnosis is suspected and glucocorticoid therapy is instituted, it can occur early in the course of treatment. Aspirin is known to be protective against ischemic events in patients with atherosclerosis and has anti-inflammatory effects in inflamed blood vessels, including inhibition of interferon-γ. A recently reported cumulative meta-analysis of retrospective studies showed that antiplatelet or anticoagulant therapy has a marginal benefit when used together with corticosteroids in patients with established GCA.[12] Although this has not been demonstrated in prospective randomized controlled trials, in most patients, adjunctive therapy with low-dose aspirin should be considered unless there is a strong contraindication to its use.

Alternative immunosuppressive agents have been tested in both PMR and GCA patients in an attempt to "spare steroids" and to control the inflammatory state. Studies examining the efficacy of methotrexate in GCA have yielded mixed results, with the largest, most recent study showing no incremental benefit from combined therapy. One individual patient meta-analysis of three randomized placebo controlled trials suggested a modest benefit to methotrexate in GCA in terms of affording a steroid-sparing benefit and reducing likelihood of flares. Given the modest nature of the benefit demonstrated and the potential toxicities of methotrexate in this elderly population, methotrexate is not routinely incorporated as first-line therapy in GCA. In PMR, methotrexate has been shown in one study to afford a benefit in terms of steroid sparing and possibly reducing numbers of flares. The magnitude of the benefit appears modest, and no reduction in corticosteroid-related side effects was demonstrated. At present, methotrexate is not routinely used in the management of either disease, but in individual patients with refractory disease or excessive corticosteroid morbidities, addition of weekly rheumatoid arthritis–level doses of methotrexate (7.5 to 20 mg/week) or azathioprine (2 mg/kg/day) is employed in selected instances. There have been case series suggesting cyclophosphamide may be of value in patients with refractory disease and/or unacceptable corticosteroid toxicity, but adverse events were common, and it is rare to use this agent in GCA.

Randomized controlled trials of TNF inhibitors including infliximab and adalimumab in GCA have failed to demonstrate benefit in terms of preventing

relapses or affording a steroid-sparing benefit.[13] Abatacept, a costimulatory molecule blocker, is presently being evaluated in large vessel vasculitis including GCA, in a large randomized trial, but those results are not yet available. Tocilizumab, an IL-6 inhibitor, has been of major interest in GCA and PMR because this cytokine seems pivotal in these disorders. Tocilizumab reliably reduces C-reactive protein and ESR, which in theory could be independent of a true beneficial effect on the underlying arteritis, but early controlled observations are encouraging.[14] This will be a challenge in assessing the true efficacy of Tocilizumab in GCA in the larger controlled clinical trials that are in progress.

FUTURE DIRECTIONS

Better understanding of the disease-causing roles of immunologically active cells and their cytokine products, along with genetics and correlations with clinical subsets, will lead to more focused treatment modalities and the avoidance of the need for long-term treatment with steroids. A recently published cohort study revealed that GCA is associated with increased risks for myocardial infarction, stroke, and peripheral vascular disease,[15] suggesting that greater attention should be paid to cardiovascular risk reduction in patients with this disease.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Kermani TA, Warrington KJ. Polymyalgia rheumatica. *Lancet*. 2013;381:63-72.
2. Carmona FD, Gonzalez-Gay MA, Martin J. Genetic component of giant cell arteritis. *Rheumatol (Oxf)*. 2014;53:6-18.
3. Samson M, Audia S, Martin L, et al. Pathogenesis of giant cell arteritis: new insights into the implication of CD161+ T cells. *Clin Exp Rheumatol*. 2013;31(1 suppl 75):S65-S73.
4. Weyand CM, Younge BR, Goronzy JJ. IFN-γ and IL-17: the two faces of T-cell pathology in giant cell arteritis. *Curr Opin Rheumatol*. 2011;23:43-49.
5. Mackie SL, Mallen CD. Polymyalgia rheumatica. *BMJ*. 2013;347:f6937.
6. Klink T, Geiger J, Both M, et al. Giant Cell Arteritis: Diagnostic Accuracy of MR Imaging of Superficial Cranial Arteries in Initial Diagnosis-Results from a Multicenter Trial. *Radiology*. 2014;273:844-852.
7. Ball EL, Walsh SR, Tang TY, et al. Role of ultrasonography in the diagnosis of temporal arteritis. *Br J Surg*. 2010;97:1765-1771.
8. Nesher G. The diagnosis and classification of giant cell arteritis. *J Autoimmun*. 2014;48-49:73-75.
9. Mackie SL, Pease CT. Diagnosis and management of giant cell arteritis and polymyalgia rheumatica: challenges, controversies and practical tips. *Postgrad Med*. 2013;89:284-292.
10. Weyand CM, Goronzy JJ. Giant-cell arteritis and polymyalgia rheumatic. *N Engl J Med*. 2014;371:50-57.
11. Waldman CW, Waldman SD, Waldman RA. Giant cell arteritis. *Med Clin North Am*. 2013;97:329-335.
12. Martinez-Taboada VM, Lopez-Hoyos M, Narvaez J, et al. Effect of antiplatelet/anticoagulant therapy on severe ischemic complications in patients with giant cell arteritis: a cumulative meta-analysis. *Autoimmun Rev*. 2014;13:788-794.
13. Yates M, Loke YK, Watts RA, et al. Prednisolone combined with adjunctive immunosuppression is not superior to prednisolone alone in terms of efficacy and safety in giant cell arteritis: meta-analysis. *Clin Rheumatol*. 2014;33:227-236.
14. Loricera J, Blanco R, Castañeda S, et al. Tocilizumab in refractory aortitis: study on 16 patients and literature review. *Clin Exp Rheumatol*. 2014;32:S79-S89.
15. Tomasson G, Peloquin C, Mohammed A, et al. Risk of cardiovascular disease early and late after a diagnosis of giant-cell arteritis: a cohort study. *Ann Intern Med*. 2014;160:73-80.

REVIEW QUESTIONS

1. A 74-year-old gentleman presents with a 7-week history of increasing generalized shoulder and hip pain, morning stiffness, low-grade fevers, and a dry cough. He was found to be anemic with a hemoglobin of 9.7 and had a thrombocytosis with a platelet count of 642 and an elevated sedimentation rate of 114. A search for infection and neoplasia has been unrevealing, including a bone marrow biopsy, which did not suggest a myeloproliferative neoplasm. He denies headaches, visual symptoms, or jaw claudication. The patient has reluctance regarding any further interventional procedures. Which of the following would be the best option to establish his diagnosis?

A. A computed tomographic (CT) angiogram with imaging of the aortic arch and its major branches
B. A conventional angiogram
C. Duplex ultrasound of his temporal arteries and axillae.
D. A positron emission tomography (PET) scan
E. Bilateral temporal artery biopsies

Answer: E The patient likely has giant cell arteritis. Although PET scanning, duplex ultrasound of temporal arteries, and CT angiography may be helpful in some patients, a temporal artery biopsy remains the gold standard and is the most specific if the biopsy can be established. Duplex sonography has yet to be proved reliably accurate. PET scanning and angiography, although potentially helpful, can provide in false-positive results in patients with severe atherosclerosis and false-negative results in patient without clear evidence of large vessel disease. Polymyalgia rheumatica (PMR) may be a possibility here (and this patient indeed has PMR), but the presence of dry cough and degree of inflammatory response makes concurrent giant cell arteritis (GCA) a pressing concern

2. A 77-year-old woman with a history of diabetes, osteoporosis, and modest alcohol use presents with 3 weeks of headaches and feeling systemically unwell and then loss of vision in her left eye 2 days ago. She is found to be anemic with a markedly elevated sedimentation rate, and a temporal artery biopsy confirms the diagnosis of GCA. She is started on 60 mg of daily prednisone. Which of the following would be an appropriate additional intervention at this point?

A. Add Infliximab 5 mg/kg monthly as a steroid-sparing intervention.
B. Add methotrexate 10 mg weekly with the hope of titrating that to 15 mg weekly to afford a steroid-sparing benefit.
C. Change her corticosteroid regimen to 120 mg on alternate days and plan an alternate day taper from there.
D. Add low-dose aspirin daily to her regimen.
E. Add tocilizumab 8 mg/kg monthly as a steroid-sparing intervention.

Answer: D Patients with GCA often struggle with comorbidities that raise concerns regarding prolonged exposure to corticosteroids. There is at this time, however, no proven steroid-sparing intervention that can decrease the risk for cranial ischemic complications or even that has been demonstrated to reduce corticosteroid-related adverse events. Low-dose aspirin has been shown in two retrospective studies to be likely protective against cranial ischemic complications, so if there is no contraindication, that would be an appropriate intervention. Methotrexate has been shown in one trial in one individual meta-analysis to have a corticosteroid benefit in terms of cumulative steroid dose and also reduced risk for flares, but in the largest multicenter study addressing its efficacy, it was not shown to be superior to placebo. Moreover, in this patient with comorbidities including diabetes and alcohol use, methotrexate could be morbid, so this would not seem an appropriate intervention at this point. Infliximab has been shown in a randomized, double-blinded, placebo-controlled trial not to be helpful in the treatment of GCA and can increase risk for infection in this patient already on high doses of corticosteroids with underlying diabetes. Alternate-day corticosteroid therapy has been shown to increase the risk for later flare and is not generally employed in GCA, particularly earlier in the course of treatment. Tocilizumab, an interleukin-6 inhibitor, is an agent of great interest in temporal arteritis, but it has not yet been shown to have a corticosteroid-sparing or disease-modifying benefit in this disorder; however, trials are presently in progress addressing this issue.

3. A 72-year-old man presents with a 2-month history of morning stiffness, mild synovitis in his knees, puffiness in his hands in the early morning hours, and severe shoulder and thigh girdle myalgias. He is mildly anemic with hemoglobin of 11.6 and has a markedly elevated sedimentation rate at 82. He is seronegative for rheumatoid factor, citric citrullinated peptide (CCP), and antineutrophil cytoplasmic antibodies. He has no evidence of peripheral joint erosions on plain radiographs. Which of the following would be the most appropriate therapeutic intervention at this point?

A. Begin naproxen 500 mg twice daily.
B. Begin methotrexate 10 mg weekly with the hope of bringing that dose up to 15 mg weekly in the next few weeks.
C. Begin methylprednisolone 12 mg daily.
D. Begin hydroxychloroquine 200 mg twice daily.
E. Begin methylprednisolone 12 mg daily along with methotrexate 10 mg weekly with a plan to taper the corticosteroids to off over the next 4 weeks.

Answer: C The patient described has PMR but presents with some peripheral synovitis, which can be seen in up to 10 to 15% of patients with PMR. The patient clinically is unlikely to have rheumatoid arthritis in the absence of seropositivity for rheumatoid factor and CCP, has typical shoulder and thigh girdle myalgias and morning stiffness, and does not have evidence of erosive disease radiographically. An initial trial of methylprednisolone 12 mg daily would have both therapeutic and likely diagnostic value because if his response is spectacular, that would strongly argue for PMR and continuing that approach to therapy. Although methotrexate and hydroxychloroquine have been agents of interest in polymyalgia rheumatica, their use has not been well established in double-blinded, placebo-controlled trials. One study suggesting a benefit to methotrexate in PMR did not demonstrate a benefit in terms of reducing numbers of corticosteroid-related adverse events, and at longer term follow-up, the initial benefits seen by the addition of methotrexate did not seem to persist. Naproxen would be unlikely to afford adequate relief to this patient and, given his comorbidities, would likely be as problematic if not more problematic than a trial of corticosteroids.

4. An 80-year-old gentleman with a prior history of giant cell arteritis successfully treated 7 years earlier presents with searing chest pain. His past medical history is otherwise only notable for borderline hypertension. He was successfully treated with a 1-year course of steroids but has been free of corticosteroids for 3 years. He has been feeling systemically well until recently. He had not seen a physician in 2 years. Which of the following explanations for his chest pain is of primary concern?

A. Dissecting thoracic aneurysm
B. Acute pericarditis
C. A myocardial infarction related to his increased risk for atherosclerosis given his prior history of GCA
D. Costochondritis related to a flare of PMR
E. Rib fracture

Answer: A Patients with a prior history of GCA are at increased risk for thoracic aortic aneurysms, even several years after completing therapy without evidence of residual active disease. The searing quality of his pain and his prior history of treated GCA make this a concern that should not be missed. The quality of his pain did not sound suggestive of a myocardial infarction, and patients with GCA have not been demonstrated to be at increased risk for atherosclerotic heart disease, although the possibility of that association has been of interest. Patients with GCA are not particularly at a higher risk for pericarditis. Although PMR can arise even years after successfully treated GCA (or even years after successfully treated PMR), the quality of his pain does not sound suggestive thereof, and costochondritis would not typically be a feature of PMR. The quality of the pain described did not sound suggestive of a rib fracture, but moreover that would be a diagnosis that could be overlooked at least initially without putting the patient at increased risk for a serious cardiovascular compromise.

5. A 72-year-old woman had presented with an 8-week history of progressively increasing morning stiffness, shoulder and thigh girdle myalgias and arthralgias, and fatigue. An extensive search for underlying infection and age-appropriate screening for underlying neoplasm were unrevealing. A serum immunofixation did not reveal a monoclonal gammopathy. She was found to be anemic with a hemoglobin of 10.7 and had a sedimentation rate of 108. She denied any headaches, scalp sensitivity, jaw claudication, or visual symptoms. She admitted to low-grade fevers but not greater than 100° F. PMR was suspected, and she was treated with methylprednisolone 12 mg daily in divided doses. Three days later, she called to report no significant improvement in her symptoms. It was suggested that her dose be raised to 16 mg daily in divided doses. She comes in for evaluation still complaining of similar myalgias and arthralgias, and her sedimentation rate remains elevated at 62, although her hemoglobin improved slightly to 10.9. Which of the following would be an appropriate next intervention?

A. Add methotrexate as a disease-controlling/steroid-sparing intervention.
B. Add infliximab to help achieve disease control.
C. Refer her for a bone marrow biopsy to exclude underlying myeloproliferative or lymphoproliferative disease.
D. Add hydroxychloroquine 400 mg daily.
E. Refer her for a temporal artery biopsy and raise the methylprednisolone to 32 mg daily.

Answer: E This patient presented with fairly typical PMR symptoms but in addition had low-grade fevers, which can be seen but are not as common as they are in GCA. Moreover, she was somewhat anemic. Her lack of response to 12 mg and then even 16 mg of daily methylprednisolone makes PMR unlikely unless in the context of a more substantial inflammatory disease such as GCA. GCA is found in approximately 10% of patients with "pure" PMR, and conversely, up to 50% of patients with GCA can present with PMR-like symptoms.

272

INFECTIONS OF BURSAE, JOINTS, AND BONES

ERIC L. MATTESON AND DOUGLAS R. OSMON

INFECTION OF BURSAE

Septic Bursitis

DEFINITION

Bursae are the satellite structures that form to protect tissues from bony prominences. The superficial bursae, including the olecranon, prepatella, infrapatella, and bursae over the first metatarsophalangeal bunions, are more likely to become infected than are the deep bursae, such as the subacromial, trochanteric, and iliopsoas bursae.[1]

EPIDEMIOLOGY

Olecranon bursitis may occur in as many as 10 in 100,000 persons. The majority of cases occur in men, and antecedent trauma to the skin is frequent.

PATHOBIOLOGY

Septic bursitis of superficial bursae is most commonly due to direct inoculation through the overlying skin; less commonly, it is secondary to overlying cellulitis. Most cases of deep septic bursitis are due to contiguous spread from adjacent infected joints or hematogenous seeding.

Predisposing risk factors for septic bursitis include trauma to the skin, as may occur in plumbers, athletes, and patients with chronic obstructive pulmonary disease (COPD) who frequently lean on the elbows; prepatellar or infrapatellar septic bursitis occurs in housecleaners, gardeners, and carpet layers. At least one third of patients with septic bursitis have an underlying comorbid illness such as diabetes mellitus, rheumatoid arthritis, gout, COPD, or alcoholism.

CLINICAL MANIFESTATIONS

In immune-competent patients, septic bursitis often but not always presents with fever and erythema and warmth of the overlying skin; there may be swelling of the bursae. In contrast to those with septic arthritis, patients with septic bursitis of superficial bursae have intact range of motion of the joints, which may be limited only at the extremes of flexion. Pain on motion of the joint and restriction of joint range of motion are highly suggestive of septic arthritis. Acute phase reactants such as C-reactive protein, the sedimentation rate, and the white blood cell count (WBC) may be elevated

DIAGNOSIS

Radiography should be performed to look for a foreign body and to evaluate the surrounding bones. Aspiration of bursal fluid is helpful in the diagnosis of patients who have pain, erythema, and/or swelling of an affected area. However, given the risk for contaminating the bursa if the aspiration occurs through skin involved with cellulitis, many clinicians choose to aspirate a bursa only if empirical antimicrobial therapy has failed. Ultrasound or computed tomography (CT) guidance greatly enhances the successful aspiration of superficial bursae. Care must be taken not to violate the joint space when aspirating a bursa to avoid inoculating the joint space.

The leukocyte count of the bursal fluid is generally lower than that seen in septic arthritis, with a mean of 13,500 cells/mm^3. Even in immune-competent hosts, cell counts can range from less than 1500 to greater than 100,000/mm^3. A leukocyte count greater than 2000/mm^3 has a sensitivity of 94% and a specificity of 79% for superficial (olecranon or prepatellar) bursitis. Bacterial culture and in vitro susceptibilities must be obtained; if additional fluid is available, a Gram stain may be obtained, although its sensitivity may be as low as 15%. The presence of crystals does not exclude the possibility of septic bursitis.

Staphylococcus aureus (Chapter 288) is the most common cause of septic bursitis, present in more than 80% of culture-proven cases, followed by β-hemolytic streptococci. Aerobic gram-negative bacilli, including *Escherichia coli*, *Campylobacter jejuni*, and *Pseudomonas* species, are rare causes of septic bursitis. Chronic bursitis may be associated with systemic infections due to *Brucella abortus*, atypical mycobacteria, or *Mycobacterium tuberculosis*, as well as fungi; the presence of these infections should raise the possibility of systemic infection.

Differential Diagnosis

In the immune-competent host, nonseptic bursitis (Chapter 263) may have a somewhat more indolent presentation than septic bursitis. The differential diagnosis includes gout, pseudogout, arthritis, and trauma with hemobursa. An overlying cellulitis may be confused with bursitis. Fever is usually not present in nonseptic bursitis due to mechanical or friction trauma.

TREATMENT

Rx

Treatment of septic bursitis is guided by knowledge of the putative underlying organisms, in most cases, *S. aureus*. Because the Gram stain is positive in less than two thirds of patients and cultures may be delayed, empirical therapy is guided by the clinical presentation. Most patients can be treated as outpatients, but those who are immunocompromised may require hospitalization for intravenous antibiotic therapy. Initial ambulatory treatment in patients without comorbidities may consist of an oral antistaphylococcal penicillin or first-generation cephalosporin. If community-acquired methicillin-resistant *S. aureus* (MRSA) is suspected, co-trimoxazole or minocycline may be added to one of these agents. In patients who are allergic to penicillin, oral clindamycin or linezolid may be used. Patients who have severe inflammation, are septic, or are immunocompromised may require hospitalization for initiation of treatment with intravenous nafcillin, oxacillin, or cefazolin; if MRSA is suspected, intravenous vancomycin, daptomycin, or linezolid should be used. Guidelines on the treatment of MRSA have been published.[2] Vancomycin can also be used in patients who are allergic to penicillin.

The duration of antimicrobial therapy is guided by the clinical response and comorbid states. It should be continued until there is no longer bursal inflammation. This may require several weeks of intravenous or oral therapy and multiple aspirations. Failure of the septic bursitis to respond to initial antibiotic therapy mandates a second course of therapy; recurrence thereafter or inability to adequately drain the bursa with needle aspiration is an indication for surgical intervention.

PREVENTION

Because superficial septic bursitis is often associated with occasional or avocational activities involving kneeling or resting on the elbows, using protective padding may be helpful.

PROGNOSIS

The optimal duration of therapy is unknown, but prognosis of superficial bursitis is generally excellent. The presence of comorbid conditions,

especially those associated with deep bursal infections, including septic arthritis, bacteremia, and osteomyelitis, is associated with more intractable and difficult disease.

INFECTION OF JOINTS

Septic Arthritis

DEFINITION

Septic arthritis refers to infection of a joint by a microorganism. It is associated with increased morbidity and mortality, as well as loss of articular integrity and function. Septic arthritis is usually caused by a bacterial infection.[3] Other microorganisms can cause infections with clinical characteristics that differ from those of bacterial infections; these are reviewed separately.

NONGONOCOCCAL SEPTIC ARTHRITIS

EPIDEMIOLOGY

The incidence of septic arthritis affecting native joints is about 5 to 8 in 100,000 patient years. Among patients presenting with an acutely swollen and painful joint, the prevalence of bacterial arthritis ranges widely, from less than 10% to as high as 27%, depending on the source population. Nongonococcal septic arthritis is the most common form of septic arthritis and is somewhat more common in men than in women.

PATHOBIOLOGY

More than 90% of cases of septic arthritis are due to staphylococci or streptococci (Table 272-1). Septic arthritis can result from direct inoculation (e.g., accidents, bites, surgery) or by extension from infected bone into an adjacent joint space. Approximately 75% of cases are due to hematogenous spread, particularly in patients with indwelling catheters and immunocompromised patients. Septic arthritis due to needle arthrocentesis (<1 in 10,000 procedures) or arthroscopy (4 cases per 1000 to 10,000 procedures) is very rare.

Bacteria causing septic arthritis produce an acute inflammatory reaction in the synovial membrane. Synovial hyperplasia and inflammatory cell immigration with the release of pro-inflammatory and cartilage-destroying cytokines and proteases result in damage to cartilage and bone. Bacterial toxins and DNA and superantigens, such as those seen in staphylococcal toxic shock syndrome, also contribute to cartilage and bone damage.

Risk factors for the development of septic arthritis include diabetes, alcoholism, cutaneous ulcers, intravenous drug use, prosthetic joints, rheumatoid arthritis, osteoarthritis, and low socioeconomic status, as well as advanced age, skin infection, indwelling intravenous catheters, cancer, and immunosuppressive therapies, including biologic response modifiers used in the management of autoimmune diseases such as rheumatoid arthritis and inflammatory bowel disease.

CLINICAL MANIFESTATIONS

Most patients with bacterial arthritis feel ill and have fever. Immunocompromised and elderly patients may not have a marked febrile response. Most cases (>80%) of septic arthritis are monoarticular; the knee is involved in more than 50% of cases. Polyarticular joint sepsis may be seen in immunocompromised patients and those with rheumatoid arthritis or systemic lupus erythematosus. Such patients frequently lack typical signs and symptoms of infection and may not appear to be particularly ill at presentation, but they may develop rapid cardiovascular decompensation. This is particularly true for patients who are taking glucocorticosteroids, other immunosuppressive agents, and biologic response modifiers, including tumor necrosis factor-α (TNF-α) inhibitors.

Patients with septic arthritis affecting nondiarthrodial joints, such as the acromioclavicular or sacroiliac joints, may have a history of intravenous drug use or may have intravenous catheters in place to treat other medical conditions. Infection of the symphysis pubis is associated with previous urinary tract surgery, pelvic malignancy, intravenous drug use, or vigorous weight-bearing physical activity, such as long-distance running, in female athletes.

A finding of microorganisms in the joint should lead to an appropriate history and physical examination to identify a source of hematogenous infection, such as cellulitis, pneumonia, or urinary tract infection. Staphylococci and β-hemolytic streptococci may enter directly through open wounds, whereas gram-negative infection may be associated with bowel or bladder disease.

DIAGNOSIS

Plain radiography should be performed to evaluate the surrounding bones and joint space and to provide a baseline for comparison after therapy is completed. Imaging modalities such as magnetic resonance imaging (MRI), CT, and plain radiography are useful to determine whether there is associated osteomyelitis and in cases of diagnostic uncertainty. Blood cultures are positive in up to 50% of patients with bacterial septic arthritis and should be obtained in all patients in whom this diagnosis is suspected.

TABLE 272-1 MICROORGANISMS RESPONSIBLE FOR ACUTE SEPTIC ARTHRITIS AND ACUTE AND CHRONIC OSTEOMYELITIS

SEPTIC ARTHRITIS		*OSTEOMYELITIS: ACUTE AND CHRONIC*	
MICROORGANISM	**FREQUENCY (%)**	**MICROORGANISM**	**FREQUENCY (%)**
Gram Positive	***60-90***	***Gram Positive***	***80-90***
Staphylococcus aureus	50-70	*Staphylococcus aureus*	60-80
Group A, B, C streptococci	15-30	Group A, B, C streptococci	10-20
Coagulase-negative staphylococci	6-20	*Staphylococcus epidermidis*	10-15
Streptococcus pneumoniae	1-3	*Streptococcus pneumoniae*	<1
Enterococcus sp	<1	*Enterococcus* sp	1-2
Corynebacterium sp	<1	*Corynebacterium* sp	1-2
Gram Negative	***5-25***	***Gram Negative***	***5-20***
Salmonella sp		*Salmonella* sp	
Pseudomonas aeruginosa		*Enterobacter* sp	
Escherichia coli		*Pseudomonas aeruginosa*	
Klebsiella pneumoniae		*Brucella* sp	
Enterobacter sp		*Pasteurella multocida*	
Kingella kingae		*Bartonella henselae*	
Haemophilus influenzae	<1-3*	*Propionibacterium* sp	
Anaerobes	***1-2***	***Anaerobes***	
Fusobacterium sp		*Bacteroides* sp	
Bacteroides fragilis			
Miscellaneous	***<5***	***Miscellaneous***	***5-7***
Mycoplasma		*Mycobacterium* sp	
Mycobacterium sp		Fungi (candidiasis, coccidioidomycosis, blastomycosis, histoplasmosis)	
Fungi			
Viruses			
Algae			

*Children.

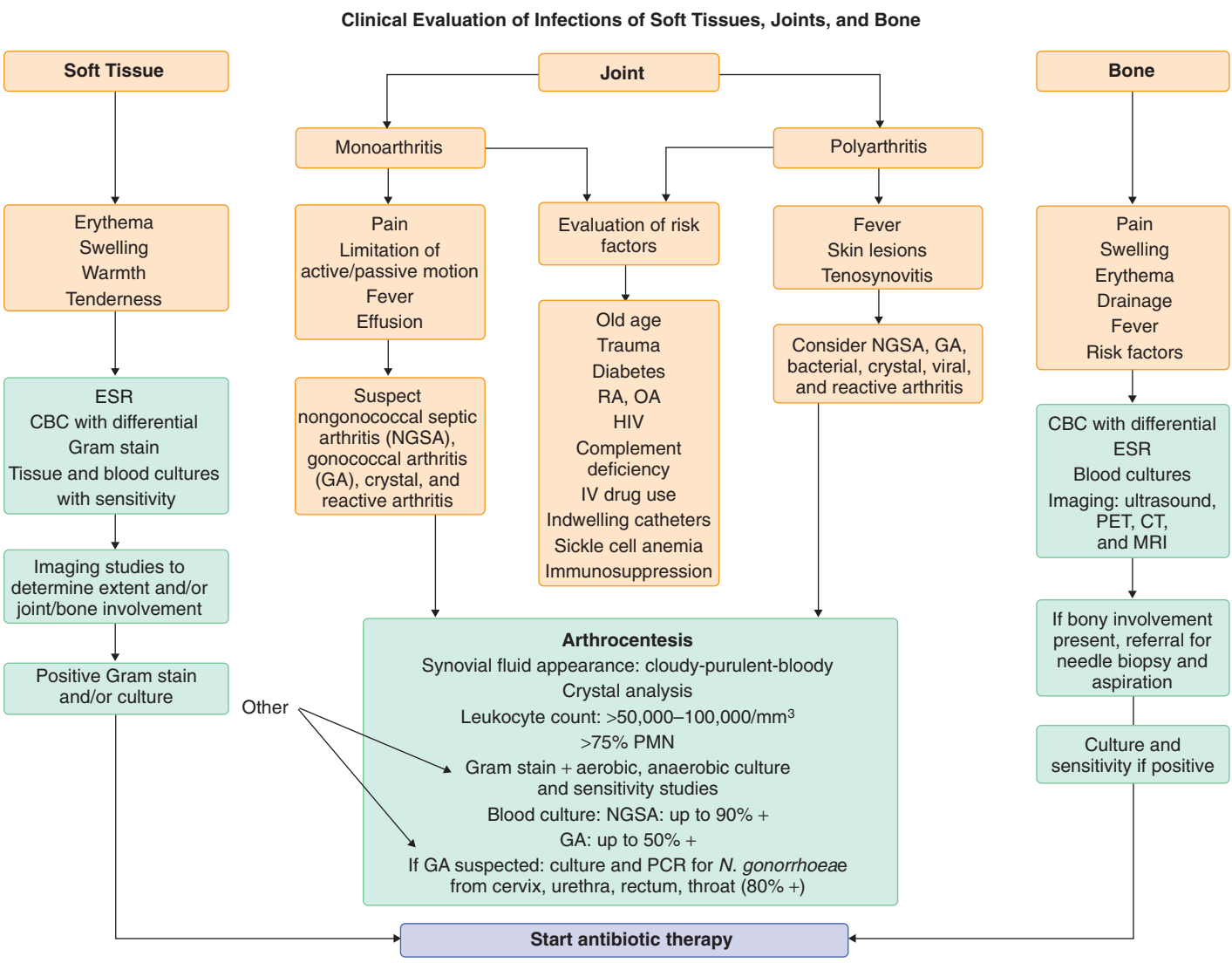

FIGURE 272-1. Clinical evaluation of infections of soft tissues, joints, and bone. CBC = complete blood count; CT = computed tomography; ESR = erythrocyte sedimentation rate; HIV = human immunodeficiency virus; IV = intravenous; MRI = magnetic resonance imaging; OA = osteoarthritis; PCR = polymerase chain reaction; PET = positron emission tomography; PMN = polymorphonuclear leukocyte; RA = rheumatoid arthritis.

If septic arthritis is suspected, synovial fluid arthrocentesis is indicated, and the fluid should be examined for bacterial culture and Gram stain; the latter is positive in only about 50% of patients. Specific cultures and stains for fungal and mycobacterial organisms should be done if there is a history of exposure or if antibacterial therapy has failed. Polymerase chain reaction (PCR) assays may be helpful for diagnosing less common joint infections such as *Borrelia,* but the value of PCR over standard culture for the diagnosis of staphylococcal or streptococcal joint infection has not yet been demonstrated. Other helpful examinations include leukocyte count and differential, as well as evaluation of the synovial fluid for the presence of crystals. The presence of gout or pseudogout crystals does not exclude the possibility of septic arthritis, particularly in patients whose WBC is above 50,000/mm^3.

A frequent clinical scenario is the patient who is anticoagulated. Because of the rapidly destructive nature of septic arthritis and the often profound systemic consequences, anticoagulation is not a contraindication to arthrocentesis. The procedure may be assisted by ultrasound guidance, especially when only small amounts of joint fluid are present or when the joint is difficult to aspirate. CT-guided arthrocentesis is particularly useful for aspiration of deep joints such as the hips and nondiarthrodial joints.

The total WBC and differential in synovial fluid are helpful in distinguishing infected from noninfected joints in immunocompetent patients. A diagnosis of septic arthritis is present in 47% of patients with a synovial WBC greater than 50,000/mm^3 and in 77% of patients with a WBC greater than 100,000/mm^3. It is important to realize that a WBC less than 50,000/mm^3, especially in immunocompromised patients, can be associated with septic arthritis, so the absolute WBC in synovial fluid is not, by itself, a reliable way to confirm or exclude a diagnosis of septic arthritis.

Differential Diagnosis

Symptoms of septic arthritis, such as acute joint pain, swelling, and even fever, with an increase in acute phase reactants, can be caused by crystalline arthritis (Chapter 273), especially pseudogout and gout, as well as psoriatic arthritis and reactive arthritis (Chapter 265). In patients who have preexisting inflammatory joint disease, such as rheumatoid arthritis, septic arthritis may be suspected if there is a sudden onset of acute or subacute monoarticular or pauciarticular joint swelling when the disease is otherwise well controlled. The presence or absence of fever is not a reliable indicator of an infected joint (Fig. 272-1).

TREATMENT

Rx

As soon as the joint has been aspirated and, ideally, after blood cultures have been obtained, prompt treatment with antibiotics must be instituted. Removal of purulent material and, where appropriate, débridement are essential. The choice of empirical antimicrobial therapy is based on which organisms are thought to be the likely cause of the septic arthritis and on the results of Gram stain and culture. No advantage of one antibiotic regimen over another has been demonstrated. If the initial Gram stain of the synovial fluid reveals gram-positive cocci, vancomycin is recommended, given the increasing frequency of infection due to MRSA and the need to initiate effective

antimicrobial therapy as soon as possible. Daptomycin and linezolid are alternative agents. If the initial Gram stain reveals gram-negative bacilli, an agent with broad coverage, including activity against *Pseudomonas aeruginosa,* is recommended. Such agents include ceftazidime, cefepime, imipenem, meropenem, piperacillin-tazobactam, and intravenous ciprofloxacin. If the Gram stain is negative, vancomycin alone in immunocompetent patients or in those unlikely to have gram-negative infection based on history and examination, or vancomycin plus one of the gram-negative antibacterials listed, is reasonable. Once culture and in vitro susceptibility results are available, therapy can be modified. The duration of antibacterial therapy usually ranges from 2 to 6 weeks.

The role of arthroscopic versus needle versus open drainage of the joint remains unsettled. Surgical management is appropriate for septic arthritis of the hip, for patients who fail to respond to serial needle aspiration and antibiotic therapy, and for patients who appear to be developing life-threatening complications such as necrotizing fasciitis. No studies have demonstrated the utility of lavage with or without synovectomy by arthroscopy versus arthrotomy or débridement. Patients should be mobilized as rapidly as possible to prevent joint contractures. In children, concomitant oral dexamethasone for 4 days can lead to more rapid symptomatic improvements.

PREVENTION

In patients requiring immunosuppression or glucocorticosteroid therapy to manage their underlying diseases, every effort should be made to use the lowest possible dose of these medications.

PROGNOSIS

Up to one third of patients with septic arthritis have a poor functional outcome, particularly older patients, patients with preexisting diseases of the joints such as osteoarthritis or rheumatoid arthritis, and patients with prosthetic joints. Poor joint outcome is associated with *S. aureus* infection in more than 50% of patients; mortality may be as high as 10 to 15%, particularly in patients who are immunocompromised or have polyarticular sepsis.

GONOCOCCAL SEPTIC ARTHRITIS

EPIDEMIOLOGY

Neisseria gonorrhoeae (Chapter 299) is a common cause of polyarthralgias and arthritis as well as oligoarticular arthritis and tenosynovitis in young, healthy patients. Disseminated gonococcal infection occurs in 0.5 to 3% of patients with gonorrhea (Chapter 299). Many of these patients have arthritis. Disseminated gonococcal infection and septic arthritis due to *N. gonorrhoeae* occur two to three times more often in women than in men. Most patients do not have a recent history of a symptomatic genital infection. The incidence of gonococcal arthritis is 133 cases per 100,000 population per year. Predisposing factors for disseminated gonococcal infection with arthritis include pregnancy, recent menstruation, complement deficiencies (C5, C6, C7, or C8), and systemic lupus erythematosus.

CLINICAL MANIFESTATIONS

Patients with gonococcal arthritis usually present with one of two clinical syndromes. The first is a purulent arthritis without skin lesions; the second is the triad of tenosynovitis, dermatitis, and polyarthralgias without purulent arthritis. The latter patients may have bacteremia and fever as well as maculopapular, vesicular, necrotic, pustular skin lesions anywhere on the integument. The arthritis is usually asymmetrical and may involve large or small joints, typically the elbows and knees or joints distal to these.

DIAGNOSIS

A high degree of clinical suspicion is required for diagnosis because many patients are asymptomatic for the primary infection. A thorough joint evaluation is important, as is an evaluation of soft tissues, particularly for tenosynovitis affecting the hands and feet. Cultures of blood, endocervix, and urethra are essential; cultures of the pharynx and rectum may be very helpful. *N. gonorrhoeae* is isolated in less than 30% of patients with the tenosynovitis-dermatitis syndrome and in about 50% of those with monoarthritis. PCR may be used to detect the gonococcal DNA in synovial fluid, skin lesions, urine, and throat samples, which are culture negative. Cultures should be submitted on Thayer-Martin media. Patients with suspected gonococcal arthritis should be screened for other coexisting sexually transmitted diseases (Chapter 285) such as syphilis, HIV, and chlamydia, as well as hepatitis B and C.

TREATMENT

Rx

Ceftriaxone is given for 2 to 4 days, followed by oral therapy to complete a minimum of 7 days of therapy, although up to 14 days of therapy is recommended. There is emerging resistance to fluoroquinolones, and unless specific in vitro susceptibility testing is available, their use is not recommended. Patients should also be treated for concomitant chlamydia with the regimens recommended by the Centers for Disease Control and Prevention (CDC). Most patients respond well to outpatient therapy, with complete resolution of the infection. Given emerging antimicrobial resistance, the most recent CDC guidelines for treating *N. gonorrhoeae* should be reviewed.

Viral Arthritis

Patients with viral syndromes may have polyarthralgias or inflammatory polyarthritis, which can mimic rheumatoid arthritis (Chapter 264). The most common viral infections associated with arthritis include hepatitis A, B, and C; cytomegalovirus; parvovirus B19; rubella; measles; and HIV. Other forms of viral arthritis are caused by adenovirus, echovirus, Epstein-Barr virus, and herpes zoster in North America and Europe; chikungunya and o'nyong-nyong viruses, especially in Africa; and Ross River virus in Australia. Chikungunya virus, which can cause severe bone and joint pain, has spread to the Caribbean basin and mainland United States.[4] Arthritis related to viral infections is likely principally reactive in nature, rather than being caused by direct synovial infection.

Other Forms of Infectious Arthritis

FUNGAL ARTHRITIS

Fungal arthritis is unusual and most commonly occurs in immunocompromised patients. Treatment with high doses of immunosuppressants, anti-TNF agents, and possibly other biologic response modifiers used in the treatment of rheumatoid arthritis and other autoimmune diseases may increase the risk for fungal infections. The infections are often systemic and may be indolent. An understanding of the epidemiology of the organisms as well as the patient's risk factors, including occupational and avocational risk factors, is essential to the diagnosis. The most common fungi in the United States include *Blastomyces dermatitidis*, *Coccidioides immitis,* and *Histoplasma capsulatum*. *Sporothrix schenckii* fungal infections may be seen, especially in gardeners. More unusual infections occur in immunocompromised patients, including *Aspergillus, Candida, Cryptococcus,* and *Nocardia.* The reader is referred to the specific chapters regarding these organisms for up-to-date antimicrobial recommendations.

LYME ARTHRITIS

Lyme disease (Chapter 321) usually causes oligoarticular arthritis, most commonly affecting the knee. Antibiotic treatment, as outlined in Chapter 321, is effective. Polyarticular disease affecting the small joints has been associated with HLA-DR4, which is found in greater frequency in patients with rheumatoid arthritis.

MYCOPLASMA ARTHRITIS

Mycoplasma hominis (Chapter 317) causes an oligoarticular or monoarticular arthritis. Risk factors include an immunocompromised state and hypogammaglobulinemia. The treatment of choice is tetracyclines, usually doxycycline; alternatively, clindamycin or fluoroquinolones can be used.

TUBERCULOUS ARTHRITIS

Most cases of tuberculosis (TB; Chapter 324) in Canada, the United States, Western Europe, Australia, and New Zealand occur in immigrants. The arthritis is usually monoarticular or oligoarticular, affecting larger joints, and TB should be suspected in patients who have refractory monoarticular or pauciarticular arthritis thought to be secondary to another bacterial infection or to a systemic inflammatory disease such as rheumatoid arthritis. TB screening is mandatory for all patients before beginning treatment with immunosuppressive drugs or biologic response modifiers.

The diagnosis of TB may be delayed because of a lack of suspicion because patients may not have pulmonary disease. Atypical mycobacterial infection may occur in fishermen and immunocompromised patients. Appropriate treatment for septic arthritis due to TB is based on guidelines and in vitro susceptibility testing, but it often includes isoniazid, ethambutol, or rifampin

and pyrazinamide as empirical therapy (Chapter 324). Atypical mycobacteria are often not susceptible to traditional antituberculous agents, and infectious disease consultation is recommended. Patients with a history of TB in whom anti-TNF therapies are being considered should be appropriately treated for TB before these drugs are started. Patients with a positive purified protein derivative (PPD) test or QuantIFERON assay for TB without a history of diagnosed tuberculosis should be treated prophylactically for several months before starting anti-TNF therapy. Clinical suspicion and culture of synovial fluid or synovial membrane obtained at biopsy are essential to the diagnosis.

SYPHILIS

Musculoskeletal involvement by syphilis (Chapter 319) is manifold in its manifestations and includes monoarticular or oligoarticular arthritis, polyarthralgias, tenosynovitis, sacroiliitis, spondylitis, chondritis, osteitis, and periostitis. Charcot's joints, osteitis, and chronic arthritis are typical of tertiary syphilis. Most patients with syphilis-related arthritis can be treated successfully. Arthritis may complicate congenital, secondary, and tertiary syphilis.

PROSTHETIC JOINT INFECTION

More than 1 million joint replacements (Chapter 276) are done each year in the United States, and this number continues to increase. Infection occurs in 0.3 to 1.7% of hip arthroplasties and 0.8 to 1.9% of knee arthroplasties, and the infection risk is two- to three-fold higher in patients with rheumatoid arthritis. Prosthetic joint infections are classified as (1) early infections, occurring within 3 months of joint replacement; (2) delayed infections, occurring 3 months to about 1 year after joint replacement; and (3) late infections, occurring more than 1 to 2 years after joint replacement.[5-7] Infections occurring within the first year are usually related to the implantation surgery itself, and late infections are usually due to hematogenous spread.

The development of a bacterial biofilm on the prosthetic joint is characteristic of prosthetic joint infection, increasing the susceptibility to infection in experimental animal models with as few as 100 colony-forming units. These biofilms are formed by bacterial glycocalyx, which increases the organisms' resistance to antimicrobial agents and likely accounts for the difficulty in obtaining viable organisms from the infected joint. More than half of all prosthetic joint infections of hips and knees are caused by staphylococci. Other organisms, including gram-negative bacilli, anaerobes, and *Candida* species, may also cause infection. In particular, *Propionibacterium* species are associated with infected shoulder arthroplasties. About 20% of cases are polymicrobial, and in 7%, cultures are negative.

Risk factors associated with the development of prosthetic joint infection include wound healing complications, prior superficial surgical site infection, prior infection of the joint, previous surgery on the joint, rheumatoid arthritis, advanced age, obesity, smoking, cancer, and diabetes mellitus. Other risk factors include simultaneous bilateral arthroplasty, prolonged operative time, requirement for blood transfusion, and infection occurring elsewhere in the body that can hematogenously seed the prosthesis.

Patients with early-onset infection may have classic symptoms and signs of septic arthritis, including joint pain, effusion, erythema, and fever. Patients with delayed infection may have only joint pain, with or without implant loosening, requiring a high degree of suspicion for the presence of infection.

The definitive diagnosis of prosthetic joint infection is based on the recovery of organisms from multiple specimens of synovial fluid and periprosthetic tissue, sonication of the prosthesis itself, acute inflammation suggestive of infection on pathologic examination of periprosthetic tissue obtained at surgery, or the presence of a sinus track communicating with the prosthesis, even in the absence of microorganisms.

An elevated sedimentation rate or C-reactive protein without another obvious cause, such as inflammatory arthritis or recent surgery, is very suggestive of infection in a patient with a painful, loose prosthesis. Plain radiographs may show loosening, new bone formation, and lucencies along the implant margin but are often nonspecific. Technetium-based scintigraphy combined with indium-labeled white blood cell scanning is suggestive of established infection but is often not performed owing to the expense. MRI and CT have low utility in diagnosing prosthetic joint infection.

Surgical treatment of prosthetic joint infection typically consists of débridement with retention of the prosthesis for acute infection, resection arthroplasty with or without staged reimplantation for chronic infection, or amputation in a few limited instances. Systemic antibiotic therapy in a patient with a prosthetic joint infection is pathogen directed and driven by the surgical therapy used to manage the infection. Following an attempt at salvage of the prosthesis with 2 to 6 weeks of effective intravenous antibiotic therapy, chronic suppressive therapy with oral antimicrobial agents is often used. In rifampin-susceptible staphylococcal infections, its addition to a companion intravenous or oral antimicrobial is recommended to avoid the emergence of resistance, to treat biofilm organisms, and to improve the chance of salvaging the prosthesis. After resection arthroplasty, 4 to 6 weeks of pathogen-directed intravenous therapy is typical before an attempt at reimplantation several weeks later. The use of depot local antimicrobial therapy with antibiotic-impregnated polymethylmethacrylate spacers following resection arthroplasty is also very common. The reader is referred to recently released guidelines for more specific information on the diagnosis and management of prosthetic joint infection (Infectious Diseases Society of America guidelines).

PREVENTION

In addition to optimizing comorbidities such as diabetes mellitus and discontinuing tobacco use preoperatively, careful screening for infection, including asymptomatic urinary tract infection, is prudent when considering joint surgery. Perioperative antibiotic treatment with cephalosporin in patients undergoing prosthetic joint replacement reduces the risk for infection by approximately three-fold. Antimicrobial therapy should be given within 60 minutes of the initial incision, ideally before tourniquet application. Cefazolin or cefuroxime can be given to patients with normal renal function, and vancomycin can be used in patients who are allergic to penicillin.

Whether an antirheumatic agent should be stopped before joint replacement surgery is unclear. The usual practice is to hold drugs such as methotrexate, anti-TNF agents, and other biologics including abatacept and tocilizumab for one to four half-lives before and after surgery. No guidelines exist for tofacitinib or anakinra, but similar suggestions may be helpful. It may be prudent to wait at least 8 to 12 weeks after rituximab therapy for rheumatoid arthritis before performing elective joint replacement; it is unclear whether B-cell reconstitution plays a role in surgical infection risk. The impact of these strategies on reducing prosthetic joint infections is unknown.

OSTEOMYELITIS

DEFINITION

Osteomyelitis is a bacterial infection of bone that causes destruction and can occur through a variety of mechanisms.[8,9]

EPIDEMIOLOGY

Osteomyelitis of the bones of the foot in adult patients with diabetes, neuropathy, and arterial insufficiency is very common. Management of osteomyelitis in diabetic feet is discussed in updates for Chapter 229. Hematogenous seeding of the spine also occurs but less frequently. The incidence of osteomyelitis due to trauma and surgery is increasing.

PATHOBIOLOGY

Osteomyelitis can develop from (1) hematogenous seeding from a distant infection, (2) contiguous spread from nearby skin and joints, and (3) penetration of microorganisms into bone at the time of trauma or surgery. Unless there is trauma or the presence of a foreign body, bone is typically very resistant to infection. Organisms such as *S. aureus* cause disease more frequently because they colonize the skin in up to 30 to 40% of individuals, frequently cause cellulitis and bacteremia, and have the ability to bind to bone through the expression of receptors for fibronectin and collagen.

Hematogenous causes of osteomyelitis typically present in elderly people; it usually involves two or more vertebrae and their intervening disc spaces. Bacteria gain access to these structures through the arterial and venous systems (Batson's venous plexus). Bacteremia from any source can cause osteomyelitis of the spine, but cellulitis, urinary tract infection, and pneumonia are the most common sources.

Contiguous focus osteomyelitis is common in adults, typically occurring in elderly people. It results from the spread of infection from nearby skin, often in the feet of patients with diabetes, neuropathy, or vascular insufficiency, or in pelvic bones in patients with decubitus due to impaired sensation from spinal cord injury or disease. Alternatively, it can occur at the time of orthopedic surgery, from contamination at the time of an open fracture, or from a human or animal bite.

Acute osteomyelitis has a duration of less than 10 days, whereas chronic infection has a duration of more than 10 days. *S. aureus* is the most common cause of hematogenous and contiguous osteomyelitis in adults.

Osteomyelitis due to β-hemolytic streptococci and aerobic gram-negative bacilli is much less common, but it can occur if infections due to these organisms result in hematogenous seeding, if nosocomial contiguous osteomyelitis occurs due to surgical site infection, or if contamination occurs at the time of traumatic open fracture. Polymicrobial infection, including infection due to anaerobes, is very common in osteomyelitis of the bones of the feet associated with diabetes and vascular insufficiency. Coagulase-negative staphylococci can be pathogenic in patients with orthopedic implants.

CLINICAL MANIFESTATIONS

Localized pain over the affected bones is a hallmark of osteomyelitis. A sinus tract or swelling and erythema due to concomitant soft tissue infection or abscess may be present in osteomyelitis due to contiguous infection. Constitutional symptoms, including fever, are present in the minority of cases and more often in hematogenous osteomyelitis. If neurologic structures are involved, neurologic signs and symptoms may be present. Signs and symptoms due to a coexisting infection that has caused hematogenous osteomyelitis may be present as well. The differential diagnosis of osteomyelitis includes diseases that can cause acute and chronic bone pain in adults, including osteoarthritis, metastatic malignancy, fractures, and SAPHO (synovitis, acne, pustulosis, hyperostosis, and osteitis) syndrome, as well as postoperative pain and soft tissue infection without concomitant osteomyelitis.

DIAGNOSIS

The WBC is often elevated in hematogenous and acute osteomyelitis. Serum inflammatory markers such as the sedimentation rate and C-reactive protein are often abnormal, particularly in cases of hematogenous infection, but may be normal in chronic contiguous osteomyelitis. Blood cultures are positive in 25 to 50% of cases of hematogenous infection but are almost always negative in chronic osteomyelitis unless concomitant soft tissue infection is present. In the setting of chronic contiguous osteomyelitis, plain radiographs often show specific abnormalities; in vertebral osteomyelitis, they are often not helpful in confirming the diagnosis of infection.

The ability to percutaneously probe or palpate bone with a probe is a simple, effective diagnostic test in patients with diabetes mellitus and possible contiguous osteomyelitis of the feet. In one prospective study, the ability to palpate bone through a contiguous ulcer had a sensitivity of 66% and a specificity of 85% for osteomyelitis.

MRI is the most sensitive and diagnostic imaging technique to identify osteomyelitis, except when orthopedic implants are present. Gallium scans are more sensitive and specific than three-phase technetium (^{99m}Tc) bone scans or indium-labeled leukocyte scans for the diagnosis of vertebral osteomyelitis. Gallium scans are used when spinal hardware is present that degrades the magnetic resonance images and in cases of skull bone osteomyelitis due to malignant external otitis.

Multiple specimens of involved bone, contiguous soft tissue, and purulence should be sent for Gram stain, aerobic and anaerobic culture, and pathologic examination at the time of bone biopsy or surgical débridement. If the history, examination, or imaging is suggestive of atypical infection, culture for fungi and mycobacteria or other unusual organisms should be performed.

TREATMENT

Rx

There are no large randomized studies comparing antimicrobial therapy for osteomyelitis.[A1] Antimicrobials for specific pathogens based on in vitro susceptibility testing are recommended, and examples of antimicrobials used to treat common pathogens causing osteomyelitis and septic arthritis are shown in Table 272-2. Oral antibiotics can achieve adequate levels in bone, and oral and parenteral therapies may achieve similar cure rates in some cases.

The duration of antimicrobial therapy is almost always dictated by surgical therapy in chronic osteomyelitis. For example, if an amputation is performed, a short course of antimicrobial therapy may be required, whereas if an extensive débridement of chronic osteomyelitis is performed, prolonged intravenous antimicrobial therapy for 4 to 6 weeks is typically recommended. If the surgical therapy could lead to a worse outcome than no surgery, chronic oral antimicrobial suppression may be recommended. Acute hematogenous vertebral osteomyelitis in adults is typically treated with 6 weeks of intravenous antimicrobial therapy, without surgical intervention, after identification of the pathogen through percutaneous or open biopsy. Hyperbaric oxygen therapy for chronic osteomyelitis is controversial.

TABLE 272-2 ANTIMICROBIAL THERAPY FOR SELECTED MICROORGANISMS IN OSTEOMYELITIS OR SEPTIC ARTHRITIS IN ADULTS

MICROORGANISM	FIRST CHOICE*	ALTERNATIVE CHOICE
Methicillin/oxacillin/nafcillin-sensitive staphylococci	Nafcillin sodium or oxacillin sodium 1.5-2 g IV q4-6h for 4-6 wk *or* cefazolin 1-2 g IV q8h	Vancomycin 15 mg/kg IV q12h for 4-6 wk
Methicillin/oxacillin/nafcillin-resistant staphylococci (MRSA)	Vancomycin† 15 mg/kg IV q12h *or* daptomycin 6 mg/kg IV q24h	Linezolid 600 mg PO/IV q12h *or* levofloxacin† 500-750 mg PO/IV daily
Penicillin-sensitive streptococci	Aqueous penicillin G 20 × 10^6 U/24 hr IV either continuously or in 6 equally divided daily doses *or* ceftriaxone 1-2 g IV q24h *or* cefazolin 1-2 g IV q8h	Vancomycin 15 mg/kg IV q12h
Enterococci	Aqueous crystalline penicillin G 20 × 10^6 U/24 hr IV either continuously or in 6 equally divided daily doses *or* ampicillin sodium 12 g/24 hr IV either continuously or in 6 equally divided daily doses; the addition of gentamicin sulfate 1 mg/kg IV or IM q8h for 1-2 wk is *optional*	Vancomycin† 15 mg/kg IV q12h; the addition of gentamicin sulfate 1 mg/kg IV or IM q8h for 1-2 wk is *optional*
Enterobacteriaceae	Ceftriaxone 2 g IV q24h	Ciprofloxacin† 500-750 mg PO q12h
Pseudomonas aeruginosa	Cefepime 2 g IV q8-12h	Ciprofloxacin† 750 mg PO q12h or ceftazidime 2 g IV q8h

*Antimicrobial selection should be based on in vitro sensitivity data, as well as allergies, intolerances, and drug interactions in individual patients.
†Doses shown are based on normal renal and hepatic function and may need to be adjusted or serum levels monitored (vancomycin).
MRSA = methicillin-resistant *Staphylococcus aureus.*
Adapted from Berbari EF, Steckelberg JM, Osmon DR. Osteomyelitis. In: Mandell GL, Bennett JE, Dolin R, eds. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases,* 7th ed. Philadelphia: Churchill Livingstone/Elsevier; 2010:1457-1467.

PREVENTION

Improving the control of diabetes and decreasing the incidence of peripheral vascular disease will reduce the incidence of diabetic foot bone infection. Optimal strategies to prevent surgical site infection after orthopedic procedures will prevent orthopedic implant infection after surgery and open fractures.

PROGNOSIS

The success of osteomyelitis management depends on the medical and surgical therapy employed and the ability to improve comorbidities such as arterial insufficiency. The ability of orthopedic surgeons to perform more extensive reconstructive surgery has allowed more extensive débridement and higher success rates, as well as restoration of function. Treatment failure can lead to relapse of infection or progression of infection to involve more of the affected bone. Long-standing osteomyelitis can be complicated by amyloidosis, squamous cell carcinoma of the skin in a chronic sinus tract, or primary bone malignancy.

Grade A References

A1. American Academy of Orthopedic Surgeons (AAOS). Advisory statement. Recommendations for the intravenous antibiotic prophylaxis in primary total joint arthroplasty. 2004. Retrieved October 6, 2014, from http://www.aaos.org/about/papers/advistmt/1027.asp.
A2. Conterno LO, da Silva Filho CR. Antibiotics for treating chronic osteomyelitis in adults. *Cochrane Database Syst Rev.* 2013;9:CD004439.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Amin AM, Cerceo EA, Deitelzweig SB, et al. Hospitalist perspective on the treatment of skin and soft tissue infections. *Mayo Clin Proc*. 2014;89:1436-1451.
2. Liu C, Bayer A, Cosgrove SE, et al. Clinical practice guidelines by the Infectious Diseases Society of America for the treatment of methicillin-resistant *Staphylococcus aureus* infections in adults and children. *Clin Infect Dis*. 2011;52:e18.
3. Matthews CJ, Weston VC, Jones A, et al. Bacterial septic arthritis in adults. *Lancet*. 2010; 375:846-855.
4. Morens DM, Fauchi AS. Chikungunya at the door—deja vu all over again? *N Engl J Med*. 2014; 371:885-887.
5. Yin JM, Liu ZT, Zhao SC, et al. Diagnosis, management and prevention of prosthetic joint infections. *Front Biosci (Landmark Ed)*. 2013;18:1349-1357.
6. Johannsson B, Taylor J, Clark CR, et al. Treatment approaches to prosthetic joint infections: results of an Emerging Infections Network survey. *Diagn Microbiol Infect Dis*. 2010;66:16-23.
7. Osmon DR, Berbari EF, Berendt AR, et al. Infectious Diseases Society of America. Diagnosis and management of prosthetic joint infection: clinical practice guidelines by the Infectious Diseases Society of America. *Clin Infect Dis*. 2013;56:e1-e25.
8. Zimmerli W. Vertebral osteomyelitis. *N Engl J Med*. 2010;362:1022-1029.
9. Spellberg B, Lipsky BA. Systemic antibiotic therapy for chronic osteomyelitis in adults. *Clin Infect Dis*. 2012;54:393-407.

REVIEW QUESTIONS

1. A 55-year-old man with a history of diabetes has had a chronic ulcer over the fifth metatarsal head with some swelling and pain for the past 5 weeks, following what he though was a minor trauma. On examination, bone can be palpated with a probe. Which of the following is most likely?

A. The white blood cell count is elevated.
B. There is a high probability that the patient has osteomyelitis.
C. The best imaging test to define the cause of the pain is a gallium scan.
D. The best imaging test to define the cause if the pain is plain bone radiography.
E. The most likely cause of a possible infection is *Fusobacterium* species infection.

Answer: B The ability to palpate bone by probing the overlying ulcer is highly suggestive of osteomyelitis. Patients with contiguous chronic osteomyelitis often have normal white blood cell counts and are often afebrile. Magnetic resonance imaging is the best (most sensitive) radiographic technique for demonstrating bone damage consistent with osteomyelitis. Overall, *Staphylococcus aureus* is the most common cause of osteomyelitis, but in patients with diabetic foot bone osteomyelitis, the infection is often polymicrobial (see Table 280-1).

2. A 32-year-old carpet layer presents with a 3-day history of swelling of the right knee. The father and grandfather of the patient developed gout as young men. The patient had a single episode of right toe pain and swelling 3 months earlier, which he attributed to "kicking carpet into place." He has a temperature of 38.7° C. Examination of the right knee reveals erythema and swelling, most prominently over the anterior aspect of the knee, but extending around the soft tissues about the knee. With some pain, he is able to completely extend the knee, although he cannot comfortably flex it beyond 80 degrees. The prepatellar bursa is aspirated. The white blood cell count is 750. Which of the following is correct?

A. The knee joint should also be aspirated and joint fluid sent for culture and sensitivity.
B. The low white blood cell count in the bursal fluid makes it unlikely that the bursa is infected.
C. Gram stain examination of the bursal fluid will almost always guide therapy.
D. Urate crystals will not be found on microscopic examination.
E. A magnetic resonance imaging (MRI) study must be obtained to rule out osteomyelitis.

Answer: B This patient has prepatellar bursitis, which is usually due to direct tissue invasion, in this case likely related to local tissue invasion in the setting of chronic soft tissue trauma associated with kneeling and local friction of the overlying skin. Bursal fluid is often unobtainable, but if it can be obtained, the white blood cell count in the aspirated bursal fluid in the setting of septic bursitis is generally low compared with the WBC count in septic arthritis, on average 13,500 cells in the bursal fluid. Gram stain is positive in only about 15% of cases of septic bursitis. Urate crystals may be seen in cases of gouty bursitis, which can occur concomitantly, but should not distract from the possibility of septic bursitis. MRI is the most sensitive imaging technique to evaluate osteomyelitis in routine clinical practice, but there is nothing in this case to cause suspicion of bone infection.

3. Which of the following is true regarding septic arthritis?

A. Most cases are due to direct inoculation of the joint from trauma or surgical procedures such as arthrocentesis.
B. Gonococcal arthritis is the most common form of septic arthritis in persons aged 20 to 40 years.
C. Hematogenous spread of tuberculosis is the most common cause of septic arthritis in persons from developing countries immigrating to Western countries of Europe, North American, and Australia.
D. Blood cultures are frequently positive in septic arthritis.
E. Intravenous vancomycin and daptomycin are the mainstays of treatment for most cases of septic arthritis in the immunocompetent host in the current era of multiple antibiotic resistance.

Answer: D About 75% of cases of septic arthritis are due to hematogenous spread, including surgical instrumentation. Although gonococcal arthritis must be considered in sexually active patients, 50 to 70% of cases of septic arthritis are due to *Staphylococcus aureus.* Although almost all cases of septic tuberculous arthritis in Western countries occur in immigrants from developing countries, most cases of septic arthritis in these patients are still due to nontuberculous bacteria. Blood cultures should always be obtained in persons with septic arthritis. They are positive in up to 50% of cases, even in cases when the joint aspirate fails to microscopically reveal the causative organism. Although multiple antimicrobial resistance is a serious problem, most cases of uncomplicated septic arthritis in the otherwise healthy host can be managed with an oral antistaphylococcal penicillin or first-generation cephalosporin.

4. A 76-year-old man with a history of osteoarthritis arthritis and total joint replacement of the right hip at the age of 70 years presents with a 2-day history of right groin pain and temperature to 39.5° C 3 days after finishing a 1-week course of levofloxacin for an infected ingrown toenail. He is allergic to penicillins. Empirical antibiotic therapy with vancomycin for presumed staphylococcal prosthetic joint infection is started. The synovial fluid aspirate from the prosthetic right hip is negative, but blood cultures reveal *S. aureus*, methicillin resistant. His prosthesis was not loose on plain-film examination. Which of the following is true regarding prosthetic infections in this patient?

A. A biofilm with polymicrobial bacterial colony-forming units is frequently present, complicating eradication of the organism. Explantation of the prosthesis is always needed in such cases.
B. The most appropriate antimicrobial therapy in this case would be linezolid 600 mg PO/IV every 12 hours.
C. Risk factors for prosthetic joint infection *include* wound healing complications, prior superficial surgical site infection, prior infection of the joint, osteoarthritis, and diabetes mellitus.
D. Because of the occurrence of prosthetic infection in this patient following hematogenous spread of the causative bacterium, the best course of action would be explantation of the prosthesis at this time, followed by 4 to 6 weeks of antibiotic therapy without reimplantation.
E. The best course of action would be to attempt salvage of the prosthesis with débridement and retention and 2 to 6 weeks of effective intravenous antibiotic therapy and rifampin based on the recently published Infectious Diseases Society of America Prosthetic Joint Infection guidelines (Osmon DR. Berbari EF, Berendt AR. et al. Infectious Diseases Society of America. Diagnosis and management of prosthetic joint infection: clinical practice guidelines by the Infectious Diseases Society of America. *Clin Infect Dis.* 2013;56:e1-e25), followed by chronic suppressive therapy with oral antimicrobial agents.

Answer: E Biofilms are often found in explanted infected prosthesis and complicate successful antimicrobial therapy. These biofilms are polymicrobial in about 20% of cases. Risk factors for prosthesis infection include rheumatoid arthritis, but not osteoarthritis. In most cases of acute infection, management efforts are directed toward treating the infection and attempting to salvage the prosthesis. When resection arthroplasty is needed, 4 to 6 weeks of pathogen-directed intravenous therapy is typical before an attempt at reimplantation several weeks later.

5. A 68-year-old woman with a history of chronic atrial fibrillation and diabetic neuropathy presents with a 5-day history of subacute swelling and pain in the left ankle. She reports that she "turned" her ankle while gardening about a week ago. Her temperature is 36.7° C. Inspection of the ankle reveals moderate erythema and swelling. The skin is intact; sensation to touch is diminished in both feet. The remaining history and physical examination are unrevealing, including examination of the oral cavity. The sedimentation rate is 42 mm/hour; white blood cell count is 9800 mm^3, and INR is 3.2. Which of the following strategies should be pursued in management of this patient?

A. Blood cultures and empirical therapy with vancomycin 15 mg/kg every 12 hours or daptomycin 6 mg/kg/IV every 24 hours for presumed methicillin-resistant *Staphylococcus aureus* (MRSA). Avoid arthrocentesis because of the elevated INR and risk for worsening joint damage due to hemarthrosis and the danger of hematogenous spread of the infection.

B. Blood cultures, empirical therapy with vancomycin 15 mg/kg every 12 hours or daptomycin 6 mg/kg/IV every 24 hours for presumed MRSA, and arthrocentesis for examination of synovial fluid Gram stain and culture

C. Blood cultures, empirical therapy with aqueous crystalline penicillin G 20×10^6 U every 24 hours IV either continuously or in six equally divided daily doses *or* ampicillin sodium 12 g every 24 hours IV either continuously or in six equally divided daily doses. Avoid arthrocentesis because of the elevated INR and risk for worsening joint damage due to hemarthrosis and the danger of hematogenous spread of the infection.

D. Blood cultures, empirical therapy with aqueous crystalline penicillin G 20×10^6 U every 24 hours IV either continuously or in six equally divided daily doses *or* ampicillin sodium 12 g every 24 hours IV either continuously or in six equally divided daily doses, and arthrocentesis for examination of synovial fluid Gram stain and culture

Answer: B A source of infection was not identified in this patient, who has risk factors for septic arthritis of diabetes mellitus and diabetic neuropathy. Because even in these cases the most common cause of septic arthritis is gram-positive cocci, particularly *Staphylococcus aureus*, initial treatment directed at this organism is appropriate. It is not unreasonable to initiate therapy with an MRSA-appropriate regimen. Arthrocentesis should always be performed if there is suspicion of septic arthritis; bleeding complications, even in patients on anticoagulation, are rare, and the information potentially gained generally justifies the risk for the procedure.

273 CRYSTAL DEPOSITION DISEASES

N. LAWRENCE EDWARDS

The destructive potential of intrasynovial crystals has been recognized for more than a century. The mechanisms by which certain crystals induce inflammation and joint destruction have become much better clarified over the past several decades. The three most common crystal-induced arthropathies are caused by precipitation of monosodium urate monohydrate (MSU), calcium pyrophosphate dehydrate (CPPD), and basic calcium phosphate (BCP) and are termed *gout, CPPD arthropathy,* and *basic calcium arthropathy*, respectively.

At the turn of the 20th century, Wilhelm His and Max Freudweiler at the University of Leipzig proved that MSU crystals could induce inflammation when injected into normal joints. In 1961, Daniel McCarty and Joseph Hollander developed a way of identifying MSU crystals in synovial fluid using compensated polarized light microscopy. This technique also allowed for distinguishing MSU crystals from the other "phlogystic" crystals such as CPPD and BCP.

The role of calcium-containing crystals in cartilage pathology and joint disease was first suggested in serial radiographs by Zitnan and Sitaj that showed accretion of calcific deposits in articular cartilage leading to accelerated joint destruction. The nature and structure of CPPD crystals in inflamed synovial fluid was first described by McCarty and collaborators in 1962. This group also established the term *pseudogout* to describe this common crystalline arthropathy.

Basic calcium crystals are ultra-microscopic in size and are not detected by the compensated polarized microscopy used to identify MSU and CPPD crystals. Paul Dieppe and Ralph Schumacher independently identified these very small crystals in synovial fluid using electron microscopy in the mid-1970s. Like MSU and CPPD crystals, BCP crystals are biologically active and can accelerate atrophic changes in bone and cartilage. This chapter defines these separate crystalline arthropathies and describes their different pathogeneses and treatments.

GOUT AND HYPERURICEMIA

DEFINITION

Gout is an inflammatory arthritis of metabolic origin that is caused by crystallization of MSU crystals within joints. It is the most common inflammatory joint disease in men and in older women. Its incidence and prevalence are increasing worldwide. The metabolic derangement responsible for gout is the supersaturation of blood and body fluids with the urate ion to the point that crystal formation is possible. At physiologic pH and at normal body temperature, urate is considered to be supersaturated at concentrations of 6.8 mg/dL or greater. Therefore, from a biologic perspective, hyperuricemia is any serum urate level greater than 6.8 mg/dL in both men and women. Although hyperuricemia is a necessary prerequisite for developing gout, only 20% of all hyperuricemic subjects will ultimately develop gout.

The natural course of gout progresses from an intermittent monoarthritis or oligoarthritis in the lower extremities to a chronic, destructive, and debilitating polyarthritis involving almost any peripheral joint in the body. Also considered within the definition of gout is the spectrum of clinical conditions resulting from the deposition of MSU crystals in the subcutaneous tissues and kidney manifesting as tophaceous deposits, inflammatory cellulitis, urate nephropathy, and kidney stones.

EPIDEMIOLOGY

Hyperuricemia is very common in Western cultures, with prevalence as high as 15 to 20% being reported in some more recent population-based studies.[1] This rate of hyperuricemia is more than twice that observed just three decades earlier. A number of factors have been proposed to explain this dramatic rise. These include the overall increase in longevity; the increased prevalence of hypertension, metabolic syndrome, and obesity; the ubiquitous use of thiazide diuretics and low-dose aspirin; changes in dietary trends, including the greater use of high-fructose corn syrup as a sweetener; and finally, the increase in survival of patients end-stage renal disease and organ transplantation.

There is a direct correlation between the degree of serum urate elevation and the likelihood of developing gout. The reported annual incidence of gout in subjects with baseline serum urate levels greater than or equal to 9 mg/dL is 4.9%, compared with only 0.5% in people with serum urate levels of 7.0 to 8.9 mg/dL.

Epidemiologic surveys since the 1970s demonstrate a continuous increase in the prevalence of gout. According to the most recent National Health and Nutrition Examination Survey (NHANES 2007-2009), 8.3 million adults in the United States are followed by a health care provider for gout, a prevalence of 3.9%. Men continue to make up the majority of the gout population (73%). The prevalence of gout in women has been rising at a disproportionately higher rate.

PATHOBIOLOGY

Uric acid is the end product of purine metabolism in humans. In most mammals, purine catabolism is taken one step further through the enzyme uric acid oxidase or uricase, with the purine end product in these species being the very soluble allantoin. Humans and most other hominoids lost the ability to produce the enzyme uricase nearly 18 million years ago. As a result, uric acid accumulation is possible. Whether caused by overproduction of uric acid or its underexcretion by the kidneys, this accumulation leads to supersaturation of urate ion in blood and the precipitation of MSU crystals in synovial fluid, soft tissues, and organs. Urate is produced by the conversion of a very soluble molecule, hypoxanthine, to the less soluble xanthine, which, in turn, is converted to the very insoluble uric acid by progressive purine ring oxidations catalyzed by the enzyme xanthine oxidase. Xanthine oxidase is present in several organs, but most activity in the body is found in the liver and intestines. Because of its potential for causing disease, urate elimination is very important. The total daily accumulation of uric acid from de novo synthesis, nucleotide degradation, and dietary consumption is between 800 and 1200 mg/day and is balanced by renal excretion of approximately two thirds of the total amount and intestinal elimination by the remaining one third.

Simply put, hyperuricemia occurs when urate production is not balanced by renal excretion. In 90% of all gout patients, the cause of this imbalance is renal underexcretion. The remaining 10% of gout cases are caused by purine overproduction or a combination of overproduction and underexcretion. The nongenetic causes of hyperuricemia include other medical conditions, dietary components, and medications (Table 273-1). These factors may result in either overproduction or diminished renal clearance of uric acid. Similarly, the genetic causes of hyperuricemia (Table 273-2) may affect either production or elimination of uric acid.[2]

TABLE 273-1 NONGENETIC CAUSES OF HYPERURICEMIA

IMPAIRED URIC ACID EXCRETION
Clinical Conditions
Reduced glomerular filtration rate
Hypertension
Obesity
Systemic sclerosis
Lead nephropathy
Drugs
Diuretics
Ethanol
Low-dose salicylates (0.06-3.0 g/day)
Cyclosporine
Tacrolimus
Levodopa
EXCESSIVE URIC ACID PRODUCTION
Clinical Conditions
Myeloproliferative and lymphoproliferative neoplasms
Obesity
Psoriasis
Diet and Drugs
Alcoholic beverages (especially beer)
Red meat, organ meat, shellfish
High fructose corn syrup
Cytotoxic drugs
Nicotinic acid
Pancreatic extract

Renal Urate Underexcretion

Because uric acid is small and not protein bound, it is completely filtered by the glomerulus. In normal persons, approximately 8 to 10% of the filtered load is ultimately cleared in the urine. The various renal tubular transporters that are responsible for determining how much of the filtered uric acid is actually excreted are located in the proximal convoluted tubules and are referred to collectively as the *transportasome* (Fig. 273-1). Both reabsorption and secretion occur in this segment through the actions of several organic acid transporters (OATs), with the net effect being the reabsorption of nearly 90% of the uric acid filtered at the glomerulus. These OATs are also responsible for eliminating organic acids other than uric acid as well as many commonly used medications. The most important tubular transporter of uric acid is URAT1. This transporter swaps urate ions for other monocarboxylate organic ions in both directions across the luminal membrane of proximal tubular cells. This system can be driven to reabsorb more uric acid from the tubular lumen by raising tubular epithelial concentrations of lactate, pyruvate, or the ketoacids acetoacetate and β-hydroxybutyrate. Certain drugs, when present in the tubular lumen, can displace uric acid from the transporter, causing more uric acid to be lost in the urine. These compounds include probenecid, losartan, and high-dose aspirin and are considered uricosuric.

When renal clearance of uric acid is compared between normal adult men and gouty men, the gouty subjects excrete only 70% as much uric acid as normal individuals at any given serum urate concentration. In general, gouty subjects required a serum urate concentration to be 1.7 mg/dL higher to obtain the same level of excretion as seen in normal subjects.

Most genetic polymorphisms associated with gout in genome-wide association studies encode for the various components of the uric acid transportasome. Polymorphisms in the glucose transporter GLUT-9 (encoded by the *SLC2A9* gene) are statistically the most significant determinants of serum urate. ABCG2 is a multifunctional transporter that belongs to the adenosine triphosphate (ATP)-binding cassette family found in the proximal tubule of the kidney as well as in the small intestine and liver. Polymorphisms in the *ABCG2* gene have a larger effect on urate levels in men than in women. Polymorphisms in the gene encoding URAT1 may lead to either hypouricemia or hyperuricemia. A loss-of-function mutation results in familial renal hypouricemia. Other mutations have been detected that lead to hyperuricemia and gout in Mexican, Asian, and African American populations.

The serum urate variance explained by these common genetic variants is only about 6% of the total variance observed between gouty and nongouty subjects. Similar risk-stratifying techniques demonstrate that 67% of the variance is caused by nongenetic factors such as serum creatinine, ethanol consumption, and the components of the metabolic syndrome.

TABLE 273-2 GENETIC CAUSES OF HYPERURICEMIA

SYNDROME	PHENOTYPE
INBORN ERRORS OF PURINE METABOLISM	
Hypoxanthine-guanine phosphoribosyl transferase deficiency	Neurologic dysfunction, renal stones, early-onset gout
Phosphoribosyl pyrophosphatase synthetase overactivity	Neurologic dysfunction, early-onset gout
EXCESSIVE CELL DEATH AND URATE GENERATION	
Glycogen storage disease I	Growth restriction, lactic acidosis, early-onset gout
Glycogen storage disease III	Early-onset gout
Glycogen storage disease V	Early-onset gout
Glycogen storage disease VII	Early-onset gout
Fructose-1-phosphate aldolase deficiency	Growth restriction, liver failure, early-onset gout
Myoadenylate deaminase deficiency	Myopathy, gout
Carnitine palmitoyltransferase II deficiency (late onset)	Rhabdomyolysis, gout
REDUCED RENAL EXCRETION OF URIC ACID	
Medullary cystic kidney disease	Renal dysfunction, early-onset gout
Familial juvenile hyperuricemic nephropathy	Renal dysfunction, early-onset gout
Uric acid transportasome mutations	
GLUT-9	Familial gout
ABCG2	Familial gout
URAT1	Familial gout

Urate Overproduction

In approximately 10% of gouty subjects, hyperuricemia is caused by uric acid overproduction rather than reduced renal excretion. In most of these people,

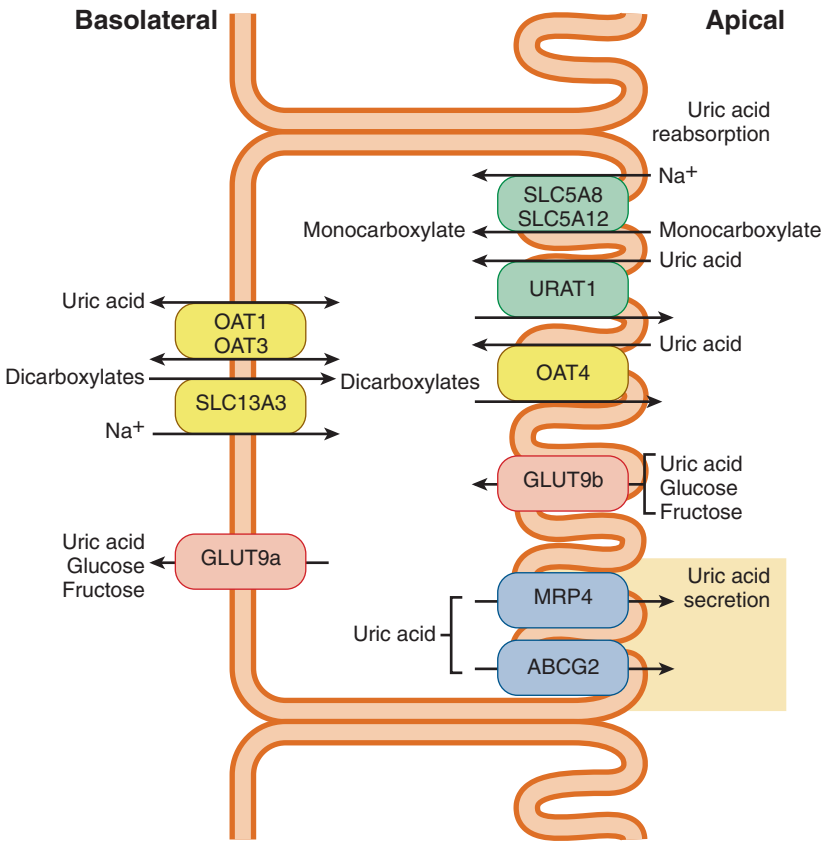

FIGURE 273-1. Renal transport of urate in proximal tubule of kidney. Serum urate reaches the tubule lumen by glomerular filtration and by secretion through the proximal tubular epithelium. Secretion of urate is facilitated in the luminal direction by MRP4, UAT, ABCG2, and NTP1, and at the basolateral membrane by OAT1 and OAT3. Reabsorption of urate from the tubular lumen is facilitated by URAT1, OAT4, OAT10, and the short isoform of GLUT (GLUT9b), and at the basolateral membrane by the long isoform (GLUT9a). ABCG2 = adenosine-triphosphate-binding cassette transporter; GLUT = glucose transporter; MRP = multidrug-resistance-related protein; NTP = sodium phosphate transport protein; OAT = organic acid transporters; URAT = uric acid transporter.

the hyperuricemia reflects accelerated cell turnover (e.g., lymphoproliferative and myeloproliferative diseases, psoriasis, chronic hemolytic states, polycythemia vera, and certain muscle glycogenoses) or other causes of enhanced purine nucleotide breakdown as seen with alcohol abuse or fructose ingestion. In addition to these secondary causes of urate overproduction, there are primary disease processes that are also responsible for urate overproduction. These are inborn errors of metabolism that result in increased de novo purine synthesis, as seen in PRPP synthetase overactivity or decreased purine salvage as seen in the complete and partial HPRT deficiencies (Lesch-Nyhan syndrome and Kelley-Seegmiller syndrome, respectively). Regardless of the cause, urate overproduction is determined by a 24-hour urine collection showing greater than 1000 mg of uric acid while eating a standard Western diet.

Pathogenesis of Monosodium Urate Monohydrate Crystals and Inflammation

MSU crystals in joints, soft tissues, and organs are the cause of pain and destruction in gout. Urate crystals will form only when physiologic conditions permit. In plasma, urate becomes insoluble at concentrations of 6.8 mg/dL (408 µmol/L) with a pH of 7.40 and normal body temperature. A reduction in pH or temperature will lower the solubility threshold even further. Not all people who are hyperuricemic will form crystals, however. There appears to be an additional requirement of a "nucleation factor" that is still poorly defined for gout and other local factors that either promote or inhibit further crystal growth.

MSU crystals form in joints and soft tissues of individuals long before they have any symptoms of gout. They deposit in small lattice structures called *microtophi* on the surface of cartilage and synovial lining. These microtophi slowly grow but are generally stable as long as the environment surrounding them does not change drastically with regard to pH, urate concentration, or temperature. At the time of the first and subsequent flares, something does change in the joint environment to cause these crystal lattice structures to break apart and shed a massive number of crystals into the joint space. These newly released and nonopsonized crystals activate receptors on synovial macrophages and are then phagocytized by monocytes and macrophages, leading to interaction with the NALP3 inflammasome. This results in the rapid production of interleukin-1 (IL-1), which is responsible for all the cardinal features of the severe inflammation associated with acute gout. The gout flare is self-limited probably as a result of a physical change in the MSU crystals after they have been repeatedly phagocytized and partially digested. A maturation of the resident monocytes in the inflamed joint also appears to help terminate the gouty flare. Even after the acute attack subsides, low-grade inflammation persists in the asymptomatic joint. It is this persistent, ongoing inflammatory process that will eventually lead to bony erosion, cartilage destruction, and synovial hypertrophy if the hyperuricemia is not treated and the MSU crystals dissolved.

CLINICAL MANIFESTATIONS

Classic Gout

The natural course of classic gout passes through three stages: asymptomatic hyperuricemia, acute intermittent gout, and chronic advanced gout.[3] The rate of progression from asymptomatic hyperuricemia to chronic advanced gout varies considerably from one person to another and is dependent on numerous factors, the most important being the degree of increase in serum urate levels.

Asymptomatic hyperuricemia[4] refers to a state in which serum urate exceeds the level of solubility (6.8 mg/dL), but symptoms of crystalline deposition have not occurred. Only 15 to 20% of all hyperuricemic people are prone to develop MSU crystals, and for this group, the period of asymptomatic hyperuricemia begins a stage of subclinical structural changes. In men, asymptomatic hyperuricemia frequently begins at puberty, whereas in women, it is usually delayed until menopause.

The initial episode of acute gout usually follows decades of asymptomatic hyperuricemia. In men, the first attack usually occurs between the fourth and sixth decades of life. In women, the age of onset is older and varies with several factors, most importantly the age of menopause. The classic acute gout attack is hallmarked by the rapid development of warmth, swelling, erythema, and exquisite pain in one or occasionally two joints. The characteristically severe pain evolves from its faintest twinge to its most intense level over an 8- to 12-hour period. Initial attacks are usually monoarticular and involve lower extremity joints. The most common joint involved is the first metatarsal phalangeal joint (termed *podagra*), followed by the ankle, midfoot, and knee (Fig. 273-2). After years of acute intermittent attacks, upper extremity joints, including the wrists, elbows, and small joints of the hands, can also become involved. Systemic symptoms of fever, chills, and malaise may accompany acute attacks, along with an intense erythema extending beyond the area of the involved joint. This may lead to confusion with a septic process.

Factors capable of provoking episodes of acute gout are those that cause fluctuations in serum urate levels, including trauma, surgery, starvation, overindulgence in certain high-purine foods, and the ingestion of any medication that raises or lowers serum urate.

The other characteristic of classic gout flares is the self-remitting nature. For the first several acute flares, the duration of the attack is 5 to 8 days. The resolution of symptoms is gradual but complete, even if no anti-inflammatory therapy is administered. The periods between the acute attacks are devoid of articular pain, although synovial fluid aspirates during this stage continue to show low-grade inflammation and the presence of MSU crystals.

Eventually, the untreated patient will progress to chronic polyarticular gout, also referred to as *advanced gout*. This stage usually develops after 10 or more years of acute intermittent gout and is evident when the pain-free intercritical periods have disappeared. Gouty flares can continue to occur against this constantly painful background. The intensity of the chronic pain is not nearly as severe as that experienced with the acute flares.

The subcutaneous tophus is the most characteristic lesion of advanced gout. The development of tophaceous deposits of MSU is a function of the duration and severity of hyperuricemia. Subcutaneous tophi may occur anywhere over the body but most commonly in the fingers, wrists, ears, knees, and olecranon bursa and at pressure points under the ulnar aspect of the forearm and Achilles tendon (Fig. 273-3). Gout at this stage can be confused with rheumatoid arthritis, especially if tophi are confused with rheumatoid nodules.

Atypical Gout Presentations

Approximately 5% of patients with gout exhibit the onset of symptoms before age 25 years. Early-onset gout represents a special subset of patients who generally have a genetic component (see Table 273-2), with a more

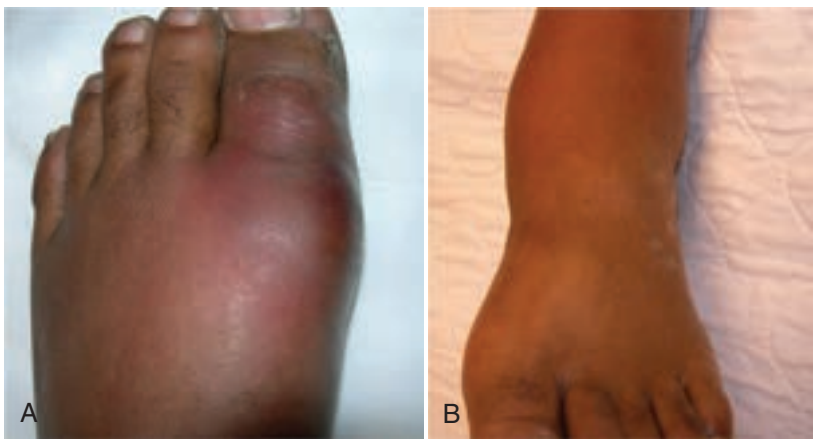

FIGURE 273-2. The intense inflammation of acute gouty arthritis. A, The marked swelling of the first metatarsophalangeal joint (podagra) is demonstrated. A dusky blue hue over an intense erythema is characteristic. **B,** Ankle swelling is shown with erythema extending beyond the area of the tibiotalar joint.

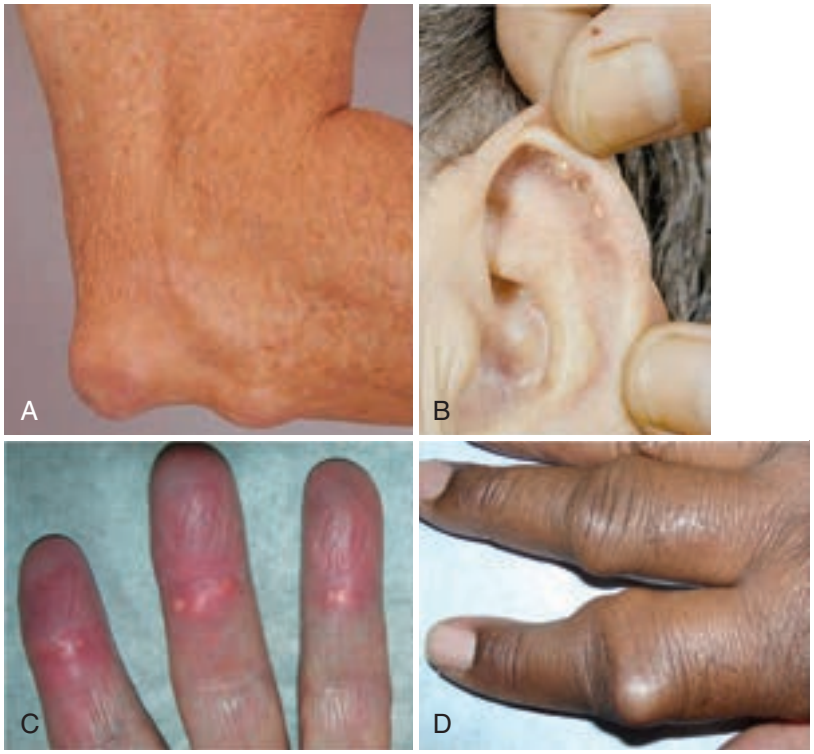

FIGURE 273-3. Characteristic locations of gouty tophi. A, At the elbow, tophi present as hard nodules along the ulnar ridge or as multiple nodules within the olecranon bursa. **B,** Ear tophi are uncommon but may be an easy source of crystal confirmation of gout when present. **C,** Small subcutaneous tophi can occur along the ventral creases of fingers. **D,** Tophi over the proximal interphalangeal or distal interphalangeal joints may be confused with Bouchard or Heberden nodes, respectively.

accelerated clinical course, requiring more aggressive antihyperuricemic therapy.

Gout develops in approximately 15% of heart transplant recipients who are taking cyclosporine to prevent allograft rejection. A slightly lower frequency of gout is seen in kidney and liver transplant recipients. Cyclosporine-induced gout has a marked shortening of the asymptomatic and acute intermittent stages, with a rapid appearance of tophi in 1 to 4 years.

In most large reviews, women account for no more than 5% of all gout subjects. This demographic is changing, with gout in older women becoming more common. Most women with gout are postmenopausal. Women who have premenopausal gout usually have renal insufficiency and hypertension and are taking thiazide diuretics, or they have a strong genetic predilection. Gout in older women may differ from classic gout in its propensity to occur in joints previously damaged by osteoarthritis such as the knees or the distal interphalangeal joints with Heberden nodes.

Saturnine gout is the term used to describe gout caused by chronic lead exposure. In the middle of the last century, this was most commonly observed in consumers of "moonshine" whiskey. Occupational exposure as seen in plumbers and environmental exposure from lead-based paints are other potential causes of this unusual form of gout.

DIAGNOSIS

Hyperuricemia is a critical risk factor for developing gout, but it is not a reliable diagnostic test because many people with elevated serum urate levels will never develop gout. Serum urate levels during an acute flare of gout are also unreliable because they may be suppressed by as much as 1.5 to 2.0 mg/dL from baseline values. The definitive diagnosis of gout is made by polarized compensated microscopy of a synovial fluid aspirate from the affected joint. The presence of intracellular needle-shaped crystals with strong negative birefringence is the diagnostic "gold standard." Similar microscopic results from a tophus aspirate or spontaneously draining fluid would also confirm the diagnosis of gout.

Synovial fluid confirmation is obtained in as few as 10% of patients said to have gout. The presumptive diagnosis of gout is based on a pattern of acute joint symptoms coupled with the patient's own medical history or a family history. The patient's medical history may reveal comorbidities frequently associated with gout or the use of medications associated with urate retention.

The characteristic clinical presentation is the rapid onset (over 8 to 12 hours) of severe pain in one or several lower extremity joints (especially the great toe, midfoot, and ankle). The presumptive diagnosis is given much greater credence if there have been similar previous attacks that spontaneously resolved to a symptom-free state.

Radiographic evaluation is not helpful in early attacks of gout except to rule out fracture. In advanced gout, affected joints may show punched-out periarticular erosions with classic overhanging edges. Ultrasonography can detect MSU crystals layered over articular cartilage in early disease and may be diagnostic. Magnetic resonance imaging is not part of a standard evaluation but will reveal soft tissue and intra-articular tophi long before they become clinically evident.

The differential diagnosis of acute gout includes bacterial infection, trauma, sarcoidosis, and CPPD arthropathy (pseudogout). Diseases occasionally confused with advanced gout include rheumatoid arthritis, reactive arthritis, and CPPD arthropathy.

TREATMENT AND PREVENTION

In 2012, the American College of Rheumatology (ACR) published its first guidelines on the management of gout.[5,6] Subsequent developments have been systematically reviewed.[7] There were several components of these guidelines that are applicable to all subjects with recurrent gouty flares. First, and foremost, is a heavy emphasis on patient education in order to obtain an optimal treatment outcome. Not only should patients be informed about dietary and other lifestyle changes that will lower their serum uric acid and lessen flares (Table 273-3), but they should also know that their disease is caused by an excessive burden of MSU crystals already present in their joints and soft tissues. This understanding of the disease process underlying gouty arthritis will help shift the focus from symptoms to "urate burden" as the real target of treatment.

Acute Gout

The treatment goal for acute gout flare is to relieve pain and terminate the flare as quickly as possible. Resting the painful joint and applying ice are generally helpful, but pharmacologic intervention is usually necessary to alter the excruciatingly painful course that may last for several days to more than a week. The therapeutic options include nonsteroidal anti-inflammatory drugs (NSAIDs),[■] oral colchicine,[■] and corticosteroids. NSAIDs are widely used but

TABLE 273-3 SPECIFIC AMERICAN COLLEGE OF RHEUMATOLOGY RECOMMENDATIONS ON LIFESTYLE AND DIET FOR GOUTY PATIENTS

Weight loss for obese patients	Avoid: Organ meats High fructose corn syrup–sweetened drinks Alcohol overuse
Healthy overall diet	Limit: Beef, pork, lamb, shellfish Beer
Exercise to achieve fitness	Encourage: Low-fat dairy
Smoking cessation	
Stay well hydrated	

may be inappropriate for patients with renal insufficiency or peptic ulcer disease. Oral colchicine administered as 1.2 mg (two tablets) at the time of flare onset followed in 1 hour by a third 0.6-mg tablet is the recommended dosing for the first 24 hours. This is followed by 7 to 10 days of once-daily or twice-daily colchicine depending on renal function. Corticosteroids can be administered orally, intramuscularly, or intra-articularly for acute gout symptoms and is a valuable option in patients with poor renal function or intolerance to colchicine. Issues critical to treatment success for a gouty flare are the early initiation of treatment, ensuring adequate dosing of anti-inflammatory therapy, and continuing the treatment until the flare has completely resolved (usually 6 to 10 days). During the acute flare, subjects already taking urate-lowering therapy (ULT) should continue the drug, whereas those not receiving this therapy should not be started.

Recurrent and Advanced Gout

The principal goal of treating gout is to lower the serum uric acid below its saturation point so that the process of crystallization will cease and the accumulated urate burden will be gradually diminished. The 2012 ACR guidelines recommend a target serum urate of less than 6 mg/dL in all subjects, with an even lower target (<5 mg/dL) for patients with advanced gout. ULT is recommended for all patients with two or more gouty flares per year, patients with advanced disease, and those with kidney stones. Allopurinol and febuxostat are both xanthine oxidase inhibitors and are considered first-line ULT. The ACR guidelines recommend that allopurinol starting dose be no greater than 100 mg/day. The dose is gradually escalated by 100 mg daily every 2 to 5 weeks, with serum urate monitoring until the target serum urate is achieved. The maximal U.S. Food and Drug Administration (FDA)-approved dose of allopurinol is 800 mg daily. In subjects with advanced chronic kidney disease, the initial dose should be reduced to 50 mg daily, with incremental dose escalations of 50 mg. Febuxostat is an alternative ULT and should be used in patients who have failed allopurinol treatment or have demonstrated sensitivity or intolerance to allopurinol.[8] The initial febuxostat dose of 40 mg daily can be increased to 80 mg daily after 2 weeks of therapy if the serum urate target is not achieved.

In patients with gout who have not attained the target serum urate despite maximal doses of either allopurinol or febuxostat, the uricosuric agent probenecid can be added to the xanthine oxidase inhibitor. Pegloticase is an intravenously administered monomethoxypoly (ethylene glycol)-conjugated recombinant uricase that dramatically lowers serum urate levels. It is approved by the FDA for the treatment of gout in patients for whom conventional therapy has been ineffective and who still have signs and symptoms of gout.

Before the initiation of any form of ULT, the patient should be placed on maintenance anti-inflammatory therapy. This is to prevent or minimize the anticipated increase in flare activity that is associated with starting ULT. Typically, this anti-inflammatory prophylaxis is in the form of colchicine once or twice daily or low-dose NSAIDs. Anti-inflammatory prophylaxis should be continued until the subject has been free of gout flares for 6 months or more.

CALCIUM PYROPHOSPHATE DIHYDRATE CRYSTAL DEPOSITION DISEASE

DEFINITIONS

The heterogeneous group of clinical conditions associated with CPPD crystals are collectively called *CPPD crystal deposition disease*. Within this spectrum is the common radiographic finding of chondrocalcinosis (CC) that is frequently asymptomatic. The acute synovitis associated with intra-articular CPPD crystals can closely mimic the findings of gout and is hence referred to as *pseudogout*. The more chronic changes associated with CPPD-induced bone and cartilage destruction is referred to as *pyrophosphate arthropathy*. These conditions are further classified as familial (genetic), metabolic, and sporadic.

EPIDEMIOLOGY

The true prevalence of CPPD crystal deposition disease is unknown, but it is generally thought to be underdiagnosed because of its confusion with other forms of arthritis.[9] The prevalence of radiographically appearing chondrocalcinosis has been studied extensively and is clearly an age-related phenomenon. Chondrocalcinosis of the meniscal and articular cartilage of the knee is seen in 4% of people between the ages of 55 and 59 years; in 18% of those 80 to 84 years; and in approximately 27% of those older than 85 years.

PATHOBIOLOGY

Clinical CPPD disease is divided into three categories based on the etiology of altered inorganic pyrophosphate (PPi) metabolism. The categories are hereditary (familial), sporadic (idiopathic), and metabolic. All three types of CPPD disease are associated with extracellular PPi accumulation around chondrocytes, and extracellular PPi is necessary for the formation of CPPD crystals. Intracellular PPi is a common byproduct of many synthetic reactions, and if allowed to accumulate intracellularly, it impedes the further synthesis of proteins, nucleotides, and lipids. Therefore, there are multiple mechanisms for removing PPi from the cellular cytoplasm, including specific and nonspecific membrane transporters and both intracellular and extracellular (ecto-) pyrophosphate hydrolysis.

Hereditary forms of CPPD crystal deposition disease may be caused by increased transmembrane transport of chondrocyte intracellular PPi to its extracellular matrix by diminished activity of the protective pyrophosphate hydrolases, or because of altered influences of factors that lead to increased extracellular PPi, including transforming growth factor-β, bone morphogenic protein-2 and -4, ascorbic acid, and osteopontin. The best characterized of the hereditary causes is a gain of function mutation of the ANK anion transporter on the chondrocyte plasma membrane.

These types of mutations are rarely observed in the sporadic form of CPPD disease. Aging chondrocytes in culture produce considerably more PPi than do younger chondrocytes, although the exact mechanism for this is unclear.

The metabolic diseases that predispose to CPPD crystal deposition disease include hemochromatosis, hyperparathyroidism, hypomagnesemia (as in Gitelman syndrome), and hypophosphatasia. All of these metabolic conditions result in increased extracellular PPi or other alterations in the cartilage matrix that are permissive for CCPD crystal formation.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

CPPD crystal deposition disease presents in a variety of fashions. It is frequently asymptomatic (lanthanic) and recognized only by the appearance of chondrocalcinosis on radiographs. The most common clinical manifestation accounting for approximately 60% of CPPD disease is a polyarticular, noninflammatory arthritis affecting joints not typically involved in primary osteoarthritis, including the wrists, shoulders, and metacarpophalangeal joints (particularly the second and third metacarpophalangeals). This form of CPPD disease is called *pseudo-osteoarthritis* and may be associated with occasional inflammatory attacks. The acute monoarticular presentation is known as pseudogout. The pain and swelling of pseudogout can be similar to that seen in gout. The onset is usually not as abrupt as with gout, and the attacks tend to last longer—frequently months. Pseudogout occurs more often in the large joints than in small joints. CPPD crystal deposition disease can occasionally present as a chronic polyarticular inflammatory disease that may mimic rheumatoid arthritis or polymyalgia rheumatica.

CPPD disease is diagnosed by identifying chondrocalcinosis by radiography in a patient with a clinical history suggestive of the disease (Fig. 273-4). Definitive diagnosis is made by the finding of CPPD crystals by compensated polarized light microscopic examination of aspirated synovial fluid.[10]

PREVENTION AND TREATMENT

Rx

There are no specific therapies for CPPD crystal deposition disease.[11] In patients with metabolic disease–associated CPPD disease, treatment and control of the metabolic disease can afford some improvement in the arthritis, although this is not the case for phlebotomy-treated hereditary hemochromatosis. For both the acute and chronic forms of CPPD crystal deposition disease, therapy is directed at symptoms. NSAIDs are the mainstay of

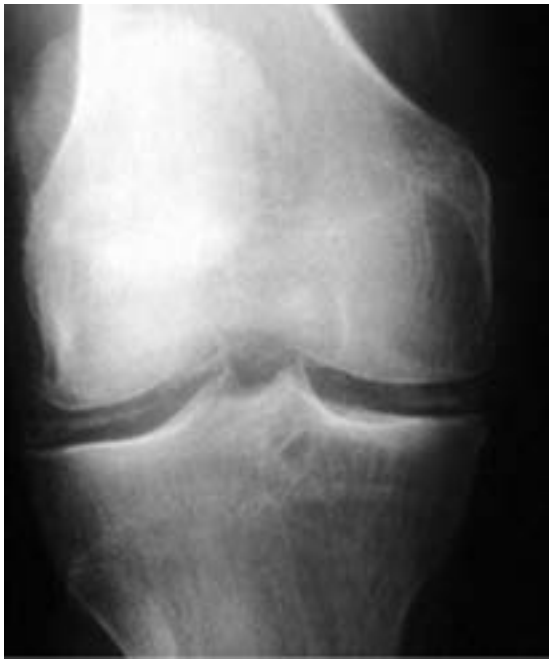

FIGURE 273-4. Calcium pyrophosphate dehydrate arthropathy of the knee. Radiographic evidence of chondrocalcinosis showing fibrocartilage calcification as thick, linear deposits parallel to and separate from subchondral bone.

treatment. Low-dose oral colchicine can be used in both acute and chronic settings. Intra-articular steroids have also been proved to be beneficial for symptomatic CPPD crystal deposition disease. The IL-1 inhibitor, anakinra, has been effectively used "off-label" for treating CPPD disease flares. On the other hand, intra-articular viscosupplementation (hyaluronic acid) may exacerbate joint symptoms.

APATITE (BASIC CALCIUM PHOSPHATE)-ASSOCIATED ARTHROPATHY

BCP crystals include several different crystal species. The most common of these is hydroxyapatite. The BCP crystals are unlike MSU or CPPD crystals in that they are not identifiable by polarized microscopy. These very tiny crystals are responsible for several important clinical conditions. Although BCP crystals are found in 50% of osteoarthritic synovial fluids, the incidence and prevalence of the individual apatite-associated clinical manifestations have not been established. This is especially true of the most severe and destructive apatite syndromes such as Milwaukee shoulder and tumoral calcinosis. The prevalence of the most common apatite-associated conditions, calcific periarthritis of the shoulder, was 3% in a large North American study.

PATHOBIOLOGY

Like MSU and CPPD crystals, BCP crystals exert their pro-inflammatory effects by being phagocytized by resident synoviocytes and influxing leukocytes. Unlike MSU and CPPD crystals, apatite crystals apparently do not act through the NALP3 inflammasome. Rather, BCP crystals are dissolved in the acidic phagosome and raise intracellular calcium levels and then activate the calcium-dependent signaling pathways. The synoviocytes are stimulated by these pathways to increase production of tumor necrosis factor-α and IL-6, and influxing neutrophils are stimulated to increase pro-inflammatory oxygen radicals. The fibroblasts in the joint lining and surrounding soft tissues are stimulated to increase production of many of the matrix metalloproteinases, such as collagenase 1, collagenase 3, and stromolysin 1.

CLINICAL MANIFESTATIONS OF BASIC CALCIUM PHOSPHATE DEPOSITION

The clinical manifestations of BCP deposition can be acute or chronic, and their cause can be idiopathic, hereditary, or secondary to other diseases that cause hypercalcemia (Table 273-4). BCP crystals can be found in 50% of synovial fluids from osteoarthritic knees. The presence of BCP crystals correlates with more severe radiographic changes secondary to more rapid deterioration in these osteoarthritic knees. Acute inflammatory arthritis associated with BCP deposition is similar in many ways to that of gout and pseudogout and has been referred to as *pseudo-pseudogout*. The patients tend to be younger and usually have evidence of soft tissue BCP deposition elsewhere in the body. Periarticular calcifications are often asymptomatic. However, BCP deposition can cause an acute and severe inflammation of the ligaments, tendons, and bursae surrounding a joint, termed *acute calcific periarthritis*. This frequently occurs around the shoulders and hips but can also occur in fingers, toes, wrists, and ankles.

TABLE 273-4 CLINICAL MANIFESTATIONS OF BASIC CALCIUM PHOSPHATE DEPOSITION DISEASE

Osteoarthritis and basic calcium phosphate deposition
Acute inflammatory arthritis
Acute calcific periarthritis
Chronic noninflammatory arthropathy
Diffuse idiopathic skeletal hyperostosis (DISH)
Tumoral calcinosis
Calcifications associated with hypercalcemic states
Hyperparathyroidism
Hypervitaminosis D
Sarcoidosis
Metastatic cancer
Myeloma
Leukemia

The most destructive BCP-associated arthropathy is Milwaukee shoulder, which is characterized by large noninflammatory effusions containing BCP crystals with or without CPPD crystals. The process results in the destruction of the rotator cuff leading to marked instability and glenohumeral cartilage dissolution. This process can also affect knees. Like the acute calcific periarthritis discussed previously, Milwaukee shoulder is observed in women four times more frequently than in men.

Diffuse idiopathic skeletal hyperostosis (DISH) predominates in men and elderly people. The radiographic appearance of this condition is flowing ossifications along the anterolateral aspect of spinal vertebrae, especially in the thoracic spine. DISH is usually asymptomatic, but very large bridging osteophytes can cause pain; in the cervical spine, it can even result in dysphagia.

Idiopathic tumoral calcinosis is rare but most prevalent in young patients of African descent. These subjects have large irregular calcifying masses in the soft tissue surrounding the shoulders, hips, and elbows. Some cases show a familial occurrence and medical conditions associated with hyperphosphatemia.

Finally, metastatic calcifications can arise in any medical condition associated with hypercalcemia such as hyperparathyroidism, hypervitaminosis D, sarcoidosis, metastatic cancer, multiple myeloma, and leukemia.

TREATMENT AND PREVENTION

Rx

The treatment of most BCP-associated syndromes is conservative. In acute inflammatory arthritis and periarthritis, low-dose oral colchicine and NSAIDs are the mainstay of a symptomatic therapy. The persistent large effusion seen in Milwaukee shoulder should be serially aspirated to decrease intracapsular pressure. The added effectiveness of corticosteroid injections in this setting is unproved. The calcinosis associated with abnormal calcium and phosphate metabolism is best managed by treating the underlying metabolic process.

Grade A References

A1. van Durme CM, Wechalekar MD, Buchbinder R, et al. Non-steroidal anti-inflammatory drugs for acute gout. *Cochrane Database Syst Rev*. 2014;9:CD010120.

A2. van Echteld I, Wechalekar MD, Schlesinger N, et al. Colchicine for acute gout. *Cochrane Database Syst Rev*. 2014;8:CD006190.

A3. Seth R, Kydd AS, Buchbinder R, et al. Allopurinol for chronic gout. *Cochrane Database Syst Rev*. 2014;10:CD006077.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Zhu Y, Pandys BJ, Choi HK. Prevalence of gout and hyperuricemia in the US general population: the National Health and Nutrition Examination Survey 2007-2008. *Arthritis Rheum*. 2011;63:3136-3141.
2. Reginato AM, Mount DB, Yang I, et al. The genetics of hyperuricemia and gout. *Nat Dev Rheumatol*. 2012;8:610-621.
3. Roddy E, Mallen CD, Doherty M. Gout. *BMJ*. 2013;347:f5648.
4. Akkineni R, Tapp S, Tosteson ANA, et al. Treatment of asymptomatic hyperuricemia and prevention of vascular disease: a decision analytic approach. *J Rheumatol*. 2014;41:739-748.
5. Khanna D, Fitzgerald JD, Khanna PP, et al. American College of Rheumatology guidelines for management of gout. Part 1: systematic nonpharmacologic and pharmacologic therapeutic approaches to hyperuricemia. *Arthritis Care Res*. 2012;64:1431-1446.
6. Khanna D, Khanna PP, Fitzgerald PP, et al. American College of Rheumatology guidelines for management of gout. Part 2: therapy and anti-inflammatory prophylaxis of acute gouty arthritis. *Arthritis Care Res*. 2012;64:1447-1461.
7. Khanna PP, Gladue HS, Singh MK, et al. Treatment of acute gout: a systematic review. *Semin Arthritis Rheum*. 2014;44:31-38.
8. Jutkowitz E, Choi HK, Pizzi LT, et al. Cost-effectiveness of allopurinol and febuxostat for the management of gout. *Ann Intern Med*. 2014;161:617-626.
9. Abhishek A, Doherty M. Epidemiology of calcium pyrophosphate crystal arthritis and basic calcium phosphate crystal arthropathy. *Rheum Dis Clin North Am*. 2014;40:177-191.
10. Löffler C, Sattler H, Peters L, et al. Distinguishing Gouty Arthritis from Calcium Pyrophosphate Disease and Other Arthritides. *J Rheumatol*. 2015;42:513-520.
11. Pascart T, Richette P, Flipo RM. Treatment of nongout deposition disease: an update. *Arthritis*. 2014;2014:375202.

REVIEW QUESTIONS

1. The prevalence of hyperuricemia in the United States has been rising dramatically over the past three decades because of all of the following *except* which one?

A. Increased longevity of the population
B. Increased use of low-dose aspirin and thiazide diuretics
C. Greater consumption of low-fat dairy products
D. Greater consumption of high-fructose corn syrup
E. Increased prevalence of the metabolic syndrome

Answer: C The aging population, the increasing prevalence of the metabolic syndrome, and increased consumption of medications and diets that increase blood uric acid levels by either inhibiting its excretion (low-dose aspirin and thiazide diuretics) or increasing its production (high-fructose corn syrup) are all factors leading to the increased prevalence of hyperuricemia and gout in the United States. Increased consumption of low-fat dairy products does not cause hyperuricemia.

2. Which one of the following is true regarding renal clearance of uric acid?

A. Comparable in gouty and nongouty subjects
B. Reduced in gouty subjects by 70%
C. Proportional to the serum urate levels in both gouty and nongouty subjects
D. Controlled by components of the "transportasome" in the distal nephron
E. Reduced by the angiotensin II receptor blocker losartan

Answer: C Gouty subjects have reduced renal clearance of uric acid, and their excretion of uric acid is reduced by only 30% (not 70%). The renal tubular transporters (collectively called the *transportasome*) that are responsible for determining how much of the filtered uric acid is actually excreted are located in the proximal convoluted tubules, not the distal nephron. Renal clearance of uric acid is not reduced by losartan.

3. Factors that may provoke flares of gout include all of the following *except* which one?

A. Initiation of allopurinol
B. Overindulgence in shellfish
C. Starvation
D. Initiation of colchicine
E. Surgery

Answer: D Initiation of colchicine can actually block flares of gout. In contrast, allopurinol or any other drug or factor that alters blood uric acid levels (either up or down) can provoke an acute gouty attack.

4. According to the American College of Rheumatology (ACR) 2012 treatment guidelines for urate-lowering therapy, which one of the following is true?

A. Prophylactic anti-inflammatory therapy should be initiated before starting allopurinol or febuxostat.
B. Patients with chronic kidney disease should not be treated with more than 300 mg of allopurinol daily.
C. Serum urate levels should be reduced to less than 5 mg/dL in all patients with gout.
D. The initial starting dose of allopurinol is 300 mg daily in most patients.
E. Probenecid is first-line therapy for patients with chronic kidney disease stage 3 or 4.

Answer: A Prophylactic anti-inflammatory drugs should be initiated before starting allopurinol or febuxostat because any acute change in uric acid levels (including acute changes that actually reduce the levels, as these drugs would do, as well as increase the levels) can precipitate an acute gout attack. The ACR guidelines recommend starting allopurinol at doses of 100 mg daily in those without kidney disease or 50 mg daily in those who do have advanced chronic kidney disease, followed by dose escalations of 100 mg or 50 mg in these situations, respectively. ACR guidelines recommend a target serum uric acid of less than 6 mg/dL, except in those with advanced gout, whose levels should be lowered to less than 5 mg/dL. The uricosuric agent, probenecid, is recommended only as addition to a first-line xanthine oxidase inhibitor if target urate levels cannot be attained with maximum doses of the latter.

5. The laboratory investigation for underlying metabolic causes of calcium pyrophosphate dehydrate (CPPD) deposition disease would include all of the following, *except* which one?

A. Serum ferritin
B. Serum calcium and phosphate
C. Serum alkaline phosphatase
D. Serum magnesium
E. Serum lead

Answer: E Metabolic diseases that predispose to CPPD crystal deposition disease include hemochromatosis, hyperparathyroidism, hypomagnesemia states, and hypophosphatasia. *Saturnine gout* is the term used to describe gout (not CPPD crystal deposition disease) caused by chronic lead exposure (e.g., by occupational exposure or drinking "moonshine" whiskey).

6. The basic calcium crystal arthropathy referred to as *Milwaukee shoulder* is characterized by which one of the following?

A. Bilateral involvement in men
B. Destruction of the glenohumeral joint
C. Markedly inflammatory synovial fluid
D. Birefringent crystals in the synovial fluid
E. Slow progression

Answer: B The form of BCP arthropathy known as Milwaukee shoulder has a predilection for elderly women, is usually unilateral but can involve the knees, and has non-inflammatory synovial fluid (WBC <1,000 per cubic mm) that is usually bloody in appearance. Despite non-inflammatory fluid the condition can be rapidly destructive.

274 FIBROMYALGIA, CHRONIC FATIGUE SYNDROME, AND MYOFASCIAL PAIN

ROBERT M. BENNETT

Fibromyalgia (FM) and chronic fatigue syndrome (CFS) are multisymptomatic syndromes defined respectively by the core features of chronic widespread pain and chronic unexplained fatigue. Although there is a significant overlap in symptomatology between FM and CFS, they are generally considered to be separate disorders. Myofascial pain is a universal experience that is usually self-limited; when it becomes persistent, it may accentuate and perpetuate the experience of chronic fatigue and FM.

FIBROMYALGIA

EPIDEMIOLOGY

Chronic musculoskeletal pain is commonly encountered in the general population with an estimated prevalence of about 20%. It is subdivided into chronic regional pain, with a prevalence of about 25%, and chronic widespread pain, with a prevalence of about 10%.

FM is considered to be a subset of chronic widespread pain and has a prevalence of about 2% in women and 0.5% in men. There is a steady increase in FM with age, with about 12% of women in their 60s being affected. Chronic musculoskeletal pain is associated with a reduction in overall health status, and FM patients have more impairment than do patients with chronic widespread pain or chronic regional pain. The prevalence of FM in the medical setting is much greater, with about 20 to 30% of the rheumatology visits in the United States made for FM.

PATHOBIOLOGY

There is persuasive evidence that FM pain results from abnormal sensory processing within the central nervous system (Fig. 274-1).[1] This is commonly referred to as *central sensitization* and results from an amplification of peripheral sensory stimuli and a deficit of descending inhibitory control from the midbrain.[2] The magnification of peripheral sensory input is readily visualized as increased activity in somatosensory areas of the brain on magnetic resonance imaging (MRI) in FM subjects versus healthy controls. Central sensitization is the result of a persistent neuronal hyperexcitability that continues long after the original sensitizing input has waned. The pathophysiology of this phenomenon is based, in part, on temporal summation of neural impulses. This occurs when pain fibers (unmyelinated C) are repetitively stimulated at a rate greater than one impulse every 3 seconds. At a biochemical level, such stimulation results in depolarization of *N*-methyl-D-aspartate (NMDA) receptors, which causes transcriptional changes that affect pain processing. Thus, persistent pain input may eventually give rise to central sensitization in some individuals, probably based, in part, on a genetic predisposition.[3]

The frequent development of FM from focal pain states, such as in rheumatoid arthritis, post-traumatic injuries, and endometriosis, is thought to be a result of chronic nociceptive input stimulating central neuroplastic changes.

The descending pain system, originating in the periaqueductal gray nucleus of the midbrain and terminating in the dorsal horn, is important in modulating the transmission of nerve impulses to the brain. This inhibitory system is defective in FM patients. Activation of the nuclei in the midbrain involved in this system occurs in response to opioids, endorphins, emotions, and the placebo response. This modulation may upregulate or downregulate sensory processing, and in part, it underlies the influence of the psyche on the pain experience.

Understanding the biologic basis for central sensitization has provided an explanation for the common association of FM with other conditions such as irritable bowel syndrome, overactive bladder, multiple chemical hypersensitivity, and chronic daily headaches.

CLINICAL MANIFESTATIONS

FM patients always have the core symptom of widespread pain but also report a wide array of other symptoms (Table 274-1).[4]

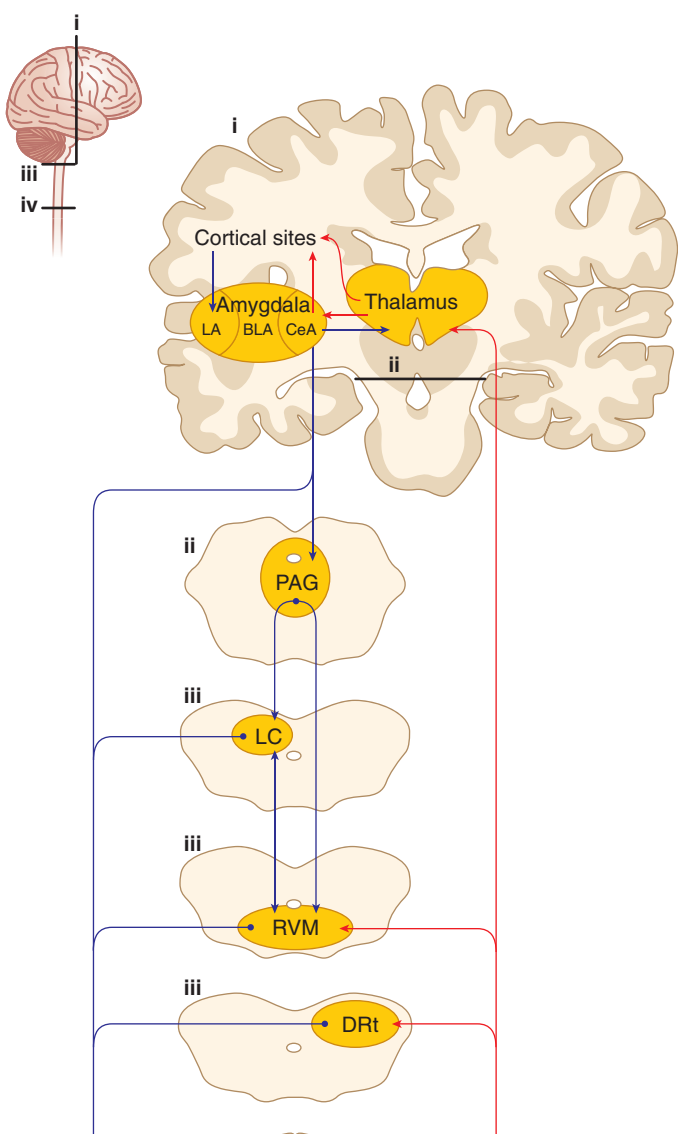

FIGURE 274-1. Neuroanatomy of pain pathways. Nociceptive inputs enter the spinal dorsal horn through primary afferent fibers that synapse onto transmission neurons. The projection fibers ascend through the contralateral spinothalamic tract. Ascending projections target the thalamus, and collateral projections also target mesencephalic nuclei, including the DRt, the RVM, and the midbrain PAG. Descending projections from the DRt are a critical component of the DNIC pathway. Rostral projections from the thalamus target areas that include cortical sites and the amygdala. The lateral capsular part of the CeA ("nociceptive amygdala") receives nociceptive inputs from the brain stem and spinal cord. Inputs from the thalamus and cortex enter through the lateral (LA) and basolateral (BLA) amygdala. The CeA sends outputs to cortical sites and the thalamus, in which cognitive and conscious perceptions of pain are integrated. Descending pain modulation is mediated through projections to the PAG, which also receives inputs from other sites, including the hypothalamus, and communicates with the RVM as well as other medullary nuclei that send descending projections to the spinal dorsal horn through the DLF. The noradrenergic locus coeruleus (LC) receives inputs from the PAG, communicates with the RVM, and sends descending noradrenergic inhibitory projections to the spinal cord. Antinociceptive and pronociceptive spinopetal projections from the RVM positively and negatively modulate nociceptive inputs and provide for an endogenous pain regulatory system. Ascending (red) and descending (blue) tracts are shown schematically. Areas labeled "i-iv" in the small diagram correspond with labeled details of the larger diagram. (From Ossipov MH, Dussor GO, Porreca F. Central modulation of pain. *J Clin Invest* 2010;120[11]:3779-3787.)

TABLE 274-1 SELF-REPORTED SYMPTOMS IN FIBROMYALGIA

CURRENT SYMPTOM	FREQUENCY (%)
Low back pain	63
Recurrent headaches	47
Arthritis	46
Muscle spasm	46
Tingling	46
Balance problems	45
Irritable bowel syndrome	44
Numbness	44
Chronic fatigue	40
Bloating	40
Depression	40
Anxiety	38
Sinus problems	37
Tooth disorders	32
Restless legs	32
Tinnitus	30
Jaw pain	29
Bladder problems	26
Rashes	25

From Bennett RM, Jones J, Turk DC, et al. An internet survey of 2,596 people with fibromyalgia. *BMC Musculoskelet Disord.* 2007;8:27.

Pain
The core symptom of FM is chronic widespread pain and stiffness. Characteristically, the pain is described as a constant dull ache that is worsened by muscle overactivity. FM-related pain is usually perceived as arising from muscle; however, many patients also report joint pain but have no objective evidence of arthritis.

Fatigue
Easy fatigability from physical exertion, mental exertion, and psychological stressors is typical of FM. Patients with CFS have many similarities with FM patients; about 75% of patients meeting the diagnostic criteria for CFS also meet criteria for the diagnosis of FM.

Disordered Sleep
FM patients characteristically have nonrestorative sleep. Even if they sleep continuously for 8 to 10 hours, they awake feeling tired. Many exhibit an alpha-delta electroencephalographic pattern that would explain their never achieving the restorative stages of non–rapid eye movement sleep. A poor night's sleep is often followed by worsening of FM symptoms the next day. Restless legs syndrome and the associated periodic limb movement disorder occur in up to 60% of FM patients. A thorough evaluation for sleep disorders, such as sleep apnea, is warranted in all FM patients.

Cognitive Dysfunction
Cognitive dysfunction is a prominent complaint of many FM patients. They commonly describe difficulties with working memory, semantic memory, concentration, logical analysis, and motivation. It has been estimated that the working memory deficit in FM is comparable to 20 years of aging. Recent studies have documented that these neurocognitive deficits correlate with local brain morphology in the frontal lobe and anterior cingulate gyrus.

Psychological Distress
Having a chronic painful disorder for which there is currently no generally accepted cure often produces an existential crisis. Approximately 30% of FM patients have significant depression at any given time, and about 60% have a lifetime prevalence of depressive illness. Although psychiatric disorders are common in FM patients, they do not seem to be intrinsically related to the pathophysiology of FM; effective treatment of depression with a selective serotonin re-uptake inhibitor (SSRI) does not eliminate FM pain. FM patients are at an increased risk for suicide and suicide ideation.[5] Fibromyalgia and depression show familial clustering, suggestive of an underlying genetic susceptibility.

Associated Problems
Fibromyalgia patients often have a spectrum of other pain disorders that share similar features, such as not having a well-defined etiology, having an association with mood disorders, and having no definitive cure. These accompanying disorders are often referred to as *central sensitivity syndromes* and include irritable bowel syndrome, overactive bladder syndrome, temporomandibular pain disorder, chronic pelvic pain, chronic daily headaches, and chemical hypersensitivity syndrome.

History
Initiation and Maintenance of Fibromyalgia
FM seldom emerges "out of the blue." Many patients relate an onset following an acute injury, repetitive workload, persistent stress, infections, and toxin exposure. It is not uncommon for a regional pain state to evolve into FM. FM is commonly found as an accompaniment of other painful disorders, such as rheumatoid arthritis, migraine, low back pain, systemic lupus erythematosus, Sjögren syndrome, inflammatory bowel disease, endometriosis, and osteoarthritis. There is increasing evidence that FM is a common occurrence in victims of post-traumatic stress disorder; for instance, both FM and CFS were common diagnoses after Operation Desert Storm in 1991. Childhood abuse and sexual trauma are often elicited during a careful and empathetic history.

DIAGNOSIS
The current "gold standard" for diagnosis of FM is based on the classification criteria of the American College of Rheumatology.[6] These criteria require persistent symptoms of at least 3 months' duration and widespread pain accompanied by tenderness at 11 or more of 18 tender point locations.

The following are the locations of the nine paired tender points:

Occiput: bilateral, at the suboccipital muscle insertions
Low cervical: bilateral, at the anterior aspects of the intertransverse spaces at C5-C7
Trapezius: bilateral, at the midpoint of the upper border
Supraspinatus: bilateral, at the origins, above the scapular spine near the medial border
Second rib: bilateral, at the second costochondral junctions just lateral to the junctions on the upper surfaces
Lateral epicondyle: bilateral, 2 cm distal to the epicondyles
Gluteal: bilateral, in the upper outer quadrants of the buttocks in the anterior fold of muscle
Greater trochanter: bilateral, posterior to the trochanteric prominence
Knee: bilateral, at the medial fat pad proximal to the joint line

Because the fibromyalgia tender point evaluation is seldom performed in clinical practice, more recent criteria have proscribed a diagnosis based on the combination of common symptoms and pain locations. Figure 274-2 is one example of such criteria that has an 80% diagnostic accuracy.

FM is not a diagnosis of exclusion; thus, laboratory tests and imaging studies play no role in establishing the diagnosis, although they are often indicated in the evaluation of concomitant peripheral pain generators and accompanying symptoms.

TREATMENT Rx

Successful management of FM requires a thorough analysis in terms of the biopsychosocial model of disease. The major management issues that require attention are listed in Figure 274-3.[7]

Education
There is evidence that higher educational attainment is associated with a better prognosis in many chronic disorders, such as FM. Education has a positive effect through cognitive behavioral strategies, such as goal setting and reassessment of priorities. Educated patients are more likely to take an active role in self-management.

Pain
In considering the management of pain in FM, it is logical to focus on the major sites of pain processing, namely, peripheral pain generation and central pain pathways. There is no specific tissue pathology, at least in peripheral tissues, that is characteristic of FM. However, once the central nervous system

Pain location inventory (PLI):

Directions: For each of the following 28 sites, select those locations where you have experienced persistent pain during the past 7 days. The score will be between 0 and 28.

Neck	Left upper back	Right wrist	Left thigh
Right jaw	Right lower back	Left wrist	Right knee
Left jaw	Left lower back	Right hand	Left knee
Mid–upper back	Right shoulder	Left hand	Right ankle
Front of chest	Left shoulder	Right hip	Left ankle
Mid–lower back	Right arm	Left hip	Right foot
Right upper back	Left arm	Right thigh	Left foot

10–item Symptom Impact Questionnaire (SIQR):

Directions: For each of the following 10 questions, check the one box that best indicates the intensity of the following common symptoms over the last 7 days. Divide the total symptom score (0–100) by 2 (i.e., the range will be 0 to 50).

		0	1	2	3	4	5	6	7	8	9	
1.	*Pain*	No pain	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Unbearable pain
2.	*Energy*	Lots of energy	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	No energy
3.	*Stiffness*	No stiffness	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Severe stiffness
4.	*Sleep*	Awoke rested	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Awoke very tired
5.	*Depression*	No depression	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Very depressed
6.	*Memory Problems*	Good memory	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Very poor memory
7.	*Anxiety*	Not anxious	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Very anxious
8.	*Tenderness to touch*	No tenderness	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Very tender
9.	*Balance problems*	No imbalance	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Severe imbalance
10.	*Sensitivity to loud noises, bright lights, odors and cold*	No sensitivity	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Extreme sensitivity

Criteria:

A patient fulfilling the following guidelines has a high likelihood of having FM:*

1. **The symptoms and pain locations have been persistent for at least the last 3 months**
2. **Pain location score is ≥ 17**
3. **SIQR symptom score is ≥ 21**

* • *Fibromyalgia patients have a continuum of symptoms; a diagnosis based on a strict numerical cutoff is subject to error.*
• *The presence of another pain disorder or related symptoms does not rule out a diagnosis of fibromyalgia.*
• *A careful clinical evaluation is always required in order to identify any condition that could fully account for the patient's symptoms and/or contribute to the severity of the symptoms.*

FIGURE 274-2. Alternative diagnostic criteria for fibromyalgia.[4] (From Bennett RM, Friend R, Marcus D, et al. Criteria for the diagnosis of fibromyalgia: validation of the modified 2010 preliminary American College of Rheumatology criteria and the development of alternative criteria. *Arthritis Care Res (Hoboken)*. 2014;66:1364-1373.)

is sensitized, not only are peripheral pain generators perceived as being more painful, but they also prolong and amplify the central neuroplastic changes. Thus, a critical first component of treating FM pain is to identify and effectively treat all peripheral pain generators, which commonly include peripheral arthritis, axial arthritis, spinal stenosis, myofascial trigger points, neuropathic pain, vascular headaches, visceral pain (e.g., irritable bowel syndrome, overactive bladder), postsurgical pain, and pelvic pain syndromes (e.g., endometriosis).

Management of central sensitization is typically initiated with heterocyclic antidepressant (HCA) medications such as amitriptyline, trazodone, cyclobenzaprine, or nortriptyline. (Chapter 397) There is evidence that antidepressant medications with both re-uptake inhibition of serotonin and norepinephrine (e.g., venlafaxine, duloxetine, milnacipran) are more effective in treating FM pain. Both duloxetine and milnacipran have U.S. Food and Drug Administration (FDA) approval for use in FM. Importantly, the mechanism of action of these drugs is independent of any antidepressant effect and results from upregulation of the descending pain system that originates in the brain stem and uses serotonin and norepinephrine as neurotransmitters at dorsal horn synapses. Anticonvulsant medications such as gabapentin, pregabalin, and topiramate are increasingly being used in FM and other chronic pain states; they inhibit the presynaptic release of glutamate and thus modulate the activation of NMDA receptors. Pregabalin has been beneficial in randomized trials and has FDA approval for use in FM. Sedative hypnotics such as zolpidem and zopiclone often help the nonrestorative sleep, but not pain. On the other hand, sodium oxybate, an FDA-approved drug for narcolepsy, has been shown to improve not only sleep but also pain, stiffness, and fatigue; it has not been approved for use in fibromyalgia.

Opioids are often used in the treatment of FM, but long-term trials are lacking. Opioids should not be the first choice for analgesia; however, they should not be withheld if less powerful analgesics have failed. Tramadol (both Ultram and Ultracet) has proved useful in reducing FM pain in two controlled trials. Tramadol is a weak opioid agonist that also inhibits the re-uptake of serotonin and norepinephrine at the level of the dorsal horn. It is metabolized by CYP2D6, as are many antidepressant medications. FM patients who are taking tramadol or any of several antidepressants that are eliminated by CYP2D6 are at risk for the development of a serotonin syndrome (Chapter 434). A careful review of concomitant medications is an important prerequisite in the prescription of any new medication, especially in difficult-to-treat patients who are often taking multiple medications.

Fatigue

A search for a treatable cause of fatigue is always indicated (E-Table 274-1). Common treatable causes of chronic fatigue in FM patients are inappropriate dosing of medications, depression, aerobic deconditioning, primary sleep disorder (e.g., sleep apnea), non-restorative sleep, a coexisting inflammatory disorder, and neurally mediated hypotension. As in CFS, the underlying cause of *primary* fatigue in FM is not known. Sodium oxybate, methyl phenidate, and modafinil have provided worthwhile improvement in fatigue in some FM patients.

Sleep

Low-dose HCAs, particularly trazodone and cyclobenzaprine, are the mainstay of sleep pharmacotherapy in FM patients. Some patients cannot tolerate HCAs because of unacceptable levels of daytime drowsiness or weight gain.

E-TABLE 274-1 DIFFERENTIAL DIAGNOSIS OF CHRONIC FATIGUE

CHRONIC INFECTIONS	**CHRONIC INFLAMMATION**
Hepatitis C Lyme disease Parasitic and fungal infections Chronic pulmonary infections Human immunodeficiency virus Tropical diseases	Rheumatoid arthritis Systemic lupus erythematosus Sjögren syndrome Polymyositis/dermatomyositis Vasculitis Sarcoidosis
SLEEP DISORDERS	**CARDIOPULMONARY**
Obstructive sleep apnea Restless leg syndrome Circadian rhythm disorder Upper airway resistance syndrome Narcolepsy/parasomnias Alpha-delta sleep disorder	Congestive heart failure Neurally mediated hypotension Postural orthostatic tachycardia syndrome Pulmonary hypertension Chronic obstructive pulmonary disease Mitral valve prolapse
ENDOCRINE/METABOLIC DISORDERS	**GASTROINTESTINAL**
Addison disease Cushing syndrome Poorly controlled diabetes Thyroid disorders Hemochromatosis Hypopituitarism Diabetes insipidus	Celiac disease Inflammatory bowel disease Autoimmune hepatitis Hepatic cirrhosis
	MALIGNANCY
	Lymphoma and occult malignancies Post-chemotherapy syndrome
GENERAL MEDICAL DISORDERS	**NEUROLOGIC DISORDERS**
Anemia (any cause) Chronic renal/hepatic failure Malnutrition Medication side effects Chronic pain disorders	Multiple sclerosis Myasthenia gravis Muscular dystrophies Parkinson disease Early dementia
PSYCHOLOGICAL	**LIFESTYLE FACTORS**
Mood disorders (depression/anxiety/bipolar) Schizophrenia Post-traumatic stress disorder Anorexia nervosa/bulimia Childhood abuse and/or neglect	Chronic overwork Persistent unresolved stress Inadequate exercise Morbid obesity (body mass index >40) Alcoholism/drug abuse

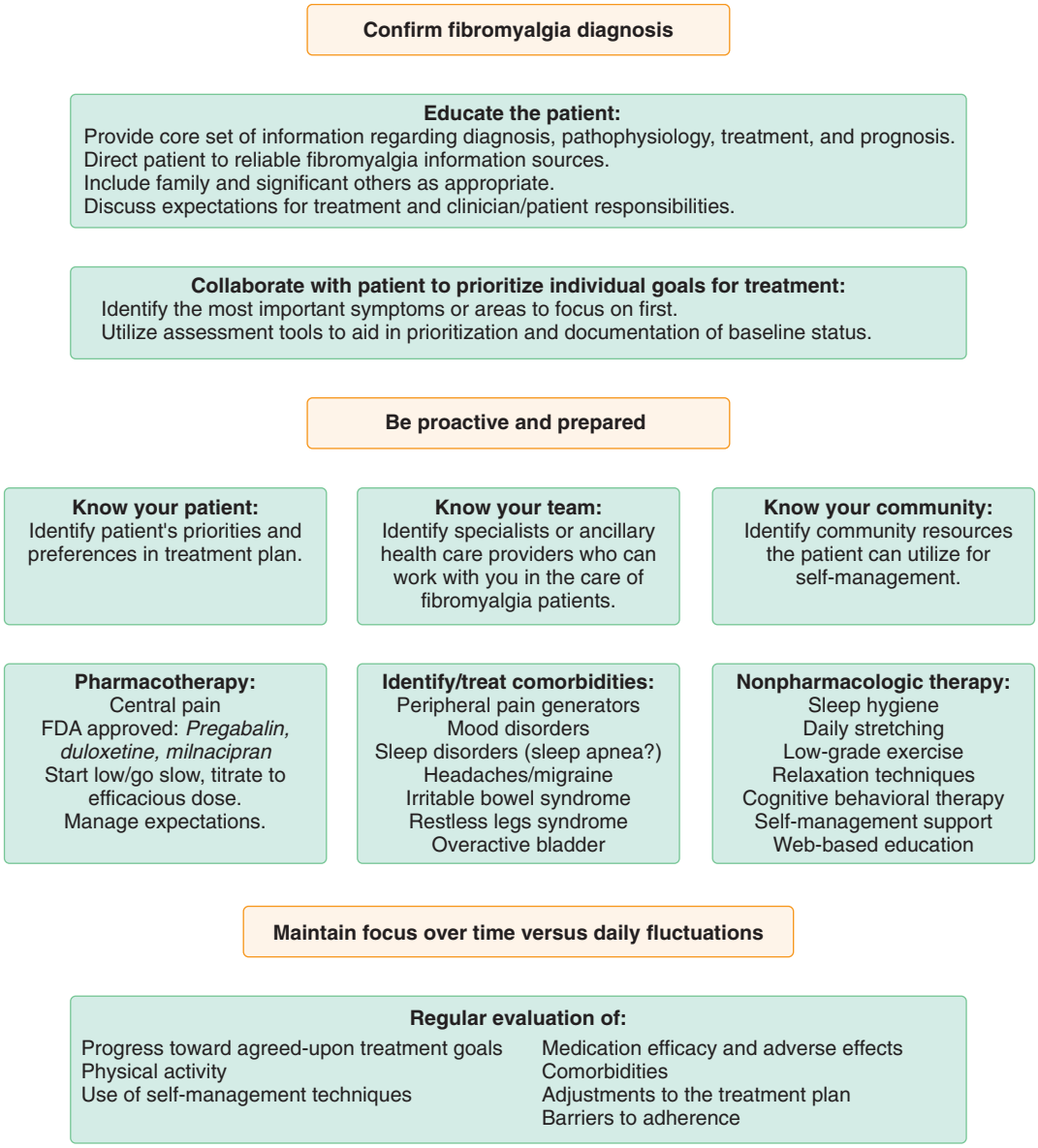

FIGURE 274-3. Management algorithm for treatment of fibromyalgia. (Adapted from Arnold LM, Clauw DJ, Dunegan LJ, Turk DC. A framework for fibromyalgia management for primary care providers. *Mayo Clin Proc.* 2012;87[5]:488-496.)

In these patients, short-acting benzodiazepine-like medications such as zolpidem, zapelon, and eszopiclone may be beneficial. Sodium oxybate is proving to be a useful medication in improving the nonrestorative sleep of FM. About 25% of male and 15% of female FM patients have sleep apnea (Chapter 405), which usually requires treatment with continuous positive airway pressure or surgery. By far the most common sleep disorder in FM patients is restless legs syndrome or periodic limb movement disorder. Treatment is with dopamine agonists such as L-dopa/carbidopa, pramipexole, or ropinirole.

Psychological Distress

Stressors related to psychosocial/economic and health issues often develop in FM patients. Psychological intervention in terms of improving the internal locus of control and more effective problem solving is important in such patients. Cognitive-behavioral therapy is particularly well suited to effect these changes, although it does not affect pain. Although antidepressant medications are commonly used in the treatment of pain and disordered sleep in FM patients, the doses are often suboptimal for treating depressive illness.

Deconditioning

FM is invariably aggravated by excessive exertion. However, a carefully graded program of aerobic conditioning and stretching is a critical component of an effective FM treatment program. Daily stretching, progressive walking, and simple strength training have been shown to be an important component of FM treatment. In general, exercise needs to be incrementally added to the program *after* some control of pain, sleep, and depression has been achieved.

Endocrine Dysfunction

There is no good evidence that FM is primarily due to an endocrine disorder. However, common problems such as hypothyroidism and menopausal symptoms often aggravate pain and fatigue, and appropriate replacement therapy is frequently worth a trial. A subset of fibromyalgia patients have adult growth hormone deficiency and show a worthwhile response to growth hormone therapy.

PROGNOSIS

FM symptoms usually persist for many years. A 5-year longitudinal study of 1550 FM patients undergoing standard treatment by U.S. rheumatologists reported substantial improvement in 10%, moderate improvement in 15%, and a worsening of symptoms in 39%. The majority of patients continued with high levels of self-reported symptoms and distress, with about 25% having a trend toward improvement.[8] In general, recommended medications can be expected to effect a 30% improvement in about 30% of patients. Difficulty in accepting an FM diagnosis, hypervigilance, high levels of psychological distress, and poor social support are poor prognostic factors. The consequences of pain, fatigue, and cognitive dysfunction negatively influence the sustained performance of physical and mental tasks. Everyday activities

TABLE 274-2 SELF-REPORTED SYMPTOMS IN CHRONIC FATIGUE SYNDROME

SYMPTOM	FREQUENCY (%)
Non-refreshing sleep	97
Memory/concentration problems	94
Pain in two or more joints	90
Muscle pain	89
Muscle discomfort	87
Difficulty thinking	85
Sleep problems	85
Fatigue after exercise (>24 hr)	81
Migratory joint pain	76
Unexplained muscle weakness	75
Intolerance to exercise	72
Anxiety	71
Malaise after exertion (>24 hr)	69
Sweatiness/cold hands and feet	66
Light/noise sensitivity	66
Headaches	65
Intolerance to be on your feet	61
Difficulty in understanding things	58
Sore throat	57
Tender glands in the neck/armpits	57
Depression	56
Confusion or disorientation	55
Mild fever or chills	53
Migraine	28

From Nacul LC, Lacerda EM, Pheby D, et al. Prevalence of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) in three regions of England: a repeated cross-sectional study in primary care. *BMC Med.* 2011;9:91.

take longer for FM patients, who need more time to get started in the morning and often require extra rest periods during the day.

CHRONIC FATIGUE SYNDROME

CFS, also referred to as myalgic encephalomyelitis (ME), is a poorly understood disorder characterized by a subacute onset of disabling fatigue and other defining symptoms (Table 274-2).[9] There are no conclusive diagnostic tests or generally effective treatments.

EPIDEMIOLOGY

Fatigue is a common symptom; indeed, some 50% of individuals report "feeling fatigued" in population surveys. Fatigue is one of the most common problems reported to primary care doctors. In a series of 1000 consecutive patients seen in primary care, 8.5% reported debilitating fatigue lasting 6 months or longer without apparent cause; only 15% satisfied the clinical definition for CFS shown in Table 274-3. In many cases, fatigue is self-limited, or the causes are self-evident (e.g., insufficient rest, a medical illness, depression, or insomnia). On the other hand, well-documented CFS is relatively rare, with a frequency of 0.006 to 3.0% in various population surveys. CFS is diagnosed predominantly in women aged 30 to 55 years. They are typically highly functional before the onset of the illness.

PATHOBIOLOGY

The cause of CFS remains poorly defined. There are many reports linking CFS with chronic infections; most have not stood up to careful scrutiny. Although many CFS patients complain of nonrestorative sleep, polysomnographic studies have been inconclusive (Chapter 405). The similarity between CFS and early Addison disease has prompted numerous neuroendocrine studies. They have failed to show a clinically treatable endocrine deficiency. About 30% of CFS patients have abnormal results on tilt-table testing with evidence of autonomic dysfunction in terms of neurally mediated hypotension or postural orthostatic tachycardia syndrome. This may be of relevance

TABLE 274-3 1994 CFS INTERNATIONAL STUDY GROUP CRITERIA

A diagnosis of CFS requires the following features:

1. Persistent chronic fatigue (at least 6 months) or intermittent, unexplained chronic fatigue, which relapses, or with a definite start, and is not the result of recent exertions. The fatigue is not improved by rest and results in a significant reduction in the patient's previous level of activity.
2. Exclusion of other diseases that may cause chronic fatigue *plus* four of the following eight minor criteria that have been present concurrently for 6 months or longer, after the onset of fatigue:
   - A. Recently impaired memory or concentration.
   - B. Pain on swallowing
   - C. Painful axillary or cervical lymph nodes
   - D. Muscle pains
   - E. Joint pain without swelling
   - F. Headache with a new pattern or increased severity
   - G. Sleep that does not improve after rest
   - H. Postexertional discomfort lasting more than 24 hours

From Morris G, Maes M. Case definitions and diagnostic criteria for myalgic encephalomyelitis and chronic fatigue syndrome: from clinical-consensus to evidence-based case definitions. *Neuro Endocrinol Lett.* 2013;34(3):185-199.

in a subpopulation of patients, but treatment with fludrocortisone was of no benefit in one large placebo-controlled study. Many studies have suggested a low level of immune stimulation in CFS, the most common abnormalities being reduced natural killer cell function and increased numbers of cytotoxic T cells. However, these findings are not diagnostically useful, and clinical improvement has not been associated with any significant change in immune markers. Extensive psychological testing in CFS patients has failed to reveal a common psychiatric denominator. MRI studies[10] have reported significant reductions in gray-matter volume, which are consistent with the complaint of impaired memory; they also suggest subtle abnormalities in visual processing and discrepancies between intended actions and consequent movements (E-Fig. 274-1).[11]

CLINICAL MANIFESTATIONS

The chief complaint is an abrupt onset of debilitating fatigue that causes significant impairment in daily activities, work ability, and social relationships. Typically, these patients report excellent health status just before the onset of fatigue, which is often preceded by a flulike prodrome. Symptoms suggestive of an infectious etiology include low-grade fever, night sweats, tender cervical lymph nodes, sore throat, myalgias, and headaches. As in FM, many CFS patients complain of generalized musculoskeletal pain, cognitive dysfunction, irritable bowel syndrome, temporomandibular pain disorders, nonrestorative sleep, and multiple chemical sensitivities (see Table 274-2).

DIAGNOSIS

There are several definitions of CFS; this has led to different clinical descriptions depending on the criteria used. The most widely used criteria are the 1994 CFS International Study Group Criteria.[12] Whereas the definition of FM contains no exclusionary criteria, the definition of CFS excludes patients with fatiguing medical disorders such as melancholic depression, psychotic disorders, substance abuse, and severe obesity (body mass index >40). Because CFS is not a diagnosis of exclusion, it is important to systematically evaluate the patient for treatable causes of fatigue (see E-Table 274-1).

TREATMENT

Rx

Currently, there is no generally accepted pharmacologic intervention for patients with CFS. Antidepressants (HCAs and SSRIs) are of minimal benefit, although two studies have reported modest benefit with the use of monoamine oxidase inhibitors in patients with significant vegetative symptoms. Patients with presumed CFS should have a general medical work-up to exclude treatable causes of fatigue (e.g., sleep apnea, hypothyroidism, chronic infection, anemia, orthostatic hypotension). Patients with poorly defined chronic illnesses require affirmation that they have a real illness and that the clinician will provide them with ongoing empathetic care. Education is a critical component in managing CFS; acceptance of having a chronic disorder, avoidance of catastrophizing, and adoption of activity pacing are important facets of

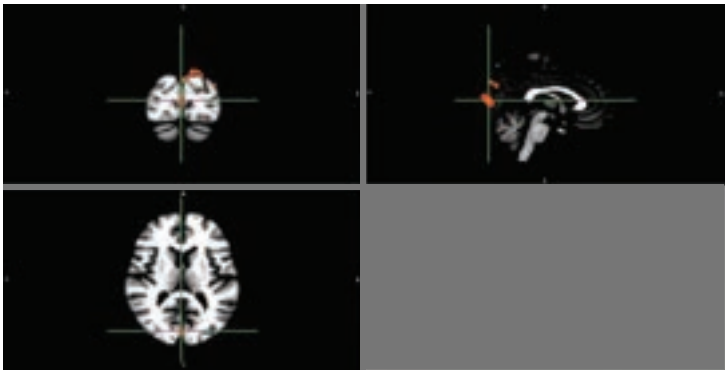

E-FIGURE 274-1. Regional gray- and white-matter volumetric changes in chronic fatigue syndrome. High-resolution structural 3T cerebral magnetic resonance images depicting voxel volume in the gray matter in chronic fatigue syndrome patients compared with healthy controls. Significantly reduced gray-matter volume *(P* < .05) was noted in the occipital lobes; left lateral occipital cortex, superior division; left primary visual cortex; right angular gyrus; and left parahippocampal gyrus. This figure shows *P*-value maps depicting clusters of gray matter volume reduction (red areas). (From Puri BK, Jakeman PM, Agour M, et al. Regional grey and white matter volumetric changes in myalgic encephalomyelitis [chronic fatigue syndrome]: a voxel-based morphometry 3 T MRI study. *Br J Radiol.* 2012;85[1015]:e270-e273.)

living with CFS[13]; such problems are often helped by cognitive behavioral therapy.[14] The mainstay of treatment is to engage the patient in a gently graded program of aerobic exercise.

PROGNOSIS

The prognosis for a full recovery varies between 0 and 37%; substantial improvement rates vary between 6 and 62%. Younger patients and those without a significant psychiatric overlay or catastrophizing have the best prognosis.

MYOFASCIAL PAIN

Pain arising from muscle is a universal human experience. In most instances, it is due to muscle *macrotrauma* (i.e., a muscle tear or sprain) or muscle *microtrauma* (i.e., injury at the sarcomere level resulting from repetitive muscle use or overexertion). After appropriate rest, the pain usually dissipates over a period of a few days to weeks. In some cases, a persistent pain focus develops that has the characteristics of a myofascial trigger point. This diagnosis should be considered if the patient complains of focal pain that is aggravated by muscle use or psychological stressors that cause increased muscle tension. Myofascial trigger points often play a role in poorly characterized focal pain syndromes such as low back pain, jaw pain, pelvic pain, and headache.[15]

A myofascial trigger point is a well-defined point of focal tenderness within a muscle (Fig. 274-4). Palpation usually reveals a ropelike induration referred to as a "taut band." In many instances, firm palpation of this area causes pain in a referred distribution that reproduces the patient's symptoms. Importantly, referred pain from a trigger point does not follow a nerve root distribution (i.e., it is not dermatomal). Trigger points frequently cause dysfunction in terms of a restriction in range of movement and weakness; the involved muscle or muscles often demonstrate easy fatigability (Table 274-4).

Electromyographic recordings from trigger points show spontaneous low-voltage activity resembling end-plate spike potentials. In normal muscle, depolarization of the motor end plate initiates the release of calcium ions from the sarcoplasmic reticulum, which in turn results in activation of the myosin-actin contractile elements ("contraction coupling"). It is thought that the areas of electrical hyperexcitability in trigger points result from a nonphysiologic focal contraction caused by an influx of calcium ions into the damaged sarcomeres. If this process is not "switched off," excessive utilization of adenosine triphosphate may result in a focal energy crisis in the injured muscle and thus perpetuate the problem. Functional MRI studies have confirmed the hyperalgesic features of myofascial trigger points in terms of increased activity of brain regions involved in processing stimulus intensity and negative affect. Microdialysis of trigger point foci have shown, compared with adjacent muscle, a more acidic pH and an increase of inflammatory cytokines, bradykinin, substance P, and calcitonin gene–related peptide.

Factors commonly cited as predisposing to trigger point formation include deconditioning, poor posture, repetitive mechanical stress, mechanical imbalance (e.g., leg length inequality), joint disorders, and nonrestorative sleep. The muscle pain experienced by FM patients is often the result of pain originating from myofascial trigger points and amplified by central sensitization.[16] It is hypothesized that myofascial pain is an important "peripheral pain generator" that may, in some FM patients, perpetuate and accentuate the process of central sensitization. The characteristic tender points used in FM diagnosis are in fact myofascial trigger points.

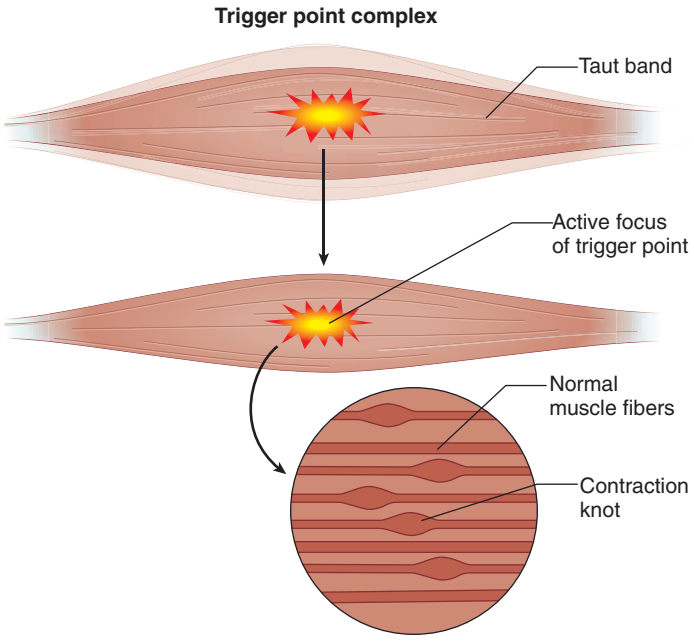

FIGURE 274-4. Schematic depiction of a myofascial trigger point. Depiction of a trigger point complex as seen in a longitudinal section of muscle. The top component represents a muscle with a taut band. The middle component represents a magnified view of the taut band containing an active trigger point focus. The lower component represents further magnification of the taut band and trigger point focus showing contraction knots (contracted sarcomere units). It is envisaged that these contraction knots are responsible for the nodularity of the taut band. Effective needling of myofascial trigger point requires piercing each contraction knot within the taut band. (From Bennett R. Myofascial pain syndromes and their evaluation. *Best Pract Res Clin Rheumatol.* 2007;21:427-445.)

TREATMENT

Rx

Management of myofascial trigger points is based on the following principles:

Postural and Ergonomic

An important issue in the effective management of myofascial pain syndromes is correction of predisposing factors (see earlier). These factors interfere with the ability of the muscle to fully recover and are the most common reason for treatment failure.

Stretching

Restoration of a muscle to its full stretch length breaks the link between the energy crisis and contraction of injured sarcomeres (see earlier). Commonly used stretching techniques include spray and stretch with ethyl chloride, acupressure, post-isometric relaxation, and deep stroking massage.

Strengthening

Muscles harboring trigger points usually become weak because of the inhibitory effects of pain. A program of slowly progressive strengthening is essential to restore full function and minimize the risk for recurrence and perpetuation of satellite trigger points.

Trigger Point Injections

Needling the myofascial trigger point with a "peppering technique" often provides a worthwhile and long-lasting benefit. Although dry needling is effective, the use of a local anesthetic (1% lidocaine or 1% procaine) helps confirm the accuracy of the injection and provides immediate relief. Validation of the accuracy of the injection is suggested by the patient's report that the pain is reproduced on entry of the needle into the trigger point; in superficial muscles a local twitch response may be observed, and this provides further evidence of an accurate injection.

Medications

Treatment of myofascial trigger points is mainly nonpharmacologic. Nonsteroidal anti-inflammatory drugs and other analgesics often provide moderate symptomatic relief. As in FM, drugs that modulate pain at the central level are a useful adjunct in difficult-to-treat patients, especially if central sensitization is suspected.

FUTURE DIRECTIONS

Both FM and CFS represent a major challenge to modern medicine in understanding a constellation of common, often disabling symptoms that originate

TABLE 274-4 DIAGNOSTIC FEATURES OF A MYOFASCIAL TRIGGER POINT

- Focal point of tenderness on palpation of the muscle involved
- Reproduction of pain complaint by trigger-point palpation (about 3-kg pressure)
- Palpation reveals an induration of the adjacent muscle (the "taut band")
- Often pseudo-weakness of the muscle involved (no atrophy)
- Often referred pain on continued (at least 5 seconds' duration) pressure over trigger point

within the central nervous system but cannot be explained in terms of a classic psychological disorder. The central sensitization demonstrated in FM patients has provided important insight into the physiologic basis for the complaint of widespread pain, but no such breakthrough has occurred in the comprehensive understanding of fatigue in either FM or CFS. Current research suggests that both these conditions have a genetic predisposition that interacts with environmental insults such as infections, trauma, myofascial foci, and psychological stressors. In the next few years, epigenetic studies should start to unravel the complex interaction of environment and genes. Both these disorders are prime examples of the need to integrate the classic biomedical model of disease with psychosocial influences (i.e., the biopsychosocial model of disease).

Grade A References

A1. Moore A, Wiffen P, Kalso E. Antiepileptic drugs for neuropathic pain and fibromyalgia. *JAMA*. 2014;312:182-183.

A2. Bernardy K, Fuber N, Kollner V, et al. Efficacy of cognitive-behavioral therapies in fibromyalgia syndrome: a systematic review and metaanalysis of randomized controlled trials. *J Rheumatol*. 2010;37:1991-2005.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ossipov MH, Dussor GO, Porreca F. Central modulation of pain. *J Clin Invest*. 2010;120:3779-3787.
2. Abeles AM, Pillinger MH, Solitar BM, et al. Narrative review: the pathophysiology of fibromyalgia. *Ann Intern Med*. 2007;146:726-734.
3. Feng J, Zhang Z, Wu X, et al. Discovery of potential new gene variants and inflammatory cytokine associations with fibromyalgia syndrome by whole exome sequencing. *PLoS ONE*. 2013;8:e65033.
4. Clauw DJ. Fibromyalgia: a clinical review. *JAMA*. 2014;311:1547-1555.
5. Calandre EP, Vilchez JS, Molina-Barea R, et al. Suicide attempts and risk of suicide in patients with fibromyalgia: a survey in Spanish patients. *Rheumatology*. 2011;50:1889-1893.
6. Bennett RM, Friend R, Marcus D, et al. Criteria for the diagnosis of fibromyalgia: validation of the modified 2010 preliminary American College of Rheumatology criteria and the development of alternative criteria. *Arthritis Care Res (Hoboken)*. 2014;66:1364-1373.
7. Arnold LM, Clauw DJ, Dunegan LJ, et al. A framework for fibromyalgia management for primary care providers. *Mayo Clin Proc*. 2012;87:488-496.
8. Walitt B, Fitzcharles MA, Hassett AL, et al. The longitudinal outcome of fibromyalgia: a study of 1555 patients. *J Rheumatol*. 2011;38:2238-2246.
9. Nacul LC, Lacerda EM, Pheby D, et al. Prevalence of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) in three regions of England: a repeated cross-sectional study in primary care. *BMC Med*. 2011;9:91.
10. Zeineh MM, Kang J, Atlas SW, et al. Right arcuate fasciculus abnormality in chronic fatigue syndrome. *Radiology*. 2015;274:517-526.
11. Puri BK, Jakeman PM, Agour M, et al. Regional grey and white matter volumetric changes in myalgic encephalomyelitis (chronic fatigue syndrome): a voxel-based morphometry 3 T MRI study. *Br J Radiol*. 2012;85:e270-e273.
12. Morris G, Maes M. Case definitions and diagnostic criteria for myalgic encephalomyelitis and chronic fatigue syndrome: from clinical-consensus to evidence-based case definitions. *Neuro Endocrinol Lett*. 2013;34:185-199.
13. Goudsmit EM, Nijs J, Jason LA, et al. Pacing as a strategy to improve energy management in myalgic encephalomyelitis/chronic fatigue syndrome: a consensus document. *Disabil Rehabil*. 2012;34:1140-1147.
14. Poppe C, Petrovic M, Vogelaers D, et al. Cognitive behavior therapy in patients with chronic fatigue syndrome: the role of illness acceptance and neuroticism. *J Psychosom Res*. 2013;74:367-372.
15. Giamberardino MA, Affaitati G, Fabrizio A, et al. Myofascial pain syndromes and their evaluation. *Best Pract Res Clin Rheumatol*. 2011;25:185-198.
16. Dall'Agnol L, Medeiros LF, Torres IL, et al. Repetitive transcranial magnetic stimulation increases the corticospinal inhibition and the brain-derived neurotrophic factor in chronic myofascial pain syndrome: an explanatory double-blinded, randomized, sham-controlled trial. *J Pain*. 2014;15:845-855.

REVIEW QUESTIONS

1. A 45-year-old woman presents with a 2-year history of pain and swelling in her hands and aching muscles. Which of the following is the most likely diagnosis?

A. Rheumatoid arthritis
B. Fibromyalgia
C. Systemic lupus erythematosus
D. Hand osteoarthritis with concomitant fibromyalgia
E. Dercum disease

Answer: D Hand osteoarthritis is especially common in women, as is fibromyalgia. Although fibromyalgia patients often complain of joint swelling, this is never observed clinically. Thus, the most likely diagnosis is a combination of fibromyalgia and hand osteoarthritis. Early rheumatoid arthritis can have this presentation, but after 2 years, other joints, especially the feet, would be symptomatic. Dercum disease is a rare disorder characterized by multiple painful lipomas.

2. A 47-year-old woman presents with a 2-year history of pain and stiffness in her right shoulder without swelling. She has a full range of motion in the shoulder but has pain when stretching forward. She reports that she feels a "bump" in the upper shoulder muscles. Which of the following is the most likely diagnosis?

A. Polyarteritis nodosa
B. Fibromyalgia
C. A focal myofascial trigger point
D. Fibrosarcoma of the trapezius muscle
E. Polymyositis

Answer: C This is the typical presentation of a myofascial trigger point in the shoulder girdle muscles. The pain associated with myofascial trigger points is aggravated by full range of motion. Most patients with myofascial pain problems are aware of a "tight knot" in the involved muscles. Polymyositis is a generalized autoimmune disorder of muscle that usually presents as proximal muscle weakness; pain is seldom a major symptom.

3. Chronic fatigue syndrome (CFS) has a population prevalence of which of the following?

A. 2%
B. 8 to 10%
C. Less than 1%
D. 0.001% to 0.003%
E. 4 to 8%

Answer: C CFS is less prevalent than fibromyalgia, which has a prevalence of about 2% in the general population. However, most population studies in CFS report a prevalence of 0.006 to 3.0%. This is in contradistinction to a complaint of fatigue, which is one of the most common symptoms seen by family doctors.

4. The clinical features of myofascial trigger points often include all *except* which of the following?

A. Pain referred distally when the trigger point is palpated for 1 second
B. Pain referred distally when the trigger point is palpated for between 5 and 10 seconds
C. Restricted movement in involved muscle
D. A tight "knot" in the involved muscle
E. Pseudo-weakness of the involved muscle

Answer: A A critical clinical feature of a myofascial trigger point is distally referred pain after about 5 or more seconds of steady pressure over the tight knot. A common mistake in evaluating myofascial trigger points is to just prod them without exerting prolonged pressure. The patient needs to be asked to report pain or discomfort occurring in an area outside the tight knot during palpation; this referred pain is usually distal to the trigger point. This feature is even more marked when a trigger point is needled.

275

SYSTEMIC DISEASES IN WHICH ARTHRITIS IS A FEATURE

STERLING G. WEST

Arthritis, arthralgias, and myalgias can be significant features of several systemic diseases and may be the presenting symptoms for some of these disorders (Table 275-1). Appropriate evaluation of these musculoskeletal symptoms, including selected laboratory tests and radiographs, can provide clues to the early diagnosis of these diseases. Synovial biopsies are rarely necessary but can be diagnostic. Brief descriptions of the arthritic manifestations of some of these systemic disorders follow; a more detailed discussion of each entity is found in the chapters devoted to these diseases. Because of the rarity of many of these diseases, evidence-based treatments with U.S. Food and Drug Administration–approved medications are lacking.

TABLE 275-1 SYSTEMIC DISEASES ASSOCIATED WITH ARTHRITIS

DISEASE	TEST*
GASTROINTESTINAL DISEASES	
Autoimmune hepatitis	Liver-associated enzymes, *ASMA*
Primary biliary cirrhosis	Alkaline phosphatase, *antimitochondrial Ab*
Pancreatitis-arthritis syndrome	Lipase, amylase, *abdominal CT scan*
Whipple's disease	*Tissue biopsy, tissue immunohistochemical stain for Tropheryma whippelii, PCR for T. whippeli DNA*
Gluten-sensitive enteropathy	*Antitransglutaminase antibody, small bowel biopsy*
Inflammatory bowel disease	Stool guaiac, *colonoscopy*
Hepatitis B/hepatitis C	Liver-associated enzymes, *hepatitis serology, cryoglobulins*
Intestinal bypass arthritis	*Cryoglobulins*
HEMATOLOGIC DISORDERS	
Hemophilia	PTT, *factor VIII and IX levels*
Hemoglobinopathies	CBC, *hemoglobin electrophoresis*
Hypogammaglobulinemia	Low total protein, *SPEP, immunoglobulins*
Plasma cell dyscrasias	High total protein, *SPEP, UPEP, IEF*
ENDOCRINE DISORDERS	
Diabetes mellitus	Glucose, *hemoglobin* A_{1c}
Thyroid disorders	TSH, *thyroxine*
Parathyroid disorders	Calcium, phosphorus, *PTH*
Acromegaly	Radiographs, *growth hormone, IGF-1*
Hyperlipoproteinemia	Lipid panel
Paget's disease	Bone-specific alkaline phosphatase, radiographs, *bone scan*
MALIGNANT DISORDERS	
Hypertrophic osteoarthropathy	Radiographs (hands, wrists, chest)
Leukemia and lymphoma	CBC, LDH, *bone marrow/tissue biopsy*
Carcinomatous polyarthritis	*Cancer screen*
Palmar fasciitis and arthritis	*CA-125, pelvic CT scan, cancer screen*
OTHER DISEASES	
Hemochromatosis	Iron studies, radiographs, *HFE gene*
Multicentric reticulohistiocytosis	Radiographs, *skin/synovial biopsy*
Sarcoidosis	Chest radiograph, *ACE level, tissue biopsy*
IgG4-related disease	Serum IgG4 level, *histopathology of biopsy specimens including IgG4 immunostaining*
Alkaptonuria	Radiographs, *urine homogentisic acid level*
Fabry's disease	Angiokeratomas, *α-galactosidase A level or gene mutation*
Relapsing polychondritis	*Cartilage biopsy*
Cystic fibrosis	Chest radiograph, *sweat chloride, CFTR gene mutation*
Tenosynovial giant cell tumor: diffuse type (diffuse pigmented villonodular synovitis)	Synovial fluid analysis, *MRI, synovial biopsy*
Systemic infections	Cultures, serologies (RPR, HIV, EBV, parvovirus)

*Tests listed are common laboratory tests and radiographs that are frequently ordered; this information should provide a clue that a systemic disease is a possible cause of the patient's musculoskeletal symptoms. These tests, coupled with the history and physical examination, should be followed by more specific tests and biopsies (listed in italics) to confirm the diagnosis.
Ab = antibody; ACE = angiotensin-converting enzyme; ASMA = anti–smooth muscle antibody; CBC = complete blood cell count; CFTR = cystic fibrosis transmembrane conductance regulator; CT = computed tomography; EBV= Epstein Barr virus; HIV = human immunodeficiency virus; IEF = immunoelectrophoresis; IGF-1= insulin-like growth factor-1; IgG4 = immunoglobulin G4; LDH = lactate dehydrogenase; MRI = magnetic resonance imaging; PCR = polymerase chain reaction; PTH = parathyroid hormone; PTT = partial thromboplastin time; RPR = rapid plasmin reagin; SPEP = serum protein electrophoresis; TSH = thyroid-stimulating hormone; UPEP = urine protein electrophoresis.

AUTOIMMUNE HEPATITIS

Patients with type I autoimmune hepatitis (Chapter 149) may present with a syndrome resembling systemic lupus erythematosus (SLE; Chapter 266). Patients with the early-onset subset are frequently young and female, with complaints of polyarthralgia and occasionally fever. Laboratory examination may show leukopenia, a positive antinuclear antibody (70 to 90%), elevated erythrocyte sedimentation rate, polyclonal gammopathy, and elevated liver-associated enzymes. Antibodies against double-stranded DNA are usually not seen, whereas antibodies against the smooth muscle antigen (F1 actin) support the diagnosis. Joint radiographs show soft tissue swelling without erosions or deformity. Joint pain resolves with corticosteroid therapy for the liver disease. Patients with autoimmune hepatitis have an increased risk for having a concurrent autoimmune disease.[1]

PRIMARY BILIARY CIRRHOSIS

Up to 50% of patients with primary biliary cirrhosis (Chapter 155) have other autoimmune disorders, including rheumatoid arthritis (RA), Sjögren's syndrome, limited scleroderma, and autoimmune thyroiditis.[2] In addition to antimitochondrial antibodies, rheumatoid factor, antinuclear antibodies, and anticentromere antibodies are often present. More than 10% of patients with primary biliary cirrhosis have a symmetrical or asymmetrical small joint inflammatory arthritis. Unlike RA, it can involve distal interphalangeal joints and is rarely erosive or deforming. Other musculoskeletal manifestations include osteomalacia related to vitamin D deficiency, osteoporosis related to renal tubular acidosis, and hypertrophic osteoarthropathy associated with liver disease.

WHIPPLE'S DISEASE

An inflammatory arthritis occurs in 60 to 90% of patients with Whipple's disease (Chapter 140) and may precede other clinical manifestations by years.[3] The joint involvement is typically an intermittent, migratory oligoarthritis affecting large joints more than small joints or the spine, lasting from several hours to days. The synovial fluid is inflammatory, with a

predominance of mononuclear cells. Subcutaneous nodules are occasionally seen, contributing to an erroneous diagnosis of rheumatic fever or RA. However, patients consistently test negative for rheumatoid factor and antinuclear antibodies. Synovial biopsies show rod-shaped bacilli on electron microscopy, which have been identified as *Tropheryma whipplei*. Diagnosis is suspected when duodenal, synovial, or lymph node biopsies show periodic acid–Schiff–positive macrophages. Infection is confirmed by demonstration of the organism in tissue by immunohistochemical staining with antisera specific for *T. whipplei*. Quantitative polymerase chain reaction to detect *T. whipplei* DNA is used as a confirmatory test performed on tissue and body fluids. Typically, the arthritis does not cause radiographic changes or deformities. Prolonged antibiotic therapy results in resolution of musculoskeletal as well as other symptoms of this disease. Relapses, especially neurologic, can occur in up to 35% of patients after cessation of antibiotic therapy.

GLUTEN-SENSITIVE ENTEROPATHY (CELIAC DISEASE)

An asymmetrical oligoarthritis or symmetrical polyarthritis occurs in up to 25% of adults with celiac disease (Chapter 140). It may precede the enteropathic symptoms by months to years. Large joints such as knees and ankles, more than hips and shoulders, are most commonly involved. Axial involvement is reported. The arthritis does not cause deformities or radiographic changes and resolves with a gluten-free diet in 40 to 50% of cases. Another musculoskeletal manifestation is osteomalacia related to vitamin D malabsorption, which may mimic diffuse fibromyalgia.

PANCREATITIS-ARTHRITIS SYNDROME

Pancreatic panniculitis is a systemic syndrome occurring in some patients with pancreatic acinar cell carcinoma and less commonly in patients with pancreatitis or hematologic malignancies. This syndrome is characterized by tender red nodules, usually on the extremities; these are frequently misdiagnosed as erythema nodosum, but biopsy shows areas of lobular panniculitis with fat necrosis. Arthritis occurs in 60% of patients and usually involves the ankles and knees. Synovial fluid is typically noninflammatory and creamy in color. It contains multiple lipid droplets because of necrosis of fat in the synovial membrane. Other manifestations include osteolytic lesions (10%) from bone marrow fat necrosis, pleuropericarditis, fever, and eosinophilia. The prominent fat necrosis is due to the release of lipase, amylase, and trypsin from the diseased pancreas. Another musculoskeletal manifestation resulting from pancreatic disease is osteomalacia from vitamin D deficiency related to malabsorption.

HEMOPHILIA

Hemophilia A (factor VIII deficiency) and hemophilia B (factor IX deficiency) (Chapter 174) are associated with hemarthrosis.[4] Almost all patients with factor levels less than 1% of normal experience recurrent hemarthroses spontaneously or after minor trauma. Large joints (knees, elbows, ankles) are most commonly involved. Intramuscular hemorrhage can also occur. Recurrent hemarthrosis can lead to proliferative synovitis and cartilage degradation, resulting in both erosive and degenerative changes on radiographs. Physical examination shows bone enlargement, crepitus, atrophic muscles, and joint contractures. Treatment of acute monoarthritis consists of factor replacement to achieve a level of 30% or greater, given at the first sign of joint swelling. Patients with fever (temperature >38° C) or who fail to respond to factor replacement need joint aspiration to rule out septic arthritis, which occurs with an increased incidence in hemophilia. Chronic arthritis is treated with nonsteroidal anti-inflammatory drugs (NSAIDs), which do not inhibit platelet function; arthroscopic or radiation synovectomy for chronic synovitis; and total joint arthroplasty for end-stage joint disease. The regular prophylactic administration of factor replacement has reduced the risk for developing chronic arthropathy. Acute and chronic arthritis is less frequent and less severe in patients with hemophilia B compared with hemophilia A.

HEMOGLOBINOPATHIES

Patients with sickle cell anemia (Chapter 163) or the heterozygous states of sickle β-thalassemia and sickle–hemoglobin C disease frequently experience polyarthralgia. Local sickling of cells leads to obstruction of the microcirculation and to bone infarctions. Patients most commonly experience painful crises causing chest, back, and joint pain. A painful large joint arthritis (usually in the knees), often with noninflammatory synovial effusions, lasting days to 2 to 3 weeks can also occur. Infarcts in the metaphyses of bones are commonly found on joint radiographs. Vertebral bodies have a characteristic "Lincoln log" appearance or a central cuplike indentation ("codfish vertebrae"). Femoral and humeral head osteonecrosis can occur in up to 33% of sickle cell anemia and sickle–hemoglobin C disease cases. Because of splenic autoinfarction, septic arthritis (*Staphylococcus aureus*) and osteomyelitis (50% caused by *Salmonella*) have been associated with sickle cell disease. In adults, gout has been reported, whereas in children younger than 2 years, an acute, painful, nonpitting swelling of the hands and feet (hand-foot syndrome) associated with fever and leukocytosis may be the first manifestation of sickle cell anemia. Treatment includes intravenous hydration, oxygen, and analgesics. Hydroxyurea can reduce the frequency of painful crises. In patients with β-thalassemia major (Cooley's anemia; Chapter 162), significant expansion of bone marrow develops as a result of increased erythroid precursors, leading to osteoporosis and microfractures that affect primarily the lower extremities. Chelation therapy with deferiprone (to reduce iron overload from transfusions) can cause arthralgias in 20% of patients.

HYPOGAMMAGLOBULINEMIA

Patients with congenital X-linked hypogammaglobulinemia (Bruton's disease) or acquired common variable immunodeficiency (CVID; Chapter 250) can develop a nonerosive, noninfectious large joint oligoarthritis that responds to intravenous gamma globulin therapy. However, septic arthritis caused by common pathogens or *Mycoplasma* can also occur and must be rigorously excluded. In adults with acquired CVID, autoimmune conditions are also common (22%), including autoimmune cytopenias and pernicious anemia. Selective immunoglobulin A (IgA) deficiency (Chapter 250) is associated with various rheumatic manifestations, including positive autoantibodies, in the absence of clinical disease. Systemic autoimmune disorders, including SLE, juvenile idiopathic arthritis, and others, as well as organ-specific autoimmune disorders such as type 1 diabetes mellitus and myasthenia gravis, also occur in IgA-deficient individuals.

DIABETES MELLITUS

Diabetic stiff hand syndrome of limited joint mobility (diabetic cheiroarthropathy) occurs in more than 30% of patients with long-standing, poorly controlled type 1 or type 2 diabetes mellitus (Chapter 229).[5] Patients present with the insidious development of flexion contractures and thickened skin of the fingers, which may be confused with scleroderma. These changes may be due to excess glycosylation of tendinous structures and accumulation of sugar alcohols, producing excess water content in the tissues and leading to increased stiffness. As a result of the inability to extend the fingers fully, the "prayer sign" is observed on physical examination. Unlike diabetic stiff hand syndrome, Dupuytren's contractures are due to a chronic thickening of the palmar aponeurosis, causing flexion deformities of the third and fourth digits. It is a frequent musculoskeletal complication, occurring in more than 20% of type 2 diabetic patients. A less common manifestation is Charcot's, or neuropathic, joints, occurring in less than 1% of all patients with long-standing diabetes. All patients have a diabetic peripheral neuropathy and typically present with painless swelling of the feet caused, most commonly, by destruction of the tarsometatarsal joints. Deformities can occur with midtarsal collapse ("rocker bottom" feet), predisposing to ulceration and infection of the skin over desensate bony prominences. Radiographs are diagnostic, and treatment should include supportive footwear and protected weight bearing.

Unlike Charcot's joint, diabetic osteolysis and diabetic amyotrophy are unique to diabetes. The osteolysis is characterized by resorption of the distal metatarsal bone and proximal phalanges of the feet, giving radiographs a characteristic "licked candy" appearance. Pain is variable, and treatment is conservative because the process may terminate on its own. Diabetic amyotrophy is a lumbar polyradiculopathy (L2 to L4) that arises with severe pain, dysesthesias, and rapid atrophy of the proximal muscles of one or both thighs. Carpal tunnel syndrome (25%), adhesive capsulitis of the shoulder (frozen shoulder), flexor tenosynovitis (trigger finger) of the hands, diffuse idiopathic skeletal hyperostosis (type 2 diabetes), osteopenia (type 1 diabetes), diabetic muscle infarction (usually of the thigh), osteomyelitis of the foot, and septic joints are all musculoskeletal conditions that occur with increased frequency in diabetic patients. Aggressive control of blood glucose helps prevent some of these musculoskeletal complications.

THYROID DISORDERS

Musculoskeletal symptoms occur in 33% of patients with clinical hypothyroidism (thyroid-stimulating hormone levels >20 μU/mL) (Chapter 226). Patients can present with carpal tunnel syndrome, Raynaud's phenomenon, or muscle aching and stiffness similar to fibromyalgia and polymyalgia

rheumatica. Patients with severe hypothyroidism can experience a noninflammatory myopathy with proximal muscle weakness and elevated creatine kinase, which may be confused clinically with polymyositis. Similarly, myxedematous patients can develop a symmetrical arthropathy of the large joints, especially the knees, associated with noninflammatory synovial fluid with increased viscosity. The association of hypothyroidism with chondrocalcinosis is controversial, but clearly patients beginning thyroid replacement therapy can experience an acute attack of pseudogout. Patients with hyperthyroidism can develop proximal myopathy (70%), adhesive capsulitis of the shoulder (10%), osteoporosis, or thyroid acropachy. Thyroid acropachy occurs in less than 1% of patients with Graves' disease and consists of soft tissue swelling of the hands, digital clubbing, and periostitis, particularly involving the metacarpal and phalangeal bone shafts. Pain is usually mild, radiographs are characteristic, and there is no effective therapy. Patients with autoimmune thyroid disease have an increased prevalence of positive antinuclear antibodies and an increased association with systemic connective tissue diseases such as Sjögren's syndrome.

PARATHYROID DISORDERS

Primary hyperparathyroidism (Chapter 245) can develop with osteoporosis and fractures or with chondrocalcinosis and episodes of acute pseudogout.[5] In severe hyperparathyroidism, which is rare, vague myalgias and arthralgias resembling fibromyalgia; a reversible, painless, proximal myopathy with normal creatine kinase; and osteitis fibrosa cystica with bone pain can be seen. Osteitis fibrosa cystica occurs primarily in patients with secondary hyperparathyroidism associated with renal failure and has a characteristic radiographic appearance, with subperiosteal resorption on the radial side of the phalanges, small erosions in the hands and distal clavicles, and discrete lytic bone lesions (brown tumors). Ectopic calcifications, joint laxity, and tendon ruptures have been reported in patients with severe hyperparathyroidism. Hypoparathyroidism has also been associated with myopathy and ectopic calcifications. Patients with type Ia pseudohypoparathyroidism and pseudo-pseudohypoparathyroidism have a shortened fourth metacarpal bone bilaterally.

ACROMEGALY

Up to 75% of patients with acromegaly (Chapter 224) develop an atypical form of osteoarthritis. The knees, shoulders, hips, and lumbosacral and cervical spine are the most frequently symptomatic areas, although the hands reveal the most characteristic radiographic changes, with widened joint spaces due to cartilage hypertrophy. Carpal tunnel syndrome (50%), Raynaud's phenomenon (33%), and proximal muscle weakness with normal creatine kinase can also occur.

HYPERLIPOPROTEINEMIA

Type IIa familial hyperlipidemia (Chapter 206) is associated with tendinous and tuberous-osseous xanthomas as well as episodic Achilles tendinitis. An acute migratory, inflammatory arthritis persisting up to a month and resembling rheumatic fever occurs in up to 50% of patients. Predominantly large joints are affected. In addition, a self-limited, acute monoarticular or oligoarticular arthritis involving the knee or ankle can occur. Patients with type III familial hyperlipoproteinemia can develop tendon and bone xanthomas. Patients with human immunodeficiency virus (HIV) infection taking protease inhibitor drugs can develop dyslipidemia leading to tendon xanthomas. In all hyperlipidemias, gout must be excluded before ascribing the symptoms to hyperlipoproteinemia. Therapy with NSAIDs and treatment of the underlying lipid disorder should be pursued. Notably, some of the therapies used to treat hyperlipidemia can cause musculoskeletal symptoms, including hyperuricemia and gout from nicotinic acid and myalgias (with or without elevated creatine kinase) from statin therapy.

PAGET'S DISEASE

Paget's disease (Chapter 247) can cause bone pain and deformity. An elevated bone-specific alkaline phosphatase and characteristic radiographic changes can help make the diagnosis. Joint pain caused by secondary osteoarthritis in areas of bone involvement by Paget's disease most commonly occurs in the hips, knees, or vertebrae. Spinal stenosis from Paget's disease of the spine has been reported. Bisphosphonate therapy is highly effective.

HYPERTROPHIC OSTEOARTHROPATHY

Hypertrophic osteoarthropathy is a syndrome that includes clubbing of the fingers and toes, periostitis of long bones (distal tibia, femur, radius), and arthritis (Fig. 275-1). Hypertrophic osteoarthropathy is classified into primary (hereditary) and secondary forms. Between 80 and 90% of secondary hypertrophic osteoarthropathy is associated with intrathoracic neoplasms, especially non–small cell lung cancer.[6] Other causes include other neoplasms, chronic pulmonary infections, congenital heart disease, cirrhosis, HIV infection, medications (voriconazole), and inflammatory bowel disease. Patients with secondary hypertrophic osteoarthropathy can present with acute, severe, burning bone pain and a noninflammatory arthritis caused by periarticular periostitis. Pain is accentuated by dependency of the limbs. Pitting edema, warmth, and tenderness of the legs and forearms can be seen. Radiographs show diagnostic changes of periosteal elevation, new bone formation, or both along the distal ends of long bones. Therapy is symptomatic, and hypertrophic osteoarthropathy improves with successful treatment of the underlying primary disease. In resistant cases, treatment with intravenous bisphosphonate has been effective in modulating symptoms.

LEUKEMIA AND LYMPHOMA

Leukemia can arise as an asymmetrical or migratory polyarthritis, monoarthritis (rare), back pain (10%), or nocturnal bone pain. Articular manifestations occur in 14 to 50% of children and 4 to 16% of adults with acute leukemia and can precede the diagnosis by months. Joint pain is attributed to leukemic synovial infiltration and usually involves the ankle or knee, but it can be polyarticular, resembling juvenile or adult RA. The joint pain is disproportionately more severe than the clinical findings. Synovial effusions are uncommon, and evidence of leukemic cells in the synovial fluid is rare. Bone pain due to subperiosteal leukemic cell infiltration occurs in up to 50% of patients, with long bone pain (lower extremities) more common in children and back pain more common in adults. Radiographs are normal in 50%

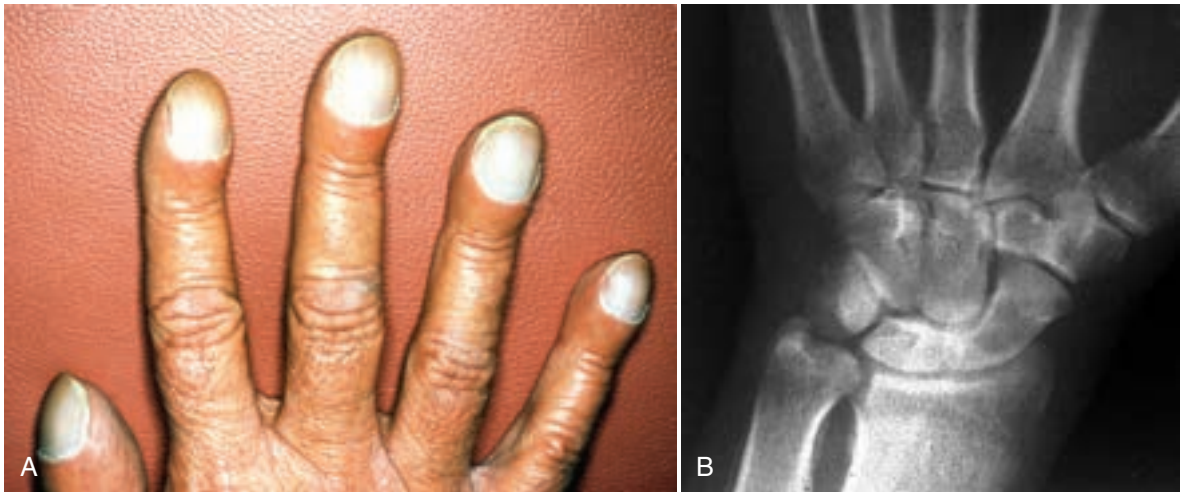

FIGURE 275-1. Hypertrophic osteoarthropathy. A, Severe clubbing of the nails. **B,** Radiograph demonstrating periosteal elevation of the distal radius and ulna.

of cases. The musculoskeletal symptoms are poorly responsive to NSAIDs but can resolve with successful therapy of the leukemia. Musculoskeletal symptoms occur in 25% of patients with non-Hodgkin's lymphoma. Nocturnal bone pain is the most common presenting musculoskeletal complaint. A seronegative monoarthritis or polyarthritis can occur and should be suspected in patients with severe constitutional symptoms or lymphadenopathy out of proportion to the degree of arthritis. Patients with angioimmunoblastic T-cell lymphoma (Chapter 185) may occasionally develop a chronic, nonerosive polyarthritis with erythroderma.

CARCINOMATOUS POLYARTHRITIS

Polyarthritis can rarely (<2%) be the presenting manifestation of an occult malignancy; it may precede the discovery of the malignancy by several months.[7] Breast, colon, lung, ovarian, and lymphoproliferative malignancies are the most commonly associated cancers. Clinical features suggesting carcinomatous polyarthritis include the explosive onset of a rheumatoid factor–negative, asymmetrical polyarthritis involving predominantly the lower extremities and sparing the hands and wrists in a patient older than 60 years. Polymyalgia rheumatica and RA must be excluded. Treatment of the underlying malignancy results in improvement of the arthritis.

PALMAR FASCIITIS AND ARTHRITIS SYNDROME

Ovarian carcinoma (Chapter 199) is the most common malignancy found in patients with palmar fasciitis and arthritis. This musculoskeletal manifestation can also be seen in patients with breast, gastric, or pancreatic cancer. Patients present with a severe, painful, symmetrical inflammatory polyarthritis and fasciitis causing contractures primarily of the hands and, less commonly, the feet. Patients may have vasomotor instability, causing diagnostic confusion with complex regional pain syndrome or RA. This syndrome portends a poor prognosis because it typically manifests after tumor metastasis. Response to treatment is poor, although clinical improvement can occur with successful eradication of the underlying tumor.

HEMOCHROMATOSIS

Joint involvement occurs in 40 to 75% of patients with hereditary hemochromatosis (Chapter 212) and may be the presenting symptom (Fig. 275-2).[8] The metacarpophalangeal (MCP) joints (especially the second and third MCP joints), wrists, knees, hips, shoulders, and ankles are most often involved in a symmetrical pattern. The arthropathy resembles osteoarthritis, with joint swelling resulting from bone enlargement, but it is distinguished clinically by the involvement of atypical joints, such as MCP joints, wrists, and ankles. Radiographs show joint space narrowing, subchondral cysts, sclerosis, and osteophytes that are hooklike at the MCP joints. Chondrocalcinosis is present in up to 50% of patients. It is typically asymptomatic, but in some patients it leads to attacks of acute inflammatory synovitis (pseudogout), which may result in the misdiagnosis of RA. The prevalence of overt arthritis increases with age, and it may be only minimally symptomatic when the disease arises in other organs. However, it is not uncommon for articular pain to be the initial presenting complaint (33%). Consequently, all patients (especially male) presenting with premature osteoarthritis occurring in atypical joints, especially MCP joints and wrists, should be screened for hereditary hemochromatosis with iron studies. The mechanism whereby iron causes arthritis is unclear, but it may be related to hemosiderin deposits in the synovial membrane and chondrocytes activating degradative enzymes. Treatment is symptomatic with NSAIDs and, when severe, total joint arthroplasties. Phlebotomy for iron removal does not alter the course of the arthritis. Additional rheumatic manifestations in patients with hemochromatosis include osteoporosis related to hypogonadotropic hypogonadism, osteomalacia related to vitamin D deficiency when liver disease is severe, and an increased susceptibility to *Yersinia* septic arthritis.

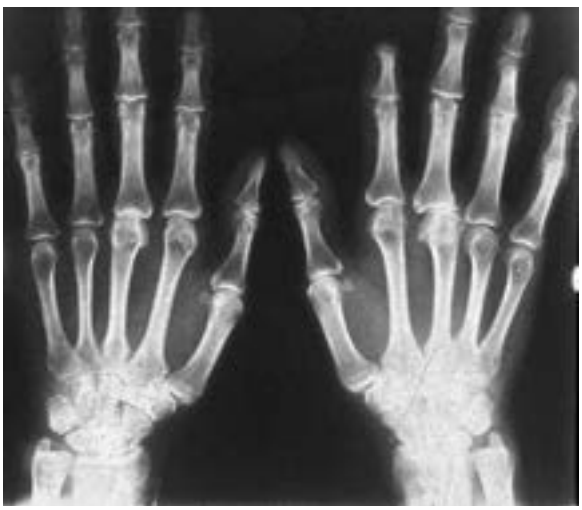

FIGURE 275-2. Hemochromatotic arthropathy. Radiograph demonstrating degenerative changes, with hooklike osteophytes of the second and third metacarpophalangeal joints bilaterally.

MULTICENTRIC RETICULOHISTIOCYTOSIS

Multicentric reticulohistiocytosis (MRH) is a chronic, symmetrical, inflammatory polyarthritis most commonly affecting the hands and cervical spine.[9] It may resemble RA but can be differentiated by its prominent distal interphalangeal joint synovitis. Joint involvement remits and relapses initially, but in 50% of cases it worsens into a severely deforming arthritis mutilans. Firm, nonpruritic, reddish brown or yellow papulonodular lesions ("coral beads") that wax and wane occur around the nail beds and on the face, hands, ears, and other areas predominantly above the waist. The skin lesions have a diagnostic histology. In 50 to 66% of patients, these diagnostic nodules follow the onset of arthritis by months to years. Additional associations include xanthelasma (33%) and malignancies of various types (25%), which may precede or follow the onset of MRH. MRH usually remits spontaneously in 8 to 10 years but often leaves permanent cutaneous and joint damage. Treatment may include methotrexate or cytotoxic therapy if the arthritis is aggressive. Anti–tumor necrosis factor-α (anti-TNF-α) therapy is reportedly beneficial in resistant cases.

SARCOIDOSIS

Joint manifestations including arthritis, periarthritis, and arthralgias occur in 4 to 38% of patients with sarcoidosis (Chapter 95).[10] Rheumatic involvement is divided into acute and chronic types. The first consists of the triad of arthritis, erythema nodosum, and hilar adenopathy on chest radiographs (Löfgren's syndrome), which may be accompanied by fever. Arthritis arises most often in the knees and ankles, and periarticular pain can be severe. Treatment is with NSAIDs, colchicine, or both, and symptoms usually remit spontaneously over several weeks. The less common type of joint involvement in sarcoidosis consists of synovitis that accompanies the slower onset, more chronic, systemic form of sarcoidosis. Polyarthritis, oligoarthritis, or monoarthritis can affect the small or large joints; it is typically nondestructive but in some cases can be aggressive. Dactylitis resulting from sarcoid bone and soft tissue involvement can occur (Fig. 275-3). In contrast to the acute type, chronic sarcoid arthropathy is characterized by mildly inflammatory synovial fluid and histologic granulomas on synovial biopsy. Treatment consists of NSAIDs, low-dose corticosteroids, hydroxychloroquine, and methotrexate or azathioprine. In refractory cases, anti-TNF-α therapy has been successful. Other musculoskeletal manifestations of sarcoidosis include lytic or sclerotic bone lesions (3 to 13%) and symptomatic acute or chronic myopathy (3%). Notably, asymptomatic lesions involving bone and muscle are much more common on magnetic resonance imaging and tissue biopsies.

IMMUNOGLOBULIN G4–RELATED DISEASE

Patients with IgG4-related disease are typically men (70 to 75%) older than 50 years. Patients present with a variety of local and systemic manifestations, some of which can resemble several rheumatic diseases (Chapter 256).[11] For example, tumefactive lesions of the salivary glands can mimic Sjögren's syndrome. Destructive sinus and middle ear lesions or periorbital masses can suggest granulomatous polyangiitis (GPA) (formerly Wegener's granulomatosis). Furthermore, patients with IgG4-related disease frequently (40%) have allergic manifestations including chronic sinusitis and pulmonary symptoms, which may add to the diagnostic confusion with GPA. IgG4-related disease can also cause an inflammatory aortitis with aneurysm formation that can be mistaken for giant cell arteritis. A fibrosclerotic presentation in the abdomen can mimic retroperitoneal fibrosis. Other organs that can be involved include the pancreas, biliary tree, kidneys, lymph nodes, meninges, thyroid, breast, prostate, pericardium, and skin. Although up to 70% of patients will have an elevated serum IgG4 level, histopathologic analysis of biopsy specimens is the "gold standard" for diagnosis. The key pathologic features are a dense lymphoplasmacytic infiltrate organized in a storiform

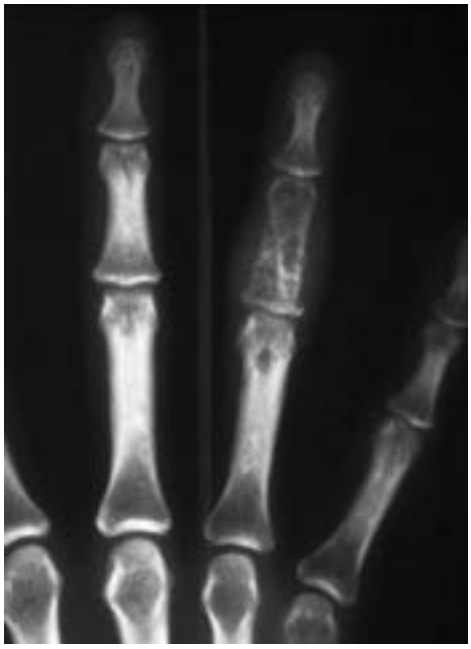

FIGURE 275-3. Sarcoid bone involvement. Punched-out lytic lesions of the middle phalanx, with soft tissue swelling.

pattern, obliterative phlebitis, and a mild or moderate eosinophilic infiltrate. The plasma cell infiltrate will show more than 30 IgG4-positive cells per high-power field and a ratio of IgG4 to IgG positive cells that is higher than 50%. Glucocorticoids are effective in most patients. Several medications (azathioprine, mycophenolate mofetil, methotrexate) have been used as steroid-sparing agents to maintain remission. For refractory disease, B-cell depletion therapy with rituximab is effective.

ALKAPTONURIA (OCHRONOSIS)

Although an inherited disorder, alkaptonuria is usually not diagnosed until the patient presents with progressive premature osteoarthritis as a young adult (before age 30 to 35 years). The spine is initially involved, followed by the knees, shoulders, and hips. Small peripheral joints are spared. Radiographs show multiple vacuum discs, disc space ossification, and osteoarthritic changes in the spine. Nonarticular features include bluish brown discoloration of ear pinna, sclera, and nasal cartilage. Deposition of ochronotic pigment onto collagen fibers causes the articular cartilage to become brittle and fragmented. The noninflammatory synovial fluid may show tiny shards of pigmented cartilage ("ground pepper"). The diagnosis of alkaptonuria is suspected when fresh urine turns dark brown or black on standing or with alkalinization. The diagnosis is confirmed by quantitative measurement of increased homogentisic acid in urine. A specific enzyme assay for homogentisic acid dioxygenase can also be performed. There is no effective therapy for alkaptonuria, although nitisinone is currently under investigation. The arthritis is treated symptomatically with analgesics.

FABRY'S DISEASE

Most patients with hereditary lysosomal storage diseases present and are diagnosed during childhood (Chapter 208). However, female heterozygotes and atypical variants of Fabry's disease may have a milder and later-onset phenotype. Fabry's disease is an X-linked lipid storage disease caused by a deficiency of lysosomal α-galactosidase A. Males with classic Fabry's disease usually develop neuromuscular symptoms in childhood, including painful crises with burning paresthesias of the distal extremities, often accompanied by fever. However, in female heterozygotes and males with low residual α-galactosidase A levels, disease manifestations may occur for the first time in adulthood. Neuromuscular manifestations can range from painful acroparesthesias to fibromyalgia. Progressive or isolated cardiac, cerebrovascular, and renal disease can develop later. Fabry's disease should be suspected in any patient with a paternal family history of early-onset renal failure. Patients should be examined for characteristic ocular stigmata (cornea verticillata) and dermal signs (angiokeratomas). The diagnosis is confirmed in males by determining α-galactosidase A activity in plasma or peripheral leukocytes. In contrast, female carriers must be tested for one of the specific gene mutations. Early diagnosis is important because enzyme replacement therapy can prevent irreversible organ damage.

RELAPSING POLYCHONDRITIS

Relapsing polychondritis is an uncommon multisystem disorder characterized by recurrent episodes of inflammation of cartilaginous tissues.[12] Patients with relapsing polychondritis typically present with the sudden onset of pain and erythema involving the cartilage of the external ear, larynx, trachea, or nose. A nonerosive, seronegative polyarthritis or oligoarthritis affecting small, large, or parasternal joints (23 to 47%); ocular inflammation, including episcleritis or scleritis; and audiovestibular disturbances may also be presenting symptoms. The arthritis is typically acute, migratory, and episodic and resolves spontaneously over days to weeks. Rarely, it can become chronic. Tenosynovitis is also common. Relapsing polychondritis is presumably due to a cell-mediated and humoral immune response against cartilage components; biopsies showing acute and chronic inflammation destroying cartilage support the diagnosis. Late sequelae of relapsing polychondritis include deformity of the pinnae or nose, reduced vision or hearing, tracheal narrowing or collapse, and aortic insufficiency resulting from aortic ring dilation as well as other cardiovascular abnormalities. Patients with relapsing polychondritis frequently have associated coexisting diseases, such as systemic vasculitis, various connective tissue diseases (e.g., RA), myelodysplastic syndromes and other cancers, and thyroid disease. Treatment depends on the severity of the presentation and whether major organs are involved. Mild episodes of inflammation are treated with NSAIDs, colchicine, dapsone, and low-dose corticosteroids. Life-threatening or organ-threatening complications are treated with high-dose corticosteroids and immunosuppressive agents such as methotrexate or cyclophosphamide. Infliximab and tocilizumab have been anecdotally effective in treatment-resistant cases.

CYSTIC FIBROSIS

In up to 10% of patients with cystic fibrosis (Chapter 89), an episodic, nondestructive, inflammatory oligoarthritis develops, most commonly involving the fingers and lower extremity joints. This arthritis is thought to be due to immune complex deposition caused by chronic lung infections. Attacks last for a few days and may be associated with fever and painful nodular skin lesions and purpura. Other musculoskeletal manifestations include osteoporosis (30-75%) and osteomalacia related to malabsorption and, more rarely, hypertrophic osteoarthropathy (5%) and a small vessel vasculitis.

TENOSYNOVIAL GIANT CELL TUMOR: DIFFUSE TYPE

The diffuse type of tenosynovial giant cell tumor (TGCT) (also called diffuse pigmented villonodular synovitis) occurs most commonly in the third and fourth decades of life. It is characterized by the onset of unilateral pain and swelling of a joint, typically the knee (80%). Unusually, a tendon, bursa, or another joint can be involved. The synovial fluid is characteristically brown or hemorrhagic, and radiographs may show soft tissue swelling, osteolysis, subchondral cysts, and bone erosions. TGCT is a nonmalignant condition due to a translocation between chromosomes 1p13 and 2q35 in which the gene coding for colony-stimulating factor-1 (CSF-1) is fused to the collagen VI alpha-3 gene. Up to 15% of cells in the TGCT overexpress CSF-1. The remaining cells in the tumor are inflammatory cells recruited into the tumor because they contain the receptor for CSF-1. TGCT is best diagnosed by synovial biopsy. Microscopic examination reveals a characteristic histology, including marked synovial cell hyperplasia and subsynovial invasion by masses of polygonal cells, multinucleated giant cells, and lipid-filled macrophages. Hemosiderin deposits are between and within cells and have a characteristic appearance on magnetic resonance imaging, with nodular foci of decreased signal on both T1- and T2-weighted images. The treatment for pigmented villonodular synovitis is synovectomy with or without postoperative radiotherapy.[13]

FUTURE DIRECTIONS

With the advances being made in immunology and genetics, there will be an increased understanding of the pathogenesis of many of these diseases. Treatments such as immunomodulating biologic agents or cartilage-preserving therapies will be developed on the basis of new discoveries elucidating the etiology of these unusual disorders. Because of the rarity of many of these diseases, the establishment of registries and international databases detailing clinical characteristics and response to therapies would be a valuable resource.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Teufel A, Weinmann A, Kahaly GJ, et al. Concurrent autoimmune diseases in patients with autoimmune hepatitis. *J Clin Gastrenterol*. 2010;44:208-213.
2. Efe C, Wahlin S, Ozaslan E, et al. Autoimmune hepatitis/primary biliary cirrhosis overlap syndrome and associated extrahepatic autoimmune diseases. *Eur J Gastroenterol Hepatol*. 2012;24:531-534.
3. Puechal X. Whipple's disease. *Ann Rheum Dis*. 2013;72:797-803.
4. Dunn AL. Pathophysiology, diagnosis and prevention of arthropathy in patients with haemophilia. *Haemophilia*. 2011;17:571-578.
5. Chakravarty SD, Markenson JA. Rheumatic manifestations of endocrine disease. *Curr Opin Rheumatol*. 2013;25:37-43.
6. Nguyen S, Hojjati M. Review of current therapies for secondary hypertrophic pulmonary osteoarthropathy. *Clin Rheumatol*. 2011;30:7-13.
7. Rugiene R, Dadoniene J, Aleknavicius E, et al. Prevalence of paraneoplastic rheumatic syndromes and their antibody profile among patients with solid tumors. *Clin Rheumatol*. 2011;30:373-380.
8. Carroll GJ, Breidahl WH, Olynyk JK. Characteristics of the arthropathy described in hereditary hemochromatosis. *Arthritis Care Res*. 2012;64:9-14.
9. Trotta F, Colina M. Multicentric reticulohistiocytosis and fibroblastic rheumatism. *Best Pract Res Clin Rheumatol*. 2012;26:543-557.
10. Drent M, Cremers JP, Jansen TL. Pulmonary meets rheumatology in sarcoidosis: a review on the therapeutic approach. *Curr Opin Rheumatol*. 2014;26:276-284.
11. Stone JH, Zen Y, Deshpande V. IgG4-related disease. *N Engl J Med*. 2012;366:539-551.
12. Arnaud L, Mathian A, Haroche J, et al. Pathogenesis of relapsing polychondritis: a 2013 update. *Autoimmun Rev*. 2014;13:90-95.
13. Colman MW, Ye J, Weiss KR, et al. Does combined open and arthroscopic synovectomy for diffuse PVNS of the knee improve recurrence rates? *Clin Orthop Relat Res*. 2013;47:883-890.

REVIEW QUESTIONS

1. A 30-year-old man complains of a several-months history of joint pain involving his hands, wrists, knees, and ankles. He reports chronic swelling but no acute attacks of synovitis. He has been taking up to eight acetaminophen tablets a day to control the pain. He denies rashes, nodules, pleuritic chest pain, or dyspnea. He reports no family history of arthritis. His father died of liver cancer. Physical examination is remarkable for synovial thickening over his metacarpophalangeal joints (MCPs), wrists, and ankles. He has small effusions in both knees. His skin examination shows darkened skin but no nodules. Radiographs show joint space narrowing, sclerosis, and osteophytes of his MCPs and radiocarpal joints. Which one of the following tests is most likely to support this patient's diagnosis?

A. Iron studies
B. Anti–cyclic citrullinated peptide (CCP) antibodies
C. Liver-associated enzymes
D. Urine homogentisic acid
E. Morning serum cortisol

Answer: A This patient has hemochromatotic arthropathy, which would be supported by abnormal iron studies (elevated iron, % transferrin saturation, and ferritin). Up to 33% of patients with hemochromatosis present with arthritis as their initial manifestation. Osteoarthritic involvement of MCPs, wrists, and ankles, coupled with dark skin demonstrating a grayish hue, are characteristic. The osteoarthritic changes on radiographs excludes rheumatoid arthritis, so an anti-CCP antibody would not be helpful. Although liver-associated enzymes may be elevated, other stigmata of liver disease are absent early in the course of hemochromatosis. However, the early diagnosis and treatment of this patient and affected family members will help prevent cirrhosis and possible liver cancer, which developed in his father. A urinary homogentisic acid level will be normal. Ochronosis can cause dark skin, but the arthritis presents in the lumbar spine with osteoarthritis and calcified vertebral discs. Addison's disease (primary adrenal insufficiency) likewise causes hyperpigmentation, but it is not typically associated with rheumatic manifestations.

2. A 32-year-old woman complains of painful lower extremity nodules and bilateral ankle arthritis over the past 5 days. She denies fever, sore throat, cough, chest or abdominal pain, or dysuria. Past history is significant for two urinary tract infections and one episode of pelvic inflammatory disease. She has a family history of a mother who died with ovarian cancer. Her only medication is birth control pills. Physical examination shows normal vital signs. She has bilateral swelling and tenderness of her ankles. Several 2- to 3-cm erythematous, tender nodules are on the extensor surfaces of both lower legs. The remainder of the examination is unremarkable. Which one of the following tests should be ordered now to support this patient's clinical diagnosis?

A. Nodule biopsy
B. Ankle joint synovial fluid analysis
C. Cervical culture for *Neisseria gonorrhoeae*
D. Chest radiograph

Answer: D This patient has a presentation consistent with acute sarcoidosis manifested by erythema nodosum and bilateral ankle arthritis (Sweiss NJ, Patterson K, Sawaqed R, et al. Rheumatologic manifestations of sarcoidosis. *Semin Respir Crit Care Med*. 2010;31:463-473). A person with bilateral ankle arthritis should always be investigated for possible sarcoidosis. Despite negative pulmonary symptoms, a chest radiograph will show bilateral hilar adenopathy supporting the clinical diagnosis of Lofgren's syndrome. A nodule biopsy would show a septal panniculitis consistent with erythema nodosum, which could be from a cause other than sarcoidosis such as birth control pills. Synovial fluid aspiration is usually unsuccessful and always nondiagnostic in a patient with an acute sarcoidosis presentation. A cervical culture in an asymptomatic patient is unlikely to be useful, and disseminated gonococcemia typically presents with fever, arthritis, and a vesiculopustular rash, but not nodules.

3. A 72-year-old woman complains of a 3-week history of bilateral progressive hand pain unresponsive to naproxen, prednisone, and oxycodone. She denies fever, rash, or Raynaud's phenomenon. She has had a 10-pound weight loss and lower abdominal pain that she attributes to constipation and lack of appetite owing to taking narcotics. Physical examination shows normal vital signs. She has bilateral metacarpophalangeal and proximal interphalangeal joint swelling and tenderness. She has diffuse swelling, erythema, and tenderness that follow the track of the flexor tendons of the digits. Straightening the fingers causes intense pain. Shoulder examination shows decreased internal rotation bilaterally. The remainder of the joint, heart, and lung examinations were normal. Tests obtained in the emergency room last week showed normal hand radiographs except for soft tissue swelling and chondrocalcinosis of the wrists. Rheumatoid factor and antinuclear antibody were negative. Which one of the following tests would be most important to obtain next?

A. Magnetic resonance imaging (MRI) of the hands
B. Anti–CCP antibody
C. Bimanual pelvis examination
D. Synovial fluid aspiration and crystal examination

Answer: C This patient is presenting with the palmar fasciitis and polyarthritis syndrome (PFPAS). This presentation is highly associated with ovarian cancer when it occurs in an elderly woman (Shah A, Jack A, Liu H, Hopkins RS. Neoplastic/paraneoplastic dermatitis, fasciitis, and panniculitis. *Rheum Dis Clin North Am*. 2011;37:573-592). Therefore, a bimanual pelvic examination would be the most important test to obtain. MRI of the hands will show palmar fasciitis, which was diagnosed on the clinical examination. Although rheumatoid arthritis can have an acute onset, the lack of response to nonsteroidal anti-inflammatory drugs (NSAIDs) and prednisone makes this diagnosis unlikely. In addition, the rheumatoid factor was negative, so an anti-CCP antibody is unlikely to be positive. Chondrocalcinosis is a common incidental radiographic finding in elderly patients. The multijoint presentation and lack of response to prednisone make acute pseudogout unlikely, and therefore putting the patient through a difficult aspiration of a finger joint looking for crystals is not indicated.

4. A 52-year-old man is admitted to the hospital for joint pain, diarrhea, and weight loss. He reports a 3-year history of palindromic, inflammatory polyarthritis that tends to affect large joints, including the knees, ankles, and wrists. He also has increased lower back pain that is worse at the end of the day. Previous evaluation showed a negative rheumatoid factor, negative antinuclear antibody, and normal hand and wrist radiographs. He states the arthritis lasts several weeks before resolving. NSAIDs, hydroxychloroquine, and methotrexate have been ineffective in reducing the frequency of attacks that occur every few months. Over the past 6 months, he has noted diarrhea and a 20-pound weight loss. He has had mild abdominal discomfort and a decreased appetite. Evaluation showed a normal colonoscopy 4 months ago. An abdominal computed tomography scan showed hepatomegaly and several enlarged periaortic lymph nodes. A 24-hour stool collection revealed steatorrhea. An antitransglutaminase antibody was negative. Physical examination shows a thin male with normal vital signs except a pulse of 100 beats per minute. He has mild hepatomegaly. He has a few enlarged axillary lymph nodes. His right wrist and left knee are swollen. Synovial fluid aspiration from his knee yesterday showed a cell count of 22,100/mm^3 with 80% neutrophils. Gram stain and crystal examinations were negative. Laboratory evaluation included the following:

Complete blood count: hematocrit 33%, white blood cell count 7800/mm^3, platelets normal
Chemistries: normal
Albumin: 2.7 g/dL
Aspartate transaminase: 51 U/L (normal, 0 to 40)
Urinalysis: normal
Erythrocyte sedimentation rate: 48 mm/hour

Which one of the following tests is most likely to confirm your clinical diagnosis?

A. Hepatitis C serologies
B. Esophagogastroduodenoscopy with small bowel biopsy
C. Anti–*Saccharomyces cerevisiae* antibody
D. HLA-B27 and sacroiliac joint radiograph
E. Blood culture

Answer: B This patient has Whipple's disease, which would be diagnosed by a small bowel biopsy (Puechal X. Whipple's disease. *Ann Rheum Dis.* 2013;72:797-803). Patients with Whipple's disease frequently have an intermittent inflammatory arthritis that precedes other manifestations of Whipple's disease by a few years. The patient's more recent weight loss, diarrhea with steatorrhea, hepatomegaly, and lymphadenopathy are other manifestations of Whipple's disease. The small bowel biopsy showing periodic acid–Schiff–positive macrophages and immunohistochemical staining with antisera specific for *Tropheryma whipplei* will confirm the diagnosis. A lymph node or synovial biopsy would show similar diagnostic findings. The recent normal colonoscopy rules out inflammatory bowel disease, so an anti–*Saccharomyces cerevisiae* antibody will not be helpful. Despite an enlarged liver and mild transaminitis, which are also seen in Whipple's disease, hepatitis C would not explain his steatorrhea. The patient's back pain is mechanical because it is worse at the end of the day. Therefore, an HLA-B27 test and sacroiliac joint radiographs will not confirm his diagnosis, nor will these tests explain his diarrhea. Although the patient has an inflammatory synovial fluid, septic arthritis would not follow this clinical course.

5. You are asked to consult on a 64-year-old man hospitalized for aortitis. He was emergently admitted 3 days ago with chest pain from a 6-cm descending thoracic aortic aneurysm and dissection. Pathologic examination of the surgically removed diseased aorta revealed a nongranulomatous aortitis with fibrosis. His past history includes diet-controlled diabetes mellitus and hypertension controlled with lisinopril. One year ago, he had an orbital pseudotumor successfully treated with 3 months of prednisone. He denies fever, headaches, or visual problems. Physical examination was limited because of his recent surgery. Temporal artery pulses were normal. He had atelectatic crackles in both bases. He has 1^+ pedal edema. No abdominal bruits or masses. Preoperative laboratories include the following:

Complete blood count: hematocrit 35%, white blood cell count 10,400/mm^3, platelets 280,000
Chemistries: normal
Rapid plasma reagin: negative

Aorta histology showed lymphoplasmacytic infiltrate without granulomas or giant cells. Adventitial phlebitis and fibrosis were present. Stains for spirochetes and fungal elements were negative. Which one of the following is most likely to help confirm this patient's diagnosis?

A. Genetic testing for transforming growth factor-β (TGF-β) receptor mutations
B. Antineutrophil cytoplasmic antibody
C. Immunohistochemical staining of tissue for immunoglobulin G4 (IgG4)-positive plasma cells
D. Temporal artery biopsy

Answer: C This patient has IgG4-related disease (Stone JH, Zen Y, Deshpande V. IgG4-related disease. *N Engl J Med.* 2012;366:539-551). The nongranulomatous, lymphoplasmacytic infiltrate in the aorta is consistent with this. The past history of an orbital pseudotumor is also compatible with this condition. Serum IgG4 level is typically elevated. Immunohistochemical studies would show that the aortic infiltrate was primarily IgG4-producing plasma cells. Genetic testing for TGF-β receptor mutations would not be helpful because this is not the histology seen in patients with Loeys-Dietz syndrome. Antineutrophil cytoplasmic antibody–related vasculitis does not cause aortitis. A temporal artery biopsy would offer no further information not already shown by the aortic histology.

276 SURGICAL TREATMENT OF JOINT DISEASES

C. RONALD MACKENZIE AND EDWIN P. SU

Estimates of the prevalence of arthritis and other rheumatic diseases demonstrate the enormous impact that these conditions have on the U.S. populace and the health care system in general. More than 21% of U.S. adults (46 million people) currently report physician-diagnosed arthritis. Although the majority of this health burden arises as a consequence of osteoarthritis, the full span of the rheumatic diseases contributes to the impact of this class of conditions. Already the leading cause of disability in the nation, the number of people with arthritis and arthritis-attributable limitation in activity is anticipated to approach 67 million affected adults by the year 2030. Ultimately, surgical intervention is required in many of these individuals. Factors such as an increased patient awareness of the benefits of surgery, the desire for higher activity levels, and improvements in surgical techniques have, in concert with the increasing prevalence of chronic arthritis, fueled the growth in utilization of orthopedic surgery. In 2010, it was estimated that about 300,000 primary total hip replacements and about 650,000 primary total knee replacements were performed in the United States. By the year 2020, it is predicted that about 500,000 hip replacements and 1.5 million knee replacements will be performed each year.[1]

PATHOBIOLOGY

The pathobiology of joint arthritis leading to surgical intervention is essentially that of articular cartilage damage resulting in the loss of mechanical properties, accompanied by inflammation of the joint lining. With continued cartilage deterioration, stiffness and pain ensue. Without the protective layer of articular cartilage, the nociceptive and proprioceptive receptors in the periosteum are activated, leading to unremitting pain.

Osteoarthritis is the most common cause of end-stage arthritis (Chapter 262). Osteoarthritis may be primary, due to biochemical changes in the cartilage, or secondary to systemic disease affecting the cartilage, joint damage from preexisting inflammatory joint disease, or trauma. Mechanical overload and imbalances lead to further cartilage degradation. Important adaptive processes such as subchondral sclerosis and osteophyte formation occur in response to joint overload, and, if chronically present, cyst formation in the subarticular bone may also result. Over time, the osteophytes or bone spurs will lead to restricted range of motion (Fig. 276-1).

Inflammatory arthritis, by contrast, is a constellation of diseases involving the synovium. Included in this class of disorders are such important conditions as rheumatoid arthritis (Chapter 264), psoriatic arthritis, and the seronegative spondyloarthropathies (Chapter 265). On a pathologic level, all involve the release of inflammatory mediators in the adjacent synovium, leading secondarily to cartilage destruction. In contrast to osteoarthritis, there is no mechanical overload as a primary mechanism, and no bone sclerosis or osteophyte formation are seen. Rather, the inflammatory synovitis leads to characteristic loss of cartilage matrix, marginal bony erosions, and osteopenia (Fig. 276-2)

Trauma is another important cause of joint destruction. Post-traumatic arthritis is initiated by cartilage damage at the time of injury or by secondary mechanical imbalances that result from fractures of juxta-articular bone. Abnormal loading conditions will subsequently lead to a wear-and-tear form of cartilage damage.

Osteonecrosis is another entity that may lead to joint arthritis. In this process, the blood supply to the bone is compromised, leading to necrosis of the bone supporting the articular surface. The most commonly affected joints are the hip, shoulder, and knee. As the disease progresses, the necrotic bone may collapse, leading to the loss of articular integrity and progressive cartilage deterioration.

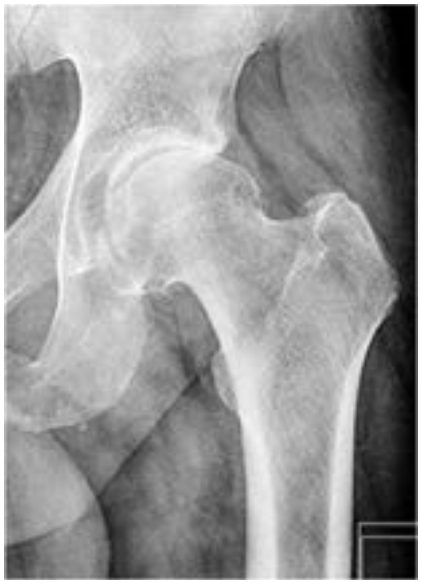

FIGURE 276-1. Radiograph of an osteoarthritic left hip. Note the asymmetrical joint space narrowing and subchondral sclerosis that are characteristic of a wear-and-tear pattern of joint deterioration.

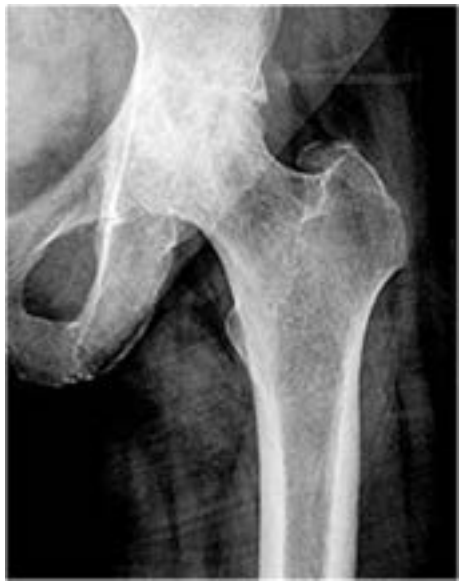

FIGURE 276-2. Radiograph of a left hip with end-stage inflammatory arthritis. Note the symmetrical pattern of cartilage loss and presence of osteopenia.

Other causes of arthritis that may lead to joint damage include metabolic disorders (chondrocalcinosis, gout), tumor (synovial chondromatosis), infections (post-septic), and bleeding disorders (hemophilia).

PREOPERATIVE CONSIDERATIONS

The indications for orthopedic surgery are refractory joint pain and disability. Ultimately, the patient and physicians need to agree that the possible benefits of surgery outweigh the risks (Chapter 431). The decision to proceed with surgery therefore reflects the outcome of a partnership among the patient, the orthopedic surgeon, and the patient's primary physician or rheumatologist (Chapter 430). Achieving the necessary decision-making balance may be complicated, especially given the increasing burden of comorbidity that accompanies the aging patient.

In the elective setting, joint replacement and spine surgery are the most common procedures under consideration. With the former, severe pain and functional limitation unrelieved by conservative treatment are the most common indications for surgical intervention. In the case of spinal surgery, severe radiculopathy, nerve dysfunction (e.g., acute foot drop), and myelopathy are additional considerations. In contrast to elective surgery, there are circumstances when a deliberative approach is not possible because of the

development of more urgent, occasionally life-threatening clinical problems. Examples include hip fracture, acute myelopathy, or the patient with an infected native or prosthetic joint. Because the patient's general health is at risk in these settings, the medical-surgical team must stabilize the patient quickly in order to optimize the outcome. Owing to the coupling of medical advances with increasing financial and resource constraints, a dominant trend toward the performance of surgery in the ambulatory setting has emerged. Indeed the percentage of all surgical procedures performed on an outpatient basis in the United States rose from 20% in 1982 to 60% in 1995, a phenomenon particularly relevant to the arthroscopic techniques of orthopedic surgery. Among the benefits of these developments has been the opportunity to move the preoperative medical evaluation to the outpatient arena. This change in practice allows time for discourse with the other physicians involved in the patient's care, for supplementary consultation and investigation, and, when necessary, for the institution of therapy directed at optimizing the patient's medical status before the contemplated surgery. Approached in this manner, the preoperative evaluation becomes a focal point of communication among all members of the medical team, enhancing the collaborative nature of the consultative process and, ultimately, the patient's care.

Although the efficacy of preoperative assessment has not been definitively established, the aging and increasing complexity of modern-day surgical patients justifies this clinical practice. Although no consensus exists regarding what constitutes the optimal preoperative medical evaluation, a growing literature pertaining to perioperative medicine supports various core principles that underlie effective medical consultation in this clinical setting (Chapter 431).

ANESTHESIA IN THE ORTHOPEDIC PATIENT

Given the protean clinical features that accompany chronic arthritis and the connective tissue diseases, a variety of issues, including airway considerations, the surgical site (joint region), the anticipated duration of surgery, and comorbidities are important determinants of the type of anesthesia to be employed, whether invasive monitoring will be necessary, and the length of time the patient will require intensive monitoring after surgery.

General anesthesia and regional anesthesia are commonly used in the orthopedic patient (Chapter 432). General anesthesia with endotracheal intubation may present a particular danger in patients with rheumatoid arthritis or ankylosing spondylitis. Patients with cervical spine instability or those with a rigid spine may require fiberoptic intubation. Regional anesthesia may involve local anesthesia or peripheral nerve block for minor procedures or epidural-spinal anesthesia for total joint arthroplasty.

Although the debate concerning the relative merits of regional versus general anesthesia remains unresolved, many procedures, particularly orthopedic surgery, are well suited for regional anesthetic techniques. Advantages of regional approaches include a reduction in blood loss, deep vein thrombosis and pulmonary embolism, adverse postoperative respiratory events, and death. Further postoperative pain, a significant problem for patients with a painful rheumatic disease, may be best managed with regional anesthesia. For example, peripheral nerve blocks using longer acting anesthetics and infusion methodologies are often employed because they provide excellent intraoperative anesthesia and postoperative pain relief.

A number of options exist for the control of postoperative pain, including the traditional intravenous and intramuscular routes of narcotic medications (systemic), the use of epidural analgesia, and the local infiltration of anesthetics into the surgical site. The direct administration of local mixtures of medications, including long-acting anesthetics and anti-inflammatory drugs, has become more popular because of the ease of use and excellent efficacy, particularly around the hip and knee joints. Patient-controlled analgesia (PCA) using an epidural route of administration is also an effective method of pain control after lower extremity surgery. Further, epidural PCA and local soft tissue injections facilitate postoperative physical therapy, which is important to the restoration of range of motion in patients undergoing orthopedic procedures. Both methods also reduce the systemic absorption of analgesics, thereby minimizing the problem of narcotic-induced respiratory depression. Parenterally administered nonsteroidal anti-inflammatory drugs (NSAIDs) are also useful and can be used to reduce narcotic requirements after major surgery. However, the common contraindications to NSAID therapy, such as peptic ulcer, renal, and ischemic heart disease, should be observed in the postoperative setting.

SURGICAL MANAGEMENT

Surgical treatment of joint disease is focused primarily on the relief of pain; secondary objectives are improvement in joint motion, reduction in swelling, return to function, and prevention of continued cartilage destruction. Realizing that surgical treatment has limitations and complications, the decision to move forward is one that must be individualized for each patient. Factors such as disease severity, the patient's desired activity level, and the anticipated longevity of the patient are all relevant to decision making. Typically, patients who are candidates for the surgical treatment of joint diseases have failed conservative measures (NSAIDs, physical therapy, intra-articular injections) and have daily pain that hinders their function and diminishes their quality of life.

ORTHOPEDIC PROCEDURES

Osteotomy

In circumstances in which a structural abnormality around a joint has led to mechanical overload, an osteotomy (cutting bone) may be an option to correct alignment problems. The most common sites for osteotomy are the hip, to treat acetabular dysplasia, and the tibia, to realign the knee. In acetabular dysplasia, the hip socket is excessively shallow, leading to abnormal stresses on the articular cartilage and premature osteoarthritis. An acetabular osteotomy can be performed in patients in whom significant cartilage still remains. By rotating the pelvic bones, a deeper socket can be formed, reducing stresses on the cartilage and thereby slowing down the arthritic process. With tibial osteotomy, the knee joint can be realigned to direct forces away from the region of cartilage damage. Usually a varus (bow-legged) deformity indicates that the medial compartment of the knee is excessively worn and, as such, a tibial osteotomy realigns the joint in such a way to direct forces to the uninvolved, lateral compartment. Typically, osteotomy is considered an option for younger patients (<40 years); beyond this age, the loss of cartilage is generally such that more reproducible results would be attained with total joint arthroplasty.

Arthroscopy

Arthroscopic surgery is performed by inserting a camera and specialized instruments into a joint through small, puncture-type incisions. Arthroscopic surgery is effective in the treatment of intra-articular pathology such as meniscal tears of the knee, labral tears of the hip, cartilage flaps, small chondral defects, and loose bodies. However, after the articular cartilage is significantly damaged, arthroscopic débridement is usually ineffective in the absence of mechanical symptoms such as locking and clicking. In some instances, underlying joint arthritis may lead to tears in the meniscus or labrum; if such a tear results in new mechanical symptoms, then arthroscopic surgery may be helpful in selected cases.

In the assessment of the hip, there has been an increased focus on the femoral head and neck architecture as a cause of osteoarthritis. In certain patients, the anatomy of the femoral head and neck may lead to impingement of the femoral neck upon the acetabular rim, typically in flexion and internal rotation. This condition, known as *femoroacetabular impingement*, results in the repetitive contact between the femoral neck and acetabular rim and is believed to result in labral tears, cartilage damage, and eventually arthritis.[2] Thus, there is currently much interest in reshaping the bones of the femur and acetabulum by so-called osteochondroplasty. This procedure is being performed as an open or arthroscopic procedure and provides good symptomatic relief in the short term. The long-term effects, specifically the impact on the future development of arthritis, have yet to be demonstrated for this procedure.

Synovectomy

Synovectomy refers to removal of the synovial lining of the joint, either through an open or an arthroscopic approach. In conditions such as rheumatoid arthritis, in which the disease process involves an actively inflamed synovium, it follows that debulking the pathologic tissue may reduce symptoms and slow the destruction of cartilage. In practice, synovectomy can be effective at relieving pain as long as there is remaining cartilage. However, the procedure is not predictable in terms of regaining joint motion. Further, after the cartilage is completely worn through, the joint deterioration is too advanced for synovectomy to be helpful. Therefore, synovectomy is generally performed in patients with rheumatoid arthritis (or other forms of inflammatory arthritis) who have active synovitis in the presence of relatively preserved articular cartilage. The most common joints that benefit from synovectomy are the knee and elbow. However, synovectomy should be considered as "buying time" because the synovium will reappear.

Arthrodesis

Arthrodesis, or fusion of a joint, achieves the goal of pain relief by creating a nonmobile joint. Rather than have the arthritic joint surfaces elicit pain with

movement, a surgical fusion (arthrodesis) of the articulating bones creates a construct that can bear weight and is stable. This is achieved by removing the articular surfaces from the joint and immobilizing the bones such that they heal in a solid union. This procedure was formerly the treatment of choice for hip and knee arthritis in young, active laborers because of its durability and avoidance of implants with their propensity to wear. However, creating stiffness at one joint will increase stresses on the joints above and below the fused joint.

Hip fusion may be performed in young patients to treat the sequelae of slipped capital femoral epiphysis, Legg-Calvé-Perthes disease, post-septic arthritis, or osteonecrosis of the femoral head. Fusion surgery can achieve a painless, supportive joint that is capable of bearing heavy loads while avoiding artificial implants. However, the gait mechanics are altered, requiring more energy for ambulation. Further, the lack of motion at the hip increases stresses on the joints above and below the hip. Thus, the natural history of a hip fusion is the development of ipsilateral knee arthritis and low back pain after 20 to 25 years, necessitating much later the conversion of a fused hip to a hip replacement (fusion takedown). Although hip arthrodesis is still a viable option in the young arthritic patient, patients' desire for maintaining hip mobility in order to sit and drive has made this largely a treatment of the past.

Fusion of the knee joint is performed less commonly than hip fusion. In addition to the lack of motion that may make it difficult to sit or climb stairs, knee fusion cannot be successfully converted to total knee replacement. Thus, knee fusion is generally considered a salvage procedure, mainly employed in situations in which replacement is not possible (e.g., lack of muscle function or persistent infection).

Ankle fusion is still commonly performed as the treatment of choice of tibiotalar arthritis. Because the historical results of ankle replacement have not been durable, fusing the ankle is the best method of creating a pain-free joint. Furthermore, the ability of the knee and subtalar joints to compensate for a stiff ankle has made this procedure more tolerable.

Total Joint Arthroplasty

Joint arthroplasty refers to the re-creation of congruent joint surfaces, typically with artificial parts. In certain patients and in non-load-bearing joints such as the elbow, interpositional arthroplasty can be performed by placing a tissue graft between the arthritic surfaces. In the case of weight-bearing joints such as the hip and knee, however, metal and plastic materials produce the most durable results. In such circumstances, the articular surfaces are replaced by shaped materials designed to recreate the joint kinematics; thus, the procedures are commonly called total hip, knee, and shoulder replacements.

In general, after the articular cartilage is completely worn or destroyed on both sides of the joint, arthroplasty is the most predictable option to relieve pain. After total joint arthroplasty, it is advisable to reduce stresses on the joint to promote implant longevity. This includes weight loss and avoidance of impact activities; walking, cycling, and gliding type activities are permitted, but in general, running and jumping should not be performed. Because a total joint arthroplasty involves artificial, moving components, the replaced joints are subject to the same wear and tear as native joints. Thus, they have a finite lifespan that is dependent on a patient's weight and activity level and the implant materials. Subsequent revisions of joint replacements can be difficult and less durable; thus, it is wise to defer joint arthroplasty until there are no other options.

Total hip replacement was first developed in the 1950s in the United Kingdom using metal and plastic components attached to the bone with cement. The early results were so predictable and reproducible that the technique rapidly spread worldwide. The National Institutes of Health, in 1994, published a consensus statement that total hip replacement "is one of the most successful surgical procedures and provides immediate and substantial improvement in a patient's pain, mobility, and quality of life. Compared with treatments for other chronic debilitating diseases, total hip replacement is highly cost effective."

Total hip replacement is the treatment of choice for end-stage arthritis caused by any of the aforementioned pathobiologic processes. It involves the exposure of the joint, removal of the arthritic femoral head at the level of the femoral neck, and removal of enough acetabular bone to place a prosthetic socket. The femoral implant is inserted into the intramedullary canal and anchored with bone-ingrowth techniques or bone cement. The typical bearing materials for the hip joint are metal-on-polyethylene, ceramic-on-polyethylene, and ceramic-on-ceramic implants. These combinations of materials may be chosen based on the patient's age, activity level, and surgeon preference (Fig. 276-3). Metal-on-metal total hip replacement, popular for a time period in early to mid-2000s, has now fallen out of favor because of findings of adverse local tissue reactions to the metal debris.

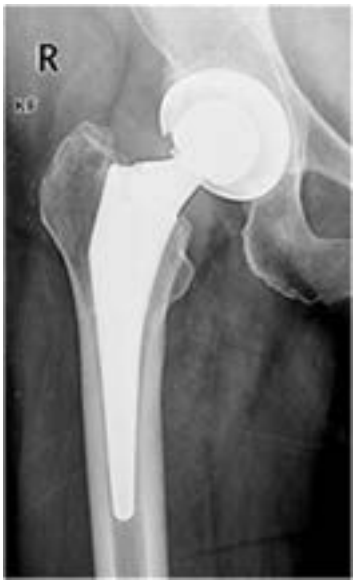

FIGURE 276-3. Radiograph of a total hip replacement consisting of uncemented acetabular and femoral components. The articulating materials are a metal ball and polyethylene liner.

Using modern implant materials and surgical technique, the implant survival rates are 90 to 95% successful at 15 years; however, longevity will vary depending on patient factors such as weight and activity. There have been cases in which total hip replacement implants have lasted more than 30 years.

A major cause of failure of total hip replacement has been wearing of the implant materials coupled with the body's reaction to the particulate debris shed into the joint space over time. In the process of immunologic uptake of the debris, inflammatory cytokines are released, causing osteoclasts to resorb periprosthetic bone. The end result of this osteolytic process is that the implant attachments to bone may be compromised, causing loosening of the prosthesis and pain. The hope is that the current generation of implant materials will reduce the amount of particulate wear debris, thereby extending longevity further.

Total knee replacement was developed in the United States shortly after total hip replacement. The surgery involves the removal of the arthritic surfaces of the tibia and femur followed by their replacement with a metal femoral implant and a metal and/or polyethylene tibial component (Fig. 276-4). It is termed "total" knee replacement to distinguish it from a unicompartmental or "partial" knee replacement.

The recovery from total knee replacement is more difficult than after total hip replacement because of greater postoperative pain and the emphasis on regaining motion. Nonetheless, pain relief and function are usually excellent when recovery is complete. Current studies demonstrate modern implant survival to be 90 to 95% at 15 years, again depending on patient factors.

The long-term outcome after total hip or knee arthroplasty has improved steadily over recent decades. Despite more pre-surgical morbidity, rates are declining especially for cardiac disease, stroke, and pulmonary embolism.[3]

Although less commonly performed than hip and knee replacement, total shoulder replacement is an excellent pain-relieving procedure for glenohumeral arthritis. Developed from experience gained for total hip replacement and total knee replacement, total shoulder replacement uses a stemmed humeral implant with a metal ball and metal and/or polyethylene glenoid socket. Total shoulder replacement requires an extensive rehabilitation protocol to regain range of motion and strength; however, 1 year after surgery, 95% of patients have pain-free use of the shoulder.

Surgical Innovations in Joint Arthroplasty

Minimally Invasive Surgery

As in all surgical subspecialties, there has been a movement toward minimally invasive surgery. This may be a misnomer because the actual work done inside the joint has not changed. Some surgeons have suggested the terminology be changed to "smaller incision surgery" or "less invasive surgery." In any case,

the idea is that the smallest incision possible be used to perform the operation, resulting in less tissue trauma. The interest in this type of surgery has resulted in improvements in instrument design and surgical training. Typically, a hip replacement, performed until recently using a 10-inch incision, can now be done through a 4- to 5-inch incision. Similarly, knee replacement incisions are about one half of their former length. Despite these surgical advances, even larger benefits have resulted from the increased attention to various nonsurgical modalities, all directed at more rapid surgical recovery. Examples include the increased use of peripheral nerve blocks and local tissue anesthetic infiltration, preemptive analgesia, and a more expeditious approach to rehabilitation. Such approaches have reduced the average hospital length of stay to 2 to 3 days after a total hip replacement and 3 to 4 days after a total knee replacement.

Improvements in Implant Technology

As the average age of patients undergoing hip and knee replacement decreases, while their activity levels increase, the number of revision surgeries is projected to grow. Therefore, much research is being performed to improve the longevity of implant materials.[4] The gold standard of arthroplasty is to use a cobalt-chrome (metal) implant against a polyethylene (plastic) surface. Unfortunately, the harder metal surface will eventually wear away the softer plastic surface. Therefore, biomechanical engineers have developed a more resistant, "highly cross-linked" polyethylene that demonstrates greater wear resistance in laboratory simulators. Such highly cross-linked polyethylene has been in clinical use for about 10 years, and the early experience suggests significantly less wear compared with standard polyethylene.[5] Other materials such as ceramics and metals are also being used in an attempt to improve longevity. Nonetheless, to date, there is no consensus concerning the optimal weight-bearing surfaces.

Bone-preserving implants have also been developed in order to maintain more options when revision surgery becomes a necessity. Such procedures may require the removal of the implant and placement of a new prosthetic device; thus, with more bone available, surgical options are enhanced. One such bone-preserving implant is the hip-resurfacing device introduced in the United States in 2006. This is discussed more fully later.

Unicompartmental knee replacement is a bone-preserving implant for the knee. As suggested by its name, it involves replacing only one of the three knee compartments with a prosthetic device (Fig. 276-5). Therefore, candidates for unicompartmental knee replacement must have arthritis limited to a single compartment. Because the surgical trauma and dissection are significantly reduced compared with total knee replacement, patient recovery tends to be less painful and quicker. However, there remains a higher failure rate for unicompartmental knee replacement compared with total knee replacement because of the possibility for developing arthritic changes elsewhere in the joint.

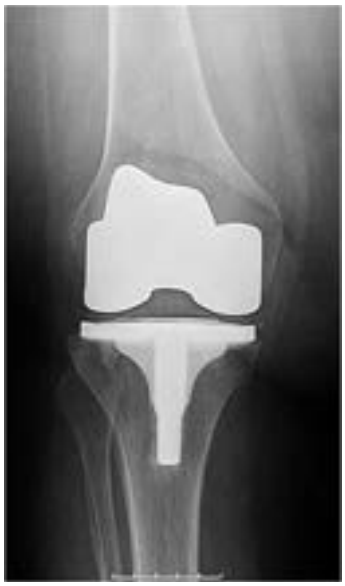

FIGURE 276-4. Radiograph of a total knee replacement. The components are metal, with a polyethylene insert between the tibial and femoral components.

Resurfacing Arthroplasty

Hip resurfacing is an alternative treatment to total hip replacement in the younger, active patient. The primary benefit is the preservation of proximal femoral bone in the event that future (revision) surgery is necessary. Rather than removing the femoral head and portion of the femoral neck as in total hip replacement, the bone is sculpted to accept a metal resurfacing cap (like a tooth), preserving an additional 4 to 5 centimeters of bone (Fig. 276-6). The acetabulum is prepared to accept a metal socket, creating a metal-on-metal joint. There are currently no alternatives to the metal-on-metal articulation of a hip-resurfacing implant; however, in contradistinction to metal-on-metal total hip replacements, there appear to be fewer issues arising from the materials in hip resurfacing. The likely explanation for this finding is that there are fewer pieces and junctions in a hip resurfacing compared with total hip replacement. The preservation of the proximal femoral bone raises an additional risk for failure (i.e., femoral neck fracture below the resurfacing implant), which is estimated to occur in 1% of cases. Although there are no defined age limits for hip resurfacing, the best candidates have been males younger than 55 years, likely because the bone quality in this demographic group is the strongest and most robust. In this age group, the 10-year results of hip resurfacing in Australia have demonstrated a 94% rate of survival (free of revision), which is superior to total hip replacement. However, concerns regarding longevity, a greater short-term failure rate than total hip

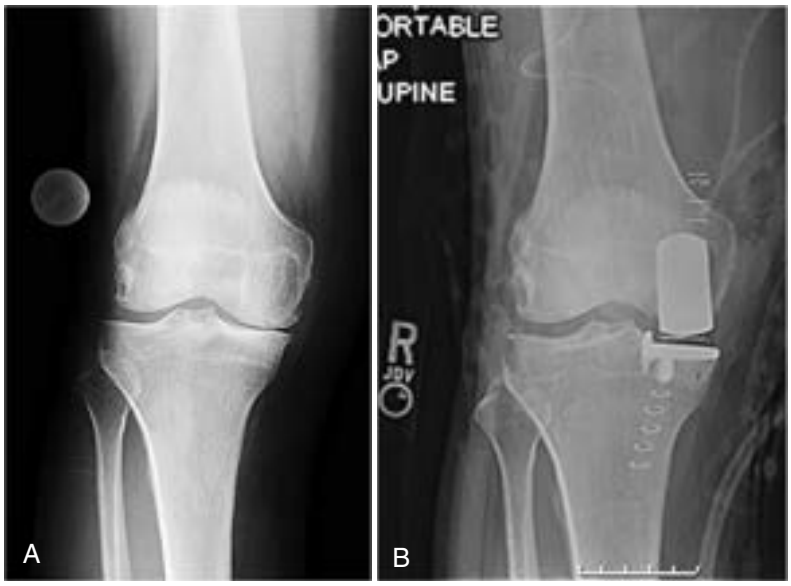

FIGURE 276-5. A, Radiograph of a knee with arthritis limited to the medial compartment. **B,** A medial unicompartmental knee replacement.

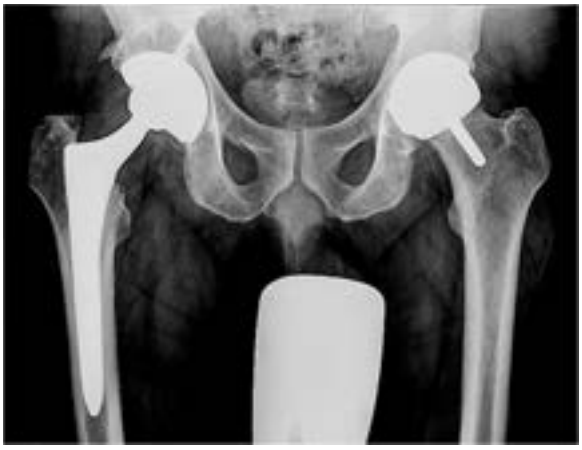

FIGURE 276-6. Radiograph of a patient with a right total hip replacement and a left hip resurfacing. The total hip replacement consists of a longer stem placed into the medullary canal of the femur. The hip resurfacing implant preserves the proximal femoral bone.

replacement, poorer results in women, and metal ion release have led to questions concerning the superiority of the procedure.

Computer Navigation

An orthopedic surgeon relies on visualization, instrument jigs, and experience in order to recreate the proper joint mechanics. Although the surgeon may know exactly how the artificial components are to be placed, it may be difficult to achieve perfect alignment in every operation. Computer navigation is a tool that can be used to aid in the reproducible positioning of implants. Although some errors are inherent in the precision of computer navigation, such techniques have reliably diminished outlier results. It has yet to be determined whether the longevity of hip and knee replacements inserted with the aid of computer navigation differ from those inserted by conventional approaches. For this reason, as well as the expense and time associated with its use, computer navigation is not universally practiced.

ORTHOPEDIC PROCEDURES ON OTHER JOINTS

Elbow, Ankle, and Wrist

The elbow, ankle, and wrist are less frequently replaced joints. With the exception of the rheumatoid wrist, these joints are less commonly afflicted by chronic arthritis. Further, the smaller bones making up these joints also translate to a diminished surface area for implant fixation, thus lowering the durability of the surgical procedures. Although total joint arthroplasty can be successful at relieving pain in the short term, 10-year results do not approach that of total hip or knee replacement. Synovectomy remains an effective surgical option in selected patients with inflammatory arthritis of the elbow, ankle, or wrist.

Spine

The spine consists of multiple levels of articulating bone, discs, and facets. Although not a "joint" in the typical sense of the word, various spinal segments may be differentially involved in chronic arthritis. Generally, the lumbar spine is the most affected, although a similar process can occur in the cervical spine; more than 95% of patients older than 50 years will demonstrate degenerative changes in the lumbar spine. With aging, the nucleus pulposus of the intervertebral disc loses water content and elasticity, resulting in disc space collapse and increased forces across the vertebral facets. As a consequence of the resultant increased pressure, bone spurs develop, leading to stenosis (narrowing) of the neural foramen. As a result of the degenerative process or *spondylosis*, patients may experience mechanical back pain with bending, extending, and twisting. Pain, numbness, or weakness radiating down the extremities in a radicular distribution can result from the neural foraminal stenosis.

Mechanical back pain is the most common affliction of the spine. Unfortunately, this constellation of back symptoms is not reproducibly relieved by surgical intervention. Physical therapy and use of proper back mechanics are the mainstay of treating mechanical back pain.

The advanced sequela of spinal arthritis is spinal stenosis, which can result in lower extremity pain and weakness. This disorder can reliably benefit from decompressive surgery with or without instrumented fusion. The spinal unit must be evaluated for stability in order to determine whether or not spinal fusion is necessary.

Newer surgical treatments such as lumbar disc replacement may be used in certain cases to treat spinal disorders. Whereas the gold standard, vertebral fusion, eliminates painful motion at degenerative disc levels, disc replacement attempts to preserve motion of the spinal unit. As in hip and knee arthrodesis, fusion of the spinal unit will increase forces at the levels above and below the fused segment; thus, further degeneration will often occur. Disc replacement surgery aims to eliminate the progression of arthritis at adjacent levels by retaining motion. Long-term studies comparing disc replacement and spinal fusion are not yet available.

MANAGEMENT ISSUES IN PATIENTS WITH ARTHRITIS UNDERGOING SURGERY

Prevention of Postoperative Infection

Efforts to prevent and detect any infectious processes before and after surgery are of utmost importance. The skin and urinary tract are sites of specific concern, and infection can be ruled out by a careful physical examination and routine preoperative urine culture. In addition, dental consultation may be appropriate in patients with poor oral hygiene and dentition.

Prophylactic antibiotic therapy for total joint arthroplasty patients should begin less than 2 hours before surgery and continue for 24 hours. A common protocol involves cefazolin (Ancef) 1 g every 8 hours (total of three doses) or, in penicillin allergic patients, vancomycin 1 g every 12 hours (total of two doses).

Peripheral Nerve Injuries

Peripheral nerve injuries arise more often after upper and lower extremity surgery because they generally result from excessive traction on the nerve or, alternatively, as a consequence of nerve compression resulting from prolonged positioning of the extremity during surgery or while in a cast. Early detection and intervention are critical to the outcome in these circumstances. Patients with chronic neurologic disorders, such as neuropathies in the setting of diabetes or spinal stenosis, are at increased risk for nerve injury.

Venous Thromboembolism

Prevention of venous thromboembolic phenomena after orthopedic surgery is the most thoroughly studied of potential postoperative complications, and pulmonary embolism remains an important cause of mortality. The orthopedic literature has concentrated on lower extremity arthroplasty, although a recent study suggests that similar approaches should also be considered after total shoulder arthroplasty, in which the risk for thromboembolism may be higher than generally appreciated.

After orthopedic surgery, a complicated balance exists between a possible life-threatening pulmonary embolus and the potential for postoperative bleeding. Numerous protocols have documented the effectiveness of prophylaxis, which should begin at the time of the procedure. Short intraoperative time reduces the risk for deep vein thrombosis, as does the type of anesthesia. Epidural anesthesia reduces the risk for proximal deep vein thrombosis following total hip replacement by two- to three-fold and also reduces the overall risk for deep vein thrombosis by at least 20%. Other intraoperative interventions, such as hypotensive anesthesia and intraoperative heparin administration, further reduce thrombogenesis. Mechanical methods also have proven efficacy at reducing risk for thromboembolism. These include compression methodologies such as stockings and various pneumatic devices, foot flexion-extension exercises, and early ambulation. These are safe, effective approaches that do not increase the risk for bleeding.

The mainstay of prevention is prophylactic anticoagulation, which should begin immediately after surgery. Regimes include aspirin, warfarin, low-molecular-weight heparin, and several new oral anticoagulants, often used in combination with various mechanical compression devices (Chapter 38). Continuing prophylaxis for 21 days rather than 7 days reduces risks of thromboembolism while increasing risks of minor bleeding.[6]

Fat Embolism Syndrome

Fat embolization, a well-described complication of skeletal trauma, may also occur after procedures involving instrumentation of the femoral medullary canal. Although the embolization of fat is believed to occur almost entirely in the setting of hip or femoral fractures, 1 to 3% of patients undergoing joint replacement surgery (particularly simultaneous bilateral procedures) develop fat embolism syndrome (FES).

The signs and symptoms of FES involve the respiratory, neurologic, and hematologic systems, as well as the skin.[7] Time of onset is variable, with hemodynamic instability developing almost immediately in some or insidiously over the first 2 to 3 postoperative days in others. In the latter, patients gradually become hypoxemic, may be hypotensive, and are often confused. Respiratory signs are the most common manifestation. Most patients develop mild to moderate hypoxemia or radiographic changes (mainly bilateral alveolar infiltrates), but only a minority will develop life-threatening adult respiratory distress syndrome. Neurologic manifestations range from mild drowsiness to acute confusional states or to severe obtundation and coma, all consequences of the hypoxemia and the direct effect of the embolization of fat on the brain. The skin eruption, which is rare in total joint arthroplasty patients, takes the form of a petechial rash involving the folds of the neck and axillae, as well as petechiae in the subconjunctiva and oral mucosa. Retinal edema and hemorrhage are also seen. Transient thrombocytopenia is common.

Although patients suspected to have developed FES need to be closely monitored, in most instances after total joint arthroplasty, the condition is relatively benign. Treatment is supportive and includes the administration of oxygen and the prevention of pulmonary hypertension (by fluid restriction and the use of diuretics and venodilators). Corticosteroids are not effective. In most patients, the condition resolves within 3 to 7 days, although in severe cases, the mortality rate has remained in the 5 to 15% range even with modern aggressive therapy.

Cervical Spine

In those rheumatoid arthritis patients who exhibit advanced destructive disease, cervical spine instability should be ruled out before surgery with flexion-extension films in patients with neck pain or crepitus on range-of-motion testing, radicular symptoms, or arm and/or leg weakness. Affected patients should wear a soft cervical collar to the operating room. When possible, epidural or spinal anesthesia should be employed.

Conversely, in patients with ankylosing spondylitis, the patient's rigid cervical spine may also present technical challenges for the anesthesiologist during intubation. Fiberoptic methods are often employed in this clinical setting.

Immunosuppressive and Anti-inflammatory Therapy

The potential contribution of corticosteroids, the disease-modifying antirheumatic drugs (DMARDs), and the newer biologic agents to the risk for postoperative infection and wound dehiscence are well recognized by clinicians, although the debate concerning how to manage these medications in the perioperative setting is not settled. The primary challenge in this context is to achieve an optimal balance between the maintenance of control of the underlying disease while minimizing the risk for postoperative wound infection and/or wound breakdown.

Methotrexate (MTX) is the most commonly used DMARD for the treatment of rheumatoid and other forms of inflammatory arthritis, so it is frequently seen in the perioperative setting. Given its importance in the maintenance of disease control, coupled with observations concerning the low rates of postoperative wound infection and dehiscence associated with its use, it appears safe, indeed desirable, to continue MTX throughout the perioperative period.

With respect to the biologic (anti–tumor necrosis factor) agents, recommendations have been formulated by international groups in which such agents are discontinued for short periods of time before surgery, generally 2 to 4 weeks. Some favor the longer 4-week interval for agents with longer half-lives. There is virtually no information concerning postoperative infection and wound healing with the newer agents such as anakinra, rituximab, and abatacept. Recommendations similar to the other biologics must suffice until more data can be gathered.

The other important medication commonly encountered in the perioperative setting is corticosteroids. In addition to problems related to wound healing, there is the additional concern of postoperative adrenal insufficiency. Traditionally, steroids have been shown to increase the rate of wound infection and dehiscence in surgical patients, although data derived from the orthopedic setting per se remain scant. Similarly insecure is the published experience concerning postoperative adrenal insufficiency. In contrast to MTX and the biologic therapies, however, the question is not whether to stop corticosteroids before surgery because this is generally not feasible. Rather, management considerations pertain to how much and for how long steroid augmentation (stress doses) will be required. In this regard, a nuanced approach, premised on patient-specific considerations, is required. For the patient chronically managed with low doses of corticosteroid (i.e., prednisone ≤7.5 mg/day) or on any dosage for less than 3 weeks, stress dose steroid therapy is unnecessary, and maintenance of the patients' usual daily dosage should be sufficient. In contrast, for patients taking larger dosages (≥20 mg/day) for longer periods (>3 weeks) of time, it is prudent to assume secondary adrenal suppression exists, justifying the use of stress dose steroid therapy. For those taking intermediate dosages (i.e., 7.5 to 20 mg/day), decision making needs to be individualized. In addition to the dosage and duration of steroid therapy, other considerations, such as the presence of diabetes, the use of other immunosuppressive therapy, hypoalbuminemia, and poor nutritional status, as well as the magnitude of the surgical procedure, need to be taken into account.

Grade A References

A1. Yadeau JT, Goytizolo EA, Padgett DE, et al. Analgesia after total knee replacement: local infiltration versus epidural combined with a femoral nerve blockade: a prospective, randomised pragmatic trial. *Bone Joint J*. 2013;95-B(5):629-635.

A2. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med*. 2002;347:81-88.

A3. Kirkley A, Birmingham TB, Litchfield RB, et al. A randomized trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med*. 2008;359:1097-1107.

A4. Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. *N Engl J Med*. 2013;369:2515-2524.

A5. Garbuz DS, Tanzer M, Greidanus NV, et al. The John Charnley Award: metal-on-metal hip resurfacing versus large-diameter head metal-on-metal total hip arthroplasty: a randomized clinical trial. *Clin Orthop Relat Res*. 2010;468:318-325.

A6. Sobieraj DM, Coleman CI, Tongbram V, et al. Comparative effectiveness of low-molecular-weight heparins versus other anticoagulants in major orthopedic surgery: a systematic review and meta-analysis. *Pharmacotherapy*. 2012;32:799-808.

A7. Adam SS, McDuffie JR, Lachiewicz PF, et al. Comparative effectiveness of new oral anticoagulants and standard thromboprophylaxis in patients having total hip or knee replacement: a systematic review. *Ann Intern Med*. 2013;159:275-284.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Kurtz SM, Ong KL, Lau E, et al. Impact of the economic downturn on total joint replacement demand in the United States: updated projections to 2021. *J Bone Joint Surg Am*. 2014;96:624-630.
2. Nepple JJ, Byrd JW, Siebenrock KA, et al. Overview of treatment options, clinical results, and controversies in the management of femoroacetabular impingement. *J Am Acad Orthop Surg*. 2013;21(suppl 1):S53-S58.
3. Lalmohamed A, Vestergaard P, de Boer A, et al. Changes in mortality patterns following total hip or knee arthroplasty over the past two decades: a nationwide cohort study. *Arthritis Rheumatol*. 2014; 66:311-318.
4. Langton DJ, Jameson SS, Joyce TJ, et al. Accelerating failure rate of the ASR total hip replacement. *J Bone Joint Surg Br*. 2011;93:1011-1016.
5. Bragdon CR, Doerner M, Martell J, et al., for the Multicenter Study Group. The 2012 John Charnley Award: clinical multicenter studies of the wear performance of highly crosslinked remelted polyethylene in THA. *Clin Orthop Relat Res*. 2013;471:393-402.
6. Sobieraj DM, Lee S, Coleman CI, et al. Prolonged versus standard-duration venous thromboprophylaxis in major orthopedic surgery: a systematic review. *Ann Intern Med*. 2012;156:720-727.
7. Kwiatt ME, Seamon MJ. Fat embolism syndrome. *Int J Crit Illn Inj Sci*. 2013;3:64-68.

REVIEW QUESTIONS

1. A 47-year-old man presents with groin pain of 2 years' duration. The pain is worse with weight-bearing activity, such as standing or walking. On physical examination, the hip demonstrates evidence of limited flexion and internal rotation. He has tried nonsurgical measures, such as oral nonsteroidal anti-inflammatory drugs (NSAIDs), physical therapy, and rest; none has provided any lasting relief. Radiographs demonstrate complete loss of the cartilage space in the hip joint. His bone structure and quality appear to be good. The patient feels extremely debilitated at this point because of his inability to perform activities of daily living (ADLs). In addition to ADLs, the patient would like to have the ability to run, in the future. Which of the following is the most appropriate recommendation?

A. Hip arthroscopy
B. Hip resurfacing
C. Total hip replacement
D. Fusion of the hip
E. Bracing of the hip

Answer: B Hip resurfacing is appropriate in this category of patient: a male patient, younger than 65 years. Hip arthroscopy is not likely to alleviate the symptoms in this patient because of full-thickness cartilage loss. Total hip replacement is an option, but the preservation of bone is desirable in younger patients, and impact sports are not advisable with a total hip replacement. Fusion of the hip has essentially become a historical operation: few surgeons perform this anymore because of the superior function of a hip replacement. Bracing of the hip would not be expected to help in situations of complete cartilage loss.

2. A 67-year-old woman presents with knee pain and an increasing deformity over the past 2 years. She notices that her knees are becoming "more bowlegged." With any prolonged weight-bearing activity, she experiences diffuse pain within the knee joint. In addition, the knee swells up on occasion, causing difficulty bending. She has only tried oral supplements such as glucosamine and chondroitin sulfate sporadically for pain relief. Physical examination reveals medial joint line tenderness and a small joint effusion. Her radiographs demonstrate moderate to severe joint space narrowing. The patient desires to return to walking activities and improve motion in her knee. What is the most appropriate recommendation for this patient?

A. Immediate total knee replacement surgery
B. Knee arthroscopy
C. Knee fusion
D. Partial knee replacement
E. A trial of nonsurgical measures, such as weight loss, physical therapy, and bracing

Answer: E Given the patient's moderate joint space narrowing and lack of prior nonsurgical interventions, a trial of weight loss, physical therapy, and bracing is worthwhile. Knee arthritis can respond well to this treatment regimen, reducing the stresses on the knee joint and helping the patient delay total knee replacement surgery.

3. Inflammatory arthritis can be distinguished from osteoarthritis by which of the following?

A. Mechanical overload
B. Sclerosis
C. Symmetrical joint space narrowing
D. Cystic changes
E. Osteophyte formation

Answer: C The hallmark of inflammatory arthritis is the presence of symmetrical joint space narrowing, a feature best appreciated radiographically. All the other characteristics are those of degenerative disease processes.

4. The advantages of the use of regional (epidural) anesthesia in lower extremity orthopedic surgery include which of the following?

A. Reduction in blood loss
B. Reduction in postoperative thromboembolic complications
C. Reduction in postoperative pulmonary complications
D. Facilitation of postoperative pain control
E. All of the above

Answer: E Regional anesthesia is a popular and commonly employed anesthesiology technique for patients undergoing lower extremity orthopedic procedures for a variety of reasons. Each of the potential advantages listed has been demonstrated with this form of anesthesia in this setting.

5. A 52-year-old obese man undergoes bilateral total hip replacement for severely symptomatic osteoarthritis of the hip. During the surgical procedure, after cementing of the first hip, the anesthesiologist notes a rise in the patient's pulmonary artery pressures that persists through the completion of the surgery. In the recovery room, the patient develops progressive hypoxemia and may require mechanical ventilation. Fat embolism is suspected. What additional sign of this condition is *least* likely to be helpful securing this diagnosis?

A. Bilateral pulmonary infiltrates
B. Mental confusion
C. Retinal hemorrhage
D. Petechial skin eruption
E. Signs of right heart strain

Answer: D All of the above phenomena have been described as important features of classic fat embolism syndrome. Nonetheless, in the postoperative arthroplasty setting, the classic petechial skin eruption is virtually never seen.

277

INTRODUCTION TO MICROBIAL DISEASE: HOST-PATHOGEN INTERACTIONS

W. MICHAEL SCHELD

Infectious diseases have profoundly influenced the course of human history. The "black death" (caused by *Yersinia pestis*) changed the social structure of medieval Europe, in the process eliminating approximately a third of the population. The outcomes of military campaigns have been altered by outbreaks of diseases such as dysentery and typhus. Examples include Napoleon's retreat from Russia, after typhus did more damage to his army than the opposition forces did; the decision by the French to sell the Louisiana Territory after French soldiers died from yellow fever in Cuba and the Gulf Coast; and the introduction of smallpox to the nonimmune population of the New World by Europeans, thus facilitating the "conquest" and the dawn of the colonial age. Malaria influenced the geographic and racial pattern and distribution of hemoglobins and erythrocyte antigens in Africa. The development of *Plasmodium falciparum* is inhibited by the presence of hemoglobin S, and Duffy blood group–negative erythrocytes are resistant to infection with *Plasmodium vivax.* Thus, populations with these erythrocyte factors are found in areas where malaria is common.

Infections are a major cause of morbidity and mortality in the world. Of the approximately 53 million deaths worldwide in 2009, at least a third were due to infectious diseases. In the United States, pneumonia is the fifth leading cause of death overall and the most common cause of death related to infection. Acquired immunodeficiency syndrome (AIDS) threatens to disrupt the social fabric in many countries of Africa and is severely distressing the health care system in the United States and other parts of the world. The year 2011 marked the 30th "anniversary" of the AIDS epidemic. Approximately 35.3 million people worldwide are currently infected with human immunodeficiency virus (HIV), and since 1981, approximately 36 million have died (≈600,000 in the United States alone). AIDS is now the leading cause of death in sub-Saharan Africa.

Infection can be defined as the multiplication of microbes (from viruses to multicellular parasites) in the tissues of the host. The host may or may not be symptomatic. For example, HIV infection may cause no overt signs or symptoms of illness for years. The definition of infection should also include the multiplication of microbes on the surface or in the lumen of the host that causes signs and symptoms of illness or disease. For example, toxin-producing strains of *Escherichia coli* may multiply in the gut and cause a diarrheal illness without invading tissues. Microbes can cause diseases without actually coming in contact with the host by virtue of toxin production. *Clostridium botulinum* may grow in certain improperly processed foods and produce a toxin that can be lethal on ingestion. A relatively trivial infection, such as that caused by *Clostridium tetani* in a small puncture wound, can cause devastating illness because of a toxin released from the organism growing in tissues. It has now become apparent that multiple virulence factors of microorganisms can be carried in tandem on so-called pathogenicity islands of the genome (the "virulome").

We live in a virtual sea of microorganisms, and all our body surfaces have indigenous bacterial flora. In fact, we are a "super organism" as our native flora outnumbers our own human cells by a ratio of 10 : 1. This normal flora actually protects us from infection. Reduction of gut colonization increases susceptibility to infection by pathogens such as *Salmonella enteritidis* serovar *typhimurium.* Bacteria that constitute the normal flora are thought to exert their protective effect by several mechanisms: (1) using nutrients and occupying an ecologic niche, thus competing with pathogens; (2) producing antibacterial substances that inhibit the growth of pathogens; and (3) inducing host immunity that is cross-reactive and effective against pathogens. These conclusions appear to be oversimplistic, however. For example, colonization of the gastrointestinal tract with *Bacteroides fragilis* expressing an immunodominant bacterial polysaccharide causes dendritic cell activation and induction of a T_H1-mediated response, leading to a splenic response characterized by normal numbers of $CD4^+$ T cells, lymphoid architecture, and systemic lymphocytic expansion. Thus, a single bacterial molecule in our gut is necessary to make us "immunologically fit." Indeed, it has become apparent that a healthy, diverse microbiome is vital to proper immune system function. The timing of changes in the microbiome can also be of crucial importance. For example, pregnant mice fed antibacterials pass along their altered gut microbiome to their offspring. The neonates, in turn, display decreased total number and composition of gut microbes that is associated with decreased numbers of circulating and bone marrow neutrophils. This disordered neutrophil homeostasis leads to impaired host defense and increased susceptibility to *E. coli* K1 and *Klebsiella pneumoniae* sepsis, classic neonatal pathogens in humans.[1] Furthermore, because children are often prescribed multiple courses of antibacterials, one must wonder if these (often unnecessary) exposures later predispose them to epidemic disorders, such as asthma, autoimmunity, inflammatory bowel disease, and obesity.

Only a small proportion of microbial species can be considered primary or professional pathogens, and even among these species, a relatively small number of clones have been shown to cause disease. For example, epidemic meningococcal meningitis and meningococcemia are due to a small number of clones of *Neisseria meningitidis,* and the worldwide explosion of penicillin-resistant *Streptococcus pneumoniae* can be traced to a few clones originating in South Africa and Spain. This observation supports the concept that pathogenic organisms are highly adapted to the pathogenic state and have developed characteristics that enable them to be transmitted, to attach to surfaces, to invade tissue, to avoid host defenses, and thus to cause disease. In contrast, opportunistic pathogens cause disease principally in impaired hosts, and these organisms, which may be harmless members of normal flora in healthy persons, can act as virulent invaders in patients with severe defects in host defense mechanisms. Although opportunistic infection has traditionally been viewed as the exploitation of a weakened host through physiologic stress or immunocompromise (or both) by relatively "avirulent" pathogens, this is an oversimplification. For example, *Pseudomonas aeruginosa* recognizes host immune activation, specifically by binding interferon-γ to a cell surface protein OprF, which in turn, through a quorum-sensing signaling system, leads to the overexpression of virulence determinants such as PA-I (lecA) and pyocyanin. Thus, bacteria have developed a "contingency system" that recognizes immunologic perturbations in the host and counters this response by the expression of virulence factors.

Pathogenic organisms may be acquired by several routes. For example, direct contact has been implicated in the acquisition of staphylococcal disease. Airborne spread, usually by droplet nuclei, occurs in respiratory diseases such as influenza, in severe acute respiratory syndrome (SARS), and in the recently recognized Middle East respiratory syndrome (MERS). Contaminated water is the usual vehicle in *Giardia* infection and typhoid fever. Food-borne toxic illnesses may be caused by extracellular toxins produced by *Clostridium perfringens* and *Staphylococcus aureus.* Blood and blood products may be vectors for transmitting hepatitis B and C viruses as well as HIV. Sexual transmission is also important for these agents and for a variety of other pathogens, including *Treponema pallidum* (syphilis), *Neisseria gonorrhoeae* (gonorrhea), and *Chlamydia trachomatis* (nonspecific urethritis). The fetus may be infected in utero, and the infection may be devastating if the agent is rubella virus, cytomegalovirus, or parvovirus B19. Arthropod vectors may be important, as illustrated by mosquitoes for malaria and dengue, ticks for Lyme disease and ehrlichiosis, and lice for typhus.

Pathogens are able to cause disease because of a finely tuned array of adaptations, including the ability to attach to appropriate cells, often mediated by specialized structures such as the pili on gram-negative rods. Microbes such as *Shigella* species have the ability to invade cells and cause damage. Toxins may act at a distance or may intoxicate only infected cells. Pathogens have the ability to thwart host defenses by a variety of ingenious maneuvers. The antiphagocytic coat of the pneumococcus is an example. Organisms may change their surface antigen display at an astonishingly rapid rate to outmaneuver the host immune system. Examples include influenza virus and trypanosomes. Certain pathogens (e.g., *Toxoplasma gondii*) have the ability to inhibit the respiratory burst of phagocytes, and others (e.g., *Streptococcus pyogenes*) can destroy phagocytic cells that have engulfed them. The environment plays an important role in infection, both in transmission and in the host's ability to combat the invader. The humidity and temperature of air may affect the infectivity of airborne pathogens. The sanitary state of food and water, woefully lacking in many areas of the developing world, is an important factor in the acquisition of enteric pathogens, one of the major causes of mortality, morbidity, and disability, such as physical and mental developmental delay leading to poor performance in school. The malaria associated with the "bad air" of swamps is, in fact, due to the mosquitoes there, but the

environmental association was appropriate. The nutritional status of the host is clearly a significant factor in certain infectious diseases. It is likely that micronutrient deficiency contributes to the invasion and multiplication of certain pathogens. A new concept is the possibility that infectious diseases cause malnutrition through a vicious circle of diarrhea leading to dehydration and poor oral intake, resulting in secondary diarrhea with a propensity for "stunting" and delaying intellectual development. Establishment of infection is a complicated interplay of factors involving the microbe, the host, and the environment.

Host reaction to infection may result in illness. For example, previous infection with *Campylobacter jejuni* is responsible for about 40% of cases of Guillain-Barré syndrome. The mechanism is thought to be the production of antibodies against *C. jejuni* lipopolysaccharides that cross-react with gangliosides in peripheral nerves. Similarly, much of the damage resulting from meningitis is due to the host's response to invading bacterial pathogens.

With some exceptions, infectious diseases are often treatable and curable. Thus, it is important to make an accurate etiologic diagnosis and to institute appropriate therapy promptly. In acute infections such as pneumonia, meningitis, or sepsis, rapid institution of therapy may be life-saving; thus, a presumptive etiologic diagnosis should be established before a definitive diagnosis. This presumptive diagnosis is based on the history, physical examination, epidemiology of illness in the community, and rapid techniques such as microscopic examination of appropriate gram-stained specimens or molecular techniques such as antigen detection or polymerase chain reaction. Antimicrobial therapy can then be instituted for the presumptive etiologic agents, but it must be reevaluated as more definitive diagnostic information becomes available.

The study as well as the understanding of infectious diseases is a dynamic process. A number of factors or themes of current interest contribute to this conclusion, including the following.

EMERGING INFECTIONS

The most obvious is AIDS, but recent examples with a major impact on the public health in the United States include community-associated methicillin-resistant *S. aureus*, a hypervirulent strain of *Clostridium difficile*, the 2009 H1N1 influenza, and multidrug-resistant gram-negative bacteria, such as carbapenamase-producing Enterobacteriaceae. More than 400 new, emerging infectious diseases have been described in the last 70 years; approximately 60% are zoonoses associated with geographic "hotspots." Their emergence is driven largely by ecologic, socioeconomic, and environmental factors.[2]

GENOMICS AND OTHER "OMICS"

The exact sequence of the genome of more than 2000 microbes relevant to humans has been determined. This new information, in concert with genomic information from multicellular organisms such as the *Anopheles* mosquito, offers significant promise for the development of new therapies and vaccines.[3] Careful analysis of the genomes of pathogens will continue to yield important information about the pathogenesis of infection. For example, genome sequencing of group A streptococci, collected over time with relevant robust clinical information, has detected the acquisition of new determinants (often by prophage) responsible for increased virulence and resulting in toxic shock syndrome, necrotizing fasciitis, or both, even within a single patient with sequential samples. Proteomics, transcriptomics, metabolomics, and virulomics have transformed research on infectious diseases and promise significant improvements in diagnostics and therapeutics in the future.

GENETIC FACTORS ALTERING SUSCEPTIBILITY TO INFECTION AND THE RESPONSE TO INFECTIOUS DISEASES

This field promises new and significant information relevant to the wide variety of responses to infectious diseases in humans. For example, an overvigorous response, with generation of tumor necrosis factor-α, may accentuate the development of cerebral complications in falciparum malaria. Analysis of single-nucleotide polymorphisms of the human genome will lead to an enhanced understanding of two fundamental issues in infectious diseases: why invasive, overt disease develops in only a small fraction of individuals colonized with a given microbe, and why infections are more severe in some people than in others. Variants in genes that encode molecules that mediate attachment, pathogen recognition, inflammatory cytokine response, and innate and adaptive immunity are being identified at an astonishing rate.

INNATE IMMUNITY

This is the most active field in the immunology of infectious diseases. The identification of pattern recognition receptors (e.g., Toll-like receptors and nucleotide oligomerization domain–like receptors) that recognize pathogen-associated molecular patterns, as well as endogenous substances reflecting tissue injury (e.g., alarmins), has revolutionized our understanding of the early host response to infection. Agonists or antagonists of Toll-like receptors have already entered clinical trials as adjuvant therapies or to improve the immunogenicity of vaccines. The other area that has exploded recently is the study of antimicrobial peptides (e.g., defensins, cathecidins, histatins, galectins) and their role in the early response to infectious disorders.

ANTIMICROBIAL RESISTANCE

The development of new antibacterial agents has slowed despite the burgeoning problem of antimicrobial resistance. This disconnect has been the focus of meetings among the pharmaceutical industry, the Infectious Diseases Society of America, the U.S. Food and Drug Administration, and internationally. Multiresistant pneumococci, vancomycin resistance in *S. aureus* and enterococci, and, perhaps most important, multidrug-resistant gram-negative bacilli are just a few examples. Some multidrug-resistant gram-negative bacilli are susceptible to only a few agents of "last resort," such as colistin or tigecycline; others are truly untreatable. Unfortunately, new agents active against these latter strains are years if not decades away from introduction.

THE ROLE OF INFECTIOUS AGENTS IN CHRONIC DISEASES

Many so-called idiopathic diseases may in fact have an infectious basis. Conditions for which there is some evidence (but not conclusive proof) of an infectious basis include diabetes, atherosclerosis, acute leukemia, collagen vascular diseases, and inflammatory bowel disease. Detection of "uncultivatable" microorganisms by newer techniques, such as 16S RNA analysis, may uncover agents responsible for "noninfectious" diseases or suggest a role in conditions that are considered infectious but in which the pathogen or pathogens are controversial (e.g., bacterial vaginosis). In addition, we know that hepatitis C virus, human papillomavirus, and *Helicobacter pylori* cause human cancers. Furthermore, changes in our own microbiome may lead to disease. Alterations in the gut microbiome are associated with obesity. Another recent example comes from experiments with mice lacking TLR5. These mice develop hyperphagia and hallmark features of the metabolic syndrome, including hyperlipidemia, hypertension, insulin resistance, and increased adiposity, associated with an altered gut microbiome. Further, transfer of this changed microbiota into germ-free wild-type mice induces most features of the metabolic syndrome in the recipients.[4] The explosion of new knowledge on the role of the human microbiome in health and disease has been so rapid and profound in the last decade that we thought a separate chapter on this subject was warranted (see Chapter 279).

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Deshmukh HS, Liu Y, Menkiti OR, et al. The microbiota regulates neutrophil homeostasis and host resistance to *Escherichia coli* K1 and sepsis in neonatal mice. *Nat Med*. 2014;20:524-530.
2. Gebreyes WA, Dupouy-Camet J, Newport MJ, et al. The global one health paradigm: challenges and opportunities for tackling infectious diseases at the human, animal, and environment interface in low-resource settings. *PLoS Negl Trop Dis*. 2014;8:e3257.
3. Relman DA. Microbial genomics and infectious diseases. *N Engl J Med*. 2011;365:347-357.
4. Vijay-Kumar M, Aitken JD, Carvalho FA, et al. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science*. 2010;328:228-231.

278

THE HUMAN MICROBIOME

ILSEUNG CHO AND MARTIN J. BLASER

Until recently, our understanding of the microbiota of humans (formerly called the normal flora) was handicapped by the limitations of traditional microbial cultivation. However, DNA-based analyses have expanded our horizons by generating enormous new data sets that can be mined for information about the composition and functional properties of our indigenous microbial communities (the human microbiome) and its collective genes (the metagenome). There has been great progress in characterizing the compositional range of the "normal" microbiome of healthy individuals.[1] Major clustering patterns at body sites such as the gastrointestinal tract provide new

ways to classify individuals and, possibly, their disease risks. Substantial progress has been made in defining the overarching concepts that advance the field. However, the subject is vast, and the implications for human health and disease are wide-ranging. Although most focus has been on bacteria, inquiries aimed at eukaryotes, archaea, viruses, and retroviruses also are needed.

CHARACTERIZING THE MICROBIOME

Animals have had residential microbes for hundreds of millions of years, and comparisons of the phylogenies of animal hosts and their microbiota suggest the existence of specific selection based on co-adaptation. Cooperative interactions between microbes and their hosts typically involve microbial participation in host functions such as defense, metabolism, and reproduction. The composition of the microbiome varies by anatomic site (Fig. 278-1). The primary determinant of community composition is anatomic location, and individuals can be grouped according to the major types present at specific sites, such as the gastrointestinal tract. However, dietary changes can rapidly cause substantial changes in intestinal composition and function. Similarly, nasopharyngeal microbiota in young children varies seasonally, and vaginal microbiota may vary with menses. The aggregate microbiota of an individual appears to have a host-specific pattern, but large perturbations, such as antibiotic exposure or enteric infections, can lead to transient disequilibria or to the development of new stable states.

Among all mammals, the microbiota is extensively conserved at high taxonomic levels, but variation increases at progressively lower taxonomic levels. Consequently, most of the sequences obtained from the mouse gut represent genera that are not detected in humans. Furthermore, intraspecies variability of the microbiota within the human population is also substantial. Indicator organisms such as *Helicobacter pylori* and *Streptococcus mutans* highlight some differences across the microbiota and metagenome among human ethnic groups; however, the extent of ethnic variation in overall metagenomic composition is unknown. The microbiomes of monozygotic twins are more closely related to one another than to those of unrelated individuals but not strikingly so, indicating important postnatal influences on composition.

The extensive lower-level taxonomic variation and large compositional differences observed even among highly related host organisms (e.g., mice and humans) are counterbalanced by the substantial conservation of metagenomic core functions (Fig. 278-2), reflecting the conservation of core bacterial properties involved in nucleic acid and protein synthesis and in metabolic and structural requirements. Of the more than 50 known phyla, most of the human microbiota is composed of fewer than 10 (and mostly six) phyla. Bacteria from other phyla, often of plant origin, that may be present in skin, nasopharyngeal, or gut samples are generally infrequent (<0.01% of the sequences) and probably represent transient carriage from food-borne and airborne exposures. The parallel needs of individual bacteria lead to competition for key substrates and to functional redundancy in the microbiota. Nevertheless, the enormous bacterial biomass also provides many unique or minimally redundant bacterial genes.

Resilience and Community Disturbance

Resilience, the ability to withstand disturbance, is a central concept in ecology. Whereas the microbiome of human adults appears highly resilient, the same may not be true for children. Because microbial population

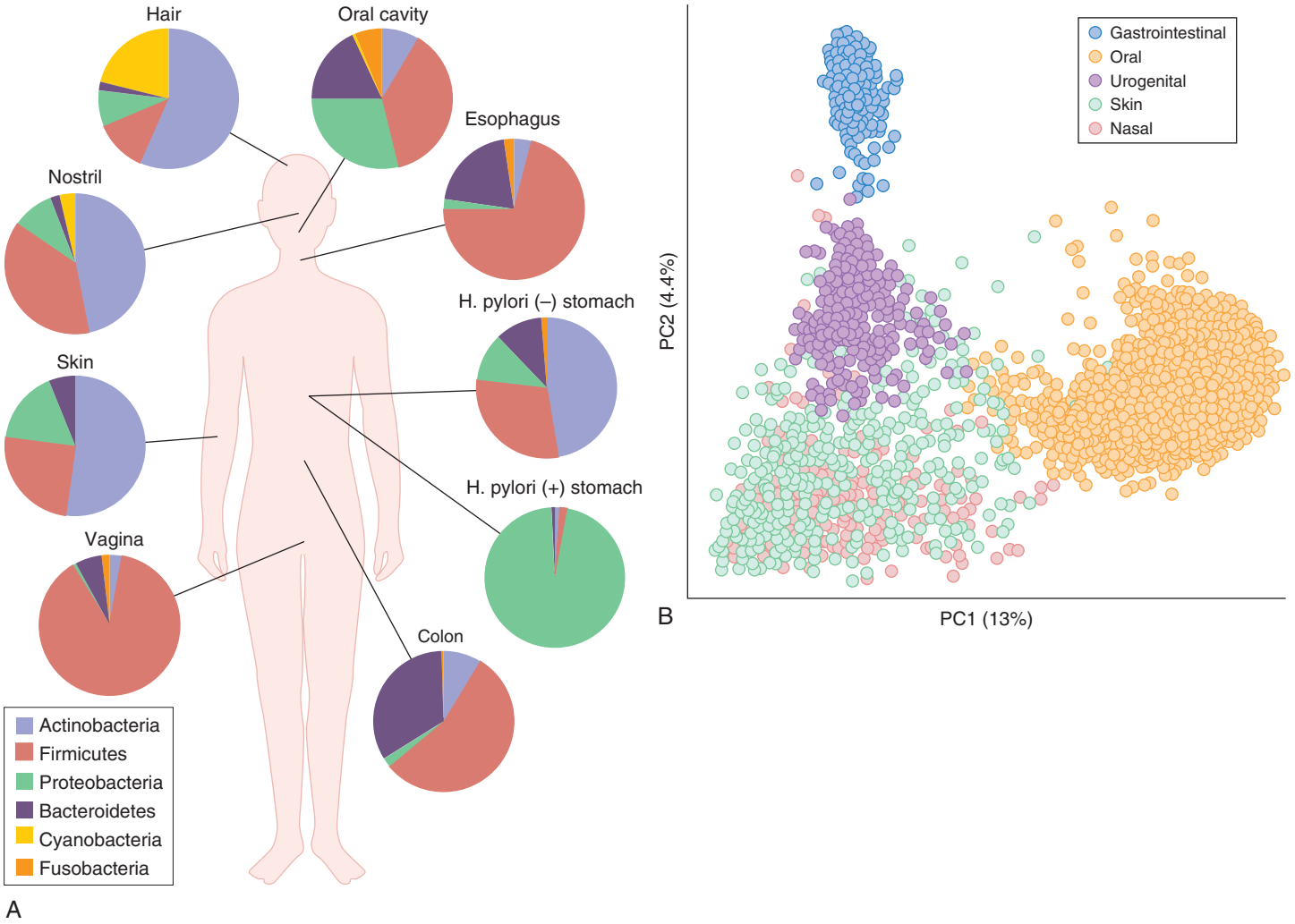

FIGURE 278-1. Compositional differences in the microbiome and metagenome by anatomic site. A, Substantial intersite microbiome variation. Higher level (e.g., phylum) taxonomic features display temporal (longitudinal) stability in individuals at specific anatomic sites. The figure indicates percentages of sequences at the taxonomic phylum level from selected references. Certain features, such as the presence or absence of *Helicobacter pylori*, are associated with alternative community composition states. (Modified from Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. *Nat Rev Genet*. 2012;13:260-270.) **B,** Taxonomic variation and spatial relationships. Primary clustering, shown by principal coordinate analysis, is by anatomic site. Gastrointestinal, oral, urogenital, and skin populations separate, but samples from the nasal cavity bridge both the oral and skin populations. Site-specific differences as well as observed interpersonal conservation provide a framework to understand biologic and pathologic significance of particular microbiome compositions. (From Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature.* 2012;486:207-214.)

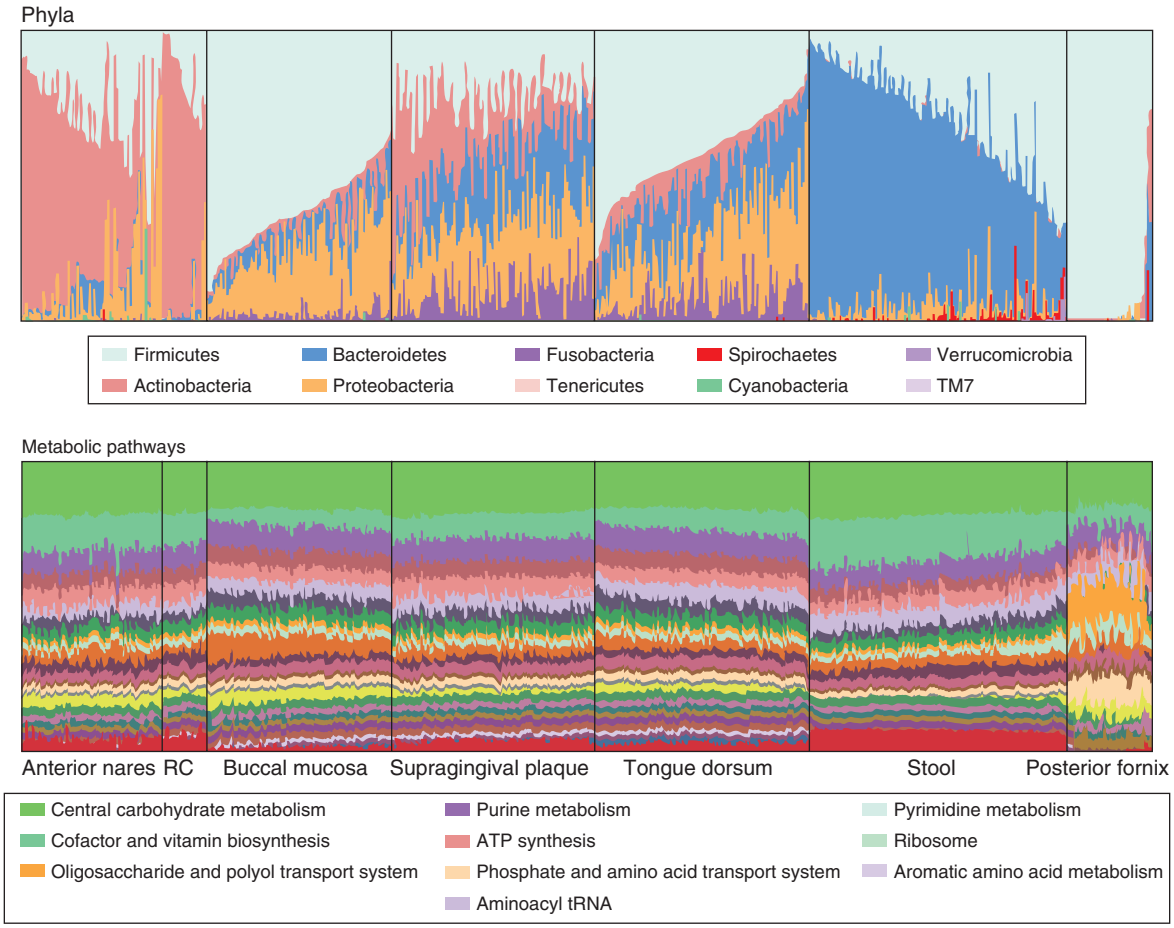

FIGURE 278-2. Variation and conservation in the microbiome and metagenomic functions of human anatomic sites. The phylogenetic composition of the microbiome at several anatomic sites shows substantial variation among individuals (*top panel*). Most but not all communities' membership consists of one or two dominant species (e.g., Firmicutes and Bacteroidetes in stool). However, the core metabolic functions of the microbial communities (*bottom panel*) are evenly distributed and highly conserved within a healthy population and across anatomic sites. Vertical bars represent microbiome samples of 242 individuals; colors represent relative abundance of microbial phyla or metabolic pathways in those samples. RC = retroauricular crease. (From Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486:207-214.)

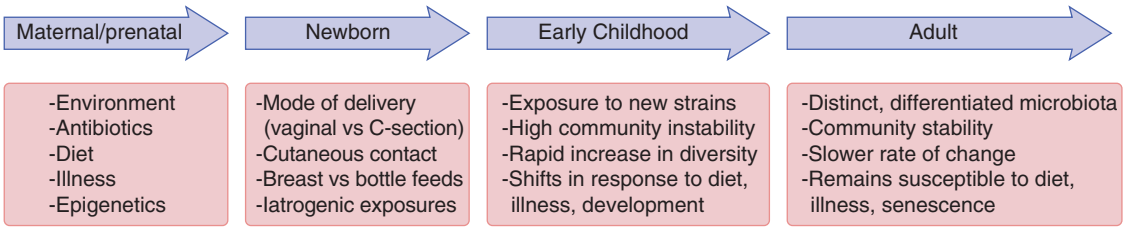

FIGURE 278-3. Factors and characteristics affecting acquisition and development of the microbiota from birth to adulthood. Numerous factors can affect the microbiome, beginning with prenatal exposures, such as to antibiotics, that can alter the composition of the maternal microbiome. Evidence exists that the mode of delivery can affect the initial composition of an infant's microbiome. The microbiome of young children is dynamic, but later in childhood (after 3 years of age), communities stabilize. At every stage of life, the microbiota are susceptible to external pressures, such as antibiotic exposure, diet, or disease.

structures appear more dynamic and developmental, resilience may be lower. An important natural experiment has been occurring during the past 70 years in which most of the world's population has been exposed to pharmacologic doses of antimicrobial agents. Such usage has been based on the implicit belief that the human microbiome is completely resilient and returns to the *status quo ante* after antibiotic-induced perturbation. These exposures also may cause medium- and long-term selection of resistant organisms and destabilization of the microbiome with new species compositions in the absence of further antibiotic exposure. Thus, despite the extensive resilience inherent in a complex ecosystem, there may be loss of recovery from continued perturbations, with important future implications for human health.

Extinctions

The human microbiome represents one or more complete ecosystems. Such communities may resist random perturbations, but if keystone species are lost, effects may cascade with secondary extinctions. High biodiversity diminishes this risk. In the short term, functional redundancy may mask extinction effects, but in the longer run, extinctions lead to loss of contingency responses that can cause ecologic crashes. Considering the importance of guilds of bacteria exploiting parallel and sequential metabolic pathways and the extent of medical interventions that are perturbing them, these concepts are germane to the human metagenome.

INFLUENCES ON THE MICROBIOTA DURING THE LIFE CYCLE

Differences in microbiota composition exist across body sites and between individuals.[2] However, changes are also evident across the human lifespan. Studies have answered important questions, such as whether temporal changes are life-stage specific and if they are predetermined by host genetic characteristics or environmental factors (Fig. 278-3).

Inheritance of Microbiota

The congruent phylogenies of mammals and their microbiota provide strong evidence for the inheritance of the microbiota. Increasing evidence supports maternal inheritance. Until the amniotic sac ruptures, a fetus has been considered to be essentially sterile. Immediately after vaginal delivery, microbial

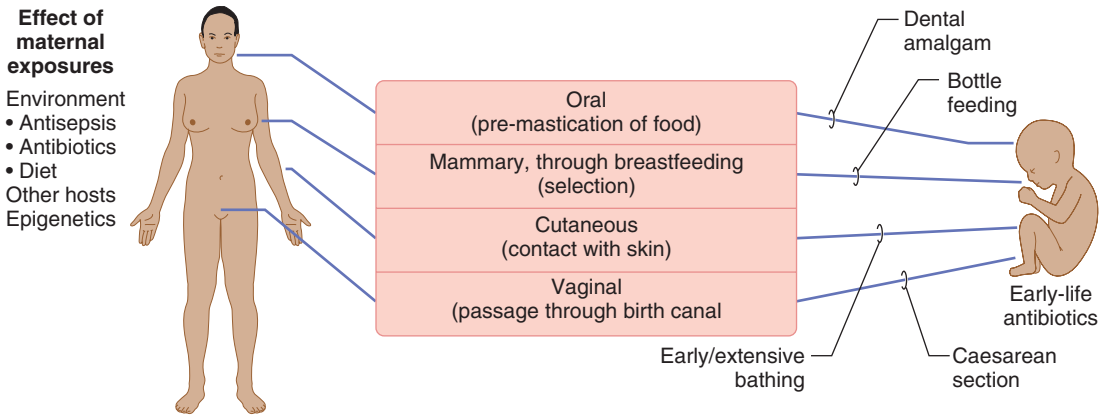

FIGURE 278-4. Factors affecting maternal exposures and vertical transmission of the microbiome between mother and child. Natural birth provides important opportunities for mother-to-child microbial transmission through a series of direct physical contacts. However, modern health practices can reduce flow of organisms and genes, thereby affecting the composition of the early infant microbiome. Many factors can affect early microbiomic development, including offspring genetics and epigenetics as well as dietary variation and environmental exposures. Organisms that have particular tissue or niche specificity explain conserved microbiome compositions at specific anatomic sites. (Modified from Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. *Nat Rev Genet*. 2012;13:260-270.)

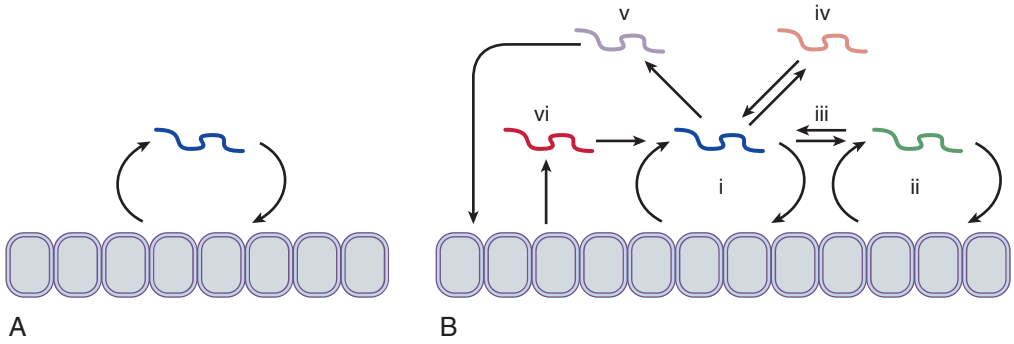

FIGURE 278-5. Equilibrium between co-evolved microbes and host cells. A, Single-organism equilibrium. In this model, there is a counter-regulatory interaction involving metabolic and physical signals between the microbe and host. **B,** Multiple organisms in equilibrium. This more complex model better approximates the interactions that occur in the human microbiome. Organisms may have individual equilibrium relationships (e.g., *i* and *ii*) with host cells. However, the interactions between these two microbes (*iii*) will affect their individual interactions. Similarly, another microbe (*iv*) may interact exclusively with an organism (*i*) but not with the host. Some interactions may be primarily unidirectional, as with a microbe (*v*) that is influenced by others and directly signals the host but does not receive direct host signals back. This may also proceed in the opposite direction as well (*vi*). (Modified from Plottel CS, Blaser MJ. Microbiome and malignancy. *Cell Host Microbe.* 2011;10:324-335.)

populations in the infant closely resemble those of the mother's vagina, with lactobacilli predominating. Because lactic acid–producing bacteria dominate both the vaginal canal and mother's milk, the initial bloom of lactobacilli cannot be considered accidental.

The multiple opportunities for the microbiota to be transferred from a mother to her baby may be disrupted by modern lifestyles.[3] Cesarean section instead of vaginal delivery is an obvious example of the potential impact of medical practice on microbiota composition, with substantial differences in founding populations that may persist for months. Such difference may alter infant immune responses, potentially leading to long-term consequences (Fig. 278-4).

Postnatal Influences on the Microbiota

Over a lifetime, each person develops a densely populated microbiome. The eruption of teeth is responsible for major successions in the oral microbiota, suggesting that succession may be a general property of human microbiome dynamics. Exposure (or not) to environmental microbes is another important but highly variable reservoir for the resident microbiota. Antibiotic use early in life produces major shifts in both microbiota characteristics and host developmental phenotypes, both on the farm and in experimental animals. Whether such precedents are applicable to human children is unknown but seems likely. If so, then both the timing of microbiome succession and the specific organisms present may affect metabolic, immunologic, and even cognitive development.

Microbiome Dynamics in Adults

Our knowledge of microbiome dynamics, especially age-related changes, during human adulthood is limited. We know that the postmenopause vaginal microbiota differs substantially from that during the reproductive period. Similarly, in the stomach, the age-related progressive development of gastric atrophy (enhanced by *H. pylori* presence) selects for gastric microbiota that are substantially different from the norm. Analogous changes may be occurring in other body sites as senescence advances. In the gut, the ratio of Bacteroidetes to Firmicutes changes with age. The specific interactions that the microbiome has with human hosts are enormously complex (Fig. 278-5), and further studies will be needed to define the specific pathways that inform the maintenance of health or the pathogenesis of disease. These concepts are particularly relevant to oncogenesis, which is generally age related. In the multistep Nordling hypothesis of oncogenesis, four to six somatic cell mutations are needed for cancer development. Microbiota shifts may contribute to this multistep process. Residential microbes can contribute to host cell mutagenesis by inflammation, increased cell proliferation, and production of promutagenic metabolites (e.g., butyrate).

DISEASE LINKS AND HEALTH IMPLICATIONS

How, then, does the microbiome affect human health? For many conditions, the challenge is to discover whether there is a causal link between microbiome variation and significant disease. Limitations in the definitions and stratification of clinical syndromes, including irritable bowel syndrome and nonulcer dyspepsia, reduce the potential of microbiome studies. Several examples of preliminary but promising observations follow.

Cutaneous Microbiome

The cutaneous microbiome could be involved in specific diseases, such as psoriasis, a chronic, idiopathic inflammatory condition. In some studies, Actinobacteria were significantly underrepresented in psoriatic lesions compared with unaffected skin both in psoriasis patients and in normal controls. Atopic dermatitis, another chronic inflammatory condition, has substantially increased in incidence in developed countries, suggesting a potential role for microbiome alterations. Classic atopic dermatitis occurs in areas, such as the antecubital and popliteal fossae, with similar microbial populations, suggesting a microbiome role. Similarly, *Propionibacterium acnes* has been implicated in the common dermatologic condition acne. Although *P. acnes* thrives in the cutaneous pilosebaceous units and secretes enzymes that cause local injury

and inflammation, continuing investigations are identifying *P. acnes* strain differences as well as other microbes in acne development. In chronic skin ulcers, often secondary to venous stasis or diabetes, cutaneous microbiome shifts occur. For example, Pseudomonadaceae are increased in patients with chronic ulcers treated with antibiotics, and Streptococcaceae are increased in diabetic ulcers.

Gastric Microbiome

The discovery of *H. pylori* overturned the dogma that the stomach is sterile. In *H. pylori*–negative persons, gastric microbiota diversity is high. Most of the prominent gastric species also are abundant in the oropharynx, indicating that many constituents are swallowed from more proximal sites or that close relatives of the oral microbiota colonize more distally. In contrast, among *H. pylori*–positive persons, *H. pylori* often accounts for more than 90% of sequence reads from the gastric microbiota, reducing the overall diversity in this discrete niche. *H. pylori* is a classical amphibont; the presence (or absence) of an *H. pylori*–dominated gastric microbiota is strongly associated with particular diseases with important age-related differences. Its presence increases risk for development of peptic ulcer disease, gastric mucosa-associated lymphoid tissue (MALT) tumors, and gastric adenocarcinoma, but its absence is associated with increased reflux esophagitis and childhood-onset asthma. These disease relationships illustrate our complex biologic interactions with our microbiota.

The Colonic Microbiota and Colorectal Cancer

Involvement of the colonic microbiota in the development of colorectal cancers has long been suspected.[4] Synthesis of short-chain fatty acids, in particular butyrate, may induce apoptosis, cell cycle arrest, and differentiation through Wnt signaling. Microbes may also be genotoxic to colonic epithelial cells, inducting polyploidy. The colonic microbiota also might promote colorectal cancer by eliciting host responses, for example, by stimulating exaggerated immune responses, potentially through T_H17 cells. Another link with colorectal cancer is suggested by antibiotic administration, altering the composition of the colonic microbiota, affecting expression of host genes involved in cell cycle regulation, and reducing epithelial proliferation. Early studies limited to identifying culture-dependent species, such as *Streptococcus bovis*, could not adequately assess anaerobic constituents. However, the anaerobic genus *Fusobacterium* has recently been associated with colorectal cancer, especially *Fusobacterium nucleatum*, a mucosally adherent, proinflammatory microbe first identified in the mouth. In colorectal cancer samples, *F. nucleatum* sequences were significantly enriched compared with control subjects, whereas both Bacteroidetes and Firmicutes were relatively depleted in those with *Fusobacterium*-rich malignant neoplasms. However, the causal direction of the association has not yet been ascertained. Finally, epigenetic phenomena, such as DNA hypermethylation, have been demonstrated to play an important role in the development of colorectal cancers through microsatellite instability and the sessile serrated adenoma pathway. It is possible that interactions with the microbiome may affect epigenetic pathways for colonic carcinogenesis.

The Colon Microbiota and Inflammatory Bowel Disease

The microbiome is essential for the activation of host immune responses.[5] For example, in mice, T_H17 cell differentiation in the lamina propria requires the presence of segmented filamentous bacteria, and polysaccharide A produced by *Bacteroides fragilis* mediates conversion of $CD4^+$ T cells into regulatory T cells. Accordingly, microbes are suspected to be involved in inflammatory bowel diseases (Chapter 141). Susceptibility to inflammatory bowel disease is associated with host polymorphisms in bacterial sensor genes such as nucleotide-binding oligomerization domain–containing protein 2 (*NOD2*; also known as *CARD15*) and toll-like receptor 4 (*TLR4*), and patients with inflammatory bowel disease sometimes improve after antibiotic treatment. Exposure to antibiotics in early childhood has been associated with significantly increased risk for Crohn's disease, suggesting that gut microbiome perturbations may enhance disease risk. Microbial diversity is significantly diminished in Crohn's disease, suggesting decreased gut microbiome resilience. Gut microbiome population structures of patients with ulcerative colitis or Crohn's disease are not normal and are clustered by disease. More recent investigations have implicated adherent-invasive *Escherichia coli* as a candidate pathogen in ileal Crohn's disease, given its ability to adhere to and invade epithelial cells and to replicate within macrophages. Specific bacteria among the Enterobacteriaceae may synergize with a disordered microbiome to increase the risk of ulcerative colitis. Among twins discordant for ulcerative colitis, those affected had significantly reduced bacterial diversity but increased Actinobacteria and Proteobacteria. Patients with Crohn's disease have overrepresentation of *Enterococcus faecium* and several Proteobacteria compared with controls. The microbial patterns observed for the conditions described are preliminary, and their specificity and causal direction have not been established.[6]

The Gut Microbiota and Diseases of the Liver

The gut microbiota may be involved in hepatic conditions, including nonalcoholic fatty liver disease, alcoholic steatosis, and hepatocellular carcinoma. The liver is the first solid organ exposed to the metabolic products generated by the gut microbiome, including acetaldehyde, ammonia, and phenols. Compared with germ-free mice, the presence of a microbiome in conventional mice led to suppression of intestinal epithelium angiopoietin-related protein 4, which inhibits lipoprotein lipase, increasing downstream triglyceride accumulation in the hepatic parenchyma and adipocytes. Chronic ethanol exposure disturbs the gut microbiome, but roles for the microbiome in steatosis are unresolved. Particular murine colonic commensals (e.g., *Helicobacter hepaticus*) promote the development of hepatocellular carcinoma. Patients with cirrhosis have a substantially altered microbiome, including community-wide changes at multiple taxonomic levels, with enrichment of Proteobacteria and Fusobacteria (phyla) and of Enterobacteriaceae, Veillonellaceae, and Streptococcaceae (family). Although many observations suggest links between microbiome composition and liver disease, definitive associations in humans are lacking.

The Gut Microbiota and Obesity

Genetically obese (*ob*/*ob*) mice have decreased ratios of Bacteroidetes to Firmicutes compared with lean (*ob*/+ and +/+ wild-type) siblings. Transplantation of gut microbiota from the obese (*ob*/*ob*) to germ-free mice conferred an obese phenotype, which shows the transmissibility of metabolic phenotypes; the transferred microbiomes had increased capacity for energy harvest. Studies have shown that transplantation of human microbiota can achieve parallel effects. In humans, relative Bacteroidetes proportions increase with weight loss. In monozygotic and dizygotic twins, obesity has been associated with decreased Bacteroidetes and diminished bacterial diversity, with enrichment of genes related to lipid and carbohydrate metabolism. In early life, the administration of antibiotics or alterations in diet may select for altered microbiota compositions, contributing to development of obesity.[7] Antibiotic use in human infancy, before the age of 6 months, has been significantly associated with obesity development. In contrast, in another study, perinatal administration of a *Lactobacillus rhamnosus* GG–based probiotic decreased excessive weight gain during childhood. These preliminary studies provide support for the concept that the early-life microbiota is modifiable, with alterations affecting risk of childhood-onset obesity. Studies are ongoing in adults to determine specific characteristics of the microbiome that may predict the risk of adiposity-related comorbidities.[8] Interventions used to treat obesity, such as the Roux-en-Y gastric bypass, substantially alter the gut microbiome, which may contribute to the metabolic effects.

The Gut Microbiota and Rheumatoid Arthritis

Altered regulation of host responses secondary to dysbiosis within the gut lumen could affect distant anatomic sites. This may be a mechanism in rheumatoid arthritis, another chronic idiopathic inflammatory condition. In mice, the presence of segmented filamentous bacteria in the gut microbiome causes local expansion of T_H17 cells that then migrate to peripheral immune compartments and activate B cells into antibody-producing plasma cells. Antibody production leads to immune-mediated destruction of the joints that mirrors rheumatoid arthritis. Substantial alterations in the gut microbiota have been identified in patients in the early stages of rheumatoid arthritis, consistent with a pathogenic role.

The Effect of the Microbiota on the Brain and Behavior

Early work suggests potential links between the gut microbiome and specific neurologic processes and disorders. Germ-free mice showed an exaggerated hypothalamic-pituitary response to stress that diminished after conventionalization with fecal matter. Subsequent investigations showed that the microbiome could alter neurotransmitter levels, such as brain-derived neurotropic factor and serotonin.[9] The rising incidence of autism in developed societies has led investigators to question whether shifts in gut microbes may play a causative role in these neurodevelopmental disorders. However, causal associations are difficult to determine from existing studies because many affected

children have concomitant gastrointestinal symptoms and were repeatedly treated with antibiotics. Microbiome effects on multiple sclerosis and other neurologic diseases have been hypothesized.

CAUSE OR EFFECT?

Microbiome analysis in humans has been largely observational, associating disease phenotypes with particular microbiota constituents. But which is causal? Does factor A cause factor B, does factor B cause factor A, or does factor C cause both A and B? Bradford Hill developed criteria to address the questions, "In what circumstances can we pass from this observed association to a verdict of causation? Upon what basis should we proceed to do so?" The criteria include the strength of association; its consistency, specificity, temporality, and biologic plausibility; and whether biologic gradients are present, experimental support exists, and support can be extrapolated from known causal relationships. These criteria are applicable to understanding either environmental or genetic (including metagenomic) causal roles.

For understanding causation and pathogenesis, model organisms provide an important approach. Animal models approximate some human diseases (e.g., asthma, atherosclerosis), but others (e.g., psoriasis) are not well reproduced. For those diseases that can be studied in model organisms, microbiota roles can be explored. Standard models of inbred mice are limited by their uncontrolled microbiome diversity. Certain disease states are well studied in these models, such as the effects of segmented filamentous bacteria on T_H17 development or the susceptibility to type 1 diabetes in nonobese diabetic mice. The use of germ-free mice eliminates microbiome variability until the introduction of "test" microbiota but requires specialized facilities, limiting their widespread use. The recent availability of germ-free animals from commercial sources permits their conventionalization without requiring such facilities, permitting direct observation of microbiota effects.

PERSPECTIVES

To better understand the implications of microbiota and metagenome variation in human health and disease, improved informatics tools are needed. The multidimensionality of the human and microbial phenotypes and the dynamic, nonlinear interactions challenge simple linear solutions.

The inherent compositional complexity of the microbiome limits investigations of microbe-associated diseases by classical approaches, such as Koch's postulates. Instead of single organisms associated with disease, community characteristics (composition and metagenomic functionality) may be more relevant. The principles of host interaction with pathogens and commensals contain many parallel features, which can help tutor the new field, but the nature of the selection for commensalism is more complex and highly dynamic. The scale of the interface suggests that microbiome-host interactions have important bearings on disease susceptibility; microbial effects on host metabolism and immunity provide "proof-of-principle" for the broader phenomenon of disease susceptibility. Modifying disease risk by altering metabolic, immunologic, or developmental pathways is an obvious strategy. Study of the microbiome, leading to new understandings of complex traits, will ultimately lead to new preventives, diagnostics, and therapies.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486:207-214.
2. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486:222-227.
3. Koren O, Goodrich JK, Cullender TC, et al. Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell*. 2012;150:470-480.
4. Plottel CS, Blaser MJ. Microbiome and malignancy. *Cell Host Microbe*. 2011;10:324-335.
5. Littman DR, Pamer EG. Role of the commensal microbiota in normal and pathogenic host immune responses. *Cell Host Microbe*. 2011;10:311-323.
6. Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology*. 2014;146:1489-1499.
7. Cho I, Yamanishi S, Cox L, et al. Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature*. 2012;488:621-626.
8. Le Chatelier E, Nielsen T, Qin J, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature*. 2013;500:541-546.
9. Mayer EA, Savidge T, Shulman RJ. Brain-gut microbiome interactions and functional bowel disorders. *Gastroenterology*. 2014;146:1500-1512.

REVIEW QUESTIONS

1. Which of the following statements is correct regarding the human microbiome?

A. The human microbiome is a DNA-based analysis of the composition of the indigenous microbial communities throughout an individual's gastrointestinal tract (formerly called normal flora).
B. In a healthy individual, the phylogenetic composition of the microbiome remains mostly constant from one anatomic site to another, but the core metabolic functions of these microbial communities vary greatly across anatomic sites.
C. Changes in microbiota composition occur during the entire life cycle of an individual, with the most dramatic changes developing after early childhood (after 3 years of age).
D. There is strong evidence for an inherited basis of the microbiota, particularly supporting maternal inheritance.
E. Metabolic and physical signaling occurs among the microbes but not between the microbes and their host, creating dynamic equilibrium relationships within the microbiome independent of its host.

Answer: D The congruent phylogenies of mammals and their microbiota provide strong evidence for the heritability of the microbiota. In humans, increasing evidence supports a maternal inheritance. Choice A is incorrect because the human microbiome involves much more than the gastrointestinal tract. Indeed, the presence of major clustering patterns at various body sites provides new ways to classify individuals and their potential for disease risks (see Fig. 278-1). Choice B is incorrect because the reverse is true: it is the phylogenetic composition of the microbiome that varies greatly from one anatomic site to another, whereas the core metabolic functions remain remarkably constant and conserved across anatomic sites (see Fig. 278-2). The most dramatic changes in the microbiome occur in the prenatal and neonatal periods, but the microbiome after the age of 3 years tends to stabilize. At every stage of life, the microbiota is susceptible to external pressures, diet or disease in the host, with constant metabolic and physical signaling occurring between microbes and host cells, not just among the microbes themselves.

2. An example of a determinant that is *not* known to be associated with variation in the composition of microbial communities at different sites is

A. Colonic microbiota varying with colon cancer
B. Vaginal microbiota varying with menses
C. Periodontal microbiota varying with tooth brushing
D. Gastrointestinal tract microbiota varying with diet
E. Nasopharyngeal microbiota in young children varying with seasonal change

Answer: C Cooperative interactions between microbes and their hosts typically involve microbial participation in host functions such as defense, metabolism, and reproduction. All of the above except C are known to cause perturbations. Large perturbations, such as antibiotic exposure or enteric infections, can lead to transient disequilibria or even the development of new stable states.

3. For which of the following conditions does the presence of *Helicobacter pylori* alter the risk?

A. Gastric cancer
B. Peptic ulcer disease
C. Reflux esophagitis
D. Childhood-onset asthma
E. All of the above

Answer: E In the absence of *H. pylori*, gastric microbiota diversity is high (although it was once believed that the stomach is sterile). In contrast, among *H. pylori*–positive individuals, this organism accounts for more than 90% of the sequence reads from gastric microbiota. The presence of *H. pylori* increases the risk for peptic ulcer disease, gastric cancer, and mucosa-associated lymphoid tissue (MALT) lymphomas. On the other hand, the absence of *H. pylori* increases the risk of reflux esophagitis and childhood-onset asthma. (See section on gastric microbiome under Disease Links and Health Implications.)

4. Which of the following disease links with gut microbiota is correct?

A. Antibiotic use in infancy is associated with obesity development.
B. Shifts in gut microbes in infancy play a role in causation of autism.
C. Changes in gut microbiota are pathogenic in the early stages of systemic lupus erythematosus.
D. Disturbances in the gut microbiome by chronic alcohol exposure contribute to development of alcoholic steatosis of the liver.
E. None of the above

Answer: A Most microbiome associations with human disease have been observational in nature and therefore do not prove cause-and-effect relationships (see Cause or Effect?). Although a causal relationship cannot be considered definitive at this time, antibiotic use in human infancy, before the age of 6 months, has been significantly associated with later obesity development. Furthermore, perinatal administration of *Lactobacillus*-based probiotics has been shown to decrease excessive weight gain during childhood. These and other observations have suggested that early-life microbiota are modifiable, with alterations affecting the risk of childhood-onset obesity (see The Gut Microbiota and Obesity). At this point, associations between shifts in gut microbes and autism or alcoholic steatosis (although chronic alcohol exposure does alter the gut microbiome) are entirely observational and preliminary. Substantial alterations in the gut microbiota have been identified in patients with early rheumatoid arthritis, consistent with a pathogenic role, but not with systemic lupus erythematosus. (See corresponding sections under Disease Links and Health Implications.)

279

PRINCIPLES OF ANTI-INFECTIVE THERAPY

GEORGE M. ELIOPOULOS

Among the pharmaceutical agents used in the treatment of human disease, antimicrobial agents are distinctive because they target invading microorganisms rather than abnormal human cellular functions. As a result, to select an appropriate antimicrobial regimen, it is necessary to consider both the activity of the agent against the known or suspected pathogen and the effects that agent might have on the individual under treatment. Although the term *anti-infective agent* can be used more broadly to include substances that ameliorate infection by altering the virulence of the pathogen or modulating the host's response to infection, for purposes of this chapter, *anti-infective agent* and *antimicrobial agent* are used interchangeably to refer to drugs that inhibit the growth of microbial pathogens. This chapter focuses primarily on agents directed against bacterial pathogens, although many parallels can be drawn to the use of antimicrobial agents for the treatment of fungal, viral, or parasitic infections.

On the time scale of human history, the modern antibiotic era is short. Since the introduction of penicillin for general clinical use in the mid-1940s, the numerous antimicrobial agents developed for human use have saved countless lives and have led to amazing advances in cancer chemotherapy, organ transplantation, and implant surgery that have improved and extended the lives of many others. Unfortunately, over time, resistance to available antibiotics has become widespread among many common bacterial pathogens,[1] making the selection of appropriate antimicrobial regimens ever more challenging and threatening to thrust an unfortunate few into a situation resembling the pre-antibiotic era.

SELECTING ANTIMICROBIAL THERAPY TARGETING THE PATHOGEN

Empirical Antimicrobial Therapy

In most instances, selection of the initial antimicrobial therapy proceeds empirically, before a causative organism is identified or tested for susceptibility to antimicrobial agents. The clinician's first decision is whether a patient's symptoms are likely to represent infection. Fever may result from neoplastic, rheumatologic, or other noninfectious processes and does not necessarily imply the presence of infection. Noninfectious causes of fever, such as deep vein thrombophlebitis, drug reaction, and vasculitis, may pose just as great a risk to the patient as infection and must not be overlooked.

Additional symptoms, signs, and laboratory or radiographic data usually help define whether infection is likely and, if so, localize the organ systems involved. This information allows an initial prediction about the organisms likely to be involved. For example, if the initial data cause one to suspect a diagnosis of community-acquired pneumonia in a previously healthy person who does not have any unusual exposures, *Streptococcus pneumoniae* and atypical bacteria such as *Mycoplasma pneumoniae* or *Chlamydophila pneumoniae* would be prominent on the list of potential pathogens to be targeted in selecting antimicrobial therapy. Examination of a gram-stained slide of expectorated sputum may provide valuable information. The prominent appearance of gram-positive cocci in clusters, for example, would alert the clinician to the possible presence of *Staphylococcus aureus*, many isolates of which are now methicillin resistant, and thus lead the clinician to select treatment options to include targeting of these organisms.

Guidance regarding the probable pathogens for site-specific infections and the susceptibility of these organisms to antimicrobial agents is available from a number of sources.[2] In some cases, the susceptibility of suspected pathogens can be predicted with a high degree of certainty. For example, *Streptococcus pyogenes* remains uniformly susceptible to penicillin G. In other instances, resistance has emerged to antimicrobials previously considered to be highly active against a species. Resistance rates for a given organism may vary widely by region, by health care institution, or even by patient care area within a hospital. For this reason, access to periodically updated, cumulative antibiotic susceptibility profile data specific to an institution can be important. Typically presented in tabular form, these "antibiograms" show the percentage of recently isolated bacterial pathogens that proved "susceptible" to the antibiotics tested and can help guide the selection of appropriate empirical regimens at that practice site.

There is mounting evidence that selection of an appropriate regimen (i.e., one that contains an antimicrobial that can be expected to inhibit the causative pathogen at the site of infection) and the prompt initiation of that empirical treatment result in improved clinical outcomes in those with serious infections. Published guidelines for the treatment of community-acquired pneumonia (Chapter 97) advise administration of the first dose of appropriate antimicrobial therapy while the patient is still in the emergency department.

Whenever possible, samples of purulent exudates, blood, or other body fluids suspected to be infected should be obtained for culture before antimicrobial therapy is started. Identification and susceptibility testing of the microorganisms detected can be used to direct subsequent definitive treatment. At times, however, this principle must be overridden. For example, when bacterial meningitis is suspected, antibiotic therapy (often with

adjunctive corticosteroids) must not be delayed when a lumbar puncture cannot be performed promptly to obtain material for culture. In such instances, blood samples taken for culture before the administration of antibiotics often reveal the causative organism, or the pathogen may grow from spinal fluid even if lumbar puncture is delayed.

Definitive Antimicrobial Therapy

Identification of the causative microorganism and determination of its susceptibility to available drugs are the basis for optimizing definitive antimicrobial regimens. Often, the antibiotics used for empirical therapy are appropriate for definitive therapy and can be continued. At other times, the results allow one to switch to a narrower spectrum, better tolerated, or less expensive antimicrobial.[3] In some instances, test results indicate the need to broaden the spectrum of an anti-infective regimen by adding or substituting agents active against pathogens inadequately targeted by the initial empirical regimen.

In almost all cases, it is desirable to test an infecting organism's susceptibility to antimicrobials that may be useful. To extend the example cited earlier, although it is not necessary to test the susceptibility of *S. pyogenes* to penicillin G, some isolates are resistant to macrolide antibiotics (e.g., erythromycin, azithromycin) and other drugs, so the testing of alternative agents might be useful for patients who are intolerant of β-lactam antibiotics. Even when the activity of certain antimicrobials can be predicted with great confidence, susceptibility testing is still useful. For example, surveillance studies examining hundreds of isolates have predicted that vancomycin or linezolid would inhibit virtually all *S. aureus* strains recovered from initial clinical specimens. Therefore, on statistical grounds, testing of these agents would not seem warranted; however, rare isolates resistant to these agents have now been encountered, and it is advantageous to detect such isolates for both therapeutic and epidemiologic purposes. For most bacterial pathogens, resistance to commonly used agents is sufficiently frequent that testing of antimicrobials being considered for definitive therapy is essential. Organisms of the family Enterobacteriaceae that are resistant to multiple antibiotics are isolated often enough, even among outpatients, that susceptibility to agents previously considered broadly active, including third-generation cephalosporins, fluoroquinolones, and aminoglycosides, is no longer ensured. Even more challenging problems of drug resistance are encountered among isolates of species such as *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, and *Stenotrophomonas maltophilia*.

Susceptibility Testing

Several methods are available for determining the susceptibility of a bacterial isolate to antimicrobial agents being considered for therapy. Tests most frequently used in clinical microbiology laboratories today are variations of three methods: serial dilution, disc diffusion, and gradient diffusion. The minimal inhibitory concentration (MIC) represents the lowest concentration of an antimicrobial tested that inhibits growth of the microorganism in test media.

In the dilution method, the antimicrobial is diluted in broth or agar to span a range of (usually) two-fold decreasing concentrations, and the medium is then inoculated with a standardized number of organisms. After incubation for a specified period (usually 16 to 24 hours) at 35° to 37° C, the series of dilution tubes or microtiter wells (for broth dilution) or agar plates (for agar dilution) is examined for growth. The MIC is determined by direct inspection as the lowest concentration that prevents turbidity of the broth or colony formation on agar. Modifications of this method allow the automation of many steps in the process, permitting more efficient test performance in clinical laboratories.

In the disc diffusion method, paper discs impregnated with a standardized amount of the antimicrobial are placed on an agar plate, the surface of which has been seeded with the bacterium to be tested. During incubation, the antimicrobial diffuses from the disc into the surrounding agar and inhibits growth of the seeded organism. After a specified period of incubation, the zone of growth inhibition around the disc is measured. By this method, the MIC is not determined directly. Instead, relying on accumulated data correlating inhibition zones with MICs, the measured zone is used to predict the susceptibility of the organism to the drug tested.

The gradient diffusion method is similar to the disc diffusion method, except that instead of using a round paper disc impregnated with a single concentration of the antimicrobial, this test uses a strip impregnated with the antimicrobial applied in a concentration gradient along its length. The strip is laid on the surface of an agar plate that has been inoculated with a suspension of the organism to be tested, and the plate is then incubated. By visually inspecting where the zone of growth inhibition on the agar surface intersects the strip (which is marked at intervals corresponding to MIC equivalents), it is possible to determine the MIC value directly.

To perform susceptibility studies and to interpret the results, it is necessary to identify the organism to be tested. This knowledge allows the selection of appropriate methods and interpretive criteria to determine whether an organism is "susceptible," "intermediate," or "resistant" to an antimicrobial on the basis of measurement of the MIC or the inhibition zone diameter. To illustrate this point, consider that an enterococcus is determined to be susceptible to penicillin if the MIC is less than or equal to 8 μg/mL, whereas for viridans streptococci, the corresponding breakpoint for susceptibility to penicillin is the MIC of 0.12 μg/mL. Thus, knowledge that the MIC of penicillin against a gram-positive coccus growing in short chains is 2 μg/mL does not allow the determination of whether it is susceptible to penicillin unless the organism has been identified.

Additional tests are sometimes required to fully assess susceptibility to an antimicrobial. For oxacillin-susceptible *S. aureus*, a test for penicillinase production is performed to assess susceptibility to penicillin G. For erythromycin-resistant, clindamycin-susceptible *S. aureus*, the laboratory may perform a supplementary D-zone test or equivalent before reporting the clindamycin result. A positive D-zone test result (i.e., blunting of the inhibition zone around a clindamycin disc in proximity to an erythromycin disc) predicts the presence of *erm* genes. Their product, a ribosomal methylase, can confer resistance to clindamycin if it is expressed; however, clindamycin is a poor inducer of this resistance trait (in contrast to erythromycin, which is a good inducer). A positive test result thus implies the presence of inducible resistance traits. Mutants with constitutive production of methylase can be selected during treatment, resulting in the emergence of clindamycin resistance and an increased risk of clinical failure when this drug is used to treat serious staphylococcal infections caused by strains with the *erm* gene.

In principle, tests for the presence of resistance genes, their products, or both can be used in place of phenotypic resistance testing. Such methods have the potential to provide answers more rapidly than can be obtained with the usual susceptibility tests of growth inhibition, which generally require several hours of incubation. At present, these tests are not yet widely employed, with the exception of testing for methicillin resistance by detection of the *mecA* gene or its product, penicillin-binding protein 2a, or testing for rifampin resistance by detection of resistance mutations in *Mycobacterium tuberculosis*. Newer technologies, such as matrix-assisted laser desorption ionization–time of flight mass spectrometry, are being explored as a means of providing not only more rapid identification of organisms but also their predicted susceptibility to antimicrobial agents.[4]

Bactericidal Activity

In some circumstances, an antimicrobial regimen that kills pathogenic microorganisms would be preferable to an alternative regimen that only inhibits growth of the pathogen. Bactericidal activity is desirable in the treatment of endocarditis or meningitis; in these infections, bacteriostatic agents have generally performed poorly, possibly because of inadequate host responses to infection at these sites. Tests to measure the bactericidal activity of an antibiotic in vitro have been developed. Bactericidal activity is usually defined as a 99.9% reduction in the number of viable colony-forming units relative to the inoculum density at a specified incubation time, which is usually 20 to 24 hours.

Despite the theoretical benefit of determining the bactericidal activity of an antibiotic or drug regimen, these tests are rarely used clinically because of several factors, including (1) the labor-intensive nature of the tests, (2) the potential for discordant results due to the various methods and criteria for determining bactericidal activity, and (3) the imperfect correlation between bactericidal activity measured in vitro and clinical outcomes observed.

SELECTING ANTIMICROBIAL THERAPY APPROPRIATE TO THE INFECTION AND PATIENT

Nature of the Infection

Determination that a pathogenic microorganism is susceptible to an antibiotic in vitro does not ensure that treatment with that drug will result in a successful clinical outcome. The antimicrobial must reach the site of infection in adequate concentration, which is generally assumed to be some multiple of the MIC, and it must demonstrate activity in the infection milieu. For some infections and antimicrobials, these requirements cannot easily be met.

A number of antimicrobials fail to penetrate into cerebrospinal fluid sufficiently well to permit their use for the treatment of bacterial meningitis in

adults. First-generation cephalosporins or aminoglycosides given intravenously do not enter the subarachnoid space well enough to allow their use as primary agents for treatment of this disease. Aminoglycosides have been administered by intrathecal or intraventricular instillation when needed for the treatment of gram-negative meningitis, but the availability of newer β-lactams with broad activity, high potency, and reasonable cerebrospinal fluid penetration when administered intravenously has largely eliminated the need for direct instillation of antimicrobials.

In other situations, antimicrobials may penetrate to the site of infection, only to be inactivated by local factors. For example, daptomycin is inactivated by interaction with pulmonary surfactant, so this antibiotic is not indicated for the treatment of bronchopneumonia, even though it is highly active against *S. pneumoniae* isolates in vitro. Antibiotics can also be inactivated by cellular debris or macromolecules present within abscesses, and some exhibit reduced potency at the low pH and reduced oxygen tensions prevailing at these sites. Finally, high densities of microorganisms within abscesses may elaborate sufficiently high concentrations of β-lactamases to inactivate some relatively labile β-lactam antibiotics. All these factors provide a rationale for the drainage of large abscesses as an adjunct to antimicrobial therapy.

Bacterial infections associated with foreign bodies such as artificial joints, cardiac pacemakers, or prosthetic heart valves can be particularly difficult to eradicate without removal of the foreign material. The reasons are not completely understood, but they relate, at least in part, to the presence of biofilm, which is composed of bacteria embedded within extracellular material that is adherent to the foreign body. Bacteria recovered from biofilms are metabolically different from and less susceptible to antimicrobial agents than planktonic cells (i.e., those freely suspended in liquid medium) of the same organism. Rifampin is often added to antimicrobial regimens for the treatment of infections involving prosthetic material. This inhibitor of RNA polymerase rapidly penetrates into biofilms and demonstrates relatively similar activity against both biofilm-associated and planktonic cells of a susceptible organism. However, because resistance to rifampin emerges rapidly, it is not used as a single agent in these circumstances; rifampin must be combined with a second active drug to minimize the risk that resistance will emerge. Despite such approaches, many infections involving implanted devices prove refractory to antimicrobial therapy alone and require removal of the foreign material for eradication.

Host Factors

After consideration of the nature of the infection and the antimicrobials determined in vitro to be active against a bacterial isolate (or likely to be active against probable pathogens when an isolate is not yet available), the ultimate choice of an antimicrobial regimen must take into account a number of additional patient-specific factors, some examples of which are examined in the following paragraphs.

Allergies

It is imperative to obtain a history of previous allergic reactions to antimicrobial agents. Some reactions are by nature so severe and potentially life-threatening that one must avoid using the same agent or drugs within the same class for which cross-reactivity is likely to occur. Examples of such reactions include an immediate hypersensitivity reaction to penicillin (e.g., hives, lip swelling, laryngeal edema, circulatory collapse) and a mucocutaneous bullous eruption from a sulfonamide (e.g., Stevens-Johnson syndrome).

In cases in which the allergic reaction was mild, such as a faint, self-limited rash in a patient receiving penicillin, the clinician may elect to use a related antimicrobial, such as a cephalosporin, when the probability of cross-sensitivity and the risk for a severe adverse outcome if a reaction were to occur are both assessed to be low. In these instances, careful monitoring of the patient for adverse reactions is essential. Rarely, for patients with significant allergies to potentially life-saving antimicrobial agents for which no alternative exists, desensitization of the patient to the antimicrobial is attempted so that the agent can be used. For example, desensitization protocols are available for penicillin and for trimethoprim-sulfamethoxazole. Because of the risks involved, these procedures may need to be performed in intensive care unit settings.

Pregnancy

A number of antimicrobial agents have the potential to cause fetal harm if they are administered to a pregnant woman. For example, tetracyclines can cause tooth discoloration and hypoplasia of dental enamel and are thus avoided in pregnant women and young children. Streptomycin given during pregnancy can cross the placenta, and evidence of eighth nerve toxicity in the child has been reported. A few other antimicrobials are labeled by the Food and Drug Administration as pregnancy category D (evidence of human risk) or are contraindicated because of fetal harm (category X). Many more antimicrobials, however, are assigned to category C; for these drugs, the potential risk to the fetus is based on animal studies. In designing antimicrobial regimens, the possibility of pregnancy should be considered in any woman of childbearing age so that the risks of candidate agents can be individually reviewed and the safest possible therapy selected.

Many antibiotics used to treat lactating women can be found in breast milk. Thus, it may be necessary to suspend breast-feeding during treatment if exposure of the infant to the drug must be avoided.

Pregnant women may be particularly susceptible to certain antimicrobial-associated toxicities. Death resulting from hepatic failure has been described in pregnant women receiving large doses of tetracycline. Potentially life-threatening hepatic steatosis has been observed in patients treated with a combination of the antiretroviral agents didanosine plus stavudine; pregnant women may be especially vulnerable to this toxic effect.

Age

For reasons discussed earlier, tetracycline antibiotics are avoided in children during tooth development to prevent discoloration and enamel hypoplasia of the permanent teeth. Because fluoroquinolone antimicrobials produce erosion of cartilage and arthropathy in juvenile animals, they are avoided in children when alternative agents are available. Recently, limited pediatric indications were added for ciprofloxacin (for complicated urinary tract infection and pyelonephritis and after inhalational exposure to anthrax) and for levofloxacin (inhalational exposure to anthrax and for plague). Musculoskeletal complaints appear to be more frequent in children treated with ciprofloxacin than with nonfluoroquinolone antimicrobials.

Pediatric dosing regimens differ from those appropriate for adults. Some agents, such as linezolid, are eliminated much more rapidly in young children (excluding preterm neonates) than in older children and adults, so higher doses may be required. In premature infants and neonates, renal function has not yet reached full capacity, and elimination of some antimicrobials may be delayed. Similarly, hepatic clearance activity is not fully developed in the very young, which has led to cardiovascular collapse and fatalities from chloramphenicol treatment. Absorption of oral antimicrobials may also differ with age if their absorption is dependent on gastric pH. The gastric pH of young children is higher than that of adults, and achlorhydria resulting in higher gastric pH is more common in adults older than 60 years than in younger adults. Thus, in young children and older adults, the absorption of oral drugs that are unstable in acid, such as penicillin G, may be higher than that in younger adults. In contrast, antimicrobials such as ketoconazole require gastric acid for absorption and may be less bioavailable in persons with reduced gastric acid production.

A curious association between the appearance of a rash and the patient's age and sex was noted during development of the fluoroquinolone antimicrobial gemifloxacin. In clinical studies, rash was more common in young women than in men and older women, suggesting that there may be hormonal influences on the risk for development of a rash.

Renal and Hepatic Function

Renal excretion and hepatobiliary excretion are the major routes of elimination for antimicrobial agents. Relatively few antibacterial agents can be administered without dosage adjustments in patients with renal dysfunction. Included among these drugs are nafcillin, ceftriaxone, doxycycline, azithromycin, and linezolid. Although linezolid exposure is not significantly altered, microbiologically inactive metabolites of the compound do accumulate in end-stage renal disease; what, if any, effect this has is unknown.

A number of antimicrobial agents require major dosage adjustments in the presence of renal dysfunction. The dosing interval for ceftazidime, usually administered every 8 hours in patients with normal renal function, is extended to once every 24 to 48 hours in persons with creatinine clearance below 10 mL/minute. Vancomycin is also administered at substantially increased dosing intervals or at smaller doses as renal function declines. Because of the increased efficiency of newer hemodialysis membranes in removing vancomycin, dosages are usually based on measured serum drug concentrations, and dosing may be required after each dialysis session.

In some instances, clearance of the antimicrobial agent is not affected by renal dysfunction, but excipients may accumulate, with the potential for toxic effects. For example, clearance of the antifungal agent voriconazole is not

dependent on renal function. However, its intravenous preparation contains the solubilizing agent sulfobutyl ether β-cyclodextrin, which does accumulate in the presence of renal insufficiency. The intravenous preparation should not be used in those with moderate to severe renal dysfunction, but the oral formulation, which does not contain β-cyclodextrin, can be administered. A number of other antimicrobials may accumulate in the presence of severe liver disease, with the possibility of an increased risk for adverse events. Antimicrobials requiring dose adjustments for various levels of hepatic insufficiency include metronidazole, chloramphenicol, tigecycline, caspofungin, and voriconazole. For ceftriaxone, dosage adjustments or careful monitoring may be required in patients with both hepatic and renal dysfunction.

Drug-Drug Interactions

One of the most important considerations in the selection of an appropriate antimicrobial regimen is to determine whether the drug or drugs will interact with other medications the patient is taking. Some drug-drug interactions can have severe or even fatal consequences. There are too many potential interactions to list comprehensively, but some examples are provided in this section. Fortunately, resources are now available that allow the clinician to check for potential drug-drug interactions when an antimicrobial agent is ordered.

A large number of antimicrobials are eliminated through cytochrome P-450 pathways. As a result, they may interfere with the elimination of other drugs cleared by these pathways, leading to their accumulation to potentially dangerous levels. Several macrolide antibacterials, some fluoroquinolones, and human immunodeficiency virus protease inhibitors are among the most likely antimicrobials to inhibit the clearance of other drugs. For example, use of the protease inhibitor darunavir/ritonavir is contraindicated with several drugs, including ergot derivatives, the neuroleptic drug pimozide, certain sedative-hypnotic agents, and others. Macrolides may result in increased levels of some 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors, which can lead to rhabdomyolysis.

In contrast, administration of rifampin induces the cytochrome P-450 system and may enhance the clearance of other drugs, some of which have narrow therapeutic windows. This may result in a number of important effects, including reduced effectiveness of oral contraceptives and increased warfarin requirements to maintain desired levels of anticoagulation. It is important to consider these potential interactions not only when starting rifampin therapy but also when *stopping* treatment. When rifampin is stopped, unless the previously increased dose of warfarin is adjusted downward accordingly, excessive anticoagulation and possibly serious bleeding can occur.

A number of other drug interactions have been described. Linezolid has weak monoamine oxidase inhibitor activity. As such, it has the potential to enhance the hypertensive effect of adrenergic agonists and has been associated with the development of serotonin syndrome in patients taking serotonergic antidepressants. Patients with this syndrome can exhibit a number of signs and symptoms, including fever, tachycardia, tremulousness, agitation, confusion, and clonus, occasionally with fatal results. Serotonin syndrome (Chapter 434) has been described in patients taking linezolid together with drugs other than selective serotonin re-uptake inhibitors; in principle, it could occur when linezolid is combined with any of a large number of agents that increase serotonin concentrations in the central nervous system.

Other Host Factors

Several additional host factors may influence the choice of a suitable antimicrobial regimen. Some antimicrobials have the potential to induce hemolysis in persons with glucose-6-phosphate dehydrogenase deficiency (Chapter 161). Among the drugs that should be avoided in these individuals are primaquine, nitrofurantoin, and various sulfonamides.

Coexisting diseases should also be taken into account. Use of fluoroquinolones or linezolid has been associated with abnormalities in glucose homeostasis. Hyperkalemia has been observed in patients with renal insufficiency during treatment with trimethoprim-sulfamethoxazole because trimethoprim blocks the renal excretion of potassium in the distal tubule.

In some cases, the patient's occupation might play a role in the selection of a treatment regimen. Antibiotics that can cause transient (minocycline) or permanent (streptomycin) dizziness or unsteadiness may create hazardous situations in those whose occupations require excellent balance. Antimicrobial agents with the potential to cause photosensitivity, such as tetracyclines, fluoroquinolones, trimethoprim, and sulfonamides, may be problematic in persons with significant sun exposure during outdoor employment or other activities.

ANTIMICROBIAL COMBINATIONS

It is common for hospitalized patients to receive more than one antimicrobial agent simultaneously. The rationale for using antimicrobials in combination is not always clearly defined, and there are a number of potential disadvantages to combination therapy. The basis for using combination therapy is considered in this section.

Reasons to Use Combination Antimicrobial Therapy

The clinical indications for using combination antimicrobial therapy can be divided into five categories. Two of these categories (empirical therapy and polymicrobial infections) relate to maximizing the likelihood that at least one agent in the combination will be active against known or suspected pathogens. The other three reasons (minimizing toxicity, preventing the emergence of resistance, and obtaining synergistic inhibition or killing) attempt to exploit the unique advantages of some combinations as compared with a component drug alone.

To Provide Broad Coverage during Empirical Therapy

A common reason for using more than one antimicrobial in hospitalized patients is to provide broad coverage against potential pathogens and to maximize the likelihood of delivering an active antimicrobial agent as quickly as possible to seriously ill patients. When the pathogen is unknown, the antimicrobial regimen often includes an agent broadly active against gram-positive bacteria, including methicillin-resistant *S. aureus* (MRSA), such as vancomycin, as well as an agent active against aerobic or facultative gram-negative bacteria. Selection of the latter is strongly influenced by local patterns of antimicrobial resistance specific to the institution and might include an extended-spectrum cephalosporin, an aminoglycoside, a fluoroquinolone, a β-lactam–β-lactamase inhibitor drug, or a carbapenem.[5] The last two choices also provide activity against gram-negative anaerobes. Alternatively, one could add an agent such as metronidazole to provide anaerobic activity. Because of the high frequency of antibiotic resistance in *P. aeruginosa* isolates, in settings in which that pathogen is encountered frequently, empirical use of two agents with antipseudomonal activity may be justified to maximize the likelihood that at least one of the agents will inhibit the organism.

Combination therapy is widely used in the initial treatment of hospitalized patients with community-acquired pneumonia. Commonly used regimens include a third-generation cephalosporin such as ceftriaxone with a macrolide or fluoroquinolone. This cephalosporin provides antimicrobial activity against *S. pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, and several other "typical" bacterial pathogens associated with community-acquired pneumonia, with the notable exception of MRSA. The macrolide azithromycin is commonly added to provide activity against "atypical" bacteria that cause pneumonia, including *M. pneumoniae*, *C. pneumoniae*, and *Legionella* species. One of the respiratory fluoroquinolones may also be added to attain coverage of the atypical organisms. Although fluoroquinolones approved for respiratory tract infections are likely to cover most or all of the organisms targeted by the cephalosporin, isolates of *S. pneumoniae* resistant to fluoroquinolones do exist, so guidelines recommend combination therapy in patients with severe pneumonia requiring hospitalization.

To Treat Documented Polymicrobial Infections

For many infections from which two or more pathogens are recovered, it is possible to provide adequate coverage with a single, broadly active antimicrobial agent. Switching to a single agent reduces the patient's exposure to potential antibiotic toxicities, is usually more convenient for nursing staff, and may be less expensive. For some patients, susceptibility profiles or allergies to broad-spectrum agents justify the use of antibiotic combinations for the treatment of polymicrobial infections.

To Attempt to Reduce Toxicity

It is theoretically possible to use two or more drugs of different classes with additive antimicrobial activities and independent toxicities, each at relatively low doses, to achieve sufficient potency while avoiding toxicity. However, there are no situations in which the approach of using submaximal doses of multiple agents is predictably effective in accomplishing this goal. This does not exclude the possibility that, in isolated instances, a successful response might be attained from additive effects when drugs with marginal activities are combined for treatment of infection due to multiply resistant organisms.

To Prevent the Emergence of Drug Resistance

The treatment of tuberculosis provides the paradigm for using combinations of drugs in an attempt to prevent the emergence of resistance to any one agent. The basis for this approach is that if resistance to two different agents occurs by independent mechanisms, the probability of resistance developing to both drugs is the product of the probability of resistance developing to each drug, which is likely to be very low, so resistance should not emerge. Similar reasoning has justified the use of combination regimens when rifampin is required for the treatment of nonmycobacterial infections. Rifampin is not used alone (with rare exceptions, such as brief courses for the eradication of meningococcal carriage) because resistance to this agent can emerge quickly. As mentioned earlier, rifampin is particularly useful in the treatment of infections related to foreign devices because of its activity against biofilm-associated bacteria. In such cases, it is combined with another active antimicrobial, such as vancomycin for coagulase-negative staphylococcal prosthetic valve endocarditis (usually with a brief course of gentamicin as well to reduce the bacterial inoculum further at the beginning of therapy) or a fluoroquinolone for orthopedic device–related infections.

It has been difficult to show unequivocally that combination therapy protects against the emergence of resistance to antimicrobial drugs in other situations, including infections caused by *P. aeruginosa* or *Enterobacter* species. There are two plausible explanations of why combinations do not prevent resistance predictably. First, there may be differential penetration of the two antimicrobials at an infected site or differences in activity at the site of infection. Thus, a more readily penetrating agent may be left relatively unprotected in a privileged site of infection. Second, for many commonly encountered bacteria, resistance mechanisms against unrelated antimicrobial classes may not be truly independent. Some bacterial efflux pumps recognize chemically unrelated substrates, so upregulation of pump activity may confer resistance to several classes of antimicrobials. In other instances, there may be coordinated upregulation of efflux mechanisms and downregulation of outer membrane protein channels (porins), again potentially conferring resistance simultaneously to two or more antimicrobial classes.

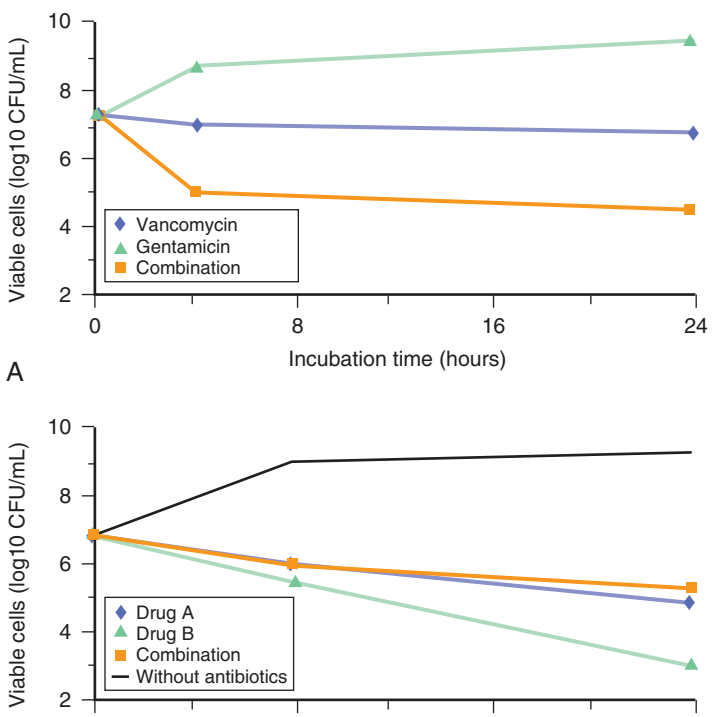

FIGURE 279-1. Bactericidal synergism and antagonism. A, Bactericidal synergism between vancomycin and gentamicin against an isolate of *Enterococcus* species. Killing by the combination of drugs is substantially greater than that by each agent alone. **B,** Antagonism of the bactericidal activity of drug B by the more slowly bactericidal drug A. Killing by the combination of drugs is less than that by drug B alone. Growth in the absence of antibiotics is also shown. CFU = colony-forming unit.

To Attain Synergism

Decades ago, the surprising benefits of using penicillin and streptomycin together for the treatment of enterococcal endocarditis were discovered empirically. Penicillin alone usually inhibits but does not kill enterococci, and failure rates were high when penicillin G was used alone to treat enterococcal endocarditis. Streptomycin has no significant activity against enterococci at clinically relevant concentrations. However, the combination results in bactericidal synergism in vitro and high cure rates in patients with enterococcal endocarditis. Detailed studies of this phenomenon demonstrated that in the presence of a cell wall–active antibiotic, uptake of the aminoglycoside into the bacterial cell increases substantially. Unfortunately, increasing rates of high-level resistance to streptomycin (MIC > 2000 μg/mL), gentamicin (MIC > 500 μg/mL), or both have nullified the benefit of such combinations against a substantial number of enterococcal isolates today. An example of bactericidal synergism between vancomycin and gentamicin against an *Enterococcus* isolate is illustrated in Figure 279-1A.

Combinations of cell wall–active agents plus aminoglycosides have been shown to achieve synergistic killing against a broad range of gram-positive and gram-negative bacteria when tested in vitro. Modest clinical benefits were shown when short courses of gentamicin were added to nafcillin for the treatment of *S. aureus* endocarditis, but at the cost of added nephrotoxicity. Against strains of viridans streptococci that are relatively insensitive to penicillin, the addition of an aminoglycoside for the first 2 weeks of a 4-week course of penicillin G is believed to result in a higher likelihood of cure.

Although it was once considered important in the treatment of gram-negative bacterial infections, especially in immunocompromised (e.g., neutropenic) patients, the clinical value of a synergistic combination of a cell wall–active agent and an aminoglycoside has been difficult to prove in recent experience. To a large extent, the introduction of agents with potent activity against gram-negative bacteria has diminished the perceived value of synergistic combinations. Nevertheless, there is some evidence that administration of two or more active drugs for empirical therapy may achieve a better outcome than is possible with a single active agent for *P. aeruginosa* infections, especially in neutropenic patients. The major value of combination therapy in this setting is to ensure that at least one active agent is administered promptly. The combination of sulfamethoxazole and trimethoprim, agents that block sequential steps in folic acid synthesis, can also achieve bactericidal (or bacteriostatic) synergism against a number of important gram-positive and gram-negative pathogens. Quinupristin and dalfopristin are streptogramin antibiotics that display inhibitory activity against gram-positive organisms. Combining these two agents, as is done in the commercial formulation, results in bactericidal synergism against organisms susceptible to both.

β-Lactam–β-lactamase inhibitor antimicrobials represent another example of synergistic combinations. Five drugs in this category are currently marketed in the United States: amoxicillin-clavulanate, ampicillin-sulbactam, ticarcillin-clavulanate, piperacillin-tazobactam, and ceftolozane-tazobactam. The β-lactamase inhibitors themselves, clavulanic acid, sulbactam, and tazobactam, are devoid of significant antimicrobial activity, with rare exceptions. However, by inhibiting common β-lactamases that are sensitive to these agents, the inhibitors restore the activity of the hydrolyzable companion penicillins against many target pathogens elaborating these enzymes.

Antagonism

Antibiotic combinations can sometimes result in microbiologic antagonism, such that the combination may have *reduced* activity compared with the most active single agent of the treatment regimen. Time-kill curves illustrating in vitro antagonism are shown in Figure 279-1B. In this example, the more slowly bactericidal drug A antagonizes the killing effect of the intrinsically more bactericidal drug B. Antagonistic interactions against *S. aureus* between less bactericidal (linezolid) and more bactericidal (vancomycin) antimicrobials have also been demonstrated in vivo in experimental endocarditis. In vitro antagonism can be demonstrated when certain β-lactams are tested in combination against gram-negative bacteria with inducible β-lactamases. Here, exposure to one β-lactam can de-repress the synthesis of inducible β-lactamases, which then degrade the second antibiotic.

It is uncommon to encounter clinically apparent antagonism between antibiotics in the patient care setting, in part because offending combinations are not likely to be used in routine clinical care today. However, if unusual antimicrobial combinations are used, in desperation, against isolates exhibiting

multiple drug resistance, it is possible that clinically relevant antagonism will be encountered more often in the future. Antagonism of bactericidal activities may also be difficult to detect in clinical practice because most common infections (with the exception of endocarditis and meningitis) do not unequivocally benefit from bactericidal therapy. As long as one agent maintains inhibitory activity, it is unlikely that failure resulting from antagonism will be observed.

CONSIDERATIONS IN ANTIMICROBIAL ADMINISTRATION

Route of Administration

In almost all instances, antimicrobial therapy for infections of mild to moderate severity that are treated in the outpatient setting can be undertaken with oral agents. There are notable exceptions, such as the use of intramuscular injections of benzathine penicillin for the treatment of syphilis or ceftriaxone for the treatment of otitis media or gonorrhea caused by strains resistant to oral agents.

Drugs such as levofloxacin and linezolid demonstrate virtually complete bioavailability when administered by the oral route in persons with normally functioning gastrointestinal tracts, and they can be used as an alternative to intravenous therapy in many patients with more serious infections. Even for these well-absorbed antimicrobials, however, treatment of seriously ill patients in the hospital is often initiated with intravenous formulations because of the uncertainty of gastrointestinal tract function under conditions of hemodynamic instability.

Antimicrobial therapy can be administered by other routes, including topical administration for the treatment of infected skin lesions (e.g., mupirocin or retapamulin ointments) and intravaginal administration for candidiasis (e.g., azole creams) or for bacterial vaginosis (e.g., metronidazole gel). Topical administration *onto* the eye is used to treat conjunctivitis or as adjunctive therapy for deeper infections; administration *into* the globe itself is a component of regimens for the treatment of endophthalmitis. Infections associated with peritoneal dialysis are frequently treated by intraperitoneal instillation of antimicrobials admixed with the dialysis solution. Rarely, direct administration into the thecal space or into the cerebral ventricles is necessary for the treatment of meningitis when the required antimicrobials do not achieve adequate concentrations in cerebrospinal fluid after systemic administration. For the treatment of *Clostridium difficile*–associated diarrhea, vancomycin reaches high concentrations in the intestine when it is given orally, but it is occasionally administered directly into the colon for the intraluminal treatment of severe infections.

The availability of protocols for the outpatient use of long venous catheters, whether inserted centrally or peripherally, has made it possible to administer antimicrobial agents that are not well absorbed orally. Thus, many patients who require long-term antibiotic treatment for infections such as endocarditis, osteomyelitis, neuroborreliosis, and other conditions can be treated as outpatients, often after an initial period of hospitalization for a full assessment of the infection, initiation of therapy, and stabilization of the medical condition. In addition to monitoring for adverse effects from the antibiotic itself, patients treated through indwelling intravenous devices require close observation for complications related to the catheter, such as thrombophlebitis, entry site infections, or line-related blood stream infections.

Pharmacodynamic Considerations

In recent years, the scientific basis for selection of a dosing regimen has extended well beyond empirical dosing strategies based primarily on the pharmacokinetic characteristics of antimicrobial agents. Pharmacodynamics relates the time course of antibiotic concentrations after dosing, the observed antimicrobial effects against likely pathogens, and the potential adverse effects of the agent.

Studies of the pharmacokinetic and pharmacodynamic properties of antimicrobial agents allow the prediction of their activities with various dosing regimens. For β-lactam antibiotics, the time during which the concentration of free drug (i.e., the non–protein-bound fraction) exceeds the MIC of the pathogen best relates to antimicrobial effectiveness in animal models. This provides the rationale for the frequent dosing schedules of β-lactams with short half-lives, such as penicillin G and the antistaphylococcal penicillins, and for the use of extended intravenous infusions when β-lactams are used to treat marginally susceptible organisms.

In contrast, the aminoglycosides and fluoroquinolones demonstrate concentration-dependent killing of bacteria. For these drugs, animal models show that the ratio of either peak concentration to MIC or the area under the 24-hour drug concentration curve to MIC better predicts effectiveness. With these agents, less frequent, higher dosing would be optimal (except perhaps when used with cell wall–active agents to achieve synergy). For the aminoglycosides, less frequent dosing may also allow more time for washout of the drug from the kidney, thus potentially minimizing the risk for nephrotoxicity.

For daptomycin, the adoption of once-daily dosing largely mitigated the muscle toxicity that had been seen with more frequent dosing and allowed the use of this agent for serious gram-positive infections.

MONITORING ANTIMICROBIAL CONCENTRATIONS

From a practical point of view, there are few situations in which assays to determine concentrations of antimicrobials in blood or body fluids are readily available. Commercial assays for the measurement of serum aminoglycoside concentrations are available and, because of these agents' great potential for toxicity, are used frequently. Commercial assays to measure vancomycin concentrations are also widely available. It may be prudent to monitor serum concentrations of vancomycin in patients with unstable renal function, those undergoing hemodialysis, patients at the extremes of body composition, or those with particularly serious infections in which high concentrations may be desirable. In some young adults, clearance of vancomycin may be so great that unexpectedly low concentrations result with the usual dosing regimens.

ADMINISTRATIVE ASPECTS OF ANTIMICROBIAL THERAPY

Formularies

In most practice settings today, the choice of antimicrobials is constrained in some way. For example, in hospitals and other facilities, institutional formularies may limit the choice of antimicrobial agents available, require special approval for the use of selected agents, or both. Such policies can, in principle, enhance efficiency by avoiding the need to stock and to dispense multiple agents with similar antimicrobial activities, minimize costs by allowing purchase of the most cost-effective alternatives, and potentially increase patient safety by encouraging clinical personnel to become familiar with a manageable number of agents. In the outpatient setting, it is common practice for health insurers to assign oral drugs to various tiers of coverage and copayment; as a result, there may be dramatic financial benefits or disincentives for the patient to receive specific agents. In both health care settings, the practitioner must be familiar with the options available to patients under these constraints.

Interchangeability

Although two drugs may have antimicrobial spectra that are so similar that only one need be included on a formulary, it is not always safe to assume that the activity of either agent can be predicted perfectly by susceptibility to the other. For example, for most bacterial species, the percentage of isolates susceptible to meropenem and imipenem will be roughly comparable. However, there are differences in mechanisms of resistance to these two carbapenems, so it is possible that a specific strain will be susceptible to one but resistant to the other. For serious infections, even when two drugs are considered interchangeable, one should determine susceptibility to the specific antimicrobial that is to be used.

Impact on the Institutional Environment

In contrast to other medications, which almost always affect only the patient receiving them, antimicrobial use can have a significant impact on the institutional environment as well.[6] As a result, it is sometimes necessary or desirable to manage the use of antimicrobial agents on an administrative level to avoid selective pressure leading to the spread of antibiotic resistance. Within institutions, antimicrobial-resistant organisms not only threaten the patient treated with the antimicrobial but also can be transmitted to other vulnerable persons, including those who have not been exposed to the drug. These considerations have given rise to the movement to promote antimicrobial stewardship in health care environments.[7]

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Llor C, Bjerrum L. Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. *Ther Adv Drug Saf*. 2014;5:229-241.
2. Gilbert DN, Chambers HF, Eliopoulos GM, et al., eds. The Sanford Guide to Antimicrobial Therapy. 44th ed. Sperryville, VA: Antimicrobial Therapy; 2014.
3. Braykov NP, Morgan DJ, Schweizer ML, et al. Assessment of empirical antibiotic therapy optimisation in six hospitals: an observational cohort study. *Lancet Infect Dis*. 2014;14:1220-1227.
4. Hrabák J, Chudácková E, Walková R. Matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry for detection of antibiotic resistance mechanisms: from research to routine diagnosis. *Clin Microbiol Rev*. 2013;26:103-114.
5. Hawkey PM, Livermore DM. Carbapenem antibiotics for serious infections. *BMJ*. 2012;344:e3236.
6. Cook PP, Gooch M. Long-term effects of an antimicrobial stewardship programme at a tertiary-care teaching hospital. *Int J Antimicrob Agents*. 2015;45:262-267.
7. Tamma PD, Cosgrove SE. Antimicrobial stewardship. *Infect Dis Clin North Am*. 2011;25:245-260.

REVIEW QUESTIONS

1. Which of the following host factors should be considered in selecting an antibiotic regimen for treatment of a patient?

A. Prior history of allergies
B. Whether the patient is currently pregnant or breast-feeding
C. Age and sex of the patient
D. Options A and B
E. Options A, B, and C

Answer: E The need to consider host factors A and B is obvious. The age and sex of the patient are also relevant. Some antibiotics (e.g., fluoroquinolones, tetracyclines) are generally avoided in children with only a few exceptions. In clinical studies, young women were more likely to develop rash to gemifloxacin than were men or older women.

2. Important justifications for use of antimicrobial combinations in clinical practice include all of the following *except*

A. To reduce drug toxicity by allowing use of half of the usual full doses of each of two agents
B. To attain bactericidal synergism for treating enterococcal endocarditis
C. To treat documented polymicrobial infections
D. To minimize emergence of resistance to rifampin in treating prosthetic device infections
E. To target a broader range of potential pathogens during empirical therapy

Answer: A Although using combinations to reduce doses of individual component drugs is a theoretically plausible consideration, it is not currently feasible to accomplish this in practice. Combinations are used to achieve bactericidal synergism to treat endocarditis (e.g., cell wall–active agent plus an aminoglycoside to treat endocarditis due to an enterococcus or a penicillin-nonsusceptible viridans streptococcus), to treat polymicrobial infections (e.g., vancomycin plus a carbapenem to treat infection due to both MRSA and extended-spectrum β-lactamase–producing Enterobacteriaceae), and to minimize resistance to rifampin for infections due to prosthetic devices such as heart valves and joint prostheses.

3. Which of the following statements about antimicrobial agents is true?

A. Regardless of the antimicrobial used, administration by the intravenous route is always more effective when doses are given at least every 4 hours.
B. Daptomycin is effective for treatment of community-acquired pneumonia if it is administered once daily.
C. In this day and age, there is no role for intramuscular administration of antimicrobial agents.
D. It is usually wise to measure serum bactericidal titers in treating endocarditis.
E. None of the above

Answer: E Answer A might be relevant to agents such as β-lactams, the effectiveness of which depends on "time above the MIC," but does not apply to agents such as fluoroquinolones or aminoglycosides. Answer B is incorrect because daptomycin is not indicated for treatment of bronchopneumonia, as it is inactivated by pulmonary surfactant. Answer C is also not correct. Benzathine penicillin and ceftriaxone are used intramuscularly for treatment of syphilis and gonorrhea, respectively, in the outpatient setting. Answer D cannot be justified either. Although studies have suggested that higher serum bactericidal titers correlate with cure in endocarditis, the correlation is imperfect, the results are not always reproducible, and the procedures are labor-intensive. For these reasons, this test is rarely used in clinical practice.

280 APPROACH TO FEVER OR SUSPECTED INFECTION IN THE NORMAL HOST

JAMES E. LEGGETT

We are constantly exposed to microorganisms through our skin or mucous membranes. Most microorganisms are adapted to niches in the environment that make them avirulent to humans (Chapter 279). Pathogens in a normal host are relatively few, and most of the time, exposure results in only transient or stable colonization. Infection can be defined as invasion of a pathogen that triggers an immune response, whether the infection is asymptomatic or symptomatic. Manifestations of infection are protean and are due as much to our immune response as to the attributes of the particular pathogen.

The inflammatory response that accompanies infection is usually marked by fever. Fever is a tightly controlled elevation in body temperature above the normal range in response to a central nervous system change in the set point. Defining normal body temperature is somewhat problematic because it is dependent on both physiology and the method of measurement. Normal oral temperature in 99% of the population ranges from 36.0° to 37.7° C, with a circadian variation of 1° C or more between the morning nadir and the evening peak. Mean oral temperature in healthy adults is 36.8° ± 0.4° C, with women exhibiting slightly higher values than men (36.9° vs. 36.7° C). In menstruating women, the morning temperature may rise by 0.6° C with ovulation and remain higher until menses occur. Measured rectal temperatures are 0.4° C higher than oral and 0.8° C higher than aural (tympanic membrane) temperatures. However, considerable individual variability exists. Clinicians generally define significant fever as a temperature higher than 38.3° C (101.0° F). Despite historical claims, fever patterns are not especially helpful in establishing a specific diagnosis.

The majority of acute febrile illnesses lasting less than 2 weeks have an infectious cause. These infections occur predominantly where body surfaces interact with the environment, such as the upper and lower respiratory tracts, gastrointestinal and genitourinary systems, and skin. The majority of acute respiratory and gastrointestinal infections are viral in nature. As the duration of the febrile illness lengthens beyond 3 weeks, other inflammatory illnesses become more prominent in the differential diagnosis. Most chronic febrile illnesses are not caused by infection.

PATHOBIOLOGY OF INFECTION AND FEVER

Infection ensues only when a pathogen overcomes both nonspecific innate and specific adaptive humoral and cellular immune responses. The normal indigenous microflora, host physical barriers (e.g., skin, mucous membranes, cilia), and soluble factors (e.g., cytokines, complement) provide important barricades to pathogen invasion. Disruption of these barriers, which provide a first line of defense, permits the invasion of pathogens. The acute phase response triggered by such disruption provides direct antimicrobial activity and prompts the development of adaptive immunity mediated by lymphocytes and macrophages. This inflammatory response plays an important role in containing infection. Unfortunately, an exaggerated response may worsen the clinical condition. It is the neutrophil response that causes the damage seen in septic arthritis, and it is the unchecked immune response that precipitates the systemic inflammatory response syndrome.

Body temperature is regulated both physiologically and behaviorally. Basal metabolic processes, governed especially by thyroid hormones but also by catecholamines and growth hormone, are responsible for the normal resting body temperature. Thermogenesis may be increased up to 80% by hyperthyroidism and decreased as much as 50% by hypothyroidism. Moderate activity increases thermogenesis and results in a transiently increased temperature until heat-dissipating processes are engaged. Each 1° F increase in temperature results in a 7% increase in the basal metabolic rate. Vaporization from the lungs and skin accounts for a third of basal body heat loss and for as much as all heat loss at ambient dry temperatures above 36° C. The elderly have a decrease in basal metabolism as well as blunted responses to thermogenetic stimuli, but they have the same average core temperature as young people.

The hypothalamus contains temperature-sensitive neurons that have receptors for pro-inflammatory and anti-inflammatory cytokines, which are continuously balanced to maintain a homeothermic set point. When body temperature becomes elevated, cutaneous vasodilation and sweating occur, and people may reduce activity and seek a cooler environment. In contrast, low body temperature is increased by shivering, piloerection, cutaneous vasoconstriction, adding clothes, and seeking a warmer environment. In a febrile illness, symptoms may be due to the underlying disease or to the fever itself. Malaise is the rule, and many febrile patients experience myalgia secondary to the muscle contractions used to generate temperature elevation. Although it was once thought that the back and thigh pain related to rigors suggests bacteremia, any febrile stimulus can produce such symptoms. The chill associated with rigors may be related to the surface vasoconstriction that accompanies the increase in core temperature.

Fever is a complex physiologic process involving metabolic and immunologic responses (Fig. 280-1). Exogenous pyrogens cause fever largely mediated by endogenous pro-inflammatory pyrogenic cytokines produced by phagocytic leukocytes, including interleukin-1, interleukin-6, tumor necrosis factor-α, and interferon-γ. These cytokines stimulate the immune responses of T and B cells, macrophages, and polymorphonuclear leukocytes. They appear to act through a common mechanism involving the activation of Toll-like receptors and induction of prostaglandin synthesis. Feedback inhibitory responses are mediated by adrenocorticotropic hormone, arginine vasopressin, serotonin, dopamine, and other homeostatic mechanisms, thus emphasizing the orchestrated nature of fever production and response to infection. These thermoregulatory mechanisms rarely allow fevers to exceed 41° C (106° F). Temperatures exceeding 41° C are often due to a drug-induced imbalance in these mechanisms and may cause direct cellular damage.

Failure of fever to develop during severe bacterial infection has, in some studies, been associated with higher morbidity and mortality. Whether this is due to the absence of fever or to associated conditions, such as chronic renal failure or corticosteroid use, has not been determined. Favorable effects of fever on host-microbe interactions are suggested by inhibited multiplication of some pathogens (such as *Streptococcus pneumoniae* and *Treponema pallidum*), reduced proliferation of pathogens in the presence of

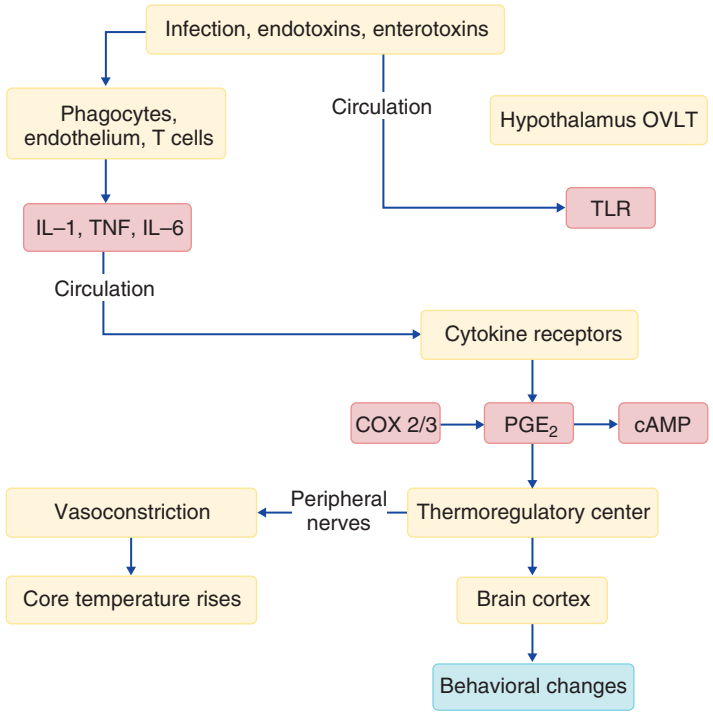

FIGURE 280-1. Pathways leading to the production of fever in bacterial infection, whether local or systemic. Bacteria release cell wall products such as peptidoglycans and endotoxin as well as enterotoxins, which bind to Toll-like receptors (TLRs) on phagocytes (neutrophils, macrophages) and endothelial cells. As a result, pyrogenic cytokines such as interleukin (IL)–1, IL-6, and tumor necrosis factor (TNF)–α are released into the circulation and bind to cytokine receptors in the hypothalamic organum vasculosum of lamina terminalis (OVLT). Bacterial products may also directly bind to TLRs on the OVLT. Activation of TLRs and cytokine receptors induces cyclooxygenase 2 (COX2), which leads to the production of prostaglandin E_2 (PGE_2) and elevates cyclic adenosine monophosphate (cAMP) in the brain. This, in turn, triggers neurons in the thermoregulatory center to raise the hypothalamic thermostatic set point. In addition, neuronal signals to the cortex prompt behavioral changes to conserve heat (e.g., posturing, adding clothing). The hypothalamus also triggers sympathetic peripheral efferent nerves that constrict peripheral blood vessels and conserve central heat until hypothalamic PGE_2 levels fall.

hypoferremia, augmented complement-mediated lysis, and increased neutrophil entry into inflammatory sites. Indeed, vertebrate endothermy restricts most fungi as potential pathogens in hosts with intact immune systems.

Temperature-pulse dissociation, in which there is relative bradycardia compared with the usual increase of 2.44 beats/minute per 1°F, has been described in typhoid fever, leptospirosis, rickettsiosis, dengue, legionellosis, and babesiosis, for unclear reasons.

No laboratory abnormalities accompany typical benign acute viral infections. Leukocytosis (Chapter 167) of various lineages is usually seen in other infections in immunocompetent adults. Neutrophilia is the norm in most acute infections, whatever the cause. The elderly, although not mounting a neutrophilic response, generally display bandemia during an acute bacterial infection. Neutropenia (Chapter 167) may be seen in rickettsial, severe viral, and overwhelming bacterial infections. Eosinophilia (Chapter 170) is typical of invasive helminthic and some protozoal infections. Lymphocytosis may accompany many viral and rickettsial infections and is common during convalescence from acute bacterial infection. Monocytosis may be seen in tuberculosis. Virtually all infections have an impact on the erythroid system, but given the long half-life of erythrocytes, usually only chronic infections or other inflammatory diseases result in anemia (Chapter 158). Few acute infections rapidly produce anemia. For instance, *Helicobacter pylori* may induce a bleeding ulcer, *Plasmodium falciparum* may directly lyse erythrocytes, overwhelming clostridial and other bacterial infections associated with disseminated intravascular coagulopathy may cause hemolytic anemia, and *Mycoplasma pneumoniae* may induce immunologically mediated hemolysis.

APPROACH TO FEBRILE ILLNESS IN OUTPATIENTS AND INPATIENTS

Infectious disease epidemiology depends on the interaction among pathogens, susceptible hosts, and environmental conditions allowing exposure. Most infections are transmitted horizontally between people by contact (e.g., hands, fomites), a common vehicle (e.g., food, water), air (e.g., tuberculosis), or vectors (e.g., mosquitoes). Evaluation of a patient with a known or possible infection should determine whether the condition might be due to a transmissible agent and its source, whether the patient has done any recent traveling, whether there are secondary causes, and what measures need to be taken to contact health department officials and to prevent additional infections.

The age of a patient influences which illnesses should be considered. Natural exposure or immunization generally limits certain illnesses, such as rubeola, rubella, and varicella. Waning of immunity may likewise lead to pertussis in young adults or reactivation of tuberculosis in the elderly. Other physiologic effects of aging, such as impaired bladder emptying, lead to increased rates of urinary tract infection in the elderly.

A patient's occupation and travel history should be noted. An abattoir worker is more likely to have been exposed to *Brucella* (Chapter 310) than is someone with another occupation. Indiana residents are more likely be infected with histoplasmosis (Chapter 332), whereas those from the Southwest desert may have coccidioidomycosis (Chapter 333), despite having a similar febrile illness. Many other illnesses are likewise directly related to specific geographic exposure, with varying incubation times before their onset (Chapter 286). Typhoid fever should be manifested within a few weeks, whereas amebic liver abscess might not cause symptoms until months after a traveler's return from an endemic area (Table 280-1). The Centers for Disease Control and Prevention website (http://www.cdc.gov) and many others provide more specific information about prevalent infections in all parts of the world (Chapter 286).

Many travelers return home with fever after a variable incubation time, generally with other symptoms and signs as well. The first consideration in evaluating such a patient is that an infection unrelated to travel is more likely to be the cause of the illness.[1] Once such routine infections have been ruled out, the differential diagnosis should include infections related to travel,[2] whether within a region of the United States (e.g., ehrlichiosis, Colorado tick fever, hantavirus) or abroad (e.g., visceral leishmaniasis, tick-borne encephalitis in Europe) (Chapter 286). For example, prompt evaluation of a patient who has traveled to a malaria-endemic area should be undertaken and blood tests performed to determine the presence of parasites.

The setting in which a febrile illness occurs influences both the diagnostic approach and the differential diagnosis. In the ambulatory arena, with a generally healthy febrile patient, the clinician should not necessarily pursue a diagnosis as aggressively as with a hospitalized or chronically ill patient. Empirical treatment of a presumed urinary tract infection is warranted in the outpatient setting, where the cost of a culture is often more than that of the antibiotic. However, the cost of a urine culture in a hospital setting is minimal in comparison to the daily cost of care, and accurate identification of the pathogen may speed hospital discharge. Likewise, the pathogens commonly causing febrile illness in health care facilities, including nursing homes, may differ from those seen in ambulatory settings. Most patients in the ambulatory setting have noncritical, self-limited infections.

TABLE 280-1 SELECTED EXAMPLES OF FEVER ASSOCIATED WITH RECENT TRAVEL

	INCUBATION PERIOD	
DISEASE	**<2-3 WEEKS**	**>3-4 WEEKS**
COMMON		
Dengue	+	
Dysentery	+	
Entamebic liver abscess		+
Enteric fever	+	
Malaria	+	+
Pulmonary tuberculosis		+
Viral hepatitis		+
LESS COMMON		
Ehrlichiosis	+	
Leptospirosis	+	
Schistosomiasis		+
Viral (hemorrhagic, encephalitic)	+	
Visceral leishmaniasis		+

Higher temperatures are usually due to invasive visceral disease, such as community-acquired pneumonia or pyelonephritis. Common viral respiratory infections and gastroenteritis as well as some cases of subacute bacterial endocarditis are accompanied by temperatures below 102°F. Moreover, many infections may not be associated with fever, such as Lyme disease, osteomyelitis, and most sexually transmitted diseases. The clinician must always keep in mind that certain infections, such as sexually transmitted diseases (Chapter 285) or herpes zoster, occur normally in immunocompetent hosts and may signal a higher risk for infection with human immunodeficiency virus (HIV) or an already established immunodeficiency.

Fever in Outpatients

In the ambulatory setting, an acutely febrile patient represents a common problem and only infrequently presents an enigmatic diagnostic challenge. In most instances, a febrile illness is accompanied by localizing symptoms and signs suggesting a specific diagnosis. For instance, leg erythema, pain, and fever in a patient with tinea pedis or a saphenous vein graft incision immediately suggests streptococcal cellulitis. Several diseases may masquerade as infectious cellulitis.[3] If the patient has had a gradual onset and does not appear toxic, only clinical observation and follow-up are required. If the patient appears toxic, with tachypnea and apprehension or confusion accompanying localized findings, clinically focused diagnostic studies should be performed immediately, and hospitalization should be considered. When a patient has fever and only nonspecific constitutional symptoms, it may be more difficult to address the problem in a single ambulatory clinic visit, requiring a balance between observation and investigation.

Fever in Inpatients

Fever and leukocytosis are probably the main clinical parameters for evaluating potential infections in hospitalized patients. However, about 10% of nosocomial bacteremias occur without fever, and health care–associated infections occur without fever in a substantial proportion of patients who are elderly or have significant comorbid conditions. Most cases of hospital-associated fever represent nosocomial infection, which typically involves the lower respiratory tract, urinary tract, or surgical wounds (Table 280-2). Some important causes of nosocomial fever may not exhibit easily discernible localizing symptoms or signs. Antibiotic-induced colitis secondary to *Clostridium difficile* (Chapter 296) is increasing in prevalence and may be characterized by little or no diarrhea. It is probably the most common cause of a leukemoid reaction in hospitalized patients. Other intra-abdominal processes involving the hepatobiliary system, bowel infarction, viscus perforation, or abscesses may have little in the way of localizing symptoms or signs.

Given the greater severity of illness and comorbid conditions in intensive care unit (ICU) patients, it is logical that fever and infection are more frequent in the ICU than elsewhere. Infection was recently found to be present in more than 80% of febrile ICU patients, although both infectious and noninfectious

TABLE 280-2 SELECTED CAUSES OF HOSPITAL-ASSOCIATED FEVER

COMMON	LESS COMMON
INFECTIOUS	**INFECTIOUS**
Clostridium difficile enterocolitis	Biliary tract disease
Pneumonia	Endometritis
Surgical wound	Intra-abdominal abscess
Urinary tract	Mediastinitis
Vascular catheter	Sinusitis
NONINFECTIOUS	**NONINFECTIOUS**
Drug-induced fever	Adrenal insufficiency
Hematoma	Gout
Immediate postoperative state	Myocardial infarction
Transfusion reaction	Organ infarction
Venous thromboembolism	Pancreatitis

causes of fever may coexist. Indeed, ischemia or devitalization of tissue provokes an inflammatory response similar to that prompted by infection. About half of patients with acute myocardial infarction have a temperature between 38.0° and 38.5° C within 2 to 3 days of their infarction. Similarly, about half of patients with deep venous thrombosis and pulmonary embolism have a temperature in the same range, most commonly in the first 3 days after diagnosis. A third or more of patients with stroke demonstrate fever, which is also a common consequence of subarachnoid or intracerebral hemorrhage and subdural hematoma, especially within 72 hours of onset.[4] The fever in such cases may result from damage to the hypothalamus or pulmonary aspiration secondary to obtundation. Iatrogenic causes of fever should be considered. Fever and chills may be seen in up to a quarter of patients receiving platelet transfusions, although the frequency is much less with other blood products.

SYNDROMIC APPROACH

Fever and Rash

A syndromic approach is valuable to narrow the many possible causes of a suspected infection. Two approaches must be juxtaposed in this evaluation, and both are key in recognizing patterns. The clinician must be aware of (1) the differential diagnosis of the particular type of lesion observed and (2) the constellation of findings produced by individual pathogens. Because of the variety of possible manifestations and the often overlapping symptoms and signs, both elements are key in arriving at a probable diagnosis. Moreover, fever and associated findings, such as exanthem, lymphadenopathy, or jaundice, may be due to noninfectious systemic diseases as well as infectious ones. For instance, leukocytoclastic vasculitis and fever may be found in meningococcemia, Rocky Mountain spotted fever, and hepatitis C, but they are also seen in noninfectious inflammatory diseases. Likewise, fever and adenopathy may be due to lymphoma or to cat-scratch disease.

A recognizable exanthem may lead to the immediate recognition of a particular pathogen (Chapter 441), but often a larger differential diagnosis must be entertained. The clinician must recognize the type or types of skin lesions present, the distribution of the exanthem, and the chronologic progression with respect to the onset of fever and other symptoms (Table 280-3). Morphologic variations in skin lesions help in the differential diagnosis. Maculopapular exanthems are frequently seen in viral illness, hypersensitivity drug reactions, and immune complex–mediated diseases. Some of the most common viral causes include the many enteroviruses, but similar lesions may also be seen with hepatitis B and West Nile viruses. Erythema multiforme, a subset of maculopapular exanthem, can result from various viral infections or drug eruptions (Chapter 440). It may have a spectrum of disease that ranges from benign to the life-threatening Stevens-Johnson syndrome/toxic epidermal necrolysis complex. Herpes simplex virus is perhaps the most common cause of erythema multiforme. Although drugs, especially antibiotics, are the major precipitating factor in the Stevens-Johnson syndrome/toxic epidermal necrolysis complex, *M. pneumoniae* has been associated with it as well. Evolution of the cutaneous findings over time may give clues to the cause; for example, the initial blanching, erythematous, maculopapular lesions may later evolve into petechiae, as seen in meningococcemia, Rocky Mountain spotted fever, and dengue. Secondary syphilis may be manifested with a multitude of morphologic skin lesions. Sometimes, many different manifestations occur simultaneously in the same patient. Most vesiculobullous skin exanthems are immunologically mediated. The few infections associated with these eruptions include herpes simplex and varicella-zoster viruses and enteroviruses such as echovirus and coxsackievirus. The poxviruses, which

TABLE 280-3 SELECTED INFECTIONS WITH FEVER AND RASH

ETIOLOGY	MACULES, PAPULES	VESICLES, BULLAE	PETECHIAE, PURPURA
BACTERIA			
Borrelia burgdorferi	+ (annular)		
Neisseria meningitidis			+
Rickettsia rickettsii	+		+
Treponema pallidum	+ (secondary)		
Vibrio vulnificus		+	
FUNGI AND MYCOBACTERIA			
Disseminated disease	+ (nodular)		
PROTOZOA			
Plasmodium falciparum			+
VIRUSES			
Enteroviruses	+	+	+
Epstein-Barr	+		+
Hemorrhagic fever			+
Herpes		+	
HIV	+		

TABLE 280-4 FEVER AND RASH INVOLVING THE PALMS AND SOLES

Erythema multiforme
Hand-foot-and-mouth disease
Neisseria infection
Rocky Mountain spotted fever
Streptobacillus moniliformis infection
Subacute bacterial endocarditis
Syphilis (secondary)
Toxic shock syndrome
Varicella-zoster infection

can also cause such exanthems, are much rarer or are associated with bioterrorism. Pustules, or vesicles containing leukocytes, are usually associated with psoriasis or infections with *Pseudomonas*, *Staphylococcus*, or *Neisseria*. Bullous exanthems in the presence of sepsis suggest severe streptococcal cellulitis or necrotizing fasciitis, staphylococcal impetigo, or *Vibrio* infections.

Petechial and purpuric eruptions are due to the extravasation of red blood cells and should always lead to consideration of a potentially serious illness. Pathogens creating such lesions most commonly include *Neisseria meningitidis*, *Rickettsia*, and *Capnocytophaga canimorsus*, but these eruptions may be seen with a variety of other pathogens, including *Staphylococcus aureus*, group B streptococci, and other gram-negative bacilli. A petechial exanthem may also be seen with enteroviruses and viral hemorrhagic fevers. The most common causes of petechiae not attributable to infections include thrombocytopenia and vasculitis.

The presence of fever and rash involving the palms and soles allows considerable narrowing of the differential diagnosis (Table 280-4). In addition to the diffuse erythema associated with toxic shock syndrome, illnesses such as Rocky Mountain spotted fever, secondary syphilis, hand-foot-and-mouth disease, *Neisseria* infections, and rat-bite fever should be considered in patients with maculopapular exanthems involving these areas.

Nodular skin lesions may be either noninfectious, as seen in malignant disease or with certain drugs (e.g., sulfonamides), or infectious, as seen in a variety of inflammatory diseases. Atypical mycobacteria and disseminated fungi often produce skin nodules. The tender nodules of erythema nodosum usually occur in crops located pretibially, but they may be solitary or occur on other parts of the body. They do not typically suppurate, and they heal without scarring. Infectious agents are the most likely cause of erythema nodosum. Diffuse erythema may be seen with scarlet fever, toxic shock syndrome, Kawasaki disease, Stevens-Johnson syndrome, and toxic epidermal necrolysis, with desquamation occurring late in all these syndromes. Sweet's syndrome, a febrile neutrophilic dermatosis, represents a hypersensitivity reaction often preceded by an upper respiratory tract infection.

Fever and Musculoskeletal Complaints

Fever and localized tenderness, swelling, or erythema generally accompany septic arthritis and often accompany osteomyelitis (Chapter 272). Septic

bacterial arthritis in adults usually is manifested acutely and involves a single large joint such as the knee, hip, or shoulder, unless the infection is directly inoculated by trauma or surgery. Septic oligoarthritis may be seen with endocarditis and rat-bite fever. Disseminated gonococcal disease is the usual cause of arthritis involving small joints of the wrist, ankle, and digits, often with tenosynovitis. Acute or subacute polyarthritis may be seen in several viral diseases, including parvovirus B19 and hepatitis B, and in Lyme disease, but it is more typical of immunologic disorders. Rheumatologic diseases generally have more subacute manifestations, with more symmetrical polyarthritis. Hematogenous osteomyelitis in adults frequently involves the vertebrae and is almost always initiated by discitis with symmetrical involvement of adjacent vertebrae (as opposed to malignant metastasis, which is asymmetrical and does not involve the disc).

Myositis secondary to clostridia, streptococci, *Aeromonas*, or mixed aerobic-anaerobic infections usually causes an acutely septic picture with painful, edematous involvement of the limb or torso. Pyomyositis frequently involves deep muscles such as the psoas or gluteus and is usually due to *S. aureus*. Diffuse myositis may be seen with leptospirosis or toxoplasmosis, and rhabdomyolysis occurs with a variety of viral infections and legionellosis.

Fever and Lymphadenopathy or Hepatosplenomegaly

Fever and lymphadenopathy may suggest a variety of illnesses, both infectious and noninfectious (Table 280-5). Lymphadenopathy (Chapter 168) may be regional or generalized. Local enlargement can occur with either a local infection or some systemic illnesses (e.g., posterior cervical lymphadenopathy with Epstein-Barr virus and other viral illnesses). Generalized lymphadenopathy usually suggests a systemic disorder, which may itself be either infectious or noninfectious. Although the combination of fever and lymphadenopathy secondary to infection is especially common during childhood, it is also frequently seen in adults. As in other syndromes, acute versus chronic adenopathy tilts the diagnosis toward different broad categories of illness. In chronic adenopathy, histopathologic evaluation of enlarged lymph nodes may point to a particular diagnosis. For instance, toxoplasmosis or cat-scratch disease can be easily differentiated from mycobacterial disease or sarcoidosis.

Fever and hepatosplenomegaly (Chapter 168) may provide an important clue to the cause of a febrile illness, which is typically either an infection or a malignant neoplasm arising from bone marrow or the reticuloendothelial system. Jaundice may also limit the differential diagnosis (Table 280-6).

TABLE 280-5 COMMON CAUSES OF FEVER AND LYMPHADENOPATHY

REGIONAL	GENERALIZED
Cervical	Cytomegalovirus
Streptococci	Epstein-Barr virus
Tuberculosis	HIV
Viral upper respiratory tract infection	Lymphoma
Peripheral	Sarcoidosis
Bartonella henselae	Syphilis (secondary)
Herpesviruses	Toxoplasmosis
Lymphoma	Viral hepatitis
Metastatic cancer	
Sporotrichosis	
Streptococci	
Inguinal	
Chancroid	
Herpes	
Lymphogranuloma venereum	
Syphilis (primary)	

Aside from the viral hepatitides and other diseases affecting primarily the liver, many pathogens producing sepsis can cause hyperbilirubinemia.

APPROACH TO FEVER OF UNKNOWN ORIGIN

The majority of febrile illnesses are short-lived, but fever may be prolonged for weeks or months as part of an infectious disease, inflammatory disorder, or occult neoplasm. When fever is caused by infection, the site is an area not easily controlled by host defenses, leading to the continued release of inflammatory cytokines. Likewise, macrophage and lymphocyte involvement in inflammatory disorders causes persistent cytokine production, as do certain neoplasms. Given this final common pathway, it is easy to understand why the majority of cases of classic fever of unknown origin (FUO), loosely defined as lasting longer than 3 weeks despite routine investigation, are found in these three broad categories.[5,6] Two other categories round out the bulk of FUO cases: miscellaneous illnesses and cases in which the fever remains undiagnosed (Table 280-7). The proportion of patients in each category varies by geographic locale, age, duration of fever, and immune status. As more effective methods of diagnosing viral and bacterial infections have become available, and with improved serologic studies to detect connective tissue disorders and better imaging techniques to detect occult malignant neoplasms, the proportion of patients with FUO in the miscellaneous and undiagnosed categories has increased to about a third of the total in developed countries. The longer a febrile illness persists without a diagnosis or appropriate therapy, the less likely it is to be due to an infection. In one study, only 6% of patients who had persistent FUO beyond 6 months were found to have infection. Most such cases resolve spontaneously, with a mortality rate of less than 3% at 5 years.

Bacterial species, particularly *Mycobacterium tuberculosis*, make up the largest category of infections that cause prolonged FUO. *M. tuberculosis* and other bacterial pathogens causing FUO have adapted to survive intracellularly or frequently change their surface antigens so that they are not readily eradicated by host defenses. Other infections causing FUO are localized in cryptic abscesses, especially intra-abdominally, or reside on heart valves, where the inflammatory response is blunted. Persistent viral infections constitute a small and shrinking subset of patients with FUO because modern techniques can more readily detect infection with Epstein-Barr virus, cytomegalovirus, and others. Among the pathogens likely to be characterized initially by fever alone, cytomegalovirus is the most common cause of mononucleosis in adults, and malaria is a common cause of fever in returning travelers.

Malignant disease may result in persistent fever due to the production of inflammatory cytokines, necrosis, or the presence of a complicating infection. The most common malignant neoplasms manifesting as FUO are lymphomas, leukemias, and solid tumors with metastases to the liver. Connective tissue disorders may lead to tissue inflammation, which produces fever as a prominent feature of the illness. In a recent series, adult Still's disease (Chapter 261) was the leading rheumatologic disorder manifesting as FUO. Temporal arteritis and polymyalgia rheumatica (Chapter 271) are seen almost exclusively in patients older than 50 years. Systemic lupus erythema-

TABLE 280-6 COMMON INFECTIOUS CAUSES OF FEVER AND JAUNDICE

Bacterial sepsis
Cholangitis
Hepatic abscess
Leptospirosis
Malaria
Viral hepatitis
Yellow fever

TABLE 280-7 FREQUENCY OF SELECTED CHRONIC FEBRILE ILLNESSES

INFECTION, 25-50%	MALIGNANT DISEASE, 20-30%	CONNECTIVE TISSUE DISEASE, 15-30%	MISCELLANEOUS, 10-20%	UNDIAGNOSED, 10-30%
Cytomegalovirus	Carcinomatosis	Polyarteritis nodosa	Drug-induced fever	
Endocarditis	Leukemia	Rheumatoid arthritis	Granulomatous hepatitis	
Intra-abdominal	Local tumor	Still's disease	Inflammatory bowel disease	
Mycoses	Lymphoma	Systemic lupus erythematosus	Pancreatitis	
Occult abscess		Temporal arteritis	Pulmonary embolism	
Tuberculosis				

tosus (Chapter 266) is an occasional cause of FUO, especially if it is manifested in an atypical fashion.

The miscellaneous category of FUO includes several disparate groups of diseases. Granulomatous diseases such as granulomatous hepatitis, Crohn's disease, or sarcoidosis may incite cellular immune responses that result in fever. Granulomatous hepatitis was present in up to 6% of National Institutes of Health cases with fever lasting longer than 6 months. Chronic pancreatitis may occasionally cause FUO, as may recurrent pulmonary embolism.

Drug-induced fever (Table 280-8) may be the only manifestation of an adverse drug event in up to 5% of cases of drug hypersensitivity. Recognition of drug-induced fever is important to avoid extra tests, additional therapy, and prolonged hospitalization.[7] The mechanisms by which drugs incite fever are not well understood in many cases. These events may result from hypersensitivity reactions, altered thermoregulatory homeostasis directly related to either drug administration or the drug's pharmacologic action, or idiosyncratic reactions. Hypersensitivity reactions are usually accompanied by an exanthem or enanthem and hepatic, renal, or pulmonary dysfunction in addition to fever. Antimicrobial agents appear to be the most common cause of drug-induced fever and are responsible for approximately a third of episodes in some studies. β-Lactams and sulfonamides account for most cases because they are among the most frequently administered antimicrobials. Anticonvulsants are also common causes of drug-induced fever secondary to hypersensitivity reactions. Altered thermoregulation is possible with a variety of drugs, including those with anticholinergic activity, such as phenothiazines and tricyclic antidepressants. Sympathomimetic agents, such as amphetamines and cocaine, may also cause fever. Drug administration itself may cause fever if the vehicle of the drug is contaminated with exogenous pyrogens or chemical phlebitis occurs. Some drugs appear to have intrinsic pyrogenic properties, such as amphotericin B and bleomycin. Others cause fever as a result of their pharmacologic activity, such as interferon alfa or interleukin-2. With antibiotics, drug-induced fever occurs with the rapid lysis of spirochetes or other bacteria, known as the Jarisch-Herxheimer reaction. Idiosyncratic drug-induced febrile reactions include malignant hyperthermia, neuroleptic malignant syndrome, and serotonin syndrome (Chapter 434). Drugs implicated in these reactions are inhaled anesthetic agents, central nervous system dopamine-depleting agents, and serotonin re-uptake inhibitors, among others. Drug-induced fever is usually a diagnosis of exclusion. The duration of drug exposure before the onset of fever, the clinical appearance of the patient, and the pattern of the fever are not particularly useful. Elimination of a single drug at a time, beginning with the one most likely to be implicated, is the usual means of identifying the causative agent. The fever abates once the drug has been eliminated from the body, usually within 3 to 4 days of discontinuing use of the drug.

Laboratory evaluation and diagnostic imaging studies should be chosen according to information derived from a detailed history and physical examination. These may initially include complete blood count and blood chemistry determinations, erythrocyte sedimentation rate or C-reactive protein, blood cultures, and antibody tests (antinuclear antibody, cytomegalovirus, Epstein-Barr virus, HIV) as well as a chest radiograph and computed tomography of the abdomen.[8] 18-Fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) may also be useful in difficult cases.[9] Despite the recent focus on "emerging" infectious diseases, the cause of FUO is still more likely to be a common pathogen presenting atypically.

TABLE 280-8 SELECTED AGENTS ASSOCIATED WITH DRUG-INDUCED FEVER

COMMON	LESS COMMON
ANTIMICROBIAL	
Amphotericin B	Clindamycin
β-Lactams	Fluoroquinolones
Sulfonamides	Rifampin
CARDIOVASCULAR	
Procainamide	Diltiazem
Quinidine	Hydralazine
CENTRAL NERVOUS SYSTEM	
Carbamazepine	Haloperidol
Phenytoin	Serotonin re-uptake inhibitors
MISCELLANEOUS	
Bleomycin	Allopurinol
Interferon alfa	Cimetidine
Interleukin-2	Tacrolimus

INITIAL MANAGEMENT OF SUSPECTED INFECTION IN THE AMBULATORY SETTING

An acutely febrile patient in the ambulatory setting presents a common but often demanding diagnostic problem. In most cases, the history and physical examination reveal diagnostic clues and may guide decisions about additional studies or therapy. More difficult to diagnose is a fever that occurs without localizing symptoms or is accompanied only by nonspecific symptoms, such as malaise or anorexia. Fortunately, most such acute, undifferentiated febrile illnesses are benign and resolve spontaneously within 1 or 2 weeks without a specific diagnosis being made. In such cases, no further evaluation beyond the initial visit is warranted. If symptoms persist, the history and physical examination should be repeated, looking for previously unsought clues and new physical findings. Laboratory studies might be required.

In patients with an illness involving cough of less than 3 weeks' duration, the evaluation should focus on ruling out a serious disorder. Normal vital signs and normal findings on a chest examination effectively rule out most cases of pneumonia. Such cough illnesses are caused by viral pathogens in more than 90% of cases. Antibiotics are ineffective in such patients, and antimicrobial therapy does not prevent bacterial complications such as pneumonia. The presence of sputum and its characteristics are not helpful in distinguishing bacterial from viral infections. Adults with prolonged coughing lasting longer than 3 weeks or with recurrent episodes should be evaluated for reactive airway disease, gastroesophageal reflux, and other illnesses. Infections rarely causing prolonged cough include *Bordetella pertussis*, *M. pneumoniae*, and *Chlamydophila pneumoniae*. Clinicians in this case should obtain a chest radiograph; treat for exacerbation of chronic obstructive pulmonary disease (fever, leukocytosis, purulent sputum), if present; treat a confirmed bacterial infection; and direct therapy to a specific underlying cause or other causes.

Symptoms and signs of pharyngitis include fever, tonsillar exudates, tender anterior cervical lymph nodes, and absence of cough. If fewer than two of these criteria are present, the patient should be managed as though viral pharyngitis were the cause. With two or more of these criteria, one should consider obtaining a rapid streptococcal antigen test.[10] Because of the low incidence of streptococcal infection and acute rheumatic fever in adults, a negative rapid test result alone is sufficient to rule out infection with *Streptococcus pyogenes*. If the antigen test result is positive, the patient can be managed with a β-lactam antibiotic if not allergic. Ninety percent of cases of pharyngitis in adults are viral in origin. In a patient with symptoms of upper respiratory tract infection and a mucopurulent nasal discharge of less than 10 days' duration, purulent nasal secretions do not predict bacterial infection. Most cases of acute rhinosinusitis seen in the outpatient setting are caused by uncomplicated upper respiratory viral infection. If symptoms have been present for more than 10 days without improvement, or if there are specific symptoms of sinusitis of any duration (purulent nasal discharge lasting 3 to 4 days, unilateral facial pain and pressure, maxillary toothache, or worsening of symptoms after initial improvement), amoxicillin or another β-lactam should be considered, with other antimicrobial classes used in penicillin-allergic patients. Most clinical outcomes are not adversely affected by delayed antibiotics for upper respiratory infections.

Community-acquired pneumonia (Chapter 97) should be suspected in a patient with cough, sputum production, or dyspnea, especially if it is accompanied by fever and altered breath sounds. A chest radiograph should be performed to confirm the diagnosis. Determining where to care for the patient is the most important immediate decision. Outpatient care generally suffices for patients younger than 50 years with no cardiopulmonary disease; for patients with no comorbid conditions (including malignant disease, heart failure, diabetes, or hospitalization within the last year); and for patients with no physical examination findings, such as altered mental status, pulse of 125 beats/minute or greater, or respiratory rate of 30/minute or greater. Recent guidelines developed by the American Thoracic Society and the Infectious Diseases Society of America suggest a β-lactam, macrolide, or doxycycline. Fluoroquinolones should be used for outpatients only when the patient has failed to respond to first-line therapy, has a significant comorbidity, or has a known allergy to a first-line agent.

Skin and soft tissue infections are caused, for the most part, by streptococci; a minority are due to *S. aureus* and, rarely, other bacteria whose presence may be suggested by epidemiologic considerations (e.g., swimming in fresh water, where *Aeromonas* may be the pathogen). Pain may be present for

12 hours or more before skin discoloration is noted. A furuncle or abscess formation should prompt consideration of *S. aureus* and, rarely, *Streptococcus anginosus* group. Incision and drainage may be sufficient for a skin abscess, although the rapidly expanding, virulent, community-acquired methicillin-resistant *S. aureus* phenotype may require antimicrobial therapy. Septic bursitis is nearly always due to *S. aureus*, and the infected bursa should be aspirated and drained, in addition to using antibiotics (Chapter 272).

Gastrointestinal infections may be due to ingested toxins, viruses, or, less commonly, bacteria, with or without associated toxin production. The appropriate approach depends on the epidemiologic setting, such as improper food storage, travel abroad, or contact with another ill person (Chapter 283). Symptoms of cystitis in a young, sexually active woman can be treated with empirical antibiotics, but when fever and flank pain are present and the patient is nauseated, consideration of a brief hospital admission or an initial intravenous dose of antibiotics may be necessary (Chapter 284). The possibility of pelvic inflammatory disease should also be entertained.

In the initial evaluation of a patient with a more chronic, persistent fever, a careful history and physical examination provide important diagnostic clues, directing further investigation. The initial goal is to characterize the illness accurately, in addition to eliciting important host and epidemiologic factors. A careful review of systems is necessary to understand the extent of involvement of various organ systems as well as to note previous medical conditions. The examination should be broader than for an acute febrile illness with localizing symptoms and signs. Laboratory tests may also play a more important role in guiding further investigation. Repeated evaluations are the norm rather than the exception in these cases.

Blindly initiating empirical therapy in febrile patients with no imminent risk of serious clinical harm or death should be discouraged because it may impede a timely diagnosis affording definitive care. Procalcitonin, which is a precursor of calcitonin, is an acute phase reactant that is more likely to be elevated with bacterial than with viral infections, and its use may reduce unnecessary antibiotics in some situations, such as patients with respiratory infections. However, procalcitonin distinguishes sepsis from nonseptic systemic inflammation poorly (71% sensitivity, 71% specificity, receiver operating characteristic curve 0.63),[11] and it appears to be less useful in such settings.

INITIAL MANAGEMENT OF SUSPECTED NOSOCOMIAL INFECTION

Determination of the nature of a febrile illness in a hospitalized patient must take into account the host, the setting, and the timing of recent trauma or type and duration of surgery, in addition to the general approach taken for ambulatory patients. A classic mnemonic—the six *w*'s—may help guide the evaluation: wind, water, wound, walk, wonder drug, and what we did. "Wind" refers to fever within the first 24 hours of surgery, when it is unusual to have an infection. A fever at this time is often thought to be related to the anesthetic agent, atelectasis, or surgical trauma. The only bacteria believed to cause significant infections within 24 hours of surgery are *S. pyogenes* and *Clostridium* species, both of which are unusual in the typical hospital patient. "Water" refers to a urinary tract infection occurring after the third day of urinary tract catheterization. Because nearly all nosocomial urinary tract infections occur in patients with indwelling urinary catheters or in those who have undergone urologic instrumentation, urinalysis or culture (or both) should be performed routinely only in febrile patients with such risk factors. There is a high prevalence of bacteriuria in patients who have been catheterized for 3 days or longer, and there is a relatively low incidence of true infection attributable to bacteriuria. "Wound" infections commonly occur about 5 to 7 days postoperatively, whether they are surface wounds or complications of dehiscence of gastrointestinal anastomoses. Some of the highest rates of skin and soft tissue infections in the National Nosocomial Infection Surveillance database are seen with gastrointestinal procedures. Toxin-producing *C. difficile* is the only significant nosocomial gastrointestinal infection seen in hospitalized patients, so a routine bacterial stool culture is not necessary. "Walk" refers to possible deep venous thrombosis or pulmonary embolism in someone who has not received appropriate prophylaxis or who is otherwise at risk for thrombosis. Fever induced by a "wonder drug" is typically seen after approximately 7 to 10 days of use if the patient does not already have an allergy to that medication, in which case it recurs immediately. An exception to this rule is sulfamethoxazole, for which approximately half of hypersensitivity reactions occur within 3 days of initiation. Finally, "what we did" alerts the clinician to the possibility of an iatrogenic infection, such as intravenous catheter–related bacteremia.

CONCLUSION

The initial management of patients with febrile illnesses requires three major considerations. First, is the illness more likely to be infectious or more likely to be related to some other process? Excessive antibiotic use when it is not warranted, such as for viral infections or collagen vascular disease, may cause an adverse reaction, in addition to contributing to the worldwide increase in antimicrobial resistance. However, an empirical antibiotic is appropriate in many cases of fever and localizing signs of bacterial infection. Second, the clinician must rapidly assess the severity of the illness and determine whether it is likely to cause significant organ damage or even death. In a febrile patient with signs of sepsis, the clinician must quickly decide which specific therapy is indicated because a delay in initiating antimicrobial therapy is correlated with increased morbidity and mortality.[12] Finally, the clinician needs to determine whether supportive care alone, including antipyretic therapy, is warranted.

The nearly universal prevalence of febrile adaptive responses to microbial challenge suggests that fever has a net benefit to the host. In addition to clinical studies correlating elevated core temperature and improved prognosis during infection, investigations of principal endogenous mediators have provided evidence of a protective effect of pyrogenic cytokines. Although the use of antipyretic medications is a long-established and widespread practice, the actual benefit of temperature reduction in febrile patients is uncertain. Antipyretic therapy does not protect against the recurrence of childhood febrile seizures, nor has its risk-benefit ratio been determined in patients with cardiopulmonary and other underlying disorders. In summary, fever is usually not harmful, and antipyretics may confuse the clinical picture by dampening it, although their anti-inflammatory effects are often beneficial.

Grade A References

A1. Lemiengre MB, van Driel ML, Merenstein D, et al. Antibiotics for clinically diagnosed acute rhinosinusitis in adults. *Cochrane Database Syst Rev*. 2012;10:CD006089.

A2. Little P, Moore M, Kelly J, et al. Delayed antibiotic prescribing strategies for respiratory tract infections in primary care: pragmatic, factorial, randomised controlled trial. *BMJ*. 2014;348:g1606.

A3. Long W, Li LJ, Huang GZ, et al. Procalcitonin guidance for reduction of antibiotic use in patients hospitalized with severe acute exacerbations of asthma: a randomized controlled study with 12-month follow-up. *Crit Care*. 2014;18:471.

A4. Shehabi Y, Sterba M, Garrett PM, et al. Procalcitonin algorithm in critically ill adults with undifferentiated infection or suspected sepsis. A randomized controlled trial. *Am J Respir Crit Care Med*. 2014;190:1102-1110.

A5. Jefferies S, Weatherall M, Young P, et al. The effect of antipyretic medications on mortality in critically ill patients with infection: a systematic review and meta-analysis. *Crit Care Resusc*. 2011;13:125-131.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Jensenius M, Han PV, Schlagenhauf P, et al. GeoSentinel Surveillance Network. Acute and potentially life-threatening tropical diseases in western travelers—a GeoSentinel multicenter study, 1996-2011. *Am J Trop Med Hyg*. 2013;88:397-404.
2. Gautret P, Cramer JP, Field V, et al. Infectious diseases among travellers and migrants in Europe, EuroTravNet 2010. *Euro Surveill*. 2012;17.
3. Keller EC, Tomecki KJ, Alraisea MC. Distinguishing cellulitis from its mimics. *Cleve Clin J Med*. 2012;79:547-552.
4. Rincon F, Lyden P, Mayer SA. Relationship between temperature, hematoma growth, and functional outcome after intracerebral hemorrhage. *Neurocrit Care*. 2013;18:45-53.
5. Yamanouchi M, Uehara Y, Yokokawa H, et al. Analysis of 256 cases of classic fever of unknown origin. *Intern Med*. 2014;53:2471-2475.
6. Robine A, Hot A, Maucort-Boulch D, et al. Fever of unknown origin in the 2000s: evaluation of 103 cases over eleven years. *Presse Med*. 2014;43:e233-e240.
7. Patel RA, Gallagher JC. Drug fever. *Pharmacotherapy*. 2010;30:57-69.
8. Baron EJ, Miller JM, Weinstein MP, et al. A guide to utilization of the microbiology laboratory for diagnosis of infectious diseases: 2013 recommendations by the Infectious Diseases Society of America and the American Society for Microbiology. *Clin Infect Dis*. 2013;57:e22-e121.
9. Gafter-Gvili A, Raibman S, Grossman A, et al. [18F]FDG-PET/CT for the diagnosis of patients with fever of unknown origin. *QJM*. 2014;[Epub ahead of print].
10. Shulman ST, Bisno AL, Clegg HW, et al. Clinical practice guideline for the diagnosis and management of group A streptococcal pharyngitis: 2012 update by the Infectious Diseases Society of America. *Clin Infect Dis*. 2012;55:e86-e102.
11. Ruiz-Esteban R, Sarabia PR, Delgado EG, et al. Procalcitonin and C-reactive protein levels as diagnostic tools in febrile patients admitted to a General Internal Medicine ward. *Clin Biochem*. 2012;45:22-25.
12. Gaieski DF, Mikkelsen ME, Band RA, et al. Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. *Crit Care Med*. 2010;38:1045-1053.

REVIEW QUESTIONS

1. A 72-year-old man was admitted with exacerbation of his chronic obstructive pulmonary disease (COPD) symptoms and temperature of 38.3° C. His respiratory symptoms improved and he defervesced with antibiotics; but on his third hospital day, he became extremely agitated, requiring haloperidol, a major tranquilizer and central nervous system dopamine-depleting agent. Within hours his temperature rose to 42° C, and he developed muscle rigidity and dysautonomia. What is the most likely etiology of his worsening clinical picture?

A. Aspiration pneumonia due to his agitatio
B. Pulmonary embolism
C. COPD exacerbation
D. Nosocomial infection
E. Neuroleptic malignant syndrome due to haloperidol

Answer: E Temperatures exceeding 41° C are often due to drug-induced imbalance in thermoregulatory mechanisms and may cause direct cellular damage (Chapter 434).

2. A 46-year-old female immigrant had returned to her work as a day-care teacher for a month after visiting her family in Central America at Thanksgiving when she developed watery diarrhea. She was otherwise well and afebrile. Her complete blood count and electrolyte panel were normal. Several of her pupils had similar illness. What is the most likely etiology of her illness?

A. *Plasmodium vivax* infection
B. Traveler's diarrhea due to toxigenic *E. coli*
C. Amebiasis
D. Shigellosis
E. Viral gastroenteritis

Answer: E No laboratory abnormalities accompany typical benign acute viral infections. The first consideration in evaluating such a patient is that an infection unrelated to travel is more likely to be the cause of the illness. Common viral respiratory infections and gastroenteritis are accompanied by temperatures below 102° F.

3. A 36-year-old man developed low-grade fever, sore throat, and a rash. He had pharyngeal erythematous-based vesicles and ulcerations as well as vesicles on his hands and palms, feet, and buttocks. His 2-year-old child was recovering from a similar illness. What is the most likely etiology of his illness?

A. Epstein-Barr virus infection
B. Herpes simplex virus infection
C. Leptospirosis
D. Lyme disease
E. Enteroviral hand-foot-and-mouth disease

Answer: E See Tables 280-3 and 280-4. The presence of fever and rash involving the palms and soles allows considerable narrowing of the differential diagnosis.

4. For which of the following ambulatory adult patients should an antibiotic be prescribed?

A. A 47-year-old woman with no underlying medical conditions who has had a persistent cough for 2 weeks.
B. A 28-year-old woman with 10-day illness and now with unilateral face pain, maxillary toothache, and worsening symptoms after initial improvement.
C. A 31-year-old woman with pharyngitis but no fever, tonsillar exudates, tender anterior cervical adenopathy, or cough.
D. A 54-year-old man with no fever and prolonged coughing lasting 6 weeks.
E. A 65-year-old man with COPD and yellow-green mucopurulent discharge and no other new symptoms.

Answer: B See section on initial management of suspected infection in the ambulatory setting.

5. A 76-year-old woman admitted for community-acquired pneumonia, treated empirically with moxifloxacin, developed severe watery diarrhea, abdominal pain, fever, and leukocytosis (white blood cell count of 30,000) on day 4 of hospitalization. What is the likely etiology of her nosocomial infection?

A. *C. difficile* infection
B. Urinary tract infection
C. Intravenous catheter–related infection
D. *Legionella* infection
E. Drug fever

Answer: A Toxin-producing *C. difficile* is the only significant nosocomial gastrointestinal infection seen in hospitalized patients.

281

APPROACH TO FEVER AND SUSPECTED INFECTION IN THE COMPROMISED HOST

KIEREN A. MARR

DEFINITION

This chapter focuses on the approach to suspected infection and fever in compromised hosts.[1] Multiple conditions, including burns, critical illness, and inherited immunodeficiencies (to name only a few), can compromise the immune system and render a person at risk for different infections, but this chapter focuses primarily on hosts who have acquired defects in immune function secondary to medical therapies, or the "medically immunosuppressed." Examples include treatment of neoplastic diseases (particularly hematologic disorders such as leukemia and lymphoma), organ transplantation, and treatment of collagen vascular or autoimmune diseases. Management of people with the acquired immunodeficiency syndrome (AIDS) is discussed in Chapters 384 through 395, and a more thorough discussion of primary immunodeficiency is provided in Chapter 250.

TABLE 281-1 CONDITIONS, INTERVENTIONS, AND IMMUNE DEFECTS TYPICALLY ENCOUNTERED IN COMPROMISED HOSTS

UNDERLYING CONDITION	INTERVENTION	TYPE OF DEFECT
Treatment of neoplastic diseases (particularly hematologic malignant neoplasms)	Underlying disease (without intervention)	Defects in production of bone marrow cells associated with defects in cellular immunity and phagocytic function (e.g., cytopenias associated with bone marrow infiltration with malignant cells)
	Cytotoxic chemotherapies	Bone marrow suppression; defects in primary and secondary humoral and cellular immunity; breach in mucosal barriers (skin, gut); impairment in mucociliary clearance; defects in other organ function (e.g., kidney, liver)
Hematopoietic stem cell transplantation	Underlying disease, without intervention (e.g., hematologic malignant neoplasms)	Defects in primary and secondary humoral and cellular immunity; defects in phagocytic cell quantity and function
	Cytotoxic conditioning therapy (± total body irradiation)	Bone marrow suppression; defects in primary and secondary humoral and cellular immunity; breach in mucosal barriers; defects in organ function
	Stem cell manipulation (e.g., T-cell depletion)	Delay in cellular engraftment
	Prophylaxis and treatment of graft-versus-host disease (e.g., corticosteroids, calcineurin inhibitors, antimetabolites, TNF-α antagonists)	Defective function in phagocytic cells and dysfunction of primary and secondary humoral and cellular immunity
Solid organ transplantation	Underlying disease, without intervention (e.g., diabetes, end-stage liver disease)	Organ dysfunction and miscellaneous immune dysfunction
	Induction therapies (e.g., corticosteroids, antilymphocyte globulin, splenectomy, anti–interleukin-2 Ab, anti-CD52 Ab, calcineurin inhibitors	Depletion and impairment in primary and secondary cellular and humoral immunity
	Surgical intervention and altered anatomy	Breach in mucosal barriers; defects in organ function
	Acute and chronic rejection prophylaxis and treatment (e.g., corticosteroids, calcineurin inhibitors, antimetabolites and alkylating agents, plasmapheresis, antithymocyte globulin, monoclonal antibodies to B and T cells, anticytokine therapies, T-cell costimulation blockers)	Defective function in phagocytic cells, primary and secondary humoral and cellular immunity
Treatment of collagen vascular and autoimmune diseases	Anti-inflammatory and immunosuppressive agents (corticosteroids, nonsteroidal anti-inflammatory drugs, calcineurin inhibitors, sirolimus, mycophenolate mofetil)	Defective function in phagocytic cells, primary and secondary humoral and cellular immunity
	Antimetabolite and alkylating agents	Bone marrow suppression, defects in primary and secondary humoral and cellular immunity
	Biologic immune response modifiers (e.g., antithymocyte globulin, monoclonal antibodies to B and T cells, anticytokine therapies, T-cell costimulation blockers)	Defective function in primary and secondary humoral and cellular immunity

Ab = antibody; TNF = tumor necrosis factor.

APPROACH TO THE PATIENT

The approach to the immunosuppressed patient requires detailed knowledge of the type of immune defect and related risks. Table 281-1 outlines general defects in host responses with the types of conditions that characterize groups of medically immunosuppressed patients discussed in the context of this chapter (treatment of neoplastic diseases, hematopoietic stem cell and solid organ transplantation, and treatment of autoimmune and collagen vascular diseases). As the management of many of these conditions becomes more and more complex and dependent on biologic agents that affect immune responses at both broad and focused targets, knowledge of prior therapies received has become critically important in developing an informed approach to fever.

Patients with Malignant Disease

In patients with neoplastic diseases, particularly hematologic malignancies, the underlying condition plays a role in dictating infectious risks. For instance, the absolute number of phagocytic cells belonging to the polymorphonuclear leukocyte series may be reduced or the function of those cells impaired in the setting of specific malignancies (e.g., acute or chronic leukemias). In conditions such as acute leukemia, in which the cells are abnormal in morphology and function and only a small proportion of normally functioning cells circulate, risks for bacterial infections are enhanced, even in the absence of administered cytotoxic therapies. In certain conditions, such as in the setting of chronic lymphocytic leukemia, there may be quantitative defects in humoral factors that are critical in host defense, such as circulating immunoglobulin G and immunoglobulin M antibodies, secretory immunoglobulin A antibodies, and components of the complement cascade that can directly lyse some bacteria. Another component of the population of phagocytic cells includes circulating monocytes and tissue macrophages and the fixed mononuclear cells of the reticuloendothelial system. These cells collaborate with helper T cells in defense against pathogens that can survive intracellularly, such as mycobacteria, fungi, and some viruses and parasites. The spectrum of infectious risks is further enhanced and prolonged after treatment with cytotoxic drugs. These therapies affect other organ functions that are critical to defense, especially the integrity of the gastrointestinal tract mucosal barrier and airway innate clearance mechanisms, posing additional susceptibilities to bacterial and fungal pathogens. In this manner, the neoplastic disorder itself and the specific therapies used to treat it combine to define both acute and chronic risks for infection.

Transplant Recipients

Hematopoietic stem cell transplantation (HSCT) (Chapter 178) posits additional risks to the patient as a result of the agents used for conditioning therapy in preparation for the stem cell transplantation, variable rate and magnitude of cellular engraftment, and, in recipients of allogeneic HSCT, administration of additional agents to modulate risks for and treatment of graft-versus-host disease (GVHD). Thus, risks for specific infections can be roughly divided on a time scale relative to engraftment. Organ dysfunction, loss of natural barriers (skin and gut), and neutropenia dictate enhanced early risks for bacteria and fungi that inhabit the gastrointestinal tract; impaired humoral and secondary immunity enhance late risks for infections caused by viruses, fungi, and encapsulated bacteria, especially in people treated aggressively for GVHD.

Immunodeficiency in solid organ transplant recipients (Chapter 49) is largely related to the acute initiation and chronic maintenance requirements of therapies to suppress T- and B-cell function to minimize the impact of allosensitization and to decrease risks for early and late graft rejection. Therapies have evolved over time, with increased use of targeted biologic therapies, but in general, risks are largely related to those associated with acute and chronic cellular and humoral dysfunction. The type and amount of therapy differ according to immunologic risk of recipients. Additional variables modulating overall risks for infection include the altered anatomy, surgical intervention, and potential of infection transmitted from the graft itself (i.e., donor derived).

Transplant recipients have increased risks both for acute infection and for reactivation of latent infections after initiation of immunosuppression. Hence, pretransplantation evaluation should be focused on detection of latent herpes viruses (e.g., cytomegalovirus [CMV]) and other pathogens (e.g., *Mycobacterium tuberculosis*) that can be transferred or reactivated with transplantation and immunosuppression.

Two important concepts regarding immunosuppression that have emerged from the field of transplantation include observations of the immunomodulatory effects of viral reactivation and infection and the "net state of immunosuppression."[2] It was long ago noted that viral infections (both reactivation and disease) enhance risks for other infections. This has been particularly well documented for CMV (Chapter 376), which is recognized as a risk for other infections in recipients of both hematopoietic stem cell and solid organ transplant grafts. Overall risks for infection are related to epidemiologic exposures and the net state of immunosuppression, dictated by multiple host, donor, and medical variables. This net state is variable in both quantity and changes in character over time, largely influenced by therapies to prevent rejection or GVHD, and other complications, such as viral reactivation. This concept, which originated from an understanding of solid organ transplantation, can perhaps be applied to the care of all immunosuppressed patients.

Patients Treated for Autoimmune Disease

Table 281-1 also outlines the types of immunosuppressive therapies frequently administered to patients for the control of connective tissue diseases and autoimmune conditions. This is detailed here to emphasize that this group of patients is growing in importance in both hospitalized and outpatient populations, with increased use of biologic immune response modifiers (Chapters 35 and 36) enhancing risks for both reactivation of latent infection (e.g., *M. tuberculosis* and *Histoplasma capsulatum*) and severe manifestations of acute infection. Infectious risks should be considered in balancing need for these therapies, designing preventive regimens, and creating differential diagnoses of suspected infection.

FEVER IN THE COMPROMISED HOST

The onset of fever in a compromised patient can be an ominous development, and depending on the nature and magnitude of the impaired host defenses, a febrile response can herald the onset of a life-threatening systemic infection. A diagnostic approach should be derived by careful consideration of the patient's signs and symptoms of infection, immunosuppression, and whether the patient is at heightened risk for reactivation of latent infection. Because infection can progress rapidly, empirical antimicrobial therapy may be indicated even before an infection is definitively diagnosed. In this situation, empirical antimicrobial therapy may be indicated even before an infection is definitively diagnosed. Here, the common scenario of fever in the neutropenic host is discussed in depth.

Fever during Neutropenia: Diagnostic Considerations

If fever occurs in the setting of chemotherapy-induced neutropenia, the risk for bacterial infection increases proportionally with the decline in neutrophil count, especially with prolonged durations of significant neutropenia. Early pivotal studies documented that infection rates increase with neutrophil levels lower than 1000 cells/mm^3, progressively increasing as counts decline to less than 100 cells/mm^3. The duration of significant neutropenia is also an important determinant of the type of infection most likely to occur, with the risk for bacterial and fungal infections increasing with each successive week in which leukocyte counts are less than 500 cells/mm^3. In these studies, neutropenia and lymphopenia played significant roles in influencing infection rates in the setting of acute leukemia; however, neutropenia alone was more important than lymphopenia alone. These studies marked some of the earliest efforts that laid the foundation for our current approach to treatment of fever during neutropenia.

Historically, the most common causes of fever during neutropenia were gram-negative bacteria arising from the gastrointestinal tract. These observations drove establishment of empirical and prophylactic antibiotic practices designed to prevent and to treat unrecognized infection caused by the most common predicted pathogens. In the 1990s, concurrent with increased use of prophylactic and empirical antibiotics, especially quinolones and extended-spectrum β-lactams, reported rates of gram-negative bacteremias declined, with proportional increases in the numbers of bacteremias caused by gram-positive organisms. Why the change in epidemiology occurred is a matter of debate, but it is likely multifactorial; in addition to increased use of effective preventive antibacterials, there may be a role played by increased use of indwelling intravascular devices, positing higher risks for gram-positive bacteremias. It has also been recognized that fever that persists despite administration of broad-spectrum antibacterial therapy may herald the onset or presence of undiagnosed invasive fungal infections.

TABLE 281-2 APPROACH TO FEVER DURING CHEMOTHERAPY-INDUCED NEUTROPENIA

PAST AND CURRENT CLINICAL CONSIDERATIONS
What is the type and duration of immunologic deficiency?
Does the patient have any organ dysfunction that would predispose to particular infection?
Does the patient have any unique environmental or epidemiologic exposures?
What are the patient's prior infections and colonizing organisms?
What are the current and recently administered antimicrobial agents?
Are there any specific presenting signs or symptoms that suggest a particular type of infection or syndrome?

The importance of mucositis in driving inflammation and leading to development of bacterial or fungal infection through mucosal barrier injury cannot be overemphasized in patients administered cytotoxic therapies. Studies have shown that mucositis can produce inflammation adequate to drive development of fever. It is also likely that some people develop fever by transient seeding of the blood stream with colonizing bacterial or fungal pathogens. Some of these infections may be caused by organisms that are less well adapted to growth with standard microbiologic methods. These concepts support liberal use of empirical antimicrobials in the febrile neutropenic setting, with the focus on administration of a compound that is active against the most likely pathogens, considering the patients' epidemiologic exposures and colonizing organisms, especially in the gastrointestinal tract.

Table 281-2 lists multiple questions and considerations that the clinician should entertain when approaching fever in the neutropenic patient. The differential diagnosis of fever in the setting of chemotherapy-induced neutropenia is influenced by local and hospital exposure and the type of preventive antibiotics administered to the patient, which serve to alter microbial epidemiology within the gastrointestinal tract (see Table 281-2). Importantly, the type and the duration of immunodeficiency can alter overall risks, with "latent" infections presenting at development of first fever. Specific organ dysfunction, such as underlying pulmonary disease or renal impairment, can predispose to unique infectious syndromes (see later). Epidemiologic exposures should be thoroughly solicited; for instance, diagnostic evaluation should consider whether the patient previously or currently resides in areas endemic for *M. tuberculosis* or other infections that become latent. Current and previously administered antimicrobial drugs both affect risks for specific infections and can alter host microbial epidemiology. With this in mind, it is useful to have some information on colonizing organisms that may display complex resistance profiles, such as vancomycin-resistant enterococci and bacteria that express extended-spectrum β-lactamases or other resistance determinants (carbapenemases). Knowledge of recent colonization with these organisms should tailor initial antibiotic management, especially in patients who present severely ill.

One early consideration in treatment of fever during neutropenia is whether the patient requires hospitalization for therapy.[3] Risk assessment is an integral part of early evaluation to determine whether outpatient therapy is feasible. Two risk assessment systems have been developed, with the Multinational Association of Supportive Care in Cancer score validated to serve as a useful predictor of outcome, potentially assisting in identifying patients who can be treated with oral antibiotics and close monitoring rather than with inpatient therapy. Although the score is useful as a general guide to risk stratification, other variables that are important to consider in making risk assessment are underlying disease (e.g., lymphoma vs. leukemia), past and anticipated duration of neutropenia, symptoms and signs of infection foci, other comorbidities, and, perhaps most important, whether the patient has access to immediate and reliable medical attention if discharged from the medical facility. Recent guidelines suggest that febrile neutropenic patients can be managed as outpatients, provided the risk index is low enough and empirical antibacterial therapy is administered within an hour of triage, with close monitoring for stability to ensure safety in outpatient management.

The onset of fever should trigger a prompt and thorough bedside evaluation of the patient. Beginning with examination of the head and neck, there should be a specific examination for evidence of central nervous system (CNS) infection as well as a general evaluation of mental status. The

oropharynx must be examined for evidence of pharyngitis and focal tenderness. Sinus membranes should be evaluated for the presence of erythema or necrosis. Complete examination of the heart, lung fields, and abdomen is critical, with attention to the potential presence of new murmurs, pneumonia, and intra-abdominal tenderness. The perirectal area and the entire integument should be examined. Intravenous catheter exit sites and tunnels should be carefully examined, and blood should be drawn through all catheter channels for culture. Because catheter exit sites and tunnels can be infected in neutropenic patients without showing early signs of inflammation and erythema and with classic signs of infection presenting only after recovery of neutrophils, examination should be performed daily and with close scrutiny for evolving localized infection that may necessitate catheter removal.

Laboratory studies should be undertaken, with emphasis placed on procedures that can yield prompt results, such as Gram stain of body fluids, exudates, or aspirates. Routine blood work should include a complete blood count with differential, serum creatinine concentration, and screening liver function studies. A chest radiograph should be part of the initial evaluation, as should routine urinalysis. Because routine radiographs are insensitive for detection of small nodular lesions, especially those caused by filamentous fungi, computed tomography (CT) should be performed in evaluating persistent fever, especially in the presence of airway symptoms. No biomarker has yet to be proved reliable in discriminating between severe infection or other causes of fever during neutropenia, although studies have focused attention on the utility of lipopolysaccharide-binding protein, interleukins 6 and 8, procalcitonin, and C-reactive protein, to name only a few.

Although fever is the hallmark of infection, it is not specific for the presence of an infectious process. The development of fever may be a result of multiple causes, including medications, reaction to blood components, Sweet's syndrome, and GVHD. The fundamental principle is that infection should be suspected as the most likely cause of fever in a compromised host, and therapy should be applied empirically, even as diagnostic tests are being performed. Multiple episodes of fever during prolonged hospitalization and neutropenia are not uncommon; each episode requires comprehensive assessment. After a documented infection, it should not be assumed that a subsequent episode of fever is caused by the same recrudescent pathogen; the law of diagnostic parsimony tends to be less reliable in immunosuppressed patients.

Fever during Neutropenia: Management

Progression of infection can occur rapidly in neutropenic hosts. Very high mortality rates associated with bacteremia, especially that caused by gram-negative bacteria, triggered the introduction of routine empirical therapies (i.e., treatment of fever before diagnosis of infection). There are now many options for initial antibiotic therapy; choice should be tailored according to patient and institutional variables, as outlined in Table 281-1 and Figure 281-1. The first therapeutic distinction is whether a patient is at high risk, warranting inpatient management and intravenous antibiotic therapy, or low risk, potentially treated with oral regimens as an outpatient. In low-risk patients, the combination of a fluoroquinolone such as ciprofloxacin with amoxicillin-clavulanate has been shown to be effective. In high-risk patients, admission for treatment and prompt administration of a broad-spectrum antibiotic regimen is necessary. An international guideline panel of the Surviving Sepsis Campaign recommends starting antibiotics as soon as possible, preferably within an hour of recognition of fever during neutropenia. Although these recommendations were not specifically developed for this population, recent outcomes studies suggest that delays in administering antibiotics may be associated with prolonged hospital stays.[4]

Early studies demonstrated that the combination of an antipseudomonal β-lactam and an aminoglycoside is effective, but a recent meta-analysis showed that monotherapy with one of the new broad-spectrum β-lactams is associated with better outcomes compared to the combination therapy. Extended-spectrum agents, such as third- and fourth-generation cephalosporins and

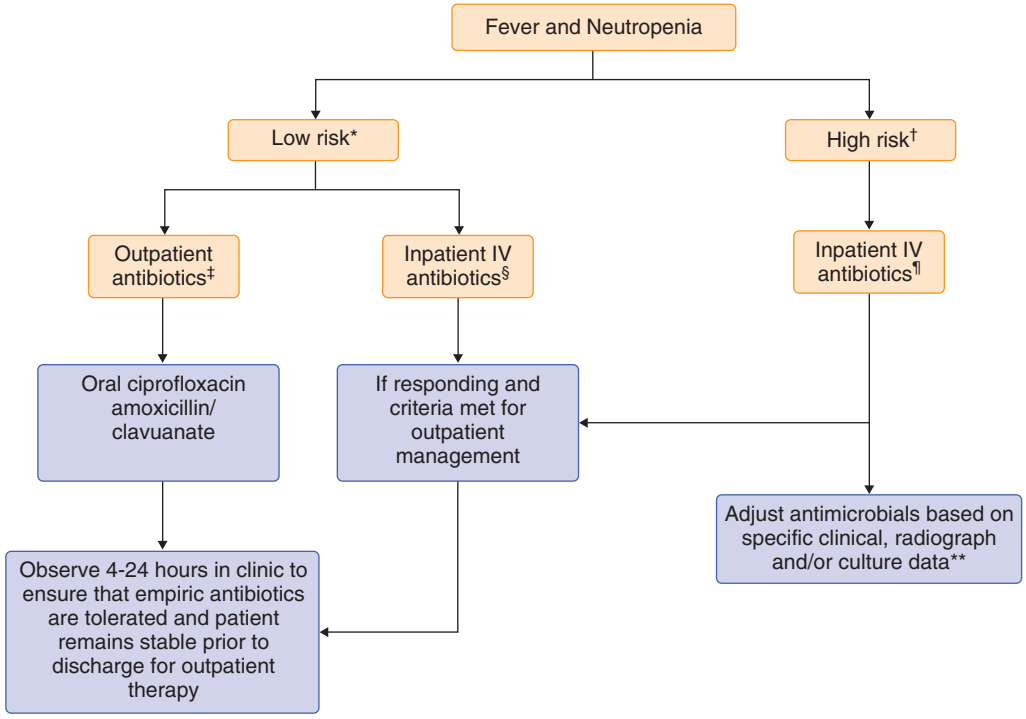

* Low Risk = anticipated neutropenia ≤7 days and clinically stable and no medical comorbidities.
† High Risk = anticipated neutropenia >7 days, *or* clinically unstable, *or* any medical comorbidities.
‡ If able to tolerate and absorb; caregiver, access, and transportation are available; patient and physician decide.
§ If there is documented infection requiring IV antibiotics; there is gastrointestinal intolerance; patient and physician decide.
¶ Empiric antibiotic monotherapy with any of the following: Piperacillin/tazobactam, *or* Carbapenem, *or* Ceftazidime, *or* Cefepime.
** For example: vancomycin and linezolid for cellulitis or pneumonia; add aminoglycoside and switch to carbapenem for pneumonia or gram negative bacteremia; metronidazole for abdominal symptoms of suspected *C. difficile* infection.

FIGURE 281-1. Initial management of fever (≥38.3°C) and neutropenia (≤0.5 × 10^9 cells/mL). Limited data to support recommendation. (Modified from Freifeld AG, Bow EJ, Sepkowitz KA, et al. Clinical practice guideline for the use of antimicrobial agents in neutropenic patients with cancer: 2010 update by the Infectious Diseases Society of America. *Clin Infect Dis.* 2011;52:e56-e93.)

carbapenems, were then shown to be effective options administered as monotherapy. Meta-analyses have now shown that the routine use of an aminoglycoside in combination may result in more toxicities and no better outcomes.

One major decision point in early therapy involves when to initiate vancomycin in high-risk patients. Advocates point to the risk for gram-positive infections that may carry a higher morbidity in neutropenic patients, including those caused by *Staphylococcus aureus* and viridans streptococci. However, because initiation of vancomycin with fever does not affect outcomes except in the case of treating breakthrough viridans streptococcal infection, in a setting where toxicities may offset benefits, most guidelines do not support its administration except in the case of documented or suspected catheter-related infection, colonization with penicillin- and cephalosporin-resistant pneumococci or methicillin-resistant *S. aureus*, positive blood cultures for gram-positive bacteria, or hemodynamic instability.

The median time to defervescence is shorter (2 days) in low-risk patients and longer (5 to 7 days) in high-risk patients. Clinical response to the first few days of therapy is a critical determinant of the course of extended antimicrobial therapy. Therapy should be tailored to the diagnostic findings. If patients are stable yet still febrile during a period of prolonged and severe neutropenia, clinical judgment must be used in deciding whether to maintain the initial regimen or to switch to an alternative regimen. If patients become afebrile after 3 to 5 days of antibacterial treatment but cultures are negative, some authorities recommend continuing the broad-spectrum intravenous coverage until recovery of the neutrophil count. This, however, may not be practical for patients with leukemia in blast crisis or refractory aplastic anemia, in whom periods of aplasia lasting weeks or more are common. Others believe that a switch to oral treatment is justifiable (e.g., a fluoroquinolone possibly paired with a β-lactam) if the patient becomes afebrile and appears to be clinically stable. If the patient remains afebrile, there is rapid improvement in the underlying condition, such as with recovery of the circulating neutrophil count to higher than 500 cells/mm^3, and no focus of infection is identified, discontinuation of treatment is an option. If the neutrophil count recovers to above 500 cells/mm^3 and fever persists, clinical judgment must be used to define needs for antimicrobial therapy while a search for the cause of the fever is continued. If the patient was initially treated with vancomycin and no confirmatory cultures supporting continued vancomycin use are obtained after 3 days (e.g., no coagulase-negative staphylococci from blood or methicillin-resistant *S. aureus*), intravenous vancomycin therapy should be discontinued.

Clinical deterioration should trigger consideration of infections resistant to the empirical regimen. Classic examples include breakthrough streptococcal bacteremia in patients not receiving vancomycin; vancomycin-resistant enterococci; breakthrough extended-spectrum β-lactamase–producing gram-negative bacteria in patients taking single-agent β-lactams; *Stenotrophomonas* species infections occurring in the setting of carbapenem monotherapy; and infections with multidrug-resistant pathogens, such as *Acinetobacter* species and organisms that produce carbapenemases.

There are many causes of persistent fever, both noninfectious and infectious. Noninfectious causes include hematomas, drug reactions, transfusion reactions, pulmonary emboli, splenic infarcts, and the underlying malignant disease. The possibility of infection caused by nonbacterial pathogens, such as fungi (especially *Candida* and *Aspergillus* species), should prompt evaluation and consideration of antifungal empirical therapies in the setting of fever that persists more than 4 to 7 days. Many drugs have been evaluated and shown to be effective in this setting, including azole drugs, echinocandins, and polyenes. Choice should be tailored to current antifungal prophylaxis, diagnostic findings, suspicion of *Candida* versus *Aspergillus* infection, and organ function. Current efforts are focused on developing reliable screening methods to negate the need for potentially toxic antifungals and to derive "preemptive" strategies for guiding use.

There are no hard and fast rules for duration of therapies; the simplest recommendation is to treat documented pathogens until the signs and symptoms of infection subside. On the other hand, persistently compromised hosts may remain febrile for weeks without identification of the cause. For a patient with persistent fever in whom no pathogen is identified, the duration of therapy must be based on integration of clinical data and the best estimate of the direction of the host's status. As mentioned previously, therapy can be discontinued in stable, afebrile patients, assuming that the absolute neutrophil count exceeds 500 cells/mm^3. Clearly, if broad-spectrum antibacterial therapy is to be discontinued, the patient must be monitored carefully thereafter. For patients whose neutrophil counts remain at levels less than 500 cells/mm^3, particularly the subset with severe profound neutropenia of less than 100 cells/mm^3, it is prudent to continue empirical antibacterial and antifungal therapy, with reappraisal of all diagnostic measures. The decision to stop therapy at whatever arbitrary interval may be justified if the patient's condition is stable.

The use of granulocyte colony-stimulating factors for the prevention of febrile neutropenia is discussed in Chapter 167. For the management of established fever in neutropenic patients, evidence-based guidelines have been published by the American Society of Clinical Oncology and by the Infectious Diseases Society of America. The National Comprehensive Cancer Network and the German Society for Haematology and Medical Oncology have likewise published recommendations for the use of myeloid growth factors in the oncology setting. [5,6] The guidelines generally agree, and they support the use of colony-stimulating factors in similar circumstances that indicate high risk for infection-associated complications and poor clinical outcomes, such as in patients with anticipated long and profound durations of neutropenia, age older than 65 years, uncontrolled primary disease, pneumonia, hemodynamic compromise, and invasive fungal infections.

The remainder of this chapter discusses diagnosis and evaluation of specific syndromes that are common in individuals who are immunocompromised because of multiple conditions, including transplantation. Table 281-3 summarizes some of the most common infectious and noninfectious syndromes that involve the skin, lungs, gastrointestinal tract, and nervous system, as discussed next.

CUTANEOUS SYNDROMES

Cutaneous abnormalities can provide a clue to bacteremia, and aspiration and culture of suspicious lesions can be as valuable as a blood culture. Ascending streptococcal or staphylococcal cellulitis can occur in both immunocompromised and non-compromised patients. Metastatic abscesses are a well-recognized part of the *S. aureus* bacteremia syndrome. Necrotizing vasculitis is classically associated with *Pseudomonas aeruginosa* infections; its cutaneous lesion of ecthyma gangrenosum is an erythematous, indurated target or "bull's-eye" lesion with an area of central necrosis that can appear in crops (Fig. 281-2). However, other gram-negative endotoxin-producing bacteria have been associated with similar cutaneous lesions.

In the neutropenic host, disseminated fungal infections may be initially recognized by characteristic cutaneous lesions. Disseminated candidiasis in neutropenic patients can be manifested with diffuse maculopapular, erythematous, and sometimes tender lesions. The appearance of cutaneous lesions typically changes in character with engraftment of neutrophils (Fig. 281-3). Disseminated infections caused by filamentous organisms such as *Aspergillus* species cause similar lesions, but usually fewer in number and more often with some component of central necrosis. Other filamentous fungi, namely, those with which infection is characterized by a high fungal burden, such as *Fusarium* species, typically cause more skin lesions in multiple stages of evolution, ranging from papules to larger erythematous lesions with central necrosis. Multiple filamentous fungi such as *Aspergillus* species and Zygomycetes can also cause primary cutaneous lesions, especially with a breech in skin integrity associated with catheter sites, trauma, and surgery. Infection with *Cryptococcus neoformans* can be accompanied by cutaneous involvement, with manifestations ranging from molluscum-like lesions to primary cutaneous cellulitis, which may be especially common in solid organ transplant recipients. Cutaneous lesions are an opportunity to establish diagnosis through aspiration, biopsy, and culture.

Morbilliform eruptions or maculopapular exanthems are frequent in neutropenic patients and transplant recipients, and they can be caused by drug reactions, GVHD, and numerous viral infections. Primary infection and reactivation with herpes viruses such as CMV and Epstein-Barr virus can be accompanied by rashes, and diagnostics should be considered in the appropriate context. Human herpesvirus 6, the primary cause of roseola infantum in childhood, leads to latency and can cause disease in immunocompromised hosts both by reactivation and by primary infection. Disease can be accompanied by fever, rash, myelosuppression, and involvement of other organ systems (e.g., CNS). In immunocompromised patients, adenovirus can be both primarily acquired, usually through the respiratory tract, and reactivated; it causes fever, rash, and potentially disease involving multiple organ systems (lungs, gastrointestinal tract, kidneys, liver, CNS). In HSCT recipients, the constellation of fever, rash, diarrhea, and hepatitis may be confused for severe GVHD. Parvovirus B19 infection can be severe in immunocompromised hosts and associated with fever, rash, and manifestations of hemophagocytosis, although there are other infectious causes of hemophagocytic syndromes as well (Chapter 169). The characteristic vesicular rashes of

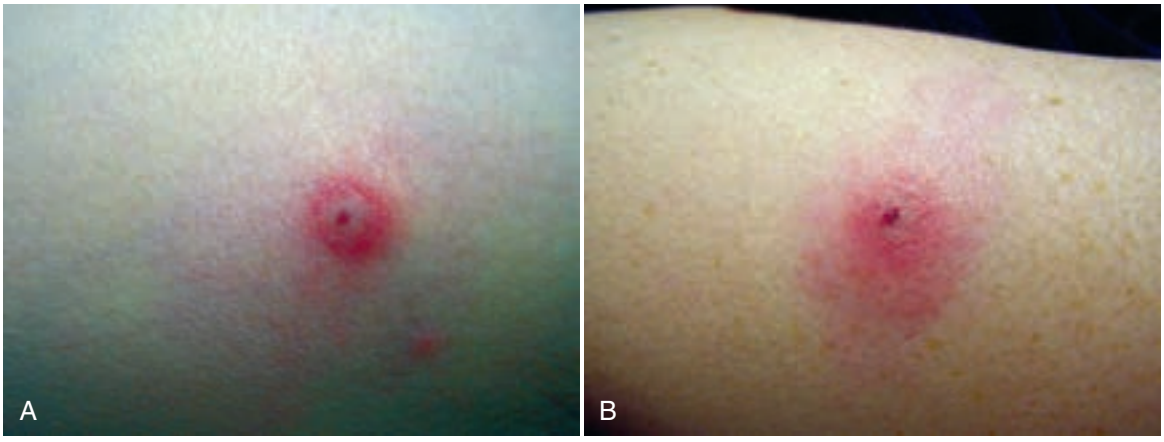

FIGURE 281-2. Ecthyma gangrenosum. A 28-year-old woman with fever and neutropenia while receiving chemotherapy for acute leukemia developed several tender edematous papules on her thighs. A, Central crust and surrounding erythema are shown. B, The papules became necrotic during 1 to 2 days, with the formation of black, well-demarcated eschar. Cultures from blood and the necrotic eschar grew *Pseudomonas aeruginosa*. (© DermAtlas; http://www.DermAtlas.org.)

TABLE 281-3 COMMON INFECTIOUS AND NONINFECTIOUS SYNDROMES IN IMMUNOCOMPROMISED HOSTS

PRIMARY ORGAN SYSTEM	BACTERIA	FUNGI	VIRUSES	NONINFECTIOUS
Cutaneous	Disseminated gram-positive and gram-negative bacteria, e.g., *Staphylococcus aureus* *Pseudomonas aeruginosa* *Mycobacterium* spp *Nocardia* spp	*Candida* spp. Filamentous fungi, e.g., *Aspergillus* spp Zygomycetes *Fusarium* spp *Scedosporium* spp *Cryptococcus* spp	Herpes simplex Varicella-zoster CMV HHV-6 Adenovirus Parvovirus B19	Drug eruptions GVHD Sweet's syndrome
Sinopulmonary	Gram-positive and gram-negative causes of sinusitis and pneumonia *S. aureus* *Streptococcus pneumoniae* *P. aeruginosa* *Haemophilus influenzae* Anaerobes Legionella *Nocardia* spp *Mycobacterium* spp	Filamentous fungi, e.g., *Aspergillus* spp Zygomycetes *Fusarium* spp *Scedosporium* spp *Cryptococcus* spp Endemic fungi, e.g., *Histoplasma capsulatum* *Coccidioides immitis* *Blastomyces dermatitidis* *Pneumocystis jiroveci*	Respiratory viruses, e.g., RSV Parainfluenza Influenza Adenovirus Reactivation herpes viruses, e.g., VZV	Drug-related pulmonary toxicities Pneumonitis (sirolimus) Diffuse alveolar damage Bronchiolitis obliterans syndromes
Gastrointestinal	Bacterial enterocolitis ("typhlitis") Mixed gram-positive, gram-negative *Clostridium difficile* colitis Enteric diarrheal pathogens *Salmonella* spp *Shigella* spp *Escherichia coli* *Campylobacter* spp	*Candida* spp.	CMV EBV-PTLD Adenovirus Coxsackievirus Rotavirus Norovirus	Drug-related toxicities, e.g., MMF
Neurologic	Gram-positive and gram-negative bacteria *Listeria* spp Pneumococcus Meningococcus *Mycobacterium tuberculosis*	Filamentous fungi *Cryptococcus* spp	Herpes viruses HSV HHV-6 VZV JC virus West Nile virus Miscellaneous viral encephalitides	Drug-related toxicities, e.g., carbapenem-related seizures PRES

CMV, cytomegalovirus; EBV-PTLD = Epstein-Barr virus–post-transplantation lymphoproliferative disorder; GVHD = graft-versus-host disease; HHV-6 = human herpes virus 6; HSV = herpes simplex virus; MMF = mycophenolate mofetil; PRES = posterior reversible encephalopathy syndrome; RSV = respiratory syncytial virus; VZV = varicella-zoster virus.

varicella reactivation are frequent in stem cell and solid organ transplant recipients with chronic T-cell deficiencies, especially in the absence of antiviral prophylaxis; an important therapeutic consideration is to recognize and aggressively treat disease that involves viscera because disseminated disease is associated with high mortality rates. Antivirals administered as prophylaxis in high-risk HSCT and solid organ transplant recipients can decrease both early morbidity and late mortality associated with herpes simplex virus (HSV), varicella-zoster virus (VZV), and CMV disease, although drug-related toxicities need to be measured in risk-benefit calculations.

There are numerous noninfectious causes of rashes and lesions that are common in immunocompromised individuals. Drug-induced hypersensitivity syndromes can be both mild and severe, potentially associated with life-threatening toxic epidermal necrolysis. As in other populations, antimicrobial agents are frequently implicated as causative agents. GVHD, especially during the acute phase, frequently is manifested as suspected infection, with fever, nonspecific rashes, and frequently disease involving the gastrointestinal tract (diarrhea, hepatitis).

Sweet's syndrome, or acute febrile neutrophilic dermatosis, is characterized by skin lesions with neutrophilic infiltration in the dermis (see Fig. 440-22). This presents a diagnostic dilemma in neutropenic patients and has been described in the setting of impending neutrophil recovery or treatment with granulocyte colony-stimulating factor. It is also associated with numerous drugs and can appear as a paraneoplastic phenomenon. Biopsy with appropriate microbial stains and culture is essential to distinguish these lesions from infectious causes of ecthyma gangrenosum and other disseminated infections, such as those caused by fungi.

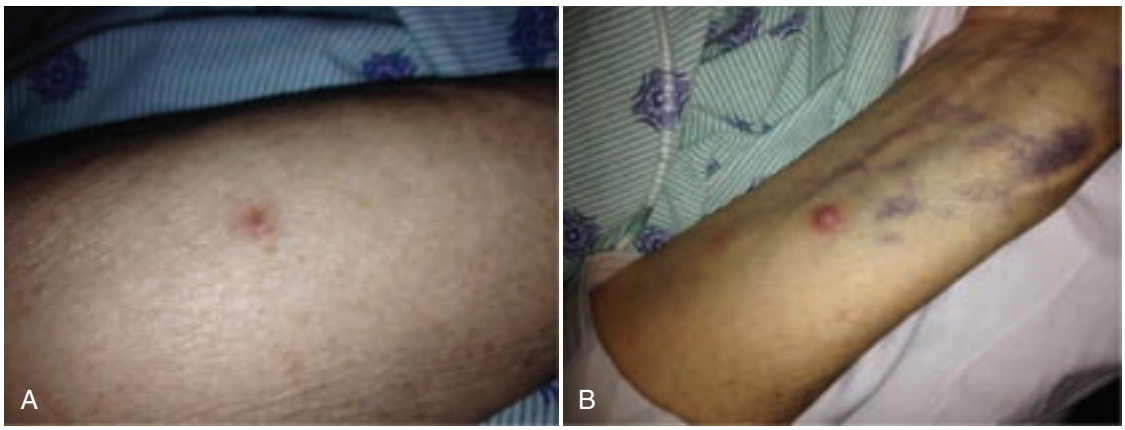

FIGURE 281-3. Disseminated candidiasis. A 60-year-old woman with fever during neutropenia that developed after receipt of therapy for acute leukemia developed tender papular lesions on her extremities, trunk, and back. **A,** Blood cultures returned positive for *Candida tropicalis*. **B,** After resolution of neutropenia, lesions developed a more pustular appearance.

RESPIRATORY SYNDROMES

The lungs are a challenging site in evaluating fever in a compromised patient because detection of abnormalities is easy but obtaining lung secretions or infected tissues can be difficult. Pneumonia should be suspected in a patient who has respiratory symptoms as manifested by cough, shortness of breath, chest pain, and hypoxia. In the early stages of pneumonitis, routine chest radiographs may be normal, whereas more expensive imaging procedures such as CT can reveal pulmonary infiltrates or abscesses. Both community-acquired pathogens such as pneumococci and *Haemophilus influenzae* can cause lobar or diffuse pneumonia. Gram-negative bacilli can cause pneumonia of a necrotizing type in severely neutropenic patients. All patients who are receiving ventilator support are at risk for secondary gram-negative bacillary pneumonia or staphylococcal pneumonia. Clusters of outbreaks of *Legionella pneumophila* infection have occurred in immunocompromised patients maintained in dialysis, transplant, or intensive care units; these outbreaks reflect institutional environmental contamination.

Opportunistic fungi have been increasingly recognized as causes of lung infection in compromised neutropenic patients and transplant recipients. A travel history is essential in a compromised patient who has evidence of lung disease; epidemic mycoses such as blastomycosis, coccidioidomycosis, and histoplasmosis may be manifested as acute pneumonia after recent exposure (although more typically, initial exposure in a normal host leads to containment in an initial focus of fungal lung disease). After immune suppression, the primary focus can be the source of reactivated disease. *Candida* species, in contrast, are uncommon primary lung pathogens. Although *Candida* species commonly colonize indwelling vascular and urinary catheters, candidal pneumonia is unusual in the absence of systemic candidiasis. Systemic candidiasis usually originates from the gastrointestinal tract if it is not secondary to vascular catheter infection, but *Candida* lung nodules can occur after systemic infection, especially with high fungal burden. Although traditionally associated with an "interstitial pattern" of lung infiltration, *Pneumocystis* species pneumonia can be manifested as local consolidation or pulmonary nodules, exhibiting granulomatous inflammation on pathologic examination. Filamentous fungi, which include *Aspergillus*, Zygomycetes, and less frequently diagnosed *Fusarium* species and *Scedosporium* species, are difficult to treat. These infections may be accompanied by chest pain and occasionally hemoptysis. From an initial focus, *Aspergillus* infection can spread through the pulmonary vasculature, which sets the stage for localized hemorrhage, creating a halo sign on CT scan, and infarction and necrosis, which can progress to cavitary lesions. Non-neutropenic hosts frequently also develop less specific radiographic findings, such as bronchopneumonia. These organisms can also cause primary airway disease, presenting with features typical of tracheobronchitis, with or without findings apparent on CT scan. This infection is particularly well described in lung transplant recipients, who may also have involvement of the bronchial anastomosis.

Attention has recently been drawn to the high risks for severe pulmonary infection caused by reactivation of *M. tuberculosis* in people who are treated with biologic immune response modifiers (Chapter 36) in the setting of autoimmune or other inflammatory diseases, such as with rheumatoid arthritis, psoriasis, and inflammatory bowel diseases. High risks are particularly well described in the setting of treatment with tumor necrosis factor antagonists, justifying enhanced vigilance, including routine screening for prior infection, and consideration of preventive chemotherapy.[7] These patients are also at increased risk for invasive fungal infections, including reactivation of endemic infections such as histoplasmosis, warranting enhanced screening and high suspicion for disease.

Viral infections that involve the lung are difficult to diagnose in immunocompromised patients. A particularly common concern is reactivation pneumonitis caused by members of the herpesvirus family, especially CMV, which occurs most frequently in the setting of chronic T-cell depression associated with transplantation. Respiratory viruses, which infect immunocompromised hosts with the same frequency as in the general population, cause lower tract disease and pneumonitis more frequently in hosts with suppressed cellular immunity. Essentially any respiratory virus can cause upper and lower tract disease, depending on geographic and seasonal epidemiology; most frequently recognized are respiratory syncytial virus, parainfluenza viruses, influenza viruses, and adenoviruses. Patients who have defects in cellular immunity typically exhibit higher viral loads and prolonged shedding, presenting considerations with regard to enhanced therapy, emergence of antiviral resistance, and infection control. The transplant population has served as a sentinel population in recognizing multiple emerging viral pathogens, including human metapneumovirus, bocavirus, and the KI and WU polyomaviruses.

There are numerous recognized noninfectious causes of pulmonary infiltrates in immunosuppressed hosts. These include early complications of chemotherapy administration, such as diffuse alveolar damage and hemorrhage, and late complications of GVHD and organ rejection (e.g., bronchiolitis obliterans syndromes). Certain drugs that are frequently administered in these populations of patients can cause direct lung toxicity; one classic example is the proliferation signal inhibitor sirolimus.

GASTROINTESTINAL SYNDROMES

There are multiple causes of diarrhea in a compromised host, including conventional enteric pathogens such as *Salmonella* (Chapter 308), *Shigella* (Chapter 309), and *Campylobacter* (Chapter 303). More recently, attention has been drawn to the frequency of noroviruses as an important cause of chronic gastroenteritis in immunocompromised patients, in whom diagnoses have been classically elusive and misapplied to noninfectious syndromes (e.g., GVHD), and outcomes can be poor.[8] In patients who have been in the hospital and have been receiving multiple courses of antibiotic treatment, *Clostridium difficile* (Chapter 296) is a common occurrence, and colitis can be both severe and persistent in immunosuppressed hosts. Two acid-fast staining parasites are *Isospora belli* and *Cryptosporidium* species, and they are associated with predisposing impairments in cell-mediated immunity. More recently recognized are the microsporidians. *Giardia lamblia* is classically associated with hypogammaglobulinemia. Individuals who have received long courses of chemotherapy, radiation, and antibiotics commonly experience *Candida* mucosal overgrowth in the mouth and esophagus. HSV and CMV can cause symptoms identical to those of *Candida* esophagitis. In severely neutropenic patients, anaerobic streptococci and gram-negative pathogens such as *P. aeruginosa* can cause severe mucositis and pharyngitis. In cancer patients, these organisms take advantage of the cytotoxic effects of chemotherapy, which promotes sloughing of mucosal surfaces and subsequently predisposes to infection.

Neutropenic patients may develop enterocolitis that can be of mixed anaerobic and aerobic bacterial origin. Neutropenic enterocolitis, also known as

typhlitis or *ileocecal syndrome*, results from chemotherapeutic damage to the intestinal mucosa in the setting of neutropenia. Presentation usually includes fever, abdominal pain, nausea, vomiting, and diarrhea. Because neutropenic enterocolitis can rapidly progress to intestinal perforation, sepsis, and multisystem organ failure, prompt diagnosis and aggressive medical or surgical intervention are required.

CMV colitis can be focal or diffuse; in some solid organ transplant recipients, disease can be present without demonstration of a positive blood polymerase chain reaction for CMV, so endoscopy is required for diagnosis. Because the differential diagnosis of colitis and diarrhea is broad and includes multiple infections, focal Epstein-Barr virus–associated post-transplantation lymphoproliferative disorder, GVHD, and drug-induced toxicity such as that caused by mycophenolate mofetil, these patients should be evaluated with endoscopy to reach definitive diagnosis.

NEUROLOGIC SYNDROMES

Both gram-positive and gram-negative bacteria (including anaerobes) can cause brain abscess or meningitis. *Listeria monocytogenes* is a common cause of meningitis in a compromised host. This pathogen is a gram-positive pleomorphic bacillus that may be difficult to identify on routine Gram stain of cerebrospinal fluid. Encapsulated bacteria such as pneumococci and staphylococci can cause metastatic CNS disease and meningitis. In patients with impaired cell-mediated immunity, *C. neoformans* is also a leading cause of CNS infection. Other fungal pathogens, such as *Aspergillus* and other molds, can invade the CNS, both by direct sinus invasion and by hematogenous spread. Focal brain lesions and meningoencephalitis in individuals who have chronic deficiency in cellular immunity can be caused by reactivation or acute severe infection by typically latent organisms such as *Toxoplasma gondii*, *M. tuberculosis,* and *H. capsulatum*. Reactivated or quiescent CNS syphilis should also be considered in patients with severe immunologic impairment.

There are numerous viral causes of meningoencephalitis. Similar to other populations, one needs to consider enteroviruses, measles, and neurotropic herpesviruses (HSV-1, CMV, VZV). Human herpesvirus 6 is a common cause of encephalitis and post-transplantation acute limbic encephalitis in transplant recipients, characterized by seizures, anterograde amnesia, and neuroimaging abnormalities involving the temporal lobes. Progressive multifocal leukoencephalopathy caused by JC polyomavirus occurs in patients with chronic CD4 deficiency, such as with human immunodeficiency virus (HIV) type 1 infection. Studies have emphasized that transplant recipients are at increased risk for meningoencephalitis caused by West Nile virus compared with the general population. Anyone presenting early after transplantation should be considered at risk for potentially severe infections acquired from the donor; West Nile virus, rabies, HIV, HSV, and multiple other viruses have been documented to be transmitted through organ donation.

There are multiple noninfectious causes of neurologic symptoms, which include drug toxicities and immunologic disorders, such as paraneoplastic syndromes and Guillain-Barré syndrome. Particularly relevant in the transplant recipient, patients with autoimmune diseases, and recipients of high doses of cancer chemotherapy is the *posterior reversible encephalopathy syndrome*. The classic presentation of the posterior reversible encephalopathy syndrome encompasses a sudden onset of severe "thunderclap" headache, seizures, confusion, and visual disturbance, accompanied by a CT or magnetic resonance imaging pattern of predominantly posterior cerebral edema and angiographic evidence of reversible vasoconstriction. It may be caused by endothelial injury, vasospasm, or edema associated with drugs such as calcineurin inhibitors (cyclosporine, tacrolimus).

CONCLUSION

Infections are a major cause of mortality in immunocompromised hosts. The approach to fever and suspected infection requires knowledge of specific risks inherent to the type and duration of immunodeficiency, diagnostic diligence, and tailored therapeutics designed to avoid rapid progression and poor outcomes.

Grade A References

A1. Flowers CR, Seidenfeld J, Bow EJ, et al. Antimicrobial prophylaxis and outpatient management of fever and neutropenia in adults treated for malignancy: American Society of Clinical Oncology clinical practice guideline. *J Clin Oncol*. 2013;31:794-810.

A2. Vidal L, Ben Dor I, Paul M, et al. Oral versus intravenous antibiotic treatment for febrile neutropenia in cancer patients. *Cochrane Database Syst Rev*. 2013;10:CD003992.

A3. Paul M, Lador A, Grozinsky-Glasberg S, et al. Beta lactam antibiotic monotherapy versus beta lactam-aminoglycoside antibiotic combination therapy for sepsis. *Cochrane Database Syst Rev*. 2014;1:CD003344.

A4. Averbuch D, Orasch C, Cordonnier C, et al. European guidelines for empirical antibacterial therapy for febrile neutropenic patients in the era of growing resistance: summary of the 2011 4th European Conference on Infections in Leukemia. *Haematologica*. 2013;98:1826-1835.

A5. Goldberg E, Gafter-Gvili A, Robenshtok E, et al. Empirical antifungal therapy for patients with neutropenia and persistent fever: systematic review and meta-analysis. *Eur J Cancer*. 2008;44:2192-2203.

A6. Yahav D, Gafter-Gvili A, Muchtar E, et al. Antiviral prophylaxis in haematological patients: systematic review and meta-analysis. *Eur J Cancer*. 2009;45:3131-3148.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Patel DM, Riedel DJ. Fever in immunocompromised hosts. *Emerg Med Clin North Am*. 2013;31: 1059-1071.
2. Fishman JA. Opportunistic infections—coming to the limits of immunosuppression? *Cold Spring Harb Perspect Med*. 2013;3:a015669.
3. Gea-Banacloche J. Evidence-based approach to treatment of febrile neutropenia in hematologic malignancies. *Hematology Am Soc Hematol Educ Program*. 2013;2013:414-422.
4. Perron T, Emara M, Ahmed S. Time to antibiotics and outcomes in cancer patients with febrile neutropenia. *BMC Health Serv Res*. 2014;14:162.
5. Crawford J, Armitage J, Balducci L, et al. Myeloid growth factors. *J Natl Compr Canc Netw*. 2013;11:1266-1290.
6. Vehreschild JJ, Böhme A, Cornely OA, et al. Prophylaxis of infectious complications with colony-stimulating factors in adult cancer patients undergoing chemotherapy—evidence-based guidelines from the Infectious Diseases Working Party AGIHO of the German Society for Haematology and Medical Oncology (DGH). *Ann Oncol*. 2014;25:1709-1718.
7. Selmi C, Ceribelli A, Naguwa SM, et al. Safety issues and concerns of new immunomodulators in rheumatology. *Expert Opin Drug Saf*. 2014;1-11.
8. Green KY. Norovirus infection in immunocompromised hosts. *Clin Microbiol Infect*. 2014;20: 717-723.

REVIEW QUESTIONS

1. A patient who has an absolute neutrophil count below 500 cells/mm^3 for 2 weeks after induction therapy for acute myelogenous leukemia is at increased risk for

A. *E. coli* bacteremia from the gastrointestinal tract
B. Influenza infection involving the upper respiratory tract
C. Varicella-zoster virus reactivation
D. Cytomegalovirus disease involving the lungs
E. *Cryptococcus neoformans* meningoencephalitis

Answer: A These patients do not have increased risk for influenza involving the upper respiratory tract compared with the population in general but have increased risk for complications (such as pneumonitis). Answers C to E are correct only for people with advanced deficits in T-cell immunity.

2. Early antibiotic therapy would be appropriate before documenting a microbial diagnosis in

A. A neutropenic patient who has developed fever, with normal hemodynamics
B. A heart transplant patient who has a new left lower lobe consolidation, fever, and leukocytosis
C. A kidney transplant patient with pyuria, fever, and elevated creatinine concentration
D. None of the above
E. All of the above

Answer: E These all outline high-risk situations in which empirical antibiotic therapy would be appropriate before documentation of disease to avoid rapid progression with sepsis.

3. A man presents with diffuse papular skin lesions 20 days after receipt of bone marrow transplantation, with fever. He has been receiving fluconazole, moxifloxacin, and acyclovir prophylactically. The least likely cause of his skin lesions is

A. *Candida glabrata*
B. Sweet's syndrome
C. *Aspergillus fumigatus*
D. Herpes simplex virus
E. A vasculitic reaction to a drug

Answer: D Herpes simplex virus rarely disseminates to cause these types of skin lesions, especially in people receiving acyclovir.

4. People who have received tumor necrosis factor inhibitors as biologic therapy for an autoimmune disease have heightened risks for infections caused by which of the following?

A. Viruses
B. Bacteria
C. Fungi
D. All of the above
E. None of the above

Answer: D Tumor necrosis factor inhibition may alter immune responses to multiple arms of the immune system.

5. What is the most likely cause of neutropenic enterocolitis that developed after receipt of cytotoxic therapy for acute myelogenous leukemia?

A. Cytomegalovirus
B. *Clostridium difficile*
C. Norovirus
D. Mixed anaerobic and aerobic bacteria
E. Adenovirus

Answer: D These infections are typically caused by organisms that are "native" in the gastrointestinal tract during episodes of mucosal injury.

282 PREVENTION AND CONTROL OF HEALTH CARE–ASSOCIATED INFECTIONS

DAVID P. CALFEE

THE BURDEN OF HEALTH CARE–ASSOCIATED INFECTIONS

The Centers for Disease Control and Prevention (CDC) defines health care–associated infections (HAIs) as infections that patients acquire during the course of receiving health care treatment for other conditions. *Nosocomial infection* is a term that refers specifically to an HAI that develops in association with hospital care. The development of infection during the course of health care is not, however, limited to the acute care hospital setting. Thus, *health care–associated infection* is the preferred term in referring to the broader spectrum of infections that develop during the course of health care, wherever that care may be provided, including acute care hospitals, long-term care facilities, rehabilitation facilities, dialysis facilities, and even the patient's home during the receipt of home care services.

The most extensive data regarding the incidence of and outcomes associated with HAIs come from the acute care hospital setting. On the basis of data reported by U.S. hospitals to the CDC in 2002, it has been estimated that 1.7 million HAIs occur in U.S. hospital patients each year, with almost 99,000 associated deaths (Table 282-1). A point prevalence survey conducted in 2010 in 183 U.S. hospitals that the prevalence of HAI was 4%.[1] Previous European studies have estimated that 4.1 million HAIs occur in European acute care hospitals each year. Thus, approximately one of every 14 to 20 patients admitted to U.S. and European hospitals develops an HAI, making HAI one of the most common complications associated with the receipt of health care. Moreover, these data indicate that HAIs are one of the top 10 causes of death in the United States. Whereas many of these HAI-associated deaths occur among patients who are already severely ill and who have a high likelihood of death due to their underlying disease, a substantial proportion of HAI-related deaths occur among persons who were otherwise expected to survive their hospitalization. In a single-center study, 31% of unexpected in-hospital deaths were determined to be possibly or probably related to an HAI. In addition to an increased risk of death, patients who develop HAI suffer a number of other adverse outcomes, including prolonged hospital stays, additional medical interventions and antibiotic treatment, discomfort, and loss of function and income. It has been estimated that these HAIs cost U.S. hospitals between $28.4 billion and $45 billion each year.[2] These statistics are particularly concerning when they are considered with the knowledge that many of these infections are preventable. In fact, a systematic review found that 55 to 70% of four of the most common types of HAIs are preventable through the use of currently available, evidence-based preventive strategies (see Table 282-1).

Although the majority of HAI statistics come from acute care hospitals, there are data to demonstrate that HAIs are significant problems in other health care settings as well. Point prevalence surveys conducted in European and the U.S. Veterans Affairs system long-term care facilities found that the prevalence of HAI among long-term care facility residents ranged from 2.4 to 5.2%. The overall burden of HAI among long-term care facility residents has been estimated to be 1.64 to 3.83 million infections per year in the United States and at least 2.6 million infections per year in Europe. Vascular access–related infections are the most common HAIs among patients requiring

TABLE 282-1 ESTIMATES OF THE BURDEN, COSTS, AND PREVENTABILITY OF HEALTH CARE–ASSOCIATED INFECTIONS IN U.S. HOSPITALS

TYPE OF INFECTION	NUMBER OF INFECTIONS PER YEAR[a]	AVERAGE ATTRIBUTABLE COST* PER INFECTION[b]	NUMBER OF DEATHS (CASE-FATALITY RATE)[a]	PROPORTION PREVENTABLE[c]
Urinary tract infection	561,667	$749-$1007	13,088 (2.3%)	65-70%
Catheter-associated urinary tract infection	449,334			
Surgical site infection	290,485	$11,087-$34,670	8205 (2.8%)	55%
Pneumonia	250,205	$14,806-$28,508	35,967 (14.4%)	55%
Ventilator-associated pneumonia	52,543			
Blood stream infection	248,678	$6461-$29,156	30,665 (12.3%)	65-70%
Central line–associated blood stream infection	92,011			
Other	386,090	$5682-$9124	11,062 (2.9%)	
C. difficile infection	178,000			

*In 2007 U.S. dollars.
[a]Klevens RM, Edwards JR, Richards CL, Jr., et al. Estimating health care-associated infections and deaths in U.S. hospitals, 2002. *Public Health Rep.* 2007;122:160-166.
[b]Scott RD. The direct medical costs of healthcare-associated infections in U.S. hospitals and the benefits of prevention. Centers for Disease Control and Prevention; 2009. Available at: http://www.cdc.gov/hai/pdfs/hai/scott_costpaper.pdf. Accessed January 25, 2015.
[c]Umscheid CA, Mitchell MD, Doshi JA, et al. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. *Infect Control Hosp Epidemiol.* 2011;32:101-114.

chronic hemodialysis for end-stage renal disease, with approximately 37,000 catheter-related blood stream infections occurring in U.S. end-stage renal disease patients each year. The magnitude of HAIs related to care provided in other settings, such as ambulatory surgery and endoscopy centers, has not been as thoroughly studied, but such infections have been well described.

Pathogenesis

HAIs can be caused by organisms that are a part of the patient's normal flora (i.e., endogenous infection) or by pathogens acquired during exposure to health care (i.e., exogenous infection) through the contaminated hands of health care workers, the environment, contaminated medical equipment, other patients, or visitors. A variety of factors can contribute to the development of an HAI, and in many cases, HAIs are multifactorial in nature. These factors can be related to the pathogen, the host, the specific health care interventions that a patient receives, the setting in which health care is received, and the methods by which these interventions are made. HAI prevention strategies focus on eliminating, reducing, or modifying one or more of these risk factors.

Pathogen-Related Factors

A variety of pathogen-related factors contribute to the ability of an organism to cause infection. These factors include the organism's normal reservoir, mode of transmission (e.g., direct or indirect contact transmission, respiratory droplets, airborne particles), ability to survive on inanimate objects and surfaces, ability to produce biofilm, virulence factors, and resistance to antimicrobial agents and, for some organisms (e.g., *Clostridium difficile*), disinfectants.

Host-Related Factors

Many host-specific factors are associated with an inherent increased risk of one or more types of infection, regardless of the receipt of health care; however, when a patient with one or more of these risk factors enters the health care system, these factors contribute to an increased risk of HAI. Such risk factors include age (with neonates and older adults having an increased risk of infection because of incomplete development or senescence of the immune system, respectively), obesity, smoking, severity of illness, and certain medical conditions (e.g., burns, end-stage liver or renal disease, poorly controlled diabetes, some cancers, congenital or acquired immune deficiency). These factors reflect suppression of the immune system or breaches of other normal host defense mechanisms. Whereas many of these factors are not amenable to intervention or cannot be effectively modified in the short term, interventions that address remediable risk factors (e.g., obesity, smoking, poorly controlled diabetes mellitus) have the potential to reduce the risk of HAI during future episodes of health care.

Health Care–Related Factors

Health care–related HAI risk factors are those resulting from interventions that are intended to treat or otherwise provide benefit for a patient's existing medical conditions but that also introduce an increase in the risk of infection. These factors may disrupt normal host defenses or alter the patient's normal microbiologic flora. Health care–related risk factors include the use of invasive devices (e.g., central venous catheters, urinary catheters, endotracheal tubes), surgical procedures, exposure to antibiotics, receipt of immunosuppressive medications, and prolonged hospitalization. Because each of these interventions poses at least some degree of increased risk of infection, the risk-to-benefit ratio of each intervention must frequently be reassessed so that patients are not exposed to unnecessary risk. For example, central venous catheters and indwelling urinary catheters are major risk factors for primary blood stream infection and urinary tract infection, respectively. In a patient who has a true medical need for one of these devices, the benefits of the catheter exceed the risk of infection. However, once the patient recovers from the condition that necessitated the catheter, the risks associated with the device then outweigh the benefits.

Exposure to antibiotics is a well-established risk factor for colonization and infection with multidrug-resistant organisms (MDROs) and development of *C. difficile* infection (CDI) through a mechanism known as antibiotic selection pressure. Antimicrobial use is common in acute care hospitals and other health care settings, such as long-term care and dialysis facilities and ambulatory care practices. A 2009 point prevalence survey of hospitals in 25 European countries found that 29% of hospitalized patients receive one or more antimicrobials during their hospital stay. Among 70 academic hospitals in the United States, 49.6 to 76% of hospitalized adults received at least one dose of an antibacterial drug, and mean total antimicrobial use was 839 days of therapy (range, 594 to 1109) per 1000 patient-days in 2009. Even more important, studies have shown that a large proportion of antimicrobial use is inappropriate. Inappropriate antimicrobial use includes administration of antimicrobial regimens that are broader in spectrum or longer in duration than necessary, use of antibiotics that do not provide activity against the causative pathogen, treatment for test results that do not reflect the presence of infection (e.g., specimen contamination, asymptomatic colonization), use of antibacterial agents for treatment of conditions that are not due to bacterial infection (e.g., viral respiratory tract infections), and prescription of inappropriate doses of an antibiotic. Other studies have found that approximately 30% of antimicrobial use in acute care hospitals is inappropriate. Similar proportions of inappropriate antimicrobial use have been reported among long-term care facility residents and chronic hemodialysis patients. Misuse and overuse of antimicrobial agents in the outpatient setting are also well-recognized problems in the United States and other countries, including many countries where antimicrobials can be obtained without a prescription. This inappropriate use of antimicrobial agents introduces unnecessary risk for the development of complications of antibiotic therapy, including CDI, MDRO infection, and toxicity, and represents an important target for intervention.

Health Care Delivery–Related Factors

This group of risk factors includes those that are introduced as a result of the way in which health care is delivered. These risk factors do not offer any potential benefit to patients but rather are associated only with risk. Health care delivery–associated risk factors include, among other things, failure to perform hand hygiene when indicated or to use aseptic or sterile technique during invasive procedures, unsafe injection practices (e.g., entering

TABLE 282-2 RATES OF ANTIMICROBIAL RESISTANCE AMONG PATHOGENIC ISOLATES FROM HEALTH CARE–ASSOCIATED INFECTIONS

ORGANISM	ANTIBIOTIC CLASS	PROPORTION OF ISOLATES RESISTANT TO ANTIBIOTIC		
		United States (2009-2010)[a]	*ICUs in 36 Developing Countries (2004-2009)*[b]	*ICUs in 13 European Countries (2007)*[c]
Staphylococcus aureus	Anti-staphylococcal penicillins (e.g., oxacillin, methicillin)	44-59%	71-84%	34.5%
Klebsiella species	Extended-spectrum cephalosporins	13-29%	72-76%	22%
	Carbapenems	8-13%	7-8%	2%
Pseudomonas aeruginosa	Carbapenems	11-30%	36-47%	38%
Enterococcus faecium	Glycopeptides (vancomycin)	62-83%	NR	2.7%
Acinetobacter baumannii	Carbapenems	37-74%	52-66%	73%
Escherichia coli	Extended-spectrum cephalosporins	11-19%	50-68%	11.7%
	Carbapenems	2-3%	4-6%	1.2%
	Fluoroquinolones	25-42%	32-55%	32%

ICUs, intensive care units; NR, not reported.
[a]Sievert DM, Ricks P, Edwards JR, et al. Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009-2010. *Infect Control Hosp Epidemiol.* 2013;34:1-14.
[b]Rosenthal VD, Bijie H, Maki DG, et al. International Nosocomial Infection Control Consortium (INICC) report, data summary of 36 countries, for 2004-2009. *Am J Infect Control.* 2012;40:396-407.
[c]European Centre for Disease Prevention and Control. Surveillance of healthcare-associated infections in Europe, 2007. Stockholm: ECDC; 2012. Available at: http://ecdc.europa.eu/en/publications/Publications/120215_SUR_HAI_2007.pdf.

a multidose vial with a used needle), and failure to adequately clean and disinfect or to sterilize the patient environment and medical equipment and instruments. These risks are all potentially modifiable and are thus important targets for HAI prevention initiatives. Antibiotic use, which has already been discussed as a patient-specific health care–related risk factor, can also be considered a health care delivery–related risk factor. Unlike with other types of drugs, use and misuse of antibiotics in one patient or population can introduce risks among the larger population through changes in microbial ecology (i.e., selection and increased prevalence of antimicrobial-resistant pathogens).

Many of these health care delivery–related factors are the result of poor adherence to recommended, evidence-based infection prevention practices. Despite recognition that poor hand hygiene practice is a leading cause of pathogen transmission, the existence of major national and international guidelines, and initiatives to improve hand hygiene practices among health care workers, compliance with recommended hand hygiene practices among health care personnel remains unacceptably low. In the United States, average rates of health care worker compliance with recommended hand hygiene practices have been reported to be less than 50%, with some individual studies reporting rates as low as 20% in some intensive care units (ICUs). Similarly, unsafe injection practices continue to be identified as the cause of health care–related transmission of blood-borne pathogens, such as hepatitis B and C viruses.

In recent years, there has been an increasing recognition of the role of environmental contamination in the transmission of health care–associated pathogens. Environmental contamination with these organisms is common, and many of these organisms can persist in the health care environment for prolonged periods. For example, environmental contamination with *C. difficile* has been detected in 100% of hospital rooms occupied by patients with CDI, whereas methicillin-resistant *Staphylococus aureus* (MRSA) has been detected on environmental surfaces in approximately 70% of hospital rooms housing patients infected and colonized with MRSA. Some but not all studies have identified similarly high rates of environmental contamination with multidrug-resistant gram-negative pathogens. This contamination can result in patient-to-patient transmission through transient contamination of health care workers' hands and equipment or by direct contact of the patient with the contaminated environment. Studies have shown that admission to a hospital room in which the prior occupant was colonized or infected with one of several MDROs is a significant risk factor for acquisition of that organism. Environmental contamination is, however, a potentially modifiable risk factor for HAI. Cleaning and disinfection of the environment and portable medical equipment that is shared among patients are often suboptimal. For example, one multicenter study conducted in 36 acute care hospitals in the United States found that at baseline, only 48% of high-risk environmental surfaces were cleaned during routine cleaning after discharge of the patient. Improvement in cleaning practices and other interventions to reduce the microbiologic burden in the health care environment has been shown to reduce the microbiologic burden of organisms in the environment and has been associated with a reduction in the risk of acquisition of MDROs and CDI. In addition to cleaning and disinfection of environmental surfaces and disinfection or sterilization of shared medical equipment, environmental infection control interventions are important for preventing patients from acquiring pathogens due to exposure to water (e.g., *Legionella* species) and air (e.g., environmental fungi) within the health care setting.

Pathogens in Health Care–Associated Infections

The organisms most commonly identified in device- and procedure-associated infections (i.e., central line–associated blood stream infection, catheter-associated urinary tract infection, ventilator-associated pneumonia, and surgical site infection) vary somewhat among the different types and sites of infection. Overall, eight pathogen groups accounted for more than 80% of pathogens identified in device- and procedure-related infections reported to the CDC through the National Healthcare Safety Network in 2009 and 2010.[3] These pathogen groups and the proportion of reported pathogens that they represented include *Staphylococcus aureus* (16%), *Enterococcus* species (14%), *Escherichia coli* (12%), coagulase-negative staphylococci (11%), *Candida* species (9%), *Klebsiella* species (8.0%), *Pseudomonas aeruginosa* (8%), and *Enterobacter* species (5%). Whereas many of these pathogens represent patients' endogenous flora, further discussion of several of the organisms that may be acquired during exposure to health care is warranted.

Multidrug-Resistant Organisms

An increasingly concerning problem is the emergence of acquired antimicrobial resistance among many of the bacterial pathogens that are common causes of HAIs (Table 282-2). MDROs represent a significant health threat because infections caused by many of these MDROs have been associated with worse outcomes than those caused by antimicrobial-susceptible strains of the same organism, including excess length of hospital stay, increased health care costs, and higher mortality, with mortality rates approaching 50% in some studies. Possible explanations for the observed increased rate of adverse outcomes associated with MDRO infections include the presence of more severe underlying disease, delays in initiating effective therapy, and the use of more toxic or less effective therapy for treatment of the infection. Regardless of their cause, the poor outcomes associated with MDRO infections highlight the critical need for effective preventive measures and development of new antibiotics with activity against these MDROs, particularly multidrug-resistant gram-negative bacilli (MDR-GNB).

Organisms can develop resistance to antimicrobial agents to which they were previously susceptible through a variety of mechanisms, including induction, genetic mutation, and acquisition of new genetic material (e.g., conjugation with cell-to-cell transfer of genetic material by plasmids or transposons). In the health care setting, however, patient-to-patient transmission of MDROs is more common than de novo development of resistance in a previously susceptible organism within the patient's existing flora. Identified risk factors for acquisition of MDROs include exposure to antibiotics, frequent or prolonged exposure to health care facilities (e.g., hospitals, nursing homes), poor infection control practices among health care workers,

environmental contamination with MDROs, and prevalence of MDROs among other patients within a health care facility.

Among the pathogens reported to the CDC between 2009 and 2010 as causes of device-associated infections and surgical site infections, approximately 20% had antimicrobial susceptibility profiles that met the CDC's definition of multidrug resistance (see Table 282-2). Of note, 62 to 83% of *Enterococcus faecium* isolates were vancomycin resistant (VRE), and 44 to 59% of *S. aureus* isolates were resistant to methicillin (MRSA). These multidrug-resistant gram-positive pathogens have been recognized as significant health care–associated pathogens for several decades. More recently, the emergence of multidrug resistance among several gram-negative pathogens has been identified as a growing global health threat among persons receiving health care. For example, 65% of *Acinetobacter baumannii* isolates reported to the CDC in 2009 and 2010 demonstrated acquired resistance to at least one drug in three or more antibiotic classes. Such multidrug resistance definitions were also met by 15% of *Klebsiella* isolates and 14% of *P. aeruginosa* isolates. Antimicrobial resistance is an important problem in many regions of the world. For example, International Nosocomial Infection Control Consortium data from 36 developing countries in Asia, Africa, Europe, and Latin America collected between 2004 and 2009 demonstrated methicillin resistance in 84% of *S. aureus* isolates and extended-spectrum cephalosporin resistance in 76.3% of *Klebsiella pneumoniae* isolates and 66.7% of *E. coli* isolates.[4] The lack of use of a standardized definition of multidrug resistance limits direct comparisons of resistance data from different populations. Standardized definitions for multidrug resistant, extensively drug resistant, and pandrug resistant have recently been proposed.[5] Adoption of these or other standardized definitions is needed to allow a more thorough understanding of the global burden of antimicrobial resistance among health care–associated pathogens.

One example of the emergence and rapid dissemination of MDR-GNB is carbapenem-resistant Enterobacteriaceae (CRE), particularly *K. pneumoniae* (Chapter 305). Carbapenem resistance among these organisms was rare in the United States before the year 2000, at which time less than 1% of *K. pneumoniae* isolates reported to the CDC demonstrated such resistance. By 2009-2010, 8 to 13% of *K. pneumoniae* isolates from hospital-associated infections were resistant to carbapenems. In the United States, carbapenem resistance among the Enterobacteriaceae is most commonly due to the production of *K. pneumoniae* carbapenemase (KPC), a class A serine β-lactamase enzyme that hydrolyzes all β-lactam antibiotics. The KPC enzyme, which was first described in 2001, is encoded by the bla_{KPC} gene carried on a transmissible plasmid that also carries additional genes that confer resistance to several other classes of antimicrobial agents. Thus, in addition to carbapenem resistance, these organisms demonstrate resistance to other β-lactam antibiotics and to several other classes of antibiotics. Other mechanisms of carbapenem resistance among the Enterobacteriaceae, including other carbapenemases such as New Delhi metallo-β-lactamase-1 (NDM-1), OXA, VIM, and IMP, contribute to the growing threat of CRE around the world.

Asymptomatic carriage of MDROs is relatively common among persons with health care exposures. In fact, patients with clinically apparent MDRO infections represent a relatively small proportion of the total burden of these pathogens. Reported rates of MRSA carriage have ranged from 4.6 to 13.6% among hospital patients, 2 to 22% among U.S. ambulatory dialysis patients, and 10 to 100% among residents of long-term care facilities. The prevalence of VRE carriage among hospital patients and ambulatory dialysis patients has been reported to range from 6.3 to 67% and 0 to 16%, respectively. It is more difficult to describe the prevalence of MDR-GNB carriage because of use of different definitions of multidrug resistance and the inclusion of multiple and different organisms (e.g., *P. aeruginosa*, *A. baumannii*, *K. pneumoniae*) among reported studies. Studies that have included a variety of MDR-GNB have reported carriage rates of 19 to 32% in ICU patients and hospital patients with diarrhea, 25% among residents of long-term care facilities, and 16% among chronic hemodialysis patients. Studies that have focused specifically on carbapenem-resistant Enterobacteriaceae have demonstrated prevalence rates ranging from 2 to 5.4% among high-risk hospital patients in the United States and 2 to 49% among post–acute care facility patients in Israel. These asymptomatic carriers play an important role in the epidemiology of MDRO infections. First, they are at substantial risk of subsequent infection with the colonizing organism, with up to one third of carriers of MRSA, VRE, and MDR-GNB developing symptomatic infection within 12 months. Second, asymptomatic MDRO carriers can contribute to MDRO transmission within the health care system through contamination of their surrounding environment and of health care workers' hands, clothing, and medical equipment. In fact, several studies have demonstrated that the risk of acquiring an MDRO, such as MRSA, VRE, and MDR-GNB, during hospitalization is related to the prevalence, or colonization pressure, of that MDRO among other patients. Finally, admission to a hospital room in which the prior patient was colonized or infected with the MRSA, VRE, or MDR-GNB has been associated with an increased risk of acquisition of those organisms.

There are, however, some encouraging data related to the incidence of some MDROs, particularly MRSA. In the United States, the incidence of hospital-onset invasive MRSA infections decreased 9.4% per year between 2005 and 2008.[6] During a similar period (2005 to 2011), the incidence of invasive MRSA infections among dialysis patients, a group with a rate of invasive MRSA infection that is approximately 100 times greater than that of the general population, was also observed to have significantly decreased in the United States, with annual decreases of 6.7% for health care–associated community-onset cases and 10.5% for hospital-onset cases.[7] These changes occurred despite the emergence of community-associated MRSA as a significant cause of skin and soft tissue infections among persons without typical health care–associated risk factors and the introduction of community-associated MRSA as a health care–acquired pathogen. In England, a 56% reduction in the number of cases of MRSA bacteremia reported through a mandatory reporting system was observed between 2004 and 2008. The specific cause of these observed decreases in MRSA HAIs is uncertain and may be the result of improvements in basic infection control practices, introduction of specific MRSA prevention practices, or other changes in the epidemiology of this pathogen.

Clostridium Difficile

C. difficile, the etiologic agent of pseudomembranous colitis, is the most common cause of health care–associated infectious diarrhea (Chapter 296). Although community-associated CDI occurs, most cases are associated with receipt of health care. Analysis of CDC data from 2010 found that 94% of cases of CDI were associated with health care exposure but that 75% of cases had their onset outside of the hospital (e.g., in the community or in long-term care facilities). Clinical manifestations of CDI range from asymptomatic carriage to mild diarrhea to life-threatening colitis, toxic megacolon, and sepsis. The overall mortality associated with CDI has been reported to range from 2 to 6%, with substantially higher mortality among patients who develop toxic megacolon and other severe manifestations of the disease. Compared with patients without CDI, hospital patients who develop CDI experience an extended duration of hospitalization of 2.8 to 6.4 days, and approximately 15 to 30% of these patients will experience at least one recurrence of the disease, typically within 1 to 2 months of the initial episode. Persons who experience one recurrence have a 50 to 60% chance of additional recurrences. U.S. data from 2009 indicate that 4.8% and 12.8% of hospital patients with CDI are readmitted to the hospital with a principal diagnosis or any diagnosis of CDI, respectively, within 30 days and that by 90 days, rates of hospital readmission increase to 6.8% and 17.2%, respectively. Estimates of CDI-attributable hospital costs range from $5682 to $9124 per case (in 2007 U.S. dollars), amounting to a total cost of $1 billion to $4.8 billion per year.

C. difficile has become one of the most common health care–associated pathogens, and several studies have suggested that more HAIs are now due to *C. difficile* than to MRSA. The incidence of CDI among hospitalized patients in the United States more than doubled between 2000 and 2009. Reported rates of hospital-onset CDI in the United States and Europe are 7.4 and 4.1 per 10,000 patient-days, respectively. In addition to an increased incidence of CDI, the rate of *C. difficile*–related death increased more than four-fold in the United States during the past decade, causing approximately 14,000 deaths per year. The observed increases in the incidence of and mortality associated with CDI were temporally associated with the emergence and dissemination of the 027/NAP1/BI strain of *C. difficile*, which has been associated with greater mortality and higher rates of recurrence than other circulating strains. Initially identified in North America, this epidemic strain has now disseminated globally. Data from England show a 50% reduction in CDI between the fourth quarter of 2009 and the first quarter of 2013 after the introduction of a national CDI surveillance and prevention program. In the United States, however, *C. difficile* remains a challenge, and reductions like those reported in the United Kingdom have not been observed. In fact, between 2008 and 2012, *C. difficile* hospitalizations increased by 11.2%, and from 2010 to 2012, the incidence of hospital-onset CDI increased by 28%.

The problem of *C. difficile* extends beyond the acute care hospital setting. For instance, data from Ohio indicate that 62% of CDI cases in 2006 occurred in nursing homes. Another study conducted in Montgomery County, New York, found that 52% of cases of CDI in long-term care facility residents developed within 4 weeks of discharge from an acute care hospitalization.[8] Data from the CDC's Emerging Infections Program showed that 20% of

hospital-onset CDI occurred among persons with recent (i.e., <12 weeks) residence in a nursing home and that 67% of nursing home–onset CDI cases occurred among residents who had been recently discharged from an acute care hospital. These latter findings demonstrate the complex epidemiology of CDI.

The development of CDI is a two-step process that first requires acquisition of *C. difficile* through fecal-oral transmission and then the presence or introduction of factors that allow progression to symptomatic disease. Although some healthy individuals without health care exposure are intestinal carriers of *C. difficile*, acquisition of the organism during contact with the health care system plays a critical role in the epidemiology of CDI. Transmission in the health care setting can occur as the result of exposure to organisms on health care workers' hands, environmental surfaces, or medical equipment. Hand contamination of health care workers is common after contact with the skin or environment of persons with CDI. For example, a prospective study of 30 CDI patients found that hand contamination occurred in 50% of health care workers after contact with patients' skin and in 50% of health care workers after contact with the patient's immediate environment, particularly the bed rails.[9] In the absence of effective hand hygiene practices, contaminated health care workers may transmit the organism to other patients. Contamination of the environment and medical equipment with *C. difficile* is also common. Testing performed in hospital rooms has identified *C. difficile* on environmental surfaces in up to 100% of rooms housing patients with active CDI and in as many as 33% of non-CDI patient rooms. Commonly contaminated surfaces include bed rails, bedside tables, telephones, call buttons, and blood pressure cuffs. *C. difficile* spores are resistant to killing by many common hospital disinfectants and can persist in the environment for long times, further contributing to the risk of exposure to a contaminated environment. Patients who are admitted to a hospital room in which the previous occupant had CDI have been found to be at greater risk for development of CDI than patients admitted to hospital rooms in which the previous occupant did not have CDI. The risk for development of CDI has also been demonstrated to be directly related to the prevalence, or colonization pressure, of *C. difficile* among other patients in a health care facility. This likely reflects an increased risk of exposure to the organism through contaminated hands of health care workers, environment, or medical equipment as the number of persons with CDI increases.

Once *C. difficile* has been ingested, several factors are known to be associated with development of symptomatic CDI. The major risk factor is receipt of antibiotics. Such exposures disrupt the normal intestinal flora and allow *C. difficile* to multiply to larger numbers and to produce toxins that result in disease. Although exposure to any antimicrobial agent may increase the risk for development of CDI, clindamycin, third-generation cephalosporins, penicillins, and fluoroquinolones may present the highest risk. Other factors that have been associated with an increased risk for development of CDI include receipt of cytotoxic chemotherapeutic agents and gastric acid suppressive medications such as proton pump inhibitors, failure to develop an antibody response to *C. difficile*, and older age (i.e., age older than 64 years).

Viruses

Respiratory Viruses

Common respiratory viruses, such as influenza, can be transmitted in the health care setting by health care workers, visitors, and patients, resulting in health care–acquired disease. Higher rates of morbidity and mortality have been observed among those who acquire infection during hospitalization, probably due to the presence of significant underlying medical illness. Despite several studies that have associated higher influenza immunization rates of health care workers with lower rates of nosocomial influenza transmission, the uptake of influenza vaccination among health care workers remains relatively low. This has led many public health agencies and professional societies to call for mandatory influenza vaccination policies for all eligible health care workers. The health care–associated transmission of newly emerged respiratory viruses, such as the severe acute respiratory syndrome coronavirus (SARS-CoV) in 2003, pandemic influenza in 2009, and the Middle East respiratory syndrome coronavirus (MERS-CoV) in 2013, highlights the importance of syndromic surveillance to allow rapid identification of patients with potentially communicable diseases and implementation of appropriate infection control precautions.

Blood-borne Viruses

Although routine screening of the blood supply for the blood-borne pathogens (BBPs) hepatitis B virus, hepatitis C virus, and human immunodeficiency virus has dramatically decreased the incidence of health care–associated BBP infections, transmission of these pathogens within health care settings continues to occur. Most health care–associated BBP transmission that occurs now is due to failure to adhere to recommended basic infection control practices. Unsafe injection practices (e.g., reuse of syringes, contamination of multidose vials, improper use and disinfection of blood glucose monitoring devices that are used for multiple patients) and inadequate cleaning, disinfection, and sterilization of medical equipment and the health care environment (e.g., dialysis facilities) have been identified in several recent outbreaks of patient-to-patient transmission of BBPs. Between 2001 and 2011, an estimated 130,198 patients from 17 U.S. states were notified of possible health care–associated exposures to BBPs as the result of 35 reported unsafe injection practice events.[10] A large proportion of documented health care–associated hepatitis B virus and hepatitis C virus transmission events has occurred in outpatient settings and long-term care facilities, highlighting the importance of infection prevention programs throughout the entire health care system. Transmission of BBPs from health care worker to patient is uncommon but can occur, typically in the setting of "exposure-prone" invasive procedures. Guidelines are available to assist health care workers and health care facilities in minimizing the risk posed to patients by a BBP-infected health care worker while allowing most such health care workers to remain involved in patient care activities (E-Table 282-1).

Fungi

Candida albicans and other *Candida* species accounted for approximately 9.5% of all pathogens reported to the CDC between 2009 and 2010 as causes of device-associated infections and surgical site infections, placing them among the most common pathogens implicated in HAIs. *Candida* species were particularly common causes of catheter-associated urinary tract infection and catheter-associated blood stream infection, accounting for 12.7% and 14.6% of pathogens reported among these types of infection, respectively. Exposure to environmental fungi, such as *Aspergillus* species, in the health care setting can result in HAI, particularly in immunocompromised hosts. Such exposure and resulting infection are most commonly associated with inadequate environmental control measures during construction, demolition, or water damage within the health care facility. A recent multistate outbreak of invasive fungal infections in the United States, mostly due to *Exserohilum rostratum*, associated with contaminated methylprednisolone injections demonstrates that contaminated medications and other medical products are additional potential sources of exposure to fungal pathogens during health care.

Device-Associated Infections

Central Line–Associated Blood Stream Infections

Central line–associated blood stream infections (CLABSIs) are blood stream infections that occur in patients with a central venous catheter and in whom there is no other identified source of the infection. Development of CLABSI has been associated with longer hospital stays, increased risk of death, and greater hospital costs than those observed among otherwise similar patients who do not develop CLABSI (see Table 282-1). Rates of morbidity and mortality, however, vary substantially, depending on the causative pathogen and characteristics of the patient in whom the infection occurs. The CLABSI definition was developed for epidemiologic surveillance purposes rather than for medical decision making for individual patients. It differs in definition from that of a catheter-related blood stream infection, a term that may be used for research or clinical purposes, in which more specific testing, such as catheter segment cultures, quantitative blood cultures, or differential time to positivity of paired blood cultures, is done to definitively determine that the catheter is the true source of a blood stream infection. Although CLABSIs are often thought of as a complication that occurs among ICU patients, the use of central venous catheters in non-ICU hospital wards has expanded substantially during the past few decades. Thus, the incidence of and the overall burden of CLABSI in some non-ICU wards now often exceeds that in ICUs. CLABSIs are also important problems among nonhospitalized persons with central venous catheters, such as persons receiving chronic total parenteral nutrition, chemotherapy, or dialysis in the outpatient or home care setting. In 2008, for example, approximately 37,000 CLABSIs occurred among outpatient hemodialysis patients in the United States.

Blood stream infections due to central venous catheters are largely the result of contamination or colonization of the external surface or the intraluminal surface of the catheter. This contamination can occur either during catheter insertion or after insertion, related to a number of aspects of catheter use and care. Research has identified a number of effective strategies to reduce the risk of catheter contamination during insertion and throughout the time that the catheter remains in situ. This research has led to the development

E-TABLE 282-1 SELECT GUIDELINES AND EVIDENCE-BASED RECOMMENDATIONS FOR THE PREVENTION AND MANAGEMENT OF HEALTH CARE–ASSOCIATED INFECTIONS

GENERAL TOPIC	SPECIFIC TOPIC	GUIDELINES AND RECOMMENDATIONS
Device-associated infections	Vascular catheter–associated infections	Marschall J, Mermel LA, Fakih M, et al. Strategies to prevent central line-associated bloodstream infections in acute care hospitals: 2014 update. *Infect Control Hosp Epidemiol*. 2014;35:753-771. O'Grady NP, Alexander M, Burns LA, et al. Guidelines for the prevention of intravascular catheter–related infections. *Clin Infect Dis*. 2011;52:1087-1099. Mermel LA, Allon M, Bouza E, et al. Clinical practice guidelines for the diagnosis and management of intravascular catheter–related infection: 2009 update by the Infectious Diseases Society of America. *Clin Infect Dis*. 2009;49:1-45.
	Catheter-associated urinary tract infections	Lo E, Nicolle LE, Coffin SE, et al. Strategies to prevent catheter-associated urinary tract infections in acute care hospitals: 2014 update. *Infect Control Hosp Epidemiol*. 2014;35:464-479. Gould CV, Umscheid CA, Agarwal RK, et al. *Guideline for Prevention of Catheter-Associated Urinary Tract Infections 2009*. http://www.cdc.gov/hicpac/pdf/CAUTI/CAUTIguideline2009final.pdf Hooton TM, Bradley SF, Cardenas DD, et al. Diagnosis, prevention, and treatment of catheter-associated urinary tract infection in adults: 2009 International Clinical Practice Guidelines from the Infectious Diseases Society of America. *Clin Infect Dis*. 2010;50:625-663.
	Ventilator-associated pneumonia/health care–associated pneumonia	Centers for Disease Control and Prevention. Guideline for preventing health care–associated pneumonia, 2003. *MMWR Recomm Rep*. 2004;53(RR-3):1-36. Klompas M, Branson R, Eichenwald EC, et al. Strategies to prevent ventilator-associated pneumonia in acute care hospitals: 2014 update. *Infect Control Hosp Epidemiol*. 2014;35:915-936.
	Cardiovascular implantable electronic devices	Baddour LM, Epstein AE, Erickson CC, et al. Update on cardiovascular implantable electronic device infections and their management: a scientific statement from the American Heart Association. *Circulation*. 2010;121:458-477.
Surgical site infections		Mangram AJ, Horan TC, Pearson ML, et al. Guideline for the prevention of surgical site infection, 1999. *Infect Control Hosp Epidemiol*. 1999;20:247-278. Anderson DJ, Podgorny K, Berrios-Torres SI, et al. Strategies to prevent surgical site infections in acute care hospitals: 2014 update. *Infect Control Hosp Epidemiol*. 2014;35:605-627. Bratzler DW, Dellinger EP, Olsen KM, et al. Clinical practice guidelines for antimicrobial prophylaxis in surgery. *Am J Health Syst Pharm*. 2013;70:195-283.
Health care–associated pathogens	*Clostridium difficile*	Dubberke ER, Carling P, Carrico R, et al. Strategies to prevent Clostridium difficile infections in acute care hospitals: 2014 Update. *Infect Control Hosp Epidemiol*. 2014;35:628-645. Cohen SH, Gerding DN, Johnson S, et al. Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA). *Clin Infect Dis*. 2010;31:431-455.
	Multidrug-resistant organisms	Centers for Disease Control and Prevention. *Management of Multidrug-Resistant Organisms in Healthcare Settings*. http://www.cdc.gov/hicpac/mdro/mdro_toc.html
	Methicillin-resistant *Staphylococcus aureus*	Calfee DP, Salgado CD, Milstone AM, et al. Strategies to prevent methicillin-resistant Staphylococcus aureus transmission and infection in acute care hospitals: 2014 update. *Infect Control Hosp Epidemiol*. 2014;35:772-796. Liu C, Bayer A, Cosgrove SE, et al. Clinical practice guidelines by the Infectious Diseases Society of America for the treatment of methicillin-resistant *Staphylococcus aureus* infections in adults and children. *Clin Infect Dis*. 2011;52:1-38.
	Carbapenem-resistant Enterobacteriaceae	Centers for Disease Control and Prevention. Guidance for control of infections with carbapenem-resistant or carbapenemase-producing Enterobacteriaceae in acute care facilities. *MMWR Morb Mortal Wkly Rep*. 2009;58:256-260. Centers for Disease Control and Prevention. *2012 CRE toolkit: Guidance for Control of Carbapenem-resistant Enterobacteriaceae (CRE)*. http://www.cdc.gov/hai/organisms/cre/cre-toolkit/index.html
	Blood-borne pathogens	Henderson DK, Dembry L, Fishman NO, et al. SHEA guideline for management of healthcare workers who are infected with hepatitis B virus, hepatitis C virus, and/or human immunodeficiency virus. *Infect Control Hosp Epidemiol*. 2010;31:203-232. Centers for Disease Control and Prevention. Updated CDC recommendations for the management of hepatitis B virus–infected health-care providers and students. *MMWR Recomm Rep*. 2012;61:1-12.
	Mycobacterium tuberculosis	Centers for Disease Control and Prevention. Guidelines for preventing the transmission of *Mycobacterium tuberculosis* in health-care settings, 2005. *MMWR Recomm Rep*. 2005;54(RR-17):1-137.
Antimicrobial stewardship		Dellit TH, Owens RC, McGowan JE, et al. Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America guidelines for developing an institutional program to enhance antimicrobial stewardship. *Clin Infect Dis*. 2007;44:159-177.
General infection prevention		Centers for Disease Control and Prevention. *Guide to Infection Prevention in Outpatient Settings: Minimum Expectations for Safe Care*. 2011. http://www.cdc.gov/hai/settings/outpatient/outpatient-care-guidelines.html Centers for Disease Control and Prevention. Guideline for hand hygiene in health-care settings: recommendations of the Healthcare Infection Control Practices Advisory Committee and the HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. *MMWR Recomm Rep*. 2002;51(RR-16):1-45. World Health Organization. *WHO Guidelines on Hand Hygiene in Health Care*. 2009. http://whqlibdoc.who.int/publications/2009/9789241597906_eng.pdf Siegel JD, Rhinehart E, Jackson M, Chiarello L; Healthcare Infection Control Practices Advisory Committee. *2007 Guideline for Isolation Precautions: Preventing Transmission of Infectious Agents in Healthcare Settings*. 2007. http://www.cdc.gov/hicpac/2007IP/2007isolationPrecautions.html Centers for Disease Control and Prevention. Guidelines for environmental infection control in health-care facilities. *MMWR Recomm Rep*. 2003;52(RR-10):1-42. Rutala WA, Weber DJ; Healthcare Infection Control Practices Advisory Committee. *Guideline for Disinfection and Sterilization in Healthcare Facilities, 2008*. http://www.cdc.gov/hicpac/pdf/guidelines/Disinfection_Nov_2008.pdf

of evidence-based guidelines for prevention of vascular catheter–related infections (see E-Table 282-1). A review of the scientific literature concluded that 65 to 70% of CLABSIs are preventable with the use of currently available strategies and technologies.[11] The "central line bundle" refers to a small number of evidence-based practices that, when used together, can reduce the risk of CLABSI even more than would be expected when the components are introduced individually. The central line bundle includes hand hygiene, maximal barrier precautions during insertion (i.e., use of sterile gown and gloves and a surgical cap and mask by the operator and covering the patient with a sterile, full-body drape), chlorhexidine skin antisepsis, optimal insertion site selection (i.e., avoidance of the femoral site in adult patients), and daily review of catheter necessity with immediate removal of catheters that are no longer necessary. Other interventions that have been associated with reductions in CLABSI rates include the use of antimicrobial or antiseptic catheters, covering of the insertion site with a sterile gauze or semipermeable transparent dressing, cleansing of the catheter insertion site with an antiseptic such as chlorhexidine, use of aseptic technique to access and to manipulate the catheter, scrubbing of the catheter hub with a disinfectant before accessing the catheter lumen for administration of medications or other products or for aspiration of blood, use of antimicrobial- or antiseptic-coated catheters, and use of chlorhexidine rather than regular soap and water for daily bathing of ICU patients.

Widespread adoption of the central line bundle and other CLABSI prevention strategies has been associated with a substantial reduction in the incidence of CLABSI in U.S. hospitals. Data from the CDC demonstrate a 58% decrease in the incidence of CLABSI in ICUs between 2001 and 2009. More recent data from the CDC's National Healthcare Safety Network show a further 41% reduction in CLABSI in U.S. acute care hospitals between 2008 and 2011, with reductions noted in both ICU and non-ICU settings. In these most recent data, the incidence of CLABSI ranged from 0.9 to 3.7 per 1000 catheter-days in ICUs. The highest rate was observed in burn units.[12] Most other types of ICUs had rates ranging from 1.0 to 1.6 per 1000 catheter-days. In non-ICU wards, the highest CLABSI rates were reported from hematopoietic stem cell units and hematology-oncology wards.

Catheter-Associated Urinary Tract Infections

A catheter-associated urinary tract infection (CAUTI) is a urinary tract infection that develops in a patient who has or who recently had an indwelling urinary catheter. Similar to the pathogenesis of catheter-related blood stream infections, CAUTIs develop by the introduction of pathogens into the bladder as a result of contamination and colonization of the internal or external surface of the catheter. Among patients with indwelling urinary catheters, the incidence of bacteriuria is 3 to 8% per day, and 10 to 25% of those with bacteriuria will subsequently develop symptoms consistent with a urinary tract infection. Compared with the mortality and costs associated with many other HAIs, those associated with individual CAUTI events are substantially lower (see Table 282-1). CAUTI, however, occurs 5 to 10 times more frequently than CLABSI and ventilator-associated pneumonia, and thus the total burden of suffering and expense due to CAUTI is substantial. It has been estimated that 65 to 70% of CAUTIs that occur in acute care hospitals are preventable. Data reported to the CDC's National Healthcare Safety Network in 2011 show that the mean incidence of CAUTI in U.S. ICUs ranged from 1.2 to 4.5 per 1000 catheter-days, with substantial variation between different types of ICUs. In non-ICU inpatient wards, the mean rate ranged from 0.2 to 3.2, with the highest rates being observed on rehabilitation units. These data demonstrate a small but statistically significant 7% reduction in CAUTI between 2009 and 2011. The large reductions observed among many published studies of CAUTI prevention initiatives compared with the relatively small overall reduction observed in U.S. hospitals suggest that substantial opportunities remain for additional CAUTI prevention.

A number of basic practices, such as aseptic technique during insertion, maintenance of proper cleanliness and hygiene, securement of the catheter to avoid piston-like movement of the catheter within the urethra, and maintenance of a closed system with unobstructed flow of urine from the bladder into the collection system, are recommended to reduce the risk of CAUTI. Several studies have demonstrated that indwelling urethral catheters are often inserted for inappropriate reasons and that many catheters that were initially inserted for an appropriate indication remain in place even after the initial indication for catheterization has resolved. This may be due in part to lack of familiarity with appropriate indications for catheter insertion or with options that exist regarding alternatives to the use of indwelling urethral catheters. Additional studies have found that physicians are often unaware that their patient has a urinary catheter. Thus, perhaps the greatest opportunity for CAUTI prevention is avoidance of unnecessary catheter insertion and prompt removal of catheters that are no longer necessary. Development of protocols that explicitly define appropriate indications for insertion of urinary catheters, introduction of interventions that remind clinicians to reassess the appropriateness of a patient's urinary catheter, and nurse-driven protocols that allow nurses to remove unnecessary urinary catheters have been associated with reduced catheter use and lower rates of CAUTI.

Ventilator-Associated Pneumonia

In hospitalized patients, mechanical ventilation is one of the most common risk factors for the development of pneumonia. Mechanical ventilation and the interventions required to provide mechanical ventilation (e.g., endotracheal intubation, sedation) increase the risk of pulmonary infection through a variety of mechanisms, including an increased risk of aspiration of oropharyngeal and gastrointestinal secretions and impairment of the cough reflex. From both a clinical and epidemiologic standpoint, ventilator-associated pneumonia (VAP) is a difficult diagnosis to establish with certainty because of the subjective nature of many of the variables considered (e.g., chest radiograph findings, changes in the characteristics of respiratory tract secretions), alternative explanations for clinical and radiographic abnormalities (e.g., acute respiratory distress syndrome, atelectasis), and difficulty in determining whether the results of respiratory tract cultures represent true infection or colonization of the airway.

In 2011, the incidence of VAP in U.S. hospital ICUs ranged from 0.2 to 3.2 per 1000 ventilator-days, with the highest rates observed in trauma and burn units.[12] These infections are associated with mortality rates and health care costs that are among the highest observed among all HAIs (see Table 282-1). Studies suggest that at least 55% of VAP cases are preventable. As with other device-associated infections, avoiding the use of the device is the most effective means of preventing infection. For preventing VAP and other complications of mechanical ventilation, the use of noninvasive methods of ventilation, daily sedation interruption and assessment of readiness to be weaned, and early mobilization can eliminate or at least reduce the duration of mechanical ventilation. For patients who do require mechanical ventilation, the following are routinely recommended for VAP prevention: proper cleaning, disinfection, sterilization, use, and maintenance of respiratory equipment; elevation of the head of the bed (unless it is medically contraindicated); and routine oral hygiene care, typically with an antiseptic such as chlorhexidine. In some randomized, controlled trials, selective oropharyngeal or digestive decontamination has been associated with a significant reduction in VAP, but this approach has not yet been widely adopted as a standard of care in the United States, at least in part because of concerns that this intervention may lead to selection of antimicrobial-resistant organisms.

Other Device-Related Infections

With the advances that have occurred in medical technology in recent years, there have been substantial improvements in the capabilities of existing medical devices and the development of new implantable devices to treat and to manage a variety of medical conditions, particularly of the cardiovascular and nervous systems. These devices include cardiovascular implantable electronic devices such as pacemakers and implantable cardioverter-defibrillators, ventricular assist devices, deep brain stimulators, and intrathecal pumps. The potential benefits that these devices may offer to patients come with at least some degree of risk of device-related infection.

Surgical Site Infections

Surgical site infections (SSIs) are infections that develop at the site of an operative procedure. Although only a relatively small proportion of patients who undergo surgery subsequently develop SSI and the overall mortality rate associated with SSI is relatively low, the absolute number of SSIs that occur and the overall cost and burden of morbidity and mortality associated with SSI are large because of the volume of surgical procedures performed each year (see Table 282-1). A number of factors contribute to the risk of SSI: the specific site and type of surgical procedure; duration of the procedure; tissue hypoxia at the surgical site; wound contamination with endogenous or exogenous organisms; surgical technique; perioperative infection prevention practices; environmental controls related to temperature, humidity, and air purity within the operating room; and the patient's underlying medical conditions, smoking status, and other factors that may increase the inherent risk of infection. Although some of these factors are not amenable to corrective intervention, many of them are, and a number of interventions have been proved to reduce the risk of SSI. In fact, through routine application of evidence-based

preventive measures, it has been estimated that 55% of SSIs could be prevented. The use of sterile technique, preoperative skin preparation with an antiseptic agent, administration of antimicrobial prophylaxis within 60 minutes before the surgical incision, and perioperative glucose control are practices that have been demonstrated to reduce the risk of infection associated with a wide variety of surgical procedures. There is also evidence that at least in some types of surgeries, maintenance of normothermia and the use of supplemental oxygen (e.g., 80% fraction of inspired oxygen) in the perioperative period may reduce the risk of SSI. Although substantial opportunities for improvement remain, a report from the CDC's National Healthcare Safety Network that compares SSI data from 2008 and 2011 shows during that 4-year period a statistically significant 17% reduction in the incidence of SSI occurring in association with several common surgical procedures.[13]

HEALTH CARE–ASSOCIATED INFECTION PREVENTION STRATEGIES

As noted in the preceding paragraphs, many HAIs are preventable. This has been recognized since at least as early as the mid-1800s, when Semmelweiss demonstrated a dramatic reduction in sepsis-related deaths among maternity ward patients after an intervention to improve hand hygiene among health care workers. The degree to which HAIs are preventable was more clearly quantified in a multicenter study conducted in the 1970s. In that study, it was found that 32% of HAIs could be prevented through establishment of an effective infection control program. More recent estimates that take into account newer research and the technology that has been developed during the following three decades suggest that at least 55 to 70% of some of the most common HAIs can be prevented. Scientific research and clinical experience have led to the publication of evidence-based guidelines for the prevention of HAIs (see E-Table 282-1).[14]

Strategies to prevent HAI are designed to interrupt one or more of the factors associated with pathogen transmission or the development of infection due to endogenous or exogenous organisms. HAI prevention strategies are sometimes described by one of two categorization schemes that classify the intervention on the basis of either the applicability of the intervention for the prevention of a variety of HAIs or the priority an intervention should be given with regard to the overall HAI prevention program within a health care facility. The first categorization scheme classifies interventions as either "horizontal" or "vertical" infection prevention strategies. Horizontal prevention strategies are those that have the potential to prevent a wide variety of HAIs or to prevent transmission of a variety of pathogens. Vertical strategies, on the other hand, are those that target a specific pathogen. Examples of horizontal prevention strategies include hand hygiene, education of health care workers, environmental cleaning and disinfection, antimicrobial stewardship, and use of aseptic technique for invasive procedures. Vertical or pathogen-specific prevention strategies include the use of transmission-based precautions that are applied to patients known or suspected to be colonized or infected with a specific pathogen to prevent transmission of that pathogen. Preventing its transmission could involve interrupting its normal routes of transmission (e.g., contact, droplet, airborne), active surveillance testing for the pathogen (e.g., MRSA) to detect asymptomatic carriers who may serve as unrecognized reservoirs for its transmission, targeted decolonization therapy for persons identified as carriers of the pathogen of interest (e.g., MRSA), and vaccination programs (e.g., health care worker influenza vaccination campaigns).

The second categorization scheme classifies interventions as either basic or advanced (or special) preventive practices. Basic practices are those that should routinely be implemented in all health care facilities. Basic practices for acute care hospitals include most of the horizontal approaches described before as well as a few vertical approaches, such as transmission-based precautions. Special or advanced interventions are additional interventions that a facility may implement when basic practices fail to adequately control or prevent HAIs or transmission of one or more specific pathogens within the facility. This approach is consistent with recommendations and guidelines from the CDC and several professional societies that promote a tiered approach to the implementation of various HAI prevention strategies (see E-Table 282-1). The selection of specific strategies to be implemented within a specific health care facility should be based on an assessment of risk, local data and epidemiology, scientific evidence and guidelines, regulatory and accreditation requirements, and facility-specific cost-benefit calculations. Many of the prevention strategies that have been proven effective for prevention of device-associated infections and SSIs have already been discussed. Several additional preventive strategies are discussed in more detail here.

Antimicrobial Stewardship

Antimicrobial stewardship refers to interventions designed to improve the appropriateness of antimicrobial use by promoting the selection of an optimal antimicrobial regimen (i.e., the most appropriate drug, dose, duration, and route of administration) for a specific patient. The goals of an antimicrobial stewardship program are to optimize clinical outcomes (e.g., cure of infection related to antimicrobial use, minimize toxicity and other adverse events) and to limit the antimicrobial selection pressure that drives the emergence of antimicrobial-resistant strains. Multidisciplinary antimicrobial stewardship programs have been associated with several desirable outcomes, including significant reductions in antimicrobial use, reduced rates of antimicrobial resistance among health care–associated pathogens, reduced incidence of adverse outcomes associated with antibiotic use (e.g., toxicity, *C. difficile* infection), and significant reductions in hospital antimicrobial-associated costs. Thus, antimicrobial stewardship can be considered a horizontal infection prevention strategy.

A variety of approaches have been used by successful antimicrobial stewardship programs, and guidelines describing these strategies have been published (see E-Table 282-1). One of the most commonly used and most effective strategies includes formulary restriction and requirement of preauthorization before prescribing of certain antibiotics (e.g., antibiotics that are broad in their antimicrobial spectrum, are associated with significant toxicity, or are expensive). A second approach that has been considered to be a core strategy for antimicrobial stewardship activities is prospective audit of the appropriateness of prescribed antimicrobial therapy with provision of feedback to the prescribing clinician if opportunities for further optimization of therapy are available (e.g., narrowing or broadening spectrum of therapy, discontinuing antimicrobial therapy, or altering drug dose or dosing interval on the basis of available clinical data). Additional approaches that have been included in successful antimicrobial stewardship programs include education, development of guidelines and clinical pathways, computer-assisted decision support, and protocols to optimize conversion from the parenteral to the oral route of administration when appropriate.

Despite a substantial amount of data indicating that antimicrobial stewardship activities are an effective means of reducing inappropriate antimicrobial use and the clinical and financial consequences of excessive antimicrobial use, many acute care hospitals do not yet have formal antimicrobial stewardship programs in place, and many of those that do are not adequately resourced to reach their full potential. In addition to the standard challenges associated with implementation of interventions that require a change in human behavior and clinical practice, antimicrobial stewardship programs must also address the complex and constantly changing problems and issues associated with antimicrobial resistance. Antimicrobial stewardship programs are most commonly found in academic hospitals and large community hospitals, but there is a recognized need for development of such programs in smaller community hospitals, long-term care facilities, dialysis facilities, and outpatient practices.

Decolonization Therapy

Decolonization refers to the administration or application of antimicrobial or antiseptic agents to a person to eliminate or to reduce the burden of carriage of one or more pathogens. Perhaps most commonly recognized as a vertical measure that has been used in some MRSA control programs, decolonization therapy has also been used as a horizontal preventive measure. For example, selective oropharyngeal or digestive decontamination has been used for prevention of VAP and SSI after colorectal surgery. More recently, topical decolonization has been studied as a horizontal intervention for prevention of a variety of HAIs and prevention of transmission of a variety of pathogens. Several quasi-experimental, before-after studies have associated the use of chlorhexidine, compared with nonantimicrobial soap, for daily bathing of patients with significant reductions in rates of blood stream infections, including CLABSI, acquisition of MDROs, blood culture contamination, and contamination of the environment and health care workers. Thus far, two higher quality studies of daily chlorhexidine bathing have been completed and published. A multicenter, cluster-randomized trial conducted in eight adult ICUs and one bone marrow transplant unit in the United States found that daily bathing with chlorhexidine, compared with nonmedicated soap, was associated with a 23% reduction in the combined outcome of MRSA and VRE acquisition. A similar study in pediatric ICUs showed a significant reduction in the incidence of bacteremia in the per-protocol analysis, although the difference observed in the intention-to-treat analysis did not

reach statistical significance.[A1] A third cluster-randomized trial conducted in 74 adult ICUs in the United States found that providing all ICU patients with decolonization therapy that consisted of intranasal application of mupirocin for 5 days and daily chlorhexidine bathing significantly reduced the MRSA-positive clinical cultures attributable to the ICU by 37% and was associated with a lower incidence of all-cause blood stream infections compared with the use of active surveillance and contact precautions for MRSA-colonized patients without decolonization therapy.[A2] Selective decontamination of the digestive tract reduces ICU-acquired bacteremia even more than does selective oropharyngeal decontamination, but it also increases carriage of aminoglycoside-resistant gram-negative bacteria.[A4] Most of the data supporting the use of universal decolonization come from studies conducted in ICUs, and the role of this intervention in preventing infection and pathogen transmission in other settings is not well established.

Active Surveillance Testing

Active surveillance testing is a vertical intervention that identifies asymptomatic carriers of a pathogen of interest (e.g., MRSA, VRE, MDR-GNB) with the intention to introduce additional interventions for identified carriers to prevent infection in the carrier or transmission to others. The interventions that may be applied to the identified carriers include transmission-based precautions (e.g., contact precautions), decolonization therapy (mostly applicable to *S. aureus*), and altered antimicrobial therapy (e.g., surgical antimicrobial prophylaxis). The role of active surveillance has long been the subject of debate and investigation. It is commonly used in outbreak control efforts in conjunction with other interventions. In the non-outbreak setting, there are numerous reports of use of active surveillance as part of a comprehensive program to reduce transmission of or infection with MDROs in individual hospitals, large hospital systems, and health care facilities within specific geographic regions, such as a number of northern European countries. A cluster-randomized trial conducted in U.S. ICUs found that there was not a significant difference in the incidence of colonization or infection with MRSA and VRE in ICUs that performed active surveillance testing with introduction of contact precautions for patients found to be carriers of MRSA or VRE compared with ICUs that did not conduct active surveillance testing.[A3] However, there was a delay in reporting the results of surveillance testing that was longer than would be anticipated in normal clinical practice, which may limit the ability to generalize the study findings to all settings. A comparative effectiveness review of MRSA screening strategies concluded that the strength of evidence for universal screening to prevent MRSA HAIs was low and that there was insufficient evidence to assess other outcomes associated with universal screening or to assess the comparative effectiveness of other MRSA screening strategies (e.g., targeted screening).[15]

As previously mentioned in the discussion of decolonization therapy, a more recently published cluster-randomized trial found that providing universal decolonization to all ICU patients without the use of active surveillance testing was associated with a significantly greater reduction in MRSA clinical cultures attributable to the ICU than was active surveillance testing with isolation of MRSA-positive patients. Although data from these recent, cluster-randomized trials of active surveillance testing and the emergence of data supporting the use of horizontal measures such as universal decolonization therapy have contributed substantial new data regarding the control of multidrug-resistant gram-positive pathogens (e.g., MRSA and VRE) in the ICU in non-outbreak situations, additional study of these interventions in other settings and for other pathogens, such as MDR-GNB, is needed.

Interfacility and Regional Collaboration

Given the frequency with which patients are shared among hospitals and other health care settings within geographic regions and the evidence of interfacility patient transfer as a risk factor of introduction of MDROs and *C. difficile* into a health care facility, implementation of comprehensive prevention programs in all facilities within a region may be more effective in preventing HAIs and MDRO transmission than implementation of the same interventions in a single health care institution. The establishment of regional infection prevention initiatives offers several potential benefits, including the ability to more effectively address the complex epidemiology of HAIs that involve multiple health care facilities (e.g., MDRO acquisition, transmission, and infection) than single-institution interventions and opportunities for collaboration and sharing of information, experiences, and resources. A number of successful regional and national initiatives for the prevention of HAI, including CLABSI, dialysis-related blood stream infections, MRSA, VRE, CRE, and *C. difficile*, have been described.

Grade A References

A1. Climo MW, Yokoe DS, Warren DK, et al. Effect of daily chlorhexidine bathing on hospital-acquired infection. *N Engl J Med*. 2013;368:533-542.

A2. Milstone AM, Elward A, Song X, et al. Daily chlorhexidine bathing to reduce bacteraemia in critically ill children: a multicentre, cluster-randomised, crossover trial. *Lancet*. 2013;381:1099-1106.

A3. Huang SS, Septimus E, Kleinman K, et al. Targeted versus universal decolonization to prevent ICU infection. *N Engl J Med*. 2013;368:2255-2265.

A4. Oostdijk EA, Kesecioglu J, Schultz MJ, et al. Effects of decontamination of the oropharynx and intestinal tract on antibiotic resistance in ICUs: a randomized clinical trial. *JAMA*. 2014;312:1429-1437.

A5. Huskins W, Huckabee C, O'Grady N, et al. Intervention to reduce transmission of resistant bacteria in intensive care. *N Engl J Med*. 2011;364:1407-1418.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GOALS AND INCENTIVES FOR HEALTH CARE–ASSOCIATED INFECTION PREVENTION

In addition to the obvious altruistic incentive of preventing harm to the patient, there are a number of additional incentives for health care providers and health care facilities to prevent HAIs. In the United States, the amount of attention given to HAIs and their prevention has greatly increased since the 1999 release of the Institute of Medicine report *To Err Is Human: Building a Safer Health System*. With this attention has come increasing pressure from the public, regulatory and accreditation agencies, and health care payers for the health care industry to improve its efforts and outcomes related to HAI prevention. Hospital accreditation agencies such as the Joint Commission have established standards for infection prevention programs and periodically survey hospitals to determine if such programs are in place. Similarly, the Centers for Medicare and Medicaid Services (CMS) Conditions of Participation list specific requirements for infection prevention and control processes in health care facilities that receive funding from CMS. Hospitals found to be deficient in the implementation of the requirements are at risk of losing the funding that they receive from the agency.

During the past decade, several legislative acts in the United States, such as the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, the Deficit Reduction Act of 2005, and the Patient Protection and Affordable Care Act of 2010, have included requirements related to HAI reporting and accountability. In an effort to address the important public health and patient safety issue of HAIs, the U.S. Department of Health and Human Services developed the National Action Plan to Prevent Healthcare-Associated Infections. Among other things, the Action Plan sets specific goals and targets for improvement during a 5-year period. Introduced in 2009, phase I identified nine national targets for HAI prevention in U.S. acute care hospitals. These nine targets addressed five specific HAIs and health care–associated pathogens, including CLABSI, CAUTI, SSI, MRSA, and *C. difficile*. Phase II of the Action Plan added targets for end-stage renal disease facilities, ambulatory surgery centers, and health care worker influenza vaccination. Phase III introduced targets for long-term care facilities.

In recent years, CMS has introduced several initiatives that provide new incentives for hospitals to demonstrate success in HAI prevention. One of these incentives is the hospital-acquired condition (HAC) program. HACs are conditions that are determined to be high cost or high volume, which result in higher hospital payment by CMS, and that are deemed to be reasonably preventable with application of evidence-based guidelines. In the HAC program, administrative data (i.e., *International Classification of Diseases, Ninth Revision* diagnosis codes assigned at the time of hospital discharge) are used to identify HACs, and as of October 2008, hospitals participating in the CMS Inpatient Prospective Payment System no longer receive additional reimbursement for these conditions when they occur. The HACs currently identified for inclusion in this program are vascular catheter–associated infection, CAUTI, and some SSIs.

In 2011, the CMS Inpatient Quality Reporting Program began to require reporting of CLABSIs that occur in ICUs in acute care hospitals with use of clinical infection surveillance data. Since 2011, more HAI-related process and outcome measures have been added to the program. As of November 2013, reporting requirements include CLABSI and CAUTI in ICU patients, SSI related to colon surgery and abdominal hysterectomy, MRSA bacteremia, *C. difficile* infection, and health care worker influenza vaccination data. Requirements for reporting additional HAI-related outcome will likely be added in subsequent years. Some of these metrics will be used along with other quality measures to determine a hospital's annual reimbursement from CMS through its value-based purchasing program. In addition to being used to determine hospital reimbursement, much of this HAI-related data is being made available to the public through the Hospital Compare website (www.hospitalcompare.hhs.gov). In addition to financial consequences for hospitals with the highest rates of HAIs, in 2011, CMS introduced the Partnership for Patients program, which provides resources to assist hospitals to improve patient safety with the goal of reducing specific hospital-associated conditions, including CAUTI, CLABSI, SSI, and VAP, by 40% by the end of the year 2013. More than 3300 U.S. hospitals elected to participate in this voluntary program.

Several individual U.S. states have introduced mandatory reporting programs for one or more HAI-related outcomes. Many of those states' mandatory reporting programs include public reporting of HAI data. Some states have also mandated implementation of specific infection prevention strategies, such as antimicrobial stewardship, active surveillance testing for MRSA, and mandatory influenza vaccination or use of masks for health care workers during influenza season. Although the reported intention of these programs is to prevent HAIs, some data suggest that mandated public reporting and nonpayment programs do not result in improved HAI-related processes or outcomes.

GENERAL REFERENCES

1. Magill SS, Edwards JR, Bamberg W, et al. Multistate point-prevalence survey of health care-associated infections. *N Engl J Med*. 2014;370:1198-1208.
2. Scott RD. The direct medical costs of healthcare-associated infections in U.S. hospitals and the benefits of prevention. Centers for Disease Control and Prevention; 2009. Available at: http://www.cdc.gov/hai/pdfs/hai/scott_costpaper.pdf. Accessed January 25, 2015.
3. Sievert DM, Ricks P, Edwards JR, et al. Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009-2010. *Infect Control Hosp Epidemiol*. 2013;34:1-14.
4. Rosenthal VD, Bijie H, Maki DG, et al. International Nosocomial Infection Control Consortium (INICC) report, data summary of 36 countries, for 2004-2009. *Am J Infect Control*. 2012;40:396-407.
5. Magiorakos AP, Srinivasan A, Carey RB, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect*. 2012;18:268-281.
6. Kallen A, Mu Y, Bulens S, et al. Health care–associated invasive MRSA infections, 2005-2008. *JAMA*. 2010;304:641-648.
7. Nguyen DB, Lessa FC, Belflower R, et al. Invasive methicillin-resistant *Staphylococcus aureus* infections among chronic dialysis patients in the United States, 2005-2011. *Clin Infect Dis*. 2013;57:1393-1400.
8. Pawar D, Tsay R, Nelson DS, et al. Burden of *Clostridium difficile* infection in long-term care facilities in Monroe County, New York. *Infect Control Hosp Epidemiol*. 2012;33:1107-1112.
9. Guerrero DM, Nerandzic MM, Jury LA, et al. Acquisition of spores on gloved hands after contact with the skin of patients with *Clostridium difficile* infection and with environmental surfaces in their rooms. *Am J Infect Control*. 2012;40:556-558.
10. Guh AY, Thompson ND, Schaefer MK, et al. Patient notification for bloodborne pathogen testing due to unsafe injection practices in the US health care settings, 2001-2011. *Med Care*. 2012;50:785-791.
11. Umscheid C, Mitchell M, Doshi J, et al. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. *Infect Control Hosp Epidemiol*. 2011;32:101-114.
12. Dudeck MA, Horan TC, Peterson KD, et al. National Healthcare Safety Network report, data summary for 2011, device-associated module. *Am J Infect Control*. 2013;41:286-300.
13. Malpiedi P, Peterson K, Soe M, et al. National and state healthcare-associated infection standardized infection ratio report. Centers for Disease Control and Prevention; 2013. Available at: http://www.cdc.gov/hai/pdfs/SIR/SIR-Report_02_07_2013.pdf. Accessed January 25, 2015.
14. Yokoe DS, Anderson DJ, Berenholtz SM, et al. A compendium of strategies to prevent healthcare-associated infections in acute care hospitals: 2014 updates. *Infect Control Hosp Epidemiol*. 2014;35:967-977.
15. Glick S, Samson D, Huang E, et al. *Screening for methicillin-resistant* Staphylococcus aureus *(MRSA). Comparative Effectiveness Review No. 102. (Prepared by the Blue Cross and Blue Shield Association Technology Evaluation Center Evidence-based Practice Center Under Contract No. 290-2007-10058-I.) HRQ Publication No. 13-EHC043-EF*. Rockville, MD: Agency for Healthcare Research and Quality; June 2013.

REVIEW QUESTIONS

1. What percentage of device- and procedure-associated health care–associated infections that occur in U.S. hospitals are potentially preventable?

A. 0-15%
B. 20-30%
C. 40-50%
D. 55-70%
E. 85-100%

Answer: D A systematic review of published studies found that 55 to 70% of central line–associated blood stream infections, catheter-associated urinary tract infections, ventilator-associated pneumonia, and surgical site infections are preventable with the use of currently available evidence-based prevention strategies and technologies. Unfortunately, many of these strategies are not consistently implemented.

2. Bathing of intensive care unit (ICU) patients with chlorhexidine on a daily basis has been associated with which of the following outcomes?

A. Reduction in MRSA-positive clinical cultures attributable to the ICU
B. Reduction in overall blood stream infection rates in ICU patients
C. Reduction in central line–associated blood stream infection rates in ICU patients
D. Reduction in environmental contamination with VRE
E. All of the above

Answer: E Several quasi-experimental studies and three cluster-randomized controlled trials have demonstrated the potential for daily bathing with chlorhexidine to reduce a variety of health care–associated infection outcomes in ICU patients. This intervention has generally been well tolerated. Further study of the risk of development of chlorhexidine resistance among common health care–associated pathogens as the use of chlorhexidine becomes more common is warranted. Daily chlorhexidine bathing in other settings has not been subject to such extensive study, and thus the role of the horizontal infection control intervention in the non-ICU setting is not yet known.

3. The increased rates of morbidity and mortality associated with health care–associated infections caused by multidrug-resistant organisms (MDROs), compared with those caused by susceptible organisms of the same species, are likely due to all *except* which of the following?

A. Greater virulence of antimicrobial-resistant organisms compared with antimicrobial-susceptible organisms
B. Delays in initiating effective antimicrobial therapy
C. The presence of more severe underlying disease in patients who develop MDRO infections
D. The availability of only less effective or more toxic antimicrobial agents for treatment of some MDRO infections compared with agents available for the treatment of more susceptible organisms

Answer: A The higher rates of morbidity and mortality associated with MDRO infections, compared with infections caused by susceptible organisms of the same species, are likely to be multifactorial in nature. In general, however, MDROs are not more virulent than susceptible organisms of the same species.

4. Which of the following scenarios does *not* represent an avoidable health care–associated risk factor for health care–associated infection or MDRO acquisition?

A. The presence of an indwelling urinary catheter in an otherwise healthy patient who underwent elective knee replacement surgery 2 days earlier
B. Insertion of a central venous catheter for administration of cytotoxic chemotherapy
C. Insertion of an indwelling urinary catheter to reduce the need for nursing assistance in a 73-year-old patient with urinary incontinence
D. Administration of antimicrobial prophylaxis for 48 hours after an elective colectomy
E. Continuation of empirically prescribed piperacillin-tazobactam therapy for a patient admitted in septic shock who has now demonstrated a clinical response to current therapy and from whom pan-susceptible *E. coli* has been isolated from blood and urine specimens obtained at the time of admission

Answer: B The appropriateness of indwelling invasive devices and ongoing antimicrobial therapy should be assessed on a regular (i.e., at least once daily) basis with discontinuation of the device when an appropriate indication is not present and adjustment of antimicrobial therapy based on the results of diagnostic testing and clinical assessments.

5. Recommendations for the prevention of central line–associated blood stream infections (CLABSIs) include all of the following *except*

A. Use of maximal barrier precautions during catheter insertion (e.g., use of a cap, mask, sterile gown, sterile gloves, and sterile full-body drape)
B. Routine replacement of central venous catheters every 7 days
C. Preparation of the catheter insertion site with an antiseptic agent (e.g., >0.5% chlorhexidine preparation with alcohol) before catheter insertion
D. Daily review of catheter necessity with immediate removal of catheters that are no longer necessary
E. Scrubbing of the access port with an antiseptic before accessing the port

Answer: B Currently available evidence-based guidelines for prevention of CLABSI provide recommendations for catheter insertion, use, and maintenance. Routine catheter replacement is not a recommended strategy for CLABSI prevention.

283

APPROACH TO THE PATIENT WITH SUSPECTED ENTERIC INFECTION

HERBERT L. DUPONT

EPIDEMIOLOGY

Enteric infections are second only to respiratory tract infections as common infectious medical problems. In certain populations, enteric infections are hyperendemic: poorly nourished infants living in developing tropical countries showing excessive rates of mortality; infants in certain daycare centers; unhygienic residents of custodial institutions for the mentally retarded; immunosuppressed persons; and visitors from industrialized areas to developing regions with traveler's diarrhea.

ETIOLOGY

In approaching a patient with an enteric infection, epidemiologic (Table 283-1) and clinical features (Table 283-2) are used to identify the type of etiologic agent responsible for illness and to develop a plan for evaluation (Table 283-3) and management (Table 283-4).

Recent travel (Chapter 286) to mountainous regions or recreational lakes of North America should raise the suspicion of infection caused by *Giardia* species.[1] When diarrhea occurs during or after travel to a developing tropical region, a bacterial enteropathogen should be suspected.[2] The leading causes of traveler's diarrhea worldwide are the diarrheogenic *Escherichia coli*[3]: enterotoxigenic *E. coli* (ETEC) and enteroaggregative *E. coli* (EAEC). The invasive bacteria (*Shigella*, *Salmonella*, and *Campylobacter* species) cause diarrhea among travelers to all regions but are more common in Asia. Infection with *Cyclospora* species should be suspected when persistent or recurrent diarrhea follows travel to Nepal, Haiti, or Peru or other regions of the developing world (travel-related infections are discussed in detail in Chapter 286).

A specific food or water vehicle cannot be suspected unless multiple cases of illness with a common exposure occur. All too frequently, persons assume that food consumed during their last meal before an illness onset is responsible for the symptoms. The highly variable incubation period for diarrheal disease, which may be as short as 2 hours after eating a food, for preformed toxins, to a week or even longer for microbial enteropathogens, makes the determination of a specific food or beverage in a single case of illness impossible. When an outbreak of diarrhea results in multiple cases, a category of etiology (preformed toxin versus enteric infection) can be determined by calculating the incubation period after looking at timing of the common exposure and the time of first symptoms. Short incubation periods are characteristic of food poisoning associated with enterotoxins (2 to 7 hours for cases caused by *Staphylococcus aureus*, 2 to 4 hours for *Bacillus cereus* enterotoxin food poisoning). Longer incubation periods (usually 12 to 72 hours or longer) are associated with most cases of intestinal infection.

The clinical expression of diarrheal illness will give clues to the etiologic agent involved in disease (see Table 283-2). In the patient with diarrhea who is receiving or recently has completed a course of an antimicrobial drug, a proton pump inhibitor, or an anticancer drug, particularly with recent or

current hospitalization, *Clostridium difficile* infection (Chapter 296) should be suspected. An increasing number of cases of *C. difficile* diarrhea are occurring in the community. When a person has close contact with an infant or infants attending a daycare center, a number of low-dose pathogens found in this setting (e.g., *Giardia*, *Cryptosporidium*, or *Shigella* species or viral pathogens, particularly norovirus) should be suspected. Some homosexuals may show high rates of enteric infection acquired through fecal-oral contamination, often associated with infection by multiple pathogens or through the practice of unprotected receptive anal intercourse leading to proctitis due to sexually transmitted organisms. When persons experience advanced acquired immunodeficiency syndrome (AIDS) or other forms of severe immunodeficiency associated with metastatic malignant disease or chronic use of immunosuppressive drugs, depressed intestinal immunity may lead to enteric infection with a variety of parasitic, bacterial, or viral pathogens (see Table 283-1) (Chapter 281). Infants with malnutrition may develop persistent diarrhea and substantial long-term morbidity due to protozoal parasites, including *Giardia* and *Cryptosporidium*.

TABLE 283-1 EPIDEMIOLOGIC FEATURES IMPORTANT IN DETERMINING POTENTIAL CAUSE OF ENTERIC INFECTION IN A PERSON OR PERSONS WITH DIARRHEA

EPIDEMIOLOGIC FEATURE	ETIOLOGIC AGENT TO SUSPECT
Travel to mountainous areas of North America	*Giardia* spp
Travel to Russia (especially St. Petersburg)	*Cryptosporidium*, *Giardia* spp
Travel to Nepal	*Cyclospora* spp
Travel to the developing tropical/semitropical world from an industrialized region	Enterotoxigenic *Escherichia coli*, enteroaggregative *E. coli*; *Shigella*, *Campylobacter*, *Salmonella* spp; other bacterial causes; *Giardia*, *Cyclospora*, *Cryptosporidium* spp and noroviruses
Presence of associated cases (an outbreak)	Use incubation period and clinical features to determine probable cause
Antibiotic, chemotherapy, or proton pump inhibitor use in the past 2 months, particularly with a history of recent or current hospitalization	*Clostridium difficile*
Contact with daycare centers	Any enteropathogen, often the low-dose organisms: *Giardia*, *Cryptosporidium*, *Shigella* spp or viral pathogens
Homosexual person practicing unprotected sex	Any organism spread by fecal-oral route; in those with proctitis, suspect *Neisseria gonorrhoeae*, *Chlamydia trachomatis*, herpes simplex, or *Treponema pallidum*
Immunosuppressed person	Any agent, especially *Cryptosporidium*, *Cyclospora*, *Isospora*, *Shigella*, and *Salmonella* spp; *C. jejuni*, *C. difficile*, *Mycobacterium avium-intracellulare*, microsporidia, herpes simplex virus, and cytomegalovirus
Recent or current cruise ship travel	Norovirus, less frequently enterotoxigenic *E. coli*

Enteric infection syndromes may be divided into at least five groups on the basis of the clinical presentation: (1) febrile systemic disease (enteric fever); (2) acute watery diarrhea (secretory diarrhea); (3) recurrent vomiting as the primary manifestation of enteric disease (gastroenteritis); (4) passage of many small-volume stools containing blood and mucus (dysentery); and (5) diarrhea lasting 2 weeks or longer (persistent diarrhea). Table 283-2 lists the major syndromes along with the expected cause.[4]

Noroviruses (Chapter 380) have become the major cause of food-borne gastroenteritis and the most commonly identified cause of waterborne enteric disease.[5] They have been identified as causes of persistent diarrhea in immunocompromised patients, especially in those undergoing hematopoietic stem cell transplantation. *Campylobacter* species (Chapter 303) is a commonly reported bacterial enteropathogen in industrialized countries and is the most important definable cause of Guillain-Barré syndrome (Chapter 420), often resulting in severe disease requiring assisted ventilation, intensive care unit confinement, and permanent neurologic sequelae. *E. coli* O157:H7 and other Shiga toxin–producing *E. coli* (STEC) (Chapter 304) are important causes of food-borne and waterborne colitis complicated by hemolytic-uremic syndrome in children and occasionally in elderly people.

The most commonly detected pathogens in endemic diarrhea in the United States in one study were noroviruses (26%), rotavirus (18%), and *Salmonella* species (5.3%). EAEC strains are being shown to be important causes of pediatric diarrhea in the United States as diagnostic methods improve.[6] In May and June of 2011, there was a large outbreak of diarrhea and hemolytic-uremic syndrome reported from Germany and France due to an *E. coli* O104:H4 strain of EAEC that had picked up the STEC phage controlling production of Shiga toxin. In the future, with improved diagnostic tools, we are likely to see more of these hybrid superpathogens with multiple virulence properties.

DIAGNOSIS

Laboratory Findings

Laboratory tests (Fig. 283-1; see Table 283-3) can be useful and are of particular value in the more severely ill patients, when subjects are forced by their illness to alter activities or are totally disabled and confined to bed or when many patients are afflicted during an outbreak. In each of these situations, the laboratory may help establish cause and allow development of a proper plan of treatment (see Table 283-3). Useful laboratory tests include procedures looking for fecal inflammatory markers, such as microscopic detection of fecal leukocytes or the more sensitive, commercially available test, fecal lactoferrin or calprotein. These tests are particularly helpful to suggest the presence of the invasive bacterial pathogens *Shigella*, *Salmonella*, and *Campylobacter* species or the noninvasive but inflammatory *C. difficile*. Stool culture is performed in more severe cases of sporadic diarrhea and in disease outbreaks

TABLE 283-2 CLINICAL FEATURES OF ENTERIC INFECTION

CLINICAL SYNDROME	ETIOLOGIC AGENTS SUSPECTED	SPECIAL CONSIDERATIONS
Sustained fever, often with systemic toxicity (enteric or typhoid fever)	*Salmonella typhi*, nontyphoid *Salmonella* spp, *Campylobacter* spp, *Shigella* spp, *Yersinia enterocolitica*	Stool and blood cultures; empirical antibiotics generally indicated
Acute watery (secretory) diarrhea	Any agent. Consider *Vibrio cholerae* (if water losses are major), enterotoxigenic or enteroaggregative *Escherichia coli*, *Shigella* spp, *Salmonella* spp, *Campylobacter jejuni*, viral or parasitic protozoal pathogen	Fluid and electrolyte therapy crucial for recovery in dehydration
Recurrent vomiting (gastroenteritis)	Viral agents (rotavirus or noroviruses) or preformed toxin (*Staphylococcus aureus* or *Bacillus cereus*)	In case of an outbreak, incubation period suggests the etiology
Bloody diarrhea (dysentery)	*Shigella* spp, *C. jejuni*, *Salmonella* spp, Shiga toxin–producing *E. coli* (e.g., O157:H7 or other serotype) or invasive *E. coli*, *Aeromonas hydrophila*, noncholera *Vibrio* spp, *Yersinia enterocolitica*, *Entamoeba histolytica*, or inflammatory bowel disease	Stool culture and occasionally parasite examination important to determining cause; hemolytic-uremic syndrome may complicate diarrheal disease caused by Shiga toxin–producing *E. coli* or rarely *Shigella dysenteriae*
Diarrhea lasting ≥2 weeks (persistent diarrhea)	*Giardia* spp and other protozoal parasites, bacterial overgrowth, bacterial diarrhea, lactase deficiency, Brainerd diarrhea, postinfectious irritable bowel syndrome (PI-IBS), unmasked inflammatory bowel disease (IBD), or celiac sprue	Stool culture and parasite examination indicated; empirical anti-*Giardia* therapy may be useful; remove milk from diet; prior raw milk or untreated (well or surface) water consumption or international travel may predispose to Brainerd diarrhea; with illness lasting >30 days, consider Brainerd diarrhea, PI-IBS, celiac disease, or IBD

TABLE 283-3 LABORATORY TESTS AND PROCEDURES USEFUL IN THE DIAGNOSIS OF INFECTIOUS DIARRHEA

SPECIFIC TEST OR PROCEDURE	WHEN INDICATED	CLINICAL SIGNIFICANCE
Fecal leukocyte test	For moderate to severe cases	When present, indicates diffuse colonic inflammation, often due to *Shigella*, *Salmonella*, *Campylobacter* spp, Shiga toxin–producing *Escherichia coli*, or *Clostridium difficile*
Fecal lactoferrin	For moderate to severe cases to help identify inflammatory forms of enteric infection, to use in health care–associated diarrhea to help determine whether *C. difficile* toxin test should be performed	More sensitive test than fecal leukocytes and will pick up the same pathogens as fecal leukocytes but also pathogens associated with less striking degrees of inflammation (enteroaggregative *E. coli* and *C. difficile*)
C. difficile toxins A and B	Diarrhea associated with use of antibiotics, chemotherapy, or proton pump inhibitors, especially associated with current or recent hospitalization	Most sensitive tests are culture and tissue culture assay. Polymerase chain reaction is sensitive but lacks specificity. Most specific tests are enzyme immunoassay for toxins A and B; a two-step procedure can be used: sensitive but nonspecific *C. difficile* glutamate dehydrogenase antigen test followed by toxin assay
Stool culture for *Shigella*, *Salmonella*, *Campylobacter* spp, and Shiga toxin–producing *E. coli* (O157:H7 and others)	Moderate to severe diarrhea and when stools are positive for inflammatory markers or contain gross blood and mucus (dysentery)	The four mucosa-inflammatory bacteria are the only bacteria routinely sought by most laboratories.
Specialized stool culture for *Vibrio* spp	For cases of profuse watery diarrhea in cholera-endemic areas and outbreaks of seafood-associated diarrhea or dysentery	Cholera cases may need aggressive fluid therapy. Non-cholera vibrios can cause dysentery.
Parasite examination: (1) enzyme immunoassay for *Giardia* spp, *Cryptosporidium* spp, or *Entamoeba histolytica*; (2) acid-fast stain for *Cyclospora* or *Cryptosporidium* spp or *Isospora*; or (3) trichome stain and microscopic examination	In any patient with persistent diarrhea and when diarrhea follows visits to mountainous or recreational lakes in North America, Nepal, Haiti, Peru, or Russia (particularly St. Petersburg)	If microscopic evaluation is performed, experience of the laboratory personnel is important. The commercially available enzyme immunoassay tests are sensitive.
Esophagogastroduodenoscopy and flexible sigmoidoscopy	Persistent diarrhea in patients without evidence of cause of illness	Identified cause of diarrhea is treated; without diagnosis, subjects may be treated symptomatically.

TABLE 283-4 THERAPY FOR AND PREVENTION OF INFECTIOUS DIARRHEA

THERAPEUTIC OPTION	INDICATION	PHARMACOLOGIC AGENT
Oral fluid and electrolyte therapy	For infants, elderly patients, and anyone with profuse watery diarrhea	Soups, soft drinks, and saltine crackers are sufficient; formal oral replacement therapy may be needed with dehydrating forms of diarrhea
Diet	In all forms of diarrhea to facilitate enterocyte renewal and recovery	Soups and broth, saltine crackers, steamed vegetables, baked or broiled meats
Nonspecific therapy	For temporary (≤48 hours) control of diarrhea in older children and adults without evidence of severe diarrhea caused by an invasive or inflammatory bacterial or parasitic pathogen	Loperamide is the most effective symptomatic treatment and will decrease number of stools passed by 60%; bismuth subsalicylate is much less effective and will reduce number of stools by 40%; the antisecretory agent crofelemer is of value in HIV-associated diarrhea
Empirical antibacterial drugs	Enteric fever with toxicity Febrile dysenteric diarrhea Traveler's diarrhea	Fluoroquinolones for 7-10 days Azithromycin is recommended when fever or dysentery complicates illness Rifaximin for 3 days, fluoroquinolone for 1-3 days, or azithromycin 1000 mg in single dose
Specific antibacterial therapy	Shigellosis, campylobacteriosis, cholera	See Chapters 302, 303, and 309
Antiparasitic drugs	Giardiasis, amebiasis, cryptosporidiosis, cyclosporiasis	See Chapters 350, 351, and 353
Prophylaxis in traveler's diarrhea	Persons traveling to developing areas on tight schedules, those with history of prior traveler's diarrhea, persons with unstable underlying medical disorders, and those interested in prophylaxis	Rifaximin, 200 mg twice daily with meals, while in a high-risk region

and is carried out with blood culture in a patient with fever and systemic toxicity. Other indications for stool culture are presence of dysentery (passage of grossly bloody stools) and when fecal inflammatory markers are found. In dysenteric diarrhea, particularly in the presence of an outbreak, the laboratory should also be instructed to look for *E. coli* O157:H7 and other Shiga toxin–producing *E. coli*. Parasite examination is indicated by diarrhea and persistent (≥14 days) illness; a recent trip to Nepal, Haiti, Peru, or Russia; evidence that the subject practices oral-anal sex or unprotected receptive anal intercourse; or associated immunosuppression. Other tests are indicated in special situations, including stool culture for *Vibrio cholerae* in a patient with severe watery diarrhea with excessive fluid losses in a cholera-endemic area and culture for *Mycobacterium avium* complex, herpes simplex virus, and cytomegalovirus in those with immunosuppression. For patients with persistent diarrhea without etiologic diagnosis when routine tests are employed, endoscopy (esophagogastroduodenoscopy and flexible sigmoidoscopy or colonoscopy) may be indicated in attempting to determine the nature and cause of illness.

The broad range of conventional diagnostic approaches currently available to identify enteropathogens associated with infectious diarrhea will be increasingly supplemented by or replaced with new molecular methods, including real-time polymerase chain reaction, quantitation of pathogen load, and next-generation sequencing.[7]

TREATMENT

Rx

Treatment of diarrhea should be tailored to the clinical syndrome. Oral rehydration therapy with fluids and electrolytes is used to treat acute watery diarrhea and gastroenteritis and all forms of enteric infection, especially when complicated with any degree of dehydration. Oral rehydration therapy is particularly important in infants; it can be life-saving in developing countries for infants with severe diarrhea. Patients with diarrhea should be fed easily digestible foods to facilitate enterocyte renewal and to speed up disease recovery.

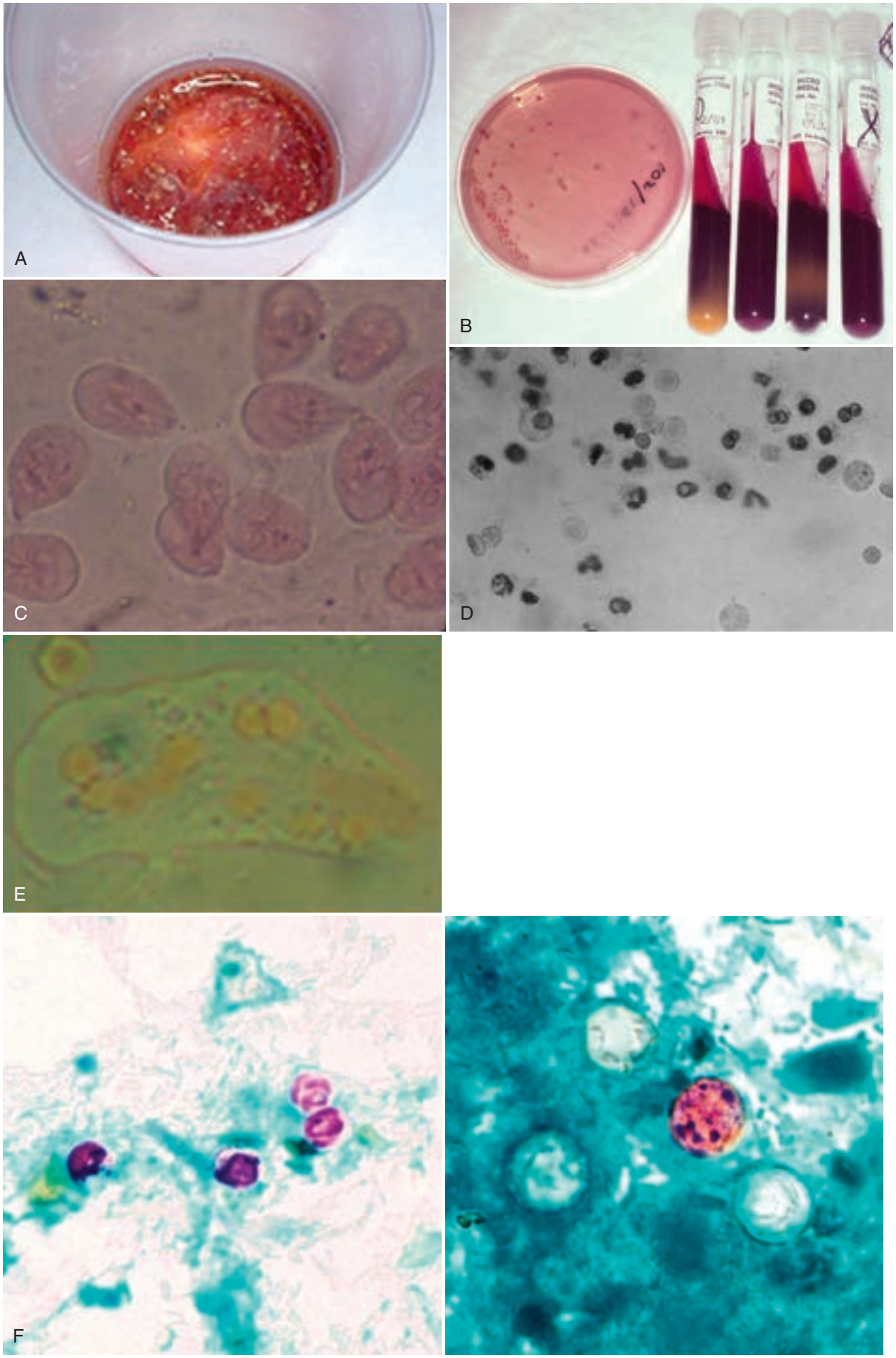

FIGURE 283-1. Laboratory tests to diagnose causes of diarrhea. A, Dysenteric stool. **B,** Stool culture and biochemical tests confirm *Salmonella*. **C,** *Giardia* trophozoites. **D,** Many leukocytes in diffuse colonic inflammation. **E,** *Entamoeba histolytica* trophozoite with ingested red blood cells. **F,** Oocysts of *Cryptosporidium* (*left*) and *Cyclospora* (*right*). (From the CDC Public Health Information Library. http://phil.cdc.gov/phil/home.asp: images 7829 and 7827.)

TABLE 283-5 NONINFECTIOUS CAUSES OF DIARRHEA

Running	Small bowel bacterial overgrowth
Fecal impaction	Systemic mastocytosis and eosinophilic gastroenteritis
Drugs and laxatives	Tropical sprue
Enteral feeding	Celiac sprue
Irradiation	Dermatitis herpetiformis
Pancreatic insufficiency	Whipple's disease
Intestinal lymphangiectasia	Thyrotoxicosis
Foods (especially dietetic)	Adrenal insufficiency
Cirrhosis and biliary obstruction	Factitious
Diabetic diarrhea	Inflammatory bowel disease
Alcoholism	Food allergy
Collagenous colitis	Carcinoid
VIPoma	Villous adenoma
Ischemic bowel disease	Stress with autonomic stimulation
Irritable bowel syndrome	

In afebrile, nondysenteric diarrhea, symptomatic drugs may allow older children and adults with illness to return earlier to school or work. Loperamide is the most active drug for improvement of symptoms. Bismuth subsalicylate can reduce diarrhea and is mildly effective in reducing nausea and vomiting associated with viral gastroenteritis.

For enteric fever, febrile dysenteric disease, and moderate to severe cases of traveler's diarrhea, empirical antimicrobial therapy is indicated (see Table 283-4).[8] For outbreaks of dysenteric diarrhea, particularly in children in whom fever is not significant, antibacterial and antimotility drugs should be initially withheld while the etiology of the outbreak is being established to prevent patients infected by STEC strains from being predisposed to hemolytic-uremic syndrome. For bacterial and parasitic pathogen-specific diarrhea, antimicrobial therapy is often advised (see other chapters in the text for specific treatments). Because of the importance of diarrhea when persons travel from industrialized regions to developing countries, prophylaxis with the orally administered, poorly absorbed rifaximin can be employed for some groups (see Table 283-4), with expected protection rates exceeding 70%.

In sporadic cases of acute or persistent diarrhea, infectious agents are not always responsible. Table 283-5 offers a partial list of the noninfectious causes of diarrhea that should be considered.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. DuPont HL. *Giardia*: both a harmless commensal and a devastating pathogen. *J Clin Invest*. 2013;123:2352-2354.
2. Steffan R, Hill DR, DuPont HL. Traveler's diarrhea: A clinical review. *JAMA*. 2015;313:71-80.
3. Gould LH, Mody RK, Ong KL, et al. Increased recognition of non-O157 Shiga toxin–producing *Escherichia coli* infections in the United States during 2000-2010: epidemiologic features and comparison with *E. coli* O157 infections. *Foodborne Pathog Dis*. 2013;10:453-460.
4. Pfeiffer ML, DuPont HL, Ochoa TJ. The patient presenting with acute dysentery—a systematic review. *J Infect*. 2012;64:374-386.
5. Ramani S, Atmar RL, Estes MK. Epidemiology of human noroviruses and updates on vaccine development. *Curr Opin Gastroenterol*. 2014;30:25-33.
6. Scallan E, Mahon BE, Hoekstra RM, Griffin PM. Estimates of illness, hospitalization, and deaths caused by major bacterial enteric pathogens in young children in the United States. *Pediatr Infect Dis*. 2013;32:217-221.
7. Platts-Mills JA, Liu J, Houpt ER. New concepts in diagnostics for infectious diarrhea. *Mucosal Immunol*. 2013;6:876-885.
8. DuPont HL. Acute infectious diarrhea in immunocomopetent adults. *N Engl J Med*. 2014;370:1532-1540.

REVIEW QUESTIONS

1. A 30-year old traveler to Goa, India, for vacation develops diarrhea and fever on the fifth day at the resort. After 36 hours of diarrhea, stools begin to contain bright red blood. She visited a U.S. travel medicine clinic before heading for India and was given two antibiotics, one for the common watery diarrhea and a second in the less likely event that she developed bloody diarrhea and fever. Which of the following antibiotics would she have correctly received from the clinic to treat dysenteric traveler's diarrhea?

A. Oral ciprofloxacin, 500 mg twice a day for 3 days
B. Azithromycin, 1000-mg single dose
C. Levofloxacin, 500 mg once a day for 3 days
D. Metronidazole, 500 mg three times a day for 10 days
E. Rifaximin, 200 mg three times a day for 3 days

Answer: B The invasive pathogens (*Shigella*, *Salmonella*, and *Campylobacter*) occur more commonly in travelers to Asia compared with destinations in Africa and Latin America. Ciprofloxacin and levofloxacin would not be the best choice in this patient because of the importance of fluoroquinolone-resistant *Campylobacter* as a causative agent of dysenteric traveler's diarrhea occurring in Asia. Metronidazole would be appropriate for persistent diarrhea caused by *Giardia* or *Entamoeba histolytica* but would not be useful in most patients with traveler's diarrhea where the typical enteropathogenic agents are bacterial in origin. Rifaximin is of value for the common watery diarrhea of travelers but is of no value for the treatment of mucosally invasive enteropathogens as the drug is poorly absorbed (<0.4%). Azithromycin has the broadest spectrum of activity against the causes of traveler's diarrhea and should be preferentially used for dysenteric traveler's diarrhea or when breakthrough diarrhea occurs in a patient taking rifaximin to prevent traveler's diarrhea.

2. Two families eat a common meal and five people from the two families develop diarrhea the next afternoon, of whom two are passing grossly bloody stools. A sixth person eating the meal without diarrhea develops motor weakness and decreased reflexes 11 days later, progressing to quadriplegia. She is confined to a hospital intensive care unit requiring mechanical ventilation. What probably caused the outbreak of diarrhea and the case of quadriplegia?

A. *Campylobacter jejuni*
B. *Shigella flexneri*
C. *Vibrio cholerae*
D. *Aeromonas sobria*
E. *Vibrio vulnificus*

Answer: A The most common causes of dysentery (passing grossly bloody stools) in the United States are *Shigella* and *Campylobacter*. *Campylobacter* is the most important definable cause of Guillain-Barré syndrome (GBS), occurring during the 2 months after *Campylobacter* infection in fewer than 2 per 10,000 cases. GBS is an example of molecular mimicry with antibodies directed to *Campylobacter* lipooligosaccharides and peripheral nerve gangliosides. The complication can occur in asymptomatic *Campylobacter* infection as was seen in this case. *Campylobacter* can be suspected by searching for serologic evidence of past infection in subjects with GBS.

3. You receive a call from your office. One of your patients has developed watery diarrhea 2 hours after eating at a local restaurant. No other persons are known to be ill. What should you tell your office?

A. Obtain a sample of food if possible for *S. aureus* enterotoxin.
B. Have the office draw a blood culture sample.
C. Do a complete clinical and epidemiologic history.
D. Recommend ciprofloxacin therapy for the patient.
E. Collect a stool sample for studies of enteric pathogens.

Answer: C A meal or beverage can be implicated as a cause of an enteric disease if there are associated cases of illness occurring with a common exposure. The only exception to this is a case of classic botulism occurring shortly after consumption of a food. The incubation period of an enteric infection/intoxication can be as long as 9 days and as short as 2 hours. Thus, the problem could have resulted with an exposure during this broad time range. By taking a careful history, it would have been learned that this patient had diarrhea as well as vomiting and a temperature of 102° F. By obtaining more information, an infectious agent is more likely than a preformed toxin of *Staphylococcus aureus* or *Bacillus cereus*, which could produce illness with this short incubation period. Clearly, the meal was not responsible because an enteric infection has a minimum incubation period of 14 hours. In this subject, a stool culture revealed a *Salmonella* spp, and he gave a history of eating a turkey meal the day before.

4. In January 2010, local authorities in Maryland were notified of a large cluster of gastroenteritis involving more than 100 persons, 24 to 36 hours after consumption of raw oysters in three Baltimore restaurants. The median incubation period was 30 hours. Vomiting and watery nonbloody diarrhea with abdominal pain were common. Some subjects complained only of nausea and vomiting. Fever was seen in approximately 30% of the affected persons. Approximately 20 subjects had stool studies performed, which were negative for bacterial and parasitic pathogens. Clinical recovery was complete in 72 hours, and there were no long-term sequelae. Which of the following organisms likely caused the outbreak?

A. *Shigella sonnei*
B. *Vibrio cholerae*
C. *Giardia*
D. Norovirus
E. *Staphylococcus aureus* enterotoxin

Answer: D In the United States each year, more than 20 million cases of acute norovirus gastroenteritis occur, with half being reported to be foodborne. The reservoir of norovirus is infected persons who may be asymptomatic. With fecal contamination of an oyster bed, norovirus contamination occurs readily. Outbreaks of norovirus gastroenteritis from oysters are not rare. The illness seen here meets the Kaplan criteria for norovirus illness: (1) median duration of illness is 12 to 60 hours; (2) median incubation period is 24 to 48 hours; (3) more than 50% of affected persons complain of vomiting; and (4) no bacterial agent is found. Norovirus is more common in winter months, which fits with the timing of the outbreak. *Shigella*, *V. cholerae*, and *Giardia* produce a different disease (febrile dysentery, dehydrating diarrhea, persistent illness). *Staphylococcus* food poisoning associated with concentration of an enterotoxin in improperly handled food has a shorter incubation period (2 to 7 hours) and is not likely to be seen in oyster-associated outbreaks.

5. On June 15, 2010, local physicians reported 11 pediatric cases of diarrhea to a county health department. A preliminary investigation found that nine of the persons recently had visited a large city park with a wading pool. The health department performed an evaluation of persons who used the park or wading pool and found that of 89 children interviewed, 69 met the case definition. Secondary spread of illness to family members and contacts was found to be common. One third of the children had temperature above 102° F; half had bloody diarrhea. A pathogen was isolated from three children providing a stool sample. Which of the following is likely to be the cause of the outbreak?

A. *Shigella sonnei*
B. Enteroinvasive *E. coli*
C. *Campylobacter jejuni*
D. *Vibrio cholerae*
E. Enterotoxigenic *E. coli*

Answer: A Outbreaks of infectious diarrhea have been reported among infants and children in wading pools where chlorine levels are not monitored and fecal contamination is expected. The likely causes of the illness are the low-dose pathogens: norovirus, *Shigella* spp, Shiga toxin–producing *E. coli*, *Giardia*, and *Cryptosporidium*. Secondary spread of enteropathogens commonly occurs only for low-dose organisms. *Salmonella* and *Campylobacter* in most patients (outside the very young infant stage) are moderate-dose pathogens. *V. cholerae* and ETEC are high-dose pathogens. The moderate- and high-dose pathogens would be rare causes of outbreaks in this setting and would not have been associated with secondary transmission. If Shiga toxin–producing *E. coli* (STEC) would have been given as an option, two answers could have been correct. STEC is a low-dose pathogen seen in wading pool outbreaks and with secondary cases. The high fever is more compatible with shigellosis than with STEC diarrhea, however. *V. cholerae*, EAEC, and ETEC are each high-dose pathogens not associated with wading pools; secondary spread does not occur for them, and none causes dysenteric illness.

284 APPROACH TO THE PATIENT WITH URINARY TRACT INFECTION

LINDSAY E. NICOLLE AND S. RAGNAR NORRBY

DEFINITIONS

Urinary tract infection (UTI) is bacterial or fungal infection of the normally sterile urine. The clinical presentation varies from asymptomatic bacteriuria, when the urine culture is positive but there are no symptoms, to cystitis (bladder or lower tract infection), pyelonephritis (renal or upper tract infection), and urosepsis (systemic inflammatory response syndrome or septic shock from a urinary source). Urethritis caused by *Chlamydia trachomatis*, *Ureaplasma urealyticum* (Chapter 285), or *Neisseria gonorrhoeae* (Chapter 299) and prostatitis (Chapter 129) and renal tuberculosis (Chapter 324) are addressed elsewhere in this text.

Uncomplicated UTI occurs in women with a normal genitourinary tract. Most episodes are manifested as cystitis; acute nonobstructive pyelonephritis also occurs in these women, at a lower frequency. Complicated UTI occurs in patients with functional or structural abnormalities of the urinary tract. Complicating factors are host factors facilitating the establishment and persistence of bacteriuria or infection (Table 284-1). Uncomplicated UTI occurs rarely in young men. Men presenting with UTI should be assumed to have complicated infection until it is proved otherwise. Recurrent infection is considered a reinfection when it is caused by a new bacterial strain and relapse when the same strain that caused preceding infections is isolated. It is clinically important to classify UTIs by site of infection, presence or absence of symptoms, tendency to recur, and presence or absence of complicating factors.

TABLE 284-1 HOST FACTORS ASSOCIATED WITH COMPLICATED URINARY TRACT INFECTION

	EXAMPLES
Obstruction	Urethral or ureteric strictures, tumor Diverticula Pelvicalyceal junction obstruction Prostate enlargement Urolithiasis Extrinsic compression
Functional	Neurogenic bladder Vesicoureteral reflux Anatomic defects Pregnancy Turbulent urethral urine flow Cystocele
Urologic interventions	Indwelling or suprapubic catheters Endourologic surgery Ureteric stents Nephrostomy tubes Cystoscopy Neobladders
Metabolic or congenital diseases	Urethral valves Polycystic kidneys Nephrocalcinosis Medullary sponge kidney
Immunologic abnormalities	Renal transplantation

EPIDEMIOLOGY

UTI is the most common bacterial infection. It is somewhat more common in boys than in girls in the newborn period because of the higher frequency of urethral malformations in boys. Later in childhood, symptomatic UTIs and asymptomatic bacteriuria are more common in girls. More than half of all healthy women experience at least one symptomatic UTI in their lifetime, and each year, 2 to 10% of women experience at least one episode. UTI is uncommon in men with a normal genitourinary tract but increases after the age of 65 years, primarily attributable to prostate hypertrophy and prostatitis. The frequency of infection in patients with complicated infection varies by the abnormality promoting infection. For instance, patients with spinal cord injury and neurogenic bladder have continuing high rates of infection, but patients with abnormalities that can be corrected will no longer be at risk of infection. UTI is also one of the most common hospital-acquired infections; about 80% of these are a consequence of the use of an indwelling bladder catheter.

Asymptomatic bacteriuria is common. The prevalence increases from 1 to 2% of schoolgirls to 3 to 5% of sexually active premenopausal women, 10 to 20% of healthy postmenopausal women, and 40 to 50% of elderly women in nursing homes. It is infrequent in men until older ages, with 5 to 10% of elderly men in the community and 35 to 40% in nursing homes having bacteriuria. Some patients with genitourinary complications also have a very high prevalence. For instance, 50% of spinal cord–injured patients with a neurogenic bladder and without an indwelling catheter and 100% of patients with chronic indwelling urethral catheters have bacteriuria.

PATHOBIOLOGY

Acute uncomplicated UTI follows ascension into the bladder or kidney of uropathogenic organisms in the normal gut flora that have colonized the vagina and periurethral mucosa. The ability of these organisms, usually *Escherichia coli*, to colonize and to persist within the urinary tract is dependent on an array of virulence factors that include adhesins, toxins, and iron-scavenging proteins. Organism virulence is a major determinant of whether infection is asymptomatic or symptomatic or is manifested as cystitis or pyelonephritis. All strains causing uncomplicated infection express the FimH adhesin, but this adhesin is not specific for UTI. *E. coli* isolated from uncomplicated pyelonephritis are characterized by the presence of the P fimbria adhesin Gal(α1-4) Galβ disaccharide globoside, which initiates mucosal inflammation. Adherence of organisms in the bladder or kidney activates the innate immune

response leading to release of cytokines, particularly interleukin-6 and interleukin-8, and mobilizes leukocytes. This results in pyuria and local or systemic symptoms, including fever in patients with pyelonephritis.

The occurrence of acute uncomplicated UTI in healthy premenopausal women is determined by both genetic and behavioral factors. A genetic predisposition is supported by observations that a history of prior UTI is consistently one of the strongest associations with recurrent uncomplicated UTI, and women who experience these infections report a higher proportion of first-degree female relatives with recurrent UTI than do those without infection. One established genetic association is being a nonsecretor of the ABH blood group antigen. There is more avid binding of uropathogenic organisms to the vaginal epithelium in nonsecretors. Genetic polymorphisms affecting the innate immune response have been correlated with increased frequency of infection as well as with specific presentations.[1] The strongest behavioral factors associated with uncomplicated UTI are sexual intercourse and spermicide use. For sexually active premenopausal women, 75 to 90% of episodes are attributed to intercourse. The normal lactobacillus flora of the vagina maintains an acid environment that prevents colonization by potential uropathogens, and spermicide suppresses these organisms. Use of the birth control pill or condoms, postcoital voiding, type of underwear used, personal hygiene after voiding or defecating, and taking a bath rather than a shower are not associated with an increased risk for recurrent episodes of UTI, despite popular perceptions. Behavioral risk factors are similar for women for all clinical presentations—asymptomatic bacteriuria, cystitis, or pyelonephritis. Sexual intercourse is not a major contributor to UTI in postmenopausal women. The most important determinants of infection in these women are having a history of UTI at a younger age and nonsecretory status.

The risk of complicated UTI is determined by the underlying abnormality. Genitourinary abnormalities facilitate infection through increased entry of organisms into the bladder, such as intermittent catheterization or urologic procedures, and persistence of organisms within the urinary tract because of incomplete voiding or in biofilm on urologic devices. The determinants that promote symptomatic rather than asymptomatic infection are not well characterized. However, obstruction and mucosal trauma with bleeding are well-recognized antecedents for bacteremia and sepsis in patients with preexisting bacteriuria. Although it is generally accepted that patients with diabetes have an increased incidence of UTI, this correlates with long-term complications of diabetes, such as neurogenic bladder, rather than with diabetes.[2] Patients with poorly controlled diabetes, however, are at risk for more severe manifestations of infection.

Acquisition of bacteriuria in individuals with indwelling urinary devices, including indwelling catheters, stents, and nephrostomy tubes, is primarily attributable to biofilm development along the device.[3] Biofilm is composed of an extracellular polysaccharide material produced by the organisms that incorporates urine components including Tamm-Horsfall protein and magnesium or calcium ions. After insertion of the device, a conditioning layer composed of proteins and other host components immediately coats the device. Organisms adhere to this conditioning layer and initiate biofilm formation. Colonization usually begins at the urethral orifice or in the drainage bag, and the biofilm then ascends the catheter. Organisms growing in the biofilm persist in an environment relatively protected from antibiotics or host defenses. For patients with an indwelling catheter, the acquisition of bacteriuria occurs at a rate of 3 to 7% per day. The initial episode that follows indwelling catheter insertion is usually with a single organism, but polymicrobial flora is the norm in mature biofilms on chronic indwelling devices. *Proteus mirabilis* is a particularly important organism for biofilm formation on chronic devices. These strains may produce copious biofilm, and urease production creates an alkaline environment leading to precipitation of calcium and magnesium ions. This creates a "crystalline biofilm" that is similar to the material causing infection stones and may cause catheter obstruction. About 80% of episodes of urinary catheter obstruction are attributed to *P. mirabilis*.

ETIOLOGY

Table 284-2 summarizes the most common infecting organisms. In all types of UTI, *E. coli* is the dominant bacterial species, causing up to 85% of all symptomatic UTIs in women with community-acquired infections.[4] The second most common species causing uncomplicated cystitis is *Staphylococcus saprophyticus*, which is isolated more frequently in later summer and early fall. In patients with recurrent complicated UTI, species such as *Enterococcus faecalis*, *Enterococcus faecium*, *Klebsiella* species, *Proteus* species, *Providencia stuartii*, and *Morganella morganii* become more common. Patients with very frequent recurrences or with bladder catheters, particularly those in hospitals and nursing homes where antimicrobials are frequently used, may have *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, *Serratia marcescens*, and *Stenotrophomonas maltophilia* isolated. In such patients, *E. coli* accounts for less than 50% of infections. Urolithiasis attributed to infection stones is associated with urease-producing organisms; the alkaline urine created facilitates struvite formation.[5] Patients with frequent recurrent complicated UTI, including those with chronic indwelling catheters, often acquire organisms of increased antimicrobial resistance because of repeated exposure to antimicrobials. In health care facilities, the catheterized urinary tract is the most common site of isolation of multiply drug–resistant gram-negative organisms, including extended-spectrum β-lactamase–producing and carbapenemase-producing Enterobacteriaceae. *Candida* sp is the most common fungal UTI. Patients with *Candida* infection are characterized by the presence of diabetes or of an indwelling urinary catheter and broad-spectrum antimicrobial exposure.[6]

TABLE 284-2 MICROBIAL ETIOLOGY OF URINARY TRACT INFECTIONS

ORGANISMS	CLINICAL CHARACTERISTICS
GRAM-NEGATIVE BACTERIA	
Escherichia coli	Typical
Klebsiella pneumoniae	Often reinfection
Enterobacter spp	Often reinfection or health care–associated infection*
Proteus spp	May indicate calculi; frequent with devices
Providencia stuartii	Often reinfection or health care–associated infection*
Morganella morganii	Often reinfection or health care–associated infection*
Serratia marcescens	Often health care–associated infection*
Acinetobacter baumannii	Often health care–associated infection*
Burkholderia spp	Often health care–associated infection*
Pseudomonas aeruginosa	Often health care–associated infection*
Stenotrophomonas maltophilia	Often health care–associated infection*
GRAM-POSITIVE BACTERIA	
Staphylococcus saprophyticus	Most common during late summer and fall
Staphylococcus aureus	May indicate focus outside the genitourinary tract
Enterococcus spp	Often reinfection
Other gram-positive bacteria	In most cases contaminants or colonizers
FUNGI	
Candida spp	May indicate focus outside the genitourinary tract

*Includes hospital and nursing home care.

CLINICAL MANIFESTATIONS

Typical symptoms of cystitis, pyelonephritis, and urosepsis are listed in Table 284-3. The onset of cystitis is rapid, and symptoms usually develop during less than 24 hours. Clinically, the lack of vaginal discharge differentiates cystitis from urethritis caused by chlamydia, ureaplasma, or gonococci. Women who experience recurrent episodes of acute uncomplicated UTI are more than 90% reliable for self-diagnosis.

Pyelonephritis may also have a rapid onset and may or may not be associated with cystitis symptoms. Bacteremia occurs in 10 to 30% of patients but does not have prognostic significance. The typical flank pain and tenderness, resulting from inflammation and edema of the renal parenchyma, may be masked by the intake of analgesic drugs such as acetaminophen, which may also reduce the fever. An important differential diagnosis is renal calculus. This may have a similar location of pain but does not cause fever unless it is complicated by infection.

Complicated UTI is manifested across a clinical spectrum from minimal voiding abnormalities to symptoms consistent with cystitis, pyelonephritis, or severe sepsis. Urosepsis is a life-threatening condition, usually associated with bacteremia.[7] Obstruction or trauma to the mucosa from an indwelling catheter or with urologic surgery may precipitate bacteremia. Patients who present with urosepsis invariably have complicated UTI rather than nonobstructive pyelonephritis.

TABLE 284-3 CLINICAL SYMPTOMS OF URINARY TRACT INFECTIONS

TYPE OF URINARY TRACT INFECTION	TYPICAL SIGNS OR SYMPTOMS
Cystitis	Frequency Dysuria Urgency Stranguria (difficulty in micturition) Suprapubic pain Hematuria or cloudy urine
Pyelonephritis	Costovertebral angle pain or tenderness Fever Chills Cystitis symptoms (may be absent)
Urosepsis	Fever Chills, rigors Sepsis syndrome

TABLE 284-4 INTERPRETATION OF THE QUANTITATIVE URINE CULTURE

	QUANTITATIVE BACTERIAL COUNT
Asymptomatic bacteriuria	$\geq 10^5$ CFU/mL (two consecutive specimens for women)
Acute uncomplicated cystitis	$\geq 10^2$ CFU/mL of *Escherichia coli* or *Staphylococcus saprophyticus*
Acute uncomplicated pyelonephritis	$\geq 10^4$ CFU/mL (95% have $\geq 10^5$ CFU/mL)
Complicated urinary tract infection	$\geq 10^5$ CFU/mL (lower counts may occur with diuresis)
Intermittent or in and out catheter collection	$\geq 10^2$ CFU/mL
Suprapubic or percutaneous aspiration	Any organisms isolated

DIAGNOSIS

Laboratory Findings

The hallmark of UTI diagnosis is demonstration of bacteriuria in a urine sample that has been incubated in the bladder for at least 2 hours to allow the growth of bacteria. A pretherapy urine culture specimen should be obtained for all patients presenting with pyelonephritis, urosepsis, or complicated UTI or when the diagnosis is uncertain. It is not generally recommended for acute uncomplicated cystitis because the clinical presentation is characteristic, and use of empirical short-course therapy means symptoms are often resolved by the time the culture is available. The urine specimen must be collected in a manner that will limit contamination. The usual collection method is a midstream urine sample. All specimens should be taken promptly to the laboratory to prevent growth during transportation. Urine specimens for culture must be collected before institution of antimicrobial therapy as the urine is rapidly sterilized after initiation of systemic antimicrobials.

Interpretation of the quantitative urine culture varies with the clinical presentation and collection method (Table 284-4). *Significant bacteriuria* is usually 10^5 CFU/mL or more, where one colony-forming unit (CFU) is one or more bacterial cells forming a colony when growing on an agar plate. In women with symptoms of uncomplicated cystitis, 10^2 CFU/mL or more in midstream urine of *E. coli* or *S. saprophyticus* is consistent with infection. Other gram-positive organisms at any quantitative count should be interpreted as contaminants.[8]

Pyuria is present in most patients with symptomatic UTI or asymptomatic bacteriuria. Many other abnormalities are, however, associated with pyuria, and the presence of pyuria does not diagnose infection or differentiate symptomatic from asymptomatic infection. The absence of pyuria has a high negative predictive value to exclude UTI for most patients. However, the absence of pyuria in a urine specimen from a woman with symptoms compatible with cystitis is not an indication to withhold empirical antimicrobial therapy.

For screening of bacteriuria, identification of nitrite in the urine may be useful. Gram-negative bacteria, with the exception of *P. aeruginosa*, will metabolize nitrate to nitrite, which can be demonstrated by a color reaction on a dipstick. Gram-positive bacteria and fungi do not metabolize nitrate. The technique is rapid (<1 minute) and inexpensive. It has a high degree of specificity but is insensitive because it does not detect infections caused by gram-positive organisms.

TABLE 284-5 DECISION PROCESS FOR DIAGNOSIS AND TREATMENT OF UPPER (PYELONEPHRITIS) VERSUS LOWER (CYSTITIS) URINARY TRACT INFECTIONS

	CYSTITIS	PYELONEPHRITIS
SIGNS AND SYMPTOMS		
Fever	No	Yes
Dysuria	Yes	May be present
Frequency	Yes	May be present
Flank pain	No	Yes
DIAGNOSIS		
Pyuria	Yes	Yes
Nitrite test result	Normally positive	Normally positive
Bacteriuria	Yes	Yes
C-reactive protein	Normal	Increased
Blood cultures	Negative	Positive in ≈10-30%
TREATMENT		
First line	Short-term oral therapy (Table 284-6)	Oral: fluoroquinolone for 7 days Parenteral: cephalosporin, fluoroquinolone, or aminoglycoside for 7-14 days (Table 284-7)
Second line	Fluoroquinolone for 3 days or cephalosporin for 7 days	Injectable cephalosporin until afebrile, followed by oral step-down for total of 2 weeks
Pregnant women	Nitrofurantoin or cephalosporin for 5-7 days	Injectable cephalosporin until afebrile, followed by oral cephalosporin for 14 days

Blood culture specimens should be obtained in all patients with suspected urosepsis. Patients with acute pyelonephritis, but not those with acute cystitis, have increased serum levels of C-reactive protein.

Imaging

Early imaging should be performed in any patient with urosepsis to identify abnormalities that require immediate source control. The optimal imaging modality is an infused computed tomography scan. Magnetic resonance imaging may not identify gas in tissues or small stones. Ultrasound may provide a rapid examination to exclude significant obstruction. Investigations would also be indicated for patients with delayed response or failure to respond to appropriate antimicrobial therapy or if there is early relapse of pyelonephritis after completion of therapy. The optimal management of complicated urinary infection requires characterization of underlying abnormalities and correction of these, whenever possible. Selected patients may require studies for diagnosis of vesicoureteral reflux or to characterize differential renal function.

Differential Diagnosis

Clinical manifestations will usually differentiate acute cystitis and acute pyelonephritis (Table 284-5). New-onset frequency, dysuria, and urgency without accompanying vaginal discharge or pain have a positive predictive value of 90% for acute cystitis. The differential diagnosis for women presenting with acute irritative lower tract symptoms includes sexually transmitted infections, vulvovaginal candidiasis, and noninfectious causes such as interstitial nephritis. Some patients who present with only lower tract symptoms may have renal infection, referred to as occult pyelonephritis. Patients with appendicitis and cholecystitis can present with flank pain similar to right-sided pyelonephritis, and pelvic inflammatory disease may be misdiagnosed as urinary infection.

TREATMENT

Rx

Symptomatic UTIs should be treated with antimicrobials to decrease the duration of symptoms and, for pyelonephritis, to limit damage to renal tissue. Antimicrobials selected for treatment should be excreted renally, so

TABLE 284-6 ANTIMICROBIALS USED TO TREAT CYSTITIS

ANTIMICROBIAL	DOSE* AND DURATION
FIRST-LINE THERAPY	
Trimethoprim	100-150 mg q12h for 3 days
Trimethoprim-sulfamethoxazole	80/400 mg q12h for 3 days or 320/1600 mg single dose
Nitrofurantoin	50 mg q8h for 5-7 days
Nitrofurantoin macrocrystals	100 mg bid for 5 days
Fosfomycin trometamol	3 g single dose
Pivmecillinam	400 mg bid for 3-5 days
OTHER	
Amoxicillin-clavulanate	500 mg (amoxicillin dose) q8h for 7 days
Amoxicillin	500 mg tid for 7 days
Cefpodoxime proxetil	100 mg bid for 3 days
Cefuroxime axetil	500 mg bid for 7 days
Cefixime	400 mg/day for 7 days
Ceftibuten	400 mg/day for 5-7 days
Norfloxacin	400 mg q12h for 7 days
Ciprofloxacin	250 mg q12h for 7 days (500 mg qd extended release)
Levofloxacin	250-500 mg/day for 7 days
Doxycycline	100 mg bid for 7 days

*Doses given are for adults with normal renal function. The need to reduce dosages because of renal impairment related to infection in the kidneys, other renal diseases, or advanced age should always be considered.

high antimicrobial concentrations are achieved in the renal parenchyma and urine.

Cystitis

Table 284-6 lists recommended choices for the antimicrobial treatment of cystitis.[9] The shortest effective treatment duration for the antimicrobial should be used. Trimethoprim, trimethoprim-sulfamethoxazole, fosfomycin, pivmecillinam, and nitrofurantoin are recommended first-line treatments because they are effective with relatively short courses, and because there is limited impact on normal flora, resistance emergence is not a concern. Trimethoprim or trimethoprim-sulfamethoxazole should be selected for initial empirical therapy only if the local prevalence of resistance to these agents in community-acquired *E. coli* infections is less than 20%. Fluoroquinolones are not recommended for first-line therapy as widespread use may lead to the emergence of resistance. β-Lactam antimicrobials are about 10% less effective than the first-line agents. Patients with recurrent cystitis can be effectively managed with a strategy of early self-treatment. Early empirical treatment usually leads to prompt improvement of symptoms. Nitrofurantoin and oral cephalosporins are preferred therapy for pregnant women as these are safe for the fetus.

Pyelonephritis

For antimicrobial treatment of pyelonephritis, the initial decision is whether parenteral treatment is needed or whether oral treatment alone will suffice. Table 284-7 lists antimicrobials suitable for the treatment of pyelonephritis. After initial treatment with a parenteral drug, a transition to oral treatment is normally possible at 24 to 48 hours if the patient has clinically improved. The recommended treatment time is 7 to 14 days.

Complicated Urinary Infection

The antimicrobial regimen selected for treatment of complicated UTI is individualized on the basis of the site of infection, severity of the manifestations, new or presumed infecting organism and susceptibility, tolerance of the patient, and nature of the underlying abnormalities.[10] When symptoms are mild, it is preferable to delay initiation of antimicrobial therapy until results of urine culture are available to allow optimal antimicrobial selection. Empirical antimicrobial therapy should be initiated when severe symptoms are present. Oral or parenteral therapy is selected on the basis of the presentation and the likelihood of resistant organisms. Previous urine culture results from the patient and recent history of antimicrobial exposure are helpful to assess the likelihood of resistant organisms. Nitrofurantoin may be used for episodes of bladder infection but is not effective for renal infection and is contraindicated in individuals with renal failure. The empirical therapy selected should be reassessed after 48 to 72 hours, by which time the urine culture result should be available and the response to initial therapy can be assessed. If the organism

TABLE 284-7 ANTIMICROBIALS USED TO TREAT PYELONEPHRITIS

ROUTE OF ADMINISTRATION AND ANTIMICROBIAL	DOSE* AND DURATION
PARENTERAL	
First-Line Therapy	
Gentamicin	4.5 mg/kg/day × 10-14 days
Tobramycin	4.5 mg/kg/day × 10-14 days
Ciprofloxacin	400 mg q12h × 7 days
Levofloxacin	750 mg/day × 5 days
Cefotaxime	1 g q8h × 10-14 days
Ceftriaxone	1-2 g qd × 10-14 days
Other	
Ceftazidime	1 g q12h × 10-14 days
Ertapenem	1g qd × 10-14 days
Meropenem	500 mg q6h × 10-14 days
Piperacillin-tazobactam	3-375 g q6h × 10-14 days
Doripenem	500 mg q8h × 10-14 days
Amikacin	15 mg/kg/day × 10-14 days
Trimethoprim-sulfamethoxazole	160/800 mg q12h × 14 days
ORAL	
First-Line Therapy	
Ciprofloxacin	500 mg q12h × 7 days
Levofloxacin	250-500 mg/day × 5-7 days
Other	
Amoxicillin-clavulanate	500 mg (amoxicillin dose) q8h × 14 days
Cefuroxime axetil	500 mg q12h × 14 days
Cefixime	400 mg/day × 14 days
Ceftibuten	400 mg/day × 14 days
Cefepime	2 g q12h × 14 days

*Doses given are for adults with normal renal function. The need to reduce dosages because of renal impairment should always be considered.

isolated from a pretherapy urine culture specimen is resistant to the empirical antimicrobial therapy initiated, the antimicrobial regimen should be modified to an agent to which the organism is susceptible, irrespective of the clinical response.[11] Antibiotic prophylaxis at the time of catheter removal can reduce the risk of subsequent symptomatic infection.

Asymptomatic Bacteriuria

Asymptomatic bacteriuria should be treated only in pregnant women[12] or when antimicrobials are given as perioperative prophylaxis before a urologic procedure with trauma to the genitourinary mucosa. For all other populations, including elderly women,[13] treatment of asymptomatic bacteriuria has not been associated with improved outcomes but is uniformly followed by reinfection with organisms of increasing antimicrobial resistance. For some populations, evidence suggests that asymptomatic bacteriuria may protect subjects from symptomatic UTI. Bacteriuria in patients with indwelling catheters should not be treated unless the patient has symptoms attributed to urinary infection. Administration of antimicrobials to catheterized patients with asymptomatic bacteriuria inevitably results in reinfection with more resistant organisms.

Urosepsis

The principles of management of urosepsis are similar to those for patients with severe sepsis from any site. Parenteral empirical antimicrobial treatment and supportive care should be initiated promptly. The antimicrobial selected should provide broad-spectrum coverage for potential uropathogens, including resistant bacteria. Antimicrobial therapy is reassessed when urine and blood culture results become available and the specific infecting organism and susceptibilities are identified.

Funguria

Funguria in catheterized patients should be treated only when there is a symptomatic UTI. Symptomatic infection is treated with fluconazole 400 mg

TABLE 284-8 PROPHYLACTIC REGIMENS TO PREVENT RECURRENT URINARY TRACT INFECTION IN WOMEN

PREFERRED	OTHER
Long-term low dose	
Nitrofurantoin 50 mg od or 100 mg qd	Cephalexin 250-500 mg qd*
Trimethoprim-sulfamethoxazole, 40/200 mg qd or every other day	Norfloxacin 200 mg qd
	Ciprofloxacin 125 mg qd
Postcoital (single dose)	
Nitrofurantoin 50 or 100 mg*	Cephalexin 250 mg*
Trimethoprim-sulfamethoxazole 40/200 mg	Ciprofloxacin 125 mg
Trimethoprim 100 mg	Norfloxacin 200 mg

*Suitable for use in pregnancy.

once daily for 1 day, followed by 200 mg once daily for 7 to 14 days. If *Candida* sp resistant to fluconazole is isolated, amphotericin B deoxycholate is the recommended alternative therapy as other antifungals have limited renal excretion.

Follow-up

Patients do not require follow-up urine cultures unless symptomatic infection persists or recurs. When there is early (<30 days) symptomatic recurrence, the infecting organism should be re-evaluated to ensure that it was susceptible to the antimicrobial given.

PREVENTION

Premenopausal women with recurrent acute uncomplicated UTI should avoid spermicide use. For women with frequent recurrent acute uncomplicated UTI (more than two in 6 months or three in 12 months) presenting as either cystitis or pyelonephritis, prophylactic antimicrobial therapy is effective. This may be given as long-term low-dose prophylaxis or as postintercourse prophylaxis. Suggested regimens for prophylactic antimicrobial therapy are provided in Table 284-8. The use of cranberry tablets or juice does not reliably decrease the frequency of recurrent infection,[A6] and probiotics are not effective. For postmenopausal women, use of topical vaginal estrogens may decrease the frequency of infection.[A5] The use of systemic estrogen, however, is associated with an increased frequency of UTI. Prophylactic antimicrobial therapy is more effective than topical vaginal estrogen in these women.

Pregnant women should be screened for asymptomatic bacteriuria early in the pregnancy, usually at 12 or 16 weeks. If bacteriuria is present, these women should be treated and have subsequent follow-up culture specimens obtained monthly. If either asymptomatic or symptomatic recurrent infection occurs, prophylactic antimicrobial therapy with either cephalexin or nitrofurantoin should be continued through the duration of the pregnancy to decrease the risk for development of pyelonephritis in later pregnancy.

Prophylactic antimicrobial therapy has not been shown to be effective for patients with complicated UTI, including those with spinal cord injury or with chronic indwelling catheters. In these patients, the abnormality leading to impaired voiding means that bacteriuria is unavoidable, and antimicrobial therapy simply promotes bacteriuria with increasingly resistant organisms.

Infection control programs of health care facilities should include practices to prevent catheter-acquired UTI. Evidence-based guidelines provide clear recommendations for program components, including ongoing surveillance.[14] The most important intervention is to avoid the use of an indwelling catheter wherever possible and, when there are clear indications for catheter use, to limit the duration to as short a time as possible. The ultimate solution to the problem of catheter-acquired UTI, however, will require development of biofilm-resistant materials.[15]

PROGNOSIS

The prognosis of uncomplicated cystitis and pyelonephritis is good. Women with acute uncomplicated cystitis who do not receive antimicrobial therapy will usually have resolution of symptoms by 1 to 2 weeks, and about half will be culture negative by 6 weeks. Women with even very frequent recurrent acute uncomplicated UTI[16] experience no long-term adverse outcomes, such as renal impairment or hypertension. A small proportion of women with severe presentations of acute nonobstructive pyelonephritis develop renal scars, but these are not associated with impaired renal function. Patients with frequent recurrent complicated UTI may experience substantial morbidity with recurrent infections, but poor long-term medical outcomes are usually determined by the underlying abnormality rather than by infection. Patients with urosepsis have a fatality rate of about 10%. Factors increasing the risk of death are advanced age and significant underlying diseases as well as inadequate initial antimicrobial treatment.

Grade A References

A1. Grigoryan L, Trautner BW, Gupta K. Diagnosis and management of urinary tract infections in the outpatient setting: a review. *JAMA*. 2014;312:1677-1684.

A2. Hooton TM, Roberts PL, Stapleton AE. Cefpodoxime vs ciprofloxacin for short course treatment of acute uncomplicated cystitis. *JAMA*. 2012;307:583-589.

A3. Vazquez JC, Abalos E. Treatments for symptomatic urinary tract infections during pregnancy. *Cochrane Database Syst Rev*. 2011;1:CD002256.

A4. Sandberg T, Skoog G, Hermansson AB, et al. Ciprofloxacin for 7 days versus 14 days in women with acute pyelonephritis: a randomised, open-label and double-blind, placebo-controlled, non-inferiority trial. *Lancet*. 2012;380:484-490.

A5. Marschall J, Carpenter CR, Fowler S, et al. Antibiotic prophylaxis for urinary tract infections after removal of urinary catheter: meta-analysis. *BMJ*. 2013;346:f3147.

A6. Jepson R, Craig J, Williams G. Cranberry products and prevention of urinary tract infections. *JAMA*. 2013;310:1395-1396.

A7. Beerepoot MA, Geerlings SE, van Haarst EP, et al. Nonantibiotic prophylaxis for recurrent urinary tract infections: a systematic review and meta-analysis of randomized controlled trials. *J Urol*. 2013;190:1981-1989.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ragnarsdottir B, Lutay N, Gronberg-Hernandez J, et al. Genetics of innate immunity and UTI susceptibility. *Nat Rev Urol*. 2011;8:449-468.
2. Nicolle LE. Urinary tract infections in special populations: diabetes, renal transplant, HIV infection, and spinal cord injury. *Infect Dis Clin North Am*. 2014;28:91-104.
3. Nicolle LE. Urinary catheter–associated infections. *Infect Dis Clin North Am*. 2012;26:13-28.
4. Hooton TM. Clinical practice. Uncomplicated urinary tract infection. *N Engl J Med*. 2012;366:1028-1037.
5. Brown PD. Management of urinary tract infections associated with nephrolithiasis. *Curr Infect Dis Rep*. 2010;12:450-454.
6. Kauffman CA. Diagnosis and management of fungal urinary tract infection. *Infect Dis Clin North Am*. 2014;28:61-74.
7. Nicolle LE. Urinary tract infection. *Crit Care Clin*. 2013;29:699-716.
8. Hooton TM, Roberts PL, Cox ME, et al. Voided midstream urine culture and acute cystitis in premenopausal women. *N Engl J Med*. 2013;369:1883-1891.
9. Gupta K, Hooton TM, Naber KG, et al. Executive summary: international clinical practice guidelines for the treatment of acute uncomplicated cystitis and pyelonephritis in women: a 2010 update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. *Clin Infect Dis*. 2011;52:561-564.
10. Dielubanza EJ, Mazur DJ, Schaeffer AJ. Management of non-catheter-associated complicated urinary tract infection. *Infect Dis Clin North Am*. 2014;28:121-134.
11. Gupta K, Bhadelia N. Management of urinary tract infections from multidrug resistant organisms. *Infect Dis Clin North Am*. 2014;28:49-60.
12. Nicolle LE. Asymptomatic bacteriuria. *Curr Opin Infect Dis*. 2014;27:90-96.
13. Mody L, Juthani-Mehta M. Urinary tract infections in older women: a clinical review. *JAMA*. 2014;311:844-854.
14. Gould CV, Umscheid CA, Agarwal RK, et al. Guideline for prevention of catheter-associated urinary tract infections 2009. *Infect Control Hosp Epidemiol*. 2010;31:319-326.
15. Siddiq DM, Darouiche RO. New strategies to prevent catheter-associated urinary tract infections. *Nat Rev Urol*. 2012;9:305-314.
16. Gupta K, Trautner BW. Diagnosis and management of recurrent urinary tract infections in non-pregnant women. *BMJ*. 2013;346:f3140.

REVIEW QUESTIONS

1. A 32-year-old gravida 2 para 1 woman presents to the obstetric emergency department at 29 weeks' gestation with temperature of 39.2° C and right costovertebral angle pain and tenderness. There is associated vomiting; the blood pressure is stable. She has no history of prior urinary tract infection and has received no antimicrobial therapy during this pregnancy. Intravenous rehydration therapy is initiated, and blood and urine culture specimens are obtained. What is the most appropriate next step?

A. Await results of cultures before initiating antimicrobial therapy.
B. Initiate therapy with ciprofloxacin, 500 mg PO bid.
C. Initiate therapy with ceftriaxone, 1 to 2 g IV q24h.
D. Initiate therapy with levofloxacin, 750 mg IV q24h.
E. Await results of computed tomography scan before initiating antimicrobial therapy.

Answer: C Pregnant women presenting with acute pyelonephritis at this time of gestation are at high risk for premature labor and delivery at a vulnerable stage for the fetus. Empirical antimicrobial therapy should be initiated promptly, and initial therapy should be parenteral. Fluoroquinolone antimicrobials are contraindicated for pregnant women because of concerns with fetal cartilage development. There is no suggestion from her history that she has an increased risk for a resistant organism, and parenteral ceftriaxone would be the treatment of choice. (Vazquez JC, Abalos E. Treatments for symptomatic urinary tract infections during pregnancy. *Cochrane Database Syst Rev.* 2011;1:CD002256.)

2. A 22-year-old woman presents to her family physician with a second episode consistent with acute cystitis in the past 6 months. The previous episode responded promptly to empirical short-term antimicrobial therapy. She is sexually active in a stable monogamous relationship. Empirical antimicrobial therapy is again prescribed for treatment of this episode. What additional advice should she be given with respect to behavioral modification to decrease her risk of subsequent recurrent infection?

A. Ensure bladder voiding immediately after sexual intercourse.
B. Discontinue use of spermicide birth control method.
C. After bowel movements, she should wipe herself from front to back.
D. Increase her intake of cranberry juice.
E. Use showers rather than baths for personal hygiene.

Answer: B The only behavioral risk factor, other than sexual intercourse, that is modifiable is discontinuation of use of spermicide for birth control. Use of spermicide for birth control is associated with at least a two-fold higher risk for recurrent urinary tract infection in sexually active young women. Whereas early studies suggested a modest benefit of cranberry capsules, more recent studies report no significant improvements with the use of cranberry products. The other interventions suggested have consistently been reported not to be associated with recurrent urinary tract infection. (Gupta K, Hooton TM, Naber KG, et al. Executive summary: international clinical practice guidelines for the treatment of acute uncomplicated cystitis and pyelonephritis in women: a 2010 update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. *Clin Infect Dis*. 2011;52:561-564; and Beerepoot MA, Geerlings SE, van Haarst EP, et al. Nonantibiotic prophylaxis for recurrent urinary tract infections: a systematic review and meta-analysis of randomized controlled trials. *J Urol.* 2013;190:1981-1989.)

3. A 42-year-old man with a high-level cervical spinal cord injury was admitted to the rehabilitation unit 2 weeks ago. There is an indwelling catheter in place for bladder management. He has a single temperature elevation to 38.2° C with no localizing findings. Complete blood count and urine and blood culture specimens are obtained. The white blood cell count was normal, and after 48 hours, the blood culture is negative but the urine specimen has returned growing *E. coli* of more than 10^8 CFU/mL and an enterococcus of more than 10^8 CFU/mL. He remains afebrile and otherwise stable. What is the appropriate approach to management at this time?

A. Remove the indwelling catheter and begin intermittent catheterization for management of bladder emptying.
B. Obtain ultrasound and urodynamic studies to exclude urolithiasis.
C. Institute antimicrobial therapy for treatment of the organisms isolated from the urine culture.
D. No changes in management are required.
E. Request urologic consultation.

Answer: D This patient has a chronic indwelling catheter and will be bacteriuric at any time. Thus, the urine culture would be expected to be positive. In the absence of clinical evidence of infection, such as elevated white blood cell count or continuing fever, this is asymptomatic bacteriuria, and antimicrobial therapy is not indicated. It is desirable to remove the indwelling catheter as soon as possible, but in a patient with a high-level cervical spine injury, intermittent catheterization by the patient himself is unlikely to be feasible. All spinal cord–injured patients require urodynamic assessment, but this single fever and positive urine culture would not be an indication for evaluation outside normal management. (Nicolle LE. Urinary tract infections in special populations: diabetes, renal transplant, HIV infection, and spinal cord injury. *Infect Dis Clin North Am.* 2014;28:91-104.)

4. A 63-year-old man presents with temperature of 39.2° C associated with rigors; blood pressure is 70/40. He was diagnosed with a ureteric stone complicated by ureteric obstruction 36 hours previously and 12 hours before fever onset had successful endourologic intervention to extract the stone. The urine culture specimen collected before the procedure is growing *E. coli* of more than 10^8 CFU/mL, but susceptibilities are not yet available. Ciprofloxacin 500 mg PO bid had been initiated before the procedure. What is the appropriate approach at this time?

A. Continue the ciprofloxacin initiated, pending susceptibility information.
B. Change to oral cefixime, 400 mg once daily, to cover potential resistant organisms.
C. Initiate broad-spectrum parenteral antimicrobial therapy effective for ciprofloxacin-resistant *E. coli* immediately.
D. Initiate antimicrobial therapy with gentamicin plus ampicillin to cover additional pathogens.
E. Give a bolus of normal saline and reassess the patient in 4 to 6 hours.

Answer: C This patient has a presumptive diagnosis of postsurgical sepsis. It is essential that effective antimicrobial therapy, together with other supportive therapy, be initiated immediately. When patients progress to septic shock, one variable associated with mortality is delayed institution of effective antimicrobial therapy. In this case, it is possible that the patient has a ciprofloxacin-susceptible organism and the obstruction prevented access of antimicrobial to the infected site. It is also possible that a new organism was introduced at the time of the surgical procedure. However, as the organism may be resistant to ciprofloxacin, broad-spectrum parenteral therapy to cover potentially resistant organisms should be selected. This could be piperacillin-tazobactam or a carbapenem. Gentamicin and ampicillin are usually also reasonable choices. However, given the likelihood of resistant organisms, this regimen may not provide full coverage, and broader spectrum agents seem more appropriate. Blood and urine cultures should be repeated before initiation of the antimicrobial therapy. Obviously, the antimicrobial therapy should be reassessed and appropriate modifications made once the culture results and susceptibilities are available. (Nicolle LE. Urinary tract infection. *Crit Care Clin*. 2013;29:699-716.)

5. An 82-year-old woman with poorly controlled type 2 diabetes is admitted to the intensive care unit with respiratory failure complicating pneumococcal pneumonia. An indwelling urethral catheter was inserted for output monitoring on admission. Because of initial uncertainty about the etiology of her pneumonia, she was treated empirically with broad-spectrum therapy of meropenem and azithromycin. These have been continued, despite *S. pneumoniae* isolated in blood cultures being susceptible to penicillin G. Seven days after admission, the urine is observed to be cloudy, and the nurse sends a urine specimen for culture. It returns with a report of isolation of *Candida albicans* of more than 10^8 CFU/mL. The patient is afebrile and, while still requiring ventilatory support, is showing improvement in oxygenation. What is the appropriate response to the urine culture report?

A. Institute fluconazole 400 mg once daily for *Candida* urinary tract infection.
B. Obtain a blood culture sample for yeast species.
C. No interventions are required.
D. Review the need for the indwelling catheter and remove it if possible.
E. Repeat the urine culture to confirm *Candida* funguria.

Answer: D This patient has diabetes, is receiving broad-spectrum antimicrobial therapy, and has an indwelling catheter in situ. These are the three most important risk factors for acquisition of candiduria. There is no clinical evidence to suggest symptomatic infection from a urinary source. The diagnosis is asymptomatic candiduria, and antifungal treatment is not indicated. For any patient with an indwelling urethral catheter, the need for continuing use of the catheter should be reassessed on a daily basis and the catheter removed as soon as it is no longer necessary to limit development of complications. If the indwelling catheter cannot be removed in this patient, there is no indication for other interventions relevant to the candiduria as long as the patient is clinically improving. (Kauffman CA. Diagnosis and management of fungal urinary tract infection. *Infect Dis Clin North Am*. 2014; 28:61-74.)

285

APPROACH TO THE PATIENT WITH A SEXUALLY TRANSMITTED INFECTION

HEIDI SWYGARD AND MYRON S. COHEN

SEXUALLY TRANSMITTED INFECTIONS

DEFINITION

Sexually transmitted infections (STIs) include a wide variety of organisms that are transmitted through intimate contact involving skin or mucosal surfaces of the oropharynx, vagina, penis, and rectum. STIs can generally be divided into five broad categories (syndromes): urethritis, genital ulcers, epithelial cell disorders, female vaginal discharge, and ectoparasites (Table 285-1).

ETIOLOGY

The interaction between the host and the STI pathogen plays a critical role, and characteristic tissue changes offer exceptionally strong clues about etiology. Several STI pathogens cause local inflammation only (*Neisseria gonorrhoeae*, *Chlamydia trachomatis*, *Trichomonas vaginalis*), with the potential for local tissue invasion (*N. gonorrhoeae*, *C. trachomatis*) or systemic dissemination (*N. gonorrhoeae*). Some STI pathogens cause tissue ulceration (*Treponema pallidum*, *Haemophilus ducreyi*, herpes simplex viruses 1 and 2). Human papillomaviruses (HPVs) cause epithelial cell changes and predispose to neoplasm. Several STI pathogens (human immunodeficiency virus [HIV], hepatitis B and C viruses, cytomegalovirus) routinely use the genital tract for access without causing any local changes.

EPIDEMIOLOGY

STIs are among the most common infections worldwide, and most are never reported. Each year there are almost 10 million new STIs reported among persons aged 15 to 24 years in the United States alone; many infections are subclinical and may escape detection, suggesting that these numbers are an underestimate. Of great concern, STIs are generally transmissible whether they are symptomatic or asymptomatic.

TABLE 285-1 SYNDROMES OF SEXUALLY TRANSMITTED DISEASES

SYNDROME	ORGANISM
URETHRITIS	
Gonococcal	*Neisseria gonorrhoeae*
Nongonococcal	*Chlamydia trachomatis* *Trichomonas vaginalis* *Mycoplasma genitalium* *Ureaplasma urealyticum* Herpes simplex (primary infection)
GENITAL ULCERS	
Syphilis	*Treponema pallidum*
Genital herpes	Herpes simplex
Chancroid	*Haemophilus ducreyi*
EPITHELIAL CELL INFECTIONS	
Genital warts	Human papillomavirus
Molluscum	Molluscum contagiosum
Cervical neoplasia	Human papillomavirus types 16 and 18
FEMALE GENITAL DISCHARGE	
Cervicitis	*Neisseria gonorrhoeae* *Chlamydia trachomatis* *Trichomonas vaginalis* Herpes simplex
Pelvic inflammatory disease	*Neisseria gonorrhoeae* *Chlamydia trachomatis*
Vaginitis	*Trichomonas vaginalis* *Candida albicans*
Bacterial vaginosis	*Gardnerella vaginalis*, anaerobes
ECTOPARASITES	
Pubic lice	*Phthirus pubis*
Scabies	*Sarcoptes scabiei*

Chlamydia, gonorrhea, and syphilis are reportable to the U.S. Centers for Disease Control and Prevention. The incidence of primary and secondary syphilis declined in the 1990s, but since 2000, rates have been increasing among men (8.2 cases per 100,000 in 2011, a 3.8% increase compared with 2010) and men aged 20 to 29 years (23.4 cases per 100,000). Rates declined 9.1% in women (from 1.1 cases per 100,000 to 1.0 case per 100,000). Men who have sex with men accounted for 72% of all cases of primary and secondary syphilis reported in 2011. Following a nadir of 98.1 cases per 100,000 in 2009 (since reporting for gonorrhea began), gonorrhea rates increased slightly (104.2 cases per 100,000 in 2011). Rates are highest among the 15- to 24-year age group for both sexes.

For other STIs, seroprevalence and national surveys provide data that suggest a high burden of viral STIs. From the 2005 to 2008 National Health and Nutrition Examination Survey (NHANES), herpes simplex virus 2 (HSV-2) was shown to affect 16.2% of the survey sample for those aged 14 to 49 years; of 20- to 49-year-old participants seropositive for HSV-2, only 18.9% had been previously diagnosed. Data from NHANES collected between 2001 and 2004 show an overall prevalence of 3.1% for women, with the highest percentage among African American women (13.3%).

The spread of STIs depends on the organism and the host, the length of time an infected person remains contagious, and the number of people exposed. These parameters have been reduced to the following formula:

$$R_o = B \times D \times C$$

where R_o is the basic reproductive rate of an infection, or the mean number of secondary cases a typical single infected person will cause in a population; B is the efficiency of transmission; D is the duration of infectiousness; and C is the number of sexual partners.

PATHOBIOLOGY

The STI pathogens depend entirely on human-human transmission, although *T. vaginalis* may have some inanimate sources. The efficiency of transmission reflects the infectiousness of the index case (which depends on the concentration and phenotype of the organism in the genital tract) and the susceptibility of the sexual partner (which reflects the resistance of the host, whether it is hereditary, acquired, or innate). Because immunity to STIs is rare, reinfections are common, and vaccine development has been difficult; the only STI vaccines available target hepatitis B and HPV.

STIs produce syndromes precisely because each pathogen has a proclivity for one or more tissues and (when symptomatic) can evoke a predictable inflammatory response. For example, gonococci that infect the male urethra generally produce an intense neutrophil response that leads to a purulent discharge and pain with urination, whereas *C. trachomatis* is less likely to produce such a response in the same tissue and is more likely to produce a mild, watery discharge or no symptoms at all.

STIs serve as markers for sexual risk-taking behavior, so coinfections are common. The detection of an STI should lead to a variety of other (seemingly unrelated) tests. STI pathogens move together: gonorrhea and chlamydia cause urethritis; genital ulcers greatly increase the probability of HIV acquisition.

DIAGNOSIS AND TREATMENT

Syndromic Strategies

For a variety of common STI syndromes, treatment of the index case (based on signs and symptoms) is empirical, as is treatment of sexual partners. This approach reflects the facts that the diagnostic accuracy of some tests is imperfect, coinfection demands concomitant therapy that overrides the search for individual pathogens, and patients who are not treated immediately may not return for therapy. Genital ulcer disease and urethral discharge have high sensitivity and specificity compared with laboratory diagnosis, and empirical therapy is so successful that follow-up care ("proof of cure") is usually unnecessary. However, the vaginal discharge syndrome is far less sensitive or specific in terms of true STI diagnosis.[1] In a study of South African women, almost 90% with a laboratory-confirmed STI diagnosis had no clinical symptoms and would therefore not have been treated.[2]

The syndromic approach is particularly critical in resource-constrained countries or in areas of the United States where laboratory tests are not available or their cost is prohibitive. In the United States, concomitant microbiologic diagnosis is preferred because it (1) confirms the choice of empirical therapy or redirects subsequent care; (2) permits the detection and monitoring of resistance to treatment; and (3) enables specific diagnoses to be reported to public health authorities, which is required by state law for many STIs. However, even when laboratory tests are ordered, the most appropriate agents should be provided empirically at the point of care to resolve infection and to reduce onward transmission.

Relationship of STIs to HIV Infection

Diagnosis of an STI demonstrates increased sexual risk-taking behavior and inconsistent condom use and serves as a marker for potential HIV infection. Any patient undergoing evaluation or treatment for an STI should be tested for HIV infection. Early diagnosis of HIV infection has major personal and public health benefits.

Nevertheless, STIs also contribute to HIV acquisition and hamper efforts at optimal prevention of HIV transmission. Genital ulcers increase HIV replication in the infected person. Genital ulcers disrupt the epithelium and allow the entry of HIV; inflammation caused by ulceration recruits macrophages and lymphocytes, increasing the number of target cells for HIV.

SYNDROMES

Urethritis

Urethritis is characterized by some combination of urethral discharge and dysuria, but prostatitis can cause similar complaints. Urethritis is caused by a limited group of pathogens (see Table 285-1) that may be difficult to visualize microscopically or to grow in culture. Accordingly, empirical therapy is provided to treat a spectrum of potentially causative organisms.

Urethritis is diagnosed when one or more of the following are demonstrated: (1) mucopurulent or purulent urethral discharge, (2) Gram stain of urethral secretions demonstrating five or more leukocytes per oil immersion microscopic field, (3) positive leukocyte esterase test result on first-void urine, or (4) microscopic examination of first-void urine demonstrating 10 or more leukocytes per high-power field. If no discharge can be expressed from the urethral meatus, a calcium alginate swab can be inserted 5 mm into the urethra; the material collected is transferred to a slide by rolling the swab along the glass.

A Gram stain of urethral discharge is a simple and rapid diagnostic test to document both urethritis and gonococcal infection (Chapter 299),

TABLE 285-2 SYNDROMIC TREATMENT OF URETHRITIS

GONOCOCCAL*
Recommended
Ceftriaxone 250 mg injected intramuscularly once, *and* Azithromycin 1 g orally (single dose)
NONGONOCOCCAL
Recommended
Azithromycin 1 g orally (single dose), *or* Doxycycline 100 mg orally twice daily for 7 days
Alternative
Erythromycin base 500 mg orally four times daily for 7 days, *or* Erythromycin ethylsuccinate 800 mg orally four times daily for 7 days, *or* Ofloxacin 300 mg orally twice daily for 7 days, *or* Levofloxacin 500 mg orally once a day for 7 days

*Uncomplicated anorectal and genital disease.

characterized by the detection of leukocytes containing intracellular gram-negative diplococci. Confirmation of gonococcal urethritis does not rule out concomitant infection with *Chlamydia* or *Mycoplasma.* As culture and Gram stain have become less popular or less available, nucleic acid amplification tests that are highly sensitive and specific for the detection of organisms have been used routinely. Nucleic acid amplification tests for gonorrhea, *Chlamydia*, and *Trichomonas* can be applied to first-void urine samples (the meatus is intentionally not cleaned so that the urine is contaminated with these organisms) or urethral swab material. Specific diagnosis may enhance the management of sexual partners, and the results from such tests should be reported to the health department. However, in practice, patients and (in most cases) sexual partners must be treated before the results of these tests are available.

Treatment for urethritis should be initiated as soon as possible after the clinical diagnosis and should be directly observed if feasible (Table 285-2). *N. gonorrhoeae* (Chapter 299) has become resistant to many antimicrobials, including quinolones and oral cephalosporins, which are no longer recommended. Thus, the choice of optimal therapies is limited. Dual therapy with azithromycin and ceftriaxone increases the cure rate of uncomplicated urogenital, anorectal, and pharyngeal gonorrhea. Azithromycin can also be expected to cure most cases of nongonococcal urethritis (NGU), including those caused by *Mycoplasma genitalium*, an increasingly recognized cause of NGU. Some studies suggest that doxycycline is more effective than azithromycin for NGU, but a longer course of treatment dependent on the patient's adherence must also be considered. Currently, there is no commercially available diagnostic test for *M. genitalium*, which complicates the treatment question. In some settings in which *M. genitalium* is a consideration, persistent or recurrent NGU should be treated with moxifloxacin for 7 to 10 days.[3] *T. vaginalis*, which is susceptible to metronidazole or tinidazole, also causes urethritis and should be considered in the face of NGU treatment failure.

Women with urethritis present with some combination of dysuria and pyuria, which must be differentiated from bacterial cystitis. Because treatment for urinary tract pathogens may also resolve sexually transmitted urethritis, the clinician treating a presumed bladder infection should consider an STI as well.

Genital Ulcers

In the United States, HSV-1 and HSV-2 (Chapter 374) and *T. pallidum* are responsible for virtually all the ulcers encountered, and HSV-1 and HSV-2 are by far the most common cause.

LYMPHOGRANULOMA VENEREUM

Lymphogranuloma venereum, caused by a serovar of *C. trachomatis* (Chapter 318), is characterized by local lymph node suppuration, ulceration and subsequent fibrosis, fistula formation, and distal edema. Lymphogranuloma venereum has increasingly been emerging as a concern in HIV-infected men who have sex with men, particularly in Europe.

GENITAL HERPES

Genital herpes usually develops after an incubation period of less than 21 days and arises as clustered vesicles on an erythematous base. The vesicles become pustular and then rupture to form shallow, painful ulcers, which may coalesce. The ulcers heal by crusting over, and the process is usually completed 2 to 3 weeks after the initial lesions. Recurrences proceed through the same stages but generally last only about 5 to 7 days. The first (incident) episode of HSV-2 infection may be accompanied by systemic signs and symptoms including fever and headache, the latter reflecting the spread of HSV to the central nervous system. HSV-2 is the putative cause of Mollaret's recurrent meningitis. It may appear after a primary genital infection or as reactivation. It may also occur in the absence of genital lesions or a known diagnosis of HSV-2 infection.

About 20% of infected individuals manifest the classic genital presentation, 60% have mild and atypical signs and symptoms, and at least 20% are completely asymptomatic. Individuals who have acquired HSV-2 shed the virus approximately 3 to 4% of the time (even while asymptomatic), posing an ongoing risk to sexual partners.

SYPHILIS

The ulcerative lesion of syphilis (Chapter 319)—the chancre—is indurated and painless, and in many cases it escapes detection. Dark-field examination of scrapings suspended in saline from a genital ulcer may reveal motile spirochetes, and this finding is diagnostic. Secondary syphilis results when the spirochetes spread systemically, leading to a characteristic rash, alopecia, oral mucous patches, or condyloma latum. These skin manifestations should prompt testing for syphilis. The serologic screening test of choice for syphilis is based on the formation of antibodies to cardiolipin, a constituent of the spirochetal cell wall (e.g., rapid plasma reagin test, Venereal Disease Research Laboratory [VDRL] test, toluidine red unheated serum test [TRUST]). Confirmatory testing requires the search for an antitreponemal antibody (e.g., microhemagglutination assay–*T. pallidum*, fluorescent treponema antibody test). The anticardiolipin test provides a titer that must be used to monitor the response to treatment.

Some larger commercial laboratories have reversed the order of testing, using an antitreponemal test followed by an anticardiolipin test, which allows automation and may be cost-effective in areas of low endemicity. This represents a change in testing and interpretation and must be done with caution[4] because such an approach cannot immediately separate old, treated infections from new infections (Fig. 285-1).

Later stages of syphilis may be identified only serologically or on pathologic specimens. Late latent syphilis and syphilis of unknown duration are managed similarly. Neurosyphilis can occur at any stage of infection and should be suspected in any patient with a positive serologic test result who also has findings suggestive of nervous system involvement, including ocular and vestibular symptoms.

Worldwide, syphilis infection has been detected in a substantial number of people with recognized or unrecognized HIV infection, especially men who have sex with men. Accordingly, the diagnosis of syphilis requires HIV testing.

CHANCROID

Chancroid (Chapter 301), infection with *H. ducreyi*, produces painful, ragged ulcers and tender inguinal lymphadenopathy, which may be fluctuant. Unlike the lesions of HSV infection, these genital ulcers are likely to vary in size.

Epithelial Cell Infections

HUMAN PAPILLOMAVIRUS

Sexually transmitted HPV infection (Chapter 373) is generally transient and asymptomatic, but some patients develop visible genital warts. These warts are painless, soft, moist, pink or flesh-colored swellings that vary in shape and can be raised or flat, single or multiple, small or large, and sometimes cauliflower shaped. Warts occur in the vulva, vagina, and anus; on the cervix; and on the penis, scrotum, groin, or thigh. Genital warts are diagnosed by visual inspection. Treatment is primarily with topical agents but is generally not curative.

Two oncogenic HPV genotypes (16 and 18) are responsible for at least 85% of all cervical neoplasia, and HPV types 6 and 11 are responsible for most cases of genital warts. HPV vaccines are available that target these genotypes. Papanicolaou smears are recommended for sexually active women beginning at the age of 21 years, regardless of risk factor or age at coitarche; HPV testing, however, is not recommended for women younger than 30 years. Frequency of rescreening is influenced by age, previous screening results, and HPV results. An increasing incidence of oral HPV is associated with head and neck squamous cell carcinomas, with more than 70% of U.S. cases demonstrating detectable HPV. Male gender, smoking, and HIV-positive status are significantly associated with prevalent oral HPV.[5]

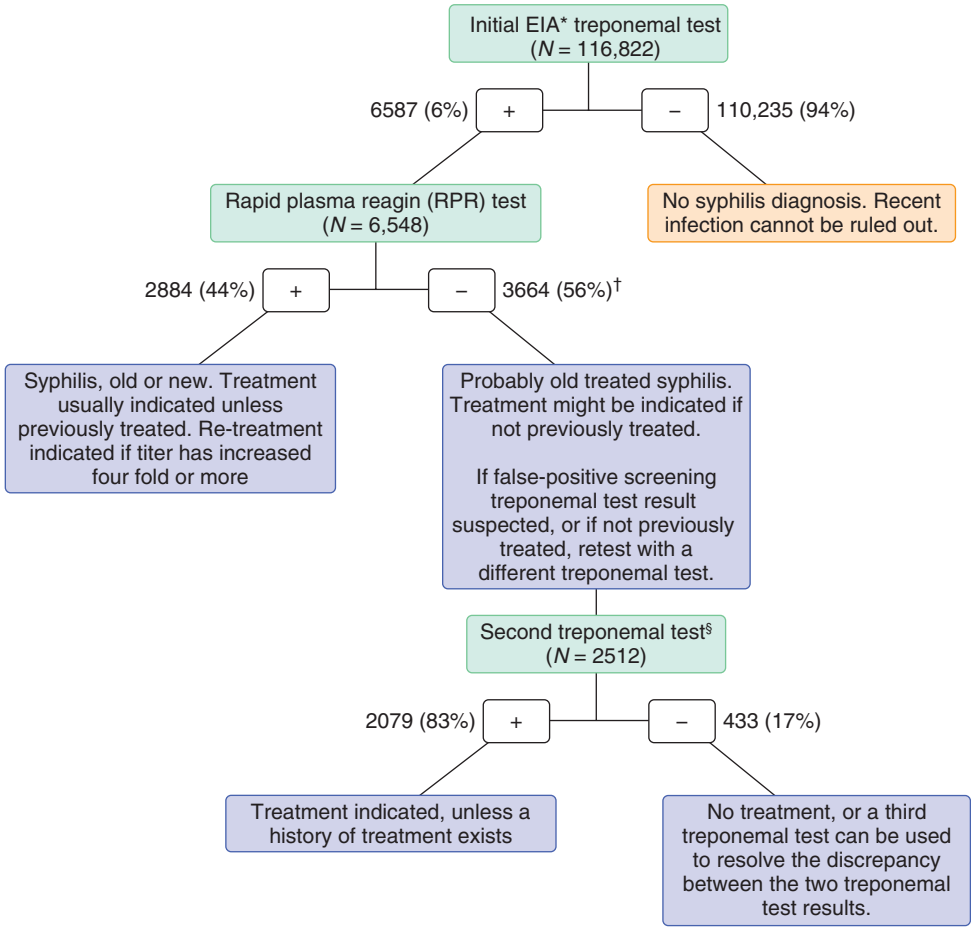

FIGURE 285-1. **Syphilis testing algorithms using treponemal tests for initial screening and recommendations from the Centers for Disease Control and Prevention, 2008. *Enzyme immunoassay. †Reactive with EIA treponemal test but nonreactive with RPR test. §Using *Treponema pallidum* particle agglutination or fluorescent treponemal antibody tests. (Redrawn from the Centers for Disease Control and Prevention. Syphilis testing algorithms using treponemal tests for initial screening—four laboratories, New York City, 2005-2006. *MMWR Morb Mortal Wkly Rep*. 2008;57:872-875.)**

Two HPV vaccines are currently approved by the Food and Drug Administration for use in males and females: a quadrivalent vaccine (protecting against HPV types 6, 11, 16, and 18), and a bivalent vaccine (protecting against HPV types 16 and 18). The quadrivalent vaccine protects against anal precancers and may influence future vaccine recommendations.

Female Genital Discharge

Infections of the female genitourinary tract produce several syndromes with overlapping symptoms (dysuria, vaginal discharge, vulvar irritation), the cause of which can usually be established with a careful history, examination, and laboratory tests. The initial approach depends on the primary anatomic site of infection—urinary tract, endocervix, or vagina. The columnar epithelium of the endocervix is susceptible to infection with *N. gonorrhoeae*, *C. trachomatis*, and *T. vaginalis*, and the vagina is susceptible to infection with *Candida albicans*, *T. vaginalis*, and the syndrome of bacterial vaginosis. The cervix may appear completely normal in women with cervical infection, but mucopurulence at the cervical os or mucosal friability suggests infection. Vaginitis is associated with a visible discharge, and the characteristics of the vaginal fluid offer diagnostic clues.

Female genital discharge is a condition in which syndromic management strategies generally lack sensitivity and specificity. In women with vaginal discharge, microscopic examination of a wet mount preparation may enhance the effectiveness of syndromic treatment, but interpretation of results is difficult and cannot exclude infection with several pathogens concurrently.

BACTERIAL VAGINOSIS

Bacterial vaginosis (BV) is the most common cause of vaginal discharge in the United States. Women with BV are often minimally symptomatic but may note mild vaginal discharge and vaginal odor (which is often increased after coitus). The normal vaginal flora contains hydrogen peroxide–producing lactobacilli such as *Lactobacillus crispatus* and *Lactobacillus jensenii*, which probably help "defend" the vagina against a number of pathogens (an example of innate immunity). *Lactobacillus acidophilus* is rarely found in the normal vagina, which explains the failure of yogurt to serve as a remedy or a preventive. BV begins with the unexplained disappearance of the normal vaginal flora and its replacement with *Gardnerella vaginalis* and many species of anaerobic bacteria. The precise mechanism causing this shift in vaginal flora is poorly understood. Most recently, previously unrecognized anaerobic bacterial species (BV-associated bacteria) have been described as potentially causative. In a study of 220 women with BV, the vaginal milieu demonstrated great species diversity. African American women without BV at the time of sampling had higher numbers of BV-associated bacteria, which could contribute to an increased risk of BV.[6]

The discharge of BV is homogeneous and may contain bubbles. Vaginal pH is elevated above the normal 4.0 to 4.5. Adding 10% potassium hydroxide to the vaginal discharge on the microscope slide or to the discharge present in the extracted speculum elicits an amine-like, fishy odor, yielding a positive "whiff" test result because of the elaboration of amines from the anaerobic flora. Examination of vaginal material as a wet mount reveals the absence of bacilli and their replacement with clumps of coccobacilli. Some vaginal epithelial cells are coated with coccobacilli, which may obscure their edges (clue cells) or the normally clear appearance of the cytoplasm. Relatively few polymorphonuclear leukocytes are observed; large numbers of leukocytes in the wet mount of a woman with BV suggest a coincident infection, possibly trichomoniasis or bacterial cervicitis.

BV is not necessarily a benign change in flora. It is associated with an increased rate of upper tract infection (endometritis, salpingitis) and, on occasion, with complications of pregnancy, including premature rupture of the membranes and preterm delivery. However, treatment of asymptomatic women with BV who are not at high risk for preterm delivery appears to confer no benefit. Women with BV may have increased risk for the acquisition of HIV. Treatment is generally directed against the anaerobic flora and

consists of metronidazole or clindamycin for 7 days. A single oral dose of metronidazole is not recommended for BV because of the high failure rate. The BV relapse rate is about 30%, and treatment of male sexual partners offers no benefit.

CANDIDIASIS

Vulvovaginal candidiasis (Chapter 338) is common and is seen most frequently in women taking antibiotics or using oral contraceptives when endogenous *Candida* species outgrow normal bacterial flora. Women usually complain of vulvar itching and discomfort and may or may not notice an accompanying discharge. The vagina generally maintains normal numbers of lactobacilli, so the vaginal pH is usually normal, which is helpful in discriminating between candidiasis and other vaginal infections. The labia and vaginal walls may be erythematous. Although classically described as "curdy," the discharge of candidiasis is frequently loose and is difficult to distinguish from other discharges. Vaginal material may be treated with 10% potassium hydroxide to destroy other cellular elements and to make the fungi easier to observe. Wet mount, however, has a sensitivity of only about 50%, and a woman with a classic clinical presentation should be treated even if fungal elements are not observed.

A wide range of topical antifungal medications are available (many without a prescription), and all these drugs are approximately equally effective, although the cure rate with some single-dose topical treatments appears to be lower than that with longer regimens. Fluconazole administered as a single oral dose of 150 mg is highly effective. Infection with yeasts other than *C. albicans* may require longer therapy. Recurrent vulvovaginal candidiasis is a problem for many women, and optimal management has not been defined. Recurrent infection should lead the clinician to consider underlying diabetes mellitus or HIV infection. Treatment of sexual partners of women with candidiasis confers no benefit.

TRICHOMONIASIS

Women with *T. vaginalis* infection complain of a purulent discharge and vulvar irritation. The vaginal walls are red, and the vagina may contain excessive yellow or green discharge displaying large bubbles. The ectocervix may also be inflamed or have punctate microhemorrhages, causing the pathognomonic "strawberry cervix" (colpitis macularis). Vaginal pH is elevated, but the whiff test result is generally negative. Wet mount reveals large numbers of polymorphonuclear leukocytes as well as motile protozoa about the same size as the leukocytes, with visible flagella; motile organisms may be recognized in about two thirds of cases. Therapy for trichomoniasis requires metronidazole or tinidazole, but resistant organisms are encountered with increasing frequency.

CERVICITIS

The diagnosis of cervicitis is suggested by tenderness on bimanual examination, visible inspection revealing inflammation, or discharge. The specific diagnosis can be made only by detecting microorganisms from the cervix. *N. gonorrhoeae* and *C. trachomatis* have tropism for cervical tissue, whereas other pathogens (including HIV) can apparently infect the vaginal tissues as well.

PELVIC INFLAMMATORY DISEASE

Each year, more than 800,000 women develop pelvic inflammatory disease (PID) in the United States. The majority of PID cases involve *N. gonorrhoeae*, *C. trachomatis*, and *Mycoplasma* (especially *M. genitalium*); however, one third of PID cases are due to other anaerobic and aerobic bacteria that ascend from the cervix into the uterine cavity, producing endometritis, and then extend to the fallopian tubes, causing salpingitis. Treatment should include therapy directed at anaerobes. Chlamydial salpingitis may be mild, and patients may not seek medical attention. Some intrauterine devices have been associated with an increased risk of salpingitis, and some data suggest that vaginal douching is a predisposing factor.

Adnexal tenderness on bimanual examination leads to the clinical diagnosis of salpingitis. Cervical tenderness, fever, leukocytosis, and an elevated sedimentation rate are sometimes observed. The clinical diagnosis is confirmed laparoscopically in only about 70% of cases, suggesting considerable error in diagnosis. Vaginal ultrasonography or computed tomography is often helpful in defining the cause of pelvic pain syndromes. Pregnant women with evidence of salpingitis should be hospitalized. Other indications for hospitalization include nonresponse to or intolerance of an oral regimen, presence of a tubo-ovarian abscess, and inability to rule out a surgical emergency, such as appendicitis.[7] Infertility complicates approximately 15% of initial attacks of salpingitis and about 75% of women who suffer three or more attacks. Ectopic pregnancy, infertility, and tubo-ovarian abscess are complications of salpingitis.

PREVENTION

STIs are preventable. The Centers for Disease Control and Prevention recommends five strategies as the foundation for an effective prevention program: (1) education and counseling of persons at risk to motivate the adoption of safer sexual behavior; (2) identification of asymptomatic infected persons and symptomatic persons unlikely to seek diagnostic and treatment services; (3) rapid and effective diagnosis and treatment of infected persons; (4) evaluation, treatment, and counseling of exposed sexual partners; and (5) preexposure vaccination of persons at risk for vaccine-preventable STIs.

Behavioral Interventions

Abstaining from sexual intercourse or being in a long-term, mutually monogamous relationship with an uninfected partner is the most reliable way to prevent STIs. Abstinence should be recommended during treatment for an STI and for anyone who wants to avoid STIs and unintended pregnancy. Both partners should be tested for STIs, including HIV infection, before initiating sexual intercourse.

Counseling is essential for people with STIs. Interactive counseling, video presentations, peer groups, and other formats that emphasize correct condom use have reduced the incidence of subsequent infections among STI clinic patients and adolescents. Randomized controlled trials demonstrate that structured risk reduction counseling can reduce the incidence of infections by 25 to 40% among some STI clinic populations. HIV testing is preceded by a counseling session, but there is little evidence of a preventive benefit from this communication.

Barrier Methods

When used consistently and correctly, male latex condoms are effective in preventing the sexual transmission of HIV infection and can reduce the risk of other STIs (gonorrhea, chlamydia, and trichomoniasis). However, because condoms do not cover all exposed areas, they are likely to be more effective in preventing infections transmitted by fluids from mucosal surfaces (e.g., gonorrhea, chlamydia, trichomoniasis, HIV infection) than in preventing those transmitted by skin-to-skin contact (e.g., HSV, HPV, syphilis, chancroid). Male condom failure usually results from inconsistent or incorrect use rather than from condom breakage. Non-latex condoms (those made of polyurethane or other synthetic material) can be used by persons with latex allergy. There is less information available on the effect of female condoms on the incidence of STIs. Although cervical caps and diaphragms cover the cervix, there is little evidence that they can prevent STIs or HIV infection.

Male Circumcision

Male circumcision reduces mucosal tissue susceptible to HIV and STI pathogens. Circumcision of adult men reduces the acquisition of HIV by more than 70% for up to 5 years after circumcision. Circumcision also appears to reduce the acquisition of other viral STI pathogens, including HSV-2 and HPV.

Partner Services

The detection of an STI demands consideration of the infected person's sexual partners, who may have undetected and serious disease. In addition, in the absence of partner treatment, reinfection of the index case can be expected. The probability that a sexual partner is also infected reflects the efficiency of transmission of the STI pathogen, as described earlier. For example, most men with gonococcal urethritis infect their partners, whereas only about half of patients with HIV infection have infected their partners at the time of outreach; partners who differ in terms of STI or HIV infection status are referred to as discordant.

Sexual partners can be notified directly by the infected person or by health care workers, sometimes through proactive contact tracing. In general, dependence on the infected person is a less reliable way to get partners treated. However, in many states it is legal to pursue expedited care by providing the infected patient with the appropriate treatment for his or her partners. Expedited partner care appears to work well for the treatment of gonorrhea and chlamydia infections.

Preexposure Interventions

Preexposure vaccination is the most effective method for preventing the transmission of certain STIs. For example, because hepatitis B virus is

frequently transmitted sexually, hepatitis B vaccination is recommended for all unvaccinated persons being evaluated for an STI. In addition, hepatitis A vaccine is recommended for men who have sex with men and for drug users (both injection and noninjection). Vaccines for HPV are now available for both females and males (Chapter 373). The HPV vaccines now have more than 5 years of follow-up and demonstrate continued high rates of protection in women. HPV vaccination for males may reduce acquisition of genital warts and is protective against anal precancerous lesions, leading to a "permissive" vaccine recommendation for boys aged 9 to 26 years from the Advisory Committee on Immunization Practices.

Several studies have evaluated the effectiveness of antiretroviral agents for HIV preexposure prophylaxis. Results have indicated HIV prevention but only in the setting of high-level (>85%) medication adherence. Furthermore, because the only currently Food and Drug Administration–approved antiretroviral for HIV prevention is also a drug combination used for therapy, regular HIV screening is necessary to prevent the development of resistance should the patient become infected.[8] Topical HIV prevention by vaginal microbicides has been only partially successful, in large part because of poor adherence. Long-acting injectable drugs, vaginal rings, and coformulated options (i.e., contraceptive and antiretroviral agents) are under evaluation.

Postexposure Prophylaxis

After consensual or nonconsensual sexual exposure, a variety of STIs can be prevented with empirical antibiotics. Prevention of HIV infection appears to require several antiretroviral agents that must be used in combination for 28 days.

Contraception

All methods of birth control can influence the acquisition and outcome of an STI. In addition, pregnancy itself (in the absence of effective birth control) affects STI acquisition and the health of the pregnancy and the neonate. Accordingly, STI management demands consideration of the reproductive health of both partners as well as family planning issues. A systematic review suggested that there was no increased risk of HIV infection in oral contraceptive users; however, the data for injectable hormonal contraceptives were more difficult to interpret.[9] Women using injectable hormonal birth control should be cautioned about consistent condom use for STI prevention until these relationships are better clarified.

TOWARD A COMPREHENSIVE MANAGEMENT STRATEGY

Although most STIs are self-limited and readily treated, the comprehensive and proper management of the patient with an STI requires considerable skill.[10] First, the correct syndrome must be recognized, and a decision about specific diagnostic tests must be made. Second, empirical therapy must be provided and must be sufficiently broad to promise cure or reduced duration of illness. Third, the clinician is obligated to search for other STIs of public health or personal significance. Fourth, the clinician must deal with the patient's sexual partners, through either referral or expedited partner therapy. Fifth, the patient needs counseling and adjunctive preventive measures, where appropriate. Such measures might include vaccination for hepatitis B or HPV or antibiotics to prevent another STI, such as incubating syphilis or HIV infection.

Grade A References

A1. Centers for Disease Control and Prevention (CDC). Update to CDC's sexually transmitted diseases treatment guidelines, 2010: oral cephalosporins no longer a recommended treatment for gonococcal infections. *MMWR Morb Mortal Wkly Rep*. 2012;61:590-594.

A2. Creighton S. Gonorrhoea. *Clin Evid (Online)*. 2014;2014.

A3. Ross JD. Pelvic inflammatory disease. *Clin Evid (Online)*. 2013;2013.

A4. Westhoff CL, Jones HE. Guiahi M. Do new guidelines and technology make the routine pelvic examination obsolete? *J Womens Health (Larchmt)*. 2011;20:5-10.

A5. Gray R, Kigozi G, Kong X, et al. The effectiveness of male circumcision for HIV prevention and effects on risk behaviors in a posttrial follow-up study. *AIDS*. 2012;26:609-615.

A6. Ferreira A, Young T, Mathews C, et al. Strategies for partner notification for sexually transmitted infections, including HIV. *Cochrane Database Syst Rev*. 2013;10:CD002843.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Fahami R. Abnormal vaginal discharge. *BMJ*. 2013;347:f4975.
2. Mlisana K, Naicker N, Werner L, et al. Symptomatic vaginal discharge is a poor predictor of sexually transmitted infections and genital tract inflammation in high-risk women in South Africa. *J Infect Dis*. 2012;206:6-14.
3. Manhart LE, Broad JM, Golden MR. *Mycoplasma genitalium*: should we treat and how? *Clin Infect Dis*. 2011;53(suppl 3):S129-S142.
4. Binnicker MJ, Jespersen DJ, Rollins LO. Treponema-specific tests for serodiagnosis of syphilis: comparative evaluation of seven assays. *J Clin Microbiol*. 2011;49:1313-1317.
5. D'Souza G, Dempsey A. The role of HPV in head and neck cancer and review of the HPV vaccine. *Prev Med*. 2011;53(suppl 1):S5-S11.
6. Srinivasan S, Hoffman NG, Morgan MT, et al. Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLoS ONE*. 2012;7:e37818.
7. Ross J, Judlin P, Jensen J. 2012 European guideline for the management of pelvic inflammatory disease. *Int J STD AIDS*. 2014;25:1-7.
8. Smith DK, Thigpen MC, Nesheim SR, et al. Interim guidance for clinicians considering the use of pre-exposure prophylaxis for the prevention of HIV infection in heterosexually active adults. *MMWR Morb Mortal Wkly Rep*. 2012;61:586-588.
9. Polis CB, Curtis KM. Use of hormonal contraceptives and HIV acquisition in women: a systematic review of the epidemiological evidence. *Lancet*. 2013;13:797-808.
10. Ahmed-Jushuf I, Lowbury R. Standards for the management of sexually transmitted infections—2014. *Sex Transm Infect*. 2014;90:444.

REVIEW QUESTIONS

1. The Centers for Disease Control and Prevention has recommended gonorrhea treatment with an injectable third-generation cephalosporin and azithromycin. Which of the following answers is the correct rationale for this recommendation?

A. Azithromycin should be used in conjunction with an injectable third-generation cephalosporin to provide therapy with two different mechanisms of action, thereby delaying emergence of resistance.
B. There is a high rate of concomitant chlamydial infection as well as an increasing recognition of *Mycoplasma genitalium* coinfection that should be treated.
C. Cefixime and azithromycin may be used in some settings without further follow-up.
D. Moxifloxacin may be used to treat *M. genitalium* as well as *Neisseria gonorrhoeae.*

Answer: A A theoretical basis does exist for using two antimicrobials with different mechanisms of action to treat gonorrhea and to prevent emergence of resistance. Although there is a high rate of concomitant chlamydia and mycoplasma coinfection with gonorrhea, it does *not* reflect the rationale for using azithromycin and an injectable third-generation cephalosporin as first-line therapy against gonorrhea. Oral cefixime therapy is a second-line therapy and should be followed with a test of cure. Fluoroquinolones are to be avoided for gonorrhea.

2. Male circumcision is currently recommended by WHO/UNAIDS. Which of the following statements is *incorrect* regarding male circumcision?

A. The long-term effectiveness of male circumcision is well described and has been demonstrated in several studies.
B. Sexually transmitted infection (STI) prevention trials that incorporate interventions like circumcision may lead to a bias/blunting of effect compared with community standards of care.
C. Male circumcision demonstrated greater than 75% reduction in human immunodeficiency virus (HIV) acquisition during a 5-year period in at least two African trials.
D. Circumcision was not accepted by men after completion of the trial.
E. Circumcised men were more likely to engage in risky sexual behaviors compared with uncircumcised men.

Answer: C The uptake of circumcision after the trial approached 80% in the Rakai study. Long-term effectiveness of circumcision has not been *proved* but may be inferred from several studies. In larger randomized controlled trials, circumcision compared with community standard of care did not appear to have a biased effect. Further, in the post-trial surveillance for the Rakai study, there were no statistically significant differences in sexual risk-taking behaviors in men subsequently choosing circumcision compared with those who did not choose circumcision, and there were no statistically significant differences in sexual risk-taking behavior in uncircumcised versus circumcised men.

3. Human papillomavirus (HPV) is a vaccine-preventable STI. Which of the following is true about HPV vaccination?

A. Oral HPV-associated head and neck cancers (associated with smoking, male gender, and HIV infection) can be prevented with vaccination.
B. Two vaccines are currently available. Both protect against HPV types 16 and 18, which are responsible for at least 85% of all cervical neoplasia; one of the vaccines includes types 6 and 11 (responsible for most cases of genital warts and precancerous anal lesions).
C. Papanicolaou smears are recommended for all sexually active women beginning at coitarche, and HPV testing is recommended for women younger than 30 years.
D. The quadrivalent HPV vaccine protects against only genital warts in males.

Answer: B Two vaccines are currently available. Both protect against HPV types 16 and 18, which are responsible for at least 85% of all cervical neoplasia; one of the vaccines includes types 6 and 11 (responsible for most cases of genital warts and precancerous anal lesions). Current HPV vaccines have not been shown to have any effect on head and neck cancers. The quadrivalent vaccine does protect against anal precancers and may influence future vaccine recommendations.

4. Which of the following are true about vaginal discharge and syndromic STI treatment?

A. Vaginal-associated and cervical-associated causes of vaginal discharge are difficult to discriminate on the basis of symptoms alone.
B. Many STIs in females are asymptomatic, leading to underidentification and undertreatment.
C. Subclinical STI may still contribute significantly to genital inflammation and risk of HIV transmission and acquisition.
D. All women possess a similar vaginal milieu that contributes significantly to bacterial vaginosis, a common noninfectious cause of vaginitis.
E. A, B, and C are true.

Answer: E All of the responses are true *except* D. Women possess a heterogeneous vaginal microbiome, some species of which are associated with an increased risk of bacterial vaginosis. Mlisana and coworkers demonstrated elevated cervicovaginal cytokine levels for women with STI, regardless of self-report of vaginal discharge. Vaginal discharge has poor sensitivity, specificity, and positive predictive value as a marker for STIs.

5. Many commercial laboratories have switched to use of treponemal-specific tests for syphilis diagnosis. Which of the following statements is true regarding the "reversed" testing algorithm?

A. Treponemal-specific testing may be used to follow treatment.
B. Some larger commercial laboratories have reversed the order of testing, using an antitreponemal test followed by an anticardiolipin test, which allows automation and may be cost-effective in areas of low endemicity.
C. If a patient has a positive treponemal antibody test result and a negative serologic test result, that patient has a false-positive syphilis test result and does not require treatment.
D. A patient presents with a positive treponemal test result and elevated serologic (1 : 64) titer 9 months ago. He was treated appropriately and lost to follow-up. The patient now presents with a positive treponemal test result and a nonreactive serologic test result. This patient should be retreated.

Answer: B Treponemal-specific tests are not used to follow treatment. Non–treponemal-specific antibody tests (e.g., VDRL, RPR) should still be used to monitor therapy; specifically, a four-fold decrease in titer indicates adequate treatment. A patient with a positive treponemal antibody test result and a negative serologic test result may have old or untreated syphilis. If a false positive is suspected, a second treponemal antibody test should be performed. This patient has evidence of appropriately treated syphilis (as evidenced by a greater than four-fold decrease in titer to a nonreactive serologic test) and does not require further therapy.

286

APPROACH TO THE PATIENT BEFORE AND AFTER TRAVEL

DAVID O. FREEDMAN

Prevention strategies and medical interventions for the traveler need to be individualized according to a risk assessment that considers both the itinerary and factors that are dependent on the prospective traveler. A structured approach to patient interaction (Table 286-1) is the most efficient way to cover the necessary educational and preventive interventions. As many of these measures will be initiated only much later at the traveler's destination, clearly printed instructions in lay language are advisable. The worldwide epidemiology of travel-related diseases is constantly changing. Special needs travelers, such as those who are immunocompromised, are pregnant, or have significant underlying disease, should be referred to a specialized travel medicine clinic.

Globally, approximately 100 million people travel from industrialized to developing countries each year. Several recent analyses have provided much needed new data on the profiles of travel-related illness determined by destination of travel.[1] Depending on destination, 22 to 64% of travelers report some illness; most of these problems are mild, self-limited illnesses, such as diarrhea, respiratory infections, and skin disorders. Infectious diseases account for up to 10% of the morbidity during travel but only 1% of the deaths, with malaria being the most common disease.

IMMUNIZATION

The choice of vaccines for an individual traveler is based on risk of exposure to vaccine-preventable diseases on the chosen itinerary, the severity of disease if it is acquired, and any risks presented by the vaccine itself. Travelers differ in their tolerance of risk. For the vaccine-preventable diseases, the monthly incidence for nonimmune travelers to developing countries is most significant for symptomatic hepatitis A at 0.03% per month overall; the risk of symptomatic hepatitis B is most significant for long-stay travelers at 0.25% per month. Enteric fever (typhoid and paratyphoid) has a risk of 0.03% per month on the Indian subcontinent and is 10 times lower in Africa and parts of Latin America. Risk of yellow fever may be as high as 0.1% per month of travel to an area with current epidemic transmission, but the risk varies greatly between destinations encompassed by the endemic area map. The risk of meningococcal meningitis, rabies, cholera, polio, measles, varicella, and Japanese encephalitis in travelers is not known but is thought to be small (<0.0001%).

Table 286-2 provides data on dosing, administration, need for boosters, and possible accelerated regimens for vaccines administered in the travel medicine setting. Details on vaccine composition, mechanism of action, use for routine adult and childhood primary vaccination, and adverse reactions can be found in Chapter 18. The following discussion focuses on indications for vaccines in the context of travel.

Verification and Update of Routine Immunizations

Because of the increased prevalence of many infections in the developing world, routine adult immunizations need to be current.[2] If no adult doses of tetanus/diphtheria/acellular pertussis (Tdap) have ever been given, a dose of Tdap should be given regardless of the time elapsed since the last tetanus/diphtheria vaccination. Persons born in the United States before 1957 or born anytime in the developing world are considered immune to measles. Other adult travelers should have received at least two doses of live measles–containing vaccine during their life unless a history of measles infection can be documented. Unvaccinated persons who have the accepted routine indications for influenza or pneumococcal vaccines (Chapter 18) should receive these during the pretravel consultation. Two doses of varicella vaccine spaced by at least 4 weeks should be considered for adult travelers without evidence of varicella immunity. Adults born before 1980 in the United States are considered immune.

Vaccines to Consider for All Developing World Travelers

A number of additional vaccines should be administered depending on the travel destinations.[3]

Hepatitis A

Hepatitis A vaccine is indicated for every nonimmune traveler to countries or areas with moderate to high risk of infection, which includes essentially

TABLE 286-1 THE PRETRAVEL CONSULTATION WITH A TRAVELER TO THE DEVELOPING WORLD—A STRUCTURED APPROACH

PERFORM RISK ASSESSMENT
The following must always be ascertained initially to determine appropriate preventive medical recommendations. Preprinted medical record forms may be used to record these.
Exact itinerary, including regions within each country to be visited
Dates of travel to assess risk of seasonal diseases
Age
Past vaccination history
Underlying illnesses
Current medications
Pregnancy status
Allergies
Purpose of trip
Risk exposures—blood, body fluids, adventure or extensive outdoor exposures
Urban versus rural travel
Type of accommodations
Level of aversion to risk
Financial limitations that may necessitate prioritization of interventions
ADMINISTER IMMUNIZATIONS
Administer routine vaccinations that are not up to date.
Administer indicated travel vaccines.
Provide to patient legally mandated Vaccine Information Statements from the Centers for Disease Control and Prevention (http://www.cdc.gov/vaccines/pubs/vis/).
Provide printed checklist to patient listing vaccines administered.
Record in the clinic record vaccines administered, lot number, and date.
Document vaccines offered to but declined by patient as well as nonrecommended vaccines administered at the patient's request.
PROVIDE MALARIA PREVENTION (IF INDICATED)
Determine whether malaria risk exists for the destination country. If yes:
Does the patient's itinerary within that country put him or her at risk? If yes:
Recommend malaria chemoprophylaxis. Several equally effective drugs of choice may be indicated. Ascertain which is best suited to the individual patient and itinerary.
Educate on personal protection against arthropods.
EDUCATE ON TRAVELER'S DIARRHEA
Recommend food and water precautions.
Prescribe and educate on standby therapy with a quinolone antibiotic or azithromycin and advise on use of loperamide and oral hydration if needed.
TEACH ESSENTIAL PREVENTIVE BEHAVIORS
Most travel-related health problems, including vaccine-preventable diseases, can be avoided through simple behaviors initiated by the traveler.
Educate on appropriate strategies in the following categories (some topics are not applicable to all destinations): blood-borne and sexually transmitted diseases, safety and crime avoidance, injury prevention, swimming safety, rabies, skin/wound care, tuberculosis, packing for healthy travel, obtaining health care abroad.
DISCUSS OTHER APPLICABLE HEATH ISSUES
Advise and prescribe for altitude illness, motion sickness, or jet leg.
Discuss prevention of specific travel-related infections that are of some risk to the traveler and have a possible preventive strategy not included in strategies above.
Discuss any minimal-risk conditions (e.g., hemorrhagic fevers) that are a frequent cause of patient anxiety.

everyone traveling outside the United States, Canada, Japan, Australia, New Zealand, Scandinavian countries, and developed countries in Europe. A single dose of hepatitis A vaccine given any time before travel provides adequate protection. Persons with a history of hepatitis or who previously lived in an endemic country for a prolonged period may benefit from prevaccination serum antibody testing.

Hepatitis B

Pretravel hepatitis B vaccination is indicated for all nonvaccinated travelers with standard indications, such as health care workers, and all longer-stay travelers who will be visiting or residing in high- or moderate-risk areas. Transmission by routes such as sexual contact, blood transfusions, contaminated medical equipment, body piercing, tattooing, acupuncture, and sharing of cooking and bathroom facilities is difficult to control or to predict in the context of travel. Vaccination is usually advocated for short-term travelers, especially younger travelers and those anticipating close contact with local populations, even if they have no specific risk factors. Adventure travelers (accident prone), backpackers, and those with underlying medical conditions are more likely to require contact with the medical system. Accelerated and hyperaccelerated schedules (see Table 286-2) are used widely in practice and are approved in many countries. These are helpful in administering all three primary doses necessary for high assurance of protection in the frequent circumstance in which the traveler is leaving in a very short time and is at risk of hepatitis B exposure.

Combination Hepatitis A and Hepatitis B Vaccine

The combined hepatitis A and hepatitis B vaccine provides convenience for travelers with an overlap of indications for use of the individual vaccines. A less well known accelerated 3-week schedule (see Table 286-2) is approved by the Food and Drug Administration.

Typhoid

Typhoid vaccine is indicated for all travelers to the Indian subcontinent and considered for those traveling to other endemic areas under all but the most deluxe and protected of conditions. Risk increases with trip duration, lodging and eating with local residents, and extent of travel off the usual tourist itineraries. Current typhoid vaccines do not protect against *Salmonella paratyphi*, which is emerging in many areas.[4] Adherence to the oral vaccine regimen may be as low as 70%.

Influenza

Influenza is transmitted year-round in the tropics. Increasing data show that influenza may be the most common vaccine-preventable illness in travelers. An increased risk of influenza has been reported among cruise ship passengers. All travelers to destinations with current influenza virus circulation, not just those with the usual risk factors, should strongly consider influenza vaccination.[5]

Vaccines for Certain Destinations

Yellow Fever

The primary indication for yellow fever vaccination is to prevent infection in individuals at risk. A map of risk areas can be found at www.cdc.gov/travel. However, yellow fever is currently the only vaccine that falls under the International Health Regulations that may necessitate vaccination purely for regulatory reasons. A number of African countries and one in South America (French Guiana) require proof of yellow fever vaccination from all arriving travelers. Other countries, both within and outside the risk zone, have submitted more complex requirements to the World Health Organization. Current country-by-country yellow fever entry requirements are at www.who.int/ith/chapters/en/index.html. A Centers for Disease Control and Prevention–designated yellow fever vaccination center should be consulted for detailed requirements. Neither yellow fever vaccine nor any other vaccine is currently required for readmission to the United States. In general, all healthy adult travelers to areas with a risk of yellow fever transmission should be vaccinated. The true duration of immunity from yellow fever vaccination is probably much longer than the stated 10 years and may exceed 30 years.[6]

Meningococcus

Meningococcal vaccine is recommended for travelers to Africa's sub-Saharan "meningitis belt" during the dry season from December through June, especially if prolonged contact with the local populace is likely. Out-of-season epidemics have occurred in Ethiopia, Somalia, and Tanzania, indicating possible changes in epidemiologic trends perhaps due to climate changes. Muslims undertaking Hajj and Umrah pilgrimages in Saudi Arabia are at a higher risk of meningococcal disease, and proof of vaccination with quadrivalent vaccine within the past 3 years is required to obtain pilgrimage visas.

Rabies

A preexposure rabies series is indicated for long-stay travel to endemic areas of Latin America, Asia, or Africa where the rabies threat is constant and where access to adequate postexposure rabies immune globulin and vaccine is likely to be limited. For short-term travel, risk groups for whom immunization should be considered include adventure travelers, bikers, hikers, cave explorers, and business travelers who travel for short but frequent trips and plan to go running outdoors on these trips.

Japanese Encephalitis

Japanese encephalitis is endemic to many rural farming areas of Southeast Asia and the Indian subcontinent. Sporadic cases with severe sequelae

TABLE 286-2 TRAVEL-RELATED VACCINES OF ADULTS

DISEASE	VACCINE	PRIMARY COURSE	ROUTE	FURTHER BOOSTERS
VACCINES TO CONSIDER FOR ALL DEVELOPING WORLD TRAVELERS				
Hepatitis A	Killed virus	0, 6-18 months	IM	None
Hepatitis B	Recombinant viral antigen	0, 1, 6 months A: 0, 1, 2 weeks and 12 months A: 0, 1, 3 weeks and 12 months*	IM IM IM	None None None
Hepatitis A/B	Combination of monovalent preparations	0, 1, 6 months A: 0, 1, 3 weeks and 12 months	IM IM	None None
Typhoid	Capsular Vi polysaccharide Live attenuated Ty21a bacteria	Single dose 0, 2, 4, 6 days	IM Oral	2-3 years 5 years
Influenza	Inactivated viral Live attenuated virus	Single dose Single dose (<50 years of age only)	IM Nasal	Annual Annual
Varicella	Live attenuated virus	0, 4-8 weeks	SC	None
VACCINES FOR CERTAIN DESTINATIONS				
Yellow fever	Live attenuated 17D virus	Single dose	SC	10 years
Meningococcus	Quadrivalent conjugated polysaccharide (A, C, Y, W135)	Single dose	IM	5 years
Rabies	Inactivated viral cell culture	0, 7, 21-28 days	IM†	None routinely but two doses after each exposure
Japanese encephalitis (Vero cell)	Inactivated viral	0, 28 days	IM	1 year if at continued risk; no data on subsequent doses
Polio‡	Inactivated viral	Single dose if adequate childhood series	SC; IM acceptable	None
Cholera§	Killed bacteria + recombinant B toxin subunit‖	0, 1 week	Oral	2 years for cholera; 3 months for ETEC

*Regimen not approved by the U.S. Food and Drug Administration for monovalent hepatitis B vaccine but approved for combination hepatitis A/B vaccine containing the same quantity of hepatitis B antigen.
†Intradermal rabies preexposure vaccine is no longer produced, and the intramuscular 1.0-mL vials are not licensed for intradermal use in a 0.1-mL dose.
‡Oral polio vaccine is no longer produced in the United States.
§Not available in the United States but available in Canada and most European countries. No cholera vaccine of any kind is currently available in the United States.
‖Also licensed in some countries for traveler's diarrhea due to enterotoxigenic *Escherichia coli*.
A = accelerated regimen to be used for imminent departures; ETEC = enterotoxigenic *E. coli*; IM = intramuscular; SC = subcutaneous.

continue to occur in travelers.[7] In temperate regions, the transmission season is from April through November. In tropical or subtropical regions of Oceania and Southeast Asia, transmission may occur year-round. Vaccination is recommended for (1) long-stay travel to an endemic rural area; (2) expatriation to anywhere in an endemic country; (3) short-term travel to endemic rural areas with extensive unprotected outdoor exposure, such as with adventure travel; and (4) short-term travel in the face of a current local epidemic.

Polio

Because of eradication efforts, poliomyelitis remains in only a few countries, but complete control remains elusive. Adults traveling to countries that are currently polio endemic (updated information at www.polioeradication.org) and who have previously completed a primary vaccine series should receive a one-time single dose of inactivated polio vaccine as a booster if the last dose or booster dose was administered at least 10 years previously.

Cholera

Cholera vaccination is no longer required by any country, and the risk to typical travelers is insignificant.[8] However, medical and aid workers staying for short periods in disaster areas or refugee camps may consider cholera vaccine. A highly effective oral killed whole cell–B subunit vaccine is available widely outside the United States.

Sequence of Travel-Related Vaccines

All currently indicated immunizations can and should be given at the same time and in any combination (Chapter 18). If two live viral antigens are not administered on the same day, they must be spaced by a month. However, yellow fever vaccine can be given at any interval with respect to single-antigen measles vaccine. Minimum intervals between vaccine doses must be respected, although 4 days or fewer before the next interval are acceptable. Regimens that involve 1-week intervals (rabies, Japanese encephalitis, accelerated hepatitis) are exceptions. There is not a maximum interval between doses of a primary vaccine series; interrupted series (except oral typhoid and rabies) need not be restarted but can be resumed beginning with the dose that is overdue.

MALARIA CHEMOPROPHYLAXIS

An average of 1500 imported cases of malaria are reported annually in the United States. Estimates of risk in travelers not taking chemoprophylaxis vary widely by destination but range from 3.4% per month in West Africa to one tenth that on the Indian subcontinent and a further 10-fold reduction in South America. The majority of cases of imported malaria in the United States and Europe occur in noncitizen immigrants visiting friends and relatives abroad.

Resources describing current country-specific malaria microepidemiology should be accessible immediately to those prescribing malaria prophylaxis. Dosing and pharmaceutical properties of antimalarial drugs are described in Chapter 345. In the limited number of countries where it is still effective, chloroquine, 500 mg salt (300 mg base) per week beginning the week before the first exposure to malaria and continuing for 4 weeks after the last exposure, is still the drug of choice. However, atovaquone/proguanil may still be used by short-stay travelers who prefer the shorter duration of that regimen.

For all other areas of the world, three drugs are equally effective, and the choice depends on both traveler and itinerary factors. Atovaquone/proguanil (250/100 mg) is a well-tolerated, once-a-day drug that should be started 1 day before arrival in the malarious area (may not coincide with first overseas destination) and continued for 7 days after the last exposure. The short period of postexposure use makes it convenient for the many travelers on typical 1- to 3-week itineraries. High cost and daily dosing make it difficult to use for extended periods. Weekly mefloquine (250 mg) is given 2 and preferably 3 weeks before the first exposure to malaria and continued for 4 weeks thereafter. Weekly dosing and a long track record of efficacy make this drug the most effective for long-stay travelers. If contraindications to mefloquine exist for long-stay travelers, daily doxycycline (100 mg) beginning 1 day before exposure can be used; unlike atovaquone/proguanil, it must be continued for 4 weeks after exposure. Approximately 5% of individuals who take either mefloquine or doxycycline discontinue therapy because of side effects. Chemoprophylaxis may be started well before departure (3 to 4 weeks for mefloquine) in those concerned about possible intolerance to any drug.

Travelers should be reminded in writing to continue antimalarial drugs for the appropriate period after the last possible exposure, that malaria can still

occur despite chemoprophylaxis, and that a malaria smear or malaria rapid diagnostic test is mandatory for any febrile illness occurring within 3 months after travel. Prevention of malaria in travelers residing in malarious areas for 6 months or more presents complex problems that have been reviewed elsewhere.

DENGUE

An estimated 100 million cases of dengue fever and 250,000 cases of dengue hemorrhagic fever occur annually (Chapter 381). The past 20 years have seen a dramatic geographic expansion of epidemic dengue fever and dengue hemorrhagic fever. Dengue accounts for up to 2% of all illness in returned travelers, and dengue is the most common systemic febrile illness in returned travelers from every region except sub-Saharan Africa, where malaria still predominates. Several dengue vaccine candidates are in advanced clinical trials. Dengue fever is transmitted by day-biting *Aedes* mosquitoes, reinforcing the need to instruct travelers to the tropics in the need for both day and night use of repellents.

TRAVELER'S DIARRHEA

The most frequent cause of traveler's diarrhea is enterotoxigenic *Escherichia coli* and in some locations enteroaggregative *E. coli*. *Salmonella*, *Shigella*, and *Campylobacter* each account for about 5 to 15%, and in Asia noncholera vibrios are significant.[9] Protozoa account for less than 5%. In adults, norovirus and rotavirus are increasingly detected. The mean duration of traveler's diarrhea, even if it is untreated, is 4 days.

All travelers to the developing world should be thoroughly educated in self-therapy for diarrheal disease and carry the appropriate agents while traveling.[10] Eighty percent of patients respond to a regimen of loperamide and an antibiotic within 24 hours. A single dose of a self-administered quinolone is usually sufficient, but patients should be instructed to complete 3 days of therapy with 500 mg of levofloxacin daily or 500 mg of ciprofloxacin twice daily should the traveler's diarrhea not resolve within 24 hours. Because of a significant increase in quinolone-resistant *Campylobacter* in Southeast Asia, India, and Nepal, travelers to those destinations should self-treat with azithromycin, 500 mg/day for 3 days or a single dose of 1000 mg.

Most guidelines do not recommend prophylaxis for the typical traveler because of potential adverse drug effects while away from medical care and because effective rapid-onset therapy is available for diarrhea should it occur. Exceptions include travelers with advanced human immunodeficiency virus (HIV) infection, those who have an underlying chronic medical problem that makes them more prone to adverse consequences from diarrhea, and travelers on a vital mission for a short period (less than 1 week) who cannot tolerate even a day of disability. Antibiotic prophylaxis should be carried out with a quinolone once per day or with rifaximin twice per day; prophylaxis should be used only for trips of 2 weeks or less.

PREVENTIVE BEHAVIORS

Most travel-related health problems, including many infectious diseases, can be significantly reduced through appropriate behavior by the traveler.

Mosquito Protection

Antimalarial chemoprophylactic drugs are less than 100% effective. Protection against arthropods will help prevent dengue, leishmaniasis, filariasis, and a number of important arboviral diseases. Travelers should be instructed to clothe themselves to reduce as much exposed skin as practicable and to apply a repellent containing DEET (concentration of 30% to 35%) to all exposed, nonsensitive areas of the body every 4 to 6 hours.[11] More frequent application is required for agents containing lower concentrations of DEET. Travelers should sleep under a permethrin-impregnated bed net in malarious areas unless they are in a sealed air-conditioned environment. Although anopheline mosquitoes are night biters, *Aedes* spp and culicine mosquitoes are usually day biters, so vigilance at all times of day is necessary.

Food and Water Precautions

Travelers to developing countries should be diligent in washing their hands frequently; avoiding food from dubious eating places, markets, and roadside vendors; avoiding buffets where there are no food covers or fly controls; avoiding high-risk food such as shellfish, reef fish (ciguatera risk), undercooked meats and poultry, dairy products, unpeeled fruits, cold sauces, and salads; avoiding both tap water and drinks or ice made from tap water; and using sealed bottled water or chemically treated, filtered, or boiled water for drinking and brushing their teeth.

Sex

Education on the incidence of HIV infection and sexually transmitted diseases among professional sex workers abroad, on the use of condoms, and on the failure rate of condoms (3 to 5% breakage/slippage) should be given regardless of the apparent circumstances of the traveler. Unprotected sex even with fellow travelers is considered high risk.[12] Travel is a disinhibiting experience in itself, and alcohol consumption tends to increase during travel.

Blood-borne Pathogens

Blood, blood products, syringes, and contaminated medical or dental instruments are a risk following accidents or trauma. Travelers should consider carrying an infusion set, needles, and a suture kit for high-risk areas. If possible, they should defer medical treatment and travel to a facility where safety can be ensured. Tattooing, acupuncture, and body piercing carry similar risks. Health care workers and others at risk in areas of high HIV prevalence without sophisticated medical infrastructure may consider carrying a 1- to 2-week supply of Truvada (200 mg of emtricitabine plus 300 mg of tenofovir), 1 tablet per day, plus raltegravir 400 mg, 1 tablet twice a day to begin immediate twice-daily postexposure prophylaxis, with the understanding that this is only an initial measure to allow time for travel to an adequate medical facility able to provide sophisticated testing and counseling.

Protection Against Skin Diseases

Infected mosquito bites are common. Practicing good hand hygiene in dirty environments and covering open wounds are preventive measures that all travelers should take. Scabies and lice infestations can be prevented by carrying out good personal hygiene. In Africa, all clothes dried outdoors should be ironed to avoid cutaneous myiasis due to the tumbu fly. Hats and sunscreen are mandatory in the tropics. Sunscreen should always be applied to skin before, and not after, an application of DEET.

Swimming and Water Exposure

Travelers should be instructed to avoid recreational (swimming, rafting, wading) or other exposure to fresh water in areas that are endemic for schistosomiasis. Hikers, bikers, and adventure travelers should consider prophylaxis with 200 mg of doxycycline once per week because of the significant risk of leptospirosis that exists in fresh water throughout the developing world. Walking barefoot in tropical areas predisposes to hookworm, *Strongyloides* infection, cutaneous larva migrans, and tungiasis.

Prevention of Tuberculosis

A predeparture baseline tuberculin skin test with annual retesting is indicated for long-stay travelers to developing countries. Aggressive treatment of skin test converters will prevent cases of active tuberculosis later. Travelers should avoid crowded public transportation or crowded public places and distance themselves immediately from anyone with a chronic or heavy cough. Expatriates should screen domestic help for tuberculosis.

NONINFECTIOUS TRAVEL PROBLEMS

Traveler's Thrombosis and Jet Lag

A causal relationship between travel-related immobility and deep venous thrombosis or pulmonary embolism in otherwise healthy travelers has become established. Risk of pulmonary embolus is essentially absent on flights lasting less than 6 hours. Those with clear, known risk factors are at highest risk. All travelers should avoid dehydration, avoid alcohol, and exercise the legs regularly in flight. Of many recommendations for prevention, only the use of graded 15 to 30 mm Hg compression stockings for those at higher risk is supported in trials, although prophylactic subcutaneous low-molecular-weight heparin just before departure and again 24 hours later for those with thrombophilia or previous thrombotic events is often used in practice. Aspirin therapy is of no proven benefit in this setting.

Jet lag (Chapter 405) occurs after crossing three or more time zones, and zolpidem (5 mg) taken for a few nights at bedtime at the destination is generally effective.

Altitude Illness

Whether ascending by car or airplane, acute mountain sickness occurs in at least 25% of people who ascend rapidly to 2500 m or more and in most people who go quickly to 3000 m or more. Gradual ascent during days is rarely practiced by modern travelers. For prevention of altitude illness, acetazolamide, 125 mg twice a day beginning the morning of the day before ascent and continuing through the day after ascent, is effective. If

symptoms of mountain sickness, such as nausea, vomiting, anorexia, lightheadedness, fatigue, or insomnia, persist beyond the day after ascent, travelers may continue to take one tablet each evening.[13] Other medications such as dexamethasone also can be used (Chapter 94). Severe complications, such as pulmonary or cerebral edema, occur uncommonly under 3500 m and are best treated by oxygen and immediate descent. Those traveling above 3500 m for longer than a brief transit of a few hours should consult an expert.

POST-TRAVEL CARE

The approach to the patient requires knowledge of world geography, the epidemiology of disease patterns in 230 or so countries, and the clinical presentation of a wide spectrum of disorders. Most illnesses are mild, most are self-limited, and many are noninfectious. Based on 43,000 ill-returned travelers seen by the GeoSentinel Surveillance Network, specific travel destinations are associated with the probability of the diagnosis of certain diseases (E-Table 286-1). Diagnostic approaches and empirical therapies can be guided by these destination-specific differences. Important region-specific disease occurrence data indicate that (1) febrile illness is most important from Africa and Southeast Asia; (2) malaria is one of the top three diagnoses from every region, yet during the past decade, dengue has become the most common febrile illness from every region outside of sub-Saharan Africa;[14] (3) in sub-Saharan Africa, rickettsial disease is second only to malaria as a cause of fever; (4) respiratory disease is most important in Southeast Asia and sub-Saharan Africa; and (5) acute diarrhea is disproportionately from South Central Asia.

Fever in a traveler who has recently returned from the tropics is a potential emergency and must be evaluated immediately so that antimalarial or other definitive treatment can be initiated rapidly if it is indicated. Persistent gastrointestinal symptoms in a returning traveler require prompt evaluation and treatment (E-Table 286-2).[15] Certain long-term travelers should be evaluated by a travel or tropical medicine specialist when they return to be screened for conditions that may be asymptomatic, such as schistosomiasis or strongyloidiasis.

Travelers who become ill during or any time up to several months after a foreign trip will frequently associate that illness with a possible travel-specific etiology. This may be the case, but often it is not. Routine things are common, and common things are common whether they are actually acquired during travel or at some time after the trip.

Grade A References

A1. Zanger P, Nurjadi D, Gabor J, et al. Effectiveness of rifaximin in prevention of diarrhoea in individuals travelling to south and southeast Asia: a randomized, double-blind, placebo-controlled, phase 3 trial. *Lancet Infect Dis*. 2013;13:946-954.

A2. Kahn SR, Lim W, Dunn AS, et al. Prevention of VTE in nonsurgical patients: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e195S-e226S.

A3. Low EV, Avery AJ, Gupta V, et al. Identifying the lowest effective dose of acetazolamide for the prophylaxis of acute mountain sickness: systematic review and meta-analysis. *BMJ*. 2012;345:e6779.

A4. Ritchie ND, Baggott AV, Andrew Todd WT. Acetazolamide for the prevention of acute mountain sickness—a systematic review and meta-analysis. *J Travel Med*. 2012;19:298-307.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

E-TABLE 286-1 SOME KEY DISEASES REPORTED IN FIVE OR MORE TRAVELERS AMONG >43,000 REPORTED ILLNESSES

DIAGNOSES	# OF CASES	TRAVEL HISTORY
GASTROINTESTINAL		
Giardia	1426	India, Thailand, Nepal, and Ghana
Campylobacter	753	India, Thailand, Indonesia, and Tanzania
Strongyloides	483	India, Vietnam, Ghana, and Dominican Republic
Salmonella enteritis	367	Thailand, India, Indonesia, and Egypt
Shigella	271	India, Egypt, Ghana, and Indonesia
Entamoeba histolytica	340	India, Indonesia, Mexico, and Thailand
Vibrio	9	5 cases in Latin America (2 cases acquired in Mexico), 3 in Asia, 1 in sub-Saharan Africa
FEBRILE		
Plasmodium falciparum	1990	Ghana, Comoros, Nigeria, and Côte d'Ivoire
Plasmodium vivax	480	India, French Guyana, Myanmar, and Papua New Guinea
Dengue	1473	Thailand, Indonesia, India, and Brazil
Enteric fever	467	India, Nepal, Pakistan, and Bangladesh
Spotted fever rickettsia	267	South Africa (68.9%), Zimbabwe, Tanzania, and Swaziland
Chikungunya	164	India, Malaysia, Indonesia, and Thailand
Hepatitis A	120	India, Morocco, Egypt, and Mexico
Acute HIV	84	Thailand, Brazil, Guinea, and Germany
Leptospirosis	83	Thailand, Laos, and Costa Rica
Hepatitis E	45	India (40%), Pakistan, and Bangladesh
Brucellosis, acute	33	India, Sudan, and Iraq
Measles	33	India, Thailand, and France
Histoplasmosis	23	Guatemala and Costa Rica
Rickettsia typhi (flea-borne)	17	Cambodia, Malaysia, Vietnam, Indonesia, and Nepal
Visceral leishmaniasis	16	India, Greece, Portugal, and Spain
Orientia tsutsugamushi (scrub typhus)	14	Thailand and Vietnam
Rubella	11	Vietnam and Thailand
Melioidosis	9	Thailand, Singapore, and Malaysia
Mumps	8	Different country for each case (2 in Western Europe, 4 in sub-Saharan Africa, 1 in Latin America, and 1 in Asia)
African trypanosomiasis	6	Zambia (3 cases), Tanzania, Zimbabwe, and Burkina Faso
Relapsing fever	6	Senegal, South Africa, and Morocco
Ross River virus	5	Australia
DERMATOLOGIC		
Rabies prophylaxis after bite or scratch	1249	Thailand, Indonesia, China, and India
Cutaneous larva migrans	806	Thailand, Brazil, Mexico, and Malaysia
Leishmaniasis (cutaneous or mucocutaneous)	264	Bolivia, Afghanistan, and Costa Rica
Myiasis	174	Senegal, Brazil, Costa Rica, and Belize
Tungiasis	87	Brazil, Madagascar, Uganda, and Ethiopia
Gnathostomiasis	12	Cambodia and Indonesia (92% from Asia)
Leprosy	11	Pakistan and Vietnam
Cutaneous atypical mycobacteria	6	All different (Caribbean, sub-Saharan Africa, and South-Central Asia)
RESPIRATORY OR PHARYNGEAL		
Influenza	367	
H1N1	176	United States, Australia, United Kingdom, and Philippines
Influenza A or B	191	Indonesia, Thailand, India, and China
Tuberculosis	170	
MDR or XDR pulmonary tuberculosis	3	1 case each for Nigeria and India (1 case unknown)
Legionellosis	35	China, Italy, and Spain
Pulmonary atypical mycobacteria	35	China, Thailand, Kenya, South Africa, and North America
Pertussis	30	India and China
NEUROLOGIC		
Ciguatera intoxication	51	Bahamas and Dominican Republic
Neurocysticercosis	21	India
Tuberculosis meningitis or tuberculoma	13	India and Pakistan
Scombroid, neurotoxic or paralytic shellfish poisoning	7	Mauritius (2 cases) and other (2 cases in Western Europe and 2 cases in Southeast Asia)
West Nile virus	5	1 case each in Afghanistan, Costa Rica, Greece, and Israel (1 unknown)

E-TABLE 286-1 SOME KEY DISEASES REPORTED IN FIVE OR MORE TRAVELERS AMONG >43,000 REPORTED ILLNESSES—cont'd

DIAGNOSES	# OF CASES	TRAVEL HISTORY
OTHER		
Schistosomiasis	792	Malawi, Uganda, Tanzania, and Ghana
Filariasis	113	Cameroon, Gabon, Democratic Republic of Congo, and Central African Republic
Lyme disease	77	United States, Germany, and Italy
Visceral larva migrans	16	6 cases in Southeast Asia (2 cases acquired in Thailand), 5 in Latin America and the Caribbean, 3 in sub-Saharan Africa, and 2 cases unknown
Fasciola	14	Australia, Germany, France, and the Netherlands
Clonorchis	12	Thailand, Laos, and China
Trichinellosis	12	Poland and Serbia

HIV = human immunodeficiency virus; MDR = multidrug-resistant; XDR = extensively drug-resistant.
Adapted from Leder K, Torresi J, Libman MD, et al. GeoSentinel surveillance of illness in returned travelers, 2007-2011. *Ann Intern Med.* 2013;158:456-468.

E-TABLE 286-2 COMMON INFECTIOUS AGENTS CAUSING CHRONIC GASTROINTESTINAL DISEASE IN THE RETURNING TRAVELER

	INCUBATION PERIOD (DAYS)	CLINICAL MANIFESTATIONS	DIAGNOSIS	TREATMENT
Giardia	7-10	Abdominal pain, nausea, persistent watery diarrhea	Stool microscopy and stool antigen	Metronidazole, 250 mg, 3 times/day for 7-10 days or 500 mg twice a day for 5-7 days
Entamoeba histolytica	11-21	Abdominal pain, fever, persistent watery diarrhea	Stool antigen assay	Metronidazole, 500-750 mg, 3 times/day for 7-10 days; plus paromomycin, 500 mg, 3 times/day for 7 days
Strongyloides	11-21	Larva currens, abdominal pain, persistent diarrhea	Stool microscopy	Ivermectin, 200 μg/kg of body weight/day for 2 days
Schistosoma	14-84	Katayama syndrome, abdominal pain, persistent diarrhea, hematuria	Stool and urine microscopy	Praziquantel, 40 mg/kg twice a day for 1 day for *S. hematobium* and *S. mansoni*, and 60 mg/kg 3 times/day for 1 day for *S. japonicum*
Campylobacter	1-4	Acute watery diarrhea, fever	Stool culture	Azithromycin, 500 mg once a day for 3 days
Shigella	1-8	Severe diarrhea, dysentery, fever	Stool culture	Ciprofloxacin, 500 mg twice a day for 3 days
Salmonella				
Serovar Typhi	5-14	Fever, headache, malaise, abdominal pain, diarrhea	Blood and stool culture	Ciprofloxacin, 20 mg/kg/day for 7 days; or azithromycin, 20 mg/kg/day for 7 days
Nontyphoidal	8-24 hours	Fever, headache, malaise, abdominal pain, diarrhea	Blood and stool culture	Ciprofloxacin, 20 mg/kg/day for 7 days; or azithromycin, 20 mg/kg/day for 7 days

Adapted from Ross AG, Olds GR, Cripps AW, et al. Enteropathogens and chronic illness in returning travelers. *N Engl J Med.* 2013;368:1817-1825.

GENERAL REFERENCES

1. Leder K, Torresi J, Libman MD, et al. GeoSentinel surveillance of returned travelers, 2007-2011. *Ann Intern Med*. 2013;158:456-468.
2. Recommendations and Guidelines. Adult immunization schedules—United States. Centers for Disease Control and Prevention. Updated annually. http://www.cdc.gov/vaccines/schedules/hcp/adult.html. Accessed January 21, 2015.
3. Travelers' Health. Destinations. Centers for Disease Control and Prevention. http://wwwnc.cdc.gov/travel/destinations/list. Accessed January 21, 2015.
4. Meltzer E, Schwartz E. Enteric fever: a travel medicine oriented view. *Curr Opin Infect Dis*. 2010;23:432-437.
5. Prevention and control of influenza with vaccines: recommendations of the Advisory Committee on Immunization Practices (ACIP)—United States, 2013-2014 influenza season. Centers for Disease Control and Prevention (CDC). *MMWR Recomm Rep*. 2013;62(RR–07):1-43.
6. Gotuzzo E, Yactayo S, Córdova E. Efficacy and duration of immunity after yellow fever vaccination: systematic review on the need for a booster every 10 years. *Am J Trop Med Hyg*. 2013;89:434-444.
7. Hills SL, Griggs AC, Fischer M. Japanese encephalitis in travelers from non-endemic countries, 1973-2008. *Am J Trop Med Hyg*. 2010;82:930-936.
8. Cholera vaccines: WHO position paper. *Wkly Epidemiol Rec*. 2010;85:117-128.
9. Zaidi D, Wine E. An update on travelers' diarrhea. *Curr Opin Gastroenterol*. 2015;31:7-13.
10. Lalani T, Maguire JD, Grant EM, et al. Epidemiology and self-treatment of travelers' diarrhea in a large, prospective cohort of Department of Defense beneficiaries. *J Travel Med*. 2014; [Epub ahead of print].
11. Lupi E, Hatz C, Schlagenhauf P. The efficacy of repellents against *Aedes*, *Anopheles*, *Culex* and *Ixodes* spp.—a literature review. *Travel Med Infect Dis*. 2013;11:374-411.
12. Vivancos R, Abubakar I, Hunter PR. Foreign travel, casual sex, and sexually transmitted infections: systematic review and meta-analysis. *Int J Infect Dis*. 2010;60:478-485.
13. Bartsch P, Swenson ER. Clinical practice: Acute high-altitude illnesses. *N Engl J Med*. 2013;368:2294-2302.
14. Kotlyar S, Rice BT. Fever in the returning traveler. *Emerg Med Clin North Am*. 2013;31:927-944.
15. Ross AG, Olds GR, Cripps AW, et al. Enteropathogens and chronic illness in returning travelers. *N Engl J Med*. 2013;368:1817-1825.

REVIEW QUESTIONS

1. A tourist visiting Burkina Faso (West Africa) for 1 week during the dry season (December to June) should receive which of the following vaccines?

A. Plague
B. Cholera
C. Hepatitis C
D. Meningococcal
E. Japanese encephalitis

Answer: D Burkina Faso is in the meningitis belt of Africa. There is no plague vaccine currently and no risk to routine tourists in Africa. Cholera vaccine is indicated only for aid and refugee workers. Japanese encephalitis occurs only in Asia.

2. A traveler to Lake Titicaca (3800 m) in Bolivia should consider pre-ascent prophylaxis with

A. Ibuprofen
B. Acetazolamide
C. Prednisone
D. Furosemide
E. Isosorbide dinitrate

Answer: B Acetazolamide hastens acclimatization by causing a bicarbonate diuresis and metabolic acidosis. Ibuprofen can prevent altitude headache but has no effect against more serious altitude manifestations. Dexamethasone but not prednisone can be used to treat but not to prevent cerebral edema. Furosemide has no effect against high-altitude pulmonary edema and does not prevent it. Nitrates are ineffective against the increased pulmonary artery pressures induced by altitude.

3. The most common vaccine-preventable disease of travelers is

A. Hepatitis A
B. Typhoid
C. Yellow fever
D. Influenza
E. Measles

Answer: D Studies indicate that up to 2% of all travelers acquire influenza during their trip. This is orders of magnitude more common than the risk of the other diseases listed.

4. Acute diarrhea occurs disproportionately in travelers returning from which region?

A. Africa
B. South America
C. South Central Asia
D. China
E. Caribbean

Answer: C Epidemiologic evidence from the GeoSentinel network is described in the text and can be found in reference 1.

5. Travelers should be instructed to avoid what exposure in areas that are endemic for schistosomiasis?

A. Recreational (swimming, rafting, wading) or other exposure to fresh water
B. Eating of unpasteurized dairy products
C. Caves
D. Undercooked shellfish
E. Ticks

Answer: A Schistosomiasis is acquired when larvae that are harbored by freshwater snails are excreted into lakes, rivers, streams, or ponds and penetrate the skin of the traveler.

287

ANTIBACTERIAL CHEMOTHERAPY

GEORGE L. DRUSANO

Antibiotics have been classified as "miracle drugs" since their introduction and have transformed our expectations regarding the outcome of infection. More recently, they have been the backbone of modern interventional medicine. Barriers not meant to be breached have been. Catheters have been inserted into veins and arteries, the bladder, and the tracheal tree. These interventions support the seriously ill patient, but they also give bacteria access to normally sterile areas. Therapies for cancer and immune-mediated disease often leave patients severely immunosuppressed. Bacterial infections in such patients are serious and, when untreated or treated late, frequently result in death. Antibacterial agents allow critical life-saving support for such patients.

What is clear is that the overuse and poor use of antibiotics have allowed many pathogens to develop resistance to drugs. Multiresistant *Staphylococcus aureus* has become a plague both in the hospital and, of late, in the community. Extended-spectrum β-lactamases and *Klebsiella pneumoniae* carbapenemase enzymes have mediated resistance to many of our most potent and broad-spectrum β-lactam agents, including the carbapenems. Consequently, it is important to understand the principles of antibacterial chemotherapy to obtain the best clinical outcomes for our patients but also, in a broader sense, to lower the probability of the emergence of resistance and to maintain the potency of the drugs we currently have in our therapeutic armamentarium.[1,2]

CHOICE OF ANTIBIOTIC, ANTIBIOTIC DOSE, AND SCHEDULE TO OPTIMIZE CLINICAL OUTCOME

As in all of clinical medicine, a well-done history and physical examination are central to proper decision making and to the achievement of optimal therapeutic outcomes in patients with infections. Key to choosing the correct drug, dose, and schedule is recognition that an infection exists. The next step is to document where the infection exists and then to identify the dominant organisms present at each site of infection. The next step is to determine whether there are any risk factors that might predict the presence of drug-resistant pathogens.

As an example, we know that community-acquired pneumonia is caused by certain traditional pathogens. *Streptococcus pneumoniae*, *Haemophilus influenzae*, and perhaps *Moraxella catarrhalis* are the "classic" bacterial pathogens associated with this entity. "Atypical" pathogens such as *Legionella* species, *Mycoplasma pneumoniae*, and *Chlamydophila pneumoniae* may also be seen. In contrast, intra-abdominal infections are dominated by *Escherichia coli* and other Enterobacteriaceae and anaerobic organisms such as *Bacteroides* species. Consequently, it is important to understand the dominant pathogens present at different infection sites so that the best drug or combination of drugs can be chosen to treat the infection.

Knowing the source of infection is also critical because drugs penetrate differently into different spaces.[3] Classically, penetration is poorer into spaces where there are tight junctions, such as the central nervous system, the eye, and the prostate. In general, the penetration of many classes of antibacterial agents is good into complicated skin and skin structure infection sites. What is often not appreciated is the divergent penetrations of different agents and even agents within the same class into the lung to treat bacterial pneumonia. For instance, the penetration by macrolide antibiotics into skin infection sites is modest, but their penetration into the lung is good, with penetration ratios (area under the concentration-time curve [AUC] in epithelial lining fluid/AUC in plasma) ranging from 4 to 20. The penetration of β-lactam drugs can range from 15 to 100%, but there is no set of variables (at least to date) that explains such a range of penetrations. Table 287-1 presents a partial list of

TABLE 287-1 PRIMARY INFECTION SITES AND THE DOMINANT BACTERIAL SPECIES PRESENT

SITE	BACTERIA
Complicated skin or skin structure infection	*Staphylococcus aureus* and *Streptococcus* spp
Diabetic foot ulcer	Organisms above plus Enterobacteriaceae
Intra-abdominal infections	*Escherichia coli* and other Enterobacteriaceae plus anaerobes
Community-acquired bacterial pneumonia	*Streptococcus pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, "atypical" pathogens
Hospital-acquired pneumonia	*S. aureus*, *Klebsiella pneumoniae*, *Enterobacter* spp, *Pseudomonas aeruginosa*, *Acinetobacter* spp
Meningitis	*S. pneumoniae*, *H. influenzae* (nontypable), meningococci; in hospital settings, *S. aureus* and gram-negatives may be seen
Urinary tract infections	Enterobacteriaceae, particularly in sexually active women; multiresistant gram-negatives in patients with complicated urinary tract infections or those instrumented; enterococci, particularly in elderly men
Prostatitis	Enterobacteriaceae, enterococci, atypical pathogens

infection sites and their dominant pathogens. The point is not to have an encyclopedic knowledge of infection sites or of pathogens but to appreciate that different infection sites require different drugs to provide adequate coverage for the most likely pathogens present.

It is also important to understand other factors that increase the probability that a resistant organism is present at the primary infection site. An example is a patient who has recently taken antibiotics before acquiring the present infection. Other examples are the acquisition of an infection in the hospital or in an extended care facility and in a patient who is immunosuppressed. In such patients, the choice of antimicrobial agents must be carefully considered to cover more resistant pathogens.

CULTURE AND GRAM STAIN

Once the site of infection has been definitively identified or the most likely source of infection has been determined, it is critical to obtain culture specimens from that site as well as to obtain blood culture samples. Coordination with the microbiology laboratory is key to make sure that culture specimens are handled appropriately. Also of great importance is the performance and interpretation of a Gram stain on the specimen. This can be straightforward if the specimen is from a normally sterile space, or it may require considerable skill in interpretation if the specimen is from an area where mixed flora is normally present, such as sputum.

Although it is not possible to make a definitive diagnosis on the basis of the cellular morphology of the organism, it is possible to combine information on the morphology, the organism's gram positivity or negativity, the infection site, and the most likely pathogens to make an initial antimicrobial choice with the highest probability of producing effective chemotherapy. The initial antibiotic chosen to cover organisms present at the primary infection site has a significant influence on the outcome of therapy. Initial choices should err on the side of caution. When definitive cultures are available, the chemotherapy can be "streamlined" to provide the most effective and least toxic antimicrobial for the patient. As the seriousness of the infection increases, providing the correct initial coverage becomes more important with regard to the ultimate outcome.

SUSCEPTIBILITY

Because infections occur at a specific place, it is important to have an understanding of the antimicrobial susceptibility patterns in one's particular hospital. Most frequently, the microbiology susceptibility patterns are different in the culture specimens taken from patients on the general wards as opposed to those in intensive care units (ICUs) because the former are generally (but not always) infected with pathogens derived from the community setting. In severely infected patients, particularly those whose infections were acquired in the ICU, it is critical to know the susceptibility patterns for these pathogens, which are likely to be multiply resistant to different classes of agents.

After the definitive culture, the identified pathogen must be examined in an antimicrobial susceptibility test. This can be automated, in which case the information returned is the minimal inhibitory concentration (MIC), or it can be reported from a disc diffusion test, where the results are typically reported as S (susceptible), I (intermediate), or R (resistant). The MIC is often misunderstood as the concentration of drug that prevents the pathogen from growing, but it is actually the concentration of drug that allows a tube (or well) containing the pathogen to remain clear by visual examination after 18 to 24 hours. If one starts with an organism concentration of 1 to 5×10^5 colony-forming units (CFUs)/mL, this criterion can actually allow almost a 1 $\log_{10}$ (CFU/mL) increase in bacterial count over this time frame and still be read as the MIC. With all its limitations, the MIC provides critical information about drug choice. For several infections, such as meningitis, endocarditis, and perhaps bacteremia, knowing the minimal bactericidal concentration (defined as the concentration required for kill of 99.9% of organisms during 18 to 24 hours) is valuable. In these cases, it is important to obtain multi-log killing of the organism to ensure a high probability of a good clinical outcome. For meningitis and endocarditis, there is also the confounding problem of penetration of the drug to the primary infection site. In other circumstances (e.g., ventilator-associated pneumonia) in which bacterial burdens are high, optimizing the probability of cure requires highly bactericidal therapy.

DETERMINING THE "CORRECT" DRUG DOSE

The correct drug dose is the one that produces a high likelihood of achieving a good clinical response with a low probability of causing a concentration-driven adverse event. Another consideration is that the optimal dose should have a high probability of suppressing the emergence of resistant mutants. To be confident that a drug dose has a high likelihood of producing a good clinical outcome requires an understanding of the relationship between the measure of drug exposure and infection outcome, that is, pharmacodynamics (Chapter 29).

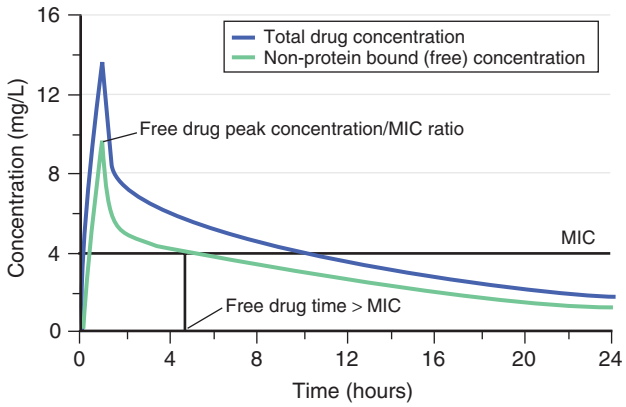

FIGURE 287-1. Three measures of drug exposure are important to the measure of potency, or the minimal inhibitory concentration (MIC). These measures are free drug peak concentration/MIC ratio, free drug time greater than MIC, and free drug area under the curve (AUC)/MIC ratio.

When we administer a drug either intravenously or orally, the drug concentration starts low, increases to a maximal value, then declines over time until another dose of the drug is given (in a multidose regimen). This drug concentration–time profile should be seen relative to a measure of drug potency against the pathogen in question: the MIC (or minimal bactericidal concentration). It is also critical to note that protein binding is important because in almost all cases, it is the free or non–protein bound drug that kills the causative pathogen. So understanding the pharmacodynamics of a drug or class of agents first requires knowledge of the concentration-time profile, protein binding, and MIC (Fig. 287-1). For drugs that kill organisms much faster as their concentrations rise (e.g., fluoroquinolones, aminoglycosides), the free drug AUC/MIC ratio is most closely linked to drug effect. For other drugs (e.g., penicillins, carbapenems, cephalosporins, monobactams), the rate of organism kill rises with concentration and plateaus quickly, in which case the free drug time greater than MIC is most closely linked to drug effect. On occasion, the free drug peak concentration/MIC ratio is linked to drug effect. This is seen when there is a rapid and frequent emergence of resistance and the peak concentration helps suppress the amplification of resistant mutant subpopulations.

These measures of drug exposure need to be linked to drug effect. There are a number of ways to do this, but two are most common. In the first and most valuable way, patients are studied during the course of a clinical trial. Multiple plasma samples are taken for documentation of the drug's pharmacokinetics and corrected for protein binding. The MIC of the infecting pathogen is determined for the drug being administered, and the three measures of drug exposure cited earlier are calculated. The outcome of the patient is then determined (usually success or failure, but in some instances, time to success or failure serves as the outcome measure). If the outcome is success or failure, the appropriate measure of drug exposure is linked to the outcome through logistic regression analysis. By creating such a relationship, we can directly estimate the probability of a patient's having a successful outcome if a certain measure of drug exposure is achieved. For example, if a patient receiving a drug that is a concentration-dependent killing agent achieves a free drug AUC/MIC ratio of 88, the logistic regression relationship allows the calculation that the probability of a good outcome is a specific number (perhaps 91%). During the past 10 years, we have been able to generate such relationships for a great number of drugs and across drug classes. The most common relationships are for the fluoroquinolone antimicrobials, but examples can be found for the aminoglycosides, oxazolidinones, β-lactams, and glycylcyclines, among others.

The second way to link exposure to antibacterial effect is through preclinical animal models of infection. Here, the end point is most often microbiologic, although survivorship models have been employed. We observe the drug concentration–time profile in the animals and at the end of the experiment determine the number of organisms left after therapy at the infection site (the baseline number of organisms having been previously determined).

We can then link the measure of drug exposure (as earlier) to the number of bacterial cells killed by a specific regimen. It is most often the case that drugs are administered to the animals in a "dose fractionation" fashion. For example, 400 mg/kg/day regimen of a β-lactam may be administered as the whole dose daily, half the dose every 12 hours, or one quarter of the dose every 6 hours. In this way, we can determine whether free drug peak concentration/MIC ratio, free drug AUC/MIC ratio, or free drug time greater than MIC is the best measure of drug exposure for the class of drug being studied.

How does knowing the relationship between drug exposure and microbiologic effect in an animal infection model help in treating patients? On the basis of experience with different drugs and drug classes, it has become possible to link a drug's microbiologic effect in an animal model to clinical outcome.[4] For instance, in the therapy of complicated skin and skin structure infections, a drug exposure in the animal model that stops organisms from growing for 24 hours or reduces their number by approximately 1 $\log_{10}$ (CFU/g) is adequate to generate a high probability of a good clinical outcome. This has been documented for oxazolidinones and glycylcyclines. For severe community-acquired pneumonia or for hospital-acquired pneumonia, a greater cell kill—generally 2 $\log_{10}$ (CFU/g) or greater—is required for a high likelihood of a good clinical outcome.

When we have clinical data with a logistic regression function, it is possible to identify an exposure target that is associated with a specific probability of a good outcome (e.g., a free drug AUC/MIC ratio of 88 generates a 91% probability of a good outcome). In the case of the animal model, it is possible to identify a target drug exposure that will achieve a certain microbiologic effect associated with a good clinical outcome (the logic is one step removed). In both cases, we have a target drug exposure, and this is used to identify the "correct" dose.

When the same dose of a drug is administered to a number of patients, no two concentration-time curves will be identical. Indeed, it is not unusual to have a 10-fold range in measures of drug exposure (e.g., peak concentration, trough concentration, AUC). When these are transformed by incorporating MIC values, ranges of 100- to 1000-fold are commonly encountered. As part of the drug development process, the pharmacokinetics of new agents are now identified in a large number of patients with the infection of interest. As part of this process, the interpatient variability in important pharmacokinetic parameters (e.g., drug clearance = 10 ± 4 L/hour) is determined. This information is then used with a tool called Monte Carlo simulation to ask and answer a key question: If I give this specific drug dose to a large number of patients and I recognize that the MIC values I am likely to encounter will range (for example) between 0.06 and 8 mg/L, how many patients will achieve the desired measure of drug exposure to give them a high likelihood of a good clinical outcome? This can be calculated directly from the data indicated earlier plus the MIC distribution for the drug and the type of infection one wishes to treat. Indeed, many of the dosage recommendations for new drugs are calculated in this manner and validated in large phase III clinical trials. It provides the clinician with a degree of confidence that the recommended doses are likely to work if the infecting pathogen is still susceptible to the drug.

EMERGENCE OF RESISTANCE

If the drug is properly chosen, the organism should have a high probability of being susceptible to it. However, during the course of therapy, the organism has many mechanisms available to make it less susceptible (the MIC increases) to the drug being used. This leads to less effect than was originally envisioned, which results in a lower probability of cure.

When are organisms likely to become resistant? There are many determinants, but five factors account for the majority of cases of resistance emerging during therapy. The first is the mutational frequency of the organism being treated. Because mistakes are made by the organism during replication, mutations occur in the genome at a specific rate, which is organism dependent. For antibiotic resistance, the mutational frequency rate is generally around $1/10^8$ to $1/10^6$. There are organisms, however—referred to as hypermutators—that have mutations in other places in the genome (often mutS) that alter their DNA replication error-checking mechanisms. These organisms' mutational frequencies are generally 10- to 100-fold higher than those of organisms without these other mutations.

The second factor is related to the bacterial burden. The higher the bacterial burden, the more likely it is that a preexistent mutant is already extant in the bacterial population. Antibiotic pressure then provides this mutant with a selective advantage, and it amplifies in the population while its more susceptible cousins are killed by the antibiotic. As the bacterial burden increases and ultimately surpasses the inverse of the mutational frequency to resistance, it becomes more and more likely that a resistant mutant exists in the population. For instance, if the mutational frequency to resistance is $1/10^7$ (1 resistant organism per a population of 10^7) and the patient has a bacterial infection in which the total bacterial burden is 10^9 organisms, it is highly likely that a resistant organism is present in the population a priori. A calculation of the actual probability can be performed by a Poisson distribution, the bacterial burden, and the mutational frequency to resistance. Consequently, infections in which the bacterial burden is high are more likely to generate resistance during therapy. For example, in clinical trials of ventilator-associated pneumonia, single-agent β-lactam drugs or fluoroquinolones allowed the emergence of resistance during therapy 33 to 50% of the time.

The third issue is drug penetration. There are circumstances (e.g., empyema) in which the bacterial burden is high but the drug's penetration to the infection site is reduced. In this case, the emergence of resistance is more likely. In contrast, poor penetration with relatively low bacterial burdens (e.g., meningitis) generally does not result in a high probability of the emergence of resistance.

The fourth issue has to do with error-prone replication in bacterial pathogens. When resistance occurs relatively late in therapy and the bacterial burden is modest, error-prone replication is often to blame. Antibiotics differ greatly with respect to their ability to induce the bacterial isolate to perform error-prone replication. Perhaps the best example is the fluoroquinolone antimicrobials because these drugs strike at the heart of DNA replication. The organism senses the attack of the antibiotic, and a whole cascade of events takes place, the most important of which is the induction of error-prone polymerases (e.g., pol V). These polymerases markedly increase the error rate in DNA replication. Most of these errors are lethal to the organism and therefore unhelpful to it. However, because this is a totally random process, by chance, a mutation can occur in a spot that provides protection from the onslaught of the antibiotic (e.g., a mutation in the DNA gyrase when the patient is receiving a fluoroquinolone). We perceive this as the organism's having a selective replication advantage (emergence of resistance), and we see an increase in the MIC of the mutant organism relative to the baseline organism.

The fifth factor has to do with mechanisms other than antibacterial target site mutations that allow the organisms to survive in the face of appropriate antibiotic chemotherapy. One extremely common mechanism seen in the majority of both gram-positive and gram-negative organisms is the upregulation of efflux pumps. These pumps are indiscriminate in their ability to pump molecules; they can eject multiple classes of antibacterials from the organism as well as natural substances that can harm it, such as metal ions. These pumps keep drug concentrations at their target sites much lower than they would be in the absence of the pumps. The pumps can be induced and then downregulated once the threat has passed, or occasionally, the organism can pick up a mutation in the part of the genome where expression of the pump is regulated, so the pump is always expressed (constitutive expression).

A similar process is seen with the production of β-lactamases. Often, these enzymes are situated on plasmids and are produced all the time. Sometimes, as with the efflux pumps, they sense the attack by the β-lactam, and their production is markedly increased (the phenomenon of induction, seen with ampC-type enzymes, which generally reside on the chromosome). Sometimes, also like the pumps, the organisms pick up a mutation in the part of the genome that regulates production of the β-lactamase. This is referred to as stable de-repression, and the enzyme is made all the time. The enzyme hydrolyzes its substrate (the β-lactam drug), thus preventing the binding to the target sites, the β-lactam–binding proteins.

Finally, in gram-negative organisms, the drug must cross the diffusional barrier of the outer membrane before binding to β-lactam–binding proteins (if the drug is a β-lactam) in the periplasm of the organism, or it must cross the inner membrane if its target site is actually inside the organism. For many agents, particularly those that are water soluble, a large percentage of their influx is due to passage through porin proteins. These proteins are water-filled channels that pass through the entirety of the outer membrane of gram-negative bacteria. Part of their function is to provide access to nutrients for the organism and allow the easy diffusion of waste products. These channels are also used to obtain passage by water-soluble antibacterial agents. The organism has the ability to downregulate these channels, either temporarily or permanently (by mutation). An example is *Pseudomonas aeruginosa*, in which downregulation of the porin channel oprD markedly diminishes the penetration of carbapenem antibiotics (β-lactam agents), with a resultant

two- to eight-fold rise in MIC. With all these mechanisms that can be brought to bear, it is no surprise that poor dosing and prolonged courses of therapy often lead to the emergence of resistance.

These mechanisms also interact. In the case of fluoroquinolone antimicrobials, where error-prone replication almost always occurs under pressure, the upregulation of efflux pumps relieves some of the antibiotic pressure, allowing the organisms to undergo more rounds of replication per unit of time, giving the error-prone replication mechanism more time to find a mutation that is not lethal and provides protection against the drug at the primary target site.

SUPPRESSING THE EMERGENCE OF RESISTANCE

Suppressing the emergence of resistance is an end point that is different from achieving a good clinical or microbiologic outcome. The antimicrobial exposure (free drug peak concentration/MIC ratio, free drug AUC/MIC ratio, or free drug time greater than MIC) necessary to suppress resistance is always at least as high as that required to attain maximal killing of bacterial cells, and in many instances, it is substantially higher.

Recently, it has been shown experimentally that it is possible to identify doses and schedules of drugs that suppress the emergence of resistance as well as provide optimal killing of bacterial cells. Unfortunately, many of the exposures necessary to achieve such an outcome are high enough to result in increased drug toxicity. Nevertheless, as new agents are developed, it will become important to identify resistance-suppressive exposures and to ascertain whether they are too toxic to be used clinically, as a way of retaining the potency of the antibacterial therapeutic armamentarium.

The other way to help suppress the emergence of resistance is to employ combination chemotherapy. Combination therapy has major advantages, but there are significant disadvantages as well. Among the advantages are that combination therapy can improve the spectrum of coverage in the setting of empirical treatment as well as help suppress the emergence of resistance. This is best seen in the treatment of *Mycobacterium tuberculosis*, although there is growing evidence that this may also be true for difficult-to-treat pathogens such as *P. aeruginosa* and *Acinetobacter* species. Combination therapy can be effective against some pathogens in serious infections for which a single agent would not be adequate. Enterococcal endocarditis is the classic example; ampicillin or vancomycin is generally static and not up to the task of curing this disease, and the second agent (an aminoglycoside, streptomycin, or gentamicin) is somewhat resistant when it is used alone (MIC values <500 mg/L). Yet the combination produces excellent bacterial kill in the endocarditic vegetation, with a high probability of cure if the patient can tolerate the combination for the requisite time. Some authors have discussed reducing drug doses in combination as a way of ameliorating toxicity, but this theoretical advantage is difficult to demonstrate convincingly in the clinical setting.

The disadvantages of combination therapy are also well known. At full doses of each, there may be added toxicity, and there are certainly greater costs associated with combinations. Finally, combinations of agents may interact antagonistically instead of either additively or synergistically. For example, the combination of tetracycline and penicillin is somewhat antagonistic and caused a number of failures in the treatment of pneumococcal meningitis in the 1950s. However, there are circumstances in which weak antagonism is tolerable. It has been shown that isoniazid plus rifampin, part of the standard therapy for *M. tuberculosis*, is somewhat antagonistic with respect to cell killing, but the combination provides superb protection against the emergence of resistance. Consequently, in discussing drug interactions (synergy, additivity, antagonism), it is important to be specific about the end point to which one is referring.

MECHANISM OF ACTION

The search for new antimicrobials begins with the principle that the target that affects pathogens must not be in the human genome or must be sufficiently different that the agent in question does not cause appreciable human toxicity. β-Lactam antibiotics are a good example; the inhibition of peptidoglycan synthesis has no effect on humans because this target is simply not present. An example of differing susceptibility to inhibition because of poor sequence homology between bacteria and humans can be seen in the inhibitors of the bacterial ribosome. The bacterial ribosome is much smaller than that in humans and has differing affinities for such drugs as aminoglycosides, tetracyclines, macrolides, and clindamycin. Consequently, such agents are an important part of the therapeutic armamentarium and cause only minor adverse effects related to their primary mode of action. In some cases, so-called off-target toxicity may take place. Again, an example is the β-lactams; a major toxicity is accelerated allergic reaction, even though the target for the agents is not present in humans. Table 287-2 lists the sites of action and the effects of many antimicrobial agents.

Often, antimicrobials are thought of as either *bacteriostatic* or *bactericidal*, for which there are standard definitions. An agent that causes a 1000-fold decline (3 $\log_{10}$ [CFU/mL] decrease) in an in vitro test system during 18 to 24 hours is defined as being bactericidal. Any agent that causes a smaller decline is defined as bacteriostatic. It is clear that these definitions are arbitrary; however, there is a clinical connection. In some circumstances, such as meningitis and endocarditis, it is critical to kill every last organism with the combination of drugs employed and the immune system. Obviously, the more organisms the antimicrobial kills, the easier this is to accomplish. In other circumstances, such as multilobar pneumonia, it is critical to kill as many of the organisms as possible so as not to overwhelm the body's immunologic defenses. For these reasons, clinicians generally prefer bactericidal to bacteriostatic agents.

MECHANISMS OF RESISTANCE

Earlier, how to suppress the emergence of resistance was examined. Here, the mechanisms by which organisms can become less susceptible to antimicrobial agents are examined. As stated earlier, in most cases, resistance is caused by an alteration in the target site; by an enzyme that alters the drug, resulting in a lack of activity; by the drug's being pumped out of the organism; or, in the case of gram-negative pathogens, by the loss or downregulation of a transmembrane porin protein. These mechanisms, along with error-prone replication, can interact to cause large increases in the MIC. For example, a gram-negative pathogen that acquired a mutation in its gyrase enzyme would generally have only a four-fold increase in its MIC value for a fluoroquinolone antimicrobial; however, if this isolate also had an efflux pump upregulated, the MIC could change 8- to 16-fold. Table 287-3 shows the mechanisms of resistance for multiple drug classes as well as the most common organisms in which these mechanisms are seen.

EFFECTS OF PHARMACOKINETIC CHANGES

Whereas antimicrobial pharmacodynamics is the study of the effect of a drug on an infecting organism, pharmacokinetics is the study of the effect of the body's processes on the drug concentration–time profile and the drug's ability to penetrate to the infection site (or the toxicity site). As such, it is a prime determinant of whether a drug will be able to kill or to inhibit the offending pathogen and, because many toxicities are concentration related, whether a serious drug-related toxicity will occur.

Earlier, the algorithm for identifying appropriate drug doses and schedules was outlined. Table 287-4 shows recommended doses and schedules of important antimicrobial agents as well as their protein-binding ability and whether alterations in renal or hepatic function generate major changes in the concentration-time profile. As in all chemotherapy, the aim is to generate a concentration-time profile in the plasma to generate a concentration-time profile at the infection site that allows the drug to inhibit or to kill the pathogen without causing toxicity.

Although there is almost always a guide to the concentration-time profile that results in an appropriate antimicrobial effect, it is more difficult to identify a linkage between drug exposure and the occurrence of toxicity. Data in this regard are available for the aminoglycosides, including a considerable amount of information on the association between drug concentrations and the likelihood of nephrotoxicity, and for daptomycin, for which a link between exposure and the likelihood of creatine kinase elevation has been elucidated.

For aminoglycosides, the relationships between drug exposure and the likelihood of a good clinical outcome and between drug exposure and the likelihood of nephrotoxicity have been determined. These are in the form of logistic regression functions, so the actual probability of both outcomes can be calculated. The one difference is that for the relationship with good outcome, the MIC value is part of the evaluation, whereas it does not figure in the toxicity relationship. Nevertheless, it is possible to derive appropriate information by making the outcome relationship MIC specific. Figure 287-2 illustrates the effect and toxicity relationships for aminoglycosides at three different MIC values. As shown, it is relatively easy to achieve a high probability of good clinical outcomes with aminoglycosides when the MIC is 0.25 mg/L, but it is virtually impossible to do so when the MIC rises to a value of 1.0 mg/L. The evaluation in the figure is based on twice-daily dosing; daily aminoglycoside dosing markedly improves this circumstance. Also,

TABLE 287-2 MECHANISMS OF ACTION OF ANTIMICROBIAL AGENTS

AGENT	SITE OF ACTION	EFFECT	BACTERICIDAL	BACTERIOSTATIC
β-Lactams (penicillins, cephalosporins, carbapenems, aztreonam)	Cell wall: penicillin-binding proteins	Inhibit cross-linking of peptidoglycan (transpeptidation), impair cell wall synthesis	+	Occasionally (enterococci)
Vancomycin, teicoplanin, dalbavancin, telavancin	Cell wall: terminal D-alanyl-D-alanine of pentapeptide peptidoglycan precursor	Inhibit polymerization of disaccharide precursors to peptidoglycan (transglycosylation), impair cell wall synthesis	+	Occasionally (enterococci)
Daptomycin	Cell membrane	Rapid depolarization of membrane potential	+	Occasionally (enterococci)
Aminoglycosides	Protein synthesis: 30S ribosome subunit	Inhibit peptide elongation, cause misreading of genetic code, inhibit protein synthesis	+	
Tetracyclines, glycylcyclines	Protein synthesis: 30S ribosome subunit	Inhibit binding of transfer RNA, inhibit protein synthesis	Occasionally	+
Chloramphenicol	Protein synthesis: 30S ribosome subunit	Blocks attachment of aminoacyl transfer RNA, inhibits protein synthesis	Occasionally	+
Macrolides, azalides, ketolides	Protein synthesis: 50S ribosome subunit	Block transfer of amino acids to peptide chain, inhibit protein synthesis	Occasionally	+
Clindamycin	Protein synthesis: 50S ribosome subunit	Blocks transfer of amino acids to peptide chain, inhibits protein synthesis	Occasionally	+
Quinupristin-dalfopristin	Protein synthesis: 50S ribosome subunit	Block extrusion of peptide chains, inhibit protein synthesis	+	+ (with quinupristin resistance)
Linezolid	Protein synthesis: 50S ribosome subunit	Blocks formation of 70S initiation complex, inhibits protein synthesis	Occasionally	+
Rifampin	Nucleic acid synthesis: β-subunit of DNA-dependent RNA polymerase	Inhibits RNA synthesis	+	
Metronidazole	Nucleic acid synthesis	Damages nucleic acids, inhibits DNA synthesis	+	
Quinolones	Nucleic acid synthesis: DNA gyrase, topoisomerase IV	Impair supercoiling of DNA, prevent decatenation of DNA molecules after replication, inhibit DNA synthesis	+	
Sulfonamides	Folic acid synthesis: dihydropteroate synthetase	Competitive inhibition of synthesis of dihydrofolate from *p*-aminobenzoic acid, pteroate, and glutamic acid	Occasionally (when used with trimethoprim)	+
Trimethoprim	Folic acid synthesis: dihydrofolate reductase	Inhibits reduction of dihydrofolate to tetrahydrofolic acid	Occasionally (when used with sulfonamide)	+

these relationships permit one to calculate the probabilities of effect and toxicity when, for example, aminoglycoside AUC is increased owing to renal dysfunction because these agents are largely eliminated renally.

When no toxicity relationship is available, there is still a target for good clinical effect. Achievement of this target should be paramount, and alterations for renal or hepatic impairment should strive to maintain the high likelihood of effect seen in patients with a relatively normal clearance function. One can change the dose or schedule of a drug to decrease its accumulation in the presence of renal or hepatic impairment (depending on the drug) and then recalculate the impact on the likelihood of attaining a good clinical outcome. One can also calculate the amount of accumulation with the proposed dose reduction or extension of the dosing interval. For the increase in exposure (relative to that in normally clearing patients), there is no clear guidance, but one can accept a certain maximal amount of accumulation as long as the proposed dose adjustment maintains a high likelihood of a good outcome. The acceptability of the increased drug exposure after dosage adjustment is usually based on a combination of preclinical toxicology and the largest exposures seen in phase I and phase II clinical trials. However, the overarching issue is that the proposed dose or schedule alteration maintains a high probability of a good clinical outcome.

DRUG CLASSES AND THEIR PROPERTIES

During the past 70 years, a large number of different classes of antimicrobial agents have been developed. These classes differ in their mechanisms of action, mechanisms of emergence of resistance, and whether they kill substantial numbers of organisms or only shut off bacterial growth. The following sections examine some of the properties of the major classes of antimicrobial agents in use today.

β-Lactam Agents

This class of drugs is arguably the most important group of antimicrobials. With chemical modification, they have an exceptionally broad spectrum of activity and, in general, an excellent safety profile. The major toxicity is related to allergic reactions to a degradation product of the drug. There are a number of different types of β-lactams, including the penicillins, the cephalosporins, the monobactams, and the carbapenems.

These agents bind to their targets, the β-lactam–binding proteins (sometimes referred to as penicillin-binding proteins). These binding proteins have an active site serine, and the drug forms a covalent bond with this site through the carbonyl of the β-lactam ring. Sometimes, the binding has direct effects on the organism's shape. For example, in gram-negative organisms, binding to penicillin-binding protein (PBP)–2 causes the organism to assume a spherical shape, whereas binding to PBP-3 causes the formation of long chains of organisms. In general, high-affinity binding to PBP-1 leads to the rapid death of the organism, sometimes accompanied by lysis. The classic response of *S. pneumoniae* to penicillin G is rapid lysis of the organism. Binding to PBP-1 (1a or 1b) leads to activation of *N*-acetylmuramic acid amidase, which destroys the bacterial cell wall, resulting in lysis.

The most common way for pathogens to protect themselves from β-lactams is by elaborating β-lactamases.[5] The genes for these enzymes may be on plasmids or other bits of transmissible DNA, or they may reside on the bacterial chromosome. Some drugs, such as the carbapenems, are resistant to hydrolysis by many enzymes (but certainly not all of them, especially the metallo-β-lactamases and *K. pneumoniae* carbapenemase-type β-lactamases). One of the ways developed to protect these drugs is to add a second agent, the β-lactamase inhibitor. Examples include potassium clavulanate,

TABLE 287-3 MECHANISMS OF ANTIMICROBIAL RESISTANCE

ANTIBACTERIAL AGENT	MECHANISM	REPRESENTATIVE ORGANISM
β-Lactams (penicillins, cephalosporins, carbapenems, aztreonam)	Altered target (penicillin-binding proteins)	Methicillin-resistant *Staphylococcus aureus* (MRSA), penicillin-resistant *Streptococcus pneumoniae, Enterococcus faecium*
	Reduced permeability	*Enterobacter* spp, *Pseudomonas aeruginosa, Acinetobacter spp*
	Enhanced efflux	*P. aeruginosa, Acinetobacter* spp
	β-Lactamases	*S. aureus*, Enterobacteriaceae (includes ESBLs), *Haemophilus influenzae, Moraxella catarrhalis, Neisseria gonorrhoeae, Enterococcus faecalis, P. aeruginosa, Acinetobacter spp*
Aminoglycosides	Inactivating enzymes (acetylation, adenylation, phosphorylation)	*S. aureus,* enterococci, *P. aeruginosa,* Enterobacteriaceae
	Reduced permeability	Enterobacteriaceae, *P. aeruginosa,* enterococci
	Enhanced efflux	*P. aeruginosa*
	Decreased ribosomal binding	*S. aureus, E. faecalis,* mycobacteria (streptomycin), gram-negative pathogens (aminoglycoside ribosomal methylase)
Chloramphenicol	Enhanced efflux	*H. influenzae*
	Reduced permeability	Enterobacteriaceae
	Inactivating enzyme (acetylation)	*S. aureus, S. pneumoniae,* enterococci
Daptomycin	Altered target	*S. aureus*
Glycylcyclines	Enhanced efflux	Enterobacteriaceae, especially *Proteus*
Macrolides, clindamycin, ketolide, quinupristin	Altered target (methylation of ribosomal RNA)	*S. aureus, S. pneumoniae* (not ketolide), streptococci, *Bacteroides fragilis*
	Enhanced efflux (not clindamycin or ketolide)	*S. pneumoniae*, streptococci
	Reduced permeability	Enterobacteriaceae
	Inactivating enzymes	*Escherichia coli, Klebsiella pneumoniae, S. aureus*
Linezolid	Altered target	Enterococci, *S. aureus*
Quinolones	Altered target (DNA gyrase, topoisomerase IV)	Enterobacteriaceae, *P. aeruginosa*
	Reduced permeability	Enterobacteriaceae, *P. aeruginosa*
	Enhanced efflux	*E. coli, P. aeruginosa*
Tetracyclines	Altered target (ribosome)	*N. gonorrhoeae*, streptococci
	Enhanced efflux	*E. coli, S. pneumoniae*
	Reduced permeability	Enterobacteriaceae
	Drug inactivation	*B. fragilis*
Rifampin	Altered target (β-subunit of polymerase)	*E. coli, S. aureus, Mycobacterium tuberculosis*
Sulfonamides, trimethoprim	Altered target (dihydropteroate synthetase or dihydrofolate reductase)	Enterobacteriaceae, *M. catarrhalis*
	Enhanced *p*-aminobenzoic acid production	*S. aureus, N. gonorrhoeae*
	Reduced permeability	*P. aeruginosa,* Enterobacteriaceae
Vancomycin	Altered target (peptidoglycan precursor binding site)	*E. faecium, E. faecalis, S. aureus*

ESBLs = extended-spectrum β-lactamases.

sulbactam, tazobactam, and, most recently, NXL-104 (an experimental drug now referred to as avibactam). These agents inhibit different types of β-lactamases, with NXL-104 being the only one able to inhibit the ampC-type enzymes carried by *P. aeruginosa*, *Enterobacter* species, *Citrobacter* species, *Serratia marcescens*, and indole-positive *Protea* (SPICE organisms).

All β-lactams are relatively non–concentration dependent in their kill rate, and free drug time greater than MIC is the important factor for increased organism killing. The classes differ somewhat, with carbapenems requiring approximately 40% free drug time greater than MIC for near-maximal cell killing. For penicillins, this percentage is approximately 50%, and for cephalosporins and monobactams, it is 60 to 70%. When these agents need to penetrate to a site of infection, such as epithelial lining fluid or central nervous system, the targets may change somewhat.

Aminoglycosides

These important agents were discovered in the late 1940s for the treatment of *M. tuberculosis* (streptomycin). Screening of natural products identified a number of different aminoglycosides, such as kanamycin, neomycin, gentamicin (actually a combination of three congeners), and tobramycin. Other semisynthetic agents, such as amikacin, netilmicin, and arbekacin (among others), have been discovered and used for therapy in the United States and elsewhere.

Nephrotoxicity and middle ear toxicity (hearing loss or loss of balance) are the defining dose-limiting toxicities of aminoglycosides and resulted in their going out of favor in the 1990s and early in the first decade of the 21st century. It has now been recognized that most (but not all) of the nephrotoxic potential can be ameliorated by intermittent dosing of these drugs (usually once daily). Even with daily therapy, however, prolonged use can still result in nephrotoxicity or ototoxicity.

The recent rise in resistance, particularly among gram-negative isolates, and new resistance mechanisms mediating resistance to even our best β-lactam agents have resulted in renewed interest in existing aminoglycosides and a search for new ones that are more resistant to inactivation by the aminoglycoside-modifying enzymes.[6]

These drugs are concentration dependent in terms of their kill rate and are rapid killers. Hence, the AUC/MIC ratio (or, as sometimes reported, peak concentration/MIC ratio) is the factor most closely associated with bacterial cell killing. Therefore, large doses administered daily to patients with normal renal function would be expected to optimize bacterial cell killing while minimizing the probability of inducing an aminoglycoside-related toxic event.

The older aminoglycosides (streptomycin and gentamicin) have the best profile for synergizing with drugs active against gram-positive streptococci (particularly enterococci). Tobramycin generally has the most potent activity against *P. aeruginosa* (depending on the aminoglycoside-modifying enzymes present in a specific locale), and less resistance is generally seen with amikacin (again, depending on the locale). Streptomycin and amikacin, because of the number of amino groups carried, are usually about four-fold less potent

TABLE 287-4 DOSAGE REGIMENS OF ANTIBACTERIALS, PHARMACOKINETICS, AND DOSE ADJUSTMENT IN PATIENTS WITH RENAL OR HEPATIC FAILURE

CLASS/AGENT	DOSE* FOR SYSTEMIC INFECTION	ORAL FORMULATION	PEAK SERUM CONCENTRATION (μg/mL)	PROTEIN BINDING (%)	NORMAL SERUM HALF-LIFE (hr)	HEPATIC FAILURE	RENAL FAILURE	SERUM LEVELS AFFECTED BY DIALYSIS
AMINOGLYCOSIDES								
Amikacin	5-6.7 mg/kg q8h or 15-20 mg/kg q24h	—	35	0	2-3	No	Major	Yes (H, P)
Gentamicin	1.7 mg/kg q8h or 5 mg/kg q24h	—	7	0	2-3	No	Major	Yes (H, P)
Netilmicin	1.7 mg/kg q8h or 5 mg/kg q24h	—	7	0	2-3	No	Major	Yes (H, P)
Tobramycin	1.7 mg/kg q8h or 5 mg/kg q24h	—	7	0	2-3	No	Major	Yes (H, P)
ANTITUBERCULOUS AGENTS								
Ethambutol	15 mg/kg q24h (PO)	Yes	2	10	3.3	No	Major	Yes (H, P)
Isoniazid	5 mg/kg q24h or 300 mg q24h (PO)	Yes	4.5	10	3	Yes	Major	Yes (H, P)
Pyrazinamide	10 mg/kg q8h (PO)	Yes	12	10	10	Yes	Yes	Yes (H)
Rifampin	10 mg/kg or 600 mg q24h (PO)	Yes	7	81-89	3	Yes	Minor	No (H)
CARBAPENEMS								
Doripenem	0.5-1.0 g q8h	—	23	<10	1	No	Yes	Yes (H)
Ertapenem	1 g q24h	—	155	95	4-5	Unknown	Yes	Yes (H)
Imipenem	0.5-1 g q6-8h	—	40	15	1	No	Avoid in severe renal dysfunction	Yes (H)
Meropenem	0.5-2 g q8h	—	50	<10	1	No	Yes	Yes (H)
FIRST-GENERATION CEPHALOSPORINS								
Cefadroxil	1000 mg q12h (PO)	Yes	16	20	1.5	No	Yes	Yes (H)
Cefazolin	0.5-2 g q8h	—	180	80	2	No	Major	Yes (H) No (P)
Cephalexin	250-500 mg q6h (PO)	Yes	18	15	1	No	Yes	Yes (H, P)
Cephradine	500-1000 mg q6-12h	Yes†	140	10	103	No	Yes	Yes (H, P)
SECOND-GENERATION CEPHALOSPORINS								
Cefaclor	250-500 mg q8h (PO)	Yes†	10	25	0.8	No	Yes	Yes (H)
Cefoxitin	1-2 g q6-8h	—	220	70	0.8	No	Yes	Yes (H) No (P)
Cefprozil	250-500 mg q12h (PO)	Yes	10	35	1.4	No	Yes	Yes (H)
Cefuroxime	750-1500 mg q8h	—	100	50	1.5	No	Yes	Yes (H, P)
Cefuroxime axetil	250-500 mg q12h (PO)	Yes	9	50	1.5	No	Yes	Yes (H, P)

TABLE 287-4 DOSAGE REGIMENS OF ANTIBACTERIALS, PHARMACOKINETICS, AND DOSE ADJUSTMENT IN PATIENTS WITH RENAL OR HEPATIC FAILURE—cont'd

CLASS/AGENT	DOSE* FOR SYSTEMIC INFECTION	ORAL FORMULATION	PEAK SERUM CONCENTRATION (μg/mL)	PROTEIN BINDING (%)	NORMAL SERUM HALF-LIFE (hr)	HEPATIC FAILURE	RENAL FAILURE	SERUM LEVELS AFFECTED BY DIALYSIS
THIRD-GENERATION CEPHALOSPORINS								
Cefdinir	300 mg q12h (PO)	Yes	2	65	1.7	Unknown	Minor	Yes (H)
Cefditoren pivoxil	400 mg q12h (PO)	Yes	4	88	1.6	No	Yes	Yes (H)
Cefixime	400 mg q24h (PO)	Yes	3-5	67	3	No	Yes	No (H, P)
Cefotaxime	1-2 g q6-8h	—	200	50	1.5	Minor	Minor	Yes (H) No (P)
Cefpodoxime proxetil	200-400 mg q12h (PO)	Yes	3	25	2.5	No	Yes	Yes (H)
Ceftazidime	1-2 g q8h	—	160	60	2	No	Major	Yes (H, P)
Ceftibuten	400 mg q24h (PO)	Yes	15	65	2.5	Unknown	Yes	Yes (H)
Ceftizoxime	1-2 g q6-8h	—	130	30	1.3	No	Major	Yes (H) No (P)
Ceftriaxone	1-2 g q12-24h	—	250	90-95	8	No	No	No (H)
FOURTH-GENERATION CEPHALOSPORINS								
Cefepime	1-2 g q8h	—	193	20	2	No	Major	Yes (H, P)
Ceftaroline	600 mg q8-12h	—	21.3	20	2. 6	No	Major	Yes (H)
PENICILLINS								
Amoxicillin	500 mg q8h (PO)	Yes	10	20	1	No	Yes	Yes (H) No (P)
Amoxicillin–clavulanic acid[‡]	875/125 mg q8-12h	Yes	2.7	25	1.3	Unknown	Moderate	Yes (H, P)
Ampicillin	1 g q6h	Yes[†]	200	20	1	No	Yes	Yes (H) No (P)
Cloxacillin	500 mg q6h (PO)	Yes[†]	9	95	0.5	No	No	No (H, P)
Dicloxacillin	500 mg q6h (PO)	Yes[†]	18	97	0.5	No	No	No (H, P)
Nafcillin	1-2 g q4-6h	—	160	90	0.5	Yes	No	No (H, P)
Oxacillin	1-2 g q4-6h	—	200	90	0.5	Yes	No	No (H, P)
Penicillin G	3-4 million units q4-6h	Yes[†]	60	60	0.5	No	Yes	Yes (H) No (P)
Penicillin V	500 mg q6h (PO)	Yes	5	80	1	No	No	Yes (H) No (P)
Piperacillin/tazobactam	3.375-4.5 g q6-8h	—	240	50	1	Minor	Minor	Yes (H)
Ticarcillin/clavulanate	3.1 g q4-8h		220	50	1	Minor	Major	Yes (H, P)
MONOBACTAMS								
Aztreonam	1-2 g q8h	—	250	60	2	No	Major	Yes (H, P)
QUINOLONES								
Ciprofloxacin	400 mg q8-q12h 500-750 mg q12h (PO)	Yes[†]	2-3	30	4	No	Minor	No (H, P)
Levofloxacin	250-750 mg q24h (IV or PO)	Yes	6-9	30	7	No	Yes	No (H, P)
Moxifloxacin	400 mg q24h (IV or PO)	Yes	4-5	50	10	No–minor	No	No (H, P)

TETRACYCLINES, GLYCYLCYCLINES								
Doxycycline	100 mg q12-24h (PO) after 200-mg loading dose	Yes	1.5-2.1	93	15-20	Avoid	No	No (H, P)
Minocycline	100 mg q12-24h (PO) after 200-mg loading dose	Yes	2.2	75	15	No	Avoid	No (H, P)
Tetracycline	500 mg q6h (PO)	Yes†	4	50	7	Avoid	Avoid	No (H, P)
Tigecycline	100 mg, then 50 mg q12h	—	0.6-0.9	70	37-38	Minor	No	No (H, P)
SULFONAMIDES								
Sulfadiazine	15 mg/kg q6h	Yes	30	50	3	Avoid	Avoid	Unknown
Sulfamethoxazole	0.5-1 g q6-8h (PO)	Yes	100	50	9	Avoid	Major	Yes (H) No (P)
Trimethoprim (with sulfamethoxazole)	3-5 mg/kg q6-8h (based on trimethoprim component)	Yes	3-9	60	10	No	Avoid	Yes (H) No (P)
MACROLIDES, LINCOSAMIDES, KETOLIDES								
Azithromycin	500 mg first dose, followed by 250 mg q24h or 500 mg ×3 days (PO) Single-dose therapy of 1-2 g for STIs	Yes†	0.4	25	12-50	Unknown	No	No (H, P)
Clarithromycin	500 mg q12h (PO)	Yes	2-3	70	7	No	Minor	Yes (H) No (P)
Clindamycin	0.3-0.9 g q8h	Yes	15	90	2.5	Minor	No	No (H, P)
Erythromycin	500 mg q6h (PO)	Yes†	1.8	70	2	Minor	No	No (H, P)
Telithromycin	800 mg q24h (PO)	Yes	2	65	10	No	No	No (H)
OTHER AGENTS								
Chloramphenicol	0.25-1 g q6h Oral administration produces higher blood concentrations than IV administration.	Yes	8-14	60	1.5	Minor	No	Yes (H) No (P)
Daptomycin	4-6 mg/kg q24h	—	58-100	90	8-9	No	Minor	No (P) Minor (H)
Linezolid	600 mg q12h	Yes	18	30	5	No–minor	No	Yes (H)
Metronidazole	500 mg q6h (anaerobes) 250 mg q8h (trichomoniasis) 750 mg q8h (amebiasis) (IV or PO)	Yes	25	20	8	Yes	No	Yes (H) No (P)
Nitrofurantoin	100 mg q6h (PO)	Yes	Nil	60	0.3	No	Avoid	Yes (H)
Quinupristin-dalfopristin (30:70)	7.5 mg/kg q8-12h	—	3.2/8§	90/30	3/1§	Minor	No	No (P)
Spectinomycin	2 g/24hr	—	100	0	2	No	Avoid	Unknown
Vancomycin	15 mg/kg q12h	Yes¶	35	50	6	No	Major	No (H, P)

*Dose in milligrams per kilogram body weight at hour intervals and/or oral dose in milligrams in patients with normal renal function; all doses are parenteral unless specified PO.
†Significant decrease or delay in absorption when administered with food.
‡Refers to information for clavulanic acid.
§Includes parent compound and active metabolites.
¶Oral vancomycin is not absorbed; it is used for intraluminal therapy only.
H = hemodialysis; P = peritoneal dialysis; STI = sexually transmitted infection.

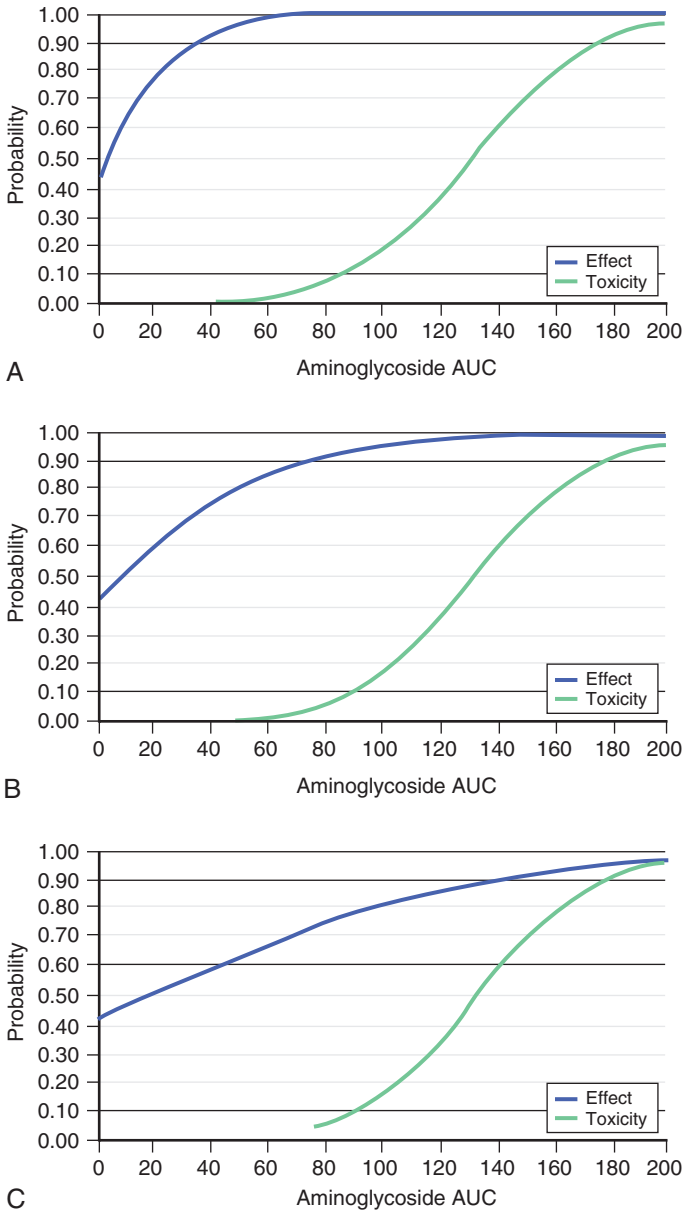

FIGURE 287-2. Probability of clinical effect versus nephrotoxicity as a function of aminoglycoside area under the curve (AUC). A, Minimal inhibitory concentration (MIC) = 0.25 mg/L. **B,** MIC = 0.5 mg/L. **C,** MIC = 1.0 mg/L.

than gentamicin or tobramycin. Consequently, their dosages are approximately three- to four-fold higher.

These agents should be thought of as components of combination regimens for seriously ill patients, particularly those thought to be infected by gram-negative organisms, in the empirical therapy setting.

Quinolones

These agents are completely synthetic and do not exist in nature. They are inhibitors (depending on the drug and the organism) of topoisomerases II and IV. This means that they strike at the heart of DNA replication, making them rapidly bactericidal. The use of a fluorine substitution markedly enhanced the microbiologic activity of these drugs, so they became useful for both community-acquired infections (especially urinary tract infections and pneumonia) and hospital-acquired gram-negative infections.

These drugs penetrate well into most spaces, with substantial concentrations inside cells. This makes them active against obligate intracellular pathogens such as *Chlamydophila*, *Legionella*, and *Mycoplasma*. They also penetrate well into spaces with tight junctions (prostate, eye, central nervous system) and into epithelial lining fluid.

Typically, the toxicities associated with these drugs are "off target." A number of these drugs were either withdrawn from the marketplace or given black box warnings by health authorities because of the infrequent occurrence of torsades de pointes (Chapter 65) or other serious and life-threatening toxicities, such as eosinophilic hepatitis.

These agents are concentration-dependent bacterial killers, indicating that the free drug AUC/MIC ratio is best associated with a regimen's ability to kill bacterial cells. Of interest, particularly for a drug that is completely synthetic, there are a number of resistance mechanisms that allow bacterial escape from drug pressure. The combination of efflux pump overexpression and error-prone replication, resulting in target site mutants as well as serious underdosing for some of the earlier drugs in the class, has resulted in the emergence of considerable resistance, particularly among gram-negative isolates and in the ICU setting.[7]

Macrolides, Tetracyclines, Ketolides, Clindamycin, and Linezolid

These agents bind to different places in the bacterial ribosome, making them inhibitors of protein synthesis and, for the most part, drugs with limited bactericidal activity. An exception may be their activity against *S. pneumoniae*.

Macrolides (particularly clarithromycin and azithromycin) are most useful for community-acquired respiratory tract infections, for two reasons. First, their spectrum is well suited to the classic and atypical pathogens encountered in these patients. Second, they concentrate well in the epithelial lining fluid, with an accumulation that ranges from 6-fold to almost 20-fold that in plasma. This accumulation also partially explains why these agents have fared much better in the respiratory tract setting than in the skin and skin structure setting. Telithromycin, a ketolide, retains activity against many (but not all) macrolide-resistant isolates.

The macrolides and telithromycin are free drug AUC/MIC ratio–driven drugs with respect to bacterial activity, but for different reasons. Telithromycin is somewhat concentration dependent with regard to its microbiologic activity, whereas classic macrolides are not. Telithromycin is more like aminoglycosides or quinolones in terms of the association between exposure indices and bacterial effect. Classic macrolides, especially azithromycin, induce a long persistent or post-antibiotic effect and are *not* concentration dependent in their bacterial effect. This post-antibiotic effect suppresses bacterial regrowth after the drug concentration declines below the MIC until the next dose of drug is administered. In this way, we perceive the linkage to be driven by the free drug AUC/MIC ratio, but with a mechanism different from that of agents whose bacterial kill rate is concentration dependent.

Most of the toxicities seen here are off target and gastrointestinal. However, there is also some elongation of the QT interval (Chapter 54).

Tetracyclines have an exceptionally broad spectrum of activity, including both gram-positive and gram-negative pathogens, and are active against atypical pathogens. The latest incarnation of a tetracycline-type agent (tigecycline, a glycylcycline) has good activity against methicillin-resistant *S. aureus* (MRSA) as well. Tigecycline differs from previous tetracyclines (e.g., doxycycline, minocycline) in that its structure prevents many efflux pumps from removing the drug from the bacteria and also provides a degree of ribosomal protection.

The lincosamide antibiotic clindamycin has a spectrum that covers most clinically significant anaerobes, including many in the *Bacteroides* group (although some resistance is being seen). It is thus a useful agent for infections in which anaerobes play a prominent role, such as lung abscesses and intra-abdominal infections. In many areas of the country, it retains activity against MRSA, but this should be verified with a D-test (inducible macrolide-lincosamide-streptogramin resistance) in the microbiology laboratory. It also has good activity against many streptococci. Because it is a protein synthesis inhibitor, it may provide improved results when the staphylococcal or streptococcal isolate elaborates a great deal of toxin.

Again, most toxicity is off target, with gastrointestinal symptoms being most frequent. Antibiotic-associated diarrhea and the more serious *Clostridium difficile* colitis can occur.

Linezolid is a member of the oxazolidinone class of antibiotics and is a protein synthesis inhibitor. It is distinguished by having excellent gram-positive activity and robust activity against MRSA. It has been used in skin and skin structure infections and has been evaluated in MRSA nosocomial pneumonia. Linezolid penetrates well into skin as well as into the epithelial lining fluid, which is the reason for its evaluation in hospital-acquired MRSA against vancomycin. It is also distinguished by having an oral formulation that is highly bioavailable.

The free drug AUC/MIC ratio is the pharmacodynamic index most associated with cell killing for linezolid. Its toxicity is reflected by dropping counts of bone marrow–derived cells, particularly platelets, seen most frequently

TABLE 287-5 DIVERSE EFFECTS OF ANTIMICROBIAL AGENTS

AGENT	GENERAL	SKIN	GI TRACT	BLOOD CELLS	KIDNEY	NERVOUS SYSTEM	OTHER
Penicillins	Hypersensitivity, anaphylaxis, serum sickness	Rash, urticaria, erythema multiforme	Diarrhea (ampicillin, amoxicillin-clavulanate), hepatitis (oxacillin)	Coombs-positive hemolytic anemia, impaired platelet function (ticarcillin), leukopenia, thrombocytopenia	Nephritis (methicillin), hypokalemia (carboxy- and ureido-penicillins)	Seizures, twitching (high doses, renal failure)	Inactivates aminoglycosides when admixed; possible with concurrent therapy in renal failure
Cephalosporins	Serum sickness (cefaclor), hypersensitivity, anaphylaxis (rare)	Rash, urticaria	Diarrhea, hepatic dysfunction, precipitates in bile (ceftriaxone), mild increase in LFTs	Neutropenia, increased prothrombin time, bleeding (due to MTT side chain), positive Coombs test	Enhances aminoglycoside toxicity, acute renal failure (rare), interstitial nephritis		Disulfiram-like reaction with alcohol use (MTT side chain)
Carbapenems	Hypersensitivity	Rash, urticaria, erythema multiforme	Vomiting with rapid infusion (imipenem), abnormal LFTs	Bone marrow suppression, positive Coombs test	Renal dysfunction	Seizures, myoclonus	
Aminoglycosides	Fever	Rash			Reversible renal failure	Irreversible vestibular toxicity and/or auditory damage, muscle blockade (with anesthetics and myasthenia gravis)	
Vancomycin	Allergy, fever	Rash		Leukopenia, thrombocytopenia	Nephrotoxic	Decreased hearing, neuropathy	Histamine release with flushing and hypotension (infusion <1 hr, antihistamines can prevent)
Quinolones	Headache, allergy, anaphylaxis (rare)	Rash (gemifloxacin), urticaria, photosensitivity (lomefloxacin)	GI distress, abnormal LFTs			Dizziness, insomnia, nervousness, tremors, visual changes, seizures	Tendon rupture arthropathy in young animals
Sulfonamides	Hypersensitivity, anaphylaxis, serum sickness, fever	Rash, Stevens-Johnson syndrome, photosensitivity	Hepatitis	Hemolysis (G6PD deficiency), agranulocytosis, marrow suppression	Crystalluria	Neuropathy	Vasculitis
Trimethoprim ± sulfamethoxazole	Fever	Rash, erythema multiforme, Stevens-Johnson syndrome, TEN	Hepatitis, pancreatitis	Marrow suppression	Hyperkalemia, acute renal failure		
Chloramphenicol	Fever			Marrow suppression (dose related), aplastic anemia		Optic neuritis, neuropathy	Circulatory collapse (gray baby syndrome in neonates)
Tetracyclines	Hypersensitivity	Photosensitization (doxycycline)	GI discomfort, hepatotoxicity in azotemia or pregnancy		Antianabolic aggravation of azotemia (except doxycycline)	Vertigo (minocycline)	Deposition in bone (dysplasia) and teeth (staining)
Macrolides	Fever	Rash	GI discomfort			Reversible decreased hearing	Phlebitis (IV erythromycin), metallic taste (clarithromycin)
Clindamycin	Fever	Rash	Diarrhea, pseudomembranous colitis				
Metronidazole	Headache, hypersensitivity		Nausea, metallic taste, pancreatitis	Leukopenia		Peripheral neuropathy, ataxia	Mutagenic, carcinogenic in rodents, disulfiram-like reaction with alcohol

GI = gastrointestinal; G6PD = glucose-6-phosphate dehydrogenase; LFT = liver function test; MTT = methylthiotetrazole; TEN = toxic epidermal necrolysis.

with prolonged use. Resistance occurs rarely and is seen with enterococci more frequently than with *S. aureus*. There are five or six gene copies of the target site. As more and more copies become mutated through homologous recombination, the MIC increases almost linearly. Because these mutated targets have some biofitness cost, there is a slow reversion to wild-type targets when therapy is discontinued. If all copies become mutated, however, they cannot revert to wild type. The early detection of MIC changes should lead to the consideration of changing or stopping therapy.

Vancomycin, Quinupristin-Dalfopristin, and Daptomycin

Although structured differently, these agents (along with linezolid) are distinguished by having reliable activity against MRSA.

Vancomycin acts at the cell wall and is bactericidal. The free drug AUC/MIC ratio is the driver of antimicrobial effect. Vancomycin has activity against many gram-positive pathogens and has lately become distinguished for its activity against *C. difficile* when it is administered orally. As with most drugs, wide use has caused the emergence of vancomycin resistance in both enterococci and *S. aureus* (although the latter is rare).

Although vancomycin has a reputation for being a somewhat slow killer, no new agent has significantly outperformed it, at least in trials of skin and skin structure infections. MIC creep has been documented. Although many *S. aureus* isolates were once susceptible with MIC values of 0.25 and 0.5 mg/L, these are now in the distinct minority; 1.0 mg/L represents the modal value, with a few isolates having values of 2.0 mg/L. The higher MICs make these isolates more difficult to treat reliably with standard dosing (1 g IV every 12 hours). Also, it has recently become clear that higher vancomycin doses are associated with a significantly higher risk of nephrotoxicity, even when the agent is administered alone. In addition to nephrotoxicity, vancomycin causes histamine release ("red man syndrome") when it is administered too quickly intravenously. Infusion times of approximately 1 hour are recommended for standard doses.

Quinupristin-dalfopristin is a combination of streptogramin A and streptogramin B antibiotics. Although each is bacteriostatic, the combination can produce some cell killing because the two drugs are synergistic. The synergy is lost, however, with the emergence of resistance to either drug. Further, the amount of bacterial cell killing is modest because of the drugs' rapid half-lives and the imposition of a 12-hour dosing interval. This combination product has activity against resistant enterococci as well as MRSA. Its toxicity includes relatively severe muscle pain in some patients. It is also phlebitogenic and has to be administered with a central venous catheter, limiting its utility.

Daptomycin is a lipopeptide antibiotic discovered in the 1980s and resurrected in the 1990s with a greater understanding of the relationship between exposure and effect versus exposure and toxicity. It was discovered that the antibacterial effect against MRSA was driven by the free drug AUC/MIC ratio. Muscle effects were seen early in its development that limited therapy. Trough concentrations of daptomycin drive this toxicity and can be ameliorated by once-daily dosing (thus minimizing trough concentrations).

This agent has been licensed for skin and skin structure infections and, of note, for complicated *S. aureus* bacteremia and right-sided endocarditis. It also has activity against enterococci. The drug is bound up in epithelial lining fluid by surfactant, excluding its use in the therapy for pneumonia.

Daptomycin may cause muscle toxicity, which is rare, but this may be preceded by the harbinger of an elevated creatine kinase level. The muscle damage is driven by daptomycin trough concentrations, generally above a concentration of about 25 mg/L. Monitoring of creatine kinase concentration is useful in the management of patients receiving daptomycin therapy.

TOXICITIES

All antimicrobials have toxicities. The most common observed toxicities for many antimicrobial agents are presented in Table 287-5.

DURATION OF THERAPY

Relatively little is known about the optimal duration of therapy. Some work has been done to define certain circumstances in which short courses of chemotherapy are effective. For instance, gonorrhea is highly likely to be cured by a single dose of drug (ceftriaxone, cefixime, fluoroquinolones), as long as the organism is susceptible to the drug given.

For trimethoprim-sulfamethoxazole and fluoroquinolones, controlled trials have shown that 3 days of therapy is adequate for uncomplicated urinary tract infections. For β-lactam antimicrobials, a somewhat longer duration is required.

In community-acquired pneumonia, controlled trials with fluoroquinolones have demonstrated optimal results with 5 days of therapy, most likely because of concentration-dependent killing. For bacterial sinusitis, direct sampling from the infected sinus has demonstrated that bacterial pathogens were eradicated by day 3 of therapy or earlier, particularly for *S. pneumoniae* infections.

In ventilator-associated pneumonia, a double-blind comparison of 8 versus 15 days of therapy demonstrated that, with a single exception, clinical outcomes were just as good, there was less emergence of resistance, fewer antibiotics were administered, and there was less toxicity in the group receiving 8 days of therapy. The single exception was when a nonfermenting gram-negative rod was recovered (*P. aeruginosa* or *Acinetobacter* species). In this circumstance, there were significantly more relapses with 8-day versus 15-day treatment, but clinical outcomes were no different.

In some infections, the organisms are slow growing and require more time for control. Both endocarditis and osteomyelitis are examples. Therapy durations of 4 to 6 weeks and occasionally longer may be required for cure in this circumstance.

Finally, one of the longest durations of therapy is seen in the treatment of tuberculosis. In this case, some organisms are in the "non-replicative persister" (NRP) state, indicating that they are not growing or metabolically active. When recovered, they are still fully susceptible to the antimicrobials being used, but while in the NRP state, they cannot be readily killed with current chemotherapy (phenotypic, not genotypic resistance). Therapy durations of 6 months for wild-type organisms and 18 to 24 months for multidrug-resistant tuberculosis are required because of the need to use suboptimal regimens. Indeed, it is a major research goal to find drugs that will readily kill organisms in the NRP state and thus substantially shorten the therapeutic course for *M. tuberculosis*.

FAILURE OF ANTIMICROBIAL THERAPY

Antimicrobial therapy occasionally fails, with failure defined as the persistence of signs and symptoms of infection or the persistence of fever. When one is confronted with failure after what was thought to be adequate antimicrobial therapy, it should set off a sequence of investigations: Did resistance emerge, or did superinfection occur? Is there an obstructed hollow viscus (e.g., urinary or gastrointestinal tract)? Is there an undrained abscess or infected collection (e.g., empyema)? Is there an infected foreign body (e.g., intravenous catheter or prosthesis), or is there devitalized tissue (e.g., sequestrum in osteomyelitis)? Is the fever (if present) attributable to a drug being administered? Thoughtful re-examination should sort out the cause in many instances.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Cook PP, Gooch M. Long-term effects of an antimicrobial stewardship programme at a tertiary-care teaching hospital. *Int J Antimicrob Agents*. 2015;45:262-267.
2. Tamma PD, Cosgrove SE. Antimicrobial stewardship. *Infect Dis Clin North Am*. 2011;25:245-260.
3. Schwameis R, Zeitlinger M. Methods to measure target site penetration of antibiotics in critically ill patients. *Curr Clin Pharmacol*. 2013;8:46-58.
4. Louie A, Boyne MT 2nd, Patel V, et al. Pharmacodynamic evaluation of the activities of six parenteral vancomycin products available in the United States. *Antimicrob Agents Chemother*. 2015; 59:622-632.
5. Gutkind GO, Di Conza J, Power P, et al. β-Lactamase–mediated resistance: a biochemical, epidemiological and genetic overview. *Curr Pharm Des*. 2013;19:164-208.
6. Poulikakos P, Falagas ME. Aminoglycoside therapy in infectious diseases. *Expert Opin Pharmacother*. 2013;14:1585-1597.
7. Redgrave LS, Sutton SB, Webber MA, et al. Fluoroquinolone resistance: mechanisms, impact on bacteria and role in evolutionary success. *Trends Microbiol*. 2014;22:438-445.

REVIEW QUESTIONS

1. For β-lactam antibiotics, the index that is most closely linked to the ability of the drug dose and schedule to kill the target pathogen is

A. Free drug Cmax/MIC ratio
B. Free drug AUC/MIC ratio
C. Time (fraction of a dosing interval) that free drug concentrations exceed the MIC
D. Total drug AUC/MIC ratio

Answer: C β-Lactam drugs are not particularly concentration dependent in their rate of kill and reach a maximal kill rate at low multiples of the MIC (usually around four-fold). Except in unusual circumstances, there is little in the way of persistent microbiologic effect after drug concentrations at the infection site decline below the MIC. Consequently, the ability to kill target pathogens is increased as time of free drug exceeding the MIC increases. Again, in the majority of instances, only free (non–protein bound drug) is microbiologically active.

2. In choosing an optimal drug dose and schedule for a patient's serious infection, which of the following is the most important consideration?

A. MIC of the pathogen
B. Drug dose chosen
C. Protein binding of the antibiotic chosen
D. Site of infection
E. Toxicity profile of the drug chosen
F. All of the above

Answer: F There are a small number of factors that clinicians can control that have an impact on the ability of a chosen drug dose and schedule to help a patient recover from an infection. Knowledge of the MIC of the infecting pathogen is critical as irrespective of the drug chosen, a lower MIC value will give a higher probability of attaining the target for a good therapeutic outcome. Likewise, the drug chosen has properties that may help or hinder in a specific case. For example, an agent that is not bactericidal is not likely to be an optimal choice in a patient with meningitis. As part of this, only free drug is (in the main) microbiologically active, so this is a serious consideration. The dose size and frequency will have a direct impact on the likelihood of achieving an exposure target that will optimize the likelihood of a good outcome. Understanding the probable infection site is important as penetration will be different into skin, cerebrospinal fluid, epithelial lining fluid, and prostate. Finally, the toxicity profile is important. Choosing a nephrotoxic agent for an intensive care unit patient would be improvident, unless there was little alternative. Consequently, all of the factors need to be considered by the clinician.

3. Which of the following is most important for resistance emergence?

A. Mobile genetic element
B. Efflux pump upregulation
C. Drug-inactivating enzymes (like β-lactamases)
D. None of the above
E. All of the above

Answer: E Each of these represents a way for the infecting pathogen to increase survivorship in the face of antimicrobial therapy. In A and C, there is most often a destruction of the drug. In B, the drug that does penetrate into the organism is actively pumped out, lowering the drug concentration below the critical level necessary for organism kill. These mechanisms also may interact with each other.

4. Of the drugs listed, which are nephrotoxic?

A. Linezolid
B. Vancomycin
C. Meropenem
D. Gentamicin
E. B and D
F. All of the above

Answer: E The nephrotoxic potential of aminoglycosides (e.g., gentamicin, tobramycin, amikacin) has been known for decades. Only more recently have we come to understand that higher doses of vancomycin (>2 g/day) are associated with a substantial risk of nephrotoxicity. (Lodise TP, Lomaestro B, Graves J, Drusano GL. Larger vancomycin doses are associated with an increased incidence of nephrotoxicity. *Antimicrob Agents Chemother*. 2008;52:1330-1336; and Lodise TP, Patel N, Lomaestro B, et al. Relationship between initial vancomycin concentration-time profile and nephrotoxicity among hospitalized patients. *Clin Infect Dis*. 2009;49:507-514.

5. Which of the following patients is likely to have the highest antimicrobial drug clearance (and hence lowest levels of the antibiotic)?

A. A 70-year-old woman with hospital-acquired pneumonia caused by an organism carrying a carbapenem-resistant Enterobacteriaceae
B. A 20-year-old patient who has crashed his Kawasaki Ninja into a tree, causing head injury, and who has developed a ventilator-associated pneumonia caused by *Pseudomonas aeruginosa* and is septic
C. A 50-year-old diabetic man with an infected foot ulcer
D. A 30-year-old sexually active woman with an uncomplicated urinary tract infection
E. A 40-year-old man with a pneumococcal superinfection after having developed an influenza virus infection (he did not get the vaccine)

Answer: B Most people think that seriously ill intensive care unit patients will, of necessity, have lower drug clearances because of illness. However, in this case, the young age of the patient, coupled with the sepsis, which will give him a hyperdynamic state, puts him at highest likelihood for having a very high drug clearance. Whereas many ascribe this to change in glomerular filtration rate (e.g., such a patient may have an estimated glomerular filtration rate in excess of 150 mL/minute), the hyperdynamic state also increases blood flow to other clearing organs, such as the liver, so even agents like a glycylcycline (e.g., tigecycline), which are not heavily renally cleared, can have very high clearances. Such patients have been shown to have a high rate of failure in clinical trials. (Ambrose PG, Bhavnani SM, Ellis-Grosse EJ, Drusano GL. Pharmacokinetic-pharmacodynamic considerations in the design of hospital-acquired or ventilator-associated bacterial pneumonia studies: look before you leap! *Clin Infect Dis*. 2010;51[Suppl]:S103-S110.)

288

STAPHYLOCOCCAL INFECTIONS

HENRY F. CHAMBERS

DEFINITION

Staphylococci are well adapted as both commensals and pathogens. Coagulase-negative species constitute a significant proportion of the normal human cutaneous microbiome. *Staphylococcus aureus,* a coagulase-positive species, is present in nasopharyngeal flora in a third of individuals, most of whom will not become infected. As a pathogen, *S. aureus* is one of the most common causes of bacterial infections, which range in severity from relatively trivial skin infections to lethal invasive disease. Coagulase-negative species are intrinsically less virulent and less invasive than *S. aureus* but are responsible for one in four health care–associated infections, frequently involving an indwelling medical device. Prevalence of antibiotic-resistant strains of staphylococci has a profound impact on therapy.

The genus *Staphylococcus* consists of more than 30 distinct species. These organisms have coevolved as normal flora of mammals and birds. They are gram-positive spherical cells (i.e., cocci) 0.5 to 1.5 µm in diameter that divide in multiple planes to form clusters resembling grapes (*staphylo-* is derived from the Greek word for "bunch of grapes") when viewed under the microscope. Staphylococci are nonmotile, nonsporulating, hardy organisms that are resistant to desiccation, extremes of pH, and high salt concentrations and are capable of growth under aerobic or anaerobic conditions. Staphylococci produce catalase, an enzyme that degrades hydrogen peroxide into water and oxygen, which definitively distinguishes them biochemically from streptococci and enterococci. The coagulase test is the basis for differentiating *S. aureus* from the numerous other nonpathogenic, coagulase-negative species. Coagulase is a secreted cell surface protein that, in the presence of a prothrombin-like plasma protein, converts fibrinogen to fibrin, forming a clot. *S. aureus* produces a variety of other species-specific surface proteins (e.g., protein A) that differentiate it from other species.

The staphylococcal chromosome is circular. Approximately 75% of the genes constitute a core genome common to all staphylococcal species. The remaining 25% contains species-defining elements and mobile genetic elements acquired by horizontal gene transfer. The *S. aureus* genome is abundant in genes encoding toxins, superantigens, and adhesins, whereas coagulase-negative species contain few adhesin and no toxin or superantigen genes. Genetically, *S. aureus* is sufficiently uniform to classify it as a single species; greater diversity among coagulase-negative species merits their classification as distinct species.

STAPHYLOCOCCUS AUREUS

EPIDEMIOLOGY

S. aureus is maintained in the human population primarily through asymptomatic colonization of the anterior nares, mucous membranes, and other moist areas of the body as well as of the groin, perineum, and perianal area in healthy children and adults. Infants become colonized by strains from their mothers within weeks of birth. Carriage rates are higher in children than in adults. Higher than average *S. aureus* carriage rates are associated with atopic dermatitis, eczema, chronic skin ulcers, and other acute and chronic skin conditions; insulin-dependent diabetes; dialysis; human immunodeficiency virus infection; and recreational injection drug use. Carriers have a several-fold higher risk for development of a subsequent *S. aureus* infection compared with noncarriers. The principal mode of transmission of *S. aureus* is direct contact with an infected individual or an asymptomatic carrier, probably through transient hand carriage. *S. aureus* may also contaminate environmental surfaces, where it can persist for days. The role of environmental contamination in transmission is not well defined, but it may be important if heavily contaminated surfaces or materials are contacted. Droplet and aerosol transmission of *S. aureus* plays little if any role.

S. aureus is responsible for millions of infections in the United States each year, most of which are community-acquired skin infections. Approximately 5 to 10% of *S. aureus* infections are invasive and are often associated with bacteremia. *S. aureus* also causes hundreds of thousands of health care–associated infections each year, 50 to 60% of which are caused by methicillin-resistant *S. aureus* (MRSA). Before the mid-1990s, MRSA strains were almost exclusively hospital or health care associated, but they have now become prevalent in the community. Community-associated MRSA strains are distinct from classic health care–associated MRSA strains in several ways (E-Table 288-1).

PATHOBIOLOGY

Virulence Factors

S. aureus is highly adapted to humans through millions of years of coevolution with hominids. Well above 50 virulence factors, including adhesins, toxins, enzymes, surface-bound proteins, and capsule polysaccharides, may be produced (E-Table 288-2). Genes encoding virulence factors may be located on the chromosome either as part of the core genome or within mobile genetic elements (or their remnants), including bacteriophages, pathogenicity islands, and cassettes, or on plasmids. Alpha-toxin, Panton-Valentine leukocidin, and phenol-soluble modulins, all of which provoke potentially deleterious host inflammatory response and cause host cell lysis, appear to be important virulence factors mediating disease severity, especially in community MRSA strains.[1] Protein A, a B-cell superantigen that promiscuously triggers B-cell proliferation and supraclonal expansion and apoptosis, interferes with host antibody-mediated adaptive immunity.[2]

Virulence factors serve to promote binding to host tissues; to evade, circumvent, or disrupt host immune responses; and to facilitate cell injury and tissue invasion. Variability in both the presence of virulence determinants and their expression among strains allows extreme diversity among clinical isolates and the remarkable adaptability and versatility of *S. aureus* as a pathogen. An extensive network of two component response systems, DNA-binding proteins and regulatory RNAs, controls the expression of virulence (and other) factors in response to environmental conditions. Principal among these is the accessory gene regulator *agr*, a two-component quorum sensing and global gene regulator that controls the expression of numerous surface and secreted proteins. Mutations in *agr* have been associated with reduced susceptibility to vancomycin and loss of virulence.

Biofilm formation occurs in the presence of foreign material, such as vascular catheters or implanted devices. Biofilm is a complex network of extracellular polysaccharides, DNA, and protein in which bacterial cells become embedded, rendering them inaccessible to clearance by host defense mechanisms. Organisms within biofilms tend to be metabolically inactive and tolerant to killing by antimicrobial agents.

Mechanisms of Disease

Pathogenesis of *S. aureus* disease occurs by two mechanisms: tissue invasion, which may be local or systemic; and toxin mediated. The characteristic lesion of tissue invasion is the abscess, a focal collection of pus (liquefied and necrotic host tissue, blood, inflammatory cells, DNA, and cellular debris) and bacterial cells surrounded by an ill-defined layer of edematous and inflamed tissue infiltrated by acute and chronic inflammatory cells.

Host defenses against *S. aureus* infection primarily consist of an intact, normal skin barrier and the innate immune system. Conditions in which these defenses are impaired are associated with increased risk of *S. aureus* infection. Among these conditions are injection drug use, presence of vascular access devices, burns, chronic skin diseases, use of systemic steroids, traumatic wounds, minor skin abrasions or trauma, surgical procedures, insulin-dependent and non–insulin-dependent diabetes, peritoneal dialysis, hemodialysis, subcutaneous and intramuscular injections, acupuncture, prosthetic implants, and congenital or acquired neutrophil disorders (e.g., chronic granulomatous disease, Job's syndrome). If the cutaneous barrier is breached, the next line of defense is the innate immune system. Neutrophils recruited to the site of infection ingest and kill staphylococci. Staphylococci elaborate numerous virulence factors specifically designed to thwart each step of the host response. If large numbers of organisms are present, the host response is overwhelmed, infection is not contained, and dissemination occurs. Endothelial cell injury and invasion can also occur. Intracellular organisms and small colony variants within phagocytes and endothelial cells may play a role in relapse and persistent bacteremia. High tissue burdens of organisms and bacteremia are usually but not always accompanied by fever, tachycardia, and other signs of the systemic inflammatory response syndrome, including frank septic shock.

The three toxin-mediated syndromes, which can occur in the absence of invasive disease, are staphylococcal food poisoning, staphylococcal toxic shock syndrome, and staphylococcal scalded skin syndrome. Staphylococcal food poisoning is caused by the ingestion of a preformed heat-stable enterotoxin. The emetogenic activity of enterotoxin is mediated by the intestinal release of 5-hydroxytryptamine and the stimulation of receptors present on afferent vagal neurons. Toxic shock syndrome is caused by a specific toxin, TSST-1, or other staphylococcal enterotoxins acting as superantigens that bind to major histocompatibility complex class II molecules of antigen-presenting cells and T-cell receptors, stimulating the massive release of cytokines from T cells and resulting in septic shock and death. Staphylococcal scalded skin syndrome and bullous impetigo are caused by either of two exfoliative toxins, A or B. These toxins are serine proteases that specifically cleave desmoglein 1, a desmosomal protein that anchors the overlying superficial epidermis to the stratum granulosum.

CLINICAL MANIFESTATIONS

Skin and Soft Tissue Infections

Skin and soft tissue infections are by far the most common infections caused by *S. aureus*; millions of cases occur annually in the United States (Chapter 441). Community MRSA strains have been associated with a 50% increase in the rates of skin and soft tissue infections in the United States. This heterogeneous group of skin diseases includes impetigo, folliculitis, furuncle, abscess, erysipelas and cellulitis, mastitis (cellulitis of the breast), necrotizing fasciitis, and wound infections.

E-TABLE 288-1 COMPARISON OF COMMUNITY AND HOSPITAL CLONES OF METHICILLIN-RESISTANT *STAPHYLOCOCCUS AUREUS* (MRSA)

PROPERTY	COMMUNITY MRSA	HOSPITAL MRSA
Methicillin-resistance gene cassette	*SCCmecIV*[*]	*SCCmecI, II, III*
Common genotypes[†]	USA300 (ST8) USA400 (ST1)	USA100 (ST5) USA500 (ST8) USA800 (ST5)
Panton-Valentine leukocidin	Present in ≈95%	Present in <5%
Antimicrobial resistances	β-Lactams, macrolides, fluoroquinolones	β-Lactams, macrolides, clindamycin, tetracycline, fluoroquinolones, aminoglycosides
Doubling time[‡]	30 minutes	45-60 minutes
agr Mutants	Uncommon	Common

*Staphylococcal chromosomal cassette type; this element contains the methicillin resistance gene *mecA*.
[†]Pulsed field gel electrophoresis type with the sequence type (ST) in parentheses.
[‡]Time for the number of cells to double in broth culture.

E-TABLE 288-2 EXAMPLES OF VIRULENCE FACTORS PRODUCED BY *STAPHYLOCOCCUS AUREUS*

ACTIVITY	GENES	PROTEIN OR MOLECULE
Inhibition of antimicrobial peptides	*icaA, icaD, icaB, icaC, icaR*	Polysaccharide intercellular adhesin (PIA)
	isdA, isdB	Iron-regulated surface determinants of *S. aureus* (IsdA and IsdB)
	mprF	Multiple peptide resistance factor (MprF)
	Sak	Staphylokinase
Inhibition of chemotaxis	*chp*	Chemotaxis inhibitory protein of *S. aureus* (CHIPS)
	ecb	Extracellular complement-binding protein (Ecb)
	efb	Extracellular fibrinogen-binding protein (Efb; inhibitor of C5a generation)
	scn	Staphylococcal inhibitor of complement (SCIN)
Inhibition of oxygen-mediated bacterial killing	*crtM, crtN*	Carotenoid pigment, staphyloxanthin (*S. aureus* golden pigment; promotes resistance to reactive oxygen species)
	isdA, isdB	Iron-regulated surface determinants of *S. aureus*, IsdA and IsdB (promote resistance to neutrophil reactive oxygen species)
	sodA, sodM	Superoxide dismutase, SodA, SodM IsdB (promotes resistance to neutrophil reactive oxygen species)
Inhibition of neutrophil function	*cap5* or *cap8*	Capsular polysaccharide (inhibits phagocytosis)
	clfA	Clumping factor A (ClfA; inhibits phagocytosis)
	eap	Extracellular adherence protein (Eap; inhibits leukocyte adhesion)
	hlgA, hlgB, hlgC	Gamma hemolysin subunits A, B, and C; HlgA, HlgB, HlgC (causes cell lysis)
	lukS-PV, lukF-PV	Panton-Valentine leukocidin (causes cell lysis)
	psm	Phenol-soluble modulin-like peptides (PSMs; causes cell lysis)
Perturbation of T-cell function	*sea, seb, sec$_n$, sed, see, seg, seh, sei, sej, sek, sel, sep*	Staphylococcal enterotoxins: SEA, SEB, SEC$_n$, SED, SEE, SEG, SEH, SEI, SEJ, SEK, SEL, SEP (superantigens)
	tst	Toxic shock syndrome toxin 1 (TSST-1; superantigen)
Perturbation of B-cell function	*spa*	Protein A (B-cell superantigen)
Adhesion	*cna*	Collagen-binding protein
	fnA, fnB	Fibronectin-binding proteins A and B
	clfA, clfB	Clumping factors A and B (binds platelets and fibrinogen)

Impetigo, folliculitis, and furuncle are superficial infections; fever and other systemic signs of infection are not present. Impetigo is a focal infection of the epidermis that occurs most commonly in children (Fig. 441-1). The typical lesion, which may be multiple or in clusters, is about 1 cm in diameter, with erythema surrounding a bulla or bullae (caused by the production of exfoliative toxin) containing cloudy fluid or with a crusty or scabbed-over appearance. Gram stain of the fluid or drainage from the lesion shows the organism. Folliculitis is a superficial infection with tender, erythematous, maculopapular or pustular lesions centered around hair follicles. Both impetigo and folliculitis readily respond to local measures, such as application of soap and water, topical antibiotics, or antiseptics; systemic antimicrobial therapy may be indicated for extensive or refractory infections.

A furuncle is simply a boil, an erythematous, suppurative infection measuring 1 to 2 cm that extends through the dermis into the subcutaneous tissue (Fig. 441-2). It may drain spontaneously with application of hot compresses or can be surgically drained with simple incision and drainage. Antimicrobial therapy is not needed. The distinction between a furuncle and an abscess is somewhat arbitrary. Abscesses tend to be larger and deeper and may be associated with systemic signs of infection and bacteremia. Furuncles may extend to fascia or deeper tissues and coalesce into carbuncles, a more severe form of infection that may be accompanied by bacteremia. Large abscesses and carbuncles, particularly in the presence of fever and the systemic inflammatory response syndrome, require surgical drainage and systemic antimicrobial therapy.

Erysipelas (Fig. 441-4) and cellulitis, which are similar in appearance, are painful, warm, indurated, erythematous, nonlocalized infections that may be accompanied by lymphangitis. Cellulitis extends into the dermis and subcutaneous fat; erysipelas is more superficial. Surrounding cellulitis is often associated with an obvious cutaneous abscess, but it may also overlie a deeper abscess, in which case the primary therapy is surgical drainage. Cellulitis without an associated abscess should be treated with systemic antimicrobial therapy. Cellulitis due to streptococci (Chapter 290) cannot reliably be distinguished from that caused by *S. aureus* owing to their similar appearance. The presence of an associated purulent lesion suggests staphylococcal infection.

Necrotizing fasciitis is an infection of the deep layers of skin and subcutaneous tissues, extending to muscle and along fascial planes. It is associated with systemic toxicity, leukocytosis, and severe pain often out of proportion to the physical findings. The overlying skin may appear to be uninvolved, belying the serious nature of this infection, which requires immediate surgical intervention for débridement of involved tissue. Necrotizing fasciitis, which is more typically caused by group A streptococci (Chapter 290) or occurs as a mixed infection, has been associated with community MRSA infection.

Pyomyositis (also termed tropical myositis) is an infection of the skeletal muscle; *S. aureus* is the most common cause. The patient presents with fever, pain, and swelling and induration that can be felt on deep palpation. The overlying skin and soft tissue may appear normal. There is often a history of trauma to the infected area. Although it can occur in otherwise normal children and adults, acquired immunodeficiency syndrome and other immunocompromising conditions are predisposing factors. This infection is thought to occur as a consequence of metastatic seeding from a subclinical bacteremia, and blood cultures may not be positive at the time of diagnosis. The diagnosis is established by culture of pus collected by needle aspirate. Surgical or percutaneous drainage should be performed, and systemic antimicrobial therapy is indicated.

S. aureus causes 30% of surgical site infections overall in the United States, and approximately 50% of those follow neurosurgical or orthopedic procedures. These infections occur at the site of incision, typically after the second or third postoperative day. Signs and symptoms are fever accompanied by erythema, edema, induration, drainage, pain, and tenderness at the surgical site. Superficial infections respond to removal of stitches, débridement of devitalized tissue, opening of the wound to allow drainage, and a short course of antimicrobial therapy. Deeper infections may require more extensive débridement and longer courses of therapy (4 to 6 weeks), particularly if bone or a prosthetic device is involved. Removal of infected prosthetic material or a foreign body greatly increases the chance of cure.

Bacteremia

Bacteremia, the presence of bacteria in the blood stream, exemplifies the pathogenicity of *S. aureus*. It is present in approximately 75% of cases of invasive, deep tissue infections. The most common sources of bacteremia are skin and soft tissue infections, central venous catheters and other intravascular devices, bone and joint infections, pneumonia, and endocarditis. Bacteremia can emanate from any source, which may not be obvious in 25% of cases. Invasion of the blood stream allows the organism to disseminate widely throughout the body, establishing multiple metastatic sites of infection and thereby perpetuating bacteremia. Fever is usually but not always present. Sepsis syndrome and septic shock are common, and death occurs in 10 to 20% of cases. Patients with occult *S. aureus* bacteremia have high rates of organ failure, septic shock, and intensive care unit (ICU) admissions, prolonged ICU stays, and high mortality.[3]

The presence of bacteremia dictates the approach to the diagnosis, management, and therapy of *S. aureus* infection. When blood cultures are positive, even if the primary source is known, there is always the possibility of endocarditis or other secondary foci of infection. Echocardiography has been generally recommended in cases of *S. aureus* bacteremia to look for valvular vegetations or other signs of endocarditis, although its role is evolving.[4] If it is available, transesophageal echocardiography is preferable to transthoracic echocardiography because it has better sensitivity. An echocardiogram should be obtained in cases of complicated bacteremia, defined by the presence of any one of the following: positive blood cultures for 3 days or more, presence of an intracardiac device (e.g., pacemaker, prosthetic valve), presence of a secondary or metastatic focus of infection, relapse or recurrence of *S. aureus* bacteremia, or clinically suspected endocarditis.

Source control is the cornerstone of management and therapy. Both primary and secondary foci of infection should be identified and eliminated whenever possible because these may lead to treatment failure or relapse once therapy is discontinued. Computed tomography or magnetic resonance imaging should be considered if signs and symptoms point to deep tissue abscesses or osteomyelitis. Follow-up blood cultures should be obtained to document clearance. Persistent bacteremia is suggestive of endovascular infection, and failure to clear blood cultures after 3 to 4 days of appropriate therapy is a strong predictor of complicated bacteremia, necessitating a longer course of therapy. Antimicrobial therapy should always be administered. A shorter duration of therapy (i.e., 14 days) is appropriate for uncomplicated bacteremia (Table 288-1). Longer courses of 4 to 6 weeks are recommended for the treatment of endocarditis or bacteremia complicated by slow resolution or the presence of metastatic infection.

Endocarditis

S. aureus is the leading cause of both native valve and prosthetic valve endocarditis (Chapter 76), accounting for approximately 30% or more of all cases. Most are community-acquired infections, often occurring as a complication of injection drug use. The number of health care–associated infections has been increasing in recent years.[5] Risk factors are diabetes mellitus, hemodialysis, and presence of a prosthetic valve. The presentation is that of an acute febrile illness with high fever developing during a few days. The patient may appear toxic and septic. The intracardiac source of infection may not be evident at first because a pathologic murmur may not be evident when the patient first presents. A quarter or more of patients have an associated infection of bone, joint, or skin and soft tissue. The aortic and mitral valves are most commonly involved in native valve infection except in injection drug users (discussed later). Systemic embolization to the brain, kidneys, spleen, gut, or other large vessels is clinically evident in about one third of cases. Peripheral manifestations, including Roth's spots, Osler's nodes, Janeway's lesions, and petechiae, occur with a similar frequency. Morbidity and mortality are high, in part due to the occurrence of this infection in older patients,

TABLE 288-1 CRITERIA FOR DIAGNOSIS OF UNCOMPLICATED *STAPHYLOCOCCUS AUREUS* BACTEREMIA*

Resolution of fever and systemic signs of infection by day 3 of therapy
Sterile blood cultures within 2 or 3 days of initiation of antimicrobial therapy
Presence of an identifiable and easily removable focus of infection
Prompt removal of the primary focus of infection
No echocardiographic or clinical signs of endocarditis
No osteomyelitis
No hematogenous secondary foci of infection
No preexisting valve abnormalities predisposing to endocarditis (e.g., prosthetic valve, rheumatic heart disease, bicuspid aortic valve)
No implanted prosthetic device (e.g., prosthetic hip)

*Uncomplicated *S. aureus* bacteremia can be treated with a shorter course of antibiotics (see text).

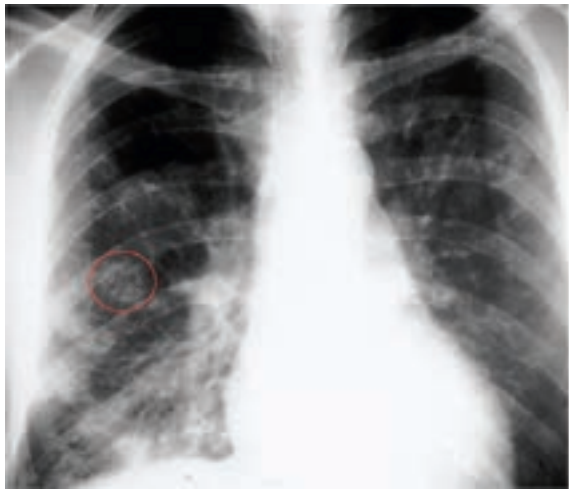

FIGURE 288-1. Chest radiograph shows multiple nodular pulmonary lesions, suggestive of septic embolization, in a patient with tricuspid valve *S. aureus* endocarditis. The *red circle* shows a lesion with signs of cavitation.

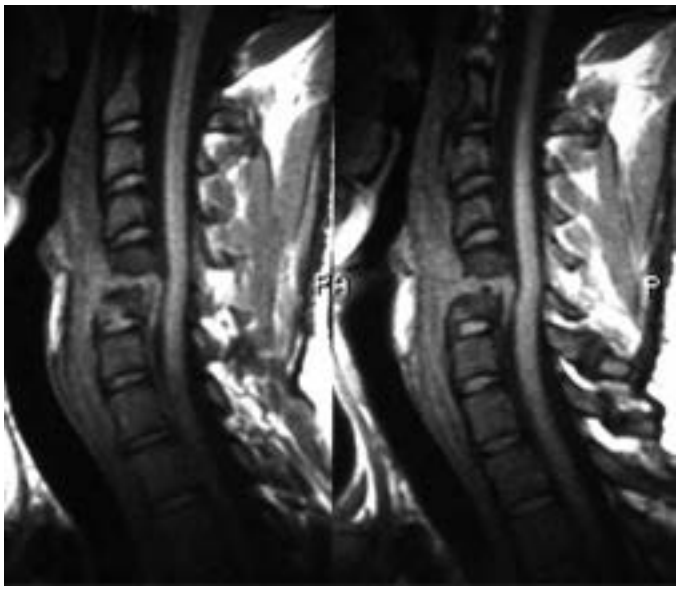

FIGURE 288-2. Noncontrast, non–fat-saturated, T1-weighted, sagittal sequence shows discitis, osteomyelitis, prevertebral and epidural abscess, and cord compression in a patient with *S. aureus* infection of the cervical spine.

many of whom have medical comorbidities. Strokes occur in approximately 20% of patients, and congestive heart failure occurs in 40 to 50%. Twenty-five percent to 30% of patients do not survive the initial hospitalization.

Native valve endocarditis in injection drug users involves the tricuspid valve in approximately three quarters of cases. Patients typically have fever, cough, hemoptysis, and pleuritic chest pain as a consequence of hematogenous seeding of the lung and septic emboli from the valve. The chest radiograph may show pulmonary infiltrates, signs of consolidation or pleural effusion, or multiple, often peripheral nodular pulmonary infiltrates with cavitation, hallmark features of septic embolization (Fig. 288-1). Patients tend to be young and otherwise healthy, so mortality is relatively low, 5% or less. Injection drug users can also have aortic valve or mitral valve endocarditis, in which case the presentation is similar to that described earlier. Conversely, patients who are not injection drug users may have tricuspid valve endocarditis with pulmonary findings.

S. aureus prosthetic valve endocarditis is associated with a 40% or higher in-hospital mortality. Although prosthetic valve endocarditis can be managed medically in some cases, outcomes tend to be worse, and surgery and valve reimplantation are usually required to cure the infection or to manage its complications.[6]

Pericarditis

S. aureus is the most common cause of purulent pericarditis in children and following cardiac surgery in adults. It may occur by contamination at the time of surgery; by bacteremic seeding from another site of infection; as a complication of endocarditis, paravalvular abscess, or myocardial abscess; or by direct extension of infection from pneumonia, lung abscess, or empyema. The presentation is that of acute pericarditis (Chapter 77), with fever and severe chest pain, tachycardia, and hemodynamic instability. The clinical course may be extremely rapid, terminating in septic shock or cardiac tamponade. Immediate drainage of the infected pericardial space and administration of systemic antimicrobial therapy are indicated. Needle pericardiocentesis is useful for confirming the diagnosis and providing temporary decompression of the pericardial space, but definitive therapy requires surgical or continuous tube drainage.

Osteomyelitis

S. aureus is the most common cause of osteomyelitis, both acute and chronic (Chapter 272). The primary mode of infection is hematogenous seeding. Up to a quarter of cases of bacteremia are complicated by osteomyelitis, and concurrent bacteremia is present in 50% or more of osteomyelitis cases. Acute osteomyelitis—defined as an initial episode with a clinical course of days to weeks, but not months—is manifested with fever and pain at the site of infection. Long bones are more commonly infected in children, whereas vertebrae are more commonly infected in adults.[7] Adults can also have long bone osteomyelitis, usually from a contiguous focus of infection or at a site of fracture or prior trauma. Vertebral osteomyelitis is frequently accompanied by paravertebral or epidural abscess (Fig. 288-2). Back pain accompanied by signs of cord compression, such as radicular pain, sensory loss, lower extremity weakness, urinary retention, and bowel or bladder incontinence, is an emergency. Magnetic resonance imaging should be performed as soon as possible to define the location and extent of infection, and neurosurgical consultation should be obtained in anticipation of surgical decompression and drainage.

Septic Arthritis

S. aureus is the most common cause of septic arthritis (Chapter 272), usually as a consequence of bacteremic seeding, trauma, or a surgical procedure. Risk factors include diabetes, recent joint surgery or joint prosthesis (Chapter 276), and rheumatoid arthritis (Chapter 264). Cardinal features are joint pain, history of joint swelling, and fever. The diagnosis is established by analysis of synovial fluid, in which the white blood cell count typically exceeds 25,000/mm^3, with 90% neutrophils. Blood cultures are positive in 30 to 50% of cases, organisms are seen on Gram stain of synovial fluid in about 50% of cases, and synovial fluid culture is almost always positive. Hip, knee, ankle, and wrist are most commonly affected. *S. aureus* also has a predilection for infecting the sternoclavicular, sacroiliac, and symphysis pubis joints. Multiple joints are involved in 5% of cases. Both antimicrobial therapy and drainage of the infected joint (by repeated needle aspiration, by arthrotomy, or arthroscopically) are required to prevent destructive arthritis.

Central Nervous System Infections

S. aureus is an uncommon cause of community-acquired central nervous system infections, such as bacterial meningitis, primary brain abscess, or subdural empyema. These infections are often associated with endocarditis or a contiguous focus of infection, such as cavernous sinus thrombosis. Mortality of these infections is as high as 30 to 50%. *S. aureus* is an important cause of nosocomial meningitis after head trauma, craniotomy, or implantation of intraventricular or extraventricular catheters.

Pulmonary Infections

S. aureus is an uncommon cause of community-acquired pneumonia (Chapter 97), accounting for 1 to 5% of cases. It is typically a severe, often fatal, fulminant necrotizing pneumonia accompanied by evidence of cavitation on chest radiographs. Production of Panton-Valentine leukocidin has been associated with severe pneumonia. Community-acquired staphylococcal pneumonia should be considered, and coverage for MRSA strains should be included in the empirical regimen, in two clinical settings: severe pneumonia requiring admission to an ICU and pneumonia in a patient with influenza.

Hospital-acquired and ventilator-associated pneumonias often are caused by *S. aureus* and MRSA in particular. Mortality can be as high as 40 to 50%, reflecting the virulence of the organism and the comorbid conditions that contribute to poor outcomes. The diagnosis is readily established by

examination of a Gram stain and culture of sputum, tracheal aspirate, or lavage fluid (obtained to avoid oropharyngeal contamination), which typically shows organisms and numerous neutrophils. The culture is less specific than the Gram stain because it can be positive in colonized patients, but it is highly sensitive. Coverage for *S. aureus* can be discontinued if the organism is not isolated in culture.

Lung abscess and pleural empyema, infections most commonly caused by oral anaerobic bacteria, are occasionally caused by *S. aureus*. The clinical course may be subacute or even indolent. Empyema occurs as a complication of prior chest tube placement, surgery, trauma, staphylococcal pneumonia, or tricuspid valve endocarditis. These infections are treated with drainage and antimicrobial therapy.

Orthopedic Device–Associated Infections

Prosthetic joint and implant-associated infections occurring within the first 12 weeks of surgery are most commonly caused by *S. aureus*. *S. aureus* is second only to coagulase-negative staphylococci as a cause of later infections. The presentation may be acute, with joint pain, evidence of arthritis, fever, and systemic signs of infection. Alternatively, the infection may run a more chronic course, with pain and loosening of the prosthesis but little or no fever. Formation of biofilm makes treatment of these infections a challenge. Intraoperative inspection, débridement, and hardware retention may be appropriate for infections of less than 3 weeks in duration or occurring within the first month of implantation.[8] Otherwise, débridement and removal of the prosthesis or implant in conjunction with antimicrobial therapy offer the best chance of cure. This can be accomplished either as a one-stage procedure, in which the infected prosthesis or implant is removed and immediately replaced, or as a two-stage procedure, in which the device is removed and replaced after completion of a 4- to 6-week course of antimicrobial therapy.

Genitourinary Infections

Genitourinary infections arise by hematogenous dissemination or as an ascending infection, usually as a result of instrumentation, urinary catheterization, or surgery. These infections include cystitis, pyelonephritis, microabscess, perinephric abscess, prostatitis, and prostatic abscess. *S. aureus* can also be a contaminant introduced during the collection of urine from a patient with asymptomatic vaginal or perineal colonization. Contamination should be suspected if the urine colony counts are low, repeated urine cultures are negative, pyuria is absent, and there are no signs or symptoms of urinary tract infection. If contamination seems unlikely and there is no well-documented history of an event that could lead to ascending infection, hematogenous dissemination should be suspected. Blood culture samples should be obtained to determine whether there is ongoing bacteremia and appropriate imaging studies performed to identify a deep focus of infection and the presence of renal and perinephric abscesses.

Toxin-Mediated Diseases

These diseases are caused by the ingestion of preformed toxin or the elaboration of toxin by *S. aureus* from a site of colonization or infection. Staphylococcal food poisoning is a gastroenteritis caused by the ingestion of a preformed enterotoxin produced in food contaminated with *S. aureus* from an infected or colonized food handler. When contaminated food is kept at room temperature for several hours, organisms replicate and produce a heat-stable toxin that is not inactivated by cooking or digestive enzymes. Nausea, vomiting, abdominal pain, and diarrhea occur within 2 to 6 hours of eating the contaminated food. There is no fever, and the illness is self-limited, generally lasting about a day. Antibiotics have no role in therapy, which consists of fluid replacement to prevent dehydration. Infants, small children, and the elderly are more severely affected and may require intravenous fluids.

Toxic shock syndrome is caused by colonization or infection with a strain that elaborates a specific toxin, TSST-1, or certain staphylococcal enterotoxins. TSST-1 is responsible for 100% of menstrual cases due to its ability to cross the vaginal mucosa and to achieve systemic concentrations. TSST-1 and the staphylococcal enterotoxins SEA, SEB, and SEC are responsible for nonmenstrual toxic shock syndrome, which is characteristically associated with an identifiable focus of infection, usually of the skin. The diagnosis is clinical, defined by fever, erythroderma, hypotension, involvement of three or more organ systems (renal, hematologic, hepatic, pulmonary, gastrointestinal, muscle, central nervous system, mucous membranes), and desquamation, especially of the palms and soles, 1 to 2 weeks after the onset of illness. Bacteremia is present in only 5% of cases. The case fatality rate is about 5%. Treatment consists of systemic antimicrobial therapy, removal of the source of toxin production, and treatment of septic shock. Streptococcal toxic shock syndrome is discussed in Chapter 290.

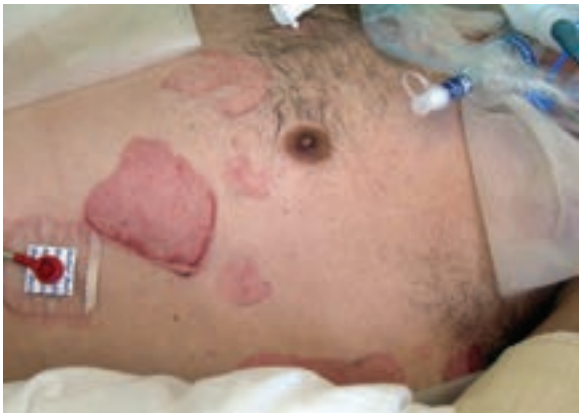

FIGURE 288-3. Scalded skin–type lesions with multiple ruptured bullae and desquamation in a patient with aortic valve endocarditis caused by an exfoliative toxin–producing strain of *S. aureus*.

Colonization or infection with an *S. aureus* strain that produces exfoliative toxin A or B may cause staphylococcal scalded skin syndrome, a disease primarily of infants but occasionally seen in adults (Fig. 288-3), and bullous impetigo, a pustular skin lesion. The appearance is that of a scald, burn, or blister. The differential diagnosis of staphylococcal scalded skin syndrome includes drug reaction (Chapter 440), toxic epidermal necrolysis (Chapter 439), Kawasaki disease (Chapter 439), and pemphigus foliaceus (Chapter 439).

DIAGNOSIS

S. aureus infection is diagnosed by isolating the organism in culture specimens of blood, tissue, or pus. Non–culture-based methods, such as nucleic acid amplification tests, are becoming available, but their utility is still being defined.[9] Culture remains the "gold standard." Gram stain is useful for making a presumptive diagnosis of staphylococcal infection and should be performed whenever possible to look for gram-positive cocci in tetrads or clusters in pus, bone or tissue samples, respiratory secretions, or body fluids such as cerebrospinal fluid, pleural or pericardial fluid, synovial fluid, or urine. Failure to isolate the organism in culture is strong evidence against *S. aureus* infection unless a patient is being actively treated with an antibiotic; even then, infected sites may remain culture positive for several days. The specificity of isolating *S. aureus* from blood or other sterile body sites is essentially 100%. Because of nasopharyngeal colonization in some uninfected individuals, isolation of *S. aureus* from culture of a respiratory specimen lacks specificity; however, if Gram stain also shows gram-positive cocci in tetrads and clusters and many neutrophils, this is suggestive of *S. aureus* pneumonia.

Susceptibility Testing

Susceptibility testing should be performed for clinically significant isolates to guide antimicrobial therapy. The critical determination is whether the isolate is methicillin resistant (i.e., resistant to β-lactam antibiotics). Resistance to macrolides and fluoroquinolones is common, and these drugs should not be used to treat suspected staphylococcal infection without confirmation of in vitro susceptibility. Clindamycin is usually active, although macrolide-resistant strains that produce a ribosomal methylase, inducibly or constitutively, are cross-resistant. Tetracyclines and trimethoprim-sulfamethoxazole are active against 80 to 90% of strains. Resistance to vancomycin, daptomycin, telavancin, or linezolid, although rare, may occur, particularly when there has been prior exposure to the drug, making susceptibility testing important for these antibiotics.

TREATMENT

Rx

The two principles of therapy for *S. aureus* infections are (1) source control by elimination of focal infection and infected foreign material whenever feasible and (2) administration of systemic antimicrobial therapy. For boils and cutaneous abscesses, incision and drainage may be all that is required. Antimicrobial therapy is indicated if the infection is not amenable to removal (e.g., cellulitis, pneumonia), drainage is impossible or inadequate, systemic signs

and symptoms of infection are present, or there is invasive disease (i.e., metastatic sites of infection; involvement of deep tissues, vital organs, sterile sites) and in all cases of bacteremia. An undrained focus of infection or retention of an infected foreign body is the most common reason for unsatisfactory clinical response, treatment failure, or relapse.

The most important consideration in selecting an antibiotic is susceptibility of the *S. aureus* isolate to β-lactams. A penicillinase-resistant penicillin, such as nafcillin (1 to 2 g every 4 to 6 hours IV, depending on the severity of the infection), oxacillin, or flucloxacillin, or a cephalosporin (e.g., cefazolin 1 to 2 g every 8 hours IV) is the agent of choice for the treatment of methicillin-susceptible *S. aureus* (MSSA) infections; no other antibiotic is as safe or as effective as a β-lactam. An orally administered β-lactam (e.g., dicloxacillin 500 mg four times daily or cephalexin 500 mg four times daily) is appropriate for most cutaneous infections; a parenteral agent is recommended, at least initially, for invasive infections. Only if the patient is allergic or has a serious reaction is an agent other than a β-lactam preferred for the treatment of infection caused by an MSSA strain.

Methicillin-resistant strains are cross-resistant to all currently available β-lactams, except for ceftaroline (dose of 600 mg IV every 12 hours), which is indicated for treatment of skin and skin structure infections caused by methicillin-resistant strains. Otherwise, a β-lactam should not be used for treatment of infection known or suspected to be caused by a methicillin-resistant strain. The prevalence of methicillin resistance in *S. aureus* isolates from hospital-acquired and health care–associated infections in the United States and in many European countries is 25 to 50% or higher. MRSA also causes a substantial proportion of community-onset infections in individuals lacking other risk factors, especially in the United States. Trimethoprim-sulfamethoxazole (one or two 80/160-mg tablets twice a day), clindamycin (300 mg three times a day), and doxycycline or minocycline (100 mg twice a day) are active in vitro against most community-acquired MSSA and MRSA strains and are effective when administered orally for the treatment of skin and soft tissue infections in outpatients. Fluoroquinolones should not be used because most methicillin-resistant strains are also fluoroquinolone resistant.

For invasive infections, vancomycin is still a drug of choice.[10] It must be administered intravenously; doses of 30 to 60 mg/kg/day, adjusted on the basis of creatinine clearance, are recommended to achieve trough serum concentrations of 15 to 20 μg/mL for patients with bacteremia, endocarditis, or other serious infections. Treatment failures are not uncommon with bacteremia or endocarditis. Patients can remain persistently bacteremic (≥3 days or longer) or relapse, even when the strain is susceptible (minimal inhibitory concentration [MIC] ≤2 μg/mL) in vitro. Other than the presence of an undrained focus of infection, the reasons for this are unclear. Possible explanations include vancomycin's slowly bactericidal activity; tolerance, in which the isolate is inhibited at low concentrations but not killed; or so-called heteroresistance, in which a small fraction of the population of organisms has a higher MIC. Vancomycin MIC of 2 μg/mL has been associated with treatment failure in some retrospective studies but not in others.[11-13] Strains with intermediate susceptibility to vancomycin (MICs of 4 or 8 μg/mL), which account for 1 to 3% of MRSA isolates, should be considered resistant because this is highly predictive of vancomycin treatment failure. Vancomycin-resistant strains (MIC >8 μg/mL), which express the *vanA* gene, are rare. Alternatives to vancomycin include quinupristin-dalfopristin (7.5 mg/kg every 8 to 12 hours, rarely used because of poor tolerability and limited efficacy data), linezolid (600 mg every 12 hours, Food and Drug Administration [FDA] approved for pneumonia and skin and soft tissue infection), daptomycin (FDA-approved doses of 4 mg/kg once daily for complicated skin and soft tissue infections and 6 mg/kg once daily for bacteremia), telavancin (10 mg/kg once daily, approved for skin and soft tissue infection and pneumonia when alternative treatments are not suitable), and ceftaroline (skin and soft tissue infection). These alternative agents have been shown to be noninferior to vancomycin in high-quality, randomized controlled trials, and none has demonstrated superiority.

Some authorities recommend higher doses of daptomycin (e.g., 10 mg/kg/day) for treatment of bacteremia or endocarditis, particularly when the infection has failed to respond to vancomycin. Daptomycin should not be used to treat primary staphylococcal pneumonia because it is inactivated by pulmonary surfactant, although it is indicated for treatment of hematogenous pneumonia, as occurs in tricuspid valve endocarditis or septic pulmonary embolization.

The role of combination therapy is ill defined. Aminoglycoside combination regimens have not been shown to improve outcomes, but there are good data demonstrating increased toxicity and adverse events with aminoglycoside combinations. Aminoglycoside combinations should not be used routinely, and if used at all, they should be reserved for patients who have failed to respond to first-line therapy. Rifampin combination therapy is recommended for the treatment of osteomyelitis, device-related bone infection and prosthetic joint infection, or prosthetic valve endocarditis, particularly that caused by methicillin-resistant strains of staphylococci. Rifampin (300 to 450 mg twice daily) must always be administered in combination with a second active agent because resistance emerges rapidly during therapy.

PREVENTION

The emergence of community MRSA and the large burden of hospital-acquired and health care–associated staphylococcal infections have stimulated renewed interest in prevention strategies. The organisms that cause these infections are usually resident flora (either *S. aureus* or coagulase-negative staphylococci), or they are acquired by direct contact with a contaminated source, such as a wound or dressing, the skin or hands of an asymptomatically colonized individual, or a contaminated health care provider. The most effective strategy is adherence to principles of basic infection control, the key component of which is hand hygiene, whether it is handwashing or use of an alcohol-based hand rub. This disrupts transmission of organisms by the hands of care providers, a well-documented source of bacterial contamination. Barrier precautions (gloves and gowns) are important for minimizing contact with infected wounds, contaminated secretions, and dressings. Isolation precautions and screening for asymptomatic carriage are more controversial and less well documented in terms of their efficacy. For patients undergoing surgical procedures, surgical hand and surgical site antisepsis, aseptic surgical technique, and antimicrobial prophylaxis are important preventive measures.

Another potentially effective means of preventing infection is screening and decolonization of *S. aureus* carriers. Studies to determine whether screening, decolonization, and isolation actually prevent MRSA infection have had mixed results. In the ICU setting, universal decolonization with 2% chlorhexidine bath cloths and 2% intranasal mupirocin ointment was found to be more effective than targeted screening and decolonization in reducing MRSA rates. This approach has yet to be widely adopted, and one concern is emergence of resistant strains.

Decolonization may be considered in two other settings: prevention of recurrent infection in individuals who have had several prior episodes and prevention of surgical site infections. The best-studied regimens are topically applied mupirocin nasal ointment (0.5 g twice daily in each nostril for 5 days), with or without bathing with chlorhexidine soap, and orally administered rifampin (600 mg in one or two divided doses) in combination with another active agent (e.g., a fluoroquinolone if the isolate is susceptible, trimethoprim-sulfamethoxazole, doxycycline). In a randomized, double-blind, placebo-controlled, multicenter trial, the number of surgical site *S. aureus* infections acquired in the hospital was reduced by the rapid screening of nasal carriers by a real-time polymerase chain reaction assay and the decolonization of carriers with mupirocin nasal ointment and chlorhexidine soap. Several antistaphylococcal vaccine candidates are currently in phase I and phase II clinical trials; the availability of an effective vaccine would be an important tool in preventing staphylococcal infections.

PROGNOSIS

Prognosis of *S. aureus* infections and outcomes depend on the site of infection, adequacy of source control, presence of comorbidities (e.g., diabetes; immunosuppression; underlying cardiac, renal, or liver disease), presence of bacteremia, presence of secondary foci of infection, presence of severe sepsis or septic shock, antibiotic effectiveness, and duration of therapy for complicated disease. Methicillin resistance is a risk factor for poorer outcome largely because of its health care association, and thus occurrence in a population that is elderly and in which comorbid medical illnesses are prevalent, and possibly because less effective antibiotics (non–β-lactams) are used to treat these infections. Historically, untreated *S. aureus* bacteremia was lethal in 85% or more of cases. Antibacterial therapy, 4 to 6 weeks or longer for complicated bacteremia or infections of deep tissues, and recognition of the importance of source control have dramatically improved outcome. Mortality remains high, in the range of 20 to 40%, in patients with severe sepsis, septic shock, or endocarditis.

COAGULASE-NEGATIVE STAPHYLOCOCCI

More than 30 different species of coagulase-negative staphylococci have been identified, and about half of these colonize humans. *Staphylococcus epidermidis* is the species that most commonly causes infection. Coagulase-negative staphylococci infrequently cause infection unless there is a foreign body in place, and although bacteremia occurs, metastatic seeding of secondary sites of infection is distinctly uncommon. Coagulase-negative staphylococci are typically resistant to methicillin and multiple other antibiotics, and they are an important reservoir of drug resistance elements that are horizontally transferrable to *S. aureus*. They are the most common cause of health care–associated infections overall in the United States (Chapter 282); they account

for one third of central line–associated blood stream infections; and they are the second most common cause of surgical site infections, particularly when a prosthetic device or other foreign material has been implanted. Infections are often indolent, causing little in the way of fever or systemic signs of infection, but they may also be acute and life-threatening, as in the case of prosthetic valve endocarditis. Coagulase-negative staphylococci are proficient biofilm producers; consequently, débridement and removal of the infected prosthetic device or foreign body are paramount. Prosthetic joint infections occurring more than 1 month after device implantation are best managed with removal of the prosthesis and reimplantation in a one-stage or two-stage procedure. Antimicrobial therapy for these infections is similar to that for *S. aureus*, except that hematogenous seeding rarely occurs, so attention is focused primarily on source control.

Coagulase-negative staphylococci, because they are normal skin flora, are the most common blood culture contaminant. In approximately 75% of cases, when the blood culture is positive for coagulase-negative staphylococci, this reflects contamination rather than infection. Sorting out whether a positive culture represents contamination or true infection can be a challenge. A single positive blood culture or blood cultures in which more than one strain is present are likely to be due to contamination. Time to blood culture positivity, quantitative blood cultures, and the presence of multiple positive cultures can be useful in determining whether a positive blood culture represents true infection. Isolation of coagulase-negative staphylococci from the blood of a patient with a prosthetic valve, intravenous pacemaker, or vascular graft can be especially problematic because these patients are at high risk for true infection. Unless the patient is hemodynamically unstable or otherwise seriously ill, it is advisable to withhold antibiotics until additional culture specimens are obtained to document the presence of true bacteremia. Isolation of coagulase-negative staphylococci from culture specimens of deep tissue, bone, prosthetic devices, or other normally sterile sites, especially if multiple cultures are positive, strongly suggests true infection.

Staphylococcus lugdunensis, in contrast to other coagulase-negative staphylococci, is pathogenic and causes infections in the absence of a foreign body and in otherwise normal hosts that clinically resemble *S. aureus* infections. These include prosthetic valve and native valve endocarditis, bacteremia, skin and soft tissue infection, septic arthritis, prosthetic joint infection, and osteomyelitis. It lacks free coagulase, but some strains produce a membrane-bound form that can lead to its misclassification as *S. aureus*. *S. lugdunensis* lacks protein A and is positive for ornithine decarboxylase and pyrrolidonyl arylamidase, differentiating it from *S. aureus*. *S. lugdunensis* is susceptible to most antibiotics, including penicillin (approximately 75% of isolates), and resistance to nafcillin and oxacillin is rare. The management of these infections is similar to that of *S. aureus* infections.

Grade A References

A1. Wang Y, Zou Y, Xie J, et al. Linezolid versus vancomycin for the treatment of suspected methicillin-resistant Staphylococcus aureus nosocomial pneumonia: a systematic review employing meta-analysis. *Eur J Clin Pharmacol*. 2015;71:107-115.

A2. Fowler VG Jr, Boucher HW, Corey GR, et al. Daptomycin versus standard therapy for bacteremia and endocarditis caused by *Staphylococcus aureus*. *N Engl J Med*. 2006;355:653-665.

A3. Corey GR, Wilcox MH, Talbot GH, et al. CANVAS 1 investigators. CANVAS 1: the first Phase III, randomized, double-blind study evaluating ceftaroline fosamil for the treatment of patients with complicated skin and skin structure infections. *J Antimicrob Chemother*. 2010;65(suppl 4): iv41-iv51.

A4. Rubinstein E, Lalani T, Corey GR, et al. ATTAIN Study Group. Telavancin versus vancomycin for hospital-acquired pneumonia due to gram-positive pathogens. *Clin Infect Dis*. 2011;52:31-40.

A5. Zimmerli W, Widmer AF, Blatter M, et al. Role of rifampin for treatment of orthopedic implant–related staphylococcal infections: a randomized controlled trial. Foreign-Body Infection (FBI) Study Group. *JAMA*. 1998;279:1537-1541.

A6. Huang SS, Septimus E, Kleinman K, et al. CDC Prevention Epicenters Program; AHRQ DECIDE Network and Healthcare-Associated Infections Program. Targeted versus universal decolonization to prevent ICU infection. *N Engl J Med*. 2013;368:2255-2265.

A7. Bode LG, Kluytmans JA, Wertheim HF, et al. Preventing surgical-site infections in nasal carriers of *Staphylococcus aureus*. *N Engl J Med*. 2010;362:9-17.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Otto M. Community-associated MRSA: what makes them special? *Int J Med Microbiol*. 2013; 303:324-330.
2. Kim HK, Thammavongsa V, Schneewind O, et al. Recurrent infections and immune evasion strategies of *Staphylococcus aureus*. *Curr Opin Microbiol*. 2012;15:92-99.
3. Fu C-M, Tseng W-P, Chiang W-C, et al. Occult *Staphylococcus aureus* bacteremia in adult emergency department patients: rare but important. *Clin Infect Dis*. 2012;54:1536-1544.
4. Khatib R, Sharma M. Echocardiography is dispensable in uncomplicated *Staphylococcus aureus* bacteremia. *Medicine (Baltimore)*. 2013;92:182-188.
5. Hoen B, Duval X. Clinical practice. Infective endocarditis. *N Engl J Med*. 2013;368:1425-1433.
6. Athan E, Chu VH, Tattevin P, et al. ICE-PCS Investigators. Clinical characteristics and outcome of infective endocarditis involving implantable cardiac devices. *JAMA*. 2012;307:1727-1735.
7. Zimmerli W. Clinical practice. Vertebral osteomyelitis. *N Engl J Med*. 2010;362:1022-1029.
8. Osmon DR, Berbari EF, Berendt AR, et al. Infectious Diseases Society of America. Diagnosis and management of prosthetic joint infection: clinical practice guidelines by the Infectious Diseases Society of America. *Clin Infect Dis*. 2013;56:e1-e25.
9. Liu C, Bayer A, Cosgrove SE, et al. Clinical practice guidelines by the Infectious Diseases Society of America for the treatment of methicillin-resistant *Staphylococcus aureus* infections in adults and children: executive summary. *Clin Infect Dis*. 2011;52:285-292.
10. Holland TL, Arnold C, Fowler VG Jr. Clinical management of Staphylococcus aureus bacteremia: a review. *JAMA*. 2014;312:1330-1341.
11. Park SY, Oh IH, Lee HJ, et al. Impact of reduced vancomycin MIC on clinical outcomes of methicillin-resistant *Staphylococcus aureus* bacteremia. *Antimicrob Agents Chemother*. 2013;57: 5536-5542.
12. Chong YP, Park SJ, Kim HS, et al. Persistent *Staphylococcus aureus* bacteremia: a prospective analysis of risk factors, outcomes, and microbiologic and genotypic characteristics of isolates. *Medicine (Baltimore)*. 2013;92:98-108.
13. Kalil AC, Van Schooneveld TC, Fey PD, et al. Association between vancomycin minimum inhibitory concentration and mortality among patients with Staphylococcus aureus bloodstream infections: a systematic review and meta-analysis.1. *JAMA*. 2014;312:1552-1564.

REVIEW QUESTIONS

1. Ms. A is a 53-year-old woman admitted 1 week ago for fevers to 102.3° F and right flank pain. A computed tomography scan showed findings consistent with a 6-cm psoas abscess. Three blood culture samples were drawn, and empirical therapy was begun with vancomycin and piperacillin-tazobactam. All three blood culture samples grew methicillin-resistant *Staphylococcus aureus* (MRSA) with a vancomycin minimal inhibitory concentration (MIC) of 2 μg/mL. Treatment was de-escalated to vancomycin alone with documented trough concentrations of 15 μg/mL. One of two blood culture samples obtained on day 5 of therapy because of recurrent fever now is reported as positive for gram-positive cocci in clusters. Which of the following is the *most likely* explanation for the persistently positive blood culture?

A. Vancomycin-nonsusceptible MRSA strain
B. Failure to obtain a transesophageal echocardiogram
C. Undrained psoas abscess
D. Subtherapeutic levels of vancomycin
E. Contamination of the blood culture sample with coagulase-negative staphylococci

Answer: C An isolate with a vancomycin MIC of 2 μg/mL is considered to be susceptible. Vancomycin-intermediate strains have low prevalence, 1 to 3%, and the MIC of these strains is 4 to 8 μg/mL. Vancomycin-resistant strains (MIC > 8 μg/mL) are rare. Thus, vancomycin nonsusceptibility is very unlikely. The vancomycin trough is in the therapeutic range, and although an echocardiogram is recommended for patients with complicated bacteremia, failure to obtain this study has not been shown to affect outcome. It is a mistake to attribute a positive blood culture to a contaminant in a patient with known *S. aureus* bacteremia. Poor source control with an undrained abscess, a well-known cause of persistent bacteremia, is the best answer.

2. Ms. B is a 58-year-old patient with degenerative arthritis who had a right hip replacement performed 3 weeks ago. She presents with about a week of slowly worsening hip pain with breakdown of the surgical site, cultures of which grew methicillin-susceptible *S. aureus*. Blood cultures are negative, and she has no fevers. She is adamant that she does not want the hardware removed. What is the best course of action in this situation?

A. Begin oral cephalexin and discharge the patient for follow-up in 2 weeks to see if she improves.
B. Begin a 2 week course of once-daily IV daptomycin and discharge the patient for follow-up in 2 weeks to assess need for further therapy.
C. Treat her with cefazolin for a week, and if she does not get better, perform incision and drainage.
D. Tell her that with incision, drainage, and debridement and a course of IV and oral antibiotics, she should do well without removal of the hardware.
E. Try to convince her to have the hardware removed because if it is left in place, the infection cannot be cured.

Answer: D The patient is likely to have a methicillin-susceptible infection of the prosthesis. Because the infection is within a month of implantation and she has had less than 3 weeks of symptoms, she should do well with incision and drainage and a prolonged course of IV and oral antibiotics (3 to 6 months). Because hardware removal is probably not necessary, trying to threaten her into a procedure she has already said she does not want is unwise. Empirical therapy with an oral (A) or parenteral (C) β-lactam is ill-advised because it is important to document whether the joint is involved and to perform incision and drainage to minimize the possibility of progression of infection and a worse outcome. Empirical therapy with daptomycin is a poor choice for the same reasons, and moreover, a β-lactam is the preferred initial therapy for methicillin-susceptible *S. aureus* (MSSA).

3. Mr. C is a previously healthy 23-year-old man who is admitted with a 1-week illness characterized by a dry cough, fever, and myalgias. Yesterday he noticed that the cough became productive of yellow sputum, and last night be noticed shortness of breath. On examination, his temperature is 104° F, pulse is 128, respiratory rate is 22, blood pressure is 100/60, and pulse oximetry is 80% saturation on room air. He has rales at the right base, and a chest radiograph shows a fluffy right lower lobe infiltrate. A rapid influenza test result is positive, and sputum Gram stain shows numerous polymorphonuclear leukocytes and a few gram-positive cocci in clusters. Two blood culture samples are obtained. Which one of the following antibiotics is the worst choice to treat this suspected pneumonia?

A. Clindamycin
B. Ceftaroline
C. Daptomycin
D. Linezolid
E. Vancomycin

Answer: C This patient has *S. aureus* pneumonia as a complication of influenza. The susceptibility of the organism is unknown, but he is at risk for MRSA infection. Either linezolid or vancomycin would be an appropriate choice as both are first-line agents. Of the three other antibiotics, all have issues, but daptomycin is the worst choice of the three because it is inactivated by pulmonary surfactant, resulting in treatment failure, and it is not indicated for treatment of pneumonia. Clindamycin could be used because most community MRSA and MSSA strains are susceptible, and although it is not a first-line agent in adults, it has been used to treat children with staphylococcal pneumonia. Ceftaroline, although not approved by the Food and Drug Administration for staphylococcal pneumonia, has an indication for treatment of community-acquired pneumonia, and it is active against both MRSA and MSSA. Thus, in a pinch, either clindamycin or ceftaroline would be preferred to daptomycin.

4. Mr. T is a 38-year-old diabetic patient who presents with a 2-day history of worsening pain and swelling of his left upper arm and subjective fevers. He tells you that he is penicillin allergic. His temperature is 99.8° F, and other vital signs are normal. On physical examination, there is a 5-cm-diameter tender, red, indurated lesion with central fluctuance over the triceps area. You perform incision and drainage, which yields about 3 mL of pus. Which one of the following antibiotics should *not* be used as empirical therapy for this community-acquired skin abscess?

A. Clindamycin
B. Doxycycline
C. Levofloxacin
D. Minocycline
E. Trimethoprim-sulfamethoxazole

Answer: C Clindamycin, doxycycline, minocycline, and trimethoprim-sulfamethoxazole are active in vitro against both MRSA and MSSA strains and have been recommended and used to treat skin and soft tissue infections in areas with high prevalence of MRSA. Many MRSA strains and some MSSA strains are fluoroquinolone resistant, and therefore fluoroquinolones should not be used in the absence of documented in vitro susceptibility.

5. Which one of the following measures is probably the most effective means of preventing MRSA transmission?

A. Chlorhexidine bath
B. Good hand hygiene
C. Isolation
D. Mupirocin nasal ointment
E. Screening for colonization

Answer: B Although randomized, controlled trials are lacking, on the basis of the known mechanisms of *S. aureus* transmission, whether MRSA or MSSA, good hand hygiene is probably the most effective strategy to prevent spread of the organism. Chlorhexidine baths alone have not proved effective. Screening and isolation of patients have had variable success. Mupirocin nasal ointment alone may reduce the risk of perioperative infection in colonized individuals but has not been shown to prevent transmission. The study of Huang and colleagues[A6] found that in an intensive care unit setting, screening and isolation and targeted MRSA decolonization were relatively ineffective in reducing MRSA rates, whereas universal decolonization did reduce rates.

289

STREPTOCOCCUS PNEUMONIAE INFECTIONS

LIONEL A. MANDELL

DEFINITION

The term *pneumococcal pneumonia* refers to infection of the pulmonary parenchyma and its associated structures by *Streptococcus pneumoniae*. There are an estimated 5 million cases annually worldwide, and community-acquired pneumonia (CAP) of all etiologies is the commonest cause of death from infection in the United States and Europe.[1]

The Pathogen

S. pneumoniae is a gram-positive coccus that typically grows in pairs or short chains. Careful examination of the diplococcal form reveals slightly tapered ends that give rise to its lancet-shaped appearance. It is a facultative anaerobe that grows best on blood agar plates in a 5% carbon dioxide ambient environment. The colonies are typically surrounded by a greenish zone of hemolysis resulting from degradation of hemoglobin by a pneumococcal toxin. The organism can be distinguished from other streptococci by its susceptibility to ethyl hydrocupreine (Optochin) and bile solubility.

The surface of the pneumococcus consists of a capsule and a cell wall. The capsule, which helps prevent phagocytosis, is composed of polysaccharides that define at least 92 different pneumococcal serotypes. The cell wall is a dynamic structure composed of more than a dozen distinct glycopeptides.

EPIDEMIOLOGY

The pneumococcus is one of the most common causes of CAP and can cause hospital-acquired and health care–associated pneumonia as well.[2] The false impression that *S. pneumoniae* is uncommon has arisen because at least one third of patients with CAP are unable to produce sputum and, even if they do, often provide an inadequate sample. Moreover, if the patient has recently taken just one dose of a drug to which the pneumococcus is susceptible, it may not be possible to isolate the organism.

The ecologic niche of the pneumococcus is the nasopharynx, and up to 80% of infants and 20% of healthy adults may be colonized. Colonizing and invasive strains of pneumococci have developed adaptive mechanisms allowing them to escape host responses. Simultaneous colonization with more than one capsular type has been reported. The horizontal transfer of genes that can occur can expand the gene pool of the colonizing pneumococci, thereby enhancing their adaptive abilities and virulence properties. Asymptomatic colonization is an immunizing experience because homologous anticapsular antibodies may be demonstrated in individuals after colonization with a specific serotype.

A particular serotype may colonize the nasopharynx for varying periods, but the average duration in infants is 7 weeks. Carriage rates are highest during the late fall, winter, and early spring. Although each of the 92 serotypes is potentially pathogenic, the most frequently encountered are types 3, 4, 6, 7, 9, 12, 14, 18, 19, and 23.

Person-to-person transmission results from close interpersonal contact. Although pneumococcal pneumonia is typically a sporadic illness, epidemics can occur in crowded settings such as daycare centers, barracks, nursing homes, and prisons.

The incidence of pneumococcal pneumonia is about 18 per 100,000 people per year. For persons 5 years or younger and 75 years or older, rates are about 23 and 35.8 per 100,000 people, respectively. For pneumococcal bacteremia, the incidence is about 7.5 per 100,000 people per year and increases with age; the case-fatality rate is 21%.

Risk factors for pneumococcal pneumonia include dementia, seizure disorders, heart failure, cerebrovascular disease, chronic obstructive pulmonary disease (COPD), human immunodeficiency virus (HIV) infection, previous viral respiratory illness, alcoholism, malnutrition, diabetes, cirrhosis, and renal insufficiency.

Certain ethnic groups such as Native Americans, particularly Alaskan natives, and Australian Aboriginals, appear to be particularly susceptible to

invasive pneumococcal infection. Any individual who has a defect in immunoglobulin G (IgG) synthesis (Chapter 250) or phagocyte function (Chapter 169) or who has undergone splenectomy is also at increased risk for invasive pneumococcal infection.

PATHOBIOLOGY

Pneumonia is the result of a breakdown in the interplay between colonizer and host as well as pneumococcal virulence factors that can influence both colonization and invasion. In pneumococcal pneumonia, the microorganism first colonizes the nasopharynx. Aspiration of small amounts of oropharyngeal contents occurs in deep sleep, even in normal individuals, but if the oropharyngeal material includes pneumococcal serotypes associated with invasive infection and if normal clearance mechanisms fail, the colonizers may become pathogens. Whether or not a particular pneumococcus results in active infection is related to a number of virulence factors that allow it to avoid or mitigate host defenses and to attack host cells. These include capsular polysaccharides, which allow avoidance of opsonophagocytosis; exoglycosidases, which enable access to cell receptors and choline-binding proteins; and divalent metal-ion binding lipoproteins, which can act as adhesins. Autolysins can cause lysis of host cells and pneumolysin, and choline-binding proteins may enable the pathogen to gain access to the vascular system, potentially resulting in bacteremia and metastatic infection.[3] The risk for infection is increased by any process or condition that exposes the host to the pathogen, that allows oropharyngeal secretions to bypass upper airway defenses, or that interferes with the host's ability to ingest and kill the pneumococci (Table 289-1).

It appears that aspirated pneumococci adhere to type II cells in the alveolus if they are not cleared by normal defense mechanisms. Attachment to resting cells is mediated by two classes of glycoconjugates, but local inflammatory mediators upregulate host cell receptors, such as those for platelet-activating factor (PAF), and provide a site of attachment for the bacteria. This interaction between PAF receptor and pathogen appears to be an important step in internalization of the bacteria by means of an endocytic vacuole and may promote invasion. Expression of pili by certain pneumococcal strains facilitates binding to epithelial cells and results in a more vigorous tumor necrosis factor (TNF)–dependent inflammatory response, which can cause further tissue damage.

In the lung, the pneumococci are able to activate complement and stimulate the cytokine response. Initially, the alveoli fill with fluid exudate (Fig. 289-1), which allows the infection to spread to adjacent uninfected alveoli. In healthy lungs, polymorphonuclear leukocytes (PMNs) constitute less than 1% to 2% of alveolar cells and normally reside in the interstitial areas of the lung and in adjacent capillaries. Recruitment of PMNs into alveoli depends on the generation of chemoattractants necessary for the directed migration of neutrophils.

Ultimately, the signs and symptoms of disease are both attributable to the pathogens themselves and to the body's response to them. The bacterial cytotoxin pneumolysin and various pneumococcal cell wall components are able to induce a variety of effects that initiate and then enhance the inflammatory response, thereby resulting in the various signs and symptoms of pneumonia.

The effects of pneumococcal infection are ultimately manifested as changes in lung mechanics secondary to reductions in lung volumes and lung compliance, as well as gas exchange problems resulting from intrapulmonary shunting and subsequent arterial hypoxemia. If severe enough, death may ensue.

Antibiotic Resistance

The phenotypic expression of antibiotic resistance corresponds to genetic alterations resulting from horizontal acquisition of foreign genetic information or from mutations in the microbial genome. The impact of antibiotic resistance as documented in vitro is not as straightforward as one might imagine, and there is not a direct relationship between resistance and clinical outcomes.

With *S. pneumoniae,* resistance may be acquired by direct DNA incorporation and remodeling from closely related oral commensal bacteria by the process of natural transformation. Pneumococcal resistance to β-lactams such as penicillin is solely attributable to the presence of low-affinity penicillin-binding proteins (PBPs). Increasing the concentration of the β-lactams can usually overcome such resistance. Revision of the break points for resistance in nonmeningeal isolates has resulted in a lower prevalence of penicillin resistance than was previously reported.[4] The PBPs themselves are *trans* and carboxypeptidase enzymes that are involved in bacterial cell wall synthesis and represent the primary sites of action for β-lactam drugs.

Resistance to macrolides, in contrast, can occur through multiple mechanisms, including target site modification or an efflux pump. Target site modification is caused by a ribosomal methylase encoded by the *ermB* gene. A change in 23S recombinant RNA mediated by this gene can result in resistance to macrolides, lincosamides, and streptogramin B–type antibiotics (MLS_B phenotype). The efflux mechanism is encoded by the *mefA* gene, which results in an M phenotype. The former is typically associated with high-level resistance with minimal inhibitory concentrations (MICs) of 64 µg/mL or higher, but the latter is usually associated with low-level resistance with MICs of 1 to 32 µg/mL. These two mechanisms account for approximately 45% and 55%, respectively, of resistant isolates. Clonal dissemination is important in the spread of macrolide resistance. Some pneumococcal isolates have been found with both the *erm* and *mef* genes, but the significance of such a finding is unknown.

Resistance to fluoroquinolones may be mediated by changes in one or both target sites (topoisomerase II and IV) of the pneumococcus, which usually result from mutations in the *gyrA* and *parC* genes, respectively. Low-level resistance may result from changes in one site, but high-level resistance can result when dual mutations occur. An efflux pump may have a role as well. Also of concern is the fact that the incidence of multidrug-resistant isolates is increasing.

TABLE 289-1 PREDISPOSING FACTORS FOR PNEUMOCOCCAL PNEUMONIA

INCREASED EXPOSURE TO *STREPTOCOCCUS PNEUMONIAE*
Prisons
Military barracks
Daycare centers
Shelters for the homeless
DECREASED HOST DEFENSES
Complement deficiency
Antibody deficiency
Functional or anatomic asplenia
Decreased numbers or function of phagocytes
SPECIFIC DISEASE ENTITIES
Multiple myeloma
Lymphoma
Chronic lymphocytic leukemia
Human immunodeficiency virus infection
RESPIRATORY AND PULMONARY PROBLEMS
Chronic obstructive pulmonary disease
Smoker
Allergies
Previous viral infection

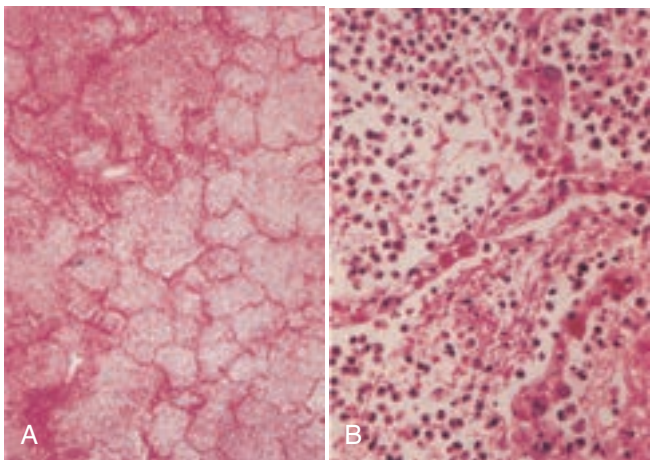

FIGURE 289-1. Pneumonia. A, Low-power magnification (×100) of hematoxylin and eosin (H&E) stain of tissue section from the left lower lobe of the lung. Note the intact alveolar walls and alveoli filled with edema and thick cellular exudates. **B,** Higher magnification (× 500) H&E stain of the same section shown in **A.** Note the heavy infiltrate of polymorphonuclear cells and the intact alveolar walls.

CLINICAL MANIFESTATIONS

The clinical features of pneumococcal pneumonia depend on factors such as whether the patient is immunocompetent or immunosuppressed, the severity of illness, and whether the patient has taken antibiotics. Temperatures can vary from 101° F to higher than 103° F (38° to 39.5° C), can be accompanied by chills and rigors, and are usually associated with a tachycardic response. A cough, typically productive of purulent and occasionally blood-tinged sputum, is often present, and as many as 46% of patients report chest pain. About 20% have gastrointestinal symptoms such as nausea, vomiting, or diarrhea. In elderly people, symptoms and signs may be more subtle; for example, older patients may have predominantly confusion.

On examination, the patient is often listless and may be cyanotic. The respiratory rate is increased; if pleuritic pain is marked, the patient may be splinting the affected side. Whereas dullness to percussion over a lung segment suggests consolidation, a flat percussion note is typically associated with a pleural effusion. Breath sounds may be "distant" if there is an overlying effusion, but they are bronchial in nature if the underlying lung is consolidated. Rales may be noted. If the patient has pleurisy without much pleural fluid, a friction rub may be heard.

No radiographic appearance is characteristic of pneumococcal pneumonia (Chapter 97). Typically, however, involvement is limited to one or more segments within a single lobe. Involvement is unilateral approximately 80% of the time, and the presence of cavitation or lung abscess is uncommon. Among patients with pneumococcal pneumonia, necrotizing changes in the lungs were seen in 6.6% of cases in a large series but were often overlooked on initial readings of chest radiographs or computed tomography (CT) scans. Forty-five percent of patients have an associated pleural effusion, but only 15% have an effusion of sufficient size to warrant drainage (>10 mm on lateral decubitus views). Purulent pericarditis can also be seen (Chapter 77).

Patients with lobar consolidation are more likely to be bacteremic, but there are no consistently significant differences in the radiologic manifestations of bacteremic and nonbacteremic pneumococcal pneumonia. Bacteremic pneumococcal pneumonia may seed distant sites and cause meningitis (Chapter 412), endocarditis (Chapter 76), or septic arthritis (Chapter 272).

With sepsis, sepsis syndrome, or septic shock (Chapter 108), the patient may be hypotensive, and the findings of organ failure vary depending on the target organ(s) involved. For example, oliguria, anuria, and acidosis suggest renal failure; myocardial impairment suggests heart failure; and jaundice is consistent with hepatic failure. Systemic activation of coagulation together with consumption of clotting proteins can result in simultaneous clotting and bleeding (Chapter 175). In some cases, peripheral gangrene and purpura fulminans may be seen.

DIAGNOSIS

Despite extensive testing, a specific etiologic agent is not found in 50% or more of patients with CAP. The history, physical examination, chest radiograph, sputum Gram stain, and blood and sputum cultures are insensitive and lack specificity, but more invasive methods (e.g., endotracheal aspirate, bronchoscopy techniques, pleural fluid aspiration, and lung biopsy) require special expertise and are infrequently indicated. As a general rule, routine diagnostic tests to identify a pathogen are optional for outpatients with CAP, but more extensive testing is recommended for hospitalized patients (Table 289-2).[5]

Ultimately, the diagnosis should be based on suggestive clinical features and a chest radiograph with or without corroborating microbiologic data. Because *S. pneumoniae* is the most common etiologic agent of CAP, virtually any recommended treatment regimen must provide adequate coverage for it, so documentation of its presence has little or no effect on ultimate clinical outcomes. From the point of view of epidemiology and antimicrobial susceptibility, however, documentation of a pathogen is desirable.

The sputum Gram stain is a relatively simple and inexpensive procedure to document the presence of certain pathogens. The adequacy of the specimen is based on the relative number of neutrophils and squamous epithelial cells (SECs). There should be at least 25 neutrophils and fewer than 10 SECs per low-power field (×100 magnification). The sensitivity of sputum Gram stain for *S. pneumoniae* is 55%, but the specificity is higher than 80%. Unfortunately, approximately 30% of patients overall are unable to produce an appropriate sputum sample. In elderly patients, this figure reaches almost 70%. Overall, only 28% at best of sputum samples are of good quality. Sputum cultures are neither sensitive nor specific, particularly when dealing with relatively fastidious pathogens such as *S. pneumoniae.*

Blood cultures are positive in only 5% to 14% of patients with CAP. If patients have taken a previous dose of antibiotic, culture results are even less useful. As a result, sputum Gram stain, sputum culture, and blood cultures are generally recommended only in patients with selected clinical indications (see Table 289-2). In patients with severe CAP, an expectorated sputum sample should be replaced by an endotracheal aspirate sample if the patient is intubated.

Pneumococcus can also be detected by polymerase chain reaction (PCR), but this is not done routinely and there currently is no evidence that such testing will change clinical outcomes. Nevertheless, an increased pneumococcal bacterial load demonstrated by real-time PCR is associated with an increased risk for septic shock, need for mechanical ventilation, and death, suggesting that such a test may be of help in identifying patients suitable for care in the intensive care unit.[6] Pneumococcal capsular polysaccharide can be detected in urine, and this may occasionally be helpful for diagnosis after antibiotic therapy has been started. The overall sensitivity of the pneumococcal urinary antigen test is less than 80% but can reach 90% or higher in patients with pneumococcal bacteremia and those with high-risk pneumonia. The specificity in adult patients with CAP can exceed 95%. Detection of antibody to pneumococcal polysaccharide is not useful.

Newer methods which may help in the diagnosis of pneumococcal pneumonia include real-time polymerase chain reaction (RT-PCR) to detect pneumococcal DNA in blood and the use of host biomarkers such as C-reactive protein (CRP) and procalcitonin (PCT).[7] These are acute phase reactants which increase in the presence of an inflammatory response, particularly to bacterial pathogens. CRP may help to identify worsening disease, and PCT may help in determining the need for antibacterial therapy.

TABLE 289-2 INDICATIONS FOR MORE EXTENSIVE DIAGNOSTIC TESTING IN COMMUNITY-ACQUIRED PNEUMONIA

Intensive care unit admission for community-acquired pneumonia
Failure of outpatient treatment
Radiographic appearance of cavities on initial evaluation
Infection resulting in neutropenia
Alcohol abuse
Chronic severe liver disease
Severe chronic obstructive lung disease
Asplenia (anatomic or functional)
Recent travel (within 2 weeks)
Positive rapid *Legionella* spp. or pneumococcal urinary antigen test result
Pleural effusion

PREVENTION

Two types of pneumococcal vaccines are available, each with its own particular advantages and disadvantages.[8]

Polysaccharide Vaccine

A polysaccharide vaccine contains 25 μg of each of the 23 capsular polysaccharides that account for 90% of invasive infections (Chapter 18). Polysaccharide vaccines stimulate B-cell responses, thereby resulting in type-specific antibody production that enhances ingestion and killing of the pathogens by phagocytes. The antigens, however, are T-cell independent and therefore do not result in long-lasting immunity. Two types of polysaccharide vaccine have been available, Pneumovax (Merck) and Pnu-Imune (Lederle).

The effectiveness of the pneumococcal polysaccharide vaccine ranges from 56% to 81%. Data suggest that it is effective in preventing both pneumococcal and all-cause pneumonia. A randomized trial demonstrated that the immunogenicity of the 23-valent polysaccharide vaccine is comparable to that of the 7-valent pneumococcal conjugate vaccine (see later) in frail, hospitalized, elderly patients. It is not effective in immunocompromised patients, such as those with sickle cell disease, chronic renal failure, immunoglobulin deficiency, Hodgkin's disease, non-Hodgkin's lymphoma, leukemia, and multiple myeloma.. The vaccine is recommended for (1) persons 65 years or older; (2) persons 2 to 64 years old with chronic illnesses such as cardiovascular disease, chronic pulmonary disease (not asthma), diabetes mellitus, alcoholism, chronic liver disease, or cerebrospinal fluid leaks; (3) persons 2 to 64 years old with functional or anatomic asplenia; and (4) persons 2 to 64 years old living in special environments or social settings (Alaskan natives, certain Native American populations, residents of long-term care facilities).

Although the effectiveness of the vaccine is less in these subgroups, the following immunocompromised patients 2 years or older should also be immunized: (1) persons with HIV infection, leukemia, lymphoma, or Hodgkin's disease and (2) those with multiple myeloma, generalized malignancy,

chronic renal failure, nephrotic syndrome, or organ or bone marrow transplants and individuals being treated with immunosuppressive chemotherapy, including steroids.

The lack of an anamnestic response with polysaccharide vaccines means that antibody levels decrease over time, and revaccination is required. Although the exact timing is unclear, most experts suggest revaccination at 5 years. For immunocompetent persons 65 years or older, a second dose is suggested if the patient was given the first vaccine 5 years earlier at an age younger than 65 years. For persons 2 to 64 years of age with asplenia, a single revaccination is suggested 5 years after the initial dose if the patient is older than 10 years. However, if the patient is younger than 10 years, revaccination should be given 3 years after the first dose. For immunocompromised patients, revaccination should be given 5 years after the first dose if the patient is older than 10 years and 3 years after the first dose if the patient is younger than 10 years.

Conjugated Vaccine

This vaccine contains capsular polysaccharide from 13 of the most frequent pneumococcal pathogens affecting children linked to an immunogenic protein, thereby producing T cell–dependent antigens, which results in long-term immunologic memory. A heptavalent vaccine was effective in reducing the risk of pneumonia in young children. Use of a 7-valent vaccine resulted in an overall decrease in the prevalence of antimicrobial resistant pneumococci and in the incidence of invasive pneumococcal disease in both children and adults and with replacement with nonvaccine serotypes such as 19A and 35B. As a result, in 2010, the U.S. Food and Drug Administration licensed a new 13-valent pneumococcal conjugate vaccine to replace the 7-valent vaccine. Overall data demonstrate a superior antibody response in adults vaccinated with the 13-valent pneumococcal conjugate vaccine compared with the pneumococcal polysaccharide vaccine, whether or not they were previously vaccinated with the latter.[9] This vaccine is recommended for children, elderly adults, and younger immunocompromised patients.[10]

TREATMENT

A number of prediction rules have been developed to help determine the appropriate site of care. This step, together with optimal antimicrobial therapy and supportive measures, can help to maximize chances for a satisfactory outcome.[11]

Empirical therapy is usually used for CAP. If pneumococcal infection can be documented, however, antibiotic therapy can be targeted against this specific pathogen.

Treatment Regimens

Directed Therapy Against Known *S. pneumoniae*

S. pneumoniae can be treated with a number of antimicrobials, including various β-lactams, macrolides, and selected fluoroquinolones (Table 289-3). Resistance has been described for virtually all of these agents, and the prevalence of resistant serotypes has risen since the introduction of the 13-valent conjugate vaccine.[12] For nonmeningeal infections such as pneumonia, susceptibility is now defined by a penicillin MIC of 2 μg/mL or less, intermediate as an MIC of 4 μg/mL, and resistant as an MIC of more than 8 μg/mL. Multidrug-resistant pathogens, which are resistant to three or more antimicrobial agents with different mechanisms of action, have been seen in a number of countries. Risk factors for drug-resistant pneumococcal infection include recent antimicrobial therapy, age younger than 2 years or older than 65 years, attendance at a daycare center, recent hospitalization, and HIV infection.

If a good-quality Gram stain of sputum reveals sheets of PMNs with lancet-shaped gram-positive diplococci as the only organism, if the patient has no risk factors for infection with resistant *S. pneumoniae,* and if the patient is not living in an area endemic for penicillin-resistant *S. pneumoniae,* it is reasonable to initiate parenteral treatment with penicillin G, 2.4 million U/day. For outpatients, oral therapy is usually given in the form of amoxicillin, 500 mg three times daily. For infection with *S. pneumoniae* with penicillin MICs of up to 1 μg/mL, penicillin is still an appropriate agent. For strains with penicillin MICs of 2 to 4 μg/mL, data regarding efficacy are conflicting, and higher doses may be better. If the patient is allergic to penicillin, a macrolide such as azithromycin (500 mg on day 1; then 250 mg/day) or clarithromycin (500 mg twice daily) may be used. Monotherapy with a macrolide should not be used in areas where there is a high prevalence of macrolide-resistant pneumococci.

Five days is the minimal recommended duration of treatment of CAP that is not complicated, severe, or associated with bacteremia.

For patients who are hospitalized because of CAP, higher doses of parenteral antibiotics are generally used, including penicillin in the range of 12 million U (2-3 million every 4 hours) or ampicillin, 4 g/day (1 g every 6 hours). Third-generation cephalosporins such as ceftriaxone (1-2 g every 24 hours) and cefotaxime (1-2 g every 8 hours) are other alternatives.

TABLE 289-3 ANTIMICROBIAL DOSES AND FREQUENCY OF ADMINISTRATION

ANTIMICROBIAL	DOSE AND ROUTE	FREQUENCY
MACROLIDES		
Azithromycin	500 mg × 1; then 250 mg PO	q24h
Clarithromycin	500 mg PO	q12h
Erythromycin	500 mg PO	q6h
TETRACYCLINE		
Doxycycline	100 mg PO	q12h
FLUOROQUINOLONES		
Moxifloxacin	400 mg PO or IV	q24h
Gemifloxacin	320 mg PO	q24h
Levofloxacin	750 mg PO or IV	q24h
β-LACTAMS		
Amoxicillin	1 g PO	q8h
Amoxicillin–clavulanate	2 g PO	q12h
Ampicillin–sulbactam	2 g IV	q6h
Cefepime	1-2 g	q12h
Cefixime	400 mg PO	q12h
Cefotaxime	1-2 g IV	q8h
Ceftriaxone	1-2 g IV	q24h
Cefpodoxime	100-200 mg	q12h
Cefuroxime	500 mg PO	q12h
Ertapenem	1 g IV	q24h
Imipenem	500 mg IV	q6h
Meropenem	1 g IV	q8h
Piperacillin–tazobactam	3.375 g IV	q6h
MISCELLANEOUS		
Linezolid	600 mg PO or IV	q12h
Vancomycin	0.75-1 g IV	q12h

IV = intravenous; PO = oral; q = every. Five days is the minimal recommended duration of treatment of community-acquired pneumonia that is not complicated, severe, or associated bacteremia.

The patient can be switched to an oral regimen when cough and shortness of breath are improving, the white blood cell count is normalizing, and oral intake and gastrointestinal tract absorption are adequate. Febrile patients are frequently treated outside the hospital, so absence of a fever is not a prerequisite for switching to an oral regimen.

Some studies have suggested that initial combination antimicrobial treatment of bacteremic pneumococcal pneumonia that includes a macrolide is associated with lower mortality rates possibly because of an immunomodulatory effect of the macrolides.

Empirical Treatment

This is discussed in the chapter dealing with CAP (Chapter 97). The body of evidence supporting the positive effects of guidelines on outcomes in CAP patients is compelling, particularly for those ill enough to require hospitalization.[13]

PROGNOSIS

In an otherwise well, relatively young patient with no comorbid conditions and with mild to moderate infection, the elevated temperature and white blood cell count usually resolve by days 2 to 4 and 4, respectively. The patient looks and feels better within a few days, but it is important to keep in mind that even in patients younger than 50 years, only 60% of cases will have resolved radiologically by 1 month. In patients older than 50 years or those with more severe infection or COPD, only 25% may clear radiographically by 1 month.

A patient who fails to respond or deteriorates after initial treatment must be carefully reassessed with a detailed review of the history and treatment course plus appropriate radiographic studies and cultures.[14] If the diagnosis is incorrect, other infectious causes of pneumonia, such as *Haemophilus influenzae* (Chapter 300) or the atypical agents (Chapter 97), must be considered. Noninfectious illnesses must also be considered; these include heart failure (Chapter 58), pulmonary embolism (Chapter 98), pulmonary neoplasm (Chapter 191), radiation injury (Chapter 20), drug reaction (Chapter 254), and inflammatory lung disease, to name a few. If the original diagnosis was correct, metastatic infection, lung abscess (Chapter 90) or empyema, and unsuspected drug resistance must be considered. Drug factors such as errors

in selection, dose, or route of administration are possible explanations, especially in patients who receive oral medication.

There also appears to be an increased risk of cardiovascular events after certain respiratory infections, including pneumococcal pneumonia. This may be attributable to an enhanced inflammatory state and its effects on coronary arteries.[15]

Grade A References

A1. Nicholson KG, Abrams KR, Batham S, et al. Randomised controlled trial and health economic evaluation of the impact of diagnostic testing for influenza, respiratory syncytial virus and Streptococcus pneumoniae infection on the management of acute admissions in the elderly and high-risk 18- to 64-year-olds. *Health Technol Assess*. 2014;18:1-274.

A2. Maruyama T, Taguchi O, Niederman MS, et al. Efficacy of 23-valent pneumococcal vaccine in preventing pneumonia and improving survival in nursing home residents: double blind, randomized and placebo control trial. *BMJ*. 2010;340:c1004.

A3. Macintyre CR, Ridda I, Gao Z, et al. A randomized clinical trial of the immunogenicity of the 7-valent pneumococcal conjugate vaccine compared to 23-valent polysaccharide vaccine in frail, hospitalized elderly. *PLoS ONE*. 2014;9:e94578.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Blasi F, Montero M, Santus P, Tarsia P. Understanding the burden of pneumococcal disease in adults. *Clin Microbiol Infect*. 2012;18(suppl 5):7-14.
2. Feldman C, Anderson R. Bacteraemic pneumococcal pneumonia. *Drugs*. 2011;71:151-163.
3. Vermatter J, Pirofsky LA. Current concepts in host-microbe interaction leading to pneumococcal pneumonia. *Curr Opin Infect Dis*. 2013;26:277-283.
4. Henriques-Normark B, Tuomanen EI. The pneumococcus: epidemiology, microbiology, and pathogenesis. *Cold Spring Harb Perspect Med*. 2013;3:a010215.
5. File T. New diagnostic tests for pneumonia: what is their role in clinical practice? *Clin Chest Med*. 2011;105:1776-1783.
6. Waterer G, Rello J. Why should we measure bacterial load when treating community-acquired pneumonia? *Curr Opin Infect Dis*. 2011;24:137-141.
7. Johansson N, Kalin M, Backman-Johansson C, et al. Procalcitonin levels in community-acquired pneumonia—correlation with aetiology and severity. *Scand J Infect Dis*. 2014;46:787-791.
8. Musher D. How effective is vaccination in preventing pneumococcal disease? *Infect Dis Clin North Am*. 2013;27:229-241.
9. Feldman C, Anderson R. Review: current and new generation pneumococcal vaccines. *J Infect*. 2014;69:309-325.
10. Tomczyk S, Bennett NM, Stoecker C, et al. Use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine among adults aged >/=65 years: recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb Mortal Wkly Rep*. 2014;63:822-825.
11. Wiemken T, Kelley R, Ramirez J. Clinical scoring tools: which is best to predict clinical response and long-term outcomes? *Infect Dis Clin North Am*. 2013;27:33-48.
12. Richter SS, Diekema DJ, Heilmann KP, et al. Changes in pneumococcal serotypes and antimicrobial resistance after introduction of the 13-valent conjugate vaccine in the United States. *Antimicrob Agents Chemother*. 2014;58:6484-6489.
13. Johnstone J, Mandell L. Guidelines and quality measures: do they improve outcomes of patients with community-acquired pneumonia? *Infect Dis Clin North Am*. 2013;27:71-86.
14. Aliberti S, Blasi F. Clinical stability versus clinical failure in patients with community-acquired pneumonia. *Semin Respir Crit Care Med*. 2012;33:284-291.
15. Singanayagam A, Singanayagam D, Chalmers JD. Is community-acquired pneumonia an independent risk factor for cardiovascular disease? *Eur Respir J*. 2012;39:187-196.

REVIEW QUESTIONS

1. Pneumonia can be caused by a variety of mechanisms. The pathobiology of pneumococcal pneumonia initially depends on

A. colonization and inhalation of the organisms.
B. inhalation of the organisms.
C. colonization and aspiration of the organisms.
D. aspiration of the organisms.
E. spread from adjacent pleural spaces.

Answer: C The pathobiology of pneumococcal pneumonia initially depends on colonization and aspiration of the organisms. The "ecologic niche" of *Streptococcus pneumoniae* is the nasopharynx, and 20% of even healthy adults may carry the organism. Carrier rates are highest in late fall to early spring. Pathogens such as bacteria, viruses, and fungi can gain access to the lower respiratory tract and the distal airways by a number of mechanisms, including inhalation, aspiration (both gross and silent), spread from adjacent sites, and rarely by direct instillation as may occur with penetrating trauma. After colonization is established, the most common means of gaining access to the lower airways is by aspiration of the pneumococcus. This can occur even during deep sleep in normal individuals. One does not have to have a seizure or be drunk and vomiting for this to happen. Pneumococcal virulence factors can influence both the colonization process and any subsequent bacterial invasion that may occur.

2. Mr. S is a 46-year-old man in excellent general health. He has no known medical problems and is a nonsmoker, a social drinker, and an avid jogger. He noted the onset of a cough productive of purulent sputum with occasional traces of blood, some right-sided pleuritic type chest pain, and shortness of breath even while walking over the past several days. He has also experienced some chills and sweats in the last day or 2. On examination, he is febrile, his respiratory rate is 24 breaths/min, and his chest radiograph shows a right middle lobe lobar consolidation without an effusion. His white blood cell count is 16,000 with a left shift, and a Gram stain of sputum showed classic gram-positive lancet-shaped diplococci with numerous polymorphonuclear leukocytes in the field. No other organisms were noted. A decision is made to admit him to the hospital. He has no drug allergies. Treatment should be started with

A. a macrolide.
B. penicillin.
C. a fluoroquinolone.
D. ceftriaxone.
E. macrolide plus ceftriaxone.

Answer: B Treatment should start with penicillin. Unfortunately, the information gained from the history, physical examination, and chest radiograph are not necessarily definitively helpful in accurately predicting the etiologic pathogen. In this case, it certainly sounds as though a diagnosis of pneumonia is correct (if pulmonary embolism has been ruled out), there are no obvious risk factors for pathogens such as *Legionella* spp., and gram-negative rods would be quite unlikely. The finding of gram-positive cocci consistent with the pneumococcus and the absence of any other pathogen seen on Gram stain argues for a pneumococcal etiology. Given his overall presentation and the results obtained so far, it would be reasonable to begin treatment with a penicillin. He does not have risk factors for drug-resistant *Streptococcus pneumoniae* (DRSP), such as recent antibiotic treatment, age older than 65 years, HIV infection, or recent hospitalization, and he does not live in an area where DRSP is endemic.

Having said this, some would argue that overall, it might be safer to initiate treatment with a macrolide and a nonpseudomonal third-generation cephalosporin or a fluoroquinolone alone for a patient requiring admission to the hospital, but not the intensive care unit. As soon as specific culture data become available (e.g., blood or sputum), treatment could be changed to penicillin. Given the specifics of this case, however, it certainly is reasonable to initiate treatment with penicillin alone.

Mandell LA, Wunderlink RG, Anzueto A, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis*. 2007; 44(Suppl):S27-S72.

3. A 20-year-old ,am who is in excellent health with no prior medical illnesses is involved in an accident while riding his motorcycle. He fractures his left femur and requires open reduction and internal fixation plus a splenectomy for his ruptured spleen. What would be the best way to prevent pneumococcal infection in this young man for the future?

A. Daily low doses of penicillin orally
B. Monthly intramuscular benzathine penicillin
C. Pneumococcal polysaccharide vaccine
D. Pneumococcal conjugate vaccine
E. Pneumococcal conjugate vaccine followed by pneumococcal polysaccharide vaccine

Answer: E Pneumococcal conjugate vaccine followed by pneumococcal polysaccharide vaccine would be used to prevent pneumococcal infection. There are currently two types of pneumococcal vaccines, a polysaccharide vaccine (PSV) and a protein conjugate vaccine (PCV). A vaccine would be the preferred method of protection by inducing immunity as opposed to relying on the use of prophylactic antibiotics. The problem is that in the case of a PSV, those who are immunosuppressed because of illness such as myeloma or lymphoma, immunosuppressing drugs, or transplantation are less likely to benefit from it. This includes patients who are either functionally or anatomically asplenic. For this 20-year-old accident victim who has never been given a pneumococcal vaccine before and is now at risk of pneumococcal infection because of his asplenic condition, he should be given a single dose of PCV13 followed at least 2 months later by a dose of PSV.

Musher D. How effective is vaccination in preventing pneumococcal disease? *Infect Dis Clin North Am*. 2013;27:229-241.

4. To diagnose pneumonia in general and pneumococcal pneumonia specifically, the best test is

A. bronchoscopy.
B. chest radiography.
C. urinary antigen.
D. biomarkers.
E. none of the above.

Answer: E Unfortunately, there is no one specific test that is definitively diagnostic of pneumonia and pneumococcal pneumonia. The diagnosis is made by taking a careful history, doing a physical examination, and then supplementing this with additional tests depending on the severity of illness and whether or not admission to the hospital is being considered. Chest radiography is usually the most helpful because if the results are negative, it usually tends to rule out pneumonia. The constellation of findings obtained from the history, physical examination, and radiography can certainly suggest pneumonia versus some other process such as pulmonary embolism, but they are not foolproof, and the pattern of infiltrate on radiographs does not signify a particular pathogen. Initial treatment can be started with broader spectrum agents and then modified depending on any additional information that becomes available (e.g., cultures of blood or sputum, urinary antigen test results).

5. Pneumococcal resistance to penicillin is the result of

A. a change in penicillin-binding proteins (PBPs).
B. an antibiotic pump mechanism.
C. an alternative metabolic pathway for the bacteria.
D. mutations in the *gyrA* or *parC* genes.
E. impaired access to the bacteria because of porin closure.

Answer: A Bacteria can protect themselves from antibiotics in a number of ways. To be effective, the antibiotic must be able to reach the target site within the pathogen, but this can be prevented by several mechanisms. These include:

1. An enzyme that inactivates the antibiotic
2. Decreased access to the bacterium via porin channels
3. Altered target sites so the antibiotic cannot recognize the target
4. Alternate metabolic pathways for the bacteria
5. Efflux pumps

In the case of β-lactam drugs such as the penicillins and the pneumococcus, the target sites are the penicillin-binding proteins, which are trans and carboxypeptidase enzymes involving bacterial cell wall synthesis. As the affinity of the drug for the target site decreases, it becomes harder for the drug to interact with the target and to induce any damage in the bacterial pathogen.

290 NONPNEUMOCOCCAL STREPTOCOCCAL INFECTIONS AND RHEUMATIC FEVER

DONALD E. LOW

THE PATHOGENS

The *Streptococcus* genus contains a number of species that inhabit a broad range of hosts, including humans and domesticated animals, where they often colonize as part of the normal flora and cause infection. They are gram-positive cocci that grow in chains. The two most useful methods of classifying streptococci are by the type of hemolytic reaction displayed on blood agar on which they are grown and the serologic reactivity of the cell wall polysaccharide antigens as originally described by Rebecca Lancefield (Table 290-1). On blood agar plates, streptococci may cause complete (β), incomplete (α), or no (γ) hemolysis. Hemolytic streptococci from humans can be classified into Lancefield groups A, B, C, F, G, and L on the basis of carbohydrate antigens of the cell wall and can be subdivided into large- and small-colony (<0.5 mm in diameter) formers. The β-hemolytic group A and group B streptococci (GBS) are considered the major pathogenic β-hemolytic streptococci. The β-hemolytic large-colony–forming species, *Streptococcus dysgalactiae* subspecies *equisimilis,* may possess Lancefield group A, C, G, or L antigens. The small-colony–forming β-hemolytic strains with Lancefield group A, C, F, or G or no group antigens belong to the *anginosus* (or previously termed *Streptococcus milleri*) group. Members of the *anginosus* group (*Streptococcus anginosus*, *Streptococcus constellatus*, and *Streptococcus intermedius*) are considered part of the viridans group streptococci (VGS), most of which display α-hemolytic or nonhemolytic reactions.

TABLE 290-1 IDENTIFICATION OF THE β-HEMOLYTIC STREPTOCOCCI

SPECIES	LANCEFIELD GROUP	COLONY SIZE	ORIGIN
S. pyogenes	A	Large	Human
S. agalactiae	B	Large	Human, bovine
S. dysgalactiae subsp. *equisimilis*	A, C, G, L	Large	Human, animals
S. equi subsp. *zooepidemicus*	C	Large	Animals, human
S. canis	G	Large	Dog, human
S. anginosus (group)*	A, C, G, F, none	Small	Human

*Group carbohydrate antigen.

STREPTOCOCCUS PYOGENES (GROUP A STREPTOCOCCUS)

The most clinically significant of the streptococcal human pathogens are *Streptococcus pyogenes* (also referred to as group A streptococci [GAS]), which contain the Lancefield group A antigen on their cell surface. GAS causes primarily infections of the upper respiratory tract and the skin. It is also responsible for a toxic shock–like syndrome; necrotizing fasciitis; and delayed nonsuppurative sequelae, including acute rheumatic fever and poststreptococcal glomerulonephritis. The major virulence factor of the organism is the M protein. This protein, a stable dimer, is anchored to the cell membrane and traverses and penetrates the cell wall. Whereas the proximal portion of the molecule is highly conserved among GAS, the distal portions contain type-specific epitopes localized on the tips that protrude from the cell surface. This variation in the distal region of the M proteins provides the basis of widely used epidemiologic typing schemes that use serologic methods (M type) or nucleotide sequence analysis of the M protein gene (*emm* type). There are more than 150 distinct *emm* and *emm*-like genes recognized. Epidemiologic studies have revealed that certain disease manifestations are commonly associated with particular M types, such as M1 and M3 types, which are associated with the most severe invasive manifestations such as toxic shock syndrome and necrotizing fasciitis. Humans are the natural host, and there is no animal or environmental reservoir. The throat and skin of the human host are the principal reservoirs for GAS. GAS is transmitted primarily person to person by either contact or droplet transmission.

EPIDEMIOLOGY

All GAS infections have the highest incidence in children younger than 10 years. The asymptomatic prevalence is also higher (15%-20%) in children than in adults. Age is not the only factor; crowded conditions in temperate climates during the winter months are also associated with epidemics of pharyngitis in school children as well as in military recruits. The peak incidence for streptococcal pharyngitis and impetigo varies with season and locale. In many temperate regions of the world, the incidence of pharyngitis peaks in winter; impetigo, although less common, peaks during the summer months. In contrast, for many tropical host populations, impetigo is hyperendemic year round, but throat infection, whether pharyngitis or asymptomatic carriage, ranges from very low to moderate levels.

CLINICAL MANIFESTATIONS

Respiratory Tract Infections

Pharyngitis

Group A streptococci are a major cause of pharyngitis and remains the only agent of this syndrome requiring etiologic diagnosis and treatment. The burden and economic costs of GAS pharyngitis are great. It has been estimated that in the United States alone, more than 7 million cases of acute pharyngitis are diagnosed by pediatricians annually. *S. pyogenes* is the cause in only 15% to 30% of them, but antibiotics are prescribed in 55% to 75% of the cases. Although a major consequence of GAS pharyngitis, acute rheumatic fever (see later), is much less common now than in the past, it is still a considerable problem in the developing world. The World Health Organization (WHO) estimates that there are about half a million cases of acute rheumatic fever worldwide annually. On clinical grounds, streptococcal pharyngitis (Fig. 290-1) is strongly suggested by the presence of fever, tonsillar exudate, tender enlarged anterior cervical lymph nodes, and absence of cough.

Because these findings are nonspecific and are also commonly found in cases of viral origin, however, even experienced physicians may accurately diagnose streptococcal pharyngitis based on the clinical findings alone in no more than 75% of the cases. Treatment of suspected GAS pharyngitis at the time of initial clinical evaluation has only a modest effect on relieving acute symptoms and perhaps on preventing suppurative complications. However, antibacterial treatment may be delayed for several days and still achieve the goal of preventing rheumatic fever and spread of disease. Prospective studies to compare the impact of various pharyngitis management strategies on clinically relevant outcomes have been recommended. A particular focus of recent guidelines has been to reduce overall use of antibiotics for treatment of pharyngitis in both children and adults in order to limit antibiotic resistance (Table 290-2).

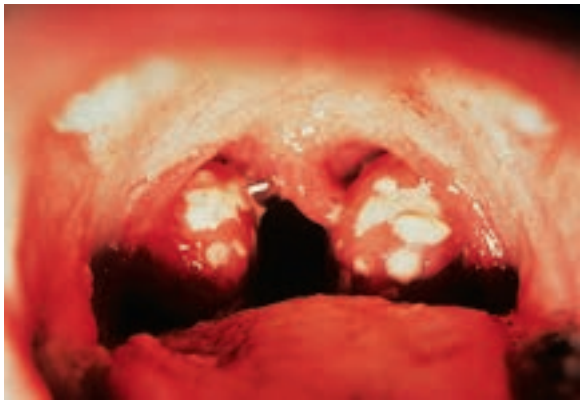

FIGURE 290-1. Acute streptococcal pharyngitis. Pus is present in the tonsillar crypts, and some palatal petechiae are seen. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine*. 3rd ed. London: Mosby; 2003).

TABLE 290-2 SCORE FOR USE IN BOTH CHILDREN AND ADULTS WITH SORE THROAT TO ESTIMATE PROBABILITY OF POSITIVE CULTURE AND GUIDE MANAGEMENT APPROACH

DIAGNOSTIC CRITERIA		*MANAGEMENT*		
FINDINGS	**POINTS**	**SCORE**	**RISK FOR STREPTOCOCCAL INFECTION (%)**	**RECOMMENDATION**
Temperature >38°C	+1	≤0	1-2.5	No culture or antibiotics
Absence of cough	+1	1	5-10	
Swollen, tender anterior cervical nodes	+1	2	11-17	Culture: antibiotics only if culture positive
Tonsillar swelling or exudate	+1	3	18-35	
Age		≥4		Antibiotics
3-14 yr	+1			
>44 yr	−1			

Adapted from McIsaac WJ, Kellner JD, Aufricht P, et al. Empirical validation of guidelines for the management of pharyngitis in children and adults. *JAMA*. 2004;291:1587-1595.

Throat culture is the conventional method for establishing the diagnosis of GAS pharyngitis. In an untreated patient with GAS pharyngitis, results of a properly obtained throat culture (by vigorous swabbing of both tonsils and posterior pharynx) are almost always positive; however, a positive throat culture may reflect chronic colonization by GAS, and the acute illness may be caused by another agent. Quantitation of GAS from the throat swab culture cannot be used to differentiate carriage from infection because sparse growth may be associated with true infection. A negative throat culture result permits the physician to withhold antibiotic therapy from most patients with sore throats.

Many GAS antigen detection tests are available commercially. These tests vary in method. Most of these tests have a high degree of specificity, but their sensitivity in clinical practice can be unacceptably low. Therefore, treatment is indicated for patients with acute pharyngitis who have a positive rapid antigen detection test result.

Otitis Media and Rhinosinusitis

During the past several decades, a number of studies have reported isolation of GAS from 2% to 5% of cultures of middle ear fluid specimens obtained from children with acute otitis media. Thus, in contrast to the very striking role of GAS as the major bacterial agent of acute pharyngitis, GAS has been fairly consistently the fourth most predominant pathogen causing pediatric acute otitis media, after *Streptococcus pneumoniae, Haemophilus influenzae,* and *Moraxella catarrhalis.* GAS acute otitis media is associated with increased risk for development of mastoiditis. Although the risk is small (<1%), it is much greater than with acute otitis media caused by *S. pneumoniae, H. influenzae,* or *M. catarrhalis.* GAS is the etiologic agent of acute bacterial rhinosinusitis in 2% to 7% of cases.

Pneumonia

The occurrence of pneumonia has increased with the resurgence of invasive GAS disease during the past several decades, with 10% of patients with invasive GAS disease presenting with pneumonia. Small outbreaks of GAS pneumonia have been described in chronic care facilities and within families, as well as sporadic cases occurring in the community. GAS pneumonia now occurs with a frequency similar to that of other well-recognized causes, such as *Staphylococcus aureus* or *Klebsiella pneumoniae*. A Canadian population-based surveillance program of invasive GAS disease confirmed that GAS pneumonia is a severe illness of sudden onset frequently associated with local and systemic complications, particularly empyema (19%), toxic shock (32%), and death (38%).

Skin and Soft Tissue Infections

Scarlet Fever

Scarlet fever is a diffuse erythematous eruption that generally occurs in association with pharyngitis, most commonly in children 5 to 15 years of age. The development of the scarlet fever rash requires prior exposure to *S. pyogenes* and occurs as a result of delayed-type skin reactivity to pyrogenic exotoxin (erythrogenic toxin, usually types A, B, or C) produced by the organism. The rash of scarlet fever is a diffuse erythema that blanches with pressure, with numerous small (1-2 mm) papular elevations, giving a sandpaper quality to the skin. It usually starts on the head and neck and is accompanied by circumoral pallor and a strawberry tongue. Subsequently, the rash expands rapidly to cover the trunk followed by the extremities and then ultimately desquamates; the palms and soles are usually spared. The rash is most marked in the skinfolds of the inguinal, axillary, antecubital, and abdominal areas and about pressure points. It often exhibits a linear petechial character in the antecubital fossae and axillary folds, known as Pastia's lines.

Erysipelas

Erysipelas is an acute, superficial, non-necrotizing dermal or hypodermal infection that is mainly caused by streptococci. The definitive diagnosis is based on clinical findings that usually include a sharply demarcated shiny erythematosus plaque associated with pain, swelling, and fever (see Fig. 441-1). Erysipelas affects predominantly adult patients in the sixth or seventh decade of life and is located on the lower limb in more than 80% of cases. A female predominance exists, except in young patients. Risk factors include disruption of the cutaneous barrier (leg ulcer, wound, fissured toe-web intertrigo, and pressure ulcer), lymphedema, chronic edema, or local surgical operations (lymph node dissection, saphenectomy). Toe-web intertrigo appears to be a major portal of entry whether or not due to dermatophytes. Erysipelas is less commonly caused by group B, C, or G streptococci and rarely by staphylococci. Bulla formation is considered as a relatively severe but frequent local complication of the disease. Although most cases of erysipelas are caused by β-hemolytic streptococci, many other bacteria can produce non-necrotizing cellulitis, which can often occur in particular circumstances, such as *Pasteurella multocida* after cat or dog bites, *Aeromonas hydrophila* after immersion in fresh water, *Vibrio* spp. after saltwater exposure, or *H. influenzae* in periorbital cellulitis in children. Recurrence is the main complication of erysipelas; it occurs in about 20% of cases. Measures to reduce recurrences of erysipelas include treatment of any predisposing factor, such as toe-web intertrigo or wound, and reducing any underlying edema. If frequent infections occur despite such measures, prophylactic antibiotics may be warranted.

Impetigo

Impetigo is a highly contagious infection of the superficial epidermis that most often affects children 2 to 5 years of age (see Fig. 441-1), although it can occur in any age group. Impetigo is classified as bullous or nonbullous impetigo; the latter is the most common form. Bullous impetigo simply means that the skin eruption is characterized by bullae (blisters). The infection usually heals without scarring, even without treatment. *S. aureus* is the most important causative organism, especially in bullous impetigo. GAS causes fewer cases, either alone or in combination with *S. aureus,* and is more often found in nonbullous impetigo. The diagnosis usually is made clinically

and can be confirmed by Gram stain and culture, although this is not usually necessary. Culture may be useful to identify patients with nephritogenic strains of GAS during outbreaks of poststreptococcal glomerulonephritis. Impetigo usually is transmitted through direct contact. Patients can further spread the infection to themselves or others after excoriating an infected area. Infections often spread rapidly through schools and daycare centers.

The Cochrane Collaboration reviewed interventions for impetigo. There was little evidence found for the use of disinfecting measures, such as chlorhexidine. There was good evidence that topical mupirocin and topical fusidic acid are equally or more effective than oral antibiotics for people with limited disease. Fusidic acid and mupirocin were of similar efficacy. It was found to be unclear whether oral antibiotics are superior to topical antibiotics for people with extensive impetigo.

Cellulitis

Bacterial cellulitis refers to a diffuse, spreading skin infection. Associated regional lymphadenopathy and lymphatic streaking are variable, and local complications (abscesses, necrosis) are more frequent than in erysipelas. Petechiae and ecchymoses with frequent bullae may develop in inflamed skin, resulting in hemorrhagic cellulitis. *Cellulitis* usually refers to a more deeply situated skin infection than erysipelas. However, the distinction between these entities is not clear-cut, and the two conditions share the typical clinical features, including sudden onset, usually with a high fever, and the tendency to recur. GAS has been considered the main causative agent of cellulitis, although groups B, C, and G streptococcus and *S. aureus* can also be a cause.

The predominant infection site for cellulitis is on the lower extremities. Lymphedema and disruption of the cutaneous barrier, which serves as a site of entry for the pathogens, are risk factors for infections. About 20% to 30% of patients have a recurrence during a 3-year follow-up period. Blood cultures are positive for β-hemolytic streptococci in fewer than 5% of cases.

Invasive Group A Streptococcus Disease

Invasive disease is defined as the isolation of GAS from an otherwise sterile site. GAS is capable of producing a variety of bacterial exotoxins, including a family of toxins known as superantigens (SAg). Severe invasive GAS disease includes streptococcal toxic shock syndrome and necrotizing fasciitis. *S. anginosus* plays a significant role as a reservoir of antimicrobial resistance genes, transferring different resistance traits to more pathogenic organisms such as *S. pneumoniae* and *S. pyogenes*. Both manifestations are associated with high morbidity and mortality.

Toxic Shock Syndrome

Streptococcal *toxic shock syndrome* is defined as hypotension accompanied by multiple organ failure, indicated by two of the following signs: renal impairment, coagulopathy, liver involvement, acute respiratory distress syndrome, a generalized rash, and soft tissue necrosis. In contrast to the other major form of toxic shock syndrome that is caused by toxin-producing *S. aureus* (Chapter 288), in which blood cultures are positive in fewer than 5% of cases, most (60%) patients with streptococcal toxic shock syndrome have positive blood cultures. In the late 1980s, patients with severe GAS infections were characterized and reported. Most patients were younger than 50 years old and otherwise healthy. All had invasive GAS infections characterized by signs that included shock; multiorgan system involvement; and rapidly progressive, destructive soft tissue infection (necrotizing fasciitis). The case-fatality rate was 30% even though most patients received appropriate antimicrobial therapy. M-types 1 and 3 were the most common type, and 80% of the isolates produced pyrogenic exotoxin A.

A focus of infection in staphylococcal toxic shock syndrome, if present at all, tends to be superficial, a complication of surgical wounds or burns or a foreign body (e.g., tampons). Streptococcal toxic shock syndrome can originate from an unknown focus or from a deep-seated soft tissue infection (e.g., necrotizing fasciitis, myositis, and cellulitis). Mortality rates are much higher with streptococcal than staphylococcal toxic shock, reported at up to 80% when associated with myositis.[1]

Necrotizing Fasciitis

Necrotizing fasciitis is defined by infection of the subcutaneous tissue and fascia that often results in necrosis with relative sparing of the underlying muscle. Histopathology demonstrates both necrosis of superficial fascia and polymorphonuclear infiltrates as well as edema of the reticular dermis, subcutaneous fat, and superficial fascia. GAS organisms in this disorder secrete proteases that disrupt host tissue and exotoxins that can cause local tissue injury or systemic manifestations as the result of an excessive and inappropriate inflammatory response, leading to a life-threatening infection.

In the early stages of necrotizing fasciitis caused by GAS, the clinical findings can be nondescript, with as many as one third of patients initially being given a diagnosis other than necrotizing fasciitis. The only feature that might alert the clinician is a complaint of pain out of proportion to that expected on examination, even for patients who report previous trauma to the area. Necrotizing fasciitis has been found to occur more frequently during the winter months and in older men. More then half of patients give a history of an antecedent skin lesion or injury at the site of infection. In adults, most cases occur in persons with at least one chronic underlying illness, and in children, the most frequent underlying illness is an infection caused by varicella-zoster virus (Chapter 375). The most common primary site of infection is the lower extremity followed by the upper extremity, trunk, and groin or perineum.

Streptococcal exotoxins, which are superantigens, have been recognized as central mediators of the systemic effects associated with severe GAS infections.[2] Superantigens interact with antigen-presenting cells and T cells to induce T-cell proliferation and massive cytokine production, which leads to fever, rash, capillary leak, and subsequent hypotension, the major symptoms of severe GAS disease (Fig. 290-2). Twelve distinct superantigens produced by GAS have been identified, and many GAS strains harbor genes encoding four to six different superantigens. The magnitude of the inflammatory response is determined by the patient's human leukocyte antigen class II type. Another important risk factor for invasive GAS infections is lack of opsonic anti-M1 protein and neutralizing antisuperantigen antibodies.

Treatment of either streptococcal toxic shock or necrotizing fasciitis has not been critically studied because of the low incidence of disease and difficulty in patient recruitment for clinical trials. The dogma for the management GAS necrotizing fasciitis is that when necrotizing fasciitis is suspected, early surgical debridement is warranted. Despite the lack of clinical or scientific evidence to support this approach, it is widely endorsed in standard textbooks and treatment guidelines. The published studies are all based on retrospective chart review of patients with necrotizing fasciitis from multiple etiologies, including mixed aerobic and anaerobic organisms.

The finding that lack of protective antibodies against streptococcal M protein and superantigens correlates with risk for developing invasive streptococcal diseases, highlights the importance of antibodies in protection against these infections and suggests that intravenous immunoglobulin (IVIG) might be a potential adjunctive therapy. IVIG exhibits high polyspecificity generated by antibodies pooled from several thousands of donors and has been shown to contain broad-spectrum antibodies against streptococcal superantigens and M proteins. In addition, IVIG has a general anti-inflammatory effect that is largely attributable to Fc receptor–mediated mechanisms. The documentation of clinical efficacy of IVIG includes several case reports and two observational cohort studies, one case-control study, and one multicenter placebo-controlled trial.

Nonsuppurative Sequelae of Streptococcal Infections

Acute Rheumatic Fever

Acute rheumatic fever is a delayed, nonsuppurative sequela of a pharyngeal infection with the GAS. After the initial pharyngitis, a latent period of 2 to 3 weeks occurs before the first signs or symptoms of acute rheumatic fever appear. The disease presents with various manifestations that may include arthritis, carditis, chorea, subcutaneous nodules, and erythema marginatum. In developing areas of the world, acute rheumatic fever and rheumatic heart disease are estimated to affect nearly 20 million people and are the leading causes of cardiovascular death during the first five decades of life. The story is much different in developed countries. In the latter half of the 20th century, rheumatic fever receded as an important health problem in almost all wealthy countries. Today, most physicians in these countries are unlikely ever to see a case of acute rheumatic fever. Most of the reduction is attributable to improved living conditions, which have resulted in less overcrowding and better hygiene, with consequent reductions in transmission of group A streptococci. In other words, rheumatic fever is now a disease of poverty.

Worldwide, at least 350,000 deaths are attributable to rheumatic fever or rheumatic heart disease each year; most occur in developing countries and among indigenous groups.[3] The observation in some studies that only a few M serotypes (types 3, 5, 6, 14, 18, 19, 24, and 29) were implicated in outbreaks of rheumatic fever suggested a particular "rheumatogenic" potential of certain strains of GAS. The decrease in the incidence of acute rheumatic fever in developed countries with the replacement of rheumatogenic types by

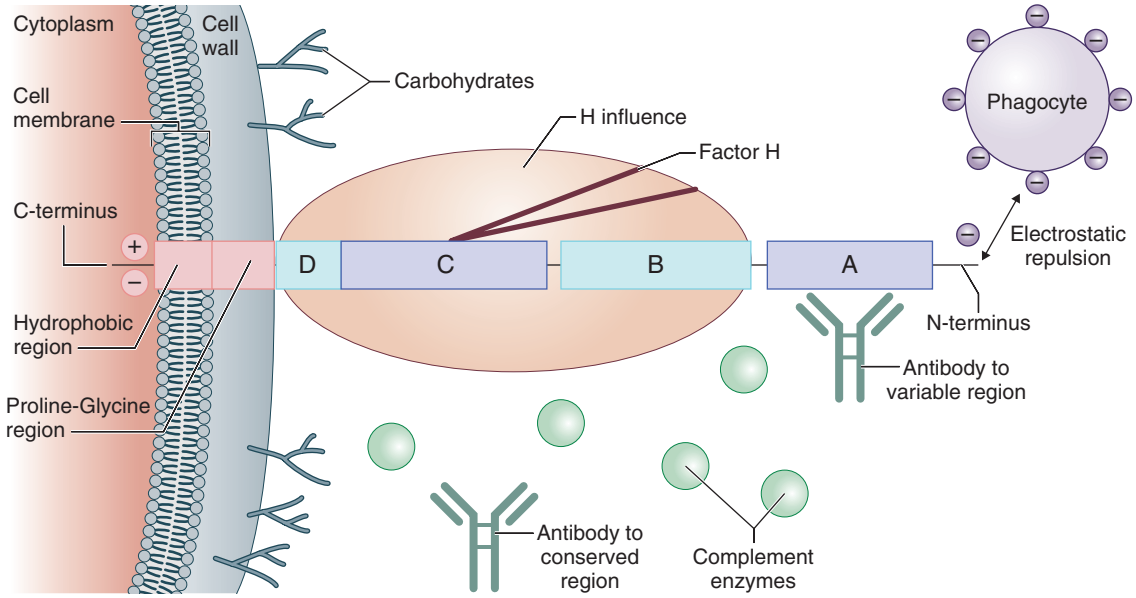

FIGURE 290-2. Thwarting the immune system is the primary job of the M protein. Negative charges at the N-terminus may repel phagocytic white blood cells. By binding with factor H—a regulatory protein produced by the human host—the M protein protects its most conserved regions from antibodies and complement enzymes. Only antibodies against the antigenically shifting hypervariable region can clear an established streptococcal infection from the host's body. (Reprinted from Fischetti VA. Streptococcal M protein. *Sci Am*. 1991;264(6):58-65; with permission from Mr. Tomoyuki Narashima from Tane+1 LLC for reproduction of illustration).

nonrheumatogenic types is not solely responsible for the decrease in acute rheumatic fever, and the issue of potential rheumatogenic strains remains unresolved. A streptococcal strain capable of causing a well-documented pharyngitis almost always is potentially capable of causing rheumatic fever. The lack of specific rheumatogenic strains also can explains the relatively high risk for recurrent disease with new streptococcal infections, in contrast to poststreptococcal glomerulonephritis, in which only a few "nephritogenic" strains appear to be capable of inducing the disease (e.g., type 12 with pharyngitis and type 49 with impetigo), and recurrent disease is uncommon.

Modified Jones criteria were first published in 1944 to be able to define persons with acute rheumatic fever. They have been periodically revised by the American Heart Association in collaboration with other groups.[4] According to revised Jones criteria, the diagnosis of rheumatic fever can be made when two of the major criteria or one major criterion plus two minor criteria are present along with evidence of streptococcal infection. Exceptions are chorea and indolent carditis, each of which by itself can indicate rheumatic fever. Major criteria include the following:

1. Migratory polyarthritis: a temporary migrating inflammation of the large joints, usually starting in the legs and migrating upward
2. Carditis: inflammation of the heart muscle, which can manifest as congestive heart failure, pericarditis with a rub, or a new heart murmur
3. Subcutaneous nodules: painless, firm collections of collagen fibers over bones or tendons, commonly appearing on the back of the wrist, the outside elbow, and the front of the knees
4. Erythema marginatum: a long-lasting rash that begins on the trunk or arms as macules and spreads outward to form a snakelike ring while clearing in the middle. This rash never starts on the face, and it is made worse with heat.
5. Sydenham's chorea (St. Vitus' dance): a characteristic series of rapid movements of the face and arms without purpose, generally occurring late in the disease

Minor criteria include the following:

1. Fever
2. Arthralgia: joint pain without swelling
3. Raised erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP)
4. Leukocytosis
5. Electrocardiogram showing features of heart block, such as a prolonged PR interval
6. Supporting evidence of streptococcal infection: elevated or rising antistreptolysin O titer or anti-DNAse B
7. Previous episode of rheumatic fever or inactive heart disease

The pathogenic mechanisms that lead to the development of acute rheumatic fever remain incompletely understood. Clearly, streptococcal pharyngeal infection is required, and genetic susceptibility may be present. On the other hand, evidence is sparse that toxins produced by the streptococcus are important. Molecular mimicry is thought to play an important role in the initiation of the tissue injury. As a result of molecular mimicry, antibodies directed against GAS antigens cross-react with host antigens. In addition to the role of antibody, observations suggest a role for cellular immunity in molecular mimicry.

Primary prevention of acute rheumatic fever is accomplished by proper identification and adequate antibiotic treatment of GAS tonsillopharyngitis.[5,6] The diagnosis of GAS pharyngitis is best accomplished by combining clinical judgment with diagnostic test results, the criterion standard of which is the throat culture. Penicillin (either oral penicillin V or injectable benzathine penicillin) is the treatment of choice because it is cost effective, has a narrow spectrum of activity, and has long-standing proven efficacy. In addition, GAS resistant to penicillin has not been documented. For penicillin-allergic individuals, acceptable alternatives include a narrow-spectrum oral cephalosporin, oral clindamycin, or various oral macrolides. An individual who has had an attack of rheumatic fever is at very high risk of developing recurrences after subsequent GAS pharyngitis and needs continuous antimicrobial prophylaxis to prevent such recurrences (secondary prevention). The recommended duration of prophylaxis depends on the number of previous attacks, the time elapsed since the last attack, the risk for exposure to GAS infections, the age of the patient, and the presence or absence of cardiac involvement. Penicillin is again the agent of choice for secondary prophylaxis, but sulfadiazine or a macrolide is an acceptable alternative in penicillin-allergic individuals.

Poststreptococcal Reactive Arthritis

The term *poststreptococcal reactive arthritis* (PSRA) was first used to describe patients with arthritis after documented GAS infection but who failed to meet the Jones criteria for the diagnosis of acute rheumatic fever. Since then, the differentiation of PSRA from acute rheumatic fever has remained unsettled.[7] Compared with acute rheumatic fever, the arthritis of PSRA is more likely to occur within 10 days after infection, be symmetrical and nonmigratory, last longer than 3 weeks, and be poorly responsive to aspirin or other nonsteroidal anti-inflammatory drugs (NSAIDs). Periarticular findings, such as tenosynovitis, have not been considered to be part of the clinical spectrum of acute rheumatic fever but have been described with PSRA. However, some studies have reported that an additive and prolonged pattern of arthritis, as well as aspirin unresponsitivity, may be seen in 19% to 36% of patients with acute rheumatic fever. Some have suggested that patients with higher ESR and CRP, a shorter duration of joint symptoms after NSAID therapy, and without relapse when NSAIDs were stopped were diagnosed as having acute rheumatic fever rather than PSRA. Although all patients with PSRA have

serologic evidence of a recent GAS infection, no more than half of these patients who have a throat culture performed have GAS isolated. Because a small proportion of patients with PSRA have been reported to develop subsequent valvular heart disease, these patients should be observed carefully for several months for clinical evidence of carditis. If valvular disease is detected, the patient should be classified as having had acute rheumatic fever and should continue to receive secondary prophylaxis.

Poststreptococcal Glomerulonephritis

Acute poststreptococcal glomerulonephritis (PSGN) is a disease characterized by the sudden appearance of edema, hematuria, proteinuria, and hypertension (Chapter 121).[8] It is now known to follow infection by nephrogenic strains of GAS, which are M types 1, 2, 4, 12, 18, 25, 49, 55, 57, and 60. These may cause skin or throat infections, but specific M types, such as 49, 55, 57, and 60, are most commonly associated with skin infections. In addition, nontypeable GAS is frequently isolated from the skin or throats of patients with glomerulonephritis, representing presumably unclassified nephritogenic strains. The overall risk of developing acute PSGN after infection by these nephritogenic strains is about 15%. The risk for nephritis may also be related to the M type and the site of infection. The risk of developing nephritis infection by M type 49 is 5% if it is present in the throat. This risk increases to 25% if infection by the same organism is present in the skin. The time lapse between infection and onset of PSGN symptoms is typically 1 to 3 weeks after a throat infection and 3 to 6 weeks after a skin infection.

The pathogenic mechanism of PSGN remains largely unknown (Chapter 121). Diffuse glomerular hypercellularity, C3 and immunoglobulin G (IgG) deposition, and proteinuria and hematuria are cardinal features. However, it can occur without all of these signs present concurrently. For a reliable diagnosis of PSGN, clear evidence that a streptococcal infection preceded the glomerulonephritis is required. However, streptococcal infection has often cleared by the time the patient appears in the clinic because the symptoms generally arise 1 to 3 weeks after the patient was infected. Acute PSGN most often follows upper respiratory tract GAS infections in colder climates and skin infections in warmer climates. Occasionally, it may also be seen after group C or G streptococcal infections. A high incidence of PSGN is noted in other parts of the world, especially in areas with tropical climates where skin infections are common. Epidemic episodes tend to occur in highly populated areas where poor hygiene, malnutrition, anemia, and parasites are common, and episodes are separated by periods of 5 to 7 years in certain communities. Epidemics occur also in the Western world, especially in closed communities. Acute PSGN affects all age groups but most commonly children and young adults. Males are affected more often than females, often in the ratio of 2 : 1. Acute PSGN rarely strikes the same individual twice. Thus, it appears likely that protective immunity is acquired as a result of an infection with a nephritogenic strain.

Although penicillin treatment of the antecedent streptococcal infections is highly efficacious in preventing acute rheumatic fever, it does not appear to be the case in PSGN. Studies carried out during epidemics of nephritis found that prior antibiotics had little, if any, effect in preventing PSGN. Penicillin is, nevertheless, effective in epidemiologic attempts to eradicate nephritogenic strains by treatment of acute glomerulonephritis patients and their colonized family contacts. Because recurrent episodes of acute glomerulonephritis are so rare, continuous antibiotic prophylaxis, such as is used in the secondary prevention of rheumatic fever, is unnecessary.

The ultimate prognosis in persons with acute glomerulonephritis depends largely on the severity of the initial insult. In an extremely small proportion of patients, the initial injury is so severe that either persistent renal failure or progression to renal failure occurs. However, in most patients, histologic regression of the disease is the rule, and the ultimate prognosis is good.

DIAGNOSIS

Streptococcal Antibody Tests

Ant streptococcal antibody titers reflect past and not present immunologic events and therefore cannot be used to determine whether an individual with pharyngitis and GAS in the pharynx is truly infected or merely a streptococcal carrier. When present, elevated or rising antistreptococcal antibody titers provide reliable confirmation of a recent GAS infection and are of value in identifying a preceding GAS infection in a patient suspected of having rheumatic fever. The most commonly used and commercially available antibody assays are antistreptolysin O and anti-DNAse B. These tests are valuable in patients who have possible nonsuppurative complications of GAS infections (acute rheumatic fever or acute glomerulonephritis). The antistreptolysin O test is usually obtained first, and if the result is not elevated, an anti-DNAse B test may be performed. Antistreptolysin O titers begin to rise approximately 1 week and peak 3 to 6 weeks after the infection. Anti-DNAse B titers begin to rise 1 to 2 weeks and peak 6 to 8 weeks after the infection. Elevated titers for both tests may persist for several months after even uncomplicated GAS infections.

It is not uncommon for laboratory personnel and physicians to misinterpret streptococcal antibody titers because of a failure to appreciate that the normal levels of these antibodies are higher among school-aged children than among adults. Single antibody titers are often misleading. Sequential samples more accurately define infection, allowing correlation of titer increases with temporal confirmation of GAS acquisition.

TREATMENT Rx

Antimicrobial Susceptibility

Penicillin and the cephalosporins continue to be the drugs of first choice for the treatment of most GAS infections. The continued susceptibility of *S. pyogenes* to β-lactams is remarkable, especially compared with the emergence of resistance to β-lactams among pneumococci and VGS over the past decade. A limited ability to acquire foreign DNA and a physiologic fitness cost associated with β-lactam resistance are possible explanations for this organism's continued susceptibility to penicillin. The macrolides are the second-line drug of choice and are used in penicillin-hypersensitive patients. Clindamycin is also the drug of choice for chronic, recurrent pharyngitis. However, resistance has emerged in this organism. Elevated rates of macrolide resistance (>25%) among GAS reported in Korea, Taiwan, Spain, and Italy are causes of concern. Limited surveillance data suggest that the overall rate of macrolide-resistant GAS in North America is 3% to 9%. There are two primary mechanisms of macrolide resistance in GAS: efflux (encoded by *mefA*) and target modification caused by ribosomal methylation (encoded by *ermB* or *ermA*). Isolates with the *mefA* gene have a pattern of resistance known as the M phenotype: they are macrolide resistant but clindamycin susceptible. GAS isolates with the *ermB* gene are usually resistant to macrolides and clindamycin. Isolates with the *ermA* gene typically have an inducible phenotype that requires exposure to a macrolide inducer before clindamycin resistance becomes evident. There are rare reports of L4 ribosomal protein and 23S recombinant RNA mutations in macrolide-resistant GAS isolates.

Fluoroquinolones are a useful therapeutic alternative for the treatment of skin and soft tissue infections in adults. Although strains of GAS nonsusceptible to the fluoroquinolones have been reported, they are rare. The mechanisms of decreased susceptibility to the fluoroquinolones among *Streptococcus* spp. are mainly mediated by point mutations in the quinolone resistance–determining region of the bacterial topoisomerase II enzymes, namely, DNA gyrase and topoisomerase IV. DNA gyrase and topoisomerase IV enzymes are homologs, with each enzyme consisting of a tetramer with two subunits (two *gyrA* and two *gyrB* molecules in DNA gyrase, two *parC* and two *parE* molecules in topoisomerase IV). *S. pyogenes* isolates that are resistant or have reduced susceptibility to fluoroquinolones have been reported, but isolates with a high level of resistance have been detected very infrequently.

Tetracycline resistance in GAS is quite common. In gram-positive organisms, resistance to tetracycline is typically conferred by ribosome protection genes, such as *tet*(M) and *tet*(O). Tetracycline resistance is usually acquired by GAS through horizontal gene transfer. A global sample of GAS revealed 80 or more separate acquisitions of tetracycline resistance. Of 244 clones, 38% and 25% displayed resistance to tetracycline and erythromycin, respectively; a relatively high proportion (15%) were resistant to both classes of drugs. *tet*(M) displayed a highly significant association with *erm*(B). Trimethoprim–sulfamethoxazole is not active against GAS and therefore should not be used as monotherapy for infections potentially caused by GAS.

STREPTOCOCCUS AGALACTIAE (GROUP B STREPTOCOCCUS)

EPIDEMIOLOGY

Group B streptococci, also known as *Streptococcus agalactiae*, was once considered a pathogen of only domestic animals, causing mastitis in cows. Although asymptomatic vaginal carriage of GBS was described in 1935, the first report of GBS sepsis in a neonate was not reported until 1964. Since the 1970s, GBS is recognized as one of the most common causes of neonatal infectious morbidity and mortality in developed countries. GBS causes significant maternal and perinatal morbidity, asymptomatic bacteriuria in pregnancy, and urinary tract and other infections in the adult nonpregnant population. The virulence of *S. agalactiae* is related to the polysaccharide

toxin it produces. Immunity is mediated by antibodies to the capsular polysaccharide and is serotype specific. Several serotypes are known, including Ia, Ib, Ic, II, III, IV, V, VI, VII, and VIII.

CLINICAL MANIFESTATIONS

Neonatal Group B Streptococcus Disease

Group B streptococci colonizes the vaginal and gastrointestinal (GI) tracts in healthy women, with carriage rates ranging from 15% to 45%. Neonates can acquire the organism vertically in utero or during delivery from the maternal genital tract. Although the transmission rate from mothers colonized with *S. agalactiae* to neonates delivered vaginally is approximately 50%, only 1% to 2% of colonized neonates go on to develop invasive GBS disease. Neonatal GBS disease is divided into early and late disease. Early GBS neonatal sepsis often presents within 24 hours of delivery but can become apparent up to 7 days postpartum. Pneumonia with bacteremia is common, but meningitis is less likely. Late GBS neonatal sepsis is defined as infection that presents between 1 week postpartum and age 3 months. Late disease commonly involves GBS serotype III, typically characterized by bacteremia and meningitis.

PREVENTION

The current approach to the prevention of early-onset GBS infection in pregnancy is to provide intrapartum antimicrobial prophylaxis in women at term with culture evidence of recent vaginal or rectal GBS infection.■ Women without a known GBS status delivering before 37 weeks' gestation with premature rupture of the membranes or intrapartum fever are also candidates for intrapartum antimicrobial prophylaxis. Penicillin or ampicillin is the initial approach. Clindamycin and macrolides are standard in individuals with penicillin allergy, but GBS infections are no longer always sensitive to these two drugs. Despite these recommendations, there is lack of evidence from well-designed and well-conducted trials to recommend intrapartum antibiotic prophylaxis to reduce neonatal early-onset GBS disease.

Invasive Group B Streptococcus Disease

S. agalactiae infection is extremely rare in healthy adults and is almost always associated with underlying abnormalities. Bacteremia with an unknown source accounts for approximately 25% of all cases of invasive GBS disease. Diabetes mellitus and malignancy are the most common underlying diseases associated with infection. Urinary tract infections are a common manifestation of GBS disease and are observed in both pregnant and nonpregnant adults. Other presentations of GBS infection include pneumonia, skin and soft tissue infections, septic arthritis, osteomyelitis, meningitis, peritonitis, and endo-ophthalmitis.

Invasive GBS disease is a major cause of illness and death in older adults and is even more frequent among long-term care facility residents, possibly because of concurrent conditions such as advanced age, diabetes, cirrhosis, and stroke, which are known risk factors for GBS infection.[9] Case-fatality rates are about 13% for persons 65 years and older. An increased prevalence of serogroup V has been reported in adult populations. Serotype V has been associated with higher rates of antimicrobial drug resistance.

TREATMENT Rx

Although β-lactams remain the preferred therapy for GBS infections, strains with elevated penicillin minimal inhibitory concentrations have been reported but are rare.

STREPTOCOCCUS DYSGALACTIAE SUBSPECIES *EQUISIMILIS* (HUMAN GROUP C AND G STREPTOCOCCI)

EPIDEMIOLOGY

Non–group A or B β-hemolytic streptococci that are frequently normal inhabitants of the oropharynx, skin, and GI and genitourinary (GU) tracts are also capable of causing significant disease. These include groups C, G, F, and L, of which the most common are groups G and C (see Table 290-1). Large-colony-forming human groups C and G streptococci are currently classified as *Streptococcus dysgalactiae* subspecies *equisimilis* (SDSE). Infections caused by SDSE are transmitted person to person; an animal reservoir for these strains has not been reported. Zoonotic group C or G streptococcal infections are comparatively rare and are mostly caused by other streptococcal species after animal contact or are associated with the consumption of unpasteurized dairy food products. Nonhuman large-colony-forming β-hemolytic group G streptococci usually belong to *Streptococcus canis,* an animal pathogen that has been found only rarely in humans.

CLINICAL MANIFESTATIONS

A population-based study carried out in North America demonstrated a substantial morbidity and mortality associated with infections caused by SDSE, comparable to that caused by invasive GAS infection during the same study period. Both SDSE and GAS affected similar adult populations, tended to be community acquired, and affected primarily persons with underlying medical conditions. The mortality rate was similar to that (8%-21%) reported in other studies. Whereas SDSE primarily presented as skin and soft tissue infection in older patients, individuals with invasive *S. anginosus* group infections were more likely to be younger patients with intra-abdominal infections; these were typically polymicrobial and represented sequelae of perforated appendicitis and associated peritonitis. This finding is supported by the fact that *S. anginosus* group isolates have been recovered from 20% to 40% of healthy appendices.

Pharyngitis is a classic presentation in adult patients, and SDSE has clearly been responsible for epidemic outbreaks of pharyngitis in children. The etiologic role of SDSE in pharyngitis is supported by reports on acute poststreptococcal glomerulonephritis and, more recently, on acute rheumatic fever after SDSE isolation from the upper respiratory tract.

In a 1-year study of 90 patients presenting with 98 disease episodes of acute bacterial cellulitis in Finland, SDSE instead of GAS was strikingly the most common finding. Some patients and household members also carried SDSE in the pharynx; it was not detected in the control participant. Whereas SDSE was isolated either from skin lesions or blood from 22% of patients, GAS was isolated from 7% of patients.

TREATMENT Rx

The non–group A or B β-hemolytic streptococci are typically susceptible to penicillins and cephalosporins, although α-hemolytic VGS have shown increasing resistance to penicillin and other β-lactams. In contrast, the resistance to erythromycin and tetracycline among SDSE is similar to rates found in GAS.

VIRIDANS GROUP STREPTOCOCCI

Streptococcus anginosus Group (*Streptococcus milleri* Group)

The *S. anginosus* group includes three distinct species and more subspecies. *S. anginosus*, *S. constellatus*, and *S. intermedius* have all been collectively known as either *S. anginosus* or *S. milleri*. Members of the *S. anginosus* group are nonmotile, facultative anaerobes that demonstrate variable hemolysis patterns (α, β, or γ) on sheep blood agar. Colonies are typically small (colony < 0.5 mm). A caramel or butterscotch smell is useful for identifying the *S. anginosus* group when present but is not a sufficiently sensitive screening test. There are β-hemolytic strains of each of the three species, and the strains may possess one of four different Lancefield group antigens (group A, C, F, or G) or no group antigen. Non-–hemolytic varieties of each of the three species are grouped into the general classification of viridans streptococci.

The organisms, although normal flora of the human oral cavity and GI tract, have the ability to cause abscesses and systemic infections, unlike *S. pyogenes* and *S. agalactiae*. The *S. anginosus* group should be considered true pathogens when isolated from humans. *S. anginosus* often presents as part of a polymicrobial infection in patients with oral, head and neck, and abdominal infections. Copathogens in such infections may include other bacteria such as *Eikenella corrodens* and *Fusobacterium nucleatum*. This group is less commonly the cause of endocarditis than other viridans streptococci. Whereas *S. intermedius* has an apparent tropism for the brain and liver, *S. anginosus* and *S. constellatus* have been isolated from a wider range of sites and infections.

Thoracic infections caused by *S. anginosus* group are largely pleural. They are polymicrobial in one third of cases and in most patients are associated with major surgery or surgical procedures of the respiratory or digestive tract. The empyema frequently requires thoracotomy for complete resolution. In the chest, infections tend to cause loculated empyemas that are not conducive to tube thoracostomy drainage or to antibiotic treatment. About 75% of infections of the chest require surgical intervention. The optimal method and timing of intervention for thoracic empyema remains debated, likely because of the presentation of pleural space disease in one of three stages: exudative effusions, fibrinopurulent collections, and organized loculations. In the fibrinopurulent stage, the lung can be thoracoscopically dissected off the chest wall without much difficulty.

Members of the *S. anginosus* group are largely susceptible to β-lactam agents. Minimal inhibitory concentrations (MICs) to penicillin G are usually 0.125 mcg/mL or less. Some strains with penicillin G MICs between 0.25 and 2 mcg/mL have been reported; rare strains have penicillin MIC 4 mcg/mL or greater. Penicillin-intermediate or -resistant strains have altered penicillin-binding proteins; these are more likely to be *S. anginosus* or *S. intermedius* than *S. constellatus*. Vancomycin is an appropriate alternative agent for penicillin-allergic patients. A parental third-generation cephalosporin is recommended for brain abscesses and bacteremia caused by members of the *S. anginosus* group.

Fluoroquinolone MICs among *S. anginosus* group members are high but in the susceptible range (0.5-1.0 mcg/mL); resistance tends to develop easily, and therefore fluoroquinolones are not appropriate first-line antimicrobial agents. Macrolide resistance appears to be emerging among the *S. anginosus* group. Most strains of the *S. anginosus* group are resistant to aminoglycosides. Sulfonamides have no activity against *S. anginosus* group isolates.

S. anginosus plays a significant role as a reservoir of antimicrobial resistance genes, transferring different resistance traits to more pathogenic organisms such as *S. pneumoniae* and *S. pyogenes*.

Other Viridans Group Streptococci

There are now at least 30 recognized species of VGS, including those discussed earlier. Although often found as commensals whose pathogenic abilities appear to be much more subtle than those of the pyogenic streptococci, they may also participate in various infections such as subacute bacterial endocarditis; catheter-related and neutropenia-related bloodstream infections; and purulent abdominal, hepatobiliary, brain, and dental infections. Antimicrobial resistance is substantial in the viridans streptococci as a group. Penicillin resistance is from 30% to 50% in strains isolated from patients in North America. *S. mitis* and *S. oralis* are the most common strains found in blood cultures of cancer patients and are commonly resistant to β-lactam antimicrobials. *S. sanguinis*, *S. oralis*, and *S. gordonii*, in descending order, are the most common strains isolated from cultures of blood of endocarditis patients.

ZOONOTIC STREPTROCOCCI

Streptococcus suis

S. suis is a pathogen in pigs that can cause severe infection in humans.[10] *S. suis* was first reported by veterinarians in 1954 after outbreaks of meningitis, septicemia, and purulent arthritis occurred among piglets. Fourteen years later, the first human *S. suis* cases were diagnosed in Denmark, and subsequently, other cases were reported in more than 30 countries, including Canada, the United States, Japan, Vietnam, and Thailand. It can be clinically manifested with meningitis, arthritis, and sepsis. *S. suis* is the name assigned to streptococci that were formally called Lancefield groups R, S, and T. Of the 35 known serotypes, the most frequent serotype identified from humans has been serotype 2 (group R). Human *S. suis* infections are most often reported from countries where pig rearing is common. The relatively high mean patient age (47-55 years) and almost complete absence of children in case series, as well as the high male-to-female patient ratio (3.5 : 1 to 6.5 : 1), support the notion that infection with *S. suis* is generally an occupational disease. The annual risk of developing *S. suis* meningitis among abattoir workers and pig breeders has been estimated to be 3 cases per 100,000 population; the risk is lower for butchers, at 1.2 cases per 100,000 population in developed countries. The incubation period of *S. suis* is from hours to 14 days (median, 2.2 days). A very short incubation time is consistent with direct entry of *S. suis* into the blood through skin wounds. Patients have generally been healthy before infection with *S. suis,* although predisposing factors, such as splenectomy, diabetes mellitus, alcoholism, malignancy, and structural heart diseases, have been reported. An outbreak in China was associated with an overall case-fatality rate of 18%, but this reached 63% among patients with severe sepsis and septic shock.

Data from Vietnam show that *S. suis* is susceptible to penicillin, ceftriaxone, and vancomycin. Resistance was seen to tetracycline (83%), erythromycin (20%), and chloramphenicol (3%). In a European study, 384 *S. suis* strains from diseased pigs were susceptible to penicillin, but resistance was detected to gentamicin (1.0%), trimethoprim–sulfamethoxazole, (6.0%), and tetracycline (75%). There may be a role for the use of dexamethasone as an adjunctive treatment to reduce mortality and improve the outcome of bacterial meningitis caused by infection with *S. suis*. In a randomized, double-blind, placebo-controlled clinical trial in Vietnam, dexamethasone (0.4 mg/kg twice daily for 4 days) resulted in a significant reduction in the risk for death at 1 month and in the risk for death and disability at 6 months in patients with confirmed bacterial meningitis; *S. suis* accounted for 25% of pathogens.

Currently, a human vaccine is not available, but simple preventive measures, such as wearing gloves during processing pig meat or slaughtering, handwashing after handling raw pork meat, and thorough cooking of pork, should prevent most cases. Travelers should be aware that dietary habits in some countries may pose a risk for infectious diseases, including *S. suis* infection.

Streptococcus canis

S. canis is an organism that was described in 1986 as having the Lancefield group G antigen and was isolated from animals, most frequently dogs. Extensive phenotypic testing of isolates from dogs was described in 1994, and the isolation of *S. canis* from a human with sepsis was reported in 1997. Despite its name, *S. canis* has also been isolated from animals other than dogs, including cats, harbor porpoises, cows, mice, rats, and rabbits. In healthy dogs, *S. canis* is found as commensal flora of the skin, oropharynx, GU tract, and anus. Rare cases of *S. canis* infection in humans have been reported in the literature. Resistance to erythromycin and tetracycline is common. In the clinical laboratory, group G streptococci are not usually identified to the species level. In addition, laboratory tests have only recently been able to identify *S. canis* accurately. Thus, the true incidence of *S. canis* infection is unknown.

Streptococcus bovis Group

Members of the *Streptococcus bovis* group have long been regarded as opportunistic pathogens in humans, causing up to 11% to 14% of all endocarditis cases and 24% of streptococcal endocarditis. *S. bovis* group isolates from human infections are divided into three biotypes, designated biotypes I, II/1, and II/2. Most *S. bovis* group endocarditis isolates belong to biotype I, for which a reclassification as *Streptococcus gallolyticus* subspecies *gallolyticus* has been proposed. The human intestinal tract is a major natural reservoir of *S. bovis* group strains with a reported carriage rate of about 10%. There is a strong association between the presence of malignant or premalignant lesions in the GI tracts of patients and isolation of *S. bovis* as a causative agent of bacteremia or endocarditis.

Streptococcus iniae

S. iniae is a β-hemolytic streptococcus without a group antigen. It is a major fish pathogen in many regions of the world. These bacteria are also zoonotic, with infections in humans associated with the handling and preparation of infected fish. The first human infections were reported in 1996. Most cases of human *S. iniae* infections have been in persons of Asian descent, who are elderly and commonly have underlying conditions such as diabetes mellitus, chronic rheumatic heart disease, cirrhosis, or other conditions.

Carrier fish have been implicated in fish-to-fish transmission of *S. iniae*, and these carriers may be responsible for human infection because fish with overt signs of disease are unmarketable. Soft tissue injuries that occur during the preparation of fresh fish from wet markets usually result in bacteremic cellulitis of the hand followed by one or more of the following conditions: endocarditis, meningitis, arthritis, sepsis, pneumonia, osteomyelitis, and toxic shock. *S. iniae* is susceptible to penicillin and other antimicrobials. The expression of a suite of virulence factors, many of them similar to those found in GAS, is responsible for successful entry, propagation, and evasion of immune defenses of the host by this bacterium. Underreporting of human cases is likely because identification of *S. iniae* is based on biochemical testing of isolates with commercial kits; the use of kits is associated with problems because *S. iniae* is not listed in commercial or clinical databases, and many atypical strains are assigned low matches.

Grade A References

A1. Koning S, van der Sande R, Verhagen AP, et al. Interventions for impetigo. *Cochrane Database Syst Rev*. 2012;1:CD003261.

A2. Van Driel ML, De Sutter AL, Keber N, et al. Different antibiotic treatments for group A streptococcal pharyngitis. *Cochrane Database Syst Rev*. 2013;4:CD004406.

A3. Ohlsson A, Shah VS. Intrapartum antibiotics for known maternal Group B streptococcal colonization. *Cochrane Database Syst Rev*. 2014;6:CD007467.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Low DE. Toxic shock syndrome: major advances in pathogenesis, but not treatment. *Crit Care Clin*. 2013;29:651-675.
2. Johansson L, Norrby-Teglund A. Immunopathogenesis of streptococcal deep tissue infections. *Curr Top Microbiol Immunol*. 2013;368:173-188.
3. Essop MR, Peters F. Contemporary issues in rheumatic fever and chronic rheumatic heart disease. *Circulation*. 2014;130:2181-2188.
4. Burke RJ, Chang C. Diagnostic criteria of acute rheumatic fever. *Autoimmun Rev*. 2013;13: 503-507.
5. Kerdemelidis M, Lennon DR, Arroll B, et al. The primary prevention of rheumatic fever. *J Paediatr Child Health*. 2010;46:534-548.
6. Gerber MA, Baltimore RS, Eaton CB, et al. Prevention of rheumatic fever and diagnosis and treatment of acute streptococcal pharyngitis: a scientific statement from the American Heart Association Rheumatic Fever, Endocarditis, and Kawasaki Disease Committee of the Council on Cardiovascular Disease in the Young, the Interdisciplinary Council on Functional Genomics and Translational Biology, and the Interdisciplinary Council on Quality of Care and Outcomes Research: endorsed by the American Academy of Pediatrics. *Circulation*. 2009;119:1541-1551.
7. van der Helm-van Mil AH. Acute rheumatic fever and poststreptococcal reactive arthritis reconsidered. *Curr Opin Rheumatol*. 2010;22:437-442.
8. VanDeVoorde RG 3rd. Acute poststreptococcal glomerulonephritis: the most common acute glomerulonephritis. *Pediatr Rev*. 2015;36:3-13.
9. Lambertsen LM, Ingels H, Schonheyder HC, et al. Nationwide laboratory-based surveillance of invasive beta-haemolytic streptococci in Denmark from 2005 to 2011. *Clin Microbiol Infect*. 2014;20:O216-O223.
10. Feng Y, Zhang H, Wu Z, et al. Streptococcus suis infection: an emerging/reemerging challenge of bacterial infectious diseases? *Virulence*. 2014;4:477-497.

REVIEW QUESTIONS

1. Group A streptococcus (GAS) is a major cause of pharyngitis and reasons for visits to health care facilities and for the prescription of antibiotics. Which of the following statements is correct?

A. The treatment of GAS pharyngitis is an important strategy for the prevention of rheumatic heart disease in developed countries.
B. Periarticular findings, such as tenosynovitis, have not been considered to be part of the clinical spectrum of acute rheumatic fever but have been described with poststreptococcal reactive arthritis.
C. Penicillin treatment of the antecedent streptococcal infections is highly efficacious in preventing acute rheumatic fever and post-streptococcal glomerulonephritis (PSGN).
D. Quantitation of GAS from the throat swab culture can be used to differentiate carriage from infection.
E. On clinical grounds, streptococcal pharyngitis cannot be strongly suggested by the presence of fever, tonsillar exudate, tender enlarged anterior cervical lymph nodes, and presence of cough without laboratory confirmation.

Answer: B The term *poststreptococcal reactive arthritis* was first used to describe patients with arthritis after documented GAS infection but who failed to meet the Jones criteria for the diagnosis of acute rheumatic fever. Since then, the differentiation of PSRA from acute rheumatic fever has remained unsettled. In the latter half of the 20th century, rheumatic fever receded as an important health problem in almost all wealthy countries. Most of the reduction is attributable to improved living conditions, which have resulted in less overcrowding and better hygiene, with consequent reductions in transmission of GAS, not the widespread use of antibiotics. In other words, rheumatic fever is a disease of poverty. Studies carried out during epidemics of nephritis found that prior antibiotics had little, if any, effect in preventing acute PSGN. Penicillin is, nevertheless, effective in epidemiologic attempts to eradicate nephritogenic strains by treatment of acute glomerulonephritis patients and their colonized family contacts. Quantitation of GAS from the throat swab culture is unable to differentiate at presentation between a child who is acutely infected and one who is a chronic carrier presenting with acute nonstreptococcal pharyngitis. The methods for making that discrimination (testing for serological response and strain typing) require time and the aid of reference or research laboratories. A position paper by the American College of Physicians–American Society of Internal Medicine/American Academy of Family Physicians/U.S. Centers for Disease Control and Prevention recommends a departure from the principle of laboratory confirmation of all adult cases with recommendations to use a clinical prediction rule.

2. Acute rheumatic fever is a delayed, nonsuppurative sequela of a pharyngeal infection with group A streptococci (GAS). Which of the following statements is correct?

A. Only "rheumatogenic" strains have been associated with acute rheumatic fever.
B. According to revised Jones criteria, the diagnosis of rheumatic fever can only be made when two of the major criteria or one major criterion plus two minor criteria are present along with evidence of streptococcal infection.
C. Penicillin (either oral penicillin V or injectable benzathine penicillin) is the treatment of choice.
D. Poststreptococcal reactive arthritis (PSRA) is an entity in patients who had arthritis after an episode of GAS pharyngitis but lacked other major criteria of acute rheumatic fever. However, they frequently go on to develop acute rheumatic fever.

Answer: C Penicillin is cost effective, has a narrow spectrum of activity, and has long-standing proven efficacy. In addition, GAS resistant to penicillin has not been documented. For penicillin-allergic individuals, acceptable alternatives include a narrow-spectrum oral cephalosporin, oral clindamycin, or various oral macrolides. Although the observation in some studies that only a few M serotypes (types 3, 5, 6, 14, 18, 19, 24, and 29) were implicated in outbreaks of rheumatic fever suggested a particular "rheumatogenic" potential of certain strains of GAS, the issue of potential rheumatogenic strains remains unresolved. A streptococcal strain capable of causing a well-documented pharyngitis almost always is potentially capable of causing rheumatic fever. Exceptions are chorea and indolent carditis, each of which by itself can indicate rheumatic fever. Compared with acute rheumatic fever, the arthritis of PSRA is more likely to occur within 10 days after infection, be symmetrical and nonmigratory, last longer than 3 weeks, and be poorly responsive to aspirin or other nonsteroidal anti-inflammatory drugs. Periarticular findings, such as tenosynovitis, have not been considered to be part of the clinical spectrum of acute rheumatic fever but have been described with PSRA. Because a small proportion of patients with PSRA have been reported to develop subsequent valvular heart disease, these patients should be observed carefully for several months for clinical evidence of carditis.

3. The current recommendations for the surgical treatment of group A streptococci (GAS) necrotizing fasciitis are based on

A. Prospective randomized controlled trials comparing immediate surgical intervention (<24 hours from time of admission) versus delayed therapy (>24 hours from time of admission).
B. Retrospective reviews of studies comparing the outcomes of patients treated early in their disease versus those who had delayed therapy.
C. Retrospective reviews of studies of patients with a diagnosis of necrotizing fasciitis from multiples etiologies.
D. Prospective studies of patients with GAS necrotizing fasciitis.

Answer: C All of the reports of outcomes of necrotizing fasciitis have been studies that have included necrotizing fasciitis from multiple causes. There have been no prospective studies comparing immediate surgical intervention (<24 hours from time of admission) versus delayed therapy (>24 hours from time of admission). The published studies are all based on retrospective chart review of patients with necrotizing fasciitis from multiple etiologies, including mixed aerobic and anaerobic organisms.

4. Since the 1970s, group B streptococcus (GBS) is recognized as one of the most common causes of neonatal infectious morbidity and mortality in developed countries. Which of the following statements is correct?

A. Although the transmission rate from mothers colonized with *S. agalactiae* to neonates delivered vaginally is approximately 50%, only 1% to 2% of colonized neonates go on to develop invasive GBS disease.
B. Late GBS neonatal sepsis is defined as infection that presents between 1 week postpartum and age 3 months. Late disease commonly involves transmission of GBS from mother to baby after discharge.
C. The current approach to the prevention of early-onset GBS infection in pregnancy requires prepartum antimicrobial prophylaxis in women in their last trimesters of pregnancy with culture evidence of recent vaginal or rectal GBS infection.
D. Early-onset GBS disease can still occur in infants whose mothers screened negative for GBS colonization. Therefore, all women should receive intrapartum antibiotic prophylaxis regardless of screening results.

Answer: A The ability of the organism to cause invasive disease is related to the virulence and preexisting immunity from the mother. Virulence of *S. agalactiae* is related to the polysaccharide toxin it produces. Immunity is mediated by antibodies to the capsular polysaccharide and is serotype specific. Several serotypes are known, including Ia, Ib, Ic, II, III, IV, V, VI, VII, and VIII. A minority of infants with GBS late-onset disease are born to GBS-colonized mothers. Intrapartum prophylaxis does not appear to prevent late-onset GBS disease, implicating infected breast milk and nosocomial or community sources in these cases.

Multiple clinical trials have demonstrated that the use of intrapartum penicillin or ampicillin significantly reduces the rate of neonatal colonization with GBS and the incidence of early-onset neonatal GBS disease. Even in the setting of a maternal GBS screening program, efforts to evaluate and treat infants with intrapartum clinical risk factors for early-onset sepsis remain important. Results of studies have demonstrated that prompt intrapartum antibiotic treatment of women with signs and symptoms of chorioamnionitis regardless of their GBS screening results and the evaluation of well-appearing infants for possible sepsis because of intrapartum clinical risk factors remain imperative.

5. The *S. anginosus* group includes three distinct species and more subspecies. *S. anginosus, S. constellatus,* and *S. intermedius* have all been collectively known as *S. anginosus*. Characteristics of this group include

A. Being part of the normal flora of the genitourinary tracts.
B. The *S. anginosus* group should be considered true pathogens when isolated from humans.
C. *S. milleri* is a closely related viridans group streptococci that is more frequently a cause of endocarditis.
D. Lung infections caused by the *S. anginosus* group are largely confined to the right middle lobe.

Answer: B Although part of the normal flora, the group has the ability to cause abscesses and systemic infections, unlike *S. pyogenes* and *S. agalactiae*. The *S. anginosus* group is a heterogeneous group of organisms that can be human commensals, colonizing the gastrointestinal tract and the oral mucosa. The *S. anginosus* group is generally considered to be of low pathogenic potential in immunocompetent individuals. The *S. anginosus* group is one that has been the source of much controversy and confusion regarding taxonomy and classification. The group includes *S. anginosus, S. constellatus,* and *S. intermedius*. However, each of these species has been known as *S. anginosus* or *S. milleri* at one time. *S. milleri* has never been accepted as a confirmed taxonomic entity, although historically, it has been widely used interchangeably with the *S. anginosus* group. Thoracic infections caused by *S. anginosus* group are largely pleural. They are polymicrobial in one third of cases, and in most patients, are associated with major surgery or surgical procedures of the respiratory or digestive tract. The empyema frequently requires thoracotomy for complete resolution. In the chest, infections tend to cause loculated empyemas that are not conducive to tube thoracostomy drainage or to antibiotic treatment alone. About 75% of infections of the chest require surgical intervention.

291 ENTEROCOCCAL INFECTIONS

TRISH M. PERL

DEFINITION

Enterococci, formerly called group D streptococci, are endogenous human gut flora that had been considered pathogens with low virulence in the past. However, more recently, they have emerged as increasingly important health care–associated pathogens. This emergence is primarily because of their inherent resistance to commonly used antimicrobials, acquisition of high-level resistance to vancomycin and aminoglycosides, persistence in the environment, and transmission from patient to patient by way of the contaminated hands of health care workers. Hence, the emergence of vancomycin-resistant enterococci (VRE) has limited therapeutic options in confirmed enterococcal infections and in empirical therapy for infections in severely ill hospitalized patients and it represents a challenge for infection control in the health care setting. This chapter reviews the most important clinical manifestations of enterococci and their diagnosis and the importance of infection prevention.

The Pathogen

Members of the genus *Enterococcus* were long classified within group D of the genus *Streptococcus*. However, in the past 30 years, they have been reclassified based on new molecular and genetic analyses. Enterococci are catalase-negative gram-positive cocci that can appear singly or in pairs or short chains. They are facultative anaerobes that grow optimally at 35° to 37° C and are usually α-hemolytic or nonhemolytic on sheep blood agar. Enterococci can grow in broth containing 6.5% NaCl and hydrolyze esculin in the presence of 40% bile salts (bile-esculin medium) that can distinguish them from most streptococci. *Enterococcus faecalis,* the most common cause of enterococcal infections in humans, is the causative agent for 80% to 90% of the enterococcal infections followed by *Enterococcus faecium,* which is found in 5% to 10% of the infections. *Enterococcus casseliflavus, Enterococcus gallinarum,* and *Enterococcus raffinosus* are less frequently associated with infections, but clusters of infections have been reported. Other species isolated from different sources in humans include *Enterococcus avium, Enterococcus caccae, Enterococcus cecorum, Enterococcus dispar, Enterococcus durans, Enterococcus gilvus, Enterococcus italicus, Enterococcus hirae, Enterococcus malodoratus, Enterococcus mundtii, Enterococcus pallens, Enterococcus pseudoavium,* and *Enterococcus sanguinicola.*

EPIDEMIOLOGY

Enterococci are part of the normal human gut flora, and infections in both hospitalized and nonhospitalized patients can arise from either an endogenous or exogenous source. The proportion of infections caused by enterococci in hospitalized patients has been increasing over the past several decades. Overall, urinary tract infections (UTIs) are the most common clinical condition caused by enterococci. Based on data reported to the Centers for Disease Control and Prevention (CDC), enterococcal species are now the second most common isolates for any health care–associated infection and prominent causes of catheter-associated bloodstream, UTIs, and surgical site infections, causing approximately 15% of these infections in North America. Enterococci are the second most common cause of catheter-associated UTIs and cause both complicated and uncomplicated UTIs. They are also the second most common cause of central line–associated blood stream infections and are involved in approximately 5% to 15% of all cases of infective endocarditis. They are also the third most frequent pathogen recovered in surgical site infections. Outside of the United States, these organisms are less common but increasingly important causes of infections.

In the 1970s, *E. faecalis* accounted for up to 95% of isolates and was associated with the introduction of third-generation cephalosporins. Increasingly, hospitalized patients, if they are colonized or infected with an *Enterococcus* species, tend to have a strain resistant to vancomycin and sometimes ampicillin (i.e., VRE).[1] *E. faecium* is the most common strain to acquire vancomycin resistance. VRE was first reported in Europe in 1986. Since the mid-, the proportion of enterococcal strains resistant to vancomycin, primarily *E. faecium,* has risen steadily. According to the CDC's National Healthcare Safety Network (NHSN), 62% to 82.6% of *E. faecium* isolates and 6.2% to 9.5% of *E. faecalis* isolates were resistant to vancomycin in 2009 to 2010.[2] Also, according to a recent report by the CDC on antimicrobial resistance in the United States, 66,000 *Enterococcus* health care–associated infections are reported yearly, and 20,000 of them are caused by VRE. There are two major genotypes for acquired vancomycin resistance, VanA and VanB. The genes encoding the VanA phenotype result in high-level resistance to vancomycin and teicoplanin and are carried on a transposon usually found on a plasmid that is transferable to other gram-positive cocci, including *Staphylococcus aureus*. VanA is mostly found in *E. faecium* and, less frequently in *E. faecalis*. VanB is associated with variable resistance to vancomycin, but isolates are usually susceptible to teicoplanin. VanB is usually recovered from *E. casseliflavus* and *E. gallinarum*. Importantly, these genetic elements have been integrated into the genome of *S. aureus* that is resistant to vancomycin. The vancomycin-resistant *S. aureus* (VRSA) acquired a vancomycin resistance gene (*VanA*) from a VRE isolate that colonized a patient coinfected with methicillin-resistant *S. aureus*.

The epidemiology of VRE differs between Europe and North America. In Europe, VRE is often detected in farm animals, likely because of the use of the antibiotic avoparcin in animal feeds until it was banned in 1997. The proportion of VRE among enterococcal clinical isolates in hospitalized Europeans has been historically lower than in the United States; however, these rates are increasing. In the United States, avoparcin was never used in animal feeds, and therefore VRE is not usually found in farm animals or healthy humans. In contrast, the proportion of enterococcal isolates resistant to vancomycin is higher in U.S. hospitals than European hospitals. The proportion of enterococcal isolates from hospitalized European patients that are highly resistant to vancomycin varies by geographic region from less than 1% to up to 35%.

Most VRE infections are associated with health care and result from an exogenous source, meaning transmission from the environment, another patient, or on the hands of a health care worker. Nearly all infections are preceded by a period of colonization, primarily in the gastrointestinal (GI) tract. A study of hospitalized patients found that the most sensitive predictor of VRE colonization was previous admission to an acute care hospital in the past year. Importantly, after being colonized with VRE, patients may harbor the strain in their GI tract for years. Similarly, a study among hemodialysis patients demonstrated that the primary risk factor for VRE colonization was hospitalization in the prior year. Although many colonized patients do not develop infections, they are still able to contaminate the environment and shed and transmit bacteria to other hospitalized patients. The organism has a predilection to contaminate the hospital environment and equipment and has been associated with outbreaks.

Apart from preexisting GI colonization, risk factors for enterococcal infections, particularly VRE infections, include severe underlying conditions such as renal failure, previous solid organ or bone marrow transplantation, solid and hematologic malignancy, diabetes mellitus, and neutropenia. Other factors associated with infection include prior surgical or GI procedures, presence of a vascular or urinary indwelling catheter, hospital factors such as location in the intensive care unit (ICU) or oncology ward, proximity to colonized patients, prolonged length of hospitalization, and recent antimicrobial exposure.

Numerous epidemiologic investigations have revealed that most classes of antimicrobials have been associated with VRE infections. In particular, vancomycin, cephalosporins, and drugs with anaerobic organism coverage use have been linked to VRE acquisition. However, measuring the attributable impact of a particular antibiotic on VRE acquisition is difficult. Increasingly, an association between VRE colonization and *Clostridium difficile*

infection is reported in high-risk patients such as those with hematologic malignancies.

PATHOBIOLOGY

Enterococci are commensal organisms that colonize the human GI tract and female genital tract. Although they are not as intrinsically virulent as other gram-positive pathogens, under certain conditions, the commensal relationship is disrupted, and serious infections occur. Several adhesion factors have been identified, including aggregation substance, which allow binding to epithelial surfaces and enhance the ability for colonization. The ability to adhere to heart valves and urinary tract epithelium enables enterococci to cause endocarditis and UTIs. Enterococci are also known to secrete potential virulence factors. These include cytolysin–hemolysin, which is a bacterial toxin that is produced in a higher proportion of infecting strains compared with stool-colonizing strains. Infecting strains also possess the ability of intestinal translocation, although the exact mechanisms of this process have yet to be determined. In addition, some cell surface determinants that are encoded may mediate adherence to host tissues that may be important in this organism's role in endocarditis. To this point, little is known about the host defense mechanisms in enterococcal infections. In addition, the exact role of capsular polysaccharides in colonization or infection is unknown. Strains have been shown to survive within phagocytic cells, yet it is unclear whether this represents successful host defense or evasion by the enterococci. The intrinsic resistance to multiple antimicrobials that enterococci possess, along with their ability to acquire new resistance mechanisms through mutation or acquisition of new genes, enhances their ability to survive and multiply in the many hospitalized patients treated with broad-spectrum antimicrobials.

CLINICAL MANIFESTATIONS

No specific clinical manifestations can help distinguish enterococcal infections from infections caused by other bacteria. Enterococci are not thought to cause lower respiratory tract infections or ventilator-associated pneumonia and, if found in these settings, likely represent colonization and not infection. Enterococci act as opportunistic pathogens in severely ill and compromised patients. They are known to cause UTIs, intra-abdominal abscesses, wound infection, bacteremia (including catheter-associated blood stream infections), and endocarditis.

Urinary Tract Infections

Urinary tract infections are the most frequent type of infection caused by enterococci.[3] Most infections are nosocomial in origin and include uncomplicated cystitis, pyelonephritis, prostatitis, and perinephric abscess. These infections are typically secondary to urinary catheterization or instrumentation. In contrast to nosocomial UTI, enterococci cause fewer than 5% of uncomplicated cystitis or pyelonephritis cases in otherwise healthy nonhospitalized women. Patients with diabetes mellitus appear to be at increased risk for enterococcal UTI. Bacteremia is only infrequently associated with enterococcal UTI.

Bacteremia

Importantly, enterococci can cause infection or contaminate cultured blood via contaminated catheter hubs and contaminated skin. Determining a true bacteremia versus a blood culture that is not clinically significant can be a challenge. Given this backdrop, the incidence of bacteremia caused by enterococci continues to increase. Specific risk factors include prolonged hospitalization, preexisting urethral catheters or intravascular lines, recent surgery, malignancy, neutropenia, and biliary pathology. Secondary bacteremia without endocarditis usually arises from the urinary tract, hepatobiliary tract, or soft tissue infection. Bacteremia secondary to an intra-abdominal source carries a high mortality rate. Risk factors for VRE bacteremia parallel those mentioned previously but additionally include severe preexisting comorbid conditions, including hematologic malignancy, human immunodeficiency virus (HIV) infection, chronic renal insufficiency, and liver transplantation. Prior exposure to broad-spectrum antimicrobials, including those with antianaerobic activity such as clindamycin or metronidazole, and exposure to multiple and prolonged antimicrobial therapy are consistent risk factors. Enterococcal bacteremia is frequently polymicrobial, and the clinical picture is often influenced by whether it is isolated alone or with other bacteria. When enterococci are isolated alone, the course is typically indolent, and frequently fever is the only sign. In contrast, polymicrobial bacteremia is more severe, often presenting with shock or disseminated intravascular coagulation. VRE bacteremia is associated with higher mortality rates than bacteremia caused by vancomycin-susceptible strains, and early treatment with an appropriate antibiotic within the first 48 hours of presentation has been associated with improved outcomes.

Endocarditis

Enterococci, particularly *E. faecalis*, are an increasingly frequent cause of endocarditis even though most enterococcal bacteremias are not complicated by endocarditis. The disease occurs most frequently in older patients, with a male predominance. Most cases appear to arise in the community. Patients with preexisting valvular heart disease, including prosthetic valves, are at highest risk, yet many patients lack underlying heart disease. Enterococci more commonly cause left-sided endocarditis primarily affecting the mitral valve. Clinically, these patients present with symptoms that closely resemble a subacute bacterial endocarditis caused by viridans streptococci. Many patients have symptoms for weeks or months before seeking medical care.

Intra-abdominal Infections

In intra-abdominal infections, enterococci are often detected as part of a polymicrobial process.[4] These infections typically arise from a hepatobiliary source, including postoperative infection in liver transplantation, and are complicated by secondary bacteremia.

Skin and Soft Tissue Infections

Enterococci rarely cause cellulitis or other soft tissue infections alone but are often isolated in mixed surgical site infections, diabetic foot infections, and decubitus ulcers along with other gram-negative bacilli, gram-positive cocci, and anaerobic bacteria. Their clinical significance in these situations has not been adequately determined. Enterococci are not thought to be primary pathogens in chronic osteomyelitis. When they are identified, it is thought they may solely represent superinfection, and thus adequate therapy may not require antibiotics directed at enterococcal eradication.

DIAGNOSIS

The diagnosis of an *Enterococcus* infection is made by isolating the organism through culture of a sterile site, such as blood or urine. Recently, molecular techniques have been developed to identify *Enterococcus* more quickly. The diagnosis and differential diagnosis of specific conditions are the same as discussed for UTIs (Chapter 284) and endocarditis (Chapter 76).

TREATMENT

Rx

Therapy for enterococcal infections is complicated by the fact that strains exhibit inherent resistance to many commonly used antibiotics, including cephalosporins.[5] In addition, enterococci can acquire resistance to a wide range of antibiotic classes, including aminoglycosides (high-level resistance), β-lactams, fluoroquinolones, and vancomycin. Thus, effective directed therapy for any severe enterococcal infection requires susceptibility testing by experienced microbiology laboratories, with therapy adjusted based on the results. Optimal therapy for most infections includes intravenous ampicillin, penicillin, or vancomycin. Given the resistance or tolerance to cell wall–targeting antibiotics, including penicillins and vancomycin, standard therapy with these antibiotics, except in UTIs, should include the addition of a synergistic aminoglycoside (i.e., gentamicin or streptomycin), as long as high-level resistance is not detected. This strategy of two-drug therapy has been associated with improved outcomes. This is particularly important in the setting of suspected endocarditis. Even when enterococcus appears susceptible to trimethoprim–sulfamethoxazole in vitro, it should not be used in therapy because clinical failures have been reported secondary to the ability of enterococci to use exogenous folate. Similarly, *E. faecalis* is intrinsically resistant to quinupristin–dalfopristin, which therefore should not be used in therapy for infections caused by this species. Recently, there has been increasing isolation of *E. faecium* strains with intrinsic resistance to penicillins. Still, most *E. faecalis* strains remain susceptible to ampicillin and the related piperacillin, in contrast to most *E. faecium* strains, which are resistant to ampicillin.

If the VRE strains are known to be susceptible, potential therapy in these infections includes linezolid, tigecycline, and daptomycin. Linezolid is commonly the drug of choice,■ although its use is associated with bone marrow suppression, including thrombocytopenia, and has only bacteriostatic activity against the enterococci. Linezolid, when used in combination with selective serotonin reuptake inhibitors, can be associated with serotonin syndrome (Chapter 434). Although daptomycin is not approved by the U.S. Food and Drug Administration (FDA) to treat VRE, it treats skin and soft tissue infections and bacteremia.[6] Tigecycline is FDA approved to treat complicated skin and soft tissue infections and complicated intra-abdominal infections caused by vancomycin-susceptible isolates of *E. faecalis,* but a recent black box warning

by the FDA limits its use. Nitrofurantoin remains an option for VRE UTIs. Given the complexity of enterococcal infections, an infectious disease consult should be considered for therapeutic guidance.

Urinary Tract Infections

A single agent can usually be used to treat enterococcal UTIs, including ampicillin, amoxicillin, penicillin, quinolones, fosfomycin, or vancomycin (E-Table 291-1). Vancomycin is typically reserved for penicillin-allergic patients or if the strain has high-level penicillin resistance. β-Lactam–β-lactamase inhibitor combinations are usually reserved for polymicrobial infections. Nitrofurantoin is also occasionally used because most strains remain susceptible. Fosfomycin is also indicated for UTIs caused by susceptible *E. faecalis*.

Bacteremia without Endocarditis

Many cases of enterococcal bacteremia are transient or self-limited, yet antibiotic therapy with penicillin or ampicillin has been shown to improve outcomes (see E-Table 291-1). Unlike in endocarditis therapy, it is not known whether patients benefit from combination therapy (penicillin or ampicillin or vancomycin plus an aminoglycoside), except perhaps when an indwelling intravascular catheter is present. In the setting of an indwelling intravascular catheter, especially for VRE, removal of the catheter is indicated. If the bacteremia is secondary to another site such as an intraabdominal abscess, drainage of the source is critical to cure.

Endocarditis

Combination therapy (intravenous penicillin, ampicillin, or vancomycin plus an aminoglycoside) is the standard therapy for enterococcal endocarditis. Various combinations have been tried with varying success depending on the species and susceptibilities of the organisms. Doses and durations are found in Chapter 76 and E-Table 291-1, but in general, consultation with infectious diseases experts is indicated. Importantly, the aminoglycoside is used to provide synergistic killing of the organism. The duration of therapy is typically 4 to 6 weeks with longer therapy given to patients who had prolonged symptoms before seeking therapy, prosthetic valve infection, or relapsed after initial treatment. If the causative enterococcal strain is highly resistant to both gentamicin and streptomycin, then alternative agents and durations must be explored, and surgery to excise infected valves should be considered. Optimal therapy of VRE strains that are resistant to ampicillin is not known but includes combination therapy under the guidance of an infectious diseases consultant. Several newly approved agents, including linezolid, and daptomycin could be considered if the strain is found to be susceptible. VRE endocarditis may require early surgery because outcomes with antibiotic therapy alone can be poor. Careful microbiologic and clinical assessment of all patients with enterococcal endocarditis and VRE in particular is helpful in deciding when surgery is necessary. If repeated blood cultures grow for more than 7 days after the initiation of medical therapy or other signs of uncontrolled infection (persistent fever or leukocytosis) are present, surgical repair of the valve or valve replacement should be considered early in the treatment course if there are no absolute contraindications for surgery.

Of note, linezolid resistance is increasingly reported even in patients without previous exposure to the antibiotic. Daptomycin resistance is also reported in both *E. faecalis* and *E. faecium* after prolonged courses of therapy.

PRIMARY PREVENTION

Optimal infection prevention for VRE, similar to many multidrug-resistant bacteria, includes proper use and compliance with hand hygiene, use of contact precautions, proper management and timely removal of urinary and vascular catheters, reduction of antibiotic selective pressure through antimicrobial stewardship, and environmental cleaning of equipment and patient rooms. The latter is particularly important for VRE. A study in a U.S. hospital's medical ICU found that the number of patients already colonized with VRE in a defined geographic area ("colonization pressure") was the most significant variable in predicting new acquisition of VRE. Hence, decontamination of patients' skin, called source control, may broadly control transmission of resistant pathogens and reduce device-related infections such as central venous catheter–associated blood stream infections. Source control using daily bathing of patients with chlorhexidine gluconate-saturated cloths has been associated with reduced VRE contamination of patients' skin and health care workers' hands. A multicenter cluster randomized, cross-over clinical study in ICUs found daily bathing with chlorhexidine gluconate (CHG) solution reduced VRE acquisition by 25% and decreased the risk for VRE bacteremia in known VRE-colonized patients,[A2] although a subsequent trial has cast doubt on the efficacy of such an approach.[7]

The methods for VRE-specific prevention are directed at preventing incident colonization in high-risk hospitalized patients. Nearly all enterococcal infections, including VRE infections, are preceded by GI colonization, and although most colonized patients do not develop infections, they are still able to shed and transmit bacteria to other hospitalized patients. At least with VRE, a colonization-to-infection ratio of 10 : 1 has been reported, suggesting for every one clinical infection in an ICU, there may be 10 colonized patients lurking undetected in the unit. Thus, the unrecognized colonized patients represent the target population for infection prevention and control efforts such as active surveillance. Active surveillance programs use rectal or perirectal surveillance swabs to detect previously unrecognized, colonized patients and isolate them to prevent further transmission. Currently, surveillance cultures are recommended at the time of hospital admission for patients at high risk for carriage of VRE, and colonized patients or infected patients should be placed in isolation using contact precautions. Contact precautions typically entail private rooms, dedicated equipment such as stethoscopes, and gloves and gowns for all patient contact, although in a recent randomized trial in medical and surgical ICUs, the use of gloves and gowns for all patient contact compared with usual care did not result in a difference in the acquisition of VRE (or methicillin-resistant *S. aureus*).[A3] Implicit in this is the continued use of standard precautions that requires cleaning and disinfection of equipment that is used in patient rooms. Many hospitals do use active surveillance in the ICU and other wards with a high prevalence of VRE, although the overall adoption of this strategy has been hindered by the perceived high costs of surveillance programs and lack of available randomized control trial data.

PROGNOSIS

Enterococcal bacteremia is associated with prolonged hospitalization and added costs compared with similar patients without enterococcal bacteremia. Still, apart from enterococcal endocarditis, the attributable mortality of enterococcal infections is difficult to quantify owing to its predilection to infect patients with preexisting comorbid conditions and high levels of illness severity. In certain patient populations, including those with liver and bone marrow transplants, studies have suggested increased morbidity, length of stay, and mortality associated with vancomycin resistance. Furthermore, a recent meta-analysis reported that the odds of dying from a vancomycin-resistant enterococcal blood stream infection were 2.5 times higher than dying from an infection caused by a susceptible enterococcus. The unadjusted mortality in susceptible enterococcal bacteremia was 20%. It is unclear why resistance is associated with higher mortality, although it is thought that delayed adequate empirical therapy may play a role. These studies should be interpreted cautiously because the clinical impact of resistance in enterococci was assessed before the availability of newer antimicrobials with activity against VRE.

Grade A References

A1. Chuang YC, Wang JT, Lin HY, et al. Daptomycin versus linezolid for treatment of vancomycin-resistant enterococcal bacteremia: systematic review and meta-analysis. *BMC Infect Dis*. 2014; 14:687.

A2. Climo MW, Yokoe DS, Warren DK, et al. Effect of daily bathing with chlorhexidine on hospital acquired infection. *N Engl J Med*. 2013;368:533-542.

A3. Harris AD, Pineles L, Belton B, et al. Universal glove and gown use and acquisition of antibiotic-resistant bacteria in the ICU: a randomized trial. *JAMA*. 2013;310:1571-1580.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

E-TABLE 291-1 TREATMENT OPTIONS FOR VARIOUS *ENTEROCOCCAL SPP.* FOR VARIOUS INFECTION SITES

	ENTEROCOCCUS FAECALIS	*ENTEROCOCCUS FAECIUM*	VANCOMYCIN RESISTANT ENTEROCOCCUS (VRE)
Urinary Tract Infection			
PCN/ampicillin susceptible*	Ampicillin (1 g IV q 6h) or amoxicillin (500 mg PO TID)	Ampicillin (1 g IV q 6h) or amoxicillin (500 mg PO TID)	Linezolid (600 mg PO/IV q 12h) or Daptomycin (8-12 mg/kg IV q 24h) or Fosfomycin (3 g PO once if catheter removed)
PCN/ampicillin resistance*	If susceptible: nitrofurantoin (100 mg PO q 12h†), or TCN (500 mg PO q6h), or fosfomycin (3 g PO once if catheter removed) if uncomplicated§ Vancomycin if resistant to other agents	If susceptible: nitrofurantoin (100 mg PO q 12h†), or TCN (500 mg PO q6h), or fosfomycin (3 g PO once if catheter removed); Vancomycin if resistant to other agents	Linezolid (600 mg PO/IV q 12h) or Daptomycin (8-12 mg/kg IV q 24h) or Fosfomycin (3 g PO once if catheter removed) if uncomplicated§
Bacteremia*			
PCN/ampicillin susceptible*	PCN (4 million IV q 4h) or ampicillin (2 g IV q 4-6h) for 7-14 days	PCN (4 million IV q 4h) or ampicillin (2 g IV q4-6h) for 7-14 days	Consider consultation with an infectious disease specialist Linezolid (600 mg PO/IV q 12h) or Daptomycin (8-12 mg/kg IV q 24h) for 14 days
With PCN allergy or PCN/ampicillin resistance	Vancomycin load with 20-25mg/kg followed by 15-20 mg/kg q 8-12 h using actual body weight to achieve a serum level of 15-20 mcg/mL (trough) Consider adding gentamicin if prolonged bacteremia and laboratory reports synergy with cell wall–active agent	Vancomycin load with 20-25mg/kg followed by 15-20 mg/kg q 8-12 h using actual body weight to achieve a serum level of 15-20 mcg/mL (trough) Consider adding gentamicin if prolonged bacteremia and laboratory reports synergy with cell wall–active agent	Linezolid (600 mg PO/IV q 12h) or Daptomycin (8-12 mg/kg IV q 24h) for 14 days
Endocarditis	Consider consultation with an infectious disease specialist	Consult an infectious disease specialist	Consult an infectious disease specialist
PCN/ampicillin and aminoglycoside susceptible*	Combination therapy indicated PCN (4 million IV q 4h) or ampicillin (2 g IV q 4h) PLUS gentamicin (1 mg/kg IV q 8h) both for 4‡-6 wk Ampicillin (2 g IV q4h) or PCN G (4 million IV q4h) PLUS ceftriaxone (2 g V q 12) for 4-6 wk	Combination therapy indicated PCN (4 million IV q 4h) or ampicillin (2 g IV q 4h) PLUS gentamicin (1 mg/kg IV q 8h) both for 4‡-6 wk	Combination therapy indicated PCN (4 million IV q 4h) or ampicillin (2 g IV q 4h) PLUS gentamicin (1 mg/kg IV q 8h) both for 6 wk or linezolid (600 mg PO/IV q 12h) PLUS another active agent or Daptomycin 8-12 mg/kg IV q 24h PLUS gentamicin (1 mg/kg IV q 8h) both for 6 wk
PCN/ampicillin susceptible and gentamicin resistant*	Consult an infectious disease specialist Combination therapy indicated PCN (4 million IV q 4h) or ampicillin (2 g IV q 4h) PLUS ceftriaxone (2 g IV q 24) both for 4‡-6 wk or combination therapy indicated PCN (4 million IV q 4h) or ampicillin (2 g IV q 4h) PLUS streptomycin (7.5 mg/kg IV q 12h) both for 4‡-6 wk	Consult an infectious disease specialist Combination therapy indicated PCN (4 million IV q 4h) or ampicillin (2 g IV q 4h) PLUS streptomycin (7.5 mg/kg IV q 12h) both for 4‡-6 wk	Consult an infectious disease specialist Combination therapy indicated PCN (4 million IV q 4h) or ampicillin (2 g IV q 4h) PLUS streptomycin (7.5 mg/kg IV q 12h) both for 6 wk
or	Ampicillin (2 g IV q4h) or PCN G (4 million IV q4h) PLUS ceftriaxone (2 g IV q 12) for 4-6 wk		Linezolid (600 mg PO/IV q 12h) PLUS another active agent or Daptomycin 8-12 mg/kg IV q 24h PLUS streptomycin (7.5 mg/kg IV q 12h) both for 6 wk
PCN/ampicillin susceptible and gentamicin intolerant	Consult an infectious disease specialist Strongly consider PCN desensitization if PCN allergy anaphylactic Ampicillin (2 g IV q4h) or PCN G (4 million IV q4h) PLUS ceftriaxone (2 g IV q 12) for 4-6 wk or	Consult an infectious disease specialist Strongly consider PCN desensitization if PCN allergy anaphylactic or	Consult an infectious disease specialist Combination therapy indicated PCN (4 million IV q 4h) or ampicillin (2 g IV q 4h) PLUS streptomycin (7.5 mg/kg IV q 12h) both for 6 wk
PCN allergic	Vancomycin load with 20-25mg/kg followed by 15-20 mg/kg q 8-12 h using actual body weight to achieve a serum level of 15-20 mcg/mL (trough) PLUS gentamicin (1 mg/kg IV q 8h) both for 4‡-6 wk	Vancomycin load with 20-25mg/kg followed by 15-20 mg/kg q 8-12 h using actual body weight to achieve a serum level of 15-20 mcg/mL (trough) PLUS gentamicin (1 mg/kg IV q 8h) both for 4‡-6 wk	Linezolid (600 mg PO/IV q 12h) PLUS another active agent or Daptomycin (8-12 mg/kg IV q 24h) PLUS either gentamicin (1 mg/kg IV q 8h) or streptomycin (7.5 mg/kg IV q 12h) both for 6 wk
PCN/ampicillin resistant and gentamicin susceptible*	Consult an infectious disease specialist Vancomycin load with 20-25mg/kg followed by 15-20 mg/kg q 8-12 h using actual body weight to achieve a serum level of 15-20 mcg/mL (trough) PLUS gentamicin (1 mg/kg IV q 8h) both for 4‡-6 wk	Consult an infectious disease specialist Vancomycin load with 20-25mg/kg followed by 15-20 mg/kg q 8-12 h using actual body weight to achieve a serum level of 15-20 mcg/mL (trough) PLUS gentamicin (1 mg/kg IV q 8h) both for 4‡-6 wk	Consult an infectious disease specialist Linezolid (600 mg PO/IV q 12h) PLUS another active agent or Daptomycin 8-12 mg/kg IV q 24h PLUS either gentamicin (1 mg/kg IV q 8h) or streptomycin (7.5 mg/kg IV q 12h) both for 6 wk
PCN/ampicillin resistant and gentamicin resistant*	Consult an infectious disease specialist Linezolid (600 mg PO/IV q 12h) or Daptomycin (8-12 mg/kg IV q 24h) for a minimum of 6 wk Failures reported; consider surgery	Consult an infectious disease specialist Linezolid (600 mg PO/IV q 12h) or Daptomycin (8-12 mg/kg IV q 24h) for a minimum of 6 wk Failures reported; consider surgery	Consult an infectious disease specialist Linezolid (600 mg PO/IV q 12h) PLUS another active agent or Daptomycin (8-12 mg/kg IV q 24h) PLUS streptomycin (7.5 mg/kg IV q 12h) if susceptible both for minimum 6 wk Failures reported; consider surgery

IV = intravenous; PCN = penicillin; PO = oral; q = every; TCN = tetracycline; TID = three times a day; UTI = urinary tract infection.
*PCN is susceptible if the minimal inhibitory concentration (MIC) ≤8 mcg/mL *and* gentamicin MIC ≤500 mcg/mL.
†Do not use with creatinine clearance less than 50 mL/min.
‡Treat for 4 weeks only when symptoms have been present for less than 3 months *and* there is a prompt response to therapy.
§For complicated UTI use 3q 12 ms (1 sachet) PO every 2-3 days (up to 21 days).

GENERAL REFERENCES

1. Arias CA, Murray BE. The rise of enterococcus: beyond vancomycin resistance. *Nat Rev Microbiol*. 2012;10:266-278.
2. Sievert DM, Ricks P, Edwards JR, et al. Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009-2010. *Infect Control Hosp Epidemiol*. 2013;34:1-14.
3. Heintz BH, Halilovic J, Christensen CL. Vancomycin-resistant enterococcal urinary tract infections. *Pharmacotherapy*. 2010;30:1136-1149.
4. Solomkin JS, Mazuski JE, Bradley JS, et al. Diagnosis and management of complicated intra-abdominal infection in adults and children: guidelines by the Surgical Infection Society and the Infectious Disease Society of America. *Clin Infect Dis*. 2010;50:133-164.
5. Bradley JS. Which antibiotic for resistant Gram-positives, and why? *J Infect*. 2014;68(suppl 1):S63-S75.
6. Patel R, Gallagher JC. Vancomycin-resistant enterococcal bacteremia pharmacotherapy. *Ann Pharmacother*. 2015;49:69-85.
7. Noto MJ, Domenico HJ, Byrne DW, et al. Chlorhexidine bathing and health care-associated infections: a randomized clinical trial. *JAMA*. 2015;313:369-378.

REVIEW QUESTIONS

1. Which of the following is a characteristic of the growth of Enterococci?

A. The ability to grow in NaCl broth
B. The β-hemolytic reaction on sheep blood agar
C. The lack of growth on media containing bile
D. The lack of esculin hydrolysis

Answer: A Enterococci are facultative anaerobes that grow optimally at 35°-37 °C and are usually β-hemolytic or nonhemolytic on sheep blood agar. Enterococci can grow in broth containing 6.5% NaCl and hydrolyze esculin in the presence of 40% bile salts (bile-esculin medium) which can distinguish them from most streptococci.

2. Vancomycin-resistant enterococci (VRE) are the result of

A. the acquisition of VanA gene from VRSA (vancomycin-resistant *Staphylococcus aureus*).
B. the selective pressure of antimicrobials on enterococci, including vancomycin.
C. an efflux pump highly active on vancomycin.
D. a porin loss limiting the vancomycin action on the cell wall

Answer: B An example of selective pressure was demonstrated in Europe when the use of specific antimicrobials in animals led to the development of resistance.

3. Among the following antimicrobials, which one is not a treatment of enterococci?

A. Daptomycin
B. Ampicillin
C. Vancomycin
D. Clindamycin

Answer: D (See section on Treatment.)

4. Enterococci are commonly associated will all of these clinical manifestations except

A. wound infection.
B. urinary tract infection.
C. bacteremia.
D. osteomyelitis.

Answer: D Enterococci are known to cause urinary tract infections, intra-abdominal abscesses, wound infections, bacteremia, including catheter-associated bloodstream infections, and endocarditis; less commonly, they cause osteoarticular infections and meningitis.

5. All of the answers below are needed to control the transmission of VRE among hospitalized patients except

A. hand hygiene.
B. contact precaution.
C. screening and treatment of colonized patients with effective antimicrobials.
D. antimicrobial stewardship programs.

Answer: C Screening or surveillance is important to identify the colonized patient and to assure isolation in contact precautions but treatment of colonization is not recommended. Limiting antimicrobial use will help decrease the selective pressure.

292

DIPHTHERIA AND OTHER *CORYNEBACTERIUM* INFECTIONS

ROLAND W. SUTTER

DEFINITION

Diphtheria is an acute infectious disease caused by toxigenic *Corynebacterium diphtheriae,* a gram-positive bacillus. The hallmark of the disease is the presence of a thick, firmly adherent pseudomembrane at the site of infection. The organism primarily infects the mucosa of the nose, pharynx, tonsils, or larynx

(respiratory diphtheria). Rarely, other mucosal sites may be infected (e.g., conjunctiva, genitals, or ear). In developing countries, a variety of indolent skin lesions (cutaneous diphtheria) are common. Absorption of toxin can result in severe complications such as life-threatening myocarditis or polyneuritis.

The Pathogen

C. diphtheriae is a member of a group of aerobic, nonmotile, unencapsulated, nonsporulating, pleomorphic gram-positive bacilli.[1] Its name comes from the Greek *korynee* (meaning "club"), which describes the shape of the organism on stained smears with one end usually being wider, and *diphtheria* (meaning "leather hide"), for the characteristic adherent membrane.[2] The genus *Corynebacterium* is characterized by bacilli that line up in parallel groups and bend when dividing to create "Chinese character" arrangements. Both nontoxigenic and toxigenic *C. diphtheria* strains exist. Toxigenicity is conferred when a nontoxigenic organism is infected with a β-phage carrying the gene for the toxin (*tox*). *C. diphtheriae* has four biotypes—*gravis, mitis, intermedius,* and *belfanti*—that are distinguished by colonial morphology and varying biochemical and hemolytic reactions. Strains may be distinguished for epidemiologic purposes by molecular techniques. Diphtheria toxin-producing strains of *Corynebacterium ulcerans* can produce classic respiratory diphtheria-like disease, including distal toxic complications.

EPIDEMIOLOGY

Humans are the only natural reservoir of *C. diphtheriae,* although the organism has occasionally been isolated from a variety of domestic and other animals, including horses. Spread occurs in close-contact settings through respiratory droplets or by direct contact with respiratory secretions or skin lesions. The organism may survive for weeks and possibly months on environmental surfaces and in dust, and fomite transmission can occur. The majority of nasopharyngeal *C. diphtheriae* infections may abort or result in asymptomatic carriage, with clinical disease developing in only about one in seven individuals. However, asymptomatic carriers are important in maintaining transmission.

In the prevaccine era, respiratory disease dominated in temperate climates, with a fall and winter peak in incidence. Most individuals acquired natural immunity by the midteen years. Cutaneous disease is more common in tropical countries, but the contribution of cutaneous diphtheria in inducing or maintaining diphtheria immunity in tropical countries is unknown. Over the past 3 decades, outbreaks of cutaneous diphtheria have occurred in the United States and Europe, typically in homeless and alcoholic inner-city adults.

Diphtheria immunization protects against disease but does not prevent carriage. Vaccination with diphtheria toxoid (formalin-treated toxin) was introduced in the 1920s. Immunization of children in an era when the majority of older individuals had natural immunity resulted in a dramatic drop in the incidence of diphtheria and an even more rapid decline in the proportion of toxigenic strains isolated, presumably because the selective advantage of the *tox* gene—promotion of greater replication and spread of the organism—is lost in an immune host. In the postvaccine era, the respiratory diphtheria has virtually been eliminated from developed countries with excellent childhood vaccination coverage.[3] In the United States, reported cases fell from 147,991 in 1920, to 15,536 in 1940, to a total of 55 cases from 1980 to 2012. The absence of reported diphtheria cases in the United States in recent years, however, does not indicate that circulation of toxigenic *C. diphtheriae* has ceased. Investigations in a Northern Plains Indian community in North Dakota and First Nations communities in Ontario, Canada, suggested that *C. diphtheriae* strains might have circulated independently for more than 2 decades in these communities despite the absence of reported respiratory diphtheria cases.

In the absence of natural environmental boosting, vaccine-induced immunity wanes with increasing age and duration since a previous vaccination dose. There is a growing cohort of individuals with no natural diphtheria immunity. Serosurveys indicate that 20% to 60% of adults in industrialized countries have diphtheria antitoxin levels below minimal protective levels. A level of 0.01 IU/mL from an in vitro neutralization assay, the "gold standard" test, is considered the lower limit of protection. Long-term protection against diphtheria requires a level of greater than 0.1 IU/mL. As long as a high proportion of the population remains susceptible, the danger of reintroduction or reemergence of toxigenic strains exists.

Large outbreaks of diphtheria have occurred in the vaccine era. In the 1990s, there was a major resurgence of diphtheria in several countries of the former Soviet Union. In Russia, the number of reported cases rose from 593 in 1989 to 39,582 in 1994, with more than two thirds of cases occurring in adults. Large-scale campaigns of mass administration of diphtheria toxoid to virtually the entire population in the affected new independent states of the former Soviet Union have since led to significant decreases in the incidence of diphtheria, from a peak of 50,449 cases in 1995, to 7197 cases in 1997, to preresurgence levels by the late 1990s (Fig. 292-1). A large outbreak occurred in Indonesia from 2010 to 2012 and in Thailand in 2012.[4] Low vaccination coverage, and non- or inadequate vaccination have been the underlying causes. In general, outbreaks of diphtheria may occur in a suceptible population caused by clonal spread of the organism or transfer of the bacteriophage to nontoxigenic strains of *C. diphtheriae*. Although the reported diphtheria cases have declined globally in the past decade, some countries (e.g., India) continue to have endemic foci.

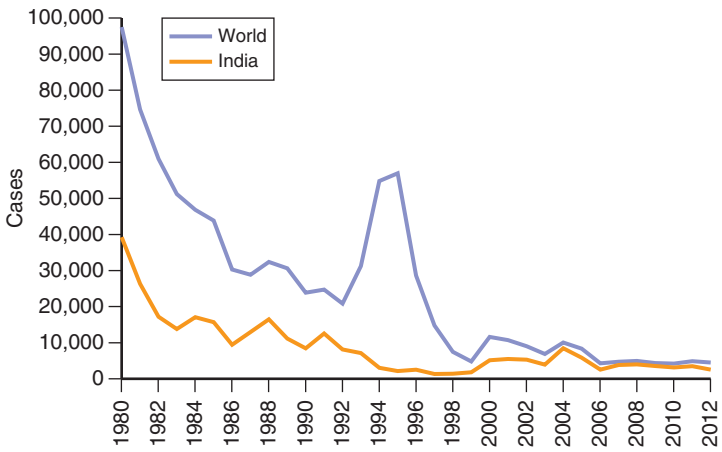

FIGURE 292-1. Reported diphtheria cases, worldwide, 1980 to 2012. (Data from the World Health Organization.)

PATHOBIOLOGY

In classic respiratory diphtheria, *C. diphtheriae* colonizes the mucosal surface of the nasopharynx or larynx and multiplies locally without blood stream invasion.[5] The symptoms and signs of diphtheria are attributable to toxin production. Diphtheria toxin is an extremely potent inhibitor of protein synthesis, and the estimated human lethal dose is 0.1 mg/kg. Released toxin causes local tissue necrosis with the formation of a tough pseudomembrane composed of a mixture of fibrin, dead cells, and bacteria that is firmly adherent to the underling submucosal tissue. The membrane usually begins on the tonsils, on the posterior pharynx, or in the nose. In more severe cases, it progressively extends over the pharyngeal wall, fauces, and soft palate and into the larynx and may result in respiratory obstruction. Toxin entering the blood stream causes tissue damage at distant sites, particularly the heart (myocarditis), nerves (demyelination), and kidney (tubular necrosis). The extent of toxin absorption varies with the site of infection, being much less from the skin or nose than from the pharynx. Nontoxigenic strains may cause mild local respiratory disease and rarely a membrane.

CLINICAL MANIFESTATIONS

Respiratory Diphtheria

Infection limited to the anterior nares (nasal diphtheria) is manifested as a chronic serosanguineous or seropurulent discharge without fever or significant toxicity. A whitish membrane may be observed on the septum. The faucial (pharyngeal) form is most common. After an incubation period of 1 to 7 days, the illness begins with a sore throat, malaise, and mild to moderate fever. Initially, there is mild pharyngeal erythema, usually followed by progressive formation of a whitish tonsillar exudate, which over a period of 24 to 48 hours consolidates into a firmly adherent grayish membrane that bleeds on attempted removal. In more severe cases, the patient appears toxic, and the membrane is more extensive. Cervical lymphadenopathy and soft tissue edema may occur and result in the typical bull neck appearance and stridor. Laryngeal involvement (laryngeal diphtheria), which may develop on its own or as a result of membrane extension from the nasopharynx, is manifested as hoarseness, stridor, and dyspnea. The following clinical classification has been proposed by the World Health Organization: (1) catarrhal form (erythema of pharynx, no membrane), (2) follicular form (patches of exudate,

no pharynx or tonsillar involvement), (3) spreading form (membranes covering the tonsils and the posterior pharynx), and (4) combined form (more than one anatomical site involved, e.g., throat and skin). The more severe clinical manifestations are associated with increasing toxin absorption levels.

The likelihood of toxic complications depends primarily on the interval between disease onset and administration of antitoxin. The severity of disease at initial evaluation closely predicts the likelihood of a severe clinical course, complications, and death. Myocarditis typically occurs in the first or second week after the onset of respiratory symptoms and develops either suddenly or insidiously with signs of low cardiac output and congestive failure. Conduction disturbances, which may occur without other signs of myocarditis, include ST-T wave abnormalities, arrhythmias, and heart block. Neurologic impairment is manifested as cranial nerve palsies and peripheral neuritis.[6] Palatal or pharyngeal paralysis (or both) occurs during the acute phase; peripheral neuritis, symmetrical and predominantly motor, occurs 2 to 12 weeks after onset of the disease. Motor deficit may range from minor proximal weakness to complete paralysis. Complete recovery is the rule. In fulminant, sometimes called "hypertoxic," diphtheria, toxic circulatory collapse with hemorrhagic features occurs.

Cutaneous Diphtheria

Cutaneous diphtheria lesions are classically indolent, deep, punched-out ulcers that may have a grayish-white membrane. However, the lesions may be indistinguishable from impetigo, or *C. diphtheriae* may infect chronic dermatoses such as stasis dermatitis. Coinfection with *Streptococcus pyogenes, Staphylococcus aureus,* or both occurs frequently. Toxic complications of only cutaneous diphtheria are rare.

DIAGNOSIS

A high index of suspicion is required. Specimens for culture should be taken from beneath the membrane, from the nasopharynx, and from any suspicious skin lesions. Because special media are required, the laboratory should be alerted to the concern about diphtheria. *C. diphtheriae* is best isolated on selective media that inhibit the growth of other nasopharyngeal organisms; one containing potassium tellurite is generally used. Based on colonial morphology and Gram stain appearance, a presumptive diagnosis may be possible within 18 to 24 hours. Culture results may be negative if the patient previously received antibiotics. Toxigenicity testing should be performed on all *C. diphtheriae* isolates. Because both nontoxigenic and toxigenic strains may be isolated from the same patient, more than one colony should be tested. Traditional testing methods include guinea pig inoculation and the modified Elek test, in which the isolate and appropriate controls are streaked on a culture plate in which a filter strip soaked with antitoxin has been embedded; toxin production is confirmed by an immunoprecipitation line in the agar. Identification of the diphtheria tox gene allowed the development of rapid and accurate polymerase chain reaction–based methods for identification of toxigenic strains.

Differential Diagnosis

The differential diagnosis includes streptococcal and viral tonsillopharyngitis, infectious mononucleosis, Vincent's angina, candidiasis, and acute epiglottitis. A history of travel to a region with endemic diphtheria or a history of contact with a recent immigrant from such an area increases the suspicion of diphtheria, as does a pre-antitoxin treatment serum antitoxin level of less than 0.01 IU/mL.

TREATMENT

Rx

The decision to initiate therapy should be based on clinical grounds because delayed treatment, especially delays in antitoxin administration, is associated with worse outcomes. The goals of treatment are to neutralize the toxin rapidly, eliminate the infecting organism, provide supportive care, and prevent further transmission (Table 292-1). The mainstay of therapy is equine diphtheria antitoxin. Because only unbound toxin can be neutralized, treatment should commence as soon as the diagnosis is suspected, and each day of delay in administration increases the likelihood of a fatal outcome. A single dose ranging in quantity from 20,000 units for localized tonsillar diphtheria up to 100,000 units is given for extensive disease with severe toxicity. Antitoxin may be administered intramuscularly or intravenously; particularly for more severe cases, the intravenous route is preferred. Tests for sensitivity to antitoxin should be performed according to package insert instructions before administering it and desensitization carried out if necessary. Antibiotic therapy, by eliminating the organism, halts toxin production, limits local infection, and prevents transmission. Parenteral penicillin (4-6 million U/day) and erythromycin (40 mg/kg/day in four divided doses; maximum of 2 g/day, usually orally if the patient can swallow) are the drugs of choice. General supportive care includes ensuring a secure airway (with tracheotomy, if necessary), electrocardiographic monitoring for evidence of myocarditis, treating heart failure and arrhythmias, and preventing secondary complications of neurologic impairment such as aspiration pneumonia. The patient should be in strict isolation until follow-up culture results are negative. Convalescing patients should receive diphtheria toxoid.

The local health department must be notified. Close contacts should have cultures performed and be administered prophylactic antibiotics and brought up to date with an age-appropriate diphtheria toxoid-containing vaccination. A positive culture result in a contact may confirm the diagnosis if the patient is culture negative. All contacts without full primary immunization and a booster within the preceding 5 years should receive diphtheria toxoid.

The availability and access to diphtheria antitoxin has become problematic in recent years. Currently, very few manufacturers exist. In the United States, because manufacturers discontinued diphtheria antitoxin production in 1997, no licensed product is available. However, diphtheria antitoxin for therapeutic purposes can be obtained from the Centers for Disease Control and Prevention (CDC; 770-488-7100), which distributes a Brazilian-produced antitoxin (Instituto Butantan, Sao Paulo, Brazil) under an Investigational New Drug protocol.

TABLE 292-1 GOALS AND PROPOSED INTERVENTIONS FOR THE MANAGEMENT OF SUSPECTED DIPHTHERIA CASES

GOALS	PROPOSED INTERVENTIONS
Neutralize toxin as soon as possible to reduce severe complications, including death.	After a presumptive diagnosis of diphtheria, immediately obtain and administer antitoxin, initiate antimicrobial treatment, and arrange for appropriate supportive care.
Prevent further spread of *Corynebacterium diphtheriae* to close contacts, including hospital staff.	Isolate the patient; strictly observe respiratory barrier procedures. Notify the health department. Review the vaccination status of the family and other close contacts and initiate postexposure prophylaxis.
Confirm the diagnosis.	Collect appropriate specimens for culture (alert the laboratory to ensure that it can prepare specific culture media).
Induce long-term protection against *C. diphtheriae* in case and close contacts.	Complete the primary series with diphtheria toxoid as needed.

PREVENTION

Immunization with diphtheria toxoid is the only effective means of primary prevention. The primary series is four doses of diphtheria toxoid (given with tetanus toxoid and pertussis vaccine) at 2, 4, 6, and 15 to 18 months; a preschool booster dose is given at 4 to 6 years of age. Thereafter, boosters should be given as part of the adolescent immunization visit (i.e., between 11 and 13 years of age) followed by doses administered every 10 years.

For additional protection against pertussis during adolescence, the CDC recommends the routine use of tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis vaccine, adsorbed (Tdap), in adolescents 11 to 18 years of age in place of tetanus and diphtheria toxoid (Td) vaccines. Similarly, for adults who have never received a dose of Tdap, the CDC recommends routine use of a single dose of Tdap for adults older than 19 years of age to replace the next booster dose of Td (Chapter 18), although a second dose five years later may be needed to protect against diphtheria.[7]

Genetically altered, fully immunogenic mutants of diphtheria toxin have been created (e.g., CRM197) CRM197 is used as a protein carrier in several polysaccharide–protein conjugate vaccines. However, the role of CRM197 in contributing to or maintaining immunity has not been evaluated.

PROGNOSIS

Diphtheria, at the beginning of the 21st century, remains a serious disease associated with a high case-fatality rate. In the United States, the diphtheria

TABLE 292-2 CORYNEBACTERIA AND RELATED ORGANISMS ASSOCIATED WITH HUMAN DISEASE

SITE OF INFECTION	PATHOGEN	CLINICAL SYNDROME	COMMENTS
Respiratory tract	*Corynebacterium diphtheriae*	Classic diphtheria	
	Corynebacterium ulcerans	Diphtheria	Zoonotic infection; may produce diphtheria toxin
	Corynebacterium pseudodiphtheriticum	Pharyngitis Rarely, pneumonia in patients with advanced AIDS	
	Arcanobacterium haemolyticum	Pharyngitis, tonsillar abscess, rash	Clinically indistinguishable from streptococcal pharyngitis
Skin and soft tissue	*Corynebacterium pseudotuberculosis*	Granulomatous lymphadenitis	Zoonotic infection, especially in sheep; occupational risk for veterinarians and butchers
	Corynebacterium minutissimum	Erythrasma	
	Corynebacterium kroppenstedtii	Granulomatous breast abscess	
Genitourinary tract	*Corynebacterium glucuronolyticum*	UTIs in men; chronic prostatitis	
	Corynebacterium urealyticum	Chronic and recurrent UTIs	More common in elderly, chronically ill, and immunosuppressed patients and in those with indwelling catheters
	Corynebacterium riegelli	Encrusted cystitis	
Health care–associated infections	*Corynebacterium jeikeium* and less commonly many others, including *Corynebacterium amycolatum, Corynebacterium striatum,* and *Corynebacterium urealyticum*	Catheter- and device-associated infections Postprocedure wound and soft tissue infections Prosthetic valve joint infections Nosocomial pneumonia CSF shunt infections	*C. jeikeium* is the most common corynebacterial pathogen in hospitals and causes severe infections in immunosuppressed patients and those with indwelling devices

CSF = cerebrospinal fluid; UTI = urinary tract infection.

case-fatality rate has remained virtually unchanged (between 5% and 10%) over recent decades.

OTHER *CORYNEBACTERIUM* SPECIES

Corynebacteria other than *C. diphtheriae* are ubiquitous in the environment and are among the normal flora colonizing humans and animals. The pathogenic potential of many of these organisms was not appreciated in the past, but many are now known to be associated with specific and often serious infectious diseases, especially in immunosuppressed, chronically ill, and hospitalized patients (Table 292-2). In general, these organisms remain susceptible to vancomycin, but resistance to other classes of antimicrobials is common and varies among species.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Poetsch A, Haussmann U, Burkovski A. Proteomics of corynebacteria: from biotechnology workhorses to pathogens. *Proteomics*. 2011;11:3244-3255.
2. Nurkovski A. Cell envelope of corynebacteria: structure and influence on pathogenicity. *ISRN Microbiol*. 2013;2013:935736.
3. Zakikhany K, Efstratiou A. Diphtheria in Europe: current problems and new challenges. *Future Microbiol*. 2012;7:595-607.
4. Wanlapakorn N, Yoocharoen P, Tharmaphornpilas P, et al. Diphtheria outbreak in Thailand, 2012: seroprevalence of diphtheria antibodies among Thai adults and its implications for immunization programs. *Southeast Asian J Trop Med Public Health*. 2014;45:1132-1141.
5. Rogers EA, Das A, Ton-That H. Adhesion by pathogenic corynebacteria. *Adv Exp Med Biol*. 2011; 715:91-103.
6. Sanghi V. Neurologic manifestations of diphtheria and pertussis. *Handb Clin Neurol*. 2014;121: 1355-1359.
7. Weinberger B, Schirmer M, Matteucci Gothe R, et al. Recall responses to tetanus and diphtheria vaccination are frequently insufficient in elderly persons. *PLoS ONE*. 2013;8:e82967.

REVIEW QUESTIONS

1. A 61-year-old Vietnam veteran with a low socioeconomic status (SES) in Seattle is brought to the emergency department (ED) because of difficulties breathing. On radiography, a lobar pneumonia is demonstrated, and appropriate treatment with antibiotics is initiated. The attending ED physician also notes a rash on the lower extremity. The lesions are deep punched-out non-itching ulcers that are indolent. On culture, *Staphylococcus aureus* is grown. What is the likely diagnosis of the skin rash?

A. Impetigo
B. Cutaneous diphtheria
C. Chronic dermatitis superinfected by *S. aureus*
D. Cat scratch disease
E. Scabies.

Answer: B Cutaneous lesions caused by *Corynebacterium diphtheriae* are typically deep punched-out ulcers in a patient of risk (low SES, probably alcohol abuse, poor hygiene). The isolation *S. aureus* (or other skin bacteria) occurs frequently. *C. diphtheriae* may not be cultured initially because of the need for special media. Impetigo and chronic dermatitis are possible diagnoses but are unlikely given the typical appearance of the lesions (usually honey-colored scabs). Similarly, cat scratch disease is unlikely in the absence of a supporting history of contact with cats. Scabies causes a papular rash with intense itching.

2. A 23-year-old Peace Corps volunteer returns from a 2-year assignment in the rural East Java Province of Indonesia. She presents with sore throat, malaise, and mild to moderate fever for 3 days. Before leaving Indonesia, the patient received erythromycin. On examination, a whitish membrane adheres tightly to the right tonsil. During the swapping of the tonsil for culture, the membrane cannot be removed and starts to bleed at the edge. What is the most likely diagnosis?

A. Streptococcal tonsillitis
B. Plaut Vincent angina
C. Respiratory diphtheria
D. Primary syphilis lesion
E. Infectious mononucleosis (MN) caused by Epstein-Barr virus (EBV)

Answer: C A greyish white pseudomembrane on a tonsil should lead to a high suspicion for diphtheria, as does the history of an extended stay in a developing country where diphtheria is endemic. None of the other diagnoses are associated with the formation of a pseudomembrane.

3. Given a high index of suspicion of respiratory diphtheria in this 23-year-old patient (see question 2), what is the immediate priority?

A. Report the suspected case to the local health department.
B. Inform the laboratory to use special media for culturing (potassium tellurite media).
C. Contact the Centers for Disease Control and Prevention and order equine diphtheria antitoxin.
D. Check the patient's diphtheria toxoid vaccination history.
E. Isolate the patient.

Answer: E In order of priority, it is essential to ensure that the airways are patent. In this case, we have no indication that these are compromised. After the diagnosis of suspected diphtheria is established, the patient needs to be isolated to prevent spread to contacts. The isolation of *Corynebacterium diphtheriae* requires a special media, so the laboratory has to be alerted. Given that the patient has been taking antibiotics for several days, it is unlikely the culture result will be positive. However, because of the potential complications of diphtheria, treatment with equine diphtheria antitoxin should be initiated on the basis of suspicion as soon as possible. Then the local health department should be contacted.

4. Which one is not a likely complication of diphtheria toxin?

A. Myocarditis
B. Kidney (tubular necrosis)
C. Neuropathy (caused by demyelination)
D. Encephalitis
E. Respiratory insufficiency

Answer: D Encephalitis is not a likely complication of diphtheria. Myocarditis occurs up to several weeks after the onset of acute respiratory diphtheria and is associated with a high fatality rate. The kidneys may be affected but recover. The most common form of neuropathy is palatal paralysis or peripheral neuropathy in the lower extremity; both are reversible. Extension of diphtheric membranes into the respiratory tree can lead to respiratory stress and even insufficiency.

5. What recommendations for diphtheria antitoxin vaccination would you make to a 26-year-old oil worker going to the Middle East for several months? The man had received two doses of diphtheria, tetanus, and pertussis (Tdap) vaccine as a child and is leaving the United States in 6 weeks.

A. Give a dose of diphtheria toxoid-containing vaccine (d).
B. Administer a dose of tetanus and diphtheria toxoid (Td).
C. Complete the vaccination schedule with adult-formulation Tdap vaccine before departure.
D. Given the possible adverse events and the unlikely exposure to diphtheria in the oil fields, no further immunization is recommended.
E. Complete the schedule with infant-formulation Tdap vaccine before departure.

Answer: C To have protective antitoxin levels, a full primary series of diphtheria toxin–containing vaccine is needed. In this instance, administer a dose now and a second dose 4 weeks later. A single dose of diphtheria toxoid would be suboptimal and would be a missed opportunity for vaccinating against tetanus and pertussis.

6. Which of the following statements about diphtheria immunization is correct?

A. DTP (diphtheria and tetanus toxoid and pertussis vaccine) should not be administered to pregnant women because of the risk of adverse pregnancy outcomes.
B. Diphtheria-specific IgG levels are higher in infants when their mothers were immunized with DTP 2 years before delivery than when their mothers received DTP in midpregnancy.
C. Adults require boosters of diphtheria toxoid every 10 years even after completing a full series of infant immunizations plus preschool and adolescent boosters.
D. A single booster administration in an adult who had never received DTP immunizations in childhood will provide adequate protection.

Answer: D For adults older than 19 years of age who have never received a dose of DTP, the U.S. Centers for Disease Control and Prevention still recommends the routine use of a single dose of DTP to replace the next booster dose. The Vaccine Adverse Event Reporting System has not identified any concerning events in maternal, infant, or fetal outcomes in pregnant women who received DTP. It has been demonstrated that protective IgG levels in the infants of women immunized with DTP within the past 2 years before delivery is the same as those in infants whose mothers were immunized in the second or third trimester. In fact, it is now suggested that vaccinating mothers in the late second or third trimester of pregnancy is safe and would more likely allow maternal antibody production to reach higher protective levels (to be transferred across the placenta to the fetus) by the time of delivery. The diphtheria prevention protocol involves a primary series of 4 doses of diphtheria toxoid (in the form of DPT) to be administered at 2, 4, 6, and 15 to 18 months; a preschool booster dose to be given at 4 to 6 years of age, and subsequent boosters to be given as part of adolescent immunization visits (between 11 and 13 years of age) followed thereafter by doses administered every 10 years. (See section on Prevention and Kao CM, Schneyer RJ, Bocchini JA. Child and adolescent immunizations: selected review of recent US recommendations and literature. *Office Pediatr*. 2014;26:383-395.)

7. Which of the following statements about diphtheria is correct?

A. Outbreaks of diphtheria can be caused either by clonal spread of toxigenic strains or by transfer of the gene for the toxin via bacteriophage to nontoxigenic strains.
B. The case-fatality rate for diphtheria has decreased dramatically since the advent of childhood immunization for the disease.
C. The extent of diphtheria toxin absorption into the circulation is not dependent on the primary site of infection.
D. Nontoxigenic strains of *Corynebacterium diphtheriae* do not cause clinical illness in humans.
E. Cutaneous diphtheria in North America and Europe occurs almost exclusively in severely immunocompromised individuals.

Answer: A Outbreaks of serious illness can be caused by either toxigenic strains or nontoxigenic strains that have been made toxigenic by transfer of the toxin gene by phages. Infection by nontoxigenic strains of *C. diphtheriae* can cause symptoms, although typically much milder than those caused by toxigenic strains. Although the incidence of diphtheria has plummeted throughout most of the world since the advent of childhood immunization, the case-fatality rate has not changed much; it remains a very serious disease (with case-fatality of 5% to 10% in the United States). The extent of toxin absorption varies with the site of infection, being much less from the skin or nose than from the pharynx. Outbreaks of cutaneous diphtheria have occurred in the United States and Europe typically in homeless and alcoholic inner-city adults, not necessarily those who are severely immunocompromised.

293

LISTERIOSIS

BENNETT LORBER

DEFINITION

Listeriosis is a food-borne infection caused by the gram-positive rod *Listeria monocytogenes*. Most patients have impaired cell-mediated immunity and are seen with life-threatening bacteremia or meningitis. However, a self-limited, febrile gastroenteritis in healthy persons also occurs.

The Pathogen

Widely distributed in nature, *L. monocytogenes* may be found in soil, on vegetation, and in the stool of healthy mammals, including humans. It causes disease in animals, especially herd animals, and in humans. The organism has been isolated from many foods, including raw vegetables, raw milk, fish, poultry, and meat. Unlike most food-borne pathogens, *L. monocytogenes* can grow at refrigerator temperatures.

EPIDEMIOLOGY

Non-perinatal listeriosis is almost always the result of food-borne infection.[1,2] Listeriosis is a relatively rare food-borne illness (≈1% of U.S. cases) but is associated with a case-fatality rate of 16 to 20% (second only to *Vibrio vulnificus* at 35 to 39%) and causes approximately 19 to 28% of all food-borne disease–related deaths.[3] Outbreaks have been documented in association with coleslaw, milk, soft cheeses, pâté, ready-to-eat pork products, deli counter meats, hot dogs, smoked fish, butter, sprouts, taco or nacho salads, and cantaloupes. In 2011, *L. monocytogenes*–contaminated cantaloupes were responsible for the deadliest food-borne disease outbreak in U.S. history, with 28 states reporting illness in 146 persons and death in 30 (21% mortality).[4]

Listeriosis was made a nationally reportable disease in 2000. The Centers for Disease Control and Prevention has established PulseNet (http://www.cdc.gov/pulsenet/), a network of public health and food regulatory laboratories that use pulsed-field gel electrophoresis to subtype food-borne pathogens to detect promptly disease clusters that may have a common source. Presently, the annual incidence of listeriosis is 3 cases per million population, and it accounts for 1800 cases per year and about 400 deaths. Neonates and adults older than 60 years have the highest infection rates. Pregnant women represent 20% of all affected individuals[5]; other adults at increased risk for invasive listeriosis (bacteremia, meningitis) include those with hematologic malignant disease, advanced acquired immunodeficiency syndrome (AIDS), a solid organ transplant, or iron overload and anyone treated with corticosteroids or an anti–tumor necrosis factor agent. However, as many as a fourth of all cases of invasive listeriosis occur in apparently healthy persons, particularly those older than 60 years.

PATHOBIOLOGY

L. monocytogenes enters the human body through the intestine, most often after the ingestion of contaminated food.[6] The bacterium induces its own uptake by gastrointestinal cells and macrophages. Mother-to-child transmission occurs transplacentally or through an infected birth canal. Within the host cell, the bacterium is enclosed in a phagolysosome, but through the production of an exotoxin called listeriolysin O, it destroys the phagolysosome membrane and gains access to the cytoplasm. All pathogenic strains of *L. monocytogenes* produce listeriolysin O, the major virulence factor. Listeriae actively divide in the cytoplasm, migrate to the periphery of the cell by polymerization of host cell actin, and then push out the cell membrane to form pseudopods, which are taken up by adjacent host cells. The bacteria move from cell to cell in this fashion and repeat their life cycle without exposure to antibodies or complement.

After invasion through the gastrointestinal tract, listeriae may disseminate hematogenously to any body site but show a particular tropism for the central nervous system (CNS). Less commonly, listeriae may spread intra-axonally through cranial nerves to reach the CNS; this mode of CNS invasion may result in rhombencephalitis (brain stem infection).

Immunity to listerial infection is handled chiefly through the cell-mediated arm of the immune system. Persons who have had splenectomy or have abnormalities solely of humoral immunity or leukocytes are not at increased risk for infection.

CLINICAL MANIFESTATIONS

The incubation period for invasive listeriosis (time from ingestion of contaminated food to illness) averages about 30 days. Invasive listeriosis in an immunocompromised adult is most often manifested as bacteremia without an obvious focus. In such cases, patients have nonspecific complaints, such as fever, malaise, myalgia, and back pain. Bacteremia is the form of invasive listeriosis that complicates pregnancy; CNS infection is extremely rare in the absence of other risk factors. Listeriosis during pregnancy may lead to spontaneous abortion or neonatal sepsis, but early antimicrobial therapy may result in the birth of a healthy child.[7] Endocarditis with *L. monocytogenes* can occur on both native and prosthetic valves and carries a high rate of septic complications. Endocarditis, but not bacteremia per se, may be a clue to underlying colon cancer; colonoscopy should be considered in all cases of listerial endocarditis.

Persons in whom *L. monocytogenes* bacteremia develops may progress to CNS infection, most commonly manifested as meningitis. *Listeria* has a predilection for infecting brain tissue as well as the meninges, and unlike other common bacterial causes of meningitis, it not infrequently causes encephalitis or brain abscess. Brain abscess as a result of infection by *L. monocytogenes* exhibits unusual features compared with other bacteria: listerial brain abscess coexists with bacteremia in nearly all cases and with meningitis in a fourth; in addition, abscesses are often subcortical.

L. monocytogenes is the most common cause of bacterial meningitis in patients with lymphomas, organ transplant recipients, and patients treated with corticosteroids for any reason. Affected persons usually have the classic acute symptoms of meningitis, but the presentation is subacute (>24 hours) in 60% of cases. Nuchal rigidity is absent in 20%. Focal neurologic findings, including ataxia, tremors, myoclonus, and seizures, may be seen, consistent with the tropism of *Listeria* for brain parenchyma. Gram stain of cerebrospinal fluid (CSF) reveals small gram-positive rods in only about one third of cases. The glucose content in CSF is normal in more than 60% of cases; mononuclear cells predominate in 30%.

Listerial rhombencephalitis[8] is an unusual form of listerial encephalitis that involves the brain stem and, unlike other listerial CNS infections, usually occurs in healthy adults. The typical clinical picture is one of a biphasic illness with a prodrome of fever, headache, nausea, and vomiting lasting about 4 days, followed by the abrupt onset of asymmetrical cranial nerve deficits, cerebellar signs, and hemiparesis or hemisensory deficits or both. Respiratory failure develops in about 40% of patients. Nuchal rigidity is present in about half, and CSF findings are only mildly abnormal, with a positive CSF culture in about 40%. Almost two thirds of patients are bacteremic. Magnetic resonance imaging is superior to computed tomography for demonstrating rhombencephalitis. Mortality is high, and serious sequelae are common in survivors.

Localized infection may occur after hematogenous seeding (e.g., liver abscess, septic arthritis) or, rarely, after direct inoculation (e.g., papulopustular rash, conjunctivitis). Osteoarticular listeriosis[9] primarily involves prosthetic joints, occurs in immunocompromised patients, and requires implant removal for cure.

Well-documented reports of food-borne outbreaks have demonstrated that ingestion of *L. monocytogenes* in a sufficiently large inoculum can result in a self-limited illness consisting of fever, chills, diarrhea, abdominal cramps, and sometimes nausea and vomiting. Symptoms follow exposure by 1 to 2 days and last for about 2 days.

DIAGNOSIS

Differential Diagnosis

Clinical situations in which a diagnosis of listeriosis should be considered include the following:

- neonatal sepsis or meningitis;
- meningitis or parenchymal brain infection in patients with subacute presentations, hematologic malignant neoplasms, AIDS, organ transplantation, corticosteroid immunosuppression, treatment with anti–tumor necrosis factor agents, or age older than 50 years;
- simultaneous infection of the meninges and brain parenchyma;
- subcortical brain abscess;
- fever during pregnancy;
- blood, CSF, or other normally sterile specimen reported to have "diphtheroids" on Gram stain or culture; and
- food-borne outbreak of febrile gastroenteritis when routine cultures fail to identify a pathogen.

The differential diagnosis of listerial CNS infection includes the more common causes of bacterial meningitis and brain abscess; indolent listerial meningitis or rhombencephalitis may mimic CNS tuberculosis.

Laboratory Findings

The diagnosis of listeriosis is made by routine bacterial culture of specimens from usually sterile sites, such as blood or CSF. The laboratory must exercise caution because *L. monocytogenes* may be mistaken for diphtheroids, streptococci, or enterococci. Specific stool culture is recommended only when routine stool cultures are negative in the setting of an outbreak of gastroenteritis; many people have enteric colonization with *L. monocytogenes* without invasive disease. The laboratory must be advised that listerial infection is suspected because the organism is unlikely to be identified with routine stool culture media.

Serologic testing (antibody to listeriolysin O) is not useful for invasive disease but may be helpful in the retrospective identification of food-borne outbreaks of febrile gastroenteritis when routine cultures are negative. Real-time polymerase chain reaction analysis of CSF for the *hly* gene, which encodes listeriolysin O, has been useful in diagnosing CNS listeriosis, including cases in which routine bacterial cultures were negative, but this test is not yet commercially available.[10]

PREVENTION

Guidelines for preventing listeriosis are similar to those for preventing other food-borne illnesses. In general, one should thoroughly cook raw food from animal sources; wash raw vegetables thoroughly before eating; keep uncooked meats separate from vegetables and from cooked and ready-to-eat foods; avoid raw (unpasteurized) milk or foods made from raw milk; and wash hands, knives, and cutting boards after each handling of uncooked foods.

People at high risk for listeriosis (those immunocompromised by illness or medications, pregnant women, and the elderly) may choose to avoid soft cheeses such as feta, Brie, Camembert, blue veined, and Mexican-style cheese such as queso fresco. Hard cheeses, processed cheeses, cream cheese, cottage cheese, and yogurt are safe. Leftover foods or ready-to-eat foods such as hot dogs should be cooked until steaming hot. It is best to avoid foods from delicatessen counters, such as prepared salads, meats, and cheeses, or at least to reheat cold cuts thoroughly until they are steaming hot before eating.

Listeriosis is effectively prevented by trimethoprim-sulfamethoxazole given as *Pneumocystis* prophylaxis to organ transplant recipients, those receiving corticosteroid immunosuppression, or individuals infected with human immunodeficiency virus. Second episodes of neonatal listerial infection are virtually unheard of, and intrapartum antibiotics are not recommended for women with a history of perinatal listeriosis.

Except for transmission from infected mother to fetus, human-to-human transmission of listeriosis does not occur; patients do not need to be isolated.

TREATMENT

Rx

Recommendations for the treatment of infection with *L. monocytogenes* derive from in vitro data, animal models, and clinical experience with small numbers of patients; no controlled trials have been performed to prove the efficacy of one drug over another. Many antimicrobials show in vitro activity against *L. monocytogenes*. Clinical utility is more relevant than in vitro susceptibility test results because cephalosporins and other drugs to which the bacterium appears to be susceptible are inadequate to treat infection.

Twenty percent of cases of bacterial meningitis in those older than 50 years are due to *L. monocytogenes*. Therefore, empirical therapy for bacterial meningitis in all adults older than 50 years should include either ampicillin or trimethoprim-sulfamethoxazole, especially in the absence of associated pneumonia, otitis, sinusitis, or endocarditis, which would suggest a cause other than *L. monocytogenes*. Cephalosporins, commonly used for the treatment of bacterial meningitis, should not be used alone when *Listeria* is a diagnostic consideration.

Ampicillin is generally considered the drug of choice for treating confirmed cases of listeriosis. In cases of meningitis and endocarditis and in patients with severely impaired T-cell function, many authorities recommend the addition of gentamicin to ampicillin for synergy on the basis of in vitro testing and animal models. For meningitis, therapy should be continued for at least 3 weeks; bacteremic patients without CNS involvement may be treated for 2 weeks. Endocarditis and brain abscess should be treated for at least 6 weeks. Meningitis doses should be used to treat all cases of invasive listeriosis, even in the absence of CNS or CSF abnormalities.

In patients with penicillin hypersensitivity, trimethoprim-sulfamethoxazole is the preferred agent. It is bactericidal and appears to be as effective as the combination of ampicillin and gentamicin. Drugs that should be avoided because of treatment failure and relapse include cephalosporins, chloramphenicol, tetracycline, vancomycin, and erythromycin.

Corticosteroids appear to be important adjunctive agents in treating the most common forms of bacterial meningitis. Their role in the treatment of listerial CNS infection is unknown.

Iron is a virulence factor for *L. monocytogenes*, and clinically, iron overload states are risk factors for listerial infection. Therefore, in patients with listeriosis and iron deficiency, it may be prudent to withhold iron replacement until antimicrobial therapy is complete.

PROGNOSIS

Listeria meningitis carries a mortality of about 25%,[11] and mortality is higher in those with underlying malignant disease. Mortality from brain abscess and endocarditis is about 50%; survivors of brain abscess commonly have significant neurologic residua.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Allerberger F, Wagner M. Listeriosis: a resurgent foodborne infection. *Clin Microbiol Infect*. 2010;16:16-23.
2. Cartwright EJ, Jackson KA, Johnson SD, et al. Listeriosis outbreaks and associated food vehicles, United States, 1998-2008. *Emerg Infect Dis*. 2013;19:1-9.
3. Goulet V, Hebert M, Hedberg C, et al. Incidence of listeriosis and related mortality among groups at risk of acquiring listeriosis. *Clin Infect Dis*. 2012;54:652-660.
4. McCollum JT, Cronquist AB, Silk BJ, et al. Multistate outbreak of listeriosis associated with cantaloupe. *N Engl J Med*. 2013;369:994-953.
5. de Noordhout CM, Devleesschauwer B, Angulo FJ, et al. The global burden of listeriosis: a systematic review and meta-analysis. *Lancet Infect Dis*. 2014;14:1073-1082.
6. Schuppler M, Loessner MJ. The opportunistic pathogen *Listeria monocytogenes:* pathogenicity and interaction with the mucosal immune system. *Int J Inflam*. 2010;2010:704321.
7. Elinav H, Hershko-Klement A, Valinsky L, et al. Pregnancy-associated listeriosis: clinical characteristics and geospatial analysis of a 10-year period in Israel. *Clin Infect Dis*. 2014;59:953-961.
8. Abbs A, Nandakumar T, Bose P, et al. Listeria rhomboencephalitis. *Pract Neurol*. 2012;12:131-132.
9. Charlier C, Leclercq A, Cazenave B, et al. *Listeria monocytogenes*–associated joint and bone infections: a study of 43 consecutive cases. *Clin Infect Dis*. 2012;54:240-248.
10. Le Monnier A, Abachin E, Beretti J-L, et al. Diagnosis of *Listeria monocytogenes* meningoencephalitis by real-time PCR for the *hly* gene. *J Clin Microbiol*. 2011;49:3917-3923.
11. Pelegrín I, Moragas M, Suárez C, et al. Listeria monocytogenes meningoencephalitis in adults: analysis of factors related to unfavourable outcome. *Infection*. 2014;42:817-827.

REVIEW QUESTIONS

1. Which one of the following individuals is at the highest risk for invasive listeriosis?

A. A 25-year-old with common variable hypogammaglobulinemia
B. A 45-year-old being treated with an anti–tumor necrosis factor agent
C. A 38-year-old with a splenectomy
D. A 52-year-old who has had leukemia and neutropenia
E. A 72-year-old on hemodialysis

Answer: B Invasive listeriosis occurs most often in those with impaired cell-mediated immunity due to underlying disease or due to treatment with drugs such as corticosteroids or anti–tumor necrosis factor agents. Splenectomy, disorders of white blood cell numbers or function, and immunoglobulin disorders are not associated with an increased risk.

2. A 66-year-old man presents with fever and altered consciousness. The findings on examination of his cerebrospinal fluid (CSF) are consistent with bacterial meningitis, and CSF Gram stain shows small, gram-positive rods. He had an anaphylactic reaction to amoxicillin 2 years ago. Which one of the following is the best treatment for this patient?

A. Ceftriaxone
B. Vancomycin
C. Imipenem
D. Trimethoprim-sulfamethoxazole
E. Aztreonam

Answer: D Gram-positive rods in the CSF of a patient with meningitis most likely represent *Listeria monocytogenes*. In those with severe penicillin allergy, trimethoprim-sulfamethoxazole is the recommended agent to treat listerial meningitis. No cephalosporin should be used to treat listeriosis. Vancomycin and imipenem are active in vitro, but treatment failures have been reported. Aztreonam has no anti-listerial activity.

3. An elderly woman is hospitalized with meningitis due to *Listeria monocytogenes*. Which one of the following types of isolation is appropriate for this patient?

A. No isolation required
B. Contact isolation
C. Airborne isolation
D. Droplet isolation

Answer: A Except for transmission from mother to fetus or newborn, human-to-human transmission of listeriosis has not occurred. No isolation is required for those infected with *L. monocytogenes.*

4. Which one of the following is the mechanism by which *Listeria monocytogenes* avoids being killed by phagocytic cells?

A. Inhibition of chemotaxis
B. Inhibition of attachment
C. Inhibition of ingestion
D. Suppression of the metabolic burst
E. Escape from the phagosome

Answer: E The exotoxin, listeriolysin O, is the major virulence factor of *L. monocytogenes* and enables the bacterium to escape from the phagocytic vacuole, whereupon it multiplies in the phagocytic cell cytoplasm.

5. An elderly man is found to have a brain abscess. Which one of the following features is strongly suggestive of *Listeria monocytogenes* as the abscess etiology rather than a more common pathogen?

A. Absence of fever
B. Absence of leukocytosis
C. Location in the pons
D. Focal neurologic findings
E. Presence of seizures

Answer: C Unlike the usual causes of brain abscess, *Listeria* has a particular tropism for the brain stem. Subcortical brain abscesses due to other pathogens are rare; the presence of a subcortical brain abscess should always raise the possibility of listerial infection.

294

ANTHRAX

DANIEL R. LUCEY AND LEV M. GRINBERG

DEFINITION

Anthrax is caused by *Bacillus anthracis*, a spore-forming, gram-positive rod that is aerobic or facultatively anaerobic. Its name derives from a Greek word for "coal," in reference to the black color of the eschar that forms in cutaneous anthrax. Although it is primarily a disease of animals (zoonosis), anthrax was developed as a biowarfare weapon by several nations in the 20th century and used for bioterrorism in the United States in 2001 when spores were mailed in letters.

The Pathogen

The bacterium is a large (1 to 1.5 by 3 to 5 μm), gram-positive rod. It has a ground-glass appearance of growth on sheep blood agar, with 2- to 5-mm, nonhemolytic, tenacious ("beaten egg white") colonies within 24 hours of culture; oval, central to subterminal spores; and a capsule that can be visualized by India ink staining.

EPIDEMIOLOGY

Human infection with *B. anthracis* is often linked to a zoonotic source, such as cattle, sheep, goats, water buffalo, and other animals. Meat, bones, hides, and hair have been reported to transmit infection. Spores can persist in soil for many years. Spores infect animals or humans, then germinate into the vegetative form of *B. anthracis* and cause disease. The World Health Organization provides an online global epidemiology database for anthrax along with guidelines for management of the disease in animals, humans, and the environment.

A systematic review of the worldwide medical literature found that between 1900 and 2005, at least 82 patients with inhalational anthrax were reported in clinical detail.[1] These 82 cases included 18 patients from the United States with naturally acquired, animal-related disease in the 20th century and 11 patients with inhalational anthrax due to the bioterrorism-related events of 2001. The dozens of patients with inhalational anthrax from the 1979 outbreak in Sverdlovsk that was linked to an accidental release of spores downwind from a military facility, however, were not included in these 82 cases because of unavailability of their clinical records. Only the autopsy records from 1979 were preserved by the pathologists (Abramova and Grinberg), and these were published first in Russian in 1993 and then in English for the 41 cases confirmed by histopathology or microbiology.[2,3]

Between 2006 and 2013, an additional five patients with naturally acquired inhalational anthrax were diagnosed, three in the United Kingdom and two in the United States. In 2009, the first human case in the United States of culture-confirmed gastrointestinal anthrax was diagnosed.[4] At least three of these six sporadic cases, including the gastrointestinal case, were linked to exposure to spores contaminating drums made from animal hides. The source of exposure was uncertain for the others. More typically, gastrointestinal anthrax is linked to consumption of meat from infected animals.

Importantly, *B. anthracis* is not transmitted from person to person through the air. Moreover, only one quarter of the 41 autopsy-confirmed inhalational cases in 1979 had any evidence of pneumonia, specifically an acute hemorrhagic pneumonia. All had the characteristic hemorrhagic mediastinal adenopathy, edema, and pleural effusions caused by the primary inhalation of spores. Infection by reaerosolization was suspected in a small number of these patients in 1979.

In support of this reaerosolization hypothesis, a detailed review of the experimental literature found evidence that spores from *Bacillus* species could reaerosolize under outdoor conditions.[5] Moreover, a study of the spores in a contaminated U.S. Senate Office concluded that spores used in a terrorist incident reaerosolized under common office activities.

The most common form of anthrax is the cutaneous form. Direct contact with infected animals or contaminated animal products is the usual mode of transmission. One of the lessons from the 2001 bioterrorist attacks in the United States, however, is that spores in the mailed letters caused cutaneous anthrax in 11 persons, none of whom had inhalational anthrax. Of note, therefore, unexplained cutaneous anthrax cases could be an early clue to an intentional release of spores, especially given that the shortest part of the incubation period in humans for cutaneous anthrax (1 to 12 days) is less than that for laboratory-documented inhalational anthrax (4 to 43 days).

In 2009-2010, at least 80 confirmed or probable cases of anthrax occurred in persons who injected heroin in Europe, primarily in the United Kingdom.[6] In 2012-2013, at least another 14 cases occurred, again in the United Kingdom[7] but also in Denmark, France, and Germany. A new term, *injection anthrax*, was applied to this novel route of infection and distinct clinical syndrome that involved severe soft tissue infection at the site of the injection and sometimes systemic disease. The single strain of *B. anthracis* in these outbreaks most closely matched that reported from a goat in Turkey, suggesting that heroin originating in Afghanistan or Pakistan and smuggled overland in an animal skin may have been the source of these spores in Europe.

PATHOBIOLOGY

Major virulence factors of *B. anthracis* include its two binary toxins, edema toxin and lethal toxin, and also its antiphagocytic poly-D-glutamic acid capsule. Edema toxin consists of edema factor bound to a third anthrax toxin component, protective antigen. Similarly, lethal toxin consists of lethal factor bound to protective antigen. These three toxin components, edema factor, lethal factor, and protective antigen, are encoded on one plasmid (pX-01). The antiphagocytic capsule is encoded on a second plasmid (pX-02). Both plasmids are necessary to cause disease.

The pathogenesis of anthrax has been attributed primarily to its two binary toxins. However, more recently, key roles for both nontoxin and toxin components have been implicated in the pathogenesis of the characteristic cardiac and endovascular abnormalities,[8] bleeding, and shock. Nontoxin components include the peptidoglycan part of the bacterial cell wall and nontoxin metalloproteinases. Moreover, inhibition of both innate and adaptive immune responses occurs by the activities of both toxins and the antiphagocytic capsule.

Similarly, the sepsis model has been proposed to explain the high lethality of inhalational anthrax, rather than the toxin model.[9] In this sepsis model, the primary role of anthrax toxin is to inhibit the immune response against the vegetative form of the bacteria, thus allowing the typically high levels of bacteremia to develop and subsequent shock, multiorgan failure, and death to occur.

Lethal toxin[10] is a metalloproteinase and inhibitor of the mitogen-activated protein kinase intracellular signal transduction pathway. It contributes to the coagulation disorder, hemolysis, and hemorrhage seen with inhalational anthrax. Whether lethal toxin contributed to the vasculitis lesions reported in the 1979 outbreak in Sverdlovsk is uncertain. Edema toxin contributes to the typical "gelatinous" fluid in the mediastinum and abdomen as well as the marked edema seen in both cutaneous anthrax and injection anthrax. The

mechanism is attributed to excessive production of cyclic adenosine monophosphate from adenosine triphosphate, by edema toxin acting as an adenylate cyclase enzyme. The result is water and calcium dysregulation with marked edema.

CLINICAL MANIFESTATIONS

Major clinical manifestations of anthrax infection—inhalational, cutaneous, gastrointestinal, injection, and meningeal—are related to the routes through which *B. anthracis* can enter the body: inhalation, contact, ingestion, or injection (e.g., by contaminated heroin).[11]

Inhalational anthrax[12] almost always causes hemorrhagic mediastinitis, gelatinous edema, and mediastinal adenopathy, resulting in mediastinal widening, as well as pleural effusions due to blockage and reversal of normal lymphatic flow and drainage within the mediastinum. These pleural effusions can be large, bloody, and recurring unless repeatedly drained by thoracentesis or chest tube. Such effusions were found to contribute to respiratory failure in the 1979 outbreak in Sverdlovsk, in part due to compression of lung parenchyma and impaired gas exchange. During the outbreak, the pleural fluid was drained by thoracentesis, rather than by chest tube, and antibiotics were injected into the pleural space. Nevertheless, nearly every patient with inhalational anthrax who underwent autopsy in this outbreak was found to have large (average, 1776 mL) pleural effusions. Although it was not known in 1979 or in 2001, these pleural effusions can serve as a reservoir for toxin. Pleural fluid toxin levels were not tested until the two U.S. patients in 2006 and 2011. High levels of lethal factor were reported.

Radiography demonstrated pulmonary infiltrates in 7 of the 11 patients in 2001 in the United States as well as in both U.S. patients in 2006 and 2011. Thus, although inhalational anthrax has been stated by some not to cause pneumonia, these radiologic findings serve to emphasize the important diagnostic point for clinicians that not only mediastinal widening and pleural effusions but also pulmonary infiltrates are often seen. Moreover, in the 1979 outbreak in Sverdlovsk, one quarter of the 41 autopsy-confirmed cases had a hemorrhagic pneumonia. Thus, pulmonary infiltrates on radiologic imaging, whether or not interpreted as pneumonia, do not rule out the diagnosis of inhalational anthrax.

In 2005, a new three-part clinical staging system for inhalational anthrax (Table 294-1) was published, adding an intermediate progressive stage based on clinical, microbiologic, and radiologic information acquired from the 11 patients in 2001. Previously, these patients would have been included in the late stage, for whom death was considered almost certain. Importantly, all six of the patients who survived the anthrax attacks in 2001 had rapid therapy initiated, including pleural drainage, during this intermediate progressive stage. This three-part symptomatic clinical staging system is cited at www.cidrap.umn.edu/idsa/bt/anthrax/biofacts/anthraxfactsheet.html.

Although in 2001 none of the five patients in the late fulminant stage survived, the two U.S. patients in 2006 and 2011 both survived even though they required mechanical ventilation, placing them in the late fulminant stage. One additional treatment they received was anthrax antitoxin, which was not available in 2001.

The shortest incubation period that has been *documented* microbiologically or histopathologically is 4 days. The range is 4 to 43 days. Although shorter incubation periods may have occurred, none have been documented in the English literature by these laboratory criteria.

Cutaneous anthrax accounts for about 95% of patients with anthrax. The incubation period ranges from 1 to 12 days. Like the three clinical stages of symptomatic inhalational anthrax, three stages of cutaneous anthrax can be delineated: (1) an initial pruritic papule progressing to (2) a central vesicular or bullous lesion with surrounding nonpitting edema and finally (3) a necrotic and hemorrhagic central lesion that evolves into the classic painless eschar with surrounding edema. Resolution can take up to 2 months. This three-stage progression can occur even if appropriate antibiotics are given. During the anthrax letter attacks of 2001, some of the cutaneous lesions were initially considered to be due to brown recluse spider bites. These lesions are usually painful, unlike the painless lesions of anthrax.

In contrast to cutaneous anthrax, injection anthrax[13] described from Europe in persons injecting spore-contaminated heroin does not typically develop a black crusted eschar, can be painful, and can present with severe gastrointestinal or central nervous system manifestations that can rapidly lead to death. The 2011 official report from Health Protection Scotland on outbreaks of this novel form of anthrax advises physicians to suspect injection anthrax in heroin users who demonstrate any of three presentations: (1) severe soft tissue infection including necrotizing fasciitis and cellulitis/abscess, especially if marked edema is present (Fig. 294-1); (2) signs of sepsis even if no soft tissue infection is evident; and (3) meningitis or subarachnoid hemorrhage/intracranial bleed.

Gastrointestinal anthrax is divided into an oropharyngeal form and an intestinal form. The oropharyngeal form has painful cervical adenopathy and an incubation period between 2 hours and 6 days. The oral lesions can ulcerate. They also can progress to cause a white pseudomembrane with dysphagia and hoarseness. The intestinal form has been described to have three clinical phases, much like the three-part progressive clinical stages of inhalational anthrax and cutaneous anthrax: (1) a prodromal phase with fever, malaise, and sometimes syncope; followed by (2) a progressive phase with abdominal

TABLE 294-1 CLINICAL STAGING SYSTEM FOR INHALATIONAL ANTHRAX

I. EARLY PRODROMAL STAGE

Nonspecific illness sometimes described as "flulike" and including any of the following: fever, cough, headache, chills, nausea, chest pain, or abdominal pain. Laboratory tests and radiographs are nondiagnostic. The prognosis for cure is good with appropriate therapy, but the diagnosis is difficult to confirm acutely in this stage.

II. INTERMEDIATE PROGRESSIVE STAGE

Any of the following findings are defining inclusion criteria for this stage:
1. Positive blood cultures (typically positive in <24 hours)
2. Mediastinal adenopathy
3. Pleural effusions: bloody, often large, require drainage, and may recur

Findings in this stage may include high fever, dyspnea, confusion or syncope, or increasing nausea and vomiting. Exclusion criteria for this stage include the following:
1. Meningitis
2. Respiratory failure requiring intubation and mechanical ventilation or
3. Shock

Importantly, patients in the intermediate progressive stage can still be cured with appropriate antibiotics and drainage of pleural effusions by repeated thoracentesis or preferably chest tube to keep the pleural space dry in order to reduce the adverse mechanical effect on respiration by large-volume effusions and to remove potentially toxin-producing *Bacillus anthracis* from the pleural space.

III. LATE FULMINANT STAGE

Inclusion criteria include any one of the following findings:
1. Meningitis
2. Respiratory failure requiring intubation and mechanical ventilation
3. Shock: end-organ hypoperfusion

Findings in this stage may also include any of those from previous stages, so there are no exclusion criteria. The probability of survival is lowest in this stage. Novel therapeutics that safely and effectively neutralize anthrax toxin may be needed to increase survival.

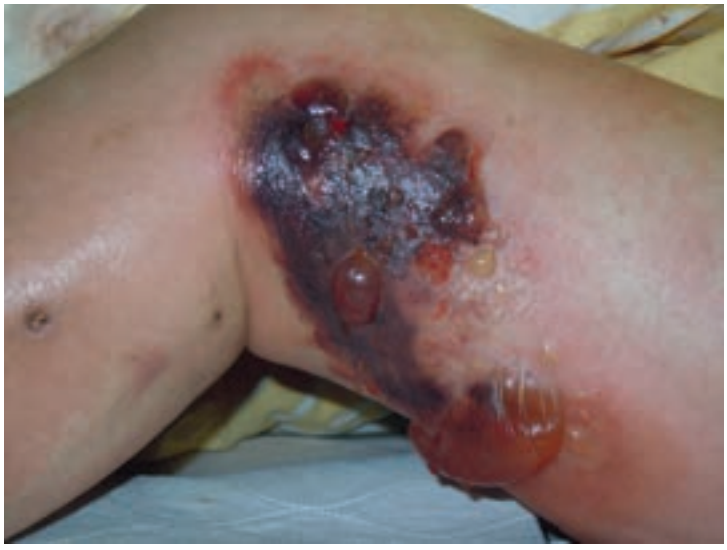

FIGURE 294-1 An example of the lesions due to injection anthrax.

pain, nausea, vomiting, abdominal distention, ascites, and severe weakness; and finally (3) a fulminant phase with rapidly increasing abdominal girth and expanding ascites, paroxysmal abdominal pain, and shock. Of note, the U.S. patient with gastrointestinal anthrax in 2010 who survived had more than 50 liters of ascitic fluid drained as part of her therapy along with antibiotics and anthrax antitoxin.

Anthrax meningoencephalitis can occur in association with any of the other forms of anthrax, inhalation, cutaneous, gastrointestinal, or injection, and rarely without a known portal of entry. Cerebral edema, parenchymal brain hemorrhage, vasculitis, and subarachnoid hemorrhage can occur. The cerebrospinal fluid is often bloody with anthrax meningitis. At autopsy, the extensive bleeding gives a characteristic macroscopic appearance termed the cardinal's cap. This form of anthrax remains fatal in more than 95% of cases.

DIAGNOSIS

If anthrax is suspected, blood should be obtained immediately before any antibiotics are given, primarily for microbiologic culture and possibly for newer assays for anthrax toxins whenever available. Notably, blood cultures from patients with inhalational anthrax will grow the large, gram-positive rods within 24 hours. Given such high levels of bacteremia, it is surprising that even one dose of an effective antibiotic has been reported to turn blood cultures negative, hence the need to draw blood culture samples before antibiotics are given. Toxin assays could still be positive after initial antibiotics, but these assays are still investigational at this time. Nasal cultures should not be performed as a routine clinical diagnostic test for individual patients because a negative result does not rule out inhalation of spores. Nasal cultures could be epidemiologically useful, however, as part of an investigation to help estimate the perimeter of exposure to spores.

In the microbiology laboratory, if the initial culture is suggestive of *B. anthracis*, three tests can be performed with a biosafety cabinet. These are assays for motility (nonmotile), catalase (positive), and hemolysis (negative). If these characteristics are found, identification is still not proved until a sample is sent to a reference laboratory where polymerase chain reaction analysis can identify *B. anthracis* and gamma-phage lysis of the encapsulated bacteria can provide confirmation. One or more antibody tests also are available, although results are unlikely to be positive during the earliest part of the disease. In contrast, a number of investigational anthrax toxin tests (e.g., for lethal factor, edema factor, or protective antigen) are being developed that should have positive results early in the disease, even if antibiotics have been given, and may also have prognostic value. At the time of this writing, these assays are available in the United States only through the Centers for Disease Control and Prevention (CDC).

A noncontrast computed tomography scan of the chest can be a valuable adjunctive diagnostic tool for inhalational anthrax. It is more sensitive than a chest radiograph for demonstrating the characteristic hyperattenuating mediastinal adenopathy causing the mediastinal widening and pleural effusions. The hyperattenuation is consistent with bleeding into lymph nodes and thus helps differentiate inhalational anthrax from tularemia, histoplasmosis, tuberculosis, sarcoidosis, and most other causes of mediastinal or hilar adenopathy.

Anthrax meningitis is typically neutrophilic and bloody. Very large gram-positive rods in the cerebrospinal fluid distinguish anthrax meningitis from other causes of gram-positive meningitis, such as the smaller *Listeria monocytogenes*. In the February 2014 *Emerging Infectious Diseases* online journal, the CDC published its first formal updated guidelines on anthrax since 2001, including diagnostic tests, prophylaxis, treatment, and monitoring.[14] New diagnostic recommendations include a lumbar puncture at admission unless it is contraindicated because new treatment recommendations will be based in part on whether meningitis has been excluded.

TREATMENT

Rx

Treatment of symptomatic anthrax includes antibiotics and, in the case of inhalational anthrax, pleural fluid drainage. In addition, in December 2012, the Food and Drug Administration (FDA) licensed the first anthrax antitoxin for treatment of inhalational anthrax. The CDC recommendations for prevention and treatment were updated in 2014, and the reader is referred to their webpage for complete details.[14]

Treatment with three intravenous antibiotics is recommended in the 2014 CDC recommendations if meningitis is possible or confirmed. These include a bactericidal fluoroquinolone (ciprofloxacin is the preferred drug), a bactericidal β-lactam (meropenem is the preferred drug), and a protein synthesis inhibitor (linezolid is the preferred drug). If meningitis is excluded, a bactericidal drug (ciprofloxacin is preferred) and a protein synthesis inhibitor (either clindamycin or linezolid) is preferred.

A total 60-day course of antibiotic therapy, of at least 2 weeks intravenously with more than one drug followed by the remainder with one oral drug, is recommended for inhalational anthrax because of concern that spores may germinate into vegetative bacteria if therapy is stopped sooner.

An essential treatment modality for inhalational anthrax is pleural drainage. The CDC 2014 guidelines state: "Drainage of pleural fluid and ascites is believed to improve survival by reducing the toxin level and by decreasing mechanical lung compression. These data support the need for early and aggressive drainage of any clinically or radiographically apparent pleural effusions; chest tube drainage is recommended over thoracentesis because many effusions will require prolonged drainage. Thoracotomy or video-assisted thoracic surgery might be required to remove gelatinous or loculated effusions."

As of 2014, there are two types of anthrax toxin antibodies in the Strategic National Stockpile. One is polyclonal, called intravenous anthrax immune globulin, and is derived from the plasma of persons who have received the anthrax vaccine. It is still investigational, and the investigational new drug protocol is held by the CDC in Atlanta. Intravenous anthrax immune globulin has been given as adjunctive therapy to several patients, including three in the United States in 2006, 2009, and 2011 as well as several patients in the United Kingdom. The other antitoxin is a humanized monoclonal antibody called raxibacumab that was licensed by the FDA in December 2012 for treatment of inhalational anthrax in both adults and children.[15] It has not been given to any patients with anthrax as of February 2015 but was licensed by the FDA on the basis of efficacy in animal models of inhalational anthrax plus human safety data in healthy volunteers. It does not cross the blood-brain[16] barrier. It is given intravenously during 2 hours and 15 minutes, after a single dose of the antihistamine diphenhydramine.

PREVENTION

Clinical anthrax can be prevented by preexposure vaccination or by postexposure prophylaxis with antibiotics. Moreover, in December 2012, the monoclonal antibody antitoxin (raxibacumab) was licensed by the FDA for postexposure prophylaxis in both adults and children "when alternative therapies are not available or not appropriate."

The current FDA-licensed anthrax vaccine (as of 1970) contains protective antigen as the vaccine antigen and alum as an adjuvant.[17] It requires five injections during an 18-month period when it is given before exposure. In contrast, postexposure prophylaxis, which is an off-label use of this same vaccine and thus needs an investigational new drug or emergency use authorization protocol, requires only three injections given during a 1-month period. This vaccine is in limited supply and is dedicated primarily for use by the military; however, it is available to civilian populations through at least one commercial travel clinic.

Antibiotics that are approved by the FDA for postexposure prophylaxis include ciprofloxacin, doxycycline, and procaine penicillin G as well as levofloxacin. Either doxycycline or ciprofloxacin is recommended as the preferred drug by the CDC for initial prophylaxis when the antibiotic susceptibility of an anthrax strain is unknown. During pregnancy, however, ciprofloxacin is preferred to doxycycline according to 2014 CDC guidelines. The CDC website on anthrax (at http://www.bt.cdc.gov/agent/anthrax) has detailed recommendations regarding the choice of antibiotics, doses, and durations for both prevention and treatment of anthrax.

Prevention of anthrax due to bioterrorism remains a national priority. A suspicion of clinical anthrax or exposure to spores warrants immediate involvement of law enforcement authorities because of the possibility of a criminal act. Evolving approaches for detecting and responding to any future bioterrorism attacks with anthrax include the BioWatch system to detect aerosolized threats and the Autonomous Detection System in postal facilities. The Cities Readiness Initiative was initially described on the CDC website in June 2004 as a program for multiple U.S. cities to help prepare for large-scale public health emergencies, including bioterrorism attacks (e.g., with aerosolized anthrax or another organism). Large volumes of medical supplies, including but not limited to antibiotics, can be delivered rapidly from the Strategic National Stockpile to one or more cities followed by local distribution. The 2012 Institute of Medicine report titled *Prepositioning Antibiotics for Anthrax* discusses the Strategic National Stockpile, Cities Readiness Initiative, BioWatch, and antibiotic issues including defining multidrug-resistant and extremely drug-resistant strains.

PROGNOSIS

The inhalational anthrax of all six patients who survived the attacks of 2001 was diagnosed during the intermediate progressive stage, and prompt therapy prevented progression beyond this stage. The 45% case-fatality rate in 2001 was much improved over the 88% rate seen in the United States from 1900 to 1976. Survival is more likely in patients who undergo pleural drainage, receive multidrug antibiotic regimens, do not require intubation or tracheotomy, and do not progress to anthrax meningoencephalitis. Of note, two patients with inhalational anthrax in the United States (in 2006 and 2011) with respiratory failure requiring mechanical ventilation have survived. Whether the addition of antitoxin to their intensive care, multiple antibiotics, and pleural drainage played a causal role in their survival is uncertain. Their survival emphasizes, however, that even patients in the late fulminant stage can sometimes survive.

The mortality from cutaneous anthrax is approximately 20% if it is untreated, particularly in patients in whom upper airway compression develops from a lesion on the neck or in whom secondary bacteremic anthrax meningitis develops. Anthrax meningitis remains fatal in more than 95% of victims, and thus better therapies are needed. The FDA licensure of the first anthrax antitoxin for both children and adults, as therapy and also as postexposure prophylaxis when alternative therapies are not available or not appropriate, may improve clinical outcome in systemic forms of anthrax. An antitoxin is still needed, however, that can treat anthrax meningitis.

FUTURE DIRECTIONS

A rapid test is still needed for the diagnosis of infection with *B. anthracis* in the early prodromal stage of the illness or in later clinical stages if the patient received antibiotics before blood culture samples were obtained. Such a test is likely to be based on a toxin or a toxin component. Ideally, it should be available at the point of care.

Anthrax toxin inhibitors, in conjunction with antibiotics, could be particularly useful in treating patients who have progressed into the systemic stages of anthrax as a result of inhalational, gastrointestinal, or injection anthrax causing serious soft tissue infection. In addition, antitoxin could be useful in the setting of infection with an engineered form of multidrug-resistant or extremely drug-resistant anthrax. New anthrax vaccines are still being tested. A shorter series of primary and booster doses for the only licensed vaccine was reported in December 2013. Criteria have been established for approval by the FDA of this vaccine specifically in the setting of postexposure prophylaxis. Meeting these criteria and obtaining licensure of the vaccine for postexposure prophylaxis will be a major advance.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Holty JE, Bravata DM, Liu H, et al. Systematic review: a century of inhalational anthrax cases from 1900-2005. *Ann Intern Med*. 2006;144:270-280.
2. Abramova AA, Grinberg LM. Pathology of anthrax sepsis according to materials of the infectious outbreak in 1979 in Sverdlovsk (macroscopic changes) [in Russian; English translation by Dr. Grinberg in 2012 and posted at www.pastasprologue.org]. *Arkh Patol*. 1993;55:12-17.
3. Meselson M, Guillemin J, Hugh-Jones M, et al. The Sverdlovsk anthrax outbreak of 1979. *Science*. 1994;226:1202-1208.
4. Klempner MS, Talbot EA, Lee SI, et al. A 24-year-old woman with abdominal pain and shock (case 25-2010/Aug 19.). *N Engl J Med*. 2010;363:766-777.
5. Layshock JA, Pearson B, Crockett K, et al. Reaerosolization of *Bacillus* spp in outdoor environments: a review of the experimental literature. *Biosecur Bioterror*. 2012;10:299-303.
6. Booth M, Donaldson L, Cui X, et al. Confirmed Bacillus anthracis infection among persons who inject drugs, Scotland, 2009-2010. *Emerg Infect Dis*. 2014;20:1452-1463.
7. Abbara A, Brooks T, Taylor GP, et al. Lessons for control of heroin-associated anthrax in Europe from 2009-2010 outbreak case studies, London, UK. *Emerg Infect Dis*. 2014;20:1115-1122.
8. Remy KE, Qui P, Li Y, et al. *B. anthracis* associated cardiovascular dysfunction and shock: the potential contributions of both non-toxin and toxin components. *BMC Med*. 2013;11:217.
9. Coggeshall KM, Lupu F, Ballard J, et al. The sepsis model: an emerging hypothesis for the lethality of inhalational anthrax. *J Cell Mol Med*. 2013;17:914-920.
10. Liu S, Moayeri M, Leppla SH, et al. Anthrax lethal and edema toxins in anthrax pathogenesis. *Trends Microbiol*. 2014;22:317-325.
11. Infectious Diseases Society of America (IDSA) website for comprehensive information on anthrax. www.cidrap.umn.edu/idsa/bt/anthrax/biofacts/anthraxfactsheet.html. Accessed February 9, 2015.
12. Sweeney DA, Hicks CW, Cui X, et al. Anthrax infection. *Am J Respir Crit Care Med*. 2011;184:1333-1341.
13. Hicks CW, Sweeney DA, Cui X, et al. An overview of anthrax infection including the recently identified form of disease in injection drug users. *Intensive Care Med*. 2012;38:1092-1104.
14. Hendricks KA, Wright ME, Shadomy SV, et al. CDC expert panel meetings on prevention and treatment of anthrax in adults. *Emerg Infect Dis*. 2014;20. http://dx.doi.org/10.3201/eid2002.130687. Accessed February 9, 2015.
15. Kummerfeldt CE. Raxibacumab: potential role in the treatment of inhalational anthrax. *Infect Drug Resist*. 2014;7:101-109.
16. Migone TS, Bolmer S, Zhong J, et al. Added Benefit of Raxibacumab to Antibiotic Treatment of Inhalational Anthrax. *Antimicrob Agents Chemother*. 2015;59:1145-1151.
17. Wright JG, Pikaytis BD, Rose CE, et al. Effect of dose schedules and intramuscular injection of anthrax vaccine adsorbed on immunological response and safety profile: a randomized trial. *Vaccine*. 2014;32:1019-1028.

REVIEW QUESTIONS

1. A 56-year-old man is diagnosed with inhalational anthrax. His blood cultures are positive for *Bacillus anthracis*. His noncontrast chest computed tomography scan shows mediastinal adenopathy and moderate-sized bilateral pleural effusions. He is receiving several appropriate antibiotics. According to the 2014 anthrax management guidelines from the Centers for Disease Control and Prevention (CDC), what is recommended to be done with regard to the pleural effusions?

A. Follow closely on a daily basis to be sure they decrease in size with antibiotics.
B. Drain by thoracentesis only if they enlarge after 48 hours with antibiotics.
C. Drain immediately by thoracentesis but without chest tube drainage.
D. Perform thoracentesis on one side and send fluid for anthrax toxin levels.
E. Perform bilateral chest tube drainage at this time.

Answer: E Retrospective analysis of the 1979 inhalational anthrax outbreak in Sverdlovsk, USSR, the 2001 outbreak in the United States, and the recent findings of high levels of anthrax toxin in the pleural fluid support the CDC's 2014 guidelines on aggressive management of pleural effusions. The CDC now advises: "Drainage of pleural fluid and ascites is believed to improve survival by reducing the toxin level and by decreasing mechanical lung compression. These data support the need for early and aggressive drainage of any clinically or radiographically apparent pleural effusions; chest tube drainage is recommended over thoracentesis because many effusions will require prolonged drainage."

2. In December 2012, the Food and Drug Administration (FDA) licensed the first anthrax antitoxin. This product is a humanized monoclonal antibody. Which of the following choices is *not* included in the FDA licensure of this antitoxin?

A. Treatment of children during any stage of inhalational anthrax
B. Treatment of adults during any stage of inhalational anthrax
C. Treatment of children with only early-stage inhalational anthrax
D. Postexposure prophylaxis when alternative therapies are unavailable
E. Postexposure prophylaxis when alternative therapies are not appropriate.

Answer: C This monoclonal anthrax antitoxin was licensed by the FDA in December 2012 for treatment of inhalational anthrax in both children and adults at any stage of the disease (early prodromal, intermediate progressive, or late fulminant). In addition, this antitoxin was licensed for postexposure prophylaxis in both children and adults "when alternative therapies are not available or not appropriate." (See discussion in the Prevention and Treatment sections.)

3. The CDC recommendations for use of the anthrax vaccine specifically for postexposure prophylaxis, in addition to an antibiotic, is for which of the following number of vaccine doses and schedules?

A. One dose of vaccine as soon as possible after exposure to spores
B. Two doses of vaccine during a 1-month period
C. Three doses of vaccine during a 1-month period
D. Three doses of vaccine during a 12-month period
E. Five doses of vaccines during an 18-month period

Answer: C Anthrax vaccine is recommended by the CDC in their 2014 guidelines specifically for postexposure prophylaxis as three doses of vaccine given during a 1-month period. In contrast, the FDA-licensed anthrax vaccine for the traditional preexposure use requires five injections given during an 18-month period. (See discussion in the Prevention section.)

4. Antibiotics that are approved by the FDA for postexposure prophylaxis include all of the following except

A. Levofloxacin
B. Ciprofloxacin
C. Procaine penicillin G
D. Clindamycin
E. Doxycycline

Answer: D Ciprofloxacin, doxycycline, levofloxacin, and procaine penicillin G are all approved by the FDA for postexposure prophylaxis use to prevent symptomatic anthrax. Clindamycin is not approved for postexposure prophylaxis; however, it is included in the 2014 CDC recommendations for treatment of some patients with symptomatic anthrax infection. (See discussion in the Prevention section.)

5. Blood culture samples drawn before antibiotics are administered typically will start growing *Bacillus anthracis* in patients with inhalational anthrax within what time frame?

A. 6 hours
B. 24 hours
C. 48 hours
D. 1 week
E. Will not grow in standard culture media

Answer: B A striking and consistent finding in all patients since the 2001 U.S. outbreak of inhalational anthrax is the growth within 24 hours of *Bacillus anthracis* from blood cultures of patients who were not already receiving antibiotics when the blood culture samples were drawn. This crucial observation emphasizes the diagnostic importance of drawing blood culture samples in any patient suspected of having inhalational anthrax before antibiotics are started. (See discussion in the Diagnosis section.)

295

ERYSIPELOTHRIX INFECTIONS

ANNETTE C. REBOLI

DEFINITION

Erysipelothrix rhusiopathiae causes three well-defined patterns of human infection: (1) erysipeloid, a cellulitis of the fingers and hands (also known as whale finger or pork finger), which is the most common manifestation of infection with *E. rhusiopathiae*; (2) a diffuse cutaneous form; and (3) a bacteremic form, with or without cutaneous involvement, usually complicated by endocarditis.

The Pathogen

E. rhusiopathiae is a thin, pleomorphic, nonsporulating, microaerophilic gram-positive rod. It may be confused with other gram-positive bacillary organisms, particularly *Listeria monocytogenes* (Chapter 293) and *Corynebacterium* species (Chapter 292). It can be differentiated from *L. monocytogenes* by its lack of motility, lack of catalase and coagulase production, and resistance to neomycin. Most strains of *E. rhusiopathiae* produce hydrogen sulfide on triple sugar iron agar slants, a feature that distinguishes *E. rhusiopathiae* from *L. monocytogenes* and from corynebacteria. Because α-hemolysis may be seen after 48 hours of incubation of *E. rhusiopathiae*, confusion with streptococci may also occur. The term *erysipeloid* refers to cutaneous infection by *E. rhusiopathiae* and should not be confused with erysipelas (see Fig. 441-4), which is a superficial cellulitis caused by streptococci or staphylococci.

EPIDEMIOLOGY

E. rhusiopathiae is found worldwide as a commensal or as a pathogen in a variety of wild and domestic animals, including swine, sheep, cattle, horses, dogs, cats, rodents, chickens, ducks, turkeys, penguins, and parrots, as well as in flies, ticks, mites, and lice. The greatest commercial impact of *E. rhusiopathiae* infection is due to disease in swine, but infection of sheep and poultry is also important economically. Environmental surfaces in contact with infected animals or their products are potential sources of *E. rhusiopathiae.* It can persist for prolonged periods in contaminated soil. *E. rhusiopathiae* is killed within 15 minutes by heating to 55° C and by several commercially available home disinfectants.

The incidence of cutaneous infection in humans seems to be decreasing because of technologic advances in animal industries. Infection is usually the result of contact with infected animals or their products. Persons at greatest risk for infection include fishers, fishmongers, farmers, butchers, slaughterhouse workers, and veterinarians.[1] The organism gains entry through cuts and abrasions on the skin. The seasonal incidence of erysipeloid parallels that of swine erysipeloid and is highest in the summer and early fall. The rare instances of systemic infection that do not have an occupational link tend to occur in immunocompromised hosts, suggesting that oropharyngeal or gastrointestinal colonization with the organism may occur. Chronic alcoholism has been acknowledged as a common underlying condition. Erysipeloid and erysipeloid with bacteremia have been reported rarely after cat and dog bites, suggesting that *E. rhusiopathiae* may be part of the oral flora of these animals.

PATHOBIOLOGY

The virulence of *E. rhusiopathiae* is associated, at least in part, with resistance to phagocytosis by polymorphonuclear leukocytes. This antiphagocytic ability results from the organism's possession of a capsule. In the absence of specific antibodies, *E. rhusiopathiae* evades phagocytosis, but even if it is phagocytosed, it is able to replicate intracellularly in these cells. Other virulence factors include enzymes (neuraminidase and hyaluronidase) and surface proteins.

CLINICAL MANIFESTATIONS

Because of the mode of acquisition (contact with infected animals or their products, with organisms inoculating abrasions on the skin), lesions are usually confined to the fingers and hands (Fig. 295-1). A well-defined, slightly elevated, violaceous lesion accompanied by a very painful, throbbing, burning, or itching sensation develops within 2 to 7 days of traumatic dermal inoculation. The infected area is swollen. Vesicles may be present, but suppuration is absent. The lesion spreads slowly to other fingers but rarely involves the fingertips or the skin above the wrist. As the lesion spreads peripherally, the central area clears.[2] Systemic signs and symptoms are rare. There may be sterile arthritis of an adjacent joint. Regional lymphadenopathy or lymphadenitis occurs in about 20% of cases, and low-grade fever develops in approximately 10%. Lesions usually resolve within 3 weeks without treatment. Relapse occurs in 1% of cases.

The diffuse cutaneous form is rare. The cutaneous lesion progresses proximally from the site of inoculation or appears at remote areas. Patients often have fever and arthralgias, but blood cultures are generally negative.

Systemic infection with *Erysipelothrix* is uncommon. More than 90 cases of bacteremia have been reported; most of the patients had endocarditis. Although cases of prosthetic valve endocarditis have been reported, most cases have involved native valves. In 60% of cases, infection developed on apparently normal heart valves. One third of patients had an antecedent or concurrent skin lesion of erysipeloid. The clinical manifestations of endocarditis secondary to *E. rhusiopathiae* and other microorganisms are similar. *E. rhusiopathiae* endocarditis correlates highly with occupation, exhibits a tropism for the aortic valve, affects more males than females, and is associated with high mortality.[3] The high mortality may reflect a delay in appropriate therapy because of the empirical use of vancomycin, which is not an effective treatment of *E. rhusiopathiae.* Cases of *E. rhusiopathiae* endocarditis have been

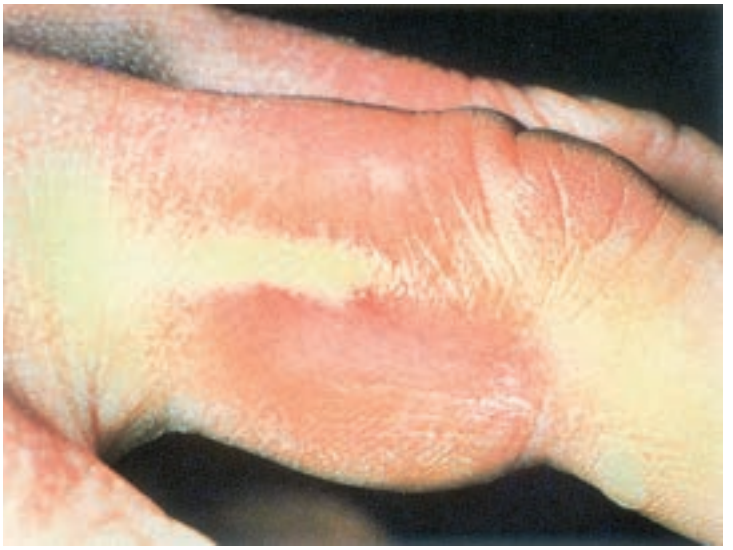

FIGURE 295-1. Erysipeloid with its characteristic purple, nonpurulent swelling of the finger. Also known as whale finger or pork finger, this form of cellulitis caused by *Erysipelothrix rhusiopathiae* should not be confused with streptococcal or staphylococcal erysipelas (see Fig. 441-4). (From Farrar WE, Wood MJ, Innes JUA, Tubbs H. *Infectious Diseases: Text and Color Atlas*. 2nd ed. New York: Gower Medical Publishing; 1992.)

complicated by paravalvular and myocardial abscess formation, cerebral emboli, congestive heart failure, valve perforation, and acute renal failure. *E. rhusiopathiae* bacteremia without endocarditis occurs more frequently than was previously believed. Bacteremia is occurring with increased frequency in immunocompromised patients, whereas endocarditis usually occurs in immunocompetent patients. Focal infections, including brain abscess, meningitis, endophthalmitis, osteomyelitis, septic arthritis, epidural and paravertebral abscesses, liver abscess, necrotizing fasciitis, intra-abdominal abscess, and peritonitis, have been reported. Some of these infections were complications of bacteremia. Septic arthritis has occurred in native joints, in prosthetic joints, and after arthroscopic surgery.[4] Peritonitis has complicated peritoneal dialysis.

DIAGNOSIS

E. rhusiopathiae grows on routine laboratory media. Because *E. rhusiopathiae* is located only in deeper parts of the skin in cases of erysipeloid, biopsy of the entire thickness of the dermis from the edge of the lesion yields maximal recovery of the organism. Definitive diagnosis by skin biopsy is rarely necessary because of the classic clinical presentation and the rapid response to therapy. Routine blood culture techniques are adequate for growth and isolation of the organism in suspected cases of bacteremia or endocarditis. Various selective media have been used to improve the isolation of *E. rhusiopathiae* from contaminated specimens. Molecular techniques, such as polymerase chain reaction with primers specific for *E. rhusiopathiae*, have been developed and improve the efficiency of detection and identification.[5]

TREATMENT

Rx

Most isolates of *E. rhusiopathiae* are susceptible to penicillin, cephalosporins, imipenem, clindamycin, ciprofloxacin, ofloxacin, and daptomycin. Some resistance has been observed with erythromycin, tetracycline, and chloramphenicol. *E. rhusiopathiae* is resistant to vancomycin, aminoglycosides, trimethoprim-sulfamethoxazole, and sulfonamides. Penicillin G is the treatment of choice. Uncomplicated cutaneous lesions generally respond well to a 5- to 7-day course of oral penicillin. Treatment hastens healing, although relapse may still occur. Bacteremia should be treated with intravenous penicillin; cases of endocarditis should be treated with 12 million to 20 million units of penicillin G daily or ceftriaxone 1 g daily for 4 to 6 weeks. Two weeks of intravenous therapy followed by 2 weeks of oral therapy has been successful. The use of quinolones or daptomycin may be considered for *Erysipelothrix* infections when the patient is allergic to β-lactams. Oral linezolid was used to complete therapy in a case of bacteremia complicated by endophthalmitis. Valve replacement may be necessary in patients with endocarditis. Infected prosthetic devices should be removed.

PREVENTION

Proper cleaning and disinfection of work surfaces and attention to hygienic work practices, including the use of gloves and hand hygiene, reduce the risk for infection. Vaccines are available for commercial use in animals only.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Wang Q, Chang BJ, Riley TV. *Erysipelothrix rhusiopathiae. Vet Microbiol.* 2010;140:405-417.
2. Veraldi S, Girgenti V, Dassoni F, et al. Erysipeloid: a review. *Clin Exp Dermatol.* 2009;34:859-862.
3. Kaya S, Gençalioğlu E, Yildirim SS, et al. Native valve endocarditis caused by Erysipelothrix rhusiopathiae in an immunocompetent individual. *J Med Microbiol.* 2013;62:1911-1913.
4. Hocqueloux L, Poisson DM, Sunder S, et al. Septic arthritis caused by *Erysipelothrix rhusiopathiae* in a prosthetic knee joint. *J Clinc Microbiol.* 2010;48:333-335.
5. Pal N, Bender JS, Opriessnig T. Rapid detection and differentiation of *Erysipelothrix* spp. by a novel multiplex real-time PCR assay. *J Appl Microbiol.* 2010;108:1083-1093.

REVIEW QUESTIONS

1. A 56-year-old fisherman, who has no significant medical problems, presents with a 2-week history of a well-defined, elevated, painful, reddish purple plaque on the dorsum of his right hand. The lesion has been slowly expanding during the course of 2 weeks and appeared approximately 1 week after his last fishing trip. He is afebrile, and there is no evidence of regional lymphadenopathy or lymphadenitis. Which of the following is the most appropriate recommendation?

A. Aspirate the leading edge and send any material for Gram stain and culture.
B. Begin oral penicillin.
C. Refer for a biopsy of the lesion.
D. Begin intravenous vancomycin.
E. No treatment

Answer: B This patient has classic erysipeloid, caused by *Erysipelothrix rhusiopathiae*. It is a zoonosis that is linked to certain occupations, including fishermen. The lesion is a cellulitis that is generally found on the hands or fingers, and the organism is introduced through cuts and abrasions of the skin. Systemic signs of infection are rare. Although uncomplicated skin lesions may resolve spontaneously during a period of 3 weeks, a short course of oral penicillin G hastens healing and is the treatment of choice. The organism is resistant to vancomycin. Because the organism is located in deeper parts of the skin and there is not any suppuration, in cases of erysipeloid, aspiration is unlikely to lead to recovery of the organism. A biopsy of the entire thickness of the dermis at the edge of the lesion would yield maximum recovery of the organism but is not indicated because of the classic presentation and the ease of treatment with penicillin. (Veraldi S, Girgenti V, Dassoni F, et al. Erysipeloid: a review. *Clin Exp Dermatol*. 2009;34:859-862.)

2. A 47-year-old man with a history of chronic liver disease, secondary to alcohol use, presents with a 3-day history of fever, chills, and generalized arthralgias. He works in a slaughterhouse that processes pigs. Approximately 3 weeks ago, he developed a cellulitis of his finger that resolved spontaneously without therapy. Blood culture samples are obtained and yield gram-positive bacilli that are identified as *Erysipelothrix rhusiopathiae*. An echocardiogram is normal, and follow-up blood cultures are negative. The antibiotic therapy of choice is

A. Vancomycin
B. Intravenous penicillin G
C. Tetracycline
D. Ampicillin and gentamicin
E. Trimethoprim-sulfamethoxazole

Answer: B This patient has bacteremia without infective endocarditis. *Erysipelothrix rhusiopathiae* bacteremia without endocarditis occurs more frequently than was previously believed and is occurring with increased frequency in immunocompromised patients, including those with liver disease. The treatment of choice for systemic infections is intravenous penicillin G. *E. rhusiopathiae* is resistant to vancomycin, aminoglycosides, tetracycline, and trimethoprim-sulfamethoxazole. (Wang Q, Chang BJ, Riley TV. *Erysipelothrix rhusiopathiae*. *Vet Microbiol*. 2010;140:405-417.)

3. A 45-year-old veterinarian presents with a 2-week history of fever, chills, and sweats. On physical examination, he is noted to have a 3/6 diastolic murmur, consistent with aortic regurgitation. Three sets of blood cultures are positive for *Erysipelothrix rhusiopathiae*, and an echocardiogram reveals a vegetation on the aortic valve. He has an allergy to penicillin, which is manifested as hives. The most appropriate alternative antibiotic therapy is

A. Vancomycin
B. Ceftriaxone
C. Daptomycin
D. Ampicillin and gentamicin
E. Tetracycline

Answer: C This patient has infective endocarditis of the aortic valve caused by *Erysipelothrix rhusiopathiae*. *E. rhusiopathiae* endocarditis correlates highly with occupation, exhibits a tropism for the aortic valve, and affects more males than females. Penicillin is the treatment of choice; daptomycin and quinolones are alternative therapies when the patient is allergic to β-lactams. For a patient with an immediate hypersensitivity reaction to penicillin, cephalosporins are contraindicated because of possible cross-reactivity. *E. rhusiopathiae* is resistant to vancomycin, tetracyclines, aminoglycosides, and trimethoprim-sulfamethoxazole. Even if there was not a concern for resistance, tetracyclines are bacteriostatic, and it is necessary to administer bactericidal agents for the treatment of endocarditis. (Piper KE, Steckelberg JM, Patel R. In vitro activity of daptomycin against clinical isolates of gram-positive bacteria. *J Infect Chemother*. 2005;11:207-209.)

4. A 60-year-old man with a history of rheumatoid arthritis, who is being treated with prednisone chronically, presents to his physician to discuss prevention of infection with *Erysipelothrix rhusiopathiae*. He has a friend who works in the slaughterhouse with him who recently was treated for endocarditis caused by *E. rhusiopathiae*. The physician should recommend the following:

A. Vaccination
B. Chronic low-dose therapy with oral penicillin while he is receiving prednisone
C. Good hand hygiene and use of protective gloves
D. Use of a gown, mask, and gloves while at work

Answer: C *Erysipelothrix rhusiopathiae* is an occupational pathogen that is transmitted through traumatic contact with infected animals or animal products. Occupations at risk include fishermen, farmers, veterinarians, and slaughterhouse workers. Immunocompromised hosts are at risk for infection. The transmission of *Erysipelothrix* generally occurs by traumatic inoculation through abrasions or cuts in the skin of the hands and rarely through consumption of contaminated meat or fish products. Proper cleaning and disinfection of work surfaces and attention to hygienic work practices, including good hand hygiene and the use of gloves, reduce the risk of infection. Vaccines are available for commercial use in animals only, and there are currently no known vaccines that are available for humans. There is no documentation of transmission by droplets or aerosols, and therefore masks are not indicated. (Wang Q, Chang BJ, Riley TV. *Erysipelothrix rhusiopathiae*. *Vet Microbiol*. 2010;140:405-417.)

5. A 63-year-old woman with a history of chronic eczema and alcoholism underwent a right knee arthroplasty for degenerative joint disease. She lived on a farm that raised swine. Six months after the arthroplasty, she developed pain and swelling of the right knee, with subsequent drainage of purulent material from a pustule in the incision. Plain radiographs revealed loosening of the prosthesis and osteomyelitis. She underwent removal of the prosthesis, and multiple operative samples of fluid and bone revealed gram-positive rods. Cultures were positive for *Erysipelothrix rhusiopathiae*. Blood cultures were negative. She was treated with intravenous imipenem for 2 weeks. Which of the following is the most appropriate recommendation for outpatient oral therapy?

A. Clarithromycin
B. Tetracycline
C. Trimethoprim-sulfamethoxazole
D. Ciprofloxacin

Answer: D Most isolates of *Erysipelothrix rhusiopathiae* are susceptible to penicillin, cephalosporins, imipenem, clindamycin, ciprofloxacin, ofloxacin, and daptomycin. Resistance has been observed with erythromycins, tetracyclines, and chloramphenicol. *E. rhusiopathiae* is also resistant to vancomycin, aminoglycosides, trimethoprim-sulfamethoxazole, and sulfonamides. An oral fluocinolone, such as ciprofloxacin, has excellent efficacy in the treatment of osteomyelitis and septic arthritis. Although arthritis is one of the most common features of *E. rhusiopathiae* infection in animals, this type of infection is relatively rare in humans. There are about 12 cases of bone and joint infection reported in the medical literature. Most patients had a skin lesion at the site of the infected joint and have an occupational or recreational exposure to an animal reservoir of infection. (Hocqueloux L, Poisson DM, Sunder S, et al. Septic arthritis caused by *Erysipelothrix rhusiopathiae* in a prosthetic knee joint. *J Clin Microbiol*. 2010;48:333-335.)

296

CLOSTRIDIAL INFECTIONS

DALE N. GERDING AND STUART JOHNSON

Clostridial infections are characterized by disease produced by toxins. They include tetanus and botulism, both caused by neurotoxins, and clostridial myonecrosis or gas gangrene, caused by the toxins of *Clostridium perfringens* as well as those of other clostridial species. Although these clostridial infections remain important clinically, their frequency has declined markedly with the advent of vaccines and better public health measures. Several previously little known species of clostridia that produce very large clostridial cytotoxins (LCCs) varying in size from 250 to 308 kD have become increasingly prominent. They include, in particular, *Clostridium difficile*, which is responsible for one of the most common health care–associated infections and for increasing mortality among elderly patients. Another LCC-producing organism, *Clostridium sordellii*, has caused devastating infections in young women in association with pregnancy and medical abortion, in injection drug users, and in patients with traumatic wounds. *Clostridium novyi* type A is a third LCC organism that has also caused severe infections in injection drug users.

CLOSTRIDIUM DIFFICILE INFECTION

DEFINITION AND PATHOGEN

C. difficile infection (CDI) is a gastrointestinal infection characterized by diarrhea (three or more loose or unformed stools in ≤24 hours) and a positive test result for *C. difficile* toxin A or toxin B in stool, evidence of a toxin-producing strain of *C. difficile* in stool, or evidence of pseudomembranous colitis on direct visualization of the colon. *C. difficile* is a spore-forming, anaerobic, gram-positive organism that survives well in water, soil, and animals and has a worldwide distribution.

EPIDEMIOLOGY

CDI occurs most frequently in health care settings, particularly in long-term care facilities and acute care hospitals, and is most frequent and lethal among the elderly, especially those older than 75 years. Rates in U.S. hospitals have nearly tripled in the decade since 2000, and it is now estimated that there are 500,000 to 750,000 CDIs each year. A specific strain of *C. difficile*—identified as the restriction endonuclease group BI, pulsed-field gel type NAP1, polymerase chain reaction ribotype 027 (BI/NAP1/027) strain—is thought to be responsible for much of the epidemic that has extended to Canada and Europe. Rates of community-onset CDI have also increased during the past decade, in association with the rising rates in hospitals and nursing homes.[1] However, 94% of community-onset cases in this population-based study were associated with receiving some kind of recent health care, including outpatient visits, recently discharged patients, and nursing home residents. It is unlikely that community-associated cases (with no health care exposure) account for more than 10 to 15% of all CDIs. Risk factors for CDI include antimicrobial use, advanced age, and stay in an acute or chronic care facility. Hospitals are considered a particularly high-risk environment because patients are elderly, antibiotic use is frequent, the environment is contaminated with *C. difficile* spores (which are difficult to eradicate), asymptomatic patients carry *C. difficile* in their stools, and health care workers carry *C. difficile* on their hands if they do not practice good hand hygiene. In about 2 to 3% of healthy adults, *C. difficile* can be cultured from their stools, but the frequency in asymptomatic hospitalized patients increases with the duration of hospitalization and may reach 20% or more. Exposure to nearly all

antimicrobials has been associated with subsequent CDI, but those with the highest risk are clindamycin, the cephalosporins, and the fluoroquinolones. CDI is rare in children and young adults, despite their frequent exposure to antimicrobials. However, children younger than 1 year are commonly colonized with *C. difficile* but remain asymptomatic, an observation that remains largely unexplained.

PATHOBIOLOGY

CDI risk appears to be minimal in the absence of antimicrobial therapy. When antimicrobials are administered, they have the unintended consequence of disrupting the normal protective bowel microbiota, which may persist for days to weeks after the antimicrobial is stopped. If *C. difficile* is ingested during this time, the spores germinate in the gut, and the vegetative form of the organism multiplies and begins to make toxins. At this point, whether the patient will develop diarrhea is dictated by the status of his or her immunity to the toxins, which is best correlated with serum immunoglobulin G antibodies directed at toxin A and toxin B of *C. difficile*. Those with good antibody responses will be asymptomatic but will remain colonized with *C. difficile*, whereas those with little or no antibody response will develop diarrhea and CDI. Toxin A is primarily an enterotoxin, and toxin B is a cytotoxin. Both act by glucosylation that disrupts the cell cytoskeleton, resulting in colonic epithelial cell rounding, fluid leakage, and cell death. In the presence of pseudomembranous colitis, the colon appears to be covered in yellow to white pseudomembranes that vary in size from punctate to completely confluent and covering the entire colon in advanced cases. On histologic evaluation, the colon demonstrates a marked neutrophil infiltration throughout the colon wall, with mucosal necrosis and volcano-like lesions from which the pseudomembrane is seen to "erupt." The pseudomembrane is composed of proteinaceous material and cellular debris.

CLINICAL MANIFESTATIONS

Clinical symptoms of CDI range from asymptomatic carriage to severe and sometimes life-threatening pseudomembranous colitis complicated by major fluid losses and systemic complications. With mild CDI, patients may simply have "nuisance diarrhea" that resolves when the implicated drug is discontinued. Others with more severe CDI have substantial fluid and protein losses combined with fever, cramps, hypoalbuminemia, leukocytosis, and hypotension. Leukocytosis is common (occurring in up to 50% of patients) and is a marker of severe CDI when it exceeds 15,000/mm^3. Extremely high levels (>50,000/mm^3) are an indication of fulminant and potentially fatal illness. Other factors that may be indicative of severe or late-stage disease include toxic megacolon, high fever, renal failure, hypotension, and a lactic acidosis greater than 5.0 mmol/L.

DIAGNOSIS

The diagnosis should be suspected in any patient who has otherwise unexplained diarrhea (three or more loose or unformed stools in ≤24 hours) in association with recent or concurrent antibiotic use. Only about 10 to 20% of patients in this category actually have CDI, but this is the group of patients that should be tested. The standard test to establish the diagnosis is to detect toxins A and B in stool or to detect a toxin-producing strain of *C. difficile* in stool. Tests that detect only toxin A are not acceptable because about 1 to 3% of strains that cause CDI produce toxin B but not toxin A. The most common laboratory method has been an enzyme immunoassay, but its sensitivity is only 50 to 80%. Repeated testing does not improve the diagnostic accuracy because enzyme immunoassays also have specificity deficiencies that increase the rate of false-positive test results with repeated testing. Polymerase chain reaction tests for *C. difficile* in stool are widely available commercially and markedly improve test sensitivity to 90 to 95% (compared with the "gold standard" of culture for a toxigenic strain). The polymerase chain reaction tests are more expensive, but their use by clinical laboratories is increasing rapidly. They increase the sensitivity of CDI diagnosis, but some studies have shown that a positive test result for stool toxin should also be present to confirm the diagnosis.[2] The standard cell cytotoxin assay is little used but more sensitive than enzyme immunoassay, but it has the disadvantages of a 24- to 48-hour turnaround and the requirement for tissue culture facilities. Stool culture for *C. difficile*, which is the most sensitive test available, is likewise slow and requires confirmation of toxin production by the organism before reporting. Testing of stools for glutamate dehydrogenase, or "common antigen," can be used as a rapid screening test, but it is only about 50% specific and requires confirmation with a toxin test, which can increase the turnaround time for reporting positive results. CDI can also be diagnosed by observing pseudomembranous colitis directly through sigmoidoscopy or colonoscopy or at surgery. When a negative test result does not confirm the diagnosis in a patient whose clinical symptoms are highly suggestive of CDI, empirical treatment of CDI should be given rather than repeating the test. The cause of infectious diarrhea in an adult patient with an onset longer than 48 hours after hospital admission is almost always CDI because other infectious enteric pathogens (with the exception of norovirus) are extremely rare in the hospital setting. Also common in the differential diagnosis are noninfectious causes of diarrhea, such as antibiotic or other drug-associated diarrhea, ischemic colitis, and idiopathic inflammatory bowel disease. For patients with antibiotic-associated diarrhea and no evidence of colitis, the cause of diarrhea with a negative *C. difficile* toxin assay is usually not defined.

PREVENTION

Two major prevention strategies are employed. The first is traditional infection control, in which barriers to transmission (gowns, gloves, isolation, hand hygiene, environmental cleaning) are used to prevent the spores of *C. difficile* from reaching the patient. The second strategy is to reduce the likelihood of infection if the patient does encounter *C. difficile* while in the hospital. The most efficacious strategy, also known as antimicrobial stewardship, is to avoid or to minimize exposure to antimicrobials, especially those with a high CDI risk, such as clindamycin, cephalosporins, and fluoroquinolones. Interventions to restrict exposure to clindamycin and to cephalosporins have been highly effective in interrupting outbreaks of CDI in hospitals.[3]

TREATMENT

Rx

Treatment of CDI begins with discontinuation of the implicated antibiotic, supportive care, and avoidance of antiperistaltic agents. Mildly ill patients may recover with these simple conservative measures, but most require specific treatment. Continuation of the offending antibiotics while CDI is being treated with vancomycin and other agents results in lower cure rates and higher CDI recurrence rates in patients. Metronidazole, at a dose of 500 mg orally three times a day for 10 to 14 days, has been the recommended treatment for patients with mild CDI because it is much less expensive and it avoids some of the concern about vancomycin-resistant enterococcus colonization. However, a large randomized prospective trial of vancomycin versus metronidazole has placed the recommendation into question as it showed vancomycin to be statistically superior to metronidazole for all patients with CDI (P = .02). Patients with severe CDI (variously defined as a white blood cell count >15,000 or a creatinine increase to >1.5-fold above baseline) should be treated with vancomycin 125 mg orally four times a day for 10 days.[4] Vancomycin can be substituted for metronidazole if there is a delayed response in patients with mild illness. A newer option is fidaxomicin (200 mg orally twice daily), a narrow-spectrum macrocycle that is as efficacious as vancomycin and reduces recurrent infection. The anticipated response to these drugs is rapid defervescence, with gradual normalization of bowel habits. Mean time to resolution of diarrhea is about 3 days; if symptoms have not resolved by day 5 or day 6 of treatment, a change in therapy should be considered. However, there are no data to support the use of more than one drug at a time to treat CDI. Failure to respond often means that either the disease has progressed too far or another condition is causing the symptoms. For patients with severe complicated or fulminant CDI, medical management includes vancomycin at a higher dose (500 mg four times a day) orally or by nasogastric tube; if ileus is present, metronidazole (500 mg IV every 8 hours) is added, with vancomycin also administered by enema. If symptoms progress with this therapy, colectomy may be life-saving and should be performed before the white blood cell count reaches 50,000/mm^3 or lactate concentration reaches 5.0 mmol/L. A colon-sparing loop ileostomy procedure followed by infusion of polyethylene glycol and vancomycin has demonstrated reduced mortality compared with historic colectomy controls and may be a preferred procedure to colectomy.[5]

Other antibiotic options include oral fusidic acid, teicoplanin, nitazoxanide, rifaximin, and bacitracin, but most of these drugs have been evaluated in only a small number of patients, and none (like metronidazole) has Food and Drug Administration approval for the treatment of CDI. There is no convincing evidence that toxin-binding agents, such as cholestyramine and probiotics, are useful in treating CDI. About 20 to 25% of patients treated with either vancomycin or metronidazole have a recurrence of symptoms when treatment is stopped because of the persistence of *C. difficile* spores or acquisition of a new strain. Treatment with vancomycin, fidaxomicin, or metronidazole is recommended, depending on the severity of the recurrence. Patients with multiple recurrences of CDI are extremely difficult to treat and may benefit from consultation with an infectious disease specialist or gastroenterologist. Observational data and one prospective randomized controlled trial suggest that fecal microbiota transplantation might be an effective treatment for recurrent *C. difficile* infection.

PROGNOSIS

The majority (≈80%) of patients respond to simple withdrawal of the implicated antibiotic combined with a single course of metronidazole, vancomycin, or fidaxomicin. Some patients with fulminant disease eventually require colectomy. The attributable mortality rate is as high as 7% in large series, and the majority of lethal cases occur in patients older than 65 years. Patients with multiple recurrences require repeated courses of antibiotics, usually vancomycin with tapering and pulse-dose regimens that may need to be continued for weeks to months, or may require resorting to fecal microbiota transplantation.

NECROTIZING CLOSTRIDIAL TISSUE INFECTION

Clostridium sordellii

DEFINITION AND PATHOGEN

C. sordellii is another clostridial species that produces LCCs, and it has become more common as a cause of septic shock and necrotizing fasciitis in association with trauma, childbirth, medical abortion, and injection drug use. *C. sordellii* antitoxin cross-neutralizes the cytotoxic effect of *C. difficile* toxins, indicating the similarity of these LCCs.

EPIDEMIOLOGY

The organism is commonly found in soil and in the intestines of animals and occasionally humans worldwide. Soil contamination of wounds is the usual suspected route of infection. Infections have been described after traumatic wounds, childbirth, medical abortion, and intramuscular or subcutaneous injection drug use.

PATHOBIOLOGY

C. sordellii produces up to seven identified toxins; of these, hemorrhagic toxin (TcsH) and lethal toxin (TcsL), which are both LCCs and analogous to TcdA and TcdB of *C. difficile*, are considered major virulence factors. TcsL has been shown to be essential for virulence.[6] The toxins produce local necrosis, progressive edema, and shock that results in high mortality. The toxins are glucosyltransferases that glucosylate the Rho, Rac, or Ras proteins, causing impaired cytoskeletal organization and massive capillary leakage, leading to the progressive edematous state.

CLINICAL MANIFESTATIONS

Initial symptoms are nonspecific and include nausea, lethargy, dizziness, and tenderness at the infection site. Tachycardia and hypotension follow within hours. Laboratory tests show a marked leukocytosis or leukemoid reaction. Hypotension and tachycardia are refractory to treatment. Edema secondary to capillary leakage is prominent, resulting in hemoconcentration but little or no fever. Peritoneal and pleural effusions are common. White blood cell elevations greater than 75,000/mm^3 are associated with a fatal outcome.

Infection Associated with Childbirth and Medical Abortion

This is a rare infection in which patients present 4 to 7 days after the administration of oral mifepristone and vaginal misoprostol for medical abortion with nausea, vomiting, weakness, abdominal pain, hypotension, and tachycardia but little or no fever and, often remarkably, a lack of findings on pelvic examination. The presentation is similar to toxic shock syndrome. There is rapid progression to vascular collapse and cardiac arrest. Leukocytosis is dramatic, with white blood cell counts near 100,000/mm^3 in most patients. After delivery, the presentation is similar, but localized swelling and discoloration of the labia and perineum may be evident if the episiotomy site is infected. All patients to date have died.

Infection Associated with Injection Drug Use

Black tar heroin (a dark, gummy, less refined and cheaper form of heroin) has been associated with necrotizing fasciitis at the subcutaneous or intramuscular injection site, presumably a result of the contaminants mixed with the heroin. Patients present with 2 to 7 days of symptoms of necrotizing fasciitis of the upper or lower extremity where they injected heroin, in some cases accompanied by hypotension. Aggressive surgical débridement is required at the infection site, together with fluids and pressors. Cultures of débrided tissue reveal multiple organisms in addition to *C. sordellii*. Mortality in these patients is 50%.

DIAGNOSIS

Diagnosis is difficult because of the lack of specific symptoms, but it is usually made by identifying the likely source of infection and isolating *C. sordellii* from the infection site or blood cultures. Polymerase chain reaction analysis for *C. sordellii* in infected tissues may be required to make the diagnosis when cultures are negative. Other bacteria are also commonly found at the infection site. Computed tomography or magnetic resonance imaging may be helpful in identifying localized infections, which can then be excised or drained surgically, providing material for microbiologic diagnosis.

PREVENTION

C. sordellii infections are very rare, and the exact mechanism of infection, especially after medical abortion, is not known. However, careful attention to wound cleansing, avoidance of injecting drugs into skin or muscle, and good hygiene during childbirth are likely to help prevent *C. sordellii* infection.

TREATMENT Rx

Definitive treatment information is lacking. The infection progresses so rapidly that therapeutic interventions are rarely successful. Surgery to remove necrotic sites of infection and administration of intravenous fluids and pressors to treat hypotension and tachycardia are supportive. Antibiotic susceptibility suggests that β-lactams, clindamycin, tetracyclines, and chloramphenicol are active, but no clinical treatment efficacy data are available. Theoretically, use of an antibiotic such as clindamycin to suppress toxin synthesis could be a useful adjunct to treatment. At present, there is no antitoxin available.

PROGNOSIS

Infections after childbirth or medical abortion have been uniformly fatal. Mortality in injection drug users or patients after trauma or surgery is about 50%.

Clostridium novyi Infection in Injecting Drug Users

C. novyi α-toxin causes an LCC-producing disease in humans. *C. novyi* has long been recognized as a cause of fatal toxemia in animals with contaminated wounds. An extended outbreak of human infections occurred in Scotland, Ireland, and elsewhere in the United Kingdom in 2000 to 2009 among persons injecting heroin extravascularly (skin or muscle popping).[7] A localized necrotizing infection with painful edema, sepsis, and significant mortality was recognized. The findings include soft tissue inflammation, edema and necrosis at the injection sites, circulatory collapse, marked leukocytosis, and pleural effusions. Treatment usually involves débridement and antibiotic administration (gentamicin, flucloxacillin, penicillin, metronidazole, and clindamycin were used during this outbreak because infections are typically polymicrobial in origin).

Clostridial Myonecrosis (Gas Gangrene)

DEFINITION

Clostridial myonecrosis, or gas gangrene, can be caused by several *Clostridium* species, most commonly *C. perfringens* after trauma or tissue injury and *Clostridium septicum* after dissemination from a colonic source.[8]

EPIDEMIOLOGY

Gas gangrene has historically been a complication of battlefield injuries and of trauma in noncombat settings. The estimated number of cases in the United States is about 1000 per year. Traumatic injuries account for about 50% of cases, with vehicular accidents accounting for the majority; the remaining cases develop in patients after crush injuries, industrial accidents, gunshot wounds, and burns. Minor injuries such as puncture wounds, intramuscular injections, simple lacerations, and subcutaneous injections with epinephrine may occasionally precipitate clostridial myonecrosis. Postoperative complications account for about 30% of cases and are most frequently associated with surgery on the appendix, biliary tract, or intestine. Approximately 20% are "spontaneous" or nontraumatic and are invariably associated with an occult colonic malignant neoplasm.

PATHOBIOLOGY

Clostridia are widely distributed in nature and can be cultured from nearly all soil samples, from environmental sites in the hospital, and from the human

intestine. A critical factor is a physiologic state of the wound with conditions that support germination and toxin production by toxigenic clostridia. Particularly critical are a low oxidation-reduction potential, hypoxia, appropriate substrates, and calcium ions. The probability of infection is substantially increased by devitalized muscle and the presence of foreign material such as soil. *C. perfringens* elaborates at least 12 recognized toxins, most importantly α-toxin and θ-toxin of *C. perfringens* type A. Although the interaction is complex, good evidence supports a central role for α-toxin, a phospholipase C, and θ-toxin or perfringolysin O, a cholesterol-dependent cytolysin, in the extensive cell death and disruption of microvascular perfusion that are characteristic of clostridial myonecrosis. The vascular perfusion changes are likely mediated by toxin-induced platelet aggregation and leukocyte margination. The α-toxin of *C. septicum*, a pore-forming cytolysin unrelated to α-toxin of *C. perfringens*, also causes cell death and microvascular perfusion changes.

CLINICAL MANIFESTATIONS

Initial symptoms of traumatic myonecrosis usually occur 1 to 4 days after the precipitating event, although the range is 8 hours to 3 weeks. The initial symptom is pain that is often sudden and severe at the site of surgery or trauma. The involved skin has intense edema and is initially pale before progressing to a bronze or magenta color, followed by the formation of bullae. The bullae contain fluid that may be clear or hemorrhagic. The discharge has an odor that is described as "foul-sweet."

Circulatory collapse and hypotension unresponsive to fluid challenge are common and may reflect the effect of α-toxin, which suppresses cardiac contractility. About 15% of patients have bacteremia that is usually complicated by rapid hemolysis with a dramatic drop in the hematocrit, which may even decrease to 0%. Common complications include jaundice, hypotension, hepatic failure, and renal failure. The renal failure is often due to hemoglobulinuria and myoglobulinuria, but it may also be due to acute tubular necrosis from hypotension. Despite the severity of the illness, the patient's mental status is usually remarkably good until very late in the disease. Surgical intervention shows necrotic muscle that does not contract with stimulation. Deeper dissection reveals beefy red necrotic muscle that becomes black and extremely friable in the late stages.

Uterine gas gangrene, which was once common after septic abortions, is now rare but may complicate normal delivery, amniocentesis, cesarean section, or abortion. The onset is usually sudden, with fever, tachycardia, hypotension, renal failure, and jaundice. Radiography may show gas in the uterine wall. The urine is often "port wine" in color as a result of hemoglobulinuria, and there is often jaundice because of massive intravascular hemolysis. The usual causes are *C. perfringens* and *C. sordellii.*

Spontaneous myonecrosis occurs in the absence of trauma and is usually caused by *C. septicum*. One distinctive association is with colon cancer and neutropenic enterocolitis, which represent portals of entry for hematogenous seeding of *C. septicum*. This infection is also seen with acute leukemia and is most common with chemotherapy for solid tumors and after stem cell transplantation. The usual portals of entry are the terminal ileum, cecum, and ascending colon, hence the term *typhlitis*. The usual manifestation is a necrotizing infection in an extremity or in the abdominal wall, accompanied by hypotension and renal failure. Examination shows spreading crepitations with rapid clinical deterioration during a period of hours, and imaging reveals gas.

DIAGNOSIS

The diagnosis of gas gangrene is usually based on a constellation of characteristic clinical features, including myonecrosis, shock, and renal failure. The patient typically complains of severe pain. Early recognition is important because early institution of treatment may strongly influence the prognosis. The diagnosis is established by examination of skin and muscle, which shows putrid discharge, characteristic bullae, and crepitations. Gram stain demonstrates abundant gram-positive bacilli and no inflammatory cells. Histopathologic examination of the lesion shows myonecrosis without polymorphonuclear leukocytes, a finding that is remarkably different from most soft tissue infections, which do not feature necrosis and have abundant inflammatory cells. Gas is present in the tissue and may be detected by physical examination, radiography, or other imaging methods.

PREVENTION

The basic principle of prevention is adequate management of traumatic wounds—establishing adequate drainage, removing foreign bodies, draining hematomas, and ensuring good hemostasis.

TREATMENT

Rx

The most important facet of treatment is prompt surgical débridement. Many cases require extensive, often mutilating surgery. Penicillin and clindamycin are recommended but are rarely adequate without radical surgery, except in patients with neutropenic enterocolitis, who can often be managed with antibiotics. The rationale for penicillin combined with clindamycin is that some strains of clostridia are resistant to clindamycin, but clindamycin is probably the superior drug for reducing toxin formation. Other antibiotics that are generally effective include metronidazole and chloramphenicol. The use of hyperbaric oxygen is controversial, in part because the therapeutic trials have been either of poor quality or not convincing.

PROGNOSIS

Factors associated with a poor prognosis include advanced age, location on the trunk, association with severe underlying disease, leukopenia, renal failure, hemolysis, and shock. The best results are seen in young patients with involvement of a single extremity. Management plays an important role, particularly the use of early and aggressive surgery as well as antibiotics. The overall mortality rate of patients with traumatic gas gangrene in tertiary centers is about 25%.

NEUROTOXIC CLOSTRIDIAL INFECTIONS

Botulism

DEFINITION

Botulism is a severe neuroparalytic disease characterized by a descending flaccid motor neuron paralysis. It is caused by botulinum toxin produced by *Clostridium botulinum.*

The Pathogen

C. botulinum is a gram-positive, spore-forming obligate anaerobe that is widely distributed in nature and is frequently found in soil, marine environments, and agricultural products. Each strain produces one of seven toxins designated by the letters A to G. A new botulism toxin designated H, the first new neurotoxin identified in more than 40 years, was recently isolated from a patient with infant botulism.[9] Botulinum toxin may also be produced by the related clostridial species *Clostridium baratii* and *Clostridium butyricum*. All these neurotoxins produce the same syndrome; the usual causes of disease in humans are types A, B, and E, with rare cases caused by type F.

EPIDEMIOLOGY

Botulism in humans is generally one of three types: food-borne botulism, infant botulism, or wound botulism. Rarely, botulism may be acquired as a result of iatrogenic misadventures with botulinum toxin, which is a potential bioterrorism agent if it is inhaled or ingested.

Food-borne botulism is the most common form of botulism in the world but is a distant second to infant botulism in the United States. Nevertheless, 20 cases from ingestion of prison-made alcohol (pruno) and the first eight cases in more than 30 years from ingestion of a commercially canned product (hot dog chili sauce) have occurred recently.[10,11] Foods most frequently implicated are home-canned vegetables or fermented foods, and most cases are sporadic single cases occasionally involving two or three people. Commercially preserved and restaurant-prepared foods are also rare causes of food-borne botulism. Type A toxin is predominant in the United States. Alaska has the highest rate of any state, with approximately 35% of all cases; 80% of events and cases in Alaska are caused by type E and are most often associated with fermentation methods used to prepare fish and marine mammals by native Alaskans.[12] Meat and meat products are frequently implicated in Europe, where the predominant toxin is type B. In China, the most frequent vehicle is a vegetable product, and type A predominates.

Infant botulism is the most frequently recognized form in the United States and is the most recently discovered type of botulism, first described in 1976. It is caused by production of botulinum toxin in the intestine after presumed spore ingestion and colonization in 2- to 36-week-old infants. Honey has been identified as the source of *C. botulinum* spores, but in most cases the source is never identified. Nearly all cases are caused by type A or type B toxin. The symptoms usually begin with constipation followed by poor feeding, weak cry, lethargy, and generalized weakness characterized as the

"floppy baby syndrome" because of loss of head control. This form of botulism is rare in adults, occurring most often in patients with anatomic or functional abnormalities of the intestines.

Wound botulism, first described in 1943, is the least frequent form of the disease and is usually caused by either type A or type B toxin. Sporadic cases in traumatic wounds contaminated by soil are rarely reported. Outbreaks have been described in black tar heroin users in the western United States, particularly if they inject the drug intramuscularly or subcutaneously (skin popping). These drug users also develop other clostridial infections, including necrotizing fasciitis caused by *C. sordellii* and *C. novyi* and tetanus caused by *Clostridium tetani.*

Inhalation or ingestion of botulinum toxin is considered one of the top six bioweapon agents in terms of probability of use. The presumed method would be contamination of the food supply, water supply, or commercial beverages or aerosolization in a highly populated area to cause inhalational botulism. It is estimated that a point-source aerosol release of the toxin could incapacitate or kill 10% of people within a 0.5-km radius.

Iatrogenic botulism results from the misuse of botulinum toxin for cosmetic or therapeutic purposes. Cosmetic treatment doses are far too low to cause systemic disease, but the use of unlicensed products with high concentrations of botulinum toxin can cause systemic symptoms. Higher doses used for the management of muscle movement disorders have caused occasional cases with systemic botulism-like symptoms.

PATHOBIOLOGY

Pathologic findings are due to absorption of toxins from the intestine (ingested preformed toxin in foods or in situ production in the intestine in infants), inhalation (aerosol from bioterrorism), absorption from cutaneous infection sources (wounds), or iatrogenic injection. The toxin is disseminated by the systemic circulation and causes flaccid paralysis by binding presynaptic motor neuron terminals and blocking acetylcholine transmission across the neuromuscular junction. The estimated lethal doses of purified botulinum toxin A for a 70-kg human are 0.09 to 0.15 μg when given intravenously, 0.8 to 0.9 μg when inhaled, and 70 μg when given orally.

CLINICAL MANIFESTATIONS

In contrast to Guillain-Barré syndrome (Chapter 420), which is an ascending paralysis, botulism is characterized by generalized weakness and a descending paralysis. Symptoms are due to absorption of botulinum toxin from the gut, the lung, or a wound. Clinical symptoms consist of highly distinctive and usually symmetrical cranial nerve palsies, followed by a symmetrical descending flaccid paralysis. Prominently involved cranial nerves III, IV, and VI cause blurred vision and diplopia; involvement of cranial nerve VII causes the characteristic expressionless facies and dysphagia; and involvement of cranial nerve IX causes dysarthria. Thus, the initial symptoms include the "four *d*'s"—diplopia, dysarthria, dysphagia, and dysphonia—although the last is rarely reported, and blurred vision is reported more commonly than diplopia. These findings are followed by a descending upper extremity paralysis and respiratory paralysis. Neurologic examination shows bilateral cranial nerve VI paresis, ptosis, dilated pupils with a sluggish reaction, and diminished gag reflex, followed by descending involvement of motor neurons. Deep tendon reflexes are diminished or absent. Mentation remains clear, vital signs are normal, and the neurologic findings are symmetrical. The most common cause of death is respiratory failure. The tempo of the disease and the extent of paralysis in the absence of treatment are highly variable. The symptoms may be restricted to a few cranial nerves, or there may be complete paralysis of all voluntary muscles. Progression may occur during a period of hours or days. The timing and extent of neurologic deficits depend on the size of the botulinum toxin inoculum.

DIAGNOSIS

Botulism should be suspected in patients with an acute flaccid paralysis involving the cranial nerves, particularly in the presence of bilateral cranial nerve VI dysfunction, associated neurologic findings, and a 10-hour to 5-day history of consuming suspect food, such as preserved or home-canned foods. Nausea, vomiting, abdominal pain, and diarrhea are common early in the illness, with constipation present when paralysis develops. The finding of two or more cases that are epidemiologically linked is virtually diagnostic of food-borne botulism because other causes of paralysis are rare and sporadic. In the absence of a history of suspect food ingestion, potentially infected wounds should be sought, including injection sites in users of black tar heroin. With bioterrorism, the epidemiology may reflect a common source exposure, such as a local water supply or an aerosolized toxin, but it could also be widely distributed with a contaminated food source, such as the milk supply.

Laboratory tests for suspected food-borne botulism include analysis of serum, stool, gastric contents, or food for botulinum toxin and culture of stool and suspect food or wounds for *C. botulinum.* Toxin assay specimens should be collected before treatment with antitoxin. With wound botulism, recovery of *C. botulinum* from wound cultures or detection of toxin in serum is diagnostic. The toxin assays are generally available only at public health laboratories. The standard is a mouse bioassay for detection and quantification of toxin. Toxin type is determined by type-specific antibody neutralization. In general, adult patients with clinical evidence of food-borne botulism have detectable toxin in sera in a third of cases and detectable toxin in stool in a third of cases, but the organism is recovered from stool in about 60%.

The differential diagnosis includes myasthenia gravis, Guillain-Barré syndrome, tick paralysis, cerebrovascular accident, trichinosis, Eaton-Lambert syndrome, hypocalcemia, hypermagnesemia, organophosphate poisoning, atropine poisoning, and paralytic poisoning by shellfish or puffer fish. Electromyography using repetitive stimulation at 2 to 50/second may be helpful in distinguishing causes of flaccid paralysis. Electromyography patterns with slow and rapid supramaximal stimulation show similar responses in botulism and Eaton-Lambert syndrome. Findings on cerebrospinal fluid analysis and cranial imaging are normal in botulism.

PREVENTION

The disease can be prevented by destroying spores in the original food source, inhibiting germination, or destroying preformed toxin.

TREATMENT

Rx

Clinicians who suspect botulism should immediately seek clinical consultation, notify public health authorities, and administer antitoxin. The agency to contact in the United States is the state health department, which will contact the Centers for Disease Control and Prevention (CDC) if needed. Additional emergency consultation is available from the CDC botulism duty officer through the CDC Emergency Operations Center (telephone: 770-488-7100); similar public health agencies should be contacted in other countries. Treatment consists of supportive care and passive neutralization with equine botulinum antitoxin.

The standard treatment of adults since 2010 in the United States is heptavalent botulinum antitoxin (HBAT) through a CDC-sponsored Food and Drug Administration investigational new drug protocol.[13] HBAT contains equine-derived antibody to the seven known botulinum toxin types (A to G) with the following nominal potency values: 7500 U anti-A; 5500 U anti-B; 5000 U anti-C; 1000 U anti-D; 8500 U anti-E; 5000 U anti-F; and 1000 U anti-G. BabyBIG (botulism immune globulin) is a human-derived treatment of infant botulism types A and B and is available for infant botulism through the California Infant Botulism Treatment and Prevention Program. The antitoxin should be given as early as possible and should not be delayed while awaiting microbiologic results. This treatment does not reverse paralysis or neutralize toxin already bound to nerve endings, but it does neutralize unbound toxin in the circulation to prevent progression. The HBAT antitoxin is derived from horses, and as a result, hypersensitivity reactions may occur.

Respiratory failure is a major risk, and patients must be monitored carefully with liberal criteria for ventilatory support. The requirement for mechanical ventilation varies from about 20% in adults with food-borne disease to 60% in patients with infant botulism. Other forms of supportive care include enteral or parenteral feeding and positioning in the reverse Trendelenburg position.

Toxin can be removed from the gastrointestinal tract by gastric lavage, cathartics, and enemas early in the course. Antibiotic treatment is unnecessary, except for wound botulism.

PROGNOSIS

The case-fatality rate for untreated food-borne botulism was once 60 to 70% but is currently 3 to 5% with treatment. Infant botulism in the United States now has a mortality rate of less than 1%; the use of human antitoxin has reduced the median duration of hospitalization from 6 to 3 weeks. Patients who survive any form of botulism generally have a complete recovery.

Tetanus

DEFINITION

Tetanus is a neurologic syndrome characterized by generalized rigidity and convulsive spasm of skeletal muscles caused by a neurotoxin elaborated at the site of injury by *Clostridium tetani.*

The Pathogen

C. tetani is an anaerobic, gram-positive, slender, motile bacillus. When it sporulates, the terminal spore gives the organism a characteristic "drumstick" or "tennis racket" shape. The vegetative form produces tetanospasmin, a protein neurotoxin with a molecular mass of approximately 151 kD, including a heavy chain (100 kD) that binds neuronal cells and a light chain that blocks the release of neurotransmitters.

EPIDEMIOLOGY

C. tetani can be found in 2 to 23% of soil samples, with the highest yield in manure-treated soil. The organism can also be found in stool from a variety of domestic and farm animals and poultry. Tetanus is most common in warm climates and in highly cultivated rural areas. The greatest problem occurs in resource-limited countries because of high numbers of unimmunized mothers and unhygienic practices. The estimated annual toll from neonatal tetanus in developing countries is greater than 257,000, mostly secondary to inadequate passive immunity caused by absence of immunity in the mother. In the United States, an average of 29 cases of tetanus were reported annually from 2001 to 2008 with a mortality of 13.2%, and almost all occurred in unimmunized or inadequately immunized persons.[14] In the U.S., patients 65 years of age and older constituted 31% of patients and had the highest mortality at 31%.

PATHOBIOLOGY

Tetanospasmin, also known as tetanus neurotoxin or TeNT, ranks with botulinum toxin as one of the most potent known microbial toxins; 2.5 ng/kg is a lethal human dose. Clinical tetanus usually results from entry of the organism into a wound and low oxygen conditions that allow spore germination and survival of the vegetative organism to produce toxin. Entry is usually through a traumatic or surgical wound, drug injection site, burn, skin ulcer, or infected umbilical cord. Tetanospasmin binds the peripheral nerve terminals and is then carried intra-axonally within membrane-bound vesicles to spinal neurons at a transport rate of approximately 75 to 250 mm/day. The light chain passes to the presynaptic terminals, where it blocks the release of neurotransmitters in inhibitory afferent motor neurons. Loss of the inhibitory influence results in sustained muscle contraction. Binding of the toxin is irreversible, so recovery requires the generation of new axon terminals.

CLINICAL MANIFESTATIONS

Forms of tetanus include generalized, local, cephalic, and neonatal. Generalized tetanus, which is the most common form, accounts for 80 to 90% of reported cases in the United States. The usual incubation period is 3 to 21 days (mean, 8 days), depending largely on the distance between the site of injury and the central nervous system. A short incubation period is associated with more severe symptoms. Generalized tetanus is characterized by a persistent tonic spasm with brief exacerbations. The neck and jaw are almost always involved. Trismus (lockjaw) is the initial complaint in 75% of cases, so the patient is often initially seen by a dentist or oral surgeon. Other early features include irritability, restlessness, diaphoresis, and dysphagia with hydrophobia and drooling. Persistent spasm of the back musculature may cause opisthotonos. These early manifestations reflect involvement of the paraspinous muscles. With progression, all muscles contract, with stronger muscles overtaking weaker muscles. Noise or tactile stimuli may precipitate spasms and generalized convulsions. Involvement of the autonomic nervous system may result in severe arrhythmias, blood pressure oscillation, profound diaphoresis, hyperthermia, rhabdomyolysis, laryngeal spasm, and urinary retention. In most cases, the patient remains lucid and afebrile. The condition may continue for 3 to 4 weeks, despite antitoxin therapy, because of the time required for intra-axonal toxin transport. Complications include fractures from sustained contractions, pulmonary emboli, bacterial infections, and dehydration.

Local tetanus, in which the patient has persistent muscle contractions in the extremity involving a contaminated wound, is rare and shows considerable variation in severity. In mild cases, a patient may simply have spasms of the involved extremity; in more severe cases, local painful spasms progress to generalized tetanus. This relatively unusual form of tetanus has an excellent prognosis, with only about 1% mortality.

Cephalic tetanus is also rare and generally follows a head injury or occurs with *C. tetani* infection of the middle ear. Clinical symptoms consist of isolated or combined dysfunction of the cranial motor nerves, most frequently cranial nerve VII. This dysfunction may remain localized or progress to generalized tetanus. The incubation period is only 1 or 2 days, and the prognosis for survival is usually poor.

Neonatal tetanus is generalized tetanus resulting from *C. tetani* infection in neonates. It occurs primarily in underdeveloped countries, where it accounts for up to half of all neonatal deaths.

DIAGNOSIS

The diagnosis of tetanus is usually based on clinical observations. The putative agent, *C. tetani*, is recovered from wound culture only about 30% of the time. Results of cerebrospinal fluid analysis are entirely normal. Diagnostic testing is usually not necessary except in cases lacking an identified portal of entry. The differential diagnosis depends on the dominant clinical features and includes dystonic reactions as a result of neuroleptic toxicity, seizure disorders, hypocalcemic or alkalotic tetany, alcohol withdrawal, and strychnine poisoning. Strychnine also antagonizes glycine, and strychnine poisoning is the only condition that truly mimics tetanus. Strychnine levels in blood and urine establish the diagnosis. Dystonic reactions may resemble tetanus and are distinguished by rapid response to anticholinergic agents.

PREVENTION

Immunization with tetanus toxoid is virtually 100% effective, so nearly all cases of tetanus occur in unimmunized or inadequately immunized individuals. The Advisory Committee on Immunization Practices has recommended diphtheria and tetanus toxoids and acellular pertussis vaccine (DTaP) for active immunization of infants and children at 2 months, 4 months, 6 months, 15 to 18 months, and 4 to 6 years of age. Protective levels of serum antitoxin in persons who complete the primary series persist for at least 10 years. Td (tetanus and diphtheria toxoids adsorbed for adult use) is recommended every 10 years, but this recommendation has been modified because of concerns about waning adult pertussis antibody protection; as a result, the Advisory Committee on Immunization Practices recommends that all adults aged 19 years and older who have not yet received a dose of Tdap (tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis) should receive a single dose regardless of the interval since last Td.[15] The recommended primary immunization series for unimmunized persons older than 7 years is Td at time 0, 4 to 8 weeks, and 6 to 12 months after the second dose, and then every 10 years. Nearly all states now require DTaP immunization for school enrollment. Immunization of childbearing women with Tdap confers protection to their infants through transplacental maternal antibody and is recommended during the third trimester of each pregnancy for optimal fetal passive antibody protection.[16]

Prevention of tetanus after injury (Table 296-1) requires appropriate wound management, assurance of adequate immunity, and consideration of antibiotic prophylaxis. The aim of surgery is to eliminate necrotic tissue, purulent collections, and foreign bodies that promote the environmental conditions necessary for spore germination. Passive immunization with tetanus immune globulin (TIG) is recommended only for "tetanus-prone"

TABLE 296-1 GUIDE TO TETANUS PROPHYLAXIS IN ROUTINE WOUND MANAGEMENT

HISTORY OF ADSORBED TETANUS TOXOID (NO. OF DOSES)	CLEAN MINOR WOUNDS		ALL OTHER WOUNDS*	
	Tdap or Td†	*TIG*‡	*Tdap or Td*†	*TIG*‡
<3 or unknown	Yes	No	Yes	Yes
≥3	No§	No	No¶	No

*Such as (but not limited to) wounds contaminated with dirt, feces, soil, and saliva; puncture wounds; avulsions; and wounds resulting from missiles, crushing, burns, and frostbite.
†For children younger than 7 years, DTaP (pediatric diphtheria–tetanus toxoid plus acellular pertussis vaccine) is recommended; if pertussis vaccine is contraindicated, DT (pediatric diphtheria–tetanus toxoid) is given. For persons aged 7 to 9 years or 65 years or older, Td (adult tetanus–diphtheria toxoid) is recommended. For persons 10 to 64 years, Tdap (adult tetanus–diphtheria toxoid plus acellular pertussis vaccine) is preferred to Td if the patient has never received Tdap and has no contraindication to pertussis vaccine. For persons 7 years and older, if Tdap is not available or not indicated because of age, Td is preferred to TT (tetanus toxoid alone). Note that pediatric formulations (DT and DTaP) contain an amount of tetanus toxoid similar to that of adult Td, but they contain three to four times as much diphtheria toxoid. DTaP and Tdap vaccines do not contain thimerosal as a preservative.
‡TIG is human tetanus immune globulin. Equine tetanus antitoxin should be used when TIG is not available.
§Yes, if more than 10 years since the last dose.
¶Yes, if more than 5 years since the last dose.

wounds in patients with an inadequate or unknown primary immunization status. The determination of whether a wound is tetanus prone depends on the interval between injury and treatment, the degree of contamination, the extent of devitalized tissue or foreign bodies at the site of injury, and the depth of injury. Antimicrobial agents, such as penicillin and metronidazole, may be given to inhibit replication of the vegetative forms of *C. tetani*, but immunization and wound cleansing are considered more important.

TREATMENT

Patients with tetanus require intensive care with particular attention to respiratory support, treatment with benzodiazepines, autonomic nervous system support, passive and active immunization, surgical débridement, and antibiotics directed against *C. tetani*.[17] There may be clinical progression for 2 to 4 weeks, despite antitoxin treatment, because of the time required to complete the transport of toxin. The severity of disease may be reduced by partial immunity; as a result, some patients have mild disease with minimal mortality, whereas others have mortality rates as high as 60% despite expert care.

Supportive Care

It is most important to assess airway function. Many patients require endotracheal intubation with benzodiazepine sedation and neuromuscular blockade; a tracheostomy should be placed if the endotracheal tube causes spasms. A feeding tube is usually required for nutritional support.

Control of Muscle Spasms

Benzodiazepines have become the mainstay of therapy to control spasms and to provide sedation. The most extensively studied is diazepam given in 5-mg increments; lorazepam and midazolam are equally effective. Patients with tetanus may have high tolerance for the sedative effects of these drugs and may require exceptionally high doses. When tetanus symptoms resolve, the drugs must be tapered during at least 2 weeks to prevent withdrawal reactions. If control of spasms cannot be achieved by benzodiazepines, long-term neuromuscular blockade is performed with vecuronium (6 to 8 mg/hour).

Passive Immunization

Human TIG should be given as soon as possible to neutralize toxin that has not entered neurons. The usual dose is 500 IU intramuscularly. Higher doses or intrathecal administration does not appear to be more effective. An alternative to TIG is pooled intravenous immune globulin. Equine TIG is equally effective, but the rate of allergic reactions is high because of the equine source; this preparation should not be used if human TIG is available.

Active Immunization

The standard three-dose schedule of immunization with tetanus toxoid should be given at an injection site separate from that used for immune globulin.

Antibiotic Therapy

C. tetani is susceptible in vitro to penicillins, cephalosporins, imipenem, macrolides, metronidazole, and tetracyclines. Clinical studies favor the use of metronidazole, which should be given in an intravenous dose of 2 g/day for 7 to 10 days.

Autonomic Nervous System Dysfunction

This complication generally reflects excessive catecholamine release and is usually treated with labetalol (0.25 to 1.0 mg/minute) for blood pressure control. Hypotension may require norepinephrine infusion. Bradycardia may require a pacemaker.

Surgery

Any wounds should be appropriately débrided.

PROGNOSIS

The overall mortality rate for generalized tetanus is 20 to 25%, even in modern medical facilities with extensive resources. Patients with moderate or severe generalized tetanus generally require treatment for 3 to 6 weeks. The highest mortality rates are at the extremes of age. The most frequent cause of death is pneumonia, but many patients have no obvious findings at autopsy, suggesting that death was directly due to the neurotoxin. Patients who recover usually recover completely.

OTHER CLOSTRIDIAL INFECTIONS

Clostridium perfringens Type C Enteritis

C. perfringens type C enteritis, also called enteritis necroticans, is a necrotizing disease involving the proximal small intestine caused by β-toxin–producing strains of *C. perfringens*. Enteritis necroticans occurs as sporadic cases or in outbreaks, most often in underdeveloped countries, most notably in Papua New Guinea in the 1960s and 1970s, where it was called pigbel because of its association with pork feasts by aboriginal people in the highlands. Outbreaks have also been reported among Khmer refugees in Thailand in the 1980s and in Sri Lanka in 2007. Enteritis necroticans also occurs rarely in isolated cases in the developed world, particularly among patients with diabetes mellitus.

PATHOBIOLOGY

Experimental and clinical evidence supports infection with *C. perfringens* type C and β-toxin as the causative agent of and key virulence factor in enteritis necroticans. The organism has been identified at the site of necrotic lesions, the disease can be reproduced in guinea pigs, isogenic β-toxin gene null mutants are avirulent, and vaccination with a toxoid preparation of β-toxin is protective. β-Toxin production is rapidly upregulated in the presence of Caco-2 enterocytes, and the toxin localizes to the endothelium in humans and piglets infected with *C. perfringens* type C.[18] These findings may explain key histopathologic hallmarks of this disease, that is, deep small intestinal necrosis with vascular necrosis and hemorrhage in the lamina propria.

CLINICAL MANIFESTATIONS

In Papua New Guinea, affected patients usually develop severe abdominal pain 12 hours to several days after a ritual pork feast (or, presumably, other infected food). Vomiting and bloody diarrhea are frequently associated findings. The abdomen becomes distended, and thickened bowel loops are sometimes appreciated on palpation. Disease severity and whether the patient experiences spontaneous recovery or bowel perforation and death depend on the extent of intestinal involvement.

DIAGNOSIS

Recognition of the clinical syndrome is critical to making the diagnosis. Culture to identify specific β-toxin–producing strains of *C. perfringens* remains a research tool and is not helpful in the management of patients. Plain radiographs of the abdomen may show dilated small bowel loops and ileus.

PREVENTION

An effective toxoid vaccine was available and used in Papua New Guinea (where the disease is endemic) as well as in the Khmer refugee camp outbreak in 1986. Vaccination was discontinued in the mid-1990s, and the vaccine is no longer available. A 2002 survey of Papua New Guinea children in the highlands suggested that pigbel was responsible for 9 to 16% of acute abdominal cases and clustered in three close geographic regions.

TREATMENT

Treatment is primarily supportive, including nasogastric suction and intravenous hydration. Surgical resection of the infected bowel is often required for those who do not initially respond to supportive measures. Antibiotics (penicillin, chloramphenicol, metronidazole) are almost always given empirically, but their role has not been defined. Prognosis depends on the extent of disease and the availability of surgery for those with more extensive intestinal involvement.

Clostridium perfringens Type A Diarrhea

C. perfringens type A is a well-recognized cause of food poisoning due to the ingestion of food, usually meat, heavily contaminated with enterotoxin-producing *C. perfringens* after storage at inappropriate temperatures. Enterotoxin production is associated with sporulation of ingested vegetative bacteria in the small intestine. The incubation period is 7 to 15 hours after ingestion, and the most prominent symptoms are diarrhea and abdominal pain. The syndrome is usually mild and self-limited.

In contrast to food poisoning due to *C. perfringens* type A, a more severe and protracted infectious diarrhea syndrome due to this organism has been recognized among hospitalized or institutionalized patients. These patients often have a history of prior or concomitant antibiotic use, and an enzyme immunoassay for *C. perfringens* enterotoxin is commercially available for investigative use. Metronidazole treatment is recommended for those with protracted diarrhea. As with CDI, diarrhea may recur after successful treatment.

Despite the self-limited nature of most food-associated diarrheal syndromes, *C. perfringens* type A has also been responsible for outbreaks of fatal illness in institutionalized mentally ill patients. A recent outbreak in a state psychiatric hospital linked to improperly prepared chicken was notable for three deaths (7% case-fatality rate) in patients taking antimotility agents and in whom necrotizing colitis was found at autopsy.[19]

Grade A References

A1. Johnson S, Louie TJ, Gerding DN, et al. Vancomycin, metronidazole, or tolevamer for *Clostridium difficile* infection: results from two multinational, randomized, controlled trials. *Clin Infect Dis*. 2014;59:345-354.

A2. Louie TJ, Miller MA, Mullane KM, et al. Fidaxomicin versus vancomycin for *Clostridium difficile* infection. *N Engl J Med*. 2011;364:422-431.

A3. Cornely OA, Nathwani D, Ivanescu C, et al. Clinical efficacy of fidaxomicin compared with vancomycin and metronidazole in Clostridium difficile infections: a meta-analysis and indirect treatment comparison. *J Antimicrob Chemother*. 2014;69:2892-2900.

A4. van Nood E, Vrieze A, Nieuwdorp M, et al. Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *N Engl J Med*. 2013;368:407-415.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Centers for Disease Control and Prevention (CDC). Vital signs: preventing *Clostridium difficile* infections. *MMWR Morb Mortal Wkly Rep*. 2012;61:157-162.
2. Planche TD, Davies KA, Coen PG, et al. Differences in outcome according to *Clostridium difficile* testing method: a prospective multicentre diagnostic validation study of *C. difficile* infection. *Lancet Infect Dis*. 2013;13:936-945.
3. Dubberke ER, Carling P, Carrico R, et al. Strategies to prevent *Clostridium difficile* infections in acute care hospitals: 2014 update. *Infect Control Hosp Epidemiol*. 2014;35(suppl 2):S48-S65.
4. Cohen SH, Gerding DN, Johnson S, et al. Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA). *Infect Control Hosp Epidemiol*. 2010;31:431-455.
5. Neal MD, Alverdy JC, Hall DE, et al. Diverting loop ileostomy and colonic lavage: an alternative to total abdominal colectomy for the treatment of severe, complicated *Clostridium difficile* associated disease. *Ann Surg*. 2011;254:423-427.
6. Aronoff DM. *Clostridium novyi, sordellii*, and *tetani*: mechanisms of disease. *Anaerobe*. 2013;24:98-101.
7. Palmateer NE, Hope VD, Roy K, et al. Infections with spore-forming bacteria in persons who inject drugs, 2000-2009. *Emerg Infect Dis*. 2013;19:29-34.
8. Stevens DL, Aldape MJ, Bryant AE. Life-threatening clostridial infections. *Anaerobe*. 2012;18:254-259.
9. Barash JR, Arnon SS. A novel strain of *Clostridium botulinum* that produces type B and type H botulinum toxins. *J Infect Dis*. 2014;209:183-191.
10. Centers for Disease Control and Prevention (CDC). Notes from the field: botulism from drinking prison-made illicit alcohol—Arizona, 2012. *MMWR Morb Mortal Wkly Rep*. 2013;62:88.
11. Juliao PC, Maslanka S, Dykes J, et al. National outbreak of type A foodborne botulism associated with a widely distributed commercially canned hot dog chili sauce. *Clin Infect Dis*. 2013;56:376-382.
12. Fagan RP, McLaughlin JB, Castrodale LJ, et al. Endemic foodborne botulism among Alaska Native persons—Alaska, 1947-2007. *Clin Infect Dis*. 2011;52:585-592.
13. Centers for Disease Control and Prevention (CDC). Investigational heptavalent botulinum antitoxin (HBAT) to replace licensed botulinum antitoxin AB and investigational botulinum antitoxin E. *MMWR Morb Mortal Wkly Rep*. 2010;59:299.
14. Centers for Disease Control and Prevention (CDC). Tetanus surveillance—United States, 2001-2008. *MMWR Morb Mortal Wkly Rep*. 2011;60:365-369.
15. Centers for Disease Control and Prevention (CDC). Updated recommendations for use of tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis (Tdap) vaccine in adults aged 65 years and older—Advisory Committee on Immunization Practices (ACIP), 2012. *MMWR Morb Mortal Wkly Rep*. 2012;29:468-470.
16. Centers for Disease Control and Prevention (CDC). Updated recommendations for use of tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis vaccine (Tdap) in pregnant women—Advisory Committee on Immunization Practices (ACIP), 2012. *MMWR Morb Mortal Wkly Rep*. 2013;62:131-135.
17. Rodrigo C, Fernando D, Rajapakse S. Pharmacological management of tetanus: an evidence-based review. *Crit Care*. 2014;18:217.
18. Uzal FA, McClane BA. Recent progress in understanding the pathogenesis of *Clostridium perfringens* type C infections. *Vet Microbiol*. 2011;153:37-43.
19. Centers for Disease Control and Prevention (CDC). Fatal foodborne *Clostridium perfringens* illness at a state psychiatric hospital—Louisiana, 2010. *MMWR Morb Mortal Wkly Rep*. 2012;61:605-608.

REVIEW QUESTIONS

1. What is the preferred treatment for a patient with a first episode of mild to moderate *C. difficile* infection (CDI)?

A. Metronidazole 500 mg orally three times a day for 10 days
B. Vancomycin 125 mg orally four times a day for 10 days
C. Fidaxomicin 200 mg orally twice a day for 10 days
D. Fecal microbiota transplant
E. A or B

Answer: E This is an evolving area of treatment recommendation that will likely change with update of the CDI guidelines, so E is the current best answer. Both metronidazole and vancomycin orally are effective for mild to moderate CDI treatment, but the publication of the largest randomized controlled trial of metronidazole versus vancomycin demonstrated that metronidazole was significantly inferior to vancomycin ($P = .02$) for all patients with CDI, which is likely to change future recommendations to vancomycin. Fidaxomicin is also effective but is much more expensive than metronidazole or vancomycin. Fecal transplants are also effective, but experience with them has been limited to treatment of patients with multiple recurrences of CDI, for which it is effective in preventing additional recurrences.

2. What is the preferred treatment for patients with severe first-episode CDI?

A. Metronidazole 500 mg orally three times a day for 10 days
B. Vancomycin 125 mg orally four times a day for 10 days
C. Fidaxomicin 200 mg orally twice a day for 10 days
D. Both A and B used together
E. Metronidazole 500 mg IV three times a day plus vancomycin 500 mg orally four times a day for 10 days

Answer: B CDI guidelines recommend vancomycin for severe CDI.[3] Metronidazole was inferior to vancomycin for severe CDI in one randomized controlled trial and was inferior to vancomycin for all CDI in another larger randomized controlled trial. Fidaxomicin is also effective in treating severe CDI, but it is no better than vancomycin in eliciting cure; however, it is superior in reducing recurrence of CDI. Because of much higher pricing than for vancomycin, it is not the preferred treatment of first-episode CDI. Combined use of metronidazole and vancomycin has not been demonstrated to be superior to vancomycin alone for severe disease; however, use of IV metronidazole and higher dose oral vancomycin is recommended for severe complicated or fulminant CDI where there is significant risk of ileus that might prevent oral vancomycin from reaching the colon.

3. A California injection heroin user is "skin popping" because of exhausting all intravenous sites of injection. He has noted increased swelling and pain at an arm injection site during the past week and now also notes that his vision is blurred and his speech is slurred, and he has chills and feels weak and feverish. The most likely cause of his symptoms is

A. *Clostridium novyi* injection site abscess
B. *Clostridium tetani* injection site infection
C. *Clostridium botulinum* infection at the injection site
D. *Clostridium perfringens* injection site abscess
E. *Clostridium sordelli* injection site infection

Answer: C Most drug injection site infections are polymicrobial, and any of the listed clostridia could be present in addition to other anaerobes and aerobic bacteria, such as staphylococci and streptococci. However, of the clostridia listed, only *C. botulinum*, the cause of botulism, is associated with the symptoms of blurred vision, slurred speech, and weakness, a hallmark of early cranial nerve involvement and descending paralysis.

4. Which of the following statements about tetanus is *not* true?

A. It is the cause of hundreds of thousands of neonatal deaths globally because of inadequate maternal tetanus vaccination.
B. Although it is rare in the United States, most cases occur in patients inadequately vaccinated.
C. A short incubation period from injury to tetanus symptoms is indicative of a rapid resolution of symptoms and increased survival.
D. Immunization of childbearing women with Tdap (tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis) is recommended during the third trimester of each pregnancy for optimal fetal passive antibody protection, which is largely directed at preventing pertussis.
E. The mortality rate for generalized tetanus is 20 to 25%, even in modern medical facilities with extensive resources.

Answer: C This statement is incorrect. The shorter the period from injury to tetanus symptoms, the worse the prognosis is and the longer symptoms are likely to last. The other statements are all true.

5. Which clostridial organism is most likely responsible for the following syndrome? A 12-year-old boy with poorly controlled type 1 diabetes mellitus was admitted to the hospital with ketoacidosis and altered mental status 2 to 3 days after eating several large meals that included chitterlings. His white blood cell count was 11,300 mm^3, and his abdominal radiograph showed dilated loops of small bowel with gas in the bowel wall. His hospital course was notable for persistent hematemesis and increasing abdominal distention. Bloody ascites and necrotic jejunum and ileum were found at laparotomy.

A. *C. difficile*
B. *C. perfringens* type A
C. *C. perfringens* type C
D. *C. sordellii*
E. *C. botulinum*

Answer: C Although it is rare in developed countries, enteritis necroticans due to *C. perfringens* type C is responsible for this unique syndrome of small intestinal necrosis. Diabetes mellitus is a common risk factor noted in cases outside endemic areas, as in this case, which was reported in the southern United States in 2000. (Petrillo TM, Beck-Sagué CM, Songer JG, et al. Enteritis necroticans [pigbel] in a diabetic child. *N Engl J Med.* 2000;342:1250-1253.) *C. difficile* may cause severe intestinal disease, but disease is limited to the colon. *C. perfringens* type A rarely causes necrotizing intestinal syndromes, but again, disease is typically limited to the colon. *C. sordellii* and *C. botulinum* are usually not associated with acute abdominal syndromes.

297 DISEASES CAUSED BY NON–SPORE-FORMING ANAEROBIC BACTERIA

ITZHAK BROOK

DEFINITION

Anaerobic bacteria are the predominant members of the indigenous, normal human flora, including the skin and the oral, gastrointestinal, and vaginal mucosa (Fig. 297-1; Table 297-1). However, the types of predominant anaerobes are different at each location.

The Pathogens

Advances in taxonomics have led to reclassification of many anaerobic species (E-Table 297-1). The genus *Bacteroides* is used only for species of the *Bacteroides fragilis* group. The "oral" *Bacteroides* and "pigmented" *Bacteroides* species have been reclassified as *Prevotella* (saccharolytic, pigmented species), *Porphyromonas* (asaccharolytic species), and other genera. Capnophilic organisms (which require an elevated concentration of carbon dioxide for growth), sometimes referred to as microaerophiles, are not true anaerobes and are often more related to *Campylobacter*, *Capnocytophaga*, and other genera. In addition, many new genera and several new species have been created to accommodate pathogens such as *Bilophila wadsworthia*, *Sutterella wadsworthensis*, *Centipeda periodontii*, and *Anaerobiospirillum thomasii*. *Fusobacterium nucleatum* is the predominant *Fusobacterium* species isolated from clinical specimens.

EPIDEMIOLOGY

Anaerobes are opportunistic pathogens that can cause serious infections, generally in synergistic infections in combination with aerobic bacteria. Because the microbiology of these infections is often complex and culture results may be delayed, awareness of the normal bacterial flora at the site of infection is an indispensable guide for selection and institution of empirical antimicrobial therapy.

TABLE 297-1 LOCATION OF VARIOUS GROUPS OF ANAEROBES AS NORMAL MICROFLORA IN HUMANS

LOCATION	NO. OF ORGANISMS PER GRAM		PREDOMINANT ANAEROBIC BACTERIA
	Aerobes	*Anaerobes*	
Skin	—	—	*Propionibacterium acnes* *Peptostreptococcus* spp
Mouth/upper respiratory tract (in saliva)	10^8-10^9	10^9-10^{11}	Pigmented *Prevotella* and *Porphyromonas* spp *Fusobacterium* spp *Peptostreptococcus* spp *Actinomyces* spp
Gastrointestinal tract (in fecal material)			
Upper	10^2-10^5	10^3-10^7	*Bacteroides fragilis* group *Clostridium* spp
Lower	10^5-10^9	10^{10}-10^{12}	*Peptostreptococcus* spp *Bifidobacterium* spp *Eubacterium* spp
Female genital tract (in vaginal secretions)	10^8	10^9	*Peptostreptococcus* spp *Prevotella bivia* *Prevotella disiens*

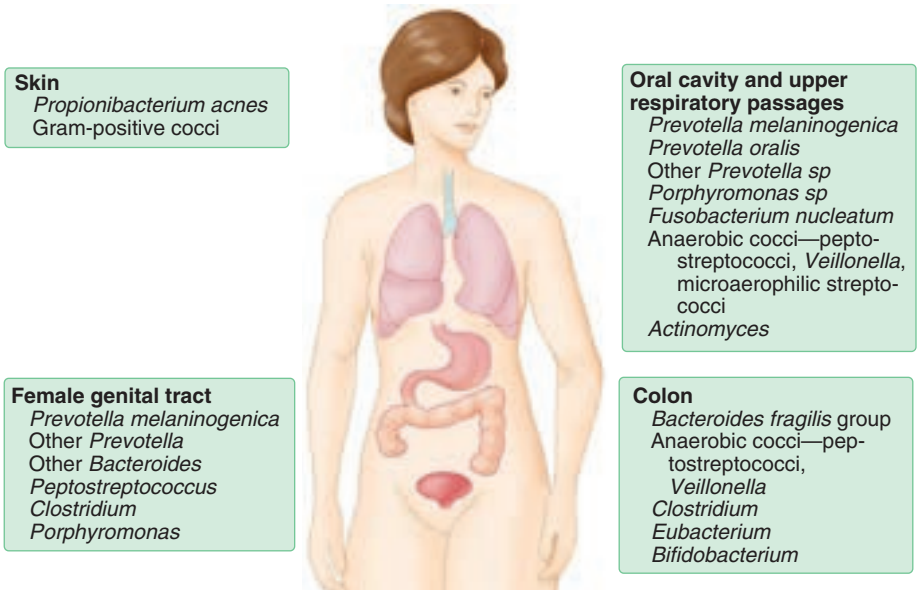

FIGURE 297-1. Anaerobes as the predominant normal microflora of the human body by general anatomic location. (Modified from Finegold SM, Sutter VL. *Diagnosis and Management of Anaerobic Infections.* Kalamazoo, MI: Upjohn; 1976. Copyright by Dr. Finegold.)

E-TABLE 297-1 TAXONOMY OF ANAEROBIC BACTERIA

CURRENT NAME	PREVIOUS NAME/COMMENT (YEAR CHANGED OR DESCRIBED)
Bacteroides fragilis	
Bacteroides caccae	
Parabacteroides distasonis	New genus (2006)
Parabacteroides merdae	New genus (2006)
Bacteroides nordii	New species (2005)
Bacteroides salyersiae	New species (2005)
Parabacteroides goldsteinii	New genus, species (2006)
Parabacteroides gordonii	New genus, species (2009)
Parabacteroides johnsonii	New genus, species (2007)
Bacteroides dorei	New species (2006)
Bacteroides massiliensis	New species, (2005)
Bacteroides eggerthii	
Bacteroides stercoris	
Bacteroides ovatus	
Bacteroides thetaiotaomicron	
Bacteroides vulgates	
Bacteroides uniformis	
Campylobacter gracilis	*Bacteroides gracilis*
Campylobacter ureolyticus	*Bacteroides ureolyticus*
Bacteroides pyogenes (2010)	*Bacteroides tectus, B. suis*
Odoribacter splanchnicus (2008)	*Bacteroides splanchnicus*
Alistipes finegoldii	New genus and species (2003)
Alistipes onderdonkii	(2006)
Alistipes putredinis	*Bacteroides putredinis* (2003)
Alistipes shahii	(2006)
Prevotella bivia	*Bacteroides bivius*
Prevotella buccae	*Bacteroides buccae (ruminicola)*
Prevotella dentalis	*Mitsuokella dentalis, Hallella sergens*
Prevotella disiens	*Bacteroides disiens*
Porphyromonas somerae	*Porphyromonas levi*
Prevotella nigrescens	(1992)
Prevotella intermedia	
Prevotella melaninogenica *Prevotella nanciensis*	*Bacteroides melaninogenicus* New species (2007)
Prevotella oralis	*Bacteroides oralis*
Prevotella oris	*Bacteroides oris (ruminicola)*
Alloprevotella tannerae	*Prevotella tannerae* (2013)
Prevotella timonensis	New species (2007)
Porphyromonas asaccharolytica	*Bacteroides asaccharolyticus*
Porphyromonas uenonis	New species (2005)
Tannerella forsythia	*Bacteroides forsythus*
Porphyromonas gingivalis	*Bacteroides gingivalis*
Porphyromonas macacae	*Bacteroides salivosus, Porphyromonas salivosa*
Fusobacterium canifelinum	New species (2004)
Fusobacterium necrophorum	
Fusobacterium ulcerans	(1988)
Anaerobiospirillum thomasii	(1997)
Bilophila wadsworthia	(1990)
Sutterella wadsworthensis	(1996)
Prevotella zoogleoformans	*Bacteroides zoogleoformans*

TABLE 297-2 POTENTIAL VIRULENCE FACTORS IN VARIOUS ANAEROBES

FACTOR	SPECIES
ADHESION	
Capsule	*Bacteroides fragilis* group, *Prevotella melaninogenica*
Pili/fimbriae	*B. fragilis* group
	Porphyromonas gingivalis
Hemagglutinin	*P. gingivalis*
Lectin	*Fusobacterium nucleatum*
INVASION/TISSUE DAMAGE	
Proteases	*Fusobacterium necrophorum*
	Bacteroides spp
	Porphyromonas spp
Hemolysins	Many species
Endotoxin	*B. fragilis*
Fibrinolysin	*B. fragilis* group
	Porphyromonas spp
Heparinase	*B. fragilis* group
	Porphyromonas spp
Neuraminidase	*B. fragilis* group
	Porphyromonas spp
ANTIPHAGOCYTIC	
Capsule	*B. fragilis* group
	P. gingivalis
Lipopolysaccharide	*B. fragilis* group
	F. necrophorum, P. gingivalis
Metabolic products	Most anaerobes
TOXINS	
Endotoxin	*B. fragilis*
	F. necrophorum
Enterotoxin	*B. fragilis*

Modified from Duerden BI. Virulence factors in anaerobes. *Clin Infect Dis.* 1994;18(Suppl 4):253.

PATHOBIOLOGY

Anaerobic bacteria range from those that cannot survive even a brief exposure to oxygen to those that can survive even in the presence of atmospheric oxygen (e.g., *B. fragilis*). Most anaerobes require an environment with a low oxidation-reduction potential (E_h gradient), which can be achieved in association with low pH, tissue destruction, byproducts from aerobic bacterial metabolism, or low oxygen content. Although they are not true anaerobes, some organisms, such as microaerophilic streptococci and other capnophilic or difficult-to-cultivate bacteria, are sometimes lumped together with anaerobes because of their fastidious nature. Some genera, such as *Lactobacillus* and *Actinomyces*, include both aerobic and anaerobic species.

Anaerobic bacteria possess a variety of virulence factors that are species specific (Table 297-2).

CLINICAL MANIFESTATIONS

Bacteremia

Transient anaerobic bacteremia occurs in about 85% of patients immediately after dental cleaning or manipulation. It is estimated that more than 200 cases of endocarditis from anaerobes are reported annually in the United States, usually in association with anatomic abnormalities or damaged cardiac valves (Chapter 76). Most anaerobic bacteremias are intermittent and associated with serious intra-abdominal or female genital tract, skin, and soft tissue infections, often proximal to the gastrointestinal tract. Which organisms are involved depends on their portal of entry and the underlying disease. The most common isolates are the *B. fragilis* group (60 to 75% of isolates). About 5 to 15% of bacteremias are caused by anaerobes, and they are the sole isolates in two thirds of these. Mortality associated with *B. fragilis* group bacteremia is 15 to 30%. Bacteremia with the *B. fragilis* group generally originates from a gastrointestinal source; with pigmented *Prevotella*, *Porphyromonas*, and *Fusobacterium* spp, from oropharyngeal and pulmonary sources; with *Fusobacterium* spp, from the female genital tract; and with *Propionibacterium acnes*, from foreign body sources. Bacteremia with peptostreptococci is associated with all sources but especially with the oropharyngeal, pulmonary, and female genital tracts.

Central Nervous System Infections

Anaerobes can cause brain abscess, subdural empyema, epidural abscess, and meningitis. The main source of brain abscess is an adjacent, generally chronic infection in the ears, mastoids, sinuses, oropharynx, teeth, or lungs. Rarely, bacteremia of another origin or endocarditis can cause such infection.

Meningitis caused by anaerobes is uncommon and can follow respiratory or dental infection or develop as a complication of a cerebrospinal fluid shunt. The isolates usually isolated from brain abscesses that complicate respiratory and dental infections include *Prevotella*, *Porphyromonas*, *Bacteroides*, *Fusobacterium*, and *Peptostreptococcus* spp. Microaerophilic and other streptococci are also often isolated. *Propionibacterium acnes* is common in shunt infections.

Head and Neck

Dental infections (Chapter 425) associated with a variety of oral anaerobes include periodontal disease, gingivitis, pulpitis, acute necrotizing ulcerative gingivitis, localized juvenile periodontitis, adult periodontitis, pericoronitis, endodontitis, periapical and dental abscesses, and postextraction infection.[1] Peritonsillar, retropharyngeal, and parapharyngeal abscesses (Chapter 429) are deep-seated, potentially life-threatening infections that may spread into the various potential spaces of the neck or mediastinum and cause jugular vein thrombosis. Oral anaerobes can be recovered in more than 50% of such cases, usually mixed with aerobes. Other regional infections include cervicofacial actinomycosis (Chapter 329), Ludwig's angina, *Fusobacterium necrophorum* sepsis with metastatic infection (Lemierre's syndrome), suppurative sialoadenitis (including parotitis), neck space infections, thyroiditis and chronic sinusitis (Chapter 426), otitis media (Chapter 426), and mastoiditis. Management involves surgical drainage and appropriate antimicrobial therapy.

Pleuropulmonary

Anaerobes predominate in oral and upper respiratory tract normal flora, and most aspiration pneumonias are due to this flora (Chapter 97).[2] Aspiration can result from altered consciousness, dysphagia, or mechanical devices such as intubation equipment. Poor oral hygiene is associated with an increased anaerobic bacterial burden, and the presence of aerobes or necrotic tissue lowers the pH, which facilitates the growth of anaerobes. Anaerobes are involved in 90% of community-acquired aspiration pneumonia and in about a third of nosocomial aspiration pneumonia, empyema, lung abscess, and pneumonia associated with tracheostomy. If the anaerobic component of aspiration pneumonia is not treated, the anaerobes can cause a lung abscess. Management requires good pulmonary toilet and antimicrobial therapy.

Intra-abdominal

Because anaerobes outnumber aerobes by 1000 to 1 in the large intestine, they play a major role in almost all intra-abdominal infections. Most visceral abscesses (e.g., hepatic; Chapter 151), chronic cholecystitis (Chapter 155), perforated and gangrenous appendicitis (Chapter 142), postoperative wound infections and abscesses, diverticulitis (Chapter 142), and any infection associated with fecal contamination of the abdominal cavity involve both aerobes and anaerobes. *B. fragilis* group members predominate because they are encapsulated, resist phagocytosis, are often resistant to many antimicrobials, and promote abscess formation. They may also be associated with concomitant bacteremia and sepsis. Randomized controlled trials have found that prophylactic antibiotics covering both anaerobic and aerobic bacteria administered orally and/or intravenously prior to elective colorectal surgery reduce the risk of surgical wound infection by as much as 75%.

Obstetric-Gynecologic

A variety of obstetric-gynecologic infections involve anaerobes. These are polymicrobial and include bacterial vaginosis; soft tissue perineal, vulvar, and Bartholin gland abscesses; endometritis; pyometra; salpingitis; tubo-ovarian abscesses; adnexal abscess; pelvic inflammatory disease, which may include pelvic cellulitis and abscess; chorioamnionitis; vaginal cuff cellulitis; septic pelvic thrombophlebitis; intrauterine contraceptive device–associated infection; septic abortion; and postsurgical obstetric and gynecologic infections. Bacterial vaginosis has been associated with preterm labor or delivery, chorioamnionitis, low birthweight, postpartum endometritis, and postabortal pelvic inflammatory disease. Bacterial vaginosis can increase the risk for infection with human immunodeficiency virus type 1 and the development of other sexually transmitted diseases (Chapter 285).

Skin and Soft Tissue

Cutaneous infections include infected ulcers, cellulitis (including synergistic necrotizing cellulitis), pyoderma, paronychia, hidradenitis suppurativa, and a variety of secondarily infected sites. Such sites include secondarily infected gastrostomy or tracheostomy site wounds, subcutaneous sebaceous or inclu-

sion cysts, eczema, psoriasis, poison ivy, atopic dermatitis, eczema herpeticum, scabies or kerion, and postsurgical wounds.

Subcutaneous infections include abscesses, decubitus ulcers, infected diabetic (vascular or trophic) ulcers, human and animal bite wounds, anaerobic cellulitis and gas gangrene, bacterial synergistic gangrene, Fournier's gangrene, infected pilonidal cyst or sinus, and burn wounds. Anaerobic soft tissue infections that occur deeper are necrotizing fasciitis, necrotizing synergistic cellulitis, and gas gangrene. These infections can involve the fascia and can induce myositis and myonecrosis.

Cultures frequently yield isolates that are members of the normal flora of the region of the infection. In addition to oral and skin flora, human bite infections often contain *Eikenella* species, and animal bites harbor *Pasteurella multocida*.

The infections are generally polymicrobial, and some (e.g., decubitus ulcers, diabetic foot ulcers) are often complicated by osteomyelitis or bacteremia. Deep tissue infections, such as necrotizing cellulitis, fasciitis, and myositis, often involve *Clostridium* species, *Streptococcus pyogenes*, or a polymicrobic aerobic and anaerobic flora. They are often associated with gas in the tissues and putrid-like pus with a gray, thin quality and have a high rate of bacteremia and mortality. Management of deep-seated soft tissue infection includes surgical débridement, drainage, and vigorous surgical management.

Osteomyelitis and Septic Arthritis

Anaerobes can be involved in osteomyelitis of the long bones after trauma and fracture, osteomyelitis related to peripheral vascular disease, decubitus ulcers, and osteomyelitis of the cranial and facial bones. Most of these infections are polymicrobial.

Cranial and facial bone osteomyelitis is generally caused by spread from a contiguous soft tissue source or from sinus, ear, or dental infection. Intestinal anaerobes originating from decubitus ulcers are involved in pelvic osteomyelitis. Osteomyelitis of long bones and septic arthritis are generally caused by hematogenous spread, trauma, or the presence of a prosthetic device.

The most commonly recovered anaerobes are peptostreptococci and *P. acnes* (often in prosthetic joint infection), *B. fragilis* group and fusobacteria (often of hematogenous origin), and clostridia (associated with trauma).

DIAGNOSIS

Anaerobic infections should be suspected in a number of specific clinical scenarios (Table 297-3). An appropriately collected microbiologic specimen (Table 297-4) is critical for accurate diagnosis.

TREATMENT

Rx

General principles of treatment (Table 297-5) include appropriate antimicrobial therapy coupled with prompt drainage, decompression of closed space infections, relief of obstructions, and surgical débridement.[3,4] The various clinically important anaerobes can be characterized by reasonably predictable antimicrobial susceptibility patterns (Table 297-6). However, some anaerobes have become resistant to antimicrobials, and many can develop resistance during therapy.[5] Reliable culture and sensitivity results should ultimately guide therapy. The efficacy of hyperbaric oxygen is unproved, but its use in conjunction with other therapeutic measures is not contraindicated.

In choosing antimicrobials for the treatment of mixed infections, their aerobic and anaerobic antibacterial spectra and their availability in oral or parenteral form should be considered. Some antimicrobials have a limited range of activity.[6] For example, metronidazole is active only against anaerobes and therefore cannot be administered as a single agent for the treatment of mixed infections. Others (i.e., carbapenems, a penicillin plus a β-lactamase inhibitor) have wide spectra of activity against aerobes and anaerobes.

Aside from susceptibility patterns, other factors influencing the choice of antimicrobial therapy include the pharmacologic characteristics of the various drugs, their toxicity, their effect on normal flora, and their bactericidal activity. Although identification of the infecting organisms and their antimicrobial susceptibility may be needed for selection of optimal therapy, the clinical setting and Gram stain preparation of the specimen may suggest the types of anaerobes present in the infection and the nature of the infectious process.

Even though the length of therapy for anaerobic infections is generally longer than that for aerobic and facultative infections, the length of treatment must be individualized, depending on the response. In some cases, treatment may require 6 to 8 weeks, but therapy may be shortened with proper surgical drainage. An anti–gram-negative enteric agent is generally added to treat Enterobacteriaceae in managing intra-abdominal infections.

The available parenteral antimicrobials for most infections are metronidazole, chloramphenicol, clindamycin, cefoxitin, a penicillin (i.e., ticarcillin, ampicillin, piperacillin) and a β-lactamase inhibitor (i.e., clavulanic acid, sulbactam, tazobactam), a carbapenem (i.e., imipenem, meropenem, doripenem, ertapenem), and tigecycline.

An agent effective against gram-negative enteric bacilli (e.g., an aminoglycoside, fluoroquinolone) or an antipseudomonal cephalosporin (e.g., cefepime) is generally added to metronidazole and, occasionally, cefoxitin in treatment of intra-abdominal infections. Penicillin can be added to metronidazole for the treatment of intracranial, pulmonary, or dental infections to cover microaerophilic streptococci and *Actinomyces* species. Penicillin is added to clindamycin to supplement its coverage against *Peptostreptococcus* species and other gram-positive anaerobic organisms. For *Chlamydia* and

TABLE 297-3 CLINICAL INDICATORS OF ANAEROBIC INFECTION

Infection adjacent to a mucosal surface
Foul-smelling discharge
Necrotic gangrenous tissue and abscess formation
Free gas or crepitus in tissue
Bacteremia or endocarditis with no growth on aerobic blood cultures
Infection related to the use of antibiotics effective against aerobes only (e.g., trimethoprim-sulfamethoxazole, aminoglycosides, older quinolones)
Infection related to tumors or other destructive processes
Infected thrombophlebitis
Infection after bites
Black discoloration of exudates containing *Prevotella melaninogenica*, which may fluoresce under ultraviolet light
"Sulfur granules" in discharges caused by actinomycosis
Clinical finding of gas gangrene or necrotizing fasciitis
Clinical condition predisposing to anaerobic infection (e.g., after maternal amnionitis, fistulous tracks, bites, dental infection, bowel perforation)

TABLE 297-4 SPECIMEN ACCEPTABILITY FOR ANAEROBIC CULTURE

SPECIMENS THAT SHOULD NOT BE CULTURED FOR ANAEROBES
Feces or rectal swabs
Throat or nasopharyngeal swabs
Sputum or bronchoscopic specimens
Routine or catheterized urine
Vaginal or cervical swabs
Material from superficial wounds or abscesses not collected properly to exclude surface contamination
Material from abdominal wounds obviously contaminated with feces, such as an open fistula
SPECIMENS APPROPRIATE FOR ANAEROBIC CULTURE
All normally sterile body fluids other than urine, such as blood, pleural fluid, and joint fluid
Urine obtained by suprapubic bladder aspiration
Percutaneous transtracheal aspiration, direct lung puncture, or double-lumen catheter bronchial brushing and bronchoalveolar lavage (both cultured quantitatively)
Culdocentesis fluid obtained after decontamination of the vagina
Material obtained from closed abscesses
Material obtained from sinus tracks or draining wounds

TABLE 297-5 GENERAL PRINCIPLES OF THERAPY FOR ANAEROBIC INFECTIONS

Decompression of closed spaces
Débridement
Drainage
Relief of obstructions
Irrigation
Provision of adequate circulation when possible
Removal of foreign bodies
Antimicrobials
Activity against most likely pathogen or pathogens: location dependent, minimal effect on normal flora
Absorption, appropriate route of administration (intravenous, oral)
Penetration into site of infection
Dosage appropriate for local tissue levels, body mass of patient, renal and liver function
Duration appropriate for condition
Susceptibility testing of isolate to guide specific therapy

TABLE 297-6 ANTIMICROBIAL SUSCEPTIBILITY PATTERNS FOR ANAEROBIC BACTERIA*

BACTERIA	PENICILLIN	β-LACTAMASE†	CEFOXITIN	CEFOTETAN	CARBAPENEMS, TIGECYCLINE	MOXIFLOXACIN	CLINDAMYCIN	METRONIDAZOLE
Bacteroides fragilis	–	+	+	+	+	+	V	+
Bacteroides thetaiotaomicron	–	+	V	V	+	V	V	+
B. fragilis group, other	–	+	V	V	+	+	V	+
Prevotella spp	V	+	+	+	+	+	+	+
Fusobacterium nucleatum	V	+	+	+	+	V	+	+
Fusobacterium necrophorum	+	+	+	+	+	V	+	+
Porphyromonas spp	+	+	+	+	+	+	+	+
Peptostreptococcus	+	+	+	+	+	+	+	V
Propionibacterium acnes	+	+	+	+	+	+	+	–
Veillonella	+	+	+	+	+	+	+	+
Actinomyces	+	+	+	+	+	+	+	–

*Based on a variety of in vitro susceptibility studies from different laboratories and using different techniques.
†β-Lactamase inhibitor–β-lactam combination (e.g., ticarcillin-clavulanate, ampicillin-sulbactam, piperacillin-tazobactam).
+ = Susceptible; – = resistant; V = variable.

Mycoplasma species, doxycycline is added to most regimens in treatment of pelvic infections. Oral therapy is often substituted for parenteral therapy. The agents available for oral therapy are clindamycin, amoxicillin and clavulanate, and metronidazole.

Grade A Reference

A1. Nelson RL, Gladman E, Barbateskovic M. Antimicrobial prophylaxis for colorectal surgery. *Cochrane Database Syst Rev*. 2014;5:CD001181.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Brook I. Anaerobic bacteria in upper respiratory tract and head and neck infections: microbiology and treatment. *Anaerobe*. 2012;18:214-220.
2. Bartlett JG. How important are anaerobic bacteria in aspiration pneumonia: when should they be treated and what is optimal therapy. *Infect Dis Clin North Am*. 2013;27:149-155.
3. Solomkin JS, Mazuski JE, Bradley JS, et al. Diagnosis and management of complicated intra-abdominal infections in adults and children: guidelines by the Surgical Infection Society and the Infectious Diseases Society of America. *Surg Infect (Larchmt)*. 2010;11:79-109.
4. Snydman DR, Jacobus NV, McDermott LA, et al. Lessons learned from the anaerobe survey: historical perspective and review of the most recent data (2005-2007). *Clin Infect Dis*. 2010;50(suppl 1): S26-S33.
5. Boyanova L, Kolarov R, Mitov I. Recent evolution of antibiotic resistance in the anaerobes as compared to previous decades. *Anaerobic*. 2015;[Epub ahead of print].
6. Brook I. Antimicrobial treatment of anaerobic infections. *Expert Opin Pharmacother*. 2011;12: 1691-1707.

REVIEW QUESTIONS

1. What culture is appropriate for isolation of anaerobic bacteria?

A. Vaginal or cervical swabs
B. Routine or catheterized urine
C. Throat or nasopharyngeal swabs
D. Aspirate of abscess
E. Feces or rectal swabs

Answer: D Obtaining an aspirate after sterilizing the skin or bypassing the normal flora avoids contamination of the specimen by normal skin or mucous membrane bacterial flora.

2. *Bacteroides fragilis* bacteremia is usually associated with

A. Trauma
B. Aspiration
C. Rupture of abdominal viscus
D. Malignant disease
E. Osteomyelitis

Answer: C Although *B. fragilis* group is the most common species found in clinical specimens, it is the least common *Bacteroides* present in fecal flora, representing only 0.5% of the bacteria present in stool. The pathogenicity of this group of organisms probably results from its ability to produce capsular material, which is protective against phagocytosis. Because of its presence in normal flora of the gastrointestinal tract, this organism is predominant in bacteremia associated with intra-abdominal infections, peritonitis, and abscesses following rupture of a viscus.

3. Which of the following antimicrobials is not effective against anaerobic bacteria?

A. Amoxicillin
B. Metronidazole
C. Gentamicin
D. Moxifloxacin
E. Cefoxitin

Answer: C Aminoglycosides, including gentamicin, are not effective against anaerobic bacteria because of their inability to reach the ribosome. Gentamicin is effective only in aerobic microorganisms because it enters cells through an electron transport–linked system that requires oxygen.

4. Which anaerobic organism cannot produce β-lactamase?

A. *Fusobacterium nucleatum*
B. *Bacteroides fragilis*
C. *Prevotella melaninogenica*
D. *Peptostreptococcus* spp
E. *Bacteroides thetaiotaomicron*

Answer: D Most *B. fragilis* group, up to 20 to 50% of *Prevotella* spp, and *Fusobacterium nucleatum* isolates can produce β-lactamase.

5. *Propionibacterium acnes* is associated with

A. Chronic sinusitis
B. Splenic abscess
C. Cellulitis
D. Abdominal trauma
E. Central nervous system shunt infection

Answer: E *Propionibacterium* species are part of the normal bacterial flora that colonizes the skin, conjunctiva, oropharynx, and gastrointestinal tract. These non–spore-forming, anaerobic, gram-positive bacilli are frequent contaminants of specimens of blood and other sterile body fluids and have been generally considered to play little or no pathogenic role in humans. *Propionibacterium acnes* has, however, been recovered in specimens obtained from patients with infections associated with a foreign body (such as an artificial valve), endocarditis, and central nervous system shunt infections.

298

NEISSERIA MENINGITIDIS INFECTIONS

DAVID S. STEPHENS

DEFINITION

Neisseria meningitidis (the meningococcus) is the cause of epidemic bacterial meningitis, fulminant sepsis (meningococcemia), milder bacteremia, and, less commonly, focal infections (such as pneumonia, septic arthritis, purulent pericarditis, and conjunctivitis).[1-3]

The Pathogen

N. meningitidis is an aerobic, diplococcal gram-negative β-proteobacterium and a member of the family Neisseriaceae, which also includes *Neisseria gonorrhoeae* (Chapter 299), another important human pathogen. The meningococcus is a commensal of the human upper respiratory tract, but it can also cause local and devastating invasive human disease. Human mucosal surfaces, most commonly the nasopharynx, are the only known reservoir. There are 12 confirmed serogroups of *N. meningitidis*, based on different capsular polysaccharide structures, but only 6 serogroups (A, B, C, W, X, and Y) cause almost all invasive meningococcal disease (Fig. 298-1). Highly pathogenic meningococci are also distinguished by genetically defined clonal complexes that can emerge and spread worldwide.[4] Dissecting the basis of meningococcal disease has provided important scientific lessons for bacterial emergence and pathogenesis, antibiotic resistance mechanisms, innate and adaptive human immune responses, and vaccine development.

EPIDEMIOLOGY

For at least 200 years, *N. meningitidis* has inflicted rapid death, disability, and fear on disparate human populations. Beginning with the initial descriptions of outbreaks in Geneva in 1805 and New Bedford, Massachusetts, in 1806, the meningococcus has caused endemic disease, case clusters, epidemics and pandemics of meningitis and septicemia, and less commonly pneumonia. An estimated 500,000 cases occur worldwide each year. The greatest burden of disease occurs in Africa and in parts of Asia, where endemic rates of disease have been 3 to 10 per 100,000 population. In sub-Saharan Africa, seasonal increases in disease and cyclic pandemics occurred every 8 to 10 years since 1905. During epidemics and cyclic pandemics, the incidence can climb to 1 per 1000 population for weeks before the frequency of disease declines in the immediate outbreak area. For example, until the introduction in 2010 of a new meningococcal conjugate vaccine for serogroup A, the dry seasons in Burkina Faso, located in the sub-Saharan meningitis belt, were accompanied by more than 5000 cases of meningitis per week (>680/100,000 population), an almost yearly occurrence in the country. Meningococcal epidemics in developing countries have been catastrophic and contribute to a cycle of poverty and hence the disorganization of social structures.

Meningococcal disease remains endemic, with focal outbreaks/clusters, in the United States, Canada, Europe, Japan, Australia, and other industrialized countries, with a lower overall incidence now at 0.1 to 2 per 100,000 population. The introduction and widespread use of new meningococcal conjugate vaccines in North America, Australia, and Europe have helped to lower the incidence,[5,6] but slow declines in industrialized countries began before new vaccine introductions. Endemic disease and outbreaks also occur in China, eastern Europe, Russia, South America, India, and Southeast Asia. Before and during World War II, large epidemic outbreaks (mainly due to serogroup A) affected the United States, Europe, Japan, and Australia; after the war, these outbreaks disappeared, for reasons that are not understood. Meningococcal disease has the highest incidence in children younger than 4 years and in adolescents, but in endemic settings, half of all cases occur in adults.

PATHOBIOLOGY

N. meningitidis is transmitted among humans through close contact by large respiratory droplets. Colonization of the upper respiratory mucosal surfaces (e.g., nasopharynx) by *N. meningitidis* is the first step in establishment of a human carrier state and invasive meningococcal disease. Acquisition of meningococci through contact with respiratory secretions or saliva can be transient, lead to colonization (carriage), or result in invasive disease. The inoculum size needed for transmission is unknown. Meningococcal disease usually occurs within 1 to 14 days of acquisition. Meningococci can be found in the urogenital tract and rectum and may be transmitted sexually.

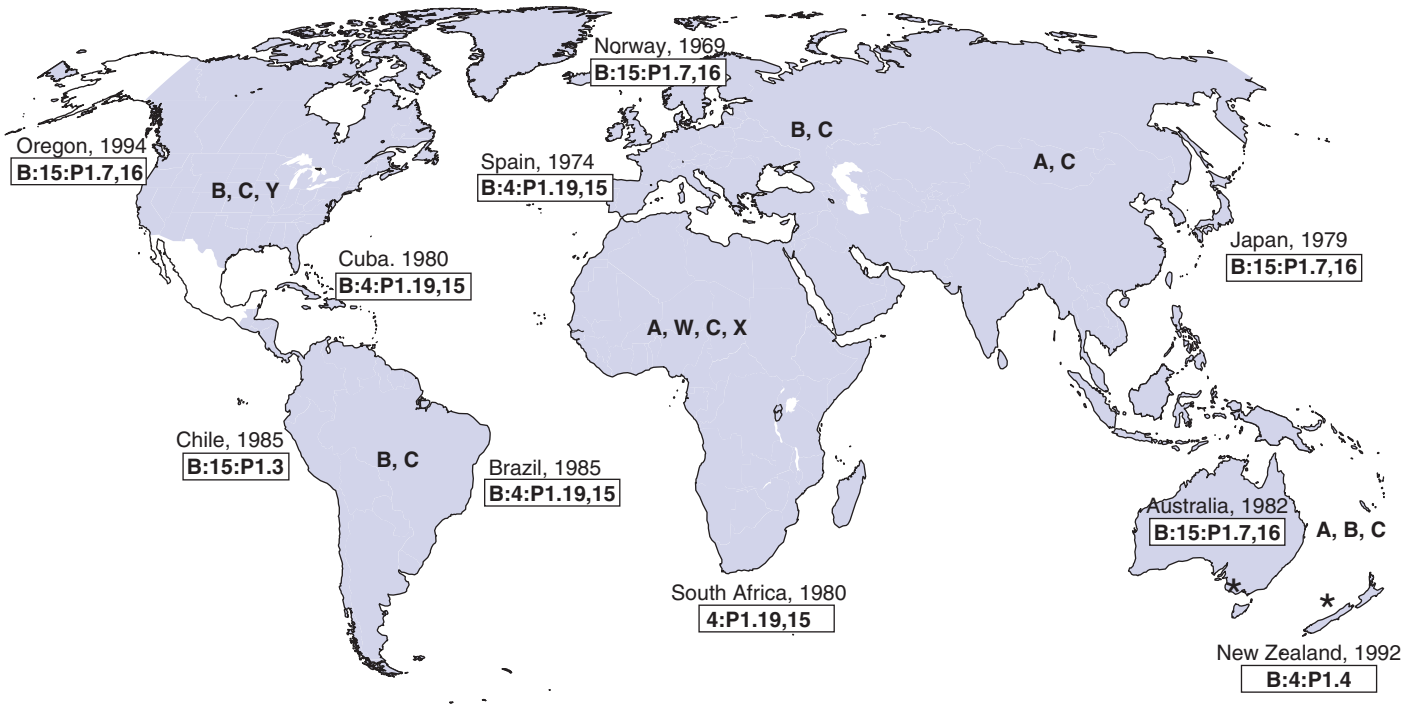

*Group B by serotype, modified from Caugant, APMIS 1998; 106:505-525

FIGURE 298-1. Epidemiology of meningococcal disease (major serogroups causing disease by region and serogroup B epidemics). (Modified from Stephens DS, Greenwood B, Brandtzaeg P. Epidemic meningitis, meningococcemia and *Neisseria meningitidis*. *Lancet*. 2007;369:2196-2210.)

Initial contact of meningococci with mucosal epithelial cells is mediated by type IV pili. These structures provide mobility ("twitching motility") to penetrate mucus and are the initial adhesins for human epithelial cells. Meningococci proceed to proliferate on the surface of human nonciliated epithelial cells, forming small microcolonies at the site of attachment; they can disseminate from colonies by post-translational glycan modifications of pili and migrate to adjacent cells by the pili-mediated motility.[7,8] Meningococci can also spread from the nasopharynx to adjacent epithelial surfaces and can infrequently cause local infections, including pneumonia, sinusitis, and otitis media. Other less common local infections include conjunctivitis, urethritis, and proctitis. Close adherence of meningococci to the host epithelial cells results in the formation of epithelial cell cortical plaques and leads to the recruitment of factors ultimately responsible for the formation and extension of host epithelial cell pseudopodia that can surround the meningococcus. Intimate meningococcal association with the epithelial cell is mediated by the bacterial opacity proteins Opa and Opc with CD66/carcinoembryonic antigen–related cell adhesion molecules and integrins, respectively, on the surface of the cell. However, other meningococcal epithelial cell mediators include the meningococcal adhesin NadA and meningococcal lipo-oligosaccharide. The formation of epithelial cell membrane protrusions and pseudopodia stems from the organization of specific molecular complexes involving the linkers ezrin and moesin along with the clustering of several membrane-integral proteins, including CD44, intracellular adhesion molecule 1, and cortical actin polymerization.[9] These events can lead to internalization of *N. meningitidis* in epithelial cells (Fig. 298-2). Intracellular meningococci reside within a membranous vacuole and are capable of translocating through the epithelial layers within 18 to 40 hours. Meningococci are capable of intracellular replication (in part because of the protective capsule), can survive under microaerophilic conditions, use lactate as a carbon source, and have the capacity to acquire iron through specialized transport systems.

Meningococci cross mucosal surfaces, enter the blood stream, and, in some individuals, produce systemic infections. Damage to the mucosal surface by coinfection, drying (e.g., very low humidity), or smoke exposure may increase this risk of meningococcal invasion. Similar molecular interactions noted for meningococci and epithelial cells also occur with endothelial cells, and meningococci can translocate across the blood-meninges barrier, possibly at the choroid plexus or by opening of intercellular junctions, and proliferate in the subarachnoid space, resulting in meningitis. In the vasculature and cerebrospinal fluid (CSF), high levels of multiplying bacteria lead to an intense inflammatory response, with pronounced increases in concentrations of tumor necrosis factor-α, interleukins (1β, 6, 8, and 10), different chemokines, and other inflammatory mediators.[10]

Resistance to complement-mediated lysis or phagocytosis is due to the expression of the capsule, lipo-oligosaccharide, and several surface-exposed proteins (factor H–binding protein, NspA, Opc, NalP). Meningococcal endotoxin released in blebs plays a major role in the inflammatory events of meningococcemia and meningococcal meningitis. Meningococcal lipid A is responsible for much of the biologic activity and toxicity of meningococcal endotoxin. The toll-like receptor 4 (TLR4) is critical to the innate immune response to bacterial endotoxins, and meningococcal endotoxin is no exception. Activation of TLR4 by endotoxin requires association with the accessory protein MD-2, an *N*-glycosylated 19- to 27-kD protein expressed in both a soluble and a membrane-bound form. Binding of endotoxin to MD-2 in association with TLR4 leads to dimerization or oligomerization of two or more TLR4s, subsequent cellular activation, and cytokine and chemokine release.

In contrast to invasive disease, an asymptomatic *N. meningitidis* carrier state is found in up to 8 to 25% of healthy individuals. Meningococcal carriage is affected by age, intimate personal contact, crowding (e.g., bars, dormitories), and vaccination or chemoprophylaxis interventions in the community. Variable carriage rates have been reported, even during epidemics. Meningococcal carriage is a dynamic process, is less common in young children (<3% and *Neisseria lactamica* predominates) than in older children, is highest in adolescents (7 to 37%), and increases in closed populations (e.g., military recruits, hajj pilgrims). Rates as high as 36 to 71% have been reported in military recruits. Damage to the upper respiratory tract by coinfections (e.g., mycoplasma, influenza, other respiratory viral infections), smoking, very low humidity, drying of mucosal surfaces, and trauma induced by dust predisposes to both meningococcal carriage and meningococcal disease. Meningococcal carriage has also been linked to status as a secretor of glycoprotein ABO blood group antigens, which are water soluble, and to ethnic background. In a large U.K. study, social behavior (e.g., attendance at pubs or clubs, intimate kissing, cigarette smoking or exposure to passive smoke) was highly associated with the risk of meningococcal carriage. Carriage can be transient or last for days, weeks, or months and is an immunizing event leading to protective immunity (e.g., serum bactericidal activity against the meningococcus).

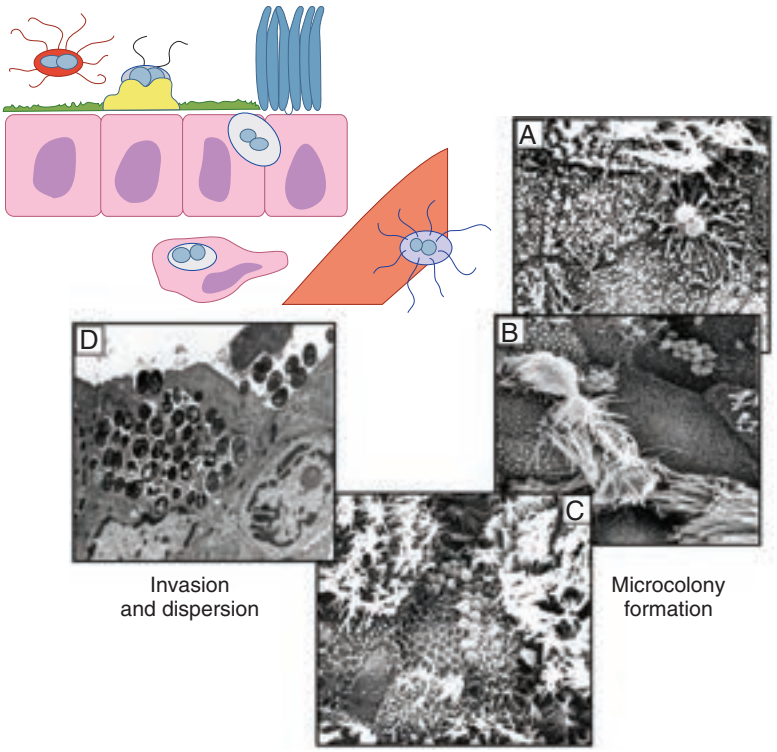

FIGURE 298-2. Steps in initiation of meningococcal colonization and invasion at the human nasopharynx. A, Adhesion and introduction of cell microvilli. **B,** Microcolony formation. **C,** Cortical plaque formation and close adherence. **D,** Human epithelial cell invasion. (Modified from Stephens DS. Biology and pathogenesis of the evolutionarily successful, obligate human bacterium *Neisseria meningitidis. Vaccine.* 2009;27S:B71-B77.)

The absence of protective bactericidal antibodies is the most important predisposing factor for systemic meningococcal disease, but complement deficiencies, genetic polymorphisms, and other host cofactors can contribute to meningococcal disease and disease severity. Disappearance of protective maternal antibodies increases the risk in older infants and young children. Congenital and acquired antibody deficiencies also increase risk. Opsonization and phagocytic function do contribute to meningococcal host defense mechanisms, as shown by disease reduction after polysaccharide vaccination in individuals with complement deficiencies. Rapidly progressive, fatal meningococcemia can arise in patients without properdin, and there is a marked risk of recurrent meningococcal infections in those with defects in the terminal complement pathway (C5-C9) and C3 deficiency.[11]

Polymorphisms in genes coding for the Fcγ-receptor II (CD32), Fcγ-receptor III (CD16), mannose-binding lectin, TLR4, and β_2-adrenoceptor gene have been associated with increased risk. Mannose-binding lectin is a plasma opsonin that initiates complement activation; specific polymorphisms in the gene are identified more frequently in children with meningococcal disease than in controls in some studies. Plasminogen activator inhibitor 1 concentrations appear to affect the severity and mortality of meningococcal sepsis, suggesting that impaired fibrinolysis is an important factor in its pathophysiology. Meningococcal disease is occasionally linked to immunosuppressive disorders, such as nephrotic syndrome, congenital or acquired hypogammaglobulinemia, splenectomy, and human immunodeficiency virus (HIV) infection/AIDS (about a 10-fold increased risk for sporadic disease). However, there has been no documented increase in epidemic outbreaks of meningococcal disease in countries with very high rates of HIV infection.

Meningococci can multiply rapidly in the vascular compartment, with an estimated doubling time of 30 to 45 minutes in some patients, or in the CSF.[12] The release of high levels of inflammatory mediators such as meningococcal endotoxin in the circulation or CSF triggers an exaggerated release of chemokines, cytokines, bradykinin, and nitric oxide. Vascular dilation, hypovolemia, capillary leak, and pronounced reduction in myocardial function are the result. At a later stage, substantial complement activation contributes to the altered endothelial barrier function and relaxation of the smooth muscles in the vessel wall through the generation of high levels of anaphylatoxins (C3a and C5a). The capillary leak syndrome results in an increased flux of albumin and water across the altered capillary wall to the extravascular space. A patient with fulminant meningococcemia accumulates a large amount of fluid in the extravascular tissue. Circulatory collapse and multiorgan dysfunction are the primary causes of death due to meningococcemia. In meningitis, morbidity and death are due predominantly to cerebral edema.

Unraveling the pathogenic mechanisms of this devastating, evolutionarily successful obligate human pathogen has significance for the understanding of human sepsis as well as for prevention through vaccines directed at mucosal pathogens.

CLINICAL MANIFESTATIONS

The meningococcus causes meningitis (37 to 50% of cases), septicemia (meningococcemia, 10 to 18% of cases), or both in 7 to 12% of cases. The presentations are less commonly a mild bacteremia or pneumonia (10% of cases) and much less commonly (<5% of cases) septic arthritis, pericarditis, chronic bacteremia, or conjunctivitis . Very rarely, meningococci can cause urethritis or proctitis. In endemic and epidemic disease outbreaks, hemorrhagic skin lesions (petechiae, purpura; Fig. 298-3) are present in 28 to 77% of patients with invasive meningococcal disease on admission, but these lesions may be absent or difficult to see in patients with dark skin. Hemorrhagic lesions sometimes occur on mucous membranes and sclera, but they are especially prevalent on the limbs. Petechiae of meningococcemia are usually larger and bluer than the pinpoint petechiae caused by thrombocytopenia, leukocytoclastic vasculitis induced by other infections or drugs, or vomiting or coughing. A nonblanching macular rash can also be a manifestation of meningococcal bacteremia. Evolving ecchymoses and purpura (diameter >10 mm) are noted mainly in patients with meningococcemia and disseminated intravascular coagulation (Chapter 175), but they may not appear until 12 hours into the illness. In addition to vasculitis, other conditions in the differential diagnosis of meningococcemia include Rocky Mountain spotted fever and enteroviral infections.

Meningitis is the most common clinical presentation of invasive meningococcal disease.[13] Headache, fever, and rash with meningismus and altered mental status are the characteristic features; however, the rash may be absent, and the presentation can resemble pneumococcal or bacterial meningitis of other causes, viral meningitis, or early-stage encephalitis. Bacteremic meningococcal pneumonia has been linked most often to serogroup Y and is more common in adolescents and adults, especially older adults (approximately one third of cases occur in those older than 65 years). Isolated septic pericarditis or septic arthritis can also be a presentation, and an autoimmune- or antibody-mediated polyarthritis can be seen in the recovery phase following invasive meningococcal disease. Chronic meningococcemia can be

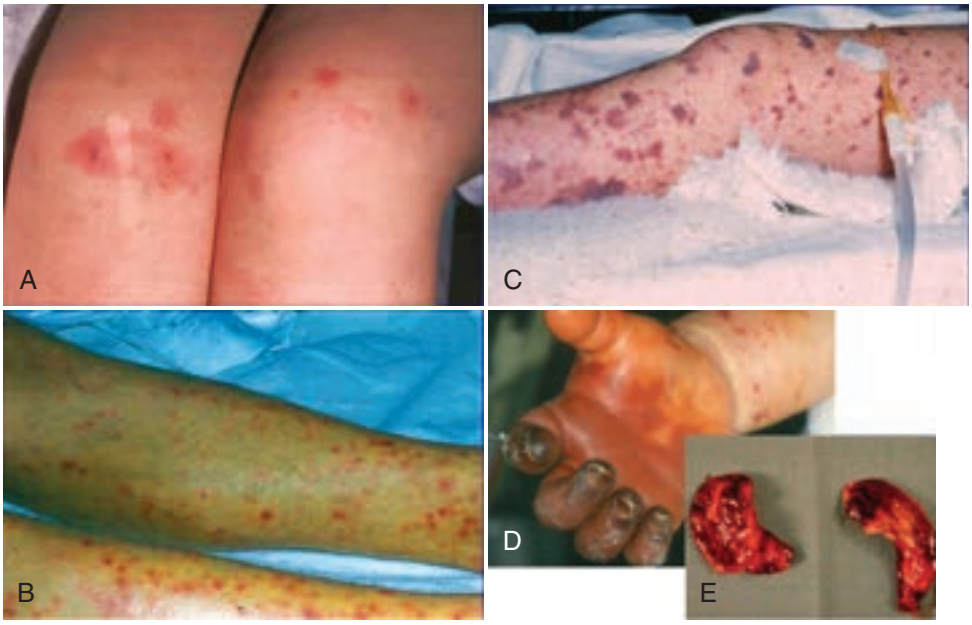

FIGURE 298-3. Clinical manifestations of meningococcal disease. A and B, Macular and petechial rashes of meningococcal bacteremia. C, Fulminant meningococcal sepsis with ecchymoses. D, Digital necrosis of meningococcemia sepsis. E, Hemorrhagic adrenals in fulminant meningococcal sepsis. (Modified from Stephens DS, Greenwood B, Brandtzaeg P. Epidemic meningitis, meningococcemia and *Neisseria meningitidis*. *Lancet.* 2007;369:2196-2210.)

manifested with low-grade fever and a polyarticular arthritis that can be confused with rheumatoid arthritis.

DIAGNOSIS

The clinical diagnosis of meningococcal meningitis relies on the recognition of fever, rash, meningeal signs, and altered mental status. The early clinical diagnosis of meningococcemia is a challenge because a rash, meningeal signs, and high fever may not be present. The course can be fulminant (<24 hours), and the early stages of disease can mimic viral infections such as those caused by enterovirus or influenza. Thus, it can be difficult to identify and to treat the disease quickly. General symptoms of sepsis (nausea and vomiting, drowsiness, irritability, leg pains, cold hands and feet, abnormal skin color) are present. However, these symptoms (in contrast to fever and rash) are not likely to be specific markers. Parents and relatives should be instructed to undress and inspect a febrile child, adolescent, or adult for a rash, and physicians and other health care providers should be alert to the concerns of parents or relatives about the abrupt or rapid deterioration of a patient.

The definitive diagnosis of invasive meningococcal disease is based on bacteriologic isolation or antigen or DNA identification of *N. meningitidis* in a usually sterile body fluid, such as blood, CSF, synovial fluid, pleural fluid, urine, or pericardial fluid. Blood and CSF are the most fruitful sources of positive cultures and for DNA identification by the polymerase chain reaction (PCR), but urine and skin lesions can also yield results in systemic meningococcal disease. The diagnosis of meningococcal meningitis is confirmed by CSF pleocytosis and Gram stain showing gram-negative diplococci (often inside neutrophils) and CSF culture, latex agglutination detecting meningococcal capsular polysaccharide in CSF, or PCR identifying *N. meningitidis* in CSF.

PCR is increasingly used for the diagnosis of meningococcal disease, including serogrouping and multilocus sequence typing, and has the potential to detect antibiotic resistance determinants. PCR techniques include real-time PCR of CSF, blood, and other sterile sites. Urine is a less sensitive fluid for PCR. An increasing number of patients are now diagnosed by PCR without culture, especially if they have received empiric prehospital antibiotic treatment.[14] The sensitivity of PCR for the diagnosis of meningococcal meningitis is more than 90 to 95%; in contrast, the sensitivity of CSF or blood culture is less than 65%.

TREATMENT

Rx

Early recognition and antibiotic administration (Table 298-1) are critical for effective treatment because effective antibiotics immediately stop the growth of *N. meningitidis*. Ceftriaxone, cefotaxime, and penicillin are effective antibiotics, as are meropenem and chloramphenicol.[15] Meningococci in CSF are killed within 3 to 4 hours after intravenous treatment with an adequate dose of a third-generation cephalosporin or penicillin, and concentrations of endotoxin in plasma fall by 50% within 2 hours. The concentrations of key cytokines and chemokines fall in parallel. Antibiotic treatment does not induce a large release of meningococcal endotoxin or lead to an increased inflammatory response.

Prehospital antibiotic treatment is advocated if the disease is suspected. One goal is to reduce the case-fatality rate for patients with fulminant meningococcal sepsis or meningitis with rapidly increasing concentrations of meningococci and inflammatory mediators in the circulation or CSF. If antibiotic treatment is initiated before admission, ceftriaxone or another effective antibiotic can be injected intravenously or intramuscularly. During epidemics in developing countries, a single injection of ceftriaxone or long-acting chloramphenicol may be sufficient for patients with meningitis,[■] and this simple treatment has saved many thousands of lives. The sensitivity of meningococci to penicillin is decreasing worldwide because of a reduced affinity for penicillin-binding protein 2, although high-level penicillin resistance remains rare in most countries. Fluoroquinolone resistance of meningococcal isolates, although rare, has also emerged. Ceftriaxone and cefotaxime can achieve CSF concentrations 45- to 8750-fold higher than the minimal inhibitory concentrations for meningococci. First-generation cephalosporins should not be used.

Patients with suspected bacterial meningitis of unknown cause are often given ceftriaxone or cefotaxime, often combined with vancomycin, until the causative agent has been identified. When *N. meningitidis* is identified, antibiotic treatment can be continued with a third-generation cephalosporin

TABLE 298-1 ANTIBIOTIC TREATMENT OF MENINGOCOCCAL MENINGITIS AND MENINGOCOCCEMIA

DRUG	AGE GROUP	DOSAGE
Ceftriaxone*	Children > 3 months Adults	50 mg/kg IV q12h 1-2 g IV q12h
Cefotaxime	Adults	50-75 mg/kg q6-8h; maximum dose, 12 g/day
Penicillin G	Adults	50,000 U/kg IV q4h; up to 4 million U q4h
Meropenem	Adults	2 g IV q8h, 6 g/day
If penicillin and cephalosporin allergic, chloramphenicol	Adults	25 mg/kg IV q6h, up to 1 g q6h

*Because of concerns in neonates from calcium/ceftriaxone precipitates and displacement of bilirubin from albumin by ceftriaxone, babies younger than 3 months should be started on cefotaxime 50 mg/kg q6-8h.

or possibly benzylpenicillin alone. Meropenem is also active clinically in the treatment of meningococcal meningitis, or chloramphenicol is a choice in penicillin- and cephalosporin-allergic patients. Traditionally, patients with meningococcal meningitis were treated for 7 days or longer, but 3 or 4 days of intravenous treatment can provide a cure without relapse.

Recognition of the different pathophysiologic processes associated with meningococcal meningitis (which causes death and morbidity predominantly by cerebral edema) and meningococcal septic shock (which causes death and morbidity predominantly through hypovolemia, capillary leak, myocardial dysfunction, and multiorgan failure) has led to improved management strategies for these two different forms of disease. Early and aggressive management of shock through the use of volume expansion, intensive care monitoring, and inotropic support can reduce fatality rates of meningococcal sepsis from higher than 30% to 5 to 10%. In meningococcemia with hypotension, the primary goal is to increase the circulating blood volume by aggressive fluid treatment. Both colloids and crystalloids (saline 0.9%) can be used without a demonstrated difference in effectiveness. Adults are given saline starting with 1 L infused intravenously during 15 to 20 minutes, followed by several liters at a reduced rate. Some patients require two to three times their own blood volume during the first 24 hours. The total fluid volume needed per 24 hours is determined by response to treatment—tissue perfusion, blood pressure, urine output, and evidence of intravascular volume overload. The volume treatment may be combined with a vasopressor such as dopamine, norepinephrine, epinephrine, or dobutamine. Fluid treatment may be complicated in patients with reduced renal function. They may need dialysis or hemofiltration to compensate for renal failure and to reduce the substantial edema that accumulates.

Patients with meningitis or mild meningococcemia should be given the normal daily fluid requirement, supplemented with the volume lost before admission unless there is evidence of the syndrome of inappropriate antidiuretic hormone. Excessive volume treatment in patients with meningitis can induce fatal brain edema and herniation. Management of raised intracranial pressure (hyperosmolar solutions, diuretics, mechanical ventilation), seizures, and hyponatremia is indicated.

Anticoagulant treatment of patients with meningococcemia and disseminated intravascular coagulation has not been documented to improve outcome. A phase III study of recombinant activated protein C in children with sepsis of all causes, including meningococcemia, was stopped because no benefit was noted, and recombinant activated protein C has been withdrawn from the market. Administration of activated protein C is associated with an increased risk of cerebral hemorrhage. The use of recombinant human tissue plasminogen activator does not appear to be beneficial. Randomized controlled clinical trials using hyperimmune serum, antibodies, or recombinant bactericidal or permeability-increasing protein (designed to inactivate *N. meningitidis* endotoxin) have not demonstrated a beneficial effect on survival. Blockade of other specific inflammatory mediators has not been adequately tested in meningococcal septic shock.

Patients with fulminant meningococcal septicemia can develop adrenal hemorrhage (Waterhouse-Friderichsen syndrome). The corticotropin concentration is higher, the cortisol concentration is lower, and the corticotropin-to-cortisol ratio is higher in patients with fatal meningococcal shock than in survivors. Adults with septic shock and indications of inadequate adrenal function are given low doses of steroids. Although the benefit has not been documented, many intensive care specialists use stress replacement doses of hydrocortisone in children with shock caused by *N. meningitidis*.

Plasmapheresis, blood exchange, and extracorporeal membrane oxygenation have been used in patients with meningococcemia; however, no controlled trials to assess the results have been done. Plasmapheresis and blood exchange appear to have little additive effect on the endogenous clearance of endotoxin and cytokines from the circulation. Extracorporeal membrane oxygenation has been used in several centers, with better results in children with acute pulmonary failure than in those with refractory septic shock. The use of insulin to control mild hyperglycemia in critically ill adults has not shown benefits (Chapter 108).

Pharmacologic doses of dexamethasone have been shown to reduce morbidity in pneumococcal and *Haemophilus influenzae* type b meningitis. The use of dexamethasone to reduce death caused by brain edema or to prevent sequelae such as deafness in patients with meningococcal meningitis remains unproved on the basis of large randomized controlled trials, but trends toward reductions in hearing loss, mortality, and arthritis after meningococcal disease are reported. Many now recommend that dexamethasone (10 mg every 6 hours for the first 4 days and in children at a dose of 0.15 mg/kg every 6 hours for 4 days beginning before or with the first dose of antibiotics) be given early in suspected or confirmed bacterial meningitis.■ Glycerol has been used to reduce intracranial pressure in bacterial meningitis, but its value is not proven. Other major life-threatening complications necessitating therapy include adult respiratory distress syndrome, neurologic sequelae ranging from coma to diabetes insipidus, pneumonia that is not necessarily meningococcal but may be secondary to aspiration during the obtunded state, and pericarditis.

Immune complex–mediated complications, such as arthritis, cutaneous vasculitis, iritis, episcleritis, pleuritis, and pericarditis, can first appear several days to 2 to 3 weeks after onset of illness, when the patient is otherwise improving. These complications, which can be multiple, are due to the deposition of antigen-antibody complexes composed of meningococcal capsular polysaccharide or other antigens, meningococcal-specific immunoglobulins, and C3 and complicate 6 to 15% of meningococcal meningitis or septicemia. Treatment is with aspirin or nonsteroidal anti-inflammatory drugs, and resolution is complete, usually within 14 days from the onset and usually without residual sequelae.

PREVENTION

Chemoprophylaxis to eliminate meningococcal carriage is recommended for close contacts of patients to prevent further transmission and disease (Table 298-2). The occurrence of meningococcal disease in household contacts is approximately 100-fold higher than in the general population. Secondary cases usually occur within 1 to 14 days of the primary case. Chemoprophylaxis can be helpful to control localized outbreaks, but it is generally not recommended for the control of large epidemics (see the later discussion of vaccines). Rifampin, ceftriaxone, azithromycin, and quinolones (but not penicillin) have the ability to eradicate meningococci in the nasopharynx. A majority of meningococcal isolates are now resistant to sulfonamides; resistance to rifampin can develop rapidly, and quinolone resistance in meningococci is reported.

Prevention through vaccination is the best option for the long-term control of meningococcal disease.[16] Capsular polysaccharide vaccines to decrease serogroup A, C, Y, and W-135 meningococcal disease were introduced in the 1970s and 1980s. These vaccines were safe, with mild local adverse events, and effective (>85%) in children older than 2 years and adults but less immunogenic in younger children; immunity to the polysaccharide vaccines was limited to 3 to 5 years of protection, and immunologic hyporesponsiveness was induced by repeated doses of the polysaccharide. Also, meningococcal polysaccharide vaccines do not induce immunologic memory and have little or no effect on nasopharyngeal carriage. Although these vaccines were used extensively to control disease in military populations and in epidemics in the African meningitis belt, in the latter they were often deployed too late in the course of an outbreak. There was no evidence that widespread use of polysaccharide vaccines reduced the frequency of epidemics in Africa.

A major advance in the past 15 years has been the development and now widespread use of meningococcal polysaccharide-protein conjugate vaccines and their introduction first into the United Kingdom and then other parts of Europe, Canada, Australia, the United States, and, more recently, the African meningitis belt. These vaccines are safe and immunogenic in young children, induce immunologic memory, and can decrease nasopharyngeal carriage of meningococci.■ In the United Kingdom, the introduction of the serogroup C conjugate meningococcal vaccine in 2000 to all children and young adults greatly reduced the rate of serogroup C disease (90% vaccine effectiveness at 3 years for patients aged 11 to 18 years). A major protective effect of the C conjugate vaccine is mediated through herd immunity.[17] Rates of serogroup C carriage and disease in nonvaccinated individuals were reduced by more than 50% through herd immunity. Polysaccharide-protein conjugate meningococcal vaccines containing serogroups A, C, Y, and W-135 were introduced for adolescents in the United States in 2005 and subsequently extended to children aged 2 months to 10 years at increased risk for meningococcal disease. In addition to routine use in older children and adolescents (first dose at the age of 11 or 12 years with booster at 16 years), populations that should benefit from the new conjugate vaccines are college freshmen, military recruits, patients with immunoglobulin or complement deficiencies (inherited or chronic deficiencies such as C3, properdin, factor D, or late complement components), patients with anatomic or functional asplenia, microbiologists who are routinely exposed to isolates of *N. meningitidis*, adults with HIV type 1 infections, and people who travel to or reside in countries where *N. meningitidis* is epidemic. An important example of a new approach to meningococcal conjugate vaccine development is the Meningitis Vaccine Program, a partnership between PATH, the World Health Organization, and the Global Alliance for Vaccines and Immunization for the development of a group A meningococcal conjugate vaccine for Africa, designated MenAfriVac, at less than $0.50 a dose.[18] Because of the huge impact of herd immunity of the serogroup C conjugate vaccines in the United Kingdom, MenAfriVac was introduced as a mass vaccination strategy for those 1 to 29 years old. The MenAfriVac vaccination campaign began in Burkina Faso in December 2010 and has been extended to Mali, Chad, Niger, Nigeria, Benin, Ghana, Senegal, Cameroon, Sudan, and other regions of the African meningitis belt. More than 100 million doses have been administered. Initial results

TABLE 298-2 CHEMOPROPHYLAXIS AGAINST MENINGOCOCCAL INFECTION

DRUG	AGE GROUP	DOSAGE	DURATION AND ROUTE OF ADMINISTRATION*	CONSIDERATIONS
Rifampin	Children <1 month	5 mg/kg q12h	2 days	
	Children >1 month	10 mg/kg q12h (maximum, 600 mg)	2 days	
	Adults	600 mg q12h	2 days	Rifampin can interfere with efficacy of oral contraceptives and some seizure prevention and anticoagulant medications; may stain soft contact lenses Not recommended for pregnant women
Ceftriaxone	Children <15 years	125 mg	Single IM dose	
	Children >15 years and adults	250 mg	Single IM dose	Ceftriaxone is recommended for prophylaxis in pregnant women.
Ciprofloxacin	Adults	500 mg	Single dose	Not recommended routinely for persons <18 years of age, but use in infants and children (20 mg/kg) may be justified after careful assessment of the risks and benefits Not recommended for pregnant or lactating women Cases of ciprofloxacin resistance have been reported, and use for prophylaxis should be based on local sensitivity of the meningococcus to the drug.
Azithromycin		10 mg/kg (maximum, 500 mg)	Single dose	Equivalent to rifampin for eradication of meningococci from nasopharynx, but data are limited

ANTIBIOTIC CHEMOPROPHYLAXIS FOR HOUSEHOLD OR INTIMATE CONTACTS:

- Household contacts and persons sharing the same living quarters, particularly young children
- Daycare center, nursery school, or child care contacts; frequent playmates of young children
- Close social contacts that were exposed to oral secretions in the week before onset, such as by kissing and sharing of eating and drinking utensils or toothbrushes
- For airline travel lasting more than 8 hours, passengers who are seated directly next to an infected person should receive prophylaxis.
- Routine prophylaxis is not recommended for health care professionals unless they have had intimate exposure to respiratory secretions.
- As the risk of secondary cases is highest during the first few days after exposure, chemoprophylaxis should be initiated as soon as possible, ideally <24 hours after identification of the index patient.
- If more than 14 days have passed since the last contact with the index patient, chemoprophylaxis is not likely to be of benefit.
- Pharyngeal cultures are not helpful in determining the need for chemoprophylaxis and may unnecessarily delay the use of effective chemoprophylaxis.
- Chemoprophylaxis has also been recommended for patients given penicillin or chloramphenicol for treatment because pharyngeal carriage may not be eliminated with these antibiotics and the patient could remain colonized with a virulent strain.
- Ceftriaxone is recommended for pregnant women.
- May want to avoid ciprofloxacin or azithromycin in individuals at risk of QT-prolongation.

Recommended groups for chemoprophylaxis based on exposure to the case in the week before onset of illness.
*Administered orally unless otherwise stated.

show the virtual elimination of serogroup A meningococcal disease in the countries vaccinated in the first 2 years after the mass campaigns.

The development of vaccines for serogroup B *N. meningitidis* has also shown progress. Serogroup B can cause prolonged outbreaks during many years, such as those seen in the past two decades in the Pacific Northwest (Oregon, parts of Washington), Brazil, Norway, and New Zealand. The serogroup B capsule has an identical structure to polysialic structures expressed in fetal neural tissue and does not induce a protective bactericidal immunoglobulin G response. Thus, strategies have been focused on noncapsular antigens, such as outer membrane proteins containing vesicles (OMV) or conserved protein antigens. The diversity of major outer membrane structures in meningococci has limited OMV approaches, but this approach has been successful in controlling serogroup B epidemics that are strain specific (e.g., in New Zealand). New serogroup B vaccines based on semiconserved surface protein antigens identified by "reverse vaccinology" are the most advanced candidates and are obtaining licensure. One of these, 4CMenB or Bexosero,■ recently approved in Europe, contains three semiconserved surface protein antigens—a member of the factor H–binding protein family, neisserial adhesin A (NadA), and neisserial heparin-binding antigen—and a meningococcal serogroup B PorA-containing OMV preparation previously used to control the serogroup B clonal outbreak in New Zealand. Alum is the adjuvant. A second serogroup B vaccine is based on two members of the factor H–binding protein family and was recently licensed in the United States.[19] Initial evaluation of the immunogenicity and safety of these vaccines appears promising. The prevention of serogroup B disease appears to be a significant step closer with these new vaccines.

PROGNOSIS

Historically, the mortality of untreated systemic meningococcal disease was 70 to 90%. Despite highly effective antibiotics and aggressive supportive care, the mortality of invasive meningococcal disease remains at about 10%. The failure to recognize disease early, the very rapid development of disease (especially meningococcemia), and the time to administration of antibiotics remain the most significant challenges. The chance of surviving shock is directly correlated to plasma concentrations of endotoxin, and half the nonsurviving patients with shock die within the first 12 hours of hospital admission.

Long-term sequelae and morbidity after invasive meningococcal disease are significant.[20] Neurologic impairment occurs in 7 to 10% with meningococcal meningitis, with palsies of the sixth, seventh, and eighth cranial nerves and hemiparesis and quadriparesis. Unilateral or bilateral sensorineural hearing loss occurs in 2 to 9% of cases, which is profound in 2% of affected individuals and necessitates cochlear implantation in 0.4%. Neurodevelopmental impairment, including behavioral and psychological problems, learning difficulties, memory deficits, executive function problems, decreased academic performance, spasticity, seizures, and focal neurologic signs, is seen in approximately 10%. Visual difficulties, seizures, and motor deficits are reported in 2 to 3%, with multiple neurologic disabilities occurring in 1 to 2% of affected individuals. Survivors of meningococcal sepsis in childhood have, in 5 to 20% as young adults, long-term behavioral and emotional problems, decreased intellectual functioning, and illness-related physical or social consequences.

Scarring of the skin, secondary to necrotic purpura, may vary from unnoticeable to requiring skin grafting. Multiple areas may be involved; the lower limbs are most frequently affected, followed by the arms, chest, and face. Amputations, of the digits or limbs, are frequently multiple; these result from necrosis of the skin, muscle, and bone of the affected parts (Fig. 298-3D) and, depending on the site and extent, may require prostheses to improve function or appearance. Bone growth disturbances, stump overgrowth, scar contractures, and soft tissue and bone infections may complicate amputations. Limb length discrepancies, which may result from the growth plate infarction, often necessitate further surgical intervention. Following acute renal failure at presentation, renal function recovers in the majority of individuals; however, evidence of renal dysfunction may persist for more than 4 years in both children and adults, with the risk being higher in those who required renal replacement therapy.

Meningococcal disease and its complications, often occurring rapidly in otherwise healthy individuals, also produce significant family, community, health care, and public health impact. The emotional toll on individuals who

survive and on the families of those with meningococcal disease in intensive care units, of those who survive with complications, and of those who die is considerable and a global phenomenon. In communities, meningococcal disease may also create considerable fear and anxiety. Post-traumatic stress disorder occurs at a higher frequency in both patients and families, often months after the illness. In one study, post-traumatic stress disorder occurred in 15% of children, in half of the mothers, and in 19% of fathers at 3 months. Meningococcal disease and its complications also result in substantial hospital and long-term health care costs. Furthermore, delay in the diagnosis of meningococcal sepsis and meningitis and septicemia is a common reason for litigation.

If parents and health care professionals recognize the importance of fever and headache with a nonblanching rash and seek treatment early, morbidity and mortality can be reduced with prehospital antibiotic treatment, rapid transportation to medical facilities, and stabilization in an intensive care unit. Prevention of meningococcal disease with new vaccines and vaccine strategies remains the major worldwide goal.

Grade A References

A1. Nathan N, Borel T, Djibo A, et al. Ceftriaxone as effective as long-acting chloramphenicol in short-course treatment of meningococcal meningitis during epidemics: a randomised non-inferiority study. *Lancet*. 2005;366:308-313.

A2. Brouwer MC, McIntyre P, Prasad K, et al. Corticosteroids for acute bacterial meningitis. *Cochrane Database Syst Rev*. 2013;6:CD004405.

A3. Cohn AC, MacNeil JR, Clark TA, Centers for Disease Control and Prevention (CDC), et al. Prevention and control of meningococcal disease: recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm Rep*. 2013;62(RR-2):1-28.

A4. Read RC, Baxter D, Chadwick DR, et al. Effect of a quadrivalent meningococcal ACWY glycoconjugate or a serogroup B meningococcal vaccine on meningococcal carriage: an observer-blind, phase 3 randomised clinical trial. *Lancet*. 2014;384:2123-2131.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Martin NG, Sadarangani M, Pollard AJ, et al. Hospital admission rates for meningitis and septicaemia caused by *Haemophilus influenzae*, *Neisseria meningitidis*, and *Streptococcus pneumoniae* in children in England over five decades: a population-based observational study. *Lancet Infect Dis*. 2014;14:397-405.
2. Castelblanco RL, Lee M, Hasbun R. Epidemiology of bacterial meningitis in the USA from 1997 to 2010: a population-based observational study. *Lancet Infect Dis*. 2014;14:813-819.
3. Pace D, Pollard AJ. Meningococcal disease: clinical presentation and sequelae. *Vaccine*. 2012; 30S:B3-B9.
4. Joseph B, Schwarz RF, Linke B, et al. Virulence evolution of the human pathogen *Neisseria meningitidis* by recombination in the core and accessory genome. *PLoS ONE*. 2011;6:e18441.
5. Cohn AC, MacNeil JR, Harrison LH, et al. Changes in *Neisseria meningitidis* disease epidemiology in the United States, 1998-2007: implications for prevention of meningococcal disease. *Clin Infect Dis*. 2010;50:184-191.
6. Chang Q, Tzeng YL, Stephens DS. Meningococcal disease: changes in epidemiology and prevention. *Clin Epidemiol*. 2012;4:237-245.
7. Coureuil M, Join-Lambert O, Lecuyer H, et al. Pathogenesis of meningococcemia. *Cold Spring Harb Perspect Med*. 2013;3.
8. Hill DJ, Virji M. Meningococcal ligands and molecular targets of the host. *Methods Mol Biol*. 2012;799:143-152.
9. Coureuil M, Join-Lambert O, Lécuyer H, et al. Mechanism of meningeal invasion by *Neisseria meningitidis*. *Virulence*. 2012;3:164-172.
10. Bryant CE, Spring DR, Gangloff M, et al. The molecular basis of the host response to lipopolysaccharide. *Nat Rev Microbiol*. 2010;8:9.
11. Ram S, Lewis LA, Rice PA. Infections of people with complement deficiencies and patients who have undergone splenectomy. *Clin Microbiol Rev*. 2010;23:740-780.
12. Brandtzaeg P, van Deuren M. Classification and pathogenesis of meningococcal infections. *Methods Mol Biol*. 2012;799:21-35.
13. Heckenberg SG, Brouwer MC, van de Beek D. Bacterial meningitis. *Handb Clin Neurol*. 2014;121:1361-1375.
14. Nemescu RE, Iancu LS, Dorneanu OS, et al. Influence of antibiotic therapy prior to admission on the efficacy of classical methods for the diagnosis of meningococcal disease. *Rev Med Chir Soc Med Nat Iasi*. 2014;118:497-502.
15. van de Beek D, Brouwer MC, Thwaites GE, et al. Advances in treatment of bacterial meningitis. *Lancet*. 2012;380:1693-1702.
16. Sáfadi MA, Bettinger JA, Maturana GM, et al. Evolving meningococcal immunization strategies. *Expert Rev Vaccines*. 2014;1-13.
17. Stephens DS. Protecting the herd: the remarkable effectiveness of the bacterial meningitis polysaccharide-protein conjugate vaccines in altering transmission dynamics. *Trans Am Clin Climatol Assoc*. 2011;122:115-123.
18. Novak RT, Kambou JL, Diomandé FV, et al. Serogroup A meningococcal conjugate vaccination in Burkina Faso: analysis of national surveillance data. *Lancet Infect Dis*. 2012;12:757-764.
19. Centers for Disease Control and Prevention (CDC). Meningococcal Disease. Serogroup B Meningococcal Vaccine & Outbreaks. http://www.cdc.gov/meningococcal/outbreaks/vaccine-serogroupB.html. Accessed January 27, 2015.
20. Edmond K, Clark A, Korczak VS, et al. Global and regional risk of disabling sequelae from bacterial meningitis: a systematic review and meta-analysis. *Lancet Infect Dis*. 2010;10:317-328.

REVIEW QUESTIONS

1. *Neisseria meningitidis* is a

A. Gram-negative rod causing pneumonia and bacteremia
B. Gram-positive diplococcus causing primarily pneumonia
C. Gram-negative diplococcus causing epidemic meningitis and septicemia
D. Gram-positive rod causing meningitis in immunocompetent patients
E. Gram-variable coccobacillus causing meningitis and otitis media

Answer: C *N. meningitidis* is an aerobic, diplococcal gram-negative β-proteobacterium and a member of the family Neisseriaceae, which also includes *Neisseria gonorrhoeae* (Chapter 299), another important human pathogen. *N. meningitidis* is the cause of epidemic bacterial meningitis, fulminant sepsis (meningococcemia), milder bacteremia, and, less commonly, focal infections (such as pneumonia, septic arthritis, purulent pericarditis, and conjunctivitis).

2. Serogroup A *N. meningitidis*

A. Causes significant disease in the United States and Japan
B. Has been a major cause of epidemic meningococcal outbreaks in sub-Saharan Africa
C. Expresses a sialic acid capsule
D. Has no effective vaccine
E. Is usually treated with sulfonamides

Answer: B In sub-Saharan Africa, seasonal increases in disease and cyclic pandemics, which have occurred every 8 to 10 years since 1905, have been predominantly caused by serogroup A *N. meningitidis*. A new serogroup A conjugate vaccine introduced in 2010 has rapidly lowered the incidence of meningococcal disease in this region.

3. Which one is not true? The rash of invasive meningococcal disease

A. Can be a blanching macular rash
B. Is typically petechial or purpuric
C. Can be confused with leukocytoclastic vasculitis, Rocky Mountain spotted fever, and enteroviral infections
D. Is almost always present and easily detected
E. Is usually a single "bull's-eye" lesion

Answer: D Hemorrhagic skin lesions (petechiae, purpura; Fig. 298-3) are present in 28 to 77% of patients with invasive meningococcal disease on admission, but these lesions may be absent or difficult to see in patients with dark skin. A nonblanching macular rash can also be a manifestation of meningococcal bacteremia.

4. Which one is not true regarding antibiotic treatment of meningococcal meningitis?

A. It should be administered as soon as the diagnosis is considered.
B. It takes 10 to 14 days to be effective.
C. Options include ceftriaxone or cefotaxime, penicillin, or meropenem.
D. Quinolone resistance has been reported.
E. Sulfonamide resistance is now common.

Answer: B Meningococci in cerebrospinal fluid are killed within 3 to 4 hours after intravenous treatment with an adequate dose of a third-generation cephalosporin or penicillin. Traditionally, patients with meningococcal meningitis were treated for 7 days or longer, but 3 or 4 days of intravenous treatment can provide a cure without relapse.

5. Which one is not true regarding prevention of meningococcal disease?

A. Chemoprophylaxis is given to prevent secondary cases that occur usually 10 to 14 days after the primary case.
B. Meningococcal polysaccharide- protein conjugate vaccines are a long-term control measure of meningococcal disease.
C. Meningococcal vaccines have effectiveness in complement-deficient patients.
D. Rifampin, ceftriaxone, ciprofloxacin, and azithromycin are used for chemoprophylaxis.
E. Meningococcal polysaccharide vaccines are now preferred for prevention strategies.

Answer: E Meningococcal conjugate vaccines are safe and immunogenic in young children, induce immunologic memory, and can decrease nasopharyngeal carriage of meningococci. Meningococcal conjugate vaccines have largely replaced the older polysaccharide vaccines and are now preferred for prevention strategies.

299 *NEISSERIA GONORRHOEAE* INFECTIONS

MATTHEW R. GOLDEN AND H. HUNTER HANDSFIELD

DEFINITION

Neisseria gonorrhoeae is a sexually transmitted organism that infects primarily the columnar epithelia of mucosal surfaces and causes urethritis in men and endocervicitis and urethritis in women. Other sites of primary infection include the rectum, pharynx, and conjunctiva, and vulvovaginitis can occur in prepubertal girls. The most common complication of gonococcal infection is pelvic inflammatory disease, which can lead to infertility, ectopic pregnancy, and chronic pelvic pain. Other much less common complications include epididymitis, posterior urethritis, urethral stricture, Bartholin's gland abscess, and perihepatitis. Bacteremia may occur, with the production of characteristic cutaneous lesions, arthritis, and, rarely, endocarditis or meningitis. Neonatal conjunctivitis (ophthalmia neonatorum) was formerly a common cause of blindness. Gonococcal infections are also thought to increase the risk of human immunodeficiency virus (HIV) transmission from persons dually infected with HIV and *N. gonorrhoeae* and to increase the risk of HIV acquisition among persons with gonorrhea who are exposed to HIV.

The Pathogen

The gonococcal envelope is similar in its basic structure to that of other gram-negative bacteria and is composed of an inner cytoplasmic membrane, a middle peptidoglycan cell wall, and an outer membrane. The outer membrane contains several surface components that play a central role in the organism's interaction with the host and its pathogenicity. Pili, hairlike projections also referred to as fimbriae, are composed of several different protein subunits and, along with other outer membrane adhesins (i.e., opacity-related proteins), facilitate attachment and invasion of host cells. Gonococci vary the composition of these proteins and surface lipo-oligosaccharides over time, allowing the organism to elude host defenses. This phase variation has also been a barrier to successful vaccine development.

EPIDEMIOLOGY

Gonorrhea is the second most commonly reported infectious disease in the United States, with 334,826 cases reported to the Centers for Disease Control and Prevention (CDC) in 2012.[1] The incidence of gonorrhea in the United States in 2012 was 107.5 per 100,000 population, although this number, which is based on cases reported to U.S. health departments, is undoubtedly an underestimate. The overall incidence of gonorrhea in the United States has been relatively stable for more than a decade, and current rates are more than 70% lower than the incidence observed in the mid-1970s. The prevalence of gonorrhea varies widely, depending on the population tested. The National Longitudinal Study of Adolescent Health (Add Health) tested a representative sample of U.S. 18- to 26-year-olds in 2001 and 2002 and found that 0.43% were infected, with wide variations geographically and among population groups. This prevalence estimate is almost identical to one derived from the National Health and Nutrition Examination Survey of 1998 to 2008, another population-based survey.[2] Among women aged 15 to 24 years tested for gonorrhea in family planning clinics in 48 U.S. states and territories in 2011, the median prevalence of gonorrhea was 0.7%, but it varied from 0 to 3.5%. By comparison, the prevalence of chlamydial infection in the Add Health study was 4.19%, and among 15- to 24-year-old females tested in U.S. family planning clinics in 2011, it was 8.3% (range, 3.8 to 15.9%).

Like virtually all other sexually transmitted infections (STIs; Chapter 285), gonorrhea rates vary widely with age, geographic location, sexual orientation, and race or ethnicity. In the United States, the rate of reported infection is highest among 15- to 24-year-olds, and rates in the South are more than twice those observed in the West and Northeast. However, the most glaring disparities are observed in comparing rates of infection among different racial and ethnic groups and in comparing heterosexuals to men who have sex with men (MSM).

Rates of reported gonorrhea among U.S. African Americans are almost 15 times higher than those among non-Hispanic whites, and rates among American Indians/Alaska Natives and Hispanics are 4 and 1.9 times higher, respectively, than those among non-Hispanic whites. Variations in reporting probably account for some of these observed differences because low socioeconomic status is associated with receiving care in public health clinics and other venues where there is more complete reporting than in the private health care sector. However, the marked disparity in reported rates is also apparent in population-based studies, clearly establishing that the different rates of infection in different racial and ethnic groups are not simply a result of reporting bias. The reasons for this profound disparity are certainly multifactorial and are not entirely clear. Different racial groups in the United States vary little in terms of their number of sex partners, so this cannot explain the different rates of STIs. Although inadequate access to medical care is likely to play a role in the racial disparities in STI rates, profound racial and ethnic disparities are also observed in the United Kingdom and the Netherlands, nations with nationalized health care systems in which access to care should be more uniform than it is in the United States. Research has highlighted the importance of concurrency (i.e., partnerships that overlap in time) and patterns of sexual mixing based on age, race, and level of sexual activity as critical determinants of a population's risk of STI. These factors are thought to be shaped by social factors (e.g., poverty, incarceration, joblessness, racism) that play a central role in defining the epidemiology of all STIs worldwide.

Gonorrhea also disproportionately affects gay and bisexual men and other MSM. Numerous cities in the United States, western Europe, and Australia have reported increases in the rate and number of gonorrhea cases among MSM since the mid-1990s. For example, in King County, Washington State, in 2012 the estimated incidence of reported gonorrhea in MSM was more than 2114 cases per 100,000, compared with 40 and 41 per 100,000 heterosexual men and women, respectively.

Gonorrhea is highly transmissible. Although it is not precisely defined, the risk of transmission from a man to a woman during a single episode of unprotected vaginal intercourse is thought to be 50 to 70%, and the risk of transmission from a woman to a man is 20%. The transmission risks associated with anal sex, fellatio, or cunnilingus are not well defined, but anal intercourse is probably a relatively efficient mode of transmission, and receipt of fellatio is likely to be an important source of gonococcal infections in at

least some populations (e.g., MSM). Gonococci die rapidly on drying, and with the exception of occasional acquisition by laboratory personnel working with the organism, nonsexual transmission does not occur in adults. Perinatal transmission causing neonatal ophthalmitis or pharyngeal infection is now rare.

PATHOBIOLOGY

After attachment to host epithelial cells, gonococci are endocytosed into the cell in a process thought to be facilitated by Por (or protein 1). Gonococci then replicate within the host cell and are released into the subepithelial space.

Typical urethral infections result in prominent inflammation, probably as a result of the release of toxic lipo-oligosaccharide and peptidoglycan fragments as well as the release of chemotactic factors that attract neutrophilic leukocytes. The reasons that some gonococcal strains selectively cause asymptomatic genital infection are poorly understood, but this propensity may be related to differences in the organism's ability to bind complement-regulatory proteins that downregulate the production of chemotactic peptides. In particular, gonococcal strains that express PorB1A appear to bind factor H and complement-binding protein and have an increased propensity for causing disseminated gonococcal infections.

Although the gonococcus is not highly mutable, many gonococci possess conjugative plasmids and are consequently able to efficiently transfer other, nontransferable plasmids, such as those conferring resistance to penicillin and tetracycline. Gonococci are also capable of efficiently transferring naked DNA (transformation). These characteristics are important in the organism's ability to develop resistance to antimicrobials. For example, recent evidence suggests that the gonococci with diminished susceptibility to oral cephalosporins possess genetic resistance mutations acquired from commensal *Neisseria* species commonly found in the oropharynx.[3,4]

CLINICAL MANIFESTATIONS

Gonococcal infections can result in a number of specific clinical syndromes, each of which has its own manifestations, differential diagnosis, and recommended evaluation. The major clinical manifestations of gonococcal infection are discussed separately later. Gonococcal ophthalmia, now a rare complication, can result from direct contact or by autoinoculation in individuals with anogenital gonorrhea and is manifested as an acute, purulent conjunctivitis that can result in corneal ulceration if it is not treated promptly.

DIAGNOSIS

Microscopy

Microscopy of a Gram-stained smear is positive when polymorphonuclear neutrophils are observed to contain intracellular gram-negative diplococci of typical morphology (Fig. 299-1). Gram-stained urethral smears are 90 to 98% sensitive in the diagnosis of symptomatic gonococcal urethritis in men and have a specificity greater than 95%. However, Gram stain is only approximately 50% sensitive for cervical or rectal infection and for asymptomatic urethral gonorrhea. Although Gram stain is often considered highly specific for such infections, the actual performance varies with the skill and experience of the examiner, and rectal and cervical smears are unreliable in many clinical settings. Smears are both insensitive and nonspecific for pharyngeal gonococcal infection and are not recommended.

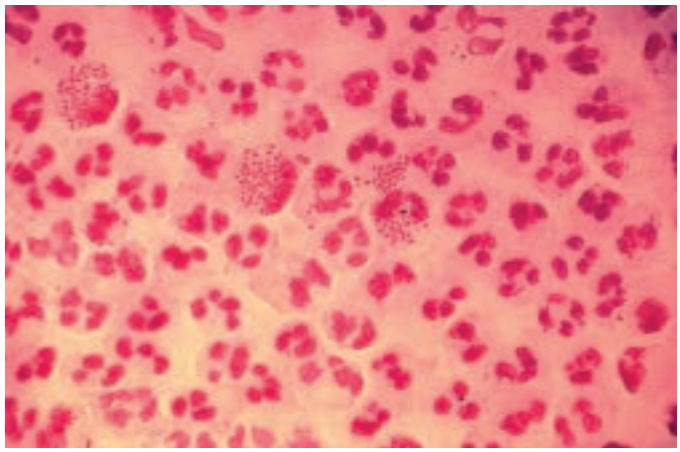

FIGURE 299-1. Gram stain in an acute case of gonococcal urethritis. This slide is used to demonstrate the nonrandom distribution of gonococci among polymorphonuclear neutrophils. Note that there are both intracellular and extracellular bacteria in the field of view.

Culture

Isolation of *N. gonorrhoeae* by culture, generally with antibiotic-containing selective media, is the historic mainstay of the diagnosis of gonorrhea. Despite the proliferation of molecular diagnostic methods, culture retains important roles in surveillance for antimicrobial resistance and selected clinical settings. Ideally, growth media should be inoculated directly and placed promptly into a humid atmosphere with increased carbon dioxide, such as a candle extinction jar. However, standard transport systems (e.g., Culturette) are acceptable if specimens are kept moist, are not refrigerated, and are processed within 6 hours. In testing of specimens not likely to be colonized by competing flora (e.g., synovial fluid), nonselective chocolate agar should be used.

Nucleic Acid Amplification Tests

In most settings in the United States, nucleic acid amplification tests (NAATs) have now supplanted culture as the dominant laboratory test used to diagnose gonorrhea.[5] NAATs approved by the Food and Drug Administration include polymerase chain reaction, transcription-mediated amplification (TMA), and DNA strand displacement (SDA). The advantages of NAATs include a slight increase in sensitivity over culture, the ability to test urine specimens and self-obtained vaginal swabs, and the fact that most NAATs are now marketed as combination assays that allow simultaneous testing for *N. gonorrhoeae* and *Chlamydia trachomatis*. The disadvantages of NAATs include the inability to perform antimicrobial susceptibility testing and the poor or inadequately defined positive predictive value of some NAATs when they are used to test low-prevalence populations. Because the prevalence of *N. gonorrhoeae* in family planning clinics in many parts of the United States and other nations is now well below 1%, the risk of false-positive screening results may be high, and reliable results depend on the use of assays with exquisite specificity. Existing evidence suggests that TMA has a high positive predictive value even in very low-prevalence settings; the positive predictive value of SDA and the current-generation Roche polymerase chain reaction is not well defined.

Although NAATs have been approved by the Food and Drug Administration only for the testing of genital tract and urine specimens, increasing evidence suggests that at least some NAATs (e.g., TMA, SDA) are substantially more sensitive than culture in detecting *N. gonorrhoeae* in pharyngeal and rectal specimens and that these NAATs are sufficiently specific to screen high-risk populations for rectal and pharyngeal gonorrhea. Given the decreasing availability of gonococcal culture, the poor sensitivity of culture on non–genital tract specimens, and the high prevalence of asymptomatic rectal and pharyngeal infections in some populations (particularly MSM), clinicians caring for patients at high risk for nongenital gonococcal infections should be able to use NAATs. Recent studies suggest that self-obtained rectal and pharyngeal specimens yield accurate results and are acceptable to MSM.

CLINICAL SYNDROMES

Urogenital Gonorrhea in Males

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Gonococcal urethritis in men is typically characterized by a purulent urethral discharge and dysuria. The usual incubation period is 2 to 6 days. A small minority of men who acquire urethral infection—generally estimated at 1 to 10%, and varying between specific strains of *N. gonorrhoeae*—remain asymptomatic.

Physical examination typically reveals purulent urethral exudate (Fig. 299-2); this is usually readily apparent, but compression of the urethra is sometimes required to express the exudate. Erythema of the meatus is sometimes present. Nongonococcal urethritis (Chapter 285) is typically characterized by less copious and less purulent discharge.

The diagnosis of gonococcal urethritis is usually suspected clinically, confirmed preliminarily by a Gram-stained smear showing leukocytes with intracellular gram-negative diplococci (see Fig. 299-1), and made definitively when *N. gonorrhoeae* is identified by culture or NAAT. Despite the usual clinical differences between gonococcal urethritis and nongonococcal urethritis, substantial overlap exists, and microbiologic diagnosis should be routine even in clinically typical cases.

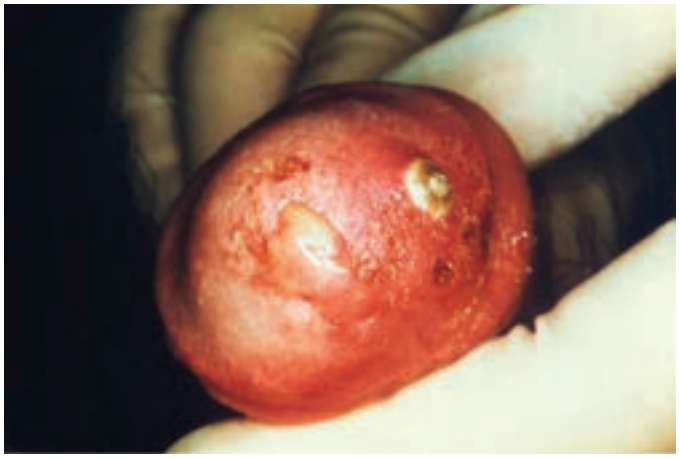

FIGURE 299-2. Male patient with a purulent penile discharge from gonorrhea and an overlying penile pyodermal lesion. Pyoderma involves the formation of a purulent skin lesion, which in this case is located on the glans penis.

PROGNOSIS

With prompt treatment, urethral gonorrhea seldom results in significant long-term morbidity. Acute epididymitis complicates gonococcal urethritis in less than 1% of cases. Patients with epididymitis usually present with unilateral testicular pain and swelling, sometimes with fever. Posterior urethritis or prostatitis, typically manifested as pelvic or perineal pain and urinary retention, was once fairly common but is now rare. Urethral stricture, another formerly common complication, is now very rare. Gonorrhea is associated with an elevated risk of HIV infection, both directly and as an epidemiologic risk factor. Diagnosis of gonorrhea should alert clinicians to counsel such patients about sexual risks, test them for HIV infection, and encourage patients to seek frequent follow-up testing for HIV infection and other STIs.

Lower Genital Tract Gonorrhea in Females

CLINICAL MANIFESTATIONS, DIAGNOSIS, AND PROGNOSIS

The primary site of infection in women is the endocervical canal. The proportion of infected women who develop symptoms is not precisely known, but probably about 50% of incident infections are symptomatic. In any case, asymptomatic infections accumulate in populations, whereas many or most symptomatically infected women present for diagnosis and treatment. Therefore, most prevalent infections among women are asymptomatic or are associated with mild symptoms not perceived by patients as abnormal or important. Gonorrhea is sometimes associated with an abnormal vaginal discharge, but other lower genital tract infections, such as bacterial vaginosis, trichomonas vaginitis, and candidal vaginitis, are much more common causes of this symptom. Lower genital tract gonorrhea in women can be associated with abnormal vaginal bleeding, which typically is manifested as metrorrhagia, scant intermenstrual bleeding, or postcoital spotting. *N. gonorrhoeae* sometimes causes dysuria and can be isolated from the urethra in up to 80% of women with gonorrhea. However, the urethra is rarely the only infected site, except in women who have had hysterectomies. In a minority of women, physical examination is notable for a purulent or mucopurulent cervical discharge, cervical edema, or easily induced cervical bleeding, signs of mucopurulent cervicitis. In uncomplicated infections, purulent exudate can sometimes be expressed from a Bartholin's gland duct, near the vaginal introitus laterally, or from Skene's glands, adjacent to the urethral meatus.

Microbiologic diagnosis usually rests on identifying *N. gonorrhoeae* in cervical secretions by NAAT or culture. Gram-stained smears are insensitive and seldom used.

The most important common complication of lower genital tract gonorrhea in women is pelvic inflammatory disease (see later). Rarer complications include Bartholin's gland abscess, which is manifested as a tender introital mass and may involve superinfection with facultative and anaerobic bacteria.

Pelvic Inflammatory Disease

CLINICAL MANIFESTATIONS

Pelvic inflammatory disease (PID) refers to an infection of the lower female genital tract and can involve the uterus (endometritis), fallopian tubes (salpingitis), and ovaries (oophoritis) as well as neighboring pelvic structures.[6] An estimated 10 to 40% of women with endocervical gonococcal infections develop PID, and gonorrhea is thought to be the cause of approximately 5 to 30% of all diagnosed cases of PID in the United States. However, this proportion varies with overall rates of gonorrhea and other causes of PID in the population, such as chlamydial infection.

Low abdominal pain is the dominant symptom of PID. Pain is variable in intensity and is often mild; it is usually bilateral, typically present for days to weeks before clinical presentation, and can be exacerbated by coitus. Symptoms often coincide with the onset or cessation of menses. Approximately one third of women have abnormal vaginal bleeding. Fever, chills, anorexia, vaginal discharge, urethritis, and proctitis occur, but they are neither sensitive nor specific in identifying women with PID. There is little if any difference in the severity of symptoms and signs of PID associated with *N. gonorrhoeae*, *C. trachomatis*, or neither of these pathogens.

The physical examination is typically notable for diffuse abdominal tenderness that is greatest in the lower quadrants and for tenderness of the pelvic organs on bimanual examination, with or without manipulation of the cervix. Most women have signs of cervicitis or bacterial vaginosis. Fever is present in a minority of cases. Abdominal or adnexal signs are occasionally unilateral, and this finding can cause confusion with appendicitis (Chapter 142), ectopic pregnancy, and other conditions. Right upper quadrant abdominal tenderness due to perihepatitis (Fitz-Hugh–Curtis syndrome) is sometimes present, which can mimic acute cholecystitis or viral hepatitis. Perihepatitis sometimes occurs in the absence of other abdominal or pelvic findings typical of PID, especially when it is caused by *C. trachomatis*. Severe PID may be accompanied by signs of generalized peritonitis.

DIAGNOSIS

The clinical diagnosis of PID is imprecise. Published studies, which have used somewhat variable criteria in populations with different prevalences of the syndrome, have reported positive predictive values of 65 to 90% compared with the "gold standard" of laparoscopically defined PID. Although nonspecific, the presence of neutrophils on a saline wet mount of vaginal secretions was 91% sensitive for endometritis in one study. The absence of white blood cells on a wet mount of vaginal secretions, particularly if mucopurulent cervicitis is also absent, should prompt the consideration of alternative diagnoses. The differential diagnosis of PID includes ectopic pregnancy; appendicitis; rupture, bleeding, or torsion of an ovarian cyst; endometriosis; urinary tract infection or pyelonephritis; renal or ureteral stones; inflammatory bowel disease; and, rarely, viral hepatitis or cholecystitis.

Because of the potential seriousness of the infection and the relative simplicity, low cost, and low toxicity of treatment, clinical diagnostic criteria emphasize sensitivity at the cost of specificity. Clinicians therefore should maintain a low threshold for tentative diagnosis and presumptive treatment of PID. The CDC recommends that all sexually active women with pelvic or lower abdominal pain and uterine, adnexal, or cervical motion tenderness be treated for possible PID if no other cause of their symptoms and signs is readily apparent. Furthermore, screening of asymptomatic sexually active young women with NAATs is recommended by the U.S. Preventive Services Task Force.[7] Factors such as fever, an elevated erythrocyte sedimentation rate or C-reactive protein level, and concurrent bacterial vaginosis or mucopurulent cervicitis further support the clinical diagnosis, but these are frequently absent. Definitive diagnosis requires histologic evidence of endometritis on endometrial biopsy; transvaginal ultrasound or other diagnostic imaging showing thickened or fluid-filled fallopian tubes or tubo-ovarian abscess; or laparoscopy demonstrating purulent tubal exudate, erythema, or edema.

PROGNOSIS

Fallopian tube scarring secondary to PID often results in tubal factor infertility (TFI), ectopic pregnancy, and chronic pelvic pain. Previous gonococcal and chlamydial infections are among the most common antecedents of these complications. Each episode of PID, whether due to *N. gonorrhoeae*, *C. trachomatis*, or neither of these organisms, significantly increases the risk for recurrent salpingitis.

The natural history of clinically apparent PID is well defined. In what is probably the best single study on the subject, TFI occurred in 8% of women after a single episode of laparoscopically proved PID, in 20% after two such episodes, and in 40% after three or more episodes. The first pregnancy after an episode of PID was ectopic in almost 8% of women, and like TFI, the risk of ectopic pregnancy increased with each successive episode of the syndrome. Chronic pelvic pain, sometimes disabling in its severity, occurs in almost 20% of women after one or more episodes of PID. Importantly, these outcomes were best studied in women with clinically apparent PID, and the pertinent studies were undertaken at a time when clinical recognition and treatment were likely delayed. Many women suffer clinically mild or silent

PID. The risk of sequelae associated with silent PID is not well defined, but more clinically severe PID is associated with a higher risk of sequelae. Most women with TFI deny a history of PID, and it seems likely that silent PID, particularly silent PID associated with *C. trachomatis*, is responsible for most cases of STI-related reproductive tract sequelae. For example, *C. trachomatis* seropositivity is strongly associated with TFI, independent of clinical or historical evidence of PID.

Rectal Infection

CLINICAL MANIFESTATIONS

Gonococcal infection of the rectum is common in women and in MSM. In women, infection is acquired either through perineal contamination with cervicovaginal secretions or by anal intercourse; the latter is the exclusive route of infection in MSM. In women with cervical gonorrhea and in MSM with gonorrhea at any anatomic site, about 40% have rectal infection. More than 80% of rectal infections are subclinical, but symptomatic proctitis sometimes is manifested as varying combinations of anal pruritus, mucopurulent discharge (often characterized by the patient as mucus-coated feces), pain, tenesmus, and bleeding. Symptomatic proctitis seems to be more common in MSM than in women with rectal gonorrhea, which suggests that the size of the infecting inoculum or trauma from anal intercourse may influence the clinical manifestations. Among MSM, rectal gonorrhea is a potent epidemiologic risk marker for the acquisition of HIV and may be a direct risk factor because anorectal inflammation enhances susceptibility to HIV infection.

DIAGNOSIS

Diagnosis of rectal gonorrhea depends on the identification of *N. gonorrhoeae*, usually by NAAT. The Gram-stained smear is insensitive and nonspecific. The differential diagnosis of symptomatic proctitis includes other traditional STIs (herpes, syphilis, and chlamydial infection, including lymphogranuloma venereum; Chapter 318) as well as ulcerative colitis, Crohn's colitis, anal fissure, rectal lacerations, and proctocolitis caused by *Shigella*, *Campylobacter*, *Yersinia enterocolitica*, and other enteric pathogens. Recent studies using NAAT among STI clinic patients suggest that approximately 5% of tested MSM and 1% of tested women have rectal infections without concurrent genital tract infections. Although less than 20% of women with gonorrhea have extragenital infections, more than half of MSM with gonorrhea have only extragenital infections, highlighting the importance of routine screening of MSM, but not women, for rectal gonorrhea.

Pharyngeal Infection

CLINICAL MANIFESTATIONS

Pharyngeal gonococcal infection results from orogenital exposure. It is more efficiently acquired by fellatio than by cunnilingus and is typically found in 3 to 7% of heterosexual men, 10 to 30% of women, and 10 to 30% of MSM with genital tract gonorrhea. Asymptomatic infection is the rule, although rare cases may exhibit exudative pharyngitis and cervical lymphadenopathy. Isolated pharyngeal infection is common is MSM and may also be common in at least some populations of heterosexuals. Complications are infrequent, and most infections eventually resolve spontaneously or in response to therapy for genital or rectal infection. Although the modest morbidity associated with pharyngeal infections probably does not justify extensive screening efforts, the oropharynx may be an important reservoir for infection in some populations, particularly MSM. The oropharynx is also a site of gene exchange between *N. gonorrhoeae* and commensal Neisseriaceae, and pharyngeal infections are thought to play a critical role in fostering the emergence of antimicrobial-resistant gonococci. Failure to identify and to eradicate pharyngeal infections probably helps sustain high levels of gonococcal transmission and may foster the spread of antibiotic-resistant gonococci. Accordingly, current guidelines suggest that MSM at risk for STIs be tested periodically for pharyngeal gonorrhea.

Gonorrhea in Children

Gonococcal conjunctivitis may develop in infants born to mothers with gonorrhea, a condition termed ophthalmia neonatorum. Formerly a common cause of blindness, gonococcal ophthalmia is now rare in industrialized countries because of both improved control of gonorrhea and routine use of neonatal ocular prophylaxis with topical antibiotics or 1% silver nitrate. Neonates may also acquire pharyngeal or rectal infection and, rarely, gonococcal pneumonia or sepsis. Beyond the neonatal period, purulent vaginitis is the most common manifestation of gonorrhea or chlamydial infection in girls, and rectal or pharyngeal infection is the most common manifestation in prepubertal boys. Most cases are acquired through sexual abuse, but occasional cases in young children may be acquired from fomites or by nonsexual personal exposure in crowded conditions, especially in tropical climates.

Disseminated Gonococcal Infection

CLINICAL MANIFESTATIONS

Disseminated gonococcal infection (DGI) usually is manifested with various combinations of polyarticular tenosynovitis, dermatitis secondary to focal septic embolization, and septic arthritis. Studies undertaken in the 1960s and 1970s estimated that DGI occurred in 1 to 3% of adults with gonorrhea, but the risk depends on characteristics of the particular strains of *N. gonorrhoeae* circulating in the population. Today, DGI probably occurs in well under 1% of gonococcal infections in most geographic areas. Women may be somewhat more susceptible to DGI than men, and the onset often coincides with menstruation. Severity varies from a mild illness with slight joint discomfort, a few skin lesions, and little or no fever to a fulminant illness with overt polyarthritis, innumerable skin lesions, high fever, and prostration. Most persons with DGI have no symptoms of genital gonorrhea, probably because some strains of *N. gonorrhoeae* that are prone to disseminate are also associated with subclinical mucosal infections. The absence of clinical symptoms and signs of mucosal infection in these strains is probably due to their ability to bind complement downregulatory molecules, thereby diminishing the local inflammatory response.

The presentation of DGI can be divided into the clinical syndromes of tenosynovitis-dermatitis and monarticular or oligoarticular arthritis, although these presentations sometimes overlap. Tenosynovitis-dermatitis is thought to predominate early in the course of dissemination, and approximately 70% of DGI cases in published series present with this syndrome. These patients usually suffer migratory polyarthralgias without purulent arthritis. There is often tendon inflammation affecting the wrists, fingers, ankles, or toes. Skin lesions are typically painless, are few in number (5 to 30), affect predominantly the extremities, and are pustular or vesiculopustular (Fig. 299-3), although petechiae, hemorrhagic macules, papules, bullae, and nodules rarely occur (see Fig. 441-5). Axial skeletal involvement is uncommon, a feature that can help differentiate DGI from reactive arthritis (Chapter 265). The tenosynovitis-dermatitis syndrome often subsides spontaneously, or it may evolve during a period of several days into an overt septic arthritis with purulent synovial fluid, usually involving only one or two joints. Approximately 25 to 50% of persons with DGI present initially with purulent monarticular or oligoarticular arthritis, often without apparent sequential evolution from arthritis-dermatitis syndrome. This form of DGI typically affects the knee, ankle, elbow, or wrist, but any joint may be involved.

DIAGNOSIS

Sexually active young persons with arthritis, tenosynovitis, or papulopustular skin lesions should be tested for *N. gonorrhoeae* at all potentially exposed anatomic sites. The diagnosis of DGI is secure when gonococci are identified by culture or NAAT in the blood, a skin lesion, or synovial fluid, but it is often made presumptively when genital, rectal, or pharyngeal gonorrhea is present in a patient with a compatible clinical syndrome that responds promptly to antibiotics.

Blood, synovial fluid, and mucosal tract cultures are positive in approximately 4 to 35%, 10 to 34%, and 80% of patients, respectively. However, the yield from culture varies with the clinical presentation. The performance of NAAT versus culture of blood, synovial fluid, or skin lesions has not been studied in patients with DGI. Patients with tenosynovitis-dermatitis more frequently have bacteremia, whereas patients with septic gonococcal arthritis are seldom bacteremic, and close to 50% have positive synovial fluid cultures; it is reasonable to suspect that the yield might be higher with NAAT. Because bacteremia is intermittent, clinicians should obtain more than one set of blood cultures to maximize the likelihood of isolating the organism. Similarly, NAAT or culture specimens should be obtained from all potentially exposed anatomic sites (genital tract, pharynx, rectum). Cultures of skin lesions are generally negative despite demonstrable gonococci by fluorescent antibody, but NAAT results may be positive in persons with negative blood and mucosal site cultures. The peripheral blood leukocyte count is generally elevated but may be normal. The synovial fluid leukocyte count is usually 20,000 to 60,000/mm^3, with higher numbers of white blood cells seen in persons with clinically apparent arthritis than in those with tenosynovitis-dermatitis. Liver function tests often show elevations in aminotransferase levels, suggestive of mild hepatitis.

The differential diagnosis of DGI includes reactive arthritis (Chapter 265), meningococcemia (Chapter 298), other kinds of septic arthritis (Chapter 272), rheumatoid arthritis (Chapter 264), systemic lupus erythematosus

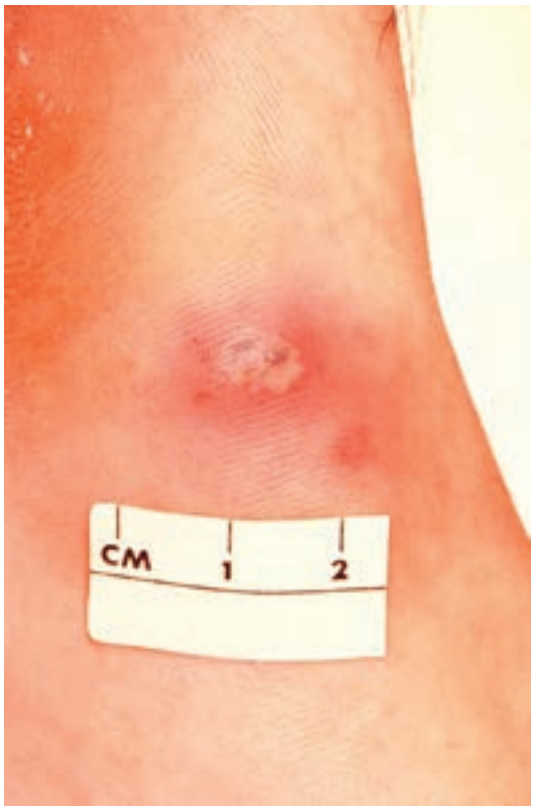

FIGURE 299-3. Cutaneous gonococcal lesion secondary to disseminated *Neisseria gonorrhoeae* infection. Although gonorrhea is a sexually transmitted disease, if it remains untreated, the *N. gonorrhoeae* bacteria responsible for the infection can disseminate throughout the body and form lesions in extragenital locations (also see Fig. 441-5). (From Handsfield HH. *Color Atlas and Synopsis of Sexually Transmitted Diseases*. 3rd ed. New York: McGraw-Hill; 2011.)

(Chapter 266), acute HIV infection (Chapter 385), syphilis (Chapter 319), and other rheumatologic conditions and infectious diseases. Reactive arthritis, often triggered by sexually acquired chlamydial infection, is the principal consideration in young adults. The skin lesions of the two conditions, when present, are generally distinct and often pathognomonic for one syndrome or the other. In addition, conjunctivitis and involvement of the axial skeleton (e.g., sacroiliitis) are common in reactive arthritis and infrequent in DGI.

PROGNOSIS

Many cases of arthritis-dermatitis syndrome resolve spontaneously; with prompt treatment, few patients suffer sequelae of DGI, but untreated septic arthritis can lead to contiguous osteomyelitis or joint destruction. Endocarditis, meningitis, and myocarditis are occasionally seen, with or without a typical DGI syndrome. Gonococcal endocarditis usually involves the aortic valve and often progresses rapidly, leading to valve destruction and heart failure.

PREVENTION

Control of gonorrhea depends on prompt diagnosis and effective treatment of infected persons, screening of sexually active women and MSM in settings where gonorrhea is prevalent, treatment of patients' partners, rescreening of persons with a recent history of gonorrhea, and efforts to promote safer sexual behaviors (e.g., condom use, abstinence, fewer partners). Sexually active MSM outside of mutually monogamous relationships, including HIV-infected men, are at high risk for gonorrhea and should be tested at least annually for rectal and pharyngeal gonococcal infection as well as for chlamydial infection, syphilis, and HIV infection. The value of screening of asymptomatic men for urethral infection (by NAAT of urine) is uncertain; the yield is low in most settings. Gonococcal screening criteria for woman are not well defined. However, because most NAATs for *C. trachomatis* also test for *N. gonorrhoeae*, most chlamydia screening includes gonorrhea testing.

Public education and personal counseling of persons with gonorrhea or at risk of contracting it should emphasize the effectiveness of mutual monogamy and condoms for vaginal or anal sex for new or casual partnerships. Every patient with gonorrhea should be counseled about the risks for HIV infection and should be tested for HIV, *C. trachomatis*, and syphilis. Because accurate epidemiologic data are essential to generate and to maintain resources for the prevention and control of STIs, all cases of gonorrhea, chlamydial infection, syphilis, and HIV infection should be promptly reported to the health department in accordance with local laws. Ultimate control of gonorrhea may require immunization, but no effective vaccine is on the horizon, despite intensive research for more than four decades.

TREATMENT Rx

Antimicrobial Susceptibility

Gonococci with chromosomal or plasmid-borne mutations that confer relative or absolute resistance to the penicillins, tetracyclines, and sulfonamides are prevalent worldwide, and none of these drugs is acceptable as empirical therapy anywhere in the world. The prevalence of β-lactamase (penicillinase) plasmids, which confer absolute resistance to penicillin, ampicillin, and amoxicillin, varies from about 10% of gonococci in the United States and western Europe to almost 50% in some developing countries. Approximately 27% of gonococcal infections in MSM and 9% of infections in heterosexuals are caused by organisms that are resistant to ciprofloxacin and other fluoroquinolones. Resistant gonococci are still more prevalent in Europe (up to 50% in some countries). Fluoroquinolone-resistant gonococci are prevalent worldwide, and fluoroquinolones are no longer suitable for routine use unless recent local data demonstrate low levels of resistance. Approximately 5% of gonococcal isolates in the United Kingdom are no longer susceptible to azithromycin, and azithromycin-resistant *N. gonorrhoeae* has been identified in the western United States, although such strains were rare as recently as 2012.

Gonococcal resistance to cephalosporins is now a major public health concern. Strains with relative resistance to oral cephalosporins (e.g., cefixime, heretofore a mainstay of gonorrhea treatment) are common in Japan, and in 2011 almost 8% of gonococcal isolates in Europe and 1.4% of isolates in the United States had reduced susceptibility to cefixime. In the United States, organisms with diminished susceptibility to oral cephalosporins are more common on the West Coast than in other parts of the country and more common among MSM than among heterosexuals.[8] Diminished susceptibility to ceftriaxone remains very rare, although isolated cases of multidrug-resistant gonorrhea—organisms that are resistant to ceftriaxone therapy—have been identified in Japan and Europe. Although recent data suggest that the proportion of gonococcal infections caused by organisms with diminished susceptibility to oral third-generation cephalosporins is now declining, ongoing surveillance for cephalosporin resistance remains a high priority, and decisions on how to treat gonorrhea need to incorporate strategies for diminishing the development of population-level antimicrobial resistance. Clinicians who treat patients for gonorrhea and other STIs should keep abreast of regional trends in resistance and be alert to modified therapeutic recommendations.[9]

Principles of Treatment

Because of the need to curtail transmission, therapy is usually based on clinical or epidemiologic suspicion before the diagnosis is confirmed microbiologically.[10] Clinicians should presumptively treat all patients evaluated as contacts to persons with known gonococcal infections, all women with PID or mucopurulent cervicitis, and, if Gram staining cannot be performed, all men presenting with a clinical syndrome of urethritis. (Men with urethritis and no evidence of gonorrhea on Gram stain should be treated for nongonococcal urethritis.) Diagnosis by NAAT precludes antimicrobial resistance testing, and even when *N. gonorrhoeae* is isolated by culture, susceptibility testing is rarely performed routinely. Accordingly, treatment of uncomplicated gonorrhea is dictated by national or local patterns of antimicrobial susceptibility, without knowledge of susceptibility in individual patients. However, susceptibility testing should be used to guide the treatment of gonococcal septic arthritis, endocarditis, or other serious complications and may increasingly become routine even in genitourinary infections.[11]

Dual treatment of uncomplicated gonorrhea, typically with both a cephalosporin and either azithromycin or doxycycline, has long been recommended to cover the possibility of simultaneous chlamydial infection, typically present in 5 to 10% of MSM, 15 to 25% of heterosexual men, and 35 to 50% of women with gonorrhea in North America. In addition, dual therapy using antibiotics with differing mechanisms of action may reduce selection pressure for antimicrobial resistance in *N. gonorrhoeae*, so that dual therapy is now recommended routinely even when chlamydial infection is unlikely or has been excluded by negative testing.

Treatment Regimens

Ceftriaxone 250 mg intramuscularly plus azithromycin 1 g orally is the preferred treatment regimen for uncomplicated gonorrhea (Table 299-1).[12] This regimen is highly effective against pharyngeal gonorrhea, which may be relatively resistant to oral therapies, and is designed to diminish the spread of cephalosporin-resistant *N. gonorrhoeae* as well as to cover simultaneous chlamydial infection. Cefixime 400 mg orally (with azithromycin 1.0 g orally)

TABLE 299-1 ANTIBIOTIC REGIMENS FOR THE TREATMENT OF GONORRHEA IN THE UNITED STATES*

UNCOMPLICATED GONORRHEA OF THE URETHRA, CERVIX, OR RECTUM
Preferred
Ceftriaxone 250 mg IM single dose *plus* azithromycin 1 g PO
Alternatives
Cefixime 400 mg PO single dose *plus* azithromycin 1 g PO *or* Azithromycin 2 g PO single dose *plus* gentamicin 240 mg IM single dose *or* gemifloxacin 320 mg PO as a single dose
INFECTION OF THE PHARYNX
Ceftriaxone 250 mg IM single dose *plus* azithromycin 1 g PO as a single dose
CONJUNCTIVITIS (NOT OPHTHALMIA NEONATORUM)
Ceftriaxone 1 g IM single dose *plus* azithromycin 1 g PO as a single dose
DISSEMINATED GONOCOCCAL INFECTION
Preferred
Ceftriaxone 1 g IM or IV q24h† *or* Ceftizoxime 1 g IV q8h *plus* Azithromycin 1 g PO as a single dose
Alternative
Cefotaxime 1 g IV q8h

*Treatment of gonorrhea in adults should always include treatment of sex partners and advice to abstain from sex for 7 days.
†Treat with intravenous therapy until the patient has been clinically improved for 24 to 48 hours. Then switch to cefixime 400 mg orally twice a day to complete a 7-day course.

remains appropriate when ceftriaxone therapy is not feasible (e.g., if a patient declines injection) or in expedited (unobserved) treatment of patients' sex partners. Clinicians should be mindful that ensuring that all patients and partners are treated is more important than always using intramuscular ceftriaxone. Persons with severe β-lactam allergies should receive azithromycin 2 g orally once, *plus* either a single dose of gentamicin 240 mg IM *or* a single dose of gemifloxacin 320 mg orally.

Women with acute PID should be treated with antibiotics active against *N. gonorrhoeae* and *C. trachomatis*.[13] The role of anaerobic bacteria in PID is uncertain, although it is not known whether anaerobic treatment is required for PID. Most women can be treated as outpatients, but the following factors should prompt hospital admission: possible surgical cause of symptoms (e.g., appendicitis), pregnancy, failure to respond to oral therapy within 72 hours of initiation, inability to tolerate or to adhere to oral therapy, severe symptoms, and tubo-ovarian abscess. The suggested outpatient regimen is ceftriaxone 250 mg IM plus doxycycline 100 mg PO twice daily for 14 days or cefoxitin 2 g IM plus probenecid 1 g PO as a single dose and doxycycline 100 mg twice daily for 14 days. Either regimen can be given with or without metronidazole (500 mg twice daily) for 14 days. For hospitalized patients or others who require parenteral therapy, the CDC recommends intravenous cefotetan or cefoxitin plus oral doxycycline or parenteral therapy with clindamycin plus gentamicin. Intravenous ampicillin-sulbactam plus oral doxycycline can also be used. Parenteral therapy is continued until improvement is observed, after which oral therapy is prescribed to complete 14 days' total treatment.

Most persons with DGI should be hospitalized and treated with a parenteral third-generation cephalosporin such as ceftriaxone, cefotaxime, or ceftizoxime. Joint irrigation or drainage appears to be unnecessary for septic arthritis, although repeated aspiration of synovial fluid may speed clinical improvement. Oral treatment (e.g., cefixime, cefpodoxime, or a fluoroquinolone) can usually be substituted after improvement begins and then continued to complete 7 days' therapy. More prolonged parenteral treatment and higher doses are indicated for the treatment of gonococcal meningitis or endocarditis, although modern data are lacking. Gonococcal epididymitis, bartholinitis, and other localized complications should generally be treated for 7 to 14 days with drugs active against both *N. gonorrhoeae* and *C. trachomatis*. Gonococcal conjunctivitis in adults can be managed with a single dose of ceftriaxone, 1 g IM, with optional saline lavage. The diagnosis of all forms of complicated gonococcal infection should be confirmed by culture with determination of antimicrobial susceptibility, which can guide completion of treatment after initial empirical therapy.

Management of Sex Partners

Failure to ensure treatment of patients' sex partners contributes to the continued transmission of gonorrhea and other bacterial STIs and often results in reinfection of the index case. For gonorrhea and chlamydial infection, all partners in the preceding 2 months should be treated; if the patient has not had sex in the preceding 2 months, the most recent partner should be treated. Very few U.S. health departments routinely make an effort to ensure that sex partners receive treatment, and the responsibility to ensure partner treatment lies jointly with the patient and the diagnosing clinician. Optimally, the partners of persons with gonorrhea should undergo diagnostic testing for gonorrhea, chlamydial infection, syphilis, and HIV infection. However, clinicians should not wait to obtain the results of diagnostic testing before treating a potentially exposed sex partner; all partners of infected persons should be treated when they initially present for evaluation.

In most settings, approximately 50% of potentially exposed sex partners go untreated, risking ongoing transmission and reinfection of the original patient. In an effort to address this problem, the CDC and several state health departments recommend that clinicians offer heterosexual patients with gonorrhea or chlamydial infection medication to give to their sex partners. This practice, termed patient-delivered partner therapy (PDPT)—sometimes termed expedited partner therapy—is supported by three randomized controlled trials showing that treating sex partners without requiring attendance for care in person decreases the risk of reinfection and increases the proportion of partners treated. PDPT requires single-dose treatment, typically with cefixime plus azithromycin for gonorrhea or azithromycin alone for the partners of patients with chlamydial infection. PDPT currently is legal in most states of the United States and is likely to become permissible in other states, and clinicians should offer PDPT as an option for most heterosexual patients with gonorrhea or chlamydial infection. The CDC website maintains up-to-date information on the legality of PDPT in U.S. states and territories (http://www.cdc.gov/STD/ept/legal/default.htm). In addition to direct delivery of drugs to the patient for PDPT, it is often feasible to write or to telephone a prescription to a cooperating pharmacy. However PDPT is implemented, when practical, clinicians should provide written information about the medication and STI prevention as well as advice to seek clinical care in addition to taking the drugs provided. Examples of forms that can be dispensed with PDPT can be found at the following websites: http://www.doh.wa.gov/YouandYourFamily/IllnessandDisease/SexuallyTransmittedDisease/ExpeditedPartnerTherapy; http://www.cdph.ca.gov/HealthInfo/discond/Pages/SexuallyTransmittedDiseases.aspx. PDPT is not recommended for MSM with gonorrhea or chlamydial infection owing to potentially high rates of syphilis and HIV infection; the partners of such patients should be evaluated and treated in person.

Follow-up

The recommended treatment regimens cure 96 to 100% of uncomplicated cases of genital or rectal gonorrhea caused by susceptible strains and at least 90% of pharyngeal infections. Retesting of infected patients to document eradication of *N. gonorrhoeae* ("test of cure") is not recommended except in pregnant women, when adherence to therapy is in doubt, or when atypical treatment regimens are employed. When test of cure is indicated, culture can be performed a week after treatment is completed, but retesting by NAAT should be delayed at least 3 weeks after treatment to reduce the possibility of detecting persistent gonococcal DNA despite the eradication of viable organisms.

Although test of cure is generally not advised, all persons diagnosed with gonorrhea should be rescreened 3 to 4 months after treatment. In prospective studies, 10 to 20% of both men and women with gonorrhea or chlamydial infection are infected when they are retested 3 to 4 months later. The likelihood of recurrent or persistent gonococcal infection appears to be reduced by up to 70% when partners are managed with PDPT. Rescreening can be done with urine or self-obtained vaginal swab testing by NAAT and does not require a second visit with a clinician.

Grade A References

A1. Creighton S. Gonorrhoea. *Clin Evid (Online)*. 2014;2014.
A2. Kirkcaldy RD, Weinstock HS, Moore PC, et al. The efficacy and safety of gentamicin plus azithromycin and gemifloxacin plus azithromycin as treatment of uncomplicated gonorrhea. *Clin Infect Dis*. 2014;59:1083-1091.
A3. Ferreira A, Young T, Mathews C, et al. Strategies for partner notification for sexually transmitted infections, including HIV. *Cochrane Database Syst Rev*. 2013;10:CD002843.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2012*. Atlanta: U.S. Department of Health and Human Services; 2013.
2. Torrone EA, Johnson RE, Tian LH, et al. Prevalence of *Neisseria gonorrhoeae* among persons 14 to 39 years of age, United States, 1999 to 2008. *Sex Transm Dis*. 2013;40:202-205.
3. Ison CA, Town K, Obi C, et al. Decreased susceptibility to cephalosporins among gonococci: data from the Gonococcal Resistance to Antimicrobials Surveillance Programme (GRASP) in England and Wales, 2007-2011. *Lancet Infect Dis*. 2013;13:762-768.
4. Allen VG, Mitterni L, Seah C, et al. *Neisseria gonorrhoeae* treatment failure and susceptibility to cefixime in Toronto, Canada. *JAMA*. 2013;309:163-170.
5. Centers for Disease Control and Prevention. Recommendations for the laboratory-based detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*—2014. *MMWR Recomm Rep*. 2014;63:1-19.
6. Mitchell C, Prabhu M. Pelvic inflammatory disease: current concepts in pathogenesis, diagnosis and treatment. *Infect Dis Clin North Am*. 2013;27:793-809.
7. Zakher B, Cantor AG, Pappas M, et al. Screening for gonorrhea and chlamydia: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014;161:884-893.
8. Kirkealdy RD, Bolan GA, Wasserheit JN. Cephalosporin-resistant gonorrhea in North America. *JAMA*. 2013;309:185-187.
9. Barbee LA. Preparing for an era of untreatable gonorrhea. *Curr Opin Infect Dis*. 2014;27:282-287.
10. Bignell C, Unemo M. 2012 European guideline on the diagnosis and treatment of gonorrhoea in adults. *Int J STD AIDS*. 2013;24:85-92.
11. Buono SA, Watson TD, Borenstein LA, et al. Stemming the tide of drug-resistant *Neisseria gonorrhoeae*: the need for an individualized approach to treatment. *J Antimicrob Chemother*. 2015;70:374-381.
12. Barbee LA, Kerani RP, Dombrowski JC, et al. A retrospective comparative study of 2-drug oral and intramuscular cephalosporin treatment regimens for pharyngeal gonorrhea. *Clin Infect Dis*. 2013;56:1539-1545.
13. Judlin P. Current concepts in managing pelvic inflammatory disease. *Curr Opin Infect Dis*. 2010;23:83-87.

REVIEW QUESTIONS

1. A 26-year-old man comes to your office complaining of pain with urination and a discharge from his penis. He reports having both insertive and receptive anal sex with two men in the last 2 months but says that he always uses condoms when having anal sex. He also reports both giving and receiving oral sex. The patient reports testing negative for HIV infection approximately 6 months ago. He thinks he had chlamydia several years ago but otherwise denies any prior history of sexually transmitted infection. Physical examination is notable for purulent urethral discharge. Your office does not have a microscope to allow you to perform a Gram stain. Which of the following would be part of the recommended management of this patient?

A. Ceftriaxone 250 mg IM plus azithromycin 1 g PO as a single dose
B. Test for HIV infection and syphilis.
C. Test for gonorrhea (rectal, pharyngeal, and urethral gonorrhea) and chlamydial infection (urethral, rectal).
D. Notify the health department of the patient's infection if he has laboratory confirmation of infection.
E. All of the above

Answer: E This patient presents with typical signs and symptoms of urethritis. Although a urethral Gram stain would be useful in determining whether this patient has gonorrhea, most clinicians in the United States cannot perform Gram stains in their office, and the patient should be empirically treated for both gonorrhea and chlamydial infection. Ceftriaxone is more active against *Neisseria gonorrhoeae* than are third-generation oral cephalosporins like cefixime and should be used in this patient. The addition of azithromycin is designed both to treat a possible concurrent chlamydial infection and as a strategy to diminish the risk of selecting for antimicrobial resistance. Of note, this patient denies having unprotected anal sex, and his only clear urethral exposure is through oral sex. Although his history may be inaccurate, an ill-defined but probably sizable proportion of all urethral gonorrhea in men who have sex with men (MSM) is likely to be acquired through oral sex, highlighting the importance of screening for pharyngeal gonorrhea among MSM. In contrast, *Chlamydia trachomatis* infrequently infects the pharynx and probably is rarely transmitted by fellatio, and no readily available tests are approved for pharyngeal chlamydial infection.

2. A 19-year-old woman comes to your office requesting contraception. She recently started a new relationship with a man and sometimes uses condoms. Before her current sex partner, she reports not having sex for about 6 months. As part of her care, you perform a routine screening test for gonorrhea and chlamydial infection on a self-obtained vaginal swab using Aptima Combo 2, a nucleic acid amplification test. The test result comes back positive for both gonorrhea and chlamydial infection. You call the patient to give her the results and arrange her follow-up. She continues to be asymptomatic. Which of the following would be part of your initial management of this patient?

A. Have the patient return to your office for treatment with ceftriaxone 250 mg IM plus azithromycin 1 g PO as a single dose.
B. Arrange for the patient to return to your office for follow-up testing in 3 months.
C. Ask the patient to bring her partner to your office with her so that he can be treated.
D. Talk to the patient and her partner about condoms and risk reduction when she returns to see you, and advise them to abstain from sex for 1 week after the completion of both of their treatment.
E. All of the above

Answer: E A regimen of ceftriaxone and azithromycin is effective for both gonorrhea and chlamydial infection. Although a test of cure is not indicated among nonpregnant women with gonorrhea or chlamydial infection, repeated testing 3 months after treatment is recommended because of the high prevalence (approximately 12%) of persistent or recurrent gonorrhea or chlamydial infection. Failure to ensure treatment of sex partners is thought to be one major cause of recurrent infection, and the management of sexually transmitted infections should include ensuring the treatment of patients' sex partners. Although gonorrhea and chlamydia are both reportable infections in all U.S. states, most U.S. health departments do not make any routine effort to contact people diagnosed with gonorrhea or chlamydial infection, and clinicians should regard partner treatment as part of their responsibility when treating patients. The best course of action here is to ask the patient to bring her partner to your office for treatment when she comes for her ceftriaxone injection. If she does not bring the partner when she returns, you should offer her patient-delivered partner therapy (PDPT) with cefixime 400 mg and azithromycin 1 g at that time. PDPT is now legal in most parts of the United States and is recommended as a routine option in the Centers for Disease Control and Prevention (CDC) treatment guidelines for sexually transmitted diseases. PDPT should be dispensed with written information about sexually transmitted infections and the drugs being provided. Helpful links for additional resources related to PDPT are available:

CDC website on expedited partner therapy, including information about the legal status of PDPT in different states: http://www.cdc.gov/STD/ept/legal/default.htm

Washington State Department of Health EPT website: http://www.doh.wa.gov/YouandYourFamily/IllnessandDisease/SexuallyTransmittedDisease/ExpeditedPartnerTherapy.aspx. Includes information sheets for patients.

3. A 30-year-old man who you have seen several times during the last few years comes to your office complaining of a urethral discharge and dysuria for 3 days. The patient only has sex with other men and reports two sex partners in the last 2 months; he uses condoms inconsistently. He has a purulent discharge on physical examination. You treat him with ceftriaxone and azithromycin, both of which he receives while in your office. He has positive nucleic acid amplification test (NAAT) results demonstrating both urethral and pharyngeal gonorrhea. His test results for *Chlamydia trachomatis*, syphilis, and HIV infection are all negative. He returns to see you 10 days later because he still has discomfort when he urinates. He reports that the discharge from his penis has improved but not resolved. On examination, his discharge is scant and cloudy, not overtly purulent. He denies having sex, including oral sex, since his treatment. A sample of the initial 30 mL of voided urine is leukocyte esterase positive. How do you manage this patient?

A. Treat him again with ceftriaxone 250 mg IM and azithromycin 1 g.
B. Perform a urethral culture and NAAT for gonorrhea.
C. Treat him with moxifloxacin 400 mg PO for 7 to 14 days.
D. Treat him with metronidazole 2 g PO once.
E. B and C

Answer: E Patients presenting with persistent symptoms of urethritis should have a Gram stain or, if that is not available, leukocyte esterase testing of their urine to identify objective evidence of inflammation. This patient has a positive leukocyte esterase test result and a visible discharge, establishing a diagnosis of persistent or recurrent urethritis. The differential diagnosis here includes reinfection, treatment failure related to gonorrhea, and persistent urethritis due to a cause other than gonorrhea. Although reinfection is always a possibility in patients whose symptoms do not completely resolve, this patient denies having sex since his treatment. Treatment failure is also possible. Treatment failure related to gonococcal resistance is now an important concern, and treatment failures have been observed, particularly among persons treated with an oral cephalosporin (i.e., cefixime) without concurrent azithromycin therapy. However, treatment failure with ceftriaxone, although conceivable, remains unlikely, given the current epidemiology of gonococcal resistance in the United States, and testing for persistent gonorrhea with both culture and a NAAT is a reasonable course of management for this patient. A positive culture for *Neisseria gonorrhoeae* would establish the diagnosis of gonorrhea, whereas a negative NAAT result would rule that diagnosis out. A positive NAAT result could reflect either persistent infection or simply the continued presence of RNA or DNA, which can persist in the absence of active infection for 7 to 21 days. Thus, a positive NAAT result with a negative culture would not establish a definitive diagnosis. Persons with possible gonorrhea treatment failures should have antimicrobial susceptibility testing performed on a cultured specimen and receive a second course of ceftriaxone 250 mg IM and azithromycin 1 g while awaiting antimicrobial susceptibility test results. The most likely diagnosis in this patient is persistent urethritis due to a pathogen that was not treated with the initial antimicrobial regimen. The most common identifiable cause of persistent urethritis in MSM is *Mycoplasma genitalium*. No Food and Drug Administration–approved commercial test is widely available for *M. genitalium*. Neither azithromycin nor doxycycline is consistently effective against this pathogen, and β-lactam antibiotics do not have activity against the organism. The best therapy for persistent urethritis in MSM is moxifloxacin 400 mg PO qd for 7 to 14 days. The CDC recommends 7 days, but treatment failures have been reported with that regimen, and longer courses may be superior. Clinicians should consider adding metronidazole 2 g PO in a single dose to the therapy for persistent urethritis in heterosexuals to treat *Trichomonas vaginalis*; *Trichomonas* is very uncommon in MSM, like the patient described here.

4. A 21-year-old women comes to your office complaining of mild low abdominal pain and vaginal discharge for 2 weeks. She denies fever, chills, nausea, vomiting, or diarrhea and has no significant past medical history. She has had an intrauterine device (IUD) in place for 2 years. She has been monogamous with a new male partner for 2 months; they used condoms for vaginal sex for 1 to 2 weeks but not since then. Her most recent sexual contact before starting her current relationship was 7 to 8 months ago. On physical examination, the patient is afebrile and well appearing. She has mild bilateral lower abdominal tenderness without rebound tenderness; her bowel sounds are normal. Her pelvic examination is notable for a yellow discharge coming out of the cervical os, cervical motion tenderness, and mild bilateral adnexal tenderness without palpable masses. How do you manage this patient?

A. Test her for gonorrhea and chlamydial infection and treat her with ceftriaxone 250 mg IM once, doxycycline 100 mg PO bid for 14 days, and metronidazole 500 mg PO bid for 14 days.
B. Arrange to see the patient again or speak to her by telephone in 48 to 72 hours to ensure that her symptoms are improving.
C. Admit her to the hospital for parenteral therapy with IV cefoxitin and oral or IV doxycycline.
D. Arrange to have her IUD removed.
E. A and B

Answer: E This patient has signs and symptoms consistent with mucopurulent cervicitis and mild pelvic inflammatory disease (PID). *Chlamydia trachomatis* and *Neisseria gonorrhoeae* are the main sexually transmitted pathogens, but neither can be identified in half or more of women with PID. PID is often considered to be a polymicrobial infection, but the role of anaerobic bacteria and the importance of treating patients with mild or moderate PID (e.g., without overt pyosalpinx or pelvic abscess) for anaerobes is controversial. Therefore, the addition of metronidazole to the patient's treatment regimen is optional. This patient does not require hospital admission but should have follow-up in the 48 to 72 hours after treatment to ensure that she is improving. An IUD called the Dalkon Shield was associated with PID many years ago, and IUD insertion is a risk factor for PID in women with gonorrhea or chlamydia at the time of device insertion. However, the diagnosis of PID does not mandate IUD removal in women with an IUD in place. At the same time, few data are available on the risk of recurrent PID in women using IUDs, and the presence of an IUD in this patient is an additional reason for close medical follow-up.

5. A 29-year-old man comes to your clinic with pain and swelling affecting his right knee. He returned from a 3-week trip to Thailand, Vietnam, and Cambodia 4 days ago. He was well throughout most of his trip and took malaria prophylaxis while traveling. He did visit some rural areas. He does not inject drugs and denies having a rash. He did have unprotected vaginal sex with two sex workers during his trip. About 1 week ago, he developed pain in his wrists, fingers, and ankles. He does not recall fever with this. This pain resolved, but he subsequently noted pain and swelling of his knee, which has affected his ability to walk. On physical examination, he is afebrile. His right knee is swollen with an effusion. His other joints are unremarkable. He has no rash. You perform an arthrocentesis, which reveals 30,000 white blood cells (90% polymorphonuclear neutrophils) with no crystals and no organism seen on Gram stain. Synovial fluid culture is pending. Which of the following is *not* indicated?

A. Call orthopedics to perform open drainage of the knee.
B. Obtain culture specimens and nucleic acid amplification test (NAAT) specimens for gonorrhea from all anatomic sites of potential sexual exposure to gonorrhea and a synovial fluid sample for NAAT.
C. Obtain blood culture samples.
D. Initiate therapy with ceftriaxone and vancomycin while awaiting additional laboratory information.
E. Test the patient for HIV infection and syphilis.

Answer: A This patient presents with a syndrome consistent with disseminated gonococcal infection (DGI). DGI is a rare complication of gonorrhea, sometimes occurring in persons with late complement component deficiencies or as a result of infections caused by gonococci that have the ability to cleave complement. DGI can be manifested as a tenosynovitis-dermatitis syndrome, a monarticular or oligoarticular arthritis, or a combination of both; 25 to 50% of patients will have a purulent arthritis, and synovial fluid cultures for gonorrhea are positive in approximately half of persons with frank arthritis. Most patients with DGI lack genital discharge or other typical symptoms of sexually transmitted infection because the strains of *N. gonorrhoeae* prone to disseminate tend to cause asymptomatic localized infections. Nevertheless, approximately 80% have gonorrhea identifiable in the genital tract, rectum, or pharynx, and all patients with suspected DGI should be tested by NAAT at all recently exposed anatomic sites. This patient's history of a migratory tenosynovitis, recent high-risk sexual exposure, and arthritis with a relatively low synovial fluid white blood cell count are all consistent with the septic joint stage of DGI, apparently preceded by the tenosynovitis-dermatitis syndrome. The most prudent course of action in this patient is to initiate therapy for both DGI and other common causes of septic arthritis, especially possible methicillin-resistant *Staphylococcus aureus*. The vancomycin can be discontinued if gonorrhea is identified in any culture or NAAT assay. Arthritis associated with DGI usually responds rapidly to antimicrobial therapy, and patients who enjoy rapid clinical improvement often receive only 2 days of intravenous ceftriaxone followed by oral cefixime 400 mg twice daily to complete a week of therapy. Septic gonococcal arthritis usually responds completely to antibiotics and does not require surgical drainage.

300 *HAEMOPHILUS* AND *MORAXELLA* INFECTIONS

MICHAEL S. SIMBERKOFF

HAEMOPHILUS INFECTIONS

DEFINITION

The name *Haemophilus* is derived from the Greek nouns *haima*, meaning "blood," and *philos*, meaning "lover." *Haemophilus* species colonize the respiratory tract; cause infections of the respiratory tract, skin, or mucous membranes of humans; and from these sites can invade and cause bacteremia, meningitis, epiglottitis, endocarditis, septic arthritis, or cellulitis (Table 300-1).

The Pathogen

Haemophilus species are small, nonmotile, aerobic or facultative anaerobic, pleomorphic, gram-negative bacilli. The prototype of this genus, *Haemophilus influenzae*, was originally recovered from patients with influenza by Pfeiffer in 1893, and it was considered the cause of that disease for many years. Primary isolation of *Haemophilus* species is best accomplished on chocolate agar medium in a carbon dioxide–enriched atmosphere.

EPIDEMIOLOGY

The precise prevalence and incidence of *H. influenzae* infections are unknown.[1] This organism can be detected in the nasopharynx of both children and adults. Before the introduction of an effective vaccine, between 3 and 5% of infants harbored *H. influenzae* type b in their nasopharynx. Children who have been immunized against *H. influenzae* type b are far less likely to be colonized and infected with this organism. However, the risk for infection in nonimmune household contacts of a patient with invasive *H. influenzae* disease is approximately 600-fold greater than the risk in the age-adjusted general population. Nontypeable *H. influenzae* can be detected in the nasopharyngeal cultures of more than 70% of young children, but infection occurs in only a small proportion of colonized individuals.

H. influenzae type b was the most common cause of meningitis in young children before effective vaccines were introduced in the 1980s. Vaccination has had a dramatic impact on the incidence of infection in this group. In a population-based study in Atlanta during the 18-month period from December 1, 1988, through May 31, 1990, invasive *H. influenzae* disease occurred in only 5.6 per 100,000 children and 1.7 per 100,000 adults. Forty of the 47 strains associated with invasive disease from adult patients in this study were serotyped. Twenty of these isolates (50%) were *H. influenzae* type b, 19 (47.5%) were nontypeable, and 1 (2.5%) was type f. A more recent study covering the period from 1989 to 2008 from the Centers for Disease Control and Prevention's Active Bacterial Core surveillance system showed that the overall incidence of invasive *H. influenzae* infections fell from 4.39 cases per 100,000 population in 1989 (shortly after introduction of the *H. influenzae* type b conjugate vaccine in the United States) to 1.55 cases per 100,000 in 2008, and the percentage of invasive infections caused by *H. influenzae* type b fell from 87 to 3%, whereas the percentage caused by nontypeable *H. influenzae* strains increased from 16.8 to 68.4%.[2]

Patients with human immunodeficiency virus (HIV) infection are at increased risk for *H. influenzae* infection. Rates of invasive *H. influenzae* infection in men aged 20 to 49 years with HIV infection and acquired immunodeficiency syndrome (AIDS) were 14.6 and 79.2 per 100,000, respectively. The majority of these infections were caused by nontypeable *H. influenzae* strains, although in a second study, 10 of 15 bacteremic *H. influenzae* type b infections observed in adults occurred in patients at risk for HIV infection, and AIDS was documented in seven of these patients.

Other factors also increase the risk for *H. influenzae* infection, including immunoglobulin deficiencies, sickle cell disease, splenectomy, malignant disease, pregnancy, cerebrospinal fluid (CSF) leaks, head trauma, alcoholism, chronic obstructive pulmonary disease (COPD), and race. Eskimo, Navajo, and Apache children have *H. influenzae* type b infection rates that are significantly greater than those in comparable non-native populations. In addition, daycare attendance, crowding, the presence of siblings, prior hospitalizations, and previous otitis media have been shown to increase the risk for *H. influenzae* type b disease in young children, whereas breast-feeding decreases this risk.

PATHOBIOLOGY

H. influenzae consists of encapsulated (typeable) and nonencapsulated (nontypeable) forms. The encapsulated forms are responsible for most of the invasive infections in children and acute epiglottitis in both children and adults, whereas the nonencapsulated forms cause respiratory mucosal infections, conjunctivitis, female genital tract infections, and invasive disease in adults. The capsules of *H. influenzae* consist of polysaccharide antigens. Six capsular serotypes (a through f) exist and are important virulence factors that inhibit opsonization, clearance, and intracellular killing of the organisms. *H. influenzae* type b, formerly the most common cause of meningitis in infancy and childhood worldwide, contains a pentose capsular polysaccharide consisting of polyribosyl ribitol phosphate (PRP). Other serotypes contain hexose polysaccharides. *H. influenzae* type b is more virulent than the other serotypes, probably because it is highly resistant to clearance once bacteremia has been initiated. Since the introduction of *H. influenzae* serotype b conjugate vaccines in the 1990s, most infections are now caused by non-b serotypes, and *H. influenzae* serotype a has emerged as a cause of serious morbidity and mortality.[3] Recent findings suggest that *H. influenzae* strains have the capacity to induce variations in their surface oligosaccharides when serially passed in the presence of human antibody and complement. It has been hypothesized that this may contribute to the organism's capacity to colonize and to survive in nares of patients before causing respiratory tract infections including otitis, sinusitis, epiglottitis, bronchitis, and pneumonia.

Fimbriae are important virulence factors that enhance the adherence of *H. influenzae* to mucosal surfaces. Both typeable and nontypeable *H. influenzae* isolates contain fimbriae. The lipo-oligosaccharides of *H. influenzae* also contribute to their virulence. Lipo-oligosaccharides appear to play a crucial role in facilitating the survival of *H. influenzae* on mucosal surfaces within the nasopharynx and in initiating invasive disease (blood stream invasion) from these sites.[4]

Outer membrane proteins also serve as virulence factors in *H. influenzae* disease. At least 15 different *H. influenzae* outer membrane proteins have been identified. One of these (P2, 39 to 40 kD) functions as a porin, and others are associated with iron binding. Successful scavenging of iron within the human host is crucial for multiplication of *H. influenzae*.

Antibodies have been recognized for decades as an important part of host defenses against *H. influenzae* diseases. The classic studies of Fothergill and Wright in 1933 demonstrated that most cases of *H. influenzae* meningitis occur in young children after they lose passively acquired maternal antibodies and before active humoral immunity to the organism develops. These protective antibodies function primarily to opsonize and to facilitate the clearance of *H. influenzae* rather than to kill virulent organisms directly.

Complement is also an essential component of host defenses against some *H. influenzae* diseases. Children with congenital deficiencies of C2, C3, and factor I have an increased incidence of *H. influenzae* infections. Patients who lack a functional spleen (e.g., those with sickle cell disease) or who have undergone splenectomy also are at risk for the development of overwhelming infection with *H. influenzae* type b.

TABLE 300-1 SITES OF COLONIZATION AND INFECTIONS BY *HAEMOPHILUS INFLUENZAE*

SPECIES	NORMAL FLORA	ASSOCIATED DISEASES
H. influenzae	Nasopharynx Upper respiratory tract	Meningitis Epiglottitis Sinusitis Otitis Pneumonia Cellulitis Arthritis Osteomyelitis Obstetric infections Endocarditis
H. influenzae, biogroup *aegyptius*		Purulent conjunctivitis Brazilian purpuric fever

CLINICAL MANIFESTATIONS

Meningitis

H. influenzae meningitis commonly occurs in children younger than 5 years and in adults with a history of skull trauma or CSF leaks. *H. influenzae* type b strains cause most of these cases. A review of 493 episodes of acute bacterial meningitis in adults at the Massachusetts General Hospital during the 27-year period from 1962 through 1988 showed that 19 cases (4%) were due to *H. influenzae*.

H. influenzae meningitis is clinically indistinguishable from other forms of acute bacterial meningitis. Most patients with *H. influenzae* meningitis have CSF white blood cell counts greater than 1000/mm^3 and hypoglycorrhachia. CSF Gram stain shows pleomorphic gram-negative bacilli in 60 to 70% of untreated cases. In some patients, however, the bipolar staining may result in a mistaken diagnosis of pneumococcal meningitis. Thus, Gram stain is neither sensitive nor specific for diagnosis of *H. influenzae* meningitis.

A diagnosis of *H. influenzae* type b meningitis can be rapidly and reliably established by detecting PRP capsular antigens in CSF. The diagnosis can be established in most cases even when antibiotics have been given before CSF is obtained. Other serotypes (most commonly type f) can also cause meningitis in adults. Therefore, serologic tests for type b antigen in CSF cannot be relied on to rule out *H. influenzae* meningitis in all cases.

Epiglottitis

H. influenzae type b is the most common cause of acute epiglottitis in both children and adults. Epiglottitis is a life-threatening infection in children that usually occurs in those younger than 5 years. The symptoms are fever, drooling, dysphagia, and respiratory distress or stridor, which appear during the course of hours. In adults, fever, sore throat, dysphagia, and odynophagia occur. Cervical tenderness and lymphadenopathy can be found at all ages. Laryngoscopy demonstrates a swollen, cherry-red epiglottis. However, this procedure should be avoided or undertaken only by experts because it may precipitate acute airway obstruction and thus make emergency tracheotomy necessary. A lateral radiograph of the neck more safely confirms the diagnosis of acute epiglottitis. The patient must be maintained in an upright position during this procedure, however, to avoid additional compromise of the airway. The cause is usually established by blood culture. Cultures of the pharynx and other mucosal surfaces are less useful because *H. influenzae* may be part of the normal flora. One review suggests that although vaccination has effectively reduced the incidence of this disease in children, it may be increasing in adults.

Pneumonia

H. influenzae is a common cause of pneumonia in both children and adults. These organisms can also cause nosocomial infections, including ventilator-associated pneumonia. The clinical features of *H. influenzae* pneumonia include fever, cough, and signs and radiographic findings of lobar consolidation.[5] Parapneumonic effusions or empyema occur commonly in patients with *H. influenzae* pneumonia. Gram-negative bacilli in sputum suggest the diagnosis, but isolation of *H. influenzae* from sputum culture alone is inadequate to prove a cause because of the high frequency with which this organism colonizes the respiratory tract. A diagnosis can be established by isolating *H. influenzae* from either blood or pleural fluid. Most isolates are nontypeable.

Tracheobronchitis

Tracheobronchitis is a condition characterized by fever, cough, and purulent sputum that occurs in the absence of radiographic infiltrates suggestive of pneumonia.[6] It frequently develops in patients with known chronic lung disease. Blood cultures are rarely positive. A combination of pleomorphic gram-negative bacilli predominating in purulent sputum, antibody titers to *H. influenzae* that rise after infection, and a response, at least transiently, to treatment of *H. influenzae* infection strongly suggests this diagnosis.

Sinusitis

H. influenzae and *Streptococcus pneumoniae* are the most frequent bacterial isolates from antral punctures or surgical specimens of patients with acute purulent sinusitis. Most *H. influenzae* isolates are nontypeable. Although patients may respond initially to treatment directed against *H. influenzae*, the response is transient if the sinus obstruction is not relieved. *H. influenzae* is not an important pathogen in patients with chronic sinusitis.

Otitis Media

H. influenzae is the most frequent cause of otitis media in young children. Approximately 90% of the *H. influenzae* isolates obtained by tympanocentesis are nontypeable; *H. influenzae* type b causes most of the remaining 10% of infections. Patients with otitis media may have ear pain or exhibit irritability. Drainage can be present. An inflamed, opaque, bulging, or perforated tympanic membrane is usually demonstrated. The cause can be proved by Gram stain and culture of purulent fluid obtained by tympanocentesis. Otitis caused by *H. influenzae* type b may occur in association with bacteremia and meningitis.

Cellulitis

H. influenzae type b is the cause of 5 to 15% of cases of cellulitis in young children. Most of the infections occur on the face or neck. *H. influenzae* cellulitis is often described as causing a distinctive blue or violaceous discoloration of the skin. However, the fever, erythema, and tenderness observed may not be distinguishable from those attributable to other causes. The diagnosis is established by culture of blood or tissue aspirates from the involved area, or both.

Bacteremia without a Primary Focus of Infection

H. influenzae causes primary bacteremia in both children and adults. In infants or children, occult meningitis or epiglottitis can be present. Rigorous clinical and laboratory evaluation is essential to avoid missing a diagnosis of life-threatening focal infection in these patients. In adults, primary *H. influenzae* bacteremia often occurs in those with underlying diseases, such as lymphoma, leukemia, or alcoholism.

Obstetric and Gynecologic Infection

Pregnancy is associated with a significant risk for *H. influenzae* infection. In the Atlanta study, 7 of 47 adult *H. influenzae* invasive infections occurred in pregnant women. Nontypeable *H. influenzae* is also an important cause of tubo-ovarian abscess and salpingitis in women. A recent study of women in England and Wales showed that pregnancy was associated with an increased risk of invasive, mostly unencapsulated *H. influenzae* infection and that these infections were associated with poor pregnancy outcomes, including fetal loss and extremely premature births or stillbirths.[7]

Pericarditis

H. influenzae type b is an important cause of primary bacterial pericarditis in children. It rarely causes this infection in adults; however, pericarditis can occur in association with pneumonia, probably as a result of contiguous spread of the infection.

Endocarditis

H. influenzae is an unusual cause of endocarditis in view of the frequency with which invasive disease occurs. Most infections occur in patients with preexisting valvular heart disease. Because of its slow initial growth in blood culture media, diagnosis of this infection may be delayed or missed. Patients with *H. influenzae* endocarditis are at high risk for arterial embolic phenomena.

Septic Arthritis

H. influenzae type b is a common cause of septic arthritis in young children; it is rare in adults. *H. influenzae* type b arthritis is clinically indistinguishable from other causes of pyogenic arthritis.

Purulent Conjunctivitis and Brazilian Purpuric Fever

H. influenzae, biogroup *aegyptius* (Koch-Weeks bacillus), causes epidemic purulent conjunctivitis in children. This disease commonly occurs in hot climates or in the summer season.

The infection is characterized by conjunctival erythema, edema, mucopurulent exudate, and varying discomfort in the eyes. An unusually virulent clone of *H. influenzae*, biogroup *aegyptius*, causes an invasive infection called Brazilian purpuric fever, which is characterized by petechial or purpuric skin lesions and vascular collapse; it occurs days to weeks after an initial episode of conjunctivitis in infants and children younger than 10 years.

TREATMENT Rx

Third-generation cephalosporins are considered the treatment of choice for serious *H. influenzae* infections, such as meningitis or epiglottitis. Treatment

with ceftriaxone (adult dose, 1 to 2 g intravenously every 12 hours) or cefotaxime (adult dose, 2 g intravenously every 6 hours) should be started in patients with proven or suspected *H. influenzae* infection, and it should be continued at least until full susceptibility data are available.

Ampicillin was effective treatment of all *H. influenzae* infections until the mid-1970s. Since the first reports of ampicillin-resistant *H. influenzae* isolates in 1972, however, the prevalence of resistance has increased dramatically. Most resistance is due to a plasmid-mediated, R-factor enzyme (tumor endothelial marker 1) β-lactamase, which can be detected rapidly in the laboratory. A small number of isolates, however, have altered penicillin-binding proteins that have decreased binding affinity to penicillin and other β-lactam antibiotics. As a consequence, the isolates may be resistant to some cephalosporins, such as cefaclor, cefamandole, and cefuroxime, in addition to ampicillin. Therefore, patients with proven or suspected *H. influenzae* infections should not be treated with ampicillin or second-generation cephalosporins until susceptibility to these antibiotics has been proved. Chloramphenicol resistance also occurs in *H. influenzae*; an inactivating enzyme, chloramphenicol acetyltransferase, causes resistance. A small number of *H. influenzae* isolates are resistant to both ampicillin and chloramphenicol.

Oral antibiotics are commonly used to treat tracheobronchitis in patients with COPD and otitis media in children, in whom *H. influenzae* isolates are common. Because of resistance, ampicillin and amoxicillin cannot be recommended for the more serious of these infections unless the susceptibility of isolates is known. Most *H. influenzae* isolates are susceptible to amoxicillin-clavulanate. They are also susceptible to azithromycin and clarithromycin, the newer macrolide antibiotics. Fluoroquinolones, such as ciprofloxacin, ofloxacin, levofloxacin, and gatifloxacin, are active against these organisms. Trimethoprim-sulfamethoxazole is also effective for most isolates.

PREVENTION

The first *H. influenzae* type b vaccines were licensed for use in the United States in 1985. They contained purified PRP antigens. However, postlicensing studies of PRP vaccines in the United States showed variable efficacy. The PRP vaccines elicit a type 2, thymus-independent B-cell response, generate few (if any) memory B cells, and fail to stimulate a response in neonates and infants.

Protein-conjugated PRP vaccines were developed to overcome the problem of lack of immune response in the most susceptible infants and some young children. Several are now licensed for use in infants.[8] At present, protein-conjugated PRP vaccines are recommended for use in all infants older than 2 months but not earlier than 6 weeks of age. Studies have shown that protein-conjugated vaccines are effective in diverse populations, including adults with COPD.

Antibiotic prophylaxis should be used for the nonimmunized household or daycare contacts of patients with invasive *H. influenzae* type b disease. Rifampin is the treatment of choice. It should be given in a dosage of 10 mg/kg once daily for 4 days to neonates younger than 1 month, 20 mg/kg (up to a maximum of 600 mg) once daily for 4 days to older children, and 600 mg/day for 4 days to adults.

Other *Haemophilus* Species

Haemophilus parainfluenzae can be found as part of the normal flora of the mouth and pharynx (Table 300-1). It is a rare cause of meningitis in children and an even rarer cause of meningitis in adults. It may cause dental infections or dental abscesses. Cases of brain abscess, epidural abscess, liver abscess, osteomyelitis, pneumonia, empyema, epiglottitis, peritonitis, septic arthritis, and bacteremia have been reported to be caused by this organism. *H. parainfluenzae* also causes subacute endocarditis, often in young adults. *Haemophilus* species are responsible for approximately 1% of cases of infective endocarditis in non–drug-abusing patients. *H. parainfluenzae*, *Haemophilus aphrophilus*, and *Haemophilus paraphrophilus* are the species most frequently recovered from these patients. *H. parainfluenzae* forms bulky vegetations on heart valves. Arterial embolization is common in patients with *H. parainfluenzae* endocarditis. Most isolates are sensitive to ampicillin, but some produce β-lactamases.

MORAXELLA INFECTIONS

DEFINITION

Moraxella species are associated with a variety of infections, the most common of which is exacerbation of chronic bronchitis by *Moraxella catarrhalis*.[9]

The Pathogen

Moraxella organisms are small, gram-negative bacteria that grow well on blood or chocolate agar. They are catalase and oxidase positive. These small diplococci are morphologically difficult to distinguish from *Neisseria*. Some *Moraxella* species are gram-negative bacilli. *M. catarrhalis* is the most important pathogen of this genus (Table 300-2).

PATHOBIOLOGY

The organism is isolated exclusively from humans and is found predominantly in the respiratory tract. *M. catarrhalis* adheres to mucosal cells with the aid of pili. Infection is believed to result from contiguous spread of the organism from sites of colonization, possibly as a result of the introduction of new, more virulent strains to which the host lacks immunity. *M. catarrhalis* possesses multiple virulence factors that can be carried through biologically active outer membrane vesicles to contribute to the pathobiology of otitis media.[10]

M. catarrhalis can often be found in respiratory secretions together with *H. influenzae*. Although the mechanism for coexistence of these pathogens is not known, evidence suggests that the outer membrane vesicles of *M. catarrhalis* inactivate complement, thus enhancing survival of *H. influenzae*.

CLINICAL MANIFESTATIONS

M. catarrhalis is associated with exacerbations of chronic bronchitis. Studies indicate that this organism can be isolated from 0.2 to 8.1% of the sputum aspirates of patients with this disease. It is the third most common pathogen isolated from these patients behind *S. pneumoniae* and *H. influenzae*.

M. catarrhalis can cause pneumonia, particularly in elderly patients with COPD and other underlying conditions such as diabetes mellitus. Sir William Osler is believed to have died as a result of *M. catarrhalis* pneumonia. Cases of bacteremic pneumonia have been reported. In addition, *M. catarrhalis* can cause nosocomial pneumonia with evidence of patient-to-patient spread of the organism.

M. catarrhalis is a common cause of otitis media in young children. Microbiologic studies indicate that this organism is present in approximately 15% of the aspirates from such patients. The organism also causes sinusitis and is a rare cause of bacteremia in children and adults.

Serious infections with other *Moraxella* species are uncommon. However, these organisms are associated with chronic conjunctivitis. Furthermore, case reports have documented the rare occurrence of invasive infections, including bacteremia, endocarditis, arthritis, pericarditis, and meningitis. Meningitis may occur in patients with complement deficiency.

TREATMENT Rx

Oral antibiotics are sufficient for the treatment of most *M. catarrhalis* infections. Inducible β-lactamases are present in many isolates. Therefore, treatment with a β-lactamase–stable antibiotic such as amoxicillin-clavulanate (usual adult dose, 500 mg every 12 hours), a cephalosporin (e.g., cefaclor, usual adult dose, 500 mg every 8 hours), or a non–β-lactam antibiotic such as trimethoprim-sulfamethoxazole (usual adult dose, 1 double-strength tablet every 12 hours) should be initiated pending susceptibility test results.

TABLE 300-2 SITES OF COLONIZATION AND INFECTION BY *MORAXELLA* SPECIES

SPECIES	NORMAL FLORA	ASSOCIATED DISEASES
M. catarrhalis	Oral cavity and upper respiratory tract	Chronic bronchitis exacerbation Otitis media Pneumonia Sinusitis Bacteremia, endocarditis Arthritis, osteomyelitis, epiglottitis (all extremely rare)
M. lacunata	Upper respiratory tract	Chronic conjunctivitis
Other *Moraxella*	Upper respiratory tract	Rare cases of bacteremia, endocarditis, arthritis, meningitis

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Laupland KB, Schønheyder HC, Ostergaard C, et al. Epidemiology of *Haemophilus influenzae* bacteremia: a multi-national population-based assessment. *J Infect*. 2011;62:142-148.
2. MacNeil JR, Cohn AC, Farley M, et al. Current epidemiology and trends in invasive *Haemophilus influenzae* disease—United States, 1989-2008. *Clin Infect Dis*. 2011;53:1230-1236.
3. Ulanova M, Tsang RS. *Haemophilus influenzae* serotype a as a cause of serious invasive infections. *Lancet Infect Dis*. 2014;14:70-82.
4. Clark SE, Eichelberger KR, Weiser JN. Evasion of killing by human antibody and complement through multiple variations in the surface oligosaccharide of *Haemophilus influenzae*. *Mol Microbiol*. 2013;603-618.
5. Okada F, Ando Y, Tanoue S, et al. Radiological findings in acute *Haemophilus influenzae* pulmonary infection. *Br J Radiol*. 2011;85:121-126.
6. Narang R, Bakewell K, Peach J, et al. Bacterial distribution in the lungs of children with protracted bacterial bronchitis. *PLoS ONE*. 2014;9:e108523.
7. Collins S, Ramsay M, Slack MP, et al. Risk of invasive *Haemophilus influenzae* infection during pregnancy and association with adverse fetal outcomes. *JAMA*. 2014;311:1125-1132.
8. Prymula R, Kriz P, Kaliskova E, et al. Effect of vaccination with pneumococcal capsular polysaccharides conjugated to *Haemophilus influenzae*–derived protein D on nasopharyngeal carriage of *Streptococcus pneumoniae* and *H. influenzae* in children under 2 years of age. *Vaccine*. 2010;28:71-78.
9. Aebi C. Moraxella catarrhalis—pathogen or commensal? *Adv Exp Med Biol*. 2011;697:107-116.
10. Hassan F. Molecular mechanisms of *Moraxella catarrhalis*–induced otitis media. *Curr Allergy Asthma Rep*. 2013;13:512-517.

REVIEW QUESTIONS

1. Which of the following statements about *Haemophilus influenzae* is true?

A. It is a gram-positive bacterium that commonly causes skin infections in infants, children, and older adults.
B. It is a gram-negative bacterium that commonly causes respiratory tract infections in infants, children, and older adults.
C. It is a gram-negative bacterium that can be isolated on and grows easily on MacConkey's agar plates without blood or other supplements.
D. It is a virus that causes epidemic influenza infections.

Answer: B *H. influenzae* is a gram-negative bacterium that causes respiratory and invasive infections in infants, children, and older adults. It is fastidious and grows best on chocolate agar in a carbon dioxide–enriched environment. It was originally recovered from patients with influenza and was considered the cause of that disease, hence its name. (See section on the pathogen.)

2. Which of the following statements about human host defenses against *Haemophilus influenzae* is true?

A. Cell-mediated immunity is the most important host defense against infection.
B. Antibody directed against the cell wall and complement are the most important host defenses against infection.
C. Antibody against the cell capsule and complement are the most important host defenses against infection.
D. Complement plays no role in host defense against infection.

Answer: C Antibody directed against the cell capsule and complement provide bactericidal immunity to *H. influenzae*. Subjects lacking either are at risk of invasive infection. Complement is an essential component of host defenses against some *H. influenzae* diseases, and complement deficiency states predispose to infection with the organism. (See section on pathobiology.)

3. Which of the following statements about treatment of *Haemophilus influenzae* infections is true?

A. Ampicillin is the initial treatment of choice for infection pending results of susceptibility tests in the laboratory.
B. Chloramphenicol is the initial treatment of choice for infection pending results of susceptibility tests in the laboratory.
C. Rifampin is the treatment of choice pending results of susceptibility tests in the laboratory.
D. A third-generation cephalosporin (e.g., ceftriaxone) is the treatment of choice pending results of susceptibility tests in the laboratory.

Answer: D Many strains of *H. influenzae* produce β-lactamases that inactivate ampicillin, and the prevalence of ampicillin resistance has increased dramatically. An inactivating enzyme, chloramphenicol acetyltransferase, causes resistance to that antibiotic. Rifampin is the treatment of choice for prophylaxis of nonimmunized contacts but not for established infection. (See sections on treatment and prevention.)

4. Which of the following statements about prevention of *Haemophilus influenzae* infections is true?

A. The unconjugated Hib polysaccharide vaccine has been effective in preventing otitis, meningitis, and bacteremic *H. influenzae* type b infections in infants, toddlers, children, and adults.
B. The conjugated Hib polysaccharide vaccine has been effective in preventing otitis, meningitis, and bacteremic *H. influenzae* type b infections in infants, toddlers, children, and adults.
C. The conjugated Hib polysaccharide vaccine has been effective in preventing otitis, meningitis, and bacteremic all serotypes of *H. influenzae* infections in infants, toddlers, children, and adults.
D. Vaccines are generally ineffective in preventing *H. influenzae* type b infection, and rifampin is the only effective prophylactic treatment.

Answer: B The conjugated vaccine protects infants, children, and, by reducing the carrier state, older adults against invasive *H. influenzae* infections. (See section on prevention.)

5. Which of the following statements about *Moraxella catarrhalis* infections is true?

A. *M. catarrhalis* is a commensal of the respiratory tract and rarely causes infection.
B. *M. catarrhalis* is commonly associated with bacteremia and meningitis.
C. *M. catarrhalis* is commonly associated with exacerbations of chronic bronchitis.
D. *M. catarrhalis* is commonly found in the respiratory tract of pet animals including dogs and cats and is transmitted from them to humans, in whom it can cause disease.

Answer: C New strains of *M. catarrhalis* are commonly associated with exacerbations of chronic bronchitis, and treatment of these speeds recovery. (See section on *Moraxella* infections).

301

CHANCROID

STANLEY M. SPINOLA

DEFINITION

Chancroid is a sexually transmitted disease characterized by painful genital ulcers and inguinal lymphadenitis. Chancroid is caused by *Haemophilus ducreyi*, a gram-negative coccobacillus that is not a true *Haemophilus* species. Within the Pasteurellaceae, *H. ducreyi* is grouped in a distinct lineage with *Mannheimia haemolytica* and *Actinobacillus pleuropneumoniae*. *H. ducreyi* is likely to have diverged from these animal respiratory pathogens to occupy its niche in the human genital epithelium.

EPIDEMIOLOGY

Chancroid is endemic in resource-poor regions of Africa and Asia and facilitates the transmission of human immunodeficiency virus (HIV-1). In the 1990s, the World Health Organization estimated the annual global prevalence of chancroid to be 4 million to 6 million cases. Because of the widespread use of syndromic management, which consists of treatment for syphilis and chancroid without diagnostic testing, the prevalence of chancroid has dramatically declined in endemic areas. Chancroid can be maintained only in networks with high sex partner change rates; female sex workers play an important role in its epidemiology. Targeted treatment of infected sex workers leads to eradication of the disease in endemic areas. Despite these successes, reports of chancroid persist from many countries. Such reports imply a reservoir of untreated sex workers. Urban outbreaks of chancroid associated with sex work occurred in the United States in the 1980s and 1990s. Owing to contact tracing and treatment efforts, the number of domestic cases of chancroid has decreased steadily, with a 65-year low of eight cases in 2013; such sporadic cases are likely to be imported after contact with infected persons in endemic areas.

The male-to-female ratio of chancroid is 3 : 1. The excess number of male cases is usually attributed to the infection of multiple partners by sex workers. However, human inoculation experiments indicate that men are twice as susceptible as women for development of pustules, suggesting that male gender is a risk factor for disease progression.[1]

PATHOBIOLOGY

Much of what is known about the pathogenesis of *H. ducreyi* is derived from experiments in which bacteria are inoculated into the skin of the upper arm of human volunteers. Puncture wounds are required to initiate infection, and the estimated infectious dose is as low as one bacterium. Papules develop within 24 hours and either spontaneously resolve or evolve into pustules in 2 to 5 days. Neutrophils and macrophages surround the organism and form an abscess that erodes the epidermis. Below the abscess, there is a collar of macrophages and regulatory T cells and a dermal infiltrate of macrophages, CD4 and CD8 T cells, natural killer (NK) cells, and dendritic cells. This histopathology resembles a suppurative granuloma and is identical to that of natural ulcers. In both experimental and natural infection, *H. ducreyi* associates with neutrophils and macrophages, which fail to ingest the organism. Mutant versus parent trials have revealed bacterial components that are required for infection; several are involved with adherence and resistance to serum killing and phagocytosis. In addition to gender, the human model has shown that there are host effects on disease progression. Differential host susceptibility is associated with distinct dendritic cell responses to the organism, which may shape T-cell and NK-cell responses that influence the ability of phagocytes to ingest the organism.

CLINICAL MANIFESTATIONS

H. ducreyi enters the skin through breaks in the epithelium that occur during intercourse.[2] Papules form within hours to days and evolve into pustules in 2 to 3 days. After a few days to 2 weeks, the pustules ulcerate. Patients typically develop one to four painful ulcers (Fig. 301-1) but do not seek treatment until they have had ulcerative symptoms for 1 to 3 weeks. By this time, 10 to 40% have suppurative inguinal lymphadenopathy or buboes (see Fig. 301-1).

FIGURE 301-1. Typical chancroidal ulcer and lymphadenitis in a man. (From Herpes-Coldsores.com. http://www.herpes-coldsores.com/std/chancroid_pictures.htm.)

Natural ulcers are classically very painful and nonindurated, with ragged edges. The ulcer may be covered by a yellow or gray necrotic exudate and bleeds when scraped. However, this presentation occurs in a minority of patients; chancroid is frequently indistinguishable from syphilis and genital herpes. Lesions in men are usually on the foreskin, coronal sulcus, or penile shaft. Lesions in women are usually on the labia; but women may have internal vaginal and cervical ulcers that are painless. Lesions also occur on the thighs and buttocks or at distant sites; extragenital lesions are thought to be due to autoinoculation. Untreated, chancroid persists for months and causes giant ulcers, erosion of the infected area, or fibrosis, leading to phimosis in men.

In the western Pacific, *H. ducreyi* causes a chronic limb ulceration syndrome that is not sexually transmitted and occurs primarily in children.[3] Close contact with family members with ulcers is implicated in these cases. In a cohort study done in yaws-endemic villages in Papua New Guinea, the prevalence of this syndrome is 3.2 cases per 100 persons. *H. ducreyi*, *Treponema pallidum* subspecies *pertenue*, and dual infections accounted for 58%, 26%, and 16% of the ulcers, respectively. The overall prevalence of *H. ducreyi* infection in children aged 5 to 15 years is an astoundingly high 7%. Similar data are emerging from other yaws-endemic areas.[4]

DIAGNOSIS

Diagnosis requires either a positive culture or a polymerase chain reaction (PCR) test. In research settings, PCR has a resolved sensitivity of 95 to 98% and a specificity of 99% for *H. ducreyi*. In comparison, the sensitivity for culture is approximately 75%, but clinical diagnosis is neither sensitive (range, 50 to 75%) nor specific (range, 50 to 75%). Unfortunately, PCR-based tests are not commercially available.[5] Most sexually transmitted disease clinics do not routinely test patients for chancroid, and the diagnosis is typically made by the exclusion of genital herpes and syphilis. If patients with genital ulcers and lymphadenitis or treatment failure for primary syphilis appear in a community, public health authorities should be notified so that specific diagnostic testing can be initiated.

The differential diagnosis includes syphilis, genital herpes, lymphogranuloma venereum, and granuloma inguinale. Mixed infections with herpes simplex virus or syphilis are common, occurring in approximately 17% of chancroid cases diagnosed by PCR. Patients with suspected chancroid

should be tested for genital herpes and have serologic tests for syphilis and HIV-1 and a dark-field examination.

TREATMENT

Rx

Because of syndromic management, little is known about the current prevalence of antibiotic resistance in *H. ducreyi*, but most clinical isolates have had plasmid-mediated resistance to ampicillin, tetracyclines, and sulfonamides. The only reliable treatment regimens are macrolides, quinolones, and third-generation cephalosporins; there have been isolated reports of erythromycin and ciprofloxacin resistance. Owing to the propensity of *H. ducreyi* to acquire plasmids, the fact that some Enterobacteriaceae harbor plasmids that encode extended-spectrum β-lactamases and quinolone resistance is a concern. Current treatment recommendations include single-dose azithromycin 1 g orally or ceftriaxone 250 mg intramuscularly, ciprofloxacin 500 mg orally twice a day for 3 days, and erythromycin base 500 mg orally three times a day for 7 days.[6]

Repeated aspiration of buboes may be required to effect a cure. In a randomized study comparing repeated aspiration versus incision and drainage, incision and drainage was considered preferable. However, incision and drainage may cause excessive scarring, especially in persons of African descent, and should be avoided according to some experts.

Initial reports of coinfection with HIV infection and chancroid suggest that such individuals have a greater number of ulcers that do not heal as readily after antibiotic treatment compared with patients infected with *H. ducreyi* alone and that single-dose regimens may not be effective in this setting. Antibiotic treatment failure is also associated with lack of circumcision. If close follow-up cannot be ensured, most experts recommend multidose regimens in HIV seropositives.

PROGNOSIS

Clinical cure correlates with a reduction in pain and purulence and re-epithelialization of the ulcer within 7 days. Patients who do not show improvement within 7 days should be regarded as treatment failures and given an alternative agent. Even if *H. ducreyi* is eradicated, ulcers may persist if genital herpes or syphilis is present and not treated. Most ulcers heal in 2 weeks; large ulcers may take 4 weeks to heal.

PREVENTION

Circumcision protects against chancroid in men. Condoms are likely to be protective. Although several antigens that may afford protection in animal models have been identified, there is no vaccine. Contacts of patients with chancroid should be treated with an approved regimen.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Janowicz DM, Ofner S, Katz BP, et al. Experimental infection of human volunteers with *Haemophilus ducreyi*: fifteen years of clinical data and experience. *J Infect Dis*. 2009;199:1671-1679.
2. Lewis DA. Epidemiology, clinical features, diagnosis and treatment of Haemophilus ducreyi—a disappearing pathogen? *Expert Rev Anti Infect Ther*. 2014;12:687-696.
3. Mitjà O, Lukehart SA, Pokowas G, et al. *Haemophilus ducreyi* as a cause of skin ulcers in children from a yaws-endemic area of Papua New Guinea: a prospective cohort study. *Lancet Glob Health*. 2014;2:e235-e241.
4. Marks M, Chi KH, Vahi V, et al. Haemophilus ducreyi associated with skin ulcers among children, Solomon Islands. *Emerg Infect Dis*. 2014;20:1705-1707.
5. Rao G, Das A, Prabhakar P, et al. Alteration in sample preparation to increase the yield of multiplex polymerase chain reaction assay for diagnosis of genital ulcer disease. *Indian J Med Microbiol*. 2013;31:15-18.
6. Centers for Disease Control and Prevention. Sexually transmitted diseases treatment guidelines, 2010. *MMWR Recomm Rep*. 2010;59(RR–12):1-116.

REVIEW QUESTIONS

1. Which of the following drugs would not be used to empirically treat a patient with suspected chancroid?

A. Ampicillin
B. Erythromycin
C. Ciprofloxacin
D. Ceftriaxone
E. Azithromycin

Answer: A Most strains of *Haemophilus ducreyi* harbor plasmids that encode β-lactamase. Although there are few data on the current state of antibiotic resistance in *H. ducreyi*, there have been few reports of resistance to quinolones, macrolides, or third-generation cephalosporins.

The following scenario applies to questions 2, 3, and 4.

A 28-year-old man who recently traveled to Africa presents with two painful, large genital ulcers and inguinal lymphadenopathy. He states that he had sex with a bar worker during the trip and subsequently had sex with his usual partner when he returned home before he noted symptoms. The lesions evolved during a period of 3 weeks. A dark-field examination for *Treponema pallidum* and a polymerase chain reaction (PCR) analysis for herpes simplex virus of the ulcer exudate are negative.

2. Serologic tests that should be ordered in this situation include those for

A. HIV-1
B. Syphilis
C. *H. ducreyi*
D. Both A and B
E. A, B, and C

Answer: D All patients with genital ulcer disease should be tested for HIV-1 and syphilis. There are no serologic tests for *H. ducreyi.*

3. Presumptive therapy should be given with

A. Doxycycline for suspected lymphogranuloma venereum
B. Acyclovir for suspected herpes simplex
C. Benzathine penicillin for suspected syphilis
D. A 3-week course of azithromycin for suspected granuloma inguinale
E. A single dose of azithromycin for suspected chancroid

Answer: E Provided his syphilis serologic test result is also negative, the patient has met clinical criteria for a presumptive diagnosis of chancroid; a single 1-g dose of oral azithromycin has cure rates of more than 90%.

4. His regular partner should be

A. Notified and be told to seek care if she develops symptoms
B. Notified and be examined but not treated if there are no lesions
C. Notified, be examined, and be treated regardless of the presence or absence of lesions or laboratory findings
D. Notified only if the patient consents
E. Notified, be examined, and be treated only if she has laboratory confirmation of infection

Answer: C The estimated transmission rate of *H. ducreyi* from infected men to women is 70%. As is true for all bacterial agents of sexually transmitted infections, all contacts should be notified, be examined, have a laboratory evaluation, and be treated regardless of clinical or laboratory findings.

5. The most sensitive and specific method for diagnosis of *H. ducreyi* infection is

A. Gram stain of the ulcer exudate
B. PCR-based test of a swab of the exudate
C. The typical clinical findings of large, nonindurated ulcers with inguinal lymphadenitis
D. Culture of the exudate
E. Direct fluorescence microscopy with specific antibodies

Answer: B PCR has a high sensitivity and specificity for *H. ducreyi*; culture is 100% specific but at most 75% sensitive. The other modalities hover in the 50% sensitivity and specificity range.

302

CHOLERA AND OTHER *VIBRIO* INFECTIONS

EDUARDO GOTUZZO AND CARLOS SEAS

CHOLERA

DEFINITION

Cholera is a feared epidemic diarrheal disease caused by *Vibrio cholerae* serogroup O1 and, since 1992, by the new serogroup O139. The disease is characterized by acute watery diarrhea. In its more severe form, a person may be severely dehydrated and in hypovolemic shock; the patient may die in a matter of a few hours after contracting the infection if treatment is not provided. Cholera is endemic today in Africa and Asia, and cases are also reported from Latin America, North America, and Europe. Seven pandemics have been registered in history since 1816; the most recent has lasted more than five decades since its recognition in Indonesia in 1961.

The Pathogen

V. cholerae is a curved gram-negative bacillus that belongs to the family Vibrionaceae and shares common characteristics with the family Enterobacteriaceae. *V. cholerae* O1 can be classified into three serotypes according to the presence of somatic antigens and into two biotypes, classic and El Tor, according to specific phenotypic characteristics. There is no evidence of different clinical spectra among the three serotypes of *V. cholerae*. The classic biotype, responsible for the first six pandemics of cholera, causes an approximately equal number of symptomatic and asymptomatic cases, whereas the El Tor biotype causes more asymptomatic infections. The classic biotype is confined to the south of Bangladesh, and the El Tor biotype is responsible for the current pandemic. The O139 serogroup is composed of a variety of genetically diverse strains, both toxigenic and nontoxigenic; it is genetically closer to El Tor *V. cholerae*.

EPIDEMIOLOGY

Cholera has both a predisposition to cause epidemics with pandemic potential and an ability to remain endemic in all affected areas. People of all ages are at risk to contract the infection in epidemic settings, whereas children older than 2 years are mainly affected in endemic areas. *V. cholerae* lives in riverine, brackish, and estuarine ecosystems, where both O1 and non-O1 strains coexist, with non-O1 and nontoxigenic O1 strains predominating over toxigenic O1 strains. In its natural environment, *V. cholerae* lives attached to algae or to crustacean shells and copepods, with which it coexists in a symbiotic manner. Several conditions, such as temperature, salinity, and availability of nutrients, determine the survival of *V. cholerae*; when these conditions are adverse, vibrios survive in a viable but nonculturable state. More recent data suggest that cholera phages modulate the abundance of *V. cholerae* in the environment and determine the beginning and end of epidemics. Phages may also play a role in the emergence of new *V. cholerae* serogroups by transferring genetic material to nontoxigenic strains.

From its aquatic environment, *V. cholerae* is introduced to humans through contamination of water sources and food. Once humans are infected, very high attack rates may take place, particularly in previously naïve populations.[1] Acquisition of the disease by drinking contaminated water from rivers, ponds, lakes, and even tube well sources has been documented. Drinking unboiled water, introducing hands into containers used to store drinking water, drinking beverages from street vendors, drinking beverages to which contaminated ice has been added, and drinking water outside the home are risk factors; these factors contributed to the acquisition of cholera during the large Peruvian epidemic of 1991. Drinking boiled water, acidic beverages, and carbonated water and using narrow-necked vessels for storing water are protective measures. Epidemics of cholera associated with the ingestion of leftover rice, raw fish, cooked crabs, seafood, raw oysters, and fresh vegetables and fruits have been documented. Person-to-person transmission is less likely to occur because a large inoculum is necessary to transmit disease. High transmission rates (approximately 50%) are reported among household contacts of patients with cholera in endemic areas.[1]

Epidemics of cholera tend to occur during the hot season. Factors affecting climate change and climate variability have an impact on the incidence of cholera. The El Niño–southern oscillation (ENSO), a periodic phenomenon representative of global climate variability, affects the transmission of cholera and vector-borne diseases. ENSO causes the warming of normally cool waters in the Pacific coastline of Peru, thereby promoting phytoplankton bloom, zooplankton bloom, and *V. cholerae* proliferation.

Some host factors are important in the transmission of cholera. The chronic gastritis associated with *Helicobacter pylori* predisposes to cholera by inducing hypochlorhydria, which reduces the ability of the stomach to contain the infection. An unexplained predisposition to severe disease in persons with the O blood group has been observed in Asia and more recently in Latin America. Thus, complex associations among climatic, seasonal, bacterial, and human factors affect cholera transmission. Although for the most part developing countries are affected by cholera, several developed countries, such as the United States, Canada, and Australia, have reported indigenous and imported cases.[2] The most recent epidemic occurred in Haiti in 2010. During the first 2 years of the epidemic, 604,634 cases were reported; the cumulative case-fatality rate was 1.2%.[3] Figure 302-1 shows the distribution of cholera in the world from 1990 to 2011.[4]

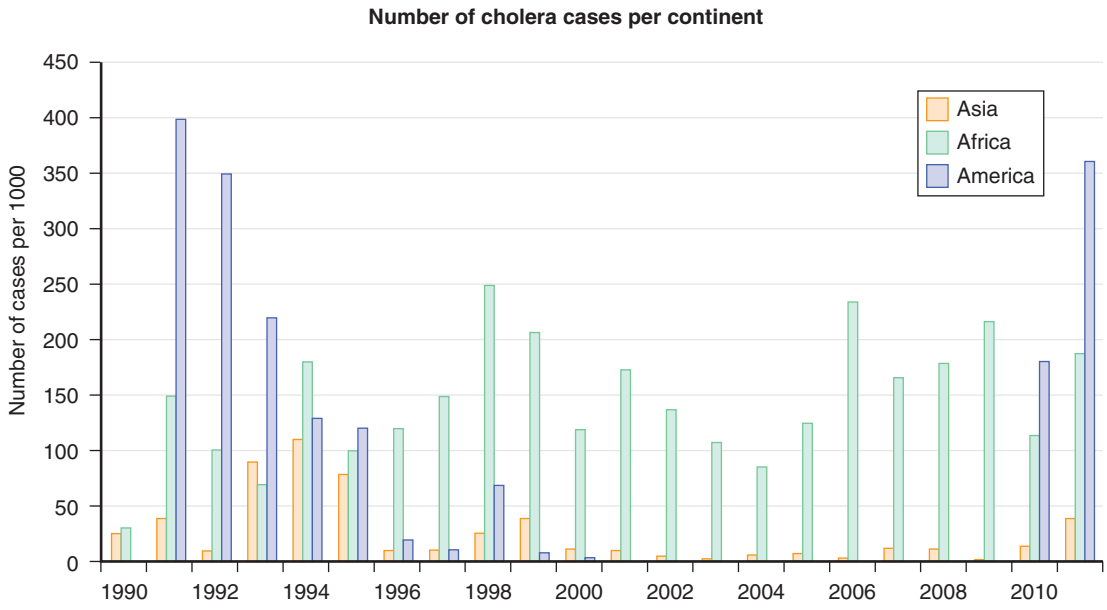

FIGURE 302-1. World distribution of cholera from 1990 to 2011 based on reports to the World Health Organization. (Reprinted with permission from the World Health Organization. Cholera, 2011. *Wkly Epidemiol Rec*. 2012;87:289-304.)

TABLE 302-1 ELECTROLYTE COMPOSITION OF CHOLERA STOOLS AND SOLUTIONS RECOMMENDED FOR TREATMENT

	Na^+	Cl^-	K^+	HCO_3^-	GLUCOSE	OSMOLARITY
Stools of adults with severe cholera	130	100	20	44		
Intravenous lactated Ringer solution	130	109	4	28*	0	271
Intravenous normal saline	154	154	0	0	0	308
Standard oral rehydration solution promoted by the WHO	90	80	20	10†	111	311
Reduced-osmolarity oral rehydration solution promoted by the WHO	75	65	20	10†	75	245
Rice-based oral rehydration solution	90	80	20	10†		270

*Lactated Ringer solution contains citrate instead of bicarbonate.
†Bicarbonate is replaced by trisodium citrate.
Glucose concentration is in mg/dL, electrolyte concentrations are in mEq/L, and osmolarity is in mOsm/L. WHO = Word Health Organization.
Modified with permission from Seas C, DuPont HL, Valdez LM, et al. Practical guidelines for the treatment of cholera. *Drugs*. 1996;51:966-973.

PATHOBIOLOGY

V. cholerae O1 and O139 cause clinical disease by secreting an enterotoxin that promotes the secretion of fluids and electrolytes by the small intestine. The infectious dose of bacteria varies with the vehicle. When water is the vehicle, more bacteria are needed to cause disease (10^3 to 10^6), but when the vehicle is food, lower amounts are needed (10^2 to 10^4). The incubation period varies from 12 to 72 hours; the median is 1.4 days.[5] Cholera toxin (CTX) has two subunits, a pentamer B subunit and a monomer A subunit. The B subunit allows binding of the toxin to a specific receptor, a ganglioside (GM_1) located on the surface of cells lining the mucosa along the intestine of humans and certain suckling mammals. The active, or A, subunit has two components, A1 and A2, linked by a disulfide bond. Activation of adenylate cyclase by the A1 component results in an increase in cyclic adenosine monophosphate in intestinal epithelial cells, which blocks the absorption of sodium and chloride by microvilli and promotes the secretion of chloride and water by crypt cells. These events lead to the production of watery diarrhea with electrolyte concentrations similar to those of plasma, as shown in Table 302-1. A few other toxins have been isolated from pathogenic *V. cholerae*, but their roles in genesis of the disease are less clear.

The genetic material of El Tor *V. cholerae* O1 is included in two circular chromosomes, the larger containing 3 megabases and the smaller containing 1.07 megabases. The main virulence genes are *ctxA* and *ctxB*, which encode CTX subunits A and B, respectively, and *tcpA*, which codes for toxin coregulated pilus. Regulation of expression of these genes is complex. Recent data suggest that vibrios are able to upregulate the expression of CTX in response to intestinal fluid components as well as in the presence of certain environmental factors. Genes unique to El Niño–southern oscillation (ENSO) *V. cholerae* may encode specific features that allow this biotype to better survive in the environment as well as to be more infectious to humans.

CLINICAL MANIFESTATIONS

Cholera is characterized by watery diarrhea and dehydration, which ranges from mild to severe and life-threatening. Patients with mild dehydration cannot be differentiated from those infected by other enteric pathogens causing watery diarrhea. In contrast, patients with severe dehydration secondary to cholera are easy to identify in that their stools have the appearance of rice water and no other clinical illness produces such severe dehydration as quickly (in a matter of a few hours) as cholera. Onset of the disease is abrupt and characterized by watery diarrhea, vomiting, generalized cramps, and oliguria. Physical examination shows a feeble pulse, fever is rarely present, patients look anxious and restless, the eyes are very sunken, mucous membranes are dry, the skin has lost its elasticity and when pinched retracts very slowly, the voice is almost nonaudible, and intestinal sounds are prominent. Although watery diarrhea is the hallmark of cholera, some patients do not have diarrhea but instead have abdominal distention and ileus, a relatively rare type of cholera called cholera sicca.

Laboratory findings in patients with severe dehydration consist of an increase in hematocrit, urine specific gravity, and total serum protein; azotemia; metabolic acidosis with a high anion gap; normal or low serum potassium levels; and normal or slightly low sodium and chloride levels. The calcium and magnesium content in plasma is high as a result of hemoconcentration. Leukocytosis is observed in patients with severe cholera. Hyperglycemia, caused by high concentrations of epinephrine, glucagon, and cortisol stimulated by hypovolemia, is more commonly seen than hypoglycemia. Acute renal failure is the most severe complication of cholera. Incidence rates of 10.6 cases per 1000 were reported in Peru during the first months of the

1991 epidemic. Patients with acute renal failure almost always have a history of improper rehydration. Cholera in pregnant women carries a poor prognosis. Pregnant women have more severe clinical illness, especially when the disease is acquired at the end of the pregnancy. Fetal loss occurs in as many as 50% of these pregnancies. Cholera in the elderly also carries a poor prognosis because of an increase in complications, particularly acute renal failure, severe metabolic acidosis, and pulmonary edema.

DIAGNOSIS

Chaotic movement under dark-field microscopy and a high number of bacteria in a stool sample from patients with diarrhea are characteristic of *V. cholerae* infection. Specific antisera against the serotype block the movement of vibrios and allow confirmation of the diagnosis. Under epidemic conditions, observing bacteria with a darting movement in a stool sample from a patient with suspected infection under dark-field microscopy is adequate to make the diagnosis. Definitive confirmation requires isolation of the bacterium in culture. Specific medium is needed to isolate *V. cholerae* from stool. Higher sensitivity and specificity have been reported with DNA amplification by polymerase chain reaction for detection of vibrios in stool and environmental samples.

TREATMENT

Rx

The objectives of therapy are to restore the fluid losses caused by diarrhea and vomiting, to correct the metabolic acidosis, to restore potassium deficits, and to replace continuous fluid losses.[6] Treatment of patients with milder forms of dehydration is easy, but treatment of patients with severe dehydration requires experience and proper training. The intravenous route should be restricted to patients with some dehydration who do not tolerate the oral route, those who purge more than 10 to 20 mL/kg/hour, and all patients with severe dehydration. Rehydration should be accomplished in two phases: the rehydration phase and the maintenance phase. The purpose of the rehydration phase is to restore normal intravascular volume, and it should last no longer than 4 hours. Intravenous fluids should be infused at a total volume of 100 mL/kg during the rehydration phase in severely dehydrated patients. Lactated Ringer solution is preferred, but other solutions may be used as well (see Table 302-1). All signs of dehydration should have disappeared and the patient should pass urine at a rate of 0.5 mL/kg/hour or greater after the rehydration phase is finished. The maintenance phase follows immediately. During this phase, the objective is to maintain normal hydration status by replacement of ongoing losses. The oral route is preferred during this phase, and the use of oral rehydration solutions is highly recommended. Oral rehydration therapy uses the principle of common transportation of solutes, electrolytes, and water by the intestine not affected by cholera toxin. People with diarrhea can undergo successful rehydration with simple solutions containing glucose and electrolytes. The World Health Organization recommends an oral rehydration solution with reduced osmolarity (245 mOsm/L) to treat all diarrheal diseases. This solution contains lower sodium than the standard oral rehydration solution promoted since 1975 (75 vs. 90 mEq/L). No more symptomatic hyponatremia is observed with the reduced-osmolarity solution than with the standard solution. The addition of L-histidine to rice-based oral rehydration solutions has been shown to reduce the volume and duration of diarrhea and the unscheduled use of intravenous therapy in adult cholera patients.[■] Patients without severe dehydration who tolerate the oral route can be rehydrated with oral rehydration solutions exclusively and discharged promptly from the health center. Recommendations for treatment of cholera patients are shown in Table 302-2. Treatment of cholera caused by *V. cholerae* O139 is the same as described earlier.

Antimicrobial agents are not life-saving and always need to be accompanied by fluid therapy.[7] Effective antibiotics in patients with severe dehydration decrease the duration of diarrhea and the volume of stool by nearly half.[■] Oral tetracycline and doxycycline are the agents of choice in areas of the globe where sensitive strains predominate. A single dose of doxycycline (300 mg) is the preferred regimen. Pregnant women can be treated with erythromycin or furazolidone. Because of the emergence of resistance to tetracyclines and other antimicrobials in many endemic areas, the quinolones and, more recently, azithromycin have been tested in clinical trials. A single-dose regimen of azithromycin (20 mg/kg) showed clinical and bacteriologic results that were comparable to a 3-day regimen with erythromycin in children and comparable to a single-dose regimen of ciprofloxacin (1 g) in adults.[■] The addition of oral zinc (30 mg/day) to an erythromycin regimen in children reduced the duration of diarrhea by 12%, with an additional 11% reduction in the volume of diarrhea in comparison to placebo.[■] Antimotility agents such as loperamide or diphenoxylate, adsorbents, analgesics, and antiemetics are not recommended. Antisecretory drugs, including racecadotril, an enkephalinase inhibitor, are not useful in patients with severe cholera. Chemoprophylaxis of household contacts of cholera cases is not routinely recommended.

TABLE 302-2 RECOMMENDATIONS FOR TREATMENT OF CHOLERA PATIENTS

Determine the degree of dehydration on arrival.
Rehydrate the patient in two phases:
- Rehydration phase—lasts 2 to 4 hours
- Maintenance phase—lasts until the end of the diarrheal episode

Register and periodically review input and output in predesigned charts.
Use the intravenous route in the following situations:
- In all severely dehydrated patients, in whom the total volume to be infused during the rehydration phase is 100 mL/kg. For patients older than 1 year, 30 mL/kg should be infused in the first 30 minutes, the remaining 70 mL/kg should be infused in 2.5 hours. For children younger than 1 year, the first 30 mL/kg should be infused in 1 hour.
- Patients with some dehydration who are unable to tolerate the oral route
- Patients with high stool output (>10 mL/kg/hour) during the maintenance phase

Use oral rehydration solutions, glucose or rice based, during the maintenance phase to match ongoing losses. Volumes of 800 to 1000 mL/hour are usually required. Low-osmolarity solutions are not recommended.
Start an oral antimicrobial agent in patients with severe cholera when full rehydration has been achieved and oral tolerance is confirmed. Single-dose doxycycline, 300 mg, is the preferred regimen. Erythromycin or a quinolone is a suitable alternative.
Discharge patients only if oral tolerance is adequate (≥1000 mL/hour), urine output is satisfactory (≥40 mL/hour), and stool volume is low (≤400 mL/hour).

Modified with permission from Seas C, DuPont HL, Valdez LM, et al. Practical guidelines for the treatment of cholera. *Drugs.* 1996;51:966-973.

PREVENTION

Access to potable water and ensuring proper management of excreta to avoid contamination of other water sources are important measures to reduce transmission of cholera. Alternative ways to prevent cholera transmission are necessary in developing countries. Water can be made safer to drink by boiling, adding chlorine, or filtering it with cloth made of cotton. An inability to implement these measures to curtail cholera transmission has prompted a search for vaccines. An ideal vaccine against cholera should elicit a fast and long-lasting immune response with minimal side effects. Parenteral vaccines are no longer recommended. Two oral vaccines, the two-dose regimen of the inactivated vaccine WC-BS (whole cell plus B subunit) and a single dose of the live attenuated CVD 103-HgR vaccine, have been tested extensively in epidemic settings and in field trials in endemic areas. Although the WC-BS vaccine showed good short-term protective efficacy (85% at 6 months), the results at 3 and 5 years were less impressive (60%), particularly in children. A large effectiveness study in Mozambique confirmed the high short-term protection against cholera (80%) by this vaccine, especially against severe dehydration (90%). In Guinea, oral vaccine was 87% protective.[8] In addition, reanalysis of data on this vaccine in field trials and in Zanzibar has shown that it may also confer herd protection in the unvaccinated population.[9] Cost-effectiveness of interventions like this require prices of the oral vaccine below 1.3 USD.[10] A large field trial of the live attenuated vaccine showed no protective efficacy. Indications for use of the currently available cholera vaccines include travel to endemic areas and situations in which high attack rates of cholera are expected, such as after environmental disasters, in refugee camps, and in urban slums in highly endemic areas. Preemptive and reactive vaccination approaches should be thoroughly evaluated in epidemic settings.[11] New oral vaccines, including both killed and live *Vibrio*, are being evaluated in endemic areas, with promising preliminary reports.

PROGNOSIS

Patients with severe cholera left untreated or improperly treated carry a poor prognosis, with mortality rates higher than 50%. However, case-fatality rates during epidemics may be reduced to values below 1% even in disaster situations, provided adequate access to health care centers and proper management of patients can be ensured. In contrast, figures higher than 10% have been reported in epidemic settings when patients had no access to health care or received improper treatment.

OTHER *VIBRIO* INFECTIONS

Noncholera vibrios have worldwide distribution and coexist in environments in which *V. cholerae* lives. They cause a spectrum of clinical syndromes, including acute diarrhea, soft tissue infections, and sepsis, especially in immunocompromised hosts. In the United States, 7700 cases of *Vibrio* infections

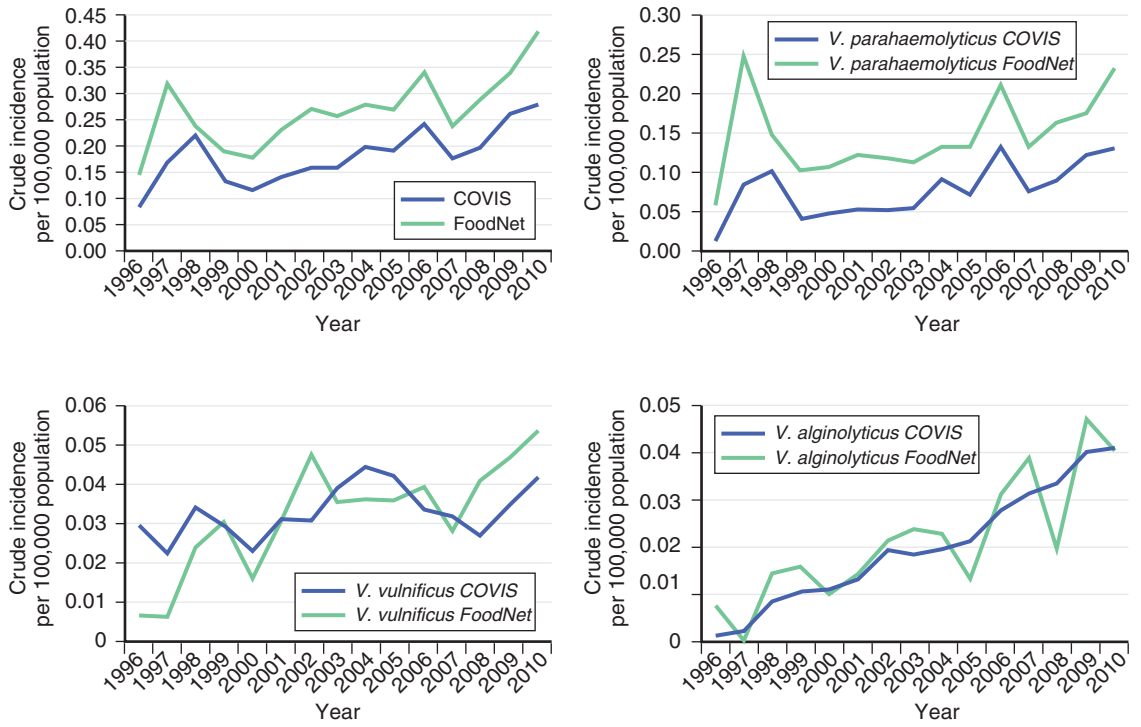

FIGURE 302-2, Crude incidence rates per 100,000 population of noncholera *Vibrio* infections based on two U.S. surveillance reports from 1996 to 2010. COVIS = Cholera and Other *Vibrio* Illness Surveillance; FoodNet = Foodborne Diseases Active Surveillance Network. (Redrawn from Newton A, Kendall M, Vugia DJ, et al. Increasing rates of vibrosis in the United States, 1996-2010: review of surveillance data from 2 systems. *Clin Infect Dis*. 2012;54[Suppl 5]:S391-S395.)

were reported by two surveillance networks covering the period from 1996 to 2010.[12] The incidence of *Vibrio* infection increased during the study period; *Vibrio parahaemolyticus* predominated (44.9% of isolates) but was associated with a low case-fatality rate of 0.7%. In contrast, *Vibrio vulnificus* accounted for 18.8% of the isolates but was associated with a case-fatality rate of 31.9%. *Vibrio* illnesses in the United States are seasonal and peak during the summer (Fig. 302-2). The incubation period for noncholeragenic *Vibrio* infection is usually 12 to 72 hours but can be as long as 1 week.

Nontoxigenic *V. cholerae* causes gastroenteritis, but unlike toxigenic *V. cholerae* O1 or O139, nontoxigenic *V. cholerae* does not cause epidemics. Illness ranges in severity from mild diarrhea to severe watery diarrhea. Fever and bloody diarrhea are unusual, but immunocompromised persons and those with liver disease can experience more severe illness, including fever, chills, and septic shock.

V. parahaemolyticus lives in marine environments and is a source of intestinal illness associated with the ingestion of contaminated shellfish. Certain serovars have shown pandemic spread (O3:K6 and O4:K68). It is not well known how this *Vibrio* causes infection in humans, but the clinical illness may mimic cholera, although most cases are milder and self-limited forms of acute watery diarrhea. Acute dysentery is reported rarely.

In the United States, *V. parahaemolyticus* and *V. vulnificus* as well as other noncholeragenic vibrios caused skin and soft tissue infections in victims and responders affected by the Gulf Coast hurricane disasters in fall 2005. *V. parahaemolyticus* wound infections are generally less severe than those caused by *V. vulnificus*.[13] However, in persons with liver disease or those who are immunocompromised, fatal infections can occur.

V. vulnificus is associated with wound infections in persons in contact with contaminated water as well as with primary sepsis in immunocompromised hosts. Wound infections follow trauma and are characterized by rapid progression of skin and soft tissue involvement, with necrosis and bulla formation occurring in more severe cases. Fever, chills, and sepsis syndrome may ensue rapidly. Primary sepsis with bacteremia and metastatic lesions on the skin, characterized by disseminated erythematous lesions that may evolve to necrotic lesions, is a distinctive clinical manifestation in patients with chronic liver illnesses, alcoholics, and patients with blood disorders such as thalassemia. A history of seafood ingestion, usually oysters, is typical. Patients are acutely ill with high fever and need to be managed aggressively with fluid resuscitation, surgical débridement, general supportive care, and antibiotic coverage. An intravenous combination of cefotaxime, 2 g four times a day, plus doxycycline, 100 mg two times a day, is recommended. This combination is synergistic in vitro. Alternative antimicrobials are ceftazidime and ciprofloxacin.

Grade A References

A1. Rabbani GH, Sack DA, Ahmed S, et al. Antidiarrheal effects of L-histidine supplemented rice-based oral rehydration solution in the treatment of male adults with severe cholera in Bangladesh: a double-blind, randomized trial. *J Infect Dis*. 2005;191:1507-1514.

A2. Leibovici-Weissman Y, Neuberger A, Bitterman R, et al. Antimicrobial drugs for treating cholera. *Cochrane Database Syst Rev*. 2014;6:CD008625.

A3. Saha D, Karim MM, Khan WA, et al. Single-dose azithromycin for the treatment of cholera in adults. *N Engl J Med*. 2006;354:2452-2462.

A4. Roy SK, Hossain MJ, Khatun W, et al. Zinc supplementation in children with cholera in Bangladesh: randomised controlled trial. *BMJ*. 2008;336:266-268.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Sugimoto JD, Koepke AA, Kenah EE, et al. Household transmission of Vibrio cholerae in Bangladesh. *PLoS Negl Trop Dis*. 2014;8:e3314.
2. Loharikar A, Newton AE, Stroika S, et al. Cholera in the United States, 2001-2011: a reflection of pattens of global epidemiology and travel. *Epidemiol Infect*. 2015;143:695-703.
3. Barzilay EJ, Schaad N, Magliore R, et al. Cholera surveillance during the Haiti epidemic—the first 2 years. *N Engl J Med*. 2013;368:599-609.
4. Cholera 2011. *Wkly Epidemiol Rec*. 2012;87:289-304.
5. Azman AS, Rudolph KE, Cummings DAT, Lessler J. The incubation period of cholera: a systematic review. *J Infect*. 2013;66:432-438.
6. Chowdhury F, Khan AI, Faruque ASG, Ryan ET. Severe, acute watery diarrhea in an adult. *PLoS Negl Trop Dis*. 2010;4:e898.
7. Harris JB, LaRocque RC, Qadri F, et al. Cholera. *Lancet*. 2012;379:2466-2476.
8. Luquero F, Grout L, Ciglenecki I, et al. Use of Vibrio cholerae vaccine in an outbreak in Guinea. *New Eng J Med*. 2014;370:2111-2120.
9. Khatib AM, Ali M, von Seidlein L, et al. Effectiveness of an oral cholera vaccine in Zanzibar: findings from a mass vaccination campaign and observational cohort study. *Lancet Infect Dis*. 2012;12:837-844.
10. Schaetti C, Weiss MG, Ali SM, et al. Cost of illness due to cholera, costs of immunization and cost effectiveness of an oral cholera mass vaccination campaign in Zanzibar. *PLoS Negl Trop Dis*. 2012;6:e1844.
11. Reyburn R, Deen JL, Grais RF, et al. The case for reactive mass oral cholera vaccinations. *PLoS Negl Trop Dis*. 2011;5:e952.
12. Newton A, Kendall M, Vugia DJ, et al. Increasing rates of vibriosis in the United States, 1996-2010: review of surveillance data from 2 systems. *Clin Infect Dis*. 2012;54(suppl 5):S391-S395.
13. Tena D, Arias M, Alvarez BT, et al. Fulminant necrotizing fasciitis due to *Vibrio parahaemolyticus*. *J Med Microbiol*. 2010;59:235-238.

REVIEW QUESTIONS

1. Cholera is an acute diarrheal disease caused by

A. *Vibrio cholerae* O1 and *Vibrio cholerae* O139
B. *Vibrio cholerae* O1 only
C. *Vibrio cholerae* O139 only
D. Non-O1 *Vibrio cholerae*
E. *Vibrio vulnificus*

Answer: A Cholera is the disease caused by both O1 *V. cholerae* and, since 1992, O139 *V. cholerae*. These two pathogens have the potential to cause local epidemics with pandemic potential.

2. The typical solute concentration of cholera stools, compared with plasma, is

A. Hypertonic
B. Hypotonic
C. Isotonic
D. Hyperosmolar
E. Hypo-osmolar

Answer: C Typical diarrhea of cholera is isotonic compared with plasma, having almost the same concentration of sodium.

3. The benefit of using effective antimicrobials in patients with severe cholera is

A. To reduce the excretion of vibrio
B. To reduce the duration of diarrhea and volume of stool by half
C. To reduce the duration of hospitalization by half
D. To reduce the need for intravenous treatment
E. To reduce the need for oral rehydration solutions

Answer: B Effective antimicrobials reduce the duration of diarrhea and the volume of stools by almost half compared with controls.

4. Intravenous fluids are indicated in patients with cholera under the following conditions:

A. In every patient with cholera irrespective of the degree of dehydration
B. In patients with some degree of dehydration
C. In patients with severe dehydration, in those who cannot tolerate the oral route, and in those with high stool output
D. In patients who do not respond to antimicrobials
E. In those thought to be infected by resistant pathogens

Answer: C Intravenous fluids should be restricted to those with severe dehydration, to those with lesser degrees of dehydration who cannot tolerate the oral route, and to those with high stool output.

5. A patient with chronic liver disease presents to the emergency department with the acute onset of fever after ingestion of oysters. The physical examination reveals a patient in shock with multiple erythematous lesions on the extremities. The most likely pathogen involved is

A. *Vibrio cholerae* O1
B. *Vibrio cholerae* O139
C. *Vibrio vulnificus*
D. *Salmonella enterica*
E. *Shigella dysenteriae* type 1

Answer: C *Vibrio vulnificus* is the most likely pathogen. It causes severe sepsis with soft tissue involvement in patients with chronic liver disease.

303

CAMPYLOBACTER INFECTIONS

BAN MISHU ALLOS

DEFINITION

Campylobacter jejuni is one of the most commonly recognized bacterial causes of diarrhea in developed and developing nations. More than 95% of campylobacters isolated in developed countries are *C. jejuni* or *Campylobacter coli.* However, other *Campylobacter* species are also associated with human disease.

The Pathogen

Campylobacters are motile, curved, gram-negative rods that are found in domestic and wild animals—especially poultry—all over the world. *C. jejuni*

is microaerophilic, requires 3 to 15% oxygen for growth, and is oxidase and catalase positive. It grows best at 42° C; however, other *Campylobacter* species that may also be pathogenic grow best at 37° C. The whole genome sequences for multiple *Campylobacter* species have been determined.

EPIDEMIOLOGY

C. jejuni infections are endemic in young children in developing nations,[1] where they may be isolated in up to 20% of children younger than 5 years with diarrhea. In developed nations, *Campylobacter* infections are among the most common bacterial causes of diarrhea in children and adults. The incidence of *C. jejuni* infection in the United States fell by more than 30% from 21.7 per 100,000 population in 1998 to 12.7 in 2008, but by 2012 the incidence had increased again to its highest level since 2000. The actual burden of disease caused by *Campylobacter* is probably much higher because even active surveillance systems substantially underreport the true incidence of infection. Epidemiologic studies have estimated that more than 2 million people in the United States are infected with *C. jejuni* each year. For reasons that are not clear, the incidence is highest in western states, such as California and Hawaii. Similarly high rates of infection are observed in Europe. In the United States, Europe, and Australia, *C. jejuni* infections show a substantial peak in warmer months. Such seasonality is not observed in tropical developing countries, perhaps because of the absence of extreme temperature variations.

The incidence of *Campylobacter* infections is highest in early childhood, an epidemiologic feature common to many food-borne bacterial pathogens. However, in the United States and other industrialized countries, the incidence of *Campylobacter* infections peaks again in early adulthood. The incidence of infection is also higher in men, a gender difference most pronounced in young adults.

Most human *C. jejuni* infections occur sporadically, with only a tiny fraction occurring as part of outbreaks. The dominant source of sporadic infections in both developed and developing countries is consumption or handling of poultry. Other sources of transmission in developed nations include foreign travel, contact with pets and other animals, contaminated drinking water, and consumption of unpasteurized milk.[2] Cross-contamination within a kitchen (e.g., use of the same utensils or cutting boards to prepare uncooked chicken and to chop fruit) has led to a variety of foods being implicated as sources of human *C. jejuni* infection. In contrast to sporadic infections, the most common source of *C. jejuni* outbreaks is unpasteurized milk; large waterborne outbreaks occasionally occur. Transmission of *C. jejuni* infection from ill food handlers is uncommon. Even in households in which an individual has culture-proven *C. jejuni* gastroenteritis, secondary transmission to other family members is unusual.

PATHOBIOLOGY

Persons become infected with *C. jejuni* as a result of orally ingesting the organism, usually in food or water. Factors that affect whether *Campylobacter* infection leads to illness include the dose of bacteria ingested, the virulence of the organism, and the specific immunity of the host to the ingested organism.[3] The minimum number of bacteria needed to cause illness varies between people, but it may be quite low; because *C. jejuni* is susceptible to gastric acidity, ingestion of very few organisms may cause illness if gastric pH is elevated as a result of illness or medication. The median incubation period is 2 to 4 days, although it may range from 1 to 7 days.

In early infection, *C. jejuni* multiplies in the bile-rich upper intestines; subsequently, tissue injury is seen in the jejunum, ileum, and colon. Gross inspection of the bowel shows a diffuse, bloody, edematous enteritis. Microscopic examination shows an inflammatory infiltrate consisting of neutrophils, mononuclear cells, and eosinophils in the lamina propria. The mucosal epithelium is ulcerated, and crypt abscesses may be seen. The pathologic appearance is nonspecific and may mimic ulcerative colitis or Crohn's disease.

Invasion of the epithelium by *C. jejuni* appears to be central to its pathogenesis, and many factors influence how *C. jejuni* adheres to and invades intestinal tissues. A superficial conserved antigen, PEB1, appears to be a major adhesin and is a target of the immune response to *C. jejuni* infection. Other factors contributing to the invasiveness and pathogenicity of *C. jejuni* that may be encoded by a virulence plasmid include type IV secretion systems and mechanisms that disrupt microtubules in host cells. The presence of the plasmid pVir in clinical isolates is significantly associated with bloody stools. *C. jejuni* also may invade in an actin- and microtubule-independent manner. Glycolipids and glycoproteins on the surface of *C. jejuni* are important in the organism's survival in the intestinal lumen and in pathogenesis because they have an impact on cell-to-cell interactions as well as on the host's immune response to infection. The bacteria's flagella facilitate its ability to colonize the gastrointestinal tract by promoting the organism's motility and chemotaxis. *C. jejuni* may produce extracellular toxins, but their role in pathogenesis has not been confirmed, with the possible exception of the cytoskeletal distending toxin (cdt), which may facilitate intracellular activities that lead to apoptosis.

Regardless of the organism's virulence, host factors are pivotal in affecting the clinical outcome of infection. In healthy volunteers fed a fixed dose of a single *C. jejuni* strain, a spectrum of illnesses develops. Patients infected with *C. jejuni* excrete the organism in feces for 2 to 3 weeks. In developing nations, where the level of immunity to *C. jejuni* is higher because of recurrent exposure, the period of convalescent excretion of *C. jejuni* is shorter.

After recovery from *Campylobacter* infection, at least short-term immunity develops. The decreasing illness-to-infection ratio with age seen in developing nations also suggests that individuals are acquiring immunity. Specific immunoglobulin A, G, and M antibodies in serum and immunoglobulin A antibodies in intestinal secretions develop in patients infected with *C. jejuni.* Patients with congenital or acquired hypogammaglobulinemia are at risk for severe or recurrent *C. jejuni* infections. Because the incidence of *C. jejuni* infection is markedly higher in persons infected with human immunodeficiency virus (HIV), cell-mediated immunity might also play a role in preventing and terminating infection.

CLINICAL MANIFESTATIONS

The clinical consequences of *Campylobacter* infection range from complete absence of symptoms to fulminant sepsis and death. In most cases, however, illnesses are brief and do not require hospitalization. In developed nations, detection of *C. jejuni* in the stool of asymptomatic persons is rare. However, in developing nations, where infections are endemic and recurrent infections occur frequently, asymptomatic infections are more common. In both developing and developed nations, persons infected with *C. jejuni* typically contract a diarrheal illness that resolves within a week. The case-fatality rate associated with this infection is low, about 0.05 death per 1000 infections, and is not surprisingly highest among the elderly and persons with comorbid conditions.

The gastroenteritis that is caused by infection with *C. jejuni* is clinically indistinguishable from that caused by other bacterial enteric pathogens, such as *Salmonella* (Chapter 308), *Shigella* (Chapter 309), or *Escherichia coli* O157:H7 (Chapter 304). The most common symptoms are diarrhea, malaise, fever, and abdominal pain (Table 303-1). Most patients with *C. jejuni* gastroenteritis experience at least 1 day with 10 or more stools; the diarrhea may be loose, watery, or bloody. Nausea is reported by some patients, but vomiting is less common. More than half the patients describe subjective fever. The abdominal cramping may be severe and is sometimes the predominant symptom. Although in most patients the symptoms resolve within 7 days, symptoms may persist in 10 to 20% of patients, and another 5 to 10% may experience a relapse.

Almost regardless of the nature of the symptoms, fecal leukocytes are found in 75% of infected patients; gross or occult blood is seen in 50%. Peripheral white blood cell counts may be elevated, but liver function test results, the hematocrit, and serum electrolyte values are usually normal.

TABLE 303-1 CLINICAL FEATURES OF *CAMPYLOBACTER* ENTERITIS DERIVED FROM OUTBREAKS IN WHICH MORE THAN 50 PATIENTS WERE INFECTED

SYMPTOM	MEDIAN FREQUENCY (%)	RANGE (%)
Fever	50	6-75
Diarrhea	84	52-100
Headache	41	6-69
Abdominal pain	79	56-99
Myalgia	42	28-59
Vomiting	15	1-42
Blood in feces	15	0.5-32

Modified from Blaser MJ, Engberg J. Clinical aspects of *Campylobacter jejuni* and *Campylobacter coli* infections. In: Nachamkin I, Szymanski CM, Blaser MJ, eds. *Campylobacter.* 3rd ed. Washington, DC: ASM Press; 2008:99-121.

Sigmoidoscopic examination reveals diffuse colonic inflammation, which is nonspecific.

Local complications of *C. jejuni* gastroenteritis are rare. In its most severe form, infection may lead to massive gastrointestinal hemorrhage or toxic megacolon. Infection of the biliary tract may result in obstructive hepatitis, cholecystitis (Chapter 155), or pancreatitis (Chapter 144). Other reported local complications include peritonitis (Chapter 142), splenic rupture, and exacerbations of inflammatory colitis. Bacteremia is detected in 1.5 per 1000 intestinal infections, with higher rates in persons who are immunocompromised or elderly, but transient bacteremia may be more common because blood cultures are infrequently obtained in patients with diarrhea and the bacteria are killed rapidly by normal human serum. Other extraintestinal complications, such as meningitis, endocarditis, osteomyelitis, and purulent arthritis, are rare.

Guillain-Barré syndrome (Chapter 420), which is a postinfectious complication of *C. jejuni* infection, occurs about once in every 2000 infections; between 30 and 50% of all cases may be triggered by a preceding *C. jejuni* infection.[4] Because the onset of neurologic symptoms occurs about 1 to 3 weeks after the onset of gastrointestinal symptoms, cross-reactivity between antibodies formed against the lipopolysaccharide and capsule of *C. jejuni* and proteins in peripheral nerve myelin or other glycolipids in peripheral nerves is probably the cause. Certain *C. jejuni* serotypes (O type 19, O type 41) are overrepresented in patients in whom Guillain-Barré syndrome develops after culture-documented *C. jejuni* infection. Other postinfectious complications of *C. jejuni* infection include reactive arthritis (seen mostly in persons with HLA-B27 histocompatibility antigens), uveitis, hemolytic-uremic syndrome, erythema nodosum, encephalitis, carditis, hemolytic anemia, and chronic gastrointestinal consequences such as irritable bowel syndrome, inflammatory bowel disease, and celiac disease.[5]

DIAGNOSIS

The diagnosis of *C. jejuni* infection should be considered in any patient with an acute febrile diarrheal illness. The diagnosis is established by culturing the organism from stool or tissue. Primary isolation of *Campylobacter* species from blood may take up to 14 days.

The presence of curved gram-negative rods on a Gram stain of stool is specific but only 50 to 75% sensitive for detecting *C. jejuni.* Examination of fecal specimens by dark-field microscopy is useful if it is done within 2 hours of passage; the characteristic darting motility of *Campylobacter* provides a presumptive diagnosis. Serum serologic studies of stool are currently available only as research tools. Use of polymerase chain reaction techniques for direct detection of organisms has been successful in research studies but has not yet been applied to the general clinical setting.

Differential Diagnosis

In patients with acute colitis and bloody diarrhea, especially those whose symptoms last longer than 1 week, *Campylobacter* enteritis may be mistaken for ulcerative colitis or Crohn's disease (Chapter 141). In such cases, it is critical to exclude infectious colitis before starting immunosuppressive therapy. In patients with severe abdominal pain, appendicitis may be suspected and unnecessary appendectomy may result (Chapter 142).

PREVENTION

Because the most common source of transmission of *C. jejuni* infection to humans in developed countries is by consumption and handling of poultry, interrupting this route of infection will probably have the greatest effect on reducing the burden of disease caused by *Campylobacter.* The nearly universal colonization of poultry flocks with *C. jejuni* makes eradication of the organism in chickens unlikely, but improvements in slaughtering plants appear to be reducing the level of contamination of products reaching humans. For the consumer, careful food preparation methods are critical; chicken must be cooked thoroughly. To avoid cross-contamination in the kitchen, cutting boards, knives, and other utensils used to prepare raw chicken should be washed with hot soapy water before being used to prepare foods eaten uncooked, such as fruits and vegetables. Person-to-person transmission of *Campylobacter* is not common; nevertheless, all persons with diarrhea, especially those who handle food, should wash their hands after using the bathroom. Travelers and campers should be cautioned against drinking untreated water. Many outbreaks of *C. jejuni* infection might also be avoided if persons abstain from drinking unpasteurized milk. Antibiotic prophylaxis for travelers is not advised. An effective anti-*Campylobacter* vaccine has not yet been developed.

TREATMENT

Rx

As is true for most patients with infectious or noninfectious diarrhea, the most important principle of treatment of *Campylobacter* gastroenteritis is restoration of proper hydration and electrolyte balance, typically with oral fluids. On occasion, intravenous fluids are needed, especially in elderly patients or young children. Most *C. jejuni* infections are self-limited and resolve without specific antibiotic treatment. Furthermore, treatment with antibiotics shortens the duration of illness by less than 48 hours.■ Prompt antimicrobial therapy is indicated for patients with high fever (>38.5° C), prolonged illness (>1 week), bloody stools, or worsening symptoms and for those who have relapsed. Antimicrobial treatment is also warranted in the elderly, infants, pregnant women, and persons who are immunocompromised, including those infected with HIV.

For many decades, the antibiotic of choice for the treatment of *C. jejuni* gastroenteritis has been erythromycin (500 mg twice daily for 5 days). A single 1-g dose of azithromycin is at least as effective,■ but it and clarithromycin are considerably more expensive. One concern with erythromycin, which is primarily metabolized by CYP3A4, is the risk of sudden cardiac death. The risk is increased five-fold when erythromycin is given with medications that inhibit CYP3A4. In patients taking one or more of these medications, azithromycin may be substituted for erythromycin.

Fluoroquinolones, carbapenems, aminoglycosides, and clindamycin may also be effective, but resistance to quinolones is now common in many parts of the world.[6] In general, rates of resistance to ampicillin, amoxicillin, and cephalosporins are too high for them to be useful in the treatment of *C. jejuni* infections.

Critically ill or septic persons with *Campylobacter* infection may benefit from carbapenems or aminoglycosides, agents to which campylobacters are exquisitely susceptible, with resistance rates consistently less than 1%. In contrast, persons with persistent or relapsing infection, especially those who are immunocompromised, may require prolonged use (sometimes months) of antibiotics. In the absence of continuing sepsis, oral agents may be used.

OTHER *CAMPYLOBACTER* SPECIES

Campylobacter fetus may cause systemic and diarrheal illnesses in compromised hosts and diarrheal illnesses in normal hosts.[7] Most *C. fetus* strains, unlike *C. jejuni*, are not susceptible to the lethal effect of normal human serum because they possess a protein capsule (S layer). In immunocompromised persons, *C. fetus* can cause extraintestinal illnesses such as bacteremia, vascular infections, and meningitis. *C. fetus* infection may also cause perinatal infection and fetal loss. Prolonged treatment with erythromycin plus either imipenem, meropenem, an aminoglycoside, or a third-generation cephalosporin is indicated for serious *C. fetus* infections.

Campylobacter upsaliensis may cause acute or chronic diarrhea in healthy or immunocompromised persons. The organism is frequently isolated from dogs with diarrhea, which could be a source for transmission to humans. Some *C. upsaliensis* strains are resistant to erythromycin, but most are susceptible to fluoroquinolones, doxycycline, third-generation cephalosporins, and amoxicillin-clavulanate.

Campylobacter hyointestinalis was first recognized as a cause of proliferative enteritis in swine; *Campylobacter lari* is most often cultured from gulls and other birds. Both organisms have now been identified as rare causes of watery diarrhea and abdominal cramping in immunocompetent children and adults. Most infected patients do not require antimicrobial therapy; all isolates studied in vitro have been susceptible to erythromycin.

Campylobacter concisus, long believed to be part of the microbiota of healthy persons, is now considered a possible cause of human gastrointestinal illness. An increasing body of evidence also has linked *C. concisus* infection with childhood Crohn's disease.

Helicobacter cinaedi and *Helicobacter fennelliae*, once called *Campylobacter*-like organisms, are causes of proctocolitis or enterocolitis and have also been reported to cause bacteremia in immunocompromised patients. The organisms are frequently resistant to erythromycin; fluoroquinolones are considered the treatment of choice in patients who require antimicrobial therapy. The organisms are also susceptible to third-generation cephalosporins, aminoglycosides, and carbapenems.

Other *Campylobacter* or related species that have been associated with human illness include *Campylobacter mucosalis, Campylobacter doylei, Campylobacter curvus, Campylobacter insulaenigrae, Campylobacter rectus, Campylobacter helveticus, Arcobacter butzleri*, and *Arcobacter cryaerophila*. Illnesses include diarrhea and localized infections, presumably as a result of transient bacteremia from intestinal sources. Recently identified *Campylobacter* species

that may be clinically relevant include *Campylobacter ureolyticus, Campylobacter troglodytis, Campylobacter lari* subspecies *concheus,* and *Campylobacter peloridis*. New pathogenic species of *Campylobacter* are being identified with some regularity.

PROGNOSIS

Even in critically ill patients, 1 week of therapy is generally sufficient to eradicate infection. *Campylobacter* infections in HIV-positive patients may be more severe, persist, recur, and be antibiotic resistant. More severe and extraintestinal illness is also more likely to occur in patients with acquired or congenital hypogammaglobulinemia. Most *C. jejuni* gastrointestinal infections in pregnant women are mild and self-limited, with no severe consequences for the mother or baby. However, if bacteremia develops in the mother, placental infection and fetal death may ensue. Infection during the third trimester can also cause neonatal sepsis and death if the woman is excreting *Campylobacter* in her stool at the time of delivery.

Grade A References

A1. Ternhag A, Asikainen T, Giesecke J, et al. A meta-analysis on the effects of antibiotic treatment on duration of symptoms caused by infection with *Campylobacter* species. *Clin Infect Dis*. 2007;44:696-700.

A2. Vukelic D, Trkulja V, Salkovic-Petrisic M. Single oral dose of azithromycin versus 5 days of oral erythromycin or no antibiotic in treatment of Campylobacter enterocolitis in children: a prospective randomized assessor-blind study. *J Pediatr Gastroenterol Nutr*. 2010;50:404-410.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Platts-Mills JA, Liu J, Gratz J, et al. Detection of *Campylobacter* in stool and determination of significance by culture, enzyme immunoassay, and PCR in developing countries. *J Clin Microbiol*. 2014; 52:1074-1080.
2. Sarkar SR, Hossain MA, Paul SK, et al. *Campylobacteriosis*—an overview. *Mymensingh Med J*. 2014;23:173-180.
3. Epps SV, Harvey RB, Hume ME, et al. Foodborne *Campylobacter*: infections, metabolism, pathogenesis and reservoirs. *Int J Environ Res Public Health*. 2013;10:6292-6304.
4. Keithlin J, Sargeant J, Thomas MK, et al. Systematic review and meta-analysis of the proportion of *Campylobacter* cases that develop chronic sequelae. *BMC Public Health*. 2014;14:1203.
5. Riddle MS, Gutierrez RL, Verdu EF, et al. The chronic gastrointestinal consequences associated with campylobacter. *Curr Gastroenterol Rep*. 2013;14:395-405.
6. Wieczorek K, Osek J. Antimicrobial resistance mechanisms among *Campylobacter*. *Biomed Res Int*. 2013;2013:340605.
7. Wagenaar JA, van Bergen MA, Blaser MJ, et al. *Campylobacter* fetus infections in humans: exposure and disease. *Clin Infect Dis*. 2014;58:1579-1586.

REVIEW QUESTIONS

1. Which one of the following statements best characterizes the epidemiology of *Campylobacter jejuni*?

A. *C. jejuni* gastroenteritis is common in developed countries such as the United States but uncommonly causes illness in developing nations.
B. In developing nations, the incidence of *C. jejuni* infection peaks during winter and spring.
C. In developed nations, the incidence of *C. jejuni* is higher in men than in women.
D. Most *C. jejuni* infections occur as a result of person-to-person transmission.
E. Most *C. jejuni* infections occur as part of outbreaks rather than as sporadic infections.

Answer: C *C. jejuni* infections are common in both developed and developing nations. In developed nations, infections peak in late summer and fall; in developing nations, no seasonality is observed. The incidence of infection in developed nations is higher in young men than in young women. Person-to-person transmission of infection is possible but not common. Most infections are sporadic. (See section on epidemiology.)

2. Which one of the following statements about the pathogenesis of *C. jejuni* is true?

A. Microscopic examination of the bowel will show an inflammatory infiltrate consisting of neutrophils, mononuclear cells, and eosinophils in the lamina propria.
B. The incubation period is short, sometimes less than 4 hours from exposure to symptom onset.
C. Extracellular toxins are important in the pathogenesis of colitis associated with *C. jejuni* infections.
D. Development of long-term immunity after a single infection makes recurrent infections unlikely.
E. *C. jejuni* exerts it pathogenic effects on superficial intestinal mucosa only; invasion to deeper tissues does not occur.

Answer: A The incubation period is usually 2 to 4 days. Extracellular toxins are elaborated but are not important in pathogenesis. Immunity to infection is short-lived; one infection does not protect against future infections. (In developing nations where recurrent infections are common, asymptomatic infections may occur in adults.) *C. jejuni* infection is capable of invasion; that is why bacteremia occasionally occurs. (See section on pathobiology.)

3. The most common clinical manifestation of *C. jejuni* infection in developed nations is

A. Asymptomatic infection
B. Chronic, watery diarrhea
C. Acute, self-limited diarrhea
D. Guillain-Barré syndrome
E. Sepsis

Answer: C Asymptomatic infections are unusual in developed nations, although they are more common in developing nations. Diarrheal illness is usually self-limited; chronic illness is not a typical feature of infection with this pathogen. Guillain-Barré syndrome is a rare complication of infection. Sepsis, although possible, also is rare. (See section on clinical manifestations.)

4. Which of the following preventive measures would be most likely to reduce the incidence of *C. jejuni* infections in developed nations?

A. More consistent use of currently available vaccines
B. Proper treatment of water from rivers or streams before being consumed by hikers and campers
C. Exclusion of persons with diarrhea from work as food handlers
D. Requirement that all eggs served in hospitals or other health care facilities be pasteurized
E. Avoidance of eating undercooked chicken in both homes and restaurants

Answer: E There is not currently a commercially available vaccine. Consumption of untreated water from rivers and streams is a risk factor but is not as important a contributor to infection as is eating undercooked chicken. Although all persons with an acute diarrheal illness should be excluded from work as food handlers, transmission from infected food handlers is not a big contributor to the incidence of *C. jejuni* infections. For example, transmission within households of infected persons is uncommon. Eggs served in hospitals should be pasteurized because of the risk of *Salmonella* infections; however, this is not likely to have a big impact on rates of *C. jejuni* infections. Consumption of undercooked chicken is the single most common risk factor for development of *C. jejuni* infections. (See section on prevention.)

5. Which one of the following antibiotics is the best choice for treatment of gastroenteritis caused by *C. jejuni* infection?

A. Amoxicillin
B. Azithromycin
C. Cephalexin
D. Gentamicin
E. Chloramphenicol

Answer: B Penicillins and cephalosporins are not effective in the treatment of *C. jejuni*. Gentamicin is effective but not an appropriate choice for an uncomplicated gastrointestinal infection. Chloramphenicol also is effective, but the side effects associated with this drug make it a less accepted choice than azithromycin. (See section on treatment.)

304 *ESCHERICHIA COLI* ENTERIC INFECTIONS

THEODORE S. STEINER

DEFINITION

Bacteria belonging to the species *Escherichia coli* are a normal component of the intestinal microbiota (Chapter 279). The majority of *E. coli* are harmless commensals, but specific isolates have acquired pathogenicity genes that enable them to cause diseases, including urinary tract infections, bacteremia, meningitis, and diarrheal illness. One particular challenge to the clinician and microbiology laboratory is how to distinguish these pathogenic *E. coli* from harmless commensal strains to better guide diagnosis and treatment.

Enteric infections caused by *E. coli* may involve the small intestine, colon, or both, depending on the organism's genetic codes for virulence traits. These virulence traits include a variety of toxins, adherence factors, and secreted mediators that work together to perturb host intestinal physiology. Specific combinations of these factors produce six major pathotypes of diarrheogenic *E. coli*: enterotoxigenic (ETEC), enteroinvasive (EIEC), enterohemorrhagic (EHEC), enteropathogenic (EPEC), enteroaggregative (EAEC), and diffusely adherent (DAEC). In addition, these pathotypes can overlap; for example, some strains can express Shiga-like toxins that are characteristic of EHEC without the usual associated adherence factors; these are collectively known as shigatoxigenic *E. coli* (STEC). Taken together, diarrheogenic *E. coli* not only constitute the major category of bacterial enteric pathogens but also provide important scientific models for the many ways in which enteric pathogens can cause disease.

The Pathogen

E. coli is a small catalase-positive, oxidase-negative, gram-negative bacillus in the family Enterobacteriaceae. It characteristically reduces nitrates, ferments glucose and usually lactose, and is either motile (with peritrichate flagella) or nonmotile. It exhibits a positive methyl red reaction and negative reactions with Voges-Proskauer, urease, phenylalanine deaminase, and citrate agents. *E. coli* is the predominant member of the Gammaproteobacteria in the intestinal tract of humans and other mammals, although it is greatly outnumbered by members of other bacterial phyla, which largely consist of strict anaerobes.

As with other gram-negative organisms, the lipopolysaccharide cell wall of *E. coli* contains immunostimulatory lipid A attached to a core oligosaccharide chain. Most *E. coli* have immunogenic carbohydrate chains known as O antigens attached to this core glycolipid to produce 173 O serogroups. There are also at least 56 distinct flagellar (H) antigens based on variable domains of the flagellin gene. Some 80 variably heat-labile capsular (K) antigens have also been described. These O, H, and K antigen combinations have allowed serotyping of thousands of different strains, which historically was the simplest way to distinguish them. Whereas serotypes are sometimes useful in identifying specific pathotypes of *E. coli*, there are numerous adherence, enterotoxic, cytotoxic, and invasiveness factors that may be gained or lost by a particular serotype because they are characteristically encoded on transmissible genetic elements such as plasmids or bacteriophages. It is these factors that convey disease pathotype because they allow colonization and perturbation of host intestinal physiology. Indeed, molecular analysis available during the past decade has shown that commensal and pathogenic *E. coli* cluster into phylogenetic groups that are often independent of O:H serotype. Nevertheless, relatively few O serogroups tend to predominate in the normal human colon (O groups 1, 2, 4, 6, 7, 8, 18, 25, 45, 75, and 81), whereas others (Table 304-1) tend to be associated with specific virulence traits and thus different types of pathogenesis in the intestine.

Well-established mechanisms of *E. coli* pathogenesis include secretion of enterotoxins (ETEC), *Shigella*-like tissue invasion (EIEC), and epithelial necrosis as a result of Shiga-like toxins (SLTs SLT-1/2 or Stx1/Stx2) causing food-borne hemorrhagic colitis (EHEC and STEC) (see Table 304-1). By comparison, the classically recognized EPEC serotypes are neither enterotoxigenic nor invasive but rather attach and efface the epithelium. Still other types of "enteroadherent" *E. coli* exhibit aggregating (EAEC) or diffuse adherence (DAEC) traits, and EAEC in particular is associated with prolonged diarrhea in children in tropical developing areas, in patients infected with human immunodeficiency virus (HIV), and in acute diarrhea in outbreak settings and travelers from developed areas.

EPIDEMIOLOGY

Part of the challenge in studying the epidemiology of enteric *E. coli* infections is that, with the exception of EHEC/STEC strains, they are not identified in routine microbiology procedures in most clinical laboratories. In addition, the diagnostic methodologies have evolved, making it difficult to compare older and more recent studies. Nevertheless, several clear epidemiologic patterns have been revealed. In addition, specific single-nucleotide polymorphisms in the lactoferrin, osteoprotegerin, CD14, and interleukin-8 genes are associated with traveler's diarrhea caused by ETEC, EAEC, or both.[1]

Enteric *E. coli* infections are acquired by the fecal-oral route, although the fecal sources and infectivity differ among the pathotypes. It is believed that a human reservoir is required for most recognized types of EPEC and ETEC, although domestic dogs and cats can also harbor human pathogenic strains.[2] Different ETEC strains can also be important veterinary pathogens, especially in calves and piglets, but the attachment and virulence traits of animal strains are different from those of strains that infect humans. The infectious dose of ETEC in volunteers is 10^6 to 10^{10} organisms, meaning that it usually requires multiplication in contaminated food or water vehicles for its transmission, rather than spreading directly from person to person. Heavy contamination with ETEC has been documented in foods prepared in homes and restaurants and by street vendors as well as in drinking water in many tropical areas. Contaminated water and food probably represent the major sources of their acquisition, primarily in warm or wet seasons.

As with most diarrheal illnesses, the highest age-specific attack rates of ETEC are found in young children, especially at the time of weaning, when ETEC accounts for anywhere from 3% to 39% (average, 13%) of acute diarrheal illnesses, depending on the population studied. Like immunologically inexperienced young children, a traveler visiting tropical areas has a 30 to 50% chance of acquiring traveler's diarrhea (Chapters 283 and 286) during a 2- to 3-week stay unless untreated water or ice and uncooked foods such as salads are strictly avoided. The most commonly recognized pathogen associated with traveler's diarrhea in most tropical areas of the world is ETEC that produces the heat-stable toxin STa, the heat-labile toxin LT, or both. A close second to ETEC as a cause of traveler's diarrhea is EAEC, now reported in 19 to 33% of affected travelers to India or Mexico.

Typical EPEC strains have been recognized primarily in poor urban areas, especially among hospitalized infants in their first year of life, with apparent

TABLE 304-1 DIFFERENT TYPES OF ENTERIC *ESCHERICHIA COLI* INFECTIONS

TYPE	MECHANISM	PREDOMINANT O SEROGROUPS	GENETIC CODE	DETECTION	CLINICAL SYNDROMES
ENTEROTOXIGENIC *E. COLI* (ETEC)					
Heat-labile toxin (LT)	Activates intestinal adenylate cyclase	6, 8, 11, 15, 20, 25, 27, 63, 80, 85, 139	Plasmid	Gene probe, PCR for LT	Watery diarrhea, traveler's diarrhea
Heat-stable toxin (STa: STh or STp)	Activates intestinal guanylate cyclase	12, 78, 115, 148, 149, 153, 155, 166, 167	Plasmid (transposon)	EIA, suckling mice, 6-hour ileal loop assay, gene probes, PCR	Watery diarrhea, traveler's diarrhea
ENTEROINVASIVE *E. COLI* (EIEC)					
	Cell invasion and spread	11, 28ac, 29, 124, 136, 144, 147, 152, 164, 167	Plasmid (140 MIa, pWR110)	Sereny test, gene probe, PCR for *ipaH*	Inflammatory dysentery
SHIGATOXIGENIC *E. COLI* (STEC)					
Enterohemorrhagic (EHEC)	Shiga-like toxins (SLTs/Stxs) and attaching/effacing ability	26, 39, 113, 121, 128, 139, 145, 157, occ 55, 111	SLT phages and adhesin plasmids; type III secretion system	EIA or PCR for SLT, serotype, cell adhesion with pedestal formation; Vero cell cytotoxicity; sorbitol agar; PCR for *eae*	Afebrile, bloody diarrhea; HUS in some cases
Non-EHEC STEC	SLTs only without attaching/effacing	26, 111, 103, 121, 45, 104, 145	SLT phage; may possess other virulence traits (e.g., O104:H4 EAEC)	SLT EIA or PCR; negative for *eae*	Hemorrhagic colitis, HUS, or benign watery diarrhea
ENTEROADHERENT *E. COLI*					
Typical enteropathogenic (EPEC)	Attach, then efface the mucosa	55, 86, 111, 114, 119, 127, 142	Bundle-forming pili on plasmid and chromosomal LEE	Serotype, focal HEp2/HeLa cell adhesion, pedestal formation, gene probe or PCR *eae*	Infantile diarrhea in developing areas
Atypical enteropathogenic (EPEC)	Attaching and effacing but different microcolony formation	26, 55, 86, 111, 119, 125, 128	Possess the LEE but not bundle-forming pili	Gene probe or PCR for LEE; cell adherence (variable)	Infantile and animal diarrhea in developed areas
Enteroaggregative (EAEC)	Colonize in aggregates; toxins (EAST, Pet), biofilm formation	3, 15, 44, 51, 77, 78, 91	Plasmid (AA); chromosome (Pic/ShET and type VI secretion)	HEp2 cell adherence; AA probe; PCR for *aggR* or other virulence genes; biofilm formation	Endemic persistent diarrhea, acute traveler's diarrhea, sporadic acute diarrhea
Diffusely adherent (DAEC)	Colonize (F 1845 afimbriate adhesin)	86, 75, 15	Chromosomal/plasmid	HEp2 cell adherence; DA gene probe/PCR	Persistent diarrhea in children >18 months old

AA = aggregative adherence; DA = diffuse adherence; EIA = enzyme immunoassay; HUS = hemolytic-uremic syndrome; LEE = locus of enterocyte effacement; PCR = polymerase chain reaction.

cross-infection in hospital nurseries. Although sporadic cases still occur, nosocomial outbreaks of EPEC diarrhea during the summer appear to have become less common and less severe in industrialized countries in the past few decades. "Atypical" EPEC strains lacking certain virulence factors have tended to predominate in developed areas but can be found in developing areas as well.[3]

EHEC frequently colonizes commercial livestock but does not infect them. EHEC (O157:H7 and others) infections were first attributed to eating undercooked hamburgers, but subsequent large outbreaks have been associated with contamination of unpasteurized apple juice, spinach, seed sprouts, and other vegetable items. Approximately 600 people were infected in a large outbreak caused by contamination of the domestic water supply in Walkerton, Ontario, in 2000. More recently, a 2011 outbreak of shigatoxigenic EAEC strain O104:H4 (ST/EAEC) associated with fenugreek sprouts sickened almost 4000 people in Europe.[4] In addition, the low infectious dose of EHEC O157:H7 means that person-to-person spread can occur, leading to secondary cases. Secondary cases of ST/EAEC O104:H4 may also have occurred but appear to be very rare. EHEC and STEC infections are especially alarming because of the risk of hemolytic-uremic syndrome (HUS) (Chapter 172). HUS can be fatal despite antimicrobial therapy, which in some instances may actually induce SLT production from bacteriophage carried within the organism and hence is generally not recommended. Patients who recover from HUS may also suffer chronic kidney injury as a result.

The natural reservoir of EAEC is not known, but outbreaks have been traced to contaminated food, and live organisms can be found in drinking water, table salsa, and other consumable items in endemic tropical areas. Volunteer studies demonstrated that a high infectious dose is required for acquisition of EAEC, suggesting that direct person-to-person spread may be difficult. In addition to its role in traveler's diarrhea, EAEC is an important cause of both acute diarrhea and persistent diarrhea and malnutrition, especially in children in tropical areas and in patients with HIV/AIDS. It was also shown in several studies to be a major cause of sporadic diarrhea in the United States and the United Kingdom.

Limited data on EIEC suggest that infectious doses are relatively high, but as with ETEC infections, adequate numbers of organisms have readily been spread in food with high attack rates in outbreak situations. This distinguishes EIEC epidemiologically from *Shigella*, which is easily spread person to person as well as in contaminated food and water.

DAEC remains the least well understood pathotype and has not consistently been found more often in diarrheal cases than in controls. Nevertheless, some studies have shown a clear association with acute diarrhea in developing areas, particularly in children 1 to 4 years of age. Part of the difficulty is the heterogeneity of strains, some of which express different types of adhesins and different groups of virulence traits, leading to inconsistent pathogenicity.

PATHOBIOLOGY

The pathogenesis of enteric *E. coli* infection begins with ingestion of the organism in contaminated food or water or rarely direct person-to person spread, in the case of EHEC. It then faces the normal gastric acid barrier. Both ETEC and EIEC appear to be sensitive to gastric acid; neutralization of gastric acid reduces the infectious dose by 100- to 1000-fold. Hypochlorhydria increased the risk of EPEC diarrhea in a volunteer study. Whereas EHEC expresses acid tolerance factors that may facilitate its survival in the stomach, hypochlorhydria was shown to be a risk factor for HUS in one study.

After ingestion and passage through the stomach, enteric *E. coli* colonize the involved part of the intestinal tract using specialized adhesins and the coordinated expression of virulence traits. This can lead to toxin production, intracellular invasion, or other disruptions of host cell physiology. These virulence traits may be shared among different enteric *E. coli* as well as related enteric pathogens, and it is their combination that leads to the characteristic pathogenic and clinical features of infection. The incubation period between ingestion and symptom development varies from pathogen to pathogen. For example, it averaged 14 hours for EAEC and 2 days for ETEC in volunteer studies; epidemiologic studies found an average incubation period of 3 to 4 days for EHEC O157:H7 but 8 days for ST/EAEC O104:H4.

ETEC colonizes the upper portion of the small bowel using fimbriate or fibrillar surface proteins known as colonization factor antigens. The colonization factor antigens bind the organism to cell surface receptors on enterocytes. Whereas this colonization itself can lead to mild inflammatory changes in the epithelium, the majority of ETEC illness is due to its enterotoxins. LT, with a molecular weight of about 86,000, has a binding and active subunit and, like the closely related choleratoxin, binds to a monosialoganglioside (GM_1) receptor. Also like choleratoxin (Chapter 302), the active subunit is an enzyme that ADP-ribosylates the regulatory subunit of adenylate cyclase, leading to constitutive production of cyclic adenosine monophosphate. The consequently increased chloride secretion and reduced sodium absorption combine to cause net isotonic electrolyte loss that can be as great as 1 liter/hour. Other human ETEC strains produce heat-stable toxin (STa), which is a much smaller molecule than LT (18 to 19 amino acids) and activates intestinal particulate guanylate cyclase. Like cyclic adenosine monophosphate, the cyclic guanosine monophosphate thus formed also causes net secretion. The roles of other enterotoxins, such as LTII, EAST, EIET, and others seen in ETEC, EAEC, and EIEC, respectively, are unclear at present. Both the colonization traits and production of enterotoxin are encoded on transmissible plasmids. Besides the complications of dehydration, the only significant pathologic change with ETEC is depletion of mucus from intestinal goblet cells.

EIEC, like the closely related *Shigella*, can invade and multiply in epithelial cells, cause experimental conjunctivitis in guinea pigs (known as the Sereny test), and produce inflammatory colitis and dysenteric or bloody diarrhea. As seen with shigellosis, a striking inflammatory response occurs, with numerous polymorphonuclear leukocytes in stool. The colon shows patchy, acute inflammation in the mucosa and submucosa with focal denuding of the surface epithelium, usually without deeper invasion or systemic spread. Although epithelial cell invasiveness in both EIEC and *Shigella* appears to be encoded on a large 120- to 140-Md plasmid, several chromosomal determinants, including the O antigen, are crucial for full invasive virulence.

Typical EPEC strains are well-established causes of infantile diarrhea. They express both plasmid-encoded localized adherence to epithelial cells (through specialized bundle-forming pili) and chromosomally mediated attachment and effacement of microvilli. The latter is characterized by the formation of cellular pedestals that hold the bacteria intimately to the cell surface. These changes in the host epithelia are mediated through protein effectors injected directly into host cells by a specialized type III secretion system encoded on the chromosomal locus of enterocyte effacement. These secreted effectors cause cellular changes that lead to villous atrophy, mucosal thinning, inflammation in the lamina propria, and variable crypt cell hyperplasia. These morphologic changes are associated with a reduction in mucosal brush border enzymes and may contribute to the impaired absorptive function and diarrhea.[5] Atypical EPEC strains are generally defined as those expressing the locus of enterocyte effacement but not the bundle-forming pili; they maintain the ability to attach/efface and cause disease.

EHEC, most notably serotype O157:H7 but also serogroups O26, O39, and others, cause type III secretion-dependent intimate adherence and microvillous effacement like EPEC, but they also produce SLTs that are responsible for the characteristic colonic mucosal disruption and hemorrhage as well as the complication of HUS. These toxins bind to the Gb3 surface ganglioside, leading to internalization and enzymatic inactivation of ribosomes, halting protein synthesis. Gb3 is highly expressed on vascular endothelial cells in the colon, kidney, and brain, which may explain the predilection for HUS to affect these organs. Organisms that produce SLTs without intimate adherence and pedestal formation are known as STEC or VTEC (Shiga-toxigenic or verotoxigenic *E. coli*), and often they lack the other virulence traits necessary for colonization and disease production. A notable exception was the ST/EAEC O104:H4 that caused the 2011 European outbreak; this strain was essentially an EAEC that also expressed SLT. The terms EHEC, STEC, and VTEC were often used interchangeably in the older literature, until the important role of the attaching/effacing ability of true EHEC strains like O157:H7 was understood.

EAEC is defined by a characteristic aggregative adherence pattern to cells and the substrata associated with biofilm formation. This adherence requires a large plasmid known as the AA plasmid, which encodes specialized aggregative adherence fimbriae and other virulence genes, including a serine protease autotransporter toxin known as Pet and an antiaggregative protein called dispersin. Chromosomal virulence traits include a mucinase, Pic, and a second enterotoxin, ShET, as well as type VI secretion systems. Limited studies on human infections with EAEC suggest that the organisms do not intimately adhere or invade but reside within a biofilm at the epithelial surface, where secreted factors contribute to a damaging host inflammatory response.

The fundamental pathogenesis of DAEC is still an area of active investigation but appears to depend on direct interactions between specialized adhesins (Afa/Dr) and host membrane proteins such as CD55 (decay-accelerating factor) or carcinoembryonic antigen. Many DAEC strains are closely related to uropathogenic *E. coli* with similar virulence traits. Some also express other virulence factors, including the serine protease autotransporter toxin SAT and a type III secretion system.

Host risk factors for diarrheogenic *E. coli* infection differ among the various bacterial pathotypes but in general include age, recent antibiotic use, and loss of gastric acid.

CLINICAL MANIFESTATIONS

The clinical manifestations of enteric *E. coli* infections differ among the pathotypes. ETEC infections generally produce watery diarrhea, particularly in young children and travelers to tropical or developing areas. Diarrhea may range from mild to severe and cholera-like; it may be life-threatening, especially in small children and elderly individuals, who are particularly prone to dehydration, undernutrition, and electrolyte imbalance (especially hypokalemia and acidosis). Other characteristic symptoms include malaise, abdominal cramping, anorexia, and occasionally nausea, vomiting, or low-grade fever. The illness is generally self-limited to 1 to 5 days and rarely extends beyond 10 to 14 days. Infections with ETEC that produce both ST and LT or ST alone may be more severe than those caused by ETEC that produce only LT. The persistence of impaired mucosal absorptive capacity for 1 to 3 weeks may further compound the cycle of malnutrition that complicates diarrheal illnesses in children in developing, tropical areas.

Infection with EIEC is characterized by inflammatory colitis, often with abdominal pain, high fever, tenesmus, and bloody or dysenteric diarrhea, essentially like that seen with *Shigella*. The incubation period is usually 1 to 3 days, with the duration generally self-limited to 7 to 10 days.

Outbreaks of EPEC infection in newborn nurseries have ranged from mild transient diarrhea to severe and rapidly fatal diarrheal illnesses, especially in premature or otherwise compromised infants. The more severe illnesses appear to have been more common in industrialized countries before 1950. However, more recent outbreaks and sporadic cases are well documented.

Hemorrhagic colitis associated with EHEC classically begins with watery diarrhea that quickly turns grossly bloody, with a conspicuous absence of fever or inflammatory exudate in stool but with significant abdominal pain. Although this diarrheal illness is self-limited, potentially fatal HUS or thrombotic thrombocytopenic purpura subsequently develops in a significant number of children and older adults (Chapter 172). Outbreaks of hemorrhagic colitis secondary to EHEC in nursing homes or other institutions may be common and severe. The incubation period in two outbreaks has been 3 to 4 days (range, 1 to 7 days), and the illness is characteristically self-limited to 5 to 12 days (mean, 7.8 days). The clinical manifestations in the ST/EAEC O104:H4 outbreak were similar, although rates of HUS were significantly higher (more than 20%) and women were disproportionately affected.[6]

EAEC has been associated with persistent diarrhea and malnutrition in children in developing areas, in HIV/AIDS patients, and in travelers who experience diarrhea (especially those genetically predisposed to greater inflammatory responses). No characteristic clinical features of EAEC have been consistently identified, although some outbreaks were associated with bloody diarrhea and several studies suggest that elevated inflammatory markers in stool are fairly common. DAEC has also been associated with diarrhea with no particular identifying features in children older than 18 months.

DIAGNOSIS

With the exception of EHEC, definitive etiologic diagnosis of *E. coli* diarrhea requires documentation of a specific virulence trait or serotype, which requires specialized immunologic tests, tissue culture, animal bioassay, and molecular testing[7] that are usually available only in research and reference laboratories. Other than for EHEC, such tests are rarely cost-effective or clinically indicated, except in outbreak or research situations. Fortunately, a probable diagnosis can often be suspected by the clinical and epidemiologic setting.

EHEC O157:H7 can be identified with reasonable accuracy by culture on sorbitol-MacConkey agar to identify nonfermenting colonies. However, it has long been recommended that any stool sample with visible blood should also be tested specifically for SLTs by enzyme-linked immunosorbent assay,

polymerase chain reaction, or other molecular methods, which can identify non-O157 serotypes and rare sorbitol-fermenting O157 strains.[8] Many experts recommend that all stools submitted for culture be tested in this way. In hemorrhagic colitis due to EHEC, sigmoidoscopy, which is rarely indicated, generally reveals only moderately hyperemic mucosa, and barium enema or CT scan may show a thumbprint pattern of segmental or diffuse colonic wall thickening. Some patients have superficial ulceration with mild neutrophil infiltration in the edematous submucosa. These changes are not pathognomonic.

Differential Diagnosis

Numerous other causes of diarrhea must be considered, depending on the clinical circumstances (Chapters 140 and 283). For example, self-limited, noninflammatory diarrhea in tropical, developing areas is most likely due to ETEC, EAEC, rotaviruses (young children), or noroviruses (older children and adults) (Chapter 380). Noninflammatory diarrhea in older children or adults in temperate areas is more likely to be due to noroviruses (Chapter 380). *Vibrio* infections (Chapter 302) are common in areas endemic for cholera or in any coastal area where inadequately cooked seafood may be eaten. If noninflammatory diarrhea persists beyond a week, especially with weight loss, other possibilities include *Giardia lamblia* (Chapter 351), *Cryptosporidium* (Chapter 350), *Cyclospora* (Chapter 353), and microsporidial infection (Chapter 353). In outbreaks of food poisoning, *Staphylococcus aureus* (Chapter 288), *Clostridium perfringens* (Chapter 296), and *Bacillus cereus* should be considered.

Inflammatory colitis with high fever and tenesmus as well as leukocytes, mucus, and blood in the stool may well be due to EIEC but should prompt a stool culture for more common invasive pathogens, such as *Campylobacter jejuni* (Chapter 303), *Shigella* (Chapter 309), *Salmonella* (Chapter 308), *Yersinia enterocolitica* (Chapter 312), or noncholera *Vibrio* (Chapter 302). Any patient with diarrhea and a history of recent antibiotic use, gastrointestinal surgery, or parturition should be screened for toxigenic *Clostridium difficile* (Chapter 296). EHEC should be strongly considered in any case of bloody diarrhea, particularly in the absence of fever; it is recommended that laboratories now routinely screen for this pathogen in all stool cultures, and they should automatically screen any grossly bloody samples. Ischemic colitis and cytomegalovirus colitis can mimic EHEC but should occur only in people at risk (vascular disease and immune compromise or inflammatory bowel disease, respectively)

TREATMENT

Rx

As with all diarrheal illnesses, the primary treatment for most *E. coli* diarrhea is replacement and maintenance of water and electrolytes, usually with a simple oral rehydration solution that uses the intact, sodium-coupled glucose or amino acid absorption (or both) to replace the fluid losses.[9] Oral rehydration solution should be given ad libitum with free water, and in breast-fed infants, continued breast-feeding and early refeeding can compensate for the nutritional losses without an adverse effect on diarrhea output. Zinc supplementation is also recommended for diarrhea in children older than 6 months in developing areas where zinc deficiency is common as it significantly reduces diarrhea volume. The enkephalinase inhibitor racecadotril, which is available in Europe but not currently in North America, also reduces diarrheal volume in children with acute gastroenteritis and is as effective as loperamide in adults but less likely to cause constipation. Certain probiotic preparations have been shown in small studies to improve symptoms when they are added to oral rehydration solution in children with infectious diarrhea but are not yet universally recommended. Antimotility agents reduce the frequency of diarrheal stools but should not be used when fever or bloody diarrhea is present as they can increase the risk of mortality due to toxic megacolon or HUS. Bismuth subsalicylate may reduce symptoms in traveler's diarrhea but should be used with caution to avoid toxic doses of salicylate.[10]

Because most *E. coli* diarrhea is self-limited, the role of antimicrobial agents is debated and remains of secondary importance to rehydration. One situation in which treatment is generally favored is traveler's diarrhea (Chapter 286) because strong clinical studies have shown a benefit of antibiotics in reducing the duration of symptoms. Unfortunately, rising antimicrobial resistance has narrowed the options for empirical therapy; currently, azithromycin, a fluoroquinolone, or rifaximin is recommended, with trimethoprim-sulfamethoxazole a somewhat less reliable alternative.[11] Antimicrobials are not recommended in EHEC infection because of the possibility of increasing the risk of HUS and a lack of evidence of efficacy.

PREVENTION

Prevention of most *E. coli* enteric infections is ultimately related to basic economic development, adequate sanitary facilities, and sufficient availability of safe water. In the interim, especially in areas where adequate water supplies and sanitary facilities are not available, measures such as exclusive breast-feeding for at least 6 to 12 months and hand hygiene reduce the likelihood of acquiring *E. coli* enteric infections. Simple, portable water filters to reduce bacterial contamination are also in development.

Travelers to developing or tropical areas should avoid drinking untreated or unboiled water or ice and eating uncooked fruits or vegetables that may have been washed with highly contaminated water. Although a number of antimicrobial agents are effective during short periods when taken prophylactically, their effectiveness is ultimately limited by rapidly emerging resistance to antimicrobial drugs as well as by the potential side effects of their indiscriminate, widespread use. For example, tetracycline resistance among ETEC is now common, and resistance to trimethoprim-sulfamethoxazole and fluoroquinolones is quickly emerging around the world. This, combined with the rapid effect of empirical antibiotics at the onset of diarrhea symptoms, has diminished enthusiasm for prophylactic antibiotics. Bismuth subsalicylate is modestly effective at preventing traveler's diarrhea, although with side effects. Another option is the killed *Vibrio cholerae*/cholera toxin B subunit vaccine (Dukoral), which provides transient, partial protection against ETEC in travelers and hence may somewhat reduce the incidence of traveler's diarrhea, although strong evidence is lacking.

Sporadic EHEC infections may be reduced by adequately cooking beef, especially hamburger, and by careful handwashing and other hygienic measures in daycare centers and nursing homes. Unfortunately, large outbreaks due to contaminated produce continue to occur and are frequently associated with items meant to be eaten raw (such as sprouts). A vaccine composed from type III secreted effectors of EHEC to reduce colonization of livestock (Econiche) is on the market but has not yet entered widespread use.

PROGNOSIS

The greatest concern with enteric *E. coli* infections in developed areas is EHEC/STEC-associated HUS, which develops in 3 to 7% of sporadic cases up to as high as 20% in outbreaks. Whereas most patients with HUS recover, they frequently require intensive medical care (including temporary hemodialysis), and as many as 30% of survivors have renal sequelae such as proteinuria, hypertension, reduced glomerular filtration rate, or, more rarely, dialysis dependence.[12]

On a global scale, the greatest impact of diarrheogenic *E. coli* is in children in poor, developing areas, who are prone to death from dehydration from these otherwise self-limited infections. Moreover, children who survive repeated episodes of infectious diarrhea (due to *E. coli* and other pathogens) may experience permanent deficits in growth and even cognitive development, the full impact of which remains unknown.[13]

Grade A References

A1. Gregorio GV, Dans LF, Silvestre MA. Early versus delayed refeeding for children with acute diarrhoea. *Cochrane Database Syst Rev*. 2011;7:CD007296.

A2. Lazzerini M, Ronfani L. Oral zinc for treating diarrhoea in children. *Cochrane Database Syst Rev*. 2013;1:CD005436.

A3. Francavilla R, Lionetti E, Castellaneta S, et al. Randomised clinical trial: *Lactobacillus reuteri* DSM 17938 vs. placebo in children with acute diarrhoea—a double-blind study. *Aliment Pharmacol Ther*. 2012;36:363-369.

A4. Taylor DN, Bourgeois AL, Ericsson CD, et al. A randomized, double-blind, multicenter study of rifaximin compared with placebo and with ciprofloxacin in the treatment of travelers' diarrhea. *Am J Trop Med Hyg*. 2006;74:1060-1066.

A5. Tribble DR, Sanders JW, Pang LW, et al. Traveler's diarrhea in Thailand: randomized, double-blind trial comparing single-dose and 3-day azithromycin-based regimens with a 3-day levofloxacin regimen. *Clin Infect Dis*. 2007;44:338-346.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Zaidi D, Wine E. An update on travelers' diarrhea. *Curr Opin Gastroenterol*. 2015;31:7-13.
2. Shabana II, Zaraket H, Suzuki H. Molecular studies on diarrhea-associated *Escherichia coli* isolated from humans and animals in Egypt. *Vet Microbiol*. 2013;167:532-539.
3. Santona S, Diaz N, Fiori PL, et al. Genotypic and phenotypic features of enteropathogenic *Escherichia coli* isolated in industrialized and developing countries. *J Infect Dev Ctries*. 2013;7:214-219.
4. Jandhyala DM, Vanguri V, Boll EJ, et al. Shiga toxin–producing *Escherichia coli* O104:H4: an emerging pathogen with enhanced virulence. *Infect Dis Clin North Am*. 2013;27:631-649.
5. Law RJ, Gur-Arie L, Rosenshine I, et al. In vitro and in vivo model systems for studying enteropathogenic *Escherichia coli* infections. *Cold Spring Harb Perspect Med*. 2013;3:a009977.
6. Zoufaly A, Cramer JP, Vettorazzi E, et al. Risk factors for development of hemolytic uremic syndrome in a cohort of adult patients with STEC 0104:H4 infection. *PLoS ONE*. 2013;8:e59209.
7. Youmans BP, Ajami NJ, Jiang ZD, et al. Development and accuracy of quantitative real-time polymerase chain reaction assays for detection and quantification of enterotoxigenic Escherichia coli (ETEC) heat labile and heat stable toxin genes in travelers' diarrhea samples. *Am J Trop Med Hyg*. 2014;90:124-132.
8. Melli LJ, Ciocchini AE, Caillava AJ, et al. Serogroup-specific bacterial engineered glycoproteins as novel antigenic targets for diagnosis of Shiga toxin-producing Escherichia coli-associated hemolytic uremic syndrome. *J Clin Microbiol*. 2015;53:528-538.
9. Zipursky A, Wazny K, Black R, et al. Global action plan for childhood diarrhoea: developing research priorities: report from a Workshop of the Programme for Global Paediatric Research. *J Glob Health*. 2013;3:010406.
10. Lalani T, Maguire JD, Grant EM, et al. Epidemiology and self-treatment of travelers' diarrhea in a large, prospective cohort of Department of Defense beneficiaries. *J Travel Med*. 2014;[Epub ahead of print].
11. Kollaritsch H, Paulke-Korinek M, Wiedermann U. Traveler's diarrhea. *Infect Dis Clin North Am*. 2012;26:691-706.
12. Spinale JM, Ruebner RL, Copelovitch L, et al. Long-term outcomes of Shiga toxin hemolytic uremic syndrome. *Pediatr Nephrol*. 2013;28:2097-2105.
13. Guerrant RL, DeBoer MD, Moore SR, et al. The impoverished gut—a triple burden of diarrhoea, stunting and chronic disease. *Nat Rev Gastroenterol Hepatol*. 2013;10:220-229.

REVIEW QUESTIONS

1. A previously healthy 22-year-old student returned home to Connecticut 2 days ago from a 2-week trip to southern India. She drank only bottled water but did eat food from street vendors. On her last day, she developed nausea followed by loose stools, which have persisted. On examination, she is afebrile with normal orthostatic vital signs and appears tired but otherwise well. Her abdomen is soft and nontender with increased bowel sounds. She has had four watery stools so far today. Which of the following is appropriate for her?

A. Oral rehydration solution and azithromycin 500 mg PO daily for 3 days for presumed ETEC or EAEC traveler's diarrhea
B. Stool ova and parasite examination and empirical metronidazole 500 mg three times a day for 7 days
C. Stool culture; alert clinical laboratory to look for ETEC and EAEC
D. Intravenous saline
E. Blood cultures and stool cultures for typhoid fever

Answer: A Traveler's diarrhea is the most common treatable illness encountered by North American visitors to developing areas. The majority of cases are caused by diarrheogenic *E. coli*, particularly ETEC and EAEC. Empirical treatment with azithromycin is appropriate, and oral rehydration should be administered. This acute presentation is unlikely to be due to a protozoan, and examination by ova and parasite and empirical antiprotozoal therapy are inappropriate. A stool culture would be appropriate if she had a fever or bloody diarrhea to look for *Campylobacter*, *Shigella*, or *Salmonella*, but clinical laboratories cannot identify ETEC or EAEC. Intravenous rehydration would be appropriate only in the setting of significant volume depletion and inability to consume oral rehydration fluid. Whereas India is endemic for typhoid fever, this illness rarely is manifested with acute diarrhea in the absence of fever.

2. How is infection with diarrheogenic *E. coli* most commonly acquired?

A. Ingestion of food or water that was contaminated with virulent *E. coli* and maintained at a temperature favorable for replication of organisms to an infectious dose
B. Exposure to fecal material from an infected person (e.g., changing diapers)
C. Transfer of genetic material encoding toxins or adherence factors from pathogenic species to commensal *E. coli* within the gut
D. Ingestion of food containing preformed heat-stable or heat-labile toxins from *E. coli*
E. Spread of virulent *E. coli* from other body sites (urinary tract or skin) through hematogenous or direct extension of infectious foci

Answer: A With the exception of EHEC, diarrheogenic *E. coli* have a high infectious dose, meaning that they must be present in large concentrations to establish disease. This is usually achievable only through environmental replication within contaminated food or water, in contrast to EHEC, which can be spread person to person. Whereas virulence genes are readily transferred between bacteria, diarrheogenic strains are distinct from commensal *E. coli* and generally exist in a relatively limited number of related phylogenetic groups. Unlike *Staphylococcus aureus* and *Clostridium perfringens* enterotoxins, the ETEC enterotoxins are produced by live organisms after they reach and adhere to the small bowel epithelium. Most diarrheogenic *E. coli* are distinct from the strains adapted to cause invasive infections, such as bacteremia, meningitis, or pyelonephritis (although many DAEC and some EAEC are phylogenetically related to uropathogenic *E. coli*).

3. An *E. coli* clone is isolated in high concentrations in stool from several patients with nonbloody diarrhea. The organism is able to adhere to HeLa cells in tissue culture but does not produce actin pedestals. Genetic analysis of the organism reveals it to be negative for AA, Afa/Dr, ST, LT, and LEE genes but positive for the SLT-1 (Stx1) and SLT-2 (Stx2) genes. To which pathotype does this organism likely belong?

A. Enterotoxigenic *E. coli* (ETEC)
B. Enteroaggregative *E. coli* (EAEC)
C. Enterohemorrhagic *E. coli* (EHEC)
D. Shigatoxigenic *E. coli* (STEC)
E. Enteropathogenic *E. coli* (EPEC)

Answer: D The ability to adhere to HeLa cells suggests the possibility of pathogenicity, but the lack of actin pedestals and LEE genes rules out EPEC and EHEC. The organism does not express the AA plasmid of EAEC, the Afa/Dr adhesin of DAEC, or the ETEC toxins. However, it does express the Shiga-like toxins, making it a non-EHEC STEC strain. These strains can cause bloody or nonbloody diarrhea or can be nonpathogenic.

4. A 6-year-old previously healthy boy is admitted with 1 day of watery and then bloody diarrhea and abdominal cramps. His temperature is 37.3° C, and he has mild tachycardia. Results of laboratory tests on admission show an elevated white blood cell count of 14.4 with 85% neutrophils. Stool culture results are pending. What is the appropriate empirical antimicrobial while awaiting stool culture results?

A. Ciprofloxacin PO
B. Trimethoprim-sulfamethoxazole PO
C. Ceftriaxone IV
D. Azithromycin PO
E. No antibiotics should be given

Answer: E This is a typical presentation of hemorrhagic colitis due to EHEC or STEC. Antibiotics are not recommended in this disease because there is inadequate proof that they are beneficial, and in fact a number of studies have shown an association with increased risk of hemolytic-uremic syndrome. Fluoroquinolones and trimethoprim-sulfamethoxazole, in particular, have been shown to increase SLT production. Whereas azithromycin was shown to reduce shedding of O104:H4 ST/EAEC, its use was not shown to reduce hemolytic-uremic syndrome risk. Although this child could have *Shigella* or *Campylobacter* infection, the lack of fever makes these less likely, and a 1-day delay in antibiotic administration until culture results return would not be harmful, given that antibiotics have a relatively small effect on those self-limited infections.

5. Which of the following would have the greatest global impact on reducing *E. coli* diarrhea infections?

A. Universal vaccination with the killed cholera/*E. coli* LT-B vaccine
B. Universal vaccination of commercial livestock with the EHEC veterinary vaccine
C. Universal antibiotic prophylaxis of travelers to developing areas
D. Universal ban on antibiotic use in animal feed
E. Universal provision of treated potable water to developing areas

Answer: E The greatest burden of all diarrheal illness, and of *E. coli* in particular, is in children in developing areas who lack regular access to treated water. The ETEC vaccine and antibiotic prophylaxis, although somewhat effective in reducing the risk of traveler's diarrhea, offer only short-term protection. The EHEC veterinary vaccine might reduce human cases of EHEC associated with contamination of food and water, but EHEC, fortunately, remains largely a rare concern and occurs mostly in developed countries. Because most *E. coli* enteric infections do not require antibiotic treatment, the emergence of antibiotic resistance due to overuse of antibiotics in people and commercial livestock is less of a concern for diarrheogenic *E. coli* than for invasive *E. coli* and other important infections.

305 INFECTIONS DUE TO OTHER MEMBERS OF THE ENTEROBACTERIACEAE, INCLUDING MANAGEMENT OF MULTIDRUG-RESISTANT STRAINS

DAVID L. PATERSON

DEFINITION

The Enterobacteriaceae are a family of gram-negative bacilli that are responsible for a broad range of infections in humans and in animals. They may be motile or nonmotile, depending on the species. They are aerobic or facultatively anaerobic in growth and have a predilection for inhabiting the gastrointestinal tract. Only extra-gastrointestinal manifestations of disease are discussed in this chapter. Enteric infections caused by *Escherichia coli* are discussed in Chapter 304.

From a microbiologic perspective, the members of the Enterobacteriaceae ferment sugars. They grow on a variety of solid media and are usually readily identified by clinical microbiology laboratories. A number of subtleties arise in detection of antibiotic resistance mechanisms by laboratories, however, sometimes leading to issues in reporting of antibiotic susceptibilities. A worldwide trend is the development of multidrug resistance in all gram-negative bacilli, including the Enterobacteriaceae. Particular emphasis is placed in this chapter on treatment and prevention of these multidrug-resistant strains.

Medically important members of the Enterobacteriaceae are listed in Table 305-1. Infections due to *Salmonella*, *Shigella*, and *Yersinia* are discussed in Chapters 308, 309, and 312, respectively.

EPIDEMIOLOGY

The Enterobacteriaceae are among the most common pathogens to infect humans worldwide. They are responsible for community-acquired, hospital-acquired, and health care–associated infections. Examples of the last category include infections acquired in nursing homes and those associated with outpatient management of cancers or hematologic malignant disease. As resident components of the flora of the gastrointestinal tract, isolates of Enterobacteriaceae may represent examples of colonization rather than true infection. This may apply to isolates from rectal swabs, urine, or respiratory secretions. In many hospitals, multidrug-resistant Enterobacteriaceae have become endemic, leading to substantial problems in the management of serious infections.[1]

E. coli is the most common cause of urinary tract infections (UTIs), accounting for more than 80% of isolates from urine in most clinical situations. Any of the remaining members of the Enterobacteriaceae can cause this infection; *Klebsiella* spp and *Proteus mirabilis* are among other common causes of UTI. *Providencia stuartii* is notable as a cause of UTI among chronically catheterized patients. The Enterobacteriaceae can cause uncomplicated UTI in healthy women as well as acute pyelonephritis and UTI complicating renal tract abnormalities or catheterization. The Enterobacteriaceae may colonize the urine or be found in contaminated, improperly collected samples.

Antibiotic-resistant, uropathogenic *E. coli* may spread clonally. In other words, *E. coli* isolates from multiple different people with UTI may be identical or closely related at a genetic level. Trimethoprim-sulfamethoxazole–resistant *E. coli* is notable for its spread in a clonal fashion in the United States (e.g., "clonal group A" and *E. coli* O15:K52:H1). A widely spread *E. coli* clone, defined by multilocus sequence typing as sequence type 131 (ST131), has been found to be associated with ciprofloxacin resistance and production of extended-spectrum β-lactamases (ESBLs).[2,3] This clone is typically associated with community-acquired UTI. It has been detected in every inhabited continent.

Given the niche of most Enterobacteriaceae in the gastrointestinal tract, it is not surprising that these bacteria are prominent as causes of peritonitis. *E. coli* ranks most common as a cause of both spontaneous bacterial peritonitis (occurring in cirrhotic patients) and bacterial peritonitis arising from visceral perforation. Pyogenic liver abscess and intra-abdominal abscess may also be due to *E. coli*. Other members of the Enterobacteriaceae may also cause these intra-abdominal infections, especially in patients with "tertiary" peritonitis occurring after prior surgery for intra-abdominal disease.

The Enterobacteriaceae may also be the causative pathogens of pneumonia. They are more frequently the cause of hospital- and health care–associated pneumonia than community-acquired pneumonia. *Klebsiella pneumoniae* was once renowned as a cause of community-acquired pneumonia in alcoholics but has declined in significance during the last few decades. One exception is the finding of community-acquired pneumonia, liver abscess, or meningitis in Asia due to highly mucoid strains of *K. pneumoniae*.[4] Hospital-acquired pneumonia due to the Enterobacteriaceae may be ventilator associated. The Enterobacteriaceae rank most common as causes of ventilator-associated pneumonia after *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and *Acinetobacter baumannii*. The Enterobacteriaceae may also cause hospital-acquired pneumonia in non–mechanically ventilated patients, such as those with neurologic impairment from head injury or cerebrovascular accident.

Outbreaks of antibiotic-resistant *K. pneumoniae* infection in hospitals have been prominent for more than three decades. In the hospital setting, *K. pneumoniae* is usually the cause of peritonitis, pneumonia, or complicated UTI. Blood stream infection arising from these sites of infection, from vascular catheters, or in association with neutropenia may also occur. In the 1970s, outbreaks due to gentamicin-resistant strains occurred. In the 1980s and 1990s, hospital outbreaks of ESBL-producing *K. pneumoniae* became commonplace. Finally, in the last decade, *K. pneumoniae* carbapenemase (KPC)–producing *K. pneumoniae* became a substantial infection control issue. The KPC-producing organisms are discussed in detail in a subsequent section of this chapter.

TABLE 305-1 SELECTED MEDICALLY IMPORTANT MEMBERS OF THE FAMILY ENTEROBACTERIACEAE

GENUS	SOME IMPORTANT SPECIES
Enterobacter	*E. cloacae*
Escherichia	*E. coli*
Klebsiella	*K. pneumoniae, K. oxytoca*
Morganella	*M. morganii*
Plesiomonas	*P. shigelloides*
Proteus	*P. mirabilis, P. vulgaris*
Providencia	*P. stuartii*
Salmonella	*S. enterica*
Serratia	*S. marcescens*
Shigella	*S. sonnei*
Yersinia	*Y. pestis, Y. enterocolitica*

PATHOBIOLOGY

The virulence factors associated with *E. coli* causing enteric infections are discussed in detail in Chapter 304. At least 40 different virulence genes have been described in *E. coli* causing extraintestinal infections. Among the virulence properties of these strains is the renowned ability of *E. coli* to adhere to uroepithelial cells. The ST131 *E. coli* clone is typically highly virulent. It belongs to "phylogenetic group" B2, which is known for extraintestinal pathogenic infections. In an evaluation of the ST131 clone, numerous extraintestinal virulence genes were found. Enterobacteriaceae other than *E. coli* also possess a number of virulence mechanisms, but it is beyond the scope of this chapter to describe these in detail.

Many Enterobacteriaceae are now multidrug resistant. Resistance of the Enterobacteriaceae to aminoglycosides is usually mediated by production of aminoglycoside-modifying enzymes. However, 16S ribosomal RNA methylation is a more recently described mechanism that results in resistance to gentamicin, tobramycin, and amikacin. Resistance of the Enterobacteriaceae to fluoroquinolones is a growing problem. In most areas, more than 20% of *E. coli* strains are resistant to fluoroquinolones. Resistance is usually mediated by mutations at target sites and is associated with the ST131 strain. Mechanisms such as efflux pump overexpression and expression of the plasmid-mediated *qnr* genes may also play a role in quinolone resistance. The mechanisms of resistance of the Enterobacteriaceae to β-lactam antibiotics are worthy of detailed description (Table 305-2).

TABLE 305-2 ANTIBIOTIC RESISTANCE ISSUES IN THE ENTEROBACTERIACEAE

ORGANISM	β-LACTAMASE	COMMENTS
NARROW-SPECTRUM β-LACTAMASES		
Escherichia coli	TEM-1	Present in at least 40% of strains globally
Klebsiella pneumoniae	SHV-1	Universally found
EXTENDED-SPECTRUM β-LACTAMASES		
E. coli	CTX-M-15	International clone ST131
K. pneumoniae	SHV, TEM, CTX-M	Hospital outbreaks
AmpC β-LACTAMASES		
Enterobacter cloacae	AmpC	Universally found
E. coli	AmpC	Occasional plasmid-mediated strains
CARBAPENEMASES		
K. pneumoniae	KPC	Endemic in many parts of the United States
K. pneumoniae	NDM	Hospital outbreaks in India

Narrow-Spectrum β-Lactamases

Ampicillin was introduced into clinical practice in the early 1960s. Enterobacteriaceae, with genes encoding β-lactamases inherent to their chromosome, were found to be intrinsically resistant to this antibiotic. Examples of these include the production of the SHV-1 β-lactamase by *K. pneumoniae* and the AmpC β-lactamase by *Enterobacter cloacae*. Other Enterobacteriaceae (e.g., *E. coli*, *Proteus mirabilis*, or *Salmonella*) lacked substantial production of chromosomally encoded β-lactamase. However, within months of the release of ampicillin, a plasmid-mediated β-lactamase leading to resistance of *E. coli* to the antibiotic was discovered. This β-lactamase was coined the TEM β-lactamase, in honor of the patient (Temoneira) from whom the TEM β-lactamase–producing *E. coli* was first isolated. Plasmids encoding resistance to ampicillin have now become widespread, with at least 40% of *E. coli* in most parts of the world now being TEM β-lactamase producers.

Extended-Spectrum β-Lactamases

Third-generation cephalosporins (such as ceftriaxone) were active against Enterobacteriaceae producing narrow-spectrum β-lactamases. However, mutant genes encoding β-lactamases capable of inactivating third-generation cephalosporins were discovered in the 1980s. The genes encoding these β-lactamases were identical to TEM or SHV, except for point mutations that led to an altered amino acid sequence. The subsequent structural change led to an ability to hydrolyze and thus to inactivate third-generation cephalosporins. In view of the extended spectrum of antibiotic-hydrolyzing abilities compared with the parent TEM and SHV enzymes, these β-lactamases were coined extended-spectrum β-lactamases (ESBLs). In addition to the TEM and SHV types of ESBLs, many new types of ESBLs have now been described, most notably the CTX-M type.

Frequently, the ST131 *E. coli* clone produces a CTX-M type of ESBL (especially CTX-M-15). Although ESBLs are typically susceptible to β-lactamase inhibitors (e.g., clavulanic acid), many ST131 *E. coli* isolates produce an additional β-lactamase (OXA-1) that confers resistance to β-lactamase inhibitors. The antibiotic resistance phenotype of the ST131-positive *E. coli* is that the organisms are typically resistant to ceftriaxone, cefotaxime, fluoroquinolones, trimethoprim, trimethoprim-sulfamethoxazole, penicillin–β-lactamase inhibitor combinations, and tetracyclines. Resistance to aminoglycosides is variable. The clone confers multidrug resistance by the presence of multiple antibiotic resistance genes. These are usually encoded on plasmids.

AmpC β-Lactamases

A number of genera within the Enterobacteriaceae have a chromosomally encoded β-lactamase capable of producing resistance to all penicillins and all cephalosporins except cefepime. In addition, these β-lactamases are not inhibited by β-lactamase inhibitors, such as clavulanic acid or tazobactam. These β-lactamases are known as AmpC β-lactamases. They are not derived from narrower-spectrum parent β-lactamases, so it is not correct to call these extended-spectrum β-lactamases. These AmpC β-lactamases may be overproduced in the presence of certain antibiotics (i.e., the genes encoding the AmpC β-lactamases are "inducible"). *Enterobacter* spp, *Citrobacter freundii*, *Serratia marcescens*, and *Morganella morganii* have inducible, chromosomally encoded AmpC β-lactamases. The genes encoding these β-lactamases have now been found on plasmids in *E. coli*, *Salmonella*, and other gram-negative bacteria.

KPC, NDM, and Other Carbapenemases

In 2001, a novel carbapenem-hydrolyzing β-lactamase from a carbapenem-resistant strain of *K. pneumoniae* was first described. This enzyme was described as the *K. pneumoniae* carbapenemase (KPC). KPC-producing organisms are typically resistant to penicillins, cephalosporins, and carbapenems and are not inhibited by clavulanic acid or other commonly used β-lactamase inhibitors, such as sulbactam and tazobactam. KPC production has been documented in *E. coli* and in many genera of the Enterobacteriaceae, such as *Enterobacter*, *Citrobacter*, *Proteus*, and *Salmonella*.

The epicenter of KPC-producing *K. pneumoniae* has been New York City. By 2004, approximately one quarter of *K. pneumoniae* isolates in a surveillance study in Brooklyn, New York, were KPC producing. The Centers for Disease Control and Prevention now reports KPC-producing organisms in numerous cities across the United States. International spread of KPC-producing organisms has been well described, with outbreaks in Israel, Italy, and Greece being particularly problematic. Parts of China and South America have also reported endemic KPC-producing organisms.[5]

Another carbapenem-hydrolyzing β-lactamase, the New Delhi metallo-β-lactamase (NDM), has emerged in the last decade as a major cause of carbapenem resistance. As its name suggests, its epicenter is the Indian subcontinent. However, it has also been found to be endemic in some hospitals in the Balkan states as well as causing outbreaks in North America and Europe. Many outbreaks of NDM-producing organisms have been associated with interhospital transfers from hospitals in the Indian subcontinent.[6] Worryingly, NDM-producing organisms have been found in drinking water in India[7] and in food-producing animals in China.[8]

Other carbapenem-hydrolyzing β-lactamases include the OXA-48 type, which is found in North Africa, the Middle East, and India and has since spread to other parts of the world, as well as a variety of metallo-β-lactamases (such as of the IMP and VIM types). These enzymes have in common the ability to render Enterobacteriaceae resistant to carbapenems and other β-lactam antibiotics.

CLINICAL MANIFESTATIONS

The clinical manifestations of UTI, peritonitis, pneumonia, and blood stream infection due to the Enterobacteriaceae are described in other chapters.

DIAGNOSIS

Infection with the Enterobacteriaceae is readily diagnosed in clinical microbiology laboratories after collection of appropriate specimens. Examination of the Gram stain enables a rapid differentiation of gram-negative bacilli from other pathogens. However, it is frequently difficult on clinical grounds and by Gram stain results to differentiate infection with the Enterobacteriaceae from other gram-negative bacilli, such as *Pseudomonas aeruginosa*.

Once a member of the Enterobacteriaceae has been identified, it is important to ensure that appropriate antibiotic susceptibility testing has been performed. In some instances, specialized tests need to be performed by the clinical microbiology laboratory to detect ESBL or carbapenemase production. Molecular epidemiologic assessment may need to be undertaken to determine if an isolate belongs to an outbreak strain.

TREATMENT

Rx

Treatment depends on the site of infection and the extent of antibiotic resistance. Empirical antibiotic choices for orally administered therapy for UTI due to *E. coli* and the other Enterobacteriaceae may include fluoroquinolones, trimethoprim-sulfamethoxazole, amoxicillin-clavulanate, and nitrofurantoin. *P. mirabilis* is resistant to nitrofurantoin. Empirical parenteral choices may include third-generation cephalosporins, penicillin–β-lactamase inhibitor combinations, aminoglycosides, fluoroquinolones, and carbapenems.

The advent of antibiotic resistance in the Enterobacteriaceae has the potential to have a huge impact on treatment of common infections. UTI is a pertinent example. Orally administered choices such as fluoroquinolones, trimethoprim-sulfamethoxazole, and amoxicillin-clavulanate are typically inactive in the ST131 *E. coli* strain. There may be a need to now admit some patients for parenteral antibiotics because of this antibiotic resistance. Worse still, some patients with hospital-acquired infection may need "last-line" antibiotics like colistin or polymyxin B. Specific details of treatment choices for antibiotic-resistant Enterobacteriaceae are given in Table 305-3.

TABLE 305-3 TREATMENT OPTIONS FOR MULTIPLY RESISTANT ENTEROBACTERIACEAE

ESBL PRODUCERS
Carbapenems (first choice, serious infections)
Piperacillin-tazobactam, cefepime (second choice, serious infections)
Nitrofurantoin, fosfomycin (first choice, UTI)
Ciprofloxacin, amoxicillin-clavulanate (second choice, UTI)
KPC OR NDM PRODUCERS
Colistin, polymyxin B, tigecycline, meropenem (as part of a combination)

ESBL = extended-spectrum β-lactamases; KPC = *K. pneumoniae* carbapenemase; NDM = New Delhi metallo-β-lactamase; UTI = urinary tract infection.

Treatment of ESBL-Producing Organisms

In vitro, the carbapenems (including imipenem, meropenem, doripenem, and ertapenem) have the most potent activity against ESBL-producing organisms. This is not surprising because these antibiotics are not inactivated by ESBLs. Carbapenems should be regarded as the drugs of choice for serious infections with ESBL-producing organisms on the basis of extensive positive clinical experience. No randomized trials have been completed comparing carbapenems with other antibiotic classes against ESBL producers. Although a substantial proportion of ESBL-producing organisms will be resistant to piperacillin-tazobactam, observational studies have shown a similar mortality from carbapenem-treated blood stream infections due to ESBL-producing compared with β-lactam/β-lactamase inhibitor–treated infections.[9] There is no evidence that combination therapy involving a carbapenem is superior to use of a carbapenem alone for ESBL producers.

Thus, meropenem or piperacillin-tazobactam is the treatment of choice for serious infections due to ESBL producers. The ability to use ertapenem once daily makes it potentially useful in serious infections with ESBL producers in nursing home residents or patients continuing parenteral therapy out of the hospital. UTIs may be treated with orally administered fosfomycin, nitrofurantoin, or amoxicillin-clavulanate if susceptible.

Treatment of Carbapenem-Resistant Organisms

Treatment of carbapenem-resistant organisms (e.g., due to KPC or NDM production) is difficult because they may lack susceptibility to all β-lactam antibiotics (including penicillins, cephalosporins, aztreonam, and carbapenems), fluoroquinolones, and aminoglycosides. Carbapenemase-producing organisms may sometimes appear susceptible to carbapenems such as meropenem and imipenem (although they are almost always recognized as resistant to ertapenem), but these antibiotics should not be relied on as monotherapy.

No randomized controlled trials have yet been performed evaluating different antibiotic options for carbapenemase producers. On the basis of observational studies, combination therapy appears to be superior to single-drug therapy. Combinations of a polymyxin, tigecycline, and meropenem have met with the greatest success.[10-12] Meropenem has been used in these combinations despite a lack of in vitro susceptibility.

A variety of new antibiotics with activity against ESBL producers and some carbapenemase producers are undergoing evaluation. These include the combination of ceftazidime and a novel β-lactamase inhibitor called avibactam and a new aminoglycoside, plazomicin. These appear to have significant activity against KPC producers. Unfortunately, avibactam does not inhibit NDM or other metallo-β-lactamases. Furthermore, plazomicin is not active against NDM producers because most exhibit aminoglycoside resistance by way of 16S ribosomal RNA methylation.

PREVENTION

Prevention of hospital-acquired outbreaks of ESBL, KPC, or NDM producers rests on a number of basic infection control principles. First, if a focus of infection exists in the hospital environment, it should be removed. Examples have included contamination of ultrasonography coupling gel and bronchoscopes. Outbreaks have been dramatically curtailed when these sources of contamination have been properly cleaned or removed from the hospital environment.

Present evidence suggests that transient carriage on the hands of health care workers is the most important means of transfer of ESBL-, KPC-, or NDM-producing Enterobacteriaceae from patient to patient. The hands of health care workers are presumably colonized by contact with the skin of patients with skin colonization of the organism or by contact with a contaminated environment around the patient. Many patients may have asymptomatic colonization with ESBL-, KPC-, or NDM-producing organisms without signs of overt infection. These patients represent an important reservoir of organisms. In some hospital wards with ongoing issues with ESBL, KPC, or NDM producers, more than 30% of patients have gastrointestinal tract colonization with these organisms at any one time. These patients should be nursed with use of contact precautions. Hand carriage by health care workers is usually eliminated by hand hygiene with alcohol-based agents. Compliance with contact isolation precautions and hand hygiene needs to be high to maximize the effectiveness of these interventions.

Changes in antibiotic policy may play a role in controlling outbreaks of ESBL, KPC, or NDM producers, but this concept remains controversial. In one reported outbreak of ESBL producers, no effort was made to change infection control procedures. Instead, at this hospital, ceftazidime use decreased and piperacillin-tazobactam was introduced into the formulary. This coincided with curtailment of the outbreak. In another institution, cephalosporins as an entire class were removed to exact control over endemic ESBL producers. The difficulty with this approach is that replacement of one antibiotic class with another may result in replacement of one antibiotic resistance issue with another. No study has demonstrated that removal of carbapenems from a hospital formulary leads to elimination of KPC or NDM producers. Because these organisms are resistant to multiple antibiotic classes, classes as diverse as β-lactam/β-lactamase inhibitor combinations and fluoroquinolones have been more commonly received than carbapenems before colonization or infection with KPC producers. Prudent use of all antibiotic classes with an emphasis on reducing duration of antibiotic use may be more useful than individual antibiotic class restriction.

PROGNOSIS

The prognosis of infection with the Enterobacteriaceae depends on multiple factors, such as site of infection, presence of underlying diseases, and adequacy of empirical antibiotic therapy. At one extreme, inadequate orally administered antibiotic therapy for uncomplicated UTI due to an ESBL producer may have no impact on mortality, although it may have an impact on duration of symptoms and need for parenteral therapy for treatment failure. At the other extreme, patients in an intensive care unit with serious infections due to KPC or NDM producers may have an in-hospital mortality rate exceeding 70%. This may compare with in-hospital mortality rates of 20 to 30% in comparable patients without infection due to a KPC or NDM producer.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Centers for Disease Control and Prevention. *New carbapenem-resistant Enterobacteriaceae warrant additional action by healthcare providers*. Available at: http://www.bt.cdc.gov/HAN/han00341.asp; Accessed March 7, 2015.
2. Rogers BA, Sidjabat HE, Paterson DL. *Escherichia coli* O25b-ST131: a pandemic, multiresistant, community-associated strain. *J Antimicrob Chemother*. 2011;66:1-14.
3. Petty NK, Ben Zakour NL, Stanton-Cook M, et al. Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci U S A*. 2014;111:5694-5699.
4. Cheng NC, Yu YC, Tai HC, et al. Recent trend of necrotizing fasciitis in Taiwan: focus on monomicrobial *Klebsiella pneumoniae* necrotizing fasciitis. *Clin Infect Dis*. 2012;55:930-939.
5. Munoz-Price LS, Poirel L, Bonomo RA, et al. Clinical epidemiology of the global expansion of *Klebsiella pneumoniae* carbapenemases. *Lancet Infect Dis*. 2013;13:785-796.
6. Wailan AM, Paterson DL. The spread and acquisition of NDM-1: a multifactorial problem. *Expert Rev Anti Infect Ther*. 2014;12:91-115.
7. Walsh TR, Weeks J, Livermore DM, et al. Dissemination of NDM-1 positive bacteria in the New Delhi environment and its implications for human health: an environmental point prevalence study. *Lancet Infect Dis*. 2011;11:355-362.
8. Wang Y, Wu C, Zhang Q, et al. Identification of New Delhi metallo-β-lactamase 1 in *Acinetobacter lwoffii* of food animal origin. *PLoS ONE*. 2012;7:e37152.
9. Rodriguez-Bano J, Navarro MD, Retamar P, et al. β-Lactam/β-lactam inhibitor combinations for the treatment of bacteremia due to extended-spectrum β-lactamase–producing *Escherichia coli*: a post hoc analysis of prospective cohorts. *Clin Infect Dis*. 2012;54:167-174.
10. Qureshi ZA, Paterson DL, Potoski BA, et al. Treatment outcome of bacteremia due to KPC-producing *Klebsiella pneumoniae*: superiority of combination antimicrobial regimens. *Antimicrob Agents Chemother*. 2012;56:2108-2113.
11. Tumbarello M, Viale P, Viscoli C, et al. Predictors of mortality in bloodstream infections caused by *Klebsiella pneumoniae* carbapenemase-producing *K. pneumoniae*: importance of combination therapy. *Clin Infect Dis*. 2012;55:943-950.
12. Daikos GL, Tsaousi S, Tzouvelekis LS, et al. Carbapenemase-producing *Klebsiella pneumoniae* bloodstream infections: lowering mortality by antibiotic combination schemes and the role of carbapenems. *Antimicrob Agents Chemother*. 2014;58:2322-2328.

REVIEW QUESTIONS

1. A 78-year-old man was admitted to a surgical intensive care unit in New York City for management of severe, necrotizing pancreatitis. He was empirically treated with meropenem. On the 18th day of surgical intensive care unit admission, while still receiving meropenem, he developed a fever (temperature of 39.5° C) and hypotension requiring inotropic support. Blood culture samples were collected, and nonmotile gram-negative bacilli were seen after incubation for 12 hours. Which of the following statements about management of the bacteremia is true?

A. He should receive ampicillin for a likely *Enterobacter cloacae* blood stream infection.
B. He should receive ampicillin-sulbactam for a likely *Escherichia coli* blood stream infection.
C. He should be maintained on meropenem for a likely ESBL-producing *Klebsiella pneumoniae* blood stream infection.
D. He should receive a polymyxin in combination with other antibiotics for blood stream infection because of a probable carbapenem-resistant organism.
E. His antibiotics should be discontinued and the blood stream infection should be managed by changing all intravascular lines.

Answer: D Carbapenem-resistant gram-negative organisms are now endemic in many parts of the world. Carbapenem resistance, mediated by the KPC carbapenemase, is widespread in many parts of the United States, Greece, Italy, and Israel. This creates significant management issues for severe sepsis due to gram-negative bacilli. At present, a polymyxin in combination with meropenem or tigecycline is the treatment of choice for KPC producers. Source control should also be used, when appropriate.

2. A 27-year-old woman has a history of recurrent urinary tract infections. She presents with symptoms of dysuria and urinary frequency. *Escherichia coli* is isolated from her urine and is found to be resistant to ampicillin, trimethoprim-sulfamethoxazole, and ciprofloxacin. The organism is susceptible to piperacillin-tazobactam and ertapenem. Which of the following scenarios is most likely?

A. TEM-1 accounts for the resistance to ciprofloxacin.
B. NDM-1 accounts for the resistance to trimethoprim-sulfamethoxazole.
C. She is most likely infected with the ST171 strain of *E. coli*.
D. She is most likely infected with the ST131 strain of *E. coli*.
E. She most likely works in a New York City nursing home and has infection with KPC-producing *E. coli*.

Answer: D ST131 *E. coli* is an international pandemic clone that is typically resistant to ciprofloxacin. It may also be ESBL producing due to production of the CTX-M-15 β-lactamase. Although there are reports of ST131 producing carbapenemases, this is not commonplace.

3. A 24-year-old medical student undertakes an elective in Nepal. While there, he goes mountain climbing but falls and breaks his leg. After undergoing surgery in Kathmandu, he is transferred back to the United States after 2 weeks of hospitalization in Nepal. A rectal swab is taken on admission to the hospital in the United States. Which of the following are most likely to be found?

A. NDM-producing *Klebsiella pneumoniae*
B. KPC-producing *Klebsiella pneumoniae*
C. AIM-1–producing *Klebsiella pneumoniae*
D. VIM-1–producing *Klebsiella pneumoniae*
E. IMP-4–producing *Klebsiella pneumoniae*

Answer: A NDM-producing organisms are endemic in many hospitals in the Indian subcontinent. Patients hospitalized in this area may become colonized with NDM producers. They may then serve as reservoirs for these organisms when transferred to hospitals in countries without endemic NDM producers. Outbreaks may result unless adequate infection control interventions are used.

4. A 66-year-old man has been ventilated for 3 weeks after a severe head injury. He develops radiologic changes in his chest consistent with right middle and lower lobe consolidation. His endotracheal aspirates have become purulent, and he has developed a new fever. The laboratory identifies a motile gram-negative bacillus. It is able to ferment sugars. It is oxidase negative. On susceptibility testing, it is resistant to ampicillin and ceftazidime but susceptible to ciprofloxacin, gentamicin, and meropenem. The most likely organism is

A. *Pseudomonas aeruginosa*
B. *Staphylococcus aureus*
C. *Streptococcus pneumoniae*
D. *Acinetobacter baumannii*
E. *Enterobacter cloacae*

Answer: E The Enterobacteriaceae are able to ferment sugars, whereas "nonfermenters" like *P. aeruginosa* (an oxidase-positive organism) and *A. baumannii* (an immotile organism) are not. Staphylococci and streptococci are gram-positive organisms. Only *E. cloacae* fits the description.

306

PSEUDOMONAS AND RELATED GRAM-NEGATIVE BACILLARY INFECTIONS

MATTHEW E. FALAGAS AND PETROS I. RAFAILIDIS

DEFINITION

Infections due to *Pseudomonas* spp are caused by members of the family Pseudomonadaceae. The Pseudomonadaceae is a group of gram-negative rods, including *Pseudomonas aeruginosa*, the most frequently recovered human pathogen in the family. Other *Pseudomonas* spp include *Pseudomonas putida, Pseudomonas alcaligenes, Pseudomonas fluorescens, Pseudomonas luteola, Pseudomonas mendocina, Pseudomonas oryzihabitans, Pseudomonas pseudoalcaligenes, Pseudomonas stutzeri, Pseudomonas chlororaphis, Pseudomonas delafieldii, Pseudomonas kingii, Pseudomonas pertucinogena,* and *Pseudomonas* CDC group 1.

Related gram-negative bacillary infections include infections due to *Stenotrophomonas maltophilia* (formerly known as *Pseudomonas maltophilia* and *Xanthomonas maltophilia*) and members of the genus *Burkholderia* (*Burkholderia pseudomallei*, *Burkholderia mallei*, and *Burkholderia cepacia* complex).

The Pathogens

Pseudomonas aeruginosa is a gram-negative, lactose nonfermenting, straight or slightly curved rod with a length ranging from 1.5 to 7 µm and a width of 0.5 to 1.0 µm. It is catalase positive, oxidase positive, and motile with one or more polar flagella. Most species oxidize glucose and reduce nitrate to nitrite or nitrogen gas. It has the ability to grow at 42° C. This pathogen is admirably armed on its exterior: a polysaccharide capsule along with lipopolysaccharides, pili, and flagella. Furthermore, the interior arsenal includes toxins such as exotoxin A, pyocyanin (blue or blue-green pigment), pyorubin (red or red-brown pigment), pyomelanin (black pigment), and pyoverdin (yellow-green pigment). *P. aeruginosa* is notorious for its ability to acquire resistance genes and to spread by horizontal transfer.

Stenotrophomonas maltophilia is a gram-negative nonfermenting bacillus motile by polar flagella. It is also catalase positive; the majority of strains are oxidase negative, but some strains are oxidase positive. *Burkholderia cepacia* and *Burkholderia pseudomallei* are also motile gram-negative lactose nonfermenting bacteria. In contrast, *Burkholderia mallei* is nonmotile.

EPIDEMIOLOGY

P. aeruginosa is one of the most common pathogens in health care–associated infections. Ventilator-associated pneumonia, bacteremia associated with central venous catheters or secondary to infections present elsewhere in the body, urinary tract infections, and surgical site infections are the main types of infections associated with *P. aeruginosa* in the hospital setting. Data from the National Healthcare Safety Network at the Centers for Disease Control and Prevention indicate that during 2009 to 2010, *P. aeruginosa* was responsible for 8% of a total of 69,475 health care–associated infections and was the sixth most frequent culprit among 81,139 pathogens.[1] Approximately 2% of these strains are resistant to carbapenems, a powerful antibiotic against them. This organism is able to survive in environments that have only minimal nutritional components. *P. aeruginosa* can colonize moist surfaces of the axilla, ear, and perineum. It is also isolated from other moist, inanimate environments within the hospital, including water in sinks and drains, mechanical ventilation equipment, dialysis equipment, toilets, showers, hydrotherapy pools, mops, water for flowers, and even cleaning solutions.

Strains of *P. aeruginosa* resistant to multiple classes of antibiotics, including quinolones and β-lactams, are a major cause of morbidity and mortality worldwide. During a relatively short time, a multidrug-resistant *P. aeruginosa* strain spread through neighboring countries.[2] The use of broad-spectrum antibiotics is certainly a risk factor for the development of multidrug-resistant *P. aeruginosa*, as are deficiencies in infection control implementation.

A range of hosts are particularly prone to infections by this pathogen: patients with neutropenia, patients with burns, cystic fibrosis (CF) patients, cancer patients, transplant recipients, diabetics, and patients with AIDS. Patients with compromised immunity due to treatment or by the disease, either humoral (hypogammaglobulinemia) or cellular (steroid treatment), as well as patients with foreign bodies (e.g., vascular grafts, orthopedic implants) are also more vulnerable to *P. aeruginosa* infections.

Community-acquired *P. aeruginosa* infection is related to exposure to water by the use of hot tubs, whirlpools, swimming pools, spas, and other types of baths as well as to the use of contact lenses, particularly the extended-wear variety. Puncture wounds including those through tennis shoes can give rise to *P. aeruginosa* infection. *P. aeruginosa* endophthalmitis after eye trauma can result in visual compromise, and *P. aeruginosa* endocarditis is frequently found in intravenous injection drug users. In addition, there is the possibility for drug-resistant *P. aeruginosa* in the community, and this risk has to be taken into account in prescribing antibiotic therapy.

S. maltophilia, once thought to be of limited virulence, is a significant emerging pathogen.[3] It is an environmental gram-negative, multidrug-resistant organism that mainly causes pneumonia and acute exacerbation of chronic obstructive pulmonary disease. Bacteremia, urinary tract infection, skin and soft tissue infections including cellulitis, osteomyelitis, endocarditis, and meningitis are among the infections caused by *S. maltophilia*.

Once isolated in the geographic setting of Southeast Asia and Australia with an incidence of approximately 50/100,000 in the general population, infections due to *B. pseudomallei* are of special interest because they are reported now more frequently also from the Indian subcontinent.[4] Fortunately, infection due to *B. mallei* is rare in humans.

Burkholderia cepacia belongs to the *Burkholderia cepacia* complex, which includes also *Burkholderia ambifaria, Burkholderia anthina, Burkholderia arboris, Burkholderia cenocepacia, Burkholderia contaminans, Burkholderia diffusa, Burkholderia dolosa, Burkholderia latens, Burkholderia lata, Burkholderia metallica, Burkholderia multivorans, Burkholderia pyrrocinia, Burkholderia seminalis, Burkholderia. stabilis, Burkholderia ubonensis,* and *Burkholderia vietnamiensis*. *B. cepacia* was previously reported under the name of *Pseudomonas cepacia*. *B. cepacia* infects mainly patients with CF and chronic granulomatous disease. However, there are many reports of *B. cepacia* infections in hospitalized patients without CF. Patients with central venous catheters, manipulation of the urogenital tract (catheterization, instrument insertion, and biopsy), burns, and wounds (surgical and other) may develop infection due to *B. cepacia*.

PATHOBIOLOGY

Pathogenesis and Pathophysiology

Innate immunity, primarily through inflammatory cytokine production and phagocytic clearance by neutrophils and macrophages, is the key to endogenous control of *P. aeruginosa* infection.[5] Patients with neutropenia and also patients in whom there are defects of recruitment and activation of polymorphonuclear neutrophils, such as those with CF, are susceptible to *P. aeruginosa* infections. This happens because the lipopolysaccharide of *P. aeruginosa* binds to the CF transmembrane conductance regulator (CFTR), which is a channel involved in chloride movement across the cell. When there is normal CFTR, release of interleukin-1 and signaling through the interleukin-1 receptor and adaptor molecule MyD88 occur. Furthermore, CFTR lipid rafts are formed and NF-κB is translocated, followed by transcription of NF-κB–dependent genes and the production of interleukin-6, interleukin-8, and intercellular adhesion molecule 1. After that follows the recruitment of polymorphonuclear neutrophils to infected tissue and the induction of apoptosis and resolution of infection. In the absence of MyD88, however, lethal pneumonia and sepsis can be induced by fewer than 60 bacterial cells applied to the noses of mice. Patients with CF have a defective CFTR and thus cannot mount a normal defensive response to *P. aeruginosa*. The major challenge has been to correlate the genetic defect leading to the synthesis of either no CFTR protein or dysfunctional CFTR to the pathogenesis of *P. aeruginosa* infection.

Other host factors that predispose to *Pseudomonas* infections are burns and wounds, especially necrotic tissue, and AIDS/HIV infection. In addition, loss of mucosal function of the tracheobronchial tree of mechanically ventilated patients and patients with cancer and disruption of the normal bacterial flora in the gastrointestinal tract are ideal settings for the growth of *P. aeruginosa*.

Bacterial Factors in Pathogenesis

P. aeruginosa is one of the ESKAPE (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa,* and *Enterobacter* species) group of bacteria that the Infectious Diseases Society of America has specifically addressed as a cause of concern. It survives in aquatic as well as in soil environments. Its armamentarium is vast and adaptable to the challenges faced, and its intraspecies but also interspecies communication in the microbial kingdom is noteworthy. Several factors including flagella, pili, exopolysaccharides, phospholipases, proteases, endotoxins, secreted toxins (types I, II, II, IV, and VI), exotoxins, and iron-binding proteins are involved in pathogenesis (E-Table 306-1). Defensive weapons of *Pseudomonas* are the production of high levels of an extracellular mucoid polysaccharide called alginate, which is the main constituent of the glycocalyx allowing growth in a biofilm. Furthermore, the expression of O side chains on the bacterium's lipopolysaccharide prevents lysis by complement. Aggressive mediators to kill immune cells and to invade and degrade tissues are the toxins and enzymes produced by the bacterium. Exotoxin A, an adenosine diphosphate–ribosylating toxin, has activity similar to that of diphtheria. The production of this cellular toxin is affected by iron levels. *P. aeruginosa* uses pyoverdin and pyochelin, typical siderophore systems, to acquire iron.

P. aeruginosa can induce hemolysis by PlcHR, a hemolytic phospholipase C. A nonhemolytic phospholipase C, PlcN, is also made by *P. aeruginosa* strains. *P. aeruginosa* type III secretion systems allow direct injection of bacterial toxins into eukaryotic cells and disrupt cellular trafficking by inhibiting the actin cytoskeleton and by affecting protein synthesis. For *P. aeruginosa*, clinical isolates expressing type III toxins are isolated significantly more frequently from patients who have poor clinical outcomes. Four major effector proteins are known: ExoS, ExoT, ExoU, and ExoY. More than these factors, the ability of *Pseudomonas* to communicate with other members of its microbial community through quorum sensing leads to the formation of biofilms, a daunting obstacle in attempting to eradicate the pathogen. Quorum sensing systems of *Pseudomonas* are lasR1, rhlR1, and *Pseudomonas* quinolone system. These systems have a complex interaction with other factors in

E-TABLE 306-1 VIRULENCE FACTORS FOR *PSEUDOMONAS AERUGINOSA*

LOCATION OR CLASS	EXAMPLES	ACTIVITY/EFFECTS ON HOST
Cell surface	Alginate	Antiphagocytic, resists opsonic killing
	Lipopolysaccharide	Endotoxic, antiphagocytic, avoids preformed antibody to previously encountered O antigens
	Pili (produced by type IV secretion)	Twitching motility, biofilm formation, adherence to host tissues
	Flagella	Motility, biofilm formation, adherence to host tissues and mucin components
	Injection of type III secretion factors	PcrG, PcrV, PcrH, PopB, and PopD proteins form injection bridge for type III effectors
Outer membrane	Siderophore receptors	Provides iron for microbial growth and survival
	Efflux pumps	Remove antibiotics
Secretion systems		
Type II	Elastase, lipase, phospholipases, chitin-binding protein, exotoxin A and others	Variety of proteolytic, lipolytic, and toxic factors; degrade host immune effectors
Type III	ExoS, ExoT, ExoU, ExoY	Intoxicates cells (ExoS, ExoT); cytotoxic (ExoU), disrupts actin cytoskeleton
Type VI	Cytoplasmic and membrane-associated proteins, ATPases, lipoproteins, Hcp1 protein	Poorly characterized but found in animal studies to be needed for optimal virulence, particularly in chronic infection
Iron acquisition	Pyoverdin, pyochelin	Scavenge iron from the host for bacterial use
Secreted toxins	Hemolysins, rhamnolipid phospholipases	Kill leukocytes, hemolysis of red cells, degrade host cell surface glycolipids
Secreted oxidative factors	Pyocyanin, ferripyochelin	Produce reactive oxygen species: H_2O_2, O_2^- Inflammatory, disrupts epithelial cell function
Quorum sensing	LasR/LasI, RhlR/RhlI, PQS	Biofilm formation, regulation of virulence factor secretion

ATPases = adenosine triphosphatases; PQS = *Pseudomonas* quinolone signal.

the context of regulation of gene transcription and production of virulence factors.

Both intrinsic (mutation-driven) and transferable (such as β-lactamase production) mechanisms confer *Pseudomonas* resistance to antibiotics, including carbapenems. Transferable resistance to aminoglycosides entails modification of aminoglycosides through phosphoryltransferases, acetyltransferases, nucleotidyltransferases, or ribosomal methyltransferases (RmtA, RmtB, RmtC, RmtD, and ArmA). Chromosomally encoded resistance involves the following:

- AmpC cephalosporinase overproduction or derepression;
- porin OprD, which facilitates diffusion of basic amino acids and carbapenems;
- class D oxacillinase (OXA-50); and
- efflux-mediated resistance through efflux pumps of the RND (resistance-nodulation-division) family, such as the MexXY, which among others affects fluoroquinolones, β-lactams, and aminoglycosides; MexAB-OprM and MexCD-OprJ, which among others affect fluoroquinolones and β-lactams; and MexEF-OprN (F), MexGHI-opmD (F), MexVW (F), and MexPQ-OpmE, which among others affect fluoroquinolones.

Chromosomally mediated resistance also is related to mutations in the quinolone resistance-determining region of gyrase A and the topoisomerase IV gene parC, which hamper the attack of fluoroquinolones on the bacterium. Aminoglycoside-inactivating enzymes are related to the *P. aeruginosa* chromosome and involve 3-*N*-aminoglycoside acetyltransferases [AAC(6)-I, AAC(6)-II, AAC(3)-I, AAC(3)-II, AAC(3)-III, AAC(3)-IV] or nucleotidyltransferases ANT(2)-I and ANT(4)-II and phosphotransferase APH(3)-VI. Impermeability resistance to aminoglycosides is also chromosomally mediated.

The production of such powerful enzymes as the metallo-β-lactamase (such as VIM-1, VIM-2, VIM-4, and IMP-29) is capable of destroying the carbapenem antibiotics, which are important for the treatment of patients with multidrug-resistant gram-negative bacteria. Furthermore, as reported in a study from French intensive care units,[6] additional intrinsic mechanisms may be present concurrently, such as the complete loss of porin OprD as a consequence of mutations or gene disruption by insertion sequences. Downregulation of OprD may be coupled with overexpression of efflux pumps (MexXY, MexEF-PprN, CzcCBA).

Pathology

The pathologic spectrum of *P. aeruginosa* infections depends on the site afflicted. Hemorrhage and necrosis may be present in severe *Pseudomonas* infections, such as in pneumonia and endocarditis. Notably, as regards the skin in the case of ecthyma gangrenosum, bacteria invade the arteries and veins of the skin, but there is little accompanying inflammation. This is reflected in the small quantity, if any, of pus present in these skin lesions. The pathology of *B. pseudomallei* infection is in sharp contrast, with intense inflammation leading to abscess formation and necrosis in the affected organs, as in the skin, liver, spleen, or lungs.

CLINICAL MANIFESTATIONS

A constellation of manifestations are included in the clinical spectrum of *P. aeruginosa* infections. There are no symptoms or signs to effectively discriminate *P. aeruginosa* infection from infections by other pathogens. Even ecthyma gangrenosum, which was once thought to represent a unique effect of *P. aeruginosa* infection, can be caused by other bacteria, such as *S. aureus* or *Citrobacter freundii*.

Febrile Neutropenia

P. aeruginosa infections during febrile neutropenia (Chapters 167 and 281) have a cardinal role. It is the organism against which empirical coverage must always be included. This principle remains unaltered in the span of time. The importance of *P. aeruginosa* infection in neutropenic patients has not diminished, and the antimicrobial resistance of the bacterium has evolved to a painstaking therapeutic challenge. Mortality is high if infection is not appropriately treated by empirical therapy. The classic clinical syndromes in febrile neutropenic patients are bacteremia, pneumonia, and soft tissue infection, mainly manifested as ecthyma gangrenosum.

Bacteremia

Bacteremia due to *P. aeruginosa* remains one of the most difficult challenges a physician may face. It is usually caused by primary infection at different sites, such as pneumonia, urinary tract infection, complicated intra-abdominal tract infection (peritonitis, abscess), and endocarditis. Manifestations can include those of sepsis and septic shock, that is, fever, tachycardia, tachypnea, hypotension, and mental status changes ranging from confusion to coma (Chapter 108). Multiorgan failure with adult respiratory distress syndrome and acute renal failure along with coagulation defects (disseminated intravascular coagulation) may occur. In the ventilated patient, an increased index of suspicion for *P. aeruginosa* must be present in case of clinical deterioration.

Eye Infections

Keratitis is among the most frequent type of disease seen, and it is associated with contact lens wear, especially extended-wear lenses. However, any form of trauma may predispose to keratitis by direct inoculation into tissue, including surgery and burns. *P. aeruginosa* keratitis is a medical emergency because of the speed with which it can progress and lead to loss of vision. Pain and redness of the eye are cardinal manifestations of keratitis. The entire cornea is opacified, and sometimes perforation occurs.

Endophthalmitis is another fulminant *P. aeruginosa* eye infection that may result from penetrating injuries, surgery, perforation of a corneal ulcer, or seeding from bacteremia. Severe pain, chemosis, decreased visual acuity or even loss of vision, anterior uveitis, vitreous involvement, and panophthalmitis are manifestations of endophthalmitis.

Other rarer eye infections include orbital cellulitis in neutropenic patients and gangrene necrosis of the eyelids, both of which are metastatic foci of bacteremia.

Ear Infections

Acute otitis externa presenting with otalgia (ear pain) is commonly seen in children and results from infection of moist, macerated skin of the external ear canal. The source of the organism is likely to be hot tubs or swimming pools (swimmer's ear), particularly if they are not sufficiently chlorinated. The natural history is usually resolution without sequelae, but chronic drainage occurs in some patients. Chronic suppurative otitis media has been associated with *P. aeruginosa*. The main clinical manifestation is drainage of fluid. Cultures are usually polymicrobial, including *P. aeruginosa*. In a third of patients, it is found in isolation.

One of the most dramatic clinical manifestations of *Pseudomonas* infections is malignant otitis externa. The diagnosis is made easily as long as there is a high index of suspicion, which should be the case in diabetics and patients with AIDS. Although literally a misnomer (as this is not a cancerous process), the fulminant evolution to death if it is not diagnosed and treated appropriately justifies this nomenclature. It usually afflicts the diabetic patient and is manifested with pain in the ear accompanied by fever (not always), drainage, and nerve palsies that may even be bilateral. Traction of the pinna will elicit pain in the majority of patients. Most commonly, cranial nerves VI to XII (in various combinations) may be involved, and thus hoarseness and dysphagia accompanying facial paralysis may be evident. Mental status may be affected (obtundation, coma) and signifies intracranial spread of the infection. A characteristic of the disease is the presence of granulation tissue at the junction of bone and cartilage in the meatus, in contrast with the generally intact tympanic membrane. Culture specimens should be taken from the external auditory canal. Computed tomography or magnetic resonance imaging findings include temporomandibular joint destruction, infratemporal fossa or nasopharyngeal soft tissue involvement, and evidence of meningitis or empyema.

Ventilator-Associated Pneumonia and Health Care–Associated Pneumonia

The respiratory tract is among the most frequent sites of infection due to *P. aeruginosa.* This organism is a well-established and frequent cause of ventilator-associated pneumonia (VAP) and health care–associated pneumonia.[7] Identification of the bacterium as a culprit can be based on culture of endotracheal tube aspirates as well as bronchoalveolar lavage fluid in the corresponding clinical setting. Ventilator-associated tracheobronchitis is usually far from innocuous, and treatment is needed to avoid progression to the more severe VAP. Obviously, for the ventilated patient, the liberation of the patient from the mechanical ventilation when feasible is a key management issue. Even when only colonization of the tracheobronchial tree is thought to be present, a watchful and vigilant eye by the clinician has to be in place as the pathogen is notorious.

Pseudomonas necrotizing pneumonia is not uncommon, and its significance is reflected by the fact that the currently suggested 8 days of antimicrobial treatment for most microbial causes of VAP does not apply to VAP related to *P. aeruginosa*. The duration of treatment of VAP due to nonfermen-

tative, gram-negative bacteria (such as *P. aeruginosa* and *A. baumannii*) has been proposed by the American Thoracic Society to be 14 days. Pneumonia due to *P. aeruginosa* can also be community acquired.

Chronic Respiratory Tract Infections

P. aeruginosa is responsible for chronic infections of the airways associated mainly with CF and chronic obstructive pulmonary disease. A description and management of *P. aeruginosa* infection in patients with CF can be found in Chapter 89. Patients with advanced chronic obstructive pulmonary disease may become infected by *P. aeruginosa* and present with an exacerbation.

Another chronic infection of the respiratory tract associated with *P. aeruginosa* is diffuse panbronchiolitis. Diffuse panbronchiolitis affects mainly Asian populations. Nevertheless, forms of this disease have also been reported in white, Hispanic, and African American patients. Criteria have been established for diagnosis of this disease[8]:

1. persistent cough, sputum and exertional dyspnea;
2. history of chronic paranasal sinusitis;
3. bilateral diffuse small nodular shadows on a plain chest radiograph or centrilobular micronodules on chest computed tomography images;
4. coarse crackles;
5. FEV_1/FVC <70% and Pao_2 <80 mm Hg; and
6. titer of cold hemagglutinin >64.

Criteria for definite diagnosis are a compilation of criteria 1 to 3 plus two criteria from criteria 4 to 6. *P. aeruginosa* is isolated in advanced stages of the disease. It shows similarities to CF in terms of respiratory tract involvement, but it is characterized by the lack of affliction of other systems (pancreas, genital tract). Furthermore, the amount of sputum produced is usually at least 50 to 100 mL/day, and the genetic background is different, with associations to HLA-Bw52 and HLA-A11 reported.

Bone and Joint Infections

P. aeruginosa is not a frequent cause of bone or joint infections (Chapter 272) in patients without foreign bodies. Such infections result from bacteremia, direct inoculation into bone, or spread from contiguous infection. Bacteremia secondary to either injection of contaminated illicit drugs or infective endocarditis in the population of intravenous drug users has been well documented to cause vertebral osteomyelitis and septic arthritis of the sternoclavicular joint. The clinical manifestations of vertebral *P. aeruginosa* osteomyelitis are more indolent than those of staphylococcal osteomyelitis. The duration of symptoms in the addict population with vertebral osteomyelitis is generally prolonged, ranging from weeks to months. Tenderness of the affected region has to be elicited, and there may be a decreased range of motion. Low-grade fever is more likely to be encountered than the high fever associated more classically with staphylococcal osteomyelitis. Sternoclavicular and sacroiliac septic arthritis from *P. aeruginosa* is seen almost exclusively in intravenous drug users. It may occur with or without endocarditis, but a primary site of infection is often not found. Its causative role seems to lag behind *S. aureus* in the affliction of the sternoclavicular joint. *P. aeruginosa* is also a cause of bone infections when medical material is present, such as that used in orthopedic surgery or neurosurgery.

Pseudomonas osteomyelitis of the foot most frequently follows puncture wounds through sneakers. The bacterium has been found between the rubber sole layers of sneakers in many cases. Most of these cases are reported in children, but it is also seen in adults. The main manifestation is pain in the foot, and there may be superficial cellulitis around the puncture wound and tenderness on deep palpation of the wound. Also, a group of patients that seems prone to *Pseudomonas* infections are those with Charcot's arthropathy. Prolonged hospital stay and more surgical operations are associated with *P. aeruginosa* secondary bone infection in patients with Charcot's arthropathy. *P. aeruginosa* also has the potential to affect the diabetic foot, especially when there is high local prevalence of the bacterium, warm climate, and frequent exposure to water.

Central Nervous System Infections

Involvement of the central nervous system is almost always secondary to a surgical procedure or penetrating head trauma and is rare after bacteremia. The entities seen most often are postoperative or post-traumatic meningitis, subdural empyema, and epidural infections resulting from initial contamination of access areas. Brain abscess secondary to embolic disease from endocarditis may be seen. Extension of the infection in necrotizing malignant otitis to the brain heralds an ominous clinical course. The cerebrospinal fluid profile of *P. aeruginosa* meningitis is that of pyogenic meningitis. Brain abscess and epidural and subdural empyema generally require surgical drainage in addition to antibiotics.

Urinary Tract Infections

P. aeruginosa urinary tract infections usually occur as a complication of the presence of a foreign body, such as a catheter or stent in the urinary tract or an obstruction (mainly stone or malignant neoplasm) of the urinary system, or after instrumentation or surgery in the urinary tract. Notwithstanding the relationship between obstructive lesions and *P. aeruginosa* urinary tract infections, there have been descriptions of *P. aeruginosa* urinary tract infections in outpatient children and adults without relevant foreign bodies, stones, or other causes of evident obstruction. *P. aeruginosa* urinary tract infection frequently is associated with bacteremia.

Skin and Soft Tissue Infections, Including Burns

P. aeruginosa causes ecthyma gangrenosum in neutropenic patients. These are small, round lesions that occur either isolated or as aggregates. There is no skin site immune to their presence. More commonly, the limbs and the perineum are affected. The mouth may be involved as well. There is an evolution from vesicles to nodules that become hemorrhagic and necrotic and eventually ulcerate. Thus they best fit in the vesiculonodular type of skin lesion description. Secondary infection of chronic skin ulcers or burns can also occur. Maceration of normal skin, such as from soaking in a hot tub, can lead to superficial infection. Folliculitis and other papular or vesicular lesions have also been attributed to *P. aeruginosa*. Other types of skin and soft tissue involvement are cellulitis, abscesses, and myositis.

Burn wound infections by *P. aeruginosa* constitute one of the most significant problems caused by this organism. A distinct clinical picture of sepsis, in which high colony counts of *P. aeruginosa* exceed 10^5 organisms per gram of tissue, is the defining feature. Patients generally exhibit the progressive formation of a black necrotic eschar, with or without bacteremia. *P. aeruginosa* remains a major pathogen in settings in which burn patients have high rates of infection. The diagnosis may be made by culture of blood or by the pathognomonic clinical picture of an expanding burn lesion caused by infection with *P. aeruginosa*.

Endovascular Infections

P. aeruginosa may cause endovascular infections, including infective endocarditis, mainly of native valves but also of prosthetic ones. In intravenous drug users, the source is generally contaminated material, needles, or other paraphernalia; in this population, *P. aeruginosa* may even lead to outbreaks of endocarditis. The manifestations of *P. aeruginosa* endocarditis resemble those of other forms of acute endocarditis in addicts except that it appears to be more indolent than *S. aureus* endocarditis. *P. aeruginosa* endocarditis occurs also in non–intravenous injection drug users.

Gastrointestinal Infections

P. aeruginosa is a pathogen involved in complicated intra-abdominal infections,[9] usually as part of a polymicrobial infection. It is recovered in cases of secondary peritonitis, tertiary peritonitis, peritonitis associated with continuous ambulatory peritoneal dialysis, and intra-abdominal abscesses. Gastrointestinal infections due to *P. aeruginosa* include necrotizing enterocolitis in children and typhlitis in neutropenic patients (neutropenic enterocolitis).

Uncommon *P. aeruginosa* Infections

P. aeruginosa can cause a number of infrequently seen syndromes: noma neonatorum, a necrotizing mucosal and perianal infection of newborns; toe web infections; the "green nail syndrome" caused by *P. aeruginosa* paronychia as a result of diffusion of pyocyanin into the nail bed; and *Pseudomonas* hot-foot syndrome, which is manifested with tender plantar nodules. Shanghai fever is a sporadic community-acquired disease of previously healthy infants that is manifested as a necrotizing enteritis with fever and diarrhea and may lead to bowel perforation, seizures, and ecthyma gangrenosum. Laboratory parameters include leukopenia, thrombocytopenia, high C-reactive protein levels, coagulopathy, and hypoalbuminemia. The mortality is approximately 15%.

Infection due to *Pseudomonas* spp Other than *P. aeruginosa*

Infection due to other *Pseudomonas* spp may occur. *P. fluorescens* may lead to bacteremia associated with central venous catheters or transfusion-related bacteremia. Notably, a multistate outbreak due to contaminated heparinized

saline flush occurred in the United States. Reports of *P. stutzeri* infections include peritonitis, meningitis, endocarditis, pneumonia, bacteremia, and endophthalmitis. *P. putida* has been reported as a causative pathogen in bacteremia, pneumonia, cholecystitis, cholangitis, and skin and soft tissue infections.

DIAGNOSIS

Diagnosis rests on the culture of the pathogen from various human biologic samples or fluids pertinent to the clinical presentation. Newer methods, including multiplex polymerase chain reaction tests, which are based on the detection of the bacterial DNA, have received clearance from the Food and Drug Administration (FDA) for use in clinical practice. They are especially of value in the setting of the intensive care unit, where speed is important in decreasing the mortality associated with delay in treatment of sepsis. Serology is not helpful in the diagnosis of *P. aeruginosa* infections. Pulsed-field gel electrophoresis, restriction fragment length polymorphism, multilocus sequence typing, and random amplified polymorphic DNA polymerase chain reaction are used mainly for epidemiologic purposes.

The differential diagnosis lies between *P. aeruginosa* and pathogens that can lead to the same spectrum of diseases, especially in the hospital environment. External otitis, infection after a nail injury, or infection after immersion in water can provide some important clues that narrow the differential diagnosis spectrum in their respective settings. Nevertheless, diagnostic considerations have to include other pathogens as well. For example, fungal otitis has to be included in the differential diagnosis of external otitis as well as of other causes of earache.

In addition, the specific population of patients (e.g., neutropenic patients, patients with burns and wounds, CF patients, patients who have medically inserted equipment or foreign bodies) should be considered to have a high pretest probability of having a *P. aeruginosa* infection. Other types of infection in the hospital environment necessitate coverage for *Pseudomonas* until microbiologic data are available either from cultures or from molecular tests detecting the bacterium's DNA. A de-escalation protocol can then be performed to narrow the antibiotic spectrum.

At times, clinical acumen has to differentiate frank infection due to *P. aeruginosa* from colonization mainly on the basis of local symptoms and signs (pain, erythema, necrosis, exudate, edema, loss of function, dysuric symptoms, productive cough) or systemic symptoms and signs (fever, hypotension, tachycardia, dyspnea, hypoxia, rash, organ failure, radiologic change of previous imaging) of infection.

TREATMENT

Rx

Treatment of *P. aeruginosa* Bacteremia

Guidelines from the Surviving Sepsis Campaign for patients with severe infections associated with respiratory failure and septic shock suggest combination therapy with an extended-spectrum β-lactam and either an aminoglycoside or a fluoroquinolone for *P. aeruginosa* bacteremia. Empirical combination therapy should not be administered for more than 3 to 5 days. De-escalation to the most appropriate single-agent therapy should be performed as soon as the susceptibility profile is known.[10] This recommendation is based on the higher possibility that at least one of the antibiotics used will have activity against the potentially multidrug-resistant pathogen. This recommendation is not based on better synergistic effect between the two antibiotics. Indeed, evidence-based data from meta-analyses suggest that a combination may not provide an advantage in terms of clinical outcomes, such as cure of the infection, or less emergence of resistance. Furthermore, a combination of a β-lactam and an aminoglycoside carries a higher rate of nephrotoxicity. There is therefore currently a trend that a β-lactam/β-lactamase inhibitor (piperacillin-tazobactam) or an antipseudomonal carbapenem (meropenem or imipenem-cilastatin) may be used alone as monotherapy without compromising patient outcomes. Nevertheless, monotherapy with an aminoglycoside is not suggested as it is associated with poor clinical outcomes. Local in vitro antimicrobial susceptibility data regarding the level of resistance of local clinical isolates of *P. aeruginosa* to various antibiotics have to be taken into consideration.[11]

Removal of an infected vascular catheter or another infected foreign device (urinary catheter, implant) may be needed to control device-related infection due to *P. aeruginosa*. Source control of an infection nidus (drainage of an abscess or empyema, excision of necrotic tissue) is also of paramount significance.

Depending on the antibiotic susceptibility of *P. aeruginosa* isolates routinely found in a specific setting, one of the following regimens would be appropriate for *P. aeruginosa* bacteremia, provided renal function as assessed by creatinine clearance is relatively normal (>50 to 60 mL/minute): intravenous piperacillin-tazobactam, 3.375 to 4.5 g every 6 to 8 hours; ceftazidime, 2 g every 8 hours; cefepime, 2 g every 8 to 12 hours; meropenem, 1 to 2 g every 8 hours; imipenem, 0.5 to 1 g every 6 hours; doripenem, 0.5 g (1-hour infusion) every 8 hours; or aztreonam, 1.5 to 2 g every 6 to 8 hours (aztreonam has been used mainly for patients with β-lactam allergy). The addition of an aminoglycoside to the other regimens depends on the level of resistance to β-lactam antibiotics seen at any given institution. If administration of a second drug is indicated, amikacin 15 mg/kg every 24 hours may be added to the β-lactam antibiotic therapy. Addition of ciprofloxacin 400 mg every 8 to 12 hours IV (instead of an aminoglycoside) has been suggested to fare equally well.

Antibiotics that have to be considered on a compassionate basis if resistance to carbapenems with antipseudomonal spectrum is encountered include colistimethate sodium (polymyxin E or colistin parenteral form)[12] and polymyxin B. Two forms of colistin are commercially available: colistin sulfate and colistimethate sodium (also called colistin methanesulfonate, pentasodium colistimethanesulfate, and colistin sulfonyl methate). Colistimethate sodium is less potent than colistin sulfate. Specifically, activity as assessed biologically is 20,500 IU/mg for colistin sulfate and approximately 12,500 IU/mg for colistimethate sodium. Colistimethate sodium is produced by the reaction of colistin with formaldehyde and sodium bisulfite. Colistin sulfate is administered orally (tablets or syrup) in bowel decontamination regimens and topically as a powder for the treatment of bacterial skin infections. Colistimethate sodium is available in parenteral formulations. The term *colistin* throughout this chapter refers to the formulation of colistimethate sodium, except if otherwise specified.

The suggested dosage of intravenous colistin for adult patients with normal renal function is different for manufacturers in the United States and the United Kingdom. Specifically, recommended dosage in the United States is 2.5 to 5 mg/kg/day of colistin base activity (75,000 to 150,000 IU/kg), divided into two to four equal doses (1 mg of colistin base activity, ≈30,000 IU; 1 mg of colistimethate sodium activity, ≈12,500 IU). In the United Kingdom, the suggested dosage is 4 to 6 mg/kg/day of colistimethate sodium activity (50,000 to 75,000 IU/kg) in three divided doses for adults and children with body weight of 60 kg or less and 80 to 160 mg of colistimethate sodium activity (1 to 2 million IU) every 8 hours for body weight of more than 60 kg. However, we and others have treated patients with higher daily doses of intravenous colistin, up to 720 mg/day (9 million IU) in three divided doses. For serum creatinine level of 1.3 to 1.5 mg/dL, 1.6 to 2.5 mg/dL, or 2.6 mg/dL and higher, the dosage of intravenous colistin recommended by the manufacturers is 160 mg (2 million IU) every 12 hours, 24 hours, or 36 hours, respectively. During hemodialysis, the recommended dose is 80 mg (1 million IU) after each hemodialysis session. However, further studies are needed to better clarify the appropriate dosing regimens of colistin, especially for patients with renal dysfunction or failure. In a small study, on the basis of pharmacokinetic and pharmacodynamic data, a loading dose of intravenous 9 million IU of colistimethate sodium followed by 4.5 million IU every 12 hours was used successfuly.[13] When dosing colistin, clinicians should be aware of the existing differences in dosage recommendations based on the specific formulation of colistin used. Lately, the parenteral formulation of fosfomycin sodium has been used in various European countries in combination with other antibiotics with an antipseudomonal spectrum.[14]

Another important consideration is the prolonged intravenous infusion of the antibiotics to exploit the pharmacodynamic properties of β-lactams in the treatment of *P. aeruginosa* and other infections. β-Lactam antibiotics have time-dependent antimicrobial activity, and thus their concentration in blood achieved by their prolonged intravenous administration is above the minimum inhibitory concentration for longer periods (proportion of the time between doses above minimum inhibitory concentration). There are data, mainly from nonrandomized studies and a relevant meta-analysis, that support the fact that a prolonged infusion of piperacillin-tazobactam during 4 hours may provide a survival benefit.[15] Meropenem is usually infused in a relatively short infusion of 30 minutes; an extended meropenem infusion is one that extends to 3 hours. Doripenem should not be used in pneumonia according to an FDA warning based on a study comparing it with imipenem-cilastatin.

Despite the introduction of new broad-spectrum antibiotics in the antibiotic arsenal, one should remember that ertapenem, ceftaroline, and tigecycline do not possess antipseudomonal activity.

Treatment of Pneumonia

The principles of treatment that apply to bacteremia are also the basis for treatment of pneumonia. Patients with nosocomial pneumonia may have *P. aeruginosa* isolates resistant to many antibiotics, a problem that appears to be becoming progressively worse. Provided renal and liver function is normal, the following recommendations have been made by the American Thoracic Society and the Infectious Diseases Society of America: IV ceftazidime, 2 g every 8 hours; cefepime, 1 to 2 g every 8 to 12 hours; imipenem, 500 mg every 6 hours or 1 g every 8 hours; meropenem, 1 g every 8 hours; or piperacillin-tazobactam, 4.5 g every 6 hours plus an aminoglycoside (amikacin 15 to 20 mg/kg once daily, with a trough level of less than 4 to 5 μg/mL for amikacin).

Aminoglycosides are not optimally active in the lungs at concentrations used for intravenous administration. The administration of aerosolized aminoglycoside may provide adequate drug levels in the tracheobronchial tree.

Tobramycin (300 mg inhaled daily) and inhaled aztreonam lysine (75 mg three times daily for 28 days) have shown safety and efficacy for CF patients and have FDA approval only in patients with CF. Another antibiotic that has been used in different parts of the world on a compassionate basis is the inhaled form of colistin. The effectiveness and safety of inhaled colistin in patients with CF have led to a revival of the use of the medication also in patients with *P. aeruginosa* infection in the critical care setting. Inhaled colistin does not have FDA approval.

Treatment of Other Infections

Standard therapy to treat meningitis due to *P. aeruginosa* includes ceftazidime or cefepime. Alternative therapies are aztreonam, ciprofloxacin, and meropenem; the addition of aminoglycosides to these alternatives should be considered as well. The duration of treatment extends to 3 weeks or 2 weeks after the first sterile cerebrospinal fluid culture. No antimicrobial agent has been approved by the FDA for intraventricular use, and the specific recommendations are not well defined. The daily intraventricular dose is 1 to 8 mg for gentamicin, 5 to 20 mg for tobramycin, 5 to 50 mg (usually 30 mg) for amikacin, 5 mg for polymyxin B, and 10 mg for colistin. The removal of all components of the infected shunt, in combination with appropriate antimicrobial therapy, appears to be the most effective treatment for cerebrospinal fluid shunt infections. The ventriculitis of the shunt infection appears to clear more rapidly with the drainage catheter still in place, and the presence of the catheter allows continued treatment of the hydrocephalus until the infection has cleared. Reshunting (placement of a new shunt) necessitates 10 to 14 days of sterile cerebrospinal fluid culture.

Formulations of aminoglycosides or fluoroquinolones for topical use are recommended for the treatment of keratitis. In cases in which involvement is extensive, ceftazidime or gentamicin may be given by subconjunctival injection. Therapy for endophthalmitis includes both systemic antibiotics at high doses to achieve better concentrations in the eye and intravitreal antibiotics. Ceftazidime has been the most frequently used antibiotic for this entity. Aminoglycosides are also injected subconjunctivally and by intraocular routes and sometimes given intravenously. Adjunctive surgery is generally performed to remove infected vitreous.

Management of otitis externa involves also the use of topical antibiotic agents (otic solutions). Protection of the ear from additional moisture and avoidance of further mechanical injury by scratching are also important. Aminoglycoside-containing otic solutions (neomycin plus polymyxin B and hydrocortisone) and quinolone-containing otic solutions (ciprofloxacin 0.2% or ofloxacin 0.3%) are the most frequently used. Gentle removal of debris and cleaning with a mixture of acetic acid, alcohol, and distilled water may also help. Chronically draining ears may require more intensive topical therapy.

Treatment of malignant external otitis involves débridement of the ear canal, including any necrotic tissue cartilage and adjacent bone, rather than extensive bone débridement or facial nerve decompression. Furthermore, treatment with ciprofloxacin or an antipseudomonal β-lactam (cephalosporin such as ceftazidime or cefepime, carbapenem, or monobactam) is necessary for a period of 6 to 8 weeks.

The majority of *P. aeruginosa* urinary tract infections are complicated. Urinary catheters, stents, or stones should be removed if possible to prevent relapse. In general, 7 to 10 days of antibiotic treatment will suffice, with up to 2 weeks for severe pyelonephritis. Quinolones (ciprofloxacin, levofloxacin) have the pharmacodynamic advantage of excellent concentrations in the urinary tract. Antipseudomonal β-lactam/β-lactamase inhibitors and carbapenems (doripenem is approved for the treatment of complicated and uncomplicated urinary tract infections and intra-abdominal infections) are equally acceptable, given their normally high levels of urinary excretion.

Management of *P. aeruginosa*–infected burn wounds is both surgical and medical. Extensive débridement of colonized eschar or necrotic tissue is required in addition to antibiotic treatment. Medical treatment follows the principles detailed in the bacteremia treatment section.

For treatment of chronic osteomyelitis due to *Pseudomonas*, ciprofloxacin has been used extensively. Experts infer that oral treatment is acceptable as an alternative to parenteral treatment.[16] Of note, an increased dosage of orally administered ciprofloxacin (750 mg twice a day) is used in the majority of the relevant studies. There are relatively few data regarding the appropriate duration of treatment. A duration of 6 to 12 weeks is most commonly reported.

Infections Caused by Organisms at One Time Classified as Pseudomonads

Most of these infections are due to members of the genus *Burkholderia*, including *B. cenocepacia*, a complex of about 10 related species previously referred to as genomovars that are phenotypically similar but distinguished primarily by nucleic acid sequences, as well as *B. mallei* and *B. pseudomallei*, causes of glanders and melioidosis, respectively.

S. maltophilia causes a variety of organ afflictions. It is manifested as pneumonia, acute exacerbation of chronic obstructive pulmonary disease, bacteremia, soft tissue and skin infection, cellulitis, myositis, osteomyelitis, catheter-related bacteremia or septicemia, meningitis, endophthalmitis, keratitis, scleritis, dacryocystitis, endocarditis, urinary tract infection, and biliary sepsis. One notices the commonalities with the clinical manifestations of *P. aeruginosa* as well as the commonalities regarding the sources of infection. Whereas it has been regarded as a rare pathogen in the past, this is far from the truth. Indeed, it is the 11th most frequently cultured microorganism in a U.S. multiple hospital study covering 1993 to 2004 (4.3% of 74,934 gram-negative bacilli). Furthermore, the prevalence of *S. maltophilia* in patients with CF has increased in a relevant study from 6% to 12.7% from 1995 to 2008. Selective media (vancomycin–imipenem–amphotericin B, gram-negative selective agar, BTB, and SM2i) have been developed to ease detection of *S. maltophilia*, especially because the pathogen is frequently co-cultured in samples of polymicrobial infections. Polymerase chain reaction amplification of 16 rRNA in blood has also been used. The bacterium is usually susceptible to trimethoprim-sulfamethoxazole, which is regarded as the first-choice antibiotic.[17] However, resistant strains are emerging. Other antibiotics that have been effective against *S. maltophilia* isolates include ciprofloxacin, ceftazidime, and ticarcillin-clavulanate; sometimes these are used in combination with trimethoprim-sulfamethoxazole. Inherent resistance to carbapenems is a notable characteristic of this bacterium.

Melioidosis is an infection due to *B. pseudomallei* and occurs mainly in Southeast Asia and Australia. A notable current addition to these geographic regions is the Indian subcontinent and Sri Lanka. One should also be aware of potential niduses in the Americas (Mexico and the northern part of South America) and Africa (Madagascar). Risk factors for acquisition of melioidosis include renal failure, diabetes mellitus, heavy alcohol consumption, chronic respiratory disease, thalassemia, glucocorticoid therapy, and cancer. It can also afflict immunocompetent travelers to the regions mentioned. Interestingly, there is a seasonal association with the rainy season in more than three quarter of cases. Clinical manifestations vary from asymptomatic infection to localized skin infection, pneumonia, and fulminant sepsis due to bacteremia. The formation of abscesses is a usual feature of the disease. Other manifestations include septic arthritis, osteomyelitis, prostatitis, neurologic manifestations such as brain stem encephalitis associated with cranial nerve palsies or myelitis with peripheral motor weakness, and kidney and spleen involvement. Suppurative parotitis, even bilateral in 10%, is a feature present in patients with melioidosis in Thailand and Cambodia. Parotitis afflicts mainly children. The bacterium may remain dormant and then reactivate. The portal of entry includes the respiratory system, the skin, and the gastrointestinal system. Indeed, recurrence of melioidosis is due to reactivation in approximately three quarters of cases. There is a need for prolonged treatment of the bacterium that entails a 2- to 4-week regimen of intravenous antibiotic administration: ceftazidime, 2 g every 8 hours; meropenem, 1 g every 8 hours; or imipenem, 1 g every 8 hours.[18] An extended course of oral eradication with trimethoprim-sulfamethoxazole for 3 to 6 months is necessary after initial intravenous antibiotics. In the case of allergy or adverse events, amoxicillin-clavulanate and doxycycline have been proposed as alternatives to trimethoprim-sulfamethoxazole. Of note, *B. pseudomallei* is a bioterrorism factor B, and thus appropriate infection control and notification of authorities are mandatory.

Glanders is an equine infection due to *B. mallei*. Humans acquire the disease from contact with horses or more rarely donkeys or mules in an occupational setting. It mainly is manifested with tracheobronchitis, pneumonia, skin lesions, or lymphadenopathy. The disease has been eradicated in many countries. Nevertheless, cases can occur especially in association with an occupational risk in veterinarians, veterinary students, farriers (hoof care workers), flayers (hide workers), transport workers, soldiers, slaughterhouse personnel, farmers, and horse fanciers. *B. mallei* remains a significant cause of zoonosis, and therefore appropriate veterinary surveillance is necessary, especially because *B. mallei* is also a significant bioterrorism agent (class B). Treatment of glanders in humans is based on limited data. The approach to treatment is similar to that used for melioidosis, with intravenous meropenem, imipenem, or ceftazidime initially and then oral trimethoprim-sulfamethoxazole for 3 to 12 months.

B. cepacia is a bacterial pathogen that is a plant and human pathogen. Its classification has evolved, as has its name. It was previously known as *Pseudomonas cepacia* and then *Xanthomonas cepacia*. *B. cepacia* belongs to the *B. cepacia* complex and is a pathogen mainly of patients with CF and chronic granulomatous disease. Relatively recent reports have implicated *B. cepacia* in infections of the lung, blood, and other sites in immunocompromised patients and even in immunocompetent patients and children in the hospital environment due to outbreaks. Aerosolized medications, chlorhexidine solutions, napkins, and prefabricated clothes along with horizontal transmission have led to significant morbidity and mortality. Patients with CF typically present initially with asymptomatic carriage. Nevertheless, progressive decline of lung function is associated with *B. cepacia*. The most dramatic presentation of the lung ailment is the cepacia syndrome, a fulminant lung infection often accompanied by bacteremia. Other types of infection include meningitis, pericarditis, endocarditis, cholangitis, peritonitis, and abscesses in the abdomen, perineum, or scrotum.

PREVENTION

Primary prevention of *Pseudomonas* infections applies to preventing pollution of water by *Pseudomonas*. This applies to both the public environment and the hospital environment. Outbreaks have been linked to aquatic

environments such as whirlpools, swimming pools, and spas. Thus, control of growth of this organism in the recreational environment by proper antibacterial treatment of water is essential, comparable to the control practiced in hospitals. Contamination of various equipment and devices (i.e., breast implants, ocular implants, and sinus irrigation devices) must be avoided. Handwashing cannot be emphasized enough in the prevention of infections due to *P. aeruginosa*, *S. maltophilia*, and *B. cepacia*. Hospital point-of-use water filtration has also been employed in the battle against *P. aeruginosa* and *S. maltophilia* infections. Barrier nursing practices may decrease horizontal transmission, especially in the critical care environment. Isolation of patients infected with extensively drug-resistant and pandrug-resistant *P. aeruginosa* and strict infection control measures (dedicated personnel and equipment, use of gowns and gloves) may be needed to avoid intrahospital spread. The most important measure is the washing of hands. Nevertheless, even antiseptic solutions may be contaminated by *P. aeruginosa*, *S. maltophilia*, and *B. cepacia*. Factors that shorten hospital length of stay and decrease the use of antibiotics are likely to decrease the incidence of these infections.

PROGNOSIS

P. aeruginosa infections carry a high mortality, even with treatment. Mortality rates of 26 to 36% have been reported. Differences also exist between appropriately and inappropriately treated *P. aeruginosa* infections, with mortality rates even double when inappropriate antibiotics are used. Furthermore, *P. aeruginosa* infections are associated with increased length of hospitalization and medical costs. Loss of vision is the grave outcome of ophthalmic infection. Antibiotic resistance associated with *P. aeruginosa* will most likely continue to pose an enormous burden on human lives and financial resources in years to come.

S. maltophilia infections are associated with attributable mortality of up to about 37% if they are not appropriately treated. Melioidosis is associated with a 14% mortality rate in Australia and up to 40% in countries in Southeast Asia when it is treated. Fortunately, reports of glanders are rare in humans. Mortality of glanders with treatment is about 40 to 50%. *B. cepacia* has a predilection for CF patients and often signals impaired overall prognosis in these patients. Furthermore, *B. cepacia* seems to adversely affect transplantation. Indeed, in patients presenting with the cepacia syndrome, survival is unusual. Furthermore, many transplant centers do not support transplantation when *B. cepacia* is present because of the grave prognosis associated with the infection in the post-transplantation period.

Grade A Reference

A1. Chetchotisakd P, Chierakul W, Chaowagul W, et al. Trimethoprim-sulfamethoxazole versus trimethoprim-sulfamethoxazole plus doxycycline as oral eradicative treatment for melioidosis (MERTH): a multicentre, double-blind, non-inferiority, randomised controlled trial. *Lancet*. 2014;383:807-814.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Sievert DM, Ricks P, Edwards JR, et al. National Healthcare Safety Network (NHSN) Team and Participating NHSN Facilities. Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009-2010. *Infect Control Hosp Epidemiol*. 2013;34:1-14.
2. Edelstein MV, Skleenova EN, Shevchenko OV, et al. Spread of extensively resistant VIM-2–positive ST235 *Pseudomonas aeruginosa* in Belarus, Kazakhstan, and Russia: a longitudinal epidemiological and clinical study. *Lancet Infect Dis*. 2013;13:867-876.
3. Brooke JS. *Stenotrophomonas maltophilia*: an emerging global opportunistic pathogen. *Clin Microbiol Rev*. 2012;25:2-41.
4. Wiersinga WJ, Currie BJ, Peacock SJ. Melioidosis. *N Engl J Med*. 2012;367:1035-1044.
5. Lovewell RR, Patankar YR, Berwin B. Mechanisms of phagocytosis and host clearance of *Pseudomonas aeruginosa*. *Am J Physiol Lung Cell Mol Physiol*. 2014;306:L591-L603.
6. Fournier D, Richardot C, Müller E, et al. Complexity of resistance mechanisms to imipenem in intensive care unit strains of Pseudomonas aeruginosa. *J Antimicrob Chemother*. 2013;68:1772-1780.
7. Wilke M, Grube R. Update on management options in the treatment of nosocomial and ventilator assisted pneumonia: review of actual guidelines and economic aspects of therapy. *Infect Drug Resist*. 2013;7:1-7.
8. Kudoh S, Keicho N. Diffuse panbronchiolitis. *Clin Chest Med*. 2012;33:297-305.
9. Solomkin JS, Mazuski JE, Bradley JS, et al. Diagnosis and management of complicated intra-abdominal infection in adults and children: guidelines by the Surgical Infection Society and the Infectious Diseases Society of America. *Clin Infect Dis*. 2010;50:133-164.
10. Dellinger RP, Levy MM, Rhodes A, et al. Surviving Sepsis Campaign Guidelines Committee including The Pediatric Subgroup. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Med*. 2013;39:165-228.
11. Nathwani D, Raman G, Sulham K, et al. Clinical and economic consequences of hospital-acquired resistant and multidrug-resistant Pseudomonas aeruginosa infections: a systematic review and meta-analysis. *Antimicrob Resist Infect Control*. 2014;3:32.
12. Falagas ME, Rafailidis PI, Ioannidou E, et al. Colistin therapy for microbiologically documented multidrug-resistant gram-negative bacterial infections: a retrospective cohort study of 258 patients. *Int J Antimicrob Agents*. 2010;35:194-199.
13. Dalfino L, Puntillo F, Mosca A, et al. High-dose, extended-interval colistin administration in critically ill patients: is this the right dosing strategy? A preliminary study. *Clin Infect Dis*. 2012;54:1720-1726.
14. Reffert JL, Smith WJ. Fosfomycin for the treatment of resistant gram-negative bacterial infections. Insights from the Society of Infectious Diseases Pharmacists. *Pharmacotherapy*. 2014;34:845-857.
15. Falagas ME, Tansarli GS, Ikawa K, et al. Clinical outcomes with extended or continuous versus short-term intravenous infusion of carbapenems and piperacillin/tazobactam: a systematic review and meta-analysis. *Clin Infect Dis*. 2013;56:272-282.
16. Spellberg B, Lipsky BA. Systemic antibiotic therapy for chronic osteomyelitis in adults. *Clin Infect Dis*. 2012;54:393-407.
17. Wang YL, Scipione MR, Dubrovskaya Y, et al. Monotherapy with fluoroquinolone or trimethoprim-sulfamethoxazole for treatment of Stenotrophomonas maltophilia infections. *Antimicrob Agents Chemother*. 2014;58:176-182.
18. Dance D. Treatment and prophylaxis of melioidosis. *Int J Antimicrob Agents*. 2014;43:310-318.

REVIEW QUESTIONS

1. Which of the following bacteria is most commonly associated with hospital-acquired infections in the United States?

A. *Stenotrophomonas maltophilia*
B. *Burkholderia cepacia*
C. *Pseudomonas aeruginosa*
D. *Burkholderia mallei*
E. *Burkholderia pseudomallei*

Answer: C Data from the National Healthcare Safety Network at the Centers for Disease Control and Prevention indicate that during 2009 to 2010, *P. aeruginosa* was responsible for 8% of a total of 69,475 hospital-acquired infections and was the sixth most frequent culprit among 81,139 pathogens. *S. maltophilia* is a pathogen responsible for hospital-acquired infections but ranks lower in frequency than *P. aeruginosa*. *B. pseudomallei* and *B. mallei* are not causes of hospital-acquired infections in the United States. However, these pathogens are bioterrorism factors B. *B. cepacia* is an important emerging pathogen that is associated with hospital-acquired infections in immunocompromised patients beyond its established role as a pathogen in cystic fibrosis patients. *B. cepacia* hospital-acquired infections rank much lower in frequency than hospital-acquired infections due to *P. aeruginosa*.

2. Which of the following statements regarding *Pseudomonas aeruginosa* is true?

A. It is a nonmotile gram-negative bacterium.
B. It is not the only cause of ecthyma gangrenosum.
C. Endocarditis due to the *P. aeruginosa* is always associated with high fever.
D. Ertapenem is effective as treatment against *P. aeruginosa*.
E. Tigecycline is effective as treatment against *P. aeruginosa*.

Answer: B Other causes of ecthyma gangrenosum include *Staphylococcus aureus, Aeromonas hydrophila, Klebsiella pneumoniae, Morganella morganii, Stenotrophomonas maltophilia, Citrobacter freundii*, and fungi (*Candida albicans, Aspergillus fumigatus*). *P. aeruginosa* is a motile bacterium. Ertapenem and tigecycline are not effective treatment against *P. aeruginosa*. Endocarditis due to *P. aeruginosa* is usually more indolent than endocarditis due to *S. aureus*. It is not always associated with high fever.

3. Which of the following patient groups is at increased risk for the development of *P. aeruginosa* infections?

A. Neutropenic patients
B. Patients with cystic fibrosis
C. Patients with Charcot's arthropathy
D. Patients requiring mechanical ventilation
E. All the above

Answer: E Neutropenia, Charcot's arthropathy, mechanical ventilation, and also burns, cystic fibrosis, cancer, transplantation, diabetes, steroid treatment, foreign bodies, hypogammaglobulinemia, and AIDS are associated with an increased risk for *P. aeruginosa* infection.

4. Which of the following is considered the drug of first choice for *Stenotrophomonas maltophilia*?

A. Doxycycline
B. Meropenem
C. Trimethoprim-sulfamethoxazole
D. Ciprofloxacin
E. Ceftazidime

Answer: C *S. maltophilia* has inherent resistance to carbapenems. Trimethoprim-sulfamethoxazole is the choice for first-line treatment. Ciprofloxacin and ceftazidime have been used in the treatment of *S. maltophilia* infections either alone or in combination but are not first-choice treatment. Tetracyclines have some activity against *S. maltophilia* but are not first-choice options for treatment.

5. The following are true regarding infection due to *Burkholderia pseudomallei* except which one?

A. Southeast Asia, Australia, the Indian subcontinent, and Sri Lanka are locations with an increased risk for this infection.
B. Risk factors for infection include renal failure, diabetes mellitus, and heavy alcohol consumption.
C. The infection is called melioidosis.
D. An extended eradication regimen with oral trimethoprim-sulfamethoxazole for 1 month is necessary to follow initial intravenous treatment.
E. Recurrence of melioidosis is due to reactivation.

Answer: D An extended oral eradication regimen with oral trimethoprim-sulfamethoxazole for 3 to 6 months (not only 1 month) is recommended to follow initial intravenous treatment for the treatment of melioidosis.

307

DISEASES CAUSED BY *ACINETOBACTER* AND *STENOTROPHOMONAS* SPECIES

KEITH S. KAYE AND ROBERT A. BONOMO

ACINETOBACTER SPECIES

DEFINITION

The Pathogen

Acinetobacter species are gram-negative aerobic bacteria that are coccobacillary in shape and are generally described as aerobic, non–lactose-fermenting, nonfastidious, nonmotile, catalase positive, and oxidase negative. Unique among microbial pathogens, the appearance of *Acinetobacter* species visualized with a Gram stain is highly dependent on the life cycle. In the early growth phases, *Acinetobacter* species appear rod shaped. In the stationary phase, they acquire a coccobacillary *Acinetobacter* spp. morphology.

In the context of this chapter, *Acinetobacter* species refers specifically to *Acinetobacter baumannii* and *Acinetobacter baumannii-calcoaceticus* complex. A full list of *Acinetobacter* species is provided in E-Table 307-1. Molecular methods used to identify and to classify *Acinetobacter* are listed in E-Table 307-2.

EPIDEMIOLOGY

Acinetobacter species can colonize all body surfaces and cause infection in almost any organ system. Consequently, there are a number of common clinical syndromes associated with infection by *Acinetobacter.* The most common infections are respiratory (pneumonia), blood stream (bacteremia), urinary tract, wound, skin and soft tissue, and burn infections; osteomyelitis secondary to trauma; and meningitis.[1,2] In general, infection is observed only in critically ill, immunocompromised, or injured hosts. Recently, infections by *Acinetobacter* spp. are being described in patients without significant medical problems from the community setting.

During the past several years, *Acinetobacter* spp emerged in the United States from a pathogen that was primarily found in intensive care units (ICUs) to one that can affect patients in non-ICU wards, patients in long-term care settings, and military personnel with combat injuries acquired in the Middle East.[3] Overall, there are few distinguishing features of *Acinetobacter* infection except for skin manifestations. The frequency of *Acinetobacter* infections is often greater in the summer than in other seasons.

CLINICAL MANIFESTATIONS

Pneumonia

Because of colonization of the oropharynx and tracheostomy tubes in patients on ventilators, the upper respiratory tract is the most common site for infection by *Acinetobacter* species. The two distinct syndromes associated with respiratory tract infection due to *Acinetobacter* are community-acquired pneumonia (CAP) and healthcare-associated pneumonia (HCAP). In tropical regions of China, Asia, Australia, and the South Pacific, CAP due to *Acinetobacter* species is increasingly recognized. In some locations, the incidence can exceed 15%. Reports have highlighted the emergence of *Acinetobacter* as a common cause of CAP in western China.[4] In Saudi Arabia, *Acinetobacter* is the most common pathogen associated with late-onset and recurrent ventilator-associated pneumonia (VAP) in one adult ICU.[5]

The common comorbid conditions predisposing to CAP due to *Acinetobacter* species are mainly chronic obstructive pulmonary disease (emphysema), renal disease, diabetes mellitus, and alcoholism. CAP due to *Acinetobacter* species appears to be associated with a high incidence of bacteremia, acute respiratory distress syndrome, sepsis, and death (mortality rates ≥50%). The reasons for these fulminant presentations are still unknown. Rarely, CAP due to *Acinetobacter* species can be manifested with consolidation and multiple lung abscesses.

More frequently, *Acinetobacter* species are a cause of HCAP, often manifested as VAP. In the United States, *Acinetobacter* species are the third leading cause of VAP. HCAP due to *Acinetobacter* largely resembles the clinical spectrum of gram-negative pneumonias (bilateral infiltrates, pleural effusion, cavitations, hypoxemia, and bacteremia). Most cases are described in patients on ventilators. The main factors associated with HCAP due to *Acinetobacter* species are mechanical ventilation, previous antibiotic exposure, ICU stay, surgery, and underlying pulmonary disease. The major challenge complicating nosocomial pneumonia due to *Acinetobacter* species is the frequent recovery of MDR and sometimes XDR strains. When isolates are MDR or XDR, the options for treatment are limited, and complications quickly arise. HCAP due to MDR *Acinetobacter* has been associated with excess lengths of stay and mortality, although some studies have reported similar outcomes compared with control patients matched by severity of illness and duration of ICU stay.

Bacteremia

Blood stream infection due to *Acinetobacter* species is often a consequence of infection of intravenous catheters (i.e., central line–associated blood stream infection or CLABSI) or is secondary to HCAP due to *Acinetobacter*. *Acinetobacter* is the ninth most common cause of CLABSI among U.S. hospitals reporting to the Centers for Disease Control and Prevention. Less commonly, wound infections can cause bacteremia. In most series,

E-TABLE 307-1 MEMBERS OF THE *ACINETOBACTER* GENUS*

Acinetobacter baumannii, *Acinetobacter* genome sp 2; reference strain ATCC 19606	*Acinetobacter* genome sp 3, reference strain ATCC 19004
Acinetobacter baylyi	*Acinetobacter* genome sp 6
Acinetobacter bouvetii	*Acinetobacter* genome sp 10
Acinetobacter calcoaceticus, *Acinetobacter* genome sp 1, reference strain ATCC 23055	*Acinetobacter* genome sp 11
Acinetobacter gerneri	*Acinetobacter* genome sp 13BJ or 14TU
Acinetobacter grimontii	*Acinetobacter* genome sp 14BJ
Acinetobacter haemolyticus, *Acinetobacter* genome sp 4, reference strain ATCC 17906	*Acinetobacter* genome sp 15BJ
Acinetobacter johnsonii, *Acinetobacter* genome sp 7, reference strain ATCC 17909	*Acinetobacter* genome sp 10
Acinetobacter junii	*Acinetobacter* genome sp 16
Acinetobacter lwoffii	*Acinetobacter* genome sp 17
Acinetobacter parvus	*Acinetobacter* genome sp 13TU, reference strain ATCC 17903 Now called *Acinetobacter* nosocomialis
Acinetobacter radioresistens	*Acinetobacter* genome sp 15TU
Acinetobacter genome sp 12	
Acinetobacter schindleri	*Acinetobacter* genome sp between 1 and 3
Acinetobacter tandoii	*Acinetobacter* genome sp close to 13TU
Acinetobacter tjernbergiae	*Acinetobacter pittii*
Acinetobacter townerii	*Acinetobacter ursingii*
Acinetobacter venetius	

*Genome species designations are added. Where appropriate, the American Type Culture Collection (ATCC) reference strain number is given.

E-TABLE 307-2 IDENTIFICATION AND TYPING METHODS USED TO CHARACTERIZE *ACINETOBACTER* SPECIES

16S rRNA gene restriction analysis
Amplified fragment length polymorphism
$bla_{\text{OXA-51}}$ gene amplification
Polymerase chain reaction coupled with electrospray ionization mass spectrometry
Multilocus sequence typing
Ribotyping
Restriction enzyme digestion of the 16S-23S rRNA intergenic spacer sequences
Sequence analysis of the 16S-23S rRNA gene spacer region
rpoB gene sequencing and its flanking spacers
Matrix-assisted laser desorption/ionization time-of-flight
Polymerase chain reaction–electrospray ionization/mass spectrometry

mortality associated with blood stream infection ranges from approximately 15 to 50%.

Urinary Tract Infection

Urinary tract infections are most commonly caused by enteric gram-negative bacilli; only rarely are these infections caused by *Acinetobacter* species. The indwelling bladder catheter has been implicated as the major risk factor for urinary tract infection (cystitis and pyelonephritis) due to *Acinetobacter* species.

Wound, Burn, and Skin and Soft Tissue Infections

In many clinical series to date, traumatic or postoperative wounds, burns, and skin and soft tissue infections (SSTIs) are the most common causes of *Acinetobacter* infections. Most likely, the combination of antibiotic use, colonization, and compromised or devitalized tissues is responsible. The spectrum of infection can extend from cellulitis to necrotizing fasciitis.

As a result of the outbreak of *A. baumannii* among military personnel in Iraq and Afghanistan, reports of severe wound infections and SSTIs caused by this pathogen are increasing in frequency. Necrotizing SSTI associated with *A. baumannii* occurs in hosts with underlying comorbidities (e.g., trauma, cirrhosis) and is often accompanied by bacteremia. *Acinetobacter* infection has become increasingly common among patients residing in burn units, sometimes resulting in unit-wide outbreaks. Multiple drug resistance and the presence of copathogens frequently complicate treatment. Most cases require surgical débridement and lead to substantial mortality. In addition, the use of central venous catheters and total parenteral nutrition is more common among patients with SSTIs. *Acinetobacter* species–associated SSTIs can be manifested with a peau d'orange appearance, with overlying tiny vesicles (Fig. 307-1), and, when untreated, can progress to necrotizing infection with bullae (hemorrhagic and nonhemorrhagic).

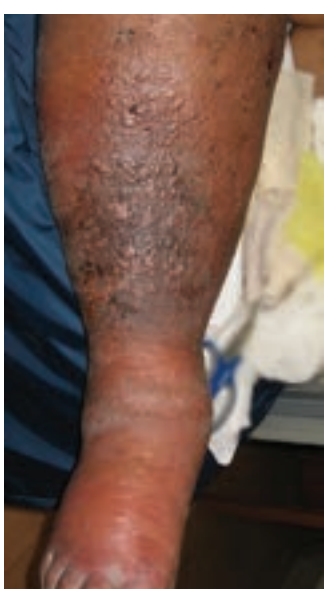

FIGURE 307-1 Cellulitis caused by *Acinetobacter baumannii*. There is characteristic edematous peau d'orange erythema, with associated vesicles that may coalesce to form nonhemorrhagic bullae. (From Guerrero DM, Perez F, Conger NG, et al. *Acinetobacter baumannii*–associated skin and soft tissue infections: recognizing a broadening spectrum of disease. *Surg Infect [Larchmt]*. 2010;11:49-57.)

Osteomyelitis

The conflicts in Iraq and Afghanistan highlighted the first cases of osteomyelitis due to *Acinetobacter* species. Before this time, rare cases of osteomyelitis occurring in soldiers were reported during the Korean and Vietnam wars. Most of the initial reports from the Middle East described "contiguous focus" osteomyelitis. These patients had open fractures or exposed bone, with gross findings of infection: purulence, necrotic tissue, or environmental contamination with exposed bone; temperature higher than 38° C; leukocyte count greater than 12,000/μL; and *Acinetobacter* species identified from culture of deep wound tissue obtained intraoperatively. Frequently, these infections require multiple surgical débridements of necrotic bone.

Meningitis

Acinetobacter meningitis is occasionally found in the post-neurosurgical setting, with mortality exceeding 15 to 30%. Infection often involves intraventricular catheters. Although *Acinetobacter* is an uncommon cause of meningitis from the community, the frequency is almost 4% in cases of hospital-acquired meningitis and almost 11% among patients with meningitis involving an indwelling intraventricular catheter. Treatment is problematic because some cephalosporins do not achieve high enough levels in cerebrospinal fluid to be effective and because of the increasing incidence of MDR and XDR *Acinetobacter* (see later).

DIAGNOSIS

There are now more than 30 different species in the genus *Acinetobacter*, and their classification and identification remain problematic for clinicians (to date, 38 species are known; 27 are named and 11 are unnamed genospecies).[6] Automated and biochemical systems are sometimes inaccurate for the identification of *Acinetobacter* species. The introduction of matrix-assisted laser desorption/ionization time-of-flight into the clinical laboratory may facilitate better diagnosis.[7]

TREATMENT Rx

An increasing number of strains of *Acinetobacter* species are resistant to all antibiotics, and these strains are often responsible for outbreaks in large hospitals. These multidrug-resistant (MDR) or extremely drug-resistant (XDR) strains are often resistant to three or more classes of antibiotics. Among the resistance genes found in *Acinetobacter* species are a large collection of genes encoding β-lactamases, aminoglycoside-modifying enzymes, and many efflux pumps. Even more concerning is the emergence of XDR strains of Acinetobacter that are resistant to carbapenems and ampicillin-sulbactam, leaving few treatment options.[8] Unfortunately, resistance to "last-line" agents, such as the cationic antimicrobial peptides colistin (polymyxin E) and its close relative polymyxin B, also occurs. Regrettably, colistin resistance is becoming more common as clinicians are forced to use this agent more frequently. In hospital environments, *Acinetobacter* species can withstand drying (desiccation) and may even be transmitted by aerosol/respiratory droplet. Combined with drug resistance, these characteristics create a formidable infection control challenge.

Increasing resistance to a variety of antimicrobial agents complicates the treatment of *Acinetobacter* species infections. In general, infections due to more resistant strains of *Acinetobacter* are associated with less favorable outcomes than are infections due to more susceptible strains. These worse outcomes are likely due in part to limited treatment options as well as delays in the time to implementation of effective antimicrobial therapy for patients with infections due to MDR and XDR *Acinetobacter* strains.

The treatment doses that follow all assume normal renal function; doses need to be adjusted on the basis of the degree of renal insufficiency. When it is susceptible, *Acinetobacter* is typically treated with sulbactam (in the formulation of ampicillin-sulbactam in the United States, 3 g IV every 6 hours); the carbapenems imipenem (500 to 1000 mg IV every 6 hours) or meropenem (500 to 1000 mg IV every 8 hours); or broad-spectrum cephalosporins.[9] Based on ampicillin component, aminoglycosides are an option for treatment of urinary tract infection. One must be cautious: susceptibility to imipenem does not always translate to susceptibility to meropenem, and susceptibility to ceftazidime does not always indicate that cefepime can be used. In addition, there are data suggesting that doripenem may be the carbapenem with the most in vitro activity.

Treatment of infection with XDR *Acinetobacter* is challenging.[10] Tigecycline (100 mg loading dose, then 50 mg IV every 12 hours) and minocycline (200 mg IV or PO initially, followed by 100 mg IV or PO every 12 hours) have good in vitro activity against many strains of XDR *Acinetobacter*. Some experts recommend higher maintenance dosages of tigecycline (100 mg IV every 12 hours) and minocycline (200 mg IV or PO every 12 hours). These agents are appropriate for treatment of SSTI, but because of poor serum concentrations and lack of clinical experience in the treatment of MDR and XDR *Acinetobacter*, monotherapy with these should be avoided for invasive infections such as bacteremia and pneumonia. Sometimes, there is no choice but to use tigecycline as it might be the only available agent with in vitro activity against infecting *Acinetobacter* strains.

Invasive infections, such as bacteremia, pneumonia, and deep wound infections, are often treated with a polymyxin antimicrobial, either colistimethate sodium (CMS, often referred to as colistin) or polymyxin B. If CMS is used, a loading dose of 5 mg/kg colistin base activity, with a maximum dose of

300 mg, should be administered, followed by daily administration of 5 mg/kg colistin base activity per day in divided doses every 8 hours. If polymyxin B is used, a loading dose of 2.5 mg/kg × 1 should be administered intravenously, followed by a dose of 1.5 to 2.5 mg/kg/day by continuous IV infusion or in divided doses every 12 hours infused during a period of 60 minutes.[11]

Clinicians often treat invasive infections due to XDR *Acinetobacter* with combination therapy, although there is a lack of prospective controlled data demonstrating the superiority of combination therapy. Agents often combined with the polymyxins include rifampin (10 mg/kg/day, not to exceed 600 mg), imipenem or meropenem, sulbactam (ampicillin-sulbactam), tigecycline or minocycline, and aminoglycosides. At the time of this writing, data obtained from clinical trials conducted in Italy suggest that rifampin may not add measurable benefit to the treatment of MDR or XDR *Acinetobacter* despite the in vitro efficacy of the combination. In addition, the benefit of aerosolized colistin has not yet been established. Interest is rising in the use of the combination of colistin and tigecycline or fosfomycin, but further studies are needed.

The major adverse effect of polymyxin therapy is nephrotoxicity.[12] Current advances have improved our understanding of the pharmacokinetics, pharmacodynamics, and dosing of the polymyxins, but this is still an emerging and changing field. There is promising interest in the use of antioxidants (e.g., ascorbic acid) to prevent nephrotoxicity, but these studies are still preliminary.

In cases of meningitis due to *Acinetobacter* species, treatment should be intravenous meropenem (2000 mg every 8 hours) plus the intraventricular administration of an aminoglycoside (gentamicin or amikacin, depending on susceptibilities). Intraventricular gentamicin is administered at a dose of 4 mg once daily every 1 to 3 days until clinical and microbiologic improvement occurs. Amikacin can be used in place of gentamicin at the dose of 30 mg once a day, if required. For cases of meningitis due to XDR *Acinetobacter*, intravenous CMS or polymyxin B should be supplemented with intraventricular or intrathecal administration of CMS 10 mg/day, colistin base activity, or polymyxin B 5 to 10 mg/day for the initial 3 days of therapy and then every other day.

PREVENTION

Infection prevention practices, including hand hygiene, use of contact precautions, maintenance of environmental cleanliness, and implementation of antimicrobial stewardship, can help prevent spread of *Acinetobacter* in the hospital. During outbreaks, cohorting of patients and use of dedicated staff to care for cohorted patients might be necessary to control spread. Removal of indwelling devices from patients, including vascular catheters and endotracheal tubes, can help prevent colonization and infection with *Acinetobacter*.

PROGNOSIS

Invasive infection due to *Acinetobacter* has been associated with crude mortality rates in excess of 50% and increases in duration of ICU stay of 6 days and duration of total hospital stay of more than 14 days. With appropriate early empiric therapy, however, the mortality rate for community-acquired infection can be as low as 11%.[13] By comparison, infections due to strains of MDR and XDR *Acinetobacter* are associated with increases in mortality and durations of hospitalization compared with more susceptible *Acinetobacter* strains.[14]

STENOTROPHOMONAS MALTOPHILIA

DEFINITION

Stenotrophomonas maltophilia has attracted significant attention because it is often MDR due to intrinsic and acquired factors. *S. maltophilia* contributes significantly to morbidity, but usually not mortality, in immunocompromised patients.

The Pathogen

S. maltophilia are gram-negative bacteria that need methionine for growth and, like *Pseudomonas aeruginosa* and *Acinetobacter* species, are non–lactose fermenters. *S. maltophilia* possess flagella that are multitrichous (more than one flagellum arising from the pole) and are distinguished from *P. aeruginosa* as oxidase negative. Colonies of *S. maltophilia* may appear pale yellow or lavender-green on blood agar plates. The organism emits a mild ammonia-like odor that is used for preliminary identification.[15]

S. maltophilia can colonize inanimate surfaces in the hospital, including catheters, intravenous fluid, water supplies, and hospital equipment. This pathogen may also survive in hospital-grade disinfectant. Other health care–associated sources of *S. maltophilia* include contaminated intravenous fluids, hospital water and ice supplies, nebulizers, dialysis machines, ventilator circuits, thermometers, blood gas analyzers, intra-abdominal balloon pumps, and central venous or arterial pressure monitors.

TABLE 307-1 RISK FACTORS FOR INFECTION BY *STENOTROPHOMONAS MALTOPHILIA*

Prolonged hospitalization or intensive care unit stay
Intravascular catheters
Indwelling devices
Mechanical ventilation or tracheostomy
Neutropenia or cytotoxic chemotherapy
Solid organ transplantation
Immunocompromised state
Mucositis
Malignant disease
Chronic lung disease (especially cystic fibrosis and chronic obstructive pulmonary disease)
HIV infection
Hemodialysis
Antibiotic exposure (especially to carbapenems, extended-spectrum cephalosporins, and fluoroquinolones)
Exposure to other patients with *S. maltophilia*

EPIDEMIOLOGY

Despite its harboring many resistance genes, little is known about the virulence of *S. maltophilia*. Nevertheless, a number of risk factors are associated with infection by *S. maltophilia* (Table 307-1).[16] Although *S. maltophilia* is regarded as primarily a nosocomial pathogen (it is the third most common nonfermenting gram-negative bacillus health care–associated pathogen), community-acquired infection can occur. *S. maltophilia* infection primarily affects patients who are immunocompromised (including patients with hematologic malignant neoplasms[17] and patients who underwent solid organ transplantation), patients cared for in ICUs who are mechanically ventilated, hemodialysis patients with intravascular catheters, neonates, and patients with cystic fibrosis. Colonization or infection with *S. maltophilia* among patients with cystic fibrosis (Chapter 89) has been associated with reduction in lung function.[18]

CLINICAL MANIFESTATIONS

Respiratory Tract

The respiratory tract is the most common site of isolation of *S. maltophilia* in the hospital. Surveillance programs reveal a rate of recovery of *S. maltophilia* from hospitalized patients with pneumonia of more than 3% in the United States. Nosocomial pneumonia due to *S. maltophilia* is often associated with significant pulmonary disease such as emphysema, bronchiectasis, lung transplantation, or endobronchial obstruction. In addition, many of these patients are on mechanical ventilators, have tracheostomy tubes in place, or are receiving broad-spectrum antibiotics. Among mechanically ventilated patients, it can be challenging to differentiate colonization due to *S. maltophilia* from infection. CAP due to *S. maltophilia* has been reported but is very rare.

The incidence of respiratory tract infection due to *S. maltophilia* in patients with cystic fibrosis (Chapter 89) is increasingly being reported. There may be clear links with resistance to antipseudomonal antibiotics (tobramycin, imipenem, ceftazidime) used to treat these patients. The presence of chronic *S. maltophilia* infection has been associated with increases in cystic fibrosis exacerbations, hospitalizations, need for lung transplantation, and mortality. It remains unclear whether *S. maltophilia* has a causative association with poor outcomes or is merely a marker for severe underlying disease.

Blood Stream Infection and Endocarditis

Bacteremia may be secondary to a respiratory, urinary, or gastrointestinal source but is most commonly due to infection of an indwelling intravascular device. In many cases, except for those involving intravenous catheters, the portal of entry is not apparent. Immunocompromised patients with indwelling intravascular catheters, including those with hematologic malignant disease, are at particularly high risk for CLABSI due to *S. maltophilia*. In cases of CLABSI due to *S. maltophilia*, removal of the infected catheter is an important component of management. On occasion, an environmental reservoir or contaminated vascular access device is linked to the presence of bacteremia. Intravenous drug users are at especially high risk for

TABLE 307-2 MECHANISMS OF RESISTANCE IN *STENOTROPHOMONAS MALTOPHILIA*

DRUG	MECHANISM OF RESISTANCE
β-Lactams, including imipenem	L1 and L2 β-lactamases Outer membrane permeability/efflux
Aminoglycosides	Aminoglycoside-modifying enzymes, transport

contaminating prosthetic valves with S. *maltophilia*. In the case of endocarditis, favorable outcomes are reported with antimicrobial therapy, but surgery may also be required.

Urinary Tract Infection

Although *S. maltophilia* is frequently recovered from urine specimens in patients with indwelling urinary catheters, the role of this organism as a pathogen in this setting is unclear.

Meningitis

Cases of central nervous system infection due to *S. maltophilia* are reported rarely. These are often associated with central nervous system devices or antecedent neurosurgery.

Skin and Soft Tissue Infection

S. maltophilia can be isolated from postoperative wounds, but its role as a pathogen in this setting is unclear. In contrast, *S. maltophilia* can cause burn wound sepsis, manifested as a syndrome very similar to ecthyma gangrenosum (Chapter 441) in immunocompromised oncology patients.

TREATMENT

Rx

As a pathogen that is an increasingly important cause of nosocomial infections, *S. maltophilia* exhibits intrinsic resistance to many antibiotics. The complexity of resistance genes and mechanisms is summarized in Table 307-2. Resistance to imipenem, piperacillin-tazobactam, ceftazidime, and aminoglycosides is common. Despite significant resistance to many agents, trimethoprim-sulfamethoxazole remains the drug of choice (10 to 15 mg/kg/day IV, based on the trimethoprim component).[19] Treatment duration is uncertain, but it is usually 14 days or longer. In addition, in vitro studies indicate that ticarcillin-clavulanate (3.1 g every 6 hours), minocycline (200 mg followed by 100 mg every 12 hours, not to exceed 400 mg in 24 hours), some of the fluoroquinolones, and tigecycline may be useful agents. Polymyxin-based regimens may serve as alternatives in the face of resistance to trimethoprim-sulfamethoxazole (see earlier for details regarding polymyxin treatment). New regimens attempting to combine tigecycline and colistin for invasive *S. maltophilia* infections appear promising. Aztreonam is effective against strains exhibiting resistance to β-lactams by metallo-β-lactamases.

PROGNOSIS

Among patients with VAP or bacteremia due to *S. maltophilia*, crude mortality rates in the range of 60% have been reported. The attributable mortality of invasive *S. maltophilia* infection has been reported to be between 12 and 37.5%,[20] but patients who receive early appropriate antibiotic therapy appear to do better.[21]

PREVENTION

Standard infection control practices including hand hygiene and thorough environmental cleaning can help prevent spread of *Stenotrophomonas* in the hospital. In outbreak settings, contact precautions and possibly cohorting of patients have been used to control spread.

Grade A **Grade A Reference**

A1. Durante-Mangoni E, Signoriello G, Andini R, et al. Colistin and rifampicin compared with colistin alone for the treatment of serious infections due to extensively drug-resistant *Acinetobacter baumannii*: a multicenter, randomized clinical trial. *Clin Infect Dis*. 2013;57:349-358.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Antunes LC, Visca P, Towner KJ. *Acinetobacter baumannii*: evolution of a global pathogen. *Pathog Dis*. 2014;71:292-301.
2. Munoz-Price LS, Weinstein RA. *Acinetobacter* infection. *N Engl J Med*. 2008;358:1271-1281.
3. Guerrero DM, Perez F, Conger NG, et al. *Acinetobacter baumannii*-associated skin and soft tissue infections: recognizing a broadening spectrum of disease. *Surg Infect (Larchmt)*. 2010;11:49-57.
4. Peng C, Zong Z, Fan H. *Acinetobacter baumannii* isolates associated with community-acquired pneumonia in West China. *Clin Microbiol Infect*. 2012;18:E491-E493.
5. El-Saed A, Balkhy HH, Al-Dorzi HM, et al. *Acinetobacter* is the most common pathogen associated with late-onset and recurrent ventilator-associated pneumonia in an adult intensive care unit in Saudi Arabia. *Int J Infect Dis*. 2013;17:e696-e701.
6. Visca P, Seifert H, Towner KJ. Acinetobacter infection—an emerging threat to human health. *IUBMB Life*. 2011;63:1048-1054.
7. Espinal P, Seifert H, Dijkshoorn L, et al. Rapid and accurate identification of genomic species from the *Acinetobacter baumannii* (Ab) group by MALDI-TOF MS. *Clin Microbiol Infect*. 2012;18:1097-1103.
8. Chopra T, Marchaim D, Awali RA, et al. Epidemiology of bloodstream infections caused by *Acinetobacter baumannii* and impact of drug resistance to both carbapenems and ampicillin-sulbactam on clinical outcomes. *Antimicrob Agents Chemother*. 2013;57:6270-6275.
9. Garnacho-Montero J, Amaya-Villar R. Multiresistant *Acinetobacter baumannii* infections: epidemiology and management. *Curr Opin Infect Dis*. 2010;23:332-339.
10. Viehman JA, Nguyen MH, Doi Y. Treatment options for carbapenem-resistant and extensively drug-resistant *Acinetobacter baumannii* infections. *Drugs*. 2014;74:1315-1333.
11. Sandri AM, Landersdorfer CB, Jacob J, et al. Population pharmacokinetics of intravenous polymyxin B in critically ill patients: implications for selection of dosage regimens. *Clin Infect Dis*. 2013;57:524-531.
12. Pogue JM, Lee J, Marchaim D, et al. Incidence of and risk factors for colistin-associated nephrotoxicity in a large academic health system. *Clini Infect Dis*. 2011;53:879-884.
13. Davis JS, McMillan M, Swaminathan A, et al. A 16-year prospective study of community-onset bacteremic *Acinetobacter* pneumonia: low mortality with appropriate initial empirical antibiotic protocols. *Chest*. 2014;146:1038-1045.
14. Lemos EV, de la Hoz FP, Einarson TR, et al. Carbapenem resistance and mortality in patients with *Acinetobacter baumannii* infection: systematic review and meta-analysis. *Clin Microbiol Infect*. 2014;20:416-423.
15. Brooke JS. *Stenotrophomonas maltophilia*: an emerging global opportunistic pathogen. *Clin Microbiol Rev*. 2012;25:2-41.
16. Abbott IJ, Slavin MA, Turnidge JD, et al. *Stenotrophomonas maltophilia*: emerging disease patterns and challenges for treatment. *Expert Rev Anti Infect Ther*. 2011;9:471-488.
17. Perez F, Adachi J, Bonomo RA. Antibiotic-resistant gram-negative bacterial infections in patients with cancer. *Clin Infect Dis*. 2014;59(suppl 5):S335-S339.
18. Hansen CR. *Stenotrophomonas maltophilia*: to be or not to be a cystic fibrosis pathogen. *Curr Opin Pulm Med*. 2012;18:628-631.
19. Wang YL, Scipione MR, Dubrovskaya Y, et al. Monotherapy with fluoroquinolone or trimethoprim-sulfamethoxazole for treatment of *Stenotrophomonas maltophilia* infections. *Antimicrob Agents Chemother*. 2014;58:176-182.
20. Hotta G, Matsumura Y, Kato K, et al. Risk factors and outcomes of *Stenotrophomonas maltophilia* bacteraemia: a comparison with bacteraemia caused by *Pseudomonas aeruginosa* and *Acinetobacter* species. *PLoS ONE*. 2014;9:e112208.
21. Lakatos B, Jakopp B, Widmer A, et al. Evaluation of treatment outcomes for *Stenotrophomonas maltophilia* bacteraemia. *Infection*. 2014;42:553-558.

REVIEW QUESTIONS

1. Your patient has a blood culture turn positive. The preliminary result from the laboratory indicates possible *Acinetobacter baumannii*. Which of the following agents is most likely to have in vitro activity against *A. baumannii*?

A. Ertapenem
B. Ampicillin-sulbactam
C. Piperacillin-tazobactam
D. Trimethoprim-sulfamethoxazole
E. Doxycycline

Answer: B Increasing resistance to a variety of antimicrobial agents complicates the treatment of *Acinetobacter* species infections today. When it is susceptible, *Acinetobacter* is typically treated with sulbactam (in the formulation of ampicillin-sulbactam in the United States at a dose of 3 g intravenously every 6 hours). Other options are the carbapenems imipenem and meropenem and the broad-spectrum cephalosporins. See treatment section under *Acinetobacter* species.

2. Your patient has a blood culture turn positive. The preliminary result from the laboratory indicates possible *Stenotrophomonas maltophilia*. Which of the following agents is most likely to have in vitro activity against *S. maltophilia*?

A. Ertapenem
B. Ampicillin-sulbactam
C. Piperacillin-tazobactam
D. Trimethoprim-sulfamethoxazole
E. Meropenem

Answer: D *S. maltophilia* is an increasingly important cause of the nosocomial infections, and it exhibits intrinsic resistance to many antibiotics. Nevertheless, trimethoprim-sulfamethoxazole remains the drug of choice, at a dose of 10 to 15 mg/kg/day intravenously, based on the trimethoprim component, and it is given for 14 days or longer with positive blood cultures. See treatment section under *Stenotrophomonas maltophilia*.

3. Your patient has been admitted to the intensive care unit because of respiratory failure. He is intubated and mechanically ventilated. The intensive care unit is in the midst of an *Acinetobacter* outbreak. What would be the most effective intervention to decrease risk for development of an infection due to *Acinetobacter* in your patient?

A. Extubate the patient and remove him from mechanical ventilation as soon as possible.
B. Institute contact precautions (i.e., gowns and glove precautions).
C. Prescribe prophylaxis for the patient with intravenous imipenem.
D. Clean frequently touched objects in the patient's room with bleach.
E. Prescribe prophylaxis for the patient with *Saccharomyces* through a nasogastric tube.

Answer: A During outbreaks in the hospital, cohorting of patients and use of dedicated staff to care for the cohorted patients might be necessary to control spread. Removal of indwelling devices from the patients, including vascular catheters and endotracheal tubes, can help prevent colonization and infection with *Acinetobacter*. See prevention section under *Acinetobacter* species.

4. You are managing a patient with a deep wound infection resulting from a combat injury that occurred in Iraq. After incision and drainage, preliminary results from operative cultures indicate growth of possible *A. baumannii*. The patient has an anaphylactic reaction to β-lactam antibiotics. Which agent would be most likely to be an effective therapeutic agent for the patient?

A. Aztreonam
B. Ciprofloxacin
C. Tigecycline
D. Chloramphenicol
E. Rifampin

Answer: C Since initial reports of the outbreak of *A. baumannii* among military personnel in Iraq and Afghanistan, reports of severe wound infections and skin and soft tissue infections by this pathogen have increased in frequency. Multiple drug resistance and the presence of copathogens frequently complicate treatment. Most cases require surgical débridement and are associated with substantial mortality. Tigecycline might be the only available agent with in vitro activity against this pathogen in this setting. See the section on wound, burn, and skin and soft tissues infections under clinical manifestations of *Acinetobacter* species.

5. Your patient with cystic fibrosis has colonization of the respiratory tract with *Stenotrophomonas maltophilia*. The patient asks you what the significance of this colonization is and what can be done. Which of the following is the most accurate answer?

A. Because *Stenotrophomonas* is only a colonizer, there is no association with poor outcomes.
B. Treatment of colonization with an active antimicrobial improves the outcomes of cystic fibrosis patients.
C. *Stenotrophomonas* increases mortality risk among cystic fibrosis patients primarily by causing necrotizing pneumonia.
D. Pneumonia due to *S. maltophilia* in a patient with cystic fibrosis is an indication for lung transplantation.
E. Colonization with *S. maltophilia* among patients with cystic fibrosis has been associated with reduction in lung function.

Answer: E Colonization or infection with *S. maltophilia* among patients with cystic fibrosis has been associated with reduction in lung function. See reference 15 and the epidemiology section under *Stenotrophomonas maltophilia*.

308

SALMONELLA INFECTIONS (INCLUDING ENTERIC FEVER)

JOHN A. CRUMP

DEFINITION

A member of the family Enterobacteriaceae, *Salmonella enterica* subspecies *enterica* includes more than 2500 serovars (also called serotypes) found in humans and other warm-blooded animals. These serovars may be associated with human asymptomatic intestinal carriage, intestinal infection, and invasive disease with extraintestinal infection. Each serovar designation follows the species name (e.g., *Salmonella enterica* subspecies *enterica* serovar Typhimurium) and is frequently abbreviated as simply *Salmonella* followed by the serovar name (e.g., *Salmonella* Typhimurium).

The Pathogen

Salmonellae are gram-negative, non–spore-forming bacilli. They are differentiated from other Enterobacteriaceae by biochemical tests. They ferment glucose, maltose, and mannitol but typically not lactose or sucrose. They reduce nitrates and do not produce cytochrome oxidase. Almost all salmonellae produce acid and gas with fermentation. Most *Salmonella* serovars cannot be distinguished by biochemical reactions. However, *Salmonella* Typhi may be preliminarily identified by its production of only trace amounts of hydrogen sulfide and diminished biochemical activity compared with other serovars.

Salmonellae can be differentiated into more than 2500 serovars by their somatic (O) antigens, which are composed of lipopolysaccharides and are part of the cell wall, and by their flagellar (H) and capsular (Vi) antigens. *Salmonella* serogroups were traditionally designated by letters based on O antigens (e.g., A, B, C1, C2). More recently, the growing number of serogroups has made it necessary to move to a numeric designation. During the transition, the traditional letter-based serogroup may be retained in brackets after the numeric designation (e.g., O:2 [A], O:4 [B], O:6,7 [C1], O:8 [C2]). Some of the important serovars and their serogroups are Typhi (group O:9 [D1]), Choleraesuis (group O:7 [C1]), Typhimurium (group O:4 [B]), and Enteritidis (group O:9 [D1]). *Salmonella* Enteritidis and Typhimurium are the most common nontyphoidal serovars causing human disease.

EPIDEMIOLOGY

Salmonella Typhi, *Salmonella* Paratyphi A, *Salmonella* Paratyphi B, *Salmonella* Paratyphi C, and *Salmonella* Sendai are either solely or almost exclusively pathogens of humans; they cause primarily enteric fever rather than diarrhea, and transmission between humans is usually through water or food. As a result of modern sewage and water treatment facilities and improved food safety practices, typhoid fever and paratyphoid fever have become rare in developed countries but remain a problem in countries that lack adequate sanitation and a safe water supply. There are usually fewer than 500 cases of typhoid fever each year in the United States, mainly acquired abroad[1]; in contrast, an estimated 26.9 million cases occurred globally in 2010.[2]

Other serovars of *Salmonella* (described here as nontyphoidal *Salmonella*) have reservoirs in warm-blooded animals and cause human foodborne illness after the consumption of contaminated meat or animal products, contamination of produce or water by animal feces or animal products, or contact with animals and their environments. Some nontyphoidal *Salmonella* serovars appear frequently in particular animal species, and human illness is often associated with exposure to these animals and their products. For example, *Salmonella* Enteritidis has a reservoir in chickens, and infection is often linked to the consumption of undercooked eggs and poultry products or exposure to live chicks. Such a relationship is less clear for some other nontyphoidal serovars (e.g., *Salmonella* Typhimurium). Foodborne nontyphoidal *Salmonella* was estimated to be associated with approximately 1.0 million domestically acquired illnesses and 378 deaths in the United States in 2006.[3] A disproportionate number of infections occur in July through October, probably related to warm weather. *Salmonella* infections are most common among infants and children younger than 5 years. Worldwide, nontyphoidal

Salmonella was estimated to cause 93.8 million diarrheal illnesses, 80.3 million food-borne illnesses, and 155,000 deaths in 2005.[4]

Antimicrobial Resistance

Salmonellae have become increasingly resistant to antimicrobial agents, usually by acquiring resistance transfer factors (e.g., plasmid mediated). It is thought that antimicrobial resistance in the human-restricted salmonellae (e.g., *Salmonella* Typhi) is driven primarily by antimicrobial use in humans, whereas antimicrobial resistance among the nontyphoidal salmonellae (e.g., *Salmonella* Typhimurium) is associated with the use of antimicrobial agents in farm animals. Among *Salmonella* Typhi isolated in the United States during 1999 to 2006, 13% were resistant to the traditional first-line antimicrobials ampicillin, chloramphenicol, and trimethoprim-sulfamethoxazole, and 38% showed decreased fluoroquinolone susceptibility. Resistance to extended-spectrum cephalosporins is rare in *Salmonella* Typhi and *Salmonella* Paratyphi. Among nontyphoidal *Salmonella* isolated from humans in the United States, resistance to three or more antimicrobial classes is common but declining, whereas resistance to ceftriaxone and nonsusceptibility to ciprofloxacin are uncommon but increasing.[5] Susceptibility breakpoints and interpretive criteria were established for azithromycin in 2015 in response to reports of *Salmonella* with elevated azithromycin minimum inhibitory concentrations.

PATHOBIOLOGY

Etiology

Salmonellae are transmitted by the ingestion of fecally contaminated food or water; contact with animals, their environments, and other fomites; and, rarely, close contact with infected persons (e.g., oral-anal intercourse). The ultimate sources of contamination are humans or animals that are acutely ill or are asymptomatic carriers.

Contaminated Animal Products

Salmonella infection in humans usually occurs from ingestion of contaminated animal food products, most often eggs, poultry, and meat. *Salmonella* Choleraesuis is associated with pig products, *Salmonella* Dublin with cattle and unpasteurized milk from cattle, and *Salmonella* Enteritidis with poultry and poultry products, including eggs. Fecal material on poultry and other animal carcasses can spread at slaughterhouses, such as when many poultry carcasses are placed in the same hot-water tank to remove feathers. *Salmonella* contaminating carcasses can multiply to high levels if meat or other animal products are not refrigerated. Human illness may result if such animal products are inadequately cooked or if utensils or other uncooked foods are cross-contaminated during preparation. A wide range of foods can be contaminated with animal or human feces, from production on the farm through consumption in the home. Reports of produce-associated *Salmonella* outbreaks, due to contamination by animal or human feces during production, are increasing.[6] *Salmonella* outbreaks have occurred from contaminated cheese, ice cream, vegetables, fruit, juice, and alfalfa sprouts.

Contaminated Food and Water

Contamination by Pets

Salmonella infections may be acquired after contamination of food or water with the feces of pet turtles, chicks, ducks, birds, dogs, cats, and many other species.

Contamination by Humans

Salmonella infection can also be acquired by eating food or by drinking water contaminated by human carriers who have not adequately washed their hands. Infection has been spread by the fecal-oral route among children, by contaminated enema and fiberoptic instruments, by diagnostic and therapeutic preparations made from animal or insect products (e.g., pancreatic extract, carmine dye), and from intentional or unintentional contamination of restaurant salad bars. Outbreaks of salmonellosis may occur in institutionalized patients, who are probably more prone to the development of *Salmonella* infections for three reasons. First, within institutionalized populations, there is an increased prevalence of underlying diseases that decrease host defense mechanisms against salmonellae, such as disorders of gastric acidity and intestinal motility; second, the use of antimicrobial agents modifies the normal, protective intestinal flora; and third, institutional food prepared in bulk may be more likely to be contaminated than are individually prepared meals. Outbreaks in nurseries and among the elderly in nursing homes are associated with the highest case-fatality ratios (>5%).

Contact with Animals and Their Environments

Both healthy and sick animals may harbor and shed *Salmonella*. Transmission of *Salmonella* from animals and their environment to humans occurs primarily by the fecal-oral route. Animal hides and saliva often harbor fecal organisms, and transmission can occur when persons pet, touch, feed, or are licked by animals. Transmission has also been associated with contaminated animal bedding, flooring, barriers, other environmental surfaces, and clothing and shoes. Contact with calves, turtles and other reptiles, rodents, and young poultry and their environments has been associated with *Salmonella* outbreaks. Humans may also become infected when animals come into contact with their food or water. Infections can be prevented by education, supervision of animal contact, provision and promotion of handwashing facilities, and separation of food handling and consumption from animal areas.[7]

Contact with Infected Persons

Close contact with persons shedding *Salmonella* is an occasional source of infection. Transmission has been documented among persons handling feces (e.g., parents changing the diapers of an infected infant) and is associated with certain sexual practices (e.g., oral-anal intercourse).

Pathophysiology

After the ingestion of organisms, the likelihood for development of infection, as well as the severity of infection, is related to the dose, the virulence of the *Salmonella* strain, and the status of host defense mechanisms. Usually at least 10^2 to 10^3 bacteria are required to produce clinical infection in a normal host. Higher inocula are associated with increased disease severity, whereas smaller inocula are more likely to result in transient intestinal carriage. Gastric acid serves as a host defense mechanism by killing many of the ingested organisms, and intestinal motility is probably another host defense mechanism. In the absence of or with a decrease in gastric acidity (e.g., in infants and the elderly; after gastrectomy, vagotomy, or gastroenterostomy; or with the use of drugs that reduce gastric acidity) and with decreased intestinal motility (e.g., the use of antimotility drugs), much smaller inocula can produce infection, and the infection tends to be more severe.

Administration of antimicrobial agents before the ingestion of salmonellae can markedly reduce the size of the inoculum needed to produce infection, presumably by reducing the concentration of protective bowel flora.

Although any *Salmonella* serovar can produce any of the *Salmonella* syndromes (transient asymptomatic carrier state, enterocolitis, enteric fever, bacteremia, and chronic carrier state), each serovar tends to be associated with certain syndromes much more frequently than with others. For example, *Salmonella* Anatum usually causes asymptomatic intestinal infection, whereas *Salmonella* Typhimurium generally causes enterocolitis. *Salmonella* Choleraesuis is more likely to produce bacteremia (often with metastatic infection) than asymptomatic infection or enterocolitis, and some serovars such as *Salmonella* Typhi and *Salmonella* Paratyphi are most likely to cause enteric fever as well as the chronic carrier state. Fortunately, most *Salmonella* serovars are of relatively low pathogenicity for humans. Therefore, although food products are commonly contaminated, large outbreaks tend to involve the more virulent serovars.

To produce infection, invasion must occur across the mucosa of the intestine. When the organisms reach the lamina propria, an influx of polymorphonuclear leukocytes serves as a host defense mechanism to prevent invasion of the lymphatics. Certain serovars seem to have a greater ability than others to invade the lymphatics and subsequently to produce bacteremia (e.g., *Salmonella* Choleraesuis and *Salmonella* Dublin, which commonly produce bacteremia after intestinal infection). Both the small intestine and the colon are involved in the inflammatory process.

In the case of *Salmonella* Typhi and other causes of enteric fever, salmonellae invade the mononuclear phagocytes in Peyer's patches in the ileum and mesenteric lymph nodes. Some intracellular salmonellae form a nonreplicating population of "persisters" that could provide a reservoir for relapsing infection; intracellular persistence is determined by conditions in the vacuolar environment of the infected cells.[8] Others multiply intracellularly and are carried through the lymphatic system and blood stream to the liver, spleen, bone marrow, and other parts of the reticuloendothelial system. Once in the reticuloendothelial system, they multiply intracellularly in mononuclear phagocytes and produce the systemic manifestations of enteric fever. The onset of fever is associated with bacteremia and the release of cytokines (e.g., tumor necrosis factor and interleukins) from mononuclear phagocytes.

Ulcerations over Peyer patches are responsible for the intestinal manifestations of enteric fever, such as pain, perforation, and bleeding.

In *Salmonella* enterocolitis, the organisms remain localized in the intestinal mucosa, and diarrhea results from the inflammation produced by polymorphonuclear leukocytes. In addition, watery diarrhea may occur, apparently the result of the secretion of water and electrolytes by small intestinal epithelial cells in response to an enterotoxin secreted by some of the *Salmonella* strains or in response to tissue mediators of inflammation.

Patients with diseases that impair host defense mechanisms seem to have an increased frequency of severe *Salmonella* infection. A striking association has been recognized between diseases producing hemolysis and *Salmonella* bacteremia. Specifically, *Salmonella* bacteremia is common in patients with sickle cell disorders, malaria, and bartonellosis. In fact, because of the frequency of *Salmonella* bacteremia in sickle cell diseases and the underlying bone disease in these patients to which salmonellae localize, these organisms are the most common cause of osteomyelitis in patients with sickle cell disorders (Chapter 163). Prolonged *Salmonella* bacteremia may occur in patients with hepatosplenic schistosomiasis, possibly related to localization on and in the intravascular schistosomes. Patients with lymphoma and leukemia are also more prone to the development of *Salmonella* bacteremia. A markedly increased frequency and severity of *Salmonella* infections in general have been observed in patients with human immunodeficiency virus (HIV) infection, particularly those with $CD4^+$ T-lymphocyte counts less than 200 cells/mm^3. Prolonged and recurrent, refractory *Salmonella* bacteremia is common among these patients. Other risk factors that increase the frequency and severity of *Salmonella* infection are extremes of age, immunocompromised states (e.g., from immunosuppressive agents), malnutrition, and probably diabetes. Nontyphoidal *Salmonella* serovars are a leading cause of community-acquired blood stream infection in sub-Saharan Africa, where children younger than 3 years and HIV-infected adults carry most of the burden of invasive disease.[9]

CLINICAL MANIFESTATIONS

Asymptomatic Intestinal Carrier State

The asymptomatic intestinal carrier state may result from inapparent infection (which is the most common form of *Salmonella* infection), or it may follow clinical disease (in which case the patient becomes a convalescent carrier). The carrier state is usually self-limited to several weeks to months, with the prevalence of positive stool cultures rapidly decreasing over time. By 1 year, far less than 1% of carriers still have positive stools. The main exception is *Salmonella* Typhi; about 3% of those infected excrete the organism for life. Women and older men are most likely to become chronic carriers of *Salmonella* Typhi, related to the presence of biliary tract disease, especially calculi. A patient who has had salmonellae in stool for 1 year (chronic carrier) is likely to become a lifelong carrier; the reservoir is in the biliary tree, usually in calculi in the gallbladder. Patients with *Schistosoma haematobium* infection are predisposed to become chronic urinary carriers of salmonellae.

Enterocolitis

After an incubation period of usually 12 to 48 hours, the illness starts suddenly with crampy abdominal pain and diarrhea. Nausea and vomiting may occur but are usually not prominent or persistent. The diarrhea may be watery and of large or small volume. Stools may contain mucus and are occasionally bloody. Polymorphonuclear leukocytes are present in the stool. Diarrhea may be mild or severe, with up to 20 to 30 stools a day. Fever is present in most patients, and the temperature may reach 40°C (104°F) or higher. The abdomen is tender to palpation. Transient bacteremia may occur and is most commonly seen in infants, the elderly, and patients with impaired host defense mechanisms.

Symptoms generally improve during a period of days, with fever lasting no more than 2 to 3 days and diarrhea lasting no more than 5 to 7 days. However, these symptoms occasionally persist for up to 14 days. More severe disease is seen with malnutrition, inflammatory bowel disease, and HIV infection. Reactive arthritis may follow enterocolitis in up to 7% of cases, especially among those with the HLA-B27 phenotype.

Enteric Fever

Enteric fever is produced by *Salmonella* Typhi (typhoid fever), *Salmonella* Paratyphi A, B, and C (paratyphoid fever), and occasionally other serovars. Sometimes it immediately follows classic enterocolitis caused by the same organism. The syndrome is characterized by prolonged, sustained fever and may be associated with relative bradycardia, splenomegaly, rose spots, and leukopenia. In Africa, the common symptoms of invasive nontyphoidal *Salmonella* disease, which is seen predominantly in patients with HIV infection, malaria, and malnutrition, are fever, hepatosplenomegaly, and respiratory symptoms; features of enterocolitis are often absent.[10]

Therapy aborts the course of the disease. The following is a description of untreated illness. After an incubation period of 5 to 21 days (generally 7 to 14 days), fever and malaise develop, often associated with cough. A small proportion of patients may have diarrhea during the incubation period. The fever tends to rise in a stepwise fashion during the first few days to a week and then becomes sustained, usually at 39.4° to 40.0°C (103° to 104°F) or higher. Relative bradycardia is seen in up to half of patients. Apathy, confusion, delirium, and even psychosis may occur. Abdominal distention, pain, and tenderness may occur in the first week and may be associated with diarrhea or constipation; these symptoms are generally more pronounced during the second week of fever. Most patients have abdominal tenderness during the course of the illness.

In about 30% of patients, rose spots develop on the abdomen or chest (or both) toward the end of the first week or during the second week of fever. These faint, salmon-colored maculopapular lesions are subtle and may be difficult to see, particularly in dark-skinned patients. Salmonellae can be cultured from punch biopsies of these lesions. Hepatosplenomegaly occurs in about half of patients. Leukopenia and neutropenia are seen in about 20%. Abnormal liver function test results are common.

After 2 weeks of illness, the severe complications of intestinal hemorrhage and perforation related to necrosis of Peyer patches may be observed in about 5% of patients. These perforations may require surgical as well as medical therapy and can occur even in a patient treated with antimicrobials. Intestinal perforation is the leading cause of death from enteric fever.

The illness usually resolves by the end of the fourth week in an untreated patient. Relapse may occur in untreated as well as in treated patients, but the illness is milder than the original episode.

Rarely, some of the following complications may occur: pancreatitis, cholecystitis, infective endocarditis, meningitis, pneumonia, hepatic or splenic abscess, orchitis, or focal infection at virtually any site.

Bacteremia

Patients with *Salmonella* bacteremia usually complain of fever and chills lasting days to weeks. Gastrointestinal symptoms are unusual, but in some patients *Salmonella* bacteremia follows classic enterocolitis. Other symptoms are nonspecific, such as malaise, anorexia, and weight loss. Metastatic infection of bones, joints, aneurysms (particularly of the abdominal aorta), meninges (mainly in infants), pericardium, pleural space, lungs, heart valves, cysts, uterine myomas, malignant neoplasms, and other sites is common, and symptoms may be related to the site of metastatic infection. Stool cultures are often negative for salmonellae, but blood cultures are positive.

Although any *Salmonella* serovar can produce bacteremia, *Salmonella* Dublin, *Salmonella* Choleraesuis, *Salmonella* Heidelberg, *Salmonella* Oranienburg, *Salmonella* Panama, and *Salmonella* Sandiego are associated with increased likelihood of bacteremia.

Salmonella bacteremia occurs with increased frequency in infants, the elderly, and patients with diseases associated with hemolysis (e.g., sickle cell diseases, malaria, bartonellosis), HIV infection, lymphoma, leukemia, disseminated histoplasmosis, and perhaps systemic lupus erythematosus. Localization to bone is common in patients with sickle cell diseases (Chapter 163).

Prolonged *Salmonella* bacteremia lasting for months may occur in patients with hepatosplenic schistosomiasis. In patients with HIV infection, recurrent, relapsing *Salmonella* bacteremia may develop, which may be difficult to cure with antimicrobial agents.

DIAGNOSIS

Although *Salmonella* enterocolitis is an invasive disease, the differential diagnosis includes all causes of acute diarrhea, including invasive bacteria such as *Campylobacter jejuni, Shigella* species, invasive *Escherichia coli, Yersinia enterocolitica,* and *Vibrio parahaemolyticus;* toxigenic bacteria such as *Vibrio cholerae,* enterotoxigenic *E. coli,* enterohemorrhagic *E. coli* (e.g., *E. coli* O157:H7), *Staphylococcus aureus, Bacillus cereus, Clostridium perfringens,* and *Clostridium difficile;* viruses; and protozoa such as *Entamoeba histolytica, Giardia intestinalis,* and *Cryptosporidium* species. The invasive bacterial causes of diarrhea, enterohemorrhagic *E. coli* and *C. difficile* infection, are also associated with polymorphonuclear leukocytes in stool, whereas bacterial toxigenic causes (other than *C. difficile* and enterohemorrhagic *E. coli*), viruses, and protozoa

generally are not. The bacterial toxigenic causes of diarrhea other than *C. difficile* and enterohemorrhagic *E. coli* do not produce fever.

Stool culture is definitive for the diagnosis of *Salmonella* enterocolitis, but by the time the results of stool culture are available, most patients are recovering. A stained smear of the stool usually demonstrates polymorphonuclear leukocytes. Serologic studies are of little clinical value in *Salmonella* enterocolitis, but they may be of use in epidemiologic studies.

The differential diagnosis of *Salmonella* bacteremia includes virtually all acute infectious and noninfectious causes of fever, including bacteremia caused by other organisms. The diagnosis is proved by isolation of the microorganism from blood or from another normally sterile site.

The differential diagnosis of enteric fever is broad and depends in part on the area of the world where the infection was acquired. All causes of sustained fever are in the differential diagnosis, including infective endocarditis, disseminated tuberculosis, brucellosis, tularemia, *Mycoplasma pneumoniae* infection, rickettsial infections, Q fever, and viral infections such as infectious mononucleosis. Depending on the site of acquisition, diseases such as malaria, amebic abscesses of the liver, and visceral leishmaniasis also enter into the differential diagnosis.

The diagnosis of enteric fever is best proved by isolation of the microorganism from blood, stool, or bone marrow.[11] During the first week of illness, blood cultures are positive in about 90% of patients, but culture positivity decreases in the next 2 weeks to less than 50% during the third week of illness. Stool cultures are usually negative during the first week but are generally positive by the third week. Bone marrow cultures give the highest yield, with up to 95% being positive; they should be considered in suspected cases with negative blood cultures. Bone marrow cultures may be positive even after several days of antimicrobial treatment, when blood cultures have become negative. Urine cultures and cultures of punch biopsies of rose spots may also be positive. The string test to obtain samples of bile from the duodenum has also yielded positive cultures.

The peripheral leukocyte count is usually normal, but leukopenia, which occurs in about 20% of cases, may be suggestive of enteric fever. Fecal leukocytes are generally present.

The Widal and other serologic tests that detect serum antibodies against *Salmonella* Typhi are limited by shortcomings of both sensitivity and specificity and rarely provide useful information to guide management of the patient. Polymerase chain reaction and other molecular techniques lack sensitivity for diagnosis from blood and other specimens, but they have been used to determine the *Salmonella* serovar of bacterial isolates.

TREATMENT Rx

Enterocolitis

The primary approach to the treatment of *Salmonella* enterocolitis is fluid and electrolyte replacement. Drugs with antiperistaltic effects, such as loperamide or diphenoxylate with atropine, can relieve cramps, but they should be used sparingly because they can prolong the diarrhea.

Salmonella enterocolitis is self-limited, and antimicrobial therapy is usually not indicated, except perhaps in groups of patients at high risk for invasive disease. Antimicrobial therapy reportedly has little effect on the clinical course, and in some studies, it has prolonged the duration of *Salmonella* excretion in stool. In addition, most patients are improving by the time salmonellae or other bacterial pathogens are isolated from stool.

The fluoroquinolones are active against virtually all bacterial pathogens that cause diarrhea (including salmonellae), except for *C. difficile* and many *Campylobacter* organisms. Thus, it is reasonable to use fluoroquinolones for patients with suspected or known *Salmonella* enterocolitis who are severely ill and suspected of being bacteremic. The threshold for antimicrobial treatment is also decreased in those at increased risk for severe illness (e.g., infants, the elderly, patients with sickle cell disease, immunosuppressed individuals). As an example, in adults, ciprofloxacin, 500 mg every 12 hours orally or 400 mg every 12 hours intravenously for 3 to 5 days, or until defervescence, has been widely used. An extended-spectrum cephalosporin such as ceftriaxone is an alternative. In the presence of gross bloody diarrhea, antimicrobial therapy should be withheld until the possibility of *E. coli* O157:H7 infection has been eliminated because antimicrobial therapy may increase the frequency of hemolytic-uremic syndrome.

Other agents, such as amoxicillin and trimethoprim-sulfamethoxazole, have also been widely used in severely ill adults. However, many strains of *Salmonella* are now resistant to these agents.

Enteric Fever

Resistance to the traditional first-line antimicrobial agents (ampicillin, chloramphenicol, trimethoprim-sulfamethoxazole) has emerged worldwide among the salmonellae causing enteric fever. Consequently, alternative antimicrobial agents are now preferred.

The fluoroquinolones have become the agents of choice for the treatment of enteric fever, for several reasons. They can be administered orally and have high bioavailability, they concentrate in bile and the bowel, and they often retain activity against multidrug-resistant strains of *Salmonella* Typhi and other causes of enteric fever. Most important, the fluoroquinolones have proved to be effective in the treatment of enteric fever, even with short courses (e.g., 3 to 7 days). The proportion of patients cured exceeds 95%, and relapse and chronic fecal carriage after therapy are uncommon. Ciprofloxacin (500 mg orally twice a day) for 7 to 14 days has been the fluoroquinolone of choice for enteric fever. If a patient cannot tolerate oral therapy, the fluoroquinolones can be administered intravenously. Reduced susceptibility and resistance to fluoroquinolones are increasingly reported in *Salmonella* Typhi and *Salmonella* Paratyphi strains both in the United States and elsewhere and are associated with treatment failure. If decreased fluoroquinolone susceptibility or resistance is suspected or demonstrated, alternative agents include extended-spectrum cephalosporins (e.g., intravenous ceftriaxone) and azithromycin.

Extended-spectrum cephalosporins such as ceftriaxone are reliable agents for the treatment of enteric fever. Ceftriaxone dosed at 1 to 2 g every 12 to 24 hours for adults and 75 mg/kg/day for children, given intravenously or intramuscularly for 10 to 14 days, results in cure of 95% of patients. Resistance to ceftriaxone has been described in clinical strains of *Salmonella* Typhi, but this occurs rarely.

If the *Salmonella* isolate is shown to be susceptible by antimicrobial susceptibility testing, ampicillin, chloramphenicol, or trimethoprim-sulfamethoxazole may be considered. The ampicillin dose is 25 mg/kg intravenously every 6 hours. The use of chloramphenicol should be weighed against the risk of bone marrow toxicity. The chloramphenicol dose is 50 mg/kg/day orally, divided into four doses. Chloramphenicol can be given intravenously at the same dose if oral therapy is not possible. Trimethoprim-sulfamethoxazole (4/20 mg/kg intravenously or orally every 12 hours) is given for 10 to 14 days.

Azithromycin (10 mg/kg/day orally for 7 days) is effective in the treatment of patients with uncomplicated typhoid fever caused by multidrug-resistant strains. The oral route of administration makes it a particularly attractive choice in settings where multidrug resistance is common and intravenous extended-spectrum cephalosporins are impractical, unavailable, or too expensive.

Patients often require supportive care with intravenous saline, correction of electrolyte and acid-base disturbances, and, in the setting of intestinal bleeding, blood transfusion. If perforation is suspected, abdominal imaging should be performed to evaluate for free air. If perforation seems likely, laparotomy should be performed as soon as possible to repair the perforation. In the setting of perforation, antimicrobial therapy should be broadened to cover bowel flora.

Steroid therapy is beneficial in some patients with severe enteric fever and coma, delirium, or shock. Dexamethasone is administered at doses of 3 mg/kg initially, followed by 1 mg/kg every 6 hours for 48 hours. Steroids can mask the signs and symptoms of abdominal perforation and should not be continued for more than 48 hours. Salicylates should be avoided.

Relapses of typhoid fever may be treated with the same antimicrobial regimen as the initial attack.

Bacteremia

The agents of choice to treat *Salmonella* bacteremia are the fluoroquinolones, such as ciprofloxacin, and the extended-spectrum cephalosporins, such as ceftriaxone. Typical doses are ciprofloxacin 400 mg every 12 hours intravenously and ceftriaxone 1 to 2 g every 12 to 24 hours intravenously. When the salmonellae are known to be susceptible, ampicillin 1 to 2 g intravenously every 4 to 6 hours or trimethoprim-sulfamethoxazole 8 mg/kg/day (of the trimethoprim component) intravenously can be used. Chloramphenicol is another option. Antimicrobial susceptibility testing is necessary because of the emergence of infections resistant to the fluoroquinolones or extended-spectrum cephalosporins.

In cases of sustained bacteremia, the possibility of endovascular infection should be investigated. For transient bacteremia or bacteremia without localization, therapy is continued for 7 to 14 days. With localization to bone, aneurysms, heart valves, and various other sites, antimicrobial therapy should be given for much longer periods (e.g., 6 weeks). Surgical drainage, removal of foreign bodies, or resection of an aneurysm is often necessary to cure localized infection. The possibility of schistosomiasis should be considered and treated, when present, in patients with sustained *Salmonella* bacteremia (Chapter 355). Patients with HIV infection tend to experience repeated relapses after treatment courses for *Salmonella* bacteremia. In this group, initial treatment with ciprofloxacin for 2 weeks or longer is recommended. Long-term suppressive therapy has been suggested for those experiencing frequent relapses.[12]

Carriers

Chronic carriers (i.e., >1 year) of salmonellae other than *Salmonella* Typhi are rare. Stools of convalescent carriers spontaneously become negative

during a period of weeks to months, and no therapy should be given. The rare chronic carrier of *Salmonella* serovars other than Typhi (usually infected with *Salmonella* Paratyphi A, B, or C) may be treated with a fluoroquinolone, amoxicillin, or trimethoprim-sulfamethoxazole in the doses listed later for 4 to 6 weeks. Patients who experience relapse usually have gallbladder disease (most often calculi) and will not be cured with antimicrobial therapy alone. Cholecystectomy plus antimicrobial therapy may cure these patients, but it is doubtful that the carrier state is a sufficient indication for cholecystectomy.

Chronic fecal carriers of *Salmonella* Typhi can be treated with ciprofloxacin (500 to 750 mg twice daily) for 6 weeks or with amoxicillin at doses of 6 g/day in three or four divided doses plus probenecid 2 g/day in divided doses for 6 weeks. Trimethoprim-sulfamethoxazole (160/800 mg twice daily) plus rifampicin (300 mg twice daily) for 6 weeks may be considered as an alternative regimen. Patients with persistent urinary carriage and *S. haematobium* infection should be treated with praziquantel before eradication of *Salmonella* Typhi is attempted. For patients with persistent carriage and anatomic abnormalities (e.g., gallstones), cholecystectomy combined with antimicrobial therapy is often necessary. For patients with persistent carriage despite adequate antimicrobial therapy and without an identifiable anatomic abnormality, chronic suppressive therapy may be considered. Chronic carriers who do not prepare food and who practice adequate personal hygiene usually do not constitute a public health hazard. Therefore, after the institution of appropriate personal hygienic precautions, and in the absence of evidence of a chronic carrier infecting others, cholecystectomy is probably not indicated to eradicate the carrier state.

PREVENTION

Salmonella infection is best prevented by protecting the water supply, preventing fecal contamination during food production, cooking and refrigerating foods, pasteurizing milk and milk products, and handwashing before preparing foods. Travelers should judiciously avoid consuming untreated water (including ice), raw vegetables, and fruits. Food should be cooked or peeled, and drinks should be boiled, carbonated, or commercially bottled. The widespread presence of salmonellae in the animal kingdom means that reducing the risk for *Salmonella* infections requires a multifaceted approach.

There is no vaccine for *Salmonella* infection other than that for *Salmonella* Typhi. Travelers should be vaccinated before going to areas that are endemic for typhoid fever.[13] Two vaccines are available in the United States. One is the typhoid Vi capsular polysaccharide vaccine, which is administered as a single intramuscular injection, with booster doses given every 2 years if needed. This vaccine provides a degree of herd protection against typhoid fever when it is used at the population level. The other licensed typhoid fever vaccine is the oral live attenuated Ty21a vaccine. Revaccination is necessary every 5 years, if indicated. Ty21a vaccine should not be used in immunocompromised persons or those receiving antimicrobials. Both these vaccines confer greater than 75% protective efficacy. Efforts are under way to develop typhoid vaccines that are effective in young children.

Vaccines afford only partial immunity to typhoid fever. Persons who have been vaccinated should still restrict their diets to avoid potentially contaminated food and fluids. When cases of imported typhoid are identified in the United States, the local health department should be informed and will monitor stool cultures. Typhoid fever acquired in the United States is typically investigated by the public health department to identify potential sources and chronic carriers.

PROGNOSIS

Mortality in patients with *Salmonella* enterocolitis is rare; infants and the elderly are at greatest risk, with death occurring as a result of dehydration and electrolyte imbalance. Mortality from *Salmonella* bacteremia is not uncommon and is most likely to occur in the very young, the very old, the malnourished, and the immunocompromised.

Before the advent of antimicrobial therapy, typhoid fever had a case-fatality ratio of 15 to 20%. This has been reduced to less than 1% in industrialized countries. However, the case-fatality ratio remains high in some developing countries. The case-fatality ratio of invasive nontyphoidal *Salmonella* in Africa is approximately 20%. In treated patients, the temperature usually returns to normal after 3 to 5 days of therapy, but this may take longer in patients treated with extended-spectrum cephalosporins than in those treated with fluoroquinolones and in those infected with isolates with decreased fluoroquinolone susceptibility who are treated with ciprofloxacin.

In the era before antimicrobial therapy, 5 to 10% of patients who recovered from typhoid fever had relapses. Relapses continued to occur in 10 to 15% of patients treated with chloramphenicol, ampicillin, and trimethoprim-sulfamethoxazole, but this seemed to be much less frequent (<5%) among those treated with ceftriaxone and fluoroquinolones. Intestinal bleeding or perforation occurs in about 5% of patients. With perforation, case-fatality ratios of 10 to 30% have been reported. Up to 3% of patients recovering from *Salmonella* Typhi infection become chronic fecal carriers.

Grade A References

A1. Effa EE, Lassi ZS, Critchley JA, et al. Fluoroquinolones for treating typhoid and paratyphoid fever (enteric fever). *Cochrane Database Syst Rev*. 2011;10:CD004530.

A2. Effa EE, Bukirwa H. Azithromycin for treating uncomplicated typhoid and paratyphoid fever (enteric fever). *Cochrane Database Syst Rev*. 2008;4:CD006083.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Imanishi M, Newton AE, Vieira AR, et al. Typhoid fever acquired in the United States, 1999-2010: epidemiology, microbiology, and use of a space-time scan statistic for outbreak detection. *Epidemiol Infect*. 2014; [Epub ahead of print].
2. Buckle GC, Walker CL, Black RE. Typhoid fever and paratyphoid fever; systematic review to estimate morbidity and mortality for 2010. *J Glob Health*. 2012;2:010401.
3. Scallan E, Hoekstra RM, Angulo FJ, et al. Foodborne illness acquired in the United States: major pathogens. *Emerg Infect Dis*. 2011;17:7-15.
4. Majowicz SE, Musto J, Scallan E, et al. The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clin Infect Dis*. 2010;50:882-889.
5. Medalla F, Hoekstra RM, Whichard JM, et al. Increase in resistance to ceftriaxone and nonsusceptibility to ciprofloxacin and decrease in multidrug resistance among *Salmonella* strains, United States, 1996-2009. *Foodborne Pathog Dis*. 2013;10:302-309.
6. Painter JA, Hoekstra RM, Ayers T, et al. Attribution of foodborne illnesses, hospitalizations, and deaths to food commodities by using outbreak data, United States, 1998-2008. *Emerg Infect Dis*. 2013;19:407-415.
7. Centers for Disease Control and Prevention. Compendium of measures to prevent disease associated with animals in public settings, 2011. *MMWR Morb Mortal Wkly Rep*. 2011;60:1-24.
8. Helaine S, Cheverton AM, Watson KG, et al. Internalization of *Salmonella* by macrophages induces formation of nonreplicating persisters. *Science*. 2014;343:204-208.
9. Reddy EA, Shaw AV, Crump JA. Community acquired bloodstream infections in Africa: a systematic review and meta-analysis. *Lancet Infect Dis*. 2010;10:417-432.
10. Feasey NA, Dougan G, Kingsley RA, et al. Invasive non-typhoidal *Salmonella* disease: an emerging and neglected tropical disease in Africa. *Lancet*. 2012;379:2489-2499.
11. Wain J, Hendriksen RS, Mikoleit ML, et al. Typhoid fever. *Lancet*. 2014;[Epub ahead of print].
12. Panel on Opportunistic Infections in HIV-Infected Adults and Adolescents. *Guidelines for the prevention and treatment of opportunistic infections in HIV-infected adults and adolescents: recommendations from the Centers for Disease Control and Prevention, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America*. Available at: http://aidsinfo.nih.gov/contentfiles/lvguidelines/adult_oi.pdf. Accessed February 9, 2015.
13. Wagner KS, Freedman JL, Andrews NJ, et al. Effectiveness of the typhoid Vi vaccine in overseas travelers from England. *J Travel Med*. 2014;[Epub ahead of print].

REVIEW QUESTIONS

1. Regarding antimicrobial resistance among *Salmonella* Typhi isolated in the United States, which of the following is *true*?

A. Ceftriaxone resistance occurs in the majority of strains.
B. Decreased fluoroquinolone susceptibility is seen in more than a third of isolates.
C. Multiple drug resistance is not observed.
D. Gentamicin is an appropriate choice for treatment of enteric fever.
E. The clinical microbiology laboratory will be unable to provide susceptibility results for azithromycin.

Answer: B Among U.S. *Salmonella* Typhi isolated during 1999 to 2006, less than 1% were resistant to ceftriaxone, 38% showed decreased fluoroquinolone susceptibility, and 13% were multdrug resistant (resistant to the traditional first-line antimicrobials ampicillin, chloramphenicol, and trimethoprim-sulfamethoxazole). For *Salmonella*, aminoglycosides may appear active *in vitro* but are not effective clinically. The Clinical Laboratory and Standards Institute (CLSI) published breakpoints and interpretive criteria for azithromycin and *Salmonella* for the first time in 2015.

2. Which of the following statements about nontyphoidal *Salmonella* is *false*?

A. Typhimurium and Enteritidis are the most common nontyphoidal *Salmonella* serovars isolated in most high-income countries.
B. Chickens are common reservoirs of *Salmonella* Enteritidis.
C. *Salmonella* Dublin has a great propensity to cause blood stream infection.
D. *Salmonella* Typhimurium is a human host–restricted serovar.
E. Produce is an increasingly important source of nontyphoidal *Salmonella* infection.

Answer: D *Salmonella* Typhimurium and *Salmonella* Enteritidis are the most commonly isolated nontyphoidal *Salmonella* from the stool of humans with diarrhea. Chickens are major reservoirs of *Salmonella* Enteritidis. *Salmonella* Dublin is commonly isolated from blood. *Salmonella* Typhi and *Salmonella* Paratyphi, but not *Salmonella* Typhimurium, are human host–restricted serovars. Uncooked produce eaten fresh is an increasingly important source of nontyphoidal *Salmonella*. This may be related to fecal contamination of produce by animals.

3. A previously healthy 23-year-old male chef develops diarrhea. Stool culture grows *Salmonella* Typhimurium but the diarrhea persists. Which of the following is the most appropriate recommendation?

A. Treat with ciprofloxacin.
B. Prescribe loperamide.
C. Encourage fluid and electrolyte replacement.
D. Suggest that he continue work if he has no fever.
E. Administer vaccine for typhoid fever.

Answer: C Nontyphoidal *Salmonella* diarrhea is usually self-limited in healthy, immunocompetent persons who are not at the extremes of age. Antimicrobials prolong *Salmonella* shedding. Drugs with antiperistaltic effects can prolong diarrhea and should be used sparingly. Individuals with diarrhea should be excluded from food handling. Typhoid vaccine does not protect against *Salmonella* Typhimurium infection.

4. A 36-year-old woman develops fever and abdominal pain after a trip to a typhoid-endemic country. Which of the following recommendations is most appropriate?

A. Order a Widal test.
B. Prescribe ampicillin for empirical treatment of typhoid fever.
C. Collect blood culture specimens.
D. Arrange for bone marrow culture.
E. Administer dexamethasone.

Answer: C Blood culture represents the conventional standard diagnostic test for typhoid fever. Bone marrow culture is more sensitive than blood culture, but it is not often used as the first test. The Widal test and other serologic approaches to typhoid fever perform poorly. The high prevalence of ampicillin resistance in *Salmonella* Typhi and *Salmonella* Paratyphi make it a poor choice for empirical treatment. Any role for dexamethasone for treatment of typhoid fever is restricted to carefully selected patients with severe disease.

5. A 45-year-old woman with a history of blood culture–confirmed typhoid fever continues to have stool cultures that grow *Salmonella* Typhi 18 months after her initial illness despite remaining well. Which of the following recommendations is *not* appropriate?

A. Do nothing because the patient feels well.
B. Treat her with antimicrobials for *Salmonella* Typhi chronic carriage.
C. Encourage handwashing.
D. Exclude the patient from food handling.
E. Determine whether she has gallbladder disease.

Answer: A This patient is a *Salmonella* Typhi chronic carrier because her stool culture is positive more than 12 months after the initial infection. Although she remains well, she continues to shed *Salmonella* Typhi in her stool, posing a risk for transmission to others. It is appropriate to treat her with antimicrobials suitable for the elimination of *Salmonella* Typhi carriage and to promote handwashing. She should be excluded from food preparation until she is demonstrated to no longer be a carrier. Gallbladder disease is a risk factor for the development of the chronic carrier state, although it is doubtful that the carrier state is a sufficient indication for cholecystectomy.

309

SHIGELLOSIS

GERALD T. KEUSCH

DEFINITION

Shigellosis is an acute infection of the large bowel due to bacteria of the genus *Shigella*, characterized by mucosal inflammation and fever. Clinical disease ranges from watery diarrhea to bloody diarrhea or dysentery, a syndrome consisting of multiple small-volume bloody stools per day, abdominal cramping, and tenesmus, a painful straining with the urge to defecate.

The Pathogen

The etiologic agents of shigellosis are gram-negative bacilli of the family Enterobacteriaceae, tribe Escherichieae, and genus *Shigella.* The organism is so closely related to *Escherichia coli* (Chapter 304) that if it were discovered today, *Shigella* would be classified as distinct serotypes of *E. coli*. Indeed, a number of *E. coli* serotypes causing *Shigella*-like illness and possessing conserved virulence factors are now well described.

EPIDEMIOLOGY

In the United States, the incidence of microbiologically confirmed cases has been steadily trending down during the past four decades on the basis of data compiled from voluntary reporting by state health departments to the Centers for Disease Control and Prevention (CDC) (Fig. 309-1). In 2011, the latest year for which data are available, 7062 cases were identified, corresponding to an all-time low incidence rate of 2.3 per 100,000.[1] However, this is not uniform across the country. For example, the incidence rate of *Shigella sonnei* infection, which accounts for more than 75% of all isolates reported to the CDC, was considerably higher in the seven southern states from Alabama across to Arizona. In addition, the incidence rate was nearly 11 per 100,000 in children 0 to 4 years old and more than 8 per 100,000 in children 5 to 9 years old, reflecting the fact that about one third of cases occur in these age groups.[2] Pediatric shigellosis in the United States is commonly associated with daycare centers, where hygiene is difficult to maintain. However, most cases of acute shigellosis are never reported because they are mild or not part of an outbreak and therefore never microbiologically investigated. For this reason, the CDC estimates that around 450,000 cases of *Shigella* infection occur each year in the United States, primarily self-limited watery diarrheas. The small inoculum, from just 10 to 10,000 organisms, documented to cause infection and illness in experimental human infections explains why shigellosis is so readily transmitted from person to person from the stool of an infected person to a susceptible individual, often by the hands and direct skin contact or indirectly through objects (fomites) previously handled by the infected person. It also explains why one case in a family is associated with transmission to 20 to 40% of household members, although children are much more likely to develop clinical illness than are adults, who may be immune as a

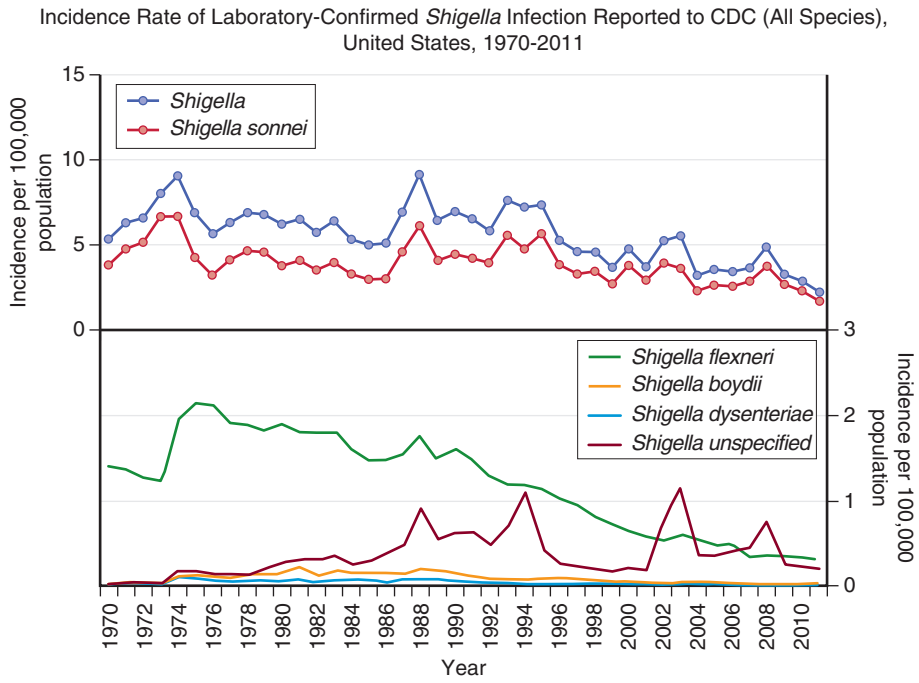

FIGURE 309-1. Incidence rate of laboratory-confirmed *Shigella* infection reported to the Centers for Disease Control and Prevention (all species), United States, 1970-2011. ***Top panel,*** **The incidence rates of infection with *Shigella* (all species) and *Shigella sonnei*. Since 1970, the incidence rate of infection with *Shigella* (all species) has been driven by the incidence of infection with *Shigella sonnei*. *Bottom panel,* The incidence rate of infection with all *Shigella* species other than *Shigella sonnei,* including infections with an unspecified species. The incidence rate of infection with *Shigella flexneri* has been decreasing since the 1980s. Since the mid-1980s, the incidence rate of *Shigella* infection in which the species is not indentified has fluctuated, likely representing, at least to some extent, outbreak situations in which public health laboratories did not characterize all outbreak-associated *Shigella* isolates to the species level. *Shigella boydii* and *Shigella dysenteriae* infections are rare in the United States.**

result of prior contact with the organism.[3] Asymptomatic carriage of *Shigella* is typically limited to a few weeks; but because an etiologic diagnosis of *S. sonnei* is rarely made, except in common-source outbreaks most likely to be investigated by public health officials, convalescent carriers often return to their normal activities while still able to transmit infection. This may account, in part, for the predilection for outbreaks in daycare centers.

Fecal contamination of food or water is another route for transmission. In settings where there are no facilities for sanitary disposal of feces, flies can serve to transfer the organism to food or water. In the United States, foodborne shigellosis results in multiperson or multistate outbreaks.[4] In 2012, the CDC Foodborne Diseases Active Surveillance Network (FoodNet) identified *Shigella* in 2128 individuals, the third leading cause, of whom 23% were hospitalized and two died. The overall incidence rate was 4.5 per 100,000 at risk in the sentinel populations, but it was four times higher in children 0 to 9 years old. These events are generally related to an infected food handler and continually seem to involve new vehicles, such as recent outbreaks traced to fresh salsa or guacamole prepared in restaurants. Multicountry outbreaks of *S. sonnei* due to contaminated fresh foods shipped from Africa or Southeast Asia demonstrate how globalization of the food supply can also globalize infection with particular strains of *Shigella*. This has important implications for the importation of antibiotic-resistant strains from areas of the world where the prevalence is particularly high. In the United States, contamination of swimming pools or bathing areas, usually by young children harboring the organism who defecate in the water, can result in common-source outbreaks. Because of the small infectious inoculum, outbreaks readily occur during complex humanitarian emergencies such as floods, landslides, earthquakes, or the gathering of refugees of conflict into crowded camps with limited sanitary facilities.

Shigella dysenteriae type 1 and *Shigella flexneri* generally cause more severe illness, resulting in bloody diarrhea or dysentery, whereas *S. sonnei* and *Shigella boydii* are typically mild infections causing self-limited watery diarrhea.[5] For still unknown reasons, *S. sonnei* is most common in high-income countries; most *Shigella* infections in developing countries are due to *S. flexneri* strains,[6] with *S. dysenteriae* type 1 periodically causing outbreaks that may last several years. *S. boydii* is present primarily in the Indian subcontinent and clinically resembles *S. sonnei*. Wherever there is severe underlying childhood malnutrition, however, any *Shigella* infection can be lethal.

In the United States, the incidence peaks in summer and fall. Females are more often infected than are males in the age range of 5 to 29 years, whereas males predominate in the 30- to 49-year age group, possibly related to anal-oral sex among men who have sex with men. Among the latter, *S. flexneri* is more common than *S. sonnei*, the reverse ratio observed in the general population. Thus, the median age of individuals with *S. sonnei* infection is 6 years, in contrast to a median age of 26 years in those harboring *S. flexneri*. In developing countries, seasonal peaks occur in the rainy season, when organisms from open defecation can be washed into the water used for drinking, or in the dry season, when water for personal hygiene is scarce. In contrast to the low case-fatality rate for *Shigella* in the United States (approximately 50/100,000 infected in the FoodNet survey), the rate in developing countries may be as high as 100-fold greater, despite the continuing decline in global diarrhea mortality during the past three decades. However, it is difficult to determine how many of the currently estimated 700,000 annual global diarrheal disease deaths are due to *Shigella* because bloody diarrhea or dysentery is not separately reported and cultures are not routinely performed, but it is likely to be a substantial proportion.[7] A careful microbiologic study of moderate to severe diarrhea in nearly 10,000 patients 0 to 5 years of age in seven sites in Africa and South Asia revealed *Shigella* as the fourth leading cause in the age group younger than 1 year, the second leading cause in the age group 1 to 2 years, and the leading cause in the age group 3 to 5 years.[8]

PATHOBIOLOGY

When *Shigella* exit an infected host in feces, they are resistant to acid pH. This facilitates their ability to survive gastric acid when ingested and accounts in large part for the small inoculum required. Once past the stomach, the organisms turn off genes governing acid resistance and turn on genes that allow them to invade the host's colon. *Shigella* initially gain entry into M cells overlying lymphoid Peyer's patches; they are then transcytosed to the lamina propria, where they are ingested by macrophages and induce inflammatory cytokines, leading to macrophage death and the recruitment and migration of polymorphonuclear leukocytes through the mucosa.[9] This forces open the tight junctions between epithelial cells and allows many more organisms from the lumen to gain access to the basolateral membrane of colonic epithelial cells, where they induce the host cells to ingest them by a mechanism resembling phagocytosis. *Shigella* then lyse the phagocytic vesicle to enter the cytoplasm and multiply; although they are nonmotile, they use their capacity to cross-link actin in the cytoplasm to propel themselves to the host cell membrane. There, by the induced phagocytosis-like mechanism, they pass laterally from one epithelial cell to another. The inflammation and death

of contiguously invaded host epithelial cells result in mucosal ulcerations and further exudation of blood and leukocytes into the colonic lumen and ultimately in stool. At the same time, regulatory pathways are engaged to transmit alarm signals that activate the innate immune system in bystander cells to confine the invading organisms to the localized area of involvement. Mucosal inflammation and ulceration are the characteristic pathologic changes caused by *Shigella*. The difference between infection by one species and that by another is the severity of the process, which is proportional to the severity of the illness they cause.

The attributes of virulence are present in a large virulence plasmid present in all *Shigella* and in enteroinvasive *E. coli* serotypes capable of causing a similar clinical illness. The virulence plasmid encodes the proteins for a type III secretion system that is capable of transferring microbial effector proteins into the host cell; these proteins regulate the cytoskeleton of the host cell to ingest the organism. Once inside the epithelial cell, shigellae are sequestered from host immune responses. Whereas the short-term gain for the organism is microbial multiplication and excretion in feces, the same mechanisms subsequently act to eliminate the organisms. For example, at later stages of infection, neutrophils play a major role in disease resolution by killing the bacteria and degrading some *Shigella* virulence proteins. *S. dysenteriae* type 1 also produces a chromosomally encoded toxin, Shiga toxin, that inhibits protein synthesis and directly causes cell death. It is one important reason that infection due to this species is so severe.

Structurally and antigenically similar toxins are produced by certain *E. coli*, notably serotype O157:H7, responsible for cases of hemorrhagic colitis (Chapter 304). All Shiga toxin–producing organisms are known causes of hemolytic-uremic syndrome (HUS; Chapter 172), a potentially lethal complication due to the toxin's effects on vascular endothelium within the kidneys.

CLINICAL MANIFESTATIONS

Infections due to *S. sonnei* or *S. boydii* typically are manifested after an incubation period of 1 to 3 days with fever and mild to moderate watery diarrhea (Table 309-1). They do not cause serious dehydration and generally resolve without treatment in 3 to 5 days. However, if the stool is examined under the microscope, white and red blood cells are found, indicative of the underlying inflammation of the mucosa. When the inflammation is more severe, as with *S. flexneri*, the watery diarrhea turns bloody and may progress to dysentery, with its characteristic small-volume bloody mucoid stool passed many (10 to >40) times per day, abdominal cramps, and tenesmus. The most severe infections are due to *S. dysenteriae* type 1 and progress rapidly from watery to bloody diarrhea and frequently to frank dysentery.

The variety of consequences and complications of *S. flexneri* and *S. dysenteriae* type 1 contributes to their importance as causes of death in children in developing countries. Bacteremia with the infecting strain or other enteric flora occurs in up to 10% of cases of severe dysentery due to *S. dysenteriae* type 1 and in some patients with *S. flexneri* infection. Intense mucosal inflammation can result in toxic megacolon, associated in many patients with leukemoid reactions and HUS. Colonic perforation and pancolitis are rare complications that may require surgical intervention. In resource-limited settings, where colostomy management may be difficult to ensure, conservative management with fluids and antibiotics or limited surgery to oversew perforations is the prudent choice; however, mortality may still be as high as 50%. In young children, in whom mesenteric support for the rectosigmoid colon is not fully developed, the intense proctitis that results in straining to pass stool may cause rectal prolapse. A rapid rise in temperature can lead to a seizure in this age group and is distinguishable from typical febrile seizures by the older age of the child and the rarity of multiple seizures. However, patients may become obtunded or even comatose, usually with *S. dysenteriae* type 1 or *S. flexneri* infections. This is often associated with hypoglycemia, due to poor food intake and inadequate gluconeogenesis, or hyponatremia, secondary to inappropriate secretion of antidiuretic hormone. Anorexia may be intense and prolonged. Combined with ongoing catabolism of host muscle protein associated with excessive pro-inflammatory cytokine production and protein-losing enteropathy due to colitis, this always results in some degree of protein-energy malnutrition (Chapter 215). Even with adequate protein-containing and energy-dense diets after infection, replacement of nutrient stores may take as long as four times the duration of the clinical illness. This is more of a problem in developing countries where the diet is poor and full nutritional recovery may not be possible before the next infection occurs, causing further nutritional deterioration. Because of its prolonged metabolic effects, shigellosis is often the precipitating cause of severe and often fatal protein-energy malnutrition or kwashiorkor. *S. flexneri* infection is also associated with reactive arthritis and other autoimmune inflammatory manifestations, such as tendinitis, conjunctivitis, uveitis, urethritis, or erythema nodosum, a constellation of findings commonly termed Reiter's syndrome (Chapter 265). This occurs primarily in individuals positive for HLA-B27 antigen and may be a consequence of molecular mimicry.

Routine laboratory studies document leukocytosis with many immature "band" forms, especially in patients with bloody diarrhea or dysentery. In as many as 20% of hospitalized patients with *S. dysenteriae* type 1 infection and 1 to 2% with *S. flexneri* infection, this may progress to a leukemoid reaction with thrombocytopenia and microangiopathic hemolytic anemia. HUS (Chapter 172) occurs in approximately 8% of patients with *S. dysenteriae* type 1 infection, usually preceded by a leukemoid reaction, and results in some degree of acute renal failure that might progress and require dialysis.

DIAGNOSIS

Shigellosis presenting with watery diarrhea is clinically indistinguishable from the many other causes of watery diarrhea, except for the frequency of fever in *Shigella*, unless the stool is examined microscopically to demonstrate the presence of red and white blood cells or infection is confirmed microbiologically. Bloody diarrhea or dysentery in patients in the developed world is most commonly due to a Shiga toxin–producing *E. coli* (STEC), such as serotype O157:H7, or *S. flexneri* and less often *Campylobacter jejuni*, nontyphoidal *Salmonella* species, or rarely *Yersinia enterocolitica* or *Entamoeba histolytica*. In patients who have recently received antibiotics, bloody diarrhea is often the consequence of *Clostridium difficile* (Chapter 296). In developing countries, *Shigella* (*S. flexneri* or *S. dysenteriae* type 1) is the most common cause of bloody diarrhea or frank dysentery. Although more sophisticated and more sensitive methods of diagnosis are available, such as quantitative

TABLE 309-1 CLINICAL SYNDROMES AND COMPLICATIONS OF SHIGELLOSIS

STAGE	TIME OF APPEARANCE AFTER ONSET OF ILLNESS	SYMPTOMS AND SIGNS	PATHOLOGY AND PATHOGENESIS
Prodrome	Earliest findings	Fever, chills, myalgias, anorexia	None or early colitis
Watery diarrhea	0-3 days	Fever, abdominal cramps, loose stools	Colitis with fecal leukocytes and erythrocytes
Bloody diarrhea	1-3 days	Frequent stools containing blood and mucus, abdominal cramps and tenderness, fever, anorexia	Colitis with fecal leukocytes and erythrocytes
Dysentery	1-5 days	Frequent small-volume stools consisting of blood, mucus, and pus; severe abdominal cramps; tenesmus	More extensive colitis with crypt abscesses and mucosal ulcerations
Acute complications	3-7 days	Seizures, obtundation, bacteremia, colonic obstruction, mucosal perforation, peritonitis	Severe colitis, terminal ileitis
Additional acute complications due to infection with *Shigella dysenteriae* type 1	3-7 days	Toxic megacolon, leukemoid reaction, hemolytic-uremic syndrome	Severe colitis, expression of Stx toxin
Postinfectious syndromes	1-3 weeks	Reactive arthritis, with or without urethritis and conjunctivitis	Autoimmune inflammatory response, most common in individuals expressing HLA-B27 antigen

polymerase chain reaction (which may double the positivity rate), the standard method remains microbiologic culture. For a genus capable of causing such severe illness, *Shigella* species are remarkably fragile outside the intestine and will rapidly die unless samples (preferentially stool and not rectal swabs) are processed rapidly, ideally at the bedside. Using more than one selective culture medium (e.g., MacConkey and xylose-lysine-deoxycholate or *Salmonella-Shigella* agar) and obtaining multiple culture specimens will increase the yield. Initially, laboratory screening is based on the failure of *Shigella* species to ferment lactose. Subculture of lactose-negative colonies to differential media such as triple sugar–iron confirms this property and reveals the ability to ferment glucose anaerobically, whereas the lack of motility or production of hydrogen sulfide distinguishes *Shigella* from *Salmonella* species. Specific diagnosis can then be made by standard serologic methods for *Shigella* O antigens. In the case of mild watery diarrhea due to *S. sonnei*, the patient is often recovered by the time the laboratory results are available, so they are of little use in guiding treatment. Culture yields are maximized when patients with bloody diarrhea or dysentery are studied. Typically in the United States, it is only these patients who are likely to be cultured, and the results are valuable for selection of empirical treatment and for epidemiologic purposes when antibiotic sensitivity is also determined to track the emergence of antibiotic resistance.

TREATMENT

Because of the typically mild nature of *S. sonnei* infection in well-nourished children, cultures are not performed, the diagnosis remains unproven, and no specific treatment is given or required. Although it might be sensible from a public health perspective to eliminate convalescent carriers with antibiotics to prevent subsequent transmission (e.g., in the setting of daycare centers), it cannot be justified because of the predilection for *Shigella* to acquire antibiotic resistance, resulting in the loss of valuable therapeutics to treat more severe illness. Hence, antibiotics should be reserved for such episodes, even in outbreak settings. Some authorities have suggested that culture-positive patients—particularly food handlers, caretakers in daycare or clinical care settings, teachers or other school employees, and the immunocompromised—should be kept at home until three consecutive stool cultures are negative; however, the question of who pays means this recommendation is unlikely to be followed in most instances.

In the case of bloody diarrhea or dysentery, antibiotics are the mainstay of treatment. This is particularly problematic in developing countries, where many of the inexpensive, safe, and useful oral drugs, such as ampicillin and trimethoprim-sulfamethoxazole, have become ineffective because of drug resistance. Presently, a fluoroquinolone, such as ciprofloxacin, or azithromycin is reliable in most patients, but these drugs are expensive (particularly azithromycin), and resistance in some isolates has been documented, especially in clusters of men who have sex with men with *S. sonnei* infection. Parenteral ceftriaxone has been useful in severe or drug-resistant illness. However, intravenous treatment requires hospitalization and monitoring and seriously increases the cost of treatment in both developed and developing countries. Isolates of *S. sonnei* expressing the CTX-M type of β-lactamase that confers resistance to all β-lactam antibiotics except cephamycins and carbapenems are increasingly being identified in Asia and elsewhere.

In severe shigellosis, the prompt initiation of effective antibiotic treatment shortens the duration of illness and achieves a more rapid improvement in symptoms such as fever, cramps, and tenesmus. Evidence from Bangladesh in patients with *S. dysenteriae* type 1 infection also demonstrates that prompt treatment reduces the frequency of HUS and presumably other complications due to colonic inflammation, such as megacolon, perforation, and rectal prolapse. In the United States and other developed countries where *S. dysenteriae* type 1 is not present, HUS is almost exclusively associated with STEC. The toxin genes in these *E. coli* are present on plasmids, in contrast to the chromosomal location of the *stx* gene in *S. dysenteriae* type 1, and are actually upregulated by commonly used antibiotics. Hence, most experts recommend that antibiotic treatment be withheld in patients with STEC infections; but if it is necessary, a drug should be selected that does not have this effect, such as azithromycin, rifaximin, or intravenous meropenem.

Given the known epidemiology of antibiotic resistance among *Shigella* species, empirical therapy for adults with bloody diarrhea or dysentery can be initiated with ciprofloxacin 500 mg orally twice daily for 3 to 5 days.[10] There is increasing evidence that shorter courses are sufficient, even for known or suspected *S. dysenteriae* type 1 infections, but given its severity, many experts would still recommend a full 5-day course. A still reliable oral alternative is azithromycin, with an initial dose of 500 mg followed by 250 mg/day for an additional 4 days, although the identification of *S. sonnei* with reduced susceptibility to azithromycin during an outbreak in Los Angeles in 2012 is of concern. Depending on the epidemiology of resistance in specific geographic areas, ampicillin (500 mg orally four times daily for 3 days) or trimethoprim-sulfamethoxazole (160/800 mg combination twice daily for 3 days) may still be effective. Doses for young children are ciprofloxacin 10 mg/kg twice daily for 3 days; azithromycin 10 mg/kg on day 1, followed by 5 mg/kg for 4 days more; ampicillin 25 mg/kg four times daily for 5 days; and a pediatric syrup of trimethoprim-sulfamethoxazole containing 4/20 mg, respectively, twice daily for 5 days. β-Lactam antibiotics, such as amoxicillin and oral cephalosporins, are not recommended because initial studies have shown that they are less effective in vivo than predicted by in vitro drug sensitivity studies. This may be due to efficient absorption, leaving insufficient amounts in the intestinal lumen. Pivamdinocillin (mecillinam; marketed as Coactin) is effective but is not available in the United States. Nonabsorbable antimicrobial agents are not recommended because optimal treatment requires therapeutic levels of the drugs in both the lumen and the mucosal compartments.

In more severe infections, complications such as hypoglycemia or hyponatremia can be managed with appropriate glucose or saline given intravenously, but patients must be monitored by trained clinical staff. Colitis with toxic megacolon and intestinal perforation represent difficult problems, especially in resource-poor settings where these complications are most likely to occur. In such settings, conservative medical or surgical management is best, even if it is not optimal. Renal failure due to HUS can generally be managed conservatively or with peritoneal dialysis, if it is available. Seizures are usually self-limited, and other neurologic complications respond to fluid and electrolyte management and correction of hypoglycemia and hyponatremia. Reactive arthritis is a greater problem as it is an autoimmune response that occurs primarily in genetically susceptible individuals positive for HLA-B27 antigen. In addition to effective antibiotic treatment to eliminate the causative organism, chronic arthritis may ensue, requiring nonsteroidal anti-inflammatory agents, steroids, or inflammatory cytokine inhibitors. In the susceptible host, infection-induced reactive arthritis can result in troublesome destructive joint disease that requires ongoing medical management. Some evidence suggests that antimotility agents are contraindicated in shigellosis, particularly in cases of bloody diarrhea or dysentery in young children, because these drugs may slow peristalsis and prolong contact between the organisms and the mucosa, thereby increasing microbial invasion, pathologic changes, and severity, including toxic megacolon.

Because of the prolonged anorexia and catabolic responses in shigellosis, attention to nutritional rehabilitation, especially in malnourished children in developing countries, is an important component of early and continuing management.

PREVENTION

Personal hygiene (in particular handwashing after handling the diapers of infected children and before food preparation), the sanitary disposal of feces, and protection of food and water sources from microbial contamination are essential to limit the spread of shigellosis. Preventing spread in daycare settings is a particular problem because it is so difficult to stop young children from constantly exploring their world, picking up bacteria on their hands, and bringing their potentially contaminated fingers or fomites to their mouths. In such settings, it is particularly important for the adult caretakers to observe good personal hygiene and to supervise children in handwashing. Soap and water are sufficient, but hand sanitizers do work and may be more convenient. In environments where soap is unavailable, water used with sand or ash for scrubbing is helpful. Keeping infected children away from daycare until their stool is negative, if indeed it was cultured, or separating recently ill children from the susceptible has been recommended but is not easy to accomplish. Household hygiene in the setting of an index case, including frequent handwashing, caution in the disposal of soiled diapers or underwear, regularly wiping down the area where these are collected with a disinfectant such as Lysol, and precautions in food preparation, can help limit intrahousehold spread.

Vaccines, particularly for the more virulent species, would be useful. However, despite much effort, a safe and effective vaccine has not been developed or approved for general use. It is doubtful whether a vaccine for *S. sonnei* infection would be recommended for children in a developed country such as the United States, even if it is safe and effective, because the illness is so mild and self-limited and because other vaccines are more important to administer in the context of an increasingly antivaccine public. Once it is available, it could be useful for travelers or among military who are deployed to high-risk locations. In contrast, effective and safe vaccines for *S. flexneri* or *S. dysenteriae* type 1 would be extremely useful in developing countries, where shigellosis accounts for a significant portion of the annual mortality due to diarrheal diseases.[11,12] Our current understanding of immunity in shigellosis indicates that serotype-specific immunity is most important; however,

whether this is best conveyed by antibody (and, if so, immunoglobulin G or A) or by cell-mediated mechanisms remains uncertain. Vaccine development has focused on attenuated live oral vaccines; unfortunately, vaccines most effective in inducing immunity have also been the most reactogenic, causing fever and often diarrhea in recipients. Recent efforts employ killed whole cells, O antigen–protein conjugates, and subunit candidates. Unless common protective antigens are identified, the most likely vaccine strategy will be based on a combination of serotype-specific antigens from the most common *S. flexneri* strains and *S. dysenteriae* type 1. We are a long way from having an approved vaccine for shigellosis.

PROGNOSIS

S. sonnei infection and most cases of *S. boydii* infection are mild and self-limited, with no sequelae. Infection with *S. flexneri* or *S. dysenteriae* responds to proper treatment with antibiotics. When treatment is not effective, either a drug-resistant strain or another etiologic agent should be suspected. In some instances, severe shigellosis with pancolitis has been misdiagnosed as inflammatory bowel disease. This can be a disaster if the patient is treated with steroids. Whereas early and effective treatment of *S. dysenteriae* type 1 reduces the risk of complications such as HUS and probably megacolon and bowel perforation, about 25% of those with HUS will have some permanent renal impairment, and a small percentage may progress to end-stage renal failure. Because of the autoimmune nature of reactive arthritis, usually associated with *S. flexneri*, early treatment might not prevent its occurrence but could mitigate its severity. Infection with one serotype of *Shigella* generally provides durable immunity to reinfection with the same strain but leaves the individual susceptible to other antigenically distinct strains and serotypes.

Grade A Reference

A1. Christopher PRH, David KV, John SM, et al. Antibiotic therapy for *Shigella* dysentery. *Cochrane Database Syst Rev*. 2010;8:CD006784.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Centers for Disease Control and Prevention (CDC). *National Shigella Surveillance Annual Report, 2011*. Atlanta, GA: US Department of Health and Human Services; 2013.
2. Scallan E, Mahon BE, Hoekstra RM, et al. Estimates of illnesses, hospitalizations and deaths caused by major bacterial enteric pathogens in young children in the United States. *Pediatr Infect Dis J*. 2013;32:217-221.
3. Boveé L, Whelan J, Sonder GF, et al. Risk factors for secondary transmission of *Shigella* infection within households: implications for current prevention policy. *BMC Infect Dis*. 2012;12:347.
4. Barton Behravesh C, Jones TF, Vugia DJ, et al. FoodNet Working Group. Deaths associated with bacterial pathogens transmitted commonly through food: foodborne diseases active surveillance network (FoodNet), 1996-2005. *J Infect Dis*. 2011;204:263-267.
5. Shakoor S, Zaidi AK, Hasan R. Tropical bacterial gastrointestinal infections. *Infect Dis Clin North Am*. 2012;26:437-453.
6. Lima IF, Havt A, Lima AA. Update on molecular epidemiology of Shigella infection. *Curr Opin Gastroenterol*. 2015;31:30-37.
7. Walker CL, Rudan I, Liu L, et al. Global burden of childhood pneumonia and diarrhoea. *Lancet*. 2013;381:1408-1416.
8. Kotloff KL, Nataro JP, Blackwelder WC, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, Gems): a prospective, case-control study. *Lancet*. 2013;82:209-222.
9. Marteyn BS, Gazi AD, Sansonetti PH. *Shigella*: a model of virulence regulation in vivo. *Gut Microbes*. 2014;3:104-120.
10. Klontz KC, Singh N. Treatment of drug-resistant Shigella infections. *Expert Rev Anti Infect Ther*. 2015;13:69-80.
11. Van de Verg LL, Venkatesan MM. A *Shigella* vaccine against prevalent stereotypes. *Clin Infect Dis*. 2014;59:942-943.
12. Kaminski RW, Wu M, Turbyfill KR, et al. Development and preclinical evaluation of a trivalent, formalin-inactivated Shigella whole-cell vaccine. *Clin Vaccine Immunol*. 2014;21:366-382.

REVIEW QUESTIONS

1. Spread of shigellosis can be difficult to control because the major route of transmission of infection is

A. Food
B. Water
C. Person to person
D. Aerosol
E. Dirty toilet seats

Answer: C Because *Shigella* resist the antibacterial effects of gastric acid and invade intestinal epithelial cells, which protect them from immune mechanisms, the infectious inoculum can be very small, which eliminates the need to multiply before ingestion and facilitates direct (skin-to-skin) or indirect (by contaminated fomites) person-to-person spread. It is true that the organism can be transferred to food or water and subsequently cause outbreaks; however, transmission by these routes can be prevented by commonsense care around food preparation and chlorination of the water supply. *Shigella* are not respiratory pathogens and so aerosol transmission does not occur. Whereas transmission by dirty toilet seats is possible, acquisition of infection by this route has not been described.

2. The most common presentation of shigellosis in the United States is

A. Watery diarrhea associated with fever
B. Dysentery
C. Hemolytic-uremic syndrome
D. Leukemoid reaction
E. Reactive arthritis

Answer: A The most common cause of shigellosis in the United States is *Shigella sonnei*, which only rarely proceeds beyond the watery diarrhea stage to bloody diarrhea or the dysentery syndrome. However, even *S. sonnei* induces an inflammatory enteritis with the production of pyrogenic cytokines such as interleukins 1, 6, and 8, and therefore elevated temperature is present in the majority of infected individuals. Some observations from experimentally infected adults suggest that the peak fever correlates with the severity of other clinical manifestations of shigellosis. The other options are all potential consequences of *Shigella* infection but largely found in cases due to either *Shigella flexneri* or *Shigella dysenteriae* type 1. *S. flexneri* in the United States is primarily found in cases involving men who have sex with men, who therefore often present with more severe presentations of shigellosis. *S. dysenteriae* type 1 is rarely found in the United States, and when such cases are identified, the infection has typically been acquired abroad.

3. The pathobiology of *Shigella* infection is a consequence of

A. Activation of intestinal secretion pathways
B. Lactose malabsorption
C. Hyperperistalsis
D. Mucosal edema
E. Mucosal invasion and inflammation

Answer: E A hallmark of pathogenic shigellae is their ability to invade cells, including phagocytic leukocytes and nonphagocytic intestinal epithelial cells, by inducing a process analogous to phagocytosis. The release of proinflammatory cytokines during this process results in fever, recruitment of leukocytes to the intestinal mucosa, epithelial cell death and mucosal ulceration, and the characteristic finding of white and red blood cells in the stool. When the genes mediating these processes are deleted or experimental infections are treated with cytokine inhibitors, the severity of the typical clinical manifestations is attenuated.

Copious watery diarrhea, typically associated with the activation of specific ion secretion pathways in intestinal mucosa and characteristic of *Vibrio cholerae* infection, is not characteristic of shigellosis. The watery diarrhea induced by *S. sonnei*, for example, is not sufficient to cause serious dehydration, except in unusual cases. Lactose malabsorption is a consequence of damage to intestinal epithelial brush border membranes, where this disaccharidase is localized, and therefore may develop during and after clinical shigellosis, but it is not a cause of altered physiology underlying the infection. Inflammation of the colonic mucosa is common in shigellosis, manifested as cramps or, in severe cases, with bloody diarrhea or dysentery and tenesmus, but again, this is a consequence of the underlying inflammatory reaction. Rather than mucosal edema resulting from the myriad inflammatory pathways being activated, the histology of the affected intestine reveals leukocytic infiltration, blood, organisms in the submucosa and within epithelial cells, and mucosal ulcerations where dead epithelial cells have been sloughed off the mucosal surface.

4. The primary treatment of bloody diarrhea due to *Shigella* species is

A. Oral rehydration fluids
B. Intravenous rehydration fluids
C. H_2 blockers
D. Antibiotics such as ciprofloxacin (Cipro)
E. Loperamide (Lomotil)

Answer: D Mild to moderate shigellosis presenting as watery diarrhea, and typically due to *S. sonnei*, is a self-limited illness, usually lasting only a couple of days and uncommonly the cause of significant dehydration requiring supervised rehydration by either the oral or the intravenous route. More severe clinical shigellosis, associated with bloody diarrhea, is uncomfortable because of the high frequency of bowel movements and associated cramps, and there is the potential for subsequent complications, depending on the etiologic agent involved. Antibiotic treatment with an agent to which the infecting organism is susceptible is well documented to resolve clinical symptoms and, if it is administered early, to prevent late complications of shigellosis. The problem has been to administer a proven effective antibiotic to which the causative organism is susceptible because of the acquisition of multidrug resistance genes over time. At present, a quinolone such as ciprofloxacin (Cipro) or a macrolide such as azithromycin (Zithromax) is the most reliable choice as the prevalence of resistance is still low in most parts of the world, although it is not unknown and is slowly increasing.

Rehydration, by either oral rehydration solutions or intravenously administered fluids, is usually not necessary as the degree of dehydration in shigellosis is mild or is readily corrected if it is more significant. However, rehydration will not reverse the clinical manifestations due to inflammation and in this sense is a secondary consideration at best. H_2 blockers are known to increase the susceptibility to acid-sensitive enteric pathogens such as *V. cholerae* but not acid-resistant organisms like *Shigella* and have no effect on already established infections. Lomotil or other gut motility–inhibiting drugs such as Imodium that reduce peristalsis are, in fact, potentially dangerous in that they can prolong the contact between invasive *Shigella* and the mucosa and worsen the clinical severity of the infection. Where *S. flexneri* or *S. dysenteriae* type 1 is common, these drugs are in fact considered to be contraindicated, especially in children. In some studies, the combined use of ciprofloxacin and loperamide reduces the duration of symptoms or the number of stools passed compared with treatment with ciprofloxacin alone.

5. The best available way to prevent shigellosis is

A. Immunization with *Shigella* vaccine
B. High standards of personal and household hygiene
C. Provision of prophylactic antibiotics to contacts of patients
D. Quarantine
E. Anti-inflammatory drugs

Answer: B Shigellosis is primarily contracted directly through fecal-oral transmission by the hands of an infected individual or indirectly by contaminated fomites to a susceptible individual. The way to break the cycle of transmission is to wash hands, especially when taking care of a patient with shigellosis (e.g., after handling soiled diapers or other clothes) or before preparing and serving food in the household. Additional measures, such as disinfecting the areas where soiled clothes are gathered with a simple disinfectant such as Lysol or diluted bleach, are also useful. Application of these measures within a household with an infected individual will reduce the transmission of infection to others in proximity to the index case. In other settings, such as developing countries with higher incidence and limited supplies of safe drinking water or facilities for the sanitary disposal of feces, use of hand sanitizers is useful and care in the consumption of nonbottled water or drinking of boiled water will be effective.

There is no available vaccine for shigellosis, although considerable effort has been expended to develop one that is both effective and safe. Prophylactic administration of antibiotics is a bad idea as hygiene is an effective way to prevent infection, and antibiotic exposure only increases the selective pressure for drug resistance to emerge. Quarantine is no longer a primary strategy for control of most infectious diseases because diagnosis and more specific measures can usually be implemented. Quarantine usually fails to prevent the spread of contagious diseases in the community as it is difficult to isolate all of the potential spreaders of infection. Whereas certain anti-inflammatory drugs (e.g., cytokine antagonists) have been shown to be effective in experimental models of shigellosis, none are currently available for clinical use.

310 BRUCELLOSIS

EDSEL MAURICE T. SALVANA AND ROBERT A. SALATA

DEFINITION

Brucellosis is a zoonotic disease with protean manifestations caused by bacteria of the genus *Brucella*. Human infection is acquired via ingestion or inhalation of bacteria in contaminated material. Although occupational exposure is a common risk factor for infection, the vast majority of disease occurs through ingestion of unpasteurized dairy products. Despite continuing efforts to control its spread, brucellosis remains a significant health and economic burden in many countries.

The Pathogen

Brucellae are slow-growing, small, aerobic, nonmotile, nonencapsulated, non–spore-forming, gram-negative coccobacilli. *Brucella abortus* (from cattle), *Brucella suis* (from pigs), *Brucella melitensis* (from sheep, goats and camels), and *Brucella canis* (from dogs) are the species that most commonly infect humans. Genome sequencing shows a high degree of homology among strains despite disparate preferred hosts.[1]

EPIDEMIOLOGY

Etiology

Virulence traditionally varies among the four major species of *Brucella*, although this concept has been recently challenged. Because of its pathogenicity and ability to remain viable in storage for long periods, *Brucella* spp. are potential agents for bioterrorism. *B. abortus* is usually associated with mild to moderate sporadic disease. *B. suis* and *B. melitensis* infections are associated with suppurative or disabling complications and can have a prolonged course. Infection with *B. canis* has an insidious onset, relapses frequently, and has a chronic but relatively mild course. Two marine species, *Brucella pinnipediae* and *Brucella cetaceae,* related to seals and cetaceans, respectively, can cause mild infection in humans. *Brucella microti,* with a high potential for pathogenicity, has been isolated from the common vole, red fox, and soil, but no instances of human infection have been reported. Other newly described *Brucella* species with known or potential human pathogenicity include *Brucella inopinata* (one human case of breast implant infection), *Brucella ovis* (sheep infection; no human cases reported), and *Brucella neotomae* (rodent infection; no human cases reported). BO2, a proposed future species closely related to *B. ovis*, has caused one human case of chronic destructive pneumonia.[1,2]

Incidence and Prevalence

More than 500,000 cases of brucellosis are reported yearly to the World Health Organization (WHO) from 100 countries. *B. melitensis* infection accounts for the majority of cases, distributed primarily in the Mediterranean region (particularly Spain and Greece), Latin America, the Arabian Gulf, and the Indian subcontinent. *B. abortus* infection occurs worldwide but has been effectively eradicated in several European countries, Japan, and Israel. Whereas *B. suis* occurs mainly in the Midwestern United States, South America, and Southeast Asia, *B. canis* infection is most common in North America, South America, Japan, and central Europe. Identification of the infecting *Brucella* species helps determine the likely source of infection.[3]

In animals, brucellosis is a chronic infection that can persist throughout life. Because of effective control programs in animals, the incidence of human brucellosis has decreased dramatically in the United States, from more than 6000 cases in 1947 to fewer than 200 cases each year since 1980. States reporting the greatest number of cases include Texas, California, Virginia, and Florida. In North America, brucellosis occurs mainly in the spring and summer, is more common in men, and is usually related to occupational exposure.

Brucella infection in the United States occurs mostly through direct contact with animals or animal secretions in high-risk groups, including slaughterhouse workers, farmers, dairy workers, veterinarians, and travelers returning from endemic areas. Laboratory workers handling infected animals or *Brucella* cultures are also at risk. More than half of the reported cases are associated with the meat-processing industry, particularly the "kill areas," where infection is spread through abraded or lacerated skin; the conjunctiva, possibly by aerosolization; and, rarely, by ingestion of infected tissue. Many cases of *B. abortus* infection in veterinarians have resulted accidentally from exposure to the strain 19 vaccine used to immunize cattle. In the southern United States, 20% of feral swine are positive for *B. suis,* and human infections in hunters have been reported. *B. melitensis* infection, transmitted through the ingestion of goat's milk cheese, has been seen in U.S. travelers to and immigrants from Mexico. Brucellosis contracted abroad may not become symptomatic until the patient returns to the United States. Although persons with human immunodeficiency virus (HIV) infection are generally at higher risk for intracellular pathogens, the clinical manifestations of brucellosis in HIV-infected and noninfected individuals are similar in the few cases of coinfection that have been reported. Human-to-human transmission is rare, but there have been increasing reports of sexually transmitted brucellosis.[4]

Brucellosis in pregnancy has been associated with spontaneous abortions, congenital abnormalities, and neonatal infections. Childhood brucellosis also occurs, mostly in school-aged children, but accounts for only 3% to 10% of all reported cases worldwide. It is more common in endemic areas, where it may account for 20% to 25% of cases, and is often a mild, self-limited process.[5]

PATHOBIOLOGY

Pathogenesis

After penetrating the epithelial cells of human skin, conjunctiva, pharynx, intestine, or lung, *Brucella* organisms in naive individuals induce a delayed inflammatory response (up to 48 hours) with polymorphonuclear leukocyte infiltration at the infection site. *Brucella* organisms are then ingested by dendritic cells, neutrophils and tissue macrophages, and these subsequently spread to regional lymph nodes. If host defenses within the lymph nodes are overwhelmed, bacteremia follows. The usual incubation period from

infection to bacteremia is 2 to 4 weeks. Bacteremia is accompanied by phagocytosis of free *Brucella* organisms by macrophages, and localization of the disease primarily to the spleen, liver, and bone marrow, with the formation of small, noncaseating granulomas, which can serve as persistent sources of infection.

As an intracellular organism, *Brucella* spp. have to avoid detection by the immune system on entry and must be able to survive a hostile intracellular environment. *Brucella* organisms avoid initial detection by the host through multiple mechanisms. Its cell wall lipopolysaccharide (smooth LPS) differs significantly from regular bacterial LPS in two important ways: it has very little effect on Toll-like receptor type 4 (TLR4) activation, and it is resistant to complement activation. In addition, *Brucella* organisms deploy a protein that interferes with TLR signaling. Upon successful entry into the host, *Brucella* organisms in phagosomes are able to survive acidification and lysosome fusion through the induction of specific virulence factors such as the VirB type IV secretion system (T4SS). To replicate, *Brucella* organisms intercept traffic between the endoplasmic reticulum and the Golgi apparatus. They also seem to inhibit apoptosis of the infected cell, thereby maintaining a persistent presence protected from the immune system.[2]

Immunity

Humoral factors play an important role in host defense against *Brucella* spp. Even in the absence of specific agglutinating antibody, normal human serum is bactericidal for *Brucella* organisms; *B. abortus* is more susceptible to serum lysis than is *B. melitensis.* The intracellular location of *Brucella* spp. within macrophages enables it to escape the lethal effects of serum to a certain extent. Specific serum agglutinating antibody has opsonic activity but does not correlate with the development of protective immunity.

A role for mononuclear phagocytes and cell-mediated immunity in brucellosis has been demonstrated. Prior infection with *Listeria monocytogenes* or *Mycobacterium tuberculosis,* both of which stimulate cell-mediated immune mechanisms, is protective against *Brucella* infection in animals. Skin testing with *Brucella* proteins elicits a typical delayed hypersensitivity response in infected individuals. Macrophages, activated with T helper 1 (T_H1)–type cytokines (including interferon-γ, tumor necrosis factor-α, interleukin-1, interleukin-12), kill *Brucella.* Studies have shown that despite high levels of T_H1 cytokine production, deficient effector phagocytic activity persists. Later in the course of infection, there is evidence of an unexpected inhibitory effect from neutrophils. Animal models have demonstrated more efficient killing of *Brucella* organisms in the absence of polymorphonuclear cells, which somehow dampen the immune response to this pathogen.[6]

CLINICAL MANIFESTATIONS

Clinically, human brucellosis can be divided into subclinical illness, acute or subacute disease, localized disease, relapsing infection, and chronic disease (Table 310-1).[7]

Subclinical Illness

Asymptomatic or clinically unrecognized human brucellosis often occurs in high-risk groups, including slaughterhouse workers, farmers, and veterinarians. The diagnosis is usually made through serologic means. More than 50% of abattoir workers and up to 33% of veterinarians have high anti-*Brucella* antibody titers but no history of recognized clinical infection. Children in endemic areas frequently have subclinical illness.

Acute and Subacute Disease

After an incubation period of several weeks or months, acute brucellosis may occur as a mild, transient illness (*B. abortus* or *B. canis*) or as an explosive, toxic illness with the potential for multiple complications (*B. melitensis*). Approximately 50% of patients have an abrupt onset over days, but the remainder have an insidious onset over weeks. Symptoms in brucellosis are protean and nonspecific. More than 90% of patients experience malaise, chills, sweats, fatigue, and weakness. More than 50% of patients have myalgias, anorexia, and weight loss. Fewer patients complain of arthralgias, cough, testicular pain, dysuria, ocular pain, or blurring of vision. Few localizing physical signs are apparent. Fever, with temperatures often greater than 39.4° C (103° F), occurs in 95% of patients. An undulating or intermittent fever pattern is not unusual. A pulse–temperature deficit (i.e., relative bradycardia) may occur. Splenomegaly is present in 10% to 15% of cases, and lymphadenopathy occurs in about 14% of patients. Axillary, cervical, and supraclavicular lymphadenopathy are most frequent and may be related to hand wounds or oropharyngeal routes of infection. Hepatomegaly is less frequent. Other laboratory findings in acute or subacute disease may include mild anemia; lymphopenia; or neutropenia (especially with bacteremia); lymphocytosis; thrombocytopenia; or, in rare cases, pancytopenia. The majority of infected individuals recover completely without sequelae if diagnosed early with prompt initiation of therapy.[8]

Localized Disease and Complications

Brucella organisms can localize to almost any organ but usually target the bones, joints, central nervous system, heart, lung, spleen, testis, liver, gallbladder, kidney, prostate, pancreas, and skin. Disease may occur at multiple sites. Complications with local manifestations most often appear in association with chronic illness, although complications may also occur with acute disease caused by *B. melitensis* or *B. suis.* In the United States, localized disease is most frequently related to *B. suis.* Osteoarticular complications account for 10% to 80% of localized disease in most reported series. Whereas sacroiliitis is the most common manifestation in young persons, spondylitis is more frequently encountered in elderly adults. Vertebral osteomyelitis, particularly in the lumbar area, is also a well-recognized complication and can be associated with paravertebral, epidural, and psoas abscesses.

Relapsing Infection

Up to 10% of patients with brucellosis relapse after antimicrobial therapy. The intracellular location of *Brucella* organisms predisposes to recurrence because the organisms are relatively protected from host defense mechanisms, and antimicrobial agents may be unable to penetrate efficiently enough to kill all the bacteria. Acquired resistance to antibiotics is another factor that can lead to treatment failure. Relapses usually occur 3 to 6 months after completion of therapy but may be seen up to 2 years after initial treatment. Relapsing

TABLE 310-1 CLINICAL CLASSIFICATION OF HUMAN BRUCELLOSIS

CLASSIFICATION	DURATION OF SYMPTOMS BEFORE DIAGNOSIS	MAJOR SYMPTOMS AND SIGNS	DIAGNOSIS	COMMENTS
Subclinical	—	Asymptomatic	Positive (low-titer) serology, negative cultures	Occurs in abattoir workers, farmers, and veterinarians
Acute and subacute	Up to 2-3 mo and 3 mo-1 yr, respectively	Malaise, chills, sweats, fatigue, headache, anorexia, arthralgias, fever, splenomegaly, lymphadenopathy, hepatomegaly	Positive serology, positive blood or bone marrow cultures	Presentation can be mild, self-limited *(B. abortus)* or fulminant with severe complications *(B. melitensis)*
Localized	Occurs with acute or chronic untreated disease	Related to involved organs	Positive serology, positive cultures in specific tissues	Bone or joint, genitourinary, hepatosplenic involvement most common
Relapsing	2-3 mo after initial episode	Same as acute illness but may have higher fever and more fatigue, weakness, chills, and sweats	Positive serology, positive cultures	May be extremely difficult to distinguish relapse from reinfection
Chronic	>1 yr	Nonspecific presentation, but neuropsychiatric symptoms and low-grade fever most common	Low titer or negative serology, negative cultures	Most controversial classification; localized disease may be associated

infection is difficult to distinguish from reinfection in high-risk groups with continued exposure. Relapses are associated with inappropriate or insufficient antimicrobial therapy, growth on blood cultures during the initial presentation, and an acute onset of disease.

Chronic Disease

Disease with a duration of more than 1 year is referred to as *chronic brucellosis.* A majority of patients classified as having chronic brucellosis really have persistent disease caused by inadequate treatment of the initial episode, or they have focal disease in bone, liver, or spleen. About 20% of patients diagnosed with chronic brucellosis complain of persistent fatigue, malaise, and depression; in many respects, this condition resembles chronic fatigue syndrome. These symptoms are frequently not associated with clinical, microbiologic, or serologic evidence of active infection and may represent a preexisting psychoneurosis.

DIAGNOSIS

Culture

Many common illnesses mimic the clinical presentation of brucellosis, and a thorough history is essential, including occupation, travel to endemic areas, avocations, and ingestion of at-risk food and beverages. The most conclusive means of establishing the diagnosis of brucellosis is the recovery of the organism from a culture from normally sterile body fluid or tissue. Sensitivity of cultures have ranged from 15% to 90%, depending on the methods used and the specimen type. In cases of suspected brucellosis, the microbiology laboratory should be asked to extend the length of incubation because it may take more than 5 days for *Brucella* organisms to grow. Handling of *Brucella* cultures is potentially hazardous to laboratory personnel.

In acute brucellosis, blood cultures are positive in 10% to 30% of cases, but this may be as high as 85% with *B. melitensis*. The sensitivity of blood cultures decreases with increasing duration of illness. With *B. melitensis* infection, bone marrow cultures are more sensitive than blood cultures. With localized brucellosis (e.g., lymph nodes, spleen, liver, skeletal system), cultures of purulent material or tissues usually yield *Brucella* organisms. Culture of cerebrospinal fluid turns positive in 45% of patients with meningitis. Antibodies against *Brucella* may be demonstrated in cerebrospinal fluid by enzyme-linked immunosorbent assay (ELISA).

Standard Tube Agglutination

In the absence of microbiologic confirmation, a presumptive diagnosis can be made by history and serology. The most frequently used test is the standard tube agglutination (STA) test, measuring antibody titers against *B. abortus* antigen. A fourfold or greater rise in titer over 2 weeks is considered significant. A presumptive case is one in which the agglutination titer is positive (1 : 160 in endemic areas; 1 : 80 in nonendemic areas) in single or serial specimens, with symptoms consistent with brucellosis. By 3 weeks of illness, more than 97% of patients demonstrate serologic evidence of infection. This test detects antibodies to *B. abortus, B. suis,* and *B. melitensis* but not to *B. canis.* Serologic confirmation of *B. canis* infection requires *B. canis* or *B. ovis* antigen. After adequate antibiotic treatment, significant STA titers can persist for up to 2 years in 5% to 7% of cases. Because of this, STA titers are not useful in differentiating relapsing infection from other febrile illnesses in patients with a history of past *Brucella* infections. Individuals with subclinical infection may demonstrate significant STA titers. In chronic localized brucellosis, STA titers may appear absent or low owing to a prozone phenomenon. This prozone effect appears to be related to the presence of immunoglobulin G or immunoglobulin A blocking antibodies; it can be eliminated if dilutions are carried out to at least 1 : 1280. False-positive STA titers related to immunologic cross-reactivity have been associated with *Brucella* skin testing, cholera vaccination, or infection with *Vibrio cholerae, Francisella tularensis,* or *Yersinia enterocolitica.*

Other Tests

Newer generation antibody tests, including ELISA, are more sensitive and specific than STA and are being used more widely. Preliminary studies using polymerase chain reaction of blood and other fluids or tissues offers rapid and highly accurate diagnosis of brucellosis.[9] Sequencing of specific gene products can identify the organism up to the species level. However, protocols still need to be standardized on a wider scale, and access to expertise and adequate laboratory facilities remains a significant limiting factor.

TREATMENT

Rx

Effective treatment of *Brucella* infection requires antibiotics that can penetrate the intracellular compartment, have little or no toxicity even with prolonged use for preventing relapse, and are bactericidal for adequate treatment of central nervous system infection and endocarditis. There remains considerable debate over which antibiotic regimen is best. In adults, the combination of oral doxycycline at 100 mg orally twice a day for 6 weeks plus intramuscular gentamicin at 5 mg/kg for 5 to 7 days is equally effective as traditional therapy using doxycycline for 6 weeks plus streptomycin 1 g intramuscularly for 14 days.■ The WHO recommends doxycycline plus rifampin 15 mg/kg orally for 6 weeks. This regimen is less effective for cases of spondylitis, which may require up to 3 months of treatment with any of the above regimens. Monotherapy with fluoroquinolones has been disappointing, and if these agents are used, they should always be combined with another active agent. Recent in vitro studies have demonstrated significant activity and synergy of tigecycline with gentamicin and rifampin; these observations must be supported in clinical trials. A prospective, non-randomized trial of triple combination doxycycline, streptomycin, and rifampicin versus the standard double combination doxycycline plus streptomycin showed a significantly higher rate of undetectable Brucella DNA on follow-up.[10] Recommendations are summarized in Table 310-2.

PREVENTION

The control of human brucellosis is directly related to prevention programs in domestic animals and the avoidance of unpasteurized milk and milk products. In slaughterhouses, important means of prevention include careful wound dressing, the use of protective glasses and clothing, the prohibition of raw meat ingestion, and the use of previously infected (immune) individuals in high-risk areas. Work is ongoing to find an effective vaccine for humans. Postexposure antimicrobial prophylaxis is controversial.

TABLE 310-2 TREATMENT FOR BRUCELLOSIS

	TREATMENT	COMMENTS
Acute, with no endocarditis or CNS involvement	Doxycycline (200 mg/day) plus rifampin (15 mg/kg/day) for 6 wk *Or* Tetracycline (2 g/day) for 6 wk plus streptomycin (1 g/day) or gentamicin (5 mg/kg/day) for 1 wk	Treatment of choice by WHO; widely used; low rate of relapse; IMG administration of streptomycin may be difficult
Alternative agents: chloramphenicol, fluoroquinolones, TMP-SMX, imipenem	Combination therapy still preferred; fluoroquinolones plus rifampin is an alternative	
In children	TMP-SMX plus rifampin	
CNS	Doxycycline plus rifampin and TMP-SMX	Third-generation cephalosporin can be substituted if susceptible in vitro
Localized	Surgically drain abscesses plus antimicrobial therapy for ≥6 wk	
Brucella endocarditis	Bactericidal drugs; early valve replacement may be necessary	Possible aortic valve destruction and/or major arterial emboli

CNS, Central nervous system; *IM,* intramuscular; *TMP-SMX,* trimethoprim–sulfamethoxazole; *WHO,* World Health Organization.

PROGNOSIS

Brucellosis treated appropriately within the first month of symptom onset is curable. Acute brucellosis often produces severe weakness and fatigue, and patients are frequently unable to work for up to 2 months. Immunity to reinfection follows initial *Brucella* infection in the majority of individuals. With early antimicrobial therapy, cases of chronic brucellosis or localized disease and complications are rare. Of the patients who die of brucellosis, 84% have endocarditis involving a previously abnormal aortic valve, often associated with severe congestive heart failure. A recent retrospective review showed a much higher risk of death with medical treatment alone compared with a combined medical and surgical approach for *Brucella* endocarditis, although this needs to be confirmed in prospective trials.[11]

Grade A Reference

A1. Roushan MR, Amiri MJ, Janmohammadi N, et al. Comparison of the efficacy of gentamicin for 5 days plus doxycycline for 8 weeks versus streptomycin for 2 weeks plus doxycycline for 45 days in the treatment of human brucellosis: a randomized clinical trial. *J Antimicrob Chemother*. 2010;65: 1028-1035.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Pappas G. The changing Brucella ecology: novel reservoirs, new threats. *Int J Antimicrob Agents*. 2010;36:S8-S11.
2. Atluri VL, Xavier MN, de Jong MF, et al. Interactions of the human pathogenic Brucella species with their hosts. *Annu Rev Microbiol*. 2011;65:523-541.
3. Seleem MN, Boyle SM, Sriranganathan N. Brucellosis: a re-emerging zoonosis. *Vet Microbiol*. 2010;140:392-398.
4. Meltzer E, Sidi Y, Smolen G, et al. Sexually transmitted brucellosis in humans. *Clin Infect Dis*. 2010;51:e12-e15.
5. Yagupsky P. Pediatric brucellosis: an (almost) forgotten disease. *Adv Exp Med Biol*. 2011;719: 123-132.
6. Barquero-Calvo E, Martirosyan A, Ordoñez-Rueda D, et al. Neutrophils exert a suppressive effect on Th1 responses to intracellular pathogen Brucella abortus. *PLoS Pathog*. 2013;9:e1003167.
7. Buzgan T, Karahocagil MK, Irmak H, et al. Clinical manifestations and complications in 1028 cases of brucellosis: a retrospective evaluation and review of the literature. *Int J Infect Dis*. 2010;14: e469-e478.
8. Eales KM, Norton RE, Ketheesan N. Brucellosis in northern Australia. *Am J Trop Med Hyg*. 2010;83: 876-887.
9. Wang S, Zhao G, Wang W, et al. Pathogenicity of two toxoplasma gondii strains in chickens of different ages infected via intraperitoneal injection. *Avian Pathol*. 2014;43:91-95.
10. Vrioni G, Bourdakis A, Pappas G, et al. Administration of a triple versus a standard double antimicrobial regimen for human Brucellosis more efficiently eliminates bacterial DNA load. *Antimicrob Agents Chemother*. 2014;58:7541-7544.
11. Keshtkar-Jahromi M, Razavi SM, Gholamin S, et al. Medical versus medical and surgical treatment for brucella endocarditis. *Ann Thorac Surg*. 2012;94:2141-2146.

REVIEW QUESTIONS

1. A 44 year-old abattoir worker of Middle Eastern origin working in Ashtabula, Ohio develops fever and back pain over the past 2 weeks. He has three pet dogs at home. A standard tube agglutination (STA) shows a *Brucella* titer of 1 : 80. Which of the following is true regarding the diagnosis of *Brucella* in this patient?

A. A single titer of 1 : 80 is sufficient to make the diagnosis of *Brucella* infection in this patient.
B. This patient may have *Brucella canis* infection based on the positive STA titer result.
C. The sensitivity of a blood culture at this time is about 30% and will likely increase with increasing length of illness.
D. A bone marrow culture will be less sensitive than blood culture.
E. The microbiology laboratory should be alerted regarding the suspected diagnosis.

Answer: E The microbiology laboratory should always be alerted in cases of suspected brucellosis. Aside from a potential biohazard, cultures may need to be kept longer than the usual incubation periods to increase yield. A single titer of 1 : 80 on STA is sufficient for a presumptive diagnosis in a nonendemic country. However, this person is from the Middle East, where Brucella is endemic, and this titer may represent previous exposure or infection. *B. canis* is not detected by the STA, and although the patient does have possible exposure from his three dogs, the STA would not be positive from this *Brucella* species. Blood culture sensitivity is about 30% and decreases with increasing duration of illness. Bone marrow culture is invasive but may be more sensitive than blood culture, especially in *B. melitensis* infection.

2. A 54-year old female veterinarian is admitted with a 1-week history of dyspnea and fever. Chest radiography shows cephalization and congestive changes with small bilateral pleural effusions. On examination, she has neck vein engorgement, fine bibasilar crackles, and a grade 4/6 systolic murmur. A transthoracic echocardiogram shows an 8-mm vegetation on the aortic valve. She is started on vancomycin 1 g intravenously every 12 hours, ceftriaxone 2 g intravenously every 24 hours, and diuretics with some improvement. Blood cultures grow *Brucella melitensis*. Which exposure is the most likely cause of this patient's brucellosis?

A. Ingestion of unpasteurized goat cheese smuggled from Europe
B. Assisting in the delivery of a stillborn calf
C. A recent encounter with seals and sea lions at the local aquarium
D. A visit to a pig farm to give vaccines
E. A bite from a dog

Answer: A *B. melitensis* is acquired from goats, sheep, and camels, including unpasteurized dairy products from these animals. Bovine hosts transmit *B. abortus,* especially from spontaneously aborted fetuses. Seals and sea lions are possible sources of *B. pinnipedia,* pigs transmit *B. suis,* and dogs harbor *B. canis*.

3. The patient in question 2 continues to have shortness of breath and signs of congestion despite adequate diuresis. Then correct management for this patient is

A. continue vancomycin but shift the ceftriaxone to ciprofloxacin and repeat transthoracic echocardiography in 6 weeks.
B. continue vancomycin and ceftriaxone and refer to a cardiothoracic surgeon for early valve replacement.
C. continue vancomycin but shift the ceftriaxone to ciprofloxacin and refer to a cardiothoracic surgeon for early valve replacement.
D. stop vancomycin and ceftriaxone and start doxycycline and intramuscular streptomycin and refer to a cardiothoracic surgeon for early valve replacement.
E. stop vancomycin and ceftriaxone and start doxycycline and intramuscular streptomycin and repeat echocardiogram in 6 weeks.

Answer: D The regimens of choice for *Brucella* infection include doxycycline in addition to either streptomycin or gentamicin. Early valve replacement has been shown to decrease mortality in *Brucella* endocarditis, so prompt referral to a cardiothoracic surgeon should be made. Vancomycin and ceftriaxone are not reliable alternative agents, and even though third-generation cephalosporins are used in alternative regimens, monotherapy is not recommended. Fluoroquinolones likewise have some activity against *Brucella* spp. but should not be used as monotherapy.

4. The intracellular nature of *Brucella* means that it must be able to avoid detection before entry into the cell and survive the defense mechanisms that the cell deploys to destroy pathogens. Which of the following statements is consistent with the strategy *Brucella* uses in its pathogenesis?

A. *Brucella* has a modified lipopolysaccharide (LPS) that has substantially decreased activity against TLR9 but is still activated by complement.
B. Pathogenic *Brucella* organisms are able to prevent acidification of the phagosome.
C. Specific virulence factors such as the VirB type IV secretion system (T4SS) are induced with phagosome acidification, which enables the bacteria to survive both acidification and lysosome fusion.
D. *Brucella* intercepts traffic between the nucleus and the ribosomes in order to replicate.
E. Neutrophils are an integral part of the response against *Brucella*, and their absence leads to substantially depressed immune activation.

Answer: C Specific virulence factors such as T4SS are activated by phagosome acidification and enable *Brucella* to survive both acidification and lysosome fusion. Modified Brucella LPS has little activity against TLR4 and is resistant to complement activation. Acidification is an integral part of inducing T4SS, and *Brucella* does not inhibit this process. *Brucella* intercepts traffic between the endoplasmic reticulum and the Golgi apparatus, not between the nucleus and ribosomes. Neutrophils have been found to be inhibitory to the immune response toward *Brucella,* and absence of these cells leads to an enhanced response.

5. A 28-year-old farmer develops fever and severe hip pain over the past 3 weeks. Magnetic resonance imaging shows sacroiliitis, and his blood cultures grow gram-negative coccobacilli. The following measures could have been taken to decrease the risk of this disease EXCEPT

A. wearing protective glasses and clothes when handling cattle and cattle products.
B. pasteurization of cheese and dairy.
C. eating only fully cooked meat.
D. vaccination against the pathogen.
E. vaccination of farm animals against the pathogen.

Answer: D There is currently no approved human vaccine against *Brucella* spp. All of the other choices are correct.

311 TULAREMIA AND OTHER *FRANCISELLA* INFECTIONS

WILLIAM SCHAFFNER

DEFINITION

Tularemia is an infectious zoonosis caused by *Francisella tularensis,* a small aerobic, pleomorphic, gram-negative bacillus. Many animal species harbor the organism, most prominently rabbits, squirrels, and muskrats. Humans acquire the infection through various means, including direct contact with infected animal tissues, ingestion of contaminated water or meat, the bite of an infected tick or deer fly, or breathing an aerosol of bacteria.[1,2] Although *F. tularensis* is highly infectious and is a well-recognized risk to laboratory personnel manipulating culture plates of the organism, it is a paradox that the illness is not communicable from person to person. Edward Francis established that deer flies can transmit the infection from animals to humans and provided detailed descriptions of its clinical manifestations. Colloquially, the disease is often referred to as *rabbit fever* or *deer fly fever.*

The Pathogen

The organism occurs in two major subspecies (biovars). *F. tularensis* biovar *tularensis* (type A) is more virulent in animals and humans, has distinctive biochemical reactions (it produces acid from glycerol and has citrulline ureidase activity), and is the common North American biovar. In contrast, *F. tularensis* biovar *holarctica* (type B) is less virulent and occurs commonly in Europe and Asia. Type B is most frequently isolated from rodent species, including muskrats *(Ondatra zibethicus),* mice *(Mus musculus),* beavers *(Castor canadensis),* voles (*Microtus* spp.), and water voles *(Arvicola terrestris),* and has been associated with an outbreak of infection in wild-caught prairie dogs. Specific virulence factors for *F. tularensis* have not been identified.

EPIDEMIOLOGY

Tularemia has been reported in the United States, Canada, Mexico, Japan, and Europe (particularly Scandinavia). It has not been reported in the United Kingdom or the Southern Hemisphere. In the United States, reported cases diminished during the second half of the 20th century from a high of 2291 cases in 1939 to the approximately 125 cases reported annually at present. Tularemia has occurred in all the continental states, but four states account for 56% of reported cases: Arkansas, Missouri, Oklahoma, and South Dakota. The island of Martha's Vineyard off the coast of Massachusetts is also a focus of tularemia.

In the United States, tularemia is usually acquired from tick bites or from contact with infected animals, especially rabbits. Tick-associated cases now constitute the majority and occur during the summer. The most common vectors in the United States are the wood tick *(Dermacentor andersoni),* the dog tick *(Dermacentor variabilis),* and the Lone Star tick *(Amblyomma americanum).* A smaller peak of autumn and winter cases is a consequence of rabbit hunters skinning and eviscerating their game. Public health education materials aimed at decreasing the hazards of handling wild animals have contributed to the reduction of tularemia in hunters. Mosquitoes are the common vectors in northern Europe. Occasional individuals acquire infection from the bite of an infected animal or, more likely, from the bite of an animal whose mouth was contaminated from recently eating a diseased animal. The latter likely explains most instances of cat-bite tularemia.

Males experience a higher incidence of disease than females in all age groups, probably as a consequence of their greater exposure to the outdoors and animal-related activities and less use of protective measures against tick bites. Persons in all age groups are affected, with children 5 to 14 years of age and older adults most prominently represented. In the United States, American Indians and Alaska natives experience the highest annual incidence (0.5 per 100,000); whites have a lower risk (0.04 per 100,000), and African Americans and Asians/Pacific Islanders have the lowest occurrence of tularemia (≤0.01 per 100,000).

Although tularemia is usually a sporadic infection, outbreaks of disease have been traced to laboratory exposure, contaminated groundwater, muskrat handling, lawn mowing, and brush cutting. In the latter two cases, primary pneumonic tularemia apparently occurred when the affected individuals created an environmental aerosol by mowing grass and cutting brush that had been contaminated with *F. tularensis* excreted in the urine and feces of infected rodents. The organism can survive in water, mud, and straw for weeks to months.

Interest in tularemia has been enhanced because of its potential use as a bioterrorism agent.[3] Its high infectivity (as few as 10 organisms have induced pneumonic disease), its ease of dissemination, and the difficulty of rapidly diagnosing acute illness are characteristics that merit its inclusion among threat agents. Thus, tularemia must be reported immediately to local public health authorities. Unusual patterns of disease will be investigated for both conventional and bioterrorist sources.

PATHOBIOLOGY

F. tularensis can infect humans through several portals of entry, including the skin, mucous membranes, and gastrointestinal and respiratory tracts. It requires intracellular residence and can multiply within macrophages and other cells. After inoculation into the skin and subcutaneous tissue, local bacterial multiplication occurs and evokes a suppurative necrotic reaction characterized by an initial polymorphonuclear response followed by an influx of macrophages and lymphocytes. These suppurative lesions evolve into granulomas. Bacteremia can occur both early and late during this process. The infection can disseminate to the lymph nodes, liver, spleen, lungs, and pleura. Viable *F. tularensis* can persist in tissues for long periods, contributing to the tendency to relapse after treatment.

CLINICAL MANIFESTATIONS

Classically, the clinical manifestations of tularemia have been separated into six categories: ulceroglandular, glandular, oculoglandular, typhoidal, oropharyngeal, and pneumonic. Although this classification has historic roots, it should not be used rigidly because many patients have features of several types. The course of illness is determined by the portal of entry, the degree of systemic involvement, and the dose and virulence of the infecting strain of *F. tularensis.*[4]

The general features of tularemia are similar regardless of the portal of entry. After exposure, the usual incubation period is 3 to 5 days (range, 1-21 days). The disease begins abruptly with the onset of fever (≥101° F), chills, malaise, and headache. Myalgia, vomiting, sore throat, and abdominal pain can also occur. Almost half the patients have a pulse rate that is substantially slower than would be anticipated based on the degree of fever (pulse–temperature dissociation). The fever may abate somewhat after 1 to 3 days, only to recur and continue along with other symptoms for 2 to 3 weeks. Untreated, weight loss, easy fatigue, and lymphadenopathy may persist for weeks longer.

Ulceroglandular Disease

Ulceroglandular disease is the form of infection most readily recognized by physicians. Along with fever and other constitutional symptoms, the patient calls attention to tender, swollen lymph nodes that drain an inoculation site. The nodes are usually axillary or inguinal, and a local lesion appears concurrently or 1 or 2 days before or after the lymphadenopathy. The lesions at the site of inoculation begin as small, red, tender, or painful papules that progress to pustules and then undergo necrosis to produce an ulcer with sharp, somewhat elevated edges and a flat base that becomes black. Untreated, the ulcers heal over a period of weeks and leave scars. Tick-induced infections produce lesions on the trunk, about the waist, and in the perineum, along with the expected local adenopathy. Children typically have occipital and cervical adenopathy from tick bites on the neck and in the hair. Animal exposure often produces lesions on the hands and forearms. Lesions may be multiple. Because the organisms evoke a localized granulomatous response, frank lymphangitis does not occur in uncomplicated tularemia, but an occasional patient manifests a chain of nodules in "sporotrichoid" fashion along the lymphatic drainage.

Patients with such apparently "localized" disease often have symptoms and findings indicating a more widespread infection. Sore throat with or without an erythematous pharynx occurs, as well as chest radiographic findings of patchy infiltrates in the lower lobes, pleural effusions, and hilar adenopathy.

Glandular and Typhoidal Disease

Glandular disease is essentially the same clinical syndrome as ulceroglandular disease but without the local lesion. Thus, the patient has fever, constitutional symptoms, and lymphadenopathy. The local lesion may have been on a part of the body where it was not seen, or it may have been small and already healed by the time the patient sought medical care. Glandular disease accounts for only 3% to 20% of cases. Typhoidal disease does not show evidence of lymphadenopathy and is essentially characterized by fever of unknown cause. These illnesses evade diagnosis unless the physician specifically considers the possibility of tularemia and inquires about tick or animal exposure. Occasionally, the diagnosis is made fortuitously when a positive blood culture is reported.

Oculoglandular Disease

Oculoglandular disease is rare (<5% of cases) and occurs when the conjunctival sac is the portal of entry via an aerosol, splash, or contaminated fingers. It is almost always unilateral and can have a dramatic manifestation with inflamed, swollen eyelids, chemosis, and painful conjunctivitis. The palpebral conjunctiva often shows small yellow nodules and ulcers, counterparts to the skin lesions of ulceroglandular disease. The affected regional lymph nodes are those of the head and neck.

Oropharyngeal Disease

Oropharyngeal disease is also uncommon in the United States and occurs when the mucous membranes of the mouth and pharynx are the portal of entry. Contaminated water or food (inadequately cooked game meat) is the source. Painful exudative pharyngitis and tonsillitis, pharyngeal ulcers, and swollen retropharyngeal and cervical lymph nodes are seen.[5]

Pneumonic Disease

Although pneumonia may be one aspect of the other tularemic syndromes, *pneumonic tularemia* refers to an illness that manifests as a distinctive pneumonia.[6] It accounts for about 10% of reported cases and occurs from inhalational exposure. This is the form of the disease that would result from bioterrorism. In addition to fever and malaise, patients may have a dry cough, substernal discomfort, pleural pain, dyspnea, and sore throat. These pulmonary symptoms may not be very prominent in the context of the systemic illness. Hemoptysis is unlikely. The results of physical examination reflect the extent and distribution of the pneumonic process, which may range from barely evident to frank consolidation with pleural effusion. Radiographic findings range from modest peribronchial infiltrates early in the illness to distinctive bronchopneumonia with effusion. Hilar adenopathy is present in more than one third of cases. Sputum examination is not helpful. Pleural effusions generally contain more than 1000 lymphocytes/mm^3. Gram stain results are negative, and pleural biopsies occasionally contain granuloma, thus inviting confusion with tuberculosis. Without a suggestive history of tick or animal exposure, patients with tularemic pneumonia may be thought to have poorly responding community-acquired pneumonia. Fluoroquinolone antibiotics are commonly used as empirical therapy in this circumstance to treat some patients with undiagnosed tularemia pneumonia.

DIAGNOSIS

The diagnosis of tularemia usually involves serologic testing with tube agglutination or microagglutination techniques. Antibody concentrations do not reach diagnostic levels until after the 11th day of illness. A single acute titer of 1 : 160 is considered presumptive; a confirmed diagnosis requires a fourfold rise in titer between acute and convalescent specimens. Titers of both immunoglobulin M and immunoglobulin G antibodies may remain elevated for many years after the illness.

Differential Diagnosis

The differential diagnosis of patients with tularemia is substantial. The local lesions can be confused with cat-scratch disease, brown recluse spider bites, *Mycobacterium marinum* infection, herpes simplex infection (Chapter 374), and even syphilis (Chapter 319) and chancroid (Chapter 301) when the lesions are in the perineum or on the penis. Pneumonic tularemia resembles common community-acquired pneumonia (Chapter 97), as well as less common infections such as psittacosis, legionellosis, and Q fever. The glandular and typhoidal forms can resemble typhoid fever (Chapter 308), brucellosis (Chapter 310), ehrlichiosis, and other illnesses accompanied by nonspecific fevers.

Routine laboratory studies do not provide specific results. Leukocyte counts may be within normal limits or elevated; thrombocytopenia, elevated liver enzymes, and sterile pyuria occur with some frequency. *F. tularensis* may be isolated from blood cultures and tissue specimens when media containing cysteine are used. Laboratory personnel should be notified when tularemia is suspected so that appropriate media can be used and to ensure that safeguards are in place to protect against the production of hazardous aerosols.

TREATMENT

Rx

Because tularemia is a relatively uncommon disease, therapeutic recommendations are based on a combination of in vitro studies and accumulated clinical experience. The preferred antimicrobials are streptomycin and gentamicin; either one is given for 10 days. Streptomycin is given at a dose of 1 g intramuscularly twice daily. Gentamicin may be more readily available and is administered at a dose of 5 mg/kg intramuscularly or intravenously once daily. Both chloramphenicol and the tetracyclines have been used in the past to treat tularemia; however, use of both these bacteriostatic agents has resulted in higher rates of relapse than treatment with streptomycin or gentamicin. Because chloramphenicol may produce bone marrow toxicity, it is rarely used today. Doxycycline is administered at a dose of 100 mg intravenously twice daily for 14 days. In recent years, ciprofloxacin has been used successfully in a growing number of patients; the dosage is 750 mg intravenously twice daily for 10 days. Both of these drugs can be switched to oral administration of the same doses as soon as tolerated by the patient. There is a need to develop new therapeutic strategies to improve the management of patients with tularemia.[7,8]

PREVENTION

Prevention of tularemia entails minimizing exposure to ticks and avoiding direct exposure to wild animals. Tick protection includes clothing that extends to the wrists and ankles, regular inspection for attached ticks, and the use of insect repellents containing diethyltoluamide (DEET). Gloves should be worn when skinning and dressing game animals, especially rabbits, and all wild rabbit and other game meats should be cooked thoroughly.

A live, attenuated vaccine has been used in the past to provide some protection to researchers working with *F. tularensis*. The vaccine is not available commercially.

PROGNOSIS

Before treatment became available, acute tularemia often lasted as long as 1 month followed by several months of debility. The mortality rate approached 10%. When appropriately diagnosed and treated, the mortality rate from tularemia is now 1% or less.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Caralho CL, Lopes des Carvalho I, Ze-Ze L, et al. Tularemia: a challenging zoonosis. *Comp Immunol Microbiol Infect Dis*. 2014;37:85-96.
2. Foley JE, Nieto NC. Tularemia. *Vet Microbiol*. 2010;140:332-338.
3. Maurin M. Francisella tularensis as a potential agent of bioterrorism? *Expert Rev Anti Infect Ther*. 2015;13:141-144.
4. Weber IB, Turabelidze G, Patrick S, et al. Clinical recognition and management of tularemia in Missouri: a retrospective records review of 121 cases. *Clin Infect Dis*. 2012;55:1283-1290.
5. Steinrucken J, Graber P. Oropharyngeal tularemia. *CMAJ*. 2014;186:E62.
6. Thomas LD, Schaffner W. Tularemia pneumonia. *Infect Dis Clin North Am*. 2010;24:43-55.
7. Boisset S, Caspar Y, Sutera V, Maurin M. New therapeutic approaches for treatment of tularaemia: a review. *Front Cell Infect Microbiol*. 2014;4:40.
8. Stevens DL, Bisno AL, Chambers HF, et al. Practice guidelines for the diagnosis and management of skin and soft-tissue infections: 2014 update by the Infectious Diseases Society of America. *Clin Infect Dis*. 2014;59:147-159.

REVIEW QUESTIONS

1. In the United States, persons most commonly acquire *F. tularensis* infection from

A. mosquitoes.
B. tick bites.
C. skinning infected rabbits.
D. person-to-person contact.
E. both B and C.

Answer: E Mosquito bites are a common mode of acquisition of infection in Europe. Although highly infectious, tularemia is not spread from person to person.

2. Tularemia is considered a potential bioterrorism agent. Among characteristic(s) of the disease that merit its inclusion among threat agents, which one of the following is not true?

A. *F. tularensis* produces an easily diagnosed illness.
B. It is highly infectious.
C. It can survive in the environment for long periods.
D. It can be disseminated readily (e.g., by aerosol).
E. A vaccine is not available.

Answer: A Indeed, because aerosol inoculation produces a nonspecific pneumonia that often defies quick diagnosis, many cases may occur before a bioterrorism event might be detected. This would have the advantage of disabling a large proportion of the target population rather quickly; the fear initiated by the recognition of the terrorism event would follow when the diagnosis became apparent.

3. It is thought that tularemic pneumonia may occur more commonly than is reported. Reasons for this include

A. underreporting of diagnosed cases to public health authorities.
B. physicians not eliciting exposure histories.
C. omission of specific serologic diagnostic testing.
D. empiric therapy of community-acquired pneumonia with fluoroquinolone antibiotics.
E. all of the above.

Answer: E Because fluoroquinolones are active against *F. tularensis,* empiric treatment may cure some patients even though a specific diagnosis was not established.

312

PLAGUE AND OTHER *YERSINIA* INFECTIONS

KENNETH L. GAGE AND PAUL S. MEAD

The genus *Yersinia* currently contains at least 18 species, only three of which are known to be significant human pathogens (*Yersinia pestis, Yersinia enterocolitica,* and *Yersinia pseudotuberculosis*). The remaining species are normally considered nonpathogenic and are most frequently isolated from environmental specimens, including fish, water, and fecal samples. Although these minor yersiniae have been identified primarily from environmental samples, they are occasionally recovered from patient samples, suggesting a possible role as human pathogens.

YERSINIA PESTIS

DEFINITION

Plague is a highly fatal flea-borne disease that is best known as the cause of the Black Death of the Middle Ages.[1] Its etiologic agent, *Y. pestis,* is a gram-negative coccobacillus belonging to the family Enterobacteriaceae.

The Pathogen

Y. pestis is microaerophilic, gram negative, nonmotile, and nonsporulating; it can exist as a facultative intracellular pathogen and exhibits bipolar staining with Wayson, Giemsa, and Wright stains (Fig. 312-1). It is fragile outside its hosts or vectors but can be grown within 24 to 48 hours on a variety of bacteriologic media at temperatures ranging from 4° to 40° C. *Y. pestis* lacks a true capsule but has a carbohydrate-protein envelope comprised of the *capsular* or *fraction 1 antigen.* Production of this antigen occurs only at temperatures above 33° C. Wild-type strains typically bear three plasmids with sizes of approximately 100 to 110 kilobases (kb), 70 to 75 kb, and 9.5 kb (19 kb if present as a dimer), respectively. Although only one serotype is thought to exist, strains can be classified into biotypes. The three classic biotypes (*antiqua, mediaevalis,* and *orientalis*) differ in their ability to ferment glycerol and reduce nitrates. All three biotypes occur in Asia, which is generally accepted as the continent where plague originated. Two biotypes exist in Africa (*antiqua* and *orientalis*), but only the *orientalis* biotype occurs in the Americas. Although these three biotypes have historical and biogeographic significance, all are highly virulent and appear to cause virtually identical signs and symptoms in humans. Recently, a fourth biotype (*microtus* or *pestoides*), which is purportedly nonpathogenic for humans, was reported from east-central Asia.[2] More modern typing methods, including ribotyping, multiple-locus variable-number tandem repeat assays, and single-nucleotide polymorphism analysis, are proving useful for molecular epidemiology studies, identification of probable environmental sources of human infection, and phylogenetic analyses.

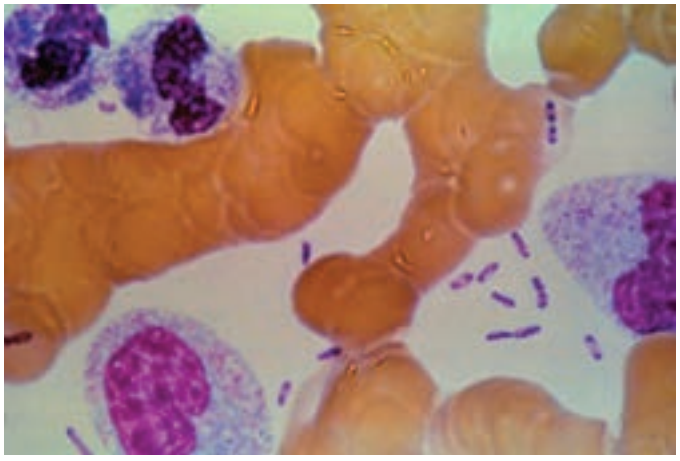

FIGURE 312-1. Wayson-stained blood smear from a fatal case of human septicemic plague (Centers for Disease Control and Prevention).

EPIDEMIOLOGY

Y. pestis is maintained in nature through transmission cycles involving certain rodent species and their fleas, which act as vectors. Although other mammals often become infected with *Y. pestis* and occasionally succumb to plague, the only nonrodent species reported to be important hosts for infecting vector fleas and perhaps contributing to the maintenance of plague in some natural foci are certain species of rabbits and hares, the steppe pika of central Asia, and the house shrew of southeastern Asia and Madagascar. Rodent-consuming carnivores and raptors might play indirect roles in spreading plague by transporting infected rodent fleas from one area to another.

Foci of *Y. pestis* infection occur in rodent populations in many regions of Africa, Asia, and the Americas, although only about 2000 to 5000 human cases were reported each year to the World Health Organization (WHO) in the period 1987 to 2009. In the 1960s and 1970s, the majority of human cases occurred in Southeast Asia, which was the site of civil wars and political unrest. However, beginning in the 1980s, this situation began to change, and most cases now occur in Africa, especially the Congo,[3] and nearby Madagascar.[4] According to the most recently released WHO statistics for 2004 to 2009, 96.6% of the world's cases occurred in the Democratic Republic of Congo, Madagascar, Uganda, and other countries in Africa (Fig. 312-2). Asian countries reported only 1.6% of the total cases, and another 1.8% were reported from five countries in the Americas, including the Brazil, Bolivia, Ecuador, Peru, and the United States. Since 1970, evidence of *Y. pestis* infection has been identified in animals or their fleas in 17 western states, and human cases have been identified in 13 of these states (Fig. 312-3). Most of these cases have occurred in three southwestern states (New Mexico, Arizona and Colorado) and California.

Humans most frequently acquire plague as a result of being bitten by infectious fleas (Fig. 312-4). On a worldwide basis, the risk of flea bite exposure is highest in poverty-stricken areas that are situated near natural plague foci and have large infestations of commensal rats (*Rattus* spp.) heavily infested with fleas, particularly the oriental rat flea (*Xenopsylla cheopis*). *X. cheopis* readily feeds on humans and is an efficient vector of *Y. pestis* to people as well as to rats. Persons are most likely to be bitten by infectious *X. cheopis* when plague epizootics cause massive mortality among susceptible rats, forcing these fleas to seek new hosts. Currently, certain regions of central and southern Africa (including Madagascar), southeastern Asia and India, and a few areas in South America remain at relatively high risk for rat-associated plague outbreaks. The spread of rat-associated plague from one region to another, perhaps by the natural movement of rats or their transport along with trade goods, poses a threat of epidemics in large rat-infested cities, as demonstrated by the repeated spread of plague from the central highlands of Madagascar to the port city of Majunga, where the disease has caused multiple outbreaks in the past 2 decades. The appearance of bubonic cases acquired via flea bite in large cities also greatly increases the possibility that secondary pneumonic plague with hemoptysis will develop in untreated bubonic cases and that the infection will spread to others through inhalation of infectious respiratory droplets expelled by coughing pneumonic plague patients. Such spread will result in cases of primary pneumonic plague, a form of the disease that is often fatal and can spread rapidly from person to person under appropriate circumstances. The persons most at risk for acquiring plague through respiratory droplet spread are those who fail to take appropriate respiratory precautions and come into close contact (≤2 m) with coughing pneumonic plague patients experiencing hemoptysis. Human-to-human spread of primary pneumonic plague has not been reported in the United States since 1924, when an outbreak occurred in Los Angeles, California. The few cases of naturally acquired primary pneumonic plague that have occurred in this country since that time, and for which likely sources of exposure were identified, have involved pet owners or veterinary staff handling cats with signs of plague pneumonia, pharyngitis, or oral abscesses. Although rare in the United States, primary pneumonic plague still poses a threat in developing countries, with outbreaks occurring in the past 20 years in India, Ecuador, Madagascar, the Democratic Republic of Congo, and Uganda. It should be noted, however, that the spread of pneumonic plague depends on many factors, and relatively simple control measures can quickly and effectively stop the respiratory droplet spread of pneumonic plague.[5]

At present, rat-associated plague poses little risk to persons in the United States, with most cases resulting from exposure to the bites of infectious wild rodent fleas (79% in one case series), although a significant number (19%)

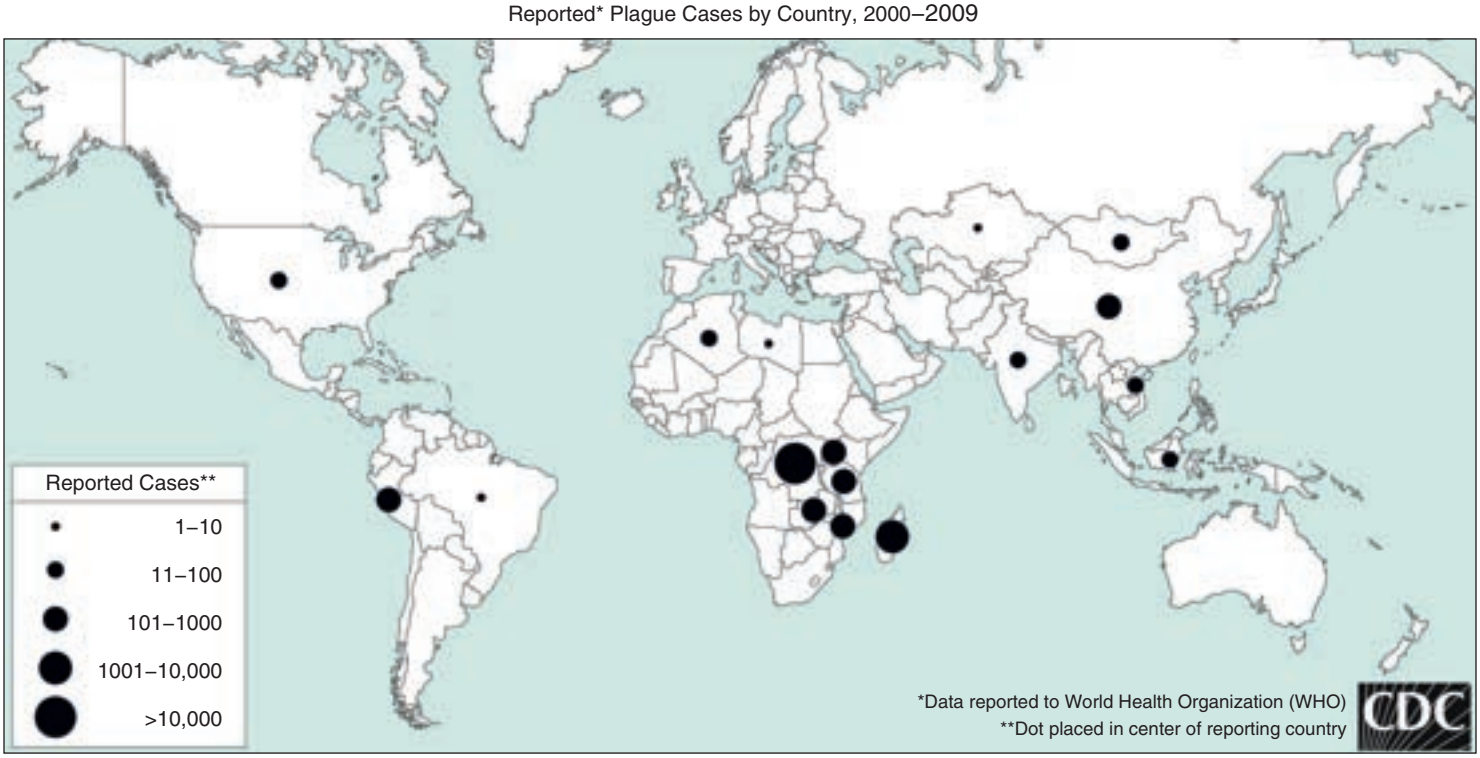

FIGURE 312-2. Worldwide distribution of plague in humans, 2000 to 2009. (Data compiled from the World Health Organization, Centers for Disease Control and Prevention, and other sources.)

were exposed while handling *Y. pestis*–infected animals, including rabbits, domestic cats, wild rodents, and wild carnivores. These direct-contact exposures occurred primarily in persons who had been bitten or scratched by infected cats or had skinned infectious rabbits or rodents for meat or certain wild carnivores for their fur. The risk of exposure to infectious wild rodent fleas or animals is typically greatest when epizootics cause high mortality rates in certain rodent species in the western United States, particularly ground squirrels, prairie dogs, wood rats, and chipmunks. Epizootics appear to progress most rapidly during late spring to early fall, when human cases also peak, probably because this is the period of greatest flea vector activity. Most cases arising from flea bite occur when persons are exposed to fleas as a result of close contact with dead hosts that are flea infested or with the nests or burrows of infected hosts. Evidence also exists that allowing dogs from plague-endemic areas to sleep in their owners' beds increases plague risk for these persons, presumably because these animals carry infectious fleas into beds. The few human cases reported during the winter months are typically acquired through direct contact with infected animals, and affected individuals have a history of hunting or trapping wild animals or handling domestic cats. The exposure sites for most U.S. cases are peridomestic environments, with smaller numbers of exposures occurring during recreational activities such as hunting or camping.[6] Occupational exposure has also rarely occurred among veterinary staff, biologists, and trappers. Rarely, primary pneumonic plague has been acquired in the United States through the inhalation of infectious materials (2% of cases) by means of close face-to-face contact with *Y. pestis*–infected cats that had oral lesions or signs of plague pneumonia, including cough.

Recently, concern has been raised that plague could be used as an agent of bioterrorism (Chapter 21). In most projected scenarios, bioterrorists would spread plague in an aerosol form, potentially resulting in numerous primary pneumonic cases, a high mortality rate, and widespread panic, especially if the *Y. pestis* strains released had been engineered to be resistant to antimicrobial agents commonly used to treat plague.

PATHOBIOLOGY

Y. pestis possesses a variety of virulence factors that promote its initial invasion of the host, evasion of the host's immune system, survival in lymph nodes, and the development of a fulminant bacteremia that can lead to fatal gram-negative sepsis and the spread of the pathogen to other organs.[7] Few bacteria rival the pathogenicity of *Y. pestis* for humans. *Y. pestis* usually enters the body at the site of a flea bite or perhaps as a result of contact between abraded or cut skin and blood or other body fluids from a *Y. pestis*–infected animal. On entering the body, plague bacteria come under attack by host phagocytes and other host defenses. Although many of the invading *Y. pestis* are killed by arriving polymorphonuclear leukocytes, some enter into and survive in mononuclear cells, where they are carried via the lymphatics to regional lymph nodes. The ability to escape from the host's innate immune defenses and disseminate to regional lymph nodes depends in part on a protease (Pla) encoded on the 9.5 kb plasmid that helps degrade fibrin clots and promote the production of excess plasmin, which can affect inflammatory exudates, break down extracellular proteins and basement membranes, and reduce levels of chemoattractants, possibly because of the inhibition of interleukin-8 production at the site of initial infection. Another virulence factor, YopM, is one of many *Yersinia* outer proteins (Yops), encoded by genes on the midsized (70-75 kb) plasmid of *Y. pestis*. Although most other Yops are degraded by the Pla protease, YopM is resistant to its activity and probably aids in the dissemination of *Y. pestis* by competing with platelets for thrombin, a factor that not only reduces clotting but also inhibits the activation of platelets, thereby lowering local inflammatory responses and promoting the dissemination of *Y. pestis* to regional lymph nodes. Initial invasion and dispersal to regional lymph nodes also depend on the ability of *Y. pestis* to survive for at least brief periods in host phagocytes. Survival in such environments is promoted by other Yops that work in concert with a type III secretory apparatus to deliver into host phagocytes those Yops that act as intracellular effectors. These effector Yops disturb the cytoskeletal dynamics of phagocytic cells and block their production of proinflammatory cytokines. Affected phagocytes are rendered incapable of killing the invading *Y. pestis,* thereby allowing this bacterium to survive extracellularly in lymphoid tissues. Survival of *Y. pestis* in mammalian hosts also depends on its ability to acquire sufficient quantities of iron for growth. The most important means of iron uptake in *Y. pestis* is a siderophore (yersiniabactin) system that can effectively compete with host iron-binding molecules for this essential nutrient. Although the ability of *Y. pestis* to resist host phagocytic killing and survive intracellularly in phagocytes is thought to be particularly important during the early stages of infection, plague bacteria also express a glycoprotein capsular antigen (caf1 or fraction 1 antigen) that confers resistance to phagocytosis. Expression of caf1 is temperature dependent, being repressed at the

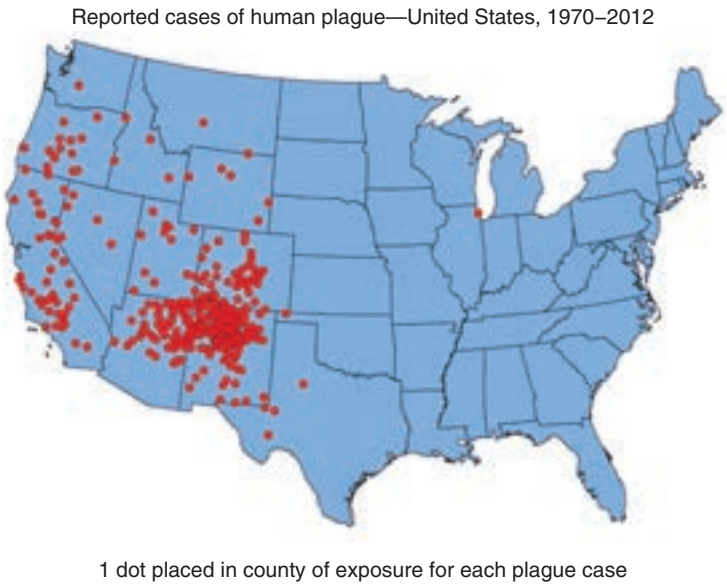

FIGURE 312-3. Distribution of human plague cases in the United States, 1970 to 2012. Case points were randomly placed within counties where exposures occurred to indicate the general distribution and clustering of cases by region. One dot placed in county of exposure for each plague case (Centers for Disease Control and Prevention).

cooler temperatures found in the flea vector and upregulated at mammalian host body temperatures.

Infected lymph nodes, termed *buboes,* can appear edematous and congested early in the course of illness but exhibit little evidence of inflammatory infiltrates or vascular injury. These buboes represent the most obvious manifestation of the lymphatic system's efforts to arrest the spread of *Y. pestis,* and within a few days after infection, they contain massive numbers of *Y. pestis* and heavy neutrophil infiltrates, which cause them to swell to the size of a hen's egg and become surrounded by serous fluid. As the illness progresses, hemorrhagic necrosis and vascular damage in the node become apparent; some nodes spontaneously rupture, and abscesses appear. Although *Y. pestis* can be present in small quantities in blood samples taken relatively early in the course of infection, large quantities of plague bacteria usually appear in the blood of patients with bubonic plague only after the lymph node defenses are overwhelmed.

As the bacteria escape from the node and proliferate in blood, patients with bubonic plague begin to exhibit evidence of plague septicemia (secondary septicemic plague). Patients with inadequately treated septicemic plague can experience widespread and overwhelming destruction of tissues as *Y. pestis* spreads to various organs, eventually resulting in their failure. Patients who die of plague often experience diffuse interstitial myocarditis, cardiac dilation, diffuse hemorrhagic splenic necrosis, renal glomeruli containing fibrin thrombi, and multifocal necrosis in the liver. Disseminated intravascular coagulation can lead to thrombosis within capillaries, vascular necrosis, ecchymoses, acral gangrene, and cutaneous, mucosal, and serosal petechiae

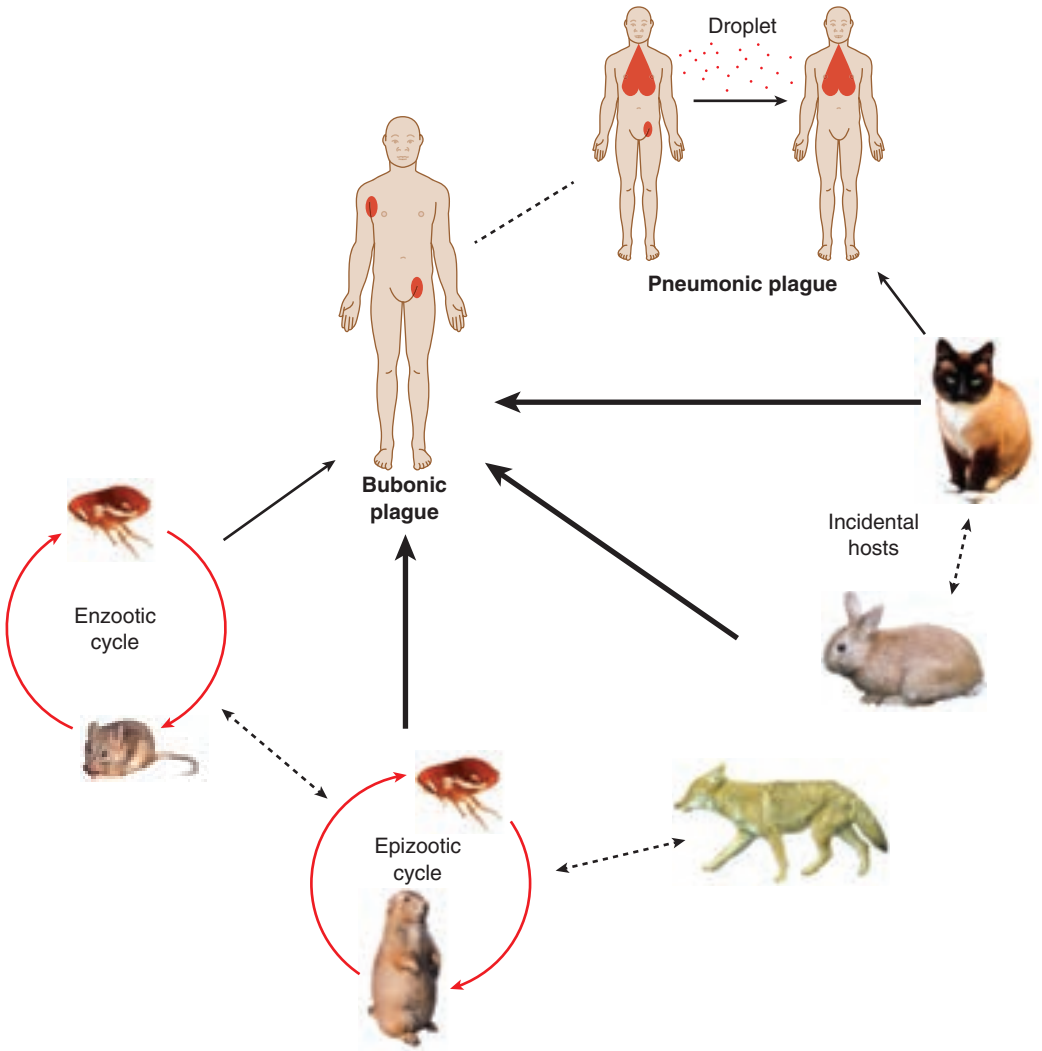

FIGURE 312-4. Ecology and transmission of plague. *Red circles* indicate zoonotic cycles. Common routes of transmission to humans indicated by *wide solid arrows* and uncommon routes by thin *arrows* (Centers for Disease Control and Prevention).

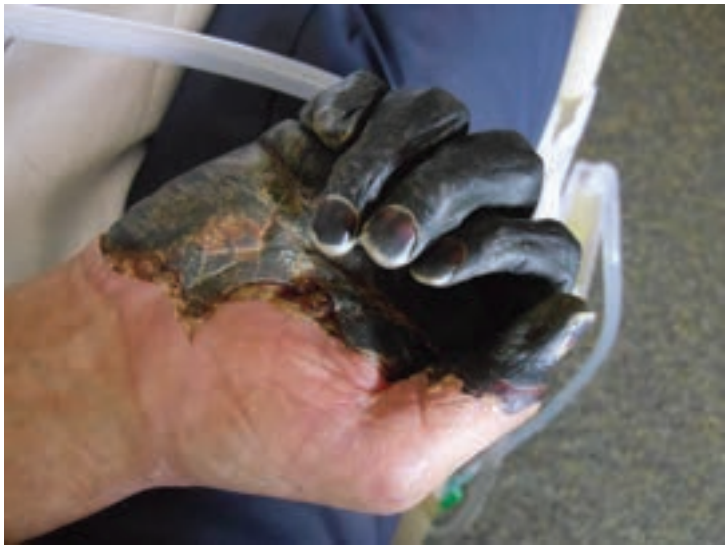

FIGURE 312-5. Hand of a patient with plague displaying acral gangrene, a manifestation that may have given rise to the term "black death" (Centers for Disease Control and Prevention).

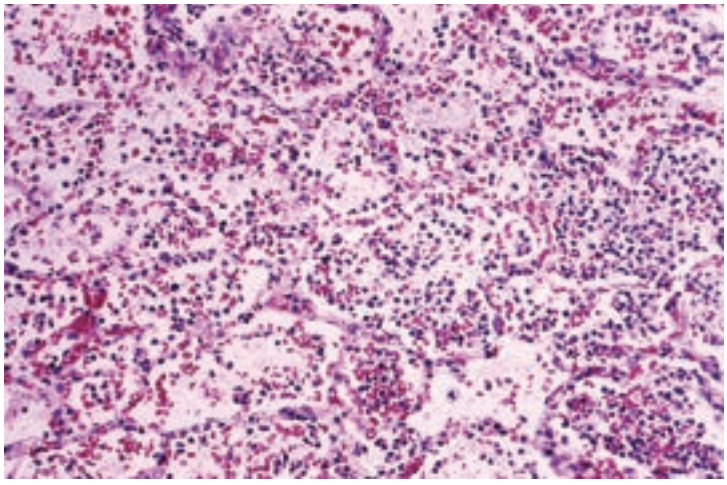

FIGURE 312-6. Photomicrograph of lung tissue from fatal case of primary pneumonic plague and septicemic plague. Note filling of alveolar spaces with inflammatory cells and debris (Centers for Disease Control and Prevention).

(Fig. 312-5). In a small proportion of cases, septicemia occurs in the absence of buboes or other signs of localized infection, a condition referred to as *primary septicemic plague.* Septicemic plague can also occur secondary to primary pneumonic plague.

Rarely, plague can be acquired through the inhalation of infectious respiratory droplets or other materials. Patients with primary pneumonic plague typically experience a rapidly progressing lung infection that is initially lobular, then lobar, and finally multilobar, with large numbers of *Y. pestis* being present in the alveoli and pulmonary secretions of affected sites (Fig. 312-6). Pneumonic plague can also occur secondary to bubonic or primary septicemic plague. In these instances, *Y. pestis* invades the lungs in a diffuse manner, with most bacilli appearing in interstitial spaces. The most common pathologic findings are edema, diffuse pulmonary congestion, limited neutrophilic infiltration, and hemorrhagic necrosis. Cavitation and liquefaction necrosis can also occur and result in scarring of the lungs, particularly at sites of consolidation.

CLINICAL MANIFESTATIONS

The three most commonly observed forms of plague (in order of decreasing occurrence) are bubonic, septicemic, and pneumonic. Unusual manifestations of plague include meningitis and pharyngitis. In rare instances, *Y. pestis* has been inoculated through the conjunctiva, resulting in oculoglandular plague. The incubation periods are 2 to 6 days for bubonic plague and 1 to 3 days for primary pneumonic plague.

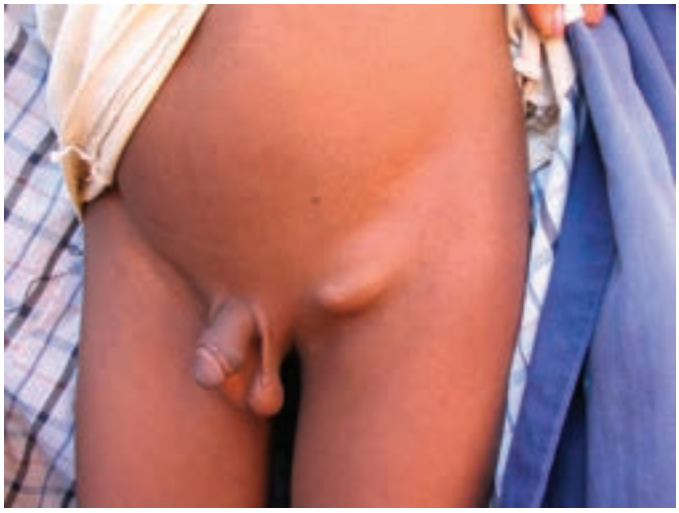

FIGURE 312-7. Young Malagasy boy with a bubo. (Courtesy Brook Yockey, Centers for Disease Control and Prevention.)

The characteristic swollen and tender lymph nodes (buboes) of bubonic plague usually appear in the node or nodes located closest to the site of initial infection (Fig. 312-7). Most cases of bubonic plague in the United States are thought to be acquired from flea bites on the legs, as indicated by the appearance of inguinal or femoral lymph node involvement on the side where the flea bite occurred. Axillary buboes are also common, often indicating the handling of an infected animal or other flea-infested object. Cervical buboes are much less common in the United States than in many developing countries, perhaps because persons in the latter are more likely to sleep on dirt floors in flea-infested huts, thus increasing their chances of being bitten about the head and neck by infectious fleas. Rarely, a skin lesion appears at the site of an infectious flea bite or other source of inoculation.

Symptoms of bubonic plague include fever, chills, myalgia, arthralgia, headache, malaise, and prostration. Untreated patients with bubonic plague become increasingly toxic, remain febrile, and experience tachycardia, agitation, confusion, delirium, and convulsions.

Septicemic plague manifests as a rapidly progressive, overwhelming endotoxemia. Patients often complain of gastrointestinal (GI) symptoms, including nausea, vomiting, diarrhea, and abdominal plain. Disseminated intravascular coagulation can also occur with the appearance of petechiae, ecchymoses, bleeding, and ischemia in the tips of the extremities. Later-stage septicemic patients are likely to experience refractory hypotension, renal shutdown, obtundation, and other signs of shock. Patients with late-stage septicemic plague can also exhibit acute respiratory distress syndrome (Chapter 104), which has occasionally been confused with hantaviral pulmonary syndrome (Chapter 381) in the American Southwest. Because septicemic plague is likely to be fulminant and fatal, favorable outcomes depend on rapid diagnosis and prompt treatment with appropriate antimicrobials.

Pneumonic plague can be accompanied by fever, cough, chest discomfort that becomes increasingly painful, tachycardia, dyspnea, bacteria-laden sputum, chills, headache, achiness, weakness, and dizziness. As the illness progresses, patients may experience increasing respiratory distress, hemoptysis, cardiopulmonary insufficiency, and circulatory collapse. In the early stages of illness, patients with primary pneumonic plague may have signs of localized pulmonary involvement, beginning in a single lung and rapidly progressing to segmental consolidation and later bronchopneumonia, with death ensuing in as little as 1 to 3 days after the onset of symptoms. Localized infection is unlikely to be observed in the lungs of patients with secondary pneumonic plague because the lung tissues are infected initially through circulatory spread, which results in a diffuse interstitial pneumonitis. The appearance of sputum also differs in primary versus secondary pneumonic plague, being watery or mucoid, frothy, and perhaps tinged with blood in primary pneumonic cases but scanter, thicker, and more tenacious in secondary pneumonic cases.

DIAGNOSIS

Plague fatalities are typically related to delay in seeking treatment or misdiagnosis. The differential diagnosis of plague in its various clinical forms

includes staphylococcal (Chapter 288) or streptococcal adenitis (Chapter 290) or pneumonia (Chapter 97); cat-scratch disease (Chapter 315); tularemia (Chapter 311); chancroid (Chapter 301); acute filarial lymphadenitis (Chapter 358); mycobacterial infection (Chapter 325); septicemia caused by other bacteria; meningococcemia; bacterial endocarditis; mycoplasmal pneumonia and other community-acquired pneumonias; legionnaires' disease (Chapter 314); Q fever; influenza pneumonitis (Chapter 364); hantaviral pulmonary syndrome (Chapter 381); viral pneumonia caused by respiratory syncytial virus (Chapter 362), cytomegalovirus (Chapter 376) or other viruses; and strangulated inguinal hernia.

Laboratory confirmation relies on bacterial culture accompanied by specific bacteriophage lysis tests or detection of a fourfold rise in antibody titer to the *Y. pestis* F1 capsular antigen over a period of 2 to 4 weeks. Although the procedures for confirming the laboratory diagnosis of *Y. pestis* infection are relatively simple, delays can occur because the relevant expertise and reagents are often limited to a few public health or reference laboratories. Because cases of plague can rapidly progress to a life-threatening illness and because culture of *Y. pestis* on bacteriologic media may require 48 hours or longer before colonies become visible, it is essential that patients suspected of having plague receive appropriate antimicrobial therapy immediately after samples have been taken. Direct fluorescent antibody assays can be used to identify *Y. pestis* bacteria in bubo aspirates, sputum samples, and tracheal washes; this procedure requires about 1 hour and provides strong presumptive evidence of infection. A presumptive diagnosis can also be obtained rapidly by detecting *Y. pestis* DNA in polymerase chain reaction (PCR) assays or *Y. pestis*–specific antigens in immunologic assays, but these tests are not widely available. Recently, rapid immunochromatographic tests have been developed that can detect plague antigen or antibodies. The potential advantages of these tests are that they require few laboratory resources, can be done in field settings, and yield results in less than 1 hour. One such test designed to detect *Y. pestis* antigens in patient samples has been used in some developing countries to diagnose plague, but these assays are still being evaluated and are not widely available. Preferred laboratory samples include blood, serum, bubo aspirates, tracheal washes, and swabs of skin lesions or pharyngeal mucosa. Cerebrospinal fluid can also be collected from patients in whom plague meningitis is suspected.

PREVENTION

No commercially available plague vaccine exists in the United States. Newer recombinant vaccines designed to stimulate immune responses to the F1 and V antigens of *Y. pestis* are undergoing clinical trials and have yielded promising results, but they are unlikely to be available in the near future.

Human plague risk can be reduced through the implementation of effective surveillance programs to rapidly identify human cases and threatening epizootics and allow effective intervention measures to be implemented. When appropriate, affected areas can be treated with insecticides to reduce the risk for flea bite exposure. Measures also should be taken to reduce the amount of food and shelter available to rodents. In rare instances, rodenticides can be used to reduce host numbers, but their use is not recommended before the implementation of effective flea control measures. Persons living in or visiting areas endemic for plague should avoid sick and dead animals and use insect repellents to reduce the risk for infectious flea bites. Dogs and cats should be prevented from roaming freely in rodent-infested areas, and these animals should be treated with flea control agents that safely and effectively kill fleas. Although most dogs appear to be somewhat resistant to plague, cats are highly susceptible, often experience severe illness, and can serve as sources of infection for their owners or veterinary staff. Cats that roam outside in endemic areas and suddenly become seriously ill should be taken to a veterinarian for evaluation.

Persons with pneumonic plague should be held in respiratory isolation for at least 48 hours after the initiation of appropriate antimicrobial therapy, and persons caring for such patients should follow respiratory droplet precautions (masks, gloves, gowns, and eye protection). Although human-to-human transmission has not been reported in the United States for many decades, rare cases of primary pneumonic plague have occurred after face-to-face contact with infected cats that had oral lesions or symptoms of pneumonic plague. Veterinary staff from plague-enzootic areas should take appropriate precautions (masks, gloves, gowns, eye protection) when handling sick cats with illnesses suggestive of plague.

Prophylactic antimicrobial therapy is generally recommended for persons with possible plague exposure only in relatively high-risk situations, such as after close contact with patients with pneumonic plague, handling of *Y. pestis*–infected animals, or being bitten by rodent fleas in an area with a recent history of epizootic activity. The most recently recommended prophylactic antimicrobials are doxycycline and ciprofloxacin (Table 312-1).

TREATMENT

Rx

The most commonly recommended agent for treating plague is streptomycin (see Table 312-1), but a randomized trial in Tanzania concluded that gentamicin and doxycycline are effective for treating plague in adults and children and cause few adverse responses.[■] A review of human cases treated in New Mexico also strongly suggests that gentamicin is effective and can be substituted for streptomycin. Tetracyclines are effective for treating uncomplicated cases of bubonic plague, and chloramphenicol is believed to be effective, particularly for plague meningitis. Although antimicrobial resistance is not believed to be a problem in the United States, strains resistant to tetracyclines and other agents have been described occasionally, and resistance to streptomycin and chloramphenicol may also occur. Recently, the Food and Drug Administration approved levofloxacin (Levaquin) to treat patients with plague under the agency's Animal Efficacy Rule, which allows evidence from animal studies, in this case those involving nonhuman primates, to be used to demonstrate the efficacy of a proposed treatment when it is not possible to conduct adequate trials in humans.

PROGNOSIS

Patients with uncomplicated bubonic plague respond quickly to appropriate antimicrobial therapy and typically defervesce, with relief from most other systemic manifestations within 2 to 5 days. Large buboes can remain swollen, however, for more than 1 week and may require incision and drainage if necrotic. In rare instances, ischemic necrosis in septicemic cases has resulted in the amputation of digits.

ENTEROPATHOGENIC YERSINIAE

DEFINITION

Y. enterocolitica and *Y. pseudotuberculosis* differ significantly from *Y. pestis* in that they are enteropathogenic, rarely cause death, and are spread via the fecal–oral route. Similar to *Y. pestis,* both of these bacteria are gram-negative members of the family Enterobacteriaceae.

The Pathogens

Y. enterocolitica is genetically quite distinct from *Y. pseudotuberculosis* and *Y. pestis;* the plague bacterium is thought to have arisen as a recent clone of *Y. pseudotuberculosis,* and the chromosomal DNA of these two species is extremely similar but markedly different from that of *Y. eneterocolitica.*[8] All three species, however, share an approximately 70-kb plasmid that encodes for various proteins (Yops) that are key virulence factors. Unlike *Y. pestis,* the enteropathogenic yersiniae are urease positive and motile at temperatures lower than 30° C. *Y. pseudotuberculosis* is rhamnose positive, thereby distinguishing it from the closely related *Y. pestis* and the more distantly related *Y. enterocolitica.* The enteropathogenic yersiniae are genetically more diverse than the more recently evolved plague bacterium. *Y. enterocolitica* contains six biogroups, five of which are known to be pathogenic for humans, and nearly 60 serogroups; *Y. pseudotuberculosis* has been classified into six distinct serogroups (O groups 1-6).

EPIDEMIOLOGY

Enteropathogenic yersiniae are transmitted via the fecal–oral route, and successful infection of the host requires large doses of bacteria (median infective dose of 10^8 to 10^9 bacteria). Typical sources of infection include the consumption of contaminated foods such as dairy products, inadequately cooked pork, and certain vegetables. Both *Y. enterocolitica* and *Y. pseudotuberculosis* can survive and proliferate slowly at refrigerator temperatures. Although far less common, person-to-person transmission has been reported, as has transmission via blood transfusion. Pigs, rodents, rabbits, sheep, goats, cattle, horses, dogs, cats, and sometimes birds serve as reservoirs for these yersiniae. Most hosts act as asymptomatic carriers, but a few human cases have been associated with handling sick animals. Symptomatic patients shed large amounts of yersiniae for as long as 2 to 3 weeks. Untreated, infected persons can become carriers and shed for 2 to 3 months.

Yersiniosis is a reportable disease in many countries. Most cases result from *Y. enterocolitica* infection, and this agent reportedly accounts for 1% to

TABLE 312-1 RECOMMENDATIONS FOR THE TREATMENT OF PATIENTS WITH PNEUMONIC PLAGUE IN CONTAINED AND MASS CASUALTY SETTINGS AND FOR POSTEXPOSURE PROPHYLAXIS*

PATIENT CATEGORY	RECOMMENDED THERAPY
CONTAINED CASUALTY SETTING	
Adults	Preferred choices: Streptomycin 1 g IM twice daily Gentamicin 5 mg/kg IM or IV once daily or 2 mg/kg loading dose followed by 1.7 mg/kg IM or IV 3 times daily Alternative choices: Doxycycline 100 mg IV twice daily or 200 mg IV once daily Ciprofloxacin 400 mg IV twice daily† Chloramphenicol 25 mg/kg IV 4 times daily‡
Children§	Preferred choices: Streptomycin 15 mg/kg IM twice daily (maximum daily dose, 2 g) Gentamicin 2.5 mg/kg IM or IV 3 times daily‖ Alternative choices: Doxycycline: if ≥45 kg, give adult dosage; if <45 kg, give 2.2 mg/kg IV twice daily (maximum, 200 mg/day) Ciprofloxacin 15 mg/kg IV twice daily† Chloramphenicol 25 mg/kg IV 4 times daily‡
Pregnant women¶	Preferred choice: Gentamicin 5 mg/kg IM or IV once daily or 2 mg/kg loading dose followed by 1.7 mg/kg IM or IV 3 times daily‖ Alternative choices: Doxycycline 100 mg IV twice daily or 200 mg IV once daily Ciprofloxacin 400 mg IV twice daily†
MASS CASUALTY SETTING AND POSTEXPOSURE PROPHYLAXIS**	
Adults	Preferred choices: Doxycycline 100 mg PO twice daily†† Ciprofloxacin 500 mg PO twice daily† Alternative choice: Chloramphenicol 25 mg/kg PO 4 times daily‡‡
Children§	Preferred choices: Doxycycline††: if ≥45 kg, give adult dosage; if <45 kg, give 2.2 mg/kg PO twice daily Ciprofloxacin 20 mg/kg PO twice daily Alternative choice: Chloramphenicol 25 mg/kg PO 4 times daily‡‡
Pregnant women¶	Preferred choices: Doxycycline 100 mg PO twice daily†† Ciprofloxacin 500 mg PO twice daily Alternative choice: Chloramphenicol 25 mg/kg PO 4 times daily‡‡

*These are consensus recommendations of the Working Group on Civilian Biodefense and are not necessarily approved by the U.S. Food and Drug Administration. In general, one antimicrobial agent should be selected, and therapy should be continued for 10 days. Oral therapy should be substituted when the patient's condition improves. Although animal and limited human studies support their use, gentamicin and ciprofloxacin are not currently approved by the U.S. Food and Drug Administration (FDA) for treatment of plague in humans. The related drugs, streptomycin and levofloxacin, are FDA approved for treatment of plague.
†Other fluoroquinolones can be substituted at doses appropriate for age. The ciprofloxacin dosage should not exceed 1 g/day in children.
‡The concentration should be maintained between 5 and 20 μg/mL. Concentrations >25 μg/mL can cause reversible bone marrow suppression.
§In children, the ciprofloxacin dose should not exceed 1 g/day, and the chloramphenicol dose should not exceed 4 g/day. Children younger than 2 years should not receive chloramphenicol.
‖Aminoglycoside dosages must be adjusted based on renal function. Evidence suggests that gentamicin 5 mg/kg intravenously or intramuscularly once daily would be efficacious in children, although this is not yet widely accepted in clinical practice. Neonates up to 1 week of age and premature infants should receive gentamicin 2.5 mg/kg intravenously twice daily.
¶In neonates, a gentamicin loading dose of 4 mg/kg should be given initially.
**The duration of treatment of plague in a mass casualty setting is 10 days. The duration of postexposure prophylaxis to prevent plague infection is 7 days.
††Tetracycline can be substituted for doxycycline.
‡‡Children younger than 2 years should not receive chloramphenicol. An oral formulation is available only outside the United States.
IM, Intramuscular; *IV,* intravenous; *PO,* oral.
Adapted from Inglesby TV, Dennis DT, Henderson DA, et al. Plague as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. *JAMA.* 2000;283:2281-2290.

3% of all cases of acute enteritis in some areas. From 1996 to 2009, the overall annual incidence in the United States was about 0.5 cases per 100,000 persons, with higher rates among children and African Americans.[9] The incidence has decreased over the past decade, especially among African American children younger than 5 years of age. Nevertheless, rates remain markedly higher among African American and Asian American children, at 3.5 and 5.1 cases per 100,000, respectively. In the past, high rates among African American populations were associated with the home preparation of chitterlings made from contaminated pork intestines. Exposure of contaminated pork products was also considered a likely source of infection for Asian Americans.

Y. enterocolitica serotypes O:3, O:8, O:9, and O:5,27 have been associated with human disease. Serotype O:3 (biotype 4) predominates in most countries and is found most commonly in swine. Serotype O:9 (biotype 2) has been isolated from sheep, cows, and goats. These last two serotypes are the most common causes of human infection worldwide but are considered less pathogenic than strains of the O:8 (biotype 1B) serotype, which are associated with the most severe outbreaks. Although serotype O:8 infections appear to be decreasing in incidence in the United States, they are becoming increasingly important in Japan, Italy, and France. In Europe, most cases involve serotype O:3 infections, although a few are associated with serotypes O:9 and O:5,27. Biotype 1A appears to be nonpathogenic, and biotype 5 has been isolated only from hares. Strains pathogenic in humans are esculin, salicin, and pyrazinamidase negative.

Y. pseudotuberculosis has been isolated from rodents, cattle, sheep, cats, dogs, and birds.[10] It occurs worldwide but human infections are most commonly reported in northern Europe and in Asia, including Japan. O group 1, 2, and 3 strains are associated with human disease, with most cases being attributed to O:1 or O:2 strains. Infected animals serve as chronic carriers and sources for the infection of water and foods such as meat, dairy products, and stored vegetables. Some cases have been associated with the handling of kittens and puppies. Recent outbreaks of *Y. pseudotuberculosis* in Finland were traced to eating iceberg lettuce and carrots (serotypes O:3 and O:1, respectively). A large Canadian outbreak was associated with milk that had been pasteurized but nevertheless became contaminated. Outbreaks in the former Soviet Union have been associated with the consumption of root vegetables that were stored underground for winter consumption and presumably became contaminated with rodent excreta containing *Y. pseudotuberculosis.*

PATHOBIOLOGY

Y. enterocolitica possesses numerous virulence factors responsible for its persistence in the GI tract and ability to cause disease in susceptible hosts. Upon reaching the ileum, *Y. enterocolitica* adheres to the mucosa, where intracellular infections in Peyer's patches, mucosal cells, and macrophages can occur. Invasion of the ileal mucosa is favored by the presence of invasin, an outer membrane protein, and a 17-kD surface factor (Ail). An inflammatory response causes abdominal pain and diarrhea, as well as ulcerative ileitis, mesenteric adenitis, and necrosis within Peyer's patches. If the regional defenses are breached, the bacteria can disseminate and cause sepsis and hepatic and splenic abscesses. Polyarthritis can also occur later in the course of illness, particularly in human leukocyte antigen (HLA)-B27–positive individuals.

Similar to *Y. pestis,* the enteropathogenic yersiniae attack host lymphoid tissues. Invasion of these tissues and resistance against host defenses depend on the possession of the approximately 70-kb plasmid that is shared by each of the three pathogenic yersiniae and bears genes encoding for various Yops and the so-called V antigen. The products of this plasmid work in concert to inhibit phagocytosis and reduce inflammation, thereby suppressing the host immune response and favoring persistence of these microbes.

When *Y. enterocolitica* and *Y. pseudotuberculosis* become established in the lymph nodes, they can resist phagocytosis, which allows them to replicate extracellularly and form aggregates in the nodes. The ability to acquire host iron can be a significant factor contributing to the virulence of *Y. enterocolitica* strains. O:8 serotypes, which are virulent for mice, have the *irp2* (iron repressible 2) gene that encodes for a pair of high-molecular-weight outer membrane proteins, but the mouse avirulent O:3 and O:9 serotypes lack *irp2.* O:3 and O:9 serotypes typically remain confined to the guts of their hosts, except when iron overload is present or patients are being treated with an iron chelator.

Patients with reactive arthritis (Chapter 265) are more likely to have fewer GI symptoms, lower T-cell proliferative responses to *Yersinia* antigens, lower initial immunoglobulin (Ig) M responses, higher and more persistent IgG and IgA responses, and increased levels of IgA with a secretory

component. *Yersinia*-specific antibody responses are also more likely to persist in patients with reactive arthritis than in those with uncomplicated GI disease. *Yersinia* antigens thought to contribute to reactive arthritis include Yops and released proteins, which stimulate host CD4 cells, and heat shock protein 60, which has been hypothesized to work in conjunction with other antigens to modulate the host immune response. Evidence exists that hosts can maintain chronic *Y. enterocolitica* infections for years after the initial infection, a factor that may induce the inflammation associated with reactive arthritis (Chapter 265).

CLINICAL MANIFESTATIONS

After an incubation period of about 3 to 7 days, a gastroenteritis typically develops that can be difficult to distinguish from *Salmonella* (Chapter 308) or *Campylobacter* (Chapter 303) gastroenteritis. The most common clinical syndromes associated with *Y. enterocolitica* infection are acute enteritis with fever, diarrhea, vomiting, right lower quadrant pain suggestive of appendicitis (Chapter 142), erythema nodosum, and reactive arthritis. Other common symptoms include associated pharyngitis, rash, joint pain, and headache. Stool examination reveals leukocytes or erythrocytes, and one fourth of patients experience bloody diarrhea.

Y. pseudotuberculosis infections most commonly manifest as enterocolitis, pharyngitis, and pseudoappendicitis.

Whereas enterocolitis is most likely to occur in young children, older children more frequently experience acute terminal ileitis, mesenteric adenitis, and systemic disease. Pseudoappendicitis has been reported in patients with mesenteric lymphadenitis. Sepsis is uncommon and is most likely to occur in persons with underlying conditions such as diabetes mellitus, cirrhosis, immunosuppression, older age, and hemochromatosis. Splenic abscesses, meningitis, or endocarditis can develop in septic patients, and the mortality rate can approach 50%. Erythema nodosum is identified in about one third of all patients and in 10% of adults.

DIAGNOSIS

Yersiniosis should be suspected in patients with abdominal pain and fever, especially if they live in high-incidence areas. The diagnosis is best accomplished by isolation of the bacterium from stool, blood, or other appropriate samples.[11] Recovery of *Y. enterocolitica* and *Y. pseudotuberculosis* from clinical and environmental samples can be greatly complicated by the presence of other bacteria that are likely to predominate. The preferred selective agar for isolation from stool specimens is CIN (cefsulodin–irgasan–novobiocin) agar prepared with relatively low cefsulodin concentrations. Isolation of yersiniae is also helped by culture at 25° to 30° C, which results in better growth than when cultures are kept at 35° C and favors growth of the more cold-tolerant yersiniae over other bacteria that require higher temperatures. After being obtained, isolates can be confirmed as enteropathogenic *Yersinia* by biochemical tests. Biotyping and serotyping, which are often available only in research or reference laboratories, can provide useful epidemiologic information. Isolation difficulties and the reported low sensitivity (about 10^3 to 10^6 colony-forming units per gram of sample) of current isolation techniques have led some to suggest that PCR or other DNA-based methods are likely to result in the more rapid and sensitive detection of these bacteria.[12]

Tube or microagglutination tests can identify antibodies to the pathogenic *Y. enterocolitica* serogroups O:3, O:9, O:5,27, and O:8, but cross-reactivity can be a problem, particularly between *Y. enterocolitica* O:9 and *Brucella* spp. PCR for O genotyping of *Y. pseudotuberculosis* could eventually replace traditional serotyping methods. Immunologically mediated *Yersinia* illnesses, including reactive arthritis, are associated with the production of IgA antibodies that can be detected by enzyme-linked immunosorbent assay (ELISA) or immunoblotting. IgG antibodies can persist for many years, but persistence of IgA antibodies for more than a few months could indicate a chronic *Yersinia* infection.

PREVENTION

Prevention depends on measures intended to protect persons from contact with contaminated environments, foods, and wastes. These include using proper sewage disposal methods, protecting water supplies from contamination with human or animal wastes, following appropriate procedures for animal husbandry and slaughtering, thoroughly cooking meats (especially pork), avoiding long-term storage of meats at temperatures higher than 39° F, and consuming only pasteurized milk. Individuals should thoroughly wash their hands after handling potentially contaminated pork or other foods. Persons with diarrhea should not work in food-handling areas, care for young children, or work with patients, and hospital staff should follow enteric precautions. Vaccines are not available.

TREATMENT

Rx

Antibiotics do not improve the course of uncomplicated enterocolitis or mesenteric adenitis, and the use of antimicrobial therapy is not generally recommended for intestinal forms of the disease. Such therapy is, however, recommended for immunocompromised patients, patients with septicemia, and those with systemic disease or extraintestinal foci of infection. It should be noted that in vitro susceptibility does not necessarily indicate efficacy in vivo.

Broad-spectrum cephalosporins, sometimes accompanied by aminoglycosides, have resulted in successful outcomes in patients with extraintestinal forms of yersiniosis, including septicemia. Ciprofloxacin, cefotaxime, and ceftriaxone are considered the most effective agents for treating *Y. enterocolitica* serogroup O:3 infection. In addition, *Y. enterocolitica* serogroup O:3 and O:9 isolates possess chromosomally determined β-lactamases that can confer resistance to ampicillin, carbenicillin, and cephalothin. Although serogroup O:8 strains, which produce type A β-lactamase, show resistance to the latter two agents, they are susceptible to ampicillin. It has yet to be determined whether antimicrobial therapy is useful for treating immunologically mediated forms of yersinial illness, including reactive arthritis.

PROGNOSIS

Cases of *Y. enterocolitica* enteritis are generally mild and self-limited after a 2- to 3-week course of illness. Sequelae most commonly include reactive arthritis and erythema nodosum.[13] Glomerulonephritis and myocarditis have also been reported, particularly with serogroup O:3, biotype 4, phage type 8 infections. Other sequelae can include endocarditis, pericarditis, and osteitis. *Y. enterocolitica*–induced reactive arthritis, which can appear 1 to 3 weeks after infection, is oligoarticular, asymmetrical, and peripheral; it occurs most frequently in the lower limbs and eventually resolves over a period of a few weeks to months. HLA-B27–positive patients are more likely to experience severe and prolonged arthritis.

Y. pseudotuberculosis infections are also generally mild and self-limited. Reported complications include erythema nodosum, iritis, reactive arthritis, and nephritis. Although infection of the bloodstream by *Y. pseudotuberculosis* is rare, one review of 72 such cases reported that 26 (36%) were fatal.

Grade A

Grade A Reference

A1. Mwengee W, Butler T, Mgema S, et al. Treatment of plague with gentamicin or doxycycline in a randomized trial in Tanzania. *Clin Infect Dis*. 2006;42:614-621.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Wagner DM, Klunk J, Harbeck M, et al. Yersinia pestis and the plague of Justinian 541-543 AD: a genomic analysis. *Lancet Infect Dis*. 2014;14:319-326.
2. Morelli G, Song Y, Mazzoni CJ, et al. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet*. 2010;42:1140-1143.
3. Butler T. Plague gives surprises in the first decade of the 21st century in the United States and worldwide. *Am J Trop Med Hyg*. 2013;89:788-793.
4. Richard V, Riehm JM, Herindrainy P, et al. Pneumonic plague outbreak, northern madagascar, 2011. *Emerg Infect Dis*. 2015;21:8-15.
5. Hinckley AF, Biggerstaff BJ, Griffith KS, et al. Transmission dynamics of primary pneumonic plague in the USA. *Epidemiol Infect*. 2012;140:554-560.
6. Kugeler KJ, Staples JE, Hinckley AF, et al. Epidemiology of human plague in the United States, 1900-2012. *Emerg Infect Dis*. 2015;21:16-22.
7. Williamson ED, Oyston PCF. The natural history and incidence of *Yersinia pestis* and prospects for vaccination. *J Med Microbiol*. 2012;61:911-918.
8. Fabrega A, Vila J. *Yersinia enterocolitica*: pathogenesis, virulence and antimicrobial resistance. *Enferm Infecc Microbiol Clin*. 2012;30:24-32.
9. Ong KL, Gould LH, Chen DL, et al. Changing epidemiology of Yersinia enterocolitica infections: markedly decreased rates in young black children, foodborne diseases active surveillance network (FoodNet), 1996-2009. *Clin Infect Dis*. 2012;54(S5):S385-S390.
10. Kaasch AJ, Dinter J, Goeser T, et al. Yersinia pseudotuberculosis bloodstream infection and septic arthritis: case report and review of the literature. *Infection*. 2012;40:185-190.
11. Pfeiffer ML, DuPont HL, Ochoa TJ. The patient presenting with acute dysentery—a systematic review. *J Infect*. 2012;64:374-386.
12. Van Lint P, De Witte E, De Henau H, et al. Evaluation of a real-time multiplex PCR for the simultaneous detection of Campylobacter jejuni, Salmonella spp., Shigella spp./EIEC, and Yersinia enterocolitica in fecal samples. *Eur J Clin Microbiol Infect Dis*. 2014;[Epub ahead of print].
13. Rosner BM, Werber D, Höhle M, et al. Clinical aspects and self-reported symptoms of sequelae of Yersinia enterocolitica infections in a population-based study, Germany 2009-2010. *BMC Infect Dis*. 2013;13:236.

REVIEW QUESTIONS

1. A 45 year-old New York City police officer is brought to the emergency department (ED) with a 2-day history of fever to 104° F (40° C) and worsening cough, dyspnea, confusion, nausea, and vomiting. The patient is intubated emergently in the ED because of respiratory distress. Chest radiographs demonstrate dense lower lobe infiltrates bilaterally. Sputum is blood tinged; Gram staining reveals numerous gram-negative bacilli with faint bipolar staining.

Appropriate actions at this time include all of following except

A. begin treatment with intravenous gentamicin and ciprofloxacin.
B. obtain blood for culture and sensitivity testing.
C. isolate the patient under airborne precautions.
D. alert infection control, public health, and law enforcement officials.
E. offer prophylaxis to ED personnel who intubated the patient.

Answer: C This scenario should raise serious concern for bioterrorism. Although naturally acquired infections can be treated with gentamicin alone, intentionally released strains may be engineered for resistance; therefore, use of multiple antimicrobials is advisable until sensitivities are known. Pneumonic plague is spread through droplets and can be prevented with droplet precautions; airborne precautions are unnecessary and likely to hinder patient care and management, especially if additional cases are identified. Although person-to-person transmission of plague has not been reported in the United States since 1925, unprotected close contact (<6 ft) is a risk factor, and prophylaxis with oral doxycycline is generally recommend.

2. A 34 year-old Los Angeles woman is hospitalized for fever and septic shock. On hospital day 2, admission blood cultures grow *Y. pestis*. On more careful reexamination, a large painful bubo is discovered in the patient's right axilla. The diagnosis of bubonic plague had not been suspected because the patient did not report traveling outside her urban Los Angeles neighborhood. Further interviews reveal that 3 days before illness onset, the patient had butchered a wild rabbit hunted by her husband and transported it to her home. In addition to handing tissues of infected wild animals, all of the following have been identified as risk factors for plague in the United States except

A. living west of the 100th meridian.
B. allowing dogs to sleep on the bed.
C. contact with an ill cat.
D. rodent die off in the area around the home.
E. failure to use insect repellent.

Answer: E Plague is endemic in the western United States. For reasons not fully known, it has not been able to establish itself east of approximately the 100th meridian. Epidemiologic studies have identified allowing dogs to sleep on the bed as a risk factor, likely by increasing opportunities for exposure to infected fleas. Unlike dogs, which are relatively unaffected by plague, cats are highly susceptible and prone to infection through capture of rodents. Plague epizootics among rodents can lead to mass mortality and an abundance of infectious fleas seeking blood meals from other sources, including humans. Although it may be helpful, insect repellent has not been demonstrated to prevent human infection.

3. A previously healthy 55 year-old wildlife biologist presents to the emergency department in Santa Fe, New Mexico, with a 4-day history of fever, severe abdominal pain, and bloody stools. While being evaluated, the patient begins vomiting blood and develops acute respiratory distress. He is admitted but dies within a few hours. Two days later, blood cultures obtained at admission reveal an unidentified gram-negative rod. The isolate is sent to three different laboratories that use automated systems to identify the organism as *Acinetobacter lwoffi*, *Pseudomonas luteola,* and *Yersinia pseudotuberculosis*. The most likely explanation for this situation is

A. laboratory contamination.
B. the correct identification is *Y. pseudotuberculosis.*
C. the correct identification is *P. luteola.*
D. The correct identification is *A. lwoffi*.
E. none of the above.

Answer: E Epidemiologic and clinical features are strongly suggestive of septicemic plague and not consistent with any of these other agents. Because of a lack robust biochemical profiles for *Y. pestis*, some automated microbiologic identification systems can misidentify the organism, usually as *Y. pseudotuberculosis*.

4. Which of the following regarding primary pneumonic plague is *false*?

A. Can be acquired from infected cats
B. Has not resulted in person-to-person transmission in North America since 1925
C. Requires isolation of patients using droplet precautions
D. Occurs in about 3% of patients with bubonic plague
E. Is a likely presentation of an intentional release

Answer: D In North America, most cases of primary pneumonic plague are acquired from ill, infected cats. The last case of person-to-person transmission occurred in Los Angeles in 1925; however, cases acquired from cats continue to occur. Person-to-person transmission can be prevented with droplet precautions; airborne precautions are not necessary. Secondary, not primary, pneumonic plague occurs after bubonic plague. Although plague may be dispersed and acquired through various routes, an aerosol release resulting in primary pneumonic plague is considered the most likely and deadly.

5. Two previously healthy sibling boys, age 6 and 16 years, are brought to the clinic by their father. Both boys are complaining of fever (102° F) and abdominal pain. The 6-year-old boy reports diarrhea with bloody stools, verified by the father. The older boy denies diarrhea but has marked right lower quadrant abdominal pain with some rebound tenderness. When asked about dietary exposures, the father mentions that both boys had eaten poorly cooked pork at a winter solstice pig roast a few days earlier. Which of the following selective media is best for isolating the most likely cause of illness?

A. Sorbitol MacConkey agar (SMAC)
B. Xylose lysine deoxycholate agar (XLD)
C. Cefsulodin–irgasan–novobiocin(CIN)
D. Campylobacter blood agar plates (Campy-BAP)

Answer: C Although several pathogens can produce bloody diarrhea and abdominal pain, features of the scenario suggest infection with *Yersinia enterocolitica*. These include the association with undercooked pork, the mixed clinical features of bloody diarrhea in a younger child and pseudoappendicitis in an older child, and onset in winter. SMAC is the preferred media for isolating Shiga toxin *Escherichia coli* O157:H7, which classically cause bloody diarrhea with little or no fever. Both *Salmonella* and *Campylobacter* infections occur more often in the summer and fall, and both are generally associated with exposures to nonpork products, especially beef, eggs, produce, and poultry.

313

WHOOPING COUGH AND OTHER *BORDETELLA* INFECTIONS

ERIK L. HEWLETT

DEFINITION

Pertussis (whooping cough), an acute respiratory illness with a potentially protracted clinical course, is caused by infection with *Bordetella pertussis* or related members of the genus, especially *Bordetella parapertussis*. The infection is highly contagious and can have both endemic and epidemic features in a population, but there are no known chronic carriers of the organism. Pertussis is currently a serious public health problem worldwide, with

reported U.S. cases in 2012 the highest in more than 50 years.[1,2] The greatest risks of morbidity and mortality are in infants, but increases in incidence in recent years include adolescents and adults and are highest in the cohort of children who have only ever received acellular pertussis vaccines.[3,4]

The Pathogens

There are now eight species in the genus *Bordetella,* but *B. pertussis, B. parapertussis,* and *B. bronchiseptica* (primarily a veterinary pathogen) are most likely to cause respiratory infections in humans. Recently identified species such as *B. holmesii* and *B. trematum* have been associated with bacteremia, meningitis, and wound infections, and *B. holmesii* has been isolated concurrently with *B. pertussis* in an outbreak of respiratory illnesses.[5] The *Bordetellae* are gram-negative, coccobacilli, and particularly fastidious and small relative to other gram-negative organisms. *B. pertussis* was first isolated by Bordet and Gengou in 1906, and a medium used for culture (Bordet-Gengou agar) still bears their names.

EPIDEMIOLOGY

In the prevaccine era (before the mid-1950s), pertussis was primarily a childhood illness; as a result, the disease received little attention in adult medicine. The extensive use of whole-cell pertussis vaccines in the United States and other parts of the world resulted in a marked reduction of pertussis. In addition, however, the vaccine elicited a change in the age-specific incidence of disease, with the majority of cases subsequently occurring in infants too young to be vaccinated and in adolescents and adults, who become susceptible because of waning immunity after childhood vaccination.[6,7] In recent years, this shift has been particularly notable, with 48,277 cases of pertussis reported in the United States in 2012, approximately 50% of those occurring in patients 7 to 19 years old. In fact, seroepidemiologic data suggest that 25% of adolescents and adults with a cough lasting more than 1 week have pertussis. The disease is endemic in the United States, but epidemics also occur every 3 to 4 years.

B. pertussis is highly contagious and the presumption that transmission can occur by aerosol droplet has recently been documented for the first time in nonhuman primate studies. There is no classic "carrier" state such as that recognized for other infectious organisms. In recent studies, however, nonhuman primates (baboons) immunized with acellular pertussis vaccine and challenged directly with *B. pertussis* were without clinical signs of disease but were able to transmit *B pertussis* to susceptible cage mates.[8] Dissemination in a community is often caused by symptomatic (or now perhaps asymptomatic) adolescents, who spread the organism during school attendance and social events. Health care personnel are also important sources of transmission, and health care–associated outbreaks can affect both vulnerable patients and their care providers.

PATHOBIOLOGY

Although *B. pertussis* and related organisms produce a number of interesting toxins, and pertussis has been described as a "toxin-mediated" disease, there is no clear pathophysiologic link between the known virulence factors and clinical manifestations of disease.[9] The infection is localized to the respiratory tract, where organisms adhere to the ciliated surface of epithelial cells. Mucous secretion is prominent, especially during later stages of the illness. Intracellular *B. pertussis* organisms have been demonstrated both *in vitro* and in samples collected from patients, but it is not an obligate intracellular organism, and the significance of these observations remains unknown.

Toxins and adhesins are important both for their potential role in the pathogenesis of disease and for their use as antigens in acellular pertussis vaccines. Filamentous hemagglutinin (FHA), pertactin (PRN), and fimbriae are adhesins that are components of some acellular vaccines, and all currently available vaccines include chemically inactivated pertussis toxin (PT). In addition to their adhesin function, FHA, PRN, and fimbriae have been shown to modulate the functions of host immune-effector cells. PT is a member of the family of adenosine diphosphate–ribosylating toxins that includes cholera toxin and diphtheria toxin, and its intracellular targets consists of several guanosine triphosphate–binding proteins such as $G_{i\alpha}$. Although PT causes lymphocytosis and enhances insulin secretion, its target cell or cells and its contribution to the clinical manifestations of pertussis are still a mystery. *Bordetellae* also possess adenylate cyclase (AC) toxin, a bacterial AC that enters host cells to produce supraphysiologic levels of cyclic adenosine monophosphate (cAMP). AC toxin acts against phagocytic cells, such as polymorphonuclear leukocytes and alveolar macrophages, whose antibacterial functions are inhibited by the increased cAMP. Additional "toxins," including the tracheal cytotoxin, tracheal colonization factor, and heat-labile or dermonecrotic toxin, have been identified by virtue of their activities in vitro or in animals. The relationship of these toxins to disease also remains ill defined.

CLINICAL MANIFESTATIONS

Although the classic paroxysmal cough of pertussis is striking and unforgettable, not all patients experience this characteristic symptom. Infants may have apnea as their only manifestation of pertussis; in others, the characteristic cough develops subsequently. Existing clinical case definitions of pertussis are decades old and based largely on presentations in infants and children. Because of the increasing incidence in adolescents and adults, who may manifest distinct symptoms and signs, the Global Pertussis Initiative has developed an algorithm based on different age groups in order to encourage expanded use of laboratory testing in suspected cases.[10]

Incubation Period

After exposure of an individual to active pertussis (typically by way of aerosol from an infected patient who is coughing), the onset of symptoms is not experienced until 1 to almost 3 weeks later. This relatively long incubation period makes it difficult to track transmission and increases the time necessary for intervention and implementation of control measures.

Catarrhal Phase

Importantly, the symptoms that reflect onset of the catarrhal phase (rhinorrhea, increased lacrimation, conjunctival injection, and sometimes low-grade fever) are nonspecific and are not suggestive of pertussis, except in the setting of an ongoing hospital or community outbreak. This phase, which can last from a few days to as long as 1 week, is often associated with a nonproductive cough.

Paroxysmal Phase

A patient's transition to the paroxysmal phase and initiation of the typical cough can suggest the diagnosis of pertussis. This striking cough consists of a series of uncontrollable expirations followed by gasping inhalation, which, depending on the airway anatomy, can result in the characteristic "whooping" sound. The whoop is more frequent in children than in adults. It is common for each episode of coughing to be associated with cyanosis and end with gagging and vomiting, which in infants can result in dehydration and malnutrition. The paroxysmal stage can last up to 4 weeks, and the development of fever or worsening pulmonary function during this time suggests the possibility of secondary pneumonia. Eighty percent or more of adolescents and adults with pertussis have paroxysmal cough, but the frequency of whoop and posttussive vomiting in this population is quite variable. In addition, during the paroxysmal stage, adults experience symptoms, such as a scratchy throat, other pharyngeal symptoms, and episodes of sweating, which are not described in children.

Convalescent Phase

A reduction in the frequency and severity of coughing attacks marks the transition to the convalescent phase, which can last weeks to months. It is often during this time that adults present to the health care system with "chronic cough" and may be evaluated for conditions such as asthma, tuberculosis, other chronic lung diseases, malignancies, and gastroesophageal reflux. After the coughing spells have ended, patients may experience a return of the paroxysmal cough in conjunction with unrelated upper respiratory illnesses or stimuli or irritants; this phenomenon is often incorrectly interpreted as a recurrence of pertussis. Eighty percent of adults with pertussis have an illness of at least 3 weeks' duration, and 27% are still coughing after 90 days.

Complications

In infants and children younger than 5 years, hospitalization is common; 5% to 10% have pneumonia, nearly 1% of infants experience seizures, and encephalopathy is seen in 0.1%. Pertussis in infants and small children can be associated with pulmonary hypertension, now known to result from the elevated white blood cell (WBC) count, which can exceed 100,000/mm^3 in a naïve host. In one study, this phenomenon occurred in fewer than 15% of cases but was present in 75% of patients who died.[11] Adolescents and adults have these same complications with a lower frequency, but adults can experience other problems associated with underlying medical conditions. For example, there are anecdotal reports of cough syncope, herniated

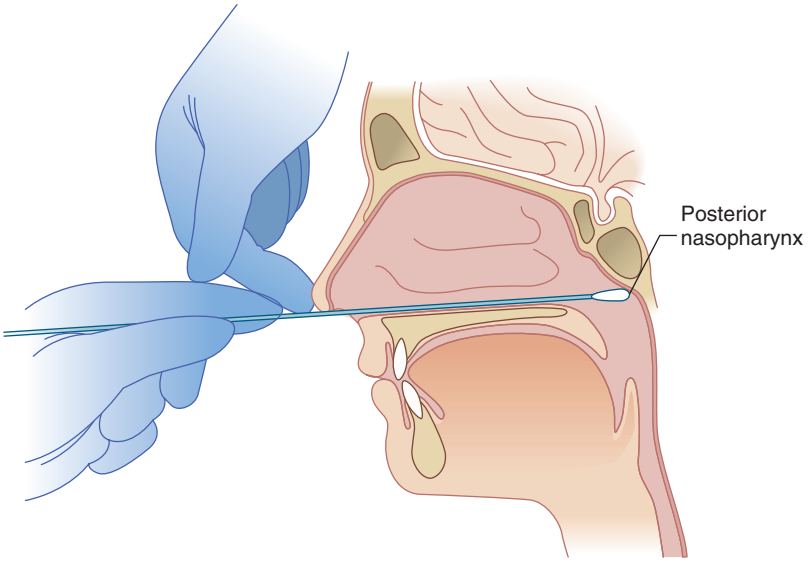

FIGURE 313-1. Technique for obtaining a nasopharyngeal specimen for the isolation of *Bordetella pertussis*. Dacron or calcium alginate swabs are recommended to obtain culture specimens. Dacron swabs are appropriate if polymerase chain reaction testing will be performed. (Redrawn from Cornia PB, Lipsky BA, Saint S, Gonzales R. Clinical problem solving: nothing to cough at. *N Engl J Med*. 2007;357:1432-1437.)

intervertebral disc, sudden-onset hearing loss, angina episodes, and carotid artery dissection.

DIAGNOSIS

Several methods can be used to detect *B. pertussis,* its products, and the host response to them, but each has its limitations. Culture of a nasopharyngeal swab or aspirate (Fig. 313-1) is the "gold standard" (specificity approaching 100% in symptomatic patients), but even with the use of specialized transport medium and processing in a careful, interested laboratory, recovery rates are often less than 50% and are completely dependent on the duration of illness and whether antimicrobial treatment was initiated before specimen collection. Polymerase chain reaction (PCR)–based diagnostic tests of nasopharyngeal swabs or aspirates are much more sensitive than culture, and results can remain positive for several days after the initiation of antimicrobial treatment. In addition, appropriately designed PCR assays can identify and distinguish among *B. pertussis, B. parapertussis,* and *B. bronchiseptica* and detect insertion sequence elements suggestive of other *Bordetella* species.[12] However, several apparent "outbreaks" of pertussis have later been found to represent false-positive PCR results. Detection of serum antibodies to products of *B. pertussis* can be used to identify patients during infection, but care must be taken to distinguish an acute response from residual antibodies elicited by prior immunization.

In view of the limited resources and equipment available for the diagnosis of pertussis, the World Health Organization has established a clinical case definition, which is 21 or more days of paroxysmal coughing with laboratory confirmation or epidemiologic linkage. Although this definition is often used for clinical trials, it is now clear that its application can result in cases of lesser severity or shorter duration being missed.

TREATMENT

Rx

Supportive Therapy

Because infants and young children have the highest risk for complications and death, supportive therapy is often the most important component of their medical care. Close observation (preferably in the hospital) is essential to ensure adequate feeding, oxygenation, and hydration to minimize complications in this age group. None of the pharmacologic interventions tested for the amelioration of cough and other symptoms is effective in pertussis. Leukophoresis has been used to reduce the WBC count and thus ameliorate pulmonary hypertension in infants and young children, but there is no agreement as to whether this approach affects mortality rates.

Antimicrobial Agents

There are two reasons to use antimicrobials in a patient with pertussis: (1) to eliminate the causative organism and reduce transmission and (2) to limit the course of illness in the treated patient. Because individuals can remain culture positive and can potentially transmit *B. pertussis* for several weeks after onset of symptoms, it is appropriate to treat patients who are seen within that time frame. Antimicrobials do not provide symptomatic relief or affect the course of the illness in the later stages of the infection, but they can do so when treatment is started within the first few days of symptoms. Although macrolide resistance is not a significant problem, it has been reported.

The recommendation of the Centers for Disease Control and Prevention for the treatment of pertussis in adults is azithromycin (500 mg on day 1 followed by 250 mg/day on days 2-5), clarithromycin (1 g/day in two divided doses for 7 days), erythromycin (2 g/day in four divided doses for 14 days), or trimethoprim–sulfamethoxazole (trimethoprim 320 mg/day, sulfamethoxazole 1600 mg/day, in two divided doses for 14 days). It has now been shown that treatment with erythromycin for 7 days is as effective as treatment for 14 days. The newer macrolides (azithromycin and clarithromycin) are better tolerated but more expensive.

PREVENTION

Chemoprophylaxis

Chemoprophylaxis with the aforementioned antimicrobial agents is an important mechanism for controlling outbreaks in hospitals or in the community. This approach is effective when initiated before the onset of symptoms and is recommended for individuals exposed within the preceding 3 weeks; high-risk persons with underlying health problems, particularly infants; or those who might have occupational or other contact with susceptible hosts.

Immunization

A killed, whole-cell vaccine was introduced in the late 1940s. Use of that product had a dramatic effect on the incidence of pertussis, with the number of reported cases in the United States falling from more than 200,000 annually to less than 2000 in 1980. In the 1970s and 1980s, recognition of the adverse events associated with the whole-cell product and public concern about the extent of those reactions led to the development of the purified acellular pertussis vaccines that are in use today.[13]

Current pediatric acellular pertussis vaccines (DTaP), which contain one or more purified protein antigens (all contain PT and other combinations of FHA, pertactin, and fimbriae types 2 and 3), are administered to infants and children up to the age of 6 years and cause significantly fewer adverse reactions than the whole-cell vaccines. In recognition of the significant incidence of pertussis in adolescents and adults and their role in transmitting the disease to infants and small children, Tdap (tetanus, diphtheria and pertussis) vaccines, which contain reduced quantities of the pertussis components, are now licensed for administration to these age groups. Their use is recommended to boost immunity in adolescents and adults who completed the recommended childhood vaccination series. Importantly, this adult booster immunization is now recommended for women during each pregnancy to prevent acquisition of *B. pertussis* and transmission to their neonates.

The increasing incidence of pertussis, including in immunized populations, appears to be attributable to a markedly reduced duration of protection relative to that anticipated from experience with whole-cell vaccines and prior infection. There is, however, an increasing number of *B. pertussis* isolates that do not express pertactin.[14] In addition, there is increasing prevalence of strains harboring the *ptxP3* allele, which is associated with increased production of PT and with alterations in expression of other virulence factors.[15,16]

In consideration of the adverse events and negative publicity associated with the whole-cell vaccines and the poor efficacy of the current acellular vaccines, there is active development of a live, attenuated pertussis vaccine, which would be administered intranasally.[17]

PROGNOSIS

Most patients will eventually clear *B. pertussis* even without antimicrobial treatment, but in most cases, the illness lasts weeks to months. Antimicrobials are of limited effectiveness in altering the course of the illness unless they are started before the paroxysmal phase. The introduction of booster doses of vaccines for adolescents and adults has not, however, resulted in the control of pertussis in well-immunized populations as had been anticipated. The next steps to address this serious public health problem are still being debated. Of major concern is the recent demonstration that baboons immunized with acellular pertussis vaccine can become infected and transmit *B. pertussis* in the absence of symptoms.

Grade A References

A1. Wang K, Bettiol S, Thompson MJ, et al. Symptomatic treatment of the cough in whooping cough. *Cochrane Database Syst Rev*. 2012;9:CD003257.

A2. Centers for Disease Control and Prevention. Recommended antimicrobial agents for the treatment and post-exposure prophylaxis of pertussis: 2005 CDC guidelines. *MMWR Recomm Rep*. 2005; 54(RR–14):1-16.

A3. Altunaiji SM, Kukuruzovic RH, Curtis NC, et al. Antibiotics for whooping cough (pertussis) [review]. *Cochrane Database Syst Rev*. 2007;3:CD004404.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Centers for Disease Control and Prevention. *Health Topic: Pertussis (Whooping Cough)*. http://www.cdc.gov/pertussis/outbreaks/trends.html. Accessed February 10, 2015.
2. Clark TA. Changing pertussis epidemiology: everything is new again. *J Infect Dis*. 2014;209:978-981.
3. Klein NP, Bartlett J, Rowhani-Rahbar A, et al. Waning protection after fifth dose of acellular pertussis vaccine in children. *N Engl J Med*. 2012;367:1012-1019.
4. Witt MA, Arias L, Katz PH, et al. Reduced risk of pertussis among persons ever vaccinated with whole cell pertussis vaccine compared to recipients of acellular pertussis vaccines in a large US cohort. *Clin Infect Dis*. 2013;56:1248-1254.
5. Rodgers L, Martin SW, Cohn A, et al. Epidemiologic and laboratory features of a large outbreak of pertussis-like illnesses associated with cocirculation *Bordetella homlesii* and *Bordetella pertussis*-Ohio, 2010-2011. *Clin Infect Dis*. 2013;56:322-331.
6. Guiso N. *Bordetella pertussis:* why is it still circulating? *J Infect*. 2014;68:S119-S124.
7. Edwards KM. Unraveling the challenges of pertussis [commentary]. *Proc Natl Acad Sci U S A*. 2014;111:575-576.
8. Warfel JM, Zimmerman LI, Merkel TJ. Acellular pertussis vaccines protect against disease but fail to prevent infection and transmission in a nonhuman primate model. *Proc Nat Acad Sci U S A*. 2014;111:787-792.
9. Hewlett EL, Burns DL, Cotter PA, et al. Pertussis pathogenesis—what we know and what we don't know. *J Infect Dis*. 2014;209:982-985.
10. Cherry JD, Tan T, Wirsing von König CH, et al. Clinical definitions of pertussis: summary of a global pertussis initiative roundtable meeting, February 2011. *Clin Infect Dis*. 2012;54:1756-1764.
11. Berger JT, Carcillo JA, Shamley TP, et al. Critical pertussis illness in children: a multicenter prospective cohort study. *Pediatr Crit Care Med*. 2013;14:356-365.
12. Leber AL. Pertussis: relevant species and diagnostic update. *Clin Lab Med*. 2014;34:237-255.
13. Lambert LC. Pertussis vaccine trials in the 1990s. *J Infect Dis*. 2014;209(suppl 1):S4-S9.
14. Pawloski LC, Queenan AM, Cassiday PK, et al. Prevalence and molecular characterization of pertactin-deficient Bordetella pertussis in the United States. *Clin Vaccine Immunol*. 2014;21:119-125.
15. King AJ, van der Lee S, Mohangoo A, et al. Genome-wide gene expression analysis of Bordetella pertussis isolates associated with a resurgence in pertussis: elucidation of factors involved in the increased fitness of epidemic strains. *PLoS ONE*. 2013;8:e66150.
16. Hegerle N, Paris AS, Brun D, et al. Evolution of French Bordetella pertussis and Bordetella parapertussis isolates: increase of Bordetellae not expressing pertactin. *Clin Microbiol Infect*. 2012;18:E340-E346.
17. Thorstensson R, Trollfors B, Al-Tawil N, et al. A Phase 1 clinical study of a live attenuated Bordetella pertussis vaccine-BPZE1; a single centre, double-blind, placebo-controlled, dose-escalating study of BPZE1 given intranasally to healthy adult male volunteers. *PLoS ONE*. 2014;e83449.

REVIEW QUESTIONS

1. The leading or most logical explanation for the current resurgence of pertussis in the United States is

A. limited efficacy and duration of protection elicited by current acellular pertussis vaccines.
B. prolonged, asymptomatic carriage of *B. pertussis* after infection.
C. enhanced detection of *B. pertussis* with the use of polymerase chain reaction (PCR) technologies.
D. development of antibiotic resistance among circulating strains of *B. pertussis*.
E. adaptation of *B. pertussis* to selective pressure imposed by acellular vaccines containing only four or five protein antigens.

Answer: A The limited protection induced by acellular pertussis vaccines and the rapid decrease in that protection with time have been documented in numerous studies. Although animal recipients of acellular vaccine can become infected and transmit to other animals without developing signs, such as cough, there is no evidence for a prolonged carrier state such as that documented for other infectious diseases. The sensitivity of PCR for identification of *B. pertussis* is greater than that of classical methods such as culture. PCR has, however, been used for a number of years and cannot explain the striking increases observed from 2010 to 2012 and the current incidence in the United States, which is the highest in more than 50 years. Antibiotic resistance among *B. pertussis* strains has been reported but is rare and not clinically relevant at present. Finally, there is an increasing number of clinical isolates of *B. pertussis* that do not express pertactin, one of the antigens in the cellular vaccine, but these existed before use of acellular vaccines, and there is no evidence that they are appearing as a result of acellular vaccine use.

2. The characteristic paroxysmal cough associated with *B. pertussis* infection is caused by

A. tracheal cytotoxin (TCT).
B. pertussis toxin (PT).
C. filamentous hemagglutinin (FHA).
D. pertactin (PRN).
E. none of the above.

Answer: E Although is has been speculated that several of the listed virulence factors could be responsible for the cough of whooping cough, none of these molecules alone can cause the cough in experimental animals (or, in some cases, in humans). It is unknown how infection with *B. pertussis* results in paroxysmal cough associated with whooping.

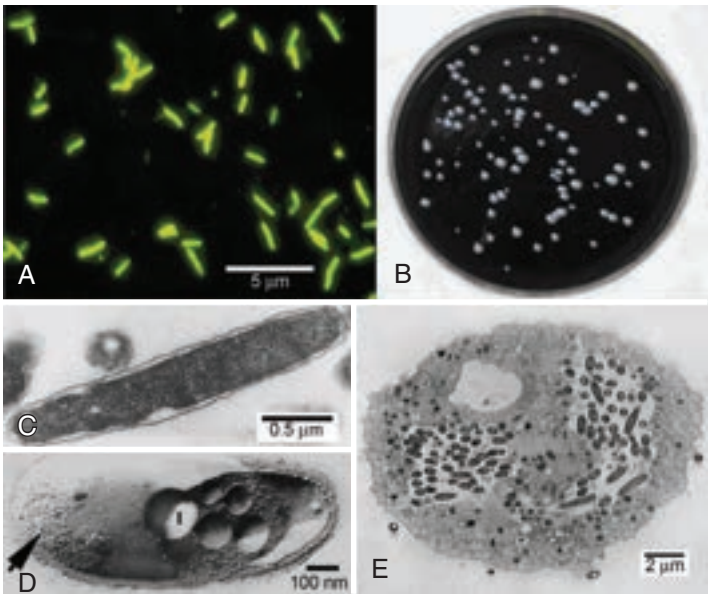

FIGURE 314-1. **A,** Direct fluorescent antibody stain of *Legionella pneumophila*. **B,** Colonies of *L. pneumophila* growing on a BCYE plate. **C,** Thin section of replicative form of *L. pneumophila*. **D,** Freeze-fracture replica of the mature infectious form obtained after growth in ameba. The latter shows prominent cytoplasmic inclusions (I) and a polar distribution of membrane proteins (*arrow*). **E,** Electron micrograph of *L. pneumophila* growing within *Acanthamoeba castellani*. (**A** Courtesy Dr. Paul Hoffman, University of Virginia. **B** courtesy Dr. Sharon Berk, Tennessee Technical University. **C** to **E** courtesy Drs. Rafel Garduno and Gary Faulkner, Dalhousie University.)

314

LEGIONELLA INFECTIONS

THOMAS J. MARRIE

DEFINITION

Legionellosis is the term for infection due to bacteria in the *Legionella* genus, of which there are two main manifestations—pneumonia (legionnaires' disease) and Pontiac fever (named after Pontiac, Michigan, where it was first recognized). Pontiac fever is usually a mild febrile illness presumed to be a reaction to lipopolysaccharide of *Legionella* species.

The 58th annual convention of the American Legion was held at a hotel in Philadelphia from July 21 to July 24, 1976. Subsequently, 182 of the delegates became ill (hence the name legionnaires' disease), and 146 were hospitalized at 87 institutions across the United States. Most had radiographic evidence of pneumonia, and 29 (16%) died. Within about 6 months of the outbreak, a new microorganism, *Legionella pneumophila*, was isolated from the pulmonary tissue of some of those who had died in the Philadelphia outbreak. In retrospect, the organism had first been isolated in 1943.

The Pathogen

Legionellae are small, gram-negative, aerobic, non–spore-forming bacilli that measure 0.3 to 0.9 µm wide by 2 to 20 µm long (Fig. 314-1A to D). These organisms require special media for growth, and many laboratories are unable to isolate legionellae; thus, when laboratory expertise is uncertain, a negative culture is meaningless. They usually do not stain with Gram stain. In tissue specimens, Dieterle or Warthin-Starry stain is used to visualize these organisms. *Legionella micdadei* retains the modified acid-fast stain and can appear as acid-fast bacilli in clinical specimens. Legionellae are aquatic organisms that thrive in both natural and man-made waterways and distribution systems, especially hot-water pipes, water heaters (electrical more so than gas heaters), cooling towers, and water fountains. In these systems, they are found in biofilms that help confer resistance to biocides and chlorine. They are also found in moist soil and mud. They can survive in these environments for long periods and can tolerate temperatures of 0° to 68° C and a pH range of 5 to 8.5. In their natural environments, they are intracellular parasites of protozoa, such as the freshwater amebae *Acanthamoeba* and *Hartmannella* species (Fig. 314-1E). Indeed, legionellae can live in at least 20 species of ameba, two species of ciliated protozoa, and one species of slime mold. Free-living legionellae in biofilms are inactivated within a few weeks, whereas those residing in amebae survive for 6 months or more. Humans are accidental hosts who become infected by inhaling *Legionella* bacteria or amebae laden with these bacteria. Genomes of four different strains of *L. pneumophila* have been sequenced, and they range in size from 3.3 to 3.5 Mb. Legionellae have many eukaryotic-like proteins, which may help the intracellular growth of these organisms in human macrophages by mimicking host proteins. *Legionella* species have type I, II, IV, and V secretion systems, which allow the efficient and rapid delivery of molecules into the phagocytic host cell.

The number of new *Legionella* species continues to grow; there are now at least 56 *Legionella* species and 73 serogroups or more, including 15 serogroups of *L. pneumophila*. A variety of typing systems can be used to further refine the isolates so that individual strains can be identified, a factor that is important for determining the source of an outbreak or understanding the difference between environmental and clinical isolates. It is noteworthy that *L. pneumophila* and all its serotypes cause disease in humans (with serotype 1 predominant), whereas only about 50% of the remaining species of *Legionella* cause such disease.[1]

Legionella longbeachae was first isolated in 1980 from a patient with pneumonia in Long Beach, California. In Australia, New Zealand, and Japan, reported cases of *L. longbeachae* infection occur as often as cases of *L. pneumophila* infection. This organism is rarely isolated from aquatic environments. The primary environmental reservoir remains unknown, but the major source of human infection is considered to be commercial potting mixes and other decomposing material, such as bark and sawdust. The *L. longbeachae* genome encodes for a range of proteins that might assist in the degradation of plant material. These enzyme systems are not present in *L. pneumophila*.

EPIDEMIOLOGY

The incubation period for legionnaires' disease (LD) is most commonly cited as 2 to 10 days, with extremes of 1 to 28 days. Person-to-person transmission

TABLE 314-1 RISK FACTORS FOR LEGIONNAIRES' DISEASE

RISK FACTOR	APPROXIMATE INCREASED RISK (-FOLD) OVER PERSONS WITHOUT THIS RISK FACTOR
HOST FACTORS	
Renal failure requiring dialysis	20
Corticosteroid therapy	5-10
Hairy cell leukemia	20
Lung or hematologic malignant neoplasm	7-20
Cytotoxic chemotherapy	5
>3 Alcoholic drinks/day	3-4
Cigarette smoking	2-10
Age older than 50 years	2
Diabetes mellitus	2
Solid organ transplantation (immunosuppression)	2
Anti–tumor necrosis factor treatment	16-21
Chronic heart or lung disease	>1
Splenectomy (non-*pneumophila* strains only)	—
ENVIRONMENTAL FACTORS	
Travel	2
Recent plumbing work in home or at work	2
Hospitalization	—
Exposure to contaminated water sources—cooling towers, hot tubs, decorative fountains	—
Exposure to potting soil (Australia) for *Legionella longbeachae*	—

does not occur. Legionellosis is found worldwide, predominantly in developed countries owing to the frequent use of cooling towers and complex plumbing systems. Underdiagnosis may be a feature in developing countries because of the laboratory facilities required. In recent years, there has been an increase in *Legionella* cases in Japan from 56 cases in 1999 to 804 cases in 2011, reaching a rate of 1.15 per 100,000 population; in Europe, the rate in 2010 was 1.25 cases per 100,000, and in the United States, the rate increased from 0.39 to 1.15 per 100,000. From 1990 to 2005, 23,076 cases of LD were reported in the United States. Only 1.7% of these cases occurred in children, whereas 63% occurred in those aged 45 to 64 years. Males accounted for 61% of cases, and rates were highest in the eastern United States, where most cases occur in the summer or fall. More recent U.S. data show an annual increase of 217% in the number of cases from 2000 to 2009, with the highest rates in the Mid-Atlantic states at 2.60 per 100,000. Twenty-four percent were travel related, and 81% of these cases involved domestic travel; noteworthy is that 5% involved cruise ship travel. The overall case-fatality rate was 8%. Eastern Canada also has higher rates of LD than the rest of that country. The epidemiology of LD in Europe is not dissimilar to that of the United States. The countries with rates of more than 2 per 100,000 were France, Denmark, Spain, Netherlands, and Italy. Twenty percent were travel associated, and the overall case-fatality rate was 11%.

There is an association between increased rainfall and cases of LD. The connection may be aerosolization of *Legionella* from rain puddles on roads. In one study, 33 (47.8%) of puddle water samples were positive for *Legionella*, yielding 325 isolates. Among the 14 sequence types of the clinical isolates, 4 were present in puddle water isolates.

The most common risk factors for the acquisition of LD are listed in Table 314-1. *Legionella* species account for 1 to 5% of community-acquired pneumonia requiring admission to a hospital. In some areas, *Legionella* accounts for about the same percentage of pneumonia treated on an ambulatory basis. *Legionella* infections can be sporadic or occur in outbreaks.[2] Outbreaks have been associated with exposure to a variety of aerosol-producing devices, including showers, a grocery store mist machine, cooling towers, whirlpool spas, decorative fountains, and evaporative condensers. Other water sources implicated in transmission of LD include water on trains, birthing pools, dental units, asphalt paving machines, and windscreen wiper fluid without added screen wash. LD can be acquired up to 10 to 11 km away from contaminated cooling towers. Aspiration of contaminated potable water by immunosuppressed patients is another mechanism by which *Legionella* is acquired. From 1985 to 2007, 14 of 2946 (0.5%) solid organ transplant recipients developed LD at one center in Barcelona; 5 (36%) were nosocomially acquired.

Exposure to contaminated potting soil is a risk factor for *L. longbeachae* infection in Australia and New Zealand.

Pontiac fever occurs predominantly in outbreaks with very high attack rates. *L. pneumophila, L. micdadei,* and *L. anisa* have been implicated in outbreaks of Pontiac fever. Among residents of nursing homes, Pontiac fever has been associated with *L. pneumophila* concentrations of greater than 10^4 colony-forming units/L in shower water. Those receiving corticosteroid therapy have a six-fold higher risk for development of Pontiac fever.

PATHOBIOLOGY

After being inhaled, legionellae are phagocytosed in the lungs by alveolar macrophages.[3] Only virulent strains of *Legionella* are capable of initiating organism-directed endocytosis when attachment to the alveolar macrophage through E-cadherin and β1 integrin receptors occurs. Legionellae abrogate phagosome-lysosome fusion and replicate in an endosome surrounded by the endoplasmic reticulum. Once it is intracellular, the bacteria-laden endosome recruits small vesicles, mitochondria, and ribosomes, and within 4 to 6 hours it becomes enveloped by the endoplasmic reticulum, thereby establishing the replicative endosome. After a latent period of about 12 hours, the bacteria start dividing. During this latent period, there is synthesis of up to 35 proteins and repression of 32 proteins. Iron must be available in the phagosome for growth. Growth continues in the macrophages for approximately 24 hours, at which time the macrophage disintegrates by apoptosis and the bacteria are released. The released bacteria are often phagocytosed by other macrophages, dendritic cells, and epithelial cells, perpetuating the infection. Cell-mediated immunity is necessary for recovery from *Legionella* infection. Production of type 1 interferons has a protective effect by promoting the activation of macrophages. Activated macrophages limit the intracellular replication of legionellae by downregulating the expression of their transferrin receptors and limiting the availability of iron to the bacteria. Toll-like receptors 2, 4, 5, and 9 are activated during infection with *L. pneumophila*. Interleukins 1α, 1β, 4, 6, 12, and 18 are detected during *Legionella* infection. Vaccination of guinea pigs with the purified major outer membrane protein OmpS protects against an LD_{100} lethal challenge, whereas immunization with purified heat shock protein 60 provides little protection. Polymorphonuclear leukocytes are present in abundance in the infected lung, but their role in clearing the infection is unclear. Bacteria can spread beyond the lung and cause metastatic infection. However, many extrapulmonary effects of LD, such as cerebellar ataxia and confusion, are not due to metastatic infection; they are presumably due to as yet unidentified toxins. The infected lung is consolidated, and there is usually no parenchymal damage once recovery occurs. Abscess formation and bronchiolitis obliterans or fibrosing alveolitis are occasionally seen.

The pathogenesis of Pontiac fever is unclear. The onset of illness occurs within 12 to 36 hours after the inhalation of, presumably, endotoxin. This period is too short for bacterial multiplication to cause the symptoms.

CLINICAL MANIFESTATIONS

Most of our knowledge of the clinical features of LD comes from studying patients who have been hospitalized with this illness, that is, those with the most severe manifestations. Fever (often high), malaise, and cough are present in most patients. Chills occur in about 75%, and dyspnea in just more than half the patients. Other features include myalgias, headache, chest pain, and diarrhea. The cough is nonproductive in 50% of patients; others have scant sputum production that is usually mucoid, rarely purulent, and very rarely bloody. There are no clinical features that distinguish individual patients with LD from those with pneumonia caused by other pathogens.[4-6] However, when patients with LD are compared with those with community-acquired pneumonia due to other agents, the patients with LD are more likely to have myalgias, headache, diarrhea, and a higher mean oral temperature at the time of presentation. They also present to the hospital sooner after the onset of symptoms, 4.7 days versus 7.7 days. When patients with LD were compared with patients with bacteremic pneumococcal pneumonia, the following features were associated with *Legionella* pneumonia: male sex (odds ratio [OR], 4.6), heavy drinking of alcohol (OR, 4.8), previous β-lactam therapy (OR, 19.9), axillary temperature higher than 39° C (OR, 10.3), myalgias (OR, 8.5), and gastrointestinal symptoms (OR, 3.5). Pleuritic chest pain and purulent sputum were less likely to be present. In a young, otherwise healthy person with rapidly progressive pneumonia (especially if the progression occurs in the setting of β-lactam therapy), LD should be strongly

suspected. Mental confusion is common, and on occasion, the presentation is dominated by extrapulmonary manifestations such as reactive arthritis, cerebellar ataxia, seizures, myoclonus, or encephalitis. Rarely, extrapulmonary infection, such as prosthetic valve endocarditis, sinusitis, dialysis shunt infection, or abscess formation, occurs.

Physical findings include fever, tachypnea, relative bradycardia, and initially only a few crackles on chest examination. Later, the findings of pulmonary consolidation are not uncommon. Abdominal examination is usually unremarkable. Rash as a manifestation of LD is very rare. Progression of the illness is not uncommon, even after the institution of antibiotic therapy. This is more likely in immunocompromised patients, who may require up to a week to respond to therapy. About half the patients with LD who require hospitalization have a complicated course.

Pontiac fever has an incubation period of about 36 hours. Fever, severe myalgia, headache, and extreme fatigue are the dominant manifestations. The illness is of short duration, lasting, on average, 3 days.

DIAGNOSIS

It is most important to have a high index of clinical suspicion that a patient might have LD. Routine laboratory test results are nonspecifically abnormal. Leukocytosis is common; leukopenia, thrombocytopenia, and disseminated intravascular coagulation also occur. Other laboratory abnormalities may include hyponatremia (in about half the patients, sometimes severe), hypophosphatemia (also common, occurring early and resolving within a few days of the initiation of treatment), mild liver function test abnormalities (except for alkaline phosphatase, which is occasionally very elevated), elevated creatine kinase (occasionally with rhabdomyolysis), microscopic hematuria, and mild proteinuria. High procalcitonin levels exceeding 1.5 are associated with a higher rate of admission to intensive care units and death. Combinations of findings may be suggestive of LD. These include high temperature, absence of sputum production, high lactate levels, increased C-reactive protein level, and low platelet counts.

There are a number of specific tests for the diagnosis of LD. Tests to detect *L. pneumophila* SG 1 antigen in urine are available commercially.[7] These are easy to use, but there is a false-negative rate of up to 26%. The sensitivity of the urinary antigen test in a review of published data was 0.74 (0.68 to 0.81), and the specificity was 0.991 (0.984 to 0.997). Rarely, the urinary antigen test result can remain positive for up to 1 year. Use of this test has allowed early diagnosis of LD because of the very short time required to do the test. This may be a factor in the lower mortality rates from LD compared with historic rates. Sputum culture has a low sensitivity but is 100% specific. It should be performed on all patients suspected of having LD. Serologic tests are not useful in the immediate management of a patient because of the long time (6 to 12 weeks) required to seroconvert; however, they do have a role in the work-up of outbreaks of LD. False-negative and false-positive serologic results do occur. A four-fold or greater increase in antibody titer between the acute and convalescent phase serum samples is diagnostic. A former criterion of a high stable antibody titer of 1 : 256 or higher is no longer considered diagnostic. Polymerase chain reaction can be used to amplify *Legionella* DNA in sputum, bronchoalveolar lavage fluid, pleural fluid, pulmonary tissue, or serum. Polymerase chain reaction can detect 1 fg of *Legionella* DNA, equivalent to one microorganism. These tests have not yet gained widespread use clinically. *Legionella* can be isolated from the blood with special media or by subculturing onto BCYE (buffered charcoal–yeast extract) agar plates, but this is not used in practice.

A chest radiograph is necessary to establish a diagnosis of pneumonia. About half the patients with LD have unilateral pulmonary involvement. The lower lobes are involved most commonly. About one third of patients have a pleural effusion. Dense opacification is common, but interstitial and nodular opacities also occur. Cavitation is uncommon; 70% of the 79 patients reported to date with lung abscess due to *Legionella* were receiving corticosteroids. Figures 314-2 to 314-5 illustrate some of the radiographic findings in LD.

The diagnosis of Pontiac fever is based on demonstration of *Legionella* in water to which the patient was exposed, seroconversion to *Legionella*, and a compatible clinical course.

Differential Diagnosis

LD should be considered in any patient with pneumonia who is admitted to the hospital, especially those who require treatment in an intensive care unit. If *Legionella* is present in a hospital's water supply, LD should be considered in all patients with nosocomial pneumonia.

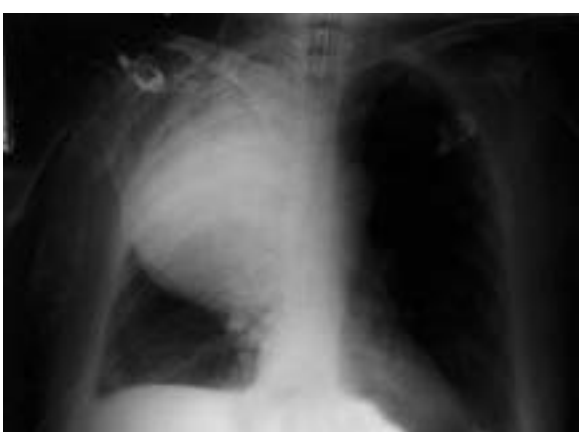

FIGURE 314-2. Posteroanterior chest radiograph of a patient with community-acquired pneumonia due to *Legionella pneumophila*. Note the dense consolidation of the right upper lobe, with bulging of the fissure. Such dense consolidation is a common radiographic appearance of legionnaires' disease.

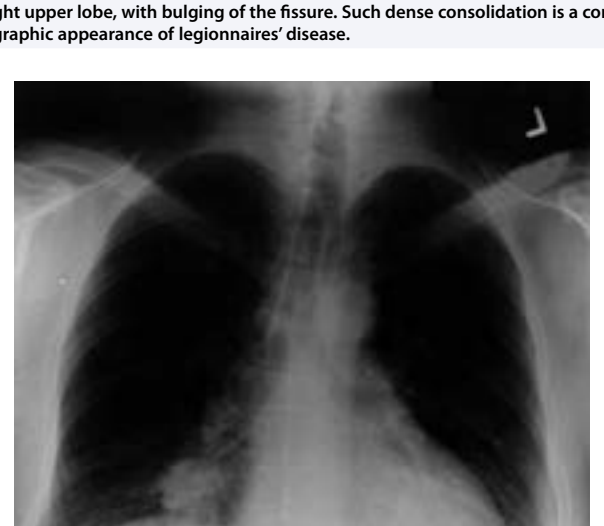

FIGURE 314-3. Posteroanterior chest radiograph of a patient with community-acquired legionnaires' disease (*L. pneumophila*) manifesting as a right lower lobe nodular opacity.

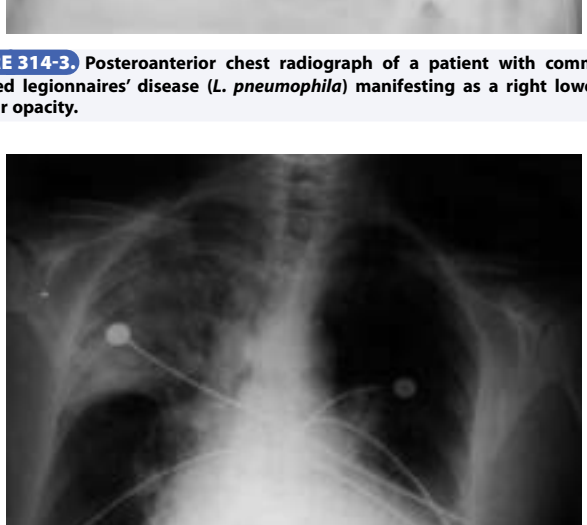

FIGURE 314-4. Posteroanterior chest radiograph of a patient with community-acquired *Legionella feeleii* pneumonia. There is patchy consolidation of the right upper lobe.

In patients who have died of legionnaires' disease, the gross pathology examination shows focal or patchy lesions in about one third of cases, lobar pneumonia in about half, and focal hemorrhages in about one quarter. On microscopic examination, there is bronchopneumonia with diffuse alveolar damage and heavy infiltration of neutrophils, macrophages, desquamation of alveolar epithelial cells, and fibrin proteinaceous debris. On occasion, there is inflammation of blood vessels mimicking a vasculitis. Organisms can be visualized with Dieterle's silver impregnation stain or by direct immunofluorescent staining (E-Fig. 314-1).

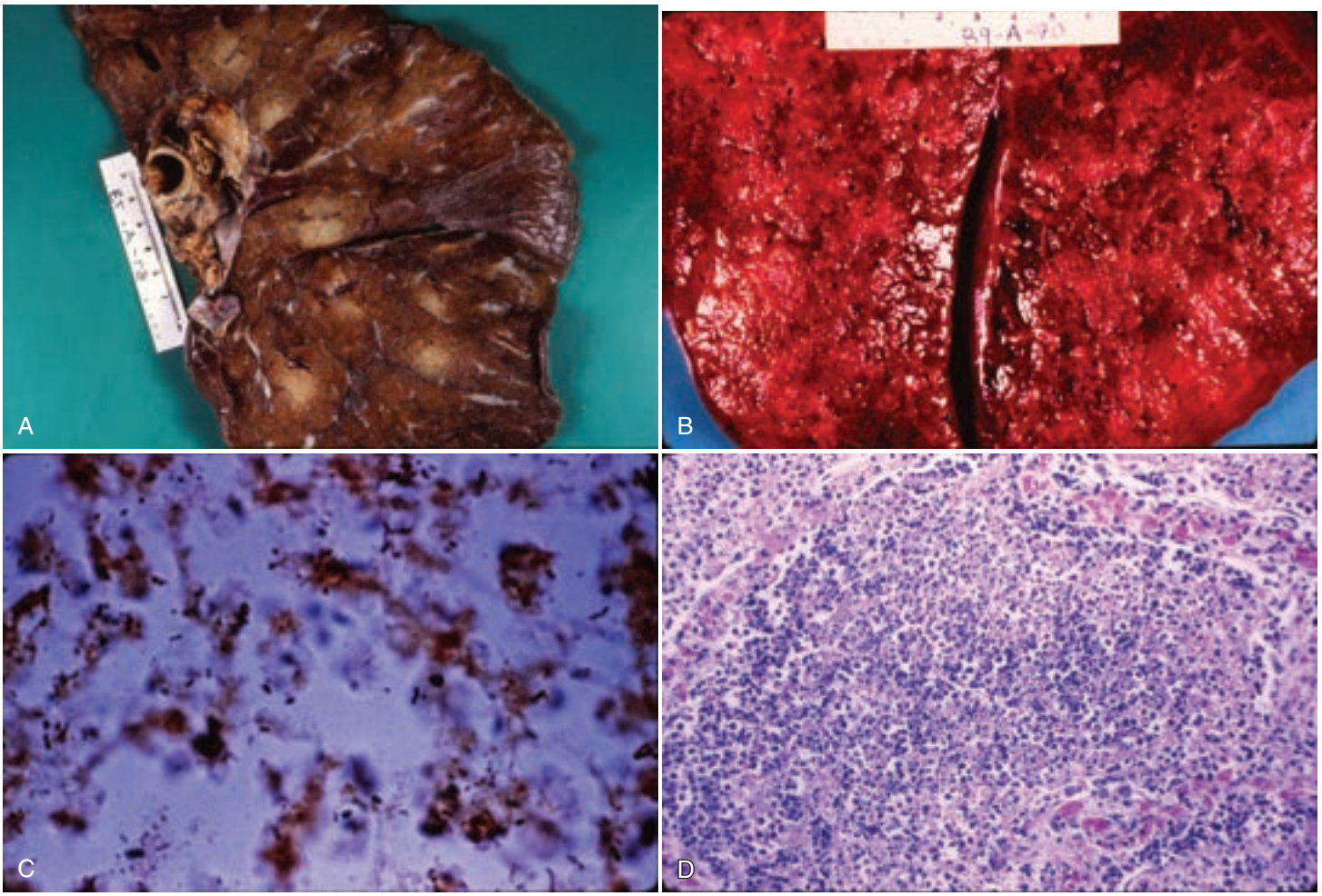

E-FIGURE 314-1. **A,** Gross pathology specimen of formalin-fixed lung from a patient with nosocomial legionnaires' disease (LD). Note the multiple nodular consolidated areas. **B,** Gross appearance of fresh lung tissue from a patient with LD. Note the "hepatization of the lung" representing consolidation. **C,** Dieterle stain on lung tissue from a patient with LD showing many microorganisms (magnification ×1000). **D,** Histologic appearance of pulmonary tissue from a patient with LD. Note the dense inflammatory infiltrate with evolving microabscess (magnification ×200).

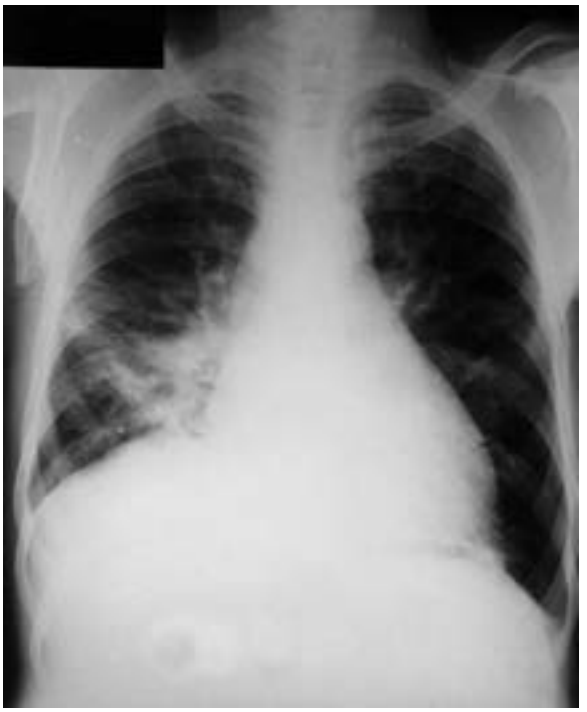

FIGURE 314-5. Posteroanterior chest radiograph of a patient with community-acquired pneumonia due to *Legionella pneumophila*. There is patchy consolidation at the right base, with subsegmental atelectasis and elevation of the right hemidiaphragm.

TABLE 314-2 TREATMENT FOR LEGIONNAIRES' DISEASE

SEVERITY OF *LEGIONELLA* PNEUMONIA	DRUG	DOSAGE
Mild pneumonia in a nonimmunocompromised person treated at home	Azithromycin	500 mg once daily for 5 days
	Clarithromycin	500 mg bid for 10 days
	Doxycycline	200 mg loading dose, then 100 mg bid for 10 days
	Levofloxacin	500 mg once daily for 10 days
	Moxifloxacin	400 mg once daily for 10 days
Pneumonia requiring hospitalization or pneumonia in an immunocompromised person	Levofloxacin	750 mg once daily (IV initially) for 10-14 days*
	Azithromycin	500 mg once daily for 10 days*
	Moxifloxacin	400 mg once daily for 10 days*
	Ciprofloxacin	400 mg q8h IV for 14 days*
	Erythromycin *plus* Rifampin	1000 mg IV qid for 3 days; then 500 mg qid for a total of 21 days 600 mg bid for 5 days

*Increase duration of treatment to 21 days for immunocompromised persons.

TREATMENT

If the diagnosis of LD is known, levofloxacin and moxifloxacin are first-choice agents (Table 314-2). Unfortunately, no randomized controlled trials have studied the therapy for LD. Data from a prospective, nonrandomized study indicate that levofloxacin is superior to macrolides for the treatment of severe LD. In this study carried out in Murcia, Spain, 3.4% of the patients receiving levofloxacin had complications, compared with 27.2% of those receiving macrolides; the levofloxacin patients had a shorter length of stay, 5.5 versus 11.3 days. The addition of rifampin to levofloxacin produced no additional benefit. In one clinical trial, 20 of 21 patients with LD were cured when treated with azithromycin. In another study, the authors selected 139 cases of *L. pneumophila* pneumonia in a prospective series of 1934 consecutive cases of community-acquired pneumonia. The overall mortality rate was 5%. Eighty patients received initial therapy with a macrolide, and 40 received levofloxacin. Those who received levofloxacin had a faster time to defervescence (2 vs. 4.5 days) and to clinical stability (3 vs. 5 days) and a shorter median length of stay (8 vs. 10 days); there were no significant differences in complications or mortality. In a review of data from six clinical trials, 75 patients with *Legionella* infection were treated with levofloxacin, and 90% of these infections had resolved clinically at the post-therapy visit 2 to 14 days after the termination of treatment. The authors concluded that treatment with levofloxacin 500 mg/day for 7 to 14 days or 750 mg/day for 5 days was effective. In a study of 33 patients admitted to an intensive care unit with LD, fluoroquinolone administration within 8 hours of arrival was associated with decreased mortality. There are no good data on the duration of treatment of LD. The consensus is that 7 to 10 days of treatment with azithromycin or a quinolone is sufficient. However, longer treatment is required if there is lung abscess, empyema, or endocarditis. In immunosuppressed patients, therapy should last for 21 days.

A meta-analysis of 24 trials involving 5015 patients with pneumonia showed that for patients with LD, there was a significant benefit when empirical therapy for pneumonia included agents active against *Legionella*. All patients who are severely ill with pneumonia should receive empirical therapy that treats *Legionella* as well as other pathogens. This is in line with the pneumonia guidelines issued by the American Thoracic Society and the Infectious Diseases Society of America.

Rifampin is frequently added to the treatment regimen of patients who are seriously ill with LD. There are no data suggesting synergy when this antibiotic is used in combination with antibiotics other than a macrolide. When it is used in this context, patients have a longer length of stay and higher bilirubin levels.

PROGNOSIS

In one study, the case-fatality rate from LD fell from 35% in 1993 to 5.6% in 2004. Such a decline in mortality was substantiated in a 15-year study of 217 cases of LD requiring admission to the hospital from Barcelona, Spain. The mortality decreased over time from 9% in 1995 to 2007 to 4% in 2007 to 2010.

In general, the fatality rate is higher in sporadic cases than in outbreak cases, and it is higher in nosocomial cases compared with community-acquired cases. The highest fatality rate is seen in nosocomial LD, for which it is still about 30%. Earlier diagnosis by the urinary antigen test and more effective antibiotic therapy account for much of the reduction in mortality. Factors associated with mortality are older age, immunosuppression, and severity of pneumonia.

Cases of LD should be reported to local health officials. An investigation is often required to determine the source of the *Legionella*.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Kanatani J, Isobe J, Kimata K, et al. Close genetic relationship between *Legionella pneumophila* serogroup 1 isolates from sputum specimens and puddles on roads by sequence-based typing. *Appl Environ Microbiol*. 2013;79:3959-3966.
2. Graham RM, Doyle CJ, Jennison AV. Real-time investigation of a *Legionella pneumophila* outbreak using whole genome sequencing. *Epidemiol Infect*. 2014;142:2347-2351.
3. Xu L, Luo ZQ. Cell biology of infection by *Legionella pneumophila*. *Microbes Infect*. 2013;15:157-167.
4. Cunha BA. Legionnaires disease: clinical differentiation from typical and other atypical pneumonias. *Infect Dis Clin North Am*. 2010;24:73-105.
5. Viasus D, Di Yacovo S, Garcia-Vidal C, et al. Community-acquired *Legionella pneumophila* pneumonia: a single center experience with 214 hospitalized sporadic cases over 15 years. *Medicine (Baltimore)*. 2013;92:51-60.
6. Haubitz S, Hitz F, Graedel L, et al. Ruling out *Legionella* in community acquired pneumonia. *Am J Med*. 2014;127:1010.e11-1010.e19.
7. Couturier MR, Graf EH, Griffin AT. Urine antigen tests for the diagnosis of respiratory infections: legionellosis, histoplasmosis, pneumococcal pneumonia. *Clin Lab Med*. 2014;34:219-236.

REVIEW QUESTIONS

1. All of the following are risk factors for legionnaires' disease, except

A. Cigarette smoking
B. Travel to Europe
C. Treatment with anti–tumor necrosis factor
D. Using potting soil for your garden in Australia
E. Being a member of the American Legion

Answer: E The disease was first recognized during a convention by the American Legion and was caused by water contamination with the organism at the hotel in Philadelphia that served as the meeting headquarters, not by any host factors. All of the other factors have been shown in one or more studies to increase one's risk of acquiring legionnaires' disease.

2. You have just seen a 55-year-old man with right lower lobe pneumonia. He has been receiving dialysis three times per week at an outpatient facility. He is quite ill, and you are concerned that he might have legionnaires' disease. The result of the *Legionella* urinary antigen test that you ordered is negative. Which of the following is correct?

A. Your patient does not have legionnaires' disease.
B. Your patient may still have legionnaires' disease.
C. Serologic testing will confirm the diagnosis.
D. Blood cultures will be positive for *Legionella*.
E. Sputum culture is a good test for *Legionella*.

Answer: B The urine antigen test detects only infection due to *L. pneumophila* serogroup 1, and even then it is only 74% sensitive. So your patient could still have legionnaires' disease, either serogroup 1 or another *Legionella* species. Sputum culture, if positive, is 100% specific. Unfortunately, most patients with legionnaires' disease do not produce sputum, and even more problematic is the fact that many laboratories are unable to grow this microorganism. Serologic diagnosis takes 4 to 6 weeks, so it is not of any use in therapeutic decision making. Also, the sensitivity can be low, especially in patients with an impaired immune response.

3. A patient has developed legionnaires' disease while hospitalized for treatment of acute pancreatitis. He spent 3 weeks in intensive care and had a nasogastric tube in place for most of that time. What was the most likely source of the patient's *Legionella*?

A. He acquired it before admission.
B. His daughter, who lives in Athens, came to visit him. She was coughing frequently during her many visits.
C. The air conditioning system
D. The potable water

Answer: D In all likelihood, the potable water is the source. You should check the protocol for nasogastric tubes in your hospital. Not uncommonly, potable water is used to initiate feeding. If the potable water is contaminated with *Legionella*, it is readily aspirated into the lungs in patients who are in a recumbent position in an intensive care unit. Person-to-person transmission of *Legionella* has not been documented. If the air conditioning system was contaminated, it would likely be through a cooling tower, and there would have likely been an outbreak of many cases of *Legionella*. Whereas it is always possible that *Legionella* was acquired before admission, the incubation period for legionnaires' disease is 2 to 10 days. It is possible to have a much longer incubation period. In the event of other cases in the hospital, molecular biology typing techniques can be used to identify the organism as coming from the potable water.

4. The following are extrapulmonary manifestations of legionnaires' disease.

A. Prosthetic valve endocarditis
B. Cerebellar ataxia
C. Reactive arthritis
D. Oral ulcers
E. A, B, and C

Answer: E Prosthetic valve endocarditis is due to direct infection of the valve by circulating *Legionella* bacteria. Cerebellar ataxia and reactive arthritis represent immunologic reactions in these tissues and are not due to direct infection.

5. The treatment of choice for moderate to severe legionnaires' disease is

A. Erythromycin
B. Amoxicillin
C. Levofloxacin
D. Vancomycin
E. Ceftriaxone

Answer: C A respiratory fluoroquinolone is probably the best antibiotic class for treatment of moderate to severe legionnaires' disease. There are not data from randomized clinical trials to provide guidance in this regard. Our recommendations are from observational studies. We do know that β-lactams are ineffective and erythromycin, although initially used to treat legionnaires' disease, takes 4 to 5 days to show improvement. Indeed, most patients get worse before they get better after treatment with erythromycin. Vancomycin would not be expected to work on a microorganism with a gram-negative cell wall.

315

BARTONELLA INFECTIONS

JEAN-MARC ROLAIN AND DIDIER RAOULT

DEFINITION

Bartonella species belong to the alpha-2 subgroup of Proteobacteria and are closely related to the genera *Brucella*, *Agrobacterium*, and *Rhizobium*. Since 1993, the genus *Bartonella* has been reorganized by addition of the genera *Rochalimaea* and *Grahamella* to the family Bartonellaceae. More than 30 known *Bartonella* species have been isolated from both animals and humans.[1] These bacteria are considered emerging pathogens that are associated with zoonosis and human infections. Among them, 14 validated species have been implicated in human diseases: *B. henselae, B. quintana, B. bacilliformis, B. elizabethae, B. clarridgeiae, B. vinsonii* subsp. *arupensis, B. vinsonii* subsp. *berkhoffii, B. alsatica, B. tamiae, B. grahamii, B. washoensis, B. rochalimae, B.*

TABLE 315-1 *BARTONELLA* SPECIES CAUSING HUMAN DISEASE

BARTONELLA SPECIES	FIRST CULTIVATION		YEAR OF DESCRIPTION	RESERVOIR HOST/VECTOR	HUMAN DISEASE
	Mammal	*Country*			
B. alsatica	Wild rabbit (*Oryctolagus cuniculus*)	France	1999	Rabbit	Endocarditis, lymphadenopathy
"*Candidatus* B. ancashi"		Peru	2013		Verruga peruana
B. bacilliformis	Human		1909	Human/sandfly	Carrión's disease, Oroya fever, verruga peruana
B. clarridgeiae	Cat		1996	Cat/cat flea	Cat-scratch disease
B. elizabethae	Endocarditis patient	United States	1993	Rat	Endocarditis, neuroretinitis
B. grahamii	Woodland mammal (*Clethrionomys glareolus*)	United Kingdom	1995	Rat, insectivore	Neuroretinitis
B. henselae	Cat		1990	Cat/cat flea	Cat-scratch disease, endocarditis, bacillary angiomatosis, bacillary peliosis, Parinaud oculoglandular syndrome, neuroretinitis, osteomyelitis, arthropathy, bacteremia with fever
B. koehlerae	Domestic cat	United States	1999	Cat	Endocarditis
B. mayotimonensis	Endocarditis patient	United States	2009	Unknown	Endocarditis
B. quintana	Human		1920	Human/body louse	Trench fever, endocarditis, bacillary angiomatosis
B. rochalimae	Human	United States	2007		Bacteremia, fever, splenomegaly
B. tamiae	Human	Thailand	2008		Febrile illness
B. vinsonii arupensis	Cattle rancher	United States	1999	Dog, rodent/ticks	Bacteremia with fever
B. washoensis			2000	Ground squirrel	Myocarditis

koehlerae, and "*Candidatus* B. ancashi" (Table 315-1). The other *Bartonella* species have been isolated only from the blood of animals, including rodents, felids, canids, dolphins, and ruminants. The route of transmission of *Bartonella* species in mammals and humans is by fleas, ticks, mites, and lice (see Table 315-1).

Bartonella infections are emerging infectious diseases that lead to a wide spectrum of either acute or chronic diseases. The status of the host immune response plays an important role in the development of the different manifestations. Four different clinical syndromes may occur with *Bartonella* infections: (1) infection of red blood cells and erythrophagocytosis, (2) granulomatous disease controlled by the immune response, (3) blood culture–negative endocarditis and bacteremia, and (4) vasculoproliferative diseases. A single *Bartonella* species can cause either acute or chronic infections and either vasculoproliferative or suppurative manifestations, but with different pathogenetic mechanisms that mainly depend on the patient's immune status. For example, *B. quintana* is responsible for trench fever as well as for endocarditis, bacteremia in the homeless population, and vasculoproliferative diseases, whereas *B. bacilliformis* is the agent of Carrión's disease, which corresponds to either an acute intraerythrocytic bacteremic disease (Oroya fever) or a chronic vasculoproliferative disease (verruga peruana). Infection of red blood cells has been well established for *B. bacilliformis* (Oroya fever) and *B. quintana* (trench fever and bacteremia in the homeless), whereas *B. henselae* and *B. koehlerae* have been seen in erythrocytes of infected cats. *B. henselae* can cause granulomatous disease, that is, cat-scratch disease (CSD), which affects lymph nodes, but can also be responsible for other clinical manifestations or complications, such as endocarditis. Vasculoproliferative diseases include bacillary angiomatosis caused by *B. henselae* and *B. quintana,* peliosis hepatis caused by *B. henselae,* and verruga peruana caused by *B. bacilliformis*. The immune status of the host plays a critical role in the development of these different forms of the disease. *B. henselae* usually causes CSD (a self-limited disease) in immunocompetent hosts, whereas it is responsible for bacillary angiomatosis in immunocompromised patients. In patients with a previous valvulopathy, any *Bartonella* infection may lead to endocarditis.

The Pathogen

Bartonella species are small, gram-negative, fastidious, pleomorphic coccobacilli or slightly curved rods (0.5 by 1 to 2 µm). Because of the slow growth of these bacteria and the lack of reproducible biochemical methods for their identification, they are usually identified by molecular methods. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry has emerged as a new technique for species identification and is an accurate and reproducible method for the rapid and inexpensive identification of *Bartonella* species. The bacteria can grow on enriched blood-containing media with a 5% carbon dioxide atmosphere after 5 to 15 days to up to 45 days on primary culture. The optimal growth temperature ranges from 28° C for *B. bacilliformis* to 35° to 37° C for the other species. *Bartonella* species can also be cocultured with endothelial cells. *Bartonella* species are either flagellated or nonflagellated cells. *B. bacilliformis* uses flagella for binding and deforming into the surface of erythrocytes. Bacteria can either persist in the blood stream of the host as intraerythrocytic parasites or colonize human endothelial cells.

EPIDEMIOLOGY

Almost all *Bartonella* species are vector-borne bacteria (see Table 315-1). Some are limited geographically, such as *B. bacilliformis,* which is found only in the Andes Mountains in South America at high altitudes, where its principal vector, *Lutzomyia verrucarum,* is distributed; others have a worldwide distribution, such as *B. henselae* and *B. quintana*. Each *Bartonella* species is highly adapted to its mammalian reservoir, in which bacteria usually cause a long-lasting intraerythrocytic bacteremia that may be asymptomatic. Humans are the hosts and reservoirs for *B. bacilliformis* and *B. quintana*. *B. quintana* is transmitted by the human body louse by inoculation of arthropod feces through broken skin (Chapter 359). Cats represent the main reservoir hosts for *B. henselae* infection; this pathogen is the agent of CSD in humans, caused by cat bites or scratches. *B. henselae* infection is transmitted from cat to cat by the cat flea. Cat fleas may also be infected by *B. quintana*. The role of dogs as reservoir hosts has been documented for several species, including *B. vinsonii* subsp. *arupensis*, *B. vinsonii* subsp. *berkhoffii*, and *B. henselae*. Wild rabbits are the reservoir hosts for *B. alsatica,* which is an agent of endocarditis and lymphadenopathy[2] in humans in close contact with rabbits. For other *Bartonella* species known to cause diseases in humans (see Table 315-1), their pathogenic role and mode of transmission are not fully understood.

INFECTION OF RED BLOOD CELLS: OROYA FEVER AND TRENCH FEVER

PATHOBIOLOGY

In Oroya fever, *B. bacilliformis* invades up to 80% of erythrocytes and produces their massive lysis, which results in severe hemolytic anemia, the major symptom of the disease.[3] Similarly, trench fever is characterized by an intracellular erythrocyte parasitism by *B. quintana*, with the percentage of infected red blood cells ranging from 0.001 to 0.005% (Fig. 315-1). Bacteria can also be seen extracellularly and in erythroblasts. This intracellular erythrocyte parasitism can presumably preserve the pathogens for efficient transmission by body lice, protect *B. quintana* from the host immune response, and

contribute to decreased antimicrobial efficacy.[4] During bacteremia in the homeless, *B. quintana* can also be seen in red blood cells.

CLINICAL MANIFESTATIONS

The main clinical manifestations of infection by *Bartonella* species are summarized in Table 315-2.[5]

Oroya fever is the acute or hemolytic phase of Carrión's disease, caused by *B. bacilliformis*; it usually develops 3 to 12 weeks after inoculation. Oroya fever results from the massive invasion of erythrocytes by *B. bacilliformis*, and without antibiotic treatment, it causes death in up to 85% of infected humans by hemolysis or when complicated by opportunistic infections such as salmonellosis. The onset is usually abrupt, with high fever, chills, headache, and anorexia. Patients have intense myalgias and arthralgias, abdominal pain, and jaundice. Complications are frequent, including meningoencephalitis, dyspnea, delirium, and superinfection leading to death. Asymptomatic persistent bacteremia may serve as the reservoir of the organism.

Trench fever is transmitted by lice (Chapter 359) and is the clinical manifestation of *B. quintana*. Trench fever affected more than 1 million people during World War I; more recently, *B. quintana* has been recognized in immunocompromised hosts, homeless people, and chronic alcoholics.[6] Clinical manifestations of trench fever may range from asymptomatic infection to severe, life-threatening illness. After an incubation period of 2 to 3 weeks, there is a sudden onset of fever that lasts 1 to 3 days associated with headache, shin pain, and dizziness. Although fatal cases have not been reported, the disease may persist for 4 to 6 weeks and result in prolonged disability. Relapses may occur years later, and in some cases there may be bacteremia with no clinical signs.

CAT-SCRATCH DISEASE

PATHOBIOLOGY

Little is known about the pathogenesis of the long-lasting lymphadenopathy in CSD. Immunopathogenesis is assumed to play an important role in CSD because bacteria have only rarely been isolated from affected lymph nodes. Thus, the disease is usually controlled by the host immune response, and there are few or no viable bacteria when lymph node biopsy specimens are analyzed; they are necrotic by pathologic examination.

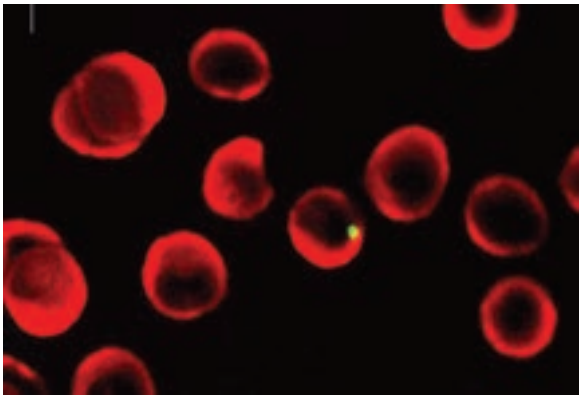

FIGURE 315-1. Section of human red blood cell infected with *Bartonella quintana* as viewed by confocal microscopy.

CLINICAL MANIFESTATIONS

Typical CSD is the most common manifestation of infection with *B. henselae* and usually is manifested as a self-limited regional lymphadenitis.[7] Transmission from cat to human occurs directly by a cat scratch or cat bite or possibly by a cat flea or tick bite. A typical papule or pustule may be seen 3 to 10 days after the scratch or bite at the site of inoculation and may last for 1 to 3 weeks. The subsequent lymphadenopathy is localized mainly to the axillary, cervical, or submaxillary nodes that drain the area where the cat scratch occurred. The enlarged lymph node is often painful and tender. Lymphadenopathy sometimes lasts for months, and in a few cases it can persist for as long as 1 to 2 years. In some cases, the lymph node may suppurate if it is not drained. Most patients are not febrile during the course of typical CSD. Systemic or severe disease may occur in about 5 to 14% of patients, with most of them suffering severe systemic symptoms due to disseminated infection. *B. alsatica* has been reported as a cause of lymphadenitis in a woman scratched on the finger while butchering a wild rabbit. After exposure to *B. henselae*, patients may develop bacteremia with or without clinical signs of typical CSD, and in patients with valvular lesions, this may result in infective endocarditis. Thus, CSD represents the primary infection of *B. henselae*, and endocarditis may follow in patients with heart valve lesions.

Approximately 10% of patients with CSD have atypical clinical manifestations, including prolonged fever (>2 weeks), malaise, neuroretinitis, encephalitis, erythema nodosum, hepatitis, fatigue, weight loss, and splenomegaly. A recent clinical study showed that musculoskeletal manifestations (myalgia, arthritis, arthralgia, tendinitis, osteomyelitis, neuralgia) were present in more than 10% of patients with CSD, demonstrating that these clinical manifestations are not as rare as might be expected from the cases reported in the past. In this series of 913 patients, myalgia and arthropathy were the most common manifestations, with an incidence of 5.8% and 5.5%, respectively. Moreover, these manifestations occurred primarily in adults whose ages ranged from 20 to 59 years. Myalgia had a mean duration of 4 weeks and was often severe. Arthropathy had a mean duration of 5.5 weeks, was more common in female patients older than 20 years, affected large and medium joints (half of those involved being weight-bearing joints), and was associated with symmetrical erythema nodosum early in the course of CSD. These musculoskeletal manifestations are often severe and may evolve into chronic forms that persist for more than a year. Tendinitis, neuralgia, and osteomyelitis are less common, with incidences lower than 1%.

Ocular and Neurologic Manifestations

Parinaud oculoglandular syndrome is a self-limited conjunctivitis associated with preauricular lymphadenopathy. Other atypical manifestations include neurologic syndromes (meningoencephalitis, meningitis, neuroretinitis). Encephalopathy may occur in 2 to 4% of CSD patients, mainly adolescents and adults. Patients usually have persistent headaches with or without fever and may develop seizures. Acute neurologic disorders range from self-limited nuchal rigidity to pupillary dilation or aphasia and hemiplegia; they may last

TABLE 315-2 CLINICAL MANIFESTATIONS ASSOCIATED WITH *BARTONELLA* SPECIES

CLINICAL MANIFESTATION	*B. BACILLIFORMIS*	*B. QUINTANA*	*B. HENSELAE*	*B. ALSATICA*	OTHERS
Intraerythrocytic bacteremia	+	+			
Chronic bacteremia	+	+	+		+
Infective endocarditis		+	+	+	+
Verruga peruana	+				
Bacillary angiomatosis		+	+		
Peliosis hepatis			+		
Lymphadenopathy		+	+ (CSD)	+	+
SENLAT with skin lesion			+		
Meningoencephalitis			+		
Uveitis-retinitis		+	+		+

CSD = cat-scratch disease; SENLAT = scalp eschar and neck lymphadenopathy.

for several weeks to months. Neuroretinitis has been associated with CSD in patients experiencing a sudden unilateral loss of visual acuity.[8] The most common picture remains papilledema associated with macular exudates causing stellar retinitis. A few reports have now established that *B. henselae* can be responsible for uveitis, along with *B. grahamii* and *B. quintana*. Patients present with either nongranulomatous or granulomatous uveitis.

Finally, tick-borne *B. henselae* infection has been described, including scalp eschar and neck lymphadenopathy after tick bites in three patients during the colder months in France. *B. henselae* was detected by molecular tools both in skin biopsy specimens (cervical and occipital) and in a *Dermacentor marginatus* tick removed from the scalp of one patient. All three patients had asthenia, but none had alopecia.

ENDOCARDITIS

PATHOBIOLOGY

B. quintana, B. henselae, B. alsatica, B. vinsonii subsp. *berkhoffii, B. elizabethae,* and "*Candidatus* Bartonella mayotimonensis" have been associated with blood culture–negative endocarditis, whereas *B. vinsonii* subsp. *arupensis* has been detected in a patient with fever and bacteremia; *B. washoensis* has been identified in one patient with myocarditis; *B. rochalimae* was reported in a patient with fever, bacteremia, and splenomegaly; and *B. tamiae* has been isolated in a patient with a febrile illness. Patients with endocarditis usually have preexisting heart valve disease that promotes the development of infective endocarditis and, in some cases, a definite risk factor for infection specifically with *Bartonella*. Endocarditis caused by *Bartonella* species exhibits slight inflammation, with a few inflammatory mononuclear cells and small vegetations; the bacteria are seen extracellularly in dense immunopositive clusters that are mainly included in vegetations and in neutrophil and macrophage cytoplasm.

CLINICAL MANIFESTATIONS

The most commonly identified agents of *Bartonella* endocarditis are *B. quintana*, followed by *B. henselae* and other *Bartonella* species.[9] Patients appear to have chronic, blood culture–negative endocarditis, usually with fever (90%). Echocardiography reveals vegetations (in 90%) that should be removed by surgery in the majority of patients. Infections with *B. henselae* are epidemiologically linked to close contact with cats or cat fleas and previous valvular heart disease, whereas *B. quintana* endocarditis is frequently described in homeless and alcoholic patients with body lice infection and can be observed in patients without previous valve lesions. The onset is usually subacute, with some patients being afebrile at the time of admission. About half the patients have embolic phenomena. Interestingly, there is a north (Europe) to south (North Africa) gradient for the proportion of *Bartonella* endocarditis in humans; thus *Bartonella* is apparently a common cause of endocarditis in North Africa. Sporadic cases of endocarditis have also been associated with *B. koehlerae, B. vinsonii* subsp. *berkhoffii, B. vinsonii* subsp. *arupensis, B. elizabethae, B. alsatica,* and "*Candidatus* Bartonella mayotimonensis."

VASCULOPROLIFERATIVE DISEASE: VERRUGA PERUANA, BACILLARY ANGIOMATOSIS, AND PELIOSIS HEPATIS

PATHOBIOLOGY

Bartonella species have the ability to cause vasculoproliferative lesions through a process of pathologic angiogenesis resulting in the formation of new capillaries from preexisting ones. These typical vasoproliferations can be expressed as skin lesions called bacillary angiomatosis, which are caused by *B. quintana* and *B. henselae*[10]; there is also a cystic form in the liver and spleen called peliosis hepatis, which is caused only by *B. henselae.*[11] Skin lesions are similar to those reported for verruga peruana, the chronic form of Carrión's disease. Bacillary angiomatosis is a neovascular proliferation that has been reported most commonly in AIDS patients and involves the skin (Fig. 315-2) and lymph nodes; it occurs less frequently in patients with other causes of immunosuppression and only exceptionally in immunocompetent patients. In bacillary angiomatosis, lesions comprise proliferating endothelial cells, bacteria, and mixed infiltrates of macrophages/monocytes and polymorphonuclear neutrophils, leading to chronic inflammation. Bacteria are clustered as aggregates both surrounding and within endothelial cells, indicating that the vascular endothelium represents a target tissue for intracellular and extracellular colonization in vivo. On histologic evaluation, bacillary angiomatosis

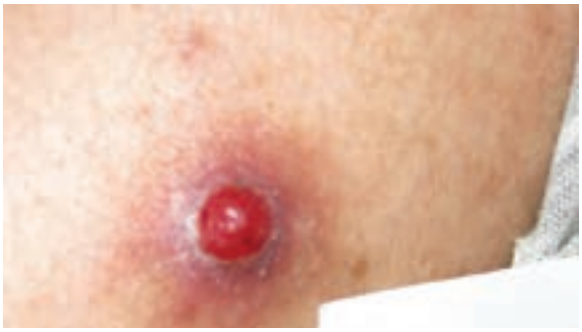

FIGURE 315-2. Skin lesion of bacillary angiomatosis.

is a lobular proliferation of small blood vessels containing endothelial cells and bacteria, usually seen in clusters when stained with Warthin-Starry. As in bacillary angiomatosis, lesions of verruga peruana are characterized by lobular proliferations and atypical endothelial cells forming both relatively solid sheets and small, well-formed vessels with patent lumens. Lesions are typically infiltrated, indicating a chronic inflammatory process.

CLINICAL MANIFESTATIONS

As already noted, bacillary angiomatosis is seen most often in AIDS patients. Cutaneous lesions often arise in crops and can be subcutaneous or dermal nodules with red or purple papules millimeters to centimeters in diameter. When cutaneous lesions are absent, the diagnosis is often difficult and delayed because signs of visceral involvement are usually nonspecific. The potentially systemic nature of bacillary angiomatosis is reflected by the involvement of brain, bone, lymph node, bone marrow, skeletal muscle, conjunctiva, and mucosal surfaces of the gastrointestinal and respiratory tracts. Peliosis hepatis affects solid internal organs, primarily the liver, with reticuloendothelial elements; in the liver, it is defined as a vascular proliferation of sinusoidal hepatic capillaries resulting in blood-filled spaces. The spleen, abdominal lymph nodes, and bone marrow may also be involved.

Following acute Oroya fever, patients usually develop angioproliferative cutaneous tumors called verruga peruana after a latent period ranging from weeks to months. The infection is characterized by benign cutaneous vascular lesions typically consisting of round papules that are frequently pruritic and bleeding. The infection is accompanied by malaise and arthralgias. Skin lesions may change over time from miliary to nodular subcutaneous lesions to large mulaire lesions. These large lesions are often engorged with blood and prone to ulceration and bleeding. This eruptive phase clinically resembles Kaposi's sarcoma or bacillary angiomatosis. However, it has a low morbidity, and there are no reports of mortality.

DIAGNOSIS OF *BARTONELLA* INFECTION

Methods used for the diagnosis of *Bartonella* infection include serology, microscopy, culture, molecular amplification of *Bartonella* species genes, direct immunofluorescence, and immunohistochemistry. The usefulness of these techniques may vary according to the disease involved (Table 315-3).

Serologic Tests

Serology remains the most widely used method for the diagnosis of CSD and *Bartonella* endocarditis because culture and isolation are difficult and time-consuming, and molecular methods are not available in all laboratories. There are currently two classic serologic methods for the diagnosis of *Bartonella* infections: enzyme-linked immunosorbent assay and immunofluorescence assay. By immunofluorescence assay, an immunoglobulin G titer of 1 : 64 or greater should be considered positive for CSD, whereas patients with endocarditis usually have higher antibody titers (≥1 : 800). In homeless patients, bacteremia has been associated with serologic tests that were positive for *B. quintana*. However, reported sensitivities of immunofluorescence assay vary considerably, from nearly 100% to less than 30%, depending on the nature of the antigens used and the selected patients. Moreover, owing to cross-reactive antibodies between *Bartonella* species, the diagnosis of *Bartonella* infection at the species level is usually not possible. More sophisticated methods should be used, especially Western blot with cross-adsorption analysis. Western blot is also useful for the differential diagnosis of endocarditis

TABLE 315-3 LABORATORY METHODS FOR THE DIAGNOSIS OF *BARTONELLA* INFECTIONS

CLINICAL MANIFESTATION	SEROLOGY	CULTURE	MOLECULAR METHODS	IMMUNOHISTOCHEMISTRY (WARTHIN-STARRY STAIN)	MICROSCOPY FOR INTRAERYTHROCYTIC ORGANISM (GIEMSA OR IF)
Oroya fever	–	+	+	–	+++
Verruga peruana	–	–	+	+	–
Trench fever	+/–	+	+	–	+
Chronic bacteremia	+/–	+++	+	–	–
Infective endocarditis	+++	+++	+++	+++	–
Bacillary angiomatosis	+/–	++	+++	+++	–
Peliosis hepatis	+/–	++	+++	+++	–
Cat-scratch disease	+	–	+++	+/–	+/–
SENLAT with skin lesion	+	+	+++	+	–
Meningoencephalitis	++	–	++	–	–
Uveitis-retinitis	++	–	++	+	–

IF = immunofluorescence; SENLAT = scalp eschar and neck lymphadenopathy.

because the profile obtained for endocarditis is specific compared with that for CSD or chronic bacteremia.

Microscopy, Immunofluorescence, and Immunohistochemistry

The diagnosis of Oroya fever is based on examination of a peripheral blood smear stained by Giemsa; the percentage of infected red blood cells is sufficiently high so that bacteria are visible. Similarly, *B. quintana* can be seen within red blood cells with use of specific monoclonal antibody by immunofluorescence and confocal microscopy (see Fig. 315-1). Microscopic examination after Warthin-Starry silver staining or immunohistochemistry of a cardiac valve or skin biopsy specimen is also useful for the detection of *Bartonella* organisms in patients with endocarditis and bacillary angiomatosis.

Culture

Bartonella species can be recovered from the blood of bacteremic patients as well as from cardiac valves, skin and liver biopsy specimens, and, rarely, lymph nodes. Bacteria can be isolated directly from these specimens after plating onto blood agar solid media, blood culture in broth, and cocultivation in endothelial cells. Cultures are usually positive after 2 weeks of incubation, but up to 45 days may be necessary for primary isolation.

Molecular Detection Methods

Direct detection and final identification of *Bartonella* species from blood and tissue specimens, including lymph node, cardiac valve, skin, and liver, can be achieved by polymerase chain reaction amplification and sequencing of various housekeeping genes. Real-time polymerase chain reaction with TaqMan probes has been added to the panel of molecular techniques for the specific detection of *Bartonella* at the species level from lymph nodes in CSD, directly from cardiac valves in patients with endocarditis, and in patients with bacillary angiomatosis or peliosis hepatis.

Differential Diagnosis

The differential diagnosis of bacillary angiomatosis is Kaposi's sarcoma, but visualization of bacteria in the tissue specimen can help distinguish between these two entities. Trench fever may be confused mainly with typhus group rickettsiosis, relapsing fever, and malaria. CSD may be confused with tularemia, pyogenic lymphadenitis, mycobacterial infection, and neoplasia.

TREATMENT

Trench Fever

Before the advent of antibiotics, acetylsalicylic acid was the most effective drug for the pain caused by trench fever. However, a randomized clinical trial in homeless persons with episodes of bacteremia demonstrated that the combination of gentamicin and doxycycline is more effective in stopping bacteremia compared with no treatment or the use of β-lactams or doxycycline alone. Because the intraerythrocytic presence of *B. quintana* decreases antimicrobial efficacy, the duration of treatment is critical in trench fever. Patients with *B. quintana* bacteremia should be treated with gentamicin (3 mg/kg body weight/day IV for 14 days), in combination with doxycycline (200 mg/day PO) for 28 days.

Oroya Fever and Verruga Peruana

In Oroya fever, treatment with penicillin, streptomycin, fluoroquinolones, tetracycline, or erythromycin produces rapid defervescence and disappearance of the organisms from the blood, usually within 24 hours. As an alternative treatment, chloramphenicol can be used alone or in combination with a β-lactam. Treatment with chloramphenicol may also have the advantage of covering commonly associated *Salmonella* species. Patients with Oroya fever should be treated with chloramphenicol (500 mg PO four times a day) for 14 days in combination with another antibiotic (especially a β-lactam compound). Streptomycin (15 to 20 mg/kg/day for 10 days) was the traditional treatment for verruga peruana. However, its use is problematic, especially in children, and rifampin has become the drug of choice for the treatment of eruptive-phase bartonellosis. Failures of rifampin treatment have been reported in verruga peruana. Finally, ciprofloxacin (500 mg PO twice daily for 7 to 10 days) has been used successfully for the treatment of multiple eruptive-phase lesions in adults and has been proposed as an appropriate alternative. Doxycycline in association with gentamicin may be used to treat the eruptive phase of Carrión's disease.

Cat-Scratch Disease

Typical CSD is a self-limited illness that resolves within 2 to 6 months and usually does not respond to therapy because the bacteria within necrotic lymph nodes are not alive. In cases of long-lasting lymphadenopathy, patients should be reassured that it is benign and will probably subside spontaneously within 2 to 4 months. Management consists of analgesics for pain and prudent follow-up.[12] However, azithromycin (500 mg PO on day 1 and 250 mg PO on days 2 to 5 as single daily doses) is an alternative for patients with large, bulky lymphadenopathy.[13] If lymphadenopathy does not resolve, lymph nodes can be removed surgically. For atypical presentations of CSD, there are no data regarding the benefit of specific antimicrobial therapy for immunocompetent patients. For neuroretinitis, the combination of doxycycline (100 mg PO or IV twice daily) with rifampin (300 mg PO twice daily) seems to promote disease resolution, to improve visual acuity, to reduce optic disc edema, and to decrease disease duration.

Endocarditis

Patients with *Bartonella* endocarditis have a higher death rate and undergo valvular surgery more frequently than patients with endocarditis caused by other pathogens. Patients with suspected (but culture-negative) *Bartonella* endocarditis or proved *B. quintana* endocarditis should be treated with oral doxycycline 100 mg twice a day for 6 weeks in combination with gentamicin 3 mg/kg/day in one intravenous daily dose for 14 days.[14] The American Heart Association consensus on treating infective endocarditis is ceftriaxone plus gentamicin, with or without doxycycline, when *Bartonella* is suspected, and doxycycline plus gentamicin when *Bartonella* endocarditis is confirmed. However, there is direct evidence that patients receiving an aminoglycoside are more likely to fully recover, and those treated with aminoglycosides for at least 14 days are more likely to survive than those receiving a shorter duration

of therapy. In the absence of any prospective study for the treatment of *Bartonella* endocarditis, the same regimen as for *B. henselae* and *B. quintana* should be used for endocarditis when another *Bartonella* species has been identified as the causative agent.

Bacillary Angiomatosis and Peliosis Hepatis

Without appropriate therapy, infection spreads systemically and can involve virtually any organ, and the outcome is sometimes fatal. Thus, antibiotic treatment is warranted in all cases of *Bartonella*-associated vasculoproliferative disease. On the basis of empirical clinical reports, erythromycin remains the treatment of choice and has been used successfully to treat many patients with bacillary angiomatosis. The response to treatment in bacillary angiomatosis can be dramatic in immunocompromised patients, with resolution of palpable subcutaneous lesions within hours. Erythromycin also has an antiangiogenic effect on microvascular endothelial cells that could explain this quick disappearance of lesions. Erythromycin (500 mg four times daily) for 3 months is first-line therapy. Doxycycline (100 mg PO or IV twice daily) is currently proposed as an appropriate alternative. In patients with serious infections, erythromycin or doxycycline can be used in combination with rifampin (300 mg PO twice daily). The duration of therapy is critical. We recommend that treatment be given for at least 3 months for bacillary angiomatosis and 4 months for peliosis hepatis. Peliosis hepatis responds slowly, and hepatic lesions continue to improve after several months of treatment, whereas cutaneous bacillary angiomatosis demonstrates improvement after 4 to 7 days of treatment and resolves after about 1 month of treatment. Relapses of peliosis hepatis and bacillary angiomatosis lesions after antibiotic treatment have been reported frequently, especially in immunocompromised patients with a shorter duration of therapy. Patients who relapse after the recommended treatment should probably be re-treated for 4 to 6 months with erythromycin (500 mg PO four times daily) or doxycycline (100 mg PO twice daily), and those with repeated relapses should receive antibiotic therapy as long as they are immunocompromised.

PREVENTION

Because arthropods play an important role in the transmission of feline *Bartonella* species to humans, rigorous arthropod control (Chapter 359) should be recommended by health care workers, particularly when advising immunocompromised individuals on the risks related to pet ownership. Cat fleas live on both cats and dogs and are best controlled by fumigating areas where these animals live.

PROGNOSIS

Mortality in those with Oroya fever was as high as 50% before the antibiotic era but is now limited by the use of antibiotics. Trench fever should be treated with antibiotics to avoid more severe disease, especially in patients with valvulopathy who are at risk for development of endocarditis. CSD usually resolves spontaneously without any treatment in immunocompetent patients. In the case of complications or in immunocompromised patients, antibiotic therapy should be given, and patients usually respond well to treatment. Finally, for vasculoproliferative diseases, antibiotics are usually effective if they are given for a long time, but cutaneous lesions in verruga peruana may benefit from surgery.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Buffet JP, Kosoy M, Vayssier-Taussat M. Natural history of *Bartonella*-infecting rodents in light of new knowledge on genomics, diversity and evolution. *Future Microbiol*. 2013;8:1117-1128.
2. Block SL. Managing cervical lymphadenitis—a total pain in the neck! *Pediatr Ann*. 2014;43: 390-396.
3. del Valle Mendoza J, Silva Caso W, Tinco Valdez C, et al. Diagnosis of Carrington's disease by direct blood PCR in thin blood smear negative samples. *PLoS ONE*. 2014;9:e92283.
4. Eicher SC, Dehio C. *Bartonella* entry mechanisms into mammalian host cells. *Cell Microbiol*. 2012;14:1166-1173.
5. Maguina C, Guerra H, Ventosilla P. Bartonellosis. *Clin Dermatol*. 2009;27:271-280.
6. Badiaga S, Brouqui P. Human louse-transmitted infectious diseases. *Clin Microbiol Infect*. 2012;18: 332-337.
7. Zangwill KM. Cat scratch disease and other *Bartonella* infections. *Adv Exp Med Biol*. 2013;764: 159-166.
8. Gan JJ, Mandell AM, Otis JA, et al. Suspecting optic neuritis, diagnosing *Bartonella* cat scratch disease. *Arch Neurol*. 2011;68:122-126.
9. Chaloner GL, Harrison TG, Birtles RJ. *Bartonella* species as a cause of infective endocarditis in the UK. *Epidemiol Infect*. 2013;141:841-846.
10. Moulin C, Kanitakis J, Ranchin B, et al. Cutaneous bacillary angiomatosis in renal transplant recipients: report of three new cases and literature review. *Transpl Infect Dis*. 2012;14:403-409.
11. Battal B, Akgun V, Sari S. Peliosis hepatis: one pathology, a thousand faces, and a clinical and radiological diagnostic challenge. *J Dig Dis*. 2014;15:281-282.
12. Prutsky G, Domecq JP, Mori L, et al. Treatment outcomes of human bartonellosis: a systematic review and meta-analysis. *Int J Infect Dis*. 2013;17:e811-e819.
13. Biswas S, Rolain JM. *Bartonella* infection: treatment and drug resistance. *Future Microbiol*. 2010;5:1719-1732.
14. Angelakis E, Raoult D. Pathogenicity and treatment of Bartonella infections. *Int J Antimicrob Agents*. 2014;44:16-25.

REVIEW QUESTIONS

1. Which of the following is the main vector of *Bartonella quintana* infections in humans?

A. Mites
B. Lice
C. Ticks
D. Bed bugs
E. Fleas

Answer: B The human body louse is the main vector (see Table 315-1), but *B. quintana* could be detected in fleas and ticks.

2. Which treatment is the most appropriate recommendation for *Bartonella* endocarditis?

A. Erythromycin
B. Doxycycline plus rifampin
C. Doxycycline
D. Doxycycline plus gentamicin
E. Streptomycin

Answer: D *Bartonella* endocarditis should be treated with doxycycline for 6 weeks plus gentamicin for 14 days. (See Treatment section.)

3. Which laboratory method is the most appropriate for the diagnosis of cat-scratch disease (CSD)?

A. Real-time polymerase chain reaction
B. Serology
C. Gram staining
D. Cell culture
E. Immunohistochemistry

Answer: A Serology remains a widely used method where molecular methods are not available for the diagnosis of CSD and *Bartonella* endocarditis because culture and isolation are difficult and time-consuming. However, the reported sensitivities of immunofluorescence assays vary considerably, depending on the nature of the antigen used and the selected patients. Furthermore, because of cross-reactivity of antibodies between *Bartonella* species, serologic methods usually cannot make the diagnosis of *Bartonella* infection at the species level. Direct detection and definitive *Bartonella* species identification are possible from blood and tissue (e.g., lymph node) specimens. See Table 315-3.

4. Which *Bartonella* species is the causative agent of peliosis hepatis?

A. *B. quintana*
B. *B. henselae*
C. *B. bacilliformis*
D. *B. grahamii*
E. *B. alsatica*

Answer: B Peliosis hepatis is characterized by the development of cystic, blood-filled cavities throughout the liver parenchyma and can be caused by a variety of drugs, malignant neoplasms, post–renal transplantation, and different infections including human immunodeficiency virus infection, tuberculosis, and *Bartonella henselae*.

5. Which treatment is the most appropriate recommendation for bacillary angiomatosis?

A. Erythromycin
B. Doxycycline plus rifampin
C. Doxycycline
D. Doxycycline plus gentamicin
E. Streptomycin

Answer: A Erythromycin (500 mg four times daily) for 3 months is first-line therapy. (See Treatment section.)

316

GRANULOMA INGUINALE (DONOVANOSIS)

EDWARD W. HOOK III

DEFINITION

Granuloma inguinale, also known as donovanosis, is a slowly progressive ulcerative disease that involves principally the skin and subcutaneous tissues of the genital, inguinal, and anal regions.

The Pathogen

The causative organism is *Klebsiella granulomatis* (formerly *Calymmatobacterium granulomatis*), a facultative intracellular parasite. The organism is challenging to cultivate but can sometimes be grown in yolk sacs, and successful cell culture has been reported from South Africa and Australia. Successful culture, in turn, has permitted the development of polymerase chain reaction assays, currently for research purposes.

EPIDEMIOLOGY

The organism is primarily transmitted sexually, but it can probably be transmitted by nonsexual contact as well. Transmission efficiency is relatively low, and multiple sexual contacts with an infected partner seem to be necessary for the transmission of infection. The disease is rarely encountered in the United States. Although still relatively rare, it is more common in other areas of the world, including India, Papua New Guinea, the Caribbean, southern Africa, and parts of Australia, and it may be becoming less common even in these areas.

CLINICAL MANIFESTATIONS

After an incubation period of up to 50 days, the initial lesion usually appears as a subcutaneous nodule that erodes through the surface and develops into a beefy, elevated granulomatous lesion (Fig. 316-1).[1,2] The lesion is painless but tends to bleed easily, and it is not associated with systemic symptoms. Lesion exudate is often described as foul smelling. Secondary bacterial infection may cause a necrotic, painful, ulcerative lesion that may be rapidly destructive. A cicatricial form may also occur, with a depigmented elevated area of keloid-like scar containing scattered islands of granulomatous tissue. About 90% of lesions occur on the genitals and are commonly associated with pseudobuboes; these swellings usually are not due to involvement of the inguinal lymph nodes but rather are due to granulomatous involvement of subcutaneous tissue. Metastatic infection of bones or viscera is occasionally seen. Clinical experience suggests that secondary carcinomas may be a rare complication of granuloma inguinale.

DIAGNOSIS

The diagnosis is made by demonstrating intracellular Donovan's bodies in histiocytes or other mononuclear cells from lesion scrapings or biopsy samples. Wright stain and Giemsa stain of fresh impression smears or unfixed biopsy specimens usually demonstrate the bacilli relatively easily, although multiple biopsies may be necessary in chronic cases. Histologic examination of biopsy specimens shows plasma cells with some infiltration by polymorphonuclear leukocytes but no giant cells. In infected lesions, *K. granulomatis* is found primarily in histiocytes or other mononuclear cells. Cell culture and polymerase chain reaction methods are currently primarily research tools. No serologic test is clinically available.

Differential Diagnosis

The differential diagnosis includes squamous cell carcinoma, chancroid (Chapter 301), lymphogranuloma venereum (Chapter 318), syphilis

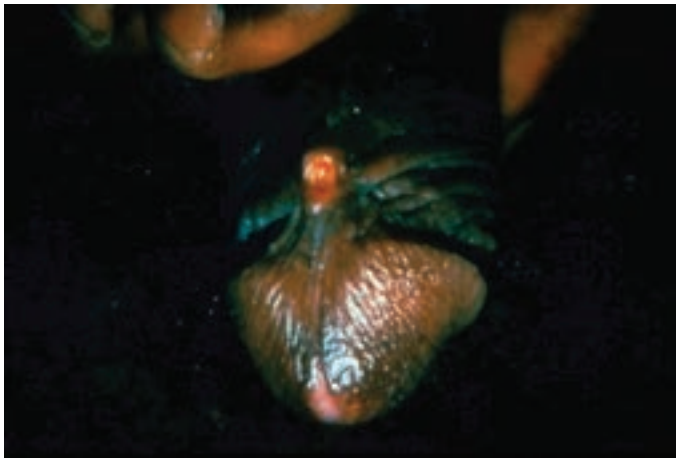

FIGURE 316-1. Typical primary lesion of granuloma inguinale. (Reproduced with permission from Herpes-Coldsores.com. http://www.herpes-coldsores.com/std/lymphogranuloma-pictures.htm.)

(Chapter 319), and other ulcerative granulomatous diseases.[3] In the absence of therapy, patients may not present until lesions have been present for months, long after the lesions of syphilis and other ulcerative sexually transmitted diseases have resolved. Chancroid is usually differentiated by its irregular undermined borders, which are not seen in typical cases of granuloma inguinale. Dark-field examination and serologic tests should help distinguish syphilis. Biopsy of lesions may be necessary to distinguish granuloma inguinale from certain tumors.

TREATMENT

Rx

There are no recent clinical trials of therapy for granuloma inguinale. Recommended treatment consists of azithromycin administered either as 1.0 g weekly or 500 mg daily or doxycycline 100 mg twice daily.[4] One double-strength trimethoprim-sulfamethoxazole tablet twice daily or erythromycin base 500 mg four times daily is recommended as alternative therapy. An aminoglycoside (e.g., gentamicin 1 mg/kg intravenously every 8 hours) may be added if these regimens do not result in clinical improvement within a few days. Treatment should be administered for at least 3 weeks and until lesions are completely healed. Patients should be monitored for at least several weeks after treatment is discontinued because of the possibility of relapse. Although the risk of communicability appears to be low, sexual contacts should also be examined; at present, treatment of contacts is not indicated in the absence of clinically evident disease.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Velho PE, de Souza EM, Belda W Jr. Donovanosis. *Braz J Infect Dis*. 2008;12:521-525.
2. Bezerra SM, Jardim MM, Silva VB. Granuloma inguinale (Donovanosis). *An Bras Dermatol*. 2011;86:585-586.
3. Basta-Juzabasic A, Ceovic R. Chancroid, lymphogranuloma venereum, granuloma inguinale, genital herpes simplex virus, and molluscum contagiosum. *Clin Dermatol*. 2014;32:290-298.
4. O'Farrell N, Moi H, IUSTI/WHO European STD guidelines Editorial Board. European guideline for the management of donovanosis, 2010. *Int J STD AIDS*. 2010;21:609-610.

REVIEW QUESTION

1. Which of the following statements regarding granuloma inguinale is correct?

A. It is transmitted efficiently and only by sexual contact.
B. The initial lesion is typically painful and associated with fever.
C. Lesions are associated with buboes.
D. Diagnosis requires demonstration of intracellular Donovan's bodies.
E. Diagnosis requires serologic confirmation.

Answer: D No serologic test is clinically available for granuloma inguinale, and definitive diagnosis depends on demonstration of Donovan's bodies in mononuclear cells from lesion scrapings or biopsy samples. Although it is primarily transmitted through sexual contact, the organism can probably be transmitted by nonsexual contact as well; transmission efficiency is low. The lesion is painless, and it is not accompanied by systemic symptoms. It is associated with a so-called pseudobubo, a swelling in the inguinal area caused by granulomatous infiltration of subcutaneous tissue—not actual inguinal lymph node involvement that would constitute a bubo.

317

MYCOPLASMA INFECTIONS

STEPHEN G. BAUM

DEFINITION

Mycoplasma organisms of the class Mollicutes are ubiquitous as pathogens and colonizing agents in the plant, animal, and insect kingdoms. They represent the smallest known free-living forms, but because they have fastidious growth requirements, they are difficult to culture. However, the presence of several species of *Mycoplasma* as commensals in animals and on human oral and genital mucosa frequently resulted in contamination of cell cultures. Such contamination led to the false implication of mycoplasmas as causative agents in many human diseases, both trivial and life-threatening. Of the human diseases proved to be due to mycoplasmas, pneumonia caused by *Mycoplasma pneumoniae* is by far the most clinically important. This infection constitutes a significant proportion of cases previously classified as atypical pneumonia, a nonspecific term for patchy pneumonias that generally do not respond to β-lactam antibiotics and have etiologic agents that are not easily cultured or visible on Gram stain. The term *atypical pneumonia* persists despite our increasing ability to identify specific etiologic agents, such as viruses, *Legionella*, and *Chlamydophila*. Other proven *Mycoplasma* infections include those in the urogenital tract caused by *Ureaplasma* species, *Mycoplasma hominis*, and *Mycoplasma genitalium*; wound infections caused by *M. hominis*; and overwhelming systemic infection in immunocompromised patients caused by *Mycoplasma fermentans* (*incognitus* strain).

The Pathogen

Mycoplasmas are short rods (10 × 200 nm) that have no cell wall and are bounded by a sterol-containing membrane; thus, they are unaffected by cell wall–inhibiting antimicrobials such as β-lactams. In tissue culture, mycoplasmas are intracellular; but in vivo, infection is primarily extracellular and affects epithelial cells and their organelles, such as cilia. Attachment to respiratory epithelium is by way of terminal adhesin proteins in specialized tip organelles.

EPIDEMIOLOGY

M. pneumoniae infection is spread person to person by respiratory droplets produced by coughing. Relatively close association with the index case appears to be required. The disease is usually introduced into families by a young child; in some studies, most of the infected adults were the parents of young children. As opposed to most viral respiratory infections, which are manifested 1 to 3 days after infection, *Mycoplasma* has an incubation period of 2 to 3 weeks. Therefore, a careful history showing several weeks between cases within a family may be an important clue to the mycoplasmal etiology. Organisms can be cultured from the sputum of infected individuals for weeks to months after clinically efficacious treatment.

Most cases of *Mycoplasma* respiratory infection occur singly or as family outbreaks. However, in closed populations, such as military recruit camps and boarding schools, *Mycoplasma* can cause mini-epidemics and may be responsible for 25 to 75% of cases of pneumonia in such settings. Serologically based epidemiologic studies have documented the high incidence of *Mycoplasma* respiratory infection throughout the world. In the United States, it is estimated that each year at least one case of *Mycoplasma* pneumonia occurs for every 1000 persons, or more than 2 million cases annually. The incidence of *Mycoplasma* nonpneumonic respiratory infection may be 10 to 20 times greater. The highest attack rates are in individuals 5 to 20 years old, but *M. pneumoniae* infection can occur at any age and may cause particularly severe disease in neonates.

As opposed to viral respiratory infections that peak in winter in temperate climates, a few studies have reported a peak incidence of *M. pneumoniae* outbreaks in the fall. Most surveys, however, show little or no seasonal predominance in sporadic cases. There is an age-related relationship of upper versus lower respiratory tract infection caused by *M. pneumoniae.* In children younger than 3 years, primarily upper respiratory tract infection develops, whereas in those 5 to 20 years old, bronchitis and pneumonia tend to occur. In older adults, pneumonia predominates.

The prevalence and incidence of clinical urogenital disease caused by *Ureaplasma urealyticum*, *Ureaplasma parvum*, *M. hominis*, and *M. genitalium* are much less well documented (Table 317-1). These organisms are rarely cultured outside the realm of clinical studies, and they exist as clinically inapparent commensals of the genitourinary tract. Diseases attributed to *Ureaplasma* species include urinary tract infection with and without calculus formation. The organism has been implicated as a cause of low birthweight in neonates. *M. hominis* is a common genitourinary and oral commensal as well and has been documented as a cause of endometritis and postpartum fever.

M. hominis has also caused sternal wound infection after cardiothoracic surgery and has been implicated in arthritis in immunocompromised patients. *Mycoplasma salivarium* may be involved in periodontal disease, and *M. genitalium* is implicated in some cases of nongonococcal urethritis and vaginitis.

TABLE 317-1 SITES AND INFECTIONS RELATED TO HUMAN MYCOPLASMAS

SUBGROUP	SITES OF ISOLATION	DISEASES	OCCURRENCE
M. hominis	GU tract (F > M)	Cervicitis, vaginitis, ?prostatitis	Common
	Conjunctivae (neonate)	Conjunctivitis	
	Blood (peripartum)	Peripartum sepsis	
	Surgical wounds, joints	Sternotomy infection, arthritis	
M. orale	Oropharynx	?	Common
M. pneumoniae	Respiratory tract	URI, pneumonia	Common
M. salivarium	Oropharynx, gingiva	?Periodontal disease	Common
M. fermentans	GU tract, blood, tissues	Multisystem disease in AIDS	Uncommon
M. genitalium	GU tract	Urethritis, cervicitis, PID	Uncommon
Ureaplasma spp	GU tract	Urethritis, upper GU infection	Common

AIDS = acquired immunodeficiency syndrome; F = female; GU = genitourinary; M = male; PID = pelvic inflammatory disease; URI = upper respiratory tract infection.

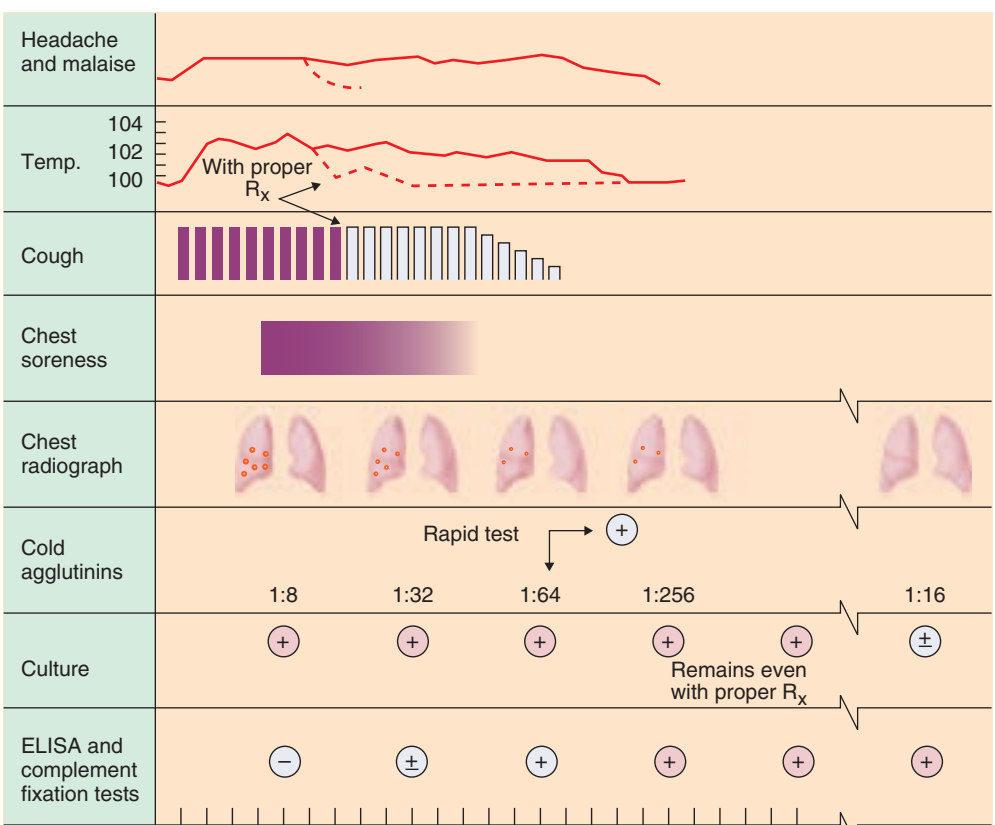

FIGURE 317-1. Major clinical and laboratory manifestations of mycoplasmal pneumonia. ELISA = enzyme-linked immunosorbent assay. (Modified from Baum SG. Mycoplasmal infections. In: Wyngaarden JB, Smith LH Jr, eds. *Cecil Textbook of Medicine*.17th ed. Philadelphia: Saunders; 1985:1506.)

M. fermentans, incognitus strain, was identified about four decades ago as an infectious agent in immunocompromised patients, in whom it causes overwhelming multisystem involvement.

PATHOBIOLOGY

Because of the low fatality of most *Mycoplasma* infections, there is little human pathologic material. Inoculation onto animal tracheal organ cultures is followed by ciliary damage and desquamation of surface epithelium. This latter effect is probably responsible for the hacking cough in *Mycoplasma* respiratory infection.

Several characteristics of *M. pneumoniae* probably play a direct role in the respiratory pathogenicity of this organism. The first is the affinity of *M. pneumoniae* for respiratory epithelial cells. Attachment appears to be between a terminal organelle at one end of the filamentous organism and a sialated glycoprotein (I-FI) on the surface of both respiratory epithelium and erythrocytes that acts as a receptor. *M. pneumoniae* attaches to ciliated epithelial cells at the base of cilia and appears to produce most of its physiologic and cytolytic changes while remaining extracellular. Hydrogen peroxide produced by *M. pneumoniae* (the only human mycoplasma to do so) may be responsible for some in vivo cell damage, as it is for the hemolysis seen when the organisms are grown on blood agar plates. *Mycoplasma* infection activates T and B cells and induces many pro-inflammatory and anti-inflammatory cytokines and chemokines that may play a role in inflammation-related cell destruction. Strain-specific elaboration of biofilms probably plays a role in protecting the organism from host immune cells and may decrease antimicrobial penetration.

In the course of *M. pneumoniae* infection, some patients produce cold agglutinins. These oligoclonal M-type immunoglobulins (IgM) cross-react with I antigens, one of the blood group antigens common to almost all mature human erythrocytes, and high titers may cause hemolysis (presumably as a result of complement-activated, Coombs-positive erythrocyte destruction) and lead to some of the complications described in the Clinical Manifestations section. Like other IgM antibodies, the *Mycoplasma*-induced cold agglutinins (Chapter 160) develop early in the disease (7 to 10 days) and are often present by the time the patient seeks medical attention. The titer of these agglutinins peaks at 2 to 3 weeks, and they persist for 2 to 3 months (Fig. 317-1).

Several theories about the factors triggering the formation of cold agglutinins in *Mycoplasma* pneumonia have been proposed. One is that the organism alters the I antigen such that it becomes antigenic to the patient. The hydrogen peroxide elaborated by *M. pneumoniae* could be responsible for this alteration. One study indicated that the I antigen in a sialated state may serve as a receptor for *M. pneumoniae* and that the cold agglutinins are directed at the modified receptor. Other studies indicate that the cold agglutinins are directed at mycoplasmal substructures themselves and merely cross-react with the I antigen (altered or native) on red cells. Given their apparent target, these antibodies could either contribute to cytolysis and exacerbate infection or interfere with cell-to-cell spread by blocking or disrupting the cell receptor for the mycoplasma. High titers of cold agglutinins have also been associated with hemolysis, presumably as a result of the activation of complement-mediated erythrocyte destruction (Chapter 160). Although clinically significant hemolysis is uncommon, subclinical levels of red cell destruction are common.

Sickle cell disease, sickle-related hemoglobinopathies, and hypogammaglobulinemia predispose to increased severity and mortality. Therefore, some of the available pathologic data may be influenced by the pathophysiologic mechanisms of these underlying conditions. Deaths have occurred in patients with diffuse pneumonia, acute respiratory distress syndrome, thromboembolism, and disseminated intravascular coagulation.

In nonfatal cases in which lung biopsy was performed, the inflammatory process involved primarily the trachea, bronchioles, and peribronchial tissue. The lumen of the respiratory tree was filled with a purulent exudate rich in polymorphonuclear leukocytes. The lining of the bronchial and bronchiolar walls showed metaplastic cells, and the walls themselves were infiltrated with monocytic elements, especially plasma cells. There was widening of the peribronchial septa and hyperplasia of type II pneumocytes.

CLINICAL MANIFESTATIONS

In view of the very high incidence of *Mycoplasma* respiratory infection when it is studied epidemiologically in large populations, versus the rarity of individual sporadic diagnoses, it appears that a specific, laboratory-confirmed diagnosis of this entity is seldom accomplished in routine clinical practice. There are probably four reasons for this. The first is that *Mycoplasma* pneumonia is usually self-limited and rarely fatal. This fact dampens the zeal to establish the cause of infection. Second, mycoplasmas are relatively fastidious and slow growing; therefore, culture results, if obtained at all, often return after the patient has recovered. Third, *M. pneumoniae* responds to the empirical antimicrobial therapy suggested for community-acquired pneumonia. Finally, knowledge of the epidemiology and clinical manifestations of infection is deficient, so the diagnosis is often not considered outside the classic age group.

Respiratory Infection

The majority of *M. pneumoniae* infections involve only the upper respiratory tract. After a 2- to 3-week incubation period, the disease has an insidious onset consisting of fever, malaise, headache, and cough (see Fig. 317-1). Cough is the clinical hallmark of *M. pneumoniae* infection. The frequency and severity of the cough increase during the next 1 to 2 days, and it may become debilitating. The gradual onset of symptoms is in contradistinction to the often fulminant manifestation of respiratory infection caused by influenza virus or adenovirus. Unfortunately, no clinical signs or symptoms reliably differentiate *M. pneumoniae* infections from other community-acquired pneumonias.[1]

In 5 to 10% of patients, depending somewhat on age, the infection progresses to tracheobronchitis or pneumonia. In these cases, the original manifestations persist, and the cough becomes more severe. It is usually relatively nonproductive but may yield white or occasionally blood-flecked sputum. Gram staining of this sputum reveals inflammatory cells but no predominant bacterial species (part of the definition of atypical pneumonia). With continued cough, parasternal chest soreness may develop as a result of muscle strain, but true pleuritic pain is unusual. Fever is usually at the level of 101° to 102° F and may be associated with chilly sensations. As opposed to pneumonia caused by *Streptococcus pneumoniae* (Chapter 289), that caused by *M. pneumoniae* rarely produces true shaking chills. In comparison to influenza, which can also be manifested as an atypical pneumonia syndrome, myalgias and gastrointestinal complaints of nausea and vomiting are unusual. Diarrhea, sometimes a concomitant of adenoviral or *Legionella* pneumonia, is uncommon in *Mycoplasma* infection.

On physical examination, the patient does not appear to be terribly ill. In fact, this disease is the paradigm of the term *walking pneumonia*. The pharynx may be injected and erythematous, usually without the marked cervical adenopathy seen with group A streptococcal pharyngitis. *M. pneumoniae* is not a common cause of isolated pharyngitis in the pediatric or adult population. Much has been made of the finding of bullous myringitis in this disease. Although this abnormality was present in about 20% of volunteers with experimentally induced *M. pneumoniae* infection, true bullous myringitis in naturally occurring *Mycoplasma* disease is rare. In a study involving children, otitis was rarely associated with isolation of *Mycoplasma* and, on the contrary, was often associated with bacterial and viral upper respiratory tract pathogens. Thus, the absence of myringitis, bullous or otherwise, should not dissuade one from a diagnosis of *Mycoplasma* pneumonia.

Examination of the chest in patients with *Mycoplasma* pneumonia is often unrevealing, even in those with severe, productive cough. There may be no auscultative or percussive findings, or only minimal rales (crackles) may be present. The disparity between physical findings and radiographic evidence of pneumonia in this condition may be the greatest of any of the atypical pneumonia syndromes (Fig. 317-2). Although wheezing can occur in this disease, in one study of asthmatic patients, the presence of wheezing had a negative correlation with isolation of *M. pneumoniae* compared with isolation of viral respiratory pathogens such as respiratory syncytial virus (Chapter 362). *M. pneumoniae* does not seem to be a common pathogen in patients with preexisting chronic obstructive lung disease, either. Bacterial superinfection after *M. pneumoniae* respiratory infection is rare. The radiographic finding of interstitial or patchy alveolar pneumonia does not allow differentiation from any of the other causes of the atypical pneumonia syndrome.

Pleural effusion (usually small) occurs in 5 to 20% of patients with *M. pneumoniae* pneumonia. This low incidence of pleural inflammation is consistent with the rarity of pleuritic pain. If effusion is present, thoracentesis reveals serous fluid that is exudative, with minimal inflammatory reaction.

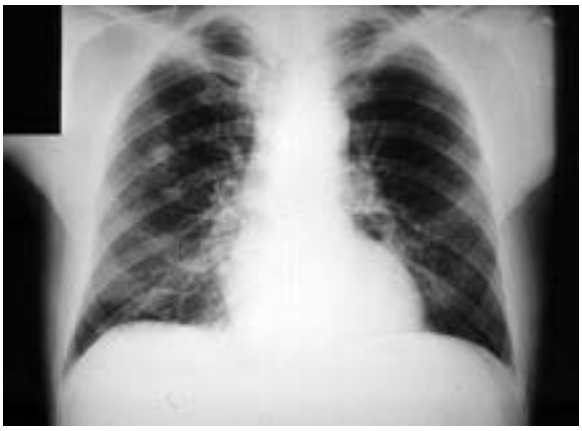

FIGURE 317-2. Chest radiograph showing moderate interstitial pneumonia due to *Mycoplasma pneumoniae* infection in a patient with a paucity of findings on chest auscultation.

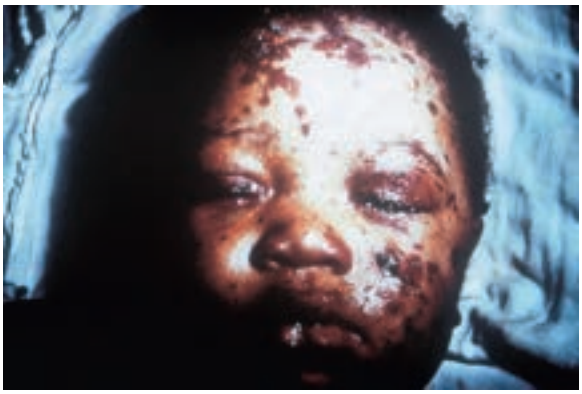

FIGURE 317-3. Stevens-Johnson syndrome in a child with *Mycoplasma* pneumonia. (From Baum SG. *Mycoplasma pneumoniae* and atypical pneumonia. In: Mandell GL, Bennett JE, Dolin R, eds. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*. 6th ed. Philadelphia: Churchill Livingstone; 2005:2274.)

The cell differential count in the fluid is variable, and bloody effusions are rare. It is unusual to isolate *M. pneumoniae* from effusions when they do occur, but several reports of such isolation exist. Although the pneumonia is generally mild and self-limited, fulminant, severe, and lethal cases have been reported in normal young adults and may be underdiagnosed.

Extrapulmonary Involvement

Abnormalities in almost every organ system have been described as examples of the extrapulmonary manifestations of *M. pneumoniae* infection. The frequency of these extrapulmonary manifestations varies greatly from one report to another, and they are much less common when viewed as part of a prospective epidemiologic study rather than as the sum of isolated case reports. The conclusion appears to be that the high prevalence of *Mycoplasma* infection predisposes to the reporting of many concurrent but perhaps unrelated events as though they were part of the mycoplasmal disease. Several clinical syndromes have been reported with sufficient frequency to provide some support for a causal relationship.

Dermatologic Involvement

A wide variety of transient dermatologic conditions have been reported in conjunction with *Mycoplasma* pneumonia, including macular, morbilliform, and papulovesicular eruptions as well as erythema nodosum and urticaria. Again, the variety and high incidence of these rashes in the absence of *Mycoplasma* infection make it difficult to define the relationships, if any, among these occurrences. Furthermore, the role of concurrent antibiotic therapy in the development of exanthems during *M. pneumoniae* infection is unknown. One skin condition that occurs often enough in concert with *M. pneumoniae* infection to provide some basis for relatedness is erythema multiforme majus, or Stevens-Johnson syndrome (Fig. 317-3). This rash has been reported in up to 7% of patients with *Mycoplasma* pneumonia and consists of erythematous vesicles, plaques, and bullae involving the skin, with particular localization at mucocutaneous junctions. The conjunctivae as well as

organs of the gastrointestinal and genitourinary tracts and the joints may also be involved. Stevens-Johnson syndrome has been associated with isolated cases of many other infections, including some that can be manifested as atypical pneumonia syndrome, such as legionnaires' pneumonia, adenovirus respiratory-conjunctivitis syndrome, and influenza B infection. However, the association with *M. pneumoniae* infection is by far the most common. Stevens-Johnson syndrome tends to occur in younger patients with *Mycoplasma* pneumonia and has a definite male preponderance (2 : 1 to 4 : 1).

The pathogenesis of Stevens-Johnson syndrome is unclear. It has long been supposed that immunity plays a major role, but several reports have noted the culture of *M. pneumoniae* from the lesions. The relationship to the level of cold agglutinins in this disease is variable. It has been suggested that the development of Stevens-Johnson syndrome may be the result of augmented sensitivity to antibiotics in the presence of *M. pneumoniae* infection, but erythema multiforme majus develops in some patients in the absence of previous or concurrent antibiotic therapy. Most patients clear the lesions in 1 to 2 weeks without scarring unless impetiginization supervenes.

A debilitating form of *M. pneumoniae*–associated mucositis without skin involvement has been described as an atypical Stevens-Johnson syndrome, but it is now increasingly recognized as a separate entity, termed *M. pneumoniae*–associated mucositis.[2]

Raynaud's syndrome (Chapter 80), a transient, reversible vasospasm of the digits that develops on exposure to cold, is not technically a dermatologic syndrome; however, it is manifested in the skin. This phenomenon occurs in many people, usually women, without any association with infection. It has been reported in patients with acute *Mycoplasma* pneumonia, regardless of whether they exhibited the syndrome before infection. Although the pathophysiologic mechanism of this condition in *M. pneumoniae* infection is unclear, it may be related to the in vivo action of cold agglutinins (see Diagnosis, later). Other vascular complications reported in *M. pneumoniae* infection include internal carotid artery occlusion and cerebral infarction.

Cardiac Complications

Cardiac abnormalities are among the most commonly reported extrapulmonary manifestations of *M. pneumoniae* infection. Signs and symptoms suggesting involvement of the heart are arrhythmia, congestive failure, chest pain, and electrocardiographic abnormalities, particularly conduction defects. Although cardiac abnormalities have been reported in as many as 10% of cases of *M. pneumoniae* infection, other reports indicate a much lower prevalence. Cardiac complications are more common with increasing age. They prolong the illness but rarely lead to death. The mechanism of heart damage is unknown, but *M. pneumoniae* has been isolated from the pericardial fluid of at least one patient.

Neurologic Complications

Neurologic complications are said to occur in about 0.1% of *Mycoplasma* infections, but proof of the cause of central nervous system involvement is somewhat tenuous. Central nervous system involvement is most common in children, with encephalitis the most common and perhaps the most devastating.[3] Aseptic meningitis and meningoencephalitis, transverse myelitis, brain stem dysfunction, Guillain-Barré syndrome, cerebral ataxia, and peripheral neuropathy have all been reported. Cerebrospinal fluid findings in these cases are variable, but the cellular response is usually minimal, with slightly elevated protein level and normal to slightly depressed glucose concentration. Most often, the diagnosis of *Mycoplasma*-related central nervous system involvement is based on the exclusion of other causes, the presence of an antecedent or intercurrent respiratory illness, and a rise in antibody titer to *M. pneumoniae* in serum. On occasion, *Mycoplasma*-specific antibodies have been demonstrated in cerebrospinal fluid, but these titers have paralleled serum antibody titers. Neurologic complications are usually reversible; however, mortality in patients with central nervous system involvement is higher than in those without such involvement. Although *M. pneumoniae* has been isolated from a few of these patients, polymerase chain reaction (PCR) failed to detect *M. pneumoniae* DNA in the cerebrospinal fluid of 11 patients deemed to have *M. pneumoniae*–related central nervous system disease on serologic grounds. Therefore, immune mechanisms of neural damage have been suggested. Some mycoplasmas elaborate a neurotoxin, but this has not been described for *M. pneumoniae.*

Musculoskeletal, Renal, and Hematopoietic Complications

Polyarthralgias are common in *Mycoplasma* pneumonia, but monarticular or migratory arthritis is rare. Immune mechanisms have been postulated, but there have been a few reports of the isolation of *M. pneumoniae* from joint fluid. Several of the cases of frank arthritis were reported in patients with hypogammaglobulinemia. Nonhuman mycoplasmas probably cause arthropathy in several animal species, and *M. hominis* has been associated with human arthritis. Rhabdomyolysis has also been increasingly reported and may be associated with very high muscle enzyme, interleukin-18, and tumor necrosis factor-α levels.[4]

Renal complications associated with immune complex deposition and high-titer cold agglutinins have been reported. There are several case reports of *M. pneumoniae*–associated aplastic anemia.

Conditions Leading to Increased Susceptibility

Several reports have emphasized the unusual severity of *M. pneumoniae* infection in patients with sickle cell disease or sickle-related hemoglobinopathies (Chapter 163). Large pleural effusions and marked respiratory distress may develop in these patients. Functional asplenia and its attendant opsonization deficiencies may contribute to overwhelming infection with *M. pneumoniae*, as they do with *S. pneumoniae* infection. Some patients with sickle cell disease and *M. pneumoniae* infection in whom extremely high cold agglutinin titers develop may experience digital necrosis, as they do with *S. pneumoniae* infection. A hypothesis about the pathogenesis is provided in the discussion on cold agglutinins in the Diagnosis section. Children with immunodeficiency syndromes have been the subjects of case reports of *M. pneumoniae* infection. Because *Mycoplasma* infections are so common in normal children, the contribution of immunodeficiency is unclear. *M. pneumoniae* is not a very common opportunistic agent in patients with acquired immunodeficiency syndrome (AIDS), but *M. fermentans* (*incognitus* strain) has been identified in these patients. Unusually severe but nonlethal *M. pneumoniae* infection has also been reported in children with Down syndrome.

DIAGNOSIS

The diagnosis of *Mycoplasma* pneumonia is made primarily on clinical grounds. The organism can be grown in cell-free media, but most hospital diagnostic laboratories are not prepared to culture mycoplasmas. Because of this, there is considerable interest in developing rapid diagnostic tests with high sensitivity and specificity. These assays fall into three categories: detection of *M. pneumoniae*–specific immunoglobulins in serum and detection of *M. pneumoniae*–specific antigens or nucleotide sequences directly in clinical specimens. Diagnostically, the most useful *M. pneumoniae*–specific immunoglobulin to detect is IgM because it is most likely to indicate recent infection. Enzyme-linked immunoassays have been developed to detect IgM and IgG directed against *M. pneumoniae*. When used in patients with positive assays for complement-fixing antibodies, the enzyme immunoassay had a specificity of more than 99% and a sensitivity of 98%. Specificity was retained but sensitivity dropped to only 46% when IgG alone was the target. Variations on this theme detect IgM antibodies directed at specific *M. pneumoniae* antigens. The tests are simple to perform and have high sensitivity and specificity, but because all these tests are designed to detect IgM antibody, results may be negative early in the course of infection (<7 to 10 days). Therefore, they do not provide the desired confirmation early enough to guide initial therapy.

Detection of *M. pneumoniae* antigens directly in sputum specimens has been accomplished with the use of an antigen capture, indirect enzyme immunoassay. The specificity of the assay was high, the reagents reacting only with *M. pneumoniae* and *M. genitalium.* Sensitivity was also relatively high (91%) when the assay was used on sputum and nasopharyngeal aspirates from patients who were shown by culture or serologically to have *M. pneumoniae* infection.

The high sensitivity and specificity of real-time PCR performed on sputum, nasopharyngeal aspirate, or throat swab material suggest that this test could serve as a specific and rapid diagnostic method. Detection of *M. pneumoniae*–specific nucleotide sequences directly in clinical material has been accomplished with the use of tests developed locally or by large reference laboratories. These tests, which can be completed in a few hours, detect either *M. pneumoniae* DNA (PCR) or ribosomal RNA (reverse transcriptase nucleic acid amplification).[5] Compared with PCR as the "gold standard" of proven infection, very few of the available antibody assays have acceptable sensitivity and specificity. Although the nucleic acid probe assays show excellent specificity, they fail to detect infection in some antibody-positive patients, and their sensitivity varies considerably, depending on the source of the specimen tested: highest for sputum, but lower for nasopharyngeal aspirate or throat swab material. This latter point is important in evaluating comparative studies, many of which use different specimen sources for their assays.[6]

In the course of *M. pneumoniae* infection, several classes of antibody are produced. Some of these fulfill the desired role of antibody production—neutralization of the agent—and others appear to be autoantibodies. The latter include agglutinins to lung, brain, cardiolipins, and smooth muscle. The best studied of these autoagglutinins are the cold isohemagglutinins. These agglutinins were found to be capable of clumping erythrocytes at 4° C. Agglutination was reversible by warming the serum-erythrocyte mixture to 37° C and, unlike hemagglutination by myxoviruses and paramyxoviruses, was repeatable with the same sample, indicating that receptor-destroying enzyme (neuraminidase) plays no role in the dissociation at 37° C.

The cold agglutinin phenomenon, when positive, occurs early, but it is neither sensitive nor specific. Other diseases that can give rise to cold agglutinins are mononucleosis caused by Epstein-Barr virus (anti-i), cytomegalovirus infection (anti-I), some other viral diseases, and lymphoma. A titer of 1 : 32 or greater is highly suggestive of infection with *M. pneumoniae.*

When this test is performed in a laboratory, the results will not be available for at least a day and in some cases a week, but a rapid version can be done at the bedside. In this test, 1 mL of the patient's blood is drawn into a tube containing anticoagulant. The type of tube used to collect specimens for prothrombin determination is preferred. Before cooling, examination shows a smooth coating of the tube by red cells. The blood is cooled to 4° C by placing it on liquid ice or in a standard refrigerator. After several minutes, the tube is examined for the presence of macroscopic erythrocyte agglutination. The tube is then rewarmed to 37° C in an incubator or by exposure to body heat and reexamined. The agglutination should dissociate at 37° C, and the appearance should be as it was before cooling. This temperature-associated agglutination and dissociation can be repeated many times on the same sample. A positive result in the bedside test correlates with a laboratory titer of 1 : 64 or greater and is therefore less sensitive than the laboratory test. It can be accomplished in minutes, however, and if positive, it is highly suggestive of *Mycoplasma*-related cold agglutination. The presence of cold agglutinins can also artifactually give rise to macrocytic indices secondary to in vitro clumping of erythrocytes, as measured by the Coulter counter method. In this case, the red cell distribution width would be high, indicative of heterogeneity in measured red cell size.

Finally, PCR should become the gold standard for diagnosis because of its high sensitivity and specificity and the rapidity with which results can be obtained. A multiplex assay kit for this test has been approved for rapid diagnosis of *M. pneumoniae* infection. However, the test may not be available in many medical centers. Widespread availability of real-time definitive diagnostic assays will enhance the ability to apply organism-specific treatment.[7]

TREATMENT

Rx

Despite the number and variety of tests for the diagnosis of *M. pneumoniae* infection, most cases are encountered in the ambulatory setting, and the institution of antimicrobial therapy is empirical and based on clinical recognition of the syndrome.

Antimicrobial therapy is not necessary for mycoplasmal upper respiratory tract infection, and the mycoplasmal cause of this syndrome most often goes undiagnosed. The pneumonia caused by *Mycoplasma* is self-limited and not life-threatening in most cases. However, treatment with effective antimicrobials may shorten the duration of illness[8] and, by reducing cough and the number of organisms per unit volume of sputum, may reduce the spread of infection to contacts. Because *Mycoplasma* pneumonia is sporadic, most of the studies of antimicrobial efficacy are directed not at this single agent but rather at the treatment of community-acquired pneumonia, of which *Mycoplasma* makes up a variable percentage. Retrospective analysis of specific causes then allows microbe-specific efficacy.

As would be predicted by the lack of a cell wall, all mycoplasmas, including *M. pneumoniae*, are unaffected by β-lactam antibiotics such as the penicillins and cephalosporins. Aminoglycosides are effective in vitro but have not been evaluated for efficacy in vivo. The mainstays of treatment of *M. pneumoniae* respiratory tract infection are macrolides and tetracyclines. Use of either drug class significantly shortens the duration of illness. The radiographic findings may take a week or longer to resolve, even with appropriate therapy (see Fig. 317-1), and organisms may continue to be cultured from sputum for several weeks after a complete course of clinically effective treatment. This may be a result of the fact that although *M. pneumoniae* causes respiratory disease as an extracellular parasite, it has the capacity to reside intracellularly as well, thus making it difficult to eradicate the organism in vivo as opposed to cell cultures. The effect of therapy on extrapulmonary manifestations is unknown. *M. hominis* is not sensitive to erythromycin.

Although the tetracyclines are very active against *M. pneumoniae*, their use is precluded in young children because of adverse effects on developing teeth and bones. Erythromycin is poorly tolerated by many people because of its gastrointestinal side effects, including nausea, vomiting, abdominal pain, and diarrhea. Erythromycin also raises theophylline levels, a consideration in the few asthmatic patients who may still be taking this drug.

Because of the adverse effects of erythromycin and tetracycline, there is considerable interest in the antimycoplasmal efficacy of other agents. Doxycycline is somewhat better tolerated than tetracycline and can be administered in two daily doses rather than three. In vitro, doxycycline is as effective as tetracycline against *M. pneumoniae*, but it is contraindicated in young children.

Several other classes of antimicrobials have significant in vitro and in vivo activity against *M. pneumoniae* and other *Mycoplasma* species, including the fluoroquinolones, broad-spectrum macrolides (azithromycin, clarithromycin), and members of the macrolide-lincomycin-streptogramin-ketolide (MLSK) class of antimicrobials. There are no good data on the optimal duration of therapy needed to minimize carriage and relapse with these agents.

The macrolides are more active in vitro than the tetracyclines are. Fluoroquinolones have adequate activity for the treatment of these infections. They are more active than the tetracyclines but are at least 100 times less active than the macrolides. The streptogramins are also less active than the macrolides but more active than the tetracyclines. There is a significant cost differential in the use of these drugs. The newer macrolides and quinolones are 50 to 60 times more expensive than the tetracyclines and 6 to 10 times more costly than erythromycin.

During the past decade, there have been increasing reports, first from Asia and later spreading to other areas including the United States,[9] describing decreasing susceptibility of *M. pneumoniae* isolates to the macrolides. By 2012, one study from China reported 95% macrolide resistance. Taking all this into account, recommended therapy includes doxycycline 100 mg every 12 hours in older children and adults,[10] or azithromycin 500 mg on day 1 and then 250 mg every 24 hours. Young children should be given erythromycin (10 mg/kg every 6 hours) or an extended-spectrum macrolide (10 to 12 mg/kg on day 1, followed by 5 mg/kg/day). The usual duration of therapy is 14 days; shorter courses may lead to relapse. The addition of corticosteroids (e.g., prednisolone 1 mg/kg twice daily for three days then tapered over one week) has been reported to shorten the duration of symptoms in patients with severe or refractory infections, especially when the serum lactate dehydrogenase level is above about 300 IU/L.[11] *M. hominis* is not sensitive to macrolides, but it is sensitive to the other antimicrobials recommended for *M. pneumoniae* infection. Azithromycin (1-g single dose) achieves a significantly higher cure rate for *Mycoplasma genitalium*–associated urethritis than does multidose doxycycline.

PREVENTION

Outbreaks of *M. pneumoniae* respiratory infection among military recruits led to interest in producing a vaccine to protect against this organism. The resulting vaccines induced specific antibody responses, but protection was limited to no more than 50% of vaccine recipients. Live vaccines using attenuated wild-type and temperature-sensitive mutant mycoplasmas have proved no more effective. In one study, volunteers who received vaccine but did not mount an antibody response had more severe disease when rechallenged with wild-type *Mycoplasma* than did nonvaccinated personnel.

Although *M. pneumoniae* is perhaps the leading cause of atypical pneumonia syndrome in closed populations, the enthusiasm for developing a vaccine for this disease has waned. New technology involving DNA expression library immunization has proved successful in animal studies with nonhuman mycoplasmas, and these methods may breathe new life into *M. pneumoniae* vaccine development.

Secondary Prevention

Prophylactic antibiotic use in family members exposed to *Mycoplasma* decreases clinical disease in these patients, but seroconversion is not prevented. One study showed that azithromycin prophylaxis, given as a 500-mg loading dose and 250 mg/day on days 2 through 5, significantly reduced the secondary attack rate of *M. pneumoniae* infection in a long-term care facility.

Grade A References

A1. Bradley JS, Arguedas A, Blumer JL, et al. Comparative study of levofloxacin in the treatment of children with community-acquired pneumonia. *Pediatr Infect Dis J*. 2007;26:868-878.

A2. Mulholland S, Gavranich JB, Gillies MB, et al. Antibiotics for community-acquired lower respiratory tract infections secondary to *Mycoplasma pneumoniae* in children. *Cochrane Database Syst Rev*. 2012;9:CD004875.

A3. Huang L, Gao X, Chen M. Early treatment with corticosteroids in patients with *Mycoplasma pneumoniae* pneumonia: a randomized clinical trial. *J Trop Pediatr*. 2014;60:338-342.
A4. Luo Z, Luo J, Liu E, et al. Effects of prednisolone on refractory *Mycoplasma pneumoniae* pneumonia in children. *Pediatr Pulmonol*. 2014;49:377-380.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Wang K, Gill P, Perera R, et al. Clinical symptoms and signs for the diagnosis of *Mycoplasma pneumoniae* in children and adolescents with community-acquired pneumonia. *Cochrane Database Syst Rev*. 2012;10:CD009175.
2. Vujic I, Shroff A, Grzelka M, et al. *Mycoplasma pneumoniae*–associated mucositis—case report and systematic review of literature. *J Eur Acad Dermatol Venereol*. 2015;29:595-598.
3. Erol I, Ozkale Y, Alkan O, et al. Acute disseminated encephalomyelitis in children and adolescents: a single center experience. *Pediatr Neurol*. 2013;49:266-273.
4. Oishi T, Narita M, Ohya H, et al. Rhabdomyolysis associated with antimicrobial drug–resistant *Mycoplasma pneumoniae*. *Emerg Infect Dis*. 2012;18:849-851.
5. Di Marco E. Real-time PCR detection of *Mycoplasma pneumoniae* in the diagnosis of community-acquired pneumonia. *Methods Mol Biol*. 2014;1160:99-105.
6. Loens K, Goossens H, Ieven M. Acute respiratory infection due to *Mycoplasma pneumoniae*: current status of diagnostic methods. *Eur J Clin Microbiol Infect Dis*. 2010;29:1055-1069.
7. Poritz MA, Blaschke AJ, Byington CL, et al. FilmArray, an automated nested multiplex PCR system for multi-pathogen detection: development and application to respiratory tract infection. *PLoS ONE*. 2011;6:e26047.
8. Biondi E, McCulloh R, Alverson B, et al. Treatment of mycoplasma pneumonia: a systematic review. *Pediatrics*. 2014;133:1081-1090.
9. Diaz MH, Benitez AJ, Winchell JM. Investigations of *Mycoplasma pneumoniae* infections in the United States: trends in molecular typing and macrolide resistance from 2006 to 2013. *J Clin Microbiol*. 2015;53:124-130.
10. Lung DC, Yip EK, Lam DS, et al. Rapid defervescence after doxycycline treatment of macrolide-resistant *Mycoplasma pneumoniae*-associated community-acquired pneumonia in children. *Pediatr Infect Dis J*. 2013;32:1396-1399.
11. Miyashita N, Kawai Y, Inamura N, et al. Setting a standard for the initiation of steroid therapy in refractory or severe Mycoplasma pneumoniae pneumonia in adolescents and adults. *J Infect Chemother*. 2015;21:153-160.

REVIEW QUESTIONS

1. *Mycoplasma pneumoniae* is one of the most common causes of the atypical pneumonia syndrome. Another organism that shares many of the characteristics of organisms causing this syndrome is

A. *Streptococcus pneumoniae*
B. *Chlamydophila pneumoniae*
C. *Staphylococcus aureus*
D. *Klebsiella pneumoniae*
E. *Acinetobacter baumannii*

Answer: B *Chlamydophila* does not show up on Gram stain or grow in common bacterial culture, two of the common characteristics of these "atypical" organisms. (See Definition section.)

2. Which of the following classes of antimicrobials would *not* be effective in treating pneumonia in a young adult caused by *M. pneumoniae*?

A. Fluoroquinolones
B. β-Lactamase–resistant penicillins
C. Extended-spectrum macrolides
D. Tetracyclines
E. MLSK class

Answer: B Mycoplasmas do not have cell walls, and the penicillins work by inhibiting cell wall synthesis. The other drugs all are effective in treating *M. pneumoniae.* (See section on treatment.)

3. Which of the following is true of cold agglutinins that occur in about 50 to 70% of patients with *M. pneumoniae* pneumonias?

A. They are IgG antibodies directed against the nucleic acid of the organism.
B. They are secreted by the organism as part of the infectious process.
C. They help to neutralize the organism and thereby protect the host.
D. They are IgM class antibodies directed at antigens on the host erythrocyte.
E. When present, they are pathognomonic of *Mycoplasma* infection.

Answer: D Mycoplasma-induced cold agglutinins are IgM antibodies directed at the I antigens of red cell membranes. They are detected by a positive direct antiglobulin test (DAT) result, also known as the direct Coombs' test. Although clinically significant hemolysis is rare, subclinical positivity for DAT is common in *Mycoplasma* pneumonia. (See Pathobiology section.)

4. Which of the following is true about *Mycoplasma hominis*?

A. Causes puerperal sepsis
B. Is sensitive to macrolides, as are the other mycoplasmas
C. Is a common cause of infectious carditis
D. Is the most common cause of nonbacterial urinary tract infection
E. Has, as its primary reservoir in nature, the feral cat

Answer: A *M. hominis* is a common genitourinary and oral commensal, but it has also been documented as a cause of endometritis and postpartum fever. (See section on epidemiology.)

5. Which of the following chest radiographic patterns would be most characteristic of *M. pneumoniae* pneumonia?

A. Pneumothorax
B. Dense, multilobar alveolar infiltrates
C. Dense unilobar infiltrate with ipsilateral pleural effusion
D. No infiltrate seen
E. Multifocal interstitial pattern

Answer: E The disparity between the paucity of physical findings and the radiographic appearance of pneumonia in this condition can be striking (see Fig. 317-2).

318 DISEASES CAUSED BY CHLAMYDIAE

WILLIAM M. GEISLER

DEFINITION

Chlamydiae are obligate intracellular bacteria that cause a variety of human and animal diseases and much morbidity. Chlamydiae were originally classified taxonomically into one genus, *Chlamydia*. On the basis of sequence analysis of 16S rRNA genes, it had been proposed that chlamydial taxonomy be revised to contain two genera: *Chlamydia* and *Chlamydophila.* However, on the basis of additional data on chlamydia genome sequences and meetings within the *Chlamydia* scientific community, it has been agreed that the family Chlamydiaceae will contain a single genus, *Chlamydia*. Within the family are now nine recognized species: *C. trachomatis, C. pneumoniae, C. psittaci, C. pecorum, C. muridarum, C. felis, C. abortus, C. suis,* and *C. caviae*. *C. trachomatis* is classified into a trachoma biovar and lymphogranuloma venereum (LGV) biovar. This chapter is limited to human diseases caused by chlamydiae.

The Pathogen

Chlamydiae have a gram-negative cell wall structure consisting of an outer membrane that contains lipopolysaccharide and an inner cytoplasmic membrane. The outer membrane contains a single 40-kD major outer membrane protein, OmpA (also known as MOMP), and two cysteine-rich minor outer membrane–associated proteins (OmcA and OmcB); through intermolecular and intramolecular disulfide bonding, these proteins form a complex that provides structural rigidity.

Chlamydiae grow only within intracellular membrane-bound vacuoles, termed inclusions,[1] that seclude the organism from extracellular and cytoplasmic environments. They share a distinct biphasic developmental cycle (Fig. 318-1) that includes an extracellular, metabolically inactive, infectious form (an elementary body) and an intracellular replicative form (a reticulate body). In vitro studies have shown that chlamydiae may enter a persistent state under certain conditions (penicillin treatment,[2] challenge with certain cytokines, restriction of select nutrients) in which they have reduced metabolic activity and may be more refractory to antibiotic treatment; whether this occurs in vivo is unclear. Chlamydiae are unable to synthesize adenosine triphosphate and therefore depend on the host cell for nutrients to meet their energy requirements.

Additional knowledge of chlamydial cell biology emerged after several chlamydial genomes were sequenced. Polymorphic outer membrane proteins were identified, but their role in pathogenesis is unclear. Chlamydiae were also found to express proteins that localize to the cytoplasmic surface of the inclusion membrane (e.g., Inc A). *C. trachomatis* was demonstrated to have type III secretion genes, which may encode a secretion apparatus providing a means for protein transport. Chlamydiae have small genomes (*C. trachomatis* contains 894 protein-coding genes, and *C. pneumoniae* contains 1052 genes). Most strains of *C. trachomatis* and some strains of *C. psittaci* contain a 7-kilobase cryptic plasmid.

PATHOBIOLOGY

Macrophages are the principal host cells for *C. psittaci* and *C. trachomatis* LGV biovar, whereas the principal host cells for *C. trachomatis* trachoma biovar and *C. pneumoniae* strains are columnar epithelial cells at mucosal sites. Host cell tropism correlates with the type of inflammation elicited by chlamydiae. The LGV biovar and *C. psittaci* produce granulomatous inflammation, characteristic of delayed hypersensitivity reactions. The trachoma biovar produces neutrophilic exudate during acute infection and submucosal mononuclear infiltration with lymphoid follicle formation during later stages of infection.

Chlamydiae elicit both humoral and cellular immune responses. Infection can persist or recur even after an adaptive immune response develops, suggesting that the organism has evolved strategies for immune evasion. Persistent or recurrent infections can elicit inflammatory cellular immune responses that cause tissue injury.

CHLAMYDIAL DISEASES

Table 318-1 summarizes the diseases caused by chlamydiae in humans and the associated clinical and laboratory characteristics.

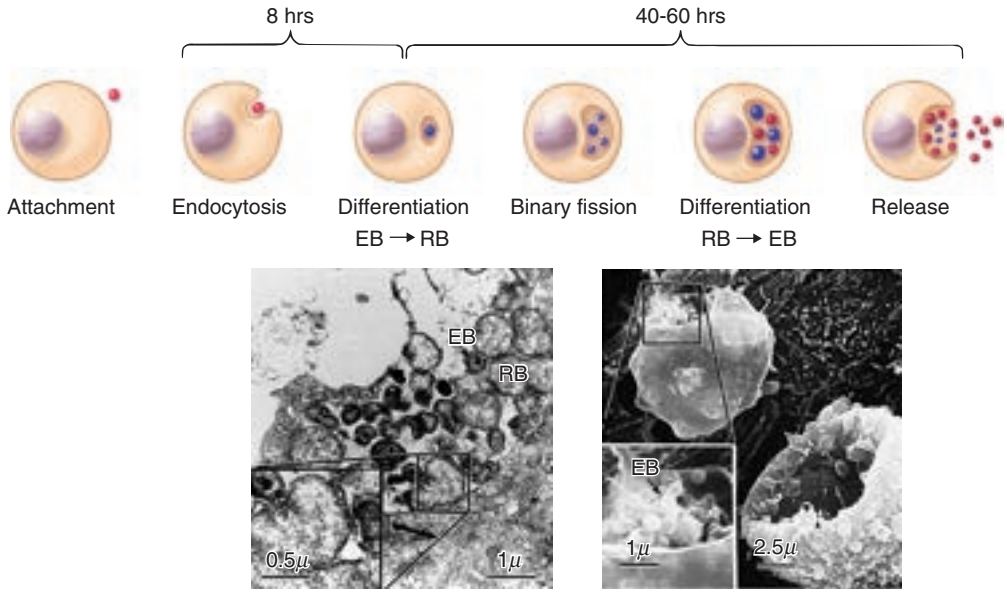

FIGURE 318-1. Developmental cycle of chlamydiae. The *top panel* shows the developmental cycle common to all chlamydiae. The *red circles* represent elementary bodies (EBs), and the *blue circles* represent reticulate bodies (RBs). Chlamydiae infect eukaryotic cells through multiple attachment mechanisms. After attachment, EBs enter the cell within a membrane-bound vacuole that remains unfused with lysosomes. EBs reorganize into RBs and asynchronously replicate 8 to 12 times, with a doubling time of 2 to 3 hours. At the conclusion of the growth cycle, RBs differentiate back to EBs, and each inclusion yields 100 to 1000 new infectious EBs. The *bottom left panel,* a transmission electron micrograph taken 40 hours after infection, shows the large RBs and the smaller EBs, which have a condensed nucleoid structure within their cytoplasm. The *bottom right panel,* a scanning electron micrograph taken 60 hours after infection, shows a membrane-bound vacuole containing many EBs exiting from an infected HeLa cell.

TABLE 318-1 MAJOR DISEASES CAUSED BY CHLAMYDIAE IN HUMANS AND ASSOCIATED CLINICAL AND LABORATORY FEATURES

SPECIES	SEROVAR	DISEASE	TRANSMISSION ROUTE	DIAGNOSIS	PREVENTION
C. trachomatis	A–C	Trachoma	Fomites, eye-seeking flies	Clinical criteria or culture/NAAT	SAFE strategy
	D–K	Urethritis, cervicitis, proctitis, epididymitis, PID	Sexual contact	NAAT	Abstinence or monogamy, education, condoms, partner treatment
	D–K	Inclusion conjunctivitis, infant pneumonia	Perinatal contact	Culture, DFA, NAAT, or serology (for pneumonia)	Prenatal chlamydia screening: treat infected mothers
	L1–L3	Lymphogranuloma venereum	Sexual contact	Serology or culture/NAAT plus OmpA typing	Abstinence or monogamy, education, condoms, partner treatment
C. pneumoniae	One	Upper respiratory infections, atypical pneumonia, asthma exacerbations	Respiratory droplet	Serology or culture/PCR	None
C. psittaci	Multiple	Psittacosis, atypical pneumonia, febrile illness	Aerosolized bird secretions, dust	Serology	Quarantine and chlortetracycline for imported birds, avoidance or precautions for at-risk subjects

DFA = direct fluorescent antibody test; NAAT = nucleic acid amplification test; PCR = polymerase chain reaction; PID = pelvic inflammatory disease; SAFE = World Health Organization's recommended strategy: acronym for surgery (for trichiasis), antimicrobials (periodic community-wide treatment), facial cleanliness, and environmental improvement.

Chlamydia trachomatis

C. trachomatis infection is a common bacterial infection in humans and accounts for significant morbidity worldwide. *C. trachomatis* isolates have been differentiated into 18 major serovars (i.e., OmpA types) based on variations in OmpA that are identified on antigen cross-reactivity in the microimmunofluorescence test. The major diseases caused by *C. trachomatis* are trachoma, caused by serovars A, B, Ba, and C; sexually and perinatally transmitted diseases, caused by serovars D through K (and, rarely, serovars B and Ba); and sexually transmitted LGV, caused by serovars L1, L2, L2a, and L3. Sequencing of the *ompA* gene has led to the recognition of more OmpA variants, including L2b. Multilocus sequencing typing is a newer tool that has been used to further discriminate between *C. trachomatis* strains of the same OmpA genotype. Trachoma and LGV are endemic in developing areas of the world (although LGV outbreaks have also occurred in populations of men who have sex with men in developed countries), whereas sexually and perinatally transmitted non-LGV chlamydial infections occur worldwide.

TRACHOMA

EPIDEMIOLOGY

Trachoma is a chronic follicular conjunctivitis. The overall incidence is unknown, but it has been estimated by the World Health Organization (WHO) that 21.4 million people have active trachoma.[3] Trachoma is endemic in more than 50 countries and is especially common in poor areas of sub-Saharan Africa, where the disease prevalence in children may exceed 40%. According to the Trachoma Atlas provided by the International Coalition for Trachoma Control, there are an estimated 110 million people living in endemic areas and another 210 million living in suspected endemic areas. Trachoma is a major public health problem because scarring from trachoma causes blindness, affecting 7.2 million people by WHO estimates. Trachoma is the most common preventable cause of blindness worldwide. Active trachoma often occurs in the first few years of life. The inflammation from recurrent or persistent trachoma can lead to conjunctival scarring, which can ultimately cause corneal damage and blindness later in life. The *C. trachomatis* serovars that produce trachoma are spread by direct contact with fingers or fomites (e.g., washcloths, handkerchiefs) contaminated with eye discharge from an infected person or by eye-seeking flies. Because of this mode of transmission, trachoma often clusters in households. Risk factors for trachoma include poor facial hygiene, limited access to water, poor sanitation, and proximity to other infected persons or to a heavy density of eye-seeking flies.

CLINICAL MANIFESTATIONS

There are two stages of trachoma disease, and they can overlap. Initially, trachoma begins as an inflammatory follicular conjunctivitis (i.e., active trachoma). On eversion of the upper eyelid, white to pale yellow lymphoid follicles can be visualized on the superior tarsal conjunctival surface, and papillae may be noted between follicles. Minimal watery or mucoid ocular discharge may also be seen. In more severe active trachoma, the conjunctiva can be thickened and edematous. Subsequently, conjunctival inflammation can progress to cause scarring of the upper tarsal conjunctiva (the cicatricial stage of disease). Scarring deforms the eyelid and can lead to an inward turning of the eyelashes, which can result in corneal abrasion (trichiasis). Over time, trichiasis causes corneal edema, ulceration, vascularization (pannus), scarring, and opacification. The corneal damage leads to decreased vision or blindness, occurring mostly in young adults and middle-aged persons. Viral conjunctivitis (e.g., adenovirus) is manifested in a clinical fashion similar to active trachoma, but it is self-limited and usually resolves within a week. Trachoma can be complicated by superinfection with other bacterial pathogens (e.g., *Haemophilus influenzae*), which should be considered when purulent ocular discharge or significant inflammation of the bulbar conjunctiva is present.

DIAGNOSIS

Because the majority of trachoma cases occur in developing countries without access to laboratory testing or the necessary resources, trachoma is often diagnosed clinically on the basis of findings of active trachoma (follicles on the upper tarsal conjunctiva or pronounced inflammatory thickening of the tarsal conjunctiva) or cicatricial disease. When laboratories are available, detection of *C. trachomatis* provides definitive evidence of active trachoma and may identify infection in subjects with minimal clinical evidence of active trachoma. Isolation of the organism in cell culture is one means to detect *C. trachomatis*, but the test's sensitivity is less than 50% and the methods are labor-intensive. Nonculture tests have higher sensitivities in active trachoma. For instance, nucleic acid amplification tests (NAATs) are the most sensitive diagnostic tests, but they are not widely available in many trachoma-endemic areas. It is uncommon for adults with late scarring to have *C. trachomatis* detected by any of these assays.

TREATMENT

Rx

Active trachoma can be treated with a tetracycline eye ointment twice daily for 6 weeks or oral macrolide therapy. The latter is preferred in part because extraocular sites such as the nasopharynx may be infected in young children. Single-dose oral azithromycin (20 mg/kg; maximum of 1 g) is as effective as tetracycline ointment and is more advantageous in terms of compliance and side effect profile; tetracycline application can irritate the ocular surface.

Surgical intervention is the only effective management for trichiasis. Eyelid rotation surgery prevents the eyelashes from abrading the cornea, preventing blindness and other nonvisual symptoms. Trichiasis recurrence after surgery is a major concern, and recurrence rates are highly variable across studies. Other concerns include accessibility to surgery and the patient's acceptance.

PREVENTION

The WHO is committed to eliminating blinding trachoma by 2020 and recommends that all countries with endemic trachoma adopt the SAFE strategy: surgery (for trichiasis), antimicrobials (periodic community-wide treatment), facial cleanliness, and environmental improvement. Mass treatment

of a community with single-dose oral azithromycin is safe and has dramatically reduced the prevalence of infection for up to 1 year after treatment, without an increase in antimicrobial resistance; it has also been shown to reduce mortality in children. Annual azithromycin treatment is recommended for trachoma-endemic areas. Reintroduction of trachoma after mass treatment has been demonstrated, which may be due in part to decreased herd immunity. Repeated mass treatment (every few months) provides herd protection to the entire community.

Mass antibiotic treatment as the sole intervention for eliminating trachoma is unlikely to be successful if other factors that facilitate transmission are not addressed. Face washing and good hygiene help reduce the risk of transmission through contact with fingers and flies. Although the promotion of facial cleanliness through educational campaigns may be one of the most important interventions, sustaining such behavioral changes can be challenging. Achieving better environmental conditions through measures that reduce household and community fly density and improve waste management and access to clean water can also limit transmission. Improvement in socioeconomic conditions in a community correlates with a decline in trachoma prevalence.

SEXUALLY AND PERINATALLY TRANSMITTED CHLAMYDIAL INFECTIONS

EPIDEMIOLOGY

Chlamydia is the most prevalent bacterial sexually transmitted infection in the United States. Currently, more than 1.4 million infections are reported to the Centers for Disease Control and Prevention (CDC) annually.[4] The number of reported cases is increasing each year; this may reflect increased chlamydia screening efforts rather than a true increase in infection burden, but this remains unclear. Taking into account underreporting and underscreening, it is estimated that more than 2.8 million new chlamydial infections occur annually in the United States. Higher chlamydia prevalence rates have been associated with younger age (sexually active adolescents and young adults), select minority ethnic groups (especially African Americans), and new or multiple sexual partners.[5] Chlamydia prevalence is higher in women than in men; it is unclear whether this is due to higher screening rates in women or whether women may be more susceptible to infection acquisition or persistence. Chlamydia prevalence is highest in the southeastern United States. The estimated total cost attributable to chlamydial disease in the United States exceeds $2.4 billion annually. From a global perspective, WHO estimated from 2005 data that there were approximately 98 million adults with chlamydia. In addition to the adverse effects on the reproductive health of women, chlamydia has a substantial impact on perinatal outcomes and facilitates the transmission of human immunodeficiency virus (HIV).

CLINICAL MANIFESTATIONS

Urethritis

C. trachomatis is the most common cause of nongonococcal urethritis in men,[6] being responsible for 20 to 55% of cases. Although the majority of men with chlamydial urethritis are asymptomatic, studies in high-prevalence venues (e.g., sexually transmitted disease clinics) have reported that 40 to 60% of men with chlamydial urethritis are symptomatic. The most frequent symptoms are urethral discomfort (itching or pain) with urination and urethral discharge. On examination, a mild to moderate amount of clear or cloudy/mucoid (rarely purulent) urethral discharge may be visualized (Fig. 318-2A). Urethral discharge sometimes becomes apparent only after stripping the urethra from the base of the penis to the glans; this should be considered in men reporting urethral symptoms but without urethral discharge on initial inspection. Gram stain from a urethral swab specimen demonstrates

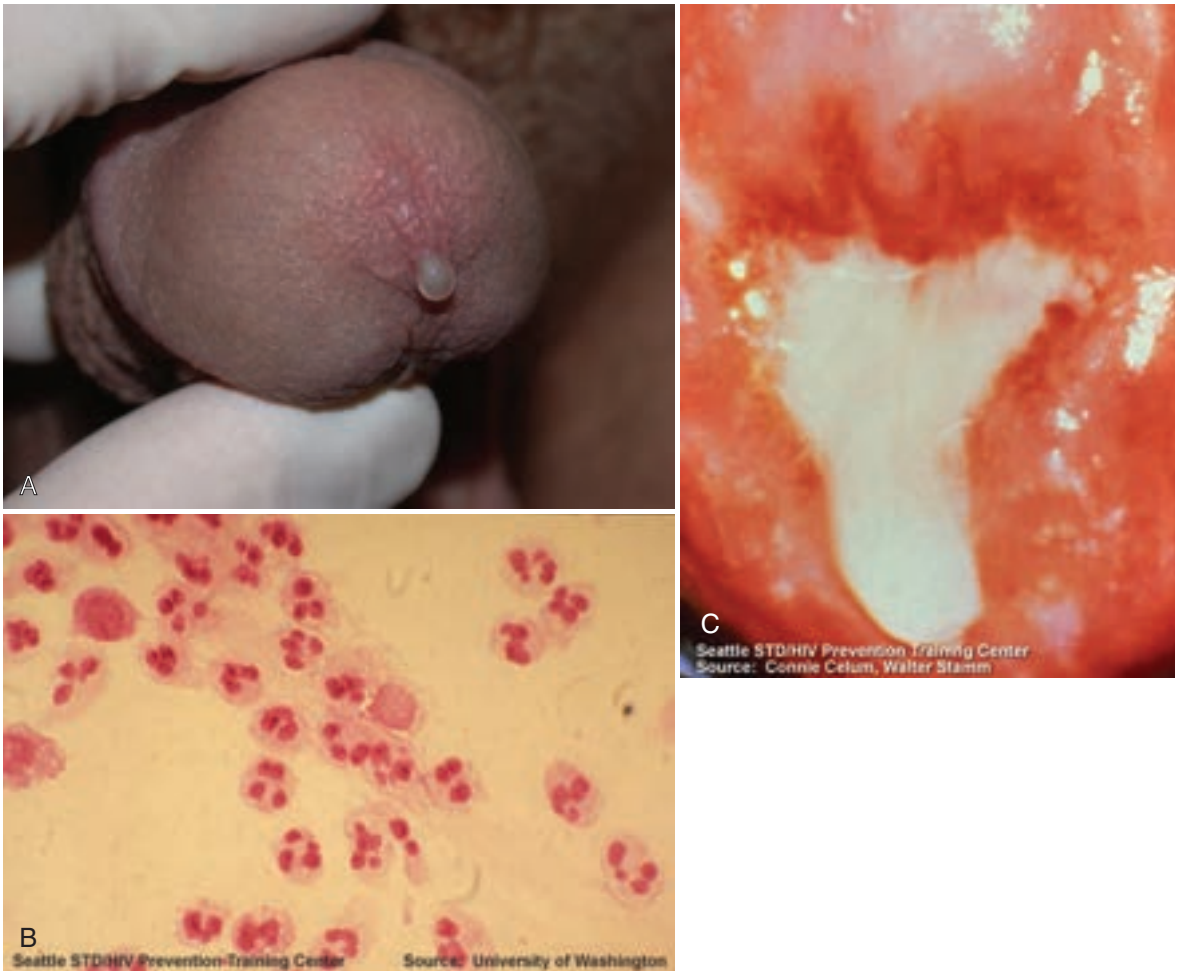

FIGURE 318-2. Clinical manifestations of genital *Chlamydia trachomatis* infection. A, Cloudy urethral discharge of urethritis. **B,** Urethral specimen Gram stain revealing nongonococcal urethritis findings: five or more polymorphonuclear cells per oil field (1000×) and the absence of intracellular gram-negative diplococci. **C,** Purulent endocervical discharge of mucopurulent cervicitis. (**A** courtesy James Sizemore, MD. **B** and **C** from *Practitioner's Handbook for the Management of Sexually Transmitted Disease*. http://depts.washington.edu/handbook. Accessed March 18, 2015.)

five or more polymorphonuclear leukocytes (PMNs) per oil field (1000×) in the majority of chlamydial infections (Fig. 318-2B). Chlamydial organisms cannot be visualized on Gram stain. Up to 20% of men with chlamydial urethritis have fewer than five PMNs per oil field on urethral Gram stain, reflecting the fact that minimal inflammation may be elicited in some cases of chlamydial urethritis. Urethral inflammation can also be detected by a positive urinary leukocyte esterase test result on unspun first-void urine.

C. trachomatis urethritis also occurs in women, who may be asymptomatic or present with an acute urethral syndrome characterized by dysuria, urinary frequency, and pyuria. This acute urethral syndrome mimics a urinary tract infection, and chlamydia should be suspected in women with pyuria but negative urine nitrite or negative urine culture, especially in sexually active adolescents and young adults; many urinary tract infection treatments are not effective against chlamydia. Mild urethral discharge may be seen. Pelvic examination should be performed in women with suspected chlamydia urethritis to search for other clinical findings of chlamydia (e.g., cervicitis).

Epididymitis

Chlamydia can spread from the urethra to the epididymis, causing epididymitis in up to 1 percent of infected men. Symptoms include testicular pain and scrotal erythema and swelling that are typically unilateral. On examination, there is palpable swelling and tenderness of the epididymis; accompanying findings may include testicular tenderness, scrotal erythema and swelling, urethral discharge, or hydrocele. In men younger than 35 years, *C. trachomatis* is the principal cause of epididymitis, whereas in men older than 35 years, complicated urinary tract infection with uropathogens is a more common cause, especially in those with prostate disorders. Up to 15% of epididymitis cases are complicated by chronic pain that is usually idiopathic and often unresponsive to antibiotics. Other complications include decreased fertility and, rarely, testicular abscess.

Cervicitis

C. trachomatis is the most common cause of cervicitis, being responsible for up to 50% of cases. The majority of women with endocervical chlamydial infection are asymptomatic. Symptoms, when present, are often mild and nonspecific for chlamydia and include the following: vaginal discharge, intermenstrual vaginal bleeding, dysuria, and pain during intercourse (dyspareunia). Approximately 10% of women with asymptomatic cervical chlamydial infection have mucopurulent cervicitis detected on pelvic examination (Fig. 318-2C); this is characterized by a purulent or cloudy endocervical discharge visible in the endocervical canal or on the tip of an endocervical swab. A similar proportion may have endocervical bleeding that is easily induced with passage of a swab through the cervical os. Nonspecific findings may include vaginal discharge and edematous ectopy (a darker red area of columnar epithelium visible on the face of the cervix). Gram stain of an endocervical sample usually shows more than 10 PMNs per oil field. A vaginal wet mount often shows more than 5 to 10 PMNs per 400× field.

Pelvic Inflammatory Disease

Chlamydia can spread from the cervix to the endometrium (causing endometritis), fallopian tubes (causing salpingitis), and peritoneum (causing peritonitis or perihepatitis). These upper genital tract infections are collectively referred to as pelvic inflammatory disease (PID). Estimates of the proportion of cervical chlamydial infections that progress to PID vary greatly but are most commonly around 10 to 20%. The majority of PID cases are subclinical or silent. PID symptoms include pelvic or lower abdominal pain (especially during menses or the first 2 weeks of the menstrual cycle) and nausea. Fever is less common. Examination findings include cervical motion tenderness and tenderness of the uterus or adnexa. Although most cases of chlamydial PID are due to the natural progression of infection, they may also occur post partum or after pregnancy termination. The long-term consequences of PID include infertility, ectopic pregnancy, and chronic pelvic pain.

Complications during Pregnancy

There is some evidence that genital chlamydial infection during pregnancy can lead to adverse outcomes, including preterm labor, low birthweight, miscarriage, and stillbirth.

Reactive Arthritis

Reactive arthritis, characterized by the classic triad of trigger infection (e.g., chlamydia), conjunctivitis, and arthritis, can complicate chlamydial infection (Chapter 265). There is a male predominance in reactive arthritis cases triggered by chlamydia, and it has been estimated that reactive arthritis occurs in about 1% of men presenting with chlamydial urethritis.

Proctitis

Proctitis caused by non-LGV *C. trachomatis* OmpA types is usually asymptomatic. Subjects with acute symptomatic proctitis may report rectal pain or bleeding, tenesmus, pruritus, rectal discharge, or diarrhea. Anoscopy or sigmoidoscopy may reveal friable mucosa and a mucoid or mucopurulent discharge. A rectal swab Gram stain often reveals many PMNs per oil field.

Conjunctivitis

An acute follicular conjunctivitis may rarely occur in adolescents or adults with genital chlamydial infection. The presumed mode of acquisition is autoinoculation with infected genital secretions. The typical clinical presentation is subacute or chronic infection characterized by unilateral conjunctival redness, mucoid or mucopurulent ocular discharge, a foreign body sensation, and preauricular adenopathy.

Oropharyngeal Infection

C. trachomatis has been detected in the oropharynx of sexually active subjects and is asymptomatic in most instances. Recent evidence suggests that *C. trachomatis* can be transmitted from oropharyngeal sites to the genital tract, providing rationale for treatment of oropharyngeal chlamydia. However, because oropharyngeal chlamydia prevalence is very low in most populations and the clinical significance of *C. trachomatis* detected in the oropharynx is unclear, routine oropharyngeal chlamydia screening is not recommended.

Infant Inclusion Conjunctivitis and Pneumonia

Because the prevalence of chlamydia in pregnant adolescents and young adults in the United States can be high (>5%), the morbidity associated with perinatally transmitted chlamydia is considerable. Neonates exposed to *C. trachomatis* during passage through the birth canal may develop inclusion conjunctivitis (≈20 to 40%) or pneumonia (≈10 to 20%). The conjunctivitis, termed inclusion conjunctivitis because the cytoplasmic chlamydia inclusion bodies demonstrated in neonatal conjunctival scrapings are the same as those seen in genital scrapings from adults with genital chlamydia, usually develops 5 to 12 days after birth but may occur as long as 4 to 6 weeks after birth. Clinical manifestations include conjunctival injection and thickening, a clear or mucopurulent ocular discharge, and eyelid swelling. Chlamydia pneumonia in infants usually occurs subacutely between 1 and 3 months of age. Characteristic clinical features include a repetitive staccato cough and absence of fever. Other clinical findings may include tachypnea, crackles on auscultation of the lungs, nasal discharge, and eosinophilia. Chest radiographs may reveal bilateral diffuse infiltrates.

Lymphogranuloma Venereum

LGV is a sexually transmitted infection caused by *C. trachomatis* LGV OmpA types. In contrast to infection with non-LGV strains, LGV is a more invasive systemic infection that involves lymphoid tissue (causing lymphadenitis) and can be ulcerative. LGV is endemic in Africa, India, Southeast Asia, South America, and the Caribbean. LGV had been uncommon in the United States, with fewer than 100 cases reported annually. However, in 2003, an LGV outbreak was reported in the Netherlands, and since then, the LGV epidemic has continued to spread throughout Europe and to North America and Australia. LGV outbreaks are primarily occurring in men who have sex with men with high-risk sexual behaviors (most are HIV infected) who are presenting with proctitis and are infected with LGV OmpA variant L2b.[7]

The clinical manifestations of LGV differ in early versus later stages of infection. In the early stage (3 to 30 days after acquisition) of genital LGV, a primary skin lesion or lesions may develop on the genital mucosa or adjacent skin in the form of a papule, ulcer, or herpetiform lesion. The lesion is usually asymptomatic and goes unnoticed, but it may be erosive; it heals quickly without scarring. Early genital LGV may also be manifested as a nonspecific inflammatory syndrome (e.g., urethritis and cervicitis), similar to infection with non-LGV strains. Genital LGV can progress to an inguinal syndrome 2 to 4 weeks later, characterized by painful, erythematous inguinal lymphadenopathy (buboes; see Fig. 312-3) and systemic manifestations including fever, headache, arthralgias, myalgias, and leukocytosis. The buboes are commonly unilateral, and about one third spontaneously rupture

and drain pus, which can be complicated by fistulas or sinus tracks; unruptured buboes usually heal. In later stages, genital tract fibrosis can lead to complications such as infertility, elephantiasis, strictures, fistulas, and subcutaneous sclerosis.

LGV may also be manifested as an anogenital syndrome with invasive proctitis. Symptoms include fever, tenesmus, anal pruritus, and a rectal discharge that may be mucoid or, less commonly, mucopurulent or bloody. Up to 20 to 30% of patients are asymptomatic. The rectal mucosa is friable, with multiple superficial ulcerations, and biopsy may reveal submucosal granulomas and crypt abscesses; these clinical and histopathologic findings resemble Crohn's disease (Chapter 141). Late complications include rectal strictures, anal fistulas, and lymphorrhoids (perianal growths of lymphatic tissue).

Natural History

The natural history of untreated genital *C. trachomatis* infection has not been fully elucidated, in part because ethical considerations limit such research. Outcomes of untreated genital chlamydia include spontaneous resolution (i.e., immune-mediated clearance) and persisting infection; the latter may be subclinical or may progress to clinical manifestations (e.g., urethritis, cervicitis), which can remain uncomplicated or lead to a complication (e.g., PID). On the basis of limited evidence in subjects with chlamydia detected initially by a screening test who returned within weeks to months for treatment and were retested, anywhere from 10 to 40% of infections resolve spontaneously before treatment.[8] Some patients with persisting infection develop clinical manifestations before returning for treatment, including a small proportion of females (up to 4%) who develop symptomatic PID. Other sparse data in females suggest that up to 50% of genital chlamydial infections resolve after a year, but a small percentage (<10%) could persist for several years. There is recent evidence to suggest that chlamydia resolution before treatment may lower the short-term risk for reinfection.[9] An improved understanding of the natural history of chlamydia could have an impact on screening and treatment recommendations.

DIAGNOSIS

C. trachomatis infections are difficult to diagnose clinically because the majority are asymptomatic, and even when symptoms or signs are present, they are nonspecific. Therefore, a definitive diagnosis relies on detecting the organism. Laboratory diagnosis confirms the clinical diagnosis in those with clinical manifestations and detects infection in asymptomatic individuals (i.e., screening). The CDC recommends annual chlamydia screening for sexually active women aged 25 years and younger as well as for older women with risk factors (e.g., new or multiple sexual partners). Select studies have demonstrated that chlamydia screening reduces the prevalence of chlamydia in women and the rate of PID. Surveillance data from the United States, England, and British Columbia during the last decade have demonstrated a steady decline in PID rates. Universal chlamydia screening in men is not recommended; selective chlamydia testing is appropriate in venues with high prevalences (e.g., sexually transmitted disease clinics, correctional facilities), for high-risk men, or for symptomatic men.

The reference standard for the detection of *C. trachomatis* has been isolation of the organism in cell culture. Development of nonculture tests was important because culture is technically demanding, expensive, and not widely available. Earlier nonculture tests included enzyme immunoassay, direct fluorescent antibody, and nucleic acid hybridization. These tests were less expensive and less technically demanding than culture but had lower sensitivities (lower limit of detection $\geq 10^3$ elementary bodies) and therefore detected fewer infections. These earlier tests have been replaced in most laboratories by NAATs, which are now the recommended diagnostic tests for chlamydia.[10] NAATs have the highest sensitivity (detecting a small number of elementary bodies) and can be performed on genital swabs and on noninvasively collected specimens (first-void urine or self-collected vaginal swabs) with similar accuracy. The CDC's recommended specimens for screening with NAATs are first-void urine in males and vaginal swabs in females. NAATs have not been cleared by the Food and Drug Administration for use with rectal, oropharyngeal, or conjunctival swab specimens; however, some laboratories have validated the assays to meet Clinical Laboratory Improvement Amendments (CLIA) requirements.

The role of serology in diagnosis of *C. trachomatis* infection is limited primarily to two indications: (1) infant pneumonia syndrome, the diagnosis of which is suggested by a *C. trachomatis* immunoglobulin M (IgM) antibody titer of 1 : 32 or greater with the microimmunofluorescence (MIF) assay; and (2) LGV, the diagnosis of which is suggested by a complement fixation (CF) antibody titer greater than 1 : 64 or MIF antibody titer greater than 1 : 256 in the appropriate clinical context. MIF has a higher specificity than CF. LGV can be more definitively diagnosed by demonstrating an LGV OmpA type on *C. trachomatis* DNA from infected material.

TREATMENT

C. trachomatis is susceptible to tetracyclines, macrolides, and select quinolones (ofloxacin and levofloxacin, but not ciprofloxacin). The CDC's recommended treatment for uncomplicated chlamydia is either doxycycline 100 mg orally twice daily for 7 days or azithromycin 1 g orally in a single dose.[11] A meta-analysis of genital chlamydia treatment trials that used primarily chlamydia culture revealed that these regimens are equally efficacious, with cure rates around 97 to 98%. A subsequent study in women using the more sensitive NAAT suggested that the cure rate in genital chlamydia may be slightly lower (≈92%).[12] Recent retrospective studies have raised concern about the efficacy of azithromycin for rectal chlamydia; however, these studies had limitations, and prospective clinical trials comparing azithromycin versus doxycycline regimens for rectal chlamydia are needed. Treatment compliance is higher with azithromycin. Alternative treatment regimens include erythromycin base 500 mg orally four times a day for 7 days, ofloxacin 300 mg orally twice daily for 7 days, and levofloxacin 500 mg orally once a day for 7 days. Azithromycin is the recommended treatment for chlamydia-infected pregnant women. *C. trachomatis* epididymitis and PID should be treated with doxycycline for 10 and 14 days, respectively. Treatment of these syndromes is usually empirical before test results are available; therefore, ceftriaxone 250 mg intramuscularly in a single dose is added to doxycycline to cover gonorrhea. LGV should be treated with doxycycline for 3 weeks; some experts recommend azithromycin 1 g orally weekly for 3 weeks as an alternative.[13] Infant chlamydial infections are treated with erythromycin base 50 mg/kg/day orally divided into four daily doses for 14 days; data on other macrolides are limited, but a shorter treatment course with azithromycin 20 mg/kg/day orally may be effective.

A test of cure (approximately 3 weeks after completion of chlamydia treatment) is recommended only for pregnant women. Sexual partners (including current partners and those with contact in the preceding 60 days) and parents of chlamydia-infected infants should be evaluated and treated empirically. Expedited partner therapy, whereby chlamydia-infected patients are offered medication or a prescription to give to their sexual partners, or clinicians provide medication to contacts without an examination, may reduce the risk for recurrent chlamydia.[14] Patients and their partners should remain abstinent until treatment is completed.

PREVENTION

Education and the provision of condoms are preventive measures that should accompany chlamydia treatment. Recurrent chlamydia occurs in approximately 10 to 20% of chlamydia-infected subjects within a few months of treatment; therefore, repeated chlamydia testing is recommended approximately 3 months after treatment. Routine screening may reduce the risk of pelvic inflammatory disease in sexually active young women.[15]

Chlamydia pneumoniae

In 1986, a new chlamydial pathogen that caused acute respiratory tract infections was identified and designated *Chlamydia* strain TWAR. It was initially thought to be a novel *C. psittaci* strain but was later recognized as the species *C. pneumoniae*. Pneumonia and upper respiratory tract infections (bronchitis, pharyngitis, laryngitis, and sinusitis) are the most frequently identified diseases caused by *C. pneumoniae*.[16] However, *C. pneumoniae* may also contribute to exacerbations of chronic bronchitis and asthma. There is evidence suggesting that *C. pneumoniae* may contribute to atherosclerosis, although large *C. pneumoniae* treatment trials have not demonstrated benefit in preventing adverse cardiovascular events. Select central nervous system disorders (e.g., Alzheimer's disease, multiple sclerosis) have been linked to *C. pneumoniae*, but a causal relationship has not been established.

EPIDEMIOLOGY

The majority of adults in the United States and other developed countries are seropositive for *C. pneumoniae*, up to 80% in some populations. Seroconversion often occurs during childhood or adolescence and may be subclinical. Studies incorporating culture or polymerase chain reaction (PCR) suggest that infection is not uncommon in children younger than 5 years. *C. pneumoniae* causes an atypical pneumonia syndrome, with an estimated annual

incidence of 1 case per 1000 population; epidemiologic studies suggest a 4-year cycle of increased pneumonia incidence. Up to 10% of community-acquired pneumonias are attributed to *C. pneumoniae*, and coinfection with other respiratory pathogens such as *Streptococcus pneumoniae* and *Mycoplasma pneumoniae* is not uncommon. The organism is believed to be acquired through the inhalation of infected respiratory droplets from persons with disease and possibly asymptomatic carriers. This mode of transmission can facilitate the spread of infection among household members and can cause epidemics in enclosed populations, such as persons in military barracks, nursing homes, and schools.

CLINICAL MANIFESTATIONS

Most *C. pneumoniae* infections occur in children, who often have mild clinical manifestations or are asymptomatic. Clinical manifestations are more evident in adults, especially the elderly, who have the highest incidence of *C. pneumoniae* pneumonia. *C. pneumoniae* causes an atypical pneumonia that is usually of mild to moderate severity. The incubation period may be several weeks, and the disease onset is gradual. A nonproductive cough is usually present and is often preceded or accompanied by nasal congestion, sore throat, and hoarseness. Headache may occur in up to half of patients. Fever and dyspnea occur less commonly. On examination, localized pulmonary crackles or rhonchi are often heard. Chest radiography shows pneumonitis, most often evident as a single subsegmental lower lobe infiltrate. The leukocyte count is usually normal. *C. pneumoniae* may also be manifested as isolated bronchitis, sinusitis, laryngitis, or nonexudative pharyngitis. The clinical course of these upper respiratory tract infections may be protracted for several weeks.

DIAGNOSIS

C. pneumoniae infection is usually diagnosed by serology or direct detection of the organism in respiratory specimens by cell culture or other nonculture methods. Serology is the test used in most clinical settings, with the MIF assay considered the reference standard for serodiagnosis. Acute infection is suggested by a four-fold rise in IgG from paired sera or a single high IgM (>1 : 16) or IgG (>1 : 512) titer. However, serology is limited by its specificity, reproducibility, and clinical correlation. *C. pneumoniae* can be isolated in cell culture, but culture is technically challenging and time-consuming. Antigen detection with use of fluorescent monoclonal antibodies has a lower sensitivity than culture and is also technically challenging. *C. pneumoniae* PCR is more sensitive than culture, and real-time PCR appears to have advantages over conventional PCR. Although there are still issues involving the standardization of PCR methods for *C. pneumoniae* detection, PCR is promising and will likely become the test of choice.

TREATMENT Rx

C. pneumoniae is susceptible to tetracyclines, macrolides, and fluoroquinolones. Treatment trials using culture have demonstrated that select macrolide and fluoroquinolone regimens eradicate *C. pneumoniae* in approximately 70 to 85% of subjects with pneumonia. The clinical response to treatment may be slow, and some patients may need re-treatment. The suggested treatment duration for most regimens is typically 10 to 14 days, except that shorter courses may be effective for azithromycin (10 mg/kg on day 1, followed by 5 mg/kg during the next 4 days; up to 1.5 g orally during 5 days). Chronic *C. pneumoniae* infections may require even longer courses of treatment (e.g., 6 weeks), and macrolides are suggested in this setting. Protective immunity after *C. pneumoniae* infection may not occur, and therefore reinfection is common.

Chlamydia psittaci

EPIDEMIOLOGY

C. psittaci naturally infects a variety of mammals and birds. *C. psittaci* strains appear to be host specific, and most human infections are linked to contact with an infected bird. *C. psittaci* infection in humans is termed psittacosis, in part because exposure to psittacine birds (parrots, parakeets, and budgerigars) is commonly implicated in infections. However, because human cases have been linked to exposure to finches, pigeons, pheasants, ducks, turkeys, chickens, seagulls, and other birds, ornithosis may be a more appropriate term. Psittacosis disease in birds ranges from an asymptomatic carriage state to a mild symptomatic illness manifested by ruffled feathers, anorexia, shivering, dyspnea, diarrhea, or depression. Infected birds shed *C. psittaci* in urine, feces, or secretions from their beaks or eyes. Their feathers and surrounding environment become contaminated. Transmission to humans is primarily by inhalation of aerosolized bird secretions or dust. Infected birds may shed the organism for months. Person-to-person transmission rarely occurs.

Human psittacosis is a rare infection, due in part to antibiotic-laced bird feed and a mandated quarantine for imported birds. The number of cases of psittacosis in the United States has been stable for the past 10 years, with fewer than 50 confirmed cases reported annually; a larger number of cases are reported but not confirmed.

CLINICAL MANIFESTATIONS

Psittacosis initially involves the lungs and then spreads to the reticuloendothelial system. The clinical spectrum of infection ranges from asymptomatic to fulminant, and clinical manifestations may resemble several other nonspecific febrile systemic illnesses, including Q fever, typhoid fever, and legionnaires' disease.[17] After an incubation period of 5 to 14 days, some patients may present with a nonspecific virus-like illness or a mononucleosis-like syndrome. The presentation most suggestive of psittacosis is an acute febrile atypical pneumonia.[18] Patients initially have an abrupt onset of shaking chills and a fever as high as 40.5° C. Temperature-pulse dissociation (i.e., elevated temperature with a normal pulse) may occur. Constitutional symptoms, including headache, myalgias, and arthralgias, are prominent. A cough, usually nonproductive, appears early in the illness and may accompany chest pain, which is usually nonpleuritic. Auscultation may be normal or reveal bilateral crackles. Chest radiograph findings are usually more dramatic than lung examination findings; the most common finding is single lower lobe consolidation, but multiple localized bronchopneumonic patches, diffuse ground-glass changes, and a miliary pattern have been described. Small pleural effusions may be seen. In contrast to *C. pneumoniae* pneumonia, psittacosis is more severe, with high fever and absent or minimal upper respiratory complaints.

Extrapulmonary findings frequently occur in psittacosis. Splenomegaly is common. A faint erythematous, blanching, maculopapular rash (Horder's spots), resembling the rose spots of typhoid fever, can occur, as can erythema nodosum. Signs of hepatitis, endocarditis (culture negative), pericarditis, myocarditis, meningoencephalitis, hemolytic anemia, or disseminated intravascular coagulation may be noted. *C. psittaci* infection has also been associated with nongastrointestinal extranodal marginal zone lymphomas of mucosa-associated lymphoid tissue, including ocular and central nervous system sites.

DIAGNOSIS

Psittacosis should be suspected in patients with a febrile illness (especially atypical pneumonia) who report exposure to sick or imported birds or who have regular exposure to birds, including bird owners, pet shop workers, veterinarians, zookeepers, and poultry processing plant workers. The diagnosis can be made with serology or by isolating the organism in cell culture. *C. psittaci* is a biocontainment level 3 agent because of its stability in the environment and aerosol transmission. Because laboratory-acquired *C. psittaci* infections are well documented, culture is discouraged and serology is preferred. If culture is attempted, laboratory staff should be notified in advance so that appropriate precautions can be taken. A serologic diagnosis of psittacosis is made by demonstrating (1) a four-fold or greater rise in CF or MIF antibody against *C. psittaci* to a titer of at least 1 : 32 from acute to convalescent sera collected at least 2 weeks apart (3 to 6 weeks is recommended) or (2) an IgM titer of 1 : 16 or greater against *C. psittaci* by MIF.

TREATMENT Rx

Untreated psittacosis can be fatal, but mortality is rare with prompt antimicrobial treatment. Because of the delay in laboratory diagnosis of psittacosis, empirical therapy should be provided on the basis of clinical suspicion. *C. psittaci* susceptibility has been demonstrated to tetracyclines, macrolides, and some of the newer fluoroquinolones. The recommended treatment regimen, based on clinical experience, is either tetracycline 500 mg four times a day or doxycycline 100 mg twice a day for 10 to 21 days. Further studies on the clinical efficacy of azithromycin and newer fluoroquinolones are needed. The initial treatment response can be dramatic, with defervescence and marked clinical improvement within 24 to 48 hours. Full recovery may take several weeks, and relapse or reinfection can occur. Endocarditis treatment includes prolonged antibiotic therapy and consideration of valve replacement.

Association with Atherosclerosis

There is evidence that *C. pneumoniae* may cause or contribute to atherosclerosis (Chapter 70) and cardiovascular disease. Seroepidemiologic studies have demonstrated a higher prevalence of *C. pneumoniae* antibodies in patients with atherosclerosis, although the serologic tests used and seropositivity criteria varied across studies. In addition, *C. pneumoniae* has been identified in atherosclerotic plaques by culture, immunohistochemical staining, and PCR. Animal studies have demonstrated that *C. pneumoniae* infection can initiate or enhance the progression of atherosclerosis, and sparse animal studies have demonstrated that antibiotics prevent atherosclerosis. However, large-scale randomized treatment trials have failed to show that *C. pneumoniae* treatment prevents cardiovascular events or mortality; some of these studies differed in terms of study populations, macrolide used, and length of treatment. Currently, antibiotics are not recommended for the prevention of adverse atherosclerotic cardiovascular events.

Association with Asthma

Epidemiologic and clinical studies have demonstrated an association between *C. pneumoniae* infection and asthma, both acute exacerbations and more severe chronic asthma. It is thought that persistent *C. pneumoniae* infection contributes to airway inflammation and hyperresponsiveness. Select randomized clinical trials have demonstrated improvement in asthma disease (symptoms, inflammatory markers, or peak expiratory flow) after antimicrobial treatment of presumed or proven *C. pneumoniae* infection, especially during acute exacerbations. However, it is difficult to determine how much of the improvement may have been due to the immunomodulatory effects of the antibiotics rather than their antimicrobial activity. Although the role of antibiotic treatment for *C. pneumoniae* infection in asthmatics remains somewhat controversial, treatment for *C. pneumoniae* infection is currently recommended for patients with acute exacerbations of asthma and clinical evidence of infection.

Association with Chronic Neurologic Diseases

C. pneumoniae infection has been implicated in the pathogenesis of multiple sclerosis and Alzheimer's disease on the basis of some studies demonstrating detection of the organism at a higher frequency in affected neurologic tissue. However, to date, the studies are contradictory, and a causal relationship between *C. pneumoniae* infection and either multiple sclerosis or Alzheimer's disease has not been established.

PREVENTION

Epidemic psittacosis is preventable by a 30-day period of quarantine for all imported psittacine birds and their treatment with feed containing chlortetracycline. The U.S. Department of Agriculture recommends extending treatment for an additional 15 days after quarantine. Prevention of epidemic and endemic psittacosis also relies on avoidance of or protection from exposure to dust or body secretions from birds or their living areas as well as avoidance of the handling of sick birds. Environmental sanitation is another important preventive measure, considering the organism's resistance to drying.

Grade A References

A1. Porco TC, Gebre T, Ayele B, et al. Effect of mass distribution of azithromycin for trachoma control on overall mortality in Ethiopian children: a randomized trial. *JAMA*. 2009;302:962-968.

A2. House JI, Ayele B, Porco TC, et al. Assessment of herd protection against trachoma due to repeated mass antibiotic distributions: a cluster-randomised trial. *Lancet*. 2009;373:1111-1118.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Nunes A, Gomes JP. Evolution, phylogeny, and molecular epidemiology of *Chlamydia*. *Infect Genet Evol*. 2014;23:49-64.
2. Wyrick PB. *Chlamydia trachomatis* persistence in vitro: an overview. *J Infect Dis*. 2010;201(suppl 2):S88-S95.
3. Taylor HR, Burton MJ, Haddad D, et al. Trachoma. *Lancet*. 2014;384:2142-2152.
4. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2013*. Atlanta: U.S. Department of Health and Human Services; 2014.
5. Torrone E, Papp J, Weinstock H. Prevalence of *Chlamydia trachomatis* genital infection among persons aged 14-39 years—United States, 2007-2012. *MMWR Morb Mortal Wkly Rep*. 2014;63:834-838.
6. Mackern-Oberti JP, Motrich RD, Breser ML, et al. *Chlamydia trachomatis* infection of the male genital tract: an update. *J Reprod Immunol*. 2013;100:37-53.
7. de Vrieze NH, van Rooijen M, van der Loeff MF, et al. Anorectal and inguinal lymphogranuloma venereum among men who have sex with men in Amsterdam, The Netherlands: trends over time, symptomatology and concurrent infections. *Sex Transm Infect*. 2013;89:548-552.
8. Geisler WM. Duration of untreated, uncomplicated *Chlamydia trachomatis* genital infection and factors associated with chlamydia resolution: a review of human studies. *J Infect Dis*. 2010;201(suppl 2):S104-S113.
9. Geisler WM, Lensing SY, Press CG, et al. Spontaneous resolution of genital *Chlamydia trachomatis* infection in women and protection from reinfection. *J Infect Dis*. 2013;207:1850-1856.
10. Gaydos CA. Nucleic acid amplification tests for gonorrhea and chlamydia: practice and applications. *Infect Dis Clin North Am*. 2005;19:367-386.
11. Centers for Disease Control and Prevention. Sexually Transmitted Diseases Treatment Guidelines, 2010. *MMWR Recomm Rep*. 2010;59(RR–12):44-47.
12. Batteiger BE, Tu W, Ofner S, et al. Repeated *Chlamydia trachomatis* genital infections in adolescent women. *J Infect Dis*. 2010;201:42-51.
13. de Vries HJ, Zingoni A, Kreuter A, et al. 2013 European guideline on the management of lymphogranuloma venereum. *J Eur Acad Dermatol Venereol*. 2015;29:1-6.
14. Hogben M, Kidd S, Burstein GR. Expedited partner therapy for sexually transmitted infections. *Curr Opin Obstet Gynecol*. 2012;24:299-304.
15. Zakher B, Cantor AG, Pappas M, et al. Screening for gonorrhea and chlamydia: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014;161:884-893.
16. Roulis E, Polkinghorne A, Timms B. *Chlamydia pneumoniae*: modern insights into an ancient pathogen. *Trends Microbiol*. 2013;21:120-128.
17. Stewardson AJ, Grayson ML. Psittacosis. *Infect Dis Clin North Am*. 2010;24:7-25.
18. Branley JM, Weston KM, England J, et al. Clinical features of endemic community-acquired psittacosis. *New Microbes New Infect*. 2014;2:7-12.

REVIEW QUESTIONS

1. Which of the following is *not* a recommended strategy for the elimination of trachoma and its complications?

A. Surgery to correct trichiasis
B. Mass antibiotic treatments within a trachoma-endemic community
C. Face washing and good hygiene
D. Achieving better sanitary conditions in the environment
E. Education to promote safer sexual practices

Answer: E The World Health Organization is committed to the elimination of trachoma and its complications by 2020 and recommends that endemic countries adopt the SAFE strategy: surgery (for trichiasis), antimicrobials (periodic community-wide treatment), facial cleanliness, and environmental sanitation improvement. Trachoma is spread by direct contact with fingers or fomites (e.g., washcloths, handkerchiefs) contaminated with eye discharge from an infected person or by eye-seeking flies, not through sexual transmission.

2. Routine annual screening for genital chlamydial infection is recommended by the Centers for Disease Control and Prevention (CDC) for the following patient group:

A. All sexually active women
B. All sexually active men
C. Sexually active women 25 years of age and younger
D. Sexually active men 25 years of age and younger
E. None of the above

Answer: C Routine genital chlamydia screening (i.e., testing of asymptomatic persons for chlamydia) in women has been shown to reduce the rate of pelvic inflammatory disease. Surveillance data from the CDC have shown that the majority of chlamydia cases reported each year are in women 25 years of age and younger. Therefore, the CDC recommends routine annual screening for genital chlamydial infection in women within this age group. Screening in women older than 25 years of age is recommended only if risk factors (e.g., new sexual partner, multiple sexual partners) are present. Routine chlamydia screening is not recommended for men; however, screening should be considered for men in venues with a high prevalence of chlamydia (e.g., STD clinics).

3. The recommended treatment for lymphogranuloma venereum (LGV) infection is

A. Doxycycline 100 mg twice daily orally for 7 days
B. Doxycycline 100 mg twice daily orally for 21 days
C. Azithromycin 1 g orally single-dose therapy
D. Erythromycin 500 mg four times a day for 21 days
E. Ciprofloxacin 500 mg twice daily for 7 days

Answer: B *Chlamydia trachomatis* strains of the LGV serovar/genotype cause more invasive disease than the non-LGV strains do, and therefore a course of doxycycline 100 mg twice daily for 21 days is recommended. There is insufficient clinical experience with macrolides to recommend them as a first-line agent for LGV. Some experts do recommend azithromycin 1 g weekly for 3 weeks as an alternative to the 3-week doxycycline regimen, but a single dose of azithromycin is believed to be insufficient to reliably cure invasive LGV infections. Ciprofloxacin has poor in vitro activity against *C. trachomatis* and is not recommended for treatment of LGV or non-LGV *C. trachomatis* infections.

4. Which of the following statements regarding *Chlamydia pneumoniae* is most accurate?

A. The seroprevalence of *C. pneumoniae* in adults in the United States is estimated to be around 15%.
B. Pneumonia caused by *C. pneumoniae* usually is manifested as an acute-onset illness with cough productive of purulent sputum.
C. *C. pneumoniae* causes upper respiratory tract infections, including pharyngitis and bronchitis.
D. Penicillin is a highly effective treatment for *C. pneumoniae* respiratory infections.
E. Large clinical trials have consistently demonstrated a reduction in adverse cardiovascular events in patients receiving antibiotic treatment directed against *C. pneumoniae.*

Answer: C The majority of adults in the United States and other developed nations are seropositive for *C. pneumoniae.* It is well known that *C. pneumoniae* can cause upper and lower respiratory tract infections. Pneumonia caused by *C. pneumoniae* usually develops gradually with a nonproductive cough, referred to as a walking pneumonia, in contrast to pneumococcal pneumonia, which is more likely to be acute in onset with a productive cough. *C. pneumoniae* respiratory infections are most effectively treated with an antibiotic from the tetracycline, macrolide, or respiratory fluoroquinolone (levofloxacin, moxifloxacin) class and are not adequately treated by a β-lactam antibiotic such as penicillin. Even though *C. pneumoniae* has been associated with atherosclerosis, large *C. pneumoniae* treatment trials have not demonstrated benefit in preventing adverse cardiovascular events.

5. Which of the following statements regarding *Chlamydia psittaci* is *not* true?

A. *C. psittaci* infection has been associated with exposure to psittacine birds (parrots, parakeets, and budgerigars).
B. *C. psittaci* infection has been associated with exposure to turkeys.
C. Birds may be carriers of *C. psittaci* but do not develop clinical disease.
D. *C. psittaci* infection can be manifested as an abrupt-onset febrile illness in humans.
E. *C. psittaci* has been classified as a biocontainment level 3 agent.

Answer: C *C. psittaci* infection has been associated with exposure to psittacine birds and several other nonpsittacine birds, including turkeys. Whereas birds may be asymptomatic carriers of *C. psittaci*, they may also develop a mild symptomatic illness manifested by ruffled feathers, anorexia, shivering, dyspnea, diarrhea, or depression. *C. psittaci* infection can be manifested as an abrupt-onset febrile illness in humans with pulmonary and extrapulmonary manifestations. *C. psittaci* has been classified as a biocontainment level 3 agent because of its stability in the environment and aerosol transmission. Because laboratory-acquired *C. psittaci* infections are well documented, culture is discouraged and serology is preferred for diagnosis.

319

SYPHILIS

EDWARD W. HOOK III

DEFINITION

Syphilis, which is a chronic infectious disease caused by the bacterium *Treponema pallidum* subspecies *pallidum*, is usually acquired by sexual contact with another infected individual. Syphilis is remarkable among infectious diseases for its large variety of clinical manifestations. If untreated, it progresses through primary, secondary, and tertiary stages. The early stages (i.e., primary and secondary), when lesions are present, are infectious. Spontaneous healing of early lesions occurs, followed by a long latent period. In about 30% of untreated patients, late disease of the heart, central nervous system (CNS), or other organs may develop years after the initial infection. Although the disease is less common now than previously, it remains a challenge to clinicians because of its protean manifestations, and it is of interest to biologists because of the prolonged, tenuous balance between the host and the invading spirochete.[1,2]

The Pathogen

The causative agent of syphilis, *T. pallidum* subspecies *pallidum*, is closely related to other pathogenic spirochetes (Chapter 320), including those causing yaws (*T. pallidum* subspecies *pertenue*) and pinta (*Treponema carateum*). *T. pallidum* is a thin, helical bacterium approximately 0.15 µm wide and 6 to 15 µm long. The organism has 6 to 14 spirals and is tapered on either end. It is too thin to be seen by ordinary Gram stain microscopy but can be visualized in wet mounts by dark-field microscopy or in fixed specimens by silver stain or fluorescent antibody methods.

Unlike most bacteria, which have protein-rich outer membranes, the *T. pallidum* outer membrane appears to be composed of predominantly phospholipids, with few surface-exposed proteins. It has been hypothesized that because of this structure, syphilis can progress despite the brisk antibody response to non–surface-exposed internal antigens, which is the basis for serologic tests for the diagnosis and management of syphilis. Between the outer membrane and the peptidoglycan cell wall are six axial fibrils; three are attached at each end, and they overlap in the center of the organism. They are structurally and biochemically similar to flagella and are in part responsible for the organism's motility.

It is possible to culture *T. pallidum*, but sustained in vitro cultivation is not yet possible, and yields are very low. Culture is of limited use in research and of no use in clinical practice. All isolates studied have been susceptible to penicillin and are antigenically similar. The only known natural hosts for *T. pallidum* are humans and certain monkeys and higher apes.

EPIDEMIOLOGY

With the exception of congenital syphilis, syphilis is acquired almost exclusively by intimate contact with the infectious lesions of primary or secondary syphilis (e.g., chancres, mucous patches, condylomata lata). Disease is usually acquired through sexual intercourse, including anogenital and orogenital intercourse. Health care workers are sometimes infected during the unsuspecting examination of patients with infectious lesions. Infection by contact with fomites is extremely uncommon. Before the advent of modern blood banking techniques, syphilis was occasionally transmitted through the transfusion of blood from persons with *T. pallidum* bacteremia, and occasional parenteral transmission still occurs as a result of the sharing of contaminated needles.

Syphilis is most common in large cities and in young, sexually active individuals. The highest rate is found in men between the ages of 20 and 29 years. In 2012, 67% of the 3142 U.S. counties reported no cases of primary or secondary syphilis, and just 28 locales accounted for about 50% of all reported infections.[3] The disease is most prevalent in the Southeast and California.

Syphilis spares no class, race, or group but is more prevalent among persons living on the margins of society. U.S. syphilis rates are about six-fold greater in African Americans than in non-Hispanic whites. In 2012, more than 75% of reported early syphilis occurred in men who acknowledged sex with other men. Increased numbers of different partners and perhaps indiscriminate choice of partners increase the risk of acquiring sexually transmitted disease (Chapter 285). A traditional cornerstone of syphilis control has been the epidemiologic investigation and treatment of sexual contacts of patients with primary or secondary lesions and patients with early latent disease. Patients with primary and secondary syphilis name, on average, nearly three different sexual contacts within the previous 90 days. As syphilis has become associated with drug use and anonymous sex, epidemiologic investigations have become less efficacious.

The incidence of syphilis has generally declined worldwide for more than 100 years, with the exception of periods of war or social upheaval. With the introduction of penicillin, there was a rapid decline in primary and secondary syphilis, to approximately 4 cases per 100,000 people in 1957. This decline was followed by reductions in federal expenditures for syphilis control, which resulted in resurgence of infectious primary and secondary syphilis in the United States; peaks of more than 12 cases per 100,000 people were attained several times from 1965 through the mid-1990s. Because many cases of syphilis are not reported, the true incidence may be much higher.

During the past 40 years, syphilis epidemics have occurred serially in at least three U.S. population subgroups. In the 1970s and 1980s, men who had sex with other men accounted for a disproportionate number of the total cases of infectious syphilis. Similar trends occurred in other countries. Then, after a period of decline, U.S. syphilis rates nearly doubled from 1986 to 1990, with 50,578 cases reported in 1990 in an epidemic disproportionately affecting multiracial heterosexual men and women and occurring contemporaneously with an epidemic of crack cocaine use. After 1990, syphilis rates again declined; in 2001, there were 6103 cases of primary and secondary syphilis reported, one of the lowest numbers since 1959. The epidemic of the late 1980s probably contributed to the spread of human immunodeficiency virus (HIV) infection (see Syphilis-HIV Interactions) and to dramatic increases in the rate of congenital syphilis. Since 2001, syphilis rates have again begun to increase in men, and now especially men infected with HIV.

In 2013, the rate of reported primary and secondary syphilis in the United States was 5.3 cases per 100,000 population, more than double the lowest-ever rate of 2.1 in 2000. During 2005 to 2013, primary and secondary syphilis rates increased among men of all ages, races, and ethnicities across all regions. Recent years have shown an accelerated increase occurring among men who have sex with men. Among women, rates increased during 2005 to 2008 and decreased during 2009 to 2013.[4]

Patients with clinically evident late syphilis, particularly those with cardiovascular or gummatous syphilis, are becoming less common, perhaps as a result of the effectiveness of penicillin therapy for early syphilis. However, surveys indicate that there are still significant numbers of patients with untreated neurologic syphilis, especially in older age groups.

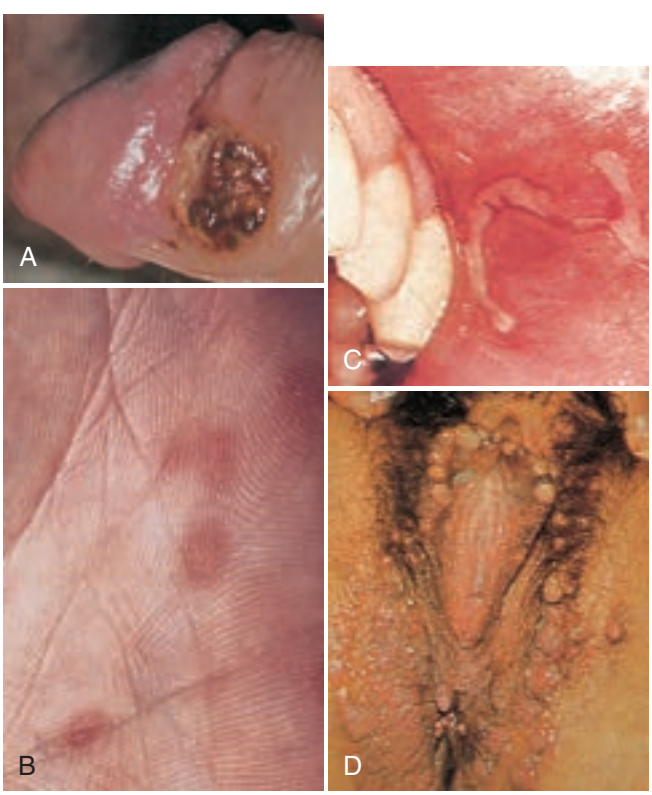

FIGURE 319-1. Syphilis lesions. A, Chancre in primary syphilis. **B,** Palmar lesions of a coppery color in secondary syphilis. **C,** Mucous patch in secondary syphilis. **D,** Condylomata lata in secondary syphilis. (**A, C, and D** from Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine*. 3rd ed. London: Mosby; 2003. **B** from Habif TP, Cambell JI, Quitadamo MJ, et al. *Skin Disease: Diagnosis and Treatment.* St. Louis: Mosby; 2001.)

Natural Course of Untreated Syphilis

The incubation period from the time of exposure to development of the primary lesion averages approximately 21 days (range, 10 to 90 days). Initially, a painless papule develops at the site of inoculation and soon breaks down to form a clean-based ulcer—the chancre—with raised, indurated margins (Fig. 319-1A). The chancre persists for 2 to 6 weeks and then heals spontaneously. Several weeks later, a secondary stage characterized by low-grade fever, headache, malaise, generalized lymphadenopathy, and a mucocutaneous rash typically develops. There may be involvement of visceral organs. The secondary eruption may occur while the primary chancre is still healing or up to several months after disappearance of the chancre. Secondary lesions also heal spontaneously within 2 to 6 weeks, and the infection then becomes latent. In more than 20% of patients with untreated latent syphilis, relapsing lesions later develop, similar to those of the secondary stage; rarely, the relapse takes the form of recurrence of the primary chancre. In the era before antibiotics, late, destructive tertiary lesions involving the eyes, the CNS, the heart, and other organs, including the skin, eventually developed in about a third of untreated patients. These lesions may occur a few years to as long as 25 years after infection.

The incidence of late complications of untreated syphilis is currently unknown, but it seems to be less than that seen previously. Cases of gumma are now so rare as to be reportable.

PATHOBIOLOGY

T. pallidum may penetrate through normal mucosal membranes and minor abrasions on epithelial surfaces. The first lesions appear at the site of direct, primary inoculation. The minimal number of treponemes needed to establish infection is not known but may be as low as one. Multiplication of organisms is slow, with a division time in rabbits of approximately 33 hours. The slow growth of treponemes in humans probably accounts in part for the protracted nature of the illness, the relatively long incubation period, and the need for relatively long duration of therapy.

Syphilis is a systemic disease from the onset. Treponemes are capable of specific attachment to host cells, but it is not known whether attachment results in damage to the host cells. Most treponemes are found in the intercellular spaces, but they are occasionally seen within phagocytic cells. However, there is no evidence of prolonged intracellular survival of treponemes. *T. pallidum* is not known to produce toxins.

The primary pathologic lesion of syphilis is a focal endarteritis with an increase in adventitial cells, endothelial proliferation, and the presence of an inflammatory cuff around affected vessels. Lymphocytes, plasma cells, and monocytes predominate in the inflammatory lesion, and polymorphonuclear cells are seen in some cases. The vessel lumen is frequently obliterated. With healing, there is considerable fibrosis. Treponemes may be seen in most early lesions of syphilis and in some of the late lesions, such as the meningoencephalitis of general paresis.

Granulomatous reaction is common in secondary and late syphilis. The granulomas are histologically nonspecific, and cases of syphilis have been incorrectly diagnosed as sarcoidosis or other granulomatous diseases. Human inoculation studies suggest that the pathogenesis of the gumma, which is a granulomatous lesion, involves hypersensitivity to small numbers of virulent treponemes introduced into a previously sensitized host.

Intracutaneous inoculation of partially purified antigens of *T. pallidum* into patients with syphilis in various stages has shown that delayed cellular hypersensitivity develops only late in secondary syphilis but is uniformly present in latent syphilis. There may be temporary hyporesponsiveness of lymphocytes to treponemal antigens in patients with primary and secondary syphilis. It is possible that the waxing and waning of lesions in early syphilis depend on the balance between the development of effective cellular immunity and the suppression of thymus-derived lymphocyte function.

The host responds to infection by producing numerous antibodies; in some instances, circulating immune complexes may be formed as well. For example, nephrotic syndrome has occasionally been recognized in secondary syphilis, and renal biopsy specimens from such patients have shown membranous glomerulonephritis characterized by focal subepithelial basement membrane deposits containing immunoglobulin G, C3, and treponemal antibody.

CLINICAL MANIFESTATIONS

Primary Syphilis

The typical lesion of primary syphilis, the chancre, is a painless, clean-based, indurated ulcer (Fig. 319-1A). The chancre starts as a papule, but then superficial erosion results in ulceration. The borders of the ulcer are raised, firm, and indurated. On occasion, secondary infections change the appearance and cause a painful lesion. Most chancres are single, but multiple ulcers are sometimes seen, particularly when skinfolds are apposed (i.e., kissing chancres). An untreated chancre heals in several weeks and leaves a faint scar. The chancre is usually associated with regional adenopathy, which may be unilateral or bilateral. The regional nodes are movable, discrete, and rubbery. If the chancre occurs in the cervix or the rectum, the affected regional iliac nodes are not palpable.

Chancres can occur at any site of potential inoculation by direct contact, with most occurring in anogenital locations. Chancres may also be seen in the pharynx, on the tongue, around the lips, on the fingers, on the nipples, and in other diverse areas. The morphology depends in part on the area of the body where they occur and on the host's immune response. Chancres in previously infected individuals may be small and remain papular. Chancres of the finger may appear more erosive and can be quite painful. Chancres of the anal canal may be missed in men who have sex with men unless a careful examination is undertaken.

Secondary Syphilis

Between 4 and 8 weeks after the appearance of the primary chancre, signs and symptoms of secondary syphilis typically develop. Symptoms may include malaise, fever, headache, sore throat, and other systemic complaints. Most patients have generalized lymphadenopathy, including involvement of the epitrochlear nodes. Approximately 30% of patients have evidence of a healing chancre, although many patients (including a disproportionate number of women and of men who have sex with men) give no history of a primary lesion.

At least 80% of patients with secondary syphilis have cutaneous or mucocutaneous lesions at some point in their illness. The diagnosis is frequently first suspected on the basis of the cutaneous eruption. The rash is often minimally symptomatic, and many patients with late syphilis do not recall primary or secondary lesions. The rashes are varied in appearance but have certain

characteristic features. The lesions are usually widespread, are symmetrical in distribution, and are frequently pink, coppery, or dusky red (particularly the earliest macular lesions). They are generally nonpruritic, although occasional exceptions have been reported, and they are rarely vesicular or bullous in adults. They are indurated, except for the very earliest macular lesions, and frequently have a superficial scale (i.e., papulosquamous lesions). The lesions tend to be polymorphic and rounded, and on healing, they may leave residual pigmentation or depigmentation. They may be faint and difficult to visualize, particularly on dark-skinned individuals.

The earliest pink macular lesions are typically seen on the trunk, with later spread to the rest of the body. The face is often spared, except around the mouth. Subsequently, a papular rash appears that is usually generalized but is marked on the palms and soles (Fig. 327-1B). These rashes are often associated with a superficial scale and may be hyperpigmented. When the rash occurs on the face, it may be pustular and resemble acne vulgaris. On occasion, the scale may be so great that it resembles psoriasis. Ulceration may occur and produce lesions resembling ecthyma. In malnourished or debilitated patients, extensive and destructive ulcerative lesions with a heaped-up crust may occur, the so-called rupial lesions. Lesions around the hair follicles may result in patchy alopecia of the beard or scalp.

Ringed or annular lesions may occur, especially around the face and particularly on dark-skinned individuals. A lesion at the angle of the mouth or the corner of the nose may have a central linear erosion, the so-called split papule.

The palate and pharynx may be inflamed. In approximately 30% of secondary syphilis patients, so-called mucous patches (Fig. 319-1C) develop; these slightly raised, oval areas are covered by a grayish white membrane that, when raised, reveals a pink base that does not bleed. These lesions may be seen on the genitalia, in the mouth, or on the tongue.

In warm, moist areas such as the perineum, large, pale, flat-topped papules may coalesce to form condylomata lata (Fig. 319-1D). Papules may also be seen in the axilla and rarely occur in a generalized form. These papules are not to be confused with the common venereal warts (i.e., condylomata acuminata), which are small, often multiple, and more sharply raised than condylomata lata. Like mucous patches, condylomata lata are highly infectious.

Other manifestations of secondary syphilis include hepatitis, which has been reported in up to 10% of patients in some series. Jaundice is rare, but an elevated alkaline phosphatase level is common. Liver biopsy reveals small areas of focal necrosis and mononuclear infiltrate or periportal vasculitis. Spirochetes can often be visualized with silver stains. Periostitis with widespread lytic lesions of bone has been reported occasionally; bone scanning appears to be a sensitive test for early syphilitic osteitis. An immune complex type of nephropathy with transient nephrotic syndrome has been documented rarely. There may be iritis or an anterior uveitis. Between 10 and 30% of patients have pleocytosis in cerebrospinal fluid (CSF), but symptomatic meningitis is seen in less than 1% of patients. Symptomatic gastritis may occur.

Relapsing Syphilis

After resolution of the primary or secondary skin lesions, 20 to 30% of patients experience cutaneous recurrences. Recurrent lesions may be fewer or more firmly indurated than the initial lesions. Like the typical lesions of primary or secondary syphilis, they are infectious for exposed sexual partners.

Latent Syphilis

By definition, latent syphilis is the stage at which there are no clinical signs of syphilis and the CSF is normal. Latency, which begins when the first attack of secondary syphilis has passed and may last for a lifetime, is usually detected by reactive serologic tests for syphilis (see Diagnosis). Congenital syphilis must also be excluded before the diagnosis of latent syphilis can be made. Patients may or may not have a clinical history of earlier primary or secondary syphilis manifestations.

Latency has been divided into two stages: early and late. Most infectious relapses occur in the first year, and epidemiologic evidence shows that the most infectious period is during the first year of infection. Early latency is therefore defined as the first year after resolution of the primary or secondary lesions or as a newly reactive serologic test response for syphilis in an otherwise asymptomatic individual who has had a negative serologic test result within the preceding year. Late latent syphilis, or, more accurately, latent syphilis of unknown duration, is ordinarily not infectious, except for pregnant women, who can transmit infection to the fetus despite long-standing infection.

TABLE 319-1 NEWLY DIAGNOSED TERTIARY SYPHILIS IN 105 PATIENTS IN DENMARK, 1961-1970

TYPE OF TERTIARY SYPHILIS	NO. OBSERVED*
Neurosyphilis	72
Asymptomatic	45
Tabes dorsalis	11
General paresis	13
Meningovascular	1
Optic atrophy	2
Cardiovascular syphilis	44
Aortic insufficiency	16
Aortic aneurysm	13
Uncomplicated aortitis†	15
Late benign syphilis (gumma)	4

*Some patients had more than one form of late syphilis.
†Autopsy diagnoses only.

Late Syphilis

Late syphilis (Table 319-1) is usually slowly progressive, although certain neurologic syndromes may have a sudden onset because of endarteritis and CNS thrombosis. Late syphilis is not infectious through sexual contact. Any organ of the body may be involved, but three main types of disease can be distinguished: late benign (gummatous), cardiovascular, and neurosyphilis.

Late Benign Syphilis

In the penicillin era, gummas are rare. They typically develop 1 to 10 years after initial infection and may involve any part of the body. Although gummas may be destructive, they respond rapidly to treatment and are therefore relatively benign. On histologic evaluation, the gumma is a granuloma.

Gummas may be solitary or multiple and most often come to medical attention as space-occupying lesions. They are usually asymmetrical and are often grouped. Gummas may start as a superficial nodule or as a deeper lesion that breaks down to form punched-out ulcers. They are ordinarily indolent, slowly progressive, and indurated on palpation. Cutaneous gummas may resemble other chronic granulomatous ulcerative lesions caused by tuberculosis, sarcoidosis, leprosy, and other deep fungal infections. Precise histologic diagnosis may not be possible. However, syphilitic gummas are the only such lesions to heal dramatically with penicillin therapy.

Gummas may also involve deep visceral organs, particularly the respiratory tract, gastrointestinal tract, and bones. In addition, they may involve the larynx or the pulmonary parenchyma. Gummas of the stomach may masquerade as carcinoma of the stomach or lymphoma. Gummas of the liver were once the most common form of visceral syphilis and often manifested as hepatosplenomegaly and anemia and occasionally as fever and jaundice. Skeletal gummas typically produce lesions in the long bones, skull, and clavicle; a characteristic symptom is nocturnal pain. Radiologic abnormalities, when present, include periostitis and lytic or sclerotic, destructive osteitis.

Cardiovascular Syphilis

The primary cardiovascular complications of syphilis are aortic insufficiency (Chapter 75) and aortic aneurysm (Chapter 78), usually of the ascending aorta. Less commonly, other large arteries may be affected, and involvement of the coronary ostia rarely results in coronary insufficiency. All these complications are caused by obliterative endarteritis of the vasa vasorum, with resultant damage to the intima and media of the great vessels. This damage results in dilation of the ascending aorta, but the valve cusps remain normal. An aneurysm occasionally is manifested as a pulsating mass bulging through the anterior chest wall. Syphilitic aortitis[5] may also involve the descending aorta proximal to the renal arteries.

Cardiovascular syphilis usually begins within 5 to 10 years of the initial infection but may not be manifested clinically until 20 to 30 years later. Cardiovascular syphilis does not occur after congenital infection, a phenomenon that remains unexplained.

Asymptomatic aortitis is best diagnosed by visualizing linear calcifications in the wall of the ascending aorta by radiography. The signs of syphilitic aortic insufficiency are the same as for aortic insufficiency of other causes. In aortic insufficiency resulting from dilation of the aortic ring, the decrescendo murmur is often loudest along the right sternal margin. Syphilitic aneurysms may be fusiform but are more typically saccular and do not lead to aortic dissection. Between 10 and 20% of patients with cardiovascular syphilis have coexistent neurosyphilis.

Neurosyphilis

CNS involvement occurs throughout the natural history of syphilis. Neurosyphilis[6] can be divided into five groups: asymptomatic, syphilitic meningitis, meningovascular syphilis, tabes dorsalis, and general paresis. Asymptomatic neurosyphilis can occur at any time, whereas syphilitic meningitis is most common during the secondary stage of infection. Meningovascular syphilis, tabes dorsalis, and general paresis are typically manifestations of late syphilis. The divisions are not absolute, and overlap between syndromes is typical. Current cases of neurosyphilis are likely to be variants of the classic syndromes, possibly as a result of the use of antimicrobial agents for other diseases.

Syphilitic Meningitis

Acute to subacute aseptic meningitis can occur at any time after the primary stage, but it usually occurs within the first year of infection. It frequently involves the base of the brain and may result in unilateral or bilateral cranial nerve palsies. Mild aseptic meningitis may be relatively common in patients with early syphilis, but severe disease occurs in only about 1.5% of untreated patients. Syphilitic meningitis typically resolves without treatment.

Meningovascular Syphilis

Some patients have sufficient endarteritis and perivascular inflammation to result in cerebrovascular thrombosis and infarction, generally 5 to 10 years after the initial infection. However, case reports suggest that in syphilis patients with coexistent HIV infection, meningovascular syphilis may be manifested earlier or may be a manifestation of treatment failure. Patients frequently have associated aseptic meningitis. Most cerebrovascular accidents are not caused by syphilitic arteritis, even in patients with a reactive serologic test result for syphilis. However, syphilis should be considered a potential cause in relatively young patients with a history of syphilis and without other risk factors for cerebrovascular accidents.

Tabes Dorsalis

Tabes dorsalis, which appears to be far less common than in the prepenicillin era, is a slowly progressive, degenerative disease that involves the posterior columns and posterior roots of the spinal cord and results in progressive loss of peripheral reflexes, impairment of vibration and position sense, and progressive ataxia. Sensory changes may lead to chronic destructive changes in the large joints of the affected limbs in advanced cases (i.e., Charcot's joints). Incontinence of the bladder and impotence are common. Sudden and severely painful crises of uncertain origin are a characteristic part of the syndrome. Severe, sharp abdominal pain may lead to exploratory surgery.

Optic atrophy is seen in 20% of cases. In 90% of patients, the pupils are bilaterally small and fail to constrict further in response to light, but they do respond normally to accommodation (i.e., Argyll Robertson pupils).

The onset of tabes dorsalis is usually first noticed 20 to 30 years after the initial infection. Its cause is unclear, and spirochetes cannot be demonstrated in the posterior column or dorsal root.

General Paresis

This form of neurosyphilis is a chronic meningoencephalitis resulting in the gradual and progressive loss of cortical function. It typically occurs 10 to 20 years after the initial infection. On pathologic examination, there is a perivascular and meningeal chronic inflammatory reaction, with thickening of the meninges, granular ependymitis, degeneration of the cortical parenchyma, and abundant spirochetes in tissues.

In its early stages, general paresis results in nonspecific symptoms of early dementia, such as irritability, fatigability, headaches, forgetfulness, and personality changes. Later, there is impaired memory, defective judgment, lack of insight, confusion, and often depression or marked elation. Patients may be delusional, and seizures sometimes occur. There may also be loss of other cortical functions, including paralysis or aphasia.

Physical signs are primarily those of the altered mental status. Cranial nerve palsies are uncommon, and optic atrophy is rare. The complete Argyll Robertson pupil is also uncommon, but irregular or otherwise abnormal pupils are not infrequent. Peripheral reflexes are often somewhat increased.

Syphilis-HIV Interactions

Syphilis, like other genital ulcer diseases, is associated with a three- to fivefold increased risk for acquisition of HIV infection. Presumably, genital ulcers act as portals of entry through which HIV may more readily infect exposed individuals. As a result, HIV serologic testing 3 months after a diagnosis of syphilis is recommended for all patients. Conversely, in individuals with HIV infection who acquire syphilis, the natural history of the infection may be modified. HIV-infected syphilis patients are somewhat more likely than non–HIV-infected patients to present initially with secondary syphilis. HIV-infected secondary syphilis patients are also more likely than HIV-negative secondary syphilis patients to have coexistent chancres, suggesting that the healing of chancres is delayed or the appearance of secondary manifestations is accelerated in the presence of HIV coinfection.

Congenital Syphilis

Congenital syphilis results from the transplacental, hematogenous spread of syphilis from the mother to the fetus. In 2012, 322 cases of congenital syphilis were reported in the United States. A serologic test for syphilis should be performed in all expectant mothers at the beginning of pregnancy and should be repeated near the end of pregnancy in women living in areas where syphilis is relatively common.[7]

The risk for fetal infection is greatest in the early stages of untreated maternal syphilis and declines slowly thereafter, but the untreated mother can infect her fetus during at least the first 5 years of her infection. Adequate treatment of the mother before the 16th week of pregnancy usually prevents clinical illness in the neonate. Later treatment may not prevent late sequelae of the disease in the child. Untreated maternal infection may result in stillbirth, neonatal death, prematurity, or syndromes of early or late congenital syphilis in surviving infants.

Manifestations of early congenital syphilis are often seen in the perinatal period but may not develop until the infant has been discharged from the hospital. The disease resembles secondary syphilis in adults, except that the rash may be vesicular or bullous. The child often has rhinitis, hepatosplenomegaly, hemolytic anemia, jaundice, and pseudoparalysis (i.e., immobility of one or more extremities) as a result of painful osteochondritis.

Late congenital syphilis is defined as congenital syphilis diagnosed more than 2 years after birth. The disease may remain latent, with no manifestations of late damage. Cardiovascular alterations have not been observed in patients with congenital syphilis. Neurologic manifestations are common and may include eighth cranial nerve deafness and interstitial keratitis. Periostitis may result in prominent frontal bones of the skull, depression of the bridge of the nose (saddle nose), poor development of the maxilla, and anterior bowing of the tibia (saber shins). There may be late-onset arthritis of the knees (Clutton's joints). The permanent dentition may show characteristic abnormalities known as Hutchinson's teeth; the upper central incisors are widely spaced, centrally notched, and tapered in the manner of a screwdriver. The molars may show multiple poorly developed cusps (mulberry molars).

DIAGNOSIS

Dark-Field Examination

The most definitive means of syphilis diagnosis is finding typical spirochetes in lesions of early acquired or congenital syphilis. Dark-field examination is often positive in cases of primary syphilis and in patients with the moist mucosal lesions of secondary and congenital syphilis. The result may occasionally be positive for aspirates of lymph nodes in secondary syphilis. False-negative results may occur in primary syphilis because of the application of soaps, antiseptics, or other compounds toxic to *T. pallidum* to the lesions. A single negative result is therefore insufficient to exclude syphilis. For high-risk individuals (e.g., drug users, homosexually active men), it is appropriate to treat presumptively on the basis of suspicious lesions after performing serologic tests. Confusion may also arise because of the presence of spirochetes that are morphologically indistinguishable from *T. pallidum* organisms in the mouth, particularly around the gingival margins. Living *T. pallidum* organisms demonstrate gradual motion to and fro, rotational movement around the long axis, and rather sudden 90-degree flexing near the center of the organism. Because most physicians do not have the proper equipment and are not familiar with dark-field microscopy techniques, public health authorities can

be called for assistance. *T. pallidum* may also be demonstrated in biopsy or pathologic specimens by fluorescent antibody stains or by silver stains.

Serologic Tests

Two basic types of serologic tests (Table 319-2) are widely used to diagnose infection with *T. pallidum:* (1) nontreponemal tests that detect antibodies reactive with diphosphatidylglycerol (cardiolipin), which is a normal component of many tissues; and (2) tests that detect antibodies to specific treponemal antigens.

Nontreponemal Tests

The standard tests to detect anticardiolipin antibody are the rapid plasma reagin (RPR) and Venereal Disease Research Laboratory (VDRL) tests, which are slide flocculation tests. The RPR and VDRL are readily quantified, so they are the tests of choice for monitoring patients' responses to treatment. The relative proportion of patients with a false-positive RPR result depends on the prevalence of syphilis in the community; the lower the prevalence of syphilis, the higher the proportion of reactive RPR test results from nonsyphilitic causes.

The RPR test result begins to turn positive less than 1 week after onset of the chancre; thus, a nonreactive RPR test result does not exclude primary syphilis, particularly if the lesion is less than 1 week old. The RPR test result is positive in 99% of patients with secondary syphilis (Table 319-3). Patients with advanced HIV infection may have negative test results, and some patients have such high titers of antibody that they are in antibody excess; dilution of their serum paradoxically results in conversion of a negative test result to a positive one, the so-called prozone reaction. RPR reactivity tends to diminish in later stages of syphilis, and only about 70% of patients with cardiovascular syphilis or late neurosyphilis have positive RPR test results.

The *quantitative titer* of the RPR or VDRL test is somewhat useful in diagnosis and is quite useful for monitoring of the therapeutic response. Most patients with secondary syphilis have titers of at least 1 : 16. Most patients with false-positive RPR test results have titers of less than 1 : 8. No single titer is diagnostic by itself. Significant rises (four-fold or greater) in paired sera, however, strongly indicate acute syphilis.

Treponemal Tests

Several types of treponemal tests are widely used. Recently, treponemal enzyme immunoassays (EIAs) using cloned treponemal antigens for treponemal antibody detection have become available from several manufacturers and have gained favor because of their low cost and ease of use. In addition to EIA tests, agglutination of particles to which *T. pallidum* antigens have been fixed is the basis of the widely used *T. pallidum* particle agglutination (TP-PA) test. The fluorescent treponemal antibody absorption (FTA-ABS) test has been widely used as well and is reported in terms of relative brilliance of fluorescence, from borderline to 4 plus; most laboratories report only test results with 2 plus or greater reactivity as positive. For patients lacking historical or clinical evidence of syphilis but with a reactive FTA-ABS test result, the test should be repeated. Use of another treponemal test may be helpful in problem cases. The TP-PA test is slightly less sensitive than the RPR or FTA-ABS test in primary syphilis. Its sensitivity and specificity are otherwise nearly identical to those of the FTA-ABS test.

Because EIA serologic tests permit the screening of large numbers of sera and have performance characteristics (sensitivity, specificity, predictive values) similar to those of other treponemal tests, they have been increasingly used recently for syphilis screening.[8,9] Persons with reactive treponemal antigen EIAs should be tested with a quantitative nontreponemal test, such as the RPR or VDRL test, for confirmation and to permit the use of that test to evaluate the subsequent response to therapy. It is not unusual for patients to have a reactive EIA test result for syphilis and a nonreactive RPR or VDRL test result. A substantial proportion of these EIA-only positive test results are falsely positive or are detecting long-standing, often previously treated syphilis, but occasionally they may detect very recent infection before RPR or VDRL test results become positive.

When nontreponemal tests such as the RPR and VDRL are used for screening, treponemal tests are used to confirm that persons with reactive nontreponemal test results have antibodies to *T. pallidum.* Results of treponemal tests are not reliably quantified. They are sensitive and have a high degree of specificity, in that only approximately 1% of normal individuals have reactive treponemal test results. They are reactive in 85% of patients with primary syphilis, 99% with secondary syphilis, and at least 95% with late syphilis. They may therefore be the only test with a positive result in patients with cardiovascular or neurologic syphilis. For patients with late syphilis, treponemal test results often remain reactive for life, despite adequate therapy.

TABLE 319-2 SEROLOGIC TESTS FOR SYPHILIS

TYPE	USE
NONTREPONEMAL (ANTICARDIOLIPIN) ANTIBODIES	
VDRL (slide flocculation)	Screening, quantitation of response to treatment
RPR (circle card) (agglutination)	Screening, quantitation of response to treatment
SPECIFIC TREPONEMAL ANTIBODIES	
FTA-ABS (immunofluorescence with absorbed serum)	Confirmatory, diagnostic; not for routine screening
TP-PA (microhemagglutination)	Similar to FTA-ABS but can be quantified and automated
EIA	Confirmatory and increasingly used for screening; automated

EIA = enzyme immunoassay; FTA-ABS = fluorescent treponemal antibody absorption test; RPR = rapid plasma regain test; TP-PA = *Treponema pallidum* particle agglutination; VDRL = Venereal Disease Research Laboratory.

TABLE 319-3 FREQUENCY OF POSITIVE SEROLOGIC TEST RESULTS IN UNTREATED SYPHILIS

STAGE	VDRL (%)	FTA-ABS (%)	TP-PA (%)
Primary	70	85	50-60
Secondary	99	100	100
Latent or late	70	98	98

FTA-ABS = fluorescent treponemal antibody absorption test; TP-PA = *Treponema pallidum* particle agglutination; VDRL = Venereal Disease Research Laboratory.

Differential Diagnosis

The differential diagnosis of a genital ulcer (Chapter 285) includes genital herpes (Chapter 374), chancroid (Chapter 301), lymphogranuloma venereum (Chapter 318), and a number of other ulcerative processes. Classically, herpetic ulcers are multiple, painful, superficial, and, if seen early, vesicular. However, atypical manifestations may be indistinguishable from a syphilitic chancre. Genital herpes is much more common than syphilis and is now the most common cause of a "typical chancre" in North America. Syphilitic chancres may also be coinfected with herpes simplex virus in about 15% of cases. The ulcers of chancroid are usually painful, often multiple, and frequently exudative and nonindurated. Lymphogranuloma venereum may produce a small, papular lesion associated with regional adenopathy. Other conditions that must be distinguished include granuloma inguinale (Chapter 316), drug eruptions, carcinoma, superficial fungal infections (Chapter 438), traumatic lesions, and lichen planus (Chapter 438). In most cases, the final distinction is based on dark-field examination, which is positive only in syphilis, and on serologic test results.

The differential diagnosis of the skin lesions of secondary syphilis includes pityriasis rosea (Chapter 438), which can be differentiated by the occurrence of lesions along lines of skin cleavage and frequently by the presence of a herald patch. Drug eruptions, acute febrile exanthems, psoriasis, lichen planus, scabies, and other diseases must also be considered in some cases. A mucous patch may superficially resemble oral candidiasis (i.e., thrush). Infectious mononucleosis (Chapter 377) may appear similar to secondary syphilis, with sore throat, generalized adenopathy, hepatitis, and a generalized rash. Hepatitis (Chapter 148) may also cause confusion.

False-Positive Serologic Test Results for Syphilis

The RPR or VDRL test result is reactive in patients with other treponemal diseases, such as pinta, yaws, and endemic syphilis (Chapter 320). These test results may also be falsely reactive in persons who do not have treponemal infections based on a negative clinical history or negative results of serum treponemal tests.[10]

The origins of false-positive results are better studied for nontreponemal tests than for treponemal tests. Acute (<6 months) false-positive RPR test results occur with low frequency in patients with atypical pneumonia, malaria,

and other bacterial or viral infections, and they may occur after smallpox or other vaccinations as well. Chronic false-positive RPR test results (persisting >6 months) are relatively common in patients with autoimmune disorders such as systemic lupus erythematosus (Chapter 266), parenteral drug users, HIV-infected patients, patients with leprosy, and the aged. Between 8 and 20% of patients with systemic lupus erythematosus have false-positive RPR test results. Chronic false-positive RPR test results in female patients 20 years or younger indicate a significant risk for the future development of systemic lupus erythematosus, thyroiditis, or other autoimmune disorders. As many as a third of parenteral drug users have false-positive RPR test results. More than 1% of persons 70 years old and 10% of those older than 80 years also have low-titer, false-positive RPR test results. In most cases of false-positive RPR test results, the titer is less than 1 : 8, although a few patients with lymphoma and other diseases have very high-titer, false-positive results.

There is also an increased incidence of false-positive treponemal test results in other chronic inflammatory diseases associated with hyperglobulinemia, including rheumatoid arthritis and biliary cirrhosis. On occasion, reproducible positive FTA-ABS results are obtained in patients with no clinical or historical evidence of syphilis and no evidence of diseases typically associated with false-positive FTA-ABS results. If the diagnosis is in doubt and if the patient is not allergic to penicillin, it is often prudent to treat for possible syphilis.

Neurosyphilis

Asymptomatic neurosyphilis is diagnosed when there are CSF abnormalities, such as lymphocytic pleocytosis, protein elevation, or a reactive VDRL test result, in a syphilis patient in the absence of signs and symptoms of neurologic disease. Unlike serologic tests, the VDRL and RPR tests do not perform equally for CSF, and only the VDRL is recommended. Although numerous other processes can cause CSF pleocytosis or protein elevations, false-positive CSF VDRL test results are rare in the absence of a traumatic tap. If the CSF is normal 2 years or longer after the initial infection, a positive CSF finding is not likely to develop later. Routine lumbar punctures to examine CSF are not indicated in early syphilis unless the patient is known to have HIV infection. Lumbar puncture in HIV-infected persons with early syphilis is the subject of controversy. Although a nonreactive CSF FTA-ABS result may be useful to rule out the diagnosis, no diagnosis of neurosyphilis should be based solely on the CSF FTA-ABS test.

In syphilitic meningitis, the CSF shows a lymphocytic pleocytosis, with increased protein and usually normal glucose concentrations. The CSF VDRL test is nearly always reactive. Rarely, the CSF glucose concentration is decreased. Without treatment, syphilitic meningitis generally resolves, similar to the course of other manifestations of early syphilis. This syndrome can mimic tuberculous or fungal meningitis or nonpurulent meningitis of various causes.

In tabes dorsalis, the VDRL test for serum is nonreactive in as many as 30 to 40% of patients, and 10 to 20% of patients (even before the advent of penicillin) have normal CSF VDRL results. The FTA-ABS test for serum is nearly always reactive. In general paresis, the CSF is nearly always abnormal, with lymphocytic pleocytosis and an increased total protein concentration. The VDRL test is usually reactive for CSF and serum.

Congenital Syphilis

Because many infants with congenital syphilis may be clinically normal at birth but develop serious, symptomatic disease some weeks later, it is important to determine whether a newborn with a reactive serologic test result for syphilis has passively transferred maternal antibody or is actively infected. If the mother has been adequately treated for syphilis during pregnancy and the infant is clinically normal at birth, one option is to monitor the infant carefully by serial examinations and RPR titers. If the reactive RPR result for the infant is caused by passively transferred maternal antibody, the titer will fall markedly in the first 2 months of life; a rising titer indicates active disease and the need for treatment. However, the risk of improper follow-up of RPR-positive but clinically normal neonates makes the immediate empirical administration of effective therapy an attractive alternative.

TREATMENT

Rx

T. pallidum is inhibited by less than 0.01 μg/mL of penicillin G. Because treponemes divide slowly and penicillin acts only on dividing cells, it is necessary to maintain serum levels of penicillin for many days (Table 319-4).

Early Infectious Syphilis

Early syphilis (<1 year) can be treated with a single injection of 2.4 million units of benzathine penicillin G, which provides low but effective serum levels for about 2 weeks and cures approximately 95% of patients.[11] It is not necessary to examine CSF at this stage because penicillin prevents the later development of neurosyphilis.[12]

Individuals with other sexually transmitted diseases may have been exposed to syphilis at the time they became infected. Treatment with a single dose of β-lactam antibiotics (penicillins, cephalosporins), which provide relatively high serum levels for a brief period, is ineffective in established early syphilis but is curative if the disease is still in the incubating stage. The ceftriaxone regimen used for gonorrhea (Chapter 299) is probably curative for incubating syphilis, but careful follow-up is indicated if there is reason to suspect exposure to syphilis in a patient treated for gonorrhea with ceftriaxone. Single-dose therapy with 2.0 g of azithromycin administered orally was as effective as benzathine penicillin therapy in several studies of early syphilis,[13] but treatment failures have been reported in persons with coexistent HIV infection. Currently, azithromycin should not be used for the treatment of early syphilis unless close follow-up can be ensured.

For patients allergic to penicillin, 100 mg of doxycycline orally twice daily for 14 days is recommended. Particularly careful follow-up is necessary for patients treated with drugs other than penicillin because they may not be fully compliant with these prolonged courses of oral therapy and these regimens have been less fully evaluated clinically. Ceftriaxone, given in doses of 500 mg to 1.0 g intramuscularly daily for 10 days, may be effective but has been studied only in small numbers of patients with syphilis. Quinolone antibiotics have essentially no effect on syphilis.

Syphilis of More than 1 Year in Duration

Prolonged therapy with intramuscular injections of 2.4 million units of benzathine penicillin G weekly for 3 weeks is recommended for the treatment of late latent syphilis and latent syphilis of unknown duration. Limited evidence suggests that treatment of latent syphilis with a total dose of 7.2 million units of benzathine penicillin during a 3-week period is curative, even if the patient has asymptomatic neurosyphilis. However, because of the possible lack of efficacy of benzathine penicillin G in some patients with CNS syphilis, CSF examination should be considered in those with latent syphilis to exclude asymptomatic neurosyphilis, particularly in HIV-positive patients, in whom the prevalence of asymptomatic neurosyphilis is higher. Alternatively, a lumbar puncture can be performed at the conclusion of the follow-up period (2 years); if the CSF is normal, the patient can be reassured that neurosyphilis will not develop.

Larger doses of penicillin are recommended for persons with proven neurosyphilis (see Table 319-4). General paresis responds well to penicillin therapy if it is administered early, although progressive neurologic decline may develop later in as many as a third of treated patients. Carbamazepine in doses of 400 to 800 mg/day reportedly treats the lightning pains of tabes dorsalis effectively. Published studies show that a total of 6.0 to 9.0 million units of penicillin G results in a satisfactory clinical response in approximately 90% of patients with neurosyphilis who do not have HIV infection. There are anecdotal reports of increased treatment failures in patients with concomitant HIV infection, and there is considerable rationale to treat these patients with intravenous penicillin G (20 million units/day for at least 10 days). Therapy for neurosyphilis can result in increased CSF pleocytosis for 7 to 10 days after treatment is started and may transiently convert a normal CSF to abnormal.

Although there is no evidence that therapy with antimicrobial drugs is clinically beneficial in patients with cardiovascular syphilis, treatment is recommended to prevent further progression of disease and because approximately 15% of patients with cardiovascular syphilis have associated neurosyphilis. If patients are allergic to penicillin, it is mandatory that the CSF be examined before therapy is undertaken; if the CSF is abnormal, desensitization to penicillin is generally recommended. With a normal CSF, tetracycline (500 mg orally four times a day) or doxycycline (100 mg orally two times a day) taken for 4 weeks is probably effective.

Syphilis in Pregnancy

Because of the risk to the fetus, evaluation and treatment of a pregnant RPR-positive patient must be rapid, particularly for those patients first seen in the later stages of pregnancy. If a confirmatory treponemal test result is positive and the patient has not been treated, penicillin should be administered in doses appropriate for early or late syphilis, as outlined earlier. For penicillin-allergic patients, penicillin desensitization is preferred; patients should not be treated with tetracycline or erythromycin because of toxicity (tetracycline) or lack of efficacy (erythromycin). For patients who are RPR positive but treponemal test negative and have no clinical signs of syphilis, treatment may be withheld; a quantitative RPR test and another treponemal test should be repeated in 4 weeks. If the treponemal titer has risen four-fold or more, or if clinical signs of syphilis have developed, the patient should be treated. If, after repeated examination, the diagnosis remains equivocal, the patient should be

TABLE 319-4 PENICILLIN TREATMENT FOR SYPHILIS AS RECOMMENDED BY THE U.S. PUBLIC HEALTH SERVICE

INDICATIONS FOR SYPHILIS THERAPY*	DOSAGE AND ADMINISTRATION†	
	Benzathine Penicillin G	*Aqueous Benzylpenicillin G or Procaine Penicillin G*
Primary, secondary, and early latent syphilis (<1 year); epidemiologic treatment	Total of 2.4 million units; single IM dose of two injections of 1.2 million units in one session	Total of 4.8 million units IM in doses of 600,000 units/day for 8 consecutive days
Late latent (>1 year) or when CSF was not examined in "latency"; cardiovascular syphilis, late benign (cutaneous, osseous, visceral gumma)	Total of 7.2 million units IM in doses of 2.4 million units at 7-day intervals during a 21-day period	Total of 9 million units IM in doses of 600,000 units/day during a 15-day period
Symptomatic or asymptomatic neurosyphilis	2-4 million units aqueous (crystalline) penicillin G IV q4h for at least 10 days	2-4 million units procaine penicillin IM daily and probenecid 500 mg orally four times daily, for 10-14 days
Congenital		
Infants	CSF normal: total of 50,000 units/kg IM in a single or divided dose at one session	CSF abnormal: total of 50,000 units/kg/day IM for 10 consecutive days‡
Older children	CSF normal: same as for early congenital syphilis, up to 2.4 million units	CSF abnormal: 200,000-300,000 units/kg/day aqueous crystalline penicillin IV for 10-14 days

*In *pregnancy,* treatment depends on the stage of syphilis.
†Individual doses can be divided for injection in each buttock to minimize discomfort.
‡For aqueous penicillin, give in two divided intravenous doses per day; for procaine penicillin, give as one daily dose intramuscularly.
CSF = cerebrospinal fluid.

treated to prevent possible disease in the neonate. After treatment, a quantitative RPR titer should be monitored monthly; if it rises four-fold, the patient should be treated a second time.

Congenital Syphilis

Proper treatment of the mother usually prevents active congenital syphilis in the neonate. However, infected infants may be clinically normal at birth, and the infant may be seronegative if the mother's infection was acquired late in pregnancy. The infant should be treated at birth if the mother has received no treatment or inadequate treatment or has been treated with drugs other than penicillin, if the mother has not yet responded to possibly effective therapy, or if the infant cannot be carefully monitored for several months after birth. The infant's CSF should be examined before treatment. If the CSF is normal, the child can be treated with a single intramuscular injection of 50,000 units/kg (up to 2.4 million units) of benzathine penicillin G. If the CSF is abnormal, the infant should be treated with 50,000 units/kg of aqueous penicillin G given intramuscularly or intravenously twice daily for a minimum of 10 days. Alternatively, a single daily intramuscular injection of 50,000 units/kg of procaine penicillin may be given for 10 days. Antimicrobial agents other than penicillin are not recommended for treatment of congenital syphilis.

Jarisch-Herxheimer Reactions

Up to 60% of patients with early syphilis and a significant proportion of patients with later stages of syphilis experience a transient febrile reaction after therapy for syphilis. The pathogenesis is unclear, but it may be caused by the liberation of antigens from spirochetes.

This reaction usually occurs in the first few hours after therapy, peaks at 6 to 8 hours, and disappears within 12 to 24 hours of therapy. On occasion, Jarisch-Herxheimer reactions are mistaken for allergic reactions to syphilis therapy. Temperature elevation is usually low grade, and there is often associated myalgia, headache, and malaise. The skin lesions of secondary syphilis are frequently exacerbated during the Jarisch-Herxheimer reaction, and cutaneous lesions that were not visible may become visible. The reaction is generally of no clinical significance and in most cases can be treated with salicylates. Corticosteroids have been used to prevent adverse effects of the Jarisch-Herxheimer reaction, but there is no evidence that they are clinically beneficial (other than reducing fever) or necessary. Institution of treatment with small doses of penicillin does not prevent the reaction.

PREVENTION

All patients with syphilis should be reported to public health authorities. In the absence of an effective vaccine, control of syphilis depends on finding and treating persons with infectious lesions of primary and secondary syphilis before they can transmit the disease as well as finding and treating individuals with incubating syphilis before infectious lesions develop. All patients with early syphilis (primary, secondary, or early latent) should be carefully interviewed by qualified persons to determine the nature of their recent sexual contacts. Approximately 16% of the named recent contacts of patients with early syphilis are found to have active, untreated syphilis on examination, and a similar proportion of individuals named as suspects or associates also have active syphilis.

Treatment of the sexual contacts of patients with early syphilis with 2.4 million units of benzathine penicillin G intramuscularly is recommended even if the contacts are clinically and serologically normal on examination. This is because syphilis eventually develops in 30% of clinically normal contacts who are untreated. In general, preventive treatment is given to all sexual contacts in the past 90 days, although nearly all cases of syphilis in contacts develop within 60 days of exposure.

PROGNOSIS

Follow-up Examinations

All HIV-seronegative patients with early or congenital syphilis should return for quantitative VDRL titers and clinical examination 6 and 12 months after treatment. For HIV-positive patients, serologic tests should be repeated at 1, 2, 3, 6, 9, and 12 months. Patients with late latent syphilis should also be examined 24 months after therapy.

In about 80 to 85% of patients with early (i.e., primary, secondary, or early latent) syphilis, quantitative RPR titers decline two or more dilutions (four-fold) by 6 and 12 months after therapy. In serofast patients, prolonged reactive RPR test results are associated with older age, lower initial RPR titers, prolonged infection, or more advanced stage (primary < secondary < early latent) infection.[13] Re-treatment of patients with serofast RPR results at 6 months leads to serologic response to syphilis in a minority of patients.[14] Chronic, low-titer RPR reactivity after therapy is much more common in cases of late syphilis and should not be viewed with alarm. Treponemal test results may remain positive for years despite adequate therapy. A four-fold or greater rise in RPR titer after therapy is sufficient evidence for repeated treatment. Patients with treated early syphilis are susceptible to reinfection, and many clinical and serologic relapses after therapy are probably reinfections. As such, they represent failures of proper epidemiologic case finding and preventive therapy for the patient's sexual contacts.

Patients with neurosyphilis should be monitored with serologic tests for at least 3 years and with repeated CSF examinations at 6-month intervals. CSF pleocytosis is the first abnormality to disappear, but cell counts may not be normal for 1 to 2 years. Elevated CSF protein levels fall even more slowly, followed by a change in the positive CSF VDRL test result, which may take years to become negative. It is not known whether high-dose intravenous penicillin therapy accelerates the return of CSF to normal. Rising CSF cell counts, protein level, and CSF VDRL titer obtained at follow-up are an indication for repeated treatment.

Antibiotic therapy should ultimately cure essentially all patients with early or secondary syphilis, although treatment failures may occur in patients with concomitant HIV infection. In tabes dorsalis, penicillin usually arrests progression but does not reverse the symptoms. Meningovascular syphilis generally responds well, except for residual damage resulting from ischemic infarcts.

Grade A Reference

A1. Bai ZG, Wang B, Yang K, et al. Azithromycin versus penicillin G benzathine for early syphilis. *Cochrane Database Syst Rev*. 2012;6:CD007270.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Shockman S, Buescher LS, Stone SP. Syphilis in the United States. *Clin Dermatol*. 2014;32:213-218.
2. Cohen SE, Klausner JD, Engelman J, et al. Syphilis in the modern era: an update for physicians. *Infect Dis Clin North Am*. 2013;27:705-722.
3. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2012*. Atlanta: U.S. Department of Health and Human Services; 2013.
4. Patton ME, Su JR, Nelson R, et al. Primary and secondary syphilis—United States, 2005-2013. *MMWR Morb Mortal Wkly Rep*. 2014;63:402-406.
5. Dietrich A, Gauglitz GG, Pfluger TT, et al. Syphilitic aortitis in secondary syphilis. *JAMA Dermatol*. 2014;150:790-791.
6. Berger JR, Dean D. Neurosyphilis. *Handb Clin Neurol*. 2014;121:1461-1472.
7. Wolff T, Shelton E, Sessions C, et al. Screening for syphilis infection in pregnant women: evidence for the U.S. Preventive Services Task Force reaffirmation recommendation statement. *Ann Intern Med*. 2009;150:710-716.
8. Binnicker MJ, Jespersen DJ, Rollins LO. Direct comparison of the traditional and reverse syphilis screening algorithms in a population with a low prevalence of syphilis. *J Clin Microbiol*. 2012;50:148-150.
9. Morshed MG. Current trend on syphilis diagnosis: issues and challenges. *Adv Exp Med Biol*. 2014;808:51-64.
10. Liu F, Liu LL, Guo XJ, et al. Characterization of the classical biological false-positive reaction in the serological test for syphilis in the modern era. *Int Immunopharmacol*. 2014;20:331-336.
11. Clement ME, Okeke NL, Hicks CB. Treatment of syphilis: a systematic review. *JAMA*. 2014;312:1905-1917.
12. Holman KM, Hook EW 3rd. Clinical management of early syphilis. *Expert Rev Anti Infect Ther*. 2013;11:839-843.
13. Sena AC, Wolff M, Martin DH, et al. Predictors of serological cure and the serofast state after treatment in HIV-negative persons with early syphilis. *Clin Infect Dis*. 2011;53:1092-1099.
14. Sena AC, Wolff M, Behets F, et al. Response to therapy following retreatment of serofast early syphilis patients with benzathine penicillin. *Clin Infect Dis*. 2013;56:420-422.

REVIEW QUESTIONS

1. A 33-year-old man is referred to you after receipt of notification that a blood test result for syphilis performed at the time of a recent blood donation was positive. You learn that the test was a treponemal enzyme immunoassay (EIA). A rapid plasma reagin (RPR) test performed in follow-up was nonreactive. He reports being in a long-standing, mutually monogamous sexual relationship. Physical examination shows no evidence of syphilis. Possible explanations for his reactive EIA and nonreactive RPR syphilis test results are

A. Residual antibody reactivity after successful treatment of syphilis in the past
B. Previously undetected latent syphilis of unknown duration
C. A falsely positive test result
D. He has acquired syphilis so recently that his RPR test result is not yet positive
E. All of the above

Answer: E Results of recently developed treponemal antigen EIA tests for syphilis are not uncommonly positive while results of nontreponemal tests, such as the RPR and Venereal Disease Research Laboratory (VDRL) tests, are negative. In most settings (and in this patient), unless there is a history of prior treated syphilis, this typically reflects a falsely positive test result; more uncommonly, it could represent previously undetected latent syphilis or, very rarely, recently acquired syphilis before the development of a reactive RPR or VDRL test result (lesions of primary syphilis are typically present in these situations).

2. You are asked to recommend management for a 27-year-old woman with secondary syphilis who is midway through her second trimester of pregnancy. The patient has a well-documented history of penicillin allergy. The recommended therapy for this patient is

A. Doxycycline 100 mg twice daily for 14 days
B. Azithromycin 2.0 g orally as a single dose
C. Ceftriaxone 500 mg IM or IV once daily for 10 days
D. Penicillin desensitization followed by a single intramuscular injection of 2.4 million units of benzathine penicillin
E. Penicillin desensitization followed by two intramuscular injections of 2.4 million units of benzathine penicillin administered a week apart

Answer: D For treatment of syphilis in pregnancy, there are no recommended alternatives to penicillin for therapy. Doxycycline could cause dental staining in her child, and neither doxycycline nor azithromycin has been studied in this situation. As a result, pregnant, penicillin-allergic patients should be desensitized and treated as recommended for the stage of disease they have. There is no evidence that pregnant patients need treatment with higher doses or prolonged therapy.

3. In 2013, early syphilis was most common among minority groups and particularly common among

A. Persons who use illicit drugs, such as cocaine or heroin
B. Men who have sex with other men
C. Adolescents
D. Commercial sex workers
E. Heterosexuals in monogamous relationships

Answer: B During the past 50 years, the epidemiology of syphilis has periodically shifted. Following an epidemic of syphilis in heterosexual men and women in the 1990s associated with a contemporary epidemic of crack cocaine use, the epidemiology of early syphilis again shifted. Since the beginning of the 21st century, syphilis in the United States has disproportionately occurred among men who have sex with men, including those infected with human immunodeficiency virus (HIV). As a result, annual (or more frequent, depending on sexual history) screening for syphilis is recommended for men who have sex with men and all persons with HIV infection.

4. A 34-year-old man with HIV infection well controlled by antiviral therapy presents with the acute onset of right-sided weakness. His CD4 lymphocyte count is 540/mm^3. In the course of his evaluation, he is found to have reactive RPR (positive at a 1 : 64 dilution) and *T. pallidum* particle agglutination (TP-PA) test results. Lumbar puncture reveals the following cerebrospinal fluid (CSF) findings: cell count, 32 cells (100% mononuclear); protein level, 54 mg/dL; glucose concentration, 63 mg/dL (simultaneous serum glucose concentration, 10 mg/dL); and a nonreactive CSF VDRL test response. The most likely diagnosis in this man is

A. Latent syphilis with CSF abnormalities attributable to his HIV infection
B. Meningovascular syphilis
C. Embolic stroke
D. Hypertensive stroke

Answer: B This patient is a member of a group at risk for acquisition of syphilis (HIV-infected men who have sex with other men), is relatively young to have atherosclerotic stroke, and has CSF abnormalities consistent with neurosyphilis. His RPR test result is consistent with active syphilis. The CSF VDRL test response is reactive in less than 70% of persons with neurosyphilis and need not be present for diagnosis of neurosyphilis. Whereas HIV infection can lead to elevations of CSF mononuclear cells, this rarely exceeds 20 cells unless other pathologic processes are present. Although other possibilities for stroke in relatively young persons should be considered, the most likely diagnosis in this man is meningovascular neurosyphilis.

5. A patient presents to you reporting that he was just notified that a casual sex partner he had sex with about 6 months ago was diagnosed with secondary syphilis shortly afterward. His exposures included multiple episodes of genital and oral intercourse. On physical examination, there is no clinical evidence of syphilis. At this point, in addition to serologic testing for syphilis, recommended management is

A. Preventive treatment with intramuscular injections of 2.4 million units of benzathine penicillin G
B. Serologic testing at this time and again 6 months in the future (treatment based on serologic test results)
C. Serologic testing at this time; no subsequent testing is needed (treatment based on serologic test results)
D. Reassurance that based on the fact that there is no history of lesions and his negative test results, there is no reason for concern and no further follow-up or treatment is needed

Answer: C Because lesions of syphilis may be painless and subtle, serologic testing is recommended for all persons reporting exposure to a partner with syphilis. Preventive penicillin is recommended treatment for persons exposed to sex partners within the preceding 90 days. In this patient, the exposure was 6 months ago, so only serologic testing is recommended. There is no need for treatment if the serologic test result for syphilis is negative and no need for testing beyond this time.

320 NONSYPHILITIC TREPONEMATOSES

EDWARD W. HOOK III

DEFINITION

The nonsyphilitic treponematoses—yaws, endemic syphilis (previously known as bejel), and pinta—are the spirochetal diseases caused by *Treponema pallidum* subspecies (yaws and endemic syphilis) or by the closely related organism *Treponema carateum* (pinta). Like syphilis, the nonsyphilitic treponematoses are usually transmitted through direct contact with an infectious cutaneous or mucosal lesion. The natural history of the nonsyphilitic treponematoses also has a number of similarities to that of syphilis (Chapter 319).

The Pathogen

Yaws is caused by *T. pallidum* subspecies *pertenue*, endemic syphilis is caused by *T. pallidum* subspecies *endemicum*, and pinta is caused by *T. carateum*. The *T. pallidum* subspecies causing nonsyphilitic treponematoses are closely related to *T. pallidum* subspecies *pallidum*, which causes venereal syphilis; there is a high degree (more than 99%) of DNA homology, and they share unique pathogen-restricted antigens.[1] Analyses of recently described genetic sequence variations among *T. pallidum* subspecies promise the eventual clarification of pathophysiologic differences among the subspecies as well as answers to the age-old question of the origins of syphilis. Like *T. pallidum*, these treponemes are spirochetal bacteria with helical structures and measure about 0.2 µm in diameter and 10 µm in length. They are visible by dark-field microscopy but cannot be cultivated for prolonged periods in vitro.

EPIDEMIOLOGY

Worldwide, the nonsyphilitic treponematoses are rare. However, rates are increasing (particularly for yaws) in some regions where previous World Health Organization (WHO)–coordinated control programs had dramatically reduced disease prevalence. Yaws is prevalent in moist, humid regions, including rural areas of tropical Africa, the Americas, Southeast Asia, and Oceania. The highest incidence occurs in children between 2 and 5 years of age. Endemic syphilis occurs in more arid climates, including Africa, eastern Mediterranean countries, the Arabian peninsula, central Asia, and Australia. Pinta occurs in rural areas of tropical Central and South America and affects mostly older children and adolescents. Humans are the only known carriers of the nonsyphilitic treponematoses, although *T. pallidum* strains associated with genital lesions have recently been described in African baboons, potentially providing a clue to the origin of human treponemal infections, including syphilis.[2] The spirochete enters the skin only after it is broken, such as by a scratch or an insect bite. Transmission is believed to occur by contact of the skin directly or by indirect contamination through hands or fomites; it is facilitated by conditions of poor personal hygiene and crowding.

PATHOBIOLOGY

Primary nodular or ulcerative lesions typically develop at sites of inoculation after an incubation period of several weeks. Untreated primary lesions serve as a source for local spread through scratching or for hematogenous dissemination, which gives rise to a secondary stage of infection characterized by the development of widespread manifestations involving the skin, lymph nodes, and bone or cartilage. Without therapy, the primary and secondary manifestations of infection resolve, and the infection becomes latent, detectable only with serologic testing; however, periodic recurrent secondary manifestations may occur for several years. A proportion of persons with long-standing untreated infection are at risk for late sequelae, which may include bone deformity, destruction of nasal cartilage, or chronic skin changes. Unlike syphilis, the nonsyphilitic treponematoses are primarily diseases of children, are not transmitted across the placenta, and do not invade the central nervous system to cause clinical disease.

CLINICAL MANIFESTATIONS

Yaws,[3] the most common nonsyphilitic treponematosis, produces a skin papule at the inoculation site after an incubation period of 3 to 4 weeks. The most common sites are the legs and buttocks. The papule enlarges, ulcerates, and forms a serous crust from which treponemes can be recovered. Regional lymphadenitis may accompany the papule, which heals spontaneously within 6 months. A generalized secondary rash occurs before or after the initial lesion heals; this rash is also papular and is often covered with brown crusts. Relapsing crops of lesions can occur. Papillomas may result, and the plantar surfaces of the feet are involved with hyperkeratotic lesions. Periostitis of the long bones leads to tenderness, and fever may be present. Relapsing lesions of early yaws may occur during a period of several years and result in chronic ulcerations and destructive gummatous lesions affecting the skin and bones.

Endemic syphilis produces patches on the mucous membranes of the oral cavity and pharynx and can cause split papules at the mucocutaneous junction of the oral angles. Anal, genital, and other intertriginous skin areas can be affected by lesions that resemble secondary syphilis. Regional lymphadenitis is common, and generalized rashes are rare. Healing of these early lesions is followed by latency, manifested as seropositivity, or by late lesions that resemble gummatous tertiary syphilis (Chapter 319). Lesions include nodular skin ulcers, bone deformities, and ulcerative lesions that can perforate the palate.

Pinta starts similarly as a cutaneous papule with regional lymphadenitis, followed by a generalized maculopapular eruption. One to 3 years after healing of the initial lesion, large hyperpigmented brown or blue macules develop; they subsequently lose their pigment and become white. The time required for lesions to pass through these stages varies, so the same patient may have coexisting areas of increased pigment and loss of pigment.

DIAGNOSIS

The clinical differentiation of the nonsyphilitic treponematoses from one another and from syphilis may be challenging, requiring the integration of epidemiologic features, clinical findings, and supportive but not diagnostic laboratory test results. The skin lesions of the endemic treponematoses may resemble other cutaneous processes, including impetigo (Chapter 439), scabies, cutaneous fungal infections (Chapter 438), and other diseases. By dark-field microscopy, the causative spirochetes from early skin lesions can be observed directly; however, dark-field microscopy is rarely available in settings where nonsyphilitic treponematoses are seen. There is no specific test for any of the nonsyphilitic treponematoses, but serologic tests for syphilis detect cross-reacting antibodies in these diseases. The rapid plasma reagin (RPR) test, the Venereal Disease Research Laboratory (VDRL) test, and the fluorescent treponemal antibody absorption (FTA-ABS) test as well as other specific treponemal tests give positive results if serum is obtained at least 2 weeks after the lesions initially appear.

PREVENTION

The prevalence of these diseases was reduced dramatically in the 1950s by mass penicillin treatment campaigns. The WHO campaign treated about 53 million cases of yaws and 350,000 cases of pinta in the 1950s, with good results. These campaigns, however, were not adequate to eradicate the disease, and in recent years, the prevalence of yaws has increased, with foci of infection now reported in sub-Saharan Africa, several Pacific islands including Indonesia, and the Amazon region of South America. Current estimates are that as many as 2.5 million persons are infected worldwide, 75% of whom are younger than 15 years. Whereas penicillin is effective for both treatment and prevention of infection, requirements for cold chain transport, parenteral administration, and allergies sometimes compromise the utility of the drug.

Recent studies demonstrating the efficacy of single-dose oral azithromycin have expanded practical intervention strategies and led to renewed emphasis on yaws eradication by the WHO with targeted mass therapy.[4]

TREATMENT AND PROGNOSIS

Rx

Single-dose, long-acting benzathine penicillin G, 1.2 million units intramuscularly, has been the preferred treatment in patients with early lesions. For patients with late manifestations, this therapy should be repeated twice at approximately 7-day intervals. The early lesions heal rapidly, and most seropositive cases convert to seronegative status. Late destructive lesions take longer to show improvement. A randomized trial demonstrated oral azithromycin, 30 mg/kg up to a maximal dose of 2.0 g, to be as effective as penicillin for yaws therapy, providing the first readily administered, single-dose alternative to penicillin for treatment and prevention of nonsyphilitic treponematoses.

Grade A

Grade A Reference

A1. Mitja O, Hays R, Ipai A, et al. Single-dose azithromycin versus benzathine benzylpenicillin for treatment of yaws in children in Papua New Guinea: an open-label, non-inferiority, randomized trial. *Lancet*. 2012;379:342-347.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Giacani L, Lukehart SA. The endemic treponematoses. *Clin Microbiol Rev*. 2014;27:89-115.
2. Harper KN, Fyumagwa RD, Hoare R, et al. *Treponema pallidum* infection in the wild baboons of East Africa: distribution and genetic characterization of the strains responsible. *PLoS ONE*. 2012;7:e50882.
3. Mitja O, Asiedu K, Mabey D. Yaws. *Lancet*. 2013;381:763-773.
4. Mitja O, Hays R, Rinaldi AC, et al. New treatment schemes for yaws: the path toward eradication. *Clin Infect Dis*. 2012;55:406-412.

REVIEW QUESTIONS

1. Nonsyphilitic treponematoses can be differentiated from syphilis by

A. Serologic testing
B. Response to therapy
C. Clinical and epidemiologic characteristics
D. Dark-field microscopy

Answer: C Traditionally, the human treponematoses have been differentiated on the basis of their clinical manifestations and epidemiologic characteristics because the etiologic agents are indistinguishable in the laboratory. They all respond to penicillin therapy.

2. The clinical course of untreated nonsyphilitic treponematoses is characterized by

A. Lesions at the site of inoculation that resolve without therapy
B. Risk for congenital transmission
C. Risk for central nervous system disease
D. Uniform progression to chronic, deforming disease in all untreated persons

Answer: A Unlike syphilis, the nonsyphilitic treponematoses are not transmitted across the placenta and do not invade the central nervous system to cause clinical disease. Only a proportion of individuals with long-standing untreated infection are at risk for late sequelae like bone deformities, nasal cartilage destruction, or chronic skin changes.

321

LYME DISEASE

GARY P. WORMSER

DEFINITION

Lyme disease (also known as Lyme borreliosis) is a zoonotic infection that is transmitted by certain *Ixodes* tick species and caused by a group of related spirochetes referred to formally as *Borrelia burgdorferi* sensu lato, or more simply as Lyme borrelia.[1] Lyme disease was first described in 1977 after an investigation of a cluster of cases of arthritis among children living in the area of Lyme, Connecticut. With more than 25,000 cases reported annually, it is the most common vector-borne infection in the United States; Lyme disease is also an infection of public health importance in both Europe and Asia. The most common clinical manifestation is a characteristic skin lesion called erythema migrans. This lesion is a result of inflammation associated with the centrifugal spread of the spirochete within the skin from the site where the tick deposited the microorganism. The spirochete may also spread hematogenously to other skin locations, resulting in secondary erythema migrans skin lesions, or to non-skin sites, such as the joints, nervous system, or heart, leading to a variety of extracutaneous clinical manifestations.[2]

The Pathogen

In the United States, the only species of Lyme borrelia known to cause human infection is *B. burgdorferi* (also referred to as *B. burgdorferi* sensu stricto). Although *B. burgdorferi* also causes Lyme disease in Europe, collectively other species of Lyme borrelia that can be distinguished genotypically account for the majority of infections there, especially *Borrelia afzelii* and *Borrelia garinii*. The fact that at least six species of Lyme borrelia may cause infection in Europe has created serodiagnostic challenges and accounts for a wider variety of possible clinical manifestations there than in the United States (see later). *B. garinii* appears to be the most neurotropic and *B. burgdorferi* the most arthritogenic among the species of Lyme borrelia.

Lyme borrelia are motile, microaerophilic, spirochetal bacteria with 3 to 10 loose coils arranged in a helical shape. The cells are 10 to 30 µm in length and 0.2 to 0.5 µm in width and contain at least seven periplasmic flagella that are responsible for the organism's motility. Borrelia are too thin to be seen by Gram stain, but live organisms can be visualized by dark-field or phase contrast microscopy. In tissues, they can be recognized by light microscopy after application of silver stains or by fluorescent microscopic methods. *B. burgdorferi* was the first spirochete for which the complete genome was sequenced. Genetic studies suggest the nearly complete absence of biosynthetic pathways, making the microorganism dependent on its environment for nutritional requirements.[3] Lyme borrelia can be grown in vitro in a highly enriched culture medium.

EPIDEMIOLOGY

In the United States, more than 95% of cases of Lyme disease are concentrated in just 14 states: 12 eastern states and two in the North Central region. The states with the largest number of cases are Pennsylvania, Massachusetts, New York, New Jersey, and Connecticut. *Ixodes scapularis* (also known as the deer tick or black-legged tick) is the tick vector in these states. *Ixodes pacificus* is the vector for cases that occur in the Northwest region. Lyme disease also occurs in limited areas of Canada. Cases of Lyme disease occur throughout the temperate regions of Europe and are especially common in Scandinavia and countries of central Europe such as Slovenia, Austria, and Germany. *Ixodes ricinus* transmits the infection in Europe, and *Ixodes persulcatus* is the vector in the Asian region of Russia, China, and Japan.

The principal reservoirs for Lyme borrelia (i.e., source of infection for ticks) in the United States and Eurasia are small mammals such as mice and certain species of birds. Deer play an essential role in the life cycle of the *I. scapularis* tick species but are not a competent reservoir for *B. burgdorferi.*

Although Lyme borrelia exist in enzootic cycles in the southern United States, cases of Lyme disease arising indigenously have not been well documented in states south of Virginia. A skin lesion that resembles erythema migrans does occur in the southern United States, however, but it is associated with the bite of the *Amblyomma americanum* tick, a tick species that is not a competent vector for *B. burgdorferi*. This condition is of unknown etiology and is referred to as southern tick-associated rash illness (STARI).

The likelihood of acquiring Lyme disease is directly related to exposure to environments in which infected ticks are present. Of the three feeding stages in the life cycle of *I. scapularis*, the second or nymphal stage is the most important epidemiologically for transmission of infection to humans. The first or larval stage is uninfected and cannot transmit this infection. Although the third stage (i.e., the adult stage of the tick) is more likely to be infected with *B. burgdorferi* than the nymphal stage is, it is less important in transmission to humans because this stage is present in smaller numbers in the environment and because there is less human activity outdoors during the time periods in the spring and fall when this stage is seeking a blood meal. In addition, adult ticks are larger and their bites cause more skin irritation than bites of nymphal ticks do, thereby increasing the likelihood that they will be noticed and detached by a person who has been bitten. If they are not removed, *Ixodes* ticks will usually feed for at least 3 days (Fig. 321-1). Transmission of *B. burgdorferi* by *I. scapularis* or *I. pacificus* ticks is typically delayed for more than 36 hours from the start of the blood meal, providing the opportunity to prevent infection simply by finding and removing the tick.[4] Transmission of *B. afzelii* by the European tick *I. ricinus* is considerably faster, however, often occurring within the first 24 hours of feeding.

Most cases of erythema migrans in the United States occur during June through August. There is a bimodal age distribution, with the highest incidences in children 5 to 9 years old and in adults 45 to 54 years of age, but individuals of all ages are at risk. The reported incidence of Lyme disease in the United States is rising, partially as a result of expansion of the deer population and the spread of infected *I. scapularis* ticks to new geographic areas.

Extracutaneous manifestations are somewhat less likely than erythema migrans to occur during June to August because the time from the tick bite until the onset of these manifestations is longer. Because adult *I. scapularis* ticks may become active on warm days during the winter, cases of erythema migrans may occasionally occur even in the colder months.

PATHOBIOLOGY

B. burgdorferi resides in the midgut of the *Ixodes* tick through attachment of its outer surface protein A (OspA) to the cells lining this site. With the onset

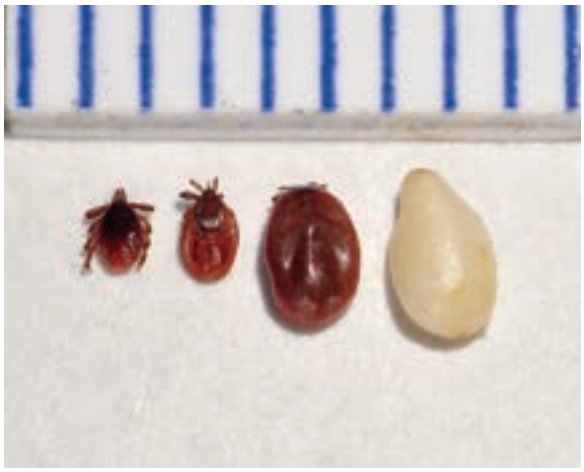

FIGURE 321-1. From left to right is an unfed nymphal stage *Ixodes scapularis* tick, a nymphal stage *I. scapularis* tick after about 48 hours of feeding, a nymphal stage *I. scapularis* tick after about 126 hours of feeding, and a sesame seed. The distance between the ruler marks is 1 mm. (Courtesy Kam Truhn of Fordham University.)

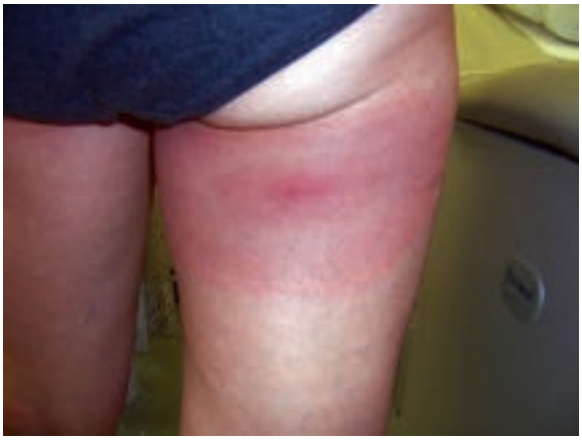

FIGURE 321-2. Erythema migrans skin lesion on the posterior aspect of the right thigh.

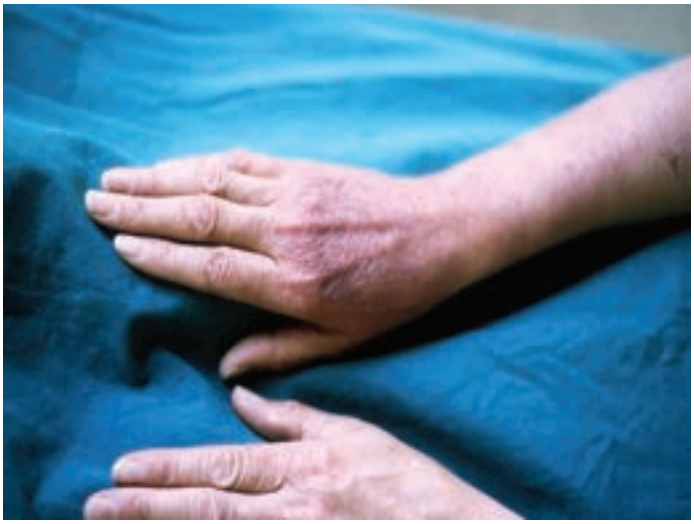

FIGURE 321-3. Acrodermatitis chronica atrophicans. This late cutaneous manifestation of Lyme borreliosis is characterized by slowly expanding red violaceous lesions that typically involve the dorsal surfaces of acral sites and do not heel spontaneously. The lesions are initially inflammatory and later on more and more atrophic. (Courtesy Dr. Franc Strle.)

of the blood meal, because of changes in temperature, pH, and probably other factors, the spirochete increases in number and undergoes a sequence of phenotypic changes including downregulation of OspA expression and upregulation of another outer surface protein, OspC.[5] This set of events releases the spirochete from the midgut and permits migration to the salivary gland. At that site, OspC binds to a salivary protein, Salp15, that is induced during the blood meal in borrelia-infected ticks; this protects the spirochete from antibody-mediated killing and aids in transmission to a vertebrate host. *Ixodes* tick saliva also contains other components with anti-inflammatory and immunomodulatory actions that serve to promote infection with Lyme borrelia and other tick-transmitted pathogens. Expression of OspC by the spirochete in the mammalian host is essential for infection to be established because in some manner this protein protects the microorganism from immediate elimination by the innate immune system.

The spirochete is deposited by the tick into the skin rather than directly into the blood stream. Hematogenous dissemination seems to be an important mechanism responsible for spread of the spirochete to other sites. Alternatively, spread of the spirochete to other sites might occur through tissue planes. The likelihood of entry into the blood stream is affected by the strain of Lyme borrelia causing the infection.[6] Unlike patients who are bacteremic with more conventional pathogens, patients with spirochetemia rarely appear "septic." In one study, only 5% of 93 spirochetemic patients were febrile when the blood culture specimen was obtained, and almost none was found to have leukocytosis. The lack of fever and other clinical signs of sepsis may be due to the absence of lipopolysaccharide within the borrelial cell wall.

Infection of humans or animals elicits innate and adaptive immune responses resulting in both macrophage and antibody-mediated killing of the spirochete. The inflammatory response in tissue typically shows an infiltration of lymphocytes, macrophages, and plasma cells, although granulocytes predominate in synovial fluid samples of Lyme arthritis patients. The potential for persistence of infection despite a robust humoral and cellular immunologic response is, however, typical of infection with Lyme borrelia, as it is with *Treponema pallidum* infection. The virulence factors responsible for persistence of infection include the spirochete's ability to downregulate expression of certain immunogenic surface-exposed proteins, including OspC, and to alter rapidly and continually by recombination the antigenic properties of a surface lipoprotein known as variable major protein–like sequence expressed (VlsE). In addition, the ability of the spirochete to bind avidly to various components of the extracellular matrix may also contribute to persistence.

All of the objective clinical manifestations of Lyme disease are thought to be due to an inflammatory response to live spirochetes or to their undegraded antigens. Obliterative endarteritis has been seen histologically in synovial tissue, but its importance in pathogenesis is unclear. Lyme borrelia are not known to produce toxins. In humans, the only role so far established for host genetic factors is in the development of antibiotic-refractory Lyme arthritis, which is seen most often in patients with certain HLA DR alleles, some of which coincide with those associated with rheumatoid arthritis.

CLINICAL MANIFESTATIONS

The clinical manifestations are often categorized as follows:

- Early localized infection, typically manifested by a single erythema migrans (Fig. 321-2) skin lesion, with or without viral infection–like symptoms, but without objective extracutaneous manifestations
- Early disseminated infection, usually manifested by multiple erythema migrans skin lesions or by an objective manifestation of early neurologic Lyme disease or Lyme carditis
- Late disease, usually manifested by arthritis but may also include certain rare neurologic manifestations or the skin condition known as acrodermatitis chronica atrophicans (Fig. 321-3)

Children and adults have similar clinical manifestations. The expected frequency of the various clinical presentations is well illustrated by a study of 313 cases of Lyme disease diagnosed in Wurzburg, Germany, during a 12-month period. In this series, erythema migrans by itself was seen in 89% of cases, early neurologic manifestations in 3%, cardiac manifestations in less than 1%, borrelial lymphocytoma in 2%, arthritis in 5%, and acrodermatitis chronica atrophicans in 1%. None of the patients had late neurologic Lyme disease. A similar distribution of cases has been seen in recent case series in the United States, except for the absence of borrelial lymphocytoma and acrodermatitis chronica atrophicans. In earlier studies in the United States from the 1980s, there was a much higher proportion of patients who had neurologic, cardiac, or joint manifestations. This may have been due to a bias of ascertainment in the older studies or to improved recognition and

TABLE 321-1 CLINICAL SYMPTOMS AND SIGNS PRESENT IN AT LEAST 20% OF PATIENTS WITH ERYTHEMA MIGRANS

CHARACTERISTIC	UNITED STATES (95% CI)
Viral infection-like	65% (52-76%)
Fatigue	47% (37-58%)
Headache	36% (27-46%)
Myalgias	35% (26-45%)
Arthralgias	35% (25-46%)
Fever	33% (23-43%)
Stiff neck	31% (21-43%)
Lymphadenopathy	22% (13-33%)
Dysesthesia	20% (12-32%)
	EUROPE (95% CI)
Viral infection-like	37% (27-49%)
Dysesthesia	35% (25-47%)
Headache	20% (14-29%)

CI = confidence interval.
Modified from Tibbles CD, Edlow JA. Does this patient have erythema migrans? *JAMA*. 2007;297:2617-2627.

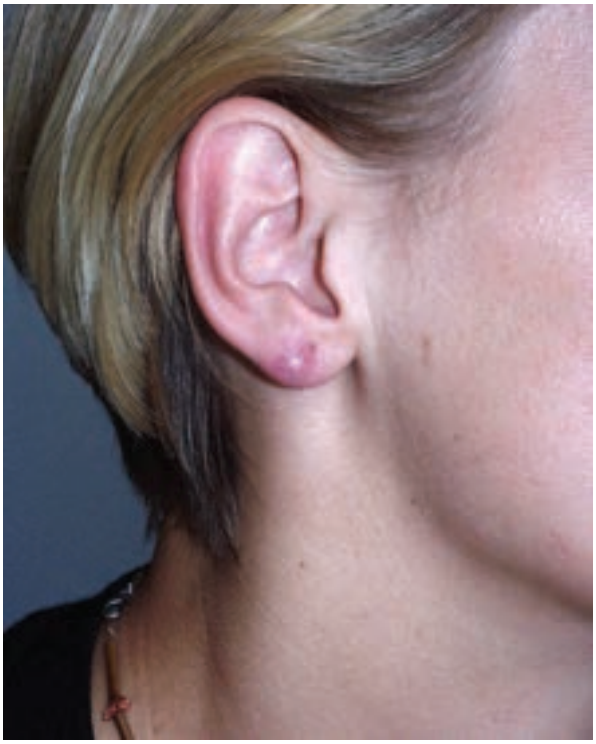

FIGURE 321-4. Borrelial lymphocytoma. A rare cutaneous manifestation of Lyme disease that is seen predominantly outside the United States, it is manifested as a solitary bluish red swelling most commonly on the earlobe of children and the breast in adults. (Courtesy Dr. Franc Strle.)

treatment of patients with erythema migrans more recently, thereby preventing the development of extracutaneous complications because they occur later.

Early Localized Infection

Single Erythema Migrans Skin Lesion

Erythema migrans is by far the most common clinical manifestation of Lyme disease. Although the appearance of the skin lesion is often distinctive (see Fig. 321-2), it is not pathognomonic for Lyme disease. Erythema migrans appears 7 to 14 days (range, 3 to 30 days) after tick detachment and is characterized by an expanding, flat to slightly raised, erythematous skin lesion (usually ≥5 cm in diameter) that is round or oval. The bite mark from the preceding tick bite can sometimes be identified at or near the center of the lesion and is called a punctum. Approximately 80% of patients in the United States with erythema migrans have just a single skin lesion. Nonspecific viral infection–like symptoms or signs, such as malaise, neck pain, headache, fatigue, migratory arthralgias, or chills and fever, may be present (Table 321-1) but are more common in patients infected with *B. burgdorferi* or *B. garinii* compared with *B. afzelii*. Prominent respiratory or gastrointestinal symptoms are highly atypical for Lyme disease. An acute febrile illness in the absence of a skin lesion or other objective clinical manifestation has been attributed to early Lyme disease, but the possibility of misdiagnosis is greater in this situation because of the potential for false-positive serologic test results.

Erythema migrans skin lesions can vary in appearance. Some (especially lesions of short duration) are nearly uniform in color, whereas others may show central clearing or a target-like appearance. About 5% have a vesicular-pustular center. Erythema migrans on the lower extremities may sometimes be purpuric. Erythema migrans lesions may be scaly when they are long-standing and fading or after topical corticosteroid creams have been applied. The most common locations include the thigh, back, shoulder, and calf. Lesions are often asymptomatic but can be mildly painful or pruritic, and tender regional lymphadenopathy may be present. The majority of U.S. patients with erythema migrans, as for all other clinical manifestations, do not recall a preceding tick bite.

Certain signs and symptoms that have been described in small numbers of patients with erythema migrans in early case series, such as hepatomegaly, splenomegaly, sore throat, conjunctivitis, or testicular swelling, may have been coincidental findings. There has been no microbiologic confirmation of borrelia infection at these sites.

Borrelial Lymphocytoma

Borrelial lymphocytoma is a rare cutaneous manifestation of Lyme disease that almost never occurs in the United States. It often is manifested at or near a preceding tick bite as a solitary bluish red swelling with a diameter of up to a few centimeters. The most common locations are the earlobe in children and the breast in adults (Fig. 321-4). Histologic examination shows a dense polyclonal infiltration of the cutis and subcutis by predominantly B lymphocytes, frequently with germinal center formation. Histologic evaluation may be necessary to exclude malignant disease for patients with suspected borrelial lymphocytoma at a location other than the earlobe.

Early Disseminated Infection

Multiple Erythema Migrans Skin Lesions

In the United States, approximately 20% of patients with erythema migrans have multiple skin lesions at the time of presentation. Secondary erythema migrans skin lesions can be smaller than 5 cm, do not have a punctum, and are usually not tender or pruritic. They arise from hematogenous dissemination to the skin rather than from additional tick bites.

Early Neurologic Lyme Disease

Within weeks to several months after infection, patients may develop neurologic manifestations, the most common of which are cranial neuropathy (particularly peripheral seventh nerve palsy that may be bilateral), lymphocytic meningitis, and sensory (often painful) radiculopathy.[7] Less common manifestations, among others, include mononeuritis multiplex (multifocal involvement of anatomically unrelated nerves) and brachial or lumbosacral plexopathies. A pseudotumor cerebri–like picture has been reported occasionally in children. The presence of a concomitant erythema migrans lesion or the patient's recollection of a recent lesion consistent with erythema migrans may be helpful diagnostically. Studies suggest that approximately 90% of children with Lyme meningitis in the United States have at least one of the following three findings: concomitant erythema migrans, cranial nerve palsy, or papilledema.

Cardiac Lyme Disease

Weeks to months after infection, patients may develop cardiac manifestations of Lyme disease, most often fluctuating degrees of atrioventricular heart block or other manifestations of a myopericarditis, which may cause the patient to complain of lightheadedness, palpitations, dyspnea, chest pain, or syncope. Heart block is typically at or above the atrioventricular node.

Erythema migrans is often but not invariably present concurrently. Valvular dysfunction is not known to occur, and chronic cardiomyopathy has been reported only rarely in Europe.

Late Lyme Disease

Lyme Arthritis

If patients with erythema migrans in the United States are not treated with antibiotics, approximately 60% will develop a monarticular or oligoarticular arthritis at a mean time of 6 months after disease onset (range, 4 days to as long as 2 years). In untreated patients, Lyme arthritis is characterized by intermittent attacks of synovitis that last for a few weeks to several months. One or two joints are involved at a time. Primarily large joints are affected,[8] but there may be involvement of the temporomandibular joint, small joints, and periarticular sites. The most commonly involved joint is the knee. Baker's cysts may form and rupture. Joint swelling is often pronounced, but pain is usually relatively modest. Particularly in children, there may be concomitant fever, but adults are often minimally symptomatic aside from the arthritis.

In about 10% of adult patients with Lyme arthritis in the United States, involvement of a large joint (almost always the knee) may persist despite appropriate antibiotic treatment. Erosion of cartilage or bone may develop in such cases.

Late Neurologic Lyme Disease

After months to years of infection, late neurologic manifestations may develop. These include encephalomyelitis, peripheral neuropathy, and encephalopathy. Because most patients with Lyme disease are now diagnosed and treated early in the course of infection, these more indolent forms of neurologic Lyme disease are rare.

In untreated patients, encephalomyelitis has been monophasic and slowly progressive, mainly involving white matter. It is the most severe neurologic manifestation and, although infrequent, is probably more common in Europe than in the United States. Cerebrospinal fluid (CSF) examination typically shows a lymphocytic pleocytosis, a moderately elevated protein level, and a normal glucose level, with evidence of intrathecal production of antibody to borrelia. Magnetic resonance imaging of the affected part of the brain or spinal cord can demonstrate areas of inflammation, typically with increased signal on T2 and fluid attenuation inversion recovery imaging and enhancement after administration of contrast material.

In the United States, peripheral neuropathy typically is manifested as a mild, diffuse "stocking-glove" process. Patients typically complain of intermittent limb paresthesias and sometimes radicular pain. The most common abnormality on neurologic examination is reduced vibratory sensation of the distal lower extremities. Electrophysiologic studies show a patchy axonal neuropathy. Nerve biopsy shows axonal loss and small perivascular collections of lymphocytes without spirochetes. CSF findings are often normal without evidence of intrathecal antibody production.

Encephalopathy is an imprecisely defined clinical entity characterized by mild abnormalities of memory or other cognitive functions that are demonstrable on either a careful mental status examination or on formal neuropsychological testing. CSF examination findings may be completely normal or may show intrathecal antibody production, mild CSF protein elevation, or a mild pleocytosis. Cranial imaging studies may occasionally demonstrate focal areas of presumed parenchymal inflammation, but findings are most often normal.

Acrodermatitis Chronica Atrophicans

Acrodermatitis chronica atrophicans is a late skin manifestation of Lyme disease that is most commonly seen in women older than 40 years.[9] This skin lesion develops insidiously several years after initial infection, usually on the extensor surfaces of the hands and feet. Early lesions are characterized by a slight bluish red discoloration and doughy swelling. Histologic examination shows lymphocytes and plasma cells in the skin and sometimes in the subcutis, with or without atrophy. Initially unilateral, the lesion may later become bilateral. Over time, there is resolution of the edema with development of skin atrophy. Nodules may develop over bone prominences. About two thirds of patients have an associated peripheral neuropathy of the affected extremity, manifested primarily as local sensory loss.

Although presumably any of the species of Lyme borrelia may cause acrodermatitis chronica atrophicans, by far the most common etiologic agent is *B. afzelii*.[1] Therefore, this manifestation is rarely seen in the United States.

DIAGNOSIS

General Laboratory Testing

White blood cell count, hemoglobin and hematocrit levels, and platelet count are usually normal in Lyme disease, unless coinfection with *Anaplasma phagocytophilum*, *Babesia microti*, or a tick-borne encephalitis virus is present (see later). Lymphopenia, however, may be found in the absence of a recognized coinfection.[10] In patients with erythema migrans, mild abnormalities of liver function tests (particularly elevations of aspartate and alanine aminotransferase levels) can be seen in approximately 35% of patients. The erythrocyte sedimentation rate may be modestly elevated in all stages of Lyme disease, but values greater than 80 mm/hour are distinctly uncommon.

CSF examination in Lyme meningitis typically shows a pleocytosis with more than 90% lymphocytes, a modestly elevated protein level, and a normal glucose level. Synovial fluid examination in Lyme arthritis typically shows approximately 25,000 white cells/mm^3 (range, 500 to 110,000/mm^3) with a polymorphonuclear predominance.

Serologic Testing

Erythema migrans skin lesions may go unnoticed by the patient because of the absence of prominent local symptoms and occurrence on parts of the body that are difficult for the patient to visualize. Therefore, a complete skin examination should be performed for any patient thought to have early localized or disseminated Lyme disease. Erythema migrans is the only clinical manifestation sufficiently distinctive to allow a clinical diagnosis in the absence of a supporting laboratory test. Erythema migrans is diagnosed on the basis of recognition of the characteristic appearance of the skin lesion in persons who live in or have recently traveled to areas endemic for Lyme disease. Because of the short duration of infection at this stage, serologic assays for antibodies to Lyme borrelia are infrequently positive and thus should be obtained only in atypical cases, then in conjunction with convalescent-phase serologic testing 2 to 4 weeks after the acute sample is obtained (Table 321-2).

For non–erythema migrans presentations of Lyme disease, the mainstay of laboratory diagnosis is two-tier serologic testing in which the first tier test is usually a sensitive enzyme-linked immunosorbent assay (EIA). If the EIA result is positive or equivocal, separate IgM and IgG immunoblots are performed on the original serum sample. If symptoms have persisted for at least 4 weeks, then specifically the IgG immunoblot should be positive for the results to be interpreted as evidence of seropositivity. Untreated patients who remain seronegative for 6 to 8 weeks are unlikely to have Lyme disease, and other possible diagnoses should be pursued.

Omitting the first-tier EIA or interpreting the immunoblot with alternative criteria that are not evidence based will potentially decrease the specificity of testing and is not recommended. False-positive results on the IgM immunoblot may be due to cross-reactive antibodies that arise from polyclonal B-cell

TABLE 321-2 DIAGNOSIS OF LYME DISEASE

DIAGNOSTIC MODALITY	APPLICATION	COMMENT
Visual inspection	Erythema migrans	Usually seronegative at time of presentation
Two-tier serology with a positive IgM and/or IgG immunoblot	Lyme carditis	Look for concomitant erythema migrans
	Early neurologic Lyme disease	Look for concomitant erythema migrans; intrathecal antibody may be detectable before serum antibody in Europe; PCR sometimes positive in CSF
	Borrelial lymphocytoma	Biopsy may be needed to exclude malignant neoplasm
Two-tier serology with a positive IgG immunoblot	Lyme arthritis	PCR often positive in synovial fluid
	Late neurologic Lyme disease	Intrathecal antibody positivity expected in Lyme encephalomyelitis
	Acrodermatitis chronica atrophicans	

CSF = cerebrospinal fluid; Ig = immunoglobulin; PCR = polymerase chain reaction.

stimulation. Probably the most common cause of false-positive results, however, is the overreading of nonspecific weak bands.[11] Background rates of seropositivity, which may exceed 4% in highly endemic areas of the United States with even higher rates than this in Europe, may also confound the interpretation of seroreactivity. Therefore, a positive serologic test result does not mean that the patient necessarily has active Lyme disease. The positive predictive value is most informative when the pretest probability based on the clinical features is at least 20%. Serologic testing is not indicated in routine follow-up of patients after treatment as either IgM or IgG borrelial antibodies may persist for many years in successfully treated patients.

Testing for borrelial antibody that is produced locally in the central nervous system (i.e., intrathecal antibody) may be helpful in the diagnosis of neurologic Lyme disease and has been reported to precede detection of serum antibody in a minority of European patients. Positive test results for intrathecal antibody may persist, however, for long periods after successful antibiotic treatment.

Other Diagnostic Modalities

Culture for Lyme borrelia is not routinely done or available to diagnose Lyme disease. It is unnecessary for patients with erythema migrans and too insensitive for patients with extracutaneous manifestations of Lyme disease. In contrast, polymerase chain reaction (PCR) for detection of borrelial DNA is positive on synovial fluid specimens in up to approximately 80% of untreated patients with Lyme arthritis, and a positive result lends support for this diagnosis in a patient who is IgG seropositive. The sensitivity of PCR in CSF tends to be much lower, however, and was only approximately 5% in a study of children from the United States with early neurologic Lyme disease. A negative PCR result on either type of fluid does not exclude Lyme disease.

Differential Diagnosis

Erythema Migrans

Tick-bite hypersensitivity reactions can be mistaken for erythema migrans, but these reactions occur within 48 hours of a tick bite, are usually pruritic, and tend to wane within a few days. In contrast, erythema migrans skin lesions in untreated patients will last for a median time of approximately 4 weeks in the United States and even longer in Europe. Bacterial cellulitis rarely occurs at the most frequent skin sites for erythema migrans and would not be expected to demonstrate central clearing or a target-like appearance. Erythema migrans, unlike erythema multiforme, does not involve the mucous membranes, palms, or soles. Southern tick-associated rash illness (STARI) is the most likely diagnosis in patients with erythema migrans–like skin lesions who were bitten by an *A. americanum* tick or developed this lesion in the southern United States (see under Pathobiology). Other considerations in the differential diagnosis of erythema migrans that usually can be readily distinguished include tinea (often pruritic with a thin, raised, scaly border), nummular eczema (symmetrical pruritic lesions with a tendency to scale and crust), granuloma annulare (acral location, especially on the dorsum of hands and feet, relatively fixed in size and <5 cm in diameter), contact dermatitis (pruritic with streaking along the area of contact and a vesicular component), urticaria (raised, pruritic, and usually <5 cm in diameter), fixed drug eruption (usually on genitals, hands, feet, or face and fixed in size), pityriasis rosea (multiple, moderately pruritic lesions with peripheral scale and relatively fixed in size), and spider bite (often very painful and necrotic with a central eschar).

Extracutaneous Lyme Disease

Included in the differential diagnosis of early neurologic Lyme disease is Bell's palsy. Even in highly endemic areas of the United States, other causes of seventh nerve palsy outnumber Lyme disease by a margin of 3 : 1. Viral meningitis and mechanical radiculopathy can also potentially be confused with manifestations of early neurologic Lyme disease. Viral or other causes of myopericarditis may resemble cardiac Lyme disease. Many causes of synovitis might be considered in the differential diagnosis of Lyme arthritis, but the pattern of joint involvement, such as symmetrical small joint involvement in rheumatoid arthritis, is often distinctly different from that found in Lyme arthritis. Lyme encephalomyelitis may occasionally be confused clinically with a first episode of relapsing-remitting multiple sclerosis or primary progressive multiple sclerosis. Testing for borrelial antibody in serum (and CSF for the last condition) usually suffices to differentiate these conditions from Lyme disease.

TREATMENT

Rx

In vitro studies have shown that Lyme borrelia are highly susceptible to tetracyclines, most penicillins, and many second- and third-generation cephalosporins. *B. burgdorferi* is resistant to certain fluoroquinolones, rifampin, and first-generation cephalosporins. Whether macrolides are active in vitro depends on the borrelial strain tested and the assay technique used. Although most manifestations of Lyme disease will resolve spontaneously without treatment, antibiotic therapy may hasten resolution and prevent progression.

Oral antibiotic therapy is used to treat patients with erythema migrans (Table 321-3). Doxycycline, amoxicillin, and cefuroxime axetil are each highly effective and are the preferred agents for this indication. Macrolides such as azithromycin are somewhat less effective than other oral antibiotics and consequently are not recommended as first-line therapy.

Doxycycline alone among the first-line agents is effective against *A. phagocytophilum* coinfection and is the only agent for which a prospective clinical trial has demonstrated that just 10 days of treatment is effective. Doxycycline, however, may cause photosensitivity, which is a concern because early Lyme disease occurs most commonly during the summer months; in addition, this drug is relatively contraindicated in children younger than 8 years and in women who are pregnant or breast-feeding. When erythema migrans cannot be reliably distinguished from community-acquired bacterial cellulitis, either cefuroxime axetil or amoxicillin–clavulanate potassium (Augmentin) is preferred because these antimicrobials are generally effective against both types of infection.

Within 24 hours of initiation of antimicrobial therapy, up to 15% of patients treated for erythema migrans will experience a Jarisch-Herxheimer–like reaction characterized by an increase in the size or intensity of erythema in the skin lesion and more intense viral infection–like systemic symptoms. Fever, if present, should resolve within 48 hours and the skin lesion itself within 7 to 14 days. Other symptoms, such as fatigue or arthralgia, tend to improve but not invariably to resolve within this time frame, lasting for more than 3 months in one quarter of patients. Extending the initial course of treatment does not cause faster relief of symptoms. Oral antibiotic therapy is also used as first-line treatment for the other cutaneous manifestations of Lyme disease discussed elsewhere in this chapter and as initial treatment of patients with Lyme arthritis.

The preferred parenteral agent for Lyme disease is ceftriaxone because it is highly active against Lyme borrelia in vitro, crosses the blood-brain barrier well, and has a long serum half-life, allowing the convenience of once-daily administration. Alternative choices for parenterally administered antibiotics are cefotaxime and intravenous penicillin. Parenteral antibiotic therapy is recommended to treat patients with late neurologic Lyme disease and those with cardiac Lyme disease who are admitted to the hospital for monitoring (see Table 321-3). Parenteral antibiotics are often given to patients with Lyme arthritis who have failed to respond to one or more courses of oral antibiotic treatment.

TABLE 321-3 RECOMMENDED THERAPY FOR ADULT PATIENTS WITH LYME DISEASE*

THERAPY	MANIFESTATION	DURATION
Doxycycline 100 mg PO bid *or*	Erythema migrans	14 days
	Borrelial lymphocytoma	14 days
Amoxicillin 500 mg PO tid *or*	Acrodermatitis chronica atrophicans	21 days
	Lyme arthritis	28 days
Cefuroxime axetil 500 mg PO bid	Lyme carditis—mild	14 days
	Cranial neuropathy	14 days†
Doxycycline 100 mg PO bid	Lyme meningitis or radiculopathy in Europe and possibly in the United States	14 days
Ceftriaxone 2 g IV daily	Lyme arthritis that failed to respond to oral therapy	14-28 days
	Late neurologic Lyme disease	14-28 days
	Lyme carditis requiring hospitalization	14 days
	Lyme meningitis or radiculopathy in the United States	14 days
Azithromycin 500 mg PO daily	Erythema migrans in a patient intolerant of doxycycline and β-lactam antibiotics	6-10 days

*Regardless of the clinical manifestations of Lyme disease, complete response to treatment may be delayed beyond the treatment duration. Relapse may occur with any of these regimens; patients with objective signs of relapse may need another course of treatment.

†Although any one of first-line oral antibiotics appears to be effective in patients with cranial neuropathy, there is only limited experience in patients with a cranial neuropathy other than seventh nerve palsy or with agents other than doxycycline.

Modified from Wormser GP, Dattwyler RJ, Shapiro ED, et al. The clinical assessment, treatment, and prevention of Lyme disease, human granulocytic anaplasmosis, and babesiosis: clinical practice guidelines by the Infectious Diseases Society of America. *Clin Infect Dis*. 2006;43:1089-1134.

In the United States, parenteral therapy has been the preferred management strategy for early neurologic Lyme disease, especially for meningitis and radiculitis, with oral therapy reserved for patients with uncomplicated seventh nerve palsy. Studies conducted in Europe, however, have provided convincing evidence that oral doxycycline is just as effective as ceftriaxone for any of the primary manifestations of early neurologic Lyme disease.[A1,A2] Although the same may be true in the United States, studies are lacking. Other oral antibiotics, such as amoxicillin, have been used successfully to treat patients with uncomplicated seventh nerve palsy, but published data on efficacy are much more limited for these agents. Available data indicate that seventh nerve palsy will resolve, with or without antibiotic treatment, and that the rate of recovery is not accelerated by antibiotics. Therefore, the primary reason to treat such patients is to prevent the subsequent development of later complications, particularly Lyme arthritis.

The presence of either papilledema or sixth cranial nerve palsy may indicate the presence of increased intracranial pressure in patients with neurologic Lyme disease. The elevated pressure will typically fall in response to antibiotic therapy, but other measures conventionally used to lower pressure may need to be considered in individual cases.

Symptomatic patients with cardiac Lyme disease and those with high-grade first-degree atrioventricular heart block (PR interval of ≥300 msec) and second- or third-degree block should be hospitalized and closely monitored. Temporary cardiac pacing may be required. In treated patients, complete heart block generally resolves within 1 week, and lesser conduction disturbances resolve within 6 weeks.

Lyme arthritis typically responds to antibiotic treatment. Patients whose arthritis is improved but not resolved after an initial course of oral therapy may be re-treated with a second course of oral antibiotics, with parenteral antibiotic therapy reserved for those without any significant clinical response. Approximately 10% of adult patients in the United States, however, do not respond clinically to antibiotic therapy and are said to have antibiotic-refractory Lyme arthritis; this condition has been defined as persistent synovitis for at least 2 months after completion of a course of intravenous ceftriaxone (or 1 month after completion of two 4-week courses of an oral antibiotic) in conjunction with negative PCR test results on synovial fluid and on synovial tissue if it is available. Because these patients are no longer believed to be actively infected, they are customarily treated with nonsteroidal anti-inflammatory agents, intra-articular injections of corticosteroids, or disease-modifying antirheumatic drugs. Arthroscopic synovectomy has also been used successfully for patients with this condition.

Pregnant patients with Lyme disease are generally treated similarly to nonpregnant patients except that doxycycline should be avoided because of the potential for adverse effects to both the fetus and the mother. No published data convincingly support a congenital Lyme disease syndrome.

Post–Lyme Disease Symptoms and Syndrome

The outcome of treatment in most patients with erythema migrans is excellent. Studies show, however, that when questioned at 6 months or more after treatment of erythema migrans, approximately 5 to 15% of patients will report purely subjective symptoms such as fatigue or musculoskeletal pains. These subjective symptoms are typically mild and may wax and wane in intensity. Patients who have them are referred to as having post–Lyme disease symptoms or syndrome, depending on the symptom duration and severity. The cause of these symptoms is currently unknown. Carefully done microbiologic evaluations in the United States have failed to find evidence of either persistent *B. burgdorferi* infection or a coinfection with a second *Ixodes*-transmitted pathogen. Furthermore, re-treatment has provided either no measurable benefit or a benefit so modest or ambiguous that it was outweighed by the risks associated with the antibiotic therapy.[A3,A4] Therefore, symptomatic treatment is recommended for such patients. In addition, a prospective study conducted in Europe that for the first time also incorporated a "healthy control" group has challenged the notion that such a syndrome even exists. In this study, the frequency of unexplained subjective symptoms in the Lyme disease group was no higher than that in the control group by 6 months into follow-up.[12]

Chronic Lyme Disease

The term *chronic Lyme disease* is poorly defined but widely used. In Europe, the term has been used to refer to the objective manifestations that most authorities prefer to call late Lyme disease. Others have used the term to refer to patients with post–Lyme disease subjective complaints. Most often, chronic Lyme disease is used as a diagnosis for patients with persistent pain, neurocognitive complaints, or fatigue, without objective clinical or serologic evidence of past or present *B. burgdorferi* infection. In this usage, the term is a misnomer and has become the latest in a series of postulated syndromes that attempt to attribute "medically unexplained symptoms" to particular infections.[13]

Coinfections

Ixodes ticks may be coinfected with and transmit Lyme borrelia along with other pathogens, such as *A. phagocytophilum*, *B. microti* (the primary cause of babesiosis), and a tick-borne encephalitis virus. The likelihood of coinfection is dependent on the particular species of *Ixodes* tick and on the geographic area. Thus, bites from *I. scapularis* ticks in certain areas may lead to the development of Lyme disease, human granulocytic anaplasmosis, or babesiosis as a single infection or less frequently as a coinfection. In addition, this tick species can potentially transmit the deer tick virus subtype of the Powassan virus, an *Ehrlichia* species referred to as the *Ehrlichia muris*-like agent, and *Borrelia miyamotoi*.[14-16] In Europe, the most common coinfection is Lyme disease with tick-borne encephalitis virus infection.

Coinfection should be considered in patients from geographic areas endemic for these pathogens who present with more severe initial symptoms than are commonly observed with Lyme disease alone. In this situation, coinfection should be considered especially in those who have high-grade fever for more than 48 hours despite antibiotic therapy appropriate for Lyme disease; those who develop recurrent fever; and those who have unexplained leukopenia, thrombocytopenia, or anemia. Coinfection might also be considered in the situation in which there has been resolution of the erythema migrans skin lesion but either no improvement or worsening of the viral infection–like symptoms.

Reinfection

Patients treated for early Lyme disease do not appear to develop an immunologic response that is adequate to protect against reinfection with a different strain of *B. burgdorferi*.[17] Therefore patients with erythema migrans may become reinfected at a different skin site if they get a bite from another infected tick.[18] Reinfection has been well documented only in patients who were treated for early infection (nearly always erythema migrans) and not after late manifestations of Lyme disease, such as Lyme arthritis. Clinical manifestations of reinfection appear to be similar to those of the primary infection.

PREVENTION

Lyme disease can be prevented by avoiding tick-infested environments and by covering bare skin and using tick repellents on skin and clothing when in such environments. The tick density around individual residences can be reduced by removing leaf litter, placing wood chips where lawns abut forests, applying acaricides to property, and constructing fences to keep out deer. Bathing within 2 hours of tick exposure has been shown to decrease the risk of Lyme disease. Daily inspections of the entire skin surface (including scalp) to remove attached ticks is recommended because of the grace period between the time of tick attachment and transmission of *B. burgdorferi*. Removal is accomplished by grasping the tick as close to its mouth parts as possible with a forceps (or tweezers) and then gently pulling it out. Clinical studies have demonstrated that without any other intervention, more than 96% of patients who find and remove an attached *I. scapularis* tick will remain free of Lyme disease, even in highly endemic geographic regions. If the tick is not found or removed, the probability of infection approaches the infection rate in the regional tick population (typically >20% of nymphal stage *I. scapularis* ticks are infected in highly endemic areas of the Northeast and Midwest United States).

Evidence shows that doxycycline chemoprophylaxis can further reduce the chance for development of Lyme disease after removal of an *I. scapularis* tick. A single 200-mg dose of doxycycline is about 90% effective in preventing erythema migrans at the tick bite site.[A4] Use of a single dose of doxycycline within 72 hours of tick removal should be considered for persons in highly endemic areas who are known to have been bitten by a nymphal or adult *I. scapularis* tick that was estimated to have been attached for at least 36 hours. Given the uncertain efficacy of a short course of amoxicillin in this situation, observation rather than chemoprophylaxis has been recommended for individuals for whom doxycycline is contraindicated. No vaccine is currently available to prevent Lyme disease in humans.

Grade A References

A1. Wormser GP, Ramanathan R, Nowakowski J, et al. Duration of antibiotic therapy for early Lyme disease. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med*. 2003;138:697-704.

A2. Ljostad U, Skogvoll E, Eikeland R, et al. Oral doxycycline versus intravenous ceftriaxone for European Lyme neuroborreliosis: a multicentre, non-inferiority, double-blind, randomised trial. *Lancet Neurol*. 2008;7:690-695.

A3. Klempner MS, Hu LT, Evans J, et al. Two controlled trials of antibiotic treatment in patients with persistent symptoms and a history of Lyme disease. *N Engl J Med*. 2001;345:85-92.

A4. Warshafsky S, Lee DH, Francois LK, et al. Efficacy of antibiotic prophylaxis for the prevention of Lyme disease: an updated systematic review and meta-analysis. *J Antimicrob Chemother*. 2010;65:1137-1144.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Stanek G, Wormser GP, Gray J, et al. Lyme borreliosis. *Lancet*. 2012;379:461-473.
2. Shapiro ED. Clinical practice. Lyme disease. *N Engl J Med*. 2014;370:1724-1731.
3. Radolf JD, Caimano MJ, Stevenson B, et al. Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes. *Nat Rev Microbiol*. 2012;10:87-99.
4. Warshafsky S, Lee DH, Francois LK, et al. Efficacy of antibiotic prophylaxis for the prevention of Lyme disease: an updated systematic review and meta-analysis. *J Antimicrob Chemother*. 2010;65:1137-1144.
5. Kenedy MR, Lenhart TR, Akins DR. The role of *Borrelia burgdorferi* outer surface proteins. *FEMS Immunol Med Microbiol*. 2012;66:1-19.
6. Brisson D, Baxamusa N, Schwartz I, et al. Biodiversity of *Borrelia burgdorferi* strains in tissues of Lyme disease patients. *PLoS ONE*. 2011;6:e22926.
7. Halperin JJ. Neurologic manifestations of Lyme disease. *Curr Infect Dis Rep*. 2011;3:360-366.
8. Bockenstedt LK, Wormser GP. Review: unraveling Lyme disease. *Arthritis Rheumatol*. 2014;66:2313-2323.
9. Strle F, Wormser GP, Mead P, et al. Gender disparity between cutaneous and non-cutaneous manifestations of Lyme borreliosis. *PLoS ONE*. 2013;8:e64110.
10. Wormser GP, Aguero-Rosenfeld ME, Cox ME, et al. Differences and similarities between culture-confirmed human granulocytic anaplasmosis and early Lyme disease. *J Clin Microbiol*. 2013;51:954-958.
11. Seriburi V, Ndukwe N, Chang Z, et al. High frequency of false positive IgM immunoblots for *Borrelia burgdorferi* in clinical practice. *Clin Microbiol Infect*. 2012;18:1236-1240.
12. Cerar D, Cerar T, Ruzic-Sabljic E, et al. Subjective symptoms after treatment of early Lyme disease. *Am J Med*. 2010;123:79-86.
13. Lantos PM. Chronic Lyme disease: the controversies and the science. *Expert Rev Anti Infect Ther*. 2011;9:787-797.
14. El Khoury MY, Hull RC, Bryant PW, et al. Diagnosis of acute deer tick virus encephalitis. *Clin Infect Dis*. 2013;56:e40-e47.
15. Pritt BS, Sloan LM, Johnson DK, et al. Emergence of a new pathogenic *Ehrlichia* species, Wisconsin and Minnesota, 2009. *N Engl J Med*. 2011;365:422-429.
16. Gugliotta JL, Goethert HK, Berardi VP, et al. Meningoencephalitis from *Borrelia miyamotoi* in an immunocompromised patient. *N Engl J Med*. 2013;368:240-245.
17. Khatchikian CE, Nadelman RB, Nowakowski J, et al. Evidence for strain-specific immunity in patients treated for early Lyme disease. *Infect Immun*. 2014;82:1408-1413.
18. Nadelman RB, Hanincová K, Mukherjee P, et al. Differentiation of reinfection from relapse in recurrent Lyme disease. *N Engl J Med*. 2012;367:1883-1890.

REVIEW QUESTIONS

1. Aside from the egg, there are three feeding stages of the deer tick, *Ixodes scapularis*: the larval, nymphal, and adult stages. Adult deer ticks have higher infection rates with *Borrelia burgdorferi* than the other stages do. Which stage is the most important epidemiologically for transmission of *B. burgdorferi* to humans?

A. Egg
B. Larval
C. Nymphal
D. Adult
E. Egg and larval stages

Answer: C The nymphal stage is the most important epidemiologically for transmission of infection to humans. The larval stage, which is the first stage, is uninfected and cannot transmit the infection. The adult stage, which is the third stage, is infected with *B. burgdorferi*, but it is less important in transmission to humans because this stage is present in smaller numbers in the environment. (See Epidemiology section.)

2. A single 200-mg dose of doxycycline should be considered to prevent Lyme disease under which circumstance?

A. Documented *I. scapularis* tick bite in which the tick had been attached for 36 hours or more and doxycycline can be given within 72 hours after tick removal
B. Clinical suspicion of a tick bite based on the presence of a scab on the leg
C. Any documented *I. scapularis* tick bite regardless of duration of attachment
D. Any documented tick bite
E. In patients with suspected erythema migrans

Answer: A Doxycycline is particularly effective for prevention of Lyme disease when there has been a documented *I. scapularis* tick bite in an area highly endemic for Lyme disease in which the tick had been attached for 36 hours or more and doxycycline can be given within 72 hours after tick removal. (See Epidemiology and Prevention sections.)

3. Which of the following would be least consistent with early or late neurologic Lyme disease?

A. Cerebrospinal fluid (CSF) with 150 white cells, 95% lymphocytes
B. Peripheral unilateral seventh nerve palsy with an 8 × 10-cm erythematous skin lesion on the right thigh
C. Bilateral peripheral seventh nerve palsies
D. CSF with 1000 white cells, 85% polymorphonuclear leukocytes
E. Six months of chronic headache with CSF showing 35 white cells, 95% mononuclear cells; test result for intrathecal production of antibody to *B. burgdorferi* is positive, and IgG antibody to *B. burgdorferi* is present in serum

Answer: D Neither early nor late neurologic Lyme disease is associated with a neutrophilic pleocytosis in the CSF. (See Clinical Manifestations section.)

4. Serologic testing to support a diagnosis of Lyme disease is usually not recommended in patients with possible

A. Erythema migrans
B. Lyme arthritis
C. Early neurologic Lyme disease
D. Lyme carditis
E. Late neurologic Lyme disease

Answer: A Because of the short duration of infection at the erythema migrans stage, serologic assays for antibodies to Lyme borrelia are rarely positive when this finding is noted. (See Diagnosis section.)

5. Which of the following antimicrobials is ineffective for Lyme disease?

A. Doxycycline
B. Ceftriaxone (third-generation cephalosporin)
C. Cefuroxime (second-generation cephalosporin)
D. Amoxicillin
E. Cephalexin (first-generation cephalosporin)

Answer: E *B. burgdorferi* is resistant to first-generation cephalosporins as well as to certain fluoroquinolones and rifampin. (See Treatment section.)

322 RELAPSING FEVER AND OTHER *BORRELIA* INFECTIONS

WILLIAM A. PETRI, JR.

DEFINITION

Relapsing fever is a spirochetal infection with bacteria of the genus *Borrelia*. There are two modes of transmission: epidemic louse-borne and endemic tick-borne relapsing fever. Disease is characterized by recurrent bouts of fever and spirochetemia separated by short fever-free periods.

The Pathogen

Members of the genus *Borrelia* are motile spirochetes that measure 0.5 µm in diameter and 5 to 40 µm in length. They are aerophilic and require long-chain fatty acids for growth. Louse-borne relapsing fever is caused by *Borrelia recurrentis*. Tick-borne relapsing fever organisms are named after their tick vector and include the closely related species *Borrelia duttonii* (Old World); *Borrelia hermsii*, *Borrelia turicatae*, and *Borrelia parkeri* (North America); and *Borrelia miyamotoi* (Old and New World).[1]

EPIDEMIOLOGY

Louse-borne epidemic relapsing fever is caused by *B. recurrentis* and is carried from person to person by the human body louse (*Pediculus humanus*). There is no animal reservoir. The spirochete lives in the louse hemolymph; infection is transmitted to humans when the louse is crushed on human skin and infective spirochetes penetrate the skin or mucous membranes. Epidemics have occurred during famines and at wartime when breakdown in sanitation favors the transmission of body lice. Louse-borne disease remains endemic in central and east Africa (Ethiopia, Somalia, Chad, and the Sudan) and in the South American Andes (Bolivia and Peru).

Tick-borne endemic relapsing fever occurs throughout the world and is transmitted to humans by *Ornithodoros* soft ticks. The ticks become infected by feeding on wild rodents (including mice, rats, squirrels, and chipmunks), which serve as natural reservoirs for the organisms. In the United States, relapsing fever is limited to humid mountainous areas of the West at altitudes of 1500 to 8000 feet, where the tick vector *Ornithodoros hermsii* resides in forests of ponderosa pine and Douglas fir trees. A key diagnostic clue has been a history of sleeping in rodent-infested rustic cabins in western U.S. national parks.[2] In Tanzania, where house infestation with *Ornithodoros* tick vectors can be very high, relapsing fever was identified in 11% of children seen at a clinic with fever. In the northeastern U.S., the prevalence of antibodies to *Borrelia miyamotoi* is nearly 50%, as high as for *Borrelia burgdorferi* (Chapter 321).[3]

PATHOBIOLOGY

Borrelia infection begins in the skin at the site of the louse or tick bite and is followed by rapid dissemination of the spirochetes through the blood stream. Spirochetes are visible on Wright-stained peripheral blood smears during the initial febrile episode and during each febrile relapse in most patients. The spirochete burden in blood positively correlates with symptom severity. Clearance of spirochetes from blood is associated with the production of serotype-specific immune sera; anti-*Borrelia* antibodies have been shown in animal models to be the major mechanism of immune clearance of infection.

Relapses are associated with cyclic antigenic variation in the variable major proteins (VMPs), which are the abundant outer membrane proteins of the spirochete that carry the serotype-specific epitopes. Antigenic variation is the consequence of recombination events that occur between VMP genes at silent and expression sites on linear plasmids. A single *B. hermsii* bacterium may produce as many as 40 distinct serotypes. Because spirochetes undergo one or several antigenic phases during infection, no specific or standard procedure has been developed for routine serodiagnosis of relapsing fever.

CLINICAL MANIFESTATIONS

An abrupt onset of fever (temperature > 39° C in most patients), headache, myalgia, and shaking chills characterizes the onset of illness. Cough, nausea and vomiting, and fatigue are less frequent complaints. Signs include fever, tachycardia, lethargy or confusion, conjunctival injection, and epistaxis. Hepatosplenomegaly, jaundice, and often a truncal petechial rash are common signs in louse-borne relapsing fever. Neurologic findings may occur,[4] including meningitis, meningoencephalitis, and facial palsy, although these entities are more common with tick-borne relapsing fever. Untreated louse-borne disease lasts 6 days, and relapses occur once after an afebrile period of 9 days. The initial illness of tick-borne relapsing fever lasts about 4 days without antibiotic treatment, with an average of two relapses (each after an average 10-day afebrile period) before the diagnosis is made.

Relapsing fever in pregnancy can cause placental damage and intrauterine growth retardation and results in miscarriage in a third of patients. Neonatal infection by both the tick- and louse-borne forms is accompanied by jaundice, hepatosplenomegaly, and often sepsis and hemorrhage. Fever and hepatosplenomegaly are also common signs in children. It has been also recognized in immunocompromised patients.[5]

DIAGNOSIS

The diagnosis should be considered in patients with fever who are returning from a stay in cabins in the mountainous and high-elevation areas of the western United States. Only a few patients will remember tick exposure, because *O. hermsii* is a night feeder, has a painless bite, and remains attached for only 15 minutes. Internationally, relapsing fever can occur sporadically wherever dwellings are infected with *Ornithodoros* ticks, as well as in epidemics with louse-borne disease.

Because the number of organisms in blood is extremely high, the diagnosis is most often made by direct visualization of the organism in a blood smear (Fig. 322-1), although the diagnosis can also be made with polymerase chain reaction and serodiagnostic tests. Spirochetes can be demonstrated in peripheral blood smears taken during febrile episodes in 70% of patients. Additional sensitivity may be gained by examination of a buffy coat preparation of peripheral blood. Because of their characteristic locomotion, spirochetes can be readily detected by direct visualization of thick blood films under low-power microscopy. Culture of the organism requires a special medium and is not practical in a clinical laboratory setting. The white blood cell count is generally normal, but platelet counts of less than 50,000/mm^3 occur in up to 90% of cases of louse-borne disease. Prothrombin and activated partial thromboplastin times are often prolonged. In louse-borne disease, elevations in liver function test results (serum transaminases and bilirubin) and blood urea nitrogen are common. Urinalysis may reveal proteinuria and microscopic hematuria. Examination of cerebrospinal fluid may show a lymphocytic pleocytosis, and spirochetes may be directly visualized.

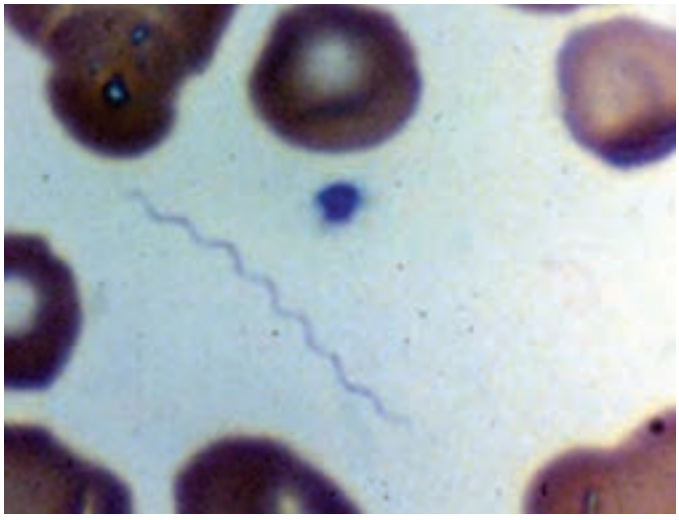

FIGURE 322-1. A single spirochete is seen in a Wright-stained thin blood smear from a patient with relapsing fever.

TREATMENT

Borrelia is generally quite sensitive to antibiotics, which has led to recommendations for single-dose treatments. Although this may be sufficient, especially for louse-borne disease, recent reports suggest that silent residual infections occur and may best be addressed by longer treatments. For tick-borne relapsing infection, treatment should extend for 7 days to reduce the risk for persistent infection. Tetracycline, doxycycline, and erythromycin are all effective antibiotics. Erythromycin should be used in pregnant women and children younger than 7 years (in whom tetracyclines can stain the permanent teeth). Penicillin treatment has been reported to clear the spirochetemia more slowly than tetracycline does.

The Jarisch-Herxheimer reaction (typically characterized by a rise in body temperature of 1°C, rigors, a rise in blood pressure followed by a fall, and transient leukopenia) occurs 2 to 3 hours after treatment in many patients with louse-borne disease, less commonly in tick-borne disease, and should be anticipated and managed supportively. Death as a result of shock from the Jarisch-Herxheimer reaction occurs rarely. The Jarisch-Herxheimer reaction has been associated with accelerated phagocytosis of spirochetes by neutrophils and transient elevations in tumor necrosis factor (TNF)-α, interleukin (IL)-6, IL-8, and IL-10. In small numbers of patients with louse-borne relapsing fever, anti–TNF-α antibodies have been effective in prevention.

PREVENTION

Prevention of louse-borne relapsing fever hinges on improving hygienic conditions, delousing affected areas, and antibiotic treatment of patients and close contacts. Tick-borne relapsing fever can be prevented by reducing the risk of contact with rodents and ticks, including repair of structural flaws in cabins and other residences so rodents cannot nest in or around them, as well as spraying infested indoor environments. Tick-bite screening and prophylactic treatment with doxycycline in highly endemic areas has been reported to be a practical, safe, and effective policy in preventing tick-borne relapsing fever.[6]

PROGNOSIS

Epidemics of louse-borne relapsing fever have been reported, with mortality rates approaching 40%; as much as 5% of the mortality is related to Jarisch-Herxheimer reactions with treatment. Mortality from tick-borne disease is less than 5%. Autopsies of patients with louse-borne disease have documented intracranial hemorrhage, brain edema, bronchopneumonia, hepatic necrosis, and splenic infarcts.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Crowder CD, Carolan HE, Rounds MA, et al. Prevalence of *Borrelia miyamotoi* in *Ixodes* ticks in Europe and the United States. *Emerg Infect Dis*. 2014;20:1678-1682.
2. Fritz CL, Payne JR, Schwan TG. Serologic evidence for *Borrelia hermsii* infection in rodents on federally owned recreational areas in California. *Vector Borne Zoonotic Dis*. 2013;13:376-381.
3. Krause PJ, Narasimhan S, Wormser GP, et al. *Borrelia miyamotoi* sensu lato seroreactivity and seroprevalence in the northeastern United States. *Emerg Infect Dis*. 2014;20:1183-1190.
4. Rajaguru S, Havlichek D, Khalife W, et al. A 20-year-old man with fever, headache, and neck stiffness: neuroborreliosis with tick-borne relapsing fever. *Clin Infect Dis*. 2011;52:271-272.
5. Gugliotta JL, Goethert HK, Berardi VP, et al. Meningoencephalitis from *Borrelia miyamotoi* in an immunocompromised patient. *N Engl J Med*. 2013;368:240-245.
6. Balicer RD, Mimouni D, Bar-Zeev Y, et al. Postexposure prophylaxis of tick-borne relapsing fever. *Eur J Clin Microbiol Infect Dis*. 2010;29:253-258.

REVIEW QUESTIONS

1. The differential diagnosis for a patient presenting with an undifferentiated febrile illness, chills, headache, and recent tick exposure include ***all but one*** of the following:

A. Lyme disease
B. Rocky Mountain spotted fever (RMSF)
C. Ehrlichiosis
D. Tick paralysis
E. Anaplasmosis

Answer: D Not tick paralysis, which is not an infection but in fact an intoxication due to a tick neurotoxin. An undifferentiated febrile illness with recent tick exposure could include Lyme disease, RMSF, ehrlichiosis, and anaplasmosis.

2. Causes of a peripheral facial nerve palsy include ***all but one*** of the following:

A. Lyme disease
B. Herpes simplex virus (HSV)
C. Sarcoidosis
D. Babesiosis
E. Relapsing fever

Answer: D Not babesiosis. HSV is the most likely cause of Bell's palsy, with the borrelial infections, Lyme disease, and relapsing fever not uncommon, and sarcoidosis rare.

3. Diagnosis of relapsing fever can include ***all but one*** of the following:

A. Serology (acute and convalescent)
B. Positive blood smear for a spirochete
C. Polymerase chain reaction (PCR) assay of blood
D. Sleeping in rustic cabins in the U.S. West
E. Blood cultures

Answer: E Blood cultures cannot culture the *Borrelia* bacteria that cause relapsing fever; however, the number of organisms per milliliter of blood is so high they can be seen on a blood smear and detected by PCR.

4. Treatment of relapsing fever is:

A. Tetracycline
B. Ceftriaxone
C. Complicated by Jarisch-Herxheimer reaction
D. A and B
E. A, B, C

Answer: E Tetracycline and ceftriaxone are equally effective, and either can cause the Jarisch-Herxheimer reaction.

5. Complications of relapsing fever include:

A. Jarisch-Herxheimer reaction
B. Meningitis
C. Jaundice
D. A and C
E. All of the above

Answer: E Relapsing fever can be complicated by meningitis, jaundice (especially the louse-borne form), and the Jarisch-Herxheimer reaction, which is treatment associated and probably due to accelerated phagocytosis of spirochetes by neutrophils and the consequent elevation in levels of tumor necrosis factor (TNF)-α and interleukin (IL)-6, IL-8, and IL-10.

323

LEPTOSPIROSIS

ATIS MUEHLENBACHS AND SHERIF R. ZAKI

DEFINITION

Leptospirosis is a zoonotic disease caused by pathogenic *Leptospira* species spirochetes. Leptospirosis is distributed worldwide and is most prevalent in tropical developing countries. Leptospires frequently infect wild and domestic mammals. Humans are infected directly by contact with infected animals or indirectly through contact with urine-contaminated soil or water. Disease severity ranges from mild and self-limiting to severe with life-threatening manifestations including massive pulmonary hemorrhage and Weil's disease (the triad of jaundice, acute renal failure, and bleeding).

The Pathogen

Leptospires are thin, coiled, highly motile spirochetes measuring 6 to 20 microns in length. They are obligate aerobes that can survive for several weeks in the environment. *Leptospira* is currently genetically classified into nine pathogenic species (*L. interrogans, L. kirschneri, L. borgpetersenii, L. santarosai, L. noguchii, L. weilii, L. alexanderi, L. alstoni, L. kmetyi*) and those of intermediate pathogenesis (*L. inadai, L. broomii, L. fainei, L. wolffii, L. licerasiae*). Pathogenic leptospires are further classified into over 25 serogroups and 250 serovars that differ by geographic range and host specificity, which is useful information for outbreak and other epidemiologic investigations.

EPIDEMIOLOGY

Over 350,000 cases of leptospirosis are estimated to occur each year and are generally underreported. In the United States, around 100 to 200 leptospirosis cases are identified each year, with most occurring in Hawaii and Puerto Rico; the incidence is likely higher.[1] The majority of infections are mild and self-limiting, but case fatality in reported cases may be as high as 10%. In endemic areas, up to 20 to 30% of acute undifferentiated fever cases may be due to leptospirosis,[2] and seroprevalence can range from 5 to 15%. The major groups at risk are slum dwellers, subsistence farmers, and animal workers, owing to exposure to rodent, domestic, and wild animal reservoirs. In both tropical and temperate climates, the urban poor are an underrecognized population at risk.[3] Fresh water exposure is a risk factor, particularly during heavy rainfall and natural disasters, in addition to exposure during travel, extreme outdoor sports activities, and military operations. Humans are considered accidental hosts; rare human-to-human transmission by transplacental infection and breast-feeding has been reported.

PATHOBIOLOGY

Leptospires can directly penetrate abraded skin and mucous membranes and spread hematogenously to target organs. The classic illness is biphasic, with the first phase characterized by leptospiremia and the second phase with organism clearance by agglutinating antibodies and an associated host response that can be immunopathogenic.[4,5] Leptospires can persist for a period of time in target organs. In asymptomatic reservoir animals, leptospires can reach massive densities within the renal tubules, resulting in continued urinary excretion.

Pathologic findings (Fig. 323-1) may include pulmonary hemorrhage, diffuse alveolar damage, mild to marked hepatocellular dissociation, mild portal hepatitis, lymphohistiocytic interstitial nephritis, renal tubular necrosis, and mild renal glomerular mesangial hyperplasia. Hemorrhage in other organs, multifocal myocarditis, myositis, and hemophagocytosis may also be present. By immunohistochemistry or silver stains (Fig. 323-2), leptospires can be seen within the renal interstitium, hepatic parenchyma, and within the walls of small, medium, and large pulmonary blood vessels.

Leptospire tissue penetration may be mediated by a burrowing motion and secreted enzymes including collagenase and sphingomyelinase. Leptospiral proteins interact directly with host extracellular matrix components such as collagen, fibronectin, and laminins. Leptospires are resistant to the alternative pathway of complement-mediated lysis and can bind inhibitory complement factor H. Leptospiral lipopolysaccharides and lipopeptides have low endotoxic potency but can activate innate immune response through toll-like receptor (TLR)-2 signaling and are thought to generate a cytokine response. Immunity is considered to be primarily humoral and serotype specific. Circulating immune complexes may contribute to renal damage and endothelial dysfunction. An expansion of γδ T cells occurs during infection. Leptospires may also directly activate plasminogen to plasmin, the main enzyme of the fibrinolytic system, which could promote hemorrhage.[6]

Genomic studies of pathogenic intermediate and saprophytic *Leptospira* species have revealed a relatively large genome that contains genes involved in environmental survival, chemotaxis, and motility that may be involved in pathogenesis.[7] Little is known about host genetic risk factors, although the HLA-DQ6 allele has been associated with increased susceptibility to infection.

CLINICAL MANIFESTATIONS

The incubation period is typically 7 to 12 days (range, 2 to 30 days). During the early phase of illness (first 3 to 7 days), the majority of patients present with high fever (38° to 40° C) and myalgia. Cough, nausea and vomiting, diarrhea, headache, photophobia, and rash may be seen. Conjunctival suffusion is a characteristic finding (Fig. 323-3), but it is only seen in a third of patients near the end of early-phase illness. Myalgia may be pronounced and most frequently involves the calves and lumbar musculature, with creatine phosphokinase elevation. Severe cervical and abdominal myalgia may mimic nuchal rigidity or an acute abdomen, respectively. Rash occurs in 10 to 20% of patients and may be urticarial, maculopapular, or purpuric in a typically

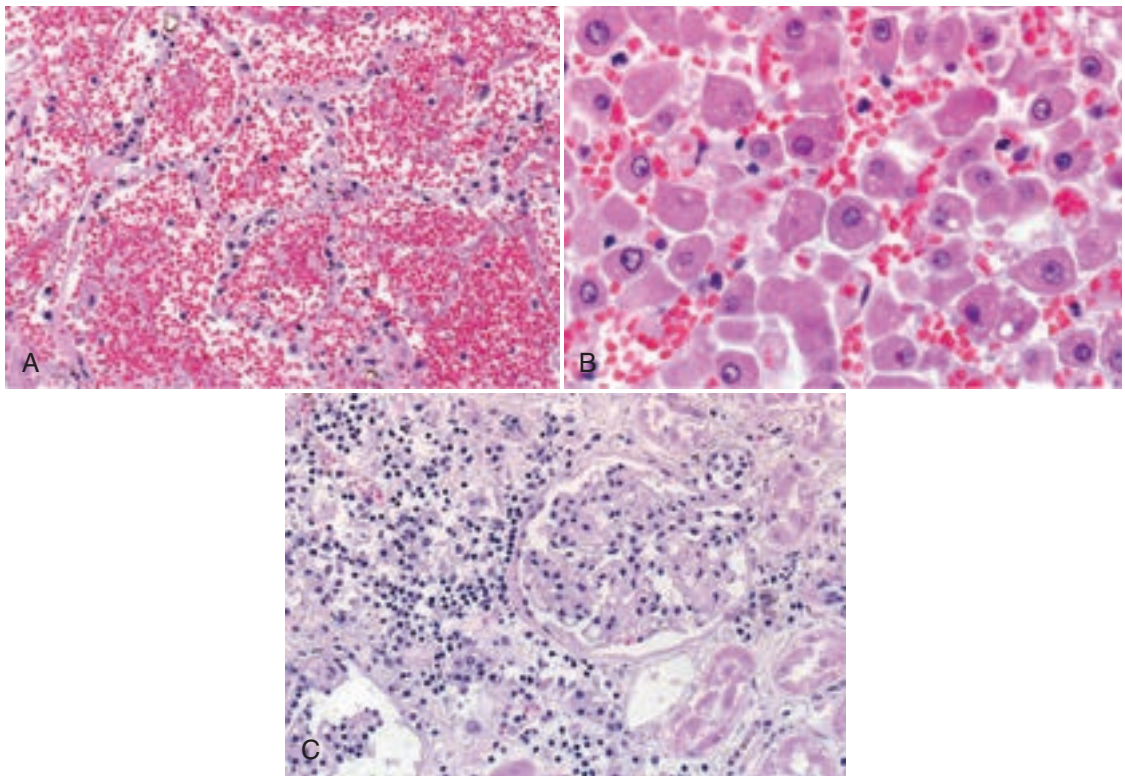

FIGURE 323-1. Pathologic features of severe leptospirosis. A, Pulmonary hemorrhage. **B,** Hepatocellular dissociation. **C,** Interstitial nephritis.

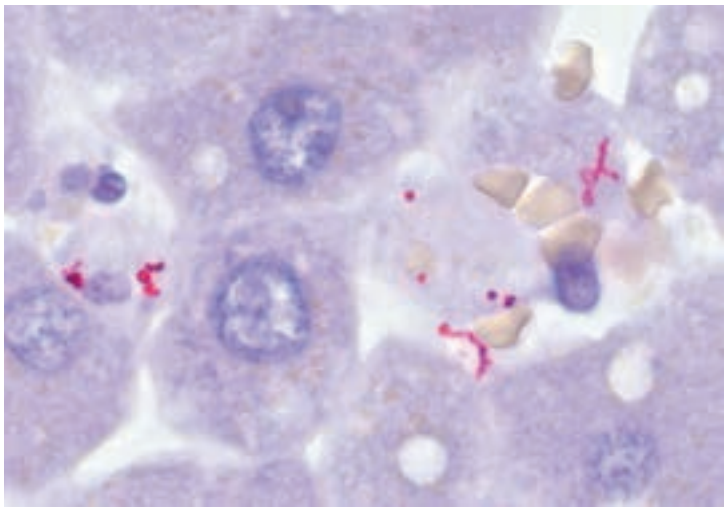

FIGURE 323-2. Immunostaining of *Leptospira* in liver. Note the coiled nature of the spirochetes. Granular antigen staining is also seen.

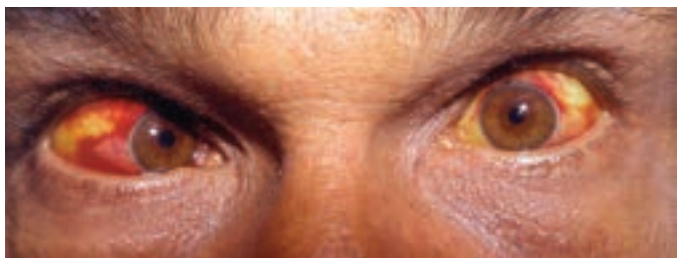

FIGURE 323-3. Conjunctival suffusion in a patient with severe leptospirosis. Conjunctival suffusion is a characteristic sign of leptospirosis and best seen along the right upper palpebral border above the subconjunctival hemorrhage. (Courtesy Antonio Seguro and Paulo Marotto, Hospital Emílio Ribas and Universidade de São Paulo.)

pretibial and truncal distribution. Hepatosplenomegaly and lymphadenopathy may also be present. Resolution of symptoms coincides with the presence of agglutinating immunoglobulin (Ig)M antibodies and reduction in leptospiremia. As a classic biphasic disease, fever may recur 3 to 4 days after remission. In this later immune phase, severe headache is often present and can be associated with photophobia, meningeal signs, and cerebrospinal fluid (CSF) pleocytosis.

Severe disease may occur progressively at initial presentation or during late-phase leptospirosis and result in 10 to 50% mortality. Life-threatening manifestations are renal failure, hypotension, hemorrhage, and respiratory failure. Jaundice occurs in 5 to 10% of patients, and serum bilirubin levels can be elevated up to 40 to 80 mg/dL, with only moderate and minor elevations in transaminase and alkaline phosphatase levels, respectively. Leptospirosis may present as acute cholecystitis. Long-term hepatic sequelae are typically not seen. Renal findings are typically of nonoliguric hypokalemic renal insufficiency and impaired tubular sodium reabsorption. Volume loss may result in oliguric renal insufficiency and acute tubular necrosis, and renal failure occurs in about half of severe cases. Common urinalysis findings are proteinuria, white blood cells, hematuria, and hyaline and granular casts.

Pulmonary findings may include cough, dyspnea, and hemoptysis. Leptospirosis-associated pulmonary hemorrhage and acute respiratory distress syndrome is now recognized as a common clinical presentation. Radiographic findings may show a patchy alveolar infiltrate to large areas of consolidation due to hemorrhage.[8] Cardiac conduction abnormalities can be seen and tend to be nonspecific in mild disease. First-degree atrioventricular block and features of pericarditis are the most common findings in severe disease. Arrhythmias including ventricular fibrillation may also occur.

Thrombocytopenia is frequent and does not appear to be due to platelet consumption, but disseminated intravascular coagulopathy can occur. Prothrombin and partial thromboplastin times are typically normal or only mildly elevated. Petechiae, conjunctival hemorrhage, and purpura can be seen in addition to the more severe hemorrhagic manifestations.

Aseptic meningitis is the most frequent neurologic manifestation. CSF findings include pleocytosis with neutrophil predominance in early disease, followed by lymphocyte predominance in later disease. Glucose is generally normal, and CSF pressure may be elevated. Less common are intracerebral hemorrhage, encephalitis, myelitis, and peripheral neuropathy. Ocular leptospirosis, including chronic uveitis, is thought to be primarily caused by an immunopathogenic mechanism and occurs late in the disease process.

DIAGNOSIS

Few clinical signs differentiate early stage leptospirosis, and severe leptospirosis may be recognized in the form of Weil's disease. Reference tests require specialized laboratories, and there is a need for point-of-care testing in tropical areas. Culture and serologic assays may provide false negatives.[9-11]

Culture of the organism requires specialized media (Ellinghausen-McCullough-Johnson-Harris (EMJH) or Fletcher media, with often more than 4 weeks needed with observation under darkfield microscopy. Cultures generally have low sensitivity, with the best yield obtained from peripheral blood during days 1 through 4 of the acute illness; urine may be positive for up to day 10. Direct darkfield microscopy of clinical specimens can also be attempted but has low sensitivity and specificity.

The gold standard serologic test is the microagglutination test (MAT), available at the Centers for Disease Control and Prevention. MAT is considered positive if there is a four-fold rise in titer between acute and convalescent sera. This assay requires cultured *Leptospira*; it provides serotype data and has a very high specificity but lower sensitivity. U.S. Food and Drug Administration–approved IgM enzyme-linked immunosorbent assay (ELISA) and indirect hemagglutination assays are also available. Multiple lateral flow rapid diagnostic "dipstick" tests are manufactured and used in other countries; however, their performance may not be well characterized. Diagnostic polymerase chain reaction (PCR) assays are being investigated and hold promise for diagnostic accuracy.[12,13] Organisms within tissues can be detected by Warthin-Starry silver stain or immunohistochemistry. In the United States, leptospirosis is a nationally notifiable disease.

DIFFERENTIAL DIAGNOSIS

As previously discussed, few clinical or laboratory findings differentiate leptospirosis from other causes of acute fever. Differential diagnosis depends on other diseases in the geographic area, most frequently malaria, dengue, typhoid fever, and rickettsial diseases including scrub typhus. Other diseases in the differential include influenza, acute viral hepatitis, yellow fever, bacterial and viral meningitis, bacterial sepsis, and hantavirus infection. A high index of suspicion is needed in endemic areas. Useful diagnostic clues are conjunctival suffusion, muscle tenderness, and pulmonary bleeding. High serum bilirubin in the presence of relatively mild transaminitis would argue against viral hepatitis and favor leptospirosis. Hypokalemia, elevated creatinine, elevated creatine phosphokinase, and thrombocytopenia are nonspecific but may also suggest leptospirosis.

TREATMENT

Rx

The World Health Organization guidelines and widespread clinical practice are to treat patients early with antibiotics for leptospirosis. Penicillin, doxycycline, or a cephalosporin appear to be equally efficacious, and doxycycline has the advantage of also treating rickettsial infections. Of note, a recent meta-analysis of seven randomized trials over the past 30 years concluded there is insufficient "grade A" evidence to recommend for or against the use of antibiotics to treat leptospirosis[A1]; antibiotic therapy may decrease the duration of clinical illness by 2 to 4 days; however, in severely ill patients, penicillin may be associated with a higher rate of dialysis. Additional randomized controlled trials are needed.

Treatment regimens for mild leptospirosis include doxycycline (100 mg PO bid), ampicillin (500 mg PO q6h), amoxicillin (500 mg orally [PO] q8h), or azithromycin (1 g followed by 500 mg daily for 2 days); and for severe disease, intravenous (IV) penicillin (1.5 million units q6h), ceftriaxone (1 g daily), cefotaxime (1 g q6h), or doxycycline (100 mg IV q12h). Treatment duration is typically 7 days. The Jarisch-Herxheimer reaction, a febrile inflammatory reaction that occurs with initiation of treatment and results from clearance of the organism from the circulation (Chapter 319), can occur shortly after antimicrobial therapy is started,[14] so patients require monitoring at initiation of antibiotics.

Prompt triage of high-risk patients and aggressive supportive care are essential. Hypotension should be treated, and volume repletion is useful in limiting renal damage. Patients with nonoliguric hypokalemic renal insufficiency may be treated by volume and potassium repletion. Prompt dialysis is indicated for oliguric renal insufficiency, either by continuous hemofiltration or by peritoneal dialysis. Serial electrocardiograms are helpful to monitor for arrhythmia. Aggressive therapy may be needed to treat hemorrhage and respiratory failure.

PREVENTION

Preventive measures include sanitation to prevent population exposure to contaminated water, and limiting water contamination by animal reservoirs such as dogs, pigs, and cattle. Vaccination is available for domestic and livestock animals. Rodent control is important.[15] Workers with occupational exposure to animals or contaminated water or soil are encouraged to use protective equipment such as gloves and boots. Travelers to endemic areas should be counseled on fresh water exposure. A vaccine is not available for human use within the United States. Vaccine trials have been performed in Cuba, Russia and China, however vaccine safety and efficacy are uncertain.

Prophylaxis with doxycycline (200 mg PO weekly) is widely used for persons with exposure to contaminated water or at high risk for leptospirosis. Prophylaxis may work better in reducing risk of clinical disease for short-term travelers with high risk, rather than residents in an endemic area. Of interest, a recent meta-analysis concluded that there is insufficient grade A evidence to recommend for or against the use of antibiotic prophylaxis[A2]; additional randomized controlled trials are needed.

PROGNOSIS

The majority of infections are self-limiting, but there may be 10% case fatality among patients who seek medical attention. Severe leptospirosis is associated with a 10 to 50% risk of death. Death is more frequent during infection if there is renal failure, altered mental status, older age, oliguria, pulmonary hemorrhage, or respiratory insufficiency. Ocular disease, including chronic uveitis, may result in severe visual impairment.

Grade A References

A1. Brett-Major DM, Coldren R. Antibiotics for leptospirosis. *Cochrane Database Syst Rev*. 2012;2:CD008264.

A2. Brett-Major DM, Lipnick RJ. Antibiotic prophylaxis for leptospirosis. *Cochrane Database Syst Rev*. 2009;3:CD007342.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Centers for Disease Control and Prevention. Notes from the field: investigation of leptospirosis underreporting–Puerto Rico, 2010. *MMWR Morb Mortal Wkly Rep*. 2012;61:421.
2. Guerra MA. Leptospirosis: public health perspectives. *Biologicals*. 2013;41:295-297.
3. Dupouey J, Faucher B, Edouard S, et al. Human leptospirosis: an emerging risk in Europe? *Comp Immunol Microbiol Infect Dis*. 2014;37:77-83.
4. Evangelista KV, Coburn J. *Leptospira* as an emerging pathogen: a review of its biology, pathogenesis and host immune responses. *Future Microbiol*. 2010;5:1413-1425.
5. Ricaldi JN, Swancutt MA, Matthias MA. Current trends in translational research in leptospirosis. *Curr Opin Infect Dis*. 2013;26:399-403.
6. Oliveira R, Domingos RF, Siqueira GH, et al. Adhesins of *Leptospira interrogans* mediate the interaction to fibrinogen and inhibit fibrin clot formation in vitro. *PLoS Negl Trop Dis*. 2013;7:e2396.
7. Lehmann JS, Fouts DE, Haft DH, et al. Pathogenomic inference of virulence-associated genes in *Leptospira interrogans*. *PLoS Negl Trop Dis*. 2013;7:e2468.
8. Marchiori E, Lourenco S, Setubal S, et al. Clinical and imaging manifestations of hemorrhagic pulmonary leptospirosis: a state-of-the-art review. *Lung*. 2011;189:1-9.
9. Budihal SV, Perwez K. Leptospirosis diagnosis: competency of various laboratory tests. *J Clin Diagn Res*. 2014;8:199-202.
10. Limmathurotsakul D, Turner EL, Wuthiekanun V, et al. Fool's gold: why imperfect reference tests are undermining the evaluation of novel diagnostics: a reevaluation of 5 diagnostic tests for leptospirosis. *Clin Infect Dis*. 2012;55:322-331.
11. Schreier S, Doungchawee G, Chadsuthi S, et al. Leptospirosis: current situation and trends of specific laboratory tests. *Expert Rev Clin Immunol*. 2013;9:263-280.
12. Agampodi SB, Matthias MA, Moreno AC, et al. Utility of quantitative polymerase chain reaction in leptospirosis diagnosis: association of level of leptospiremia and clinical manifestations in Sri Lanka. *Clin Infect Dis*. 2012;54:1249-1255.
13. Waggoner JJ, Balassiano I, Abeynayake J, et al. Sensitive real-time PCR detection of pathogenic *Leptospira spp.* and a comparison of nucleic acid amplification methods for the diagnosis of leptospirosis. *PLoS ONE*. 2014;9:e112356.
14. Guerrier G, D'Ortenzio E. The Jarisch-Herxheimer reaction in leptospirosis: a systematic review. *PLoS ONE*. 2013;8:e59266.
15. Kamath R, Swain S, Pattershetty S, et al. Studying risk factors associated with human leptospirosis. *J Glob Infect Dis*. 2014;6:3-9.

REVIEW QUESTIONS

1. Which group is least at risk for leptospirosis?

A. Hospital workers
B. Veterinarians
C. Farmers
D. Sewage workers
E. Urban homeless

Answer: A Health care workers are at low risk for leptospirosis because human-to-human transmission is rare and primarily involves vertical transmission from mother to infant. Because *Leptospira* species are zoonoses that can persist in the environment, persons with exposure to domestic or wild animals, soil, or fresh water are at greatest risk. The urban poor of both temperate and tropical climates are an underappreciated population at risk.

2. Which is the least likely presentation of leptospirosis?

A. High fever and headache
B. Aseptic meningitis
C. Cough and hemoptysis
D. Localized seizure
E. Acute cholecystitis

Answer: D Localized seizure would be an atypical presentation of leptospirosis, which is generally not considered to be associated with focal neurologic findings. Hemoptysis and renal and hepatic failure are classic presentations of severe leptospirosis. Fever is very common, and aseptic meningitis is well described. Leptospirosis can be associated with acute abdominal pain, and acute cholecystitis has been described as a disease presentation.

3. What is the most likely renal finding in leptospirosis?

A. Cortical necrosis
B. Massive proteinuria
C. Hyperkalemic nonoliguric renal failure
D. Pyelonephritis
E. Allergic nephritis

Answer: C Hyperkalemic nonoliguric renal failure is the typical renal presentation of leptospirosis, which can progress to oliguric renal failure. Hypokalemia is an important laboratory value that can support the diagnosis of leptospirosis. Treatment involves volume resuscitation to prevent oliguric renal failure, electrolyte replacement, and prompt consideration of dialysis. Massive proteinuria would be atypical and suggests a different etiology. The pathologic findings are of interstitial nephritis; cortical necrosis, pyelonephritis, and eosinophilic infiltrates are typically not seen.

4. Of the following, what is the most likely clinical presentation and associated finding in leptospirosis?

A. Jaundice with markedly elevated transaminases
B. Myalgia with elevated creatine kinase
C. Petechiae and markedly prolonged prothrombin time
D. Fever and pancytopenia
E. Severe headache and papilledema

Answer: B Myalgia in leptospirosis is associated with elevated creatine kinase. Myalgia can be extremely painful and mimic an acute abdomen or the nuchal rigidity seen in meningitis. Jaundice is classically seen in leptospirosis, but transaminases are generally low despite markedly elevated bilirubin; these findings can be an important clue to support the diagnosis of leptospirosis. Petechiae and a bleeding diatheses, including massive pulmonary hemorrhage, can be present, but coagulation parameters are typically not markedly abnormal. Fever and headache are very common, but pancytopenia and papilledema would be unusual.

5. Of the following, which are the most likely pathologic findings in leptospirosis?

A. Massive hepatic necrosis and acute renal tubular injury
B. Pulmonary hemorrhage, interstitial nephritis, and hepatocellular dissociation
C. Neutrophilic bronchopneumonia and interstitial nephritis
D. Vasculitis with abundant intravascular spirochetes
E. Pulmonary and renal infarction

Answer: B Massive pulmonary hemorrhage, interstitial nephritis, and hepatocellular dissociation are the typical findings seen at autopsy. Massive hepatic necrosis is generally not seen and would suggest dengue or yellow fever virus infection (depending on the geographic location) or other causes of acute fulminant liver failure. Leptospiral organisms do not cause neutrophilic bronchopneumonia. Vasculitis is generally not a prominent feature of leptospirosis, although a degree of inflammation has been described within pulmonary vessel walls. *Leptospira* species spirochetes are not intravascular but associated with the interstitium; these organisms have several well-defined mechanisms to bind collagen and other components of the extracellular matrix. Infarcts are not a typical feature and would suggest a thromboembolic process.

324 TUBERCULOSIS

JERROLD J. ELLNER

DEFINITION

Tuberculosis (TB) is a chronic granulomatous disease with a unique latent stage usually caused by the acid-fast bacillus (AFB) *Mycobacterium tuberculosis* (Mtb). The most common site of disease is the lung; frequent extrapulmonary sites are the lymph nodes, pleura, bones, and joints. TB is spread from person to person by inhalation of infectious droplet nuclei aerosolized by patients with pulmonary TB (PTB). TB is a major cause of morbidity and mortality worldwide, with over 95% of cases and 99% of deaths occurring in resource-limited settings. The human immunodeficiency virus (HIV) pandemic led to a resurgence of TB and promoted explosive nosocomial outbreaks of multiple-drug resistant TB. The result was increased attention to TB as a global public health emergency and increased funding for TB control and research. The problems posed for TB control are compounded by increasing drug-resistant disease that is expensive to treat and may be refractory to available drugs.

The Pathogen

TB is caused by infection with one of the three members of the Mtb complex: Mtb, *Mycobacterium africanum*, or *Mycobacterium bovis*. The causative organism is a slender, non-motile, non–spore-forming, non–toxin-producing bacillus that may be beaded and is approximately 2 to 4 μm in length. It is a slow-growing (doubling time of 18 to 24 hours) facultative aerobe that can persist intracellularly for prolonged periods. The organism is identified in clinical specimens as an acid-fast bacillus (AFB). Mtb can be stained with carbol fuchsin by either alkalinization (Kinyoun) or heat (Ziehl-Neelsen) methods. The waxy coat of Mtb, composed of mycolic acid and other complex

TABLE 324-1 RISK FACTORS FOR TUBERCULOSIS

RISK FACTOR	INCREASED RISK OF RECENT INFECTION*	INCREASED RISK OF PROGRESSION FROM INFECTION TO DISEASE	TST CUT POINT
Household contact of PTB	X		>5 mm
Solid organ transplant recipients, immunosuppressive treatment (TNF inhibitors, prednisone > 15 mg/day for > 1 month), fibrotic lesions on chest radiograph consistent with prior TB		X	>5 mm
HIV infection	X	X	>5 mm
Foreign-born, injecting drug users, TST-positive children, adolescents, young adults	X		>10 mm
Residents or workers in hospitals, homeless shelters, correctional facilities, nursing homes, residences for the HIV-infected	X		>10 mm
Underweight (>15%), silicosis, diabetes mellitus (particularly insulin-dependent or poorly controlled), renal failure, hemodialysis, gastrectomy, jejuno-ileal bypass, carcinoma of the head and neck, lung cancer, lymphoma, leukemia		X	>10 mm
None			>15 mm

*Recent infection per se increases risk of progressing from infection to disease (12.9 cases per 1000 person-years in the first year compared to 1.6 per 1000 person-years in the subsequent 7 years).
HIV = human immunodeficiency virus; PTB = pulmonary TB; TNF = tumor necrosis factor; TST = tuberculin skin test.

lipopolysaccharides, precludes decolorization of the stain with a mixture of acid and alcohol.

DNA sequencing of Mtb and genetic manipulations have promoted basic understanding of the metabolism and virulence of the organism, its immunodominant antigens, and capacity to survive adverse conditions and persist intracellularly. Clinical isolates of Mtb differ in their virulence in experimental models, potential for transmission in humans, and interaction with the host (immunopathology, induction of host cytokines, delayed-type hypersensitivity [DTH]). For example, the hypervirulent Beijing strain family overexpresses a phenolic glycolipid that inhibits innate immunity and may thereby contribute to its pathogenicity.

There are six main phylogeographic lineages of Mtb, each associated with a specific human population. The families differ in geographic distribution and in some cases the potential for transmission and pathogenesis. Strain typing is particularly useful in outbreak investigations and can be performed by several techniques, restriction fragment length polymorphism (RFLP) of the insertional element IS 6110 or spoligotyping. The finding that multiple cases of TB are caused by the same strain and constitute a "cluster" suggests that they are epidemiologically linked. Whole-genomic sequencing has emerged as a more powerful tool to establish transmission even in the absence of epidemiologic links.

TB caused by *M. africanum* is clinically identical to that caused by Mtb. *M. bovis* has greater than 95% DNA homology with Mtb and causes disease in humans, cattle, deer, badgers, and other animals. The main route of transmission of Mtb is person to person through respiratory aerosols generated by coughing. Bacilli in small droplet nuclei (1 to 5 μm in diameter) remain suspended in air for long periods and once inhaled can reach the airways, where only 1 to 5 organisms are sufficient to cause infection. Laryngeal involvement renders the patient highly infectious. Direct cutaneous inoculation ("prosector's wart") does occur. *M. bovis* can be transmitted by the gastrointestinal route, usually through ingestion of contaminated milk.

EPIDEMIOLOGY

In 2012, there were an estimated 8.6 million new cases of TB (13% in HIV-infected persons) and 1.3 million deaths as a result of TB.[1] Fifty-eight percent of cases of TB occurred in Asia and 25% in Africa. The largest number of cases was in India (26% of all cases) and China (12%). Worldwide, TB occurs more commonly in males than females, but 210,000 women died of TB; the prevalence of disease peaks in young adults, with major economic consequence. Globally, both TB incidence (declining 2% per year) and case-fatality rates have been falling.

HIV infection has a profound effect on the epidemiology of TB, promoting and accelerating progression from infection to active TB and both reactivation and reinfection disease. Some 80% of HIV-infected TB cases are in sub-Saharan Africa, resulting in TB case rates as high as 1% in South Africa and Swaziland.

In the United States in 2013, the incidence of TB was 3.0 per 100,000, with 9588 new cases reported (6.8% were HIV-infected, 5.7% homeless, and 3.8% incarcerated).[2] The incidence rate in foreign-born individuals was 13-fold higher than in those born in the United States and foreign-born individuals accounted for 64% of new cases. Rates were 26-fold higher in non-Hispanic Asians than non-Hispanic whites. The age-specific prevalence of TB in the United States is skewed toward older adults, presumably representing reactivation disease in the foreign born and U.S. born that were infected when TB was more common.

Country of origin is a large determinant of both the risk of latent TB infection (LTBI) and of disease. In a low-prevalence setting such as the United States, the prevalence of LTBI (defined as a positive tuberculin skin test [TST] or interferon-γ release assay [IGRA] and no active disease) is approximately 4%. Those infected are at markedly increased risk of disease compared to uninfected, a risk that is further increased by medical comorbidities and other factors shown in Table 324-1. The risk is not homogeneous within groups affected, for example, by extent of immunosuppression in HIV or duration, severity, and control of diabetes. Smoking and alcoholism also confer an increased risk for TB, although smaller than the conditions listed.

TB caused by drug-resistant organisms is a particular and emerging threat to public health.[3] Multidrug-resistant (MDR) TB (resistant to isoniazid [INH] and rifampin [RIF]) and extensively drug-resistant (XDR) TB (MDR plus resistance to fluoroquinolones [FQs] and a second-line injectable (kanamycin, amikacin, or capreomycin) are much more difficult and expensive to treat and in some cases may be incurable. There are 450,000 incident cases of MDR TB each year, with an increasing prevalence because most are untreated, and 170,000 deaths. About 5 to 10% of these new cases of MDR TB are XDR. Twenty-seven countries account for 85% of the cases of MDR TB, with most occurring in India, China, and the Russian Federation. Eighty-six U.S. cases of MDR TB and two cases of XDR TB were reported in 2012. Most cases of MDR TB are not treated, which results in increasing transmission and prevalence.

PATHOBIOLOGY

Typically, the chain of transmission of TB begins with an infectious case of pulmonary TB (PTB) (Fig. 324-1). Infectiousness of a patient is determined by sputum smear status (3 to 4+ AFB), cough strength and frequency, the presence of cavitary lung disease, and the characteristics of the physical space shared with the source (ventilation and air recirculation). However, not all strongly AFB smear–positive patients with PTB are equally infectious, and there may be high transmitters, owing to host or bacterial factors or both. Recent studies, in fact, show that only about 50-60% of strongly sputum smear–positive persons with PTB generate aerosols that contain viable organisms.[4]

In both low- and high-prevalence countries, exposure/infection usually occurs in the household. In this setting, where exposure may be intense and protracted, 50 to 75% of contacts become infected. The higher numbers result from studies in which repeated testing identifies all TST convertors. In outbreaks occurring in residential shelters, hospitals, and prisons, Mtb infection or disease has been documented after brief exposure. Important variables that may explain differences in transmission include virulence of the organism, innate immunity, and susceptibility of the exposed populations (e.g., HIV infected). Human genetic factors such as polymorphisms in expression or regulation of toll-like receptors (TLR), pattern recognition

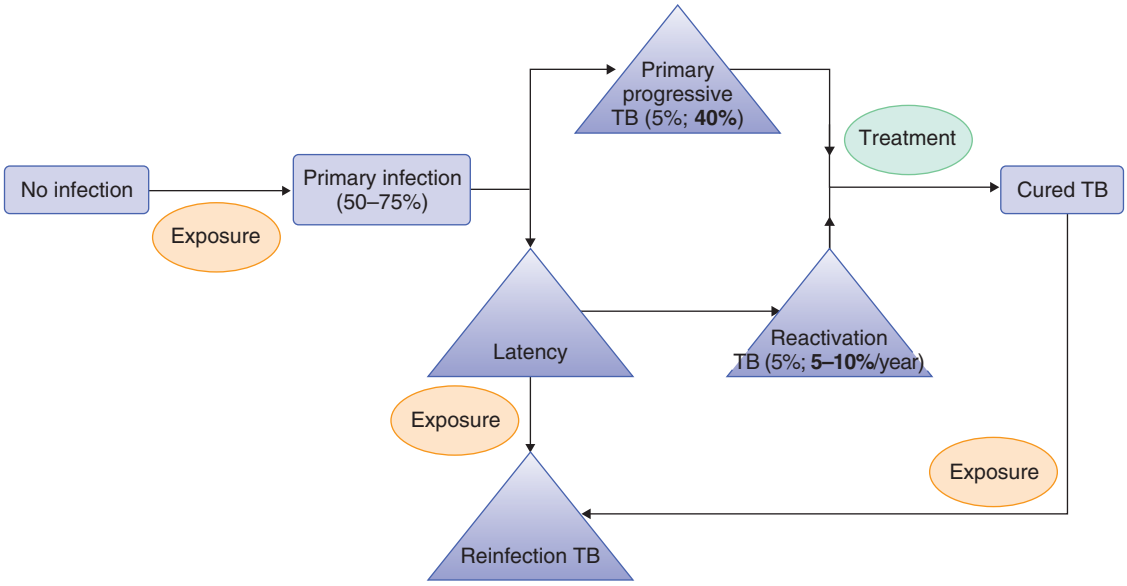

FIGURE 324-1. Natural History of TB. The proportion of individuals affected are shown in parentheses. Bolded figures are for HIV infection with severe immunosuppression. A number of medical risk factors besides HIV promote progression from *Mycobacterium tuberculosis* infection to disease (see Table 324-1).

receptors important to innate immunity, may modulate risk of infection and expression of infection (TST) as well as risk and expression of disease. Predisposition to disease is seen with defects in interferon (IFN)-γ and interleukin (IL)-12 receptors, consistent with their role in immunity. Polymorphisms that modify inflammation (e.g., by affecting leukotriene A_4 hydrolase) also may affect disease manifestations and response to therapy.

Two models of the natural history of TB are shown in Figures 324-1 and 324-2. There is increasing evidence that that the natural history represents a continuum rather than distinct entities of latent and active TB.[5] Diagnostic biomarkers that stratify risk of progression from LTBI to active TB would be of enormous value for targeting public health interventions.

The lung is the site of most cases of reactivation TB. The hallmark of the pathology is granuloma formation with caseation necrosis and multinucleated Langerhans giant cells. The caseous material found in necrotic cavities contains AFB. Mtb multiplies exuberantly in the liquid caseum. Immunologically, expectorated sputum contains cytokines and both upregulators and downregulators of the immune and inflammatory response, the downregulation being dominant. Bronchoalveolar lavage shows a lymphocytic alveolitis, with an influx of immature macrophages representing monocytes attracted from blood.[6] In sum, there is an active but downregulated immune response concomitant with bacterial replication. As a consequence of the inflammation, extensive apoptosis occurs and may lead to the deletion of Mtb-responsive T cells, which may play a role in the requirement for a long duration of therapy and in the susceptibility to reinfection TB. In HIV-infected persons with advanced immunosuppression, granulomas may be poorly formed or absent. Lung tissue is infiltrated with foamy epithelioid cells, which are macrophages laden with AFBs. Caseation may or may not be present, but there is extensive inflammation and necrosis.

Extrapulmonary TB can involve any organ. Persistence of organisms in areas that are relatively well oxygenated may explain the more frequent sites of reactivation, such as the apices of the lung, cortices of the kidney, and vertebral bodies.

Several forms of extrapulmonary TB have a shared pathogenesis: discharge of a contiguous tuberculous focus into a serosal cavity, a brisk inflammatory reaction based on preexisting delayed-type hypersensitivity (DTH), fever, frequently negative smears of exudative fluid for AFBs, and sometimes a transiently negative TST. The focus that discharges may be a long-standing focus or one seeded during recent dissemination associated with primary infection. This basic scenario also occurs in pleural TB, TB pericarditis, TB meningitis (the parameningeal focus is called a "Rich focus"), TB peritonitis, and TB arthritis.

Pleural TB represents an in situ DTH reaction with activation of T_H1 helper T cells, abundant cytokines, including IFN-γ and tumor necrosis factor (TNF)-α, and apoptosis. In non–HIV-infected individuals, organisms are sparse, which may be why self-cure of pleural TB can take place. However, in the absence of chemotherapy for TB, there is a high risk for TB recurrence, usually as PTB on the side contralateral to the effusion.

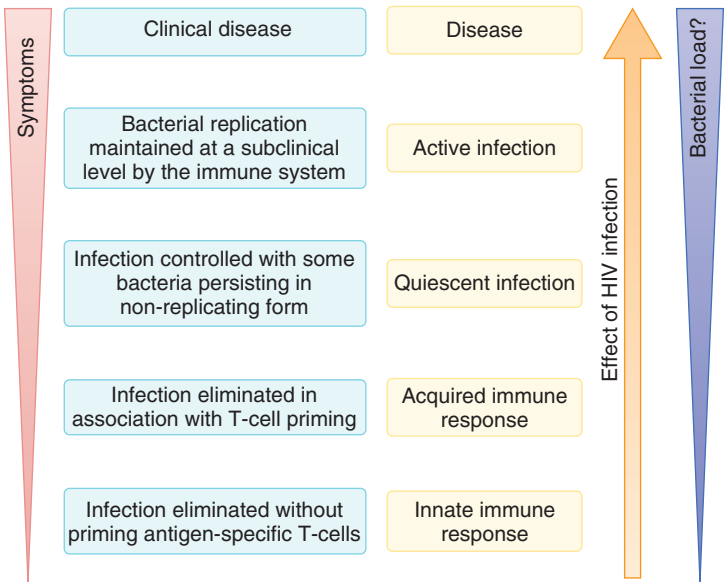

FIGURE 324-2. Tuberculosis infection as a spectrum. The outcome of infection by *Mycobacterium tuberculosis* is generally represented as a bimodal distribution between active tuberculosis (TB) and latent TB on the basis of the presence or absence of clinical symptoms. It is proposed that latent TB is usefully represented as part of a spectrum of responses to infection. One consequence of this model is that there may be a subpopulation within the group that is currently defined as having latent TB that should be preferentially targeted for preventive therapy. A second consequence is that efforts to develop drugs for effective treatment of latent TB would overlap the search for drugs that shorten treatment times for active TB. (From Barry CE 3rd, Boshoff HI, Dartois V, et al. The spectrum of latent tuberculosis. Rethinking the biology and intervention strategies. *Nat Rev Microbiol.* 2009;7:845-855, Figure 1.)

In addition to reactivation of a latent focus, reinfection with Mtb may occur and progress to disease. Reinfection is more likely if the host is immunosuppressed or there is repeated or intense exposure. Treated cases of PTB also are predisposed to reinfection disease as discussed earlier. LTBI is over 70% protective against reinfection TB.

CLINICAL MANIFESTATIONS

Primary Tuberculosis

Most cases of primary TB are unrecognized clinically except for conversion of the TST. There may be fever, shortness of breath, nonproductive cough, and rarely erythema nodosum. Crepitations and focal wheezes may be present. Chest radiographs show small patchy opacities in the mid-lung fields, often with unilateral hilar lymphadenopathy. Upper or middle lobe collapse may also be seen as a result of bronchial compression by enlarged nodes or transient pleural effusion. Recent studies with positron emission

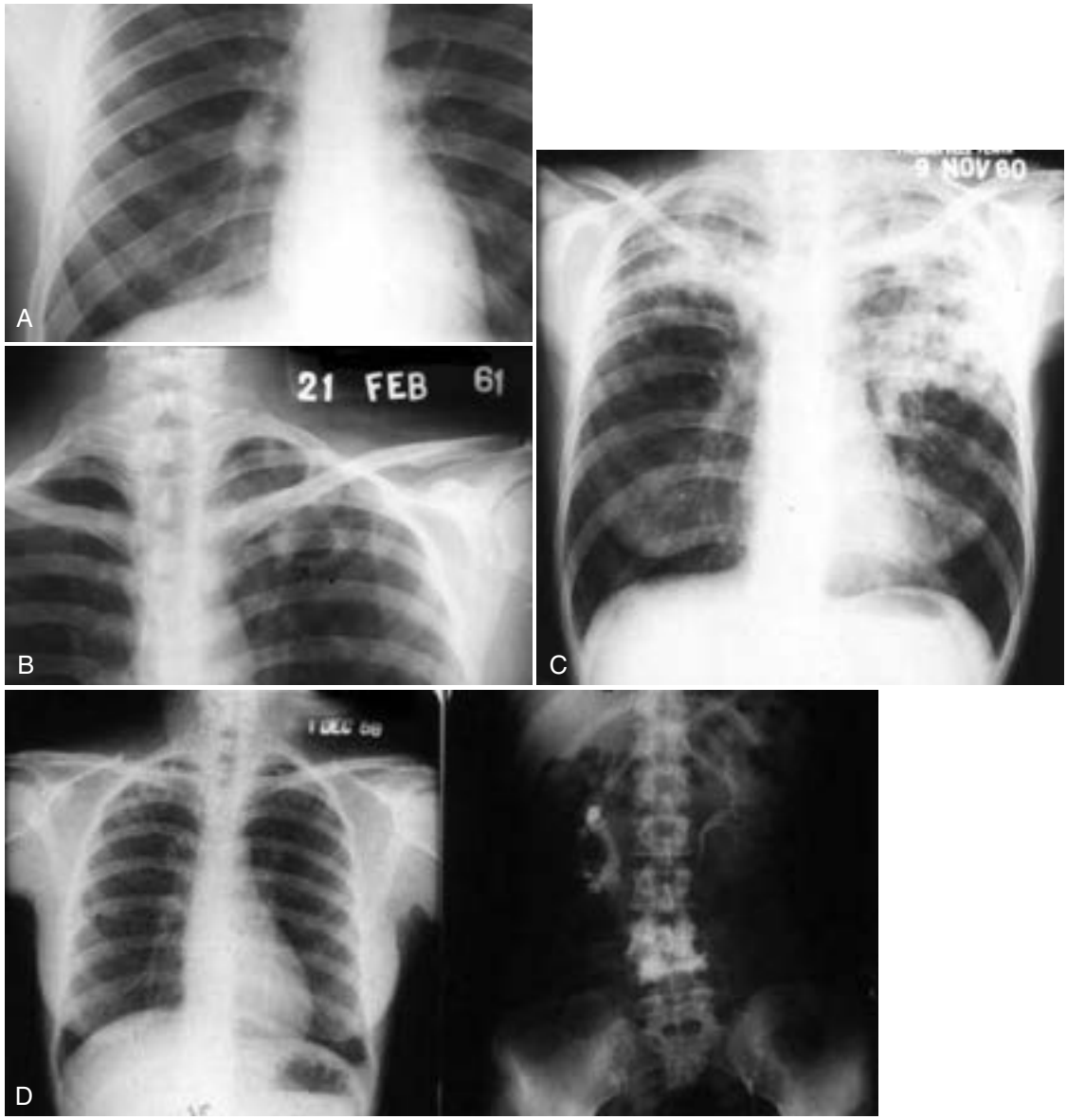

FIGURE 324-3. **A, Ghon complex. B, Moderately advanced pulmonary tuberculosis (TB). C, Far advanced pulmonary TB. D, Pulmonary (*left*) and extrapulmonary (*right*) TB. (Radiographs courtesy Thomas M. Daniel, MD.)**

tomographic computed tomography (PET-CT) scan show that most household contacts of infectious TB cases with a positive TST have mediastinal adenopathy that resolves with INH preventive therapy. HIV-infected persons with LTBI studied with this modality may show parenchymal uptake consistent with subclinical disease. In most individuals (immunosuppression being the exception), the manifestations of primary TB resolve without treatment, concurrent with the development of an adaptive immune response. During the subsequent period of clinical latency, evidence of the primary infection may be found as a small calcified parenchymal scar in the mid-lung fields (Ghon complex), sometimes associated with similar findings in the draining hilar nodes (Ranke complex) (Fig. 324-3A). A small scar caused by an arrested lesion in the apices of the lung is called a Simon focus.

Progressive Primary Tuberculosis

Failure to develop adaptive immunity is most common in young children, the elderly, and the immunocompromised. Progressive primary TB manifested as TB meningitis may develop in this setting. Primary infection also may progress to PTB within the first 1 to 2 years. In this case, PTB usually is upper lobe and cavitary, distant from the site of primary infection. Recent data indicate that in high-prevalence areas, most PTB cases represent progressive primary disease. Clinically, it is not possible to distinguish between progressive primary and "post-primary" or reactivation TB.

Reactivation Tuberculosis

The terms *reactivation TB* and *post-primary TB* are used interchangeably to connote that primary TB is followed by a variable period of at least 2 years of clinical latency, after which TB develops in the setting of existing DTH/adaptive immunity, which contributes importantly to the pathogenesis and clinical manifestations.

PTB is the most common form of reactivation TB. Typical clinical findings in PTB consist of the insidious onset of a productive cough, night sweats, anorexia, and weight loss. Fever is present in approximately half of those affected. Patients may be asymptomatic and the diagnosis suggested by a chest radiograph obtained for other reasons (subclinical TB). The sputum may be purulent, blood streaked, or frankly bloody. Pleuritic chest pain may occur when there is subpleural inflammation. Dyspnea is not a hallmark of PTB, in part because thrombosis of vessels limits the perfusion of inflamed areas, so hypoxemia is not a prominent clinical feature. Physical examination may show dullness to percussion, low-pitched amphoric (hollow-sounding) breath sounds, and occasionally crepitations that may be post-tussive. Chest radiographs often reveal more disease than suggested by physical examination (see Fig. 324-3B and C). Typically (>95% of cases), lesions are found in the apical and posterior segments of the upper lobes and the superior (dorsal) segment of the lower lobe. There is a progression from patchy opacities and consolidation to cavitation reflective of caseation and liquefaction. Advanced imaging with PET-CT shows a heterogeneity in metabolic activity of different lesions within the same patient (Fig. 324-4), which may be associated with variable response to treatment. Rupture and discharge into bronchi and intrabronchial spread may lead to disease in multiple areas, including the other lung (so-called TB bronchopneumonia). There may be involvement of the larynx and middle ear. Early cavities are thin walled and evolve into characteristic chronic thick-walled cavities. Ten percent of all cavities have an

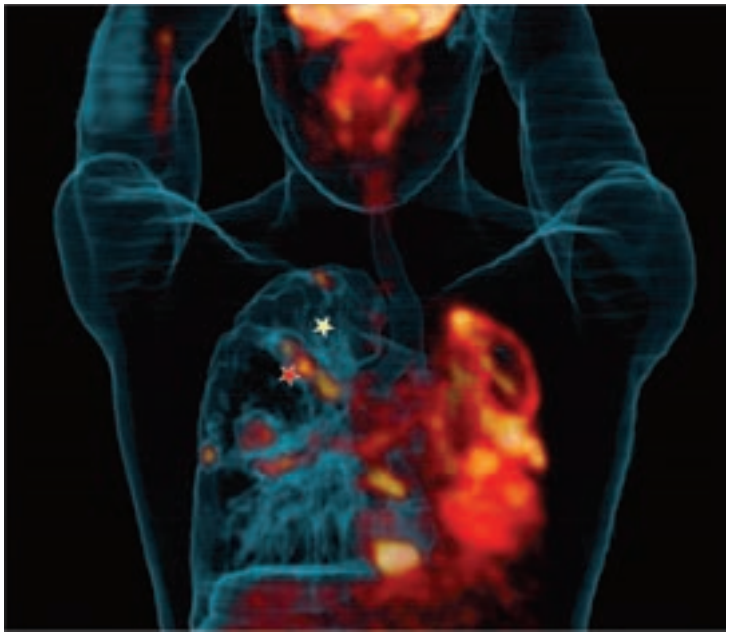

FIGURE 324-4. Positron emission tomographic computed tomography (PET-CT) imaging. An ^{18}F-fluorodeoxyglucose (FDG) PET-CT scan of a patient with tuberculosis with extensive bilateral disease and a complete collapse of the left lung. The right lung also shows extensive disease throughout and illustrates the variability of FDG-PET uptake among lesions within even a single infected patient. The yellow star illustrates one lesion that fails to take up FDG that lies immediately adjacent to a string of three lesions that take up label avidly (*red star*). These different types of lesions respond to chemotherapy with different kinetics, indicating that they represent distinct bacterial subpopulations in different microenvironments. (From Barry CE 3rd, Boshoff HI, Dartois V, et al. The spectrum of latent tuberculosis. Rethinking the biology and intervention strategies. *Nat Rev Microbiol.* 2009;7:845-855, Figure 2.)

air-fluid level. There may be an associated pleural effusion or rarely, with rupture of cavities into the air space, pyopneumothorax. If the disease is minimal, it may best be seen on apical lordotic chest radiographs or on CT. Rarely, chest radiographs are normal, and the accompanying symptoms and positive sputum smears may be the result of endobronchial lesions or rupture of a tuberculous node into bronchi. Healing, fibrosis, and contraction obliterate small cavities, although large cavities may persist and even become the eventual nidus for an aspergilloma or a "scar" carcinoma.

In immunocompromised persons, the opacities may be located in the mid- and lower lung fields and be manifested as poorly resolving lobar or segmental pneumonitis, atelectasis, nodules, and cavities. Early in the course of HIV infection when the $CD4^+$ count exceeds 200/μL, PTB may be typical in its manifestation. At lower $CD4^+$ counts, mid- and lower lung abnormalities are more common. At a $CD4^+$ count below 100/μL, the findings may be quite atypical, with prominent hilar and mediastinal adenopathy, pleural disease, interstitial or miliary opacities, or any combination of these manifestations. This picture resembles primary TB and, in fact, may represent progressive primary TB or reinfection disease. Chest radiographs are normal in up to 20% of persons with culture-confirmed TB, sometimes in the presence of a smear that contains AFB. In this CD4 strata, disseminated and extrapulmonary TB are the rule, with or without concurrent PTB.

Reinfection Tuberculosis

Reinfection TB is clinically indistinguishable from other forms and an important pathogenetic mechanism in high transmission settings. A shift in the drug susceptibility profile, a documented change in the DNA fingerprint or sequence, or occurrence in an outbreak setting may be the only evidence supporting the diagnosis of reinfection TB. Reinfection TB typically affects those with preexisting LTBI or following treatment of PTB.

Extrapulmonary Tuberculosis

Approximately 20% of cases of TB in non–HIV-infected populations are extrapulmonary (see Fig. 324-3D). In areas endemic for TB, extrapulmonary TB often occurs concurrently with PTB and is more common in children and young adults, in whom it represents progressive primary infection. By contrast, in low-prevalence areas, isolated extra-pulmonary TB is more common, and there is a shift to the elderly that represents reactivation TB. HIV infection is associated with a higher frequency of extrapulmonary disease, including the more serious forms, disseminated TB and TB meningitis.

Pleural Tuberculosis

Pleural TB occurs by direct extension when a subpleural caseous focus discharges into the pleural space or through hematogenous seeding. There may be concurrent PTB. Its peak occurrence is 3 to 6 months after primary infection. The typical manifestation is abrupt onset of fever, pleuritic chest pain, and cough. Occasionally there is an insidious presentation consisting of fever, weight loss, and malaise. If the pleural effusion is large enough, shortness of breath may be seen. Physical examination shows dullness to percussion and decreased breath sounds. Above the area of dullness there may be true egophony. Chest radiographs typically show unilateral pleural effusion, more frequently in the right hemithorax. Bilateral disease occurs in 10% of cases. Pleural effusions may be medium-sized, large, or, uncommonly, massive.

Miliary Tuberculosis

Miliary TB usually has an insidious manifestation consisting of fever, weight loss, night sweats, and little in the way of localizing symptoms or signs. There may be concurrent TB meningitis with associated symptoms. Physical examination may show choroidal tubercles (raised white-yellow plaques on funduscopic examination, present in 15% of cases), lymphadenopathy, and hepatomegaly. Chest radiographs may show multiple bilateral small opacities termed *miliary infiltrates* because of their resemblance to millet seeds. The findings on initial chest radiographs are often subtle and may be clear-cut only in retrospect after 3 months of follow-up. Performance of CT or high-resolution CT is useful because of its increased sensitivity (Fig. 324-5). A variant of miliary TB is disseminated are active TB, as may occur in HIV-infected patients or those treated with TNF inhibitors. In this entity, chest radiographic findings may be even more minimal. In the HIV-infected individual with advanced immunodeficiency, blood cultures are positive for Mtb in 20 to 40% of patients and may be the only manifestation of TB.

Tuberculous Meningitis

TB meningitis is usually characterized by less than 2 weeks of fever, headache, and meningismus. There may be depressed levels of consciousness, diplopia, and (rarely) hemiparesis. Physical examination shows a stiff neck and occasionally cranial neuropathy (VI, III, IV, VII in order of frequency) and long-tract signs. Chest radiographs may be consistent with PTB or miliary TB. CT of the head may show contrast enhancement over the basilar meninges, hypodense areas consistent with infarcts, hydrocephalus, and sometimes focal inflammatory lesions (tuberculomas). CT angiography may show entrapment of vessels or vasculitis.

Tuberculous Lymphadenitis

Lymphadenitis may be the sole manifestation of TB or, more frequently, particularly in the HIV infected, may accompany PTB. Patients with isolated lymph node disease may be afebrile. The supraclavicular and posterior cervical lymph nodes are most frequently involved. This is in contrast to scrofula caused by atypical mycobacteria or *M. bovis* and often seen in children, in whom submandibular and high anterior cervical adenopathy predominates. The lymphadenitis is not usually painful, and aspiration of the lymph node with the finding of AFB is an excellent approach to establish the diagnosis.

Tuberculous Pericarditis

The usual manifestation of TB pericarditis is chronic but may occasionally be subacute with fever, night sweats, chest pain, shortness of breath, pedal edema, and other signs of right heart failure. Physical examination shows signs of pericardial disease, right-sided heart failure, and tamponade (in ≈ 10%). Pericardial aspiration and biopsy are the diagnostic procedures of choice. When tamponade is present, a pericardial window can be both diagnostic and therapeutic.

Tuberculous Peritonitis

TB peritonitis may be accompanied by abdominal pain and fever, at times mimicking an acute abdomen. Alternatively, there may be an insidious manifestation consisting of abdominal pain, swelling, night sweats, and weight loss. The clinical syndrome is caused by discharge of tuberculous lymph nodes into the peritoneal space. Exudative ascites is usually present unless TB is superimposed on preexisting transudative ascites, as in alcoholic liver

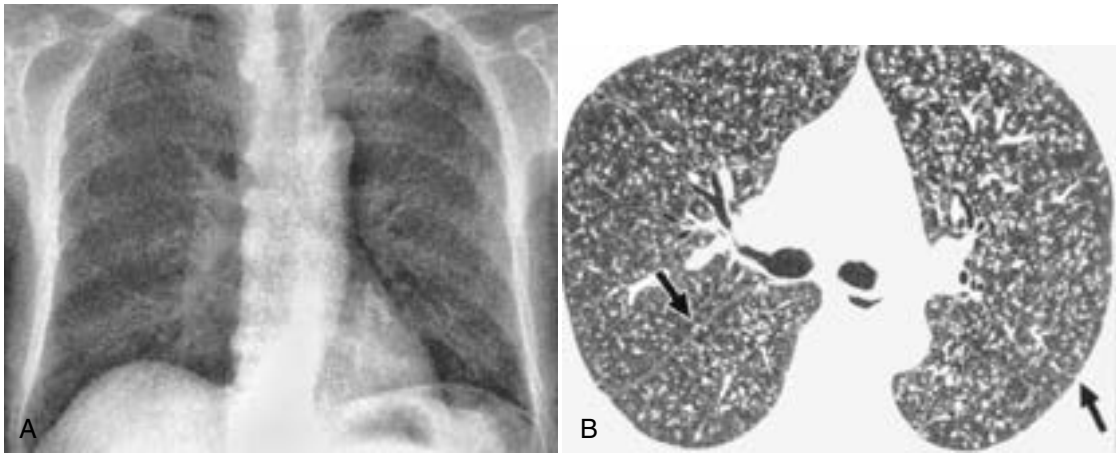

FIGURE 324-5. Miliary tuberculosis in a 70-year-old man. **A**, Posteroanterior chest radiograph shows evenly distributed, discrete, uniformly millet-sized nodular opacities in both lungs. **B**, High-resolution computed tomography (1.0-mm section thickness) at the level of the right upper lobar bronchus shows uniformly sized small nodules randomly distributed throughout both lungs. Note the subpleural and subfissural nodules (*arrows*). (From Jeong YJ, Lee KS. Pulmonary tuberculosis: up-to-date imaging and management. *AJR Am J Roentgenol.* 2008;191:834-844.)

disease. On physical examination, the abdomen has been described as "doughy," because matted loops of bowel may be palpable. A variant of this syndrome is perhaps best termed *abdominal TB*. In this case, the abdominal pain is subacute, the associated findings on physical examination less striking, and ascites less prominent or absent. The best method for diagnosis when ascites is present is laparoscopically guided peritoneal biopsy. In areas endemic for TB and HIV, the finding of intra-abdominal lymphadenopathy on abdominal ultrasound or CT is often used to support the diagnosis of abdominal TB.

Gastrointestinal Tuberculosis

Patients with gastrointestinal TB have fever, abdominal pain, diarrhea, and gastrointestinal bleeding or obstruction. Roentgenograms of the small bowel and abdominal CT show involvement of the terminal ileum, similar to Crohn's disease. The diagnosis is made on clinical suspicion in areas endemic for TB and HIV or by the finding of TB elsewhere. Occasionally, intraluminal biopsy of the terminal ileum or other involved sites is used to establish the diagnosis.

Renal Tuberculosis

There may be few symptoms and signs associated with renal TB, although occasionally dysuria, hematuria, and flank pain are present. The diagnosis is often suggested by the finding of sterile pyuria or hematuria as initial abnormalities that trigger evaluation. Physical examination is usually unremarkable. CT shows renal cortical scarring, occasionally with mass or cavitary lesions, papillary necrosis with calyceal and ureteral dilation, or "beading" of the ureter because of ureteral strictures.

Vertebral Osteomyelitis

The initial site of disease is the subchondral region of the anterior portion of the vertebral body. The lower thoracic and lumbar vertebrae are involved most commonly. The disc space is initially spared but becomes involved late with spread to adjacent vertebrae. Paravertebral "cold abscesses" may dissect through tissue planes. Patients have back and sometimes radicular pain. Occasionally and more often with cervical disease, there may be weakness of the legs and incontinence of stool and urine. Physical examination may show a gibbus deformity caused by anterior compression fractures or paraparesis. Radiographs of the spine, as well as CT and magnetic resonance imaging, may show abnormalities in adjacent vertebrae, with anterior compression (see Fig. 324-3D). Cold abscesses may be appreciated as well.

Other Forms of Extrapulmonary Tuberculosis

TB of the bone or joints may be manifested subacutely as a combination of synovitis and osteomyelitis. The joints involved may have sustained previous trauma. TB of the female genital tract may result in pelvic pain, menorrhagia, vaginal discharge, or infertility. Males may have an epididymal mass, sometimes seen in patients with miliary TB. TB can cause granulomatous uveitis.

DIAGNOSIS

Infection with *Mycobacterium Tuberculosis*

The diagnosis of latent TB infection (LTBI) is one of exclusion, based on the finding of delayed-type hypersensitivity (DTH) and the absence of active TB. The tuberculin skin test (TST) has been used widely, and there is strong epidemiologic evidence supporting its interpretation. Tuberculin purified protein derivative (PPD) derived from autoclaved culture filtrates of Mtb is used to elicit DTH. The response elicited by PPD is nonspecific because of broad cross-reactivity among tuberculous and nontuberculous mycobacteria and other organisms as well. The TST is performed by injecting 5 tuberculin units of PPD in 0.1 mL intradermally. The reaction is assessed as induration after 48 to 72 hours. Problems with the TST are legion. It is the only bioassay used in clinical medicine, and there is no quality control of the application or reading of the TST. The sensitivity of the TST is less in immunosuppressed patients, such as those with HIV infection, and also less in the presence of active TB. The TST may revert to negative over time, and it may be boosted by the application of PPD. On repeat testing the result is a "pseudoconversion" that does not represent new infection with Mtb. The greatest limitation of the TST is its nonspecificity. There has been uncertainty in its interpretation, particularly in the setting of previous vaccination with *M. bovis* bacille Calmette-Guérin (BCG). In fact, BCG administered once at birth has little effect on the TST beyond the first year. By 10 years of age, only 1% of positive TSTs can be ascribed to previous BCG administration.

The TST remains of value in the diagnosis of LTBI in at-risk individuals who are candidates for treatment (preventive therapy). Interpretation of the TST is based on a "sliding scale" that takes into account an individual's a priori risk for Mtb infection and, if infected, the risk for progressing to disease (see Table 324-1). Changing the cut point for positives, in effect, modifies the sensitivity and specificity of the TST. Routine testing is not recommended for low-risk populations; however, testing may be performed because of employment. Certain populations should undergo annual testing, including staff or individuals living or working in congregate settings (incarcerated, homeless, HIV-infected, correctional facility staff), injection drug users, and others at risk because of sociodemographic factors. For individuals who will undergo an annual TST, a true baseline should be established by two-step skin testing. After the initial negative TST, the test is repeated in 1 to 3 weeks (there is no need to repeat the test if the first TST is positive). An increased reaction size on the second TST is known as "boosting" and may be due to previous infection with non-tuberculous mycobacteria or Mtb or vaccination with BCG. TST conversion from negative to positive is the best indicator of intervening new infection with Mtb and is defined as an increase in reaction size of 6 mm or greater from less than 10 mm to 10 mm or greater. It also may be of value to perform a two-step TST on individuals older than 60 years and therefore at risk for TST reversion.

In HIV infection, the TST may be negative before administration of antiretroviral therapy (ART) and convert to positive with treatment. For this reason, it is recommended that the TST be repeated in TST-negative

HIV-infected persons once their $CD4^+$ count reaches 200/μL and annually thereafter.

The sensitivity of the TST is decreased in individuals with active TB, more so in certain forms of extrapulmonary TB. Though insensitive for the diagnosis of active TB, the TST has another application in low-prevalence areas. If the differential diagnosis of a clinical condition includes TB, establishment of previous Mtb infection increases the likelihood that the clinical findings represent TB. The TST is of particular value in the evaluation of smear-negative patients with pulmonary disease suggestive of PTB and in patients suspected of having extrapulmonary TB. Confirmatory findings such as positive culture or histology, response to therapy, and lack of an alternative diagnosis are necessary to establish the diagnosis of TB.

Interferon gamma releasing assays (IGRAs) (QuantiFERON-TB Gold test) have also been approved for the diagnosis of LTBI, and the CDC considers them interchangeable with the TST.[7] The assays represent in vitro cell culture in which blood cells are stimulated with a mix of antigens present in Mtb but not in BCG and most nontuberculous mycobacteria. The main advantage of IGRAs is specificity. The antigens used to stimulate cells are present in Mtb but not BCG. IGRAs also require only a single patient visit. The disadvantages of IGRAs are expense, technical requirements, and controversy over test sensitivity in certain situations such as HIV infection and in household contacts of patients with PTB. Another issue has been instability of the result in individuals undergoing annual testing with IGRAs close to the cut point. The main indication for performing an IGRA, therefore, is to diagnose LTBI in a person who received BCG vaccination beyond the first year of life. There are several other situations in which IGRAs may be useful, for example, confirmation of a positive TST.

Because treatment of LTBI is effective in TST-positive HIV-infected persons, and the risk for progression of infection is inordinately high in such individuals, it is important to perform tests with high negative predictive value. Therefore, both a TST and IGRA should be performed in HIV-infected persons and others at high risk of progression from Mtb infection to disease or of a poor outcome. If either TST or IGRA is positive, the individual is a candidate for treatment of LTBI. Recent studies indicate that IGRAs may be less sensitive than TST in household contacts recently infected with Mtb, a particularly high-risk group.[8] This may be due to delayed conversion of IGRAs relative to TST. Mathematical modeling suggests IGRAs should replace TST in immunocompromised and perhaps all individuals, and that despite their increased cost, they are cost-effective in the United States. Unfortunately, modeling currently is based on data that are quite variable between studies. A fourth-generation test, Quantiferon-Gold Plus will undergo evaluation shortly and may show improved and stable accuracy.

In high TB–prevalence areas, the World Health Organization (WHO) recommends preventive therapy for all HIV-infected persons because of the difficulty in implementing a TST program that would identify those that will achieve the greatest benefit of INH preventive therapy.

Active Tuberculosis

In countries with a high TB burden, the diagnosis of TB is often based on clinical symptoms and sputum microscopy. Clinical diagnosis without the benefit of culture confirmation or radiography is the norm in endemic countries where access to diagnostics is limited. The diagnosis is also made on clinical grounds alone when smears are negative and suspicion for TB is high. Cultures, if available, require several weeks to months, and the decision to begin therapy must be made promptly. Clinical diagnosis is particularly important in HIV-infected persons because of the risk for rapid progression of the TB if left untreated, the more frequent occurrence of AFB smear–negative PTB, and in those with forms of TB that are "paucibacillary" (pediatric, meningeal, miliary, abdominal, pleural, pericardial), in which bacteria are few and AFB smears typically negative. The diagnosis of miliary, abdominal, pleural, and pericardial TB may be confirmed by the finding of AFBs in biopsied tissue or by culture. In the absence of bacteriologic confirmation, either because cultures are negative or because they are unavailable, the final diagnosis often relies on response to therapy or establishment of an alternative diagnosis. It should be noted that the empirical approach, taken of necessity in resource-limited settings, leads to overdiagnosis and overtreatment of TB, which expends TB program resources and delays treatment of other infections. It is therefore preferable to attempt to establish a definite diagnosis based on demonstration of Mtb by smears, cultures, or nucleic acid amplification tests of infected secretions or tissue specimens.

Sputum microscopy is the standard approach to the diagnosis of PTB. A smear requires 1000 to 10,000 bacilli/mL to be read as positive. Both hot and cold carbol fuchsin methods (Ziehl-Neelsen and Kinyoun) are used extensively. The use of fluorochrome stains such as auramine-rhodamine allows more rapid screening of sputum smears and improves sensitivity by about 10%. Three specimens, preferably early morning samples, should be examined to establish the diagnosis. Yield is higher in the presence of cavitary lung disease. Approximately half of individuals with PTB are AFB sputum smear negative, and this proportion is higher in the HIV infected. The quantity of AFBs present in the sputum smear is a rough measure of the infectiousness of patients with PTB, and it is a convenient way to monitor response to treatment. On this basis, additional roles have evolved for the sputum smear as a tool to monitor the potential for transmission and response to therapy.

HIV-infected persons with PTB carry a particular risk for transmission to health care workers and have been documented as sources of nosocomial outbreaks. Therefore, in the United States, in areas of high prevalence of TB, HIV-infected persons with pulmonary symptoms may be placed into respiratory isolation until infectious TB can be reasonably excluded by three negative sputum smears for AFB on specimens separated by at least 8 hours.

The diagnosis of pediatric TB has always been problematic. Children do not produce sputum readily, and TB is often noncavitary, extrapulmonary, or both. Sputum samples may not be readily obtained from infants and children. Options in this case include sputum induction and gastric aspiration. The sensitivity of AFB smears of gastric aspirates and induced sputum is 25 to 30%. Therefore, the diagnosis often is based on clinical and epidemiologic features as well as response to therapy.

Bronchoscopy with bronchoalveolar lavage or transbronchial biopsy is another option for the diagnosis of TB and is useful in all severely ill individuals and the immunocompromised, in whom the diagnosis of TB or an alternative infection must be made quickly if treatment is to have an impact on patient outcome.

AFB smears should be performed on normally sterile fluid obtained from all patients suspected of having TB. The yield of smears from pleural fluid, pericardial fluid, ascitic fluid, and cerebrospinal fluid (CSF) is low in patients with TB but may be higher in those with HIV coinfection, particularly if immunodeficiency is advanced. In TB meningitis, CSF may clot spontaneously, and AFB stains of the clot have increased yield. Rapid diagnosis of some forms of extrapulmonary TB may be made by biopsy of tissue (pleural, pericardial, peritoneal, synovial, terminal ileum); the presence of granulomas, particularly if necrotizing, virtually confirms the diagnosis. Necrotizing granulomas are seen in TB and fungal diseases (particularly histoplasmosis, blastomycosis, coccidioidomycosis, and sporotrichosis). AFBs may also be seen in tissue histology, and Mtb cultured from the specimen.

The diagnosis of miliary TB can be suggested by CT of the chest and confirmed by transbronchial lung biopsy (highest yield), as well as biopsy of the liver, bone marrow, or abnormal lymph nodes. If TB meningitis is suspected and the patient is immunosuppressed, it is particularly important to exclude cryptococcal meningitis by performing a cryptococcal polysaccharide antigen test, as well as an India ink preparation on CSF sediment.

The diagnosis of TB from specimens of normally sterile fluid can be difficult and is increased by culturing relatively large volumes. In addition, the yield of biopsy and culture of tissue (pleura, pericardium) is additive. There are particular diagnostic features for various forms of extrapulmonary TB. In TB meningitis, the initial CSF examination may show neutrophil predominance, but this evolves into a lymphocytic meningitis (100 to 500 cells/μL) with high protein and depressed glucose. TB of the pleura, pericardium, and peritoneum is associated with an exudative effusion, often with a lymphocyte predominance. Low glucose may be found in 20% of TB effusions but limits the differential diagnosis considerably. For example, malignancy, empyema, and rheumatoid arthritis are the other causes of pleural effusion with low glucose. Pericardial fluid in patients with TB pericarditis may be bloody. Eosinophilic meningitis and chylous pleural effusions or ascites may also be seen in TB.

Currently, the gold standard for the diagnosis of TB is culture on solid (Löwenstein-Jensen) or in liquid (BACTEC MGIT 960 system) media. The mycobacteria growth indicator tube (MGIT) system, which has replaced BACTEC, is non-radiometric and based on oxygen quenching in the presence of replicating mycobacteria. When compared with solid media, culture with liquid media is more sensitive and growth is more rapid (1 to 3 weeks vs. 3 to 8 weeks for solid media). Once an isolate is available, drug susceptibility testing should be performed to guide therapy. This takes an additional 2 to 4 weeks on solid media, although isoniazid (INH) and rifampicin (RIF) susceptibility results are available in several days when the molecular line probe assay (described later) is used. Liquid medium can be inoculated with smear-positive specimens for direct drug susceptibility testing, which also accelerates the process. Once mycobacterial growth occurs, speciation is possible

within hours with commercially available DNA probes. Nuclear acid amplification tests are approved and commercially available for use in TB diagnostics. Their sensitivity is somewhat higher than that of AFB smears, and their specificity is excellent. Expense precludes routine use of such tests, however.

The Xpert MTB-RIF has transformed TB diagnosis globally, although there is limited uptake in the United States because of delays in approval. Through in situ DNA amplification reaction, this test allows a specific diagnosis of TB and determination of susceptibility to rifampicin within 90 minutes.[9] After minimal processing, sputum is added to a cartridge. Gene amplification is done with primers based on the *rpoB* gene, which encodes the target of RIF, and resistance-conferring mutations are detected. This method is capable of establishing the diagnosis of TB in 97% of patients with PTB, including 98% of AFB sputum smear–positive and 73% of smear-negative individuals, thus rivaling the sensitivity of solid culture. It does not require molecular expertise by the technician and is not subject to amplicon (DNA) contamination, because it is a closed system. The uptake of GenXpert has been remarkable. The government of South Africa is replacing sputum smear analysis with GenXpert for the diagnosis of TB. The government of Brazil is developing a similar policy. In Uganda, there will be a single reference laboratory for culture and drug susceptibility testing (DST). GenXpert will be available regionally and used mainly for the diagnosis of smear-negative cases. GenXpert is priced differently for low-income countries, but cost still may be prohibitive in some settings. The test allows more rapid diagnosis and initiation of treatment but may not affect the outcome of treatment depending on a site's effectiveness in treating smear-negative patients.[10]

There are other candidate nucleic acid amplification tests (NAATs) about to undergo evaluation that are less expensive than GenXpert and may be truly point of care. Diagnostics for pediatric and extrapulmonary TB more sensitive than NAATs are likely to be based on host responses. As regards host-based diagnostics, there are promising data based on transcriptomics, proteomics, and metabolomics.

TB in the HIV-infected patient poses particular diagnostic issues because of increased likelihood of smear-negative PTB, and in advanced HIV, atypical presentation and extrapulmonary disease.

The HIV-infected TB suspect poses unique difficulties in diagnosis. Recent data indicate that a dipstick to detect lipoarabinomannan, the major cell wall glycolipid of Mtb, has a sensitivity of 60% when $CD4^+$ cell count is below 100 cells/μL.[11] This would be the first point-of-care test.

The development of new diagnostics also targets rapid determination of drug susceptibility. The Xpert MTB-RIF test described earlier is promising in this regard. The commercially available molecular line probe assay (Hain Genotype MTBDR plus) has been approved in Europe and is recommended by the WHO. This assay is based on probing for mutations in the targets of INH and RIF. The assay requires the technician be trained in molecular methods, and results require several days. This assay also is useful for rapidly establishing drug susceptibility once cultures are positive.

TREATMENT

Comprehensive reviews of TB treatment have been published and provide complete information on drugs for TB, monitoring of therapy, management of adverse events, and treatment of pregnant women, children, and other special populations:

- http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5211a1.htm (treatment)
- http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5804a1.htm (treatment and prevention in HIV-infected persons)

Treatment of TB is both a clinical and a public health issue. The goals are to cure the patient and minimize transmission. For that reason, the treating physician has the obligation to ensure that treatment is completed with good adherence to medications. Because of the declining number of TB cases in the United States, with subsequent decline in expertise in TB management, treatment is more likely to occur in a public health clinic than in the private sector. The cornerstone of TB treatment is multidrug therapy. This is necessary because Mtb undergoes spontaneous mutation to drug resistance at a frequency such that most patients with cavitary lung disease—and therefore patients with a high burden of organisms—are likely to harbor resistant mutants.

For the treatment of TB caused by drug-susceptible organisms, there is an intensive phase of therapy for the first 2 months aimed at the rapidly dividing and metabolizing organisms and usually resulting in sterilization of sputum in those with PTB. This is followed by a 4- to 6-month continuation phase that kills the slowly metabolizing persisting organisms. The five first-line anti-TB drugs, INH, RIF, ethambutol (EMB), pyrazinamide (PZA), and streptomycin (SM), form the foundation of chemotherapy for TB. The precepts of therapy have been largely defined by controlled clinical trials. Standard short-course therapy for PTB requires a 2-month intensive phase of four drugs (INH, RIF, EMB, and PZA), followed by a 4-month continuation phase with INH and RIF.

The use of intermittent dosing regimens increases the feasibility of directly observed therapy (DOT) but is controversial. Intermittent regimens may be slightly less effective than daily 5/7 regimens (5 weekdays observed, 2 weekend days self-administered). For patients with extensive cavitary disease and the severely immunosuppressed, daily therapy should be administered throughout the entire course. For HIV-infected persons with TB who are not severely immunosuppressed, treatment should be daily during the intensive phase and at least three times weekly during the continuation phase. Non–HIV-infected persons without extensive cavitary disease can be treated with intermittent regimens throughout. Originally, EMB was perceived to be a "place holder" should the isolate prove to be resistant to INH, and the recommendation was to eliminate its use once drug susceptibility test results were available. There is recent evidence that EMB may be important because of its synergistic or additive activity with other first-line drugs, thereby providing an argument against its premature discontinuation before a full 2 months of therapy has been received.

The problem of nonadherence has led to the directive to provide DOT. The addition of PZA to the intensive phase allows for the so-called short-course chemotherapy; if PZA is not tolerated, comparable outcomes can be obtained with 9 months of INH-RIF. Treatment of drug-susceptible TB with this standard regimen can be expected to cure approximately 90 to 95% of cases. Routine monitoring of liver function test results is not recommended unless there is preexisting liver disease. Patients should return to the clinic promptly with any signs of drug toxicity, particularly those of early hepatotoxicity (nausea, malaise, anorexia, upper abdominal discomfort). Visits should be scheduled monthly and include clinical assessment and sputum examinations. For those with PTB, sputum cultures are continued until two consecutive cultures are negative. In uncomplicated PTB, defervescence is expected within 2 weeks, and there should be weight gain and diminution of cough and chest pain. About 20% of patients with cavitary PTB remain sputum culture positive after 2 months of therapy. A positive sputum culture at 3 months or failure to improve on chest radiographs suggests nonadherence with therapy, drug-resistant TB, or an alternative or complicating diagnosis. Positive sputum culture at 4 months is defined as treatment failure.

Extrapulmonary TB is usually associated with a smaller bacterial burden than is the case with PTB and can be treated with standard short-course regimens of 6 to 9 months' duration. However, because of the serious ramifications of treatment failure, more prolonged treatment of at least 9 to 12 months is recommended for miliary, meningeal, and skeletal TB. Adjunctive surgical débridement and stabilization may be necessary for skeletal TB. Adjunctive corticosteroids are indicated for TB pericarditis and severe pleurisy, as well as extensive PTB with clinical toxicity or respiratory failure. In TB meningitis, dexamethasone improves survival but has not had an impact on the proportion surviving with severe neurologic sequelae. Ventricular shunting may be necessary to relieve hydrocephalus.

HIV-TB coinfection creates additional management issues. Fortunately, the response to treatment of TB is comparable to that in HIV-uninfected individuals, except for higher early mortality. Intermittent regimens have a tendency to lead to RIF resistance; therefore, daily administration of drugs is recommended in the intensive phase, and drugs should be administered no less than three times weekly in the continuation phase. In resource-limited settings, the administration of cotrimoxazole prophylaxis is associated with improved survival. Integration of the treatment of TB and HIV rather than sequential treatment of first TB and then HIV is associated with a 56% reduction in mortality. The results of three randomized controlled trials support the current Department of Health and Human Services and Infectious Diseases Society of America guideline to start ART 2 weeks after initiation of TB treatment if $CD4^+$ count is below 50 cells/μL. Early initiation of ART does incur a risk for paradoxical reaction, a form of immune reconstitution inflammatory syndrome (IRIS) (Chapter 395). TB-IRIS is more likely to occur when the $CD4^+$ count is low, the viral load is high, and the interval between starting TB drugs and ART is short.[12] Immune reconstitution is associated with inflammation and transient exacerbation of disease mimicking progression of TB. There may, for example, be fever, lymphadenitis, and pleural or worsening parenchymal disease, including consolidation and new or progressing nodular opacities on chest radiographs. In about one third of dually infected patients, IRIS will develop within the first 2 months of treatment and often within the first 2 to 3 weeks of starting ART. TB-IRIS usually is not an important cause of mortality. A controlled trial has shown that corticosteroid therapy limits morbidity, although it is not useful for treating tuberculous pericarditis. More problematic is IRIS associated with respiratory failure or neurologic involvement. Initiation of ART in HIV-infected persons may also be associated with "unmasking TB," which occurs within 3 months. This may be due to a missed diagnosis of TB in screening, the development of inflammation at sites of mycobacterial replication in tissue, or progression of latent infection consequent to ART. Unmasking TB is common in countries endemic for TB (5 to 10% in Uganda, for example, but about 25% in South Africa) and associated with morbidity, mortality, and risk for nosocomial transmission.

In general, TB does not affect the response of HIV to ART. A major issue, however, is the interactions between rifamycins that induce CYP3A hepatic microsomal enzymes and protease inhibitors and some non-nucleoside reverse transcriptase inhibitors (NRTIs). Because of its potency, simplicity, and proven clinical efficacy, efavirenz 600 mg with two NRTIs, along with rifampin-based TB regimens, is the preferred strategy for co-treatment of HIV and TB. Rifabutin is off-patent and increasingly available as a less potent inducer of cytochrome enzymes, but its own pharmacokinetics can be affected by certain ARTs. Rifampin is the preferred rifamycin for efavirenz containing regimens, whereas rifabutin should be used at a does of 150 mg daily with a boosted protease inhibitor. Guidelines for other drug combinations for managing adverse events were updated June 2013 and are available at http://www.cdc.gov/tb/TB_HIV_Drugs/default.htm.[13]

Management of TB-HIV coinfection may be complicated for the clinician. For example, a new fever may be due to drug reaction, TB-IRIS, drug resistance, or a complicating opportunistic infection.

Drug-resistant TB is more difficult to cure than drug-susceptible TB and in some instances may be incurable. DOT is particularly important to prevent acquisition of additional drug resistance. Selection and monitoring of treatment for drug-resistant TB should be the responsibility of those experienced with the unique drug regimens and issues involved. General issues concerning the management of drug-resistant TB are listed in Table 324-2. A major problem, however, is the lack of information on drug susceptibility at the time of initiation of treatment. In resource-limited settings, drug susceptibility testing may not be available at all. Drug resistance is more likely in retreatment cases, so standardized retreatment regimens are administered. The risk for drug resistance is greater if the initial regimen was not administered in a DOT program. For patients who relapse, did not receive DOT, were not treated with a RIF-containing regimen, were known or presumed to have irregular treatment, or have severe disease, an expanded regimen should be started. An example would be INH, RIF, and PZA plus an injectable (streptomycin if not previously used) and a fluoroquinolone, with or without an additional oral agent (ethionamide, para-aminosalicylic acid [PAS], or cycloserine). Ideally, the retreatment regimen should contain at least three drugs to which the isolate is likely to be susceptible. In resource-poor environments, streptomycin alone is often added to INH, RIF, EMB, and PZA. Although most retreatment cases remain drug susceptible, about 25% do not respond adequately to the standard regimen.[14] This approach is clearly inadequate for MDR or XDR TB. Ideally, drug susceptibility testing should be performed for all patients being retreated to ensure success of treatment and prevent acquisition of additional resistance.

Monoresistance to INH or streptomycin has little effect on the outcome of TB treatment. INH-resistant TB can be treated with RIF, PZA, and EMB for 6 months. Ofloxacin may be added for extensive or refractory disease. RIF-resistant TB can be treated with 12 to 18 months of INH, EMB, ofloxacin, and PZA (plus an injectable for the first 2 months if there is extensive disease). Acquisition of additional drug resistance is more likely if the initial isolate was resistant to one drug (6%) or more than one drug (14%) versus being pansusceptible (0.8%).

MDR TB represents a problem of a different order of magnitude. Not only is cure difficult and extremely expensive, but unrecognized MDR TB can also lead to nosocomial outbreaks and rapid and high rates of fatality in the HIV infected.

TABLE 324-2 PRINCIPLES OF MANAGEMENT OF TUBERCULOSIS CAUSED BY DRUG-RESISTANT ORGANISMS

- Do not add a single drug to a failing regimen.
- When starting or modifying therapy, add three previously unused drugs (one an injectable) to which there is susceptibility.
- In MDR TB where there is resistance to first-line drugs, in addition to INH and RIF, treat with 4-6 drugs.
- Patients with MDR TB have highest priority for DOT, because treatment failure may mean XDR TB.
- Intermittent therapy should not be used (except for injectables after 2-3 mo).
- Do not use drugs to which the Mtb isolate is resistant. Low-level INH resistance may be the exception.
- There is cross-resistance among rifamycins but not between streptomycin and other aminoglycosides.
- There may be susceptibility to moxifloxacin but resistance to other fluoroquinolones.
- Drug susceptibility testing for PZA is complex technically and not performed in most laboratories. Monoresistance to PZA suggests *Mycobacterium bovis*.
- Always consult with an expert. There is predisposition to acquisition of additional drug resistance that will decrease the chances of cure.

DOT = directly observed therapy; INH = isoniazid; MDR TB = multidrug-resistant tuberculosis; Mtb = *Mycobacterium tuberculosis*; PZA = pyrazinamide; RIF = rifampin; XDR TB = extensively drug-resistant tuberculosis.

There is no substitute for performing drug susceptibility testing that includes all first-line and available second-line drugs. MDR TB is not a homogeneous entity. Frequently, the first diagnosis of MDR TB occurs in the setting of residual susceptibility to most other drugs. However, after repeated treatments, the isolate usually acquires additional drug resistance and, regardless of whether it fulfills the definition, is essentially XDR.

The cornerstone of treatment of MDR TB is administration of at least four drugs to which the isolate is susceptible (see Table 324-2). Typically, the regimen will include first-line drugs with retained activity, a fluoroquinolone (FQ), an injectable (amikacin, kanamycin, or capreomycin), and a second-line drug if disease is extensive (ethionamide, cycloserine, PAS). The duration of treatment is set at 12 to 18 months (12 to 15 months after sterilization of sputum). Residual susceptibility to FQ treatment and capreomycin has been shown to be a determinant of treatment outcome. Surgical resection is a consideration if the disease is localized, sputum remains culture positive, medical therapy is not tolerated, or massive hemoptysis is present. A recent review of published case series and cohort studies showed that surgical resection is beneficial in the adjunctive treatment of drug-resistant TB; however, the results might not be applicable in all settings, and well-designed studies are needed. Cure rates of 60 to 80% can be expected if therapy is targeted to drug susceptibility test results and MDR organisms remain sensitive to enough chemotherapeutic drugs that a reasonable regimen can be established. The U.S. Food and Drug Administration (FDA) recently approved both bedaquiline and delamanid for the treatment of MDR TB.[15] Bedaquiline, a diarylquinoline, targets mycobacterial ATP synthase, and delamanid, a nitrodihydroimidazooxazole inhibits Mtb mycolic acid synthesis. Both accelerate sterilization of sputum when added to regimens for treatment of MDR TB. It is recommended that bedaquiline be added to three drugs active against the isolate. These are the first new classes of TB treatment drugs to be approved by the FDA in 40 years. They should revolutionize management of MDR TB. Spectinamides, a new class of semisynthetic anti-TB agents, have been shown to overcome the efflux pump that is upregulated in MDR strains.[16]

The definition of XDR TB requires that an organism be MDR with additional resistance to an FQ plus at least one of three injectables (amikacin, kanamycin, capreomycin). The outcome of treatment of XDR TB has been variable. The best reported results have been obtained in Peru, with a comprehensive approach that included therapeutic regimens that were tailored according to drug susceptibility testing. Effective regimens included cycloserine, capreomycin, and PAS. Moxifloxacin may be active even if the isolate is resistant to first-generation FQs. Adjunctive surgery should be considered. A recent study indicates that linezolid has remarkable activity against XDR TB,[ ] and other oxazolidinones are in development for TB treatment.

The emergence of drug resistance has emphasized the need for new drugs and new drug regimens. A drug regimen that shortens the course of TB would lessen the emergence of drug resistance, because adherence would increase and DOT would be simplified. Once MDR TB develops, particularly in the setting of XDR TB, new drugs are necessary to improve efficacy and shorten treatment.

There are promising developments with existing classes of drugs, as well with as new classes about to enter or already in clinical trials. Rifapentine, a rifamycin with a longer half-life than RIF, shows remarkable enhancement of sterilization of sputum at 2 months. Moxifloxacin has increased activity against Mtb, although data of its efficacy in 4-month regimens have been disappointing to date. PA-824 and delamanid are nitroimidazole derivatives that are active against slowly replicating bacilli. Also in development are SQ-109 (ethylenediamine), an ethambutol derivative, and sutezolid (an oxazolidinone more active than linezolid).

PREVENTION

The issues surrounding prevention are presented in a comprehensive statement published in 2000 (http://www.cdc.gov/mmwr/preview/mmwrhtml/rr4906a1.htm), and prevention in HIV-infected persons is presented at the website http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5804a1.htm.

The approach taken in a low-prevalence setting such as the United States is to target tuberculin skin testing to those at high risk for recent Mtb infection and to those with comorbid conditions that predispose to progression from infection to disease. In either category, a positive TST becomes an indication for treatment of LTBI. The cut point of TST is adjusted according to risk of MTB infection and risk of progressing from infection to disease (see Table 324-1).

LTBI can be treated with rifampicin for 3 to 4 months, INH for 6 to 12 months, or the combination for 3 to 4 months,[17] thereby decreasing the lifetime risk for development of TB by approximately 75 to 90%, depending on the level of adherence to treatment. Pyridoxine should be administered to prevent INH-induced peripheral neuropathy. Monitoring of liver function is indicated in older individuals (the risk for hepatotoxicity increases beyond the age of 35 years) and in those with significant alcohol intake or underlying

liver disease (or both). INH should be discontinued if symptoms develop (see Treatment) or hepatic transaminase levels rise to more than three to five times the upper limit of normal. Rifapentine plus INH once weekly for 3 months (12 doses) is as effective as 9 months of INH alone in preventing active TB in patients with latent infection.[A7] Administration of this regimen should be directly observed. In HIV-infected persons, treatment of LTBI confers short-term efficacy (≈1 year), but this may be extended if ART is administered. Treatment for 3 years confers an additional benefit that is most marked in the TST positives. Unfortunately, treatment of LTBI has not been applied broadly. This is in part due to the concern that if screening is inadequate, patients with active TB will be treated with a single drug and thereby acquire drug resistance. The available data do not support this concern. In a recently reported cluster-randomized study, a trial of mass screening and INH preventive therapy for TB control among gold mine workers in South Africa (where TB is epidemic) had no significant effect on TB control despite the successful use of INH in preventing TB during treatment.[A8] In a low-resource setting in the HIV infected, the absence of current fever, weight loss, or night sweats has a negative predictive value of 97%, may obviate the need for chest radiography if it is not routinely available, and identifies a group that should be treated with INH preventive treatment regardless of TST status. Although INH preventive therapy is the preferred regimen, there is renewed interest in the possibility that rifamycin-containing regimens may confer more sustained protection. There are no data on how to treat LTBI caused by a known or probably drug-resistant isolate. PZA plus EMB or ofloxacin has been used after known exposure to MDR TB that is susceptible to the combination, but gastrointestinal intolerance is very common. Studies of linezolid are planned in this population.

Secondary preventive therapy after completion of TB treatment is indicated in HIV-infected persons in the setting of intense exposure to Mtb. Data demonstrating efficacy of secondary prevention largely result from studies of adults in Congo and gold miners in South Africa.

A randomized controlled trial has now demonstrated that in HIV-positive patients, early initiation of ART as compared with delayed treatment (where there was a decline in CD4 counts or AIDS-related illness had already occurred) led to a significant decrease in the development of TB.[A9]

Tracing of household contacts is a critical element of TB programs. The household is a major site of TB transmission, particularly in an area of low endemicity, so treatment of TST converters is an important strategy for elimination of TB. It is essential that transmission be limited in hospitals and other settings in which infectious PTB patients may comingle with susceptible hosts. The risk of transmission to health care workers should be ascertained annually by skin testing. The finding of excessive risk for new TB infection should lead to focused measures for reducing transmission. All TB suspects should be placed in respiratory isolation in negative-pressure rooms with at least six air exchanges per hour and high-efficiency particulate air (HEPA) filtration or ultraviolet irradiation. N-95 personnel respirator devices that are fit-tested are necessary for individuals entering areas with a known high risk for exposure.

BCG vaccine is widely used and relatively safe except in the setting of immunosuppression. Unfortunately, efficacy has been variable by age and by latitude. A meta-analysis indicated an overall efficacy of 50%. It is approximately 80% effective in preventing the severe forms of TB in childhood, miliary TB, and TB meningitis. However, its failure to prevent adult TB, particularly at low latitudes that include the most endemic areas, means that BCG vaccine does not have an impact on the public health problem of TB. The development of a more effective protective vaccine has been hampered by the absence of good animal models and the lack of correlates of protective immunity in humans. The issue is further compounded by the natural history of TB such that large, lengthy, and expensive trials may be needed to establish protective efficacy. Nonetheless, several vaccine candidates have been developed and their potential efficacy supported by preclinical studies in animal models. Two of the 12 currently in clinical trials are MVA85A, a recombinant modified vaccinia virus expressing the 85A antigen, a major secretory product of Mtb, and Mtb72f, a combination of two immunogenic antigens, Mtb 32a and Mtb 39a, in two adjuvants, ASO2a and ASO1B. A recent trial of the MVA85A vaccine in infants to boost the BCG response failed to show efficacy.

PROGNOSIS

In the pre-chemotherapeutic era, minimal PTB stabilized in about 50% of cases. Pleural TB also self-cured, with a risk for reactivation as noted previously. With treatment, the prognosis of patients with TB depends on the extent of PTB, the sites of extrapulmonary TB, drug susceptibility of the isolate, and the presence of HIV infection and other comorbid conditions. With extensive PTB, respiratory failure may supervene with a poor prognosis. There is also increased risk for serious, sometimes lethal, hemoptysis and pneumothorax. Miliary TB is associated with a high case-fatality rate, in part related to delays in diagnosis. TB meningitis is associated with serious neurologic residua, as well as high mortality. MDR TB and XDR TB are accompanied by high rates of treatment failure, morbidity, and mortality. TB in HIV-infected persons is associated with high early mortality that is poorly characterized. In the absence of ART, there is increased risk for other opportunistic infections and progression of HIV disease. The addition of ART, however, results in the morbidities associated with the concurrent administration of TB and HIV medications and with TB-IRIS. Long-term outcomes of patients with XDR TB were recently reported from South Africa. Their outcomes were poor irrespective of HIV status, and because of the scarcity of long-stay or palliative care facilities, substantial numbers of patients with XDR TB who had failed treatment and had positive sputum cultures were being discharged to likely transmit disease into the wider community.[18]

Grade A References

A1. Havlir DV, Kendall MA, Ive P, et al. Timing of antiviral therapy for HIV-1 and tuberculosis. *N Engl J Med*. 2011;365:1482-1491.

A2. Blanc F-X, Sok T, Laureillard D, et al. Earlier vs later start of antiviral therapy in HIV-infected adults with tuberculosis. *N Engl J Med*. 2011;365:1471-1481.

A3. Meintges G, Wildkinson RJ, Morroni C, et al. Randomized placebo-controlled trial of prednisone for paradoxical tuberculosis-associated immune reconstitution syndrome. *AIDS*. 2010;24:2381-2390.

A4. Mayosi BM, Ntsekhe M, Bosch J, et al. Prednisolone and *Mycobacterium indicus pranii* in tuberculous pericarditis. *N Engl J Med*. 2014;371:1121-1130.

A5. Diacon AH, Pym A, Grobusch M, et al. The diaryquinoline TMC 207 for multidrug resistant tuberculosis. *N Engl J Med*. 2009;360:2397-2405.

A6. Lee M, Lee J, Carroll MW, et al. Linezolid for treatment of chronic extensively drug resistant tuberculosis. *N Engl J Med*. 2012;367:1508-1518.

A7. Sterling TR, Villarino ME, Borisov AS, et al. Three months of rifapentine and isoniazid for latent tuberculosis infection. *N Engl J Med*. 2011;365:2155-2166.

A8. Churchyard GJ, Fielding KL, Lewis JJ, et al. A trial of mass isoniazid preventive therapy for tuberculosis control. *N Engl J Med*. 2014;370:301-310.

A9. Grinsztejn B, Hosseinipour MC, Ribaudo HJ, et al. Effects of early versus delayed initiation of antiretroviral treatment on clinical outcomes of HIV-1 infection: results from the phase 3 HPTN 052 randomised controlled trial. *Lancet Infect Dis*. 2014;14:281-290.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. World Health Organization. *Global Tuberculosis Report 2014*. Geneva: WHO; 2014. http://www.who.int/tb/publications/global_report/en/. Accessed January 28, 2015.
2. Alami NN, Yuen CM, Miramontes R, et al. Centers for Disease Control and Prevention. Trends in tuberculosis–United States, 2013. *MMWR Morb Mortal Wkly Rep*. 2014;63:229-233.
3. Matteelli A, Roggi A, Carvalho AC. Extensively drug-resistant tuberculosis: epidemiology and management. *Clin Epidemiol*. 2014;6:111-118.
4. Fennelly KP, Jones-López EC, Ayakaka I, et al. Variability of infectious aerosols produced during coughing by patients with pulmonary tuberculosis. *Am J Respir Crit Care Med*. 2012;186:450-457.
5. Dutta NK, Karakousis PC. Latent tuberculosis infection: myths, models, and molecular mechanisms. *Microbiol Mol Biol Rev*. 2014;78:343-371.
6. Cambier CJ, Takaki KK, Larson RP, et al. Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. *Nature*. 2014;505:218-222.
7. Mazurek GH, Jereb J, Vernon A, et al. Updated guidelines for using interferon gamma release assays to detect *Mycobacterium tuberculosis* infection–United States, 2010. *MMWR Recomm Rep*. 2010;59:1-25.
8. Ribeiro-Rodrigues R, Kim S, Coelho da Silva FD, et al. Discordance of tuberculin skin test and interferon gamma release assay in recently exposed household contacts of pulmonary TB cases in Brazil. *PLoS ONE*. 2014;9:e96564.
9. Boehme CC, Nabeta P, Hillemann D, et al. Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med*. 2010;363:1006-1015.
10. Theron G, Zijenah L, Chanda D, et al. Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. *Lancet*. 2014;383:424-435.
11. Nakiyingi L, Moodley VM, Manabe YC, et al. Diagnostic accuracy of a rapid urine lipoarabinomannan test for tuberculosis in HIV-infected adults. *J Acquir Immune Defic Syndr*. 2014;66:270-279.
12. Luetkemeyer AF, Kendall MA, Nyirenda M, et al. Tuberculosis immune reconstitution inflammatory syndrome in A5221 STRIDE: timing, severity, and implications for HIV-TB programs. *J Acquir Immune Defic Syndr*. 2014;65:423-428.
13. Centers for Disease Control and Prevention. Managing drug interactions in the treatment of HIV-related tuberculosis (on-line). http://www.cdc.gov/tb/publications/guidelines/tb_hiv_drugs/default.htm/. Accessed January 28, 2015.
14. Peterson S, Nyakoojo G, Fennelly K, et al. Effectiveness of the standard WHO recommended retreatment regimen (category II) for tuberculosis in Kampala, Uganda: a prospective cohort study. *PLoS Med*. 2011;8:e1000427.
15. Centers for Disease Control and Prevention. Provisional CDC guidelines for the use and safety monitoring of bedaquiline fumarate (Sirturo) for the treatment of multidrug-resistant tuberculosis. *MMWR Recomm Rep*. 2013;62:1-12.
16. Lee RE, Hurdle JG, Liu J, et al. Spectinamides: a new class of semisynthetic antituberculosis agents that overcome native drug efflux. *Nat Med*. 2014;20:152-158.
17. Stagg HR, Zenner D, Harris RJ, et al. Treatment of latent tuberculosis infection: a network meta-analysis. *Ann Intern Med*. 2014;161:419-428.
18. Pietersen E, Ignatius E, Streicher EM, et al. Long-term outcomes of patients with extensively drug-resistant tuberculosis in South Africa: a cohort study. *Lancet*. 2014;383:1230-1239.

REVIEW QUESTIONS

1. The following is true of tuberculosis (TB) in the United States:

A. Most cases are in the foreign born.
B. Most cases are in the HIV infected.
C. The elderly are not at risk.
D. The period of greatest risk of progression is 5 years and longer after infection.
E. Outbreaks are a major source of multidrug-resistant (MDR) TB cases.

Answer: A Foreign-born individuals account for over 60% of U.S. TB cases. The greatest risk of progressing to disease is in the first and second years after infection. Outbreaks have been limited by application of infection control measures.

2. Advantages of interferon gamma release assays (IGRAs) such as Quantiferon-Gold for the diagnosis of latent TB infection include all the following ***except***:

A. More specific than tuberculin skin test (TST)
B. Reproducible results in health care workers
C. Does not require return visits
D. End points are objective.
E. Antigens are not found in BCG.

Answer: B There is lack of reproducibility in serial testing, particularly if the "positive" is close to end point.

3. The advantages of GenXpert MTB/RIF for the diagnosis of TB include all the following ***except***:

A. Useful in monitoring therapy
B. More sensitive than smear
C. Does not require proficiency in molecular technology
D. More rapid than culture
E. Provides rifampin susceptibility

Answer: A Not useful in monitoring therapy; the presence of *Mycobacterium tuberculosis* DNA in sputum does not indicate that there are viable organisms.

4. Which of the following is true concerning treatment of MDR TB:

A. The strategy of adding a single drug to retreatment regimens is cost-effective.
B. Empirical therapy is as good as tailoring regimen to drug susceptibility.
C. Acquisition of additional drug resistance is unlikely.
D. Standard duration of therapy (6 months) is adequate.
E. New drugs with potent activity now are available.

Answer: E Bedaquiline and delamanid have been approved by the FDA. Adding a single drug to a retreatment regimen is inadequate in one fourth of cases.

5. Which of the following is true regarding treatment of latent TB infection (LTBI)?

A. There is a prolonged effect in HIV-infected persons.
B. A 12-dose regimen of rifapentine-isoniazid is effective.
C. Treatment requires both a positive TST and IGRA.
D. There is no need to monitor for adverse events.
E. Screening in HIV-infected patients in resource-limited settings requires chest x-ray and culture.

Answer: B A 12-dose course of rifapentine-isoniazid is effective. The efficacy of isoniazid preventive therapy in HIV-infected patients in high-transmission settings wanes after 6 to 12 months.

325 THE NONTUBERCULOUS MYCOBACTERIA

STEVEN M. HOLLAND

DEFINITION

Nontuberculous mycobacteria generally include the growing number of mycobacteria other than *Mycobacterium tuberculosis* and its close relatives (Chapter 324) and *Mycobacterium leprae* (Chapter 326). Other names that have been used include *atypical mycobacteria, mycobacteria other than tuberculosis,* and *environmental mycobacteria*. The number of nontuberculous mycobacteria is growing rapidly as a result of the advent of DNA sequence typing for determining criteria for speciation. Accordingly, the number of species of nontuberculous mycobacteria has increased to almost 170 and will continue to increase for the near future.

The Pathogens

Identification of any mycobacterium requires that the appropriate tests be thought of ahead of time and be performed, because routine microbiological testing does not identify mycobacteria. Nontuberculous mycobacteria are typically first detected on acid-fast smears of sputum or other body fluids. When levels of organisms are high, mycobacteria may be seen on Gram stain

TABLE 325-1 COMMON NONTUBERCULOUS MYCOBACTERIA

ORGANISM	DISEASE
RAPIDLY GROWING NONTUBERCULOUS MYCOBACTERIA	
M. abscessus	Lung, disseminated, lymph node
M. chelonae	Skin
M. fortuitum	Line infections, lung
M. smegmatis	Almost never associated with disease
SLOWLY GROWING NONTUBERCULOUS MYCOBACTERIA	
M. avium complex	Lung, disseminated, lymph node
M. kansasii	Lung
M. marinum	Skin, tendons (fish tank granuloma)
M. xenopi	Lung
M. simiae	Lung
M. szulgai	Lung
M. malmoense	Lung
M. scrofulaceum	Lymph node
M. haemophilum	Disseminated, skin
M. genavense	Disseminated
M. ulcerans	Skin (Buruli ulcer; toxin producing)
M. neoarum	Disseminated
M. celatum	Disseminated
M. gordonae	Almost never causes disease
M. terrae complex	Disseminated

M = *Mycobacterium*.

as gram-positive beaded rods, but this finding is unreliable. The first step in identification is to request the appropriate smear (acid-fast or fluorochrome) and culture. Nontuberculous mycobacteria are broadly differentiated into rapidly growing (<7 days) and slowly growing (>7 days) forms. *M. tuberculosis*, by contrast, typically takes 2 or more weeks to grow. Formation of pigment in light (photochromogens) or dark (scotochromogens) and lack of pigment (nonchromogens) have also been used to help categorize nontuberculous mycobacteria. Current diagnostics use biochemical, nucleic acid, or cell wall composition on high-performance liquid chromatography for speciation (Table 325-1). For purposes of diagnosis, prognosis, and therapy, identification of nontuberculous mycobacteria should be taken to the species level.

EPIDEMIOLOGY

As a group, the nontuberculous mycobacteria are ubiquitous in soil and water and are often found in certain animals, but they rarely cause disease in humans. There are very few instances of human-to-human transmission of nontuberculous mycobacteria. However, *Mycobacterium massiliense* has caused outbreaks of infection in cystic fibrosis centers. Because these infections are not reported to health agencies and their identification is sometimes problematic, reliable data on incidence and prevalence are lacking. In the United States, however, isolates of nontuberculous mycobacteria have exceeded those for *M. tuberculosis* for many years. In patients with cystic fibrosis (Chapter 89), for example, rates of clinical nontuberculous mycobacterial infection range up to 40%, but even more patients harbor the organism. Differentiating active disease from commensal harboring of the organism remains problematic. Other patient groups, such as those with bronchiectasis, also have elevated rates of nontuberculous mycobacterial infection, but the rates are undefined.[1] The bulk of nontuberculous mycobacterial disease in North America is due to *Mycobacterium kansasii*, *Mycobacterium avium* complex (MAC), and *Mycobacterium abscessus*.

PATHOBIOLOGY

Because exposure is essentially universal and disease is rare, normal host defenses against nontuberculous mycobacteria must be highly effective. Therefore, otherwise healthy individuals in whom disease develops must have specific susceptibility factors that permit these infections to become established, multiply, and cause disease.

With the advent of human immunodeficiency virus (HIV) infection, $CD4^+$ T lymphocytes were identified as key effectors against nontuberculous mycobacteria. Much of the genetic basis of susceptibility to disseminated nontuberculous mycobacterial infection outside HIV infection has been found to be due to specific mutations in the interferon (IFN)-γ/interleukin (IL)-12 synthesis and response pathways. However, only about 70% of disseminated cases unassociated with HIV infection have a genetic diagnosis, and genetic causes of predisposition to nontuberculous mycobacterial lung disease are still very few.

Mycobacteria are typically phagocytosed by macrophages, which respond with the production of IL-12, a heterodimer composed of p35 and p40 moieties that together constitute IL-12p70 (Fig. 325-1). IL-12 activates T lymphocytes and natural killer (NK) cells through binding to its receptor (composed of IL-12Rβ1 and IL-12Rβ2/IL-23R) and results in phosphorylation of STAT4 (signal transducer and activator of transcription 4). IL-12 stimulation leads to production and secretion of IFN-γ, which activates neutrophils and macrophages to produce reactive oxidants and increase major histocompatibility complex display and Fc receptors. IFN-γ signals through its receptor (composed of IFN-γR1 and IFN-γR2), thereby leading to phosphorylation of STAT1, which in turn regulates IFN-γ–responsive genes such as those for the production of IL-12 and tumor necrosis factor (TNF)-α. Therefore, the positive feedback loop between IFN-γ and IL-12/IL-23 is pivotal in the immune response to mycobacteria and other intracellular infections (most importantly *Salmonella*, *Histoplasma*, *Coccidioides*). The advent of potent TNF-α inhibitors such as infliximab, adalimumab, certolizumab, etanercept, and golimumab (Chapter 36) has provided the ability to neutralize this critical cytokine, which has occasionally resulted in mycobacterial and fungal infections.

CLINICAL MANIFESTATIONS

Disseminated Disease

Disseminated nontuberculous mycobacterial disease secondary to MAC used to occur commonly in the setting of advanced acquired immunodeficiency syndrome (AIDS) but is now uncommon in North America because of MAC prophylaxis and improved treatment of HIV infection. The portal of entry was the bowel, with spread to bone marrow and the blood stream. Rapidly growing mycobacteria such as *Mycobacterium fortuitum* sometimes infect deep indwelling lines. The severe disseminated infection seen with immune defects is typically associated with malaise, fever, and weight loss, and it is often accompanied by organomegaly and lymphadenopathy. Disseminated (two or more organ) involvement in a child without an underlying iatrogenic cause should always prompt an investigation of the IFN-γ/IL-12 pathway.[2] Nontuberculous mycobacterial osteomyelitis is especially common with dominant negative mutations in IFN-γR1. A male with conical or peg teeth or an abnormal hair pattern and disseminated nontuberculous mycobacterial infection should be evaluated for defects in the pathway that activates nuclear factor (NF)κB. Some patients with disseminated rapidly growing infections (predominantly *M. abscessus*) have high-titer autoantibodies to IFN-γ.

Pulmonary Disease

Lung disease caused by nontuberculous mycobacteria is by far the most common form of the infection in North America. Predisposing factors include underlying lung disease, such as bronchiectasis (Chapter 90), pneumoconiosis (Chapter 93), chronic obstructive pulmonary disease (Chapter 88), primary ciliary dyskinesia, and cystic fibrosis.[3] The manifestations of *M. kansasii* infection can be very similar to those of tuberculosis (Chapter 324) and consist of hemoptysis, chest pain, and cavitary lung disease. MAC infection most commonly occurs in women in their sixth or seventh decade who have had months to years of nagging intermittent cough and fatigue, with or without sputum production or chest pain. Bronchiectasis and nontuberculous mycobacterial infection often coexist and progress in tandem, thus making causality difficult to determine. When compared with male smokers with upper lobe cavitary disease, who tend to carry the very same single strain of MAC indefinitely, nonsmoking females with nodular bronchiectasis tend to have several strains simultaneously and change them over the course of their disease process. Patients with pulmonary alveolar proteinosis (Chapter 91) are prone to pulmonary nontuberculous mycobacterial and *Nocardia* infections, likely reflecting their association with anti-GM-CSF autoantibodies and impaired alveolar macrophage function. Esophageal motility disorders such as achalasia (Chapter 138) have been associated with pulmonary disease, especially that caused by rapidly growing nontuberculous mycobacteria such as *M. abscessus*. It is important to note that lung disease rarely disseminates, illustrating that the defects leading to isolated pulmonary involvement are specific to the respiratory epithelium, whereas those defects leading to disseminated disease affect immune cells.

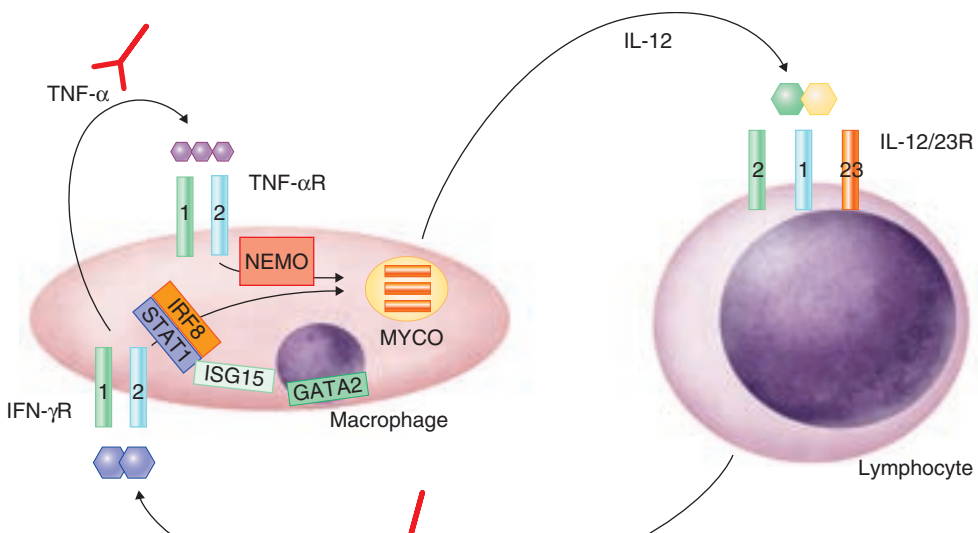

FIGURE 325-1. Schematization of the critical cytokine interactions between infected macrophages and T and natural killer lymphocytes. Organisms (MYCO) infect macrophages, which release heterodimeric interleukin (IL)-12. This acts on the IL-12/23 receptor complex and leads to the production of homodimeric interferon (IFN)-γ. IFN-γ acts on its receptor to stimulate the production of tumor necrosis factor (TNF)-α and kill intracellular organisms such as mycobacteria, salmonellae, and some fungi. Homotrimeric TNF-α acts on its own receptor and also contributes to killing of intracellular organisms. Both IFN-γ and TNF-α lead to upregulation of IL-12. TNF-α–blocking antibodies work either by blocking the ligand (infliximab, adalimumab, certolizumab) or by providing soluble receptor (etanercept). Mutations in both chains of IFN-γR, IL-12p40, and IL-12Rβ1, IL-12Rβ2 and signal elements for IFN-γR and TNF-αR have been identified through their predisposition to mycobacterial infections. IRF8 = interferon regulatory factor 8; ISG = interferon stimulated gene; NEMO = nuclear factor kappa-B essential modulator; STAT1 = signed transducer and activator of transcription 1.

Therefore, evaluation of isolated lung disease should focus on respiratory tract causes.[4]

Cervical Lymph Nodes

Isolated cervical lymphadenopathy, most frequently caused by MAC, is the most common form of nontuberculous mycobacterial infection in young children in North America. The organism is generally MAC, but other nontuberculous mycobacteria can also cause disease. The cervical swelling is often firm and relatively painless with a paucity of systemic signs. Because the differential diagnosis of painless adenopathy includes malignancy, many of these infections are incidentally diagnosed at biopsy. Local fistulas usually resolve completely with resection or antibiotic therapy or both.

Skin and Soft Tissue Disease

Mycobacterium marinum causes skin infections, usually papules or ulcers, associated with water exposure and is known as "fish tank granuloma." Numerous outbreaks of skin infections caused by rapidly growing mycobacteria (especially *M. abscessus*, *M. fortuitum*, and *Mycobacterium chelonae*) have been due to skin contamination from instruments used for surgical procedures (especially cosmetic surgery), injections, and other procedures.[5] These infections are typically accompanied by painful, erythematous, draining subcutaneous nodules, usually without associated fever or systemic symptoms.

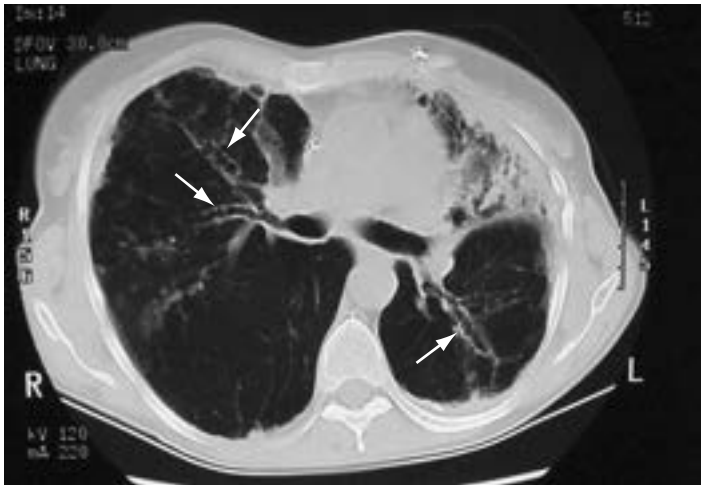

FIGURE 325-2. Chest computed tomography in a patient with severe pulmonary *Mycobacterium abscessus* infection. Arrows indicate bronchiectasis. Note the extensive left upper lobe destruction and diffuse pleural reaction. In addition, the left lung is smaller than the right as a result of extensive loss of lung parenchyma.

DIAGNOSIS

With the continued decline in cases of tuberculosis, nontuberculous mycobacteria are now the most common mycobacteria isolated from humans in North America. The conventional tuberculin skin test (purified protein derivative [PPD]) evokes a cell-mediated response to secreted mycobacterial antigens. Unfortunately, the PPD test does not differentiate well between nontuberculous mycobacterial and tuberculosis infection, although large PPD reactions (>15 mm) more commonly signify tuberculosis. With the progressive decline in active tuberculosis in the United States, nontuberculous mycobacteria are likely to account for significant proportions of PPD reactivity. Newer IFN-γ release assays (IGRAs) incubate blood with relatively tuberculosis-specific recombinant proteins and elicit T-cell secretion of IFN-γ, thereby helping to clarify whether PPD reactivity is due to tuberculosis.

Isolation of nontuberculous mycobacteria from blood specimens is clear evidence of disease. However, because the slow-growing nontuberculous mycobacteria typically do not grow well in routine blood culture media, the diagnosis must be suspected. Isolation of nontuberculous mycobacteria from a biopsy specimen is strong evidence of infection, but cases of laboratory contamination do occur. Identification of organisms on stained sections of biopsy material confirms the authenticity of the culture. Some unusual nontuberculous mycobacteria require lower incubation temperatures or special additives for growth (e.g., *Mycobacterium hemophilum*).

The radiographic appearance of nontuberculous mycobacterial disease in the lung ranges from normal to nodules, bronchiectasis, air space disease, and extensive cavity formation, similar to that seen in tuberculosis (Fig. 325-2). Isolation of nontuberculous mycobacteria from respiratory samples presents special problems in both sensitivity and specificity. *Mycobacterium gordonae* is often recovered from respiratory samples and is almost never thought to be a real pathogen. Many patients, especially those with bronchiectasis, will occasionally have nontuberculous mycobacteria recovered from sputum culture without such mycobacteria being seen on smear. Specific criteria for definitive diagnosis of nontuberculous mycobacterial lung disease exist for MAC, *M. abscessus*, and *M. kansasii*, but they are probably good guidelines for other nontuberculous mycobacteria as well. A positive diagnosis requires

that two of three sputum samples grow nontuberculous mycobacteria, regardless of smear findings; a positive bronchoscopic alveolar sample, regardless of smear findings; or a biopsy specimen of pulmonary parenchyma with granulomatous inflammation or mycobacteria found on section and nontuberculous mycobacteria on culture.

Once isolated, identification of nontuberculous mycobacteria is important because it will determine the broad class of antimycobacterial therapy to be used. Many laboratories now use DNA probes to identify MAC, *M. gordonae*, and *M. kansasii*.[6] Drug susceptibility testing is of limited and largely unproven value, although clarithromycin susceptibility testing for MAC and rifampin susceptibility testing for *M. kansasii* are indicated. Initial isolates of MAC that have not been exposed to macrolides are almost always susceptible to macrolides. Any nontuberculous mycobacteria that have resisted a course of antimicrobials should probably be tested for antibiotic susceptibility as well.

PREVENTION

Prophylaxis of MAC disease in patients infected with HIV is started when the $CD4^+$ T-lymphocyte count is less than 50 cells/μL. Azithromycin 1200 mg weekly, clarithromycin 1000 mg daily, and rifabutin 300 mg daily are effective.

TREATMENT

Rx

It is rarely an emergency to initiate treatment of nontuberculous mycobacterial infections, which are relatively slow-growing chronic infections that evolve over a period of weeks to years, not hours to days. Therefore, empirical therapy is not usually needed, and identification of the species is advisable before starting complex, often poorly tolerated and potentially toxic, regimens. Similar to the case with tuberculosis, single-drug therapy is almost always associated with the emergence of antimicrobial resistance and is strongly discouraged.

MAC infection frequently requires complex multidrug therapy, the foundation of which is a macrolide (clarithromycin or azithromycin), ethambutol, and a rifamycin (rifampin or rifabutin). For disseminated nontuberculous mycobacterial disease in HIV-infected patients, the use of rifamycins poses special problems of drug interactions with protease inhibitors. For pulmonary MAC disease, three-times-weekly administration of drugs has been used successfully. The duration of therapy is prolonged, generally for 12 months after culture conversion and typically for a total of at least 18 months. Other drugs with activity against MAC include aminoglycosides, fluoroquinolones, and clofazimine.

M. kansasii lung disease is similar to tuberculosis in many ways and is also effectively treated with isoniazid (300 mg/day), rifampin (600 mg/day), and ethambutol (15 mg/kg/day). Treatment should continue until cultures have been negative for at least 1 year. Other drugs with very high activity against *M. kansasii* include clarithromycin, fluoroquinolones, and aminoglycosides.

Rapidly growing mycobacteria pose special therapeutic problems. Extrapulmonary disease in an immunocompetent host is usually due to inoculation (e.g., surgery, injection, trauma) or line infection and is often treated successfully with a macrolide and another drug (based on in vitro susceptibility), along with removal of the offending focus. By comparison, pulmonary disease, especially that caused by *M. abscessus*, is extremely difficult to eradicate, although repeated courses of treatment are usually effective in reducing the infectious burden and symptoms. Therapy generally includes a macrolide along with an intravenous agent such as amikacin, a carbapenem, cefoxitin, or tigecycline.[7] Other oral agents used according to in vitro susceptibility testing and tolerance include fluoroquinolones, doxycycline, and linezolid. Inhaled amikacin may be an option for treatment-refractory pulmonary infections.[8]

Treatment of the other nontuberculous mycobacteria is less well defined, but macrolides and aminoglycosides are usually effective, with other agents added as indicated. Expert consultation is strongly encouraged for difficult or unusual nontuberculous mycobacterial infections.

PROGNOSIS

The effect of nontuberculous mycobacterial infection on longevity is closely tied to the underlying condition (e.g., IFN-γ/IL-12 pathway defect, cystic fibrosis). With no or inadequate treatment, symptoms are intrusive, and the infections can lead to fatal complications, including overwhelming infection or severe lung destruction.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Adjemian J, Olivier KN, Seitz AE, et al. Prevalence of nontuberculous mycobacterial lung disease in U.S. Medicare beneficiaries. *Am J Respir Crit Care Med*. 2012;185:881-886.
2. Rosenzweig SD, Holland SM. Recent insights into the pathobiology of innate immune deficiencies. *Curr Allergy Asthma Rep*. 2011;11:369-377.
3. Alvarez-Uria G. Lung disease caused by nontuberculous mycobacteria. *Curr Opin Pulm Med*. 2010; 16:251-256.
4. Fowler CJ, Olivier KN, Leung JM, et al. Abnormal nasal nitric oxide production, ciliary beat frequency, and Toll-like receptor response in pulmonary nontuberculous mycobacterial disease epithelium. *Am J Respir Crit Care Med*. 2013;187:1374-1381.
5. Atkins BL, Gottlieb T. Skin and soft tissue infections caused by nontuberculous mycobacteria. *Curr Opin Infect Dis*. 2014;27:137-145.
6. Somoskovi A, Salfinger M. Nontuberculous mycobacteria in respiratory infections: advances in diagnosis and identification. *Clin Lab Med*. 2014;34:271-295.
7. Johnson MM, Odell JA. Nontuberculous mycobacterial pulmonary infections. *J Thorac Dis*. 2014;6:210-220.
8. Olivier KN, Shaw PA, Glaser TS, et al. Inhaled amikacin for treatment of refractory pulmonary nontuberculous mycobacterial disease. *Ann Am Thorac Soc*. 2014;11:30-35.

REVIEW QUESTIONS

1. Initial *Mycobacterium avium* complex (MAC) susceptibility testing is recommended and validated for which of the following?

A. Rifamycins
B. Macrolides
C. Aminoglycosides
D. Ethambutol
E. All of the above

Answer: B The only antibiotic class for which there is evidence for in vitro testing to predict in vivo value is macrolides. The other agents have limited predictive value in in vitro testing for MAC. Therefore, only macrolide testing is recommended and interpretable with break points.

2. Pulmonary MAC infection is associated with which of the following underlying conditions?

A. Cystic fibrosis
B. Primary ciliary dyskinesia
C. Bronchiectasis
D. Pneumoconiosis
E. All of the above

Answer: E Isolated pulmonary MAC infection has been clearly associated with defects in respiratory ciliary function and channel defects affecting the respiratory epithelium.

3. Disseminated MAC infection is associated with which of the following conditions?

A. Mutations in the interleukin (IL)-12 receptor
B. Mutations in the interferon (IFN)-γ receptor
C. Mutations in *STAT1*
D. Tumor necrosis factor (TNF)-inhibiting antibodies
E. All of the above

Answer: E The IL-12/IFN-γ pathway is critical for the control of intracellular organisms including nontuberculous mycobacteria. IFN-γ also induces TNF-α, linking these two pathways together.

4. Nontuberculous mycobacteria (NTM) are relatively ubiquitous in the environment. Which organism is the most common cause of nontuberculous mycobacterial infection in humans?

A. *Mycobacterium gordonae*
B. *Mycobacterium smegmatis*
C. MAC
D. *Mycobacterium abscessus* complex
E. *Mycobacterium kansasii*

Answer: C MAC is a compilation of *M. avium*, *M. intracellulare*, and *M. avium* X-cluster. The leading cause of disseminated NTM infection is *M. avium*, whereas the majority of isolated pulmonary NTM is due to *M. intracellulare*. *M. avium* X-cluster is a relatively infrequent cause of disease that is usually extrapulmonary.

5. Nontuberculous mycobacterial osteomyelitis is rare outside of genetic immunodeficiency. Which immunodeficiency is it most associated with?

A. Chronic granulomatous disease
B. Anti–TNF antibody therapy
C. Recessive deficiency of IFN-γR
D. Dominant negative deficiency of IFN-γR
E. Advanced HIV

Answer: D Osteomyelitis due to NTM and bacille Calmette Guérin (BCG) is strongly associated with the autosomal dominant form of IFN-γ receptor 1 deficiency. Isolated osteomyelitis due to NTM is much less common in the recessive forms of IFN-γ receptor 1 deficiency and very unusual in other genetic defects.

326

LEPROSY (HANSEN DISEASE)

JOEL D. ERNST

DEFINITION

Leprosy (Hansen disease) is a chronic infection caused by *Mycobacterium leprae*, an acid-fast slowly growing bacterium that cannot yet be cultured in vitro. Leprosy is found worldwide, although three countries of high prevalence (India, Brazil, and Indonesia) currently account for more than 80% of reported cases.[1] The primary manifestations of infection with *M. leprae* occur in the skin and peripheral nerves. The skin lesions of leprosy are classically hypopigmented, hypoesthetic or anesthetic, and nonpruritic. Peripheral nerves can be damaged by direct infection with *M. leprae* or by the immune response to the infection; the result is loss of sensation and motor function. Additional morbidity is due to the peripheral nerve dysfunction, including painless traumatic and burn injuries, secondary bacterial infections, and muscle atrophy and contractures. Leprosy per se is not a cause of death, but the debility associated with leprosy contributes to the severity of poverty and the likelihood of death from malnutrition or other infections. Despite the low transmissibility of *M. leprae* and the ability of multiple-drug therapy to cure leprosy, it remains a stigmatized disease that can pose a challenge to diagnosis and therapy.

The Pathogen

M. leprae is an acid-fast bacillus that contains a mycolic acid–rich cell wall and a single membrane. Despite nearly 150 years of effort, *M. leprae* remains uncultivatible in vitro. For biochemical and structural characterization, *M. leprae* can be grown in large quantities in nine-banded armadillos (*Dasypus novemcinctus*), and inoculation of the footpads of athymic mice allows semiquantitation of viable bacilli.

The genome of *M. leprae* consists of 3,268,203 base pairs (bp), compared with the *Mycobacterium tuberculosis* genome of 4,411,529 bp, and the number of expressed genes of *M. leprae* is approximately 60% fewer than that of *M. tuberculosis*. Because *M. leprae* and *M. tuberculosis* probably evolved from a common mycobacterial ancestor, *M. leprae* appears to have lost approximately 2000 genes since this divergence, leaving it dependent on specialized ecologic niches for its survival. Among the genes lacking in *M. leprae* are those of the *mbt* complex, whose products are involved in bacterial acquisition of iron. *M. leprae* also lacks many of the genes for lipid biosynthesis and modification that are characteristic of *M. tuberculosis*. The genome sequence has also allowed a directed approach to identification of 16 strains of *M. leprae* from geographically diverse sources. Genome sequence analyses have revealed that leprosy was introduced to the United States from Europe,[2] and also facilitated discovery that armadillos in the southeastern United States have the same unique strain as the U.S.-born leprosy patients have in the same region, suggesting that armadillos may be a reservoir for the pathogen in that region.

EPIDEMIOLOGY

Leprosy is found worldwide, although endemic leprosy is absent from northern Europe, where it was present in epidemic form as recently as the 19th century. The global prevalence of leprosy is about 180,000 known cases, and the current incidence is about 220,000. By definition of the World Health Organization (WHO) Strategic Plan for the Elimination of Leprosy, a newly diagnosed patient who has been treated with multidrug therapy is removed from the prevalence registry, which explains the lower prevalence than incidence of this chronic infection. Since initiation of the WHO Strategic Plan (whose goal is to eliminate leprosy as a public health problem, i.e., a prevalence of < 1 in 10,000 in all regions), an estimated 14 million people have been cured of leprosy.

The success of multidrug therapy notwithstanding, leprosy remains a public health problem in Brazil, Indonesia, Philippines, Democratic Republic of Congo, India, Madagascar, Mozambique, Nepal, and the United Republic of Tanzania. India and Brazil currently have the largest number of cases. Although domestic transmission of leprosy is extremely rare in the United States, 82 cases of leprosy were diagnosed in 2011, including cases in

immigrants from India, Brazil, the Philippines, the Dominican Republic, and Mexico. Because leprosy is not highly transmissible, it is not considered a disease of travelers other than immigrants.

Inability to culture *M. leprae* in vitro has been a major hindrance to understanding the modes of transmission and reservoirs of the organism. Observational studies reveal a low frequency of leprosy in casual travelers or temporary residents of high-incidence regions, thus indicating that *M. leprae* is not highly transmissible. Even in areas of high incidence, clusters of leprosy are rare outside families or others with prolonged close contact. It is believed that transmission of *M. leprae* commonly occurs through the respiratory route, because nasal secretions of people with lepromatous leprosy may contain 10^7 viable bacilli per milliliter. In addition, transmission of *M. leprae* is thought to occur through contact with contaminated soil, although soil has not been found to be a reservoir for the bacilli.

PATHOBIOLOGY

Immunology

There is an inverse correlation between the number of lymphocytes and the number of acid-fast bacteria present in skin lesions. Tuberculoid lesions have abundant lymphocytes, well-formed granulomas, and few bacteria (hence this form of leprosy is also termed *paucibacillary*). In contrast, lepromatous lesions have very few lymphocytes, poorly organized or no granulomas, and large numbers of bacteria (also termed *multibacillary leprosy*). Between these polar extremes are intermediate forms that represent a continuum of the histopathologic and bacteriologic findings, termed *borderline tuberculoid*, *borderline*, and *borderline lepromatous* (Fig. 326-1). In addition to correlating with the number of bacteria in individual lesions, the polar forms of leprosy correlate with the total number of skin lesions in an individual patient: tuberculoid leprosy exhibits few (<five) lesions, whereas lepromatous leprosy is characterized by multiple lesions (≥five, up to hundreds).

Leprosy provides a paradigm for the effect of the cellular immune response to a bacterial pathogen on the clinical manifestations of the infection.[3] Individuals in whom a T helper 1 (T_H1) immune response (characterized by antigen-specific T cells that produce interferon [IFN]-γ, lymphotoxin [LTA], or interleukin [IL]-2 and no IL-4 or IL-5) develops to *M. leprae* exhibit few skin lesions and few bacteria within the lesions (paucibacillary leprosy). In contrast, persons in whom a T_H2 immune response develops (T cells that produce little IFN-γ, lymphotoxin, or IL-2, but produce IL-4, IL-5, and IL-13) have larger numbers of skin lesions and large numbers of bacteria within lesions (multibacillary leprosy). The roles of other T-cell subsets, such as T_H17 or T-regulatory cells, remain to be defined in leprosy. The primary determinant of the differential immune response to *M. leprae* is incompletely understood, although substantial evidence indicates that host genetic polymorphisms contribute to the likelihood of paucibacillary versus multibacillary leprosy.

Pathogenesis of Nerve Damage

Peripheral nerve damage, the most important consequence of infection with *M. leprae*, occurs in all forms of leprosy and underlies the complications of the infection. *M. leprae* invades Schwann cells, the glial cells of peripheral nerves. Schwann cells form a functional unit with peripheral nerve axons and are surrounded by laminin-2, a neural-specific extracellular matrix protein. The G domain of laminin-2 can bind simultaneously to *M. leprae* and to the Schwann cell laminin receptor α-dystroglycan, which promotes binding of *M. leprae* to Schwann cells by use of laminin-2 as a bridging molecule. Laminin-2 interacts with two distinct molecules on the surface of *M. leprae*, a 21-kD protein and phenolic glycolipid-1 (PGL-1); either can mediate internalization by Schwann cells. Once *M. leprae* is bound and internalized by Schwann cells, it can cause direct demyelination of peripheral nerves in the absence of an immune response, apparently by signaling through ErbB2 and Erk1/2. *M. leprae*–mediated demyelination occurs without early cell death or toxicity, although Schwann cells and neurons can die by apoptosis later after infection. In addition, dead *M. leprae* or PGL-1 shed from live or dying *M. leprae* can mediate peripheral nerve demyelination and may contribute to the ongoing nerve damage that can follow initiation of chemotherapy. In addition to PGL-1–mediated demyelination, an *M. leprae* 19-kD lipoprotein can mediate Schwann cell apoptosis in vitro, and apoptotic Schwann cells can be found in human leprosy lesions. These mechanisms may be responsible for the nerve damage in multibacillary leprosy.

In addition to direct damage to peripheral nerves by *M. leprae*, the immune response in leprosy also contributes to nerve damage, especially in paucibacillary (tuberculoid) leprosy, in which the bacteria or PGL-1 or both are present in insufficient quantity to cause widespread nerve damage, and in reversal reactions, in which inflammation is particularly prominent.[4] Several distinct immunologic mechanisms probably contribute to the nerve damage in leprosy. Pro-inflammatory cytokines such as tumor necrosis factor (TNF)-α, IL-1β, and IFN-γ are especially prominent in lesions during reversal reactions, when irreversible nerve damage can occur. Because these molecules can contribute to inflammatory tissue damage and can induce apoptosis of Schwann cells in vitro, it is likely these mediators play an active role in nerve damage. Reversal reactions are also characterized by an increase in the number of $CD4^+$ T lymphocytes in lesions, and at least some of these $CD4^+$ cells exhibit a cytotoxic phenotype and kill *M. leprae*–infected Schwann cells through antigen- and major histocompatibility complex class II–dependent secretion of cytotoxic granule contents. Whether these mechanisms of nerve damage occur in chronic tuberculoid leprosy is not established, but similar cytokines and T lymphocytes are found in tuberculoid lesions.

Genetics

Cases of leprosy cluster in families, partly because of shared environments and similar exposure but also probably because of genetic determinants of susceptibility. Genetic loci whose allelic variants are related to altered susceptibility to infection with *M. leprae* include *PARK2*, *NRAMP1*, *TNF*, *TLR1*, *FCN2*, *LTA*, *NOD2*, *RIPK2*, *IL23R*, and *RAB32*. Parkin, the product of PARK2, is an E3 ubiquitin ligase that promotes autophagy and killing of intracellular bacteria, including mycobacteria. *NRAMP1*, *NOD2*, *RIPK2*, and *TNF* are expressed by macrophages, and quantitative or temporal differences in their expression may account for differences in innate susceptibility to infection. TLR1 and ficolin-2 are pattern recognition molecules that recognize bacterial lipopeptides and polysaccharides, respectively, whereas NOD2 and its downstream kinase RIPK2 are involved in responses to bacterial peptidoglycan. The overlap in susceptibility genes for leprosy and for Crohn's disease suggests one or more shared mechanisms of pathogenesis.[5] Genes with polymorphisms associated with a predisposition to distinct clinical forms of leprosy (i.e., lepromatous vs. tuberculoid) include HLA-DRB1*1501 and HLA-DRB1*1502, *NRAMP1*, *TNF*, *IL-10*, and *TAP2*.

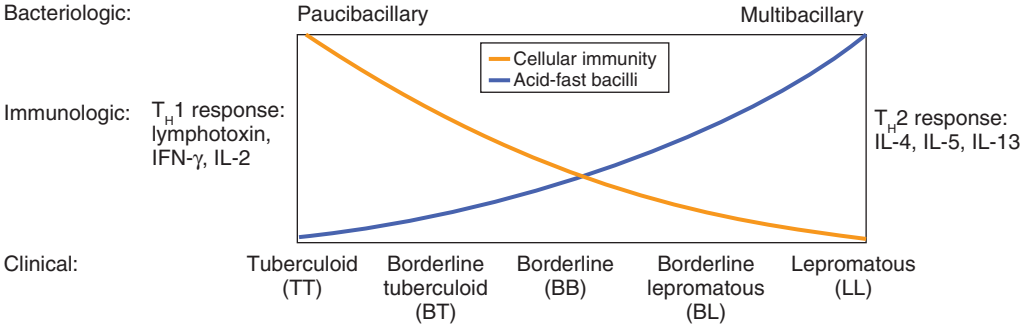

FIGURE 326-1. Bacteriologic, immunologic, and clinical spectrum of leprosy. Tuberculoid (paucibacillary) leprosy is characterized by a T_H1 immune response and few or no detectable bacilli in biopsy specimens of skin lesions or smears of skin slits. At the opposite pole of the spectrum, lepromatous (multibacillary) leprosy is accompanied by a T_H2 immune response, numerous skin lesions, and numerous acid-fast bacilli on skin biopsy specimens or smears. The intermediate forms can be classified according to their resemblance to tuberculoid or lepromatous leprosy. IFN = interferon; IL = interleukin.

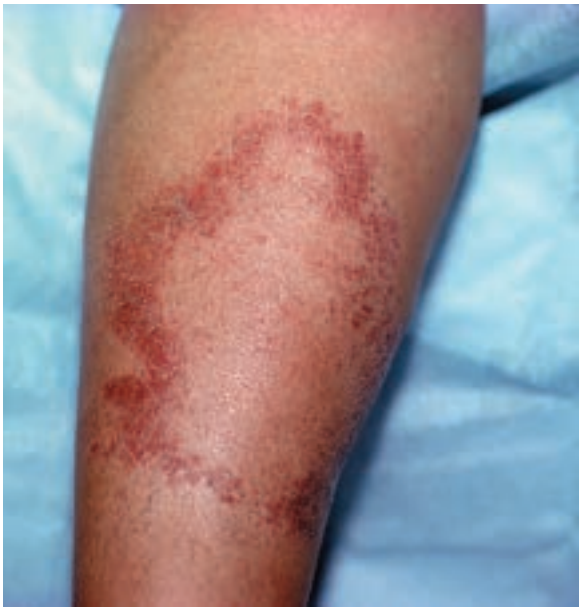

FIGURE 326-2. Tuberculoid leprosy. A single large lesion with irregular, raised, erythematous borders and a depressed, hypopigmented center is shown. (From Hansen's disease [leprosy]. In: James WD, Berger TG, Elston DM, eds. *Andrews' Diseases of the Skin.* 10th ed. Philadelphia: Elsevier; 2006.)

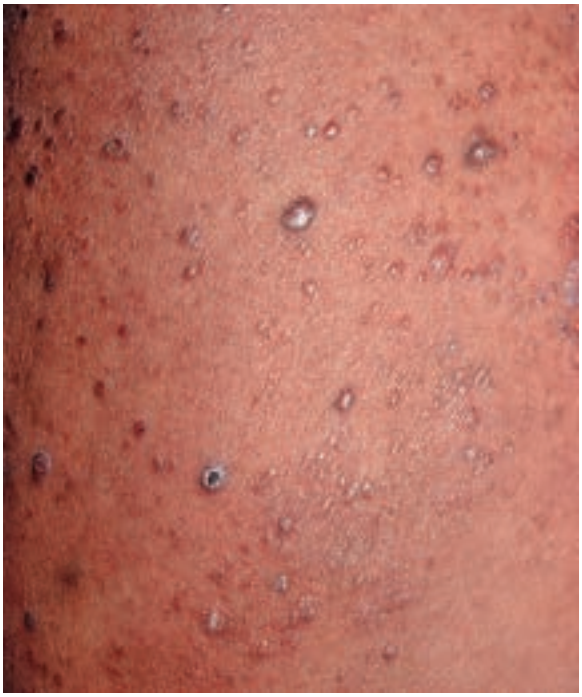

FIGURE 326-3. Lepromatous leprosy. Numerous papules and nodules are apparent. Lepromatous macules have ill-defined borders, with normal sensation usually maintained. (From Hansen's disease [leprosy]. In: James WD, Berger TG, Elston DM, eds. *Andrews' Diseases of the Skin.* 10th ed. Philadelphia: Elsevier; 2006.)

CLINICAL MANIFESTATIONS

The most common manifestations of leprosy involve the skin and peripheral nerves and are determined by the polarity of the disease: paucibacillary (tuberculoid) or multibacillary (lepromatous). The onset of leprosy is usually insidious. Depending on the form of leprosy, numbness may be an initial complaint or finding, and skin lesions may become apparent only months or years later. The classification of leprosy as tuberculoid, lepromatous, or one of the borderline forms is based on the combination of clinical examination, the number of bacteria seen on skin slit smears or skin biopsy specimens, and the histologic appearance. Because the nature of the potential complications and the specific course of chemotherapy are determined by the form of leprosy, accurate diagnosis and classification are essential.

Tuberculoid Leprosy

Tuberculoid leprosy is characterized by the presence of fewer than five skin lesions, which are typically hypopigmented or erythematous macules with raised erythematous borders and an atrophic center (Fig. 326-2). The skin lesions in tuberculoid leprosy are usually hypoesthetic or anesthetic; when multiple lesions are present, their distribution is asymmetrical. The skin lesions may be large and are most commonly found on the face, trunk, or extremities; however, they are not found in the axillae, groin, perineum, or on the scalp, presumably because of the preference of *M. leprae* for lower temperatures.

Local peripheral nerve involvement is common in tuberculoid leprosy and is asymmetrical. In addition to hypoesthesia or anesthesia of the skin lesions, nerve involvement in tuberculoid leprosy is manifested as enlargement or tenderness (or both) of the peripheral nerves that serve the region of the skin lesions. Superficial nerves such as the ulnar, superficial peroneal, or greater auricular nerves may be visibly enlarged, depending on the location of the skin lesions. Functional complications of nerve involvement, such as muscle atrophy and contractures, may be present at the time of diagnosis of tuberculoid leprosy. Tuberculoid leprosy is a stable form; it does not convert to borderline or lepromatous forms.

Lepromatous Leprosy

Lepromatous leprosy is characterized by multiple skin lesions that are smaller than those observed in tuberculoid leprosy (Fig. 326-3). Although the sites of skin lesions are similar to those of tuberculoid leprosy, the multiple lesions of lepromatous leprosy are often symmetrically distributed. Lepromatous macules may have poorly defined borders and no loss of sensation; local nerve enlargement is not characteristic. In addition to macules, lepromatous skin lesions may be nodules or plaques, or they may diffusely infiltrate the skin, especially on the face (which may cause loss of eyebrows and "leonine facies").

Nerve involvement in lepromatous leprosy is characteristically symmetrical and exhibits a stocking-glove distribution unrelated to the location of skin lesions. Peripheral nerve involvement may initially be manifested as loss of temperature sensation, followed by loss of light touch, pain, and deep pressure sense. In addition, dysesthesia is common. Motor complications, including muscle weakness and atrophy of the muscles of the hands, feet, and face, develop in the absence of effective antileprosy chemotherapy. Involvement of the facial nerve may result in corneal exposure, ulceration, and blindness. Persons with lepromatous leprosy may also have prominent rhinorrhea as a result of nasal mucosal involvement, and they may shed large numbers of *M. leprae* in their nasal secretions—one of the major sources of bacilli for transmission to other individuals. Like tuberculoid leprosy, lepromatous leprosy is stable; conversion to other forms does not occur.

Borderline Forms of Leprosy

Borderline tuberculoid leprosy is characterized by skin lesions similar to those of tuberculoid leprosy, but they are more numerous and may be accompanied by satellite lesions around large lesions. In borderline leprosy, skin lesions are numerous but remain asymmetrical. The lesions are usually plaques rather than macules and exhibit satellite lesions. Nerve involvement in borderline leprosy is manifested as thickening or tenderness of local nerves, but the skin lesions retain sensation. Borderline lepromatous leprosy is characterized by numerous symmetrical small macules, papules, plaques, and nodules but not the diffuse skin infiltration found in full-blown lepromatous leprosy. Unlike tuberculoid and lepromatous leprosy, the borderline forms are unstable and progress to the lepromatous form over time unless effective treatment is provided. Reactional states, including both reversal reactions and downgrading reactions, occur in those with borderline forms of leprosy.

Reactional States

Individuals with leprosy who may otherwise avoid care may exhibit acute reactional symptoms and signs. Physicians in developed countries may encounter patients with reactional states in acute care settings.[6]

Type 1 reactions, which are mediated by cellular immune responses to *M. leprae* antigens in skin lesions and nerves, occur in borderline tuberculoid, borderline, and borderline lepromatous leprosy. Type 1 reactions are frequently accompanied by worsening of peripheral nerve manifestations and may result in permanent nerve damage; they should be considered medical emergencies. Reversal reactions, which are type 1 reactions that occur after

initiation of therapy for leprosy or for human immunodeficiency virus (HIV) infection, are associated with enhanced T_H1 immune responses that develop in patients with a large burden of *M. leprae*; they are most severe in those with borderline lepromatous leprosy. Downgrading reactions occur in association with the transition of borderline disease toward the lepromatous form. Although the immune mechanisms that underlie reversal reactions and downgrading reactions are believed to be distinct, their clinical manifestations are indistinguishable. Type 1 reactions may have an acute or insidious onset, and they are characterized by inflammation of preexisting skin and nerve lesions. Skin lesions, which become erythematous and edematous, may also become tender and thereby resemble cellulitis, but type 1 reactions are not accompanied by fever or other systemic symptoms or signs. Increased expression of TNF has been found in lesions during type 1 reactions and may contribute to the clinical and functional consequences. Moreover, type 1 reactions have been reported when TNF antagonist therapy has been withdrawn after diagnosis of borderline lepromatous leprosy. Recent transcriptomic analyses have also implicated the complement system in the pathogenesis of type 1 and type 2 leprosy reactions.[7]

Type 2 reactions, also known as erythema nodosum leprosum (ENL), occur in persons with borderline lepromatous and lepromatous leprosy; they may be mediated by immune complexes rather than cellular immune responses. Type 2 reactions, which occur most often after initiation of antileprosy chemotherapy or during pregnancy, are generally accompanied by fever and arthralgias. Additional signs of systemic inflammatory disease may appear, including hepatosplenomegaly, lymphadenopathy, arthritis, nephritis, keratitis, and iritis. The skin lesions of ENL resemble those of classic erythema nodosum (Chapter 440), with widely distributed erythematous dermal and subcutaneous nodules whose location is unrelated to the leprosy lesions. Biopsy of ENL lesions shows leukocytoclastic vasculitis.

Leprosy and Human Immunodeficiency Virus

$CD4^+$ T-cell–mediated immunity is essential for control of *M. leprae,* and the extent of T_H1 immunity determines whether an individual will have tuberculoid or lepromatous leprosy. However, coinfection with HIV and depletion of $CD4^+$ T lymphocytes does not affect the rate of progression of leprosy, nor does it cause conversion of tuberculoid leprosy to the lepromatous form. In contrast, type 1 reversal reactions may accompany immune reconstitution subsequent to initiation of effective antiretroviral therapy for HIV infection (Chapter 395). The manifestations of these reversal reactions are similar to those observed in patients who are not infected with HIV.

DIAGNOSIS

The diagnosis of leprosy should be considered in any patient with skin and peripheral nerve manifestations, especially those who have lived in countries where leprosy is endemic. Although leprosy is a chronic infection, its acute complications require prompt diagnosis and therapy to prevent irreversible peripheral nerve damage. It is also important to classify a patient's disease as tuberculoid, lepromatous, or one of the specific borderline forms, because correct classification is necessary for selecting optimal therapy and anticipating potential reactional states. Because *M. leprae* cannot be cultured in vitro and there is currently no reliable serologic test or other diagnostic biomarker for leprosy, diagnosis and classification of leprosy depend on the combination of clinical examination, histopathologic evaluation, and acid-fast staining of skin slit or biopsy specimens.

Clinical Examination

Examination of an individual with possible or confirmed leprosy must include evaluation and documentation of the number, location, and characteristics of skin lesions. In addition to descriptions of the skin lesions, accurate classification of leprosy depends on whether the lesions are distributed symmetrically and whether they are hypoesthetic or anesthetic. The examination must also include a search for (1) enlarged and tender peripheral nerves, (2) the presence of sensory deficits (especially temperature sensation and pain) and skin ulcerations, and (3) the nature and distribution of motor deficits, muscle atrophy, and contractures. Because some medications used for the treatment of leprosy are contraindicated in pregnancy, women of childbearing age should be evaluated for pregnancy.

Skin Smears and Biopsies

In developing countries, classification of multibacillary or paucibacillary leprosy is made by the combination of clinical examination findings and bacterial counts as determined on acid-fast–stained smears made from skin slits of lesions and skin from cool areas of the body, such as the earlobes. In developed countries, skin biopsies are usually performed instead of skin slits. Skin specimens should be obtained from the active borders of lesions and should include subcutaneous tissue. On hematoxylin-eosin staining, tuberculoid leprosy is characterized by granulomas with giant cells, aggregates of epithelioid macrophages that are neither vacuolated nor foamy, and lymphocytes at the periphery. Although granulomas may be found in other skin diseases, selective destruction of nerve trunks and perineural fibrosis are specific features of leprosy. Acid-fast stains (preferably done with the Fite procedure) show rare or undetectable bacilli in tuberculoid leprosy. Lesions of lepromatous leprosy show poorly organized granulomas without giant cells or lymphocytes; macrophages are foamy and lipid laden. Acid-fast staining of lepromatous leprosy lesions reveals abundant bacilli that usually appear in large clumps ("globi"). The borderline forms of leprosy exhibit less well-organized granulomas with fewer giant cells and lymphocytes but more foamy macrophages and acid-fast bacilli as the spectrum varies from borderline tuberculoid to borderline lepromatous.

Specialized immunohistochemistry stains, such as for $CD4^+$ T lymphocytes or cytokine expression, are useful in research studies but are not currently used for the diagnosis or classification of leprosy. Polymerase chain reaction amplification of *M. leprae* genomic DNA from skin slits or skin biopsy specimens has not yet contributed to enhanced sensitivity or specificity of diagnosis or classification.

Diagnosis of Reactional States

The diagnosis of type 1 reactions is based on clinical findings in a patient with borderline tuberculoid, borderline, or borderline lepromatous leprosy and acute inflammation of preexisting skin lesions, with or without worsening of nerve lesions. Type 1 reactions are not accompanied by systemic findings such as fever or arthritis. At highest risk for type 1 reactions are patients who have recently initiated antileprosy chemotherapy, although type 1 reactions can occur spontaneously. Diagnosis of a type 2 reaction (ENL) is also based on clinical findings of new erythematous subcutaneous or dermal nodules in a patient with borderline lepromatous or lepromatous leprosy. There are currently no diagnostic tests or biomarkers for ENL, and skin biopsy will not distinguish ENL from classic erythema nodosum.

TREATMENT Rx

Agents to Treat Leprosy

The first-line antimicrobial agents for leprosy are dapsone and rifampin. Clofazimine, minocycline, certain fluoroquinolones, and clarithromycin are also useful in specific contexts, including drug intolerance or resistance.

Dapsone is inexpensive and well tolerated, has a long serum half-life (≈28 hours), and is safe for use during pregnancy. Glucose-6-phosphate dehydrogenase (G6PD)-deficient individuals (Chapter 161) are susceptible to dapsone-induced methemoglobinemia and hemolysis, and all patients should be screened for G6PD deficiency before starting dapsone. Patients with mild G6PD deficiency (the African type, caused by mutations that lead to instability of the enzyme) can begin dapsone at 25 mg/day but require close monitoring for hemolytic anemia. Dapsone can also cause bone marrow suppression and profound neutropenia. Other rare adverse effects of dapsone include hepatitis, cholestatic jaundice, and a hypersensitivity syndrome that usually occurs within 4 to 6 weeks of initiation of dapsone and is characterized by exfoliative dermatitis, generalized lymphadenopathy, fever, and hepatosplenomegaly. A recent study in China found that the presence of the HLA class I allele B*13:01 confers a seven-fold higher risk of the dapsone hypersensitivity syndrome, strongly implicating $CD8^+$ T cells in its pathogenesis.[8]

Rifampin, the most bactericidal drug against *M. leprae*, is well absorbed after oral administration and has a serum half-life of approximately 3 hours. Rifampin should never be used as monotherapy because resistance can develop with single point mutations in its target, RNA polymerase II. Because rifampin is bactericidal and rapid release of components from dead bacteria can have pro-inflammatory effects, some experts withhold rifampin during reversal reactions. Adverse effects of rifampin include maculopapular rash, hepatotoxicity, an influenza-like syndrome (most frequent with intermittent therapy), and orange discoloration of tears, urine, saliva, and sweat. Thrombocytopenia occurs occasionally but is not usually severe. Rifampin also induces metabolism and decreases serum concentrations of other drugs, including antiretroviral protease inhibitors and non-nucleoside reverse transcriptase inhibitors, methadone, and oral contraceptives. Rifampin decreases serum concentrations of dapsone, but this effect is not clinically significant with a dapsone dose of 100 mg/day.

Clofazimine is a lipophilic dye that is bacteriostatic against *M. leprae.* It has a very long (≈70 days) half-life and appears to have anti-inflammatory activity

as well as direct bacteriostatic activity. Because of its anti-inflammatory activity, clofazimine is useful in the treatment of type 1 reactional states. Clofazimine is generally well tolerated; its major side effect is discoloration of the skin, which occurs in nearly all clofazimine-treated patients. The skin discoloration can range from reddish tan to bluish black and can be blotchy, but it is reversible within 6 to 12 months of discontinuation of the drug. In chronic reactional patients maintained with high doses of clofazimine (200 to 300 mg/day), enteropathy with crampy abdominal pain, mild nausea, or diarrhea (or both) and even bowel obstruction can develop.

Regimens to Treat Leprosy

Chemotherapy for leprosy involves the use of multiple drugs to optimize the rate of cure and prevent emergence of drug resistance.[9] The regimen currently recommended in the United States for paucibacillary (tuberculoid and borderline tuberculoid) leprosy in adults consists of dapsone, 100 mg, and rifampin, 600 mg, both given daily for 12 months. The U.S. recommended regimen for multibacillary leprosy in adults is dapsone, 100 mg, plus rifampin, 600 mg, and clofazimine, 50 mg, each given daily for 24 months (available at http://www.hrsa.gov/hansensdisease/diagnosis/recommendedtreatment.html). Clofazimine is currently not commercially available, but it can be obtained in the United States through the National Hansen's Disease Program (1-800-642-2477). In resource-limited countries where the burden of leprosy is highest, the WHO regimen for paucibacillary leprosy is dapsone, 100 mg daily, plus rifampin, 600 mg once a month, for 6 months. The WHO regimen for adults for multibacillary leprosy (which differs from the U.S. recommendation) is dapsone, 100 mg daily, plus rifampin, 600 mg once a month, plus clofazimine, 50 mg daily, each given for 12 months, plus an additional dose of clofazimine, 300 mg once a month (available at http://www.who.int/lep/mdt/regimens/en/index.html). The monthly doses of rifampin and clofazimine should be administered under supervision. Alternative agents for patients with drug intolerance or drug resistance include clarithromycin (may be substituted for any of the first-line drugs), minocycline (may be substituted for dapsone or clofazimine), and ofloxacin (may be substituted for clofazimine).

Response to Therapy

Response to effective therapy for leprosy is seen clinically as flattening and resolution of the papules, nodules, or plaques, with or without improvement in nerve function. Clinical improvement may begin within the first months of therapy, but resolution of skin lesions is often delayed as long as 1 to 2 years after completion of therapy. Quantitation of the bacillary load to assess response to treatment is cumbersome, semiquantitative, and not recommended.

Patients who have been adequately treated may experience worsening of nerve and skin symptoms, perhaps because of a late reversal reaction or relapsed leprosy. If skin specimens do not reveal acid-fast organisms, a therapeutic trial of corticosteroids, which will ameliorate the symptoms of reversal reactions but not those of relapsed leprosy, can help make the distinction and assist in choosing subsequent therapy. Patients who experience relapse after treatment of paucibacillary disease should be treated for multibacillary disease, because the most likely cause of relapse is previous multibacillary disease that was misclassified. Patients with multibacillary disease who relapse should be retreated with a regimen containing dapsone and rifampin with the addition of at least two drugs that were not used in the initial treatment regimen, unless susceptibility testing is available and confirms that the organisms remain susceptible to dapsone and rifampin. The choices among additional drugs include minocycline, ofloxacin or moxifloxacin, and clarithromycin. Relapsed multibacillary patients may benefit from lifelong maintenance therapy after completing 2 years of a salvage regimen. Because susceptibility testing cannot be performed with in vitro assays, an alternative approach to determination of susceptibility by detecting mutations in the targets of dapsone and rifampin (*folP1* and *rpoB*, respectively) is beginning to be widely used and is commercially available. The mouse footpad assay has been used, but it is becoming less available.

Treatment of Reactional States

Type 1 reactions may develop before, during, or years after completion of antileprosy chemotherapy.[10] Type 1 reactions that involve worsening of nerve symptoms are medical emergencies because permanent nerve damage can occur. Type 1 reactions usually respond to prednisone at a daily dose of 60 to 80 mg, which can be tapered slowly once symptoms are controlled. Type 1 reactions can also respond to high-dose clofazimine (200 to 300 mg/day), although reactions with worsening nerve symptoms should be treated initially with prednisone. Patients who have type 1 reactions that occur before or during antileprosy chemotherapy and whose reactions include nerve involvement should have rifampin withheld until the worsened nerve symptoms resolve, because release of pro-inflammatory components from dying bacteria may contribute to inflammation and nerve damage. Dapsone and clofazimine should be continued during treatment of type 1 reactions.

The treatment of choice for severe type 2 reactions (ENL) is thalidomide, except in pregnant or potentially pregnant women. Thalidomide requires that patients and the prescribing physician be enrolled in the System for Thalidomide Education and Prescribing Safety (STEPS) program to avoid the drug's teratogenic effects. The mechanism of action of thalidomide is incompletely understood but is likely to include inhibition of TNF. The dose of thalidomide for ENL varies, depending on the severity of the reaction. In patients with ENL and high fever, frank arthritis, and large subcutaneous plaques, up to 100 mg four times daily may be required to achieve a clinical response. Once a clinical response is achieved, the dose of thalidomide may be tapered to a maintenance dose of 50 to 100 mg given once daily at night (because thalidomide is sedating). For milder cases of ENL, 50 to 100 mg per night may be sufficient to achieve and maintain control. ENL in women of childbearing age and thalidomide-unresponsive cases may respond to corticosteroids. Methotrexate may be efficacious in otherwise treatment-resistant ENL, but methotrexate has been assigned by the U.S. Food and Drug Administration to pregnancy category X (teratogenic risks "clearly outweigh" potential benefits) and is also contraindicated in nursing mothers. Antileprosy chemotherapy, including rifampin, should be continued in patients with ENL.

Other Therapy

Nerve damage in leprosy, which can result in muscle atrophy, contractures, and autoamputation, is the major cause of debility. Supportive care, reconstructive surgery, physical and occupational therapy, and rehabilitation can be extremely valuable in allowing patients to achieve and maintain optimal function.

PREVENTION

There is currently no effective specific vaccine for leprosy, but several trials have observed a partial protective effect of bacille Calmette-Guérin vaccination.[11] It is likely that improved understanding of transmission of *M. leprae* will be generated by the use of DNA-based strain typing, so better preventive measures are likely to become available in the near future.

PROGNOSIS

Multidrug chemotherapy cures a high proportion of people with paucibacillary and multibacillary leprosy. The currently recommended regimens provide high rates of response, with relapse rates of approximately 0.1% per year in paucibacillary cases and up to 5% per year in multibacillary cases. Some cases of paucibacillary leprosy may enter remission or even self-cure, but all cases of multibacillary leprosy are progressive. Because of its efficacy and low toxicity and to minimize long-term morbidity, multidrug chemotherapy should be used in all persons in whom leprosy is diagnosed.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. World Health Organization. Global leprosy situation, 2012. *WHO Weekly Epidemiological Record (WER)*. 2012;87:317-328.
2. Schuenemann VJ, Singh P, Mendum TA, et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science*. 2013;341:179-183.
3. Polycarpou A, Walker SL, Lockwood DN. New findings in the pathogenesis of leprosy and implications for the management of leprosy. *Curr Opin Infect Dis*. 2013;26:413-419.
4. Kamath S, Vaccaro SA, Rea TH, Ochoa MT. Recognizing and managing the immunologic reaction in leprosy. *J Am Acad Dermatol*. 2014;71:795-803.
5. Zhang F, Liu H, Chen S, et al. Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. *Nat Genet*. 2011;43:1247-1251.
6. Scollard DM, Martelli CM, Stefani MM, et al. Risk factors for leprosy reactions in three endemic countries. *Am J Trop Med Hyg*. 2015;92:108-114.
7. Dupnik KM, Bair TB, Maia AO, et al. Transcriptional changes that characterize the immune reactions of leprosy. *J Infect Dis*. 2014;[Epub ahead of print].
8. Zhang FR, Liu H, Irwanto A, et al. HLA-B*13:01 and the dapsone hypersensitivity syndrome. *N Engl J Med*. 2013;369:1620-1628.
9. Williams DL, Lewis C, Sandoval FG, et al. Drug resistance in patients with leprosy in the United States. *Clin Infect Dis*. 2014;58:72-73.
10. White C, Franco-Paredes C. Leprosy in the 21st Century. *Clin Microbiol Rev*. 2015;28:80-94.
11. Shetty VP, Mistry NF, Wakade AV, et al. BCG immunotherapy as an adjunct to chemotherapy in BL-lL patients–its effect on clinical regression, reaction severity, nerve function, lepromin conversion, bacterial/antigen clearance and 'persister' *M. leprae*. *Lepr Rev*. 2013;84:23-40.

REVIEW QUESTIONS

1. A 23-year-old Brazilian man was seen by a dermatologist for evaluation of multiple bilateral plaques and nodules; biopsy revealed poorly formed granulomas with multiple foamy cells, and Fite stain revealed multiple acid-fast bacteria. Additional investigation revealed glucose-6-phosphate dehydrogenase (G6PD) deficiency with the Mediterranean variant. What are the potential treatment options for this patient's leprosy?

A. Rifampin and dapsone
B. Rifampin and clarithromycin
C. Rifampin
D. Rifampin and clofazimine
E. Rifampin and ofloxacin

Answer: B The patient has lepromatous or borderline lepromatous leprosy with a high bacterial burden. His G6PD mutation is sufficiently severe (Chapter 161) that he is at high risk of significant hemolysis if he is treated with dapsone. Rifampin monotherapy is likely to select for rifampin-resistant variants, especially in someone with the high bacterial burden present in multibacillary leprosy. According to current guidelines, the only recommended drug to substitute for dapsone is clarithromycin. Reference: http://www.hrsa.gov/hansensdisease/diagnosis/recommendedtreatment.html

2. A 27-year-old Brazilian woman was diagnosed with borderline tuberculoid leprosy on the basis of clinical and histopathologic criteria and was begun on multidrug therapy with dapsone and rifampin. Six months after initiation of therapy, she complains of worsening of her skin lesions, which are now more erythematous, and worsening neuritic pain with weakness of her left hand. She is afebrile, and a complete blood count and urinalysis are normal. What is the most appropriate therapeutic intervention?

A. Add ofloxacin and clarithromycin for drug-resistant leprosy with progression.
B. Discontinue rifampin and treat with thalidomide.
C. Hold rifampin and add clofazimine and prednisone.
D. Continue rifampin and dapsone and treat with prednisone until symptoms resolve.
E. Biopsy the edge of one of the skin lesions and treat according to the results.

Answer: C The patient has typical symptoms and findings of a reversal reaction, which requires effective anti-inflammatory therapy to minimize progressive peripheral nerve damage. Reversal reactions are currently diagnosed clinically; a skin biopsy is not necessary. When there is evidence of nerve involvement during a reversal reaction, most experts recommend holding rifampin to minimize the ongoing release of pro-inflammatory components during bacterial death. The preferred treatment is clofazimine, owing to its anti-inflammatory activity; its additional antimycobacterial activity is thought to help minimize emergence of dapsone-resistant bacteria. When there is evidence of neural involvement, prompt addition of prednisone is indicated. Emergence of drug resistance during treatment of leprosy with a low bacterial burden, such as in borderline tuberculoid leprosy, occurs rarely and more often presents as a relapse rather than progression with evidence of increased inflammation. Thalidomide is used to treat erythema nodosum leprosum (ENL), not reversal reactions, and must be used with extreme caution and multiple methods of contraception in a woman of child-bearing age.

3. A 32-year-old man originally from Indonesia was found to have borderline tuberculoid leprosy during evaluation of a mononeuropathy. Results of a complete blood count, serum chemistry panel, and G6PD assay were all within normal limits. Four weeks after beginning treatment with dapsone and rifampin, he had gradual onset of malaise, fever, and diffuse rash. Physical examination is remarkable for a toxic appearance, fever (38.6° C), tachycardia, exfoliative dermatitis, diffuse lymphadenopathy, and hepatosplenomegaly. Lab studies are remarkable for leukocytosis, thrombocytopenia, and elevated transaminases. What intervention is most important for his recovery from the acute illness?

A. Addition of thalidomide to his current regimen.
B. Substitute clarithromycin for rifampin.
C. Add clofazimine for antimycobacterial and anti-inflammatory activity.
D. Add prednisone.
E. Discontinue dapsone.

Answer: E The patient has dapsone hypersensitivity syndrome, which is an immunologically mediated reaction that is likely due to activation of $CD8^+$ T cells. This syndrome has a case fatality rate as high as 10%, and when it occurs, dapsone must be discontinued immediately. A syndrome of systemic toxicity with the acute onset of rash in a patient receiving treatment for leprosy could also be erythema nodosum leprosum, but ENL occurs rarely if ever in patients with a low bacterial burden, as in tuberculoid or borderline tuberculoid leprosy.

4. A 72-year-old man from the Philippines was diagnosed with borderline lepromatous leprosy; treatment with rifampin, dapsone, and clofazimine was initiated. After 4 weeks, a reference laboratory reported that polymerase chain reaction (PCR) amplification of two separate skin biopsy samples revealed mutations in *Mycobacterium leprae polB* that are diagnostic of high-level rifampin resistance. How should his treatment be altered?

A. Rifapentine should be substituted for rifampin.
B. The dapsone dosage should be increased.
C. Clarithromycin should be substituted for rifampin.
D. Ofloxacin should be substituted for rifampin.
E. Minocycline should be substituted for rifampin.

Answer: C Multidrug treatment of leprosy is essential for optimal relapse-free outcomes, especially in multibacillary leprosy, which includes borderline lepromatous leprosy. Although experience is limited, it is unlikely that treatment with dapsone and clofazimine is adequate for an optimal outcome. Mutations in *polB* that confer resistance to rifampin also confer resistance to other rifamycins, including rifabutin and rifapentine. Of the alternative drugs, clarithromycin is the only one currently recommended to substitute for rifampin.

5. In patients with lepromatous leprosy, in whom lesions contain very few T lymphocytes, nerve damage is thought to be due to which of the following mechanisms?

A. Demyelination of peripheral nerve fibers by interaction of phenolic glycolipid with Schwann cells
B. Autoantibody destruction of peripheral nerve axons
C. T_H1-mediated inflammatory tissue damage
D. Cytolytic $CD8^+$ T-cell expression of granzyme B
E. Complement-mediated cytotoxicity

Answer: A In multibacillary (including lepromatous) leprosy, existing evidence is most consistent with nerve fiber damage by products of *M. leprae* rather than by host-mediated mechanisms. *M. leprae* phenolic glycolipid interacts with Schwann cell laminin 2 and activates the receptor tyrosine kinase ErbB2. Downstream signaling that requires activation of the Erk1/2 kinase(s) results in demyelination. In contrast, nerve damage in paucibacillary leprosy is believed to be mediated by host T-cell responses.

327

RICKETTSIAL INFECTIONS

DIDIER RAOULT

DEFINITION

Rickettsioses are emerging infectious diseases. Because of better diagnostic tools and changes in tick exposure, many new rickettsial diseases have been described in the past 20 years. Three families of diseases are grouped under this name: (1) rickettsioses, (2) ehrlichioses and anaplasmoses, and (3) Q fever.

The Pathogens

The agents of rickettsial diseases (formerly grouped in the order Rickettsiales) are small gram-negative bacteria that grow within eukaryotic cells. They have never been grown in axenic media thus far and for culture require living hosts such as cell cultures, embryonated eggs, or susceptible animals. With the exception of *Rickettsia prowazekii*, the agent of epidemic typhus, these

TABLE 327-1 GENETIC CLASSIFICATION OF RICKETTSIALES

	GENUS	GROUP	SPECIES	SUBSPECIES	FIRST YEAR OF ISOLATION OR DISCOVERY
Rickettsiae	*Rickettsia*	Typhus	*R. prowazekii*		1916
			R. typhi		1920
		Spotted fever	*R. conorii*	*conorii*	1932
				israeli	1974
				caspia	1991
				indica	2001
			R. rickettsii		1919
			R. sibirica	*sibirica*	1946
				mongolotimonae	1996
			R. slovaca		1997
			R. honei		1991
			R. japonica		1992
			R. parkeri		2003
			R. massiliae		2006
			R. monacencis		2007
			R. heilongjiangensis		1998
			R. aeschlimannii		2001
			R. helvetica		2000
			R. australis		1950
			R. felis		2001
			R. akari		1946
			R. raoultii		2008
	Orientia	Scrub typhus	*O. tsutsugamushi*		1920
Ehrlichiae	*Ehrlichia*		*E. chaffeensis*		1991
			E. ewingii		1999
			E. canis		1996
	Anaplasma		*A. phagocytophilum*		1992
	Neorickettsia		*N. sennetsu*		1957
	Wolbachia		*W. pipientis*		2001
Coxiellae	*Coxiella*		*C. burnetii*		1931

bacteria infect humans incidentally and are mainly animal pathogens. On the basis of molecular phylogeny, the bacteria causing rickettsial diseases have been reclassified into three phyla (Table 327-1).

Because of their difficult growth in vitro, the main diagnostic tool for rickettsioses is serology. Serologic evaluation is frequently hampered by late positivity and cross-reactivity. The development of direct staining in blood smears or skin biopsy samples, as well as polymerase chain reaction (PCR) amplification of DNA in blood samples or biopsy specimens, has considerably helped identification at the species level and led to the description of emerging pathogens.[1]

RICKETTSIOSES (DISEASES CAUSED BY *RICKETTSIA* SPECIES AND *ORIENTIA TSUTSUGAMUSHI*)

DEFINITION

Rickettsia species are small gram-negative bacteria that multiply free in the cytoplasm of their host cells. The target cells in humans are endothelial cells or monocytes, and vasculitis is the most prominent clinical manifestation. These bacteria invade cells by phagocytosis and escape the phagosome vacuole.[2]

The genome of *Rickettsia* is small, between 1.1 and 1.6 Mb; some have plasmids and potential for conjugation. These bacteria have a family of outer membrane proteins of the surface cell antigen family, including rOmpA (lacking in typhus group) and rOmpB. These proteins are major antigens that help identify the rickettsial species, and their encoding genes are used for amplification and sequencing for diagnostic or taxonomic purposes. Among rickettsiae, two subgroups, the typhus group and the spotted fever group, were identified on the basis of growth conditions and antigenicity. A specific group antigen determined to be lipopolysaccharide has been identified. The optimal growth temperature is 37° C for the typhus group and 32° to 35° C for the spotted fever group. The complete genome sequencing of *R. prowazekii* (from the typhus group) showed that it is mainly a subset of *Rickettsia conorii* (a member of the spotted fever group).

Tick-Borne Rickettsioses

ROCKY MOUNTAIN SPOTTED FEVER

EPIDEMIOLOGY

Rocky Mountain spotted fever (RMSF), the most severe of the rickettsioses, is caused by *Rickettsia rickettsii* (Table 327-2).[3] It is the major tick-transmitted rickettsiosis (Chapter 359) recognized in America, with *Rickettsia africae* in the West Indies, *Rickettsia parkeri* in the southern states of the United States,[4] and perhaps *Rickettsia amblyommii*. It was described first in the 19th century in the western United States. RMSF is prevalent in at least 44 U.S. states (Fig. 327-1) and in Central and South America (Argentina, Brazil, Colombia, Costa Rica, Mexico, and Panama).[5]

Rickettsia is transmitted transovarially to tick progeny from one generation to the next. The infecting ticks are mainly *Dermacentor andersoni* (a wood tick) in the western United States; *Dermacentor variabilis* (the American dog tick) in the East, the Midwest, and the South; and *Rhipicephalus sanguineus* in Arizona. In Central and South America, *Amblyomma cajennense* is the major vector. Humans are infected through infected saliva after a tick bite. The duration of attachment is critical in any tick-borne rickettsiosis, and transmission is unlikely when the tick feeds for less than 20 hours. The tick

TABLE 327-2 RICKETTSIAL DISEASES IN HUMAN BEINGS

DISEASE	ORGANISM	ARTHROPOD HOST	GEOGRAPHIC AREA	RASH	ESCHAR TACHE NOIRE	REGIONAL LYMPH NODE	HIGH FEVER	FATALITY RATE
TICK-TRANSMITTED SPOTTED FEVERS								
Rocky Mountain spotted fever	*R. rickettsii*	*Dermacentor andersoni* *Dermacentor variabilis* *Rhipicephalus sanguineus* *Amblyomma cajennense*	America (North, Central, and South)	Yes, may be purpuric	Very rare	No	Yes	High
Mediterranean spotted fever, Astrakhan fever, Israeli spotted fever	*R. conorii*	*Rhipicephalus sanguineus*	Mediterranean, India, Caspian Sea, Africa	Yes, papular; may be purpuric	Yes	No	Yes	Moderate
African tick bite fever	*R. africae*	*Amblyomma hebraeum* *Amblyomma variegatum*	Sub-Saharan Africa, West Indies	Yes, half of cases may be vesicular	Yes (frequently multiple)	Yes	No	Low
Queensland tick typhus	*R. australis*	*Ixodes holocyclus*	Eastern Australia	Yes, may be vesicular	Yes	?	Yes	Moderate
Siberian tick typhus	*R. sibirica*	*Dermacentor nuttallii*	Siberia, China, Mongolia	Yes	Yes	No	Yes	Low
Scalp eschar, neck lymphadenopathy after tick bite (SENLAT)	*R. slovaca or R. raoultii*	*Dermacentor marginatus* *Dermacentor reticulatus*	Europe, Pakistan	Very rare	Yes, may be erythematous	Yes (painful)	No	Low
Lymphangitis-associated rickettsiosis (LAR)	*R. sibirica mongolotimonae*	*Hyalomma asiaticum*	Mongolia, Africa, Europe	Yes	Yes	Yes	Yes	Low
Unnamed	*R. aeschlimannii*	*Hyalomma* sp.	Mediterranean, Africa	Yes	Yes	Yes	Yes	Unknown
Flinders Island spotted fever	*R. honei*	*Ixodes granulosus*	Flinders Island, eastern Australia	Yes	Yes	Yes	Yes	Low
Japanese spotted fever	*R. japonica*	*Ixodes ricinus*	Japan, Korea (China?)	Yes	Yes	No	Yes	Low
Unnamed	*R. parkeri*	*Amblyomma maculatum*	America	Yes	Yes	No	Yes	
Unnamed	*R. helvetica* *R. massiliae* *R. monacencis*	*Ixodes ricinus* *Rhipicephalus sanguineus* *Ixodes ricinus*	Europe, Asia Europe, United States Europe	No Yes Yes	Yes Yes Yes	No No No	No Yes Yes	 Unknown Unknown
FLEA-TRANSMITTED DISEASES								
Murine typhus	*R. typhi*	*Xenopsylla cheopis* *Ctenocephalides felis* (mosquitoes)	Worldwide	Yes	No	No	Yes	Low
Flea-borne spotted fever	*R. felis*	*Ctenocephalides felis*	Worldwide	Sometimes	Sometimes	Unknown	Yes	Unknown
LOUSE-TRANSMITTED DISEASE								
Epidemic typhus	*R. prowazekii*	*Pediculus humanus corporis* *Amblyomma* ticks (?)	Worldwide	Yes	No	No	Yes	High
American sylvatic typhus	*R. prowazekii*	Flying squirrel ectoparasites	United States	Yes	No	No	Yes	Low
Brill-Zinsser disease (relapse of epidemic typhus)	*R. prowazekii*		Worldwide	Yes, could lack	No	No	No	Low
MITE-TRANSMITTED DISEASE								
Rickettsialpox	*R. akari*	*Liponyssoides sanguineus*	Worldwide	Yes, vesicular	Yes	Yes	Yes	Low
Scrub typhus	*Orientia tsutsugamushi*	*Leptotrombidium* sp. (chiggers)	Central and eastern Asia, Australia	Yes	Yes	Yes	Yes	High, may relapse

bite is painless and frequently unnoticed. Rarely, an eschar at the site of the tick bite is observed in RMSF. The epidemiology of RMSF undergoes largely unexplained yearly variations. This temporal repartition is determined by tick activity and human encounter. More than 500 cases occur each year, and more than 90% are reported from April to September. The disease is more prevalent in children younger than 10 years. A recent increase has been reported, but the current diagnostic tools do not allow discrimination between RMSF and other rickettsioses.

CLINICAL MANIFESTATIONS

Two to 14 days after the tick bite, fever and headaches appear. The fever is high (temperature > 102°F) and associated with nonspecific symptoms

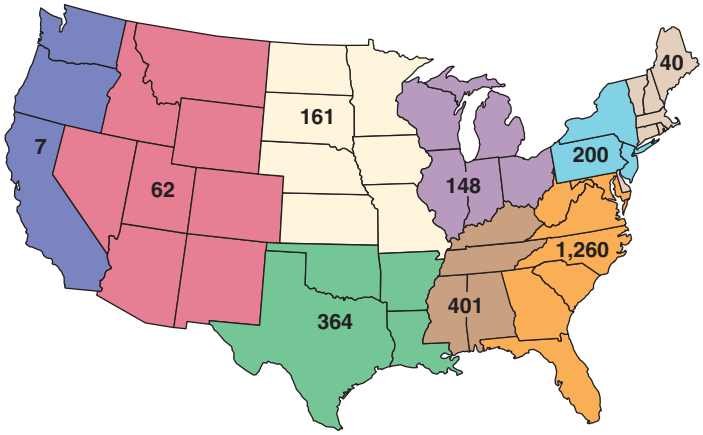

FIGURE 327-1. Number of reported cases of Rocky Mountain spotted fever by region, 1994 to 1998.

including malaise, myalgias, nausea, vomiting, anorexia, and diarrhea. At this stage, RMSF is not frequently diagnosed, but during the "tick season," patients with high fever who live in or have a history of travel to an endemic location and possibly a history of tick bite should be considered as possibly having RMSF.

The most characteristic feature is a rash. However, the classic triad of fever, headache, and rash is present in only 44% of confirmed cases. Rash is found in 14% of cases on the first day of disease and in less than 50% in the first 3 days. The rash is macular; it appears first on the ankles and wrists and then generalizes. Spots are 1 to 5 mm in diameter and can evolve from pink to purpuric. A rash can appear later or even not at all; Rocky Mountain "spotless" fever represented 34% of cases in a series from the Centers for Disease Control and Prevention (CDC). Involvement of the palms and soles theoretically differentiates the typhus diseases (in which it is absent) from the spotted fevers.

Untreated patients worsen progressively. The disease is associated in various degrees with general manifestations related to vascular inflammation and increased vascular permeability and with multiple organ involvement that can lead to multiple organ dysfunction syndrome (MODS). In severe forms, patients suffer from edema, hypovolemia, hypoalbuminemia, and hypotension leading to shock. In very severe cases, necrosis and gangrene of the extremities occur. In some instances, noncardiogenic pulmonary edema develops; pulmonary involvement leading to respiratory distress can cause death. Renal failure can result either from hypovolemia and shock and be reversible or from acute tubular necrosis and require hemodialysis. The usual neurologic symptoms are confusion, lethargy, and stupor. In severe cases, delirium, coma, and seizures are observed. Cerebrospinal fluid (CSF) sampling exhibits meningitis in a third of cases; in general, a few monocyte cells (10 to 100) are observed, along with increased protein but normal glucose levels. Heart involvement can cause arrhythmia. Liver involvement is manifested as an increase in transaminases in a third of patients and jaundice in 8%. Jaundice can also reflect hemolysis. Intestinal tract involvement is manifested as abdominal pain, diarrhea, vomiting, and severe bleeding (upper gastrointestinal hemorrhage can cause death). Ocular involvement consists of conjunctivitis and retinal abnormalities, including hemorrhages, papilledema, and arterial occlusion.

The blood cell count shows a normal number of white blood cells but often immature myeloid cells. Thrombocytopenia is observed in 30 to 50% of cases and may be marked in severe cases. Anemia develops in 30% of patients. Coagulopathy with decreases in clotting factors (including fibrinogen) and prolonged coagulation times may contribute to bleeding; serum albumin may be low and proteins of the acute phase response increased (C-reactive protein, ferritin, fibrinogen). Hyponatremia and hypocalcemia may be noted and correlate with severity, as with an increase in creatinine. Increased concentrations of serum enzymes such as aminotransferases (aspartate [AST] and alanine [ALT] aminotransferase), lactate dehydrogenase (LDH), and creatine kinase usually reflect the severity of organ involvement—including the lung, heart, and liver—and multifocal rhabdomyolysis.

DIAGNOSIS

The diagnosis of RMSF should be based on clinical and epidemiologic findings and lead to early use of doxycycline. The most important clue is unexplained fever in a patient with a history of tick exposure in an endemic area. When a rash is present, RMSF should be suspected and the patient treated accordingly unless another cause is demonstrated. The differential diagnosis includes other rickettsioses (e.g., those caused by *R. parkeri* in southeastern states), meningococcemia, enterovirus infections, typhoid, leptospirosis, ehrlichiosis, gonococcemia, toxic shock syndrome, syphilis, rubella, measles, and the Kawasaki syndrome. Drug hypersensitivity, especially after antimicrobial use for febrile illness, is sometimes confused with RMSF.

The main diagnostic test relies on serology, and treatment should never be delayed to obtain diagnostic confirmation. Criteria for laboratory confirmation include a four-fold or greater change in antibody titer determined by serology (measured by immunofluorescence assay [IFA], complement fixation, or latex agglutination) and direct detection of the bacterium by demonstration of specific antigens by immunodetection, genomic amplification by PCR, or culture. A biopsy specimen of a skin lesion is the best sample for this purpose. Culture of *Rickettsia* takes 3 to 7 days and is restricted to specialized laboratories. It is performed on cell lines such as Vero, L929, or HEL cells. Immunodetection by IFA or immunohistochemistry is sensitive and specific. It can be performed with frozen or fixed and paraffin-embedded material and allows retrospective diagnosis. PCR amplification and identification give promising results in rickettsioses in general but have not been properly evaluated for diagnosis of RMSF. Skin biopsy and direct detection in removed ticks yield the best results because blood contains inhibitors and only few copies of rickettsial DNA.

Two serum samples should be tested (early and convalescent). The early serum is usually negative because patients seroconvert between the 7th and 15th days. IFA is highly sensitive and specific. A cutoff value of 1 : 64 for total immunoglobulin and 1 : 32 for IgM antibodies is required for diagnosis. The latex agglutination cutoff is 1 : 64 or 1 : 128. Cross-reactive antibodies have been reported with infections caused by other rickettsioses, *Ehrlichia*, *Bartonella*, *Legionella*, and *Proteus*. False-positives, including IgM, may be observed when rheumatoid factor is present in serum and in patients with viral infection generating nonspecific B-lymphocyte proliferation (cytomegalovirus, Epstein-Barr virus). Complement fixation (which lacks sensitivity) and the Weil-Felix test (using antibodies that cross-react with *Proteus* strains) should not be used.

TREATMENT

Rx

The prognosis for patients with RMSF depends on the timing of antimicrobial treatment. Doxycycline saves patients with RMSF. The recommended dose is 100 mg twice daily, and treatment should be continued for at least 3 days after the fever resolves. Oral treatment is effective, but in patients with gastric intolerance or coma, the intravenous route is advised. Several antimicrobials are effective in vitro against *R. rickettsii*, including fluoroquinolones, rifampin, and new macrolide antimicrobials (but not erythromycin), but lack of clinical experience precludes their use for RMSF. β-Lactam antimicrobials, aminoglycosides, and cotrimoxazole are not effective.

Severely ill patients should be treated in intensive care units and fluid administration carefully monitored. Mechanical ventilation is used in case of respiratory distress, hemodialysis in patients with renal insufficiency, and antiseizure drugs in patients with seizures. Anemia and coagulation abnormalities may also be corrected. For patients with gangrene of the extremities, amputation may be necessary. Glucocorticoids have not proved useful.

PREVENTION

Prevention is based on avoidance of tick bites (Chapter 359) by use of repellents, protective garments, or both. To discourage tick attachment, repellents containing permethrin can be sprayed on boots and clothing and will last for several days. Repellents containing DEET (*N,N*-diethyl-*m*-toluamide) can be applied to the skin but will last only a few hours before reapplication is necessary. It is also useful to check for ticks after exposure. Careful examination of the scalp, groin, and axillae is recommended. The tick can be removed by forceps, and the skin should be disinfected (Fig. 327-2).

PROGNOSIS

The evolution of RMSF depends strongly on the timing of diagnosis and antimicrobial treatment. The current fatality rate is 2.4% on the basis of a 4-year national survey in the United States (27 deaths were attributable to RMSF during this period). This rate is currently declining, but this may result

from reporting of confounding rickettsial diseases. No significant difference in outcome was observed between blacks and whites, but the case-fatality rate was highest in individuals older than 70 years (9%). Patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency are more susceptible to severe infection. Treatment with chloramphenicol has been associated with a poorer outcome than treatment with doxycycline. Recovery from RMSF is usually complete, but neurologic sequelae can remain, and amputation of extremities may be necessary after gangrene.[6]

OTHER TICK-BORNE RICKETTSIOSES

EPIDEMIOLOGY

Like other tick-transmitted diseases (Chapter 359), rickettsioses have a limited geographic distribution that is determined mainly by the tick vector ecology (Fig. 327-3). *R. parkeri* has recently been identified in the United States and South America. *R. conorii* is found in Europe around the Mediterranean and Caspian seas (*caspia* subspecies); *Rickettsia slovaca*, *Rickettsia raoultii*, and possibly *Rickettsia helvetica* in western and central Europe; and *Rickettsia sibirica mongolotimonae* in France and Greece. Elsewhere, a number of specific agents of rickettsial disease have been identified (see Table 327-2).

CLINICAL MANIFESTATIONS

R. conorii comprises different but closely related subspecies. Many names are given to the infection caused by *R. conorii:* Mediterranean spotted fever (MSF), boutonneuse fever, Marseilles fever, Kenya tick typhus (caused by the subspecies *R. conorii conorii*), Astrakhan fever (caused by *R. conorii caspia*), Israeli spotted fever (caused by *R. conorii israeli*), and Indian tick typhus (caused by *R. conorii indica*). *R. conorii* is closely related to *R. rickettsii*, with which it shares many common antigens that generate cross-reactive antibodies. MSF resembles RMSF but has several distinct characteristics. The spontaneous evolution is milder, but a fatality rate of 1.5 to 2.5% in hospitalized patients is still observed. A malignant form of the disease that includes purpuric rash, shock, and MODS has been described in alcoholic, diabetic, human immunodeficiency virus (HIV)-infected, and elderly or debilitated patients. The typical clinical manifestation is that of a patient with fever, a rash, and a tache noire (i.e., a black eschar at the site of the tick bite). A tache noire is found in 50 to 80% of cases. Multiple lesions do not occur because the dog tick vector, *R. sanguineus*, seldom bites humans. The rash is frequently clearly papular, which led to one of the names of the disease, boutonneuse fever. Israeli tick bite fever and Astrakhan fever appear to be milder than typical MSF, and tache noire is usually lacking.

R. africae, which causes African tick bite fever, may be responsible for most of the rickettsioses worldwide. It is extremely common in travelers visiting southern Africa. It is transmitted by African ticks, *Amblyomma hebraeum* and *Amblyomma africanum.* These ticks are often infected; as many as 60% can harbor *R. africae.* They usually feed on ungulates but attack human beings in groups and cause a high prevalence of infection in rural Africa (60% of tested patients exhibit antibodies) and in travelers. The tick attacks typically generate clusters of cases in Safari tourists. The disease differs from MSF in that it is much milder, fever is frequently absent, a rash is observed in only half the patients, and the rash may be vesicular (which has never been reported in confirmed MSF). Moreover, several taches noires are frequently observed. They are prevalently found on the lower limbs and often associated with draining lymphadenopathy in the groin.

Japanese spotted fever (caused by *Rickettsia japonica*) and Siberian tick typhus (caused by *R. sibirica*) resemble MSF. Infections caused by *R. sibirica mongolitimonae* resemble MSF but in some cases exhibit specific clinical features, including a tache noire, groin lymphadenopathy, and lymphangitis joining these two lesions. The disease has recently been named lymphangitis-associated rickettsiosis. *Rickettsia australis* (Queensland tick typhus) and *Rickettsia honei* (Flinders Island spotted fever) cause diseases resembling MSF, but their rash can be vesicular.

R. slovaca and *R. raoultii* cause a disease apparently common in Europe named scalp eschar and neck lymphadenopathy transmitted by ticks (Hungary, Germany, France, Spain). Its tick vectors, *Dermacentor marginatus* and *Dermacentor reticulatus*, preferentially bite in cold months and bite the scalp because they prefer hairy prey. In contrast to other tick-borne rickettsioses, the disease is more prevalent in children and women. It is rarely exanthematic; the typical clinical picture consists of an erythematous skin lesion at

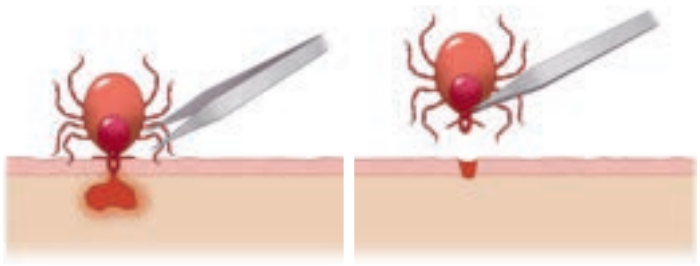

FIGURE 327-2. Tick removal technique.

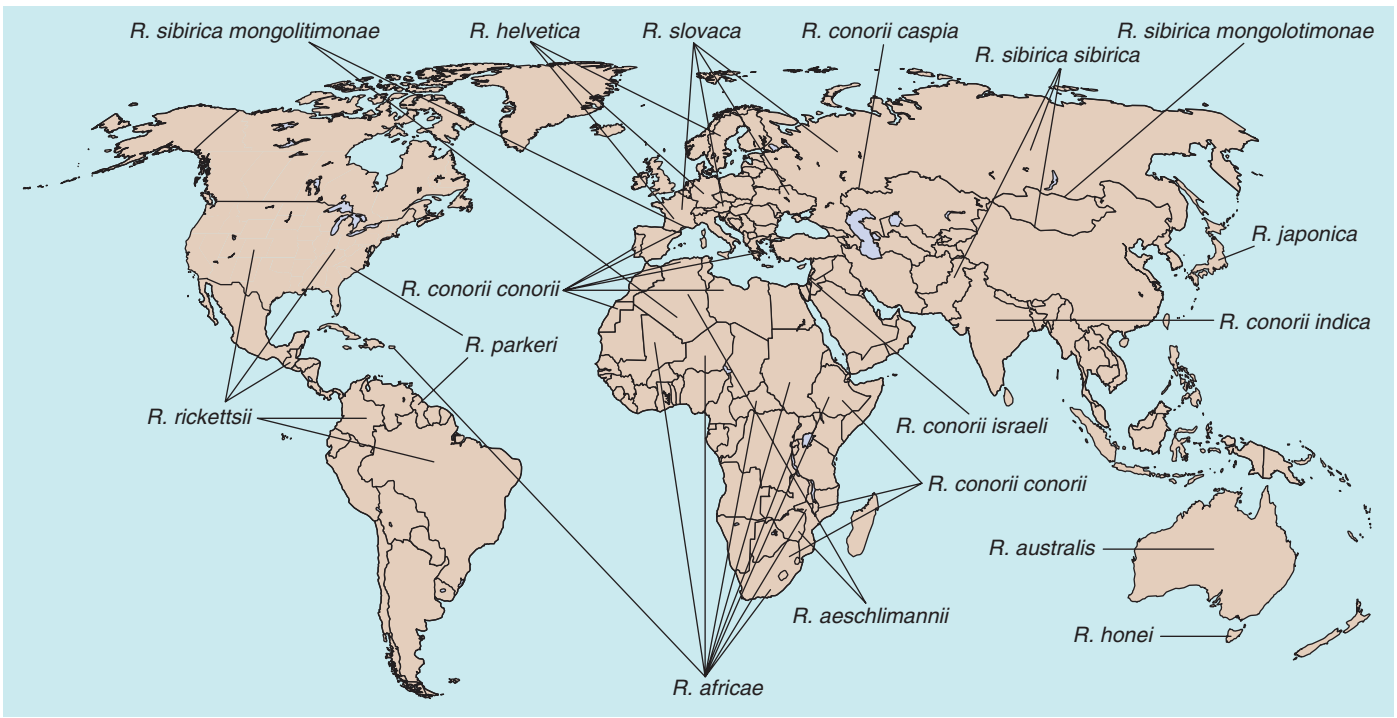

FIGURE 327-3. Geographic distribution of tick-borne rickettsioses.

the site of the tick bite on the scalp that ranges from 2 to 8 cm in diameter, and a draining neck lymphadenopathy that may be painful. Rarely, patients may exhibit fever and a rash. Deep postinfectious asthenia and residual alopecia at the site of the tick bite can be observed. The occurrence of this rickettsiosis without rash may stimulate research on other new rickettsial diseases with only localized manifestations.

DIAGNOSIS

The diagnosis of other tick-borne rickettsioses is similar to that of RMSF, mainly by serology (IFA; see earlier). An exception is *R. slovaca* infection, in which the serologic response is weak, possibly because of its lack of general infection; in this case, PCR of a skin lesion sample by a swab or a lymph node aspirate is the best solution. In *R. africae* infection, the serologic response occurs later than in RMSF and MSF, and late serum samples are therefore recommended.

TREATMENT Rx

Doxycycline (100 mg twice daily for adults or 4.4 mg/kg body weight per day in two divided doses for children under 45.4 kg [100 lb]) is the drug of choice for treatment. A single day of therapy usually suffices, but in adults with more severe disease, it should be administered until the patient is afebrile for 24 hours. In pregnant women, josamycin, a macrolide antimicrobial, has proved efficient at a dose of 3 g daily for 7 days for MSF; quinolones and newer macrolide antimicrobials give results comparable to those of doxycycline but with longer regimens.

Flea-Transmitted Diseases

Fleas (Chapter 359) can harbor two rickettsial species: *Rickettsia typhi*, the agent of murine typhus, and *Rickettsia felis*, the agent of flea-borne spotted fever. Both rickettsiae can be transmitted transovarially in the flea. Vectors are *Xenopsylla cheopis* and *Pulex irritans* but also *Ctenocephalides felis*, a cat flea. Rats, cats, opossums, and dogs can propagate infected fleas. These reservoirs and vectors are distributed worldwide, and thus these diseases have a global distribution. Fleas can be infected by both species at the same time.

MURINE TYPHUS

DEFINITION

Fleas are usually infected by *R. typhi* when feeding on apparently healthy rats that have blood-borne infection. Humans and other mammals are infected through autoinoculation by scratching a fleabite that is contaminated with feces from an infected flea. Murine typhus, because of its cycle, is more prevalent in hot and humid areas when rats proliferate.

EPIDEMIOLOGY

In the United States, 50 to 100 cases are reported yearly, mainly in southern California and southern Texas. In California, a transmission cycle involving opossums and cat fleas has been demonstrated. Murine typhus is extremely common in Southeast Asia and North Africa and is a common cause of fever in travelers.

CLINICAL MANIFESTATIONS

On the basis of studies of infected volunteers, the incubation period is generally 8 to 16 days. The disease begins with abrupt fever, nausea, vomiting, myalgias, arthralgias, and headache. A rash is observed in 40 to 50% of patients about 6 days after the onset.[7] It is detected even less frequently in patients with dark skin. The rash begins as pink maculae that can evolve to be maculopapular. It is often discrete, starting in the axilla; it generalizes to the trunk but does not usually involve the face, palms, and soles. In severe cases, it can become purpuric. The most frequently involved organ is the lung. A third of patients have a cough, and in a fourth, a nonspecific interstitial pneumonia develops that is sometimes associated with a pleural effusion. In severe forms, respiratory failure occurs. In patients with severe disease, neurologic symptoms range from confusion and stupor to coma and seizures. Cerebral hemorrhages may occur. Gastrointestinal involvement can be manifested as vomiting, abdominal pain, jaundice, and in severe cases, hematemesis.

The white blood cell count initially shows leukopenia and then leukocytosis. Thrombocytopenia can be noted as well as anemia, specifically when hemolysis is observed (frequently in patients with G6PD deficiency). A moderate increase in serum liver enzymes is common. In patients with severe disease, hyponatremia and hypoalbuminemia are observed.

DIAGNOSIS

The diagnosis of murine typhus is based mainly on serology (IFA), with titers similar to those of RMSF. On serologic evaluation, *R. typhi* cross-reacts with *R. prowazekii*; it can be differentiated either by comparing titers (two dilutions or more if IgG and IgM titers are discriminative) or by cross-adsorption. In this technique, the serum is absorbed with either antigen and then retested, and the causative agent is that removing antibodies to both bacteria. Skin biopsies and blood samples for culture and PCR may be valuable.

TREATMENT Rx

Treatment is the same as that for RMSF.

PROGNOSIS

The prognosis is usually favorable, but 10% of patients require intensive care and 1% die. Older patients and those with G6PD deficiency (Chapter 161) or chronic debilitating conditions are at higher risk.

FLEA-BORNE SPOTTED FEVER CAUSED BY *RICKETTSIA FELIS*

R. felis is mainly transmitted transovarially. Its genome comprises one or two plasmids, one being apparently conjugative. This is a new, incompletely defined disease. The bacterium is found in fleas in the Americas, Asia, Europe, Africa, and New Zealand. Isolated cases have been reported from Texas, Mexico, Brazil, France, and Germany. Reported cases all exhibited fever, a rash in six of seven cases, and inoculation eschar in some cases. The diagnosis can be based on serologic evaluation using specific *R. felis* antigen or PCR of blood or skin biopsy samples. Treatment has not been established, but the bacterium is highly susceptible to doxycycline and resistant to erythromycin. *R. felis* has been found at very high prevalence in the blood of febrile sub-Saharan Africans and is suspected to be transmitted by mosquitoes.

Louse and Mite Infections

EPIDEMIC LOUSE-BORNE TYPHUS

EPIDEMIOLOGY

The human body louse (Chapter 359) lives in clothes and multiplies rapidly when cold weather and lack of hygiene allow it to. The body louse transmits three bacterial diseases: (1) trench fever (caused by *Bartonella quintana*), (2) relapsing fever (caused by *Borrelia recurrentis*), and (3) exanthematic typhus (caused by *R. prowazekii*). The name *typhus* is derived from the Greek *tuphos*, which describes the neurologic condition associated with this disease and with typhoid. The body louse is prevalent during war, in poor countries, and in the homeless population of rich countries, including the United States and Europe. A 100,000-person outbreak of typhus was reported during the civil war in Burundi in 1997, and cases were reported in Russia, Peru, the United States, Algeria, and France in the 1990s. Louse-transmitted diseases killed more people than weapons did during central and eastern European wars in the 19th and 20th centuries.

The epidemiology of *R. prowazekii* is mainly related to humans as reservoirs and lice as vectors. In the United States, the eastern flying squirrel *(Glaucomys volans volans)* is also a reservoir, and its fleas, lice, and mites can be infected.

R. prowazekii has also been found in *Amblyomma* ticks, but their role is not known. The louse is infected when feeding on blood, which it does five times a day. *R. prowazekii* multiplies in the gut of the louse and is released in feces; after a few days, it destroys intestinal epithelium, which causes bright red blood to spread from the gut (typhus was also named the red louse disease). The patient is usually contaminated by infected feces (in which *R. prowazekii* survives for weeks), through aerosols, or by skin autoinoculation after scratching. Patients who recover from typhus may harbor the bacterium in a dormant form and suffer relapses under stressful conditions years later; this relapsing form is called Brill-Zinsser disease. During the relapse, a bacteremia occurs that may allow the start of a new outbreak if lice bite the patient.

CLINICAL MANIFESTATIONS

Typhus begins abruptly with fever, headaches, and myalgias, which may have led to the crouched posture termed *sutama* in the largest recent outbreak in Burundi. Cough and neurologic signs (stupor, confusion, or coma) are common. A rash is observed in 20 to 80% of patients, depending on the population studied; it is probably commonly underobserved on dark skin. It generally starts in the axilla and then spreads. The rash is usually macular but can be papular or purpuric in severe cases. In some cases, diarrhea and jaundice are reported. Splenomegaly is infrequently found. In severe cases, shock occurs and the fatality rate is 20 to 30%. Leukopenia, thrombocytopenia, and anemia as well as an increase in serum hepatic enzymes may be noted.

Sylvatic typhus in the United States is caused by an *R. prowazekii* variant and is a milder disease. The most prominent clinical features are neurologic. Few cases have been described, and nearly all occurred in areas where the eastern flying squirrel is found, east of the Mississippi.

Brill-Zinsser disease is difficult to diagnose because rash is rare and recent exposure to lice can be lacking. Interviewing the patient may reveal prior exposure to lice, associated or not with a diagnosis of typhus in previous years. The disease is mild and the prognosis is good.

DIAGNOSIS

The diagnosis of typhus should be considered when grouped cases of high fever with confusion are observed in patients exposed to lice. The most common diagnostic error is to attribute the findings to typhoid, which can have fatal consequences because the antimicrobials typically prescribed for that condition (β-lactams, cotrimoxazole, and quinolones) are ineffective treatment of typhus. In tropical countries, typhus is frequently confused with malaria, hemorrhagic fever, and dengue. In persons with lice, it can be confused with trench fever and relapsing fever, but treatment for both can be prescribed.

The diagnosis of typhus should be clinical because the fatality rate is high and the treatment safe and efficient. Any outbreak of unexplained fever in unhygienic environments may suggest typhus, including outbreaks during civil wars (such as in Algeria, Rwanda, and Burundi), during social collapses (such as in Russia and Ukraine), in jails (such as in Rwanda and Burundi), and in chronically poor and cold countries. The diagnosis is based mainly on serology, in which there is cross-reaction with *R. typhi* (see earlier). When the investigation is performed under difficult field conditions, a drop of blood applied on filter paper and sent to a reference laboratory is valuable for serologic testing. Culture and PCR are helpful and can be performed with a skin biopsy sample or blood. Lice are good diagnostic tools because they can be tested even when dry and can be sent in closed containers without specific temperature conditions.

TREATMENT Rx

Treatment of typhus is extremely simple, cheap, and effective; 200 mg of doxycycline orally in two divided doses is life-saving. Comatose patients should be treated parenterally. In allergic patients, chloramphenicol is the only known alternative, prescribed at a dose of 2 g/day for 10 days. There is no current vaccination, and the fight against lice is the major prevention strategy. Because lice are fragile, changing and boiling clothes are efficient. When this is not possible, insecticides (primarily permethrin) or ivermectin orally should be used.

SCRUB TYPHUS (*ORIENTIA TSUTSUGAMUSHI*)

EPIDEMIOLOGY

Scrub typhus is transmitted by the bite of trombiculid mite (Chapter 359) larvae infected by *O. tsutsugamushi.* These mites, also named chiggers, are vertically infected through their mother. Scrub typhus distribution is limited to a triangle extending between northern Japan, eastern Australia, and eastern Russia and includes the Far East, China, and the Indian subcontinent. All together, 1 billion people may be exposed. Seasonality is determined by the emergence of larvae. It is one of the three most common causes of prolonged fever in rural Asia; in temperate zones, it occurs mainly in autumn and to a lesser extent in spring. *O. tsutsugamushi* species have a wide heterogenicity that may allow the definition of several species, but a single species is currently recognized with many serotypes.

CLINICAL MANIFESTATIONS

The disease occurs in patients exposed to rural or urban foci of scrub typhus after a delay of 10 or more days. The onset is usually sudden and includes fever, headache, and myalgias. Attentive examination may reveal an inoculation eschar at the site of the mite bite and tender draining lymph nodes. Generalized lymphadenopathy and rash may be observed. The symptoms vary according to organ involvement. Neuromeningeal symptoms are relatively common. Severe forms can be manifested as septic shock. Abortion commonly occurs in pregnant women.

Leukopenia, thrombocytopenia, and increased levels of hepatic enzymes can occur. Evolution depends on the hosts and strains, and the fatality rate ranges from 0 to 30%. Scrub typhus is not more severe in HIV-infected patients, and surprisingly, HIV suppressive factors appear to be produced during infection. Relapses may occur in this disease.

DIAGNOSIS

Diagnosis may be difficult. Because the clinical features are frequently not specific, epidemiologic factors are critical.[8] A diagnosis of infectious mononucleosis has erroneously been made in patients with scrub typhus. The bacterium can be detected by culture (in cells or mice) or by PCR in blood and biopsy specimens. The serologic technique first used was agglutination of *Proteus mirabilis* serotype OXK in the Weil-Felix reaction. This test lacks sensitivity and specificity and should be replaced by IFA or enzyme-linked immunosorbent assay tests using the three or four major serotypes.

TREATMENT Rx

Chloramphenicol was the mainstay of treatment for many years, but now doxycycline is recommended. Single-day treatment with doxycycline is followed by relapses, and even repeated treatment for 2 days at a 7-day interval does not prevent all relapses. Hence, the currently recommended regimen is doxycycline, 100 mg orally twice a day for 7 days. Cases resistant to doxycycline have been reported, and rifampin (600 mg orally daily) is a reasonable alternative. Quinolones should be avoided. Prophylaxis is based on the use of repellents.

RICKETTSIALPOX (*RICKETTSIA AKARI*)

EPIDEMIOLOGY

Rickettsialpox was first described by a general practitioner in 1946 in New York City, where it is still prevalent. *R. akari*, the causal agent, is transmitted by the bite of the mouse mite (*Liponyssoides sanguineus*). Its prevalence is probably underestimated; an active search revealed 13 cases in a New York hospital in the 1980s. Cases have been reported in Arizona, Utah, and Ohio. After the terrorist attacks of 9/11/01, cases of black skin eschars were investigated as possible anthrax in New York but were in fact rickettsialpox. High seroprevalence was reported among intravenous drug users in Baltimore. Cases have also been reported from Russia, Ukraine, Slovenia, and Korea.

CLINICAL MANIFESTATIONS

Ten days after the mite bite, the beginning of the illness is marked by fever, headache, and myalgia. Careful examination reveals an inoculation eschar and a draining lymphadenopathy that could be mistaken for cutaneous anthrax. Two to 6 days later, a rash appears and comprises 5 to 40 macular then papular and vesicular spots. This aspect led to the name of the disease. It is frequently mistaken for chickenpox. The disease is usually mild.

DIAGNOSIS

The diagnosis can be made by serologic testing with IFA. Specific antigens react with high titer, but antibodies to other *Rickettsia* may be detected. The diagnosis may also be made on skin specimens by culture, immunodetection, or PCR.

TREATMENT Rx

Doxycycline is highly effective in these patients. Prevention is based on the control of mice.

EHRLICHIOSES AND ANAPLASMOSES

DEFINITION

The index case of modern ehrlichiosis was reported in the United States in 1987.[9] The patient died of fever, presumably acquired after a tick bite in Arkansas, despite receiving chloramphenicol. The patient had several initially confusing diagnostic features; on blood smears, morulae in polymorphonuclear (PMN) cells were seen, and antibodies to *Ehrlichia canis*, a pathogen of dogs but not humans, were detected. He was then thought to have *Ehrlichia chaffeensis*, but this bacterium infects monocytes, not PMN cells. A diagnosis of *Anaplasma phagocytophilum* infection (or human granulocytic ehrlichiosis [HGE]) was considered, but the tick vector of this disease is absent in Arkansas. The most likely diagnosis is currently believed to be infection with *Ehrlichia ewingii*, an agent transmitted by *Amblyomma americanum* that is prevalent in Arkansas, infects PMN cells, and cross-reacts with *E. canis*, although it typically affects immunocompromised hosts. This case illustrates the progress in knowledge on ehrlichioses and how difficult it is to conclude the etiology of an atypical infection definitively on the basis of serology alone.

All *Ehrlichia* pathogenic for humans—except *E. ewingii*—can be cultured.[10] The ehrlichiae have been reclassified into four genera, mainly on the basis of 16S ribosomal RNA–derived phylogenetic analysis. Two are the tick-associated genera *Ehrlichia* and *Anaplasma* (*A. phagocytophilum*, or the HGE agent that was formerly named *Ehrlichia phagocytophila*). One is a helminth-associated genus, *Neorickettsia*, including *Neorickettsia sennetsu* (formerly *Rickettsia sennetsu*, then *Ehrlichia sennetsu*). The fourth is *Wolbachia pipientis*, a bacterium associated with arthropods (insects, crustaceans, and acarids) and helminth worms (mainly filaria). These organisms elicit cross-reactive antibodies.

Ehrlichiae multiply exclusively in vacuoles of their eukaryotic cell host, where they form clusters known as morulae. The vacuoles are derived from phagosomes and help the organism escape bactericidal lysosomal fusion. In humans, ehrlichiae are associated with monocytes (*E. chaffeensis*, *E. canis*, *N. sennetsu*) or PMN cells (*A. phagocytophilum*, *E. ewingii*).

Ehrlichioses can be acquired through tick bites, by ingestion of nematodes through contaminated water or animals (fish, snails), or as a consequence of filariasis.

American Human Monocytic Ehrlichiosis (*Ehrlichia chaffeensis*)

EPIDEMIOLOGY

Human monocytic ehrlichiosis (HME) is caused by *E. chaffeensis.* This organism has been isolated or identified by PCR mainly in the United States in the southeastern, south central, and mid-Atlantic states and California (Table 327-3). In the United States, *A. americanum* (Lone Star tick) is the vector (Chapter 359), and the white-tailed deer is the main mammalian reservoir. Immature ticks are infected by blood while feeding on persistently bacteremic reservoirs. *E. chaffeensis* is transmitted transstadially (remaining with the same vector from one life stage to the next) in the tick and infects its next host (deer or human) during its next blood meal. The disease epidemiology reflects the tick habitat and activity, with most cases being contracted in the southern United States, in rural areas, and from April to September. In highly endemic areas, the incidence can reach 100 cases per 100,000 inhabitants. The severity is age dependent, which may explain the lower incidence reported in children. Males are more often affected than females, with a sex ratio of 4 : 1.

TABLE 327-3 EHRLICHIOSES AND ANAPLASMOSES

DISEASE	AGENT	VECTOR	GEOGRAPHIC REPARTITION
American monocytic ehrlichiosis	*Ehrlichia chaffeensis*	*Amblyomma americanum*	South central, southeastern, mid-Atlantic coastal states
Human granulocytic ehrlichiosis	*Anaplasma phagocytophilum*	*Ixodes ricinus*	Europe, China
		Ixodes scapularis	Northeast, upper Midwest, northern California
E. ewingii	*Ehrlichia ewingii*	*Amblyomma americanum*	South central, southeastern, mid-Atlantic coastal states
Japanese monocytic ehrlichiosis	*Neorickettsia sennetsu*	Helminth of the gray mullet?	Japan
E. canis	*Ehrlichia canis*	*Rhipicephalus sanguineus*	Venezuela

CLINICAL MANIFESTATIONS

The incubation lasts for 7 to 10 days after an identified tick exposure in 80% of cases. Patients have fever, headache, malaise, nausea, and anorexia. Untreated patients worsen and may require intensive care. Gastrointestinal tract involvement consisting of nausea, vomiting, diarrhea, and abdominal pain is common. Central nervous system infection is manifested in many forms from confusion to coma. A rash is observed in a third of cases and lymphadenopathy in a fourth. In severe forms, sepsis syndrome and multiple organ dysfunction syndrome may occur.

The white blood cell count typically shows leukopenia, caused by both lymphopenia and neutropenia. Thrombocytopenia is also frequently found; anemia may appear later. Coagulopathy may be observed in severe forms. Increases in serum enzymes, including AST, ALT, and LDH, may reflect organ involvement, as does an elevated serum creatinine. CSF examination in patients with neurologic symptoms reveals pleocytosis and increased protein levels. Cells may be monocytic or PMN. The prognosis depends on early antimicrobial treatment, but the fatality rate is still high at 2.5%. In persons coinfected with HIV, it may be most severe; in one series, 6 of 13 patients died.

DIAGNOSIS

The diagnosis of HME should be considered in patients with a history of tick exposure and unexplained fever. HME resembles RMSF, but rash is less frequent. Later in the disease, it can be misdiagnosed as anything that causes severe sepsis.

Leukopenia associated with thrombocytopenia and an increase in liver enzyme levels may establish the etiology. Careful examination of blood and CSF smears may help identify typical morulae. Treatment should be started in any suspected case. The diagnosis can be confirmed by culture in specialized laboratories using a canine cell line, DH82. However, PCR is more practical; confirmatory PCR using a second target gene is useful. Most cases are currently diagnosed serologically by a four-fold or greater increase in antibody titer or by seroconversion. The reference technique is IFA. A single titer of 25 is indicative of the diagnosis. There are cross-reactive antibodies among *Ehrlichia* species and with *A. phagocytophilum.* Western blotting may be valuable to distinguish among these bacteria.

TREATMENT

Rx

Doxycycline (100 mg twice daily for adults) is the drug of choice for patients with ehrlichiosis. The optimal duration of therapy has not been established, but current regimens recommend continuation of treatment for at least 3 days after the fever subsides and until evidence of clinical improvement, for a minimum total course of 5 to 7 days. Severe or complicated disease may require longer treatment courses. Because tetracyclines are contraindicated in pregnancy, rifampin has been used successfully in a limited number of pregnant women with documented HME.

Human Granulocytic Ehrlichiosis (*Anaplasma phagocytophilum*)

EPIDEMIOLOGY

The first human case of *A. phagocytophilum* infection was recognized in 1990. The disease is found in America, Asia, and Europe (Fig. 327-4). It is transmitted by *Ixodes scapularis* (eastern North America), *Ixodes pacificus* (western North America), *Ixodes ricinus* (Europe), and *Ixodes persulcatus* (Asia), the vectors of Lyme disease (Chapter 321), and its epidemiology is similar. Coinfection with the two diseases may occur. The temporal distribution of the disease parallels that of nymph tick activity, with two peaks in spring and autumn. Ticks are born free of *Ehrlichia* and are infected while feeding on bacteremic small mammals. Deer play a major role as hosts of adult ticks and reservoirs. In highly endemic areas, the incidence can reach 50 per 100,000

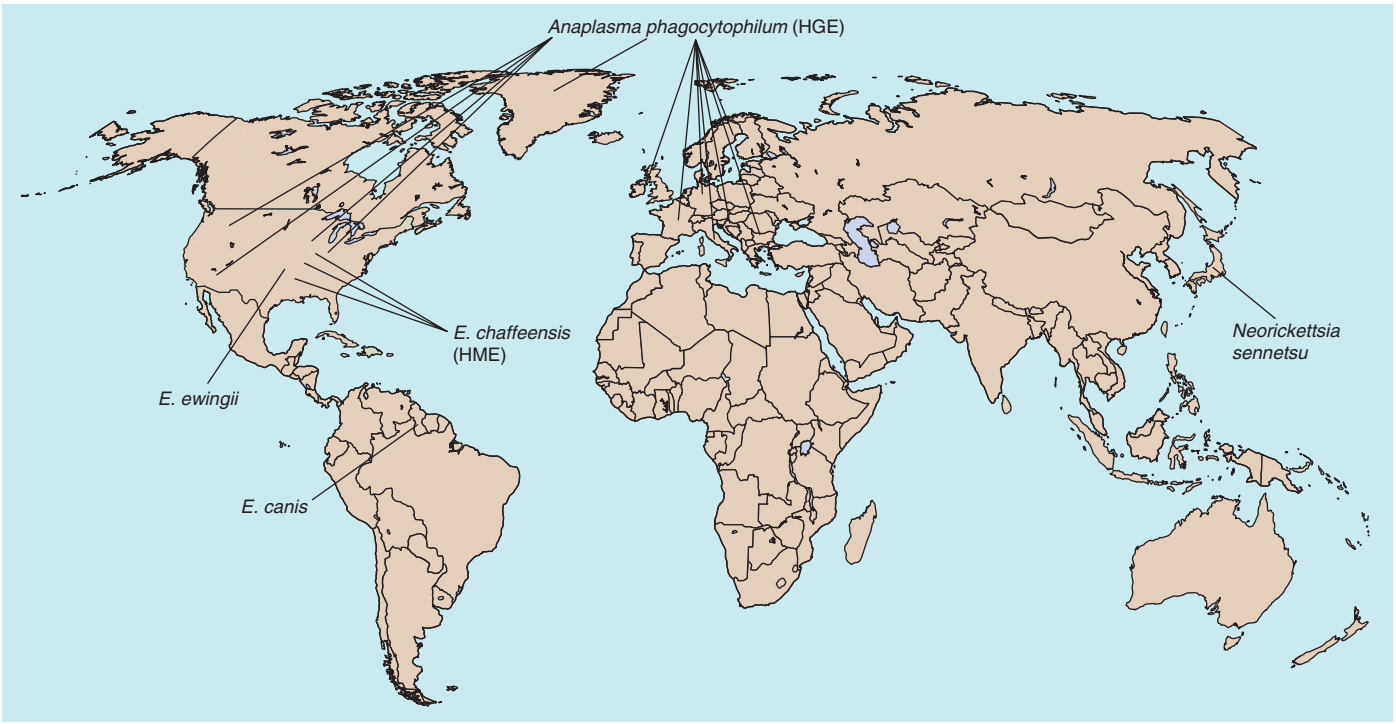

FIGURE 327-4. Geographic distribution of ehrlichioses. HGE = human granulocytic ehrlichiosis; HME = human monocytic ehrlichiosis.

inhabitants per year. The mean age of diagnosed patients is high, and males are more frequently infected than females, with a sex ratio of 3 : 1.

CLINICAL MANIFESTATIONS

The incubation time is usually between 7 and 10 days, and 80% of patients report a history of tick exposure. Many infections may be asymptomatic or too mild to require a diagnostic procedure. The disease frequently begins abruptly, with fever, headache, malaise, and myalgias that may be particularly severe. Rash is found in less than 10% of cases. Visceral involvement may be observed and includes digestive symptoms such as nausea, vomiting, and diarrhea. Neurologic symptoms may include confusion, meningitis, and meningoencephalitis.

The evolution of the disease is favorable in most cases, even without specific therapy, but the disease may evolve to septic shock in some individuals. Patients with underlying diseases are more at risk of dying. Most deaths are the consequence of *Anaplasma*-induced immunosuppression, and patients may experience invasive aspergillosis, candidiasis, cryptococcosis, or herpes esophagitis.

DIAGNOSIS

Laboratory findings consist of the association of thrombocytopenia and leukopenia (lymphopenia or neutropenia). An increase in serum transaminases is also frequent. The diagnosis can be made by careful examination of blood smears for morulae within PMN cells (Fig. 327-5). Culture from blood is possible in appropriate cells (HL-60), and PCR is useful as for HME. Most cases are diagnosed by serologic testing with IFA, which is comparable to that in HME (see earlier).

TREATMENT

Rx

Treatment is also similar to that of HME except that *A. phagocytophilum* is susceptible to fluoroquinolones in vitro, but these drugs have not been tested in patients.

Ehrlichia ewingii

Canine granulocytic ehrlichiosis, reported in the United States in 1972, is caused by *E. ewingii.* This bacterium was characterized by amplification and sequencing of the 16S ribosomal RNA gene. The vector of *E. ewingii* is *A. americanum* (Chapter 359), which also transmits *E. chaffeensis.* Among 60 cases of ehrlichiosis in Missouri in 1999, four were caused by *E. ewingii;* four other cases have been reported since by the CDC. The disease was prevalent in immunocompromised hosts (seven of eight) coinfected with HIV or receiving immunosuppressive drugs. Patients who report tick exposure are noted to have fever, thrombocytopenia, leukopenia, and various symptoms, including meningitis. Morulae may be seen on blood smears in PMN cells. The evolution in reported cases was good; patients responded dramatically to doxycycline. Patients have antibodies to *E. chaffeensis*, and PCR has been shown to be useful when it is applied to blood samples. This diagnosis should be considered when ehrlichiosis is suspected in immunocompromised patients exposed to *A. americanum* ticks.

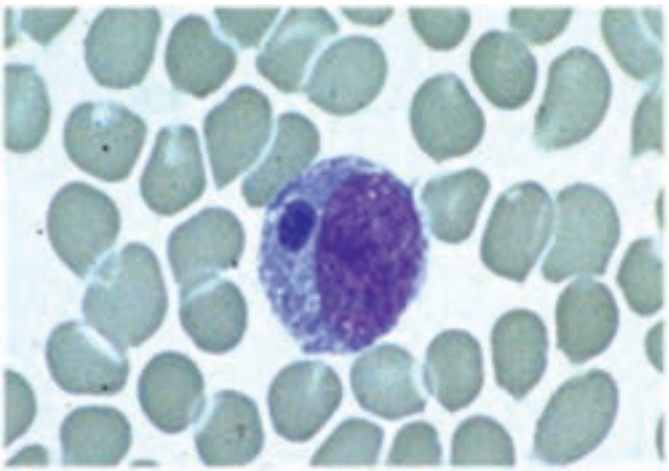

FIGURE 327-5. Peripheral blood leukocyte (monocyte) containing ehrlichial morula in a patient with human monocytic ehrlichiosis. (Courtesy Centers for Disease Control and Prevention. http://www.cdc.gov/ehrlichiosis/symptoms.)

Ehrlichia canis

Canine monocytic ehrlichiosis was reported first in Algeria in the 1930s. It is caused by *E. canis* and transmitted by the dog tick *R. sanguineus* (Chapter 359). This tick is found worldwide and is prevalent in temperate and hot areas. In 1996, a single case of infection was reported in an asymptomatic

man from Venezuela who owned an infected dog. Recently, cases have been reported in patients in South America.

A New *Ehrlichia* in Wisconsin

In four patients from Wisconsin, a new *Ehrlichia* species closely related to *E. muris* was cultured and identified[11] as *Ixodes scapularis* ticks. All patients had fever, malaise, headache and lymphopenia, and some also had thrombocytopenia and elevated transaminases.

Candidatus Neoehrlichia mikurensis

Candidatus Neoehrlichia mikurensis, transmitted by Ixodes ticks, has been reported based on PCR on the blood of febrile immunocompromised European patients and in China.

Wolbachia Species

Wolbachia bacteria are endosymbionts of arthropods and nematodes. They were known to be present in filarial worms, but it was later shown that they may play a role in human disease. These bacteria manipulate the fertility of their host. Eradication of *Wolbachia* in filariae may lead to infertility and stop the microfilariae from spreading. This effect was demonstrated by field treatment with doxycycline in patients with onchocerciasis. The patients improved when treated with this drug, which is effective on *Wolbachia* and subsequently on the worm's fertility but not on the worm itself. In 2001, it was shown that the adverse reactions observed after treatment of lymphatic filariasis may be caused by the release of *Wolbachia* from destroyed worms. Some authors suggested that eradicating *Wolbachia* before the anthelmintic prescription would avoid these reactions. For some reason, *Loa loa* (Chapter 358) do not harbor *Wolbachia*, and the genome of *Wolbachia* integrated in the *Brugia malayi* genome makes it inaccessible to therapy.

Q FEVER

DEFINITION

Q fever is a worldwide zoonosis caused by *Coxiella burnetii*. The name Q fever is derived from "query" to emphasize the surprising aspect of the disease first described in Queensland, Australia, in 1935 by Derrick. The infection in humans is variable in its severity, clinical expression, and natural course (i.e., acute or chronic). It is considered by the CDC to be a potential agent of bioterrorism. Ungulates and pets are the major sources of human infection.[12]

The Pathogen

C. burnetii is a gram-negative bacterium that naturally infects its host's monocytes. It multiplies in an acidic vacuole. Strains are heterogeneous genetically and antigenically and are associated with acute infections of variable severity. *C. burnetii* in vitro generates a deleted avirulent mutant also named phase II. This mutant exhibits diagnostic antigens that are useful because they are more reactive during acute infection.

C. burnetii is incompletely eliminated after acute infection. In immunocompromised hosts and patients with cardiac valve lesions, *C. burnetii* continues to multiply despite high levels of antibodies and causes chronic infection. Control of the disease in acute Q fever is associated with formation of a granuloma.

EPIDEMIOLOGY

C. burnetii infects a wide range of animals, including mammals, birds, and ticks. Ungulates and pets (cats and dogs) are the most common source of the disease. Mammals are infected through aerosols and may shed *Coxiella* in feces, urine, milk, and birth products. Humans are usually infected by aerosols or less frequently by exposure to milk products. Interhuman infections through sexual intercourse, during delivery, or by blood transfusion have been reported. *Coxiella* survives in the environment and can be spread far by the wind. In the past few years, major outbreaks were related to sheep and goats. The disease is partly seasonal and related to lambing time. The current geographic repartition is largely unknown. Males have more severe disease but are not more often exposed to Q fever, and middle-aged people are more frequently affected and hospitalized. The number of reported cases recently increased dramatically in Europe and Asia. Many American soldiers were infected during the war in Iraq. A giant outbreak was observed more recently in the Netherlands.

CLINICAL MANIFESTATIONS

After contamination by *C. burnetii,* 60% of patients seroconvert without apparent disease, 38% experience a self-limited disease, and only 2% require diagnostic evaluation. Months to years after the primary infection, a chronic infection associated with an immunocompromised situation, a cardiac valve lesion, or a vascular prosthesis or aneurysm develops in 0.2 to 0.5% of patients.

TABLE 327-4 SITUATIONS THAT SHOULD PROMPT SEROLOGIC TESTING FOR Q FEVER

ACUTE Q FEVER (PHASE II ANTIGEN AND IgG ≥ 200 AND IgM ≥ 50)
Fever in a patient in contact with ungulates
Unexplained prolonged fever (>7 days)
Granulomatous hepatitis
Fever and thrombocytopenia
Meningoencephalitis
Myocarditis
Erythema nodosum
Fever during pregnancy
Fever in a patient in contact with a parturient pet
Unexplained atypical pneumonia
Fever and an increase in transaminases (2-5 times the normal level)
Aseptic meningitis
Guillain-Barré syndrome
Pericarditis
Spontaneous abortion
CHRONIC Q FEVER (PHASE I ANTIGEN AND IgG ≥ 800 AND IgA ≥ 100)
Blood culture–negative endocarditis
Patient with a valvulopathy and unexplained
Fever
Weight loss
Fatigue
Increased erythrocyte sedimentation rate
Increased transaminases
Thrombocytopenia
Patient with unusually rapid degradation of a prosthetic valve
Fever in a patient with a vascular aneurysm or prosthesis
Aseptic osteomyelitis
Chronic pericarditis
Multiple spontaneous abortions

Patients with diagnosed acute infection may have a variety of symptoms (Table 327-4). Isolated prolonged fever was observed in 14% of more than 1000 patients. Pneumonia was found in 37% and was the only symptom in 17%. This percentage may vary according to the place of study and reach 90% of diagnosed cases. Some cases may be associated with respiratory distress. Hepatitis is found in 60% of patients and is the sole manifestation in 40%. The association of fever and a moderate increase in transaminases is an important clue. Some hepatitides, specifically in middle-aged men, are associated with an inflammatory syndrome and autoantibodies and may be resistant to antimicrobial treatment. Liver biopsy, when it is performed, exhibits granulomas that may be typified by a lipid vacuole and surrounded by a fibrinoid ring in the form of a doughnut. Less frequently (1.5% of cases), patients exhibit a rash. Patients can have specific neurologic manifestations such as meningitis, encephalitis, meningoencephalitis, or peripheral neuropathy. In 1 to 2% of cases, patients have cardiovascular manifestations such as pericarditis or, more rarely, myocarditis.

Evolution is usually favorable even without treatment, except in special hosts. In pregnant women, symptomatic or not, Q fever compromises the pregnancy. When infected during the first trimester, the patient usually aborts spontaneously. When the patient is infected later, the disease can result in fetal death or prematurity, or the outcome may be normal. Chronic uterine infection may develop in half the patients infected during pregnancy, and they may later experience multiple spontaneous abortions. From 30 to 50% of patients with heart valve or vascular lesions may experience chronic endocarditis within 2 years. This evolution is not prevented by regular treatment.

Patients with Q fever endocarditis have a chronic infection with low-grade fever, progressive degradation of valve function, and progressive heart failure. Fever is intermittent, and vegetations are frequently absent on echocardiography. Endocarditis is therefore not frequently considered in the initial differential diagnosis. If it is not diagnosed, the disease progressively worsens, and emboli (mainly cerebral) as well as renal insufficiency, splenomegaly, and hepatomegaly may be observed. Digital clubbing may also be seen. The main clue to the diagnosis in a patient with a valvulopathy is unexplained sickness (unexplained fatigue, weight loss, fever), a biological abnormality (leukopenia, increased erythrocyte sedimentation rate, thrombocytopenia, increase in

hepatic enzymes), or rapid degradation of a prosthetic valve. Chronic osteomyelitis, hepatitis, and infection of an aneurysm and vascular prosthesis have been reported.

Leukopenia may be observed; thrombocytopenia is frequent, as are increases in hepatic enzymes. Circulating anticoagulant associated with antiphospholipid antibodies may be observed, as may anti–smooth muscle antibodies. During endocarditis, antinuclear antibodies, microhematuria, and rheumatoid factor are frequently found.

DIAGNOSIS

The diagnosis is based mainly on serology (see Table 327-4). Direct detection by culture and PCR or immunochemistry in valve, liver, or blood samples is also useful, but serologic evaluation by IFA is the best method. Two antigens (phase I and phase II) can be tested. Acute Q fever is diagnosed when seroconversion or a four-fold increase is obtained with phase II antigen. A single serum test exhibiting IgG antibodies of 200 or greater and IgM of 50 or greater against phase II is also diagnostic. During chronic Q fever, antibodies are at higher titer and directed against both phase I and phase II. IgG against phase I at a titer of 800 or 1600 is diagnostic of chronic infection, as is IgA at 100 or greater. Serology is useful for follow-up of patients with acute Q fever and underlying disease and those with treated chronic Q fever.

TREATMENT

Rx

Treatment is easy during acute Q fever. Doxycycline is the most efficient antimicrobial and should be prescribed for 2 weeks. Some patients with hepatitis do not respond well because of an excessive immune response. They rapidly improve with a short course of glucocorticoids. In pregnant women, cotrimoxazole during the entire pregnancy may decrease the chance of an unfavorable outcome. As for endocarditis, bactericidal treatment is necessary. In vitro, antimicrobial efficacy is impaired by the low pH of the vacuoles in which *C. burnetii* resides. Hydroxychloroquine increases the pH of these vacuoles and restores the bactericidal effect of doxycycline. In patients with endocarditis, the recommended treatment is a combination of doxycycline (200 mg daily) and hydroxychloroquine (600 mg/day, then adjusted to reach a 1-mg/mL plasma concentration). This regimen is prescribed for 18 to 36 months according to serologic results. We recently observed that a more rapid favorable outcome was obtained with doxycycline serum levels higher than 5 µg/mL. Some strains may be resistant to doxycycline, and new macrolides may be an alternative. The major problem with this treatment is photosensitivity; sun exposure should be avoided. An alternative treatment is a combination of doxycycline and ofloxacin for 3 years or more.

PREVENTION

Prevention is based on veterinary control in animals. A vaccine is currently available in Australia.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Merhej V, Angelakis E, Socolovschi C, et al. Genotyping, evolution and epidemiological findings of Rickettsia species. *Infect Genet Evol*. 2014;25:122-137.
2. Sahni SK, Narra HP, Sahni A, et al. Recent molecular insights into rickettsial pathogenesis and immunity. *Future Microbiol*. 2013;8:1265-1288.
3. Openshaw JJ, Swerdlow DL, Krebs JW, et al. Rocky mountain spotted fever in the United States, 2000-2007: interpreting contemporary increases in incidence. *Am J Trop Med Hyg*. 2010;83:174-182.
4. Vaughn MF, Delisle J, Johnson J, et al. Seroepidemiologic study of human infections with spotted fever group Rickettsiae in North Carolina. *J Clin Microbiol*. 2014;52:3960-3966.
5. Parola P, Paddock CD, Socolovschi C, et al. Update on tick-borne rickettsioses around the world: a geographic approach. *Clin Microbiol Rev*. 2013;26:657-702.
6. Mediannikov O, Socolovschi C, Edouard S, et al. Common epidemiology of *Rickettsia felis* infection and malaria, Africa. *Emerg Infect Dis*. 2013;19:1775-1783.
7. Tsioutis C, Chaliotis G, Kokkini S, et al. Murine typhus in elderly patients: a prospective study of 49 patients. *Scand J Infect Dis*. 2014;46:779-782.
8. Janardhanan J, Trowbridge P, Varghese GM. Diagnosis of scrub typhus. *Expert Rev Anti Infect Ther*. 2014;12:1533-1540.
9. Ismail N, Bloch KC, McBride JW. Human ehrlichiosis and anaplasmosis. *Clin Lab Med*. 2010;30:261-292.
10. Pujalte GC, Chua JV. Tick-borne infections in the United States. *Prim Care*. 2013;40:619-635.
11. Pritt BS, Sloan LM, Johnson DK, et al. Emergence of a new pathogenic *Ehrlichia* species, Wisconsin and Minnesota, 2009. *N Engl J Med*. 2011;365:422-429.
12. Das I, Guest N, Steeds R, et al. Chronic Q fever: an ongoing challenge in diagnosis and management. *Can J Infect Dis Med Microbiol*. 2014;25:35-37.

REVIEW QUESTIONS

1. A 50-year-old man with acute Q fever presents after 7 days of fever, when he became spontaneously afebrile. You should:

A. Prescribe 15 days of doxycycline.
B. Evaluate clinically and/or echocardiographically if he has an underlying cardiovascular predisposing factor to chronic endocarditis.
C. Perform hepatic echography to detect hepatomegaly.
D. Explain to the patient that as he is spontaneously cured, he does not need to follow up.
E. Investigate the possible environmental source of the disease.

Answer: B It is critical in such patients to determine the risk of endocarditis. Progression to chronic endocarditis may be accompanied by only low-grade or intermittent fever, but if it is not diagnosed it can cause severe valve dysfunction, heart failure, and embolic complications.

2. A 6-year-old boy is diagnosed with Rocky Mountain spotted fever and presents with high fever. You should prescribe:

A. Treatment with doxycycline adjusted to weight
B. Chloramphenicol intravenously
C. Ciprofloxacin orally
D. Erythromycin intravenously
E. Ceftriaxone intravenously

Answer: A Doxycycline is recommended even in children for the treatment of potentially severe diseases. Oral treatment is effective.

3. A patient with body lice exposure and fever returns from Eastern Congo and you suspect louse-borne typhus. You should:

A. Immediately prescribe two pills of 100 mg doxycycline.
B. Prescribe intravenous chloramphenicol.
C. Sample blood for serology and wait for the result to treat.
D. Sample blood for polymerase chain reaction and wait for the result to treat.
E. Sample a stain biopsy and wait for the result to treat.

Answer: A This situation is an emergency; the patient should be treated as soon as the diagnosis is suspected. The diagnosis of typhus should be clinical because the fatality rate is high and the treatment safe and effective.

4. A pregnant women who travelled in Thailand received a diagnosis of scrub typhus. You should treat with which of the following:

A. Doxycycline
B. Ciprofloxacin
C. Chloramphenicol
D. Rifampicin
E. Streptomycin

Answer: D Because doxycycline is a U.S. Food and Drug Administration Pregnancy Category C drug ("use not recommended unless essential for the patient's welfare."), rifampicin is the reference treatment for scrub typhus when doxycycline is contraindicated.

5. Which of these tick bite diseases may be commonly superinfected by yeast or opportunistic bacteria?

A. Human monocytic ehrlichiosis
B. Infection by *E. canis*
C. Rocky Mountain spotted fever
D. Human granulocytic anaplasmosis
E. African tick bite fever

Answer: D Granulocytopenia favors superinfection by opportunistic pathogens.

328

ZOONOSES

STUART LEVIN AND KAMALJIT SINGH

DEFINITION

Zoonoses are classically defined as infections or diseases naturally transmitted between humans and vertebrate animals. There are approximately 1400 human pathogens, of which about 800 species are zoonotic. Zoonoses may be bacterial, viral, fungal, parasitic, or due to unconventional agents such as prions.[1] Of the approximate 400 emerging or reemerging pathogens in the past 70 years, 60% are known to be zoonotic, with a disproportionate number of the new zoonoses being caused by RNA viruses. Arthropod-transmitted diseases are prototypic zoonoses, but whereas infections such as West Nile fever, *Schistosoma japonicum*, *Trypanosoma rhodesiense*, and *Plasmodium knowlesi* malaria are zoonoses, other related diseases such as *Schistosoma mansoni*, *Trypanosoma gambiense*, and *Plasmodium vivax* malaria are not zoonoses, because these pathogens are transferred from human to human and do not depend on vertebrate reservoirs. Human infestations with ectoparasites are likewise not considered to be zoonoses.

We have little ability to predict how an emerging zoonotic pathogen will act. Most zoonotic pathogens like rabies enter a human and cause disease; however, in these cases, humans are dead-end hosts and are unlikely to transmit the disease to others. Less commonly, pathogens such as human immunodeficiency virus (HIV) or recombinant swine H1N1 (pH1N1/2009) enter the human population with great difficulty but are then successfully transmitted from human to human to cause global epidemics. The recent emphasis on emerging infections that are predominantly zoonoses and the potential for new pandemics markedly enhance the need for human medicine to be integrated with veterinary medicine and veterinary research. This is the concept of One Health, which calls for enhanced collaboration between physicians, veterinarians, and environmental health professionals for the optimal health of all species on the planet.[2,3]

EPIDEMIOLOGY

During the course of evolution, humans have sought the company of vertebrates for many reasons, including their companionship as pets. In the United States, fewer than 60% of households have pets; a surprising 56% of dog owners sleep with their dog next to them, whereas among cat owners, 75% sleep with their pets.[4] The over 100 million dogs and cats in the United States facilitate the transmission of more than 250 different infectious species and cause more than 1 million bite injuries each year. Other pets include birds, fish, and reptiles (particularly common are small pet turtles that can transmit salmonellosis). Another popular trend is the keeping of exotic animals as pets, many of which are important reservoirs of a variety of viral and bacterial human pathogens.

Almost all arthropod-transmitted infectious agents acquired from vertebrates in the United States are due to either ticks or mosquitoes. Ticks are the more common villain,[5] and Lyme disease is the most frequent arthropod-transmitted infectious disease in the United States (Chapter 321). At least five tick-borne pathogens are known to be transmitted by *Ixodes scapularis* (the principal tick vector for Lyme disease): *Borrelia burgdorferi*, *Anaplasma phagocytophilum* (an agent for anaplasmosis, previously known as human granulocytic ehrlichiosis), *Babesia microti* (an agent of human babesiosis), *Borrelia miyamotoi* (an agent of relapsing fever), and the Powassan encephalitis virus. Two or more of these pathogens can be transmitted by a single tick bite, resulting in 30 different permutations of mixed infections. In addition, two geographically distinct species of *Borrelia* (Chapter 322) cause relapsing fevers in the United States, with *Borrelia hermsii* described in the mountain states west of the Mississippi River and *Borrelia turicatae* in the southwest plains states. Threats from mosquito bites include almost all the vector-borne encephalitides, such as St. Louis encephalitis and West Nile virus (Chapter 383).

The recognition of novel zoonoses such as Hendra and Nipah (henipaviruses) and the SARS (severe acute respiratory syndrome) coronavirus has intensified research in wildlife reservoirs especially among the more than 1000 species of Chiroptera or bats. Bats have been known as the origin of a number of "old" viruses such as yellow fever virus (Chapter 381) and Japanese encephalitis virus but are now also recognized as the source of a succession of highly pathogenic emerging zoonoses, including Ebola, Marburg, and MERS (Middle East respiratory syndrome) coronavirus. (Dromedary camels have also been identified as hosts for MERS.) Another area of great interest is in the highly pathogenic avian influenza viruses (HPAIV) that spread from wild birds and ducks to chickens and humans, such as A/H5N1, with over 640 cases and 380 deaths since 2003, and other avian viruses, such as A/H7N9, with 132 laboratory-confirmed cases and 38 fatalities since March 2013. Emerging swine zoonoses include the 2009 novel swine-origin A/H1N1 virus that began in February 2009 in Veracruz, Mexico. Other novel swine influenza virus infections (e.g., vH3N2) have been documented in fairgoers and pigs in agricultural fairs in the United States since 2007.[6] In addition, living in close contact with swine, a common husbandry practice in Asia, has been associated with an emerging rapidly fatal meningitis due to *Streptococcus suis*.

The risk of contracting a zoonosis is increased by direct animal contact in meat handlers, poachers, hunters, exotic animal smugglers, and international travelers; by exposure to and inhalation of infectious air particles; by insect

TABLE 328-1 NEWER ZOONOSES IN THE UNITED STATES (EMERGING INFECTIONS)

DISEASE*	INFECTIOUS AGENT	CLINICAL FINDINGS	VECTOR/ACQUISITION
Ehrlichiosis, monocytic	*Ehrlichia chaffeensis* (Chapter 327)	Fever, myalgia, leukopenia; monocyte inclusions not often seen; maculopapular rash in a minority	*Amblyomma* (Lone Star) tick bite
Human granulocytic anaplasmosis (HGA)	*Anaplasma phagocytophilium* (Chapter 327)	Fever, myalgia, leukopenia; granulocyte inclusions often seen on blood smear	*Ixodes* (deer) tick bite
"Flu syndrome"	*Ehrlichia ewingii* (Chapter 327)	Immunocompromised host—fever, myalgia	*Amblyomma*
STARI (Southern tick-associated rash illness)	Unknown	Lyme disease–like illness	*Amblyomma*
Relapsing fever	*Borrelia hermsii* & *Borrelia turicatae* (Chapter 322)	Fever, headaches, myalgias	*Ornithodorus* ticks
Cat-scratch disease	*Bartonella* spp. (there are 8 identified zoonotic *Bartonella* spp.) (Chapter 315)	Cervical lymphadenopathy in normal hosts and cutaneous and hepatic angiomatosis in AIDS patients	Cat scratch or bite
Hemorrhagic diarrhea	Enterohemorrhagic *Escherichia coli* O157:H7 (other Shiga toxin–producing strains) (Chapter 304)	Rectal bleeding, dysentery, hemolytic-uremic syndrome	Contaminated undercooked meat
Hantavirus pulmonary syndrome	*Hantavirus*—Sin Nombre (Chapter 381)	Noncardiac pulmonary edema, elevated hematocrit	Fomites of wild rodents
Cryptosporidium diarrhea	*Cryptosporidium parvum* (Chapter 350)	Prolonged watery diarrhea	Contaminated water—cattle and sheep
Paragonimiasis	*Paragonimus kellicotti* (Chapter 356)	Diarrhea, cough, pleuritic chest pain, hemoptysis	Freshwater crayfish
Dysentery	*Campylobacter jejuni* (Chapter 303)	Dysentery, reactive arthritis, Guillain-Barré syndrome	Contaminated chicken
Pyogenic skin ulcer	*Capnocytophaga canimorsus* (Chapter 280)	Sepsis, skin infection	Dog bites
West Nile fever encephalitis	West Nile virus (Chapters 382 and 383)	Encephalitis, myelitis, Guillain-Barré syndrome, West Nile fever	Mosquito bite, birds
SARS	Coronavirus (Chapter 366)	Severe lower respiratory tract infection	Horseshoe bats
Cutaneous leishmaniasis	*Leishmania mexicana* (Chapter 348)	Skin nodules and ulcers	Wood rat

*See table of contents and index to locate a more detailed discussion of each disease.
AIDS = acquired immunodeficiency syndrome; SARS = severe acute respiratory syndrome.

bites, contact with previously infected human blood products, contact with and ingestion of infectious agents transmitted by animal-contaminated water; by insufficiently cooked meat, eggs, dairy products, and fish; and by acts of bioterrorism. Patrons of petting zoos, pet owners, farmers, hunters, laboratory researchers, cave explorers, hikers, and veterinarians, among others, are at higher risk for a zoonosis than the general population is. Infectious agents transmitted by these routes from animal sources essentially include members of all microbial classes: viruses and prions, bacteria and rickettsia, fungi, helminths, and protozoa. Immunocompromised hosts, such as splenectomized individuals, transplant recipients, patients with acquired immunodeficiency syndrome (AIDS), patients on chemotherapeutic agents (including biologics used in rheumatology and oncology), and pregnant women and their fetuses, are at higher risk for clinical disease when exposed to these various infectious agents. Chagas disease, a protozoal infection, is a zoonosis in South America but is not endemic to the United States (with only seven Autochthonous cases described since 1955); however, immigration from Latin America has led to "globalization" of this disease, with an estimated 300,000 persons with *Trypanosoma cruzi* infection in the United States and the potential for transmission of the infection through blood, birth, or organ donations. *Leishmania donovani* is almost always a small-animal reservoir zoonosis, but in India, it has become established in the human population, with female sandflies as the vector for human-to-human transmission.

As the world shrinks, warms, flattens, and remains in conflict, new infectious diseases seem inevitable, and previously rare ones are seen in unexpected places (Table 328-1) because of global warming, human intrusion into previously underexplored or never explored sites, world travel, deforestation, and the increasing threat of biological terrorism or warfare. Tens of thousands of years ago, human communities consisted of small groups of people living in vastly disparate regions that posed a very low risk for global epidemics. Increasing urbanization, intermingling of humans and domestic animals, and lack of adequate sanitation threaten to dramatically increase the incidence and severity of epidemic infectious diseases.

CLINICAL MANIFESTATIONS

Zoonoses can manifest as a variety of clinical syndromes, including respiratory disease (Table 328-2), central nervous system disease (Table 328-3), and rash or skin lesions (Table 328-4). Other zoonotic clinical syndromes may present as sepsis, polyarthritis, jaundice, acute renal failure, endocarditis, or diarrhea (Table 328-5).

Emerging zoonotic infectious diseases include Nipah virus and Hendra virus encephalitis, *Hantavirus* pneumonia and *Hantavirus* fever with renal failure, the SARS and MERS coronaviruses, West Nile encephalitis (which spread coast to coast in the United States in 3 years and is now apparently a permanent resident), monkeypox (an outbreak of which in the upper Midwest was caused by pet prairie dogs housed next to an imported Gambian rat infected with the virus), *Leishmania mexicana* in Texas, and at least 8 to 10 newly identified tick-borne rickettsial spotted fevers worldwide, including the flea-borne *Rickettsia felis* in California and tick-borne *Rickettsia parkeri* in the Gulf Coast states. Avian H5N1 and A/H7N9, pH1N1/2009 (swine) influenza, SARS and MERS coronaviruses, and West Nile fever have recently exposed the fragility of our world village. Nevertheless, exotic zoonoses remain a less frequent cause of fever in travelers than the well-known ordinary gastrointestinal and pulmonary pathogens.

DIAGNOSIS

Non–animal-associated environment- or travel-related infectious diseases can be confused with zoonoses. The vast majority of clinical diseases caused by *Legionella pneumophila*, *Entamoeba histolytica*, *Giardia lamblia*, *Burkholderia pseudomallei*, *Chromobacterium violaceum*, *Aeromonas hydrophila*, and airborne fungi such as *Blastomyces dermatitidis*, *Coccidioides immitis*, and *Histoplasma capsulatum* are acquired through environmental exposure and are only rarely related to animal hosts. *Sporothrix schenckii*, almost always an environmentally acquired pathogen stemming from vegetation-related injuries, has also been transmitted from cats with draining cutaneous ulcers to owners and animal handlers. Histoplasmosis has been acquired by explorers (spelunkers) in caves contaminated by bat guano.

Unfortunately, some descriptive disease titles can be misleading to clinicians and interfere with reaching the correct diagnosis. The transmission of tick-borne Rocky Mountain spotted fever actually occurs much more commonly in the southeastern United States than in the Rocky Mountains and has even occurred in the middle of New York City. In turn, the geographic range of Southern tick-associated rash illness (STARI), a Lyme disease–like infection transmitted by the Lone Star tick, *Amblyomma americanum*, has expanded to areas where *Ixodes scapularis* transmission of *B. burgdorferi*

TABLE 328-2 RESPIRATORY TRACT ZOONOSES

DISEASE*	MICROORGANISM†	CLINICAL FINDINGS	RESERVOIR AND/OR VECTOR
Psittacosis‡	*Chlamydophila psittaci* (Chapter 318)	Pneumonia, often severe	Aerosols from parrots, ducks, turkeys
Q fever	*Coxiella burnetii* (Chapter 327)	Pneumonia, hepatitis, myocarditis	Airborne from soil contaminated by sheep, goats, and cats, particularly if parturient
Tularemia	*Francisella tularensis* (Chapter 311)	Cutaneous ulcer and regional node, pneumonia and hilar node, pleural effusion	Rabbit contact (winter) and tick/fly bites
Plague	*Yersinia pestis* (Chapter 312)	Inguinal nodes, bubonic plague (basilar pneumonia develops in 10%), hilar node enlargement	Fleas from prairie dogs, rock squirrels, rats
Hantavirus cardiopulmonary syndrome	*Hantavirus* (Chapter 381)	Upper respiratory to lower respiratory to adult respiratory distress syndrome (ARDS) to death	Deer mouse fomites: urine, feces, saliva
Rhodococcus pneumonia	*Rhodococcus equi*	Pneumonia often cavitates in those with AIDS and in other immunosuppressed patients.	Horse manure, soil
Mycoplasma arginini pneumonia	*Mycoplasma arginini* (Chapter 317)	Pneumonia, sepsis, neutropenia	Sheep, goats
Foot-and-mouth disease	*Aphthovirus*	Nonspecific upper respiratory tract infection, oral vesicles	Cloven-footed mammals
Whooping cough	*Bordetella bronchiseptica* (Chapter 313)	Pneumonia, bronchitis, whooping cough	Dogs
Histoplasmosis	*Histoplasma capsulatum* (Chapter 332)	Pneumonia or fever of unknown origin	Bats
Anthrax	*Bacillus anthracis* (Chapter 294)	Mediastinal widening with CT scan; pneumonia often absent	Herbivore mammals
Glanders	*Burkholderia mallei* (Chapter 306)	Pneumonia, erosive tracheobronchitis	Horses, mules

*See table of contents and index to locate a more detailed discussion of each disease.
†Because of the fastidious nature of some organisms, the rapid development of diagnostic tools, and the risk some agents pose to laboratory workers, a clinical microbiologist should be consulted if these agents are considered in a patient's differential diagnosis.
‡Occurs in more than 1000 animal species.
AIDS = acquired immunodeficiency syndrome; CT = computed tomography.

TABLE 328-3 CENTRAL NERVOUS SYSTEM ZOONOSES

DISEASE*	ORGANISM	CLINICAL FINDINGS	ACQUISITION
Listeriosis	*Listeria monocytogenes* (Chapter 293)	Purulent meningitis during pregnancy, in patients > 65 yr, in neonates, and immunosuppressed	Unpasteurized cheese and other dairy products; cattle, goats
Leptospirosis	*Leptospira interrogans* (Chapter 323)	Aseptic meningitis, hepatorenal syndrome	Asymptomatic dogs, cattle; common water source
Herpes B encephalitis	Cercopithecine herpesvirus (herpes B virus) (Chapter 414)	Diffuse progressive encephalitis	Macaca monkey bites or scratches
Lyme disease	*Borrelia burgdorferi* (Chapters 321 and 322)	Lymphocytic meningitis, motor-sensory neuropathy, facial palsy	Tick bite
Lymphocytic choriomeningitis	Lymphocytic choriomeningitis virus (Chapter 412)	Lymphocytic meningitis, occasionally with pneumonia	Inhalation of mouse secretions: urine, feces, saliva
Mosquito-borne encephalitis (U.S.)	Eastern, Western equine, St. Louis, California encephalitis; West Nile virus (Chapter 383)	Diffuse encephalitis least severe; California encephalitis most severe; Eastern equine encephalitis, myelitis, meningitis	Mosquito-borne from horses, birds
Rabies encephalitis	Rabies virus (Chapter 414)	Encephalitis; almost always fatal	Bites from dogs, skunks, bats, raccoons, foxes
Toxoplasmosis	*Toxoplasma gondii* (Chapter 349)	Multiple brain masses in AIDS patients	Cat feces or ingestion of undercooked lamb or pork
Cerebral cysticercosis	*Taenia solium* (Chapter 354)	Epilepsy, CNS cysts, eosinophilic meningitis, hydrocephalus	Fecal-oral; contamination of food with pork tapeworm eggs
New-variant Creutzfeldt-Jakob disease	Prion (proteinaceous infectious particle) (Chapter 415)	Dementia, ataxia, myoclonus	Beef from cattle fed scraps from contaminated sheep carcasses
Nipah virus (Chapter 414)	Paramyxovirus	Acute encephalitis, death	Pig contact, fruit bats
Hendra virus	Paramyxovirus	Pneumonitis, encephalitis	Horses, fruit bats

*See table of contents and index to locate a more detailed discussion of each disease.
AIDS = acquired immunodeficiency syndrome; CNS = central nervous system.

occurs, resulting in misdiagnosis as Lyme borreliosis. Urban New York City continues to be a major source of rickettsialpox, where it was first described 60 years ago.[7] Vegetarians and other strict non–pork-eating persons have been seriously infected with the pig tapeworm *Taenia solium* as a result of fecal contamination of food from infected human food handlers. Recent outbreaks of food-borne zoonotic disease with Shiga toxin–producing *Escherichia coli*–infected brussels sprouts in Germany and *Listeria*-infected cantaloupes in the United States in 2011 highlight the difficulties in identifying the original source of the disease outbreak. The separation of an infection from the source of transmission, or a lack of knowledge, often blurs the diagnosis of a zoonoses. For example, unexpected outbreaks of leptospirosis among triathlon athletes exposed to water bodies that are contaminated by animal urine may not be recognized by physicians in a nonendemic setting. Emerging parasitic zoonosis such as the lung fluke, *Paragonimus kellicotti* (acquired by ingestion of contaminated crayfish from the Mississippi River basin), although unknown to most physicians and microbiologists, is well recognized by veterinarians in North America.

Until recently, human influenza A was not well recognized as a zoonosis; however, interspecies spread and mixing of swine, avian, and human influenza viruses can occur in unique geographic areas such as southern China or Mexico (pH1N1/2009), where dense concentrations of ducks, pigs, and people cohabit. Viral incubation of the three influenza species in the pig, with

TABLE 328-4 ZOONOSES CAUSING RASH OR SKIN NODULE, ULCER

DISEASE*	MICROORGANISM†	CLINICAL FINDINGS	RESERVOIR AND/OR VECTOR
Ehrlichiosis (monocytic)	*Ehrlichia chaffeensis* (Chapter 327)	Macular rash (seen in < half of patients) with a central distribution; prevalent in south-central United States	Tick bite
Leptospirosis	*Leptospira interrogans* (Chapter 323)	Central macular rash in 20%; occasional enanthem, conjunctival suffusion, hepatorenal syndrome	Urine-contaminated water; dogs, cattle
Lyme disease	*Borrelia burgdorferi* (Chapters 321 and 322)	Erythema migrans; 20% have multiple lesions	Mouse reservoir—tick bite
Rocky Mountain spotted fever	*Rickettsia rickettsii* (Chapter 327)	Acral or peripheral distribution of maculopapular to hemorrhagic rash to gangrenous lesions—no eschar	Tick bite
Typhus (epidemic)	*Rickettsia prowazekii* (Chapter 327)	Macular rash with a central distribution (can be hemorrhagic)	Flying squirrel fleas or fomites
Spotted fever	*Rickettsia parkeri* (Chapter 327)	Eschar, fever, headache, generalized rash involving palms and soles	*Amblyomma maculatum*
Cat-scratch disease	*Bartonella* spp. (Chapter 315)	Bacillary angiomatosis, peliosis hepatis in HIV patients, cervical adenopathy, subacute bacterial endocarditis, fever of unknown origin	Cat scratch or bite
Tularemia	*Francisella tularensis* (Chapter 311)	Ulcer and node, typhoidal syndrome, pneumonia	Rabbit contact and tick bite
Anthrax	*Bacillus anthracis* (Chapter 294)	Painless, edematous, nonpurulent ulcer that becomes black and necrotic over days; mediastinitis	Herbivore animal products, including animal hides
Rickettsialpox (spotted fever *Rickettsia*)	*Rickettsia akari*, *Rickettsia conorii*, *Rickettsia africae* (8 others found on all continents) (Chapter 327)	Fever, rash, eschar (multiple), tache noire (*R. africae*, often multiple)	House mouse mite
Monkeypox	Orthopoxvirus (Chapter 372)	Multiple maculopustular lesions with lymphadenopathy	Prairie dog (U.S.), African rodents
Erysipeloid	*Erysipelothrix rhusiopathiae* (Chapter 295)	Red, tender, swollen finger; subacute bacterial endocarditis	Skin pricked while cleaning fish, domestic meat animals

*See table of contents and index to locate a more detailed discussion of each disease.
†Because of the fastidious nature of some organisms, the rapid development of diagnostic tools, and the risk some agents pose to laboratory workers, a clinical microbiologist should be consulted if these agents are considered in a patient's differential diagnosis.

TABLE 328-5 OTHER ZOONOTIC CLINICAL SYNDROMES

DISEASE*	ORGANISM
Sepsis syndrome	*Streptococcus suis*, *Capnocytophaga canimorsus*, *Yersinia enterocolitica*, *Pasteurella multocida*, *Rickettsia rickettsii*, *Streptobacillus moniliformus*, *Rickettsia conorii*, *Leptospira icterohaemorrhagiae*
Polyarthritis	*Streptobacillus moniliformus*, *Borrelia burgdorferi*, *Spirillum minus*, *Brucella* spp., alpha virus—six groups (Ross River, Sindbis, Mayaro, chikungunya, o'nyong-nyong, igbo-ora)
Jaundice	*Coxiella burnetii*, *Leptospira icterohaemorrhagiae*, yellow fever (sylvatic), hepatitis E, echinococcosis, *Fasciola hepatica*
Renal failure (acute)	*Leptospira*, *Hantavirus*, *Rickettsia rickettsii*, *Rickettsia conorii*, *Escherichia coli* O157
Endocarditis	*Bartonella*, *Coxiella burnetii*, *Erysipelothrix rhusiopathiae*, *Brucella*, *Streptococcus suis*, *Listeria monocytogenes*, *Chlamydia psittaci*
Diarrhea	*Salmonella*, *Campylobacter jejuni*, Shiga toxin–positive *E. coli*, *Giardia intestinalis*, cryptosporidia, *Yersinia pseudotuberculosis*, *Yersinia enterocolitica*

*See table of contents and index to locate a more detailed discussion of each disease.

a reassortment of antigens and subsequent spread of virulent "new" influenza strains to humans, can lead to massive influenza pandemics that in sheer numbers (billions) surpass any past epidemics of smallpox or plague.[8] Initially, HIV-1 and HIV-2 were transmitted as simian immunodeficiency viruses from chimpanzees and mangabey primates, respectively, to humans. As with influenza, the subsequent 60 million (and counting) cases of HIV no longer require an animal reservoir for continuing transmission. Because these pandemics of AIDS and influenza now spread from human to human without the help of the initiating animal host, the World Health Organization no longer considers them zoonoses.

Leprosy, an illness of biblical notoriety transmitted from human to human, is endemic in at least three animal species, including the armadillo. This animal has rarely been implicated in the transmission of leprosy to humans in the United States.

Despite the large number of zoonoses described, clinicians evaluating an individual patient usually need to consider only a limited number of historical details to arrive at an appropriate differential diagnosis:

1. A history of direct contact with animals or animal products, animal bites, arthropod exposure, and food ingestion may offer clues to the correct cause.
2. All hobbies such as hunting and pet exposures (including contact with other owners' pets) should be elicited. Ownership of exotic pets or reptiles is often denied.
3. The patient's travel history must be considered because a number of zoonoses are still quite limited in geographic distribution.
4. Occupational and recreational high-risk activities such as cave exploration must be ascertained.
5. The patient's clinical manifestations (course and organ involvement) are used to focus on the most likely cause (see Tables 328-2 through 328-5).
6. Though neither specific nor sensitive, the inoculation nodule (eschar, tache noire, chancre, necrotic ulcer), with or without fever, rash, or local lymphadenopathy, is the only general clue from the physical examination that might alert the physician to a zoonosis when the history does not.
7. Significant unilateral hilar adenopathy in the presence of acute pneumonia can be a clue to anthrax, plague, or tularemia.

TREATMENT

Rx

The approach to treatment of the various zoonoses is presented in their respective chapters.

PREVENTION

Guidelines have been published to help prevent nosocomial transmission of zoonotic diseases. Preventive measures to decrease infection in compromised hosts include the routine immunization of pets, neutering of pets, use of caution when handling pet fomites, rigorous handwashing practices, and avoidance of ingesting undercooked meat, fish, shellfish, and eggs.

Isolation is recommended for anthrax, Andes *Hantavirus* disease, herpes B, monkeypox, Q fever, rabies, plague, SARS and MERS coronavirus (Chapter 366), influenza, and the hemorrhagic fever illnesses (Chapter 381)

TABLE 328-6 HIGHLY FATAL ZOONOSES

DISEASE*	FATALITY RATE (%)
Creutzfeldt-Jakob disease (new variant)	100
Rabies	100
Anthrax, inhalational	80-90
Herpes simiae	50-75†
Ebola virus	70
Eastern equine encephalitis	50-70
Hantavirus pulmonary syndrome (U.S.)	60‡
H5N1 influenza	60
Yellow fever (sylvatic)	20-50§
Lassa fever	15-25§
Plague	50-80†
Rocky Mountain spotted fever	20-60†
East African sleeping sickness	20-30†
Anthrax, cutaneous	20†
Tularemia, pneumonic	30-60†
Tularemia, cutaneous	2-10
Visceral leishmaniasis	5-25†
Louse-borne relapsing fever	5-40†

*See table of contents and index to locate a more detailed discussion of each disease.
†If untreated.
‡Case mortality of hospitalized patients.
§If jaundiced.

caused by Argentine, Bolivian, Crimean-Congo, Ebola, Lassa, and Marburg viruses.

PROGNOSIS

The prognosis of zoonoses varies widely, but a number of these diseases have very high case-fatality rates (Table 328-6).

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Akritidis N. Parasitic, fungal and prion zoonoses: an expanding universe of candidates for human disease. *Clin Microbiol Infect*. 2011;17:331-335.
2. Gebreyes WA, Dupouy-Camet J, Newport MJ, et al. The Global One Health paradigm: challenges and opportunities for tackling infectious diseases at the human, animal, and environment interface in low-resource settings. *PLoS Negl Trop Dis*. 2014;8:e3257.
3. Bidajsee S, Macpherson CN. Zoonoses and One Health: a review of the literature. *J Parasitol Res*. 2014;2014:8743-8745.
4. Chomel BB, Sun B. Zoonoses in the bedroom. *Emerg Infect Dis*. 2011;17:167-172.
5. Parola P, Paddock CD, Socolovschi C, et al. Update on tick-borne rickettsioses around the world: a geographic approach. *Clinical Microbiol Rev*. 2013;26:657-700.
6. Christou L. The global burden of bacterial and viral zoonotic infections. *Clin Microbiol Infect*. 2011;17:326-330.
7. Himsworth CG, Parsons KL, Jardine C, et al. Rats, cities, people, and pathogens: a systematic review and narrative synthesis of literature regarding the ecology of rat-associated zoonoses in urban centers. *Vector Borne Zoonotic Dis*. 2013;13:349-359.
8. Smith TC, Harper AL, Nair R, et al. Emerging swine zoonoses. *Vector Borne Zoonotic Dis*. 2011;11:1225-1234.

REVIEW QUESTIONS

1. The deer tick (*Ixodes scapularis*) is a known vector for which of the following infections:

A. Babesiosis
B. Human monocytic ehrlichiosis (HME)
C. Rocky Mountain spotted fever (RMSF)
D. Leishmaniasis
E. Chagas Disease

Answer: A HME and RMSF are transmitted by the bite of the Lone Star tick. Leishmaniasis is transmitted by the bite of a sandfly, and Chagas disease is spread by the kissing bug (Triatome).

2. Which of the following viral infections is associated with bats?

A. Henipaviruses
B. Dengue fever
C. West Nile virus
D. Avian H5N1 influenza
E. Herpes simiae

Answer: A Henipaviruses, which include the Hendra virus and Nipah virus, are paramyxoviruses found in fruit bats in Australia and Asia, respectively.

3. A 32-year-old man with a history of Hodgkin's lymphoma in remission presents to the emergency department with a 3-day history of fever, cough, and left-sided pleuritic chest pain. He reports no recent travel or pets. One week prior to presentation, he had Thanksgiving dinner at his mother's house, where he remembered feeding table scraps to her dog. On admission, he is found to be febrile with a left lower lobe pneumonia and pleural effusion. Blood cultures drawn on admission are positive for a fastidious gram-negative rod that is most likely:

A. *Streptococcus* pneumonia
B. *Haemophilus influenzae*
C. *Pseudomonas aeruginosa*
D. Methicillin-resistant *Staphylococcus aureus*
E. *Pasteurella multocida*

Answer: E Transient contact with pets, including being licked by an animal, can result in transmission of pet-associated zoonoses such as *Pasteurella multocida*. Presumably in this case, the patient contacted the dog's saliva with his feeding hand and proceeded to eat without washing his hands. Patients with prior hematologic malignancies may be slow to fully recover immune function after completion of chemotherapy, and this may represent an added risk factor for infection.

4. A 12-year-old Kenyan boy presents with a necrotic ulcer on his right arm that has been present for approximately 2 weeks. The lesion started as a nonpainful papule, which gradually enlarged and eroded to leave a painless ulcer with a black necrotic-appearing base. He does not recall any trauma and works on the family farm herding cattle. On examination, the patient also has a low-grade fever with associated swollen right axillary lymph nodes. The most appropriate treatment would be:

A. Penicillin
B. Trimethoprim-sulfamethoxazole (TMP-SMX)
C. Clindamycin
D. Azithromycin
E. Pentavalent antimony

Answer: A The patient has cutaneous anthrax, probably acquired from contact with cattle, and the treatment of choice is penicillin. TMP-SMX and clindamycin are reasonable treatment options for community-associated methicillin-resistant *Staphylococcus aureus* skin–soft tissue infections and impetigo. Azithromycin is used for treatment of cat-scratch disease or *Mycobacterium marinum* infection. Pentavalent antimony is generally used for treatment of cutaneous leishmaniasis (particularly New World cutaneous leishmaniasis).

5. A 50-year-old woman is admitted to the hospital for fevers and diffuse arthralgias and myalgias. She denies recent travel or pets. On examination, the patient has a faint diffuse rash over the extremities. Blood cultures drawn on admission are positive only at day 5 for a pleomorphic branching gram-variable rod that fails to grow on routine media. On further questioning, the patient admits to keeping a pet python and feeds it with rats. The most likely cause of her infection is:

A. *Neisseria gonorrhoeae*
B. *Streptococcus suis*
C. *Bacillus anthracis*
D. *Streptobacillus moniliformis*
E. *Borrelia burgdorferi*

Answer: D The patient has rat-bite fever due to *Streptobacillus moniliformis*.

329

ACTINOMYCOSIS

ITZHAK BROOK

DEFINITION

Actinomycosis is an uncommon, subacute to chronic bacterial infection that induces both suppurative and granulomatous inflammation. Localized swelling with suppuration, abscess formation, tissue fibrosis, and draining sinuses characterize this disease. The infection spreads contiguously and often forms draining sinuses that extrude characteristic but not pathognomonic "sulfur granules." Infections of the oral and cervicofacial regions are the most common, but any site in the body can be infected, including the thoracic region, abdominopelvic region, and central nervous system (CNS). Musculoskeletal or disseminated disease is rare but does occur.

The Pathogen

Actinomycetes of the genera *Actinomyces, Propionibacterium,* and *Bifidobacterium* act as the principal pathogens. However, 98 to 99% of actinomycoses are caused by non–spore-forming anaerobic or microaerophilic bacterial species of the genus *Actinomyces,* family Actinomycetaceae, order Actinomycetales.[1] Of the 30 *Actinomyces* spp, 8 may cause disease in humans: the strictly anaerobic *A. israelii, A. gerencseriae* (formerly known as *A. israelii* serotype II), *A. odontolyticus, A. naeslundii, A. meyeri, A. viscosus, A. pyogenes,* and *A. georgiae. A. israelii* is the most common species causing human disease. *Propionibacterium propionicum* (formerly known as *Arachnia propionica*) and *Bifidobacterium dentium* (formerly known as *Actinomyces eriksonii*) also are associated with clinically indistinguishable infection. The organisms are filamentous, branching, gram-positive, pleomorphic, non–spore-forming, non–acid-fast anaerobic or microaerophilic bacilli. *Actinomyces* organisms are fastidious bacteria that require enriched culture media; 6 to 10% ambient CO_2 may aid in their growth, which takes up to 2 to 3 weeks in culture. Most actinomycotic infections are polymicrobial and involve other aerobic and anaerobic bacteria. The most common co-isolates depend on the infection site and are *Actinobacillus actinomycetemcomitans, Aggregatibacter aphrophilus, Eikenella corrodens, Bacteroides, Fusobacterium, Capnocytophaga,* aerobic and anaerobic streptococci, *Staphylococcus,* and Enterobacteriaceae.

EPIDEMIOLOGY

Actinomyces spp are members of the endogenous mucous membrane flora in the oral cavity, lower gastrointestinal tract, bronchi, and female genital tract. No external environmental reservoir, such as soil or straw, has been documented, nor has person-to-person transmission of pathogenic *Actinomyces* spp been demonstrated. Although infection can occur in all age groups, it is rarely seen in children or patients older than 60 years. Most cases are encountered in individuals in the middle decades of life. A male-to-female infection ratio of 3 : 1 is reported in most series. The explanation for this ratio is the higher prevalence of poor oral hygiene and oral trauma in men. The annual reported incidence in the United States is fewer than 100 cases. However, because of the fastidious nature of the organism, many cases are undiagnosed and the true incidence is probably much higher.

PATHOBIOLOGY

Actinomyces spp are agents of low pathogenicity and require mucosal barrier disruption to cause disease.[2] Actinomycosis usually occurs in immunocompetent persons but may afflict those with diminished host defenses. Risk factors include steroids, bisphosphonates, leukemia with chemotherapy, human immunodeficiency virus (HIV), alcoholism, lung and renal transplant receipt, and local tissue damage caused by trauma, recent surgery, or irradiation. Oral and cervicofacial diseases are commonly associated with dental caries and extractions, gingivitis and gingival trauma, infection in erupting secondary teeth, chronic tonsillitis, otitis or mastoiditis, diabetes mellitus, immunosuppression, immunodeficiency, malnutrition, and neoplastic disease. Pulmonary infections generally arise after aspiration of oropharyngeal or gastrointestinal secretions and has been reported in patients with underlying lung disorders, such as emphysema, chronic bronchitis, and bronchiectasis. Gastrointestinal infection frequently follows loss of mucosal integrity, such as with surgery, trauma, foreign bodies, perforated appendix or diverticulitis, neoplasia, foreign bodies, and emergency colonic surgery. Extended use (>2 years) of intrauterine contraceptive devices (IUDs) increases risk for the development of actinomycosis of the female genital tract.

Other bacterial species that are frequently copathogens with *Actinomyces* spp may assist in the spread of infection by inhibiting host defenses and reducing local oxygen tension. Once the organism is established locally, it may spread progressively to surrounding tissues. The infection tends to spread without regard for anatomic barriers, including fascial planes and lymphatic channels. The end result is a chronic, indurated, suppurative infection (usually with draining sinuses and fibrosis, especially in pelvic and abdominal infection). The fibrotic walls of the mass before suppuration are "wooden" in nature and may be confused with a neoplasm. Hematogenous spread can be fulminant but is rare.

Actinomyces spp grow in microscopic or macroscopic clusters of tangled filaments surrounded by neutrophils. Plasma cells and multinucleated giant cells are often observed with lesions, as are large macrophages with foamy cytoplasm around purulent centers. When visible, these clusters are pale yellow and exude through sinus tracks; they are called sulfur granules (originally called drusen). These granules (1 to 2 mm in diameter) are made of aggregates of organisms and contain calcium phosphate. A central purulent loculation surrounds the granules. Their centers have a basophilic staining property, with eosinophilic rays terminating in pear-shaped "clubs." One to six granules can be present per loculation, and up to 50 loculations can be present in a lesion. Multicenter giant cells can be seen as well.

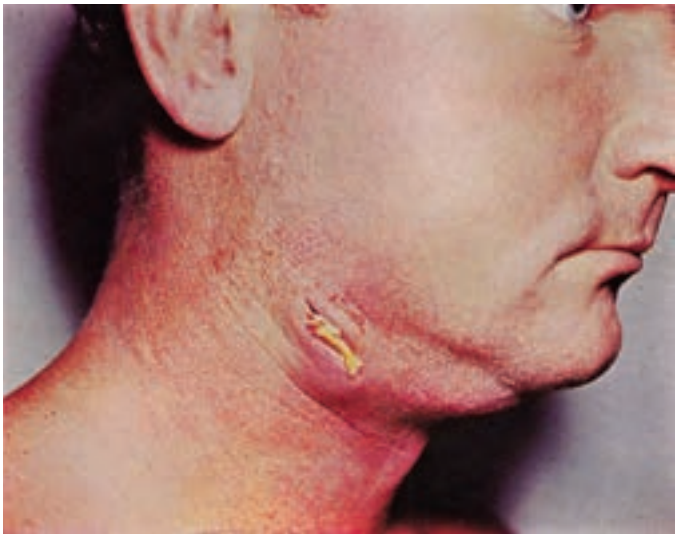

FIGURE 329-1. Actinomycosis of the jaw, observed at Letterman General Hospital, San Francisco, Calif., in a sergeant who had punctured the floor of his mouth while picking his teeth. (Courtesy Office of Medical History, Office of the Surgeon General, U.S. Army.)

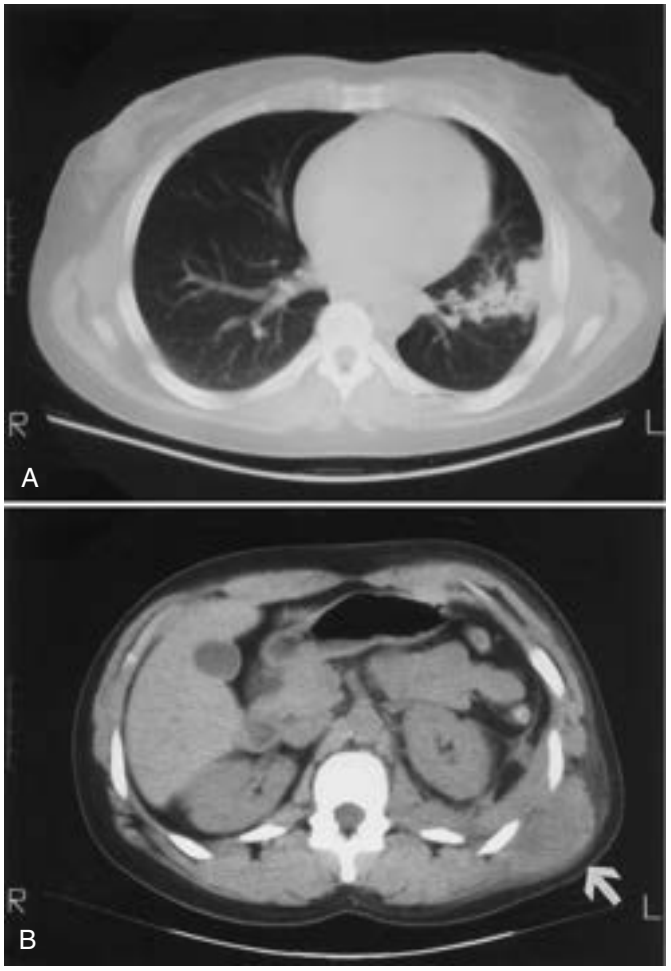

FIGURE 329-2. Thoracic computed tomography scan of a 43-year-old woman with pulmonary actinomycosis. A, There is consolidation of the lung with pleural thickening adjacent to the parenchymal disease. **B,** Abscess extended into the left breast and inferiorly to the costophrenic sulcus, to the retroperitoneum, and into the lateral abdominal wall (*arrow*).

CLINICAL MANIFESTATIONS

Cervicofacial

Cervicofacial infection is the most common manifestation of actinomycosis (Fig. 329-1).[3] It is generally odontogenic and evolves as a chronic or subacute painless or painful soft tissue slowly progressive, nontender swelling or mass involving the submandibular or paramandibular region. However, the submental and retromandibular spaces, temporomandibular joint, cheek, chin, and upper jaw can be involved. The swelling may have a ligneous consistency caused by tissue fibrosis. Depending on the composition of the concomitant synergistic flora, the onset of actinomycosis may be acute, subacute, or chronic. When *Staphylococcus aureus* or β-hemolytic streptococci are involved, an acute painful abscess or a phlegmatous cellulitis may be the initial manifestation. Pain and trismus can be disproportionate to the degree of inflammation apparent. The chronic form of the disease is the most common form and is characterized by painless infiltration and bluish or reddish induration that generally progresses to form multiple abscesses and draining sinus tracts discharging pus that may contain sulfur granules in up to 25% of instances. Periapical infection, trismus, dyspnea, dysphagia, fever, pain, and leukocytosis may be present. The infection can extend to the carotid artery, tongue, sinuses, ears, mastoid, orbit, salivary glands, pharynx, masseter muscle, thyroid, larynx, trachea, or thorax. Bone (most commonly the mandible) may be invaded from the adjacent soft tissue and results in periostitis or osteomyelitis. Cervical spine or cranial bone infection may lead to subdural empyema and invasion of the CNS. The differential diagnosis includes tuberculosis (scrofula), fungal infections, nocardiosis, suppurative infections by other organisms, and neoplasms.

Thoracic

Thoracic actinomycosis is an indolent, slowly progressive process involving the pulmonary parenchyma and pleural space.[4,5] This form accounts for 15 to 30% of actinomycosis cases and is caused by aspiration of infective material from the oropharynx, as well as rarely after esophageal perforation, by extension into the mediastinum from the neck, by spread from an abdominal site, or by hematogenous spread to the lung. Infection can spread from a pneumonic focus across lung fissures to involve the pleura and the chest wall, with eventual fistula formation and drainage containing sulfur granules (Fig. 329-2). The mediastinum, pericardium, and myocardium also rarely can be affected. Granules are seldom present in sputum. The incidence of this complication, as well as the destruction of thoracic vertebrae and adjacent ribs, has declined in the antimicrobial era.

The complaints of patients with thoracic actinomycosis are nonspecific. The most common are chest pain, productive cough, dyspnea, weight loss, and fever. Anemia, mild leukocytosis, and elevated sedimentation rate are relatively common. There is often a history of underlying lung disease, and patients are rarely initially seen in an early stage of infection. The pulmonary lesion is either a mass lesion or pneumonitis and may resemble tuberculosis, especially when cavity formation occurs, or blastomycosis, which may destroy ribs posteriorly but rarely forms sinuses.[6] Nocardiosis, bronchogenic carcinoma, cryptococcosis, aspiration pneumonia, pulmonary infection, and lymphoma also can mimic thoracic actinomycosis. Pleural thickening, effusion, or emphysema is common.

Abdominal

Abdominal actinomycosis makes up approximately 20% of actinomycosis.[7] It is a chronic, localized inflammatory process that can occur weeks, months, or years after the integrity of the gastrointestinal mucosa is breached by surgery or trauma. Extension from the thorax or pelvis or through hematogenous dissemination also can occur. The ileocecal region is involved most frequently (usually after a perforated appendix) with the formation of a mass lesion. The infection extends slowly to contiguous organs, especially the liver, and may involve retroperitoneal tissues, the spine, or the abdominal wall. Hepatic, renal, or splenic dissemination is an uncommon complication. Persistent draining sinuses may form, and those involving the perianal region can simulate Crohn's disease (Chapter 141) or tuberculosis. The extensive fibrosis of actinomycotic lesions, recognized by the examiner as a mass, often suggests tumor. A frequent finding on computed tomography (CT) is an infiltrative mass with dense inhomogeneous contrast medium enhancement. Constitutional symptoms and signs are nonspecific, the most common being fever, diarrhea or constipation, weight loss, nausea, vomiting, pain, and sensation of a mass.

Pelvic

Pelvic infection is observed in patients with prolonged use of IUDs and also may occur from extension of intestinal infection, commonly from indolent ileocecal disease.[8] Manifestations may range from a chronic vaginal discharge

to pelvic inflammatory disease with tubo-ovarian abscesses or pseudomalignant masses. Patients generally have abnormal vaginal bleeding or discharge, abdominal or pelvic pain, menorrhagia, fever, and weight loss.

Endometritis is the earlier form of the infection, followed by tubo-ovarian abscesses. Extension to the uterus, bladder, rectal area, abdominal wall, peritoneum, pelvic bones, thorax, and systemic circulation also can occur.

Central Nervous System

Infections of the CNS are very rare and generally manifest as single or multiple encapsulated brain abscesses that appear as ring-enhancing lesions with a thick wall that may be irregular or nodular on contrast-enhanced CT scans. There are no features that readily distinguish actinomycosis from other causes of brain abscesses. Rarely, solid nodular or mass lesions termed *actinomycetomas* or *actinomycotic granulomas* are found. Headache and focal neurologic signs are the most common finding. Most actinomycotic infections of the CNS are seeded hematogenously from a distant primary site, but direct extension of cervicofacial disease also occurs. Sinus formation is not a characteristic of CNS disease. The rare meningitis caused by *Actinomyces* is chronic and basilar in location, and the cerebrospinal fluid pleocytosis is usually lymphocytic. Thus, it may be misdiagnosed as tuberculous meningitis. Extension to the cranial epidural or subdural space and spinal epidural space also can occur from adjacent foci.

DIAGNOSIS

Appropriate microbiologic and pathologic studies are essential for diagnosis. A high index of suspicion should be communicated to the microbiology laboratory, along with material from draining sinuses, from deep-needle aspiration, or from biopsy specimens. CT or ultrasound needle aspiration can be used to obtain a biopsy specimen. It is important to avoid contamination of the specimen by normal flora and administration of antimicrobial therapy before a specimen is obtained. Anaerobic culture is required, and no selective media are available to restrict overgrowth of the slow-growing *Actinomyces* by associated microflora. The presence, in pus or tissue specimens, of non–acid-fast, gram-positive organisms with filamentous branching is suggestive of the diagnosis. The characteristic morphologic features of sulfur granules within are helpful. In tissue sections stained with hematoxylin-eosin, sulfur granules are round or oval basophilic masses with a radiating arrangement of eosinophilic terminal clubs. However, *Actinomyces* spp are infrequently visible in sections stained with hematoxylin-eosin; visualization is facilitated by special stains such as Gomori methenamine silver, *p*-aminosalicylic acid, McCallen-Goodpasture, and Brown-Brenn. Multiple biopsy sections from different tissue levels are recommended to improve the histopathologic diagnosis. The granules must be distinguished from similar structures that are sometimes produced in infections caused by *Nocardia, Monosporium, Cephalosporium, Staphylococcus* (botryomycosis), and others. *Actinomyces* and *Arachnia* can generally be differentiated from other gram-positive anaerobes by means of their growth rate (slow), by catalase production (negative, except for *A. viscosus*), and by gas-liquid chromatographic detection of the acetic, lactic, and succinic acids produced in peptone-yeast-glucose broth. Specific staining with fluorescent-conjugated monoclonal antibody testing can be used, but this method is not readily available to clinical microbiology laboratories.

Imaging methods such as conventional radiography, CT, and magnetic resonance imaging do not provide a specific diagnosis but allow more accurate definition of the dimensions and extension of the infection.

TREATMENT

Rx

Prolonged antimicrobial therapy (i.e., 6 to 12 months) has typically been recommended for patients with all clinical forms of actinomycosis to prevent disease recrudescence. However, individualization of courses of therapy is recommended because the duration of treatment depends on the initial burden of disease, the site of infection, and the clinical and radiologic response. Adequate drainage is indicated if abscesses are present.

Penicillin G is the drug of choice for treatment of an infection caused by any of the *Actinomyces.* It is given in high dosage during a prolonged period because the infection has a tendency to recur. Most deep-seated infections can be expected to respond to intravenous penicillin G, 18 to 24 million units/day given for 2 to 6 weeks, followed by an oral phenoxypenicillin in a dosage of 2 to 4 g/day. A few additional weeks of oral penicillin therapy may suffice for uncomplicated cervicofacial disease; complicated cases and extensive pulmonary[9] or abdominal disease may require treatment for 12 to 18 months. Little evidence exists of acquired resistance to penicillin G by *Actinomyces* during prolonged therapy. The combination of a penicillin (i.e., amoxicillin, piperacillin) and a β-lactamase inhibitor (i.e., clavulanate, tazobactam) offers the advantage of coverage against penicillin-resistant aerobic and anaerobic copathogens. Alternative first-line antibiotics include amoxicillin, tetracycline, erythromycin, and clindamycin. Ceftriaxone, imipenem, and fluoroquinolones have also been used successfully. Metronidazole, aminoglycosides, oxacillin, and cephalexin are not effective. In vitro antimicrobial susceptibility testing of *Actinomyces* is difficult, and the results may not be predictive of antimicrobial effects in vivo.

The need to use combination antimicrobial therapy to eradicate microorganisms that are isolated in association with *Actinomyces* has not been established. However, because many of these organisms are known pathogens, treatment is usually appropriate, especially with lower abdominal infections. Surgical removal of infected tissue also may be necessary in some cases, especially if extensive necrotic tissue or fistulas are present, if malignant disease cannot be excluded, and if large abscesses cannot be drained by percutaneous aspiration. When well-defined IUD-related symptoms and Papanicolaou smears demonstrate *Actinomyces* by specific fluorescence-labeled antibody, the IUD should be removed. Antimicrobial administration for a 2-week period may be indicated. More serious infections require prolonged therapy.

PROGNOSIS

The availability of antimicrobial treatment has greatly improved the prognosis for all forms of actinomycosis. At present, cure rates are high and neither deformity nor death is common.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Sullivan DC, Chapman SW. Bacteria that masquerade as fungi: actinomycosis/nocardia. *Proc Am Thorac Soc*. 2010;7:216-221.
2. Wong VK, Turmezei TD, Weston VC. Actinomycosis. *BMJ*. 2011;11:343.
3. Valour F, Sénéchal A, Dupieux C, et al. Actinomycosis: etiology, clinical features, diagnosis, treatment, and management. *Infect Drug Resist*. 2014;7:183-197.
4. Kim SR, Jung LY, Oh IJ, et al. Pulmonary actinomycosis during the first decade of 21st century: cases of 94 patients. *BMC Infect Dis*. 2013;13:216.
5. Song JU, Park HY, Jeon K, et al. Treatment of thoracic actinomycosis: a retrospective analysis of 40 patients. *Ann Thorac Med*. 2010;5:80-85.
6. Heo SH, Shin SS, Kim JW, et al. Imaging of actinomycosis in various organs: a comprehensive review. *Radiographics*. 2014;34:19-33.
7. Triantopoulou C, der Molen AV, Es AC, et al. Abdominopelvic actinomycosis: spectrum of imaging findings and common mimickers. *Acta Radiol Short Rep*. 2014;3:2047981614524570.
8. Choi MH, Hong DG, Seong WJ, et al. Pelvic actinomycosis confirmed after surgery: single center experience. *Arch Gynecol Obstet*. 2010;281:651-656.
9. Park JY, Lee T, Lee H, et al. Multivariate analysis of prognostic factors in patients with pulmonary actinomycosis. *BMC Infect Dis*. 2014;14:10.

REVIEW QUESTIONS

1. Which of the following statements is correct regarding treatment of actinomycosis?

A. Treatment should be guided by in vitro antimicrobial susceptibility testing.
B. Prolonged therapy with penicillin is limited by increasingly frequent acquisition of resistance to it.
C. Cervicofacial and thoracic actinomycosis generally requires combination antimicrobial therapy.
D. Penicillin G is the drug of choice and should be given in high doses for a prolonged period because the infection has a tendency to recur.
E. Surgical intervention is almost never required if antibiotic therapy selection and duration are correctly prescribed.

Answer: D Penicillin G is the drug of choice for treatment of an infection caused by any of the actinomycoses. The infection has a tendency to recur, so high doses for prolonged periods are generally required. Little if any evidence exists for significant acquired resistance to penicillin G during prolonged therapy. The need for combination antimicrobial therapy to eradicate actinomycosis has not been established; however, it is often considered clinically appropriate, especially with lower abdominal infections. Effective antibiotic therapy does not replace the need for surgical intervention when needed, such as removal of necrotic tissue, fistula repair, and drainage. In vitro antimicrobial susceptibility testing of *Actinomycosis* is difficult and the results may not be predictive of in vivo effects.

2. Cervicofacial actinomycosis is:

A. A rare manifestation of the infection.
B. Usually caused by extension of primary intrathoracic involvement.
C. Acutely accompanied by disproportionately severe symptoms of pain and trismus.
D. Not prone to spread beyond the soft tissues of the neck, although its extension to those soft tissues can be extensive.
E. A chronic disease.

Answer: C Cervicofacial infection is the most common manifestation of actinomycosis. It is generally odontogenic (not intrathoracic) in origin. Actinomycosis can be acute, subacute, or chronic. In its acute form, the symptoms of pain and trismus are disproportionate for the signs of inflammation that is apparent. Chronic actinomycosis can extend widely, including to the carotid arteries, tongue, sinuses, mastoid, orbit, thyroid, larynx, trachea, and contiguous muscle and bone.

3. Which of the following statements is correct regarding the epidemiology of actinomycosis?

A. There is no person-to-person transmission.
B. External reservoirs include soil and straw.
C. Most diagnosed individuals are elderly.
D. Males and females are comparably affected.
E. It predominantly occurs in tropical climates.

Answer: A There is no evidence for person-to-person transmission or for any external environmental reservoirs (e.g., soil or straw) for actinomycosis. Ages of infected patients can range from childhood to old age, but peak incidence is in the middle decades of life. A male-to-female predominance (3 : 1 ratio) has been attributed to the increased prevalence of poor oral hygiene and oral trauma in males (because *Actinomyces* spp are endogenous flora of the oral cavity).

330

NOCARDIOSIS

FREDERICK S. SOUTHWICK

DEFINITION

Nocardiosis refers to infections caused by *Nocardia* spp. *Nocardia* most commonly causes pneumonia but also can infect the central nervous system (CNS) and the skin. Less commonly, this organism can disseminate throughout the body. These infections usually occur in patients with defective immunity.

Etiology

Nocardia spp are thin, aerobic, gram-positive bacilli that form branching filaments. The bacteria stain irregularly and appear beaded on Gram stain. The speciation of *Nocardia* has been problematic. The original classification was based on the ability to use specific nutrients and decompose substrates such as adenine, casein, urea, gelatin, and xanthine. However, gene sequencing and DNA-DNA hybridization have now defined the true taxonomy. The species called *N. asteroides* was previously reported to be the most common cause of human disease. However, the majority of these bacteria were misidentified by today's standards. The number of species causing human disease is large and includes *N. abscessus, N. brevicatena/paucivorans* complex, *N. nova* complex, *N. transvalensis* complex, *N. farcinica, N. cyriacigeorgica, N. otitidiscaviarum, N. veterana, N. brasiliensis,* and *N. pseudobrasiliensis*.

EPIDEMIOLOGY

Nocardia spp are ubiquitous and primarily originate in soil. Despite being found throughout the environment, they rarely cause symptomatic infection in humans. Because nocardiosis is not a reportable disease, the frequency of this disease is unknown. The annual incidence has been estimated to be 0.4 in 100,000. The risk for symptomatic *Nocardia* infection is greatly increased

(estimated to be 140 to 340 times greater) in individuals who are immunocompromised, including patients who are receiving immunosuppressive agents following bone marrow or solid organ transplant, and patients with acquired immunodeficiency syndrome (AIDS). Corticosteroids are the most frequent immunosuppressant associated with nocardiosis[1]; however, cases also have been reported in patients receiving anti–tumor necrosis factor-α antibody (infliximab) as well as other immunosuppressants. It is important to keep in mind that trimethoprim-sulfamethoxazole prophylaxis does not always protect against *Nocardia*. Other risk groups include patients with cancer, Cushing's disease, chronic granulomatous disease, and dysgammaglobulinemia. Patients with chronic pulmonary disorders, particularly alveolar proteinosis, are also more susceptible to this infection. In approximately one third of patients with nocardiosis, no predisposing condition can be identified.

PATHOBIOLOGY

Most *Nocardia* spp gain entry to the host via the respiratory tract or less commonly by skin inoculation. Invading bacteria elicit a neutrophil response that inhibits but does not kill the organism. The bacteria are phagocytosed by neutrophils and macrophages and become enclosed in a membrane-bound phagolysosome. In this closed environment neutrophils and macrophages are able to kill many species of bacteria by synthesizing superoxide and hydrogen peroxide. However, *Nocardia* are able to survive in this hostile environment by producing superoxide dismutase, an enzyme that inactivates these toxic oxygen byproducts. In addition, *Nocardia* spp produce a mycolic acid called cord factor that inhibits the fusion of lysosomes with the phagolysomal compartment, preventing toxic proteases and other antibacterial products from reaching the intracellular bacteria. Cell wall extractable lipids impair phagocytosis and also inhibit bacterial killing.[2] In addition to neutrophils and macrophages, cell-mediated[3] and humoral immunity also play roles in protecting the host against *Nocardia* invasion, explaining the wide range of immunocompromised patients that are at increased risk for contracting nocardiosis.

CLINICAL MANIFESTATIONS

Nocardiosis has no pathognomonic characteristics, and delays in diagnosis are common. Failure of a pulmonary or skin infection to respond to conventional antibiotic therapy should raise the possibility of a *Nocardia* spp infection. Nocardiosis always should be considered in the immunocompromised patient.

Pulmonary Nocardiosis

Approximately two thirds of patients with nocardiosis present with pulmonary infection.[4,5] Pulmonary disease is usually subacute in onset, mimicking a fungal or mycobacterial infection, and is most commonly misdiagnosed as tuberculosis. The most common complaints are a persistent cough producing purulent sputum, fever, anorexia, and weight loss. Less commonly, patients may report pleuritic chest pain and dyspnea. Hemoptysis is rare but can develop in patients with large cavitary lesions. Acute onset of pneumonia has occasionally been reported in the immunocompromised host.

Central Nervous System Infection

Approximately 5% of patients with a *Nocardia* infection have CNS involvement.[6] Multilocular brain abscess is the most common CNS manifestation[7] and is usually the consequence of transient bacterial dissemination from the lung. Lesions can occur in any region of the brain, and symptoms depend on location. Headache is the usual initial complaint and frequently localized to the site of the abscess. Patients also may present with neurologic deficits and seizures. The combined findings of a lung nodule on chest x-ray and a ring-enhancing CNS lesion are often mistaken for metastatic lung carcinoma. Other diagnoses that should be considered when the immunocompromised host presents with both a lung and CNS focus are disseminated aspergillosis and toxoplasmosis. Meningitis is a less common CNS manifestation and is often associated with brain abscess (40% of meningitis cases). The CSF cell count usually reveals neutrophils and the CSF culture may be positive, particularly if the culture is held for a prolonged period.

Cutaneous Infection

Skin infection is usually caused by *N. brasiliensis* and typically follows a break in the skin that is contaminated by soil. Cutaneous disease has been reported in association with trauma, a postoperative wound, insect bites, thorn bush scratches, or even a cat scratch. Initially a pustule or a moderately erythematous, nonfluctuant nodule develops at the site of inoculation. Erythema can extend along the lymphatic system and is associated with tender lymphadenopathy.[8] This form of cutaneous infection has been termed *lymphocutaneous* or *sporotrichoid* disease. Similar skin manifestations are seen with other etiologies, including cat scratch disease, tularemia, *Mycobacterium marinum*, and sporotrichosis. In the immunocompromised host, disseminated infection may be manifest by multiple erythematous raised nodules and is an ominous finding. In tropical regions of South and Central America *Nocardia* spp can cause ulcerations and large tumor-like lesions called *mycetomas* that are usually found on the lower legs.

DIAGNOSIS

Radiology

In pulmonary disease, chest x-ray findings are variable, with pulmonary nodules or mass lesions being seen most often (Fig. 330-1). Less frequently, consolidation, cavitary lesions with air-fluid levels, interstitial infiltrates, and pleural effusions are found. Chest CT often demonstrates areas of low attenuation within consolidations, multiple nodules, and chest wall extension of the infection. Patients with AIDS are more likely to have multiple pulmonary nodules, cavitary lesions, and upper lobe infiltrates. In some patients the infiltrate may resolve, particularly in patients with normal immune function. However, the patient may present with brain abscess several months later as a consequence of transient dissemination. In CNS infection CT or magnetic resonance imaging with contrast usually demonstrates one or more ring-enhancing lesions (Fig. 330-2). *Nocardia* brain abscess is more commonly multiloculated; otherwise the radiologic findings are similar to those in other

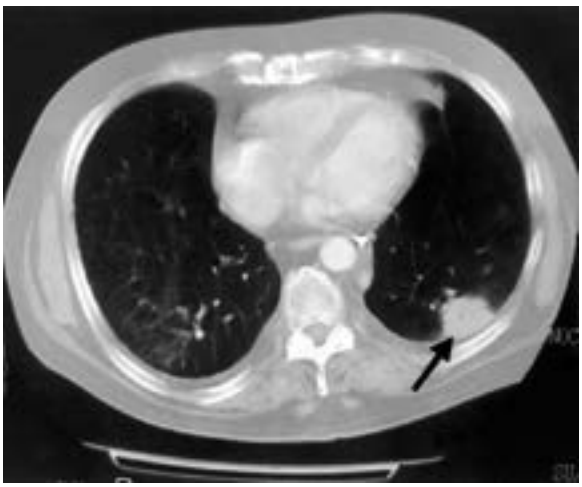

FIGURE 330-1. Chest computed tomography showing a peripheral nodular lung lesion (*arrow*) caused by *Nocardia* infection.

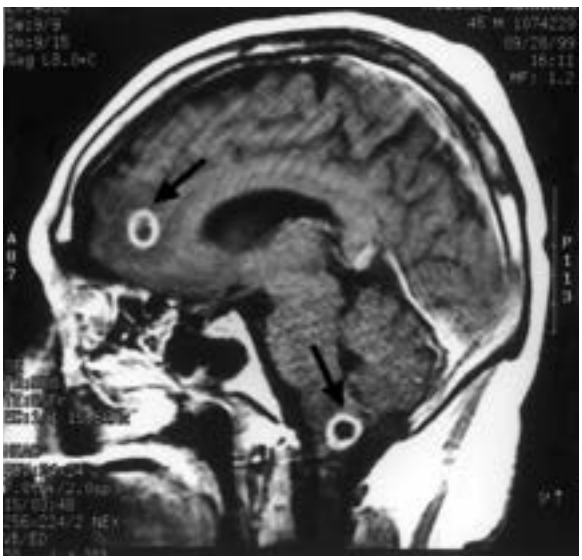

FIGURE 330-2. Multiple, gadolinium-enhanced brain lesions (*arrows*) caused by *Nocardia* infection.

bacterial causes of brain abscess. Positron emission tomography is not helpful in differentiating *Nocardia* brain abscess from tumor; both demonstrate increased uptake.

Histopathology

Invasive procedures are generally required for specific diagnosis. For pulmonary infection, bronchoscopy with transbronchial biopsy or skinny needle biopsy is recommended. CT-guided needle aspirate is the diagnostic procedure of choice for brain abscess. Histopathologic examination usually reveals an acute inflammatory response with a predominance of neutrophils. Micronodular abscesses with minimal capsular formation are usually found. Gram stain or Brown-Brenn stains reveal gram-positive, beaded, branching forms. The morphology is identical to that of *Actinomyces;* however, the high lipid content of its cell wall often renders *Nocardia* modified acid-fast positive, whereas *Actinomyces* spp are modified acid-fast negative. However, acid fastness may be variable when staining *Nocardia* colonies from cultures and is unreliable for direct clinical samples.

Culture

Isolation of the organism on culture provides a definitive diagnosis from needle aspirate samples of brain abscess, and a presumptive diagnosis when grown from respiratory and cutaneous samples. *Nocardia* grows best under aerobic conditions with 5 to 10% carbon dioxide. Because the organism grows slowly on blood agar, taking 3 to 5 days to form colonies, other organisms can overgrow. When *Nocardia* is suspected the clinical laboratory should be notified to allow the use of selective media and prolonged incubation. 16S ribosomal RNA gene sequencing allows rapid speciation of isolates. Antibiotic susceptibility testing always should be performed to guide the choice of therapy.

TREATMENT

Rx

Because of the rarity of nocardiosis, no prospective treatment trials have been performed, and all recommendations are based on retrospective studies and in vitro sensitivity testing. Sulfonamides remain the treatment of choice for pulmonary and cutaneous disease. Trimethoprim-sulfamethoxazole given orally at a dose of 160/800 mg (one double strength) three times daily is the most commonly used adult regimen. When patients with disseminated and/or CNS *Nocardia* are treated with sulfonamides alone, survival has been less than 50%. One of the most common species to cause disseminated diseases, *N. farcinica,* is also one of the most common *Nocardia* spp to be resistant to sulfonamides. In these more serious conditions combination therapy is generally recommended, the exact regimen being guided by antibiotic susceptibility testing. One recommended empirical regimen is trimethoprim-sulfamethoxazole 15 mg/kg/day IV of the trimethoprim component divided into 2 to 4 doses, amikacin (7.5 mg/kg every 12 hours), and either ceftriaxone (2 g twice daily) or imipenem (500 mg four times daily). In a retrospective study, patients failing trimethoprim-sulfamethoxazole responded to imipenem with or without amikacin.[9] Linezolid (600 mg twice daily) has been used successfully in a small number of CNS infections. However, prolonged therapy with this agent can lead to bone marrow toxicity and warrants weekly monitoring of the peripheral cell counts.[10] The newer fluoroquinolone moxifloxacin (400 mg/day PO) has been shown to have activity against several strains of *Nocardia,* including *N. farcinica* and *N. brasiliensis,* and this agent may prove useful in patients who cannot tolerate sulfonamides; however, relapse has been reported. Minocycline (100 mg PO twice daily) and amoxicillin-clavulanate (875/125 mg PO twice daily) are other potentially effective alternative treatments for *Nocardia.* Because of the intracellular nature of *Nocardia* and the slow rates of bacterial growth, antibiotic treatment for 6 to 12 months in the immunocompromised host and 4 to 6 months in the normal host is usually required to prevent relapse.

In addition to antibiotics for patients with brain abscess or subcutaneous abscesses, surgical drainage is required for cure.

PROGNOSIS

The overall mortality for nocardiosis is approximately 25%. In otherwise healthy individuals pulmonary nocardiosis has a better prognosis (15% mortality). The survival rate is worse in patients with bacteremia, patients with acute infection (symptoms for < 3 weeks), patients receiving corticosteroids or cytotoxic agents, patients with disseminated disease involving two or more noncontiguous organs, and patients with meningitis.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Lebeaux D, Morelon E, Suarez F, et al. Nocardiosis in transplant recipients. *Eur J Clin Microbiol Infect Dis*. 2014;33:689-702.
2. Trevino-Villarreal JH, Vera-Cabrera L, Valero-Guillen PL, et al. *Nocardia brasiliensis* cell wall lipids modulate macrophage and dendritic responses that favor development of experimental actinomycetoma in BALB/c mice. *Infect Immun*. 2012;80:3587-3601.
3. Rosas-Taraco AG, Perez-Linan AR, Bocanegra-Ibarias P, et al. *Nocardia brasiliensis* induces an immunosuppressive microenvironment that favors chronic infection in BALB/c mice. *Infect Immun*. 2012;80:2493-2499.
4. Chen J, Zhou H, Xu P, et al. Clinical and radiographic characteristics of pulmonary nocardiosis: clues to earlier diagnosis. *PLoS ONE*. 2014;9:e90724.
5. Kurahara Y, Tachibana K, Tsuyuguchi K, et al. Pulmonary nocardiosis: a clinical analysis of 59 cases. *Respir Investig*. 2014;52:160-166.
6. Anagnostou T, Arvanitis M, Kourkoumpetis TK, et al. Nocardiosis of the central nervous system: experience from a general hospital and review of 84 cases from the literature. *Medicine*. 2014;93:19-32.
7. Nandhagopal R, Al-Muharrmi Z, Balkhair A. Nocardia brain abscess. *QJM*. 2014;107:1041-1042.
8. Dodiuk-Gad R, Cohen E, Ziv M, et al. Cutaneous nocardiosis: report of two cases and review of the literature. *Int J Dermatol*. 2010;49:1380-1385.
9. Ameen M, Arenas R, Vasquez del Mercado E, et al. Efficacy of imipenem therapy for *Nocardia actinomycetomas* refractory to sulfonamides. *J Am Acad Dermatol*. 2010;62:239-246.
10. Ambrosioni J, Lew D, Garbino J. Nocardiosis: updated clinical review and experience at a tertiary center. *Infection*. 2010;38:89-97.

REVIEW QUESTIONS

1. How can *Nocardia* be contracted by humans?

A. Spread from person to person
B. Spread from animals to humans (zoonotic infection)
C. By inhaling dust
D. By being bitten by mosquitoes
E. By being bitten by ticks

Answer: C *Nocardia* usually gains entry through the lungs. Patients inhale dust contaminated with *Nocardia,* a common soil organism.

2. A 50-year-old male bicycler fell on a dirt path, scraping his arm. Over the next 2 weeks he noted progressive swelling of his arm associated with swollen lymph nodes and streaks of erythema. He also developed a fever to 101° F. His peripheral WBC was 12,500 (90% PMN). How should you manage this patient (*best* answer)?

A. Begin an antistaphylococcal antibiotic
B. Swab the surface of his erythematous skin for culture
C. Draw blood cultures
D. Obtain a tissue biopsy for histopathologic examination and culture
E. Obtain a urine antigen test for histoplasmosis

Answer: D Diagnosis of *Nocardia* is difficult and often delayed. Obtaining a biopsy sample for histopathologic examination and staining, as well as culture, is the primary method for making the appropriate diagnosis. Empirical antibiotics often fail to treat *Nocardia*. A surface swab culture is likely to grow only normal skin flora because *Nocardia* is slow growing. *Histoplasma* rarely causes soft tissue infection, and therefore urine antigen is highly likely to be negative.

3. A 55-year-old renal transplant recipient, who has been receiving high-dose corticosteroids for transplant rejection, develops severe headaches that persist for 2 weeks. His headaches are sharp and localized to the left temporal region. They are not relieved by Tylenol or aspirin. He is admitted to the hospital after having a grand mal seizure that began with tonic-clonic movements of the right arm and leg. CT with contrast shows a multiloculated ring-enhancing abscess in the left temporal region. A CT-guided needle biopsy is performed. Gram stain of the sample demonstrates gram-positive beaded branching forms accompanied by many PMNs. What is the simplest and most rapid way to differentiate whether your patient has nocardiosis or actinomycosis?

A. Modified acid-fast stain *Nocardia* is acid-fast positive
B. Modified acid-fast stain *Actinomyces* spp are acid-fast positive
C. Polymerase chain reaction of the sample
D. Aerobic and anaerobic culture
E. PET scan

Answer: A The *Nocardia* cell wall has a high lipid content that tightly binds carbol fuchsin (the red stain used for acid-fast staining) and is resistant to acid decolorization while *Actinomyces* spp is readily decolorized. PCR is most useful for speciation. It is more complicated to perform and is not readily available in most hospital-based diagnostic laboratories. *Nocardia* is slow growing and may take over a week to grow. PET scan can falsely diagnose *Nocardia* brain abscess as a brain tumor and is not a helpful diagnostic tool.

4. What would be the best treatment plan for a patient with a *Nocardia* brain abscess?

A. Ceftriaxone and metronidazole combined with surgical drainage
B. High-dose penicillin alone
C. Trimethoprim-sulfamethoxazole combined with imipenem
D. Clindamycin combined with surgical drainage
E. Trimethoprim-sulfamethoxazole combined with imipenem and surgical drainage

Answer: E Trimethoprim-sulfamethoxazole continues to cover many strains of *Nocardia;* however; combination therapy is favored for brain abscess. Outcome is better when the abscess is also surgically drained. Penicillins have limited activity against *Nocardia,* as does ceftriaxone. Metronidazole has no activity and clindamycin fails to cross the blood-brain barrier and is not recommended for CNS infections.

5. How long should a patient with a *Nocardia* soft tissue infection be treated?

A. 7 to 10 days
B. 2 weeks
C. 1 month
D. 2 months
E. 6 months

Answer: E *Nocardia* is an intracellular pathogen that grows slowly. Relapses have been observed if treatment is not extended for 6 months. In patients who are severely immunocompromised, 12 months of treatment may be required.

331

SYSTEMIC ANTIFUNGAL AGENTS

DAVID A. STEVENS

Methods for in vitro susceptibility testing are available as standardized tools. A variety of assays are available for therapeutic drug monitoring of serum and other body fluids. As more common mutations inducing resistance are uncovered,[1] it becomes possible to develop molecular screening methods that can detect the resistance genes in clinical isolates, in advance of susceptibility test results.[2]

AMPHOTERICIN B–BASED PREPARATIONS

Mechanism of Action

Amphotericin B is a lipophilic molecule that exerts its antifungal effect by insertion into the fungal cytoplasmic membrane. Amphotericin B causes membrane permeability to increase. Loss of intracellular molecules impairs fungal viability. The onset of action is rapid. Amphotericin B also has effects on oxidation that may enhance antifungal activity and in vivo may activate endothelial cells.

Spectrum of Activity and Mechanisms of Resistance

Amphotericin B is active against most fungi, and its spectrum of activity is not influenced by the choice of formulation. When resistance occurs, it is generally attributed to reductions in ergosterol biosynthesis and the synthesis of alternative sterols that lessen the ability of amphotericin B to interact with the fungal membrane; oxidant scavengers also may be produced. Primary resistance is common for *Scedosporium* and *Trichosporon* spp. Among the *Candida* spp, primary resistance is noted at meaningful frequencies most often for *Candida lusitaniae*. Development of resistance in isolates of normally susceptible species is uncommon. In some studies, the principal pharmacodynamic driver of in vivo response has been the ratio of the peak serum concentration to the minimal inhibitory concentration (MIC).

Available Formulations

There are four commercially available amphotericin B formulations: amphotericin B deoxycholate (ABD) and three lipid-associated formulations—amphotericin B colloidal dispersion (ABCD), amphotericin B lipid complex, and liposomal amphotericin B. All formulations must be infused in 5% dextrose with no electrolytes added. Infusion bottles need not be protected from light. In attempts to produce less expensive lipid-associated formulations, some have advocated mixing ABD with a parenteral fat emulsion. Although less nephrotoxicity has been observed in adults given this preparation at a dose of 1 mg/kg/day compared with infusions of ABD in 5% dextrose, no advantage was found in children. Serum amphotericin B concentrations were also lower with the fat emulsion, raising the possibility that amphotericin B was simply aggregating in the fat emulsion, but the cloudiness could not be perceived in the milky-looking lipid. Use of such preparations should be reserved for investigational settings.

Amphotericin B Deoxycholate

Formulation

ABD for intravenous use is a colloidal suspension. If a filter with a 0.22-μm pore diameter is placed in the infusion line, considerable drug is removed by the filter. The addition of electrolyte aggregates the colloids, so the solution becomes cloudy; this is to be avoided. ABD is available from several generic manufacturers, and significant differences in the formulations have been reported, which may account in part for the intersubject variation in toxicities observed.

Pharmacology

Most of the drug leaves the circulation promptly, with only a small percentage being excreted in urine or bile. Amphotericin B is stored in the liver and other organs; the drug appears to re-enter the circulation slowly. Blood levels are not influenced by hepatic or renal failure. Hemodialysis does not alter blood levels, except in an occasional patient with lipemic plasma who may be losing drug owing to its adherence to the dialysis membrane. Concentrations of amphotericin B in fluid from inflamed areas, such as pleura, peritoneum, joint, vitreous humor, and aqueous humor, are roughly two thirds of the nadir serum level. Amphotericin B penetrates poorly into either normal or inflamed meninges, saliva, bronchial secretions, brain, pancreas, muscle, bone, vitreous humor, and normal amniotic fluid. Urine concentrations are similar to serum concentrations. Peak serum concentrations with conventional intravenous doses are roughly 0.5 to 2 μg/mL; these concentrations fall rapidly and then slowly approach a plateau of roughly 0.2 to 0.5 μg/mL. The initial half-life is approximately 24 hours; the β-phase half-life is roughly 15 days. Serum concentrations can be detected for at least 7 weeks after the end of therapy, presumably reflecting release from cell membranes. The drug also has complex immunomodulatory properties that are potentially of clinical significance.

Nephrotoxicity

ABD causes a dose-dependent decrease in the glomerular filtration rate. The direct vasoconstrictive effect of amphotericin B on afferent renal arterioles results in reduced glomerular and renal tubular blood flow. Other effects on the kidney include potassium, magnesium, and bicarbonate wasting and decreased erythropoietin production. Loss of renal function is due to the destruction of renal tubular cells, disruption of tubular basement membrane, and loss of functioning nephron units. Saline loading, such as the infusion of 1 L saline before ABD, has been associated with reduced nephrotoxicity in some studies. Potassium wasting often requires supplemental oral or intravenous potassium. Renal tubular acidosis from bicarbonate wasting rarely requires base replacement, but other drugs and diseases that promote acidosis may act synergistically.

Azotemia caused by amphotericin B is often worse in patients taking other nephrotoxic drugs. Hypotension, intravascular volume depletion, and other preexisting renal disease all magnify the management problems associated with amphotericin B–induced azotemia. These toxicities are lessened by use of the lipid-associated formulations of amphotericin B.

Early in the course of therapy with ABD, azotemia may increase rapidly; it often improves a little and then stabilizes after several days. Adults with no other renal disease have an average serum creatinine level of 2 to 3 mg/dL at therapeutic doses, and therapy should not be withheld unless azotemia exceeds this level. Attempting to give ABD to an adult without causing azotemia usually leads to inadequate therapy.

Other Chronic Toxicity

Nausea, anorexia, and vomiting are common. Phlebitis occurs if peripheral vein catheters are used. Normocytic normochromic anemia occurs gradually. The hematocrit rarely falls below 20 to 25% unless other causes of anemia are present. Rarely, thrombocytopenia, modest leukopenia, arrhythmias, coagulopathy, hemorrhagic enteritis, tinnitus, vertigo, encephalopathy, seizures, hemolysis, or dysesthesia of the soles of the feet may be observed. Amphotericin B remains the first choice parenteral agent in pregnancy despite its potential toxicities.[3]

Acute Reactions

Approximately 30 to 45 minutes after beginning the first few ABD infusions, chills, fever, and tachypnea may occur, peak in 15 to 30 minutes, and then slowly abate over 2 to 4 hours. A patient with underlying cardiac or pulmonary disease may have hypoxemia. These reactions are less common in young children or in patients receiving corticosteroids. Subsequent infusions of the same dose cause progressively milder reactions. Premedication with acetaminophen or the addition of hydrocortisone 25 to 50 mg to the infusion solution can diminish the reaction. Meperidine given early in a chill shortens the rigors but may induce nausea or emesis. Concern about this kind of reaction in an unstable patient has led some physicians to use a test dose of 1 mg given over a 15-minute period to assess the subsequent reaction over 1 hour before deciding whether to administer a full therapeutic dose. Patients with rapidly progressive mycoses should receive a full therapeutic dose within 24 hours. These reactions should not be mistaken for anaphylaxis or considered a contraindication to further amphotericin B. True allergic reactions are extremely rare.

Administration

ABD is infused over a 2- to 4-hour interval. Infusions 1 hour in duration generally appear to be safe for persons who have tolerated slower infusions. Rapid infusion in patients with severely compromised renal function may lead to acute, marked hyperkalemia and ventricular fibrillation.

Patients receiving a stable daily dose can be changed to a double dose on alternate days to reduce the frequency of infusion-associated toxicity, particularly anorexia, and to make outpatient therapy more convenient. Doses above 1.5 mg/kg are generally not given on this schedule because the toxicity of such infusions is not well described. Continuous infusion of amphotericin B with doses up to 2 mg/kg/24 hours is another approach (based on limited data) to reducing toxicity, but this is not consistent with the observation that the principal pharmacodynamic driver of efficacy for amphotericin B is peak drug concentration.[4]

Dosage

Daily ABD doses of 0.3 mg/kg often suffice for esophageal candidiasis. A dose of 0.5 mg/kg is appropriate for blastomycosis, disseminated histoplasmosis, and extracutaneous sporotrichosis. Patients with cryptococcal meningitis are generally given doses of 0.6 to 0.8 mg/kg; those with coccidioidomycosis may require doses of 1 mg/kg. Patients with mucormycosis or invasive aspergillosis are given daily doses of 1 to 1.5 mg/kg until improvement is clearly present. Doses of 0.5 to 1 mg/kg are often used in neutropenic patients receiving empirical amphotericin B. Local instillation of amphotericin B into cerebrospinal fluid (CSF), joints, or pleura is rarely indicated. One exception is coccidioidal meningitis, which may be treated with intrathecal ABD because it may produce superior results to systemic azole therapy, particularly in the long term, although with far greater toxicity. Intraocular administration for fungal endophthalmitis is occasionally used; doses of 10 μg appear to avoid retinal toxicity. Corneal baths with 1 mg/mL in sterile water are useful for fungal keratitis but are irritating.

Lipid-Associated Formulations of Amphotericin B

Pharmacology and Toxicity

The three lipid-associated formulations have quite different pharmacokinetic patterns. When compared on the basis of equal dosages (milligram/kilogram), the lipid-associated formulations produce tissue amphotericin B concentrations that range from 90% lower to 500% higher than those seen with ABD, with the most consistent relative reduction seen in the kidney. The lipid-associated formulations are typically given at doses that are 3- to 12-fold higher than those used for ABD. All three formulations generally require higher doses in experimental animals to achieve the same therapeutic effect as ABD.

These higher but equipotent doses are notably better tolerated than ABD, with reductions in both the frequency and the severity of acute infusion-related reactions and chronic nephrotoxicity. An exception is ABCD, which generally induces acute infusion-related reactions similar to those for ABD.

Randomized clinical trials comparing ABD as therapy for a defined mycosis are limited to demonstrations that liposomal amphotericin B has a similar efficacy for cryptococcal meningitis and a greater efficacy for histoplasmosis. Randomized comparisons with ABD in persistently neutropenic and febrile cancer patients consistently demonstrate a better tolerability profile, but few data on differential antifungal effect. The aggregate open-label data on efficacy rates for the lipid-associated formulations are similar to those for ABD. Although the lipid-associated formulations are notably more costly than ABD, the purchase cost of the compound must be balanced against the morbidity of ABD-related nephrotoxicity and the financial costs of monitoring, treating, and managing it.

FLUCYTOSINE

Formulation and Pharmacology

Flucytosine (5-fluorocytosine) is the fluorine analogue of a normal body constituent, cytosine. Flucytosine is marketed as 250- and 500-mg capsules. Absorption from the gastrointestinal tract is rapid and complete, and approximately 90% is excreted unchanged in the urine. CSF concentrations are approximately 74% of simultaneous serum concentrations; the drug also penetrates well into aqueous humor, joints, bronchial secretions, peritoneal fluid, brain, bile, and bone, and it is readily cleared by hemodialysis and peritoneal dialysis.

The half-life of the drug in the serum of patients with normal renal function is 3 to 5 hours and is longer in newborns. Abnormal hepatic function has no influence, but decreased renal function prolongs the half-life.

Mechanisms of Action and Resistance

Isolates of *Candida* spp are usually susceptible, as are most isolates of *Cryptococcus neoformans*. Flucytosine is often active against isolates of *Aspergillus* spp and against the melanin-pigmented molds that cause chromoblastomycosis. The mechanism of antifungal action appears to be deamination to 5-fluorouracil and then conversion through several steps to 5-fluorodeoxyuridylic acid monophosphate, a noncompetitive inhibitor of thymidylate synthetase, which interferes with DNA synthesis, or conversion to 5-fluouridine triphosphate, which is incorporated into RNA and causes aberrant transcription. In some studies, the principal pharmacodynamic driver of response was the proportion of time the blood level exceeded the minimum inhibitory concentration (MIC). Resistance may be due to loss of the cytosine permease that permits flucytosine to cross the fungal cell membrane or loss of any of the enzymes that lead to its conversion into forms that interfere with DNA or RNA synthesis.

Administration and Dosage

Flucytosine is administered orally, 100 to 150 mg/kg/day, in four divided doses. As an approximation, the total daily dose should be reduced to 75 mg/kg with a creatinine clearance of 26 to 50 mL/minute and to 37 mg/kg when the creatinine clearance is 13 to 25 mL/minute. Ideally, the blood level should be measured in azotemic patients 2 hours after the last dose and immediately before the next dose. These values should range between 10 and 100 μg/mL. Patients requiring hemodialysis can be given a single post-dialysis dose of 37.5 mg/kg. Further doses are adjusted by blood level.

Flucytosine given alone to patients with normal renal, hematologic, and gastrointestinal function is infrequently associated with adverse effects, including rash, diarrhea, and hepatic dysfunction. In the presence of azotemia (such as that caused by concomitant amphotericin B), leukopenia, thrombocytopenia, and enterocolitis may appear. These complications seem to be far more frequent among patients whose flucytosine blood levels attain, and especially if they exceed, 100 to 125 μg/mL. Patients receiving flucytosine whose renal function is changing should have their serum flucytosine concentrations determined twice weekly, and the leukocyte count, platelet count, alkaline phosphatase, and aminotransferase levels should be measured at a similar frequency. Patients in whom loose stools or dull abdominal pain suddenly develops or who have laboratory evidence consistent with flucytosine toxicity should have their flucytosine blood levels determined, and consideration should be given to withholding the drug until the situation is clarified. Patients with bone marrow and gastrointestinal toxicity from flucytosine often tolerate the drug at a reduced dosage. Uncommonly, vomiting, bowel perforation, confusion, hallucinations, headache, sedation, and euphoria have been reported. Flucytosine is contraindicated in pregnancy.

Conversion of flucytosine to 5-fluorouracil within the human body occurs to a sufficient degree to possibly account for the drug's toxicity to bone marrow and the gastrointestinal tract. It is likely that the drug is secreted into the gut, where flucytosine becomes deaminated by intestinal bacteria and is reabsorbed as 5-fluorouracil.

Flucytosine has a beneficial effect in patients with cryptococcosis, candidiasis, and chromoblastomycosis. It is not the drug of choice for any infection, however, because its clinical efficacy in the first two mycoses is inferior to that of amphotericin B, primary drug resistance is not uncommon in *Candida* infection, and secondary drug resistance is common in cryptococcosis and chromoblastomycosis.

Flucytosine and amphotericin B are at least additive in their effects. Flucytosine permits a lower dose of amphotericin B to be used to achieve the same therapeutic effect, and amphotericin B prevents the emergence of secondary drug resistance. These advantages have been confirmed in large multicenter studies of cryptococcal meningitis, and a more rapid antifungal effect has been shown.

Flucytosine therapy is more difficult to manage in patients with diminished bone marrow reserve. Intravenous flucytosine is no longer available in the United States, but it is used at the same dosage as the capsule formulation.

Flucytosine resistance has occurred, albeit uncommonly, during combination therapy. Use of combination therapy in such patients incurs the risk for toxicity without evidence that flucytosine adds to the therapeutic effect. An MIC of 20 μg/mL or less is considered susceptible.

AZOLE ANTIFUNGAL AGENTS

Mechanism of Action

The azole ring confers antifungal activity on a variety of synthetic organic compounds. *N*-substitution of imidazoles has created a family of drugs called *triazoles* that have the same mechanism of action as imidazoles, a similar or broader spectrum of activity, and less effect on human sterol synthesis. Both imidazoles and triazoles inhibit C-14α demethylation of lanosterol in fungi; this leads to reduced concentrations of ergosterol, which is essential for a normal fungal cytoplasmic membrane. In some studies, the principal pharmacodynamic driver for response to the triazole antifungal agents has been the ratio of total drug exposure (area under the curve) to the MIC. Because of their interaction with the cytochrome P-450 (CYP) system and, in some cases, the P-glycoprotein pumps, the azoles as a class have a large number of drug-drug interactions, only some of which are covered here. In any patient in whom azole therapy is contemplated, package inserts for the drug should be consulted to determine which of the patient's other medications could result in a significant drug-drug interaction.

Newer triazoles have properties that make them preferable to ketoconazole. These include not only less hormonal inhibition but also better distribution into body fluids, both parenteral and oral formulations, less hepatotoxicity, and a broader spectrum. Some have fewer drug interactions. Resistance to azoles in previously susceptible species is emerging. Resistance mechanisms include increased drug efflux and altered or increased C-14α demethylase. It has been speculated that some of the common mutations in this demethylase in *Aspergillus* isolates from certain geographic areas, resulting in resistance, may have arisen owing to the widespread agricultural use of azole fungicides. All agents in this class have the potential for embryotoxicity and teratogenicity and should be avoided during pregnancy, particularly the first trimester.

Ketoconazole

Formulations and Pharmacology

Ketoconazole is available in tablet form. It is metabolized in the liver and excreted as inactive drug in bile and, to a small extent, urine. The drug is not removed significantly by hemodialysis or peritoneal dialysis. Decreased renal or hepatic function does not alter plasma drug levels. The initial half-life is approximately 2 hours, with a β-phase half-life of approximately 9 hours commencing 8 to 12 hours after ingestion.

Oral absorption of ketoconazole differs among individuals. Serious gastrointestinal disease may lead to low blood levels. Inhibitors of gastric acid secretion should not be given to patients taking ketoconazole because blood levels of the latter drug are drastically reduced. The drinking of citrus fruit juices or cola beverages with ketoconazole improves absorption in hypochlorhydric patients. Patients who need antacids should take them 1 to 2 hours after ketoconazole. Low concentrations are found in vaginal secretions, saliva, and breast milk, but penetration into inflamed joints is better.

Uses

The main advantage of ketoconazole is its lower cost compared with the cost of triazoles. Although at a dose of 400 mg/day the drug is effective in many mycoses, usage has been supplanted by the triazoles, but least so in cutaneous mycoses. Therapy is continued for 6 to 12 months or longer, if the response is slow, to prevent relapse. Improvement may not be evident for weeks to months. Although the dose can be advanced to 600 or 800 mg/day in patients who are not responding to therapy, there is more evidence of increased toxicity than increased efficacy. Aspergillosis does not respond to ketoconazole; additionally, the subsequent use of amphotericin B may be antagonized.

Adverse Effects

The most frequent toxic effects are anorexia, nausea, and vomiting. Dividing doses greater than 400 mg/day is not recommended because hormonal suppression is prolonged. Ketoconazole causes a dose-dependent depression in the serum testosterone-stimulated and adrenocorticotropic hormone–stimulated cortisol response. The effect of doses of 800 to 1200 mg/day is profound enough to have prompted trials in the treatment of Cushing's syndrome and prostate cancer. Hypertension is seen in a few of these high-dose patients, associated with increased levels of mineralocorticoid precursors.

Gynecomastia, oligospermia, and menstrual irregularities may be seen during prolonged therapy. Allergic rash and pruritus have been noted.

Perhaps the most serious complication is hepatitis. Fortunately, this complication is estimated to occur in only 1 in 15,000 exposed individuals. A slight, asymptomatic elevation of transaminase levels occurs in 5 to 15% and is generally transient. Ketoconazole-associated hepatitis begins as anorexia, malaise, nausea, and vomiting. Eighty percent of cases occur within the first 3 months. Progression can be swift. If hepatotoxicity is suspected, serum transaminase and alkaline phosphatase levels should be measured. Some authorities have recommended that liver function be measured periodically. This procedure does not protect a patient who has a rapid onset of hepatitis in the interval between tests, but it does require that all patients with abnormalities be contacted to inquire about symptoms and to arrange for repeat testing.

Itraconazole

Formulations and Pharmacology

Itraconazole (Sporanox) is marketed as a capsule and as an oral suspension in cyclodextrin (an oligosaccharide ring). The ring entraps the hydrophobic, water-insoluble drug, thus making it soluble; it is then released either at the lipid membrane of the enterocyte after oral administration or directly into tissues after intravenous administration. The solution can be delivered through a nasogastric tube in intubated patients, and it makes the dosing of infants and small children more convenient. Oral absorption of the capsule is significantly enhanced by food, whereas absorption of the solution is best on an empty stomach. Coadministration of a cola beverage with itraconazole capsules increases absorption. Peak levels with either preparation are achieved 4 to 6 hours after a dose. Steady state is achieved only after 13 to 15 days, at which time the β-elimination half-life is approximately 19 to 22 hours. Absorption of the capsule is markedly depressed in bone marrow transplant recipients, probably because of hypochlorhydria, mucositis, and graft-versus-host intestinal changes, and in patients with acquired immunodeficiency syndrome (AIDS) because of enteritis, but this problem can be alleviated by using the solution.

For deep mycoses, an initial itraconazole dose of 200 mg three times daily is recommended for the first 3 days to quickly achieve high serum and tissue levels. Hydroxyitraconazole, a metabolite of itraconazole, appears in the blood in amounts roughly twice that of the parent drug; it has antifungal activity and pharmacokinetics similar to those of the parent compound.

Bioassays of itraconazole yield much higher concentrations than measurement by high-pressure liquid chromatography; the difference depends on the susceptibility of the bioassay organism to hydroxyitraconazole. Concentrations of itraconazole in tissue, pus, and bronchial secretions are generally higher than those in plasma. Ocular levels are low. Saliva concentrations persist for 8 hours after administration of the solution and may be beneficial in treating oral disease or eradicating oral colonization. The drug is metabolized in the liver and excreted in feces as metabolites. Of the liquid administered, less than 0.5% of the cyclodextrin is absorbed. No significant amount of bioactive itraconazole appears in urine. Plasma concentrations do not increase in patients with renal insufficiency or decrease with hemodialysis. The half-life is prolonged in those with cirrhosis.

Adverse Effects

The most common adverse effect is dose-related nausea and abdominal discomfort, but symptoms rarely necessitate stopping therapy. Dividing the dose and administering the drug twice daily can improve tolerance and raise blood levels. Hypokalemia and edema may occur at doses of 400 mg/day or higher. Allergic rash is seen occasionally. Itraconazole is rarely hepatotoxic. Diarrhea, nausea, and other gastrointestinal complaints are more frequent with the solution. A negative inotropic effect is rarely seen.

Drug Interactions

Blood levels are reduced by about half in patients taking drugs that decrease gastric acidity. Some of the most notable drug interactions are rifampin, rifabutin, isoniazid, phenytoin, carbamazepine, and phenobarbital, which decrease itraconazole blood levels. Itraconazole decreases rifampin blood levels and increases blood levels of some antihistamines, potentially causing polymorphic ventricular tachycardia (torsades de pointes), as well as increasing levels of warfarin, benzodiazepines, hepatic hydroxymethylglutaryl–coenzyme A (HMG-CoA) reductase cholesterol-lowering agents, dihydropyridine calcium-channel blockers, digoxin, quinidine, cyclosporine, tacrolimus, methylprednisolone, human immunodeficiency virus (HIV) protease inhibitors (ritonavir, indinavir), and vinca alkaloids (vincristine, vinblastine).

Uses

Itraconazole is useful for the treatment of invasive aspergillosis, allergic bronchopulmonary aspergillosis, blastomycosis, histoplasmosis, meningeal and nonmeningeal coccidioidomycosis, paracoccidioidomycosis, sporotrichosis, phaeohyphomycosis, mucosal candidiasis, ringworm (including onychomycosis), and tinea versicolor. Itraconazole is also useful for the prevention of relapse of disseminated histoplasmosis in patients with AIDS. Itraconazole may be useful for prophylaxis against fungal infections during neutropenia and as empirical therapy for febrile episodes in neutropenia.

Fluconazole

Formulations and Pharmacology

Fluconazole is currently available in oral and vaginal tablets, as a powder for oral suspension, and as an intravenous formulation. Fluconazole is well absorbed from the gastrointestinal tract. Of the oral dose, 60 to 75% appears unchanged in the urine. Oral absorption is not decreased in patients with AIDS or patients taking H_2-blocking agents. Fluconazole penetrates into the brain, saliva, sputum, and urine.

The half-life in patients with normal renal function is 27 to 34 hours; it increases to 59 and 98 hours in groups with creatinine clearances of 35 and 14 mL/minute, respectively. The normal dose should be reduced to 50% when the creatinine clearance is reduced to 50 mL/minute and to 25% when the creatinine clearance is below 20 mL/minute. A loading dose of twice the daily dose is recommended. Patients receiving hemodialysis should have one daily dose after each session.

Drug Interactions

Among other drug interactions, fluconazole can cause significant increases in the blood level of phenytoin, glipizide, glyburide, tolbutamide, warfarin, rifabutin, cisapride, quinidine, or cyclosporine. Rifampin lowers fluconazole blood levels by approximately one fourth.

Adverse Effects

Adverse effects are uncommon. Even with chronic therapy, including doses exceeding 400 mg/day, headache, hair loss, and anorexia were the most common symptoms; 10% of patients had rises in aspartate aminotransferase levels. Alopecia is reversible, in some cases, even when the drug is continued at lower doses. Neurotoxicity has been described after heroic doses of 2000 mg/day. Rarely, anaphylaxis after the first dose or Stevens-Johnson syndrome has been observed.

Indications

Candidiasis

Provided the infection is not caused by fluconazole-resistant *Candida,* fluconazole is effective for the treatment of oropharyngeal and esophageal candidiasis. A single dose of 150 mg is approximately as effective as topical treatment of vulvovaginal candidiasis. Patients with candidemia who are not neutropenic or otherwise seriously immunosuppressed respond as well to intravenous fluconazole therapy as to amphotericin B. A study comparing fluconazole with a combination of fluconazole plus amphotericin as initial therapy of candidemia found that the combination might produce more rapid clearance of the pathogen from the blood stream. In a few patients with *Candida* endocarditis, long-term fluconazole therapy has been used to prevent relapse after amphotericin B therapy. For immunosuppressed patients and for rapidly progressing or severely ill patients with deep-seated candidiasis, amphotericin B or an echinocandin is preferred.

Cryptococcal Meningitis

Clinical trials in AIDS have engendered the conventional practice of therapy with amphotericin B or amphotericin B plus flucytosine for at least the first 2 weeks. Therapy can then be changed to fluconazole 400 mg/day for 2 months if the patient is clinically stable. The propensity of patients with AIDS to relapse has led to maintenance therapy with fluconazole 200 mg/day, at least until antiretroviral therapy causes CD4 counts to approach normal. Itraconazole capsules are inferior to fluconazole for maintenance therapy. Relapse because of fluconazole resistance is rare. Fluconazole is effective for the eradication of genitourinary foci. For patients without AIDS, fluconazole is useful for those who have completed a course of amphotericin B and seem to be at high risk for relapse.

Other Mycoses

Fluconazole is useful for coccidioidal meningitis and disseminated nonmeningeal coccidioidomycosis, but a direct comparison with itraconazole found a trend favoring itraconazole owing to its superior efficacy for skeletal infections. The two drugs were similarly efficacious for soft tissue and pulmonary infections. Cutaneous sporotrichosis, ringworm, histoplasmosis, and blastomycosis may respond to fluconazole, but the results are inferior to those with itraconazole.

Prophylaxis in Neutropenic Patients

Fluconazole decreases the incidence of death from deep-seated candidiasis in bone marrow transplant recipients. In a long-term follow-up of one study, bone marrow transplant recipients who received at least 75 days of fluconazole prophylaxis also had improved survival. Use of this drug for prophylaxis may result in shifts to less susceptible species in patient flora.

Prophylaxis in Patients with Acquired Immunodeficiency Syndrome

Fluconazole has reduced the incidence of oral and vulvovaginal candidiasis and cryptococcosis in patients with advanced HIV infection, but it has negligible effects on other mycoses. Cost, lack of effect on survival, and the possibility of azole resistance have led the advisory committee of the Infectious Diseases Society of America to recommend against routine fluconazole prophylaxis in patients with AIDS.

Voriconazole

Formulations and Pharmacology

Voriconazole is marketed as a tablet and as a solution in sulfobutyl ether β-cyclodextrin for intravenous administration. Voriconazole is cleared by hepatic metabolism, with less than 2% of the dose excreted unchanged in the urine. Voriconazole exhibits nonlinear pharmacokinetics owing to saturation of the clearance pathways at higher doses. The principal enzyme involved in clearance is hepatic CYP2C19. This enzyme has significant genetic polymorphisms that affect the metabolism of this drug. Despite these differences in metabolism, the achieved plasma levels overlap, and an initial dose adjustment based on genotype or racial group is not necessary. Different studies have suggested that favorable outcomes correlate with serum concentrations greater than 1 to 2 μg/mL, and neurologic/psychiatric, hepatic or cardiac side effects occur with levels greater than 5 μg/mL, so therapeutic drug monitoring is recommended, with dose adjustments as necessary.

Standard loading doses followed by maintenance doses that are 50% of normal are recommended for individuals with mild-to-moderate hepatic cirrhosis. Dosage adjustments are not required for renal dysfunction, and voriconazole is not significantly cleared by hemodialysis. Because of nephrotoxicity from the cyclodextrin, the intravenous formulation should not be used in patients with a creatinine clearance less than 50 mL/minute.

Drug Interactions

Voriconazole has many drug interactions. Rifampin, carbamazepine, long-acting barbiturates, glucocorticoids, and ritonavir induce the hepatic enzymes responsible for the clearance of voriconazole and thereby reduce voriconazole levels. Sirolimus levels are increased dramatically. Reduction in the clearance of pimozide, quinidine, and some antihistamines may be sufficient to place the patient at risk for QTc prolongation. Reduction in the clearance of ergot alkaloids could lead to chronic ergot poisoning. Cyclosporine, tacrolimus, warfarin, oral coumarins, lipid-lowering statin agents, benzodiazepines, calcium-channel blockers, sulfonylureas, methadone, and vinca alkaloids can be coadministered, but the dosages of these drugs may need to be reduced, and clinical or laboratory monitoring is suggested. Voriconazole significantly reduces the rate of clearance of omeprazole. Voriconazole with rifabutin, phenytoin, or efavirenz results in lower voriconazole levels and elevated levels of the other drug. Other interactions are possible, and any cytochrome P450 (CYP 450) inhibitors, blockers, or inducers could have an interaction with voriconazole.

Adverse Effects

The most frequently reported adverse event is a transient, reversible visual disturbance beginning approximately 30 minutes after a dose. Patients should be advised to avoid activities that require keen visual acuity while experiencing visual changes. Hallucinations and confusion also have been reported, associated with higher blood levels. Liver enzyme abnormalities appear to correlate with higher blood levels. Photosensitivity can be severe and has been associated with skin cancers in rare instances. Prolongation of the QT interval and tachyarrhythmias have been noted in patients with proarrhythmic risk factors, such as electrolyte abnormalities. Periostitis has been noted after prolonged therapy, apparently related to the fluoride atoms in the molecule. Leukoencephalopathy has been rarely reported.

Indications

Aspergillosis

Voriconazole was licensed for the treatment of invasive aspergillosis[5] on the basis of a randomized, unblinded comparative trial in which patients were randomized to receive initial therapy with either voriconazole or amphotericin B. Following the initial randomization, patients could be switched to other licensed therapies, as dictated by clinical events. After 12 weeks, 53% of the patients randomized to voriconazole, but only 32% of those randomized to amphotericin B, had a successful outcome.

Other Mycoses

Voriconazole is also licensed for the treatment of invasive fusariosis and scedosporiosis, based on high response rates for these diseases. Infections by *Scedosporium apiospermum* complex may respond, but *Scedosporium prolificans* infections are commonly resistant. The drug is efficacious in esophageal candidiasis, invasive candidiasis, and refractory candidiasis. Although the drug has appeal for use in prophylaxis, there has been concern about breakthrough zygomycoses with long-term use.

Posaconazole

Formulations and Pharmacology

Posaconazole has been available as an oral suspension, and a tablet form has been introduced. Absorption of the suspension is usually maximized by dividing the daily dose into four administrations and enhanced by taking the suspension with or after fatty food. Some commercial dietary supplements have the same effect as fatty meals and are better tolerated. The half- life is 20 to 35 hours. Clearance is primarily by fecal excretion, with only 13% excreted via renal clearance. Hepatic metabolism via uridine diphosphate glucuronidation plays only a small role in clearance, and CYP-mediated oxidation does not occur. The drug is concentrated in phagocytes. The dose is 800 mg/day for treatment of deep-seated mycoses, although 600 mg has been used successfully for prophylaxis, and 100 to 200 mg has been used for mucosal candidiasis. At steady state, 800 mg/day produces a peak concentration (C_{max}) of 4.2 μg/mL. Blood levels appear to be rather unpredictable, and outcomes in the treatment of aspergillosis correlate with levels. Therapeutic drug monitoring is advisable, and levels greater than 0.7 μg/mL have correlated with efficacy in prophylaxis. Dosage adjustment is not needed for hepatic or renal failure or hemodialysis. The dosing for the 100-mg oral tablet is 300 mg twice daily as a first day loading dose, then 300 mg/day. Blood concentrations after administration of the tablets exceed those after optimal dosing of the solution and appear negligibly affected by food or by medications that affect gastric pH or motility. Side effects of the tablets reported include somnolence, diarrhea, and flatulence. An intravenous formulation of posaconazole was introduced in early 2014.

Drug Interactions

Although posaconazole has fewer interactions than itraconazole and voriconazole, there are many significant ones, including increased levels of calcium-channel blockers, cyclosporine, sirolimus, tacrolimus, quinidine, atazanavir, amiodarone, cisapride, corticosteroids, digoxin, benzodiazepines, and drugs for erectile dysfunction. Posaconazole levels are reduced by cimetidine, efavirenz, metoclopramide, rifampin, and, in some studies, with diarrhea and proton pump inhibitors. Posaconazole with phenytoin, carbamazepine, or rifabutin result in lower posaconazole levels and elevated levels of the other drug.

Adverse Effects

The paucity of side effects is similar to that of fluconazole. Gastrointestinal symptoms, headache, and acneiform rashes of the face are occasionally seen. Electrolyte abnormalities should be rectified before therapy to avoid arrhythmias.

Indications

Posaconazole appears to possess a very broad spectrum of activity against molds, including many zygomycetes, a unique property among azoles, and

the least cross-reactivity among the azoles to mutations in the ergosterol synthesis pathway that cause resistance to other azoles. The main interest in the drug has been as prophylaxis and for salvage therapy of refractory mycoses. Efficacy in preventing opportunistic mycoses has been shown in the patients at highest risk, transplant patients with graft-versus-host disease, and those with hematologic malignancies and neutropenia. Impressive results in salvage have been demonstrated in aspergillosis and coccidioidomycosis. Superficial candidiasis is also responsive.

ECHINOCANDIN ANTIFUNGAL AGENTS

General Features

The echinocandin antifungal agents act by inhibiting the synthesis of 1,3-β-D-glucan in the fungal cell wall. In studies, the principal pharmacodynamic driver of in vivo response has been the ratio of the peak concentration to the MIC. There are now three licensed echinocandins: caspofungin, micafungin, and anidulafungin. These agents are similar to one another, having very few drug interactions, an admirably low adverse event rate, and essentially identical antifungal spectra. Of the three drugs, caspofungin has the broadest array of approved indications and the most clinical data. Micafungin has the most detailed data on its use in neonates and children. Anidulafungin appears to have somewhat fewer drug-drug interactions. All are cyclic lipopeptides that must be given intravenously.

They are fungicidal against all *Candida* spp, including isolates resistant to other agents. Reduced activity against isolates of *Candida parapsilosis* and *Candida guilliermondii* and a paradoxical effect whereby high concentrations that permit growth in vitro have been noted in a minority of *Candida* isolates and do not appear to be clinically relevant. Resistance, while uncommon, is increasingly noted, and instances of resistance development on therapy have been described. All are active against *Aspergillus* spp, but activity is limited to growing and dividing hyphal elements. Their activity against other fungi is variable.

Caspofungin

Formulations and Pharmacology

Caspofungin is marketed for intravenous infusion. The clearance of caspofungin is through a combination of spontaneous chemical degradation, hydrolysis, and *N*-acetylation. Dose adjustments are not required for impaired renal function or hemodialysis. The clearance of caspofungin is modestly reduced in subjects with moderate hepatic insufficiency, and a reduction of the usual daily dose is recommended. Penetration into infected tissues appears to be good.

Drug Interactions

Caspofungin has few meaningful drug interactions. Cyclosporine coadministration increases caspofungin exposure and has been associated with increased hepatic transaminase levels, so concomitant use requires caution. Caspofungin reduces tacrolimus exposure by approximately 20%, and dosage adjustments may be required. Rifampin reduces caspofungin blood levels by approximately 30%, so the daily dosage of caspofungin should be increased from 50 to 70 mg if these drugs are co-administered. Likewise, limited data on other inducers of drug clearance (efavirenz, nevirapine, phenytoin, dexamethasone, carbamazepine) suggest that reduced caspofungin levels are possible and that increasing the daily dose to 70 mg should be considered.

Adverse Effects

Overall, the adverse reactions with caspofungin have been infrequent and minor. Symptoms possibly related to histamine release during rapid infusion have been reported.

Indications

Caspofungin is indicated for the treatment of invasive candidiasis and esophageal candidiasis. Caspofungin is also indicated for invasive aspergillosis in patients who are refractory to or intolerant of other therapies. With its activity against these two major opportunistic pathogens, the use of caspofungin for empirical therapy in high-risk situations is logical and is supported by clinical trials.

Micafungin

Micafungin has an in vitro spectrum and properties similar to those of caspofungin. It is similarly cidal for *Candida* spp and noncidal in vitro for *Aspergillus*.

Formulations and Pharmacology

The drug is light sensitive. After an intravenous infusion, the terminal half-life is approximately 15 hours. The recommended dose is 100 mg/day. Once daily dosing produces steady state in 3 days. Metabolism appears to be mostly through fecal excretion; less than 1% is excreted in the urine. The drug appears to penetrate particularly well into the lungs, liver, kidneys, and gastric mucosa. No adjustment for hepatic or renal insufficiency is needed.

Drug Interactions

Micafungin has fewer drug interactions than caspofungin. There is no interaction with cyclosporine or rifampin.

Adverse Effects

Micafungin has an excellent safety profile. Histamine release, manifested most prominently by erythema over the body, can be avoided by slow infusion. Phlebitis has been reported.

Indications

Efficacy is good in invasive candidiasis, similar to that of amphotericin or caspofungin. In clinical dose-finding studies against *Candida* esophagitis, doses of 75 to 150 mg/day were particularly effective. A trial of prophylaxis in neutropenic patients with hematopoietic stem cell transplant showed efficacy at 50 mg/day. Limited experience suggests comparable activity to caspofungin against aspergillosis.

Anidulafungin

Formulations and Pharmacology

Anidulafungin is slowly degraded by the chemical opening of its ring structure. A 100-mg dose produces a C_{max} (peak concentration achieved) of 8.6 μg/mL at a time of peak concentration (T_{max}) of 6 to 7 hours. The half-life is 30 to 50 hours. Steady state is achieved in 3 to 10 days. Dose adjustments are not required for renal or hepatic insufficiency.

Drug Interactions

Anidulafungin has no meaningful CYP enzyme interactions. It is not an inducer, inhibitor, or substrate and has no interaction with drugs cleared by these pathways or drugs that induce these pathways. The lack of metabolism suggests this would be the best echinocandin to use in the presence of liver failure.

Adverse Effects

Histamine-mediated symptoms (e.g., rash, urticaria, flushing, pruritus, dyspnea, hypotension) have been noted on occasion, but they are infrequent when the rate of infusion does not exceed 1.1 mg/minute. Other adverse reactions to anidulafungin have generally been infrequent and minor (diarrhea, hypokalemia, elevated liver function tests, neutropenia, nausea, headache, dermatitis).

Indications

Anidulafungin is indicated for the treatment of invasive candidiasis and esophageal candidiasis. Randomized trials indicate that anidulafungin is at least as efficacious as fluconazole. Inefficacy against *C. parapsilosis* may be of greatest concern with this echinocandin.

OTHER AGENTS

Combinations of antifungal agents are increasingly being used in an effort to improve the very poor response rates in some diseases and to combine drugs with different mechanisms of action. Although synergy in vitro may be relatively easy to demonstrate, there are few examples of synergy in animal model studies. In patients, the costs are high and the side effects are likely to be increased. The clinical advantages of the amphotericin-flucytosine combination in cryptococcosis have yet to be documented definitively in other situations.

Immunomodulators, particularly cytokines, hold promise, but insufficient clinical data exist to determine where and how these drugs might be used.[4] Granulocyte-macrophage colony-stimulating factor was shown to be effective in one trial when given prophylactically during induction therapy of patients with leukemia who are high risk.

Therapy of the various forms of tinea and of onychomycosis with systemic agents is discussed in Chapter 438. Although *Pneumocystis jiroveci* (formerly

Pneumocystis carinii) is now classified among the fungi, the drugs used to treat it are principally those used to treat parasitic infections (Chapter 344).

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Pfaller MA. Antifungal drug resistance: mechanisms, epidemiology, and consequences for treatment. *Am J Med*. 2012;125:s3-s13.
2. Pappas PG. Antifungal clinical trials and guidelines: what we know and do not know. *Cold Spring Harb Perspect Med*. 2014;4:a019745.
3. Pilmis B, Jullien V, Sobel J, et al. Antifungal drugs during pregnancy: an updated review. *J Antimicrob Chemother*. 2015;70:14-22.
4. Hope WW, Goodwin J, Felton T, et al. Population pharmacokinetics of conventional and intermittent dosing of liposomal amphotericin B in adults: a critical step for rational design of alternative regimens. *Antimicrob Agents Chemother*. 2012;56:5303-5308.
5. Sandherr M, Maschmeyer G. Pharmacology and metabolism of voriconazole and posaconazole in the treatment of invasive aspergillosis: review of the literature. *Eur J Med Res*. 2011;16:139-144.

332

HISTOPLASMOSIS

CAROL A. KAUFFMAN

DEFINITION

Histoplasmosis is the most common endemic mycosis in the United States. Most infections are self-limited, but the organism has the ability to cause acute and chronic pulmonary infections and disseminated infection.[1,2]

The Pathogen

Histoplasma capsulatum var *capsulatum* is a thermally dimorphic fungus. In the environment and at temperatures lower than 35° C, it exists as a mold that produces conidia. It produces both tuberculate macroconidia, which are helpful for identification purposes in the laboratory, and microconidia, which are the infectious form. In tissues and at 35 to 37° C, *H. capsulatum* transforms into tiny 2- to 4-μm oval yeasts that reproduce by budding and parasitize macrophages. African histoplasmosis is caused by a different subspecies, *H. capsulatum* var *duboisii,* and has different disease manifestations, particularly involvement of skin, subcutaneous tissues, lymph nodes, and bone, and only rarely the lungs and other internal organs.

EPIDEMIOLOGY

Histoplasmosis, though found worldwide, is primarily a disease of North and Central America. *H. capsulatum* is endemic in the Mississippi and Ohio River valleys, with extension into the St. Lawrence basin; microfoci exist in discrete isolated areas in several eastern states. Soil, caves, and abandoned buildings containing high concentrations of bird or bat guano support luxuriant growth of the organism. Every year, hundreds of thousands of individuals who live in areas endemic for *H. capsulatum* are infected. Most cases are sporadic and the exact source of exposure is unknown. Point source outbreaks that have included as few as 4 persons and as many as 100,000 have been well described in association with disruption of soil; cleaning attics, bridges, or barns; renovating or tearing down old structures laden with guano; and spelunking.[3]

PATHOBIOLOGY

After inhalation of microconidia into the alveoli, a localized pulmonary infection ensues. Neutrophils and macrophages phagocytize the organism, now in the yeast phase; the organism is able to survive and travels within macrophages to the hilar and mediastinal lymph nodes and throughout the reticuloendothelial system by hematogenous dissemination. Such dissemination probably occurs in most persons who are infected and in normal hosts is associated with no symptoms. After several weeks, T cells specifically sensitized by *H. capsulatum* antigens activate macrophages, which are then able to kill the intracellular fungi. Histoplasmosis is a classic example of the pivotal importance of the cell-mediated immune system in containing intracellular pathogens.

The extent of disease is determined by both the number of conidia inhaled and the immune response of the host. A small inoculum can cause severe pulmonary infection or progress to acute symptomatic disseminated histoplasmosis in immunosuppressed patients. Conversely, severe life-threatening pulmonary infection may develop in a healthy individual if a large number of conidia are inhaled, as might occur during the demolition of old buildings or while spelunking in a heavily infested cave.

Reinfection can occur in persons who previously had histoplasmosis but is uncommon and almost always occurs in the setting of heavy exposure. Reactivation of latent infection takes place in patients who have deficient cell-mediated immunity, as evidenced by the occurrence of histoplasmosis in immunosuppressed persons who grew up in the endemic area but have not been back in that area for years.

CLINICAL MANIFESTATIONS

Acute Pulmonary Histoplasmosis

Infection is asymptomatic in most people infected with *H. capsulatum*. Those who do have symptomatic pulmonary infection usually have a self-limited illness that begins several weeks after exposure and is characterized by fever, chills, fatigue, nonproductive cough, anterior chest discomfort, and myalgias. A patchy lobar or multilobar nodular infiltrate is noted on chest radiographs.

The differential diagnosis of acute pulmonary histoplasmosis includes pneumonia from *Blastomyces dermatitidis, Mycoplasma pneumoniae, Legionella* sp, and *Chlamydia pneumoniae*. When enlarged hilar or mediastinal lymph nodes are present, histoplasmosis should be strongly considered. The most difficult to differentiate is acute pulmonary blastomycosis (Chapter 334) because the endemic areas overlap, a comparable history of outdoor activities is often obtained, and radiographs show similar findings.

In patients who have experienced heavy exposure to *H. capsulatum* and in those who are immunosuppressed, acute pulmonary histoplasmosis can be life-threatening. High spiking fevers, chills, prostration, dyspnea, and cough are prominent.[4] Chest radiographs show diffuse reticulonodular pulmonary infiltrates, and acute respiratory distress syndrome (ARDS) can occur.

Chronic Pulmonary Histoplasmosis

Chronic cavitary pulmonary histoplasmosis is a progressive, often fatal form of histoplasmosis that develops almost exclusively in older patients who have chronic obstructive pulmonary disease.[5] Symptoms include fever, fatigue, anorexia, weight loss, cough productive of purulent sputum, and hemoptysis. On chest radiography the usual findings are unilateral or bilateral upper lobe infiltrates with multiple cavities and extensive fibrosis in the lower lobes. Chronic pulmonary histoplasmosis mimics tuberculosis, other fungal pneumonias (especially blastomycosis and sporotrichosis), and nontuberculous mycobacterial infections with regard to symptoms, signs, and radiographic findings.

Complications of Pulmonary Histoplasmosis

The mediastinal and hilar lymph nodes frequently calcify as the infection resolves; years later they can erode into bronchi and cause hemoptysis and expectoration of broncholiths. Granulomatous mediastinitis is an uncommon syndrome characterized by continuing inflammation and necrosis in the mediastinal lymph nodes. The enlarged nodes are readily apparent on chest radiographs, and computed tomography (CT) shows central necrosis and, in some cases, impingement on adjacent structures, including the esophagus, airways, and blood vessels. Although the symptoms usually resolve without treatment, obstructive syndromes can be severe and the nodes can persist for years.

Fibrosing mediastinitis[6] is a rare complication of histoplasmosis in which the host responds to the infection with an inappropriate excessive fibrotic response. Obstruction of the airways, superior vena cava, or pulmonary arteries and veins can occur with resultant progressive right heart failure and respiratory insufficiency. Bilateral obstruction of the pulmonary vasculature is less common than unilateral involvement and carries a worse prognosis. Mediastinal widening is seen on chest radiographs, and CT and angiography define the extent of invasion and obstruction of mediastinal structures.

Pericarditis is a manifestation of a local inflammatory reaction to adjacent histoplasmosis. Patients respond promptly to anti-inflammatory medications without antifungal therapy. Hemodynamic compromise, though unusual, requires drainage of pericardial fluid; only rarely has progression to constrictive pericarditis been documented.

Disseminated Histoplasmosis

Symptomatic disseminated histoplasmosis occurs mostly in immunosuppressed patients. Patients with acquired immunodeficiency syndrome (AIDS) and CD4+ counts lower than 150/μL, infants, and those who have a hematologic malignancy, have received an organ transplant, or are taking corticosteroids or tumor necrosis factor antagonists are at greatest risk for acute disseminated histoplasmosis.[7] Symptoms and signs include chills, fever, anorexia, weight loss, hypotension, dyspnea, hepatosplenomegaly, and skin and mucous membrane lesions. Pancytopenia, diffuse pulmonary infiltrates

on chest radiography and CT, findings of disseminated intravascular coagulation, and acute respiratory failure are common. This syndrome is indistinguishable from sepsis of any bacterial or viral cause. In patients with acquired immunodeficiency syndrome (AIDS), the differential diagnosis includes cytomegalovirus, disseminated *Mycobacterium avium* complex infection, and tuberculosis.

Chronic progressive disseminated histoplasmosis is a fatal form of histoplasmosis that occurs mostly in middle-aged to elderly men who have no known immunosuppressive illness. The illness is characterized by fever, night sweats, weight loss, anorexia, and fatigue. Patients appear chronically ill, hepatosplenomegaly and mucocutaneous ulcerations are common, and adrenal insufficiency can develop. An increased erythrocyte sedimentation rate, elevated alkaline phosphatase, pancytopenia, and diffuse reticulonodular infiltrates on chest radiography and CT are typical. Patients with this form of histoplasmosis often have fever of unknown origin. Miliary tuberculosis, lymphoma, and sarcoidosis must be excluded.

Involvement of almost every organ system has been reported with disseminated infection. Adrenal insufficiency must be sought in any patient who has unexplained hypotension, hyponatremia, and hyperkalemia. Abdominal CT shows markedly enlarged adrenal glands. Central nervous system involvement manifests as either meningitis or focal lesions seen on magnetic resonance imaging and is more common in patients with AIDS. Skin lesions, also more common in patients with AIDS, can be papular, pustular, or ulcerated. *Histoplasma* endocarditis is a rare form of disseminated infection.[8]

DIAGNOSIS

The definitive diagnostic test for histoplasmosis is growth of the organism in culture. Unfortunately, *H. capsulatum* may take as long as 6 weeks to grow in vitro. Tissue samples, bronchoalveolar lavage fluid, sputum, and blood are appropriate for culture. For patients who have evidence of dissemination, blood cultures are best performed with the lysis-centrifugation (Isolator tube) system; bone marrow and liver biopsy material often yield *H. capsulatum* in the setting of dissemination. If pulmonary histoplasmosis is a diagnostic consideration, the laboratory should be informed so that a special medium that decreases the growth of commensal fungi can be used for the culture of pulmonary samples. As soon as growth of a mold has been detected, highly specific DNA probes for *H. capsulatum* allow rapid identification of the organism.

If the patient is acutely ill, tissue biopsy should be performed to search for the distinctive 2- to 4-µm oval yeasts with single buds, which allows a tentative diagnosis to be made as quickly as possible. Routine tissue stains do not show the tiny yeasts; biopsy material must be stained with methenamine silver or periodic acid–Schiff stains. In patients with disseminated disease, bone marrow, liver, skin, and mucocutaneous lesions usually reveal many organisms. The organisms also can be seen within neutrophils in peripheral blood smears from patients with acute disseminated infection. In patients with chronic pulmonary histoplasmosis or granulomatous mediastinitis, biopsy of the lung or lymph nodes may reveal the organism.

Serology plays an important role in the diagnosis of some forms of histoplasmosis. Complement fixation assays that use two different antigens—mycelial and yeast—and immunodiffusion tests are available. Serology is helpful in diagnosing acute pulmonary histoplasmosis when a four-fold rise in complement fixation titer, a complement fixation titer higher than 1:32, or the appearance of an M precipitin band by immunodiffusion assay is found. These tests also are useful in patients with chronic forms of pulmonary or disseminated histoplasmosis, but they are rarely helpful in immunosuppressed patients, who often cannot mount an antibody response. Because serologic tests are not definitive in patients with mediastinal lymphadenopathy, the results always should be confirmed by tissue biopsy. An enzyme immunoassay for *H. capsulatum* polysaccharide antigen in urine and serum is extremely helpful in patients with disseminated infection and in those with acute pulmonary histoplasmosis.[9] Almost 75% of patients who have experienced heavy exposure and demonstrate diffuse pulmonary infiltrates and more than 90% of patients who have disseminated histoplasmosis will have a positive urinary antigen test. However, the assay is not useful for patients who have chronic cavitary histoplasmosis, granulomatous mediastinitis, or fibrosing mediastinitis. False-positive reactions have been noted commonly with blastomycosis, paracoccidioidomycosis, and penicilliosis and occasionally with acute coccidioidomycosis. Antigen levels usually become undetectable with successful therapy. Polymerase chain reaction has been used for diagnosis in some cases, but the assays are not standardized.

TREATMENT

Rx

Guidelines for the treatment of histoplasmosis have been published by the Infectious Diseases Society of America (IDSA).[10] Itraconazole is the drug of choice for mild-to-moderate histoplasmosis, and lipid formulations of amphotericin B should be used as initial therapy for severe, life-threatening infections. Fluconazole is less active and should be considered a second-line agent. There is less experience using voriconazole or posaconazole, although both appear to be effective therapy. The echinocandins are not effective for histoplasmosis.

For all patients who are treated with itraconazole, serum levels should be determined when a steady state has been reached after 2 weeks of therapy to ensure adequate absorption. Serum concentrations should be greater than 1 µg/mL.

Pulmonary Histoplasmosis

Treatment is not usually given for acute pulmonary histoplasmosis; many times the diagnosis is not made until after the symptoms have resolved. However, if the patient remains symptomatic after 4 weeks, therapy with itraconazole, 200 mg once or twice daily for 6 to 12 weeks, should be given. All patients who have severe pulmonary histoplasmosis and all immunosuppressed patients should be treated with an antifungal agent. Lipid formulation amphotericin B, 3 to 5 mg/kg/day, is recommended for several weeks until a response is noted, at which time therapy can be changed to oral itraconazole, 200 mg three times daily for 3 days and then 200 mg twice daily. A short course of methylprednisolone is recommended in the IDSA guidelines for patients in whom respiratory distress develops in association with acute pulmonary histoplasmosis.

Antifungal therapy should be given to all patients with chronic pulmonary histoplasmosis. Itraconazole, 200 mg twice daily for 12 to 24 months, is recommended. A trial of itraconazole for 6 to 12 weeks is often given to patients with symptomatic granulomatous mediastinitis, although there are no data proving such therapy to be effective. Surgical resection of nodes causing obstructive symptoms may be beneficial. Antifungal therapy offers no benefit for patients with fibrosing mediastinitis. Surgery is not indicated and carries a high operative mortality rate. Intravascular stents have been used successfully in patients who have vascular obstruction.

Disseminated Histoplasmosis

All patients with symptomatic disseminated histoplasmosis should receive antifungal therapy. Patients who have only mild-to-moderate symptoms with acute disseminated disease and most patients with chronic progressive disseminated histoplasmosis can be treated with itraconazole, 200 mg twice daily after a loading dose of 200 mg three times daily for 3 days. A total of 12 months of therapy is generally adequate, but for those with chronic progressive disease, the duration of treatment may need to be longer.

Patients who have moderately severe to severe disseminated histoplasmosis should be treated with liposomal amphotericin B, 3 mg/kg/day. Therapy can be changed to itraconazole after the patient has responded to therapy, generally in a few weeks. Therapy should continue for a total of 12 months. For those who are immunosuppressed, therapy with itraconazole, 200 mg/day, should continue until the immunosuppression resolves. For patients with AIDS, suppressive therapy can be safely stopped in those who have had a year of therapy, are receiving antiretroviral therapy, and have CD4+ counts greater than 150 cells/µL.[11]

Prevention of histoplasmosis is difficult because exposure can occur without the person's knowledge in highly endemic areas. Persons who are immunosuppressed should be advised to avoid demolition areas, spelunking, and cleaning of farm buildings or attics. A randomized, blinded, placebo-controlled trial in patients with AIDS found that itraconazole, 200 mg/day, was effective in preventing infection. Recommendations are to use prophylaxis only for patients whose CD4+ counts are lower than 150/µL and who live in a highly endemic area with a rate of histoplasmosis that is greater than 10 cases per 100 patient years. There are no recommendations for prophylaxis for other immunosuppressed patients.

PROGNOSIS

Acute pulmonary histoplasmosis is usually a self-limited disease. Patients who require treatment generally respond promptly to antifungal agents. However, the response of patients with chronic cavitary pulmonary histoplasmosis is often poor, primarily because of their severe underlying pulmonary disease. Mediastinal fibrosis has a poor prognosis, but intravascular stenting has led to improvement in many patients. Patients with disseminated histoplasmosis, even those with advanced AIDS, usually respond promptly to antifungal therapy; most deaths in immunosuppressed patients occur when the diagnosis is delayed. Older patients with chronic progressive disseminated histoplasmosis have a slower, but usually complete response to therapy.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Kauffman CA. Histoplasmosis. *Clin Chest Med*. 2009;30:217-225.
2. Knox KS, Hage CA. Histoplasmosis. *Proc Am Thorac Soc*. 2010;7:169-172.
3. Centers for Disease Control and Prevention. Histoplasmosis outbreak associated with the renovation of an old house: Quebec, Canada, 2013. *MMWR Morb Mortal Wkly Rep*. 2014;62: 1041-1044.
4. Hage CA, Bowyer S, Tarvin SE, et al. Recognition, diagnosis, and treatment of histoplasmosis complicating tumor necrosis factor blocker therapy. *Clin Infect Dis*. 2010;50:85-92.
5. Ledtke C, Tomford JW, Jain A, et al. Clinical presentation and management of histoplasmosis in older adults. *J Am Geriatr Soc*. 2012;60:265-270.
6. McNeeley MF, Chung JH, Bhalla S, et al. Imaging of granulomatous and fibrosing mediastinitis. *AJR Am J Roentgenol*. 2012;199:319-327.
7. Kauffman CA, Freifeld AG, Andes DR, et al. Endemic fungal infections in solid organ and hematopoietic cell transplant recipients enrolled in the Transplant Associated Infection Surveillance Network (TRANSNET). *Transpl Infect Dis*. 2014;16:213-224.
8. Riddell J, Kauffman CA, Smith JA, et al. Histoplasma capsulatum endocarditis: multi-center case series with review of current diagnostic techniques and treatment. *Medicine (Baltimore)*. 2014;93: 186-193.
9. Hage CA, Ribes JA, Wengenack NL, et al. A multicenter evaluation of tests for diagnosis of histoplasmosis. *Clin Infect Dis*. 2011;53:448-454.
10. Wheat LJ, Freifeld AG, Kleiman MJ, et al. Clinical practice guidelines for the management of patients with histoplasmosis: 2007 update by the Infectious Diseases Society of America. *Clin Infect Dis*. 2007;45:807-827.
11. Myint T, Anderson AM, Sanchez A, et al. Histoplasmosis in patients with human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS): multicenter study of outcomes and factors associated with relapse. *Medicine (Baltimore)*. 2014;93:11-18.

REVIEW QUESTIONS

1. A 72-year-old man from Kentucky who has emphysema and coronary artery disease comes in with fever, night sweats, 20-lb weight loss, cough, and sputum production with occasional hemoptysis for the last 4 to 5 months. Chest x-ray shows bilateral upper lobe cavitary infiltrates and emphysematous changes. The *most* likely diagnosis is:

A. Coccidioidomycosis.
B. Sporotrichosis.
C. Blastomycosis.
D. Histoplasmosis.

Answer: D. Any of the above fungal infections can manifest by upper lobe cavitary lesions. However, the most likely diagnosis is histoplasmosis because he is from Kentucky, where *H. capsulatum* is highly endemic. Blastomycosis is also endemic in Kentucky but is less common than histoplasmosis. Pulmonary sporotrichosis could manifest with a similar picture, but is very uncommon. Coccidioidomycosis occurs in the desert southwest and is not possible unless he had traveled to that area.

2. Which test would be most useful for the diagnosis of chronic cavitary pulmonary histoplasmosis in the patient noted in question 1?

A. Serum antigen assay for *H. capsulatum* cell wall polysaccharide
B. CT scan of the chest
C. Sputum culture for fungus
D. PCR for *H. capsulatum* on serum
E. Urine enzyme immunoassay

Answer: C. The definitive diagnostic test is growing *H. capsulatum* in culture, and in this form of histoplamosis, sputum cultures usually yield the organism. PCR tests can be useful when performed on a tissue biopsy, but not on serum. The serum and urine antigen assays are not sensitive for the chronic cavitary form of histoplasmosis. CT scan is useful to define the extent of pulmonary disease, but it does not define the causative agent.

3. Which form of histoplasmosis does not respond to antifungal therapy?

A. Chronic cavitary pulmonary
B. Mediastinal fibrosis
C. Disseminated infection in immunocompetent patients
D. Acute respiratory distress syndrome (ARDS)
E. Disseminated infections in patients with AIDS

Answer: B. Mediastinal fibrosis is characterized by an abnormal host response with the production of excessive fibrosis that can lead to obstruction of pulmonary arteries, the superior vena cava, and mainstem bronchi. The organism is rarely seen within the fibrous tissue, and it is thought that there is no longer active infection when the patient presents with obstructive symptoms. Disseminated histoplasmosis, including in patients with AIDS, and acute and chronic pulmonary histoplasmosis, including in patients with AIDS, respond to antifungal therapy.

4. A 65-year-old man is admitted with fevers and hypotension. He has a 3-month history of night sweats and weight loss, and he is found to have anemia (Hb 9.5 g/dL), thrombocytopenia (platelets 112,000/μL), and leukopenia (2600 WBC/μL). His serum sodium is 126 mEq/L and potassium is 5.2 mEq/L. What is the most likely fungal infection that could explain these findings?

A. Coccidioidomycosis
B. Sporotrichosis
C. Histoplasmosis
D. Cryptococcosis
E. Candidiasis

Answer: C. The patient has adrenal insufficiency, and the fungus that most often causes this is *H. capsulatum*. The other fungal infections have been reported very rarely to cause adrenal insufficiency. This manifestation is not uncommon in patients with AIDS who develop disseminated histoplasmosis and in older adults who develop the chronic progressive form of disseminated histoplasmosis.

5. A 37-year-old woman is being treated with etanercept for rheumatoid arthritis. She developed disseminated histoplasmosis 3 weeks after an old barn on her farm had been demolished. On admission to hospital, she has hypoxemia, pancytopenia, and signs of disseminated intravascular coagulation. She should be treated with:

A. Liposomal amphotericin B.
B. Itraconazole.
C. Amphotericin B deoxycholate.
D. Fluconazole.

Answer: A. Liposomal amphotericin B has been shown to be superior to amphotericin B deoxycholate for the treatment of severe disseminated histoplasmosis in patients with AIDS. Mortality was decreased, patients became afebrile sooner, and toxicity was less with the liposomal formulation. Any immunosuppressed patient who has severe histoplasmosis should be treated with this agent. Itraconazole is appropriate therapy for mild-to-moderate histoplasmosis and for step-down therapy after the patient has responded to amphotericin B. Fluconazole is less active against *H. capsulatum* and is a second-line agent.

333

COCCIDIOIDOMYCOSIS

JOHN N. GALGIANI

DEFINITION

Coccidioidomycosis is a systemic fungal infection caused by *Coccidioides* spp endemic to some arid regions of the Western Hemisphere (Table 333-1).

The Pathogen

Coccidioides immitis and *C. posadasii* are dimorphic fungi classified as Ascomycetes by ribosomal gene homology. In their vegetative state, mycelia with true septations mature to produce arthroconidia, single cells approximately 2 to 5 µm in diameter. After infection, an arthroconidium enlarges to a spherule up to 75 µm in diameter and undergoes internal septation to produce scores of endospores. When the spherules rupture, packets of endospores are released and produce more spherules in infected tissue or revert to mycelia if removed from the body.

EPIDEMIOLOGY

Coccidioides organisms can be recovered from the soil of the low deserts of Arizona; the Central Valley of California; parts of other states, including New Mexico and Texas; and parts of Central and South America.[1] Isolated pockets of endemnicity have been found unexpectedly elsewhere such as in Washington State.[2] Endemic regions follow the climatologic Sonoran life zone, which is characterized by modest rainfall, mild winters, and low humidity. However, even in the most highly endemic areas, fungal colonies are sparse and occupy only a tiny fraction of the total acreage. Mycelia bloom beneath the surface during periods of rain, and arthroconidia develop as the earth dries.[3] Rates of infection are highest during dry months, occasionally accentuated when soil is disturbed by windstorms or construction equipment. Exposure to contaminated bales of cotton or other fomites can rarely result in infection beyond the endemic regions. Person-to-person transmission of pulmonary infection has not been reported, and isolation precautions are unnecessary, even in acute care areas. As of 2013, *Coccidioides* spp are no longer listed and controlled by the Centers for Disease Control and Prevention as select agents.

TABLE 333-1 COCCIDIOIDOMYCOSIS: CLINICAL CHARACTERISTICS AND TREATMENT

CHARACTERISTIC	DESCRIPTION
Causative fungi	*Coccidioides immitis* and *Coccidioides posadasii*
Primary geographic distribution	Lower Sonoran deserts of the Western Hemisphere, including parts of Arizona, California, and New Mexico; western Texas; parts of Central and South America
Primary route of acquisition	Respiratory (inhalation of arthroconidia)
Principal site of disease	Lungs most common; spread to skin, bones, meninges, and other viscera uncommon but serious
Opportunistic infection in compromised hosts	Diffuse pneumonia and widespread infections common in patients with T-lymphocyte defects or during high-dose corticosteroid therapy
Drug of choice for most patients	No antifungal is required for uncomplicated pneumonia; fluconazole or itraconazole for progressive forms of infection
Alternative therapy	Amphotericin B (especially with diffuse pneumonia or rapidly progressive infections), voriconazole, posaconazole

Incidence and Prevalence

In general, the annual risk for infection within the most strongly endemic areas is 3%, resulting in approximately 150,000 new infections per year. Reported clinical illness following infection is increasing. For example, from 1998 to 2011, reported infections increased more than eight-fold.[4] With unusually intense exposure, such as at archaeology sites or during military maneuvers within endemic regions, infections can develop in the majority of persons exposed for only a matter of days. Arizona and California contribute 66% and 31%, respectively, of all U.S. infections.

PATHOBIOLOGY

Inhaling even a single arthroconidium to the level of the terminal bronchiole initiates virtually all coccidioidal infections. Fungal proliferation engenders both granulomatous inflammation, which is associated with intact spherules, and acute inflammation, including eosinophils, which is associated with spherule rupture and proliferation. Focal pneumonia is often associated with ipsilateral hilar adenopathy; less frequently, infection enlarges the paratracheal, supraclavicular, and cervical nodes. Lesions occurring elsewhere are the result of hematogenous dissemination, and most become apparent within 2 years of the initial infection. Although progressive dissemination occurs in less than 1% of infections, as many as 8% of persons with self-limited infection manifest asymptomatic chorioretinal scars, suggesting that subclinical hematogenous spread may be frequent. Within weeks after infection, durable T-cell immunity normally arrests fungal proliferation, which allows the inflammation to resolve and prevents reinfection in the future. However, control of the infection occurs without sterilizing lesions, and reactivation of the dormant infection is possible even many years later in patients whose cell-mediated immunity becomes deficient.

CLINICAL MANIFESTATIONS

Two thirds of infections are asymptomatic and are detected only by finding dermal hypersensitivity to coccidioidal antigens. Those who become ill usually experience pulmonary syndromes that eventually are self-limited. However, some patients develop complications or progressive forms of infection that display a broad variety of manifestations and pose difficult management problems.

Primary Pulmonary Infections

Symptoms develop within 5 to 21 days after exposure. For residents of or recent visitors to southern Arizona, coccidioidomycosis accounts for approximately one third of cases of community-acquired pneumonia.[5] Fever, weight loss, fatigue, dry cough, and pleuritic chest pain are common but nonspecific complaints. Arthralgia of multiple joints without significant effusion is also frequent and is referred to as "desert rheumatism." Occasionally, skin manifestations develop, including a short-lived nonpruritic maculopapular rash, erythema multiforme, or erythema nodosum. These arthritic and dermatologic manifestations are mediated by circulating immune complexes or other immunologic phenomena rather than by fungal dissemination and resolve without tissue destruction. Radiographs of the chest may not detect any abnormalities or may demonstrate pulmonary infiltrates that are either segmental or lobar. Hilar adenopathy is often a distinctive finding and may suggest lymphoma by its appearance. Peripneumonic pleural effusions may occur and generally resolve without intervention, although cultures of pleural biopsies usually yield *Coccidioides* spp. Eosinophilia may be a prominent finding in differential leukocyte counts of peripheral blood, and the erythrocyte sedimentation rate is generally elevated. Symptoms frequently persist for many weeks before improvement is clearly under way, and the illness, especially lassitude, may persist for months.[6]

The primary pulmonary process produces a variety of sequelae. The most frequent is the development of a pulmonary nodule (Fig. 333-1), typically measuring 1 to 4 cm and lying within 5 cm of the hilum. Despite their harmless nature, coccidioidal nodules may cause concern because of their similarity to a malignant mass (Chapters 84 and 191). Positron emission tomography scans are typically positive. For these reasons, management usually requires percutaneous needle aspiration or resection. Another consequence of pulmonary coccidioidomycosis is cavitation of the infiltrate, which occurs in approximately 5% of cases of pneumonia.[7] Most cavities are solitary and thin walled and reside in an upper lobe, close to the pleura. Occasionally, they produce pain, hemoptysis, or adjacent infiltrates. Cavities may acquire mycetomas from either *Coccidioides* spp or some other colonizing mold. Infrequently, a cavity ruptures and forms a pyopneumothorax. Half of the time,

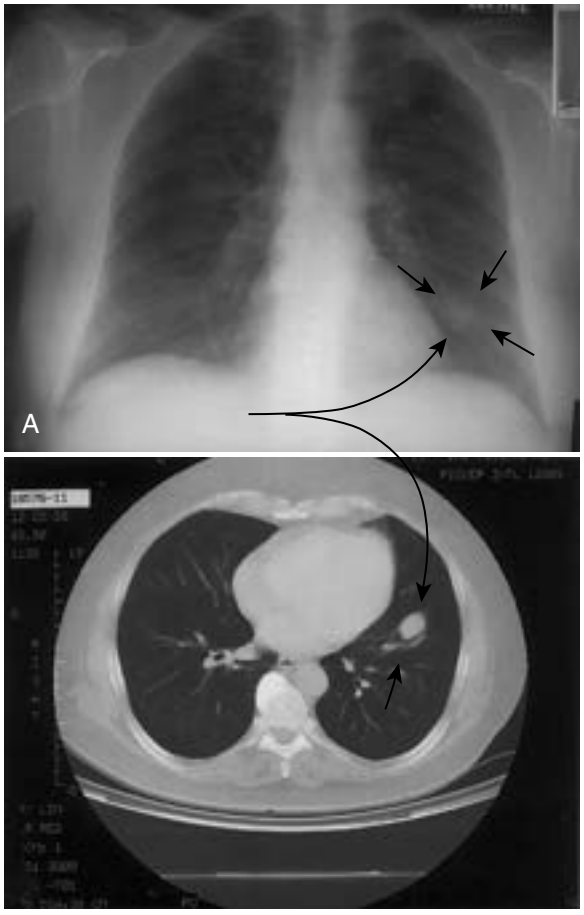

FIGURE 333-1. Coccidioidomycosis. A, Benign nodule secondary to coccidioidomycosis (*arrows*). **B,** Computed tomography scan of the nodule shown in **A** (*arrows*).

this is the first symptom of coccidioidal infection and typically occurs in otherwise healthy young men. An air-fluid level in the pleural space, detectable by roentgenography, often helps differentiate this problem from a spontaneous pneumothorax. Prompt surgical resection of the cavity is the preferred treatment of this complication. The least common pulmonary complication is persistent fibrocavitary infection that progresses to involvement of both lungs. Rarely, a mutation of the *STAT1* gene results in chronic progressive noncavitary pulmonary destruction.[8]

Extrapulmonary Dissemination

Coccidioidomycosis in patients with deficiencies in cellular immunity, such as solid organ recipients, those with acquired immunodeficiency syndrome (AIDS) or lymphoma, and women during their third trimester of pregnancy, usually results in dissemination beyond the lungs. Mutations of the genes for γ-interferon or the interleukin-12 receptor also predispose persons to dissemination.[9] However, disseminated infection occurs in some patients who have no underlying disease and do not manifest heightened susceptibility to other infections. Disseminated infection is more likely in men than in women and in persons of certain ancestry such as Africans, Filipinos, or Native Americans compared to Caucasians. The most common locations for disseminated lesions are the skin (cutaneous papules or subcutaneous abscesses); joints (especially the knee); bones, including the vertebrae; and basilar meninges.[10] Such infections may produce one or many lesions and are frequently subacute or chronic in their manifestation. In broadly immunosuppressed patients, coccidioidal infections may be more fulminant, with fungemia detectable in blood cultures and the development of diffuse reticulonodular embolic pulmonary infiltrates. Although the kidneys and the urinary bladder are rarely involved, *Coccidioides* may be recovered from concentrated specimens of urine, usually because of focal dissemination to the prostate. In contrast to histoplasmosis, the gastrointestinal tract is rarely involved in coccidioidomycosis.

DIAGNOSIS

The diagnosis is definitively established by recovering *Coccidioides* spp from clinical specimens. On direct examination of respiratory specimens or tissue, spherules appear as large structures with refractile walls and internal organization; they are also seen on hematoxylin-eosin, silver, or periodic acid–Schiff stains of histologic preparations. Gram stain does not detect spherules. In culture, mycelial growth is often evident within the first week of incubation, and DNA probing with commercially available kits allows rapid genus identification. Recovery of *Coccidioides* spp from patients with only scant respiratory secretions associated with the initial pneumonia and from the cerebrospinal fluid of patients with meningitis may be difficult.

A presumptive diagnosis of coccidioidal infection is often based on the detection of specific antibodies in serum. Within the first weeks of the initial infection, a precipitin-type antibody is detected, usually by immunodiffusion techniques. Later, complement fixation–type antibodies generally appear. However, these tests may be falsely negative as often as half the time during the first weeks of illness.[11] When reported quantitatively, concentrations of complement fixation antibodies are usually highest in the most extensive infections and decrease in patients whose infections are controlled. An important means of diagnosing coccidioidal meningitis is the detection of complement fixation antibodies in cerebrospinal fluid, along with other abnormalities such as leukocytosis, elevated protein concentration, or low glucose concentration. Newer enzyme immunoassay commercial kits are also available, are generally more sensitive than the older serologic tests, but occasionally produce falsely positive results. Coccidioidal antigens are sometimes found in the urine or serum of patients with widespread infection. A *Coccidioides* genus–specific, quantitative, real-time polymerase chain reaction assay has been developed for the early diagnosis of coccidioidomycosis.[12] In 2014, a coccidioidal skin test became clinically available and, if positive, indicates past infection.

TREATMENT

Rx

The role of antifungal therapy for primary uncomplicated infections is unsettled because clinical trials have not been performed to determine whether treatment either shortens the course of symptoms or diminishes the chance of complications.[13,14] However, the value of treatment is clear for patients with progressive illness.[15] Because many coccidioidal infections are chronic, initial treatment often consists of oral azole antifungal agents, such as fluconazole and itraconazole. Responses to these two drugs are similar, but itraconazole is preferred in patients with skeletal infections. Doses of these azoles are 400 mg/day or higher, and treatment is usually continued for a year or more. Satisfactory responses are obtained in approximately two thirds of patients. Fluconazole is effective therapy for coccidioidal meningitis and has greatly reduced the number of patients treated with intrathecal amphotericin B. Unfortunately, cessation of azole therapy is often followed by a recurrence of symptoms, especially in those with coccidioidal meningitis. Therefore, many patients need protracted or even lifelong therapy to control disease activity. The limited evidence available for the newer azole antifungals (voriconazole, posaconazole) indicates that they are also effective in some patients and are sometimes useful for refractory infections. Amphotericin B remains a rational choice when treatment with azole antifungals has failed. Daily doses range from 0.4 to 1 mg/kg for the original deoxycholate formulation and up to 5 mg/kg per day for newer lipid formulations. Occasionally, in a patient with rapid disease progression, amphotericin B may produce a more rapid therapeutic response and is therefore the preferred initial therapy. In addition to antifungal agents, surgical removal of necrotic tissue may be essential to control the progressive damage from specific lesions.

PROGNOSIS

After resolution of the initial infection, most patients maintain lifelong immunity, and infections after re-exposure are rare. However, cessation of symptoms is frequently accomplished without eradicating *Coccidioides* completely, and recurrence of the original infection many years after the original episode is a well-recognized risk for intercurrent profound immunosuppression. Re-treating patients with rheumatic disease with biologic response modifiers and/or disease-modifying antirheumatic drugs after coccidioidomycosis appears to be safe in some patients.[16] For patients in whom the initial infection cannot be resolved, the disease typically follows a protracted course. Although infection is more often debilitating than fatal, fulminant respiratory failure can occur, and, if untreated, coccidioidal meningitis is nearly always fatal within 2 years.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Brown J, Benedict K, Park BJ, et al. Coccidioidomycosis: epidemiology. *Clin Epidemiol*. 2013;5:185-197.
2. Litvintseva AP, Marsden-Haug N, Hurst S, et al. Valley fever: Finding new places for an old disease: *Coccidioides immitis* found in Washington State soil associated with recent human infection. *Clin Infect Dis*. 2015;60:e1-e3.
3. Nguyen C, Barker BM, Hoover S, et al. Recent advances in our understanding of the environmental, epidemiological, immunological, and clinical dimensions of coccidioidomycosis. *Clin Microbiol Rev*. 2013;26:505-525.
4. Centers for Disease Control and Prevention. Increase in reported coccidioidomycosis: United States, 1998-2011. *MMWR Morb Mortal Wkly Rep*. 2013;62:217-221.
5. Malo J, Luraschi-Monjagatta C, Wolk DM, et al. Update on the diagnosis of pulmonary coccidioidomycosis. *Ann Am Thorac Soc*. 2014;11:243-253.
6. Blair JE, Chang YH, Cheng MR, et al. Characteristics of patients with mild to moderate primary pulmonary coccidioidomycosis. *Emerg Infect Dis*. 2014;20:983-990.
7. Sobonya RE, Yanes J, Klotz SA. Cavitary pulmonary coccidioidomycosis: pathologic and clinical correlates of disease. *Hum Pathol*. 2014;45:153-159.
8. Sampaio EP, Bax HI, Hsu AP, et al. A novel *STAT1* mutation associated with disseminated mycobacterial disease. *J Clin Immunol*. 2012;32:681-689.
9. Vinh DC, Schwartz B, Hsu AP, et al. Interleukin-12 receptor β1 deficiency predisposing to disseminated coccidioidomycosis. *Clin Infect Dis*. 2011;52:e99-e102.
10. Murthy JM, Sundaram C. Fungal infections of the central nervous system. *Handb Clin Neurol*. 2014;121:1383-1401.
11. Malo J, Luraschi-Monjagatta C, Wolk DM, et al. Update on the diagnosis of pulmonary coccidioidomycosis. *Ann Am Thorac Soc*. 2014;11:243-253.
12. Gago S, Buitrago MJ, Clemons KV, et al. Development and validation of a quantitative real-time PCR assay for the early diagnosis of coccidioidomycosis. *Diagn Microbiol Infect Dis*. 2014;79:214-221.
13. Blair JE, Chang YH, Cheng MR, et al. Characteristics of patients with mild to moderate primary pulmonary coccidioidomycosis. *Emerg Infect Dis*. 2014;20:983-990.
14. Galgiani JN. Elements of style in managing coccidioidomycosis. *Clin Infect Dis*. 2013;56:1586-1588.
15. Limper AH, Knox KS, Sarosi GA, et al. An official American Thoracic Society statement: Treatment of fungal infections in adult pulmonary and critical care patients. *Am J Respir Crit Care Med*. 2011;183:96-128.
16. Taroumian S, Knowles SL, Lisse JR, et al. Management of coccidioidomycosis in patients receiving biologic response modifiers or disease-modifying antirheumatic drugs. *Arthritis Care Res (Hoboken)*. 2012;64:1903-1909.

REVIEW QUESTIONS

1. A patient in Seattle, Washington is diagnosed with community-acquired pneumonia that developed 2 weeks after attending a business conference in Phoenix, Arizona. The likelihood that the etiology of the infection is *Coccidioides* is:

A. 80%
B. 30%
C. 15%
D. 5%
E. Less than 1%

Answer: B Two prospective studies demonstrate this frequency within the endemic region. It is equally applicable to visitors with symptoms soon after endemic exposure.

2. A 56-year-old man, a long-time smoker who lives in the California central valley, is found to have a 2.5-cm right middle lobe nodule. He was diagnosed with valley fever 5 years ago. Evaluation should include:

A. Repeated CT scans every 6 months to determine if the nodule is enlarging.
B. Initiation of fluconazole 400 mg/day and re-evaluation for regression of the nodule.
C. Obtain tissue for microscopic examination, either by bronchoscopy or percutaneous fine-needle aspiration.
D. Perform edge resection by thoracotomy.
E. Obtain a coccidioidal serology.

Answer: C There is insufficient information to assume the nodule is related to the prior coccidioidal diagnosis. It is too large to simply observe in a patient at risk for lung cancer. There is a high probability (perhaps 50%) that the nodule is not cancer, which would make the thoracotomy overly invasive without first establishing that the lesion is a cancer. Antifungal treatment would be unlikely to affect the nodule size. Whether the coccidioidal serologic test was positive should not change the need for tissue to exclude cancer.

3. A patient develops coccidioidal meningitis, is started on fluconazole 400 mg/day and over the next 6 months completely resolves her symptoms. An MRI of the brain was never abnormal, and a follow-up lumbar puncture produces CSF with essentially normal laboratory values. Treatment is continued unchanged for 2 years. At that point, management should be:

A. Stop treatment and follow-closely for possible relapse.
B. Stop treatment only if reevaluation shows no evidence of pulmonary or extrapulmonary infection.
C. Continue treatment at a reduced dose.
D. Continue treatment at the current dosage indefinitely

Answer: D Relapses are frequent and potentially lethal. Even reducing the dose may lead to recurrence of infection.

4. A patient is found to have a thin-walled pulmonary cavity, 2 cm in diameter, abutting the visceral pleura in the left upper lobe. He has no symptoms associated with this. A coccidioidal complement fixation antibody test of serum returns a positive result at a 1 : 4 dilution. Management of this situation should be:

A. No treatment, with periodic reassessment with chest x-ray.
B. Start fluconazole 400 mg/day to see if the cavity responds to treatment by decreasing.
C. Start amphotericin B intravenously.
D. Resect the cavity by visually assisted transthoracic surgery (VATS) to avert the risk for cavity rupture and consequent bronchopleural fistula formation.

Answer: A Such cavities often follow an uncomplicated course. Treatment with antifungal agents does not affect the likelihood of their getting smaller. Although rupture occurs in some, most do not.

5. A woman in El Paso, Texas develops arthralgias in both ankles and erythema nodosum. She has no pulmonary symptoms. An enzyme immunosorbant assay for coccidioidal antibodies is positive for IgG antibodies. Which of the following is the *best* assessment of this current illness?

A. Because the EIA IgG finding indicates past infection, the etiology still is not clear.
B. This is likely the "desert rheumatism syndrome" of coccidioidomycosis.
C. The skin lesions are likely to become progressive destructive lesions, and biopsy should be performed to determine how best to treat the patient.
D. This presentation increases the likelihood that the patient eventually will develop serious complications.

Answer: B Because coccidioidal serology findings usually return to undetectable levels months following a coccidioidal infection, B is the correct answer. Patients with *E. nodosum* seldom develop future complications. A biopsy sample of the skin lesions would show an inflammatory process that is not specific to a particular etiology.

334

BLASTOMYCOSIS

CAROL A. KAUFFMAN

DEFINITION

Blastomycosis (North American blastomycosis) is an endemic mycosis that primarily causes infection of the lungs and skin and, less commonly, infection of the osteoarticular and genitourinary systems.[1]

The Pathogen

Blastomyces dermatitidis is a thermally dimorphic fungus. In the environment in the mold phase, the organism produces conidia, which when aerosolized and inhaled cause infection. At 37°C on culture media and in tissues, the organism is a yeast that is 5 to 20 µm in diameter, has a thick refractile cell wall, and produces single broad-based buds (Fig. 334-1)

EPIDEMIOLOGY

B. dermatitidis exists in many diverse geographic areas worldwide, but most cases of blastomycosis are reported from the south central and north central regions of the United States[2] and the Canadian provinces surrounding the Great Lakes. The natural niche of *B. dermatitidis* is thought to be soil and decaying vegetation, especially in areas associated with rivers and lakes. Although most cases occur sporadically, several well-described outbreaks have occurred in association with activities along waterways. The largest community outbreak was reported in 2009 and 2010 in Wisconsin, with geographic and ethnic clustering likely related to multifocal environmental sources.[3] The typical patient in whom blastomycosis develops is a middle-aged man who has an outdoor occupation or hobby. The association of blastomycosis developing in both hunters and their dogs is well known in endemic areas.

PATHOBIOLOGY

After the inhalation of conidia, *B. dermatitidis* transforms into the yeast phase and causes pulmonary infection. Although many patients manifest only pulmonary symptoms, others have cutaneous lesions in the absence of other organ involvement or have disseminated infection. It is likely that most patients have asymptomatic hematogenous dissemination after the initial pulmonary infection. Thus, cutaneous lesions should be viewed as a manifestation of hematogenous spread of the organism. Except in rare instances, blastomycosis is not acquired by inoculation. Cellular immunity involving T lymphocytes and macrophages is an important component of the host response to infection with *B. dermatitidis,* but neutrophils also play a role. Most patients with blastomycosis are healthy hosts. Patients who are immunosuppressed are more likely to have severe disease. Infection in an immunosuppressed host can occur after new exposure to *B. dermatitidis* or from reactivation of a latent focus of infection acquired years earlier.[4]

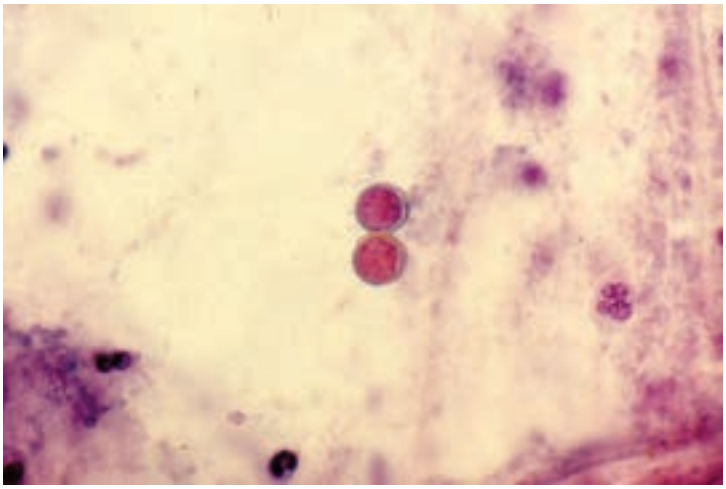

FIGURE 334-1. Papanicolaou stain of sputum showing a thick-walled yeast with broad-based budding typical of Blastomyces dermatitidis.

CLINICAL MANIFESTATIONS

Pulmonary

Most patients with acute pulmonary blastomycosis are asymptomatic or are thought to have community-acquired pneumonia. Patients with acute pneumonia have fever, malaise, a nonproductive cough, and a pulmonary infiltrate that shows lobar or multilobar patchy or nodular infiltrates on chest radiographs. Development of skin lesions is a strong clue for blastomycosis. Chronic pulmonary blastomycosis must be differentiated from tuberculosis, other fungal infections, and lung cancer.[5] Fever, night sweats, weight loss, fatigue, cough, sputum production, hemoptysis, and dyspnea are commonly noted. On chest radiograph the lesions are cavitary, nodular, fibrotic, or mass-like in appearance. Hilar and mediastinal lymphadenopathy and pleural effusions are not commonly seen. Overwhelming pulmonary disease with acute respiratory distress syndrome (ARDS) occurs infrequently but has a high mortality rate. Whether this is due to inhalation of a large number of conidia or to an exuberant host response is not known. Given improvement in some patients with corticosteroids, the latter may be important.

Disseminated Infection

Cutaneous lesions are the most common manifestation of disseminated blastomycosis. The lesions are usually well-circumscribed, painless papules, nodules, or plaques that become verrucous and develop multiple punctate draining areas in the center. Some patients have predominantly ulcerative lesions. Cutaneous lesions, sometimes single but more often multiple, are most common on the face and extremities but can appear anywhere. The skin lesions of blastomycosis clinically mimic those associated with nontuberculous mycobacteria, other fungal infections, pyoderma gangrenosum, and bromide use. An uncommon manifestation, seen more often in immunocompromised patients, is the appearance of hundreds of pustular lesions that readily reveal the organism when aspirated.

Another manifestation of disseminated blastomycosis is osteoarticular involvement. Osteomyelitis can be associated with contiguous skin lesions or can appear at sites distant from cutaneous lesions. It is helpful to obtain a bone scan in all patients with disseminated blastomycosis because of the propensity of the organism to infect bone. Genitourinary involvement may be asymptomatic or be associated with signs of prostatism and the presence of a nodule on digital examination. Infrequently occurring findings include laryngeal and oropharyngeal nodules; ocular lesions; central nervous system (CNS) involvement, either meningitis or intracerebral mass lesions; and dissemination to the liver, spleen, and lymph nodes.

DIAGNOSIS

The definitive diagnostic test for blastomycosis is growth of the organism from an aspirate, tissue biopsy specimen, sputum, or body fluid. Urine obtained before and after prostatic massage should be sent for fungal culture in those with disseminated blastomycosis. The mold phase takes several weeks to grow at room temperature. Once growth has occurred, the organism can be rapidly identified as *B. dermatitidis* with a highly specific and sensitive DNA probe. Histopathologic examination of cutaneous or pulmonary lesions, cytologic examination of sputum, bronchoalveolar lavage fluid, or other tissue fluids, and calcofluor fluorescent staining of sputum or purulent material from pustular lesions should be performed to look for the distinctive large, thick-walled yeast with a single broad-based bud. Identification of characteristic organisms allows a tentative diagnosis of blastomycosis and initiation of antifungal therapy before culture results are known. An enzyme immunoassay for *B. dermatitidis* cell wall antigens can be performed on urine and serum and is a useful rapid diagnostic test in patients who have severe pulmonary or disseminated blastomycosis.[6] Because *B. dermatitidis* and *H. capsulatum* share many cell wall antigens, this assay is often positive in patients with histoplasmosis, as well as in those with blastomycosis. Antibody tests for blastomycosis are neither sensitive nor specific. Polymerase chain reaction on tissue samples has proved useful if histopathology and culture are not diagnostic.[7]

TREATMENT Rx

Guidelines for the treatment of blastomycosis have been published by the Infectious Diseases Society of America (IDSA),[8] and the American Thoracic Society.[9] With the exception of patients who have acute pulmonary

blastomycosis that has totally resolved before the diagnosis is established, all patients with blastomycosis should be treated with an antifungal agent. Patients who have mild-to-moderate pulmonary or disseminated blastomycosis should be treated with itraconazole, 200 mg once or twice daily. The length of treatment is 6 to 12 months to achieve mycologic cure and prevent relapse. Fluconazole is not as effective as itraconazole. However, if the patient is unable to take itraconazole, fluconazole can be used, but the dosage should be 400 to 800 mg/day for 6 to 12 months. Voriconazole is increasingly reported to be effective in patients unable to tolerate itraconazole. Successful use of posaconazole has been reported in only a few patients. The echinocandins are not active against *B. dermatitidis* and should not be used.

Patients who have severe pulmonary or disseminated blastomycosis, all patients with CNS infection,[10] and most immunosuppressed patients should be treated initially with a lipid formulation of amphotericin B. The dosage is 3 to 5 mg/kg daily, except for CNS infection, for which 5 mg/kg daily should be used. After clinical improvement has occurred, usually within several weeks, therapy can be changed to itraconazole, 200 mg twice daily for a total of at least 12 months of therapy. For all patients who are treated with itraconazole, serum levels should be determined when steady state has been reached after 2 weeks of therapy to ensure adequate absorption. Serum concentrations should be greater than 1 µg/mL. Corticosteroids have been helpful as adjunctive therapy for patients with ARDS associated with blastomycosis, but this practice remains controversial.

PROGNOSIS

The prognosis for patients with pulmonary or disseminated blastomycosis treated with itraconazole is excellent; more than 90% are cured.[11] If relapse does occur, a second course of itraconazole is usually successful. Most reported deaths occur in patients with overwhelming pneumonia and ARDS.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Smith JA, Kauffman CA. Blastomycosis. *Proc Am Thorac Soc*. 2010;7:173-180.
2. Carlos WG, Rose AS, Wheat LJ, et al. Blastomycosis in Indiana: digging up more cases. *Chest*. 2010;138:1377-1382.
3. Roy M, Benedict K, Deak E, et al. A large community outbreak of blastomycosis in Wisconsin with geographic and ethnic clustering. *Clin Infect Dis*. 2013;57:655-662.
4. Grim SA, Proia L, Miller R, et al. A multicenter study of histoplasmosis and blastomycosis after solid organ transplantation. *Transpl Infect Dis*. 2012;14:17-23.
5. Bradsher RW Jr. The endemic mimic: blastomycosis an illness often misdiagnosed. *Trans Am Clin Climatol Assoc*. 2014;125:188-202.
6. Bariola JR, Hage CA, Durkin M, et al. Detection of *Blastomyces dermatitidis* antigen in patients with newly diagnosed blastomycosis. *Diagn Microbiol Infect Dis*. 2011;69:187-191.
7. Sidamonidze K, Peck MK, Perez M, et al. Real-time PCR assay for identification of *Blastomyces dermatitidis* in culture and in tissue. *J Clin Microbiol*. 2013;50:1783-1786.
8. Chapman SW, Dismukes WE, Proia LA, et al. Clinical practice guidelines for the management of blastomycosis: 2008 update by the Infectious Diseases Society of America. *Clin Infect Dis*. 2008; 46:1801-1812.
9. Limper AH, Knox KS, Sarosi GA, et al. An official American Thoracic Society statement: Treatment of fungal infections in adult pulmonary and critical care patients. *Am J Respir Crit Care Med*. 2011;183:96-128.
10. Bariola JR, Perry P, Pappas PG, et al. Blastomycosis of the central nervous system: a multicenter review of diagnosis and treatment in the modern era. *Clin Infect Dis*. 2010;50:797-804.
11. Khuu D, Shafir S, Bristow B, et al. Blastomycosis mortality rates, United States, 1990-2010. *Emerg Infect Dis*. 2014;20:1789-1794.

REVIEW QUESTIONS

1. A 64-year-old previously healthy man from upper Michigan developed two skin lesions on his face and one on his thigh in late summer. He remembered having several days of feeling feverish with myalgias and a dry cough a few weeks before the first skin lesions appeared. He had been clearing brush around his home in early summer. On examination, he had three heaped-up, nontender lesions, two of which had punctate areas draining a small amount of purulent material on his face and left thigh. Chest radiograph revealed a left upper lobe infiltrate. Biopsy of one of the skin lesions revealed a pyogranulomatous tissue response and special stains showed thick-walled yeasts with a single broad-based budding daughter cell. The *most* likely etiology of this patient's skin lesions and pulmonary infiltrate is:

A. Coccidioidomycosis.
B. Histoplasmosis.
C. Blastomycosis.
D. Sporotrichosis.
E. Disseminated cryptococcosis.

Answer: C Blastomycosis occurs throughout the Mississippi River Valley, in the north central states around the Great Lakes, and along the St. Lawrence Seaway. Blastomycosis begins as a pulmonary infection, but this may cause few symptoms and go undetected until a skin lesion develops. The skin lesions can be ulcers, papules, nodules, or plaques and often evolve to become verrucous with punctate draining areas in the center. Biopsy of a skin lesion shows an exuberant epidermal reaction known as *pseudoepitheliomatous hyperplasia.* Special stains should be performed to look for the distinctive large round, thick-walled yeasts that have a single broad-based budding daughter cell attached, allowing a tentative diagnosis of blastomycosis. Coccidioidomycosis occurs in the desert Southwest and not the upper Midwest. The skin lesions of histoplasmosis are generally smaller papules and usually appear in patients with systemic symptoms of disseminated infection. The skin lesions of sporotrichosis are rarely disseminated; they usually occur in the distribution of the lymphatic drainage from the initial inoculation site. Skin lesions are a prominent clue to disseminated cryptococcal infection in individuals with AIDS and other immunocompromised conditions.

2. The treatment of choice for mild-to-moderate disseminated blastomycosis is:

A. Fluconazole.
B. Itraconazole.
C. Micafungin.
D. Voriconazole.
E. Amphotericin B.

Answer: B Itraconazole is the drug of choice for the treatment of most of the endemic mycoses, and the most experience has accrued with this agent. If itraconazole is not tolerated, the second choice for blastomycosis should be voriconazole, which is increasingly used for the treatment. Fluconazole is not as active, and high doses must be given if the patient is treated with this agent. Micafungin and the other echinocandins are not active against the endemic mycoses. Amphotericin B is reserved as initial therapy for severe blastomycosis.

3. A patient presents with a single skin lesion on her left cheek that yields *B. dermatitidis.* Chest radiograph reveals no pulmonary infiltrate. The most likely mechanism of infection with this organism was:

A. Inoculation when a twig scratched her cheek.
B. Ingestion of mold-contaminated unwashed vegetables.
C. Inoculation of conidia when she rubbed her face while gardening.
D. Inhalation of conidia of the mold phase from the soil.

Answer: D Blastomycosis results when conidia are inhaled into the lungs from the soil or decaying vegetation where the organism is growing. The organism is ingested by macrophages and converts into the yeast phase. Spread occurs by the hematogenous route, and the skin is the most common site of dissemination. The pulmonary infection may be healed by the time the patient presents with skin lesions. Even one skin lesion is a manifestation of hematogenous spread unless a clear-cut inoculation event has been documented. The organism is never acquired by ingestion.

4. Which is the most definitive diagnostic test for blastomycosis?

A. Culture of a skin lesion
B. PCR for *B. dermatitidis* in serum
C. Antigen detection in a sputum sample
D. Complement fixation antibody tests

Answer: A Culture is always definitive for the endemic mycoses, which are environmental organisms and never colonizers of humans. PCR can be useful if performed on tissue samples, but is not useful for serum. A *B.dermatitidis* antigen assay is available for urine and serum, but cannot be performed on sputum samples. Antibody assays are not specific or sensitive for infection with *B. dermatitidis.*

5. A patient presents with acute respiratory distress syndrome (ARDS) several weeks after returning from a hunting trip in northern Wisconsin. Bronchoalveolar lavage reveals large yeasts with single broad-based buds seen on calcofluor white stain. The most appropriate treatment would be:

A. Itraconazole, 200 mg PO twice daily
B. Lipid formulation amphotericin B, 5 mg/kg/day IV
C. Posaconazole, 400 mg PO twice daily
D. Caspofungin, 50 mg/day IV
E. Voriconazole, 200 mg PO twice daily

Answer: B Persons seriously ill with blastomycosis, which is the most likely diagnosis given the organism seen on the calcofluor stain of the BAL, require treatment with amphotericin B. Most physicians now use a lipid formulation to decrease the risks for nephrotoxicity. Oral azole agents should not be relied on for therapy in a seriously ill patient. Caspofungin and the other echinocandins are not active against *B. dermatitidis* or other endemic mycoses.

335

PARACOCCIDIOIDOMYCOSIS

CAROL A. KAUFFMAN

DEFINITION

Paracoccidioidomycosis (South American blastomycosis) is a subacute to chronic mycosis that is endemic in Central and South America. The disease is characterized primarily by pulmonary, mucous membrane, and cutaneous lesions, but disseminated disease also occurs.[1]

The Pathogen

Paracoccidioides brasiliensis is a thermally dimorphic fungus. In the environment and at temperatures below 35° C, the organism is a mold that produces conidia, the infectious form. In tissues and at 37° C in vitro, the organism assumes the yeast form with multiple narrow-based daughter cells attached to the mother cell.

EPIDEMIOLOGY

P. brasiliensis exists only in humid areas of Central and South America. More than 80% of cases are from Brazil. The presumed ecologic niche is in soil, but the exact conditions that favor growth of the organism have not been elucidated. The disease is most prevalent in middle-aged to elderly men from rural areas. The reason for the sexual imbalance (male-to-female ratio of 13 : 1 in many reports, but as high as 70 : 1 in one report from Colombia) is possibly related to the inhibitory effects of estrogens on transition from the mold to the yeast phase of the organism,[2] rather than solely environmental exposure. Although the disease classically develops later in life, it is likely that initial exposure occurs many years earlier. Cases seen in areas outside Central and South America have all been linked to previous residence in the endemic area.

PATHOBIOLOGY

Paracoccidioidomycosis develops after the inhalation of aerosolized conidia encountered in the environment. Once in the alveoli, the mycelial phase converts to the yeast phase. The infection may remain localized to the lungs, although it is likely that asymptomatic hematogenous dissemination occurs during most infections. In most patients, manifestations of disease do not develop at the time of the initial infection. The primary host defense mechanism against *P. brasiliensis* appears to be cell-mediated immunity, but neutrophils also play a role in host defense. The histopathologic picture includes both neutrophilic and granulomatous responses. There have been increasing reports of paracoccidioidomycosis in patients infected with human immunodeficiency virus (HIV)[3] and in other immunosuppressed patients; in these patients there is widespread dissemination, and histopathologic examination shows poorly formed granulomas. Reactivation of latent infection acquired years earlier is the presumed pathogenesis of most cases of the chronic adult form of paracoccidioidomycosis and cases that appear years after the patient has left the endemic area.

CLINICAL MANIFESTATIONS

Acute-Subacute (Juvenile) Paracoccidioidomycosis

The acute-subacute form of paracoccidioidomycosis occurs in less than 10% of patients. It is a disease of the reticuloendothelial system with widespread dissemination to the liver, spleen, lymph nodes, and bone marrow. Patients younger than 30 years typically have this form of paracoccidioidomycosis; however, older adults, especially those who are immunosuppressed, also can manifest this type of rapidly progressive disease. In patients with HIV infection, rapid progression occurs with multiple cutaneous lesions, lymphadenopathy, hepatosplenomegaly, and severe pulmonary involvement with hypoxemia.

Chronic (Adult) Paracoccidioidomycosis

Chronic paracoccidioidomycosis is slowly progressive over many years and is the form seen in more than 90% of patients. Most patients with this type of paracoccidioidomycosis are older men. Pulmonary involvement is prominent and clinically mimics tuberculosis and other chronic fungal pneumonias.[4] Radiographically, nodular, interstitial, or cavitary lesions are seen but differ from those of tuberculosis and histoplasmosis in that the infiltrates tend to be in the middle and lower lung fields rather than the apices. Many patients with the adult form of paracoccidioidomycosis also have ulcerative or nodular mucous membrane lesions, primarily in the anterior nares, the oral cavity, and the larynx; these are slowly destructive and can lead to dysphonia and stenosis of the airway. Cutaneous lesions, particularly on the face, are also common and may be papular, nodular, ulcerative, or plaquelike. The mucocutaneous lesions must be differentiated from mucocutaneous leishmaniasis and squamous cell carcinoma. Adrenal involvement has been reported in more than 90% of cases at autopsy, but adrenal insufficiency is only noted in about half.

DIAGNOSIS

Definitive diagnosis of paracoccidioidomycosis is established by growth of *P. brasiliensis* in culture. The organism may take as long as 4 weeks to grow. For seriously ill patients, direct examination of body fluids, sputum, or purulent material treated with potassium hydroxide or calcofluor fluorescent stain or histopathologic examination of tissue biopsy samples can provide a presumptive diagnosis while awaiting culture results. The characteristic appearance of *P. brasiliensis* consists of thick-walled yeast cells that have multiple small, circumferentially attached, narrow-based budding daughter yeast cells—a distinctive morphologic picture likened to a ship's steering wheel.

A variety of immunodiffusion assays, enzyme immunoassays, and complement fixation assays are available in endemic areas, but sensitivity and specificity vary greatly. The immunodiffusion assay appears to be most useful.[5] An assay for circulating cell wall antigens of *P. brasiliensis* has also been developed in endemic areas, but its role has not been established.

TREATMENT

Rx

The drug of choice for the treatment of paracoccidioidomycosis is itraconazole (200 mg/day for 6 to 12 months). Ketoconazole at a dosage of 200 to 400 mg daily for 1 year is effective and certainly less expensive than itraconazole, but the incidence of side effects is greater and relapses occur more

frequently than with itraconazole. Fluconazole is less effective and should not be used unless no other agent is available. Voriconazole has been shown to be as effective as itraconazole in a randomized open-label pilot study, and there are scattered reports of the successful use of posaconazole as salvage therapy. Sulfonamides have been used for years to treat paracoccidioidomycosis and are clearly the most inexpensive form of treatment. However, relapse rates are higher than with the azoles. Amphotericin B is effective but rarely required, except in immunosuppressed patients with life-threatening disseminated disease. Most HIV-infected patients with paracoccidioidomycosis have been treated with amphotericin B as initial therapy, followed by lifelong suppressive therapy with either itraconazole or trimethoprim-sulfamethoxazole. Adjunctive corticosteroids have been used in a few patients with severe infection and appeared to be useful, but this remains controversial.[6]

PROGNOSIS

Patients with paracoccidioidomycosis have an excellent response to antifungal therapy, with overall relapse rates of about 5% with itraconazole therapy.[7] Although HIV-infected patients respond less well. Patients who have extensive pulmonary involvement at the time of diagnosis are at high risk for progressive fibrosis despite antifungal therapy.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bocca AL, Amaral AC, Teixeira MM, et al. Paracoccidioidomycosis: eco-epidemiology, taxonomy and clinical and therapeutic issues. *Future Microbiol.* 2013;8:1177-1191.
2. Shankar J, Restrepo A, Clemons KV, et al. Hormones and resistance of women to paracoccidioidomycosis. *Clin Microbiol Rev*. 2011;24:296-313.
3. Morejon KM, Machado AA, Martinez R. Paracoccidioidomycosis in patients infected with and not infected with human immunodeficiency virus: a case-control study. *Am J Trop Med Hyg*. 2009;80: 359-366.
4. Queiroz-Telles F, Escuissato DL. Pulmonary paracoccidioidomycosis. *Semin Respir Crit Care Med*. 2011;32:764-774.
5. Webber LP, Martins MD, de Oliveira MG, et al. Disseminated paracoccidioidomycosis diagnosis based on oral lesions. *Contemp Cin Dent*. 2014;5:213-216.
6. Benard G, Campos AF, Netto LC, et al. Treatment of severe forms of paracoccidioidomycosis: is there a role for corticosteroids? *Med Mycol*. 2012;50:641-648.
7. Sylvestre TF, Franciscone Silva LR, Cavalcante Rde S, et al. Prevalence and serological diagnosis of relapse in paracoccidioidomycosis patients. *PLoS Negl Trop Dis*. 2014;8:e2834.

REVIEW QUESTIONS

1. To entertain a diagnosis of paracoccidioidomycosis, one should obtain a history of residence in:

A. The Mideast.
B. South America.
C. Southeaast Asia.
D. North America.
E. Africa.

Answer: B The geographic distribution of *Paracoccidioides brasiliensis* is in several countries in South America. It does not occur in any of the other areas listed.

2. The usual hosts in whom chronic paracoccidioidomycosis is seen are:

A. Infants.
B. Teenage girls.
C. Middle-aged men.
D. Older women.
E. Immunocompromised individuals.

Answer: C Chronic paracoccidioidomycosis is seen almost entirely in middle-aged to older men. The male-to-female ratio is usually quoted as 13 : 1, but has been reported to be as high as 70 : 1 in some countries of South America. The infection begins earlier in life but is slowly progressive, and symptoms appear in middle age. The propensity for infection to occur in men has been explained both by greater exposure to the organism and by the protective effect of estrogens on infection in women.

3. The treatment of choice for paracoccidioidomycosis is:

A. Posaconazole.
B. Amphotericin B.
C. Caspofungin.
D. Itraconazole.

Answer: D Itraconazole is the preferred treatment of paracoccidioidomycosis. There is little experience with posaconazole, but it has been shown to be effective. Caspofungin and other ecinocandins have no activity against *P. brasiliensis* or other endemic mycoses. Amphotericin B is reserved for treatment of severe infection, such as occurs in immunocompromised patients.

4. A major complication of chronic paracoccidioidomycosis in spite of antifungal therapy is:

A. Pulmonary fibrosis.
B. Granulomatous mediastinitis.
C. Constrictive pericarditis.
D. Pneumothorax.
E. Retroperitoneal fibrosis.

Answer: A Pulmonary fibrosis can progress in patients who have the chronic form of paracoccidioidomycosis, in spite of treatment. Progression is more likely in those who have more extensive pulmonary infection. Granulomatous mediastinitis occurs with histoplasmosis, but is uncommon with paracoccidioidomycosis. Constrictive pericarditis is not a complication of paracoccidioidomycosis, and pneumothorax is not commonly seen.

5. In patients who have HIV infection, paracoccidioidomycosis can be a life-threatening acute infection and should be treated with:

A. Itraconazole.
B. Voriconazole.
C. Amphotericin B.
D. Trimethoprim-sulfamethoxazole.
E. Combination antifungals.

Answer: C HIV-infected patients develop acute disseminated paracoccidioidomycosis and should be treated with amphotericin B initially. Step-down therapy to itraconazole can be given after they have improved. Trimethoprim-sulfamethoxazole is effective for paracoccidioidomycosis, but should not be used for severe infection as initial therapy. Voriconazole also is effective for paracoccidioidomycosis, but should not be used for severe infection as initial therapy.

336

CRYPTOCOCCOSIS

CAROL A. KAUFFMAN

DEFINITION

Cryptococcosis occurs most often in persons who are immunosuppressed, especially those infected with human immunodeficiency virus (HIV). Meningitis is the most common clinical manifestation, but pulmonary and other organ involvement occur as well.

The Pathogen

Among the approximately 40 species of *Cryptococcus, Cryptococcus neoformans* and, much less often, *Cryptococcus gattii* are the predominant pathogens in humans.[1] In the environment, *Cryptococcus* species exist as yeasts that have minimal capsules and are easily aerosolized and inhaled. In tissues, *C. neoformans* and *C. gattii* are enveloped by a large polysaccharide capsule that is a major virulence factor. *C. neoformans* is found in the soil and grows well in avian droppings that have high nitrogen content. *C. gattii* is more restricted geographically and has been described mostly in Australia and Southeast Asia, where the ecological niche is the eucalyptus tree. However, the epidemiology of *C. gattii* is evolving with its emergence in the past decade in the Pacific Northwest and now in other areas of North America; the ecologic niche has not been definitively established in these areas. Most of this chapter focuses on *C. neoformans,* which is more common and which has been studied extensively.

EPIDEMIOLOGY

Before the widespread availability of antiretroviral therapy (ART), cryptococcosis occurred in 5 to 10% of patients with acquired immunodeficiency syndrome (AIDS), and almost always in those with fewer than 50 CD4 cells/μL. Cryptococcosis is less commonly seen now in Europe and North America but is extremely common in Africa, where it is estimated that the prevalence among HIV-infected patients is as high as 30%.

In the non-AIDS population, cryptococcosis is a frequent opportunistic infection in patients who have received a solid organ transplant,[2] have been treated with corticosteroids, or have diabetes mellitus, renal failure, cirrhosis, or chronic pulmonary disease. For some patients, the only risk factor appears to be older age. In every reported series, separate from those dealing only with HIV infection, approximately 20% of patients have no known underlying illness. *C. gattii* most often causes illness in normal hosts.[3]

PATHOBIOLOGY

The organism is inhaled from the environment and causes pulmonary infection initially. The primary host defense at this stage is complement-dependent macrophage and neutrophil phagocytosis and killing. Natural killer cells also have the ability to kill the organism. Ultimately, however, T-cell immunity is the most important host determinant in limiting the replication of *C. neoformans*. In most normal hosts, the infection remains localized to the lungs and does not cause symptomatic infection. It is likely that a few organisms exist as walled-off subpleural granulomas in many who have had pulmonary infection. If the host becomes immunosuppressed, the organism can then reactivate and disseminate to other sites. *C. neoformans* is clearly neurotropic, and the primary disease manifestation is meningoencephalitis. However, dissemination to many organs is likely, especially in those with deficient T-cell immunity.

Virulence factors for *C. neoformans* include the capsule, which requires opsonization for efficient phagocytosis, and the production of melanin, which has been shown to occur in vivo and enables the organism to resist intracellular killing. Both of these factors may help explain the virulence of the organism once it has reached the central nervous system (CNS). Antibody and complement levels are low in the brain, and thus phagocytosis of the organism is minimal. Brain tissue provides high concentrations of substrates, such as catecholamines, for the phenol oxidase enzyme systems of *C. neoformans* that produce melanin, thereby aiding survival of the organism.

CLINICAL MANIFESTATIONS

Central Nervous System Infection

The most common manifestation of cryptococcosis is CNS infection.[4] The typical picture is subacute to chronic meningoencephalitis. Patients usually have increasingly severe headaches over a period of several weeks. Other symptoms and signs include nuchal rigidity, lethargy, personality changes, confusion, visual abnormalities (photophobia, diplopia, decreased visual acuity, papilledema, extraocular nerve palsies), and nausea and vomiting. Less commonly, hearing loss, ataxia, and seizures occur. Fever is present in only approximately half the patients. Elderly persons with cryptococcal meningitis may have just dementia, without other neurologic findings. AIDS patients often have subtle CNS symptoms but usually have fever and other constitutional symptoms and rapidly manifest signs of dissemination.

Pulmonary Infection

In non-HIV-infected patients, the most common underlying risk factor for pulmonary cryptococcal infection is chronic obstructive pulmonary disease, followed by corticosteroid use and receipt of a solid organ transplant. *C. neoformans* may merely be an airway colonizer in some patients; in others, symptomatic infection, manifested by fever, cough, and dyspnea, requires treatment with an antifungal agent. The typical lesion noted with pulmonary cryptococcosis is a pleural-based nodular lesion. However, patchy pneumonitis, multiple nodular lesions, cavitary lesions, masslike lesions, and diffuse pulmonary infiltrates have all been noted with pulmonary cryptococcosis. Patients with advanced HIV infection are likely to have diffuse infiltrates that can progress rapidly to acute respiratory insufficiency. All immunosuppressed patients who have pulmonary cryptococcosis and all patients with any CNS symptoms should undergo lumbar puncture to be certain that meningitis is not present. Whether normal hosts with isolated pulmonary cryptococcosis and negative serum antigen tests require lumbar puncture remains controversial.

Involvement of Other Organs

C. neoformans has been reported to infect most organs during the course of disseminated infection, especially in AIDS patients. Skin lesions are a prominent clue to dissemination. Papules that resemble molluscum contagiosum or an acneiform rash, nodules, ulcers, plaques, draining sinuses, and cellulitis have all been reported. Focal involvement can occur in the prostate and other organs of the genitourinary tract, in osteoarticular structures, in the breast, and in the eye, larynx, and other head and neck structures. The prostate, in particular, has been noted as a sanctuary from which persisting organisms can later disseminate.

DIAGNOSIS

The diagnosis of cryptococcosis is established when the yeast is grown in culture. Appropriate specimens for culture include cerebrospinal fluid (CSF),

blood, sputum, material from skin lesions, and other body fluids or tissues that appear to be infected. The organism grows in several days on most standard agar media. Most automated blood culture systems allow rapid growth of *C. neoformans*. Visualization of the capsule and performance of a few simple tests differentiate *C. neoformans* from other yeasts. Tissue biopsy shows the 5- to 10-µm yeast surrounded by the capsule. Definitive diagnosis of cryptococcosis can be made by mucicarmine staining, which selectively stains the polysaccharide capsule a deep rose color. In CSF or other body fluids, an India ink preparation allows visualization of the budding yeast cells surrounded by the large capsule, but this test is not currently done by many laboratories.

The latex agglutination assay for cryptococcal polysaccharide antigen (CRAG) is a highly sensitive and specific diagnostic test.[5] CRAG is positive in CSF in almost 100% and in serum in about 75% of patients who have meningitis. In AIDS patients, serum CRAG is almost always positive and is an excellent screening tool, and in these patients, titers in both CSF and serum are exceptionally high because of the enormous burden of organisms. In non-AIDS patients who have pulmonary cryptococcosis, the CRAG assay is positive in only 25 to 50% of cases. False-positive results with the CRAG assay are uncommon but have been reported in patients with *Trichosporon asahii* infections because of cross-reacting antigens shared by both fungi.

A newer technique, lateral flow analysis (LFA) to detect cryptococcal polysaccharide antigen has been developed as a dipstick assay, similar to that of pregnancy tests, that can be performed in serum or urine at the point of care by clinicians caring for the patient. This technique appears to be as sensitive and specific as the classic CRAG test.

The CSF of patients with cryptococcal meningitis typically has an increased number of white blood cells (but rarely >500/µL), a predominance of lymphocytes (although neutrophils are sometimes prominent early in the course), elevated protein, and decreased glucose. AIDS patients most often have normal or only mildly abnormal findings as a result of their markedly defective immune response. Despite normal CSF findings with regard to cells, protein, and glucose, every AIDS patient with a headache must have a CRAG or LFA test and culture performed on CSF. It is extremely important that an opening pressure be obtained when lumbar puncture is performed. Especially in AIDS patients, extremely high intracranial pressure (>350 mm H_2O) has been associated with poor outcomes and must be aggressively lowered.

All patients with cryptococcal meningitis should undergo computed tomography or magnetic resonance imaging of the brain to look for mass lesions and to assess ventricular size. Obstructive hydrocephalus is uncommon but requires a shunting procedure to decrease the pressure. More commonly, the increased intracranial pressure with cryptococcal infection is associated with normal-sized ventricles and is due to blockage at the arachnoid villi or increased brain edema (or both), perhaps related to the osmotic effect of the polysaccharide capsule. Different methods for reducing pressure are used in this situation.

TREATMENT

Rx

Guidelines for the treatment of cryptococcal infection have been published by the Infectious Diseases Society of America (IDSA).

Central Nervous System Infection

Early multicenter randomized trials in non-AIDS patients showed superiority of the combination of amphotericin B and flucytosine for 6 weeks over amphotericin B alone for 10 weeks. Subsequent randomized trials in the azole era have been performed only in the AIDS population. They have confirmed the benefit of flucytosine added to amphotericin B for induction therapy and have shown that initial therapy with fluconazole alone or with amphotericin B alone is not as effective as therapy with amphotericin B and flucytosine. The combination of amphotericin B and flucytosine has been shown to be the most rapidly fungicidal regimen, and increasing number of reports have documented that rapid fungicidal activity that clears the organism from the CSF is associated with improved outcomes. Regimens using amphotericin B with fluconazole or flucytosine with fluconazole are less effective but reasonable alternatives when the preferred regimen of amphotericin B plus flucytosine is not available.

Current recommendations for AIDS patients are to give induction therapy with intravenous amphotericin B deoxycholate, 0.7 to 1 mg/kg daily, combined with oral flucytosine, 100 mg/kg daily given in four divided doses for at least 2 weeks, followed by consolidation therapy with oral fluconazole, 400 mg daily for a minimum of 8 weeks, and then suppressive therapy with fluconazole, 200 mg daily. Lipid formulations of amphotericin B at daily dosages of 3 to 5 mg/kg daily are increasingly used because they are less nephrotoxic; however, they are often not available in developing countries. Induction therapy with amphotericin B (1 mg/kg per day for 4 weeks) plus flucytosine (100 mg/kg per day for 2 weeks) is associated with improved survival among HIV-positive patients with cryptococcal meningitis compared with either amphotericin B alone or amphotericin B plus fluconazole (400 mg twice daily for 2 weeks).

For patients who have undergone 12 months of antifungal therapy, who have $CD4^+$ counts higher than 100/µL, and whose HIV viral load is undetectable on antiretroviral therapy, the suppressive therapy can be stopped. Suppressive therapy with fluconazole for transplant recipients is recommended for 6 to 12 months. The IDSA guideline recommendations[6] for non-HIV-infected, non–transplant recipients are to treat with amphotericin B deoxycholate, 0.7 to 1.0 mg/kg daily, plus flucytosine, 100 mg/kg daily, in four divided doses for at least 4 weeks for induction therapy, followed by consolidation therapy with fluconazole, 400 mg daily for 8 weeks, and suppressive therapy with fluconazole, 200 mg daily for 6 to 12 months. Again, many physicians use lipid formulations of amphotericin B, 3 to 5 mg/kg daily, in this population, many of whom are older and have underlying medical illnesses.

Only one treatment trial used voriconazole in combination with amphotericin B, and there are just a few case reports on the use of voriconazole and posaconazole for salvage treatment of cryptococcal meningitis. These are reasonable alternatives if no other azoles can be used. The echinocandins are not active against *C. neoformans* and should not be used.

A significant observation from the AIDS treatment trials was the role of increased intracranial pressure as a cause of early death from cryptococcal meningitis. An aggressive approach to the diagnosis and treatment of increased intracranial pressure in both AIDS and non-AIDS patients is mandatory and should include daily lumbar puncture or placement of a temporary lumbar drain or ventriculostomy until the opening pressure remains lower than 190 mm H_2O. Treatment with corticosteroids, acetazolamide, or mannitol has not proved efficacious in this setting.

Another issue that has emerged is the development of immune reconstitution inflammatory syndrome (IRIS), which can occur in patients with AIDS who are receiving effective antiretroviral therapy (ART) that increases the CD4 count (Chapter 395). Symptoms of meningitis reappear and are due to the inflammatory response and not to a relapse of disease. In a recently reported trial, HIV-positive patients with cryptococcal meningitis who had not previously received ART were randomized to initiate either earlier ART (1 to 2 weeks after meningitis diagnosis) or deferred ART (5 weeks after diagnosis), with all patients treated for cryptococcal meningitis with amphotericin B (0.7 to 1.0 mg/kg/day) plus fluconazole (800 mg per day) for 14 days, followed by fluconazole consolidation therapy. Although the incidence of recognized cryptococcal IRIS did not differ significantly between the earlier and deferred ART groups, deferring ART for 5 weeks after diagnosis was associated with significantly improved survival, especially among patients with a paucity of CSF white cells. IRIS can also occur in transplant recipients in whom immunosuppressive therapy is decreased rapidly. Generally, no specific therapy is needed for mild IRIS, but sometimes corticosteroids are needed if increased intracranial pressure occurs.

Pulmonary and Other Nonmeningeal Infections

Treatment of nonmeningeal cryptococcosis depends on the severity of the infection. Many patients with isolated pulmonary or other focal infections are not severely ill, and oral fluconazole, 400 mg daily for 6 to 12 months, is recommended. For patients who are severely ill, therapy is the same as noted earlier for CNS infection.

PROGNOSIS

The outcome for both AIDS and non-AIDS patients with cryptococcal meningitis has improved markedly in the developed world. However, in Africa the mortality from cryptococcal meningitis in AIDS patients approaches 100% in some areas because of lack of access to specific therapy. Dementia, which usually occurs in older patients, hearing loss, and visual loss may not be reversed even though mycologic cure is achieved. Fluconazole, 200 mg three times per week, is safe and effective as primary prophylaxis against cryptococcal disease in cryptococcal antigen-negative, HIV-infected adults with CD4 counts lower than 200 cells/µL, both before and during early antiretroviral treatment.

Grade A References

A1. Nussbaum JC, Jackson A, Namarika D, et al. Combination flucytosine and high-dose fluconazole compared with fluconazole monotherapy for the treatment of cryptococcal meningitis: a randomized trial in Malawi. *Clin Infect Dis*. 2010;50:338-344.

A2. Day JN, Chau TT, Wolbers M, et al. Combination antifungal theory for cryptococcal meningitis. *N Engl J Med*. 2013;368:1291-1302.
A3. Boulware DR, Meya DB, Muzoora C, et al. Timing of antiretroviral therapy after diagnosis of cryptococcal meningitis. *N Engl J Med*. 2014;370:2487-2498.
A4. Parkes-Ratanshi R, Wakeham K, Levin J, et al. Primary prophylaxis of cryptococcal disease with fluconazole in HIV-positive Ugandan adults: a double-blind randomised, placebo-controlled trial. *Lancet Infect Dis*. 2011;11:933-941.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Kwon-Chung KJ, Fraser JA, Doering TL, et al. Cryptococcus neoformans and Cryptococcus gattii, the etiologic agents of cryptococcosis. *Cold Spring Harb Perspect Med*. 2014;4:a019760.
2. Baddley JW, Schain DC, Gupte AA, et al. Transmission of *Cryptococcus neoformans* by organ transplantation. *Clin Infect Dis*. 2011;52:e94-e98.
3. Harris JR, Lockhart SR, Debess E, et al. *Cryptococcus gattii* in the United States: clinical aspects of infection with an emerging pathogen. *Clin Infect Dis*. 2011;53:1188-1195.
4. Sloan DJ, Parris V. Cryptococcal meningitis: epidemiology and therapeutic options. *Clin Epidemiol*. 2014;6:169-182.
5. Makadzange AT, McHugh G. New approaches to the diagnosis and treatment of cryptococcal meningitis. *Semin Neurol*. 2014;34:47-60.
6. Perfect JR, Dismukes WE, Dromer F, et al. Clinical practice guidelines for the management of cryptococcal disease: 2009 update by the Infectious Disease Society of America. *Clin Infect Dis*. 2010; 50:291-322.

REVIEW QUESTIONS

1. A patient with cryptococcal meningitis had improvement of headache and nausea following the initial diagnostic lumbar puncture. Amphotericin B and flucytosine were begun immediately after the tap, but 2 days later, he has recurrence of these symptoms and also has some blurring of vision. What is the likely reason for the return of his symptoms?

A. Reaction to amphotericin B
B. Secondary bacterial infection related to the lumbar puncture
C. Increased intracranial pressure
D. Resistance of the organism to amphotericin B
E. Resistance of the organism to flucytosine

Answer: C This is a typical picture in that the lumbar puncture takes off cerebrospinal fluid and relieves some of the pressure briefly. Increased intracranial pressure is most likely due to blockage of the arachnoid villi and/or increased brain edema and is thought to be secondary to the large capsule surrounding the yeast. Resistance to amphotericin B and flucytosine is very uncommon. The symptoms, especially blurring of vision, are unlikely due to amphotericin B toxicity. Secondary bacterial infection after a spinal tap is also rare.

2. What is the treatment for the condition noted in question 1?

A. Perform repeated lumbar punctures to bring down the increased pressure
B. Change therapy to fluconazole
C. Continue amphotericin B but stop flucytosine
D. Add ceftriaxone 2 g every 12 hours
E. Symptomatic treatment only

Answer: A Treatment of increased intracranial pressure, which is seen commonly with cryptococcal meningitis, is to perform repeated spinal taps until the pressure remains low. There is no reason to change antifungal therapy or add an antibiotic.

3. Which of the following is the preferred treatment for an AIDS patient with cryptococcal meningitis?

A. Fluconazole and flucytosine
B. Voriconazole and amphotericin B
C. Posaconazole and caspofungin
D. Amphotericin B and flucytosine
E. Fluconazole

Answer: D Several randomized blind and open-label treatment trials have shown the benefit of combined therapy with amphotericin B and flucytosine for cryptococcal meningitis. Echinocandins have no role in treating this infection. The other suggested regimens are inferior to amphotericin B and flucytosine.

4. What is the major host defense against *Cryptococcus* species?

A. Immunoglobulin A antibody
B. Natural killer cells
C. T lymphocytes
D. Complement
E. Neutrophils

Answer: C Patients who have deficient T-cell immunity, such as AIDS patients with low CD4 cell counts, transplant recipients, and patients on corticosteroids, are at high risk for developing infection with *Cryptococcus.*

5. The diagnosis of cryptococcal meningitis can be made in a few hours by performing which of the following?

A. Culture of cerebrospinal fluid
B. Magnetic resonance imaging
C. Lateral flow dipstick test for cryptococcal antigen
D. Cryptococcal immunoglobulin M antibody assay
E. Gram stain of cerebrospinal fluid

Answer: C Culture is definitive but takes several days. Antibody assays for *Cryptococcus neoformans* are not helpful for diagnosis. Magnetic resonance imaging may show brain abscesses and meningeal enhancement, but these are not specific for cryptococcosis. The dipstick method for detecting antigen in cerebrospinal fluid is a rapid, sensitive, and specific test to make the diagnosis of cryptococcal meningitis. An India ink preparation (but not Gram stain) allows visualization of the budding yeast cells surrounded by the large capsule, but this test is not currently done by many laboratories.

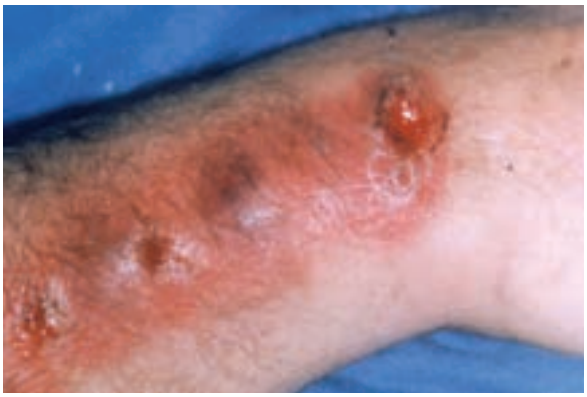

FIGURE 337-1. Lymphocutaneous sporotrichosis. The lesion at the inoculation site has ulcerated. (From Watanakunakorn C.: Photoquiz. *Clin Infect Dis.* 1996;22:765.)

337

SPOROTRICHOSIS

CAROL A. KAUFFMAN

DEFINITION

Sporotrichosis is a subacute or chronic infection that is usually localized to cutaneous and lymphocutaneous structures, but pulmonary, osteoarticular, and disseminated infection can occur in patients who have certain underlying diseases.

The Pathogen

Sporothrix schenckii is a thermally dimorphic fungus. In the environment at temperatures lower than 35° to 37° C, the organism is a mold and produces conidia, the infectious form. In tissues and at 35° to 37° C, *S. schenckii* transforms into the yeast phase; the yeasts are 4 to 6 μm in diameter; are cigar shaped, round, or oval; and reproduce by budding.

EPIDEMIOLOGY

S. schenckii is found worldwide in climates ranging from temperate to tropical.[1] The organism exists in a variety of environmental niches, including soil, sphagnum moss, hay, decaying wood, and other vegetation. Infection is seen almost entirely in persons whose vocation, avocation, or living condition brings them into contact with the organism in the environment. Landscaping activities, gardening, farming, and motor vehicle collisions have been associated with sporotrichosis. Inhalation of *S. schenckii* conidia occurs less commonly and results in pulmonary and, rarely, disseminated sporotrichosis. Most cases of sporotrichosis are sporadic, but outbreaks have been described. An extensive outbreak extending over many years in Rio de Janeiro and occurring mostly in children and women has been traced to infected domestic cats.

PATHOBIOLOGY

Infection is almost always acquired by inoculation of conidia and remains localized to the immediate and contiguous cutaneous, subcutaneous, and lymphatic structures. Some strains of *S. schenckii* grow poorly at temperatures higher than 35° C; these strains usually cause fixed cutaneous lesions without lymphatic spread. The typical host response to infection with *S. schenckii* is a mixed neutrophilic and granulomatous reaction. Antibody is not protective; cell-mediated immunity is important in containing the infection. In individuals who have underlying illnesses, including alcoholism, diabetes mellitus, and chronic obstructive pulmonary disease, *S. schenckii* is more likely to involve osteoarticular structures and the lungs. Widespread dissemination develops in persons infected with human immunodeficiency virus (HIV) but is a distinctly unusual event in normal hosts.[2]

CLINICAL MANIFESTATIONS

Lymphocutaneous

Days to weeks after inoculation of *S. schenckii* conidia, a papular lesion develops at the inoculation site; the lesion becomes nodular and often ulcerates (Fig. 337-1). Drainage is not grossly purulent, and the lesion is not terribly painful. Similar lesions occur along the lymphatic distribution proximal to the primary lesion. Verrucous or ulcerative fixed cutaneous lesions do not exhibit lymphatic extension. The differential diagnosis of lymphocutaneous sporotrichosis includes *Nocardia* infections (Chapter 330), particularly *Nocardia brasiliensis*; atypical mycobacterial infections (Chapter 325), especially *Mycobacterium marinum*; *Leishmania brasiliensis* infections (Chapter 348); and tularemia (Chapter 311).

Visceral and Osteoarticular

Pulmonary sporotrichosis[3] occurs most often in middle-aged men who have chronic pulmonary disease and abuse alcohol. Fever, night sweats, weight loss, fatigue, dyspnea, cough, purulent sputum, and hemoptysis are common. Chest radiographs show unilateral or bilateral upper lobe cavities with variable amounts of fibrosis and nodular lesions. The disease mimics reactivation tuberculosis in many aspects. Osteoarticular sporotrichosis is found most often in middle-aged men and occurs more frequently in alcoholics. Infection may involve one or multiple joints; the joints most commonly affected are the knee, elbow, wrist, and ankle. Isolated bursitis, tenosynovitis, and nerve entrapment syndromes have been reported. Osteoarticular infection can follow local inoculation, but in most patients this occurs secondary to hematogenous spread. Isolated case reports document sporotrichosis involving the pericardium, eye, perirectal tissues, larynx, breast, epididymis, spleen, liver, bone marrow, lymph nodes, and meninges. Disseminated sporotrichosis, manifested as widespread ulcerative cutaneous lesions with or without visceral involvement, is uncommon; most cases have been reported in patients with advanced HIV infection.

DIAGNOSIS

Growth of *S. schenckii* from culture of material aspirated from a lesion, a tissue biopsy specimen, sputum, or body fluid is the most effective method of establishing the diagnosis of sporotrichosis. Growth of the mold phase of the organism is usually evident within a few days. Histopathologic examination of biopsy material shows a mixed granulomatous and pyogenic process; however, the organisms are often present in small numbers and are frequently not visualized. Serology is not useful in the diagnosis of sporotrichosis. Polymerase chain reaction testing has been used to confirm the diagnosis[4] but must be obtained at a fungal reference laboratory.

TREATMENT

Rx

Because sporotrichosis is usually a localized subacute to chronic infection, oral antifungal agents are preferred; amphotericin B is reserved for severe visceral infections. Guidelines for the management of sporotrichosis have been published by the Infectious Diseases Society of America.[5] Itraconazole is the drug of choice for lymphocutaneous sporotrichosis.[6] The usual dosage is 200 mg daily, and treatment should continue for several weeks after all lesions have disappeared, generally for a total of 3 to 6 months. Saturated solution of potassium iodide (SSKI) has been used to treat sporotrichosis for almost a century. The initial dose is 5 to 10 drops three times daily in water or juice, with the dose increasing over a period of several weeks to a maximum of 40 to 50 drops three times daily. SSKI has many side effects, including salivary gland swelling, metallic taste, rash, and fever; the only advantage is that it is inexpensive. Fluconazole is less effective than itraconazole but for occasional patients can be used at a dosage of 400 to 800 mg daily. Voriconazole is not active against *S. schenckii*, and there is minimal experience using

posaconazole. Terbinafine appears to be effective for sporotrichosis at a dosage of 500 mg twice daily. Local hyperthermia, induced by a variety of different warming devices or baths, has been shown to be effective in some patients with fixed cutaneous lesions.

Osteoarticular and pulmonary sporotrichosis are usually treated with itraconazole, 200 mg twice daily for 1 to 2 years. Other azoles are less effective, and SSKI is ineffective. In a seriously ill patient with pulmonary sporotrichosis, a lipid formulation of amphotericin B at a dosage of 3 to 5 mg/kg daily should be used as initial therapy. After the patient has shown improvement, therapy can be changed to itraconazole. A lipid formulation of amphotericin B, at a dosage of 3 to 5 mg/kg daily, is the drug of choice for disseminated sporotrichosis. Therapy can be changed to itraconazole, 200 mg twice daily, once the patient has stabilized. Patients with HIV infection and disseminated sporotrichosis should remain on lifelong maintenance therapy with itraconazole, 200 mg daily.

For all patients who are treated with itraconazole, serum levels should be determined when steady state has been reached after 2 weeks of therapy to ensure adequate absorption. Serum concentrations should be greater than 1 µg/mL.

PROGNOSIS

The prognosis for patients with cutaneous and lymphocutaneous sporotrichosis is excellent. Almost all patients are cured with one course of therapy; relapses occur in only a small proportion of patients. Extracutaneous forms of sporotrichosis do not respond well to therapy, partly because of delays in diagnosis and partly because of the underlying diseases that are frequently found in those who have extracutaneous sporotrichosis. The outcome of disseminated sporotrichosis in patients with HIV infection has improved in recent years with effective antiretroviral therapy.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Chakrabarti A, Bonifaz A, Gutierrez-Galhardo MC, et al. Global epidemiology of sporotrichosis. *Med Mycol*. 2015;53:3-14.
2. Freitas DF, Hoagland BS, Valle AC, et al. Sporotrichosis in HIV-infected patients: report of 21 cases of endemic sporotrichosis in Rio de Janeiro, Brazil. *Med Mycol*. 2012;50:170-178.
3. Aung AK, Teh BM, McGrath C, et al. Pulmonary sporotrichosis: case series and systematic analysis of literature on clinic-radiological patterns and management outcomes. *Med Mycol*. 2013; 51:534-544.
4. Rodriguez-Brito S, Camacho E, Mendoza M, et al. Differential identification of *Sporothrix* spp. and *Leishmania* spp. by conventional PCR and qPCR in multiplex format. *Med Mycol*. 2015;53:22-27.
5. Kauffman CA, Bustamante B, Chapman SW, et al. Clinical practice guidelines for the management of sporotrichosis: 2007 update by the Infectious Diseases Society of America. *Clin Infect Dis*. 2007;45:1255-1265.
6. Barros MB, Schubach AO, Oliveira RV, et al. Treatment of cutaneous sporotrichosis with itraconazole: study of 645 patients. *Clin Infect Dis*. 2011;52:e200-e206.

REVIEW QUESTIONS

1. A 26-year-old man from Michigan had a motorcycle collision; he was thrown off and scraped his left shoulder in the dirt. Two weeks later, he developed a nodular lesion at the site of the injury. This ulcerated, and then several other nodular lesions that became ulcerated appeared proximal to the initial lesion over the next 3 weeks. He had no systemic symptoms or signs of infection. What is the most likely organism causing this infection?

A. *Leishmania brasiliensis*
B. *Mycobacterium marinum*
C. *Nocardia asteroides*
D. *Sporothrix schenckii*

Answer: D This is a typical picture of lymphocutaneous sporotrichosis due to inoculation of *S. schenckii* from soil into the skin and subcutaneous tissues. Days to weeks after inoculation, a papule develops at the site of inoculation. The primary lesion can become nodular, but most often, it ulcerates. The drainage is not grossly purulent and has no odor, and the lesion is not terribly painful. *M. marinum* is usually associated with water exposure. *L. brasiliensis* is not found in North America. *Nocardia* species usually cause infection in immunocompromised hosts.

2. Which of the following is the most appropriate method for establishing the diagnosis of sporotrichosis?

A. Culture of the fungus from a skin lesion
B. Polymerase chain reaction (PCR) on a serum sample
C. Antibody testing
D. Antigen assay in urine
E. Histopathology

Answer: A There is no sensitive or specific antibody or antigen assay for *S. schenckii*. PCR can be performed at reference laboratories on tissue, but not on serum. Histopathology of biopsy material shows a mixed granulomatous and pyogenic process, but the organisms are often present in small numbers and frequently not visualized. Culture of a lesion is the definitive diagnostic procedure for sporotrichosis.

3. Which of the following is the treatment of choice for lymphocutaneous sporotrichosis?

A. Posaconazole
B. Itraconazole
C. Voriconazole
D. Fluconazole
E. Saturated solution of potassium iodide (SSKI)

Answer: B Itraconazole is the treatment of choice for lymphocutaneous sporotrichosis. Fluconazole has some activity, but large doses are needed to be effective. Voriconazole does not have activity against *S. schenckii*, and there are minimal data with posaconazole. SSKI has been used to treat sporotrichosis for almost a century, but it has many side effects, and its only advantage is that it is not expensive.

4. Which animal is most often implicated in zoonotic transmission of *S. schenckii* to humans?

A. Armadillos
B. Dogs
C. Pigs
D. Cats
E. Birds

Answer: D Cats have been associated with an ongoing outbreak since 1998 of sporotrichosis in Rio de Janeiro; this outbreak involves mostly children and women who care for the cats, which are infected and generally have a high burden of organisms. Dogs have only occasionally been involved in transmission of sporotrichosis to humans. Armadillos can transmit *S. schenckii* to humans by inoculating the organisms from dirt on their claws. Pigs and birds have not transmitted *S. schenckii*.

5. Saturated solution of potassium iodide (SSKI) is effective in treating which manifestation of sporotrichosis?

A. Osteoarticular
B. Pulmonary
C. Lymphocutaneous
D. Disseminated
E. Breast

Answer: C SSKI has been used for a century to treat sporotrichosis but is only effective against the cutaneous and lymphocutaneous forms. Pulmonary, osteoarticular, and disseminated infection, including unusual sites like breast, must be treated with either itraconazole or amphotericin B.

338

CANDIDIASIS

CAROL A. KAUFFMAN

DEFINITION

Candidiasis encompasses a wide variety of clinical syndromes caused by yeasts of the genus *Candida.* Of the species that cause infection in humans, *Candida albicans* is the most common; *Candida glabrata, Candida parapsilosis,* and *Candida tropicalis* are responsible for most of the remaining infections. Organisms such as *Candida krusei, Candida lusitaniae,* and *Candida guilliermondii* are less common causes of infection.

The Pathogen

Candida species are 2- to 6-µm yeastlike organisms that reproduce by budding. Most species, with the exception of *C. glabrata,* form pseudohyphae (elongated buds that remain attached to the mother cell) and hyphae in tissues.

Candida species cause a wide spectrum of diseases ranging in severity from localized mucous membrane infection to life-threatening disseminated disease. The major determinant of the severity of infection is the host's immune response. Local infections are often related to overgrowth of *Candida* as a result of changes in the normal microbiota. Invasive infections that remain within an organ system, such as urinary tract infections, usually occur because of local anatomic abnormalities. In an immunosuppressed host, especially a patient with neutropenia, widespread visceral dissemination is common.

EPIDEMIOLOGY

Candida species reside normally in the gastrointestinal and genitourinary tracts and on the skin. As colonizers, *Candida* species do not cause infection unless there is a defect in host defense mechanisms or unless exogenous factors, such as antibiotic use, have upset the ecology of the normal microbiota. *C. albicans* is the species most commonly found colonizing humans; *C. glabrata* is the second most common species, and *C. tropicalis, C. parapsilosis,* and others are found less often. The species of *Candida* colonizing and infecting patients has changed in recent decades in that *C. glabrata,* a species that is increasingly resistant to fluconazole, has become a prominent pathogen in many hospitals.

Though uncommon, acquisition of *Candida* from environmental sources has been noted. The *Candida* species most often associated with transmission from contaminated fluids or devices, especially central intravenous catheters, has been *C. parapsilosis.*

Candidiasis is the most common opportunistic fungal infection as a result of both the organisms' ubiquity and the increasing number of patients with risk factors for infection with these organisms.[1] The classic immunosuppressed host at risk for serious *Candida* infections is a neutropenic patient with a hematologic malignancy who has received cytotoxic agents and corticosteroids. Increasingly, however, candidiasis is seen in patients in intensive care units (ICUs). Risk factors for the development of serious *Candida* infections in ICU patients include the use of broad-spectrum antimicrobials, indwelling central venous catheters, previous surgical procedures, renal failure, parenteral nutrition, and high Acute Physiology and Chronic Health Evaluation (APACHE) score. Certain ICU populations, especially very-low-birthweight neonates and burn victims, are at even higher risk for *Candida* infection than is the typical ICU patient.

The primary manifestation of *Candida* infection in patients with HIV/AIDS is mucocutaneous infection, primarily oropharyngeal candidiasis. The development of mucosal *Candida* infection is related to deficient T-cell immunity as reflected by a low CD4 lymphocyte count. With appropriate antiretroviral therapy, oropharyngeal candidiasis has become an uncommon opportunistic infection.

PATHOBIOLOGY

The usual mode of infection with *Candida* is egress from its normal niche into the bloodstream or other tissues; the source is usually the gastrointestinal tract, but the skin and genitourinary tract are other sources. The primary host defense in response to this event is phagocytosis and killing by neutrophils, monocytes, and macrophages. C-C chemokine receptor 2 (CCR2)-expressing inflammatory monocytes and their tissue-resident derivatives play an essential antifungal role, particularly in the first 48 hours after *Candida* infection.[2] Phagocytosis is enhanced in the presence of specific anti-*Candida* antibody and complement. Several different mechanisms are operative within neutrophils and macrophages that allow the killing of yeasts. Thus, patients who are leukopenic, especially those with chemotherapy-induced disruption of the gut mucosa, are at great risk for invasion with *Candida* species. Once *Candida* gains access to the bloodstream, widespread hematogenous dissemination is the rule. Biopsy of involved organs shows multiple microabscesses composed of neutrophils (in a host who has these cells), budding yeasts, and often pseudohyphae or hyphae. Over time, the lesions show a mixed neutrophilic and granulomatous response.

T-cell immunity is an important host defense against infection with *Candida* at mucosal surfaces. In contrast to those with neutropenia, patients with deficient T-cell immunity are at risk for persistent and recurrent mucocutaneous candidiasis, but invasive infection rarely develops.

CLINICAL MANIFESTATIONS

Mucocutaneous Candidiasis

Oropharyngeal Candidiasis

Local mucous membrane and cutaneous lesions are the most common forms of *Candida* infection. Oropharyngeal candidiasis, or *thrush* (Chapter 425), can be due to either local factors or T-cell dysfunction. Local factors include the use of broad-spectrum antimicrobials or inhaled corticosteroids, xerostomia, and radiation treatment of the head and neck. Denture stomatitis occurs frequently in persons who wear full upper dentures, especially those who do not remove their dentures at night.

Thrush secondary to T-cell dysfunction is most commonly seen in patients with HIV infection (Chapter 390) and is the most frequent opportunistic infection noted in patients with AIDS. The appearance of thrush in a

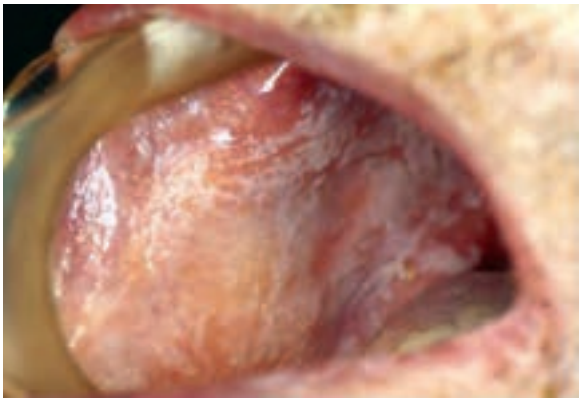

FIGURE 338-1. Thrush.

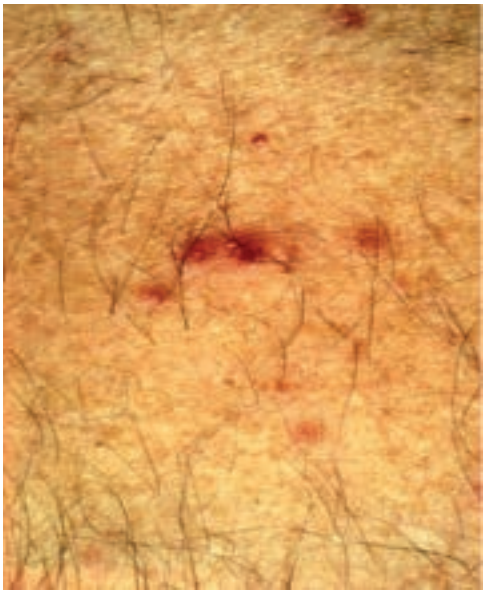

FIGURE 338-2. Skin lesions in invasive candidiasis.

previously healthy individual with no known risk factors should immediately raise suspicion of HIV infection.

Thrush manifests with white plaques on the buccal mucosa, palate, oropharynx, or tongue (Fig. 338-1). Scraping the lesions with a tongue depressor reveals an erythematous, nonulcerated mucosa under the plaques. Denture stomatitis almost always manifests as a painful erythematous palate without plaques. Angular cheilitis, or perlàche, which is the presence of painful cracks at the corners of the mouth, can occur with or without thrush.

Esophagitis

Esophagitis may accompany oropharyngeal candidiasis or may occur independently of lesions in the oropharynx (Chapter 138). The development of *Candida* esophagitis is almost always related to immune dysfunction and not simply to local factors. *Candida* esophagitis occurs in AIDS patients with low CD4 counts, patients with leukemia, and others taking immunosuppressive agents. The classic symptom of *Candida* esophagitis is odynophagia localized to a discrete substernal area. The differential diagnosis includes ulcerations due to herpes simplex or cytomegalovirus and, in AIDS patients, idiopathic ulcers.

Vulvovaginitis

Candida vulvovaginitis is a common infection in women of childbearing age and is the most frequent mucocutaneous manifestation of *Candida* infection.[3] Risk factors include conditions associated with increased estrogen levels, such as the use of oral contraceptives and pregnancy, diabetes mellitus, therapy with corticosteroids or broad-spectrum antimicrobials, and HIV infection. Symptoms include vaginal discomfort, discharge, and vulvar pruritus. The discharge is usually curdlike, but it can also be thin and watery. The labia are erythematous and swollen, and the vaginal walls show erythema and white plaques. Although most women have only a few episodes throughout their lives, a minority have frequent recurrences; in most of these patients, no discrete risk factor is found, and the cause is presumed to be local immune dysregulation.

Cutaneous Candidiasis

Candida infection of the skin (Chapter 441) occurs mostly in the intertriginous areas or under a large pannus or pendulous breasts. The lesions are erythematous, pruritic, and frequently pustular; have a distinct border; and are almost always associated with smaller satellite lesions, which helps distinguish candidiasis from tinea cruris or corporis. *Candida* onychomycosis results in thickened, opaque, and onycholytic nails. *Candida* can also cause paronychia, especially in those whose occupation involves frequent immersion of the hands in water.

Chronic Mucocutaneous Candidiasis

This uncommon syndrome usually begins in childhood and is characterized by recalcitrant and relapsing thrush, vaginitis, onychomycosis, and hyperkeratotic skin lesions on the face, scalp, and hands. Autosomal dominant chronic mucocutaneous candidiasis is associated with mutations in the CC domain of *STAT1* leading to defective T_H1 and T_H17 responses.[4] Some patients have associated autoimmune endocrinopathies, including hypoparathyroidism, hypothyroidism, and hypoadrenalism (autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy [APECED], which is caused by a loss-of-function mutation of the autoimmune regulator gene, *AIRE*, and in these patients autoantibodies against interleukin-17 (IL-17) and IL-22 are found.[5] (See Autoimmune Polyglandular Syndrome Type 1 in Chapter 231.)

Disseminated Infections

Candidemia

The most common manifestation of disseminated *Candida* infection is candidemia. However, candidemia merely implies the presence of *Candida* in blood; it does not define the extent of visceral involvement. *Candida* obtained from a blood culture should never be considered a contaminant and should always prompt a search for the probable source and the extent of infection. Risk factors for candidemia include broad-spectrum antimicrobial therapy, central intravenous catheters, parenteral nutrition, renal failure, surgical procedures involving the gastrointestinal tract, neutropenia, and corticosteroid therapy. The attributable mortality from candidemia approaches 40%; overall mortality is higher in elderly patients and neonates.

Although candidemia is the most obvious manifestation of serious infection with *Candida* species, septic shock can occur, along with invasion of multiple viscera, in the absence of positive blood cultures.[6] The clinical picture of invasive candidiasis is indistinguishable from that of bacterial infection. The characteristic histologic picture consists of multiple microabscesses in many organs. The eyes, kidneys, liver, spleen, and brain are the most commonly involved sites, but virtually all organs can be involved. Clinical clues to the diagnosis of invasive candidiasis include the appearance of skin and retinal lesions. The nonpainful, nonpruritic skin lesions are papular to pustular and surrounded by a zone of erythema (Fig. 338-2). The eye lesions appear as distinctive white exudates in the retina (Fig. 338-3); with extension into the vitreous body, the retina becomes obscured.

Endocarditis

Candida endocarditis is an uncommon and often fatal complication of candidemia. It occurs most often in intravenous drug users, patients who have prosthetic cardiac valves, and those with central venous catheters in place. Blood cultures are usually persistently positive, and echocardiography reveals large vegetations that can readily embolize to major vessels.

Chronic Disseminated (Hepatosplenic) Candidiasis

This syndrome almost always occurs in leukemic patients who have had an episode of neutropenia. After the neutrophil count returns to normal, fevers that are often quite high, right upper quadrant tenderness, and nausea develop. The alkaline phosphatase level is generally elevated, and distinctive punched-out lesions are seen in the liver, spleen, and sometimes the kidneys on computed tomography (Fig. 338-4). Biopsy of these lesions shows microabscesses that contain budding yeasts.

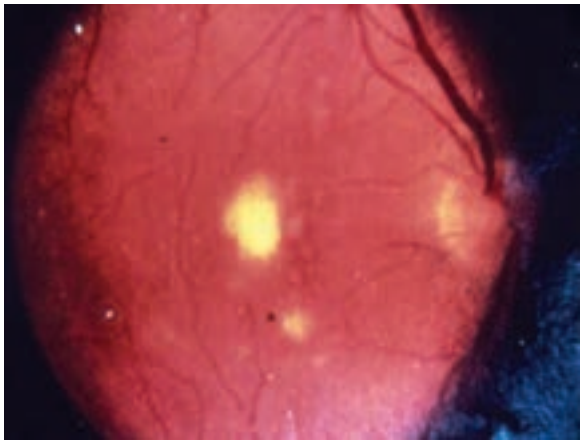

FIGURE 338-3. Retinal involvement.

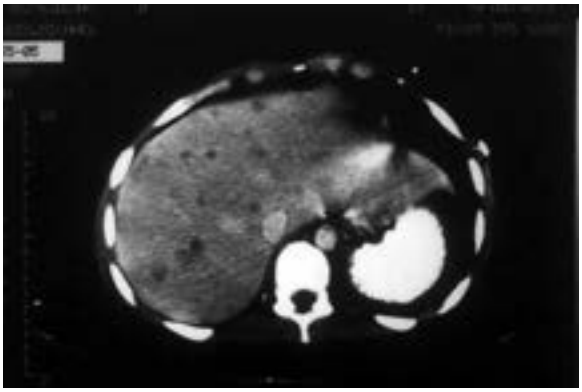

FIGURE 338-4. Computed tomography scan of a patient with chronic disseminated candidiasis (hepatosplenic candidiasis). Note the distinctive punched-out lesions in the liver.

Focal Invasive Infections

These forms of candidiasis result from local inoculation, contiguous spread, or hematogenous spread. Hematogenous spread, which often goes undetected, is probably the most common pathogenetic mechanism.

Urinary Tract Infections

Candiduria is a frequent finding in hospitalized patients and is related to factors such as diabetes mellitus, broad-spectrum antimicrobial treatment, indwelling urinary devices, and genitourinary tract structural abnormalities. Most patients with candiduria have only bladder colonization and not infection. Urinary tract infection with *Candida* species can arise by two mechanisms. Patients with candidemia can develop multiple microabscesses secondary to hematogenous spread to the kidneys. Other patients, who have the risk factors noted earlier, can develop cystitis or ascending infection with pyelonephritis. Patients with cystitis or pyelonephritis have symptoms indistinguishable from those of bacterial infections. A fungus ball composed of fungal hyphae can develop at any level of the collecting system and lead to obstruction, with subsequent infection.

Osteoarticular Infections

Osteoarticular infections arise secondary to hematogenous seeding or exogenous inoculation during intra-articular injection, a surgical procedure, or trauma. Vertebral osteomyelitis is the most common manifestation of osteoarticular candidiasis. The symptoms of back pain and fever may occur many weeks after an episode of fungemia.

Endophthalmitis

Exogenous endophthalmitis occurs secondary to trauma or ophthalmic surgery. Most often, the procedure involved is cataract extraction, with or without lens implantation, and the most common infecting species is *C. parapsilosis.* Primary infection occurs in the anterior chamber, but ultimately the posterior chamber is also involved. Endogenous *Candida* endophthalmitis results from hematogenous seeding of the choroid and retina and is one of the most serious complications of candidemia. Characteristic white lesions are visible in the retina, and with progression of the infection, vitritis occurs; the risk for loss of vision is quite high.

Peritonitis

Candida peritonitis can follow bowel surgery or perforation. Symptoms are the same as those noted in bacterial peritonitis. Usually, this type of infection is polymicrobial, and abscess formation is common. In patients maintained on continuous ambulatory peritoneal dialysis, *Candida* peritonitis generally develops as a late infection after previous episodes of bacterial peritonitis. A cloudy dialysate, abdominal pain, and fever are typically noted.

Meningitis

Acute *Candida* meningitis occurs as part of disseminated infection, especially in low-birthweight neonates. Chronic meningitis, an uncommon manifestation of candidiasis, resembles cryptococcal or tuberculous meningitis with regard to symptoms and cerebrospinal fluid findings.

DIAGNOSIS

The diagnosis of mucocutaneous candidiasis is often made clinically. Culture is rarely indicated. Confirmation can be sought by scraping the lesions and performing either a potassium hydroxide preparation or a Gram stain to look for budding yeasts (Chapter 436). In cases in which the disease is recurrent or unresponsive to standard therapy, lesions should be cultured to establish whether a more resistant species, such as *C. glabrata* or *C. krusei,* is the causative agent. In the event of suspected esophagitis, endoscopy shows plaque-like lesions or ulcerations, and biopsy shows mucosal invasion with budding yeasts and pseudohyphae.

The diagnosis of invasive candidiasis is more difficult. Evidence of dissemination is usually sought by culturing blood or other sterile body sites. The automated blood culture systems used by most hospitals are as sensitive as the lysis-centrifugation system for growing *Candida* from blood. However, no system is sensitive enough for clinicians to rely on blood cultures to establish the diagnosis of invasive candidiasis in all cases or to rule out candidiasis as a diagnostic possibility. In addition, 1 to 4 days is required for growth to occur; in a desperately ill patient, this delay is problematic.

The tips of intravenous catheters that have been removed should be sent for culture. However, no studies have evaluated the number of yeasts that is indicative of infection, and many physicians accept the growth of any yeast as affirming infection that requires treatment. Many focal forms of candidiasis are indistinguishable from bacterial infection, and biopsy should be performed for histopathologic and culture studies.

In a seriously ill patient suspected of having candidiasis, the development of pustular skin lesions or typical retinal lesions can be helpful. Budding yeasts typical of *Candida* species should be sought by smearing material from a pustule on a slide and staining it with Gram stain or by performing a biopsy of a lesion and staining the tissue section with a silver stain. All patients who are candidemic or suspected of having disseminated *Candida* infection should undergo a dilated ophthalmologic examination, preferably by an ophthalmologist, to look for typical retinal lesions.

Imaging studies are invaluable for certain forms of candidiasis, especially chronic disseminated candidiasis, and they can be of major help in defining the extent of infection in other types of *Candida* infection, such as osteoarticular and urinary tract infections and endocarditis.

There are increasing reports of using an assay for β-D-glucan, a cell wall component of fungi, or polymerase chain reaction (PCR) as diagnostic aids for invasive candidiasis.[7,8] The β-D-glucan assay is commercially available; sensitivity and specificity vary depending on the patient populations studied and further studies need to be performed to establish its role in diagnosis. PCR is not standardized, but some studies show that it is more sensitive than β-D-glucan and it may prove useful for earlier diagnosis.

TREATMENT

Rx

Guidelines for treatment of the various forms of candidiasis have been published by the Infectious Diseases Society of America (IDSA).[9] Mucocutaneous disease is obviously treated in a much different fashion than disseminated

life-threatening infection. Because diagnostic tests are not sensitive, empirical therapy is indicated in some circumstances, and for patients at the highest risk for *Candida* infection, antifungal prophylaxis can decrease that risk.

Mucocutaneous Infections

Most mucocutaneous infections should initially be treated with local creams, solutions, troches, or suspensions.[10] For thrush, clotrimazole troches (10 mg four or five times daily) are preferred to nystatin suspension (commonly given as "swish and swallow" four times daily). Patients with AIDS may not respond to local therapy, especially when their CD4 counts are low; in this situation, oral fluconazole 100 to 200 mg daily is given. For vaginitis, a variety of creams and vaginal tablets (miconazole, clotrimazole, and others) are effective, but many women prefer to take a single 150-mg fluconazole tablet orally. Recurrent vaginitis is a more complicated therapeutic issue and often requires chronic suppressive therapy with fluconazole. Esophagitis should always be treated with a systemically absorbed agent; the usual treatment is fluconazole 200 mg/day for 14 days.

In patients with advanced AIDS and low CD4 counts, who are often taking fluconazole to prevent recurrent candidiasis, fluconazole-refractory disease may develop. For these patients, increasing the dosage of fluconazole or switching to itraconazole suspension 200 mg daily, voriconazole 200 mg twice daily, or posaconazole suspension 400 mg daily should be effective. If oral tablets and solutions are no longer effective, intravenous amphotericin B, caspofungin, anidulafungin, and micafungin are alternative agents that can be used. Patients with the syndrome of chronic mucocutaneous candidiasis require lifelong suppressive therapy with oral azole agents.

Candidemia and Invasive Candidiasis

All patients with candidemia should be treated with an antifungal agent, including patients who have only one blood culture that yields *Candida* and those with a vascular catheter tip that yields *Candida*. The rationale for this recommendation is related to the high rate of metastatic foci in major organs associated with hematogenously disseminated candidiasis. Randomized controlled trials have shown the effectiveness of the following antifungal agents for the treatment of candidemia: fluconazole 400 or 800 mg/day; the three echinocandins—caspofungin 50 mg/day, anidulafungin 100 mg/day, and micafungin 100 mg/day; voriconazole 3 mg/kg twice daily; amphotericin B 0.7 mg/kg/day; and a lipid formulation of amphotericin B 3 to 5 mg/kg/day. The IDSA guidelines recommend fluconazole for patients who are not severely ill and have not had recent azole exposure and an echinocandin for severely ill patients and those who have had recent azole exposure. Patients who have stabilized clinically and are found to have an isolate, such as *C. albicans*, that is likely to be susceptible to fluconazole can be transitioned to fluconazole from an echinocandin. Voriconazole is recommended for step-down therapy rather than initial therapy, and amphotericin B formulations are used infrequently, except for patients who are neutropenic and for neonates.

All vascular catheters should be removed because removal has been shown to help clear *Candida* from blood more quickly. Repeated blood cultures should be obtained to ascertain that the fungemia has resolved, and treatment should continue for 2 weeks after the date of the first negative blood culture. An individual patient-level quantitative review of seven randomized trials for the treatment of invasive candidiasis found an overall mortality in the entire data set of 31.4%.[11] Significant predictors of mortality included increasing age; use of immunosuppressive therapy; and infection with *C. tropicalis*. Improved survival and clinical success was found with the use of an echinocandin and the removal of central venous catheters.

Because diagnostic tests are not sensitive, seriously ill patients who could have invasive candidiasis may need to be treated before culture confirmation. This approach is used frequently in neutropenic patients and is increasingly used in the ICU setting.[12] Liposomal amphotericin B, caspofungin, and voriconazole have been shown to be effective for empirical use in neutropenic patients. A placebo-controlled, randomized trial of fluconazole empirical therapy in ICU patients failed to show a benefit; however, there were acknowledged problems with the chosen end point, and the rate of candidemia was too low to allow a proper evaluation of empirical therapy. The IDSA guidelines recommend that empirical therapy be reserved for febrile, critically ill patients who have risk factors for invasive candidiasis. The agents recommended are either fluconazole or an echinocandin, with the caveats noted earlier for treating patients with documented candidemia. Compelling data for early treatment come from a study of 224 candidemic patients who had septic shock. Mortality rates as high as 98% were found in patients in whom there was a delay beyond 24 hours of the onset of shock in initiating antifungal therapy and in effecting source control, defined as draining abscesses and removing central venous catheters.

Endocarditis should be treated with a lipid formulation of amphotericin B, with or without flucytosine. Echinocandins are an acceptable alternative. Infected valves should be replaced. In a few patients for whom valve replacement was not an option, lifelong suppression with fluconazole appeared to be effective.

Chronic disseminated candidiasis generally requires months of therapy for cure. Most patients begin therapy with a lipid formulation of amphotericin B and are then switched to fluconazole and treated until the lesions disappear on computed tomography scanning.

Focal Invasive Infections

Treatment of focal infections depends on the organ system involved. Perhaps the simplest to treat are urinary tract infections. Most patients with candiduria are not infected but merely colonized; removing the selective pressure of antimicrobials and indwelling catheters eliminates candiduria in many of these patients. For those who have infection, oral fluconazole at a dosage of 200 mg/day for 2 weeks is recommended. Bladder irrigation with amphotericin B should not be used because it eradicates only bladder colonization, requires that a catheter be placed in the bladder, and is associated with a high recurrence rate. None of the newer antifungal agents has a role in the treatment of urinary tract infections.

Osteoarticular infections require months of therapy; a lipid formulation of amphotericin B or an echinocandin can be given initially, followed by long-term therapy with an azole. Peritonitis associated with chronic ambulatory peritoneal dialysis can be treated with amphotericin B, fluconazole, or an echinocandin, depending on the species of *Candida* causing infection. Intraperitoneal administration of amphotericin B can be extremely irritating and should not be attempted. The dialysis catheter should be removed. Meningitis should be treated initially with a lipid formulation of amphotericin B and flucytosine; patients with more chronic disease can be switched to fluconazole for a longer duration of therapy.

Treatment of *Candida* eye infections varies with the extent of ocular involvement.[13] Lesions discovered early at the stage of choroidal or retinal involvement perhaps can be treated effectively with systemic antifungal agents (amphotericin B, an echinocandin, fluconazole, or voriconazole) alone. Many experts prefer to use an agent, such as voriconazole or fluconazole, that achieves higher concentrations in the eye. Lesions extending into the vitreous require more aggressive therapy. The best results have been obtained with pars plana vitrectomy, injection of amphotericin B or voriconazole into the vitreous; and a systemic antifungal agent such as fluconazole or voriconazole. Management must be individualized and performed in concert with an ophthalmologist experienced in the treatment of this infection. Treatment of endophthalmitis associated with an intraocular lens implant requires removal of the implant, vitrectomy, and local amphotericin B injections, as well as therapy with fluconazole or voriconazole.

PREVENTION

For certain populations at the highest risk for invasive fungal infection, prophylactic antifungal agents can prevent infection. The populations for whom prophylaxis is recommended include stem cell transplant recipients, patients with acute leukemia who are undergoing induction chemotherapy, high-risk liver transplant recipients, and pancreas and small bowel transplant recipients; in these groups, a variety of different agents are effective. In the ICU population, prophylaxis with fluconazole can be effective, but it is recommended only in units that have a high rate of invasive candidiasis, and only in those patients at the highest risk for infection. In a placebo-controlled trial of caspofungin as antifungal prophylaxis in adults who were in the ICU for at least 3 days, were ventilated, received antibiotics, had a central line, and had at least one additional risk factor, caspofungin was safe and tended to reduce the incidence of invasive candidiasis when used for prophylaxis, but the difference was not statistically different.[A1] Restricting the use of prophylaxis is essential to prevent the widespread use of azoles, with subsequent selection of resistant species.

PROGNOSIS

The prognosis for patients with mucocutaneous infections is excellent. The prognosis for focal invasive infections depends on the organ involved and the patient's immune status. For example, whereas pyelonephritis may respond well to antifungal therapy, endocarditis and meningitis are more difficult to treat and have poor outcomes. Invasive candidiasis has a high mortality rate. Early treatment with an effective antifungal agent is extremely important for a favorable outcome.

Grade A References

A1. Schuster MG, Edwards JE Jr, Sobel JD, et al. Empirical fluconazole versus placebo for intensive care unit patients: a randomized trial. *Ann Intern Med*. 2008;149:83-90.

A2. Ostrosky-Zeichner L, Shoham S, Vasquez J, et al. MSG-01: a randomized, double-blind, placebo-controlled trial of caspofungin prophylaxis followed by preemptive therapy for invasive candidiasis in high-risk adults in the critical care setting. *Clin Infect Dis*. 2014;58:1219-1226.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Quindos G. Epidemiology of candidaemia and invasive candidiasis. A changing face. *Rev Iberoam Micol*. 2014;31:42-48.
2. Ngo LY, Kasahara S, Kumasaka DK, et al. Inflammatory monocytes mediate early and organ-specific innate defense during systemic candidiasis. *J Infect Dis*. 2014;209:109-119.
3. Ilkit M, Guzel AB. The epidemiology, pathogenesis, and diagnosis of vulvovaginal candidosis: a mycological perspective. *Crit Rev Microbiol*. 2011;37:250-261.
4. van de Veerdonk FL, Plantinga TS, Hoischen A, et al. *STAT1* mutations in autosomal dominant chronic mucocutaneous candidiasis. *N Engl J Med*. 2011;365:54-61.
5. Sarkadi AK, Tasko S, Csorba G, et al. Autoantibodies to IL-17A may be correlated with the severity of mucocutaneous candidiasis in APECED patients. *J Clin Immunol*. 2014;34:181-193.
6. Kollef M, Micek S, Hampton N, et al. Septic shock attributed to *Candida* infection: importance of empiric therapy and source control. *Clin Infect Dis*. 2012;54:1739-1745.
7. Nguyen MH, Wissel MC, Shields RK, et al. Performance of Candida real-time polymerase chain reaction, beta-D-glucan assay, and blood cultures in the diagnosis of invasive candidiasis. *Clin Infect Dis*. 2012;54:1240-1248.
8. Karageorgopoulos DE, Vouloumanou EK, Ntziora F, et al. β-D-glucan assay for the diagnosis of invasive fungal infections: a meta-analysis. *Clin Infect Dis*. 2011;52:750-770.
9. Pappas PG, Kauffman CA, Andes D, et al. Clinical practice guidelines for the management of candidiasis: 2009 update by the Infectious Diseases Society of America. *Clin Infect Dis*. 2009;48:503-535.
10. Summers PR. Topical therapy for mucosal yeast infections. *Curr Probl Dermatol*. 2011;40:48-57.
11. Andes DR, Safdar N, Baddley JW, et al. Impact of treatment strategy on outcomes in patients with candidemia and other forms of invasive candidiasis: a patient-level quantitative review of randomized trials. *Clin Infect Dis*. 2012;54:1110-1122.
12. Bassetti M, Mikulska M, Viscoli C. Bench-to-bedside review: therapeutic management of invasive candidiasis in the intensive care unit. *Crit Care*. 2010;14:244.
13. Riddell J, Comer GM, Kauffman CA. Treatment of endogenous fungal endophthalmitis: focus on new antifungal agents. *Clin Infect Dis*. 2011;52:648-653.

REVIEW QUESTIONS

1. *Candida glabrata* infections are more difficult to treat for which of the following reasons?

A. *C. glabrata* is more virulent.
B. *C. glabrata* cannot be phagocytized by neutrophils.
C. *C. glabrata* is resistant to azole antifungals.
D. *C. glabrata* readily forms hyphae and invades tissues.

Answer: C *C. glabrata* infections are difficult to treat because of both inherent and acquired resistance to fluconazole and other azoles. It is not more virulent and may actually be less so. It does not form hyphae in contrast to the other species of *Candida*. Phagocytosis of *C. glabrata* occurs just as it does with other *Candida* species.

2. An AIDS patient comes in because he has retrosternal pain when he swallows. He does not have thrush, and there is no cervical lymphadenopathy. What is the most likely diagnosis?

A. Gastroesophageal reflux disease
B. Pericarditis
C. Pulmonary embolus
D. Esophageal candidiasis
E. Angina

Answer: D Odynophagia is the classic symptom shown by AIDS patients who have *Candida* esophagitis. The other diseases generally do not cause pain when swallowing food. Esophagitis can occur without thrush, although frequently both are present concomitantly. Response to treatment with fluconazole can be used as a diagnostic test for *Candida* esophagitis; endoscopy can be done, also, but is usually reserved for patients who do not respond to fluconazole therapy.

3. Which is *not* a risk factor for candidemia?

A. High APACHE II score
B. Central venous catheter
C. Deficient T-cell immunity
D. Prior surgical procedure
E. Neutropenia

Answer: C The risk factors for candidemia include everything noted above except deficient T-cell immunity. T cells are important in protection against mucosal infection with *Candida* species but do not play a role in prevention of invasive candidiasis and candidemia.

4. A 70-year-old man who is in the intensive care unit after having abdominal surgery for a perforated bowel becomes hypotensive and febrile. He has a central catheter in place, is receiving parenteral nutrition and broad-spectrum antibiotics, and has kidney failure requiring renal replacement therapy. What is the most appropriate approach to treatment after you obtain blood cultures?

A. Treat with fluconazole when the laboratory tells you that yeasts are growing in the cultures.
B. Treat with an echinocandin after the laboratory identifies the yeast in the blood cultures as *C. glabrata.*
C. Treat with amphotericin B immediately after obtaining the blood cultures.
D. Treat with fluconazole after the laboratory identifies the yeast in the blood culture as *C. glabrata.*
E. Treat with an echinocandin immediately after obtaining the blood cultures.

Answer: E In a patient at great risk for candidemia, as is this patient (who is on renal replacement therapy, has a central venous catheter plus another central line for dialysis, is receiving broad-spectrum antibiotics and parenteral nutrition, has a perforated bowel, and has undergone recent abdominal surgery), empirical treatment with an antifungal agent should be started as soon as the blood cultures and any other pertinent cultures have been obtained. Amphotericin B is rarely used in this setting, and echinocandins or fluconazole are generally used. In a patient who is not stable and is hypotensive, an echinocandin is preferred. Waiting for the cultures to grow, which will take 1 to 3 days, and for identification of species, which may add another 2 to 3 days, increases the mortality rate enormously in candidemic patients.

5. Which finding is helpful in directing the clinician to a diagnosis of invasive candidiasis?

A. Nonpainful papular skin lesions on an erythematous base on the chest and left arm
B. Painful erythematous nodules developing on the shins
C. Bilateral spreading erythema over the lower legs and feet
D. Molluscum-type lesions on the forehead accompanied by fever
E. Pruritic blisters appearing in crops over the trunk

Answer: A Candidemia and invasive candidiasis can present with the sudden appearance of nonpainful, nonpruritic papules that evolve to become pustules on an erythematous base. These can occur anywhere on the body. All of the other lesions do not occur with candidemia.

339

ASPERGILLOSIS

THOMAS J. WALSH

DEFINITION

Aspergillosis is defined as an infection with one or more of the species of the genus *Aspergillus.* Sporelike structures called *conidia* are aerosolized from the mold form of the organism growing in the environment. When conidia reach tissue, they germinate to form invasive filaments called *hyphae.*

The Pathogens

The most common species infecting humans are *Aspergillus fumigatus, Aspergillus flatus, Aspergillus terreus, and Aspergillus niger.* The species are usually identified in culture by characteristic microscopic features of hyphae and the structures producing conidia. When some species are not readily identifiable, they may be reported by the clinical laboratory as "*Aspergillus* species or *Aspergillus* sp." Molecular methods are increasingly used for identification. *Aspergillus fumigatus* may be reported as "*A. fumigatus* species complex." Some species within *A. fumigatus* complex may be particularly drug resistant. *Aspergillus terreus* is resistant to amphotericin B. Aspergilli within tissue appear as dichotomously branched (Y-shaped) septate hyphae. *Scedosporium* and *Fusarium* species also may produce septate hyphae in tissue. The presence of septa and dichotomous branching differentiates *Aspergillus* species from the Mucorales, which are the causative organisms of mucormycosis (Chapter 340).

EPIDEMIOLOGY

Aspergillus species. are ubiquitous organisms in the external environment, including soil, decaying matter, and air in temperatures as high as 40 to 50° C.[1] Aspergilli are easily isolated from houses, particularly from basements, crawl spaces, bedding, humidifiers, ventilation ducts, potted plants, dust, condiments (e.g., pepper), and marijuana samples. Aspergilli cause abortion in cattle and are important pathogens of marine organisms, insects, and domesticated and wild birds. Aflatoxin, which is one of the most potent carcinogens known, is produced by strains of *Aspergillus flavus* at ambient temperature on stored grain, spices, and nuts. Foodborne ingestion of preformed aflatoxin may cause hepatic necrosis or hepatocellular carcinoma (Chapter 196) in animals and humans.

Aspergillus species may be acquired from airborne conidia in inpatient and outpatient settings. Nosocomial aspergillosis is associated with building renovation, new construction, unfiltered air, contaminated ventilation systems, and fireproofing materials.[2] Hospital water, which may become aerosolized during activities such as showering, is a newly described potential source of aspergilli. As human pathogens, *Aspergillus* species may cause acute invasive disease, chronic infection, or allergic symptoms. A classification of aspergillosis is presented in Table 339-1.

Acute invasive aspergillosis develops in immunocompromised patient populations, particularly those with hematologic malignancies, hematopoietic stem cell transplantation (HSCT), severe aplastic anemia, primary immunodeficiencies, and solid organ transplantation, especially of heart, lung, and liver.[1] Genetic deficiency of the soluble pattern-recognition receptor called PTX3 (long pentraxin 3) caused by homozygous haplotype (h2/h2) in the *PTX3* gene of donor cells has been found to lead to impaired neutrophil antifungal capacity and increased risk for invasive aspergillosis in recipients of HSCT.[3] Persistent neutropenia, corticosteroids, other immunosuppressive agents, graft versus host disease (GVHD), and cytomegalovirus (CMV) disease are the most frequently observed clinical risk factors. The mortality of acute invasive aspergillosis varies from as much as 100% with central nervous system (CNS) infection to approximately 65% with pulmonary infection in HSCT recipients. Early recognition of clinical manifestations followed by initiation of antifungal therapy may improve the ominous prognosis of acute invasive aspergillosis.

TABLE 339-1 CLASSIFICATION OF ASPERGILLOSIS

CATEGORY	SPECIFIC FORMS OF ASPERGILLOSIS
Acute invasive aspergillosis	Invasive pulmonary aspergillosis Empyema Tracheobronchial infection Extrapulmonary aspergillosis Acute sinusitis Focal rhinitis Cerebral, cerebellar, or brain stem infarction Endophthalmitis Osteomyelitis Epidural abscess Cardiac aspergillosis Myocarditis Endocarditis Pericarditis Gastrointestinal aspergillosis Renal infection Cutaneous lesions (nodules, ulcers) Disseminated aspergillosis
Chronic aspergillosis	Aspergilloma Chronic necrotizing pulmonary aspergillosis Chronic cavitary pulmonary aspergillosis *Aspergillus* otomycosis
Allergic forms of aspergillosis	Allergic bronchopulmonary aspergillosis Extrinsic allergic alveolitis Allergic *Aspergillus* sinusitis

CLINICAL MANIFESTATIONS

Invasive Aspergillosis

The classic clinical manifestations of *invasive pulmonary aspergillosis* (IPA) in immunocompromised hosts are fever and focal pulmonary infiltrates, nodules, or wedge-shaped densities resembling infarcts (see Table 339-1).[4] Cough, pleuritic pain, and hemoptysis also may be present. Focal pulmonary infiltrates may progress to a cavity on recovery from neutropenia. Pulmonary infiltrates may also present as bronchopneumonia in an immunosuppressed patient. The pulmonary pathology in all these entities is that of hemorrhagic infarction caused by the organism's capacity to invade blood vessel walls (angioinvasion). These processes lead to formation of a necrotic center surrounded by a ring of hemorrhage and edema, which correlates with a "halo sign" surrounding the nodular density. Concomitant pleural effusion may develop and represent *Aspergillus* empyema. Tracheobronchial aspergillosis in immunocompromised patients presents as ulcerative, pseudomembranous, or plaquelike large airway disease that may presage pulmonary parenchymal invasion.

Acute Aspergillus sinusitis may occur concomitantly or independently of IPA. Although symptoms may include fever, localized pressure, and pain, they may be absent in severely immunocompromised patients. Eschar may be observed by speculum examination or endoscopy on the nasal septum and turbinates. Acute *Aspergillus* sinusitis of the ethmoid and sphenoid sinuses may progress to cavernous sinus thrombosis with symptoms referable to cranial nerves III, IV, $V_{1,2}$, and VI. *Aspergillus flavus* has a high propensity for causing acute sinus infection.

The tissue targets of *extrapulmonary* and *disseminated aspergillosis* most commonly include the CNS, where abscesses and infarcts are characteristic.[5] Patients with CNS aspergillosis present with focal paresis, cranial nerve deficits, and seizures. The glucose level in cerebrospinal fluid (CSF) is usually normal, and cultures of CSF are negative. Other extrapulmonary manifestations include endophthalmitis, myocardial infarction, gastrointestinal disease, renal infarction, cutaneous lesions, and Budd-Chiari syndrome. Esophageal ulcers and mesenteric thrombosis may produce gastrointestinal bleeding. Renal infection may present as flank pain and hematuria.

Aspergillus endocarditis usually begins as an isolated infection in intravenous drug users or after cardiac valvular surgery.[6] *Aspergillus* endocarditis most commonly presents as major arterial emboli. Blood cultures, which are

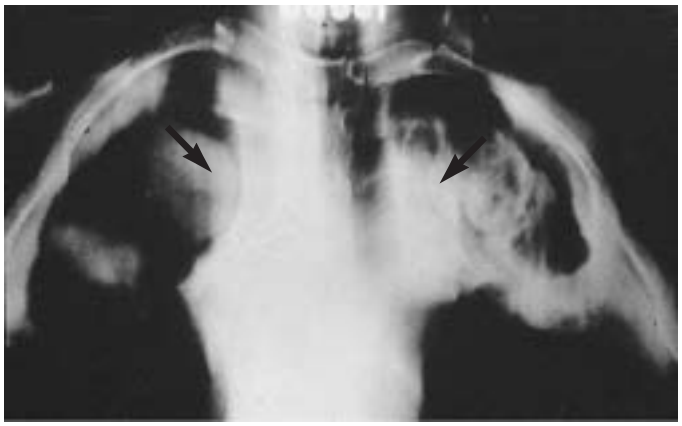

FIGURE 339-1. Tomogram of pulmonary aspergillomas *(arrows)*.

seldom positive, may be delayed in growth by as much as 14 to 21 days. Diagnosis is difficult, and despite valve replacement with antifungal therapy, mortality approaches 100%. *Aspergillus pericarditis* may arise from contiguous pulmonary lesions or through transmural infection from endocardial infection.

Locally invasive aspergillosis usually develops in immunocompromised patients as cutaneous ulcers, focal rhinitis, osteomyelitis, and septic arthritis.[7] Cutaneous ulcers have been associated with use of contaminated adhesive tape and arm boards. Blood-borne infection in illicit intravenous drug users may present as foci of dissemination in brain, lung, kidney, and bone. Keratitis, endophthalmitis, and infection of burn wounds may develop from traumatic inoculation in otherwise immunocompetent patients.

Chronic Pulmonary Aspergillosis

Aspergilloma appears on chest radiograph as a ball in a cavity. The fungus ball consists of matted hyphae and debris in a preformed cavity from previous pulmonary tuberculosis, histoplasmosis, or fibrocystic sarcoidosis (Fig. 339-1). Symptomatic patients present with cough, hemoptysis, dyspnea, weight loss, fatigue, chest pain, or fever. Sputum culture is typically positive for *Aspergillus* species, particularly *Aspergillus niger*. Pleural aspergillosis may complicate surgical resection of aspergilloma or develop spontaneously as a bronchopleural fistula or concomitantly with tuberculosis.

As a stage in the repair process of infarcted lung tissue in neutropenic patients, one or more apparent "aspergillomas" may develop in consolidated lesions during recovery from neutropenia. These apparent aspergillomas do not develop in preexisting cavities and create an "air-crescent sign," or Monod sign, during their formation.

Chronic necrotizing pulmonary aspergillosis (CNPA) and *chronic cavitary pulmonary aspergillosis* (CCPA) occur in patients with underlying chronic lung disease, chronic immunosuppression, such as that due to prolonged use of systemic corticosteroids, or both. CNPA characteristically causes a slowly progressive inflammatory destruction of lung tissue superimposed on chronic lung disease. Clinical manifestations of worsening pulmonary function, cough, and dyspnea in CNPA may be indistinguishable from concomitant primary chronic respiratory disease.

CCPA is defined as the presence of multiple *Aspergillus*-related cavities, which may or may not contain an aspergilloma. Patients with CCPA may have genetically mediated deficits in innate host defenses. Occurring in association with symptoms of cough, hemoptysis, and dyspnea, the progressive cavities of CCPA tend to coalesce with the loss of functional lung tissue.

Aspergillus otomycosis is a chronic infection that usually involves the external auditory canal with symptoms of pain, pruritus, hypoacusis, and otic discharge in patients with impaired mucocutaneous immunity, such as those with chronic eczema, hypogammaglobulinemia, diabetes mellitus, or HIV infection and those receiving corticosteroids. *Aspergillus* may involve the middle ear and extend into the mastoid sinus if the tympanic membrane has been perforated.

Allergic Forms of Aspergillosis

Allergic bronchopulmonary aspergillosis (ABPA) develops most frequently in patients with a history of chronic asthma or cystic fibrosis.[8,9] Occurring in genetically susceptible patients exposed to specific *Aspergillus* antigens, ABPA is characterized by episodic airway obstruction, fever, eosinophilia, positive sputum cultures, mucous plugs containing hyphae, the presence of grossly visible brown flecks in sputum (hyphae), transient infiltrates and parallel "tramline" or ring markings on chest radiographs, proximal bronchiectasis, upper lobe contraction, and elevated levels of total immunoglobulins G and E (IgG and IgE). Eosinophilia may be present in blood, sputum, and lung tissue. Mucous plugs contribute to development of pulmonary infiltrates, atelectasis, and peribronchial inflammation. The parallel or ring markings are caused by thickened ectatic bronchi, whereas the upper lobe changes are due to progressive apical fibrosis. Pulmonary infiltrates in ABPA may be nonsegmental and transient in association with eosinophilia and asthma; alternatively, they may be segmental and associated with bronchial obstruction by mucous plugs, wherein asthma and eosinophilia may be absent.

Extrinsic allergic alveolitis is an unusual allergic form of *Aspergillus* lung disease that has been most frequently associated with *Aspergillus clavatus* in malt workers. A hypersensitivity pneumonitis with dyspnea and fever develops approximately 4 hours after exposure. Diffuse reticulonodular interstitial infiltrates may be present at the time of symptoms. Patients have IgG precipitins and cell-mediated immune reactions against *Aspergillus* antigens. Granulomas are present in lung tissue. In comparison to ABPA, eosinophilia is not a feature of *Aspergillus* extrinsic allergic alveolitis.

Allergic Aspergillus sinusitis (AAS) is a noninvasive form of sinus disease that typically presents in patients with asthma, nasal polyps, sinus opacification, and eosinophilia. Sinus aspirate yields mucinous material containing eosinophils, Charcot-Leyden crystals, and hyphal elements. AAS and ABPA may coexist in some patients. Advanced forms of AAS may present with proptosis and optic neuropathy, necessitating prompt surgical intervention.

DIAGNOSIS

Invasive Aspergillosis

Diagnosis of IPA and disseminated aspergillosis is difficult. None of the aforementioned clinical manifestations are diagnostic for invasive aspergillosis. Advances in computed tomography (CT) have revealed characteristic features of nodules, halo signs, wedge-shaped infiltrates, and air-crescent signs during IPA in immunocompromised patients (Fig. 339-2). However, infections caused by *Fusarium* species, *Scedosporium* species, the Mucorales, as well as *Pseudomonas aeruginosa* may be radiologically indistinguishable from IPA. Microbiologic confirmation, where possible, is important to differentiate aspergillosis from other filamentous fungal infections. Bronchoalveolar lavage (BAL), percutaneous needle aspiration, video-assisted thoracoscopic (VATS) biopsy, and if necessary, open lung biopsy are standard procedures for establishing a microbiologic diagnosis of invasive aspergillosis. Specimens obtained from these procedures may demonstrate dichotomously branching septate hyphae by direct microscopy or grow *Aspergillus* species in culture. Each of these procedures is associated with false-negative results, as well as with complications. Conversely, the presence of *Aspergillus* by direct examination or culture in an immunocompromised host with pulmonary nodules or well-circumscribed infiltrates carries a high probability for diagnosis of invasive aspergillosis.

Galactomannan, which is a heteropolysaccharide of the *Aspergillus* cell wall, is a useful biomarker that is released into the circulation and alveolar spaces during IPA.[10] Detection of galactomannan by enzyme immunoassay (EIA) in serum or BAL above certain thresholds is strong microbiologic evidence for a diagnosis of invasive aspergillosis in immunocompromised patients with characteristic clinical manifestations. Nonetheless, false-positive results have been reported in patients who have received piperacillin-tazobactam or amoxicillin-clavulanate, in cases in which Plasmalyte was used for BAL, and in other deeply invasive mycoses such as blastomycosis and histoplasmosis. Serum galactomannan may be falsely negative in patients receiving antifungal prophylaxis or empirical therapy.

(1→3)-β-D-glucan is another *Aspergillus* cell wall polysaccharide that is detected in serum during invasive disease. The sensitivity of the *Limulus* spectrophotometric assay for detection of (1→3)-β-D-glucan in patients with invasive aspergillosis appears to be comparable to that of the galactomannan EIA. However, because (1→3)-β-D-glucan also is present in the cell wall of other medically important fungi, including *Candida* species, the specificity of (1→3)-β-D-glucan for *Aspergillus* species is less than that of galactomannan.

Although molecular diagnostic tools, such as quantitative polymerase chain reaction (PCR), for diagnosis of invasive aspergillosis are promising, they also remain investigational for this infection. Ultimately, the combined use of improved diagnostic imaging, microscopy, culture methodology, cell

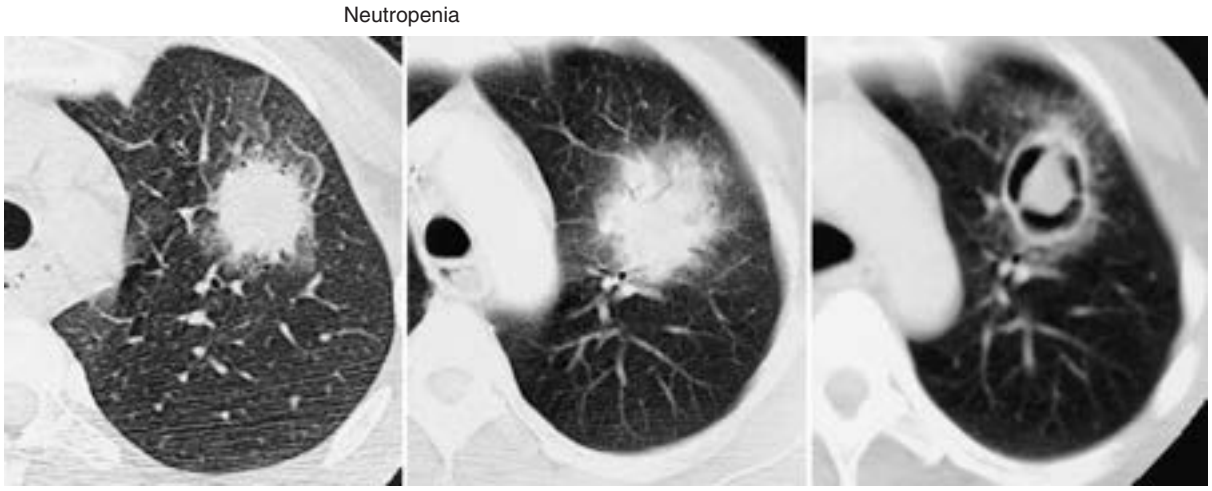

FIGURE 339-2. Evolution of radiography of invasive aspergillosis in an immunocompromised host. D = days after the lesion is first noted.

wall biomarkers, and possibly PCR in conjunction with careful bedside assessment of risk factors and clinical manifestations will improve the diagnosis of invasive aspergillosis and early initiation of therapy.

Diagnosis of locally invasive extrapulmonary aspergillosis causing mucocutaneous lesions, osteomyelitis, and septic arthritis is best accomplished with biopsy, direct microscopy, and culture. A diagnosis of *Aspergillus* keratitis is established by careful culture of corneal lesions by an ophthalmologist.

Allergic Forms of Aspergillosis

Diagnosis of ABPA is based on the presence of a combination of clinical, biologic, and radiologic, criteria (Fig. 339-3). The consensus criteria for establishing a diagnosis of ABPA differ depending on the presence of cystic fibrosis. For patients with ABPA without the presence of cystic fibrosis (Chapter 89), criteria for ABPA include asthma, an immediate cutaneous reaction to *A. fumigatus* antigen, total serum IgE concentration higher than 1000 ng/mL, elevated *A. fumigatus*–specific serum IgE levels, precipitating serum antibodies to *A. fumigatus*, central bronchiectasis, peripheral blood eosinophilia, and characteristic pulmonary infiltrates. The latter two features (eosinophilia and pulmonary infiltrates) are considered nonessential because they only may be present during an acute phase of ABPA.

Among patients with cystic fibrosis, distinguishing between ABPA and an episode of clinical deterioration with colonization by *Aspergillus* species is challenging. The current criteria of the Cystic Fibrosis Foundation help to define ABPA in that setting: clinical deterioration (coughing, wheezing, increased sputum production, exercise intolerance, and decrease in pulmonary function); immediate hypersensitivity to *A. fumigatus* (positive skin test or IgE response); total serum IgE concentration higher than 1000 kUI/L; precipitating antibodies to *A. fumigatus*; abnormal chest radiograph (infiltrate, mucous plugs, or unexplained changes compared with previous chest radiograph).

A biphasic skin test response may assist in the diagnosis. A scratch test with *Aspergillus* antigens produces an immediate wheal-and-flare reaction that is mediated by IgE and blocked by antihistamines but not by corticosteroids. An intracutaneous test with the antigens produces a later (6 to 8 hours) reaction that is mediated by IgG and complement and blocked by corticosteroids.

An occupational history of exposure is critical to the diagnosis of *extrinsic allergic alveolitis*. A typical history of recurrent episodes developing within 24 hours after inhalation of conidial antigens in an agricultural environment in conjunction with a negative scratch test, a positive intradermal test, and granulomas with immunoglobulins and complement in tissue is diagnostically consistent with *Aspergillus* extrinsic allergic alveolitis.

Recurrent sinusitis in a patient with asthma, nasal polyps, eosinophilia, sinus opacification, and a sinus aspirate yielding mucinous material containing eosinophils, Charcot- Leyden crystals, and hyphal elements establishes a diagnosis of AAS.

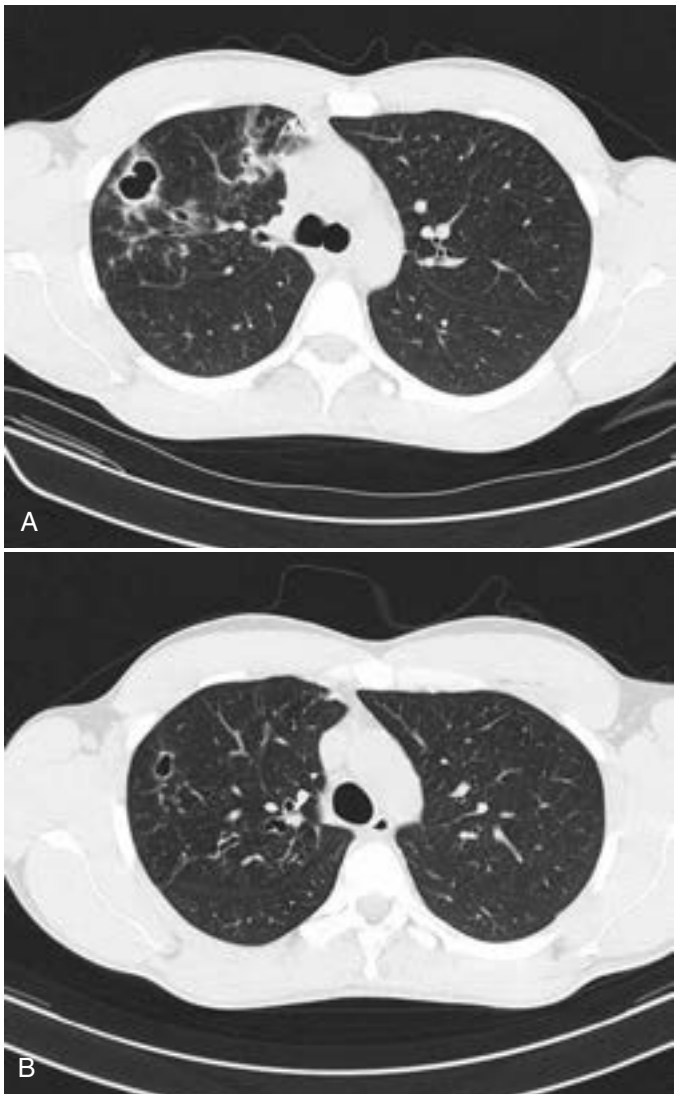

FIGURE 339-3. **Allergic bronchopulmonary aspergillosis in a patient with a long history of asthma. (A)** is a thin slice CT image showing bronchiectasis with cystic changes in the right upper lobe. This patient had an underlying history of asthma and a markedly elevated IgE level and other findings consistent with allergic bronchopulmonary aspergillosis. **(B)** is the same patient after treatment with systemic steroids; the cystic bronchiectasis has markedly improved. (Courtesy of Anne E. O'Donnell, MD.)

TABLE 339-2 ANTIFUNGAL THERAPY OF INVASIVE ASPERGILLOSIS*

FIRST-LINE TREATMENT IN ADULTS†	
Drug of choice:	
Voriconazole	IV therapy: 6 mg/kg q12h for two doses, then 4 mg/kg q12h Oral therapy: 300 mg or 4mg/kg bid
Alternate (see text for conditions):	
Liposomal amphotericin B	3-5 mg/kg IV daily
SECOND-LINE OR SALVAGE TREATMENT IN ADULTS†	
Amphotericin B lipid complex	5 mg/kg IV daily
or	
Caspofungin	70 mg IV daily for first dose, then 50 mg IV daily
or	
Posaconazole	200 mg PO qid or 400 mg PO bid
or	
Itraconazole	400 mg PO (capsules) daily (in either one or two doses); or 2.5 mg/kg PO (solution) bid

*Refer to package insert for dosage modification of antifungal agents in liver disease or renal impairment.
†Duration of antifungal therapy depends on therapeutic response of documented lesions, burden of disease, host immunocompetence, and type of aspergillosis (e.g., acute invasive vs. chronic, vs. allergic). Guidelines of Infectious Diseases Society of America recommend at least 6 to 12 weeks for invasive pulmonary aspergillosis. Patients who are immunosuppressed continue treatment throughout the period of immunosuppression and until resolution of lesions. In patients with previously diagnosed invasive aspergillosis, antifungal therapy should be continued or reinitiated during subsequent periods of immunosuppression (e.g., chemotherapy, stem cell transplantation, graft vs. host disease) to prevent recrudescence
(see Walsh TJ, Anaissie EJ, Denning DW, et al. Treatment of aspergillosis: clinical practice guidelines of the Infectious Diseases Society of America (IDSA). *Clin Infect Dis.* 2008: 46:327-360.)

TREATMENT

Rx

Invasive Aspergillosis

The foundation of treatment of invasive aspergillosis consists of (1) antifungal medical therapy (Table 339-2), (2) reversal of immunosuppression and, where appropriate, (3) surgical resection of infected lesions (see Table 339-2).[11,12] Dosages of antifungal agents are listed in Table 339-2. Voriconazole is recommended in most patients for the primary treatment of invasive aspergillosis, including pulmonary, disseminated, and extrapulmonary isolated infection. This recommendation is based on the randomized controlled trial showing that voriconazole is superior to deoxycholate amphotericin B (D-AmB) as primary treatment for invasive aspergillosis. However, not all patients are candidates to receive voriconazole. This includes patients with substantially elevated hepatic transaminases, hepatic dysfunction, and a history of hypersensitivity to or intolerance of voriconazole. Such patients should receive liposomal amphotericin B (L-AmB) as primary therapy. This recommendation is based on the randomized trial that demonstrated comparable efficacy of approximately 70% using dosages of 3 mg/kg/day and 10 mg/kg/day of LAmB in patients who had predominantly hematologic malignancies. L-AmB also is indicated as primary therapy for patients in whom there is a suspicion for or documentation of concurrent mucormycosis.

Second-line or salvage antifungal therapy is indicated in patients who are intolerant of or whose infection is unresponsive to primary therapy. Among the antifungal agents used in this setting are a lipid formulation of amphotericin B, posaconazole, itraconazole, or an echinocandin (caspofungin is the only agent licensed for this indication). For patients who are already receiving voriconazole, a change of class to a lipid formulation, addition of an echinocandin, and use of another azole are alternative possibilities. Such decisions are complicated and warrant infectious diseases consultation with consideration of pharmacokinetic and host factors.

Antifungal therapy for invasive aspergillosis should be continued until lesions have resolved, cultures and biomarkers are negative, and reversible underlying predispositions have abated. Reinstating therapy in patients who have previously responded should be considered if immunosuppression is reinstituted or if neutropenia recurs.

Preliminary reports of a randomized multicenter study demonstrate that a new antifungal triazole known as isavuconazole is as effective but better tolerated than voriconazole in the primary treatment of invasive aspergillosis.

Reversal of immunosuppression is a critical factor in the successful management of invasive aspergillosis. Recovery from neutropenia and decreasing the daily dosage or discontinuation of corticosteroids, where feasible, are two of the most important forms of improving host response. Depending on the protocol used, granulocyte transfusions may stabilize *Aspergillus* lesions until recovery from neutropenia. Granulocyte colony-stimulating factor (G-CSF) and granulocyte-macrophage colony-stimulating factor (GM-CSF) may accelerate recovery from neutropenia. The role of GM-CSF, G-CSF, or interferon-γ in immunocompromised non-neutropenic patients with invasive aspergillosis remains to be further defined.

Surgical management of infected lesions is an important adjunctive component of primary therapy for several forms of invasive aspergillosis: endocarditis, pericarditis, osteomyelitis, epidural abscess, infected vascular catheters and prosthetic devices, and skin and soft tissue infection. Surgical management also is important for several conditions of invasive pulmonary aspergillosis: recurrent hemoptysis from a single cavitary lesion, invasion of a pulmonary lesion into the chest wall, and pulmonary lesions contiguous with great vessels or the pericardium. *Aspergillus* empyema requires closed chest tube drainage and possibly débridement of the infected pleural cavity. Débridement of sinus aspergillosis, particularly when the ethmoid and frontal sinuses are infected, may prevent extension into the orbit or into the cavernous sinus. Surgical resection of selected lesions of the CNS may be indicated for establishing a diagnosis, reducing increased intracranial pressure, and/or protecting critical neural centers. Location of CNS lesions and neurologic sequelae after resection are critical factors in neurosurgical management of aspergillosis.

Local infusion of antifungal agents, particularly intravitreal therapy for endophthalmitis, provides high concentrations to compartments that may not be reached by systemic therapy. Topical irrigation with voriconazole, amphotericin B, or if available, pimaricin is an important adjunct to management of *Aspergillus* keratitis.

Chronic Aspergillosis

Medical therapy has limited benefit in treatment of aspergilloma; however, some patients may benefit from extended use of an antifungal triazole. Lifelong commitment to an antifungal triazole should be balanced against the natural history of approximately 10% of aspergillomas resolving spontaneously. By comparison, medical therapy with itraconazole or voriconazole is the standard of treatment for CCPA. Patients with CCPA typically achieve improvement of symptoms and stabilization or improvement of radiologic changes. Patients with CNPA also receive an antifungal triazole; however, assessment of response is more difficult because of the underlying chronic lung disease. The role of surgical resection in patients with solitary aspergilloma or CCPA is limited because of development of bronchopleural fistula, *Aspergillus* infection of the pleural space, and potentially further worsening of already compromised pulmonary function. However, surgical resection may have a more important role in treating patients with recurrent and severe hemoptysis, for which the benefits of removing the cavity usually outweigh the known risks. Bronchial artery embolization and transthoracic direct intracavitary instillation of antifungal agents for aspergilloma have only transient benefits but substantial risk.

Topical irrigating solutions of boric acid, acetic acid, or an antifungal azole cream may be effective in treating *Aspergillus* otomycosis. Voriconazole, posaconazole, or itraconazole may be necessary for refractory cases or perforated tympanic membranes.

Allergic Forms of Aspergillosis

ABPA is treated with a combination of corticosteroids and itraconazole. This recommendation is based on two double-blind, randomized, placebo-controlled trials for treatment of ABPA. These studies demonstrated that itraconazole (200 mg twice daily orally for 16 weeks) resulted in significant amelioration of disease, as evidenced by improvement in exercise tolerance and pulmonary function, reduction in corticosteroid dose, increased interval between corticosteroid courses, as well as decreased eosinophilic inflammatory parameters and IgE concentration. Although corticosteroid therapy is the mainstay of treatment of ABPA, chronic administration of corticosteroids causes severe immunosuppression and multisystem metabolic abnormalities. Addition of itraconazole reduces organism burden, attenuates antigenic stimulus for destructive bronchial inflammation, and provides a corticosteroid-sparing effect. Intermittent use of corticosteroids or substantially raising the dose in patients receiving chronic therapy can produce rapid resolution of marked symptomatic episodes or deteriorating forced expiratory volume in 1 second (FEV_1).

Extrinsic allergic alveolitis is best managed by removing patients from the allergenic environment. An accurate occupational history is critical to this intervention.

AAS is treated by endoscopic drainage to relieve obstruction by tenacious mucin. Itraconazole, nasal corticosteroids, and systemic corticosteroids, alone or in combination, may be beneficial in some patients with AAS. Caution is warranted with chronic use of systemic or nasal corticosteroids. Itraconazole may have a corticosteroid sparing effect.

Combination Antifungal Therapy

Combinations of polyenes, triazoles, and echinocandins are being explored. The aggregate of well-conducted in vitro, in vivo, and clinical observational

studies support the additive or synergistic interaction of a triazole and echinocandin in primary treatment of invasive pulmonary aspergillosis. In a randomized trial designed to study combination therapy for invasive aspergillosis, patients with hematologic malignancies and/or allogeneic HSCT were randomized at study entry to receive initial treatment with a combination of voriconazole and anidulafungin or voriconazole monotherapy.[A1] Combination therapy was associated with reduced all-cause mortality at 6 weeks compared with voriconazole monotherapy, but this difference did not reach statistical superiority. In patients with probable invasive aspergillosis, combination therapy was associated with a significant survival benefit.

PREVENTION

Several strategies may be used for prevention of invasive aspergillosis in immunocompromised patients: primary prophylaxis, empirical therapy, and secondary prophylaxis. Posaconazole is licensed for prophylaxis of invasive aspergillosis in patients with hematologic malignancies and in HSCT recipients.[A2] This recommendation in hematologic malignancies is based on a randomized clinical trial in patients undergoing chemotherapy for acute myelogenous leukemia or myelodysplasia. Posaconazole significantly prevented invasive fungal infections more effectively than did either fluconazole or itraconazole and improved overall survival. There were, however, more adverse events with posaconazole. Voriconazole also is used for this indication but with less evidence.[A3] A multicenter, randomized, double-blind trial comparing fluconazole versus voriconazole in HSCT recipients for the prevention of mycoses found nonsignificant trends of fewer *Aspergillus* infections with voriconazole. The study also demonstrated that in the context of intensive monitoring and structured empirical antifungal therapy, 6-month fungal-free survival did not differ in allogeneic HSCT recipients given prophylactic fluconazole or voriconazole. Empirical antifungal therapy provides early treatment for persistently febrile immunocompromised patients and systemic prophylaxis for high-risk hosts with or without pulmonary infiltrates. Liposomal amphotericin B (D-AmB), caspofungin, and voriconazole have been used for this strategy. Secondary prophylaxis of invasive aspergillosis with voriconazole is used for patients with a history of previous aspergillosis who are scheduled for a subsequent cycle of immunosuppression that may increase the risk for recurrence.

Reduction of exposure to airborne conidia, such as by HEPA filtration of hospital air, avoiding activities that increase conidial aerosols (room maintenance, dust exposures, and contaminated materials (e.g., potted plants), as well as providing clean water distribution systems, may reduce acquisition of *Aspergillus* by immunosuppressed or neutropenic patients.

For patients with allergic forms of aspergillosis, use of corticosteroids and itraconazole, alone or in combination, may prevent debilitating exacerbations. The toxicity of chronic administration of prednisone warrants strategies for intermittent administration of prednisone or corticosteroid-sparing use of itraconazole.

PROGNOSIS

Untreated invasive aspergillosis is associated with severe morbidity and high mortality in immunocompromised patients. Prognosis is improved by both early initiation of antifungal therapy, reversal of immunosuppression, and successful treatment of the underlying primary disease. For patients with chronic aspergillosis, multidisciplinary specialized supportive care may improve outcome and quality of life.

Grade A References

A1. Cornely OA, Maertens J, Bresnik M, et al. Liposomal amphotericin B as initial therapy for invasive mold infection: a randomized trial comparing a high-loading dose regimen with standard dosing (AmBiLoad trial). *Clin Infect Dis*. 2007;44:1289-1297.

A2. Marr KA, Schlamm HT, Herbrecht R, et al. Combination antifungal therapy for invasive aspergillosis: a randomized trial. *Ann Intern Med*. 2015;162:81-89.

A3. Cornely OA, Maertens J, Winston DJ, et al. Posaconazole compared with fluconazole or itraconazole prophylaxis in high-risk neutropenic patients receiving chemotherapy. *N Engl J Med*. 2007;356:348-359.

A4. Wingard JR, Carter SL, Walsh TJ, et al. Randomized, double blind trial of fluconazole vs. voriconazole for the prevention of invasive fungal disease after allogeneic hematopoietic cell transplantation. *Blood*. 2010;116:5111-5118.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Segal BH. Aspergillosis. *N Engl J Med*. 2009;360:1870-1884.
2. Alangaden GJ. Nosocomial fungal infections: epidemiology, infection control, and prevention. *Infect Dis Clin North Am*. 2011;25:201-225.
3. Cunha C, Aversa F, Lacera JF, et al. Genetic PTX3 deficiency and aspergillosis in stem-cell transplantation. *N Engl J Med*. 2014;370:421-432.
4. Kosmidis C, Denning DW. The clinical spectrum of pulmonary aspergillosis. *Thorax*. 2015;3: 270-277.
5. McCarthy M, Rosengart A, Schuetz AN, et al. Mold infections of the central nervous system. *N Engl J Med*. 2014;371:150-160.
6. Kalokhe AS, Rouphael N, El Chami MF, et al. Aspergillus endocarditis: a review of the literature. *Int J Infect Dis*. 2010;14:e1040-e1047.
7. Gamaletsou MN, Rammaert B, Bueno MA, et al. Epidemiology, clinical manifestations, management and outcome of 179 cases of *Aspergillus* osteomyelitis. *J Infect*. 2014;68:478-493.
8. Agarwal R, Chakrabarti A. Allergic bronchopulmonary aspergillosis in asthma: epidemiological, clinical and therapeutic issues. *Future Microbiol*. 2013;8:1463-1474.
9. Desoubeaux G, Bailly E, Chandenier J. Diagnosis of invasive pulmonary aspergillosis: updates and recommendations. *Med Mal Infect*. 2014;44:89-101.
10. Schuetz AN. Invasive fungal infections: biomarkers and molecular approaches to diagnosis. *Clin Lab Med*. 2013;33:505-525.
11. Walsh TJ, Anaissie EJ, Denning DW, et al. Treatment of aspergillosis: clinical practice guidelines of the Infectious Diseases Society of America (IDSA). *Clin Infect Dis*. 2008;46:327-360.
12. Koulenti D, Garnacho-Montero J, Blot S. Approach to invasive pulmonary aspergillosis in critically ill patients. *Curr Opin Infect Dis*. 2014;27:174-183.

REVIEW QUESTIONS

1. Which of the following statements is correct regarding the treatment of invasive aspergillosis?

A. Reversal of immunosuppression is critical for successful therapy.
B. Combination antifungal therapy, including a triazole, is the treatment of choice because monotherapy is usually not effective.
C. There are no biomarkers that can be clinically used for following response to treatment.
D. Surgical treatment is rarely needed when long-term, uninterrupted antifungal treatment can be administered.
E. Systemic antifungal therapy is sufficient to treat ocular involvement.

Answer: A Invasive aspergillosis characteristically occurs in immunocompromised hosts. In these individuals, reversal of immunosuppression (e.g., recovery from neutropenia, discontinuation of corticosteroids) is a critical factor in successful management. The antifungal treatment of choice for invasive aspergillosis, according to Infectious Diseases Society of America clinical practice guidelines, is monotherapy with voriconazole; the alternate drug is liposomal amphotericin B. Second-line and salvage treatments likewise involve monotherapy. Surgical treatment is an important adjunctive component of primary therapy for several forms of invasive aspergillosis (e.g., endocarditis, osteomyelitis, epidural abscess, soft tissue infection, and others). Intravitreal antifungal therapy for endophthalmitis and topical irrigation with an antifungal agent for keratitis can provide high concentrations of these agents to ocular compartments that may not be reached by systemic therapy. Galactomannan and (1→3)-β-D-glucan in serum or bronchoalveolar lavage are useful biomarkers.

2. Which of the following statements is correct regarding treatment of different forms of chronic aspergillosis?

A. Long-term voriconazole has a high success rate in curing aspergilloma.
B. Bronchial artery embolization is an effective alternative for curing aspergilloma.
C. Response to antifungal therapy for chronic necrotizing pulmonary aspergillosis is conveniently monitored by serial computed tomography (CT).
D. Surgical resection is recommended for definitive therapy of solitary aspergilloma and chronic cavitary pulmonary aspergillosis.
E. Surgical resection plays an important role in treating disease complicated by hemoptysis.

Answer: E Surgical resection may have an important role in treating patients with recurrent and severe hemoptysis. However, in general, surgical resection in patients with solitary aspergilloma or chronic necrotizing pulmonary aspergillosis is limited by complications of bronchopleural fistula development, *Aspergillus* infection of the pleural space, and potential further worsening of already compromised pulmonary function. Medical therapy has limited benefit in the treatment of aspergilloma. Although some benefit may be derived from extended use of an antifungal triazole, lifelong commitment to it should be balanced against the natural history of spontaneous resolution in about 10% of patients. Bronchial artery embolization (or transthoracic direct intracavitary instillation of antifungal agents) for aspergilloma has only transient benefits but substantial risk. CT or radiologic assessment of response to antifungal therapy in chronic necrotizing pulmonary aspergillosis is difficult because of the underlying chronic lung disease.

3. Allergic bronchopulmonary aspergillosis (ABPA) occurs most frequently in patients with which of the following?

A. Intravenous drug abuse
B. Immunocompromised status
C. Asthma
D. α_1-Antitrypsin deficiency
E. Occupational exposure (e.g., malt workers)

Answer: C ABPA characteristically develops in patients with a history of chronic asthma or cystic fibrosis (not α_1-antitrypsin deficiency). ABPA is characterized by episodic airway obstruction, fever, eosinophilia, mucous plugs containing hyphae, transient infiltrates and parallel "tramline" or ring markings on chest radiographs, bronchiectasis, and upper lobe contraction. Extrinsic allergic alveolitis, an unusual allergic form of *Aspergillus* lung disease, is associated with *Aspergillus clavatus* in malt workers. In contrast, invasive aspergillosis is seen in immunocompromised patients and those with intravenous drug abuse.

340

MUCORMYCOSIS

D.P. KONTOYIANNIS

DEFINITION

Mucormycosis is the accurate unifying term used to describe infections caused by fungi belonging to the order Mucorales. Zygomycosis, an alternative term used to describe these life-threatening infections, has become less accurate based on a recent taxonomic reclassification (using molecular methods) that abolished Zygomycetes as a class (and placed the order Mucorales in the subphylum Mucormycotina).[1] Mucorales typically cause aggressive, acute-onset, frequently fatal angioinvasive infections, especially in immunosuppressed hosts.

EPIDEMIOLOGY

Mucorales fungi are distributed worldwide and found in decaying organic substrates. The true incidence of mucormycosis is not known and probably is underestimated because of difficulties in antemortem diagnosis. The relative frequency of Mucorales families causing infection differs. In a recent review of more than 900 reported cases, the most common microbiologically confirmed infecting species were *Rhizopus* (47%), *Mucor* (18%), *Cunninghamella bertholletiae* (7%), *Apophysomyces elegans* (5%), *Absidia* (5%), *Saksenaea* (5%), and *Rhizomucor pusillus* (4%). Some Mucorales causing infection have specific geographic and host associations. For example, the thermophilic Mucorales *Saksenaea vasiformis elegans* have specific geographic distributions, as recently shown in victims of combat-related injuries from Afghanistan who developed necrotizing soft tissue infections.[2] Also, major natural disasters have been associated with rapidly progressing necrotizing soft tissue infections by infrequently isolated species, such as those caused by *Apophysomyces elegans* in the Joplin tornado victims in 2011.[3]

The classic risk factors for mucormycosis include hematologic malignancy, hematopoietic stem cell or solid organ transplantation, poorly controlled diabetes mellitus, chronic acidemia, prematurity, profound chronic debilitation, trauma, burns, and very rarely intravenous drug use.[4] Nosocomial cutaneous infections can develop at surgical wound and intravenous catheter insertion sites. Finally, breakthrough mucormycosis has been increasingly observed in patients with leukemia and in recipients of hematopoietic stem cell transplants receiving *Aspergillus*-active drugs such as voriconazole (which has no anti-Mucorales activity). This association has been a topic of debate.

PATHOBIOLOGY

Mucorales species are saprophytic, rapidly growing fungi. Angioinvasive growth results in infarction and necrosis of surrounding tissue, which is the hallmark of mucormycosis. The major modes of transmission are inhalation, ingestion, and cutaneous inoculation, with inhalation of spores from environmental sources being the most common. Cutaneous or percutaneous transmission occurs with traumatic disruption of skin barriers, and it is the most important mode of transmission in immunocompetent hosts. Gastrointestinal acquisition, while less common, has occurred in patients with repeated ingestion of spores during severe malnutrition, non-nutritional substances (pica), contaminated herbal/homeopathic products, or allopurinol tablets.

Host immunity in healthy hosts prevents germination of fungal spores unless the inoculum is heavy.[5] To establish invasive infection, spores must overcome both innate and adaptive immune responses to germinate into hyphae. Defects in phagocytic activity caused by insufficient numbers (i.e., neutropenia) and functional defects caused by glucocorticoids, hyperglycemia, and/or acidosis allow unimpeded proliferation of fungi because of the absence of coordinated, effective host responses. Unsurprisingly, mucormycosis is often disseminated in severely immunosuppressed patients, with high

mortality rates. However, lymphopenia may not be as critical in patients with acquired immunodeficiency syndrome.

Free iron is an essential component of the pathogenesis of mucormycosis, as suggested by the predisposition of patients with iron overload and ketoacidosis to such infections. These patients often receive the iron chelator deferoxamine. Both iron overload and use of deferoxamine (which is also used to treat aluminum overload in dialysis recipients) are risk factors for angioinvasive mucormycosis. *Rhizopus oryzae* can utilize deferoxamine as a xenosiderophore to form a ferrioxamine complex and to obtain more iron for use. Reassuringly, the newer iron chelators agents (e.g., deferasirox) are not associated with increased risk for mucormycosis; on the contrary, in preclinical models, deferasirox has exhibited direct fungicidal effects against Mucorales via iron starvation.

Historically, poorly controlled diabetes mellitus (types 1 and 2) has been a major predisposing factor, reported in 36 to 88% of all cases of mucormycosis. In particular, diabetic patients with ketoacidosis are susceptible to mucormycosis. Normal human serum cannot support the growth of *R. oryzae*, whereas serum in diabetic patients can do so. Acidosis disrupts the normal inhibitory activity of serum by attenuating the ability of transferrin to bind iron from the fungus. In addition, quantitative and qualitative neutrophil and phagocytic cell dysfunction occurs in diabetic patients with ketoacidosis and may play a role in the pathogenesis.

CLINICAL MANIFESTATIONS

The clinical presentation of mucormycosis depends on the host's underlying immune and medical condition. Hence, pulmonary mucormycosis is most common in neutropenic or corticosteroid-treated patients (e.g., hematopoietic stem cell and solid organ transplant recipients). In contrast, rhino-orbital or rhinocerebral mucormycosis is the characteristic presentation in patients with diabetic ketoacidosis. Finally, cutaneous mucormycosis in both immunocompetent and immunocompromised hosts is typically seen following local trauma or burns resulting in breakdown of skin integrity and/or subcutaneous tissue injuries. Infectious syndromes associated with Mucorales are grouped based on clinical presentation into one of six categories: (1) rhinocerebral, (2) pulmonary, (3) cutaneous, (4) gastrointestinal, (5) disseminated, and (6) unusual presentations, as follows:

Rhinocerebral Mucormycosis

Rhinocerebral mucormycosis describes an infection originating in the paranasal sinuses after inhalation of Mucorales spores and extending to the orbit (sino-orbital) or brain (rhinocerebral), particularly in patients with diabetic ketoacidosis or those with profound neutropenia.[6] Rhinocerebral mucormycosis is the most common manifestation. Early signs and symptoms of sinus invasion may be indistinguishable from common causes of sinusitis. Common symptoms include sinus pain, congestion, headache, mouth pain, otologic symptoms, and hypo-osmia/anosmia. Involved tissues become red and then violaceous and, finally, black with thrombosis, and tissue necrosis. Necrotic eschar of the nasal cavity and turbinates, facial lesions, and exophytic or necrotic lesions of the hard palate are signs of extensive, rapidly progressing infection. A painful black eschar on the palate or nasal mucosa is a classic diagnostic but late sign. Absence of this finding does not rule out rhinocerebral infection, as necrotic nasal or palate lesions occur in only 50% of patients within 3 days of onset of infection. Extension into the periorbital region is not uncommon at presentation. Signs and symptoms of periorbital and orbital involvement include periorbital swelling, preseptal and/or orbital cellulitis, proptosis, chemosis, blurred vision or rapidly progressing external ophthalmoplegia, diplopia, eyelid gangrene, retinal detachment, and endophthalmitis. Also, patients with extensive rhino-orbital or rhinocerebral disease may present with trigeminal or other cranial nerve palsy, which is consistent with frequent histologic findings of perineural invasion. Infection can rapidly progress through the cavernous sinuses into the central nervous system, resulting in cavernous sinus and internal carotid artery thrombosis. A bloody nasal discharge may be the only sign indicating that the infection has invaded through the nasal turbinates and into the brain. Patients with advanced infection may have cranial neuropathies and/or altered consciousness; bone destruction; retinal artery, internal carotid artery, cavernous, and less often, sagittal sinus, thrombosis; frontal lobe necrosis; epidural and subdural abscesses; and/or basilar artery aneurysm.

Plain films and cerebrospinal fluid findings lack sensitivity in diagnosing rhinocerebral mucormycosis. Computed tomography (CT) and magnetic resonance imaging (MRI) are more useful for revealing soft tissue involvement around the nerve sheaths and bone destruction. CT frequently shows mucosal thickening, air-fluid levels, and bony erosion. Orbital thickening may be detected earlier by using MRI. CT and MRI scans of the orbits may be unremarkable during the initial stages of mucormycosis, highlighting the importance of serial radiographic imaging in monitoring disease progression. Extraorbital muscle thickening is often the first sign of orbital involvement and should prompt empirical antifungal therapy followed by surgical exploration or biopsy.

Accurate diagnosis and prompt medical and surgical intervention are critical because of the rapid progression of the infection. Definitive diagnosis of necrotic lesions using biopsy and rapid histologic assessment of frozen sections should be performed as soon as possible because it directly impacts outcome. In a review of 929 documented mucormycosis cases, the mortality rate was as follows: 62% in rhinocerebral mucormycosis, 24% in sino-orbital involvement, and 16% in isolated sinus disease. Isolated sinusitis is curable by following timely surgical intervention and systemic antifungal therapy.

Pulmonary Mucormycosis

The clinical manifestations of pulmonary mucormycosis are indistinguishable from those of invasive pulmonary aspergillosis. Patients may present with fever refractory to broad-spectrum antibiotics, nonproductive cough, and progressive dyspnea. Less commonly, pleuritic chest pain, hemoptysis, and pleural effusion are seen. If the major pulmonary blood vessels are invaded by fungal hyphae, massive, potentially fatal hemoptysis can occur. Pulmonary mucormycosis can progress and invade adjacent organs via traverse tissue planes, including the diaphragm, chest wall, and pleura (Fig. 340-1). Clues for distinguishing pulmonary mucormycosis from invasive pulmonary aspergillosis include the presence of pansinusitis, a history of prophylaxis with antifungals against *Aspergillus* but not Mucorales (e.g., voriconazole, echinocandins), and possibly continual absence of detectable *Aspergillus galactomannan* antigen in serum. In rare circumstances, pulmonary mucormycosis can present as an endobronchial or tracheal lesion with a less fulminant course, especially in diabetics. Endobronchial mucormycosis may cause airway obstruction or erosion of major pulmonary blood vessels and fatal hemoptysis. In more immunocompetent hosts, pulmonary mucormycosis may present with more atypical, slowly progressing forms. Like *Aspergillus* species, Mucorales can form mycetomas in preexisting lung cavities and cause slowly necrotizing pneumonia and hypersensitivity syndromes. Investigators have also implicated *Rhizopus* species in allergic alveolitis among farm workers or sawmill workers (wood-trimmer's disease).

Because the first-line antifungal typically used for aspergillosis is voriconazole, which lacks activity against Mucorales, failure to achieve a timely diagnosis of pulmonary mucormycosis and delayed antifungal therapy (e.g., amphotericin B) rapidly worsens outcome.

Skin and Soft Tissue Mucormycosis

Cutaneous mucormycosis typically occurs in victims of severe skin or muscular injury. It starts as erythema and a skin induration at a puncture site and progresses to necrosis with a black eschar. Cutaneous infections can quickly extend into the deep fascia and muscle layers. Necrotizing fasciitis is rare and has a poor prognosis. Neutropenic patients in particular are susceptible to lymphatic and blood vessel invasion, infarction, and necrosis with eventual dissemination. Interestingly, the skin appears to be a less common site of secondary involvement with disseminated mucormycosis than of infections by other hyaline molds such as *Fusarium* or *Scedosporium* species. Even so, skin lesions in patients with suspected mucormycosis should raise concerns about disseminated disease and prompt, careful clinical work-up. Because the differential diagnosis of necrotic skin lesions is broad, especially in neutropenic patients, biopsy specimens should be obtained from the center of the lesion down to the subcutaneous fat. Excision and wide débridement of cutaneous lesions, coupled with systemic antifungal therapy and, on occasion, hyperbaric oxygen therapy, can further reduce mortality rates.

Gastrointestinal Mucormycosis

Primary gastrointestinal mucormycosis is rare and can present as necrotizing enterocolitis and involve any part of the alimentary system with mortality rates of more than 85%. It occurs primarily in malnourished patients and premature infants, in which the stomach is the most commonly affected site, followed by the colon and ileum. The liver, spleen, and pancreas also can be involved. Physicians have described liver abscesses following ingestion of herbal products contaminated by *Mucor indicus*. Fungi can invade the bowel wall and blood vessels, resulting in bowel perforation, peritonitis, and massive gastrointestinal hemorrhage. In neutropenic patients, seeding of the gastrointestinal tract is likely more common than previously thought because 75% of gastrointestinal mucormycoses are diagnosed postmortem. Symptoms and

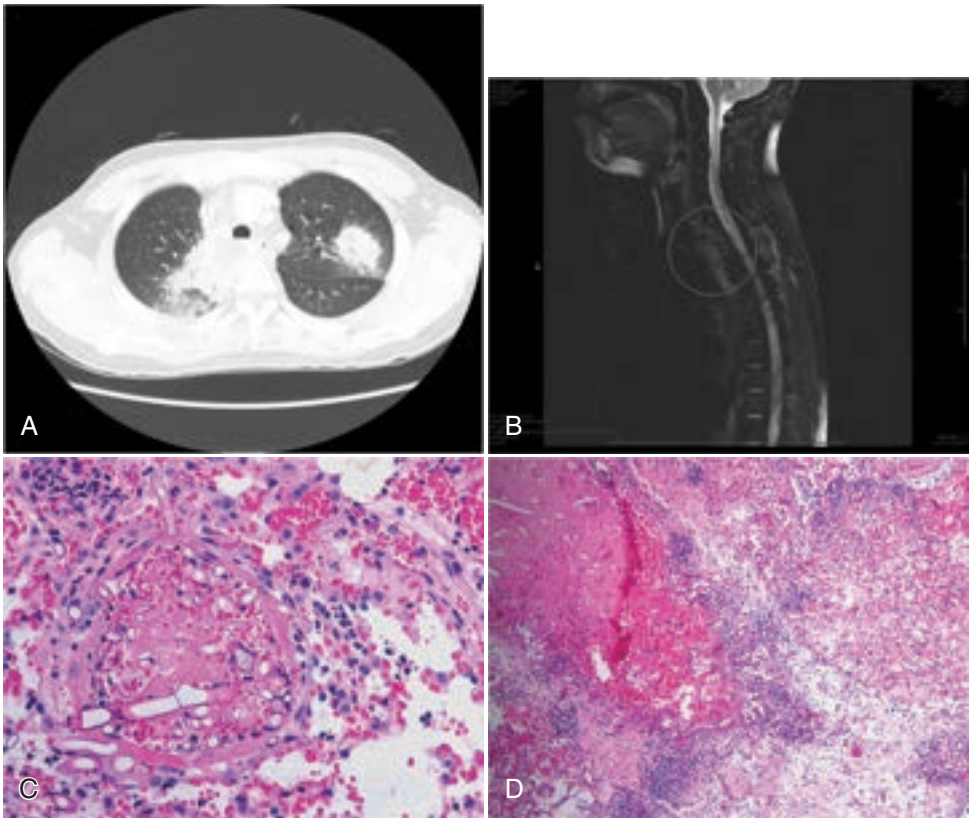

FIGURE 340-1. Extensive, progressive pulmonary mucormycosis in a patient with active leukemia and neutropenia. Characteristic extension of the infection across tissue planes to trachea (producing a fistula) and mediastinum (**A**) and adjacent spine (**B**) are shown. The histopathologic characteristics of profound necrosis and hemorrhage and pauciseptate, broad-based, ribbon-like Mucorales are also shown (**C** and **D**). Culture of a tissue biopsy specimen remained negative. The patient died 3 weeks after diagnosis, despite aggressive use of a high-dose lipid formulation of amphotericin B (AMB) and adjunct immune therapy.

signs of gastrointestinal mucormycosis include fever, abdominal distention, nausea, vomiting, abdominal pain, diarrhea, melena, hematemesis, hematochezia, and masslike appendiceal and ileal lesions.

Disseminated Mucormycosis

Disseminated mucormycosis is rarely apparent antemortem. Severely immunosuppressed patients (e.g., those with prolonged and profound neutropenia, allogeneic stem cell transplant recipients with severe graft-versus-host disease) and patients receiving deferoxamine are at the highest risk. Symptoms vary depending on the site of dissemination and degree of vascular infarction of the affected organs. The most common organ as source of dissemination is the lung, and the most common site of spread is the brain. Diagnosis of disseminated mucormycosis is challenging and requires a high level of suspicion because the infection may present as an unexpected acute vascular event. Biopsy of suspected sites is critical because of the low yield of blood cultures and suboptimal recovery of the fungus from respiratory specimens. Without appropriate timely treatment, virtually all patients with disseminated mucormycosis die.

Rare Clinical Presentations of Mucormycosis

Mucormycosis has protean manifestations that involve any organ. Authors have reported isolated cases of tracheal, mediastinum, bone, heart, kidney, otitis externa, and corneal involvement. More recently, there have been reports of renal mucormycosis in patients with intravenous drug abuse and/or those receiving corticosteroids. Cerebral mucormycosis often presenting as brain abscess involving the basal ganglia, and in conjunction with infective endocarditis, has been typically observed in patients using illicit intravenous drugs. Reports of peritonitis in patients undergoing continuous ambulatory peritoneal dialysis have been rare. In all cases of device-related mucormycosis, prompt removal of the device and several weeks of systemic antifungal therapy are essential for resolution of the infection.

DIAGNOSIS

The clinical signs and symptoms of mucormycosis are nonspecific. Therefore, a high level of suspicion in susceptible patient populations is of paramount importance.[7] Biopsy analysis and culture from sterile sites remain critical. Tissue swabs and cultures of sputum, sinus secretions, and bronchoalveolar lavage fluid are usually nondiagnostic. For example, fungal contamination of clinical specimens occurs because the small size of sporangiospores (approximately 6 μm in width) allows easy dispersion via the airborne route. Particles of this size may remain airborne even with very slight movements in air and contaminate clinical samples. Therefore, growth in culture may not represent clinically significant invasive mucormycosis. However, the value of Mucorales-positive cultures (especially repetitive cultures) as an important indication of infection in immunocompromised patients is quite high. The site of infection has a major impact on the likelihood of histopathologic diagnosis. With their ease of accessibility, sinuses are the major site of definite infection.

Histopathology

A variety of stains, including hematoxylin and eosin, Grocott-Gomori methenamine-silver nitrate, and periodic acid–Schiff, reveal characteristic hyphal elements in tissue. Histopathologic examination of infected tissue typically shows characteristic broad (3 to 25 μm in diameter), thin-walled, primarily aseptate hyphae; focal bulbous dilation; and nondichotomous irregular branching at occasional right angles accompanying tissue necrosis and fungal angioinvasion (see Fig. 340-1). Perineural invasion is found in 90% of tissues containing nerves. The inflammatory responses to mucormycosis can range from neutrophilic, granulomatous, and/or pyogranulomatous to minimal inflammation with hemorrhage. Also, fungal hyphae can be examined directly using a potassium hydroxide preparation of a tissue specimen or bronchial alveolar lavage fluid. Although contamination is always a possibility, discovery of fungal elements in a specimen obtained from an immunocompromised host is considered significant. Treatment with fluorescent stains such as Calcofluor White and Blankofluor may enhance detection of hyphal elements during microscopic examination. Improved staining procedures may be important when the number of organisms is small or the amount of tissue is limited.

Culture

Mucorales fungi characteristically produce large, ribbon-like hyphae with irregular diameters and only occasionally septa, resulting in characterization of these organisms primarily as aseptate fungi. Identification can be

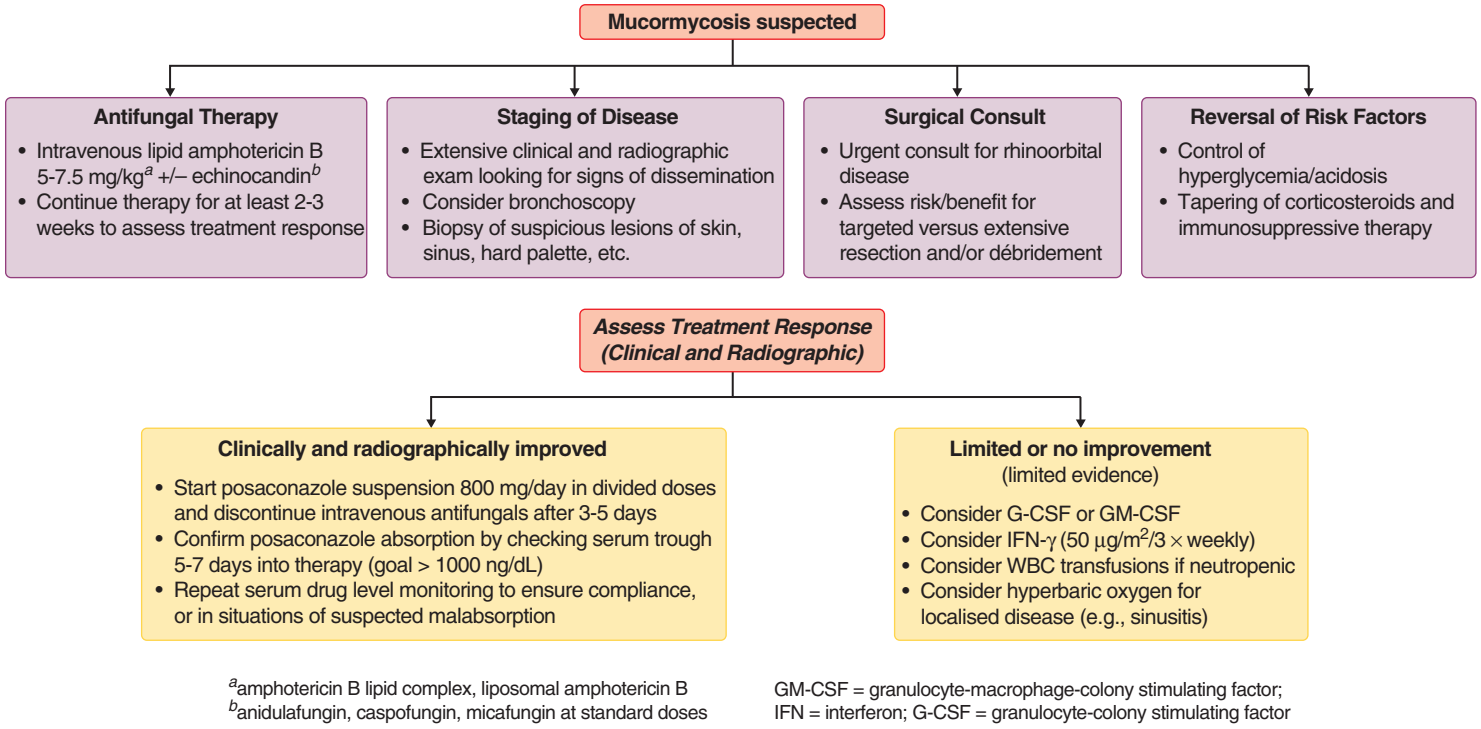

[a]amphotericin B lipid complex, liposomal amphotericin B
[b]anidulafungin, caspofungin, micafungin at standard doses

GM-CSF = granulocyte-macrophage-colony stimulating factor; IFN = interferon; G-CSF = granulocyte-colony stimulating factor

FIGURE 340-2. Diagram of the management approach to patients with suspected mucormycosis.

confirmed by observing the characteristic saclike fruiting structures (sporangia), which produce internally spherical yellow or brown spores (sporangiospores). Spores range from 3 to 11 μm in diameter and are easily aerosolized. Blood cultures are rarely positive for these pathogens despite their angioinvasive nature. Paradoxically, even when fungal hyphae are seen in histopathologic analysis, fungal cultures may not be positive because of the friability of nonseptated hyphae, making them more susceptible to damage during tissue manipulation. However, collection of several proper clinical specimens is important. Recovery of Mucorales from tissue can be improved by mincing (not homogenizing) tissue specimens and using culture techniques that simulate in vivo fungal growth, including incubation at 35° to 37°C under relatively semianaerobic conditions.

Morphologic identification of Mucorales requires their cultivation to examine reproductive fruiting structures. Most of these fungi grow rapidly on most fungal media (e.g., Sabouraud dextrose agar) when incubated at 25° to 30° C. These fungi are sensitive to the protein inhibitor cycloheximide, and addition of this agent to fungal media may not ensure optimal recovery. Morphologic features alone, especially when assessed by individuals with expertise in fungal identification, can provide a high level of accuracy comparable to that of molecular methods.

Data on the antifungal susceptibility of Mucorales species. are limited, and minimal inhibitory concentration (MIC) testing is rarely available outside research or research laboratories. The MIC end points for these rapidly growing fungi are inconsistent, not standardized, and at times difficult to interpret. Because interpretive MIC break points for Mucorales have yet to be defined, the correlation between clinical responses and MIC values is uncertain. Mucorales are resistant to many antifungals, including flucytosine, ketoconazole, fluconazole, voriconazole, and the echinocandins. Also, they have variable susceptibility to itraconazole. Amphotericin B (AMB) and posaconazole, a new triazole, are the most active agents in vitro, although their activities differ among different Mucorales families. The activity of antifungal combinations against these fungi has yet to be proved in vivo.

The importance of early differentiation of Mucorales from more common opportunistic molds such as *Aspergillus* species has generated considerable interest in development of culture- or histopathology-independent diagnostic tests such as detection of specific antigens or nucleic acids using polymerase chain reaction or in situ hybridization techniques. Molecular techniques for detecting Mucorales are few, not widely available, and investigational. This is an important unmet need for the management of mucormycosis.

TREATMENT

Approach to the management of suspected mucormycosis and assessment of treatment response are shown in Figure 340-2. Successful treatment of mucormycosis relies on a multifaceted strategy that includes (1) aggressive attempts at diagnosis and rapid initiation of effective antifungal therapy and (2) extensive surgical débridement, and (3) rapid control of underlying medical conditions.[8]

Again, early diagnosis is critical to the outcome. Small focal lesions can be surgically resected before they progress to involve critical structures or distal organs. Patients often have indolent clinical presentations until extensive invasion or dissemination of the infection occurs.[9]

Antifungal Therapy

Delayed administration of systemic antifungal therapy increases the probability of patient death. Most of the knowledge about the activity of currently used antifungals comes from small case series, anecdotes, and animal models of infection. Therefore, the optimal treatment approach is uncertain.[10] Most of the clinical experience has been with AMB. Previously, the recommended antifungal therapy for mucormycosis included AMB deoxycholate at the maximum tolerated dosage, usually 1.0 to 1.5 mg/kg/day. The nephrotoxic and systemic toxic effects of regular AMB led to the development of the lipid formulations of AMB (liposomal AMB, AMB lipid complex, AMB colloidal dispersion). These agents are less nephrotoxic than regular AMB and can be given at higher doses (e.g., 5 to 10 mg/kg per day). Lipid formulations of AMB are now considered the drugs of choice for mucormycosis. Furthermore, use of percutaneous or aerosolized AMB in conjunction with concomitant systemic therapy has been successful in selected patients with pulmonary mucormycosis. Topical therapy with AMB as well as other polyenes (natamycin) may be effective against primary cutaneous and ocular mucormycosis. Treatment of mucormycosis with AMB-based combinations has been successful in small retrospective case series. In particular, a benefit has been suggested for echinocandin-liposomal AMB combination in 41 diabetic patients with rhino-orbital mucormycosis compared with the ones who received amphotericin B lipid complex (ABLC) or liposomal AMB alone. This benefit was most pronounced in patients with cerebral involvement.

Although azoles traditionally have been inactive against Mucorales, the new broad-spectrum triazole posaconazole demonstrated promising activity. Among open-label studies and retrospective surveys evaluating posaconazole suspension as salvage therapy (800 mg/day) in patients with refractory mucormycosis, the agent showed a response rate approaching 70%. Furthermore, posaconazole has been well tolerated. Determining whether posaconazole alone or combined with a lipid formulation of AMB or other agent (e.g., deferasirox) is of value requires further study. Posaconazole has limitations because absorption of the oral suspension is suboptimal in patients with mucositis,

severe diarrhea, acid suppression therapy, or poor oral intake. Absorption of oral posaconazole is maximized when administered with high-fat foods in separate doses (four times daily). Finally, steady-state plasma concentrations of posaconazole are not reached until around 1 week of therapy. The new formula of posaconazole (posaconazole tablets 300 mg/daily) has not been studied adequately in mucormycosis.

The duration of antifungal therapy should be determined on an individual basis. Near normalization of radiographic imaging, negative follow-up biopsy specimens, and cultures from the affected site, as well as recovery from immunosuppression, are important indicators for stopping antifungal therapy.

Surgery

Surgical débridement of cutaneous lesions is crucial and must be done without delay because of the aggressively invasive nature of mucormycosis. A coordinated effort among all subspecialties involved (surgery, infectious diseases, head and neck, ophthalmology, pathology, clinical microbiology, and plastic surgery) is crucial, and the internist can play a vital role coordinating it.

Repeated removal of necrotic tissue or aggressive surgical measures such as enucleation of the eye may be required for control of the infection. Decisions regarding the extent of débridement are often made at the bedside. A CT or MRI scan before surgery and intraoperative frozen section analysis help determine the extent of tissue and tissue margin involvement. Low platelet counts, as may be seen in patients with underlying hematologic malignancies, must be corrected with transfusions before surgical intervention. Unfortunately, bleeding problems can limit surgical options. Surgery in conjunction with systemic antifungal therapy has been shown to significantly improve survival rates.

Management of Comorbidity and Adjunct Treatments

Adjunct measures have been proposed to improve host immunity, tissue viability, and impeding fungal proliferation. Rapid correction of underlying conditions, such as control of hyperglycemia, reversal of ketoacidosis, rapid tapering of glucocorticoid therapy, and discontinuation of deferoxamine-based treatment, can influence outcomes. Hyperbaric oxygen is a beneficial adjunct therapy for mucormycosis, particularly in diabetic patients with rhinocerebral disease. Specifically, the increased oxygen pressure achieved seems to improve neutrophil activity and oxidative killing by polyene antifungals. Also, high concentrations of oxygen can inhibit growth of the organism in vitro and improve the rate of wound healing by increasing the release of tissue growth factors. However, this treatment has not been studied vigorously to determine efficacy and cannot be routinely recommended. Investigators have proposed several immune augmentation strategies as adjunct therapy, including administration of cytokines (e.g., granulocyte colony-stimulating factor [G-CSF], interferon). In refractory neutropenic patients, granulocyte transfusion may be beneficial until granulocyte recovery. These adjunct measures, although promising, are yet to be studied sufficiently. Finally, the new iron chelator deferasirox has been considered as an adjunct antifungal agent based on preclinical studies and very limited human experience with patients with refractory mucormycosis. Results of the small randomized, double-blinded DEFEAT Mucor trial were published in 2012.[A1] Twenty patients with proven or probable mucormycosis were randomized to treatment with liposomal amphotericin B plus deferasirox (20 mg/kg per day for 14 days) or liposomal amphotericin B plus placebo. Although reported adverse events were similar between the two study groups, significantly higher mortality rates were found in patients randomized to receive deferasirox at 30 (45% vs. 11%) and 90 days (82% vs. 22%, $P = .01$). However, patients in the deferasirox arm were more likely than patients in the placebo arm to have active malignancy, neutropenia, and/or corticosteroid therapy, and less likely to have received additional antifungals, making the results of this pilot trial less conclusive. Nevertheless, currently available data do not support a role for initial deferasirox therapy for mucormycosis. Further knowledge of the unique virulence attributes of Mucorales based on genomic analysis might aid the development on novel therapeutic targets.

Grade A Reference

A1. Spellberg B, Ibrahim AS, Chin-Hong PV, et al. The Deferasirox-AmBisome Therapy for Mucormycosis (DEFEAT Mucor) study: a randomized, double-blinded, placebo-controlled trial. *J Antimicrob Chemother*. 2012;67:715-722.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Lewis RE, Kontoyiannis DP. Epidemiology and treatment of mucormycosis. *Future Microbiol*. 2013;8:1163-1175.
2. Warkentien T, Rodriguez C, Lloyd B, et al. Infectious Disease Clinical Research Program Trauma Infectious Disease Outcomes Study Group. Invasive mold infections following combat-related injuries. *Clin Infect Dis*. 2013;121:2385-2392.
3. Neblett Fanfair R, Benedict K, Bos J, et al. Necrotizing cutaneous mucormycosis after a tornado in Joplin, Missouri, in 2011. *N Engl J Med*. 2012;367:2214-2225.
4. Ibrahim AS, Kontoyiannis DP. Update on mucormycosis pathogenesis. *Curr Opin Infect Dis*. 2013;26:508-515.
5. Millon L, Larosa F, Lepiller Q, et al. Quantitative polymerase chain reaction detection of circulating DNA in serum for early diagnosis of mucormycosis in immunocompromised patients. *Clin Infect Dis*. 2013;56:e95-e101.
6. McCarthy M, Rosengart A, Schuetz AN, et al. Mold infections of the central nervous system. *N Engl J Med*. 2014;371:150-160.
7. Skiada A, Lanternier F, Groll AH, et al. Diagnosis and treatment of mucormycosis in patients with haematological malignancies: guidelines from the 3rd European Conference on Infections in Leukemia (ECIL 3). *Haematologica*. 2013;98:492-506.
8. Cornely OA, Arikan-Akdagli S, Dannaoui E, et al. ESCMID and ECMM joint clinical guidelines for the diagnosis and management of mucormycosis 2013. *Clin Microbiol Infect*. 2014;20(suppl 3):5-26.
9. Tacke D, Koehler P, Markiefka B, et al. Our 2014 approach to mucormyocis. *Mycoses*. 2014;57:519-524.
10. Kontoyiannis DP, Lewis RE. How I treat mucormycosis. *Blood*. 2011;118:1216-1224.

REVIEW QUESTIONS

1. What is the best way to diagnose mucormycosis?

A. Serum galactomannan
B. Serum β-D-glucan
C. Anti-Mucorales antibodies
D. Polymerase chain reaction (PCR)
E. Tissue for biopsy and culture

Answer: E Unfortunately, there are no biomarkers to diagnose mucormycosis, and this is a major unmet need for the management of the disease. Specifically, galactomannan and β-D-glucan do not detect Mucorales. PCR, although promising, remains investigational. There are no data about the value of anti-Mucorales antibodies for diagnosis. Tissue for histopathologic analysis and culture remains the only method for documenting invasive disease. (See Skiada A, Lanternier F, Groll AH, et al. Diagnosis and treatment of mucormycosis in patients with haematological malignancies: guidelines from the 3rd European Conference on Infections in Leukemia [ECIL 3]. *Haematologica.* 2013;98[4]:492-506; and Tacke D, Koehler P, Markiefka B, Cornely OA. Our 2014 approach to mucormyocis. *Mycoses.* 2014;57[9]:519-524.)

2. All the conditions below are common risk factors for mucormycosis *except* which one of the following?

A. Iron overload
B. Poorly controlled diabetes
C. Crowding conditions
D. Severe trauma of soft tissues
E. Chronic severe immune suppression

Answer: C Classic risk factors for the disease include iron overload that results in increased availability of host iron to support invasive fungal growth, significant and chronic immunosuppression such as severe cytopenia, systemic corticosteroids, crushing soft tissue injury (e.g., following severe trauma), and poorly controlled diabetes, especially in the setting of acidosis. Mucorales are extremely rare causes of catheter related infections not known to be associated with crowding conditions in the absence of trauma. (See Epidemiology.)

3. What is the most common Mucorales species causing mucormycosis?

A. *Mucor* spp
B. *Cunninghamella bertholletiae*
C. *Rhizomucor* spp
D. *Rhizopus* spp
E. *Apophysomyces* spp

Answer: D Although there is notable geographic and host-related variability of infecting Mucorales, on a global scale *Rhizopus* species are the most common cause of mucormycosis, accounting for nearly 50% of such cases. (See Epidemiology.)

4. Which of the following azoles has the most potent activity against Mucorales?

A. Fluconazole
B. Posaconazole
C. Voriconazole
D. Itraconazole
E. Ketoconazole

Answer: B Mucorales are innately resistant to most antifungal drugs. Specifically, the only agents that have activity are amphotericin B and the triazole posaconazole. (See Skiada A, Lanternier F, Groll AH, et al. Diagnosis and treatment of mucormycosis in patients with haematological malignancies: guidelines from the 3rd European Conference on Infections in Leukemia [ECIL 3]. *Haematologica.* 2013;98[4]:492-506; Tacke D, Koehler P, Markiefka B, Cornely OA. Our 2014 approach to mucormyocis. *Mycoses.* 2014;57[9]:519-524; and Kontoyiannis DP, Lewis RE. How I treat mucormycosis. *Blood.* 2011;118[5]:1216-1224. Also see section Treatment.)

341

PNEUMOCYSTIS PNEUMONIA

JOSEPH A. KOVACS

DEFINITION

Pneumocystis jirovecii is a fungus that causes pneumonia almost exclusively in immunodeficient patients. In the 1950s, it was recognized as the cause of an epidemic interstitial plasma cell pneumonia that occurred primarily in premature and malnourished infants. During the 1960s and 1970s, *Pneumocystis* pneumonia (PCP) was seen with increasing frequency in immunodeficient patients, especially those receiving immunosuppressive chemotherapy. In the early 1980s, an epidemic of PCP in previously healthy adults heralded the onset of the HIV/AIDS epidemic; in the United States and in many other parts of the world, PCP has been the most common life-threatening opportunistic infection in this population. Although the frequency of PCP decreased in HIV-infected patients first in association with the widespread use of anti-*Pneumocystis* prophylaxis and later with the introduction of effective combination antiretroviral therapy for HIV/AIDS, it continues to be seen with regularity in HIV-infected and other immunodeficient patients.[1] Over the past decade, there has been a marked increase in outbreaks of PCP among renal transplant recipients, especially in Europe and Australia.

The Pathogen

Pneumocystis, which was long thought to be a protozoan, has now been definitively classified as an ascomycete fungus, based primarily on molecular studies demonstrating that a large number of genes have a high level of homology to other fungi rather than to protozoa. Molecular as well as antigenic studies have further demonstrated that the genus *Pneumocystis* includes a group of closely related organisms that are unique species, each of which can infect only a single host species. This has led to the application of species names to the group of organisms previously called *Pneumocystis carinii*. *P. carinii* is reserved for a species that infects rats (there is also a second species, *P. wakefieldiae*, that can infect rats). The organism infecting humans has been renamed *P. jirovecii,* in honor of Otto Jirovec, one of the investigators who recognized that this organism is the causative agent of interstitial plasma cell pneumonia. Despite the name change, the acronym PCP (for *Pneumocystis* pneumonia) continues to be used to designate the disease in humans.

Pneumocystis species show a strict host specificity: attempts to transmit, for example, rat or human *Pneumocystis* to mice have been unsuccessful. Molecular evolutionary studies suggest that *Pneumocystis* species coevolved with their hosts, with rat and mouse *Pneumocystis* diverging an estimated 33 million years ago, similar to the time rat and mouse species diverged.

Studies of *Pneumocystis* have been substantially hampered by an inability to grow any species in culture for a sustained period. Thus, the life cycle of *Pneumocystis* is unknown, although putative life cycles based on morphologic studies have been proposed; there is accumulating evidence supporting a sexual phase in the life cycle. There are two easily recognized forms of the organism: trophic forms (≈2 to 6 μm in diameter) and cysts (also called asci; ≈6 to 8 μm in diameter), which can contain up to eight intracystic bodies (ascospores); additional intermediate forms are also seen. Trophic forms, which have an amorphous shape, are estimated to outnumber cysts, which are spherical, by approximately 10 : 1 in an infected lung. β-1,3-Glucan is found in the cyst form only, where it contributes to the rigidity of the cell wall. The estimated genome size is approximately 8 million base pairs.[2] The most abundant surface protein of *Pneumocystis*, the major surface glycoprotein, is found on both cysts and trophic forms and is encoded by a multicopy gene family, only one of which is apparently expressed in a given organism; this provides *Pneumocystis* with the potential for antigenic variability. Although, to date, only a single *Pneumocystis* species has been found that infects humans, molecular typing techniques have demonstrated a high level of diversity among human *Pneumocystis* isolates.

EPIDEMIOLOGY

Pneumocystis has a worldwide distribution. Serologic studies show a high prevalence of anti-*Pneumocystis* antibodies in all populations studied to date.

In America and Europe, serologic studies have demonstrated that most humans develop antibodies to *Pneumocystis* by an early age, suggesting that this is a ubiquitous organism. In support of this, an autopsy study in infants younger than 1 year with no major underlying medical problems identified *Pneumocystis* infection in lung tissue by polymerase chain reaction (PCR) in 100% of cases.

Animal studies have demonstrated that *Pneumocystis* is transmitted by the respiratory route. Human infection appears to be transmitted in the same manner. There is no evidence that water or fomites play a role in transmission. A small number of animal studies and very limited human data suggest that transmission can occur transplacentally, although the clinical relevance is unknown.

Given the strict host specificity of *Pneumocystis*, the source of organisms infecting humans is presumably other humans, by either direct or indirect exposure; *Pneumocystis* is not a zoonosis. PCR-based studies have identified *Pneumocystis* DNA in air sampled in close proximity to patients with PCP, and outbreaks, especially in renal transplant patients, have been linked to a single strain of *Pneumocystis*, both of which suggest that patients with PCP can transmit it to other susceptible patients.[3] Because clinically apparent PCP is rare, and given the high penetration of infection into healthy human populations at a young age, *Pneumocystis* infection is also likely acquired from apparently healthy humans, in whom subclinical infection (either asymptomatic or minimally symptomatic) must be quite common to allow such rapid and broad dissemination. The high frequency of detection in infants suggests an important role in transmission.

Although infection with *Pneumocystis* is widespread, in immunocompetent hosts, it does not appear to cause significant disease. Clinically significant PCP occurs exclusively in patients with severe levels of immunodeficiency that are usually associated with a high risk for other opportunistic pathogens. Populations at risk include those with congenital immunodeficiencies (Chapter 250), especially severe combined immunodeficiency (SCID) and hyper–immunoglobulin M (IgM) syndrome; patients with HIV infection (Chapter 384) and human T-lymphotropic virus-1–associated lymphoma (Chapter 378); patients receiving chemotherapy for the treatment of malignancies, especially lymphoma; transplant patients receiving immunosuppressive therapy; and patients being treated with prolonged courses of immunosuppressive drugs (Chapter 35), especially corticosteroids, for inflammatory diseases such as Wegener's granulomatosis or systemic lupus erythematosus. Biologic agents such as rituximab and those targeting tumor necrosis factor-α (TNF-α) are also associated with an increased, though low, absolute risk for PCP (e.g. 0.18 to 0.4% in patients receiving anti-TNF-α agents in Japan).[4] Among patients with HIV infection, the best predictor of the risk for developing PCP is the CD4 count: patients with CD4 counts less than 200 cells/mm^3, and especially those with CD4 counts less than 100 cells/mm^3, are at greatest risk. Patients with a prior history of PCP and those with unexplained fever, weight loss, or thrush are also at increased risk. Among other patient populations, laboratory parameters are not as useful in quantifying risk, although non-HIV patients with CD4 counts less than 200 cells/mm^3 may be at increased risk. Additional information on risk is provided in the section on prophylaxis.

For many years, development of PCP was thought to result from a reactivation of latent infection by organisms that remained viable following infection at an early age, similar to tuberculosis. However, recent molecular epidemiologic studies based on the detection of mutations in the dihydropteroate synthase (DHPS) gene of *Pneumocystis*, as well as genotyping of isolates from outbreaks of PCP primarily in renal transplant patients, have provided compelling evidence that the infecting strain is often recently acquired. Molecular studies have further documented that in the majority of non-outbreak cases of PCP, more than one strain can be identified in respiratory samples. In patients who develop recurrent PCP, this recurrence can be due to relapse, especially for early recurrences, or to reinfection with a novel strain. Recurrent PCP occurs almost exclusively in HIV-infected patients, in whom the risk was greater than 50% early in the AIDS epidemic, before the availability of highly active antiretroviral therapy (HAART) and the broad use of anti-*Pneumocystis* prevention.

PATHOBIOLOGY

Animal studies have provided important insights into the pathogenesis of *Pneumocystis* infection. Exposure for as little as 1 day to a *Pneumocystis*-infected animal results in transmission of infection. In healthy animals, an adaptive immune response develops by approximately 5 to 6 weeks, which leads to control and clearance before the organism burden produces symptoms. CD4 cells are critical to this control, although other populations, including B cells and macrophages, also play important roles. CD8 cells can contribute to the associated inflammation. In immunodeficient animal models, inability to control *Pneumocystis* replication leads to severe pneumonia by 2 to 3 months. Limited human data suggest a similar time course. Host inflammatory responses appear to play a critical role in the development of pulmonary symptoms. This may account for the development of symptoms of PCP in patients in whom corticosteroids are being tapered, as well as the exacerbation of hypoxia that develops approximately 4 days after the initiation of anti-*Pneumocystis* therapy (in the absence of concomitant corticosteroid therapy).

The lung pathology in patients with PCP is characteristic. Staining with hematoxylin and eosin demonstrates an acellular, foamy, eosinophilic, intra-alveolar exudate associated with mild interstitial inflammation (Fig. 341-1). With disease progression, hyaline membrane formation and interstitial as well as intraluminal fibrosis develop. Although methenamine silver stain highlights cysts scattered throughout the eosinophilic exudate, based on Giemsa staining of thin sections as well as electron micrographs, this exudate is composed almost entirely of *Pneumocystis* organisms. Atypical pathology, including noncaseating granulomas and intrapulmonary cystic changes, can also be seen.

As noted earlier, patients with certain congenital immunodeficiencies, especially SCID patients, who have global T- and B-cell defects, and hyper-IgM syndrome patients, whose primary defect is in CD40-CD40L signaling,

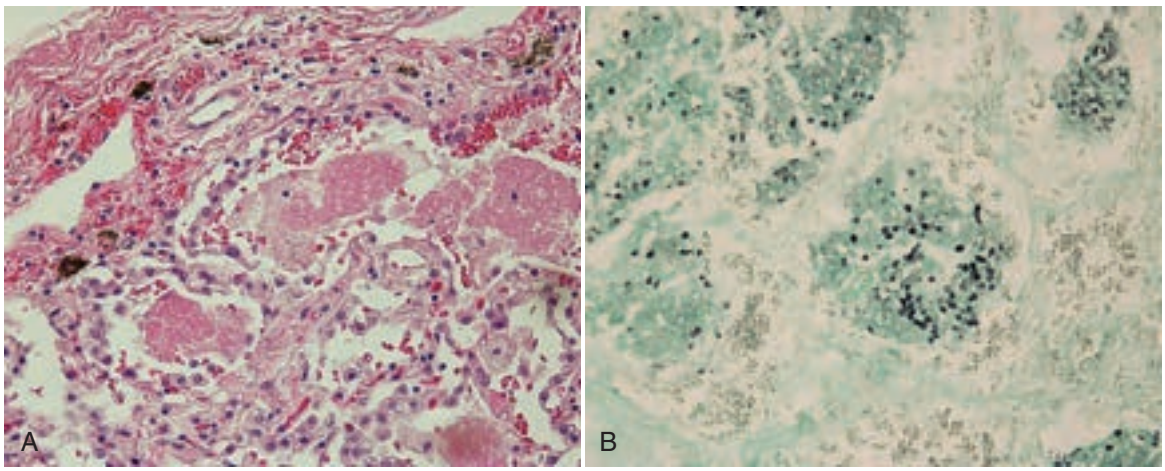

FIGURE 341-1. Histopathology of the lung of a patient who died from *Pneumocystis* pneumonia (PCP). A, The hematoxylin and eosin–stained section shows the characteristic acellular, eosinophilic, intra-alveolar exudate that is typical of PCP. **B,** Methenamine silver staining of lung tissue from the same patient demonstrates black-staining cysts scattered throughout the intra-alveolar exudates.

are at increased risk for developing PCP. Among HIV-infected patients, polymorphisms in the FcγRIIa gene and in the chemokine receptor gene for CCRL2 were each associated with an increased risk for developing PCP in single studies.[5]

CLINICAL MANIFESTATIONS

In nonimmunosuppressed humans, no well-defined clinical syndrome is associated with *Pneumocystis* infection. *Pneumocystis* has been identified in infants by PCR and may be associated with a mild respiratory syndrome, but a postulated association with sudden infant death syndrome has not been supported by data from well-controlled studies. *Pneumocystis* has been detected by PCR in pulmonary samples from patients with chronic obstructive pulmonary disease (COPD), but what role, if any, it plays in the development or progression of COPD (Chapter 88) remains to be elucidated.[6]

Pneumonia is the primary clinical manifestation of *Pneumocystis* infection in immunosuppressed patients. PCP typically presents with a fever, nonproductive cough, and shortness of breath that initially occurs only on exertion but, without therapy, inevitably progresses to dyspnea at rest. Only one or two of these symptoms may be present initially. Development of symptoms may be insidious, over the course of a few weeks, as is common in patients with HIV infection; a more rapid onset, over the course of only a few days, is more common in non-AIDS patients. Purulent sputum production is unusual, and chills and chest pain occur in a minority of patients. Patients with HIV infection may present with other manifestations of immunodeficiency, including weight loss and thrush.

In non-AIDS patients, corticosteroids are a common risk factor.[7] Clinical manifestations may develop as corticosteroids are being tapered, which presumably represents the unmasking of an inflammatory response to the infection as immunosuppression is decreased.

Rarely, extrapulmonary disease can involve the skin, eye (choroiditis), central nervous system, bone marrow, thyroid, spleen, liver, gastrointestinal tract, lymph node, or multiple organs in disseminated disease. Extrapulmonary disease can occur with or without concurrent pneumonia. Use of aerosol pentamidine for prophylaxis has been associated with an increased risk for extrapulmonary disease in HIV-infected patients, but even in that circumstance, it remains extremely rare. Symptoms are related to the specific site involved and may be nonspecific; diagnosis is often made at autopsy.

DIAGNOSIS

Given the nonspecific symptoms, especially early in the disease process, clinicians must have a high index of suspicion for PCP even in patients not known to be immunosuppressed; many patients with HIV infection are unaware of their status until they present with an opportunistic infection. Knowledge of the most recent CD4 count is helpful in assessing HIV-infected patients because PCP is rare in those whose CD4 counts are above 200 cells/mm^3.

Physical examination and routine laboratory tests are usually not helpful in making the diagnosis because many pulmonary processes, both infectious and noninfectious, can present in a similar manner. Moreover, even though patients may be tachypneic and appear to be in respiratory distress, those presenting early in the disease course may have an entirely normal lung examination. Lymphopenia is common but is a manifestation of the underlying disease rather than PCP. Lactate dehydrogenase levels may be elevated but have poor specificity.

The initial evaluation should include a chest radiograph and assessment of arterial oxygenation, either by blood gas measurement or by pulse oximetry. The chest radiograph typically shows perihilar or diffuse bilateral interstitial infiltrates that progress to a diffuse alveolar pattern (Fig. 341-2). However, PCP has been associated with unilateral disease, focal disease, consolidation, nodules, cavities, pneumothorax, and, rarely, pleural effusions. In up to 30% of patients with HIV infection, the chest radiograph appears normal; in this situation, a computed tomography (CT) scan of the chest, especially high-resolution CT, is invariably abnormal, usually showing a patchy or diffuse ground-glass pattern (see Fig. 341-2).[8]

Arterial oxygenation at rest is often abnormal, although in 30% or more of cases, it is within normal limits. Exercise testing induces desaturation and an increase in the alveolar-arterial oxygen (A-a O_2) gradient in the majority of patients with PCP, even those with normal oxygenation at rest or a normal chest radiograph. An abnormal resting diffusing capacity is also common. However, although many of these tests have a high sensitivity for PCP, their specificity is poor because other respiratory processes show similar abnormalities.

Serologic tests are not helpful in diagnosing PCP. Although antibody titers against recombinant *Pneumocystis* proteins such as major surface glycoprotein may be increased in patients with PCP, such tests have not shown utility for diagnosis in individual patients. Although serum and bronchoalveolar lavage (BAL) β-D-glucan levels are elevated in many patients, this again is a nonspecific test because other fungal infections can also lead to elevations, and conditions unrelated to fungal infection can lead to false-positive results. To date, there are inadequate data from well-conducted prospective trials to support routine use of this assay for definitive diagnosis of PCP.

Because *Pneumocystis* cannot be cultured, a definitive diagnosis of PCP requires detection of the organism in a pulmonary sample. This can be accomplished by any number of colorimetric or immunologic stains or by molecular techniques. Until the development of anti-*Pneumocystis* monoclonal antibodies in the 1980s, colorimetric stains were routinely used, and they continue to be used at many centers owing to cost considerations. Such stains include Gomori methenamine silver, toluidine blue O, Gram-Weigert, and cresyl echt violet, which stain the cyst wall of *Pneumocystis*, as well as Giemsa-type stains, including Diff-Quik, which can stain both trophic forms, the more abundant form of the organism, and intracystic bodies within cysts, but not the cyst wall. *Pneumocystis* can also be detected by Calcofluor white, Papanicolaou, periodic acid–Schiff, and, rarely, Gram stain. None of the colorimetric stains is specific for *Pneumocystis*. Cyst wall stains such as Gomori methenamine silver and toluidine blue O can stain other fungi, and Giemsa-type stains also stain background cells and cellular debris. The latter requires substantial expertise for correct interpretation.

Immunofluorescent assays using anti-*Pneumocystis* monoclonal antibodies provide a number of advantages over colorimetric stains. They are specific for *Pneumocystis* and do not show cross-reactivity with other organisms, including fungi; they can be performed and interpreted rapidly; and they have increased sensitivity, especially when examining induced sputum samples (Fig. 341-3).

Molecular techniques, primarily PCR-based assays, have been extensively evaluated for diagnosing PCP. PCR assays are 10- to 100-fold more sensitive than stains for the detection of *Pneumocystis;* this may allow for diagnosis using samples such as oral washes, which have a lower organism burden than induced sputum or BAL.[9] This increased sensitivity, however, is associated with decreased specificity because it allows the detection of organisms in patients ultimately shown not to have PCP.[10] The latter situation presumably reflects colonization or subclinical infection that does not require specific anti-*Pneumocystis* therapy. PCR-based assays are currently not broadly used because of limited availability, lack of standardization among laboratories, and the lack of a commercial product with U.S. Food and Drug Administration approval.

As detection methods for *Pneumocystis* improved, there was a parallel improvement in sample acquisition. Before the AIDS epidemic, open lung biopsy was required for diagnosis. During the 1980s, bronchoscopy, initially with brushings and biopsy, and subsequently with BAL, was shown to have a greater than 90% sensitivity for diagnosing PCP; BAL continues to be the primary diagnostic modality at many centers today. Although expectorated sputum has a low diagnostic yield, induced sputum, especially when combined with immunofluorescent staining, can have a sensitivity approaching 90%, although in many centers the diagnostic yield is much lower, likely due in part to variability in the methods used for induction and processing. Ideally, sputum induction should be the first step in diagnosis, followed by bronchoscopy with BAL (Fig. 341-4). Bronchoscopic or open lung biopsy is only rarely needed to make the diagnosis.

The differential diagnosis of pulmonary infiltrates in immunosuppressed patients is very broad and includes infections such as adenovirus, cytomegalovirus, tuberculosis, cryptococcosis, histoplasmosis, aspergillosis, and toxoplasmosis, as well as noninfectious processes such as tumor, congestive heart failure, pulmonary emboli, radiation- and chemotherapy-induced pneumonitis, and, especially in HIV-infected patients, nonspecific interstitial pneumonitis and Kaposi's sarcoma.

TREATMENT

Rx

Specific anti-*Pneumocystis* therapy should be initiated promptly when the diagnosis is suspected in a potentially susceptible patient. Dosing regimens with documented efficacy are listed in Table 341-1. Although there are no controlled studies defining the optimal duration of therapy, HIV-infected patients should be treated for 21 days and non-HIV patients for at least 14

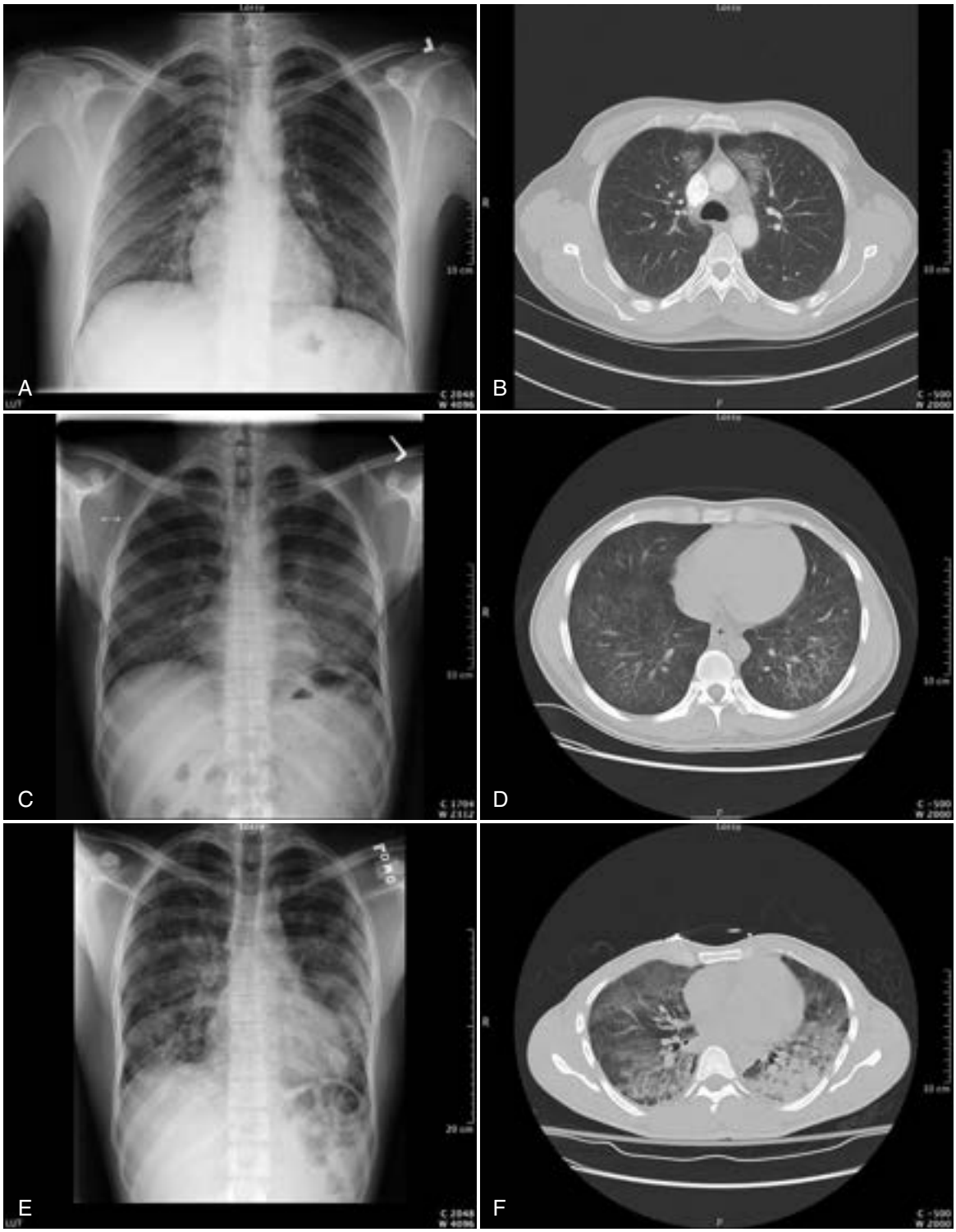

FIGURE 341-2. **Chest radiographs (A, C, E) and corresponding computed tomography (CT) scans (B, D, F) from three HIV patients diagnosed with laboratory-confirmed *Pneumocystis* pneumonia. Patient 1 presented with no symptoms; had minimal abnormalities on an incidental chest radiograph (A), suggesting an interstitial process; and had focal infiltrates on CT (B). Patient 2 presented with fever and weight loss but no shortness of breath. Pulse oximetry demonstrated 99% saturation at rest, with a decrease to 89% with exercise. Chest radiograph (C) showed bilateral lower lobe infiltrates, and CT (D) showed bilateral lower lobe interstitial infiltrates with patches of ground-glass attenuation. Patient 3 presented with a 2-week history of fever, night sweats, shortness of breath, and weakness. His Pao_2 at diagnosis was 67 mm Hg, and his A-a O_2 gradient was 53 mm Hg. Chest radiograph (E) showed bilateral mid and lower lung field infiltrates, and CT (F) showed bilateral infiltrates with lower lung consolidation.**

days.[11,12] For HIV-infected patients with moderate to severe disease (Pao_2 <70 mm Hg or A-a O_2 gradient >35 mm Hg), concomitant corticosteroid therapy should be administered. In contrast, the use of corticosteroids in non-HIV patients is not well defined, as discussed later. The diagnosis should be definitively confirmed by sputum induction or bronchoscopy as soon as possible; however, delaying such confirmation for a few days after the initiation of therapy will not decrease the diagnostic yield because organisms can be detected in clinical samples for more than 3 weeks after the initiation of therapy. Empirical therapy in the absence of a confirmed diagnosis runs the risk of delaying appropriate therapy for another infection, giving inappropriate therapy with known toxicities, and performing a definitive procedure such as bronchoscopy when the patient is failing therapy and consequently has more severe pulmonary compromise.

The treatment of choice for PCP or extrapulmonary disease, regardless of severity, is trimethoprim-sulfamethoxazole, which combines inhibitors of two enzymes in the folate synthetic pathway of *Pneumocystis*: sulfamethoxazole, an inhibitor of DHPS, and trimethoprim, an inhibitor of dihydrofolate reductase (DHFR). Trimethoprim-sulfamethoxazole is available in both oral and intravenous formulations. Oral therapy should be reserved for patients with mild to moderate disease in whom poor absorption is not a concern.

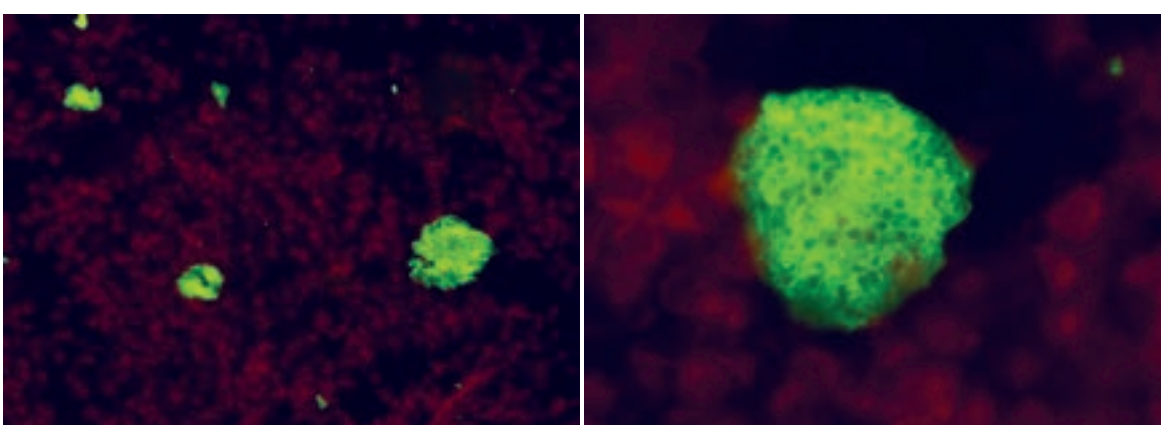

FIGURE 341-3. **Immunofluorescent detection of *Pneumocystis* using a direct fluorescent antibody test to examine an induced sputum sample. A,** Low power (100× original) allows visualization of multiple clusters of organisms staining bright green. **B,** Under high power (400× original), individual organisms can be seen within a single cluster.

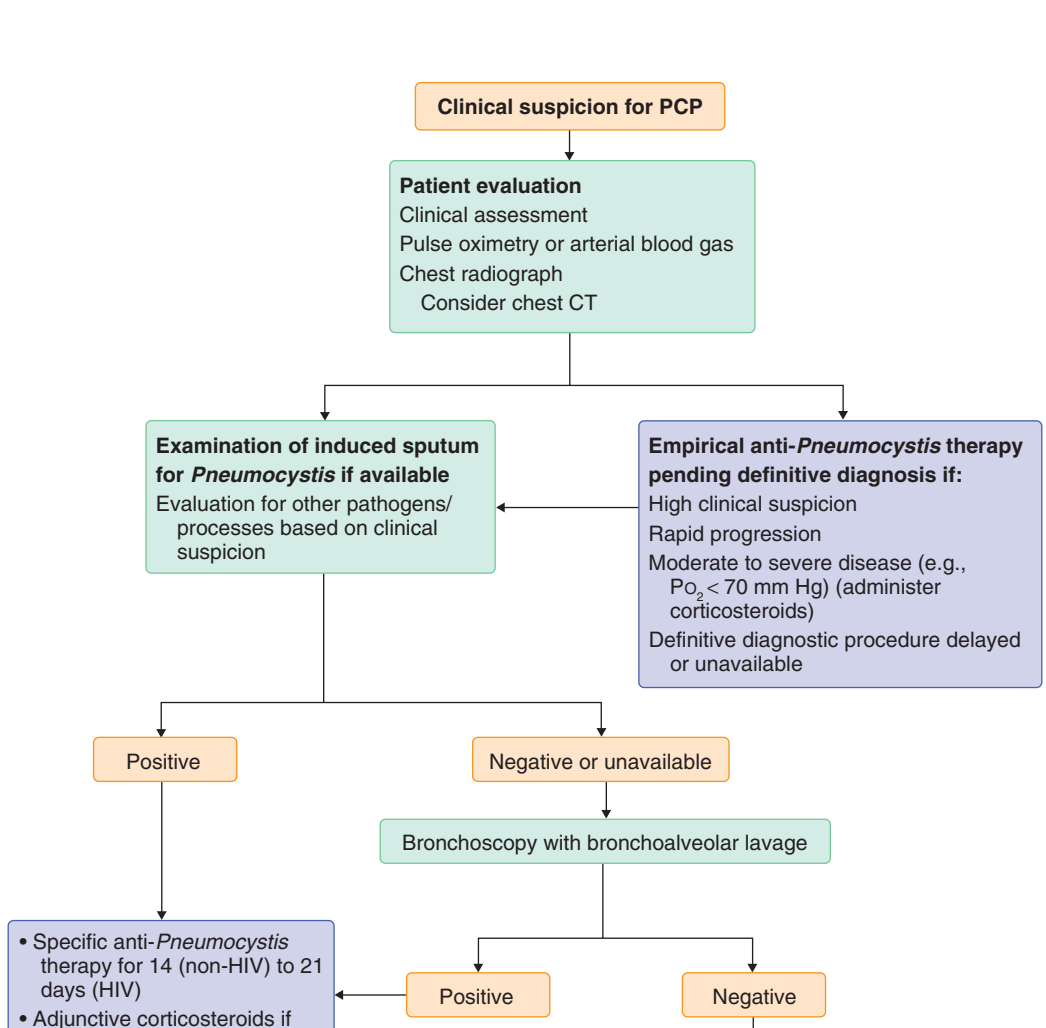

FIGURE 341-4. **Algorithm for the evaluation of patients with suspected *Pneumocystis* pneumonia (PCP). CT = computed tomography; HIV = human immunodeficiency virus.**

TABLE 341-1 DRUG REGIMENS FOR TREATMENT OF *PNEUMOCYSTIS PNEUMONIA* (PCP)

INDICATION	REGIMEN PREFERENCE	DRUG	ROUTE	DOSE	COMMENTS
Mild PCP: PaO_2 ≥70 mm Hg or (A-a) O_2 gradient ≤35 mm Hg	Preferred	Trimethoprim-Sulfamethoxazole (TMP-SMX)	PO	2 double-strength (160 TMP+800 SMX) tablets TID	
	Alternative	Trimethoprim plus	PO	5 mg/kg TID (15 mg/kg/d)	
		Dapsone	PO	100 mg QD	If possible test for G6PD deficiency before use
	Alternative	Clindamycin plus	PO	450 mg QID or 600 mg TID	
		Primaquine	PO	30 mg (base) QD	If possible test for G6PD deficiency before use
	Alternative	Atovaquone	PO	750 mg BID with food	
Moderate to severe PCP: PaO_2 <70 mm Hg or (A-a) O_2 gradient >35 mm Hg; moderate PCP ((A-a) O_2 gradient 35 to 45 mm Hg); can be treated with an oral regimen	Preferred	Trimethoprim-Sulfamethoxazole (TMP-SMX)	IV	5 mg/kg Q8H TMP and 25 mg/kg Q8H SMX (15 mg/kg/d TMP and 75 mg/kg/d SMX)	May switch to oral therapy following clinical improvement
	Alternative	Pentamidine	IV	3-4 mg/kg QD	Infuse over >60 minutes
	Alternative	Clindamycin plus	IV	600 mg Q6H or 900 mg Q8H	May switch to oral therapy following clinical improvement
		Primaquine	PO	30 mg (base) QD	No parenteral formulation is available
Adjunctive therapy for moderate to severe PCP: PaO_2 <70 mm Hg or (A-a) O_2 gradient >35 mm Hg	Preferred	Prednisone	PO	40 mg BID, days 1-5; 40 mg QD, days 6-10; 20 mg QD days 11-21	Begin as early as possible and within 72 hours; efficacy if started later has not been demonstrated
	Preferred	Methylprednisolone	IV	30 mg BID, days 1-5; 30 mg QD, days 6-10; 15 mg QD days 11-21	Use if parenteral therapy is necessary

Note: HIV-infected patients should receive 21 days of therapy; non-HIV patients should receive at least 14 days of therapy.

Outpatient therapy should be reserved for patients with mild to moderate disease who will reliably return for follow-up.

The efficacy of trimethoprim-sulfamethoxazole for both the treatment and the prevention of PCP was first demonstrated in a pediatric cancer population in the 1970s; it has been the first-line therapy since then, when the only alternative was pentamidine. The high incidence of PCP during the early years of the AIDS epidemic led to the identification of a number of new agents with anti-*Pneumocystis* activity, and these were extensively evaluated in randomized controlled trials, primarily in HIV-infected patients. In patients with mild to moderate disease (A-a O_2 gradient <45 mm Hg), trimethoprim-sulfamethoxazole has superior efficacy compared with atovaquone and similar efficacy compared with trimethoprim-dapsone and clindamycin-primaquine. In patients with moderate to severe disease (A-a O_2 gradient >30 mm Hg), it showed superior efficacy to trimetrexate, a potent inhibitor of *Pneumocystis* DHFR that is approved for the treatment of PCP but is no longer commercially available. In smaller, lower power studies, trimethoprim-sulfamethoxazole and pentamidine showed similar efficacy.

The major toxicities associated with trimethoprim-sulfamethoxazole include fever, rash, neutropenia, thrombocytopenia, nausea, vomiting, and transaminase elevations. Hyperkalemia and crystalluria have also been reported, and hyponatremia is seen primarily in association with intravenous administration. Toxicities usually appear after the first week of therapy. Toxicities are much more common in HIV-infected patients, occurring in about 50 to 60%; 15 to 35% of these patients discontinue therapy because of the adverse events. Although folinic acid can decrease the toxicities associated with some DHFR inhibitors such as pyrimethamine, it should not be administered with trimethoprim-sulfamethoxazole; in one placebo-controlled trial, it did not decrease side effects but was associated with an increased risk for therapeutic failure and death.

Alternative regimens for patients with mild to moderate disease include trimethoprim-dapsone, clindamycin-primaquine, and atovaquone. Like sulfamethoxazole, dapsone is an inhibitor of *Pneumocystis* DHPS. Adverse reactions to trimethoprim-dapsone include rash, fever, nausea and vomiting, transaminase elevations, methemoglobinemia, anemia, and mild hyperkalemia. Approximately 20 to 30% of patients with adverse reactions to trimethoprim-sulfamethoxazole experience adverse reactions to trimethoprim-dapsone. Toxicities associated with clindamycin-primaquine include fever, rash, diarrhea, anemia, neutropenia, transaminase elevations, and methemoglobinemia. For both dapsone and primaquine and, to a lesser extent, sulfamethoxazole, glucose-6-phosphate dehydrogenase deficiency (Chapter 161) can increase the risk for hemolytic anemia and methemoglobinemia. In a randomized trial comparing trimethoprim-sulfamethoxazole, trimethoprim-dapsone, and clindamycin-primaquine in 181 HIV-infected patients with mild to moderate disease, response rates and toxicities were similar among the three arms, with an overall therapeutic failure rate of 9% by day 21 and a dose-limiting toxicity rate of 31%. Serious transaminase elevations were more common in the trimethoprim-sulfamethoxazole group, and serious hematologic toxicities were more common in the clindamycin-primaquine group.

Atovaquone is a hydroxynaphthoquinone with activity against *Toxoplasma* and malaria as well as *Pneumocystis*. Atovaquone (tablet formulation) was less effective than trimethoprim-sulfamethoxazole in a randomized trial in HIV-infected patients with mild to moderate disease and showed a trend toward lesser efficacy compared with pentamidine in another study. Low serum atovaquone levels and preexisting diarrhea were associated with therapeutic failure. The current formulation is a suspension that has approximately 50% greater bioavailability than the tablet formulation, which may result in improved responses. Atovaquone should be taken with food because this increases its bioavailability, and it should be avoided in patients with potentially decreased gastrointestinal absorption (e.g., diarrhea). Toxicities of atovaquone include rash, fever, transaminase elevations, nausea, vomiting, diarrhea, neutropenia, and anemia.

Therapeutic alternatives to trimethoprim-sulfamethoxazole for patients with disease requiring parenteral therapy are limited to clindamycin-primaquine (but only clindamycin is available for intravenous administration) and pentamidine. Caspofungin and other echinocandins, which are β-1,3-glucan synthase inhibitors, cannot be recommended because there have been no clinical trials documenting their efficacy; β-1,3-glucan is present in the cyst but not in the trophic form of the organism.

Pentamidine was the first drug demonstrated to have anti-*Pneumocystis* activity. Available data suggest that trimethoprim-sulfamethoxazole and pentamidine have similar efficacy; trimethoprim-sulfamethoxazole is the preferred regimen because the toxicities associated with pentamidine are more frequent and potentially more severe. Intravenous pentamidine was originally associated with severe hypotension, but a slow (>1 hour) infusion is usually well tolerated. Intramuscular administration is associated with a high frequency of sterile abscesses at the injection site. Toxicities associated with pentamidine, which occur in about 50 to 60% of patients and frequently result in discontinuation of the drug, include nephrotoxicity, hypoglycemia, hyperglycemia, fever, neutropenia, thrombocytopenia, hypotension, hyperkalemia, transaminase elevations, and pancreatitis. Hypoglycemia may be life-threatening and may precede the development of hyperglycemia; hyperglycemia may be irreversible. Torsades de pointes (Chapter 65) has also rarely been reported.

Adjunctive Corticosteroid Therapy

Initiation of specific anti-*Pneumocystis* therapy is associated with deterioration in oxygenation after approximately 3 to 4 days; this likely results from a host inflammatory response to organisms damaged by therapy. Randomized controlled trials have demonstrated that the early addition of corticosteroids to specific anti-*Pneumocystis* therapy in HIV-infected patients can prevent this deterioration and improve survival, without a significant increase in opportunistic complications other than localized herpes simplex infection. In the largest such study, corticosteroid therapy was associated with a 50% decrease in respiratory failure and mortality; this benefit was limited to patients with moderate to severe disease (E-Fig. 341-1). For HIV-infected patients, corticosteroids and specific anti-*Pneumocystis* therapy should be started at the same time; the addition of corticosteroids after 72 hours has shown no benefit,

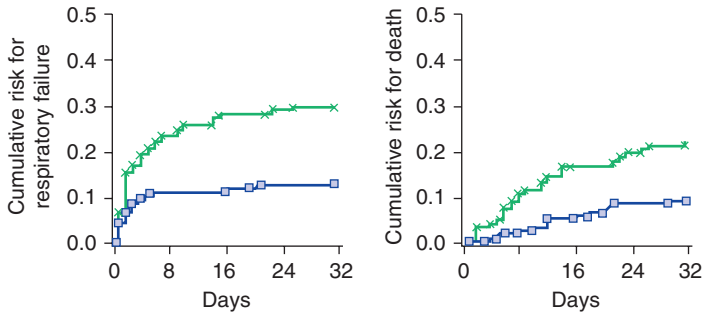

E-FIGURE 341-1. Cumulative risks for an unfavorable outcome during a randomized trial of corticosteroid therapy in patients with HIV infection and *Pneumocystis* pneumonia measured over a period of 31 days. The risk for respiratory failure *(left)* was 0.14 in the corticosteroid group (*blue square*) and 0.30 in the standard-treatment group (*green*) ($P = .004$). The risks for death *(right)* were 0.11 and 0.23, respectively ($P = .009$). Corticosteroid therapy was associated with a significant decrease in both outcomes. (Modified from Bozzette SA, Sattler FR, Chiu J, et al. A controlled trial of early adjunctive treatment with corticosteroids for *Pneumocystis carinii* pneumonia in the acquired immunodeficiency syndrome. *N Engl J Med*. 1990;323:1451-1457. Copyright 1990, Massachusetts Medical Society. All rights reserved.)

although it is reasonable to add them if patients exhibit deterioration after this time. Although the optimal regimen has not been defined by controlled trials, the tapering regimen from the largest study is most commonly used (see Table 341-1).

The optimal utilization of corticosteroids in non-HIV patients who are often receiving them as part of the treatment regimens for their underlying disease is less clear because data from randomized controlled trials are not available; dosing may need to be individualized to balance the immunosuppressive effects that potentially contributed to the development of PCP against the anti-inflammatory effects that may ameliorate life-threatening pulmonary dysfunction. One retrospective analysis of 31 patients suggested that increasing corticosteroids to a prednisone equivalent of 60 mg/day or more was associated with clinical benefit. It is reasonable to administer corticosteroids to non-HIV patients with moderate or severe disease if they were not receiving corticosteroids, using the same regimen as for HIV-infected patients; for patients already taking corticosteroids at lower doses, the dosage can be increased to those levels.[13]

Initiation of Antiretroviral Therapy

Given that many patients with HIV infection who are diagnosed with PCP are not receiving antiretroviral therapy, an important issue is how soon to start HAART after a diagnosis of PCP. Retrospective and prospective studies have suggested that, in general, it is safe to initiate HAART while patients are being treated for PCP, and early HAART may be associated with an improved outcome. In a randomized 282-patient trial that examined early versus late initiation of HAART in patients with acute opportunistic infections, 63% of whom had PCP, the early initiation arm (HAART started a median of 12 days after starting therapy for the opportunistic infection) had a decreased rate of AIDS progression or death. Similar findings for tuberculosis in developing countries have emphasized the benefit of early initiation of HAART.

Major concerns about initiating HAART include the risk for adverse drug reactions, which may be confused with adverse reactions to anti-PCP therapy; the risk for overlapping toxicities, which may complicate management; and the risk for immune reconstitution. There are several reports describing apparent immune reconstitution inflammatory syndrome (Chapter 395), which can be life-threatening, in patients who started HAART soon after being diagnosed with PCP. Thus, many clinicians initiate HAART during or immediately after completion of anti-*Pneumocystis* therapy, assuming the patient has shown clinical improvement, is able to tolerate oral medications, and accepts the commitment to lifelong therapy. However, the parameters for such an approach are difficult to define precisely. Patients who start HAART early should be closely monitored for a recurrence of symptoms that may represent immune reconstitution.

Treatment Failure

The optimal approach to the management of patients who are failing therapy has not been well defined. In patients with progressive respiratory deterioration, it is critical that the diagnosis of PCP be confirmed rather than presumptive and that other concurrent processes (e.g., other infections, congestive heart failure, pulmonary emboli) have been ruled out; bronchoscopy should be considered to facilitate these determinations. Parenteral therapy should be used to eliminate absorption concerns, and corticosteroid medications should be added if this has not already been done. Because patients who will ultimately respond can show clinical deterioration at 3 to 4 days, as noted earlier, it is reasonable to wait 5 to 8 days before considering a change in drug therapy.

Only trimethoprim-sulfamethoxazole and pentamidine are available in parenteral formulations. Parenteral clindamycin is available, but primaquine is available only as a tablet. No randomized trials have examined the relative efficacy of these agents in patients who are failing therapy. For patients who have not received trimethoprim-sulfamethoxazole, this should be the first choice as an alternative agent, assuming the patient did not have a life-threatening adverse reaction previously. Rapid desensitization (similar to penicillin desensitization), ideally in consultation with an allergy specialist, can be considered in patients with prior adverse reactions; however, patients with a history of Stevens-Johnson syndrome or toxic epidermal necrolysis should not be rechallenged. Although retrospective cohort studies and meta-analyses have found that clindamycin-primaquine is superior to pentamidine in patients failing a first-line regimen, there are potential biases in such analyses (e.g., severity of illness or ability to take oral medications may have affected the choice of salvage regimen). There are no data to recommend switching to an alternative agent rather than adding an alternative agent (if toxicity is not an issue); both approaches have been used.

Resistance

Although *Pneumocystis* cannot be cultured, molecular studies have identified mutations in genes that are the targets of anti-*Pneumocystis* therapy, and these mutations appear to represent the development of resistance by *Pneumocystis* to these agents. The best-characterized mutations have been identified in the DHPS gene of *Pneumocystis*, which is the target of sulfamethoxazole and dapsone. Two mutations at the active site of this enzyme, which can occur either individually or together, have been identified with greater frequency in patients receiving trimethoprim-sulfamethoxazole or dapsone for prophylaxis; in vitro studies suggest that these mutations confer resistance. The clinical relevance of these mutations remains uncertain; some studies have found worse outcomes in patients with these mutations, but others have found no such association.[14] Most patients in whom these mutations were identified retrospectively were successfully treated with sulfa-containing drugs. In contrast to DHPS, there are very limited reports suggesting that the DHFR gene of *Pneumocystis*, which is the target of trimethoprim and pyrimethamine, has developed potential drug-resistant mutations.

Atovaquone presumably binds to the mitochondrial bc_1 complex of *Pneumocystis* and thus inhibits electron transport. Multiple mutations have been identified in the cytochrome B gene of *Pneumocystis,* which presumably represent resistance in patients receiving atovaquone for prophylaxis; these mutations have not, however, been associated with clinical outcome.

Because the presence of these mutations has not been definitively associated with worsening prognosis, clinical decisions should not be based on their identification. Methods for identifying these mutations are not readily available, and their detection should remain a research tool until their clinical relevance can be better defined.

PREVENTION

Although *Pneumocystis* is transmitted by the airborne route, exposure to the organism appears to be ubiquitous in humans, suggesting that avoidance of exposure may be difficult. In animal models, it takes 2 to 3 months following exposure to develop a heavy infection. If the pattern of growth were similar in humans, clinical symptoms would develop only months after exposure to a source that is likely unknown. Currently, respiratory isolation of patients with active PCP is not required, although it is reasonable to avoid having a susceptible patient share a room with a PCP patient. Recent outbreaks in renal transplant patients strongly suggest a common source of infection; a better understanding of the patterns of transmission in these settings may lead to improved guidelines for preventing the spread of infection. It is noteworthy that in these outbreaks, broad institution of anti-*Pneumocystis* prophylaxis was the intervention that terminated the outbreaks.

A major advance in the management of patients at risk for the development of PCP was the demonstration that trimethoprim-sulfamethoxazole was highly effective in preventing the disease in a susceptible pediatric population. Subsequent studies, primarily in HIV-infected patients, demonstrated that additional drug regimens were also effective. This has led to the broad use of anti-*Pneumocystis* prophylaxis in a wide range of susceptible populations.

Two important issues in administering prophylaxis are identifying populations at risk and defining the period of risk during which prophylaxis should be provided. AIDS patients are at especially high risk; before the use of prophylaxis or HAART, the lifetime incidence of PCP in this population was estimated at 60 to 80%. The most recent CD4 count is a validated surrogate marker for HIV-infected patients: patients with CD4 counts below 200 cells/mm^3 are at substantially increased risk for developing PCP, and prophylaxis is recommended for this group.[15] Although 10 to 15% of patients who develop PCP have higher CD4 counts, the incidence is very low in this population, given the large number of patients who fall in this category. Patients with CD4 counts greater than 200 cells/mm^3 but a CD4 percentage of less than 14% or a history of an AIDS-defining illness are also candidates for prophylaxis. For pediatric patients with HIV infection, in whom the normal CD4 count changes with age, guidelines are based on age. Prophylaxis is recommended for children older than 6 years with CD4 counts below 200 cells/mm^3 or 15%, for children between 1 and 5 years old with CD4 counts below 500 cells/mm^3 or 15%, and for all children younger than 12 months.[16]

Before the availability of HAART, when patients with HIV infection initiated prophylaxis, they were committed to continuing it for life because immunologic decline was irreversible. With HAART, however, control of HIV replication leads to an increase in the CD4 count, which is associated with a concomitant decrease in the risk for developing PCP. Multiple studies have shown that when the CD4 count has been above 200 cells/mm^3 for at least 3 months (ideally in the setting of controlled HIV replication), prophylaxis can be safely discontinued because the risk for developing PCP is no greater than in patients whose CD4 counts never fell below 200 cells/mm^3. In most of these studies, the median CD4 count was greater than 300 cells/mm^3, and HIV viral loads were below detection limits in the majority of patients. Recent uncontrolled studies have suggested that prophylaxis can also be safely discontinued in patients with CD4 counts between 100 and

200 cells/mm^3 who have virologically suppressed HIV, but specific criteria for discontinuation (e.g., duration of viral suppression) were not defined in these studies.[17]

Among non-HIV-infected patients, the CD4 count is not routinely measured, and it has not been shown to have the same predictive value for the development of PCP as in HIV-infected patients; however, CD4 counts below 200 cells/mm^3 do appear to increase their susceptibility. Recommendations for PCP prophylaxis in these populations are based on clinical parameters, including empirical identification of periods of risk and estimation of levels of immunosuppression. Very broad prophylaxis has not been implemented because of the side effects associated with these regimens; for example, there is concern that trimethoprim-sulfamethoxazole might cause bone marrow suppression that would interfere with engraftment or cause nephrotoxicity that would damage a transplanted kidney.

Risk factors for non-HIV-infected patients include underlying disease, older age, use of immunosuppressive drugs, radiation therapy, graft-versus-host disease, and concomitant cytomegalovirus infection (Chapter 376). Patients with malignancies, especially hematologic malignancies, but increasingly solid tumors as well, are at risk for PCP primarily owing to the therapies they receive; the incidence can range from 1 to 43% in the absence of prophylaxis and is highly dependent on the intensity and duration of immunosuppression. In the absence of prophylaxis, the risk for developing PCP in transplant patients, whether hematopoietic stem cell or solid organ transplants, is reportedly about 5 to 15%, although lung and heart-lung transplant patients appear to have a higher incidence (up to 43%). Among patients with collagen vascular disease, the reported risk is less than 2% without prophylaxis, although patients with Wegener's granulomatosis (Chapter 270) reportedly have a risk up to 12%, presumably because of the use of more immunosuppressive treatment regimens.[18] In patients with inflammatory bowel disease, the incidence in one large retrospective cohort study, was about 1% per year, with a greater risk for Crohn's disease compared with ulcerative colitis.[19] The current risk for developing PCP in non-HIV populations is difficult to quantify because of the widespread use of prophylaxis and because immunosuppressive regimens are evolving.

To facilitate the management of prophylaxis in at-risk populations, a number of guidelines have been developed by expert panels that have made recommendations based on the strength of the available data. HIV guidelines have already been summarized. For allogeneic stem cell transplant recipients, prophylaxis is recommended from the time of engraftment to at least 6 months after transplantation, and longer for patients who continue receiving immunosuppressive therapy or who have chronic graft-versus-host disease. For autologous stem cell transplant recipients, who have a lower risk for PCP, prophylaxis for 3 to 6 months should be considered if the degree of immunosuppression is substantial owing to underlying disease or therapy (e.g., patients with leukemia or lymphoma receiving intensive conditioning or immunosuppressive therapy).

For solid organ transplant patients, prophylaxis has not been universally adopted; it has been used primarily in centers with a known incidence greater than 3%. Guidelines recommend the administration of prophylaxis for 3 to 12 months in renal transplant patients and for longer periods, up to lifelong, in heart, lung, liver, and intestine transplant recipients. As noted earlier, a number of outbreaks of PCP have recently been reported in renal transplant centers, and many of those patients developed disease more than 1 year after transplantation. Risk factors identified in case-control studies have included older age, recent or concurrent cytomegalovirus infection, and treatment for rejection. Specific immunosuppressive drugs, such as mycophenolate mofetil and cyclosporine, have not been consistently implicated.

For patients with inflammatory bowel disease (Chapter 141), who appear to be at increased risk as newer immunosuppressive agents are used, data are limited; however, consensus-based guidelines recommend prophylaxis for patients receiving triple immunomodulators that include either a calcineurin inhibitor or an anti-TNF agent. No consensus was reached for less intensive regimens. For patients with connective tissue disorders or vasculitis, there are currently no consensus guidelines.

Corticosteroid therapy (Chapter 35) is a well-described risk factor in non-HIV-infected patients, with approximately 90% of patients receiving such therapy before developing PCP in some studies. Higher dose and longer duration increase the risk. Not all patients who receive corticosteroids are at risk, however; for instance, asthmatic patients receiving corticosteroid therapy are at low risk. Although there are no consensus guidelines on the use of prophylaxis for patients receiving corticosteroids, one reasonable approach is to provide prophylaxis to patients with an underlying immunosuppressive or inflammatory disease who receive at least 20 mg of prednisone or equivalent for longer than 1 month. Other immunosuppressive agents (Chapter 35), such as calcineurin inhibitors, sirolimus, TNF antagonists, and rituximab, also appear to increase the risk for developing PCP, primarily in the patient populations noted earlier. Based on the manufacturer's recommendations, patients receiving temozolomide plus radiotherapy for glioblastoma multiforme should receive anti-*Pneumocystis* prophylaxis.

Trimethoprim-sulfamethoxazole is the first-line agent for prophylaxis in all populations (Table 341-2). Alternatives include dapsone alone or combined with pyrimethamine plus leucovorin, atovaquone, and aerosol pentamidine administered by the Respirgard II nebulizer. In a randomized trial of 843 HIV-infected patients comparing trimethoprim-sulfamethoxazole with dapsone and aerosol pentamidine, no significant differences were seen on an intent-to-treat basis, but the lowest failure rates were seen while patients were receiving trimethoprim-sulfamethoxazole. Conversely, trimethoprim-sulfamethoxazole was superior to aerosol pentamidine in another randomized study. In other large randomized trials in HIV-infected patients, the following regimens showed similar efficacy: atovaquone suspension and dapsone, atovaquone suspension and aerosol pentamidine, and dapsone-pyrimethamine and aerosol pentamidine. No randomized trials of these regimens have been conducted in non-HIV-infected populations, but clinical experience suggests that they are effective in these populations as well. Patients receiving pyrimethamine plus sulfadiazine plus leucovorin for treatment of toxoplasmosis do not require additional anti-*Pneumocystis* prophylaxis as this regimen also prevents PCP.

Although the combination of sulfadoxine and pyrimethamine is also efficacious, it is contraindicated in patients with sulfonamide allergies. Moreover, because Stevens-Johnson syndrome and other potentially life-threatening cutaneous reactions are more common with this combination than with trimethoprim-sulfamethoxazole, and because its long half-life results in slow

TABLE 341-2 DRUG REGIMENS FOR PREVENTION OF *PNEUMOCYSTIS PNEUMONIA* (PCP)

INDICATION	DRUG	ROUTE	DOSE	COMMENTS
Preferred	Trimethoprim-Sulfamethoxazole (TMP-SMX)	PO	1 double-strength tablet (160 mg TMP+800 mg SMX) or one single-strength tablet (80 mg TMP+400 mg SMX) QD	Also active in preventing toxoplasmosis
Alternative	Trimethoprim-Sulfamethoxazole (TMP-SMX)	PO	1 double-strength tablet (160 mg TMP+800 mg SMX) 3 times weekly	Also active in preventing toxoplasmosis
Alternative	Dapsone	PO	100 mg QD or 50 mg BID	Test for G6PD deficiency before use
Alternative	Dapsone plus	PO	50 mg QD	Also active in preventing toxoplasmosis
	Pyrimethamine plus	PO	50 mg once weekly	
	Leucovorin	PO	25 mg once weekly	Should be administered with pyrimethamine to minimize toxicity
Alternative	Atovaquone	PO	1,500 mg QD with food	Likely active in preventing toxoplasmosis
Alternative	Pentamidine	Aerosol	300 mg via the Respirgard II nebulizer once monthly	Not active in preventing toxoplasmosis

Note: Patients receiving pyrimethamine-sulfadiazine and atovaquone therapy for toxoplasmosis do not appear to need additional prophylaxis for PCP; patients receiving clindamycin-pyrimethamine therapy for toxoplasmosis will need additional prophylaxis for PCP. GGPD = glucose-6-phosphate dehydrogenase.

clearance after the drug is discontinued, sulfadoxine plus pyrimethamine should probably not be used in sulfa-tolerant patients if trimethoprim-sulfamethoxazole is available.

HIV-infected patients with a prior history of a mild sulfa allergy (e.g., mild rash, excluding those with prior Stevens-Johnson syndrome or toxic epidermal necrolysis) can often be safely rechallenged with trimethoprim-sulfamethoxazole. Randomized trials have demonstrated that dose escalation over a 6- to 13-day period is associated with better tolerance than direct rechallenge with full-dose trimethoprim-sulfamethoxazole, and that up to 75% of patients can continue to receive trimethoprim-sulfamethoxazole for at least 6 months.

PROGNOSIS

Mortality for untreated PCP approaches 100%. With therapy, the survival rate for HIV-infected patients with confirmed PCP is now as high as 95%, but a poorer survival rate of 75% has been reported in patients without HIV infection.[20] Risk factors for death in HIV-infected patients include more severe hypoxia, older age, recurrent episodes of PCP, low hemoglobin, and the presence of comorbid conditions. Although mortality for patients admitted to an intensive care unit is high, survival for HIV-infected patients has improved in recent years, now approaching 75%.[21]

Grade A References

A1. Briel M, Bucher HC, Boscacci R, et al. Adjunctive corticosteroids for *Pneumocystis jiroveci* pneumonia in patients with HIV-infection. *Cochrane Database Syst Rev*. 2006;3:CD006150.

A2. Stern A, Green H, Paul M, et al. Prophylaxis for Pneumocystis pneumonia (PCP) in non-HIV immunocompromised patients. *Cochrane Database Syst Rev*. 2014;10:CD005590.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Maini R, Henderson KL, Sheridan EA, et al. Increasing *Pneumocystis* pneumonia, England, UK, 2000-2010. *Emerg Infect Dis*. 2013;19:386-392.
2. Cisse OH, Pagni M, Hauser PM. De novo assembly of the *Pneumocystis jirovecii* genome from a single bronchoalveolar lavage fluid specimen from a patient. *mBio*. 2012;4:e00428-12.
3. Sassi M, Ripamonti C, Mueller NJ, et al. Outbreaks of *Pneumocystis* pneumonia in 2 renal transplant centers linked to a single strain of *Pneumocystis*: implications for transmission and virulence. *Clin Infect Dis*. 2012;54:1437-1444.
4. Watanabe K, Sakai R, Koike R, et al. Clinical characteristics and risk factors for *Pneumocystis jirovecii* pneumonia in patients with rheumatoid arthritis receiving adalimumab: a retrospective review and case-control study of 17 patients. *Mod Rheumatol*. 2013;23:1085-1093.
5. An P, Li R, Wang JM, et al. Role of exonic variation in chemokine receptor genes on AIDS: CCRL2 F167Y association with *Pneumocystis* pneumonia. *PLoS Genet*. 2011;7:e1002328.
6. Norris KA, Morris A. *Pneumocystis* infection and the pathogenesis of chronic obstructive pulmonary disease. *Immunol Res*. 2011;50:175-180.
7. Iriart X, Challan Belval T, Fillaux J, et al. Risk factors of pneumocystis pneumonia in solid organ recipients in the era of the common use of posttransplantation prophylaxis. *Am J Transplant*. 2015;15:190-199.
8. Vogel MN, Vatlach M, Weissgerber P, et al. HRCT-features of *Pneumocystis jiroveci* pneumonia and their evolution before and after treatment in non-HIV immunocompromised patients. *Eur J Radiol*. 2012;81:1315-1320.
9. Robert-Gangneux F, Belaz S, Revest M, et al. Diagnosis of *Pneumocystis jirovecii* pneumonia in immunocompromised patients by real-time PCR: a 4-year prospective study. *J Clin Microbiol*. 2014;52:3370-3376.
10. Hauser PM, Bille J, Lass-Florl C, et al. Multicenter, prospective clinical evaluation of respiratory samples from subjects at risk for *Pneumocystis jirovecii* infection by use of a commercial real-time PCR assay. *J Clin Microbiol*. 2011;49:1872-1878.
11. Cooley L, Dendle C, Wolf J, et al. Consensus guidelines for diagnosis, prophylaxis and management of *Pneumocystis jirovecii* pneumonia in patients with haematological and solid malignancies, 2014. *Intern Med J*. 2014;44:1350-1363.
12. Roux A, Gonzalez F, Roux M, et al. Update on pulmonary *Pneumocystis jirovecii* infection in non-HIV patients. *Med Mal Infect*. 2014;44:185-198.
13. Martin SI, Fishman JA. *Pneumocystis* pneumonia in solid organ transplantation. *Am J Transplant*. 2013;13(suppl 4):272-279.
14. Yoon C, Subramanian A, Chi A, et al. Dihydropteroate synthase mutations in *Pneumocystis* pneumonia: impact of applying different definitions of prophylaxis, mortality endpoints and mutant in a single cohort. *Med Mycol*. 2013;51:568-575.
15. Panel on Opportunistic Infections in HIV-Infected Adults and Adolescents. Guidelines for the prevention and treatment of opportunistic infections in HIV-infected adults and adolescents: recommendations from the Centers for Disease Control and Prevention, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. Available at http://aidsinfo.nih.gov/guidelines/html/4/adult-and-adolescent-oi-prevention-and-treatment-guidelines/0. Accessed January 28, 2015.
16. Panel on Antiretroviral Therapy and Medical Management of HIV-Infected Children. Guidelines for the use of antiretroviral agents in pediatric HIV infection. Available at http://aidsinfo.nih.gov/contentfiles/lvguidelines/pediatricguidelines.pdf. Accessed January 28, 2015.
17. Mocroft A, Reiss P, Kirk O, et al. Is it safe to discontinue primary *Pneumocystis jiroveci* pneumonia prophylaxis in patients with virologically suppressed HIV infection and a CD4 cell count <200 cells/microL? *Clin Infect Dis*. 2010;51:611-619.
18. Mori S, Sugimoto M. *Pneumocystis jirovecii* infection: an emerging threat to patients with rheumatoid arthritis. *Rheumatology (Oxford)*. 2012;51:2120-2130.
19. Long MD, Farraye FA, Okafor PN, et al. Increased risk of *Pneumocystis jiroveci* pneumonia among patients with inflammatory bowel disease. *Inflamm Bowel Dis*. 2013;19:1018-1024.
20. Roux A, Canet E, Valade S, et al. *Pneumocystis jirovecii* pneumonia in patients with or without AIDS, France. *Emerg Infect Dis*. 2014;20:1490-1497.
21. Barbier F, Roux A, Canet E, et al. Temporal trends in critical events complicating HIV infection: 1999-2010 multicentre cohort study in France. *Intensive Care Med*. 2014;40:1906-1915.

REVIEW QUESTIONS

1. A 47 year old previously healthy male comes into the emergency room complaining of a 2 week history of shortness of breath, non-productive cough, and fevers to 103° C. On physical exam he is tachypneic, with no rales or rhonchi, and has thrush on his buccal mucosa. His chest X-ray and pulse oximetry are normal. Which of the following is true?

A. The normal chest X-ray and pulse oximetry rule out *Pneumocystis* pneumonia.
B. A chest CT can help in the diagnosis of *Pneumocystis* pneumonia since it is almost invariably abnormal, even in the setting of a normal chest X-ray.
C. A positive PCR test for *Pneumocystis* definitively confirms the diagnosis of *Pneumocystis* pneumonia as the specificity of the PCR test for PCP approaches 100%.
D. Culture of induced, but not expectorated, sputum on Sabouraud agar will confirm the diagnosis of *Pneumocystis* pneumonia in 2 to 3 days.
E. Bronchoscopy with bronchoalveolar lavage is helpful if positive for *Pneumocystis* but has a low sensitivity (~40-50%); open lung biopsy is required to confirm the diagnosis of PCP in about 50% or patients.

Answer: B This patient likely has undiagnosed HIV infection and *Pneumocystis* pneumonia based on the symptoms and presence of thrush. Patients presenting with *Pneumocystis* pneumonia can have a normal chest x-ray (CXR) and normal pulse oximetry or arterial oxygenation at rest. Even in patients with a normal CXR, a CT scan of the chest will invariably show infiltrates, as illustrated in Figure 341-2. PCR-based assays have a high sensitivity but the specificity does not approach 100% as it can also be positive in individuals who are colonized or subclinically infected with *Pneumocystis*, but do not have *Pneumocystis* pneumonia and do not require specific anti-*Pneumocystis* therapy. Pneumocystis cannot be cultured, so definitive diagnosis depends on direct detection of the organism in a clinical sample, usually an induced sputum or bronchoalveolar lavage sample. Bronchoscopy with bronchoalveolar lavage has a >90% sensitivity for *Pneumocystis* pneumonia and is usually the definitive diagnostic procedure for *Pneumocystis* pneumonia. Prior to the AIDS epidemic and the demonstration of the utility of bronchoscopy, open lung biopsy was often required to make the diagnosis, but currently it is rarely necessary; negative results by bronchoscopy with BAL is sufficient to rule out *Pneumocystis* pneumonia in most circumstances.

2. A 25 year old female with a history of prior intravenous drug use comes in with a 1 week history of fever and a non-productive cough. CXR shows diffuse bilateral infiltrates. Resting PaO_2 at presentation is 96 mm Hg with an A-a gradient of 15 mm Hg. A rapid HIV test is positive. An induced sputum stain for *Pneumocystis* is positive, and the patient is started on trimethoprim-sulfamethoxazole, 2 double strength tables 3 times a day. She is sent home to be followed in clinic. Three days later she returns complaining of increasing shortness of breath with walking. Resting PaO_2 has decreased to 86 mm Hg, with an A-a gradient of 30 mm Hg. What is the most appropriate management course for this patient?

A. Discontinue the trimethoprim-sulfamethoxazole and begin pentamidine.
B. Discontinue the trimethoprim-sulfamethoxazole and begin clindamycin-primaquine.
C. Bronchoscope the patient to rule out other infections, and evaluate the patient for pulmonary emboli.
D. Begin prednisone 40 mg/day in addition to the trimethoprim-sulfamethoxazole.
E. Continue the trimethoprim-sulfamethoxazole; no change in therapy is needed.

Answer: E Deterioration in oxygenation at day 3 to 4 after start of therapy is common and expected in patients with PCP and does not signify clinical failure. Thus there is no reason to change therapy to an alternative regimen, or to search for alternative diagnoses at this point. While prednisone has been shown to decrease the frequency of deterioration in oxygenation in patients with PCP, a large randomized trial found that clinical benefit in terms of respiratory failure and survival outcome was limited to patients with moderate to severe disease, defined as a $PaO_2 < 70$ mm Hg or an A-a gradient >35 mm Hg. Moreover, addition of prednisone after 3 days of specific anti-*Pneumocystis* therapy has not been shown to be beneficial.

3. A 38 year old male with a recent history of thrush is admitted with a fever and non-productive cough. He is diagnosed with HIV infection and *Pneumocystis* pneumonia, and is started on trimethoprim-sulfamethoxazole. He would like to start anti-HIV medications soon. How should his antiretroviral therapy be managed?

A. Initiate HAART 2 months after discontinuation of anti-*Pneumocystis* therapy.
B. Initiate HAART immediately if HIV plasma viral load is >100,000 copies/ml, and 2 months after discontinuation of anti-*Pneumocystis* therapy if HIV plasma viral load is ≤100,000 copies/ml.
C. Timing of HAART should be based on the CD4 count. If the CD4 count is ≤350 cells/mm^3, HAART should be initiated within 1 month of completion of therapy. If the CD4 count is >350 cells/mm^3, therapy can be deferred until the cell count drops below that level.
D. Initiate HAART within 2 weeks of starting anti-*Pneumocystis* therapy if possible.
E. Don't initiate HAART until the CXR has normalized to minimize the risk of developing immune reconstitution inflammatory syndrome (IRIS).

Answer: D A randomized trial in patients with an opportunistic infection, 63% of whom had *Pneumocystis* pneumonia, found that initiation of HAART early (~12 days) vs. late (~45 days) after the start of therapy for the opportunistic infection was associated with a lower risk of AIDS progression or death. Patients with *Pneumocystis* pneumonia should be started on HAART regardless of the CD4 count or viral load, since they have already shown that their immune systems are severely compromised; without HAART or anti-*Pneumocystis* prophylaxis they are at high risk for recurrence of *Pneumocystis* pneumonia. Of note, most patients diagnosed with *Pneumocystis* pneumonia will have a CD4 count under 200 cells/mm^3. While cases of IRIS in the setting of *Pneumocystis* pneumonia have been reported, the incidence appears to be relatively low and rarely appears to be life-threatening; in the randomized trial noted above, the incidence of IRIS in patients with *Pneumocystis* pneumonia was 7.3%.

4. A patient who received a renal transplant 6 months ago and is receiving mycophenolate mofetil, cyclosporine, and prednisone for chronic immunosuppression is diagnosed with *Pneumocystis* pneumonia. This is the fourth renal transplant patient diagnosed with *Pneumocystis* pneumonia during the past year. Renal transplant patients at this center do not routinely receive anti-*Pneumocystis* prophylaxis because no cases of *Pneumocystis* pneumonia were seen in the prior ten years. What is the best management strategy for stopping this outbreak of *Pneumocystis* pneumonia?

A. Place the patient on respiratory isolation to prevent spread of the organism to other transplant patients.
B. Clean the walls, floors, and ceilings of the patient rooms in the clinic and inpatient service with bleach to eliminate the source of *Pneumocystis*.
C. Have the patients at risk for infection wash daily with a povidone-iodine antiseptic solution to minimize colonization with *Pneumocystis*.
D. Begin prophylaxis with trimethoprim-sulfamethoxazole in all transplant patients for at least 6 months, and longer for high risk patients (e.g. those requiring additional immunosuppression for graft rejection).
E. Begin prophylaxis with clindamycin-pyrimethamine in all transplant patients for at least 4 months, and longer for high risk patients (e.g. those requiring additional immunosuppression for graft rejection).

Answer: D Outbreaks of *Pneumocystis* pneumonia have been reported from multiple renal transplant centers over the past 10 years. As demonstrated in molecular typing studies, these outbreaks are caused by a single or a limited number of strains of *Pneumocystis* at each center. This strongly implies recent transmission of the infection, and potential acquisition as a nosocomial infection. Current data suggest that *Pneumocystis* is acquired from other humans via transmission by the respiratory route. There is no evidence for an environmental reservoir, thus cleaning a room will not eliminate the sources of infection. Although colonization or subclinical infection of the respiratory tract with *Pneumocystis* has been documented by sensitive PCR-based methods, there is no evidence that the skin is colonized, and thus washing with povidone-iodine antiseptic solution will not eliminate potential reservoirs for this specific organism. While it is logical to utilize respiratory isolation to prevent the spread of *Pneumocystis*, there are no data to demonstrate that this is an effective strategy to halt an outbreak. Infection may be spread during the incubation period, which in animal models, can be 8-12 weeks before the development of severe pneumonia. Clindamycin-pyrimethamine has not been demonstrated to have activity in either treatment or prophylaxis of *Pneumocystis* pneumonia. Trimethoprim-sulfamethoxazole is a highly effective prophylactic regimen and institution of prophylaxis with trimethoprim-sulfamethoxazole has been successful in halting the outbreaks of *Pneumocystis* pneumonia reported to date.

5. A 35 year-old female comes in with a headache and weakness in her right arm. An MRI scan shows multiple ring-enhancing lesions. Her anti-toxoplasma IgG antibody test is positive, but her IgM is negative. Cerebrospinal fluid is positive for *Toxoplasma* by PCR. Her HIV test is positive, and her baseline labs include a CD4 count of 27 cells/mm^3 and an HIV viral load of 42,500 copies/ml. The patient noted that she had received trimethoprim-sulfamethoxazole many years ago for a urinary tract infection and said she thought she developed a rash around her elbows but did complete her therapy. She is started on pyrimethamine-sulfadiazine-leucovorin for treatment of CNS toxoplasmosis. How should anti-*Pneumocystis* prophylaxis be managed in this patient?

A. No specific anti-*Pneumocystis* prophylaxis is required.
B. Begin dapsone 100 mg/day since she may have an allergy to trimethoprim-sulfamethoxazole.
C. Begin atovaquone 1,500 mg per day since she may have an allergy to trimethoprim-sulfamethoxazole and patients with sulfa allergies are frequently allergic to dapsone.
D. Begin azithromycin, 1,200 mg per week, to provide prophylaxis for both *Pneumocystis* and *Mycobacterium avium* complex.
E. Begin trimethoprim-sulfamethoxazole, one double-strength daily; change to dapsone if the patient develops a rash.

Answer: A This severely immunocompromised patient with HIV infection and toxoplasmosis is at high risk for *Pneumocystis* pneumonia based on her CD4 count. Although she describes a possible rash during prior treatment with trimethoprim-sulfamethoxazole, a rash limited to the elbows would be unusual for a drug reaction, and reactions to sulfamethoxazole, the component that appears to be most commonly responsible for reactions to trimethoprim-sulfamethoxazole, are rarely immediate or life-threatening. Thus therapy with a sulfa drug is not contraindicated, and the patient was started on pyrimethamine-sulfadiazine-leucovorin for treatment of toxoplasmosis. Pyrimethamine and sulfadiazine target the same enzymes as trimethoprim and sulfamethoxazole (dihydrofolate reductase and dihydropteroate synthase, respectively), and studies have shown that patients receiving pyrimethamine-sulfadiazine-leucovorin are at low risk for developing *Pneumocystis* pneumonia and thus do not require additional prophylaxis. While azithromycin prophylaxis should be started in this patient due to the low CD4 count, it has not been studied as an anti-*Pneumocystis* agent and thus additional prophylaxis for *Pneumocystis* would be required if it were indicated.

342

MYCETOMA

D. P. KONTOYIANNIS

DEFINITION

Mycetoma (a tumor produced by fungi) was first described in 1842 in the Madura district of India, hence the terms "Madura foot," "maduromycosis," and "maduromycetoma." However, there is evidence of its existence from as far back as the Byzantine era and ancient India.

Mycetoma is a chronic, slowly progressive infection that starts in the subcutaneous tissue and spreads across tissue planes to contiguous structures. The disease has diverse etiology; also, the offending organism is inoculated into the subcutaneous tissues by trauma typically associated with soil contamination. The hallmarks of mycetoma are the presence of "grains" that consist of colonies of the infectious organism and chronically draining sinus tracts. There is some confusion in the literature, however, because the term pulmonary mycetoma is used inappropriately to describe fungus balls typically caused by *Aspergillus* species that colonize a preexisting lung cavity; the term *aspergilloma* is more appropriate for this entity, the pathogenesis of which is distinctly different from that of true mycetoma.

The Pathogen

Two groups of soil-inhabiting pathogens, each of which accounts for approximately 50% of the cases, cause mycetoma: (1) the filamentous aerobic actinomycetes, hence the term actinomycetoma, and (2) a wide range of saprophytic soil and woody plant fungi, hence the term eumycetoma. Eumycetoma accounts for approximately 50% of cases of mycetoma.[1,2]

A variety of *Nocardia* species (e.g., *Nocardia brasiliensis, Nocardia asteroides*), *Actinomadura* species (e.g., *Actinomadura pelletierii, Actinomadura madurae*), and *Streptomyces* species (e.g., *Streptomyces somaliensis*) have been reported to cause actinomycetoma. Even more numerous are the agents that cause eumycetoma, such as *Madurella* species (e.g., *Madurella mycetomatis* causes 70% of all cases of eumycetoma,), which are probably the most prevalent mycetoma-causative fungal species worldwide. Other eumycetoma include *Fusarium* species, *Acremonium* species, *Pseudallescheria boydii, Exophiala* species, and *Curvularia* species. There is controversy about whether the various dermatophytes and *Aspergillus* species can cause mycetoma. Eumycetoma is often further characterized on the basis of the color of the grains; specifically, white- to yellow-grain mycetomas (white piedra) are typically caused by hyalohyphomycetes (e.g., *P. boydii*, *Fusarium* species, *Acremonium* species), and black-grain eumycetomas (black piedra) are caused by *Madurella* species and other less common fungi. However, the geographic distribution of the fungi that cause black grain eumycetoma is variable.

EPIDEMIOLOGY

Although mycetoma has a global distribution, it occurs primarily in the tropical and, to a lesser extent, the temperate zones. More specifically, the infection is quite prevalent in India, Mexico, Central America, South America, the Middle East, and especially sub-Saharan Africa (the "mycetoma belt"); Sudan has a particularly high burden of mycetoma. Indigenously acquired mycetoma is sporadic in North America and Europe. However, the globalization of tourism and the increase in immigration from countries of high endemicity of mycetoma to the western countries necessitates awareness of this entity, even in the developed word.

The relative frequency of actinomycetoma and eumycetoma differs among geographic areas. Hence, eumycetoma is more common in India and Africa, and actinomycetoma is more common in Central and South America. Furthermore, the causative agents of mycetoma differ in their geographic distribution. For example, *Scedosporium apiospermum* (*P. boydii*) is the most common agent of mycetoma in North America, and *Actinomadura* and *Nocardia* species are predominant in Central and South America. Finally, *Luidia senegalensis* and *M. mycetomatis* are predominant in sub-Saharan Africa and India (Fig. 342-1). The recent development of molecular typing procedures such as polymerase chain reaction restriction fragment length polymorphism holds promise in expanding our knowledge on the environmental sources and the pathogenesis of some agents of eumycetoma such as *M. mycetomatis*.

PATHOGENESIS

Local trauma (e.g., wood splinters) introduces a mycetoma-causative organism into the skin and subcutaneous tissues and initiates a chain of events that leads to chronic, suppurative granulomatous inflammation, tumefaction, formation of multiple fistulous tracts and sinuses, deep abscesses, fibrosis and scar formation, and extension to adjacent connective tissue across the lines of least resistance (fascia) and ultimately to bones, muscles, nerves, and tendon sheaths, leading to gross anatomic distortion of the affected site. In addition, a chronic suppurative granuloma featuring reactive fibrosis and grains (sclerotia), which is a matrix consisting of vegetative aggregates of the etiologic agents and host-derived inflammatory response, is characteristic of mycetoma in histologic sections. This infection is not contagious, however. Even though the genetics and immunopathogenesis of mycetoma are not well defined, it appears that there are differences in host susceptibility because some affected persons have impaired or delayed hypersensitivity reactions or polymorphisms in genes encoding for chemokines (e.g., CCL50) and cytokines (e.g., IL-10).[3] However, mycetoma does not appear to be more common in immunocompromised hosts. The lack of appropriate animal models that simulate the macroscopic features of subcutaneous human infection limits our understanding of the pathogenesis of mycetoma.

CLINICAL MANIFESTATIONS

The clinical manifestations and natural history of mycetoma are variable and, to some degree, related to the pathogenic agent involved.[4] For example, the progression of eumycetoma tends to be slower than that of actinomycetoma. In addition, eumycetoma lesions tend to be more confined and have less inflammation and fewer granulomas and fistulas but more fibrosis compared with actinomycetoma lesions. Furthermore, male mycetoma patients predominate (5 : 1 over female patients), and the disease is typically seen in rural areas and in persons susceptible to local trauma and contamination from soil (e.g., thorns). Hence, farmers, gardeners, woodcutters, herders, and people who work outside while barefoot are more susceptible to this infection. Not surprisingly, the foot is the most common site involved in mycetoma (Fig. 342-2), but any other part of the body, such as the hands, thighs, torso, and back of the head, may become involved. Location is typically solitary. Painless

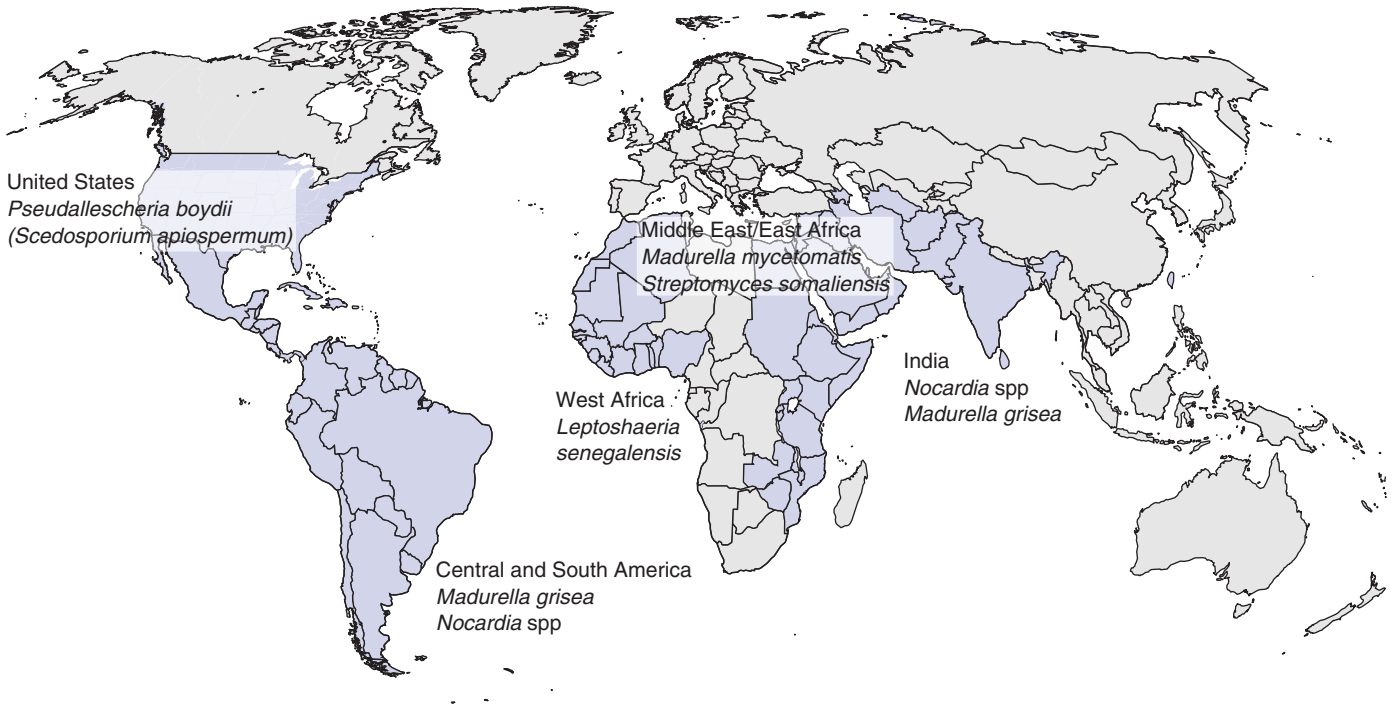

FIGURE 342-1. **Predominant agents of mycetoma according to region.**

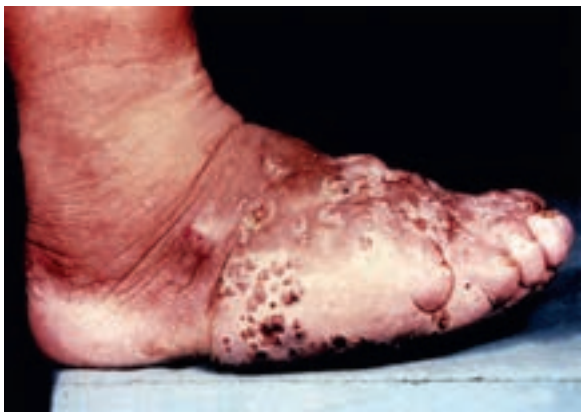

FIGURE 342-2. **Madura foot.** A 40-year-old farmer from rural Venezuela with a 10-year history of foot edema and slowly progressive deformity following an injury caused by being struck by a hammer presented with chronic crusted plaques, multiple confluent tender abscesses with fistulization, and release of black grains. The range of motion of the patient's ankle and foot joints was limited, but the joints were not painful. A deep skin biopsy with hematoxylin and eosin staining, periodic acid–Schiff staining, foot radiography, sampling of black grain smears, and a mycology culture were performed. The foot radiograph showed osteofibrosis, destruction of articular surfaces, osteoporosis, and ankylosis, and *Madurella* species grew in the culture. (Courtesy Dr. M. Mendoza, Instituto De Biomedicina, Laboratorio de Micología, San Jose Caracas, Venezuela.)

nodular and/or papular swelling is the most common early manifestation of mycetoma, which is followed by a slow evolution to painless, fixed woody induration. This infection typically runs a chronic, relentless course, sometimes spanning several decades. It is characterized by recurring, vicious cycles of suppuration, draining sinuses, bacterial superinfection, and scar formation. Old sinuses may close up, but new ones may occur, and satellite lesions may be seen. Constitutional symptoms are surprisingly rare. In particular, the presence of fever indicates bacterial superinfection. Bone involvement mimicking clinically chronic osteomyelitis with osteolytic cavitary bone lesions, periosteal reaction or sclerosis (seen with radiography, computed tomography [CT], or magnetic resonance imaging [MRI] studies), osteoporosis, and reactive periosteal bone formation may occur; such involvement can be substantial. However, pathologic fractures are rare. In addition, because nerves are relatively spared from involvement, neuropathic manifestations are uncommon. Inexorable limb deformity and misuse because of destruction of deeper tissues may be seen in chronic, refractory, and advanced cases. Finally, because mycetoma does not spread hematogenously, visceral dissemination is not seen. However, because lymphatic spread may occur (typical incidence of 1 to 3% but more common with actinomycetoma and especially after surgery), regional lymphadenitis may develop.

DIAGNOSIS

Mycetoma, especially in its advanced forms, has a rather characteristic presentation. The classic triad of painless soft tissue swelling, draining sinus tracts, and extrusion of grains facilitates diagnosis with a high degree of accuracy, especially in endemic areas. For instance, the macroscopic and microscopic appearance of grains in pus-filled draining sinuses frequently allows presumptive diagnosis of the offending pathogen. However, basing the diagnosis on the presence of grains in tissue may be difficult because these grains might be composed of dead organisms. Furthermore, grains may be contaminated by surface bacteria or fungi. Therefore, a deep-tissue biopsy specimen is ideal for staining with hematoxylin and eosin, and appropriate selective bacterial (e.g., Löwenstein-Jensen culture medium) and fungal (e.g., blood agar and modified Sabouraud dextrose agar with antibiotics) cultures and stains (Gram stain, modified Ziehl-Neelsen stain, Gomori methenamine silver, periodic acid–Schiff stain) are preferable for primary detection. Alternatively, aspiration of grains from unopened sinus tracts might provide material for culture.

The culture should be maintained for several weeks because some of the causative agents of mycetoma (e.g., *Nocardia* and *Streptomyces* species) are slowly growing and can take 4 to 6 weeks to grow. Histopathology differentiates eumycetoma from actinomycetoma; however, the multitude of fungi causing eumycetoma necessitates culture identification because agents of eumycetoma might respond differently to antifungal agents. Similarly, cross-reactivity, lack of standardization, and the multitude of agents causing eumycetoma limit the practical value of serology. Finally, there are no studies correlating in vitro susceptibility testing with outcome in either actinomycetoma or eumycetoma.

Differential Diagnosis

The specific manifestations of mycetoma are sometimes confused with other entities.[5] For example, early lesions could be confused with soft tissue neoplasms or foreign body granulomas. In addition, mycetoma without fistulas must occasionally be distinguished from chronic cutaneous fungal infections such as sporotrichosis (mycetomatous lymphatic sporotrichosis) and dermatophytic mycetoma. The latter infection, which is typically seen in Africans and sometimes called pseudomycetoma, is a painless granulomatous

induration of the skin and subcutaneous tissues caused by ringworm that may be associated with grains consisting of fungi. However, unlike mycetoma, dermatophytic mycetoma is confined to the skin and subcutaneous tissue and does not spread to fasciae or bone. Similarly, chronic severe botryomycosis (typically caused by gram-positive cocci) with purulent exudates, grains, and draining sinus tracts may be confused with mycetoma; however, the presence of visceral dissemination supports a diagnosis of (severe) botryomycosis. Actinomycosis (Chapter 329), which is caused by endogenous microaerophilic actinomycetes (part of normal mucosa flora), also has a propensity for grains and formation of draining sinus tracts, but unlike mycetoma, its location (e.g., neck, chest, pelvis) is rather characteristic. In addition, differentiation between mycetomas with bone involvement and chronic osteomyelitis or osseous tumors may be difficult. Ultrasonography has been used to reliably differentiate mycetoma from either tumor or osteomyelitis. The dot-in-circles sign seen at MRI (tiny hypodense foci, believed to be grains, within high-hyperintensity spherical lesions, believed to be granulomas scattered by areas of fibrosis) might provide an early and specific diagnostic clue for mycetoma, Finally, in cases of mycetoma without draining sinus tracts, benign or malignant skin tumors, chronic granulomatous lesions (e.g., thorn granuloma, cutaneous tuberculosis), chromomycosis, and verrucous leishmaniasis are diagnoses that should be excluded. Unfortunately, delays of specific diagnosis by many months or even years is not uncommon, even in areas of high endemicity.

TREATMENT

Rx

Therapy for mycetoma should be individualized.[6] Optimal management has not been well defined, however, because the literature consists of rather heterogeneous and uncontrolled small studies. There is no "gold standard" approach for therapy, and no single agent is effective against all causative agents of mycetoma. Hence, successful treatment necessitates a reliable diagnosis, differentiation between actinomycetoma and eumycetoma, assessment of the extent of the lesion, and identification of the causative agent. Specifically, the degree of tissue invasion, especially bone involvement (as determined by radiology studies), site affected, and specific etiologic diagnosis determine the type and intensity of therapy. Recent studies indicate that MRI and CT (more sensitive especially for mycetoma involving bone) are especially useful in "staging" the disease.

In general, considering the refractoriness of eumycetoma to medical therapy, surgery plays a more prominent role.[7] In contrast, considering the satisfactory response of actinomycetoma to medical therapy (success rate of up to 90%) and recognized risk for lymphatic spread following surgery, chronic antibiotic administration is the mainstay of actinomycetoma management.

For actinomycetoma, treatment consists primarily of chronic antibiotic therapy (for at least 9 to 12 months) in conjunction with limited debulking surgery in selected cases; combination therapy designed for potential synergy is preferred. A variety of drugs (trimethoprim-sulfamethoxazole, tetracyclines, dapsone, streptomycin) have been used in different sequences and combinations according to the specific cause of actinomycetoma (e.g., trimethoprim-sulfamethoxazole with or without dapsone for *Nocardia* species, streptomycin with dapsone for *A. madurae*). Parenteral streptomycin is usually reserved for cases that do not respond to oral therapy. Responses to these drugs tend to occur slowly (within at least 1 month). Also, relapse is not uncommon, and multiple cycles of therapy may be needed for chronically recurrent disease. Side effects and compliance issues after prolonged administration of antibiotics are common problems.

For eumycetoma, however, medical therapy has produced mixed results. The best results were obtained with prolonged (9 to 12 months) use of oral imidazoles (e.g., ketoconazole 200 to 400 mg/day or itraconazole 200 to 400 mg/day). The experience using newer triazoles (e.g., voriconazole or posaconazole) or the allylamine terbinafine (with or without concomitant itraconazole), although encouraging, is not as extensive, however.[8] Posaconazole or voriconazole might prove to be especially useful for treatment of eumycetoma due to *S. apiospermum (P. boydii),* a fungus that is not susceptible to either ketoconazole or itraconazole. Also, intravenous amphotericin B and its lipid formulations have been used for refractory cases with rather disappointing results. This is not surprising because most eumycetoma-causative agents (e.g., *P. boydii*) are resistant to amphotericin B in vitro.

The need for and extent of surgery for mycetoma depend on the etiologic agent and, more important, extent of the lesion. Early wide-margin surgery for early localized lesions is curative. Although it is potentially curative, major disfiguring or mutilating surgery (e.g., amputation) is reserved for very advanced or refractory cases. Furthermore, primary reliance on surgery could result in recurrence or even spreading of the disease because of incomplete excision.

The prognosis for mycetoma depends on the site and degree of tissue involvement (e.g., worse with involvement of the back because of poor healing or in the presence of bone destruction) and, more important, the timeliness of the diagnosis and monitoring of recurrence and extension to other tissues, especially bone. CT and MRI are important modalities for early detection of bone involvement.

Finally, since mycetoma represents a significant, yet frequently ignored, socioeconomic burden for tropical and subtropical countries,[1] every effort for its prevention should be made. There is vaccine for that entity, and education of persons at risk in endemic areas is crucial (e.g., avoidance of walking barefoot). Finally, prompt recognition for early curative chemotherapy and/or surgery could lead to a better outcome of this difficult-to-manage, chronic infection.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Van Belkum A, Fahal A, van de Sande WW. Mycetoma caused by *Madurella mycetomatis*: a completely neglected medico-social problem. *Adv Exp Med Biol*. 2013;764:179-189.
2. Estrada R, Chavez-Lopez G, Estrada-Chavez G, et al. Eumycetoma. *Clin Dermatol*. 2012;30: 389-396.
3. Mhmoud NA, Fahal AH, van de Sande WW. The association between the interleukon-10 cytokine and CC chemokine ligand 5 polymorphisms and mycetoma granuloma formation. *Med Mycol*. 2013;51:527-533.
4. Bonifaz A, Tirado-Sánchez A, Calderón L, et al. Mycetoma: experience of 482 cases in a single center in Mexico. *PLoS Negl Trop Dis*. 2014;8:e3102.
5. van de Sande WW, Fahal AH, Goodfellow M, et al. Merits and pitfalls of currently used diagnostic tools in mycetoma. *PLoS Negl Trop Dis*. 2014;8:e2918.
6. Welsh O, Al-Abdely HM, Salinas-Carmona MC, et al. Mycetoma medical therapy. *PLoS Negl Trop Dis*. 2014;8:e3218.
7. Fahal AH, Shaheen S, Jones DH. The orthopaedic aspects of mycetoma. *Bone Joint J*. 2014;96B(3): 420-425.
8. Chowdhary A, Meis JF, Guarro J, et al. ESCMID and ECMM joint clinical guidelines for the diagnosis and management of systemic phaeohyphomycosis: diseases caused by black fungi. *Clin Microbiol Infect*. 2014;20(suppl 3):47-75.

REVIEW QUESTIONS

1. All is true for eumycetoma *except* which one of the following?

A. The condition is most commonly found in tropical and subtropical regions.
B. Dissemination of subcutaneous lesion is common.
C. Examination of grains from draining sinuses can aid in identification of the offending fungus.
D. Local trauma and exposure to soil are typical initiating factors.

Answer: B Local trauma (e.g., wood splinters) introduces a mycetoma-causative fungus into the skin and subcutaneous tissues and initiates a chain of events that leads to chronic, suppurative granulomatous inflammation, tumefaction, and formation of multiple fistulous draining tracts in which fungal colonies in the form of grains are found. The color and morphologic characteristics of these grains could contribute to the identification of the offending fungal pathogen. Eumycetoma has a global distribution, but it is more common in areas with tropical or subtropical climate. Although extensive soft tissue involvement and adjacent bone involvement can be seen in advanced cases, visceral dissemination is typically absent (Estrada R, Chavez-Lopez G, Estrada-Chavez G, Lopez-Martinez R, Welsh O. Eumycetoma. *Clin Dermatol.* 2012;30:389-396).

2. Which antifungal has in vitro activity against *Scedosporium apiospermum (Pseudallescheria boydii)*?

A. Terbinafine
B. Fluconazole
C. Amphotericin B
D. Posaconazole
E. Ketoconazole

Answer: D *S. apiospermum (P. boydii)* is a cause of eumycetoma for which therapeutic options are limited because the fungus is resistant to most antifungals with the exception of the new triazoles, posaconazole and voriconazole (Negroni R, Tobon A, Bustamante B, et al. Posaconazole treatment of refractory eumycetoma and chromoblastomycosis. *Rev Inst Med Trop Sao Paulo.* 2005;47:339-346).

343 DEMATIACEOUS FUNGAL INFECTIONS

PETER G. PAPPAS

DEFINITION

Dematiaceous fungi represent a large group of fungal organisms characterized by the presence of abundant melanin in the cell wall, which gives rise to a brown-black coloration on artificial culture media and which can be seen on histopathologic specimens. A related term, *phaeohyphomycosis,* refers broadly to infection by these pigmented fungi. The two terms are often used interchangeably, but when dematiaceous fungi are reviewed, three distinct clinical conditions are encountered: eumycetoma (e.g., Madura foot), chromomycosis (also known as chromoblastomycosis), and phaeohyphomycosis. Eumycetoma are covered in Chapter 342. This chapter focuses on the latter two entities.

The Pathogens

More than 100 dematiaceous fungi have been identified as causes of colonization or disease in humans. The most common organisms and their related conditions are listed in Table 343-1. The taxonomy of the dematiaceous fungi is somewhat confusing because these agents belong to different classes, including Hyphomycetes, Ascomycetes, Basidiomycetes, Coelomycetes, and Zygomycetes. The most common agents of phaeohyphomycosis include species in the following genera: *Bipolaris, Curvularia, Exophiala, Cladosporium, Cladophialophora, Alternaria, Exserohilum, Ochroconis, Wangiella, Phialophora, Scedosporium, Phaeoacromonium,* and *Chaetomium.* These agents are ubiquitous saprophytes of soil and decaying matter, and some are important plant pathogens. In tissue, these organisms exist as yeastlike cells, septated hyphae, or a combination of yeast and hyphae. Many have a histologic appearance similar to *Aspergillus* and *Fusarium* species, but they can be easily distinguished on the basis of positive melanin staining with the Fontana-Masson procedure.

Chromomycosis (formerly known as chromoblastomycosis) is a chronic skin and subcutaneous infection that is observed most frequently in the tropics. Virtually all cases of chromomycosis are caused by three species: *Fonsecaea pedrosoi, Cladosproium carrionii,* and *Phialophora verrucosa.* The distinctive histologic appearance is characterized by the presence of thick-walled, dark brown bodies known as *sclerotic cells* or *copper pennies,* which represent individual organisms and may be seen in clusters or as single cells. The etiologic fungi causing chromomycosis are indistinguishable on histologic examination of tissue.

EPIDEMIOLOGY AND PATHOBIOLOGY

The agents of chromomycosis and phaeohyphomycosis are found worldwide. Although there is no unique endemic area for most of these infections, some

TABLE 343-1 DEMATIACEOUS FUNGI AND ASSOCIATED DISEASES

CLINICAL CONDITION	COMMON ETIOLOGIC AGENTS
Chromomycosis	*Fonsecaea pedrosoi* *Cladophialophora carrionii* *Phialophora verrucosa*
Cutaneous or subcutaneous disease	*Exophiala jeanselmei* *Wangiella dermatitidis* *Phialophora* spp *Bipolaris* spp *Alternaria* spp
Sinusitis	*Bipolaris* spp *Curvularia* spp *Exserohilum* spp *Alternaria* spp
Central nervous system	*Cladophialophora bantiana* *Ochroconis gallopavum* *Rhinocladiella mackenziei* *Chaetomium altrobrunnium*
Healthcare-associated	*Exserohilum rostratum* *Exophilia* spp
Disseminated	*Wangiella dermatitidis* *Exophiala jeanselmei* *Bipolaris* spp *Ochroconis gallopavum* *Phialophora* spp *Scedosporium prolificans*

observations are relevant. Allergic fungal sinusitis associated with dematiaceous fungi appears to be more common in the southern United States. Chronic infections of the lower extremities are more commonly seen in men and in tropical areas. Chromomycosis is more prevalent in rural populations in the tropics and is hyperendemic in certain geographic areas such as Madagascar, Brazil, and other Latin American countries, where most infections are caused by *Fonsecaea pedrosoi* and *Cladosporium carrionii.*[1]

Most cutaneous infections occur as a result of minor skin trauma and direct inoculation of the organism. Other risk factors include intravenous drug abuse, chronic sinusitis, freshwater immersion, and chronic corticosteroid therapy.

In the developed world, phaeohyphomycosis is an important emerging fungal infection, particularly among immunocompromised patients such as solid organ and hematopoietic stem cell transplant recipients, patients with prolonged neutropenia, and other chronically immunocompromised individuals. Phaeohyphomycosis is reported in HIV-infected patients but is far less common than other opportunistic fungi. Extracutaneous invasive disease occurs in otherwise normal patients but is much less common.

The recent U.S. epidemic of fungal meningitis, epidural abscess, sacroiliitis, vertebral osteomyelitis, discitis, and peripheral arthritis caused by *Exserohilum rostratum* following injection of contaminated methylprednisolone acetate from a single compounding pharmacy is a striking example of the risk for dematiaceous fungal infections following invasive procedures in the health care setting.[2-4] In this epidemic, more than 750 persons with injection associated infection were identified, of whom almost 10% died as a direct result of the infection. Previous reports of infection due to *Exophiala* species following contaminated steroid injections, infected breast implants, other prosthetic materials, and, rarely, contaminated intravascular catheters and intravenous fluids further underscore the importance of these pathogens as potential health care–associated infections.

CLINICAL MANIFESTATIONS

Chromomycosis is manifested as a cutaneous or subcutaneous lesion that may range in size from a small papule to a large confluent plaque involving a major portion of an extremity. Single or multiple lesions may be seen, and ulceration may occur. Lesions may remain unchanged in size and consistency for months or years, although most tend to progress in the absence of specific therapy. Chronic lesions may become dry and crusted with a raised border, which may be smooth or irregular. Multiple lesions can coalesce to form larger plaques in which central scarring may develop. Occasionally, lesions assume a verrucous, warty appearance. The differential diagnosis includes other fungal infections such as blastomycosis, coccidioidomycosis, sporotrichosis, histoplasmosis, and paracoccidioidomycosis. Nocardiosis and cutaneous mycobacteriosis can also mimic the lesions of chromomycosis. Cutaneous lesions usually remain confined to one anatomic site, although nodular lymphangitis and autoinoculation resulting in multifocal cutaneous disease may occur. Common complications include local disfigurement due to scarring and extensive tissue involvement. Disseminated disease involving visceral organs may occur, but this is rare.

Phaeohyphomycosis is associated with several well-described clinical syndromes. *Superficial* infection is characterized by tinea nigra and black piedra. Tinea nigra is a darkening of the skin caused by growth of *Phaeoannellomyces werneckii* in the stratum corneum. Black piedra is associated with the development of focal thickening on the hair shaft and results from colonization of the shaft by *Piedraia hortae. Cutaneous* phaeohyphomycosis involves deeper skin structures and results in dermatomycosis and onychomycosis; this is frequently due to agents such as *Scytalidium* and *Phyllosticta* species.

Subcutaneous phaeohyphomycosis is relatively common and may be confused with chromomycosis. Patients have discrete subcutaneous nodules or cysts that result from direct inoculation or penetrating trauma. The most common organisms are *Exophiala jeanselmei, Wangiella dermatitidis, and Phialophora* species. *Mycotic keratitis* as a result of infection with *Curvularia, Exophiala,* and *Exserohilum* species may occur after corneal trauma or surgery.

Foreign body–related infections are seen in patients undergoing chronic ambulatory peritoneal dialysis in whom fungal peritonitis develops, in patients with indwelling intravenous catheters, and rarely in other devices such as breast implants.

Fungal sinusitis is commonly associated with dematiaceous fungi and can be manifested as allergic fungal sinusitis, a fungus ball (eumycetoma) in a sinus cavity, and invasive fungal sinusitis associated with extension into bone, soft tissue, and the central nervous system.[5] This latter manifestation is indistinguishable from rhinocerebral zygomycosis or invasive *Aspergillus* sinusitis. *Bipolaris, Curvularia,* and *Alternaria* species are the most common organisms causing invasive fungal sinusitis.

Systemic phaeohyphomycosis may result from direct extension from a colonized area or dissemination from a distant source. Most patients with systemic disease have significant underlying immunosuppression, and the organisms have a proclivity for involvement of the brain, lungs, endocardium, and other visceral organs. Among patients with primary central nervous system disease, *Cladophialophora bantiana, Ochroconis gallopavum, Rhinocladiella mackenziei,* and *Chaetomium altrobrunnium* are the most common etiologic agents, and the majority of these are otherwise healthy patients with no known underlying immunodeficiency. Among immunocompromised patients, *Ochroconis gallopavum*, *Bipolaris* species, and *Exophiala* species are seen more commonly.

A unique form of systemic phaeohyphomycosis has been seen recently in the fungal meningitis epidemic due to *Exserohilum rostratum* from contaminated methylprednisolone acetate injections as mentioned earlier. Patients with the meningeal form of this infection presented with symptoms ranging from local symptoms due to epidural abscess to devastating complications such as severe meningitis and basilar infarctions. Persons with extraneural involvement included those with focal sacroiliitis and peripheral joint arthritis.[6]

DIAGNOSIS

The diagnosis of phaeohyphomycosis is suggested by direct examination of a clinical specimen with a 10% potassium hydroxide preparation or special stains to demonstrate pigmentation in the cell walls of these organisms. For patients with chromomycosis, the finding of sclerotic cells or copper pennies on skin biopsy is characteristic, and special stains are usually unnecessary. For patients with other forms of phaeohyphomycosis, the Fontana-Masson stain is useful in distinguishing organisms with significant melanin content. Culture remains the "gold standard" by which a specific etiologic diagnosis is established, and the identity of the organism is largely based on colony and microscopic morphology. A polymerase chain reaction–based diagnostic assay for *Exserohilum rostratum* was developed in the context of the fungal meningitis outbreak, and it has served as a reliable marker of infection in patients who were exposed to contaminated steroid injections.[7] Serologic studies and molecular diagnostics for other organisms are not generally available.

TREATMENT

Rx

For chromomycosis, surgical excision of a cutaneous or subcutaneous lesion is often curative, although antifungal therapy is usually given in conjunction with surgery.[8] There are limited clinical studies and no large randomized trials that have assessed the efficacy of antifungal therapy for this condition. Historically, 5-flucytosine (5-FC, 150 mg/kg/day) has been advocated for the oral treatment of chromomycosis based on moderate in vitro activity and clinical experience. Because of limited availability, the need for prolonged therapy, and the necessity of monitoring serum levels, 5-FC is uncommonly used for this purpose. The triazoles, including itraconazole (200 mg orally twice daily), voriconazole (200 mg twice daily), and posaconazole (300 mg daily), demonstrate the best in vitro activity, although clinical studies with these agents are very limited. Terbinafine (500 mg orally twice daily) has also been used successfully for the treatment of chromomycosis and is an effective alternative to azole therapy.

Amphotericin B has modest in vitro activity against most of the dematiaceous fungi and is most often reserved for patients with life-threatening or disseminated disease. Among the triazoles, posaconazole offers the most potential in the clinical setting based on scattered reports from patients with central nervous system infection, but comparative clinical data are not available because of the relative rarity of these infections. Most patients with epidemic *Exserohilum rostratum* infections have been treated with voriconazole (intravenous, then oral) with or without a lipid formulation of amphotericin B for severe central nervous system involvement, and anecdotal reports suggest that this approach has been generally successful. The length of antifungal therapy for any of the systemic dematiaceous fungal infections is unclear but should probably be continued for at least 6 months or until 1 month after resolution of all signs and symptoms of disease.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Revankar SG, Sutton DA. Melanized fungi in human disease. *Clin Microbiol Rev*. 2010;23:884-928.
2. Kainer MA, Reagan DR, Nguyen DB, et al. Fungal infections associated with contaminated methylprednisolone in Tennessee. *N Engl J Med*. 2012;367:2194-2203.
3. Chiller TM, Roy M, Nguyen D, et al. Clinical findings for fungal infections caused by methylprednisolone injections. *N Engl J Med*. 2013;369:1610-1619.
4. Smith RM, Schaefer MK, Kainer MA, et al. Fungal infections associated with contaminated methylprednisolone injections. *N Engl J Med*. 2013;1598-1609.
5. McCarthy M, Rosengart A, Schuetz AN, et al. Mold infections of the central nervous system. *N Engl J Med*. 2014;371:150-160.
6. Ritter JM, Muehlenbachs A, Blau DM, et al. Exserohilum infections associated with contaminated steroid injections: a clinicopathologic review of 40 cases. *Am J Pathol*. 2013;183:881-892.
7. Centers for Disease Control and Prevention (CDC). Spinal and paraspinal infections associated with contaminated methylprednisolone acetate injections: Michigan, 2012-2013. *MMWR Morb Mortal Wkly Rep*. 2013;62:377-381.
8. Chowdhary A, Meis JF, Guarro J, et al. ESCMID and ECMM joint clinical guidelines for the diagnosis and management of systemic phaeohyphomycosis: diseases caused by black fungi. *Clin Microbiol Infect*. 2014;20(suppl 3):47-75.

REVIEW QUESTIONS

1. Which of the following organisms is most commonly associated with central nervous system infections?

A. *Exserohilum rostratum*
B. *Alternaria alternans*
C. *Scedosporium prolificans*
D. *Cladophialophora bantiana*
E. *Exophiala* spp

Answer: D *Cladophialophora bantiana* is commonly associated with central nervous system infections, especially brain abscess, in normal hosts and in immunocompromised hosts. The other organisms frequently associated with central nervous system infections include *Ochroconis gallopavum* and *Rhinocladiella mackenziei*.

2. Which of the following organisms was associated with the U.S. epidemic involving injectable methylprednisolone acetate, leading to hundreds of cases of meningitis, parameningeal infections, and peripheral joint infections?

A. *Exserohilum* spp
B. *Alternaria* spp
C. *Curvularia* spp
D. *Bipolaris* spp
E. *Fonsecaea* spp

Answer: A The major cause of the recently reported U.S. epidemic of contaminated methylprednisolone injections was *Exserohilum rostratum*. This organism was isolated by culture and/or polymerase chain reaction in the majority of patients in for whom positive tests results were available. The epidemic led to 751 proven and probable cases, of which approximately 8% died as a direct result of the infection.

3. Which of the following statement is *not* true concerning dematiaceous fungi?

A. These organisms contain high concentrations of melanin in the cell wall.
B. These organisms belong to the same class.
C. These organisms are unusual causes of invasive human infection.
D. These organisms are commonly found in the environment.
E. More than 100 of these organisms have been associated with invasive human disease or colonization.

Answer: B These organisms belong to several classes of fungi. More than 100 dematiaceous fungi have been associated with human disease or colonization. These organisms are ubiquitous in the environment, and most have the capability to cause invasive disease in the appropriate clinical setting.

4. Which of the following features of chromomycosis is true?

A. This disorder is characterized by discreet nodular skin and/or subcutaneous lesions.
B. The infection often follows cutaneous inoculation through minor trauma.
C. Most disease is caused by three organisms: *Fonsecaea*, *Cladosporium,* and *Phialophora*.
D. The characteristic histologic findings are sclerotic cells or "copper pennies."
E. All of the above are true.

Answer: E Chromomycosis is a disease largely seen in the tropics and follows minor skin trauma. The lesion is usually unilateral and is typically confined to an extremity. The lesions have a characteristic nodular, cutaneous, or subcutaneous appearance, and a histologic finding of sclerotic cells or copper pennies is characteristic of this condition. Chromomycosis is usually caused by one of the three organisms listed.

5. Which of the following is the best choice for therapy for chromomycosis?

A. Fluconazole
B. Amphotericin B
C. Flucytosine
D. Miconazole topical cream
E. Itraconazole

Answer: E Among the choices given, itraconazole is the most effective agent. Among the azoles, posaconazole and voriconazole have excellent in vitro activity against most of the dematiaceous fungi, including those that cause chromomycosis. Small clinical trials evaluating these azoles and terbinafine suggest that each of these agents is effective when administered for an appropriate duration of treatment, usually months or even years.

344

ANTIPARASITIC THERAPY

RICHARD D. PEARSON

International travel, widespread immigration, and a growing number of persons with compromised immunity due to HIV/AIDS, organ transplants, corticosteroids and other conditions have resulted in increased attention to parasitic diseases worldwide. A number of drugs are available to treat them, but physicians practicing in industrialized countries often are not familiar with their use. The focus of this chapter is on their therapeutic indications, pharmacology, and major side effects. Generalizations emerge that help in organizing an otherwise vast amount of information.

In considering the chemotherapy of parasitic diseases, it is helpful to classify infections into those caused by helminths, multicellular worms with complex internal structures, and protozoa, single-celled organisms that multiply by cell division. Helminths are further divided into nematodes, or roundworms, which can be grouped into those that live in the gastrointestinal tract and those found elsewhere in the body, and platyhelminths, or flat worms, which are subdivided into cestodes, or tapeworms, and trematodes, or flukes.

The protozoa can be grouped into those that reside under anaerobic conditions in the gastrointestinal tract or vagina and those that live aerobically in the body. The most prevalent systemic protozoan infections are attributable to members of the Apicomplexa, which cause malaria, babesiosis, and toxoplasmosis, and the Kinetoplastida, which are responsible for Chagas disease, human African trypanosomiasis (sleeping sickness), and leishmaniasis.

The Centers for Disease Control and Prevention (CDC) provide detailed information about the diagnosis and treatment of parasitic diseases through their website (www.cdc.gov). An excellent compilation of therapeutic recommendations is available in tabular form in *The Medical Letter on Drugs and Therapeutics*, "Drugs for Parasitic Infections" (available at www.medicalletter.org). Many antiparasitic drugs are commercially available in the United States, whereas others can only be obtained from the manufacturer, special pharmacies, or the CDC Drug Service. Some are available through investigational new drug protocols. Several drugs that are used for the treatment of parasitic diseases are also active against common bacterial or fungal pathogens. They are discussed elsewhere (Chapters 287 and 331).

TREATMENT OF HELMINTHS

Intestinal Roundworms (Nematodes)

Soil-transmitted intestinal roundworms (Chapter 357) are among the world's most prevalent parasites, and there are good therapeutic options for them. *Ascaris lumbricoides*, the hookworms *Ancylostoma duodenale* and *Necator americanus*, and *Trichuris trichiura* each infect on the order of 1 billion people worldwide. Many residents of impoverished areas harbor more than one soil-transmitted pathogen.

Albendazole has a broad spectrum of activity against intestinal roundworms. It is active against *A. lumbricoides*, the hookworms, and *T. trichiura*. Administered as a single 400-mg dose, it has been used successfully in mass treatment programs for children living in high prevalence areas. However, reinfection is common, and treatment is often repeated at 3- to 4-month intervals. The CDC recommends presumptive treatment of refugees from endemic regions with a single dose of albendazole (600 mg), administered overseas before departure for the United States. It has been highly effective and well tolerated in refugees from Africa and Southeast Asia.[1] Daily doses of albendazole for 3 days are recommended for persons with heavy *T. trichiura* infection. Twice-daily doses of albendazole, 600 mg for 7 days, are used as an alternative to ivermectin for the treatment of *Strongyloides stercoralis* infection. Failures can occur with either drug, and they are often used together for longer periods to treat those with disseminated hyperinfection. Albendazole is effective against pinworms in a single dose that is repeated in 2 weeks. It can be used for cutaneous larva migrans, which is caused by migrating stages of *Ancylostoma braziliense* and other intestinal helminths of animals. It is the drug of choice for trichinosis, and it is an alternative for the treatment of *Trichostrongylus* species and *Capillaria philippinensis*.

Mebendazole, administered twice daily at a dosage of 100 mg orally for 3 days, has a similar spectrum of activity as albendazole against *A. lumbricoides*, hookworms, and *T. trichiura*. In this regimen, it is more effective than a single dose of albendazole and is considered the treatment of choice for *T. trichiura*. A single 500-mg dose of mebendazole has been used in mass treatment programs. Mebendazole is effective in treating pinworm when given at 100 mg orally in one dose followed by a second dose after 2 weeks. It is an alternative to albendazole for the treatment of trichinosis. Mebendazole is poorly absorbed and not effective against *S. stercoralis*.

Pyrantel pamoate is a relatively safe, poorly absorbed, over-the-counter drug with activity against *A. lumbricoides*, hookworms, and pinworms, but it is not effective against *T. trichiura* or *S. stercoralis*. When used for pinworms, it is administered as an oral suspension at a dose of 11 mg/kg (to a maximum of 1 g), which is repeated after 2 weeks. The combination of ***oxantel pamoate***, 20 mg/kg, and ***albendazole***, 400 mg, taken on consecutive days, has been recently found to result in higher cure and egg-reduction rates for *T. trichiura* infection then the rates with standard therapy.

Ivermectin at an oral dose of 200 µg/kg daily for 2 days is considered the treatment of choice for *S. stercoralis*.[2] It is also effective against cutaneous larva migrans and *A. lumbricoides*, but not hookworms. It is considered an alternative to mebendazole for the treatment of *T. trichiura*. Albendazole, mebendazole, pyrantel pamoate, and ivermectin replaced a number of older anthelmintics that were more toxic, such as piperazine and thiabendazole, or less effective.

Systemic Roundworms (Nematodes)

Diethylcarbamazine is the drug of choice for lymphatic filarial infections (Chapter 358) caused by *Wuchereria bancrofti*, *Brugia malayi,* and *Brugia timori,* as well as for tropical pulmonary eosinophilia. It promotes the host's killing of microfilariae of these species and also damages or kills the adult worms. Inflammatory side effects are common and due in part to the release of lipopolysaccharide from endosymbiotic *Wolbachia* bacteria within dying filaria. *Wolbachia* are necessary for filarial development and are a potential drug target. Long-term therapy with doxycycline has been shown to result in their elimination and has been used for therapy. Diethylcarbamazine is also used for *Loa loa* infections in persons with acceptably low microfilaremia (<8000/mL). Encephalopathy may result from treatment of those with

higher levels. In that case, apheresis or treatment with albendazole is used first to reduce the number of microfilaria. Diethylcarbamazine can also be used prophylactically for *L. loa*. Ivermectin has activity against the microfilariae of *W. bancrofti*, *Brugia* species, and *L. loa*, but it does not kill adult worms and is not recommended for treatment of these organisms.

Ivermectin is the treatment of choice for onchocerciasis. It is administered as a single dose of 150 μg/kg. It does not kill adult *Onchocerca volvulus*, but it decreases ova production and reduces microfilariae in the skin and eyes. Retreatment is usually necessary at 6- to 12-month intervals until the patient is free of symptoms. Profits from the use of ivermectin for treatment of the dog heartworm *Dirofilaria immitis* have permitted the manufacturer to provide the drug free to persons with onchocerciasis in developing areas. Diethylcarbamazine should not be used for onchocerciasis. It kills the microfilariae of *O. volvulus* rapidly, but the release of parasite and *Wolbachia* antigens can result in severe ocular and systemic inflammatory responses. The latter is known as the Mazzotti reaction. Ivermectin is associated with less rapid killing of microfilariae and less severe reactions.

Tapeworms (Cestodes)

Praziquantel has a broad spectrum of activity against tapeworms (Chapter 354) and flukes. It is the drug of choice for adult tapeworms in the human intestinal tract. It is effective against *Taenia solium* (pork tapeworm), *Taenia saginata* (beef), *Diphyllobothrium latum* (fish), and *Hymenolepis nana* (dwarf tapeworm) when administered as a single dose. Niclosamide, which is not absorbed, is an effective alternative for the treatment of *T. saginata* and *D. latum*. It also kills adult *T. solium*, but disintegration of the worm and release of viable ova into the intestinal lumen raise the theoretical possibility of autoinfection. In the case of *H. nana*, a dose of nitazoxanide, 500 mg twice daily for 3 days, provides an alternative to praziquantel. It also has activity against *T. saginata* and potentially other tapeworm species.

Neurocysticercosis caused by the larval or tissue phase of *T. solium* is a major cause of seizures and other central nervous system (CNS) abnormalities in residents of and immigrants from endemic areas in Latin America and elsewhere. Symptoms can result from the physical presence of cysticerci, but the inflammation elicited by the release of antigens from dying cysticerci is often more important. Both albendazole and praziquantel are capable of killing cysticerci in the brain; albendazole is the drug of choice for pharmacokinetic reasons. Their use depends on the clinical syndrome. Corticosteroids are administered concurrently—and sometimes alone—to reduce the inflammatory response and the increase in intracranial pressure associated with it. Albendazole is administered for 15 to 30 days at 400 mg twice a day. Praziquantel is administered at 100 mg/kg in three divided doses the first day and then at 50 mg/kg in three divided doses for 29 days. The concurrent use of corticosteroids increases the serum level of albendazole but decreases that of praziquantel. Neither albendazole nor praziquantel should be used in persons with cysticerci in the eye or spinal cord because the release of antigens can trigger a locally destructive inflammatory reaction.

Surgery or PAIR (*p*ercutaneous *a*spiration, *i*njection of chemicals, and *r*easpiration) is the preferred approach for *Echinococcus granulosus* cysts. Albendazole is used for inoperable *E. granulosus* and *Echinococcus multilocularis* disease and in persons in whom medical treatment is preferred for other reasons. Administered at a dose of 400 mg twice daily for adults, generally for 1 to 6 months, albendazole may cure one third of uncomplicated *Echinococcus granulosus* liver cysts. Albendazole is also administered before ultrasound-guided PAIR or surgery to prevent seeding of the peritoneum should the cyst's contents spill. Potentially fatal bone marrow suppression and hepatitis are concerns in persons receiving high-dose, prolonged albendazole therapy.

Flukes (Trematodes)

Praziquantel is the drug of choice for the treatment of all forms of schistosomiasis (Chapter 355), as well as intestinal, lung, and liver flukes (Chapter 356), with the exception of the liver fluke *Fasciola hepatica*. For schistosomiasis, two or three doses are given in 1 day depending on the species. The flukes are treated with three doses on 1 or 2 days, depending on the species. Oxamniquine is an alternative for *Schistosoma mansoni*, but it is more toxic and less effective. Higher doses of oxamniquine are recommended for *S. mansoni* infections acquired in areas of Egypt or equatorial Africa where the parasite is less susceptible. Either praziquantel or albendazole can be used to treat the liver fluke *Clonorchis sinensis*. *F. hepatica* responds to the veterinary agent triclabendazole. Bithionol, which is more toxic, and nitazoxanide are alternatives.

Drugs Used for Helminthic Infections

Albendazole is poorly soluble in water, but it is well absorbed when administered with a fatty meal. It undergoes rapid first-pass metabolism in the liver to albendazole sulfoxide, which has excellent anthelmintic activity. The serum half-life of albendazole sulfoxide is 8 to 9 hours. Elimination of albendazole sulfoxide and other metabolites is achieved primarily through the kidney. Albendazole binds to tubulin in susceptible parasites, inhibits microtubule assembly, and decreases glucose absorption. It does not affect human tubulin. It also inhibits fumarate reductase in helminths. Concurrent administration of dexamethasone, which is frequently given to prevent cerebral edema in persons with neurocysticercosis, increases serum levels by approximately 50%. The cerebrospinal fluid (CSF) concentration of albendazole is approximately 40% of that in serum.

Albendazole is generally well tolerated when given as a single dose for the treatment of intestinal nematode infections, although gastrointestinal discomfort may develop or patients may experience migration of adult *A. lumbricoides* from the nose or mouth or see them in their stools. Albendazole at higher doses and for longer duration is also used for persons with neurocysticercosis. Corticosteroids are administered concurrently to reduce intracranial inflammation and resulting increased pressure. Albendazole is contraindicated in those with cysticerci in the eye or spinal cord. High-dose, prolonged therapy with albendazole, such as that recommended for echinococcal disease, can be complicated by alopecia, hepatitis, or bone marrow suppression, which is not always reversible after discontinuation of the drug. Albendazole is embryotoxic in animals and contraindicated during pregnancy.

Mebendazole is only slightly soluble in water and is relatively poorly absorbed from the gastrointestinal tract. This is advantageous for the treatment of intestinal parasites but limits its effectiveness against tissue-dwelling helminths. Absorbed drug is metabolized in the liver and excreted in urine. Mebendazole selectively binds to helminthic tubulin, blocks its assembly into microtubules, and inhibits glucose uptake, thereby leading to depletion of glycogen stores and ultimately death of the parasite. It is relatively well tolerated in the doses used to treat intestinal helminths. Transient abdominal pain and diarrhea occur in a small number of recipients. Mebendazole is contraindicated during pregnancy.

Ivermectin is a macrocyclic lactone produced by *Streptomyces avermitilis*. It has a broad spectrum of activity against helminths and arthropods, including *Sarcoptes scabiei*, the cause of scabies. It is well absorbed after oral administration. Ivermectin is highly protein bound, has a serum half-life of 12 hours, and accumulates in adipose tissue and the liver. It is subject to enterohepatic recirculation and ultimately eliminated in stool. Ivermectin activates the opening of gated chloride channels in susceptible helminths and arthropods. The result is an influx of chloride ions and paralysis of the pharyngeal pumping mechanism of helminths. Ivermectin is generally well tolerated in humans, although inflammatory reactions can result in response to antigens released from dying parasites.

Diethylcarbamazine, a piperazine derivative, is well absorbed orally and has a half-life of 8 hours. The parent drug and its metabolites are excreted through the kidney. Although the mechanism of action is uncertain, the piperazine moiety may result in paralysis of sensitive helminths. Diethylcarbamazine also alters the surface membranes of susceptible microfilariae, thereby resulting in destruction by the host's immune system. Side effects include those caused by the drug and those that result from release of the parasite antigens and lipopolysaccharide from filaria-harbored, endosymbiotic *Wolbachia*. Adverse effects include nausea, vomiting, anorexia, headache, malaise, weakness, arthralgias, and rarely, acute psychotic reactions. In patients with lymphatic filariasis, localized swelling or nodules may develop along the lymphatics during treatment, or transient lymphedema or hydrocele formation may occur.

Praziquantel is well absorbed after oral administration. It undergoes extensive first-pass metabolism, and the metabolites, which are inactive, are excreted in urine. Praziquantel is approximately 80% protein bound, with a serum half-life of 4 to 6 hours. It is rapidly taken up by susceptible cestodes and trematodes. In the case of schistosomes, praziquantel damages the tegument, which results in intense vacuolation and increased permeability to calcium. Adult schistosomes are paralyzed and translocated to the liver through the portal circulation. Sequestered antigens are exposed on their surface, permitting binding of antibodies and phagocytes and resulting in immune destruction. Praziquantel is an alternative to albendazole for the treatment of neurocysticercosis. The concurrent administration of

corticosteroids, which are necessary to decrease inflammation and edema in the brain, reduces the serum concentration of praziquantel. The concentration in CSF is approximately 15 to 20% that of serum.

Praziquantel is frequently associated with mild, transient side effects, including headaches, lassitude, dizziness, nausea, vomiting, and abdominal discomfort, but they are seldom severe enough to interrupt therapy. Untoward reactions attributed to release of parasite antigens have been reported in patients treated for schistosomiasis and pulmonary paragonimiasis. Increased intracranial pressure resulting from release of cysticercal antigens is a potentially life-threatening consequence in patients receiving praziquantel for neurocysticercosis. Corticosteroids should be administered concurrently. Praziquantel is contraindicated in persons with cysticerci in the eye or spinal cord.

TREATMENT OF PROTOZOAL DISEASES

Intestinal and Vaginal Protozoa

Several major luminal pathogens, including *Entamoeba histolytica* (Chapter 352), *Giardia lamblia* (Chapter 351), and *Trichomonas vaginalis* (Chapter 353), live in anaerobic conditions in the intestine or vagina and are susceptible to metronidazole and tinidazole. The latter has favorable pharmacodynamics and is generally better tolerated. Because neither metronidazole nor tinidazole reliably eradicates cysts of *E. histolytica*, either paromomycin or iodoquinol, which are active in the lumen of the bowel, is administered as well. Either one of these drugs or diloxanide furoate can be used alone to treat persons with asymptomatic cyst excretion.

Giardiasis can also be treated with nitazoxanide. It is well tolerated, and an oral formulation is available for children. Nitazoxanide is the only drug available for the treatment of cryptosporidiosis (Chapter 350). It is effective in immunocompetent persons, but not in those with AIDS. Trimethoprim-sulfamethoxazole, which inhibits successive steps in the folic acid pathway, is the drug of choice for *Cystoisospora* (*Isospora*) *belli* and *Cyclospora cayetanensis*. Ciprofloxacin, a fluoroquinolone antibiotic, is an alternative. Tetracycline is the treatment of choice for *Balantidium coli*; metronidazole and iodoquinol are alternatives. Finally, albendazole is effective in the treatment of intestinal and disseminated microsporidiosis caused by *Encephalitozoon* (*Septata*) *intestinalis* and for some other microsporidial species that cause disease in persons with AIDS.

Metronidazole, a nitroimidazole, is rapidly absorbed after oral administration and has a half-life of 8 hours. More than half of each dose is metabolized in the liver. The metabolites and remaining parent drug are excreted in urine. Metronidazole is activated by reduction of its 5-nitro group through a sequence of intermediate steps involving microbial electron transport proteins of low redox potential. It is concentrated in susceptible anaerobic organisms and serves as an electron sink. Nausea, vomiting, diarrhea, and a metallic taste are often associated with the use of metronidazole. They are less common with the lower doses recommended for the treatment of giardiasis than with the higher doses used for amebiasis. Other untoward effects include headache, dizziness, vertigo, and numbness. Potentially severe disulfiram-like reactions occur in patients who ingest alcohol while taking metronidazole.

Tinidazole, another 5-nitroimidazole, has a similar mechanism of action and spectrum of activity as metronidazole but more favorable pharmacodynamics, and it is generally better tolerated. It has been used widely around the world for the treatment of giardiasis, intestinal amebiasis, and trichomoniasis. In comparison with metronidazole, it has a longer half-life, a shorter and less complicated dosing regimen, and fewer gastrointestinal side effects. It, too, can cause severe disulfiram-like reactions after alcohol ingestion.

Nitazoxanide, a 5-nitrothiazole salicylamide derivative, has a broad spectrum of activity against protozoa and helminths. It is formulated as a liquid for use in children. Nitazoxanide is well absorbed orally and hydrolyzed to its active metabolite tizoxanide, which undergoes conjugation to tizoxanide glucuronide. The parent compound is not detectable in serum. Maximum concentrations of the metabolites are observed in 1 to 4 hours. They are excreted in urine and bile. Tizoxanide is highly protein bound. Although its antiparasitic mechanism of action is uncertain, tizoxanide inhibits pyruvate : ferredoxin oxidoreductase–dependent electron transport reactions essential for the metabolism of susceptible anaerobic organisms. It is very well tolerated in children and adults.

Malaria: Prophylaxis and Treatment

As discussed in this section, a number of drugs are available for the prophylaxis and treatment of malaria. Most act against *Plasmodium* stages within erythrocytes. Only primaquine kills *Plasmodium vivax* and *Plasmodium ovale* hypnozoites in the liver. The antimalarial drug of choice depends on the geographic site visited and the *Plasmodium* species encountered (Chapter 345). Antimicrobial resistance is now widespread among *Plasmodium falciparum* isolates and well documented in *P. vivax* from some regions. Country-specific recommendations for prophylaxis and treatment are provided by the CDC at www.cdc.gov/travel/ and in "CDC Health Information for International Travel 2014."[3,4] Major additions to the therapeutic armamentarium in the United States include the fixed drug combination of artemether and lumefantrine (Coartem), which is used for the oral treatment of acute malaria acquired in areas with chloroquine-resistant *P. falciparum*, and artesunate for intravenous administration in those with severe malaria who cannot take oral medications and for whom quinidine is contraindicated or not available. Artesunate can be obtained from the CDC on an emergency basis as an investigational new drug.

Atovaquone plus ***proguanil*** (adult tablets contain 250 mg of atovaquone and 100 mg of proguanil) is used for prophylaxis and treatment of chloroquine-resistant and sensitive malaria. Atovaquone is a highly lipophilic compound with low aqueous solubility. Administration with food enhances its absorption two-fold. Plasma concentrations do not increase proportionately with dose. Atovaquone is highly protein bound with a half-life exceeding 60 hours. It undergoes extensive enterohepatic cycling and is eventually excreted unchanged in feces. Atovaquone selectively inhibits electron transport in the mitochondria of susceptible *Plasmodium* species at the level of the cytochrome bc_1 complex, which results in collapse of mitochondrial membrane potential. It also affects pyrimidine biosynthesis, which is obligatorily coupled to electron transport in *Plasmodium*. Resistance develops rapidly when atovaquone is used alone to treat malaria. Atovaquone is generally well tolerated but can cause nausea, vomiting, diarrhea, rash, and pruritus.

Proguanil is absorbed slowly after oral administration. Its serum level falls to zero within 24 hours, so it must be administered daily. Its triazine metabolite, cycloguanil, inhibits dihydrofolate reductase in susceptible *Plasmodium* species. Resistance is well documented when proguanil is used alone. Proguanil also acts synergistically with atovaquone to collapse mitochondrial membrane potential in susceptible *Plasmodium* species.

The combination of atovaquone and proguanil is considered the best tolerated of the options for prevention of chloroquine-resistant malaria. It is begun 1 to 2 days before departure and continued during the time of exposure and for 7 days thereafter. Higher doses are administered over a period of 3 days to treat acute, uncomplicated malaria. Potential side effects include abdominal pain, nausea, vomiting, diarrhea, headache, pruritus, and rash. Asymptomatic, transient elevations in liver enzymes have been observed with treatment doses.

Doxycycline 100 mg taken daily by adults provides effective prophylaxis against all *Plasmodium* species. It is begun 1 to 2 days before departure and continued during the time of exposure and for 4 weeks after leaving the malaria-endemic area. Doxycycline or tetracycline is also often administered with quinine for the treatment of acute chloroquine-resistant malaria, but neither drug acts rapidly enough to be used alone for treatment. Doxycycline is generally well tolerated, although it can cause gastrointestinal symptoms and "pill" esophagitis. To avoid the latter, it should be taken with a full glass of water, and the recipient should remain upright for an hour or more after ingestion. Other potential side effects include photosensitivity dermatitis, *Candida albicans* vaginitis, and antibiotic-associated colitis. Finally, doxycycline and tetracycline should not be used in children younger than 8 years or in women who are pregnant or breastfeeding.

Mefloquine, a quinoline methanol compound derived from quinine, was once used widely for the prophylaxis and occasionally treatment of chloroquine-resistant *P. falciparum* malaria. Mefloquine is available for oral administration only. Slowly and incompletely absorbed, it is 99% protein bound. It has a variable half-life ranging from 6 to 23 days with a mean of approximately 14 days. It is metabolized and excreted slowly through bile and feces.

Concern about neuropsychiatric and other toxicities and the availability of better-tolerated alternatives have limited its use in recent years. It is associated with nausea, dizziness, vivid dreams, fatigue, and lassitude. Less common but of greater concern are anxiety, depression, acute psychosis, and seizures. Mefloquine is contraindicated in persons with a history of epilepsy or psychiatric disorders. It now carries a U.S. Food and Drug Administration (FDA) black-box warning. It also depresses atrioventricular conduction and should not be used in persons taking β-blockers for cardiac indications.

Although not approved for use during pregnancy or in children weighing less than 15 kg, mefloquine has been used in situations in which its potential benefits were judged to outweigh the risks.

Chloroquine, a 4-aminoquinoline, has a bitter taste but is well absorbed from the gastrointestinal tract. Its half-life, which varies among persons, averages 4 days, thus permitting once-weekly administration for prophylaxis. Chloroquine is concentrated in the hemoglobin-containing digestive vesicles of asexual intraerythrocytic parasites. It inhibits the parasite's heme polymerase that incorporates ferriprotoporphyrin type IX complexes, which are potentially toxic to the parasite, into insoluble, nontoxic, crystalline hemozoin. Chloroquine-resistant strains of *P. falciparum* actively transport chloroquine out of the intraparasitic compartment. Although this action can be blocked by calcium-channel inhibitors in vitro, chloroquine resistance has not been effectively reversed in humans. Hydroxychloroquine (Plaquenil), which is used for rheumatologic diseases, is also effective against chloroquine-sensitive *Plasmodium* species.

Chloroquine is generally well tolerated when used at the doses recommended for the prophylaxis and treatment of malaria. Side effects include headache, nausea, vomiting, blurred vision, dizziness, and fatigue. Some Africans and African Americans experience pruritus, which responds to antihistamines. Rare side effects include depigmentation of hair, exacerbation of psoriasis, blood dyscrasias, seizures, neuropsychiatric effects, and reactions in persons with porphyria. Retinal damage has occurred in persons receiving chloroquine at high doses for the treatment of rheumatologic disorders, but it has not been documented as a problem in those taking it weekly over a period of many years for malaria prophylaxis. Cardiopulmonary collapse and death have occurred after accidental overdose and in adults attempting suicide. As little as 5 g of chloroquine can be fatal unless treatment is initiated with mechanical respiration, medications to control seizures, and blood pressure support.

Primaquine, an 8-aminoquinoline, eradicates the hepatic hypnozoite stage of *P. vivax* and *P. ovale* and is used as a 14-day course at the end of treatment or prophylaxis to prevent late relapses in persons who are or may be infected with these *Plasmodium* species. It is also an alternative for daily prophylaxis for *Plasmodium vivax* and other species. In that case, it is begun 1 or 2 days before exposure and continued during and for 7 days after the traveler leaves the endemic area.

Primaquine is well absorbed orally and rapidly converted to carboxyprimaquine, which has a half-life of approximately 7 days. It is generally well tolerated, although some recipients experience abdominal cramps, epigastric distress, and nausea. The major concern is hemolysis in persons with glucose-6-phosphate dehydrogenase (G6PD) deficiency (Chapter 161). The G6PD status of the recipient should be determined before it is administered. Rarely, primaquine causes neutropenia, methemoglobinemia, hypertension, or arrhythmias. Primaquine is contraindicated during pregnancy and in breast-feeding mothers because life-threatening hemolysis may occur if the fetus or baby is deficient in G6PD. Travelers should be warned not to give the drug to fellow travelers who have not been screened for G6PD deficiency.

Quinine sulfate, a cinchona alkaloid, is the oldest of the antimalarials. It has a very bitter taste. It is rapidly absorbed after oral administration and has a half-life of 16 to 18 hours in persons with malaria. Quinine has the poorest therapeutic-to-toxicity ratio of any antimalarial drug. The side effects, known collectively as cinchonism, include tinnitus, decreased hearing, headache, nausea, vomiting, dysphoria, and visual disturbances. They are dose related and reversible. Quinine has also been associated with severe hypoglycemia in persons with heavy *P. falciparum* infection as a result of the utilization of glucose by the parasites and release of insulin from the pancreas. Hypoglycemia can be prevented or treated by the intravenous administration of glucose. Rare complications with quinine include massive hemolysis in patients with heavy *P. falciparum* infection resulting in hemoglobinuria and renal failure (blackwater fever), cutaneous hypersensitivity reactions, agranulocytosis, and hepatitis. Quinine can cause respiratory paralysis in persons with myasthenia gravis. It stimulates uterine contractions and can produce abortions, but it has saved the lives of many pregnant women with *P. falciparum* malaria. Quinine dihydrochloride given intravenously can cause myocardial depression, peripheral vascular collapse, respiratory depression, and potentially death.

Quinidine gluconate, the stereoisomer of quinine, is recommended for the intravenous treatment of patients with severe malaria and in those who cannot take antimalarials orally. Quinidine gluconate was once widely used for the treatment of ventricular ectopy, but it has been replaced by newer antiarrhythmic agents, which has decreased its availability in many hospitals. Side effects include prolongation of the QT interval, arrhythmias, and hypotension, particularly if it is infused too rapidly. Persons receiving intravenous quinidine should be monitored in an intensive care setting, and therapy should be switched to an oral antimalarial medication as soon as possible.

Artemether,[5] ***artesunate***,[6] and other artemisinins are sesquiterpene lactone derivatives of the wormwood plant *Artemisia annua*, from which qinghaosu, the Chinese herbal medication for fever, is derived. They are endoperoxide-containing compounds. In the presence of intraparasitic iron, they are converted into free radicals and other intermediates that alkylate specific malarial proteins and act rapidly to kill intraerythrocytic parasites. The artemisinins have been used widely around the world for the treatment of acute malaria caused by chloroquine-resistant *P. falciparum*, as well as other *Plasmodium* species. They are usually administered with a second antimalarial drug that has a different mechanism of action and longer half-life to prevent the development of resistance. The route of administration with artemisinins varies; some are well absorbed orally, whereas others are administered intravenously, intramuscularly, or by suppository. Their short half-lives preclude their use for prophylaxis. Side effects in humans are common but seldom result in discontinuation of the drug. Neurologic toxicity and cerebellar dysfunction have been observed in dogs receiving chronic, high-dose therapy.

Artemether plus lumefantrine (Coartem), a fixed drug combination, is available in the United States and has been widely used around the world to treat chloroquine-resistant malaria. It should be taken with food, but grapefruit juice should be avoided. Common adverse reactions in adults are headache, anorexia, dizziness, asthenia, arthralgia, and myalgia. The most common in children are fever, cough, vomiting, anorexia, and headache. They do not usually require discontinuation of therapy. Of greater concern, artemether-lumefantrine can result in prolongation of the QT interval and is contraindicated in persons with abnormal QT. It also inhibits CYP206 and can hereby reduce the metabolism of other medications that prolong the QT interval. It can also decrease the effectiveness of birth control pills. Care must be taken in reviewing the recipient's medication list for potential interactions before artemether-lumefantrine is prescribed.

Artesunate is available from the CDC as an investigational new drug for the intravenous treatment of severe malaria in persons who cannot take quinidine gluconate, in those who have failed quinidine, or when it is not available (contact the CDC Malaria Hot Line for information). Artesunate is rapidly hydrolyzed to dihydroartemisinin, which is responsible for the antimalarial effect. Studies in malaria endemic regions suggest that parenteral artesunate has a higher success rate and lower adverse event rate than quinidine. In addition to the side effects with Coartem described previously, artesunate has been associated with delayed hemolysis approximately 2 weeks after completion of therapy.

Toxoplasmosis, Babesiosis, and Amoebic Encephalitis

Toxoplasma gondii (Chapter 349) and *Babesia* species (Chapter 353) are other important pathogens of the phylum Apicomplexa. Pyrimethamine and sulfadiazine are recommended for the treatment of toxoplasmosis. They inhibit sequential steps in the folic acid metabolic pathway. Pyrimethamine preferentially inhibits dihydrofolate reductase in susceptible parasites. It is well absorbed orally. The major side effect is macrocytic anemia, which can be prevented by the concurrent administration of leucovorin. Sulfonamides reduce the activity of dihydropteroate synthetase and the binding of *p*-aminobenzoic acid to it. In ocular toxoplasmosis with macular involvement, corticosteroids are used along with anti-*Toxoplasma* therapy to minimize the local inflammatory response. Clindamycin plus pyrimethamine and atovaquone plus pyrimethamine are therapeutic options in sulfonamide-intolerant patients. Persons with AIDS, $CD4^+$ counts less than $100/mm^3$, and serologic evidence of *T. gondii* infection should receive prophylaxis with one of the following regimens: daily trimethoprim-sulfamethoxazole, pyrimethamine plus dapsone, pyrimethamine plus atovaquone, or atovaquone alone. Spiramycin, a macrolide, is used for the treatment of toxoplasmosis during pregnancy. Two therapeutic regimens are available for babesiosis. The greatest experience has been with the combination of clindamycin plus quinine, but side effects are common. Atovaquone plus azithromycin is effective and better tolerated (Chapter 353). Finally, the addition of miltefosine, an antileishmanial drug (see later), to multidrug regimens has improved the outcome of persons with encephalitis due to the free-living *Acanthamoeba* and *Balamuthia*. It is now recommended as part of regimens to treat them as well as *Naegleria* infections. It is available from the CDC for these infections.

CHAGAS' DISEASE, AFRICAN TRYPANOSOMIASIS, AND LEISHMANIASIS

The *Leishmania* species that cause cutaneous, mucosal, and visceral leishmaniasis; *Trypanosoma cruzi*, the etiology of Chagas' disease; and *Trypanosoma brucei rhodesiense* and *Trypanosoma brucei gambiense,* which are responsible for human African trypanosomiasis (sleeping sickness); pose difficult therapeutic challenges.

Nifurtimox and benznidazole orally are the only drugs available to treat *T. cruzi* (Chapter 347).[7] Benznidazole has been used widely in endemic areas in Latin America and nifurtimox in the United States. They lower mortality and shorten the duration of acute Chagas' disease. Treatment is also recommended for persons with recent infection as well as asymptomatic children and adults through middle age with indeterminate-stage *T. cruzi* infection. The percentage of those who are parasitologically cured by treatment has been debated. Neither benznidazole nor nifurtimox can reverse the manifestations of chronic Chagas' disease after they have developed. Side effects are common with both drugs and increase in frequency and severity with age. Benznidazole is associated with allergic dermatitis, peripheral neuropathy, insomnia, and gastrointestinal symptoms, including anorexia and weight loss. Nifurtimox causes anorexia, nausea, vomiting, weight loss, headache, dizziness or vertigo, paresthesias, weakness, and polyneuropathy.

Suramin, pentamidine, eflornithine, and melarsoprol are used for the treatment of human African trypanosomiasis (Chapter 346). Suramin is recommended for the hemolymphatic stage of *T. brucei rhodesiense* infection, and melarsoprol is used in those with CNS involvement. Both drugs are associated with potentially severe side effects. Eflornithine, which is much better tolerated, is effective against both the hemolymphatic and CNS stages of *T. brucei gambiense* infection. It does not have activity against *T. brucei rhodesiense*. Unfortunately, eflornithine is costly and not available in many endemic areas, and supplies are limited. When eflornithine is not available, pentamidine, which has substantial untoward effects, is used for the hemolymphatic stage of *T. brucei gambiense* infection, with suramin being used as an alternative. Melarsoprol is used for CNS disease.

Liposomal amphotericin B (AmBisome) and miltefosine are the only drugs approved by the FDA for treatment of visceral leishmaniasis (Chapter 348) in the United States. Amphotericin B deoxycholate is also effective. Liposomes deliver amphotericin to macrophages and are theoretically attractive because leishmania reside within them. Liposomal amphotericin B is also better tolerated than amphotericin B. Other lipid-associated amphotericin B preparations appear to be effective, but they have been less extensively studied and are not FDA approved for this indication.

For many years, two pentavalent antimony drugs, stibogluconate sodium and meglumine antimoniate, were used to treat visceral leishmaniasis, but resistance is now common among *Leishmania donovani* isolates in India, and therapeutic failures occur in other areas. In addition, pentavalent antimonials require parenteral administration. Side effects increase with age and include gastrointestinal complaints, pancreatitis, myalgias, headache, malaise, elevated liver enzyme levels, and occasionally, bone marrow suppression. Nonspecific ST-T wave changes are common. Sudden death has been reported in persons receiving more than the recommended dose.

Miltefosine,[8] an alkylphospholipid and phosphocholine analogue, that was initially developed as an antineoplastic drug, has proved effective for the treatment of antimony-resistant visceral leishmaniasis in the Indian subcontinent and for cutaneous leishmaniasis in a number of other geographic locations. A major advantage is oral administration. The pharmacokinetics are characterized by a long elimination half-life and extensive drug accumulation. The mechanism of action is uncertain, but it induces apoptosis-like changes in the parasite and has immune modulatory effects in the host. Side effects are frequent but typically are mild to moderate. Dose-dependent gastrointestinal toxicity can result in nausea, vomiting, and diarrhea. They tend to decrease with continued administration of the drug. Elevations in liver enzymes and creatinine are common but usually transient. Miltefosine is embryotoxic and is thus contraindicated during pregnancy. Contraceptive cover is mandatory in females of childbearing years during and for 4 months after therapy. Although miltefosine has proved effective in the treatment of visceral leishmaniasis on the Indian subcontinent, relapses have been reported in 10 to 20% of cases and are even more frequent in persons coinfected with HIV/AIDS. Resistance has also been reported.

Treatment of cutaneous leishmaniasis depends on the size, number, complexity, and location of the skin lesions, their cosmetic impact, the infecting *Leishmania* species, and its propensity to cause mucosal disease. Simple lesions acquired in Europe, Africa, and Asia where mucosal dissemination is rare are often treated topically, or if they are spontaneously healing, followed without therapy. Lesion-directed treatment options include cryotherapy; heat therapy, which requires a specialized delivery system; or intralesional injection of pentostam, which is not available in the United States. Topical therapy with direct application of paromomycin is another option. An ointment containing 15% paromomycin and 12% methylbenzethonium chloride in white paraffin developed in Israel has been the most widely used. Recent results with a U.S. Army formulation of topical paromomycin appear very promising.[9]

Parenteral or oral antileishmanial therapy is used for persons with complicated cutaneous disease and those who are or may be infected with a New World *Leishmania* species potentially associated with mucosal leishmaniasis. Stibogluconate sodium and meglumine antimoniate have been used widely over the years, but toxicity and the requirement for parenteral administration are problematic. Miltefosine has also been used in the treatment of cutaneous and mucosal leishmaniasis. The efficacy has varied with the *Leishmania* species and geographic location. The imidazole antifungals vary in their activity against different *Leishmania* species. Recent clinical experience in Brazil suggests that higher doses of fluconazole are more effective than the 200-mg/day dose used in earlier trials. Liposomal amphotericin B and amphotericin B deoxycholate are effective but more toxic and expensive parenteral alternatives.

Mucosal leishmaniasis is less responsive than cutaneous leishmaniasis to treatment, and relapses are common. Therapeutic options include stibogluconate sodium, meglumine antimoniate, liposomal amphotericin B, amphotericin B deoxycholate, and miltefosine.

Grade A References

A1. Speich B, Ame SM, Ali SM, et al. Ozantel pamote-albendazole for *Trichuris trichiura* infection. *N Engl J Med*. 2014;370:610-620.

A2. Basanez MG, Pion SD, Boakes E, et al. Effect of single-dose ivermectin on *Onchocerca volvulus*: a systematic review and meta-analysis. *Lancet Infect Dis*. 2008;8:310-322.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Swanson SJ, Phares CR, Mamo B, et al. Albendazole therapy and enteric parasites in United States-bound refugees. *N Engl J Med*. 2012;366:1498-1507.
2. Omura S, Crump A. Ivermectin: panacea for resource-poor communities? *Trends Parasitol*. 2014;30:445-455.
3. Centers for Disease Control and Prevention. CDC Health Information for International Travel 2014. Atlanta: U.S. Department of Health and Human Services, Public Health Service; 2014. Available at: http://wwwnc.cdc.gov/travel/. Accessed March 9, 2015.
4. CDC Treatment Guidelines. Treatment of malaria (based on drugs available for use in the United States—updated July 1, 2013). Available at: http://www.cdc.gov/malaria/resources/pdf/clinicalguidance.pdf. Accessed March 9, 2015.
5. Stover KR, King ST, Robinson J. Artemether-lumefantrine: an option for malaria. *Ann Pharmacother*. 2012;46:567-577.
6. Hess KM, Goad JA, Arguin PM. Intravenous artesunate for the treatment of severe malaria. *Ann Pharmacother*. 2010;44:1250-1258.
7. Bern C. Antitrypanosomal therapy for chronic Chagas' disease. *N Engl J Med*. 2011;364:2527-2534.
8. Dorlo TP, Balasegam M, Reijnen JH, et al. Miltefosine: a review of its pharmacology and therapeutic efficacy in leishmaniasis. *J Antimicrob Chemother*. 2012;67:2576-2597.
9. Ben Salah A, Ben Messaoud N, Guedri E, et al. Topical paromomycin with or without gentamicin for cutaneous leishmaniasis. *N Engl J Med*. 2013;368:524-532.

REVIEW QUESTIONS

1. A regional health district in a resource limited area in rural Africa desires to treat school children empirically for intestinal helminthes to enhance their growth and intellectual development. Which of the following would be the best choice?

A. Albendazole 400 mg once
B. Ivermectin 150 µg/kg once
C. Diethylcarbamazine 50 mg once
E. Pyrantel pamoate 100 mg once
F. Praziquantel 50 mg/kg once

Answer: A The most likely intestinal helminthes in areas of poor sanitation are *Ascaris lumbricoides*, hookworms, and *Trichuris trichiura*. A single dose of albendazole 400 mg is active against all of them. Ivermectin can be used against *Ascaris* and *Trichuris* but lacks activity against hook worms. Pyrantel pamoate has activity against *Ascaris* and hookworms, but not *Trichuris*. Diethylcarbamazine is used for the treatment of lymphatic filariasis. Praziquantel is used for the treatment of intestinal tapeworms (cestode) and most fluke (trematode) infections.

2. Plans are underway to start a new hospital in Southeast Asia where *Schistosoma japonicum*, *Fasciolopsis buski,* and *Clonorchis sinensis* are endemic. Which of the following antihelminthic medications should be available to treat these diseases?

A. Nitazoxanide
B. Praziquantel
C. Albendazole
D. Atovaquone/proguanil
E. Tinidazole

Answer: B Praziquantel is the treatment of choice for these three pathogens and other flukes (trematodes) with the exception of *Fasciola hepatica*, the liver fluke. Nitazoxanide is indicated for intestinal protozoa, including *Giardia lamblia* and *Cryptosporidium,* and has activity against some intestinal nematodes and cestodes. Albendazole is used primarily for intestinal round worms and neurocysticercosis. Atovaquone/proguanil is effective for the prophylaxis and treatment of malaria. Tinidazole is used for the treatment of anaerobic luminal pathogens including *Entamoeba histolytica, Giardia lamblia,* and *Trichomonas vaginalis*.

3. A 24-year-old man is planning a 7-day trip to a game park in Tanzania, East Africa. He has a history of depression, which is well controlled with daily fluoxetine (Prozac). Which of the following would be the best choice for malaria prophylaxis?

A. Mefloquine (Lariam) weekly
B. Artemether/lumefantrine (Coartem) daily
C. Chloroquine weekly
D. Pyrimethamine daily
E. Atovaquone/proguanil (Malarone) daily

Answer: E The options for prevention of chloroquine-resistant malaria include atovaquone/proguanil, doxycycline, mefloquine, and primaquine. They are of comparable efficacy. In respect to side effects, atovaquone/proguanil is the best tolerated. Artemether/lumefantrine and other artemisinin derivates have short half-lives and cannot be used for prophylaxis. Chloroquine cannot be used in sub-Saharan Africa where chloroquine-resistant *Plasmodium falciparum* is endemic. Likewise, resistance to pyrimethamine is widespread in Africa. Mefloquine is effective against chloroquine-resistant malaria but is associated with serious neuropsychiatric and other effects, including depression, psychosis, and seizures.

4. A 55-year-old woman presented in May with fever and headache of 1 day's duration. She returned 7 days earlier from a 3-week trip to game parks in Tanzania, East Africa. She is otherwise in good health and taking no medications. She denies cough, sputum production, nausea, vomiting, diarrhea, dysuria, and skin rash. On examination, her temperature is 40° C, blood pressure is 120/85 mm Hg, heart rate is 120 beats per minute, and respiratory rate is 18 breaths per minute. She has no rash, her neck is supple, chest is clear to auscultation and percussion, and abdomen is soft and nontender with normal bowel sounds. The laboratory technician on call has never done a malaria smear, and the hospital does not offer rapid diagnostic tests. Results from a reference laboratory will be available in 24 hours. Which of the following would be the best course of action?

A. Treat with mefloquine
B. Treat if parasites are identified at the reference laboratory
C. Treat with chloroquine
D. Treat with artemether/lumefantrine (Coartem)
E. Treat with doxycycline

Answer: D Fever in a returning traveler from Africa is malaria until proved otherwise! A delay in treatment can result in death, and therapy should not be delayed. The fastest acting drug for acute malaria is the fixed combination artemether/lumefantrine. Chloroquine cannot be used because of widespread chloroquine-resistant *Plasmodium falciparum* in East Africa. High-dose mefloquine has the risk for neurologic toxicity and would be used only if other antimalarial drugs were not available. Doxycycline does not act rapidly enough to be used alone for the treatment of acute malaria.

5. A 27-year-old woman presents with diarrhea of 3 weeks' duration. She recently returned after a 6-week visit to the Amazon Basin of South America. She stayed in local houses, drank tap water, and was an adventurous eater. She has had four to six loose stools a day, bloating, and intermittent abdominal crampy pain. On physical examination, her temperature is 37.0° C, blood pressure is 120/62 mm Hg, heart rate is 70 beats per minute, and respiratory rate is 12 breaths per minute. She is in no distress. The bowel sounds are normal. Her abdomen is soft, nontender, and without mass or hepatosplenomegaly. The rectal examination is normal. She is HIV negative. Stool tests are positive for cryptosporidial antigen and giardia antigen. What is the best choice for treatment?

A. Nitazoxanide
B. Tinidazole
C. Metronidazole
D. Albendazole
E. Praziquantel

Answer: A Nitazoxanide is the only drug with activity against *Cryptosporidium* species, and it is also effective against *Giardia*. Tinidazole and metronidazole are active against luminal protozoa, including *Entamoeba histolytica*, *Giardia lamblia,* and some others, but not *Cryptosporidium*. Albendazole is active against most intestinal nematodes. Although it has some activity against *Giardia*, albendazole is not the drug of choice for giardiasis and is not effective for *Cryptosporidium*. Praziquantel has activity against tapeworms (cestodes) and most flukes (trematodes), with the exception of *Fasciola hepatica.*

345

MALARIA

PHILIP J. ROSENTHAL AND MOSES R. KAMYA

DEFINITION

Malaria is caused by infection with protozoan parasites of the genus *Plasmodium,* all of which are transmitted by bites of infected anopheline mosquitoes.[1] Malaria is typically characterized by an acute febrile illness, with parasites infecting large numbers of erythrocytes, and classically entails recurrent episodes of fever and chills. It was first described thousands of years ago and is named on the basis of the belief that it was caused by the bad air of the marshes surrounding Rome. Malaria is common, causing hundreds of millions of illnesses each year throughout most of the tropics. Severe disease can occur, primarily with *Plasmodium falciparum* infection, with the acute development of serious organ dysfunction or when chronic and repeated infection leads to severe anemia. *P. falciparum* malaria kills an estimated 660,000 people each year, mostly children in sub-Saharan Africa.

The Pathogen

P. falciparum is responsible for most episodes of severe malaria. It is endemic in most malarious areas and is by far the predominant species in Africa. *Plasmodium vivax* is about as common as *P. falciparum,* except in Africa, but causes severe disease much less commonly. However, studies suggest that severe illness associated with *P. vivax* infection is more common than had previously been appreciated.[2] *Plasmodium ovale* and *Plasmodium malariae* are

much less common causes of disease and generally do not cause severe illness. A fifth parasite causing human infections is *Plasmodium knowlesi*, a parasite of macaque monkeys that is a fairly common zoonosis in parts of Southeast Asia and has been responsible for malarial illnesses, including severe disease, in individuals exposed to macaque-biting vectors in forested areas.

EPIDEMIOLOGY

Malaria in Endemic Countries

Malaria is the most important parasitic disease of humans, causing hundreds of millions of illnesses and hundreds of thousands of deaths each year. The disease is endemic in most of the tropics, including many parts of South and Central America, Africa, the Middle East, the Indian subcontinent, Southeast Asia, and Oceania. Transmission, morbidity, and mortality are greatest in Africa, where infection with *P. falciparum* predominates. In most other endemic areas, disease caused by both *P. falciparum* and *P. vivax* is common. In highly endemic areas, the group at greatest risk is young children, who experience most episodes of disease and the most deaths. A second high-risk group is pregnant women, with high risks of maternal and fetal morbidity from *P. falciparum* malaria, including many deaths secondary to low birthweight. In highly endemic areas, in addition to extensive mortality, malaria exerts a massive toll through adverse effects on child development; contributions to school and work absenteeism; and, overall, billions of dollars in lost income among the poorest citizens of the poorest countries of the world. In areas of developing countries with lower levels of malaria transmission, malaria can be epidemic, with intermittent increases in transmission that cause major morbidity in relatively nonimmune populations.

Malaria in Travelers

Malaria is also common in travelers of any age from nonendemic areas to the tropics and may be manifested in those who have returned to nonendemic areas up to many months after travel. Indeed, malaria is the most common documented cause of febrile illness in travelers returning from the tropics to developed countries. Malaria is also rarely transmitted in areas considered nonendemic, including the United States, when imported parasites are transmitted by local anopheline mosquitoes, by blood products, or through congenital spread of infection.

Malaria Transmission

Malaria is transmitted by multiple species of mosquitoes of the genus *Anopheles,* which vary in geographic distribution, ecologic preferences, and susceptibility to mosquito control measures. Anopheline mosquitoes bite at night, so personal mosquito control measures focus on avoidance of mosquito bites overnight. Levels of malaria transmission in endemic areas vary greatly, from areas where residents experience only rare infectious bites to regions of Africa where individuals may receive hundreds of infectious bites each year.

Recent Changes in the Epidemiology of Malaria

A major effort to eradicate malaria after World War II led to elimination of the disease in many areas, including the United States and Europe, but eventually failed to control the disease in most of the tropics. During the ensuing decades, improvements in malaria control were few, and malarial morbidity worsened in many areas, driven by the loss of enthusiasm for vector control; increasing resistance of mosquitoes to insecticides; and, in particular, resistance of parasites to commonly used antimalarial drugs. More than 50 years later, a new large effort to control and eventually to eradicate malaria has been initiated.[3] Key control measures include indoor residual spraying of insecticides, distribution of insecticide-impregnated bed nets, prompt provision of effective drugs to those with malaria, and targeted administration of drugs to prevent infection in high-risk groups. Recent efforts have led to documented decreases in levels of malarial morbidity and mortality in some areas, notably parts of Africa,[4] Asia, and Oceania with relatively low levels of transmission intensity. It is of great interest to learn whether recent advances due to intensified control efforts can bring significant improvements to those parts of the world most affected with malaria, in particular, highly endemic regions of Africa and Asia.

Malaria and Human Immunodeficiency Virus Infection

Malaria does not differ as markedly between those with human immunodeficiency virus (HIV) infection and others as is the case with typical opportunistic infections. However, several interactions between malaria and HIV infection have been established. First, HIV infection appears to disrupt the acquired immune response to malaria and thereby increases the incidence and severity of malaria. Second, acute malaria elevates HIV viral load and so may increase the risk of HIV transmission. Third, HIV infection may be associated with reduced efficacy of antimalarial treatment, especially in the setting of severe immune suppression. Fourth, therapies for each infection may have an impact on the other, leading to unanticipated effects on drug efficacy or toxicity. Fifth, routine interventions for HIV infection may affect the incidence of malaria; notably, daily trimethoprim–sulfamethoxazole, a routine regimen in HIV-infected patients, offers partial protection against malaria. Because both HIV infection and malaria are common in many areas, in particular sub-Saharan Africa, even modest associations are important. Thus, malaria coinfection in HIV-infected individuals may be an important factor in promoting the spread of HIV infection in Africa. However, increasing implementation of insecticide treated bed nets, prophylaxis against opportunistic infections with trimethoprim–sulfamethoxazole, and antiretroviral therapy will likely substantially lessen malaria risk in HIV-infected patients, such that the risk in those receiving optimal management will not be substantially greater than that in the general population.

PATHOBIOLOGY

Parasite Life Cycle

Malaria is transmitted by the bite of infected female anopheline mosquitoes. During feeding, mosquitoes inject sporozoites, which circulate to the liver and infect hepatocytes, causing asymptomatic liver infection (Fig. 345-1). Merozoites are subsequently released from the liver, and they rapidly infect

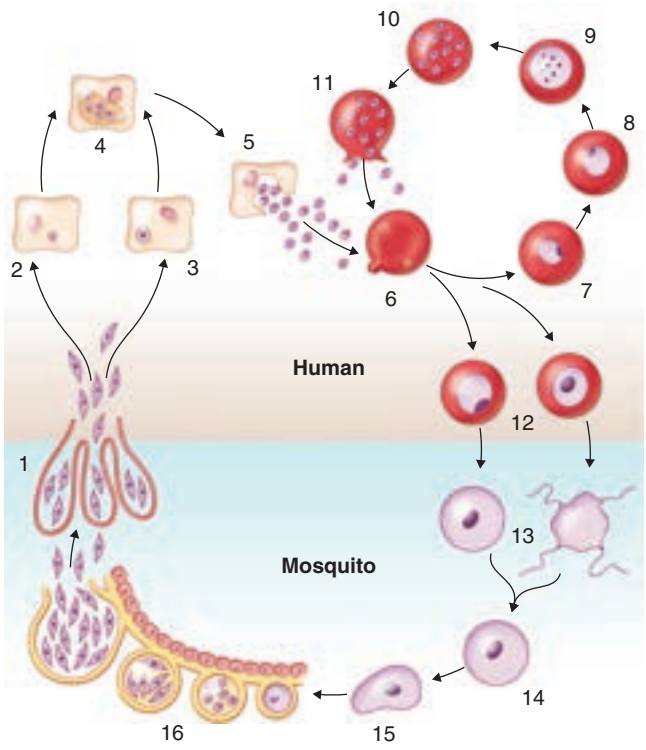

FIGURE 345-1. Life cycle of the malaria parasite. The *upper* and *lower halves* of the diagram indicate the human and anopheline mosquito parts of the cycle, respectively. Sporozoites from the salivary gland of a female *Anopheles* mosquito are injected under the skin (1). They then travel through the blood stream to the liver (2) and mature within hepatocytes to become tissue schizonts (4). Up to 30,000 parasites are then released into the blood stream as merozoites (5) and produce symptomatic infection as they invade and destroy red blood cells. However, in *Plasmodium vivax* and *Plasmodium ovale* infection, some parasites remain dormant in the liver as hypnozoites (3), which can later develop to tissue schizonts and merozoites, leading to relapse. Within the blood stream, merozoites (5) invade erythrocytes (6) and mature to the ring (7, 8), trophozoite (9), and schizont (10) asexual stages. Mature schizonts lyse their host erythrocytes and release the next generation of merozoites (11), which invade previously uninfected cells. Some erythrocytic parasites differentiate to sexual forms (male and female gametocytes) (12). When taken up by a mosquito, the gametocytes mature to male and female gametes (13), fuse to form zygotes (14), and then develop into ookinetes that invade the gut of the mosquito (15) and develop into oocysts (16). Mature oocysts produce sporozoites, which migrate to the salivary gland of the mosquito (1) to allow another human infection. (From Krogstad DJ: Blood and tissue protozoa. In: Schaechter M, Medoff G, Eisenstein BI, eds. *Mechanisms of Microbial Diseases.* 2nd ed. Baltimore: Williams & Wilkins; 1993:600.)

erythrocytes[5] to begin the asexual erythrocytic stage of infection that is responsible for human disease. Multiple rounds of erythrocytic development, with production of merozoites that invade additional erythrocytes, lead to large numbers of circulating parasites and clinical illness. Each erythrocytic cycle lasts approximately 24 hours for *P. knowlesi*; 48 hours for *P. falciparum*, *P. vivax*, and *P. ovale;* and 72 hours for *P. malariae*. Some erythrocytic parasites also develop into sexual gametocytes, which are taken up by mosquitoes. In the mosquito, gametocytes mature to gametes, and after fusion of male and female gametes to produce zygotes, parasites develop into ookinetes, oocysts, and then salivary gland sporozoites that are infectious for humans, allowing completion of the life cycle and infection of others. *P. vivax* and *P. ovale* also cause a chronic liver infection, in which hypnozoites persist in hepatocytes in a dormant state not eradicated by most therapies for acute disease and subsequently can progress to erythrocytic infection and a relapse of clinical illness.

Pathogenic Features of Malaria Parasites

The most common clinical feature of malaria is fever. Fever coincides with rupture of large numbers of schizont-infected erythrocytes at the completion of the erythrocytic cycle and with high circulating levels of tumor necrosis factor-α (TNF-α). Severe falciparum malaria is associated with very high levels of TNF-α and other inflammatory cytokines, but the specific roles of cytokines in pathogenesis are not well understood. *P. falciparum* infects erythrocytes of all ages, so it is capable of routinely causing high parasitemias, with infection of more than 1% of erythrocytes and more than 10^5 infected erythrocytes per microliter of blood. Non–*P. falciparum* parasites infect smaller numbers of erythrocytes, limiting the extent of infection and morbidity. Non–*P. falciparum* parasites are more likely to cause highly synchronous infections, leading if untreated to regular cycles of fever every 48 (*P. vivax* and *P. ovale*) or 72 (*P. malariae*) hours, often with minimal symptoms between fever episodes.

The contribution of parasite virulence determinants to the severity of malaria is poorly understood. A key biologic feature of *P. falciparum* infection is the ability of parasites to mediate the adherence of infected erythrocytes to a number of ligands on endothelial cells. By this mechanism, erythrocytes infected with the more mature stages of erythrocytic parasites do not circulate but rather adhere within small blood vessels in the brain and other organs. This phenomenon, termed cytoadherence, allows parasites to avoid passing through the spleen, where abnormal erythrocytes would be cleared. Cytoadherence is also likely to play a major role in mediating severe manifestations of *P. falciparum* malaria, with local inflammatory changes mediated by large numbers of adherent parasites leading to organ dysfunction. In particular, *P. falciparum* malaria can progress to cerebral malaria, including coma; noncardiogenic pulmonary edema, including severe respiratory compromise; and renal failure, severe anemia, acidosis, hypoglycemia, and other syndromes of organ dysfunction. Pregnancy selects for a subset of *P. falciparum* strains that specifically bind to ligands in the placenta. Pregnant women in endemic areas, in particular those in their first pregnancy who lack antibodies specific for placenta-binding parasites, are at high risk of morbidity, including anemia, from high parasite loads in the placenta and of poor fetal outcomes, including intrauterine growth retardation, spontaneous abortion, and low birthweight.

A common cause of death from *P. falciparum* malaria, in particular in children in endemic areas, is severe anemia. Anemia is caused by destruction of infected and uninfected erythrocytes, decreased hematopoiesis, and bleeding. In many endemic areas, most asymptomatic young children are infected with *P. falciparum,* with chronic infection contributing to chronic anemia. Other factors contributing to anemia are nutritional deficits and intestinal roundworm infections. With frequent malarial infections and chronic anemia, children are ill equipped to manage acute worsening of the anemia caused by acute malarial illness. With limited access to health care, children often present for medical care late in the course of illness if at all, and many deaths result.

P. falciparum parasites use antigenic variation to evade the host immune response. The principal protein that mediates the cytoadherence of infected erythrocytes to endothelial cells, *P. falciparum* erythrocyte membrane protein-1 (PfEMP-1), is transported to the erythrocyte surface and is a target of host immune responses that limit infection. The PfEMP-1 family comprises about 60 proteins, but only one PfEMP-1 variant is expressed on the surface of infected erythrocytes at a time. During the course of an infection, parasites frequently vary the expression of PfEMP-1s to stymie host responses. This factor and the high variability in sequence of many PfEMP-1 molecules present a broad repertoire of antigens and probably help explain the slow acquisition of protective antimalarial immunity. Antigenic variation and other aspects of immunologic diversity are poorly understood for non–*P. falciparum* malaria parasites, but each species appears capable of repeated infections.

Plasmodial parasites other than *P. falciparum* do not cause cytoadherence of infected erythrocytes, infect lower numbers of erythrocytes, and are much less commonly responsible for complicated and severe disease. However, recent reports suggest that *P. vivax* can cause severe disease, in particular respiratory compromise, more commonly than has previously been appreciated. *P. vivax* is also particularly prone to cause splenic rupture, although this complication can be seen with all malarial species. *P. malariae* most commonly causes a mild febrile illness, but chronic or repeated infection has been associated with an immune complex–mediated glomerulonephritis with nephrotic syndrome (Chapter 121). Recent reports indicate that *P. knowlesi* can cause severe illness, including deaths. The short (24-hour) erythrocytic cycle of *P. knowlesi,* which allows more rapid replication of these parasites compared with others, may partly explain the propensity of this zoonotic parasite to cause severe illness.

Host Immunity and Genetics

The nature of human immune responses to malaria remains poorly characterized, but protective responses require multiple infections and apparently humoral and cell-mediated responses. Where *P. falciparum* malaria is common, disease occurs primarily in children. After some protection during the first few months of life, probably because of protective effects of maternal antiplasmodial antibodies and fetal hemoglobin, young children are infected frequently, experience repeated febrile malaria illnesses, and are at high risk of severe disease. With repeated episodes of malaria, children develop partial immunity. Immunity develops gradually, with some protection against severe malaria after only a few infections, increasing protection against symptomatic illness in children, and eventually strong protection against infection. Thus, in highly endemic areas, young children experience frequent episodes of malaria, especially at about age 6 months to 5 years; older children are frequently infected but uncommonly symptomatic; and adults experience identifiable malarial parasitemia only uncommonly. However, antimalarial immunity is not complete; malaria can occur in individuals of any age. In addition, immunity requires boosting by repeated infections, so adults are at increased risk of disease if they return to a highly endemic site after an extended stay in a nonendemic area.

A number of human genetic polymorphisms offer protection against malaria.[6] The best characterized is sickle hemoglobin (Chapter 163). Hemoglobin S heterozygotes are partially protected against severe *P. falciparum* malaria, leading to a balanced polymorphism in which the survival advantage of the polymorphism allows persistence of sickle cell disease in homozygotes. Other erythrocyte polymorphisms that also are likely to offer protection against malaria include hemoglobin C and E, thalassemias (Chapter 162), glucose-6-phosphate dehydrogenase deficiency (Chapter 161), and ovalocytosis (Chapter 161). The Duffy antigen, an erythrocyte chemokine receptor of uncertain function, is the principal receptor on human erythrocytes for attachment and subsequent invasion of *P. vivax*. Most Africans lack the erythrocyte Duffy antigen, explaining the nearly complete absence of *P. vivax* malaria in most of Africa.

CLINICAL MANIFESTATIONS

Uncomplicated Malaria

Most malarial episodes, even with *P. falciparum* infection, are uncomplicated. The incubation period after an infectious bite is usually 10 to 14 days for *P. falciparum* and about 2 weeks for other species, but this can be much longer, especially in non–*P. falciparum* malaria and in individuals with prior immunity. The hallmark of malaria is fever, often with a nonspecific influenza-like prodrome including headache and fatigue followed by a classic malarial paroxysm including chills, high fever, and then sweats. Patients may be remarkably well between febrile episodes. Fevers are typically irregular early in the illness but without therapy may become regular, with 48-hour (*P. vivax* and *P. ovale*) or 72-hour (*P. malariae*) cycles, especially with non–*P. falciparum* disease. Headache, malaise, myalgias, arthralgias, rigors, confusion, cough, chest pain, abdominal pain, anorexia, nausea, vomiting, and diarrhea are common. Seizures are often simple febrile convulsions, especially in young children, but they also may represent evidence of severe neurologic disease. Physical findings may be absent or include signs of anemia, jaundice, splenomegaly, and mild hepatomegaly. Rash and lymphadenopathy are not typical

in malaria and thus are suggestive of another cause of fever. Laboratory studies commonly show anemia, thrombocytopenia, and liver and renal function abnormalities.

Severe *P. falciparum* Malaria

Severe malaria can be defined as presentation with signs of severe illness or organ dysfunction (including prostration, impaired consciousness, convulsions, respiratory distress, shock, acidosis, severe anemia, excessive bleeding, hypoglycemia, jaundice, hemoglobinuria, and renal impairment) or a high parasite load (generally peripheral parasitemia >5% or >200,000 parasites/μL). Cerebral malaria, the most common severe complication in children, is generally defined as altered consciousness in the setting of *P. falciparum* malaria. Seizures are common, and deep coma, abnormal posturing, focal neurologic findings, and abnormal respiratory patterns can be seen. The mortality rate is 15% to 25%, with about 10% with persistence of neurologic sequelae, but many patients do remarkably well with appropriate therapy.

Severe anemia is a common presentation, particularly in young children. Transfusions are avoided when possible but play a key role in management of those with severe compromise. Respiratory failure is caused by noncardiogenic pulmonary edema and is more common in adults than in children. Mechanical ventilation can be life saving, if available. Acute renal failure is also more common in adults and generally due to hypoperfusion and acute tubular necrosis; hemofiltration and hemodialysis are valuable, if available. Blackwater fever, including intravascular hemolysis and hemoglobinuria, has an uncertain etiology but can be caused by quinine. Hepatic dysfunction, including jaundice, can be seen; jaundice may also be caused by hemolysis. Splenic enlargement is common, and splenic rupture can be seen. Hypoglycemia is common because of glucose consumption by parasites, increased demand, impaired gluconeogenesis, and quinine-induced insulin secretion; blood glucose levels should be observed closely, with glucose replacement as needed. Metabolic acidosis caused by lactic acidosis and other factors is common; the value of specific therapies for acidosis or aggressive fluid resuscitation is uncertain. Electrolyte derangements can be seen. Coagulopathy, caused by consumption of clotting factors, and marked thrombocytopenia, caused by increased platelet turnover, can lead to significant bleeding. Bacterial infection and sepsis can coexist with malaria; presumptive use of antibiotics is warranted when signs of sepsis are seen.

Complications of Non–*P. falciparum* Malaria

The large majority of infections with non–*P. falciparum* parasites are uncomplicated both in endemic areas and in nonimmune travelers. Nonetheless, *P. vivax* infection is common in many areas, and studies from a number of sites in Asia and Oceania have shown that it makes up about a quarter of children hospitalized with severe malaria, with a mortality rate of about 1%. Important features of severe *P. vivax* malaria include severe anemia and respiratory distress. All malarias, but in particular *P. vivax* infection, can be complicated by splenic rupture. Chronic malaria infections can be complicated by hyperreactive malarial splenomegaly, with massive splenomegaly and findings of hypersplenism. Chronic infections can also lead to the nephrotic syndrome, particularly with *P. malariae* infection.

DIAGNOSIS

Clinical Features

In individuals with febrile illness and malaria risk, it is essential to make a diagnosis promptly and important to distinguish the different species that infect humans because management differs according to the infecting species. Malaria is the most common cause of febrile illness in many areas, and with limited diagnostic capabilities, it is frequently diagnosed based only on presentation with a febrile illness. However, formal diagnosis is preferred. In travelers returning from endemic areas with fever, historical details can aid in the diagnosis. Malaria is most likely in individuals who failed to use measures to prevent infection and in travelers to the most highly endemic areas, such as rural sub-Saharan Africa. *P. falciparum* malaria generally has a fairly short incubation period in nonimmune individuals, so it presents within 1 to 2 months of return in more than 90% of travelers. Infection with the other malarial species can present a number of months and uncommonly more than 1 year after exposure.

Blood Smears

The standard means of diagnosis in malaria-endemic areas is by thick blood smear. In this procedure, 1 drop of blood is allowed to dry on a slide, erythrocytes are lysed, and parasites are then stained with Giemsa. Parasites are easily identified by trained microscopists, and parasite density can be estimated on the basis of counts relative to those of leukocytes. However, thick blood smears do not allow identification of erythrocyte morphology, which is helpful in species diagnosis, and are difficult for those with limited training. Giemsa-stained thin blood smears offer an improved means of characterizing parasite morphology, but the process is much less efficient than for thick smears. Thus, thick smears are the standard means of diagnosis in highly endemic areas, and thin smears are preferred where malaria is uncommon and where laboratory personnel have ample time to examine multiple microscopic fields. It is important to distinguish infecting species of malaria parasites. In *P. falciparum* infection, generally only ring-form asexual parasites are seen. Trophozoites of *P. vivax* and *P. ovale* are present in enlarged (and ovoid in the case of *P. ovale*) erythrocytes that contain inclusions known as Schüffner's dots. Intraerythrocytic *P. malariae* and *P. knowlesi* trophozoites are often elongate in shape. Sexual stage gametocytes (which have a characteristic oblong shape in *P. falciparum*) are also seen on blood smears; most treatments do not eradicate gametocytes, so persistence of these forms for a few weeks is not a sign of treatment failure.

Antigen Detection

An important new means of malaria diagnosis is antigen detection. Multiple simple tests are now available, incorporating colorimetric detection of one or two antigens in an assay that requires limited training and only a few minutes.[7] The most used assays in Africa use histidine-rich protein-2, a protein that is abundant and long-lived but expressed only by *P. falciparum*. Other assays identify plasmodial lactate dehydrogenase and aldolase, which are produced by all human malarial species. Some tests use two antigens to separately identify *P. falciparum* and all-species plasmodial infection. Rapid diagnostic tests have recently become a standard component of many malaria control programs. One test is approved in the United States. However, with many different tests available around the world, standardization is not optimal, and the specific role of rapid antigen tests for the diagnosis of malaria in different epidemiologic settings is not yet established.

Other Diagnostic Tests

Serologic tests are available to identify prior malaria infection, but responses develop slowly and persist for an extended period, so these tests have limited clinical value unless it is helpful to diagnose the cause of a febrile illness retrospectively in an individual without prior history of malaria. Malaria parasites can be identified readily with polymerase chain reaction (PCR) using primers encoding genus- and species-specific sequences. These tests are convenient for research purposes because they can be conducted on DNA extracted from blood spotted onto filter paper in field settings. PCR is more sensitive than other diagnostic modalities. It is not practical for routine diagnosis because of the time required to complete an assay, the uncertain significance of a positive test result in endemic areas where low-level parasitemia may be common and clinically insignificant, and the cost and logistical requirements of the procedure in developing countries.

TREATMENT Rx

The treatment of falciparum malaria has been challenged by drug resistance for many years[8,9] and has changed dramatically recently, with recommendations for artemisinin-based combination therapy (ACT) in nearly all countries endemic for *P. falciparum* malaria (Table 345-1). Non–*P. falciparum* malaria is still generally treated with chloroquine, although chloroquine resistance is increasing in *P. vivax*; resistant vivax malaria can be treated with other drugs that are standard to treat falciparum malaria.

Chloroquine and Other Aminoquinolines

Chloroquine has been widely used to treat malaria for more than 60 years. It remains the treatment of choice for non–*P. falciparum* malaria and *P. falciparum* malaria in the few areas where resistance has not been seen (primarily Central America and the Caribbean) and is generally rapidly effective and well tolerated. For *P. vivax* and *P. ovale* infections, primaquine must also be given to eradicate dormant liver forms and thereby prevent subsequent relapse. Chloroquine remains effective as weekly chemoprophylaxis to prevent malaria in areas without resistance. Amodiaquine and piperaquine probably share mechanisms of action with chloroquine, but they are routinely active against chloroquine-resistant parasites because of increased potency compared with chloroquine and some differences in mechanisms of resistance. Monotherapy with either drug is discouraged, but each is now a component of a leading ACT (Table 345-2). Amodiaquine is somewhat less well tolerated than other

TABLE 345-1 TREATMENT OF MALARIA*

CHLOROQUINE-RESISTANT *PLASMODIUM FALCIPARUM*, RESISTANT *PLASMODIUM VIVAX*, OR SPECIES UNIDENTIFIED	
UNCOMPLICATED DISEASE	
Coartem (artemether 20 mg, lumefantrine 120 mg)	4 tablets orally twice daily for 3 days
or	
Malarone (atovaquone 250 mg, proguanil 100 mg)	4 tablets daily for 3 days
or	
Quinine	650 mg quinine sulfate 3 times daily for 3-7 days
plus	
Doxycycline	100 mg twice daily for 7 days
or plus	
Clindamycin	600 mg twice daily for 7 days
or	
Mefloquine	750 mg followed by 500 mg in 6-8 hours; can also be given as a single 1250-mg dose, although this is less well tolerated than the divided dose
COMPLICATED *P. FALCIPARUM* MALARIA OR UNABLE TO TOLERATE ORAL MEDICATIONS†	
IV artesunate‡	2.4 mg/kg every 12 hr on day 1; then daily for 2 additional days
or	
IV quinidine gluconate§	10 mg/kg over 1 to 2 hr; then 0.02 mg/kg/min *or* 15 mg/kg over 4 hr; then 7.5 mg/kg over 4 hr every 8 hr
or	
IV quinine dihydrochloride§,‖	20 mg/kg over 4 hr; then 10 mg/kg every 8 hr
or	
IM artemether‖	3.2 mg/kg IM; then 1.6 mg/kg/day
CHLOROQUINE-SUSCEPTIBLE *P. FALCIPARUM* AND OTHER SPECIES	
Chloroquine phosphate	1 g, followed by 500 mg at 6, 24, and 48 hr *or* 1 g at 0 and 24 hr; then 0.5 g at 48 hr
plus (for *P. vivax* and *P. ovale* only)	
Primaquine¶	30-mg base (52.6 mg primaquine phosphate) daily for 14 days

*Dosages refer to salts unless indicated and are for adults. For pediatric dosing and full Centers for Disease Control and Prevention (CDC) recommendations, see http://www.cdc.gov/malaria/pdf/treatmenttable.pdf.
†IV regimens should be given until the patient can tolerate oral agents and then followed by a course of oral therapy (doxycycline, clindamycin, or full treatment courses of other drugs, as listed) when patients can tolerate this.
‡Available in the United States on an investigational basis through the CDC.
§Cardiac monitoring should be in place during IV administration of quinidine or quinine.
‖Not available in the United States.
¶Use primaquine only after demonstrating normal levels of glucose-6-phosphate dehydrogenase.
IM, Intramuscular; *IV,* intravenous.

TABLE 345-2 RECOMMENDATIONS FOR THE TREATMENT OF *PLASMODIUM FALCIPARUM* MALARIA IN DEVELOPING COUNTRIES*

REGIMEN	NOTES
Artemether–lumefantrine (Coartem, Riamet)	First-line therapy in many countries; FDA approved
Artesunate–amodiaquine (ASAQ)	First-line therapy in many African countries; efficacy limited by resistance to amodiaquine in many areas
Artesunate–mefloquine	Standard therapy in parts of Southeast Asia
Artesunate–sulfadoxine–pyrimethamine	Efficacy low compared with other regimens in most areas
Dihydroartemisinin–piperaquine	Newer regimen; highly effective in studies in Asia and Africa

*Recommendations modified from World Health Organization. *Guidelines for the Treatment of Malaria.* Geneva: World Health Organization; 2010.
FDA, Food and Drug Administration.

aminoquinolines. It is generally safe with short-term use but can cause rare serious toxicities, including hepatic and bone marrow toxicity, especially with chronic use, and so it is not recommended for chemoprophylaxis. Piperaquine was well tolerated during extensive use in China a few decades ago; although resistance was reportedly common in China, it does not appear to be a problem at present.

Mefloquine, Halofantrine, and Lumefantrine

Mefloquine offers effective therapy and chemoprophylaxis for most chloroquine-resistant strains of *P. falciparum* and for other species. Resistance to mefloquine is uncommon but has been seen in parts of Southeast Asia. Mefloquine is now used to treat *P. falciparum* malaria in combination with artesunate and is one of three drugs recommended for chemoprophylaxis against *P. falciparum* by the Centers for Disease Control and Prevention (CDC). Tolerability of chemoprophylactic and especially treatment doses of mefloquine is often limited by neurologic and gastrointestinal (GI) toxicity. Halofantrine is an effective drug; however, its use is limited by uncommon but dangerous cardiac rhythm disturbances. Lumefantrine is a related drug that does not share this toxicity and offers effective therapy in combination with artemether as Coartem.

Quinine and Quinidine

Quinine has been used to treat malaria for hundreds of years. It offers rapid action against all species, with limited known resistance except in Southeast Asia, where failures against *P. falciparum* malaria are fairly common. Quinine can be used to treat uncomplicated malaria, but GI and other nonspecific toxicities lead to difficulty in tolerating a full 7-day treatment course. This problem is circumvented by combining a 3-day course of quinine with other agents. For severe disease, intravenous (IV) quinine or, in the United States, IV quinidine has been standard therapy for many years; therapy should be accompanied by cardiac monitoring because of cardiac toxicities of the drugs.

Primaquine

Primaquine is the only available drug to eradicate dormant liver forms of *P. vivax* and *P. ovale* that are not cleared by other drugs and can lead to relapses after therapy with chloroquine and other agents. Primaquine is also an alternative agent for chemoprophylaxis against *P. falciparum* and other species. Primaquine use is limited principally by hemolysis or methemoglobinemia (Chapter 158) in individuals with deficiency in glucose-6-phosphate dehydrogenase (Chapter 161). Testing for deficiency should be performed before use of the drug.

Inhibitors of Folate Metabolism

Inhibitors of two parasite enzymes involved in folate metabolism, dihydrofolate reductase and dihydropteroate synthase, are used in fixed-dose combination regimens for the treatment and prevention of malaria. For treatment, sulfadoxine–pyrimethamine (Fansidar) was heavily used to treat uncomplicated *P. falciparum* malaria, but resistance has increased markedly in most endemic areas. The dihydrofolate reductase inhibitor proguanil is combined with atovaquone in Malarone (see later). For chemoprophylaxis, sulfadoxine–pyrimethamine is no longer recommended because of drug resistance and rare life-threatening dermatologic toxicities. However, less frequent dosing in intermittent preventive therapy regimens has been well tolerated and has decreased malaria in high-risk African groups, particularly pregnant women and young children. Seasonal malaria chemoprevention with monthly combined sulfadoxine–pyrimethamine and amodiaquine during the transmission season is now recommended for malaria control in areas of Africa with highly seasonal transmission and limited drug resistance. Daily trimethoprim–sulfamethoxazole, a common prophylactic to prevent opportunistic infections in individuals with HIV infection, has offered some protection against malaria in Africa.

Artemisinins

Artemisinin, the active component of an herbal medicine from China, and a number of its analogues offer rapid elimination of circulating malaria parasites and activity against gametocytes to limit disease transmission. The drugs are all short-acting, leading to frequent recrudescences of infection after short-course monotherapy. For this reason and to limit selection of resistance, artemisinins are now used in combination with longer acting drugs to treat malaria in 3-day regimens. A number of these combinations have become the standard therapies for *P. falciparum* malaria in most endemic

countries (see Table 345-2). Leading regimens are fixed-dose combinations of artemether–lumefantrine, artesunate–amodiaquine, artesunate–mefloquine, and dihydroartemisinin–piperaquine, and artesunate–pyronaridine will soon be available.[A1] In some settings rank orders of efficacy are seen for leading ACTs, typically with optimal efficacy for artemether–lumefantrine or dihydroartemisinin–piperaquine, although results have varied based on study details and geography. Of these, artemether–lumefantrine is most widely advocated, is now approved in the United States, and includes a partner drug that has never been available as monotherapy, but it requires twice-daily therapy. The other regimens offer once-daily therapy but have potential problems with resistance to partner drugs. Resistance to artemisinins is a recent concern, with evidence for prolonged times to parasite clearance in Southeast Asia suggestive of diminished drug responsiveness of *P. falciparum*.[9-11] Artemisinins also have an increasing role for the treatment of severe malaria. IV artesunate has been shown to be superior to quinine for the treatment of severe malaria in a mostly adult population in Asia[A2] and in African children,[A3] notably with survival advantages over quinine in both populations. For settings with limited infrastructure, intramuscular artemether and intrarectal artesunate have also shown good efficacy. IV artesunate should now be considered the first-line therapy for severe malaria in the United States, where it is available from the CDC. Artemisinins are generally very well tolerated, with minimal toxicity.

Atovaquone–Proguanil (Malarone)

This fixed-dose combination of a dihydrofolate reductase inhibitor and atovaquone, which has a unique antimalarial mechanism, has excellent efficacy against most *P. falciparum* infections. It is approved for both treatment and chemoprophylaxis of *P. falciparum* malaria and other species in the United States, where it is now widely used for both indications. Malarone offers excellent efficacy with minimal toxicity. Adverse effects may include GI symptoms, elevations in liver enzymes, headache, and rash. Widespread use in developing countries is limited by high cost and concerns about resistance because resistance to each component drug is readily selected.

Antibiotics

A number of antibacterials are slow-acting antimalarials. Tetracyclines and clindamycin should not be used alone to treat malaria but are combined with quinine to allow a shorter duration of therapy with that drug. In addition, doxycycline is effective in chemoprophylaxis of most *P. falciparum* malaria and is recommended for this purpose by the CDC, in particular for travelers to regions of Southeast Asia with high-level resistance to other drugs. Macrolides also have antimalarial activity and are currently under study for use against malaria.

Treatment of Severe Malaria

Severe malaria is a medical emergency and requires parenteral therapy. With appropriate prompt therapy and supportive care, rapid recoveries may be seen even in very ill individuals. As noted before, standard therapy for severe malaria has been IV quinine or quinidine, with continuous cardiac monitoring if possible. However, as noted above, IV artesunate showed superior efficacy and decreased toxicity. Although it is not yet available in all endemic areas, it has become the worldwide standard to treat severe malaria. In the United States, IV quinidine is not available in all hospitals and IV artesunate must be obtained from the CDC; initial treatment should be with whichever drug is most rapidly available. Appropriate care of severe malaria includes close nursing care; maintenance of fluids, electrolytes, and glucose; respiratory and hemodynamic support; and consideration of blood transfusions, anticonvulsants, antibiotics for bacterial infections, and hemodialysis or hemofiltration. Aggressive fluid resuscitation, blood transfusion for moderate anemia, exchange transfusion, and specific treatment of acidosis are of uncertain value. After the acute illness, parenteral quinine, quinidine, or artesunate should be followed with oral longer acting drugs, typically a full course of an oral ACT, Malarone, mefloquine, or quinine plus doxycycline or clindamycin.

PREVENTION

Recent years have been marked by a dramatic increase in interest and investment in the control of malaria. Key malaria control interventions in malaria-endemic regions are control of mosquito vectors by indoor residual spraying of insecticides; personal protection against mosquito bites with insecticide-impregnated bed nets; routine use of artemisinin-based combination therapies, which offer prompt control of malaria infections and activity against gametocytes to limit transmission to mosquitoes; and selected use of intermittent preventive therapies to decrease malarial incidence in high-risk groups. No vaccine to prevent malaria is yet available, but extensive research on potential vaccines is under way. RTS,S, which is based on an immunogenic sporozoite antigen, is the most advanced vaccine candidate. Recent clinical trials have consistently shown protection in children immunized with RTS,S, with protection against malaria of about 30% to 60% in the year after immunization, but lower levels of protection in very young children, in areas of highest malaria exposure, and when assessed over longer periods of time after immunization.[A1]

TABLE 345-3 CHEMOPROPHYLAXIS OF MALARIA*

AREAS WITH CHLOROQUINE-RESISTANT *PLASMODIUM FALCIPARUM*	
Malarone	1 tablet (250 mg artesunate/100 mg proguanil) daily
Mefloquine	250 mg weekly
Doxycycline	100 mg daily
Primaquine†	30 mg daily during exposure (chemoprophylaxis) or 30 mg daily for 14 days (terminal prophylaxis against *P. vivax* and *P. ovale*)
AREAS WITHOUT CHLOROQUINE-RESISTANT *P. FALCIPARUM*	
Chloroquine phosphate	500 mg weekly

*Recommendations may change on the basis of drug resistance patterns. For additional details and pediatric dosing, see Centers for Disease Control and Prevention guidelines (http://www.cdc.gov). Begin 1 to 2 weeks before travel for mefloquine and 2 days before for doxycycline, Malarone, and primaquine; continue for 4 weeks after leaving the endemic area (1 week for Malarone; 2 weeks for primaquine). All doses refer to salts unless indicated.
†Use primaquine only after demonstrating normal levels of glucose-6-phosphate dehydrogenase.

Preventive Measures for Travelers to Malaria-Endemic Regions

It is important for nonimmune travelers (Chapter 286) to endemic areas to be protected against potentially lethal malaria. Travelers should decrease exposure to night-biting anopheline mosquitoes by use of insecticide repellants and sleeping in rooms that are screened or equipped with insecticide-impregnated bed nets. Standard advice for travelers to endemic areas is also to use low doses of preventive drugs chosen on the basis of the resistance profile of the particular region. Chloroquine is still recommended for malaria-endemic regions of Central America and the Caribbean. For nearly all other areas, the CDC recommends use of daily Malarone, weekly mefloquine, or daily doxycycline; details of dosing vary, but it is important to continue therapy after return from travel to eliminate parasites as they emerge from the liver (Table 345-3). For areas with high risk of *P. vivax* malaria, some authorities recommend a full treatment course of primaquine after travel to eliminate dormant liver stages. For all chemoprophylaxis, it is important to appreciate that no mosquito avoidance methods or drug regimens are fully protective, so consideration of malaria as a cause of fever in returned travelers is essential (Chapter 286).

PROGNOSIS

Patients with malaria caused by *P. vivax, P. ovale,* or *P. malariae* generally respond well to chloroquine and make an uneventful recovery. Chloroquine-resistance is increasing with *P. vivax* from many areas; failures of initial treatment are usually not dangerous but should be followed by treatment with another regimen, such as an ACT, Malarone, or mefloquine. Patients with *P. falciparum* malaria also generally respond well to prompt therapy as long as the disease is not overly advanced at presentation. The mortality rate in those with uncomplicated *P. falciparum* malaria is about 0.1%. Key contributing factors to most deaths from *P. falciparum* malaria are probably a delay between the appearance of symptoms and presentation for definitive therapy and the use of suboptimal therapies. Presentation with high-level parasitemia (>200,000 parasites/μL or >5% parasitemia) or signs of severe malaria are predictive of a poor outcome. However, with aggressive support, even individuals with severe disease can often experience complete recoveries.

Grade A References

A1. Mutabingwa TK, Anthony D, Heller A, et al. Amodiaquine alone, amodiaquine+sulfadoxine-pyrimethamine, amodiaquine+artesunate, and artemether-lumefantrine for outpatient treatment of malaria in Tanzanian children: a four-arm randomised effectiveness trial. *Lancet.* 2005;365:1474-1480.

A2. Smithuis F, Kyaw MK, Phe O, et al. Efficacy and effectiveness of dihydroartemisinin-piperaquine versus artesunate-mefloquine in falciparum malaria: an open-label randomised comparison. *Lancet.* 2006;367:2075-2085.

A3. Yeka A, Lameyre V, Afizi K, et al. Efficacy and Safety of Fixed-Dose Artesunate-Amodiaquine vs. Artemether-Lumefantrine for Repeated Treatment of Uncomplicated Malaria in Ugandan Children. *PLoS ONE.* 2014;9:e113311.

A4. Gogtay N, Kannan S, Thatte UM, et al. Artemisinin-based combination therapy for treating uncomplicated Plasmodium vivax malaria. *Cochrane Database Syst Rev*. 2013;10:CD008492.
A5. Zongo I, Dorsey G, Rouamba N, et al. Randomized comparison of amodiaquine plus sulfadoxine-pyrimethamine, artemether-lumefantrine, and dihydroartemisinin-piperaquine for the treatment of uncomplicated *Plasmodium falciparum* malaria in Burkina Faso. *Clin Infect Dis*. 2007;45:1453-1461.
A6. Esu E, Effa EE, Opie ON, et al. Artemether for severe malaria. *Cochrane Database Syst Rev*. 2014;9:CD010678.
A7. Tshefu AK, Gaye O, Kayentao K, et al. Efficacy and safety of a fixed-dose oral combination of pyronaridine-artesunate compared with artemether-lumefantrine in children and adults with uncomplicated Plasmodium falciparum malaria: a randomised non-inferiority trial. *Lancet*. 2010;375:1457-1467.
A8. Four Artemisinin-Based Combinations (4ABC) Study Group. A head-to-head comparison of four artemisinin-based combinations for treating uncomplicated malaria in African children: a randomized trial. *PLoS Med*. 2011;8:e1001119.
A9. Zani B, Gathu M, Donegan S, et al. Dihydroartemisinin-piperquine for treating uncomplicated Plasmodium falciparum malaria. *Cochrane Database Syst Rev*. 2014;1:CD010927.
A10. Bukirwa H, Unnikrishnan B, Kramer CV, et al. Artesunate plus pyronaridine for treating uncomplicated Plasmodium falciparum malaria. *Cochrane Database Syst Rev*. 2014;3:CD006404.
A11. Dondorp A, Nosten F, Stepniewska K, et al. Artesunate versus quinine for treatment of severe falciparum malaria: a randomised trial. *Lancet*. 2005;366:717-725.
A12. Dondorp AM, Fanello CI, Hendriksen IC, et al. Artesunate versus quinine in the treatment of severe falciparum malaria in African children (AQUAMAT): an open-label, randomised trial. *Lancet*. 2010;376:1647-1657.
A13. Agnandji ST, Lell B, Soulanoudjingar SS, et al. First results of phase 3 trial of RTS,S/AS01 malaria vaccine in African children. *N Engl J Med*. 2011;365:1863-1875.
A14. Agnandji ST, Lell B, Fernandes JF, et al. A phase 3 trial of RTS,S/AS01 malaria vaccine in African infants. *N Engl J Med*. 2012;367:2284-2295.
A15. Olotu A, Fegan G, Wambua J, et al. Four-year efficacy of RTS,S/AS01E and its interaction with malaria exposure. *N Engl J Med*. 2013;368:1111-1120.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. White NJ, Pukrittayakamee S, Hien TT, et al. Malaria. *Lancet*. 2014;383:723-735.
2. Anstey NM, Douglas NM, Poespoprodjo JR, Price RN. Plasmodium vivax: clinical spectrum, risk factors and pathogenesis. *Adv Parasitol*. 2012;80:151-201.
3. Cotter C, Sturrock HJ, Hsiang MS, et al. The changing epidemiology of malaria elimination: new strategies for new challenges. *Lancet*. 2013;382:900-911.
4. Noor AM, Kinyoki DK, Mundia CW, et al. The changing risk of Plasmodium falciparum malaria infection in Africa: 2000-10: a spatial and temporal analysis of transmission intensity. *Lancet*. 2014;383:1739-1747.
5. Walker DM, Oghumu S, Gupta G, et al. Mechanisms of cellular invasion by intracellular parasites. *Cell Mol Life Sci*. 2014;71:1245-1263.
6. Taylor SM, Parobek CM, Fairhurst RM. Haemoglobinopathies and the clinical epidemiology of malaria: a systematic review and meta-analysis. *Lancet Infect Dis*. 2012;12:457-468.
7. Wilson ML. Malaria rapid diagnostic tests. *Clin Infect Dis*. 2012;54:1637-1641.
8. Lin JT, Juliano JJ, Wongsrichanalai C. Drug-resistant malaria: the era of ACT. *Curr Infect Dis Rep*. 2010;12:165-173.
9. Rosenthal PJ. The interplay between drug resistance and fitness in malaria parasites. *Mol Microbiol*. 2013;89:1025-1038.
10. Ashley EA, Dhorda M, Fairhurst RM, et al. Spread of artemisinin resistance in Plasmodium falciparum malaria. *N Engl J Med*. 2014;371:411-423.
11. Ariey F, Witkowski B, Amaratunga C, et al. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature*. 2014;505:50-55.

REVIEW QUESTIONS

1. A young American adult consults with you before travel to Kenya. Appropriate regimens for the prevention of malaria in travelers from a developed country to areas with chloroquine-resistant *P. falciparum* malaria include

A. Malarone, mefloquine, or artemisinin.
B. Malarone, mefloquine, or doxycycline.
C. Coartem, Malarone, or doxycycline.
D. quinine, artemisinin, or Fansidar.

Answer: B These three drugs are all listed as first-line regimens for the prevention of malaria in travelers. Artemisinin and artemisinin-based combination therapy (Coartem) are not appropriate because of the short half-lives of artemisinins. Quinine is not appropriate because of a short half-life and poor tolerability.

2. Reasons for the predilection for *P. falciparum* to cause severe malaria include all of the following except

A. cytoadherence of *P. falciparum*–infected erythrocytes to vascular endothelium.
B. infection of erythrocytes of all ages.
C. a high prevalence of resistance to available antimalarial drugs.
D. increasing pathogenicity with increasing age of the patient.

Answer: D Major contributors to the pathogenesis of *P. falciparum* malaria include cytoadherence of P. falciparum–infected erythrocytes to vascular endothelium, infection of erythrocytes of all ages, and high prevalence of resistance to available antimalarial drugs. *P. falciparum* malaria does not generally show increasing pathogenicity with increasing age. Rather, the disease is typically most severe in young children because of their lack of prior exposure to the pathogen and thus lack of antimalarial immunity and possibly other factors.

3. The region with the greatest morbidity and mortality from malaria in the world is

A. Africa.
B. Southeast Asia.
C. South America.
D. Oceania.
E. the Middle East.

Answer: A The greatest morbidity and mortality from malaria is in Africa, likely because of many factors, including a preponderance of *P. falciparum* malaria, particularly competent anopheline mosquito vectors, and limited medical infrastructure.

4. The region with the most highly resistant malaria parasites in the world is:

A. Africa
B. Southeast Asia
C. South America
D. Oceania
E. Middle East

Answer: B Explanation. The greatest prevalence of parasites resistant to available antimalarial drugs is in Southeast Asia. Resistance to chloroquine, antifolates, mefloquine, and quinine is seen in this area, and new data demonstrate decreasing efficacy of artemisinins, manifested as delayed clearance of parasites after treatment with artesunate or artemisinin-based combination therapies.

5. A child from Ghana is admitted with fever, altered consciousness, acute renal failure, and 7% parasitemia with *P. falciparum*. Severe malaria should be treated with:

A. Oral artemisinin-based combination therapy
B. Malarone or mefloquine
C. Quinine or quinidine
D. Intravenous artesunate, if available, and as a back-up intravenous quinine or quinidine

Answer: D The new international standard-of-care for the treatment of severe malaria is intravenous artesunate. Oral therapy should be avoided in very ill individuals. Intravenous quinine or quinidine is highly effective, but intravenous artesunate was superior to intravenous quinine in randomized trials of Asians and Africans with severe malaria.

346

AFRICAN SLEEPING SICKNESS

WILLIAM A. PETRI, JR.

DEFINITION

Human African trypanosomiasis (HAT), commonly known as sleeping sickness, is a vector-borne parasitic disease transmitted to humans and animals by the bite of the tsetse fly (genus *Glossina*). Infection is caused by protozoa of the genus *Trypanosoma* and species *brucei*. In humans, there are two forms of illness caused by two distinct subspecies that are morphologically identical but differ in their geographic range and clinical presentations. *Trypanosoma brucei gambiense* is typically found in west and central Africa and *Trypanosoma brucei rhodesiense* in east Africa. *T. b. gambiense* has a more chronic course, and *T. b. rhodesiense* causes a rapid disease course; both have late stages marked by meningoencephalitis, resulting in coma and death if untreated. There is a third subspecies, *Trypanosoma brucei brucei,* known to cause chronic infection called nagana in cattle; however, humans are not susceptible to this organism.

The Pathogen

As an extracellular microbe, the parasite must evade immune clearance to establish a persistent infection. The surface of the blood stream form of the trypanosome is covered by a dense, homogeneous coat of variant surface glycoproteins (VSGs), which are immunodominant. Each individual trypanosome expresses only one VSG at a time but possesses more than 1000 different silent copies of the VSG gene, and switching to a new VSG occurs at a frequency of about 1×10^{-6} parasites. Hence, the trypanosomes expressing a given VSG will eventually be cleared by the host antibody response, but any individual trypanosome that has switched to a new VSG will evade immune clearance, resulting in a new peak of parasitemia. Recombination between VSG alleles ensures a virtually limitless repertoire of new VSGs; therefore, antibody-mediated parasite eradication is impossible.

EPIDEMIOLOGY

Human African trypanosomiasis is relatively less of a problem now than it has been in the prior century, partly because of cyclic epidemics in the past as well as recent increases in public health efforts for control. However, it remains a looming threat to an estimated 60 million people living in tsetse fly–inhabited areas among 36 countries of sub-Saharan Africa and causes significant morbidity, accounting for 1.5 million disability-adjusted life years in Africa as a whole.

The causative organism was first identified during the latter half of the 19th century, roughly corresponding to the first of three epidemics occurring during the past century. The first epidemic occurred in the Congo basin and Uganda between 1896 and 1906, driven in part by various natural disasters that decimated populations of livestock and regional droughts as well as changes in population distribution influenced by colonialism. The second epidemic occurred in numerous endemic countries in the 1920s; control was achieved after major efforts to systematically screen and treat individuals followed by extensive vector control programs, including bush clearing and insecticide spraying. These acts were nearly successful in halting transmission by the 1960s; however, the independence of many African nations around this time hindered the sustainability of prevention and control programs. The incidence began to rise by the 1970s and surged a decade later, leading to the third epidemic of the 20th century. However, in the past 15 years, increased access to at-risk populations for diagnosis and treatment has led to a 68% reduction in the incidence of reported cases.

Despite increased surveillance efforts during the past 15 years, there are still at-risk areas that lack active monitoring programs, and many people are thought to die of HAT without an accurate diagnosis. The World Health Organization reported an annual incidence of 12,000 cases in 2007, a dramatic decrease from the estimated incidence of 300,000 cases a decade earlier. It is estimated that 300,000 to 500,000 people are infected, contributing to approximately 100,000 deaths each year. More than 90% of reported cases are due to *T. b. gambiense,* the majority being from the Democratic Republic of the Congo.

The geographic distribution includes areas where the vector, parasite, reservoir hosts, and human hosts cohabit (Fig. 346-1). In general, these include focal areas on the African continent within 15 degrees North and 15 degrees South latitude, with a predilection for rural areas. Humans at greatest risk for infection are those who rely on animal husbandry, agriculture, fishing, and hunting for their livelihoods. Often, disease is concentrated among foci of rural areas, having significant socioeconomic impact on affected villages. With a few exceptions, this infection is never found in urban areas. There have

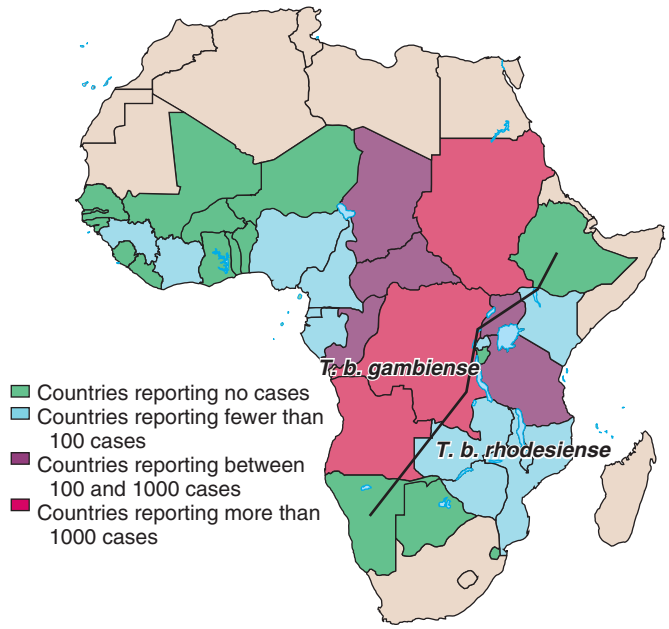

FIGURE 346-1. Map of human African trypanosomiasis (HAT). These 36 sub-Saharan African countries are considered endemic for HAT. *Shaded areas* represent the reported incidence from 1997 to 2006. The *black line* roughly represents the dividing line for the two parasites, although some overlap may already occur.

been fewer than 50 cases reported annually outside of Africa, usually the result of travel by North Americans or Europeans to African game reserves.[1]

T. b. gambiense is transmitted by tsetse flies from the *Glossina palpalis* group, the annotated genome sequence of which was recently reported. Teste flies are typically found along riverbanks among wooded areas in the more tropical regions of central and West Africa. The reservoir hosts for *T. b. gambiense* and the focus of public health campaigns are mainly humans. Because of this, *T. b. gambiense* is not generally considered a zoonotic disease; the role of the animal reservoir for this organism remains undetermined, although natural infections have been reported in domestic animals such as dogs, sheep, cattle, and pigs.

T. b. rhodesiense is transmitted by tsetse flies belonging to the *Glossina morsitans* group,[2] which are commonly found among woodland and savannah areas of east and central Africa. *T. b. rhodesiense* is a zoonotic disease with numerous wild and domestic species of animals acting as reservoirs. With this infection, the animal reservoir serves an important role in the cycle, sustaining parasite transmission and human infections. Domestic species, especially cattle, have the potential to drive outbreaks and, not surprisingly, have served as the focus of successful prevention campaigns.

PATHOBIOLOGY

After the bite of a tsetse fly carrying metacyclic trypomastigotes, a local reaction (chancre) may form at the inoculation site. This symptom is seen with *T. b. rhodesiense* infection and is more frequently observed in travelers but is rarely seen with *T. b. gambiense* infection. Parasites subsequently disseminate into the blood and lymphatic systems in what is considered stage I of the disease. Spread of the parasites into the central nervous system (CNS) defines stage II of the disease, which is invariably fatal if untreated. The parasite appears to remain extracellular throughout the course of infection.

Peaks and waves of parasitemia occur during stage I disease and result in the classic symptom of intermittent fever. These bouts of fever correspond to a type 1 inflammatory response (Chapter 48), in which classically activated macrophages produce high levels of tumor necrosis factor (TNF) and nitric oxide. This helps in the control of parasitemia but also contributes to tissue damage. Type 2 inflammatory responses with high interleukin-10 production may subsequently occur, which limit TNF and nitric oxide production after initial parasitemia has been controlled. Antibody responses are directed to VSGs and other trypanosome antigens (e.g., antigens from lysed parasites), but autoantibodies are also produced. Generalized febrile episodes are observed together with lymphadenopathy and myocardial and pericardial inflammation. Cardiac involvement is typically more severe with *T. b. rhodesiense* infection. Anemia, thrombocytopenia, disseminated intravascular coagulation, and renal disease may also be observed.

In stage II of the disease, parasites cross the blood-brain barrier and invade the CNS. Acute meningoencephalitis develops, with a variety of inflammatory cells infiltrating the brain, including macrophages, lymphocytes, plasma cells, Mott cells, and morular cells. These inflammatory cells are found in the meninges, which become thickened, as well as in the perivascular spaces and neuropil. Edema, hemorrhage, and granulomatous lesions are often present; thrombosis and neuronal degeneration may also be observed.

CLINICAL MANIFESTATIONS

The clinical manifestations of HAT differ on the basis of the infecting organism.[3] *T. b. gambiense* is more of a chronic disease, with an estimated average duration of 3 years; infection with *T. b. rhodesiense* progresses much more rapidly, leading to coma and death within weeks to months. However, infection with *T. b. gambiense* has been known to cause a rapid decline. Other similarities exist, including an early hemolymphatic stage (stage I) and a late stage characterized by prominent CNS disease including meningoencephalitis (stage II).[4,5]

West African Sleeping Sickness

Infection starts after the bite of a fly infected with *T. b. gambiense*. At the site of inoculation, a painful, indurated, and erythematous trypanosomal chancre may develop 1 to 2 weeks after the bite and resolves spontaneously after several weeks. On occasion, the chancre may ulcerate. However, these features are rarely seen at the time of clinical presentation and sometimes are not known to have occurred on history; thus, many develop disseminated disease without awareness of a localized infection.

The hemolymphatic stage, when parasites disseminate throughout the body, may not be manifested clinically until weeks or months after the initial bite. Typical symptoms include intermittent spiking fevers, occasionally accompanied by headaches and malaise. These features may persist for weeks or months because of the cyclic nature of parasitemia and antibody production against the various antigens sequentially expressed by the parasite.

Lymphadenopathy (Chapter 168) is a common finding in West African sleeping sickness. Whereas regional lymphadenopathy may develop after the initial bite, generalized lymphadenopathy around the head and neck is often found as chronic illness develops. Enlargement of nodes in the posterior cervical triangle, commonly known as Winterbottom's sign, is the classic finding, but other cervical and supraclavicular nodes may also be involved. Affected nodes are discrete, movable, rubbery nodes that are nontender to palpation; over time, they may become more indurated because of fibrosis.

Other symptoms reported include pruritus, occasionally accompanied by rash, arthralgias, and periarticular swelling as well as by transient edema of the extremities and face. Less common symptoms include features consistent with neuroendocrine dysfunction, including loss of libido and impotence, amenorrhea and infertility, alopecia, and gynecomastia. Signs of disease include hepatomegaly and splenomegaly; cardiac dysfunction, including tachycardia; electrocardiographic abnormalities such as prolonged QTc intervals and repolarization changes; and, less commonly, pericarditis or myocarditis. Hemolytic anemia and derangements in liver function test results may occur.

Months or even years after the initial infection, stage II disease develops and is characterized by headaches, daytime somnolence, and neuropsychiatric features. Behavioral changes, including irritability, confusion, inability to concentrate, and lassitude, are among the first signs of CNS disease; psychosis has also been described. Neurologic findings are numerous and include a variety of motor and sensory disturbances, including extrapyramidal features, dysesthesias, and visual impairment. These symptoms have given the infections its common name of sleeping sickness and are manifested by daytime somnolence and nocturnal irritability. Late-stage disease is marked by cerebral edema and meningoencephalitis. Progressively, a loss of neurologic function can lead to paralysis, and many may succumb to aspiration pneumonias or malnutrition; otherwise, coma leads to death in the absence of treatment.

East African Sleeping Sickness

Compared with Gambian sleeping sickness, East African sleeping sickness is more rapidly progressive. The infective bite is more frequently associated with the development of a chancre, although some studies report this in only 20% of patients. An incubation period lasting days to weeks is needed before symptoms are demonstrated. Initial symptoms include severe intermittent fevers that may resemble those found with malaria. Lymphadenopathy is not as common with this infecting organism; Winterbottom's sign is typically absent. Skin changes are more prominent, and rashes in the early stage of infection are particularly common in expatriates with the infection. In addition, cardiac manifestations are more commonly and clinically relevant; tachycardia, arrhythmias, myocarditis, and congestive heart failure have been documented and may be severe enough to cause death before the development of severe CNS disease. CNS disease mirrors that of West African sleeping sickness, but the onset occurs earlier, and the rate of deterioration is more rapid. Hematologic abnormalities include anemia, thrombocytopenia, and disseminated intravascular coagulation. Death may ensue after weeks or months without treatment.

DIAGNOSIS

Epidemiologic clues and clinical findings may together suggest the diagnosis of HAT, but definitive diagnosis relies on demonstration of the parasite. In the early stage of disease, light microscopy and Giemsa stain may be used to visualize the highly motile parasites directly from fresh specimens of fluid expressed from chancres or lymph node aspirates. Peripheral blood smears, including Giemsa-stained thick and thin smears and bone marrow aspirates, have been successful. Blood smears have increased sensitivity when they are performed during stage I disease, when parasitemia is high (Fig. 346-2); the threshold for visualization of parasites by thick smear is approximately 5000 parasites/mL. Performance is superior with *T. b. rhodesiense* infection, given the higher parasite load. If the initial smear analysis is negative, subsequent examinations should be pursued. Concentration techniques, including buffy coat examination, should be used when technically feasible. Culture of any of these fluids may yield higher sensitivity than smear preparations.

Cerebrospinal fluid (CSF) must also be analyzed to determine the most appropriate treatment. Abnormalities in CSF analysis often start with increased cell counts and progress to include an elevated opening pressure

and total protein levels with an increased polyclonal IgM. Stage II disease is defined by the presence of more than 5 white blood cells/μL, the presence of trypanosomes, or an elevated total protein count (>370 mg/L) in the CSF. Newer diagnostic methods, including polymerase chain reaction analysis of CSF and a latex-agglutination test for CSF IgM, hold promise but require validation to determine outcomes after treatment of patients with positive test results.

Allocation of resources to this neglected tropical disease in recent years has allowed slow but exciting progress in the field of HAT diagnostics. The genomes of *T. brucei* species have been mapped, and molecular assays have been developed that can distinguish among species of HAT with a single polymerase chain reaction test. Novel uses of mass spectrometry have been developed that use proteomic signature analysis to identify specific fingerprints of HAT with the host.

In contrast to these highly technical and expensive methods, other tests suitable for the field are being validated, including the dot-ELISA test, which would be able to provide information on stage of disease. Serology tests are also available for diagnosis of *T. b. gambiense* infection. The card agglutination test for trypanosomiasis with *T. b. gambiense* (CATT) is commonly used by screening programs; the sensitivity ranges from 87% to 98%, depending on the population under study, and the specificity may be as high as 95%. No serology tests are available for *T. b. rhodesiense* infection. Rapid diagnostic tests for infection with *T. b. gambiense* include HAT Sero-Strop (which uses a dipstick method) and the HAT Sero + K SeT test (which uses a lateral flow device) for testing blood or plasma, respectively, with results available in 15 minutes at an estimated price of $2.50 each.[6]

TREATMENT

Rx

There are very few drugs for treatment of HAT, and those that are commonly used are quite toxic (Table 346-1).[7] Treatment depends on the infecting organism and the stage of disease. For stage I disease with *T. b. gambiense* infection, pentamidine is the drug of choice. The standard regimen consists of daily parenteral administration for 1 week; however, studies are ongoing to determine the efficacy of shortening therapy to three doses.

Suramin is also used for stage I disease with *T. b. rhodesiense*. This medication is complex to mix and to administer. The drug must be administered in a slow intravenous infusion periodically during 3 weeks. Although anaphylaxis is rare (approximately one in 20,000 patients), a test dose is recommended before full treatment is initiated. A number of side effects require close monitoring, the most important being nephrotoxicity. Urinalysis is recommended before each dose, and the drug should be discontinued if proteinuria persists and casts are seen in the urine sediment.

Melarsoprol is a highly effective but impressively toxic drug used for stage II disease with either organism. The side effects are numerous, but the most important is the life-threatening encephalopathy that may develop from the arsenic (highly fatal) or as an inflammatory reaction. Concomitant steroid use is helpful in reducing the risk of encephalopathy and death without compromising the efficacy of melarsoprol. Gradual increase in the first round of treatment between 2 and 3.6 mg/kg in divided doses three times daily during 3 days has also been shown to reduce the risks of encephalopathy. Melarsoprol has been replaced by eflornithine for *T. b. gambiense* infection but remains the only agent available for the treatment of stage II *T. b. rhodesiense* infection.

Developed more than 20 years ago, eflornithine is the newest agent and drug of choice for treatment of stage II disease with *T. b. gambiense*. Unfortunately, it is not effective as monotherapy for *T. b. rhodesiense* because of reduced susceptibility of this organism to the drug. Eflornithine requires frequent intravenous administration, and up to 40% of patients may report some adverse effects. The drug is much more tolerable than its predecessor melarsoprol and is associated with a comparatively reduced mortality rate.

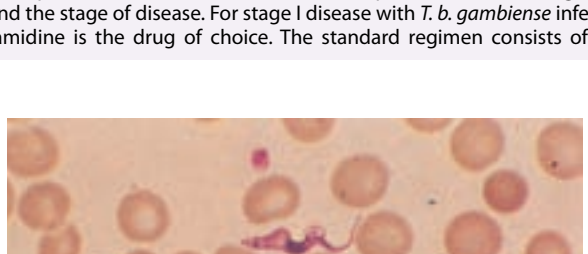

FIGURE 346-2. *Trypanosoma rhodesiense* in peripheral blood. It has a nucleus, posterior kinetoplast, undulating membrane, and flagellum (×1500).

PREVENTION

To date, there is no vaccine against HAT. The mainstays of prevention include active case finding with early treatment and vector control. Given the nature of each infection, active case surveillance is more suitable for infection with *T. b. gambiense*, but vector control is more effective for prevention of infection with *T. b. rhodesiense.* The goal of active case finding is to identify infected individuals who may still be in the asymptomatic stage or early stage. This approach, which is more suited for disease from *T. b. gambiense*, typically consists of a screening examination for lymphadenopathy followed by a CATT test. If results are positive for both, the patient should undergo further evaluation with lymph node aspirate and blood testing. If trypanosomes are found, the patient should be treated. Vector control measures include tsetse fly traps and insecticide-impregnated screens; the traps are easily maintained by locals, but the screens require regular retreatment and thus are more labor intensive and costly to maintain. Mass spraying, once a successful method until the 1960s, is no longer practiced. However, if an epidemic does occur, ground or aerial spraying combined with disruption of the animal reservoir's habitat may be the most effective method of achieving rapid vector control.

Travelers to endemic areas should be aware of this disease and should use basic protective measures. Protective clothing of at least medium weight should be worn; neutral colors are most effective because flies are attracted

TABLE 346-1 DRUGS USED TO TREAT HUMAN AFRICAN TRYPANOSOMIASIS

DRUG	CLASS	STAGE	ROUTE	ADULT DOSE	ADVERSE EFFECTS
TRYPANOSOMA BRUCEI GAMBIENSE					
Pentamidine	Aromatic diamidine	I	IM or IV	4 mg/kg daily for 7 days	Pain, GI symptoms, hypoglycemia or hyperglycemia, electrolyte abnormalities, leukopenia, thrombocytopenia
Eflornithine	Ornithine carboxylase inhibitor	II	IV	100 mg/kg every 6 hr for 14 days	GI symptoms, bone marrow toxicity, seizures
Melarsoprol (alternative option)	Trivalent arsenical	II	IV	2.2 mg/kg daily for 10 days (maximum daily dose, 180 mg)	Encephalopathic syndromes, peripheral neuropathy, paralysis, cardiac dysrhythmias, GI symptoms, rash, pruritus, thrombophlebitis
TRYPANOSOMA BRUCEI RHODESIENSE					
Suramin	Polysulfonated naphthylamine derivative of urea	I	IV	1 g IV on days 1, 3, 7, 14, and 21 (after test dose of 100-200 mg)	Anaphylaxis, nephrotoxicity, fever, rash, pruritus, arthralgias, reversible peripheral neuropathy, and bone marrow toxicity
Melarsoprol	Trivalent arsenical	II	IV	1.2 mg/kg every 8 hr for 3 consecutive days each week for 3 weeks (maximum daily dose, 180 mg)	Encephalopathic syndromes, peripheral neuropathy, paralysis, cardiac dysrhythmias, GI symptoms, rash, pruritus, thrombophlebitis

GI, Gastrointestinal; IM, intramuscular; IV, intravenous.

to bright and dark colors. Tsetse flies are attracted to moving vehicles but rest in the shade or bushes. Although use of insect repellant is prudent for other vector-borne illnesses that may be endemic in these areas, it has not been proven to substantially reduce the risk of tsetse fly bites. There is no recommended chemoprophylaxis for travelers.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Neuberger A, Meltzer E, Leshem E, et al. The changing epidemiology of human African trypanosomiasis among patients from nonendemic countries—1902-2012. *PLoS ONE*. 2014;9:e88647.
2. International Glossina Genome Initiative. Genome sequence of the fly (Glossina morsitans): vector of African trypanosomiasis. *Science*. 2014;344:380-385.
3. Brun R, Blum J, Chappuis F, et al. Human African trypanosomiasis. *Lancet*. 2010;375:148-159.
4. Kennedy PG. Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *Lancet Neurol*. 2013;12:186-194.
5. Lejon V, Bentivoglio M, Franco JR. Human African trypanosomiasis. *Handb Clin Neurol*. 2013;114:169-181.
6. Buscher P, Gilleman Q, Lejon V, et al. Rapid diagnostic test for sleeping sickness. *N Engl J Med*. 2013;368:1069-1070.
7. Eperon G, Balasegaram M, Potet J, et al. Treatment options for second-stage gambiense human African trypanosomiasis. *Expert Rev Anti Infect Ther*. 2014;12:1407-1417.

REVIEW QUESTIONS

1. Risk factors for sleeping sickness is/are

a. tsetse fly bite.
b. residence in urban areas of North Africa.
c. rural areas of sub-Saharan Africa.
d. a and c.
e. a and b.

Answer: D African sleeping sickness is rare in urban areas or in North Africa. This vector-borne parasitic disease is transmitted to humans and animals by the bite of the tsetse fly.

2. Clinical presentation of sleeping sickness is/are

a. chancre at site of bite.
b. fever and headache.
c. lymphadenopathy.
d. a and c.
e. a, b, and c.

Answer: E Sleeping sickness can present acutely with a chancre at the site of the bite, fever, headache, and lymphadenopathy.

3. Diagnostic test(s) of sleeping sickness is/are

a. visualize parasite in chancre, node, or blood.
b. polymerase chain reaction (PCR) on cerebrospinal fluid.
c. serology.
d. a and c.
e. a, b, and c.

Answer: E Diagnosis of African sleeping sickness can be accomplished by visualizing the parasite by PCR and serology.

4. What drug is used to treat patients with stage 1 sleeping sickness caused by *T. b. gambiense*?

a. Pentamidine
b. Melarsoprol
c. Eflornithine
d. a and b
e. a, b, and c

Answer: A Therapies with demonstrated effectiveness include melarsoprol, pentamidine, and eflornithine. For stage I disease, pentamidine is the treatment of choice.

5. East African sleeping sickness is

a. more rapidly progressive than West African.
b. characterized by severe fevers.
c. commonly associated with lymphadenopathy.
d. a and b.
e. a, b, and c.

Answer: D East African sleeping sickness tends to be more severe and rapidly progressive, with intermittent fevers that resemble those seen in malaria. Unlike West African sleeping sickness, in which lymphadenopathy (often generalized) is common, lymphadenopathy is not typically associated with lymphadenopathy in East African sleeping sickness.

347

CHAGAS DISEASE

LOUIS V. KIRCHHOFF

DEFINITION

Chagas disease, or American trypanosomiasis, is caused by the protozoan parasite *Trypanosoma cruzi*. The terms *Chagas disease*, *American trypanosomiasis*, and *T. cruzi infection* are synonyms.

The Pathogen

Several dozen species are included in the genus *Trypanosoma*, but only the African trypanosome *Trypanosoma brucei* (subspecies *T. b. gambiense* [West African] and *T. b. rhodesiense* [East African]) (Chapter 346) and the American trypanosome *T. cruzi* cause disease in humans. Many species of triatomine insects, also called kissing bugs, act as vectors for *T. cruzi*, and many species of wild and domestic mammals, as well as humans, are involved in the complex life cycle of this fascinating organism. The vectors become infected by ingesting blood from mammals that have parasites in their blood stream. The parasites then multiply in the gut of the insects and are ultimately discharged in the feces of the vector. Transmission to a new mammalian host occurs when parasite-laden vector feces contact vulnerable surfaces such as the mucosae of the mouth or nose, the conjunctivae, or breaks in the skin. When in contact with tissues of the new host, the contaminating parasites enter local cells and multiply intracellularly, and as parasitized cells rupture, are released into the lymphatics and blood stream. The circulating organisms enter new cells at distant sites and in this manner maintain an endless process of asynchronous multiplication. The life cycle is completed as parasites are swept up in blood meals taken by vectors. In addition to vector-borne transmission, *T. cruzi* can be transmitted by blood or organs donated by infected persons, from mother to unborn child, by the ingestion of contaminated food or drink,[1] and in laboratory accidents.

EPIDEMIOLOGY

Epizootiology of *T. cruzi*

The triatomine vectors that transmit *T. cruzi* are found in the Americas from southern Argentina through the southern half of the United States. The parasite has been isolated from more than 100 species of domestic and wild mammals, which for the most part probably become infected when they eat infected vectors. Armadillos, wood rats, raccoons, and opossums are typical wild mammalian reservoirs, and these and other species that harbor *T. cruzi* can be found in large numbers in the southern and western parts of the United States.

Typically, humans acquire *T. cruzi* infection when they live in houses in enzootic areas where the sylvatic cycle of transmission is active. The process begins when vector species adaptable to living in human dwellings take up residence in niches in the primitive wood, mud, and stone houses that are typical in many regions of Latin America. These vectors become domiciliary and they then take blood meals, mostly at night, from the humans who occupy the dwellings that they have invaded, as well as from domestic animals that sleep there, particularly dogs. Thus, Chagas disease is primarily a public health problem among poor people who live in rural areas.

Epidemiology of Chagas Disease in Latin America

Chagas disease is a zoonosis that is endemic in Mexico and all countries of Central and South America. None of the Caribbean islands are endemic. In 2006, the Pan American Health Organization estimated that 8 million people are chronically infected with *T. cruzi* and that up to 12,000 deaths result from Chagas disease annually.[2] Since 1991 a major international vector control program in the southern cone countries of South America (Chile, Argentina, Paraguay, Brazil, Bolivia, and Uruguay) has achieved a marked reduction in transmission of *T. cruzi* through housing improvement, education of people at risk for acquiring the infection, and spraying of residual insecticides. Substantial reductions in prevalence rates in school-aged children and in blood donors constitute clear evidence of the success of the program. Uruguay, Chile, and Brazil were certified as being free of vector transmission in 1997, 1999, and 2006, respectively. Marked reduction in transmission has been achieved in Argentina as well. Similar programs have been initiated in Central America and the Andean countries. In parallel with the vector control programs, donor screening has been implemented throughout almost the entire endemic range, and with the notable exception of Mexico, transmission of *T. cruzi* by transfusion has largely been eliminated.

Epidemiology of Chagas Disease in the United States

As noted, the sylvatic cycle of *T. cruzi* exists in much of the southern and western regions of the United States, but only six cases of autochthonous transmission have been reported: three in Texas and one each in Tennessee, Louisiana, and California. Moreover, the nationwide blood donor screening that started in January 2007, in which more than 50 million units have been tested, has uncovered only 17 *T. cruzi*–infected donors, who appear to have acquired the infection autochthonously.[3] In the past 30 years, only about 15 laboratory-acquired and imported cases of acute Chagas disease have been reported to the Centers for Disease Control and Prevention (CDC). Only one of the latter infections occurred in a tourist returning to the United States, but three such instances have been reported in Europe as well as one in Canada. Thus, acute Chagas disease is extremely rare in the United States.

Census data indicate that including illegal and legal immigrants, 23 million persons born in countries endemic for Chagas disease currently live in the United States. Roughly 17 million of these individuals are from Mexico, where the overall prevalence of *T. cruzi* infection is thought to range from 0.5% to 1.0%. Moreover, more than half of the non-Mexican immigrants have come from Central America, a region where the prevalence of *T. cruzi* infection is relatively high. A recent estimate puts the number of *T. cruzi*–infected persons currently living in the United States at 300,000. Several studies done before blood donor screening began in 2007 identified *T. cruzi*–infected persons in the donor pool, and nine instances of transmission by transfusion in the United States and Canada were described. Since screening began in 2007, more than 50 million units have been screened, and nearly 3000 *T. cruzi*–infected donors have been identified and permanently deferred from donation. The confirmed rate of *T. cruzi* infection in donors is about 1 in 13,300.[4] With the goal of reducing the enormous cost of universal screening ($100 to $200 million per year), a selective screening protocol based on previous negative test results has been implemented.

The transplantation in the United States of organs from three persons with undiagnosed chronic *T. cruzi* infection resulted in acute Chagas disease in five recipients, one of whom died of the infection. To date, only one instance of congenital transmission of *T. cruzi* here has been reported. A reasonable estimate of the number of babies born in the United States each year with congenital Chagas disease puts it in the range of 63,000. The fact that most babies with congenital Chagas disease are asymptomatic[5] and the low level of knowledge about Chagas disease among caregivers likely underlie the dearth of reported cases.

PATHOBIOLOGY

In acute Chagas disease, an inflammatory lesion, called a chagoma, may appear at the site of entry of the parasites. Local histologic changes include intracellular parasitism of muscle and other subcutaneous tissues, lymphocytic infiltration, interstitial edema, and hyperplasia of lymph nodes that drain the area. As the parasites spread systemically through the lymphatics and blood stream, muscles, including the myocardium, are the most heavily parasitized tissues, but the organisms can invade essentially any tissue. Myocarditis may develop in association with focal areas of infected cardiomyocytes, inflammation, and necrosis. The characteristic pseudocysts seen in sections of *T. cruzi*–infected tissues are actually host cells filled with

multiplying forms of the parasite (Fig. 347-1). In some patients, parasites can be seen in cerebrospinal fluid (CSF).

In persons with chronic Chagas disease, the heart is the organ most commonly affected. Hearts obtained at autopsy from patients who died of Chagas cardiomyopathy usually have a global appearance reflecting biventricular enlargement and thinning of ventricular walls (Fig. 347-2). Mural thrombi are frequently present, and an apical aneurysm of the left ventricle is typical in patients with advanced disease. At the cellular level, the process that underlies these gross pathologic abnormalities is a chronic inflammation with mononuclear cell infiltration, diffuse interstitial fibrosis, and atrophy of myocardial cells. The chronic inflammation affects the conduction system as well and causes a variety of rhythm disturbances, including atrial bradyarrhythmias and fibrillation; premature ventricular contractions; bundle branch blocks, often of the right bundle; ventricular tachycardia; and third-degree atrioventricular block. Parasites are rarely seen in diseased tissues by conventional histologic methods, but several studies using polymerase chain reaction (PCR) assays found a correlation between the intensity of inflammation and the presence of parasites. Evidence accumulated to date implicates the persistence of parasites and the resulting chronic inflammation in affected tissues—rather than autoimmune mechanisms—as the basis for the pathogenesis in patients with chronic *T. cruzi* infection.

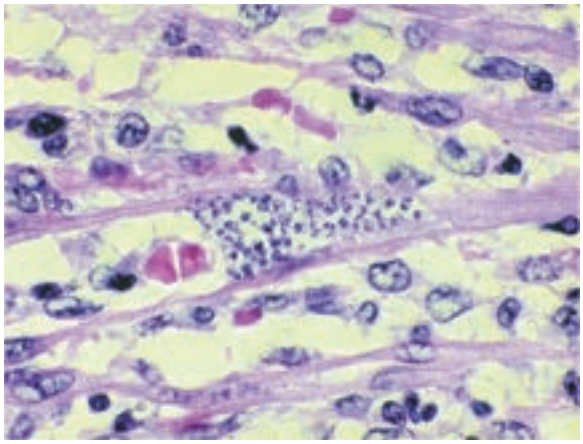

FIGURE 347-1. ***Trypanosoma cruzi*** **in cardiac muscle of a child who died of acute Chagas myocarditis.** An infected myocyte containing several dozen *T. cruzi* amastigotes is in the center of the field (hematoxylin-eosin staining, ×900).

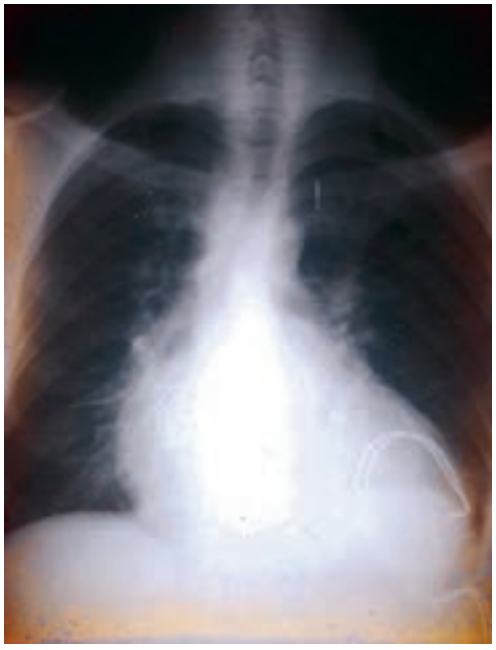

FIGURE 347-2. Chest radiograph of a patient from Bolivia with chronic *Trypanosoma cruzi* infection, rhythm disturbances, and cardiomyopathy. Pacemaker wires can be seen in the area of the left ventricle.

The dilation and hypertrophy observed on gross examination of the esophagus or colon of a patient with chronic Chagas disease of the digestive tract (megadisease) is striking. Focal inflammatory lesions with lymphocytic infiltration are seen on microscopic examination of affected tissues. In addition, the number of neurons in the myenteric plexus is reduced, and periganglion and intraganglion fibrosis with Schwann cell proliferation and lymphocytosis is present. In most patients, the clinical consequences of this parasympathetic denervation are limited to the esophagus or colon (or to both), but the ureters, biliary tree, and other hollow viscera can be affected as well.

CLINICAL MANIFESTATIONS

Acute Chagas Disease

Acute Chagas disease is usually an illness of children but can occur at any age.[6] Symptoms are typically mild and nonspecific. When the parasite has entered through a break in the skin or the site of a vector's puncture, as noted, a chagoma may appear with local lymphadenopathy. The Romaña sign, the classic finding in acute Chagas disease, consists of painless edema of the palpebrae and periocular tissues and may appear when the conjunctiva is the portal of entry. These initial local signs can be followed by fever, malaise, anorexia, and edema of the face and lower extremities. Generalized lymphadenopathy and hepatosplenomegaly may also be present. Severe myocarditis may develop as well, and most deaths are caused by the resulting congestive heart failure. Meningoencephalitis is a rare complication. In untreated patients the acute illness resolves spontaneously over a period of 6 to 8 weeks as the patient enters the indeterminate phase of Chagas disease, which is characterized by subpatent parasitemia, absence of associated signs and symptoms, and easily detectable antibodies to *T. cruzi*.

Chronic Chagas Cardiopathy

In only 10% to 30% of persons chronically infected with *T. cruzi* does clinically manifested disease develop. It most often involves rhythm disturbances or cardiomyopathy.[7,8] Symptoms of cardiac Chagas disease can develop insidiously over years and often decades after the initial infection. Clinical findings reflect the rhythm disturbances, congestive heart failure, and thromboembolism that characterize the illness. Dizziness, syncope, and even seizures can result from a wide variety of arrhythmias. The cardiomyopathy often leads to biventricular failure, and right-sided heart failure can predominate in patients with advanced disease. A validated risk score assessment tool has been developed to gauge prognosis. Chronic Chagas disease is an independent risk factor for stroke.

Chronic Gastrointestinal Chagas Disease (Megadisease)

The esophagus and colon are the segments of the gastrointestinal (GI) tract most commonly affected in persons with chronic *T. cruzi* infection. In patients with megaesophagus, the symptoms are similar to those of idiopathic achalasia (Chapter 138) and may include cough, dysphagia, odynophagia, and regurgitation. Hypersalivation and consequent salivary gland hypertrophy develop in some patients with advanced esophageal dysfunction. Aspiration can occur, especially during sleep, and in untreated patients, repeated episodes of aspiration pneumonitis are common. Weight loss and even cachexia in patients with severe megaesophagus can combine with pneumonitis to cause death. Patients with chagasic megacolon have intermittent abdominal pain and chronic constipation and in advanced cases can go for several weeks between bowel movements. Rarely, acute obstruction, occasionally with volvulus, can lead to perforation, sepsis, and death.

IMMUNOSUPPRESSION AND TRANSPLANTATION IN PATIENTS INFECTED WITH *T. CRUZI*

When persons with chronic carriage of *T. cruzi* become immunosuppressed, reactivation of the infection can occur, sometimes with an intensity that is atypical of acute Chagas disease in immunocompetent persons.[9] The overall incidence of reactivation in immunosuppressed persons who harbor the parasite chronically is not known. Reactivation after renal transplantation has been reported, and in rare instances central nervous system abscesses and skin lesions were involved. The consensus view is that Chagas disease should not be a contraindication to kidney transplantation. In *T. cruzi*–infected patients who do undergo the procedure, periodic monitoring for signs and symptoms of acute Chagas disease should nonetheless be carried out, including careful neurologic evaluation, and parasitologic testing should be performed when acute illnesses occur postoperatively.

Reactivation of *T. cruzi* infection can also occur in persons coinfected with the parasite and human immunodeficiency virus (HIV). Dozens of such

patients have been described. It is striking that in many of these patients *T. cruzi* brain abscesses developed, which do not occur in immunocompetent patients with chronic Chagas disease. It has been shown that HIV viral loads increase in the context of reactivated acute Chagas disease. Calculations based on the overlapping epidemiologies of HIV and *T. cruzi* infections in the endemic countries suggest that the incidence of reactivation of the latter in coinfected persons is low.

DIAGNOSIS

Acute Chagas Disease

The first step in considering the diagnosis of acute Chagas disease is establishing that a person is at risk for *T. cruzi* infection. Risk factors include recent residence or blood transfusion in an endemic area, birth to a mother with geographic- or transfusion-associated risk in the case of a newborn, or a laboratory accident involving the parasite. Definitive diagnosis of acute Chagas disease can be made only by detecting parasites. Serologic assays for *T. cruzi*–specific IgM are not accurate enough to justify their use. In immunocompetent persons suspected of having acute Chagas disease, the most productive approach is examination of wet preparations of anticoagulated blood or buffy coat for the highly motile blood stream parasites. They can be seen in Giemsa-stained smears as well. In infected immunocompromised patients, moreover, parasites can sometimes be found in other specimens such as lymph node aspirates, biopsy specimens of skin lesions, bone marrow, endomyocardial tissue, CSF, and pericardial fluid.

When these direct methods fail to detect organisms in an at-risk person, samples should be tested with a PCR assay (see later).[10] PCR assays have been shown to be more sensitive than the direct methods described earlier for detecting *T. cruzi*. Another method is to culture blood or other samples in specialized liquid medium, but the usefulness of this approach is limited by low sensitivity (50%-70% for hemoculture) and by the fact that cultures take a minimum of 2 weeks before turning positive. In newborns whose blood is negative both by direct examination and in a PCR assay right after birth, serologic evaluation for *T. cruzi*–specific IgG should be performed 6 to 9 months later, by which time maternal antibodies will have disappeared.

Chronic Chagas Disease

Chronic *T. cruzi* infection is usually diagnosed by detecting IgG antibodies that specifically bind to parasite antigens, and in almost all instances isolation of the organism is not necessary. More than 30 serologic assays for diagnosing Chagas disease are currently available commercially in endemic countries, where they are used widely for testing clinical specimens and screening blood donors. Even though these tests usually have good sensitivity and reasonable specificity, false-positive reactions do occur, typically with specimens from people who have other infectious diseases or autoimmune conditions. The World Health Organization has recommended that testing be done with two assays based on different formats. In the United States, the Ortho *T. cruzi* ELISA Test System (Ortho-Clinical Diagnostics, Raritan, NJ) and the Abbott Prism Chagas Assay (Abbott Laboratories, Abbott Park, IL) are used to screen donated blood. The Abbott ESA Chagas and the Chagas RIPA[11] are used for confirmatory testing of donor samples that are positive in the screening tests.

Since the publication in 1989 of two reports describing the use of PCR tests to detect *T. cruzi*, well over 100 articles dealing with this approach have appeared. In human studies, the sensitivity of the PCR assays ranged from 44.7% to 100%, with most being higher than 90%. It is generally accepted that the level of sensitivity of these assays is not high enough to justify their use for confirmatory testing of serologically positive donor samples. In contrast, PCR assays may be useful in persons who have borderline serologic results, in patients suspected of having congenital or acute Chagas disease in whom parasites are not detected microscopically, and in infected patients who have received specific treatment. In all such persons, because of the sensitivity issue, only positive PCR results can be taken as being truly indicative of infection status.

TREATMENT Rx

Antiparasitic Drugs (also see Chapter 344)

The two drugs currently available for treating Chagas disease are unsatisfactory, and the need for a parasitologically curative drug regimen is the most important current challenge in Chagas disease research.[12] Nifurtimox (Lampit, Bayer 2502, Leverkusen, Germany) is a nitrofuran derivative that has been used for more than 3 decades. Nifurtimox reduces symptoms and decreases mortality rates in patients with acute Chagas disease, approximately 70% of whom are cured parasitologically. Nifurtimox also can cure a substantial portion of children in the indeterminate phase, but unfortunately, cure rates may be less than 10% in adults with long-standing chronic *T. cruzi* infection.

Disadvantages of nifurtimox include its long course of treatment and occasionally bothersome side effects, including GI complaints such as anorexia, nausea, vomiting, weight loss, and abdominal pain. Patients taking the drug may also have neurologic symptoms such as insomnia, restlessness, paresthesias, twitching, polyneuritis, and even seizures.[13] For adults, the recommended oral dosage is 8 to 10 mg/kg body weight per day. For adolescents, the dose is 12.5 to 15 mg/kg/day, and for children 1 to 10 years of age, it is 15 to 20 mg/kg/day. The drug should be given each day in four divided doses, and treatment should be continued for 90 to 120 days. In the United States, nifurtimox can only be obtained from the CDC Drug Service (404-639-3670).

The second drug for treating *T. cruzi* infection is the nitroimidazole derivative benznidazole (Rochagan, Radinil, Roche 7-1051; LAFEPE, Pernambuco, Brazil). Cure rates are similar or perhaps a bit higher than those achieved with nifurtimox. A cure rate higher than 90% in babies with congenital infection has been observed with benznidazole. Side effects can include rash, peripheral neuropathy, and granulocytopenia.[14] Benznidazole is considered the drug of choice by most Latin American experts. The recommended oral dosage of benznidazole is 5 to 10 mg/kg body weight per day for children and 5 mg/kg body weight per day for adults, in both cases for 60 days.

There is broad agreement among experts that treatment is indicated in all patients with acute or congenital infections, as well as in chronically infected children up to 18 years old. This recommendation is supported by several studies showing that a majority of such patients can be cured parasitologically. By extension, it would be reasonable to treat anyone 18 years or older known to have acquired *T. cruzi* infection within the past 17 years. There is also broad agreement that persons with advanced symptomatic *T. cruzi* infection should not be given specific treatment. The remaining question, then, is whether adults with long-standing indeterminate-phase infections, who by far constitute the largest group of *T. cruzi*–infected persons, should be treated. This is a thorny question because the burden of taking a full course of either drug can be substantial and because parasitologic cure rates are so low. Importantly, there is no clear evidence from randomized controlled trials that therapy with either drug delays the onset of symptoms or the progression of pathology or reduces mortality rates in adults with chronic *T. cruzi* infection. In 2006, a panel of experts convened at the CDC recommended that adults 19 to 50 years old in the indeterminate phase or with mild symptoms be offered treatment. A large randomized trial (the BENEFIT multicenter trial) designed to assess the clinical and parasitologic efficacy of benznidazole in patients with Chagas disease 18 to 75 years old without advanced disease is being conducted in Colombia, Bolivia, Brazil, and Argentina. Initial findings will be available in 2015. A recently reported randomized trial comparing benznidazole (150 mg twice daily) with posaconazole, which had shown trypanocidal activity in mouse models, given at two doses (400 mg twice daily or 100 mg twice daily) all for 60 days for chronic Chagas disease showed that benznidazole had a significantly smaller percentage of treatment failures than did either treatment dose of posaconazole.■ A recently reported study from Argentina showed that treatment with either benznidazole or nifurtimox before pregnancy can reduce the probability of subsequent congenital transmission of the parasite.[15]

Management of Symptomatic Chagas Disease

An algorithm for the evaluation of persons with newly diagnosed Chagas disease has been developed (Fig. 347-3). *T. cruzi*–infected patients in whom symptomatic cardiac or GI disease develops should be referred to appropriate subspecialists. Beyond the possible use of nifurtimox or benznidazole, treatment of acute and chronic Chagas disease is symptomatic. In patients with symptomatic chronic Chagas cardiac disease, treatment should be directed at managing symptoms with the anticoagulants and cardiotropic drugs used in patients with cardiomyopathy of other causes. Pacemakers are useful in patients with ominous arrhythmias. The usefulness of implantable cardioverter-defibrillators in patients with Chagas heart disease has not been established and needs further investigation in prospective randomized trials.

Heart transplantation (Chapter 82) is an option in patients with end-stage Chagas cardiac disease, and more than 150 such patients have undergone the procedure in Brazil and the United States.[16] As is the case with other *T. cruzi*–infected patients who are immunosuppressed, reactivation is a risk but it is manageable. The usefulness and side effects of long-term prophylaxis for reactivation with either benznidazole or nifurtimox in *T. cruzi*–infected patients after heart transplantation have not been evaluated. The long-term survival of Chagas patients with heart transplants appears to be longer than that in patients undergoing cardiac transplantation for other reasons, probably because the lesions of *T. cruzi*–associated pathology affect mostly the heart.

Chagas megaesophagus should be treated as normally done for idiopathic achalasia (Chapter 138), which usually responds to balloon dilation of the lower esophageal sphincter when symptoms are mild. Surgical treatment may

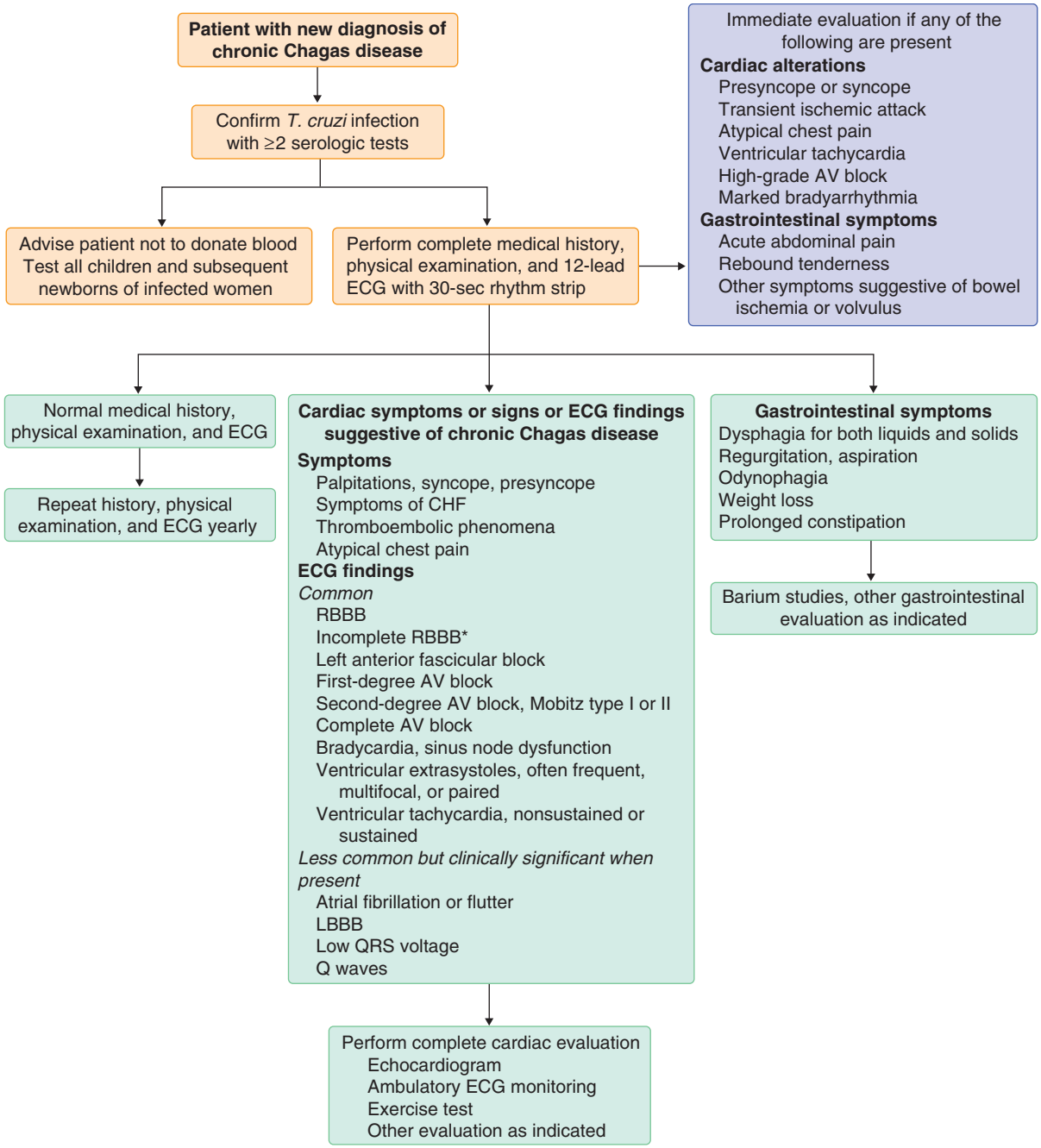

*QRS interval of 0.10 to 0.11 seconds in adults. Criteria based on the *Minnesota Code Manual of Electrocardiographic Findings* with modifications from Maguire et al. Different criteria may be required for ECGs in children.

FIGURE 347-3. **Algorithm for baseline evaluation of a patient with newly diagnosed chronic *Trypanosoma cruzi* infection. *AV,* Atrioventricular; *CHF,* congestive heart failure; *ECG,* electrocardiogram; *LBBB,* left bundle branch block; *RBBB,* right bundle branch block. (From Bern C, Montgomery SP, Herwaldt BL, et al. Evaluation and treatment of Chagas disease in the United States: a systematic review. *JAMA*. 2007;298:2171-2181.)**

be required in patients who do not respond to repeated attempts at balloon dilation. Laparoscopic myotomy is being used with increasing frequency to treat Chagas megaesophagus, as is the case with achalasia.

Chagas megacolon in its early stage can be treated with a high-fiber diet and occasional laxatives or enemas. Fecal impaction requiring manual disimpaction can occur, and toxic megacolon requires surgery. In patients with advanced megacolon, volvulus (Chapter 142) can develop when an enlarged and lengthened sigmoid colon twists and folds on itself; volvulus causes a constellation of symptoms and in many cases requires immediate surgery. Even if the symptoms associated with volvulus are resolved without operative intervention, however, surgical treatment is usually ultimately necessary because the volvulus tends to recur. Several surgical procedures are used to treat advanced Chagas megacolon, all of which include resection of the sigmoid and removal of part of the rectum.

PREVENTION

Reducing human contact with triatomine vectors through education of at-risk persons, housing improvement, and spraying of residual insecticides in endemic countries has resulted in reduction or elimination of vector transmission of *T. cruzi* in a major part of the endemic range, and progress in this regard is expected to continue. Serologic screening of donated blood has essentially eliminated transfusion-related transmission of the parasite in most endemic areas. Outbreaks of acute Chagas disease through oral transmission can be avoided by the implementation of better food safety standards. Drug treatment of *T. cruzi*–infected women before pregnancy reduces the likelihood of congenital transmission. No protocols have been defined and validated for preventing reactivation of *T. cruzi* infection in chronically infected persons who are immunosuppressed iatrogenically or by HIV. A treatment regimen that reliably results in parasitologic cure is

needed to prevent the onset or progression of chronic symptomatic Chagas disease.

PROGNOSIS

The prognosis for patients with acute Chagas disease is generally excellent because most acutely infected persons have only mild symptoms that resolve spontaneously, even without specific treatment. The occasional patient who has symptomatic acute Chagas myocarditis should generally do well if treated early. In persons with chronic *T. cruzi* infection, the lifetime risk for the development of related cardiac or GI dysfunction is only 10% to 30%.[17] A validated risk score assessment tool has been developed to gauge prognosis.[18] It has not been determined in randomized controlled trials whether this rate of progression to clinical disease or overall mortality is significantly affected by treatment with nifurtimox or benznidazole.

Visit *expertconsult.com* for the e-expanded chapter.

Grade A Reference

A1. Molina I, Gomez I, Prat J, et al. Randomized trial of posaconazole and benznidazole for chronic Chagas' disease. *N Eng J Med*. 2014;370:1899-1908.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Alarcon de Noya B, Diaz-Bello Z, Colmenares C, et al. Large urban outbreak of orally acquired acute Chagas disease at a school in Caracas, Venezuela. *J Infect Dis*. 2010;201:1308-1315.
2. Schmunis G. Status of and cost of Chagas disease worldwide. *Lancet Infect Dis*. 2013;13: 283-284.
3. Cantey PT, Stramer SL, Townsend RL, et al. The United States Trypanosoma cruzi Infection Study: evidence for vector-borne transmission of the parasite that causes Chagas disease among United States blood donors. *Transfusion*. 2012;52:1922-1930.
4. Custer B, Agapova M, Bruhn R, et al. Epidemiologic and laboratory findings from 3 years of testing United States blood donors for Trypanosoma cruzi. *Transfusion*. 2012;52:1901-1911.
5. Gebrekristos HT, Buekens P. Mother-to-child transmission of *Trypanosoma cruzi*. *J Pediatric Infect Dis Soc*. 2014;3:S36-S40.
6. Andrade DV, Gollob KJ, Dutra WO. Acute Chagas disease: new global challenges for an old neglected disease. *PLoS Negl Trop Dis*. 2014;8:e3010.
7. Rassi A Jr, Rassi A, Marin-Neto JA. Chagas disease. *Lancet*. 2010;375:1388-1402.
8. Nunes MCP, Dones W, Morillo CA, et al. Chagas disease: an overview of clinical and epidemiological aspects. *J Am Coll Cardiol*. 2013;62:767-776.
9. Lattes R, Lasala MB. Chagas disease in the immunosuppressed patient. *Clin Microbiol Infect*. 2014;20:300-309.
10. Schijman AG, Bisio M, Orellana L, et al. International study to evaluate PCR methods for detection of Trypanosoma cruzi DNA in blood samples from Chagas disease patients. *PLoS Negl Trop Dis*. 2011;5:e931.
11. Otani MM, Vinelli E, Kirchhoff LV, et al. WHO comparative evaluation of serologic assays for Chagas disease. *Transfusion*. 2009;49:1076-1082.
12. Bern C. Antitrypanosomal therapy for chronic Chagas' disease. *N Engl J Med*. 2011;364: 2527-2534.
13. Jackson Y, Alirol E, Getaz L, et al. Tolerance and safety of nifurtimox in patients with chronic Chagas disease. *Clin Infect Dis*. 2010;51:e69-e75.
14. Altcheh J, Moscatelli G, Moroni S, et al. Adverse events after the use of benznidazole in infants and children with Chagas disease. *Pediatrics*. 2011;127:e212-e218.
15. Fabbro DL, Danesi E, Olivera V, et al. Trypanocide treatment of women infected with *Trypanosoma cruzi* and its effect on preventing congenital Chagas. *PLoS Negl Trop Dis*. 2014;8:e3312.
16. Kransdorf EP, Czer LS, Luthringer DJ, et al. Heart transplantation for Chagas cardiomyopathy in the United States. *Am J Transplant*. 2013;13:3262-3268.
17. Sabino EC, Ribeiro AL, Salemi VM, et al. Ten-year incidence of Chagas cardiomyopathy among asymptomatic *Trypanosoma cruzi*-seropositive former blood donors. *Circulation*. 2013;127: 1105-1115.
18. Rassi A Jr, Rassi A, Little WC, et al. Development and validation of a risk score for predicting death in Chagas' heart disease. *N Engl J Med*. 2006;355:799-808.

REVIEW QUESTIONS

1. Which of the following measures have been important over the past few decades in reducing the transmission of *Trypanosoma cruzi* (Chagas disease) in endemic countries?

A. Improvement of housing conditions
B. Identification and treatment of babies with congenital Chagas disease
C. Spraying of insecticides in at-risk houses to eliminate insect vectors
D. All of the above
E. A and C

Answer: E Improvement of housing conditions and insecticide spraying, along with education of at-risk populations, all of which reduce contact with infected vectors, have been at the core of the successful programs that reduce transmission of *T. cruzi*. Programs focused on identifying and treating babies with congenital *T. cruzi*-infection have not been implemented widely and in any event would have relatively little effect on reducing transmission of the parasite.

2. In which of the following patients is chronic Chagas disease (*Trypanosoma cruzi* infection) likely to be the cause of the patient's signs and symptoms?

A. A 56-year-old man from rural Brazil with early-onset dementia
B. A 58-year-old Bolivian woman with abdominal pain caused by a small bowel obstruction
C. A 62-year-old man with weight loss, hemoptysis, and anemia
D. A 20-year-old Bolivian man with recurrent syncope and third-degree heart block and an ectopic ventricular pacer on ECG
E. A 7-year-old child with intermittent diarrhea and eosinophilia

Answer: D In the endemic countries, particularly in Bolivia, where the overall prevalence is about 7%, chronic Chagas disease is a major cause of serious rhythm disturbances, heart block and sudden death among young people. The other options are unlikely because a link between Chagas disease and dementia has not been established; chronic Chagas disease is sometimes manifest as obstructive megacolon but not small bowel obstruction; weight loss, hemoptysis, and anemia are more suggestive of tuberculosis than Chagas disease; and intermittent diarrhea and eosinophilia are not associated with chronic Chagas disease in patients of any age.

3. Which of the following Latin American countries has the lowest prevalence of Chagas disease?

A. Bolivia
B. Dominican Republic
C. Colombia
D. Argentina
E. Ecuador

Answer: B Chagas disease is endemic in all the countries of Central and South America, as well as in Mexico. None of the Caribbean islands are endemic for Chagas disease. The only persons with Chagas disease in the Dominican Republic would be immigrants from the endemic countries, and thus the prevalence there would be quite low and far less than in the other countries listed where vector-borne transmission continues.

4. Which of the following statements is false?

A. Of all the countries in which Chagas disease is endemic, Bolivia has the highest prevalence.
B. Several hundred cases of transfusion transmission of Chagas disease were reported in the United States and Canada before the implementation of donor screening in 2007.
C. Transfusion transmission of *Trypanosoma cruzi* (Chagas disease) has been largely eliminated in the endemic countries by serologic screening of blood donors.
D. Highly accurate assays for diagnosing Chagas disease are available in the United States, Canada, and the endemic countries.
E. *Trypanosoma cruzi* is widely distributed among wild mammals and insect vectors in the southern and southwestern United States.

Answer: B Before the implementation of blood donor screening for Chagas disease in the United States and Canada in 2007, only nine instances of transfusion transmission of the infection had been reported. The other statements are all true: Bolivia's Chagas disease prevalence of about 7% is the highest of all the endemic countries; To the credit of politicians and public health authorities in the endemic countries, and because of the availability of accurate serologic assays, transfusion transmission of Chagas disease is largely a thing of the past. Mexico is a notable exception to this achievement, however, because universal donor testing for Chagas disease has not been implemented there yet. To the surprise of many, in the southern third of the United States *T. cruzi* is widely enzootic in many species of wild mammals and insect vectors, as well as in domestic dogs.

5. The current consensus of experts in Chagas disease holds that the following groups of persons with *T. cruzi* infection should be given specific treatment (benznidazole or nifurtimox) except

A. babies with congenital Chagas disease.
B. patients older than 50 years of age with advanced symptomatic Chagas heart disease.
C. children and adolescents up to age 18 years.
D. patients with AIDS and reactivated acute *T. cruzi* infection.
E. adults with acute Chagas disease and mild symptoms.

Answer: B A panel of experts in Chagas disease from the endemic countries as well as the United States and Canada, convened by the Centers for Disease Control and Prevention in 2006 with the goal of developing evidence-based guidance on diagnosis and treatment, concluded that all groups of patients listed above should be treated, with the exception of patients older than 50 years of age or those with advanced symptomatic Chagas cardiac disease (Bern C, Montgomery SP, Herwaldt BL, et al. Evaluation and treatment of Chagas disease in the United States: a systematic review. *JAMA*. 2007;298: 2171-2181).

348

LEISHMANIASIS

SIMON L. CROFT AND PIERRE A. BUFFET

DEFINITION

Leishmaniasis is caused by protozoan parasites of the genus *Leishmania* that are generally transmitted between mammalian hosts by female phlebotomine sandflies. The disease in humans has a number of clinical manifestations ranging in severity from self-curing, limited cutaneous lesions to disseminated, potentially fatal visceral disease. The parasite exists in the sandfly gut as an extracellular flagellated form, the promastigote, and as an intracellular form, the amastigote, that survives and multiplies in a phagolysosomal compartment of macrophages in the mammalian host.

This disease complex is caused by 17 species of *Leishmania*, which are widely distributed in Europe, Asia, Africa, and South and Central America, with limited foci in Southeast Asia.[1] The characteristics of the main *Leishmania* species are summarized in Table 348-1. There are an estimated 1.5 to 2.0 million new cases each year, with up to 70,000 deaths, although this is probably an underestimate as leishmaniasis is not a reportable disease in many of the 101 countries and territories in which it is known to occur. Many *Leishmania* infections are either asymptomatic or misdiagnosed.

Because leishmaniasis is a disease complex, clinical aspects are considered under separate sections for visceral leishmaniasis (VL) and cutaneous leishmaniasis (CL).

EPIDEMIOLOGY

Infection is established in the mammalian host following a bite of the female sandfly belonging to either *Phlebotomus* spp in Europe, Asia, and Africa or *Lutzomyia* spp in the Americas. Different species of sandfly are associated with transmission of different *Leishmania* species. Most species that cause CL have a zoonotic (acquired from another mammal) transmission cycle, with the exception of *Leishmania tropica*, which is sometimes anthroponotic (transmitted between human beings). VL is either normally anthroponotic (in the case of *Leishmania donovani*) or zoonotic (in the case of *Leishmania infantum*). The predominant mammalian hosts (the reservoirs) are associated with different *Leishmania* species in diverse ecosystems (Fig. 348-1).

VL is caused by either *L. donovani* or *L. infantum* (which is identical to *Leishmania chagasi* in South America). These species have different geographic distributions, with the highest incidence found in the poorest communities in six countries (Bangladesh, Nepal, India, Sudan, Ethiopia, Brazil), and infection is potentially fatal if it is untreated. An estimated 1 in 5 to 1 in 50 infections are asymptomatic, depending on the parasite species and host immunity. Since 2005, there has been a regional program to eliminate VL (aiming to reduce annual incidence to <1 per 10,000) in the Indian subcontinent.[2]

CL, which undergoes self-cure in more than 90% of patients within 3 to 18 months, is widely distributed, but its prevalence is difficult to estimate because of underreporting. Prevalence is associated with age, possibly related to the acquisition of immunity and risk factors, including the presence of domestic animals, rodents, or other mammalian hosts.[3] Ecologic conditions for sandflies, including shaded and humid habitats in crevices and mammal burrows, have been identified. The disease also must be considered in travelers returning from endemic areas.[4]

Urbanization, deforestation, and migration have resulted in changing patterns of disease, with transmission occurring in peridomestic cycles.[5] Other forms of transmission, such as through intravenous needles shared by drug users, have been reported in Spain.

PATHOBIOLOGY

The infection is initially established in the skin after the inoculation of infective metacyclic promastigotes by the sandfly. These infective forms have a glycoprotein coat (a lipophosphoglycan) that enables them to resist complement and to attach to and invade host cells. Peptides in sandfly saliva (e.g., maxadilan) cause vasodilation and erythema and help establish infection in the dermal layer of the skin. Early responses to infection involve neutrophil infiltration and invasion of resident macrophages. Progress of the disease depends on the parasite species and host responses. For both VL and CL, disease progression depends on the maintenance of a parasite-specific immunosuppressive state. Host cell macrophages are in a deactivated state, but cure follows when macrophages become activated and kill the parasites, which are sensitive to nitric oxide and oxygen radicals, in the phagolysosomal compartment. Resolution of disease, after the activation of macrophages, is mediated by a T-helper cell T_H1 response after interaction of antigen-presenting cells (e.g., dendritic cells) with $CD4^+$ and $CD8^+$ T cells and subsequent secretion of pro-inflammatory cytokines (e.g., interleukin 2, interferon-γ, tumor necrosis factor-α). However, in clinical forms such as active VL or diffuse CL, a T_H2 cell response predominates whereby downregulation of macrophage activity follows the production of cytokines such as interleukins 4, 10, and 13 and transforming growth factor-β. This profile has been defined in experimental models, mainly inbred mice, and ongoing clinical studies support the notion of a similar profile in human infection.

In patients with VL, the absence of a T cell–specific immune response to leishmanial antigens is associated with uncontrolled progression of infection. This is linked to elevated levels of interleukin 10 and decreased production of interferon-γ. Genetic susceptibility to *L. donovani* in Sudan has been associated with a solute carrier family (formerly NRAMP1) that regulates macrophage activation and with a polymorphism in the interleukin 4 gene. In localized simple CL, patients show a T_H1-type response and a delayed-type hypersensitivity (DTH) response. DTH is frequently measured by a Montenegro skin test, which can also be used in epidemiologic prevalence studies. Chronic infections show a T_H2-type response, most predominantly in patients with diffuse CL, in whom there is complete anergy to leishmania antigen and no DTH response. Patients with mucosal leishmaniasis (ML) have both T_H1 and T_H2 cell responses and a strong DTH response. Post–kala-azar dermal leishmaniasis (PKDL), a rare sequela to cure from VL, is poorly understood. The roles of $CD4^+$ and $CD25^+$ T cells appear to be different in Indian and Sudanese forms of PKDL.[6]

VISCERAL LEISHMANIASIS

CLINICAL MANIFESTATIONS

The onset of VL, often referred to as kala-azar when it is caused by *L. donovani*, occurs weeks to months after the initial infection. Clinical signs and symptoms do not distinguish VL from hyperreactive malarial splenomegaly or other infectious or hematologic conditions. Anemia, leukopenia, thrombocytopenia, systemic inflammation, or polyclonal hypergammaglobulinemia, either isolated or combined, suggests but does not confirm the diagnosis. Parasitologic tests are therefore indispensable before a therapeutic decision is made.[7]

TABLE 348-1 CHARACTERISTICS OF THE MAIN *LEISHMANIA* SPECIES

LEISHMANIA SPP	*LEISHMANIA* SUBGENUS	DISTRIBUTION: OLD WORLD	DISTRIBUTION: NEW WORLD	PRIMARY FORM	SECONDARY FORMS	ANTHROPONOTIC: AREAS OF TRANSMISSION	ZOONOTIC: RESERVOIR	ALTERNATIVE NAME
L. donovani	*Leishmania*	Indian subcontinent E. Africa		VL	PKDL CL, ML OIVL	Indian subcontinent E. Africa		Kala-azar
L. infantum (*L. chagasi*)	*Leishmania*	Europe Asia	S. & C. America	VL	CL, ML OIVL		Canid	
L. major	*Leishmania*	Asia N. & E. Africa Europe		CL			Rodent	
L. tropica	*Leishmania*	Asia Europe		CL	Recidivans	Syria Afghanistan	Rodent	Aleppo boil
L. aethiopica	*Leishmania*	Ethiopia		CL	DCL		Hyrax	
L. mexicana	*Leishmania*		C. America	CL			Rodent	Chiclero's ulcer
L. amazonensis	*Leishmania*	C. & S. America		CL	DCL		Rodent	
L. braziliensis	*Viannia*	S. America		CL ML	DissCL Lymph		Rodent, marsupial	ML-espundia
L. panamensis	*Viannia*	C. & S. America		CL	ML Lymph		Edentate rodent	Ulcera de bejuco
L. guyanensis	*Viannia*	S. America		CL	ML Lymph		Rodent, edentates	Pian bois
L. peruviana	*Viannia*	S. America		CL			Canids	Uta
L. siamensis		Southeast Asia		CL VL				

CL = cutaneous leishmaniasis; DCL = diffuse cutaneous leishmaniasis; DissCL = disseminated cutaneous leishmaniasis; Lymph = nodular lymphangitis; ML = mucosal leishmaniasis; OIVL = opportunistic infection with VL in HIV-infected patients; PKDL = post–kala-azar dermal leishmaniasis; VL = visceral leishmaniasis.

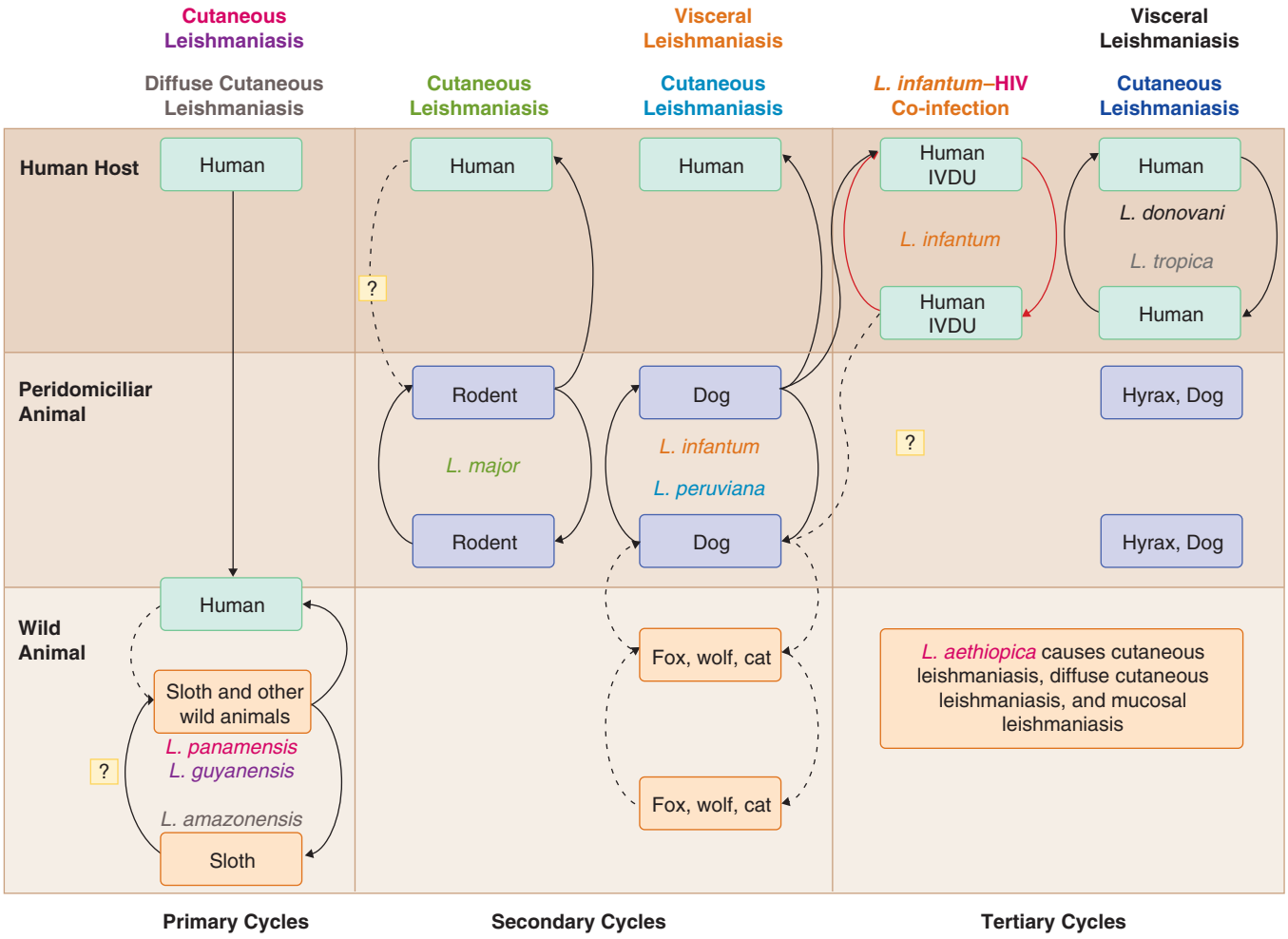

FIGURE 348-1. Old World and New World zoonotic and anthroponotic life cycles of the main species of *Leishmania*. Leishmaniasis is often referred to as a disease complex because different forms of disease can be caused by the same parasite species and similar forms of disease can be caused by different parasite species. IVDU = intravenous drug use.

DIAGNOSIS

Parasitology

Microscopic visualization of amastigotes in samples from the lymph nodes, bone marrow, liver, spleen, or other organs was usually the first step. Because spleen aspiration causes life-threatening complications in approximately 0.1% of patients, it should be performed only in trained facilities and only if other, lower-risk methods cannot be used. It is still used in the field because of the higher cost, logistic constraints, and lower sensitivity of bone marrow aspiration. Polymerase chain reaction (PCR) is more sensitive than microscopic examination and has become the first-line test in several referral hospitals and research centers.[8] Quantitative PCR with validated thresholds allows accurate diagnosis with venous blood samples, thereby avoiding bone marrow aspiration.

Serology

Serologic tests based on indirect fluorescent antibody, enzyme-linked immunosorbent assay, or Western blot display high performance but require equipment that is poorly adapted to field settings. The field-friendly direct agglutination test and immunochromatography (dipstick) with the rK39 antigen have high diagnostic accuracy and can be used in peripheral health centers.[9] Whatever the serologic test used, specific antibodies remain detectable for several years after cure or asymptomatic infection.

Antigen Detection Tests

A latex agglutination test that detects a heat-stable, low-molecular-weight carbohydrate antigen in the urine of patients with VL showed good specificity but low to moderate sensitivity in East Africa and the Indian subcontinent. Further optimization is required.

Complex Manifestations of Visceral Leishmaniasis

VISCERAL LEISHMANIASIS–HUMAN IMMUNODEFICIENCY VIRUS COINFECTION

Although the clinical manifestations in VL patients infected with human immunodeficiency virus (HIV) without severe immunosuppression are similar to those in immunocompetent patients, atypical clinical features can be found in patients with low $CD4^+$ T-cell counts (<200/μL). In the latter group, physicians may order investigations for VL even in the absence of classic signs (e.g., lack of splenomegaly). A substantial proportion of HIV-VL–coinfected patients may have other opportunistic infections that complicate the clinical diagnosis. The parasitic load is usually higher and parasites may be found in tissues other than spleen, liver, bone marrow, or lymph nodes (e.g., in the gut or lung), especially in severely immunosuppressed patients. Therefore, the sensitivity of microscopic examination, culture, or PCR of blood (plain blood or buffy coat) or bone marrow aspirates is generally higher than that in immunocompetent VL patients. Limited data have also shown high sensitivity of the latex agglutination test in the urine of HIV-VL–coinfected patients. In contrast, the sensitivity of serologic tests is decreased in coinfected patients, although study results are equivocal and depend on several factors, such as the test's format, region of endemicity, and level of immunosuppression. For example, the direct agglutination test has shown high sensitivity in Ethiopia. Increased sensitivity can be achieved by using a sequential combination of different serologic tests.

POST–KALA-AZAR DERMAL LEISHMANIASIS (PKDL)

After successful treatment of visceral disease due to *L. donovani*, a proportion of patients progress to disseminated cutaneous disease. In a small proportion of patients with PKDL, uveitis occurs with poor prognosis for preservation of eyesight. This is reported in up to 20% of patients in Sudan and 0.5% in India and Bangladesh. Studies in India have shown that smears are more likely to show amastigotes if specimens are taken from nodular lesions rather than from papular or macular lesions. Serologic tests such as the direct agglutination test, enzyme-linked immunosorbent assay, and rK39 immunochromatography are of limited value because a positive finding may be the result of persistent antibodies after the past episode of VL. Nevertheless, serology can be helpful when a previous history of VL is uncertain.

TREATMENT Rx

General Principles

The therapeutic management of patients with VL requires renutrition; broad-spectrum antibiotics if bacterial superinfection is suspected; transfusion in case of severe anemia; and proper hydration, especially when amphotericin B is used as the specific antileishmanial treatment.

The clinical response to antileishmanial agents depends on the clinical form and the infecting species or even the subspecies (zymodeme). Many antileishmanial agents are toxic, expensive, or difficult to administer in field conditions. No single satisfactory option for treating most of the clinical forms or species has been validated. Although therapeutic decisions and follow-up have become relatively simple for the treatment of VL in nonendemic and some endemic countries based on the powerful, well-tolerated agent liposomal amphotericin B,[10] more complexity persists for the treatment of VL in East Africa and in endemic countries where liposomal amphotericin B is not available.[11] The treatment decision for CL or ML often requires expert advice.[12]

Visceral Leishmaniasis in Immunocompetent Patients

Single-Agent Therapies

Pentamidine is efficient in treating VL only when high doses (more than seven injections of 4 mg/kg) are used. These are toxic, and pentamidine is no longer used for the treatment of VL. Lower doses (fewer than four injections of 4 mg/kg) induce much less severe adverse events and are still used for the secondary prophylaxis of VL in HIV-infected patients (fortnightly to monthly injections) and in those with CL caused by *Leishmania guyanensis* or *Leishmania panamensis* (one to three injections).

Pentavalent antimonials (sodium stibogluconate, meglumine antimoniate) are still prescribed as first-intention drugs in many areas. In the Indian subcontinent (mainly in northern Bihar), *L. donovani* is resistant to pentavalent antimonials.[13] Their efficacy has decreased from 90% to less than 40% during the past 40 years. In other VL foci, failure rates of initial treatment do not exceed 10% as long as the dosage is respected (20 mg of pentavalent antimony per kilogram per day for 28 days). In an endemic country, mortality in treated patients may exceed 10%, with toxicity positively correlating with age. Pentavalent antimonials are contraindicated in patients with heart, kidney, or liver disease or advanced age and in pregnant women. Generic formulations have generally but not constantly demonstrated activity and tolerance identical to those of sodium stibogluconate.

Miltefosine is an alkyl phosphocholine. The oral form (2.5 mg/kg/day for 28 days) is effective for VL in India. In Ethiopian patients with VL (28% with HIV infection), there was significantly higher mortality with sodium stibogluconate (generic) than with miltefosine (9.7% vs. 2.1%) despite the lower parasitologic efficacy of miltefosine (92.1% vs. 99.3%). Miltefosine is contraindicated in pregnant women or those who may become pregnant. Because of persistent levels of the drug, contraceptive measures must be observed for 3 months after therapy. Compared with a decade ago, substantial increase in the failure rate of oral miltefosine has been noted in the treatment of visceral leishmaniases in India and Nepal since 2012.

Paromomycin (aminosidine sulfate) (15 mg/kg in sulfate form equivalent to 11 mg of base per kilogram intramuscularly for 21 days) is highly effective in India, where it was registered in 2006. In East Africa, its efficacy is significantly lower. Similar to the situation with miltefosine, the longevity of the product would probably be better preserved in combination than as a single-agent therapy.

Amphotericin B (cumulative dose of 7 to 15 mg/kg) cures more than 98% of patients in India, whereas treatment of visceral infection with *L. infantum/chagasi* requires at least 14 mg/kg.

Liposomal amphotericin B achieves very high cure rates in India with lower cumulative doses than are needed with the deoxycholate formulation. In preliminary studies in East Africa, the doses of liposomal amphotericin B required to cure VL (about 30 mg/kg cumulatively) are markedly higher than those used in India. Liposomal amphotericin B is associated with less infusion-related fever and chills, less renal toxicity, and a reduction in the number of infusions and length of hospitalization. The better renal tolerance of liposomal amphotericin B is especially beneficial in patients with renal failure or a kidney graft and in those with an increase in serum creatinine concentration during amphotericin B deoxycholate therapy. Liposomal amphotericin B is the antileishmanial agent with the best benefit-risk ratio. It is now the first-line option in most nonendemic countries, both in children and in adults. The very high cost of liposomal amphotericin B has been reduced for use in endemic countries, but wider implementation would be greatly facilitated by donation. A single infusion of liposomal amphotericin B at a dose of 10 mg/kg of body weight was reported in 2010 not to be inferior to and less expensive than conventional treatment with amphotericin B deoxycholate. In this study, conducted in India, 304 of 304 patients (100%) with VL in the liposomal therapy group and 106 of 108 (98%) in the conventional therapy group had apparent cure responses at day 30. Cure rates at 6 months were similar: 95.7% with liposomal therapy and 96.3% with conventional therapy.

Combinations and Coadministration

To limit extension of the resistance to pentavalent antimonials and to prevent the emergence of resistance to paromomycin or miltefosine, shorter courses of therapy are being investigated. Because only miltefosine may be administered orally, this approach is based at least partly on products administered parenterally. As high cure rates have been reported, the combination of antimony and paromomycin is approved for use in East Africa but still

requires many injections. Combinations of amphotericin with oral miltefosine (sequential) and of intramuscular paromomycin with miltefosine can achieve cure rates of up to 98% with 7- to 10-day courses of therapy. A single infusion of liposomal amphotericin B followed by oral miltefosine has been highly effective in India, but confirmation studies are required here and elsewhere.

Visceral Leishmaniasis in Immunodeficient Patients

Leishmania infantum–HIV Coinfection

As with other major opportunistic infections during HIV infection, treatment may be subdivided into initial course and secondary prophylaxis. With initial treatment, the efficacy of meglumine antimoniate and that of amphotericin B are similar. The severe adverse effects of antimony derivatives are more common in this context. Doses of liposomal amphotericin B administered to immunodeficient patients are higher (30 to 40 mg/kg in cumulative dose) than those administered to immunocompetent patients. When therapeutic immunosuppression may not be reduced or optimization of highly active antiretroviral therapy is not possible, secondary prophylaxis is often proposed. Amphotericin B lipid complex moderately reduces the frequency of recurrences. Discontinuous administration of liposomal amphotericin B is generally used, but a progressive reduction in its efficacy has been reported, although apparently not associated with decreased parasitic sensitivity in vitro. Miltefosine is another option, especially if the reduction in parasitic load is backed up by quantitative PCR with a validated threshold (low residual load probably being associated with a lower risk for resistance). Administration of pentamidine once or twice a month is another potentially interesting option because drug levels persist for weeks or months after a single administration. Pancreatic tolerance should be monitored closely.

Leishmania donovani–HIV Coinfection

Patients should benefit from effective antiretroviral treatment. The experience with *L. infantum*–HIV coinfection is in part transposable. The potential risk for the emergence of resistance to antileishmanial agents is still more significant in this context because *L. donovani* (but not *L. infantum*) can be transmitted from human to human. Important clinical research is ongoing in East Africa.

CUTANEOUS LEISHMANIASIS

CLINICAL MANIFESTATIONS

Although the signs and symptoms of CL vary considerably (Fig. 348-2)—from pure nodular lesions to developing ulceration through dry crusty lesions and squamous plaques—there are some fairly constant features. First, firm infiltration is almost constant (the exception being the initial macule of PKDL). Second, the evolution is subacute, which is a useful criterion in practice. A lesion reaching its maximum size in less than a week is most likely not due to CL. Finally, except in patients with numerous satellite papulopustules, the lesion or lesions are sharply defined. Colonization of the CL ulceration with bacteria may give the lesion a purulent appearance, whereas patent superinfection adds an erythematous ring distinctly overflowing the infiltrated edge of the ulceration and making a usually cold and painless lesion feel hot and painful. Several dermatologic conditions, such as staphylococcal or streptococcal infection, mycobacterial ulcer, leprosy, fungal infection, cancer, sarcoidosis, and tropical ulcer, can mimic CL or ML lesions (see Fig. 348-2). Because treatment is costly and potentially toxic, diagnostic confirmation is necessary.

DIAGNOSIS

Parasitology

Scraping, fine-needle aspiration, or biopsy of lesions provides appropriate samples in which amastigotes can be identified (Fig. 348-3). Scraping should be performed at both the center and edges of the lesion with a curved scalpel blade. Local anesthesia considerably reduces patients' discomfort and increases sensitivity. Use of an epinephrine-containing local anesthetic (contraindicated for lesions on extremities) or pinching of the lesion between the thumb and finger until blanching appears helps obtain a bloodless scraping, thus optimizing microscopic examination. The 2- to 4-mm skin fragment obtained by punch biopsy provides abundant material, which facilitates the search for scarce parasites and for an alternative diagnosis by culture (e.g., mycobacteria, fungi) as well as by histopathologic examination. Culture of a biopsy sample requires homogenization in saline or culture medium under sterile conditions.

The material obtained by any of these methods can be used for microscopic examination, culture, and PCR. Microscopic examination of Giemsa-stained material is the most widely available method. Culture of the parasite in specific media (such as fetal calf serum–supplemented Schneider's or Novy-Nicolle-McNeal media) allows identification, characterization, and storage of the isolate. Detection of parasitic nucleic acids by molecular diagnosis (mainly PCR) increases sensitivity and allows identification of the *Leishmania* species (see Table 348-1). This is particularly useful in regions (e.g., New World) where several *Leishmania* species—with various clinical outcomes and responses to treatment—coexist. Both culture and molecular-based diagnosis require substantial laboratory infrastructure and technical expertise.

Serology

Serologic diagnosis is of limited use for CL because of low sensitivity and variable specificity. The leishmanin (or Montenegro) skin test (LST) evaluates the cell-mediated response against *Leishmania* spp. The LST requires culture and preferentially fixation of local species of *Leishmania* and therefore lacks standardization. The production of commercial formulations of the LST lacks sustainability. Like serologic tests, the LST does not distinguish between past and present infections.

Complex Manifestations of Cutaneous Leishmaniasis

MUCOSAL LEISHMANIASIS

A proportion of CL infections (about 1 to 10% in Brazil, Bolivia, and Peru) caused by *Leishmania braziliensis* or *L. guyanensis* progress to a metastatic infection of the mucosa of the oral or nasal cavity or larynx, often 1 to 5 years after healing of the initial simple cutaneous lesion. An immunopathologic response results in extensive destruction of local tissue. Allergic rhinitis, paracoccidioidomycosis or other deep mycosis, cancrum oris, leprosy, and sarcoidosis may mimic the lesions of ML. Positive serology (e.g., indirect fluorescent antibody, enzyme-linked immunosorbent assay) or LST indicates possible ML. Parasites are scarce in mucosal lesions. Therefore, a search for parasites in mucosal samples—obtained by scraping or biopsy—by microscopic examination or by culture lacks sensitivity. PCR has proved to be the most sensitive approach to confirm ML.

DIFFUSE CUTANEOUS LEISHMANIASIS

Patients with diffuse CL have an anergic response to *Leishmania* antigens, and nonulcerative nodules, loaded with parasites, disseminate from the initial site of infection to multiple sites on the patient's body. There is no self-cure, and treatment is difficult. This form of the disease is found in South America and East Africa, often associated with *Leishmania amazonensis* and *Leishmania aethiopica* infection.

RECIDIVANS CUTANEOUS LEISHMANIASIS

Recidivans CL is characterized by the development of lesions containing granulomatous tissue. The lesions often take many years to heal and may arise years after healing of a simple localized lesion. New ulcers and papules may form over the edge of the old scar. Infections are normally associated with *L. tropica* infection and are difficult to treat.

TREATMENT Rx

The clinical consequences of CL are dermatologic; concomitant visceral impairment is exceptional. The intensity of the discomfort, related to one or several oozing or unsightly lesions, and the impact of atrophic, hypopigmented or hyperpigmented scars depend on the lesions' topography. In the New World, mucosal impairment may affect up to 1 to 15% of patients (higher in Bolivia). The risk for metastasis (initially nasal and then affecting the entire otorhinolaryngeal zone) has strongly influenced therapeutic decisions. Systemic treatment of any New World CL has been recommended, but recent data indicate that a different strategy should probably be considered.[14] Inadequate surgical excision and pentavalent antimonials or systemic pentamidine administered in excessive doses or without sufficient follow-up may paradoxically be a major risk for patients with CL. Analysis of the number and size of the lesions and their topography, the clinical signs of spread (cutaneous, mucous, or lymphatic), and the predictable cosmetic or functional impact helps in estimating the benefit-risk ratio.[15]

Oral therapy with relatively nontoxic drugs (ketoconazole, fluconazole, itraconazole, miltefosine) would be the simplest option, and oral miltefosine monotherapy has been successful in Brazil.■ Local therapy with intralesional injections of pentavalent antimonials (with or without cryotherapy), photodynamic therapy, and thermotherapy are therefore attractive options to avoid potentially toxic, expensive, or impractical systemic schedules. However, implementation of these methods is hampered by logistic constraints and requires skilled health care providers. An efficient topical ointment would

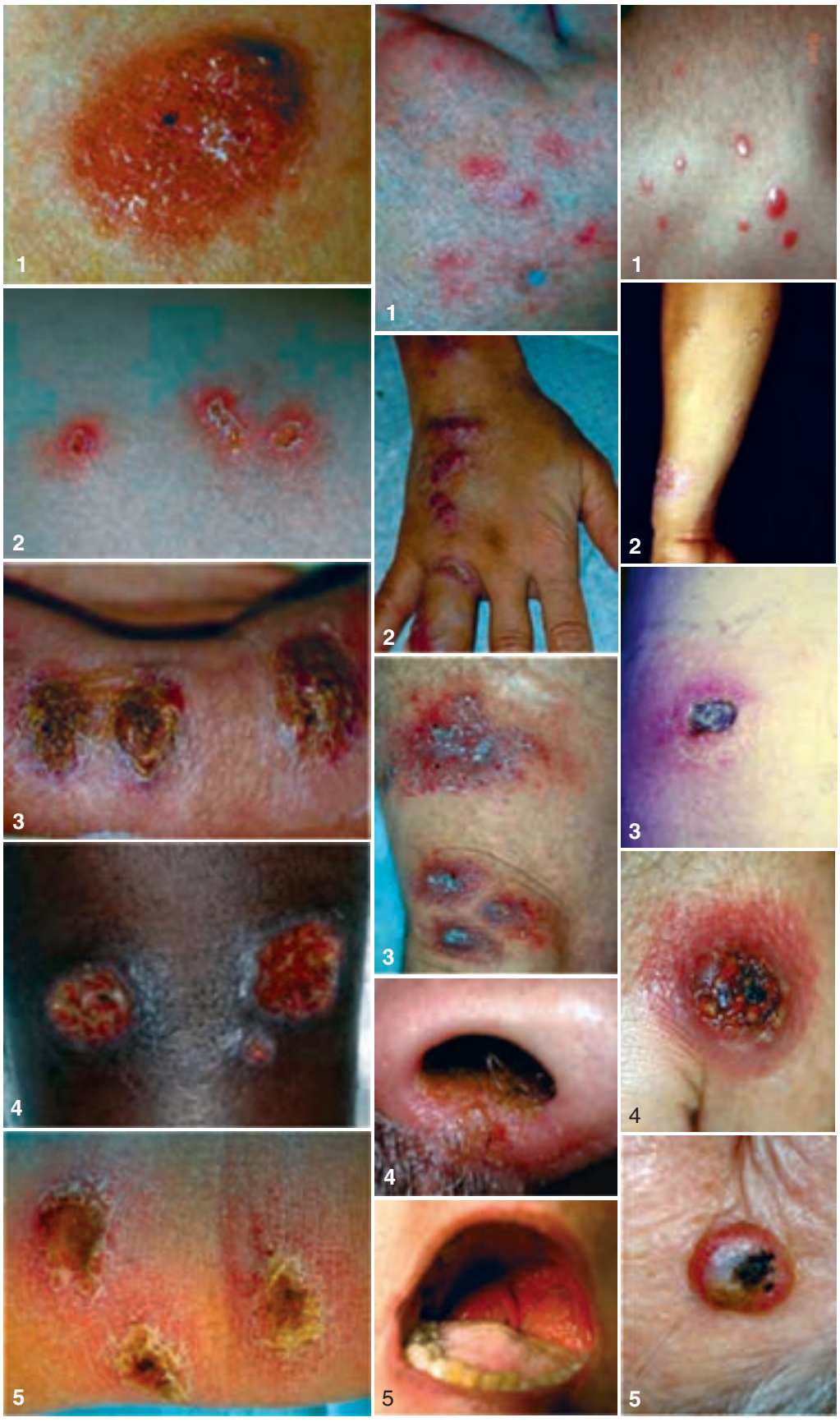

FIGURE 348-2. Clinical features of cutaneous leishmaniasis (CL). A, Typical forms of CL. **A1,** Papular-nodular lesion. **A2,** Squamous lesion. **A3,** Crusty lesion. **A4,** Ulcerated lesion. **A5,** Superinfected lesion. **B,** Atypical forms of CL and mucosal leishmaniasis. **B1,** Multiple papules on the face (*L. infantum*, Balearic Islands). **B2,** Nodular lymphangitis (*L. braziliensis*, Brazil). **B3,** Multiple lesions with numerous peripheral papules (*L. major*, Tunisia). **B4,** Initial spread to the nasal mucosa (anterior septum, *L. braziliensis*, Bolivia). **B5,** Infiltration and ulceration of the tonsils (*L. infantum*, France). **C,** Clinical manifestations that are not CL. **C1,** Multiple papules (late secondary syphilis) (Chapter 319). **C2,** Nodular lymphangitis (sporotrichosis) (Chapter 337). **C3,** Single crusty ulceration (*Mycobacterium ulcerans* infection) (Chapter 325). **C4,** Ulcerated acute *Staphylococcus aureus* infection (Chapter 288). **C5,** Ulcerated nodule keratoacanthoma (Chapter 203).

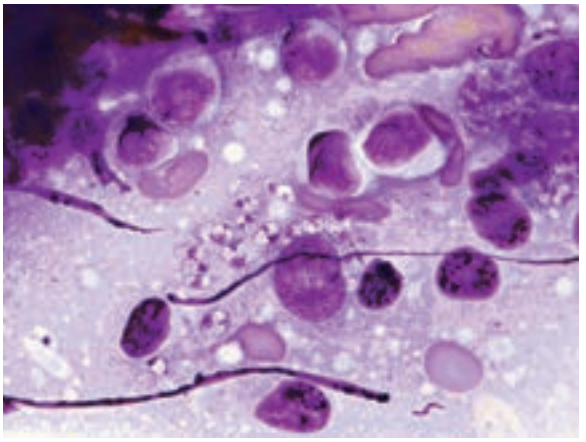

FIGURE 348-3. **Amastigotes of *Leishmania major* in scrapings from a skin ulcer.**

resolve most of these issues, but no efficient formulation is widely available. Several formulations of topical paromomycin have been tested with variable results. Recently, an ointment containing paromomycin and gentamicin has shown efficacy with acceptable local tolerance in patients with CL caused by *L. major* and *L. panamensis*.[A3,A4] An algorithm for accurate therapeutic decisions recently validated in travelers is proposed in E-Figure 348-1.

PREVENTION

There are no modern vaccines for human disease and no prophylactic drug regimens. Leishmanization, or inoculation of people with live virulent parasites to cause a local limited lesion and provide protection, was used for CL (e.g., in Iran) but is not recommended by the World Health Organization. Sandflies are sensitive to most insecticides, and transmission has been controlled with malaria eradication campaigns and residual spraying in houses. Insecticide-impregnated bed nets and dog collars (to prevent transmission of zoonotic CL) have been investigated. The high incidence of vertical and venereal transmission in dogs may cause elimination programs to fail in *L. infantum* foci.

PROGNOSIS

Most untreated patients with established disease ultimately die of the disease. VL has an excellent prognosis with less than a 2% death rate in patients treated soon enough with liposomal amphotericin B. Mortality increases when bleeding (mainly from the digestive tract or lung) or secondary bacterial infection occurs, usually after prolonged untreated evolution or unresponsiveness of VL to first-line agents. Post-therapeutic relapse is frequent in HIV-coinfected patients with low initial or persistently low $CD4^+$ counts. In immunosuppressed patients, VL is rarely the direct cause of death but may give rise to complex therapeutic situations. A proportion of *L. donovani*–infected patients experience PKDL weeks to years after the initial episode, which may rarely lead to severe ocular involvement. CL caused by *L. braziliensis* may metastasize to the nose and other mucous membranes with variable frequency in different areas. CL caused by *L. tropica* often relapses (recidivans CL). A small proportion of patients infected with *L. amazonensis* and *L. aethiopica* experience diffuse CL. Even when treated, most CL lesions leave disfiguring, hypotrophic, hypopigmented, or hyperpigmented scars.

Grade A References

A1. Sundar S, Chakravarty J, Agarwal D, et al. Single-dose liposomal amphotericin B for visceral leishmaniasis in India. *N Engl J Med*. 2010;362:504-512.

A2. Chrusciak-Talhari A, Dietze R, Chrusciak Talhari C, et al. Randomized controlled clinical trial to access efficacy and safety of miltefosine in the treatment of cutaneous leishmaniasis Caused by *Leishmania (Viannia) guyanensis* in Manaus, Brazil. *Am J Trop Med Hyg*. 2011;84:255-260.

A3. Ben Salah A, Ben Messaoud N, Guedri E, et al. Topical paromomycin with or without gentamicin for cutaneous leishmaniasis. *N Engl J Med*. 2013;368:524-532.

A4. Sosa N, Capitan Z, Nieto J, et al. Randomized, double-blinded, phase 2 trial of WR 279,396 (paromomycin and gentamicin) for cutaneous leishmaniasis in Panama. *Am J Trop Med Hyg*. 2013;89:557-563.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

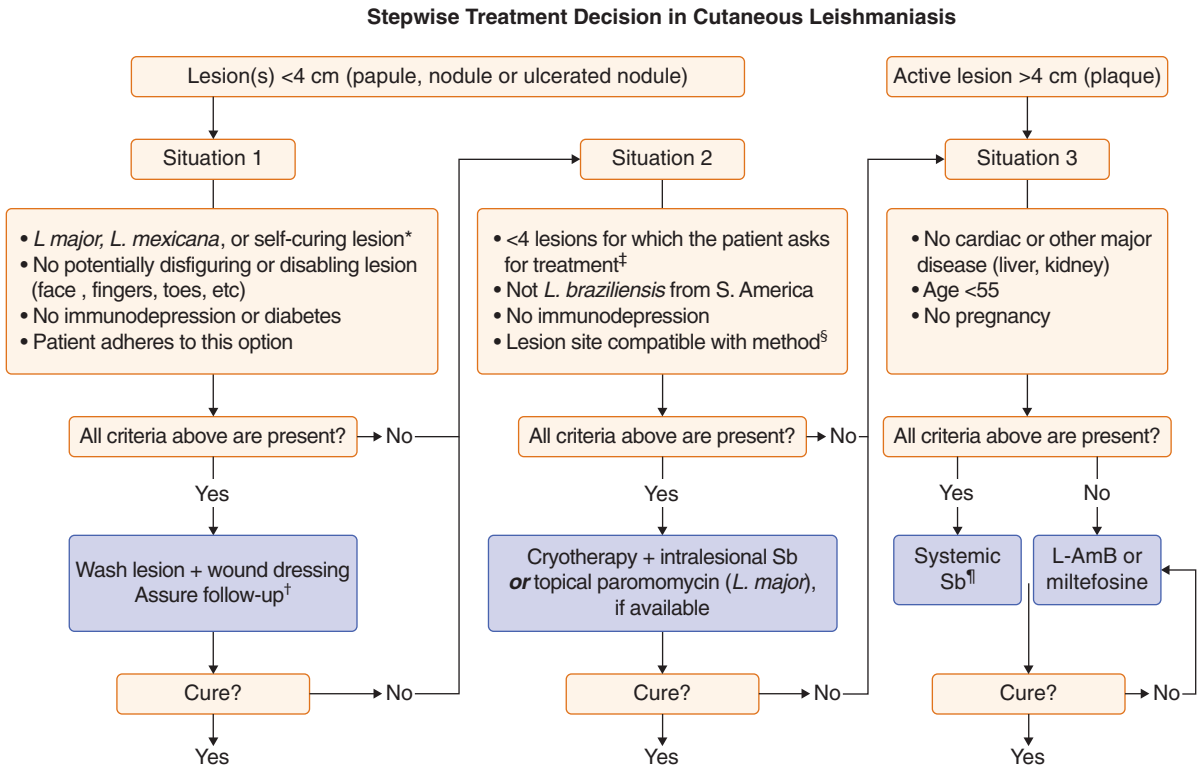

E-FIGURE 348-1. Algorithm of treatment options for cutaneous leishmaniasis (CL)—an example of decision making. (1) Abstention can probably be proposed to patients with scarring lesions. The healing rate in *L. major*–infected patients is 50 to 75% after 3 to 6 months of evolution. (2) Many patients do not understand that no therapy is proposed and will therefore search for a physician ready to recommend immediate therapy, thus raising the risk for an inadequate therapeutic choice (e.g., surgical excision, toxic overdosed systemic course of the reference antileishmanial agent). Genuine adherence of the patient to abstention is therefore essential. As time passes, lack of cure makes abstention more difficult to maintain. Close follow-up is thus important. (3) Most local therapies require injections either for intralesional injection or for local anesthesia. Pain is therefore a strong limiting factor, which makes treatment of more than three lesions in one session almost impracticable in most patients. The crude number of lesions is, however, a poor determinant of treatment choice because most patients with numerous lesions have only two or three large exposed lesions requiring immediate therapy, the other lesions being small, nodular, or unexposed and therefore amenable to abstention or expectant therapy. The number of lesions that can be treated with a topical ointment is much higher. (4) Most lesions on the face (cheek, over the zygomatic bone) can easily be injected. Lesions on nose, lips, eyelids, and ears are, however, in general difficult to inject. Lesions on joints cannot usually be injected but are in general easily treated by applying ointment. (5) Most local methods require heavy or expensive equipment and trained health care providers; they are therefore resource limited. L-AmB = liposomal amphotericin B; Sb = pentavalent antimony. (From Morizot G, Kendjo E, Mouri O, et al. Travelers with cutaneous leishmaniasis cured without systemic therapy. *Clin Infect Dis*. 2013;57:370-380.)

*Self-curing lesions show flattening or reduction in the surface of the ulceration or induration.

†Washing of the lesion, wound dressing, and follow-up were performed in all situations.

‡Lymphatic dissemination did not influence treatment decision. However, because it increased the number of lesions requiring therapy, it sometimes justified the use of systemic therapy.

§Most lesions of limbs, trunk, check, upper cheek, chin, and front were injected, including those close to large joints. Ears, fingers, or toes were not injected. In children, premedication facilitated the injection.

¶Systemic pentamidine for *L. guyanensis/panamensis.*

GENERAL REFERENCES

1. Alvar J, Velez ID, Bern C, et al. Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE*. 2012;7:e35671.
2. Matlashewski G, Arana B, Kroeger A, et al. Visceral leishmaniasis: elimination with existing interventions. *Lancet Infect Dis*. 2011;11:322-325.
3. Lachaud L, Dedet JP, Marty P, et al. Surveillance of leishmaniases in France, 1999 to 2012. *Euro Surveill*. 2013;18:20534.
4. Mansueto P, Seidita A, Vitale G, et al. Leishmaniasis in travelers: a literature review. *Travel Med Infect Dis*. 2014;12:563-581.
5. Ready PD. Epidemiology of visceral leishmaniasis. *Clin Epidemiol*. 2014;6:147-154.
6. Mukhopadhyay D, Dalton JE, Kaye PM, Chatterjee M. Post kala-azar dermal leishmaniasis: an unresolved mystery. *Trends Parasitol*. 2014;30:65-74.
7. World Health Organization. Control of the leishmaniases: report of a meeting of the WHO Expert Committee on the control of leishmaniases, Geneva, 22-26 March 2010. *WHO Tech Rep Series*. 2011;949:1-185.
8. Molina I, Fisa R, Riera C, et al. Ultrasensitive real-time PCR for the clinical management of visceral leishmaniasis in HIV-infected patients. *Am J Trop Med Hyg*. 2013;89:105-110.
9. Elmahallawy EK, Sampedro Martinez A, Rodriguez-Granger J, et al. Diagnosis of leishmaniasis. *J Infect Dev Ctries*. 2014;8:961-972.
10. Croft SL, Olliaro P. Leishmaniasis chemotherapy—challenges and opportunities. *Clin Microbiol Infect*. 2011;17:1478-1483.
11. van Griensven J, Balasegaram M, Meheus F, et al. Combination therapy for visceral leishmaniasis. *Lancet Infect Dis*. 2010;10:184-194.
12. Morizot G, Kendjo E, Mouri O, et al. Travelers with cutaneous leishmaniasis cured without systemic therapy. *Clin Infect Dis*. 2013;57:370-380.
13. Mohapatra S. Drug resistance in leishmaniasis: newer developments. *Trop Parasitol*. 2014;4:4-9.
14. Blum J, Lockwood DN, Visser L, et al. Local or systemic treatment for New World cutaneous leishmaniasis? Re-evaluating the evidence for the risk of mucosal leishmaniasis. *Int Health*. 2012;4:153-163.
15. Goto H, Lindoso JA. Current diagnosis and treatment of cutaneous and mucocutaneous leishmaniasis. *Expert Rev Anti Infect Ther*. 2010;8:419-433.

REVIEW QUESTIONS

1. A 60-year-old woman has been experiencing fever without chills for 10 days, starting 1 month after a 6-month travel from China to Europe. Clinical examination shows splenomegaly and liver enlargement. Laboratory parameters show anemia, leukopenia, and thrombocytopenia. Which of the following statements is correct?

A. The most likely diagnostic possibilities are a hematologic condition and visceral leishmaniasis.
B. The most likely diagnostic possibilities are bacterial infection and Chagas' disease.
C. The most likely diagnostic possibilities are viral infection and histoplasmosis.
D. None of the above is accurate.

Answer: A The association of fever and splenomegaly with pancytopenia is highly suggestive of but not specific for visceral leishmaniasis. Several hematologic conditions (myeloproliferative neoplasms, splenic lymphoma, Castleman's syndrome) are manifested with the same association of signs and laboratory abnormalities. Bacterial conditions and acute Chagas' disease are infrequently associated with pancytopenia, so this option is possible but less frequent. Viral infections (human immunodeficiency virus, human herpesvirus 8, cytomegalovirus) and histoplasmosis may be manifested with splenomegaly and pancytopenia, but these presentations are less prevalent in series of fever of unknown origin (at least in the European context), so this option is possible but less likely.

2. A 3-year-old child has fever, splenomegaly, anemia, leukopenia, and thrombocytopenia. Direct examination of a bone marrow aspirate (Giemsa-stained smear) for *Leishmania* amastigotes is negative. The most sensitive and specific, safest laboratory test to confirm the diagnosis of visceral leishmaniasis is

A. Wait for *Leishmania* culture on the bone marrow aspirate to become positive
B. Test for antileishmanial antibody
C. Quantitative polymerase chain reaction (PCR) for *Leishmania* on a blood sample
D. Direct examination of a splenic aspirate (Giemsa-stained smear)
E. Direct examination of the bone marrow aspirate

Answer: C Quantitative PCR is a highly sensitive and specific test as soon as thresholds for symptomatic disease have been defined (low-grade positivity is frequent in endemic areas). Antileishmanial antibody can be positive because of past infection. Culture is also a sensitive test although generally less so than an optimized PCR based on the amplification of kinetoplastic DNA. Splenic aspiration is not a safe procedure.

3. A 22-year-old HIV-seropositive man with fever, splenomegaly, and pancytopenia has visceral leishmaniasis confirmed by quantitative PCR on a peripheral blood sample. The following items suggest a complicated form.

A. Mucosal hemorrhages
B. Presence of splenomegaly 5 cm below the costal margin
C. Cough
D. Moderate thrombocytopenia

Answer: A Hemorrhage and bacterial superinfection are the two main complications of advanced visceral leishmaniasis. Splenomegaly and mild to moderate thrombocytopenia are present in a majority of patients with uncomplicated visceral leishmaniasis. Dry cough is present in 10 to 15% of patients with visceral leishmaniasis (but will justify diagnostic study to rule out a bacterial superinfection in the lung).

4. A 2-year-old child has parasitologically confirmed, uncomplicated visceral leishmaniasis. The treatment option with the most favorable therapeutic window is

A. Miltefosine 2.5 mg/kg for 28 days orally
B. Pentamidine mesylate 4 mg/kg for 10 days by slow IV infusion
C. Amphotericin B deoxycholate 0.5 mg/kg/day for 14 days by slow IV infusion
D. Meglumine antimoniate (pentavalent antimony) 20 mg SbV/kg/day for 28 days by slow IV infusion
E. Liposomal amphotericin B 10 mg/kg D1 and D2 by slow IV infusion

Answer: E All these options have been used in the therapy of visceral leishmaniasis. The 2-day schedule of liposomal amphotericin B has been validated in children.

5. A 35-year-old woman has a subacute, painless, crusted ulceration of the forearm that appeared 1 month after travel to Israel. Which of the following findings makes the diagnosis of cutaneous leishmaniasis unlikely?

A. The lesion has an oval shape.
B. The lesion is infiltrated.
C. The lesion had a red, painful border for the last 4 days.
D. The lesion reached its maximum size in less than a week.
E. The lesion has small peripheral papules.

Answer: D Cutaneous leishmaniasis evolves subacutely, and therefore a lesion that reaches its maximum size within a week of its appearance makes the diagnosis unlikely. Firm, sharply defined lesions are characteristic, and numerous satellite papulopustules can occur. Although the uncomplicated lesions are "cold and painless," superinfection can make them "hot and painful" with surrounding erythema.

349

TOXOPLASMOSIS

JOSE G. MONTOYA

DEFINITION

The ubiquitous parasite *Toxoplasma gondii* is the etiologic agent of toxoplasmosis. The term *toxoplasmosis* is reserved for the disease process in which clinical manifestations are present, whereas *Toxoplasma infection* best describes the asymptomatic presence of the parasite. Toxoplasmosis may result in significant morbidity and mortality of the fetus, newborn, and immunocompromised patient. However, *T. gondii* can also be responsible for chorioretinitis, lymphadenopathy, pneumonia, brain abscesses, myositis, myocarditis, and hepatitis in immunocompetent patients.

A more aggressive form of congenital and adult toxoplasmosis appears to occur in certain geographic locales in Latin America, where pneumonia, fever of unknown origin, brain abscesses, and death have been reported in HIV-negative and otherwise immunocompetent individuals. Recent epidemiologic studies have established the role of novel risk factors for the acquisition of the acute infection, including the ingestion of untreated water, oysters, mussels, or clams.

The Pathogen

T. gondii is an intracellular parasite with a high capacity for host cell invasion due to a motile invasive form (tachyzoite or trophozoite) characterized by an evolutionarily unique apical complex and a mechanism of actin-based motility.

T. gondii maintains a highly clonal population structure in North America and Europe that consists of three main lineages: types I, II, and III. This relatively low genetic diversity comes as a surprise, given the fact that the parasite has the capacity to infect any warm-blooded animal. It has a sexual life cycle that takes place in the small intestine of cats. In Europe, type II strains predominate and are most commonly associated with human toxoplasmosis, both in congenital infections and in immunocompromised patients. In North America, types II and I appear to be equally common. In Latin America, type I strains are common, but *Toxoplasma* strains appear to be genetically more diverse, with many different genotypes described mainly in Brazil and French Guiana. These atypical Latin American strains, initially called exotic strains, belong to several haplogroups that are endemic to Latin America and have been found to be associated with more severe clinical manifestations; physicians need to entertain the diagnosis of toxoplasmosis in their patients presenting with pneumonia, fever of unknown origin, brain abscesses, lymphadenopathy, or chorioretinitis who are from or were traveling in these endemic areas.

In nature, the parasite exists in several forms, including the tachyzoite (Fig. 349-1A), the oocyst that contains sporozoites (Fig. 349-1B), the tissue cyst that contains bradyzoites (Fig. 349-1C), and the sexual forms (macrogametes and microgametes). The tachyzoite is the rapidly proliferating form of the parasite responsible for the clinical manifestations of toxoplasmosis observed in the setting of the acute infection or reactivation of a latent infection. The tissue cyst is the slower metabolic form of the parasite responsible for chronic infection and for its transmission through meat consumption in humans and animals. Tissue cysts persist in tissues for the life of the host and cannot be eradicated by drugs. Tissue cysts vary in shape and size from younger ones that contain only a few bradyzoites to older tissue cysts that may contain several thousand bradyzoites and may reach more than 100 μm in size. The central nervous system (CNS), eye, and skeletal, smooth, and heart muscles appear to be the most common sites of tissue cyst formation (i.e., latent infection). Oocysts are primarily responsible for the worldwide and large-scale spread of the parasite among different populations of other animals and humans. Domestic and feral animals belonging to the Felidae family shed

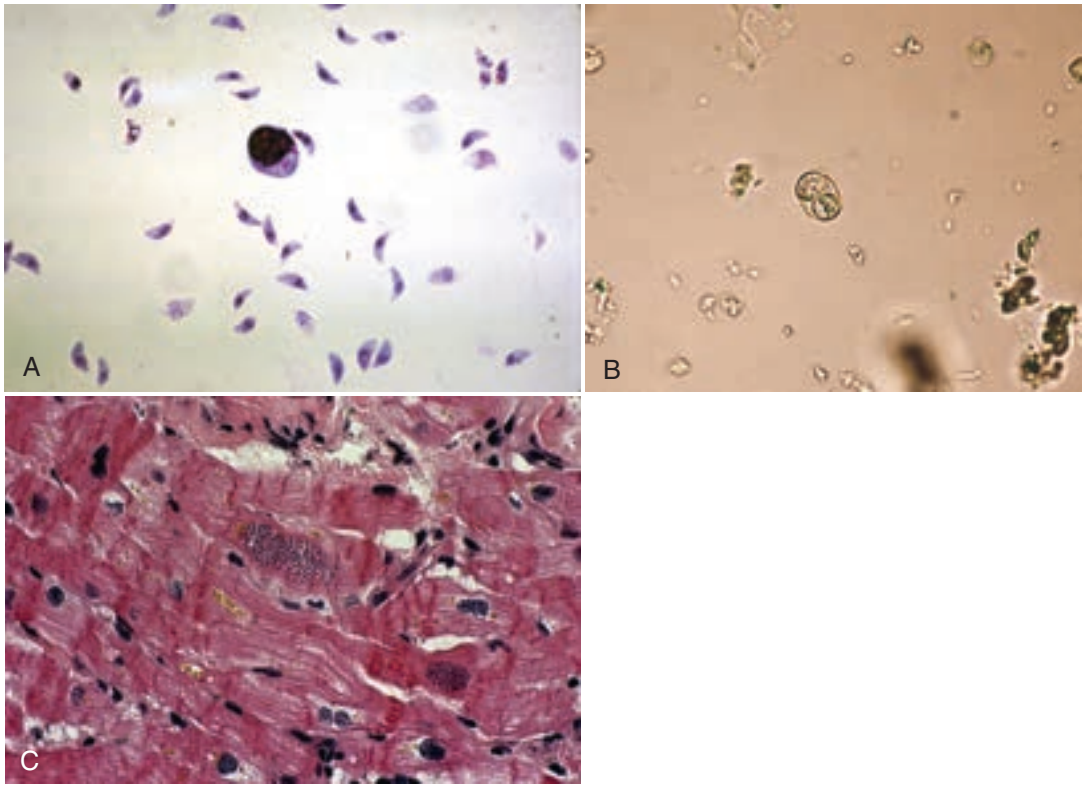

FIGURE 349-1. ***Toxoplasma gondii*** **exists in nature primarily in three forms. A,** Tachyzoites are bow or banana shaped, measure 2 to 3 µm wide and 5 to 7 µm long, and can be stained with Wright-Giemsa stain. **B,** Oocysts isolated from cat's feces are subspherical to spherical and measure 10 × 12 µm in diameter. **C,** Tissue cyst observed in human myocardial tissue stained with hematoxylin and eosin. Tissue cysts vary in shape and size and may reach more than 100 µm.

oocysts after they ingest any of the infectious forms of the parasite: tachyzoites, tissue cysts, and oocysts. As many as 10 million oocysts may be shed in the feces of an infected animal in a single day for periods varying from 7 to 20 days. Oocysts may remain viable for as long as 18 months in moist soil; this results in an environmental reservoir from which incidental hosts may be infected.

EPIDEMIOLOGY

The prevalence of *T. gondii* infection varies significantly according to geographic locale and the socioeconomic status of the population. It can be as low as 7% in England and as high as 80% in the Central African Republic. Seroprevalence increases with age because of increasing length of exposure with age, and it is inversely associated with socioeconomic status because of the strong influence of hygienic and alimentary habits in the transmission of the parasite. The overall age-adjusted seroprevalence of *T. gondii* infection in the United States has been recently reported at 11%, but it may be higher in certain geographic areas or socioeconomic groups. The seroprevalence of the parasite has declined during the past 30 years in the United States[1] and several other countries but appears to be stable or increasing in certain geographic locales, such as in the tropics (e.g., Latin America).

Humans and non-felid animals are incidental hosts and become infected primarily by the ingestion of infected meat containing tissue cysts or of contaminated food, water, or soil material containing oocysts. They can also become infected during gestation by vertical transmission of the parasite from the mother to her offspring. In addition, humans can become infected through organ transplantation and, more rarely, in the setting of laboratory accidents. Ingestion of raw or undercooked meat contaminated with tissue cysts is probably one of the major routes of transmission in humans. Ingestion of untreated water, food, or soil contaminated with oocysts is another significant route of infection with the parasite. Untreated water has been found to be the source of large epidemics of toxoplasmosis in Canada and Brazil.

The main risk factors for *T. gondii* infection in the United States include eating raw ground beef; eating rare lamb; eating locally produced cured, dried, or smoked meat; working with meat; drinking unpasteurized goat's milk; having three or more kittens. In this study, eating raw oysters, clams, or mussels was also identified as a novel risk factor. Untreated water as a potential vehicle for the transmission of *T. gondii* has been established in several large epidemiologic studies and was found to have a trend toward increased risk for acute infection in the United States.

In up to 50% of individuals acutely infected with *T. gondii*, it is not possible to identify the presence of a known risk factor for their acute infection. Thus, attempting to establish whether a patient is at risk for toxoplasmosis solely on the basis of the epidemiologic history is a futile task. Patients may have been infected with *T. gondii* even if they have not owned or been in contact with cats, have not eaten undercooked meat or shellfish, and have not ingested untreated water.

The seroprevalence of *T. gondii* infection in immunocompromised patients reflects the seroprevalence of the particular population from which they come. Latent *T. gondii* infection can reactivate in these patients, particularly in those with the acquired immunodeficiency syndrome (AIDS) and in hematopoietic stem cell transplant (HSCT, including bone marrow transplant) and liver transplant patients. In these patients, it is important to establish whether they have been infected with the parasite before severe immunosuppression ensues or their transplant procedure because serologic testing in severe immunosuppression or after transplantation may be unreliable. Approximately 30% of AIDS patients who are infected with *T. gondii* will develop toxoplasmosis by reactivation of their chronic infection if their CD4 count falls below 200 cells/µL and they are not taking anti-*Toxoplasma* primary prophylaxis. The advent of highly active antiretroviral therapy, in addition to the use of primary anti-*Toxoplasma* prophylaxis, has clearly contributed to the decline in the incidence of toxoplasmosis in AIDS patients. Among HSCT (including bone marrow transplant) patients, those recipients who are *Toxoplasma* seropositive before transplantation, receive an allogeneic graft, and develop graft-versus-host disease have the highest risk of reactivation.

For solid organ transplants, the highest risk for toxoplasmosis is observed when an allograft from a *Toxoplasma*-seropositive donor (D^+) is transplanted into a seronegative recipient (R^-). In D^+/R^- patients, there is a 25% risk for development of potentially life-threatening toxoplasmosis if effective anti-*Toxoplasma* prophylaxis is not instituted. It is highly advised that the *Toxoplasma* serologic status of the donor and the recipient be established before transplantation. Serologic test results are less reliable in the post-transplantation period, and they may significantly vary without any clinical relevance.

Transmission of *T. gondii* to the fetus can occur during pregnancy when a woman acquires her primary infection during gestation. The incidence of seroconversion for pregnant women in the United States has been estimated at 0.27%. The overall rate of transmission of the parasite (prevalence of congenital toxoplasmosis) in seroconverting women has been estimated between 50 and 60% before spiramycin was instituted as an attempt to decrease vertical transmission and 25 to 30% thereafter. The transmission rate increases with the gestational age at which maternal infection is acquired. In women who have been treated for toxoplasmosis during gestation, it can be as low as 4.5% during the first trimester, 31.7% during the second trimester, and as high as 63% during the third trimester. The likelihood of severe disease is inversely proportional to the gestational age at which maternal infection was acquired. Although objective data are lacking on the prevalence of congenital toxoplasmosis in the United States, it has been estimated that among the approximately 4.2 million live births per year, congenital *T. gondii* infection occurs in 500 to 5000 newborns. The global annual incidence of congenital toxoplasmosis has been estimated to be 190,100 cases. This is equivalent to a burden of 1.2 million disability-adjusted life years. Particularly high burdens are seen in South America and in some Middle Eastern and low-income countries.[2]

Congenital transmission in women who were infected before conception has only rarely been reported in immunosuppressed patients, in those who were acutely infected shortly before conception (i.e., within 3 months of conception), and in those who have been reinfected with a more virulent strain.

PATHOBIOLOGY

Pathogenesis

After oral infection with tissue cysts (e.g., contaminated meat) or oocysts (e.g., contaminated soil, water, or food), the wall of both infectious forms is disrupted by the digestive juices of the gastrointestinal tract. Bradyzoites (from cysts) and sporozoites (from oocysts) are released and converted to the tachyzoite form. Tachyzoites have the capacity to infect contiguous cells or distant tissues by hematogenous or lymphatic spread. Tachyzoites appear to actively and rapidly migrate across epithelial cells and may traffic to distant sites while they are extracellular (acute infection).[3] Their histologic hallmark is necrosis surrounded by inflammation.

In immunocompetent individuals, the immune system controls the proliferation of the tachyzoite and induces its conversion to bradyzoites, facilitating the final formation of tissue cysts (chronic infection). Tissue cysts persist for the life of the host, and *T. gondii* can be isolated from tissues of individuals who have died from causes other than toxoplasmosis.

It appears that the activation of well-orchestrated immune responses is required for the successful resistance against *T. gondii*. Innate, humoral, and cellular immune responses likely to be involved in preventing the uncontrolled proliferation of tachyzoites include activation of the monocyte-macrophage system, dendritic cells, natural killer cells, and αβ and γδ T cells; *T. gondii*–specific and cytotoxic $CD4^+$ and $CD8^+$ T cells; and interferon-γ, interleukin 12, tumor necrosis factor-α, interleukin 10 and other cytokines, transforming growth factor-β, costimulatory molecules (e.g., CD28, CD40 ligand), and, to some degree, immunoglobulins. Recent studies have shown that innate type 1 immune responses that involve interferon-γ–producing natural killer cells and neutrophils, rather than interferon-γ–producing T cells, predetermine host resistance to *T. gondii*.[4]

In immunocompromised patients previously infected with *T. gondii*, decreased T cell–mediated or severe impairment in B cell–mediated immune responses can facilitate the reactivation of their infection (i.e., conversion of bradyzoites in their tissue cysts into rapidly proliferating tachyzoites). Toxoplasmosis in this setting is 100% lethal if untreated.

Pathology

Most of the data on the pathology of toxoplasmosis come from studies of congenitally infected babies and immunosuppressed patients. CNS lesions of patients with toxoplasmosis are characterized by significant necrosis and surrounding inflammation. In congenitally infected cases, necrotic areas may calcify and lead to typical radiographic findings suggestive but not diagnostic of toxoplasmosis. Hydrocephalus may result from obstruction of the aqueduct of Sylvius or foramen of Monro. Tachyzoites and tissue cysts may be visualized near necrotic foci, near or in glial nodules, in perivascular regions, and in cerebral tissue uninvolved by inflammatory changes.

Formation of multiple brain abscesses is relatively common in patients with AIDS. In the areas around the abscesses, edema, vasculitis, hemorrhage, and cerebral infarction secondary to vascular involvement may also be present. Important associated features in toxoplasmic encephalitis are arteritis, perivascular cuffing, and astrocytosis. A "diffuse form" of toxoplasmic encephalitis has been described with histopathologic findings of widespread microglial nodules without abscess formation in the gray matter of the cerebrum, cerebellum, and brain stem.

Pulmonary involvement by *T. gondii* in the immunodeficient patient can lead to interstitial pneumonitis, necrotizing pneumonitis, consolidation, pleural effusion or empyema, or all of these. Chorioretinitis in AIDS patients is characterized by segmental panophthalmitis and areas of coagulative necrosis associated with tissue cysts and tachyzoites.

Toxoplasmic lymphadenitis in immunocompetent individuals may result in patterns of findings that are often diagnostic of the disease: a reactive follicular hyperplasia; irregular clusters of epithelioid histiocytes encroaching on and blurring the margins of the germinal centers; and focal distention of sinuses with monocytoid cells.

CLINICAL MANIFESTATIONS

Toxoplasmosis should be entertained in the differential diagnosis of several clinical syndromes in immunocompetent, unborn, newborn, infant, pediatric, adult, and immunocompromised patients (Table 349-1). Symptoms result from the primary infection or reactivation of the parasite due to T cell–mediated or severe B cell–mediated immunodeficiency. Primary infection can be asymptomatic in a significant number of individuals, and conventional risk factors for the acute infection may not be present in a particular patient. Thus, the possibility of acute toxoplasmosis or *T. gondii* infection should not be ruled out because of the absence of epidemiologic risk factors (e.g., exposure to cats or undercooked meat) or symptoms in a given patient. For this reason, if the goal is to detect each case of primary *T. gondii* infection in a population of patients (e.g., pregnant women), only systematic and universal screening methods can achieve such an objective; testing of only symptomatic patients or those with conventional epidemiologic risk factors will miss a significant number of acute cases.

Severity of toxoplasmosis due to primary infection or reactivation in a given patient or population may be influenced by the infecting strain (e.g., type I strain), size of the inoculum, infectious form (e.g., oocyst vs. cyst), or genetics of the host (e.g., presence of HLA-DQ3). During the past decade, it has become clear that patients infected in certain geographic locales (e.g., South America) have more aggressive clinical presentations, including a more severe primary infection and disease due to reactivation. These observations need to be kept in mind on seeing ill travelers returning from the endemic areas or patients who were born in these areas and in whom toxoplasmosis by reactivation has been included in their differential diagnosis.

Lymphadenopathy due to toxoplasmosis may be completely asymptomatic or be accompanied by other symptoms, such as fever (temperature as high as 104° F), headache, general malaise, and fatigue. It can be localized or generalized. A solitary, occipital, and painlessly enlarged lymph node can be the sole manifestation of toxoplasmosis in a child, pregnant woman, or adult. However, more generalized cervical, axillary, and abdominal lymphadenopathy has also been reported. Lymph nodes are usually 1 to 3 cm in size, nonsuppurative, and nontender on palpation. They usually regress within 12 weeks, but a mild relapse of the lymphadenopathy has been observed between months 3 and 6. Recurrence of toxoplasmic lymphadenopathy beyond the sixth month is extremely rare.

Ocular disease due to *T. gondii* can be asymptomatic or symptomatic and can be the result of congenital or postnatally acquired infection. In both settings (congenitally and postnatally acquired), toxoplasmic chorioretinitis can be discovered at the time of the diagnosis of the infection or as a reactivation of the subsequent latent infection months to years later.[5] Up to 17% of patients acutely infected with the parasite in Brazil and in a Canadian outbreak of toxoplasmosis presented with concurrent symptomatic toxoplasmic chorioretinitis at the time their acute infection was diagnosed. Similar cases have been described in Europe and the United States. *T. gondii* strain type appears to be a contributing factor determining severity and recurrence of ocular toxoplasmosis.[6] Symptomatic ocular disease primarily consists of a retinochoroiditis that can result in blurred vision, eye pain, decreased visual acuity, floaters, scotoma, photophobia, or epiphora. The morphology of the retinal lesions on funduscopic examination is thought to be characteristic of toxoplasmosis. An active whitish infiltrate is usually attached to the darkly pigmented border of an older scar (Fig. 349-2). However, retinal lesions tend to be less typical in older or immunocompromised patients.

TABLE 349-1 CLINICAL MANIFESTATIONS OF TOXOPLASMOSIS IN HUMANS

CLINICAL CATEGORIES	CLINICAL MANIFESTATIONS AND SYNDROMES
Primary infection	
Immunocompetent individuals and pregnant women	Most patients are asymptomatic. However, in ≈10% of patients, the following symptoms or syndromes, alone or in various combinations, have been reported: fever, lymphadenopathy, headache, myalgias, arthralgias, sore throat, stiff neck, nausea, abdominal pain, anorexia, rash, confusion, earache, eye pain, general malaise, fatigue. Chorioretinitis resulting in blurred vision, eye pain, decreased visual acuity, floaters, scotoma, photophobia, or epiphora Hepatitis; myositis; myocarditis Disseminated disease, pneumonia, brain abscesses, and even death have been observed in immunocompetent individuals infected with atypical strains of *T. gondii* (e.g. in Latin America).
Congenital toxoplasmosis	
Fetus	Ultrasound study can be normal or reveal hydrocephalus, brain or hepatic calcifications, splenomegaly, ascites, pericarditis. Fetal death can also result from overwhelming infection.
Newborn	Newborn can be entirely normal, have a nonspecific illness, or have abnormal findings on physical examination including chorioretinitis, strabismus, blindness, seizures, encephalitis, abnormal cephalic perimeter (microcephaly or hydrocephalus), psychomotor or mental retardation, hepatosplenomegaly, pneumonitis, diarrhea, hypothermia, jaundice, petechiae, rash. Intracranial calcifications can be present in brain imaging studies. Newborns can also die as a result of overwhelming infection.
Children and adults	Children can continue to suffer the chronic sequelae of the congenital disease. However, children may be born apparently normal and become symptomatic for the first time during childhood, adolescence, or adulthood, primarily in the form of reactivation of congenitally acquired chorioretinitis.
Chronic infection	Asymptomatic. However, some investigators have proposed a role of chronic infection in individuals with schizophrenia, bipolar disease, and behavioral issues including a higher incidence of motor vehicle accidents. Chorioretinitis can occur as a reactivation of congenital or postnatally acquired disease in otherwise immunocompetent individuals.
Reactivation of chronic infection in immunocompromised patients	Multiple brain abscesses, diffuse encephalitis, seizures, chorioretinitis, fever of unknown origin, pneumonia, myocarditis, hepatosplenomegaly, lymphadenopathy, rash

Other less common but well-documented syndromes have been associated with the acute infection, including hepatitis, myositis, myocarditis, and skin lesions. More aggressive disease, including pneumonia, brain abscesses, and death, has been observed in immunocompetent patients in Latin America.

Primary infection can be observed in solid transplant patients when an allograft from a seropositive donor is transplanted into a seronegative recipient (D^+/R^-). Disseminated and localized toxoplasmosis has been reported in this setting, including myocarditis, pneumonia, fever of unknown origin, and encephalitis.

Congenital disease can be asymptomatic in the fetus, newborn, child, or adult. However, most of the infected offspring will eventually develop signs and symptoms of toxoplasmosis (see Table 349-1). The classic triad of chorioretinitis, hydrocephalus (Fig. 349-3A and B), and brain calcifications is highly suggestive of toxoplasmosis and is primarily seen in babies whose mothers have not been treated against the parasite during gestation. Eye examination by an experienced pediatric ophthalmologist may reveal active or inactive toxoplasmic chorioretinitis. New lesions have been reported in up to 30% of congenitally infected children observed up until 11 years of age when their mothers have been treated but in up to 70% when their mothers have not.

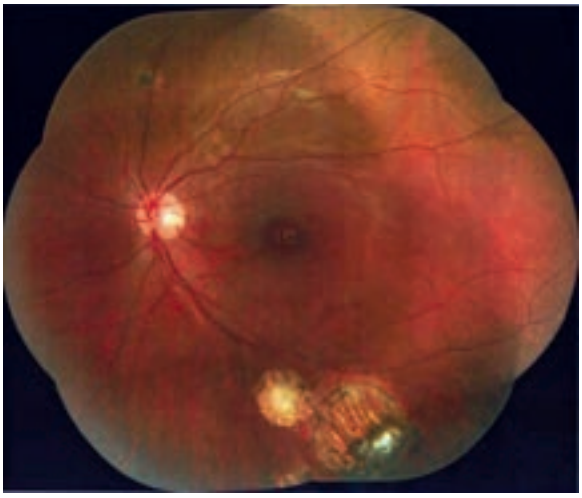

FIGURE 349-2. The morphology of the retinal lesions on funduscopic examination believed to be characteristic of toxoplasmic retinochoroiditis. An active whitish infiltrate is usually attached to the darkly pigmented border of an older scar.

Chronic infection is believed to be asymptomatic. However, several investigators have recently suggested the possibility that chronic infection may play a role in the predisposition of infected individuals to have a higher frequency of traffic accidents, mental illness (e.g. schizophrenia, bipolar disease), and behavioral abnormalities.

Reactivation of the chronic infection is usually observed in patients with significant impairment of T cell–mediated immunity or severe impairment of B cell–mediated immunity. Toxoplasmosis by reactivation can cause brain abscesses, diffuse encephalitis, seizures, chorioretinitis, fever of unknown origin, pneumonia, myocarditis, hepatosplenomegaly, lymphadenopathy, and rash. Although multiple brain abscesses (Fig. 349-3C) are commonly described in patients with toxoplasmic encephalitis, diffuse encephalitis without space-occupying lesions by magnetic resonance imaging has been reported with a very high mortality. Fever with pneumonia can be the sole manifestation of toxoplasmosis in immunocompromised patients, including HSCT and solid organ transplant recipients. Toxoplasmic pneumonitis can be manifested with cough, dyspnea, hypoxia, and diffuse bilateral or localized infiltrates. Most patients with pneumonia have been reported to have bilateral ground-glass opacities that can be confused with *Pneumocystis* pneumonia or viral etiologies. Fever alone has frequently been described in patients with allogeneic HSCT and liver transplant patients. Reactivation in heart tissue causing congestive heart failure, arrhythmias, and pericarditis has been described.

DIAGNOSIS

Laboratory methods for the diagnosis of *T. gondii* infection and toxoplasmosis include serologic tests, polymerase chain reaction (PCR), microscopy examination of tissue and body fluids, and attempts to isolate the parasite (Table 349-2).

The first step is to establish whether the patient has never been infected with *Toxoplasma* or has an acute or latent *T. gondii* infection; this can be accomplished by serologic testing. Serologic tests can establish whether a patient has never been infected or is acutely or chronically infected regardless of the presence or absence of symptoms. Available serologic tools include methods to detect *T. gondii*–specific IgG-, IgM-, IgA-, IgE-, and IgG-based avidity and differential agglutination (AC/HS).

With the use of commercial kits for the detection of IgG and IgM, most hospital-based or commercial laboratories can reliably diagnose the absence of *T. gondii* infection (negative IgG/negative IgM) and the presence of chronic infection (positive IgG/negative IgM). However, the diagnosis of acute infection is more challenging. A positive IgM test result is observed during the acute infection, but it may remain positive for months to years in

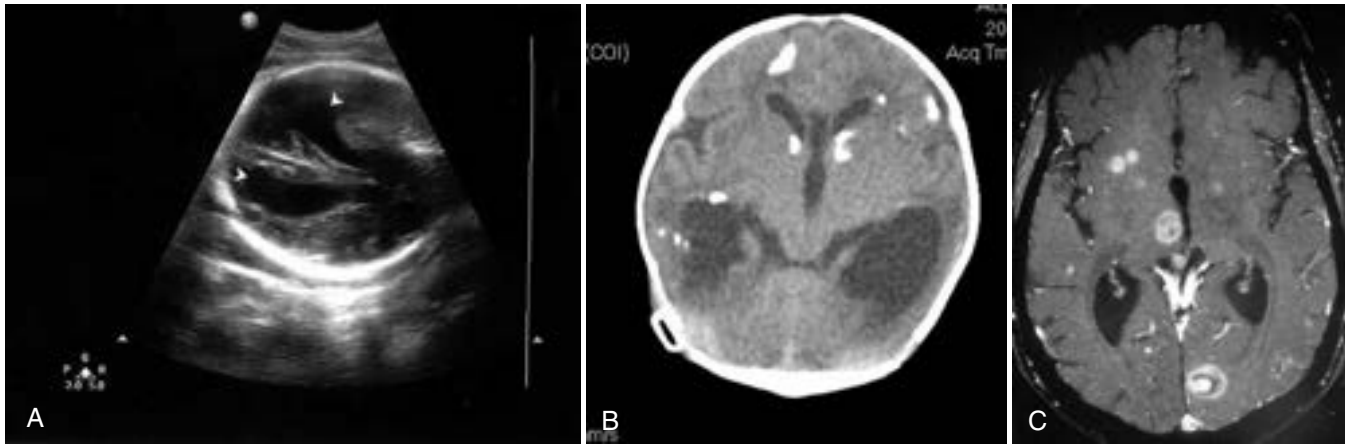

FIGURE 349-3. Radiologic manifestations of central nervous system toxoplasmosis. A, Fetal ultrasound of a fetus congenitally infected with *T. gondii* in the United States reveals hydrocephalus. **B,** Computed tomography scan of the brain of a newborn congenitally infected with *T. gondii* in the United States reveals hydrocephalus and calcifications. **C,** Magnetic resonance image of the brain of an AIDS patient revealing multiple ring-enhancing brain lesions.

TABLE 349-2 LABORATORY METHODS FOR THE DIAGNOSIS OF *T. GONDII* INFECTION AND TOXOPLASMOSIS* IN HUMANS

METHOD	CLINICAL INTERPRETATION
Serologic tests	
IgG	A positive test result establishes that the patient has been infected with *T. gondii*. However, a negative test result can be seen in patients infected within 4 weeks before serum sampling or in patients unable to produce IgG (e.g., immunocompromised hosts).
IgM	A positive test result suggests but is not necessarily diagnostic of an acute infection. Sera with positive IgM test results should be sent to a reference laboratory for confirmatory testing that includes a more specific IgM assay and additional tests, including avidity, acetone (AC)–fixed versus formalin (HS)–fixed (AC/HS) differential agglutination test of tachyzoites, and IgA and IgE.† Positive IgM test results can be seen in chronically infected patients because of persistence of the IgM response or false-positive results observed in certain commercial kits.
Confirmatory testing for positive IgM test results	IgG avidity test; differential agglutination (AC/HS); IgA, IgE performed at a reference laboratory.† At PAMF, a high IgG avidity test result‡ or a nonacute AC/HS test result indicates that the patient has been infected for more than 4 months (avidity) or 12 months (AC/HS).
Polymerase chain reaction (PCR)	B1 and AF487550 genes are the most commonly used targets for amplification. PCR test can be performed in any body fluid, including amniotic fluid, peripheral blood, cerebrospinal fluid, bronchoalveolar lavage fluid, vitreous fluid, aqueous humor, peritoneal fluid, pleural fluid, ascitic fluid, and urine. PCR can also be performed in any tissue. A positive test result in any body fluid establishes that the patient has either acute or reactivated toxoplasmosis. However, a positive PCR test result in tissue is more difficult to interpret because it does not differentiate symptomatic toxoplasmosis from a latent infection. Although DNA extraction is more cumbersome, it can be attempted in paraffin-embedded tissue.
Direct visualization of the parasite	Identification of tachyzoites in any body fluid or tissue is diagnostic of toxoplasmosis due to acute infection or reactivation of a chronic infection. Tachyzoites can be identified by hematoxylin and eosin or cytologic studies but are better visualized with Wright-Giemsa and immunoperoxidase stains. Identification of cysts by hematoxylin and eosin or immunoperoxidase stains confirms the presence of *T. gondii* in the host but does not necessarily establish that the patient has toxoplasmosis. However, a strong inflammatory response surrounding the cysts is highly suggestive of toxoplasmosis, possibly explaining the patient's symptoms.
Attempts to isolate the parasite	A positive isolation study in any body fluid establishes the diagnosis of toxoplasmosis. Isolation of *T. gondii* in cell cultures or peritoneal cavity of mice can be attempted at reference laboratories. These studies can be important in trying to establish a correlation between the genetics of the parasite and its clinical manifestations.
Histology of the lymph node	Classic histologic triad is considered diagnostic: follicular hyperplasia, epithelioid histiocytes impinging on the margins of the germinal centers, and monocytoid cells focally distending sinus walls.

**T. gondii* infection = asymptomatic presence of the parasite. Toxoplasmosis = active symptoms and signs are present.
†For example, Palo Alto Medical Foundation Toxoplasma Serology Laboratory (PAMF-TSL), Palo Alto, Calif; *www.pamf.org/serology*; 650-853-4828; *toxolab@pamf.org*.
‡The window of exclusion for acute infection varies for different avidity kits (usually between 3 and 5 months).

certain individuals without any apparent clinical relevance. In addition, commercial IgM kits have been designed to be extremely sensitive so that an acute infection will be rarely missed; as a consequence, their specificity is somewhat sacrificed. Of patients who are found to be IgM positive at hospital-based or commercial laboratories, 60% are found to be chronically infected when their serum is tested at the national reference laboratory for the study and diagnosis of toxoplasmosis in the United States (Palo Alto Medical Foundation Toxoplasma Serology Laboratory [PAMF-TSL], Palo Alto, Calif; www.pamf.org/serology; 650-853-4828; toxolab@pamf.org). At PAMF-TSL, a battery of confirmatory tests (avidity, differential agglutination, IgA, IgE) are performed in addition to the "gold standard" dye test for IgG and the "double" sandwich capture enzyme-linked immunosorbent assay for IgM for confirmatory testing of positive IgM test results obtained at hospital-based or commercial laboratories. These tests are used in various combinations, depending on the clinical scenario of each patient and the questions of the treating physician. For an appropriate interpretation of the serologic test results obtained at PAMF-TSL, it is also crucial to have relevant clinical information available for the medical consultants (for instance, low positive IgG and positive IgM test results with a high IgG avidity test result will mean no risk of congenital toxoplasmosis for a 16-week pregnant woman, but the same results can be highly supportive of the diagnosis of toxoplasmic encephalitis for an AIDS patient with multiple brain lesions). At PAMF-TSL, three interpretations can be given to final serologic test results: (1) acute, results are consistent with a recently acquired infection; (2) chronic, consistent with an infection acquired in the distant past; and (3) equivocal, cannot exclude a recently acquired infection; an earlier or subsequent sample is required to attempt to establish whether the infection is acute or chronic. For serologic test results consistent with an acute infection, an attempt is made by the

medical consultants at PAMF-TSL to establish the approximate date that the infection was acquired.

The definitive diagnosis of toxoplasmosis (due to primary infection or reactivation of a chronic infection) requires the identification of tachyzoites in tissues or body fluids or the amplification of parasite DNA in any body fluid (see Table 349-2). Tachyzoites can be visualized in histologic sections stained with hematoxylin and eosin or in cytologic preparations without any specific staining, but they are better visualized with Wright-Giemsa (see Fig. 349-1A) and *T. gondii*–specific immunoperoxidase stains. Real-time PCR has become a useful method for the diagnosis of toxoplasmosis in the cerebrospinal fluid (CSF), and it is the diagnostic tool of choice for toxoplasmic encephalitis in immunocompromised patients (in CSF, assuming the lumbar puncture is deemed safe and feasible) and for the prenatal diagnosis of congenital toxoplasmosis (in amniotic fluid). Isolation of the parasite in any body fluid is also diagnostic of toxoplasmosis and can be attempted at reference laboratories. The diagnosis of toxoplasmosis can be indirectly supported by the use of serologic tools, demonstration of cysts in tissues (see Fig. 349-1C) surrounded by an inflammatory response, and attempts to isolate the parasite (see Table 349-2); in cases of toxoplasmic lymphadenitis, histologic features can be diagnostic.

Immunocompetent Patients, Pregnant Women, and Patients with Lymphadenopathy

The first diagnostic goal in these patients is to establish whether they have ever been infected with *T. gondii*. If *T. gondii*–specific IgG and IgM test results are negative, the possibility that the patient's symptoms are due to the parasite can be ruled out. During pregnancy, these results confirm that the mother has not been exposed to *T. gondii* but that she is at risk, if exposed, to acquire the primary infection during pregnancy and therefore can potentially transmit *T. gondii* to her offspring.[7] In attempting to determine whether the patient is infected with *T. gondii*, it is important to perform both IgG and IgM tests because during the first 4 weeks of the acute infection, the IgG can still be negative while the IgM will be positive. In these cases, seroconversion can be diagnosed by having a new positive IgG test result in a subsequent serum sample. In rare instances, infected patients may be IgG negative because of their incapacity to produce IgG.

If the patient is found to be IgG positive, the next goal is to determine whether the patient is having an acute infection or has been chronically infected (e.g., >6 months). If the IgG titer is low (e.g., a dye test at PAMF-TSL ≤512) and the IgM test result is negative, the patient has essentially been infected for more than 6 months. With these results, a patient whose symptoms or lymphadenopathy had a date of onset within 6 months of serum sampling will be considered unlikely to have toxoplasmosis. For a pregnant woman whose serum was obtained within 6 months of gestation, these results will mean that her infection was acquired before conception and that the risk for congenital toxoplasmosis is essentially zero.

If the patient is found to have a positive IgM test result confirmed to be indicative of a recently acquired infection at a reference laboratory and the onset of symptoms or lymphadenopathy falls within the time predicted by the serologic test results for the acquisition of *T. gondii*, the patient will be diagnosed as having acute toxoplasmosis. For a pregnant woman, if the predicted time for when the infection was acquired falls within her gestational age, she will be diagnosed with toxoplasmosis during pregnancy and at risk for transmitting the parasite to her baby.

In patients with lymphadenopathy, the histologic examination of the lymph node tissue obtained by excisional biopsy can be diagnostic or pathognomonic of toxoplasmic lymphadenitis (see earlier under Pathology).

Prenatal and Postnatal Diagnosis of Congenital Toxoplasmosis

Once the diagnosis of acute toxoplasmosis or *T. gondii* infection has been confirmed or is highly suspected in the mother, the next step is to attempt to establish whether her offspring has been infected. Consultation with reference centers for the study and diagnosis of congenital toxoplasmosis is highly recommended.

Ultrasound abnormalities can be consistent with or suggestive of congenital toxoplasmosis (see Fig. 349-3A), but they are not diagnostic. The method of choice for the prenatal diagnosis of congenital toxoplasmosis is a PCR in amniotic fluid obtained at 18 weeks of gestation. Attempts to diagnose congenital toxoplasmosis from amniotic fluid obtained before 18 weeks of gestation should be avoided because the studies reported to date have included only pregnant women whose gestational age was 18 weeks or more. In addition, false-negative results have been reported in women whose amniocentesis was performed before 18 weeks of gestation. The overall sensitivity of the amniotic fluid PCR has been reported between 64 and 92% and is highly dependent of the gestational age at which the mother acquired the infection.

In the newborn, congenital toxoplasmosis can be confirmed by positive *T. gondii*–specific serologic test results or PCR.[8] Samples for serology should be obtained in the peripheral blood of the baby. Cord blood should be avoided because of the high rate of maternal blood contamination. However, there is still a small degree of maternal blood contamination in newborn blood obtained by peripheral venipuncture, during the first 5 days of life for IgM antibodies and the first 10 days of life for IgA antibodies. A positive IgM immunosorbent agglutination assay (after 5 days of life) or IgA enzyme-linked immunosorbent assay (after 10 days) is diagnostic of congenital disease. Congenitally infected babies can be positive for both; positive for either one, but negative for the other test; or negative for both. A positive *T. gondii*–specific IgM in CSF is diagnostic of congenital disease, but testing of the CSF by PCR rather than for IgM is strongly recommended because of the higher sensitivity of the PCR test. The diagnosis can also be made by a positive PCR in peripheral blood, CSF, or urine. The CSF of infected infants may exhibit very high levels of protein (e.g., 1000 mg/dL). Cellular response in CSF is characterized by lymphocytosis, and eosinophilia may be present. Brain imaging studies may reveal calcifications or hydrocephalus; computed tomography scan is superior to ultrasound examination in the detection of these CNS abnormalities (see Fig. 349-3B).

Ocular Disease

Serologic and PCR testing can be helpful in the diagnosis of toxoplasmic chorioretinitis. An IgG-negative/IgM-negative patient is unlikely to have ocular disease due to toxoplasmosis. However, patients should be tested at reference laboratories (e.g., PAMF-TSL) because their *T. gondii*–specific IgG can be present but at very low levels such that only a gold standard method like the dye test can detect it. In patients with eye lesions typical of toxoplasmic chorioretinitis (see Fig. 349-2), a positive IgG test result at a relatively low titer (e.g., a dye test at PAMF-TSL ≤512) and a negative IgM test result are diagnostic of ocular disease due to the parasite reactivation. If the serologic test reveals a positive IgM result and confirmatory testing at PAMF-TSL establishes the diagnosis of an acute infection in patients 1 year of age or older, the eye disease is most likely the result of eye involvement in the setting of a recent and postnatally acquired infection.

In patients with atypically appearing lesions or in whom the response to anti-*Toxoplasma* drugs is atypical or absent, a *T. gondii*–specific immune load (aqueous humor) or PCR in ocular fluids (vitreous fluid is preferable to aqueous humor because of probable higher sensitivity, but it is riskier to obtain) should be considered.

Immunocompromised Patients

It appears that acute infection rarely occurs in immunocompromised patients. However, life-threatening disease can occur when the patient's latent infection is reactivated by AIDS, HSCT, liver transplantation, or other diseases characterized by severe T-cell deficiency.

Toxoplasmosis can also develop when *T. gondii* is transmitted from a seropositive donor to a seronegative recipient through an infected allograft (e.g., heart, liver, kidney).[9] Therefore, to establish the risk for toxoplasmosis and to have a high index of suspicion when patients develop illnesses suggestive of toxoplasmosis, all immunocompromised patients should be tested for *T. gondii*–specific IgG as soon as they have been diagnosed with the underlying disease or it has been established that they will be subsequently immunosuppressed. In addition, serologic testing may not be reliable when immunosuppression is advanced or severe. Post-transplantation serologic test results for IgG antibodies may remain positive or may rise, decrease, or even become negative. Thus, pretransplantation *Toxoplasma* serologic studies are critical for interpretation of subsequent test results and clinical evaluation. Solid organ donors should also be tested for *T. gondii*–specific IgG as their allograft has the potential to transmit the parasite to the transplanted patient (e.g., heart, heart-lung, kidney, kidney-pancreas, liver, liver-pancreas). Toxoplasmosis in solid organ transplant recipients causes substantial morbidity, including disseminated disease, and mortality.

In AIDS patients suspected of having toxoplasmic encephalitis with the presence of multiple brain-occupying and ring-enhancing lesions (see Fig. 349-3C), a CD4 count below 200 cells/μL, and a positive *T. gondii*–specific IgG, the response to anti–*Toxoplasma*-specific treatment is considered an

additional "diagnostic" indicator of toxoplasmic encephalitis. In these patients, invasive diagnostic tests (e.g., lumbar puncture, brain biopsy) are considered unnecessary unless they do not respond to treatment within a 7- to 10-day period. This diagnostic paradigm should not be applied to other populations of immunosuppressed patients (e.g., transplant patients) because their differential diagnosis often includes other pathogens, such as invasive mold infections. In those patients, examination of the CSF by PCR or brain biopsy should be attempted at the outset of the illness.

The definitive diagnosis of toxoplasmosis in immunosuppressed patients relies on PCR, direct visualization of the parasite, and attempts for isolation of the organism (see Table 349-2). PCR testing in body fluids is the diagnostic method of choice for immunosuppressed patients at risk for toxoplasmosis who develop unexplained fever (e.g., in whole blood), pneumonia (e.g., in bronchoalveolar lavage fluid), brain lesions (e.g., in CSF), or other compatible syndromes. Theoretically, PCR can be performed in any body fluid or tissue, and laboratories have validated its use in most fluids and some tissues. Attempts to identify the tachyzoite or tissue cyst in tissues by microscopy can be enhanced with the use of the *T. gondii*–specific immunoperoxidase stain. CSF examination by PCR or brain biopsy should be initially considered in AIDS patients who have a low likelihood of having toxoplasmic encephalitis, such as those who have a single lesion by magnetic resonance imaging examination, have tested seronegative for *T. gondii* infection, have a CD4 count of more than 200 cells/μL, or who are not responding to an appropriate anti-*Toxoplasma* regimen.

TREATMENT

Rx

Principles of antiparasitic therapy are discussed in Chapter 344. Treatment of toxoplasmosis is indicated for immunocompetent patients with acute infection in the setting of ongoing fever, myocarditis, myositis, hepatitis, pneumonia, brain lesions or skin lesions, and lymphadenopathy accompanied by severe or persisting symptoms. Treatment is indicated as well for patients with active chorioretinitis due to primary infection or reactivation of a latent infection (Table 349-3).[10] In immunocompetent patients, treatment is prescribed for 3 to 4 weeks or until symptoms have subsided, whichever is longer. For toxoplasmic lymphadenitis, trimethoprim-sulfamethoxazole (TMP-SMZ) (8 mg TMP/40 mg SMZ per kilogram per day divided into two doses for 1 month) increased the cure rate to 65% compared with a 13% resolution rate with placebo.[11] Treatment is also often recommended for all pregnant women suspected of having or diagnosed with primary infection during gestation (Table 349-4) in an attempt to prevent transmission of the parasite to the fetus (spiramycin) or, if congenital infection has occurred, to start treatment of the fetus in utero (pyrimethamine, sulfadiazine, and folinic acid). During pregnancy, treatment regimens are prescribed for the duration of the gestation. There was a worldwide controversy about the efficacy of spiramycin to decrease the incidence of congenital toxoplasmosis and of pyrimethamine, sulfadiazine, and folinic acid to decrease the frequency of clinical signs in infected offspring. Although no definitive studies ever disproved their efficacy, several epidemiologic studies erroneously concluded that there was no evidence of benefit. However, since 2006, several studies have reported a strong association between prenatal treatment of women infected during gestation (with spiramycin or pyrimethamine, sulfadiazine, and folinic acid) and decreases in the incidence of congenital toxoplasmosis and frequency of clinical signs in infected offspring. Spiramycin is recommended for pregnant women who have been definitively diagnosed to have or are highly suspected of having an acute infection during pregnancy and acquired before 18 weeks of gestation. Spiramycin should be given throughout pregnancy unless fetal infection is suspected or documented. Fetal infection should be investigated with amniotic fluid PCR at 18 weeks of gestation to see whether *Toxoplasma* DNA is amplified and with monthly follow-up ultrasound examinations. Therapy with pyrimethamine, sulfadiazine, and folinic acid is recommended for pregnant women who have been definitively diagnosed to have or are highly suspected of having an acute infection during pregnancy and acquired after18 weeks of gestation, whose amniotic fluid PCR is positive for the presence of *Toxoplasma* DNA, or whose follow-up ultrasound examinations are suggestive of congenital toxoplasmosis in the setting of acute *Toxoplasma* infection during gestation. In addition, newborns and infants diagnosed with or suspected of having congenital toxoplasmosis should also be treated during their first year of life (see Table 349-4).

Treatment at higher doses is urgently indicated for all immunocompromised patients with toxoplasmosis due to reactivation of their latent infection or primary infection acquired by natural exposure to the parasite or by solid organ transplantation (see Table 349-3). If untreated, toxoplasmosis in these patients has a very high rate of morbidity and mortality.

TABLE 349-3 TREATMENT REGIMENS FOR PATIENTS WITH ACUTE OR PRIMARY TOXOPLASMOSIS AND IMMUNOCOMPROMISED PATIENTS WITH TOXOPLASMOSIS DUE TO REACTIVATION*

	IMMUNOCOMPETENT PATIENTS WITH ACUTE INFECTION†	IMMUNOCOMPROMISED PATIENTS
Pyrimethamine (PO)	50 mg every 12 hr for 2 days, followed by 25 to 50 mg daily	200 mg loading dose, followed by 50 mg/day (<60 kg) to 75 mg/day (>60 kg)
plus		
Folinic acid‡	10-20 mg daily (during and 1 week after therapy with pyrimethamine)	10-20 mg daily (up to 50 mg/day)
	plus either sulfadiazine or clindamycin or atovaquone as follows:	*plus either sulfadiazine or clindamycin or atovaquone as follows*:
Sulfadiazine (PO)	75 mg/kg (first dose), followed by 50 mg/kg every 12 hr (maximum 4 g/day)	1000 mg (<60 kg) to 1500 mg (>60 kg) every 6 hr
or		
Clindamycin (PO or IV)	300 mg every 6 hr	600 mg every 6 hr (up to 1200 mg every 6 hr)
or		
Atovaquone (PO)	1500 mg orally twice daily	1500 mg orally twice daily
Trimethoprim-sulfamethoxazole (PO or IV)	10 mg/kg/day (trimethoprim component) in two or three doses	10 mg/kg/day (trimethoprim component) in two or three doses (doses as high as 15-20 mg/kg/day have been used)
Pyrimethamine and folinic acid	Same doses as above	Same doses as above
plus		
Clarithromycin (PO)	500 mg every 12 hr	500 mg every 12 hr
or		
Dapsone (PO)	100 mg/day	100 mg/day
or		
Azithromycin (PO)	900 to 1200 mg/day	900 to 1200 mg/day

After the successful use of a combination regimen during the acute or primary therapy phase, the same agents at half-dose are usually used for maintenance or secondary prophylaxis.
*Preferred regimens: pyrimethamine, sulfadiazine, and folinic acid or trimethoprim-sulfamethoxazole. Assistance is available for the diagnosis and management of patients with toxoplasmosis at the Palo Alto Medical Foundation Toxoplasma Serology Laboratory (PAMF-TSL), Palo Alto, Calif; *www.pamf.org/serology*; 650-853-4828; *toxolab@pamf.org*.
†Particularly in the setting of myocarditis, myositis, hepatitis, pneumonia, brain or skin lesions, and lymphadenopathy accompanied by severe or persisting symptoms. Also indicated for those with active ocular disease due to primary infection or reactivation.
‡Folinic acid = leucovorin; folic acid must not be used as a substitute for folinic acid.

PREVENTION

Primary Infection

Because approximately 50% of patients may inadvertently become infected with the parasite without having a recognized risk factor for acute infection, only systematic serologic testing can establish whether a patient has been exposed to *T. gondii*. Thus, each pregnant woman and immunocompromised patient should be screened for *T. gondii*–specific IgG and IgM regardless of their epidemiologic history. Seronegative pregnant women and immunocompromised individuals should be counseled on how to maximize their prevention efforts to avoid infection with *T. gondii*. In addition, seronegative pregnant women should be tested serially during gestation in an attempt to diagnose seroconversion at the earliest time possible. In some countries, such as France, seronegative pregnant women are mandated by law to be tested every month for *T. gondii*–specific IgG and IgM. Women who seroconvert are offered spiramycin (if infected before 18 weeks of gestation) or pyrimethamine, sulfadiazine, and folinic acid (if infected after 18 weeks). Mothers whose amniotic fluid is found to be positive by PCR or those in whom fetal ultrasound study is highly suggestive of congenital toxoplasmosis are offered pyrimethamine, sulfadiazine, and folinic acid. Although infection often occurs in the absence of known risk factors for the acute infection, educational interventions to avoid exposure to the parasite have been shown to be effective in decreasing the incidence of seroconversion during gestation.

The majority of epidemiologic studies worldwide have recognized contaminated and undercooked meat as one of the main risk factors for the

TABLE 349-4 TREATMENT REGIMENS FOR PREGNANT WOMEN WHO HAVE LIKELY ACQUIRED *T. GONDII* INFECTION DURING GESTATION AND INFANTS SUSPECTED OF HAVING OR CONFIRMED TO HAVE CONGENITAL TOXOPLASMOSIS

	DURING PREGNANCY	IN CONGENITAL DISEASE
Spiramycin (oral)	Recommended for pregnant women suspected of having or confirmed to have acquired the infection during gestation and before 18 weeks of gestation. Spiramycin should be administered until delivery in those with negative amniotic fluid PCR test results and normal follow-up ultrasound studies or low suspicion of fetal infection. Spiramycin is not teratogenic, and it is available in the United States only through the Investigational New Drug (IND) process at the Food and Drug Administration (301-796-1600). Prior medical consultation is required.* Dosage: 1 g (3 million units) every 8 hr (for a total of 3 g or 9 million units per day)	Not recommended during pregnancy if the fetus has been documented to be or suspected to have been infected. In the setting of fetal infection, pyrimethamine, sulfadiazine, and folinic acid should be instituted (see below)
Pyrimethamine (oral) plus sulfadiazine (oral) plus folinic acid† (oral)	Recommended for women ≥18 weeks of gestation in whom it is suspected or confirmed that the acute infection has been acquired at or after 18 weeks of gestation or who have a positive amniotic fluid PCR test result or an abnormal ultrasound study suggestive of congenital toxoplasmosis. Pyrimethamine is teratogenic and should not be used during pregnancy before week 18 (in some centers in Europe, it is used as early as week 14). Sulfadiazine should not be used alone. Dosages: Pyrimethamine: 50 mg every 12 hr for 2 days followed by 50 mg daily Sulfadiazine: 75 mg/kg (first dose) followed by 50 mg/kg every 12 hr (maximum 4 g/day) Folinic acid† (leucovorin): 10-20 mg daily (during and for 1 week after pyrimethamine therapy)	**Infant** (treatment regimen is usually recommended for 1 year): Pyrimethamine: 1 mg/kg every 12 hr for 2 days; followed by 1 mg/kg/day for 2 or 6 months; followed by 1 mg/kg/day every Monday, Wednesday, Friday Sulfadiazine: 50 mg/kg every 12 hr Folinic acid† (leucovorin): 10 mg three times weekly Prednisone (if CSF protein ≥1 g/dL or severe chorioretinitis): 0.5 mg/kg every 12 hr (until CSF protein <1 g/dL or resolution of severe chorioretinitis) **Older children with active disease** (usually 1-2 weeks beyond resolution of clinical manifestations): Pyrimethamine: 1 mg/kg every 12 hr (maximum 50 mg) for 2 days followed by 1 mg/kg/day (maximum 25 mg) Sulfadiazine: 75 mg/kg (first dose) followed by 50 mg/kg every 12 hr Folinic acid† (leucovorin): 10-20 mg three times weekly Prednisone (severe chorioretinitis): 1 mg/kg/day, divided bid, maximum 40 mg/day, rapid taper

*Palo Alto Medical Foundation Toxoplasma Serology Laboratory (PAMF-TSL), Palo Alto, Calif; *www.pamf.org/serology*; 650-853-4828; *toxolab@pamf.org*; or U.S. (Chicago) National Collaborative Treatment Trial Study (NCCTS), 773-834-4152.
†Folic acid should not be used as a substitute for folinic acid.
CSF = cerebrospinal fluid; PCR = polymerase chain reaction.

transmission of the parasite. This appears to be the case in Europe, North America, and Latin America. Tissue cysts in meat are rendered nonviable by γ-irradiation (0.4 kGy), heating throughout to 67° C, or freezing to −20° C for 48 hours and then thawing. Cured, dried, or smoked meat has been associated with the acute infection and should not be considered *Toxoplasma* free. Soil exposure and soil-related activities have been reported to play a more prominent role in transmission in certain geographic areas, such as Latin America.

In seronegative recipients of a solid organ from a seropositive donor, TMP-SMZ for at least 6 months or pyrimethamine for at least 6 weeks has been reported to be effective in the prevention of primary infection in the newly immunosuppressed patient.

Reactivation of Latent Infection in Immunocompromised Patients and Those with Ocular Disease

Drugs used to prevent reactivation of the latent infection in immunosuppressed hosts include TMP-SMZ (e.g., single strength or 80 mg TMP and 400 mg SMZ, 1 tablet per day) and atovaquone (1500 mg/day). Dapsone-pyrimethamine and sulfadoxine-pyrimethamine have been also reported to be effective, but their use appears to be limited because of potential hematologic toxicity.

Prophylaxis against reactivation of latent infection has been successful in AIDS patients dually infected with HIV and *T. gondii* (*Toxoplasma* IgG seropositive) and whose $CD4^+$ T-cell counts are below 200 cells/μL. For prophylactic purposes, TMP-SMZ should probably not be used below a minimum dose of 160 mg TMP/800 mg SMZ orally twice a day on a thrice-weekly regimen or 80 mg TMP/400 mg SMZ once a day. In AIDS patients, 100 mg of dapsone plus 50 mg of pyrimethamine orally twice weekly or atovaquone (1500 mg/day) has also been effective in preventing toxoplasmic encephalitis. Findings in these studies have been extrapolated to non-AIDS immunosuppressed patients because of the absence of data in this population of patients.

Toxoplasma-seropositive recipients of an allogeneic HSCT (Chapter 178) who develop graft-versus-host disease represent a unique challenge. Reactivation of toxoplasmosis can be manifested by a nonspecific illness (e.g., fever or pneumonia) and be life-threatening. The disease is often not recognized. Atovaquone prophylaxis has been proposed as an alternative regimen in these patients, given the potential bone marrow toxicity of TMP-SMZ. Some investigators have proposed a preemptive strategy in which *Toxoplasma*-seropositive patients who receive an allogeneic HSCT are monitored on a routine basis (e.g., weekly for the first 100 days) with *T. gondii* PCR. Those found positive would receive preemptive prophylaxis with TMP-SMZ or atovaquone.

Discontinuation of prophylaxis against toxoplasmic encephalitis has proved safe in AIDS patients receiving highly active antiretroviral therapy who demonstrate an increase in their $CD4^+$ T-cell counts to at least 200 cells/μL and whose viral load has been undetectable for at least 6 months.

In patients with ocular toxoplasmosis who experience frequent relapses (e.g., more than two episodes per year), 80 mg TMP/400 mg SMZ daily for at least 1 year has been shown to be effective in the prevention of their recurrences.

PROGNOSIS

The primary infection has a wide spectrum of manifestations in humans, from asymptomatic in most individuals to pneumonia or life-threatening disease if it is acquired in certain areas of the world. Primary infection can also be fatal in the fetus and in immunocompromised individuals. Early diagnosis and treatment can make a significant difference in the prognosis of these patients.

It is not clear at this time whether chronic infection in immunocompetent individuals is clinically irrelevant. Several investigators have proposed that latent infection with the parasite may play a significant role in mental illness (e.g., schizophrenia) or in the propensity of the infected individual to incur motor vehicle accidents. Immunocompetent patients can reactivate chronic infection in their retina, and the prognosis is influenced by the proximity of the lesions to the macula, involvement of one or both eyes, and number of relapses. It is believed that treatment can slow the progression of these lesions and expedite their healing.

Reactivation of latent infection in immunosuppressed individuals with significant defects in their T cell–mediated or B cell–mediated immunity, if untreated, is 100% fatal. Even when treated in an intensive care unit, disseminated toxoplasmosis in immunocompromised patients has a mortality rate of about 80%.[11]

Grade A References

A1. Alavi SM, Alavi L. Treatment of toxoplasmic lymphadenitis with co-trimoxazole: double-blind, randomized clinical trial. *Int J Infect Dis*. 2010;3:e67-e69.

A2. Felix JP, Lira RP, Zacchia RS, et al. Trimethoprim-sulfamethoxazole versus placebo to reduce the risk of recurrences of *Toxoplasma gondii* retinochoroiditis: randomized controlled clinical trial. *Am J Ophthalmol*. 2014;157:762-766.e1.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Jones JL, Parise ME, Fiore AE. Neglected parasitic infections in the United States: toxoplasmosis. *Am J Trop Med Hyg*. 2014;90:794-799.
2. Torgerson PR, Mastroiacovo P. The global burden of congenital toxoplasmosis: a systematic review. *Bull World Health Organ*. 2013;91:501-508.
3. Walker DM, Oghumu S, Gupta G, et al. Mechanisms of cellular invasion by intracellular parasites. *Cell Mol Life Sci*. 2014;71:1245-1263.
4. Yarovinsky F. Innate immunity to *Toxoplasma gondii* infection. *Nat Rev Immunol*. 2014;14:109-119.
5. Kijlstra A, Petersen E. Epidemiology, pathophysiology, and the future of ocular toxoplasmosis. *Ocul Immunol Inflamm*. 2014;22:138-147.
6. Vasconcelos-Santos DV. Ocular manifestations of systemic disease: toxoplasmosis. *Curr Opin Ophthalmol*. 2012;23:543-550.
7. Stillwaggon E, Carrier CS, Sautter M, McLeod R. Maternal serologic screening to prevent congenital toxoplasmosis: a decision-analytic economic model. *PLoS Negl Trop Dis*. 2011;5:e1333.
8. Kieffer F, Wallon M. Congenital toxoplasmosis. *Handb Clin Neurol*. 2013;112:1099-1101.
9. Fernandez-Sabe N, Cervera C, Farinas MC, et al. Risk factors, clinical features, and outcomes of toxoplasmosis in solid-organ transplant recipients: a matched case-control study. *Clin Infect Dis*. 2012;54:355-361.
10. Rajapakse S, Shivanthan MC, Samaranayake N, et al. Antibiotics for human toxoplasmosis: a systematic review of randomized trials. *Pathog Glob Health*. 2013;107:162-169.
11. Schmidt M, Sonneville R, Schnell D, et al. Clinical features and outcomes in patients with disseminated toxoplasmosis admitted to intensive care: a multicenter study. *Clin Infect Dis*. 2013;57:1535-1541.

REVIEW QUESTIONS

1. Toxoplasmosis should be suspected in the following groups of patients.

A. Only those who own cats
B. Only those who eat undercooked or raw meat
C. Those with clinical syndromes suggestive of toxoplasmosis regardless of their epidemiologic history
D. Those who own cats or eat undercooked or raw meat

Answer: C One of the greatest misconceptions in the care of patients in whom the diagnosis of toxoplasmosis should be entertained is the belief among clinicians that *Toxoplasma* infection should be suspected only in patients with epidemiologic risk factors. In all epidemiologic studies addressing risk factors for *Toxoplasma* infection, a risk factor could not be established in up to 50% of individuals with toxoplasmosis. Toxoplasmosis should be suspected on the basis of clinical manifestations and should not be excluded solely by the absence of epidemiologic risk factors. (See Epidemiology section.)

2. The diagnosis of lymphadenitis due to toxoplasmosis can be established by

A. Amplification of *Toxoplasma* DNA from the affected lymph node tissue by the polymerase chain reaction (PCR)
B. Isolation of the *Toxoplasma* strain by inoculation of macerated lymph node tissue in the peritoneal cavity of mice in reference laboratories
C. Presence in lymph node biopsy specimen of a classic histologic triad: reactive follicular hyperplasia, irregular clusters of epithelioid histiocytes encroaching on the margins of the germinal centers, and focal distention of sinuses with monocytoid cells.
D. Presence in lymph node biopsy specimen of caseating and necrotizing granulomas
E. Microscopic examination of lymph node tissue with Wright-Giemsa stain to identify presence of tachyzoites or tissue cysts

Answer: C Toxoplasmic lymphadenitis can be diagnosed by the identification of this classic inflammatory response in the microscopic examination of lymph node tissue. This classic histologic triad (reactive follicular hyperplasia, irregular clusters of epithelioid histiocytes encroaching on or blurring the margins of the germinal centers, and focal distention of sinuses with monocytoid cells) has not been observed in other diseases causing lymphadenitis. The diagnosis requires intact lymph node tissue (e.g., excisional lymph node biopsy) and cannot be made with the fine-needle aspiration of a lymph node. Toxoplasmic lymphadenitis typically does not result in formation of caseating or necrotizing granulomas. Attempts at identifying the presence of tachyzoites or tissue cysts by various stain methods, amplifying the parasite's DNA by PCR, or isolating the parasite in reference laboratories from lymph node tissue have all yielded negative results or have been successful in only a handful of case reports. The most effective approach to diagnosis of toxoplasmic lymphadenitis is by performing *Toxoplasma* serologic tests at a reference laboratory and, if indicated, lymph node excisional biopsy.

3. The definitive diagnosis of toxoplasmosis in hematopoietic cell, stem cell, or bone marrow transplant recipients can be made by the following methods, *except*

A. Amplification of *Toxoplasma* DNA in cerebrospinal fluid by PCR in patients with encephalitis
B. Amplification of *Toxoplasma* DNA in bronchoalveolar lavage fluid by PCR in patients with pneumonia
C. Amplification of *Toxoplasma* DNA in vitreous fluid by PCR in patients with atypical retinal lesions
D. Amplification of *Toxoplasma* DNA in peripheral blood by PCR in patients with fever of unknown origin
E. Amplification of *Toxoplasma* DNA in brain tissue by PCR in patients with encephalitis

Answer: E PCR is the primary method for the diagnosis of toxoplasmosis in immunocompromised patients. A positive *Toxoplasma* PCR test result in cerebrospinal fluid, bronchoalveolar lavage fluid, vitreous fluid, and peripheral blood and other body fluids is diagnostic of toxoplasmic encephalitis, pneumonia, chorioretinitis, and dissemination, respectively, in these patients. Ideally, all immunocompromised patients should be tested for *Toxoplasma* IgG and IgM before they become immunosuppressed or as soon as the diagnosis of the underlying immunosuppressive disorder has been made because serologic testing is less reliable or not at all reliable once the patient is in a more advanced immunosuppressive state. The identification of patients at risk of toxoplasmosis is facilitated by the identification of those who are *Toxoplasma* IgG positive and should not be solely based on epidemiologic history because up to 50% of at-risk patients may be missed. (See Diagnosis section, Immunocompromised Patients, and Table 349-2.)

4. All pregnant women should be tested for *Toxoplasma* IgG and IgM as early as possible during gestation to determine whether they have an acute infection (positive *Toxoplasma* IgG/IgM confirmed with additional tests at a reference laboratory) or are at risk of acquiring an acute infection during gestation (negative *Toxoplasma* IgG/IgM). Those identified as having acute *Toxoplasma* infection or toxoplasmosis during pregnancy acquired before 18 weeks of gestation should be treated with the following drug.

A. Spiramycin
B. Pyrimethamine
C. Sulfadiazine
D. Folinic acid
E. Clindamycin

Answer: A Spiramycin is the drug of choice during pregnancy in an attempt to prevent transmission of the parasite to the fetus. Spiramycin is not teratogenic and can be obtained at no cost in the United States through the Investigational New Drug (IND) process at the Food and Drug Administration (301-796-1600). Medical consultation with the Palo Alto Medical Foundation Toxoplasma Serology Laboratory ([PAMF-TSL], Palo Alto, Calif; *www.pamf.org/serology*; 650-853-4828; *toxolab@pamf.org*) is required. If *Toxoplasma* infection was acquired after 18 weeks of gestation (when vertical transmission rates become significantly higher) or fetal infection is highly suspected (e.g., ultrasound abnormalities suggestive of congenital toxoplasmosis) or established (positive *Toxoplasma* PCR test result in amniotic fluid), the combination of pyrimethamine, sulfadiazine, and folinic acid is the treatment of choice. None of these drugs should be used alone. Clindamycin could theoretically replace sulfadiazine in patients with significant sulfa drug allergy. (See Treatment section and Table 349-3.)

350 CRYPTOSPORIDIOSIS

ALDO A. M. LIMA AND RICHARD L. GUERRANT

DEFINITION

Cryptosporidiosis is the clinical disease in humans and animals caused by protozoal parasites of the genus *Cryptosporidium* (Apicomplexa). *Cryptosporidium* spp are recognized as major waterborne parasites worldwide, and 14 species of 30 to date have been documented to infect humans. Two named species, *C. hominis* and *C. pestis*, are considered of major public health significance. Four (*C. cuniculus, C. meleagridis, C. viatorum,* and *C. felis*) and eight (*C. parvum, C. fayeri, C. canis, C. suis, C. ubiquitum, C. scrofarum, C. muris,* and *C. andersoni*) of 14 named species are considered of moderate and minor public health significance, respectively. There are nine species of 30 that are shared between humans and cattle.

The Pathogen

The family Cryptosporidiidae has a hidden sporocyst and undergoes monoxenous completion of its cycle in one host, where it can cause predominantly intestinal, cloacal, and gastric infections.

EPIDEMIOLOGY

The life cycle begins with ingestion of *Cryptosporidium* oocysts (2 to 5 μm) by the vertebrate host, with subsequent excystation within the lumen of the small intestine to release four sporozoites (Fig. 350-1). The sporozoites attach to and enter the host's epithelial cells to form intracellular but extracytoplasmic parasitophorous vacuoles, where they develop into trophozoites and subsequently type 1 meronts (schizonts). By asexual nuclear division, they multiply and release six to eight type 1 merozoites that invade neighboring host cells and develop into type 2 meronts, or trophozoites, to complete the asexual reproductive cycle. Type 2 meronts undergo two nuclear divisions and release four type 2 merozoites that can infect the host's cells and further develop into male (microgamont) or female (macrogamete) forms. Microgametes released from the microgamont can penetrate the macrogametes to form zygotes. Approximately 20% of the zygotes develop into thin-walled autoinfectious oocysts; some 80% become thick-walled oocysts, which are excreted in stool.

Cryptosporidium oocysts have at least five characteristics that make this organism a common problem and that help define the potential risk for person-to-person spread and for waterborne and food-borne disease outbreaks. First, *Cryptosporidium* oocysts are resistant to many chemical disinfectants, such as chlorine. Second, the organism is highly infectious, with the median *C. parvum* infectious dose being 132 or fewer oocysts. Third, the size of the oocysts, 2 to 5 μm, allows them to pass through many conventional filters. Fourth, the monoxenous life cycle allows infectious oocysts to be excreted in large numbers in feces, which can easily spread. Fifth, the organism is associated with geographic, seasonal, and socioeconomic differences in the distribution of *Cryptosporidium* spp.

Cryptosporidiosis is seasonal and related to precipitation and temperature fluctuations worldwide.[1] Excystation of *C. parvum* increases in water temperatures up to 46° C (natural sunlight for 12 hours). Two waterborne outbreaks brought cryptosporidiosis to public health attention. In January 1989, an increased number of cases of cryptosporidiosis were reported in Swindon and Oxfordshire, United Kingdom, with a peak reached in March. Mapping of the addresses of early cases showed associations with water supplies. In late February, the water authority detected *Cryptosporidium* oocysts in samples from the water that supplied the affected areas and in the treated water. Contamination of the raw Thames River water used at three water treatment plants was found, with evidence of pollution by farm effluent. This outbreak resulted in 516 cases and thus ignited public interest and led to an official inquiry. The second outbreak occurred in early spring 1993 in Milwaukee, Wisconsin, and was the largest documented outbreak of waterborne disease ever in the United States, with an estimated 403,000 people having acute watery diarrhea and potentially 112 deaths. Findings from this outbreak indicated that *Cryptosporidium* oocysts passed through the filtration system of one of the city's water treatment plants. Again, this suggested that the water quality standards and detection methods for *Cryptosporidium* were inadequate. Swimming pool contamination, especially of bigger pools, pools with more heterogeneous mixing such as municipal pools, and pools catering to young children (wading pools), is associated with more cases.[2] Food-borne outbreaks have also been reported in association with contaminated apple cider, unpasteurized milk, chicken salad, vegetables, raw produce, and shellfish.[3]

The prevalence of cryptosporidiosis varies by geographic region, with the highest rates seen in developing countries. Most cases of cryptosporidiosis occur in young children, thus suggesting that the host's immune response plays a role in susceptibility. Furthermore, cryptosporidiosis in individuals infected with human immunodeficiency virus (HIV) and in those with acquired immunodeficiency syndrome (AIDS) is persistent and occasionally life-threatening and may involve infections of the hepatobiliary and respiratory tracts in addition to the entire gastrointestinal tract. Only effective antiretroviral therapy with restoration of immune function controls this devastating disease. The prevalence of cryptosporidiosis in patients with AIDS varies from 31% in those with $CD4^+$ cell counts equal or higher than 200/μL to 38% in those with $CD4^+$ cell counts lower than 100/μL.

PATHOBIOLOGY

Cryptosporidium oocysts and sporozoites interact with host cells, including the processes of excystation, gliding motility, attachment, invasion, parasitophorous vacuole formation, intracellular maintenance, and host cell damage.[4] Oocysts of *Cryptosporidium* spp use their cysteine and serine proteases and aminopeptidase for excystation in the upper part of the small bowel and release infective sporozoites that invade the mucosal epithelium and occasionally Peyer's patch M cells, often extending to the terminal ileum and colon. The sporozoites secrete proteins from the apical organelles for locomotion and attachment. In immunocompromised patients, the organisms can be found throughout the gut, biliary tract, pancreas, and respiratory tract. As noted earlier, trophozoites undergo asexual reproduction by merogony to form type 1 and then type 2 meronts. Sporozoites and merozoites are internalized by similar invasion machinery and actin reorganization. Two classes of proteins, mucin-like glycoproteins and thrombospondin-related adhesive proteins, mediate adhesion of the parasite. The parasite uses proteases for the proteolytic processing of surface and apical complex proteins for invasion and for egress from the host cells. Entry into the host's cell occurs within 30 seconds and is dependent on the parasite's actinomyosin cytoskeleton to enter host-derived bimembrane parasitophorous vacuoles in a unique intracellular but extracytoplasmic niche.[5] Dense polymerized actin forms at the base fusion of the host-parasite bimembranes. Invasion of cells of the host leads to displacement of the microvillous border, villous atrophy, blunting and crypt cell hyperplasia, and marked infiltration by lymphocytes, plasma cells, and some neutrophils into the lamina propria, with apoptosis of infected cells and significant alteration in intestinal permeability, as measured by mannitol and lactulose markers. This noninvasive functional test shows both reduced absorptive area and disrupted barrier function, which is even more severe in patients with HIV/AIDS. Upregulation of nuclear factor κB and the pro-inflammatory cascade causes secretory and mildly inflammatory diarrhea. Pro-inflammatory cytokines and potential biomarkers, such as tumor necrosis factor-α, interleukins 1β and 8, and lactoferrin, are significantly increased in murine and human infections. Interleukins 1β and 8 both upregulate cyclooxygenase 2, which results in prostaglandin synthesis in the epithelial cells and production of substance P by the inflammatory cells; these products together decrease net sodium absorption and increase net chloride secretion, which causes the secretory diarrhea often seen with this infection.

Both innate and adaptive immune responses are associated with immunity to cryptosporidiosis. Two chemokines, CXCL-10 and perhaps CXCL-8, are initially implicated in the attraction of pro-inflammatory cells. Activation of

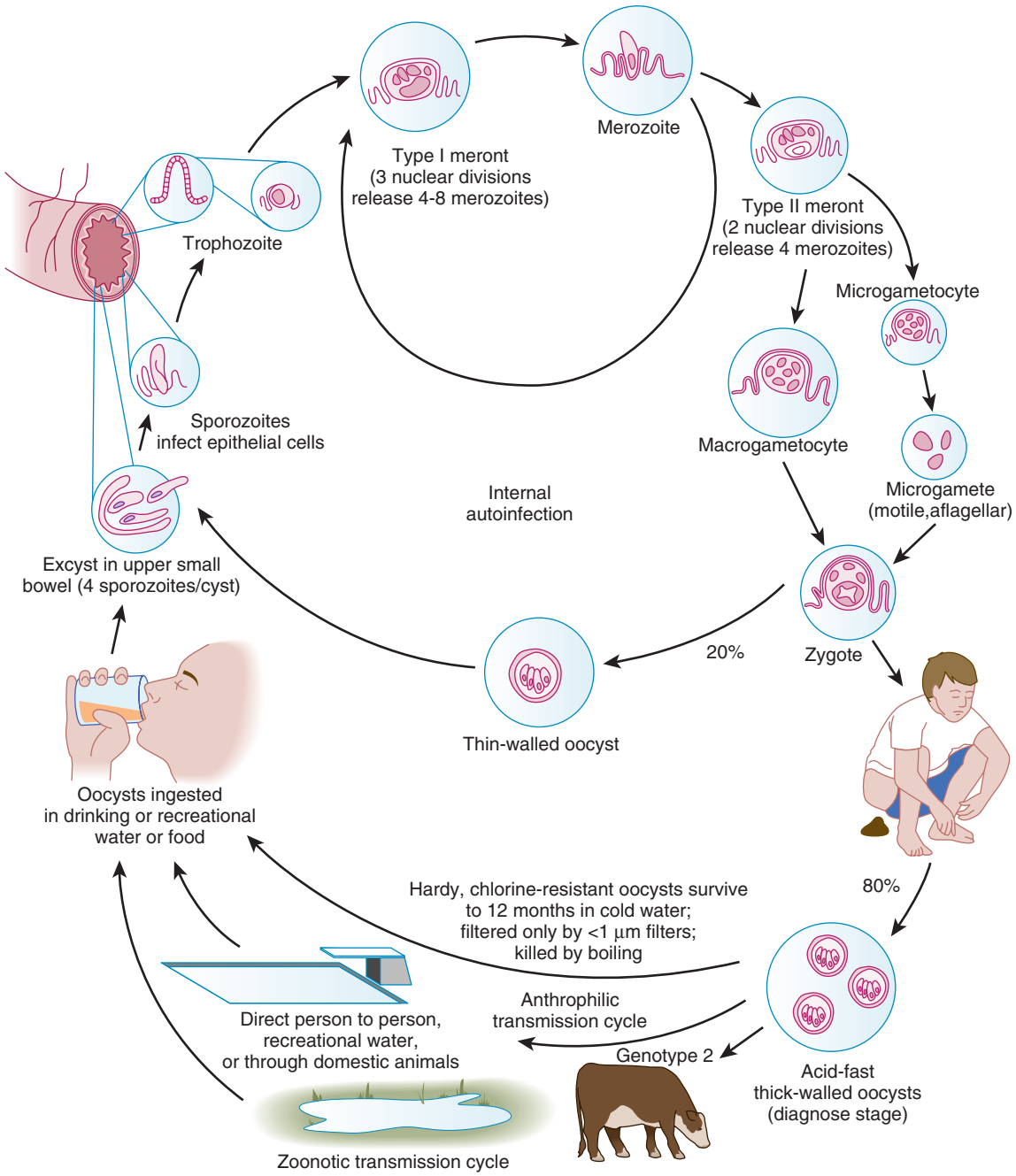

FIGURE 350-1. **Life cycle of *Cryptosporidium* spp. (Redrawn with permission from Kosek M, Alcantara C, Lima AAM, et al. Cryptosporidiosis: an update. *Lancet Infect Dis.* 2001;1:262-269.)**

the immune system involves early stimulation by interleukin 15, which may be important for initial clearance of the parasite. $CD4^+$ T cells and interferon-γ play important roles in the eventual immune defense against this infection, probably more so than B cells or $CD8^+$ cells.

CLINICAL MANIFESTATIONS

Cryptosporidiosis is a cosmopolitan, self-limited infection in immunocompetent hosts that affects all age groups and both sexes. In developing countries, the disease occurs most frequently in children younger than 5 years because of high rates of fecal-oral exposure and the development of partial immunity in older children and adults. In developed countries, the disease occurs at all ages, and most cases are associated with small waterborne outbreaks, frequently with contamination of recreational water. Patients from developed countries are more often adults. The incubation period is approximately 1 week, with a range of 1 to 30 days.[6] The diarrhea can have a sudden onset, frequently with voluminous watery stools, abdominal pain, nausea or vomiting, bloating, malaise, fatigue, weight loss, and occasional fever. In young children at risk in developing countries, it may have a long-term impact on physical activity, school performance, and development of cognitive function. Rarely, respiratory infection and cough have been reported. In normal hosts, the disease usually lasts 1 to 3 weeks but sometimes for more than a month. Oocyst shedding typically lasts 1 to 2 weeks but can occur for up to 2 months. Patients with T-cell immune deficiency, such as those with hematologic malignant neoplasms (particularly children), primary T-cell deficiencies (severe combined immunodeficiency and CD40 ligand deficiency), and HIV/AIDS, are at high risk for the development of more severe disease and have an increased risk for death.

Immunocompetent Host

Data on the natural history of cryptosporidiosis in immunocompetent hosts have been obtained mostly from developed countries, and such data are derived from waterborne outbreaks, patients seeking medical assistance, travelers, animal workers, children in daycare, and their contacts. Most patients in outbreaks and travelers are adults and usually have diarrhea with a median duration of 14 days and a range of 1 to 100 days. The duration and severity of diarrhea appear to be similar for *C. parvum* and *C. hominis* infections. Recurrence of diarrhea is common and occurs in 30 to 41% of patients. Reports from the United Kingdom describe abdominal pain (96%), vomiting (65%), fever (59%), and bloody diarrhea (11%). The clinical manifestations in 285 people surveyed with a confirmed laboratory diagnosis of *Cryptosporidium* infection in the massive outbreak in Milwaukee showed a median duration of 9 days (range, 1 to 55 days), with watery diarrhea in 93%,

abdominal cramps in 84%, fever in 57%, and vomiting in 48%. Patients may continue to shed oocysts for seven months despite being asymptomatic.[7]

Childhood Cryptosporidiosis in Developing Countries

In developing countries, cryptosporidiosis is associated with prolonged (7 to 13 days) or persistent (≥14 days) diarrhea, increased overall burden of diarrhea, risk for malnutrition, and infant mortality. A nested case-control study from a cohort of children in Brazil has shown that children younger than 1 year with cryptosporidiosis have subsequent increased diarrheal burdens and growth shortfalls. These findings have been extended in a shantytown in Peru, where cryptosporidiosis has been associated with growth faltering and stunting. This was also true for children with asymptomatic infections, which are even more common in endemic areas. The long-term impact of childhood diarrhea and cryptosporidiosis has also been evaluated in a cohort study in Brazil, in which it was shown that children with more diarrhea morbidity or cryptosporidiosis early in life (0 to 2 years) have reduced fitness and impaired cognitive function at 6 to 9 years of age. Several studies have now shown that oocysts are shed longer and the number of oocysts is higher with *C. hominis* than with *C. parvum* infection. In Brazil, a birth cohort study demonstrated that children with *C. hominis* infection have higher fecal lactoferrin and delayed growth in comparison to those with *C. parvum* infections. Diarrhea, nausea, vomiting, and general malaise are more frequently associated with *C. hominis* infection. Subtype family analysis of *C. hominis* (Ia, Ib, Id, and Ie) has indicated that Ib is more commonly associated with nausea, vomiting, and general malaise.

Immunocompromised Hosts

Low $CD4^+$ lymphocyte counts in HIV-infected patients are associated with severe cryptosporidial diarrhea. For example, patients with $CD4^+$ counts higher than $180/mm^3$ usually have transient or self-limited disease, whereas those with $CD4^+$ counts lower than $50/mm^3$ often have severe disease with more than 2 L of stool per day. The introduction of highly active antiretroviral therapy (HAART) heralded reduced rates of cryptosporidiosis in patients with HIV/AIDS. HAART may also directly inhibit sporozoite development and invasion in the host's cells. Patients with other immune disorders, such as T-cell immune deficiency (severe combined immunodeficiency and CD40 ligand deficiency) or hematologic malignant neoplasms, particularly children, are at higher risk for more severe, prolonged, or extensive disease, with infection occasionally extending to the gallbladder, the pancreatic duct, and even the bronchial tree. Several complications in these patients have been described, including pancreatitis, cholecystitis, sclerosing cholangitis, papillitis, and terminal bile duct stenosis with subsequent biliary cirrhosis.

DIAGNOSIS

The clinical manifestations and physical examination findings in patients with cryptosporidiosis are not unique, and the differential diagnosis should include other causes of infectious gastroenteritis, such as *Giardia*, *Cyclospora*, *Isospora*, microsporidia, *Escherichia coli* pathotypes (enteropathogenic [EPEC], enteroaggregative [EAEC], diffusely adherent [DAEC], enterohemorrhagic [EHEC], enteroinvasive [EIEC], and enterotoxigenic [ETEC]), *Campylobacter*, *Salmonella*, *Shigella*, rotavirus, norovirus, and others. Definitive diagnosis of *Cryptosporidium* enteric infection is made by stool examination. Up to three fecal samples with fixation and concentration before permanent staining or polymerase chain reaction analysis[8,9] may increase detection rates. Oocysts are stained with the acid-fast stain, modified Ziehl-Nielsen stain, fluorescence staining with auramine-phenol, or immunofluorescent stains. Acid-fast staining requires around 500,000 oocysts per gram for detection in formed stools, whereas immunofluorescence is at least 10 times more sensitive, and commercially available enzyme-linked immunoassays have a sensitivity and specificity approaching 100% for *Cryptosporidium*. Polymerase chain reaction testing can detect as few as 50 to 500 oocysts per milliliter of liquid stool and can be used to differentiate or even to quantify *Cryptosporidium* species and genotypes.

TREATMENT Rx

Nitazoxanide has emerged as the only promising candidate for treatment of cryptosporidiosis. It is licensed in the United States for treatment of cryptosporidiosis in nonimmunodeficient children and adults and has reportedly reduced the duration of diarrhea and oocyst shedding in several double-blind, placebo-controlled clinical trials. A clinical trial in Egypt in adults (500 mg twice daily for 3 days) and children (200 mg for 4- to 11-year-olds and 100 mg for 1- to 3-year-olds, twice daily for 3 days) with cryptosporidiosis showed that 80% exhibited resolution of diarrhea versus 41% of the placebo group.[A1] A second clinical trial of a 3-day course of treatment showed resolution of symptoms at 4 days in 96% of the patients receiving nitazoxanide tablets (500 mg twice daily for 3 days) versus only 41% of those receiving placebo tablets.[A2]

Prevention and treatment of cryptosporidiosis in immunocompromised patients have been reviewed; there has been no evidence of a reduction in the duration or frequency of diarrhea with nitazoxanide. However, nitazoxanide has resulted in significantly greater oocyst clearance than has placebo in all children.[A3] In patients with HIV/AIDS, HAART has emerged as the most important therapy to prevent and to reduce the severity and frequency of cryptosporidiosis. Oral rehydration therapy or intravenous fluid replacement for more severe disease is key to preventing dehydration and risk for immediate death.

PREVENTION

Prevention of person-to-person spread should be achieved with personal hygiene guidelines such as frequent handwashing after using or cleaning the toilet, changing diapers, and caring for a person with diarrhea. It is recommended that people with cryptosporidiosis be excluded from the workplace setting until 48 hours after the last diarrheal episode. Handwashing facilities should be available and used at farms to facilitate personal hygiene. Because most outbreaks of cryptosporidiosis are linked to oocysts in source water, routine survey of treated water for this parasite is key to preventing major spread of this disease. Optimization of multibarrier approaches, including chemical treatment and water filtration and treatment systems, is highly recommended. Ultraviolet irradiation and ozone are effective in inactivating *Cryptosporidium* and *Giardia* cysts in water and may prove useful in controlling the transmission of waterborne protozoa.

Grade A References

A1. Rossignol JF, Ayoub A, Ayers MS. Treatment of diarrhea caused by Cryptosporidium parvum: a prospective randomized, double-blind, placebo-controlled study of nitazoxanide. *J Infect Dis*. 2001;184:103-106.

A2. Rossignol JF, Kabil SM, Younis AM, et al. Effect of nitazoxanide in diarrhea and enteritis caused by Cryptosporidium species. *Clin Gastroenterol Hepatol*. 2006;4:320-324.

A3. Abubakar I, Aliyu SH, Arumugam C, et al. Treatment of cryptosporidiosis in immunocompromised individuals: systemic review and meta-analysis. *Br J Clin Pharmacol*. 2007;63:387-393.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Lal A, Hales S, French N, et al. Seasonality in human zoonotic enteric diseases: a systematic review. *PLoS ONE*. 2012;7:e31883.
2. Polgreen PM, Sparks JD, Polgreen LA, et al. A statewide outbreak of *Cryptosporidium* and its association with the distribution of public swimming pools. *Epidemiol Infect*. 2012;140:1439-1445.
3. Yoder JS, Beach MJ. *Cryptosporidium* surveillance and risk factors in the United States. *Exp Parasitol*. 2010;1244:31-39.
4. Bouzid M, Hunter PR, Chalmers RM, Tyler KM. *Cryptosporidium* pathogenicity and virulence. *Clin Microbiol Rev*. 2013;26:115-134.
5. Lendner M, Daugschies A. Cryptosporidium infections: molecular advances. *Parasitology*. 2014; 141:1511-1532.
6. Chalmers RM, Davies AP. Minireview: clinical cryptosporidiosis. *Exp Parasitol*. 2010;124: 138-146.
7. Chappell CL, Okhuysen PC, Langer-Curry RC, et al. Cryptosporidium muris: infectivity and illness in health adult volunteers. *Am J Trop Med Hyg*. 2015;92:50-55.
8. Mary C, Chapey E, Dutoit E, et al; ANOFEL *Cryptosporidium* National Network. Multicentric evaluation of a new real-time PCR assay for quantification of *Cryptosporidium* spp. and identification of *Cryptosporidium parvum* and *Cryptosporidium hominis*. *J Clin Microbiol*. 2013;51:2556-2563.
9. Checkley W, White AC Jr, Jaganath D, et al. A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for cryptosporidium. *Lancet Infect Dis*. 2015;15:85-94.

REVIEW QUESTIONS

1. *Cryptosporidium* infection

A. May be asymptomatic but still impair a child's growth
B. Produces solid protective immunity
C. Is readily treated with broad-spectrum antimicrobial agents
D. Is primarily a food-borne pathogen
E. Can be diagnosed with an iodine stain for parasites

Answer: A Asymptomatic infections can impair growth, even without causing overt diarrheal symptoms. In fact, because they are more common, the total stunting impact of asymptomatic infections can exceed that of the symptomatic ones. Furthermore, recurrent infection is common, especially where water and environmental fecal contamination is often seen, and it is difficult to treat, especially in the immunocompromised host. Finally, diagnosis requires special acid-fast or fluorescence antibody staining, enzyme-linked immunosorbent assay, or polymerase chain reaction analysis as routine iodine stains for ova and parasites typically do not identify this protozoan parasite.

2. Cryptosporidiosis

A. Usually causes bloody diarrhea in immunocompromised patients
B. Cannot be spread by direct contact because it requires maturation outside the host
C. Is a major cause of moderate to severe diarrhea in developing countries
D. Does not cause relapsing disease
E. Is an acute disease

Answer: C *Cryptosporidium* can cause severe, watery, cholera-like diarrhea in immunocompromised patients, such as patients with AIDS. Because the infectious dose is low and sporozoites develop into the mature oocysts in the infected host, they are infectious when shed in the feces (unlike *Cyclospora*, which do require maturation outside the host). In the recent GEMS study in seven sites in Africa and Asia, *Cryptosporidium* was second only to rotaviruses as an attributable cause of moderate to severe diarrhea that can often recur. (Kotloff KL, Nataro JP, Blackwelder WC, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries [the Global Enteric Multicenter Study, GEMS]: a prospective, case-control study. *Lancet*. 2013;382:209-222.)

3. The risk of acquiring *Cryptosporidium* infections is

A. Reduced by chlorination of water
B. Greater in immunocompromised patients
C. Reduced by filters that remove particles larger than 6 μm
D. Requires ingestion of more than 10^5 parasite oocysts
E. Greater in adults

Answer: B *Cryptosporidium* is a hardy, chlorine-resistant oocyst that can be infectious with doses as low as 150 or fewer oocysts and probably even lower doses in immunocompromised individuals. It is also able to pass through common filters unless they are definitely smaller than 1-μm filter size.

4. The major outbreaks of cryptosporidiosis were associated with

A. Ingestion of contaminated drinking water
B. Ingestion of contaminated recreational water
C. Exposure to infected animals
D. Ingestion of contaminated food
E. Contact with infected persons or animals

Answer: A *Cryptosporidium* infection is transmitted by the fecal-oral route and results from the ingestion of *Cryptosporidium* oocysts through the consumption of fecally contaminated food or water (drinking and recreational water) or through direct person-to-person or animal-to-person contact. The major outbreaks of cryptosporidiosis were associated with fecally contaminated drinking water.

5. Which of the following is true regarding the treatment or prevention of cryptosporidiosis?

A. Nitazoxanide reduces the shedding of oocysts in immunocompetent hosts.
B. Disinfection by chlorination prevents cryptosporidiosis.
C. Nitazoxanide cures cryptosporidiosis in immunocompromised patients.
D. Filtration is an efficient method to prevent cryptosporidiosis.
E. Ozone and ultraviolet irradiation are not adequate to prevent cryptosporidiosis.

Answer: A Nitazoxanide has emerged as the drug of choice for treatment of cryptosporidiosis in immunocompetent children and adults. In immunocompromised patients, treatment or prevention with nitazoxanide shows no evidence of a reduction in the duration or frequency of diarrhea. Chlorination alone is ineffective against *Cryptosporidium* oocysts. Ultraviolet irradiation and ozone are effective in inactivating *Cryptosporidium* and *Giardia* cysts in water and can be used to prevent transmission of waterborne protozoa.

351

GIARDIASIS

THEODORE E. NASH AND DAVID R. HILL

DEFINITION

Giardia lamblia (Giardia duodenalis, Giardia intestinalis) is a ubiquitous, small intestinal protozoan parasite of humans and other mammals. It is the most common parasitic infection of the gastrointestinal tract in the United States as well as worldwide and is responsible for outbreaks of diarrhea and sporadic endemic disease.[1,2]

The Pathogen

Giardia has a simple life cycle. The trophozoite, which is 9 to 21 µm long, 5 to 15 µm wide, and 2 to 4 µm thick (Fig. 351-1), resides in the small intestine and is responsible for manifestations of disease. It has four pairs of flagella, two nuclei, and a ventral sucking disc by which it may adhere to intestinal epithelial cells. The dorsal surface is pear shaped and bilaterally symmetrical, with the two highly characteristic nuclei best visualized after staining. In the lower small intestine, the trophozoite develops into an environmentally resistant cyst. Detection of soluble cyst wall proteins in the feces forms the basis of many stool antigen assays.

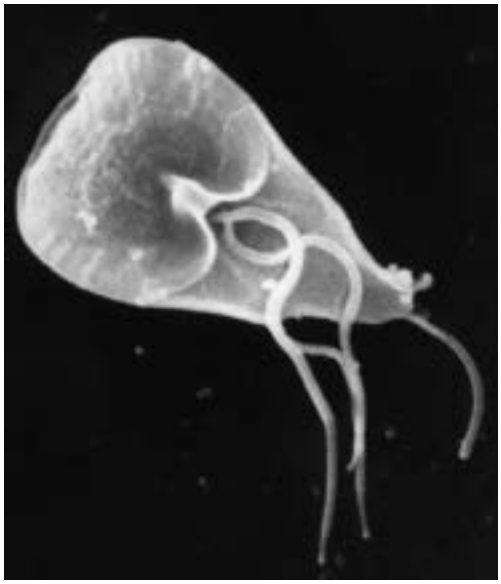

FIGURE 351-1. ***Giardia lamblia*.** This scanning electron micrograph reveals some of the external ultrastructural details displayed by a flagellated *G. lamblia* protozoan parasite.

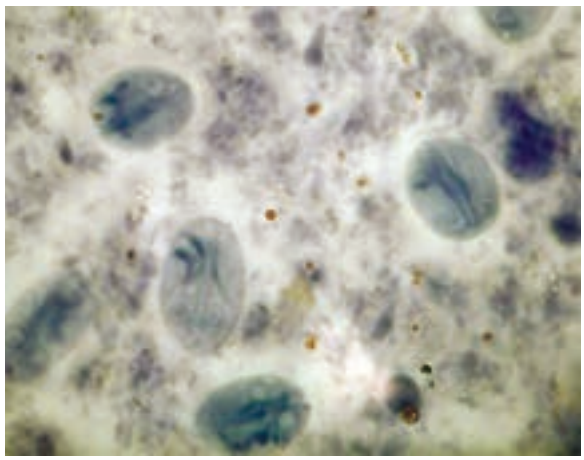

FIGURE 351-2. **Iron hematoxylin stain of *Giardia lamblia* cysts from stool.**

Excreted cysts are mature and infectious. They are oval and about 8 to 12 µm in length and 7 to 10 µm in width (Fig. 351-2). After ingestion and exposure to acid and proteases in the stomach and intestines, they excyst in the small intestine, yielding two trophozoites from each cyst, which quickly divide again. In vitro trophozoites double in number every 6 hours for the fastest growing isolates.

A number of morphologically identical but genetically distinct *Giardia* infect humans and animals that are now divided into eight assemblages. Humans and some animals are infected with either assemblage A or B. These two assemblages are genetically and biologically diverse and appear to be two separate species.[3]

Giardia is well adapted to its existence as a parasite. It has two equal functioning nuclei and lacks mitochondria and peroxisomes. It has a simplified metabolism and is dependent on the host for nutrients such as purines, pyrimidines, cysteine, and cholesterol. The WB isolate, assemblage A, has a compact genome (11.7 Mb) with unusually short promoters. The parasite's rigid cytoskeleton is composed of unique families of structural proteins and carbohydrates.

Giardia is the only bowel-dwelling parasite that undergoes antigenic variation. Only one of a family of about 250 variant-specific proteins (VSPs) is expressed on the surface of the trophozoite at a time.[4] Both immune and biologic selection of trophozoites that express specific VSPs occurs in humans and animals with giardiasis. Expressed VSPs must be compatible with the host's intestinal environment and are probably not recognized by the host's immune system because antibodies to VSPs are inhibitory or cytotoxic. Whereas all VSPs are transcribed, all but one of the transcripts are eliminated by RNA interference–based mechanisms, resulting in expression of a single VSP surface protein.[5] Exactly how selection and switching occur is unclear.

EPIDEMIOLOGY

Giardia is among the most common parasitic infections of humans; it is highly infectious, and cysts are frequently excreted in large numbers (as high as 10^7 cysts per gram of feces), especially in young children. Cysts can survive for months in cold water, are relatively resistant to chlorination, but are intolerant of desiccation and heat in comparison to the relatively resistant ova of cryptosporidia and helminths. Experimentally, between 10 and 100 cysts is sufficient to establish infection 100% of the time. Consequently, ingestion of water or food that contains small levels of contamination can result in infection. Approximately 20,000 infections are reported annually in the United States, but because of underreporting, actual infections are estimated at more than 100,000 cases per year.

Infections are most common in young children and are more frequent in summer and fall months. Giardiasis is acquired after ingestion of contaminated water or food or through person-to-person contact. In past decades, large outbreaks in high-income countries such as the United States occurred after ingestion of contaminated drinking water obtained from surface sources such as reservoirs, lakes, and mountain streams. However, with improved water treatment measures, ingestion of contaminated recreational water from pools or lakes is now a more common source of outbreaks. Backpackers who ingest untreated surface water remain at risk of infection.

Although outbreaks from contaminated food or infected food handlers are well described, they are not often documented. Worldwide, person-to-person transmission may be the most frequent mode of infection and is the primary way children are infected in daycare centers, where infection can be common. Person-to-person spread occurs among family members with infected children and following sexual practices that lead to fecal-oral contact.[6] In low-income, highly endemic regions, almost all children are infected by 2 to 3 years of age. Although partial immunity can develop in previously exposed adults, reinfection of children after treatment is common in highly endemic regions. Longer-term travel, particularly to south Asia, increases the risk of giardiasis. For the returned traveler with persistent diarrhea, giardiasis should be ruled out.

The understanding of immunity in humans is based mostly on animal models that have limited applicability to human infection and disease. In addition, some of the findings are conflicting. In the classic human experimental infections reported by Rendtorff and colleagues in the 1950s, self-cure occurred in about 84% of the persons. In a more recent experimental human challenge study, in which the infecting parasite was well characterized and the inoculum known, rechallenge with the same isolate after treatment resulted in brief, asymptomatic infections in two persons, suggesting development of partial immunity. Humans with hypogammaglobulinemia are susceptible to *Giardia* and have more severe disease that is resistant to treatment. Studies in animals support a major role of intestinal antibodies (particularly IgA) in protective immunity and an essential role of T-cell immunity in the control of *G. lamblia* infection.

Giardia infections are neither more severe nor more common in most other immunosuppressed states and in persons with selective IgA deficiency. Even though most human immunodeficiency virus–infected patients respond to usual treatment, a subset develop recurrent or repeated infections that are difficult to treat.

PATHOBIOLOGY

Giardia infections involve complex interactions between host and parasite. The two assemblages (A and B) that infect humans are composed of genetically distinguishable isolates, which may vary in infectivity, antigenicity, and virulence. In addition, human hosts vary in susceptibility to infection and disease and in the response or tolerance to infection. Pathogenic mechanisms need to explain the varied clinical manifestations as well as the contrasting situation in which there are high rates of infection and disease seen in waterborne outbreaks of giardiasis in regions where infections are sporadic, like in the United States, compared with the large number of asymptomatic infections in children in low-income regions. In addition, infection with *Giardia* can lead to malabsorption, weight loss, and nutritional deficiencies in some settings and little effect on nutritional parameters in other settings.

Giardia is strictly an intraluminal parasite that adheres to the epithelium by way of an adhesive or sucking disc. Invasion of the epithelium either does

not occur or is rare. The number of trophozoites in the intestine can be so large that adherent organisms cover much of the epithelial surface. This could disrupt the epithelial brush border and contribute to disaccharidase deficiency seen in some patients. A number of studies demonstrate direct epithelial cell barrier dysfunction in vitro and in vivo in humans. Exactly how this occurs is not known. There is no evidence of production of a classic enterotoxin, although it is possible that secreted or surface proteins may be injurious to cells. Of patients with persistent giardiasis after treatment, nearly half exhibited inflammation on small bowel biopsy specimens, supporting the view that chronic inflammatory responses contribute to disease at least in this subset.

CLINICAL MANIFESTATIONS

The clinical manifestations, course, and duration of *Giardia* infections are variable. Infections may be self-limited or persistent, asymptomatic or symptomatic. In general, patients are not as sick as those with bacterial diarrheas. Acute disease manifestations occur commonly in travelers and in outbreaks and are characterized by diarrhea, nausea, anorexia, dehydration, flatulence, eructation, foul-smelling stools, distention, abdominal cramping, and weight loss. Malabsorption is more commonly seen in chronic infection. Fever and vomiting are uncommon. Blood, mucus, and polymorphonuclear cells in the stool, which are not usual features of small bowel infections, should suggest an alternative or additional diagnosis. Dehydration, although uncommon, may be severe and require hospitalization; hospitalization for giardiasis has occurred in the United States with a frequency similar to that for shigellosis. On occasion, nausea and vomiting will predominate and suggest other causes.

In experimental infections using inoculated cysts, presence of cysts in the stool occurred 6 to 15 days (mean, 9 days) after inoculation. In more recent experimental infections, *Giardia* antigen was detected 1 day before cyst excretion. In one well-documented food-borne outbreak, 74% became ill, with an incubation period of 2 to 19 days and peak symptoms at 5 to 6 days. Symptoms continued for a median of 18 days.

Acute symptoms can resolve, wax and wane, or settle into a chronic phase, which can be prolonged and last weeks to months. Long-lasting symptoms should indicate a search for the parasite. Lactose deficiency is common and can persist for some weeks after treatment, and it needs to be distinguished in symptomatic patients from relapse or reinfection. In extreme cases, malabsorption and weight loss are severe and mimic sprue. Infants and children are particularly susceptible to infection and disease, which can result in growth failure that is reversed on successful treatment.

A typical scenario is a mildly to moderately ill person who complains of an increased number of urgent loose stools, with flatus, cramping, anorexia, and weight loss. There may be periods when the person feels better only to relapse and become noticeably worse. After days to several weeks, the individual will seek medical help. Similar to other causes of infectious diarrheas, symptoms can continue after successful treatment and evolve into irritable bowel syndrome (Chapter 136).[7] Uncommonly, *Giardia* is found in biliary and pancreatic ducts and can cause cholecystitis and pancreatitis. Extraintestinal manifestations and long-term consequences are usually uncommon, but a series of sporadic cases documented them in a third of the patients.[8] The manifestations can include rash, reactive arthritis, eye complaints, and cognitive deficiencies.[9]

Disease and symptoms caused by *Giardia* in children in low-income regions are variably present despite almost universal infection by the age of 3 years and frequent reinfection with prevalence rates that are commonly above 20%. Although there are convincing studies showing detrimental growth and nutritional effects in some populations, other studies show little or no effects. An analysis of published studies indicated that persistent diarrhea as a result of infection with *Giardia* is a cause of malnutrition.[10] The reasons for these disparate results may be due to differences in the populations, such as prior exposures, maternal factors, nutritional status, and diet, or variability in the organism, including the genetics of the isolate (e.g., assemblage type) and properties of the expressed VSP. In contrast, *Giardia*-naïve visitors frequently develop symptomatic giardiasis while visiting or working in highly endemic regions, in contrast to a mostly asymptomatic population, suggesting that the endemic population has developed an accommodation to the disease. Some studies suggest giardiasis protects against other acute diarrhea episodes.[11]

DIAGNOSIS

The diagnosis of giardiasis is based on detection of cysts, trophozoites, or, more recently, parasite-specific antigens in fecal samples. Polymerase chain reaction has been largely experimental but is being increasingly used in field and laboratory settings. Polymerase chain reaction is sensitive and specific and can determine the infecting assemblages.[12] Because excretion of cysts may be variable or in low concentrations, a single examination for ova and parasites is only 50 to 80% sensitive, and two or three examinations may be necessary. Stool antigen tests are standard in most laboratories and are highly sensitive (>90%), specific (close to 100%), and relatively inexpensive and do not require a trained microscopist. Examination of small intestine biopsy specimens or intestinal contents for trophozoites was the previous "gold standard" for diagnosis but is now uncommonly needed to establish or to confirm the diagnosis. In low-intensity infections, all testing methods can be falsely negative and require repeated testing.

The laboratory findings are nonspecific. The white blood cell count and liver function test results are usually normal. Electrolyte disturbances can be present if diarrhea and vomiting are severe. White blood cells, lactoferrin, blood, and mucus are not found in the stool. Immunoglobulin levels are usually normal but abnormally low or absent in susceptible hypogammaglobulinemic individuals.

TREATMENT

Rx

Details of antiparasitic therapy in general are provided in Chapter 344. Tinidazole (Tindamax), a Food and Drug Administration (FDA)–approved nitroimidazole drug similar to metronidazole (Flagyl), has been the treatment of choice; other nitroimidazoles (e.g., ornidazole and secnidazole) that are not approved in the United States are also effective. Tinidazole is given as a single dose, and compared with metronidazole, it has fewer side effects and greater efficacy. In adults, the dose is 2 g orally; in children, the dose is 50 mg/kg as a single dose with a maximum of 2 g. Metronidazole has been used to treat giardiasis for decades but has never been approved by the FDA for this indication; it requires multiple dosing at 250 mg orally three times a day for 5 to 7 days for adults and 15 mg/kg/day in three divided doses for 5 to 7 days for children. Gastrointestinal side effects of metronidazole are relatively common, and alcohol should not be taken concomitantly because of the possibility of a disulfiram reaction with both drugs. A meta-analysis indicated that albendazole (400 mg/day for 5 days), not presently approved by the FDA for treatment of giardiasis, has efficacy similar to that of metronidazole and fewer side effects. However, many studies test for efficacy shortly after therapy is stopped, so recurrent infection cannot be adequately determined, and there is only limited experience with albendazole. Nitazoxanide is a drug with broad activity against protozoa, helminths, and bacteria and is FDA approved for the treatment of giardiasis. It is given at a dose of 100 mg orally every 12 hours for 3 days for children aged 12 months to 3 years, 200 mg orally every 12 hours for 3 days for children 4 to 11 years of age, and 500 mg orally every 12 hours for 3 days for persons older than 12 years. Because it is available in a liquid suspension as well as in a 500-mg tablet, it may be easier to give to young children. It should be given with food. Most of the side effects are gastrointestinal symptoms and headache.

Paromomycin, a nonabsorbable aminoglycoside, has activity against *Giardia* and has been used in pregnant women to avoid the theoretical fetal adverse events of nitroimidazoles, particularly during the first trimester. It is given to adults at 500 mg three times a day for 5 to 10 days and to children at 25 to 35 mg/kg/day orally in three divided doses for 5 to 10 days. Quinacrine and furazolidone (FDA approved but not usually available) are also active against *Giardia* but should be reserved for use in particular situations.

Patients usually experience relief of symptoms on treatment. Failure of treatment is frequently heralded by a return of symptoms days to weeks after cessation of therapy and requires either re-treatment with an alternative class of drug or an increased dosing of the initial therapy. Clinically resistant cases usually respond to combination treatment; quinacrine and metronidazole have been most effective, with metronidazole plus albendazole also showing efficacy.[13]

PREVENTION

Infection is prevented by scrupulous personal hygiene, proper disposal of sewage, removal or killing of cysts from water supplies, and preventing contamination of food and water. Cysts are relatively labile and are susceptible to heating and filtration with small water volume filters of 0.2 to 1 μm. Heating (bringing water to a boil) is preferred because other pathogens found in feces are also inactivated. Cysts are not reliably susceptible to chlorination because the concentrations of chlorine, water temperatures, turbidity, and pH present when treating commercial water supplies are subopitmal. Four drops of 5.25% bleach to 1 liter for 1 hour at room temperature is sufficient for killing. At lower temperatures, inactivation may not be complete.

Grade A References

A1. Escobedo AA, Alvarez G, Gonzalez ME, et al. The treatment of giardiasis in children: single-dose tinidazole compared with 3 days of nitazoxanide. *Ann Trop Med Parasitol*. 2008;102:199-207.

A2. Pasupulet V, Escobedo AA, Despande A, et al. Efficacy of 5-nitromidazoles for the treatment of giardiasis: a systematic review of randomized controlled trials. *PLoS Negl Trop Dis*. 2014;8:e2733.

A3. Solaymani-Mohammadi S, Genkinger JM, Loffredo CA, et al. A meta-analysis of the effectiveness of albendazole compared with metronidazole as treatments for infections with *Giardia duodenalis*. *PLoS Negl Trop Dis*. 2010;4:e682.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Plutzer J, Ongerth J, Karanis P. *Giardia* taxonomy, phylogeny and epidemiology: facts and open questions. *Int J Hyg Environ Health*. 2010;213:321-333.
2. Brunkard JM, Ailes E, Roberts VA, et al. Surveillance for waterborne disease outbreaks associated with drinking water—United States, 2007-2008. *MMWR Surveill Summ*. 2011;60:38-68.
3. Feng YY, Xiao LH. Zoonotic potential and molecular epidemiology of *Giardia* species and giardiasis. *Clin Microbiol Rev*. 2011;24:110-140.
4. Adam RD, Nigam A, Seshadri V, et al. The *Giardia lamblia* vsp gene repertoire: characteristics, genomic organization, and evolution. *BMC Genomics*. 2010;11:424.
5. Prucca CG, Rivero FD, Lujan HD. Regulation of antigenic variation in *Giardia lamblia*. *Annu Rev Microbiol*. 2011;65:611-630.
6. Escobedo AA, Almirall P, Alfonso M, et al. Sexual transmission of giardiasis: a neglected route of spread? *Acta Trop*. 2014;132:106-111.
7. Hanevik K, Wensaas KA, Rortveit G, et al. Irritable bowel syndrome and chronic fatigue 6 years after giardia infection: a controlled prospective cohort study. *Clin Infect Dis*. 2014;59:1394-1400.
8. Cantey PT, Roy S, Lee B, et al. Study of nonoutbreak giardiasis: novel findings and implications for research. *Am J Med*. 2011;124:1175.e1-1175.e8.
9. Halliez MC, Buret AG. Extra-intestinal and long term consequences of *Giardia duodenalis* infections. *World J Gastroenterol*. 2013;19:8974-8985.
10. Muhsen K, Levine MM. A systematic review and meta-analysis of the association between *Giardia lamblia* and endemic pediatric diarrhea in developing countries. *Clin Infect Dis*. 2012;55: S271-S293.
11. Muhsen K, Cohen D, Levine MM. Can *Giardia lamblia* infection lower the risk of acute diarrhea among preschool children? *J Trop Pediatr*. 2014;60:99-103.
12. Vanni I, Caccio SM, van Lith L, et al. Detection of *Giardia duodenalis* assemblages A and B in human feces by simple, assemblage-specific PCR assays. *PLoS Negl Trop Dis*. 2012;6:e1776.
13. Watkins RR, Eckmann L. Treatment of giardiasis: current status and future directions. *Curr Infect Dis Rep*. 2014;16:396.

REVIEW QUESTIONS

1. A 24-year-old woman spends 4 days in Cancun and within 2 days of return to the United States develops diarrhea consisting of multiple episodes per day of unformed stool. The patient is normally well but 2 years ago was diagnosed with giardiasis. As her symptoms are now similar, she seeks evaluation for it again. The health care provider orders a stool examination for ova and parasites. Which is the correct answer?

A. Giardiasis is the likely diagnosis because the symptoms are similar.
B. Ordering a stool examination for ova and parasites is the best way to establish the diagnosis for her symptoms.
C. The diagnosis of giardiasis is unlikely.
D. Amebiasis is more common than giardiasis in travelers to Mexico.
E. None of the above

Answer: C Giardiasis is unlikely. First, the incubation period is too short for giardiasis, amebiasis, and other gastrointestinal parasites. Second, most diarrhea episodes of travelers to Mexico are due to bacterial infections (e.g., *Escherichia coli*, *Campylobacter*, and *Salmonella*). If she is sufficiently unwell, evaluation should also be directed at these causes or empirical antibiotic treatment considered. Unless giardiasis is confirmed, treatment should not be initiated for this.

2. A 25-year-old man developed diarrhea and weight loss. He was diagnosed with giardiasis after an evaluation for diarrhea that also included stool culture and testing for *Clostridium difficile*, as he has a history of chronic sinusitis that has been intermittently treated with antibiotics. He was treated for *Giardia* with metronidazole and initially responded. However, his symptoms have returned, and he has again been diagnosed with giardiasis. What is the most appropriate step in his evaluation?

A. Repeat evaluation for *C. difficile*.
B. Re-treat with a second course of metronidazole.
C. The *Giardia* is resistant and should be tested for drug sensitivities.
D. Order quantitative immunoglobulin levels.
E. None of the above.

Answer: D The patient has recurrent giardiasis, weight loss, and chronic sinusitis, which is consistent with a diagnosis of common variable immunoglobulin deficiency. Giardiasis in these patients is difficult to treat and recurrent. Although his diarrhea could be due to several causes, *C. difficile* is not likely because it was not diagnosed initially and would not have occurred while the patient was taking metronidazole. Another course of treatment is not unreasonable, but in general, re-treatment with the same drug fails in this situation.

3. A mother of 2- and 7-year-old children develops severe diarrhea. Five weeks earlier, the family stayed at a campground over the weekend. There was supplied faucet water from a deep well that is protected from surface contamination. Three weeks before becoming ill, both children began attending a new daycare school for the summer. She has had two episodes of dehydration requiring fluid resuscitation. On evaluation, she is found to be infected with *Giardia*. The most likely source of her infection is

A. The camping trip
B. The faucet water
C. Her children
D. The water in the home
E. None of the above

Answer: C In addition to the patient, both of her children were subsequently diagnosed with giardiasis. Frequently, the parents of young children are diagnosed, and infection is then traced to a child who is a carrier or only mildly symptomatic. The entire family may become infected. Surface water from lakes or streams during camping can be a source of exposure compared with a relatively secure water source, such as a faucet at park destinations. Given the possible sources for her infection, the daycare center is most likely.

4. A 45-year-old woman who was diagnosed with and treated successfully for symptomatic giardiasis 6 months ago now returns to your office complaining of listlessness, intermittent abdominal pains, and occasional loose stools. Which is the most likely diagnosis?

A. Lactose intolerance
B. Recurrent giardiasis
C. Bacterial diarrhea
D. Post-giardiasis irritable bowel syndrome

Answer: D Although the symptoms are vague and could be due to recurrent giardiasis, the more likely diagnosis is irritable bowel syndrome secondary to prior symptomatic giardiasis. Irritable bowel syndrome and fatigue were found to occur more frequently than in controls after a large epidemic of giardiasis in Bergen, Norway. Prior giardiasis does not preclude other causes of her symptoms, such as depression. Symptoms and findings of lactose intolerance due to shortening of villi is commonly found during and shortly after giardiasis but would not be expected to be present 6 months after infection.

5. A 10-year-old Filipino child who lives in a rural area of the Philippines is taking part in a study surveying children for enteric parasites. He is found to have *Giardia* in his stool but is asymptomatic and is well nourished. Which of the following is a reason for not treating this child?

A. The child is likely to be reinfected in his home setting.
B. Treatment of giardiasis in children is associated with severe adverse drug events.
C. Treatment is usually unsuccessful in endemic settings.
D. The child is too old for treatment to be helpful.

Answer: A Prevalence of *Giardia* infection can be 20% or more in cross-sectional studies from highly endemic regions. Many of these persons are asymptomatic, and reinfection after treatment can be common, particularly in young children. Although he could be treated, there is debate in the literature about the benefits of treatment of asymptomatic children who are likely to be reinfected.

352

AMEBIASIS

WILLIAM A. PETRI, JR. AND ALDO A.M. LIMA

DEFINITION

Amebiasis is due to infection with the enteric protozoan parasite *Entamoeba histolytica*. Amebiasis can cause asymptomatic colonization, diarrhea, dysentery, and colitis as well as spread extraintestinally to cause liver and rarely brain abscess (Fig. 352-1).

The Pathogen

E. histolytica has a low infectious dose (<100 organisms), is resistant to chlorine, and is environmentally stable. These properties make it a threat to food and water supplies, as the 1998 municipal water outbreak of amebic liver abscess in Tbilisi, Georgia, demonstrated. Its tissue-destructive properties are the reason for the parasite's being named *histolytica*.

EPIDEMIOLOGY

Most *E. histolytica* infections occur in the developing world, including the Indian subcontinent, Southeast Asia, sub-Saharan Africa, and Central and South America, as a result of fecal-oral transmission. A national serologic survey in Mexico demonstrated antibody to *E. histolytica* in 8.4% of the population. In an urban slum of Fortaleza, Brazil, 25% of the population tested carried antibody to *E. histolytica*, and the prevalence in children 6 to 14 years of age was 40%. In Dhaka, Bangladesh, where diarrheal diseases are the leading cause of childhood death, the annual incidence of *E. histolytica* infection in a cohort of preschool children was 40%. The annual incidence of amebic liver abscess was 21 cases per 100,000 inhabitants in Hue City, Vietnam. The best current estimate by the World Health Organization is that *E. histolytica* infection results in 34 to 50 million symptomatic cases each year worldwide and as many as 100,000 deaths.

In the United States, amebiasis is the third most common parasitic infection after giardiasis and cryptosporidiosis (1.2 cases/100,000 U.S. population). Most cases in industrialized countries occur in travelers to and immigrants from endemic regions as well as in institutionalized individuals. In returning travelers, diarrhea is the predominant reason for a patient to visit a physician, and amebiasis is the second most common cause of diarrhea in returning travelers. Previously reported high rates of *E. histolytica* infection in homosexual men in the United States actually reflect a high prevalence of *Entamoeba dispar* infection in this population. In contrast, in Asia, amebiasis is more frequently an initial symptom of human immunodeficiency virus (HIV) infection and acquired immunodeficiency syndrome because of the common risk for acquisition of both HIV infection and amebiasis through the sexual practices of men who have sex with men. The typical patient with an amebic liver abscess in the United States is a 20- to 40-year-old Hispanic male immigrant. Several groups are at increased risk for severe amebiasis, including the very young or old, malnourished persons, pregnant women, and patients treated with corticosteroids.

PATHOBIOLOGY

Killing of host cells is required for invasion of the intestine by the parasite. The process of host cell destruction has been experimentally separated into sequential steps of adherence, contact-dependent cytotoxicity, followed finally by phagocytosis of the host cell corpse. The initial contact of parasite to host is mediated by the parasite's galactose and *N*-acetyl-D-galactosamine (Gal/GalNAc)–specific lectin, which binds to carbohydrate determinants on the host's intestinal epithelium. Human cells die by apoptosis induced by *E. histolytica*, a process that requires attachment of the Gal/GalNAc lectin to a host cell's receptor, as well as the parasite's acid intracellular vesicles, which may serve as delivery vehicles for an amebic pore-forming protein. *E. histolytica* initiates apoptosis in host cells by directly activating the host cell's distal apoptotic machinery; caspase 3 is activated within minutes of *E. histolytica* adherence, a caspase 3 inhibitor blocks *E. histolytica* killing, and caspase 3–deficient or bcl-2–overexpressing mice are resistant to amebic infection. Recognition and ingestion of the apoptotic host cell's corpse are required for colonic infection by the parasite, and multiple ligands and receptors are involved, including the Gal/GalNAc lectin, a phosphatidylserine receptor, serine-rich *E. histolytica* protein, and collectins. After ingestion of the corpse of the host cell, additional parasitic factors participate in invasion into the intestinal mucosa. For example, *E. histolytica* encodes at least 44 cysteine proteinase genes that have been implicated in degradation of colonic mucin glycoproteins; digestion of extracellular matrix, hemoglobin, and villin; and inactivation of interleukin-18 (IL-18).

The innate immune response to amebic infection includes activation of the alternative complement pathway, with C3a and C5a recruiting neutrophils to the site of infection but with amebae resisting killing by the membrane attack complex through the Gal/GalNAc lectin. In the murine model of intestinal amebiasis, innate resistance is conferred by nonhematopoietic cells, thus suggesting importance of the epithelial production of cytokines such as tumor necrosis factor-α (TNF-α), IL-1α, IL-6, IL-8, growth-related oncogene-α (GROα), and granulocyte-macrophage colony-stimulating factor. Neutrophils are the earliest innate cellular immune response (within 1 to 2 days) for both intestinal and hepatic amebiasis. Depletion of neutrophils with anti–Gr-1 neutralizing antibody in a murine model results in exacerbated amebic hepatic and intestinal disease. Macrophages and T lymphocytes are also recruited by day 3 of an infection. Macrophages acquire amebicidal activity after in vitro stimulation with interferon-γ (IFN-γ), TNF-α, or colony-stimulating factor 1. Natural killer cells may be important in part as a source of IFN-γ, as well as infiltrating mast cells for their ability to contribute to the innate immune response by the production of IL-6 and TNF-α.

The acquired immune response reflects the opposing roles of IL-4 and IFN-γ in persistence and clearance of amebic infection, respectively. Inbred mice of the CBA strain are susceptible to intestinal amebiasis and develop a rapid T_H2 phenotypic immune response, and this response is deleterious insofar as inhibition of IL-4 can convert the response to a healing IFN-γ response. Effective acquired immunity in humans is associated with both a systemic IFN-γ and a mucosal IgA response directed at the Gal/GalNAc lectin. Children with mucosal IgA against the Gal/GalNAc lectin were found to have 86% fewer new *E. histolytica* infections in the following year. Similarly, the risk for amebiasis was 50% lower in children who were in the 50th percentile and above for the production of IFN-γ by peripheral blood mononuclear cells stimulated with soluble amebic antigen. In contrast, there is an association between higher TNF-α production and *E. histolytica* diarrhea,[1] thus indicating that there is a fine line between a pro-inflammatory cellular immune response that is protective and one that is disease enhancing.

CLINICAL MANIFESTATIONS

Asymptomatic Intraluminal Amebiasis

The asymptomatic cyst-passing carrier state is the most common type of amebic infection. All *Entamoeba moshkovskii* and *E. dispar* infections and as many as 80% of *E. histolytica* infections are asymptomatic. Asymptomatically infected individuals represent a risk to the community because they are a source of new infections and a risk to themselves because 1 in 10 to 20 colonized individuals progress to symptomatic infection. The host has a bearing on whether infection is asymptomatic in that children heterozygous for the HLA class II DQB1*0601/DRB1*1501 haplotype have been found to be protected from symptomatic infection with amebiasis. In addition, certain genotypes of *E. histolytica* appear to be associated with the propensity for colonization as opposed to invasion.

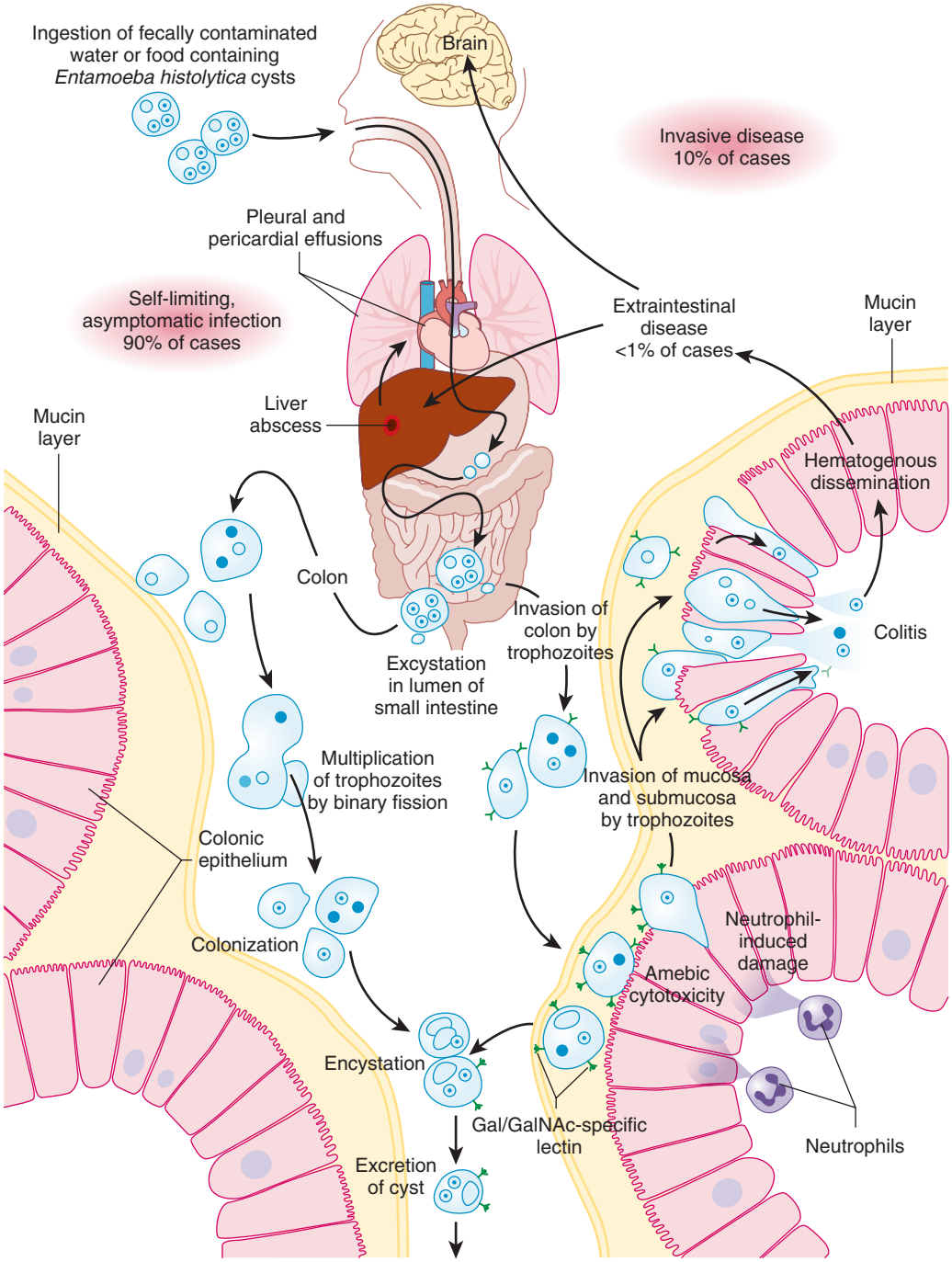

FIGURE 352-1. Life cycle of *Entamoeba histolytica*. Infection is normally initiated by the ingestion of fecally contaminated water or food containing *E. histolytica* cysts. The infective cyst form of the parasite survives passage through the stomach and small intestine. Excystation occurs in the bowel lumen, where motile and potentially invasive trophozoites are formed. In most infections, the trophozoites aggregate in the intestinal mucin layer and form new cysts, which results in a self-limited and asymptomatic infection. In some cases, however, adherence to and lysis of the colonic epithelium, mediated by the galactose and *N*-acetyl-D-galactosamine (Gal/GalNAc)–specific lectin, initiates invasion of the colon by trophozoites. Neutrophils responding to the invasion contribute to cellular protection at the site of invasion. Once the intestinal epithelium is invaded, extraintestinal spread to the peritoneum, liver, and other sites may follow. Factors controlling invasion, as opposed to encystation, probably include parasitic "quorum sensing" signaled by the Gal/GalNAc-specific lectin, interactions of amebae with the bacterial flora of the intestine, and innate and acquired immune responses of the host. (Redrawn with permission from Haque R, Huston CD, Hughes M, et al. Current concepts: amebiasis. *N Engl J Med.* 2003;348:1565-1573.)

Amebic Diarrhea

Amebic diarrhea without dysentery is the most common amebic disease. It is defined as diarrhea in an *E. histolytica*–infected individual. There is no requirement for the presence of mucus or visible or microscopic blood in stool for the diagnosis of amebic diarrhea. In one community-based study of a cohort of preschool children in Bangladesh, the annual incidence of amebic infection, diarrhea, and dysentery was 45%, 9%, and 3%, respectively. The mean duration of amebic diarrhea was 3 days in one study. It causes approximately 2% of cases of diarrhea severe enough to warrant hospital evaluation in developing countries such as Bangladesh.

Amebic Dysentery or Colitis

Diarrhea with mucus or visible or microscopic blood in a patient with *E. histolytica* infection is the definition of amebic dysentery or colitis. Approximately 15 to 33% of patients with *E. histolytica* diarrhea also have amebic dysentery. The onset of symptoms is typically gradual during a period of 3 or 4 weeks after infection, with abdominal tenderness and increasingly severe diarrhea being the primary complaints. Patients with bacterial causes of dysentery usually have only 1 or 2 days of symptoms. Surprisingly, fever is present in only a minority of patients with amebic colitis. In young children, intussusception, perforation, peritonitis, or necrotizing colitis may develop

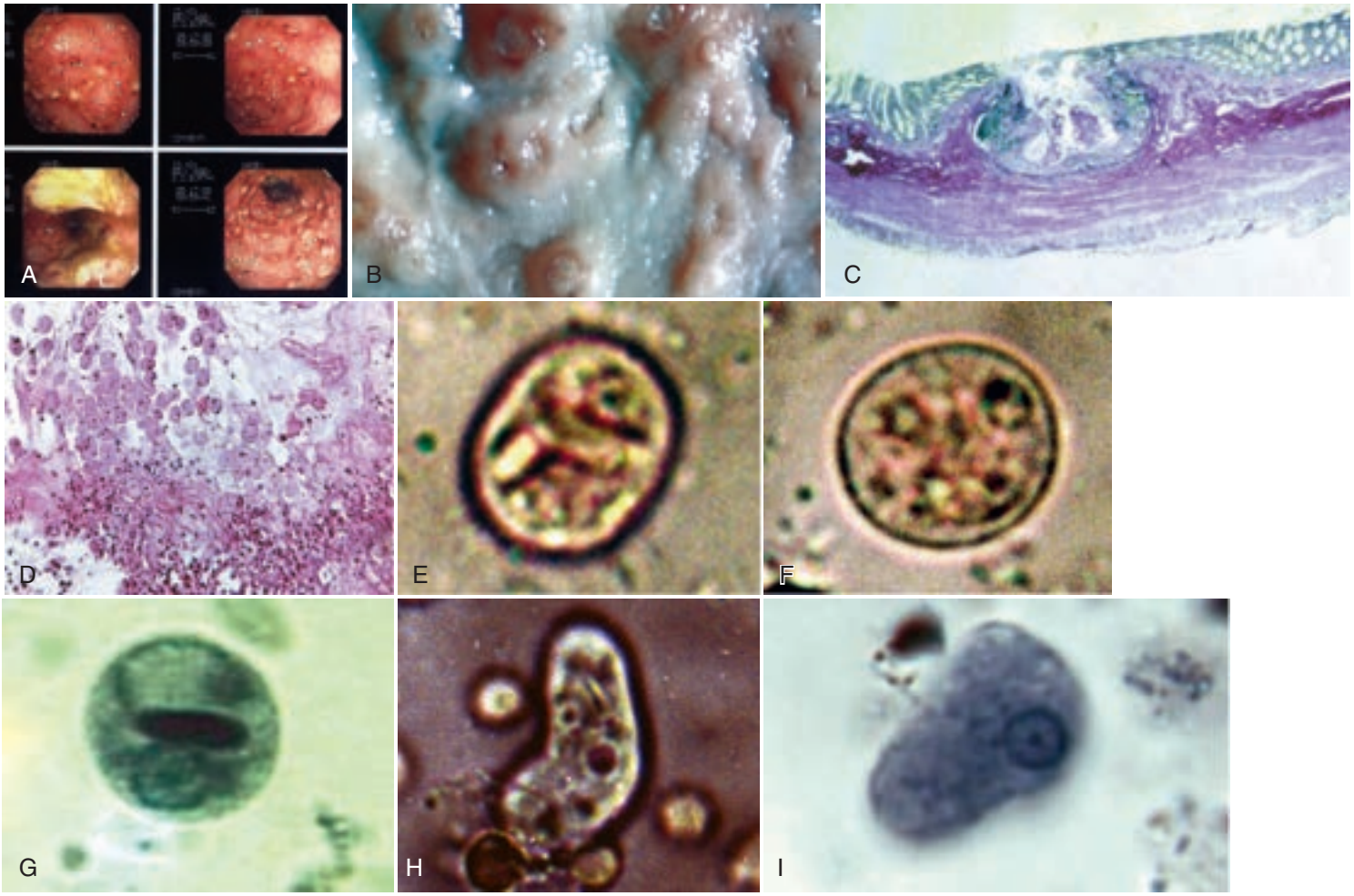

FIGURE 352-2. **Endoscopic and pathologic features of intestinal amebiasis. A,** Colonoscopic appearance of intestinal amebiasis. **B,** Colonic ulcers averaging 1 to 2 mm in diameter on gross pathologic examination. **C,** Cross section of a flask-shaped colonic ulcer (hematoxylin-eosin stain, magnification ×20). **D,** Inflammatory response to intestinal invasion by *Entamoeba histolytica* (hematoxylin-eosin stain, magnification ×100). **E** and **F,** *E. histolytica* cysts in a saline preparation (magnification ×1000). **G,** Iodine-stained cyst from stool (magnification ×1000). **H,** *E. histolytica* trophozoite with an ingested erythrocyte in a saline preparation from stool (magnification ×1000). **I,** Trophozoite from stool stained with trichrome (magnification ×1000). (B-D from the slide collection of the late Dr. Harrison Juniper.) (From Haque R, Huston CD, Hughes M, et al. Current concepts: amebiasis. *N Engl J Med.* 2003;348:1565-1573.)

rapidly (Fig. 352-2). Unusual manifestations of amebic colitis include toxic megacolon (0.5% of cases, usually requiring surgical intervention) and ameboma (granulation tissue in the colonic lumen mimicking colon cancer in appearance).[2]

Amebic Liver Abscess

Amebic liver abscess is 10 times as common in men as in women and is unusual in children. The typical patient with an amebic liver abscess in the United States is an immigrant from an endemic area, a man aged 20 to 40 years with fever, right upper quadrant pain, leukocytosis, abnormal serum transaminase and alkaline phosphatase levels, and a defect seen on hepatic imaging studies.[3] Most patients have 2 to 4 weeks of prior fever, cough, and abdominal pain in the right upper quadrant or epigastrium. Involvement of the diaphragmatic surface of the liver may lead to right-sided pleural pain or referred shoulder pain and an elevated right hemidiaphragm seen on chest radiography (Fig. 352-3). Hepatomegaly with point tenderness over the liver, below the ribs, or in the intercostal spaces is a typical finding.

If a space-filling defect in the liver is observed, the differential diagnosis includes (1) amebiasis (most common in men with a history of travel to or residence in a developing country), (2) pyogenic or bacterial abscess (particularly suspected in women, patients with cholecystitis, the elderly, individuals with diabetes, and patients with jaundice), (3) echinococcal abscess (which would be an incidental finding because echinococcal abscess should not cause pain or fever), and (4) cancer. Most patients with amebic liver abscess will have detectable circulating antigen in serum as well as serum antiamebic antibodies.

In children, abdominal pain is reported infrequently with amebic liver abscess. More commonly, high fever, abdominal distention, irritability, and tachypnea are noted. Some of these children are admitted to the hospital with fever of unknown origin. Hepatomegaly occurs frequently, but elicitation of hepatic tenderness is not well documented. In one report, four of five children younger than 5 years died of amebic liver abscess because the diagnosis was not suspected.

Unusual extraintestinal manifestations of amebiasis include direct extension of the liver abscess to the pleura or pericardium and brain abscess. Death usually results from rupture of the liver abscess into the peritoneum, thorax, or pericardium, but it may also be caused by extensive hepatic damage and liver failure.

Other Extraintestinal Infections

Thoracic amebiasis is the most common type of extra-abdominal amebiasis after liver abscess and occurs in about 10% of patients with amebic liver abscess. It develops by direct extension from the liver. Pericardial amebiasis is the next most common form of extraintestinal involvement and may result from rupture of a liver abscess in the left lobe of the liver into the pericardium or from extension of the right-sided pleural amebiasis. Cerebral amebic abscesses have been found in about 0.5 to 5% of patients with amebic liver abscess. In one series of 18 patients with proven cerebral amebiasis, findings on the initial neurologic examination were normal in 13. Other foci of infection are rare, but amebic rectovesical fistula formation and involvement of the pharynx, heart, aorta, and scapula have been reported. Cutaneous infection may arise from trophozoites emerging from the rectum.

DIAGNOSIS

Diagnosis of amebiasis is best accomplished by the combination of serology and identification of the parasite in feces or at extraintestinal sites of invasion

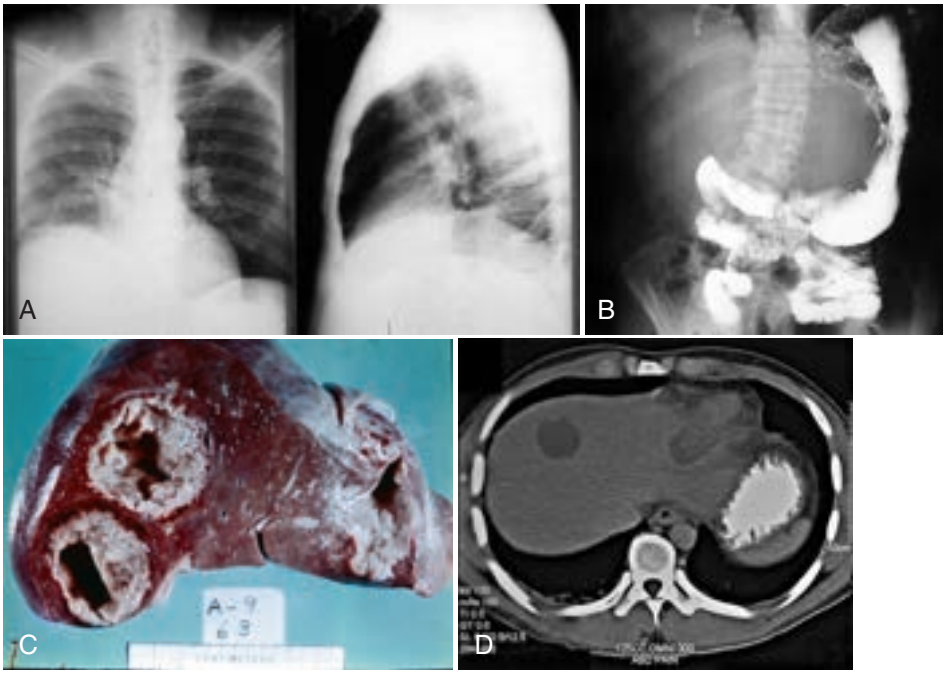

FIGURE 352-3. Radiographic and pathologic features of extraintestinal amebiasis. A, Left posteroanterior and right lateral chest radiographs in a patient with amebic liver abscess. The findings include an elevated right hemidiaphragm and evidence of atelectasis. **B,** Luminal narrowing revealed by barium enema examination in a patient with an ameboma. **C,** Two abscesses in the right lobe and one abscess in the left lobe of a patient with amebic liver abscess. **D,** Abdominal computed tomography showing one abscess in the right lobe and one abscess in the left lobe in a patient with amebic liver abscess. (From Haque R, Huston CD, Hughes M, et al. Current concepts: amebiasis. *N Engl J Med.* 2003;348:1565-1573.)

TABLE 352-1 SENSITIVITY OF TESTS FOR DIAGNOSIS OF AMEBIASIS

TEST	COLITIS	LIVER ABSCESS
Microscopy: stool	25-60%	10-40%
Stool antigen detection	80%	≈40%
Serum antigen detection	65%	>95%
Microscopy: abscess fluid	N/A	≤20%
Real-time PCR	>95%	>95%
Serologic testing (indirect hemagglutination)		
Acute	70%	70-80%
Convalescent	>90%	>90%

N/A = not available; PCR = polymerase chain reaction.
Modified from Haque R, Huston CD, Hughes M, et al. Current concepts: amebiasis. *N Engl J Med.* 2003;348:1565-1573.

(such as pus obtained by fine needle aspiration of a liver abscess).[4] Examination of stool for ova and parasites should *not* be used to diagnose amebiasis (Table 352-1). The most sensitive diagnostic approach is the combined use of *E. histolytica*–specific antigen detection or polymerase chain reaction plus serology.

TREATMENT (also see Chapter 344)

Rx

Therapy for invasive infection differs from that for noninvasive infection, which may be treated with paromomycin (Table 352-2). Invasive infections require treatment with nitroimidazoles, particularly metronidazole, tinidazole, secnidazole, or ornidazole. For amebic colitis, tinidazole reduces treatment failure rates and adverse effects compared with metronidazole. In the rare case of fulminant amebic colitis, it is prudent to add broad-spectrum antibiotics to treat intestinal bacteria that may spill into the peritoneum. Parasites persist in up to half of the patients who are treated with a nitroimidazole, so treatment should be followed with paromomycin or the second-line agent

TABLE 352-2 DRUG THERAPY FOR TREATMENT OF AMEBIASIS

DRUG	ADULT DOSAGE	SIDE EFFECTS
AMEBIC LIVER ABSCESS		
Metronidazole	750 mg PO tid × 10 days	Primarily GI side effects: anorexia, nausea, vomiting, diarrhea, abdominal discomfort, or unpleasant metallic taste Disulfiram-like intolerance reaction to alcoholic beverages Neurotoxicity, including seizures, peripheral neuropathy, dizziness, confusion, irritability
or		
Tinidazole	2 g PO once daily × 5 days	Primarily GI side effects and disulfiram-like intolerance reaction to alcoholic beverages as for metronidazole
Followed by a luminal agent		
Paromomycin	30 mg/kg/day PO in 3 divided doses per day × 5-10 days	Primarily GI side effects: diarrhea, GI upset
or		
Diloxanide furoate	500 mg PO tid × 10 days	Primarily GI side effects: flatulence, nausea, vomiting Pruritus, urticaria
AMEBIC COLITIS		
Metronidazole	750 mg PO tid × 5-10 days	Same as for amebic liver abscess
Plus a luminal agent (same as for amebic liver abscess)		
ASYMPTOMATIC INTESTINAL COLONIZATION		
Treatment with a luminal agent as for amebic liver abscess		

GI = gastrointestinal.
Modified from Haque R, Huston CD, Hughes M, et al. Current concepts: amebiasis. *N Engl J Med.* 2003;348:1565-1573.

TABLE 352-3 FREE-LIVING AMEBAE

ORGANISM	DISEASE	EPIDEMIOLOGY	DIAGNOSIS	CLINICAL COURSE	THERAPY
Naegleria fowleri	Primary amebic encephalitis	Warm freshwater exposure	CSF wet mount for ameba, PCR	Death within 1-2 weeks of onset	Amphotericin B
Acanthamoeba spp	Keratitis	Corneal trauma, usually from contact lens	Corneal scraping for amebae and cysts	Subacute	Polyhexamethylene biguanide, chlorhexidine, propamidine, hexamidine
Acanthamoeba spp	Granulomatous amebic encephalitis	Immunodeficient (organ transplants, HIV/AIDS)	Biopsy of brain or skin abscess—IFA or PCR	Subacute	Combination therapy with pentamidine, an azole (fluconazole or itraconazole), flucytosine, and sulfadiazine
Balamuthia mandrillaris	Granulomatous amebic encephalitis	Immunodeficient but also immunocompetent	Biopsy of brain	Subacute	Combination therapy with flucytosine, pentamidine, fluconazole, and sulfadiazine plus either azithromycin or clarithromycin
Sappinia	Amebic encephalitis	Single patient was not immunodeficient			Azithromycin, pentamidine, itraconazole, flucytosine

AIDS = acquired immunodeficiency syndrome; CSF = cerebrospinal fluid; HIV = human immunodeficiency virus; IFA = indirect fluorescent antibody; PCR = polymerase chain reaction.
Modified from Visvesvara GS, Moura H, Schuster FL. Pathogenic and opportunistic free-living amoebae: *Acanthamoeba* spp, *Balamuthia mandrillaris*, *Naegleria fowleri*, and *Sappinia diploidea*. *FEMS Immunol Med Microbiol*. 2007;50:1-26.

diloxanide furoate to cure luminal infection. High-throughput drug screening has identified auranofin, a Food and Drug Administration–approved drug used for treatment of rheumatoid arthritis, as a potentially active agent against *E. histolytica*. Drainage of a liver abscess should be considered in patients who do not show a clinical response to drug therapy within 5 to 7 days or in those with a high risk for rupture of the abscess, as defined by a cavity with a diameter of greater than 5 cm or by the presence of lesions in the left lobe. Percutaneous needle aspiration or catheter drainage is the procedure of choice for drainage of a liver abscess. Surgical intervention is occasionally required for drainage of a liver abscess, acute abdomen, gastrointestinal bleeding, or toxic megacolon.

PREVENTION

The feasibility of prevention by vaccination with the parasite's Gal/GalNAc lectin is supported by substantial data from human, animal model, and in vitro studies.[5] This vaccine is in the late stages of preclinical development for the prevention of amebiasis in infants and children in the developing world. Provision of sanitation and clean water and safe sexual practices to prevent fecal-oral transmission are of great importance but not universally effective because of the low infectious dose and chlorine resistance of the cyst.

PROGNOSIS

Therapy for amebiasis is highly effective. Drug resistance is not reported.

FREE-LIVING AMEBAE

Rare infections of the central nervous system can be seen with infection by free-living amebae of the genera *Naegleria*, *Balamuthia*, *Acanthamoeba*, and *Sappinia*. *Naegleria fowleri* is the agent of primary amebic meningoencephalitis, which occurs in previously healthy children and young adults who have swum in fresh water 2 to 5 days before the onset of meningoencephalitis. Cerebrospinal fluid has a polymorphonuclear predominance, and motile amebae can be seen in a wet mount of cerebrospinal fluid. The disease is relentlessly progressive to death in most patients. In one case of successful treatment, a combination of intrathecal and intravenous amphotericin B and miconazole and oral rifampin was used. *Acanthamoeba* can cause keratitis[6] in individuals with corneal injuries (usually from contact lens use) as well as granulomatous amebic encephalitis in the immunocompromised. Granulomatous amebic encephalitis can also be caused by *Balamuthia* and *Sappinia*; it is usually associated with focal neurologic findings and has a subacute course (Table 352-3).

Grade A Reference

A1. Gonzales ML, Dans LF, Martinez EG. Antiamoebic drugs for treating amoebic colitis. *Cochrane Database Syst Rev*. 2009;2:CD006085.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Peterson KM, Shu J, Duggal P, et al. The role of TNF-α in susceptibility to amebiasis. *Am J Trop Med Hyg*. 2010;82:620-625.
2. Boopathy V, Alexander T, Balasubramanian P, et al. Amoeboma: resurfacing of a vanishing illness. *BMJ Case Rep*. 2014.
3. Hoque MI, Uddin MS, Sarker AR, et al. Common presentation of amebic liver abscess—a study in a tertiary care hospital in Bangladesh. *Mymensingh Med J*. 2014;23:724-729.
4. Mokhtari M, Kumar PV. Amebic liver abscess: fine needle aspiration diagnosis. *Acta Cytol*. 2014;58:225-228.
5. Quach J, St-Pierre J, Chadee K. The future of vaccine development against *Entamoeba histolytica*. *Hum Vaccin Immunother*. 2014;10:1514-1521.
6. Ross J, Roy SL, Mathers WD, et al. Clinical characteristics of *Acanthamoeba* keratitis infections in 28 states, 2008 to 2011. *Cornea*. 2014;33:161-168.

REVIEW QUESTIONS

1. Individuals at increased risk for amebiasis include all *but*

A. Diabetics
B. Pacific Islanders
C. Men who have sex with men
D. Residents of institutions for the mentally retarded
E. Children in the developing world

Answer: A All of the other answers are risk factors for amebiasis, with children in low-income settings in the developing world at highest risk.

2. Which statement is *false* for the parasite *Entamoeba dispar*?

A. It is as prevalent as *Entamoeba histolytica* worldwide.
B. It is a cause of amebic colitis and liver abscess.
C. It is more common than *E. histolytica* in North America.
D. It is identical in appearance to *E. histolytica* on stool ova and parasite examination.
E. It is fecal-orally transmitted.

Answer: B *E. dispar* is not known to cause disease and is more common in temperate climates than the identical-appearing *E. histolytica*.

3. Which statement is true for the diagnosis of intestinal amebiasis?

A. Serology is positive acutely in more than 90% of cases.
B. Test results for occult blood in stool are positive in most.
C. Stool ova and parasite examination is sensitive and specific.
D. *E. histolytica*–specific stool antigen detection test is more than 80% sensitive for the diagnosis.

Answer: D Stool ova and parasite examination is sensitive and specific. Serologic testing by indirect hemagglutination is positive during acute illness in no more than 70% of cases (but >90% in convalescence). Blood in the stool occurs in virtually all patients with amebic colitis, but stools do not have to be positive for occult blood in amebic diarrhea. The stool ova and parasite examination is neither sensitive nor specific for amebiasis.

4. Which statement is false about extraintestinal amebiasis?

A. Amebic liver abscess is much more common in men than in women and rare in children.
B. Amebic brain abscess is seen only in patients with amebic liver abscess.
C. Amebic liver abscess can be treated without drainage in many cases.
D. Bacterial superinfection of amebic liver abscess is common.

Answer: D For unclear reasons, bacterial superinfection of an amebic liver abscess is rare.

353

BABESIOSIS AND OTHER PROTOZOAN DISEASES

SAM R. TELFORD III AND PETER J. KRAUSE

BABESIOSIS

Babesiosis is a tick-borne malaria-like disease caused by sporozoan parasites of the genus *Babesia*.

EPIDEMIOLOGY

Three worldwide epidemiologic patterns are apparent. The first involves the rodent-maintained *Babesia microti*, which is a species complex distributed across the Holarctic. On average, more than 1000 cases of *B. microti* babesiosis have been reported from the northeastern United States and upper Midwestern states from 2011 through 2013, the first 3 years that human babesiosis has been designated as a notifiable infectious disease by the Centers for Disease Control and Prevention (CDC). By comparison, about 30,000 Lyme disease cases are reported each year in the United States. The vector for *B. microti* is the same as that for Lyme disease (Chapter 321), the deer tick, *Ixodes dammini*, also known as northern populations of *I. scapularis*. Indeed, concurrent babesiosis and Lyme disease is common. Immune-intact as well as immunocompromised individuals are at risk. Within the last decade, *B. microti* has been increasingly reported in an expanded distribution from the original foci in coastal New England and the Upper Midwest, and it is now possible that babesiosis cases may be found wherever Lyme disease is intensely zoonotic. In addition, cases of *B. microti* or *B. microti*–like babesiosis have been reported from Australia, China, Germany, Japan, and Taiwan; the vectors have not been definitively identified.[1] The second pattern is represented by fewer than 50 cases of babesiosis due to *Babesia divergens*, *B. divergens*–like, or closely related species (e.g., *Babesia venatorum* or EU1) that have been reported from Europe. Almost all have been in splenectomized patients residing in sites where European castor bean ticks (*Ixodes ricinus*) and deer are common.[2] A few cases have been described in the United States and the Canary Islands. The third pattern of babesiosis involves sporadic cases due to diverse *Babesia* spp. These include a *Babesia duncani* (WA-1) and CA-type parasites of the western United States; a *B. divergens*–like species (MO-1); a *Babesia motasi*–like infection (KO-1) in Korea; and unidentified

Babesia spp from Colombia, Egypt, India, Mexico, and South Africa. There are at least 100 described *Babesia* spp from mammals and birds, and these hemoparasites are common animal infections on all continents. With few exceptions, *Babesia* spp are transmitted by ixodid ticks. Thus, wherever humans are intensely exposed to hard-bodied ticks, babesiosis should be part of a differential diagnosis for a patient presenting with fever and hematologic abnormalities.

Although the known zoonotic tick vectors (*I. dammini, I. ricinus*) have marked seasonal periods of activity (May to August) and the majority of reported cases are acquired during these times, babesiosis may be diagnosed at any time of the year. More than 150 cases of transfusion-acquired babesiosis due to *B. microti* and three due to *B. duncani* have been reported. The actual number of cases is thought to be much greater. *B. microti* is currently the most commonly reported transfusion-transmitted pathogen in the United States, and the number of such cases is increasing, including those ending in death.[3] Cases occur throughout the year, and about 10% of cases occur in nonendemic areas because *Babesia*-infected blood is exported to nonendemic areas or persons become infected in endemic areas and subsequently donate blood in nonendemic areas. A few cases of transplacentally transmitted babesiosis have been reported.

PATHOBIOLOGY

The pathophysiology of *Babesia* infection is directly related to the development of parasitemia. Peripheral blood parasitemias of 70% or greater have been reported, although most cases sustain parasitemias on the order of 0.5 to 5%. In hamsters infected by inoculation of a human-derived *B. microti* strain, intravascular hemolysis develops as the parasitemia rises and results in profound anemia. The hematocrit may fall to less than 20%. During this acute phase of disease, there is extramedullary hematopoiesis and hyperplasia of the splenic red pulp. Livers of infected hamsters contain hypertrophied Kupffer cells, many with ingested parasitized erythrocytes but little hemoglobin breakdown products. The proximal convoluted tubules of the kidneys contain abundant hemosiderin, an observation consistent with the occurrence of marked intravascular hemolysis.

Excessive production of pro-inflammatory cytokines seems to best explain the most common clinical manifestations, which include fever, sweats, chills, headache, myalgia, nausea, vomiting, diarrhea, and pallor. Administration of recombinant tumor necrosis factor (TNF) to human volunteers induces most of the symptoms of babesiosis and malaria. Such findings are not seen when erythrocyte lysis is due to noninfectious causes, which suggests that the release of merozoites serves as a trigger for the pro-inflammatory cascade. Elevated serum concentrations of TNF as well as of interferon-γ, interleukins 2 and 6, E-selectin, vascular cell adhesion molecule 1, and intracellular cell adhesion molecule 1 are detected during the acute phase of human *B. microti* infection and return to baseline within 3 months after resolution of infection.

Severe illness caused by infection with *Babesia* includes a complex array of metabolic abnormalities and organ dysfunction. Pulmonary disease is the most common complication in people experiencing severe *Babesia* infection, with up to 20% of patients suffering from noncardiogenic pulmonary edema. Pro-inflammatory cytokines appear to mediate the pulmonary complications of *Babesia* infection, at least in part. TNF and interferon-γ mRNA are upregulated in the lungs of *B. duncani*–infected mice, whereas TNF-knockout mice are less likely to die of fulminating *B. duncani* infection than are those with an intact TNF response. It is also likely that lung and other end-organ disease is mediated, at least in part, by vascular stasis, which has been described in the lungs of hamsters and mice infected with *B. duncani*.

CLINICAL MANIFESTATIONS

About a quarter of *B. microti* infections in adults and half of those in children are subclinical. This estimate is derived from an epidemiologic study that determined the frequency of people who seroconverted during the course of the summer transmission season but reported no illness, coupled with a careful accounting of symptomatic cases. Most people experience a mild to moderate illness lasting about a week. There is a gradual onset of malaise, anorexia, fatigue, fever (temperature as high as 40° C), sweats, and myalgia. Nausea, vomiting, headache, shaking chills, emotional lability, depression, hemoglobinuria, and hyperesthesia also have been reported. Findings on physical examination consist of fever, pallor, splenomegaly, and hepatomegaly. Laboratory abnormalities include anemia, thrombocytopenia, and leukopenia. Parasitemia generally ranges from barely detectable on blood smear to 5% in previously healthy people but may reach 85% in asplenic and other immunocompromised patients. Lactate dehydrogenase, bilirubin, and transaminase levels may be elevated in more severe cases. Persistent relapsing illness may occur in highly immunocompromised people who fail to clear the infection for months or more than a year despite multiple courses of antibiotics. The case-fatality rate for *B. microti* babesiosis has been estimated to be 6 to 9% in hospitalized patients but may be as high as 20% in immunocompromised hosts, including those who acquire the infection through blood transfusion. Severe babesiosis usually occurs only in people with asplenia, malignant disease, coinfection with human immunodeficiency virus (HIV), immunosuppressive treatment, or age younger than 2 months or older than 50 years.

Cases of babesiosis caused by species other than *B. microti* tend to be severe, at least in part because they are primarily reported in immunocompromised patients. Virtually all European patients experiencing *B. divergens* infection have been splenectomized, and about a third of the patients died. In these patients, there was an acute onset of illness with hemoglobinuria, a persistent nonperiodic high fever (temperature of 40° to 41° C), shaking chills, intense sweats, headaches, and myalgia as well as lumbar and abdominal pain. Vomiting and diarrhea may occur. Pulmonary, renal, or liver failure may develop rapidly. In fatal cases, patients become comatose with multiorgan failure. *B. duncani*, *B. venatorum*, and *B. divergens*–like infections also have often been reported in immunocompromised hosts with a similarly severe course of illness.

DIAGNOSIS

The diagnosis of babesiosis is based on epidemiologic and clinical findings and confirmed by laboratory testing. Given clinical findings consistent with babesiosis, the diagnosis may be confirmed by examination of a Giemsa-stained thin blood smear for the presence of parasites within erythrocytes. In an immunocompromised patient, parasitemias are likely to exceed one infected cell per oil immersion field and thus are quickly detected. For *B. microti* babesiosis (Fig. 353-1), examination of a slide for 10 minutes or as many fields as needed to tally 200 leukocytes (that are not infected but serve as a marker for effort) and repeated smears performed twice a day may be required. Standard Romanowsky stains (Giemsa, Wright) with malaria protocols are optimal. Artifactual inclusions are limited mainly to stain precipitates (which can be determined by their presence in the plasma spaces between cells), Howell-Jolly or Heinz bodies (Chapter 157), or platelets superimposed on erythrocytes, which always have a light colored halo when visualized this way. *Babesia* spp have clearly defined chromatin with a lighter-colored cytoplasm (Fig. 353-2A). They may be mistaken for early malarial trophozoites. Neither malarial nor babesial rings have hemozoin (malarial pigment), so this is not a good feature to distinguish between the two. Paired piriform parasites, arranged in a *v*, are suggestive of *B. divergens* or *B. divergens*–like infection (Fig. 353-2B). Rings of all sizes may be seen in all species. Multiple parasites may frequently be seen in single erythrocytes, as well as clumps of extracellular parasites. Tetrad forms (Fig. 353-2C) and Maltese cross forms (Fig. 353-2D) are diagnostic but are rarely seen in *B. microti* babesiosis. They seem to be more common with *B. duncani* or CA-type infections.

Polymerase chain reaction (PCR) assays are an important adjunct to blood smears. PCR is usually more sensitive than blood smears in cases in which parasitemias are sparse. Real-time PCR assays performed in-house would

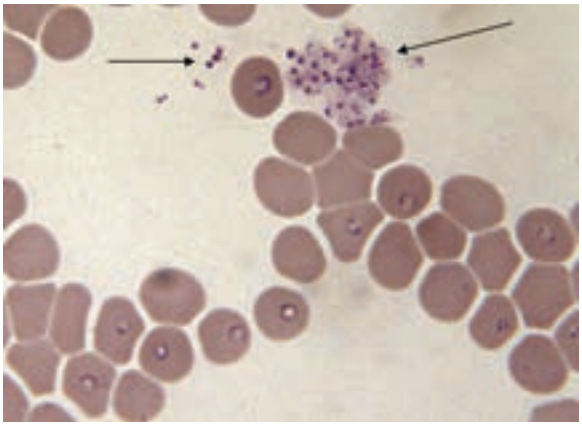

FIGURE 353-1 ***Babesia microti*.** Human infection, Nantucket Island. Predominance of ring forms with a cluster of extraerythrocytic parasites (*arrow*) free in the plasma.

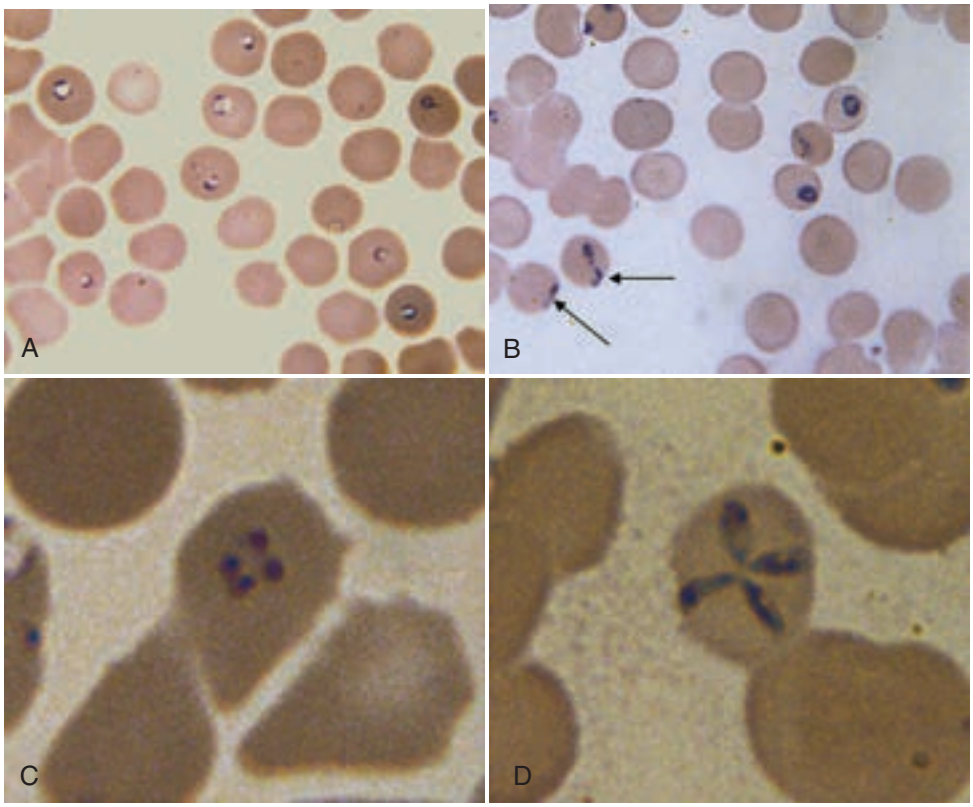

FIGURE 353-2. Diagnosis of *Babesia* infection. A, Typical thin-film field demonstrating ring forms of *B. microti* with vacuole or "whitish" cytoplasm demarcated by a dark-staining, defined chromatin. **B,** *B. Divergens*–like (MO-1) with robust rings; accolé form and paired piriform parasites are marked by *arrows*. **C,** Tetrad forms of *B. microti*. **D,** MO-1, classic Maltese cross. (Microscope slide from human case of MO-1 kindly provided by Dr. J. F. Beattie, Department of Pathology, The Medical Center, Bowling Green, Ky).

provide confirmation nearly as quickly as microscopy with the added advantage of increased sensitivity.[4] As with many molecular diagnostic assays, use of PCR for babesia (either conventional or real-time quantitative) is limited by false positives, validation of experimental assays, and insurance reimbursement. PCR sequencing, performed by collaborating research laboratories, is extremely valuable in retrospectively identifying the species of *Babesia* when the identity is not clear.

Serologic testing is useful for confirming *B. microti* infection. The indirect immunofluorescence test, using antigen from infected hamster red cells, is sensitive and specific and is currently the serologic method of choice. Analysis of paired acute and convalescent serum samples is most useful for a confirmation of *B. microti* infection. The presence of parasite-specific IgM may indicate that the patient has an acute infection even in the absence of readily demonstrable parasitemia. Serology is not generally useful for *B. divergens* babesiosis, given its fulminant natural history. Because parasitemia occurs before an antibody response and the doubling time of *B. divergens* can be as short as 8 hours, treatment needs to be initiated immediately on the basis of clinical suspicion and initial laboratory results.

The known vectors for human babesiosis are ticks that also transmit the agents of Lyme disease, human granulocytic anaplasmosis, *Borrelia miyamotoi* infection, *Ehrlichia muris*–like infection, and tick-borne encephalitis virus. Thus, coinfections should be considered in all patients with babesiosis. Acute illness in patients coinfected with Lyme disease and babesiosis is more severe and more persistent than in patients experiencing Lyme disease alone.

TREATMENT

Rx

Therapy for mild to moderate *B. microti* cases should consist of the combination of atovaquone (750 mg orally twice daily for 7 to 10 days) and azithromycin (500 to 1000 mg initial dose followed by 250 mg orally daily for 7 to 10 days; for immunocompromised hosts, 600 to 1000 mg orally daily for 7 to 10 days). A prospective randomized trial demonstrated that patients treated with atovaquone and azithromycin cleared parasitemia as effectively as did those receiving clindamycin and quinine and with fewer side effects. For severe babesiosis, a 7- to 10-day course of the combination of clindamycin (either 300 to 600 mg every 6 hours intravenously or 600 mg orally every 8 hours) and quinine (650 mg orally every 8 hours) should be used. The pediatric regimen is clindamycin, 7 to 10 mg/kg given every 6 to 8 hours (maximum of 600 mg/dose) and quinine 8 mg/kg given every 8 hours orally (maximum of 650 mg per dose).

Treatment may occasionally fail in high-risk patients or in those who must discontinue quinine because of side effects, such as severe tinnitus and gastrointestinal distress. Multiple courses of treatment for a prolonged duration may be required to clear parasitemia in immunocompromised patients; combination therapy should be used that may include two or more of the following antimicrobials: artemisinin, atovaquone, azithromycin, clindamycin, doxycycline, atovaquone-proguanil (Malarone), pentamidine, quinine, and trimethoprim-sulfamethoxazole. Once an effective combination is identified, it should be continued for at least 6 weeks and 2 weeks beyond the time when *Babesia* can no longer be visualized on blood smear or blood samples become PCR negative.

Exchange transfusion should be considered in severely ill patients with parasitemias in excess of 10%, evidence of severe hemolysis, or organ compromise. In particularly severe babesiosis cases, partial or complete blood exchange transfusion (1 to 3 blood volumes) should be undertaken, in addition to treatment with clindamycin and quinine.

PREVENTION

Prevention depends on reducing the risk for tick bites. Immunocompromised individuals should be especially careful to use personal protection and may even consider avoiding highly endemic sites such as coastal New England and Long Island during May through July, when risk is the greatest.[5] Use of repellants such as DEET or application of permethrin to clothing will greatly reduce tick attachment. Such products should be applied to shoes, socks, and trouser cuffs. Wearing light-colored long pants and tucking the cuffs into socks will also help prevent ticks from gaining access to attachment sites. Daily examination for attached ticks should be performed; the best way to do this is to feel for new bumps on a soapy body in the shower. Any attached

ticks should be promptly removed by simple traction, which is best accomplished with the use of tweezers. As with the agent of Lyme disease, ticks must be attached at least 36 to 48 hours before a sufficient inoculum of *Babesia* sporozoites is delivered. Community-level prevention should focus on public education about the risks of tick-borne infection, reducing habitat for ticks (brush removal and landscaping around yards), or reducing the reproductive hosts for the tick. Deer reduction will reduce the abundance of the deer tick vector for *B. microti* babesiosis. Currently, screening of blood donations for *Babesia* spp consists only of targeted questions about a history of previous *Babesia* infection, but laboratory screening methods are being developed.

PROGNOSIS

Death may occur in patients with severe babesiosis, but other long-term sequelae have not been reported for patients who have been adequately treated. In most patients who complete a full treatment regimen, *B. microti* DNA becomes undetectable by PCR within 3 months. Infection does not imply protective immunity based on laboratory rodent models, although subsequent infections are limited in duration and intensity. Recrudescent infections have been reported, mainly in immunocompromised individuals.

MISCELLANEOUS ENTERIC PROTOZOA

The gastrointestinal and urogenital tracts may contain representatives of the four major groupings of protozoa (amebae, sporozoa, flagellates, and ciliates). Diarrhea and other lower gastrointestinal signs and symptoms may be caused by diverse protozoa. Specific clinical diagnosis is not possible; expert clinical parasitology support is required to determine whether an agent that has been detected in a stool sample is a pathogenic species.[6] Identification is necessary because treatment options differ by the agent. With the exception of *Trichomonas vaginalis* infection (sexually transmitted), all of the enteric protozoa are acquired by the ingestion of food or materials contaminated by human feces; a small subset may have extraintestinal manifestations. Given a shared mode of transmission (fecal-oral), demonstrating the presence of any one of these protozoa within a stool sample from a patient is justification for an intensified search for those that are recognized as clinically significant pathogens (*Entamoeba histolytica*, *Giardia lamblia/intestinalis*, *Cyclospora cayetanensis*, *Cystoisospora belli*, and *Cryptosporidium parvum/hominis*). Other protozoa, many of which morphologically resemble true pathogens, are commonly detected within stools of patients with lower gastrointestinal disturbances, but support for their role as etiologic agents is weak.

Cryptosporidiosis (Chapter 350), giardiasis (Chapter 351), and amebiasis (Chapter 352) are discussed in separate chapters. Trichomoniasis and coccidian enteritis are discussed here because they are relatively common infections.

Trichomoniasis

EPIDEMIOLOGY

Trichomonas vaginalis is among the most prevalent of all pathogenic protozoa and is one of the most common sexually transmitted infections in the United States and likely worldwide.[7] As many as 30% of female college students and 40% of pregnant Nigerian women were found to be infected. The highest incidence of infection occurs in women with multiple sexual partners and those with other sexually transmitted diseases (Chapter 285). *T. vaginalis* can also be passed from infected mothers to their newborn daughters, but it is seldom symptomatic in girls before menarche. The parasite is able to survive for some time in moist environments, and nonvenereal transmission, although uncommon, can occur. Trichomoniasis, like other sexually transmitted diseases, may increase the likelihood of transmission of HIV.

PATHOBIOLOGY

T. vaginalis, known colloquially as a flagellate, is classified in the phylum Metamonada and class Parabasalia along with another human pathogen, *Dientamoeba fragilis* (previously thought to be an ameba). The 10- to 15-µm-long trophozoites multiply by longitudinal binary fission on the epithelial surface of the vagina or urethra as well as in vaginal or urethral secretions and are thereby transmitted by sexual intercourse. No cyst form is known, and the trophozoites are easily killed by drying.

The parabasalids are a phylogenetic sister group to the class Eopharingia, which includes *Giardia* spp. The parabasalids lack mitochondria and are anaerobic; they all have a unique cellular organelle, the hydrogenosome, which is a relic of the mitochondrion and serves as the site of anaerobic pyruvate metabolism. A hemolysin is produced and may cause epithelial damage.

CLINICAL MANIFESTATIONS

Trichomoniasis is one of three common causes of vaginitis or vaginosis (along with bacterial vaginosis and vulvovaginal candidiasis). It is characterized by a thin gray to yellowish green frothy discharge; vulvovaginal erythema; ectocervical erythema or "strawberry cervix," observable mainly by colposcopy; pH higher than 4.5; increased presence of polymorphonuclear leukocytes; and a positive result of the whiff test, in which a foul fishy odor is intensified on addition of potassium hydroxide. The incubation period for trichomoniasis is 5 to 28 days. In addition to a frothy discharge, vaginitis can be accompanied by vulvovaginal irritation, dyspareunia, abdominal pain, and dysuria. Symptoms may worsen during menstruation. Population-based studies indicate that as many as half of *T. vaginalis* infections in women and the majority in men are asymptomatic. *T. vaginalis* can frequently be isolated from the male partners of infected women and can produce symptomatic urethritis. Urethral discharge is generally scant in these cases. Rarely, *T. vaginalis* is associated with epididymitis, superficial penile ulcerations that are usually located under the prepuce, or prostatitis.

DIAGNOSIS

The CDC guidelines (http://www.cdc.gov/std/treatment/2010/vaginal-discharge.htm#a2) state that all women with a sexually transmitted infection should be specifically tested for evidence of *T. vaginalis* infection, and HIV-positive women should be tested annually. In women, vaginal and urethral secretions should be examined. *T. vaginalis* is seen in wet mounts of vaginal secretions in approximately 60% of infected women, thus confirming the diagnosis. Live *T. vaginalis* have a twitching or tumbling motion in wet mounts, and polymorphonuclear leukocytes are usually present. Direct immunofluorescent antibody staining is more sensitive than wet mounts but technically more difficult. Culture is an even more sensitive method of diagnosis; commercial kits for culture are available, but the results are not available for 3 to 7 days. *T. vaginalis* is occasionally identified in Papanicolaou-stained smears; Giemsa stain may also be used (Fig. 353-3). For men, a wet mount of material from a platinum loop scraping of the anterior urethra reveals the organism in approximately half the cases. Prostatic massage before collection of urine for *Trichomonas* culture is a more sensitive diagnostic approach. *T. vaginalis* is not found in the gastrointestinal tract, and the presence of trichomonads in wet fecal mounts or stained fixed fecal smears (iron hematoxylin or trichrome) most likely represents the commensal *Pentatrichomonas* (formerly *Trichomonas*) *hominis*. Serology has limited clinical use because of issues of sensitivity and specificity and because evidence of exposure does not imply current disease. New modes of testing are becoming available and will eventually supplant microscopy. One Food and Drug Administration (FDA)–approved (and CLIA waived) point-of-care antigen assay (OSOM Trichomonas Rapid Test; Sekisui Diagnostics, Framingham, Mass) has

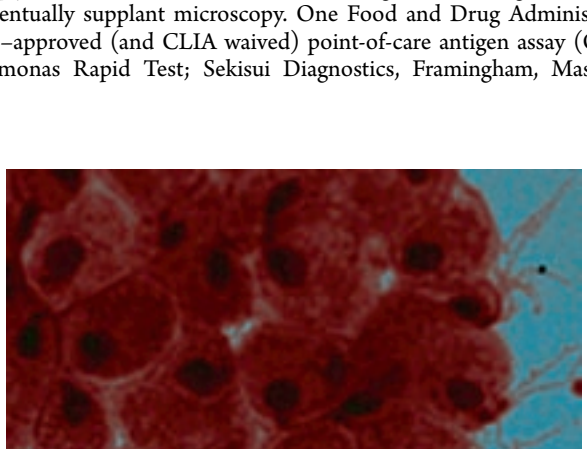

FIGURE 353-3. ***Trichomonas vaginalis*, Giemsa-stained smear of cultivated trophozoites.** (Bright-field microscopy, ×630. Scale bar is 15 µm.)

greater sensitivity and specificity in comparison with microscopy. The FDA has also cleared a single PCR-based assay (APTIMA *Trichomonas vaginalis* Assay; Hologic GenProbe, San Diego, Calif) with excellent sensitivity and specificity, but the assay requires specimens (swab samples) fixed in a proprietary solution.

TREATMENT (also see Chapter 344)

Rx

Tinidazole, a single 2-g oral dose in adults, or metronidazole, either as a single 2-g oral dose or 500 mg twice daily for 7 days, is the treatment of choice. Tinidazole is the better tolerated of the two. Single-dose therapy (metronidazole or tinidazole) ensures compliance of the patient but can produce nausea and a metallic taste, particularly with metronidazole. Both tinidazole and metronidazole have a disulfiram-like effect, and patients who consume alcohol within 24 hours of metronidazole or 72 hours of tinidazole may experience severe nausea, vomiting, and flushing. The use of tinidazole and metronidazole is relatively contraindicated during pregnancy, given the lack of well-controlled studies. Treatment failures with metronidazole are uncommon but well documented. HIV-positive women should be treated for 7 days with 500 mg of metronidazole because of frequent recrudescence with the single-dose therapy. Some instances of treatment failure in immune-intact women result from reinfection, others from poor compliance, but some are caused by metronidazole-resistant parasites. A repeated course of metronidazole (2 g orally daily for 5 days) may be tried. If patients remain refractory to appropriate treatment, metronidazole sensitivity can be tested by the CDC (available at www.dpd.cdc.gov/dpdx/HTML/DiagnosticProcedures.htm).

PREVENTION

Condoms (male or female) reduce risk of acquiring trichomoniasis. Sexual partners should be treated concurrently to prevent reinfection because nearly 20% of male partners are coinfected.

PROGNOSIS

Rare complications include pelvic inflammatory disease. Infection during gestation may lead to fetal growth retardation. There is no natural or acquired immunity, so reinfection may be common.

Coccidian Enteritis

EPIDEMIOLOGY

The coccidia, with 43 genera and more than 1700 recognized species, are well-known veterinary pathogens.[8] At least two of the coccidia, *Cyclospora cayetanensis* (an eimeriid) and *Cystoisospora belli* (a sarcocystid), are causative agents of enteritis in humans. Despite their ubiquity (nearly 100 species have been described), *Sarcocystis* spp have been rare causes of enteritis in humans and an even less common cause of myositis, although *Sarcocystis nesbitti* caused a disease comprising fever, myalgia, headache, and myositis in 89 college students who had traveled to Malaysia.[9] *C. cayetanensis* may have both an animal and human reservoir, but *C. belli* is thought to be an anthroponosis. Cyclosporiasis is a cause of gastroenteritis in tropical and subtropical areas, with Peru, Mexico, Haiti, Caribbean countries, and Nepal commonly reporting such cases. Since 1990, at least 11 food-borne outbreaks affecting approximately 3600 persons have been documented in the United States and Canada among persons who have eaten contaminated raspberries, fresh basil, snow peas, or mesclun. Cyclosporiasis is not uncommonly diagnosed in international travelers, and large outbreaks have been reported from cruise ships.[10]

PATHOBIOLOGY

Oocysts are passed in feces and must sporulate for at least a day (*C. belli*) or 5 to 11 days (*C. cayetanensis*) before attaining infectivity. Sporozoites are liberated in the small bowel, penetrating enterocytes (for *C. cayetanensis*, mainly in the jejunum). *C. cayetanensis* appears to have a complicated developmental cycle with at least two merogonic cycles in the bowel, leading to the formation of gametes. The gametes fuse within the enterocyte cytoplasm, and an oocyst wall is deposited around the zygote. The sexual cycle begins about a week after infection, with oocysts sloughing into the bowel lumen and subsequently out of the body through feces. Biopsy specimens from infected patients demonstrate mononuclear and eosinophilic infiltrates in the lamina propria as well as alterations to the morphology of villi. Humans appear to be intermediate hosts for *S. nesbitti*, with sporozoites liberated in the small bowel, entering the vasculature, and forming cysts in muscle.

CLINICAL MANIFESTATIONS

After an incubation period of approximately 1 week, either organism produces watery diarrhea, nausea, vomiting, abdominal pain, myalgias, anorexia, and fatigue. Illness is usually self-limited, but symptoms can be prolonged (10 to 12 weeks) and associated with steatorrhea, flatulence, and substantial weight loss in persons who are immunocompromised, especially those with AIDS. Infections may also be asymptomatic. With *S. nesbitti*, fever and prominent myalgias are the dominant presentation.

DIAGNOSIS

The diagnosis is confirmed by identifying coccidia in stool samples stained with modified acid-fast or modified safranin preparations or by phase contrast microscopy or bright-field microscopy (using iodine as a contrast medium) of wet mounts (Fig. 353-4). *C. cayetanensis* and *C. belli* may be sensitively detected by fluorescent microscopy of wet mounts. PCR is specific and can be sensitive, depending on the mode of DNA extraction, but there are no FDA-approved assays available; PCR support may be requested from the CDC through state public health departments. *S. nesbitti* appears to undergo an aberrant asexual cycle only within humans and may not produce oocysts to be liberated into the bowel; therefore, diagnosis has required muscle biopsy and demonstration of sarcocysts by histology or detection of the agent's DNA by PCR.

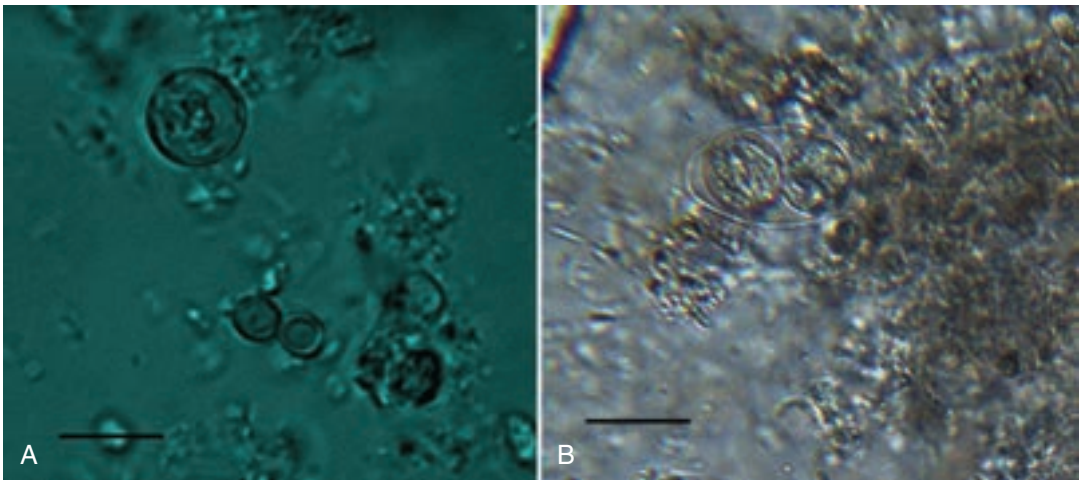

FIGURE 353-4. Diagnosis of coccidian enteritis. A, *Cyclospora cayetanensis*, seen on formalin-fixed human feces, unsporulated. (Bright-field microscopy, ×1000 with green contrast filter. Scale bar is 10 µm.) Note the doublet of yeast cells in lower right of the photomicrograph. **B,** *Cystoisospora belli*, seen on formalin-fixed human feces, partially sporulated. (Bright-field microscopy, ×400. Scale bar is 15 µm.)

TABLE 353-1 OTHER ENTERIC PROTOZOA

ORGANISM	EPIDEMIOLOGY	MANIFESTATIONS	THERAPY*
Balantidium coli	Primarily an infection of animals, especially pigs, but also affects humans	Asymptomatic or mild and self-resolving; occasionally more severe with abdominal pain, blood, and mucus in stool	Tetracycline (500 mg qid for 10 days) Alternative: metronidazole (750 mg tid for 5 days) or iodoquinol (650 mg tid for 20 days)
Blastocystis hominis	Probably worldwide, including North America; often found concomitantly with *Giardia lamblia*	Pathogenicity is debated	The need for treatment is debated, but symptomatic improvement has been reported with metronidazole (750 mg tid for10 days), or trimethoprim-sulfamethoxazole (160 mg TMP/800 mg SMX bid for 7 days)
Dientamoeba fragilis	Worldwide distribution; frequently found concomitantly with the pinworm *Enterobius*	Often asymptomatic; diarrhea reported	Paromomycin (25-35 mg/kg body weight per day in 3 doses for 7 days), tetracycline (500 mg qid for 10 days), metronidazole (500-750 mg tid for 10 days), or iodoquinol (650 mg tid for 20 days)
Microsporidia† (*Enterocytozoon bieneusi* and *Encephalitozoon intestinalis*)	Apparent worldwide distribution	AIDS patients with persistent diarrhea and wasting; self-limited cases in immunocompetent persons	Oral fumagillin (20 mg tid) has been effective for *E. bieneusi*, but it has been associated with thrombocytopenia. Albendazole (400 mg bid) has been effective for *E. intestinalis*. Treatment with HAART may lead to clinical response in HIV-infected patients with microsporidial diarrhea.
Sarcocystis species	Common pathogens of animals; rare in humans; acquired by ingesting contaminated beef or pork	Often asymptomatic; nausea, vomiting, abdominal pain, and diarrhea may occur; eosinophilic necrotizing enteritis has been reported	No specific therapy

*Based on CDC recommendations, www.cdc.gov/parasites/az/index.html. Accessed March 9, 2015. The dosages and durations are for adults.
†Associated with persistent, severe diarrhea in persons with AIDS.
AIDS = acquired immunodeficiency syndrome; HAART = highly active antiretroviral therapy; HIV = human immunodeficiency virus.

TREATMENT

Rx

Rehydration is important, as it is for any severe diarrheal disease. Both infections may respond to treatment with 160 mg trimethoprim and 800 mg sulfamethoxazole taken twice daily for 7 to 10 days. HIV-infected patients may require a longer course of therapy. The widespread use of trimethoprim-sulfamethoxazole (Bactrim) prophylaxis for *Pneumocystis* has reduced the incidence of coccidial diarrhea in patients with HIV infection. No specific treatment has been recommended for sarcocystosis.

PREVENTION

At the community level, preventing the contamination of water and food (mainly vegetables and fruits) by animal or human feces reduces the risk of transmission. Washing vegetables and fruits in water will reduce the potential inoculum but does not eliminate all risk. Sarcocystosis may be prevented by ensuring that meat is well cooked.

PROGNOSIS

Reactive arthritis, Guillain-Barré syndrome, Reiter's syndrome, cholecystitis, and cholangitis have been reported as complications of either coccidian enteritis, mainly in patients with AIDS. Otherwise, treatment appears to eradicate the organism. Whether reinfection may occur is not known. Reports of extraintestinal development of *C. belli* and *C. cayetanensis* suggest the possibility of reinvasion of the bowel with ensuing recrudescence of signs and symptoms.

Other Enteric Protozoans

A number of other protozoa transmitted by fecal-oral contamination have been associated with enteric disease (Table 353-1). Some reside in the lumen of the bowel, and others invade and multiply within enterocytes. Enteric protozoa should be considered in the differential diagnosis of patients with persistent diarrhea and abdominal symptoms, particularly those with a history of recent international travel. A clinical diagnosis is rarely possible; laboratory tests, mainly for ova and parasites in stools, establish the diagnosis. Expert microscopists are required because these parasites may be confused with fecal debris. Pathogenic protozoa must also be differentiated from commensals such as *Entamoeba coli, Endolimax nana, Iodamoeba bütschlii, Pentatrichomonas hominis,* and *Chilomastix mesnili.* Therapy includes administration of the appropriate antiprotozoal drug and rehydration, as listed in Table 353-1[11] (also see Chapter 344).

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Vannier E, Krause PJ. Human babesiosis. *N Engl J Med*. 2012;366:2397-2407.
2. Hildebrandt A, Gray JS, Hunfeld KP. Human babesiosis in Europe: what clinicians need to know. *Infection*. 2013;41:1057-1072.
3. Herwaldt BL, Linden JV, Bosserman E, et al. Transfusion-associated babesiosis in the United States: a description of cases. *Ann Intern Med*. 2011;155:509-519.
4. Bloch EM, Lee TH, Krause PJ, et al. Development of a real-time polymerase chain reaction assay for sensitive detection and quantitation of *Babesia microti* infection. *Transfusion*. 2013;53: 2299-2306.
5. Centers for Disease Control and Prevention (CDC). Babesiosis surveillance—18 states, 2011. *MMWR Morb Mortal Wkly Rep*. 2012;61:505-509.
6. McHardy IH, Wu M, Shimizu-Cohen R, et al. Detection of intestinal protozoa in the clinical laboratory. *J Clin Microbiol*. 2014;52:712-720.
7. Edwards T, Burke P, Smalley H, et al. Trichomonas vaginalis: Clinical relevance, pathogenicity and diagnosis. *Crit Rev Microbiol*. 2014;1-12.
8. Legua P, Seas C. Cystoisospora and cyclospora. *Curr Opin Infect Dis*. 2013;26:479-483.
9. AbuBakar S, Teoh BT, Sam SS, et al. Outbreak of human infection with *Sarcocystis nesbitti*, Malaysia, 2012. *Emerg Infect Dis*. 2013;19:1989-1991.
10. Gibbs RA, Nanyonjo R, Pingault NM, et al. An outbreak of *Cyclospora* infection on a cruise ship. *Epidemiol Infect*. 2013;141:508-516.
11. Centers for Disease Control and Prevention. Parasites. http://www.cdc.gov/parasites/az/index.html. Accessed March 9, 2015.

REVIEW QUESTIONS

1. A 35-year-old woman presented to the emergency department in Toronto with a 2-day history of fever, headache, and malaise. She had just returned from spending 2 weeks during June on Nantucket Island. A 3-inch expanding circular rash was noted on one of her extremities, and she was diagnosed with acute Lyme disease and provided with a prescription for 100 mg of doxycycline, to be taken twice a day for 21 days. A month later, she presents again with fever, headache, chills, myalgia, and fatigue. She had not traveled again and works as a law clerk. Complete blood count demonstrates a hematocrit of 35, hemoglobin of 12, white blood cell count of 3000 (60% neutrophils, 35% lymphocytes, and 5% monocytes), and platelet count of 120,000; a blood smear is normal. What is the most likely cause of the recurrent symptoms?

A. Epstein-Barr virus infection
B. Human anaplasmosis
C. Myelodysplastic syndrome
D. Human ehrlichiosis
E. Babesiosis

Answer: E Symptomatic Epstein-Barr virus infection is unlikely in a 35-year-old patient and usually is accompanied by lymphocytosis with atypical lymphocytes. Although thrombocytopenia or leukopenia may be seen with anaplasmosis and ehrlichiosis, these infections are susceptible to doxycycline and would have been adequately treated with the course prescribed for the Lyme disease. Myelodysplastic syndrome typically is manifested with anemia and without fever. Treatment failure may rarely occur with acute Lyme disease, mostly as a result of poor compliance of the patient with therapy, but later manifestations of Lyme disease are more likely to be a monarticular arthritis or radiculopathy. Babesiosis may be coacquired with Lyme disease, but babesia are not susceptible to the tetracyclines. The negative blood smear is not unusual, and the severity of illness is not necessarily related to parasitemia. Thrombocytopenia and leukopenia may be evident in the absence of hematologic signs of anemia, particularly in early infection. Polymerase chain reaction will often be positive when blood smears are negative, given its greater sensitivity. Serology with a high anti-*Babesia* titer would provide evidence of infection if polymerase chain reaction is not performed, given that the patient was likely infected 6 weeks previously.

2. A 30-year-old biologist presented with a 3-week history of watery, nonbloody diarrhea with nausea and low-grade fever. He had previously been in good health and had returned 6 weeks ago from a field trip to Panama, where he had stayed for 1 month. He had several episodes of diarrhea in Panama that had responded quickly to Pepto-Bismol. He had remained hydrated and thus his activity was not curtailed. He now seeks medical attention because the diarrhea fails to respond to Pepto-Bismol or Imodium. Physical examination findings are normal, as is a complete blood count with the exception of slight eosinophilia. The patient admits to eating local foods and failing to always drink bottled water. Leukocytes are not noted in a stool examination, but Charcot-Leyden crystals are present. Stool analysis for ova and parasites demonstrates 30 × 15-µm oocysts with a sharply defined cell wall and a single sporoblast. Rare *Entamoeba* spp cysts with eight nuclei are also found. Which of the following would provide the most critical information for case management?

A. Testing for *Clostridium difficile* toxin
B. Human immunodeficiency virus (HIV) infection status
C. Testing for norovirus
D. Specific identification of the *Entamoeba* sp
E. Whether the Pepto-Bismol–treatable diarrhea was similar in quality to that experienced now

Answer: B The oocyst that was detected is likely *Cystoisospora belli*, which typically causes a self-limited diarrhea in immunocompetent younger patients. Norovirus and *C. difficile* disease should also terminate within a few days in a younger, healthy individual. The presence of eight nuclei identifies *Entamoeba coli*. Amebic infection reflects fecal-oral contamination, although the nonpathogenic *Escherichia coli* may also be found as a gut commensal. It is possible that the diarrhea episodes in Panama were an uncomplicated traveler's diarrhea and that the current disease is not related. Persistent enteritis due to *C. belli* is commonly seen in HIV-infected patients, who may develop life-threatening persistent diarrhea, malabsorption, and dehydration and require lifelong suppression of *C. belli* with trimethoprim-sulfamethoxazole (Bactrim).

3. A 40-year-old married woman presents with a foul-smelling vaginal discharge of 3 days' duration. She is sexually active with her husband and complains of a recurring history of discharge during the last 6 months, typically a week after intercourse, but not with every instance of intercourse. The discharge would resolve within 2 weeks if she was abstinent. A wet mount of the discharge demonstrates flagellated cells moving with a tumbling motion, and a diagnosis of trichomoniasis is made. Metronidazole 250 mg orally three times a day is prescribed for 7 days. She returns 2 months later with the same discharge, stating that she had improved after the treatment course but that the discharge returned within the last week. Which of the following is the most likely explanation for the recurrence of symptoms?

A. The patient's husband has *Trichomonas* infection and is reinfecting the patient.
B. The patient has antibiotic-resistant trichomoniasis.
C. The original diagnosis based on the web mount examination alone was incorrect.
D. The patient has been noncompliant with treatment.

Answer: A Patients may be reinfected by their untreated sexual partners, and thus a diagnosis of trichomoniasis should imply presumptive treatment of the partner. The majority of infected men are asymptomatic. The current Centers for Disease Control and Prevention recommendations for treatment failure are to re-treat with tinidazole. Metronidazole susceptibility may be established, but there is no guidance on minimum inhibitory concentration thresholds to define resistant strains. Although *Trichomonas vaginalis* is detected on wet mount examination of vaginal secretions in only about 60% of infected women, it confirms the diagnosis when it is seen in the appropriate clinical setting.

354

CESTODES

A. CLINTON WHITE AND ENRICO BRUNETTI

DEFINITION

The Pathogens

Cestode parasites are members of the animal kingdom, subphylum Cestoda. The organisms are characterized by several life cycle stages, which typically develop in distinct hosts. The adult stage is the tapeworm, which is acquired by ingestion of uncooked tissues harboring larval forms. After ingestion, the larvae excyst and the scolex attaches to the intestines. Segments, termed *proglottids*, develop at the base of the scolex and are displaced from the scolex by new proglottids to form a chain or tapeworm. The host in which the tapeworm develops is termed the *definitive host*. The proglottids contain male and female sexual organs and produce large numbers of ova. The proglottids or their ova are shed in stools. Humans are the definitive hosts for a number of different tapeworms, including the *Taenia* species, *Diphyllobothrium* species, and *Hymenolepis nana*. Humans can also be an accidental host for the dog and cat tapeworms of the genus *Dipylidium* (Table 354-1).

The intermediate hosts harbor the larval form of the parasite. Infection follows ingestion of the ova. Under the influence of gastric and intestinal fluids, the ova hatch, releasing the invasive larvae (oncospheres), which migrate to tissues, forming tissue forms. The forms in tissue vary between

TABLE 354-1 COMMON HUMAN TAPEWORM INFECTIONS

ORGANISM	INTERMEDIATE HOST	COMMON NAME	CLINICAL PRESENTATION	TREATMENT
Diphyllobothrium spp	Fish	Fish tapeworm	Passing segments, pernicious anemia	Praziquantel, niclosamide
Hymenolepis nana	Humans	Dwarf tapeworm	Asymptomatic, diarrhea	Praziquantel, niclosamide
Taenia saginata	Cattle	Beef tapeworm	Asymptomatic, passing segments	Praziquantel, niclosamide
Taenia asiatica	Pigs	Asian tapeworm	Asymptomatic, passing worms	Praziquantel, niclosamide
Taenia solium	Pigs	Pork tapeworm	Asymptomatic, passing segments	Praziquantel, niclosamide
Dipylidium caninum	Fleas	Dog tapeworm	Passing segments	Praziquantel, niclosamide

TABLE 354-2 HUMAN LARVAL CESTODE INFECTIONS

ORGANISM	COMMON NAME	ORGANS INVOLVED
Taenia solium	Cysticercosis	Brain, spinal fluid, eye, muscle
Echinococcus granulosus group	Cystic hydatid disease	Liver, lung, other
Echinococcus multilocularis	Alveolar hydatid disease	Liver
Taenia multiceps, *Taenia* spp	Coenurosis	Brain, eyes
Spirometra species	Sparganosis	Subcutaneous tissue, viscera

TABLE 354-3 THERAPY FOR INTESTINAL TAPEWORM INFECTIONS

	PRAZIQUANTEL	NICLOSAMIDE	NITAZOXANIDE
Dosage			
Adults	5-10 mg/kg for all age groups (25 mg/kg for *Hymenolepis nana*)	2 g (4 tablets)	500 mg
Children >34 kg		1.5 g (3 tablets)	200 mg
Children 11-34 kg		1 g (2 tablets)	100 mg
Administration	Taken as a single dose	Taken as a single dose; tablets must be chewed and swallowed	Taken twice a day for 3 days
Side effects	Mild but frequent, including dizziness, myalgias, nausea, vomiting, diarrhea, abdominal pain	Nausea, vomiting, abdominal pain, diarrhea, drowsiness, dizziness, headache, pruritus	
Pregnancy		No known mutagenic effects; considered safe if indicated	

organisms and may include the cysticercus (a bladder containing a single invaginated scolex), the coenurus (a bladder with multiple scolices), the hydatid (a cystic structure with a germinal layer, which forms numerous protoscolices), or the plerocercoid (a solid form seen in *Spirometra* species). Humans can harbor the intermediate forms of *Taenia solium* (cysticercosis), *Echinococcus granulosus* group (cystic hydatid disease), *Echinococcus multilocularis* (alveolar hydatid disease), and rarely other organisms (Table 354-2). Humans can serve as both the definitive host and an intermediate host for two species, *T. solium* and *H. nana*. In the case of *T. solium*, humans are the obligate host for the tapeworm stage (pork tapeworm) but can also harbor the cystic form (cysticercosis). In the case of *H. nana*, both stages typically develop in a single person, with the cysticercoid form in the intestinal wall and the tapeworm in the lumen.

INTESTINAL TAPEWORM INFECTIONS

Diphyllobothrium Species (Fish Tapeworm)

Diphyllobothrium tapeworms are large segmented parasites that are acquired by ingestion of undercooked or pickled freshwater fish dishes (sushi, sashimi, ceviche, carpaccio, gefilte fish). The tapeworms develop within a few weeks and can live for more than 10 years.

EPIDEMIOLOGY AND PATHOBIOLOGY

Diphyllobothrium species are found worldwide, including foci in Europe, North and South America, and Asia. Perhaps 20 million people are thought to be infected worldwide. Major foci include Russia, Japan, and South America. Disease was formerly highly endemic in Scandinavia, where it is now rarely diagnosed.

In most cases, infection has little impact on the host. However, one species, *Diphyllobothrium latum*, contains vitamin B_{12} receptors on the surface of the tapeworm, which can out-compete the host, leading to vitamin B_{12} deficiency (Chapter 164). This manifestation has been described only in Scandinavia.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

In most of those infected, *Diphyllobothrium* species produce few or no symptoms. Some may complain of gastrointestinal symptoms (abdominal discomfort, nausea, weight loss). The main clinical manifestation is the observation of proglottids being passed in stool. Pernicious anemia with symptoms of anemia or peripheral neuropathy may develop with *D. latum* infection. The diagnosis depends on observation of the characteristic operculated eggs in stool.

TREATMENT AND PREVENTION

Rx

A single oral dose of praziquantel (5 to 10 mg/kg) is usually adequate for therapy (Table 354-3). Niclosamide can be used as an alternative (2 g [adults] or 50 mg/kg [children] in a single dose chewed and swallowed), but it is not available in the United States. Parasites in fish can be killed by cooking (>56°C, >5 minutes) or freezing (–20°C, 24 hours). Infected fish may also be identified by inspection.

Hymenolepis nana

Hymenolepis nana is the human dwarf tapeworm. *Hymenolepis diminuta*, a rat tapeworm, can also cause human infection.

EPIDEMIOLOGY AND PATHOBIOLOGY

H. nana is widely prevalent worldwide, with estimates of 50 to 75 million people infected. Infection follows ingestion of ova. The larvae are released, invade, and develop into cysticercoid forms in the intestinal villi. After a few days, the cysticercoids mature, invade the lumen, and are transformed into a scolex, forming small tapeworms (up to 5 cm long), which begin producing eggs within 2 to 3 weeks. Autoinfection either in the intestines or by the fecal-oral route can lead to heavy infection.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Most infections are asymptomatic. However, some children may be infected by hundreds or thousands of worms, which can cause abdominal pain, loose stools, diarrhea, and malabsorption. Diagnosis depends on observation of the characteristic eggs in stool. More than one specimen may be required.

TREATMENT AND PREVENTION

Rx

Praziquantel (15 to 25 mg/kg as a single oral dose) is usually effective in treating *H. nana* infection,[1] but it may need to be repeated in heavy infection (see Table 354-3). Nitazoxanide (100 mg by mouth twice daily for 3 days for children 1 to 3 years of age, 200 mg by mouth twice daily for 3 days for children 4 to 11 years of age, and 500 mg by mouth twice daily for 3 days for older children) is a reasonable alternative therapy; efficacy is about 75 to 82%. Niclosamide can be used as an alternative. Transmission is by the fecal-oral route and could be prevented by improved hygiene. Mass chemotherapy has been used to control infection in some populations.

Dipylidium caninum

Dipylidium caninum is a common tapeworm of dogs and cats. Dogs are infected by ingestion of fleas, which carry the cysticercoid form in their body cavities. The tapeworms can also develop in children who have ingested the fleas. It is widespread worldwide, but human infections are unusual.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Infection may be asymptomatic. In some cases, the motile proglottids may be noted in stool. The proglottids are similar in size and shape to rice grains. Diagnosis depends on identification of the ova in stool or identification of the proglottids.

TREATMENT AND PREVENTION

Rx

There are no controlled trials of treatment for *Dipylidium* infection, but infection is likely to respond to regimens used for other tapeworms (see Table 354-3). The main measure for prevention is treatment of pets for fleas and tapeworms.

Taenia saginata

Taeniasis refers to infection with the tapeworm form of one of three *Taenia* species. *Taenia solium* and *Taenia asiatica* are acquired from ingestion of undercooked pork. *Taenia saginata*, called the beef tapeworm, is a common intestinal infection worldwide. Cattle are the intermediate hosts, harboring the tissue cysticerci in their muscle. Humans are the obligate definitive host, harboring the tapeworm form.

EPIDEMIOLOGY AND PATHOBIOLOGY

T. saginata is common worldwide in areas where cattle are raised and human fecal material contaminates the pastures. Approximately 45 to 60 million people are thought to be infected. It is found on most continents. Very high rates (>20% of the population) have been noted in east Africa, Bali, and Tibet. It is likewise endemic in the Middle East, the Americas, and Europe. *T. saginata* is also common in other parts of Asia, but many of the epidemiologic studies did not differentiate *T. saginata* from *T. asiatica*.

T. saginata tapeworms are acquired by ingestion of undercooked beef. The scolex attaches to the intestinal wall, and proglottids form at the base of the scolex. The proglottids gradually enlarge as they are displaced from the scolex by newer proglottids. The chain of proglottids can reach a length of up to 30 feet. The terminal proglottids are shed periodically in the stool. Terminal proglottids are typically off-white, 2 to 3 cm long, 0.5 to 1 cm wide, and 1 to 2 mm thick.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Mild symptoms (e.g., nausea, abdominal discomfort, anorexia, and pruritus) may be noted. The motile proglottids may cause discomfort as they exit the anus or may be noted in stool.

Ova may be noted in stool. The ova are 40 µm in diameter, surrounded by brown radial striations, and the embryos have six hooks. However, the ova of the three *Taenia* species are morphologically indistinguishable. The proglottids can be distinguished from those of *T. solium* by counting the number of uterine branches (≥14 branches suggests *T. saginata*). However, the proglottids of *T. saginata* cannot be readily distinguished from *T. asiatica*.

TREATMENT, PREVENTION, AND PROGNOSIS

Rx

Taeniasis can be treated with praziquantel in a single dose (see Table 354-3). Single doses of niclosamide are also effective. Nitazoxanide has also been used for *T. saginata*. Taeniasis can be prevented by inspection of beef. Also, cooking to 56°C for 5 minutes or freezing at –20°C for 7 to 10 days destroys the infective larvae. Only minor symptoms are noted and are eventually self-limited with or without treatment.

Taenia asiatica

T. asiatica is a cause of taeniasis in Asia, termed *Asian taeniasis*. Infection is acquired by ingestion of undercooked pork. Pigs are infected by ingestion of the ova from tapeworm carriers. *T. asiatica* has been widely described in China, Taiwan, Korea, Indonesia, and Southeast Asia. The clinical manifestations, diagnosis, treatment, and prevention of *T. asiatica* infection are similar to those noted for *T. saginata* infection.

Taenia solium

T. solium, also known as the pork tapeworm, can cause both tapeworm infection and larval infection termed *cysticercosis*. *T. solium* tapeworm infections are caused by ingestion of infected undercooked pork. The scolex evaginates and attaches to the intestines, forming proglottids. The proglottids gradually mature as they are separated from the scolex by new proglottids. The adult worms are often 10 to 20 feet long. Within the mature proglottids, thousands of microscopic ova develop. The ova are either excreted into the stool or shed with the proglottids. By contrast, ingestion of the ova results in development of larval infection, termed cysticercosis (see later). Thus, the tapeworm carrier poses risk of self-infection as well as infection to other people.

EPIDEMIOLOGY AND PATHOBIOLOGY

T. solium is common worldwide in areas where pigs are raised and where pigs have access to human fecal material. Only a few million people are thought to harbor the tapeworm form. Pork tapeworm infection is highly endemic in Latin America, sub-Saharan Africa, south Asia, and Southeast Asia.

T. solium tapeworms are acquired by ingestion of undercooked pork. The scolex attaches to the intestinal wall and proglottids form at the base of the scolex. The proglottids gradually enlarge as they are displaced from the scolex by newer proglottids. The terminal proglottids are shed periodically in the stool. Terminal proglottids are typically off-white, 2 cm long, 0.5 to 1 cm wide, and 1 to 2 mm thick.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Mild symptoms (e.g., nausea, abdominal discomfort, anorexia, and pruritus) may be noted. The proglottids may be noted in stool. Diagnosis is made by the finding of ova in stool. The ova are 40 µm in diameter, surrounded by brown radial striations, and embryos have six hooks and are morphologically indistinguishable from the other *Taenia* species. The proglottids can be distinguished from those of *T. saginata* and *T. asiatica* by counting the number of uterine branches (<14 branches suggests *T. solium*).

TREATMENT AND PREVENTION

Rx

Taeniasis can be treated with praziquantel in a single dose. Single doses of niclosamide are also effective. Taeniasis can be prevented by inspection of pork. Also, cooking to 56°C for 5 minutes or freezing at –20°C for 7 to 10 days destroys the infective larvae. Current control measures include mass chemotherapy for entire populations with praziquantel. Comprehensive control programs include mass chemotherapy to eliminate tapeworm carriers, treatment of porcine cysticercosis with drugs such as oxfendazole, and improved hygiene to limit access of pigs to human fecal material.

TISSUE CESTODE (CYST) INFECTION

Taenia solium (Cysticercosis)

DEFINITION

T. solium is the cause of human larval infection termed cysticercosis. The normal hosts for the larval (cysticercus) forms are pigs. When ingested by

pigs, the ova hatch, releasing the invasive larvae (termed *oncospheres*), which invade the intestines, migrate to tissues (especially muscle), and mature into cysticercus forms within the tissues. The cysticercus consists of a thin translucent bladder containing an invaginated scolex, which is poised to form a tapeworm after being ingested by a human host. The ova are also infectious to people, including the tapeworm carrier. The sticky ova attach to the hands of the tapeworm carrier and are transmitted by the oral route to the carrier or close contacts. After ingestion, the ova can migrate to tissues and form cysts (cysticercosis). The presence of cysticerci in the central nervous system is termed *neurocysticercosis*.[2] Neurocysticercosis includes cysticerci in the brain parenchyma (parenchymal neurocysticercosis) and cysticerci in the ventricles, subarachnoid space, spine, and eye (extraparenchymal neurocysticercosis).

EPIDEMIOLOGY

Cysticercosis is found in all regions of the world where pigs are raised with access to human fecal material. Some estimate that 60 million people are infected with *T. solium* cysticerci. However, exact data on incidence and prevalence are available from only a limited number of studies because of the requirement for neuroimaging studies to make a diagnosis. In the 19th century, infection was highly endemic in Europe. However, with improving standards of living, local transmission is now limited to a few rural areas in southern and eastern Europe. Cysticercosis is widespread in rural areas of Latin America. High prevalence rates have been documented in parts of Mexico, Guatemala, Honduras, Ecuador, Peru, Brazil, and Bolivia. In endemic villages, more than 10% of the population may have abnormalities on neuroimaging studies consistent with neurocysticercosis. Studies have highlighted the importance of cysticercosis in sub-Saharan Africa. Cysticercosis is widespread in India, Nepal, Southeast Asia, and parts of China. In India, neurocysticercosis most commonly is manifested with single enhancing lesions, which are the main lesion associated with seizure disorders in that country. In the United States, approximately 2000 cases are diagnosed each year. Most cases are in immigrants from pig-raising villages in Mexico and Latin America. However, there are also imported cases from Asia and a few locally acquired infections.

PATHOBIOLOGY

For cysticercosis, the pathogenesis and pathophysiology vary with the location of the cysticerci and the host inflammatory response. Cysticerci in the brain parenchyma initially suppress the host inflammatory response. After a silent period, estimated to be several years, the cysticerci lose the ability to suppress the host inflammatory response, leading to parenchymal inflammation, which typically is manifested by seizures. The cysticerci induce a granulomatous response, which gradually degrades the parasites. In some cases, the lesions resolve. However, in others, degradation leads to formation of calcified granulomas. These calcified lesions may intermittently become inflamed (as evidenced by edema or contrast enhancement on magnetic resonance imaging [MRI] scan) and may cause recurrent seizures during a period of years. In some cases, cysticerci develop within the ventricles of the brain and can mechanically cause obstructive hydrocephalus. Cysticerci in the subarachnoid space may cause a chronic arachnoiditis, which can be manifested by vasculitis and stroke, communicating hydrocephalus, basilar meningitis, and, in some cases, mass effect. Cysticerci can also develop in the spine (manifested as radiculitis), eye, subcutaneous tissue, and muscle.

CLINICAL MANIFESTATIONS

The clinical manifestations vary with the location of the cysticerci and the associated host response (Fig. 354-1).[3,4] Neurocysticercosis can lead to a spectrum of cognitive abnormalities, ranging from impairment in a single domain, to cognitive impairment, and occasionally to dementia. All forms of disease may be associated with headaches. In general, parenchymal cysticerci are associated with seizures,[5] whereas ventricular and subarachnoid cysticercosis are associated with hydrocephalus.

Single Enhancing Lesion

A single enhancing lesion is the most common manifestation of cysticercosis in India and the United States. Patients typically present with seizures, which can be focal or focal with secondary generalization. Many will have a single seizure or a few seizures during the period when the cysticercus is degenerating, but the duration of seizures is eventually self-limited in most cases. However, a few go on to develop calcified lesions, which are a risk factor for recurrent seizures.

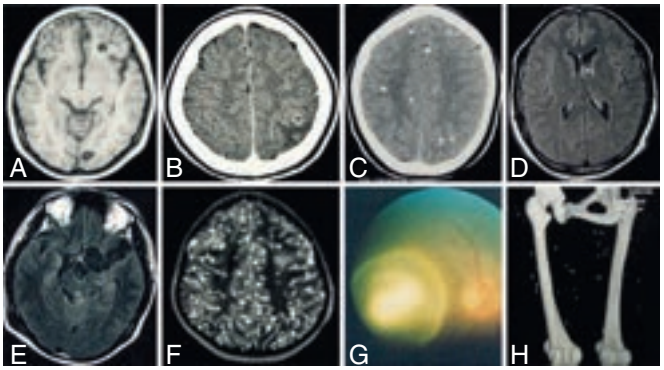

FIGURE 354-1. Human neurocysticercosis can be classified on the basis of neuroimaging studies. A, Multiple cystic lesions. **B,** Single enhancing lesion. **C,** Multiple calcifications. **D,** Intraventricular cysticerci. **E,** Subarachnoid cysticerci. **F,** Diffuse infection with cerebral edema, termed cysticercal encephalitis. **G,** Ocular cysticerci. **H,** Diffuse muscle calcifications. (Reprinted from Garcia HH, Del Brutto OH. Neurocysticercosis: updated concepts about an old disease. *Lancet Neurol.* 2005;4:653-661.)

Multiple Parenchymal Cysticerci

In patients with multiple lesions, the main presentation is with seizures, associated with parenchymal inflammation. In contrast to those with single lesions, seizures are more likely to recur.

Calcified Lesions

Many patients do not present until after they have calcified lesions. Patients with calcified lesions may develop recurrent seizures during a period of years.

Ventricular Cysticerci

The cysticerci typically are manifested by obstructive hydrocephalus. The patient may present with headache, nausea and vomiting, dizziness, altered mental status, or papilledema with altered vision. This is a medical emergency and can be fatal if it is not treated.

Subarachnoid Cysticerci

Cysticerci in the basilar cisterns are often accompanied by cysticerci in other locations, including parenchymal cysticerci or calcifications, ventricular cysticerci, and spinal or ocular cysticerci; patients may present with disease attributable to cysticerci at these sites. Cysticerci in the basilar cisterns are particularly prone to cause arachnoiditis. Manifestations of arachnoiditis may include vascular involvement (large- or small-vessel strokes), meningeal signs, or communicating hydrocephalus (headaches, nausea, vomiting, dizziness, and altered mental status).

DIAGNOSIS

The major clinical manifestations of neurocysticercosis (e.g., seizures and hydrocephalus) are not specific, and it is difficult to identify the parasites. The main tools used to diagnose neurocysticercosis are neuroimaging studies. Computed tomography (CT) scans are sensitive for identification of parenchymal calcifications, which appear as 2- to 5-mm nodules. CT may also reveal parenchymal cysticerci or obstructive hydrocephalus. MRI scans are more sensitive for identification of the cysticerci, especially in the subarachnoid space and ventricles. Three-dimensional fast imaging employing steady-state acquisition (FIESTA) sequences are particularly effective for ventricular cysticerci. The cysticerci are typically round, 1 to 2 cm in diameter. The cyst fluid is usually isodense with spinal fluid. In uninflamed cysticerci, the walls may not be visible. However, most cases demonstrate enhancement of the cyst walls or surrounding tissues and associated edema. In some cases, the scolex may be visible as a 1- to 2-mm solid nodule, cylinder, or spiral on the side of the cystic lesions. Serologic tests are useful to confirm the diagnosis. Assays using crude antigen, including enzyme-linked immunosorbent assay, are associated with poor sensitivity and specificity and are not reliable. An immunoblot assay using semipurified membrane glycoproteins is highly specific for the diagnosis. The sensitivity is excellent in cases with extraparenchymal or multiple parenchymal cysticerci. However, the sensitivity is poor in those with single enhancing lesions or just calcifications. Antigen detection assays are sensitive and more specific for viable cases and are increasingly available in the United States.

TREATMENT

Rx

Treatment varies with the clinical manifestations and form of infection. Seizures should be treated with antiepileptic drugs (Chapter 403). Phenytoin and carbamazepine are typically used and can control the seizures. Newer antiepileptic drugs may be more effective. There are no viable parasites in those with just calcified lesions, so the main measure is to control symptoms (e.g., antiepileptic drugs for those with seizures). In patients presenting with hydrocephalus, surgery to reestablish cerebrospinal fluid flow is the critical initial step in management. The role of antiparasitic drugs (Chapter 344) varies with the form of infection. Randomized controlled trials have demonstrated more rapid resolution of parenchymal cystic lesions with fewer generalized seizures in those treated with corticosteroids and albendazole (15 mg/kg/day in two daily doses) compared with those treated with placebo. Praziquantel (50 to 100 mg/kg/day in three daily doses) can be used as an alternative. There are emerging data on the use of the two drugs in combination, which may have greater cysticidal activity. A number of clinical trials in subjects with single enhancing lesions have demonstrated a slightly more rapid radiologic resolution and fewer seizures in those treated with steroids or antiparasitic drugs. However, the benefits of even high-dose steroids are not dramatic. For patients with cysticerci in the ventricles, management usually involves removal of the cysticerci, which can be best achieved by neuroendoscopy.[6] However, the main alternative approach is placement of a ventriculoperitoneal shunt. Chronic steroids or antiparasitic drugs may decrease the rate of shunt failure, which usually results from clogging of the shunts by the cysticerci or proteinaceous debris. There are no controlled trials on the management of subarachnoid cysticercosis. However, expert opinion supports treatment of subarachnoid cysticercosis with prolonged courses of antiparasitic drugs (e.g., albendazole for months), higher doses of albendazole, or combinations of albendazole and praziquantel. Chronic anti-inflammatory medications (e.g., prednisone 1 mg/kg/day, or 24 mg/day of dexamethasone) are also critically important. Patients who will be treated with chronic steroids should be screened for *Mycobacterium tuberculosis* and *Strongyloides* infections before initiation of steroid therapy. Methotrexate is increasingly recommended as a steroid-sparing agent. Patients with hydrocephalus should be treated by cerebrospinal fluid diversion (e.g., ventriculoperitoneal shunting). Before treatment with antiparasitic drugs, patients should undergo a funduscopic examination. Intraocular parasites may develop brisk inflammatory responses after treatment with antiparasitic drugs. Because this inflammation could lead to blindness, most authorities recommend extraction of the parasites before antiparasitic therapy. However, there are also reports of treatment of intraocular parasites with antiparasitic drugs.

PROGNOSIS

The prognosis varies significantly between the different forms of neurocysticercosis. Parenchymal enhancing lesions and parenchymal cystic lesions will eventually resolve, but this may take months to years. Patients who have or develop calcifications are at lifelong risk for recurrent seizures. Patients with ventricular or subarachnoid disease are at high risk for morbidity and mortality. However, a recent case series noted no deaths with optimal management.

Cystic Hydatid Disease (*Echinococcus granulosus* group)

DEFINITION

Cystic echinococcosis (CE), also called cystic hydatidosis, is caused by the larval stage of cestodes of the *Echinococcus granulosus* complex. All of these parasites were initially thought to be a single species, *E. granulosus*. However, molecular studies demonstrate that *E. granulosus* comprises a number of different species and genotypes. In humans, the clinical manifestations range from asymptomatic infection to severe, potentially fatal disease.

Echinococcal cysts consist of a periparasitic host tissue (pericyst or adventitia), which encompasses the larval endocyst, and the endocyst itself. The endocyst has an outer, acellular laminated layer and an inner, or germinative, layer that gives rise to brood capsules and protoscolices. The cyst is filled with clear fluid, numerous brood capsules, and protoscolices. Some cysts may also harbor daughter cysts of variable size. The protoscolices convert to tapeworms in the canine definitive hosts but can also form new cysts when released in mammalian tissues.

EPIDEMIOLOGY

E. granulosus species occur on all continents and in circumpolar, temperate, subtropical, and tropical zones. The highest prevalence of the parasite is found in parts of Eurasia, Africa, Australia, and South America. Within the endemic zones, the prevalence of the parasites varies from sporadic to high, but only a few countries can be regarded as being free of *E. granulosus*.

It is difficult to determine the true incidence of CE because of the slow rate of growth and variable clinical presentation. Most epidemiologic reports are based on hospital- and surgery-based surveys that greatly underestimate the actual rates of infection, especially in low socioeconomic groups with limited access to diagnosis and treatment.

Since the mid-1980s, however, mass community-based surveys using portable ultrasound scanners have been conducted in many remote, rural areas of the world. The sensitivity and specificity of ultrasound have been shown to be superior to those of serology in prevalence surveys. These studies showed the real burden of disease, uncovering population infection rates of up to 6.6%.

E. granulosus exists as a complex of species and strains that differ in a variety of criteria that may have an impact on the epidemiology, pathology, and control of CE. To date, 10 distinct genotypes (G1 to G10) have been identified. Some distinct species have been identified (*Echinococcus equinus*, *Echinococcus ortleppi*). The great majority of *E. granulosus* isolates from human patients thus far have been of the sheep genotype (G1).

CLINICAL MANIFESTATIONS

The presentation of human CE is protean.[6] Patients seek medical attention when a large cyst has some mechanical effect on organ function or rupture of a cyst causes acute hypersensitivity reactions. The cyst is often diagnosed incidentally during ultrasound examination, chest radiography, or body scanning performed for other clinical reasons. The liver is the most frequent location of echinococcal cysts, representing approximately 70% of cases.[7,8] The lungs are the second most common location. However, CE can occasionally occur in virtually any other organ.

Common symptoms are upper abdominal discomfort and pain, poor appetite, and a mass in the abdomen. Physical findings are hepatomegaly, a palpable mass on the surface of the liver or other organs, and abdominal distention. Other manifestations include jaundice, colic-like pains, portal hypertension, ascites, and compression of the inferior vena cava. If cysts in the lung rupture into the bronchi, symptoms may include intense cough, a salty taste in the mouth, or vomiting of hydatid material and cystic membranes. Patients may present with a chest mass, chest pain, chronic cough, pneumothorax, eosinophilic pneumonitis, pleural effusion, parasitic lung embolism, hemoptysis, or biliptysis. Cysts in the heart can cause a cardiac mass, pericardial effusion, and embolism. Cysts in the breast must be differentiated from neoplasms. Cysts located in the spine and in the brain can cause serious neurologic symptoms, including paralysis and seizures.

DIAGNOSIS

The diagnosis of CE is based on imaging methods and on serology, but serology has only a confirmatory role. Routine laboratory tests are nonspecific. Cyst rupture into the biliary tree may cause elevation of alkaline phosphatase, sometimes in association with hyperamylasemia and eosinophilia (up to 60%). Unless the cyst has ruptured, eosinophilia is low grade or absent.

Imaging

Modern imaging tools (ultrasound, CT, and, to a lesser extent, MRI) are central to the diagnosis and clinical management of CE. Ultrasound is the procedure of choice for diagnosis of asymptomatic CE. Ultrasound is also useful for longitudinal studies, such as monitoring the response of cysts to treatment and recording cyst growth rate. In 2003, the World Health Organization (WHO) Informal Working Group on Echinococcosis proposed a standardized ultrasound classification (Fig. 354-2). This classification defines six cyst stages that are assigned to three clinical groups. The active group comprises developing cysts, which may be unilocular (CE1) or multivesicular with daughter cysts (class CE2) and are usually found to be viable. The transitional group (class CE3) contains cysts that are usually starting to degenerate. There are two types of CE3: the "water lily sign" for floating membranes, which is now known as subclass CE3a; and predominantly solid cysts with daughter cysts, or subclass CE3b. This subdivision is based on their different response to percutaneous treatment (see later) and albendazole, which is generally good for CE3a and poor for CE3b. A study using nuclear magnetic resonance spectroscopy has found that CE3a and CE3b may have different metabolic characteristics. The inactive group (classes CE4 and CE5) exhibits involution and signs of solidification of cyst content with increasing degrees of calcification and is nearly always found to be nonviable.

CT scanning has the advantage of inspecting any organ, detecting smaller cysts located outside the liver, locating cysts precisely, and sometimes differentiating parasitic from nonparasitic cysts. MRI may have some advantages

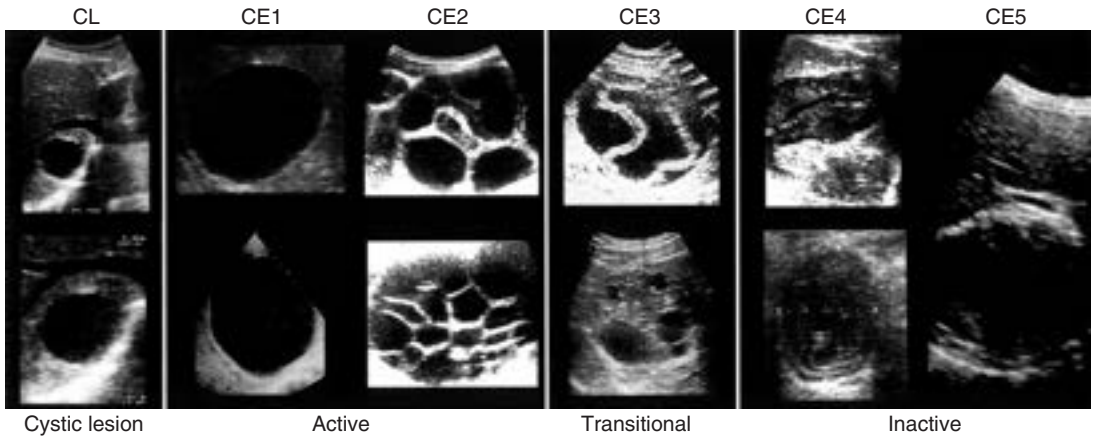

FIGURE 354-2. WHO Informal Working Group on Echinococcosis standardized ultrasound classification of cystic echinococcosis. CL lesions are cystic lesions lacking a distinct wall and may have other diagnoses. CE1 lesions are cystic lesions with a visible wall that may demonstrate protoscolices ("hydatid sand"). CE2 lesions include internal septation. CE3 lesions may be detached from the wall (CE3a, top row) or have daughter cysts with internal thickening (CE3b, bottom row). CE4 lesions are heterogeneous lesions with degeneration. CE5 lesions show thick calcification.

over CT scanning in the evaluation of postsurgical residual lesions, recurrences, and selected extrahepatic infections, such as cardiac infections. Furthermore, a study has shown that MRI reproduces the ultrasound-defined features of CE better than CT does. If ultrasound cannot be performed because of cyst location or patient-specific reasons, MRI with heavily T2-weighted series is preferable to CT. Plain radiographs are used for cysts in the lungs, bone, and muscle and for detection of calcified cysts.

Serology

Serologic tests are useful for confirmation of presumptive imaging diagnoses. However, many tests are available, and they are not standardized. Their sensitivity varies with the location of the cysts. Hepatic cysts are more likely to elicit an immune response than are pulmonary, brain, or splenic cysts. Serologic test results are usually positive when the endocyst is detached (CE3a) and in active (CE2) and transitional (CE3b) stages. Serologic test results are generally negative in patients with inactive cysts (CE4 and CE5). Titers tend to slowly decrease when a cyst becomes inactive (CE4, CE5) and after radical surgery. Titers may remain positive after conservative surgery in which the antigen source (the germinal layer) is not completely removed. Antibody titers usually increase immediately after medical or percutaneous treatments because of the mobilization of the antigen following disruption of cyst integrity.[9]

Other Diagnostic Procedures

Fine-needle aspiration of the cyst performed under ultrasonographic guidance, with a transhepatic approach, under anthelmintic coverage, is useful for differentiation of CE, malignant neoplasms, abscesses, and nonparasitic cysts. The procedure must be carried out in the presence of an anesthesiologist ready to manage the rare but possible anaphylactic reaction.

TREATMENT

Rx

The appropriate treatment depends on factors of the individual patient, the characteristics of the cyst, the therapeutic resources available, and the physician's preference. There are few randomized clinical trials evaluating treatment options, so a low level of evidence supports one therapeutic modality over another.

Surgery

Surgery has long been the only option in the treatment of CE. However, in the past two decades, medical treatment, percutaneous procedures, and a "watch and wait" approach have been successfully introduced and replaced surgery as the treatment of choice in selected cases. Surgery remains the main therapy in complicated cysts (i.e., those with rupture, biliary fistula, compression of vital structures, superinfection, or hemorrhage), cysts at high risk of rupture, or large cysts with many daughter vesicles that are not suitable for percutaneous treatments. Surgery can be performed as an open procedure, with either radical or conservative techniques, or laparoscopically, but there are controversies as to the safest and most effective technique and in which cases it should be applied. In all cases, perioperative albendazole prophylaxis, from 1 week before surgery until 4 weeks postoperatively, is recommended as a cautionary measure to minimize the risk of fluid spillage and consequent secondary echinococcosis from seeding of protoscolices in the abdominal cavity. Some authorities treat with praziquantel as well.

Percutaneous Treatments

Percutaneous techniques provide an alternative to surgery and benzimidazole derivatives. These treatment modalities aim either to destroy the germinal layer with scolicidal agents or to evacuate the entire endocyst. The most popular method aimed at destroying the germinal layer is PAIR (puncture, aspiration, injection of a scolicidal agent, and reaspiration). Many modified catheterization techniques are used to evacuate the endocyst and are generally reserved for cysts that are difficult to drain or tend to relapse after PAIR, such as multivesiculated cysts or cysts with predominantly solid content and daughter cysts. A growing number of articles have reported its safety in treating abdominal, especially liver, echinococcal cysts. In a study on 5943 percutaneous punctures of echinococcal cysts, lethal anaphylaxis occurred in 0.03% of procedures, whereas reversible allergic reactions complicated 1.7% of procedures. Prophylactic administration of albendazole for at least 30 days after puncture is a cautionary measure that should always accompany PAIR. PAIR is generally successful at inducing permanent solidification in CE1 and CE3a cysts. A few reports with long-term follow-up indicate that multivesicular cysts (i.e., CE2 and CE3b) tend to relapse repeatedly after PAIR.

Chemotherapy

Albendazole (Chapter 344) is the antiparasitic drug of choice for CE. It is administered orally at a dosage of 10 to 15 mg/kg/day; administration should be continuous without treatment interruptions. However, the optimal dosage and optimal duration of treatment with albendazole have not been formally assessed, and data from the small clinical trials generally fail to take into account the cyst's characteristics. A recent systematic review on the effect of albendazole showed that the efficacy of the drug may have been overstated in previous retrospective, nonrandomized studies. Albendazole induces solidification in small and medium-sized CE1 and CE3a cysts, whereas it has generally little effect on giant (diameter > 10 cm) CE1 and CE3a cysts. It has no effect on most cases of CE2 and CE3b cysts.

Adverse effects of benzimidazoles (Chapter 344) include hepatotoxicity, leukopenia, thrombocytopenia, and alopecia. Increases in aminotransferases may be due to drug-related efficacy or real drug-related toxicity. Whereas teratogenic risks are theoretical, it is nonetheless good practice to avoid use during pregnancy when possible and to delay treatment until after delivery unless it is absolutely necessary.

Watch and Wait

Inactive liver cysts that are free of complications, such as compression on neighboring organs, are increasingly monitored without being treated.[10] Prospective studies need to be carried out to confirm the safety of this option.

Follow-up

Follow-up is crucial to evaluate the efficacy of treatment. Long-term follow-up, generally more than 5 years, is required to evaluate local recurrences, which have been reported up to 10 years after apparently successful treatment. When the combination of imaging and serology is inconclusive, fine-needle aspiration should be performed to ascertain the viability of the cyst contents.

Alveolar Hydatid Disease (*Echinococcus multilocularis*)

Alveolar hydatid disease is caused by the tissue forms of *Echinococcus multilocularis*. In tissues, typically the liver, *E. multilocularis* grows as a budding mass rather than as a large cystic lesion. The tissues resemble lung tissues, hence the name "alveolar." The normal definitive hosts are canines, including wolves and foxes. The normal intermediate hosts are rodents. Humans are accidentally infected by contact with soil containing the ova.

EPIDEMIOLOGY

E. multilocularis is endemic in arctic and alpine areas of the Northern Hemisphere. It is highly endemic in western China, Tibet, and central Asia. In recent years, *E. multilocularis* has emerged as an important problem in Alpine areas of central Europe and adjacent forested areas.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Human *E. multilocularis* infection almost invariably involves the liver, in which it is manifested as a tumor-like mass that gradually expands during decades. The main symptoms are liver discomfort and swelling. Diagnosis is by demonstration of a characteristic mass on imaging studies, with the etiology confirmed by serologic tests.[11]

TREATMENT AND PROGNOSIS

Rx

Surgery remains the mainstay of treatment of *E. multilocularis*. When feasible, all infected tissues should be removed. Apparently curative therapy should be followed by a 2-year course of albendazole to decrease the risk of relapse. In some cases, resection is feasible only when it is accompanied by liver transplantation. In cases that are not amenable to surgical resection, prolonged courses of albendazole can suppress growth of the lesion. After treatment with benzimidazoles, mortality is similar to that of the age- and-sex-matched general population.

Other Larval Cestode Infections

Sparganosis is caused by infection with the larval (plerocercoid) stage of *Spirometra mansonoides*. Infection is acquired by ingestion or application of infected meat (frogs, birds, fish) or exposure of skin to infected flesh (e.g., poultices of infected tissues). After infection, plerocercoids develop in the tissues, typically presenting as subcutaneous or central nervous system nodules or occasionally larva migrans symptoms. Treatment usually involves removal of the nodule.

Coenurosis is a rare larval cestode infection caused by human infection with the larval stage of the dog tapeworms *Taenia multiceps* and *Taenia serialis*. In the tissue, the larva forms a cystic lesion containing multiple scolices (the coenurus). The cystic lesion is usually single and most frequently identified in brain, eye, or soft tissues. Treatment usually involves removal.

Echinococcus oligarthrus and *Echinococcus vogeli* have been associated with polycystic hydatid disease in northern South America. *Taenia crassiceps* has been identified in the eye and in tissues of compromised hosts. *Hymenolepis*-like organisms have also been identified in tissues of AIDS patients.

Grade A References

A1. Baird A, Wiebe S, Zunt JR, et al. Evidence-based guideline: treatment of parenchymal neurocysticercosis: report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology*. 2013;80:1424-1429.

A2. Garcia HH, Gonzales I, Lescano AG, et al. Efficacy of combined antiparasitic therapy with praziquantel and albendazole for neurocysticercosis: a double-blind, randomised controlled trial. *Lancet Infect Dis*. 2014;14:687-695.

A3. Otte WM, Singla M, Sander JW, et al. Drug therapy for solitary cysticercus granuloma: a systematic review and meta-analysis. *Neurology*. 2013;80:152-162.

A4. Garcia HH, Gonzales I, Lescano AG, et al. Enhanced steroid dosing reduces seizures during antiparasitic treatment for cysticercosis and early after. *Epilepsia*. 2014;55:1452-1459.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Chai JY. Praziquantel treatment in trematode and cestode infections: an update. *Infect Chemother*. 2013;45:32-43.
2. Del Brutto OH. Neurocysticercosis. *Handb Clin Neurol*. 2014;121:1445-1459.
3. Garcia HH, Nash TE, Del Brutto OH. Clinical symptoms, diagnosis, and treatment of neurocysticercosis. *Lancet Neurol*. 2014;13:1202-1215.
4. Serpa JA, Graviss EA, Kass JS, et al. Neurocysticercosis in Houston, Texas: an update. *Medicine (Baltimore)*. 2011;90:81-86.
5. Moyano LM, Saito M, Montano SM, et al. Neurocysticercosis as a cause of epilepsy and seizures in two community-based studies in a cysticercosis-endemic region in Peru. *PLoS Negl Trop Dis*. 2014;8:e2692.
6. Brunetti E, Kern P, Vuitton DA. Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. *Acta Trop*. 2010;114:1-16.
7. Rinaldi F, Brunetti E, Neumayr A, et al. Cystic echinococcosis of the liver: a primer for hepatologists. *World J Hepatol*. 2014;6:293-305.
8. Touma D, Sersté T, Ntounda R, et al. The liver involvement of the hydatid disease: a systematic review designed for the hepato-gastroenterologist. *Acta Gastroenterol Belg*. 2013;76:210-218.
9. McManus DP, Gray DJ, Zhang W, et al. Diagnosis, treatment, and management of echinococcosis. *BMJ*. 2012;344:e3866.
10. Piccoli L, Tamarozzi F, Cattaneo F, et al. Long-term sonographic and serological follow-up of inactive echinococcal cysts of the liver: hints for a "watch-and-wait" approach. *PLoS Negl Trop Dis*. 2014;8:e3057.
11. Piarroux M, Piarroux R, Giorgi R, et al. Clinical features and evolution of alveolar echinococcosis in France from 1982 to 2007: results of a survey in 387 patients. *J Hepatol*. 2011;55:1025-1033.

REVIEW QUESTIONS

1. Active cysts in the WHO Informal Working Group on Echinococcosis classification are

A. CE1
B. CE1, CE2
C. CE3a, CE3b
D. CE4
E. CE4,CE5

Answer: B CE1 (unilocular) and CE2 (cysts with daughter cysts) are categorized as viable or active in the WHO Informal Working Group on Echinococcosis. When aspirated, viable protoscolices are found in the fluid.

2. The risk of lethal anaphylactic shock after an echinococcal cyst puncture is

A. 2%
B. 3%
C. 4%
D. 5%
E. <1%

Answer: E In an analysis of the literature that included 5943 percutaneous puncture procedures of 5517 hepatic and nonhepatic echinococcal cysts, two cases of lethal anaphylaxis and 99 reversible anaphylactic reactions were reported. Lethal anaphylaxis occurred in 0.03% of percutaneous puncture procedures, corresponding to 0.04% of treated cysts, whereas reversible allergic reactions complicated 1.7% of percutaneous punctures, corresponding to 1.8% of treated echinococcal cysts. Lethal anaphylaxis related to percutaneous treatment of cystic echinococcosis is an extremely rare event and is observed no more frequently than drug-related anaphylactic side effects are.

3. The best imaging technique for staging of cystic echinococcosis (CE) cysts of the liver is

A. Computed tomography (CT)
B. Magnetic resonance imaging (MRI)
C. Ultrasound
D. Positron emission tomography (PET) scan
E. Scintigraphy

Answer: C Imaging plays the key role in diagnosis and staging of CE. The description of CE-specific imaging features and the WHO CE cyst classification are based on ultrasound. The reproducibility of the ultrasound-defined features of CE cysts is variable in MRI and CT. This is of particular importance for cysts that are not accessible by ultrasound and because of the increasing availability and overuse of CT and MRI.

A retrospective study was conducted in patients with abdominal CE cysts seen on ultrasound who had additional CT or MRI scans performed; it showed that MRI reproduces the ultrasound-defined features of CE better than CT does. If ultrasound cannot be performed because of cyst location or patient-specific reasons, MRI with heavily T2-weighted series is preferable to CT.

4. In a patient presenting with seizures and calcified cysticercosis with surrounding edema on MRI, the most appropriate therapy is

A. Antiepileptic drugs alone
B. Antiepileptic drugs and albendazole
C. Antiepileptic drugs and prednisone
D. Antiepileptic drugs, albendazole, and prednisone
E. Placement of a ventriculoperitoneal shunt

Answer: A Patients with calcified lesions no longer have viable cysticerci and thus do not benefit from antiparasitic drugs. Prednisone may transiently decrease edema, but there is no evidence of clinical benefit, and steroid withdrawal has been associated with flares of disease.

5. In a patient presenting with multiple subarachnoid cysticerci complicated by basilar meningitis, which of the following approaches have been associated with the best outcomes?

A. Treatment with a 2-week course of albendazole and dexamethasone
B. Removal of the cysticerci by open craniotomy
C. Treatment with prolonged courses of albendazole, dexamethasone, and methotrexate
D. Neuroendoscopic removal of intraventricular cysticerci
E. Cerebrospinal fluid diversion alone for hydrocephalus

Answer: C Subarachnoid neurocysticercosis responds poorly to the doses of albendazole used for parenchymal infection. Use of cerebrospinal fluid diversion alone carries a high case-fatality rate. The optimal approach includes prolonged courses of antiparasitic drugs along with anti-inflammatory treatment. Higher doses of albendazole, combinations of praziquantel and albendazole, and endoscopic debulking (removal of cysticerci in the basilar cistern, not just the ventricles) have also been reported as helpful.

355

SCHISTOSOMIASIS (BILHARZIASIS)

EDGAR M. CARVALHO AND ALDO A. M. LIMA

DEFINITION

Schistosomiasis, which is caused by trematodes of the genus *Schistosoma*, is one of the most important parasitic diseases of humans and is a global public health problem in the developing world.[1] *S. mansoni, S. haematobium, S. japonicum, S. intercalatum,* and *S. mekongi* are the five major species of *Schistosoma* affecting humans. Other *Schistosoma* species that occasionally infect humans include *S. bovis, S. mattheei,* and some avian schistosomes.

EPIDEMIOLOGY

Schistosomiasis occurs mainly in tropical and subtropical areas, especially in poor communities without access to safe drinking water and adequate sanitation.[2] It is estimated that 200 million people are infected by the helminth and 600 to 799 million are at risk of infection.[3] *S. mansoni* is found in 55 countries, mainly in Africa, the Middle East, the Caribbean, Brazil, Venezuela, and Suriname. *S. haematobium* is endemic in 53 countries in the Middle East and most of the African continent. *S. japonicum* is endemic in China, Indonesia, and the Philippines. *S. intercalatum* has been reported from 10 countries in Africa. *S. mekongi* is found in Cambodia and Laos.

The endemicity of schistosomiasis depends on the urban disposal of urine (*S. haematobium*) and feces (*S. mansoni, S. japonicum, S. intercalatum, S. mekongi*), the presence of suitable snail hosts, and human exposure to cercariae.[4] The freshwater snail intermediate hosts are *Biomphalaria* sp in Africa and *Biomphalaria glabrata (Australorbis)* and *Tropicarbis* in South America and the West Indies. In some cases, the endemicity of schistosomiasis may be maintained by animal reservoirs. Such is the case with *S. japonicum*, which infects dogs and cows. Rodents, monkeys, and baboons have been found infected in nature, but the role of these animals as reservoirs does not seem to be epidemiologically important.

Etiology and Life Cycle

The schistosomes are digenetic parasitic trematodes (Fig. 355-1). Although they are morphologically distinct, the species of *Schistosoma* that infect humans share some common characteristics. The large male (0.6 to 2.2 cm × 2 to 4 mm) has a ventral gynecophoric canal, in which the female (1.2 to 2.6 cm × 1 to 2 mm) is held during copulation. The sequencing of the *S. mansoni* genome has been determined.

Adult worms live in the mesenteric veins (*S. mansoni, S. japonicum, S. mekongi,* and *S. intercalatum*) or in the venous plexus around the lower ends of the ureters and the urinary bladder (*S. haematobium*). In these sites, they start their sexual reproduction by releasing eggs. Once deposited in the host, eggs can stay in the mesenteric vein, be trapped in the intestines, escape to the

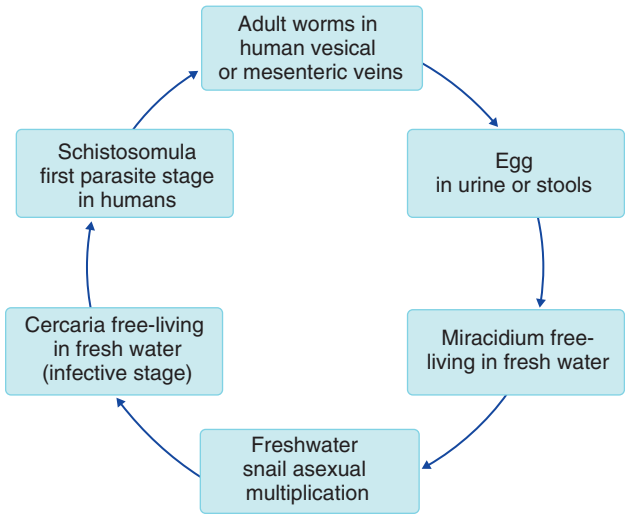

FIGURE 355-1. Schistosome life cycle.

intestinal lumen, and migrate by portal blood to the liver (*S. mansoni, S. japonicum*). Eggs of *S. haematobium* can be trapped in the intestines and bladder, escape to the intestinal or bladder lumen, and be trapped in the female genital tract. After being excreted with feces or urine into fresh water, the eggs hatch and release ciliated motile miracidia that penetrate into the snail intermediate host. Following asexual multiplication in the snail, the development of cercariae, the infective forms for humans, takes 4 to 7 weeks. After leaving the snails, the cercariae can survive in fresh water for almost 72 hours. When penetration of the skin in the human host occurs, the cercariae lose their tails and change into schistosomula. Schistosomula migrate to the lungs and in about 6 weeks mature into adult worms, after which they descend to their final habitat. Viable eggs can be seen in excretions (i.e., stool or urine) 5 to 9 weeks after cercarial penetration. The lifespan of the worms ranges from 5 to 10 years.

PATHOBIOLOGY

The pathogenesis of acute human schistosomiasis is mainly related to egg deposition and liberation of antigens of adult worms and eggs. A strong inflammatory response characterized by high levels of pro-inflammatory cytokines, such as interleukins 1 and 6 and tumor necrosis factor-α, and by circulating immune complexes participates in the pathogenesis of the acute phase of the disease.

In chronic schistosomiasis, tissue injury is mediated by egg-induced granulomas and the subsequent appearance of fibrosis. Because the habitat of *S. mansoni, S. japonicum, S. mekongi,* and *S. intercalatum* worms is the mesenteric blood vessels, the intestines are involved primarily, and egg embolism results in secondary involvement of the liver. Host genetics, immunologic response, and parasite load measured by egg count in the stool are associated with a greater chance of liver fibrosis resulting in hepatosplenomegaly. Polymorphisms within the interferon-γ receptor 1 (*IFNGR1*) gene may be related to severe hepatic disease caused by *S. mansoni* infection. There is also evidence that increased levels of interleukin 5 and tumor necrosis factor-αare associated with fibrosis.[5] The granuloma is mainly composed of macrophages, eosinophils, fibroblasts, and collagen deposition. The size of these granulomas and the resulting fibrosis lead to most of the chronic fibro-obstructive lesions in schistosomiasis. In the liver, the granulomas result in perisinusoidal obstruction of portal blood flow, portal hypertension, splenomegaly, esophageal varices, and portosystemic collateral circulation. Liver cell perfusion is consequently preserved, and liver function test results remain normal well into the course of the disease. In *S. haematobium* infection, eggs and granuloma formation are predominantly found in the ureters, bladder, and genital tract.

In schistosome-infected persons, the intensity of infection increases during the first two decades of life as children accumulate worms, with infection intensity declining thereafter. In the *S. haematobium*–infected population, IgE increases progressively with age, and IgE antibodies directed against adult worm antigens are associated with subsequent low intensities of reinfection. Alternatively, in subjects who are highly exposed to contaminated water but have negative stool examination results, there is evidence of higher interferon-γ production in response to *S. mansoni* antigens.[6] The existence of a major codominant gene, called *SM1*, which controls the intensity of infection by *S. mansoni*, has been demonstrated. *S. mansoni* infection levels are controlled by a locus that maps to chromosome 5q31-q33, which is close to a locus regulating IgE levels, indicating that genetic factors are probably critical to susceptibility and resistance to schistosome infection.

Because modulation of the immune response is a characteristic of chronic schistosomiasis, *S. mansoni* infection attenuates the clinical manifestations of autoimmune and inflammatory diseases, but it may also impair the immunologic response to vaccines and change manifestations of other infectious diseases.

CLINICAL MANIFESTATIONS

Clinical manifestations of schistosomiasis are divided into schistosome dermatitis, acute schistosomiasis, and chronic schistosomiasis. Schistosome dermatitis, or swimmer's itch, is an uncommon manifestation seen mainly when avian cercariae penetrate the skin and are destroyed. Schistosome dermatitis is a sensitization phenomenon occurring in previously exposed persons. The cercariae evoke an acute inflammatory response with edema, early infiltration of neutrophils and lymphocytes, and later invasion of eosinophils. A pruritic papular rash occurs within 24 hours after the penetration of cercariae and reaches maximal intensity in 2 to 3 days.

Acute schistosomiasis occurs usually 20 to 50 days after primary exposure. Although it is asymptomatic in endemic areas, acute schistosomiasis is becoming a frequent and major clinical problem in nonimmune individuals from urban regions who are exposed for the first time to a heavy infectious dose in an endemic area. The clinical syndrome (i.e., fever, chills, liver and spleen enlargement, and marked eosinophilia) originally described for *S. japonicum* infection, and still common for this species, is increasingly being diagnosed in Brazil in individuals with *S. mansoni* infection. Malaise, diarrhea, weight loss, cough, dyspnea, chest pain, restrictive respiratory insufficiency, and pericarditis are important findings in this phase. High levels of circulating immune complexes correlate with respiratory manifestations, and tumor necrosis factor-αlevels correlate with the presence of abdominal pain, diarrhea, and weight loss. Abdominal ultrasound may show hepatosplenomegaly. Acute disease is not observed in individuals living in endemic areas of schistosomiasis because of the down-modulation of the immune response by antigens or idiotypes transferred from mother to child.

In chronic schistosomiasis, abdominal pain, irregular bowel movements, and blood in the stool are the main symptoms of intestinal involvement. Colonic polyposis may occur, especially in Egypt. Hepatosplenic involvement is the most important cause of morbidity with *S. mansoni* and *S. japonicum* infection. Patients may remain asymptomatic until the manifestation of hepatic fibrosis and portal hypertension develops. Hepatic fibrosis is caused by a granulomatous reaction to *Schistosoma* eggs that have been carried to the liver. Hematemesis from bleeding esophageal or gastric varices may occur. In such cases, anemia and decreasing levels of serum albumin are observed. Some patients have severe hepatosplenic disease with decompensated liver disease. Jaundice, ascites, and liver failure are then observed. Concomitant infection by *Salmonella* species, and less extensively by other gram-negative bacteria, with *S. mansoni* or *S. haematobium* leads to a picture of prolonged fever, hepatosplenomegaly, and mild leukocytosis with eosinophilia. Glomerulonephritis, infantilism, and hypersplenism are other complications associated with hepatosplenic schistosomiasis. The detection of pulmonary hypertension is increasing with the use of more advanced diagnostic technology. Pulmonary hypertension, which used to be exclusively linked to the hepatosplenic form of the disease, has been documented in patients without liver fibrosis.[7] In hospitalized adult patients with *S. japonicum* infection, cerebral schistosomiasis occurs in 1.7 to 4.3%.[8] It may occur as early as 6 weeks after infection, and the most common sign is focal jacksonian epilepsy. Signs and symptoms of generalized encephalitis may occasionally be found. In *S. mansoni* infection, neurologic involvement is rare and mainly characterized by transverse myelitis, which occurs mainly in patients without liver fibrosis and hepatosplenomegaly.[9]

In *S. haematobium* infection, the main organ system involved is the urogenital tract. The acute granulomatous response to parasite eggs in the early stages causes urinary tract disease, such as urethral ulceration and bladder polyposis. In chronic disease, usually in older patients, granulomas at the lower end of the ureters obstruct urinary flow and may cause hydroureter and hydronephrosis. Bladder fibrosis and calcification are also seen in this phase. Up to 70% of infected individuals have hematuria, dysuria, or urinary frequency. Urine examination reveals proteinuria and hematuria. Radiologic findings include hydronephrosis; hydroureter; ureteral strictures, dilation, or distortion; ureteral calcifications; ureterolithiasis; calcified bladder; polyps; reduction in bladder capacity; irregular contraction of the bladder wall; or a dilated bladder because of bladder neck fibrosis. An increased incidence of squamous cell carcinoma of the bladder has been reported in endemic areas of *S. haematobium* infection.[10] *S. haematobium* eggs have occasionally been found in the lungs, with subsequent focal pulmonary arteritis and pulmonary hypertension. Genital schistosomiasis has been documented in up to 75% of women in *S. haematobium*–endemic areas[11] and has also been documented in girls.[12] Spontaneous bleeding, burning sensation in the genitals, dyspareunia, itching, tumors due to granulomas, and infertility are the more common complaints. The association of genital schistosomiasis and HIV infection has been recognized. Monocytes and $CD4^+$ T cells as well as high expression of chemokine receptors CCR5 and CXCR4 documented in schistosomiasis may facilitate binding of the virus after penetration through an ulcerated friable epithelium.

DIAGNOSIS

A definitive diagnosis of schistosomiasis can be made only by finding schistosome eggs in feces, urine, or a biopsy specimen, usually from the rectum (Table 355-1). However, a steep decrease of sensitivity is found in low-endemicity areas. A history of contact with contaminated water and appropriate clinical manifestations are important steps in establishing the diagnosis. Because schistosome eggs may be few, concentration by sedimentation should be performed. All eggs from feces, urine, or tissues should be examined under high power to determine their viability by visualizing the activity of cilia of the excretory flame cells of the enclosed miracidium. Dead eggs

TABLE 355-1 DIAGNOSIS OF SCHISTOSOMIASIS

SCHISTOSOME	EGGS	DIAGNOSIS
S. haematobium	Mainly found in urine but may be found in stools or rectal biopsy specimens Eggs: 143 × 50 µm; spindle shaped: rounded anterior, conical posterior, tapering to a terminal delicate spine	Obtain urine sample at midday (when eggs are excreted); more than one sample may be needed Examine urine directly or by filtering 10 mL of urine through a Nuclepore membrane Rectal biopsy in suspected cases with normal urine Serologic testing to diagnose early or light infection
S. mansoni	Eggs: 155 × 66 µm; oval with lateral, long spine	Examine stool for eggs Use the Kato-Katz thick smear method for quantification purposes Rectal biopsy or serologic testing to diagnose stool-negative cases, particularly in lightly infected patients
S. japonicum	Found in stool Eggs: 89 × 67 µm; oval or rounded with a lateral, short, sometimes curved spine	Examine stool for eggs Kato-Katz thick smear (for quantitative assessment) Rectal biopsy for those with light infections, especially with less common manifestations (i.e., cerebral schistosomiasis)
S. mekongi	Found in stool Eggs: 60 × 32 µm; smaller than eggs of *S. japonicum*	Examine stool for eggs
S. intercalatum	Found in stool Eggs: 180 × 65 µm; terminal spine	Examine stool for eggs

may persist for a long time after successful therapy or natural death of the worms. Because the intensity of infection is associated with morbidity, quantitative techniques such as the Kato-Katz thick smear method are recommended for *S. mansoni* and *S. japonicum*. Rectal biopsy may be used for those with light infection. Ultrasonography allows determination of the degree of liver fibrosis. *S. mekongi* and *S. intercalatum* infection is diagnosed by examination of the stool for eggs.

Urine examination for *S. haematobium* eggs can be performed by direct or concentration methods. Samples should be obtained at midday, when excretion of eggs is maximal. Rectal biopsy may be performed in patients with negative urine results. Schistosome real-time polymerase chain reaction is sensitive and specific in urine and stool.[13] After *S. haematobium* infection is diagnosed, assessment of urinary tract disease by ultrasonography is recommended. Because of an increased incidence of carcinoma of the bladder, cancer surveillance should be performed in patients with *S. haematobium* infection.

Serologic tests may help the diagnosis of acute infection because the symptoms are not specific and the finding of eggs in stool may reflect chronic infection.

Quantification of circulating antigens in serum and urine is an alternative for the diagnosis of schistosome infection. However, the sensitivity of the method decreases in patients with light infection (<100 eggs per gram of feces). This test has also been used to monitor the efficacy of antischistosome chemotherapy. A significant decrease in antigen levels or negativity of the test is observed as early as 10 days after therapy.

TREATMENT

Chemotherapy is by far the major method used for prophylaxis, control, and cure of schistosomiasis. Several compounds are in use: metrifonate, oxamniquine, praziquantel, and artemisin derivatives (artesunate and artemether).[14] Praziquantel, a pyrazinoisoquinoline derivative, is the drug of choice for the treatment of schistosomiasis for four reasons: high efficacy against all schistosome species and against cestodes, lack of serious short-term and long-term side effects, administration as a single oral dose, and competitive cost. Recent reviews confirm that a single dose of praziquantel (40 mg/kg) is an effective treatment for *S. mansoni* infection.[A1,A2] Doses lower than 40 mg/kg may be less effective, with no additional benefit for higher doses. Oxamniquine (40 mg/kg) is also effective, and on the basis of current limited evidence, it is uncertain which intervention is more effective. A meta-analysis study also confirmed that artemisinin derivatives, unlike oxamniquine, used in combination with praziquantel increased the cure rates in schistosomiasis treatment, but artemisinin derivatives or oxamniquine alone did not.[A3]

The standard recommended treatment consists of a single dose of praziquantel, 40 mg/kg, for *S. mansoni, S. haematobium,* and *S. intercalatum* infection. In *S. japonicum* infection, a total dose of 60 mg/kg is recommended, split into two or three doses in a single day. Although no significant difference was found in the overall cure rates between single-dose (40 mg/kg) and double treatment (40 mg/kg with 2-week interval) regimens of praziquantel for *S. haematobium*, the effect of double treatment resulted in significant reduction in infection intensity and microhematuria, which may have an impact in reducing morbidity.[A4] *S. mekongi* may require two treatments at 60 mg/kg body weight. With these dosages of praziquantel, recorded cure rates are 75 to 85% for *S. haematobium*, 63 to 85% for *S. mansoni*, 80 to 90% for *S. japonicum*, 89% for *S. intercalatum*, and 60 to 80% for double infections with *S. mansoni* and *S. haematobium*. A decrease in the efficacy of praziquantel has been observed in patients coinfected with human T-cell leukemia virus type 1.

Praziquantel is well tolerated and effective in patients of all ages and for different clinical forms of schistosomiasis, including advanced hepatosplenic cases (*S. mansoni*), cerebral schistosomiasis (*S. japonicum*), and neurologic syndromes (*S. mansoni* and *S. haematobium*), possibly in association with corticosteroids.[15] However, praziquantel has a poor prophylactic effect, which reduces its efficacy in areas of high transmission. There have been several reports of persistent schistosome egg shedding after treatment, thus posing a concern about the emergence of drug resistance.[16] A systematic review and meta-analysis show that either artesunate or artemether has a confirmed prophylactic effect across different trials performed in China. Further studies will be necessary to examine the combination of antischistosome chemotherapy for *S. haematobium* infection in which repeated standard treatment fails to clear the infection.

The most common adverse events observed with praziquantel or oxamniquine are related to the gastrointestinal tract: abdominal pain or discomfort, nausea, vomiting, anorexia, and diarrhea. These symptoms can be observed in up to 50% of patients but are usually well tolerated. Other side effects are related to the central nervous system (e.g., headache, dizziness, drowsiness) and the skin (e.g., pruritus, eruptions) or may be nonspecific (e.g., fever, fatigue). In general, the cumulative experience from a large number of studies allows the conclusion that praziquantel is an extremely well tolerated drug that requires minimal medical supervision and is therefore particularly suitable for mass chemotherapy programs. Although a reduction in the intensity of infection and morbidity has been documented after mass chemotherapy, to control the disease, provision of clean water, use of molluscicides, implementation of adequate sanitation, and improvement of socioeconomic conditions should also be undertaken.

Grade A References

A1. Danso-Appiah A, Olliaro PL, Donegan S, et al. Drugs for treating *Schistosoma mansoni* infection. *Cochrane Database Syst Rev*. 2013;2:CD000528.

A2. Kramer CV, Zhang F, Sinclair D, et al. Drugs for treating urinary schistosomiasis. *Cochrane Database Syst Rev*. 2014;8:CD000053.

A3. Pérez del Villar L, Burguillo FJ, López-Abán J, et al. Systematic review and meta-analysis of artemisinin based therapies for the treatment and prevention of schistosomiasis. *PLoS ONE*. 2012;7:e45867.

A4. Sacko M, Magnussen P, Traore M, et al. The effect of single dose versus two doses of praziquantel on Schistosoma haematobium infection and pathology among school-aged children in Mali. *Parasitology*. 2009;136:1851-1857.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

TREATMENT

Rx

The effects of praziquantel on schistosomes can be summarized into three mechanisms: muscle contraction, tegumental damage (i.e., vacuolization and blebbing), and metabolic alterations (i.e., decreased glucose uptake, lactate excretion, and glycogen content). Schistosome calcium ion channels are so far the only molecular targets of praziquantel action identified. The drug may also interfere with adenosine uptake, which is relevant because schistosomes are unable to synthesize purines de novo. Several drugs (e.g., phenytoin, rifampin, and azole antifungals) decrease plasma concentrations of praziquantel because they inhibit the cytochrome P450 pathway 3A4. Chloroquine and cimetidine also decrease praziquantel bioavailability. Praziquantel's activity also depends on the patient's immune system. Several studies have reported that praziquantel induces the exposure of worm surface antigens that may function as targets for an immune response.

GENERAL REFERENCES

1. Colley DG, Bustinduy AL, Secor WE, et al. Human schistosomiasis. *Lancet*. 2014;383:2253-2264.
2. Grimes JE, Croll D, Harrison WE, et al. The relationship between water, sanitation and schistosomiasis: a systematic review and meta-analysis. *PLoS Negl Trop Dis*. 2014;8:e3296.
3. World Health Organization. Schistosomiasis: population requiring preventive chemotherapy and number of people treated in 2010. *Wkly Epidemiol Rec*. 2012;87:37-44.
4. Cavalcanti MG, Silva LF, Peralta RH, et al. Schistosomiasis in areas of low endemicity: a new era in diagnosis. *Trends Parasitol*. 2013;29:75-82.
5. De Souza RP, Cardoso LS, Lopes GTV, et al. Cytokine and chemokine profile in individuals with different degrees of periportal fibrosis due to *Schistosoma mansoni* infection. *J Parasitol Res*. 2012;2012:1-10.
6. Oliveira RR, Figueiredo JP, Cardoso LS, et al. Factors associated with resistance to *Schistosoma mansoni* infection in an endemic area of Bahia, Brazil. *Am J Trop Med Hyg*. 2012;86:296-305.
7. Hoette HS, Fernandes CJC, Jardim C, et al. Review article: schistosomiasis associated pulmonary hypertension. *Int J Clin Pract*. 2010;64:25-28.
8. Coyle CM. Schistosomiasis of the nervous system. *Handb Clin Neurol*. 2013;114:271-281.
9. Ferrari TC, Moreira PR. Neuroschistosomiasis: clinical symptoms and pathogenesis. *Lancet Neurol*. 2011;10:853-864.
10. Honeycutt J, Hammam O, Fu CL, et al. Controversies and challenges in research on urogenital schistosomiasis-associated bladder cancer. *Trends Parasitol*. 2014;30:324-332.
11. Kjetland EF, Leutscher PDC, Ndhlovu PD. A review of female genital schistosomiasis. *Trends Parasitol*. 2012;28:58-65.
12. Hegertun IEA, Gundersen KMS, Kleppa E, et al. *S. haematobium* as a common cause of genital morbidity in girls: a cross-sectional study of children in South Africa. *PloS Negl Trop Dis*. 2013;7:2104-2111.
13. Cnops L, Tannich E, Polman K, et al. *Schistosoma* real-time PCR as diagnostic tool for international travelers and migrants. *Trop Med Int Health*. 2012;17:208-216.
14. Wikman-Jorgensen PE, Henríquez-Camacho CA, Serrano-Villar S, et al. The role of artesunate for the treatment of urinary schistosomiasis in schoolchildren: a systematic review and meta-analysis. *Pathog Glob Health*. 2012;106:397-404.
15. Erfe JM, Belizario VY, Chua PL, et al. Validating the WHO dose pole in the Philippines for school-based mass drug administration of praziquantel for morbidity control of schistosomiasis. *Trans R Soc Trop Med Hyg*. 2013;107:620-626.
16. Wang W, Wang L, Liang YS. Susceptibility or resistance of praziquantel in human schistosomiasis: a review. *Parasitol Res*. 2012;111:1871-1877.

REVIEW QUESTIONS

1. A 20-year-old man was admitted with a fever (temperature of 39° C), abdominal pain, dyspnea, malaise, and weight loss of 30 pounds in the last 10 days. He has lived in Italy but traveled to areas endemic for *S. mansoni*, where he swam in a lagoon 4 weeks before the start of symptoms. A mild enlargement of the liver and spleen was noted on examination, and chest radiographs showed interstitial infiltrates. Schistosomiasis is suspected. Which of the following is the most appropriate test for diagnosis?

A. Stool examination
B. Urine examination
C. Real-time polymerase chain reaction in stool
D. Serologic test
E. Rectal biopsy

Answer: D Egg deposition can be documented only up to 4 weeks after exposure. Although serologic tests do not rule out previous *S. mansoni* infection, the patient did not have history of previous exposure to *S. mansoni*, and therefore serology is the best test for the diagnosis of acute disease.

2. Which of the following statements is correct about human schistosomiasis?

A. It may be prophylactically treated with praziquantel.
B. Praziquantel is currently the treatment of choice.
C. The combination of praziquantel and oxamniquine is recommended.
D. There is no combination treatment available.
E. Susceptibility or resistance to praziquantel has not been reported.

Answer: B Praziquantel has a poor prophylactic effect, which reduces its efficacy in areas of high transmission of *Schistosoma*. However, praziquantel is considered the drug of choice to treat schistosomiasis for its efficacy, low adverse events including severe adverse events, single oral dose, and competitive cost. There is a combination of praziquantel and artemisin derivatives that shows increased cure rates in schistosomiasis treatment.

356

LIVER, INTESTINAL, AND LUNG FLUKE INFECTIONS

EDUARDO GOTUZZO

Flukes belong to the phylum Platyhelminthes and class Trematoda. The major flukes of medical importance, other than *Schistosoma,* are discussed in this chapter; they belong to the families Fasciolidae, Opisthorchiidae,

TABLE 356-1 CLINICAL MANIFESTATIONS, DIAGNOSIS, AND TREATMENT OF LIVER, INTESTINAL, AND LUNG FLUKE INFECTIONS IN HUMANS

FLUKE, NUMBER OF PEOPLE INFECTED, AND GEOGRAPHIC DISTRIBUTION	CLINICAL MANIFESTATIONS ACUTE AND CHRONIC PHASES	DIAGNOSIS*	TREATMENT†
LIVER FLUKES			
Fasciola spp 17 million Cosmopolitan	Acute ‡Hepatomegaly, eosinophilia, fever, abdominal pain, metastatic-like lesions on liver CT Chronic §Biliary obstruction, gallstones, fibrosis, cholangitis	Stools are negative Serology (Fas2 ELISA) Eggs in stool Fas2 ELISA	TCZ
Opisthorchis spp 11.2 million Asia	‡Fever, malaise, arthralgia, lymphadenopathy, and rash §Jaundice, cholangitis, cholangiocarcinoma	Serology Eggs in stool	PZQ
Clonorchis spp 35 million Asia Eastern Europe	‡Fever, rash, malaise, and RUQ discomfort §Choledocholithiasis, cholangitis, cholecystitis, liver abscess, and possible cholangiocarcinoma	Serology Eggs in stool	PZQ
INTESTINAL FLUKES			
Fasciolopsis buski and others¶ 50 million Asia and North Africa	Small bowel inflammation, ulceration, mucus secretion, protein-losing enteropathy, malabsorption	Eggs in stool	PZQ
LUNG FLUKE			
Paragonimus spp 22 million Asia, Americas, Africa	‡Abdominal pain, pleuritic pain, cough, eosinophilia §Hemoptysis, cough, chronic pleural effusions, pulmonary cysts, abscess	Serology Eggs in sputum or stool	PZQ or TCZ

*For egg morphology and size, refer to text in the Diagnosis section.
†For drug regimens, refer to text in the Treatment section.
‡Acute.
§Chronic. For chronic infection, a sedimentation technique is preferred (suggested technique: rapid sedimentation technique or Kato-Katz technique).
¶Diagnosis and treatment also apply for other intestinal flukes.
CT = computed tomography; ELISA = enzyme-linked immunosorbent assay; PZQ = praziquantel; RUQ = right upper quadrant; TCZ = triclabendazole.

Heterophyidae, Echinostomatidae, and Troglotrematidae. With the exception of schistosomes, all flat worms of clinical significance are hermaphroditic. Blood flukes, or schistosomes, are discussed in Chapter 355. The epidemiology, clinical manifestations, diagnosis, and treatment are summarized in Table 356-1.

LIVER FLUKES

Fascioliasis

Fascioliasis is a zoonosis caused by *Fasciola hepatica* (adult: 30 by 13 mm) or *Fasciola gigantica* (adult: 75 by 20 mm). The most common natural hosts are sheep, cattle, and goats.

EPIDEMIOLOGY

The infection is distributed globally. The highlands of Peru, Bolivia, Ecuador, Egypt, and Vietnam are the most affected regions in the world. Prevalence rates higher than 60% have been reported in Peru and Bolivia. Estimates of the number of people infected in some countries are 830,000 in Egypt, 742,000 in Peru, 360,000 in Bolivia, 37,000 in Yemen, 20,000 in Ecuador, and 10,000 in Iran. It has been estimated that between 2.4 and 17 million people are infected worldwide.[1]

PATHOBIOLOGY

The life cycle begins when the parasite's eggs in stool are deposited in water; miracidia appear, develop, and hatch in 9 to 14 days and invade many species of freshwater snails (*Lymnaea* spp), in which they multiply as sporozoites, rediae, and cercariae during a period of 4 to 7 weeks. They then leave as free-swimming cercaria that subsequently attach to watercress, water lettuce, alfalfa, mint, parsley, or khat. The main source of infection is consumption of raw vegetables or water contaminated with metacercariae. Women have a higher incidence of the disease, with more severe infections and complications than seen in men.

CLINICAL MANIFESTATIONS

After consumption of contaminated vegetables, the larvae excyst in the duodenum and then migrate through the bowel wall to the liver through the peritoneal cavity. In 4 weeks, they reach the liver, penetrate Glisson's capsule, and cause inflammation and pain. An acute diarrhea of 2 to 5 days' duration may occur before liver invasion. During their migration through the liver, the ongoing inflammatory process is accompanied by fever, pain, and hypereosinophilia. In a few cases, intense hemorrhage manifested as subcapsular liver hematoma may develop. Computed tomography (CT), magnetic resonance imaging, or ultrasonography can detect these initial lesions. Furthermore, migration of the parasite leaves a trail or track that can be observed in histologic sections or by imaging (CT). The flukes sometimes die and leave cavities filled with necrotic debris that are eventually replaced by scar tissue and then become calcified. After 3 to 5 months of migration in the liver, the juvenile larvae finally reach the bile ducts. During this invasive, migratory, or acute phase the clinical manifestations are prolonged fever, hepatomegaly, abdominal pain, and eosinophilia. Multiple hypodense lesions are seen on CT, similar to metastases, but they change in position, attenuation, and shape in time because the parasites are still migrating. Acute fascioliasis is clinically similar to acute cholecystitis but with the addition of significant eosinophilia. It can occur in travelers with acute subcapsular hematoma or "metastatic-like lesions" seen on CT of the liver. Hyperbilirubinemia is notably absent in this phase. Other manifestations are anorexia, weight loss, nausea, vomiting, cough, diarrhea, urticaria, lymphadenopathy, and arthralgias. Occasionally, the juvenile larvae reach other ectopic or extrahepatic locations, such as subcutaneous tissue, the pancreas, the eye, the brain,[2] and the stomach wall, among others. In endemic areas, the acute phase manifestations can be superimposed on chronic infection.

Arrival of the parasite in the bile ducts marks the beginning of the chronic phase. Mature flukes consume hepatocytes and duct epithelium and reside for years in the hepatic and common bile ducts and sometimes in the gallbladder. In this chronic phase, the liver contains large dilated, thick-walled, and calcareous bile ducts with yellowish-brown bile. The bile ducts have a thickened hyperplastic wall with marked fibrosis. Symptoms usually reflect biliary obstruction with colicky pain in the right upper quadrant and epigastric area. Eosinophilia is absent in half of the chronic cases. Bacterial superinfection of these cysts and consequent cholangitis can develop. Other complications are hemobilia and liver fibrosis. Alkaline phosphatase is commonly elevated because of biliary obstruction, which sometimes requires surgical intervention. On imaging, the initial lesions may be often confused with hepatic metastases. Other findings on CT are hepatomegaly, tracklike

hypodense lesions in subcapsular locations, multiple hypodense nodular areas (abscess-like lesions), or low-density, serpiginous, tortuous, tunnel-like branching lesions ranging from 2 to 10 mm. CT also can show subcapsular hematoma, enhancement of Glisson's capsule, necrotic granuloma, and cystic calcifications. After maturation, the adult flukes start laying eggs, which are passed from the sphincter of Oddi to the intestines and evacuated to the environment along with stool. Adult parasites can live in the bile ducts for up to 13 years. In endemic populations, chronic infection has been reported as mildly symptomatic, whereas in travelers or temporary residents, it has been reported to cause biliary obstruction, with adult parasites being seen on endoscopic retrograde cholangiopancreatography (ERCP).

In summary, the typical clinical presentation of acute fascioliasis must be differentiated from cholecystitis; "liver metastasis" with fever and hypereosinophilia should raise the possibility of this infection; and in children and adolescents, systemic toxocariasis will be the differential diagnosis

Clonorchiasis and Opisthorchiasis

Clonorchiasis is the disease caused by *Clonorchis sinensis,* also called the Chinese or oriental liver fluke (adult: 10 to 25 mm by 3 to 5 mm). Opisthorchiasis is caused by *Opisthorchis viverrini* (adult: 5 mm to 10 mm by 1 mm to 2 mm) and *Opisthorchis felineus* (adult: 7 to 12 mm by 2 to 3 mm). The most common natural hosts are dogs, cats, pigs, and some small wild mammals.

EPIDEMIOLOGY

The global estimate of the number of people infected is 46.2 million: 35 million with *C. sinensis* (15 million in China), 10 million with *O. viverrini* (8 million in Thailand), and 1.2 million with *O. felineus.* There are 601 million and 79.8 million people at risk for infection with *Clonorchis* and *Opisthorchis*, respectively. Both infections are endemic in the Far East, Southeast Asia, and Eastern Europe. *C. sinensis* is endemic in northeast China, southern Korea, Japan, Taiwan, northern Vietnam, and the far eastern part of Russia, whereas *O. viverrini* is endemic in Laos, Thailand, Vietnam, and Cambodia. *O. felineus* infection is prevalent in Russia, Ukraine, and Kazakhstan.

PATHOBIOLOGY

The life cycle starts when the adult worm deposits fully developed eggs, which are then passed to the environment through feces. They hatch in water and the miracidia infect their first intermediate host, a freshwater snail (*Bithynia* spp or *Parafossarulus* spp), where they transform into sporocysts, rediae, and cercariae. Cercariae are released from the snail and then penetrate freshwater fish, which are the second intermediate host (*Cyclocheilichthys* spp, *Puntius* spp, *Hampala dispar*); the cercariae encyst as metacercariae in the muscles or under the scales. In general, the infection is acquired by eating raw or uncooked cyprinoid fish products in rural areas or dishes such as koi-pla, a salad made with raw fish. The metacercariae pass through the stomach and reach the small intestine unharmed. Then, through the ampulla of Vater, they reach the bile ducts, where they mature into adult worms within 4 weeks and deposit yellow, operculated eggs. The parasites may live for up to 45 years in a human host.

CLINICAL MANIFESTATIONS

Clonorchiasis as an acute infection caused by *C. sinensis* is usually asymptomatic, but some patients may have fever, rash, malaise, and right upper quadrant abdominal discomfort.[3] Chronic infections may be manifested as recurrent pyogenic cholangitis, cholecystitis, obstructive jaundice, hepatomegaly, cholecystitis, multiple hepatic tumors, cholelithiasis, or pancreatitis. In chronic carriers with a high load of parasites, cholangiocarcinoma may develop, especially in Thailand.

Opisthorchiasis as an acute infection caused by *O. viverrini* can cause right upper quadrant abdominal pain, flatulence, fatigue, and a hot sensation over the abdomen. In the chronic phase, mild hepatomegaly occurs, mainly in more heavily infected patients (egg counts >10,000/g). Jaundice and splenomegaly are not observed. Intrahepatic duct stones and recurrent suppurative cholangitis are common manifestations of opisthorchiasis. Whenever jaundice and ascending cholangitis are detected, fluke-related cholangiocarcinoma should be suspected.[4]

In opisthorchiasis caused by *O. felineus,* infestation usually follows the consumption of raw, slightly salted, and frozen fish ("stroganina"), and acute symptoms occur 2 to 4 weeks later, including high-grade fever, nausea, vomiting, abdominal pain, malaise, arthralgias, lymphadenopathy, and rash. Peripheral eosinophilia is a common finding, especially during the initial 2 to 6 weeks of the infection, together with raised liver enzyme levels. In chronic infection, the eosinophilia is usually milder. Patients may have suppurative cholangitis and liver abscesses because of biliary obstruction. Ultrasonography or CT demonstrates the pathologic changes in the liver, including intrahepatic duct dilation and periductal changes.

The pathologic and clinical consequences of these flukes are related to the intensity and duration of cumulative infestations. In general, they cause inflammation around the biliary tree, severe hyperplasia of epithelial cells, metaplasia of mucin-producing cells in the mucosa, and progressive periductal fibrosis. There are clear associations between *O. viverrini* infection and cholangiocarcinoma.[5] Several *N*-nitroso compounds and their precursors occur at low levels in fermented food, such as preserved mud fish paste (*pla ra*), a condiment that is a ubiquitous component of the cuisine of northeastern Thailand and Laos.

INTESTINAL FLUKES

The most common human intestinal trematode is *Fasciolopsis buski* (adult: 20 to 75 mm by 8 to 20 mm). It is found mainly in the central and southeast parts of Asia. *F. buski* is a common parasite in pigs. Others are *Heterophyes* (adult: 1 to 2 mm in length), *Metagonimus yokogawai* (adult: 1 to 2.5 mm by 0.4 to 0.75 mm), and *Echinostoma* spp (adult: 6.5 by 1 to 2 mm).

EPIDEMIOLOGY

More than 50 species of intestinal trematodes from the Far East, Middle East, and North Africa have been reported to cause human infection. *H. heterophyes* also can be found in the Nile delta region of Egypt. An estimated 40 to 50 million people are infected with one or several species of intestinal flukes. Their life cycles are similar. The adult worm, attached to the intestinal wall of humans, produces eggs that are passed in feces. The eggs reach water, and miracidia develop and penetrate the first intermediate host—snails. During the course of 6 to 7 weeks inside the host snails, they develop into sporocysts, rediae, and cercariae. The cercariae leave the snails to encyst in the second intermediate host, which can be freshwater snails, fish, tadpoles, or vegetables. Humans are infected by the ingestion of raw stems, leaves (especially bamboo shoots), watercress, or water chestnuts with encysted metacercariae. In the human duodenum, the metacercariae attach to the walls and become adult worms in approximately 3 months.

PATHOBIOLOGY AND CLINICAL MANIFESTATIONS

Despite the fact that the majority of intestinal fluke infections are asymptomatic, they may cause inflammation, ulceration, and mucus secretion at the site of attachment, particularly in the duodenum and jejunum. In fact, gastrointestinal hemorrhage, perforation, and abscesses have been observed. The differential diagnoses include typhoid fever, intestinal tuberculosis, and amebiasis. In these endemic cases, ulcerative colitis and other inflammatory diseases of the bowel are uncommon. In heavy infection, intestinal obstruction, protein-losing enteropathy, malabsorption, impaired vitamin B_{12} absorption, hypoalbuminemia, and anasarca have been reported. The adult worm causes traumatic, toxic, and obstructive damage to the intestinal mucosa. Some cases have been diagnosed by direct visualization of the adult parasite via esophagogastroduodenoscopy.

PULMONARY FLUKES

Paragonimiasis is a zoonosis caused by *Paragonimus* spp (adult size: 10 by 5 mm).[6] Reservoir hosts include felids, canids, viverrids, mustelids, some rodents, and pigs. At least 10 species of *Paragonimus* are known to cause human disease; of these, *Paragonimus westermani* is the most common.

EPIDEMIOLOGY

An estimated 22 million people are infected worldwide with *Paragonimus* spp, and 293 million are at risk. Human paragonimiasis is distributed mainly in Southeast Asia, Japan, Korea, China, and the Philippines, where *P. westermani* is the main species. In other areas of low endemicity, other species have been reported, such as *Paragonimus mexicanus* in Latin America, *Paragonimus kellicotti* in North America, *Paragonimus heterotremus* in India, and *Paragonimus africanus,* and *Paragonimus uterobilateralis* in West Africa.

PATHOBIOLOGY

The life cycle starts when the eggs are excreted unembryonated in sputum, or alternatively they can be swallowed and passed in stool. In the external environment the eggs become embryonated, and miracidia hatch, seek the first intermediate host, a snail (families Pleuroceridae and Thiaridae), and penetrate its soft tissues. Within the snail, asexual reproduction occurs for

several weeks, with transformation into sporocysts, rediae, and cercariae; the last emerge from the snail and invade the second intermediate host, a crustacean such as a crab or crayfish, where they encyst and become metacercariae. This is the infective stage for the mammalian host. Human infection occurs by eating inadequately cooked or pickled crab or crayfish that harbor metacercariae of the parasite. The metacercariae excyst in the duodenum, penetrate the intestinal wall, and migrate through the peritoneal cavity toward the lungs. During migration through the peritoneum and diaphragm, the inflammatory process causes abdominal pain and dry cough. When invading the lungs, they become encapsulated and develop into adults. Infections may persist for 20 years in humans.

CLINICAL MANIFESTATIONS

Paragonimiasis typically results from the consumption of raw or improperly cooked crustaceans, especially crabs. Recently in the United States, some autochthonous cases have been reported in the Midwest in people who consumed raw crayfish while camping during the summer.[7] Most of the infected people are asymptomatic and have subclinical disease. During the first month of infection, abdominal pain may represent the juvenile larvae migrating through the abdominal cavity before reaching the lungs. Irritation of the diaphragm or pleura may cause dry cough. Fever, chest pain, fatigue, and urticaria may follow, as well as eosinophilia. Pleural effusions may be seen at this stage, with significant eosinophilia noted on analysis of pleural fluid, which can be the first clue to the diagnosis. In fact, pleural manifestations predominate early in the disease process, whereas lesions of the pulmonary parenchyma predominate later in the course of disease. Moreover, pneumothorax and mild eosinophilia may occur only 1 month after the initiation of infection. The migrating worms may cause bronchiectasis, interstitial pneumonitis, transient hemorrhage, or bronchopneumonia. Again, pulmonary lesions and eosinophilia raise the possibility of paragonimiasis. Cough and recurrent hemoptysis are common clinical findings in this phase. The chronic stage occurs when the worms are paired in a cyst in the pulmonary parenchyma. Eggs are produced 6 weeks after infection, and if there is communication with the bronchial tree, eggs may be seen in a sputum sample under microscopy, or they can be swallowed and passed with stool. The rusty discoloration of sputum is caused by the presence of the tan- to brown-pigmented *Paragonimus* eggs; the sputa of these patients have been classically described as resembling "iron filings." Charcot-Leyden crystals can be seen.

Peripheral blood eosinophilia and elevated total serum IgE levels are observed in approximately 80% of patients. Common findings on CT are pleural effusion, hydropneumothorax, pulmonary nodules or consolidation of air spaces, and cysts. The most common ectopic form is cerebral paragonimiasis, which is manifested as eosinophilic meningitis or meningoencephalitis, brain tumor, or just residual calcifications from a past infection.

DIAGNOSIS AND TREATMENT OF LIVER, INTESTINAL, AND LUNG FLUKE INFECTIONS

DIAGNOSIS

In general, transmission of food-borne trematodiases is restricted geographically and distribution is one of the most important factors in suspecting this diagnosis. In the appropriate clinical setting, laboratory and imaging data can add information to narrow the differential diagnosis. Acute fluke infections require a high level of suspicion. Serologic tests, direct visualization of the migrating larvae, or empirical therapy with a significant clinical response (including reduction of eosinophilia) are major criteria to confirm the diagnosis. In the chronic phase, the diagnosis is usually made by visualization of the eggs in stool or, in the case of *Paragonimus,* in sputum. A sedimentation technique must be performed on a series of at least three stool specimens from alternate days or even weeks. Some of the sedimentation or concentration parasitologic techniques used for these infections include the Lumbreras rapid sedimentation technique, the Kato-Katz technique, and the ether-formalin concentration technique. Immunodiagnostics is an excellent tool, particularly for patients who do not have demonstrable eggs in clinical specimens.

If acute fascioliasis is strongly suspected, serology would be the next step. Cathepsin L1–based antibody enzyme-linked immunosorbent assay (ELISA) has a sensitivity of 92% and a specificity of 84%. If negative, the diagnosis is unlikely. If serology is not available, CT of the liver can visualize the characteristic tracklike lesions. However, because the parasitic lesions are very similar to metastases, liver biopsy may be necessary. If serology or CT is not available, a trial of triclabendazole with clinical (including eosinophilia) resolution is the major criterion for diagnosis. In chronic fascioliasis, the Lumbreras rapid sedimentation technique is the method of choice to detect the eggs in stool. At least three stool examinations are preferred. If negative, serology can be helpful. Ultrasonography and CT have low sensitivity in this phase. ERCP usually finds the adult parasites in the bile duct when performed for other reasons. Nonetheless, ERCP can be useful to eliminate the adult parasites causing biliary obstruction.

In opisthorchiasis, serology or stool examinations can be performed to approach the diagnosis. The Ov-CP-1–based enzyme-linked immunosorbent assay (ELISA) has a sensitivity of 95% and a specificity of 96%. For *C. sinensis,* ELISA has a sensitivity between 81.3 and 96% and a specificity between 92.6 and 96.2%. For detection of eggs and to measure the intensity of infection, the Kato-Katz technique is preferred. Intestinal fluke eggs can be detected by performing a sedimentation stool technique, preferably in consecutive stool samples.

For paragonimiasis, an immunoblot assay performed with a crude antigen extract of *P. westermani* has been in use at the Centers for Disease Control and Prevention; the sensitivity of the test is 96%, and its specificity is 99%. This would be the ideal first step to confirm the diagnosis. In the acute phase, the precise location of the migrating larva is unknown and a biopsy may not necessarily target the parasite. When serologic findings are negative (or not available), a trial of praziquantel or triclabendazole with a positive clinical response in 48 to 72 hours is a major criterion for diagnosis. In the chronic phase, several sputum samples have to be examined by a sedimentation technique to increase sensitivity. Stool examinations are complementary because the eggs can be swallowed by the host and then passed through stool. If a pulmonary cyst contains adult parasites with eggs but they do not have communication to the main bronchi, serologic examination is indicated to confirm the diagnosis. If not available, biopsy is warranted.

Under the light microscope, the morphologic characteristics and size of the eggs may be sufficient to identify the specific fluke. For *Fasciola,* the egg is large, ellipsoid, and oval, with an indistinct operculum and thin shell. The *F. hepatica* egg measures 120 to 150 µm by 63 to 90 µm, the *F. gigantica* egg measures 160 to 190 µm by 70 to 90 µm, and the *F. buski* egg measures 130 to 159 µm by 78 to 98 µm. *Opisthorchis* eggs are elongated with an operculum on the anterior end and a pointed terminal "knob" on the posterior end (26 to 30 µm by 11 to 15 µm). *C. sinensis* eggs (27 to 35 µm by 11 to 20 µm) are small, ovoid, or elongated, with broad rounded posterior ends, a convex operculum resting on "shoulders," and a small "knob" on the posterior ends. *H. heterophyes* and *M. yokogawai* eggs are similar in size, 26 to 30 µm by 15 to 17 µm. *Echinostoma* eggs are brownish, operculated, and measure 83 to 116 µm by 58 to 69 µm. *Paragonimus* eggs measure 68 to 118 µm by 39 to 67 µm, are reddish-brown and ovoid or elongated, and have a thick shell; the operculum is slightly flattened and fits into the shoulder area of the shell, and the posterior end is thickened. The egg is often asymmetrical, with one side slightly flattened.

TREATMENT

Rx

The general principles and details of antiparasitic therapy are provided in Chapter 344. For fascioliasis, 10 mg/kg of triclabendazole once or twice has a cure rate higher than 90%, and it is the treatment of choice. Cure is achieved if stool examinations remain negative for at least 3 months. Serology usually can take more than a year to resolve. In treatment of the chronic phase, the dead parasites can occasionally cause biliary obstruction, which may need surgical consultation. In case of failure, some experts recommend double doses of triclabendazole for 2 days. Recently, in Mexico, nitazoxanide has appeared promising, but more experience is needed. For *O. viverrini,* a single dose of praziquantel (40 to 50 mg/kg) has a cure rate of 91 to 97%. For clonorchiasis, the recommended dose of praziquantel is 25 mg/kg three times for 1 day (total dose of 75 mg/kg), which has a cure rate of 83 to 85%. Tribendimidine has been found to be comparably effective as praziquantel in the treatment of *C. sinesis* infection with fewer adverse events in an open-label trial in 74 patients. For intestinal flukes, praziquantel, 25 mg/kg by mouth three times daily for 1 day, is recommended. For paragonimiasis, praziquantel, 25 mg/kg by mouth three times daily for 3 days, or triclabendazole, 10 mg/kg by mouth once or twice, is highly effective. For ectopic cases, surgery may be necessary. In follow-up, negative stool examinations in the ensuing weeks can confirm cure. However, because the rate of reinfection is high in individuals from endemic areas, a suddenly positive stool examination is highly suggestive of a new infection rather than failure to respond to treatment.

PREVENTION

Prevention of infection with these flukes depends on several factors, including recognizing their geographic distribution and avoiding the consumption of raw vegetables, fish, crayfish, or contaminated water in endemic areas. Proper medical advice must be given to individuals traveling or planning to reside in endemic areas, not only to prevent fluke infection but also to inform them about the risk for coinfection with other parasites. Control of these flukes in animals is impractical because of wild animal reservoirs. However, control of human infections is challenging because it involves changing long-established cultural, dietary, and sanitary habits. Massive chemotherapy in highly endemic populations may reduce the infection in humans and in selected animals.

Grade A References

A1. Hien TT, Truong NT, Minh NH, et al. A randomized controlled pilot study of artesunate versus triclabendazole for human fascioliasis in central Vietnam. *Am J Trop Med Hyg*. 2008;78:388-392.

A2. Qian MB, Yap B, Yang YC, et al. Efficacy and safety of tribendimidine against *Clonorchis sinesis*. *Clin Infect Dis*. 2013;56:e76-e82.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Furst T, Keiser J, Utzinger J. Global burden of human food-borne trematodiasis: a systematic review and meta-analysis. *Lancet Infect Dis*. 2012;12:210-221.
2. Mas-Coma S, Agramunt VH, Valero MA. Neurological and ocular fascioliasis in humans. *Adv Parasitol*. 2014;84:27-149.
3. Saijuntha W, Sithithaworn P, Kaitsopit N, et al. Liver flukes: Clonorchis and Opisthorchis. *Adv Exp Med Biol*. 2014;766:153-199.
4. Sripa B, Brindley PJ, Mulvenna J, et al. The tumorigenic liver fluke *Opisthorchis viverrini:* multiple pathways to cancer. *Trends Parasitol*. 2012;28:395-407.
5. Tyson GL, El-Serag HB. Risk factors for cholangiocarcinoma. *Hepatology*. 2011;54:173-184.
6. Blair D. Paragonimiasis. *Adv Exp Med Biol*. 2014;766:115-152.
7. Diaz JH. Paragonimiasis acquired in the United States: native and nonnative species. *Clin Microbiol Rev*. 2013;26:493-504.

REVIEW QUESTIONS

1. Hepatic *Fasciola* is acquired by ingestion of:

A. Uncooked shellfish
B. Watercress
C. Fresh fish meat products
D. Fresh fish
E. All the above

Answer: B Hepatic *Fasciola* do not invade fish or meat. After invading freshwater snails, they leave as free-swimming cercaria that subsequently attach to watercress (as well as water lettuce, alfalfa, mint, parsley, or khat).

2. *Paragonimus* is acquired by ingestion of:

A. Uncooked crab or crayfish
B. Watercress and fresh fish meat products
C. All the above

Answer: A *Paragonimus* emerges from the snail and invades the second intermediate host, a crustacean such as a crab or crayfish, where they encyst and become the infective stage for a mammalian host.

3. The treatment of choice for *O. viverrini* or clonorchiasis is:

A. Albendazole
B. Praziquantel
C. Ivermectin
D. Triclabendazole
E. Emetina

Answer: B For *O. viverrini* or clonorchiasis, praziquantel regimens have been shown in several studies to have cure rates of 91 to 97% and 83 to 85%, respectively. Albendazole has poor activity. Ivermectin is effective for onchocerciasis and strongyloides. Triclabendazole is the drug for *Fasciola,* but it is second-line option here. Emetina is a parenteral drug that is useful only for *Fasciola* with high toxicity.

4. Cholangiocarcinoma is associated in Southeast Asia with chronic carriage of:

A. *Fasciola hepatica*
B. *Strongyloides*
C. *Toxoplasma*
D. *O. viverrini*
E. *Taenia*

Answer: D Several studies in Thailand have demonstrated a strong association of *O. viverrini* with cholangiocarcinoma. *Fasciola hepatica* is associated with fibrosis but not cancer. *Strongyloides* produces hyperinfection but not cancer. Toxoplasma does not invade the biliary tract. Taenia is not associated with cancer.

5. A 33-year-old man presents with pleural effusion and eosinophilia. In the following weeks he developed a recurrent hemoptysis. The most likely diagnosis is:

A. Tuberculosis
B. Strongyloidiasis
C. Toxocara
D. *Fasciola hepatica*
E. *Paragonimus*

Answer: E The typical clinical pattern for *Paragonimus* is the damaged lung, where hemoptysis and eosinophilia are the more common presentation. Evaluation of the sputum will be the key diagnostic test in this case. TB is the main differential diagnosis, but eosinophilia is not common in TB. Strongyloidiasis produces lung disease (migratory eosinophilic pneumonia) but is rarely associated with pleural effusion. *Toxocara* usually affects children and produces systemic disease with liver involvement; lung involvement is rare and hemoptysis even more so. *Fasciola hepatica* is associated with eosinophilia, but its typical presentation is a liver "abscess"–like or metastatic-like syndrome, and lung involvement with hemoptysis is rare.

357 INTESTINAL NEMATODE INFECTIONS

DAVID J. DIEMERT

DEFINITION

Nematodes are nonsegmented roundworms belonging to the phylum Nematoda. Most species are free-living in soil or water, but a few parasitize humans. Nematode infections are highly prevalent and affect millions worldwide. They are complex multicellular organisms with specialized organs that include a protective outer coating or cuticle, a complete and functional gastrointestinal tract, and muscular, nervous, and reproductive systems.

Nematodes of medical importance can be divided into those that primarily affect the gastrointestinal tract, where adult worms become established and cause disease, and those that affect other tissues and organ systems. The former group is covered in this chapter and includes the roundworm *Ascaris lumbricoides,* the hookworms *Ancylostoma duodenale* and *Necator americanus,* the pinworm *Enterobius vermicularis,* the whipworm *Trichuris trichiura,* and the threadworm *Strongyloides stercoralis*. Zoonotic intestinal nematodes such as *Trichostrongylus* and *Anisakis* also occasionally infect and cause disease in the gastrointestinal tract of humans. Nematodes that invade and cause disease primarily in tissues outside the gastrointestinal tract include those that cause lymphatic filariasis (*Wuchereria bancrofti, Brugia malayi,* and *Brugia timori*), *Onchocerca volvulus, Loa,* the guinea worm *Dracunculus medinensis,* as well as *Trichinella* and *Angiostrongylus* spp (Chapter 358).

Nematodes that infect humans measure from several millimeters to more than a meter in length and often survive for months to years within their host. With the exception of *S. stercoralis* and *Capillaria philippinensis,* adult worms cannot complete their life cycle within a human host. Instead, sexually mature adult worms mate and produce eggs or larvae that must have at least one stage of development outside the host, either in the environment or in an intermediate host.

Nematode infections are rarely fatal; they more commonly result in chronic morbidity such as iron deficiency anemia caused by hookworm or blindness due to onchocerciasis. For most nematodes, the severity of the clinical manifestations of infection is proportional to the number of worms harbored by a given host; although light infections with only a few worms are usually asymptomatic, pathologic features appear with heavier worm burdens.

Nematode infections are prevalent in the temperate and tropical regions of Africa, Asia, and Latin America. They are transmitted by the oral ingestion of embryonated eggs or by penetration of infective larvae through the skin, either by direct contact with contaminated soil or by arthropod vectors. Nematode infections are most common in areas with poor sanitation, where the environment is contaminated by human waste, and in climates that support survival of the insect vector, if one is involved in the life cycle.

In endemic areas most individuals harbor low numbers of worms and a minority have relatively high worm burdens and contribute disproportionally to both transmission and morbidity.

ASCARIASIS

The Pathogen

A. lumbricoides, colloquially known as *roundworm,* is acquired by oral ingestion of embryonated eggs. In the stomach, the egg's protective outer shell is dissolved by gastric acid, releasing larvae into the small intestine, where they penetrate the intestinal wall and enter the portal circulation. The larvae migrate to the liver and then the pulmonary vasculature, where they break into the alveolar spaces, ascend the bronchial tree, and are swallowed to re-enter the small intestine, developing there into adult worms approximately 9 to 11 weeks after egg ingestion. Adult worms (Fig. 357-1) range in length between 15 and 50 cm and survive in the host for approximately 18 months.[1] Female adult *Ascaris* worms produce approximately 240,000 eggs per day, which are expelled in feces. Fertilized eggs embryonate in warm, moist, shady soil, after which they are infectious (Fig. 357-2). Eggs can survive up to 15 years in the environment, being extremely hardy and resistant to extreme temperatures and desiccation.

EPIDEMIOLOGY

A. lumbricoides is the most prevalent nematode infection worldwide, with approximately a billion people chronically infected.[2] Infection is common in sub-Saharan Africa, south and Southeast Asia, and Latin America, especially in rural areas of high population density with inadequate sanitation or treatment of sewage and where untreated human feces are used as fertilizer.

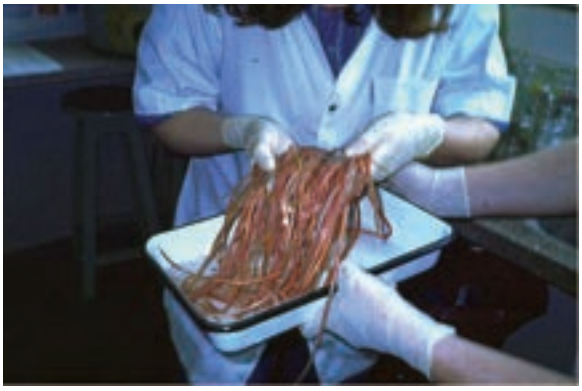

FIGURE 357-1. Mass of adult *Ascaris lumbricoides* worms recovered from a child after the administration of mebendazole. (Reproduced with permission from Dickson Despommier.)

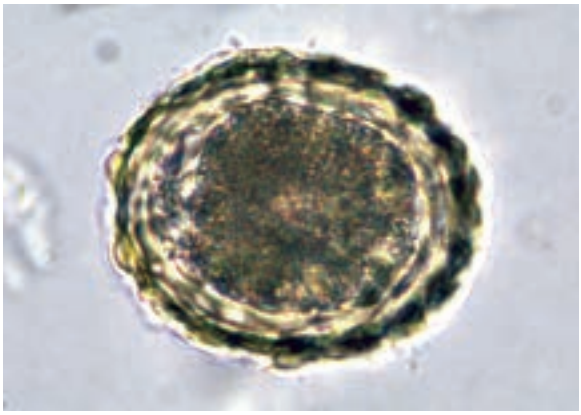

FIGURE 357-2. Fertilized, unembryonated egg of *Ascaris lumbricoides*. (Reproduced with permission from Dickson Despommier.)

Climate is an important determinant of disease in that warm temperature and adequate moisture are required for embryonation of eggs in soil. In endemic areas, the prevalence and intensity of infection with *A. lumbricoides* increase dramatically during the first 2 to 3 years of life; it remains high between the ages of 4 and 15 years and then declines during adulthood.[3]

CLINICAL MANIFESTATIONS

Infection with *A. lumbricoides* is usually asymptomatic. Clinical manifestations are associated with heavy worm burdens and can be classified into those resulting from larval migration through the lungs and those due to parasitism of the gastrointestinal tract by adult worms.

When migrating through the lungs, *A. lumbricoides* larvae can induce an intense reaction that is due to both the physical disruption that their migration causes and a dramatic inflammatory hypersensitivity response to secreted antigens. This phenomenon is more common in areas in which transmission is seasonal, such as on the Arabian Peninsula, where outbreaks of pneumonitis typically follow the rainy season because of resumption of transmission. Symptoms may last for 2 to 3 weeks and include the sudden onset of wheezing, dyspnea, paroxysmal nonproductive cough, and high fever. Respiratory symptoms may coincide with or be preceded by urticarial rash, angioedema, abdominal pain, and vomiting. These symptoms usually resolve spontaneously.

With moderate or heavy infections, signs and symptoms can result from obstruction caused by a mass of worms in the small intestine or migration of worms to the biliary tree, pancreatic duct, or appendix. Intestinal obstruction is more common in young children because of their smaller lumen size and is characterized by colicky abdominal pain and vomiting that may progress to signs of intestinal perforation. Hepatobiliary and pancreatic ascariasis is more common in adults, presumably because their biliary tree is large enough to accommodate a migrating worm. Chronic intestinal infection also can manifest as abdominal pain and distention, diarrhea, and nausea. More insidious effects, especially in children, include decreased protein and fat absorption, development of vitamin A and C deficiencies, and lactose intolerance, which together lead to stunted growth and impaired cognitive development.

DIAGNOSIS

The diagnosis of ascariasis is usually made by microscopic examination of a sample of feces for characteristic thick-shelled eggs.[4] Pulmonary ascariasis cannot be diagnosed on the basis of identification of ova in feces because adult worms have not yet matured and begun producing eggs; instead, larvae, as well as eosinophils or Charcot-Leyden crystals (formed from the breakdown of eosinophils), may be visualized on microscopic examination of sputum. Pulmonary disease is also usually marked by peripheral eosinophilia and transient infiltrates on chest radiographs. The diagnosis of intestinal or biliary obstruction caused by *A. lumbricoides* relies increasingly on radiologic evaluation with ultrasound or endoscopic retrograde cholangiopancreatography (ERCP).

TREATMENT

Rx

Intestinal ascariasis is usually treated with a single oral dose of albendazole (Table 357-1). Alternatives include mebendazole, ivermectin, or pyrantel pamoate. No specific treatment is recommended for symptoms of pulmonary ascariasis because the condition is self-limited. In severe cases of biliary obstruction, including cholangitis, ERCP with or without resection of the ampulla of Vater is highly successful and may reduce the need for surgical intervention.

PREVENTION

The definitive means of preventing *Ascaris* infection is improvement of hygiene and proper disposal of human waste. In endemic communities where this is not feasible, control efforts center on regular (at least annual) mass administration of an anthelmintic medication such as albendazole or mebendazole.[5]

HOOKWORM

The Pathogen

Hookworm infection in humans is due almost exclusively to two species: *N. americanus* and *A. duodenale*. Humans may also be incidentally infected by the zoonotic hookworms *Ancylostoma caninum, Ancylostoma braziliensis, Bunostomum phlebotomum,* and *Uncinaria stenocephala*, which can cause self-limited dermatologic lesions known as cutaneous larva migrans (Fig. 357-3). Additionally, *Ancylostoma ceylanicum,* normally a hookworm infecting cats, has been reported to cause intestinal hookworm disease in humans, especially in Asia, whereas *A. caninum* has been implicated as a cause of eosinophilic enteritis in Australia.

Infection occurs when exposed skin comes in contact with infective filariform larvae in contaminated soil or grass. Larvae penetrate the skin, enter the afferent circulation, and migrate to the pulmonary vasculature, where they break into the alveolar spaces, ascend the bronchial tree to the trachea, and are swallowed into the gastrointestinal tract. Larvae undergo two molts to mature into adult worms approximately 5 to 9 weeks after skin penetration.

TABLE 357-1 TREATMENT OF INTESTINAL NEMATODES

NEMATODE	TREATMENT
Ascaris lumbricoides	Albendazole, 400 mg once. Alternatives: Mebendazole, 100 mg bid for 3 days; ivermectin 150-200 μg/kg once, or pyrantel pamoate, 11 mg/kg for 3 days with the maximum daily dose not to exceed 1 g
Hookworm (*Necator americanus* and *Ancylostoma duodenale*)	Albendazole, 400 mg/day for 3 days. Alternatives: Mebendazole, 100 mg bid for 3 days, or pyrantel pamoate, 11 mg/kg for 3 days with the maximum daily dose not to exceed 1 g
Trichuris trichiura	Mebendazole, 100 mg bid, or albendazole, 400 mg/day, for 3 days; or oxantel pamoate, 20 mg/kg, plus albendazole, 400 mg, administered daily for 2 days*
Enterobius vermicularis	Pyrantel pamoate, 11 mg/kg once, with a second dose 2 wk later; maximum dose of 1 g. Alternatives: Mebendazole, 100 mg once, or albendazole, 400 mg once, repeated in 2 wk
Strongyloides stercoralis	Uncomplicated infection: Ivermectin, 200 μg/kg/day for 2 days.† Alternative: Albendazole, 400 mg/day for 7 days
Trichostrongylus spp	Pyrantel pamoate, 11 mg/kg once; maximum dose of 1 g. Alternatives: Albendazole, 400 mg once, or mebendazole, 100 mg bid for 3 days
Capillaria philippinensis	Albendazole, 400 mg/day for 10 days. Alternative: Mebendazole, 200 mg bid for 20 days

*Oxantel pamoate is not approved by the U.S. Food and Drug Administration for this use at the time of this writing.
†Treatment may need to be extended in immunocompromised patients with disseminated disease.

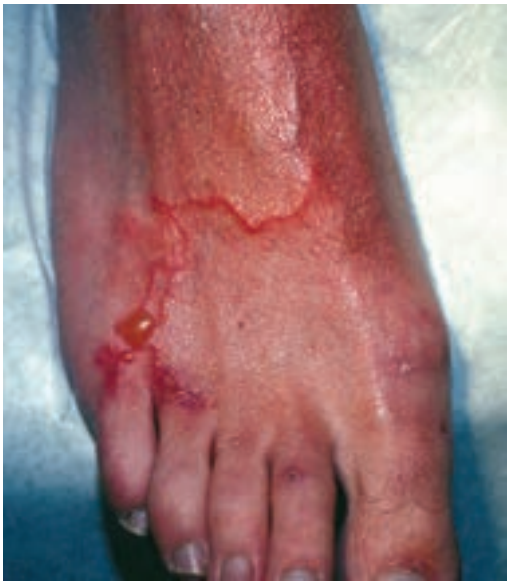

FIGURE 357-3. Typical lesion of cutaneous larva migrans. An erythematous, serpiginous track caused by intradermal migration of a dog (*Ancylostoma caninum*) or cat (*Ancylostoma braziliense*) hookworm larva is apparent. (Reproduced with permission from Gregory L. Zalar.)

Adult hookworms reside in the lumen of the upper part of the small intestine, where they attach to the mucosa by means of cutting teeth (*A. duodenale*) or a rounded cutting plate (*N. americanus*). After mating in the host intestinal tract, a female adult worm produces thousands of eggs per day, which then exit the body in feces; *A. duodenale* female worms lay approximately 28,000 eggs daily, whereas the output from *N. americanus* worms is considerably less, averaging around 10,000 a day. Hookworm eggs hatch in warm, moist soil and release larvae that can infect another host. Humans are the only major definitive host for these two parasites, and there are no intermediate or reservoir hosts. *A. duodenale* survives on average for 1 year in the human intestine, whereas *N. americanus* lives for 3 to 5 years.

EPIDEMIOLOGY

More than 500 million people are infected with hookworms worldwide. *N. americanus* is the most widespread hookworm, whereas *A. duodenale* is more geographically restricted in distribution. The highest prevalence of infection occurs in rural areas of tropical and less developed countries, where environmental and socioeconomic conditions favor transmission. Climate is an important determinant of hookworm transmission, with adequate moisture and warm temperature being essential for larval development in soil. An equally important determinant of infection is poverty and inadequate sanitation and supply of clean water. In endemic areas, prevalence increases with age in young children and reaches a plateau by approximately 10 years of age, whereas the intensity of infection rises at a slower rate during childhood, reaches a plateau by around 20 years, and then increases again in the elderly. Whether such age dependency reflects differences in exposure, acquired immunity, or a combination of both is controversial.

Although cutaneous larva migrans is found throughout the tropics, in the United States it is diagnosed primarily in tourists who have recently returned from a vacation to a tropical beach destination, especially in the Caribbean, Brazil, Mexico, and Southeast Asia. Occasionally, autochthonous cases (originating where found) have been reported in the United States, usually from southeastern coastal states such as Florida and South Carolina. Most commonly, cutaneous larva migrans occurs when exposed skin comes in contact with the larval stages of the dog or cat hookworms *A. caninum* or *A. braziliense,* respectively, present in moist soil or sand (especially on beaches) contaminated with animal feces. Other animal hookworms such as *U. stenocephala* and *B. phlebotomum* are less common causes.

PATHOBIOLOGY

The major pathology of hookworm infection is due to the associated gastrointestinal blood loss and the resulting iron deficiency anemia. Hookworms attach to the intestinal mucosa and secrete enzymes that enable them to invade submucosal tissues and ingest villous tissue and blood. Hemoglobinases within the hookworm's digestive canal enable degradation of human hemoglobin for use as an essential nutrient source. The amount of blood loss is directly related to the total worm burden. *A. duodenale* causes more blood loss than *N. americanus;* each *N. americanus* worm results in a daily blood loss of 0.03 to 0.1 mL, and the corresponding figure for *A. duodenale* is between 0.15 and 0.26 mL.

CLINICAL MANIFESTATIONS

The clinical features of hookworm infection can be separated into acute manifestations associated with larval migration through the skin and other tissues and acute and chronic manifestations resulting from parasitism of the gastrointestinal tract by adult worms. Repeated skin exposure to hookworm larvae can result in a hypersensitivity reaction known as "ground itch," a pruritic erythematous and papular rash that appears most commonly on the hands and feet. In contrast, when zoonotic hookworm larvae penetrate the skin to produce cutaneous larva migrans, most commonly on the feet, thighs, and buttocks, they are unable to complete their life cycle in the human host and eventually die after causing a typical clinical syndrome of intensely pruritic, erythematous serpiginous tracks (see Fig. 357-3). Tracks appear after an incubation period of a few days, can be single or multiple, and advance by millimeters to a few centimeters each day. Vesiculobullous or papular lesions may develop along the tracks, as can secondary bacterial infection as a result of scratching. Untreated, lesions usually heal spontaneously within weeks to months following death of the larvae in the skin.

Migration of hookworm larvae through the lungs may induce mild and transient pulmonary symptoms consisting of dry cough, sore throat, wheezing, and low-grade fever. Acute symptomatic disease may uncommonly result from the oral ingestion of *A. duodenale* larvae, referred to as the Wakana syndrome, which is characterized by nausea, vomiting, pharyngeal irritation, cough, dyspnea, and hoarseness.

Abdominal symptoms and signs caused by hookworm infection are rare. Instead, the manifestations of hookworm disease occur when intestinal blood loss exceeds the nutritional reserves of the host and results in iron deficiency anemia. Usually only moderate- and high-intensity (≥2000 eggs per gram of feces) hookworm infections produce clinical disease, which resembles that of iron deficiency anemia secondary to other causes (Chapter 159). In addition, the protein losses associated with heavy hookworm infection can result in hypoproteinemia and anasarca. As iron deficiency anemia develops and worsens, an infected individual may have weakness, palpitations, fainting, dizziness, dyspnea, lassitude, and headache. Uncommonly, there may be constipation or diarrhea with occult blood in the stool or frank melena, especially in children; there also may be an urge to eat soil (pica). Overwhelming hookworm infection may cause listlessness, coma, and even death, especially in infants. Because children and women of reproductive age have reduced iron reserves, they are at particular risk for symptomatic disease. Severe iron deficiency anemia caused by hookworm during pregnancy can result in adverse consequences for the mother, her unborn fetus (miscarriage, intrauterine growth restriction), and the neonate (anemia, failure to thrive). In children, the anemia and protein malnutrition associated with chronic intestinal parasitism cause long-term impairments in physical and cognitive development.

DIAGNOSIS

The diagnosis of hookworm infection is made by microscopic identification of characteristic eggs in the stool. The eggs of *N. americanus* and *A. duodenale* cannot be distinguished because both are colorless and have a single thin hyaline shell with blunted ends; they range in size from 55 to 75 µm by 36 to 40 µm. Egg concentration techniques, such as the formalin–ethyl acetate sedimentation method, can be used to detect even light infections, although a direct wet mount examination is adequate for detecting moderate-to-heavy infections. In addition, eosinophilia is a common finding in chronic infection and also during larval migration through the lungs.

TREATMENT

Rx

Three daily oral doses of albendazole, 400 mg, is the recommended treatment for intestinal hookworm infection (see Table 357-1). Less effective alternatives include mebendazole, pyrantel pamoate, and single-dose albendazole. For cutaneous larva migrans, although the disease is self-limited and will resolve spontaneously within weeks to a few months, treatment with a single dose of ivermectin will lead to more rapid resolution of symptoms and skin manifestations. Albendazole is an alternative treatment of cutaneous larva migrans.

PREVENTION

The ideal method for preventing hookworm infection is improvement in hygiene and proper disposal of human waste. Until this occurs, in endemic communities control of disease consists of regular (at least annual) mass administration of an anthelmintic medication such as albendazole or mebendazole. For cutaneous larva migrans, tourists should be advised to wear shoes or sandals when walking to and on beaches and to avoid beaches frequented by stray cats and dogs.

TRICHURIASIS

The Pathogen

Infection with *T. trichiura,* also known as *whipworm,* does not have a tissue migratory phase like *A. lumbricoides* and hookworm, and its entire life cycle in the host is limited to the gastrointestinal tract. After embryonated eggs are ingested orally, larvae are released into the small intestine, where they undergo a series of molts before being carried passively to the transverse and descending colon. The narrow anterior end of adult worms embeds in the columnar epithelium, with the posterior portion protruding into the lumen, thereby allowing eggs to be released into feces, which are then passed into the environment, where they must embryonate in warm, moist soil to complete the life cycle. Adult worms can measure up to 50 mm in length and survive in the host for approximately 1.5 to 2 years. The period between ingestion of eggs and detection of eggs in feces is approximately 90 days.

EPIDEMIOLOGY

Like the other soil-transmitted helminths, trichuriasis is most common in poor, mostly rural areas of the tropics and subtropics where disposal of human waste is inadequate. The estimated prevalence worldwide is 800 million. Children are more frequently infected than adults and more likely to have higher worm burdens. Humans are the only host.

CLINICAL MANIFESTATIONS

Most *T. trichiura* infections are asymptomatic. Disease is mostly seen in children because the majority of heavy infections (>10,000 eggs per gram of feces) occur in this age group. Heavy infections can be accompanied by acute dysentery or chronic colitis resembling inflammatory bowel disease and result in abdominal pain and diarrhea. Chronic mucosal inflammation and edema of the colon and rectum can lead to protracted tenesmus that results in rectal prolapse. Chronic *Trichuris* colitis also can lead to malnutrition, impaired growth, and anemia.

DIAGNOSIS

Infection is diagnosed by microscopic identification of the typical barrel-shaped eggs with bipolar plugs in direct or concentrated smears of fecal specimens.

TREATMENT Rx

Although *T. trichiura* responds less effectively than *A. lumbricoides* or hookworm to treatment with albendazole or mebendazole, a 3-day course of one of these two benzimidazole drugs has been the recommended therapy, as listed in Table 357-1. The addition of ivermectin (200 µg/kg) to either drug increases the response rate significantly. More recently, the combination of oxantel pamoate, 20 mg/kg, plus albendazole, 400 mg, administered daily for 2 days, was found to result in higher cure and egg-reduction rates for *T. trichiura* infection then the rates with standard therapy. (Oxanetel is not approved for this by the U.S. Food and Drug Administration at the time of this writing.)

PREVENTION

As for *A. lumbricoides* and hookworm, control of trichuriasis in endemic areas centers on regular mass anthelminthic drug administration, primarily to school-age children, although single doses of albendazole or mebendazole are poorly effective for this intestinal nematode.

ENTEROBIASIS

The Pathogen

E. vermicularis, or pinworm, is transmitted by the fecal-oral route. Embryonated eggs on fingernails, bedding, or clothing are ingested and hatch in the upper small intestine, where they develop into adults before taking residence in the large intestine. Adult worms measuring between 2 and 5 mm live freely in the lumen of the colon, where they mate. Gravid female worms migrate nightly out of the rectum and deposit large numbers of eggs (11,000 per worm) on the perianal and perineal skin, where they rapidly embryonate within 6 hours of being deposited. If they are still on the skin, infective larvae are released that can migrate back through the anus into the rectum (retroinfection); alternatively, autoinfection occurs when eggs are transferred to the mouth via scratching the perianal skin on which eggs have been deposited, commonly in children. In infected females, larvae may also migrate into the genital tract and establish an ectopic infection.

EPIDEMIOLOGY

E. vermicularis is found worldwide and is the most prevalent nematode infection in temperate climes. Transmission is especially frequent in primary schools and daycare centers, where children are in close contact. Up to a quarter of all children worldwide are estimated to be infected with pinworm.

CLINICAL MANIFESTATIONS

Although most pinworm infections are asymptomatic, perianal pruritus is the most common symptom and is caused by an allergic response to worm proteins. The pruritus can be intense and result in chronic sleep deprivation. Rarely, adult *E. vermicularis* may precipitate appendicitis. When hatched larvae migrate into the female genital tract, vulvovaginitis, salpingitis, or peritonitis may develop.

DIAGNOSIS

Pinworm infection is diagnosed by identifying eggs by examination under a microscope on a piece of cellophane tape applied to the perianal region immediately after waking and before bathing. Characteristic *E. vermicularis* eggs are oval and slightly flattened on one side. It is unusual to find eggs in feces or adult worms in the perianal area. Repeated examination may be necessary.

TREATMENT Rx

Pinworm infection is treated with a single dose of pyrantel pamoate, mebendazole, or albendazole, which must be repeated 2 weeks later because the drugs do not kill eggs or developing larvae (see Table 357-1). Given the high rate of transmission, all household members and individuals in close contact with the patient (e.g., other children attending the same daycare center) should be treated as well. Bedding and underclothes should be thoroughly laundered in hot water followed by a hot dryer to kill the eggs.

STRONGYLOIDIASIS

The Pathogen

S. stercoralis, or threadworm, is endemic in warm climates worldwide, including parts of the United States.[6] Infection occurs when exposed skin comes in contact with free-living filariform larvae in soil contaminated with human feces. After penetrating the skin, larvae enter the afferent circulation and travel to the pulmonary vasculature, where they rupture into the alveolar spaces, ascend the bronchial tree to the pharynx, and are swallowed into the gastrointestinal tract. Further development into adult worms occurs in the upper small intestine, where parasites live embedded in the mucosa. Unlike most nematodes, *S. stercoralis* reproduces by parthenogenesis, with no apparent parasitic male worm present in human infection. Female worms begin laying eggs within 25 to 30 days after infection. The embryonated eggs hatch rapidly in the lumen of the small intestine and release noninfectious rhabditiform larvae that migrate to the colon and are excreted in feces. Alternatively, larvae may directly penetrate the colonic mucosa or perianal skin after migrating out of the anus and enter the circulation directly, a mechanism known as *autoinfection.* This phenomenon can lead to maintenance of parasitism in the host for decades.

Infectious filariform larvae develop in the soil by direct transformation from rhabditiform larvae or indirectly from eggs produced by free-living adult worms that have developed from rhabditiform larvae in warm, moist, sandy soil.

A less common type of strongyloidiasis, called *swollen belly syndrome,* has been attributed to infection with *Strongyloides fuelleborni* in infants living in sub-Saharan Africa and Papua New Guinea.

EPIDEMIOLOGY

S. stercoralis infection is endemic in the tropical and subtropical regions of sub-Saharan Africa, Asia, Latin America, and areas of eastern and southern Europe, with a worldwide prevalence of between 30 and 100 million. In the United States, infection is diagnosed most frequently in immigrants, commonly from Southeast Asia, although strongyloidiasis is still endemic in parts of rural Appalachia. *S. stercoralis* also be transmitted sexually through oral-anal contact, most often among men who have sex with men. Cases of transmission through solid organ transplantation have also been reported.

PATHOBIOLOGY

In immunologically competent individuals, infection is usually asymptomatic or is associated with mild gastrointestinal symptoms. The main complications of infection, however, occur in individuals with cell-mediated immunodeficiency such as those chronically taking corticosteroids, renal transplant recipients, patients with Hodgkin disease and other lymphomas, leukemic patients, and those infected with human T-cell lymphotropic virus type 1.[7] In these individuals, the autoinfection cycle of the *S. stercoralis* life cycle can become amplified and lead to a hyperinfection syndrome with a large increase in the total worm load in the infected person. Hyperinfection can lead to life-threatening dissemination of larvae and adult worms to aberrant sites such as the brain, pancreas, and kidneys. For unknown reasons, advanced acquired immunodeficiency syndrome has not been associated

with hyperinfection syndrome or with disseminated strongyloidiasis. Disseminated strongyloidiasis is frequently accompanied by bacterial sepsis, probably as a result of translocation of enteric organisms carried by migrating larvae.

CLINICAL MANIFESTATIONS

Most *S. stercoralis* infections, especially in immunocompetent hosts, are asymptomatic or are associated with only mild gastrointestinal manifestations such as abdominal pain, bloating, and watery diarrhea. Gastrointestinal bleeding, manifested by hematochezia or melena, occurs in less than 20% of those infected. Rare causes of morbidity include small bowel obstruction, paralytic ileus, and a malabsorption syndrome (especially in children).[8]

During the migratory phase of larvae through the lungs, symptoms are rare in immunocompetent patients, although there may be peripheral eosinophilia. However, pulmonary signs and symptoms in immunocompromised persons with hyperinfection syndrome can be severe and resemble those of acute respiratory distress syndrome with dyspnea, productive cough, and hemoptysis accompanied by fever, tachypnea, and hypoxemia.

Migration of filariform larvae from the anus can lead to a dermatologic manifestation known as larva currens, which is characterized by serpiginous erythematous maculopapular tracks that migrate by 5 to 15 cm/hour, primarily on the skin of the buttocks, upper aspect of the thighs, and lower part of the abdomen.

Autoinfection leading to exceptionally high worm loads (i.e., hyperinfection) and disseminated strongyloidiasis can occur in persons with deficient cell-mediated immunity. Because asymptomatic infection with *S. stercoralis* may persist for decades after initial infection, it is important to remember that a change in immune status associated with conditions such as the administration of immunosuppressive drugs following solid organ transplantation may result in hyperinfection syndrome even though the infection was previously asymptomatic. Massive increases in the number of *Strongyloides* larvae because of hyperinfection can present as acute enteritis with severe diarrhea and ulcerative disease of the small and large intestine. During disseminated infection, larvae and sometimes adult worms penetrate the intestinal mucosa, migrate to aberrant sites, including the central nervous system, and result in metastatic abscesses and gram-negative meningitis due to enteric bacteria being carried by the migrating parasites. Less common complications of disseminated disease include glomerulonephritis and minimal-change nephrotic syndrome, acute respiratory distress syndrome, and alveolar hemorrhage. Mortality from hyperinfection and disseminated disease can be high, although early diagnosis and prompt initiation of treatment are associated with improved outcomes.

Infants with swollen belly syndrome caused by *S. fuelleborni* are often seen acutely with abdominal ascites that is not accompanied by diarrhea or fever. The ascites is due to gastrointestinal protein loss; it can be significant enough to cause respiratory impairment and is associated with a high rate of mortality.

DIAGNOSIS

Definitive diagnosis of *S. stercoralis* infection relies on microscopic identification of larvae in feces or other fluids (such as sputum) or tissues. Detection of filariform larvae in feces implies active autoinfection. Intestinal strongyloidiasis can be diagnosed by identification of larvae in direct smears of freshly passed stool, although the sensitivity of examining a single fecal sample is as low as 30%. Sensitivity can be increased by examining multiple fecal specimens, by using concentration techniques, and by plating feces on an agar plate and inspecting for tracks of colonies created by bacteria being dragged by migrating larvae.

Hyperinfection syndrome and disseminated strongyloidiasis can be diagnosed by detection of filariform larvae in duodenal fluid obtained by endoscopy, in sputum, or in bronchoalveolar lavage specimens. Larvae have also been recovered from cerebrospinal fluid, urine, peritoneal washings, skin, and the brains of immunocompromised persons.

Fluctuating eosinophilia is common with uncomplicated intestinal strongyloidiasis, especially during the pulmonary migration phase of initial infection. However, eosinophilia may be absent in patients with hyperinfection and dissemination. In fact, those with hyperinfection and eosinophilia appear to have a better prognosis than do those without eosinophilia.

Serologic diagnosis using an enzyme-linked immunosorbent assay (ELISA) that detects antibodies to filariform larval antigens is very sensitive, even in immunocompromised hosts with disseminated strongyloidiasis, although false positives may occur in cases of coinfection with other nematodes, particularly filaria. Specificity is improved by using the newer luciferase immunoprecipitation systems (LIPS) assays that incorporate *Strongyloides*-specific recombinant antigens. LIPS assays have the additional advantage of rapid reversion to seronegativity following treatment compared to the slow decline in ELISA titers.[9]

TREATMENT

Rx

Uncomplicated intestinal strongyloidiasis can be treated effectively with ivermectin (200 µg/kg body weight daily for 2 days), with cure rates in excess of 90%. Albendazole is an alternative treatment (see Table 357-1). Decreases in serologic antibody titer and eosinophilia indicate a treatment response in the absence of continued exposure. After 6 months, ELISA titers should decrease significantly whereas LIPS assays should revert to negative.

In immunocompromised patients with hyperinfection or disseminated disease, daily treatment should be extended. Some experts recommend continuing treatment until 2 weeks after fecal examinations have become negative (i.e., for one autoinfection cycle). For severely ill patients who are unable to tolerate oral therapy, parenteral veterinary and enema ivermectin preparations have been used. Combination therapy with ivermectin and albendazole may also be used to treat disseminated strongyloidiasis, but data on whether this improves prognosis over monotherapy are lacking.

PREVENTION

In endemic areas, the risk for infection can be reduced by minimizing skin contact with contaminated soil, although elimination of this infection will occur only with improvements in sanitation and treatment of human waste. To prevent hyperinfection in individuals already infected, diagnosis should be attempted before the onset of immunosuppression if possible, such as before organ transplantation or cancer chemotherapy. Anyone who has resided in or traveled to an endemic area should undergo screening for asymptomatic infection, preferably by serologic tests or, if not possible, by microscopic examination of at least three fecal samples for the presence of larvae. Patients with positive screening test results should be treated empirically with ivermectin. Individuals with negative screening test results but unexplained eosinophilia and a history of exposure should also be considered for empirical treatment. In those undergoing hematopoietic stem cell transplantation, documentation of cure with at least three consecutive negative fecal examinations or a negative LIPS assays is recommended before proceeding with transplantation.

UNCOMMON INTESTINAL NEMATODIASES

Humans may serve as accidental hosts for several nematodes that ordinarily parasitize the intestines of other mammals.

Trichostrongylus

Human infection with several different species of the genus *Trichostrongylus* has been reported in Iran, the Far East, and Australia. Humans are incidentally infected when larvae are ingested with leafy vegetables that have been contaminated with soil containing the feces of herbivorous animals. *Trichostrongylus* worms are similar to hookworms in their morphology, appearance of their eggs on fecal examination, and the pathology that they induce. Heavy infections may be accompanied by diarrhea and anemia. Drugs recommended for treatment are pyrantel pamoate, albendazole, or mebendazole (see Table 357-1).

Anisakiasis

Anisakiasis results from ingestion of the larvae of nematodes that normally infect sea mammals such as dolphins, whales, and seals. Larvae of the genera *Anisakis, Phocanema,* and *Pseudoterranova* infect the flesh of a number of saltwater fish species as intermediate hosts. Consumption of raw or undercooked fish, often in the form of sushi or sashimi, results in release of the infective larvae into the stomach, followed by invasion of the stomach or duodenal wall, which causes upper abdominal pain that can be intense. Anisakid worms cannot further develop in humans and die within a few days; an eosinophilic granulomatous reaction may result that mimics a gastric tumor. Diagnosis and treatment are accomplished by endoscopic removal of the parasite. Infection is prevented by cooking or freezing seafood before consumption. Of note, salting, smoking, and marinating fish do not kill anisakid larvae.

Capillaria philippinensis

C. philippinensis can cause a serious intestinal infection that has been reported primarily in the Philippines and Thailand, although it has also been observed in Japan, Taiwan, Korea, and Egypt. Adult worms resemble those of *Trichinella spiralis,* although biologically they mimic *S. stercoralis* in that they have an autoinfectious cycle of reproduction in which larvae can develop into adult worms without having to leave the host. Even though the life cycle is not completely known, this nematode probably parasitizes waterfowl that feed on fish and crustaceans, which serve as intermediate hosts. Humans become infected by eating raw or undercooked infected shrimp or fish. Adult worms travel to the mucosal crypts of the small intestine, where they deposit living larvae, sometimes resulting in overwhelming infection. Clinical disease consists of severe diarrhea associated with anorexia, vomiting, and weight loss. Mortality rates as high as 10% have been reported, with death resulting from severe malabsorption and protein-losing enteropathy. Diagnosis depends on finding eggs or larvae in feces. The treatment of choice is albendazole or mebendazole (see Table 357-1).

Other Intestinal Nematode Infections

Several nematodes that normally parasitize the intestine of nonhuman primates very rarely infect the human gastrointestinal tract. *Oesophagostomum bifurcum* infections have been reported from Africa, Asia, and South America and result from the oral ingestion of infective larvae. Adult worms can result in the formation of nodules in the intestinal wall that may be manifested as abdominal masses. Finally, infection with *Ternidens diminutus* has been reported to result in colonic ulcerations and nodular lesions in people living in southern Africa.

Grade A References

A1. Keiser J, Utzinger J. Efficacy of current drugs against soil-transmitted helminth infections: systematic review and meta-analysis. *JAMA*. 2008;299:1937-1948.

A2. Steinmann P, Utzinger J, Du ZW, et al. Efficacy of single-dose and triple-dose albendazole and mebendazole against soil-transmitted helminths and *Taenia* spp.: a randomized controlled trial. *PLoS ONE*. 2011;6:e25003.

A3. Knopp S, Mohammed KA, Speich B, et al. Albendazole and mebendazole administered alone or in combination with ivermectin against *Trichuris trichiura:* a randomized controlled trial. *Clin Infect Dis*. 2010;51:1420-1428.

A4. Speich B, Ame SM, Ali SM, et al. Oxantel pamoate-albendazole for *Trichuris trichiura* infection. *N Engl J Med*. 2014;370:610-620.

A5. Suputtamongkol Y, Premasathian N, Bhumimuang K, et al. Efficacy and safety of single and double doses of ivermectin versus 7-day high dose albendazole for chronic strongyloidiasis. *PLoS Negl Trop Dis*. 2011;5:e1044.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Dold C, Holland CV. Ascaris and ascariasis. *Microbes Infect*. 2011;13:632-637.
2. Betson M, Nejsum P, Bendall RP, et al. Molecular epidemiology of ascariasis: a global perspective on the transmission dynamics of ascaris in people and pigs. *J Infect Dis*. 2014;210:932-941.
3. Knopp S, Steinmann P, Keiser J, et al. Nematode infections: soil-transmitted helminths and trichinella. *Infect Dis Clin North Am*. 2012;26:341-358.
4. Schmitt BH, Rosenblatt JE, Pritt BS. Laboratory diagnosis of tropical infections. *Infect Dis Clin North Am*. 2012;26:513-554.
5. McCarty TR, Turkeltaub JA, Hotez PJ. Global progress towards eliminating gastrointestinal helminth infections. *Curr Opin Gastroenterol*. 2014;30:18-24.
6. Greaves D, Coggle S, Pollard C, et al. *Strongyloides stercoralis* infection. *BMJ*. 2013;347:f4610.
7. Marcos LA, Terashima A, Canales M, et al. Update on strongyloidiasis in the immunocompromised host. *Curr Infect Dis Rep*. 2011;13:35-46.
8. Mejia R, Nutman TB. Screening, prevention, and treatment for hyperinfection syndrome and disseminated infections caused by *Strongyloides stercoralis*. *Curr Opin Infect Dis*. 2012;25:458-463.
9. Levenhagen MA, Costa-Cruz JM. Update on immunologic and molecular diagnosis of human strongyloidiasis. *Acta Trop*. 2014;135:33-43.

REVIEW QUESTIONS

1. A 21-year-old woman living in rural Indonesia is in her third trimester of pregnancy and is found to have severe microcytic anemia, with a hematocrit of only 19%. She denies vaginal bleeding but consistently has darkened stools that are positive when tested for occult blood. Blood laboratory testing is also remarkable for eosinophilia and low serum albumin concentration. Which of the following parasitic infections is the most likely to result in this syndrome?

A. Enterobiasis
B. Amoebiasis
C. *Diphyllobothrium latum* infection
D. Hookworm infection
E. Strongyloidiasis.

Answer: D Intestinal blood loss is the principal clinical manifestation of hookworm infection, which can lead to iron deficiency and microcytic anemia. Pregnant women and children are at greater risk due to their higher iron needs. *Diphyllobothrium latum* infection is associated with development of megaloblastic anemia as a result of vitamin B_{12} deficiency and is thus associated with an elevated, rather than low, red blood cell mean corpuscular volume. Infection with *Enterobius vermicularis, Entamoeba histolytica,* or *Strongyloides stercoralis* is not usually associated with intestinal blood loss or iron deficiency.

2. A couple has just returned from their honeymoon in the Caribbean complaining of an intensely pruritic rash on their feet, buttocks, and legs. On examination, the rash consists of multiple raised, serpiginous red tracks. Most of their time was spent relaxing on a beach owned by the resort where they were staying. Which of the following is the most likely causative organism?

A. *Dracunculus medinensis*
B. *Ancylostoma duodenale*
C. *Ancylostoma braziliense*
D. *Strongyloides stercoralis*
E. *Enterobius vermicularis*

Answer: C The description of the skin lesions is typical of cutaneous larva migrans. This syndrome is most commonly caused by penetration of the skin by the larvae of the dog or cat hookworms, *Ancylostoma caninum* or *Ancylostoma braziliense*. *Strongyloides stercoralis* also may cause a migrating serpiginous dermatologic rash known as larva currens, but it is found most commonly on the buttocks and lower back because the larvae that cause the tracks originate from the anus. *Dracunculus medinensis* results in a dramatic ulcerative, painful, nonmigratory skin lesion from which the adult female worm emerges to release eggs into the environment. Neither *Ancylostoma duodenale* nor *Enterobius vermicularis* produce rashes.

3. A 60-year-old male kidney transplant recipient from rural Eastern Kentucky is admitted to the hospital with a 14-day history of fever and profuse diarrhea. His medications include prednisone, tacrolimus, mycophenolate mofetil, and lisinopril. Laboratory examination is remarkable for a leukocyte count of 13,000/mm^3, of which 25% are eosinophils. In the hospital, blood cultures are repeatedly positive for *Escherichia coli*. Besides treating the bacteremia with antibacterial therapy, which of the following medications is indicated to treat the most likely underlying predisposing parasitic infection?

A. Pyrantel pamoate
B. Ivermectin
C. Praziquantel
D. Mebendazole
E. Metronidazole

Answer: B The patient's clinical picture is suggestive of disseminated strongyloidiasis. *Strongyloides stercoralis* is endemic in rural parts of Appalachia and infection can be maintained in a host for decades due to the autoinfection cycle of this parasite. Immunosuppression, particularly with corticosteroids, may lead to hyperinfection and dissemination of larvae to aberrant locations in the body. Enteric bacteria such as *E. coli* can be carried by the migrating larvae throughout the body, resulting in gram-negative sepsis, meningitis, and abscess formation. Ivermectin is the treatment of choice for strongyloidiasis. Although albendazole is a second-line therapy, mebendazole, also a benzimidazole, has poor oral bioavailability and would therefore not be an ideal choice. Pyrantel pamoate and praziquantel are anthelminthics but have poor activity against *S. stercoralis*.

4. A young boy who recently emigrated from Guatemala to the United States with his parents is brought by them to the emergency room because of lethargy and severe abdominal pain that has persisted for about 2 days. He is unable to walk due to the pain. On physical examination, he is afebrile; bowel sounds are absent, and there is significant diffuse abdominal tenderness that is more pronounced in the right lower quadrant. Radiographic studies suggest acute intestinal obstruction at the level of the ileum. How did this child become infected with the most likely causative organism?

A. Drinking water contaminated with cysts
B. Ingesting soil contaminated with embryonated eggs
C. Walking barefoot outside, thus permitting larvae in contaminated soil to penetrate the skin
D. Swimming in lakes that harbor snails and cercariae
E. Eating insufficiently cooked meat infected with cysts

Answer: B Infection with *Ascaris lumbricoides* causing intestinal obstruction is the most likely diagnosis. This nematode is a soil-transmitted helminth that to be infective to humans requires that eggs undergo a phase of development and embryonation in warm, moist soil. Infection results from ingestion of contaminated soil. Drinking contaminated water containing cysts is the means of transmission for protozoan organisms such as *Giardia intestinalis* and *Entamoeba histolytica*. Meat is the vehicle for *Taenia* tapeworm or *Trichinella spiralis* infections, skin exposure to water containing the intermediate snail host is the means of acquiring schistosomiasis, and skin contact with contaminated soil pertains to the transmission of hookworm or *Strongyloides stercoralis*.

5. A young boy in Kenya is taken to the doctor because of episodes of blood-streaked diarrhea, fever, and lower abdominal pain. Examination of a smear of his feces under the microscope reveals many oval eggs with bipolar plugs. Which of the following complications might occur if the child is not treated for the causative helminth?

A. Obstruction of the duodenum
B. Rectal prolapse
C. Asthma-like symptoms
D. Gram-negative meningitis
E. Severe megaloblastic anemia

Answer: B Usually, light infections with the whipworm *Trichuris trichiura* are asymptomatic. However, heavy infections may produce symptoms that include those of dysentery, as in this child. Failure to treat with an anthelminthic could result in rectal prolapse due to the chronic straining induced by the sensation of rectal fullness caused by the presence of many whipworms in the rectum and colon. Although microcytic anemia may develop due to intestinal blood loss associated with *T. trichiura* infection, megaloblastic anemia resulting from vitamin B_{12} deficiency is associated with infection with the tapeworm *Diphyllobothrium latum*. *T. trichiura* is small and the adult worms are located in the large intestine, making obstruction of the duodenum unlikely. Also, whipworm has no pulmonary phase like *Ascaris lumbricoides* or hookworm and is therefore not associated with respiratory symptoms. Meningitis due to gram-negative bacilli is associated with disseminated strongyloidiasis but not with whipworm infection.

358

TISSUE NEMATODE INFECTIONS

DAVID J. DIEMERT

The tissue nematodes can be broadly divided into those for which humans serve as the principal host (the filariases) and those that usually infect animals but can incidentally infect humans. Several zoonotic nematodes, such as *Toxocara, Trichinella,* and *Angiostrongylus,* can infect humans through accidental oral ingestion of helminth eggs or larvae but are unable to complete their life cycle in the host. Clinical manifestations are primarily due to the aberrant migration of larvae through various tissues.

TOXOCARIASIS

DEFINITION

Accidental ingestion of embryonated eggs of the dog roundworm *Toxocara canis,* or less frequently of the cat ascarid *Toxocara cati,* can lead to the clinical syndromes of visceral larva migrans and ocular larva migrans.[1] Symptoms are caused by the migration of larvae through the body, invading organs to cause serious disease and even death.

EPIDEMIOLOGY

Toxocara infections in animals are ubiquitous in their distribution throughout the world and are more common in younger animals than in adults. In humans, children are most frequently infected, probably due to exposure to soil contaminated with dog or cat feces when playing in sandboxes or playgrounds. Visceral larva migrans occurs most commonly in children younger than 5 years of age, whereas ocular larva migrans typically affects older children between the ages of 5 and 10 years.

PATHOBIOLOGY

The life cycle of *Toxocara* in the animal host resembles that of *Ascaris lumbricoides* in humans; larvae penetrate the intestinal wall after being released into the lumen from ingested eggs, migrate through the vasculature to the lungs, enter into the alveolar space, and ascend the bronchial tree until they are swallowed back into the gastrointestinal tract, where they develop into adult worms that can produce eggs. However, when embryonated *Toxocara* eggs are ingested by humans, the released larvae migrate throughout the body (most commonly to the lungs, the liver, the central nervous system [CNS], and occasionally the eyes) but cannot develop into adult worms. Ultimately, the larvae die, inducing significant immediate-type and delayed-type hypersensitivity reactions that result in eosinophilic granuloma formation. Visceral larva migrans and ocular larva migrans seem to be mutually exclusive, suggesting that different *Toxocara* strains may have different tissue tropisms. Alternatively, visceral larva migrans may result from repeated infections whereas ocular larva migrans may be a manifestation of infection in children who have not been previously sensitized.

CLINICAL MANIFESTATIONS

Most *Toxocara* infections in humans are asymptomatic. Visceral larva migrans is characterized by low-grade fever, pulmonary symptoms including cough and wheeze,[2] and less frequently hepatosplenomegaly accompanied by right upper quadrant pain. Symptoms appear gradually and resolve over 4 to 8 weeks. Myocarditis, nephritis, and CNS disease are less common. CNS involvement can result in seizures, encephalopathy, neuropsychiatric symptoms, or eosinophilic meningoencephalitis. A more subtle syndrome referred to as *covert toxocariasis* has been described that may result from less dramatic migration of larvae through organs. For example, *T. canis* has been suggested as an environmental risk factor for asthma in inner-city populations.

Ocular larva migrans typically manifests as unilateral visual impairment that is sometimes accompanied by strabismus.[3] The degree of vision loss depends on the particular ocular structure involved, and permanent blindness can occur. Ocular larva migrans involving the retina can be difficult to distinguish from other causes of focal intraretinal lesions, such as retinoblastoma or tuberculosis.

DIAGNOSIS

Toxocariasis can be presumed on the basis of a compatible clinical presentation and history of exposure to dogs or cats. Eosinophilia and hypergammaglobulinemia are often present. Serologic diagnosis by an enzyme-linked immunosorbent assay (ELISA) that employs antigens secreted by second-stage larvae to measure anti-*Toxocara* antibodies may be informative. Although establishment of sensitivity and specificity is difficult because of the inability to make a definitive parasitologic diagnosis, it is estimated that when a titer cutoff of 1 : 32 is used, the sensitivity of the ELISA is 78% for visceral larva migrans but lower for ocular larva migrans; specificity for both syndromes is greater than 90%. Biopsy of tissues to document the presence of larvae is not recommended because of low sensitivity.

Computed tomography and fluorescein angiography may be helpful in the diagnosis of ocular larva migrans, especially to differentiate it from retinoblastoma and other causes of intraocular space-occupying lesions. Anti-*Toxocara* antibodies can be detected in aqueous and vitreous humor fluid, and elevated levels relative to serum are suggestive of ocular larva migrans.

TREATMENT

Rx

Albendazole (400 mg twice daily for 5 days) is the treatment of choice for toxocariasis (Table 358-1). Mebendazole is not recommended because of its poor oral bioavailability. In patients with severe pulmonary, cardiac, or neurologic involvement, corticosteroids may reduce the severity and duration of symptoms. Ocular larva migrans is treated by vitrectomy, corticosteroids, or albendazole. See also Chapter 344.

TABLE 358-1 TREATMENT OF TISSUE NEMATODE INFECTIONS

NEMATODE INFECTION	TREATMENT
Toxocariasis	Albendazole, 400 mg twice daily for 5 days
Trichinellosis	Albendazole, 400 mg twice daily for 8-14 days*
Angiostrongyliasis	Treatment with albendazole or mebendazole is controversial but may relieve symptoms.
Gnathostomiasis	Albendazole, 400 mg daily for 3 wk Alternative: Ivermectin, 200 µg/kg/day for 2 days ± Surgical removal
Lymphatic filariasis	Diethylcarbamazine, 6 mg/kg/day divided in three doses for 12 days† Alternative: Doxycycline, 100-200 mg/day for 4-8 wk
Onchocerciasis	Ivermectin, 150 µg/kg once, repeated every 6-12 mo until resolution of symptoms
Loiasis	Diethylcarbamazine, 9 mg/kg/day divided in three doses for 21 days†,‡
Mansonella perstans	Doxycycline, 200 mg/day for 6 wk
Mansonella ozzardi	Ivermectin, 200 µg/kg once
Mansonella streptocerca	Diethylcarbamazine, 6 mg/kg/day divided in three doses for 12 days Alternative: Ivermectin, 150 µg/kg once
Dracunculiasis	Extraction of the adult worm

*Treatment is effective only if it is initiated during the intestinal phase of infection.
†Start at a dose of 50 mg on the first day, 50 mg three times daily on the second, 100 mg three times daily on the third, and then 6 mg/kg/day on day 4 onward.
‡Repeated treatment after 6 mo is often necessary if symptoms and eosinophilia persist.

PREVENTION

Visceral larva migrans and ocular larva migrans may be prevented by periodic anthelminthic treatment of dogs and cats, proper disposal of pet feces, covering sandboxes, washing hands after playing with dogs or cats, and keeping children from playing in areas where pets have defecated.

BAYLISASCARIASIS

Baylisascariasis is a rare zoonosis caused by infection with the ascarid *Baylisascaris procyonis,* which is primarily a nematode parasite of raccoons and other small carnivores. In North America, infection is most commonly associated with contact with raccoons or environments contaminated with their feces and occurs predominantly in infants and young children who ingest the embryonated eggs while playing with soil. Clinically, disease is manifested as neural larva migrans because of the larval invasion of the CNS after release from ingested eggs in the gastrointestinal tract. Characteristic findings include fever and altered mental status accompanied by focal neurologic deficits and seizures; examination of the cerebrospinal fluid (CSF) reveals eosinophilic meningitis. *B. procyonis* also has been associated with ocular larva migrans, usually in otherwise healthy adults. Infection can be fatal or result in significant permanent neurologic or visual impairment. Once neural larva migrans is present, response to treatment with anthelmintics is poor, although corticosteroids may be helpful. Successful use of photocoagulation in ocular larva migrans has been reported. Prophylactic albendazole (25 mg/kg/day for 20 days) started within days of an exposure may prevent clinical disease.

TRICHINELLOSIS

DEFINITION

Trichinella infects a range of mammalian hosts, with the domestic pig serving as the most important reservoir worldwide. Humans are infected through eating raw or undercooked pork or other meats of domestic or wild animals that are contaminated with larvae that are encysted in muscle tissue. Although larvae develop into adults in the human intestinal tract, mate, and produce offspring larvae, clinical disease is characterized not so much by the intestinal infection as by the newborn larvae that penetrate the intestinal wall and disseminate throughout the body.

EPIDEMIOLOGY

Several different species of *Trichinella* can cause disease in humans, although *Trichinella spiralis* is the most important. *T. spiralis* is enzootic throughout the world in omnivorous and carnivorous wild animals, including bears, boars, and rats. *Trichinella nativa* affects predominantly carnivores (e.g., walruses, polar bears, and seals) living in the Arctic and subarctic regions of North America, Europe, and Asia. *Trichinella* is introduced into domestic animal populations, usually pigs or horses, by giving them unprocessed feed containing meat scraps of infected animals, most commonly rats. Because of regulations banning this practice in the United States, Canada, and the European Union, human infection by consumption of undercooked or smoked pork products or beef contaminated with the encysted larvae has been virtually eliminated, although it still occurs throughout the rest of the world. Instead, ingestion of poorly cooked wild game, especially bear or boar meat, is now the most common source of infection in these places. An important source of infection with *T. nativa* in Alaskan and Canadian Arctic native populations is eating of uncooked walrus meat.

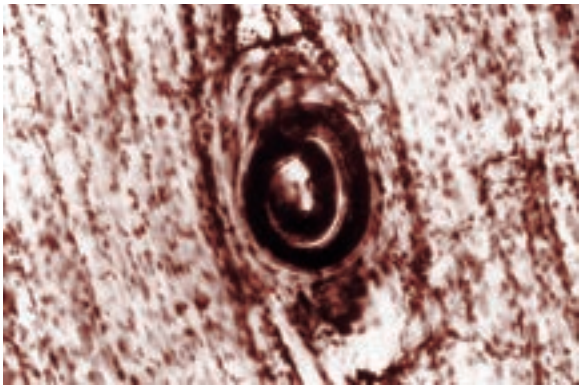

FIGURE 358-1 Nurse cell in muscle tissue containing a larva of *Trichinella spiralis*. (Courtesy Dr. I. Kagan, Centers for Disease Control and Prevention, Atlanta, GA.)

PATHOBIOLOGY

Trichinellosis results from ingestion of striated muscle containing encysted infective larvae.[4] Larvae are released from muscle tissue by digestive enzymes in the stomach and then migrate to the upper two thirds of the small intestine, where they rapidly develop into sexually mature adult worms after only 2 days. Adults live embedded in the columnar epithelium, where they grow to a length of 3 mm (females) or 1.5 mm (males). Females begin producing newborn larvae within 5 days of mating. Adult worms remain viable for an additional 3 to 5 weeks, after which acquired immunity develops that leads to their expulsion from the host.

Newborn larvae possess a sword-like stylet in their oral cavity that permits them to penetrate the lamina propria and enter the lymphatic and blood vessels of the host, allowing them to migrate throughout the body. Larvae enter all types of cells, where they usually die, with the exception of striated skeletal and cardiac muscle cells. Unique among nematodes, mature *Trichinella* larvae have an intracellular phase, developing and transforming muscle cells into "nurse cells" that support larval growth and development (Fig. 358-1). In nurse cells, *Trichinella* larvae can survive for decades. Although nurse cells do not result in any disease in most mammals, they can induce an eosinophilic granulomatous reaction in humans that may result in significant tissue damage and dysfunction.

CLINICAL MANIFESTATIONS

Clinical disease in humans can be divided into an initial intestinal phase followed by a systemic or muscle phase. The initial phase of infection that occurs within days after ingestion of larvae may be associated with mild diarrhea, abdominal pain, and vomiting. This phase is self-limited and usually resolves spontaneously within 10 days.

The systemic dissemination of *Trichinella* larvae can result in myocardial, pulmonary, and focal neurologic manifestations,[5] although usually only in the most heavily infected persons. This systemic phase of infection usually begins 2 to 3 weeks after ingestion of infective larvae and may persist for several weeks. Clinical manifestations typically include fever, periorbital or facial edema, a diffuse inflammatory myositis (Chapter 421) that is characterized by myalgias and muscle tenderness, and petechial hemorrhages most easily observed in the subungual skin and conjunctivae. Larval invasion of the myocardium can lead to myocarditis (Chapter 60) that may result in heart failure or arrhythmias.

As with most nematodes, the severity of symptoms is related to the total worm burden. Because adult worms are incapable of reproducing within the host, the number of encysted larvae ingested is the most important determinant of the number of larvae that invade muscle and other tissues.

DIAGNOSIS

A diagnosis of trichinellosis should be suspected in individuals with a compatible clinical presentation, a history of eating raw or undercooked meat, eosinophilia, and increased muscle enzymes such as creatine kinase and lactate dehydrogenase. Definitive diagnosis depends on visualization of nurse cells in a muscle biopsy specimen or detection of *Trichinella*-specific DNA by the polymerase chain reaction (PCR) technique, although this tool is not widely available. Findings on muscle biopsy may be normal even in heavily infected patients because of sampling error. Detection of anti-*Trichinella* antibodies can be very useful in making a diagnosis; ELISA is the most commonly used method. Antibodies can be detected as early as 12 days after initial infection.

TREATMENT

Rx

If patients present during the intestinal phase of infection, albendazole is recommended at a dosage of 400 mg twice daily for 8 to 14 days to kill the adult worms and thus prevent release of more newborn larvae (see Table 358-1). Although it is not known if albendazole is effective against newborn larvae, administration of this drug during the systemic phase of infection could potentially worsen symptoms by exacerbating the host inflammatory response to dying larvae. Treatment of severe systemic disease, including myocarditis and neurologic disease, should be directed toward reducing inflammation, most commonly with corticosteroids, although albendazole also should be given in such cases because corticosteroids may delay expulsion of adult worms from the intestine, thus increasing the number of newborn larvae that may be released. Symptomatic treatment with antipyretics and analgesics also should be considered. See also Chapter 344.

PREVENTION

Trichinella infection is prevented by thoroughly cooking meat products to kill the encysted larvae. Freezing meat solid at −20° C for at least 3 days will kill *T. spiralis* but not all other species of *Trichinella*. Of note, curing and smoking techniques do not reliably kill this nematode.

ANGIOSTRONGYLIASIS

DEFINITION

Angiostrongylus cantonensis and *Angiostrongylus costaricensis* are nematodes that normally infect rodents, primarily rats.[6] The adult worms of *A. cantonensis,* or rat lungworm, inhabit the pulmonary arteries of rodents; larvae are produced that migrate to the pharynx, are swallowed, and then are passed in the feces. Mollusks such as snails, slugs, and prawns serve as intermediate hosts until they are ingested by definitive hosts. Released larvae migrate to the brain, where they develop into immature adult worms before traveling to the pulmonary vasculature to become sexually mature adults. Humans are incidentally infected after eating poorly cooked or raw intermediate mollusk hosts; larvae can migrate to the CNS but cannot develop further. Fresh vegetables also may serve as a vehicle of human infection if they are contaminated with parts of mollusks containing infective larvae. As opposed to *A. cantonensis,* the larvae of *A. costaricensis* can develop into sexually mature adult worms in the local lymphatics and mesenteric arterioles of humans and release eggs and larvae into the intestinal tissue, causing an intense eosinophilic granulomatous reaction.

EPIDEMIOLOGY

Human infections with *A. cantonensis* occur mainly in Southeast Asia and the South Pacific and less frequently in Brazil, the Caribbean, and recently, in the United States in Louisiana. Abdominal angiostrongyliasis due to *A. costaricensis* has been reported mainly in Latin America, mostly in young children.

CLINICAL MANIFESTATIONS

In human cases of *A. cantonensis* infection, ingested larvae penetrate the intestinal wall and migrate to the brain, the meninges, and less commonly the spinal cord and eye. Fever, severe headache, meningismus, nausea, vomiting, seizures, and focal neurologic deficits may develop. *A. costaricensis* infection can mimic appendicitis with right-sided abdominal pain, vomiting, and fever. Less frequently, gastrointestinal bleeding may occur.

DIAGNOSIS

Diagnosis of *A. cantonensis* infection is based on a history of ingestion of potentially contaminated food, the presence of peripheral eosinophilia, and detection of eosinophils and rarely larvae in the CSF.[7] In neither CNS nor intestinal angiostrongyliasis are larvae or eggs found in the feces, although both may be seen in tissue specimens for *A. costaricensis*. Serology is not commercially available.

TREATMENT AND PREVENTION

Rx

Most patients infected with either species of *Angiostrongylus* recover completely after approximately 2 weeks. The use of anthelmintics (Chapter 344) is controversial, with only a few reports of benefit with albendazole or mebendazole, usually administered in combination with analgesics and corticosteroids to relieve symptoms. Serial lumbar punctures to remove CSF can relieve symptoms of raised intracranial pressure caused by infection with *A. cantonensis*. Proper cooking of food and washing of vegetables can prevent this infection.

GNATHOSTOMIASIS

Gnathostoma spinigerum is an intestinal nematode of dogs and cats; intermediate hosts include tiny crustaceans (copepods), amphibians, freshwater fish, and birds.[8] Accidental human infection occurs throughout the Far East, Thailand, and Latin America, particularly Mexico, on eating of raw or undercooked invertebrate hosts harboring larvae. Larvae are released in the intestine and subsequently migrate through the body but are unable to reach sexual maturity in humans. The most common clinical presentation is migrating painful and pruritic subcutaneous swellings. Eosinophilic meningitis and ocular larva migrans also may occur, with potentially devastating results, including paralysis, subarachnoid hemorrhage, and permanent visual loss. Peripheral eosinophilia, often marked, is usually present; with meningitis, eosinophils are also present in the CSF. Although serologic testing is not available in the United States, laboratories in Thailand and Mexico can perform it. Treatment of cutaneous disease with either a 3-week course of albendazole or a 2-day course of ivermectin (200 µg/kg/day) is recommended (see Table 358-1); for neurologic or ocular involvement, anthelminthics are not advised because they may worsen manifestations. Gnathostomiasis may be prevented by thoroughly cooking fish.

FILARIASES

DEFINITION

The filariases are a group of arthropod-borne nematode infections that are endemic mostly in tropical areas of the world. Instead of residing in the intestine, mature adult filarial worms live in the lymphatics or in connective tissue (Table 358-2). Eight filarial species infect humans: *Wuchereria bancrofti, Brugia malayi, Onchocerca volvulus, Brugia timori, Loa loa, Mansonella streptocerca, Mansonella perstans,* and *Mansonella ozzardi.* The first three are the most common filariases worldwide. Although not usually fatal, these infections can result in significant disability and disfigurement, such as irreversible limb lymphedema (*W. bancrofti* and *B. malayi*) or blindness (*O. volvulus*). Most of the filariases require prolonged exposure for disease to manifest and are therefore uncommon in short-term travelers to endemic areas. The one exception is loiasis, which can occur in returned travelers and expatriates who have spent extended periods in endemic regions.

For all of the filarial nematodes, infection begins with the bite of an infected arthropod vector that deposits infective larvae called *microfilariae* into the skin or blood. Over several months, microfilariae mature into adult worms capable of mating to produce microfilariae that can be ingested by another arthropod vector to complete the life cycle. Adult worms can survive for 5 to 17 years in the human host; microfilariae live for between 5 months and 5 years. For most of the filarial nematodes except *B. malayi* and *M. perstans,* humans are the only definitive host.

Clinical manifestations of infection are varied and are caused by either adult worms or migrating microfilariae. Severity of disease is in most cases

TABLE 358-2 FILARIAL PARASITES OF HUMANS

			MICROFILARIAE		
SPECIES	**DISTRIBUTION**	**VECTOR**	**PRIMARY LOCATION**	**PERIODICITY**	**PRESENCE OF SHEATH**
Wuchereria bancrofti	Tropics worldwide	Mosquitoes	Blood	Nocturnal, subperiodic	+
Brugia malayi	India, Southeast Asia	Mosquitoes	Blood	Nocturnal, subperiodic	+
Brugia timori	Indonesia	Mosquitoes	Blood	Nocturnal	+
Onchocerca volvulus	Africa, Central and South America	*Simulium* black flies	Skin, eye	None or minimal	–
Loa loa	West and Central Africa	*Chrysops* flies	Blood	Diurnal	+
Mansonella perstans	Africa, South America, Caribbean	Midges	Blood	None	–
Mansonella ozzardi	Central and South America, Caribbean	Midges, *Simulium* black flies	Blood	None	–
Mansonella streptocerca	West and Central Africa	Midges	Skin	None	–

proportional to the worm burden harbored by an individual, with relatively light infections commonly being asymptomatic. For several of the filariases, the host inflammatory response to infection becomes apparent only on the death of the adult worm or microfilariae. This may be triggered by exposure to filarial antigens that were previously hidden from the immune system or by release of bacterial endosymbionts of the genus *Wolbachia* that live inside several of the filariae. *Wolbachia* are of the order Rickettsiales and are found in the hypodermis of adult worms and in oocysts, embryos, and microfilariae; they play a critical role in worm viability and fertility.[9]

Diagnosis of filarial infections usually depends on the microscopic examination of either blood or skin specimens for characteristic microfilariae (see Table 358-2). Microfilariae of the different filarial species measure between 170 and 320 µm in length and can be distinguished on the basis of the tissue source of the specimen, the presence or absence of a sheath, and the arrangement of nuclei in the tail. For some filarial species, microfilariae are present in the blood only during certain periods of the day to coincide with biting habits of the arthropod vector, which must be taken into account in timing blood sampling for diagnosis based on microscopy. Serology is not useful in endemic areas because a positive result does not distinguish between previous and current infection, and there is considerable antigenic cross-reactivity between the filariae and other nematodes. Detection of antifilarial antibodies may, however, be useful in returned long-term travelers or expatriates who are not originally from endemic areas.

Diethylcarbamazine (DEC), ivermectin, and albendazole are the principal antifilarial drugs, although they have varying efficacies against the different filarial species (see Table 358-1). DEC is macrofilaricidal (active against the adult worm) for *W. bancrofti*, *Brugia* spp, and *L. loa*, although prolonged or repeated courses are required for this effect. More commonly, the goal is to suppress microfilaria production by adult female worms, which can be achieved by single doses of antifilarial drugs administered alone or in combination annually or biannually. In some cases, reduction of microfilariae in the blood or skin can ameliorate symptoms or prevent progression of disease and interrupt transmission. Furthermore, targeting the *Wolbachia* endosymbionts of some filarial species with extended courses of antibiotics such as doxycycline can be macrofilaricidal.

Lymphatic Filariasis

DEFINITION

The three etiologic agents of lymphatic filariasis, *W. bancrofti, B. malayi,* and *B. timori,* are transmitted to humans through the bite of an infected mosquito.[10] Microfilariae deposited at the bite wound subsequently migrate through the subcutaneous tissue to the lymphatic system, where adult worms develop after approximately 4 to 12 months. The worms reside coiled in lymph nodes and may extend into afferent lymph vessels and surrounding subcutaneous tissue. The lymphatics of the lower and upper extremities and male genitalia are most commonly affected. After mating, females, which measure between 4 and 10 cm in length, twice the length of males, release more than 10,000 microfilariae per day that migrate into the blood stream until ingestion by mosquito intermediate hosts taking a blood meal. In most endemic areas, microfilariae are present in the peripheral blood only at night, when mosquito vectors are most likely to bite. Adult filariae live between 5 and 8 years within the host, although infections lasting for decades have been reported.

EPIDEMIOLOGY

An estimated 120 million people are affected by lymphatic filariasis worldwide; most cases are caused by *W. bancrofti,* and only approximately 10 to 20 million are due to *B. malayi*. *B. timori* is of minor importance, being restricted to southeastern Indonesia. *W. bancrofti* is widely distributed in the tropics, especially in Southeast Asia, the Indian subcontinent, Africa, South America, the Caribbean, and the South Pacific. The major vectors of bancroftian filariasis are *Culex* mosquitoes in urban areas, anopheline mosquitoes in rural areas of Africa, and *Aedes* species in the Pacific.

Humans are the only definitive host for *W. bancrofti*. *B. malayi,* however, can be zoonotic, with both monkey and feline species serving as reservoir hosts and transmission to humans by *Mansonia* mosquitoes. Brugian filariasis is found primarily in India, Malaysia, and other areas in Southeast Asia.

PATHOBIOLOGY

The pathologic process of filarial infections is primarily due to obstruction of lymphatic circulation resulting from damage induced by adult worms, specifically a local inflammatory lymphangitis with components of the innate and adaptive immune response that lead to hypertrophy of the vessel walls.[11] This inflammatory response can be triggered by release of antigens from dead or dying worms, although evidence suggests that it also is induced by living worms and *Wolbachia* antigens that are excreted or secreted into the surrounding milieu. Inflammatory damage is also exacerbated by secondary bacterial and fungal infections.

The initial inflammatory response leads to endothelial and connective tissue proliferation and vessel dilation, which impairs normal lymphatic function and results in lymphedema that is initially reversible. However, worm death results in a granulomatous reaction to released worm and *Wolbachia* antigens. The infiltration of giant cells, plasma cells, eosinophils, and neutrophils can completely occlude the lumen of the lymphatic vessel. Over time, progressive fibrosis and obstruction of lymph flow result in irreversible edema.[12] Although recanalization and collateralization of lymph vessels may occur, lymphatic function remains compromised.

CLINICAL MANIFESTATIONS

The clinical manifestations of lymphatic filariasis cover a wide spectrum from asymptomatic infection to severe chronic lymphatic obstruction accompanied by lymphedema and enlargement of the affected limb or body part (referred to as *elephantiasis*). Other common clinical outcomes include acute episodic lymphadenitis (also called *filarial fever*) and tropical pulmonary eosinophilia. Most infected individuals living in endemic regions are clinically asymptomatic, although microfilariae can be observed in their blood. Despite the absence of a significant inflammatory response, these individuals may nevertheless exhibit dilation of the affected lymphatics on ultrasound, which precedes the onset of clinically apparent disease.

For unknown reasons, newly exposed individuals may develop acute inflammatory reactions that can rapidly progress to chronic or irreversible changes compared with those born in endemic areas. Severe episodes of lymphadenitis, often with genital involvement, may lead to the relatively rapid development of lymphedema and elephantiasis within a year of arrival. Findings usually resolve quickly if the individual is promptly removed from the endemic area. Microfilariae are usually not detected in these patients.

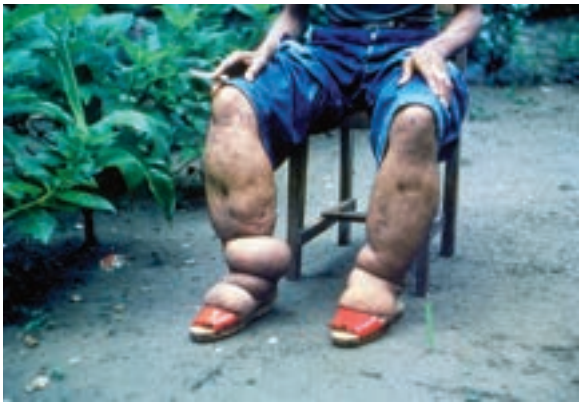

FIGURE 358-2. Elephantiasis, or chronic lymphedema due to infection with *Wuchereria bancrofti*. (Courtesy Centers for Disease Control and Prevention, Atlanta, GA.)

ACUTE LYMPHADENITIS

Acute episodes of retrograde lymphadenitis occur most commonly in adolescents in endemic areas, often in response to dying adult worms. Painful, erythematous enlargement of an affected lymph node, most commonly inguinal, precedes the onset of lymphangitis and is accompanied by fever and chills. Episodes usually last for approximately a week, frequently recur, and can be incapacitating. Defervescence is abrupt and associated with desquamation of the overlying skin. In men, inguinal lymphadenitis can be complicated by epididymitis and orchitis. Patients with filarial fevers may be microfilaremic but often are not.

ELEPHANTIASIS

Repeated episodes of lymphadenitis eventually lead to dilation of the lymphatic vessels, resulting in chronic lymphedema over the course of months to years (Fig. 358-2). The extremities, breasts, and genitalia are most commonly affected, although with *B. malayi* infection, usually only the lower parts of the legs are involved. The edema is initially pitting, but the subcutaneous tissue eventually loses its elasticity, resulting in woody edema with thickening of subcutaneous tissue and hyperkeratosis. Secondary bacterial or fungal infection contributes significantly to the chronic pathologic process of elephantiasis.

In bancroftian filariasis, development of hydrocele is a common manifestation of chronic filariasis in men and sometimes can become massive and debilitating; lymphedema of the vulva is less commonly seen in women. Involvement of the retroperitoneal lymphatics can lead to their rupture to produce intermittent chyluria or chylocele.

TROPICAL PULMONARY EOSINOPHILIA

Tropical pulmonary eosinophilia develops in a small minority of individuals with filarial infections. The syndrome is most commonly seen in young men living in southern India, although it also occurs in Pakistan, Sri Lanka, Southeast Asia, and Brazil. Characteristic clinical findings include nocturnal paroxysmal cough, wheeze, and low-grade fever that are accompanied by weight loss and prominent peripheral eosinophilia. Levels of both total IgE and antifilarial antibodies are typically high. Chest radiographs may show diffuse interstitial infiltrates or mottled opacities in the middle and lower lung fields. Without treatment, chronic restrictive lung disease may develop.

DIAGNOSIS

Definitive diagnosis usually relies on microscopic examination of a Giemsa-stained blood smear for microfilariae. Although thick blood smears are relatively insensitive except in cases of high microfilaremia, concentration or filtering techniques can increase diagnostic yield. Characteristic microfilariae are 250 to 320 μm in length. Collection of blood should be timed according to the known periodicity of the microfilariae.

A rapid immunochromatographic card test is available for *W. bancrofti* (there is no equivalent test for *Brugia* infections), and has the advantage of not requiring nocturnal collection of blood because it detects circulating antigen of the adult worm and not microfilariae. PCR methods also have been developed to detect filarial antigens in blood, although these are not widely available. Serologic detection of antifilarial antibodies is of limited value because of extensive antigenic cross-reactivity with other nematodes. Furthermore, actively infected individuals cannot be distinguished from those previously infected, and those merely exposed but not infected also may have positive serologic test results.

Ultrasound examination of the lymphatic vessels of the spermatic cord of men can be used to visualize the "filarial dance sign," which is pathognomonic for a nest of filarial parasites.

Individuals with elephantiasis may be amicrofilaremic. Diagnosis therefore depends on a compatible clinical history and physical examination in the context of the appropriate epidemiology, and it may be supported by a positive antigen test or, in men, by a suggestive scrotal ultrasonogram. It should be distinguished from podoconiosis, a tropical lymphedema that results from long-term barefoot exposure to red-clay soil derived from volcanic rock and that may be a T-cell–mediated inflammatory disease.

TREATMENT

Rx

Management of lymphatic filariasis differs according to whether the aim is disease control or curative treatment of an individual patient. In endemic areas, annual mass drug administration with a combination of two antifilarial drugs can reduce transmission by decreasing the number of microfilariae in the blood available to biting mosquitoes. These programs use different combinations of single-dose DEC, ivermectin, and albendazole, administered at least once a year. DEC is administered with albendazole except in areas where onchocerciasis or loiasis is also found, in which case ivermectin plus albendazole is used.

It is recommended that all individuals with active infection by a lymphatic filarial parasite, whether symptomatic or asymptomatic, be treated with an antifilarial medication (see Table 358-1). The treatment of choice is DEC (6 mg/kg/day in three divided doses for 12 days). In the United States, DEC is available only through the Centers for Disease Control and Prevention (CDC) drug service (http://www.cdc.gov/laboratory/drugservice/formulary.html). For patients with high levels of microfilariae in the blood, treatment can be started at a low dose of 50 mg daily and scaled up during the first 3 days to reduce side effects of treatment such as fever, headache, dizziness, nausea, vomiting, rash, myalgias, and arthralgias. These normally resolve after a few days of treatment and can be treated with antipyretics, antihistamines, and, if symptoms are severe, corticosteroids.

DEC is both microfilaricidal and partially macrofilaricidal. In individuals who will not be returning to endemic areas, repeated treatments with DEC are often attempted to kill the adult worms instead of just reducing the levels of microfilariae in the blood. Typically, courses of DEC are repeated every 6 to 12 months. Although the adult worm burden is reduced in most treated individuals, all parasites are eliminated in less than a quarter. In men with dancing live adult worms visible by ultrasound in the scrotal lymphatics, serial studies may be performed to monitor the effects of therapy.

Unfortunately, lymphedema due to lymphatic filariasis usually is not reversible with DEC treatment, except in the very early stages. Nevertheless, several management modalities can be employed to limit the chronic sequelae of lymphatic filariasis. Critical among these is the prevention of secondary bacterial and fungal infection through meticulous hygiene and prompt treatment of suspected infections with antimicrobials. Limb elevation, physiotherapy, and use of elastic stockings may slow the worsening of lymphedema. Surgery is usually not indicated except for cases of hydrocele.

Interestingly, treatment directed against the *Wolbachia* endosymbiont has been shown to be effective in killing adult *W. bancrofti* and *Brugia* worms. Doxycycline, 100 or 200 mg daily for 4 to 8 weeks, reduces female worm fertility, with a resulting suppression of microfilaremia for up to a year, and reduces the number of live adult worms. Given the duration of treatment, these regimens are not ideal for disease control programs in endemic countries.

DEC is highly effective in the treatment of tropical pulmonary eosinophilia. Treatment with 6 mg/kg/day for 14 to 21 days results in resolution of symptoms within a week, although relapse may occur even after an interval of years. See also Chapter 344.

PREVENTION

Annual mass treatment with single doses of two antifilarial drugs can significantly reduce the prevalence of infection within a community.[13] In some areas, DEC-fortified table salt has been used to reduce the levels of microfilaremia in affected communities to interrupt the transmission cycle. Vector control through use of insecticide-treated bed nets and residual indoor spraying of insecticides appears to be effective.[14]

Onchocerciasis

DEFINITION

Onchocerciasis, or river blindness, caused by the nematode *O. volvulus,* is transmitted to humans by *Simulium* black flies. Infective microfilariae develop into male and female adult worms over several months and live for 9 to 14 years coiled within subcutaneous fibrous nodules (onchocercomas). Adult females measure between 20 and 70 cm in length and remain confined to the nodules; males are only 3 to 5 cm long and freely migrate through the subcutaneous tissues between nodules to inseminate females. Mature female worms produce up to 1500 microfilariae per day, which leave the nodule to migrate primarily through the skin and ocular tissues.[13] Microfilariae live within the host for 12 to 18 months.

EPIDEMIOLOGY

Onchocerciasis is endemic in equatorial Africa, with small foci in four Latin American countries (Guatemala, southern Mexico, Venezuela, and Brazil) and in Yemen. More than 37 million people are estimated to be infected, 500,000 of whom have significant visual impairment and 270,000 of whom are blind.[15] More than 99% of cases occur in sub-Saharan Africa, with Nigeria being the most highly endemic country. Because *Simulium* black flies require fast-flowing, well-oxygenated water for egg laying and reproduction, onchocerciasis is concentrated around streams and rivers, often in the most fertile farming areas. In communities bordering such waterways in endemic areas, up to 50% of the population can be affected.

Blindness caused by *O. volvulus* results in significant morbidity, long-term disability, and reduced economic productivity. In addition, onchocerciasis has been associated with a reduced life expectancy of at least 10 years compared with that of uninfected individuals in the same area, an effect that appears to be independent of the blindness that develops.

PATHOBIOLOGY

The pathologic changes of onchocerciasis are primarily due to an inflammatory reaction elicited by microfilariae, mostly in the skin, eyes, and lymph nodes. Adult worms contained in nodules are relatively isolated from the host immune response. Tissue damage results from a cell-mediated immune response to dying microfilariae, which becomes more pronounced as infection persists. The degree of tissue damage is directly related to the intensity of infection and magnitude of the host response. Sclerosing keratitis, the major cause of blindness, is caused by an inflammatory reaction to dying intraocular microfilariae that appears to be dependent on T helper cell type 2 (T_H2) cytokines. With time, neovascularization and scarring of the cornea lead to corneal opacification and eventual blindness. In the skin, similar immune responses result in pruritus and angioedema. Ongoing low-grade inflammation in the skin eventually leads to loss of elasticity and atrophy. Chronic inflammatory changes and fibrosis are also seen in lymph nodes.

Like the nematodes responsible for lymphatic filariasis, *O. volvulus* adult worms contain endosymbiotic *Wolbachia* bacteria that are obligatory for the development, survival, and fertility of these worms. Pro-inflammatory *Wolbachia* proteins released by dying microfilariae may be responsible for a significant amount of the immunopathology associated with onchocerciasis. For example, *Wolbachia* antigens have been shown to interact with the innate immune system through a toll-like receptor 2–mediated mechanism (Chapter 45).

CLINICAL MANIFESTATIONS

Onchodermatitis

Onchocerciasis commonly manifests with a diffuse papular dermatitis that is intensely pruritic. In heavily infected individuals in endemic areas, the pruritus is intractable, leading to scratching and excoriation to the point of bleeding and even suicide. Hypersensitivity reactions, scabies, insect bites, and atopic or contact dermatitis should be considered in the differential diagnosis of the acute papular dermatitis seen with onchocerciasis. The skin of affected areas becomes edematous and thickened, losing its elasticity and taking on an orange-peel texture. A lichenified dermatitis (referred to as *sowda*) may occur; it consists of an intensely pruritic eruption limited to one extremity, usually a leg, with hyperpigmented papules and plaques accompanied by edema of the entire limb. Over time, the skin will atrophy and fine wrinkles appear, especially over the buttocks. Pruritus is uncommon at this point. Areas of depigmentation may occur most commonly over the shins, a phenomenon called *leopard skin.*

Subcutaneous Nodules

Subcutaneous onchocercomas containing adult worms are most often palpable over bony prominences. In Africa, the nodules are most commonly found over the hips and lower limbs; in Latin America, they are often located on the head and upper part of the body. Nodules usually measure between 0.5 and 3 cm in diameter and are freely mobile. In lightly infected individuals, such as expatriates, nodules are usually not detectable.

Ocular Lesions

Initial ocular involvement is characterized by conjunctivitis, excess tearing, and photophobia in response to dying microfilariae. At this time, the corneal disease consists of a punctate keratitis or snowflake corneal opacities. Over 20 to 30 years, this leads to sclerosing keratitis, neovascularization, and corneal opacification. The anterior chamber of the eye also may be involved, with iritis, iridocyclitis, and secondary glaucoma. Posterior ocular disease can manifest as chorioretinitis, optic neuritis, and optic atrophy.

Lymphadenopathy

Lymphadenopathy is frequently found in the inguinal and femoral areas in Africa and in the head and neck in Latin America. Advanced disease in the inguinal region can result in the so-called hanging groin, with elongated atrophic skin containing nontender and fibrotic lymph nodes.

DIAGNOSIS

Definitive diagnosis has traditionally been made by observing unsheathed motile microfilariae measuring 200 to 300 µm in length that are released from superficial skin snips. To take a skin snip, a thin piece of skin overlying a bone prominence that has been tented up with a needle is sliced with a scalpel blade, or a corneal-scleral punch instrument is used to obtain a small piece of skin without drawing blood. Avoidance of blood contamination is critical so as to avoid confusion with blood-borne microfilariae in cases patients are coinfected with other filariases. Typically, six snips are taken, one from over each scapula, iliac crest, and lateral aspect of each calf, and then incubated with warm physiologic saline and examined microscopically for motile microfilariae after at least 30 minutes of incubation, although longer periods of up to 24 hours may be necessary. Newer techniques include PCR amplification of filarial DNA directly from skin snips that are far more sensitive than direct visualization. With ocular disease, free microfilariae may be visible by slit lamp examination in the anterior chamber or aqueous humor.

Subcutaneous nodules can be sampled or examined by ultrasound to demonstrate the presence of adult worms. Serologic tests are usually positive for antifilarial antibodies but are not specific because of extensive antigenic cross-reactivity with other nematodes. Eosinophilia is a common but inconsistent finding.

In the past, the Mazzotti test was used to diagnose onchocerciasis. In this test, a challenge dose of DEC was administered to patients suspected of having onchocerciasis; with *O. volvulus* infection, an intense pruritic skin reaction would develop within hours. However, in patients with high-intensity infections, the Mazzotti reaction could be severe and even worsen ocular disease, resulting in permanent visual loss. Therefore, this test is no longer recommended, although some instead suggest the application of a small amount of cream that contains DEC to the skin to provoke a localized Mazzotti reaction.

TREATMENT

Rx

Ivermectin is the treatment of choice for onchocerciasis (see Table 358-1).[16] Administration of a single dose of ivermectin (150 µg/kg) is effective in ameliorating ocular and dermatologic disease by destroying microfilariae and suppressing their release from female worms. Because ivermectin is not active against the encapsulated adult worms, treatment must be repeated every 6 to 12 months, probably for at least 10 years in those without further exposure. For unknown reasons, pruritus in lightly infected expatriates may require more aggressive and frequent treatment for the first 2 years. Within 24 hours of treatment with ivermectin, fever and pruritus may occur in reaction to the dying microfilariae or released *Wolbachia* antigens, especially in those with high pretreatment levels of microfilariae. Use of ivermectin in areas where *L. loa* (see later) is co-endemic should be undertaken with caution because treatment may precipitate severe reactions including encephalopathy in those with high levels of microfilaremia. DEC should never be used for treatment of onchocerciasis because of frequent unacceptable reactions to dying

microfilariae ranging from urticaria and angioedema to hypotension and death. Although the drug suramin (available from the CDC drug service) is active against adult *O. volvulus* worms, because of its excessive toxicity and potentially life-threatening effects, it is used only in rare situations. Surgical removal of palpable nodules has been successful in resolving the infection in some areas, notably in Central America.

Doxycycline, 200 mg/day administered for 4 to 6 weeks, followed by single-dose ivermectin, has been shown to deplete *Wolbachia* endosymbionts from adult worms and to suppress *O. volvulus* embryogenesis and production of microfilariae for up to 18 months. Some experts now recommend this regimen for patients with onchocerciasis who have left an endemic area and will not be re-exposed. See also Chapter 344.

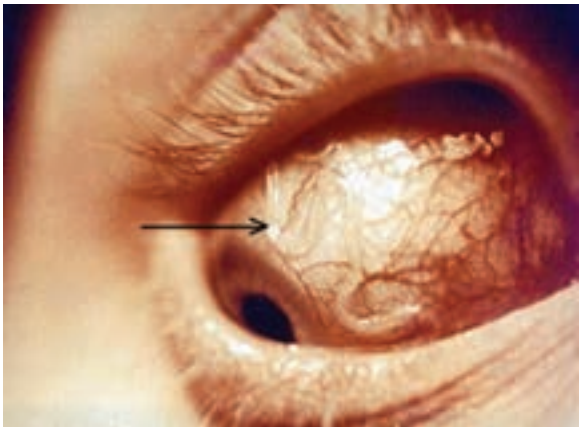

FIGURE 358-3. Adult *Loa loa* worm migrating across the eye (*arrow*).

PREVENTION

Regular mass administration of ivermectin to affected communities forms the core of the global eradication strategy for onchocerciasis.[17] Implementation of this program has been made easier because the drug is donated by the manufacturer. In addition to benefiting infected individuals, mass drug administration reduces the microfilariae available to vectors and thus interrupts the transmission cycle. For travelers to endemic areas, use of insect repellent may be beneficial.

Loiasis

DEFINITION

Loiasis is caused by infection with the filarial nematode *Loa,* otherwise known as the African eye worm. *L. loa* are transmitted by flies of the genus *Chrysops* during a blood meal. Adult worms develop during a period of 1 to 4 years and live for up to 17 years.[18] They migrate freely in the subcutaneous tissue, including the subconjunctiva or sclera of the eye. Adult females measure between 40 and 70 mm in length; males are shorter, measuring between 25 and 35 mm. After mating, females release microfilariae into the blood. *L. loa* microfilariae exhibit a diurnal periodicity coinciding with the feeding habits of *Chrysops,* with microfilaremia peaking around noon.

EPIDEMIOLOGY

Loiasis is endemic in the rain forest regions of central and western Africa. Although accurate numbers are not available, loiasis appears to be most prevalent in Gabon, Cameroon, the Democratic Republic of the Congo, Nigeria, and the Central African Republic. Loiasis requires a shorter period of exposure than other filarial infections and can be seen in returning travelers or expatriates who have spent extended periods in Africa.

PATHOBIOLOGY

Neither adult *L. loa* worms nor microfilariae have any direct pathologic effects. In a subset of infected individuals, a hypersensitivity response, termed a *Calabar swelling,* develops to secretions from adult worms or released microfilariae, resulting in recurrent localized angioedema that often precedes the migrating worm. These patients have very high serum levels of immunoglobulin E (IgE) antibodies and eosinophilia. This reaction is more commonly observed in visitors to endemic areas rather than in native residents. Unlike other filariae, *L. loa* does not contain *Wolbachia* endosymbionts.

CLINICAL MANIFESTATIONS

Most individuals with loiasis are asymptomatic despite being microfilaremic. Clinical manifestations of infection are more common in long-term visitors to endemic areas than in people native to the regions. Recurrent Calabar swellings are the most common finding in these individuals, who are not usually microfilaremic. They are nonerythematous swellings measuring 5 to 20 cm in diameter that typically occur on the extremities and the face and last for a few days. The onset is often preceded by pruritus and pain. On occasion, adult worms may migrate across the subconjunctiva or sclera of the eye in both groups of patients, causing severe pain and inflammation (Fig. 358-3). Rare complications of infection include nephropathy and encephalitis, which usually develop in those with high levels of microfilariae after receiving DEC or ivermectin treatment for other filarial infections. Endomyocardial fibrosis resulting from eosinophilic infiltration of the myocardium has been reported in association with loiasis.

DIAGNOSIS

Definitive diagnosis depends on microscopic examination of a Giemsa-stained blood film for characteristic sheathed microfilariae. Blood should be collected between 10 AM and 2 PM because of the diurnal periodicity of the microfilariae. Because individuals who are not native to endemic areas usually are not microfilaremic, diagnosis relies on a compatible history, clinical findings, peripheral eosinophilia, and elevated antifilarial antibody levels. Adult worms sometimes can be surgically removed while migrating across the eye or through subcutaneous tissues. Calabar swellings must be distinguished from onchocercomas and other causes of angioedema.

TREATMENT

Rx

DEC (9 mg/kg/day for 21 days) is active against both adult *L. loa* worms and microfilariae (see Table 358-1). Treatment usually is increased from a dose of 50 mg/day on the first day to the full dose on the fourth day to minimize the likelihood of treatment-associated complications, the most serious of which are glomerulonephritis and potentially fatal encephalopathy. Treatment-associated complications are more common with high pretreatment microfilarial levels and result from host allergic reactions to dying microfilariae. Antihistamines and corticosteroids may be employed to reduce allergic side effects. Alternatively, apheresis can be used to remove circulating microfilariae before initiation of DEC in these individuals. Albendazole, which is microfilaricidal but has no activity against adult worms, also has been used to reduce microfilaria levels before treatment with DEC. Repeated courses of DEC may be necessary in approximately half of patients before clinical manifestations completely resolve. Persistent or increasing eosinophilia or levels of antifilarial antibodies 6 months after treatment also should prompt re-evaluation for repeated treatment. Adult worms in the eye may be surgically removed.

Ivermectin is microfilaricidal but has no macrofilaricidal effect and may cause toxic encephalopathy in individuals with high microfilaria levels. In areas where onchocerciasis is co-endemic, this infection should be ruled out before the initiation of DEC for loiasis to prevent toxicity from dying *O. volvulus* microfilariae. See also Chapter 344.

PREVENTION

Weekly chemoprophylaxis with DEC administered as a 300-mg dose is effective in preventing loiasis in long-term residents of endemic areas.

Less Common Filarial Infections

MANSONELLA PERSTANS

M. perstans infection occurs throughout central Africa, in northeastern South America, and in parts of the Caribbean. Microfilariae are transmitted by *Culicoides* midges and develop into adult worms that live in serous body cavities, such as the pleural, pericardial, and peritoneal spaces, as well as in mesenteric and retroperitoneal tissues. Most infections are asymptomatic, although painless conjunctival nodules with eyelid edema have been reported. Transient angioedema and Calabar-like swellings, fever, headache, arthralgias, and neurologic manifestations also may occur. Microfilariae do not exhibit periodicity and can be observed on stained blood films. Eosinophilia is common.

M. perstans harbor *Wolbachia* endosymbionts, and treatment with doxycycline, 200 mg/day for 6 weeks, has been shown to be highly effective in suppressing microfilaremia for up to 3 years, suggesting that the treatment is macrofilaricidal.■

MANSONELLA OZZARDI

Infections with *M. ozzardi* occur in Central and South America and parts of the Caribbean, especially Haiti. Vectors include *Simulium* black flies and midges. Adult worms locate to the peritoneal and thoracic cavities or the lymphatics; microfilariae circulate in the blood without periodicity. Infection usually results in asymptomatic eosinophilia, although arthritis and allergic symptoms such as urticaria and lymphadenopathy may occur. Administration of ivermectin as a single dose of 200 µg/kg has been reported to provide long-term suppression of microfilaremia and improvement of symptoms. Neither DEC nor the benzimidazoles are effective.

MANSONELLA STREPTOCERCA

M. streptocerca is endemic in the tropical forest zone of western and central Africa and is transmitted by biting midges. Similar to *O. volvulus,* adult worms live in the subcutaneous tissues, as do microfilariae. In contrast to onchocerciasis, however, microfilariae do not invade the eye. Infection is usually asymptomatic, although a pruritic dermatitis with depigmentation similar to onchodermatitis can affect the trunk and upper extremities. Associated axillary or inguinal adenopathy is common. Microfilariae have characteristic hooked tails and can be visualized in skin snips. In areas where onchocerciasis is co-endemic, skin specimens must be stained to differentiate *M. streptocerca* from *O. volvulus*. DEC is microfilaricidal and macrofilaricidal and is given as 6 mg/kg/day for 12 days. Ivermectin is effective against microfilariae but not adult worms.

ZOONOTIC FILARIAL INFECTIONS

A rare accidental filarial infection of humans with the dog heartworm *Dirofilaria immitis* occurs worldwide. Transmitted by mosquitoes, *D. immitis* microfilariae cannot reach maturity in humans but embolize to the lung after dying in the right ventricle. Most infections are asymptomatic, but some people experience cough, chest pain, and hemoptysis consistent with lung infarction. Chest radiographs demonstrate typical coin lesions that may be mistaken for carcinoma. Other animal filariae, including *Dirofilaria repens* of dogs and *Dirofilaria tenuis* of raccoons, can infect humans and result in subcutaneous nodules that may be migratory. Eosinophilia and antifilarial antibodies are not usually present in zoonotic filarial infections. Surgical removal of lesions is both diagnostic and curative.

DRACUNCULIASIS

Dracunculiasis is a disfiguring disease caused by the nematode *Dracunculus medinensis,* also known as the Guinea worm. Although previously found in India, Pakistan, and Latin America, it is now endemic in only five countries in sub-Saharan Africa (South Sudan, Sudan, Chad, Mali, and Ethiopia) because of concerted efforts at eradication. As of 2014, fewer than 200 cases are thought to exist, with the largest concentration being in South Sudan. Transmission to humans occurs through ingestion of tiny crustacean intermediate hosts called *copepods* that harbor infective larvae. Released larvae penetrate the intestinal wall and migrate to the subcutaneous tissues, where they develop into adult worms. After approximately a year, female worms induce vesicular skin lesions, usually on the lower extremities, that eventually ulcerate. On direct contact with fresh water, the female worm releases thousands of motile larvae that can then complete the transmission cycle by infecting copepods in the water. Adult worms can measure up to a meter in length. Fever and allergic symptoms, including wheezing and urticaria, may precede rupture of the blister or occur with attempts to extract the worm. Secondary bacterial infection of the skin lesions is frequent. Although not usually fatal, dracunculiasis can result in significant disability.

Traditionally, emerging worms are extracted by slowly winding a few centimeters of the parasite on a stick each day, taking care not to break it. Surgical removal can be attempted but may exacerbate allergic symptoms. There is no effective chemotherapy for this infection. Prevention efforts have been highly successful in breaking the transmission cycle and have led to eradication of the parasite from many countries. Strategies include filtering of drinking water through finely woven cloth, education of infected individuals not to enter fresh water, treatment of water sources with larvicides, and provision of safe drinking water from wells.

Grade A References

A1. Taylor MJ, Makunde WH, McGarry HF, et al. Macrofilaricidal activity after doxycycline treatment of *Wuchereria bancrofti*: a double-blind, randomised placebo-controlled trial. *Lancet*. 2005;365:2116-2121.

A2. Hoerauf A, Specht S, Büttner M, et al. *Wolbachia* endobacteria depletion by doxycycline as antifilarial therapy has macrofilaricidal activity in onchocerciasis: a randomized placebo-controlled study. *Med Microbiol Immunol*. 2008;197:295-311.

A3. Coulibaly YI, Dembele B, Diallo AA, et al. A randomized trial of doxycycline for *Mansonella perstans* infection. *N Engl J Med*. 2009;361:1448-1458.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Rubinsky-Elefant G, Hirata CE, Yamamoto JH, et al. Human toxocariasis: diagnosis, worldwide seroprevalences and clinical expression of the systemic and ocular forms. *Ann Trop Med Parasitol*. 2010;104:3-23.
2. Ranasuriya G, Mian A, Boujaoude Z, et al. Pulmonary toxocariasis: a case report and literature review. *Infection*. 2014;42:575-578.
3. Schneier AJ, Durand ML. Ocular toxocariasis: advances in diagnosis and treatment. *Int Ophthalmol Clin*. 2011;51:135-144.
4. Knopp S, Steinmann P, Keiser J, et al. Nematode infections: soil-transmitted helminths and trichinella. *Infect Dis Clin North Am*. 2012;26:341-358.
5. Bruschi F, Brunetti E, Pozio E. Neurotrichinellosis. *Handb Clin Neurol*. 2013;114:243-249.
6. Wang QP, Wu ZD, Wei J, et al. Human *Angiostrongylus cantonensis:* an update. *Eur J Clin Microbiol Infect Dis*. 2012;31:389-395.
7. Senthong V, Chindaprasirt J, Sawanyawisuth K. Differential diagnosis of CNS angiostrongyliasis: a short review. *Hawaii J Med Public Health*. 2013;72:52-54.
8. Herman JS, Chiodini PL. Gnathostomiasis, another emerging imported disease. *Clin Microbiol Rev*. 2009;22:484-492.
9. Bouchery T, Lefoulon E, Karadjian G, et al. The symbiotic role of *Wolbachia* in Onchocercidae and its impact on filariasis. *Clin Microbiol Infect*. 2013;19:131-140.
10. Knopp S, Steinmann P, Hatz C, et al. Nematode infections: filariases. *Infect Dis Clin North Am*. 2012;26:359-381.
11. Taylor MJ, Hoerauf A, Bockarie M. Lymphatic filariasis and onchocerciasis. *Lancet*. 2010;376:1175-1185.
12. Babu S, Nutman TB. Immunology of lymphatic filariasis. *Parasite Immunol*. 2014;36:338-346.
13. Keating J, Yukich JO, Mollenkopf S, et al. Lymphatic filariasis and onchocerciasis prevention, treatment, and control across diverse settings: a systematic review. *Acta Trop*. 2014;135C:85-95.
14. Reimer LJ, Thomsen EK, Tisch DJ, et al. Insecticidal bed nets and filariasis transmission in Papua New Guinea. *N Engl J Med*. 2013;369:745-753.
15. Tamarozzi F, Halliday A, Gentil K, et al. Onchocerciasis: the role of *Wolbachia* bacterial endosymbionts in parasite biology, disease pathogenesis, and treatment. *Clin Microbiol Rev*. 2011;24:459-468.
16. Banla M, Tchalim S, Karabou PK, et al. Sustainable control of onchocerciasis: ocular pathology in onchocerciasis patients treated annually with ivermectin for 23 years: a cohort study. *PLoS ONE*. 2014;9:e98411.
17. Cupp EW, Sauerbrey M, Richards F. Elimination of human onchocerciasis: history of progress and current feasibility using ivermectin (Mectizan(®)) monotherapy. *Acta Trop*. 2011;120(suppl 1):S100-S108.
18. Metzger WG, Mordmüller B. *Loa loa:* does it deserve to be neglected? *Lancet Infect Dis*. 2014;14:353-357.

REVIEW QUESTIONS

1. A 40-year-old woman from southern India has severe lymphedema of her right lower extremity and left breast. She has had this condition for at least a decade, and it is a common condition in her rural village. She now has difficulty going to work and is extremely embarrassed by her deformity to the point at which she tries not to venture outside of her home except at night. What is the most likely diagnosis?

A. Lymphatic filariasis
B. Loiasis
C. Onchocerciasis
D. Dracunculiasis
E. Podoconiosis

Answer: A The most likely diagnosis is lymphatic filariais due to chronic infection with *Wuchereria bancrofti*. Although loiasis, onchocerciasis, and dracunculiasis all have dermatologic manifestations, none result in chronic, irreversible lymphedema. *Loa loa* is associated with transient and migratory Calabar swellings that are hypersensitivity reactions to migrating adult worms; *Onchocerca volvulus* infection can result in generalized chronic dermatitis, and adult *Dracunculus medinensis* worms can rupture through intact skin from their location in the subcutaneous tissues to release eggs into the environment. Podoconiosis can result in chronic lymphedema due to irritant skin exposure to red clay, although this occurs mainly in the lower extremities, almost always starting in the foot.

2. A 28-year-old graduate student in physical anthropology has been experiencing unusual swellings over different parts of her body for the past 2 months. The swellings range in size from 5 to 15 cm and generally last 2 to 4 days before resolving; they are preceded by mild pain and intense pruritus that lasts a couple of hours. The field research for her dissertation is based on observational studies of gorillas in Gabon. A complete blood count is remarkable for very high eosinophilia (4000/mm^3) and an unusual finding on the peripheral blood smear (collected at mid-day) that prompts an urgent page from the laboratory technologist. What is the most likely explanation of the phenomenon seen on the blood smear?

A. Microfilaria of *Onchocercia volvulus*
B. Microfilaria of *Wuchereria bancrofti*
C. Promastigote of *Trypanosoma brucei gambiense*
D. Larva of *Dracunculus medinensis*
E. Microfilaria of *Loa loa*

Answer: E The skin manifestations are most likely Calabar swellings due to migrating adult *Loa loa* worms in the subcutaneous tissue. These are usually transient and are preceded by pain and pruritus. Loaiasis is endemic in Gabon. *Onchocerca volvulus* is also endemic in this country and may result in subcutaneous nodules, although these are usually neither migratory nor transient. In addition, microfilariae of *O. volvulus* are visualized after exiting from skin snip specimens, whereas those of *Loa loa* are seen in blood. The microfilariae of *Wuchereria bancrofti,* which is endemic in Africa, are present in the blood, although their periodicity differs from that of *L. loa*. *W. bancrofti* microfilariae are best visualized in blood collected at night whereas those of *L. loa* are more likely to be seen in a sample collected during the day, preferably between 10 AM and 2 PM. *Trypanosoma brucei,* the causative agent of African sleeping sickness, does not cause migratory skin swellings and is not associated with eosinophilia, although promastigotes can be observed on a Giemsa-stained blood smear. Finally, *Dracunculus medinensis* is associated with nonmotile skin lesions due to the adult female worm rupturing through subcutaneous tissue to the external environment to release larvae; however, no part of the life cycle occurs in the blood.

3. Within 2 weeks of eating jerky that he made from a cougar he shot while hunting, an Oklahoma man develops severe vomiting and diarrhea, followed 2 days later by a fever of 38.7° C, throbbing headache, and myalgias. After a few more days of continued symptoms, he seeks medical attention. His white blood cell count at presentation is 17,300/mm^3, with 25% eosinophils. Which of the following laboratory tests would be most useful in making the correct diagnosis?

A. Serum creatinine concentration
B. Enzyme-linked immunosorbent assay (ELISA) for the suspected pathogen
C. Examination of a Giemsa-stained blood smear using a light microscope
D. Serum creatine kinase concentration
E. Skin biopsy

Answer: D The most likely diagnosis is trichinellosis, probably due to *Trichinella nativa*. Infected individuals, especially those with myalgias and muscle tenderness, usually have elevated levels of serum creatine kinase, likely because *Trichinella* larvae invade striated muscle cells, inducing myositis. Unless the myositis is severe and rhabdomyolysis develops, associated renal dysfunction is rare. At this point in the disease process, seroconversion has likely not yet occurred, and, therefore, the ELISA for *Trichinella* would be negative. Larvae are not usually detectable in a stained blood smear or a skin biopsy, although nurse cells, which are striated muscle cells containing encysted larvae, can be visualized on muscle biopsy specimens.

4. A 7-year-old boy who has never traveled outside of North America develops decreased visual acuity in his right eye. Visible leukokoria is evident to his parents, which prompts a visit to their pediatrician, who consults an ophthalmologist. Fundoscopy reveals a single lesion of the retina. Assuming that his ocular condition is caused by a parasitic worm, which is the most likely means that this child became infected?

A. Being bitten by an infected black fly
B. Eating dirt while playing in a Boston public park
C. Swimming in a lake containing snails
D. Walking barefoot outside while visiting the family's summer cabin in rural West Virginia
E. Eating ceviche while on a family vacation to Mexico.

Answer: B Although several nematode species have been associated with ocular manifestations in humans, *Toxocara canis* and *Toxocara cati* are the most common etiologies of ocular larva migrans and are found worldwide. Children who eat dirt because of pica are most at risk. Dogs and cats naturally harbor the adult parasitic worms, but humans—especially children—may become accidentally infected when embryonated eggs are ingested, even though humans are unsuitable hosts for maturation into adult worms. Retinal involvement occurs when excysted larvae penetrate the small intestine and disseminate throughout the body, including the eye. Due to the inflammatory response, visually distinguishing the condition in situ from a retinoblastoma or ocular tuberculosis can be challenging. *Onchocera volvulus,* which is transmitted by the bite of a *Simulium* black fly, can cause ocular lesions, but these are almost exclusively corneal. In addition, onchocerciasis is endemic now mostly in Africa. Skin exposure to fresh water containing certain species of snails is the means of acquiring schistosomiasis, which does not result in retinal manifestations. Nor does *Strongyloides stercoralis,* which is endemic in certain rural parts of Appalachia and is acquired through skin exposure to soil containing infective larvae.

5. A 35-year-old woman who has recently immigrated to the United States from Cameroon presents with chronic atrophic dermatitis and punctate keratitis. Laboratory investigations reveal a peripheral eosinophil count of 3500/mm^3. Microfilaria are observed in the anterior chambers of both eyes on slit lamp examination. When considering management of this patient, which of the following antiparasitic medications should be contraindicated?

A. Ivermectin
B. Diethylcarbamazine (DEC)
C. Albendazole
D. Pyrantel pamoate
E. Praziquantel

Answer: B This woman most likely has onchocerciasis due to chronic infection with *Onchocerca volvulus,* resulting in the typical skin and ocular manifestations. DEC should never be used for treatment of onchocerciasis because of frequent unacceptable hypersensitivity reactions to dying microfilariae, which can range from urticaria and angioedema to hypotension and death. Ivermectin is the treatment of choice for *O. volvulus,* whereas albendazole, pyrantel pamoate, and praziquantel are neither effective nor detrimental in this condition.

359

ARTHROPODS AND LEECHES

DIRK M. ELSTON

ARTHROPODS

The Pathogens

Arthropods act as disease vectors and cause human injury by means of the direct toxic effects of their venom or via an immune response to their antigens (Tables 359-1 to 359-4). Immediate reactions to stings may be related to histamine, serotonin, formic acid, or kinins contained within the venom. Delayed reactions to bites and stings generally represent a host response to proteinaceous allergens contained within venom or saliva.[1]

Whereas many diseases are spread by a single vector, others have multiple potential vectors. For example, tularemia is commonly acquired from ticks or handling of infected carcasses but also may be spread by deerflies and horseflies. Rickettsiae are spread by both ticks and fleas. It is important to note that the vector may influence manifestations of the disease. For example, *Bartonella* transmitted by a flea may produce bacillary angiomatosis, whereas the same organism transmitted by a louse is much more likely to manifest as endocarditis. Disease manifestations also may reflect the underlying immune status of the host, previous sensitization, or associated comorbidities. For example, hypersensitivity reactions to insect stings are more common in those with an atopic diathesis and anaphylaxis may be a manifestation of underlying mastocytosis. Exaggerated bite reactions (Fig. 359-1) may be a manifestation of an underlying lymphoproliferative disease.

TABLE 359-1 MEDICALLY IMPORTANT ARTHROPODS

Arachnida Acari—mites, ticks Araneida—spiders Scorpionida
Pentastomida—tongue worms
Chilopoda—centipedes
Diplopoda—millipedes
Crustacea Copepoda—*Cyclops, Diaptomus* Decapoda—shrimp, lobster, crayfish, crab
Insecta Anoplura—lice Coleoptera—beetles Diptera—mosquitoes, black flies, midges, horse flies, deer flies, greenheads, tsetse flies, stable flies, sand flies, houseflies, bluebottle flies
Hemiptera—bed bugs, reduviids
Hymenoptera—ants, bees, wasps
Lepidoptera—moths, caterpillars
Siphonaptera—fleas

TABLE 359-2 ARTHROPOD VECTORS

DISEASE	VECTOR
African trypanosomiasis	Tsetse flies
American trypanosomiasis	Triatome bugs
Arboviridae	*Culex* mosquitoes
Babesiosis	*Ixodes scapularis* (deer tick)
Bartonellosis	Fleas, lice, sandflies
Dengue	*Aedes* mosquitoes
Endemic typhus	*Ctenocephalides felis* and *Xenopsylla cheopis* (fleas)
Filariasis	Anopheline and *Aedes* mosquitoes
Human anaplasmosis	*Ixodes scapularis*
Human monocytic ehrlichiosis	*Amblyomma americanum* (lone star tick)
Leishmaniasis	Sandflies
Lyme disease	*Ixodes scapularis*
Malaria	Anopheline mosquitoes
Onchocerciasis	*Simulum* flies
Plague	*Xenopsylla cheopis* and *Pulex irritans* (human flea)
Rickettsial pox	*Liponyssoides sanguineus* (house mouse mite)
Rocky Mountain spotted fever	*Dermacentor variabilis* (dog tick) *Dermacentor andersoni* (wood tick) *Amblyomma americanum*
Tick paralysis	*Dermacentor andersoni* *Dermacentor variabilis*
Tick-borne relapsing fever	*Ornithodoros* genus (soft tick)
Trypanosomiasis	*Tsetse* fly, hemipterids
Tularemia	*Amblyomma americanum* *Dermacentor andersoni* *Dermacentor variabilis* *Chrysops* deer flies Horseflies
Tyhpus	*Pediculus humanus* (lice)
Typhus, endemic	Fleas
Typhus, scrub	Chigger mites
Viral encephalitis	*Aedes* mosquitoes
West Nile fever	*Culex* mosquitoes
Yellow fever	*Aedes* mosquitoes

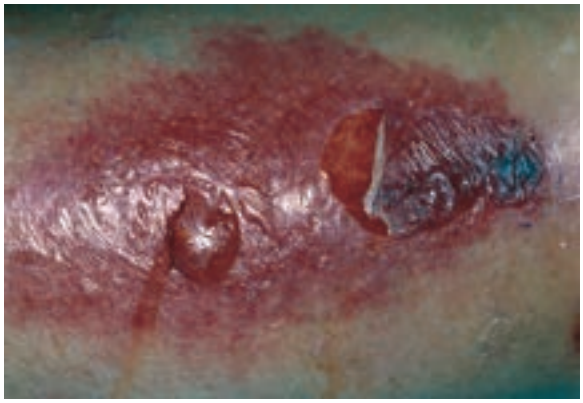

FIGURE 359-1. Exaggerated arthropod bite response in a patient with chronic lymphocytic leukemia.

TABLE 359-3 MEDICALLY IMPORTANT MITES

FAMILY	
Acaridae	Commonly found in grain, flour, and other foods
Cheyletidae	Walking dandruff in dogs, cats, and rabbits
Demodicidae	Human hair follicle mites
Dermanyssidae	Bird and rodent mites
Glycyphagidae	Associated with food, animals, and leaf litter
Hemoganasidae	Straw mites
Psoroptidae	Scab mites
Pyemotidae	Grain and straw mites
Sarcoptidae	Scabies mites
Trombiculidae	Chigger or harvest mites

TABLE 359-4 CAUSES OF DIRECT TOXIC INJURY BY ARTHROPODS AND LEECHES

ORGANISM	TYPE OF INJURY	DISTRIBUTION AND POPULATION AFFECTED
Australian funnel web spiders (*Hadronyche* and *Atrax*)	Neurotoxin producing tingling around the mouth, muscle twitching, nausea, vomiting, profuse sweating, coma, myocardial injury	Australians
Blister beetles	Bullous reactions	Widely distributed, but more common in hot climates
Brazilian wandering spider (*Phoneutria* spiders)	Neurotoxin resulting in paralysis	Brazilians
Brown *Loxosceles* spiders	Dermonecrotic reactions	Contact through attics and woodpiles
Centipedes	Neurotoxin that may produce pain, paresthesia, erythema, edema, and profuse bleeding. Rarely, coronary ischemia, rhabdomyolysis, proteinuria, and renal failure.	Most toxic species in Asia
Hymenopterids	Anaphylactic reactions	Wide distribution
Leeches	Bleeding	Fresh water exposure
Lepidopterids *Lonomia* caterpillars	Rash, ophthalmia nodosa, keratoconjunctivitis Fatal bleeding diathesis	Wide distribution Brazil
Millipedes	Chemical burns	Most toxic species in Australia and Africa
Pacific funnel web spider (*Tegenaria agrestis*)	Dermonecrotic reactions	Basements in the Pacific Northwest
Scorpions *Centruroides sculpturatus* Buthid scorpions	 Mild systemic symptoms, sometimes severe in children Severe cardiac toxicity	 Southwestern United States Asia, Africa, and Latin America
Tarantulas	Contact urticaria and ophthalmia nodosa	Hot climates
Widow spiders (*Latrodectus*)	Neurotoxin mimicking acute abdomen	Contact through outhouses and woodpiles

INFECTION BY ARACHNIDA

SCABIES

DEFINITION

Human scabies is caused by the mite *Sarcoptes scabiei* variant *hominis,* an obligate human pathogen, belonging to the class Arachnida, subclass Acari, order Astigmata, and family Sarcoptidae. Animal mange or scabies is caused by related *Sarcoptes* or *Psoroptes* mites. These zoonotic mites can produce transient symptoms in exposed humans, but are rarely capable of producing sustained infestation. *Sarcoptes scabiei* variant *hominis* is round to oval with a small anterior portion, dorsal spines, and hairlike projections on the rudimentary legs.

PATHOBIOLOGY

After mating on the skin surface, the female mite begins to burrow and lay eggs at a rate of 1 to 3 per day. The eggs mature and hatch over 3 to 4 days. The larvae pass through protonymph and tritonymph stages before molting into adults. The entire life cycle ranges between 30 and 60 days. All stages of the mite are capable of penetrating the stratum corneum via secretion of enzymes. In immunosuppressed individuals, mite populations may reach staggering proportions within thick, white, highly infectious crusts.

In immunocompetent patients, scabies mites induce a strong host response characterized by a superficial and deep perivascular inflammatory infiltrate composed of lymphocytes, histiocytes, and eosinophils. There is marked production of interleukin-6 (IL-6) resulting in proliferation of keratinocytes, activation of Th1 CD4+ cells causing production of IL-2, and activation of Th2 CD4+ cells causing production of IL-4. Disease manifestations relate to the complex balance of the immune response.

CLINICAL MANIFESTATIONS

In immunocompetent adults, the eruption manifests with severe pruritus with nocturnal exacerbation and a symmetrical erythematous, papulovesicular eruption. Pruritus can be present on both affected and unaffected skin. There is a predilection for involvement of the interdigital web spaces, wrists (Fig. 359-2), anterior axillary folds, periumbilical skin, the areolae in females, and the penis and scrotum in males. In the pediatric population, head, neck, face, palms, and soles are commonly involved. Nodular scabies is a clinical variant accounting for 7% of all cases. Those affected present with intensely pruritic red-brown nodules 2 to 20 mm in size on the genitalia, buttocks, groin, and axillary region. These nodules do not contain mites but represent hypersensitivity to mite products. Crusted scabies manifests in immunosuppressed individuals as silver to white crusts involving any portion of the body, but with a predilection for the hands and ears. In elderly patients, the manifestations may be quite subtle with mild pruritus and prurigo-like papules on the trunk and extremities. In patients with human T-lymphotropic virus type I infection, scabies often manifests with linear, psoriasiform, hyperkeratotic lesions on the dorsal hands and feet. Crusted scabies can result in sepsis, with a 5-year mortality rate of up to 50%.

The most pathognomonic findings are the presence of burrows and genital nodules. The burrow is a short serpiginous gray to white keratotic line measuring 1 to 4 mm in length. They are most common on the hands and feet, particularly on the finger web spaces, the thenar and hypothenar eminences, and the wrists. Genital nodules may involve the glans, scrotum, or labia majora. A history of itch in other family members or sexual partners is highly suggestive of the diagnosis.

DIAGNOSIS

The diagnosis is established by identification of the female mite, ova, or feces in a scraping from a burrow (Fig. 359-3). Definitive diagnosis relies on direct visualization of mites, eggs, or mite pellets under microscopy. Additional diagnostic methods include the burrow ink test, video-dermoscopy, and polymerase chain reaction (PCR) and enzyme-linked immunosorbent assay (ELISA) for mite product or specific immunoglobulin E (IgE).

TREATMENT

Rx

Commonly employed topical agents include permethrin 5% and precipitated sulfur 2 to 10% in petrolatum. Topical agents used worldwide include benzyl benzoate 10 to 25%, monosulfiram 5 to 25%, malathion 0.5%, and esdepallethrin 0.63%. The topical agent crotamiton 10% has relatively weak antiscabetic activity but has antipruritic as well as antibacterial actions. An alternative to topical therapy is the use of oral ivermectin, 200 to 400 µg/kg given as a single dose, then repeated in 10 days.[2]

Proper application of any topical must include the umbilicus, genitalia, under the nails, and the skin up to the edge of all body orifices. Fomites are important in the spread of crusted scabies, with hospital linens being of particular concern. Caregivers involved in bathing or lifting, sexual partners, and family members are at high risk for infection and should be treated at the same time as the patient regardless of symptomatology. Clothes and bed linens should be machine washed at 60°C followed by heated drying.

OTHER MITES

Mites are ubiquitous in the environment and are a frequent cause of dermatitis. The most common medically important mites are listed in Table 359-3.

The majority of mites are free-living, but thousands of species are obligate parasites of animals or plants. Like ticks, mites mature through various life stages. Typically, there is a single six-legged larval stage, followed by several nymphal stages and an adult stage. Both nymphs and adults typically have eight legs. Although most mites lay eggs, the spiny rat mite (*Laelaps echidninus*) is viviparous, and the straw itch mite (*Pyemotes tritici*) protects its young internally throughout their life cycle until they emerge as sexually mature adults. Manifestations of mite infestations are protean and include papular, vesicular, urticarial, and morbilliform eruptions. Chigger mites most often affect the lower legs, the edges of underwear, and the genitalia. Summer penile syndrome is a manifestation of chigger mite infestation. It presents with tender to itchy papules of the glans and scrotum and may mimic scabies infestation. Chigger (Trombiculid) mites in Asia are vectors for scrub typhus,

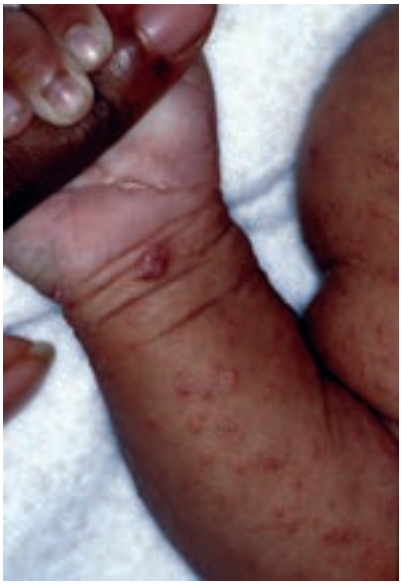

FIGURE 359-2. Scabies infestation in children commonly manifests with burrows (shown here between the thenar and hypothenar eminences), papulovesicles, and crusts involving the wrists and flexures.

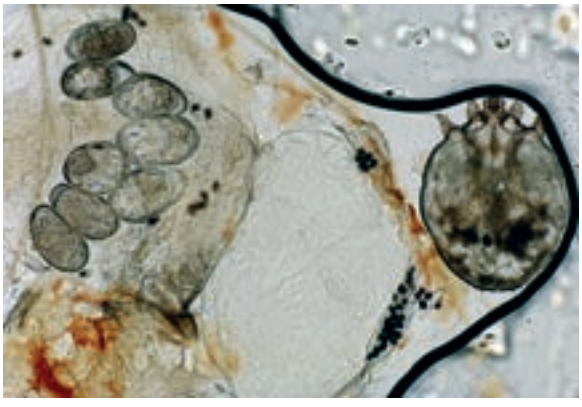

FIGURE 359-3. Scabies preparation demonstrating an adult mite, ova, and feces.

whereas house mouse mites in New York City transmit rickettsial pox. In endemic areas, the presence of an eschar in a patient with acute febrile illness has a high diagnostic value for rickettsial infection.

Most mite-induced rashes represent a "bite and run" injury and can be treated with topical corticosteroids or camphor and menthol lotion for symptomatic relief.

TICKS

Ticks are important disease vectors, and *Dermacentor* ticks can cause fatal tick paralysis in children. It is caused by a neurotoxin secreted by the salivary glands of an engorged female tick. Tick paralysis manifests as the development of unsteady gait, followed by progressive, ascending flaccid paralysis. As the tick attaches to the scalp, often behind the ear, it frequently goes unnoticed, and death occurs in over 10% of affected children due to respiratory failure. Nodular lesions are commonly pseudolymphomatous histologically.

Rickettsial illnesses typically manifest with fever and a headache, and treatment with a tetracycline should never be withheld because of the absence of a rash. Any delay in the initiation of antibiotic therapy may prove fatal, and empirical treatment should never be delayed until confirmatory tests are available. Doxycycline is generally recommended in both adults and children.

Rickettsial diseases, including Rocky Mountain spotted fever, continue to emerge in much of the world, especially in South America. Most of these agents produce milder syndromes similar to Rocky Mountain spotted fever.

Exclusion of animal hosts from recreational areas is key to control of tick-borne illness.[3] Oral agents, including avermectin-laced feed corn can be used to kill ticks that feed on deer. Removal of leaf debris leads to a reduction in tick numbers via dehydration of adults and ova. Permethrin, a pyrethroid insecticide with neurotoxic activity, is widely marketed as a topical acaricide to be applied to clothing. It is stable through several wash cycles. Some North African *Hyalomma* ticks have demonstrated high-level resistance, and permethrin may produce a pheromone-like attachment response in these ticks. Resistance in North American ticks can be esterase-based or related to sodium channel gene mutations, and permethrin resistance has been linked to outbreaks of bovine babesiosis and anaplasmosis.

A veterinarian should be consulted concerning control of ticks in pets and livestock.

Antibiotic prophylaxis after tick bites is rarely justified; it is indicated mostly in highly endemic areas where the ticks are heavily engorged. When signs and symptoms appear, a 10-day course of antibiotic is typically sufficient. A controlled trial comparing 10 days of oral doxycycline (with or without intravenous ceftriaxone) with 20 days of oral doxycycline for the treatment of early Lyme disease found a similar response rate in all three treatment groups.

SPIDERS

Toxic spiders are found worldwide.[4] The most feared include the funnel web spiders of Australia and *Phoneutria* wandering spiders in Brazil, which can be fatal. In North America, brown recluse spiders can produce a severe dermonecrotic reaction, but it should be noted that most spider bites result only in a mild local cutaneous reaction.

Black and brown widow spiders have a worldwide distribution. *Latrodectus mactans* is the most widely distributed black widow spider in North America. *Latrodectus curacaviensis* is common in South America and *Latrodectus tredecimguttatus* in Europe. Both black (*Latrodectus indistinctus*) and brown (*Latrodectus geometricus*) widow spiders are found in Africa and Madagascar, whereas Australia and New Zealand have red-black spiders (*L. mactans hasselti*).

Widow spiders are commonly encountered in woodpiles, and bites are typically defensive, frequently involving the hand after the spider has been disturbed. Latrotoxins act by increasing intracellular calcium concentrations, depolarizing neurons, and stimulating uncontrolled release of neurotransmitters, particularly norepinephrine and acetylcholine. Divalent cation-dependent tetramers of α-latrotoxin produce membrane pores in mammalian cells. Only the females have the ability to deliver venom and can be identified by the red hourglass pattern and shiny, black abdomen. Bites cause an envenomation syndrome comprising localized pain, diaphoresis, hypertension, weakness, muscle cramping, and severe pain in areas such as the abdomen and back. The syndrome may mimic an acute surgical abdomen. Both benzodiazepines and intravenous calcium gluconate have been used to treat associated tetany, but antivenin produces more rapid relief and is indicated for unresponsive tetany and priapism. Purified F(ab)2 fragment antivenin is now available and associated with a low risk for adverse reactions.

FIGURE 359-4. Brown recluse spiders are characterized by a dorsal violin case pattern.

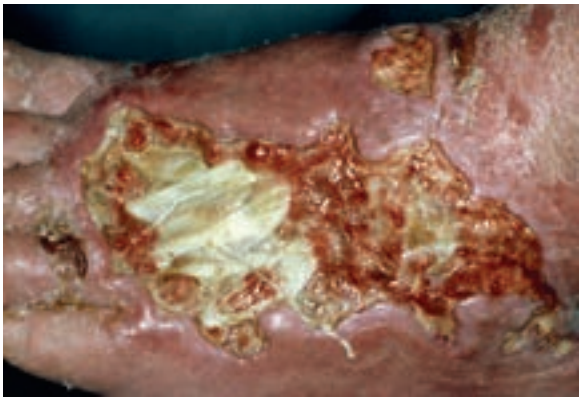

FIGURE 359-5. Dermonecrotic reactions from brown recluse spiders can be severe, but most are mild conditions. Furuncles or pyoderma gangrenosum are common misdiagnoses in patients who actually have brown recluse spider bites. (Courtesy Larry Becker, MD.)

Brown *Loxosceles* spiders are found throughout the world, with the most famous being the brown recluse spider, *Loxosceles recluse* (Fig. 359-4).[5] The resulting necrosis can be severe (Fig. 359-5), but most bites do not result in skin necrosis, and those of related spiders, including *Loxosceles rufescens, Loxosceles deserta,* and *Loxosceles arizonica,* are even milder. Brown recluse spiders are endemic to southeastern and midwestern United States. They are quite shy and bites are very uncommon; however, envenomation can cause severe dermatonecrosis and hemolytic anemia. Most skin reactions diagnosed as "brown recluse bites" represent furuncles or pyoderma gangrenosum. The diagnosis can be confirmed by an enzyme immunoassay, glycophorin A measurement, or a passive hemagglutination inhibition test.

Sphingomyelinase D is the major toxin in brown recluse venom, and hyaluronidase allows the eschars to spread in a gravity-dependent fashion. Systemic reactions include disseminated intravascular coagulation and Coombs'-positive hemolytic anemia, but these reactions are rare as well. Most bites can be treated supportively with rest, ice, and elevation. Intradermal injection of polyclonal anti-*Loxosceles* Fab fragments or antivenin may reduce the ultimate size of the necrotic area, but these agents are not readily available. Studies with dapsone, colchicine, hyperbaric oxygen, and prednisone have generally been disappointing. Dapsone produces hemolysis, especially in individuals who have inherited G6PD deficiency (Chapter 161). Based on limited animal data and anecdotal human data, intralesional triamcinolone appears to be appropriate for patients with rapidly expanding areas of necrosis. The complement inhibitor eculizumab has demonstrated the ability to prevent venom-induced hemolysis in vitro.

Hadronyche and *Atrax* funnel-web spiders are endemic to a region in eastern Australia. Their venom contains small peptide neurotoxins, including the δ-atracotoxins that slow tetrodotoxin-sensitive voltage-gated sodium channel inactivation. Catecholaminergic and cholinergic excess can result in bradycardia or tachycardia, hypertension, shock, hypersalivation, and diaphoresis. Paresthesia, fasciculations, and muscle spasms may occur. Severe myocardial injury also has been reported.

Spiders belonging to the genus *Phoneutria* are found in South America, Costa Rica, and, most notably and clinically important, Brazil. Bites cause localized pain of varying severity that radiates proximally from the bite. Ninety percent of cases show at least mild systemic envenomation. Patients are symptomatic with nausea, vomiting, dizziness, tachycardia, hypertension, diaphoresis, visual changes, and priapism. Severe envenomation results in pulmonary edema and shock. Antivenin therapy is available.

Tegenaria agrestis is found in basements in the Pacific Northwest. In Europe, it exists as a rural spider, because *Tegenaria gigantea* and *Tegenaria domestica* compete for home habitats. Most reactions are mild, although some dermonecrotic reactions have been attributed to these spiders. Tarantulas can cause contact urticaria and ophthalmia nodosa related to hairs that become embedded in the cornea.

SCORPIONS

Local and systemic symptoms of scorpion stings are typically out of proportion to cutaneous signs. *Centruroides sculpturatus* produces somewhat more severe envenomation in the American Southwest, but the most dangerous Buthid scorpions are not found in the United States.[6] Most fatalities are caused by cardiorespiratory manifestations such as cardiogenic shock and pulmonary edema after envenomation and involve children younger than 10 years of age. Pancreatitis is an important cause of morbidity after scorpion envenomation. Toxic Buthid scorpions occur in Asia, Africa, and Latin America. Prazosin reverses the autonomic storm characteristic of the Indian red scorpion (*Mesobuthus tamulus*), and accelerated recovery with preserved myocardial function has been demonstrated when it is used in combination with antivenin.

Infection via Pentastomida, Chilopoda, Diplopoda, and Crustacea

TONGUE WORM (PENTASTOMIASIS)

Pentastomiasis is a zoonotic infestation that sometimes affects humans. The pentastomids *Armillifer* and *Linguatula* are wormlike arthropods endemic to Asia and Africa. They commonly infest the respiratory tracts of birds, reptiles, and small mammals. In humans, severe inflammation associated with the infestation can result in coughing, hemoptysis, lacrimation, coryza, and facial edema. Rarely, asphyxiation has been reported. This syndrome is known as *halzoun* or *marrara,* meaning suffocation.

CENTIPEDES AND MILLIPEDES

Centipedes have powerful jaws that inject a neurotoxic venom. Centipede bites may produce pain, paresthesia, erythema, edema, and profuse bleeding. Rarely, bites have been associated with coronary ischemia, rhabdomyolysis, proteinuria, and renal failure. Millipedes do not bite, but cause arcuate chemical burns via a caustic substance that they secrete when threatened. They may be found in line-dried clothing and the resulting burns may mimic signs of child abuse.

CRUSTACEANS AS A SOURCE OF INFECTION

Copepods are tiny aquatic crustaceans that serve as the intermediate hosts for nematodes including the guinea worm *Dracunculus medinensis* and the sushi worm *Gnathostoma spinigerum.* They may also transmit the cestodes *Spirometra mansonoides* and *Diphyllobothrium latum.*

Pediculosis

Pediculosis is the result of infestation by sucking lice of the phylum Arthropoda, class Insecta, order Phthiraptera, suborder Anoplura, family Pediculidae, or family Pthiridae. Lice are obligate ectoparasites, and three types infest humans: *Pediculus humanus capitis* (the head louse), *Pediculus humanus* (the body louse), and *Pthirus pubis* (the crab louse).

HEAD LICE

EPIDEMIOLOGY

Head lice affect 6 to 12 million people per year in the United States alone. Worldwide, much of the human population is affected, with highest prevalence among children between the ages of 3 and 12. Transmission occurs by means of direct contact or close contact of hats and scarves.

DIAGNOSIS

Head louse infestation should be suspected in any child with scalp pruritus and cervical adenopathy. Visual inspection can demonstrate nits, especially in the retroauricular area. Small bites and crusts of blood are common on the scalp. Definitive diagnosis involves finding at least one live louse on visual inspection of the scalp.

TREATMENT Rx

Current recommendations favor agents that kill lice by occlusion of the respiratory spiracles. One such agent (Ulesfia lotion) uses benzyl alcohol to stun the spiracle open and allow occlusion. For routine infection, a single, 10-minute, at-home application of ivermectin 0.5% lotion is 95% effective at 1 day, 85% effective at 7 days, and 74% effective at 14 days after treatment for eliminating head-louse infestations.■ For difficult-to-treat head lice infestation in children, oral ivermectin (400 µg/kg), given twice at a 7-day interval, had superior efficacy compared with topical 0.5% malathion lotion, suggesting that it could be an alternative treatment.■

Malathion is a prescription organophosphate that acts by interfering with acetylcholinesterase. In the United States, it is marketed in a flammable vehicle that contributes substantially to its efficacy. There is relatively little resistance to this agent, but the trend is away from the use of neurotoxins in children.

Lindane is an organochloride compound that is still available by prescription. Central nervous system toxicity appears rare but together with environmental concerns resulted in banning of the product in California.

BODY LICE

EPIDEMIOLOGY

Body louse infestation is common in refugee situations, during war, and among the homeless in urban areas. This is a major public health concern because the louse can serve as a vector for at least three intracellular disease-causing bacteria: *Bartonella quintana* (causing trench fever and endocarditis), *Rickettsia prowazekii* (causing epidemic typhus), *and Borrelia recurrentis* (causing louse-borne relapsing fever).

CLINICAL MANIFESTATIONS

Early infestation produces a widespread dermatitis that may mimic atopic dermatitis, allergic contact dermatitis, or a viral exanthem. In patients with widespread excoriations, the most likely differential diagnosis is scabies. The diagnosis is made by finding bluish hued maculae cerulea (Fig. 359-6), prompting a search for body lice or nits in the seams of clothing.

TREATMENT Rx

Washing and hot drying or ironing of clothing will kill lice. Permethrin has also been used on clothing; however, a pediculocide is not always necessary.

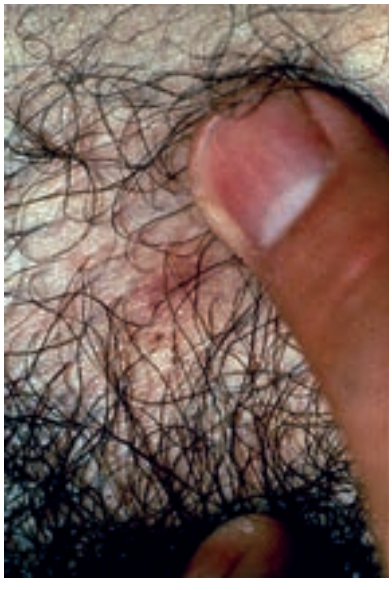

FIGURE 359-6. Maculae cerulae are bluish red spots that represent sites where body lice have fed.

CRAB LICE

PATHOBIOLOGY

Crab louse infestation is spread as a sexually transmitted disease. Roughly 30% of patients will have another concurrent sexually transmitted disease and adolescents with pubic lice are twice as likely as uninfested adolescents to have chlamydial or gonorrheal infection.

CLINICAL MANIFESTATIONS

Patients present with intense itching, and both lice and nits may be found on pubic hairs as well as hair on the chest, abdomen, and legs. Pubic lice occasionally may cause scalp infestation in patients with thick curly hair. They also may infest the eyelashes, causing phthiriasis palpebrarum infection. When observed in children, sexual exposure should be considered.

TREATMENT Rx

Treatment is by means of shaving or with topical agents used to treat head lice. Oral ivermectin has also been used at a dose of 200 to 400 µg/kg, repeated in 10 days.

Beetle-Related Dermatitis

Blister beetles (order Coleoptera) are widespread throughout the world and include members of the families Meloidae and Staphylinidae. Members of the family Oedemeridae are classified as "false" blister beetles, but can produce similar skin manifestations. Cantharidin accumulates during the larval stages of the beetle and gives it protection from predatory birds. When threatened, the beetles protect themselves by exuding a cantharidin-containing liquid. Contact with blister beetles results in vesicles and bullae. "Nairobi eye," or rove beetle dermatitis, in Northern Kenya is caused by *Paederus eximius*. Rove beetles were implicated in a recent epidemic of pustular eruption among U.S. forces in Pakistan. Treatment for all manifestations of blister beetles is largely symptomatic.

FLEAS

Fleas (order Siphonaptera) are ubiquitous among pets and wild animals. They serve as vectors for endemic typhus, bubonic plague, brucellosis, melioidosis and erysipeloid. The most common flea on dogs is the cat flea, *Ctenocephalides felis*. *Pulex irritans,* the human flea, can be found on dogs and became a major plague vector during the great epidemics. Flea bites cluster on the lower legs and present as intensely pruritic papulovesicles. Tungiasis is caused by the jigger or sand flea, *Tunga penetrans.* It typically presents with a necrotic pustule adjacent to the great toenail.

A veterinarian should be consulted about agents for flea control, including lufenuron and fipronil. Beach areas should restrict pets to prevent infestation.

FLIES

Mosquitoes and other flies belong to the order Diptera and represent major disease vectors. *Anopheles* mosquitoes bite mostly at night and are the major vectors of malaria. They also transmit filariasis. *Aedes* mosquitoes bite during the day and transmit dengue, viral encephalitis, yellow fever, and filariasis. *Culex* mosquitoes bite at night and transmit West Nile encephalitis, filariasis, and viral encephalitis. Cul*ex* mosquitoes also bite at dusk or early evening. Many arthropod diseases have a principal vector but also can be spread by other vectors. For example, although *Culex* mosquitoes are the major vector for West Nile virus in the United States, *Aedes* and *Anopheles* spp also can spread the disease. *Simulium* flies (black flies, humpback flies, buffalo gnats, turkey gnats) transmit onchocerciasis. *Lutzomyia* and *Phlebotomus* sandflies transmit leishmaniasis, *Bartonella bacilliformis,* and sandfly fever. Biting midges (punkies, gnats, no-see-ums) are small enough to pass through screens and cause papulovesicular and nodular reactions in sensitive individuals. Tabanid flies include the horse flies, deer flies, and greenheads. As a group, they are large and colorful and deliver painful bites. *Chrysops* deer flies transmit loiasis in Africa and tularemia in the United States. *Glossina* tsetse flies transmit trypanosomiasis.

Furuncular myiasis caused by the botfly is common in Latin America. *Dermatobia hominis,* glues its eggs to a mosquito, which serves as an intermediate vector. *Cordylobia anthropophaga,* the mango fly, tumbu fly, putzi fly, or skin maggot fly, is native to Africa and lays its eggs on drying clothing. It produces plaque-type myiasis with many maggots. *Cochliomyia hominivorax,* the screwworm, lays its eggs on wounds or mucous membranes. Screwworms are feared because they leave necrotic flesh and continue to travel through viable tissue. Sarcophagid flies deposit their living larvae on the hosts, where they produce a type of larva migrans. Myiasis is treated by means of mechanical removal with tweezers or by incision of the wound. Covering the embedded larvae with an occlusive substance such as surgical lubricant, petrolatum, or bacon encourages them to move upward facilitating removal.

Prevention of vector-borne disease requires control of standing water, use of screening and mosquito netting, and personal protection with repellents. Screening impregnated with pyrethroids is more effective than untreated screening. Secondary prevention includes malaria chemoprophylaxis and early treatment of illness. Other interventions include insecticidal sprays, and gas-powered mosquito traps. Most types of mosquito traps generate carbon dioxide and use chemical attractants such as octenol and butanone. Some *Culex* mosquitoes are repelled by octenol.

N,N-diethyl-3-methylbenzamide (DEET) remains the most widely used insect repellent, although picaridin is widely used in Europe and is gaining market share in the United States. IR3535, KBR 3023, and *para*-menthane diol are also used as repellents for application to skin. Permethrin is largely used for fabric impregnation. DEET has a long safety record, but occasionally can cause bullous dermatitis, anaphylaxis, or toxic encephalopathy. The American Academy of Pediatrics recommends slow-release products below 30% concentration. Neem oil products perform reasonably well against various mosquitoes, but citronella has limited efficacy.

HEMIPTERA: TRUE BUGS

Bedbugs have become pandemic and are increasingly recognized in suburban homes, hotels, and, especially, on college campuses.[7] They are red-brown and about the size of small ticks (Fig. 359-7). The wings are vestigial. *Cimex lectularius* and *C. hemipterus* (the tropical bedbug) parasitize birds and bats, who introduce them into buildings. They also spread readily via luggage. These organisms hide in cracks and crevices of beds, particularly along the seams of mattresses. Hepatitis B has been found in bedbugs, but they have not shown vector competence to transmit human disease. Bites are characteristically found on the face, neck, and arms. In covered areas, a string of bites termed "breakfast, lunch, and dinner" is typical (Fig. 359-8). Treatment of bites is with topical corticosteroids. Disinfestation involves extermination, washing and drying of all bedding on a hot setting, and placing the mattress and box spring in a zippered plastic case. Room hyperthermia has proved to be a viable alternative.

Triatome reduviids are important vectors for American trypanosomiasis (Chagas disease). Their wings overlap and are sclerotic proximally. The striped abdomen is visible lateral to the wings. Unilateral eyelid swelling (Romana sign) is caused by a conjunctival reaction to the bite and to triatome feces infected with trypanosomes.

HYMENOPTERA STINGS

Hymenoptera include membranous winged insects such as wasps, bees, and fire ants. They all result in painful stings and present a potential for anaphylactic reactions.[8] Unexpected stings occur frequently; therefore, allergic individuals must always carry a source of epinephrine such as an EpiPen auto-injector. Bees are eviscerated during the sting and the venom gland

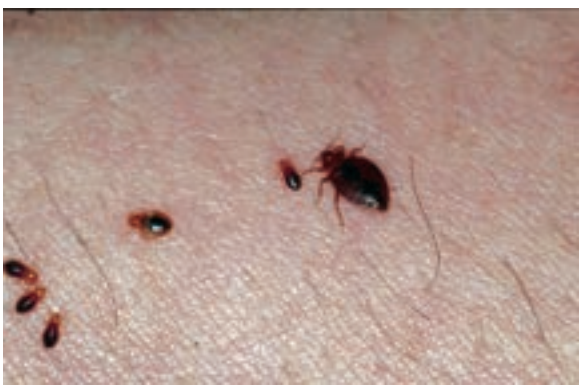

FIGURE 359-7. Bedbugs are red-brown and about the size of small ticks.

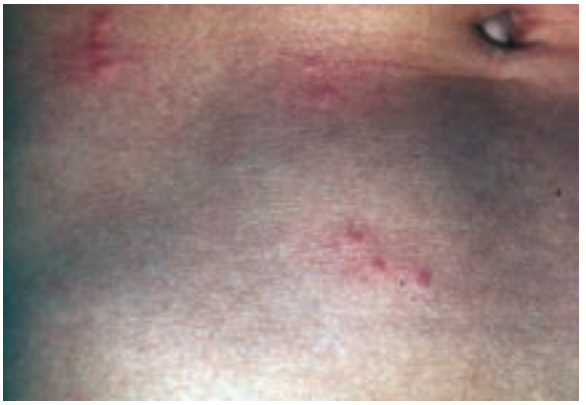

FIGURE 359-8. Bedbug bites are often arranged in a characteristic "breakfast, lunch, and dinner" pattern.

continues to inject additional venom. Therefore, bee stingers should be removed as rapidly as possible, with little concern for the exact method of removal. Venom immunotherapy can improve quality of life. Rush immunotherapy (a method for rapidly desensitizing patients to allergens) can be performed in those with severe reactions. Omalizumab has been used in conjunction with immunotherapy and has shown clinical effectiveness but further investigation is warranted to determine its exact role.

LEPIDOPTERIDS

Moths and caterpillars may have toxic hairs that may contain histamine, kinins, plasminogen activators, and other proteinaceous toxins. The reactions are usually toxic rather than allergic in effects. Clinical manifestations include pain, erythema, swelling, and hemorrhage. Systemic reactions may include renal failure. *Lonomia* caterpillars may cause a fatal bleeding diathesis with intracranial hemorrhage. Ocular reactions to hairs include ophthalmia nodosa, keratoconjunctivitis, subconjunctival nodules, iritis, and vitreoretinal involvement. Children who ingest caterpillars may develop edema of the lips, tongue, and buccal mucosa. Esophageal and tracheobronchial involvement occurs less commonly.

DELUSIONS OF PARASITOSIS

Patients with delusions of parasitosis present with an unshakable belief that they are infested with pathogens such as insects, worms, bacteria, parasites, or fungi.[9] Some believe they are infested with inanimate materials such as hairs, filaments, fibers, and particles. It is difficult to help them. A careful evaluation should exclude formications secondary to cocaine addiction and pruritus related to systemic diseases such as renal failure or hyperthyroidism. Although delusions of parasitosis have traditionally been treated with pimozide, newer atypical antipsychotic agents such as risperidone, olanzapine, quetiapine, ziprasidone, and aripiprazole reduce the need for electrocardiographic monitoring.

LEECHES AND OTHER ANNELIDS

Leeches occur in fresh water and may be associated with *Pseudomonas* or *Aeromonas* wound infections. Leeches contain anticoagulants, and the attachment sites may bleed freely when the leech is removed. More serious coagulation disorders with massive bleeding have been reported. Saturated salt solutions have been used to facilitate leech removal, but do not reduce the risk for infection.

Grade A References

A1. Pariser DM, Meinking TL, Bell M, et al. Topical 0.5% ivermectin lotion for treatment of head lice. *N Engl J Med*. 2012;367:1687-1693.

A2. Chosidow O, Giraudeau B, Cottrell J, et al. Oral ivermectin versus malathion lotion for difficult-to-treat head lice. *N Engl J Med*. 2010;362:896-905.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Golden DB. Advances in diagnosis and management of insect sting allergy. *Ann Allergy Asthma Immunol*. 2013;111:84-89.
2. Mounsey KE, McCarthy JS. Treatment and control of scabies. *Curr Opin Infect Dis*. 2013;26:133-139.
3. Pages F, Dautel H, Duvallet G, et al. Tick repellents for human use: prevention of tick bites and tick-borne diseases. *Vector Borne Zoonotic Dis*. 2014;14:85-93.
4. Kularatne SA, Senanayake N. Venomous snake bites, scorpions, and spiders. *Handb Clin Neurol*. 2014;120:987-1001.
5. Payne KS, Schilli K, Meier K, et al. Extreme pain from brown recluse spider bites: model for cytokine-driven pain. *JAMA Dermatol*. 2014;150:1205-1208.
6. Isbister GK, Bawaskar HS. Scorpion envenomation. *N Engl J Med*. 2014;371:457-463.
7. Bernardeschi C, Le Cleach L, Delaunay P, et al. Bed bug infestation. *BMJ*. 2013;346:f138.
8. Casale TB, Burks AW. Clinical practice: hymenoptera-sting hypersensitivity. *N Engl J Med*. 2014;370:1432-1439.
9. Levin EC, Gieler U. Delusions of parasitosis. *Semin Cutan Med Surg*. 2013;32:73-77.

REVIEW QUESTIONS

1. Which of the following is *not* a mechanism by which arthropod venoms cause human injury?

A. Neurotoxins
B. Extracellular matrix enzymes
C. Dermonecrotic reactions
d. Chemical burns
E. Anticoagulants

Answer: B Whereas snake venoms typically mediate tissue necrosis and extracellular matrix degradation via enzymes such as metalloproteases, hyaluronidase, and myotoxic phospholipase A_2, arthropods are not known to cause injury through these mechanisms. Causes of toxic injury by arthropods (see Table 359-4) include neurotoxins (e.g., Australian funnel web spiders, Brazilian wandering spiders, centipedes, widow spiders); dermonecrotoxins (e.g., brown *Loxosceles* spiders, Pacific funnel web spiders); chemical burns (e.g., millipedes); and anticoagulants (centipedes, leeches, *Lonomia* caterpillars).

2. A 20-year-old college student presents to her student health service clinic with an intensely pruritic papulovesicular eruption of both lower legs. She and her three roommates had recently adopted a stray dog despite their crowded and cluttered apartment. Which of the following should be recommended for management?

A. Treat with topical agents and machine wash clothes and bed linens at 60° C followed by heated drying
B. Evaluate her for underlying inflammatory bowel disease
C. Consult a veterinarian about agents for flea control
D. Apply topical steroids and exterminate or wash and dry all bedding on a hot setting
E. Search for offending agents in the seams of clothing and burn such clothes

Answer: C The patient most likely has flea bites from the stray dog. As in this case, flea bites tend to cluster on the lower legs and manifest as intensely pruritic papulovesicles. A veterinarian should be consulted about agents for flea control, including lufenuron and fipronil. The management in option A applies to scabies. The pruritic lesions of scabies have a predilection for involvement of the interdigital web spaces, wrists, anterior axillary folds, periumbilical skin, areolae in females, and penis and scrotum in males. Option B would be the approach to a patient with classic pyoderma gangrenosum, which usually occurs in the lower extremities. Pyoderma gangrenosum can be confused with brown recluse spider bites. Option E applies to the management of body lice; infestation with body lice produces a widespread dermatitis with widespread excoriations (that can be similar to those seen in scabies) and the diagnostic finding of bluish maculae cerulae (see Fig. 359-6). Bed bugs have become endemic and are increasingly recognized on college campuses, among other sites. Their bites are characteristically found on the face, neck, and arms, typically in strings of bites termed "breakfast, lunch, and dinner." Option D would apply to bed bugs.

3. Which of the following pairings of vector and disease it causes is correct?

A. *Aedes* mosquitoes and malaria
B. Fleas and yellow fever
C. Sandflies and bartonellosis
D. Wood tick and rickettsial pox
E. Tsetse flies and African trypanosomiasis

Answer: E Tsetse flies are vectors for African trypanosomiasis. Triatome bugs are vectors for American trypanosomiasis. *Aedes* mosquitoes are vectors for dengue and yellow fever; anopheline mosquitoes are the vectors for malaria. Fleas are vectors for bartonellosis and endemic typhus; *Aedes* mosquitoes are the vectors for yellow fever. Sandflies are vectors for leishmaniasis, not bartonellosis. Wood ticks are vectors for Rocky Mountain spotted fever (along with dog ticks and *Amblyomma americanum*. See Table 359-2.

360

ANTIVIRAL THERAPY (NON-HIV)

JOHN H. BEIGEL

Although some viral infections are self-limited, others can cause significant morbidity and mortality. Effective therapy is available for many of these infections. This chapter reviews currently available antiviral agents for the treatment of infections caused by viruses other than human immunodeficiency virus (HIV). Not all agents discussed are licensed in all countries.

Currently available agents can be classified into those that directly inhibit viral replication at the cellular level (antivirals), those that modify the host response to infection (immunomodulators), and those that directly inactivate viral particles (microbicides/virucides). Antiviral agents can be classified based on their mechanism of action. For example, nucleic acid analogues inhibit viral DNA or RNA synthesis by competing with endogenous nucleic acids and block the viral DNA polymerase or RNA transcriptases. By comparison, protease inhibitors prevent viral replication by binding to the enzymes that cleave viral protein precursors into active proteins.

Antiviral strategies that are not covered in this chapter include local destructive measures that destroy both host tissues and virus simultaneously, such as cryotherapy, laser, and podophyllin treatment of warts. Although effective, such measures are useful only for discrete or localized mucocutaneous infections.

ANTIVIRALS FOR HEPATITIS B VIRUS INFECTIONS

Acute hepatitis B infection (Chapter 148) generally does not require antiviral treatment. Currently approved antivirals for chronic hepatitis B (Chapter 149) include five nucleic acid analogues (adefovir, entecavir, lamivudine, telbivudine, and tenofovir) and two immune modulators (interferon alfa-2b and pegylated [PEG]-interferon alfa-2a) (Tables 360-1 to 360-3). Treatment may be initiated with any approved antiviral medications, but tenofovir, entecavir, and PEG-interferon alfa-2a are generally the preferred agents.[1] Tenofovir or entecavir is preferred for patients with compensated cirrhosis (Chapter 153).

TABLE 360-1 ANTIVIRALS FOR HEPATITIS VIRUS INFECTIONS

VIRAL INFECTION	DRUG	ROUTE	USUAL ADULT DOSAGE
Chronic hepatitis B	Tenofovir	PO	300 mg/day
	Entecavir		
	Naïve virus	PO	0.5 mg daily; optimal duration of therapy unknown
	Lamivudine-resistant virus	PO	1 mg daily; optimal duration of therapy unknown
	Interferon alfa-2b	SC	6 MU/m^2 (up to 10 MU) 3 times weekly for 16-24 wk
	PEG-interferon alfa-2a	SC	180 µg weekly for 48 weeks
	Adefovir	PO	10 mg/day
	Lamivudine	PO	100 mg/day
	Telbivudine	PO	600 mg/day
Chronic hepatitis C	Ledipasvir/Sofosbuvir	PO	90 mg/400 mg once daily for 12-24 weeks
	Ombitasvir/ Paritaprevir/ Ritonavir plus Dasabuvir	PO	12.5 mg/75 mg/50 mg once daily plus dasabuvir 250 mg twice daily
	Sofosbuvir	PO	400 mg once daily for 12-24 weeks
	Simeprevir	PO	150 mg once daily for 12 weeks
	Boceprevir	PO	800 mg every 8 hours for 24-44 weeks
	PEG-interferon alfa-2a	SC	180 µg weekly for 48 weeks
	or PEG-interferon alfa-2b	SC	1.5 µg/kg weekly for 48 weeks
	plus Ribavirin	PO	800-1200 mg/day, depending on weight

TABLE 360-2 MECHANISMS OF EXCRETION AND THRESHOLDS FOR DOSE ADJUSTMENT

	MAJOR ROUTE OF ELIMINATION	THRESHOLD FOR ADJUSTMENT IN RENAL INSUFFICIENCY OR FAILURE	ADJUSTMENT FOR HEPATIC FAILURE	SPECIAL ADJUSTMENT FOR ELDERLY
Adefovir	Renal	CrCl < 50 mL/min	No adjustment	
Entecavir	Renal	CrCl < 50 mL/min	No adjustment	
Lamivudine	Renal	CrCl < 50 mL/min	No adjustment	
Ledipasvir/Sofosbuvir	Renal	CrCl < 20 mL/min	No adjustment	
Ombitasvir/Paritaprevir/ Ritonavir plus Dasabuvir	Renal	No adjustment	No adjustment	
Ribavirin	Renal	CrCl < 50 mL/min	No adjustment	
PEG-interferon alfa-2a	Renal	CrCl < 50 mL/min	Progressive rise in alanine transaminase	>60 years, consider reduction
Telbivudine	Renal	CrCl < 50 mL/min	No adjustment	
Sofosbuvir	Renal	CrCl < 50 mL/min	No adjustment	
Simeprevir	Hepatic	No adjustment	No adjustment	
Boceprevir	Hepatic	No adjustment	No adjustment	

TABLE 360-3 SIGNIFICANT ADVERSE EFFECTS (U.S. FDA BLACK BOX WARNING)

DRUG	BLACK BOX SYNOPSIS
Adefovir	Severe acute exacerbations of hepatitis B may occur with cessation of therapy. Nephrotoxicity may occur in patients at risk for or undergoing renal dysfunction. Lactic acidosis and severe hepatomegaly with steatosis
Entecavir	Lactic acidosis and severe hepatomegaly with steatosis; severe acute exacerbations of hepatitis B may occur with cessation of therapy.
Lamivudine	Severe acute exacerbations of hepatitis B may occur with cessation of therapy. Lactic acidosis and severe hepatomegaly with steatosis
Ribavirin	Monotherapy for hepatitis C is not effective. Hemolytic anemia Teratogenic and embryocidal
Interferon alfa	May cause or aggravate neuropsychiatric, autoimmune, ischemic, and infectious disorders
Telbivudine	Severe acute exacerbation of hepatitis B may occur with cessation of therapy; lactic acidosis and severe hepatomegaly with steatosis.

Patients with chronic hepatitis B (hepatitis B surface antigen [HBsAg] positive for > 6 months, detectable serum hepatitis B virus (HBV) DNA > 20,000 IU/mL, and an alanine aminotransferase [ALT] level > twice the normal level) should be evaluated for treatment. Patients with clinically decompensated hepatitis B (e.g., icterus or other signs) generally require antiviral treatment. Therapy in hepatitis B e antigen (HBeAg)-positive chronic hepatitis B should be continued until the patient has achieved HBeAg seroconversion and serum HBV DNA is undetectable, followed by at least 6 months of additional treatment after the appearance of anti-HBe. Therapy in HBeAg-negative chronic hepatitis B should continue for at least a year. Patients with decompensated cirrhosis or recurrent hepatitis B after liver transplantation should receive lifelong treatment.

Tenofovir

Tenofovir, which is a nucleotide analogue of adenosine monophosphate, was first approved for the treatment of HIV infection. The commercially available agent, tenofovir disoproxil fumarate, is an ester prodrug of tenofovir, with an effective tenofovir bioavailability of 25%. Administration following a high-fat meal increases the oral bioavailability.

Clinical Uses

Tenofovir is approved for the treatment of chronic hepatitis B in adults with evidence of active viral replication and either persistent elevations in serum aminotransferase levels or histologically active disease. Treatment with tenofovir is more effective than adefovir in producing histologic improvement and viral suppression in patients with HBeAg-negative or HBeAg-positive chronic hepatitis B.[1] Tenofovir also has demonstrated efficacy in patients with lamivudine-resistant HBV.[1]

Toxicity

Tenofovir is generally safe and well tolerated for up to 5 years.[2] The most common side effects are nausea, diarrhea, vomiting, and anorexia. Lactic acidosis with hepatic steatosis has been reported, primarily when it is used in combination with other nucleoside analogues. Acute exacerbations of hepatitis B have been reported after discontinuation of tenofovir in patients who are coinfected with HIV and HBV.

Antiviral Resistance

Mutations in HBV polymerase that confer reduced susceptibility to tenofovir occur during prolonged use (>12 months). In vitro studies showed that adefovir-resistant HBV mutations are associated with three- to five-fold decrease in response to tenofovir, though clinical implications are not known.

Entecavir

Entecavir, which is a deoxyguanosine nucleoside analogue with specific antiviral activity for hepadnaviruses, is more potent than lamivudine and retains some activity against lamivudine-resistant HBV variants. It is well absorbed after oral administration, and its prolonged half-life (128 to 149 hours) allows once-daily dosing.

Clinical Uses

Entecavir is approved for the treatment of chronic hepatitis B in adults with evidence of active viral replication and either persistent elevations in serum aminotransferases or histologically active disease. Compared with lamivudine or telbivudine, entecavir is more efficacious in reducing HBV DNA levels and normalizing serum aminotransferases, as well as in improving histologic abnormalities.[1,2] Like tenofovir, entecavir can be used for lamivudine-resistant HBV infections, but higher doses and longer durations of therapy are needed.

Toxicity

Adverse effects reported during entecavir therapy include headache, fatigue, dizziness, nausea, abdominal pain, rhinitis, fever, diarrhea, cough, and myalgia. Lactic acidosis and severe hepatomegaly with steatosis have been reported. Severe exacerbations of hepatitis B have been observed after cessation of therapy.

Antiviral Resistance

Virologic breakthrough can occur in up to 4% of patients but is usually not indicative of resistant virus. True entecavir resistance, which is caused by specific mutations in HBV polymerase, is uncommon (1.2% after 5 years of treatment). In vitro studies show that entecavir-resistant mutations are susceptible to adefovir and tenofovir, but there is very little supportive clinical data.

Interferons

Interferons are glycoprotein cytokines with a complex array of antiviral, immunomodulating, and antineoplastic properties. Interferons are currently classified as alfa, beta, or gamma. The natural sources of these classes in general are leukocytes, fibroblasts, and lymphocytes, but they now can be produced by recombinant DNA technology. Although the full mechanism of

interferon's action is not defined, interferons generally induce synthesis of new cellular RNA and proteins that mediate antiviral effects through multiple different mechanisms.

Interferons generally must be administered daily or several times per week. However, the combination of interferon with polyethylene glycol to form PEG-interferon significantly prolongs absorption and provides higher, more sustained plasma levels that enable once-weekly administration.

Clinical Uses

In chronic active hepatitis B, treatment with interferon alfa leads to loss of HBV DNA and biochemical and histologic improvement in about 25 to 40% of patients. Administration of PEG-interferon alfa-2a or alfa-2b for 48 weeks converts about 30% of patients to seronegative status after 6 months of treatment. Whether combination therapy with interferons and antivirals confers an additional benefit compared to monotherapy for treating chronic hepatitis B is not known.

Adefovir

Adefovir, an acyclic analogue of adenosine monophosphate, is administered orally as a prodrug, adefovir dipivoxil, which is rapidly converted enzymatically to adefovir in intestinal epithelium.

Clinical Uses

In chronic hepatitis B, prolonged administration of adefovir is effective in improving histologic abnormalities of the liver, decreasing HBV DNA levels, and normalizing biochemical (ALT) markers in patients with HBeAg-positive and HBeAg-negative chronic hepatitis B. Adefovir is a slightly weaker antiviral and therefore not considered first-line therapy, but it is useful for resistant viruses (effective against chronic hepatitis B resistant to lamivudine) and HIV/HBV coinfection (see later).

Toxicity

The major adverse effect is nephrotoxicity, which is manifested by increased serum creatinine levels and sometimes hypophosphatemia, both of which are usually reversible with discontinuation of the drug. Common side effects include asthenia, headache, nausea, vomiting, and diarrhea. In addition, severe exacerbations of hepatitis B have been observed after cessation of therapy.

Antiviral Resistance

Adefovir resistance due to point mutations in HBV polymerase develops in about 6% of patients after 3 years of therapy. Lamivudine generally retains activity against adefovir-resistant variants.

Lamivudine

Lamivudine is a deoxycytidine nucleoside analogue active against retroviruses and hepadnaviruses. The triphosphate inhibits HBV polymerase, and its incorporation into viral DNA results in termination of the DNA chain.

Clinical Uses

Lamivudine suppresses hepatitis B viral replication, improves histologic abnormalities of the liver, reduces progression of fibrosis, and decreases the risk for late complications. Monotherapy with lamivudine appears to be inferior to the newer antivirals, as well as monotherapy with interferon, for sustained control of HBV replication. Combination therapy with lamivudine and interferon has shown inconsistent benefit compared with either drug alone.

Toxicity

Adverse effects of lamivudine include diarrhea, headache, and elevated liver enzymes. Severe post-treatment exacerbations of hepatitis B, including fatalities, have occurred after discontinuation of lamivudine, especially in patients who are coinfected with HBV and HIV.

Antiviral Resistance

Lamivudine resistance caused by mutations in HBV polymerase is common during prolonged treatment of hepatitis B and emerges in about 20% of treated patients annually. Resistance is associated with increases in viral replication and aminotransferases.

Telbivudine

Telbivudine is a synthetic thymidine nucleoside analogue with activity against HBV, including some lamivudine-resistant variants. The triphosphate form competitively inhibits the HBV DNA polymerase (reverse transcriptase).

Clinical Uses

In comparative trials against lamivudine or adefovir, telbivudine demonstrated greater virologic response at week 52 (60% vs. 40% of subjects HBV DNA negative by polymerase chain reaction analysis). The development of resistance with telbivudine is up to 5% after 1 year and 25% after 2 years. Telbivudine-resistant viruses are cross-resistant with lamivudine. Owing to the high development of resistance and cross-resistance, telbivudine is not considered a first-line therapeutic for hepatitis B.

Toxicity

Common side effects include headache, nausea, and vomiting. Severe acute exacerbations of hepatitis B have been reported in patients who have discontinued anti-HBV therapy. Myopathy, manifested by muscle aches or weakness with increased serum creatine kinase levels, has rarely been reported.

Special Considerations—HIV/HBV Coinfection

Lamivudine and tenofovir are licensed and have activity against both HIV and HBV. Entecavir is licensed for the treatment of HBV, but it is also weakly active against HIV. Monotherapy treatment of HBV with these agents should not be used in HIV coinfected patients, owing to the development of HIV resistance. Emtricitabine is licensed only for the treatment of HIV but is also weakly active against HBV. Withdrawal of this medication, like withdrawal of tenofovir or lamivudine, may cause acute, sometimes fulminant, exacerbations of hepatitis B in HIV/HBV coinfected patients.

If concurrent treatment of HBV and HIV is warranted, tenofovir plus emtricitabine or tenofovir plus lamivudine should be considered as the nucleoside backbone. In coinfected patients who require treatment for HBV but not HIV (e.g., HBeAg-positive patients with high CD4 counts) or who are already well controlled on anti-HIV viral therapy but now need hepatitis B treatment as well, PEG-interferon alfa-2a or adefovir should be considered to avoid the addition of one antiviral with HIV and HBV activity and the potential development of HIV resistance. If the antiretroviral regimen must be changed owing to HIV virologic failure and the HBV is adequately suppressed, antiviral drugs active against HBV should be continued for HBV treatment.

ANTIVIRALS FOR HEPATITIS C VIRUS INFECTIONS

The treatment of chronic hepatitis C (Chapter 149) has significantly changed in the last few years. The goal of HCV therapy is sustained virologic response (absence of HCV RNA for at least 12 weeks after completion of therapy). Treatment has been shown to decrease liver inflammation and cirrhosis, as well as reduce the risk of hepatocellular carcinoma.

Interferon free regimens are now preferred over interferon based regimens. (Tables 360-1 to 360-3). Ledipasvir/sofosbuvir, sofosbuvir plus simeprevir, or ombitasvir/paritaprevir/ritonavir plus dasabuvir are preferred regimens for HCV genotype 1. Sofosbuvir and weight-based ribavirin are preferred for HCV geneotypes 2 and 3. Ledipasvir/sofosbuvir, or sofosbuvir plus simeprevir are preferred regimens for HCV genotype 4. Little data are available to guide recommendations for HCV genotype 5 or 6.

Interferon based regimens now are generally reserved for patients that do not have access to interferon free regimens. Sofosbuvir in combination with ribavirin and pegylated interferon has good efficacy for all genotypes.[3,4] Simpeprevir may be used as alterative regimens for genotypes 1 and 4.

Treatment of acute hepatitis C remains symptomatic, and antiviral therapy generally is not indicated.

Ledipasvir/Sofosbuvir

Ledipasvir is a HCV NS5A inhibitor. Sofosbuvir is a nucleotide prodrug that, when triphosphorylated, inhibits HCV NS5B, a RNA dependent RNA polymerase. Both NS5A and NS5B are important for HCV viral replication. Ledipasvir/sofosbuvir is available as a combination fixed dose once daily oral tablet.

Clinical Uses

Ledipasvir/sofosbuvir is currently licensed in the treatment of genotype 1 chronic HCV. For treatment naïve patients with or without cirrhosis, treatment should continue for 12 weeks. This regimen has been associated with a sustained virologic response of 99% at 12 weeks. The duration of treatment may be reduced to as low as 8 weeks without significant changes in efficacy. For treatment-experienced patients, 12 weeks of treatment is recommended for those without cirrhosis and 24 weeks for those with cirrhosis; the

sustained virologic response is also 94-99%.[5] Ledipasvir/sofosbuvir has also been recommended in the treatment of genotype 4 chronic HCV.

Toxicity
The most common side effects are headaches and fatigue. Significant drug interactions may occur with medications such as rifampin or St. John's wort.

Resistance
Resistance to the individual component drug may occur but is rare. Ledipasvir is active against sofosbuvir-resistant viruses, and sofosbuvir retains activity against ledipasvir-resistant viruses, so the clinical significance of resistance mutations is not known.

Ombitasvir/Paritaprevir/Ritonavir plus Dasabuvir
Ombitasvir is a HCV NS5A inhibitor and paritaprevir is an inhibitor of the HCV NS3/4A protease. Ritonavir has no activity against HCV, but it is an inhibitor of hepatic CYP3A and is added to increase paritaprevir levels. These three drugs are coformulated into one tablet that is packaged with dasabuvir, an inhibitor of the HCV NS5B polymerase.

Clinical Uses
Ombitasvir/paritaprevir/ritonavir plus dasabuvir is currently licensed in the treatment of genotype 1 chronic HCV. In patients without cirrhosis, treatment for 12 weeks has been associated with a sustained virologic response of 97-99%. For geneotype 1a, the addition of ribavirin may decrease virologic failures.[5]

Toxicity
The many drug interactions that exist due to the ritonavir inhibition of CYP3A should be closely evaluated prior to initiation of therapy. The most common side effects include nausea and diarrhea, as well as rash, headache, and fatigue.

Resistance
Treatment-emergent mutations were observed in patients with virologic failure, though the clinical significance of this is not known.

Sofosbuvir
Clinical Uses
Sofosbuvir is currently licensed as a component of a combination antiviral treatment regimen for HCV genotypes 1 to 4. For genotypes 1 and 4, sofosbuvir is used in combination with PEG-interferon and ribavirin for 12 weeks. This regimen has been associated with a sustained virologic response of about 90% at 12 weeks.[5] In genotypes 2 and 3, sofosbuvir in combination with ribavirin for 12 or 24 weeks also has been associated with about a 90% sustained virologic response.[5] For patients who cannot tolerate an interferon-based regimen (autoimmune disorders, decompensated hepatic disease, leukopenia, thrombocytopenia, anemia, or preexisting cardiac disease), sofosbuvir may be used with simeprevir or ribavirin for 24 weeks.

Toxicity
Common side effects include headache, fatigue, anemia, and diarrhea. Severe pancytopenias and severe depression have been reported. Sofosbuvir is a substrate of drug transporter P-gp, and drugs that are potent P-gp inducers (e.g., rifampin or St. John's wort) may decrease sofosbuvir's plasma concentration.

Resistance
The emergence of resistance while on sofosbuvir treatment is extremely rare.

Simeprevir
Simeprevir is a protease inhibitor that binds to the NS3/4a of hepatitis C.

Clinical Uses
Simeprevir is currently licensed for the treatment of chronic hepatitis C genotype 1 as either initial therapy or for use after failure of prior interferon-based therapy. Simeprevir is given for 12 weeks in combination with peginterferon alfa and ribavirin, followed by peginterferon alfa and ribavirin for an additional 12 weeks (treatment naïve) or 36 weeks (prior nonresponse). In naïve patients, virologic response is about 80%.[5] Although not as effective as sofosbuvir, simprevir is superior to other protease inhibitors and has become second-line therapy in genotype 1 (FDA approved) and genotype 4 (not FDA approved) hepatitis C.

Toxicity
Dermatologic (including rash, pruritus, and photosensitivity) and gastrointestinal side effects are the most common adverse reactions.

Resistance
Intrinsic resistance of the NS3 to simeprevir can be caused by the Q80K polymorphism. Patients should be screened for baseline mutation of this gene prior to initiation of therapy.

Boceprevir
Boceprevir is a protease inhibitor that binds to the NS3/4a of hepatitis C.

Clinical Uses
Boceprevir is currently licensed only for the treatment of genotype 1 in patients with compensated liver disease.[5] It is used in combination with PEG-interferon and ribavirin in treatment-naïve patients as well as patients who have not responded adequately to PEG-interferon and ribavirin. Boceprevir should not be used as monotherapy. Boceprevir is started 4 weeks after initiation of therapy with PEG-interferon and ribavirin, and combination therapy is continued for an additional 24 to 44 weeks based on virologic response.[5] Subjects with undetectable HCV RNA level at weeks 8 and 24 may be considered for a shorter duration of treatment (28 weeks in total). In patients with evidence of virologic failure (HCV RNA level is > 100 IU/mL at treatment week 12 or detectable at treatment week 24), treatment with all three drugs should be stopped. Because of its inferior efficacy, boceprevir is not considered a first-line therapy for hepatitis C.

Toxicity
Common side effects include fatigue, nausea, shivering, alopecia, and dysgeusia. Patients should be monitored for serious side effects such as anemia and neutropenia, which occasionally can be dose-limiting. Patients who develop anemia on combination therapy including boceprevir may be managed by reducing the ribavirin dose. Boceprevir is a strong inhibitor of cytochrome P450 3A4/5, so drug-drug interactions (including over-the-counter and oral contraceptives) must be considered.

Resistance
Resistance to boceprevir can occur on therapy. Because boceprevir and telaprevir are structurally similar, cross-resistant mutations may emerge, and patients should not be retreated with another protease inhibitor.

Telaprevir
Telaprevir was previously licensed for the treatment of chronic HCV but was withdrawn from the market in 2014.

Pegylated Interferon 2b and Ribavirin
Clinical Uses
Neither ribavirin nor interferon should be used as monotherapy for hepatitis C. In genotype 1, the combination of PEG-interferon and ribavirin may be considered in treatment-naïve subjects who cannot be given interferon-free regimens, although sustained virologic response rates are significantly inferior. Longer durations of treatment, up to 72 weeks, may improve virologic response.[5] PEG-interferon and ribavirin should be administered for 24 weeks for HCV genotypes 2 and 3 and for 48 weeks for HCV genotypes 4, 5, and 6.

Toxicity
Common side effects of interferon administration include influenza-like symptoms (fever, chills, headache, malaise), but these symptoms usually become less severe with repeated treatments. Major toxicities include bone marrow suppression, especially granulocytopenia and thrombocytopenia, which are generally reversible when therapy is discontinued. Neuropsychiatric disturbances may be manifested by depression, anxiety, somnolence, confusion, and behavioral changes. Other side effects include profound fatigue and anorexia, weight loss, hypothyroidism or hyperthyroidism, alopecia, and cardiotoxicity with arrhythmias and reversible cardiomyopathy. Some side effects (e.g., hypothyroidism) can present a year or longer after therapy has finished.

Systemic ribavirin is frequently associated with hemolytic anemia (in up to 60% in some series) and sometimes with electrolyte abnormalities, including hypocalcemia and hypomagnesemia. Arrhythmias, pruritus, rash, nausea, and myalgia have been reported, as have neurologic side effects including insomnia and irritability. Ribavirin may be gonadotoxic and teratogenic in humans.

Special Considerations

HIV/HCV coinfection can be treated concurrently, but treatment can be complicated by drug interactions and significant toxicities.[6] Ledipasvir/sofosbuvir is considered first line therapy. Ledipasvir can increase tenofovir levels, and HIV regimens without tenofovir should be considered. Ombitasvir/paritaprevir/ritonavir plus dasabuvir will have interactions with many antiretroviral drugs. This regimen can be used with most nucleoside inhibitors and raltegravir, but use of other protease inhibitors, especially ritonavir-boosted regimens, requires expert consultation.

The HIV antiviral didanosine (DDI [Chapter 388]) should not be used with ribavirin because DDI levels and its active metabolite are increased when administered with ribavirin and can cause hepatic failure, severe peripheral neuropathy, pancreatitis, and lactic acidosis. Zidovudine (or azidothymidine [AZT]) administered with ribavirin can cause severe neutropenic anemia, and the combination should not be used. If ribavirin is the preferred therapy, discontinuation of DDI and AZT should be considered prior to initiation of ribavirin.

ANTIVIRALS FOR HERPESVIRUS INFECTIONS

Acyclovir and Valacyclovir

Acyclovir, which is an acyclic analogue of the nucleoside guanosine, is converted to its active form through initial monophosphorylation by a virus-encoded thymidine kinase (TK). Although normal human cells possess TK, the affinity of acyclovir for viral TK is approximately 200 times greater than for human TK. The monophosphate then undergoes two additional host cell enzyme-mediated phosphorylations to acyclovir triphosphate (acycloguanosine triphosphate), which preferentially inhibits viral DNA polymerase. The higher concentrations of the activated form in infected cells and its affinity for viral polymerases result in low toxicity to normal host cells.

Valacyclovir is the L-valyl ester prodrug of acyclovir. Addition of the L-valyl ester fosters greater oral absorption, after which valacyclovir is converted to acyclovir; the prodrug provides three to five times greater bioavailability than oral acyclovir.

Clinical Uses

Acyclovir and valacyclovir (Tables 360-4 to 360-6) are used principally to treat infections caused by herpes simplex virus (HSV [Chapter 374]) and varicella-zoster virus (VZV [Chapter 375]). Depending on the country, acyclovir is available in a topical ointment and cream, oral capsules, and intravenous and ophthalmic formulations. Valacyclovir is available only as an oral capsule.

Oral acyclovir or valacyclovir decreases the duration of symptoms by approximately 50% and reduces the duration of viral shedding by about 90% in initial episodes of genital herpes. Two or 3 days of therapy appears to be sufficient for recurrent genital herpes. Chronic suppression is highly effective in reducing clinical and viral recurrences, and valacyclovir reduces the risk of

TABLE 360-4 ANTIVIRALS FOR HERPESVIRUS INFECTIONS

VIRAL INFECTION	DRUG	ROUTE	USUAL ADULT DOSAGE
HERPES SIMPLEX VIRUS			
Genital herpes			
First episode	Acyclovir	PO	400 mg tid or 200 mg 5 times/day for 7-10 days
	Famciclovir	PO	250 mg tid for 7-10 days
	Valacyclovir	PO	1 g bid for 7-10 days
Recurrent	Acyclovir	PO	800 mg tid for 2 days, or 400 mg tid or 200 mg 5 times/day for 5 days
	Famciclovir	PO	1000 mg bid for 2 doses
	Valacyclovir	PO	500 mg bid for 3 days or 1 g/day for 5 days
Suppression	Acyclovir	PO	400 mg bid or 200 mg tid
	Famciclovir	PO	250 mg bid
	Valacyclovir	PO	500 mg/day or 1 g/day (10 or more episodes/year)
Orolabial herpes	Penciclovir 1%	Topical	Apply cream every 2 hr while awake for 4 days
	Acyclovir 5%	Topical	Apply cream 5 times/day for 4 days
	Docosanol 10%	Topical	Apply cream 5 times/day until healing
	Valacyclovir	PO	2 g q12h × 2 doses
	Acyclovir	PO	400 mg 5 times/day for 5 days
	Famciclovir	PO	1500 mg single dose
Mucocutaneous disease	Acyclovir	IV	5 mg/kg/8 hr for 7-14 days
	Acyclovir	PO	400 mg 5 times/day for 7-14 days
	Valacyclovir	PO	500 mg or 1 g bid for 7-10 days
Encephalitis	Acyclovir	IV	10 mg/kg/8 hr for 10 days
Neonatal	Acyclovir	IV	10-20 mg/kg/8 hr for 14-21 days
Keratoconjunctivitis	Trifluridine	Topical	1 drop of 1% solution q2h up to 9 drops/day
	Vidarabine	Topical	½-inch ribbon of 3% ointment 5 times daily
Acyclovir-resistant HSV	Foscarnet	IV	40 mg/kg q8-12 h until healed.
CYTOMEGALOVIRUS			
CMV retinitis	Ganciclovir	IV	Induction: 5 mg/kg/12 hr for 14-21 days (maintenance: 5 mg/kg/day)
	Valganciclovir	PO	Induction: 900 mg bid for 21 days (maintenance: 900 mg/day)
	Cidofovir	IV	5 mg/kg once weekly × 2, then every other week (maintenance: 5 mg/kg q2wk)
	Foscarnet	IV	60 mg/kg/8 hr or 90 mg/kg q12h for 14-21 days (maintenance: 90-120 mg/kg daily)
	Fomivirsen	Intravitreal	330 μg q2wk × 2, then q4wk (maintenance: 330 μg every month)
HIV infection, CMV colitis or esophagitis	Ganciclovir	IV	5 mg/kg/12 hr for 14-28 days (until symptoms resolved)
Prophylaxis (transplantation)	Ganciclovir	IV	5 mg/kg/12 hr for 7-14 days, then 5 mg/kg IV once a day
	Valganciclovir	PO	900 mg daily
Prophylaxis (advanced HIV infection)	Ganciclovir	IV	5 mg/kg daily
VARICELLA-ZOSTER VIRUS			
Varicella	Acyclovir	PO	800 mg qid for 5 days
Varicella in immunocompromised hosts	Acyclovir	IV	10 mg/kg/8 hr for 7-10 days
Herpes zoster in immunocompromised hosts	Acyclovir	IV	10 mg/kg/8 hr for 7-10 days
Herpes zoster in normal hosts	Acyclovir	PO	800 mg 5 times daily for 7-10 days
	Valacyclovir	PO	1 g tid for 7 days
	Famciclovir	PO	500 mg tid for 7 days

TABLE 360-5 MECHANISMS OF EXCRETION AND THRESHOLDS FOR DOSE ADJUSTMENT

	MAJOR ROUTE OF ELIMINATION	THRESHOLD FOR ADJUSTMENT IN RENAL INSUFFICIENCY OR FAILURE	ADJUSTMENT FOR HEPATIC FAILURE	ADJUSTMENT FOR OBESITY
Acyclovir IV	Renal	CrCl < 50 mL/min	No adjustment	Dose by ideal body weight
Acyclovir PO	Renal	CrCl < 25 mL/min	No adjustment	Dose by ideal body weight
Valacyclovir	Renal	CrCl < 50 mL/min	No adjustment	Unknown
Famciclovir	Renal	CrCl < 60 mL/min	No adjustment	Unknown
Foscarnet	Renal	CrCl < 1.4 mL/min/kg	No adjustment	Unknown
Ganciclovir IV	Renal	CrCl < 70 mL/min	No adjustment	Unknown
Valganciclovir	Renal	CrCl < 60 mL/min	No adjustment	Unknown
Cidofovir	Renal	CrCl < 55 mL/min	No adjustment	Unknown

TABLE 360-6 SIGNIFICANT ADVERSE EFFECTS (U.S. FDA BLACK BOX WARNING)

DRUG	BLACK BOX SYNOPSIS
Cidofovir	Renal impairment, including renal failure; prehydrate and use probenecid Neutropenia May be carcinogenic and teratogenic and may cause hypospermia or aspermia
Foscarnet	Nephrotoxicity; prehydrate Seizures related to minerals and electrolyte disturbances
Ganciclovir	Neutropenia, anemia, thrombocytopenia May be carcinogenic and teratogenic and may cause hypospermia or aspermia
Valganciclovir	Neutropenia, anemia, thrombocytopenia May be carcinogenic and teratogenic and may cause hypospermia or aspermia

transmission of genital HSV between heterosexual partners by 48%. For herpes labialis (cold sores), 1 day of therapy with oral valacyclovir improves time to healing and reduces pain, whereas acyclovir ointment has no consistent clinical benefit.

Parenteral acyclovir is indicated for the initial treatment of mucosal or cutaneous HSV infection in immunocompromised patients, neonatal HSV infections, and disseminated or organ-invasive infections in immunocompetent patients. A subsequent switch to oral valacyclovir is possible in some circumstances. High-dose parenteral acyclovir is the therapy of choice for treatment of HSV encephalitis.

Acyclovir is also effective treatment of VZV infections, although higher doses are needed than for mucosal HSV infections. In adults treated within 24 hours of the development of a varicella rash, acyclovir decreases the severity of disease and number of lesions, but oral valacyclovir may be more effective than oral acyclovir. Intravenous acyclovir is warranted for initial treatment of varicella and zoster in immunocompromised hosts. Both acyclovir chemoprophylaxis and valacyclovir chemoprophylaxis reduce the incidence of recurrent HSV in recipients of stem cell and solid organ transplants, but valacyclovir is superior for prevention of cytomegalovirus (CMV) disease.

Toxicity

Acyclovir and valacyclovir have excellent safety profiles and are generally well tolerated. Common side effects include nausea, vomiting, and headaches. Major adverse effects include renal dysfunction and central nervous system (CNS) toxicity. Dehydration and preexisting renal dysfunction predispose to the development of renal impairment. Neurologic side effects include tremor, myoclonus, confusion, lethargy, agitation, and hallucination. Renal dysfunction predisposes to the development of neurotoxicity. Neutropenia and other signs of bone marrow toxicity have also been reported rarely.

Antiviral Resistance

Despite widespread use of acyclovir, the development of HSV resistance in immunocompetent subjects is uncommon (prevalence < 1%). However, antiviral resistance is higher in immunocompromised subjects, including those with HIV infection (prevalence of 5%) or bone marrow transplants (prevalence of up to 30%). Drug-resistant refractory VZV infections can occur in highly immunocompromised patients. Intravenous foscarnet or cidofovir may be effective for infections caused by acyclovir-resistant viruses.

Penciclovir and Famciclovir

Penciclovir is an acyclic guanine analogue that unlike acyclovir is not an obligate chain terminator and may be incorporated into DNA. Penciclovir is phosphorylated by viral TK to penciclovir monophosphate, which is then converted to penciclovir triphosphate. Penciclovir demonstrates in vitro activity against VZV and HSV comparable to that of acyclovir. The bioavailability of penciclovir after oral administration is less than 2%. In contrast, famciclovir is an oral prodrug that is deacetylated and oxidized in the liver to form penciclovir; the bioavailability of penciclovir averages 77% after administration of famciclovir.

Clinical Uses

Penciclovir and famciclovir are used to treat HSV and VZV infections. Penciclovir is available as a topical cream and in some countries as an intravenous formulation. Famciclovir is available as a capsule.

Topical penciclovir, which is approved for the treatment of recurrent HSV labialis, reduces pain and lesions by about 1 day. Famciclovir is approved for the treatment of recurrent HSV labialis and genital infections and for herpes zoster, and its effectiveness is similar to valacyclovir or acyclovir. Famciclovir may also be used for suppressive therapy.

Toxicity

Topical penciclovir is well tolerated; the majority of adverse reactions are local irritation and mild erythema. Adverse effects of oral famciclovir include headache, dizziness, nausea, and diarrhea.

Antiviral Resistance

Penciclovir resistance in HSV has been uncommon in immunocompetent subjects but, like acyclovir resistance, more frequent in immunocompromised hosts (2.1%). Most acyclovir-resistant HSV isolates are cross-resistant to penciclovir.

Ganciclovir and Valganciclovir

Ganciclovir is an acyclic deoxyguanosine analogue with antiviral activity against multiple herpesviruses, including HSV, VZV, CMV (Chapter 376), Epstein-Barr virus (EBV [Chapter 377]), and human herpesvirus 8. It is much more active than acyclovir against CMV and EBV. The bioavailability of oral ganciclovir is less than 10%. Valganciclovir, the L-valyl prodrug of ganciclovir, increases the bioavailability of ganciclovir to approximately 60% after oral administration.

Clinical Uses

Ganciclovir is available as an oral capsule, a parenteral injection, and an ocular implant; valganciclovir is available only as a tablet. Ganciclovir and valganciclovir are effective for treatment of CMV retinitis, for which they are comparably active. In the absence of immune reconstitution, long-term suppression therapy is necessary. They are also used for life-threatening CMV diseases in patients with acquired immunodeficiency syndrome (AIDS) and other immunocompromised conditions and for prevention of CMV disease in transplant patients. For immunocompromised patients with organ-invasive CMV infections, intravenous ganciclovir provides clinical response rates of 70 to 90%, although response rates are lower for CMV pneumonitis after stem cell transplantation or CMV encephalitis in patients with AIDS. Oral valganciclovir provides long-term outcomes similar to those of intravenous ganciclovir for the treatment of CMV disease.

Long-term prophylaxis with ganciclovir or valganciclovir reduces the incidence of CMV disease after solid organ and stem cell transplantation,

but this therapy has substantial side effects, including bone marrow suppression. These drugs can also be used as preemptive therapy for patients who have CMV viremia or antigenemia. In the prevention of CMV disease, preemptive valganciclovir therapy may be equally effective to chronic valacyclovir prophylaxis.

Toxicity

The most common adverse effect with ganciclovir and valganciclovir is bone marrow suppression, particularly neutropenia and thrombocytopenia, which occur in up to 50% of patients given intravenous ganciclovir. Fever, edema, phlebitis, headache, neuropathy, disorientation, nausea, anorexia, rash, and myalgias have also been reported with ganciclovir therapy. Intravitreal ganciclovir implants may cause vitreous hemorrhage and retinal detachment.

Antiviral Resistance

Ganciclovir resistance secondary to mutations in CMV kinase and sometimes DNA polymerase is related to the length of ganciclovir exposure and the degree of immunosuppression. Resistance may be associated with progressive disease during continued ganciclovir use; foscarnet and cidofovir are alternative treatments.

Cidofovir

Cidofovir, which is an acyclic phosphonate derivative of cytosine, is phosphorylated to its active diphosphate form by host cellular enzymes. Cidofovir diphosphate competitively inhibits viral DNA polymerase and viral DNA synthesis. Despite a short serum half-life, the antiviral effects are protracted because of prolonged intracellular concentrations of the phosphorylated metabolite.

Clinical Uses

Cidofovir is commercially available as an intravenous infusion, and investigational formulations have included topical gel and intravitreal and intralesional injections. Intravenous cidofovir is licensed for the treatment of cytomegalovirus retinitis. Because of its toxicities, it is generally reserved for when ganciclovir or foscarnet therapy is failing. Limited data suggest that intravenous cidofovir may be effective in other CMV infections (pneumonitis, gastroenteritis), acyclovir- or foscarnet-resistant HSV infections, certain forms of human papillomavirus disease, invasive adenoviral infections in transplant recipients, and possibly BK virus infection in renal transplant patients. In addition, in vivo and animal data suggest efficacy of cidofovir against smallpox, vaccinia, and monkeypox infections.

Toxicity

Dose-related nephrotoxicity, characterized by increased serum creatinine, proteinuria, and tubular dysfunction, is the main side effect of intravenous cidofovir. Adequate hydration and concomitant oral probenecid reduce the risk. Other common side effects include diarrhea, asthenia, nausea, vomiting, neutropenia, fever, and rash. Iritis, intraocular pressure changes, loss of visual acuity, and uveitis have been reported with intravenous cidofovir. Intravitreal cidofovir is effective but locally toxic.

Antiviral Resistance

Sustained exposure to cidofovir does not easily induce resistance, although resistance has infrequently been described in HSV and CMV.

Foscarnet

Foscarnet is a pyrophosphate analogue that acts as a noncompetitive inhibitor of many viral RNA and DNA polymerases. When a nucleotide is incorporated into a DNA or RNA strand by a polymerase, pyrophosphate is released. Foscarnet directly inhibits viral polymerases without phosphorylation, so TK-deficient acyclovir-resistant HSV and VZV are susceptible to this agent.

Clinical Uses

Foscarnet is as effective as ganciclovir for the treatment of CMV retinitis in patients with AIDS, and combination therapy with ganciclovir may be superior to monotherapy with either agent for recalcitrant retinitis. For extraretinal CMV disease, foscarnet has demonstrated efficacy similar to that of ganciclovir. The choice of agent may be dictated by the side-effect profile. Foscarnet is also effective for the treatment of acyclovir-resistant HSV and VZV infections.

Toxicity

Nephrotoxicity with azotemia and proteinuria is dose limiting and occurs in more than a third of patients. A slow infusion rate and saline hydration reduce the risk. Other common side effects include anemia (30 to 50% of patients), granulocytopenia, diarrhea, nausea, vomiting, fever, seizures, paresthesias, headache, and genital ulcers. Marked electrolyte disturbances may develop, including hypophosphatemia, hypocalcemia, hypokalemia, and hypomagnesemia. Foscarnet can prolong the QT interval and be associated with cardiac arrhythmias, including ventricular tachycardia, ventricular fibrillation, and torsades de pointes.

Antiviral Resistance

The development of CMV resistance to foscarnet as a result of mutations in viral DNA polymerase is uncommon except after prolonged administration. In AIDS patients with retinitis, foscarnet resistance is detectable in 13% of patients at 6 months and in 37% at 12 months.

Fomivirsen

Fomivirsen, which is an antisense oligonucleotide that inhibits CMV replication, is currently available as an intravitreal injection that is effective for both newly diagnosed CMV retinitis and CMV retinitis failing usual therapies, although direct comparisons with other agents are lacking. Intravitreal administration of fomivirsen may cause increased intraocular pressure, iritis, vitreitis, and cataracts in 10 to 20% of patients. Topical corticosteroids may be useful for treating inflammatory changes.

Docosanol

Docosanol, which is a 22-carbon saturated fatty alcohol that inhibits intracellular penetration of lipid-enveloped viruses, is approved as a cream for the treatment of herpes labialis. Frequent topical applications have shown reductions in time to cessation of pain and healing, but direct comparison to other agents is lacking. Local reaction, rash, and pruritus are common side effects.

ANTIVIRALS FOR INFLUENZA VIRUS INFECTIONS

Currently approved antivirals for the treatment of influenza include two neuraminidase inhibitors (oseltamivir and zanamivir) and two adamantanes (amantadine and rimantadine) (Tables 360-7 and 360-8). The choice of treatment for influenza should be dictated by circulating strains, antiviral resistance within those strains, and side-effect profiles.

Oseltamivir, Zanamivir, and Peramivir (Neuraminidase Inhibitors)

Oseltamivir, zanamivir, and peramivir are sialic acid analogues that inhibit influenza virus by competitively interacting with the neuraminidases of influenza A and B viruses. Influenza neuraminidase cleaves terminal sialic acid residues and destroys the receptors recognized by viral hemagglutinin. By this mechanism, the drugs inhibit the release of virus from infected cells, thereby preventing viral aggregates and spread within the respiratory tract.

Oseltamivir is administered orally as the phosphate prodrug, which is rapidly absorbed and hydrolyzed to the active form oseltamivir carboxylate. Bioavailability exceeds 75%. Conversely, the oral bioavailability of zanamivir is poor, so it is delivered as an orally inhaled powder. Peramivir is available only as a parenterally-infused solution.

Clinical Uses

Oseltamivir and zanamivir are effective for the treatment and prophylaxis of acute influenza A and B infections. Early treatment in adults decreases the duration and severity of illness and reduces lower respiratory tract complications, antibiotic use, and, with oseltamivir, hospitalizations.[7,8] In low-risk ambulatory subjects, oseltamivir alleviates symptoms faster than placebo, and it also decreases the risk for lower respiratory complications. In cohort studies of hospitalized subjects, treatment with oseltamivir has been associated with a significant reduction in death.[9] Zanamivir is also effective in alleviating symptoms and decreasing the risk for lower respiratory complications. Both zanamivir and oseltamivir are also highly effective for the

TABLE 360-7 ANTIVIRALS FOR INFLUENZA VIRUS INFECTIONS

VIRUS	DRUG	ROUTE	USUAL ADULT TREATMENT DOSAGE
Influenza A and B virus	Oseltamivir	PO	75 mg bid for 5 days
	Peramivir	IV	600 mg single IV dose
	Zanamivir	Inhalation	10 mg bid by inhaler for 5 days
Influenza A virus	Amantadine	PO	100 mg bid or 200 mg daily for 5 days
	Rimantadine	PO	100 mg bid for 5 days

TABLE 360-8 MECHANISMS OF EXCRETION AND THRESHOLDS FOR DOSE ADJUSTMENTS

	MAJOR ROUTE OF ELIMINATION	THRESHOLD FOR ADJUSTMENT IN RENAL INSUFFICIENCY OR FAILURE	ADJUSTMENT FOR HEPATIC FAILURE	SPECIAL ADJUSTMENT FOR THE ELDERLY
Amantadine	Renal	CrCl < 50 mL/min	No adjustment	>65 years: 100 mg daily
Rimantadine	Hepatic and renal	CrCl < 10 mL/min/1.72 m^2	100 mg daily	100 mg daily
Oseltamivir	Renal	CrCl < 30 mL/min/1.72 m^2	No adjustment	
Peramivir	Renal	CrCl < 50 mL/min	No adjustment	
Zanamivir	Renal	No adjustment	No adjustment	

prevention of influenza. Peramivir is licensed as a single-dose IV infusion for uncomplicated influenza. However, as a daily infusion in patients hospitalized with influenza, its efficacy is similar to oseltamivir.[A1]

Toxicity

The most common side effects with oseltamivir are nausea and vomiting. They also may be associated with headache, rash, and possibly abnormal aminotransferase levels. Zanamivir is generally well tolerated, but severe bronchospasm has been reported primarily in patients with underlying airway disease. The most common side effects of peramivir are nausea, diarrhea, and mild neutropenia.

Antiviral Resistance

Oseltamivir resistance can be can be preexisting (widespread or local) or can emerge during therapy. In the immunocompromised host and possibly in individuals with H5N1 or H1N1, the development of resistance is associated with treatment failure. Zanamivir resistance is rare, and zanamivir retains clinical effectiveness against the most common oseltamivir-resistant variants.

Amantadine and Rimantadine (Adamantanes)

Amantadine and rimantadine are symmetrical tricyclic amines with activity against many influenza A viruses (Chapter 364). By inhibiting the ion channel function of the M2 protein of influenza A, they interfere with uncoating of the virus and release of the viral genome.

Clinical Uses

Amantadine and rimantadine decrease the length and severity of uncomplicated influenza A virus infection by susceptible strains if they are initiated within the first 2 days after the onset of symptoms, but it is uncertain whether they reduce the risk for complications. Both drugs are formulated for oral administration, and amantadine also has a pediatric syrup formulation. In recent years, marked increases in antiviral resistance in community isolates have limited the utility of these drugs.

When adamantine-susceptible viruses are circulating, both rimantadine and amantadine are effective when they are used for prophylaxis (overall 66% average for rimantadine and 74% for amantadine). Despite prophylaxis, subclinical infections may still develop and elicit immune responses that are protective against antigenically related viruses.

Toxicity

Amantadine causes CNS side effects in 10 to 30% of healthy young adults who take the standard adult dose of 200 mg/day; the frequency is significantly lower with rimantadine. Neuropsychiatric side effects include anxiety, nervousness, insomnia, and, particularly in the elderly or those with renal insufficiency, hallucinations, confusion, disorientation, and psychosis or coma. Amantadine (or less often rimantadine) is associated with an increased risk for seizures. Both drugs cause gastrointestinal side effects. Orthostatic hypotension occurs in 1 to 5%. Anticholinergic side effects, including dry mouth, occur in amantadine recipients.

Antiviral Resistance

Single point mutations in M2 confer high-level resistance to these drugs and make them ineffective. Such resistant variants emerge commonly during treatment and are transmissible. Currently circulating H1N1 and H3N2 viruses are highly resistant to adamantanes.

OTHER ANTIVIRALS

Ribavirin

Ribavirin is a purine nucleoside with antiviral activity against some DNA viruses and many RNA viruses, including influenza A and B, parainfluenza, coronaviruses including Middle East respiratory syndrome (Chapter 366), measles (Chapter 367), respiratory syncytial virus (RSV [Chapter 362]), retroviruses (Chapter 378), arenaviruses such as Lassa virus (Chapter 381), and some hantaviruses (Chapter 381).

Clinical Uses

Aerosol administration of ribavirin has been used to treat RSV bronchiolitis and pneumonia in children and to treat influenza A and B infections. Limited benefit has been seen with oral ribavirin in uncomplicated influenza. Aerosol ribavirin combined with intravenous immune globulin, particularly with the anti-RSV monoclonal antibody palivizumab, appears to reduce the mortality of RSV infection in bone marrow transplant and other highly immunocompromised patients.

Systemic ribavirin reduces the mortality associated with Lassa fever and Asian (Korean) hemorrhagic fever with renal syndrome (Chapter 381), although not mortality in patients with hantavirus-associated cardiopulmonary syndrome. It appears to have activity in Congo-Crimean hemorrhagic fever and in Nipah virus infections. Ribavirin is often recommended as treatment of hemorrhagic fevers of unknown etiology or secondary to arenaviruses or bunyaviruses in the event these viruses are used as biological weapons.

Given the breadth of ribavirin's activity against many viruses, it is often used for emerging infectious diseases (e.g., severe acute respiratory syndrome [SARS] and Middle East respiratory syndrome). However, the concentrations needed to show efficacy in cell cultures and animal models may be significantly higher than can be obtained in humans. Caution is therefore urged when ribavirin is suggested for treating these diseases.

Imiquimod

Imiquimod and the related compound resiquimod are topical immune response modifiers that lack direct antiviral effects. Instead, these agents induce activation of immune cells (monocytes, macrophages, natural killer cells) to produce antiviral cytokines, particularly interferon-α and tumor necrosis factor-α.

Topical imiquimod cream is approved for the treatment of condyloma acuminatum (Chapter 373). In immunocompetent patients, imiquimod leads to complete clearance of warts in 37 to 52% of patients. It is administered as a topical cream three times weekly for a maximum of 16 weeks and is washed off 6 to 10 hours after application.

Side effects are primarily local and include erythema, irritation, tenderness, and (less often) erosion. The side effects usually resolve with cessation of the drug.

Grade A References

A1. Marcellin P, Heathcote EJ, Buti M, et al. Tenofovir disoproxil fumarate versus adefovir dipivoxil for chronic hepatitis B. *N Engl J Med*. 2008;359:2442-2455.

A2. Kim YJ, Sinn DH, Gwak GY, et al. Tenofovir rescue therapy for chronic hepatitis B patients after multiple treatment failures. *World J Gastroenterol*. 2012;18:6996-7002.

A3. Patterson SJ, George J, Strasser SI, et al. Tenofovir disoproxil fumarate rescue therapy following failure of both lamivudine and adefovir dipivoxil in chronic hepatitis B. *Gut*. 2011; 60:247-254.

A4. Shim JH, Suh DJ, Kim KM, et al. Efficacy of entecavir in patients with chronic hepatitis B resistant to both lamivudine and adefovir or to lamivudine alone. *Hepatology*. 2009;50:1064-1071.

A5. Tsai MC, Lee CM, Chiu KW, et al. A comparison of telbivudine and entecavir for chronic hepatitis B in real-world clinical practice. *J Antimicrob Chemother*. 2012;67:696-699.

A6. Hadziyannis SJ, Tassopoulos NC, Heathcote EJ, et al. Adefovir dipivoxil for the treatment of hepatitis B e antigen-negative chronic hepatitis B. *N Engl J Med*. 2003;348:800-807.

A7. Marcellin P, Chang TT, Lim SG, et al. Adefovir dipivoxil for the treatment of hepatitis B e antigen-positive chronic hepatitis B. *N Engl J Med*. 2003;348:808-816.

A8. Liaw YF, Sung JJ, Chow WC, et al. Lamivudine for patients with chronic hepatitis B and advanced liver disease. *N Engl J Med*. 2004;351:1521-1531.

A9. Liaw YF, Gane E, Leung N, et al. 2-Year GLOBE trial results: telbivudine is superior to lamivudine in patients with chronic hepatitis B. *Gastroenterology*. 2009;136:486-495.
A10. Afdhal N, Zeuzem S, Kwo P, et al. Ledipasvir and sofosbuvir for untreated HCV genotype 1 infection. *N Engl J Med*. 2014;370:1889-1898.
A11. Kowdley KV, Gordon SC, Reddy KR, et al. Ledipasvir and sofosbuvir for 8 or 12 weeks for chronic HCV without cirrhosis. *N Engl J Med*. 2014;370:1879-1888.
A12. Afdhal N, Reddy KR, Nelson DR, et al. Ledipasvir and sofosbuvir for previously treated HCV genotype 1 infection. *N Engl J Med*. 2014;370:1483-1493.
A13. Ferenci P, Bernstein D, Lalezari J, et al. ABT-450/r-ombitasvir and dasabuvir with or without ribavirin for HCV. *N Engl J Med*. 2014;370:1983-1992.
A14. Lawitz E, Mangia A, Wyles D, et al. Sofosbuvir for previously untreated chronic hepatitis C infection. *N Engl J Med*. 2013;368:1878-1887.
A15. Zeuzem S, Dusheiko GM, Salupere R, et al. Sofosbuvir and ribavirin in HCV genotypes 2 and 3. *N Engl J Med*. 2014;370:1993-2001.
A16. Jacobson IM, Dore GJ, Foster GR, et al. Simeprevir with pegylated interferon alfa 2a plus ribavirin in treatment-naive patients with chronic hepatitis C virus genotype 1 infection (QUEST-1): a phase 3, randomised, double-blind, placebo-controlled trial. *Lancet*. 2014;384:403-413.
A17. Poordad F, McCone J Jr, Bacon BR, et al. Boceprevir for untreated chronic HCV genotype 1 infection. *N Engl J Med*. 2011;364:1195-1206.
A18. Sarrazin C, Schwendy S, Moller B, et al. Improved responses to pegylated interferon alfa-2b and ribavirin by individualizing treatment for 24-72 weeks. *Gastroenterology*. 2011;141:1656-1664.
A19. Corey L, Wald A, Patel R, et al. Once-daily valacyclovir to reduce the risk of transmission of genital herpes. *N Engl J Med*. 2004;350:11-20.
A20. Ison MG, Hui DS, Clezy K, et al. A clinical trial of intravenous peramivir compared with oral oseltamivir for the treatment of seasonal influenza in hospitalized adults. *Antivir Ther*. 2013;18:651-661.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Halegoua-De Marzio D, Hann HW. Then and now: the progress in hepatitis B treatment over the past 20 years. *World J Gastroenterol*. 2014;20:401-413.
2. Marcellin P, Gane E, Buti M, et al. Regression of cirrhosis during treatment with tenofovir disoproxil fumarate for chronic hepatitis B: a 5-year open-label follow-up study. *Lancet*. 2013;381:468-475.
3. EASL clinical practice guidelines: management of hepatitis C virus infection. *J Hepatol*. 2014;60:392-420.
4. AASLD, IDSA, IAS–USA. Recommendations for testing, managing, and treating hepatitis C. http://www.hcvguidelines.org; Accessed January 21, 2015.
5. Pawlotsky JM. New hepatitis C therapies: the toolbox, strategies, and challenges. *Gastroenterology*. 2014;146:1176-1192.
6. Naggie S, Sulkowski MS. Management of patients coinfected with HCV and HIV: a close look at the role for direct-acting antivirals. *Gastroenterology*. 2012;142:1324-1334 e1323.
7. Muthuri SG, Myles PR, Venkatesan S, et al. Impact of neuraminidase inhibitor treatment on outcomes of public health importance during the 2009-2010 influenza A(H1N1) pandemic: a systematic review and meta-analysis in hospitalized patients. *J Infect Dis*. 2013;207:553-563.
8. Santesso N, Hsu J, Mustafa R, et al. Antivirals for influenza: a summary of a systematic review and meta-analysis of observational studies. *Influenza Other Respir Viruses*. 2013;7(suppl 2):76-81.
9. Viasus D, Pano-Pardo JR, Pachon J, et al. Timing of oseltamivir administration and outcomes in hospitalized adults with pandemic 2009 influenza A(H1N1) virus infection. *Chest*. 2011;140:1025-1032.

REVIEW QUESTIONS

1. A 47-year-old man who has sex with men recently was diagnosed with HIV (HIV RNA = 22,400 copies and CD4 = 70 cells/mL) and hepatitis B (hepatitis B surface antigen [HBsAg] positive). He was started on tenofovir, emtricitabine, and efavirenz with a good response (HIV viral load < 50 and CD4 = 220). Two months later, he now has developed several days of severe nausea and vomiting and is unable to keep his medications down. The patient recalls your counseling about the risks of developing HIV resistance with intermittent doses and stops all his medications. The nausea and vomiting resolve over the subsequent few days.

He is between jobs and has travel planned. Since he now feels well with the good control of his HIV, he decides to remain on a drug holiday. Seven weeks later he develops malaise and general weakness. Biochemical testing of his blood now reveals the following: total bilirubin, 7.7 mg/dL; alanine aminotransferase (ALT), 2157 U/L; aspartate aminotransferase (AST), 2693 U/L; and alkaline phosphatase (Alk-P), 137 U/L. Which of the following is most likely to be responsible for his current clinical condition?

A. HIV
B. Hepatitis B
C. Hepatitis D
D. Syphilis
E. Herbal medications

Answer: B The patient's HIV regimen of tenofovir, emtricitabine, and efavirenz is also an effective treatment for hepatitis B. The withdrawal of medications during the drug holiday will lead to active HIV replication. Although moderate increases in AST and ALT can occur, fulminant hepatitis as described in this case is unusual. Severe acute exacerbations of hepatitis, including fatalities, have been reported in hepatitis B virus (HBV)-infected patients who stop anti-hepatitis B therapy. Hepatitis D, syphilis, or herbal medication can all cause an acute severe hepatitis but would be much less common than reactivated hepatitis B.

2. A 48-year-old man has HIV (CD4 560 and RNA 14,000) and hepatitis B (ALT 5 times the upper limit of normal and HBV DNA 35,000 IU/mL). Treatment for both HIV and HBV is recommended. However, the patient is concerned about drug toxicities. Because of his high CD4 count and relatively low HIV viral load, he does not want to begin treatment for HIV but is willing to be treated for HBV. Which of the following would be the most appropriate treatment?

A. Tenofovir
B. Entecavir
C. Ribavirin
D. PEG-interferon alfa-2a
E. Emtricitabine

Answer: D With the elevated ALT and HBV DNA, treatment is needed for this patient's hepatitis B. Treatment may be initiated with any approved antiviral medications, but tenofovir, entecavir, or PEG-interferon alfa-2a are the generally preferred agents. Emtricitabine, lamivudine, and tenofovir have activity against both HIV and HBV, and monotherapy (for HBV) with these agents would likely lead to HIV resistance. Therefore, the most appropriate treatment is PEG-interferon. Ribavirin may be used with interferon for the treatment of hepatitis C, but it has no role in the treatment of hepatitis B.

3. A 44-year-old heterosexual man with chronic hepatitis C and asthma (on inhaled corticosteroids) has been treated with telaprevir, interferon, and ribavirin for 7 weeks with good virologic response (hepatitis C virus [HCV] RNA level is > 1000 IU/mL at week 4). He routinely engages in high-risk sexual activity and now presents with 3 days of fever and fatigue as well as 1 day of pain with swallowing. On examination, he is afebrile, he has no cutaneous rashes, but examination of his mouth shows patchy desquamation of his tongue. Which of the following is the most appropriate recommendation?

A. This could be acute HIV, so an HIV viral load should be sent.
B. This could be a severe drug reaction, so the telaprevir, interferon, and ribavirin should be stopped.
C. This could be severe leukocytoclastic vasculitis from hepatitis C, so the patient should receive intravenous solumedrol.
D. This could be mucocutaneous candidiasis, so the patient should receive fluconazole.
E. This could be secondary syphilis, so the patient should receive benzathine penicillin.

Answer: B The patient is at risk for HIV and secondary syphilis, and both can both be associated with pharyngitis, myalgia, and a mucocutaneous rash. However, isolated oral involvement with frank desquamation would be very unusual for these diseases. Inhaled corticosteroids can cause oropharyngeal candidiasis, which would present with pharyngitis, dysphagia, and white mucosal plaques on examination, but desquamation of the mucosal membrane would be unusual. Hepatitis C is associated with leukocytoclastic vasculitis, usually with cutaneous findings. Telaprevir has been associated with both mild and severe dermatologic manifestations. Mild rashes can be treated with withdrawal of telaprevir. Desquamation of the tongue would suggest a more severe dermatologic reaction such as the Stevens-Johnson syndrome, which can present with mucosal desquamation before any cutaneous eruption. In this case, stopping telaprevir, interferon, and ribavirin would be the appropriate recommendation.

4. A 61-year-old man with severe chronic obstructive pulmonary disease (COPD) presents with a 36-hour history of fever, cough, myalgia, and mild shortness of breath. Rapid antigen assay from a throat swab is positive for influenza B. His current medications include tiotropium, prednisone 10 mg daily, and albuterol as needed. When advised of the diagnosis of influenza B, he asks for the least expensive medication because he recently lost his health insurance. Which of the following is the most appropriate treatment?

A. Amantadine
B. Rimantadine
C. Oseltamivir
D. Zanamivir
E. No treatment is required

Answer: C Oseltamivir, which is active against influenza B and is well tolerated, is the appropriate treatment. The patient has influenza B. Because of his history of severe COPD, his influenza B should be treated. Amantadine and rimantadine are both available as generic medications, but all strains of influenza B are resistant to amantadine and rimantadine. Inhaled zanamivir powder can cause bronchospasm in patients with underlying asthma or COPD and would not be an appropriate treatment in this patient.

361 THE COMMON COLD

RONALD B. TURNER

DEFINITION

The common cold is an upper respiratory syndrome of rhinorrhea and nasal obstruction, frequently accompanied by sore throat, sneezing, and cough. This viral syndrome is among the most common illnesses of humankind.

The Pathogens

The rhinoviruses cause at least half of all common cold illnesses. Rhinoviruses, which are RNA viruses that infect the respiratory epithelium, have long been known as common cold viruses, but they also are important causes of exacerbations of chronic bronchitis and asthma. More recent data suggest that rhinoviruses may also be associated with bronchiolitis in young children.

The coronaviruses (Chapter 366), parainfluenza viruses (Chapter 363), respiratory syncytial virus (RSV [Chapter 362]), metapneumoviruses[1], adenoviruses (Chapter 365), and influenza viruses (Chapter 364) may also cause common cold illnesses (Table 361-1). Bocavirus has recently been associated with the common cold, but these viruses are also frequently isolated from healthy control subjects, so their role as pathogens is uncertain. Bacterial pathogens such as *Bordetella pertussis* (Chapter 313) and group A streptococcus (Chapter 290) are occasionally associated with rhinorrhea, but these illnesses are generally readily distinguished from the common cold.

EPIDEMIOLOGY

The incidence of common colds decreases with age, from at least six episodes per year in young children to approximately two episodes per year in adults. The incidence of illness is higher in adults who have occupational or household exposure to children and in children who are cared for in childcare centers. Common cold illnesses occur year-round in temperate climates but have a substantially increased incidence between the early autumn and late spring. This common cold "season" consists of sequential outbreaks caused by the different respiratory viruses. In tropical climates, colds occur year-round without defined seasonality.

TABLE 361-1 VIRUSES ASSOCIATED WITH THE COMMON COLD

VIRUS GROUP	ANTIGENIC TYPES	PERCENTAGE OF CASES
Rhinoviruses	>100	40-50
Coronaviruses	5	10-15
Parainfluenza virus	5	5
Respiratory syncytial virus	2	5
Influenza virus	3	25-30
Adenovirus	47	5-10
Metapneumovirus	2	5
Other viruses: enteroviruses, bocavirus		

From Mandell GL, Bennett JE, Dolin R, eds. *Principles and Practice of Infectious Diseases.* 7th ed. Philadelphia: Churchill Livingstone; 2010:810.

PATHOBIOLOGY

Respiratory pathogens are spread from person to person by direct contact with either infected individuals or contaminated objects in the environment, by large-particle aerosols, or by small-particle aerosols. The rhinoviruses may be spread by direct contact, but recent data suggest a role for other mechanisms.[2] RSV (Chapter 362) may be spread by either direct contact or large-particle aerosols, and influenza (Chapter 364) may be spread by small-particle aerosols.

Regardless of the route of spread, the common cold syndrome is initiated by infection of the nasal epithelium. Influenza and adenovirus produce obvious damage to the respiratory epithelium. Rhinovirus and RSV, in contrast, have little or no detectable impact on the epithelium. Regardless of the histopathology, all these viruses stimulate a nonspecific host inflammatory response that appears to be responsible for many of the symptoms associated with the common cold.

The nasal obstruction of the common cold appears to result primarily from increased nasal blood flow and pooling of blood in the capacitance vessels of the nose. The increase in nasal secretion associated with the common cold may also contribute to the nasal obstruction. Rhinorrhea is primarily a result of increased vascular permeability, with leakage of serum into the nasal secretions. Increased mucus production contributes to the secretions during the later stages of the illness.

Multiple factors may play a role in the pathogenesis of cough. Cough may be related to infection of the lower airway, irritation of upper airway receptors with neurologically mediated airway reactivity, or postnasal drip with pharyngeal irritation.

The risk of infection after exposure to the respiratory viruses is primarily dependent on the presence of specific neutralizing antibodies. Antibody responses to the rhinoviruses, adenoviruses, and influenza viruses are protective against subsequent infection. The frequency of infection with these viruses is a result of the large number of distinct serotypes of rhinovirus and adenovirus and the ability of the influenza viruses to behave as though there are multiple virus serotypes by virtue of the rapid mutation of the antigens presented on the surface of the virus. The parainfluenza viruses, RSV, and metapneumoviruses do not produce protective immunity, so reinfection is common, although preexisting antibody moderates the severity of illness.

Mannose-binding lectin deficiency has been associated with an increased incidence of common colds in young children. Protection by this innate response may become less important as children experience a variety of infections and develop specific immunity. Polymorphisms that enhance inflammatory cytokine responses may be associated with more severe respiratory illness.

CLINICAL MANIFESTATIONS

The incubation of common cold illness is generally short, ranging from 2 to 8 days, although the adenoviruses may have an incubation of as long as 13 days. A sore or scratchy throat is frequently reported as the first manifestation. Sneezing is also a common early symptom. Nasal obstruction and rhinorrhea develop rapidly and, by day 2 or 3 after the onset of illness, are the most

bothersome symptoms. Cough generally develops later in the illness and frequently is the most bothersome symptom as the cold resolves. Common cold illnesses generally persist for about 1 week, although about 25% may persist for as long as 2 weeks.

Physical findings are restricted to the upper respiratory tract. Increased nasal secretion may be obvious to the examiner. A change in the color or consistency of nasal secretions is common during the course of the illness and is not indicative of sinusitis or bacterial superinfection.

DIAGNOSIS

The differential diagnosis of the common cold includes noninfectious disorders as well as other upper respiratory tract infections. Allergic rhinitis (Chapter 251) has a symptom complex similar to that of the common cold, although the presence of nasal or conjunctival itching suggests allergic disease. Most patients can reliably differentiate these illnesses.

Sinus involvement is present in uncomplicated cold illnesses, and superimposed bacterial sinusitis (Chapter 426) is difficult to differentiate from an uncomplicated cold.[3] Rhinorrhea that persists without improvement for more than 10 days may suggest the presence of bacterial sinusitis that will respond to antibiotics.

Routine laboratory studies are not helpful for the diagnosis or management of the common cold. Although the viral pathogens associated with the common cold may be detected by culture, antigen detection, polymerase chain reaction, or serologic methods, these studies are of little value unless treatment with an antiviral agent is contemplated.

TREATMENT

Rx

Specific antiviral therapy is generally not useful for the treatment of common cold illnesses. The neuraminidase inhibitors oseltamivir and zanamivir (Chapter 360) have a modest effect on influenza virus infections, but the difficulty of distinguishing influenza from other common cold pathogens and the need to start treatment early in the illness for maximum benefit are practical limitations to the use of these agents for mild upper respiratory infections. Antibacterial therapy is of no benefit in the treatment of the common cold.

Management of the common cold relies on symptomatic remedies. For example, a combination of acetaminophen, chlorpheniramine, and phenylephrine (an adrenergic agonist) is safe and effective in treating the symptoms of the common cold.[A1] However, it is preferable to use specific therapies targeted just to an individual adult's specific symptoms of nasal obstruction, rhinorrhea, or sore throat so as to avoid the side effects of unnecessary medications.

Nasal Congestion

Both topical and oral adrenergic agents (Table 361-2) are effective nasal decongestants. Although direct comparison has not been performed in the common cold, it is generally accepted that topical over-the-counter agents such as intranasal xylometazoline are more effective than oral drugs for nasal congestion. Prolonged use of the topical adrenergic agents should be avoided to prevent development of an apparent rebound effect when the drug is discontinued. Systemic absorption of oxymetazoline and xylometazoline has rarely been associated with bradycardia, hypotension, and coma. The systemic side effects of the oral adrenergic agents are central nervous system stimulation, hypertension, and palpitations. The antihistamines have no effect on nasal congestion.

TABLE 361-2 TREATMENTS FOR THE COMMON COLD

DRUG	DOSE AND DURATION	SIDE EFFECTS
Topical adrenergic agents: oxymetazoline	2-3 sprays of 0.05% solution every 12 hr as needed for up to 3 days	Rebound nasal congestion with prolonged use Nasal stinging or burning
Oral adrenergic agents: pseudoephedrine	60 mg every 4-6 hr up to 240 mg/day as needed for nasal congestion	Insomnia, agitation
Antihistamines: chlorpheniramine	4 mg orally every 4-6 hr up to 24 mg/day as needed for rhinorrhea or sneezing	Sedation
Anticholinergics: ipratropium bromide	2 sprays per nostril of 0.06% solution every 6-8 hr as needed for rhinorrhea	Nasal dryness

Rhinorrhea

The treatment of rhinorrhea is primarily by blockade of cholinergic stimulation of glandular secretion. Intranasal ipratropium bromide reduces rhinorrhea in colds by 22 to 31% compared with placebo.[A1] The most common side effects of intranasal ipratropium are nasal irritation and bleeding.

The first-generation (sedating) antihistamines reduce rhinorrhea by approximately 25% compared with placebo.[A1] These observations, the absence of histamine in the secretions of most subjects with colds, and the similarity of the response to ipratropium and first-generation antihistamines suggest that any effect of antihistamines on rhinorrhea is related to their anticholinergic rather than their antihistaminic properties. The major side effects associated with the use of the antihistamines are sedation and drying of the eyes, mouth, and nose.

Cough

Cough during colds is produced by several different mechanisms, and treatment should be directed at the most likely underlying cause (Chapters 83 and 429). If cough is caused by nasal obstruction or postnasal drip, it may respond to treatment with an antihistamine or antihistamine-decongestant combination.[A1] If a more persistent cough is the result of virus-induced reactive airway disease or viral infection of the lower airways, patients may benefit from bronchodilator therapy (Chapter 87). Cough that persists after the resolution of other cold symptoms or that persists in association with unremitting rhinorrhea may be due to sinusitis and may respond to antibiotic therapy (Chapter 426). Nonspecific cough suppression with codeine or dextromethorphan hydrobromide has not been demonstrated to be efficacious, and expectorants such as guaifenesin are not effective antitussive agents.

Pain-Related symptoms

Nonsteroidal anti-inflammatory agents are effective for sore throat, headache, ear pain, and myalgia but do not relieve nasal congestion, rhinorrhea, or cough.[A1]

Other Remedies

Zinc is an inhibitor of rhinovirus 3C protease, which is essential for virus replication. However, zinc does not have a significant antiviral effect at doses that can be used for treatment and has shown either no or relatively modest effects in reducing the severity of the common cold.[A1] This uncertain benefit must also be viewed in light of zinc's side effects; oral zinc lozenges may be associated with sore mouth and occasional nausea, whereas intranasal zinc may cause nasal irritation. Echinacea has not been shown to be beneficial in reducing symptoms of the common cold in several studies.[A1] Given the variation in echinacea products, it is possible that echinacea preparations with different phytochemical profiles might be helpful. The accumulating evidence, however, suggests that it is prudent to assume that echinacea has no beneficial effect until positive evidence of a treatment effect is produced. Antibiotics are not useful, even in patients with a purulent nasal discharge, and may cause adverse effects.[A1]

PREVENTION

Chemoprophylaxis or immunoprophylaxis is generally not available for the common cold. Immunization or chemoprophylaxis against influenza (Chapter 364) may be useful for prevention of colds caused by this pathogen, but influenza is responsible for only a small proportion of all colds. Vitamin C, even in megadoses, is of no benefit. Other nonpharmacologic interventions touted as effective prophylaxis for the common cold but of unproven benefit include zinc, vitamin E, echinacea, ginseng, exercise, and handwashing. Handwashing and exercise have undeniable benefits for health in general and can be recommended despite the paucity of evidence specific to common cold prevention. The other interventions, although probably safe, have no demonstrable benefit and simply contribute to the unnecessary health care expenditures related to the common cold.

PROGNOSIS

The common cold generally has little medical significance. However, these illnesses are frequently complicated by otitis media (Chapter 426) or sinusitis (Chapter 426) that may be a direct result of the viral infection or may be due to bacterial superinfection. Exacerbations of asthma (Chapter 87) and chronic bronchitis (Chapter 88) are also important complications of the common cold.

Grade A References

A1. De Sutter AI, van Driel ML, Kumar AA, et al. Oral antihistamine-decongestant-analgesic combinations for the common cold. *Cochrane Database Syst Rev*. 2012;2:CD004976.

A2. Picon PD, Costa MB, da Veiga Picon R, et al. Symptomatic treatment of the common cold with a fixed-dose combination of paracetamol, chlorpheniramine and phenylephrine: a randomized, placebo-controlled trial. *BMC Infect Dis*. 2013;13:556.
A3. AlBalawi ZH, Othman SS, Alfaleh K. Intranasal ipratropium bromide for the common cold. *Cochrane Database Syst Rev*. 2013;6:CD008231.
A4. Sutter AI, Lemiengre M, Campbell H, et al. Antihistamines for the common cold. *Cochrane Database Syst Rev*. 2003;3:CD001267.
A5. Smith SM, Schroeder K, Fahey T. Over-the-counter (OTC) medications for acute cough in children and adults in community settings. *Cochrane Database Syst Rev*. 2014;11:CD001831.
A6. Kim SY, Chang YJ, Cho HM, et al. Non-steroidal anti-inflammatory drugs for the common cold. *Cochrane Database Syst Rev*. 2013;6:CD006362.
A7. Science M, Johnstone J, Roth DE, et al. Zinc for the treatment of the common cold: a systematic review and meta-analysis of randomized controlled trials. *CMAJ*. 2012;184:E551-E561.
A8. Barrett B, Brown R, Rakel D, et al. Echinacea for treating the common cold: a randomized trial. *Ann Intern Med*. 2010;153:769-777.
A9. Kenealy T, Arroll B. Antibiotics for the common cold and acute purulent rhinitis. *Cochrane Database Syst Rev*. 2013;6:CD000247.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Panda S, Mohakud NK, Pena L, et al. Human metapneumovirus: review of an important respiratory pathogen. *Int J Infect Dis*. 2014;25:45-52.
2. Turner RB, Fuls JL, Rodgers ND. Effectiveness of hand sanitizers with and without organic acids for removal of rhinovirus from hands. *Antimicrob Agents Chemother*. 2010;54:1363-1364.
3. Chow AW, Benninger MS, Brook I, et al. IDSA clinical practice guideline for acute bacterial rhinosinusitis in children and adults. *Clin Infect Dis*. 2012;54:e72-e112.

362

RESPIRATORY SYNCYTIAL VIRUS

EDWARD E. WALSH

DEFINITION

Respiratory syncytial virus (RSV), which causes yearly winter outbreaks in temperate climates, is the most important cause of bronchiolitis and pneumonia in young infants, a common cause of illness in older children and young adults, and can be severe in elderly persons, adults with underlying cardiopulmonary disease, and the severely immunocompromised.[1]

The Pathogen

RSV is a single-stranded RNA-enveloped virus of the family Paramyxoviridae, genus *Pneumovirus*; it is related to human metapneumovirus. Two major transmembrane glycoproteins (G, attachment protein; F, fusion protein) carry neutralizing epitopes, and two nonstructural proteins (NS1 and NS2) block the antiviral activity of type I interferons, whereas a secreted form of G bearing a CX3C chemokine motif may modulate immune responses. Two major virus groups (A and B), each with multiple genotypes, are distinguishable.

EPIDEMIOLOGY

In the United States, epidemics begin in the south in late fall, move steadily north, and peak in February and March in colder climates.[2] In tropical areas, RSV may occur throughout the year, with peaks during the rainy season. RSV annually causes approximately 100,000 hospitalizations and accounts for 60% of bronchiolitis and 25% of pneumonia cases in infants. Mortality is rare in the United States (<400 deaths annually), but deaths are substantially greater in underdeveloped countries.[3] About half of infants become infected in their first winter and all by age 2. RSV is transmitted principally by direct contact with large-particle fomites of respiratory secretions rather than by small-particle aerosolization.

Between 1 and 3% of primary infections result in hospitalization, but lower socioeconomic status, crowding, underlying prematurity, congenital cardiac abnormalities, bronchopulmonary dysplasia, and immunosuppression are each associated with increased risk for serious disease. Severe disease is also associated with specific polymorphisms in the promoter regions of cytokine genes. Hospitalization is most frequent between the ages of 1 and 6 months, peaking at 2 months of age with a rate of 25.9 per 1000 children.[4] However, the majority of hospitalized infants are normal healthy infants without identifiable risk factors. The overall burden of RSV in infants 0 to 6 months of age is 132 office visits, 55 emergency room visits, and 17 hospitalizations per 1000. Reinfection occurs frequently throughout life, although subsequent illness is less severe and hospitalization infrequent, except for persons with underlying cardiac or pulmonary conditions.

Though often not considered in adults, RSV infection is common and may be severe in the elderly.[5,6] In a British study, RSV accounted for 17% of medically attended outpatient respiratory illnesses in persons older than 45 years, and a recent study from The Netherlands calculated that RSV-associated mortality among persons age 65 and older is about 90% as high as influenza A and greater than double that of influenza B.[7] In two recent U.S. studies, RSV infection was implicated in 6 to 10% of hospitalizations for acute pulmonary symptoms in the winter among community-dwelling elderly persons, numbers that are similar to those for influenza.[8,9]

PATHOBIOLOGY

Virus most commonly enters through the nose or eye and then spreads from the upper to the lower respiratory tract. Pathologic findings include a lymphocytic peribronchiolar infiltration with edema, obstruction, and necrosis. Bronchiolitis with multiple areas of atelectasis, and pneumonia with interstitial infiltration of mononuclear cells, as well as alveoli filled by edema and necrosis, develop in infected patients.

CLINICAL MANIFESTATIONS

Infants experience upper respiratory symptoms of conjunctival injection, mucopurulent nasal discharge, cough, and low-grade fever after an incubation period of 2 to 8 days. Otitis media (Chapter 426) is often associated with secondary bacterial infection. After several days, lower respiratory tract symptoms appear in 25 to 50% of infants, with cough, wheezing, tachypnea, and use of accessory muscles as the disease progresses. Expiratory wheezes, rhonchi, and fine rales are the most common findings on lung examination. Sudden apnea may develop in the youngest infants. Hyperinflation and diffuse interstitial pneumonitis are the most frequent radiographic findings. High-titer virus shedding lasts 7 to 10 days, although immunocompromised infants may excrete virus for a month or longer, even when asymptomatic. Coinfection with other respiratory viruses occurs in up to 30% of patients, but it usually is neither clinically discernible nor definitively associated with more severe illness.

Adults with RSV typically begin with upper respiratory symptoms, but many patients have lower respiratory symptoms, especially wheezing. Low-titer virus shedding often persists for 10 days or longer. In elderly persons, RSV attack rates are 3 to 5% annually; wheezing is more common than with influenza (Chapter 364), whereas fever is less common. In frail elderly persons or in patients with underlying chronic obstructive pulmonary disease or heart failure, severe disease can develop. Attack rates in nosocomial nursing home outbreaks average 10 to 15%, with crackles and wheezes evident in a third of patients and radiographically confirmed pneumonia in approximately 10%. The precise incidence of bacterial superinfection in hospitalized adults is unknown but can range from 15 to 40%, similar to what is seen with influenza.

RSV infection has been documented in up to 10% of bone marrow transplant recipients (Chapter 178), patients with acute leukemia (Chapter 183), and heart/lung transplant recipients (Chapters 82 and 101) during the winter months.[10] Upper respiratory symptoms lead to lower respiratory tract symptoms in about 30% of patients.

DIAGNOSIS

In infants, a presumptive diagnosis is suggested by typical symptoms during the epidemic season, but the diagnosis often is not considered in adults. Diagnosis can be made by reverse transcription–polymerase chain reaction (RT-PCR), which is the test of choice and has high sensitivity and specificity. Viral culture, which can take up to 10 days, has a sensitivity of only about 75%.

In adults, viral culture has poor sensitivity (30%), whereas RT-PCR detects infection in three quarters of culture-negative cases that are seropositive, using acute and convalescent serum. Owing to very poor sensitivity, antigen detection by immunofluorescence and enzyme immunoassay are not useful in adults, even if immunocompromised. In immunocompromised patients, chest radiographs demonstrate diffuse interstitial and alveolar infiltrates. A useful clinical clue to the presence of RSV is the frequent presence of radiographically proven sinusitis. Upper respiratory tract symptoms distinguish this illness from cytomegalovirus pneumonia (Chapter 376).

TREATMENT

Rx

Therapy for most infants is symptomatic and generally is limited to hydration and supplemental oxygen.[11] Bronchodilators have not been demonstrated to be effective, although they are occasionally used. Similarly,

glucocorticosteroids have not been shown to have benefit in most studies. Inhaled ribavirin (20-mg/mL solution administered by aerosol for 6 hours three times daily for 3 to 5 days, or 60 mg/mL for 2 hours three times daily) may be beneficial for infants at high risk for serious disease and those who are severely ill, although clinical trails have not demonstrated benefit for most infants.

No placebo-controlled studies of the effect of inhaled ribavirin have been conducted in adults with severe RSV disease, but the evidence suggests benefit in immunocompromised adults with RSV pneumonia (2 g at 60 mg/mL for 5 to 10 days), especially if begun before development of lower respiratory symptoms.[12]

Some studies using antibody therapy (the monoclonal antibody palivizumab, 15 mg/kg once, or polyclonal immunoglobulin) combined with inhaled ribavirin suggest benefit in the treatment of RSV pneumonia in immunosuppressed adults, although conclusive randomized studies have not been performed.[13] A new oral RSV-entry inhibitor has shown promise in reducing the viral load.[14] Bacterial superinfection can develop, with *Streptococcus pneumoniae* (Chapter 289) and *Haemophilus influenzae* (Chapter 300) being the most frequent organisms. For such patients, appropriate antimicrobial treatment (Chapter 97) is mandatory.

PREVENTION

Adherence to standard infection-control principles (e.g., gloves, gowns, frequent handwashing) substantially reduces nosocomial spread. A vaccine is not available. Parenteral humanized RSV monoclonal antibody (palivizumab, 15 mg/kg monthly during the RSV season) has demonstrated benefit when administered prophylactically to specific high-risk infant groups.■ Palivizumab prophylaxis in healthy late-preterm infants (32 to 35 weeks' gestation) also can reduce subsequent wheezing by 50% during the first year of life.■

PROGNOSIS

RSV infection in severely immunocompromised adults such as bone marrow transplant recipients (Chapter 178) and those with acute leukemia (Chapter 183) undergoing cytotoxic chemotherapy can carry a 60% mortality rate when pneumonia develops.[15] Lymphopenia (<100/μL) and high-dose total body irradiation are associated with more severe disease, but progression to pneumonia is very unusual in patients with an absolute lymphocyte count above 1000/μL. In lung transplant recipients, bronchiolitis obliterans syndrome may occur in 10 to 50% of patients as a late sequela of RSV infection.

Mortality is rare in otherwise healthy infants but can reach 37% in infants with cardiac disorders. In children, a link between severe bronchiolitis and subsequent asthma remains unresolved. In normal infants and healthy adults, the outcome is generally good.

Grade A References

A1. Andabaka T, Nickerson JW, Rojas-Reyes MX, et al. Monoclonal antibody for reducing the risk of respiratory syncytial virus infection in children. *Cochrane Database Syst Rev*. 2013;4:CD006602.

A2. Blanken MO, Rovers MM, Molenaar JM, et al. Respiratory syncytial virus and recurrent wheeze in healthy preterm infants. *N Engl J Med*. 2013;368:1791-1799.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Meng J, Stobart CC, Hotard AL, et al. An overview of respiratory syncytial virus. *PLoS Pathog*. 2014;10:e1004016.
2. Tang JW, Loh TP. Correlations between climate factors and incidence–a contributor to RSV seasonality. *Rev Med Virol*. 2014;24:15-34.
3. Nair H, Nokes DJ, Gessner BD, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. *Lancet*. 2010;375:1545-1555.
4. Hall CB, Weinberg GA, Blumkin AK, et al. Respiratory syncytial virus-associated hospitalizations among children less than 24 months of age. *Pediatrics*. 2013;132:e341-e348.
5. Widmer K, Zhu Y, Williams JV, et al. Rates of hospitalizations for respiratory syncytial virus, human metapneumovirus, and influenza virus in older adults. *J Infect Dis*. 2012;206:56-62.
6. Volling C, Hassan K, Mazzulli T, et al. Respiratory syncytial virus infection-associated hospitalization in adults: a retrospective cohort study. *BMC Infect Dis*. 2014;14:665.
7. van Asten L, van den Wijngaard C, van Pelt W, et al. Mortality attributable to 9 common infections: significant effect of influenza A, respiratory syncytial virus, influenza B, norovirus, and parainfluenza in elderly persons. *J Infect Dis*. 2012;206:628-639.
8. Falsey AR, Becker KL, Swinburne AJ, et al. Bacterial complications of respiratory tract viral illness: a comprehensive evaluation. *J Infect Dis*. 2013;208:432-441.
9. Seo S, Campbell AP, Xie H, et al. Outcome of respiratory syncytial virus lower respiratory tract disease in hematopoietic cell transplant recipients receiving aerosolized ribavirin: significance of stem cell source and oxygen requirement. *Biol Blood Marrow Transplant*. 2013;19:589-596.
10. Kim YJ, Guthrie KA, Waghmare A, et al. Respiratory syncytial virus in hematopoietic cell transplant recipients: factors determining progression to lower respiratory tract disease. *J Infect Dis*. 2014;209:1195-1204.
11. Chu HY, Englund JA. Respiratory syncytial virus disease: prevention and treatment. *Curr Top Microbiol Immunol*. 2013;372:235-258.
12. Shah DP, Ghantoji SS, Shah JN, et al. Impact of aerosolized ribavirin on mortality in 280 allogeneic haematopoietic stem cell transplant recipients with respiratory syncytial virus infections. *J Antimicrob Chemother*. 2013;68:1872-1880.
13. Bawage SS, Tiwari PM, Pillai S, et al. Recent advances in diagnosis, prevention, and treatment of human respiratory syncytial virus. *Adv Virol*. 2013;2013:595768.
14. DeVincenzo JP, Whitley RJ, Mackman RL, et al. Oral GS-5806 activity in a respiratory syncytial virus challenge study. *N Engl J Med*. 2014;371:711-722.
15. Shah DP, Ghantoji SS, Ariza-Heredia EJ, et al. Immunodeficiency scoring index to predict poor outcomes in hematopoietic cell transplant recipients with RSV infections. *Blood*. 2014;123:3263-3268.

REVIEW QUESTIONS

1. A 78-year-old woman with a history of chronic obstructive pulmonary disease (COPD) and heart failure presents to her physician 5 days after the New Year complaining of 3 days of nasal congestion, worsening cough, increased sputum, and increasing dyspnea. She has a temperature of 38.8° C and diffuse inspiratory and expiratory wheezing. SaO_2 on 2 L/min nasal O_2 is 90%. She received influenza vaccine 2 months ago but states that over the holidays she was with several family members, including young children, with the "flu." Which is the optimal diagnostic test to identify respiratory syncytial virus (RSV) infection?

A. Serology demonstrating a single high titer of serum immunoglobulin (Ig)G to RSV
B. A rapid RSV antigen detection assay on a nasal swab or nasal wash specimen
C. Viral culture of nasal swab or nasal wash specimen
D. Reverse transcription–polymerase chain reaction (RT-PCR) of nasal swab or nasal wash specimen

Answer: D Adults with RSV have relatively high levels of serum and mucosal antibody and subsequently shed relatively low titers of virus compared with infants. Rapid antigen tests and culture are significantly less sensitive than RT-PCR. Although serology is highly sensitive in adults, both acute and convalescent samples are required, so this an impractical diagnostic method for clinical use.

2. A 42-year-old man with acute myeloid leukemia received an allogeneic stem cell transplant from his brother 20 days ago and now presents with nasal congestion and rhinorrhea. His white blood cell (WBC) count is 400/μL with a lymphocyte count of 50/μL. A nasal swab sample is positive for RSV by multiplex PCR. His lung examination is normal, and his chest radiograph is clear. SaO_2 is 96% while breathing ambient air. Which one of the following statements is most accurate regarding this patient?

A. Therapeutic intervention is never indicated for upper respiratory tract symptoms associated with RSV infection in such a patient.
B. Palivizumab (an RSV neutralizing monoclonal antibody) infusion will clear the virus from the upper respiratory tract and hasten recovery.
C. Progression of RSV infection to lower respiratory tract involvement is uncommon unless the lymphocyte count is less than 50/μL.
D. In nonrandomized studies, administration of inhaled ribavirin is associated with a decreased risk of progression to lower respiratory tract disease
E. The use of intravenous IgG infusion in combination with inhaled ribavirin has been proven to provide optimal outcomes.

Answer: D RSV infection in hematopoietic cell transplants recipients can be severe. Disease will progress to lower respiratory involvement in about 25% of such patients and carries a 20 to 40% mortality. Factors associated with progression include total body irradiation and low absolute lymphocyte counts (<100/μL). In one analysis, RSV did not progress in patients with lymphocyte counts above 1000/μL. Adequately sized studies of palivizumab treatment, given alone or in combination with ribavirin, have not been performed. An analysis of 288 immunocompromised subjects from one large center found that treatment with inhaled ribavirin was associated with decreased risk of disease progression.

3. A 56-year-old man with COPD on 2 liters/min of home O_2 presents in February with a 3-day illness characterized by rhinitis, sore throat, increase in his daily cough, and change in color and amount of sputum production. He is afebrile and has a respiratory rate of 26 and pulse of 110. He is breathless and is wheezing on examination. His WBC count is 6700 with a normal differential, his chest radiograph is unchanged from prior films, and his SaO_2 is 78% on 2 L/min O_2, which is significantly below his baseline of 92%. He is admitted to the hospital and cautiously placed on 4 liters of nasal O_2. A nasal swab is positive for RSV by RT-PCR, and his sputum Gram stain shows more than 25 WBCs per high-power field, with mixed flora. In addition to standard hand hygiene, which of the following infection-prevention measures should be implemented?

A. Use of gowns and gloves for all patient contact
B. Isolation of the patient in a negative pressure room
C. Use of gloves, gowns, and facemask for all contact
D. None of the above
E. All of the above

Answer: A RSV can persist on hard surfaces for several hours, so gowns and gloves should be used to prevent transmission. RSV is transmitted by large fomites, not by small-particle aerosol, so neither negative pressure nor masks are necessary to prevent transmission to health care workers or other patients.

4. A 24-year-old medical student is seen in the student health service for nasal congestion, mild sore throat, and wheezing in late July. She has a history of mild untreated asthma since childhood. A rapid antigen test for influenza A is negative, and a rapid antigen test for RSV is weakly positive. On examination, her lungs show expiratory wheezes but are otherwise clear. Three days earlier she returned from rural Thailand, where she had worked for 2 months in a family medicine clinic. Which of the following statements is most correct:

A. The rapid antigen test for RSV is reliable and confirms RSV as the cause of her symptoms.
B. RSV infection is uncommon in July, so the positive antigen test should be considered a false positive.
C. RSV is the most common cause of asthma exacerbations in young adults, so the positive antigen test should be considered definitive.
D. To confirm the diagnosis of RSV, an RT-PCR assay should be performed.

Answer: D The most reliable diagnostic assay for RSV in adults is RT-PCR. In adults, the rapid antigen test for RSV has poor sensitivity and cannot be considered definitive. July is atypical for RSV in the United States, but the student had recently traveled to a tropical area in which RSV can circulate throughout the year. Although RSV can cause exacerbations of asthma, rhinovirus is a much more common precipitant of asthma in older children and adults.

363

PARAINFLUENZA VIRAL DISEASE

GEOFFREY A. WEINBERG AND KATHRYN M. EDWARDS

DEFINITION

Human parainfluenza viruses (hPIVs) are important causes of a wide spectrum of respiratory illnesses. In infants and young children, they produce acute upper and lower respiratory tract infections ranging from the common cold (Chapter 361) and otitis media (Chapter 426) to severe croup, bronchiolitis, and pneumonia (Chapter 97). In older children and adults, hPIV infections are usually limited to the upper respiratory tract (Chapter 96), although immunocompromised individuals may develop fatal respiratory failure.

The Pathogen

The hPIVs are enveloped, single-stranded, nonsegmented RNA viruses that belong to two genera in the family Paramyxoviridae. Members of this family also include respiratory syncytial virus (RSV [Chapter 362]), human metapneumovirus (Chapter 361), measles virus (Chapter 367), mumps virus (Chapter 369), Hendra and Nipah viruses, and the pathogenic animal viruses of Newcastle disease, canine distemper, and rinderpest. The hPIV genome encodes six structural proteins. The hemagglutinin-neuraminidase (HN) and fusion (F) proteins, which are exposed on the bilayered lipid envelope surrounding the helical nucleocapsid–RNA complex, mediate both attachment to host sialic acid–containing glycoproteins and penetration of the virus into susceptible mammalian cells.[1] These proteins have retained their antigenic stability for many years, unlike the "drift and shift" of the hemagglutinin and neuraminidase of influenza viruses. The four serotypes of hPIV are called types 1 to 4, including two subgroups (A and B) of type 4 virus.

The hPIVs replicate in the ciliated epithelial cells that line the respiratory tract on its luminal surface. This selective tropism is consistent with the absence of invasive disease or viremia in the immunocompetent host. Syncytium formation, which is noted in viral cell cultures and in the lungs of immunocompromised patients with severe pneumonia, is not thought to be important in typical infections of previously healthy individuals.

EPIDEMIOLOGY

The hPIVs are ubiquitous, with a worldwide geographic distribution. Their transmission is principally by large-particle fomites via close person-to-person contact. Parainfluenza virus activity displays both endemic and epidemic patterns, with each serotype favoring different age groups and distinct clinical syndromes but with enough overlap to preclude specific diagnosis based solely on clinical and epidemiologic grounds.

Primary infection with hPIV occurs early in childhood.[2,3] Of the parainfluenza viruses, type 3 (hPIV-3) generally infects infants first, such that 50 to 67% of infants demonstrate serologic evidence of infection by 1 year of age. Parainfluenza viruses 1 and 2 (hPIV-1 and hPIV-2, respectively) then most commonly infect children between 2 and 5 years of age. Parainfluenza virus 4 (hPIV-4) less commonly causes symptomatic respiratory infections; in some studies hPIV-4 is found more often in viral coinfection with other pathogens.[4]

Until the early 1960s, hPIV-1 caused endemic annual disease in the United States. For the past several decades, however, hPIV-1 has been associated with biennial outbreaks in the fall of odd-numbered years. Infections with hPIV-2 largely follow this same curious pattern but occur much less commonly than hPIV-1 infections. Infections with hPIV-3 have remained endemic throughout the year, with peaks in the late spring. The incubation period for all serotypes of hPIV is between 3 and 6 days in experimentally infected adults, but natural infections in children have incubation periods of 2 to 4 days.

The hPIVs are second only to RSV as the cause of acute upper and lower respiratory tract infections in young children in the United States. In a population-based study of children younger than 5 years hospitalized with fever or acute respiratory tract infection, 7% of children had laboratory-confirmed hPIV infection (by cell culture and molecular amplification techniques), compared with 19% with RSV and 6% with influenza A or B virus infections. The hospitalization rate for hPIV infection in children younger than 5 years was 1.02 per 1000 children per year. Extrapolating to the entire U.S. population, these data suggest that approximately 23,000 hPIV annual hospitalizations occur in children younger than 5 years. The rates of emergency department and outpatient health care visits attributable to hPIV infections in young children are 10- to 50-fold greater than hospitalization rates. Further, immunity to hPIV infection is neither complete nor durable, so older children and adults can exhibit symptomatic infection, sometimes leading to office visits and hospitalization.

CLINICAL MANIFESTATIONS

Primary Infection

Illness associated with primary hPIV infection varies by age and viral serotype. Underlying medical conditions such as cardiopulmonary compromise and immune disorders increase the severity of disease. Large cohort and longitudinal family studies have shown that hPIV infections are responsible

for about 65% of croup, 20 to 40% of lower respiratory infection, and 20% of upper respiratory infection in young children. The majority of children have experienced primary hPIV infection by the time they enter elementary school. Thus, although hPIV also causes upper respiratory infections in adults, these are reinfections rather than primary infections (see later). In general, hPIV-1 and hPIV-2 are most commonly associated with croup, whereas hPIV-3 infection often presents as undifferentiated febrile illness, bronchiolitis, or pneumonia.

Infection typically begins with upper respiratory signs and symptoms, notably coryza, rhinorrhea, pharyngitis without cervical adenopathy, and low-grade fever. Symptoms, which typically persist for 3 to 5 days, may be unpredictable and can result in sudden respiratory failure. About 15 to 25% of infected children develop signs of croup or progress to lower respiratory tract disease (e.g., bronchiolitis, pneumonia) indistinguishable from RSV infection.

Croup is characterized by a raspy barking cough with notable inspiratory stridor, dyspnea, and respiratory distress. These symptoms, which are generally spasmodic, result from subglottic inflammation and edema. On occasion, severe stridor may develop and make differentiation from epiglottitis caused by *Haemophilus influenzae* type b (Chapter 300) difficult, although epiglottitis is much less common in the United States since universal Hib vaccination became routine in 1990.

Croup is very rare in adults but has been associated with hPIV in isolated case reports. Epiglottitis is seen more often in adults, but it is more commonly associated with staphylococcal or streptococcal infection. Epiglottitis is a medical emergency for patients of all ages and must be distinguished from croup to allow proper therapy, including emergent intubation of the airway (Chapter 429). For patients who are judged to have a stable enough airway to permit urgent radiography rather than emergent tracheostomy, the differences between these two entities are clearly shown in lateral neck radiographs, on which subglottic edema and narrowing are seen with croup, and swelling of the epiglottis is seen with epiglottitis (Figs. 363-1 and 363-2). Children with bronchiolitis or pneumonia caused by hPIV have cough, rales, and wheezing associated with hypoxia, and their chest radiographs often exhibit air trapping.

Reinfection

Reinfection with hPIV is less severe and typically causes rhinorrhea in normal children and adults. Reinfection with hPIV is estimated to account for 1 to 15% of all acute respiratory illnesses in adults, the majority of whom present with simple upper respiratory tract infections.[5] However, as with RSV, some adults may develop severe disease that requires hospitalization and even ventilatory support. In such patients, fever, cough, dyspnea, and wheezing are common. Radiographic changes, primarily lobar or interstitial infiltrates, are seen in more than 50% of adults hospitalized with hPIV infection.

Elderly and Immunocompromised Patients

Nursing home outbreaks of hPIV infection may include a high incidence of pneumonia, just as hPIV infections often cause severe pneumonia in immunocompromised children and adults. Among healthy elderly residents of long-term care facilities, hPIV infections are as common as infections caused by influenza A virus, influenza B virus, or RSV. Ten percent of children and 2 to 7% of adults with leukemia or hematopoietic stem cell transplant recipients develop hPIV infections, with about 90% of those due to hPIV-3.[6-8] Although 80 to 90% of hPIV infections among patients with malignancy are community acquired, nosocomial outbreaks have occurred even in stem cell transplant units.

Approximately 25% of hPIV-infected stem cell transplant recipients develop lower respiratory disease, mostly in the first 100 days after transplantation, when lymphopenia and neutropenia are most pronounced. The strongest identified risk factors for hPIV infection in this setting are neutropenia, lymphopenia, more severe disease, and the use of corticosteroids, especially at higher doses. Fever, cough, dyspnea, and sputum production are the most common symptoms. Coinfection with *Aspergillus* (Chapter 339) and other pathogens has been reported in hematopoietic stem cell transplant recipients with pneumonia as well.

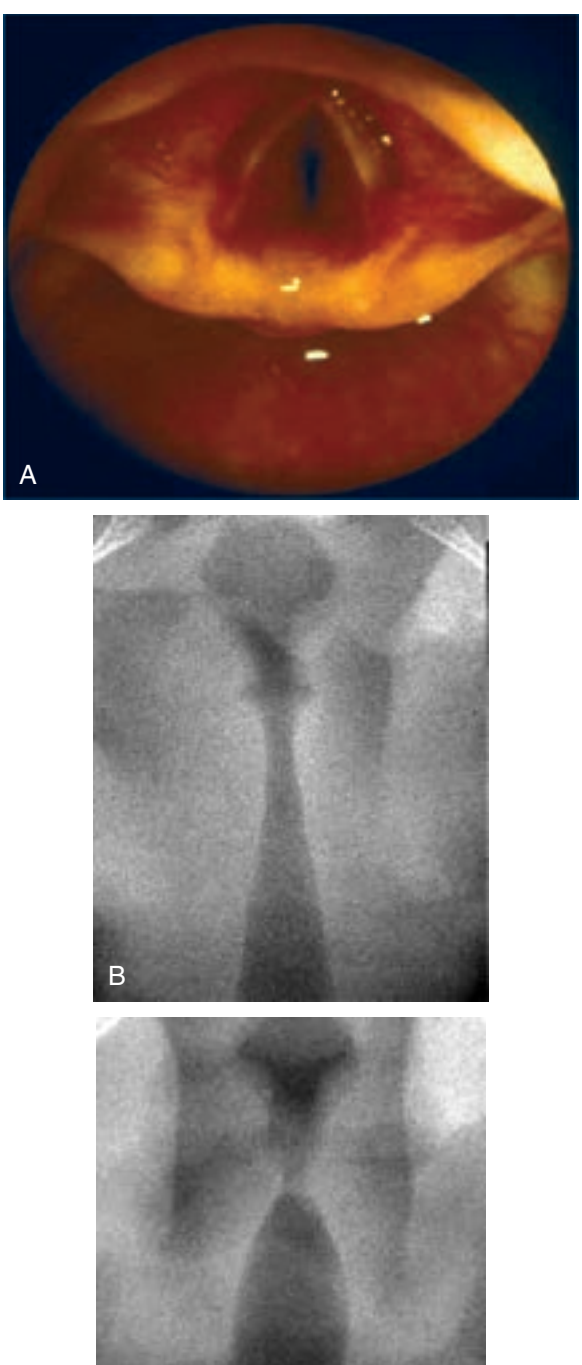

FIGURE 363-1. Subglottic edema. A, Endoscopic view of subglottic edema in viral croup. **B**, Radiologic presentation of subglottic edema in viral croup, causing narrowing ("steeple sign") of the tracheal air shadow, compared with (**C**) a normal tracheal air shadow. (From Hammer J. Acquired upper airway obstruction. *Paediatr Respir Rev.* 2004;5:25-33.)

DIAGNOSIS

Although hPIV may be suspected on clinical and epidemiologic grounds, specific diagnosis requires isolation of the virus or detection of viral antigen or RNA in respiratory secretions. Reverse transcription–polymerase chain reaction (RT-PCR) assays are highly sensitive and specific for diagnosis. Alternatively, cultures of monkey kidney or human embryonic kidney cells can be used to grow the virus, with cytopathic effects detected in 5 to 10 days (except for hPIV-4, which requires up to 3 weeks). Rapid direct or indirect immunofluorescence tests are less sensitive than culture but still are used in many clinical laboratories.

Definitive diagnosis of parainfluenza infection in adults, even by RT-PCR, may be more difficult than in children, presumably because less virus is shed in partially immune adults. However, virus can usually be recovered from the

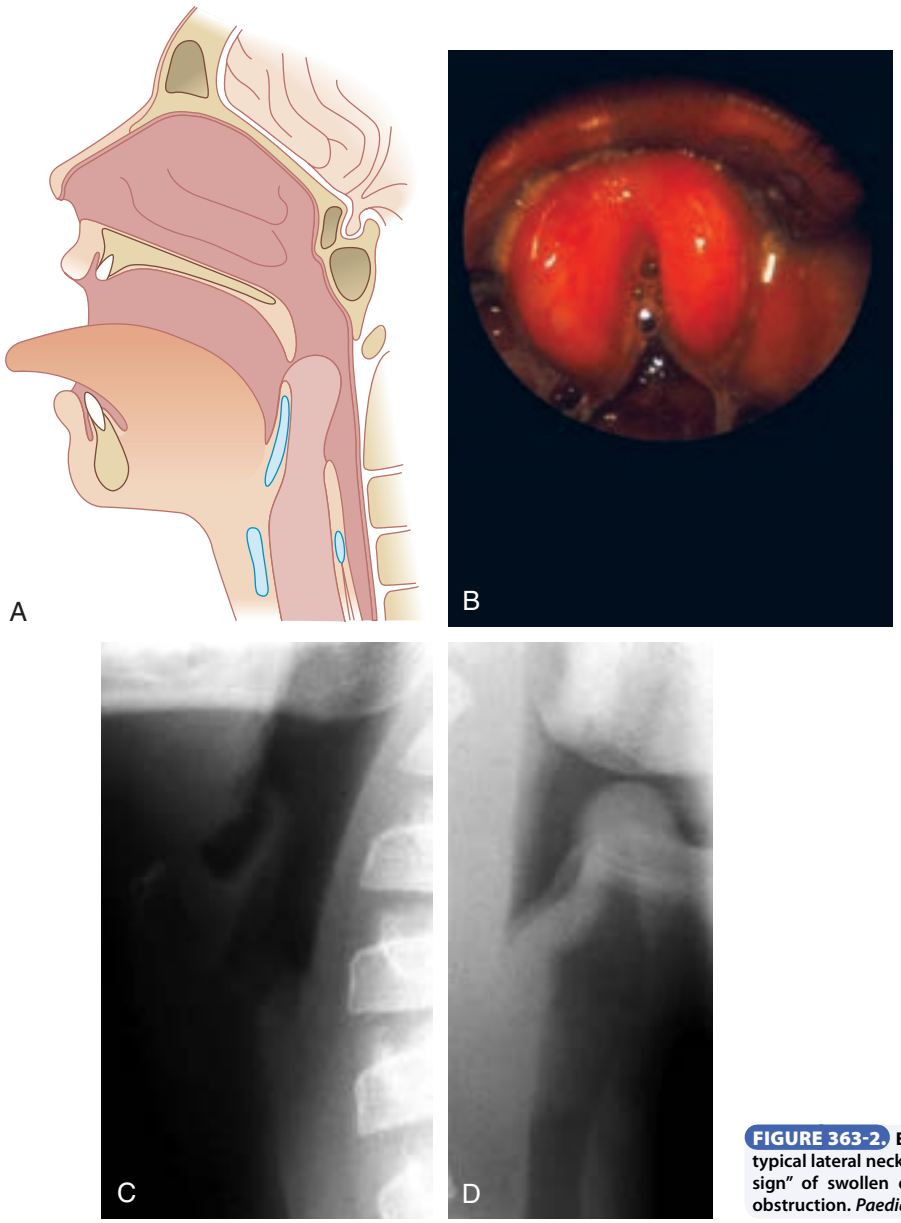

FIGURE 363-2. Epiglottitis. Schematic (**A**) and endoscopic view (**B**) of epiglottitis. **C**, The typical lateral neck radiographs of a normal child and (**D**) a child with epiglottitis ("thumb sign" of swollen epiglottis) are displayed. (From Hammer J. Acquired upper airway obstruction. *Paediatr Respir Rev.* 2004;5:25-33.)

nasopharyngeal secretions or bronchoalveolar lavage fluid in hematopoietic stem cell transplant recipients with pneumonia.

TREATMENT

Specific antiviral treatment of hPIV infection is currently unavailable. Aerosolized ribavirin has in vitro activity against hPIV and is approved for use in RSV infection. Although it has shown few benefits in immunocompromised children and adults with severe hPIV pneumonia, even when administered with concomitant intravenous immunoglobulin, some experts will offer aerosolized or intravenous ribavirin, with or without intravenous immunoglobulin, to immunocompromised patients with severe disease.

For previously healthy children with croup, general comfort, which usually means sitting in the lap of a parent or caregiver, is widely recommended.[9] Humidified air (mist) repeatedly has been shown to have no benefits and actually is associated with adverse effects, including anxiety, difficulty with cardiorespiratory monitoring, and bacterial and fungal contamination of both hot and cold mist humidifiers. Oxygen may be given by holding the end of the tubing near the nose and mouth ("blow-by" oxygen).

Children with mild to moderate croup (i.e., without stridor or significant chest wall indrawing at rest, or with stridor and indrawing but without agitation) may be given oral dexamethasone, 0.6 mg/kg body weight, and observed; if they improve, they may be discharged to home. Children with severe croup (stridor, chest wall indrawing, and agitation) may benefit from blow-by oxygen, inhaled racemic epinephrine (0.05 mL/kg body weight of a 2.25% solution of racemic epinephrine for nebulization, up to a maximum dose of 0.5 mL) or nebulized L-epinephrine (0.5 mL/kg of 1:1000 L-epinephrine, up to a maximum dose of 5 mL), and concomitant dexamethasone. Children should be admitted to the hospital if no significant clinical improvement is seen after several hours of observation and therapy.

Oral dexamethasone is as effective as intramuscular dexamethasone or nebulized budesonide and easier to administer. Corticosteroids decrease the need for hospitalization and return visits, decrease length of stay both in emergency departments and after hospital admission, and decrease the need for intubation. In general, there are no reasons to institute either antibiotic therapy or short-acting bronchodilator therapy in children with croup.

Investigational therapies for hPIV infection include both nonspecific and specific measures. Inhalation of heliox (helium-oxygen mixture) may decrease the work of breathing and improve gas exchange in children with moderate to severe croup, but adequately controlled clinical trial data are not yet available.[10] The novel antiviral agent DAS181, a sialidase fusion protein with activity against both influenza and parainfluenza viruses, has shown beneficial effects in a few adults with severe hPIV infection and is undergoing clinical trials.[11]

PREVENTION

Clinical trials of live attenuated hPIV vaccines are ongoing. Two approaches have been taken to produce live-attenuated intranasal hPIV-3 vaccines. One approach has been to attenuate an hPIV-3 isolate by repeated cold-temperature passage and manipulation by recombinant technology. A second has been to use a related but nonvirulent bovine PIV-3 as a backbone in which to insert

hPIV-3 genes to produce an attenuated chimeric vaccine virus.[12] Another vaccine candidate under study uses the chimeric bovine-human PIV-3 virus, but with an additional RSV gene inserted into the vaccine virus genome in an attempt to prevent both of these common infections in young children.[13]

PROGNOSIS

In otherwise healthy children, mortality after hPIV infection is less than 0.1%. By comparison, mortality after hPIV pneumonia in patients undergoing stem cell transplantation is as high as 10 to 30%.[14]

Grade A References

A1. Russell KF, Liang Y, O'Gorman K, et al. Glucocorticoids for croup. *Cochrane Database Syst Rev*. 2011;1:CD001955.

A2. Bjornson C, Russell K, Vandermeer B, et al. Nebulized epinephrine for croup in children. *Cochrane Database Syst Rev*. 2013;10:CD006619.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Schomacker H, Schaap-Nutt A, Collins PL, et al. Pathogenesis of acute respiratory illness caused by human parainfluenza viruses. *Curr Opin Virol*. 2012;2:294-299.
2. Morgan OW, Chittaganpitch M, Clague B, et al. Hospitalization due to human parainfluenza virus-associated lower respiratory tract illness in rural Thailand. *Influenza Other Respir Viruses*. 2013; 7:280-285.
3. Weinberg GA, Hall CB, Iwane MK, et al. Parainfluenza virus infection of young children: estimates of the population-based burden of hospitalization. *J Pediatr*. 2009;154:694-699.
4. Frost HM, Robinson CC, Dominguez SR. Epidemiology and clinical presentation of parainfluenza type 4 in children: a 3-year comparative study to parainfluenza types 1-3. *J Infect Dis*. 2014;209: 695-702.
5. Schmidt AC, Schaap-Nutt A, Bartlett EJ, et al. Progress in the development of human parainfluenza virus vaccines. *Expert Rev Respir Med*. 2011;5:515-526.
6. Srinivasan A, Wang C, Yang J, et al. Parainfluenza virus infections in children with hematologic malignancies. *Pediatr Infect Dis J*. 2011;30:855-859.
7. Chemaly RF, Hanmod SS, Rathod DB, et al. The characteristics and outcomes of parainfluenza virus infections in 200 patients with leukemia or recipients of hematopoietic stem cell transplantation. *Blood*. 2012;119:2738-2745.
8. Hirsch HH, Martino R, Ward KN, et al. Fourth European Conference on Infections in Leukaemia (ECIL-4): guidelines for diagnosis and treatment of human respiratory syncytial virus, parainfluenza virus, metapneumovirus, rhinovirus, and coronavirus. *Clin Infect Dis*. 2013;56:258-266.
9. Bjornson CL, Johnson DW. Croup in children. *CMAJ*. 2013;185:1317-1323.
10. Moraa I, Sturman N, McGuire T, et al. Heliox for croup in children. *Cochrane Database Syst Rev*. 2013;12:CD006822.
11. Chalkias S, Mackenzie MR, Gay C, et al. DAS181 treatment of hematopoietic stem cell transplant patients with parainfluenza virus lung disease requiring mechanical ventilation. *Transpl Infect Dis*. 2014;16:141-144.
12. Englund JA, Karron RA, Cunningham CK, et al. Safety and infectivity of two doses of live-attenuated recombinant cold-passaged human parainfluenza type 3 virus vaccine rHPIV3cp45 in HPIV3-seronegative young children. *Vaccine*. 2013;31:5706-5712.
13. Bernstein DI, Malkin E, Abughali N, et al. Phase 1 study of the safety and immunogenicity of a live, attenuated respiratory syncytial virus and parainfluenza virus type 3 vaccine in seronegative children. *Pediatr Infect Dis J*. 2012;31:109-114.
14. Seo S, Xie H, Karron RA, et al. Parainfluenza virus type 3 Ab in allogeneic hematopoietic cell transplant recipients: factors influencing post-transplant Ab titers and associated outcomes. *Bone Marrow Transplant*. 2014;49:1205-1211.

REVIEW QUESTIONS

1. A 2-year-old girl presents to the emergency department with stridor and dyspnea of 4 hours' duration. Her parents relate that she has had about 3 days of preceding low-grade fever, coryza, and rhinorrhea. When she woke up this morning, her parents heard a barking cough coming from the child's room and noted that her breathing was progressively noisier on the way to the hospital. On physical examination, the child is not toxic appearing, but she is uncomfortable, with obvious inspiratory stridor and a raspy barking cough. Which represents the most likely diagnosis and most appropriate management plan?

A. The child likely has community-acquired bacterial pneumonia; obtain a chest radiograph and administer oral antibiotics.
B. The child likely has community-acquired viral pneumonia; obtain a chest radiograph but do not administer oral antibiotics.
C. The child likely has epiglottitis; obtain a lateral neck radiograph, begin intravenous antibiotics, then secure the airway.
D. The child likely has croup; keep her quiet by allowing her to stay in mother's lap, provide blow-by oxygen, administer oral dexamethasone, and observe for improvement.
E. The child likely has croup; perform a careful pharyngeal examination in an attempt to visualize the epiglottis, provide oxygen by rebreather mask, and administer intramuscular dexamethasone and subcutaneous epinephrine.

Answer: D This child likely has mild to moderate croup, based on the history and physical examination findings of stridor but her nontoxic appearance. The best course is to keep her quiet in a comfortable setting such as her parent's lap, give her oxygen in the least intrusive manner (blow-by), and give oral dexamethasone 0.6 mg/kg. Oral dexamethasone can decrease symptoms, shorten the stay in the emergency department or hospital, and reduce the need for intubation. Stridor and barking cough are generally not associated with community-acquired bacterial pneumonia. Viral pneumonia with hPIV can accompany croup; but if so, dexamethasone and oxygen also would be appropriate. Epiglottitis should always be considered when examining a young child with respiratory distress, but these children tend to be more toxic in appearance; ideally the airway should be secured before performing radiographs, intravenous lines, and other procedures that might upset the child and provoke airway collapse.

2. The grandfather of the child in Question 1 underwent bone marrow transplantation for malignancy 4 weeks before the child became ill. She visited with him extensively at home a few days before she herself became ill. Now, 4 days after that contact, the grandfather is reporting fever, cough, dyspnea, and increased sputum production. His wife is concerned because he has not "looked this bad" since the transplantation was performed. Which statement best characterizes his illness?

A. He likely has hPIV pneumonia, having acquired the infection from his granddaughter; he should be admitted to the hospital and treated with aerosolized or intravenous ribavirin and intravenous immunoglobulin in view of his high risk for mortality.
B. He likely has acquired hPIV infection from his granddaughter but because adult airways are larger, he is unlikely to exhibit symptomatic croup; reassurance and continued monitoring at home will be sufficient.
C. He likely has community-acquired pneumonia and should be given antibiotics.
D. He likely has *Pneumocystis pneumonia* and should be hospitalized for intravenous trimethoprim-sulfamethoxazole. Further diagnostic studies are unnecessary and would be not be cost-effective.
E. None of the above

Answer: A This adult is likely significantly immunocompromised (first 100 days after bone marrow transplantation) and has a 10 to 30% of mortality with hPIV infection. The incubation period of hPIV and the timing of the illness in the child are consistent with transmission. Although the efficacy of ribavirin and intravenous immunoglobulin have not been proven in randomized clinical trials, their use should be considered because of the severity of his illness and his high risk of mortality. Hospitalization is likely warranted, with oxygen and supportive care given at minimum. Although adults with hPIV infection do not generally present with croup, this immunocompromised patient's condition warrants an urgent medical visit, likely with subsequent hospitalization. Many other causes of respiratory infection are possible in the first 100 days following bone marrow transplantation, but administering antibiotics without considering the contact history or obtaining further diagnostic specimens would be inappropriate.

364 INFLUENZA

FREDERICK G. HAYDEN

DEFINITION

Influenza is an acute febrile respiratory viral illness that usually occurs in annual outbreaks of varying severity and occasionally occurs in worldwide epidemics (pandemics) and as sporadic zoonotic infections. Influenza viruses infect the respiratory tract, are highly contagious, and typically produce prominent systemic symptoms early in the illness. Influenza infection causes various clinical syndromes in adults, including nonfebrile common colds (Chapter 361), pharyngitis (Chapters 290 and 429), tracheobronchitis (Chapter 96), pneumonia (Chapter 97), and a range of nonrespiratory complications. Conversely, infections with other respiratory viruses, such as respiratory syncytial virus (RSV [Chapter 362]) or adenovirus (Chapter 365), may produce influenza-like illness. Influenza A viruses have caused five pandemics of varying severity within the past 120 years (Table 364-1). The pandemic in 1918-1919 caused at least 500,000 deaths in the United States and more than 40 million worldwide, whereas the 2009 H1N1 pandemic was associated with substantially less mortality. Seasonal epidemics may cause enormous morbidity, economic loss, and often substantial mortality. Recent epidemics have caused on average more than 24,000 respiratory- and circulatory-related deaths and more than 200,000 hospitalizations in the United States alone.

TABLE 364-1 ANTIGENIC SUBTYPES OF INFLUENZA A VIRUS ASSOCIATED WITH PANDEMIC AND PANDEMIC-LIKE INFLUENZA

YEAR	INTERVAL (YR)	SUBTYPE DESIGNATION	EXTENT OF ANTIGENIC CHANGE IN INDICATED SURFACE PROTEIN*	SEVERITY OF PANDEMIC (MORTALITY)
1889	?42-59	H3N?	H+++N?	Severe
1918	18	H1N1†	H+++N+++	Very severe
1957	39	H2N2	H+++N+++	Severe
1968	11	H3N2	H+++N–	Moderate‡
1977	9	H1N1	H+++N+++	Negligible§
2009	32	H1N1	H++N++	Mild-moderate§

*Compared with antecedent or cocirculating virus.
†Formerly designated H0N1 (swine virus prototype) or Hsw1N1.
‡The population had some immunity to N2 neuraminidase.
§The older population was largely immune because of previous infection with earlier circulating, antigenically identical (1977) or related (2009) viruses; those born after 1957 were primarily affected. The impact of the 2009 pandemic virus, based on estimated years of life lost, was comparable to that observed in the 1968 pandemic in the United States.
+ = Minor change; ++ = moderate change; +++ = major change; – = no change.

The Pathogen

Influenza viruses belong to the family Orthomyxoviridae and are divided into three types (A, B, and C) based on their protein composition and other properties (Table 364-2). The virion (Fig. 364-1) is a medium-sized enveloped pleomorphic particle covered with two types of surface glycoprotein spikes, the trimeric hemagglutinin (H or HA) and the tetrameric mushroom-shaped neuraminidase (N or NA). The envelope is composed of a lipid bilayer overlying the matrix (M1) protein that surrounds the viral genome, which consists of eight segments of single-stranded negative sense RNA in influenza A and B viruses. Influenza C viruses have seven segments and only a single surface glycoprotein. Genomic replication occurs in the nucleus of infected cells, and multiple cellular proteins and pathways are involved during the infection of host cells.

Whereas influenza B and C viruses are principally human pathogens, influenza A viruses primarily infect aquatic birds and sometimes establish lineages that circulate in other animal hosts, including other avians, swine, horses, marine mammals, and dogs. Influenza A viruses are further classified into subtypes on the basis of their HA and NA glycoproteins. Sixteen HA and nine NA subtypes are currently recognized in avians, and two others are recognized in bats. Only three HAs (H1, H2, and H3) and two NAs (N1 and N2) have been documented thus far in epidemic and pandemic human influenza A viruses, although other HAs (e.g., H5, H6, H7, H9, H10) and NAs (e.g., N4, N7, N8, N9) have been found in zoonotic infections. Each strain is identified by type, subtype if influenza A, site, sample number, and year of isolation.

EPIDEMIOLOGY

Antigenic and Genetic Variation

Influenza viruses are unique among the respiratory viruses with regard to their extent of genetic and antigenic variation, epidemic behavior, and frequent association with excess mortality during community outbreaks—characteristics that exemplify influenza virus's dependence on efficient host-to-host transmission for its survival. The changing antigenicity of the surface glycoproteins largely accounts for repeated influenza epidemics. Antibody to HA can neutralize viral infectivity and is thus the major determinant of immunity. Current vaccines are based largely on inducing hemagglutination-inhibition (HAI) or neutralizing antibodies to HA. Anti-NA antibody limits viral replication and probably reduces the severity of infection. Variation involves either relatively minor (antigenic drift) or major (antigenic shift) changes in antigenicity. Significant antigenic variation is much less frequent with influenza B or C than with influenza A.

Antigenic drift results from point mutations in the HA gene segment that cause amino acid substitutions in at least one of five key antigenic sites on HA. Drift can also occur in NA and in T-cell epitopes on internal proteins. Antigenic variants emerge frequently (every year or every few years) within an influenza A or B virus. For example, the original H3N2 variant, A/Aichi/68, has undergone successive drifts resulting in epidemic strains that include the recent circulation of A/Switzerland/9715293/2013-like-viruses. Immunologic selection favors transmission of the new variant over the old because of the less frequent presence of antibody to the new virus in the population.

Antigenic shift results from the appearance of a novel influenza A virus with HA or HA and NA glycoproteins that are new to humans or that reappear after decades of absence. Because of the lack of population immunity, a new strain that is transmissible from person to person can cause pandemic disease (see Table 364-1). The origin of new pandemic strains and the basis for their possible recirculation remain incompletely understood. Avian influenza viruses have served as the reservoir of new genes for human pandemic viruses. Reassortment of gene segments may occur when two influenza viruses simultaneously infect a single cell, and reassortment events in which human viruses acquired avian genes led to both the 1957 and 1968 pandemic viruses. Because swine can support replication of both human and avian viruses, they have been postulated to serve as a mixing vessel for the generation of new strains or as the host in which avian viruses can adapt to mammals. The 2009 H1N1 pandemic virus arose as a quadruple reassortant that derived

TABLE 364-2 INFLUENZA A VIRUS PROTEINS

DESIGNATION	LOCATION (APPROXIMATE SIZE)	FUNCTION	OTHER
Hemagglutinin (HA)	Surface, transmembrane (566 aa)	Cell attachment via sialic acid–bearing receptors and penetration; fusion of host cell and viral membranes	Type-, subtype-, and strain-specific antigens; key antigen in inactivated vaccines
Neuraminidase (NA)	Surface, transmembrane (454 aa)	Virus release and spread; enzymatic activity causes removal of sialic acid residues from receptors	Type-, subtype-, and strain-specific antigens; activity linked to risk for pneumonia; site of action of neuraminidase inhibitors
M1 or matrix	Internal (252 aa)	Major structural envelope protein; virus assembly	Type-specific antigen; conserved T-cell epitopes
M2	Surface, transmembrane (97 aa)	Ion channel; virus uncoating and budding	Influenza A only; site of action of amantadine/rimantadine; ectodomain as possible vaccine candidate
Nucleoprotein (NP)	Internal (498 aa)	Major ribonucleoprotein complex component; associated with RNA and polymerase proteins	Type-specific antigen; conserved T-cell epitopes
Polymerase proteins (PB1, PB2, PA)	Internal (PB1-757 aa, PB2-759 aa, PA-716 aa)	Viral RNA replication and mRNA transcription (endonuclease, cap-snatching, nucleotide addition)	Determinant of replication efficiency; specific PB2 mutations associated with mammalian adaptation; site of action of favipiravir
NS1	Nonstructural (230 aa)	Multifunctional; regulation of virus RNA and protein synthesis; host protein interactions	Interferon antagonist; inhibition of innate immune signaling
NEP (NS2)	Internal (121 aa)	Nuclear export factor	From splicing of NS1 mRNA
PB1-F2	Nonstructural (87 aa)	Proapoptotic factor; proinflammatory effects; interferon antagonist	Expressed from 1+ reading frame of PB1 of certain viruses
PA-X	Nonstructural (41-61 aa)	Endonuclease activity; repression of cellular gene expression	Modulation of the host response to infection; expressed by 1+ ribosomal frame-shifting

Modified from Krug RA, Fodor E. Chapter 4, The virus genome and its replication. In: Webster RG, Monto AS, Braciale TJ, Lamb RA, eds. *Textbook of Influenza*. 2nd ed. Oxford: John Wiley and Sons; 2013:57-66.
NOTE: Influenza B also has eight gene segments, one of which encodes both NA and NB proteins, the latter a membrane protein of uncertain function, and another encoding the M1 (matrix) and BM2 proteins, the latter functioning as an ion channel but not inhibited by adamantanes. Nonstructural proteins are present only in infected cells. Other small nonstructural proteins (PB1-N40) have been described in influenza A and are under study.

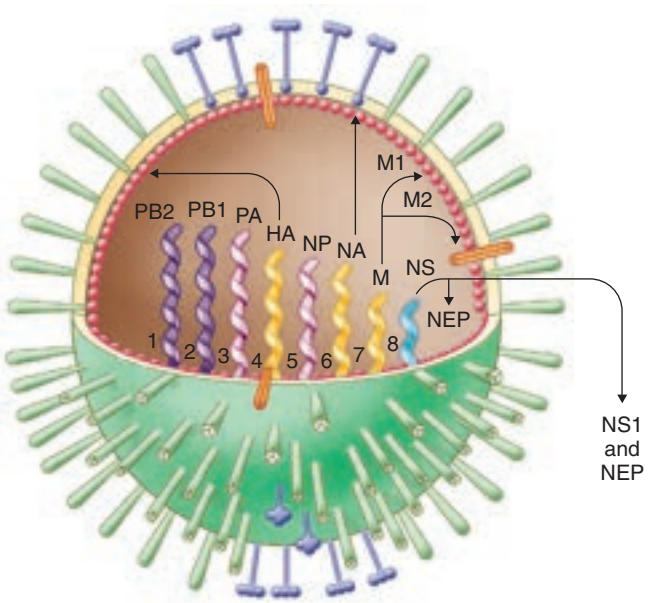

FIGURE 364-1. Diagram of influenza virus structure. Eight segments of viral RNA are contained within the lipid envelope and matrix (M1) shell. Each codes for one or more proteins that form the virus or regulate its intracellular replication. The presumed functions of each are listed in Table 364-2. (Courtesy Dr. Robert G. Webster.)

gene segments from Asian and North American swine lineage viruses that harbored genes derived from swine, avian, and human viruses.

Multiple reassortment events may occur over a period of years before a pandemic emerges. For example, the 1918 pandemic virus was probably a reassortant composed of both human and swine genes, including some of avian origin, that underwent adaptation in a mammalian host, perhaps human or swine, before causing the pandemic. Frequent intrasubtypic reassortment also occurs among seasonal human influenza A viruses and sometimes leads to new antigenic variants or altered virulence.

TABLE 364-3 AGE-SPECIFIC RATES FOR ILLNESS AND MORTALITY DURING URBAN INFLUENZA EPIDEMICS

AGE (YR)	PHYSICIAN VISITS PER 100	ACUTE RESPIRATORY DISEASE HOSPITALIZATIONS PER 10,000	PNEUMONIA- AND INFLUENZA-RELATED MORTALITY PER 100,000
<5	28	43	3
5-14	14	5	1
15-44	10	8	1
45-54	9	13	10
55-64	10	21	10
≥65	Not stated	73	104

Modified from Glezen WP. Anatomy of an urban influenza epidemic. In: Hannoun C, Kendal AP, Klenk HD, et al, eds. *Options for the Control of Influenza II*. Amsterdam: Elsevier; 1993:12.

Epidemic or Interpandemic Influenza

An epidemic is an outbreak of influenza confined to one geographic location. In temperate climates, community epidemics of influenza A virus infection often have a characteristic pattern, typically reaching a sharp peak in 2 or 3 weeks after initial recognition and persisting for 6 to 10 weeks. Increased numbers of schoolchildren with febrile respiratory illness are often the first indication of influenza in a community, soon followed by illnesses in adults and, 1 to 2 weeks later, by increased hospital admission of patients with influenza-related complications. Hospitalization rates in high-risk persons increase two- to five-fold during major epidemics (Table 364-3). School and employment absenteeism increases, as does mortality from pneumonia and underlying conditions, especially in older adults during A/H3N2 epidemics. Epidemics occur almost exclusively during the late autumn and winter months in temperate areas, but influenza activity may occur year-round in the tropics or display other patterns. The reasons for the distinct seasonality of influenza in temperate climates are uncertain but may include the school calendar, increased close indoor contact, and absolute humidity, which affects airborne transmissibility. In the United States, the onset of wintertime influenza has been associated with low absolute humidity levels during the

preceding weeks. Outbreaks sometimes occur in tour groups (land or ship) and in chronic care facilities during summer months, particularly after the appearance of a drift variant. Regional differences in the timing, magnitude, and causative viruses of influenza outbreaks are common. During epidemics, overall attack rates typically range from 5 to 20% in adults. Attack rates of 40 to 50% may occur in semiclosed populations, such as in hospitals and chronic care facilities, and in highly susceptible age groups such as children. Influenza A and B viruses, two different influenza A subtypes, two different influenza B lineages, or two different strains within a single subtype may cocirculate or occur sequentially during one season in a given location. In addition, simultaneous outbreaks of influenza A virus and RSV or other respiratory viruses occur. Strains circulating at the end of one season's epidemic are sometimes responsible for the next season's outbreak (the so-called herald wave phenomenon).

Pneumonia- and influenza-related deaths fluctuate annually, with peaks in the winter months in temperate climates. When such deaths exceed the expected threshold, the cause is typically influenza A, but influenza B virus or RSV can occasionally be responsible. Influenza A/H3N2-dominant seasons are associated with two to three times higher mortality rates than are H1N1 and B-dominant seasons. Although mortality is usually greatest during pandemics, substantial mortality occurs with epidemics. During seasonal influenza, more than 85% of pneumonia- and influenza-related deaths occur in persons aged 65 years and older. Mortality risk is especially high in those aged 85 years and older. Other cardiopulmonary and chronic diseases also result in increased mortality after influenza epidemics, so the overall influenza-associated mortality is two- to four-fold higher than pneumonia- and influenza-related deaths.

Pandemic Influenza

Pandemics of influenza A result from the emergence of a new virus capable of sustained person-to-person transmission and to which the population contains no or limited immunity. The virus spreads worldwide, often circulating outside of the usual influenza season, and usually infects persons of all ages. Preexisting immunity due to prior infection with antigenically related viruses can provide partial protection, as occurred in older adults during the 2009 H1N1 pandemic. Pandemics are associated with high morbidity rates, especially in children, and sometimes notably increased mortality rates in pregnant women and young and middle-aged adults. For example, more than 90% of deaths in the 1918 and 2009 pandemics occurred in persons younger than 65 years. In the 2009 pandemic, adults hospitalized with H1N1 were more likely to have severe pneumonia, shock, sepsis, and organ failure, and to require ICU care than were patients with seasonal influenza.

The interval between pandemics is variable (10 to 40 years) and unpredictable. The most severe pandemics have been associated with major antigenic alterations in both major surface antigens. Depending on population susceptibility and perhaps changes in the virus, one or more waves, sometimes with increased severity, may follow the initial one. As the level of immunity in the population increases, antigenic drift within the subtype may cause repeated epidemics in subsequent years, and excess mortality in persons younger than age 65 may continue for some years. For example, the pandemic 2009 H1N1 virus has continued to cause fatal infections in younger adults at least 4 years after its initial emergence.

Zoonotic Influenza

Zoonotic infections may be acquired from swine, poultry, or rarely other animals. Although most avian influenza viruses do not cause infections directly in humans, zoonotic infections due to avian H5, H7, H9, and rarely other subtypes continue to represent potential threats to global health. Initially recognized by a cluster of human cases in Hong Kong in 1997, an epizootic of avian H5N1 infections has affected poultry in many areas of Asia, the Middle East, Europe, and Africa and continues to cause sporadic human illnesses with high mortality and occasional instances of nonsustained human-to-human transmission. Since 2013, an outbreak of avian H7N9 infections in eastern China has been associated with case-fatality of approximately 30% in hospitalized patients.[1] Temporary closure of live bird markets and perhaps seasonal decreases in avian infections rapidly reduced the number of persons affected initially, but a second larger wave developed during late 2013. This outbreak has included limited family clusters, many milder illnesses in the community, and increasing geographic spread; its trajectory remains uncertain.

In the United States, infrequent zoonotic infections have been recognized for decades in people exposed to swine. Since 2012, reassortant swine H3N2 viruses that acquired the M gene and sometimes other genes from the pandemic 2009 H1N1 virus (designated variant H3N2, or H3N2v) have caused hundreds of sporadic infections, particularly in the context of agricultural fairs, sometimes followed by limited human-to-human transmission and serious infections. Other variant swine-origin viruses (H1N1v, H1N2v) have also caused zoonotic infections.

PATHOBIOLOGY

Influenza virus infection is transmitted from person to person by virus-containing respiratory secretions. Large-droplet and small-particle aerosols over short distances (1 to 2 meters) both appear to contribute,[2] but transmission by other routes, including hand contamination from secretion-laden fomites followed by self-inoculation into the eye or nose, may be possible. Infection by avian viruses can occur after direct contact with infected birds or their excreta, exposure to contaminated environments, ingestion of inadequately cooked food, and sometimes by inoculation into the conjunctiva.

The cellular receptor binding patterns and tissue tropism of influenza viruses are key determinants in transmissibility and pathogenesis. Efficient virus transmission between humans depends on virus attachment to and replication in cells bearing α-2,6-linked sialosaccharides in the upper respiratory tract and tracheobronchial tree. By comparison, the α-2,3-linked sialosaccharides, which are the preferred receptors for avian viruses, are concentrated on cells in the distal bronchioles, alveoli, and conjunctiva. Once the virus initiates infection of the respiratory tract epithelium, successive cycles of viral replication infect large numbers of cells and result in destruction of respiratory epithelium and sometimes pneumocytes through direct cytopathic effects or apoptosis. Virulence, which is a multigenic characteristic with contributions from HA, NA, NS, and polymerase proteins (see Table 364-2), varies widely among strains and is not necessarily linked to transmissibility.

The incubation period averages 2 days and varies from about 1 to 4 days for seasonal influenza but may be up to 1 week and possibly longer in infections caused by avian viruses. The quantity of virus in respiratory tract specimens generally correlates with the severity of illness and levels of host pro-inflammatory cytokine-chemokine responses—findings that support the importance of ongoing viral replication in producing acute illness. Rapid innate cellular responses are induced through Toll-like receptors and retinoic acid inducible gene I (*RIG-1*) that detect viral RNA and lead to the production of cytokines and interferons. Elevated levels of mediators such as interferon (IFN)-α, interleukin (IL)-6, and tumor necrosis factor (TNF)-α occur in blood and respiratory secretions and probably contribute to systemic symptoms and fever. Deficient IFN responses have been associated with severe influenza. The duration of viral replication depends on age, immune status, underlying conditions, viral strain, and assay method. In seasonal influenza, upper respiratory viral detection generally continues for 3 to 5 days in adults but is longer in the elderly and hospitalized patients and may persist for weeks to months in immunocompromised hosts. Higher level and more prolonged viral replication, sometimes for weeks, occurs in the lower respiratory tract of patients with viral pneumonia, including pandemic 2009 and avian H5N1 and H7N9 infections. This course typically occurs in association with high plasma pro-inflammatory cytokine-chemokine responses,[3] particularly IL-6, IL-8, and macrophage inflammatory protein (MIP)-1β, which likely reflect production in the infected lung. Viremia or extrapulmonary dissemination is rarely detected in typical human influenza, but both occur in some patients with avian H5N1 infections, in whom gastrointestinal replication may also occur.

Nasal and bronchial biopsy specimens from persons with uncomplicated influenza reveal desquamation of the ciliated columnar epithelium and loss of cilia. The lungs in fatal influenza may show necrotizing bronchitis, diffuse alveolar damage with epithelial necrosis, alveolar edema and hemorrhage, and hyaline membrane formation, followed later by squamous metaplasia and fibrosis. Secondary bacterial infections develop as a result of altered bacterial flora, damage to bronchial epithelium with depressed mucociliary clearance, decreased polymorphonuclear and alveolar macrophage functions, accumulation of alveolar fluid, and suppression of other host immune responses.

Humoral immunity to influenza appears to be largely subtype specific and durable for a particular strain. Neutralizing, hemagglutination-inhibiting (HAI), anti-NA, complement-fixing, enzyme-linked immunosorbent assay and immunofluorescent antibodies begin to develop in the sera of persons with primary influenza virus infection during the second week after infection and reach a peak by 4 weeks. Secretory antibodies develop in the upper respiratory tract and consist predominantly of immunoglobulin (Ig)A

antibodies. Antibody responses are brisker in subsequent infections. Protective immunity against influenza virus infection is mediated by neutralizing antibodies, and protection against illness is generally associated with serum HAI titers of 1 : 40 or greater, serum-neutralizing antibody titers of 1 : 8 or greater, or nasal-neutralizing antibody titers of 1 : 4 or greater.

Cell-mediated immune responses usually develop by 1 week after infection and are considered important for termination of viral replication.[4] Memory lymphocytes from previous infections can limit the severity of disease even in the absence of specific antibodies and may confer some degree of heterosubtypic immunity. Both virus-specific $CD4^+$ and $CD8^+$ T-lymphocyte responses to conserved epitopes on internal proteins appear to contribute.

An increasing number of host genetic factors are being recognized as increasing the risk of severe influenza. One allele in the IFN-induced transmembrane 3 (*IFITM3*) gene has been linked to both severe pandemic H1N1 and avian H7N9 illness.

CLINICAL MANIFESTATIONS

Influenza Syndrome

An abrupt onset of feverishness, chilliness, rigors, headache, myalgia, and malaise is characteristic of influenza but occurs in less than two thirds of cases. Systemic symptoms predominate initially, and prostration occurs in more severe cases. Myalgia, arthralgias, malaise, and headache are usually the most troublesome early symptoms, and their severity is generally related to the level of fever. Ocular symptoms, including photophobia, tearing, burning, and pain on moving the eyes, are sometimes present. Conjunctivitis is characteristic in some avian H7 virus infections, although not in recent H7N9 cases. Respiratory symptoms, particularly dry cough and nasal discharge, are typical early in the illness but are overshadowed by the systemic symptoms. Nasal obstruction, hoarseness, and sore throat are likewise common. Pandemic H1N1 2009 and avian influenza illness are associated with nausea, vomiting, and diarrhea in some adults. As systemic illness diminishes, respiratory complaints and findings become more apparent. Cough is the most frequent and troublesome symptom and may be accompanied by substernal discomfort or burning. Cough, lassitude, and malaise may persist for several weeks before full recovery.

Fever is the most common initial physical finding, but it may be minimal or absent, especially in elderly patients or immunocompromised hosts. The temperature usually rises rapidly to a peak of 38° to 40° C within 12 hours of onset, concurrently with systemic symptoms. Fever is usually continuous but may be intermittent, especially if antipyretics are administered. Typically, the duration of fever in adults is about 3 days, but it may persist from only 1 to 5 or more days. Early in the course of illness, the patient appears toxic, the face is flushed, and the skin is hot and moist. The eyes are watery and reddened. Clear nasal discharge is common. The mucosa of the nose and throat is hyperemic, but exudate is not observed. Small, tender cervical lymph nodes are often present. Transient scattered rhonchi or localized areas of rales are found in less than 20% of cases.

The same pattern of illness occurs with any strain of influenza A or B virus. However, illness is more frequent and severe in smokers. In children, maximum temperatures are often higher, cervical adenopathy may be more frequent, and nausea, vomiting, and abdominal pain are more common. Persons at higher risk for influenza-associated complications and hospitalization (Table 364-4) include pregnant women (especially during the second and third trimesters or early postpartum period), morbidly obese patients, immunosuppressed persons, and patients with various comorbidities. Older adults, especially the infirm elderly, develop fever, muscle aches, sore throat, and headache less often but have higher rates of altered mental status and pulmonary complications. Immunocompromised hosts may also have an apparently mild initial illness but subsequently progress to severe lower respiratory tract involvement. In sporadic illness attributable to H5N1 or H7N9 virus, upper respiratory complaints are less frequent, diarrhea is more common, and progressive viral pneumonia with high mortality is much more likely. Influenza C virus generally causes sporadic upper respiratory tract illness or febrile bronchitis. Most adults hospitalized with seasonal influenza have exacerbations of underlying cardiopulmonary (e.g., myocardial ischemia, heart failure, chronic obstructive pulmonary disease) or metabolic (e.g., diabetes) conditions, and about one third have pneumonia.

TABLE 364-4 PARTICULAR TARGET GROUPS FOR SEASONAL INFLUENZA IMMUNIZATION*

GROUPS AT INCREASED RISK FOR INFLUENZA COMPLICATIONS
Persons aged 50 years and older
Residents of nursing homes and other chronic care facilities
Patients with a chronic pulmonary (including asthma) or cardiac (except isolated hypertension) disorder
Patients with chronic metabolic (including diabetes), renal, hepatic, hematologic (including hemoglobinopathies), or neurologic (including those that compromise respiratory function or increase aspiration risk) disorder
Patients with immunosuppression (including that caused by medications or HIV infection)
Women who will be pregnant or within 2 weeks postpartum during the influenza season
Persons who are morbidly obese (body mass index ≥ 35)
Persons who are American Indians or Alaskan Natives
Children aged 6 months through 59 months, especially those aged < 24 months
Children and teens receiving long-term aspirin
GROUPS IN CONTACT WITH HIGH-RISK PERSONS†
Physicians, nurses, and other health care providers, including those in training
Employees of nursing homes and chronic care and assisted-living facilities
Household contacts (including children) and caregivers of at-risk persons, including children aged < 5 years (particularly contacts of children aged < 6 months), adults aged ≥ 50 years, and persons with medical conditions associated with risk of severe influenza complications (above)

*Routine annual immunization is recommended for all persons aged 6 months or older who do not have contraindications. Those listed are priority groups in the circumstance that vaccine supply is limited or delayed. The risk groups for influenza complications have been expanded because of observations from the 2009 influenza pandemic.
†Inactivated influenza vaccine is recommended for vaccination of household members, health care personnel, and others who have close contact with severely immunosuppressed persons (e.g., patients with hematopoietic stem cell transplants) during those periods in which the immunosuppressed person requires care in a protective environment.
HIV = human immunodeficiency virus.
Modified from Centers for Disease Control and Prevention. Summary recommendations: prevention and control of influenza with vaccines: recommendations of the Advisory Committee on Immunization Practices (ACIP)—United States, 2013-14. Available at: http://www.cdc.gov/flu/professionals/acip/2013-summary-recommendations.htm.

Respiratory Complications

Three pneumonic syndromes have been described—primary influenza viral pneumonia, secondary bacterial pneumonia, and mixed viral and bacterial pneumonia (Chapter 97)—but their clinical presentations and courses overlap considerably. Influenza A and B virus infections are often associated with other respiratory tract complications, including exacerbations of chronic bronchitis, asthma, or cystic fibrosis; croup and bronchiolitis in young children; and otitis media, sinusitis, and rarely parotitis or bacterial tracheitis. Apparently uncomplicated influenza frequently causes tracheobronchitis and is often accompanied by abnormal tracheobronchial clearance, airway hyperactivity, and dysfunction of small airways lasting weeks. A syndrome mimicking pulmonary embolism with transiently altered perfusion scans has also been described.

Severe primary influenza viral pneumonia and associated acute respiratory distress syndrome (ARDS [Chapter 104]) is uncommon during epidemics but accounts for 20 to 50% of pneumonias during pandemics and has been the principal manifestation of severe pandemic H1N1 or avian H5N1 and H7N9 illness. Severe viral pneumonia occurs predominantly in persons with underlying pulmonary and cardiac disorders, pregnancy, or immunodeficiency states, although up to 40% of reported cases and almost all patients with H5N1 have no recognized underlying disease. After a typical onset of influenza, patients often have progressive cough, dyspnea, sometimes hemoptysis, and even cyanosis within 3 to 7 days. Bilateral pulmonary infiltrates and hypoxemia, often indicative of ARDS, may evolve rapidly. Gram staining of sputum may show abundant polymorphonuclear leukocytes but scant bacterial flora. Sputum and endotracheal aspirates usually yield high titers of influenza virus by RNA detection or culture, but upper respiratory samples are sometimes negative.

In patients with classic bacterial superinfection, transient improvement for 1 to 4 days may be followed by recrudescence of fever, increased cough, sputum production, pleuritic chest pain, and a localized area of consolidation. Some patients develop fulminant pneumonia. Gram staining and culture of sputum or blood cultures most often reveal *Streptococcus pneumoniae* (Chapter 289), *Staphylococcus aureus* (Chapter 288) including community-acquired methicillin-resistant *S. aureus*, *Haemophilus influenzae* (Chapter 300), or

Streptococcus pyogenes. Such patients usually respond to specific antibiotic therapy, although staphylococcal infections may be particularly virulent and cause destructive pulmonary lesions. Corticosteroid use is a risk factor for invasive aspergillosis (Chapter 339), which occurs rarely after influenza.

In addition, during an outbreak of influenza, many less distinct syndromes are observed; patients may have viral tracheobronchitis (Chapter 96), milder forms of localized viral pneumonia, or mixed viral and bacterial infection. Many such patients respond to antibiotics. Immunocompromised hosts, including transplant recipients and acute leukemia patients undergoing chemotherapy, may have high rates of pneumonia and associated mortality after influenza if they are not treated with appropriate antiviral and antimicrobial medications.

Nonrespiratory Complications

Severe influenza, including both pandemic H1N1 and avian H5N1 or H7N9 disease, may be associated with sepsis syndrome (Chapter 108), acute renal insufficiency (Chapter 120), and multiorgan failure. Lymphopenia and thrombocytopenia are common in severe influenza; the hemophagocytic syndrome and disseminated intravascular coagulation (Chapter 175) can occur. Myositis with tender leg muscles and elevated serum creatine kinase levels is uncommon in adults, but rhabdomyolysis (Chapter 113) can be severe and cause myoglobinuria. Pregnant women have increased risk of premature labor and spontaneous abortion. Toxic shock syndrome (Chapter 288) caused by respiratory tract infection with toxin-bearing *S. aureus* or *Streptococcus pyogenes* can occur, and outbreaks of meningococcal infection (Chapter 298) have been associated with both influenza A and B virus infections. Myocarditis (Chapter 60) or pericarditis (Chapter 77) occurs uncommonly. Neurologic complications can include aseptic meningitis (Chapter 412), myelitis (Chapter 411), encephalopathy (Chapter 414), necrotizing encephalitis, postinfluenzal Guillain-Barré syndrome (Chapter 420), or immune-mediated encephalitis or cerebellitis. Reye's syndrome (Chapter 150), which is a well-recognized hepatic and central nervous system complication of influenza A and B virus infections in children and rarely in adults, is associated with salicylate use. Possible associations with late-onset Parkinson's syndrome or with neuropsychiatric disorders in the offspring of women experiencing intrapartum infection remain uncertain.

DIAGNOSIS

In an individual case, influenza often cannot be distinguished from infection with a number of other viruses or nonviral pathogens that produce similar clinical manifestations (Chapters 96 and 429). Conversely, when public health authorities report an epidemic of influenza in a given community and an adult patient is seen with typical febrile respiratory illness, it is highly likely that these symptoms are caused by an influenza A or B virus infection.[5] In such circumstances, the presence of fever and cough has a positive predictive value of about 80% for laboratory-proven influenza in ambulatory adults. However, the predictive value of influenza-like illness in hospitalized patients is lower.

Detection of viral RNA by nucleic acid amplification testing (NAAT) is more sensitive than viral culture and is the rapid test of choice for patients who are hospitalized with suspected influenza infection. Some commercial multiplex assays that detect influenza and a variety of other respiratory viruses can be completed within 1 to 2 hours.

By comparison, commercially available rapid influenza diagnostics tests, which can detect influenza virus antigens in less than 15 minutes, have low sensitivity (<20 to 70%) in adults, so negative results should not guide individual treatment decisions. However, such tests may be helpful in investigating outbreaks while awaiting more definitive test results. The limited specificity (generally 90 to 95%) of some rapid tests makes their predictive value low outside the influenza season, although optical readers appear to enhance their reliability.

Viral culture of nasal, sputum, or tracheal secretions during the first 2 or 3 days of illness is more sensitive than rapid influenza diagnostic tests, but results usually take 48 to 72 hours. Serologic methods are less useful clinically because they require convalescent serum obtained 14 to 21 days after the onset of infection.

Detection of secondary bacterial infections generally relies on standard microbiological studies (e.g., blood culture, sputum Gram stain and culture, urinary antigen testing). In hospitalized patients, a low serum procalcitonin level may help discriminate viral from mixed influenza-bacterial pneumonia (Chapter 97).

PREVENTION

Vaccination

Annual influenza immunization is recommended for all persons aged 6 months or older in the United States (Chapter 18).[6] Persons at higher risk for influenza-related complications or who care for or are exposed to them remain particular priorities of immunization efforts (see Table 364-4), especially when vaccine delays or shortages occur. Influenza vaccine policies differ across countries. Because influenza immunization benefits both mother and infants, the WHO has ranked pregnant women as the highest priority, followed (in no particular order) by health care workers, children 6 to 59 months of age, the elderly, and individuals with high-risk conditions.

Seasonal vaccine should be given each year in the fall as soon as available, preferably by October before the influenza season in northern temperate areas. Quadrivalent vaccines with two A (H1N1, H3N2) and two B lineage (Yamagata, Victoria) antigens are increasingly replacing trivalent vaccines. In recent seasons, pandemic 2009 virus has been incorporated as the H1N1 component.

Egg-grown intramuscularly administered inactivated influenza vaccines for persons 6 months of age and older, a high-dose inactivated vaccine for persons aged 65 years and older, and a live-attenuated intranasal vaccine for otherwise healthy persons aged 2 to 49 years are currently licensed in the United States (E-Table 364-E1). An intradermally administered inactivated vaccine, a mammalian cell culture-grown vaccine, and a recombinant DNA-produced HA vaccine, particularly useful in those with serious egg allergies, are also available for seasonal use. Racial and ethnic disparities in influenza vaccination rates persist, and rates for other target groups, including health care workers, remain suboptimal. Immunization of health care personnel, which represents an important patient safety issue, can be enhanced by strategies to improve access, and especially by employer mandates.

Inactivated vaccines (Chapter 18) given by intramuscular injection provide about 50 to 70% protection against seasonal influenza illness in young and middle-aged adults, albeit with substantial year-to-year variations in effectiveness,■ and reduce work absenteeism, use of health care resources, and antibiotics when the vaccine is well matched to the epidemic strain. Live-attenuated vaccines are highly protective against influenza in children but appear less effective in adults compared with inactivated vaccine.■ Inactivated pandemic 2009 H1N1 vaccines are highly immunogenic in older children and adults after one standard 15-μg HA dose, perhaps in part related to preexisting immunity, and the effectiveness of monovalent pandemic 2009 H1N1 vaccine ranges from 60 to 93%.

Immunogenicity and hence protection rates with inactivated seasonal vaccines are lower in elderly persons, particularly infirm nursing home residents, and in immunosuppressed patients. The effectiveness of influenza vaccine for prevention of medically attended acute respiratory illness among the elderly in nursing homes is estimated to be 20 to 40%. In ambulatory high-risk patients, immunization reduces hospitalizations from pneumonia, influenza, and major cardiovascular events,■ as well as all-cause mortality during the influenza season. Higher-dose inactivated vaccines (four-fold more HA per strain) are more immunogenic and protective than standard vaccine against seasonal influenza illness in older ambulatory adults.■ Immunization of children appears to reduce respiratory illness in household and community contacts.■ Immunization of health care providers reduces the risk of transmission to patients as well as their own risk of infection.[7] Maternal immunization benefits both mother and child.

Because protection is greatly reduced or absent in some seasons and some patient groups, new vaccines with improved immunogenicity are being developed. Oil-in-water adjuvants (e.g., MF-59, AS03) are immunogenic at lower HA doses and appear to be especially effective for inducing adequate immune responses to novel influenza viruses like avian H5N1 and H7N9. For H5 and H7 vaccines, two doses with an oil-in-water adjuvant appear to be necessary for adequate immunogenicity. An AS03-adjuvanted H5N1 vaccine (containing only 3.75 μg HA antigen) was approved for adults in the United States in 2013. Seasonal MF59-adjuvanted vaccine is approved in Europe and Canada for adults age 65 years and older, and a virosome-adjuvanted vaccine is approved in Europe. An AS03-adjuvanted vaccine provides somewhat greater protection in the ambulatory elderly.■

With inactivated vaccine, fever and systemic symptoms occur at rates comparable to those of adults given placebo but are more common in young children. Among adults, 25% or more may have mild local reactions at the site of injection. Ocular or respiratory symptoms have occasionally been

E-TABLE 364-1 CURRENTLY APPROVED INFLUENZA VACCINES IN THE UNITED STATES (2013-14 SEASON)*

VACCINE TYPE	SUBSTRATE FOR PRODUCTION	NO. COMPANIES	FDA-APPROVED AGE RANGE (YR)	ROUTE	DOSE IN ADULTS
Inactivated influenza vaccine, trivalent (IIV3), standard dose	Eggs	Multiple	≥0.5†	IM	0.5 mL containing 15 µg per HA antigen
Inactivated influenza vaccine, quadrivalent (IIV4)	Eggs	Multiple	≥0.5†	IM	0.5 mL containing 15 µg per HA antigen
Inactivated influenza vaccine, trivalent (IIV3), high dose	Eggs	1	≥65	IM	0.5 mL containing 60 µg per HA antigen
Inactivated influenza vaccine, trivalent (IIV3), cell culture (ccIIV3)	MDCK cells	1	≥18	IM	0.5 mL containing 15 µg per HA antigen
Inactivated influenza vaccine, trivalent (IIV3), intradermal	Eggs	1	18-64	Intradermal with special microinjection system	0.1 mL containing 9 µg per HA antigen
Recombinant inactivated influenza vaccine, trivalent (rIIV3)	Insect cells	1	18-49	IM	0.5 mL containing 45 µg per HA antigen
Live-attenuated influenza vaccine, quadrivalent (LIV4)	Eggs	1	2-49‡	Intranasal	0.2 mL

*Package inserts for U.S.-licensed vaccines are available at http://www.fda.gov/BiologicsBloodVaccines/Vaccines/ApprovedProducts/ucm093833.htm.
†Lower age limit of approval depends on specific vaccine product.
‡LAIV is approved only for otherwise healthy nonpregnant persons. It should be avoided in close contacts and caregivers of severely immunosuppressed persons who require a protected environment and in children and adolescents receiving concomitant aspirin therapy.
FDA = U.S. Food and Drug Administration; IM = intramuscular.

TABLE 364-5 ANTIVIRAL DOSE RECOMMENDATIONS IN ADULTS

DRUG	ROUTE	TREATMENT	PROPHYLAXIS	DOSE REDUCTIONS	COMMENT
Oseltamivir	Oral	75 mg bid	75 mg once daily†	CrCl ≤ 60 mL/min	Gastrointestinal and occasional CNS side effects
Zanamivir	Inhaled	10 mg bid	10 mg once daily†	Avoid in those with underlying airways disease	Training in use of the inhaler device is important
Peramivir*	Intravenous	600 mg single IV dose	Not studied	CrCl < 50 mL/min	Single dose approved for uncomplicated influenza in adults in US
Laninamivir*	Inhaled	40 mg once (Japan)	20 mg once daily for two days (Japan)	Avoid in those with underlying airways disease	Currently investigational in United States (US)
Zanamivir*	Intravenous	600 mg q12h	Not studied	CrCl < 80 mL/min	Currently investigational in US
Favipiravir*	Oral	1600 mg twice on day 1, 600 mg bid (Japan)	Not studied	Pending; avoid in pregnancy	Currently investigational in US; inhibitory for NAI-resistant viruses
Amantadine	Oral	100 mg bid	100 mg bid	Age > 64 yr, CrCl < 50-80 mL/min	Gastrointestinal and CNS side effects
Rimantadine	Oral	100 mg bid	100 mg bid	Age > 64 yr, CrCl < 10 mL/min, or severe hepatic dysfunction	Lower risk of CNS side effects than with amantadine

*Currently, intravenous zanamivir is available on a compassionate use/emergency investigational new drug (eIND) basis in the United States. Inhaled laninamivir is licensed in Japan; higher treatment doses (80 mg) are being studied in the United States. Oral favipiravir is approved in Japan but only for treating novel or reemerging influenza infections when other drugs are ineffective or not sufficiently effective; higher dose regimens are being studied in the United States.
†Full therapeutic doses for postexposure prophylaxis (i.e., twice-daily dosing) are appropriate for infections due to novel influenza viruses and in immunocompromised hosts.
The standard duration of treatment in uncomplicated illness is 5 days, except for laninamivir or IV peramivir, for which single doses suffice. A longer treatment duration (e.g., 10 days) should be considered for infections in seriously ill patients, immunosuppressed hosts, or infections caused by novel viruses. Higher doses (e.g., 150 mg or 225 mg twice daily) of oseltamivir have been used in patients with serious lower respiratory illness, immunocompromise, or infections by novel viruses. The duration of prophylaxis depends on the epidemiologic setting; durations of 2 weeks after immunization or 7 to 10 days for postexposure prophylaxis are appropriate. Antiviral prophylaxis may interfere with response to intranasal live-attenuated vaccine but not to intramuscular inactivated vaccine.
CNS = central nervous system; CrCl = creatinine clearance; NAI = neuraminidase inhibitor.

reported within 24 hours after administration but typically resolve quickly without specific treatment. Injection site reactions and systemic complaints (headache, fatigue, myalgia, arthralgia) are more frequent after the high HA content or oil-in-water adjuvanted inactivated vaccines. An AS03-adjuvanted pandemic 2009 H1N1 vaccine was associated with narcolepsy in children and adolescents, apparently related to a specific genetic predisposition.[8] Intranasal vaccine causes coryza and sore throat in adults.

Hypersensitivity reactions to residual egg proteins or other vaccine components occur rarely, and vaccine is contraindicated in persons with chicken egg anaphylactic hypersensitivity unless the patient has been desensitized. Inactivated vaccine does not cause exacerbation of asthma but is rarely associated with Guillain-Barré syndrome in older adults.

Medications

Inhaled zanamivir and oral oseltamivir are effective for chemoprophylaxis of both influenza[9] A and B virus infections, including postexposure prophylaxis in households (Table 364-5), with each reducing the risk of symptomatic influenza infections by about 60 to 80%. Because of widespread resistance, rimantadine and amantadine are no longer effective for prophylaxis or treatment. When an outbreak develops, unimmunized high-risk persons can be given chemoprophylaxis and inactivated vaccine simultaneously, with cessation of chemoprophylaxis after 14 days. Alternatively, if vaccine is not available or contraindicated, is a poor match, or if the patient is highly immunosuppressed, chemoprophylaxis may be continued for the duration of the community outbreak. When given to patients and staff alike, these drugs may be helpful in managing nosocomial outbreaks. Therapeutic doses (i.e., twice daily) should be considered for postexposure prophylaxis in immunocompromised hosts and in individuals exposed to novel influenza viruses to increase effectiveness and possibly reduce antiviral resistance emergence.

Most public health authorities do not advise routine use of antiviral chemoprophylaxis, largely because of concerns regarding promotion of antiviral resistance and drug availability. In the community setting, close monitoring and early initiation of antiviral treatment is an alternative approach after suspected influenza exposures. Chemoprophylaxis is generally not recommended if more than 2 to 3 days have elapsed since exposure to a person with seasonal influenza.

Precautions

Upon entry to a health care facility, source control (e.g., facemask placement) and instruction in respiratory hygiene and cough etiquette should be implemented for symptomatic patients. Patients hospitalized with suspected or proven influenza should be managed with standard and droplet precautions and, when possible, private rooms. Facemasks likely provide some degree of protection during routine patient care, but compliance with a properly fitted N95-type respirator provides better protection. A N95-type respirator should be used with other precautions (including gloves, gown, and eye protection with face shield or goggles; an airborne infection isolation room) during aerosol-generating procedures. Other strategies for preventing nosocomial influenza include influenza vaccination of both health care providers and patients, appropriate management of ill health care providers, and engineering infection-control measures (e.g., adequate air exchanges). Early implementation and compliance with masks and hand hygiene appears to reduce the risk of secondary infections in household contacts.

The risk of zoonotic infections can be reduced by avoiding exposure to potentially infectious poultry or swine or their environments in affected countries. Of note, while the highly pathogenic H5N1 virus causes lethal poultry outbreaks, the low pathogenic H7N9 virus does not cause discernible illness in affected birds.

TREATMENT Rx

Inhaled zanamivir and oral oseltamivir (see Table 364-5) are active against both influenza A and B viruses.[10] Antiviral treatment is recommended as soon as possible for patients who have proven or suspected influenza and who manifest severe, complicated, or progressive illness; are hospitalized; or have underlying conditions that increase the risk for complications (see Table 364-4), including adults older than 64 and children younger than 5 (especially those < 2 years). In patients with serious illness or high-risk conditions, decisions about initiating antiviral therapy should not wait for laboratory confirmation of influenza. Treatment within 2 days of the onset of symptoms also may be considered in low-risk outpatients with uncomplicated febrile illness, in whom therapy reduces the duration of symptoms by 1 to 2 days, the time to functional recovery, and the risk for lower respiratory complications that lead to antibiotic use. Early oseltamivir therapy also decreases the risk for pneumonia and hospitalization, and even delayed use appears to reduce mortality in hospitalized patients. Double-doses of oseltamivir do not appear more effective than standard doses even though they clear influenza B viruses more rapidly. Longer courses (e.g., 10 days) and perhaps higher doses are warranted in critically ill patients and immunocompromised hosts, in whom monitoring for virologic clearance is warranted. A single dose of intravenous peramivir appears to be as effective as a 5-day regimen of oseltamivir in treating uncomplicated influenza.

During therapy, the emergence of oseltamivir resistance is infrequent but is more common in children and immunocompromised hosts; it has developed in H5N1 and pandemic 2009 H1N1 infections owing to a specific NA substitution (H275Y). Global circulation of oseltamivir-resistant seasonal H1N1 viruses in 2008-2009 and other resistance-conferring NA substitutions in avian H7N9 virus during neuraminidase therapy highlight the need for ongoing susceptibility monitoring for it and other neuraminidase inhibitors.[11] Zanamivir is inhibitory for most oseltamivir-resistant variants, including those with

H275Y, and intravenous zanamivir is currently preferred for treating severe infections that are suspected or proven to be oseltamivir-resistant. Agents with other mechanisms of action, such as the polymerase inhibitor favipiravir, are inhibitory for neuraminidase-resistant variants. Inhaled zanamivir may be infrequently associated with bronchospasm, sometimes severe, and nebulization of the lactose-containing commercial form is contraindicated in intubated patients. Oseltamivir is associated with nausea, vomiting, rash, and possibly rare neuropsychiatric symptoms. Oral rimantadine or amantadine shortens the duration of fever and symptoms in uncomplicated influenza A caused by susceptible strains (see Table 364-5), but these drugs are not currently recommended for use because of widespread antiviral resistance.

Symptomatic measures include antipyretics and cough suppressants. Salicylates should not be used, especially in children younger than 16 years, because of their association with Reye's syndrome.

Fulminant influenza viral pneumonia, particularly following pandemic H1N1 or avian influenza infection, requires intensive care including ventilatory support (Chapter 105), often renal replacement therapy, and sometimes even extracorporeal membrane oxygenation. Early oseltamivir appears to reduce mortality in patients if it is initiated before onset of respiratory failure. Intravenous formulations of zanamivir provide reliable delivery of high drug levels in seriously ill persons and are available for compassionate use in the United States. Intravenous peramivir, intravenous ribavirin, aerosolized ribavirin, nebulized zanamivir, and a triple drug regimen of amantadine, ribavirin, and oseltamivir have been used with uncertain benefit. Timely administration of convalescent plasma or hyperimmune globulin-containing neutralizing antibodies appeared effective in reducing mortality in critically ill patients during the H1N1 pandemic. Adding sirolimus to oseltamivir accelerated viral clearance and clinical recovery in mechanically ventilated pandemic 2009 H1N1 patients, but this combination remains investigational for treating influenza.[12] By comparison, the combination of oseltamivir and zanamivir appears to be inferior to oseltamivir monotherapy. Systemic corticosteroids given for influenza-associated pneumonia or ARDS have been associated with prolonged viral replication, adverse effects, and increased mortality, so their routine use should be avoided.[13]

Adults admitted with community-acquired pneumonia (Chapter 97) during the influenza season should initially receive both antiviral therapy and antibiotics, with subsequent therapy guided by microbiological study results. Intensive care unit complications, including ventilator-associated pneumonia, are common.

PROGNOSIS

Most influenza patients make a full recovery, but they may require several weeks to return to full function. Influenza in elderly persons can cause prolonged loss of function and impairment of activities of daily living. The mortality from seasonal influenza or pandemic 2009 H1N1 illness has been low (≈1 in 10,000 persons), but their impact differs across age and risk groups. Although approximately 90% of seasonal influenza-related deaths occur in persons who are older than 65 years or at high risk because of comorbid conditions, most pandemic 2009 virus deaths have been in patients younger than 65. Mortality occurs in about 5 to 10% of adults hospitalized with seasonal influenza but is as high as 40 to 60% in zoonotic H5N1 and H7N9 infections. Bacterial infections (Chapter 97) were associated with more than 90% of fatal pneumonias in the 1918 pandemic and about 20 to 40% in the 2009 H1N1 pandemic.[14] Detection of viral RNA in the blood is a poor prognostic finding that may reflect high viral burden in the lower respiratory tract.[15] Once viral pneumonia progresses to respiratory failure, the mortality rate is often 50% or higher, but the risk is reduced by early antiviral therapy and high quality intensive care support.

Grade A References

A1. Osterholm MT, Kelley NS, Sommer A, et al. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. *Lancet Infect Dis*. 2012;12:36-44.

A2. Monto AS, Ohmit SE, Petrie JG, et al. Comparative efficacy of inactivated and live attenuated influenza vaccines. *N Engl J Med*. 2009;361:1260-1267.

A3. Udell JA, Zawi R, Bhatt DL, et al. Association between influenza vaccination and cardiovascular outcomes in high-risk patients: a meta-analysis. *JAMA*. 2013;310:1711-1720.

A4. DiazGranados CA, Dunning AJ, Kimmel M, et al. Efficacy of high-dose versus standard-dose influenza vaccine in older adults. *N Engl J Med*. 2014;371:635-645.

A5. Loeb M, Russell ML, Moss L, et al. Effect of influenza vaccination of children on infection rates in Hutterite communities: a randomized trial. *JAMA*. 2010;303:943-950.

A6. McElhaney JE, Beran J, Devaster JM, et al. AS03-adjuvanted versus non-adjuvanted inactivated trivalent influenza vaccine against seasonal influenza in elderly people: a phase 3 randomised trial. *Lancet Infect Dis*. 2013;13:485-496.

A7. Khazeni N, Bravata DM, Holty JE, et al. Systematic review: safety and efficacy of extended-duration antiviral chemoprophylaxis against pandemic and seasonal influenza. *Ann Intern Med*. 2009; 151:464-473.

A8. Loeb M, Dafoe N, Mahony J, et al. Surgical mask vs N95 respirator for preventing influenza among health care workers: a randomized trial. *JAMA*. 2009;302:1865-1871.

A9. Muthuri SG, Venkatesan S, Myles PR, et al. Effectiveness of neuraminidase inhibitors in reducing mortality in patients admitted to hospital with influenza A H1N1pdm09 virus infection: a meta-analysis of individual participant data. *Lancet Respir Med*. 2014;2:395-404.

A10. South East Asia Infectious Disease Clinical Research Network. Effect of double dose oseltamivir on clinical and virological outcomes in children and adults admitted to hospital with severe influenza: double blind randomised controlled trial. *BMJ*. 2013;346:f3039.

A11. Whitley R, Laughlin A, Carson S, et al. Single dose peramivir for the treatment of acute seasonal influenza: integrated analysis of efficacy and safety from two placebo-controlled trials. *Antivir Ther*. 2014; [Epub ahead of print].

A12. Hung IF, To KK, Lee CK, et al. Hyperimmune IV immunoglobulin treatment: a multicenter double-blind randomized controlled trial for patients with severe 2009 influenza A(H1N1) infection. *Chest*. 2013;144:464-473.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Li Q, Zhou L, Zhou M, et al. Epidemiology of human infections with avian influenza A(H7N9) virus in China. *N Engl J Med*. 2014;370:520-532.
2. Bischoff WE, Swett K, Leng I, et al. Exposure to influenza virus aerosols during routine patient care. *J Infect Dis*. 2013;207:1037-1046.
3. Wang Z, Zhang A, Wan Y, et al. Early hypercytokinemia is associated with interferon-induced transmembrane protein-3 dysfunction and predictive of fatal H7N9 infection. *Proc Natl Acad Sci U S A*. 2014;111:769-774.
4. Sridhar S, Begom S, Bermingham A, et al. Cellular immune correlates of protection against symptomatic pandemic influenza. *Nat Med*. 2013;19:1305-1312.
5. World Health Organization website. http://www.who.int/topics/influenza/en. Accessed March 11, 2015.
6. Centers for Disease Control and Prevention. Prevention and control of influenza with vaccines: recommendations of the Advisory Committee on Immunization Practices. http://www.cdc.gov/flu/professionals/acip/2013-summary-recommendations.htm. Accessed March 11, 2015.
7. Ahmed F, Lindley MC, Allred N, et al. Effect of influenza vaccination of healthcare personnel on morbidity and mortality among patients: systematic review and grading of evidence. *Clin Infect Dis*. 2014;58:50-57.
8. Partinen M, Kornum BR, Plazzi G, et al. Narcolepsy as an autoimmune disease: the role of H1N1 infection and vaccination. *Lancet Neurol*. 2014;13:600-613.
9. Centers for Disease Control and Prevention. Antiviral agents for the treatment and chemoprophylaxis of influenza. Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morbid Mortal Wkly Rep*. 2011;60(No. RR–1):1-25.
10. Centers for Disease Control and Prevention Influenza website. http://www.cdc.gov/flu/. Accessed March 11, 2015.
11. Hu Y, Lu S, Song Z, et al. Association between adverse clinical outcome in human disease caused by novel influenza A H7N9 virus and sustained viral shedding and emergence of antiviral resistance. *Lancet*. 2013;381:2273-2279.
12. Dunning J, Baillie JK, Cao B, et al. Antiviral combinations for severe influenza. *Lancet Infect Dis*. 2014;14:1259-1270.
13. Hui DS, Lee N. Adjunctive therapies and immunomodulating agents for severe influenza. *Influenza Other Respir Viruses*. 2013;7(suppl 3):52-59.
14. Jain S, Benoit SR, Skarbinski J, et al. Influenza-associated pneumonia among hospitalized patients with 2009 pandemic influenza A (H1N1) virus—United States, 2009. *Clin Infect Dis*. 2012;54:1221-1229.
15. Fielding JE, Kelly HA, Mercer GN, et al. Systematic review of influenza A(H1N1)pdm09 virus shedding: duration is affected by severity, but not age. *Influenza Other Respir Viruses*. 2014;8:142-150.

REVIEW QUESTIONS

1. A 23-year-old woman who is 6 months pregnant presents with fever, increasing cough, and dyspnea of 4 days' duration during a known community outbreak of influenza. Physical examination shows T 39° C, P 110, R 25, Sao_2 (room air) 92%, bibasilar crackles, 2/6 systolic murmur at left heart border, and 1+ pretibial edema bilaterally without cords or calf tenderness. Abdomen shows gravid uterus with normal fetal heart sounds. Nose swab is negative for influenza antigen by rapid test. WBC count is 4000 with 75% PMNs, 15% lymphs, and 10% monos. Chest radiograph shows patchy bilateral mid- and lower lung zone densities. Lower extremity ultrasound shows no evidence for deep venous thrombosis. You administer oxygen, hospitalize her, request urgent maternal-fetal and pulmonary consultations, and in the meantime decide to start which of the following regimens:

A. Antibiotics for community-acquired pneumonia (CAP), acetaminophen, and oral oseltamivir
B. Antibiotics for CAP, acetaminophen, and systemic glucocorticoids
C. Inhaled zanamivir and oral oseltamivir
D. Oral oseltamivir, systemic corticosteroids
E. Antibiotics for CAP, acetaminophen, and antitussives

Answer: A Pregnant women are at increased risk for complications of influenza, with their risk increasing across trimesters and extending into the early postpartum period. This patient appears to be rapidly deteriorating, with progression toward acute respiratory distress syndrome (ARDS) and requires urgent intervention. The negative rapid influenza antigen test and the delay in presentation from onset of symptoms should not dissuade empirical use of antiviral therapy. The possibility of bacterial coinfection mandates the use of antibiotics pending microbiological studies. The combined use of two neuraminidase inhibitors has not been shown to be superior to oseltamivir alone, and in one trial an inhaled zanamivir-oral oseltamivir regimen was inferior in uncomplicated influenza in adults. Systemic corticosteroids have not been associated with improved outcomes in severe influenza, and observational studies indicate an increased risk of secondary infections, other complications, and even mortality in patients with pandemic H1N1 pneumonia and/or ARDS.

2. A 55-year-old woman presents for routine follow-up in late November. Aside from moderate obesity (BMI 32) and long-term smoking of ½ pack per day, she has no recognized chronic medical conditions. She has refused influenza vaccines in the past, claiming egg allergies, and is often noncompliant with her scheduled visits. She reports experiencing tongue swelling after ingesting eggs. You advise:

A. Intramuscular recombinant hemagglutinin vaccine
B. Oral egg protein challenge in the office, followed by standard intramuscular vaccine
C. Intranasal live-attenuated influenza vaccine
D. Intradermal influenza vaccine
E. Allergy consultation, followed by desensitization if necessary and standard intramuscular vaccine

Answer: A Standard inactivated vaccine, intradermal vaccine, and live-attenuated intranasal vaccine are all produced in eggs and contain egg proteins. The risk of anaphylactic egg allergy may be significant in this patient, and an allergy evaluation is one course of action. However, the influenza season may well begin shortly, and her history of missing appointments may indicate that this may be the only opportunity to immunize her in advance of the season. The recombinant vaccine is produced in insect cells and does not contain egg proteins.

3. A 60-year-old man who returned 2 days ago from a business trip to Shanghai presents to the emergency department with 3 days of fever, fatigue, and nonproductive cough in January. He denies sore throat or coryza but reports nausea. None of his travelling companions report illness. He recalls having walked through various marketplaces during his trip but does not recall seeing live poultry. His physical examination shows T 38.5° C, P 95, R 18, chest clear. A rapid antigen test is negative for influenza. After consultation with local public health authorities, you submit a nasopharyngeal swab for diagnostic testing, start oral oseltamivir, and hospitalize him for close observation. On the following day, the swab is reported negative for respiratory viruses, but he has developed worsening cough and respiratory distress, with Sao_2 of 85% on a 5 L/min facemask. Chest radiographs show bilateral ground glass opacifications. WBC count is 3.0 with 10% lymphocytes; platelets 95,000; serum creatinine 1.2 mg/dL. Electrocardiogram shows nonspecific ST-T wave changes; serum troponin is normal. Antibiotics are started, pulmonary consultation is requested, and he is intubated for mechanical ventilator support. Gram stain of a tracheal aspirate shows some PMNs but only rare bacterial forms. Reverse-transcription polymerase chain reaction (RT-PCR) on the aspirate is positive for influenza A. You decide to adjust his regimen by:

A. Switching to intravenous zanamivir
B. Doubling oral oseltamivir dose by nasogastric tube
C. Adding inhaled zanamivir
D. Adding amantadine
E. Adding systemic glucocorticoids

Answer: A His recent travel history raises concern of avian H7N9 infection (also consistent with the paucity of upper respiratory symptoms and negative initial nasopharyngeal RT-PCR for influenza and other respiratory viruses), but other influenza A viruses including A(H1N1)pdm09 are also circulating, and the first case of avian H5N1 in North America occurred in a Canadian traveler returning from Beijing. The patient is rapidly progressing, with apparent viral pneumonia despite initial oseltamivir therapy. Although multiple conditions warrant consideration (e.g., secondary bacterial infection, pulmonary emboli, heart failure from infarct or myocarditis), there are also the concerns that oseltamivir may not be adequately absorbed in critically ill patients and that oseltamivir resistance can emerge rapidly. Limited data indicate that extemporaneous preparations of oral oseltamivir can be effectively administered by nasogastric tube in mechanically ventilated patients, but double doses of oseltamivir do not appear more effective than standard ones. Intravenous zanamivir provides reliable delivery of high drug levels and is active against most but not all oseltamivir-resistant variants. Amantadine is not active for currently circulating human influenza A viruses, avian H7N9, or most avian H5N1 viruses. The commercial formulation of inhaled zanamivir should not be used in mechanically ventilated patients.

365 ADENOVIRUS DISEASES

MICHAEL G. ISON

DEFINITION

Human adenoviruses, which are members of the Adenoviridae family and the *Mastadenovirus* genus, are divided into 57 serotypes and seven species (A, B, C, D, E, F, and G; Table 365-1). Adenoviruses are double-stranded DNA, nonenveloped viruses that code for 20 early and 15 late proteins. The knobbed fiber protrudes from the fiber base (E-Figure 365-E1).

EPIDEMIOLOGY

Adenoviruses cause a range of infections from mild self-limited respiratory viral infections, conjunctivitis, and diarrhea to severe disseminated disease.[1] Adenoviruses have a worldwide distribution, and infections occur throughout the year without significant seasonal variability. Most infections occur as sporadic events, although local or regional epidemics have been described. Asymptomatic respiratory infection is common and is associated with

TABLE 365-1 ADENOVIRUS SUBGROUPS & SEROTYPES WITH THEIR MAJOR SITES OF INFECTION

SUBGROUP	SEROTYPE	MAJOR SITE OF INFECTION
A	12, 18, 31*	Respiratory, urinary, GI
B1	3, 7, 16, 21, 50	Respiratory, eye (including pharyngoconjunctival fever), GI
B2	11,† 14,‡ 34,† 35†	Urinary tract infections
C	1, 2, 5, 6, 57	Respiratory, urinary, GI (especially hepatitis)
D	8§-10, 13, 15, 17, 19,§ 20, 22-30, 32, 33, 36, 37§-39, 42-49, 51, 53, 54, 56	Eye, GI
E	4	Upper respiratory tract infection, pneumonia
F	40,* 41*	GI
G	52	GI

*Associated with infantile gastroenteritis.
†Associated with hemorrhagic cystitis and interstitial nephritis.
‡Associated with epidemic pneumonia with high mortality.
§Associated with epidemic keratoconjunctivitis.
GI = gastrointestinal.

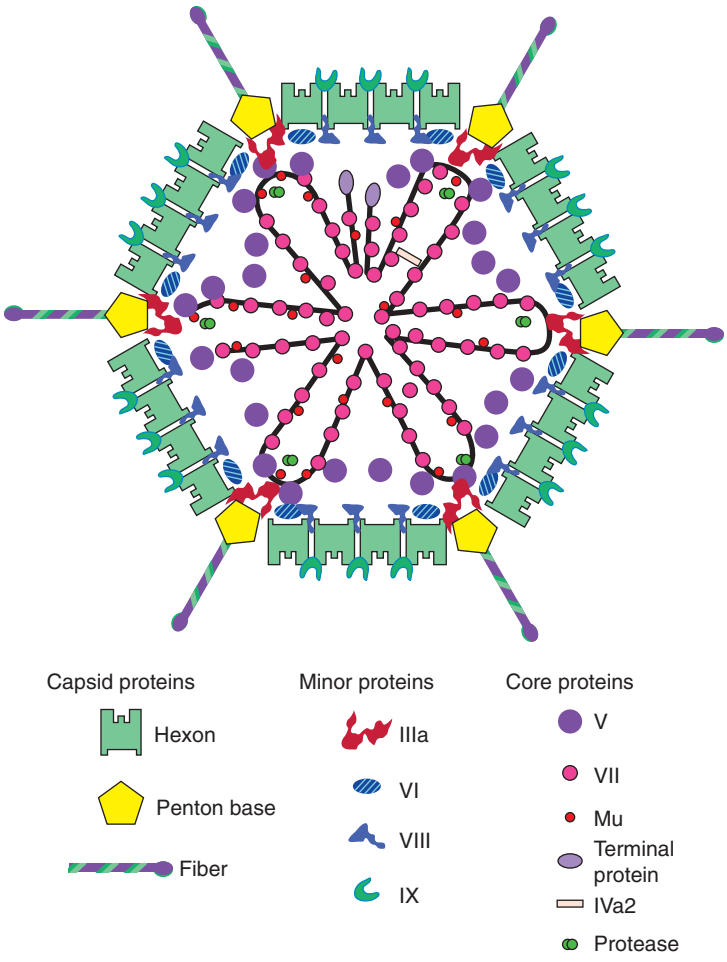

E-FIGURE 365-1. Adenovirus.

prolonged carriage, particularly in feces and tonsilar tissue. Most patients with adenovirus infection are children, generally younger than 5 years, and most individuals have serologic evidence of exposure to adenovirus by age 10. Only about 25% of symptomatic cases occur in adults. AdV1, 2, and 5 are most common among children, whereas AdV3, 4, and 7 are more common in adults.

Recent outbreaks in the United States have resulted in cases requiring hospitalization, and nosocomial transmission can cause severe infections in health care workers. Adenovirus is the leading cause of respiratory viral infections among military recruits. Unlike the year-round circulation of adenovirus in civilians, peaks of disease are recognized within the first 4 weeks of fall training exercises among military recruits.[2] Rates of adenovirus among military recruits declined significantly with widespread use of an oral live-attenuated vaccine in 1971, increased when supplies of vaccine were depleted by 1999, then declined again after an oral live-attenuated adenovirus type 4 and 7 vaccine was approved in 2011 (E-Figure 365-E2).

PATHOBIOLOGY

Adenoviruses enter susceptible hosts by the mouth, nasopharynx, or ocular conjunctiva. The fiber protein of the virus binds to a cellular receptor that varies by serotype. Upon initial infection, local replication can induce considerable cellular damage owing to tissue invasive infection. Necrotizing bronchitis, bronchiolitis, interstitial pneumonia, and fibrin and hyaline membranes are seen within the alveoli during pulmonary infections. In ocular infections, exudative and mononuclear infiltrates develop beneath the epithelium. Replication results in desquamated epithelial cells, which induce hypertrophy in regional lymphatic tissue and active proliferative germinal centers. This reaction can cause swollen adenoids or intussusception (Chapter 142) associated with enlarged mesenteric lymph nodes, particularly in children.[3]

Components of both the innate and adaptive immune responses are important for the control of adenovirus replication. Alveolar macrophages and Kupffer cells help eliminate adenovirus from the lung and liver and also secrete inflammatory cytokines such as tumor necrosis factor (TNF), interferon (IFN)-γ, interleukin (IL)-1β, IL-6, IL-8, and IL-12. $CD4^+$ and $CD8^+$ T lymphocytes play a particularly important role in the control and clearance of replicating adenovirus in humans, and the absolute lymphocyte and $CD4^+$ T-lymphocyte levels correlate inversely with adenovirus infection and the risk for developing disseminated adenoviral infections in immunosuppressed transplant patients.[4]

Both group- and type-specific neutralizing and non-neutralizing antibodies are produced in response to infection and also play a role in limiting infection. Group-specific antibodies do not neutralize the virus but can confirm infection. Neutralizing antibodies may protect against disease manifestations in the previously infected host or against reinfection with the same serotype, but they do not eliminate the carrier state.

CLINICAL MANIFESTATIONS

Adenovirus infections have a wide range of clinical manifestations that are linked to the virus type (see Table 365-1) as well as to the age and immunocompetence of the host. Common syndromes include infections of the respiratory tract, eye, gastrointestinal tract, genitourinary tract, and central nervous system (CNS). Disseminated infection also occurs, especially in immunosuppressed patients.

Respiratory Tract Infections

About 10% of pneumonias in childhood and up to 5% of community-acquired pneumonias in some series of adults[5] may be attributable to adenovirus. Typical symptoms include nasal congestion, coryza, and cough, which may mimic pertussis infections (Chapter 313). Generalized malaise, fever, chills, myalgia, headache, and abdominal pain are common systemic symptoms. Exudative tonsillitis and cervical adenopathy are also seen. If conjunctivitis accompanies these signs and symptoms, the disease is designated as pharyngoconjunctival fever.[6] Otitis media is a common presentation, particularly among infants under 1 year of age. Cases in children and in adults, especially military recruits, are indistinguishable from other viral respiratory infections such as influenza (Chapter 364), parainfluenza (Chapter 363), and respiratory syncytial virus (Chapter 362).

Ocular Infections

The two most common manifestations of ocular adenovirus infection are pharyngoconjunctival fever and epidemic keratoconjunctivitis. Pharyngoconjunctival fever is typically a mild form of acute follicular conjunctivitis that accompanies febrile pharyngitis or cervical adenitis following a 6- to 9-day incubation period and may affect both eyes. Both bulbar and palpebral conjunctival involvement may occur. Swimming pools or lakes are a common source of infection. Symptoms typically resolve without treatment and without sequelae.

Unlike pharyngoconjunctival fever, epidemic keratoconjunctivitis is a more serious disease, with edema of the eyelids, pain, lacrimation, and photophobia. Painful corneal opacities may occur, although they typically resolve over time. The disease typically involves just one eye, is self-limited, and rarely results in permanent corneal damage, although recovery may take up to 4 weeks.[7]

Gastrointestinal Tract Infections

Adenovirus can infect any part of the gastrointestinal tract and cause fever, nausea, vomiting, and diarrhea. Up to 10% of pediatric cases of diarrhea are caused by adenovirus, often with symptoms persisting for 8 to 12 days. Some adenoviruses can also cause mesenteric adenitis, which may clinically mimic appendicitis and rarely result in intussusception (Chapter 142). Adenovirus inclusions are seen in about one third to one half of appendices removed at surgery.

Genitourinary Tract Infections

Hemorrhagic cystitis, which typically manifests as microscopic or macroscopic blood in the urine and pain and cramping of the bladder, has been reported in children with AdV 11 and 21 infection. Adults can present with urethritis with or without cystitis due to AdV 19 and 37. Both diseases typically resolve over time without intervention.

Other Manifestations of Adenovirus

Meningitis (Chapter 412) and encephalitis (Chapters 383 and 414) have been reported occasionally in association with adenovirus infection. Adenovirus infection also can cause myocarditis and dilated myocardiopathy (Chapter 60).

A number of studies have found an association between detection of adenovirus in amniotic fluid associated with abnormal fetuses. In particular, echogenic liver lesions with or without hydrops and neural defects in fetuses were more common with patients with adenovirus detected in the amniotic fluid.

Adenovirus Diseases in Immunocompromised Patients

Primary adenovirus infection or reactivation of infection causes a wide range of infectious syndromes in immunocompromised children and adults. In the hematopoietic stem cell transplantation (HSCT) population (Chapter 178), the incidence of disease ranges from 3 to 47%.[8] Risk factors include an allogeneic transplant, T cell–depleted grafts, use of alemtuzumab, and acute graft-versus-host disease (GVHD).

Severe respiratory disease, hepatitis, colitis, hemorrhagic cystitis, and adenoviral keratoconjunctivitis can occur. Severe lymphopenia (<300 cells/μL) is associated with disseminated disease.[9] Although disseminated disease only effects 1 to 7% of HSCT recipients, it is associated with a significant risk of mortality (8 to 26%).

Among various solid organ transplant recipients, asymptomatic adenovirus viremia is common. However, progression to symptomatic disease is more common in small bowel and liver transplant recipients, pediatric transplant recipients, patients who receive antilymphocyte antibodies, and patients with donor-positive/recipient-negative adenovirus status.[10] Adenoviral hepatitis, typically caused by AdV 1, 2, or 5, is most common among liver transplant recipients and can be diagnosed by detection of viremia or visualization of viral intranuclear inclusions on biopsy. Adenovirus enterocolitis occurs in small bowel transplant recipients and may mimic rejection. Adenoviral pneumonia is associated with graft loss, death, or progression to obliterative bronchiolitis for lung transplant recipients. Adenovirus causes hemorrhagic cystitis with or without interstitial nephritis in kidney transplant recipients. Adenovirus DNA in biopsy specimens is associated with worse outcomes among pediatric heart transplant patients.

Primary adenoviral infections can cause severe, frequently fatal disease in children with primary immunodeficiency syndromes, including severe combined immunodeficiency disease (Chapter 250). Fatal cases have been reported in patients with acquired immunodeficiency syndrome (AIDS), but most are self-limited, particularly in patients on highly active antiretroviral therapy or without significant CD4 lymphopenia.

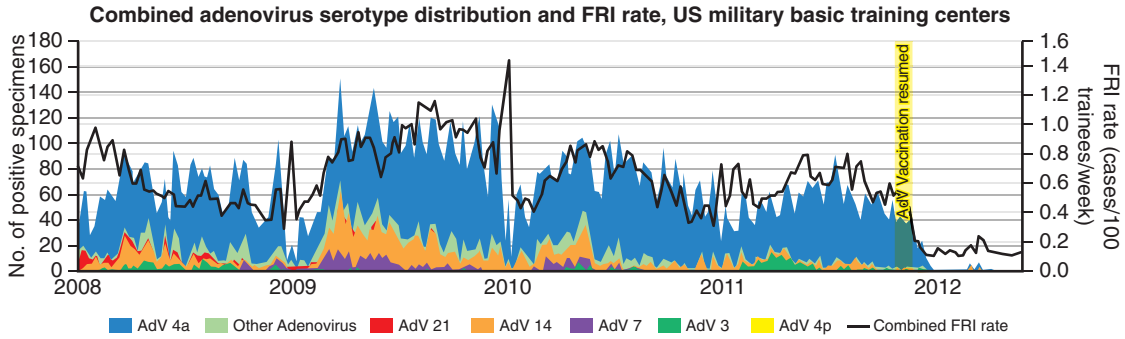

E-FIGURE 365-2. Adenovirus serotype distribution in U.S. military basic training centers before and after reintroduction of adenovirus vaccination. FRI = febrile respiratory illness. (From Hoke CH Jr, Snyder CE Jr. History of the restoration of adenovirus type 4 and type 7 vaccine, live oral [Adenovirus Vaccine] in the context of the Department of Defense acquisition system. *Vaccine.* 2013;31:1623-1632.)

DIAGNOSIS

Diagnosis of adenovirus depends on isolation of the virus from infected tissue and either histopathologic evidence of local replication or clinical symptoms consistent with infection. Detection of virus alone is not diagnostic of an adenovirus disease, because the virus can be latent in some tissues and can be intermittently shed from the throat or in the stool for months to years after primary infection. Quantitative polymerase chain reaction (PCR) has a higher diagnostic yield than culture or direct fluorescent antigen detection. Quantitative PCR assays are also useful to predict progression to disseminated disease in pediatric and, to a lesser extent, adult HSCT recipients and to determine response due to therapeutic interventions.

TREATMENT

Rx

No antiviral agents are specifically approved for the treatment of adenovirus. Cidofovir, which is a potent inhibitor of adenovirus in cell culture, has been used (either 5 mg/kg weekly for 2 weeks, then every other week or 1 mg/kg three times a week), but data suggest that its efficacy/toxicity (predominantly nephrotoxicity) ratio is too narrow to be clinically useful, except perhaps as preemptive treatment of pediatric HSCT patients with persistent or rising adenovirus viral loads. Ribavirin has not been helpful.

Brincidofovir, an experimental drug, is not nephrotoxic and is under active study in pediatric HSCT patients, in whom it has been generally well tolerated and shown to suppress adenovirus compared with placebo.[11] Further trials are ongoing. A newer approach using adenovirus- and multivirus-specific T cells is also being studied in prospective trials.[12]

PREVENTION

Strict attention to contact and droplet precautions can prevent health care–associated and institutional outbreaks of adenovirus infections, including epidemic keratoconjunctivitis. An oral live-attenuated AdV 4 and 7 vaccine is well tolerated, with a vaccine efficacy of 99.3%.[A1] Currently, the vaccine is only indicated and available to U.S. military personnel, ages 17 to 50 years.

PROGNOSIS

Prognosis is strongly linked to immune competence. Immunocompetent patients typically have self-limited illnesses that can be generally managed with symptomatic care, although some epidemics, including an AdV 14 epidemic, have resulted in high case-fatality rates.[13] In immunocompromised adults, persistent or rising adenovirus viremia predicts progressive infection and may encourage reduction of immunosuppression and institution of antiviral therapy in some cases, especially because viremia can clear without specific intervention when lymphocyte recovery occurs.

Grade A Reference

A1. Kuschner RA, Russell KL, Abuja M, et al. A phase 3, randomized, double-blind, placebo-controlled study of the safety and efficacy of the live, oral adenovirus type 4 and type 7 vaccine, in U.S. military recruits. *Vaccine*. 2013;31:2963-2971.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ghebremedhin B. Human adenovirus: viral pathogen with increasing importance. *Eur J Microbiol Immunol (Bp)*. 2014;4:26-33.
2. Hoke CH Jr, Snyder CE Jr. History of the restoration of adenovirus type 4 and type 7 vaccine, live oral (Adenovirus Vaccine) in the context of the Department of Defense acquisition system. *Vaccine*. 2013;31:1623-1632.
3. Arbizu RA, Aljomah G, Kozielski R, et al. Intussusception associated with adenovirus. *J Pediatr Gastroenterol Nutr*. 2014;59:e41.
4. Matthes-Martin S, Boztug H, Lion T. Diagnosis and treatment of adenovirus infection in immunocompromised patients. *Expert Rev Anti Infect Ther*. 2013;11:1017-1028.
5. Cao B, Huang GH, Pu ZH, et al. Emergence of community-acquired adenovirus type 55 as a cause of community-onset pneumonia. *Chest*. 2014;145:79-86.
6. Xie L, Yu XF, Sun Z, et al. Two adenovirus serotype 3 outbreaks associated with febrile respiratory disease and pharyngoconjunctival fever in children under 15 years of age in Hangzhou, China, during 2011. *J Clin Microbiol*. 2012;50:1879-1888.
7. Centers for Disease Control and Prevention. Adenovirus-associated epidemic keratoconjunctivitis outbreaks–four states, 2008-2010. *MMWR Morb Mortal Wkly Rep*. 2013;62:637-641.
8. Matthes-Martin S, Feuchtinger T, Shaw PJ, et al. European guidelines for diagnosis and treatment of adenovirus infection in leukemia and stem cell transplantation: summary of ECIL-4 (2011). *Transpl Infect Dis*. 2012;14:555-563.
9. Breuer S, Rauch M, Matthes-Martin S, et al. Molecular diagnosis and management of viral infections in hematopoietic stem cell transplant recipients. *Mol Diagn Ther*. 2012;16:63-77.
10. Florescu DF, Hoffman JA, for the AST Infectious Diseases Community of Practice. Adenovirus in solid organ transplantation. *Am J Transplant*. 2013;13(suppl 4):206-211.
11. Florescu DF, Pergam SA, Neely MN, et al. Safety and efficacy of CMX001 as salvage therapy for severe adenovirus infections in immunocompromised patients. *Biol Blood Marrow Transplant*. 2012;18:731-738.
12. Leen AM, Bollard CM, Mendizabal AM, et al. Multicenter study of banked third-party virus-specific T cells to treat severe viral infections after hematopoietic stem cell transplantation. *Blood*. 2013;121:5113-5123.
13. Lewis PF, Schmidt MA, Lu X, et al. A community-based outbreak of severe respiratory illness caused by human adenovirus serotype 14. *J Infect Dis*. 2009;199:1427-1434.

REVIEW QUESTIONS

1. In early October, a U.S. Army recruit in week 3 of his basic training exercises presents with a febrile respiratory illness consisting of minimally productive cough, fever, rhinorrhea, mild conjunctival injection, and cervical lymphadenopathy. A chest radiograph demonstrates bilateral scant infiltrates. Several other members of his troop have had a similar illness. He was appropriately vaccinated, including oral adenovirus vaccine. Which virus is likely the cause of this recruit's illness:

A. Influenza A/H1N1
B. Influenza A/H3N2
C. Adenovirus 14
D. Influenza B
E. Rhinovirus

Answer: C Adenovirus 14. Adenovirus is the most common cause of respiratory viral infection among military recruits. Typical symptoms include minimally productive cough, fever, rhinorrhea, mild conjunctival injection, and cervical lymphadenopathy. Influenza typically occurs from late November until early March. Although rhinovirus also circulates during this season, it is unlikely to cause an infection as severe as described. Additionally, because recruits are now getting adenovirus live-attenuated vaccine that includes types 3 and 7, adenovirus type 14 has become a frequent cause of infection.

2. A 30-year-old woman presents to your office with conjunctivitis, edema of the eyelids, pain, lacrimation, and photophobia. What is the most likely information to discuss with the patient?

A. The disease is likely caused by adenovirus 8, 19, or 37.
B. Blurry vision is common and likely will require missing work.
C. The problem will typically involve only one eye.
D. The infection is contagious, so careful hand hygiene and not sharing anything that touches her eye is critical.
E. All of the above

Answer: E All of the above. This is a case of epidemic keratoconjunctivitis (EKC). EKC typically is associated with a follicular conjunctivitis with edema of the eyelids, pain, lacrimation, and photophobia that develop after an 8- to 10-day incubation period. The hallmarks of the disease include significant pain and blurry vision that typically prevents patients from being able to work. Although bilateral disease has been described, involvement of a single eye is more typical. As symptoms improve over time, so will the vision, generally without any long-term effects on the eye or vision. The infection is highly contagious and can easily be spread to other members of the household or other patients if infection control is not fastidious.

3. An otherwise healthy 54-year-old woman presents to your office in July with a 3-day history of cough, nasal congestion, generalized malaise, fever, myalgias, tonsillitis, and cervical adenopathy. Her chest radiograph is normal. You perform a multiplex respiratory virus polymerase chain reaction (PCR), which returns positive for adenovirus. What do you recommend?

A. Admission to the hospital for evaluation of fever
B. A single dose of intravenous cidofovir (1 mg/kg)
C. Moxifloxacin 400 mg daily for 7 days for prevention of bacterial superinfection
D. Symptomatic treatment with cough drops and acetaminophen
E. Report the case to the health department for contact tracing.

Answer: D Symptomatic treatment with cough drops and acetaminophen. This is a classic case of adenovirus upper respiratory tract infection. This infection is associated with cough, nasal congestion, generalized malaise, fever, myalgias, tonsillitis, and cervical adenopathy; if there is concomitant conjunctivitis, the syndrome is termed *pharyngoconjunctival fever*. The infection is clinically indistinguishable from other viral causes of upper respiratory tract infection. Infection will typically resolve over the course of several days, and symptomatic treatment is all that is indicated. This patient has no signs of lower tract disease (pneumonia on chest radiograph, shortness of breath), so more severe infection is unlikely.

366

CORONAVIRUSES

SUSAN I. GERBER AND LARRY J. ANDERSON

DEFINITION

Human coronaviruses were, until 2003, recognized as a frequent cause of common cold symptoms, occasionally a cause of lower respiratory tract disease, but rarely if ever a cause of serious disease. In 2003, a novel coronavirus was recognized in humans from animals as the etiologic agent of the outbreak of severe acute respiratory syndrome (SARS).[1,2] The SARS outbreak demonstrated that coronaviruses can be serious human pathogens and led to discovery of other novel human coronaviruses as well as multiple novel coronaviruses in bats, the likely reservoir for SARS coronavirus. Furthermore, in 2012, the Middle East respiratory syndrome (MERS) coronavirus emerged and provided another example of coronavirus's ability to cause severe human disease.[3,4]

The Pathogens

Coronaviruses are members of the family Coronaviridae, which includes two subfamilies, Coronavirinae and Torovirinae. Coronaviruses are single-stranded positive-sense RNA viruses with a genome of approximately 30 kD, the largest genome among RNA viruses. These viruses were named coronaviruses because by electron microscopy they have club-shaped surface projections that give them a crown-like appearance. The genome encodes four or five structural proteins (a spike protein [S], a small envelope protein [E], a membrane protein [M], a nucleocapsid protein [N], and sometimes a hemagglutinin-esterase protein [HE]), a varying number of open reading frames scattered among the structural genes, and a polyprotein that is processed into multiple (usually 16) nonstructural proteins. These nonstructural proteins participate in virus replication but are not incorporated into the virion. Coronaviruses have also been isolated from a variety of animals and birds and, in their respective species, cause a wide range of respiratory, gastrointestinal (GI), neurologic, and systemic illnesses. The coronaviruses are divided into four genera: alpha, beta, gamma, and delta. The 229E and NL63 viruses are alphacoronaviruses, whereas OC43 and HKU1 are betacoronaviruses. SARS and MERS are both betacoronaviruses but belong to different lineages. Detection and characterization of novel coronaviruses in bats has greatly expanded our understanding of diversity among coronaviruses and will likely continue to do so.

EPIDEMIOLOGY

The common human coronaviruses—229E, OC43, NL63, and HKU1—appear to be transmitted through close contact that probably includes contamination of hands from person-to-person contact or from fomites, followed by autoinoculation to the mucosal surfaces of the mouth, nose, or eyes or inhalation of infectious droplets and possibly aerosols. Symptoms occur 2 to 4 days after infection. These coronaviruses are detected in patients with acute respiratory illnesses, most often a mild upper respiratory tract illness (i.e., common cold) but also in patients with more serious respiratory illnesses including pneumonia, bronchiolitis, and croup. Coronavirus infections are detected early in childhood, and repeated infections can occur throughout life. About 50% of children have antibodies against OC43 by 3 years of age, and about 70% of adults have such antibodies. Up to 75% of children have antibodies against NL63 and 229E by 3 to 4 years of age. Studies looking for 229E- and OC43-like infections suggest that coronaviruses are associated with about 15% of cases of the common cold and with up to 10% of cases of acute respiratory illnesses in children and adults. Individually, 229E, OC43, HKU1, and NL63 are detected in less than 1 to 4% of cases, and their individual contributions will vary by location and year. Serious illness has been reported in outbreaks among elderly patients in nursing homes. In one outbreak associated with OC43 infection, for example, 23 residents and 24 staff reported influenza-like illness, and three residents died. However, some reports have found rates of detection of coronaviruses among hospitalized children with acute respiratory illness and/or fever to be similar to the rates of asymptomatic controls, thereby raising questions about the virus's role in more severe disease and hospitalization. The 229E, OC43, NL63, and HKU1 coronaviruses can be detected throughout the year, but peak detection is often during fall and winter months in temperate climates. A second respiratory viral pathogen can be detected in 20 to 60% of specimens positive for one of these coronaviruses.

Most documented SARS coronavirus infections in humans occurred in persons ill with a SARS-like illness during the 2002-2003 global outbreak. It is likely that wild animal markets in Guangdong Province, China, played a key role in amplifying and introducing the virus into humans, but the original source of the outbreak virus was likely bats. Detection of multiple SARS-like and other coronaviruses in bats suggests that they are a rich source of coronaviruses. A coronavirus recently isolated from bats has 95% nucleotide sequence identity to the SARS viruses and can infect humans via the ACE2 receptor. Although animals were the original source of human infections, global spread of SARS coronavirus occurred through human-to-human

transmission and involved droplet, fomite transmission and, in some instances, probably small-particle aerosol transmission. Most transmission occurred within households, hospitals, or other health care facilities; little transmission occurred in the community. The fact that most transmission occurred after the patient's illness had become serious and required hospitalization helps explain the importance of transmission in health care settings and the relative lack of transmission in the community.

More recently, MERS coronavirus emerged in the Arabian peninsula, with the first documented cases retrospectively identified in April 2012 in Jordan. Like the SARS coronavirus, detection of coronaviruses similar to the MERS coronavirus suggests that bats may be a source of this virus. There is, however, likely an intermediate host species. MERS coronavirus gene sequences and antibodies that neutralize MERS coronavirus have been reported in dromedary camels. These data suggest a role for camels in the transmission of MERS coronavirus. Person-to-person spread of virus has been documented primarily among family members in households and within healthcare settings,[5] although as of December 2013, there has been no evidence of sustained community transmission.[6] The incubation period has been estimated to be just over 5 days (range, 2 to 14 days), and sporadic cases of MERS coronavirus infections continue to be seen.

TABLE 366-1 PERCENTAGE OF HOSPITALIZED SARS AND MERS CORONAVIRUS–INFECTED PATIENTS WITH SELECTED CLINICAL AND LABORATORY FEATURES OF SARS AND MERS CORONAVIRUS INFECTIONS

	SARS	MERS
Clinical or Laboratory Finding	***At Hospital Admission***	***At Presentation***
Fever	90-100%	≈90-100%
Cough or shortness of breath	40-75%	83%
Diarrhea	20-30%	26%
Chest radiograph abnormalities	65-90%	100%
Lymphopenia*	50-90%	34%

*Both $CD4^+$ and $CD8^+$ lymphocyte counts are decreased.
MERS = Middle East respiratory syndrome; SARS = severe acute respiratory syndrome.
From Assiri A, Al-Tawfiq JA, Al-Rabeeah AA, et al. Epidemiological, demographic, and clinical characteristics of 47 cases of Middle East respiratory syndrome coronavirus disease from Saudi Arabia: a descriptive study. *Lancet Infect Dis.* 2013;13:752-761.

PATHOBIOLOGY

The human coronaviruses characterized to date infect humans through the respiratory tract. The sites where the virus then replicates is determined at least in part by which cells express the respective receptors. The receptors for 229E and NL63 coronaviruses are aminopeptidase N and angiotensin-converting enzyme 2 (ACE2), respectively. The receptors for OC43 and HKU1 coronaviruses have not yet been determined, but OC43 may use several cell surface molecules as receptors, including 9-*O*-acetylated neuraminic acid. The primary receptor for SARS coronavirus is ACE2, but the virus also binds to two C-type lectins expressed on dendritic cells, DC-SIGN and L-SIGN. Aminopeptidase N is expressed in various cells, including respiratory, GI, kidney epithelial, and myeloid cells, but 229E is known to infect only respiratory epithelial cells. ACE2 is found in various tissues, including the lung, GI tract, heart, and kidneys. The SARS coronavirus has consistently been detected in pneumocytes in the lung and enterocytes in the GI tract and is occasionally found in other cells, including distal tubular cells in the kidney and macrophages in various tissues. Autopsy studies suggest that infection in the lung leads initially to diffuse alveolar damage and later may lead to a repair process that includes fibrosis in the alveolar walls. It is not known whether NL63, which also uses ACE2 as its receptor, infects sites other than the respiratory tract. The MERS coronavirus receptor is the exopeptidase dipeptidyl peptidase 4 (DPP4), also known as CD26. DPP4 is found on many different cell types including non-ciliated bronchial epithelial cells, bronchiolar epithelial cells, alveolar epithelial cells, endothelial cells, and lung ex vivo organ cultures. In addition, DPP4 is expressed on the epithelial cells in kidney, small intestine, liver, and prostate as well as in activated leukocytes.

It is likely that the illness associated with coronavirus infections results from both the cytopathic effect of the virus and the host immune and inflammatory response to the viral infection. How this interplay contributes to disease, however, is not understood. The biphasic course of SARS in some patients, with the onset of severe disease in the second week of illness and the decrease in lymphocyte numbers, suggest a role for the host response and virus-induced immune suppression in the disease process. It is likely that the host response and virus-induced immune suppression also contribute to MERS coronavirus disease.

CLINICAL MANIFESTATIONS

229E, OC43, NL63, and HKU1 coronavirus infections are commonly associated with acute respiratory illnesses that are usually mild and consistent with the common cold (Chapter 361) but can also result in the full range of acute respiratory illnesses, including pneumonia (Chapter 97), croup (Chapter 429), bronchiolitis (Chapter 92), and bronchitis (Chapter 96). The best studied of these coronaviruses, human coronaviruses 229E and OC43, cause respiratory symptoms (e.g., rhinorrhea, nasal congestion, sore throat, cough) as well as systemic symptoms (e.g., fever, headache, malaise) when they are inoculated intranasally into adult volunteers. Symptoms develop 2 to 4 days after inoculation, but about 30% of volunteers who excrete virus have no associated illness. Symptoms usually persist for about 1 week but sometimes for as long as 3 weeks. Previous infection does not induce high levels of protective immunity. Humans can be reinfected with respiratory coronaviruses throughout life, and human volunteers can be symptomatically reinfected with the same strain of coronavirus 1 year after the first infection. As with other infections, the severity of disease varies among individual patients during the same outbreak and among groups of patients during different outbreaks in the same community.

In contrast to the mild illness associated with 229E and OC43, SARS coronavirus infection nearly always results in a serious illness that requires hospitalization, often in an intensive care unit (ICU), and a high fatality rate. Radiologic evidence of pneumonia was seen in nearly all SARS coronavirus–infected persons, and acute respiratory distress syndrome (Chapter 104) requiring admission to an ICU and mechanical ventilation developed in 20% or more of patients. The initial clinical manifestation of SARS was often systemic symptoms of fever, malaise, and myalgias from 2 to 10 days (rarely > 10 days) after exposure. Several days after the onset of systemic symptoms, lower respiratory tract symptoms of nonproductive cough and shortness of breath were noted. Unlike patients with other respiratory virus infections, the majority of patients never experience upper respiratory tract symptoms such as rhinorrhea, sore throat, or nasal congestion (Table 366-1). As with other infections, the expected signs and symptoms with SARS coronavirus infection initially may be obscured or not present in elderly patients or patients with underlying chronic illnesses. During the course of their illness, most SARS coronavirus–infected patients had elevated liver enzyme levels and lymphopenia, including a substantial drop in numbers of both $CD4^+$ and $CD8^+$ T cells. SARS coronavirus–infected patients with severe complications also often suffered complications associated with intensive supportive care, such as secondary bacterial infections. A more severe illness with infection has been associated with older age, underlying chronic illness, higher liver enzyme values, lower lymphocyte levels and platelet counts, and higher titers of virus or viral RNA. Asymptomatic or mild illness occurred but was uncommon. In studies of exposed health care workers and patients who did not have a SARS-like illness, less than 1% of those tested had definitive serologic evidence of SARS coronavirus infection (i.e., a positive neutralization test result). In general, children had less severe illness than adults did.

The clinical spectrum of MERS coronavirus illness ranges from asymptomatic infection to severe illness.[7] Symptoms include cough, fever, malaise, chills, arthralgias, rigors, and dyspnea. Approximately 25% of patients have GI symptoms that include diarrhea, vomiting, and abdominal pain. Patients who are severely ill have pneumonia that sometimes progresses to acute respiratory distress syndrome.[8]

Laboratory findings include leukopenia and lymphopenia, and some patients have thrombocytopenia and abnormal liver enzymes. Chest radiographic findings have included patchy infiltrates, lobar opacities, and similar to the SARS coronavirus, a ground-glass appearance.

DIAGNOSIS

Because illness is usually mild and there is no effective treatment, the diagnosis of 229E, OC43, NL63, and HKU1 coronavirus infections have not been important to the management of patients. The accurate diagnosis of SARS and MERS coronavirus infections, however, are critical for the management of individual patients and for the public health response to

reemergence of SARS coronavirus and the current circulation of MERS coronavirus in the Middle East.

229E, OC43, HKU1, and NL63 Coronavirus Infections

Coronavirus polymerase chain reaction (PCR) assays, which can be designed to detect both known and novel coronaviruses, are the assays of choice for diagnosis of infection. Most coronavirus PCR assays are type specific—that is, specific to MERS coronavirus, SARS coronavirus, 229E, OC43, HKU1, or NL63 RNA. Coronavirus diagnostic assays are becoming more generally available and are sometimes part of a PCR panel designed to detect respiratory viruses. Presence of the virus can also be inferred by electron microscopy and confirmed by in situ or immunohistologic assays of affected tissues. Positive immunohistologic and in situ hybridization studies document the site of infection and help support a link between the virus and the disease process. A variety of enzyme or fluorescent immunoassays for antibodies have been used successfully to detect infection. Most assays detect immunoglobulin (Ig)G antibodies, but virus neutralization antibody assays are more specific. The need to detect a diagnostic rise in antibodies between acute and convalescent serum specimens for 229E, OC43, HKU1, and NL63 coronavirus infections means that serologic tests are not helpful for managing an acute illness but can be helpful for epidemiologic studies.

SARS

Three features of SARS cases help guide approach to its diagnosis. First, SARS has occurred only in persons who have some potential exposure—that is, to patients with SARS, to a location with SARS transmission, to a laboratory working on SARS coronavirus, or to a setting where SARS-infected animals might be located (e.g., southern China). Second, nearly 100% of infected patients develop chest radiographic abnormalities by day 10 of their illness. Finally, SARS nearly always develops within 10 days of exposure. Thus, a suspicion of SARS and a diagnostic evaluation can be limited to patients who have a severe lower respiratory tract illness and some potential exposure to SARS within 10 days before the onset of illness.

Laboratory confirmation of SARS coronavirus infection early in the illness proved to be problematic even with sensitive real-time PCR assays. Unlike in most respiratory viral infections, the highest titer of virus or viral RNA was found in clinical specimens from the second week of illness. During the first week of illness, the best way to detect infection is by a sensitive PCR assay or a sensitive enzyme immunoassay for N protein antigen applied to respiratory and serum specimens. During the second week of illness, respiratory and stool specimens are most likely to be positive for viral RNA. Because most populations are negative for SARS coronavirus antibodies, detection of serum SARS coronavirus antibodies is a reliable indicator of infection. Antibodies were sometimes detected early in the second week of illness but at times were not detected until 4 weeks into the illness. Because antibodies persist in serum, a specimen can be collected late without compromising the ability to detect antibodies. Serologic assays, unlike PCR assays, are not confounded by problems with specimen contamination. Because antibodies to SARS coronavirus were rarely present before the 2003 outbreak, a single positive antibody test result from an ill person could be considered diagnostic of an acute SARS coronavirus infection. However, because the reemergence of SARS coronavirus will have substantial public health, social, and economic impact, and because of occasional cross-reacting antibodies induced by other coronaviruses, a neutralization antibody test and confirmatory testing by a reference laboratory are required to confirm the diagnosis. Public health departments should be consulted for questions about SARS diagnostic tests.[9]

MERS

Though less is known about transmission of MERS coronavirus, an approach similar to SARS is appropriate. A diagnosis of MERS coronavirus infection should be considered in patients with severe acute respiratory infection of unknown cause and a possible exposure or epidemiologic link to the Middle East. PCR assays have been used to detect RNA in upper and lower respiratory tract specimens (nasopharyngeal, oropharyngeal, sputum, tracheal aspirate, and bronchoalveolar lavage fluid), serum, stool, and urine. A confirmed case of MERS coronavirus infection requires a positive PCR on at least two specific gene targets or a single positive target with sequencing on a second.

A number of serologic assays can detect antibodies to the nucleocapsid and spike proteins. The sensitivity and specificity of these assays for diagnosing current or past MERS coronavirus infection have not yet been determined. Public health departments should be consulted for questions about MERS diagnostic tests.

TREATMENT

Rx

There is no virus-specific treatment for 229E, OC43, HKU1, and NL63 coronavirus infections, but the illnesses are mild and usually resolve in a few days to a week. Patients require symptomatic therapy or, uncommonly, management of complications of infection.

Treatment of SARS and MERS coronavirus infections are more complex. Currently, no antiviral drug has proved to be effective. With the high death rate for both of these viruses and the lack of clinical or in vitro data to guide treatment, supportive measures, including mechanical ventilation and oxygenation regimens (Chapter 104), are used. In the SARS coronavirus outbreak, in vitro data showed little if any antiviral effect with ribavirin and suggested that interferon alfa, SARS convalescent phase immune globulin, and lopinavir plus ritonavir might have been useful. Although many people were treated during the outbreak, lack of control groups makes it impossible to determine which if any therapies were beneficial. For MERS coronavirus, in vitro data demonstrate inhibitory effects for a number of antiviral agents, including interferons, ribavirin, cyclosporine A, and mycophenolic acid. However, no data are currently available regarding their efficacy for treating human infection.

PREVENTION

Handwashing and other infection-control measures probably decrease the spread of coronaviruses in the home, health care facilities, and other settings. These strategies focus on reinforcing the need for patients with respiratory illnesses to cover the nose and mouth when coughing or sneezing, to use tissues to contain respiratory secretions, and to wash the hands after contact with respiratory secretions. Staff should use good infection-control practices.

Within 4 months of the initiation of the 2003 SARS outbreak, the outbreak was contained and human-to-human transmission stopped without a vaccine or effective antiviral therapy but thorough implementation of the classic public health tools of case finding, isolation, and contact tracing and management, including quarantine of contacts. The low risk of SARS transmission early in the illness and the very low rate of asymptomatic infection were key factors to the success of these preventive measures.

The cases of laboratory-acquired SARS coronavirus infection and the subsequent transmission of disease to others after one such case reinforces the importance of strict attention to safe laboratory practices. Because the reemergence of SARS could lead to global spread, the local, national, and global public health and health care communities must be alerted quickly and updated regularly about new cases and the status of transmission.

Strict attention to standard contact and airborne precautions are recommended for SARS and MERS coronavirus–infected patients within hospital settings. MERS coronavirus infections continue to occur in the Middle East, and local, national, and global public health and health care communities should be immediately notified of a case. Vaccine development is underway for both SARS and MERS coronaviruses.

PROGNOSIS

Patients with typical community-acquired coronavirus infections typically recover completely. However, patients with compromised cardiac, pulmonary, or immune systems are at increased risk of more serious lower respiratory tract illness, and outbreaks of human coronavirus infections in elderly patients in chronic care facilities can cause severe lower respiratory illnesses and deaths. In the SARS outbreak, nearly 10% of patients died. The death rate was especially high, approaching 50%, in elderly patients and patients with underlying illnesses. Although most survivors of SARS coronavirus infection appeared to achieve full recovery, as many as 25% had abnormal pulmonary findings such as ground-glass opacities on chest radiograph or abnormal pulmonary function test results (e.g., decreased diffusing capacity) 6 months or more after their illness. The MERS fatality rate initially was reported to be about 40%, but many of the patients were older adults with comorbid conditions. The case-fatality rate will likely be lower as less severe cases are identified.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. http://www.cdc.gov/sars/index.html. Accessed March 11, 2015.
2. www.who.int/csr/sars/en/. Accessed March 11, 2015.
3. http://www.cdc.gov/coronavirus/mers/index.html. Accessed March 11, 2015.
4. http://www.who.int/csr/disease/coronavirus_infections/en/. Accessed March 11, 2015.
5. Breban R, Riou J, Fontanet A. Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. *Lancet*. 2013;382:694-699.
6. Drosten C, Meyer B, Muller MA, et al. Transmission of MERS-coronavirus in household contacts. *N Engl J Med*. 2014;371:828-835.
7. Assiri A, Al-Tawfiq JA, Al-Rabeeah AA, et al. Epidemiological, demographic, and clinical characteristics of 47 cases of Middle East respiratory syndrome coronavirus disease from Saudi Arabia: a descriptive study. *Lancet Infect Dis*. 2013;13:752-761.
8. Arabi YM, Arifi AA, Balkhy HH, et al. Clinical course and outcomes of critically ill patients with middle East respiratory syndrome coronavirus infection. *Ann Intern Med*. 2014;160:389-397.
9. Graham RL, Donaldson EF, Baric RS. A decade after SARS: strategies for controlling emerging coronaviruses. *Nat Rev Microbiol*. 2013;11:836-848.

REVIEW QUESTIONS

1. Middle East respiratory syndrome (MERS) coronavirus may be diagnosed from which patient specimens?

A. Respiratory tract specimen
B. Serum
C. Stool
D. Urine
E. All of the above

Answer: E Virus has been demonstrated to be detectable in all of these patient specimens. Although respiratory tract specimens are most important for diagnosis, specimens from additional sites may aid in the diagnosis of patients.

2. What statement is most true about common human coronaviruses (OC43, 229E, NL63, and HKU1)?

A. Whenever a common human coronavirus is detected, there are associated severe clinical symptoms.
B. Occasionally, common human coronaviruses are detected in individuals with mild or no clinical symptoms.
C. All common colds are associated with common human coronaviruses.
D. Common human coronaviruses have never been associated with croup or pneumonia.

Answer: B Human coronaviruses may be detected from clinical specimens from mildly ill or asymptomatic individuals. Human coronaviruses are only one cause of the common cold, and these viruses also have been associated with croup and pneumonia.

367

MEASLES

MARTIN WEISSE AND MARK PAPANIA

DEFINITION

Measles (synonyms: rubeola, red measles, hard measles, 7-day measles) is a highly contagious systemic viral disease characterized by generalized maculopapular rash, fever, cough, coryza (rhinitis), and conjunctivitis. The measles virus is a negative-stranded RNA virus of the genus *Morbillivirus*, family Paramyxoviridae, with only one serotype and 24 genotypes.[1]

EPIDEMIOLOGY

In the prevaccine era, measles was a nearly universal rite of childhood, affecting almost every child by 15 years of age. As a result of high vaccination coverage, measles has not been endemic in the United States for more than a decade (incidence < 1 case/million people).[2] However, measles cases still occur annually in the United States because of the importation of the virus from other countries, followed by limited local transmission.[3] Many other countries have eliminated endemic measles, including almost all countries in the Americas. Unfortunately, measles remains endemic in many developed and developing countries and still caused an estimated 145,000 deaths worldwide in 2013, down from an estimated 545,000 deaths in 2000. Countries in the Southeast Asian region now account for about 25% of measles mortality, whereas countries in the African region account for about 50%. An estimated 10,000 measles deaths occurred in India in 2013, the highest in any country and about 7% of the global total estimated deaths.[4]

Measles, which is one of the most contagious diseases, infects more than 90% of susceptible people who come into close contact with an infectious patient. The measles virus is transmitted mainly by direct contact with large respiratory droplets, but airborne spread is also possible. Persons with measles are considered infectious from 4 days before the onset of the rash until 4 days after its onset. Immunocompromised patients may shed virus for a prolonged time and should be considered infectious as long as they have catarrhal symptoms.

Measles infection and measles vaccination usually confer lifelong immunity. However, antibody titers are lower following vaccination and wane over time in the absence of reexposure.[5] Serologic surveys in the United States show a measles antibody seroprevalence of over 95% among people older than age 6, which is above the threshold population immunity level (90 to 95%) required to sustain interruption of endemic measles transmission. Maternal measles antibody, which is transferred through the placenta, protects infants of immune mothers during the first months of life. However, lower antibody titers among mothers born in the vaccine era result in less antibody transfer and increased susceptibility to measles among infants.

PATHOBIOLOGY

Measles infection usually occurs through inhalation of infectious droplets into the respiratory epithelium or direct inoculation onto the conjunctiva. The virus attaches to host cells by its hemagglutinin protein, and its fusion protein allows viral entry and facilitates spread between cells. The virus initially replicates in the respiratory epithelial tissues or the adjacent lymph nodes. A primary viremia disseminates the virus to the reticuloendothelial tissues of the tonsils, lungs, gastrointestinal tract, spleen, and distant lymph nodes. Several days later, the onset of prodromal symptoms heralds a secondary viremia that results in further spread of the virus to the skin, viscera, kidney, and bladder. Histologic changes in the skin and mucous membranes consist of epithelial giant cells with surrounding mononuclear infiltrates and edema. Koplik spots, which are inflammatory lesions of the buccal submucous glands, have similar histologic features.

The development of immunoglobulin (Ig)M antibody coincides with the appearance of the rash; during the next week, IgG antibodies appear and usually persist for life. Cell-mediated immunity is also critical for resolution of acute infection. Measles virus infection causes a significant immunosuppression that predisposes infected patients to secondary bacterial and viral infections, which increase the mortality rate of measles. The exact mechanisms of immunosuppression are unknown.

CLINICAL MANIFESTATIONS

The incubation period of measles is usually 8 to 12 days after exposure until the development of the first symptoms. The measles prodrome includes fever associated with cough, coryza, conjunctivitis, and Koplik spots, often accompanied by malaise, myalgia, and headache. Fever increases with the severity of catarrhal symptoms and peaks at the onset of the exanthem; it frequently reaches 40° or 41° C and may persist for 6 days. Koplik spots, an enanthem of small bluish white spots on a red background on the buccal mucosa lateral to the molar teeth (Fig. 367-1), are considered to be pathognomonic for measles.[6] Koplik spots precede the rash by 1 day and are no longer evident 2 days into the rash.

The measles rash (Fig. 367-2) typically arises 2 to 6 days after the onset of the catarrhal symptoms. It begins on the face or behind the ears as individual red macules that blanch with pressure. During 12 to 24 hours, the macules become papular and tend to coalesce. The rash progresses from head to trunk to extremities and maintains the same pattern of progression from macules to coalescing papules. By the fourth day of the rash, the entire body is typically involved, although the hands and feet may be spared. The rash fades in the same head-to-toe order in which it appeared and often peels with a brawny desquamation during the subsequent weeks. In darkly pigmented individuals, the erythema may be difficult to appreciate, and the rash may be better felt than seen.

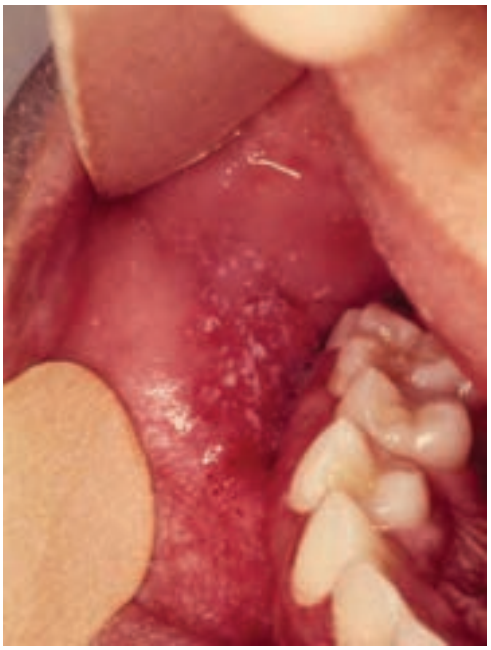

FIGURE 367-1. Typical Koplik spots.

FIGURE 367-2. Typical morbilliform eruption of measles, with broad areas of confluent red maculopapular lesions interspersed with islands of normal skin.

Important variations in the classic presentation of measles include "modified measles" in patients who have received immune globulin prophylaxis and "mild measles" in previously vaccinated persons. In both variations, any of the classic symptoms may be attenuated or absent. Such cases are extremely difficult to diagnose in the absence of known exposure to measles. Also, immunocompromised patients can have severe measles pneumonia without rash. Measles should be considered in any immunocompromised patient with pneumonia of unknown etiology.

Complications

The most common complications of measles are diarrhea, otitis media, and pneumonia. Diarrhea occurs early in the disease and is probably the result of intestinal measles virus infection. Measles-associated diarrhea significantly contributes to morbidity and mortality in malnourished patients.[7] Pneumonia is the most severe common complication and is responsible for the majority of deaths associated with measles. Measles-associated pneumonia may be a giant cell pneumonitis consistent with direct infection by measles virus or be caused by a secondary bacterial pathogen, most often *Staphylococcus aureus* (Chapter 288) or *Streptococcus pneumoniae* (Chapter 289).

Overall, children younger than 5 years and adults older than 20 years have the highest complication rates. Complication rates do not vary substantially by sex except that pregnant women are at high risk. Measles is especially hazardous during the first trimester, often leading to fetal demise. Later in pregnancy, the rate of prematurity is high, as is maternal pneumonia.

Malnutrition is associated with increased fatality rates, and children with vitamin A deficiency are at high risk for severe keratoconjunctivitis, which can lead to blindness. Immunocompromised patients (e.g., those with human immunodeficiency virus [HIV] infections, with cancer, or receiving chemotherapy or immunosuppressive drugs such as steroids) are at higher risk for severe measles complications. Interstitial pneumonia occurs 1 to 2 weeks after the onset of measles and is the most common cause of measles-related death in immunocompromised patients. Measles inclusion body encephalitis occurs in immunocompromised patients 1 to 6 months after acute measles and is usually fatal.

Postinfectious encephalomyelitis occurs in 1 in 1000 measles cases. Patients present during the 2 weeks after the onset of the rash with headache, recurrence of fever, vomiting, and stiff neck. Up to 25% of patients die, and 33% of survivors suffer neurologic sequelae. Elevated IgG and IgM can be found in cerebrospinal fluid, but measles virus cannot be isolated, thereby suggesting an autoimmune pathologic process. A rare late central nervous system complication of measles is subacute sclerosing panencephalitis (Chapter 370), a degenerative demyelinating disease due to chronic infection. Subacute sclerosing panencephalitis occurs years after the acute measles infection and is almost universally fatal.[8]

DIAGNOSIS

In many countries, measles has become such a rare disease that most physicians have never seen a patient. One key to making the diagnosis is to consider measles in any patient with a generalized maculopapular rash and fever. The index of suspicion should be raised if cough, coryza, and conjunctivitis are present and if measles virus transmission is occurring in the area or if the patient has traveled internationally or had any association with international travelers. Although Koplik spots are considered pathognomonic, the absence of Koplik spots does not exclude the diagnosis. The lack of vaccination should increase the level of suspicion, but measles can occur despite two documented doses of measles vaccine.

Confirmation of measles requires laboratory testing. The standard laboratory test for measles is an IgM measles antibody assay on a single serum specimen. However, during the first 3 days of rash, up to 25% of cases may not have detectable IgM; the test can be repeated if the rash and fever persist for 3 days. Acute and convalescent sera to test for a rise in IgG antibody are sometimes useful. Urine, throat swabs, or blood specimens can be sent for viral culture or for detection of viral RNA by reverse transcription–polymerase chain reaction (RT-PCR).[9] Genotype sequence analysis aids in tracking the source of the virus, monitoring transmission patterns, and distinguishing a mild case of measles from a vaccine reaction in patients who have recently received measles vaccine.

Differential Diagnosis (Table 367-1)

Other viral infections that may be confused with measles are rubella (Chapter 368), human herpesvirus 6 and 7 infection (Chapter 374), infectious mononucleosis (Chapter 377), and parvovirus B19 infection (erythema infectiosum [Chapter 371]). In general, these infections have a milder course. Bacterial infections that may mimic measles include secondary syphilis (Chapter 319), scarlet fever (Chapter 290), streptococcal or staphylococcal toxic shock syndrome (Chapter 439), and leptospirosis (Chapter 323). Drug eruptions (Chapter 440) and Kawasaki disease (Chapter 270) must also be considered. It can be extremely difficult to distinguish measles vaccine reactions, which may include fever and rash, from a mild case of measles. In highly vaccinated populations, measles is rare and clinical symptoms may be attenuated; serologic and virologic assays are critical in distinguishing measles from other diseases and conditions.

TREATMENT Rx

Vitamin A can reduce blindness due to severe keratoconjunctivitis and decrease mortality by as much as 50%. Vitamin A treatment is recommended for all children with measles, at doses of 50,000 IU for infants younger than 6 months of age, 100,000 IU for children 6 to 12 months of age, and 200,000 IU for children older than 1 year of age. An additional age-specific dose should be given 2 to 4 weeks later to children with clinical evidence of vitamin A deficiency. There are no recommendations for or against vitamin A treatment of adults with measles.

No specific antiviral or other therapies are recommended for measles. Ribavirin, isoprinosine, and interferon alfa have demonstrated modest benefit in small studies, but no controlled trials have been conducted. For most patients, treatment of measles consists of symptomatic therapy with antipyretics and analgesics and encouraging fluids to maintain hydration.

PREVENTION

Vaccination

Live-attenuated measles virus vaccines (Chapter 18) are the best means of preventing measles. Although the vaccine virus strains vary from different manufacturers, all provide equivalent protection. Measles vaccine is typically given as a component of the combination measles-mumps-rubella (MMR) vaccine or the measles-mumps-rubella-varicella (MMRV) vaccine. Measles-containing vaccines can be given at the same time as other live or killed

TABLE 367-1 A GUIDE TO THE DIFFERENTIAL DIAGNOSIS OF MEASLES

	CONJUNCTIVITIS	RHINITIS	SORE THROAT	ENANTHEM	LEUKOCYTOSIS	DIAGNOSTIC TESTS AVAILABLE
Measles	++	++	0	+	0	Yes
Rubella	±	0	0	0	0	Yes
Exanthem subitum	0	0	0	0	0	Yes
Enterovirus infection	0	0	±	±	0	Yes
Adenovirus infection	+	+	+	0	0	Yes
Scarlet fever	0	0	++	+	+	Yes
Infectious mononucleosis	0	0	++	+	±	Yes
Drug rash	0	0	0	0	0	No

0 = not usually present or no test available; ± = variable in occurrence; + = present: test available (virus or bacterial culture, serology); ++ = present and severe.
Modified from Brunell PA. Measles. In: Ausiello D, Goldman L, eds. *Cecil Textbook of Medicine.* 23rd ed. Philadelphia: Saunders Elsevier; 2008:2477.

vaccines. In the United States, the first dose of measles vaccine is usually given at 12 to 15 months of age, but it can be given as early as 6 months of age in infants who have a high risk of measles exposure, such as with international travel. A second dose is recommended when a child enrolls in school at 4 to 6 years of age. HIV-infected children should be vaccinated on the same schedule if they are not severely immunosuppressed. Children with perinatal HIV infection who received measles vaccination before they received antiretroviral therapy (ART) should be revaccinated with two doses after effective ART has been established.[10,11]

The first dose of vaccine is given at 9 months of age in developing countries where transmission in infancy is widespread and mortality is high. The World Health Organization recommends infants with known or suspected HIV infection also should be vaccinated at age 6 months.■ (Chapter 18). Most countries also recommend a routine second dose at various ages or provide a second opportunity for vaccination in mass campaigns to ensure protection. However, a high rate of first-dose coverage is always the top priority.

Contraindications to measles vaccination include pregnancy, immunodeficiency, leukemia and other immunosuppressive diseases, administration of immunosuppressive drugs such as systemic corticosteroids or antimetabolites, and untreated active tuberculosis. Extreme caution should be used in considering vaccination in patients with severe allergies to gelatin, neomycin, or other vaccine components or a history of thrombocytopenic purpura. Persons with HIV infection should be vaccinated against measles unless they have signs of severe immunosuppression.

Common adverse events after vaccination with measles vaccine include transient fever and mild rash, each of which occurs in roughly 5% of vaccine recipients. More serious adverse events, which are rare, include febrile seizures (<1 case/10,000 doses) and immune thrombocytopenic purpura (<2 cases/100,000 doses).[12,13]

Prevention of Measles Virus Transmission

To prevent transmission of measles, contact should be minimized between persons with suspected or confirmed measles and all persons who might be susceptible to measles; isolation and airborne precautions are indicated for hospitalized patients. When a case of measles is suspected, the local health department should be contacted immediately to assist with measures to prevent virus transmission.

Postexposure Prophylaxis

Unless otherwise contraindicated, vaccination is recommended for susceptible persons aged 6 months or older who are identified within 72 hours of exposure to measles to prevent or alter the course of the disease.■ Immune globulin should be given to susceptible household contacts of measles cases within 6 days of exposure, especially contacts who are too young (birth to 6 months) for measles vaccination, have contraindications to vaccination (e.g., pregnancy, immunodeficiency), or were not identified within 72 hours of exposure. Immune globulin may prevent measles or result in a less severe case of measles.

PROGNOSIS

In well-nourished populations with access to modern medical care, the case-fatality rate is typically 1 to 3 deaths/1000 cases, but case-fatality rates as high as 20% are seen in some developing countries as a result of malnutrition, limited access to health care, and immunodeficiency. Most measles deaths are associated with pneumonia or encephalitis. Antimicrobial treatment of secondary infections, especially pneumonias, can significantly reduce measles deaths. Vitamin A deficiency is associated with more severe disease, blindness, and increased mortality from respiratory and diarrheal causes, but vitamin A supplementation can reduce the severity of these complications. Patients with HIV infection and other immunocompromising conditions often have more severe cases with higher complication rates and case-fatality rates.

Grade A References

A1. Sudfeld CR, Navar AM, Halsey NA. Effectiveness of measles vaccination and vitamin A treatment. *Int J Epidemiol*. 2010;39(suppl 1):i48-i55.

A2. Chandwani S, Beeler J, Li H, et al. Safety and immunogenicity of early measles vaccination in children born to HIV-infected mothers in the United States: results of Pediatric AIDS Clinical Trials Group (PACTG) protocol 225. *J Infect Dis*. 2011;204(suppl 1):S179-S189.

A3. Young MK, Nimmo GR, Cripps AW, et al. Post-exposure passive immunisation for preventing measles. *Cochrane Database Syst Rev*. 2014;4:CD010056.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Moss WJ, Griffin DE. Measles. *Lancet*. 2012;379:153-164.
2. Papania MJ, Wallace GS, Rota PA, et al. Elimination of endemic measles, rubella, and congenital rubella syndrome from the Western hemisphere: the US experience. *JAMA Pediatr*. 2014;168: 148-155.
3. Gastañaduy PA, Redd SB, Fiebelkorn AP, et al. Measles—United States, January 1-May 23, 2014. *MMWR Morb Mortal Wkly Rep*. 2014;63:496-499.
4. Perry RT, Gacic-Dobo M, Dabbagh A, et al. Progress toward regional measles elimination—worldwide, 2000-2013. *MMWR Morb Mortal Wkly Rep*. 2014;63:1034-1038.
5. Scalia Tomba G, Manfredi P. Quantifying the re-exposure process to an infectious agent. Measles and varicella as examples. *Math Biosci*. 2013;245:31-39.
6. Lefebvre N, Camuset G, Bui E, et al. Koplik spots: a clinical sign with epidemiological implications for measles control. *Dermatology*. 2010;220:280-281.
7. Mahamud A, Burton A, Hassan M, et al. Risk factors for measles mortality among hospitalized Somali refugees displaced by famine, Kenya, 2011. *Clin Infect Dis*. 2013;57:e160-e166.
8. Schonberger K, Ludwig MS, Wildner M, et al. Epidemiology of subacute sclerosing panencephalitis (SSPE) in Germany from 2003 to 2009: a risk estimation. *PLoS ONE*. 2013;8:e68909.
9. Michel Y, Saloum K, Tournier C, et al. Rapid molecular diagnosis of measles virus infection in an epidemic setting. *J Med Virol*. 2013;85:723-730.
10. Abzug MJ, Qin M, Levin MJ, et al. Immunogenicity, immunologic memory, and safety following measles revaccination in HIV-infected children receiving highly active antiretroviral therapy. *J Infect Dis*. 2012;206:512-522.
11. McLean HQ, Fiebelkorn AP, Temte JL, et al. Prevention of measles, rubella, congenital rubella syndrome, and mumps, 2013: summary recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm Rep*. 2013;62:1-34.
12. Rowhani-Rahbar A, Fireman B, Lewis E, et al. Effect of age on the risk of fever and seizures following immunization with measles-containing vaccines in children. *JAMA Pediatr*. 2013;167:1111-1117.
13. O'Leary ST, Glanz JM, McClure DL, et al. The risk of immune thrombocytopenic purpura after vaccination in children and adolescents. *Pediatrics*. 2012;129:248-255.

REVIEW QUESTIONS

1. A 21-year-old man presents to your travel clinic complaining of fever to 104° F and rash. He returned 5 days ago from a 1-month trip to Indonesia, where he served in a remote village. He was scrupulous about taking his malaria prophylaxis but was bothered by many mosquito bites per day. His fever started 3 days ago and was accompanied by headache, runny nose, and watery eyes. The rash started on his trunk 2 days ago and is also evident on his upper extremities. He was vaccinated for measles, mumps, and rubella at age 1 and again at age 5. The most likely etiology of his fever and rash are:

A. Measles
B. Malaria
C. Dengue
D. Parvovirus B19
E. Rubella

Answer: C Dengue is the most likely cause of this young man's illness. Dengue is spread by mosquitoes and is the most common arboviral infection in the world. There is no specific therapy, so treatment is supportive with analgesics, hydration, and antipyretics. Measles is not very likely because the patient had received two MMR vaccines, the onset of the rash was only 1 day into the fever, and the rash did not progress cephalocaudally. However, the presentation of measles can be altered in vaccinated people. Given the risk of transmission to other people, it would be prudent to send a serum specimen for measles IgM testing. Malaria is a less likely possibility because of the malaria prophylaxis, but malaria testing also would be recommended in this situation. Parvovirus B19 infection is unlikely because its rash typically follows the prodrome by approximately 1 week, as compared with 1 day in this patient. Rubella is unlikely because its fever is rarely above 38.5° C and the rash progresses cephalocaudally.

2. A 5-year-old girl presents to your office with a 5-day history of increasing fever, runny nose, mild photophobia, conjunctivitis, and a brassy cough. A rash that started on her face 2 days ago has now spread to her neck and upper torso. On physical examination, she has white plaques on an erythematous base on her buccal mucosa. Her lungs are clear, and her chest radiograph is consistent with a diffuse patchy bronchopneumonia. The most appropriate therapy to start is:

A. Ceftriaxone
B. Vitamin A, 200,000 IU/day for 2 days
C. Azithromycin
D. Antipyretics and symptomatic cough suppressant alone
E. Ribavirin

Answer: B Although measles can present with a secondary bacterial pneumonia, it is common to have a viral pneumonia due to the measles virus. The absence of a consolidation on chest radiograph or abnormal lung findings on physical examination suggests a viral etiology. There is no indication for ceftriaxone or azithromycin. Vitamin A treatment of children with measles has been associated with decreased morbidity and mortality rates. The World Health Organization currently recommends two doses of vitamin A for all children with acute measles, regardless of their country of residence. Antipyretics, symptomatic cough suppressants, and analgesics are recommended for patients with uncomplicated measles. Although ribavirin has been demonstrated to offer modest benefits in a few small clinical trials, it is not recommended for treatment of measles.

3. A 33-year-old woman is admitted to the hospital on her fourth day of measles rash with a secondary bacterial pneumonia due to methicillin-resistant *Staphylococcus aureus* (MRSA). The most appropriate isolation precautions for this patient consist(s) of:

A. Contact isolation precautions
B. Airborne isolation precautions
C. Droplet isolation precautions
D. Contact and airborne isolation precautions
E. Contact and droplet isolation precautions

Answer: D Patients with measles should be on airborne precautions for at least the first 4 days of rash. Immunocompromised patients may shed measles virus for extended periods and should remain on airborne precautions for the duration of their hospitalization for the acute illness. Contact precautions are necessary for MRSA.

4. A 35-year-old mother comes to your travel clinic in the United States with her healthy 9-month-old infant in anticipation of a leaving in 2 weeks for a trip to France. To reduce the risk of measles associated with international travel you recommend:

A. No preventive measures are needed for measles for travel to Europe.
B. Confirm the mother has received at least one dose of measles-containing vaccine. If not, give her a dose of measles, mumps, rubella vaccine. Give the infant immune globulin prophylaxis, followed by revaccination according to the routine schedule after return from the trip.
C. Confirm the mother has received at least two doses of measles-containing vaccine. If not, give her a dose of measles, mumps, rubella vaccine. Vaccinate the infant with single-antigen measles vaccine, followed by revaccination according to the routine schedule after return from the trip.
D. Confirm the mother has received at least two doses of measles-containing vaccine. If not, give her a dose of measles, mumps, rubella vaccine. Vaccinate the infant with measles, mumps, rubella vaccine, followed by revaccination according to the routine schedule after return from the trip.
E. Give both the mother and the infant immune globulin prophylaxis, followed by revaccination of the infant according to the routine schedule after return from the trip.

Answer: D Measles remains endemic in many developed countries, and many of the measles cases imported into the United States originate in Europe, including U.S. residents returning from travel, as well as foreign visitors. For adults preparing for international travel, documented receipt of two doses of measles containing vaccine is recommended. Single-antigen measles vaccine is not available in the United States. Vaccination is preferred over immune globulin in infants as young as 6 months of age without contraindications. In infants younger than 6 months, immune globulin may be indicated.

5. A 28-year-old mother and her 4-month-old infant were exposed to measles in a daycare setting 2 days before coming into your office. The mother has a documented history of receiving two doses of measles vaccine, the last dose at age 6. How do you provide postexposure prophylaxis for this mother and child?

A. Give the mother another dose of measles, mumps, rubella vaccine. Vaccinate the infant with single-antigen measles vaccine, followed by revaccination according to the routine schedule.
B. No prophylaxis is necessary; the mother is considered immune with two doses of vaccine, and the infant is protected by maternal antibody.
C. Give the infant immune globulin prophylaxis, and consider the mother immune with two doses of vaccine. Vaccinate the infant according to the routine schedule after 12 months of age.
D. Give the mother immune globulin prophylaxis. Vaccinate the infant with measles, mumps, rubella vaccine, followed by revaccination according to the routine schedule.
E. Give both the mother and the infant immune globulin prophylaxis, followed by revaccination of the infant according to the routine schedule.

Answer: C The mother should be considered immune because she has received two doses of vaccine. However, the amount of antibody she transfers to her infant is less than would be transferred by a mother who had measles disease. As a result, the infant may become susceptible to measles in the first 6 months of life, when the risk of severe measles is high. Immune globulin is recommended for postexposure prophylaxis; measles vaccine is not given before 6 months of age.

368

RUBELLA (GERMAN MEASLES)

SUSAN E. REEF

DEFINITION

Rubella, also known as German measles, is an acute viral illness that usually presents with a generalized maculopapular rash of 1 to 3 days' duration, low-grade or no fever, and associated clinical symptoms such as lymphadenopathy, arthropathy, and conjunctivitis. However, about 20 to 50% of persons infected with rubella may present without a rash or other symptoms.

The Pathogen

Rubella virus is a member of the Togaviridae family and the genus *Rubivirus*. Rubella virus is a single-stranded enveloped RNA with a single antigenic type. It measures 50 to 70 nm in diameter and has two envelope proteins (E1, E2) and a core protein (c). The core protein is surrounded by a single-layer lipoprotein envelope with spike-like projections that contain the two glycoproteins, E1 and E2. Humans are the only known reservoir.

EPIDEMIOLOGY

In the prevaccine era, rubella epidemics occurred approximately every 6 to 9 years in the United States. The last major American epidemic, which occurred in 1964-1965, resulted in an estimated 12.5 million cases and approximately 20,000 cases of congenital rubella syndrome. In 1969, live-attenuated rubella vaccines were licensed in the United States and were introduced into the routine childhood immunization program. The initial rubella vaccination strategy was to target children from 12 months of age to puberty with a catch-up mass campaign. During the first 8 years of the vaccination program, the reported number of cases of rubella and congenital rubella syndrome decreased by 78 and 66%, respectively. When a rubella resurgence occurred mainly among adolescents and young adults in 1977-1978, however, the rubella vaccination program was modified to include susceptible postpubertal girls, persons in military service, college students, and persons in certain work settings (e.g., health care). In 1979, the RA 27/3 rubella virus vaccine replaced the other rubella virus vaccines (HPV-77 and Cendehill). Since 2003, 18 or fewer cases have been reported annually in the United States, and rubella is no longer endemic in the United States.[1]

With the success of the vaccination strategy in the Americas, the number of reported rubella cases decreased from 135,947 in 1998 to 11 in 2013, and in 2009, Argentina reported the last endemic rubella case. In Europe, nearly all cases of rubella in 2013 occurred in Poland.[2]

Although rubella and congenital rubella syndrome are no longer endemic in the the Western Hemisphere, they continue to be of global public health importance. For most countries, the epidemiology of rubella mirrors the epidemiology seen in the United States during the prevaccine era, when rubella occurred mainly among children. The number of reported rubella cases globally has decreased from 875,000 in 1999 to about 101,331 in 2013, but the incidence is substantially underestimated because rubella cases in many countries are identified through case-based measles surveillance, which focuses only on the detection of suspected measles cases. Even developed countries like Japan continue to report outbreaks.[3] The annual number of new cases of congenital rubella syndrome globally is estimated to be about 100,000 in 2010.

During the 1970s and 1980s, the use of rubella-containing vaccine (RCV) was mainly limited to industrialized countries. The number of World Health Organization (WHO) member countries using RCV in national childhood immunization schedules increased from 83 (43%) in 1996 to 141 in 2014. As of the end of 2013, rubella control and congenital rubella syndrome prevention or elimination goals have been established in four of six WHO regions (Americas, European, western Pacific, Southeast Asia). In 2013, the other two WHO regions (i.e., eastern Mediterranean, Africa) have not

established regional rubella control or congenital rubella syndrome prevention goals.[4]

PATHOBIOLOGY

Rubella

Rubella virus is transmitted through person-to-person spread by droplets shed from the respiratory secretions of infected persons. The first point of entry is the nasopharynx, where replication occurs and then spreads to the lymph nodes. Subsequent viremia may seed multiple organs, including the placenta. Viremia occurs between 8 and 9 days after exposure and peaks at 10 to 17 days, just before the onset of the rash, which usually occurs 16 to 18 days after exposure. Although individuals with rubella are considered to be only moderately contagious, they may shed virus from 7 days before the onset of the rash to approximately 5 to 7 days or more after its disappearance. Persons with both clinical and subclinical infections are considered contagious.

Congenital Rubella

Rubella virus viremia can infect the placenta of pregnant women, and viral replication can infect all fetal organs. In tissue specimens, infections with rubella virus have diverse effects ranging from small foci of infected cells in apparently normal tissue to hypoplasia, generalized vasculitis, and cell destruction. The hallmark of fetal infection is chronic infection that persists throughout fetal life, with shedding of virus up to 1 year of age.

Infants with congenital rubella syndrome may shed large quantities of virus from body secretions, particularly from the throat. Rubella virus can be found in the nasopharyngeal secretions of more than 80% of these infants during the first month of life. Rubella virus is found in 11% of infected infants between 9 and 12 months of age and in only 3% in the second year of life. Viral shedding by infants with congenital rubella syndrome can result in nosocomial outbreaks, so only individuals immune to rubella virus should be in contact with infants with congenital rubella syndrome or with congenital infection even in the absence of clinical signs of congenital rubella syndrome.

CLINICAL MANIFESTATIONS

Postnatally Acquired Rubella

Acquired rubella, which occurs in 50 to 80% of persons infected with rubella virus, is characterized by a generalized maculopapular rash that usually persists for 1 to 3 days (Fig. 368-1). The rash usually starts on the face and neck and progresses downward. The rash is fainter than the rash of measles (Chapter 367) and does not coalesce. Because of the mildness of the rash, it may be difficult to detect in persons with darker skin. Children usually have few or no prodromal symptoms, so the rash is usually the first sign of illness. However, in older children and adults, a 1- to 5-day prodrome of low-grade fever, malaise, and upper respiratory symptoms often precedes the rash. The incubation period is 14 days, with a range of 12 to 23 days. Lymphadenopathy, particularly occipital and postauricular, may be noted during the second week after exposure.

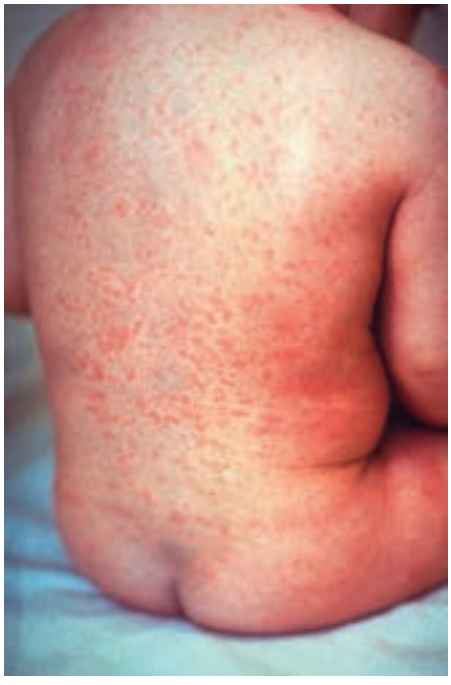

FIGURE 368-1. Rash of rubella on the skin of a child's back. The distribution is similar to that of measles, but the lesions are less intensely red. (From the Centers for Disease Control and Prevention Public Health Image Library, ID #: 712.)

Rubella disease is usually mild and results in few complications. Arthralgia and arthritis are commonly observed in infected adults, particularly in postpubertal females. Other less common complications are thrombocytopenia (1 in 3000 rubella cases) and encephalitis (1 in 6000 rubella cases).

Congenital Rubella Syndrome

The most serious consequences of rubella virus infection occur when a woman becomes infected during pregnancy, particularly during the first trimester. Complications can include miscarriage, fetal death, or a live birth with a constellation of congenital defects known as congenital rubella syndrome. The most common defects of congenital rubella syndrome affect the eyes (e.g., cataracts, pigmentary retinopathy, microphthalmos, congenital glaucoma), the ears (e.g., sensorineural hearing impairment), and the heart (e.g., patent ductus arteriosus, pulmonary arterial stenosis). Other clinical manifestations of congenital rubella syndrome may include microcephaly, developmental delay, and purpura, including dermal erythropoiesis (blueberry muffin syndrome).

Among pregnant women infected with rubella virus during the first 10 weeks of gestation, up to 90% of their live born infants may have congenital rubella syndrome. Among women infected during the first 20 weeks of pregnancy, the rate of congenital rubella syndrome in live born infants is 20%. Infants who are born with rubella virus infection but who do not have any apparent signs or symptoms of congenital rubella syndrome are referred to as infants with congenital rubella infection only.

DIAGNOSIS

Postnatally Acquired Rubella

Because many rash illnesses may mimic rubella virus infection and because 20 to 50% of rubella virus infections may be subclinical, laboratory testing is the only way to confirm the diagnosis. Some illnesses with clinical presentation similar to that of rubella include scarlet fever (Chapter 290), roseola (Chapter 439), fifth disease (Chapter 371), and measles (Chapter 367).

Serologic testing is the most common method of diagnosis of acquired rubella. Immunoglobulin (Ig)M antibody is generally detectable for up to 6 weeks after the onset of the rash. Because IgM antibodies may not be detectable before day 5 of the rash, serologic testing should be repeated no sooner than then. Diagnosis also can be made on the basis of a significant rise in the IgG antibody titer in paired acute and convalescent specimens. The acute serum specimen should be collected within 7 to 10 days after the onset of the rash, and the convalescent serum specimen should be collected 14 to 21 days after the first specimen.

Rubella virus can be isolated from nasopharyngeal (i.e., nasal, throat), blood, urine, and cerebrospinal fluid specimens of persons with rubella and congenital rubella syndrome. The most frequently positive results are from throat swabs. Rubella virus can be isolated during the prodromal period and up to 2 weeks after onset of the rash. For viral cultures, specimens should be obtained during the time of maximum virus secretion—up to 4 days after the onset of rash. During the first 4 days after the onset of the rash, rubella RNA detection by reverse transcription–polymerase chain reaction (RT-PCR) is more sensitive than rubella IgM testing.

Diagnosis in Pregnant Women

In the United States, all pregnant women should be screened for rubella IgG antibodies as part of routine prenatal care. Pregnant women who have a positive serologic test result for IgG antibody to rubella virus are considered to be immune if they do not have a recent history of exposure to rubella virus. Susceptible pregnant women should be vaccinated postpartum. Pregnant women exposed to rubella virus should be evaluated for evidence of acute infection by testing for presence of IgM antibodies in sera or a significant rise of IgG antibodies in acute and convalescent sera. Pregnant women with evidence of acute infection should be monitored clinically and evaluated for gestational age at infection to assess the risk of fetal infection.

Congenital Rubella

Diagnosis of congenital rubella syndrome in infants can be confirmed by either serologic or virologic methods. Serum IgM antibodies may be present

in an infant with congenital rubella syndrome for up to a year after birth; however, IgM antibody may not be detectable during the first month of life. Thus, infants younger than 1 month and with symptoms consistent with congenital rubella syndrome should be tested at 1 month of age. Congenital rubella syndrome also can be confirmed by documentation of a persistent rubella serum IgG titer beyond the time expected from passive transfer of maternal IgG antibody (i.e., a rubella titer that does not decline at the expected rate of a two-fold dilution per month).

Rubella virus from congenitally infected infants can be isolated most commonly from throat swabs and less commonly from urine and cerebrospinal fluid specimens. Infants with congenital rubella may excrete virus for up to 1 year, but specimens for virus isolation are most likely to be positive if they are obtained within the first 6 months after birth. Rubella virus in infants with congenital rubella syndrome also can be detected by RT-PCR with use of the same specimens as for viral isolation.

TREATMENT

Rx

There is no specific treatment for rubella or congenital rubella syndrome. For persons with rubella, symptomatic treatment may be warranted for clinical manifestations such as arthralgias, myalgias, and fever. Infants with congenital rubella syndrome should be evaluated and treated by specialists for their specific clinical manifestations.

PREVENTION

Passive Immunization

Administration of immune globulin after exposure to rubella virus will not prevent infection or viremia but might modify or suppress symptoms. Therefore, immune globulin is not recommended for routine postexposure prophylaxis of rubella in any circumstance.

Active Immunization

One dose of live-attenuated rubella vaccine (RA 27/3, RCV) induces rubella IgG antibody seroconversion in 95% or more of persons. Immunity is considered long term, probably lifelong. RCV is available as a monovalent formulation or in combination with measles-mumps (MMR) and measles-mumps-varicella (MMRV) vaccines (Chapter 18).

In the United States, the routine rubella vaccination policy is to immunize children with the first dose of MMR at 12 to 15 months of age and to provide a second dose at 4 to 6 years of age. The MMRV vaccine also can be given to children up to age 12 years. Persons who were born in 1957 or later (except women of childbearing age who are or could become pregnant) and who do not have a medical contraindication (e.g., pregnancy) should receive at least one dose of MMR unless they have documentation of one dose of live rubella virus vaccine or laboratory evidence of immunity or laboratory confirmation of disease. High-risk groups who should be targeted for vaccination include health care personnel, persons attending post–high school educational facilities, the military, international travelers, and nonpregnant women of childbearing age.

In follow-up studies of more than 2700 susceptible women who were unknowingly pregnant and who received a live-attenuated rubella vaccine, none of their infants was born with congenital rubella syndrome. Nevertheless, because of the theoretical risk of congenital rubella syndrome in infants born to pregnant women vaccinated with RCV, the vaccine should not be given to pregnant women, and pregnancy should be avoided for at least 28 days after receipt of vaccine. Receipt of RCV during pregnancy is not generally considered to be an indication for termination of pregnancy.

PROGNOSIS

Because rubella is usually a mild disease, the prognosis is excellent, with complete recovery in almost all persons. Deaths are seen in 0 to 50% of patients who develop rubella encephalitis. The prognosis for infants with congenital rubella syndrome is dependent on their clinical manifestations and access to quality medical care.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Centers for Disease Control and Prevention (CDC). Rubella and congenital rubella syndrome control and elimination–global progress, 2000-2012. *MMWR Morb Mortal Wkly Rep*. 2013;62:983-986.
2. Muscat M, Shefer A, Ben Mamou M, et al. The state of measles and rubella in the WHO European Region, 2013. *Clin Microbiol Infect*. 2014;20(suppl 5):12-18.
3. Nationwide rubella epidemic–Japan, 2013. *MMWR Morb Mortal Wkly Rep*. 2013;62:457-462.
4. Centers for Disease Control and Prevention. *Manual for the Surveillance of Vaccine-Preventable Diseases*. http://www.cdc.gov/vaccines/pubs/surv-manual/index.html; Accessed March 11, 2015.

369

MUMPS

JOHN W. GNANN, JR.

DEFINITION

Mumps is an acute systemic viral infection that occurs most commonly in children, is usually self-limited, and is clinically characterized by nonsuppurative parotitis.

The Pathogen

Mumps virus is a member of the family Paramyxoviridae. Mumps virions are pleomorphic, roughly spherical enveloped particles with an average diameter of 200 nm. Glycoprotein spikes project from the surface of the envelope. A helical nucleocapsid composed of nucleoproteins and of linear, nonsegmented, single-stranded, negative-sense RNA approximately 15.3 kilobases in size encodes seven major proteins as well as several minor proteins. Humans are the only natural hosts for mumps virus, although infection can be induced experimentally in a variety of mammalian species. In vitro, mumps virus can be cultured in many mammalian cell lines and in embryonated hens' eggs.

EPIDEMIOLOGY

In unvaccinated urban populations, mumps is a disease of school-aged children (5 to 9 years); more than 90% of individuals have mumps antibodies by 15 years of age. Before the mumps vaccine was released in the United States in 1967, mumps was an endemic disease with a seasonal peak of activity between January and May. The largest number of cases reported in the United States occurred in 1941, when the incidence of mumps was 250 cases per 100,000 population. In 1968, when the mumps vaccine was first entering clinical use, the incidence of mumps was 76 cases per 100,000 population. In 1985, only 2982 cases of mumps were reported, an incidence of 1.1 per 100,000 population, which represents a 98% decline from the number of cases reported in 1967. Sporadic outbreaks of mumps in secondary schools between 1985 and 2005 were attributed to primary vaccine failure. After 1989, when the Centers for Disease Control and Prevention (CDC) Advisory Committee on Immunization Practices (ACIP) issued a recommendation that all children receive a second dose of the measles-mumps-rubella (MMR) vaccine at the time of school entry, the problem of primary vaccine failure was reduced. In 2003, the CDC reported only 231 cases of mumps in the United States, the lowest annual total ever recorded. In 2006, however, 6584 cases of mumps were reported, with a national incidence of 2.2 cases per 100,000 population. Since that time, several outbreaks have been reported. In 2010, another mumps outbreak resulted in 3500 cases in New York and New Jersey, with the highest rate among boys 13 to 17 years of age who attended tradition-observant Jewish schools[1]; 89% of cases had received at least two doses of a mumps-containing vaccine, suggesting their illness was because of intense exposure, particularly among boys in school. In another outbreak in New York City, 90% of cases had received one dose of vaccine, and 77% had received two doses. The effectiveness of a two-dose regimen for preventing mumps was 86%.[2] In an outbreak in Guam during 2009 and 2010, 505 mumps cases occurred among school-aged children, even though 93% had received two doses of vaccine; crowding at home and high student contact rates were identified as risk factors for transmission.[3] Although some cases in recent outbreaks can be attributed to failure to vaccinate, most represented primary (insufficient initial immune response) or secondary (waning immune response) vaccine failure, thereby suggesting the need either for a mumps vaccine with a longer duration of protection or for a revised vaccination policy focusing on young adults.

PATHOBIOLOGY

Mumps is highly contagious and can be transmitted experimentally by inoculation of virus onto the nasal or buccal mucosa, suggesting that most natural infections result from droplet spread of upper respiratory secretions. The mean incubation period for mumps is 18 days. Primary viral replication takes place in epithelial cells of the upper respiratory tract, followed by spread of virus to regional lymph nodes and subsequent viremia.[4] Because virus can be

isolated from saliva for 5 to 7 days before and up to 9 days after the onset of clinical symptoms, an infected individual is potentially able to transmit mumps for up to 2 weeks. An estimated 30% of mumps infections in children are subclinical or associated only with nonspecific upper respiratory infection symptoms. Transient IgM antibody responses are detected early in the course of mumps infection, followed by the appearance of IgG antibody and cytotoxic T lymphocytes. Mumps-specific IgG can be detected during the first week of acute infection, peaks at 3 to 4 weeks, and persists for decades. Lifelong immunity follows natural infection. Patients who report more than one episode of mumps probably had parotitis of another cause.

CLINICAL MANIFESTATIONS

Parotitis

Mumps usually begins with a short prodromal phase of low-grade fever, malaise, headache, and anorexia.[5] Young children may complain of ear pain initially. The characteristic parotid tenderness and enlargement, in which the earlobe is lifted forward and obscures the angle of the mandible, then develops (Fig. 369-1). The parotid glands are involved most commonly, although other salivary glands may occasionally be enlarged. Parotitis may initially be unilateral, with swelling of the contralateral parotid gland occurring 2 to 3 days later; bilateral parotitis with symptomatic salivary gland involvement eventually develops in 70% of patients. Painful parotid gland enlargement progresses for a period of about 3 days, followed by defervescence and resolution of the parotid pain and swelling within about 7 days. Long-term sequelae of mumps parotitis are uncommon.

Aseptic Meningitis

Symptomatic meningitis occurs in 15% of cases and is the second most common manifestation of mumps. About 50% of patients with mumps parotitis have cerebrospinal fluid (CSF) pleocytosis, but many have no clinical evidence of meningitis. Signs and symptoms of meningeal inflammation (headache, neck stiffness, vomiting, and lethargy) plus high fever usually develop 4 to 5 days after the onset of parotitis, although the meningitis may occasionally precede the parotitis. Indeed, 40 to 50% of all cases of documented mumps meningitis occur in patients who never exhibit clinical parotitis. For unexplained reasons, symptomatic central nervous system (CNS) involvement with mumps is two to three times more common in boys than in girls. Examination of CSF usually reveals normal opening pressure and a mononuclear cell pleocytosis with an average cell count of 450/μL. Polymorphonuclear leukocyte predominance in CSF may be seen in some patients early during the course of mumps meningitis. CSF protein is generally normal or mildly elevated (<100 mg/dL). Hypoglycorrhachia, which is not usually seen in viral meningitis, may be present in 10 to 30% of patients with mumps meningitis. Mumps virus can be recovered from CSF. Whereas the symptoms of mumps meningitis typically resolve within 7 to 10 days, the CSF abnormalities may persist for up to 5 weeks. Mumps meningitis is generally benign, and significant neurologic complications are rare.

Encephalitis

The spectrum of mumps-induced CNS disease ranges from mild "aseptic" meningitis (which is common) to severe encephalitis (which is rare). Some cases of encephalitis develop concurrently with the parotitis and are thought to result from direct extension of viral infection from the choroid plexus ependyma into parenchymal neurons. Other cases of mumps encephalitis occur 1 to 2 weeks after the onset of parotitis and may represent a demyelinating postinfectious encephalitis. Clinical findings in mumps encephalitis include obtundation (and less commonly delirium), generalized seizures, and high fever. Other neurologic findings can include focal seizures, aphasia, paresis, and involuntary movements. Recovery from mumps encephalitis is generally complete, although complications such as aqueductal stenosis with hydrocephalus, seizure disorders, and psychomotor retardation have been reported. The overall mortality rate from mumps encephalitis is 0.5 to 2.3%.

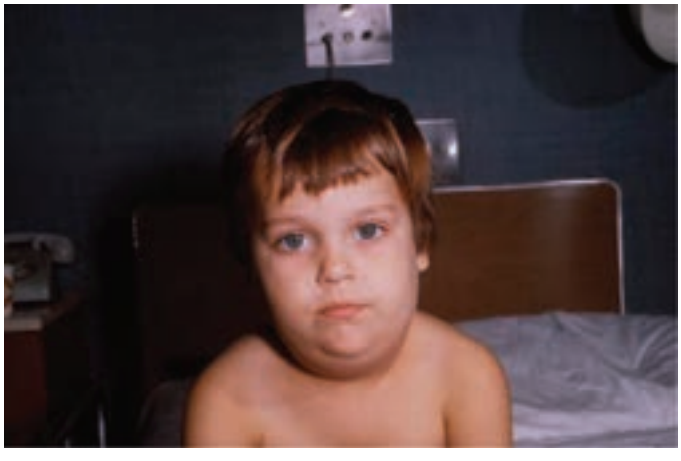

FIGURE 369-1. Mumps. Child with submandibular swelling due to mumps parotitis. (From the Centers for Disease Control and Prevention Public Health Image Library, ID #: 4491.)

Orchitis

Epididymo-orchitis is rare in young boys with mumps, but it occurs in 15 to 35% of postpubertal males with mumps.[6] Orchitis is most often unilateral (bilateral involvement occurs in 17 to 38% of cases) and results from replication of mumps virus in the seminiferous tubules, with resulting lymphocytic infiltration and edema. Orchitis typically develops within 1 week of the onset of parotitis, although orchitis, like mumps meningitis, can develop before or even in the absence of parotitis. Mumps orchitis is characterized by marked testicular swelling and severe pain accompanied by fever, nausea, and headache. The pain and swelling resolve within 5 to 7 days, but residual testicular tenderness can persist for weeks. Testicular atrophy may follow orchitis in 35 to 50% of cases; however, sterility is an uncommon complication, even in men with bilateral orchitis.

Other Manifestations

Mumps can cause inflammation of other glandular tissues, including pancreatitis and thyroiditis. Oophoritis and mastitis have been reported in postpubertal women with mumps. Transient renal function abnormalities are common in mumps, and virus can be isolated readily from urine; significant renal damage is rare, however. Other infrequent manifestations of mumps include sensorineural deafness (either transient or permanent), arthritis, myocarditis, and thrombocytopenia. Maternal mumps infection during the first trimester of pregnancy results in an increased frequency of spontaneous abortions, but no clear association between congenital malformations and maternal mumps has been demonstrated.

DIAGNOSIS

The diagnosis of mumps is usually based on clinical findings in a child with fever and parotitis, particularly if the individual is known to be susceptible and has been exposed to mumps during the preceding 2 to 3 weeks. An atypical clinical presentation (e.g., meningitis or orchitis without parotitis) requires laboratory confirmation. Culturing for mumps virus has largely been replaced by reverse transcription–polymerase chain reaction (RT-PCR) assays. Detection of mumps virus RNA is diagnostic of infection.[7] RT-PCR appears to be more sensitive than culture for detection of mumps virus from oropharyngeal swabs or CSF samples. Serologic demonstration of mumps immunoglobulin M (IgM) antibody provides good evidence of recent infection; IgM is detectable during the first week of illness and persists for at least 6 weeks. Alternatively, testing of acute and convalescent sera by enzyme-linked immunosorbent assay (ELISA) should demonstrate a diagnostic four-fold rise in mumps IgG antibody titer. About 30% of patients have an elevated serum amylase level that may be due to parotitis or pancreatitis.

The differential diagnosis of a mumpslike syndrome includes infections caused by other viruses, such as Epstein-Barr virus (Chapter 377), human herpesvirus type 6, adenovirus (Chapter 365), influenza A virus (Chapter 364), parainfluenza virus (Chapter 363), coxsackievirus (Chapter 379), or lymphocytic choriomeningitis virus. Among 101 sporadic cases of parotitis assessed in the United States in 2009 to 2011, for example, the most commonly detected viruses were Epstein-Barr virus (23%) and human herpesvirus type 6 (10%); mumps virus was not detected in any specimens, although 17% of cases demonstrated positive mumps IgM.[8] Bacteria such as *Staphylococcus aureus* can cause suppurative parotitis. Parotid gland enlargement can also occur in patients with AIDS, particularly children. Parotid gland enlargement can also be associated with Sjögren's syndrome (Chapter 268), sarcoidosis (Chapter 95), amyloidosis (Chapter 188), thiazide ingestion, iodine sensitivity, tumor, or salivary duct obstruction. Careful examination should distinguish parotitis from lymphadenopathy.

TREATMENT

Rx

Management of a patient with mumps consists of conservative measures to provide symptomatic relief, adequate hydration, and nutrition. Treatment of orchitis includes bed rest, scrotal support, analgesics, and ice packs. Patients with significant CNS involvement require hospitalization for observation and supportive care. There is currently no established role for antiviral drugs, interferon, corticosteroids, or passive immunotherapy in the treatment of mumps.

PREVENTION

The cornerstone of mumps prevention is active immunization with the live attenuated mumps vaccine (Chapter 18). In the United States, mumps vaccine is administered in combination with the measles and rubella vaccines (MMR) to children at 12 to 15 months of age, with a second dose at 4 to 6 years of age.[9] Efficacy of the mumps vaccine is estimated to be 70 to 80% after two doses.[10] The mumps vaccine is also indicated for susceptible adults. Administration of the live mumps vaccine is contraindicated in pregnant women. Vaccination is also not recommended in persons who have received immunoglobulin therapy, which might interfere with the immune response to the vaccine, within the preceding 3 months or in persons with severe systemic immunosuppression caused by disease or medical therapy.

The Jeryl-Lynn strain of attenuated mumps virus used in the United States since 1967 is a very well tolerated vaccine, although rare instances of fever, parotitis, and possibly aseptic meningitis have been reported after immunization. Beginning in 1988, an increased frequency of cases of vaccine-related mumps meningitis was recognized in other countries. These cases occurred after the administration of MMR vaccines containing other mumps strains, such as Urabe AM9 or Leningrad 3. This problem has not been seen in the United States, where the Jeryl-Lynn mumps vaccine continues to be used.

Questions about prevention often arise when an individual with no history of mumps (typically an adult male) is exposed to a patient with active mumps. The immune status of the exposed individual can be determined by serologic testing (mumps IgG by ELISA), although there may be some delay. The mumps skin test is not a reliable indicator of immune status. Most adults (>90%) born in the United States before 1957 were naturally infected and are therefore immune. Mumps vaccine can be safely administered to an individual of unknown immune status, although the efficacy of postexposure vaccination for preventing mumps has not been determined.

For infection control purposes, patients with mumps require both standard precautions and droplet precautions for at least 5 days after the onset of parotitis.[11] In the outpatient setting, a patient with suspected mumps should wear a mask and be isolated from other potentially susceptible persons. When a patient with mumps is hospitalized, a private room is required; caregivers should wear masks, and the patient should wear a mask while being transported.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Barskey AE, Schulte C, Rosen JB, et al. Mumps outbreak in Orthodox Jewish communities in the United States. *N Engl J Med*. 2012;367:1704-1713.
2. Livingston KA, Rosen JB, Zucker JR, et al. Mumps vaccine effectiveness and risk factors for disease in households during an outbreak in New York City. *Vaccine*. 2014;32:369-374.
3. Nelson GE, Aguon A, Valencia E, et al. Epidemiology of a mumps outbreak in a highly vaccinated island population and use of a third dose of measles-mumps-rubella vaccine for outbreak control: Guam 2009 to 2010. *Pediatr Infect Dis J*. 2013;32:374-380.
4. Rubin S, Eckhaus M, Rennick LJ, et al. Molecular biology, pathogenesis and pathology of mumps virus. *J Pathol*. 2015;235:242-252.
5. Hviid A, Rubin S, Mühlemann K, et al. Mumps. *Lancet*. 2008;371:932-944.
6. Ternavasio-de la Vega HG, Boronat M, Ojeda A, et al. Mumps orchitis in the post-vaccine era (1967-2009): a single-center series of 67 patients and review of clinical outcome and trends. *Medicine*. 2010;89:96-116.
7. Maillet M, Bouvat E, Robert N, et al. Mumps outbreak and laboratory diagnosis. *J Clin Virol*. 2015;62:14-19.
8. Barskey AE, Juieng P, Whitaker BL, et al. Viruses detected among sporadic cases of parotitis, United States, 2009-2011. *J Infect Dis*. 2013;208:1979-1986.
9. McLean HQ, Fiebelkorn AP, Temte JL, et al. Prevention of measles, rubella, congenital rubella syndrome, and mumps. 2013: summary recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR*. 2013;62(RR04):1-34.
10. Livingston KA, Rosen JB, Zucker JR, et al. Mumps vaccine effectiveness and risk factors for disease in households during an outbreak in New York City. *Vaccine*. 2014;32:369-374.
11. Kutty PK, Kyaw MH, Dayan GH, et al. Guidance for isolation precautions for mumps in the United States: a review of the scientific basis for policy change. *Clin Infect Dis*. 2010;50:1619-1628.

REVIEW QUESTIONS

1. A 19-year-old college student presents with a 1-day history of fever, mild headache, and bilateral parotitis. She relates that she received all of the appropriate childhood vaccinations. After examining her, you suspect mumps. What is your best option for diagnostic testing?

A. Collect urine for mumps virus culture.
B. Collect throat swab for mumps virus culture.
C. Collect throat swab for mumps virus polymerase chain reaction (PCR).
D. Collect cerebrospinal fluid (CSF) for mumps virus PCR.
E. Collect acute and convalescent sera for mumps immunoglobulin G (IgG) by enzyme-linked immunosorbent assay (ELISA).

Answer: C All of the suggested testing modalities could provide confirmation of the diagnosis of mumps. However, mumps viral culture may not be routinely available and is expensive. Real-time PCR is a rapid and very sensitive assay for detecting mumps RNA and has replaced viral culture in most situations. CSF is a suitable specimen for PCR, but there is no evidence for serious CNS involvement in this case; lumbar puncture is not indicated. Acute and convalescent sera (collected after an interval of 4 to 6 weeks) will provide a retrospective diagnosis. Serologic testing for mumps IgM would also be an acceptable diagnostic approach but may be negative very early in the clinical course.

2. A 62-year-old man who has been in good health calls your office for advice. His 15-year-old grandson visited his home 3 days ago and has now been diagnosed with mumps. Your patient grew up in the United States and relates that his mother told him that he never had mumps. He does not recall receiving a mumps vaccine. He has heard stories about mumps orchitis and is very concerned. What do you recommend?

A. Administer measles-mumps-rubella vaccine (MMR) as soon as possible.
B. Administer mumps hyperimmune globulin as soon as possible.
C. Administer the mumps skin test.
D. Send blood for mumps IgG serology.
E. Send blood for mumps IgM serology.

Answer: D More than 90% of adults who grew up in the United States between 1957 and 1967 (when the mumps vaccine was introduced) are seropositive for mumps, so the probability that this man is susceptible to mumps is low. Because approximately 30% of mumps infections are asymptomatic, many adults are seropositive with no history of disease. His serologic status can be confirmed by testing for mumps IgG. Mumps IgM testing is appropriate for diagnosing acute mumps but not for determining mumps susceptibility. Mumps vaccine (or MMR) could be safely administered, but there are no data regarding the efficacy of postexposure immunization. Mumps hyperimmune globulin was used in the past as therapy for mumps orchitis, but it is no longer recommended or available. The mumps skin test is an unreliable indicator of mumps immune status, and the antigen is no longer widely available.

3. There is an ongoing outbreak of mumps in your community, primarily involving high school and college students. Several of your adult patients have requested that they be tested for mumps susceptibility because they do not recall having mumps and can provide no documentation of vaccination. Based on mumps IgG serologic results, you have identified a few patients who are seronegative and therefore susceptible to mumps. Which of the following patients should receive mumps vaccination?

A. A 31-year-old man with type 1 diabetes mellitus, chronic kidney disease, and hypertension
B. A 23-year-old man who was recently diagnosed with AIDS ($CD4^+$ lymphocyte count = 41 cells per μL) and is receiving antiretroviral therapy
C. A 36-year-old woman who recently emigrated from Central America, is pregnant (gestational age 20 weeks), and has a teenaged child at home
D. A 55-year-old man who is 2 years status post–cardiac transplantation and is currently doing well on an antirejection regimen of mycophenolate, tacrolimus, and prednisone
E. A 40-year-old woman who has had a flare of her sarcoidosis with pulmonary and CNS involvement and is on a prolonged corticosteroid taper and is currently taking prednisone 30 mg daily

Answer: A Mumps is a live-virus vaccine containing replication-competent virus. Its use is contraindicated in (1) pregnant women; (2) persons with primary or acquired immunodeficiency syndromes; (3) persons with hematologic or lymphoproliferative malignancies; (4) persons receiving systemic immunosuppressive therapy (including prednisone ≥20 mg/day or equivalent for ≥2 weeks); or (5) persons with allergies to any mumps vaccine components. Mumps immunization (using MMR) is recommended for children or adults who are infected with HIV and who do not have evidence of severe immunosuppression (defined as CD4 count <200 lymphocytes/μL or CD4 <15%). Diabetes mellitus is not a contraindication to vaccination.

370

CYTOMEGALOVIRUS, EPSTEIN-BARR VIRUS, AND SLOW VIRUS INFECTIONS OF THE CENTRAL NERVOUS SYSTEM

JOSEPH R. BERGER AND AVINDRA NATH

CYTOMEGALOVIRUS INFECTION

Human cytomegalovirus (CMV) is a ubiquitous herpesvirus that is acquired throughout life (Chapter 376). In children, CMV is an important and relatively common cause of congenital neurologic deficits. In the United States, seroprevalence rates are 40% in adolescents and 60 to 90% in adults. Primary infection is usually asymptomatic in young, healthy adults but may be associated with a transient mononucleosis-like syndrome. CMV results in major neurologic disability in the setting of immunosuppression, particularly in transplant recipients, in persons with acquired immunodeficiency syndrome (AIDS), and in those with hematologic malignant neoplasms. In addition to the retina, CMV may involve the meninges, brain, spinal cord, peripheral nerves, and muscle.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Cytomegalovirus Encephalitis

In patients with AIDS, CMV encephalitis generally occurs only in the presence of profound immunosuppression (CD4 T lymphocyte counts <100 cells/μL). The most typical presentation in patients with AIDS is a subacute, diffuse encephalopathy evolving over a period of weeks and characterized by headache, impaired cognition and sensorium, apathy, and social withdrawal. Neurologic examination reveals abnormal mentation and variable motor features, including hyperreflexia, ataxia, and weakness. CMV ventriculitis is characteristically present, and progressive ventricular enlargement may be observed. Other features may suggest brain stem encephalitis, including internuclear ophthalmoplegia, nystagmus, cranial nerve palsies, gaze paresis, ataxia, and quadriparesis. Other findings in these patients include cerebral infarction resulting from CMV vasculitis, acute subarachnoid hemorrhage, and intracerebral hemorrhage. Virtually all patients with CMV encephalitis have systemic CMV infection. CMV myelitis, polyradiculitis, and multifocal neuritis may also occur with CMV encephalitis. Distinctive retinal lesions can often be seen ophthalmoscopically (Chapter 376) and may serve as a useful diagnostic clue. CMV can also be vertically transmitted and cause a congenital infection in the fetus with a classic triad of chorioretinitis, microcephaly, and cerebral calcifications.

Cerebral imaging studies are of limited sensitivity and low specificity in patients with CMV encephalitis. Ependymal or meningeal enhancement as well as areas of focal infarction or necrosis may be visualized. Restricted diffusion or high signal intensity lesions lining the ependyma of the lateral ventricles may be best seen on diffusion-weighted images.[1] Progressive ventricular enlargement should suggest CMV ventriculitis. Rarely, CMV infection may be manifested as a cerebral mass lesion.

Cerebrospinal fluid (CSF) findings are variable. Most patients have elevated protein levels and a CSF pleocytosis, but leukocytes may be absent and glucose levels may be normal or decreased. In contrast to other viral infections that usually cause a lymphocytic predominance in the CSF, a marked pleocytosis with a polymorphonuclear leukocyte preponderance may occur in patients with CMV ventriculoencephalitis. CMV can rarely be cultured from CSF. CSF real-time polymerase chain reaction (PCR) is the most specific diagnostic method, but the diagnosis is often difficult and relies on a high index of clinical suspicion. CMV serology is not helpful for diagnosing an active infection. Histopathologic examination reveals multinucleate cells with intranuclear cytomegalic inclusions (Fig. 370-1).

Myelitis

CMV transverse myelitis, which may be seen in both immunocompetent and immunosuppressed patients, is indistinguishable neurologically from other forms of transverse myelitis (Chapters 400 and 411). A severe necrotizing CMV myelitis, which is usually longitudinally extensive and involves several spinal segments, may occur in the immunosuppressed host, particularly with human immunodeficiency virus (HIV) infection, and is commonly associated with polyradiculitis. Some cases of necrotizing myelitis in the absence of a typical polyradiculitis syndrome have been described, with patients displaying acute or progressive paraplegia and disturbances in urinary and rectal sphincter function. Reflexes are preserved or enhanced in the legs unless concurrent neuropathy is present. A sensory level may be demonstrable. The combination of radicular complaints and a CSF polymorphonuclear pleocytosis may serve as a clue to the diagnosis. Serology for CMV may be useful in immunocompetent individuals.

Polyradiculomyelitis

As many as 25% of patients dying of AIDS have neuromuscular disease from CMV, predominantly localized to the perineurial and epineurial regions. CMV polyradiculomyelitis in HIV-infected patients is manifested subacutely during a period of days to weeks. Initial symptoms of paresthesias or dysesthetic pain localized to the perineum and lower extremities are followed by rapidly progressive paraparesis with hypotonia and diminished or absent lower extremity reflexes. Urinary retention is characteristic, and rectal

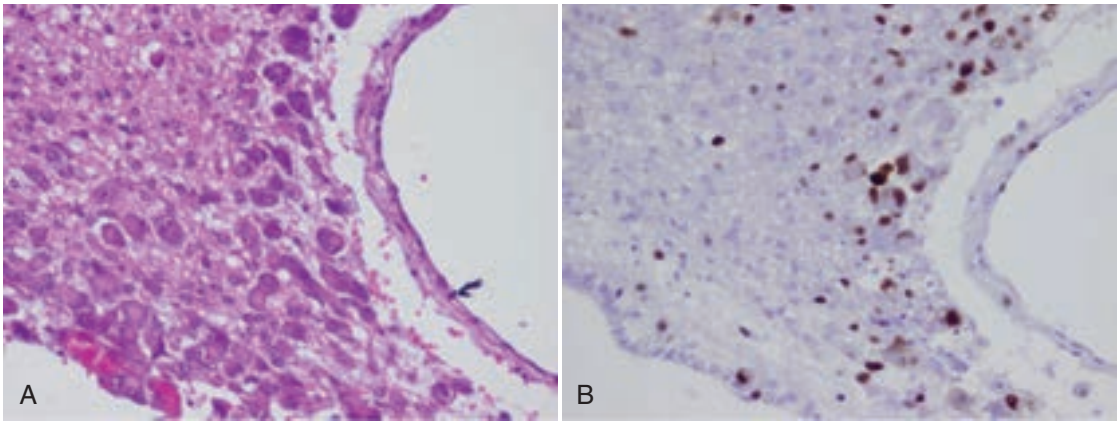

FIGURE 370-1. Pathology of CMV encephalitis. A, Large cells are seen in the perivascular region, some are multinucleated. **B**, The cytomegalic inclusions are present in the nucleus and stain with an antibody to CMV *(brown)*. Magnification is 40×. (Courtesy Martha Quezado, National Institutes of Health, Bethesda, MD.)

sphincter incontinence is common. Variable sensory findings are typically overshadowed by weakness. Babinski signs and diminished sensation below a discrete level across the trunk are evidence of an associated myelitis. With time, symptoms progress by ascending to involve the upper limbs and sometimes the cranial nerves. CSF examination generally shows polymorphonuclear pleocytosis, prominent elevation of protein levels, and a low glucose level. Spinal magnetic resonance imaging (MRI) may be normal or show enhancement of the conus medullaris, cauda equina, meninges, and nerve roots. Electrophysiologic studies reveal axonal neuropathy with evidence of acute denervation. Variable slowing of nerve conduction may occur.

The appearance of acute cauda equina syndrome in a patient with AIDS or in a solid organ or bone marrow transplant recipient is suggestive of CMV infection when a polymorphonuclear pleocytosis is present in CSF; however, the syndrome is not pathognomonic. Other conditions that may produce a cauda equina syndrome in AIDS patients include lymphomatous meningitis, syphilis, toxoplasmosis, other herpesvirus infections, and cryptococcal or bacterial meningitis. Progressive multifocal motor and sensory neuropathy that evolves during a period of weeks to months has also been seen in patients with CMV infection. Paresthesia and dysesthesia are quickly followed by prominent motor weakness involving both the upper and lower limbs asymmetrically. Neurogenic atrophy may be prominent. Nerve biopsy reveals necrotizing neuritis with mononuclear and polymorphonuclear infiltrates and cytomegalocytes localized around endoneurial capillaries in the nerve trunks and roots. Some patients may have necrotizing arteritis. PCR for CMV in CSF and comparison of levels of antibody to CMV in the serum and CSF may be useful in establishing the diagnosis.

TREATMENT

Rx

CMV neurologic complications should be treated with ganciclovir (5 mg/kg intravenously every 12 hours) plus foscarnet (60 mg/kg intravenously every 8 hours or 90 mg/kg intravenously every 12 hours until symptomatic improvement), followed by maintenance therapy with oral valganciclovir (900 mg daily) and intravenous foscarnet (90 to 120 mg/kg intravenously during 2 hours, every 24 hours); however, evidence of efficacy in these conditions is limited chiefly to case reports and small series. Cidofovir is a second-line agent (5 mg/kg by intravenous infusion during 1 hour once a week for 2 consecutive weeks, with saline hydration and probenecid, 2 g orally 3 hours before the dose and 1 g orally at 2 hours and 8 hours after the dose). CMV strains resistant to these agents have emerged, and CMV encephalitis has developed in the presence of maintenance ganciclovir therapy for CMV retinitis. Combination therapy (foscarnet and ganciclovir) or different drugs should be considered in patients already undergoing suppressive monotherapy or in those with persistent CSF pleocytosis. Maintenance therapy, which is the same regimen given every other week, is required unless the patient experiences immune reconstitution, such as after highly active antiretroviral therapy (HAART) in an AIDS patient or discontinuation of immunosuppressive regimens in a transplant or cancer patient. CMV retinitis can be treated with intravitreal ganciclovir or fomivirsen.

PROGNOSIS

The prognosis for long-term survival, especially with AIDS, is very poor, and most patients have only limited neurologic recovery.

EPSTEIN-BARR VIRUS INFECTION

Epstein-Barr virus (EBV), the major cause of infectious mononucleosis (Chapter 377), is distributed worldwide. Individuals in areas of high population density and lower social strata acquire the virus in early childhood. However, seroepidemiologic studies indicate that virtually all persons are infected by EBV by 30 years of age.

Neurologic manifestations occur in 1 to 5% of patients with primary EBV infection and may be the predominant clinical finding. The most common neurologic disorder associated with infectious mononucleosis is meningoencephalitis, which is rare in early childhood and is most often observed in persons between the ages of 15 and 25 years. Its onset may be gradual during a span of several days or be explosive. Fever, headache, mild stiff neck, confusion, lethargy, seizures, and hyperreflexia are the most typical features. Some patients may present predominantly with ataxia, cerebellitis, or other focal neurologic features, including hemiparesis, focal seizures, and brain stem findings. Hyperintense signal abnormalities on T2-weighted and fluid attenuated inversion recovery (FLAIR) cranial MRI are frequently observed but are nondiagnostic. PCR for EBV in CSF and comparison of levels of antibody to EBV in the serum and CSF may be diagnostically useful.

Ganciclovir treatment (10 mg/kg/day intravenously for 3 weeks followed by 1000 mg/day orally for another 3 weeks or until the virus is cleared) has been used in some cases but is of unproven value. The prognosis for patients with EBV meningoencephalitis is excellent, with complete resolution anticipated in 1 to 2 weeks.

SLOW VIRUS INFECTIONS OF THE CENTRAL NERVOUS SYSTEM

Human T-Lymphotropic Virus Type 1 and Human Immunodeficiency Virus

These viruses and their neurologic sequelae are considered in Chapters 378 and 394, respectively.

Subacute Sclerosing Panencephalitis

EPIDEMIOLOGY AND PATHOBIOLOGY

Subacute sclerosing panencephalitis (SSPE), which is caused by the measles virus (Chapter 367), usually affects children but can present in young adulthood. Patients generally have a history of measles within the first 2 years of life, and it is speculated that such early host exposure allows the emergence of persistent but defective virus replication because the SSPE virus genome, particularly the matrix gene, differs from wild-type measles. SSPE occurs after a latency period of months to years following acute measles infection. As a result of effective vaccination strategies against measles virus, the incidence of SSPE has decreased markedly to about 4 to 5 cases per year in the United States, but it remains 21 cases per million population in India.[2]

Gray matter is most prominently involved. The pathologic features of SSPE include gliosis, loss of myelin, and perivascular infiltrates of lymphocytes and plasma cells in white and gray matter. Neuronal cell loss is seen in later stages of the illness. Intranuclear Cowdry type A inclusions containing viral nucleocapsids are identified in neurons and glia.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

SSPE usually begins with cognitive and behavioral changes, and cortical blindness is often an early feature. Progression is associated with motor dysfunction, including prominent myoclonus, cognitive decline, choreoathetosis, dystonia, and rigidity. Its course progresses during a period of 1 to 3 years to rigid quadriparesis and a vegetative state, frequently accompanied by autonomic features, such as hyperthermia, excessive sweating, and altered pulse and blood pressure. The condition is more common in rural settings and affects boys more often than girls. Retinal changes such as macular retinitis and pigmentary changes can precede the neurologic manifestations by several months.

The electroencephalogram typically reveals unilateral or bilateral periodic complexes with synchronous bursts of two or three high-amplitude slow waves per second, with recurrence at regular intervals of 5 to 8 seconds and a 1 : 1 relationship with myoclonic jerks.[3] MRI of the brain shows high signal intensity lesions that are diffuse in the subcortical and periventricular white matter with cortical atrophy, but also rarely may involve the basal ganglia and brain stem. In early stages of the illness, the MRI may be normal. Computed tomography (CT) of the brain shows generalized atrophy. CSF protein, glucose, and cell levels are usually normal; CSF is characterized by a high immunoglobulin concentration, oligoclonal bands, and intrathecal synthesis of antibody to measles virus antigens. Serum measles antibody titers are also high. These findings are usually sufficiently characteristic for diagnosis, but measles RNA can be detected in the brain by PCR. Rarely, brain biopsy is needed for definitive diagnosis in atypical cases. Measles virus may also cause subacute encephalitis in an immunocompromised host. The prominence of cognitive and motor dysfunction in these patients resembles that of SSPE, but in the clinical setting, its subacute onset and more rapid evolution and the presence of generalized seizures rather than myoclonus are distinctive. Brain abnormalities include abundant intranuclear inclusions, but inflammation is minimal, and neither serum nor CSF antibody titers against measles virus are high. For this reason, brain biopsy is generally needed for diagnosis.

TREATMENT AND PROGNOSIS

Rx

There is no established, unequivocally effective treatment of SSPE, but arrest of the disease has been reported in some patients with SSPE after long-term treatment with intrathecal interferon-α with intravenous ribavirin or oral inosine pranobex. About 5% of patients remit spontaneously, but SSPE progresses inexorably to coma, brain stem involvement, and death in 2 to 5 years in the remainder of patients. In immunosuppressed children, such as those with HIV infection, SSPE may be fulminant and may result in death over weeks to 3 to 4 months.

Progressive Rubella Panencephalitis

Progressive rubella panencephalitis is a rare disorder resembling SSPE but caused by rubella virus (Chapter 368). It occurs as a complication of congenital rubella syndrome or, more typically, after childhood rubella. With the advent of widespread rubella immunization, this disorder has been nearly eliminated in the United States.

A hiatus of years separates early infection from the onset of neurologic deterioration, which is characterized by behavioral changes, cognitive impairment, cerebellar ataxia, spasticity, and sometimes seizures. Myoclonus is a less prominent feature than it is in SSPE. Serology or isolation of the virus from brain or peripheral blood lymphocytes confirms the cause. There is no effective treatment, and the prognosis is similar to that for SSPE.

Progressive Multifocal Leukoencephalopathy

DEFINITION

This demyelinating disease is associated with infection of oligodendrocytes by JC virus, a papovavirus that is widely distributed in humans and must undergo genetic rearrangements in its noncoding control region to enable it to replicate efficiently in glial tissue.[4] Progressive multifocal leukoencephalopathy (PML) was the first demyelinating disease to be unequivocally associated with a viral infection.

EPIDEMIOLOGY

Serologic studies indicate that the infection predominantly occurs during childhood, and more than half of the population has been infected by age 20 years. Lesser increments of JC virus seropositivity are observed in each decade thereafter. Despite the wide dissemination of JC virus infection, PML is rarely observed in the absence of underlying cellular immunosuppression. It is also rarely observed in childhood. Until the AIDS epidemic, PML was most commonly observed in patients with lymphoproliferative disorders (62% of cases) and less commonly with myeloproliferative diseases (7%), carcinomatous diseases (2%), other immunodeficiency states, and granulomatous disorders such as tuberculosis and sarcoidosis. The prevalence of PML has increased dramatically during the AIDS pandemic: as many as 5% of AIDS patients develop PML, and AIDS is now the most common underlying disorder associated with it. PML also occurs in association with the administration of monoclonal antibodies for diseases that had not previously been associated with it, including natalizumab, an α4β1 and α4β7 integrin inhibitor used in the treatment of multiple sclerosis (Chapter 411) and Crohn's disease (Chapter 141), and efalizumab, an anti-CD11a antibody that is effective in the treatment of psoriasis (Chapter 438). Other therapeutic agents (e.g., rituximab, an anti-CD20 antibody; brentuximab vedotin, a monoclonal antibody used in treatment of lymphoproliferative disorders; and mycophenolate mofetil) have also been associated with PML, but usually when used to treat disorders that already carry a risk for PML and with incidence rates that are orders of magnitude lower than are seen with natalizumab.[5]

PATHOBIOLOGY

The cardinal feature of PML is demyelination, which is typically multifocal but occasionally unifocal (Fig. 370-2). These lesions may occur in any location in the white matter but have a predilection for the parieto-occipital regions. The lesions range in size from 1 mm to several centimeters; larger lesions may reflect the coalescence of multiple smaller lesions. The other histopathologic hallmark of PML is the presence of hyperchromatic, enlarged oligodendroglial nuclei and enlarged bizarre astrocytes with lobulated hyperchromatic nuclei. Electron microscopic examination reveals the JC virions, which are 28 to 45 nm in diameter and appear singly or in dense crystalline arrays in oligodendroglial cells and, less frequently, in reactive astrocytes. Inflammatory infiltrates are typically absent, except in patients who have reconstitution of their immune system, such as HIV-infected patients being treated with HAART, in whom macrophages and lymphocytes may be found.

CLINICAL MANIFESTATIONS

The clinical hallmark of PML is the presence of focal neurologic symptoms and signs associated with radiographic evidence of white matter disease in the absence of a mass effect.[6] The most common initial symptoms include weakness, speech and language abnormalities, and behavioral and cognitive disturbances. Gait disturbances, sensory loss, and visual impairment all occur in approximately 20 to 30%. Seizures and brain stem symptoms are less common. Signs noted on physical examination parallel the reported symptoms, with weakness, typically a hemiparesis, detected in more than half of patients at initial evaluation. Gait abnormalities, cognitive problems, and speech and language disorders (i.e., dysarthria and dysphasia) are observed in about 25% of patients at initial contact. Limb and trunk ataxia, which reflects cerebellar involvement, is detected in as many as 10% of patients but may occasionally result from severe impairment in position sense (i.e., sensory ataxia). Neuro-ophthalmic symptoms occur in 50% of patients with PML and are often the initial manifestation of the disorder. The most common visual deficit is homonymous hemianopia or quadrantanopia secondary to lesions of the optic radiations. Cortical blindness may develop. Other neuro-ophthalmic manifestations include optic agnosia, alexia without agraphia, and oculomotor abnormalities. Sensory disturbances occur with PML but are distinctly less common than impairment of strength or visual function.

DIAGNOSIS

The diagnosis of PML may be strongly suggested by the clinical manifestations and the radiographic imaging. When the clinical manifestations are coupled with a positive JC virus PCR finding in CSF, the diagnosis of PML is virtually certain. Brain biopsy with demonstration of the characteristic

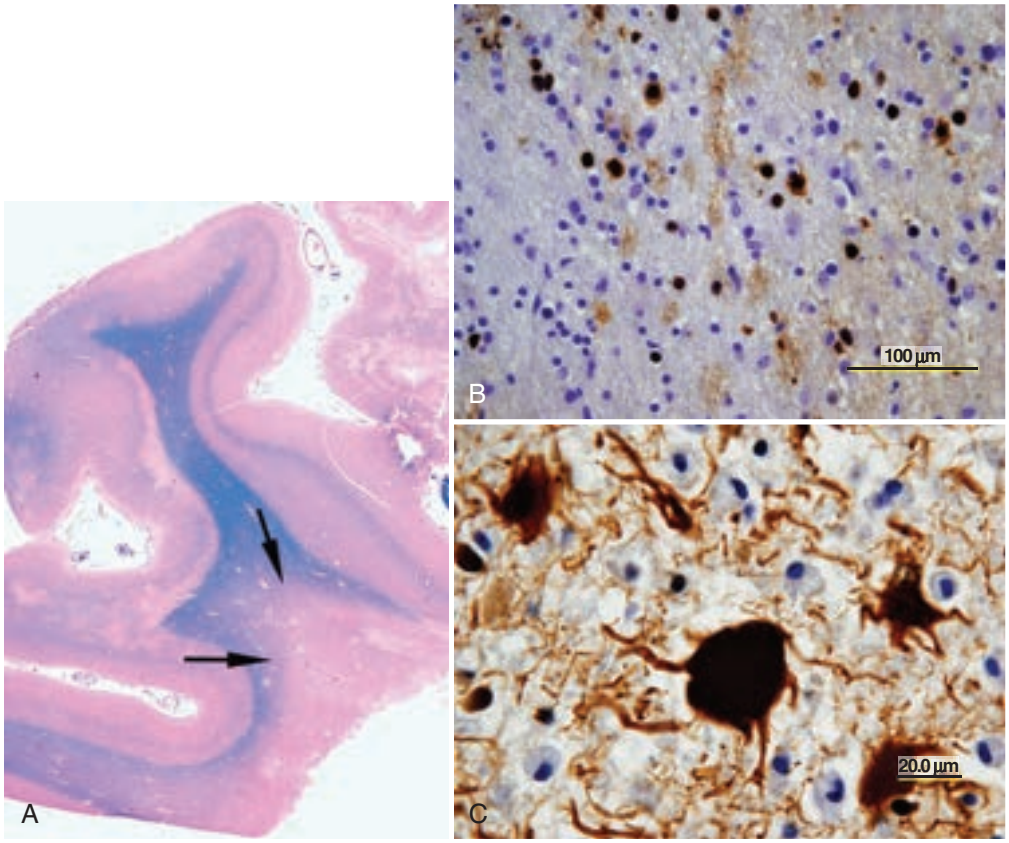

FIGURE 370-2. Pathology of progressive multifocal leukoencephalopathy. A, An area of demyelination is seen in the white matter that fails to stain with Luxol fast blue dye. **B**, Immunohistochemical staining with antibody to papovavirus shows brown-staining nuclei in oligodendrocytes, indicative of JC virus infection. **C**, Immunohistochemical staining for glial fibrillary acidic protein shows large bizarre astrocytes. (Courtesy Dr. Carlos Pardo, Johns Hopkins University, Baltimore, MD.)

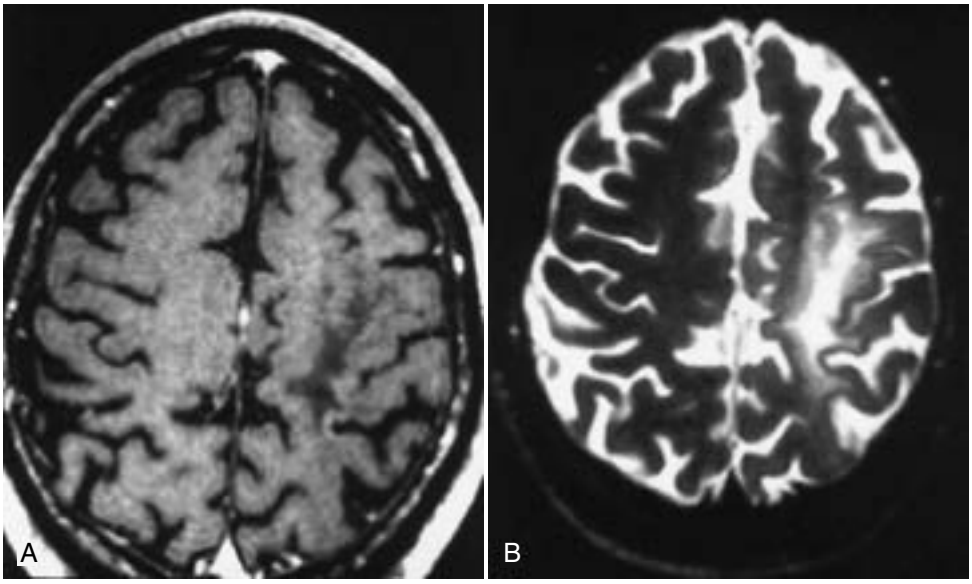

FIGURE 370-3. Cranial magnetic resonance images of progressive multifocal leukoencephalopathy. A, A T1-weighted image shows a hypointense signal abnormality of the left frontal lobe white matter. **B**, On T2-weighted imaging, the lesion is hyperintense.

histopathologic triad of PML coupled with immunohistochemical or electron microscopic evidence of JC virus remains the "gold standard" for diagnosis.

CT of the brain reveals hypodense lesions of the affected white matter that generally have a "scalloped" appearance because of involvement of the subcortical arcuate fibers lying directly beneath the cortex. Cranial MRI shows a hyperintense lesion on T2-weighted or FLAIR images in the affected regions (Fig. 370-3) and usually shows a hypointense lesion on T1-weighted images. Faint contrast enhancement, typically at the periphery of lesions, is seen in approximately 5 to 10% of pathologically confirmed cases of AIDS-associated PML and in 40 to 50% of natalizumab-associated PML on MRI. Intense intralesional nodular enhancement may be seen in patients with immune reconstitution. Frontal and parieto-occipital lobe lesions predominate, but the lesions may be observed in other sites, including the basal ganglia, the internal and external capsules, and the posterior fossa structures (i.e., cerebellum and brain stem).

The results of routine analysis of CSF are not diagnostic, but CSF protein may be elevated. CSF PCR for JC virus is of great value in diagnosis. Currently employed ultrasensitive quantitative PCR for JC virus in the CSF is not only highly sensitive but also specific.[7]

TREATMENT

Rx

Although a number of compounds can prevent JC viral replication in vitro, there currently is no effective medication for PML. The best treatment is restoration of a normal immune system by treating the underlying cause. In AIDS-related PML, an effective antiretroviral regimen improves prognosis.[8] In natalizumab-associated PML, plasma exchange can hasten the drug's elimination. When the immune system recovers, patients may develop a PML immune reconstitution inflammatory syndrome, in which a paradoxical clinical and radiographic worsening accompanies the return of a robust immune response.[9]

PROGNOSIS

Previously, PML was regarded as a fatal illness, with survival averaging 3 to 4 months in the typical patient. After the introduction of HAART, approximately 50% of patients with AIDS-associated PML survive for more than 12 months, often with partial or nearly complete clinical and radiographic recovery.[10] Factors associated with a more benign course include the presence of PML as the heralding manifestation of AIDS, high or climbing $CD4^+$ T-lymphocyte counts, contrast enhancement of the lesions on radiographic studies, and any clinical or radiographic evidence of recovery. JC virus–specific T lymphocytes appear to be critical for control of the infection.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Anderson AM, Mosunjac MB, Corey AS, et al. Simultaneous typical and extraordinary imaging findings of AIDS-associated cytomegalovirus encephalitis. *J Neurol Sci*. 2011;307:174-177.
2. Buchanan R, Bonthius DJ. Measles virus and associated central nervous system sequelae. *Semin Pediatr Neurol*. 2012;19:107-114.
3. Demir N, Cokar O, Bolukbasi F, et al. A close look at EEG in subacute sclerosing panencephalitis. *J Clin Neurophysiol*. 2013;30:348-356.
4. Beltrami S, Gordon J. Immune surveillance and response to JC virus infection and PML. *J Neurovirol*. 2014;20:137-149.
5. Zaheer F, Berger JR. Treatment-related progressive multifocal leukoencephalopathy: current understanding and future steps. *Ther Adv Drug Saf*. 2012;3:227-239.
6. Berger JR. Progressive multifocal leukoencephalopathy. *Handb Clin Neurol*. 2014;123:357-376.
7. Berger JR, Aksamit AJ, Clifford DB, et al. PML diagnostic criteria: consensus statement from the AAN Neuroinfectious Disease Section. *Neurology*. 2013;80:1430-1438.
8. Clifford DB. Progressive multifocal leukoencephalopathy therapy. *J Neurovirol*. 2014;[Epub ahead of print].
9. Bahr N, Boulware DR, Marais S, et al. Central nervous system immune reconstitution inflammatory syndrome. *Curr Infect Dis Rep*. 2013;15:583-593.
10. Yamamoto K, Watanabe K, Kikuchi Y, et al. Long-term functional prognosis of patients with HIV-associated progressive multifocal leukoencephalopathy in the era of combination ART. *AIDS Patient Care STDS*. 2015;29:1-3.

REVIEW QUESTIONS

1. Clues to the diagnosis of cytomegalovirus (CMV) encephalitis in an AIDS patient presenting with cognitive impairment include all but which one of the following?

A. Profound immunosuppression (CD4 count <100 cells/μL)
B. Ring-enhancing lesions in the white matter of the brain on magnetic resonance imaging (MRI)
C. The presence of CMV retinitis characterized by cotton-wool spots and/or the presence of CMV pneumonitis.
D. A polymorphonuclear preponderance in the cerebrospinal fluid (CSF)
E. Progressive ventricular enlargement on brain computed tomography (CT) or MRI

Answer: B Ring-enhancing lesions are not a feature of CMV encephalitis. Brain imaging usually shows progressive increase in the size of the ventricles, as well as ependymal and meningeal enhancement. CMV encephalitis is one of the few viral encephalitides in which polymorphonuclear cells may predominate with the CSF pleocytosis. Evidence of widely disseminated CMV, such as pneumonitis and retinitis, is commonly seen in profoundly immunosuppressed patients in whom CMV encephalitis develops.

2. Which of the following disorders carries the highest risk for progressive multifocal leukoencephalopathy (PML)?

A. Sarcoidosis
B. Multiple sclerosis treated with natalizumab
C. Chronic lymphocytic leukemia
D. AIDS
E. Small cell cancer of the lung

Answer: D Almost 1 in 20 individuals, particularly in the pre-HAART era, developed PML in association with HIV infection. This incidence dwarfs that of any other disorder, although the risk exceeds 1 in 100 in individuals who are JC virus antibody positive and whose multiple sclerosis has been treated with natalizumab for more than 2 years.

3. Which of the following is the most effective treatment for PML?

A. Restoring the immune system by eliminating an offending therapeutic agent or treating HIV infection with antiretroviral therapy
B. Cyclophosphamide
C. Camptothecin
D. Mefloquine
E. Mirtazapine

Answer: A All of the drugs listed have been tried in the treatment of PML, but none has demonstrated a meaningful benefit in clinical studies. However, restoration of the immune system allows for the return of a robust JC virus–specific CD8 response in the central nervous system and may arrest the disease.

371 PARVOVIRUS

NEAL S. YOUNG

DEFINITION

B19 parvovirus, which was discovered in the mid-1970s by electron microscopic observation of an anomalous precipitin reaction of a normal blood donor's serum (occupying position 19 in plate B), was first linked to human disease by the observation of virus-specific immunoglobulin M (IgM) antibody or the virus itself in the sera of sickle cell disease patients suffering transient aplastic crisis (Chapter 163). The common illness caused by the virus was identified later during outbreaks of fifth disease, a highly contagious rash illness of childhood long suspected of having a viral etiology. The ability of parvovirus to persist and to be manifested as an isolated hematologic syndrome was demonstrated by the presence of the virus in fetal liver at autopsy of hydropic newborns and in immunosuppressed patients with chronic pure red cell aplasia (Chapter 165).

The Pathogen

The parvoviruses form small icosahedral capsids of about 25 nm. They have a limited genome of single-stranded DNA. The approximately 5600 nucleotides of B19 parvovirus show remarkably little sequence variation among isolates; two variants, V9 and A6, are of uncertain clinical significance.

The Parvoviridae family contains many pathogenic animal viruses: feline panleukopenia virus, the cause of a fatal agranulocytosis in cats; canine parvovirus, which probably arose from the cat virus as a host range variant in the 1970s to produce a global pandemic and can cause fatal myocarditis in puppies; Aleutian mink virus infection, a model of immune complex disease; and porcine parvovirus, responsible for fetal wastage in pig litters. Antibodies to human adeno-associated viruses, which are dependoparoviruses that are used as gene therapy vectors, occur naturally in humans, but B19 is the only parvovirus known to be pathogenic in humans.

EPIDEMIOLOGY

B19 infection is global; infectivity rates, inferred from the presence of antiparvovirus IgG antibody in sera, are similar worldwide. Only isolated populations, Amazonian tribesman, and residents of remote islands off the coast of Africa have escaped exposure. B19 parvovirus infection is common in childhood, and half of 15-year-old adolescents have specific antiparvovirus B19 antibodies. Infection continues throughout adult life, and most elderly people are seropositive. In temperate climates, most infections occur in the spring, with small epidemics every few years being typical. Transmission is respiratory by droplet spread, and secondary infection rates among household contacts are high. Nosocomial infection can occur, and B19 parvovirus has been transmitted in blood products, especially pooled components such as factor VIII and IX concentrates. Producers of plasma derivatives now voluntarily screen by quantitative measurement of B19 DNA to reduce the risk for iatrogenic transmission.[1] The lack of a lipid envelope and the stable DNA genome make parvoviruses notoriously resistant to heat inactivation and solvent detergents.

PATHOBIOLOGY

The biology of the Parvoviridae makes them especially dependent on helper function from host cells or other viruses.[2] The autonomous parvoviruses propagate in actively dividing cells; the family Parvoviridae includes disease-causing animal parvoviruses. Adeno-associated viruses grow in tissue cultures infected with adenoviruses and herpesviruses and are popular vectors for gene transduction and therapy. B19 is the type member of the *Erythrovirus* genus, which includes very similar simian viruses, all of which are best propagated in the erythroid progenitor cells that are responsible for red blood cell production in the bone marrow. Active replication of virus can be detected by the presence of double-stranded intermediate forms by simple DNA hybridization methods. The transcription map of the erythroviruses differs markedly from that of other Parvoviridae. Only three genes produce proteins of known function. Many antigenic determinants recognized by the host immune system are located in helical loops that form the surface of each capsomere. Most of the capsid is composed of a major structural protein, called VP2, but about 5% of the capsid is the minor structural protein, VP1, which differs from VP2 only by an additional 226 amino acids at the amino terminus; this VP1 unique region is located external to the capsid surface and contains linear epitopes recognized by neutralizing antibodies.

The only known natural host cell of B19 parvovirus is the human erythroid progenitor. The tropism of the virus for an erythroid cell host results from its cellular receptor, globoside, a neutral glycolipid also known as erythrocyte P antigen. Rare individuals with the p phenotype, who congenitally lack globoside on their erythrocytes, are genetically insusceptible to B19 parvovirus infection; they show no serologic evidence of previous infection, and their marrow erythroid progenitors proliferate normally in the presence of high concentrations of virus. Parvovirus kills erythroid progenitors by expression of its nonstructural protein, and it is possible that some cells, such as megakaryocytes, may be lysed by restricted expression of viral proteins in the absence of viral propagation. B19 can be efficiently propagated in tissue culture of primary human hematopoietic cells in which erythropoietic differentiation is stimulated by erythropoietin.

The humoral immune response is dominant in B19 parvovirus infection. Natural antibody production correlates with disappearance of the virus from blood, and the presence of IgG appears to confer lasting protection against a second infection. Parvovirus infection can persist if immunoglobulin production is defective such that antibody fails to neutralize the virus; reactivity of antibodies to the unique amino-terminal region of VP1 is especially important.

CLINICAL MANIFESTATIONS

Fifth Disease

Most B19 parvovirus infections are asymptomatic. The most common clinical manifestation of infection is erythema infectiosum, or fifth disease, a rash illness of childhood characterized by a "slapped cheek" appearance (Fig. 371-1). In adult volunteers inoculated intranasally with B19, nonspecific influenza-like complaints occurred early along with viremia; the cutaneous eruption a week later corresponded to the appearance of antiviral antibodies. These more specific symptoms of B19 parvovirus infection are secondary to immune complex formation and deposition. Serologic testing generally shows seroconversion, IgM antibody or the appearance of IgG antibody to parvovirus. The rash of fifth disease may be evanescent, and recurrences can be provoked by sunlight, heat, emotion, or exercise. Fifth disease can be confused with rubella. In adults, the rash is less characteristic, can present as palpable purpura, may be associated with pruritus in up to 50% of patients,[3] and can be difficult to visualize in dark-skinned individuals.

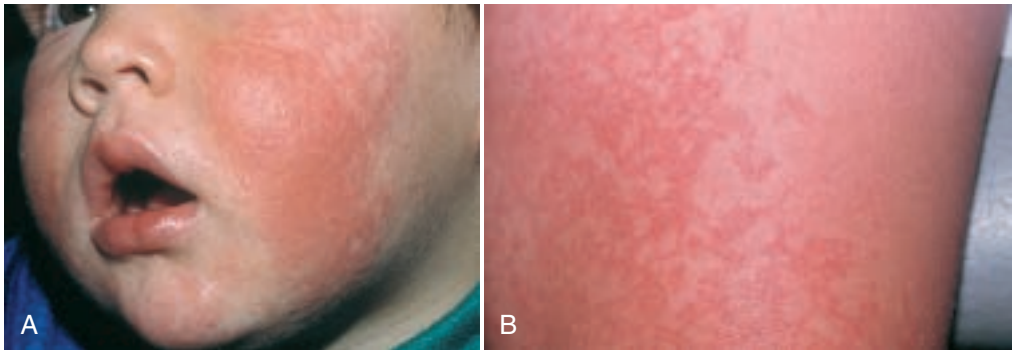

FIGURE 371-1. Erythema infectiosum. In this infection by parvovirus B19, a child will develop prominent erythema of the cheeks, "slapped cheeks" (A), followed by a lacelike erythema on the extremities (B) and buttocks. It is also known as fifth disease.

B19 Arthropathy

In contrast to the mild course in children with fifth disease, acute parvovirus infection in adults, particularly middle-aged women, may cause significant arthropathy.[4] Not only arthralgia but also a true inflammatory arthritis occurs in about 50% of older patients. Symmetrical joint involvement of the hands, ankles, knees, and wrists can resemble rheumatoid arthritis (Chapter 264), and the test result for rheumatoid factor may be positive. B19 arthropathy usually resolves within a few weeks; joint destruction does not occur. Parvovirus is not the cause of rheumatoid arthritis, but case reports suggest that B19 infection may mimic, precipitate, or worsen a variety of rheumatologic diseases, including juvenile rheumatoid arthritis, systemic lupus erythematosus, and fibromyalgia.

Transient Aplastic Crisis

In persons with underlying hemolysis or a high demand for production of circulating erythrocytes, acute B19 parvovirus infection causes transient aplastic crisis, an abrupt cessation of red blood cell production that exacerbates or, in previously compensated states, provokes severe anemia. Erythropoiesis is temporarily suppressed in all B19 parvovirus infections, but hemoglobin levels remain stable because of the long lifespan of erythrocytes. The anemic crises associated with low or absent reticulocytes in hereditary spherocytosis and sickle cell disease are virtually always secondary to B19 parvovirus infection. Parvoviremia is present in patients with transient aplastic crisis, and red cell production resumes once antibodies to the virus are produced and the infection is cleared. Transient aplastic crisis is generally a unique event in the patient's life, thus suggesting induction of long-lasting protective immunity. Although it is self-limited, aplastic crisis often requires transfusion and can lead to severe, occasionally fatal anemia that precipitates congestive heart failure and cerebrovascular accidents. Transient aplastic crisis is associated with a stereotypical bone marrow morphology, absence of maturing erythroid precursors, and the presence of "giant pronormoblasts" (Fig. 371-2) that are the cytopathic effect of parvovirus infection.

White blood cell and platelet counts may fall modestly during transient aplastic crisis, especially in patients with functioning spleens. Occasional cases of agranulocytosis may be due to B19; thrombocytopenia and pancytopenia have been reported, and B19 can precipitate a benign virus-associated hemophagocytic syndrome.

Persistent Infection

In patients who cannot mount an appropriate host antibody response, B19 parvovirus persists in the circulation, often at extremely high levels ($>10^{12}$ genome copies per milliliter). Patients do not develop the clinical features of fifth disease but instead have an entirely hematologic syndrome of pure red cell aplasia. The anemia is severe and requires transfusion; reticulocytes are absent from blood, as are erythroid precursors from marrow. Observation of giant pronormoblasts in the marrow may lead to the diagnosis. The failure to produce neutralizing antibodies to B19 parvovirus occurs in patients with congenital immunodeficiency (Nezelof's syndrome), with iatrogenic immunodeficiency (chemotherapy or immunosuppressive drugs), and with acquired immunodeficiency. Pure red cell aplasia secondary to parvovirus may be the first manifestation of the acquired immunodeficiency syndrome (AIDS), but this presentation is less common in the era of highly active antiretroviral therapy. Epidemiologic studies have suggested that parvovirus infection may worsen the manifestations of malaria (Chapter 345), especially the severity of anemia.[5]

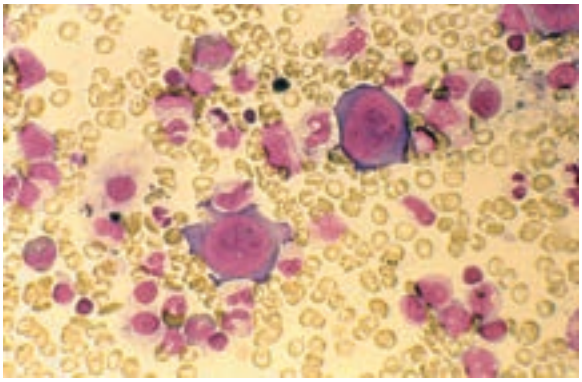

FIGURE 371-2. Bone marrow aspirate of a patient with chronic pure red cell aplasia secondary to persistent B19 parvovirus infection. Mature erythroid precursors are absent, and the prominent giant pronormoblasts are typical of B19 infection.

Hydrops Fetalis

B19 parvovirus infection of the pregnant mother followed by transplacental transmission to the fetus can lead to an adverse outcome, either miscarriage or hydrops fetalis.[6] Parvovirus infects the fetal liver, the site of erythrocyte production during early development. Hydrops is the result of severe anemia as well as perhaps myocarditis, contributing to congestive heart failure. Prospective studies have led to an estimated 30% risk for transplacental infection and 9% risk for fetal loss in women who are exposed to B19 during pregnancy. Infection during the second trimester poses the greatest risk for birth of a hydropic infant; B19 parvovirus accounts for 10 to 20% of all cases of nonimmune hydrops fetalis. The risk for spontaneous abortion resulting from first-trimester infections has been more difficult to quantitate. The likelihood of an infection increases in epidemic years and correlates with the level of contact of the pregnant woman with children. Although most B19 infections during pregnancy probably do not lead to either loss of the fetus or congenital anomalies, B19 infection is a cause of fetal death. Congenital malformations have not been consistently associated with intrauterine parvovirus infection. However, severe anemia at birth with bone marrow histology consistent with either constitutional pure red cell aplasia (Diamond-Blackfan anemia) or congenital dyserythropoietic anemia has occurred in infants salvaged by in utero blood transfusions or exchange transfusion at birth.

Other Syndromes

Elevated hepatic aminotransferase levels can accompany fifth disease, and parvovirus infection has been associated with severe but usually self-limited hepatitis in some children.[7] The presence of B19 genetic sequences in cardiac tissue has led to a diagnosis of parvovirus myocarditis.[8] Serologic and DNA evidence of B19 infection implicated parvovirus in some patients with necrotizing vasculitis, Kawasaki disease, Henoch-Schönlein purpura, and giant cell arteritis. Glove-and-sock syndrome, an exanthem localized to the hands and feet and consisting of edema, erythema, paresthesia, and pruritus has

TABLE 371-1 DIAGNOSIS OF PARVOVIRUS B19

DISEASE	IgM	IgG	B19 DOT BLOT*	B19 PCR
Fifth disease	+++	++	–	+
Polyarthropathy syndrome	++	+	–	+
Transient aplastic crisis	+/–	+/–	++	++
Persistent anemia	+/–	+/–	++	++
Hydrops/congenital infection	+/–	+	+/–	++
Previous infection	–	++	–	+/–

*Sensitivity about 10^6 genome copies per milliliter.
Ig = immunoglobulin; PCR = polymerase chain reaction.

been linked to B19. Chronic fatigue syndrome may follow parvovirus infection. Meningitis, encephalitis, and a variety of neurologic complications may occur with fifth disease and parvovirus infection.[9]

False-positive results arise when the diagnosis of infection rests on detection of amplified B19 genome by polymerase chain reaction, and, furthermore, B19 parvovirus can persist at low levels in normal individuals for many months after infection.

DIAGNOSIS

Laboratory diagnosis relies on serologic and DNA tests[10] (Table 371-1). Virus-specific antibodies are measured in standardized commercial solid-phase enzyme-labeled immunoassays, generally using recombinant capsid proteins. "Capture" formats are preferred to detect serum IgM, which is first bound to a solid phase coated with anti-μ-chain antibodies, followed by the addition of viral antigen and an antiviral monoclonal antibody. IgM antibodies are diagnostically positive in almost all cases of fifth disease at initial evaluation and appear within a few days of the onset of transient aplastic crisis; IgM may persist for months after acute infection. IgG is usually assayed in conventional indirect assays. IgG circulates later than IgM, generally at the end of the first week of illness. Although titers of IgG are generally highest in the year after an acute infection, substantial interindividual variation and the presence of IgG in a large proportion of the population make measurement of IgG less helpful than other tests for diagnosis of parvovirus. DNA assays are required for persistent B19 infection, in which antibody production is absent or minimal. Parvovirus can also be found in the sera of patients with early transient aplastic crisis. Direct hybridization methods are reliable, and they detect clinically relevant viral titers of greater than 10^6 international units (orders of magnitude below levels present in both acute and persistent infection). Gene amplification methods are more sensitive but less reliable because of false-positive results. Virus can be detected in amniotic fluid, and both virus and IgM antibody to B19 are found in umbilical cord blood; the mother's serum will show seroconversion during pregnancy, but maternal IgM may be absent at the onset of hydrops fetalis.

PREVENTION

Effective vaccines exist for animal parvoviruses, and human B19 infection can also probably be prevented. A recombinant immunogen in development for the human virus lacks DNA and is therefore noninfectious; the empty capsids have been engineered to overexpress the highly immunogenic minor structural protein VP1, and a single 2.5-μg dose of empty capsids elicited excellent neutralizing antibody responses in normal volunteers. Vaccination could prevent transient aplastic crisis in patients with sickle cell disease and other hemolytic anemias, pure red cell aplasia in some immunodeficient individuals, and hydrops if seronegative mothers were inoculated early in pregnancy.

TREATMENT

Most parvovirus infections in normal children and adults do not require specific therapy. Isolation of infected individuals is impractical, with the exception of hospitalized cases. Pure red cell aplasia and the underlying persistent B19 parvovirus infection can be dramatically terminated by discontinuation of immunosuppressive therapy or institution of effective antiretroviral drugs in patients with AIDS. Commercial immunoglobulins are a good source of antibodies to parvovirus, and persistent B19 infection responds to a 5- or 10-day course of IgG at 0.4 g/kg with a prompt decline in serum viral DNA, as measured by hybridization methods, accompanied by reticulocytosis and increased hemoglobin levels. This regimen has been curative in congenital immunodeficiency, but parvovirus in AIDS patients can persist at lower levels, and relapses of anemia may require repeated IgG administration. Immunoglobulin therapy can precipitate fifth disease rash and arthralgia.[11] Hydrops fetalis may resolve spontaneously, but intrauterine blood transfusions have been used with apparent success. Chronic arthropathy has been treated symptomatically with anti-inflammatory drugs, and there is not a role for the administration of immunoglobulin. As important as recognizing parvovirus infection is avoiding misinterpretation of laboratory studies, such as positive IgG serology or borderline IgM and DNA test results, and misguided maneuvers that delay appropriate alternative treatments.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Soucie JM, Monahan PE, Kulkarni R, et al. Evidence for the continued transmission of parvovirus B19 in patients with bleeding disorders treated with plasma-derived factor concentrates. *Transfusion*. 2013;53:1143-1144.
2. Rogo LD, Mokhtari-Azad T, Kabir MH, et al. Human parvovirus B19: a review. *Acta Virol*. 2014;58:199-213.
3. Mage V, Lipsker D, Barbarot S, et al. Different patterns of skin manifestations associated with parvovirus B19 primary infection in adults. *J Am Acad Dermatol*. 2014;71:62-69.
4. Tello-Winniczuk N, Díaz-Jouanen E, Díaz-Borjon A. Parvovirus B19-associated arthritis: report on a community outbreak. *J Clin Rheumatol*. 2011;17:449-450.
5. Toan NL, Sy BT, Song le H, et al. Co-infection of human parvovirus B19 with Plasmodium falciparum contributes to malaria disease severity in Gabonese patients. *BMC Infect Dis*. 2013;13:375.
6. Dijkmans AC, de Jong EP, Dijkmans BA, et al. Parvovirus B19 in pregnancy: prenatal diagnosis and management of fetal complications. *Curr Opin Obstet Gynecol*. 2012;24:95-101.
7. Bihari C, Rastogi A, Saxena P, et al. Parvovirus B19 associated hepatitis. *Hepat Res Treat*. 2013;2013:472027.
8. Molina KM, Garcia X, Denfield SW, et al. Parvovirus B19 myocarditis causes significant morbidity and mortality in children. *Pediatr Cardiol*. 2013;34:390-397.
9. Barah F, Whiteside S, Batista S, et al. Neurological aspects of human parvovirus B19 infection: a systematic review. *Rev Med Virol*. 2014;24:154-168.
10. Doyle S. The detection of parvoviruses. *Methods Mol Biol*. 2011;665:213-231.
11. Crabol Y, Terrier B, Rozenberg F, et al. Intravenous immunoglobulin therapy for pure red cell aplasia related to human parvovirus B19 infection: a retrospective study of 10 patients and review of the literature. *Clin Infect Dis*. 2013;56:968-977.

REVIEW QUESTIONS

1. A 5-year-old is brought by her concerned parents because of sudden onset of a rash over her face and chest. The patient has no other complaints or physical findings. Her older sibling suffered the same symptoms, and a fifth disease "epidemic" has been publicized in the local school district. Which is the appropriate action?

A. Order appropriate serologic tests for viruses that can cause rash, including measles, rubella, human herpesvirus 6, and parvovirus; then treat her accordingly.
B. Reassure the child and family.
C. Reassure the family, but recommend that the child not attend school until the rash clears.
D. Test all family members for antibodies to parvovirus B19 and for the presence of virus by sensitive DNA assays.
E. Assess for B19 DNA; if positive, administer a short course of immunoglobulin.

Answer: B Expensive and intrusive testing is to be avoided in the evaluation of acute parvovirus infection in children, who are almost always minimally symptomatic and will have an uncomplicated clinical course. Fifth disease frequently appears as an epidemic, often at 3-year intervals in a population. Further evaluation would be warranted only in the unlikely event of more serious symptoms in any family member. Because the cutaneous eruption occurs after the formation of antibodies to parvovirus, the child is no longer infectious and may attend school.

2. You follow an 18-year-old girl who has sickle cell disease with only occasional pain crises on hydroxyurea therapy. Her parents inform you that she is more tired than usual, unwilling to attend school, and sleeping much of the day. You are aware of an epidemic of fifth disease in the community. What is your response to the child's parents?

A. Reassure them that she is suffering from parvovirus infection, which is not a serious infection of childhood.
B. Immediately hospitalize the child and administer human immunoglobulin in order to clear any active parvovirus infection.
C. Recommend laboratory testing in the next few days, including a blood count and parvovirus B19 serology. If the hemoglobin is below normal, transfuse her; then if anti-B19 immunoglobulin G (IgG) is detected, administer immunoglobulin.
D. Have the child seen that day to determine her hemoglobin level and reticulocyte count. If both are lower than expected, transfuse packed red blood cells.
E. Ask the parents to be alert to new, worrisome complaints, either a skin rash or joint symptoms, that would then trigger a more extensive evaluation.

Answer: D The symptoms and setting are highly suggestive of transient aplastic crisis of sickle cell disease, in which marked worsening of anemia can be life-threatening. The child should be seen emergently; a low reticulocyte count establishes the diagnosis, and erythrocyte transfusion will alleviate the symptoms. In transient aplastic crisis, serology is not useful because specific IgG and IgM to parvovirus likely will not be detected. Viral levels of B19 would be expected to be very high. However, children with sickle cell disease mount normal responses to B19 antigens and will resolve the infection within a few days or a week, so immunoglobulin therapy is unnecessary.

3. A 38-year-old mother of three presents with a history of several weeks of severe joint pains and swelling, especially of her hands and knees, as well as morning stiffness and fatigue. She relates that two of her children had a rash a few weeks preceding her illness and that a community outbreak of parvovirus was suspected. On physical examination, she has mild swelling and marked tenderness over the proximal phalangeal joints, tenderness on palpation of her knees, but no cutaneous eruption. Laboratory screening is largely unrevealing, including normal blood counts and routine chemistries, sedimentation rate, antinuclear antibody, and rheumatoid factor. However, serologic testing for antibodies to B19 parvovirus shows a high level of anti-IgG and above normal IgM. What do you recommend?

A. Reassure the patient that she likely has a self-limited illness and prescribe nonsteroidal anti-inflammatory drugs for symptom control.
B. Perform polymerase chain reaction testing to detect B19 parvovirus; if the virus is detected, administer immunoglobulin.
C. Perform an invasive evaluation—including arthroscopy, synovial biopsy, bone marrow biopsy, and cardiac imaging—to detect other evidence of parvovirus infection.
D. Monitor serologic assays and perform B19 by gene amplification before deciding on therapy.
E. Refer her to a specialist because there is insufficient evidence to link her symptoms to a parvovirus infection.

Answer: A Arthopathy after a parvovirus infection may be debilitating in an adult. The pattern of symptoms and joint involvement can mimic rheumatoid arthritis, but without synovial destruction. The diagnosis is clinical, and tissue sampling is not known to be helpful. The typical patient relates a convincing history of exposure, as for example to children with fifth disease, and the laboratory shows evidence of past infection without viremia. The pathophysiology is not well understood, but immunoglobulin is not indicated, nor has it been effective. In most patients, only symptomatic treatment is required, and gradual resolution without sequelae is expected.

4. A 23-year-old woman in the second trimester of pregnancy reports that her children were exposed to a parvovirus B19 outbreak at school and now show the "slapped cheek" rash. She feels well. What is your advice?

A. Because parvovirus leads to early fetal loss, mid-trimester abortion, and congenital malformation, you should recommend early termination of the pregnancy.
B. Watchful waiting for symptoms
C. Intrauterine red blood cell transfusion
D. Immunoglobulin injections
E. Serologic testing and ultrasound

Answer: E Most women who are exposed to B19 during pregnancy do not become infected, and only a minority of those who are infected have adverse consequences in their fetuses. The most prudent approach is to determine whether infection has occurred by evidence of IgM antibody or IgG seroconversion; if infection is documented, the course of pregnancy should be followed with ultrasounds of the fetus. The mother will mount a neutralizing antibody response, so immunoglobulin is not necessary. Intrauterine transfusion is of uncertain benefit even to the hydropic fetus.

5. A 30-year-old man with severe anemia is found on screening to be HIV positive. Laboratory testing shows a very low reticulocyte count, and there are no erythroblasts on his bone marrow examination. He is seronegative for parvovirus B19. What therapy do you recommend?

A. Blood transfusions
B. Immunoglobulin therapy followed by highly active antiretroviral therapy
C. Corticosteroids
D. Thymectomy
E. Highly active antiretroviral therapy

Answer: B Pure red cell aplasia in an HIV-infected individual is likely secondary to persistent parvovirus infection, and anemia may be the presenting manifestation of AIDS. In the absence of an adequate immune response, the virus infects red cell progenitors. Serology is often negative because of the absence of an immune response. However, the diagnosis can be established by detecting the virus itself in blood, where it should be abundant. This syndrome has become less prevalent in the era of highly active antiretroviral therapy, which is indicated in this patient. In addition, a course of immunoglobulin therapy is likely to clear the parvovirus and quickly resolve the pure red cell aplasia.

372

SMALLPOX, MONKEYPOX, AND OTHER POXVIRUS INFECTIONS

INGER K. DAMON

DEFINITION

Human illness caused by a poxvirus is characterized by a cutaneous manifestation; illness may be localized or systemic, depending on the particular poxvirus and the route of introduction. DNA-based assays, including DNA sequencing, are the most precise methods for identification and differentiation of poxvirus genera, species, strains, and variants. The guanosine plus cytosine content of orthopoxviruses, yatapoxviruses, molluscum contagiosum virus, and parapoxviruses is approximately 33%, 32%, 60%, and 63%, respectively.

The Pathogens

All poxviruses described in this chapter (Table 372-1) belong to the family Poxviridae, subfamily Chordopoxvirinae.

EPIDEMIOLOGY

Recognition of the epidemiologic characteristics of poxvirus diseases is valuable in assessing the potential etiologic agent of a particular suspected poxvirus lesion. Knowledge of zoonotic reservoirs, geographic localizations, and capacity for epidemic transmission is critical for clinical assessment and control measures. All human poxvirus infections are zoonotic in nature, with the exception of molluscum contagiosum and variola, which are solely human pathogens. Transmission of virus to humans and to animals is through direct contact with lesions or by fomites.

Parapoxvirus and molluscipoxvirus infections are endemic worldwide; orthopoxvirus and yatapoxvirus infections are geographically restricted, probably by the distribution of competent reservoir hosts. With the exception of variola virus (the causative agent of smallpox, a disease declared eradicated in 1980), none of these diseases is required to be reported to public health systems; furthermore, diagnostics are not widely available, so it is difficult to estimate disease incidence and prevalence with any certainty. In addition to variola virus, Congo Basin clade monkeypox viruses are considered to be select agents by the U.S. government; they must be reported and appropriately handled if discovered and if samples are maintained in the United States.

Orthopoxvirus

The epidemiology of smallpox, caused by the orthopoxvirus variola, is understood through detailed studies conducted during the end of the eradication campaign. Interhuman transmission of variola virus generally occurred through the inhalation of large airborne respiratory droplets of infectious

TABLE 372-1 TAXONOMY OF POXVIRUSES KNOWN TO INFECT HUMANS

GENUS	SPECIES
Orthopoxvirus	Variola virus, vaccinia virus, cowpox virus, monkeypox virus
Parapoxvirus	Orf virus, milker's node virus, bovine papular stomatitis virus, sealpox virus
Yatapoxvirus	Tanapox virus (Yaba-like disease virus), Yaba monkey tumor virus
Molluscipoxvirus	Molluscum contagiosum virus

variola virus. Transmission usually required prolonged face-to-face or other close contact, although airborne transmission over longer distances had been reported. Transmission by fomites or contact with infectious material from the rash also occurred. Aggregate data, collected during the smallpox eradication campaign, suggest a secondary attack rate of 58.4% in unvaccinated close or household contacts and a secondary attack rate of 3.8% in previously vaccinated close or household contacts. Case-fatality rates for variola major varied with the type of disease manifested, but aggregate rates of 10 to 30% in various outbreaks have been recorded. Severity of disease correlated with rash burden and was also more severe in children and pregnant women. Variola alastrim minor, a variant of variola with a case-fatality rate of less than 1%, has similar human-to-human disease transmission characteristics.

Monkeypox has a more complex epidemiology. The virus is zoonotic, and two genetically discrete virus clades have been described, each with apparent distinct clinical and epidemiologic parameters. Human infections in western and central Africa were first identified in 1970. Investigations in the Congo basin country Zaire, now the Democratic Republic of Congo, demonstrated that human-to-human transmission of monkeypox was less prevalent than that of smallpox. The secondary attack rate in unvaccinated contacts of monkeypox cases was calculated to be 9.3% versus 37 to 88% for smallpox. Previous smallpox vaccination (3 to 19 years previously) appeared to be 85% protective in preventing disease acquisition in contacts and also ameliorated the severity of disease. Overall, most identified cases acquired disease from presumed animal exposure; only 28% of cases were ascribed to person-to-person transmission. A case-fatality rate of approximately 10% was observed in unvaccinated persons, and the majority of fatalities and the most severe disease manifestations were observed in children younger than 5 years. Serosurveys suggested that subclinical infection may have occurred in up to 28% of close contacts of monkeypox patients in some communities; this relatively low rate may contribute to the rarity of sustained generations of human-to-human transmission in household and other close contact situations.

Among primary cases, recent close contact—through hunting, skinning, killing, cooking, or playing with carcasses—was identified with *Cercopithicus, Colobus,* and *Cercocebus* (primate); *Cricetomys* (terrestrial rodent); and *Funisciurus* and *Heliosciurus* (squirrel species). Samples of animals collected in areas of western and central Africa surrounding human cases demonstrated orthopoxvirus and sometimes monkeypox-specific seroprevalence in various members of these species, except in *Cricetomys* species. The prevailing hypothesis was that squirrel species were the probable reservoir of disease.

The disease reemerged in 1996 in the Democratic Republic of Congo, this time with 88% of cases derived from secondary human-to-human contact presumed to be because of the cessation of routine smallpox vaccination in 1980 after the eradication of smallpox. Mortality was only 1%. Ecologic serosurveys showed orthopoxvirus seroprevalence in terrestrial rodents *(Cricetomys emini)* and in one domestic pig *(Sus scrofa).*

Monkeypox virus was introduced to the United States in 2003 through a consignment of animals from the West African country of Ghana. The virus was identified as belonging to a distinct clade of monkeypox that included previous West African monkeypox isolates as well as isolates derived from earlier outbreaks in primate colonies. The U.S. cases had a less pronounced rash and a less severe illness, with no mortality or human-to-human transmission. When cases in the United States (2003) were compared with cases in Congo Basin (1980 to 1986), disease in the United States was less severe based on clinical criteria, the extent of the rash, and the case-fatality rate after controlling for age and vaccination status.[1] These data, along with comparison of the genomes of West African and Congo Basin monkeypox viruses, suggest at least two populations or clades of monkeypox virus. The apparent decreased pathogenicity and transmissibility of West African clade monkeypox virus infection led, in part, to its declassification as a U.S. select agent in 2013.

Cowpox virus is found in Europe and Asia and is maintained in rodents; in Britain, the reservoirs are bank voles and wood mice. Human infection is a zoonosis. The domestic cat has been a common source of human infection, which probably explains the occurrence of cases in children; 26% of 54 cases occurred in children younger than 12 years. Most feline and human cases occur between July and October, with only occasional cases between January and June. Recent outbreaks in Europe have documented pet rats, or feeder rats, to be a source of transmission to humans. No case of bovine cowpox has been detected since 1976. Cowpox virus also is prevalent in European zoos, where cheetahs, lions, anteaters, rhinoceros, elephants, and okapi have occasionally transmitted infection to animal handlers.

Vaccinia is the live virus contained in preparations of the smallpox vaccines used to eradicate smallpox. In the United States, the vaccine is recommended for laboratory workers who use replicative orthopoxviruses and for selected military personnel. Contacts of vaccinees occasionally develop vaccinia infections. The origin of vaccinia is unknown, and no natural host for the virus is known. Recent studies have further distinguished cowpox and vaccinia viruses, both of which have been used as smallpox vaccines, and have identified at least three groups of cowpox and most vaccinia viruses group with one of these cowpox clades.[2] Vaccinia "variants" have been described, including buffalopox from contact with infected animals in India and vaccinia viruses in cattle handlers in Brazil.

Parapoxvirus

Human infection is an occupational hazard of farm workers, abattoir workers, veterinary surgeons, and students. It is most common in the lambing and calving seasons and among sheep workers.

Factors responsible for ongoing transmission have been attributed both to the environmental stability of orf virus in scab material and to the manifestation of chronic infections in some animals. A new parapoxvirus has been identified in deer hunters found to have cutaneous lesions after hunting and field dressing deer.[3]

Molluscipoxvirus

Molluscum contagiosum virus occurs worldwide, and increasing reports of the disease have paralleled the number of reported cases of acquired immunodeficiency syndrome (AIDS).[4] Traditional modes of transmission are associated with mild skin trauma such as abrasions, direct contact with a lesion, and fomites (e.g., shared towels) in some cases. However, the disease appears to be sexually transmitted, and genital lesions are more common than lesions elsewhere on the body. Children in daycare or school situations may transmit the disease to other children. Secondary spread of lesions may occur by autoinoculation (excoriation of primary lesions and spread to areas of normal skin) as well as by shaving. No known animal reservoir exists.

Yatapoxvirus

Tanapox virus is restricted to Africa, principally to Kenya and the Democratic Republic of Congo, and probably has a simian reservoir. Direct primate-to-human transmission through a break in skin has rarely been described in animal handlers, but an insect or arthropod intermediary may be involved in the transmission of tanapox virus to humans. No human-to-human transmission has been reported. Yaba monkey tumor virus causes localized infections after contact with infected primate lesions. Little is known about the epidemiology of this virus.

PATHOBIOLOGY

The majority of smallpox infections were initiated by inhalation of respiratory droplets and implantation of virus on the oropharyngeal and respiratory mucosa. No primary localized site of infection was evident if the route of exposure was by inhalation. Disease could also be introduced through suspensions of virus obtained from scabs of patients that were introduced percutaneously and constituted the practice of variolation. In these cases (when skillfully administered), illness was usually less virulent, a localized primary infectious lesion was present, and the asymptomatic incubation period was truncated.

After entry, virus moves to local lymph nodes and then disseminates to the reticuloendothelial system to replicate further. At this time, the individual is asymptomatic. In 10 to 14 days, secondary viremia occurs and heralds the prodrome of symptomatic illness. During this time, virus seeds the oropharynx and epidermis. The absence of a keratinized structure in the mucosa of

the oropharynx leads to ulceration and release of virus in saliva; virus replicates in the epidermis to cause the characteristic macular, papular, and vesicular eruptions of smallpox.

In experiments in monkeys, high levels of type I interferons, interleukin-6, and interferon-γ are seen; D-dimers and thrombocytopenia suggest disseminated intravascular coagulation. Apoptosis with loss of T cells in lymphoid organs was also observed. Of note, tumor necrosis factor-α (TNF-α) levels were minimal in the infected animals, with a notable decrease in the expression of genes regulated by nuclear factor κB and TNF-α.

In humans, the viral lesions characteristic of illness primarily develop in the epidermis, where the cells of the malpighian layer swell and vacuolate to undergo ballooning degeneration. The cytoplasm continues to enlarge, loss of nuclear material is noted, and coalescence of vacuoles through cell rupture creates reticulating degeneration of the middle and upper layers of the stratum spinosum. In the next stages, the vesicle is formed. High titers of virus are found within the lesions. In mucosal surfaces, the absence of a horny layer allows the necrosis caused by proliferation of virus within the epithelium to create ulcers and leads to liberation of large quantities of virus into the oropharynx. Evaluation of other organs in human smallpox has been done only in select autopsy cases. Mild pathologic changes are seen in the lungs.

CLINICAL MANIFESTATIONS

Orthopoxvirus

Smallpox

Naturally acquired variola virus infection is characterized by fever and a distinctive rash, with several different clinical presentations (Table 372-2).[5] When the disease still existed, an asymptomatic incubation period of 10 to 14 days (range, 7 to 17) was followed by a fever that quickly rose to about 103° F (38° to 40° C), sometimes with dermal petechiae. Associated constitutional symptoms included backache, headache, vomiting, and prostration. Within a day or two after incubation, a systemic rash with a characteristic centrifugal distribution (i.e., lesions present in greater numbers on the oral mucosa, face, and extremities than on the trunk) appeared. The fever typically abated as the rash developed. Lesions commonly appeared on the palms and soles. The rash lesions were initially macular and then advanced to the papular stage, at which point they enlarged and progressed to a vesicle by day 4 to 5 and a pustule by day 7. When lesions became pustular, the fever typically returned. Lesions became encrusted and scabbed by day 14 and then sloughed off. Skin lesions in the vesicular and pustular stages were deep seated and in the same stage of development in any one area of the body (Fig. 372-1). The ordinary disease type was subgrouped into three categories based on the extent of rash on the face and the body: confluent, semiconfluent, and discrete. In *ordinary confluent disease,* no area of skin was visible between vesiculopustular rash lesions on the trunk or the face. In *ordinary semiconfluent* and *discrete disease,* patches of normal skin were visible between rash lesions on the trunk and face, respectively. Less severe manifestations (modified smallpox or variola sine eruptione) occurred in both unvaccinated and, more commonly, vaccinated individuals. The mortality rate correlated with the rash burden.

Four main clinical types of variola can be subgrouped according to the World Health Organization (WHO) classification schema: *ordinary smallpox* (≈90% of cases) produced viremia, fever, prostration, and rash, and mortality rates were generally proportional to the extent of rash and ranged as described earlier; *(vaccine) modified smallpox* (5% of cases) produced a mild prodrome with few skin lesions in previously vaccinated people and had a mortality rate well below 10%; *flat smallpox* (≈5% of hospitalized cases) produced slowly developing lesions that were difficult to ascertain because they appeared flush with the (edematous) skin at the vesicular stage, and it was almost always fatal; and *hemorrhagic smallpox* (<1% of cases) induced bleeding into the skin and the mucous membranes and was invariably fatal. A discrete type of the ordinary form, with a typical febrile prodrome and rash, resulted from alastrim variola minor infection. Individuals with this form of disease were not nearly as moribund or "toxemic" as individuals with variola major infection. Previous vaccination was not necessarily protective against the hemorrhagic forms of disease but seemed to be protective against flat forms of disease.

In flat smallpox, illness was heralded by the abrupt onset of fever with temperatures of 38.3° to 38.9° C and the appearance of the rash after 3 to 4 days. The oral enanthem was often confluent, and sloughing of rectal mucosal membranes was also reported. At the papulovesicular stage of disease, lesions appeared as small indentations (day 6) with hemorrhages in the bases and were surrounded by an erythematous ring. By day 7 or 8, the lesions appeared flat. Bullous lesions that would slough were reported. Fever persisted throughout the disease course, and respiratory complications were often observed by day 7 or 8 of illness. Thrombocytopenia, neutropenia, and lymphocytosis were reported.

In hemorrhagic forms of smallpox, illness began with fever and typical prodromal symptoms; the fever never abated. Early after the onset of fever, petechiae and purpuric rashes became apparent; subconjunctival hemorrhages, hematuria, and vaginal bleeding were also seen. Patients usually died by day 6 of illness, well before any classic vesiculopustular rash was evident. In late hemorrhagic disease, after the onset of fever, typical maculopapular lesions developed, but the fever did not abate. The lesion evolved slowly, and areas of hemorrhage were evident at the base of the lesions. Bleeding occurred in the mucous membranes, thrombocytopenia was profound, and death occurred between days 8 and 10 of illness.

Monkeypox

After an incubation period of 7 to 17 days (mean, 12 days), a prodrome of fever, headache, backache, and fatigue begins.[6] The cutaneous eruption evolves similar to that of smallpox. Lesions evolve in the same stage in any one part of the body from macules, papules, and vesicles to pustules, and then they crust and scar (Fig. 372-2). After resolution of the rash, hypopigmentation is followed by hyperpigmentation of the scarred lesions. Pronounced

TABLE 372-2 WORLD HEALTH ORGANIZATION SMALLPOX TYPES

WHO SMALLPOX TYPE	CLINICAL DEFINITION
Variola sine eruptione	Fever, no rash
Modified	Like ordinary, with an accelerated course
Ordinary discrete	Fever, rash; areas of normal skin between pustules, even on the face
Ordinary semiconfluent	Fever, rash; pustules confluent on the face, discrete elsewhere
Ordinary confluent	Fever, rash; pustules confluent on the face and forearms
Flat	Fever, erythema, and edema of the skin; vesicles soft, flat, and bullous
Hemorrhagic, early	Fever (persistent), hemorrhages and petechiae, purpuric rash at illness onset
Hemorrhagic, late	Fever (persistent), rash, hemorrhage into the base of vesicles late in illness

FIGURE 372-1. Pustular lesions of smallpox and beginning of scarring on the face and upper part of the torso. (From the Centers for Disease Control and Prevention Public Health Image Library, ID #: 7055. Photograph by Stan Foster.)

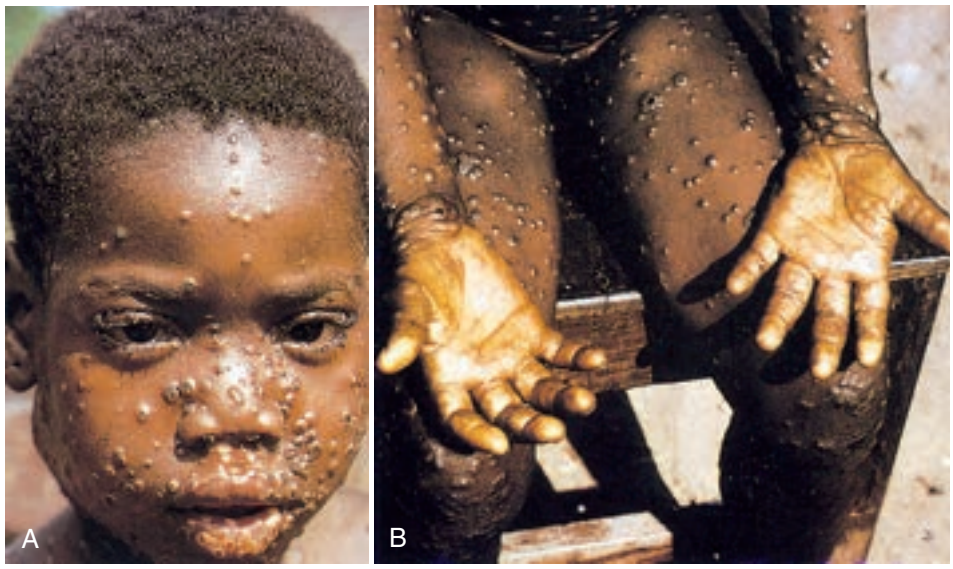

FIGURE 372-2. Rash of monkeypox of the head (A) and extremities (B) in a 7-year-old girl in central Zaire. (From Peters W, Pasvol G. *Tropical Medicine and Parasitology.* 5th ed. New York: Mosby; 2002:238.)

cervical, postauricular, submandibular, and inguinal lymphadenopathy clinically distinguishes monkeypox from smallpox.

Vaccinia

Multiple-puncture vaccinia virus infection by a bifurcated needle is the current smallpox vaccination regimen used for laboratory personnel working with orthopoxviruses, public health care personnel, and military in the United States. Most commonly, the infection progresses through a standard course of events from vesicle to pustule. However, of all vaccines used today, the smallpox vaccine, which is composed of live, replicative vaccinia virus, has one of the highest rates of adverse events. Major complications include progressive vaccinia, eczema vaccinatum, generalized vaccinia, postvaccinial encephalitis, accidental infection, and carditis.[7]

Progressive vaccinia, which is a rare and often fatal vaccine complication in persons with severe deficiencies in cellular immunity, occurs in about 1 per million vaccines, with a case-fatality rate of about 35%. Progressive vaccinia is characterized by frequently painless growth and spread of the vaccine virus beyond the inoculation site, often leading to necrosis and sometimes metastases to other body sites. The possibility of progressive vaccinia should be considered if the vaccination site lesion continues to progress and expand without apparent healing more than 15 days after vaccination.

Eczema vaccinatum[8] can occur in people with a history of atopic dermatitis (eczema), irrespective of its severity or activity, owing to local spread or dissemination from the primary vaccination site or contact with the unscabbed vaccination site of another person. A localized or generalized papular, vesicular, or pustular rash can develop anywhere on the body or be localized to previous eczematous lesions. Systemic illness with fever, malaise, and lymphadenopathy may occur. In the 1968 national U.S. surveillance of smallpox vaccination, there were 66 cases (no deaths) of eczema vaccinatum among 14.5 million vaccinees (4.6 cases/million) and 60 cases (1 death) among their several million contacts.

Generalized vaccinia describes the vesicular rash that develops after vaccination. Excluding dissemination associated with eczema vaccinatum and progressive vaccinia, it has been extremely rare to document virus in these lesions. True generalized vaccinia is believed to represent the end product of viremic spread of virus, and no predisposing factors have been identified. Generalized vaccinia was estimated to occur in about 242 of every million primary vaccinations. Studies done during the vaccination program in 2002 in the United States indicate that the majority of cases previously reported to be generalized vaccinia likely were generalized rashes caused by inflammatory or allergic responses to the vaccine, and not true generalized vaccinia. *Postvaccination encephalomyelitis* is a rare but serious complication that usually occurs only in primary vaccinees. Patients have variably displayed clinical and diagnostic features suggestive of a postimmunization demyelinating encephalomyelitis or direct viral invasion of the nervous system. This postvaccination reaction typically occurs 11 to 15 days after vaccination. Symptoms include fever, headache, vomiting, confusion, delirium, disorientation, restlessness, drowsiness or lethargy, seizures, and coma. Cerebrospinal fluid can demonstrate elevated pressure but generally has a normal cell count and chemistry profile. The diagnosis is one of exclusion, and no specific tests are available to confirm it. However, a few cases have been demonstrated to have anti-orthopoxvirus immunoglobulin M (IgM) or IgG responses in their cerebrospinal fluid.

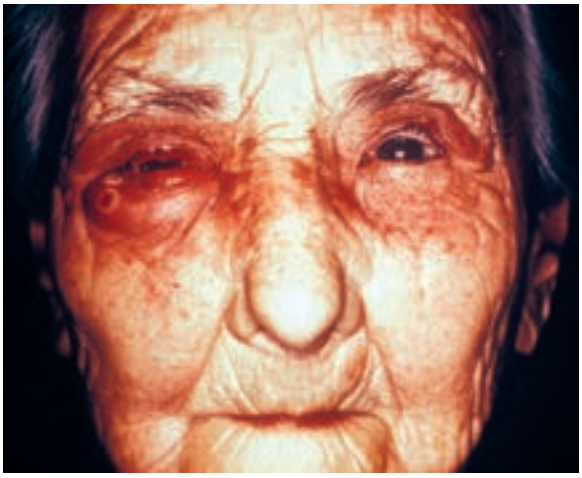

FIGURE 372-3. Vaccinia autoinoculation of an eye. (From the Centers for Disease Control and Prevention Public Health Image Library, ID #: 3322.)

Accidental infection occurs when virus from the vaccination site is transferred to another site or to another person through intimate skin contact. This complication generally occurs with primary vaccinees rather than revaccinees. Accidental self-inoculation, which most commonly occurs on the face, mouth, lips, or genitalia, is not usually serious and requires no specific treatment. Inoculation of the eye or eyelid is more serious and can be sight threatening if it is not evaluated and treated appropriately (Fig. 372-3). Between 1963 and 1968, ocular vaccinia was observed in 348 persons, including 22 who had evidence of corneal involvement and 11 who experienced permanent defects.

With buffalopox viruses in India and the Cantagalo virus in Brazil, up to 10 lesions have been described on the hands or arms of human handlers; fever, lymphadenopathy, backache, and fatigue are also associated symptoms. Transmission is believed to occur by unprotected contact with active lesions present on animal teats and udders. Interhuman transmission of buffalopox to family members has been reported to occur through contact.

Cowpox

Cowpox lesions are generally restricted to the hands and face; most patients (72%) have only one lesion. Multiple lesions may be caused by multiple primary inoculations, by autoinoculation, and very infrequently by lymphatic or viremic spread. The cowpox lesion passes through macular, papular, vesicular, and pustular stages before forming a hard black crust. The lesion is generally very painful, and erythema and edema are common at the late vesicular and pustular stages. Patients usually have lymphadenitis, fever, and general malaise. These features are generally severe in children, and absence from school or work is common. About 30% may be hospitalized. Most patients take 6 to 8 weeks to recover, but up to 12 weeks may be required. On occasion, a very severe infection and death may occur, typically in immunosuppressed individuals. Scarring is usually permanent.

Parapoxvirus

Lesions, such as from the orf virus, start as erythematous papules and progress in 1 to 2 weeks to target lesions with a red center surrounded by a white halo and an outer inflamed halo. Lesions then progress to a nodular and then a papillomatous stage, which often has a "weeping" surface. In some patients, lesions may enlarge and persist for weeks before resolving by a crusting stage, which also may persist for weeks. Very large granulomatous lesions may occasionally require surgical removal.

Most patients have only one lesion, but multiple primary lesions may develop. Systemic reaction is relatively uncommon, and the lesion is often not particularly painful. Lymphadenopathy is present in some patients, and lymphangitis is relatively uncommon. Erythema multiforme (Chapter 439) can develop in up to one third of patients.

Molluscum Contagiosum Virus

Molluscum infection occurs when molluscum contagiosum virus comes into contact with nonintact skin. The characteristic lesion begins as a small papule and, when mature, is a discrete, 2- to 5-mm-diameter, smooth, dome-shaped, pearly or flesh-colored nodule that is often umbilicated (see Fig. 440-9 in Chapter 440). A cheesy, off-white or yellowish material is easily expressed from lesions. Most patients have 1 to 20 lesions, but hundreds of lesions may occasionally be present. Because of multiple simultaneous infections or mechanical spread, the lesions may become confluent along the line of a scratch, and satellite lesions are sometimes seen.

In children, molluscum lesions occur mainly on the trunk and proximal ends of the extremities; in adults, lesions tend to occur on the trunk, pubic area, and thighs. In all cases, however, infection can be transmitted to other areas by autoinoculation. In men infected with human immunodeficiency virus (HIV), molluscum lesions can occur along the beard line and result in ocular involvement.

Individual lesions persist for about 2 months, but the disease usually persists for 6 to 9 months. Individuals with impaired cell-mediated immunity, including persons with HIV infection, tend to have more severe and prolonged infection.

Yatapoxvirus

Tanapox infection begins with a short febrile (38° to 39° C) illness that persists for 2 to 4 days and is sometimes accompanied by headache, backache, or prostration. The eruption of a lesion is frequently heralded by pruritus at the site of the outbreak. The lesion appears as a hyperpigmented macule, which often has central elevation and evolves to a papule with palpable induration. About 80% have a solitary nodule, but as many as 10 lesions can occur. Most lesions (72%) occur on the lower extremities, and the fewest occur on the face and areas normally covered by clothing.

Fever and systemic symptoms wane as the lesion erupts. The papule then becomes more "pocklike" but contains no fluid; umbilication or a pseudocrust may develop. Typically, the papule evolves into a firm, deep-seated, elevated nodule. At the end of the first week, the lesion is surrounded by erythema and indurated skin, and regional lymphangitis is common. Lesions then either ulcerate or become larger nodules (up to 2 cm in diameter) within about 2 weeks, after which the local inflammatory response wanes and the lesions began to granulate. Resolution of lesions occurs within 6 weeks. Infection appears to confer lifelong immunity.

DIAGNOSIS

Before its eradication, smallpox was relatively easy to recognize. Chickenpox (Chapter 375) produces a centripetally distributed rash and rarely appears on the palms and soles; in chickenpox, prodromal fever and systemic manifestations are mild, the lesions are superficial in nature, and lesions in different developmental stages may be present in the same area of the body. Other diseases and conditions that could be confused with vesicular-stage smallpox include monkeypox, generalized vaccinia, disseminated herpes zoster or herpes simplex virus infection (Chapters 374 and 375), drug reactions (eruptions), erythema multiforme (Chapter 439), enteroviral infections (Chapter 379), scabies (Chapter 359), insect bites, impetigo (Chapter 439), and molluscum contagiosum. Diseases confused with hemorrhagic smallpox included acute leukemia (Chapter 183), meningococcemia (Chapter 298), and idiopathic thrombocytopenic purpura (Chapter 172). The Centers for Disease Control and Prevention have developed a protocol for evaluation of patients for potential smallpox (available at http://emergency.cdc.gov/agent/smallpox/diagnosis/evalposter.asp) and an algorithm for laboratory assessment of the vesiculopustular stage of rash (available at http://emergency.cdc.gov/agent/smallpox/diagnosis/pdf/poxalgorithm1-5-12.pdf).

Orthopoxvirus

Viral culture or electron microscopic evaluation of virion particles from clinical rash specimens has been the approach to diagnosis of orthopoxvirus infections. Serologic tests are largely genus specific, detect IgG, and usually cannot differentiate among species, although IgM levels can help differentiate recent orthopoxvirus infection from remote vaccination. Nucleic acid–based tests and polymerase chain reaction (PCR) assays are other options. If the diagnosis of smallpox is being considered, viral culture requires a biosafety level 4 (BSL-4) containment facility that is sanctioned by the World Health Organization to use variola virus.

Parapoxvirus

The differential diagnosis of parapoxvirus lesions can include ecthyma gangrenosum as a result of a *Pseudomonas aeruginosa* infection (Chapter 306), vaccinia or cowpox infection, cutaneous anthrax (Chapter 294), erysipeloid (Chapter 295), tularemia (Chapter 311), and tumor. Farm workers recognize the infection and tend not to seek medical attention for routine cases, so about 45% of reported cases may have no known contact with infected animals, and the clinical diagnosis of such cases may be difficult. With negative-stain electron microscopy, virions with the characteristic morphology of parapoxviruses are usually seen easily in lesion extracts, thereby providing a rapid, certain diagnosis of the genus. The virus can be grown in cell culture, and PCR detection can be performed. Species-specific and species-generic protein-based diagnostics have also been developed for parapoxviruses.

Molluscum Contagiosum Virus

The clinical appearance of molluscum lesions usually is sufficiently characteristic to permit a clinical diagnosis. Brick-shaped virions typically can be seen in large numbers if the cheesy material expressed from the lesion is examined by electron microscopy. The virus has not been cultured in standard tissue culture systems. The characteristic histopathology of these lesions is diagnostic, but PCR methods can also identify molluscum contagiosum.

Yatapoxvirus

The limited geographic distribution of tanapox virus and the patient's travel history help with the diagnosis of tanapox infection. Unique clinical features that differentiate tanapox from other orthopoxvirus infections are the nodular nature of the rash lesion, the paucity of lesions, the benign disease course, and the protracted resolution of the rash. The solid nodular and ulcerated lesions are larger and develop more slowly than those of monkeypox, but they are smaller and develop more rapidly than those of tropical ulcers.

Tanapox virus can be detected by electron microscopy, but the appearance of the virions cannot exclude infection with other morphologically similar brick-shaped poxviruses. Nucleic acid testing or cell line culture can make the definitive diagnosis.

TREATMENT AND PREVENTION Rx

Orthopoxvirus

Vaccination with smallpox vaccine is the mainstay for prevention of orthopoxvirus infection and was the primary method used to eradicate smallpox. Stockpiles of vaccine are available should smallpox recur. In the United States, however, smallpox vaccine is currently recommended only for laboratory

personnel who work with infectious orthopoxvirus, certain public health personnel, and certain members of the military. In part because of stringent prescreening procedures, recent adverse events are rare with vaccination, but myocarditis or pericarditis was documented in 18 of 230,734 (8 per million) U.S. military primary vaccinees immunized in 2002 and 2003. In recent trials, both an attenuated tissue-cultured smallpox vaccine (LC16m8) and a replication-deficient smallpox vaccine (MVA) were about 95% effective both in newly vaccinated adults and as a booster.[A1,A2] Animal studies have demonstrated protective efficacy against a lethal monkeypox challenge.[9]

Currently, no antiviral drugs are licensed to treat orthopoxvirus or other poxvirus illnesses. Early administration of vaccinia immune globulin (VIG) may reduce the mortality of eczema vaccinatum from 30 to 40% to 7%, and VIG may also be useful for other complications (e.g., progressive vaccinia, severe generalized vaccinia, or contact infection) of vaccinia (smallpox) vaccine administration and for other orthopoxvirus infections. However, VIG alone has no clear benefit for treatment of smallpox infection itself.

Antiviral compounds (Chapter 360) with in vitro and in vivo activity against poxviruses include specifically 5-iodo-2′-deoxyuridine, adenine arabinoside, and trifluorothymidine. Because of their systemic toxicity, these compounds have been used topically for the treatment of orthopoxvirus ocular infections. In vitro, cidofovir is active against cowpox, vaccinia, monkeypox, and variola; in vivo, it protects challenged animals when it is given prophylactically or early in the evolution of disease. Cidofovir has known renal toxicity and is administered with hydration and probenecid. CMX-001/brincidofovir has been effective in treatment of systemic rabbitpox infection in rabbits and was used as part of a multidrug regimen to treat a case of progressive vaccinia.

Tecovirimat is an effective small-molecule antiviral in animal models of systemic orthopoxvirus infection[10] and has been used successfully as part of a multidrug regimen to treat a human case of eczema vaccinatum and a case of progressive vaccinia. Although not currently licensed, it is part of the U.S. Strategic National Stockpile and available for compassionate use to treat orthopoxvirus infections, including smallpox, as an investigational new drug sponsored by the Centers for Disease Control and Prevention.

Parapoxvirus

Most workers at risk for parapoxvirus become infected, and reinfection also occurs. The vaccine used to control orf in sheep is fully virulent and has caused human infection. Treatment options are limited; anecdotal reports have described the use of topical cidofovir, and other options may be topical formulations of interferon-modulating compounds such as imiquimod.

Molluscum Contagiosum Virus

Molluscum contagiosum infection is benign, and recovery is usually spontaneous, but treatment may be sought for cosmetic reasons, particularly for facial or multiple lesions. Options include cryotherapy, mechanical curettage, and chemical treatments such as podophyllin or podofilox, cantharidin, iodine, and tretinoin. Irritation has been a side effect of many of the chemical treatments. Topical application of a 3% cidofovir antiviral cream or suspension has been reported to be beneficial, as has the use of potentially immune-modulating cimetidine or topical imiquimod therapy. However, no therapy is documented to be beneficial by well-controlled randomized trials,[A3] although topical 10% potassium hydroxide solution, applied twice daily, is sometimes recommended. Covering of lesions and the use of proper hand hygiene after contact with lesions should prevent transmission in most situations. For individuals with AIDS and molluscum, highly active antiretroviral therapy with a resulting improvement in the $CD4^+$ cell count appears to be efficacious.[11]

PROGNOSIS

Monkeypox and smallpox both cause human illness, with mortality rates ranging from 10 to 40%; variola minor variants, however, have mortality rates of less than 1%. Yatapoxvirus infections are self-limited, and the illness resolves in the course of a few weeks. Parapoxvirus infections are manifested chiefly by localized symptoms, and the lesions resolve within a month or so in nonimmunocompetent hosts. Molluscum contagiosum infection is benign, usually with a spontaneous recovery, but the infection can persist for months. Variola minor has a mortality rate of less than 1%.

Grade A References

A1. Saito T, Fujii T, Kanatani Y, et al. Clinical and immunological response to attenuated tissue-cultured smallpox vaccine LC16m8. *JAMA*. 2009;301:1025-1033.

A2. von Krempelhuber A, Vollmar J, Pokorny R, et al. A randomized, double-blind, dose-finding Phase II study to evaluate immunogenicity and safety of the third generation smallpox vaccine candidate IMVAMUNE. *Vaccine*. 2010;28:1209-1216.

A3. van der Wouden JC, van der Sande R, van Suijlekom-Smit LW, et al. Interventions for cutaneous molluscum contagiosum. *Cochrane Database Syst Rev*. 2009;4:CD004767.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Reynolds MG, Carroll DS, Karem KL. Factors affecting the likelihood of monkeypox's emergence and spread in the post-smallpox era. *Curr Opin Virol*. 2012;2:335-343.
2. Haller SL, Peng C, McFadden G, et al. Poxviruses and the evolution of host range and virulence. *Infect Genet Evol*. 2014;21:15-40.
3. Roess AA, Galan A, Kitces E, et al. Novel deer-associated parapoxvirus infection in deer hunters. *N Engl J Med*. 2010;363:2621-2627.
4. Chen X, Anstey AV, Bugert JJ. Molluscum contagiosum virus infection. *Lancet Infect Dis*. 2013; 13:877-888.
5. Moore ZS, Seward JF, Lane JM. Smallpox. *Lancet*. 2006;367:425-435.
6. McCollum AM, Damon IK. Human monkeypox. *Clin Infect Dis*. 2014;58:260-267.
7. Cono J, Casey CG, Bell DM, Centers for Disease Control and Prevention. Smallpox vaccination and adverse reactions: guidance for clinicians. *MMWR Recomm Rep*. 2003;52:1-28.
8. Reed JL, Scott DE, Bray M. Eczema vaccinatum. *Clin Infect Dis*. 2012;54:832-840.
9. Hatch GJ, Graham VA, Bewley KR, et al. Assessment of the protective effect of Imvamune and Acam2000 vaccines against aerosolized monkeypox virus in cynomolgus macaques. *J Virol*. 2013;87:7805-7815.
10. Mucker EM, Goff AJ, Shamblin JD, et al. Efficacy of tecovirimat (ST-246) in nonhuman primates infected with variola virus (Smallpox). *Antimicrob Agents Chemother*. 2013;57:6246-6253.
11. Nguyen HP, Franz E, Stiegel KR, et al. Treatment of molluscum contagiosum in adult, pediatric, and immunodeficient populations. *J Cutan Med Surg*. 2014;18:1-8.

REVIEW QUESTIONS

1. A 24-year-old man presents to your office. He has a 2-cm nodular lesion on his right forearm. He does not appear systemically ill. He reports he was recently treated for malaria; he had a temperature of 103° F a week prior. He recently returned from work in Africa at a primate preserve in the Republic of Congo. Biopsy and electron micrographs of the lesion demonstrate poxvirus-particles. What is your diagnosis?

A. Tanapox virus (*Yatapoxvirus* species)
B. Monkeypox (*Orthopoxvirus* species)
C. Myiasis
D. Orf virus (*Parapoxvirus* species)
E. Smallpox (*Orthopoxvirus variola*)

Answer: A The electron micrograph demonstrates poxvirus particles. These are of a characteristic size and shape that renders them distinguishable from herpesvirus particles and bacterial and parasitic particles. Tanapox virus infections typically present with high fever followed by the development of one or a few nodular lesions. The lesions are distinct from the classic vesiculopustular lesions of orthopoxvirus infections (such as cowpox, vaccinia, monkeypox, or smallpox). Monkeypox virus typically also causes lymphadenopathy. Myiasis is caused by the bot fly and could be found in a returning African traveller, but the electron micrograph would not be consistent with it. Pain would also be associated with the lesion. Orf lesions typically are not associated with a high-antecedent fever and are usually associated with exposure to a ruminant (commonly sheep or goat). Smallpox disease is typically a fulminant disease, with a more generalized vesiculopustular rash, especially in someone unlikely to have been previously vaccinated.

2. A 35-year-old woman is seen in the emergency department with an extensive pustular rash that covers more than 30% of her body. Lesions are apparent on the palms of her hands and soles of her feet. The distribution is centrifugal. The lesions are round and feel like buckshot. She is having difficulty swallowing because of pain. She has track marks on her arms. On further questioning, she reports "having the flu" 7 days ago, despite having received the flu vaccine 2 months prior. She reveals a history of illicit intravenous drug use and claims to use clean needles from the laboratory. She works as a clinical microbiologist in a virology laboratory that has an archival collection of materials. She has recently been characterizing various unknown historic clinical specimens to update the inventory. What would be your next step?

A. Institute airborne precautions and call the health department.
B. Take pictures of the lesions, get blood cultures, sample specimens and send for bacterial analysis, and start treatment for endocarditis.
C. Perform a rapid plasma reagin test.
D. Ask about her vaccination status.
E. Get an infectious disease consult.

Answer: A The signs and symptoms are consistent with smallpox. Institution of airborne precautions will provide infection control and limit potential exposure to this patient. Notification of the health department will enable notification of the Centers for Disease Control and Prevention's Emergency Operations Center. A complete assessment of the patient can then ensue, using appropriate personnel protective equipment, ideally by a person previously vaccinated against smallpox.

3. A 25-year-old man reports to your office with a single vesiculopustular lesion on his forehead. He works on a farm with sheep and goats, none of which are ill. His wife is a laboratory postdoctoral fellow who works in a laboratory that studies monkeypox virus. You find out that she was recently vaccinated, 3 weeks ago, with smallpox vaccine before beginning work in the laboratory. What is the most likely cause of the forehead lesion?

A. Vaccinia virus
B. Orf virus
C. Monkeypox virus
D. Smallpox
E. Pseudo cowpox

Answer: A Vaccinia virus is the live virus in smallpox vaccine. Although it is possible he could have a parapoxvirus infection acquired from the sheep or goats on the farm, none of the animals appears ill. Smallpox vaccine does not contain smallpox virus, which is the variola virus. The likely source of the vesiculopustular lesion is accidental implantation of vaccinia virus during close contact with his wife.

4. An epidemiology graduate student comes to your emergency department with a 5-day history of rash. The rash initially manifest as small red spots, which then became vesicular. She also notes painful adenopathy in her cervical and axillary lymph nodes. She describes a fever before the onset of the rash. She returned from Africa 1 week ago. On the plane, she had a fever. She had been working in the Democratic Republic of Congo (DRC) on maternal-fetal health issues of indigenous people in the equatorial region of the DRC. She has been living within the community and helping with all household and village activities. You suspect she may have monkeypox. What antiviral agent might you consider using if the diagnosis is confirmed?

A. Brincidofovir or tecovirimat
B. Acyclovir
C. Ganciclovir
D. Adefovir
E. Stavudine

Answer: A Both brincidofovir (CMX-001), the orally bioavailable derivative of cidofovir, and tecovirimat (ST-246 or Arestyvyr) have activity against orthopoxvirus infections.

5. In the above case, what samples would you take for diagnosis?

A. Rash lesions and sera
B. Blood culture
C. Serum chemistries
D. Complete blood count
E. Thick and thin smears

Answer: A Poxvirus infections are usually identified by evaluation of material from the rash lesion itself. Polymerase chain reaction and culture can identify and characterize the virus. Electron microscopy can also help to exclude a herpesvirus infection. Serology can be useful if the rash lesions have resolved.

373

PAPILLOMAVIRUS

JOHN M. DOUGLAS, Jr.

DEFINITION

Human papillomaviruses (HPVs) are a group of small DNA viruses that cause a variety of benign and malignant lesions of the skin and mucous membranes. The most commonly recognized HPV-associated diseases include warts (Chapter 440) at anogenital sites (condyloma acuminatum), other skin surfaces (common warts or verruca vulgaris), and the plantar surface of the foot (verruca plantaris). In addition, HPV infection causes squamous intraepithelial lesions of the cervix, also known as cervical intraepithelial neoplasia (CIN), and of other anogenital sites. It is considered the etiologic agent of a variety of cancers, especially cervical cancer.

The Pathogen

HPV is a member of the family Papillomaviridae. Like all papillomaviruses, HPV is nonenveloped, measures 55 nm in diameter, and has a double-stranded circular DNA genome of approximately 7900 base pairs enclosed by an icosahedral capsid. The HPV genome contains three functional regions: early genes (six total—E1, E2, E4, E5, E6, E7), which are expressed soon after infection and control replication, transcription, and cellular proliferation; late genes (two total—L1, L2), which are expressed in later stages of infection and encode the structural capsid proteins; and the long control region, which contains regulatory sequences that control the replication and transcription of early and late genes. Papillomaviruses complete their life cycle only in terminally differentiated epithelial cells and thus are difficult to grow in cell culture. Papillomavirus taxonomy is based on a genotyping system involving the use of DNA sequence relatedness of the gene encoding L1, the major capsid protein, with different types defined as having less than 90% homology.[1]

Papillomaviruses are classified taxonomically by genus (Greek letters) and species (numbered), each containing one or more types. Most HPV types are included in three large genera: alpha (primarily mucosal or genital types), beta, and gamma (both of which cause cutaneous lesions). Currently, more than 150 types of HPV have been identified, over 40 of which infect genital skin and mucosa. Of the genital types, approximately 15 are considered high risk because they are associated with high-grade squamous intraepithelial lesions and cancers of the cervix, anus, penis, vulva, vagina, and oropharynx, whereas others are considered low risk because they are largely associated with genital warts and low-grade squamous intraepithelial lesions.

EPIDEMIOLOGY

HPV infections are primarily transmitted by direct contact of skin or mucous membranes with an infected lesion. Genital HPV infection is typically contracted through sexual intercourse, although nonpenetrative genital contact, oral-genital contact, and manual-genital contact are also possible routes of transmission. In addition, genital HPV infection can be transmitted to the mouth and upper respiratory tract perinatally from infected mothers to newborns. For nongenital HPV infection, personal skin-to-skin contact also plays a primary role, although for plantar warts, fomite transmission from moist surfaces is likely to be an important source of infection. Both genital and nongenital infection can be transmitted to new sites by autoinoculation.

Regarding genital HPV, in the United States, an estimated 80 million or so persons are infected and about 14 million new infections occur annually, thereby making genital HPV the most common sexually transmitted infection.[2] The prevalence of anogenital warts is estimated to be approximately 1% in the sexually active adult population. Acquisition of infection begins shortly after sexual debut, with an estimated 40 to 60% incidence of at least one type within 2 years of initiation of sex. Risk factors for infection include variables

related to probable exposure (e.g., younger age at onset of sexual activity, increased number of recent and lifetime partners, and number of partners of the sex partners), susceptibility (e.g., lack of circumcision for men), and absence of prevention factors (e.g., lack of consistent condom use or immunization). Most infections are asymptomatic and clear without treatment; only 10% are estimated to persist longer than 2 years.[3] The incidence of genital HPV infection and genital warts appears to be declining in some countries that have initiated immunization programs.[4]

Oral HPV infection is usually asymptomatic. Its prevalence in U.S. adults is approximately 5 to 7%, almost half of which are high-risk types.[5] Newly acquired oral oncogenic HPV infections are rare in healthy men, and most are cleared within 1 year.[6] Cutaneous HPV infection is most typically recognized as common and plantar warts, especially in children, in whom annual incidence rates of up to 30% have been reported.

All types and manifestations of HPV infection are more common in persons with impaired cell-mediated immunity, such as those infected with human immunodeficiency virus (HIV) or receiving immunosuppressive therapy. Among HIV-infected men who have sex with men, the prevalence of anal high-grade squamous intraepithelial lesions is almost 30%, and the anal cancer incidence is 46 per 100,000.[7] Of note, genital HPV infection may increase susceptibility to HIV infection, analogous to other sexually transmitted infections, raising the possibility that HPV vaccines could help prevent HIV infection.

Cervical cancer has declined in developed countries since the initiation of cytologic screening programs, although an estimated 12,000 cases and 4000 deaths still occur in the United States annually.[8] However, the disease is a major problem in the developing world, where screening is limited, and it is the third most common cancer in women worldwide, with an estimated 500,000 cases annually. Considering all anatomic sites, it is estimated that more than 600,000 HPV-associated cancers (5% of all cancers) occur globally each year.[9] In the United States, HPV causes about 18,000 cancers annually among females and about 8000 annually among males. HPV-associated oral cancers have been rising in the United States and are projected to become the most common HPV-associated cancer by 2020.[10]

PATHOBIOLOGY

HPV infections cause disease by producing aberrant cell growth. In the case of cutaneous and low-risk types, lesions such as warts result from HPV-induced benign proliferation of epidermal layers. For high-risk genital types, precancerous and cancerous lesions result from replacement of the epithelium by undifferentiated cells as a result of HPV-induced interference with normal cellular growth.

Infection begins in the lowest and least well-differentiated layer of the epithelium, the basal cells, where exposure is facilitated by microtrauma. Transcription and protein expression are highly coordinated with the level of cellular differentiation. In the basal layer, the viral genome becomes established in the nucleus as an episome that replicates in tandem with cellular replication, thus maintaining a stable copy number of viral genomes. As basal cells migrate up and differentiate in the superficial layers of the epithelium, full vegetative viral DNA replication and expression of structural proteins occur, with assembly of infectious virions in the most superficial layer of the epithelium, where they are released with the sloughing of dead cells during normal cellular turnover.

Persistent infection with various high-risk types of genital HPV is firmly established as the cause of squamous cell carcinoma and adenocarcinoma of the cervix, and HPV 16 in particular plays a causal role in other anogenital and oropharyngeal squamous cell cancers. There are also associations of beta-HPV types with squamous cell cancer of the skin. HPV DNA can be detected in more than 99% of cervical cancer cases, with 70% of cancers caused by the two most common high-risk types, HPV 16 and 18. The pathogenesis of HPV-induced cancer involves viral integration into the host genome with resulting disruption of the E2 transcription regulatory gene and increased expression of E6 and E7 proteins. These proteins have oncogenic activity and affect cell growth by binding with tumor suppressor proteins, E6 with p53 and E7 with the retinoblastoma tumor suppressor protein, thereby disrupting apoptosis and cell cycle regulation.

Although persistent infection with high-risk types is "necessary" for the development of cervical cancer, it is not considered "sufficient" because cancer does not develop in most infected women. Possible cofactors include cigarette smoking, prolonged hormonal contraceptive use, multiparity, micronutrient deficiency, immunodeficiency (e.g., HIV infection), and other infections, (i.e., *Chlamydia trachomatis* and herpes simplex virus type 2). In addition, data supporting a familial risk for cervical cancer point to possible genetic factors, including genes controlling the immune response (e.g., *HLA, TNF*) and cell cycle (e.g., *p53*).

Cervical cancers (Chapter 199) most commonly arise in the cervical transformation zone, the border between the squamous epithelium of the ectocervix and columnar epithelium of the endocervix. Analagously, anal cancer (Chapter 145) occurs primarily at the anatomically similar anocolumnar transition zone.

The immune response to HPV infection is less robust than for most viral infections. Viral proteins and infectious virions develop in superficial cells with limited contact with the immune system, and there is no cell lysis or viremia to trigger an inflammatory response. In addition, HPV suppresses several components of the immune response, including the interferon pathway and the expression of inflammatory cytokines and major histocompatibility complex class I. Antibody to HPV develops in only an estimated 60% of infected individuals, often as long as 6 to 12 months after infection. In contrast, the dynamics of the immune response are quite different after immunization, with almost 100% seroconversion within several months and antibody levels many-fold higher than those after natural infection. The high efficacy of HPV vaccines, which are believed to produce primarily humoral immunity, supports the importance of the antibody response in protection from infection. In contrast, once infection occurs, cellular immunity appears to be critical for clearance of infection.

The histopathologic changes of warts include epithelial papillomatosis and acanthosis, with hyperkeratosis, parakeratosis, and hyperplasia of the parabasal cells. A characteristic feature is the presence of koilocytes, which are large atypical keratinocytes with irregular, hyperchromatic nuclei surrounded by a perinuclear halo, in the upper epidermis. Squamous intraepithelial lesions are characterized by hyperkeratosis, parakeratosis, koilocytosis, and epidermal hyperplasia, with increased mitotic figures in the upper half of the epidermis. Several classification systems have been used to classify these lesions and their risk for progression based on the proportion of the epithelium replaced by undifferentiated cells. The CIN system grades lesions as CIN 1 (with undifferentiated cells occupying the lower third), CIN 2 (with undifferentiated cells in the lower third to two thirds), and CIN 3/carcinoma in situ (CIS, with undifferentiated cells across the full thickness of the epithelium). Alternatively, the Bethesda system, originally developed for use with cytology but increasingly used for histologic classification, includes only two categories: low-grade squamous intraepithelial lesions (equivalent to CIN 1) and high-grade squamous intraepithelial lesions (equivalent to CIN 2 and CIN 3/CIS).

CLINICAL MANIFESTATIONS

The clinical manifestations of HPV infection vary by anatomic site and viral type. Common warts are exophytic, hyperkeratotic papules that typically occur on the hands but can appear on any skin surface, including the genital skin; they are most commonly caused by HPV types 1, 2, 4, 27, and 57. Plantar warts, which are caused by similar types of HPV, are hyperkeratotic, endophytic, and often very painful. In contrast, flat warts (verruca plana), which are small flat-topped papules that occur more commonly on the face, hands, and legs, are caused by a different group of nongenital HPV types (e.g., types 3, 10, 28, 38, 42, 49, 75, and 76).

Epidermodysplasia verruciformis is an uncommon autosomal recessive disease that is usually manifested in childhood as diffuse warts that respond poorly to treatment and typically are caused by beta-HPV types, most commonly type 5, that are uniquely associated with it. This disease, which is thought to be due to a selective defect in cell-mediated immunity because other opportunistic infections do not occur, is associated with two types of lesions: flat warts caused by the same HPV types as in normal hosts, and scaly tinea versicolor–type lesions caused by epidermodysplasia verruciformis–associated types. The latter are associated with the development of squamous cell cancer in sun-exposed areas in 30 to 70% of persons. Similar skin lesions and, rarely, associated skin cancers can develop in other patients with acquired defects in cell-mediated immunity.

The most common clinical manifestations of oral HPV infection include oral squamous cell papillomas and condyloma acuminatum, which are caused by HPV types 6 and 11. Less frequent are common skin warts and focal epithelial hyperplasia, which are round, flat papules caused by HPV types 13 and 32. Warts caused by genital HPV types can also rarely occur in the upper respiratory tract, where they can cause a serious condition known as recurrent respiratory papillomatosis, which can result in hoarseness and even airway compromise. Oral leukoplakia (Chapters 190 and 425), which

presents as a white patch or plaque and is considered to have premalignant potential, has been associated with HPV 16 and 18 infection.

Anogenital warts are papillomatous growths that occur throughout the anogenital skin and mucosa, typically at sites of genital friction. Most such warts are caused by HPV types 6 and 11, with approximately half of infected persons developing warts. Perianal warts are most common in persons with a history of anal intercourse and are often associated with intra-anal warts, but they also can occur without such contact, presumably through autoinoculation. Anogenital warts can range from flat or papular lesions to the classic pedunculated, cauliflower-shaped condyloma acuminatum. Warts are typically asymptomatic, noticed either by the patient as a "bump" or inadvertently during a genital examination, although they can cause itching, burning, pain, or, rarely, bleeding or mechanical obstruction of the birth canal in pregnant women.

Squamous intraepithelial lesions are most commonly found on the cervix as a result of screening for cervical cancer precursors by cytology (Papanicolaou [Pap] test) or HPV molecular testing, with confirmation by colposcopy and biopsy. Like warts, they also occur at other anogenital sites; and like cervical lesions, they are categorized either as low-grade or high-grade squamous intraepithelial lesions or as various stages of intraepithelial neoplasia (e.g., vulva—VIN; vagina—VaIN; anus—AIN; penis—PIN). Low-grade squamous intraepithelial lesions can be caused by either low- or high-risk types, whereas high-grade squamous intraepithelial lesions are primarily due to high-risk types. Most squamous intraepithelial lesions are not visible on mucosal surfaces without the application of 3 to 5% acetic acid and magnification, although they can appear as flat hyperpigmented papules known as bowenoid papulosis on the external genitalia.

DIAGNOSIS

Both cutaneous (Fig. 373-1) and genital (Fig. 373-2) warts generally present an easily recognized clinical picture and can be diagnosed by history and physical examination without laboratory testing. The application of 3 to 5% acetic acid causes whitening of HPV lesions on genital mucosa ("acetowhitening") and may be useful with magnification in women; however, because sensitivity and specificity of this application have not been defined, it is not recommended for routine use. Oral lesions are also generally recognized by physical examination.

The differential diagnosis of cutaneous warts includes seborrheic and solar keratoses, nevi, irritated acrochordons, clavi, and squamous cell carcinoma (Chapters 203 and 440); lichen planus (Chapter 438) can mimic flat warts and calluses of the foot or plantar warts. Genital warts must also be distinguished from the condyloma latum lesions of secondary syphilis (Chapter 319) and molluscum contagiosum (Chapters 372 and 440). Biopsy for histopathologic examination may be helpful for lesions at all anatomic sites that are atypical or not responsive to therapy, those suggestive of high-grade squamous intraepithelial lesions or cancer (e.g., pigmented, indurated, fixed, bleeding, or ulcerated), or lesions in immunocompromised patients.

Cervical squamous intraepithelial lesions have traditionally been detected by cervical cytology through Pap tests, with assessment of abnormal results by colposcopy and biopsy for histopathologic examination. However, molecular testing for high-risk HPV is beginning to be used both for screening and for management of low-grade cytologic abnormalities.[11,12] Several large comparative trials have shown HPV testing to be more sensitive than Pap tests for detection of CIN 2/3 and more effective in prevention of cervical cancer.[■■] HPV testing in combination with cytology (co-testing) has been approved to enhance the sensitivity of screening in women older than 30 years. HPV tests are also recommended for triage of women whose Pap test results show atypical squamous cells of undetermined significance, which is an equivocal test result. Although HPV testing has been recently approved by the U.S. Food and Drug Administration as a single test for primary cervical cancer screening, at present, its use as a single test in the absence of cytology is not recommended. Furthermore, HPV screening is not recommended for detecting possible infection at other sites (e.g., anus, oropharynx).

TREATMENT

Rx

Management of HPV infection is directed toward diagnosis and treatment of the lesions themselves because there are no virus-specific therapies. Many lesions resolve spontaneously, so the goal of treatment is amelioration or

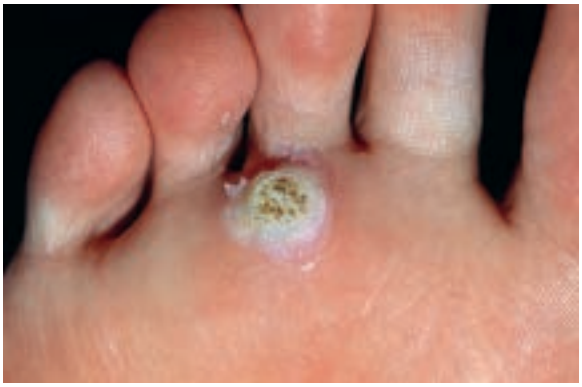

FIGURE 373-1. Plantar wart. A hyperkeratotic, verrucous papule or plaque beneath a pressure point on the sole of the foot is characteristic. Human papillomavirus types 1 (myrmecia), 2 (mosaic), and 4 are most common. Because plantar warts are driven into the skin by the pressure of walking or standing, they are usually the most treatment resistant.

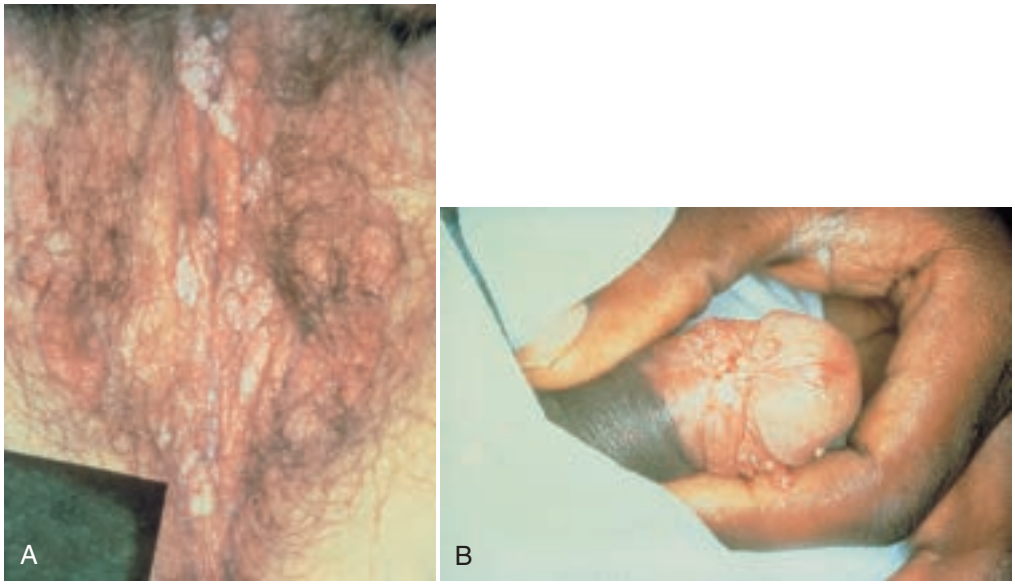

FIGURE 373-2. Genital human papillomavirus (HPV) infection. A, Vulvovaginal HPV infection. **B,** Penile HPV infection. (From Vermund SH, Bhatta MP. Papillomavirus infections. In: Cohen J, Powderly WG, eds. *Infectious Diseases,* 2nd ed. St Louis: Mosby; 2004.)

TABLE 373-1 RECOMMENDED TREATMENT OF GENITAL WARTS

PATIENT APPLIED
Podophyllotoxin (podofilox) 0.5% solution or gel; to be applied in up to 4 weekly cycles (twice daily for 3 days, followed by 4 days without treatment)
Imiquimod 3.75% cream (to be applied once daily at bedtime) or 5% cream (to be applied once daily at bedtime 3 times per week) for 6 to 10 hours for up to 16 weeks
Sinecatechins 15% ointment; to be applied 3 times daily for up to 16 weeks
PROVIDER ADMINISTERED
Cryotherapy with liquid nitrogen or cryoprobe; to be applied once every 1 to 2 weeks*
Trichloroacetic or bichloroacetic acid 80 to 90% solution; to be applied once weekly*
Office surgery* (excision, electrocautery, curettage)

*Safe for use in pregnancy.
Modified from Workowski KA; Centers for Disease Control and Prevention. Sexually transmitted disease treatment guidelines, 2014. *MMWR Morb Mortal Wkly Rep.* 2014 (in press).

TABLE 373-2 U.S. ADVISORY COMMITTEE ON IMMUNIZATION PRACTICES RECOMMENDATIONS FOR QUADRIVALENT (HPV 6, 11, 16, 18) AND BIVALENT (HPV 16, 18) HPV VACCINES

FEMALES

- Routine vaccination with three-dose series of quadrivalent or bivalent vaccine recommended in females aged 11 to 12 years, starting as young as 9 years
- Catch-up vaccination recommended in females aged 13 to 26 years who have not previously received full vaccine series

MALES

- Routine vaccination with three-dose series of quadrivalent vaccine recommended in males aged 11 to 12 years, starting as young as 9 years
- Catch-up vaccination recommended in immunocompromised persons and in males aged 13 to 21 years who have not previously received full vaccine series
- Catch-up vaccination recommended in men who have sex with men through age 26 years who have not previously received full vaccine series

Recommended dosage and schedule: 0.5 mL intramuscularly at 0, 2, and 6 months
Schedule modifications: if doses are missed, series does not need to be restarted, but second and third doses should be given as soon as possible.
Cervical cancer screening: no change in recommended interval
Special situations: females with genital warts, abnormal Papanicolaou test results, or positive human papillomavirus test results are unlikely to be infected with all four vaccine types and should be immunized per other recommendations; males with genital warts should also be immunized per other recommendations.
Pregnant and lactating women: not recommended for use in pregnancy based on lack of data; may be used in lactating women
Immunocompromised persons: no safety concerns because vaccine is noninfectious, but immune response and effectiveness might be reduced
Not recommended for females or males <9 and >26 years of age

Modified from Markowitz LE, Dunne EF, Saraiya M, et al; Centers for Disease Control and Prevention. Human papillomavirus vaccination. Recommendations of the Advisory Committee on Immunization Practices. *MMWR.* 2014;63:1-30.

prevention of symptoms or, in the case of high-grade squamous intraepithelial lesions, prevention of progression to cancer. Treatment involves destruction of lesions by physical techniques or topically applied or injected cytotoxic agents. Because treatment does not eradicate infection in surrounding tissues, recurrent lesions are common.

Treatment of warts depends on their location and size, the patient's preferences, and the provider's experience. Recommended first-line treatment of common, flat, and plantar warts includes the application of topical salicylic acid and cryotherapy; second-line treatment of recalcitrant lesions includes topical imiquimod (5% cream applied once daily at bedtime three times a week for up to 16 weeks), intralesional bleomycin, pulsed dye laser therapy, and surgical excision. Recommended treatment of genital warts (Table 373-1) includes patient-applied podophyllotoxin (0.5 or 0.15% solution or gel applied twice a day for 3 days, repeated weekly for four cycles), imiquimod, or sinecatechins, as well as provider-administered cryotherapy, trichloroacetic acid, or surgical excision; alternative treatments include intralesional interferon and laser surgery.[13] Oral lesions can be treated by locally destructive physical techniques.

For cervical lesions, treatment depends on histologic staging after colposcopy and biopsy and clinical context. Because CIN 1 usually regresses spontaneously, follow-up (by cytology, HPV test, and/or colposcopy) without treatment is recommended.[14] Treatment is generally recommended for all CIN 2/3 lesions, except in pregnant women, who have higher rates of spontaneous regression and a greater risk for reproductive tract complications after treatment, and in young woman with CIN 2. Treatment options include a variety of ablative and excisional techniques, such as cryosurgery, loop electrosurgical excision procedure, and laser surgery. Treatment is 90 to 95% effective in preventing the recurrence of lesions, and comparative clinical trials have shown similar efficacy for different treatment modalities.

PREVENTION

Primary Prevention

Primary prevention of HPV infection depends on avoidance of contact with infectious lesions and reduction of susceptibility through immunization. For example, the use of footwear in locker rooms may prevent plantar warts. For genital HPV infection, correct and consistent condom use can reduce the risk for both HPV infection and the HPV-associated diseases of genital warts, cervical squamous intraepithelial lesions, and cervical cancer. Male circumcision reduces the prevalence of both high-risk and low-risk types of genital HPV infection in men and transmission of HPV to their sex partners.

Of greatest importance for prevention are HPV vaccines (Chapter 18), which are composed of virus-like particles assembled from the major capsid protein, L1. Recommendations have been made for two currently licensed vaccines, a quadrivalent vaccine (types 6, 11, 16, and 18) and a bivalent vaccine (types 16 and 18). Recommendations are pending for a newly-licensed 9-valent vaccine (types 6, 11, 16, 18, 31, 33, 45, 52, and 58) which could prevent a higher proportion of HPV-associated disease outcomes.[15,16] The quadrivalent vaccine, given in three doses during a period of 6 months, is highly effective. In women, it prevent CIN 1 and CIN 2/3 as well as genital warts, VaIN, and VIN caused by any of the four types. In men, it prevent genital warts and AIN. The bivalent vaccine provides a high level of protection against CIN but not genital wart. Both vaccines are recommended for routine immunization of 11- to 12-year-old girls as well as females 13 t 26 years of age (Table 373-2). However, these vaccines provide protection against only two of the types of HPV that are associated with cancer, so immunized women continue to need cervical cancer screening.

The quadrivalent vaccine is also recommended for routine immunization of 11- to 12-year-old boys as well as males aged 13 to 21 years of age for prevention of genital warts (Chapter 18). Among men who have sex with men, immunization is recommended up to 26 years of age because of the additional benefit of preventing AIN (see Table 373-2). Both girls and boys 9 to 10 years of age may also be vaccinated at the discretion of the provider.

These vaccines have no therapeutic benefit against existing infection or lesions, and they are most effective if given before initiation of sexual activity. Because the HPV vaccines contain no live virus, they can be safely given to persons with impaired immunity (e.g., HIV infection). Research priorities include determining the duration of immunity after immunization, the efficacy of a two-dose regimen,[17] the extent of cross-protection against nonvaccine types, and benefit in prevention of nongenital HPV–associated cancers.

Secondary Prevention

Screening plus treatment of high-grade cervical squamous intraepithelial lesions is one of the most successful of all cancer prevention strategies. Screening has been conventionally conducted using Pap tests, but HPV tests are increasingly used as well for both primary screening and management of low-grade cytologic abnormalities. Current guidelines in the United States have evolved and now recommend screening at 3-year intervals beginning at age 21 years.[18] HPV testing is not recommended in younger women because of the high prevalence of often recently acquired and likely transient HPV infection. In contrast, among women older than 30 years, co-testing with HPV plus cytology is a recommended alternative to cytology alone because its enhanced sensitivity allows extension of screening to 5-year intervals (Chapter 199). Women with negative cytology who test HPV positive can be followed with repeat co-testing in 1 year or triaged by use of an HPV

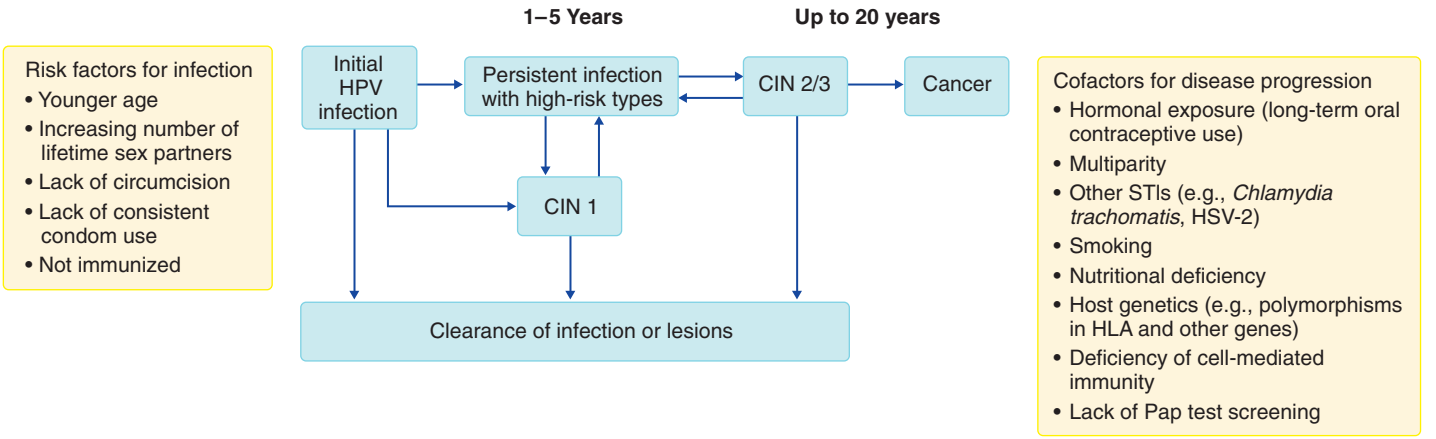

FIGURE 373-3. Natural history of genital human papillomavirus (HPV) infection and cervical cancer. CIN = cervical intraepithelial neoplasia; HSV = herpes simplex virus; STIs = sexually transmitted infections.

16/18 test. Management of other abnormalities is based on the degree of cytologic finding and age, with a higher threshold for treating younger women in whom abnormalities are more likely to resolve. With improvement in screening tests, a major prevention challenge remains increasing access to and use of screening services because approximately half of cervical cancers in the United States occur in women who were infrequently or never screened.

The value of cytologic screening for anal intraepithelial lesions in HIV-positive men who have sex with men is controversial but not currently recommended because of limited data on the natural history of these precursor lesions, the reliability of screening methods, and the safety and effectiveness of treatment.[13]

PROGNOSIS

Although the natural history of HPV infection is not fully characterized, the large majority of infections and premalignant lesions are self-limited in most immunocompetent patients. Whether infections no longer detectable have been cleared by the immune system or remain latent in the basal layer of the epithelium with the potential for reactivation is not clear, but the higher prevalence of detectable infection in advanced than in early HIV infection supports the possibility of long-term infection. Many if not most clinical lesions resolve spontaneously after the patient develops cell-mediated immunity. Spontaneous regression is estimated to occur in 25% of genital warts and more than 50% of common warts in children.

The natural history of CIN has been most intensively studied because of its relationship to cervical cancer, although many questions remain unanswered. Estimates of the likelihood of regression versus the risk for progression to invasive cancer are 90% and 1% for CIN 1, 40% and 5% for CIN 2, and 32% and 30% for CIN 3. The various stages of precursor lesions were traditionally viewed as a biologic continuum, with CIN 1 progressing through higher grades to cancer. However, newer data indicate that low-grade and high-grade squamous intraepithelial lesions may be distinct processes, in which low-grade lesions (CIN 1) represent a usually transient infection characterized by production of capsid protein (and probably infectious virions) and only minor cellular abnormalities, whereas high-grade lesions (CIN 2/3) represent proliferation of immature cells as a result of the activity of oncogenic proteins of high-risk types (Fig. 373-3). Initial infection frequently leads to a transient low-grade squamous intraepithelial lesion, with persistent infection in less than 10% of cases. Persistent infection can, in turn, lead directly to high-grade squamous intraepithelial lesions within several years of initial infection and can progress to invasive cancer after several decades. The natural history of squamous intraepithelial lesions at other anogenital sites is less well defined, but they may be associated with higher rates of spontaneous regression.

The majority of patients with HPV infection have an excellent prognosis, with the most serious outcomes of cancer occurring infrequently among the large number of persons infected. Treatment can hasten the resolution of cutaneous and genital warts and is highly effective for cervical lesions. Among women with CIN 3, long-term studies indicate that a 30% risk for cancer in untreated women can be reduced to less than 1% with treatment.

Grade A References

A1. Ronco G, Dillner J, Elfstrom KM, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. *Lancet*. 2014;383:524-532.

A2. Bouchard-Fortier G, Hajifathalian K, McKnight MD, et al. Co-testing for detection of high-grade cervical intraepithelial neoplasia and cancer compared with cytology alone: a meta-analysis of randomized controlled trials. *J Public Health (Oxf)*. 2014;36:46-55.

A3. Kwok CS, Gibbs S, Bennett C, et al. Topical treatments for cutaneous warts. *Cochrane Database Syst Rev*. 2012;9:CD001781.

A4. Martin-Hirsch PP, Paraskevaidis E, Bryant A, et al. Surgery for cervical intraepithelial neoplasia. *Cochrane Database Syst Rev*. 2013;12:CD001318.

A5. Tobian AAR, Serwadda D, Quinn TC, et al. Male circumcision for the prevention of HSV-2 and HPV infections and syphilis. *N Engl J Med*. 2009;360:1298-1309.

A6. Wawer MJ, Tobian AA, Kigozi G, et al. Effect of circumcision of HIV-negative men on transmission of human papillomavirus to HIV-negative women: a randomised trial in Rakai. Uganda. *Lancet*. 2011;377:209-218.

A7. Munoz N, Kjaer SK, Sigurdsson K, et al. Impact of human papillomavirus (HPV)-6/11/16/18 vaccine on all HPV-associated genital diseases in young women. *J Nat Cancer Inst*. 2010;102:325-339.

A8. Giuliano AR, Palefsky JM, Goldstone S, et al. Efficacy of quadrivalent HPV vaccine against HPV infection and disease in males. *N Engl J Med*. 2011;364:401-411.

A9. Palefsky JM, Giuliano AR, Goldstone S, et al. HPV vaccine against anal HPV infection and anal intraepithelial neoplasia. *N Engl J Med*. 2011;365:1576-1585.

A10. Lehtinen M, Paavonen J, Wheeler CM, et al. Overall efficacy of HPV-16/18 AS04-adjuvanted vaccine against grade 3 or greater cervical intraepithelial neoplasia: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. *Lancet Oncol*. 2012;13:89-99.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Doorbar J, Quint W, Banks L, et al. The biology and life-cycle of human papillomaviruses. *Vaccine*. 2012;30(suppl 5):F55-F70.
2. Satterwhite CL, Torrone E, Meites E, et al. Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008. *Sex Transm Dis*. 2013;40:187-193.
3. Bosch FX, Broker TR, Forman D, et al. Comprehensive control of human papillomavirus infections and related diseases. *Vaccine*. 2013;31(suppl 7):H1-H31.
4. Hariri S, Markowitz LE, Dunne EF, et al. Population impact of HPV vaccines: summary of early evidence. *J Adolesc Health*. 2013;53:679-682.
5. Chung CH, Bagheri A, D'Souza G. Epidemiology of oral human papillomavirus infection. *Oral Oncol*. 2014;50:364-369.
6. Kreimer AR, Pierce Campbell CM, Lin HY, et al. Incidence and clearance of oral human papillomavirus infection in men: the HIM cohort study. *Lancet*. 2013;382:877-887.
7. Machalek DA, Poynten M, Jin F, et al. Anal human papillomavirus infection and associated neoplastic lesions in men who have sex with men: a systematic review and meta-analysis. *Lancet Oncol*. 2012;13:487-500.
8. Saslow D, Solomon D, Lawson HW, et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *CA Cancer J Clin*. 2012;62:147-172.
9. Forman D, de Martel C, Lacey CJ, et al. Global burden of human papillomavirus and related diseases. *Vaccine*. 2012;30(suppl 5):F12-F23.
10. Gillison ML, Alemany L, Snijders PJ, et al. Human papillomavirus and diseases of the upper airway: head and neck cancer and respiratory papillomatosis. *Vaccine*. 2012;30(suppl 5):F34-F54.
11. Schiffman M, Solomon D. Clinical practice. Cervical-cancer screening with human papillomavirus and cytologic cotesting. *N Engl J Med*. 2013;369:2324-2331.
12. Priebe AM. 2012 Cervical cancer screening guidelines and the future role of HPV testing. *Clin Obstet Gynecol*. 2013;56:44-50.
13. Workowski KA, Berman S. Sexually transmitted diseases treatment guidelines, 2010. *MMWR Recomm Rep*. 2010;59:1-110.
14. Massad LS, Einstein MH, Huh WK, et al. 2012 Updated consensus guidelines for the management of abnormal cervical cancer screening tests and cancer precursors. *Obstet Gynecol*. 2013;121:829-846.
15. Hariri S, Unger ER, Schafer S, et al. HPV type attribution in high grade cervical lesions: assessing the potential benefits of vaccines in a population-based evaluation in the United States. *Cancer Epidemiol Biomarkers Prev*. 2015;24:393-399.
16. Serrano B, Alemany L, Ruiz PA, et al. Potential impact of a 9-valent HPV vaccine in HPV-related cervical disease in 4 emerging countries (Brazil, Mexico, India and China). *Cancer Epidemiol*. 2014;38:748-756.
17. Herweijer E, Leval A, Ploner A, et al. Association of varying number of doses of quadrivalent human papillomavirus vaccine with incidence of condyloma. *JAMA*. 2014;311:597-603.
18. Moyer VA. Screening for cervical cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2012;156:880-891.

REVIEW QUESTIONS

1. A 24-year-old woman comes into a university health service for counseling. She has become romantically involved with a 25-year-old man and is contemplating the initiation of a sexual relationship but is afraid to do so because he has genital warts and she has heard that human papillomavirus (HPV) is an incurable infection. She received thee doses of the quadrivalent HPV vaccine 5 years ago and wants to know if her potential partner can be tested to determine whether she is protected against the HPV type with which he is infected. Which of the following is the most appropriate recommendation?

A. Her partner should come to the health service for assessment and treatment of his warts and typing of his HPV. If it is a type that is included in the HPV vaccine, sexual activity is safe.
B. Both she and her partner should be vaccinated, after which sexual activity would be safe.
C. She should avoid vaginal intercourse but can safely engage in oral sex.
D. Her partner should come to the health service for assessment and treatment of his warts but not testing. Sexual activity would then be safe with condoms.
E. She should be tested to determine whether she has protective levels of immunity to see if sexual activity would be safe.

Answer: D It is highly likely that her potential partner's warts are caused by HPV 6 or 11, both of which are included in the quadrivalent vaccine and against which she likely has a very high level of protection. Tests to provide specific typing information on HPV-related lesions or levels of HPV-antibody are not commercially available or recommended. In addition to vaccine, lesion treatment and condom use can further reduce the risk for transmission. Booster doses of HPV vaccine are not currently recommended, and the vaccine does not have therapeutic benefit against existing infection. Genital HPV can be transmitted by oral-genital sex, although oral warts occur only infrequently. See Veldhuizen NJ, Snijders PJF, Reiss P. Factors affecting transmission of mucosal papillomaviruses. *Lancet Infect Dis.* 2010;10:862-874; and Workowski KA; Centers for Disease Control and Prevention. Sexually transmitted disease treatment guidelines 2014. *MMWR Morb Mortal Wkly Rep.* 2014 (in press).

2. A 28-year-old woman presents with genital bumps 6 weeks after a brief sexual encounter while on vacation. The lesions are slightly itchy but are otherwise asymptomatic. She is concerned that this represents a recurrence of a prior episode of genital warts and would like to refill a prescription for imiquimod cream. Her physical examination is notable for moist bilateral condylomatous lesions on her labia minora and bilateral nontender inguinal lymphadenopathy. Which of the following would be the most appropriate diagnostic test?

A. A skin biopsy of the lesion
B. A blood test for syphilis
C. A swab of the lesion for HPV testing
D. A Pap test and an HPV co-test of her cervix
E. It would be most cost-effective not to perform any test, but an empirical course of imiquimod should be initiated.

Answer: B Although a skin biopsy is the definitive method for diagnosing genital lesions, such as suspected warts, genital warts can often be diagnosed by clinical examination without diagnostic testing. In this case, the clinical presentation (moist bilateral painless lesions with nontender lymphadenopathy) would be suspicious for the condyloma lata lesions of secondary syphilis, which can be easily diagnosed by a serologic test for syphilis; empirical treatment without such a test could delay appropriate treatment. HPV testing is recommended only as part of cervical cancer screening, not to diagnose a lesion. Cervical cancer screening would not help to assess her clinical presentation, and HPV co-testing is not recommended for women younger than 30 years because of its poor specificity. See Workowski KA; Centers for Disease Control and Prevention. Sexually transmitted disease treatment guidelines 2014. *MMWR Morb Mortal Wkly Rep.* 2014 (in press).

3. You have recommended a quadrivalent HPV vaccine for a 17-year-old bisexual man visiting your adolescent health clinic. He has read that this vaccine, like many others, can have serious side effects and wants to know why this vaccine might be beneficial for him to receive. Which of the following is *not* a potential benefit of HPV vaccine for him?

A. Prevention of genital warts
B. Prevention of cervical cancer in a future female partner
C. Prevention of plantar warts
D. Prevention of anal cancer
E. Prevention of oropharyngeal cancer

Answer: C The quadrivalent HPV vaccine is highly effective in preventing anogenital lesions due to the four HPV types contained in the vaccine (types 6/11, which commonly cause genital warts, and types 16/18, which cause a variety of types of genital and anal cancer). The vaccine is recommended in the United States for all males because of its benefit in preventing genital warts, anal intraepithelial neoplasia, and probably anal cancer. Preventing infection in young males has the potential to reduce the chance of infection in their future sexual partners, a phenomenon that has already been demonstrated for vaccinated females. Because HPV types 16/18 are also a cause of oropharyngeal cancer, the vaccine could potentially provide benefit in preventing this outcome as well. Plantar warts are generally caused by HPV 1 and other nongenital types, and the vaccine has no potential for preventing these lesions. See Bosch FX, Broker TR, Forman D, et al. Comprehensive control of human papillomavirus infections and related diseases. *Vaccine.* 2013;31S:H1-31.

4. HPV molecular testing is recommended in which of following clinical situations?

A. Combined with a Pap test to screen for cervical cancer in a 35-year-old woman
B. Screening for oral cancer in a 50-year-old man
C. Combined with a Pap test to screen for cervical cancer in a 26-year-old woman who had previously received the HPV vaccine
D. Combined with a Pap test to screen for cervical cancer in a 22-year-old woman
E. Screening for anal cancer in a 45-year-old man who has sex with men

Answer: A The explosion of knowledge regarding HPV and its role in a variety of cancers has led to great interest in the use of molecular tests for screening and diagnosis. Testing is currently recommended only in limited circumstances, specifically for cervical cancer screening in combination with a Pap test for women 30 years of age or older and for triage of women with cytologic findings of atypical squamous cells of undetermined significance. Because of the higher prevalence of HPV in women younger than 30 years, HPV testing is not recommended for screening owing to its low specificity for detecting significant cervical lesions; prior vaccination status does not affect this recommendation. There are no data to support the use of HPV testing in screening for cancer at sites other than the cervix. See Bosch FX, Broker TR, Forman D, et al. Comprehensive control of human papillomavirus infections and related diseases. *Vaccine.* 2013;31S:H1-31; Schiffman MH, Solomon D. Cervical cancer screening with human papillomavirus and cytologic co-testing. *N Engl J Med.* 2013;369: 2324-2331; and Workowski KA; Centers for Disease Control and Prevention. Sexually transmitted disease treatment guidelines 2014. *MMWR Morb Mortal Wkly Rep.* 2014 (in press).

5. A 45-year-old HIV-infected man recently saw a news report that warned about HPV-related health issues in HIV infected men. He now comes to you for counseling. In considering how to address his concerns, all of the following are true except which of the following?

A. HIV-infected persons are at increased risk for cervical and anal cancer.
B. HPV-related lesions are more difficult to treat in HIV-infected persons.
C. HPV may increase the risk that HIV-uninfected persons will acquire HIV.
D. Cancer screening with anal Pap tests has been proved to prevent anal cancer in HIV-infected men who have sex with men.
E. Although its protective benefit is unproved, HPV vaccines can be safely given to HIV-infected persons.

Answer: D Although cancers of both the cervix and anus occur with increased frequency in HIV-infected persons, cytologic screening has not been proved to prevent anal cancer and is not a recommended prevention strategy. Like many infections controlled by cell-mediated immunity, HPV infections are more common and difficult to treat in HIV-infected persons, especially those with advanced immunodeficiency. Although there is no proven benefit of HPV immunization in HIV-infected persons, there are no safety risks because the vaccine contains no live virus. Recent data indicate that HPV infection can increase the risk for HIV transmission, analogous to what is observed for many other sexually transmitted infections. See Denny LA, Franceshci S, de Sanjose S, et al. Human papillomavirus, human immunodeficiency virus, and immunosuppression. *Vaccine.* 2012;30S:F168-F174; and Workowski KA; Centers for Disease Control and Prevention. Sexually transmitted disease treatment guidelines 2014. *MMWR Morb Mortal Wkly Rep.* 2014 (in press).

374

HERPES SIMPLEX VIRUS INFECTIONS

RICHARD J. WHITLEY

DEFINITION

The Pathogen

Herpes simplex virus (HSV), a member of the family Herpesviridae, has been implicated in human infections since descriptions of cutaneous spreading lesions in ancient Greek times. In 1968, well-defined antigenic and biologic differences were demonstrated between HSV type 1 (HSV-1) and HSV type 2 (HSV-2). Of all the herpesviruses, HSV-1 and HSV-2 are most closely related, with approximately 60% genomic homology. Historically, HSV-1 was more frequently associated with nongenital infection and HSV-2 with genital

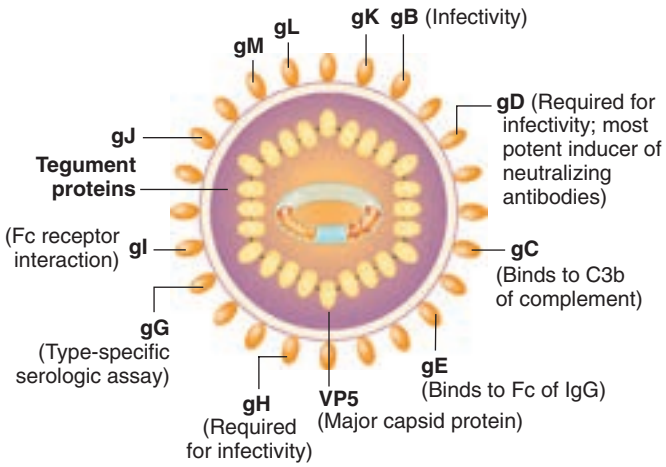

FIGURE 374-1. Schematic diagram of the herpes simplex virion. IgG = immunoglobulin G.

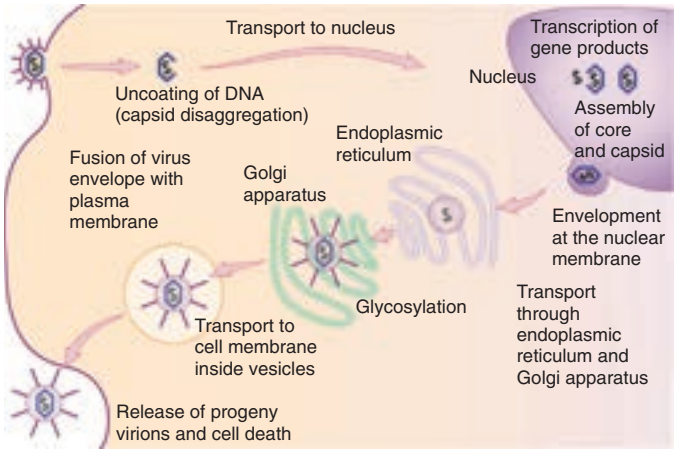

FIGURE 374-2. Schematic diagram of replication of herpes simplex virus.

disease, but these distinctions are much less relevant now that HSV-1 causes more than 50% of genital infection in some populations. These two viruses can be distinguished most reliably by DNA restriction enzyme analysis, but differences in antigen expression and biologic properties also serve as methods for differentiation.

Inclusion in the family Herpesviridae is based on the structure of the virion (Fig. 374-1). HSV contains double-stranded DNA at its central core, has a molecular weight of approximately 100 million, and encodes at least 80 polypeptides. The DNA core is surrounded by a capsid that consists of 162 capsomers arranged in icosapentahedral symmetry. The capsid is 100 to 110 nm in diameter. Tightly adherent to the capsid is the tegument, which consists of amorphous material. Loosely surrounding the capsid and tegument is a lipid bilayer envelope derived from host cell membranes. The envelope consists of polyamines, lipids, and glycoproteins. These glycoproteins confer distinctive properties to the virus and provide unique antigens to which the host is capable of responding. Notably, glycoprotein G (gG) provides antigenic specificity to HSV and therefore results in an antibody response that allows distinction between HSV-1 (gG-1) and HSV-2 (gG-2).

EPIDEMIOLOGY

HSV infections occur worldwide and have been reported in both developed and developing countries. Animal vectors for human HSV infections have not been described, and there is no seasonal variation in the incidence of HSV infections. HSV is transmitted from infected to susceptible individuals during close personal contact, and the virus must come in contact with mucosal surfaces or abraded skin for infection to be initiated. Because approximately a third of the world's population has recurrent HSV infections and because infection is rarely fatal, a large reservoir of HSV exists in the community.

Seroprevalence studies have demonstrated that acquisition of HSV-1 infection is related to socioeconomic factors. Antibodies, which indicate past infection, are found early in life among individuals of lower socioeconomic groups, presumably a consequence of crowded living conditions that provide a greater opportunity for direct contact with infected individuals. Antibodies develop in as many as 75 to 90% of individuals from lower socioeconomic populations by the end of the first decade of life. In contrast, only 30 to 40% of persons in the middle and upper socioeconomic groups are seropositive by the middle of the second decade of life.

Because infections with HSV-2 are usually acquired through sexual contact, antibodies to this virus are rarely found until the onset of sexual activity. There is a progressive increase in infection rates with HSV-2 in all populations beginning in adolescence. Overall, about one in five Americans has genital HSV-2 infection. As with HSV-1 infections, the rate of acquisition of HSV-2 infection appears to be related to socioeconomic factors. The number of sexual partners is also an important risk factor for the acquisition of HSV-2. Genital herpes infection has been found to be a risk factor for another sexually transmitted virus, human immunodeficiency virus (HIV; Chapter 386).

Localized, recurrent HSV-2 infection is the most common form of HSV infection during gestation. Transmission of infection to the fetus is most frequently related to shedding of the virus at the time of delivery. The incidence of cervical shedding in pregnant women with asymptomatic HSV infection is approximately 1%. Most infants in whom neonatal disease develops are born to women who are completely asymptomatic for genital HSV infection at the time of delivery and who have neither a past history of genital herpes nor a sexual partner reporting a genital vesicular rash. These women account for 60 to 80% of all women whose children acquire neonatal HSV infection. Women who experience a symptomatic or asymptomatic primary infection in the third trimester of gestation have a 30 to 50% risk for transmitting infection to the child.

PATHOBIOLOGY

Replication of HSV is a multistep process (Fig. 374-2). After the onset of infection, DNA is uncoated and transported to the nucleus of the host cell. This step is followed by transcription of immediate-early genes, which encode the regulatory proteins, and is followed by the expression of proteins encoded by early and then late genes.[1] These proteins include enzymes necessary for viral replication and structural proteins.

Assembly of the viral core and capsid takes place within the nucleus. Envelopment at the nuclear membrane and transport out of the nucleus occur through the endoplasmic reticulum and the Golgi apparatus. Glycosylation of the viral membrane occurs in the Golgi apparatus. Mature virions are transported to the outer membrane of the host cell inside vesicles. Release of progeny virus is accompanied by cell death. Replication for all herpesviruses is considered inefficient, with a high ratio of noninfectious to infectious viral particles.

A critical factor for transmission of HSV, regardless of virus type, is intimate contact between a person who is shedding virus and a susceptible host. With inoculation onto skin or mucous membranes, HSV replicates in epithelial cells; the incubation period is 4 to 6 days (Fig. 374-3). As replication continues, cell lysis and local inflammation ensue and result in the characteristic vesicles on an erythematous base. Regional lymphatics and lymph nodes become involved as a result of draining of infected secretions from the area of viral replication. Viremia and visceral dissemination may develop, depending on the immunologic competence of the host. In all hosts, the virus generally ascends peripheral sensory nerves to reach the dorsal root ganglia. Replication of HSV within neural tissue is followed by spread of the virus to other mucosal and skin surfaces by means of peripheral sensory nerves. HSV replicates further in epithelial cells and reproduces the lesions of the initial infection until infection is contained through host immune responses.

The histopathologic changes induced by HSV replication are similar in both primary and recurrent infection. Changes induced by viral infection include ballooning of infected cells and the appearance of condensed chromatin within the nuclei of cells, followed by subsequent degeneration of cellular nuclei. Cells lose intact plasma membranes and form multinucleated giant cells. They may also demonstrate intranuclear inclusions known as Cowdry type A bodies, which are suggestive but not diagnostic of HSV infection. With cell lysis, clear vesicular fluid containing large quantities of virus accumulates between the epidermis and dermal layer. The dermis reveals an intense inflammatory response, more so with primary infection than with recurrent disease. As healing progresses, the clear vesicular fluid becomes

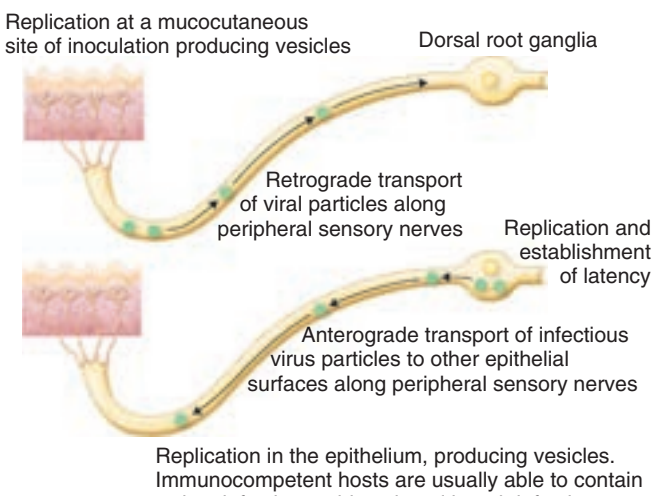

FIGURE 374-3. Schematic diagram of primary herpes simplex virus infection.

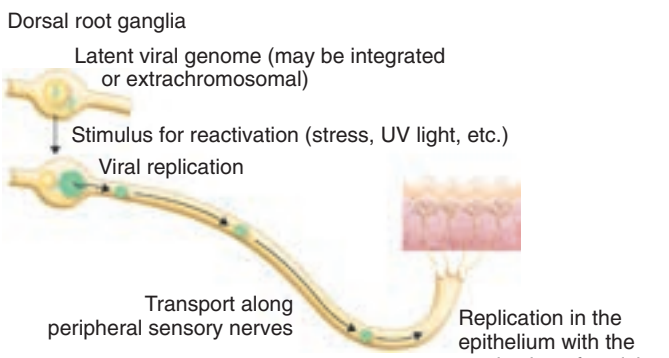

FIGURE 374-4. Schematic diagram of herpes simplex virus latency and reactivation. UV = ultraviolet.

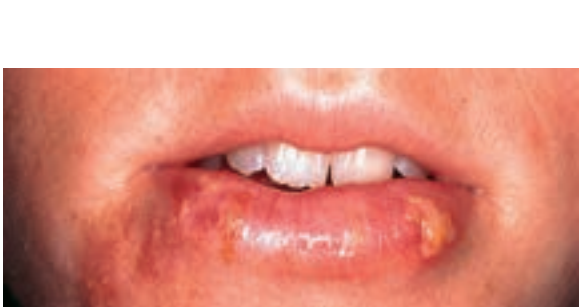

FIGURE 374-5. Herpes labialis, classic grouped blisters.

pustular with the recruitment of inflammatory cells. The pustule then forms a scab; scarring is uncommon.

Vascular changes in the area of infection include perivascular cuffing and hemorrhagic necrosis. These changes are particularly prominent when organs other than skin are involved, as is the case with herpes simplex encephalitis or disseminated neonatal HSV infection. Local lymphatics can show evidence of infection with intrusion of inflammatory cells because of draining of infected secretions from the area of viral replication. As host defenses are mounted, an influx of mononuclear cells can be detected in infected tissue.

A unique characteristic of all herpesviruses is their ability to establish latent infection, to persist in an apparently inactive state for varying lengths of time, and then to be reactivated (Fig. 374-4).[2] The latent viral genome may be either extrachromosomal or integrated into host cell DNA, depending on the virus.

Latency is established when HSV reaches the dorsal root ganglia after retrograde transmission through sensory nerve pathways. Latent virus may be reactivated and enter a replicative cycle at any point in time. Reactivation of latent virus is a well-recognized biologic phenomenon but not one that is understood from a molecular standpoint. Stimuli associated with the reactivation of latent HSV have included stress, fever, menstruation, and exposure to ultraviolet light. Precisely how these factors interact at the level of the ganglia remains to be defined. Reactivation may be clinically asymptomatic, or it may produce life-threatening disease.

CLINICAL MANIFESTATIONS

Gingivostomatitis

Gingivostomatitis (usually caused by HSV-1) occurs most frequently in children younger than 5 years. Illness is characterized by fever, sore throat, pharyngeal edema, and erythema, followed by the development of vesicular or ulcerative lesions on the oral and pharyngeal mucosa. Recurrent HSV-1 infections of the oropharynx are most frequently manifested as herpes simplex labialis (cold sores) and usually appear on the vermilion border of the lip (Fig. 374-5). Recurrences are triggered by fever, stress, and exposure to ultraviolet light as well as by other factors. Intraoral lesions as a manifestation of recurrent disease are uncommon.

Genital Herpes

Genital herpes is often caused by HSV-2, but approximately 50% of all new primary cases in young adults are caused by HSV-1. Primary infection in women usually involves the vulva, vagina, and cervix. In men, initial infection is most often associated with lesions on the glans penis, prepuce, or penile shaft. In individuals of either gender, primary disease is associated with fever, malaise, anorexia, and bilateral inguinal adenopathy. Women frequently have dysuria and urinary retention as a result of urethral involvement. Aseptic meningitis develops in as many as 10% of individuals with primary infection. Sacral radiculomyelitis may occur in both men and women and results in neuralgias, urinary retention, or obstipation. Complete healing of primary infection may take several weeks. The first episode of genital infection is less severe in individuals who have previously had HSV-1 infections at other sites. Antibodies to HSV-1 appear to ameliorate the expression of HSV-2 clinical disease, although this effect is controversial.

Recurrent genital infections in either men or women can be particularly distressing. The frequency of recurrence varies significantly from one individual to another. Of note, viral DNA can be detected by polymerase chain reaction (PCR) in genital secretions three- to four-fold more frequently than symptomatic recurrences. Furthermore, recurrences detected by PCR can occur frequently during 24 hours and persist for brief periods. A third of infected individuals have virtually no or few clinical recurrences, a third have approximately three recurrences per year, and another third have more than three per year. Several seroepidemiologic studies have found that between 25 and 65% of individuals in the United States have antibodies to HSV-2 and that seroprevalence is correlated with the number of sexual partners.

Shedding of HSV from the genital tract can be either symptomatic or, more frequently, asymptomatic. Transmission can occur in the absence of symptoms. Genital ulcerative disease attributed to HSV is a risk factor for the acquisition of HIV infection.

Herpetic Keratitis

Herpes simplex keratitis (Chapter 423) is usually caused by HSV-1 and is accompanied by conjunctivitis in many cases. It is considered the most common infectious cause of blindness in the United States. The characteristic lesions of HSV keratoconjunctivitis are dendritic ulcers best detected by fluorescein staining of the cornea. Deep stromal involvement has also been reported and may result in visual impairment.

Other Cutaneous Manifestations

HSV infections can occur at any skin site. Common among health care workers are lesions on abraded skin or the fingers, known as herpetic whitlow. Similarly, in wrestlers, disseminated cutaneous lesions known as herpes gladiatorum may develop as a result of physical contact.

HERPES SIMPLEX VIRUS INFECTIONS IN IMMUNOCOMPROMISED HOSTS

HSV infections in immunocompromised hosts, including patients with acquired immunodeficiency syndrome, are usually due to reactivation of

latent infection and are clinically more severe, may be progressive, and require a longer time to heal. Manifestations of HSV infections in this population of patients include pneumonitis, esophagitis, hepatitis, colitis, and disseminated cutaneous disease. Individuals suffering from HIV infection may have extensive perineal or orofacial ulcerations. HSV infections are also noted to be of increased severity in individuals with extensive burns.

DIAGNOSIS

Definitive diagnosis of HSV infection requires detection of viral DNA by PCR testing or isolation of the virus. DNA amplification has become the diagnostic method of choice in assessing cerebrospinal fluid (CSF) specimens for evidence of HSV infection of the central nervous system and has significantly improved sensitivity for confirmation of HSV as the cause of lip and genital herpes infection. The main role for viral cultures is to assess possible resistance to antiviral therapy.

In the absence of PCR or diagnostic virologic facilities, cytologic examination of cells scraped from a clinical lesion may be useful in making a presumptive diagnosis of HSV infection. Material obtained from scraping the base of a lesion should be smeared on a glass slide and promptly fixed in cold ethanol. The slide can be stained according to the methods of Papanicolaou, Giemsa, or Wright. The presence of intranuclear inclusions and multinucleated giant cells is indicative but not diagnostic of HSV infection. This method has a sensitivity of only 60 to 70% and should not be the sole diagnostic method used.

In addition to new tests for viral DNA, type-specific serologic assays are commercially available. These tests are based on differences between HSV-1 and HSV-2.

TREATMENT

Rx

For *immunocompromised* patients, such as patients who have cancer or are receiving immunosuppressive drugs, patients with disseminated mucocutaneous infections, or patients with herpes encephalitis, the treatment of choice is intravenous acyclovir (5 to 10 mg/kg every 8 hours for 5 to 7 days). Caution must be exercised when acyclovir is used intravenously because it may crystallize in the renal tubules when it is given too rapidly or to dehydrated patients. Valacyclovir (500 to 1000 mg two or three times daily) and famciclovir (250 to 500 mg three times daily), depending on severity, are equally efficacious and have improved pharmacokinetics, as demonstrated by improved biodistribution and less frequent dosing intervals compared with oral acyclovir. Thus, they should preferentially be used for the treatment of non-life-threatening HSV infection.[3] For life-threatening disease, only intravenous medications should be given. Immunocompromised individuals with non-life-threatening mucocutaneous HSV infections can be given oral valacyclovir (500 to 1000 mg once or twice daily for 5 to 7 days), famciclovir (250 to 500 mg three times daily for 7 days), or acyclovir (400 mg two or three times daily for 7 to 10 days).

High-risk immunocomprised hosts are at risk for developing an acyclovir-resistant infection, usually as a consequence of either an altered or deficient thymidine kinase enzyme that activates acyclovir in infected cells. Such patients can be managed with foscarnet or cidofovir, continued until there is evidence of healing, with the doses based on renal function (see Tables 360-4 and 360-5 in Chapter 360).

In *immunocompetent* hosts with mucocutaneous infections, oral acyclovir (200 mg five times daily for 5 days) reduces lesions and speeds recovery. Oral valacyclovir (1 g/day for 5 days or 500 mg twice daily for 3 days) and oral famciclovir (1 g twice daily for 1 day) are equally effective. Pritelivir (an inhibitor of the viral helicase-primase complex at 25 to 75 mg daily or 400 mg weekly) has recently been shown to reduce viral shedding and the duration of genital lesions in men and women, but is on clinical hold on the part of the FDA because of potential toxicity issues. Topical acyclovir is not as reliable as oral acyclovir and is not recommended.[4]

HSV has been used for experimental gene therapy. By removal of the $\gamma_1 34.5$ gene, both neurovirulence and the propensity to establish latency are ablated. These engineered viruses are being experimentally tested in patients with glioblastoma multiforme and colorectal metastases to liver.

PREVENTION

Some patients with particularly disabling oral recurrences or genital recurrences are candidates for chronic suppressive therapy. Potential regimens include oral acyclovir 400 mg twice daily, oral valacyclovir 500 mg or 1 g/day, and oral famciclovir 250 mg twice daily. Such regimens are approved by the U.S. Food and Drug Administration for genital herpes but not for herpes labialis. However, such therapy does not prevent recurrent HSV-2 meningitis.

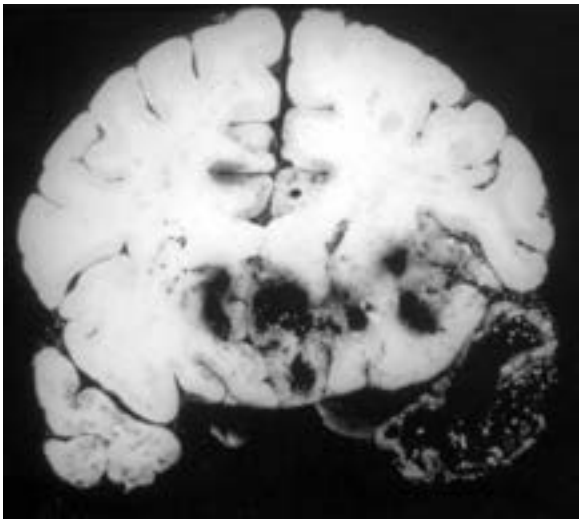

FIGURE 374-6. Hemorrhagic necrosis in herpes simplex encephalitis.

At present, experimental vaccines for HSV-1 and HSV-2 remain under investigation.[5] Acyclovir, valacyclovir, and famciclovir are given to recipients of solid organ and bone marrow transplants in the immediate post-transplantation period in an effort to prevent reactivation of latent disease at the doses noted previously. Valacyclovir (500 mg daily) decreases person-to-person transmission of HSV-2. Suppressive therapy of an HSV-2-infected but HIV-seronegative patient does not prevent HIV acquisition from HIV-seropositive partners.[6]

HERPES SIMPLEX ENCEPHALITIS

Herpes simplex encephalitis[7] is characterized by hemorrhagic necrosis of the temporal lobe (Chapter 414). Disease begins unilaterally and spreads to the contralateral temporal lobe (Fig. 374-6). It is the most common cause of focal, sporadic encephalitis in the United States today and occurs in approximately 1 in 150,000 individuals. Most cases are caused by HSV-1. The actual pathogenesis of herpes simplex encephalitis requires further clarification, although it has been speculated that primary or recurrent virus can reach the temporal lobe by ascending neural pathways, such as the trigeminal tracts or the olfactory nerves.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Clinical manifestations of herpes simplex encephalitis are characteristic of temporal lobe involvement and include headache, fever, altered consciousness, aphasia, and behavioral abnormalities. Focal seizures also may occur. CSF findings in these patients are variable but usually consist of a pleocytosis with both polymorphonuclear leukocytes and monocytes present. The protein concentration is characteristically elevated, and the glucose level is usually normal. Magnetic resonance imaging can suggest the diagnosis by demonstrating edema of either temporal lobe. Diagnosis can be confirmed by PCR detection of HSV DNA in the CSF by experienced laboratories.

TREATMENT AND PROGNOSIS

Rx

Mortality and morbidity are high, even with appropriate acyclovir antiviral therapy. At present, the mortality rate is approximately 30% 1 year after treatment. In addition, approximately 60% of survivors have moderate or severe neurologic impairment.

NEONATAL HERPES SIMPLEX VIRUS INFECTION

Neonatal HSV infection is estimated to occur in approximately 1 in 3500 deliveries in the United States each year.[8] Approximately 70% of cases are caused by HSV-2 and usually result from contact of the fetus with infected maternal genital secretions at the time of delivery. Manifestations of neonatal HSV infection can be divided into three categories: skin, eye, and mouth disease; encephalitis; and disseminated infection. As the name implies, skin, eye, and mouth disease consists of cutaneous lesions and does not involve other organ systems. Involvement of the central nervous system may occur

with encephalitis or disseminated infection and generally results in diffuse encephalitis. CSF assay characteristically reveals elevated protein levels and mononuclear pleocytosis. Disseminated infection involves multiple organ systems and can cause disseminated intravascular coagulation, hemorrhagic pneumonitis, encephalitis, and cutaneous lesions. Diagnosis is difficult in the absence of skin lesions, which occurs in as many as 36% of cases. The mortality rate for each disease classification varies from zero for skin, eye, and mouth disease to 5% for encephalitis and 25% for neonates with disseminated infection, even with appropriate antiviral treatment. In addition to the high mortality associated with these infections, morbidity is significant in that children with encephalitis or disseminated disease develop normally in only 40% of cases, even with appropriate antiviral therapy (acyclovir, 20 mg/kg every 8 hours for 14 days for skin, eye, and mouth infections and 21 days for central nervous system or disseminated disease).

Grade A References

A1. Glenny AM, Fernandez Mauleffinch LM, Pavitt S, et al. Interventions for the prevention and treatment of herpes simplex virus in patients being treated for cancer. *Cochrane Database Syst Rev*. 2009;1:CD006706.

A2. Cernik C, Gallina K, Brodell RT. The treatment of herpes simplex infections: an evidence-based review. *Arch Intern Med*. 2008;168:1137-1144.

A3. Abudalu M, Tyring S, Koltun W, et al. Single-day, patient-initiated famciclovir therapy versus 3-day valacyclovir regimen for recurrent genital herpes: a randomized, double-blind, comparative trial. *Clin Infect Dis*. 2008;47:651-658.

A4. Corey L, Wald A, Patel R, et al. Once-daily valacyclovir to reduce the risk of transmission of genital herpes. *N Engl J Med*. 2004;350:11-20.

A5. Wald A, Corey L, Timmler B, et al. Helicase-primase inhibitor pritelivir for HSV-2 infection. *N Engl J Med*. 2014;370:201-210.

A6. Aurelius E, Franzen-Rohl E, Glimaker M, et al. Long-term valacyclovir suppressive treatment after herpes simplex virus type 2 meningitis: a double-blind, randomized controlled trial. *Clin Infect Dis*. 2012;54:1304-1313.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Eisenberg RJ, Atanasiu D, Cairns TM, et al. Herpes virus fusion and entry: a story with many characters. *Viruses*. 2012;4:800-832.
2. Ma Y, He B. Recognition of herpes simplex viruses: toll-like receptors and beyond. *J Mol Biol*. 2014;426:1133-1147.
3. Vere Hodge RA, Field HJ. Antiviral agents for herpes simplex virus. *Adv Pharmacol*. 2013;67:1-38.
4. Patel R, Alderson S, Geretti A, et al. European guideline for the management of genital herpes, 2010. *Int J STD AIDS*. 2011;22:1-10.
5. Belshe RB, Leone PA, Bernstein DI, et al. Efficacy results of a trial of a herpes simplex vaccine. *N Engl J Med*. 2012;366:34-43.
6. Celum C, Wald A, Lingappa JR, et al. Acyclovir and transmission of HIV-1 from persons infected with HIV-1 and HSV-2. *N Engl J Med*. 2010;362:427-439.
7. Sili U, Kaya A, Mert A. Herpes simplex virus encephalitis: clinical manifestations, diagnosis and outcome in 106 adult patients. *J Clin Virol*. 2014;60:112-118.
8. Pinninti SG, Kimberlin DW. Maternal and neonatal herpes simplex virus infections. *Am J Perinatol*. 2013;30:113-119.

REVIEW QUESTIONS

1. A 55-year-old woman presents in the emergency department with fever, expressive aphasia, and altered behavior during the month of September. Her husband reports that the symptoms began approximately 6 hours before presentation. She is an avid hiker and spelunker. Her physical examination, aside from fever and expressive aphasia, is otherwise normal. Her cerebrospinal fluid findings are as follows: protein 75 mg/dL, 36 white blood cells (equally divided between lymphocytes and neutrophils), 0 red blood cells, and a normal glucose level. A computed tomographic scan of her head is normal. Which of the following is the most likely cause of her illness?

A. Enterovirus encephalitis
B. Rabies
C. Acute HIV infection
D. Herpes simplex virus
E. West Nile virus

Answer: D This is a classic presentation of a focal encephalopathic illness that is characteristic of herpes simplex virus encephalitis. Expressive aphasia and altered behavior are key to the diagnosis. About 20% of patients will not have red blood cells in their cerebrospinal fluid. Enterovirus very rarely causes a focal encephalopathic process; and when it does, it is in children. Rabies involves the cerebellum, whereas West Nile virus presents with anterior horn motor disease without altered behavior or expressive aphasia. Acute HIV encephalitis is a diffuse symmetrical encephalitis.

2. You elect to perform diagnostic studies on the aforementioned patient in order to determine the best management. Which of the following is most likely to confirm the diagnosis?

A. A serum immunoglobulin M (IgM) level
B. An electroencephalogram
C. A serum IgG (acute and convalescent) level
D. Biopsy of the nape of the neck
E. Polymerase chain reaction (PCR) analysis of her cerebrospinal fluid

Answer: E The "gold standard" for diagnosing herpes simplex encephalitis is PCR analysis of cerebrospinal fluid for detection of HSV DNA. Serology is of no value in making this diagnosis acutely. A neck biopsy is reserved for patients suspected of having rabies. An electroencephalogram might show spike and slow wave activity characteristic of HSV infection, but the sensitivity and specificity are low.

375

VARICELLA-ZOSTER VIRUS (CHICKENPOX, SHINGLES)

JEFFREY COHEN

DEFINITION

Primary infection with varicella-zoster virus (VZV) results in the rash of varicella (chickenpox). The virus establishes a latent infection in the nervous system and can reactivate later in life to cause zoster (shingles).[1]

The Pathogen

VZV is a member of the alpha herpesvirus family and has a DNA core surrounded by a nucleocapsid, which is in turn surrounded by a viral envelope that is studded with glycoproteins. Antibody to viral glycoproteins is important for neutralizing the virus's infectivity and for protecting against primary infection. The virus encodes a thymidine kinase, which phosphorylates acyclovir, which in turn inhibits viral DNA replication by inhibiting the VZV DNA polymerase.

EPIDEMIOLOGY

Before the advent of an effective vaccine, more than 95% of children in temperate climates were infected with VZV. By comparison, infection is usually delayed until adulthood in tropical climates. Varicella usually occurs in children younger than 5 years. Zoster is less common in tropical areas, probably because of a delay in acquisition of varicella. Varicella is more common in the winter and spring, whereas zoster has no seasonal predilection.

Primary varicella infection can occur after exposure to either chickenpox or zoster. The virus is spread by droplets and aerosols from patients or by contact with vesicular lesions. Persons are infectious beginning about 2 days before the rash appears and continuing until all lesions have crusted. Although 60 to 90% of susceptible household contacts develop varicella, only 20 to 30% of susceptible persons exposed to zoster become infected. More than 95% of primary infections result in the symptoms of varicella, and second episodes of varicella are rare. Varicella is more severe in persons with impaired cellular immunity, including patients with acquired immunodeficiency syndrome (AIDS) and infants whose mothers present with varicella 5 days before to 2 days after delivery.

About 50% of persons who have had varicella and live to age 85 years will develop zoster. The risk for zoster rises with increasing age (especially 50 years and older) and with increasing impairment of cellular immunity. For example, men with AIDS have a 20-fold higher risk for development of zoster than age-matched controls. Less than 5% of persons have a second episode of zoster, but recurrent zoster is more common in persons with impaired cellular immunity.

PATHOBIOLOGY

Varicella is transmitted by the respiratory route. The virus is thought to infect epithelial cells and lymphocytes in the oropharynx and upper respiratory tract or in the conjunctiva, and then infected lymphocytes subsequently spread the virus throughout the body. The virus then enters the skin through endothelial cells in blood vessels and spreads to epithelial cells, where it causes the vesicular rash of varicella. Lesions initially are vesicular but become pustular after the infiltration of inflammatory cells. Later the lesions break open and dry to form crusts that usually heal without scarring. During primary infection, neurons in cranial nerve ganglia and dorsal root ganglia become latently infected with the virus.

If VZV-specific cellular immunity declines, the virus can reactivate from a ganglion, travel down the axon, and replicate in epithelial cells to cause dermatomal zoster. In highly immunocompromised persons, high-grade viremia during reactivation causes disseminated zoster.

Antibodies, which usually are present at the time varicella presents clinically, persist for life. Antibody is important for protection against varicella, as evidenced by the ability of varicella immune globulin to attenuate the disease. Cytotoxic T cells are present within 2 to 3 days after the onset of varicella and limit its severity. Varicella is more severe in persons with impaired cellular immunity but not in patients with hypogammaglobulinemia. Cellular immunity, not antibody, is required to prevent reactivation of virus and zoster.

CLINICAL MANIFESTATIONS

Varicella

Varicella begins with fever and malaise followed 1 to 2 days later by a disseminated, pruritic, vesicular rash (Fig. 375-1).[2] The usual incubation period for varicella is 2 weeks (range, 10 to 21 days) after exposure to an infected person. The lesions begin as papules that become vesicles, followed by pustules and then crusts. Lesions appear on the head and then spread to the trunk and then to the extremities; the mucosa can also be involved. There are typically 200 to 500 lesions in different stages on the skin. New lesions occur for up to 5 days in normal hosts, and crusting is complete within 2 weeks.

The most common complication of varicella is bacterial superinfection of skin lesions. Group A streptococcus (Chapter 290) or *Staphylococcus aureus* (Chapter 288) infections can cause cellulitis, bacteremia, and necrotizing fasciitis. Other complications include cerebellar ataxia, viral pneumonitis, hepatitis, and thrombocytopenia. Less frequent complications include viral meningitis, encephalitis, vasculopathy (which presents as a stroke), disseminated intravascular coagulopathy (Chapter 175), and Reye syndrome (more common in children receiving aspirin). Complications involving the lungs and liver are more common in children with impaired cellular immunity including those receiving systemic steroids, children with chronic pulmonary or skin disease, adults, and pregnant women during the third trimester. The fetal varicella syndrome, which occurs in fetuses infected during the first trimester, is characterized by atrophy of limbs with scarring of skin, chorioretinitis or cataracts, and central nervous system abnormalities. Patients with AIDS and moderately reduced CD4 cell counts may develop recurrent varicella lesions in the absence of new exposures, and patients with CD4 cell counts below 200/μL may develop progressive varicella with new lesions occurring for at least 1 month or chronic verrucous lesions.

Zoster

In healthy persons, zoster presents with localized pain and increased sensation for 1 to 3 days before the development of a dermatomal vesicular rash

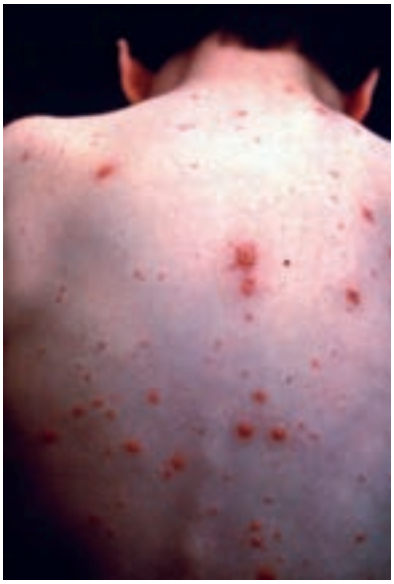

FIGURE 375-1. Child with varicella. (Courtesy of Centers for Disease Control and Prevention.)

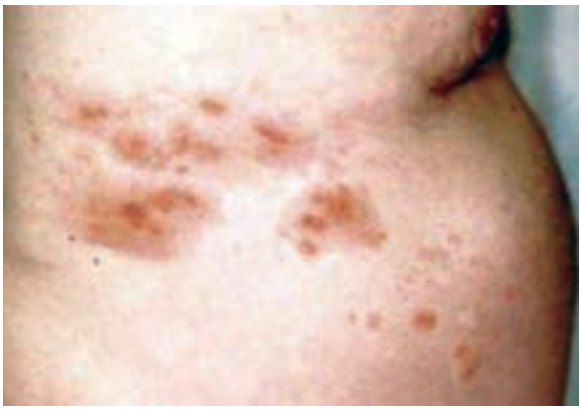

FIGURE 375-2. Dermatomal zoster. (Courtesy of Centers for Disease Control and Prevention.)

that does not cross the midline (Fig. 375-2). Zoster most frequently presents in the dermatomes innervated by trigeminal or thoracic ganglia. The rash is usually accompanied by itching, tingling, or pain. The lesions evolve from vesicles to pustules, and crusting is usually complete by 10 days. In normal hosts, a few lesions may develop outside of the dermatome owing to low-grade viremia. Some patients with zoster sine herpete never develop a rash. In persons with very impaired cellular immunity, reactivation is often associated with high-grade viremia with dissemination to large areas of the skin and involvement of multiple organs. As a result, patients with underlying malignancies are more likely to develop more serious complications from their zoster infections.[3]

A dreaded complication of zoster is post-herpetic neuralgia (Chapter 30) with pain persisting for at least 1 month after the rash has resolved. Patients may have allodynia (sensation of pain after nonpainful stimuli), paresthesias, dysesthesias, or severe neuropathic pain. Post-herpetic neuralgia is more common in persons older than 50 years. Other complications of zoster include bacterial superinfection; ocular disease, involving any of the structures of the eye, due to reactivation in the ophthalmic branch of the trigeminal ganglia; facial palsy caused by reactivation in the VII cranial nerve; Ramsay Hunt syndrome, with pain and vesicles in the ear, numbness of the anterior tongue, and ipsilateral facial palsy due to reactivation in the geniculate ganglion of the VII cranial nerve; motor neuropathy; and meningitis. Zoster vasculopathy, occurring at the time of zoster or a few months later, can present with stroke due to inflammation of the cerebral arteries.

Progressive outer retinal necrosis, with few inflammatory ocular cells, occurs when VZV reactivates in the eye of severely immunocompromised persons, including patients with AIDS and low CD4 cell counts. In contrast, acute retinal necrosis with a marked inflammatory response occurs when the virus reactivates in otherwise healthy persons. Patients with AIDS or recipients of hematopoietic stem cell transplants can have pancreatitis, hepatitis, and pneumonitis in the absence of or preceding a rash.

DIAGNOSIS

Most cases of varicella and zoster are diagnosed on the basis of their clinical presentation. A disseminated vesicular rash with lesions in various stages of evolution is usually sufficient for a diagnosis of varicella. The differential diagnosis includes impetigo (Chapter 441), enterovirus infections (Chapter 379), herpes simplex (Chapter 374), Stevens-Johnson syndrome (Chapter 440), and guttate psoriasis (Chapter 438). A dermatomal vesicular rash that does not cross the midline in a patient with a prior history of pain in the area is usually diagnostic of zoster. Herpes simplex is the most common disease that resembles zoster.

When the diagnosis of varicella or zoster must be confirmed definitively, polymerase chain reaction (PCR) for VZV from vesicular fluid is the most sensitive and specific test. PCR for VZV in the blood can be useful for diagnosis of visceral zoster in highly immunocompromised persons before the onset of rash. PCR for VZV in the cerebrospinal fluid and intrathecal synthesis of VZV-specific antibody is useful for diagnosis of VZV neurologic diseases. Culture is less sensitive than PCR because the virus is very labile. Direct fluorescent antibody testing of vesicle fluid is rapid but less sensitive than PCR. Detection of multinucleated giant cells (Tzanck smear) is less specific because lesions of herpes simplex virus have a similar appearance. Biopsy specimens show eosinophilic intranuclear inclusion bodies and multinucleate giant cells.

Serology for VZV is useful to determine the need for postexposure prophylaxis in persons who are at high risk for disease after exposure to varicella or zoster. Enzyme-linked immunosorbent assay tests are less sensitive than latex agglutination assays and may not detect antibodies in vaccinees.

TREATMENT Rx

Varicella

Symptomatic treatment includes acetaminophen for fever and lotion or baths for pruritus. Although acyclovir is licensed for the treatment of varicella, the drug is not recommended for otherwise healthy children because it only modestly decreases symptoms by about 1 day. Acyclovir reduces visceral dissemination in immunocompromised persons, in whom intravenous acyclovir (500 mg/m^2 every 8 hours for children, 10 mg/kg every 8 hours for adults) is recommended for 7 to 10 days or until all lesions have crusted. Oral acyclovir (20 mg/kg four times daily for children or 800 mg five times daily for adults) given within 24 hours after the onset of rash reduces the duration of symptoms and is recommended for treatment of adolescents, adults, newborns whose mothers developed varicella near the time of delivery, immunocompromised persons, children with chronic pulmonary or skin disease, and persons with complications of varicella. Acyclovir also should be considered for household contacts of persons with varicella or for pregnant women in the third trimester; these patients often have more severe disease. Oral valacyclovir is also approved for treatment of children ages 2 years to <18 years with varicella (20 mg/kg three times daily with a maximum dose of 1 g). Oral valacyclovir, 1 g three times daily, or famciclovir, 500 mg three times daily result in higher antiviral drug levels than does oral acyclovir and can be used in nonpregnant adults.

Zoster

Acyclovir, valacyclovir, and famciclovir (for 7 days at the same doses as for varicella) are licensed for the treatment of zoster. Oral valacyclovir and famciclovir result in higher levels of antiviral activity than oral acyclovir does. Although therapy should be started within 3 days of the rash, therapy may still be of benefit if new lesions continue to occur after this time. Because patients younger than 50 years usually have little pain associated with zoster, antiviral therapy is often not used in these patients unless they have moderate to severe pain, have disease involving the eye, have other complications, or are immunocompromised. Antiviral drugs (see earlier) reduce the duration of lesions and zoster-associated pain but not the incidence of post-herpetic neuralgia. In severely immunocompromised persons, intravenous acyclovir (7 to 10 days or until all lesions have crusted) reduces the risk for visceral dissemination. Oral valacyclovir or famciclovir may be used in persons who are less severely immunocompromised.

Corticosteroids (e.g., prednisone, 60 mg/day and tapered over 21 days), in combination with acyclovir, reduce acute pain and improve the quality of life in persons older than 50 years but do not reduce the risk for post-herpetic neuralgia. Patients with moderate to severe pain often require narcotics.

Treatment of post-herpetic neuralgia is challenging.[4] Gabapentin (initiated at a dose of 300 mg at bedtime and titrated to a maximum dose of 1200 mg three times daily) or pregabalin (initiated at a dose of 75 mg at bedtime and titrated to a maximum dose of 300 mg twice daily) may reduce pain. Additional agents include nortriptyline (initiated at a dose of 25 mg at bedtime and titrated to a maximum dose of 150 mg daily), lidocaine patches, and topical capsaicin (which itself causes pain that is not tolerated in up to one third of patients). Opioid analgesics (see Table 30-5) may be needed, but there are concerns about long-term efficacy and safety.

Treatment of VZV Complications and Acyclovir-Resistant VZV

Intravenous acyclovir is recommended for persons with acute retinal necrosis. Corticosteroids (e.g., prednisone, 1 mg/kg per day for 3 to 5 days) and intravenous acyclovir (10 to 15 mg/kg every 8 hours for 14 days) are recommended for nonimmunocompromised persons with VZV vasculopathy. Zoster involving the eye should be evaluated by an ophthalmologist to assess the potential value of topical or intraocular therapy, such as the need to reduce intraocular pressure to treat glaucoma or to use mydriatics to prevent synechiae.

Acyclovir-resistant VZV infections are rare and are limited almost exclusively to patients with AIDS or recipients of transplants. Foscarnet (40 mg/kg every 8 hours) for 2 weeks or until all lesions have crusted is the treatment of choice for acyclovir-resistant VZV.

PREVENTION

Varicella

Patients with varicella or zoster are considered infectious until all lesions have completely crusted. Airborne and contact precautions are recommended for varicella, whereas only contact precautions are necessary for immunocompetent persons with localized zoster.

The live attenuated varicella vaccine is recommended for children aged 1 to 12 years and for persons 13 years and older without immunity to the virus. The vaccine is 75 to 90% effective in protection against symptomatic varicella and more than 95% effective in protecting against severe disease.[A] Two doses of vaccine are given subcutaneously. The varicella vaccine is also given as part of a combined measles, mumps, rubella vaccine (MMRV) for children 1 to 12 years of age in the United States. The rate of disease due to varicella declined by 90% in the United States during the first 13 years after the vaccine was licensed.[5]

The most common complications of varicella vaccination are pain at the injection site, fever, and a mild rash within 2 weeks after vaccination. The rash is usually localized to the area of vaccination and is often papular; in some healthy persons, the rash can be disseminated, although there are fewer lesions and symptoms are much less severe than with wild-type virus. In persons with severely impaired cellular immunity, rash is more common, can be extensive, and may be accompanied by organ dysfunction. The varicella vaccine establishes latency and can cause shingles, although this complication occurs less commonly with vaccine virus than with wild-type virus. The vaccine strain of varicella has been transmitted to third parties only by vaccinees who developed a rash. Varicella vaccine is contraindicated in pregnant women and persons receiving high-dose immunosuppressive therapy (e.g., ≥2 mg/kg of prednisone daily) or with hematologic malignant neoplasms. Vaccination should be considered for human immunodeficiency virus (HIV)–infected children with age-specific $CD4^+$ T cells of 15% or more and adolescents and adults with $CD4^+$ T-cell counts of 200 cells/μL or higher. Serologic testing to verify immunity is not recommended for health care workers who have received two doses of vaccine because the currently available commercial antibody assays are not sensitive enough to detect protective levels of antibody.

Zoster

Zoster vaccine is approved by the U.S. Food and Drug Administration (FDA) for persons 50 years of age or older[A] and recommended by the Advisory Committee on Immunization Practices for persons 60 years of age or older (Chapter 18). This vaccine is similar to the varicella vaccine, except that the titer of virus is about 14-fold higher. The vaccine is about 50% protective in preventing zoster and 66% effective in preventing post-herpetic neuralgia.[6] Although the vaccine's efficacy to prevent zoster declines in persons older than 70 years, efficacy to prevent post-herpetic neuralgia does not decline.[A] Transient pain and erythema at the injection site are not uncommon, but serious complications attributable to the vaccine have not been reported. The vaccine also can be given safely to adults with a prior history of zoster.[7] The vaccine is contraindicated in persons with hematologic malignant neoplasms, AIDS, or HIV infection with CD4 count of 200/μL or lower and in persons receiving high-dose immunosuppressive therapy (e.g., ≥20 mg of prednisone daily) or anti–tumor necrosis factor-α therapy.

Postexposure Prophylaxis

In persons exposed to varicella, three options are available. Varicella vaccine is preferred if exposure occurred within the prior 3 days and the patient is not immunocompromised.[A] Vaccine is estimated to be 70 to 90% effective in healthy persons. VariZIG (formerly available as varicella immune globulin) prevents or attenuates varicella in 90% of susceptible persons if it is given within 4 days of exposure. The FDA recently approved VariZIG for use within 10 days of exposure, but it should be given as soon as possible after exposure.[8] VariZIG (given intramuscularly at 125 units/10 kg body weight, up to a maximum of 625 units) is indicated for susceptible persons at risk for severe varicella (e.g., pregnant women, preterm infants, neonates whose mothers have varicella between 5 days before and 2 days after delivery, immunocompromised persons) who are in close contact with patients with varicella or zoster. VariZIG has no effect in the treatment of zoster.

Oral acyclovir (40 to 80 mg/kg for 1 week beginning 7 to 9 days after exposure) is estimated to be 80 to 85% effective as postexposure prophylaxis. It is often used when the exposure occurred too long ago for vaccination or VariZIG.

PROGNOSIS

Before vaccination, about 100 children died of varicella each year in the United States. Now varicella is the underlying cause of death in an average of less than 3 Americans younger than 20 years annually.

Grade A References

A1. Klassen TP, Hartling L, Wiebe N, et al. Acyclovir for treating varicella in otherwise healthy children and adolescents. *Cochrane Database Syst Rev*. 2005;4:CD002980.

A2. Chen N, Li Q, Yang J, et al. Antiviral treatment for preventing postherpetic neuralgia. *Cochrane Database Syst Rev*. 2014;2:CD006866.

A3. Committee on Infectious Diseases. Policy statement. Prevention of varicella: update of recommendations for use of quadrivalent and monovalent varicella vaccines in children. *Pediatrics*. 2011;128:630-632.

A4. Schmader KE, Levin MJ, Gnann JW Jr, et al. Efficacy, safety, and tolerability of herpes zoster vaccine in persons aged 50-59 years. *Clin Infect Dis*. 2012;54:922-928.

A5. Gagliardi AM, Gomes Silva BN, Torloni MR, et al. Vaccines for preventing herpes zoster in older adults. *Cochrane Database Syst Rev*. 2012;10:CD008858.

A6. Macartney K, McIntyre P. Vaccines for post-exposure prophylaxis against varicella (chickenpox) in children and adults. *Cochrane Database Syst Rev*. 2014;6:CD001833.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Cohen JI. Herpes zoster. *N Engl J Med*. 2013;369:1766-1767.
2. Breuer J, Fifer H. Chickenpox. *Clin Evid (Online)*. 2011;4:912.
3. Tran TN, Ray GT, Horberg MA, et al. Complications of herpes zoster in cancer patients. *Scand J Infect Dis*. 2014;46:528-532.
4. Johnson RW, Rice AS. Clinical practice. Postherpetic neuralgia. *N Engl J Med*. 2014;371: 1526-1533.
5. Baxter R, Ray P, Tran TN, et al. Long-term effectiveness of varicella vaccine: a 14-year, prospective cohort study. *Pediatrics*. 2013;131:e1389-e1396.
6. Oxman MN. Zoster vaccine: current status and future prospects. *Clin Infect Dis*. 2010;51:197-213.
7. Morrison VA, Oxman MN, Levin MJ, et al. Safety of zoster vaccine in elderly adults following documented herpes zoster. *J Infect Dis*. 2013;208:559-563.
8. Centers for Disease Control and Prevention. FDA approval of an extended period for administering VariZIG for postexposure prophylaxis of varicella. *MMWR Morb Mortal Wkly Rep*. 2012;61:212.

REVIEW QUESTIONS

1. Which of the following people is at lowest risk for severe zoster and would not necessarily need to receive antiviral medication?

A. A 40-year-old woman with zoster involving the eye
B. A 65-year-old man with no underlying disease
C. A 30-year-old woman being treated for Hodgkin lymphoma
D. A 40-year-old woman with severe pain associated with the rash
E. A 45-year-old man with no other underlying disease

Answer: E The 45-year-old man is at least risk for complications associated with zoster. Antiviral therapy can reduce the risk for disseminated disease in immunocompromised persons and also reduce organ disease, including the eye. Antiviral therapy can reduce the acute pain associated with zoster. Persons older than 50 years are more likely to have complications associated with zoster and should be treated with antiviral therapy. (Recommendations for antiviral therapy are reported in Cohen JI. Herpes zoster. *N Engl J Med.* 2013;369:1766-1767.)

2. Which of the following persons should receive the zoster vaccine?

A. An 80-year-old woman
B. A 65-year-old man with Hodgkin disease
C. A 45-year-old man
D. A 66-year-old woman receiving 25 mg of prednisone daily for lupus
E. A 60-year-old man with HIV and a CD4 cell count <200/μL

Answer: A Vaccine is approved by the Advisory Committee on Immunization Practices for persons older than 60 years. The vaccine is contraindicated in persons with hematologic malignancies and patients receiving 20 mg or more of prednisone daily. It is not approved by the U.S. Food and Drug Administration for persons younger than 50 years. The vaccine is contraindicated in persons with CD4 cell counts of ≤200/μL. (Recommendations for the zoster vaccine are reported in Harpaz R, Ortega-Sanchez IR, Seward JF, et al. Prevention of herpes zoster: recommendations of the Advisory Committee on Immunization Practices [ACIP]. *MMWR Recomm Rep.* 2008;57:1-30.)

3. A 30-year-man who received a hematopoietic stem cell transplant 2 months ago has a disseminated vesicular rash. He had stopped taking his valacyclovir prophylaxis. Which is the most sensitive and specific test to determine whether his rash is due to varicella-zoster virus (VZV)?

A. Direct fluorescent antibody staining of cells from a skin lesion
B. Culture of a skin lesion for VZV
C. Polymerase chain reaction testing of a skin lesion for VZV
D. Tzanck smear (cytology) of a skin lesion
E. Antibody test for VZV

Answer: C The patient's rash could be due to herpes simplex or enterovirus. Polymerase chain reaction is the most sensitive and specific test for VZV. The direct fluorescent antibody test is very specific but requires sufficient numbers of cells and lesions and therefore is less sensitive. VZV is very labile, and cultures are often negative, especially if not received in the laboratory rapidly. The Tzanck smear does not distinguish HSV from VZV. Antibody testing is not useful to make a diagnosis of the cause of the rash. (Comparison of the VZV polymerase chain reaction test with other tests is reported in Sauerbrei A, Eichhorn U, Schacke M, et al. Laboratory diagnosis of herpes zoster. *J Clin Virol.* 1999;14:31-36.)

376

CYTOMEGALOVIRUS

W. LAWRENCE DREW

DEFINITION

Cytomegalovirus (CMV) is a member of the herpesvirus family and shares, with the other members, the ability to establish a long-lived latent infection. Most of the clinical disease caused by this virus results from reactivation of latent virus in immune-impaired patients, although primary infection in such patients can also be devastating.

The Pathogen

CMV has a linear, double-stranded DNA genome with about 250,000 base pairs that encode about 160 proteins. On microscopic examination, the hallmark of CMV infection is a large (cytomegalic), 25- to 35-μm cell containing

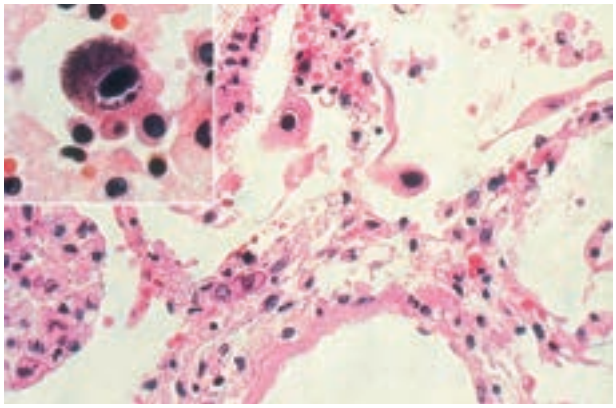

FIGURE 376-1. Cytomegalovirus (CMV) pneumonia. A lung biopsy specimen was stained with hematoxylin and eosin and magnified 250-fold. The *inset* shows a CMV "owl's eye" inclusion.

a large central, basophilic intranuclear inclusion (Fig. 376-1), referred to as an owl's eye.

EPIDEMIOLOGY

Multiple mechanisms account for the spread of this virus, including vertical (in utero, during vaginal delivery, and by breast milk) and horizontal (saliva, genital, urine) contact. These routes of transmission lead, collectively, to a 15 to 20% seroprevalence by 15 years of age in developed countries, with a higher seroprevalence in lower socioeconomic settings.[1] From that age on, there is a steady upward trend of 1 to 2% per year that is due in part to sexual transmission. As a result, approximately 50% of the general population of the United States is antibody positive by 35 years of age, and a 1% per year rate of increase occurs thereafter. In underdeveloped countries, up to 90% of persons may be seropositive by 2 years of age. Presumably, crowded living conditions permit spread of the virus through close contact with body fluids. A recent study showed that CMV was viable on metal and wood for up to 1 hour, glass and plastic to 3 hours, and rubber, cloth, and crackers to 6 hours.[2] CMV was more likely to be isolated from wet, highly absorbent surfaces. These considerations were felt to be particularly important because children may actively shed CMV in saliva and urine for months to years, and exposure to bodily fluids from young children poses substantial risk for CMV exposure among women of reproductive age. Two additional mechanisms of transmission are blood transfusion and organ transplantation. A final important epidemiologic fact is that reinfection with a different strain of CMV may occur in CMV-seropositive persons, especially those who are immunocompromised, pregnant, or sexually promiscuous.

PATHOBIOLOGY

In fully immunocompetent individuals, CMV rarely causes clinically evident end-organ disease. When immune mechanisms are deficient, especially those mediated by $CD4^+$ and $CD8^+$ lymphocytes, latent virus replicates and causes both direct and indirect effects.[3] Examples of direct virally mediated diseases are necrotizing CMV retinitis and esophagitis. In contrast, CMV pneumonitis is frequently manifested as subtle histologic alterations accompanied by limited viral replication, thus suggesting that immune-mediated injury may be the primary pathologic mechanism. Such injury may result from the upregulation and release of cytokines, including tumor necrosis factor-α, interferon-γ, and interleukin-2. Immune-mediated tissue injury may also be effected by $CD8^+$ cytotoxic T lymphocytes directed against CMV-infected target cells. The clinical manifestations of CMV infection, including meningoencephalitis, retinitis, enteritis, vasculitis, pneumonitis, myocarditis, lymphadenitis, hepatitis, adrenalitis, and pancreatitis, reflect the range of cell types that CMV is capable of infecting.

The immune response to CMV infection involves both the humoral and cell-mediated arms, but the $CD8^+$ cytotoxic T-cell response appears to be the most important. The CMV envelope glycoproteins that participate in viral entry are gB, gH/gL, and gCII. Humoral immunity directed at gB has been detected in convalescent phase sera and has been shown to block viral entry, cell-to-cell transmission, and syncytium formation in CMV-infected cells.

Fundamental to the pathogenesis of CMV is latency, or persistence of the viral genome in host cells without evidence of productive viral replication. It is thought that monocytes and bone marrow progenitor cells are sites of human CMV latency. Reactivation from the latent state has classically been associated with immunosuppression. Exposure to a rich milieu of cytokines and growth factors results in the activation of signal transduction pathways, generation of increased levels of intracellular transcription factors, and production of viable virus.

CLINICAL MANIFESTATIONS

Congenital and Neonatal Infection

In the developed world, congenital infection occurs in approximately 0.2 to 0.7% of newborns.[4] Thus in the United States each year, approximately 40,000 infants are born excreting CMV, and about 4000 (or ~10%) of these newborns show clinical evidence of congenital disease, such as microcephaly, intracerebral calcification, hepatosplenomegaly, and rash. About 90% of these clinically infected newborns will survive, but half of the survivors will have unilateral or bilateral hearing loss, mental retardation, or both. Mothers of most infants with these stigmata had a primary infection during pregnancy, although it is now well known that clinically evident congenital infection also occurs in infants born to mothers with past as well as with primary CMV infection.

Infection in Immunocompetent Persons

Virtually all CMV infections occurring in immunocompetent persons are asymptomatic. In some patients, a clinical illness resembling infectious mononucleosis may develop (Chapter 377), but with minimal pharyngitis and lymphadenopathy. Atypical lymphocytosis develops in these patients, similar to Epstein-Barr virus infection, but they have a negative heterophil antibody test result. CMV reactivation is common (33%) in critically ill immunocompetent patients, in whom it is associated with prolonged hospitalization and mortality, but whether it causes these effects is unclear.

Infection in Transplant Recipients

When a CMV-seronegative recipient receives a solid organ from a CMV-seropositive donor, the resulting illnesses include the "CMV syndrome," characterized by fever, neutropenia, atypical lymphocytes, and often hepatosplenomegaly. CMV disease may also develop in the transplanted organ. For example, CMV hepatitis in liver transplant recipients is associated with fever, hyperbilirubinemia, and elevated liver enzymes; liver failure may ensue and necessitate retransplantation. CMV-seropositive recipients of a solid organ transplant may develop CMV disease by either reactivation of latent infection or by transmission of CMV from a seropositive donor. Disease in CMV-seropositive recipients is, fortunately, less severe than that resulting from primary infection. CMV infection occurs more commonly in recipients of lung or liver transplants than in recipients of kidney transplants.

CMV pneumonia may occur after solid organ transplantation but is most common after stem cell transplantation. Fever, nonproductive cough, and dyspnea can progress rapidly. The diagnosis is suggested by interstitial to nodular infiltrates rather than by alveolar densities on chest radiographs. In contrast to solid organ transplantation, CMV disease after stem cell transplantation usually results from reactivation of latent CMV in a seropositive recipient rather than from a new, primary infection.

CMV may cause disease throughout the gastrointestinal tract. Colitis, which is a common syndrome in transplant recipients, is manifested as diarrhea, weight loss, and fever. It is characterized by diffuse submucosal hemorrhages and ulcerations.

Infection in Patients with Acquired Immunodeficiency Syndrome

In the era before highly active antiretroviral therapy, CMV retinitis occurred in approximately one third of patients with acquired immunodeficiency syndrome (AIDS), most often in those with CD4 counts below $50/mm^3$. It usually begins unilaterally with visual blurring, floaters, decreased acuity, and loss of visual fields and progresses to blindness if it is untreated. The retinal examination is abnormal, and the finding of apparent hemorrhages and exudates is the best diagnostic test (Fig. 376-2). CMV colitis is similar to that seen in transplant recipients, but esophagitis is also common and characterized by distal ulceration, which may be single but extensive. CMV neurologic disease occurs in multiple forms, including encephalitis and a polyradiculopathy/myelitis syndrome. With the efficacy of combination antiretroviral therapy, the incidence of all of these CMV syndromes has decreased dramatically, but they are still seen before HIV treatment or when such treatment is interrupted or ineffective.

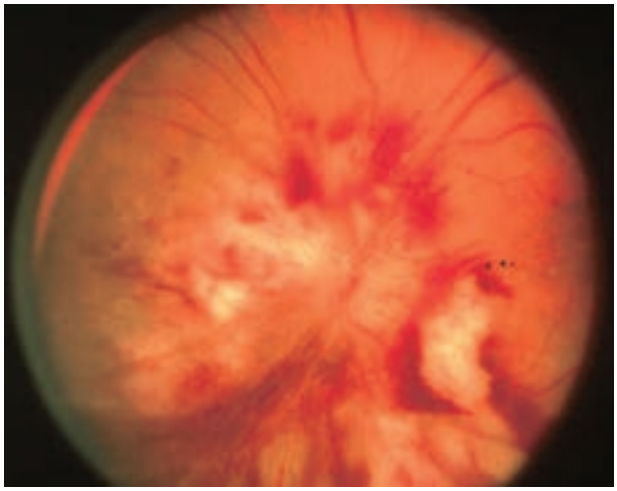

FIGURE 376-2. Cytomegalovirus retinitis as seen by direct ophthalmoscopic examination.

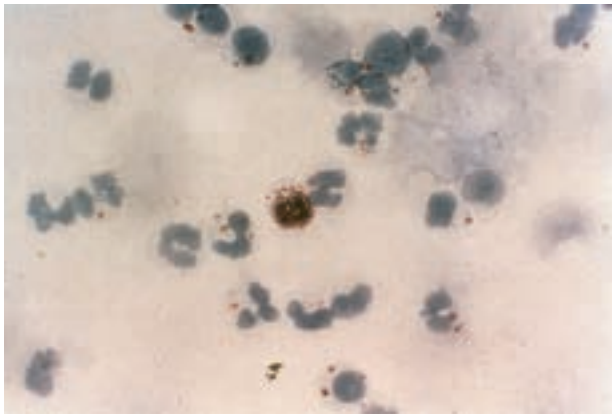

FIGURE 376-3. Peripheral blood leukocytes stained with monoclonal antibody to cytomegalovirus pp65 antigen by the immunoperoxidase technique (magnification ×500).

DIAGNOSIS

Assay of viral DNA by the polymerase chain reaction (PCR) is more sensitive than viral culture and is the best assay for the early detection of CMV disease. The assay can quantify the CMV viremia and thereby help to determine its clinical significance and to monitor therapy. Monoclonal antibodies can also be used to quantify viremia by counting CMV antigen–positive cells directly in peripheral blood leukocytes (antigenemia) (Fig. 376-3).

Seroconversion is an excellent marker for primary CMV infection, but increases in immunoglobulin G (IgG) titers, even four-fold or greater, are not diagnostic of newly acquired infection. CMV-specific IgM antibody develops during primary infection but may reappear during reactivation of latent CMV. The presence of IgG antibody is a sensitive marker of past infection and is used to screen transplant recipients and donors as well as certain blood product recipients and donors. An avidity assay can help in determining whether an infection is recent.

Viral culture, which was the prior "gold standard" for diagnosing CMV, has been supplanted by the assays described previously. Routine culture may require at least 4 to 6 weeks, whereas newer methods incorporating immunofluorescence can yield results in several days. The clinical significance of positive CMV cultures may be difficult to ascertain, particularly in immunosuppressed patients. For example, CMV may be present in the saliva or urine of up to 60 to 90% of transplant recipients and patients with AIDS, and virus in these sites does not prove that CMV is the cause of a patient's illness. Cytologic and histologic abnormalities are not sensitive measures of CMV infection, but they are specific and indicative of CMV disease.

PREVENTION

A nonviable CMV vaccine, containing the gB antigen, decreases primary infection in young women by 50% but is not commercially available. Other vaccine candidates are being developed but remain experimental.

Because CMV is transmitted by exchange of secretions or excretions, infection can be diminished by reducing exposure to body fluids. For example, transmission by both vaginal and anal intercourse, which are bidirectional risks, can be diminished by "safe sex." Similarly, limiting the contact of seronegative pregnant women with the secretions and excretions of children, especially preschoolers in daycare, can decrease primary infection and, in turn, congenital disease.

The risk for acquiring CMV disease can be reduced in seronegative, immunosuppressed patients through the use of blood products or organ grafts from CMV-seronegative donors. Valganciclovir and ganciclovir provide effective prophylaxis in solid organ transplantation.[4,5] Prophylaxis has been uncommon for stem cell transplant recipients, and these patients typically have been monitored weekly (from day 10 to day 100 after transplantation) for CMV DNA or antigenemia, with antiviral therapy introduced preemptively if seroconversion occurs or PCR testing is positive.[4] With this strategy, infection is not prevented, but end-organ disease is avoided. However, late-onset CMV end-organ disease may occur when the antiviral is eventually discontinued.[6] More recently, letermovir, a CMV terminase substitute, at a dose of 120 mg or 240 mg daily has been shown to reduce the incidence of CMV infection in such patients.[4] CMV-specific hyperimmune globulin (CMVIG) may be effective in treating pregnant women with primary CMV infection and preventing congenital infection, but controlled trials are needed. CMVIG has also been used prophylactically in high-risk seronegative organ transplant recipients, in whom it reduced CMV disease and CMV mortality. It is very expensive, however, and antivirals are more commonly used alternatives.

TREATMENT

Rx

Oral valganciclovir, intravenous ganciclovir, intravenous ganciclovir followed by oral valganciclovir, intravenous foscarnet, intravenous cidofovir, and ganciclovir intraocular injection coupled with valganciclovir are all effective treatments of CMV syndromes (Table 376-1).

Ganciclovir (dihydroxypropoxymethylguanosine [DHPG], Cytovene) is given intravenously, 5 mg/kg two times daily during initial induction (2 to 3 weeks); maintenance therapy consists of 5 mg/kg once daily (Chapter 360). Valganciclovir achieves levels comparable to intravenous ganciclovir at 5 mg/kg when it is given orally in a 900-mg dose. Initial response in retinitis (improvement or stabilization of vision or ophthalmoscopic appearance) occurs in approximately 75% of patients treated with ganciclovir or valganciclovir.[4] CMV retinitis can also be treated locally by intraocular ganciclovir injection, but this approach should be accompanied by valganciclovir to treat and/or prevent extraocular end-organ disease. Ganciclovir together with CMV hyperimmune globulin may reduce the mortality of CMV pneumonia after stem cell transplantation from approximately 85 to 40%, although the benefit of the antibody is unproved. Ganciclovir resistance may occur as a result of mutations in the phosphorylating gene (*UL97*) and/or in the DNA polymerase gene (*UL54*).[7] Granulocyte colony-stimulating factor may be needed to offset neutropenia.

Foscarnet, or phosphonoformic acid, blocks the pyrophosphate-binding site of viral DNA polymerase, thereby preventing cleavage of pyrophosphate from deoxyadenosine triphosphate. The recommended initial therapy with foscarnet is 60 mg/kg intravenously every 8 hours or 90 mg/kg every 12 hours. The maintenance dose ranges from 90 to 120 mg/kg daily. Adverse effects include renal impairment, anemia, hypocalcemia (especially ionized calcium), hypomagnesemia, and hypophosphatemia. Resistance to foscarnet can develop because of mutations in DNA polymerase. Although it is effective for treating CMV retinitis, its toxicity and the absence of an oral formulation make foscarnet a second choice for treatment of CMV disease. It is sometimes used in combination with ganciclovir for infections such as central nervous system disease or for treatment of ganciclovir-resistant virus.[8]

Cidofovir, or 3-hydroxy-2-phosphonomethoxypropyl cytosine (HPMPC), appears to the cell as a nucleotide and does not require phosphorylation by virus-encoded enzyme. It is therefore active against ganciclovir-resistant CMV strains that have resistance mutations only in *UL97,* the phosphorylating gene. When DNA polymerase *(UL54)* mutations occur in ganciclovir-treated patients, cross-resistance to cidofovir is frequent. These resistance mutations also occur in patients treated with cidofovir alone. The drug has an extremely long half-life that permits intravenous administration as infrequently as every 2 weeks during maintenance treatment.

Cidofovir is nephrotoxic, especially to the proximal renal tubule, but this side effect appears to be diminished by prehydration and concomitant probenecid therapy. Cidofovir toxicities make it a second- or third-line agent for CMV. Other drugs under investigation include brincidofovir, maribavir, and letermovir.[9]

TABLE 376-1 TREATMENT OF CYTOMEGALOVIRUS INFECTION

	PREFERRED THERAPY	ALTERNATIVE THERAPY
Cytomegalovirus (CMV) retinitis*		
Sight-threatening lesions	Valganciclovir 900 mg bid PO plus ganciclovir intraocular injection	Ganciclovir IV; foscarnet IV; plus ganciclovir intraocular injection
Peripheral lesions	Valganciclovir 900 mg bid PO	Ganciclovir IV or foscarnet IV
Maintenance therapy	Valganciclovir 900 mg od PO	Ganciclovir IV or foscarnet IV
Relapsing	Reinduction with ganciclovir IV or valganciclovir 900 mg bid PO ± ganciclovir intraocular injection	
Ganciclovir resistant	Foscarnet IV ± ganciclovir intraocular injection	Cidofovir (if only *UL97* mutation)
CMV gastrointestinal disease	Ganciclovir IV for 3-6 wk or valganciclovir 900 mg bid PO for 3-6 wk	Foscarnet IV for 3-6 wk
CMV neurologic disease	Ganciclovir IV + foscarnet IV	
CMV viremia syndrome	Valganciclovir 900 mg bid PO or ganciclovir IV until viremia clears	Foscarnet IV
Ganciclovir resistant	Foscarnet IV	

*If not already begun, antiretroviral therapy should be initiated concurrently with anti-CMV therapy, except possibly when treating central nervous system disease. For retinitis, anti-CMV therapy should be continued until the CD4 count has exceeded 100-150 cells/mm^3 for ≥6 months and the retinitis is inactive. If anti-CMV therapy is discontinued, regular monthly eye examinations should be continued. Early relapses of CMV retinitis in patients treated systemically are usually due to inadequate drug penetration, and reinduction with the same drug is often effective. Drug resistance may occur in patients treated for ≥3 months. Therapy of these patients may be guided by antiviral susceptibility testing.

Adapted from Drew WL, Erlich KS. Management of herpesvirus infections (cytomegalovirus, herpes simplex virus, and varicella-zoster virus). In: Volberding PA, Greene WC , Lange J, et al, eds. *HIV/AIDS Medicine Medical Management of AIDS 2012*. Philadelphia: Saunders Elsevier; 2012:433.

PROGNOSIS

In immunocompetent patients, the mononucleosis-like CMV syndrome resolves spontaneously. Infections in immunocompromised patients are much more serious and may result in failure of a transplanted solid organ and/or systemic CMV disease. For CMV pneumonia, death often occurs even with antiviral therapy, especially after stem cell transplantation. In AIDS patients, CMV infection generally resolves when CD4 counts exceed 100/mm^3, but it is a grave prognostic sign if counts do not recover to those levels.

Grade A References

A1. Hodson EM, Ladhani M, Webster AC, et al. Antiviral medications for preventing cytomegalovirus disease in solid organ transplant recipients. *Cochrane Database Syst Rev*. 2013;2:CD003774.

A2. Owers DS, Webster AC, Strippoli GF, et al. Pre-emptive treatment for cytomegalovirus viraemia to prevent cytomegalovirus disease in solid organ transplant recipients. *Cochrane Database Syst Rev*. 2013;2:CD005133.

A3. Boeckh M, Nichols WG, Chemaly RF, et al. Valganciclovir for the prevention of complications of late cytomegalovirus infection after allogeneic hematopoietic cell transplantation: a randomized trial. *Ann Intern Med*. 2015;162:1-10.

A4. Chemaly RF, Ullmann AJ, Stoelben S, et al. Letermovir for cytomegalovirus prophylaxis in hematopoietic-cell transplantation. *N Engl J Med*. 2014;370:1781-1789.

A5. Martin DF, Sierra-Madero J, Walmsley S, et al. A controlled trial of valganciclovir as induction therapy for cytomegalovirus retinitis. *N Engl J Med*. 2002;346:1119-1126.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Cannon MJ, Schmid DS, Hyde TB. Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. *Rev Med Virol*. 2010;20:202-213.
2. Stowell JD, Forlin-Passoni D, Din E, et al. Cytomegalovirus survival on common environmental surfaces: opportunities for viral transmission. *J Infect Dis*. 2012;205:211-214.
3. Boeckh M, Geballe AP. Cytomegalovirus: pathogen, paradigm, and puzzle. *J Clin Invest*. 2011;121:1673-1680.
4. Townsend CL, Forsgren M, Ahlfors K, et al. Long-term outcomes of congenital cytomegalovirus infection in Sweden and the United Kingdom. *Clin Infect Dis*. 2013;56:1232-1239.
5. Meije Y, Fortun J, Len O, et al. Prevention strategies for cytomegalovirus disease and long-term outcomes in the high-risk transplant patient (D+/R-): experience from the RESITRA-REIPI cohort. *Transpl Infect Dis*. 2014;16:387-396.
6. Eid AJ, Razonable RR. New developments in the management of cytomegalovirus infection after solid organ transplantation. *Drugs*. 2010;70:965-981.
7. Drew WL. Cytomegalovirus resistance testing: pitfalls and problems for the clinician. *Clin Infect Dis*. 2010;50:733-736.
8. Drew WL, Liu C. Repopulation of ganciclovir-resistant cytomegalovirus by wild-type virus. *Clin Transplant*. 2012;26:949-952.
9. Griffiths P, Lumley S. Cytomegalovirus. *Curr Opin Infect Dis*. 2014;27:554-559.

REVIEW QUESTIONS

1. Which of the following is the best diagnostic test to document clinically important cytomegalovirus (CMV) infection?

A. Tissue culture of blood
B. Tissue culture of urine
C. Cytology of urine
D. CMV-specific immunoglobulin M (IgM) antibody in blood
E. Quantitative polymerase chain reaction (PCR) for CMV DNA

Answer: E Tissue culture of blood is very insensitive compared with either antigenemia or DNAemia. Urine may remain positive by any assay for months to years after primary infection and is therefore not a specific marker of clinically important CMV disease. Cytology of urine is an insensitive assay. IgM antibody may be undetectable during infections associated with CMV reactivation or may be detectable in the absence of CMV disease. CMV PCR is a sensitive assay for CMV viremia, which usually is positive in clinically significant CMV infection of any organ.

2. CMV retinitis is best diagnosed by which of the following?

A. PCR of vitreous fluid
B. Tissue culture of blood
C. Ophthalmologic examination
D. Retinal biopsy
E. CMV PCR on blood

Answer: C PCR of vitreous fluid or retinal biopsy requires a specimen obtained by a highly skilled operator and is not readily available. Tissue culture of blood is an insensitive assay for active CMV infection. CMV PCR on blood is a sensitive assay for CMV viremia but does not identify specific end organ disease.

3. Which is the best choice for treating systemic CMV infection?

A. Oral valganciclovir
B. Cidofovir
C. Foscarnet
D. Intravenous valganciclovir
E. Oral ganciclovir

Answer: A Cidofovir and foscarnet are more toxic than valganciclovir and are thus not first-line therapy. Valganciclovir is not available in an intravenous form, and oral ganciclovir has been supplanted by oral valganciclovir and is no longer commercially available in many countries.

4. Which of the following is true?

A. Nearly 90% of U.S. adults are infected with CMV by age 25 years.
B. Primary CMV infection is usually asymptomatic.
C. Aerosols are the primary route of CMV spread in the population.
D. All CMV-seronegative patients should receive blood transfusions only from CMV-seronegative donors.
E. Retinitis is the most frequent manifestation of CMV infection in transplant recipients.

Answer: B Approximately 50% of U.S. adults are CMV-seropositive by age 35 years. The primary means of spread is by direct contact with bodily secretions. CMV-seronegative blood is preferable for immunocompromised or pregnant seronegative recipients. In developed countries, little, if any, CMV is transmitted by blood transfusion. Retinitis is a rare manifestation of CMV disease in transplant recipients, in marked contrast to untreated patients with AIDS.

5. Which of the following is true for stem cell transplant (SCT) recipients?

A. Only CMV-seronegative recipients need to be monitored for CMV viremia.
B. Valganciclovir is the best prophylactic medication.
C. Hyperimmune CMV immunoglobulin should be given to all recipients before transplantation.
D. Three months of therapy with valganciclovir prevents 95% of late-onset CMV disease.
E. None of the above

Answer: E *All* STC recipients need to be monitored for the development of CMV viremia, irrespective of CMV serostatus. CMV disease occurs more commonly by reactivation of latent virus in a seropositive patient than primary infection from a seropositive donor to a seronegative recipient. Valganciclovir prophylaxis for SCT recipients should be used only if the patient cannot be monitored for the development of CMV viremia. Many SCT patients will not become viremic and would not have required prophylactic drug with the attendant toxicities, especially to bone marrow. Hyperimmune CMV immunoglobulin is very expensive and not superior to monitoring SCT patients for CMV viremia and treating when blood assays are positive. "Late-onset" CMV viremia and disease is a frequent event after valganciclovir treatment has been discontinued.

377

EPSTEIN-BARR VIRUS INFECTION

ROBERT T. SCHOOLEY

DEFINITION

Epstein-Barr virus (EBV), a member of the gamma human herpesvirus family, is the etiologic agent of infectious mononucleosis and of a diverse assortment of neoplastic syndromes.

EPIDEMIOLOGY

Ubiquitous in the human population, EBV is found in 90 to 95% of adults throughout the world. As in the case of other herpesviruses, infection with EBV is lifelong. The virus resides in B lymphocytes and is intermittently shed asymptomatically in oropharyngeal secretions, which accounts for the bulk of its transmission in the human population. The virus is not highly contagious; it is usually acquired during early childhood through sharing of saliva-bearing fomites or during adolescence through kissing, although it can be acquired at any decade of life. Thus, for example, EBV-seronegative platonic roommates of patients with acute infectious mononucleosis in college dormitory settings are not at higher risk for acquiring the virus than others in the college population. In addition to oropharyngeal spread, the virus can be transmitted by blood transfusion or through organ donation.

Most childhood EBV infections are clinically silent, but infection of adolescents and adults results in the clinical syndrome of infectious mononucleosis between 25 and 50% of the time, depending on the setting. The incidence of infectious mononucleosis is highest in the 15- to 24-year-old age group. Incidence rates in men and women are equal, but the peak incidence is 2 years earlier in women than in men. Incidence rates are lower in lower socioeconomic populations, in whom the likelihood of acquisition is greater in childhood than in adolescence.

PATHOBIOLOGY

EBV enters B lymphocytes through the CD21 molecule (also known as the C3d receptor) on the surface of B cells or nasopharyngeal epithelial cells. Major histocompatibility complex class II molecules serve as secondary receptors on B cells. Once inside the cell, the virus expresses several nuclear proteins (termed Epstein-Barr nuclear antigens [EBNAs]) that activate EBV-encoded latent membrane proteins and other gene products responsible for regulation of B-cell growth. These events are associated with the transformation or immortalization of the B cell that is the phenotypic hallmark of B-cell infection. EBV-transformed B cells proliferate vigorously and maintain EBV DNA within progeny cell nuclei in an episomal state. During acute EBV infection, up to 20% of peripheral blood B cells express EBNA.

The host response to acute EBV infection consists of a vigorous and coordinated cellular and humoral immune response. The humoral immune response includes IgM and IgG antibodies directed at the viral capsid (VCA) and to EBNA, as well as *heterophile* antibodies to surface antigens of sheep red blood cells. Heterophile antibodies are useful diagnostically and are present at some point in up to 90% of cases. These antibodies are an epiphenomenon in host defense and are not cross-reactive with any known viral antigens.

The cellular immune response includes both natural killer (NK) and EBV-specific $CD4^+$ and $CD8^+$ T lymphocytes. The expansion of the $CD8^+$ subset of T lymphocytes during acute EBV infection includes a subset of large, activated cells demonstrable on standard peripheral blood smears as "atypical" lymphocytes. This vigorous cellular immune response is associated with an outpouring of cytokines, including tumor necrosis factor, interleukin-1, and interleukin-6, that are responsible for many of the symptoms and signs of infectious mononucleosis. Over a period of 4 to 6 weeks after initial evaluation in most patients, immune response mechanisms gain control of the EBV-driven B-cell proliferation, and the virus enters into a lifelong period of symbiosis with the host. The virus is asymptomatically shed approximately 15% of the time in the oropharyngeal fluids of healthy human immunodeficiency virus type 1 (HIV-1)-seronegative adolescents and adults. The shedding rate increases significantly in patients with defects in cellular immunity, such as those that occur with HIV-1 infection or immunosuppression associated with organ allografts.

CLINICAL MANIFESTATIONS

Most cases of acute EBV infection are clinically silent. The syndrome of infectious mononucleosis consists of the clinical triad of fever, sore throat (Chapter 429), and lymphadenopathy, in association with an atypical lymphocytosis and the transient appearance of heterophile antibodies. The incubation period between exposure and the onset of symptoms is generally 30 to 50 days. The onset of symptoms may be abrupt, or it may be heralded by a several-day nonspecific prodrome of malaise and low-grade fever. Although the classic syndrome includes fever, sore throat, and adenopathy, the findings may be dominated by only one or any combination of these symptoms. Other common clinical manifestations include headache, malaise, and anorexia. On physical examination, patients are usually febrile. Pharyngeal erythema, tonsillar enlargement (see Fig. 429-5 in Chapter 429), and cervical adenopathy are generally present. Mild periorbital edema may also be observed. Abdominal findings may include splenomegaly or hepatomegaly, or both. Splenomegaly can be demonstrated by ultrasonographic examination in virtually all patients with infectious mononucleosis, although palpable splenomegaly is only present in about 20% of patients. Splenic enlargement is usually maximal in the second or third week of illness and might not be detectable at the initial presentation. Adenopathy may be observed in noncervical regions, but it is usually much less prominent than in cervical regions. Approximately 5% of patients will exhibit a rash that may be macular, scarlatiniform, or urticarial in nature. If patients with acute EBV infection are given ampicillin or its derivatives, a pruritic maculopapular eruption will develop in 90 to 100% of them. Patients with an ampicillin-induced rash during acute EBV infection generally tolerate the drug and other penicillin products when administered later in life.

DIAGNOSIS

Because clinical manifestations of acute EBV infection are variable and other organisms may cause similar clinical syndromes, laboratory tools are required to confirm an etiologic diagnosis. Heterophile antibodies to sheep red blood cells are classically used to diagnosis EBV-induced infectious mononucleosis. Although ultimately demonstrable in approximately 90% of symptomatic acute EBV infections, these antibodies are present in only about two thirds of patients at initial encounter. If antibodies are negative at the outset and clinical suspicion is high, repeat testing in the second or third week of the illness is warranted. Although EBV-specific antibodies remain the "gold standard" for the diagnosis of acute EBV infection, if heterophile antibodies are demonstrated in a straightforward case of infectious mononucleosis, it is not generally necessary to order EBV-specific serologic studies. IgM antibodies to the EBV capsid antigen (VCA) are the most useful serologic study in the diagnosis of acute EBV infection. Relatively high titers of IgG antibodies to VCA persist for life after initial infection and are not useful in making the diagnosis of acute EBV infection. Antibodies to EBNA are slower to arise than those to capsid antigens, and acute infection may be diagnosed by demonstration of seroconversion to this antigen. If the diagnosis is based on the emergence of antibodies to EBNA, both tests should be performed in the same laboratory.

Among pathogens causing clinical syndromes that can be mistaken for acute EBV infection, cytomegalovirus (Chapter 376) is the most frequent. Patients with cytomegalovirus infection are less likely to have an acute onset of illness, and pharyngitis is less frequently a prominent manifestation of the illness. *Toxoplasma gondii* (Chapter 349) infection can also present as a nonspecific febrile illness that can be confused with infectious mononucleosis. Streptococcal pharyngitis (Chapter 290) and primary herpes stomatitis (Chapter 374) may occasionally cause symptoms that are mistaken for acute EBV infection. None of these syndromes is associated with heterophile antibodies or with other serologic evidence of acute EBV infection. The differential diagnosis is generally made by serologic studies directed at these organisms or by culture. Nonetheless, physicians should be cognizant that organisms such as group A β-hemolytic streptococci (Chapter 290) and herpes simplex virus are also common in the human population and may be demonstrated in people whose symptoms are nonetheless due to acute EBV infection.

PREVENTION AND TREATMENT

Because the virus is usually transmitted by asymptomatic oral shedders and is so common in the human population, epidemiologic interventions to prevent spread are not practical. No vaccine yet has been developed. The clinical course is generally self-limited and does not usually require specific therapeutic intervention beyond the use of aspirin or acetaminophen for antipyresis and mild pain relief, except in the presence of specific complications such as when lymphadenopathy threatens the airway or in certain cases of autoimmune hemolytic anemia (Chapter 160) or thrombocytopenia (Chapter 172). Short courses of corticosteroids have been used to hasten symptomatic recovery in cases in which the symptoms are severe or refractory. Corticosteroids should not, however, be used routinely and should be given for no longer than a 10- to 14-day tapering course that begins at a dose equivalent of 0.5mg/kg of prednisone. Although EBV replication can be inhibited in vitro or in vivo by acyclovir and related antiviral agents, the symptoms of infectious mononucleosis are primarily driven by the immune response to the virus and come later than the time of maximal viral replication. Antiviral agents have not been demonstrated to significantly accelerate resolution of symptoms or prevent complications of the disease. One small randomized trial has reported 1-day reduction in length of hospital stay when patients with severe infectious mononucleosis are given metronidazole, but confirmatory data are required before this therapy becomes routine.[1]

PROGNOSIS

Most patients recover uneventfully from the acute symptoms and signs of infectious mononucleosis over a 2- to 3-week period, although many patients may have a variable period of malaise and fatigue that can last for another 3 to 4 weeks. Some patients may take longer to make a full recovery and experience fatigue and difficulty concentrating for up to 6 months after diagnosis. Symptoms often wax and wane and can be extremely troublesome. Reassurance is usually the best approach to these patients. Corticosteroids are not of benefit in this setting. Recovery may be less straightforward in patients with certain specific complications of acute EBV infection (outlined in the next section). Death from infectious mononucleosis is rare. When it does occur, it is most frequently associated with neurologic complications of the illness, splenic rupture, or the X-linked lymphoproliferative syndrome (discussed later).

Complications

Although most patients recover spontaneously from acute EBV infection, a number of complications may arise. In some patients, these complications dominate the clinical findings, and seroconversion to EBV may be the only evidence of acute EBV infection. The most serious complication of acute EBV infection arises in individuals with the X-linked lymphoproliferative syndrome. This syndrome occurs in males with mutations in the signaling lymphocyte activation molecule (SLAM)-associated protein (SAP) that regulates T and natural killer cells. These otherwise healthy individuals have severe clinical symptoms, a large lymphocytosis consisting of T and B cells, and severe hepatitis. If patients survive the acute infection, the syndrome may evolve into progressive agammaglobulinemia or lymphoma in the following months. The genetic defect associated with this syndrome can be diagnosed in utero, and early bone marrow transplantation has been recommended for the prevention of the devastating clinical syndrome associated with acquired EBV infection.

A number of less severe, organ system–specific complications are seen substantially more frequently than the X-linked lymphoproliferative syndrome. Patients should specifically be warned about splenic rupture, a complication attributable to splenomegaly (Chapter 168) and associated stretching of the splenic capsule that occurs most frequently in the second or third week of the illness, when other symptoms of the disease are abating. It may be accompanied by major or minor trauma but may also occur without an obvious antecedent event. Patients should be counseled against activities that might result in abdominal trauma for 6 to 8 weeks after the onset of symptoms. Left upper quadrant pain, especially pain radiating to the subscapular region, should raise this diagnostic consideration. As with other complications of acute EBV infection, splenic rupture may occur occasionally in patients without other prominent clinical manifestations of acute EBV infection. Other hematologic complications include autoimmune hemolytic anemia (Chapter 160), thrombocytopenia (Chapter 172), and neutropenia (Chapter 167). These complications usually arise from a combination of self-reactive antibodies and hypersplenism and are generally self-limited and resolve with resolution of the illness. Corticosteroids may be of benefit in more severe cases of autoimmune hemolytic anemia or thrombocytopenia.

A variety of neurologic syndromes are unusual complications of acute EBV infection. EBV DNA has been detected in brain tissue from rare patients with clinical manifestations compatible with herpes simplex encephalitis (Chapter 414). Although these patients have a much better

prognosis than patients with herpes simplex encephalitis, they should receive parenteral acyclovir or ganciclovir as for herpes simplex (Chapters 360 and 414). Other neurologic complications include aseptic meningitis (Chapter 412), cerebellitis, mononeuritis multiplex (Chapter 420), Bell's palsy (Chapter 420), Guillain-Barré syndrome (Chapter 420), and transverse myelitis (Chapters 400 and 411). These complications may be clinically dramatic but are usually self-limited and associated with full recovery in 85% of patients without specific antiviral therapy.

Mild hepatomegaly can occur in acute infectious mononucleosis, and biochemical evidence of hepatitis (Chapter 148) should be expected in virtually every case of acute infection. More severe hepatic complications are uncommon, and renal, cardiac, pulmonary, and skeletal muscle complications are rare.

OTHER CLINICAL MANIFESTATIONS

In addition to infectious mononucleosis, EBV is also associated with neoplasia and lymphoproliferative disorders, which are seen most frequently in patients with defects in cellular immunity but are not restricted to such patients.

Post-transplantation Lymphoproliferative Disease

EBV-driven B-cell proliferation that is insufficiently regulated in the presence of prolonged periods of severe T-cell immunodeficiency may result in a polyclonal proliferation of B cells that is initially similar to that seen in acute infectious mononucleosis. Although most frequently occurring in the setting of organ transplantation, especially when patients are immunosuppressed with agents directed specifically at T lymphocytes, such as anti-CD3 antibodies or cyclosporine, this syndrome can be seen in other conditions with similar levels and durations of immunodeficiency such as HIV-1 infection. Post-transplantation lymphoproliferative disease (PTLD)[2] is seen more often when the donor is EBV seropositive and the recipient is seronegative, with the more intense immunosuppression of hematopoietic stem cell transplantation[3] (Chapter 178) in association with graft-versus-host disease, in patients who undergo splenectomy before transplantation, and in patient-donor pairs with higher degrees of HLA mismatch. These tumors are less frequently seen in the current era, in which allograft-associated immunosuppression is better targeted and less intense.

Patients with PTLD often present with fever, adenopathy, and splenomegaly. If the immunodeficiency persists, these disorders often proceed from a polyclonal stage, which can be reversed with restoration of immunity, to a monoclonal or oligoclonal stage that is progressive despite restoration of cellular immunodeficiency.

The diagnosis is not generally difficult to make in the appropriate clinical setting and can be made histopathologically. Elevated plasma levels of EBV DNA are associated with increased risk for PTLD after hematopoietic stem cell transplantation,[4] but their predictive value is less well established after solid organ transplantation.

There is some evidence that PTLD may be less frequent in patients who have received acyclovir or ganciclovir after transplantation, but these agents are less useful after the syndrome develops. Successful management depends on the extent to which the immunosuppressive condition can be reversed before the evolution of restricted clonality. Therapy with anti-CD20[5] antibodies with or without chemotherapy[6] is the treatment of choice for PTLD (Chapter 185). Radiation therapy is also used in some patients.

Burkitt's Lymphoma

EBV was initially described in patients with African Burkitt's lymphoma (Chapter 185). The tumor is composed of small, noncleaved B cells and, unless aggressively treated, is rapidly fatal. This aggressive B-cell lymphoma with a predilection for the head and neck is endemic in equatorial Africa and is geographically linked to *Plasmodium falciparum* malaria. EBV DNA is readily demonstrable in tumor biopsy specimens, and high titers of antibodies to EBV structural antigens are found in plasma. Sporadic cases of abdominal B-cell lymphomas with a histologic appearance compatible with Burkitt's lymphoma are also observed but are associated with EBV only about 25% of the time. The tumor is likewise seen in patients with HIV-1 infection. Although an etiologic role for EBV in Burkitt's lymphoma is widely accepted, the molecular basis by which EBV causes the neoplasm has not yet been fully delineated.

The risk for HIV-associated Burkitt's lymphoma increases with advancing immunodeficiency, but it may also be seen in patients with relatively preserved CD4 cell counts. Antiretroviral therapy reduces but does not eliminate the risk for Burkitt's lymphoma in HIV-1-infected persons. Despite the high-grade clinical behavior of the tumor, it should be vigorously treated because it is usually quite responsive to combination chemotherapy with or without radiation therapy (Chapter 185).

Hodgkin's Lymphoma

EBV is also associated with a subset of Hodgkin's lymphomas (Chapter 186), especially those of the lymphocyte-depleted or mixed-cellularity histologic subtypes. EBV DNA and proteins are detected in the Reed-Sternberg cells that are characteristic of Hodgkin's lymphoma. Therapy for EBV-associated Hodgkin's lymphoma is directed at the tumor's histology and stage (Chapter 186) and is not determined by whether it is related to EBV.

Central Nervous System Lymphoma

EBV is also associated with central nervous system (CNS) lymphoma (Chapter 185).[7] This tumor was most frequently observed in the post-transplantation setting before the HIV epidemic but is now the most frequent CNS neoplasm in HIV-1-infected individuals. The major differential diagnostic challenge is with *T. gondii* infection (Chapter 349). Although a tissue-based diagnosis is definitive, noninvasive neurodiagnostic approaches, coupled with the demonstration of EBV DNA in cerebrospinal fluid by polymerase chain reaction, can strongly support the diagnosis of lymphoma over that of *T. gondii* infection. Radiation therapy may be used, but its effects are generally palliative.

Nasopharyngeal Carcinoma

EBV has also been associated with certain cases of nasopharyngeal carcinoma (Chapter 190). This tumor is rare in Western countries, but it is much more frequent in southern China and in the Inuit population of Alaska. EBV-associated cases are generally less histologically differentiated than sporadic forms of nasopharyngeal carcinoma. EBV DNA is demonstrable in tumor tissue, and high titers of immunoglobulin A (IgA) and IgG antibodies to the EBV capsid antigens are found in plasma. The prognosis for this tumor is poor, although it is often treated with radiation therapy (Chapter 190).

Other EBV-Associated Neoplasms

EBV collaborates with another human gamma-herpesvirus, human herpesvirus type 8, to cause lymphoma in the HIV-1-infected population. These aggressive tumors present in body cavities such as the pleural, peritoneal, and pericardial spaces.

EBV DNA may also be found in moderate to slowly progressive destructive midline facial angiocentric tumors of T- and NK-cell phenotypes. This neoplasm presents clinically as a syndrome that was previously known as lethal midline granuloma (Chapter 185). EBV also plays a key role in the pathogenesis of an angiocentric EBV-associated B-cell tumor that presents clinically as lymphomatoid granulomatosis (Chapter 270).

Oral Hairy Leukoplakia

This clinical manifestation of EBV infection is characterized by a corrugated or "hairy" plaquelike lesion that extends around the lateral aspects of the tongue (Chapter 425). Oral hairy leukoplakia is most often observed in individuals with chronic forms of cellular immunodeficiency, especially those with HIV-1 infection and CD4 cell counts less than 200/μL. It is most often clinically confused with mucocutaneous candidiasis (Chapter 338) but can be differentiated because its distribution is restricted to the lateral surface of the tongue. Unlike thrush, it does not involve the buccal mucosa, palate, or pharynx and is not readily removed by superficial scraping. Biopsies demonstrate a characteristic histopathologic pattern, as well as the presence of EBV antigens and DNA within squamous epithelial cells. Although the lesions may be cosmetically troublesome, they are not generally painful. In the case of HIV-1-associated oral hairy leukoplakia, the lesions resolve with successful antiretroviral chemotherapy. If the immunosuppression cannot be reversed, oral hairy leukoplakia usually responds to valacyclovir, valganciclovir, or foscarnet (Chapter 425).

Chronic Active EBV Infection

Infrequent patients with no apparent defect in cellular immunity have been described in which chronic EBV infection has been associated with persistent or intermittent hepatitis or interstitial pulmonary disease (or both). These rare patients with bona fide organ system disease should not be confused with patients who have chronic fatigue syndrome or fibromyalgia rheumatica (Chapter 274), which have no relationship with EBV.

Grade A Reference

A1. Candy B, Hotopf M. Steroids for symptom control in infectious mononucleosis. *Cochrane Database Syst Rev*. 2006;3:CD004402.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Lennon P, O'Neill JP, Fenton JE. Effect of metronidazole versus standard care on length of stay of patients admitted with severe infectious mononucleosis: a randomized controlled trial. *Clin Microbiol Infect*. 2014;20:O450-O452.
2. Green M, Michaels MG. Epstein-Barr virus infection and posttransplant lymphoproliferative disorder. *Am J Transplant*. 2013;13(suppl 3):41-54.
3. Uhlin M, Wikell H, Sundin M, et al. Risk factors for Epstein-Barr virus-related post-transplant lymphoproliferative disease after allogeneic hematopoietic stem cell transplantation. *Haematologica*. 2014;99:346-352.
4. Liu Q, Xuan L, Liu H, et al. Molecular monitoring and stepwise preemptive therapy for Epstein-Barr virus viremia after allogeneic stem cell transplantation. *Am J Hematol*. 2013;88:550-555.
5. Styczynski J, Gil L, Tridello G, et al. Response to rituximab-based therapy and risk factor analysis in Epstein Barr virus-related lymphoproliferative disorder after hematopoietic stem cell transplant in children and adults: a study from the Infectious Diseases Working Party of the European Group for Blood and Marrow Transplantation. *Clin Infect Dis*. 2013;57:794-802.
6. Zimmermann H, Trappe RU. EBV and posttransplantation lymphoproliferative disease: what to do? *Hematology Am Soc Hematol Educ Program*. 2013;2013:95-102.
7. Tselis AC. Epstein-Barr virus infections of the nervous system. *Handb Clin Neurol*. 2014;123:285-305.

378 RETROVIRUSES OTHER THAN HUMAN IMMUNODEFICIENCY VIRUS

WILLIAM A. BLATTNER

TABLE 378-1 TRANSMISSION OF HTLV-1 AND HTLV-2

MODE OF TRANSMISSION	HTLV-1	HTLV-2
MOTHER TO INFANT		
Transplacental	Yes	Not known
Breast milk	Yes	Probable
SEXUAL		
Male to female	Yes	Yes
Female to male	Yes	Yes
Male to male	Yes	Not known
PARENTERAL		
Blood transfusion	Yes	Yes
Intravenous drug use	Yes	Yes
COFACTORS		
Ulcerative genital lesions	Yes	Not known
Cellular transfusion products	Yes	Yes
Sharing of "works"*	Yes	Yes
ELEVATED VIRUS LOAD		
Mother to infant	Yes	Not known
Heterosexual	Yes	Not known

*Intravenous paraphernalia, such as needles.

DEFINITION

There are now four members of the human T-lymphotropic virus (HTLV) family: HTLV-1, discovered in 1979; HTLV-2, discovered in 1982; and HTLV-3 and HTLV-4, discovered in 2005. HTLV-1 has been causally linked to adult T-cell leukemia/lymphoma (ATL) and to several chronic degenerative conditions, most notably HTLV-1–associated myelopathy/tropical spastic paraparesis (HAM/TSP), whereas disease associated with HTLV-2 is rare, and no disease associations have been established with HTLV-3 or HTLV-4.

The Pathogens

Within the taxa of RNA reverse transcribing viruses, the HTLV viruses, along with bovine leukemia virus, are classified in the subfamily Retroviridae within the genus Deltaretrovirus (formerly termed *oncovirus*). The oncogenic properties of these viruses and their molecular structure distinguish them from the human immunodeficiency retroviruses HIV-1 and HIV-2 (Chapter 386), which are members of the genus Lentivirus. Both deltaretroviruses and lentiviruses are capable of prolonged asymptomatic infection. In vitro, however, HIV-1 and HIV-2 have cytopathic effects on human T cells, whereas HTLV-1 and HTLV-2 are capable of transforming T cells into immortalized cell lines. The HTLVs are diploid single-stranded RNA viruses that replicate through cDNA, a proviral intermediate, by reverse transcriptase, a viral polymerase.

EPIDEMIOLOGY

HTLV-1 is widely disseminated worldwide and is estimated to infect 10 to 25 million persons, with the aggressive T-cell malignant neoplasm ATL developing in 2 to 6% and chronic inflammatory diseases, mainly HAM/TSP, developing in another 1 to 5% in their lifetime.[1] Similar to HIV, molecular epidemiology suggests that the four major subtypes of HTLV identified in humans arose from separate interspecies transmission from simians to humans. The discovery of HTLV-3 and HTLV-4 was made in Cameroon, where closely related viruses in nonhuman primates led to discovery in humans with exposure as bush meat hunters. Related to such interspecies transmission, there are four major geographic subtypes of HTLV-1: Cosmopolitan subtype A, Central African subtype B, Australio-Melanesian (Papua New Guinea, Melanesia, and Australian aborigines) subtype C, and Central African/Pygmies subtype D. Central Africa also carries a few rare subtypes (E, F, G). Within the cosmopolitan group are four subgroups: transcontinental, Japanese, West African, and North African. The virus from Australio-Melanesia differs molecularly from the Japanese and African strains by 5 to 10%, the result of independent evolution of the virus in these populations separated for tens of thousands of years. The stability of HTLV-1 in comparison to HIV-1 reflects the observation that HTLV favors viral expansion through proliferation of proviral DNA-harboring cells rather than infection of new cells by cell-free virions. The HTLV subtypes differ phylogenetically by approximately 30 to 40% among each other.

HTLV-1 is not universally present in all human populations but rather clusters geographically: southern Japan; Melanesia; Australia, in aboriginal peoples; West Africa and, by the slave trade from Africa, the Caribbean and the United States in African Americans; Central and South America; and the Mashhad region of Iran. In the United States, HTLV-1 infection is often found in persons who migrate from these regions. HTLV-2 is found in Native American people throughout North, Central, and South America and in West Africa. Most HTLV-2 infections in the United States and Europe occur in injection drug users, in whom the virus is spread by needle sharing and other injection practices. HTLV-3 and HTLV-4 were originally detected in Cameroon, but their extent in West Africa is not yet known.

Routes of Transmission

HTLV, like HIV-1, is transmitted sexually, perinatally, and by transfusion or injection drug use (Table 378-1).

Sexual Transmission

Sexual transmission of HTLV-1 from male to female and from female to male as well as from male to male has been documented. HTLV-1 transmission is cell associated and appears to be at least an order of magnitude less infectious than HIV-1. Coincidental infection with other sexually transmitted diseases, particularly those associated with ulcerative and inflammatory genital lesions, amplifies the risk of transmission. For HTLV-1, elevated viral load is linked to heightened transmission. In regions endemic for the virus, there is a characteristic age-dependent rise in HTLV-1 seroprevalence. This increase first becomes evident in the adolescent years; it is steeper in women than in men and continues in women after 40 years of age, whereas rates in men plateau around the age of 40 years. This pattern reflects more efficient male-to-female transmission. For HTLV-2, the rates for both genders are equal, thus suggesting that there may be differences in the kinetics of transmission between the two viruses.

Perinatal Transmission

For HTLV-1, transmission through breast-feeding is more efficient than in utero or perinatal transmission. Major risk factors that increase the efficiency of transmission include high proviral loads and increased duration of breast-feeding (>6 months). On average, 20% of infants breast-fed by HTLV-1–positive mothers seroconvert to HTLV-1, whereas only 1 to 2% of bottle-fed infants of HTLV-1–positive mothers become infected. In contrast, in utero and perinatal transmission accounts for virtually all HIV-1 transmission in the West, and breast-feeding accounts for an additional 15 to 20% of infant HIV infection in Africa. This difference may reflect the fact that maternal antibody to HTLV-1 transmitted across the placenta appears to neutralize perinatal HTLV-1 but not the highly mutable HIV-1. HTLV-2 is detectable in breast milk and, similar to HTLV-1, accounts for many childhood infections.

Transfusion and Injection Drug Use

Parenteral transmission, through either transfusion or injection drug use, is a major source of HTLV infection. Among blood donors in the United States, more than half of HTLV infections are due to HTLV-2. Among injection drug users, most infections are due to HTLV-2, and HTLV-2 is more efficiently transmitted by this route than HTLV-1 is.

Both HTLV-1 and HTLV-2 are transmitted in association with cellular components through a "virologic synapse,"[2] unlike HIV-1, which is transmitted by cells, plasma, or plasma products. Approximately 50% of the recipients of HTLV-1/HTLV-2–positive blood seroconvert, compared with more than 95% for HIV-1.

The only documented illness linked to HTLV-1 or HTLV-2 transfusion-associated transmission is the HTLV-associated demyelinating neurologic syndrome HAM/TSP. Leukemia has not been associated with transfusion of HTLV-positive blood. Among U.S. blood donors who are confirmed to be HTLV positive (slightly less than half are HTLV-1 positive and the others are HTLV-2 positive), the major risk factors are intravenous drug use, birthplace in an area in the Caribbean or Japan endemic for the virus, and sexual contact with a person with this profile.

Coinfection with HTLV-1 and HIV-1 appears to increase the progression to acquired immunodeficiency syndrome (AIDS) through unexplained mechanisms, possibly related to the cell-proliferative effects of HTLV-1 on HIV-1–infected T cells. Such a relationship has not been shown for HTLV-2. Other modes of transmission involving "casual contact" are not a source of infection. Health care and laboratory workers who experience a needlestick or skin or mucous membrane exposure in the absence of protective barriers have little or no risk for infection but should be monitored.

PATHOBIOLOGY

Virology

The HTLV viruses, which are single-stranded RNA viruses that contain a diploid genome, replicate through a DNA intermediary that integrates into the genome of the target T cell as a provirus, thereby resulting in lifelong infection. HTLV-1 is approximately 100 nm in diameter and has a thin, electron-dense outer envelope and an electron-dense, roughly spherical core. The total provirus genome contains 9032 nucleotides with two identical sequences termed *long terminal repeats* (LTRs) at the 5′ and 3′ ends of the genome, which contain regulatory elements that control virus expression and virion production. The retroviral structural genes (*gag* and *pol*) code for large overlapping polyproteins that are later processed into functional peptide products by virally encoded protease and cellular proteases. The encoding genes of the virus are *gag* (group-specific antigen), *pol* (polymerase/integrase/protease), and *env* (envelope), and it has a series of regulatory genes, *tax* and *rex*, and several smaller gene products that regulate infection and virus expression. A newly discovered viral gene is the basic leucine zipper factor (*HBZ*) gene, encoded by the minus strand of the HTLV-1 provirus and transcribed from the 3′ LTR. Tax protein plays a central role in enhancing the transcription of viral and cellular gene products that promote viral replication and transformation of human T lymphocytes. Through binding to the LTR, Tax promotes transcriptional activation of the viral genome, and by binding to key regulatory proteins of the NF-κB signaling pathway, it promotes cell activation and disease pathogenesis. Through binding to regulatory enhancers of the cell and through abrogation of key suppressor genes, Tax initiates the immortalization of the infected T cells. However, the Tax gene is susceptible to genetic mutations and is expressed in only about 60% of ATL cases. Rex stabilizes viral mRNA, essential for export of full-length Gag/Pol and single-spliced Env mRNA from the nucleus to the cytoplasm. *HBZ*, which is the only viral gene consistently expressed in all ATL patients, appears to be essential for persistent HTLV-1 infection.[3] HBZ inhibits the Tax-mediated activation of viral gene transcription through the 5′ LTR, which ultimately represses expression of viral proteins while simultaneously promoting the proliferation and survival of infected cells, regardless of Tax expression. An estimated 500 to 5000 HTLV clones exist in persons with asymptomatic infection, with a large proportion of clones detectable during years of follow-up.[4]

Viral Life Cycle

The initial stage of HTLV infection engages several mechanisms: *viral synapse*, cell-to-cell contact; *cellular conduits*, transient membrane extensions; *extracellular viral assembly*, membrane-bound virus; and *transinfection through dendritic cells*, virus captured on the cell surface of dendritic cells. After infection, the viral life cycle involves membrane fusion, followed by reverse transcription from an RNA template to a circular DNA provirus that is transported to the cell nucleus and integrated into the host genome. After the cell is infected, viral expansion is primarily achieved through proliferation of proviral DNA-harboring cells rather than through repeated cycles of cell-to-cell infection. For viral entry, HTLV-1 uses three distinct molecules for infection of activated $CD4^+$ cells: heparin sulfate proteoglycans, neuropilin 1, and glucose transporter 1. HTLV-2 uses neuropilin 1 and glucose transporter 1 for entry into activated $CD8^+$ T cells. After uptake and uncoating, viral RNA is transcribed by reverse transcriptase, an RNA-dependent DNA polymerase complexed to the RNA in the core of the virus particle, into double-stranded DNA. This double-stranded viral DNA is integrated into the host cell nucleus by the virally encoded integrase, which results in lifelong cell infection. The viral LTR elements are essential for integration and regulation of viral genome expression, which is controlled mainly by Tax and HBZ.

Pathogenesis of Adult T-Cell Leukemia

HTLV-1 is integrally involved in the pathogenesis of ATL through a multistage process that includes clonal integration of virus (sometimes including partial viral sequences that always include *HBZ*) into active cellular gene sequences. Interestingly, the *HBZ* gene is always present in ATL, and its quantification may be a useful marker for monitoring of the response to treatment. The clonal pattern of integration indicates that ATL is derived from a single transformed tumor cell that evolved from a virus infection *before* transformation rather than afterward as a passenger virus. Tax is an oncoprotein that interacts with numerous cellular proteins to reprogram cellular processes to alter transcription, cell cycle regulation, DNA repair, and apoptosis, thereby allowing cells with potential carcinogenic mutations to survive and to escape cell death. HBZ is also emerging as a key oncogenic protein involved in modulating cellular pathways related to cell growth, immune response, and T-cell differentiation.

In some healthy carriers, T-cell polyclonal and oligoclonal proliferations develop that can later progress to malignant transformation or may disappear spontaneously. Morphologically distinct "flower cells" (Fig. 378-1), which represent T cells with deeply lobulated nuclei resembling ATL leukemic cells, are seen on peripheral blood smears of healthy carriers, and increased numbers are detected in persons with higher HTLV viral loads.

Pathogenesis of Myelopathy/Spastic Paraparesis

Viral overproduction, as measured by high viral loads, appears to result from defective host immune responses characterized by very high levels of cytotoxic T cells. Local pathologic changes in neuronal tissue may result from immune-mediated damage caused by misdirected responses to molecular mimics of viral proteins or by local damage due to cytokine-induced damage of neuronal tissue as HTLV-1 cells infiltrate the perivascular space and the neurons of the central nervous system (CNS), particularly the spinal cord.

ADULT T-CELL LEUKEMIA/LYMPHOMA

EPIDEMIOLOGY AND PATHOBIOLOGY

The cumulative lifetime incidence of adult T-cell leukemia/lymphoma (ATL) in persons infected with HTLV-1 is between 2 and 6%, so about 2500 to 5000 cases per year occur in the approximately 10 to 25 million infected persons worldwide. The latency period is approximately 20 to 30 years after infection, with a slightly higher risk among HTLV-infected males. The age-adjusted incidence rate of ATL in the United States is 0.05 case for men and 0.03 for women per 100,000 people.[5] In areas endemic for HTLV-1, such as southern Japan and the Caribbean Islands, ATL accounts for half or more of adult lymphoid malignant neoplasms. ATL is rarely seen in children, but in one series of pediatric cases of ATL, four of the eight patients shared a homozygous deletion in the p16 gene locus, and deletion of exons 7 and 8 of p53 was detected in another child, thereby suggesting that a genetic predisposition interacts with viral infection to accelerate the progression of the disease.

CLINICAL MANIFESTATIONS

The most common disease caused by HTLV-1 is ATL, a type of T/natural killer–cell lymphoma in the new World Health Organization classification (Table 378-2). ATL is a high-grade lymphoma (Chapter 185), usually of large, medium, or pleiotropic morphology (or combined morphology) and advanced clinical stage.

The acute form of ATL, which accounts for about 55% of cases, is characterized by an aggressive, mature T-cell lymphoma that presents with a high

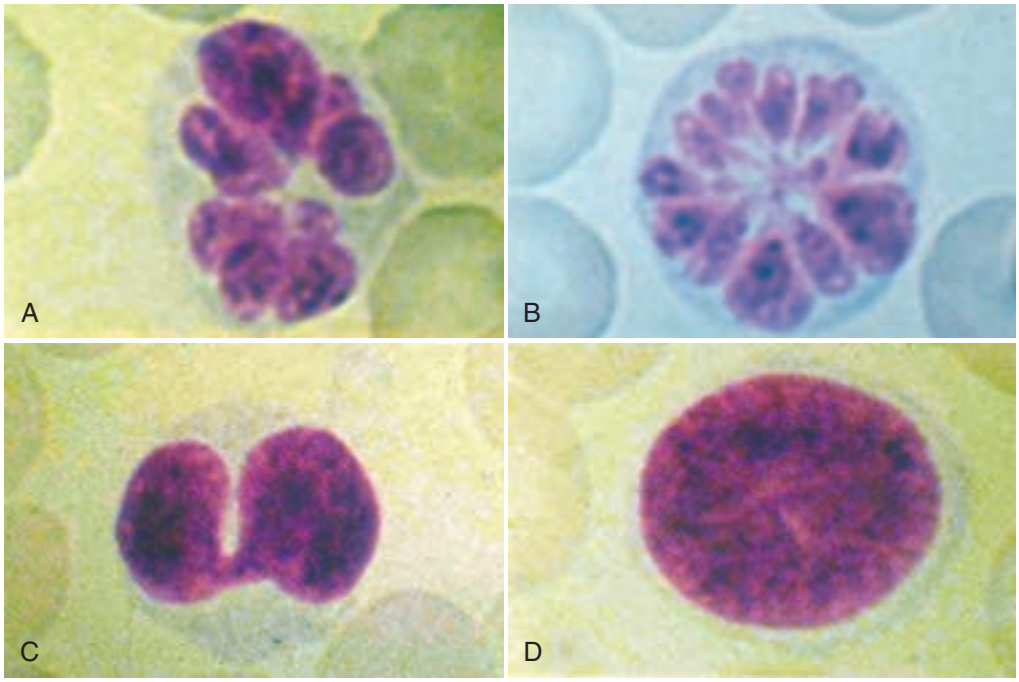

FIGURE 378-1. **Photomicrographs demonstrating the morphologic features of leukemic cells observed in different subtypes of adult T-cell leukemia/lymphoma (ATL). A and B, Polylobulated morphology of the acute type, with the highly characteristic "flower cell" shown in B. C, Typical cleaved cell seen in chronic-type ATL. D, Typical morphology of smoldering ATL. (Courtesy K. Yamaguchi and K. Takatsuki.)**

TABLE 378-2 HTLV-ASSOCIATED DISEASES

DIAGNOSIS	NATURE OF SYNDROME	STRENGTH OF ASSOCIATION
HTLV-1–ASSOCIATED DISEASES		
Adult T-cell leukemia/lymphoma	Aggressive lymphoproliferative malignant disease of mature T lymphocytes	Strong
HTLV-associated myelopathy/tropical spastic paraparesis (HAM/TSP)	Chronic progressive demyelinating syndrome of long motor tracts of spinal cord	Strong
Polymyositis	Degenerative inflammatory syndrome of skeletal muscles	Probable
Sporadic inclusion body myositis	Recently described HTLV-associated inflammatory muscle disease	Possible
Infective dermatitis	Chronic generalized eczema in adults and children; potential for pre-leukemia and immunodeficiency	Strong
Uveitis	Inflammatory infiltration of the uvea of the eye	Strong
Sjögren syndrome/keratoconjunctivitis sicca	Loss of tear production and dry eyes and dry mouth	Probable
Pulmonary lymphocyte alveolitis/ cryptogenic fibrosing alveolitis	Pulmonary infiltrate involving T lymphocytosis in lungs of patients with HAM/TSP and HTLV uveitis	Possible
HTLV-associated arthritis	Large-joint polyarthropathy; rheumatoid factor positive, with HTLV-1–positive cells infiltrating the synovia	Probable
Immunodeficiency	Subclinical (e.g., decreased PPD response) or clinical (e.g., association with clinical tuberculosis and poor response to therapy for symptomatic strongyloidiasis)	Probable
Miscellaneous clinical conditions	Case reports of small cell lung cancer with monoclonal HTLV-1 integration and invasive cervical cancer	Uncertain
HTLV-2–ASSOCIATED DISEASES		
HTLV-associated myelopathy	Increased numbers of cases among blood donors	Definite but rare

PPD = purified protein derivative.

white blood cell count, hypercalcemia, and cutaneous involvement. Lymphoma type (about 20% of cases) shares all features of acute ATL except for peripheral blood involvement. Other cases resemble T-prolymphocytic leukemia and are termed *chronic ATL* (about 20% of cases). A subset of patients with chronic ATL and with high serum levels of lactate dehydrogenase (LDH) and blood urea nitrogen (BUN) and low levels of albumin have a poor response to treatment, and their survival rates are similar to those of patients with the more aggressive acute and lymphoma types of ATL. Smoldering ATL (about 5%) may clinically resemble mycosis fungoides/Sézary syndrome (Chapter 185), with cutaneous involvement manifested as erythema or as infiltrative plaques or tumors (Fig. 378-2). A long prodrome of signs (e.g., cutaneous rashes) and symptoms (e.g., fevers) is sometimes noted in chronic and smoldering ATL before transformation to an acute or lymphoma-type ATL that is rapidly fatal.

DIAGNOSIS

The diagnosis should be considered in an adult with mature T-cell lymphoma and hypercalcemia or cutaneous involvement (or both) with characteristic flower cells (see Fig. 378-1), particularly if the individual is from a known risk group or endemic region. The diagnosis is established by testing of serum for HTLV-1 antibodies. Polymerase chain reaction (PCR) can detect infection and distinguish the type of virus. The cytologically distinct flower cell,

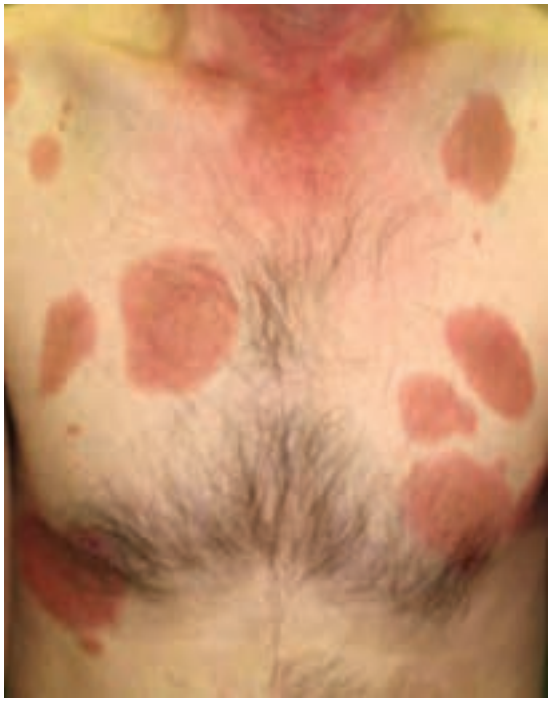

FIGURE 378-2. Cutaneous involvement in adult T-cell leukemia/lymphoma. (From Tomita H, Fumihide O, Kuwatsuka S, et al. Attenuation of an adult T-cell leukemia lesion after treatment of a concomitant simplex infection: a case study. *Virol J*. 2012;9:224. http://www.virologyj.com/content/9/1/224. Creative Commons Attribution License).

a sine qua non of HTLV-1–associated leukemia, is also seen in apparently healthy carriers. Occasional cases with characteristic clinical features are antibody negative but provirus positive as detected by PCR in blood cells or in biopsy specimens.

TREATMENT

Rx

Treatment is based on the type of ATL and its natural history, but it is also influenced by the patient's age and the presence of selected biomarkers.

Smoldering and Chronic ATL

Watchful waiting traditionally has been recommended for patients with smoldering ATL and some patients with chronic ATL with a favorable biomarker profile (normal LDH, BUN, and albumin levels) because cytoablative therapy increases the risk of lethal opportunistic infections. Some observational data suggest, however, that 5-year survival could be improved by treating smoldering and selected chronic ATL with high doses of the antiviral agent zidovudine (800 to 1000 mg/day in divided dosage) in combination with interferon alfa (6 to 9 million units)[6] on a daily basis for 30 to 60 days, but controlled trials are needed to define the optimal dose and regimen.[7]

Acute, Lymphoma-type, and Aggressive Chronic ATL

In a randomized trial, a nine-drug regimen (vincristine, cyclophosphamide, doxorubicin, and prednisone [VCAP]; doxorubicin, ranimustine, and prednisone [AMP]; and vindesine, etoposide, carboplatin, and prednisone [VECP] [Chapter 185]) was significantly better than biweekly cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) for inducing a complete remission.■ The nine-drug regimen provided nonsignificant improvements in progression-free survival at 1 year (28% vs. 16%), in median survival time (13 months vs. 11 months), and in 3-year overall survival (24% vs. 13%), but it was associated with more hematologic and infectious complications. Retrospective case series report a long-term survival of 20 to 40% after allogeneic hematopoietic stem cell transplantation (Chapter 178) but with significant treatment-related mortality.[8] Small case series have reported some benefit with zidovudine and interferon alfa, with potentially incremental benefit from the addition of arsenic triphosphate. Newer therapies, such as defucosylated humanized anti-CCR4 antibody (mogamulizumab) and denileukin diftitox (interleukin-2–diphtheria toxin conjugate), that target the high-affinity interleukin-2 receptor on ATL cells have shown some promise but need further study. In the absence of definitive data from randomized trials, a reasonable approach for patients with aggressive forms of ATL and favorable prognostic factors is VCAP-AMP-VECP alone. For patients with an unfavorable prognostic profile (thrombocytopenia, eosinophilia, bone marrow involvement, elevated LDH levels, high interleukin-5 serum level, C-C chemokine receptor 4 expression, lung resistance–related protein, p53 mutation, or p16 deletion), however, VCAP-AMP-VECP chemotherapy followed by allogeneic stem cell transplantation is recommended.

PROGNOSIS

Smoldering ATL has a relatively good prognosis, with a 5-year survival rate of 70%. For the chronic subtype, the 5-year survival is only 20%, even including the more favorable subset of patients with normal LDH, BUN, and albumin levels. For the aggressive forms of acute and lymphoma-type ATL, median disease-free survival is 0.6 year, with an overall survival of 0.8 year with high mortality linked to rapid tumor growth, infectious complications, and metabolic complications, especially hypercalcemia. Significant negative prognostic factors also include poor performance status at diagnosis, older age, advanced stage, and elevated serum LDH levels. Death usually results from rapid growth of tumor cells, hypercalcemia, bacterial sepsis, and opportunistic and other infectious complications.

HTLV-ASSOCIATED MYELOPATHY/TROPICAL SPASTIC PARAPARESIS

EPIDEMIOLOGY AND PATHOBIOLOGY

HTLV-1 causes an inflammatory neurologic syndrome known as HAM/TSP.[9] The lifetime incidence is approximately half the rate for ATL, with approximately 1 to 2% of carriers affected. HAM/TSP is approximately two times more likely to develop in females. The majority of adult cases occur in the 30- to 50-year age group, but cases have occurred in children as young as 3 years. These patterns suggest that the latency period for HAM/TSP is shorter than that for ATL and that both early-life and adult exposure causes disease. More than a dozen HAM/TSP cases associated with HTLV-2 have been reported, but the occurrence is infrequent compared with HTLV-1 carriers.

The pathogenesis results from trafficking of infected T cells in the perivascular areas and parenchyma of the spinal cord, where they cause astrocytosis. Ongoing inflammation of spinal gray and white matter results in a progressive and preferential demyelination and degeneration of the lateral and posterior columns, followed over time by loss of myelin and axons in the anterior columns.

CLINICAL MANIFESTATIONS

Symptoms include stiff gait, spasticity, lower extremity weakness, back pain, urinary incontinence and impotence, and (rarely) ataxia. The presenting symptom is stiff gait that progresses (usually slowly) to increasing spasticity and weakness, with incontinence and impotence developing later. In contrast to classic multiple sclerosis (Chapter 411), HAM/TSP is characterized by a generally slow and progressive course, absence of waxing and waning symptoms, and demyelination of long motor neurons rather than the CNS. However, some cases are acutely progressive, especially those associated with transfusion of HTLV-1–positive blood. It is not uncommon for other manifestations of HTLV-1–associated "autoimmune" disease to present coincident with the neurologic syndrome.

DIAGNOSIS

The diagnosis is suspected in patients with unexplained CNS disease and loss of pyramidal tract functions and is confirmed by testing of sera for HTLV-1 antibodies. Oligoclonal immunoglobulin bands in the cerebrospinal fluid (CSF) of patients with HAM/TSP react to HTLV-1 antigens, and the CSF to serum ratio of HTLV-1 antibodies is greater than 1. PCR quantification of the HTLV-1 cell-associated provirus in the CSF is also diagnostic. Spinal cord lesions often appear hyperintense on T2-weighted magnetic resonance imaging.

TREATMENT

Rx

Corticosteroids (e.g., intravenous methylprednisolone 1.5 g for 3 consecutive days in the first week, 2 consecutive days in the second week, 1 day in the third week, and then monthly for 6 months; an alternative is oral prednisone, 1 mg/kg/day in tapering doses for 2 to 4 months) reduce symptoms in

approximately 40% of cases, especially in early disease or in patients who are progressing rapidly. Studies evaluating antiviral drugs, such as lamivudine and zidovudine, have been inconclusive. Interferon alfa (3 MU three times per week) can improve neurologic symptoms. Pentoxifylline (400 to 1200 mg daily) decreases tumor necrosis factor-α and interferon-γ levels and has improved motor disability, especially spasticity, in uncontrolled trials. Cyclophosphamide also benefits some patients. Spasticity can be treated with tizanidine (2 to 12 mg daily) and botulinum toxin injection (100 to 400 IU every 4 to 6 months). Urinary incontinence is treated with oxybutynin (5 to 30 mg daily), imipramine (10 to 75 mg daily), or doxazosin mesylate (1 to 6 mg daily), and urinary retention is treated with bethanechol (10 to 50 mg daily). Treatment with danazol (100 to 400 mg daily in two divided doses) has resulted in improvement in urinary and fecal incontinence but not in the underlying neurologic deficit. Clinical trials are under way with Hu-Mik-(Beta)1, a genetically engineered antibody that blocks the action of interleukin-15 with the goal of blunting the autoimmune response that results in HAM/TSP.

PROGNOSIS

The prognosis of HAM/TSP is poor, with inexorable progression of neurologic deterioration.

OTHER HTLV-ASSOCIATED CONDITIONS

HAM/TSP is the prototype for a series of immune-mediated syndromes characterized by a high viral load, immune activation, and an indirect pathogenic mechanism produced by virally induced perturbations in immune function. Examples include skeletal muscle polymyositis, sporadic inclusion body myositis, uveitis (30 to 40% of cases in areas endemic for HTLV), large-joint arthropathy, pulmonary lymphocyte alveolitis/cryptogenic fibrosing alveolitis, Sjögren syndrome/keratoconjunctivitis sicca, and infective dermatitis, which is a pediatric syndrome characterized by inability to clear saprophytic skin infections (see Table 378-2). Carriers of HTLV-1 also have elevated rates of invasive cervical cancer, tuberculosis, parasitic infestations (e.g., strongyloidiasis), scabies, and refractory generalized eczema associated with infective dermatitis. One prospective study of HTLV-2–positive drug users showed an excess of asthma-related deaths and an increased frequency of skin and soft tissue infections.

PREVENTION

Patients often seek medical attention when confronted with a positive HTLV test result based on blood bank screening.[10] Confirmation with Western blot or PCR is needed. Confirmed positive patients must be told that complications related to HTLV-1 infection are rare and that HTLV-2 is hardly ever responsible for clinical disease. Second, it should be emphasized that these viruses are not easily transmitted. Third, the patient should be clearly counseled concerning the distinction between HTLV and HIV because the greatest fear that patients may experience is that they have the "AIDS virus." Other guidelines for prevention include the following:

- Blood for donation should be screened before transfusion, and positive donors should be deferred from donating.
- HTLV-1/HTLV-2–positive mothers should be discouraged from breast-feeding to prevent mother-to-infant transmission (except in particular settings, such as in the tropics, where diarrheal disease in non–breast-fed infants presents a high risk for morbidity and mortality).
- Condoms should be used by discordant couples, but given the relatively low frequency of sexual transmission per sexual encounter, couples who desire a pregnancy could time unprotected sexual intercourse to coincide with periods of maximal fertility. Such decisions, however, require careful discussion between the physician and the patient.

Postexposure prophylaxis with zidovudine is not recommended because the efficacy of such prophylaxis for HTLV infection has not been established. Vaccines containing whole virus and recombinant HTLV-1 envelope antigens have successfully prevented HTLV-1 infection in monkeys and in a rabbit model. However, a vaccine for humans is unlikely to be a high priority because of the relatively low incidence of clinical disease.

Grade A Reference

A1. Tsukasaki K, Utsunomiya A, Fukuda H, et al. VCAP-AMP-VECP compared with biweekly CHOP for adult T-cell leukemia-lymphoma: Japan Clinical Oncology Group Study JCOG9801. *J Clin Oncol*. 2007;25:5458-5464.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Gessain A, Cassar O. Epidemiological aspects and world distribution of HTLV-1 infection. *Front Microbiol*. 2012;3:388.
2. Pique C, Jones KS. Pathways of cell-cell transmission of HTLV-1. *Front Microbiol*. 2012;3:378.
3. Cook LB, Elemans M, Rowan AG, et al. HTLV-1: persistence and pathogenesis. *Virology*. 2013; 435:131-140.
4. Gillet NA, Malani N, Melamed A, et al. The host genomic environment of the provirus determines the abundance of HTLV-1–infected T-cell clones. *Blood*. 2011;117:3113-3122.
5. Iwanaga M, Watanabe T, Yamaguchi K. Adult T-cell leukemia: a review of epidemiological evidence. *Front Microbiol*. 2012;3:322.
6. Bazarbachi A, Plumelle Y, Carlos Ramos J, et al. Meta-analysis on the use of zidovudine and interferon-alfa in adult T-cell leukemia/lymphoma showing improved survival in the leukemic subtypes. *J Clin Oncol*. 2010;28:4177-4183.
7. Tsukasaki K, Tobinai K. Clinical trials and treatment of ATL. *Leuk Res Treatment*. 2012;2012:101754.
8. Kawada H, Yoshimitsu M, Nakamura D, et al. A retrospective analysis of treatment outcomes in adult T Cell leukemia/lymphoma patients with aggressive disease treated with or without allogeneic stem cell transplantation: A single-center experience. *Biol Blood Marrow Transplant*. 2015;21: 696-700.
9. McKendall RR. Neurologic disease due to HTLV-1 infection. *Handb Clin Neurol*. 2014;123: 507-530.
10. Chang YB, Kaidarova Z, Hindes D, et al. Seroprevalence and demographic determinants of human T-lymphotropic virus type 1 and 2 infections among first-time blood donors—United States, 2000-2009. *J Infect Dis*. 2014;209:523-531.

REVIEW QUESTIONS

1. A 35-year-old white woman from suburban Chicago arrives at your office having called for an emergency appointment because she was refused as a first-time blood donor owing to a positive virus test result. She received a blood transfusion as a child in the early 1980s and is concerned that she is positive for AIDS. When you contact the blood bank, the test result is positive for HTLV-2 (based on multiple bands on Western blot) and negative for HIV. After you explain to her the difference between HIV and HTLV, you make the following recommendations.

A. She should be seen every 6 months for a complete blood count to monitor her for possible leukemia.
B. She should see a neurologist annually to look for neurologic disease.
C. If she becomes pregnant, she should avoid breast-feeding her baby.
D. She can resume sexual activity with no restrictions.
E. She should have a CD4 test to make sure that she does not have AIDS.

Answer: C In the United States, all blood donors are screened for HTLV in addition to other blood-borne viruses. A positive HTLV test result is often misinterpreted as HIV, and the first task of the physician is to reassure the patient about the difference between HIV and HTLV. Because the patient is infected with HTLV-2, she is not at risk for hematologic disease. For HTLV-1, the risk for leukemia is mainly in patients with early-life exposure, whereas HTLV-associated myelopathy can occur as a result of later blood transfusion with HTLV-1–infected blood. There is no need for frequent monitoring for disease among patients with either HTLV-1 or HTLV-2, although a baseline complete blood count with differential is warranted, as is careful attention to neurologic signs and symptoms during routine annual physicals. Because HTLV-1 and probably HTLV-2 can be transmitted through breast milk, breast-feeding should be avoided. Because HTLV-1 and probably HTLV-2 can be sexually transmitted, couples may consider using condoms. A CD4 test is not warranted for HTLV infection.

2. Which geographic region is not endemic with HTLV-1?

A. United States
B. West Africa
C. Southern Japan
D. Caribbean
E. Iran

Answer: A The United States has a low prevalence and incidence of HTLV-1, and most of the cases arise from people who have migrated from endemic regions, such as the West Indies, Japan, and West Africa. In a recent serosurvey of first-time blood donors, the prevalence of HTLV-1 (5.1 per 100,000) and HTLV-2 (14.7 per 100,000) was low. The strongest risk factors for both HTLV-1 and HTLV-2 seropositivity are female sex, older age, and nonwhite race/ethnicity.

3. Which factor uniquely distinguishes HTLV-1 from HIV-1?

A. Prolonged asymptomatic infection
B. Sexual transmission
C. Integration into the host genome
D. Targets $CD4^+$ T cells
E. Immortalized cells

Answer: E HIV-1 and HIV-2 belong to the genus Lentivirus, whereas HTLV viruses belong to the genus Deltaretrovirus. Both deltaretroviruses and lentiviruses are associated with prolonged asymptomatic infections, both target $CD4^+$ T cells, both integrate into the host genome, and both are transmitted sexually. However, HIV-1 and HIV-2 have cytopathic effects on T cells, whereas the HTLVs transform T cells into immortalized cell lines.

379 ENTEROVIRUSES

JOSÉ R. ROMERO

DEFINITION

The enteroviruses belong to the genus Enterovirus in the family Picornaviridae. With the advent of molecular virology, the more than 100 recognized strains are classified on the basis of phylogenetic analysis of the nucleic acid sequence of VP1, the major enteroviral capsid protein (Table 379-1).

The Pathogens

Enteroviruses are small (30 nm in diameter), nonenveloped, icosahedral-shaped viruses. The viral capsid is composed of four viral proteins (VP1 to VP4). The enteroviruses possess an approximately 7.4-kilobase positive-sense single-stranded RNA genome. The 5′ end of the genome is covalently linked to a small protein, VPg. The genome is organized into a long (about 740 nucleotides) 5′ nontranslated region that precedes a single continuous open reading frame measuring about 6.63 kilobases. The open reading frame, which is followed by a short 3′ nontranslated region and a terminal polyadenylate tail, yields a single large polyprotein that is post-translationally modified to produce four capsid proteins, seven nonstructural proteins, and several functional protein intermediates. The 5′ and 3′ nontranslated regions of VPg participate in replication of the viral genome. The 5′ nontranslated region of the enterovirus is essential for translation and contains determinants of neurovirulence in the polioviruses.

EPIDEMIOLOGY

Worldwide, an estimated 1 billion or more enteroviral infections occur annually. In the United States, about 30 to 50 million annual infections result in approximately 10 to 15 million symptomatic cases, with coxsackievirus B1, echovirus 6, echovirus 9, echovirus 18, and coxsackievirus A9 accounting for more than 50% of these enteroviral infections.[1]

Humans are the only known reservoir for enteroviruses. Enteroviral infections are seasonal, and the majority of infections occur during the summer and early autumn in temperate regions. For example, more than 80% of infections occur in the United States from June through October. However, winter outbreaks highlight their panseasonal occurrence. In tropical and subtropical regions, infections continue year-round, with an increased incidence during the rainy season.

Globally, the dominant circulating enterovirus serotypes may vary annually by geographic region. In the United States, 15 serotypes accounted for about 90% of all isolates reported (Table 379-2).

TABLE 379-1 CLASSIFICATION OF ENTEROVIRUSES

SPECIES	SEROTYPES
Enterovirus A	CV- A2-A8, A10, A12, A14, A16 EV- A71, A76, A90, A91, A92, A114, A119, A120
Enterovirus B	CV- A9 CV- B1-B6 E- 1-7, 9, 11-21, 24-27, 29-33 EV- B69, B73-B75, B77, B78, B79-B88, B93, B97, B98, B100, B101, B106, B107, B110, B111
Enterovirus C	PV- 1-3 CV- A1, A11, A13, A17, A19-A22, A24 EV- C95, C96, C99, C102, C104, C105, C109, C113, C116-C118
Enterovirus D	EV- D68, D70, D94, D111

TABLE 379-2 THE 14 MOST COMMON ENTEROVIRUS SEROTYPES REPORTED BY NATIONAL ENTEROVIRAL SURVEILLANCE SYSTEM LABORATORIES TO THE CDC, 2006-2008

ENTEROVIRUS SEROTYPE	PERCENTAGE
Coxsackievirus B1	16.5
Echovirus 6	10.7
Echovirus 9	10.7
Echovirus 18	8.8
Coxsackievirus A9	7.5
Coxsackievirus B4	6.7
Echovirus 11	5.8
Coxsackievirus B3	5.4
Echovirus 30	4.5
Coxsackievirus B5	4.4
Coxsackievirus B2	3.4
Echovirus 25	1.8
Echovirus 7	1.7
Coxsackievirus A16	1.7
Total	89.6

From Centers for Disease Control and Prevention (CDC). Nonpolio enterovirus and human parechovirus surveillance—United States, 2006-2008. *MMWR Morb Mortal Wkly Rep.* 2010;59:1577-1580.

More than 80% of infections occur in individuals younger than 20 years, with the highest incidence in infants and children 4 years and younger. Nearly 45% of all infections occur in infants younger than 1 year. Among household members of infected children, clinical or serologic evidence of secondary infection can be seen in more than 50% of susceptible individuals. A male preponderance is noted in persons younger than 20 years (male-to-female ratio of 1.4 : 1), but not in older individuals.

Localized enteroviral outbreaks have been reported in neonatal units, nurseries, daycare centers, schools, camps, and sports teams. Community-wide outbreaks are common. Extensive regional outbreaks of EV-A71 have occurred in the Asia-Pacific region. Occasional pandemics, such as acute hemorrhagic conjunctivitis caused by EV-D70 and CV-A24, also have occurred.

Effective antipolio immunization programs have eradicated wild-type poliovirus serotype 2 worldwide and have also eradicated serotypes 1 and 3 from almost all countries except Afghanistan, Pakistan, and Nigeria, where they unfortunately remain endemic.[2] As a result, sporadic poliovirus outbreaks continue to occur in Africa, west Asia, China,[3] and war-torn Syria[4] from local reservoirs or imported infections. The use of live attenuated poliovirus vaccines has led to the problem of vaccine-derived polioviruses (VDPVs) as a result of the excretion of neurorevertant vaccine (Sabin) strains from individuals who have primary humoral immunodeficiencies, but not secondary humoral or other immunodeficiencies, or as a result of natural recombination between Sabin strains and members of the Enterovirus C species.[5] VDPVs that can circulate in the environment with evidence of person-to-person transmission have been termed circulating vaccine-derived polioviruses (cVDPVs). A third group of VDPVs, designated ambiguous VDPVs, are clinical isolates from individuals without known immunodeficiency or sewage isolates whose ultimate source is unknown. Similar to wild-type poliovirus, cVDPVs can cause acute flaccid paralysis in unimmunized or incompletely immunized individuals and have caused multiple outbreaks worldwide.

PATHOBIOLOGY

The polioviruses and the majority of the nonpolio enteroviruses are transmitted through a fecal-oral route. Notable exceptions include CV-A21 and EV-D68, which are spread by the respiratory route, and EV-D70, which may spread by contaminated fomites or ocular and respiratory secretions. Evidence also supports transplacental transmission of the enteroviruses.

Ingestion of the enteroviruses results in infection of cells of the pharynx and, because the virus is acid resistant, the lower gastrointestinal tract. Initial viral replication, which is thought to occur in the mucosal tissues of the nasopharynx and intestinal tract (i.e., tonsils and Peyer patches), leads to seeding of the deep cervical and mesenteric lymph nodes. Further replication at these sites results in a minor viremia with seeding of multiple organs, including the liver, lungs, heart, and central nervous system (CNS). Viral replication at these sites causes many of the clinical manifestations of infection and is followed by a major viremia that may infect the CNS if it was spared during the initial viremia. The virus is cleared by type-specific neutralizing antibodies directed at the capsid proteins by day 7 to 10 after infection. IgA antibodies appear in the respiratory and gastrointestinal tracts 2 to 4 weeks after infection.

The host humoral immune response is pivotal in the prevention and eradication of enteroviral infections. Congenital or acquired B-cell immunodeficiencies may result in chronic or prolonged infection. Experimental evidence suggests that the interferons are important in limiting the spread of poliovirus once infection has occurred. Natural killer cells and gamma or delta T cells may play roles in regulating the host T-cell response.

Histopathologic findings in patients who died of poliomyelitis reveal neuronal necrosis in association with mononuclear and polymorphonuclear infiltrates that are initially perivascular in distribution but are later found diffusely within the gray matter of the anterior horns of the spinal cord, the reticular formation of the hindbrain, the vestibular nuclei, and the roof nuclei of the cerebellum. In nonpolio enterovirus CNS infections in immunocompetent hosts, findings include edema of the meninges and cerebral parenchyma, with microscopic perivascular lymphocytic infiltration, increased numbers of oligodendrocytes, and focal areas of necrosis and hemorrhage. In cases of enteroviral myocarditis (Chapter 60), a mononuclear cell inflammation is associated with widespread myocardial necrosis followed by fibrosis, which may be focal but results in myocardial damage.

CLINICAL MANIFESTATIONS

The incubation period for enterovirus infections is generally 3 to 6 days, with a range of 2 days to 2 weeks. Depending on serotype and the age of the patient, as many as 90% of infected individuals may have subclinical infections. The enteroviruses are responsible for a wide array of clinical syndromes affecting nearly every organ system, and no enterovirus serotype is uniquely associated with a single disease or clinical syndrome (Table 379-3).

The most frequent enteroviral syndrome, seen in about 50 to 80% of cases, is nonspecific febrile illness, which occurs most commonly in infants, toddlers, and young children. The onset of illness is abrupt, with fever, poor appetite, lethargy, irritability, emesis, diarrhea, and upper respiratory tract symptoms. Physical findings are minimal and consist of mild pharyngeal and conjunctival injection and lymphadenopathy. Exanthems may be present in about 25% of cases.

DIAGNOSIS

Nucleic acid amplification techniques (e.g., reverse transcription–polymerase chain reaction [RT-PCR] and nucleic acid–based sequence amplification) are the preferred methods for detection and identification of all enteroviruses. Multiple studies have documented that nucleic acid amplification techniques are more sensitive and rapid than cell culture for the detection of enteroviruses in cerebrospinal fluid (CSF). RT-PCR can detect enteroviruses in CSF, blood, tissue, stool, and other body fluids within hours, and the results can shorten hospitalizations, decrease the use of antibiotics, and reduce health care costs.

Serologic testing is of limited use, although a four-fold change in antibody titer to a specific serotype of enterovirus in paired acute and convalescent sera can establish the diagnosis. Cell culture is not recommended because of limited sensitivity, prolonged positivity even after the related clinical syndrome may have resolved, and the several days required for viral detection.

TREATMENT AND PROGNOSIS

Rx

Nonspecific febrile enteroviral infections generally resolve in less than 5 days without sequelae. Because of concern for possible occult bacterial infection, however, significant numbers of young infants and children are hospitalized for evaluation and empirical therapy. After infection, virus may be shed into the nasopharynx for 2 to 6 weeks and in feces for several months.

TABLE 379-3 CLINICAL MANIFESTATIONS OF NONPOLIO ENTEROVIRUS INFECTIONS*

CLINICAL SYNDROME	GROUP A COXSACKIEVIRUSES†	GROUP B COXSACKIEVIRUSES	ECHOVIRUSES	ENTEROVIRUSES
Asymptomatic infection	All serotypes	All serotypes	All serotypes	All serotypes
Undifferentiated febrile illness ("summer grippe") with or without respiratory symptoms	All serotypes	All serotypes	All serotypes	68, 70, 71
Aseptic meningitis (often associated with an exanthem)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 16, 17, 18, 22, 24	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,12, 14, 16, 17, 18, 19, 20, 21, 25, 30, 31, 33	70, 71
Encephalitis	2, 4, 5, 6, 7, 9, 10, 16	1, 2, 3, 4, 5	2, 3, 4, 6, 7, 9, 11, 14, 17, 18, 19, 25, 30, 33	70, 71
Acute flaccid paralysis (poliomyelitis-like)	4, 5, 6, 7, 9, 10, 11, 14, 16, 21, 24	1, 2, 3, 4, 5, 6	1, 2, 4, 6, 7, 9, 11, 14, 16, 17, 18, 19, 30	68, 70, 71
Myopericarditis	1, 2, 4, 5, 7, 8, 9, 14, 16	1, 2, 3, 4, 5, 6	1, 2, 3, 4, 6, 7, 8, 9, 11, 14, 16, 17, 19, 25, 30	
Pleurodynia	1, 2, 4, 6, 9, 10, 16	1, 2, 3, 4, 5, 6	1, 2, 3, 6, 7, 8, 9, 11, 12, 14, 16, 19, 25, 30	
Herpangina	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 16, 22	1, 2, 3, 4, 5	6, 9, 11, 16, 17, [22], 25	71
Hand-foot-and-mouth disease	4, 5, 6, 7, 9, 10, 16	2, 5	7	71
Exanthems	2, 4, 5, 6, 7, 9, 10, 16	1, 2, 3, 4, 5	2, 4, 5, 6, 9, 11, 16, 18, 25	71
Common cold	2, 10, 21, 24	1, 2, 3, 4, 5	2, 4, 8, 9, 11, 20, 25	
Lower respiratory tract infections (bronchiolitis, pneumonia)	7, 9, 16	1, 2, 3, 4, 5	4, 8, 9, 11, 12, 14, 19, 20, 21, 25, 30	68, 71, 104
Acute hemorrhagic conjunctivitis§	24			70
Generalized disease of the newborn	3, 9, 16	1, 2, 3, 4, 5	3, 4, 6, 7, 9, 11, 12, 14, 17, 18, 19, 20, 21, 30	

*A great many enterovirus serotypes have been implicated in most of these syndromes, at least in sporadic cases. The serotypes listed are those that have been clearly or frequently implicated. Serotypes with a strong association are underlined.
†Because detection of many of the group A coxsackieviruses originally required suckling mouse inoculation, they are likely to be underreported as causes of illness.
§Conjunctivitis without hemorrhage is frequently seen in association with other manifestations in patients infected with many group A and group B coxsackieviruses and echoviruses, especially coxsackieviruses A9, A16, and B1 to B5 and echoviruses 2, 7, 9, 11, 16, and 30.From Modlin JR. Enterovirus. *Cecil Textbook of Medicine.* 23rd ed. Philadelphia: WB Saunders; 2008, with minor changes.

SPECIFIC CLINICAL SYNDROMES

Central Nervous System Infections

ACUTE FLACCID PARALYSIS

CLINICAL MANIFESTATIONS AND DIAGNOSIS

During poliovirus epidemics, 90 to 95% of infections are subclinical. In another 4 to 8% of patients, infection results in fever, fatigue, headache, anorexia, myalgia, and sore throat, which resolve in 2 to 3 days. Paralytic polio develops in less than 1 to 2% of infected individuals.

Sporadic cases of acute flaccid paralysis can also be seen with other enteroviruses, especially coxsackievirus CV-A7 and enterovirus EV-A71 and occasionally EV-D68. With the exception of EV-A71, the paralysis associated with nonpolio enteroviruses tends to be milder, and fever is absent at the time of onset of the paralysis.[6] The upper extremities and face are more commonly involved. Sensory pathways remain intact.

CSF may reveal a mild lymphocytic pleocytosis (<100 cells/μL) in association with mildly increased protein and normal glucose concentrations. Neuroimaging is not generally useful, but increased signal may be seen on T2-weighted magnetic resonance imaging in the anterior horn regions of the spinal cord in patients whose acute flaccid paralysis is caused by poliovirus and nonpolio enteroviruses.

TREATMENT AND PROGNOSIS

Rx

No specific therapy exists. Efforts should focus on monitoring for the development of respiratory failure or airway compromise as well as control of pain associated with muscle spasms.

The mortality associated with spinal poliomyelitis is about 5%. The mortality rates for enteroviruses associated with nonpolio enteroviral acute flaccid paralysis are not known. Before modern methods of respiratory and cardiovascular support, mortality rates higher than 50% were common in patients with bulbar or medullary poliomyelitis. The ultimate outcome of the paralysis is highly variable and can range from complete resolution to lifelong persistence. The greatest gains in recovery of strength occur during the first 6 months of convalescence. Paralytic limbs become atrophic, thereby leading to skeletal deformities. Patients with nonpolio enteroviruses usually have a more rapid recovery and less atrophy than do those with classic polio.

A syndrome of postpoliomyelitis muscle atrophy, which may be seen in 25 to 85% of individuals 2 to 3 decades after recovery from paralytic disease, is characterized by the gradual development of weakness, pain, and atrophy. Possible mechanisms include aging and neuronal dropout in compromised neuromuscular connections[7] or, less likely, reactivation/ongoing poliovirus infection.

MENINGITIS

The enteroviruses, especially echoviruses and group B coxsackieviruses, are the dominant cause of viral meningitis (Chapter 412) in all ages.[8]

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The clinical picture varies with age. The predominant symptoms of meningitis in neonates are nonspecific fever, irritability, lethargy, and poor feeding, often with a full fontanelle and a generalized rash. In neonates with meningoencephalitis, clinical findings may consist of fever, lethargy, seizures, full fontanelle, and focal neurologic abnormalities. Hepatitis, myocarditis, or pneumonitis, singly or in combination, may be present in severe cases.

In older infants and children, an abrupt onset of fever is the most frequent initial symptom. The fever may persist for 1 to 5 days and may exhibit a biphasic pattern. Irritability or lethargy is common. Other nonspecific symptoms include poor feeding, vomiting, diarrhea, and rash. Headache is present in nearly all children old enough to report it. Rash, malaise, sore throat, abdominal pain, and myalgia are common, and photophobia may be reported. Seizures occur in less than 5% of cases. Examination may reveal a full fontanelle. Signs of meningeal irritation (i.e., nuchal rigidity, Brudzinski and Kernig signs) occur in less than 10% of infants younger than 3 months and increase with age.

In adolescents and adults, headache is nearly always present and severe enough to require narcotic analgesics for control. Some patients report temporary relief of headache after lumbar puncture. Fever is not universal, but photophobia, signs of meningeal irritation, nausea, emesis, and neck stiffness occur in more than two thirds of patients. Myalgia is reported in 20 to 90% of patients. Less frequent findings include rash and abdominal pain.

CSF analysis generally reveals a mild to moderate lymphocytic pleocytosis (<500 cells/μL). Some patients, however, may have lymphocyte counts higher than 1000 or have neutrophilic pleocytosis early in the course of illness and then progress to a predominance of lymphocytes hours to days later. In a small percentage of patients, particularly infants, no pleocytosis is present even though enterovirus can be detected. The protein concentration

may be increased. Although the glucose concentration is generally normal, hypoglycorrhachia may occur, particularly in association with group B coxsackievirus meningitis. Neuroimaging in cases of meningitis is generally unrevealing. Nucleic acid amplification testing can detect enteroviruses in CSF. The serotypes most frequently identified from CSF specimens are, in descending order, E-9, E-11, E-30, CV-B5, E-6, CV-B2, CV-A9, E-4, CV-B4, E-7, E-18, E-5, and E-13.

TREATMENT AND PROGNOSIS

Rx

Treatment is supportive, with control of fever and pain. Currently available antiviral agents are not helpful. Hospitalization may not be necessary in adolescents and adults who appear well if a bacterial cause can be confidently excluded. In children and infants or when bacterial infection cannot be confidently excluded in adults, hospitalization and initial empirical antimicrobial therapy are advisable (Chapter 412) while awaiting the results of bacterial cultures of blood and CSF. Intravenous fluids may be required to prevent dehydration. Control of headache pain may require narcotic analgesics. Uncommon complications in all ages include coma, increased intracranial pressure, and inappropriate secretion of antidiuretic hormone.

The overwhelming majority of patients recover fully. The duration of illness in infants and children is generally less than 1 week. In adults, full recovery may take up to 3 weeks.

ENCEPHALITIS

The enteroviruses are responsible for up to 22% of identifiable causes of viral encephalitis (Chapter 414). In the single largest report of enterovirus-related encephalitis, 73% of confirmed cases occurred in individuals younger than 20 years, including about 40% of cases in patients younger than 10 years.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The onset of neurologic findings may be abrupt or be preceded by fever, headache, malaise, myalgia, upper respiratory symptoms, rash, nausea, emesis, or diarrhea. Somnolence, lethargy, and altered consciousness are common. Patients may exhibit irritability, changes in personality, or hallucinations. Generalized or focal seizures occur in up to 30% of patients, and a minority of patients may progress to coma. Neck stiffness and ataxia are frequent physical findings. Focal neurologic findings such as hemiplegia, hemichorea, and paresthesias are reported in nearly 30% of patients. The focal nature of the seizures and abnormal neurologic findings may be reminiscent of herpes simplex virus encephalitis (Chapter 374).

Among enteroviral encephalitides, EV-A71 is uniquely associated with severe brain stem encephalitis (rhombencephalitis), primarily in children. The initial manifestation may be a prodrome of either hand-foot-and-mouth disease or herpangina, which is followed by myoclonus that may be associated with ataxia, tremors, or cranial nerve abnormalities and a rapid onset of neurogenic pulmonary edema, shock, coma, and apnea.

Evaluation of CSF may reveal mild to moderate lymphocytic pleocytosis (<500 cells/μL), but the CSF white cell count may be normal. The protein concentration may be increased and may be the sole abnormality. The glucose concentration is generally normal. Enteroviruses may frequently be detectable by nucleic acid amplification testing of CSF.

TREATMENT AND PROGNOSIS

Rx

Management is supportive, with monitoring for the development of respiratory failure or airway compromise. Nearly 50% of patients require intensive care. In patients with focal neurologic findings, empirical acyclovir (10 to 15 mg/kg every 8 hours) is warranted until herpes simplex virus is excluded.

The median duration of hospitalization is less than 1 week, and mortality is below 10%. For severe EV-A71 rhombencephalitis, however, mortality may approach 70%. Long-term sequelae after rhombencephalitis include myoclonus, abducens nerve palsy, facial diplegia, ataxia, dysarthria, internuclear ophthalmoplegia, and central apnea.

Myopericarditis

Members of the Enterovirus B species and especially the group B coxsackieviruses are responsible for approximately a third of cases of acute myocarditis (Chapter 60). The majority of cases occur in young adults.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

An upper respiratory tract infection may precede the onset of cardiac symptoms by 1 to 2 weeks. Fever may be present, and common initial symptoms include dyspnea, chest pain, and fatigue. Physical examination may reveal a gallop rhythm or a pericardial friction rub.

Cardiomegaly may be seen on the chest radiograph. Electrocardiographic findings vary and include low-voltage QRS complexes, ST segment depression, T wave inversion, pathologic Q waves, ventricular arrhythmias, and heart block. Echocardiographic findings include a decreased ejection fraction, ventricular dilation, and pericardial effusion. Blood troponin levels are frequently elevated. Cardiac magnetic resonance imaging may help localize areas of myocarditis. Myocardial biopsy is recommended in selected patients who have refractory heart failure or suspected giant cell myocarditis (Chapter 60); enterovirus can be detected by nucleic acid amplification tests.

TREATMENT AND PROGNOSIS

Rx

Supportive treatment includes bedrest and management of heart failure (Chapter 59), arrhythmias (Chapters 64 and 65), and pericarditis with or without pericardial effusion (Chapter 77). Immunosuppressive therapy is not generally recommended (Chapter 60). In approximately a third of patients with acute myocarditis, chronic dilated cardiomyopathy develops. In patients with pericarditis, recurrent pericardial effusions or chronic constrictive pericarditis may develop in the future.

Exanthems and Enanthems

Enterovirus infections may result in a wide spectrum of febrile exanthems and enanthems, including macular, papular, maculopapular, morbilliform, rubelliform, vesicular, urticarial, papulopustular, and scarlatiniform types. Exanthems are more commonly observed in children 15 years or younger. Any serotype is capable of causing several different rashes. With the exception of CV-A16, no serotype is associated with a unique rash. Echovirus 9 and CV-A9 can cause petechial or purpuric rashes reminiscent of meningococcemia (Chapter 298).

Hand-Foot-and-Mouth Disease and Herpangina

Hand-foot-and-mouth disease is typically associated with Enterovirus A species, particularly CV-A16, CV-A6, and EV-A71.[9] Herpangina is also most commonly caused by the group A coxsackieviruses in the Enterovirus A species, but it has also been associated with group B coxsackieviruses, echoviruses, and enteroviruses from the Enterovirus B species.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Hand-foot-and-mouth disease begins with low-grade fever, malaise, anorexia, and oral soreness. In 1 to 2 days, oral macules appear and then rapidly vesiculate and ulcerate. Oral lesions are typically distributed on the buccal mucosa and tongue, but they may also occur on the palate, uvula, anterior pillars, and gums. In approximately two thirds of patients, the enanthem is accompanied by an exanthem, with tender 3- to 7-mm vesicles on the dorsum of the hands and feet, frequently involving the palms and soles. Lesions may appear on the buttocks but tend not to be vesicular. Hand-foot-and-mouth disease associated with CV-A6 has a wider distribution of skin lesions that enlarge and vesiculate. Onychomadesis (loss of fingernails) can occur 1 to 2 months after the infection.

Herpangina begins with high fever, particularly in young patients. Additional findings include sore throat, mild cervical lymphadenopathy, sialorrhea, anorexia, dysphagia, abdominal pain, and emesis. Examination of the mouth and throat reveals 1- to 2-mm papulovesicular, grayish white lesions with an areola of erythema, primarily located on the anterior pillars of the tonsillar fauces. The soft palate, uvula, and tonsils may also be involved. Rarely, the posterior buccal surfaces and dorsal tip of the tongue may be involved. During a period of 2 to 3 days, the lesions increase to 3 to 4 mm in size. On average, five lesions are present.

TREATMENT AND PROGNOSIS

Rx

Hand-foot-and-mouth disease usually resolves in less than 1 week and herpangina generally resolves in 10 days, both typically uneventfully, without the need for hospitalization and without sequelae. Infants and young children may require hospitalization for administration of parenteral fluids. Disease

associated with CV-A6 has a higher rate of hospitalization. Hand-foot-and-mouth disease due to EV-A71 may precede the development of life-threatening rhombencephalitis (see Encephalitis).

Acute Hemorrhagic Conjunctivitis

Acute hemorrhagic conjunctivitis (Chapter 423) is associated with EV-D70 and CV-A24. The illnesses caused by the two serotypes are indistinguishable from each other. However, acute hemorrhagic conjunctivitis caused by CV-A24 may be more commonly accompanied by upper respiratory and systemic symptoms and may be associated with less severe conjunctival hemorrhage. High secondary attack rates within households are common. Other enteroviruses can cause acute conjunctivitis or keratoconjunctivitis but generally without hemorrhagic manifestations.

CLINICAL MANIFESTATIONS

An incubation period of about 1 to 2 days precedes the rapid onset of palpebral swelling associated with lacrimation, photophobia, blurring of vision, and severe ocular pain. The hallmark subconjunctival hemorrhages vary in size from petechiae to large blotches. Although transient keratitis occurs frequently, it seldom results in subepithelial opacities. Preauricular adenopathy is common, but fever is not. An ocular mucopurulent discharge may occasionally be present.

TREATMENT AND PROGNOSIS

Rx

Management is supportive. The illness usually persists for 1 to 2 weeks, but complete recovery is generally the rule. A transient lumbar radiculomyelopathy and acute flaccid paralysis–like illness may develop in some patients.

Respiratory Tract Syndromes

The enteroviruses may cause upper and lower respiratory tract syndromes, alone or accompanying other syndromes. As determined by nucleic acid amplification testing, enteroviruses are responsible for up to 15% of upper respiratory tract syndromes. They also cause 18% of lower respiratory tract syndromes in hospitalized children and 25% of hospitalizations in patients with acute wheezing. Recently, EV-D68 and EV-C104 are increasingly recognized as causes of respiratory tract disease.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The enterovirus "summer cold" (Chapter 361) consists of nasal congestion, rhinorrhea, and sneezing. Malaise and cough may be present. Fever and sore throat are typically absent or minimal.

Pharyngitis, tonsillitis, or pharyngotonsillitis begins abruptly with fever and sore throat. The nasopharynx, tonsils, uvula, and soft palate demonstrate erythema and inflammation. Petechiae may be present, and cervical lymphadenitis is common. Other syndromes associated with the enteroviruses include bronchitis (Chapter 96) and bronchiolitis.

Enteroviral pneumonias begin gradually with coryza, anorexia, and low-grade fever. A nonproductive cough, tachypnea, retractions, nasal flaring, and wheezing may be present. In severe cases, cyanosis may develop. The chest radiograph may demonstrate perihilar infiltrates, patchy consolidation, air trapping, and atelectasis. Recently, EV-D68 has become a cause of significant respiratory disease, primarily in young children and infants but also in adolescents and adults. An underlying pulmonary condition such as asthma or wheezing has been reported in 70% to 80% of cases. EV-D68-associated respiratory syndromes include pneumonia, bronchiolitis, asthmatic bronchitis, asthma exacerbation, and wheezing. Signs and symptoms include cough, wheezing, dyspnea, tachycardia, and inter- and subcostal retractions. Acute flaccid paralysis is a rare complication. Interestingly, fever may be absent, but hypoxia is common. The chest radiograph may show infiltrates and atelectasis.[10]

TREATMENT AND PROGNOSIS

Rx

Treatment is supportive and consists of control of fever and pain. In older children and adults, hospitalization is not usually required. Resolution occurs in 7 days or less. Hospitalization and admission to an intensive care unit may be required for respiratory support of EV-D68 infection. Death is uncommon but does occur.

Myositis

PLEURODYNIA

The group B coxsackieviruses are the major cause of sporadic and epidemic pleurodynia, but the syndrome may also be caused by a limited number of echoviruses and group A coxsackieviruses within the Enterovirus A and B species.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The onset of illness is abrupt in approximately 75% of patients. In the remainder, the onset of pleuritic chest pain is preceded by a prodrome of headache, malaise, anorexia, and vague myalgia lasting up to 10 days. Pain may be referred to the lower ribs or the sternum, and it can radiate to the shoulders, neck, or scapula. Pain is exacerbated by deep breathing, coughing, sneezing, or movement. During paroxysms, patients tend to be tachypneic and to have shallow breathing. Additional findings can include abdominal pain, headache, cough, anorexia, nausea, vomiting, and diarrhea. Fever may be biphasic.

Physical examination does not generally reveal muscle tenderness, obvious myositis, or muscle swelling. A pleural friction rub may be present in 25% of patients. The chest radiograph is typically normal.

TREATMENT AND PROGNOSIS

Rx

Treatment is supportive, with an emphasis on nonsteroidal analgesics (e.g., ibuprofen, 200 to 400 mg per dose every 4 to 6 hours) or hydrocodone (5 to 10 mg four times per day) alone or in combination with acetaminophen to control pain. The symptoms may persist for 1 to 14 days (mean of 3.5 days) and resolve without sequelae, although recurrent symptoms may occur in 25% of patients.

INFLAMMATORY MYOSITIS

Multiple enteroviral serotypes are associated with focal or generalized myositis. In patients with B-cell immunodeficiencies, a dermatomyositis-like syndrome may develop.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Nonspecific findings include fever and chills. Involved muscles may be weak, tender, and edematous. Chemical evidence of myositis may be evidenced by elevated serum levels of creatine kinase, myoglobinemia, and myoglobinuria.

TREATMENT AND PROGNOSIS

Rx

Treatment is supportive. With the exception of patients with B-cell immunodeficiencies, recovery is complete and rapid.

Enterovirus Infections in Special Populations

PATIENTS WITH B-CELL IMMUNODEFICIENCIES

Nonpolio enteroviruses and polioviruses may result in chronic or prolonged infections in patients with congenital or acquired B-cell immunodeficiencies (Chapter 250), such as those with X-linked agammaglobulinemia, hyper-IgM syndrome, severe combined immunodeficiency syndrome, or common variable immunodeficiency, or in patients receiving chemotherapy or immunomodulatory therapies, especially rituximab, who are undergoing bone marrow or solid organ transplantation.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Meningoencephalitis, pulmonary infections, and severe gastroenteritis can occur. The initial symptoms may consist of only persistent headaches and lethargy. As the disease progresses, additional neurologic findings develop and may include ataxia, loss of cognitive skills and memory, dementia, emotional lability, paresthesias, weakness, dysarthria, and seizures. Non-CNS manifestations include a dermatomyositis-like syndrome, edema, exanthems, and hepatitis. CSF demonstrates a persistently elevated protein concentration and pleocytosis. Enterovirus is detectable in CSF by RT-PCR.

TREATMENT AND PROGNOSIS

Rx

Children with humoral immunodeficiency (Chapter 250) should receive lifelong immunoglobulin replacement therapy in an attempt to prevent chronic infection. However, chronic meningoencephalitis develops in some patients and is usually ultimately fatal. Severe or fatal enteroviral infections have been reported in individuals receiving rituximab.

PREVENTION

Handwashing is the primary method for the prevention of enteroviral infections. Only poliovirus infections are currently preventable through vaccination. However, successful trials of inactivated EV-A71 vaccines have been completed in China.[A1,A2] These vaccines could significantly reduce the incidence of severe EV-A71 CNS disease in China. Whether the vaccine will be effective in other regions of the world will depend on its ability to protect from disease caused by the different genotypes found worldwide.

Grade A References

A1. Zhu FC, Meng FY, Li JX, et al. Efficacy, safety, and immunology of an inactivated alum-adjuvant enterovirus 71 vaccine in children in China: a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2013;381:2024-2032.

A2. Zhu F, Xu W, Xia J, et al. Efficacy, safety, and immunogenicity of an enterovirus 71 vaccine in China. *N Engl J Med*. 2014;370:818-828.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Centers for Disease Control and Prevention (CDC). Nonpolio enterovirus and human parechovirus surveillance—United States, 2006-2008. *MMWR Morb Mortal Wkly Rep*. 2010;59:1577-1580.
2. World Health Organization. Progress towards global interruption of wild poliovirus transmission, January 2012–March 2013. *Wkly Epidemiol Rec.* 2013;88:181-187.
3. Luo HM, Zhang Y, Wang XQ, et al. Identification and control of a poliomyelitis outbreak in Xinjiang, China. *N Engl J Med*. 2013;369:1981-1990.
4. Aylward RB, Alwan A. Polio in Syria. *Lancet*. 2014;383:489-491.
5. World Health Organization. Update on vaccine-derived polioviruses detected worldwide, April 2011–June 2012. *Wkly Epidemiol Rec.* 2012;87:358-368.
6. Rhoades RE, Tabor-Godwin JM, Tsueng G, et al. Enterovirus infections of the central nervous system. *Virology*. 2011;411:288-305.
7. Bickerstaffe A, van Dijk JP, Beelen A, et al. Loss of motor unit size and quadriceps strength over 10 years in post-polio syndrome. *Clin Neurophysiol*. 2014;125:1255-1260.
8. de Ory F, Avellon A, Echevarria JE, et al. Viral infections of the central nervous system in Spain: a prospective study. *J Med Virol*. 2013;85:554-562.
9. Notes from the field: severe hand, foot, and mouth disease associated with coxsackievirus A6—Alabama, Connecticut, California, and Nevada, November 2011–February 2012. *MMWR Morb Mortal Wkly Rep*. 2012;61:213-214.
10. Midgley CM, Jackson MA, Selvarangan R, et al. Severe respiratory illness associated with enterovirus D68—Missouri and Illinois, 2014. *MMWR Morb Mortal Wkly Rep*. 2014;63:798-799.

REVIEW QUESTIONS

1. An 18-year-old female college freshman presents to the emergency department in early autumn for evaluation of severe headache of 36 hours' duration. She has had two episodes of nonprojectile emesis and complains of diffuse myalgia, photophobia, and phonophobia. She has had no ill contacts and has never been sexually active. She received conjugated meningococcal vaccine the summer before beginning college, and her childhood immunizations are complete. She has no significant past medical history and takes no medications other than oral contraceptives. Physical examination findings: temperature, 38.5° C; heart rate, 100/minute; respiratory rate, 20/minute; blood pressure, 110/70 mm Hg. She appears mildly distressed but is awake, alert, and talkative. Her pharynx is slightly injected. Her neck is tender to flexion. Her neurologic examination reveals positive Kernig and Brudzinski signs. Funduscopy reveals normal optic discs. Which of the following is the most likely etiology of this woman's syndrome?

A. Echovirus 9
B. *Neisseria meningitidis* type C
C. Human immunodeficiency virus
D. *Haemophilus influenzae* type b
E. None of the above

Answer: A The symptoms and mild clinical findings together with the time of year are consistent with an enteroviral infection. *Neisseria meningitidis* type C would be unlikely in this patient. She received conjugated meningococcal vaccine before starting college. Strains A, C, W-135, and Y are covered by the conjugated meningococcal vaccine. The patient has no risk factors for the acquisition of human immunodeficiency virus. The patient has received all of the childhood vaccinations, which would have included conjugated *Haemophilus influenzae* type b vaccine.

2. An 18-year-old man is brought to the emergency department for evaluation of fever of 2 days' duration and headache. The onset of the fever was sudden. Eight hours ago, he developed a headache and neck stiffness, and he has vomited three times. He complains that the light hurts his eyes. In the last 2 weeks, multiple siblings have had nonspecific febrile illnesses. His childhood immunizations are up to date, including conjugated meningococcal vaccine. Physical examination findings: temperature, 38.2° C; heart rate, 96/minute; respiratory rate, 29/minute; blood pressure, 118/78 mm Hg. He appears moderately distressed but is awake and alert. His neck is tender to flexion, and his neurologic examination reveals positive Kernig and Brudzinski signs. A lumbar puncture is performed and reveals the following: red blood cells, 0; white blood cells, 75/μL: 95% lymphocytes, 5% neutrophils; protein, 65 mg/dL; glucose, 90 mg/dL; Gram stain, no organisms seen. Which of the following studies would most likely establish the etiology of this adolescent's syndrome?

A. Cell culture detection of the etiologic agent from the cerebrospinal fluid (CSF)
B. Reverse transcription–polymerase chain reaction (RT-PCR) detection of the etiologic agent from the CSF
C. Magnetic resonance imaging of the brain
D. Serologic testing for the etiologic agent from the CSF
E. All of the above

Answer: B RT-PCR is the method of choice for the detection of the enteroviruses. RT-PCR can detect even noncultivable enterovirus serotypes. It has greater sensitivity than cell culture and can be performed in hours. Cell culture has been replaced by RT-PCR as the diagnostic method. The clinical usefulness of cell culture is limited by its lack of sensitivity, the time required to obtain a positive result, and the lack of growth of some enterovirus serotypes. Neuroimaging in cases of viral meningitis is generally unrevealing. Serologic testing is of limited use because of the large number of serotypes that compose the genus Enterovirus. A four-fold change in antibody titer to a specific serotype of enterovirus in paired acute and convalescent sera is required for diagnosis. The clinical utility of serologic testing is limited by the need for acute and convalescent specimens.

3. A 31-year-old man is being evaluated for sudden onset of chest pain involving the lower ribs and sternum. The pain, which is described as stabbing and which occasionally radiates to his shoulder, is exacerbated by deep inspiration or movement. The onset of the pain was preceded by 5 days of fever, headache, malaise, and anorexia. His past medical history is noncontributory. Physical examination findings: temperature, 38.5° C; heart rate, 116/minute; respiratory rate, 32/minute; blood pressure, 128/88 mm Hg. He is in obvious pain but is awake and alert. His ears, nose, and throat are normal. His neck is supple. Heart sounds are normal. He is tachypneic and breathing shallowly, but his lungs are clear. His abdominal examination findings are normal. A chest radiograph and electrocardiogram are both normal. Which of the following is the most likely syndrome?

A. Pneumonia due to enterovirus 68
B. Myocarditis due to coxsackievirus B3
C. Pleurodynia due to coxsackievirus B5
D. Herpangina due to coxsackievirus A4
E. None of the above

Answer: C The sudden onset of chest pain, preceded by fever and vague constitutional symptoms, is characteristic of pleurodynia. The pain may be exacerbated by deep breathing or movement. Radiation of the pain to the shoulders, back, or neck may occur. The absence of pulmonary findings (cough, wheezing) and a normal chest radiograph argue against an enteroviral pneumonia. Enteroviral myocarditis may be preceded by an upper respiratory tract infection. Common symptoms include dyspnea, chest pain, and fatigue. Physical examination may reveal a gallop rhythm or a pericardial friction rub. Cardiomegaly may be seen on the chest radiograph. Electrocardiographic findings vary and include low-voltage QRS complexes, ST segment depression, T wave inversion, pathologic Q waves, ventricular arrhythmias, and heart block. Herpangina begins abruptly with high fever; sore throat, mild cervical lymphadenopathy, sialorrhea, and dysphagia are common. An enanthem consisting of papulovesicular, grayish white lesions with an areola of erythema is seen in the mouth and throat.

4. A large community epidemic of echovirus 30 meningitis is affecting all age groups. Which of the following is the most likely mode of transmission of the virus?

A. Respiratory
B. Fomites
C. Blood-borne
D. Fecal-oral
E. All of the above

Answer: D The majority of enteroviruses are transmitted through the fecal-oral route. A few enteroviruses responsible for upper and lower respiratory tract infection are transmitted by the respiratory route. Fomites have been associated with the transmission of enteroviruses associated with hemorrhagic conjunctivitis but are not the major mode of transmission. Blood-borne transmission of the enteroviruses has not been reported.

380

ROTAVIRUSES, NOROVIRUSES, AND OTHER GASTROINTESTINAL VIRUSES

MANUEL A. FRANCO AND HARRY B. GREENBERG

DEFINITION

Viruses are a principal cause of acute infectious gastroenteritis, a syndrome of vomiting, watery diarrhea, or both that begins abruptly in otherwise healthy persons. Two distinct viruses account for the majority of cases. Rotaviruses are the most frequent cause of sporadic, severe gastroenteritis in young children and are responsible for the death of approximately 1200 children daily worldwide,[1] mainly in developing countries. Noroviruses are the primary cause of epidemic infectious gastroenteritis in both infants and adults in developed countries. For example, outbreaks of gastroenteritis in closed settings, such as cruise ships and nursing homes, are a typical manifestation of norovirus infections. However, noroviruses are also a common cause of sporadic, severe gastroenteritis in young children.[2]

The Pathogens

Noroviruses

Noroviruses, which are one of the five genera of the Caliciviridae family, are nonenveloped, icosahedral viruses with a relatively small, positive-sense, single-stranded RNA genome. The norovirus genus is further classified into five genogroups (GI to GV), only three of which (GI, GII, and GIV) are known to infect humans. GIII and GV viruses infect bovines and mice, respectively, and to date these animal viruses have not been shown to infect humans. Viruses in each genogroup are further divided into genotypes (more than 25 have been described) and subgroups. Norwalk virus is a prototype genogroup I genotype 1 (GI.1) virus. The norovirus genome is approximately 7.7 kilobases in size and consists of three open reading frames, the first of which encodes the nonstructural proteins that are essential for virus replication. The second open reading frame encodes the major capsid protein, viral protein 1 (VP1). When it is expressed as a recombinant protein, 180 molecules of VP1 autoassemble into virus-like particles (VLPs) that are critical to the study of noroviral epidemiology and immunity. Human noroviruses have not yet been adapted to cell culture, so diagnosis generally depends on amplification of virus genes by the polymerase chain reaction (PCR; see later) or use of VLPs as recombinant antigens for serologic analysis.

Rotavirus

Rotaviruses, which belong to the family Reoviridae, are large, icosahedral, nonenveloped viruses with a segmented, double-stranded RNA genome and a triple-layered protein coat. Rotaviruses are classified into groups A through G on the basis of the presence of cross-reactive antigenic epitopes and their overall genetic relatedness. Group A rotaviruses are the most commonly encountered viral enteric pathogens of young humans and many other species. Group B viruses have been identified sporadically in outbreaks of adult diarrheal illness in China and more recently in studies of children with sporadic gastroenteritis, principally in India. Group C rotaviruses are primarily veterinary pathogens and are infrequently associated with diarrheal disease in humans and animals around the world compared with group A rotaviruses. Groups D through G rotaviruses have been isolated only from animals, primarily avian species. Rotaviruses are 100-nm particles that have three concentric layers of proteins: the core is composed of VP1, VP2, and VP3 and the segmented, double-stranded RNA genome; the intermediate layer is formed by VP6, the most abundant and antigenic structural viral protein; and the external layer is composed of VP7 and VP4. The genome, which is composed of 11 segments of double-stranded RNA that together are approximately 18 kilobases in length, encodes six structural and six nonstructural proteins. As is the case among virtually all other RNA viruses, the rotavirus RNA polymerase is error prone and, along with selective pressure such as the evolution of immunity, drives viral diversity. For rotaviruses, gene reassortment, which is the mixing of gene segments from different parental viruses in cells coinfected by two or more strains, and rearrangement of the viral genome also contribute to genetic diversity. Reassortment of gene segments between animal and human rotavirus strains also occurs in natural settings, especially in less developed countries.

Other Agents

Other viral agents that cause human acute infectious gastroenteritis that is difficult to distinguish from disease caused by rotaviruses and noroviruses include the sapovirus (like norovirus, a member of the Caliciviridae family), enteric adenoviruses (Chapter 365) belonging to types 40 and 41, and astroviruses (Table 380-1). The frequency of detection (by PCR assays) of these viruses in individuals with acute gastroenteritis depends on the setting, but they are almost always detected much less frequently than are rotaviruses and noroviruses.[3] Coronaviruses (Chapter 366), toroviruses, picobirnaviruses, picornavirus (Chapter 379), bocavirus, parechoviruses, and pestiviruses have also been isolated occasionally from persons with acute gastroenteritis, but their roles as causative agents of enteric disease remain uncertain. Among patients with acute gastroenteritis, no etiologic agent is found in approximately 25 to 50% of cases.

EPIDEMIOLOGY

Norovirus

Over time, noroviruses appear to undergo antigenic drift in response to the acquisition of immunity in the general population, much like influenza viruses.[4] At present, gastroenteritis cases around the world are most frequently caused by the GII.4 norovirus strain, but new strains generally evolve every 2 to 4 years owing to antigenic drift. Outbreaks frequently take place in settings of close human contact, such as military establishments, cruise ships, nursing homes, and schools, especially in cold and dry weather (see Table 380-1). Viral spread is enhanced by the very high level of infectivity of noroviruses, as data suggest that 1 to 10 particles constitute an infectious dose.

Noroviruses of genotypes GII.4 and GII.3 also are responsible for about 12% of sporadic gastroenteritis in children younger than 5 years in both developed and developing countries. In the United States, noroviruses have recently surpassed rotavirus as the principal cause of medically attended visits for gastroenteritis in children younger than 5 years.[5] In the United States, noroviruses cause an estimated average of 570 to 800 deaths, 56,000 to 71,000 hospitalizations, 400,000 emergency department visits, 1.7 to 1.9 million outpatient visits, and 19 to 21 million total illnesses per year. The

TABLE 380-1 EPIDEMIOLOGIC AND CLINICAL FEATURES OF NOROVIRUS AND ROTAVIRUS

	NOROVIRUS	ROTAVIRUS	ASTROVIRUS
Epidemics	Occurs year-round; outbreaks tend to peak in cold weather	Year-round in equatorial countries; winter peak in others	Winter epidemics in children and endemic year-round
Key driver of epidemics	Antigenic drift strains promoted by population-based immunologic pressure	Size of the susceptible birth cohort	Unknown
Transmission	Fecal-oral, water, and food-borne outbreaks	Fecal-oral	Fecal-oral
Severity of diarrhea in children	Generally mild but can be severe	Most severe	Milder than rotavirus or noroviruses
Reservoir	Humans are the only known reservoir of noroviruses that infect humans	Mostly humans, but rotaviruses from farm animals and pets (especially in developing countries) infect humans	Humans are the only known reservoir of astroviruses that infect humans
Prevention	Viral protein 1–based vaccine in development	Several vaccines available	No vaccine in development
Age predisposition	All ages	Children <5 years; disease transmission in older family contacts is low (<25%)	Generally children, but adults may also get disease

death toll principally occurs in the elderly, and the health care costs principally occur in children younger than 5 years.[6]

Rotavirus

The incidence of rotaviral disease is similar in children in both developed and developing countries, suggesting that measures such as access to clean water will not replace the need for an effective vaccine. Before the introduction of an effective vaccine, rotavirus was estimated to be responsible for about 600,000 annual deaths worldwide. In developed countries, rotavirus rarely is fatal; but before the introduction of vaccines in the United States, it resulted in hospitalization for rotavirus-mediated gastroenteritis in about 1 in 75 children by 5 years of age.

In the temperate zones of the world, rotaviral infection occurs primarily during epidemic peaks in the cooler months of the year (see Table 380-1). This pattern is not seen, however, in countries within 10 degrees of the equator, where infection occurs in an endemic fashion year-round. Before the introduction of rotavirus vaccination, a yearly wave of rotaviral illness spread across the United States and Europe following peculiar spatiotemporal patterns. In the United States, this pattern of spread has been correlated with variation in birth rates, thereby suggesting that the number of babies experiencing their first infection is one of the primary drivers of rotavirus epidemics. The high birth rates in developing countries may also influence the differential epidemiologic distribution of rotaviruses. The widespread use of rotavirus vaccine has greatly reduced or eliminated this spatiotemporal spread of rotavirus in the United States.

Antibodies against the outer capsid proteins are the basis of serotypic classification of rotaviruses into G (glycoprotein, VP7) and P (protease-sensitive, VP4) serotypes. For technical reasons, P serotyping reagents are infrequently available, and classification is based on the P genotype (provided in brackets). Worldwide, most human infections are caused by five types of group A rotavirus; P[8]G1 is by far the most common (approximately 53% of strains), followed by P[8]G3, P[4]G2, P[8]G9, and P[8]G4. In some developing areas like India, Brazil, and Africa, P[6]G9, G5, and G8 rotaviruses, respectively, are frequently encountered. Some human rotavirus strains may have arisen after reassortment with bovine or porcine rotaviruses. A high prevalence of G12 viruses has recently been observed in several countries, thereby suggesting that this serotype may be an emerging rotavirus strain. Results from a large multicenter study in sub-Saharan Africa and south Asia, where most of the gastroenteritis deaths in children younger than 5 years of age occur, confirmed that rotavirus is the most common etiologic agent of this syndrome and causes an important nutritional burden in children.[7]

PATHOBIOLOGY

Norovirus

Histo-blood group antigens (HBGAs) are the receptors for noroviruses and determine susceptibility to disease in a strain-specific manner. The HBGAs are complex carbohydrate oligosaccharides linked to proteins or lipids that are expressed on the mucosal epithelia of the digestive tract. All three major families of HBGA, the ABO, Lewis, and secretor families, are involved in binding noroviruses. The secretor status of a person is controlled by the fucosyltransferase 2 (*FUT2*) gene. Secretor-negative individuals are specifically resistant to infection with the Norwalk virus (GI.1) and some GII viruses.

Although norovirus RNA has been detected in the blood stream of up to 15% of patients with norovirus gastroenteritis, the site of primary viral replication is most probably in the gastrointestinal tract. Consistent with the strong association of vomiting with norovirus disease, gastric emptying is delayed. Proximal jejunal biopsy specimens show blunting of the villi with crypt cell hyperplasia and cytoplasmic vacuolation, sometimes with an increase in epithelial cell apoptosis. A functional alteration of the epithelial barrier is likely to occur. Unknown at present is the effectiveness and persistence of long-term immunity in the context of natural infection, in which the infectious dose is generally quite low.

Rotavirus

For rotaviruses, HBGAs have also been recently proposed as receptors that determine susceptibility to disease in a strain-specific manner. Rotaviruses replicate in the villus tip cells of the small bowel, where the pathologic process includes shortening and atrophy of the villi, vacuolization of enterocytes, mononuclear infiltration in the lamina propria, and distention of the cisternae of the endoplasmic reticulum. However, the severity of clinical disease has not been directly related to the extent of intestinal disease; rather, it is related to levels of viral RNA in stool.

During the initial phases of the disease, altered intestinal secretion, motility, and permeability contribute to the pathophysiologic mechanism of diarrhea. Later in the course of disease, malabsorption can occur. Rotaviral NSP4, encoded by rotavirus gene 10, is a viral enterotoxin that mediates, at least in part, the early secretory components of the diarrhea. It has also been postulated that viral infection increases intestinal motility by stimulating the enteric nervous system, possibly through NSP4. Whether and to what degree the enterotoxic effect of NSP4 is clinically relevant in children or other animal species remains to be determined. Infected individuals have a short period of viremia, but its clinical consequences are unclear other than correlating with the level of fever. However, most rotavirus-infected children have mild elevations in hepatic enzymes, thereby suggesting that low-level hepatitis is a common occurrence.[8]

Rotavirus serum IgA levels measured shortly after natural infection in children generally correlate with intestinal IgA levels and appear to correlate with protection.[9] One explanation for recurrent rotavirus (and norovirus) infections is that protection from reinfection is mediated by intestinal IgA, which is not long-lasting in humans. Another explanation is that protection is dependent on neutralizing antibodies to one or both of the highly variable outer rotaviral proteins. However, a monovalent P[8]G1 vaccine induces significant protection against strains with different serotypes, thereby supporting the conclusion that protective immunity to rotavirus infection is, in large part, heterotypic.

CLINICAL MANIFESTATIONS

Norovirus

The clinical manifestations of norovirus infection are variable and depend in part on the age of the individual infected. About one third of infections are asymptomatic, but symptoms include diarrhea, nausea, vomiting, abdominal cramps, fever, and malaise that generally persist for 1 to 3 days. In children younger than 11 years, disease typically begins with the sudden onset of vomiting and can last 4 to 6 days. Virus can be shed in low titers for up to 8 weeks from otherwise healthy individuals and for more than a year in patients

with severe immunodeficiency syndromes. In neonates and premature infants, vomiting often is not a symptom, and infection has been associated with necrotizing enterocolitis. To support the diagnosis of norovirus outbreaks, the following four criteria have been proposed: (1) vomiting in more than half of affected persons; (2) mean (or median) incubation period of 24 to 48 hours; (3) mean (or median) duration of illness of 12 to 60 hours; and (4) absence of bacterial pathogens in stool culture.

Rotavirus

Rotavirus diarrhea and dehydration tend to be more severe than illness caused by the other childhood enteric pathogens. Rotavirus diarrhea is watery, persists for approximately 5 days, is often preceded by the sudden onset of vomiting, and is frequently accompanied by fever and dehydration. The incubation period of rotavirus is estimated to be less than 48 hours. Viral excretion in feces persists for 10 days in the majority of children and can persist for up to 57 days. Excretion times are longer on examination by sensitive PCR-based assays rather than solid-phase immunoassay. By the age of 5 years, virtually all children have acquired immunity to rotavirus, and severe disease after this age is uncommon.

DIAGNOSIS

Norovirus

Reverse transcription–PCR (RT-PCR) is currently the procedure of choice to detect norovirus in clinical specimens, in food, and in water. Although enzyme-linked immunosorbent assays (ELISAs) to detect noroviruses are available in Europe, their sensitivity is genotype dependent, and diagnostic specificity and sensitivity vary on the basis of the diversity of the circulating strains in the population. Moreover, these immunoassays are not easily adaptable for detection of new strains. Norovirus RNA is detected by RT-PCR in stool samples of up to 16% of healthy individuals, a finding that complicates the diagnosis of norovirus gastroenteritis. Although the relationship between disease symptoms and viral load has not been fully established, a quantitative real-time RT-PCR has been proposed to establish a relative threshold of positivity for attributing disease to norovirus.

Rotavirus

Before the introduction of the rotavirus vaccine in developed countries, well above 50% of the moderate to severe diarrheal episodes in young children during the rotavirus "season" were due to rotavirus. In tropical countries, the presence of other enteric pathogens and the absence of seasonal occurrence of rotaviral disease make it more difficult to determine which diarrheal episodes are caused by rotavirus without a diagnostic assay. Numerous ELISAs for rotavirus are commercially available, and these are generally sensitive, specific, and easy to use under most conditions. PCR has increased sensitivity for detection of rotavirus and has permitted easy typing of viruses. With PCR-based methods, however, up to 29% of healthy children younger than 1 year may be rotavirus positive, so it is difficult to associate the detection of the virus with gastroenteritis. Thus, ELISA or quantitative RT-PCR (using a threshold level as for norovirus) is preferable for diagnosis of rotavirus gastroenteritis.

PREVENTION

Norovirus

The development of a norovirus vaccine for humans is likely to be a challenge owing to the antigenic heterogeneity among circulating strains, the propensity of noroviruses to undergo antigenic drift, waning immunity, and the lack of well-established correlates of protection. Nevertheless, a new VLP-based vaccine candidate has recently shown promising efficacy in the experimental viral challenge setting.

High percentage alcohol-based sanitizers (99.5% ethanol) and 10% povidone-iodine antiseptics are superior to other alcohol-based sanitizers at reducing norovirus contamination. Simple household antimicrobial hand soap and handwashing with tap water also decrease viral contamination.

Rotavirus

The first vaccine (Rotashield [Wyeth-Lederle]), which was a quadrivalent mixture of rhesus-human rotavirus reassortants, each containing a G protein from a common human rotavirus serotype, was licensed for use in the United States but subsequently withdrawn from the market because of its association with intussusception. Subsequently, two second-generation vaccines were shown in large studies to be safe, effective, and cost-effective in developed and developing countries. The two current rotavirus vaccines have been based on two different approaches. One type of vaccine (RV5, RotaTeq [Merck]) uses a modified pentavalent vaccine made of a mixture of bovine and human reassortant rotaviruses. Another approach (RV1, Rotarix [GlaxoSmithKline]) is a monovalent attenuated human virus vaccine. Neither type of vaccine prevents subsequent rotavirus infection or mild illness, but both types effectively prevent severe illness, especially in developed countries.

Protection rates in developed and middle-income countries for each rotavirus vaccine are similar, varying from 70 to 80% against any rotavirus disease and 90 to 100% against severe gastroenteritis. Recent evidence suggests that both licensed vaccines slightly increase the risk of intussusception,[10,11] but their benefits far outweigh this low-level risk. Even in the United States, where the burden of severe disease from rotavirus is lowest, the rotavirus vaccine has significantly reduced health care utilization and expenses for diarrhea in children.

In the United States, vaccination also has had an unexpected effect on rotavirus-induced diarrhea among unvaccinated persons, thereby suggesting the induction of herd immunity. The U.S. Advisory Committee on Immunization Practices (Chapter 18) and the World Health Organization[12] now recommend the routine use of these vaccines.

Currently available vaccines are only about 50% efficacious against severe diseases in the poorest developing countries. Even with this reduced effectiveness, however, the two licensed vaccines are still cost-effective in the less developed world.

TREATMENT Rx

Because both norovirus and rotavirus disease resolves within days without treatment, the basic therapeutic goal is to prevent acute dehydration. The recommended oral rehydration salts solution, which now has an osmolarity of 331 mmol/L, is as effective as higher osmolarity solutions. After rehydration, rapid age-appropriate refeeding is recommended. Rotavirus disease induces self-limited intestinal lactase deficiency, but lactose-containing products, particularly maternal milk, should not be withheld.

Passive oral immunotherapy with diverse preparations of immunoglobulins can shorten the duration of rotavirus infection but probably is economically feasible only for immunodeficient patients or low-birthweight infants in the developed world. Recently in Bangladesh, llama-derived, heavy-chain antibody fragment specific for rotavirus was able to reduce stool output in male infants with severe rotavirus-associated diarrhea. *Lactobacillus*, a bacterium present in yogurt, is safe and, in limited studies, appears moderately effective for treatment of acute rotavirus gastroenteritis. Nonetheless, different preparations of lactobacilli vary greatly in dose of bacteria, and a general recommendation on their use has not been issued. Several studies in developing countries have shown that zinc supplementation (10 mg/day for infants younger than 6 months and 20 mg/day for older children) is useful for the treatment and prevention of diarrhea, but further studies are needed to determine whether treatment will be useful in all developing and developed countries.

At present, no pharmacologic treatment of rotavirus or norovirus diarrhea is recommended.[13] Racecadotril (4.5 mg/kg per day), an enkephalinase inhibitor that acts on the enteric nervous system, has been shown to be useful as an adjunct to treat rotaviral diarrhea in several small studies. Ondansetron (0.15 mg/kg per day), a serotonin antagonist, is effective in reducing the emesis from gastroenteritis during the phase of oral rehydration. In several small studies, nitazoxanide (15 mg/kg per day) was helpful in the treatment of rotavirus gastroenteritis. More studies are needed before any of these various preparations can be generally recommended for treatment of rotavirus diarrhea.

Grade A References

A1. Atmar RL, Bernstein DI, Harro CD, et al. Norovirus vaccine against experimental human Norwalk virus illness. *N Engl J Med*. 2011;365:2178-2187.

A2. Vesikari T, Matson DO, Dennehy P, et al. Safety and efficacy of a pentavalent human-bovine (WC3) reassortant rotavirus vaccine. *N Engl J Med*. 2006;354:23-33.

A3. Vesikari T, Karvonen A, Prymula R, et al. Efficacy of human rotavirus vaccine against rotavirus gastroenteritis during the first 2 years of life in European infants: randomised, double-blind controlled study. *Lancet*. 2007;370:1757-1763.

A4. Ruiz-Palacios GM, Pérez-Schael I, Velázquez FR, et al. Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis. *N Engl J Med*. 2006;354:11-22.

A5. Linhares AC, Velázquez FR, Pérez-Schael I, et al. Efficacy and safety of an oral live attenuated human rotavirus vaccine against rotavirus gastroenteritis during the first 2 years of life in Latin American infants: a randomised, double-blind, placebo-controlled phase III study. *Lancet*. 2008;371:1181-1189.

A6. Madhi SA, Cunliffe NA, Steele D, et al. Effect of human rotavirus vaccine on severe diarrhea in African infants. *N Engl J Med*. 2010;362:289-298.

A7. Zaman K, Dang DA, Victor JC, et al. Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in Asia: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2010;376:615-623.
A8. Armah GE, Sow SO, Breiman RF, et al. Efficacy of pentavalent rotavirus vaccine against severe rotavirus gastroenteritis in infants in developing countries in sub-Saharan Africa: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2010;376:606-614.
A9. Bhandari N, Rongsen-Chandola T, Bavdekar A, et al. Efficacy of a monovalent human-bovine (116E) rotavirus vaccine in Indian infants: a randomised, double-blind, placebo-controlled trial. *Lancet*. 2014;383:2136-2143.
A10. CHOICE Study Group. Multicenter, randomized, double-blind clinical trial to evaluate the efficacy and safety of a reduced osmolarity oral rehydration salts solution in children with acute watery diarrhea. *Pediatrics*. 2001;107:613-618.
A11. Sarker SA, Jäkel M, Sultana S, et al. Anti-rotavirus protein reduces stool output in infants with diarrhea: a randomized placebo-controlled trial. *Gastroenterology*. 2013;145:740-748.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Tate JE, Burton AH, Boschi-Pinto C, et al. 2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis. *Lancet Infect Dis*. 2012;12:136-141.
2. Glass RI, Parashar UD, Estes MK. Norovirus gastroenteritis. *N Engl J Med*. 2009;361:1776-1785.
3. Chhabra P, Payne DC, Szilagyi PG, et al. Etiology of viral gastroenteritis in children <5 years of age in the United States, 2008-2009. *J Infect Dis*. 2013;208:790-800.
4. Debbink K, Lindesmith LC, Donaldson EF, et al. Emergence of new pandemic GII.4 Sydney norovirus strain correlates with escape from herd immunity. *J Infect Dis*. 2013;208:1877-1887.
5. Payne DC, Vinje J, Szilagyi PG, et al. Norovirus and medically attended gastroenteritis in U.S. children. *N Engl J Med*. 2013;368:1121-1130.
6. Hall AJ, Lopman BA, Payne DC, et al. Norovirus disease in the United States. *Emerg Infect Dis*. 2013;19:1198-1205.
7. Kotloff KL, Nataro JP, Blackwelder WC, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet*. 2013;382:209-222.
8. Akelma AZ, Kutukoglu I, Koksal T, et al. Serum transaminase elevation in children with rotavirus gastroenteritis: seven years' experience. *Scand J Infect Dis*. 2013;45:362-367.
9. Steele D, Franco M, Angel J. Correlates of protection for rotavirus vaccines: possible alternative trial endpoints, opportunities, and challenges. *Hum Vaccin Immunother*. 2014;10:3659-3671.
10. Weintraub ES, Baggs J, Duffy J, et al. Risk of intussusception after monovalent rotavirus vaccination. *N Engl J Med*. 2014;370:513-519.
11. Yih WK, Lieu TA, Kulldorff M, et al. Intussusception risk after rotavirus vaccination in U.S. infants. *N Engl J Med*. 2014;370:503-512.
12. Rotavirus vaccines WHO position paper: January 2013—recommendations. *Vaccine*. 2013;31:6170-6171.
13. Bruzzese E, Lo Vecchio A, Guarino A. Hospital management of children with acute gastroenteritis. *Curr Opin Gastroenterol*. 2013;29:23-30.

REVIEW QUESTIONS

1. The principal agent of sporadic, severe gastroenteritis in young children worldwide is

A. Rotavirus
B. Norovirus
C. Astrovirus
D. All of the above
E. None of the above

Answer: A With the advent of rotavirus vaccines, norovirus has become the leading cause of medically attended acute gastroenteritis in the United States. Worldwide, however, rotavirus continues to be the first cause of severe gastroenteritis.

2. The most probable diagnosis of an adult with viral acute gastroenteritis related to a food-borne outbreak is

A. Rotavirus
B. Norovirus
C. Poliovirus
D. All of the above
E. None of the above

Answer: B Of these infections, only noroviruses are associated with food-borne gastroenteritis outbreaks and disease in adults.

3. Diagnosis of both norovirus and rotavirus acute gastroenteritis can optimally be performed with

A. Viral culture
B. Enzyme-linked immunosorbent assay (ELISA)
C. Quantitative reverse transcription–polymerase chain reaction (RT-PCR) testing using a threshold level
D. B or C
E. None of the above

Answer: C Quantitative RT-PCR (using a threshold level) is an optimal diagnostic strategy for both rotavirus and norovirus. RT-PCR for rotavirus without use of a threshold level will identify viral RNA in asymptomatic children. ELISA can be used for the diagnosis of rotavirus but not norovirus because these assays have variable sensitivity, depending on the norovirus genotype.

381 VIRAL HEMORRHAGIC FEVERS

DANIEL G. BAUSCH

DEFINITION

Viral hemorrhagic fever is an acute systemic illness classically involving fever, a constellation of initially nonspecific signs and symptoms, and a propensity for bleeding and shock. It may be caused by more than 30 different viruses from four taxonomic families, Filoviridae, Arenaviridae, Bunyaviridae, and Flaviviridae (Table 381-1), although not every virus in these families causes the syndrome. Hemorrhagic fever viruses are often named after the site of the first recognized case. All are single-stranded lipid-enveloped RNA viruses with small genomes (10 to 19 kilobases) that can be relatively easily inactivated in the environment. Pathogenicity varies widely by the specific virus and sometimes among strains of the same virus. Many of the hemorrhagic fever viruses have been placed on the Centers for Disease Control and Prevention select agents list of pathogens that pose a potential bioterrorism threat (Chapter 21).

EPIDEMIOLOGY

Maintenance in Nature and Transmission to Humans

With the exception of dengue virus, for which humans can now be considered the reservoir, hemorrhagic fever viruses are zoonotic, maintained in nature in mammalian reservoirs (see Table 381-1). Although viral hemorrhagic fevers collectively can be found worldwide, the endemic area of any given hemorrhagic fever virus is usually smaller than the extent of its natural reservoir or arthropod vector. With the exception of dengue and some hantaviruses, human infection is generally infrequent. Humans are dead-end hosts.

Hemorrhagic fever viruses may be transmitted to humans by usually inadvertent, direct exposure of mucous membranes or broken skin to the infected blood or excreta of its animal reservoir or, in the case of the flaviviruses and most of the bunyaviruses, by the bite of an arthropod vector. The infectious dose for most hemorrhagic fever viruses appears to be low, sometimes on the order of just a few virions. Aerosol transmission is not a predominant mode of spread, if it occurs at all, but studies in nonhuman primates show that transmission of many hemorrhagic fever viruses is possible through artificial aerosols, thereby raising the possibility of their potential use as bioweapons (Chapter 21).

Bat-Borne Viruses

The filoviruses (from the Latin *filo,* "thread," referring to their filamentous shape), Marburg and Ebola, are perhaps the most feared of all hemorrhagic fever viruses.[1] Fruit bats appear to be the filovirus reservoir, with transmission to humans likely from exposure to infected bat excreta or saliva. Nonhuman primates, especially gorillas and chimpanzees, and other wild animals may become infected, presumably from similar bat exposure, and transmit filoviruses to humans through contact with blood and body fluids of these animals, usually in association with hunting. Nonhuman primates, which are also dead-end hosts who develop severe and usually fatal disease similar to that seen in humans, may be easier prey for hunters when sick. Because hemorrhagic fever viruses are rapidly inactivated by heating, infection probably occurs by exposure during butchering and preparation. In the Philippines, Ebola Reston virus has been isolated from pigs that were presumably infected from exposure to bats.

In the recent outbreak in West Africa, which began in March, 2014, more than 20,000 cases and 8,000 deaths were confirmed by the end of the calendar year—numbers that dwarf all prior Ebola outbreaks combined.[2] The response has been hampered by limited facilities and trained medical personnel, some of whom have themselves become infected and died.

Rodent-Borne Viruses

Arenaviruses (from the Latin *arena,* "sand," referring to their sandy appearance on electron microscopy) are divided into two groups: the Old World (or lymphocytic choriomeningitis/Lassa) complex and the New World (or Tacaribe) complex.[3] Lassa virus and Lujo virus are found in Africa, whereas Junín, Machupo, Guanarito, Sabiá, and Chapare viruses are found in South America. Although there may be subtle differences among the syndromes produced by the New World arenaviruses, they are usually grouped together simply as the South American hemorrhagic fevers.

The genus Hantavirus of the Bunyaviridae family is similarly divided into Old and New World groups. The Old World hantaviruses, such as Hantaan, Seoul, and Puumala, among many others, cause hemorrhagic fevers with prominent renal involvement across Europe and Asia. New World hantaviruses, such as Sin Nombre and Andes, among many others, cause a viral hemorrhagic fever named hantavirus pulmonary syndrome, sometimes also called hantavirus cardiopulmonary syndrome to emphasize the significant cardiogenic component of this disease.

Pathogenic arenaviruses and hantaviruses are maintained in nature through chronic asymptomatic infection in rodents of the *Muridae* family, with a strict pairing between the specific virus and rodent species. Transmission between rodents may be by vertical or horizontal transmission or both, depending on the specific virus. Transmission to humans occurs through exposure to rodent excreta, either from aerosols produced when rodents urinate or by direct inoculation to the mucous membranes. Secondary aerosol generation is notoriously inefficient, so disturbing shed urine is a less likely mechanism of infection. In West Africa, Lassa virus is sometimes contracted when rodents are trapped and prepared for consumption or, more rarely, through a rodent bite. Experimental data suggest that humans may be infected with arenaviruses by the oral route.

The rodents that transmit Lassa, Machupo, and many of the Old World hantaviruses commonly invade the peridomestic environment, thereby putting housewives, children, and others who spend time at home at risk. In contrast, the reservoirs for Junín, Guanarito, and most of the New World hantaviruses typically inhabit agricultural fields, wood lots, or other rural habitats, thereby putting outdoor workers, campers, and hikers at risk.

Mosquito-Borne Viruses

Rift Valley fever virus is maintained in domestic livestock, such as cattle, buffalo, sheep, goats, and camels, in which it often provokes spontaneous abortion. The virus may be transmitted to humans by direct exposure to these animals, especially during parturition, or by mosquitoes. Farmers, abattoir workers, and veterinarians are at particular risk.[4]

Yellow fever virus is maintained in a cycle between monkeys and forest canopy mosquitoes. Sporadic cases occur when humans are bitten by these mosquitoes. Larger outbreaks occur when humans bring the virus back to more settled environments, where the urban mosquito *Aedes aegypti* can spread the virus directly between humans. *Ae. aegypti*, which typically lay eggs in artificial containers around the home and bite during the day, become infective a few weeks after feeding on a viremic monkey or human. Sanitation and mosquito control measures have virtually eliminated urban yellow fever in the Americas, but urban outbreaks continue to occur in Africa. For example, 849 cases and 171 deaths were reported in a recent outbreak in Sudan.[5]

Although nonhuman primates are also a reservoir for sylvatic strains of dengue, the virus is now largely maintained in humans, with a regular transmission cycle akin to that of urban yellow fever. Despite the presence of dengue virus in the tropics worldwide, less than 10% of infected persons develop hemorrhagic fever, primarily children between the ages of 4 and 12 years.

TABLE 381-1 PRINCIPAL VIRUSES CAUSING HEMORRHAGIC FEVER

VIRUS	DISEASE	GEOGRAPHIC DISTRIBUTION OF DISEASE	PRINCIPAL RESERVOIR/VECTOR	ANNUAL CASES	CASE TO INFECTION RATIO	HUMAN-TO-HUMAN TRANSMISSIBILITY
FILOVIRIDAE						
Ebolavirus[a]	Ebola HF	Sub-Saharan Africa	Fruit bat?	—[b]	1 : 1	High
Marburgvirus	Marburg HF	Sub-Saharan Africa	Fruit bat: Egyptian fruit bat (*Rousettus aegyptiacus*), perhaps others	—[b]	1 : 1	High
ARENAVIRIDAE[c,d]						
Old World Group						
Lassa	Lassa fever	West Africa	Rodent: natal mastomys or multimammate rat (*Mastomys natalensis*)	50,000-100,000	1 : 5-10	Moderate
Lujo[e]	Lujo HF	Zambia	Unknown, presumed rodent	Unknown	Unknown	Moderate to high
New World Group						
Junín	Argentine HF	Argentine pampas	Rodent: corn mouse (*Calomys musculinus*)	≈100	1 : 1.5	Low
Machupo	Bolivian HF	Beni department, Bolivia	Rodent: large vesper mouse (*Calomys callosus*)	≤50	1 : 1.5	Low
Guanarito	Venezuelan HF	Portuguesa state, Venezuela	Rodent: cane mouse (*Zygodontomys brevicauda*)	≤50	1 : 1.5	Low
Sabiá[f]	Proposed name: Brazilian HF	Rural area near São Paulo, Brazil?	Unknown, presumed rodent	Unknown	1 : 1.5	Low?
Chapare[g]	Chapare HF	Cochabamba, Bolivia	Unknown, presumed rodent	Unknown	Unknown	Unknown
BUNYAVIRIDAE[c]						
Old World Group						
Hantaan, Seoul, Puumala, Dobrava-Belgrade, others	HF with renal syndrome	Hantaan: northeast Asia Seoul: urban areas worldwide Puumala and Dobrava-Belgrade: Europe	Rodent Hantaan: striped field mouse (*Apodemus agrarius*) Seoul: brown or Norway rat (*Rattus norvegicus*) Puumala: bank vole (*Clethrionomys glareolus*) Dobrava-Belgrade: yellow-necked field mouse (*Apodemus flavicollis*)	50,000-150,000	Hantaan: 1 : 1.5 Others: 1 : 20	None

New World Group						
Sin Nombre, Andes, Laguna Negra, others	Hantavirus cardiopulmonary syndrome	Americas	Rodent Sin Nombre: deer mouse (*Peromyscus maniculatus*) Andes: long-tailed colilargo (*Oligoryzomys longicaudatus*) Laguna Negra: little laucha or small vesper mouse (*Calomys laucha*)	50,000-150,000	Sin Nombre: 1 : 1 Others: up to 1 : 20	None, except for Andes virus
Rift Valley fever	Rift Valley fever	Sub-Saharan Africa, Madagascar, Saudi Arabia, Yemen	Domestic livestock/mosquitoes (sylvatic *Aedes* and others)	100-100,000[b,h]	1 : 100	None
Crimean-Congo HF	Crimean-Congo HF	Africa, Balkans, southern Russia, Middle East, India, Pakistan, Afghanistan, western China	Wild and domestic vertebrates/tick (primarily *Hyalomma* species)	≈500	1 : 1-2	High
FLAVIVIRIDAE						
Yellow fever	Yellow fever	Sub-Saharan Africa, South America up to Panama	Monkey/mosquito (*Aedes aegypti*, other *Aedes* and *Haemagogus* species)	5000-200,000[i]	1 : 2-20	None
Dengue	Dengue HF	Tropics and subtropics worldwide	Human/mosquito (*Ae. aegypti* and *albopictus*)	100,000-200,000[i]	1 : 10-100, depending on age, previous infection, genetic background, and infecting serotype	None
Omsk HF	Omsk HF	Western Siberia	Rodent/tick (primarily *Dermacentor* and *Ixodes* species)	100-200	Unknown	Not reported
Kyasanur Forest disease	Kyasanur Forest disease	Karnataka state, India; Yunnan Province, China; Saudi Arabia	Vertebrate (rodents, bats, birds, monkeys, others)/tick (*Haemaphysalis* species and others)	≈500	Unknown	Not reported, but laboratory infections have occurred
Alkhumra HF[j]	Proposed name: Alkhumra HF	Saudi Arabia, Egypt	Ticks?	≤50	Unknown	Not reported

[a]Six species or subtypes of Ebolavirus are recognized with varying associated case-fatality ratios (see Table 381-2). All are endemic to sub-Saharan Africa, with the exceptions of Reston ebolavirus, which is found in the Philippines, and Lloviu ebolavirus, which was detected in bats in Spain.
[b]Although some endemic transmission of the filoviruses (Ebolavirus > Marburgvirus) and Rift Valley fever virus occurs, these viruses have most often been associated with outbreaks.
[c]The virus families Arenaviridae and Bunyaviridae are serologically, phylogenetically, and geographically divided into Old World (i.e., Africa) and New World (i.e., the Americas) complexes.
[d]In addition to the arenaviruses listed in the table, Flexal and Tacaribe viruses have caused human disease as a result of laboratory accidents. Another arenavirus, Whitewater Arroyo, has been noted in sick persons in California, but its role as a pathogen has not been clearly established.
[e]Discovered in 2008. Only five cases (four of them fatal) from one outbreak have been noted. The index case came to South Africa from Zambia.
[f]Discovered in 1990. Only three cases (one fatal) have been noted, two of them from laboratory accidents.
[g]Discovered in 2003 from a small outbreak from which blood was obtained from one fatal case and Chapare virus isolated. Few other details have been reported.
[h]Although Rift Valley fever virus can be found throughout sub-Saharan Africa, large outbreaks usually occur in East Africa's Rift Valley region.
[i]Based on estimates from the World Health Organization. Significant underreporting occurs. Incidence may fluctuate widely in place and time.
[j]Alkhumra is considered by some to be a variant of Kyasanur Forest disease virus. Disagreement exists over the proper spelling of the virus, written as Alkhurma in some publications.
HF = hemorrhagic fever.

Tick-Borne Viruses

The viruses that cause Crimean-Congo hemorrhagic fever, Omsk hemorrhagic fever,[6] Kyasanur Forest disease, and Alkhumra hemorrhagic fever are maintained in small mammals, such as rodents, hares, and hedgehogs, among which the viruses are spread by ticks. Humans are infected either by tick bites or by exposure to contaminated blood or excreta of the reservoir animals. Ticks also spread Crimean-Congo hemorrhagic fever virus to large mammals, including cattle and other domestic livestock, whose transient and asymptomatic viremia puts farmers, abattoir workers, and veterinarians at risk.

Human-to-Human Transmission

Secondary human-to-human transmission occurs with many of the hemorrhagic fever viruses, but tertiary transmission is unusual and often associated with milder disease (see Table 381-1). Secondary attack rates for hemorrhagic fever viruses are generally low (15 to 20% for Ebola Zaire virus), probably because transmission between humans requires direct contact with contaminated blood or body fluids. Human-to-human infection probably usually occurs through oral or mucous membrane exposure, most often in the context of providing care to a sick family member (community) or patient (nosocomial transmission), and occasionally during funeral rituals that entail the touching of the corpse, especially for the filoviruses. Infection through fomites cannot be excluded. Aerosol infection is thought to be rare or nonexistent. Large outbreaks almost always involve amplification in health care settings in which basic infection control measures have broken down, usually owing to extreme poverty or civil strife.

With the exception of hantaviruses and some of the flaviviruses, infectivity generally parallels the clinical state. Persons are generally most infectious late in the course of severe disease, especially when bleeding. The risk of transmission during the incubation period or from asymptomatic persons is negligible, although a case of Argentine hemorrhagic fever occurred from blood transfusion from an asymptomatic donor. Rarely, Ebola, Marburg, Lassa, and Junín viruses have been sexually transmitted during the first 3 months of convalescence owing to delayed viral clearance from the gonads, which is an immunologically protected site. Despite modern-day travel, imported cases of viral hemorrhagic fever remain extremely rare.

PATHOBIOLOGY

Although the precise mechanism varies with the specific virus, microvascular instability and impaired hemostasis are the pathobiologic hallmarks of viral hemorrhagic fever. Data from animal models suggest that cardiac inotropy may also be directly or indirectly inhibited in some viral hemorrhagic fevers, especially Lassa fever.

After inoculation, virus first replicates in dendritic cells and other local tissues, with subsequent migration to regional lymph nodes and then dissemination through the lymph and blood monocytes to a broad range of tissues and organs, including the liver, spleen, lymph nodes, adrenal glands, lungs, and endothelium. Migration of tissue macrophages results in secondary infection of permissive parenchymal cells. During the acute illness, virus can be found in a wide variety of body fluids, including blood, saliva, stool, and breast milk.

The interaction of virus with immune cells, especially macrophages and endothelial cells, results directly or indirectly (through soluble mediators) in cell activation and the unleashing of an inflammatory and vasoactive process consistent with the systemic inflammatory response syndrome. The synthesis of cell surface tissue factor triggers the extrinsic coagulation pathway. Impaired hemostasis may entail endothelial cell, platelet, or coagulation factor dysfunction. Disseminated intravascular coagulopathy (DIC) is frequently noted, especially with Ebola, Marburg, and Crimean-Congo hemorrhagic fever virus infections.

Tissue damage may be mediated through direct necrosis of infected cells or indirectly through apoptosis of immune cells, as seen in other forms of septic shock. The most affected organs vary with the virus (Table 381-2). For example, renal tubular necrosis and retroperitoneal edema are seen in hemorrhagic fevers with renal syndrome, whereas interstitial pneumonitis and myocardial depression are the hallmarks of hantavirus pulmonary syndrome. The liver is particularly affected in yellow fever, with fatty degeneration, coagulative midzonal necrosis of hepatocytes, and the presence of Councilman bodies. The brain and meninges are particularly affected in Kyasanur Forest disease and Omsk hemorrhagic fever and often in the South American hemorrhagic fevers as well. Reticuloendothelial proliferation is seen in Kyasanur Forest disease, with marked erythrophagocytosis in the spleen.

With the exception of disease caused by the hantaviruses and some of the flaviviruses, the pathogenesis of viral hemorrhagic fever appears to be related to unchecked viremia, with most fatal cases failing to mount a significant antibody response. By comparison, virus is cleared rapidly from the blood in survivors. Inflammatory cell infiltrates, which are usually mild, consist of a mix of mononuclear cells and neutrophils. In some viral hemorrhagic fevers, such as Ebola, virus replication and dissemination are facilitated by virus-induced suppression of the host adaptive immune response. For example, failure of the immune response to adequately respond appears to be a major determinant of severity in Lassa fever.[7]

In dengue, yellow fever, and hantavirus infections, in which viremia is usually cleared before the most severe phase of the disease, the host immune response may play a detrimental role. The unique process of antibody-mediated immune enhancement, in which secondary infection with a different dengue virus serotype is more severe than the primary one, may play a role in the pathogenesis of dengue hemorrhagic fever.

CLINICAL MANIFESTATIONS

Viral hemorrhagic fever is seen in both genders and all age groups, with a spectrum from relatively mild or even asymptomatic infection to severe vascular permeability resulting in shock, multiorgan system failure, and death. Although the clinical presentation may differ for each viral hemorrhagic fever as disease progresses, the limited data do not permit clear distinctions in most cases, especially in the early phases of disease. Dengue and Rift Valley fever viruses cause a range of syndromes, including rash and central nervous system involvement (Chapters 382 and 383). Hemorrhagic fever occurs in a minority of infections with these viruses.

After an incubation period ranging from days to weeks, most patients present with nonspecific signs and symptoms difficult to distinguish from a host of other febrile illnesses (see Table 381-2)[8], including fever, general malaise, anorexia, headache, chest or retrosternal pain, sore throat, myalgia, arthralgia, and lumbosacral pain. Conjunctival injection or hemorrhage is frequent but is not accompanied by itching, discharge, or rhinitis (Fig. 381-1). Relative bradycardia (Faget sign) and orthostatic hypotension may be noted, especially in yellow fever and dengue virus infections. The pharynx may be erythemic or, less frequently, exudative, especially in Lassa fever, and incorrectly lead to a diagnosis of streptococcal pharyngitis or mononucleosis. Gastrointestinal signs and symptoms readily ensue, including nausea, vomiting, epigastric and abdominal pain, abdominal tenderness (especially over the liver in filovirus infection), and nonbloody diarrhea or constipation. A misdiagnosis of appendicitis or other acute abdominal emergency (Chapter 142) sometimes prompts potentially hazardous surgical interventions.

In the recent Ebola outbreak in West Africa, the incubation period averaged 6 to 8 days. Most patients presented with some combination of fever (90%), headache (80%), diarrhea (50%), or vomiting (35%), often associated with myalgias, weakness, or abdominal pain. Hiccups also may be seen early in Ebola hemorrhagic fever.[9]

Neck pain and stiffness, retro-orbital pain, photophobia, and other meningeal signs are common in Rift Valley fever, Kyasanur Forest disease, and Omsk hemorrhagic fever. A dry cough, sometimes accompanied by a few scattered rales on auscultation, is common, but prominent pulmonary symptoms are uncommon early in the course of the disease, except with hantavirus pulmonary syndrome. Pregnant women often present with spontaneous abortion and vaginal bleeding. With the exception of yellow fever, jaundice is not typical except in patients with underlying Gilbert syndrome, drug reactions, or coinfection. Hepatosplenomegaly is frequent, but whether it is specific to the viral hemorrhagic fever or simply represents the high underlying prevalence of hepatosplenomegaly in populations in sub-Saharan Africa is unknown.

Various forms of rash, including morbilliform, maculopapular, petechial, and ecchymotic, may be seen (see Table 381-2). A maculopapular rash on the torso or face may be one early and relatively specific although insensitive indicator of Ebola or Marburg hemorrhagic fever. Rash almost always occurs in fair-skinned persons with Lassa fever but, for unclear reasons, rarely in blacks.

In severe cases, patients progress after 7 to 10 days of illness to vascular instability, which may be manifested by conjunctival injection and hemorrhage, facial flushing, edema, bleeding, hypotension, shock, and proteinuria. Facial and neck swelling are classic and relatively specific signs of Lassa fever and Lujo hemorrhagic fever.[10] The likelihood of clinically discernible hemorrhage varies with the infecting virus (see Table 381-2) and may be manifested as hematemesis, melena, hematochezia, metrorrhagia, petechiae, purpura,

TABLE 381-2 PATHOBIOLOGIC AND CLINICAL ASPECTS OF VIRAL HEMORRHAGIC FEVERS

DISEASE	INCUBATION PERIOD (DAYS)	ONSET	BLEEDING	RASH	JAUNDICE	HEART	LUNG	KIDNEY	CENTRAL NERVOUS SYSTEM	EYE	CASE-FATALITY RATIO	CLINICAL MANAGEMENT
FILOVIRIDAE												
Ebola HF	3-21	Variable	++	+++	+	++?	+	+	+	+	40-85%[a]	Supportive
Marburg HF	3-21	Abrupt	++	+++	+	++?	+	+	+	+	22-85%[b]	Supportive
ARENAVIRIDAE												
Lassa fever	5-16	Gradual	+	+[c]	0	++	+	0	+	0	20%	Ribavirin
Lujo HF	9-13	Abrupt	++	+	0	?	+	+	+	0	80%	Ribavirin
South American HFs[d]	4-14	Gradual	+++	+	0	++	+	0	+++	0	15-40%	Ribavirin, convalescent plasma
BUNYAVIRIDAE												
Hemorrhagic fever with renal syndrome	9-35	Abrupt	+++	0	0	++	+	+++	+	0	<1-50%. depending on specific virus	Ribavirin
Hantavirus pulmonary syndrome	7-35	Gradual	0 (except for Andes virus infection)	0	0	+++	+++	+	+	0	<1-50%. depending on specific virus	Supportive, ECMO?
Rift Valley fever[e]	2-5	Abrupt	++	+	++	+?	0	+	++	++	Up to 50% in severe forms	Ribavirin?
Crimean-Congo HF	1-12[f]	Abrupt	+++	0	++	+?	+	0	+	0	15-30%	Ribavirin
FLAVIVIRIDAE												
Yellow fever	3-6	Abrupt	+++	0	+++	++	+	++	++	0	20-50%	Supportive
Dengue HF	3-15	Abrupt	++	+++	+	++	+	0	+	0	Untreated: 10-15% Treated: ≤1%	Supportive
Omsk HF	3-8	Abrupt	++	0	0	+	++	0	+++	+	1-3%	Supportive
Kyasanur forest disease	3-8	Abrupt	++	0	0	+	++	0	+++	+	3-5%	Supportive
Alkhumra HF[g]	3-8	Abrupt	++	+	+	+	+	0	++	+	20-25%	Supportive

[a]Six species or subtypes of Ebolavirus are recognized with varying associated case-fatality ratios: Zaire, 85%; Sudan, 55%; Bundibugyo, 40%; Tai Forest (also called Côte d'Ivoire), 0% (only one recognized case, who survived); Reston, 0% (not pathogenic to humans); Lloviu, no human infections recognized.
[b]The case-fatality ratio was 22% in the first recognized outbreak of Marburg HF in Germany and Yugoslavia in 1967 but has been consistently above 80% in outbreaks in central Africa, where the virus is endemic. Possible reasons for this discrepancy include differences in quality of care, strain pathogenicity, route and dose of infection, underlying prevalence of immunodeficiency and comorbid illnesses, and genetic susceptibility.
[c]A morbilliform or maculopapular rash almost always occurs in persons with lighter skin, who are usually expatriates, but for unclear reasons is rarely present in darker-skinned Africans from the endemic area.
[d]Data are insufficient to distinguish between the syndromes produced by the various arenaviruses found in the Americas. They are thus frequently grouped as the South American hemorrhagic fevers.
[e]Hemorrhagic fever, encephalitis, and retinitis may be seen in Rift Valley fever independently of each other.
[f]The incubation period of Crimean-Congo HF varies with the mode of transmission: typically 1 to 3 days after tick bite and 5 to 6 days after contact with infected animal blood or tissues.
[g]Based on preliminary observations. Fewer than 100 cases have been reported.
ECMO = extracorporeal membrane oxygenation; HF = hemorrhagic fever; 0 = sign not typically noted/organ not typically affected; + = sign occasionally noted/organ occasionally affected; ++ = sign commonly noted/organ commonly affected; +++ = sign characteristic/organ involvement severe.

epistaxis, and bleeding from the gums and venipuncture sites (Fig. 381-2). Hemoptysis and hematuria are infrequent. Hemorrhage is almost never present in the first few days of illness. Large ecchymoses are characteristic of Crimean-Congo hemorrhagic fever. Central nervous system manifestations, including delirium, tremor, gait anomalies, convulsions, and hiccups, may be noted in end-stage disease, especially in Kyasanur Forest disease, Omsk hemorrhagic fever, and the South American hemorrhagic fevers, particularly Argentine hemorrhagic fever. Nevertheless, the cerebrospinal fluid findings are usually normal, with the exception of patients who have meningoencephalitis due to Kyasanur Forest disease and Omsk hemorrhagic fever, in which an elevated protein level is common. Renal insufficiency or failure is common, especially in hemorrhagic fever with renal syndrome.

Biphasic illnesses are classically noted for the flavivirus hemorrhagic fevers, in which a quiescent period of days (yellow fever, dengue hemorrhagic fever, and Rift Valley fever) to weeks (Kyasanur Forest disease and Omsk hemorrhagic fever) precedes the most severe manifestations, including hemorrhage, shock, renal failure, and meningoencephalitis. Distinct progressive phases of disease and recovery are classically described for hemorrhagic fever with renal syndrome (prodrome, hypotension, oliguria/renal failure, diuresis, and convalescence) and yellow fever (infection, intoxication, recovery) but are not seen in all cases. The initial manifestations for hantavirus pulmonary syndrome may be mild and nonspecific, but the disease may progress to require mechanical ventilation and pressor support within 24 hours; sinus bradycardia and ventricular tachycardia or fibrillation may occur. Encephalitis and retinitis may develop in Rift Valley fever independently of the presence or absence of viral hemorrhagic fever.

Bilateral noncardiogenic interstitial pulmonary edema consistent with the adult respiratory distress syndrome (ARDS) is the hallmark of hantavirus pulmonary syndrome, although chest radiographs may be normal early in the disease even when the patients complain of shortness of breath. Only about 30% of patients with hantavirus pulmonary syndrome have radiographic evidence of pulmonary edema on initial evaluation, although it develops in virtually all persons within 48 hours.

DIAGNOSIS

Because of their associated severity, risk of secondary spread, high degree of public scrutiny, and unfamiliarity to most physicians, consultation with a specialist who has experience with viral hemorrhagic fevers should be sought as soon as the diagnosis is considered. When to "sound the alarm" of viral hemorrhagic fever is a case-by-case decision left to the treating physician in consultation with experts in the field. Most viral hemorrhagic fevers are rare, and routinely practiced universal precautions are protective in most cases.

The early nonspecific presentation of viral hemorrhagic fevers makes them extremely difficult to diagnose clinically, especially outside of the setting of a recognized outbreak, which is usually detected when clusters of cases occur, especially when they involve health care workers. The differential diagnosis includes a broad array of febrile illnesses that varies by geographic region (Table 381-3). A complete epidemiologic history (including details of travel,

FIGURE 381-1. Subconjunctival hemorrhage and facial swelling in a boy with Lassa fever in Sierra Leone.

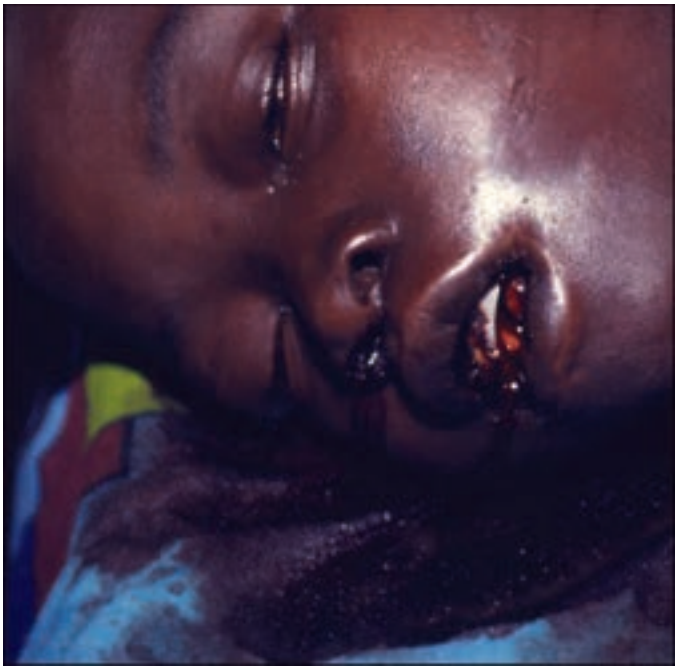

FIGURE 381-2. Bleeding in a patient with Ebola hemorrhagic fever. (From Bausch DG. Viral hemorrhagic fevers. In: Schlossberg D, ed. *Clinical Infectious Disease.* New York: Cambridge University Press; 2008.)

TABLE 381-3 DIFFERENTIAL DIAGNOSIS OF THE VIRAL HEMORRHAGIC FEVERS

DISEASE	DISTINGUISHING CHARACTERISTICS AND COMMENTS
PARASITES	
Malaria	Classically shows paroxysms of fever and chills; hemorrhagic manifestations less common; malaria smears or rapid test result usually positive; coinfection (or baseline asymptomatic parasitemia) common; responds to antimalarials
Amebiasis	Hemorrhagic manifestations other than bloody diarrhea generally not seen; amebic trophozoites identified in the stool by microscopy or antigen assays; responds to antiparasitics
Giardiasis	Positive stool antigen test result or identification of trophozoites or cysts in stool; responds to antiparasitics
African trypanosomiasis (acute stages)	Especially the East African form; examination of peripheral blood smear/buffy coat may show trypanosomes
BACTERIA (INCLUDING SPIROCHETES, RICKETTSIA, EHRLICHIA, AND COXIELLA)	
Typhoid fever	Hemorrhagic manifestations other than bloody diarrhea generally not seen; responds to antibiotics
Bacillary dysentery (including shigellosis, campylobacteriosis, salmonellosis, and enterohemorrhagic *Escherichia coli* and others)	Hemorrhagic manifestations other than bloody diarrhea generally not seen; responds to antibiotics
Capnocytophaga canimorsus	Associated with dog and cat bites, typically in persons with underlying immunodeficiency, notably asplenic patients; responds to antibiotics
Meningococcemia	Bacterial-induced DIC may mimic the bleeding diathesis of VHF; bleeding within the first 24-48 hours after onset of illness and rapidly progressive illness typical; large ecchymoses typical of meningococcemia are unusual in the VHFs except for Crimean-Congo HF; rapid serum latex agglutination tests can be used to detect bacterial antigen in meningococcal septicemia; may respond to antibiotics (critical to administer early)
Staphylococcemia	Bacterial-induced DIC may mimic the bleeding diathesis of VHF; may respond to antibiotics
Septic abortion	History of pregnancy and positive pregnancy test
Septicemic or pneumonic plague	Bacterial-induced DIC may mimic the bleeding diathesis of VHF; large ecchymoses typical of plague are unusual in the VHFs except for Crimean-Congo HF; pneumonic plague may mimic hantavirus pulmonary syndrome; may respond to antibiotics
Streptococcal or Epstein-Barr virus pharyngitis	May mimic the exudative pharyngitis sometimes seen in Lassa fever
Tuberculosis	Hemoptysis of advanced pulmonary tuberculosis may suggest VHF, but tuberculosis generally has a much slower disease evolution
Tularemia	Ulceroglandular and pneumonic forms more common; responds to antibiotics
Acute abdominal emergencies	Appendicitis, peritonitis, and bleeding upper gastrointestinal ulcer
Pyelonephritis and post-streptococcal glomerulonephritis	May mimic HF with renal syndrome
Anthrax (inhalation or gastrointestinal)	Prominent pulmonary manifestations and widened mediastinum on chest radiograph in inhalation form; responds to antibiotics
Atypical bacterial pneumonia (*Legionella, Mycoplasma, Chlamydophila pneumoniae* and *C. psittaci*, others)	May mimic hantavirus pulmonary syndrome; exposure to birds; symptoms often not present until late in the illness in psittacosis; responds to antibiotics

TABLE 381-3 DIFFERENTIAL DIAGNOSIS OF THE VIRAL HEMORRHAGIC FEVERS—cont'd

DISEASE	DISTINGUISHING CHARACTERISTICS AND COMMENTS
Relapsing fever	Recurrent fevers and influenza-like symptoms, with direct neurologic involvement and splenomegaly; spirochetes visible in blood while febrile; responds to antibiotics
Leptospirosis	Jaundice, renal failure, and myocarditis in severe cases; responds to antibiotics
Spotted fever group rickettsiae (including African tick bite fever, boutonneuse fever, Rocky Mountain spotted fever)	Incubation period of 7-10 days after tick bite, compared with 1-3 days in Crimean-Congo HF; necrotic lesions (eschar) typically seen at site of tick bite in some rickettsial diseases, whereas there may be only slight bruising at the bite site in Crimean-Congo HF; rash (if present) of rickettsial infection classically involves palms and soles
Q fever (*Coxiella burnetii*)	Broad spectrum of illness, including hepatitis, pneumonitis, encephalitis, and multisystem disease with bleeding; responds to antibiotics
Ehrlichiosis	Diagnosis by serology and PCR; blood film may be useful; responds to antibiotics
VIRUSES	
Influenza	Prominent respiratory component to clinical presentation; no hemorrhagic manifestations; influenza rapid test result may be positive; may respond to anti-influenza drugs
Arbovirus infection (including dengue and West Nile fever)	Encephalitis unusual but when present may mimic the VHFs with significant neurologic involvement (Kyasanur Forest disease, Omsk HF); usually less severe than VHF; hemorrhage not reported
Viral hepatitis (including hepatitis A, B, and E; Epstein-Barr; and cytomegalovirus)	Jaundice atypical in HF except yellow fever; test results for hepatitis antigens positive; fulminant infection resembling VHF may be seen in persons with underlying immune deficiencies
Herpes simplex or varicella-zoster	Fulminant infection with hepatitis (with or without vesicular rash); elevated transaminases and leukopenia typical; disseminated disease may be noted in otherwise healthy persons; poor response to acyclovir drugs unless recognized early
HIV/AIDS	Seroconversion syndrome or HIV/AIDS with secondary infections, especially septicemia
Measles	Rash may mimic that seen in early stages of some VHFs and may sometimes be hemorrhagic; prominence of coryza and upper respiratory symptoms in measles should help differentiate; vaccine preventable
Rubella	Rash may mimic that seen in early stages of some VHFs; usually a mild disease; vaccine preventable
Hemorrhagic or flat smallpox	Diffuse hemorrhagic or macular lesions; in contrast to the VHFs, the rash may involve the oral mucosa, palms, and soles; smallpox in the wild has been eradicated
Alphavirus infection (including chikungunya and o'nyong-nyong)	Joint pain typically a predominant feature
FUNGI	
Histoplasmosis	Pulmonary disease may mimic hantavirus pulmonary syndrome; recent entry into mines or caves
NONINFECTIOUS ETIOLOGIES	
Heat stroke	History for extreme heat exposure; absence of sweating; bleeding not typical, but DIC may occur
Idiopathic and thrombotic thrombocytopenic purpura (ITP/TTP)	Presentation usually less acute than in VHF; may have prominent neurologic symptoms in TTP; coagulation factors normal and DIC absent; often respond to corticosteroids (ITP) or plasma exchange (TTP)
Acute glaucoma	May mimic the acute ocular manifestations of Rift Valley fever
Hematologic malignant neoplasms (leukemia, lymphoma)	May resemble leukemoid reaction occasionally seen in HF with renal syndrome
Drug sensitivity or overdose	Stevens-Johnson syndrome and anticoagulant (warfarin) overdose
Industrial and agricultural chemical poisoning	Especially anticoagulants, although other symptoms of VHF absent
Hematoxic snake bite envenomation	History of snake bite

DIC = disseminated intravascular coagulopathy; HF = hemorrhagic fever; PCR, polymerase chain reaction; VHF, viral hemorrhagic fever.

possible exposures, occupational risks, and the progression of illness), physical examination, and preliminary basic laboratory results (Table 381-4) are critical. A diagnosis of viral hemorrhagic fever should be considered in patients with a clinically compatible syndrome who, within the incubation period for the particular viral hemorrhagic fever in question, (1) reside in or traveled to an endemic area (see Table 381-1); (2) had potential direct contact with blood or body fluids of someone who was ill with an acute viral hemorrhagic fever, such as health care workers, persons caring for family members at home or preparing bodies for burial, and laboratory personnel; (3) had contact with live or recently killed wild animals (especially nonhuman primates) in or recently arriving from an area where a viral hemorrhagic fever is endemic (although direct contact with the animal reservoir is not usually reported even in confirmed cases); (4) worked in a laboratory or animal facility where hemorrhagic fever viruses are handled; or (5) had sexual relations with someone recovering from a viral hemorrhagic fever in the last 3 months.

The index of suspicion should be especially high for persons in specific high-risk occupations, including health care workers, abattoir workers, veterinarians, farm workers, hunters, taxidermists, and travelers who have recently returned from endemic areas.[11] ARDS (Chapter 104) or other respiratory compromise in a person living in an endemic area for New World hantaviruses should prompt consideration of hantavirus pulmonary syndrome. Risk of tick infection, including physical examination for an eschar, should be assessed if a tick-borne viral hemorrhagic fever is suspected. However, most viral hemorrhagic fevers are rare even in persons possessing one of the risk factors, so alternative diagnoses should always be aggressively sought, especially malaria and typhoid fever in areas where they are endemic. Acts of bioterrorism (Chapter 21) must be considered if viral hemorrhagic fever is strongly suspected in a patient without any of the aforementioned risk factors, especially if clusters of cases occur. All cases should be immediately reported to local, state, and federal health authorities.

Laboratory Testing

Prompt laboratory confirmation is imperative, but testing is unfortunately only available in a few specialized laboratories except for kits with varying sensitivity and specificity for the serologic diagnosis of dengue fever and hantavirus pulmonary syndrome. Assays commonly used in the diagnosis of viral hemorrhagic fever include polymerase chain reaction,[12] enzyme-linked immunosorbent assays for viral antigen and immunoglobulin M antibody, virus culture, and immunohistochemistry on postmortem tissues. These assays generally appear to have sensitivities and specificities above 90%, although serologic diagnosis of flavivirus infection is often complicated by cross-reactions. In the United States, testing can be arranged through the Centers for Disease Control and Prevention (phone: 404-639-1115, after hours: 770-488-7100; e-mail: dvd-1spath@cdc.gov).

TABLE 381-4 INDICATED CLINICAL LABORATORY TESTS AND CHARACTERISTIC FINDINGS IN PATIENTS WITH VIRAL HEMORRHAGIC FEVER

TEST	CHARACTERISTIC FINDINGS AND COMMENTS
Leukocyte count	Early: moderate leukopenia (except for hantavirus infection, in which early leukocytosis with immunoblasts is classically noted) Later: leukocytosis with left shift; granulocytosis more suggestive of bacterial infection
Hemoglobin and hematocrit	Hemoconcentration (especially noted in hemorrhagic fever with renal syndrome and hantavirus pulmonary syndrome)
Platelet count	Mild to moderate thrombocytopenia
Electrolytes	Sodium, potassium, and acid-base perturbations, depending on fluid balance and stage of disease
BUN/creatinine	Renal failure may occur late in disease.
Serum chemistries (AST, ALT, amylase, γ-glutamyltransferase, alkaline phosphatase, creatinine kinase, lactate dehydrogenase, lactate)	Usually increased, especially in severe disease; AST > ALT A lactate level >4 mmol/L (36 mg/dL) may indicate persistent hypoperfusion and sepsis. Lactate dehydrogenase is typically markedly increased in hantavirus pulmonary syndrome.
Sedimentation rate	Normal or increased
Blood gas	Metabolic acidosis may be indicative of shock and hypoperfusion
Coagulation studies (PT, PTT, fibrinogen, fibrin split products, platelets, D-dimer)	DIC common in Ebola, Marburg, Lujo virus, Crimean-Congo HF, and New World arenavirus infections
Urinalysis	Proteinuria common; hematuria may be occasionally noted Sediment may show hyaline-granular casts and round cells with cytoplasmic inclusions.
Blood culture	Useful early to exclude VHF and later to evaluate for secondary bacterial infection Blood should be drawn before antibiotic therapy is instituted.
Stool culture	Useful to exclude VHF (in favor of hemorrhagic bacillary dysentery)
Thick and thin blood smears	May aid in the diagnosis of blood parasites (malaria and trypanosomes), bacterial sepsis (meningococcus, capnocytophaga, and anthrax), and ehrlichiosis All negative in VHF unless coinfection
Rapid test, PCR, or other assay for malaria	Negative in VHF unless coinfection with malaria

ALT = alanine aminotransferase; AST = aspartate aminotransferase; BUN = blood urea nitrogen; DIC = disseminated intravascular coagulation; PCR, polymerase chain reaction; PT = prothrombin time; PTT = partial thromboplastin time; VHF = viral hemorrhagic fever.

TREATMENT

Rx

The rarity of most viral hemorrhagic fevers and their typical occurrence in remote and resource-poor settings make controlled studies on treatment difficult. Treatment guidelines generally follow those recommended for septic shock (Chapter 108). Patients should be placed in isolation in an intensive care unit. Intramuscular and subcutaneous injections should be minimized because of the risk of hematoma.

Clinical Management Guidelines

Fluid and Electrolyte Management

Severe microvascular instability, often complicated by vomiting, severe and sometimes voluminous diarrhea, and decreased fluid intake, typically requires aggressive fluid replacement to prevent shock. In the recent Ebola outbreak, patients often required up to 5 liters of intravenous fluids daily and potassium supplementation in addition to oral rehydration.[13,14] However, overaggressive and unmonitored rehydration may lead to significant third-spacing and pulmonary edema, especially in hantavirus pulmonary syndrome.

Early goal-directed therapies with crystalloids, blood products, and vasopressors can mitigate organ dysfunction and probably reduce mortality (Chapter 108). Peritoneal dialysis and hemodialysis (Chapter 131) have been extensively used in patients with hemorrhagic fever with renal syndrome without frequent complications, but there is little published experience with the other viral hemorrhagic fevers. Specific World Health Organization guidelines for the fluid management of dengue shock syndrome in resource-limited environments specify a more conservative fluid repletion strategy because of the inability to monitor hydration status carefully and the resulting risk of fluid overload or pulmonary edema in this syndrome. Significant electrolyte imbalance may be noted, and hypokalemia often requires potassium supplementation.

Blood Products and Management of Disseminated Intravascular Coagulation

Although bleeding may be profuse in some viral hemorrhagic fevers, especially Crimean-Congo hemorrhagic fever and Ebola and Marburg hemorrhagic fevers, blood products should not be given empirically but only to meet defined clinical and laboratory parameters in the face of clinically significant hemorrhage. Transfusions, preferably with packed red blood cells, should be used to maintain a hemoglobin concentration above 7.0 g/dL while avoiding volume overload, taking into account that chronic anemia due to malaria and malnutrition may be frequent in patients in certain geographic areas. Whole blood is a reasonable alternative if packed cells are not available.

The possibility of DIC (Chapter 175) should be assessed by the relevant laboratory parameters (see Table 381-4), such as D-dimer levels. Transfusion of platelet concentrate (1 to 2 U/10 kg) should be considered when the platelet count is less than 50,000/μL in a bleeding patient or less than 20,000/μL without bleeding. The platelet count should generally rise by at least 2000/μL per unit of platelets transfused, although the response may be less if there is ongoing DIC and platelet consumption. Impaired platelet aggregation may promote hemorrhage in some viral hemorrhagic fevers, especially Lassa fever, even when platelet counts are not drastically low. Transfusion of fresh-frozen plasma (FFP) (15 to 20 mL/kg) should be considered when bleeding is present and fibrinogen levels are less than 100 mg/dL. Fibrinogen concentrate (total dose of 2 to 3 g) or cryoprecipitate (1 U/10 kg) may be administered instead of FFP, although FFP has the theoretical advantage of containing all coagulation factors and inhibitors deficient in DIC but no activated coagulation factors. Vitamin K (10 mg intravenously or orally on 3 consecutive days) may be given, especially if underlying malnutrition or liver disease is suspected. Folic acid has also sometimes been added to prevent the detrimental effect of acute folate deficiency on platelet production, especially in malnourished patients, although the efficacy of this treatment is unknown.

Oxygenation and Ventilation

In the early phases of disease and in the absence of iatrogenic pulmonary edema, most patients can be supported with oxygen administered by nasal cannula or face mask. The exception is hantavirus pulmonary syndrome, for which early endotracheal intubation and mechanical ventilation (Chapter 105) are often life-saving. Patients with viral hemorrhagic fever have an elevated risk of ventilator-induced lung injury (i.e., barotrauma) and pulmonary hemorrhage. In neurologically intact patients with hypoxemia, noninvasive positive-pressure ventilation may be a useful adjunct to forestall intubation. When mechanical ventilation is required, lung-protective tidal volumes of 6 to 8 mL/kg of ideal body weight should be employed. Extracorporeal membrane oxygenation has been used with apparent benefit in hantavirus pulmonary syndrome. Because of the risk of bleeding, arterial puncture for blood gas determination should be kept to a minimum.

Antibiotics and Secondary Infection

Patients should be immediately covered with appropriate antibacterial or antiparasitic therapy, with specific consideration of malaria (Chapter 345) and tick-borne rickettsial diseases (Chapter 327), until a diagnosis of viral hemorrhagic fever can be confirmed. These drugs should then be stopped unless there is evidence of coinfection. Secondary bacterial infection should be suspected when patients have persistent or new fever after about 2 weeks of illness, a time when most viral hemorrhagic fevers either have resulted in death or are resolving.

Antiviral Therapy

The only currently available specific antiviral therapy for any viral hemorrhagic fever is the guanosine analogue ribavirin, although it is not approved by the Food and Drug Administration for this indication (Table 381-5). The best data are for Lassa fever and hemorrhagic fever with renal syndrome, for which early treatment is imperative for maximum benefit. Anecdotal data suggest efficacy in other arenavirus hemorrhagic fevers, but controversial controlled trials of ribavirin in Crimean-Congo hemorrhagic fever and Rift Valley fever did not show efficacy. In vitro data generally show activity of ribavirin against dengue, yellow fever, and Omsk hemorrhagic fever viruses, but clinical studies have not been performed. The drug is not efficacious and should not be used for Ebola or Marburg hemorrhagic fevers. The main side effects of intravenous ribavirin are a mild to moderate hemolytic anemia, which infrequently necessitates transfusion and disappears with cessation of treatment, and rigors when the drug is infused too rapidly. ZMapp, which is a combination of three monoclonal antibodies, is effective for treating Ebola virus in non-human primates.[15] It has also been used in humans on a compassionate

TABLE 381-5 RIBAVIRIN THERAPY FOR VIRAL HEMORRHAGIC FEVER

INDICATION	ROUTE	DOSE[a]	INTERVAL
Treatment	IV[b]	30 mg/kg (maximum 2 g)[c]	Loading dose, followed by:
	IV[b]	15 mg/kg (maximum 1 g)[c]	every 6 hours for 4 days, followed by:
	IV[b]	7.5 mg/kg (maximum 500 mg)[c]	every 8 hours for 6 days
Prophylaxis	PO	35 mg/kg (maximum 2.5 g)[c]	Loading dose, followed by:
	PO	15 mg/kg (maximum 1 g)[c]	every 8 hours for 10 days:

[a]Pharmacokinetic and sensitivity testing for ribavirin has not been extensively performed for each viral hemorrhagic fever. The intravenous dose used is derived from that found efficacious in Lassa fever. Oral ribavirin has also been reported to be efficacious in many viral hemorrhagic fevers, especially for Crimean-Congo hemorrhagic fever, but few controlled data are available. Intravenous administration is strongly suggested whenever possible.
[b]The drug should be diluted in 150 mL of 0.9% saline and infused slowly.
[c]Reduce the dose in persons known to have significant renal insufficiency (creatinine clearance of less than 50 mL/minute).

use basis, but controlled trials to demonstrate efficacy definitively have not yet been performed. Other experimental therapies include TKM-Ebola (an RNA inhibitor), brincidofovir, and favipiravir.

A number of experimental antiviral drugs have shown in vitro activity and, in some cases, therapeutic benefit in animal studies, including various nucleoside analogues, inhibitors of *S*-adenosyl-L-homocysteine hydrolase, small interfering RNAs, phosphorodiamidate morpholino oligomers, antisense compounds, tyrosine kinase inhibitors, and other small molecules.[16] The Chinese drug chongcao shenkang has been reported to be efficacious in hemorrhagic fever with renal syndrome. None of these drugs are yet widely approved or available for treatment of viral hemorrhagic fever in humans.

Convalescent Plasma and Antibody Therapy

Although cellular immunity is thought to be the primary protection in most viral hemorrhagic fevers, transfusion of appropriately titered convalescent plasma within the first 8 days of illness has been reported to reduce the case-fatality of Argentine hemorrhagic fever from 15 to 30% to less than 1%. However, this therapy has been associated with a convalescent phase neurologic syndrome characterized by fever, cerebellar signs, and cranial nerve palsies in 10% of treated patients. Animal studies show convalescent plasma to be efficacious in Lassa fever as well, but only if it contains a high titer of neutralizing antibody and if there is a close antigenic match between the infecting viruses of the donor and recipient. Convalescent plasma also appears to be efficacious for Crimean-Congo hemorrhagic fever and Rift Valley fever, but there are no controlled data. Convalescent plasma or blood has been given to numerous patients with Ebola hemorrhagic fever, but its efficacy is still unknown. Because of the significant medical and logistical challenges to the use of convalescent immune plasma, including risk of concomitant transmission of other blood-borne pathogens, this therapy should be reserved for Argentine hemorrhagic fever and for severe and refractory cases when ribavirin is not an option. Numerous monoclonal and polyclonal antibody preparations have shown promise in animal models and may soon be ready for safety trials in humans.

Coagulation and Immune Modulators

A growing body of literature suggests that disturbances in the procoagulant-anticoagulant balance play an important role in the mediation of septic shock. A modest survival benefit (33%) was seen in monkeys treated with rNAPc2, a potent experimental recombinant inhibitor of the tissue factor/factor VIIa coagulation pathway.

Various immune modulators, including ibuprofen, corticosteroids, anti–tumor necrosis factor-α, nitric oxide inhibitors, statins, and interleukins, have not shown conclusive benefit in the treatment of sepsis. In a small study, recombinant interleukin-2 reduced the degree of acute renal insufficiency in hemorrhagic fever with renal syndrome, but further studies are needed before it can be considered the standard of care. Clinical trials of corticosteroids in hemorrhagic fever with renal syndrome have shown mixed results. Corticosteroids (e.g., 200 mg intravenous hydrocortisone per day, divided into two to four daily doses or administered by continuous infusion) are not recommended unless adrenal insufficiency is strongly suspected, the target blood pressure is not maintained despite adequate fluid repletion and vasopressors, or cerebral edema is suspected.

Management of Pregnancy

Uterine evacuation in pregnant patients appears to lower maternal mortality and should be considered, given the extremely high maternal and fetal mortality associated with viral hemorrhagic fever. However, this procedure must be performed with extreme caution because it can be considered high risk with regard to potential nosocomial transmission. Although it is technically contraindicated in pregnancy (Food and Drug Administration Category X), ribavirin should nevertheless be considered, in consultation with the patient, as a life-saving measure for the mother who has a viral hemorrhagic fever for which the drug is efficacious (see Tables 381-2 and 381-5).

Other Considerations

Oral or parenteral acetaminophen, tramadol, opiates, or other analgesics should be used as needed for pain control (see Tables 30-4 and 30-5), adjusting as necessary for hepatic insufficiency. Use of salicylates and nonsteroidal anti-inflammatory drugs should be avoided because of the risk of bleeding. Prophylactic therapy for gastrointestinal stress ulcers with proton pump inhibitors or histamine H_2-receptor antagonists is recommended (see Table 138-1). Antiemetics, such as the phenothiazines, are frequently warranted. Seizures can usually be managed with standard medications (Chapter 403).

PREVENTION

Patient Isolation, Personal Protective Equipment, and Nursing Precautions

Normal barrier nursing precautions to prevent parenteral and droplet exposure to blood and body fluids suffice in most instances. Once the diagnosis is suspected, however, precautions should be upgraded to "viral hemorrhagic fever precautions," which include patient isolation and the use of surgical masks, face shields, double gloves, gowns, head and shoe covers, and protective aprons.[17] It is prudent to place the patient in a negative airflow room if it is available, but hermetically sealed isolation chambers are not required. Insecticide-treated bed nets and room screens should be used in open-air settings to prevent transmission of arthropod-borne hemorrhagic fever viruses. Access to the patient should be limited to a small number of designated staff and family members with specific instructions and training on infection control guidelines and the use of personal protective equipment. Small-particle aerosol precautions should be used when procedures are performed that may generate aerosols, such as endotracheal intubation. Disinfection with bleach or a number of other commercially available disinfectants is advised, including chemical or heat inactivation of human waste.

Contact Tracing

The early nonspecific presentation of the viral hemorrhagic fevers poses a serious challenge to effective epidemiologic surveillance. Fortunately, the low secondary attack rates afford a measure of reassurance even when cases go unrecognized as long as proper barrier nursing is maintained. Furthermore, because mild cases, which may be still more difficult to recognize, are usually not very infectious, missed or delayed diagnosis of these patients is unlikely to pose a problem from an infection control standpoint.

Persons with unprotected direct contact with a patient during the symptomatic phase of a human-to-human communicable viral hemorrhagic fever should be monitored daily for evidence of disease for the duration of the longest possible incubation period, starting after their last contact (see Table 381-2). Given the generally low secondary attack rates, especially outside of caretakers, widespread contact tracing, laboratory testing, and postexposure prophylaxis are not indicated for casual contacts. Contacts should check their temperature daily and record the results in a log. Despite lack of evidence for transmission during the incubation period, it is usually recommended that exposed persons avoid close contact with household members that might result in exposure to body fluids, such as sex, kissing, and sharing of utensils for the duration of the incubation period. Confinement of asymptomatic persons is not warranted, but persons who develop fever or other signs and symptoms suggestive of viral hemorrhagic fever should be immediately isolated until the diagnosis can be ruled out.

Vaccines

Vaccines for viral hemorrhagic fevers are at various stages of development. The 17D live attenuated yellow fever vaccine has a generally excellent protection and safety profile, despite recent recognition of rare serious adverse events in elderly persons. Confirmed previous vaccination with 17D should essentially exclude the diagnosis of yellow fever unless the patient was immunocompromised at the time of vaccination. A highly efficacious live attenuated vaccine, Candid 1, also exists for Argentine hemorrhagic fever, although it is licensed only in Argentina. Candid 1 may also be effective in Bolivian hemorrhagic fever but does not protect against other arenavirus infections. Experimental vaccines for hemorrhagic fever with renal syndrome, Rift Valley fever, Omsk hemorrhagic fever, and Kyasanur Forest disease may be efficacious, although most have not been widely tested and are not widely approved or available. A number of vaccine candidates have recently been shown to be efficacious in animal models of Ebola, Marburg, and Lassa virus

infection.[18] Clinical trials of various Ebola vaccines are underway, with early results suggesting that they can safely elicit immunogenicity.[19]

Postexposure Prophylaxis

Postexposure prophylaxis should be considered only in persons with distinct high-risk exposure, defined as follows: (1) penetration of skin by a contaminated sharp instrument (e.g., needlestick injury); (2) exposure of mucous membranes or broken skin to blood or body secretions (e.g., blood splashing in the eyes or mouth); (3) participation in emergency procedures without appropriate personal protective equipment (e.g., resuscitation after cardiac arrest, intubation, or suctioning); and (4) prolonged (i.e., hours) and continuous contact in an enclosed space without appropriate personal protective equipment. The most infectious patients are those with severe clinical conditions, usually late in the course of illness. Prophylaxis should not be used when the only exposure was during the incubation period or after fever has subsided.

Postexposure prophylaxis with oral ribavirin has been recommended for Lassa fever, other arenavirus infections, and Crimean-Congo hemorrhagic fever, although no systematic data are available on its efficacy, and the frequent minor adverse events sometimes can be mistaken for the early signs of disease.[20] Oral ribavirin should be started immediately after the exposure, but not before counseling between the patient and the physician. The drug should be taken with food. Baseline hemoglobin, hematocrit, bilirubin, and creatinine levels should be determined, and therapy should be adjusted or reconsidered if significant anemia or renal insufficiency develops.

Convalescent plasma is given as postexposure prophylaxis for Argentine hemorrhagic fever. Numerous experimental vaccines, monoclonal antibodies, and other compounds have shown efficacy as postexposure prophylaxis in animal models, especially for filovirus infections, but are not yet approved for use in humans.

Reservoir and Vector Control

Avoiding contact with bats, primarily by avoiding entry into caves and mines in endemic areas, is a key prevention measure for Ebola and Marburg viruses. Personal protective equipment may be indicated for miners and other persons who work in these environments. Humans should also avoid exposure to fresh blood, body fluids, or meat of wild animals, especially nonhuman primates.

For rodent-borne viruses whose reservoirs often colonize human dwellings, improved "village hygiene" is recommended, such as eliminating unprotected storage of garbage and foodstuffs and plugging holes that allow rodents to enter homes. Prevention of the mosquito-borne hemorrhagic fever viruses hinges on controlling *Aedes* mosquitoes in and around the home, primarily by elimination of clean standing water containers that serve as larval habitats and by use of screened windows and doors, insecticide-treated bed nets, protective clothing, mosquito repellent, and aerosol bomb insecticides in enclosed spaces. Analogous measures can help protect against tick bites.

PROGNOSIS

The clinical course of viral hemorrhagic fever unfolds quite rapidly. In fatal cases, death usually occurs within 7 to 10 days after the onset of hemorrhagic fever symptoms in filovirus infection and in about 2 weeks with the arenaviruses and some of the other viruses. Mortality usually does not result directly from exsanguination, and external bleeding is seen in a minority of cases of viral hemorrhagic fever. Most deaths result from an intense inflammatory process akin to septic shock (Chapter 108) when insufficient effective circulating intravascular volume leads to hypotension, cellular dysfunction, and multiorgan system failure.

Common indicators of a poor prognosis include shock, bleeding, neurologic manifestations, high levels of viremia (or surrogate measurements of antigen or genome copies), elevated levels of aspartate aminotransferase (>150 IU/L), and pregnancy, especially during the third trimester, in which maternal and fetal mortality may be above 90%.

However, mild and even asymptomatic cases have been reported for what are considered the most virulent viral hemorrhagic fevers. Reasons for this heterogeneity are largely unknown, although differences in route and dose of infection, underlying comorbid illness, and host genetic predisposition have been postulated.

Survivors usually suffer no obvious long-term sequelae. Notable exceptions include deafness in up to 30% of patients after Lassa fever and sometimes after Venezuelan hemorrhagic fever and optic retinopathy with vision loss in Rift Valley fever. Nevertheless, convalescence may be prolonged, especially for Ebola and Marburg hemorrhagic fevers, with persistent myalgia, arthralgia, anorexia, weight loss, alopecia, pancreatitis, uveitis, and orchitis up to a year after infection. The psychological effects of viral hemorrhagic fever may include significant irritability, depression, post-traumatic stress disorder, and social stigmatization. For Ebola virus, mortality rates have been about 70% in West Africa.[21] In contrast, a very high proportion of patients who have been treated in the U.S. and Europe have survived with modern supportive care and experimental therapies.[22] Because these patients received multiple therapies in combination and in an uncontrolled fashion, the efficacy of any one therapy cannot be assessed.

Clinical management during convalescence includes the use of warm packs, acetaminophen, nonsteroidal anti-inflammatory drugs, cosmetics, hair growth stimulants, anxiolytics, antidepressants, nutritional supplements, and nutritional and psychological counseling as indicated. Uveitis in patients recovering from Ebola hemorrhagic fever, recently renamed Ebola virus disease, responds to topical steroids and atropine.

Because the patient's clinical status and infectivity generally correlate with the level of viremia, patients who have recovered from their acute illness can safely be assumed to have cleared their viremia and discharged from the hospital without concern of subsequent transmission at home. RT-PCR testing of blood and other body fluids has sometimes revealed residual nucleic acids, but their significance is unclear without cell culture confirmation of the presence of infectious virus. Clearance of virus may be delayed for weeks to months from a few immunologically protected sites, such as the central nervous system, chambers of the eye, and gonads, with the last resulting in rare sexual transmission months after recovery from acute disease. Consequently, abstinence or condom use is recommended for 3 months after acute illness. Although transmission through toilet facilities or ocular secretions has not been noted, simple precautions to avoid contact with these potentially infected body fluids are prudent, including separate toilet facilities and regular handwashing. Breast-feeding should be avoided during convalescence unless there is no other way to support the baby.

Grade A References

A1. Huggins JW, Hsiang CM, Cosgriff TM, et al. Prospective, double-blind, concurrent, placebo-controlled clinical trial of intravenous ribavirin therapy of hemorrhagic fever with renal syndrome. *J Infect Dis*. 1991;164:1119-1127.

A2. Koksal I, Yilmaz G, Aksoy F, et al. The efficacy of ribavirin in the treatment of Crimean-Congo hemorrhagic fever in Eastern Black Sea region in Turkey. *J Clin Virol*. 2010;47:65-68.

A3. Gotuzzo E, Yactayo S, Cordova E. Efficacy and duration of immunity after yellow fever vaccination: systematic review on the need for a booster every 10 years. *Am J Trop Med Hyg*. 2013;89:434-444.

A4. Enria DA, Ambrosio AM, Briggiler AM, et al. Candid#1 vaccine against Argentine hemorrhagic fever produced in Argentina. Immunogenicity and safety. *Medicina (B Aires)*. 2010;70:215-222.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Kortepeter MG, Bausch DG, Bray M. Basic clinical and laboratory features of filoviral hemorrhagic fever. *J Infect Dis*. 2011;204(suppl 3):S810-S816.
2. Centers for Disease Control and Prevention. 2014 Ebola Outbreak in West Africa. http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/. Accessed January 12, 2015.
3. McLay L, Liang Y, Ly H. Comparative analysis of disease pathogenesis and molecular mechanisms of New World and Old World arenavirus infections. *J Gen Virol*. 2014;95:1-15.
4. Sow A, Faye O, Ba Y, et al. Rift Valley fever outbreak, southern Mauritania, 2012. *Emerg Infect Dis*. 2014;20:296-299.
5. Markoff L. Yellow fever outbreak in Sudan. *N Engl J Med*. 2013;368:689-691.
6. Ruzek D, Yakimenko VV, Karan LS, et al. Omsk haemorrhagic fever. *Lancet*. 2010;376:2104-2113.
7. Yun NE, Walker DH. Pathogenesis of Lassa fever. *Viruses*. 2012;4:2031-2048.
8. Boisen ML, Schieffelin JS, Goba A, et al. Multiple circulating infections can mimic the early stages of viral hemorrhagic fevers and possible human exposure to filoviruses in Sierra Leone prior to the 2014 outbreak. *Viral Immunol*. 2015;28:19-31.
9. Schieffelin JS, Shaffer JG, Goba A, et al. Clinical illness and outcomes in patients with Ebola in Sierra Leone. *N Engl J Med*. 2014;371:2092-2100.
10. Sewlall NH, Richards G, Duse A, et al. Clinical features and patient management of Lujo hemorrhagic fever. *PLoS Negl Trop Dis*. 2014;8:e3233.
11. Bauer MP, Timen A, Vossen AC, et al. Marburg hemorrhagic fever in returning travellers: an overview aimed at clinicians. *Clin Microbiol Infect*. 2014; [Epub ahead of print].
12. Trombley AR, Wachter L, Garrison J, et al. Comprehensive panel of real-time TaqMan polymerase chain reaction assays for detection and absolute quantification of filoviruses, arenaviruses, and New World hantaviruses. *Am J Trop Med Hyg*. 2010;82:954-960.
13. Lyon GM, Mehta AK, Varkey JB, et al. Clinical care of two patients with Ebola virus disease in the United States. *N Engl J Med*. 2014;371:2402-2409.
14. Bah EI, Lamah MC, Fletcher T, et al. Clinical presentation of patients with Ebola virus disease in Conakry, Guinea. *N Engl J Med*. 2015;372:40-47.
15. Qiu X, Wong G, Audet J, et al. Reversion of advanced Ebola virus disease in nonhuman primates with ZMapp. *Nature*. 2014;514:47-53.
16. Ippolito G, Feldmann H, Lanini S, et al. Viral hemorrhagic fevers: advancing the level of treatment. *BMC Med*. 2012;10:31.
17. World Health Organization. Interim infection prevention and control guidance for care of patients with suspected or confirmed filovirus haemorrhagic fever in health-care settings, with focus on Ebola. September 2014. http://www.who.int/csr/resources/publications/ebola/filovirus_infection_control/en/. Accessed January 12, 2015.
18. Olschlager S, Flatz L. Vaccination strategies against highly pathogenic arenaviruses: the next steps toward clinical trials. *PLoS Pathog*. 2013;9:e1003212.
19. Kibuuka H, Berkowitz NM, Millard M, et al. Safety and immunogenicity of Ebola virus and Marburg virus glycoprotein DNA vaccines assessed separately and concomitantly in healthy Ugandan adults: a phase 1b, randomised, double-blind, placebo-controlled clinical trial. *Lancet*. 2014;[Epub ahead of print].
20. Bausch DG, Hadi CM, Khan SH, et al. Review of the literature and proposed guidelines for the use of oral ribavirin as postexposure prophylaxis for Lassa fever. *Clin Infect Dis*. 2010;51:1435-1441.
21. WHO Ebola Response Team. Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *N Engl J Med*. 2014;371:1481-1495.
22. Fowler RA, Fletcher T, Fischer WA 2nd, et al. Caring for critically ill patients with Ebola virus disease. Perspectives from West Africa. *Am J Respir Crit Care Med*. 2014;190:733-737.

REVIEW QUESTIONS

1. Two distinct diseases caused by hantaviruses are hantavirus pulmonary syndrome and

A. Argentine hemorrhagic fever
B. Lassa fever
C. Hantavirus hepatic syndrome
D. Hemorrhagic fever with renal syndrome
E. Adult T-cell leukemia

Answer: D The genus Hantavirus of the Bunyaviridae family is divided into Old and New World groups. The Old World hantaviruses, such as Hantaan, Seoul, and Puumala, among many others, cause hemorrhagic fevers with prominent renal involvement across Europe and Asia. New World hantaviruses, such as Sin Nombre and Andes, among many others, cause a viral hemorrhagic fever named hantavirus pulmonary syndrome, sometimes also called hantavirus cardiopulmonary syndrome to emphasize the significant cardiogenic component of this disease. A thorough travel history can help narrow the differential diagnosis, although there can be, at times, overlap between the two syndromes. One Old World hantavirus, Seoul, can be found virtually worldwide because its reservoir, the common Norway rat, has been spread globally through transport on ships.

2. The antiviral drug ribavirin is most clearly indicated for which of the following viral hemorrhagic fevers?

A. Ebola hemorrhagic fever
B. Yellow fever
C. Lassa fever
D. Omsk hemorrhagic fever
E. Rift Valley fever

Answer: C The only currently available specific antiviral therapy for any viral hemorrhagic fever is the guanosine analogue ribavirin. The best data are for Lassa fever, in which intravenous ribavirin has been shown to decrease mortality in severe disease from 55% to 5% when it is begun within the first 6 days of illness.[1] Ribavirin efficacy has also been demonstrated for hemorrhagic fever with renal syndrome due to Hantaan virus. The drug appears to be efficacious in other arenavirus hemorrhagic fevers and Crimean-Congo hemorrhagic fever, but few randomized controlled clinical trials have been performed for these syndromes. Ribavirin is not efficacious and should not be used for Ebola or Marburg hemorrhagic fevers.

1. McCormick JB, King IJ, Webb PA, et al. Lassa fever. Effective therapy with ribavirin. *N Engl J Med.* 1986;314:20-26.

3. Viral hemorrhagic fever syndromes

A. Can usually be easily distinguished from other common febrile diseases on clinical grounds early in the course of disease
B. Can be confirmed only through virus culture in a high-containment laboratory
C. Are usually characterized by copious bleeding, often leading to exsanguination and death
D. Typically have incubation periods of days to weeks
E. Should be routinely treated with intravenous corticosteroids

Answer: D After an incubation period ranging from days to weeks, most patients with viral hemorrhagic fever present with nonspecific signs and symptoms difficult to distinguish from a host of other febrile illnesses. Laboratory confirmation is therefore imperative and can be performed through a variety of methods, including the enzyme-linked immunosorbent assay, polymerase chain reaction, virus culture, and immunohistochemistry on postmortem tissues. Most of these assays do not involve live virus and thus can be performed outside of a high-containment laboratory, noting that the specimen to be tested may nevertheless be infectious. Despite the name, external bleeding is seen in a minority of cases of viral hemorrhagic fever, especially early in the course of disease. Death in fatal cases is not typically due to exsanguination but rather to an intense inflammatory process akin to septic shock leading to multiorgan system failure. Adrenal or pituitary gland necrosis with consequent vascular collapse has been postulated but not specifically demonstrated. Corticosteroids are not indicated unless adrenal insufficiency is strongly suspected, the target blood pressure is not maintained despite adequate fluid repletion and vasopressors, or cerebral edema is suspected.

4. Which of the following is the most essential component of control of Lassa virus transmission between humans?

A. Immediate isolation of all patients in the incubation period
B. Placement of insecticide-treated bed nets and room screens in patient rooms
C. Donning of "space suits" by health care personnel working in patient isolation wards
D. Placement of suspected case-patients in hermetically sealed rooms
E. Routine barrier nursing precautions (gloves, gowns, and masks)

Answer: E Human-to-human transmission of Lassa virus, as for the other directly communicable hemorrhagic fever viruses, is through direct contact with infected blood and body fluids. Routine barrier nursing precautions to prevent parenteral and droplet exposure are protective in most cases, although "viral hemorrhagic fever precautions" consisting of the use of surgical masks, face shields, double gloves, gowns, head and shoe covers, and protective aprons are implemented for added safety once the syndrome is suspected. Patients are not infectious during the incubation period. There is little evidence for aerosol transmission between humans. Hermetically sealed isolation chambers are not required, although it is prudent to place the patient in a negative airflow room if it is available. The reservoir for Lassa virus is the rodent *Mastomys natalensis*, commonly called the natal mastomys or multimammate rat. There is no arthropod vector.

5. A 23-year-old otherwise healthy woman is seen with complaints of fever, malaise, headache, and muscle aches for the past 5 days. She recently returned from a church mission trip in which she assisted physicians and nurses treating patients in a resource-poor area of Bolivia. She is in moderate distress with vital signs as follows: temperature, 102.5° F; blood pressure,105/65 mm Hg; pulse, 120; and respiratory rate, 24. Her conjunctivae are injected, and a petechial rash is noted over her torso. Initial laboratory findings show moderate leukopenia, hemoconcentration, and mild to moderate thrombocytopenia. You suspect infection with which of the following hemorrhagic fever viruses?

A. Machupo virus
B. Lujo virus
C. Ebola virus
D. Junín virus
E. Sin Nombre virus

Answer: A With the exception of dengue virus, hemorrhagic fever viruses are zoonotic and are maintained in nature in mammalian reservoirs. The endemic area of any given hemorrhagic fever virus is restricted by the distribution of its natural reservoir or arthropod vector. Knowledge of the geographic limits of the reservoir, along with a detailed travel history, is key to assessing the likelihood of exposure to a given hemorrhagic fever virus. In this case, the patient may have been exposed to Machupo virus, the causative agent of Bolivian hemorrhagic fever, which is endemic only in Bolivia. Junín virus is also found in South America but is restricted to the Argentine pampas. Ebola and Lujo viruses are endemic in Africa. Sin Nombre virus, which is one of the causative agents of hantavirus pulmonary syndrome, is found in North America. Other hantaviruses are found in Bolivia, but the patient's presentation and laboratory results are not consistent with hantavirus pulmonary syndrome, in which rash would be unusual and leukocytosis with immunoblasts is classically noted. The fact that she had potential exposure to human blood or body fluids while assisting physicians and nurses should heighten suspicion of a viral hemorrhagic fever.

382

ARBOVIRUSES CAUSING FEVER AND RASH SYNDROMES

STANLEY J. NAIDES

COLORADO TICK FEVER

DEFINITION

Colorado tick fever (mountain fever, American mountain fever) is an acute, often self-limited, typically biphasic febrile illness that is common in the Rocky Mountain areas, the Sierra Nevada and Wasatch ranges, and the Black Hills mountain areas.[1] The virus is transmitted through the bite of the hard-shelled tick *Dermacentor andersoni* (Rocky Mountain wood tick), and the disease's range corresponds to the vector's range. Other coltiviruses, such as the Salmon River, Eyach, Banna, Beijing, and Gansu viruses, have also been implicated in human disease.

The Pathogen

The causative agent, Colorado tick fever virus, is a member of the genus Coltivirus, family Reoviridae. Coltiviruses have a genome consisting of 12 double-stranded RNA segments. Colorado tick fever virus is the prototype member.

EPIDEMIOLOGY

D. andersoni is found at elevations of 4000 to 10,000 ft. Seasonal temperatures tend to influence the range, with the vector being found at higher elevations in warmer seasons and at lower elevations in colder seasons. Human exposure usually occurs during outdoor recreational activities in these areas. Occasional exposure occurs in nonendemic areas from ticks exported out of the endemic region in clothes, hiking equipment, or baggage. Infections generally take place between March and September, when the adult tick is most plentiful. Ticks are most abundant in south-facing dry and rocky slope habitats that favor small rodents (e.g., chipmunks, ground squirrels, marmosets), with underbrush cover, burrows, and humidity for the ticks. Colorado tick fever virus is found in nymphal and adult ticks that overwinter on the rodent host, in which viremia persists for weeks to months. In the endemic area, as many as 14% of *D. andersoni* ticks carry Colorado tick fever virus. Humans are an incidental host. Fewer than 50 cases have been reported annually in Colorado beginning in 1992. The actual number of cases is probably significantly larger and includes subclinical, mild, and unreported cases.

The geographic range of Colorado tick fever may be larger than the well-recognized endemic mountain areas. Serologically confirmed cases in California have been attributed to the Colorado tick fever–related virus S1-14-03, which is transmitted by *Dermacentor variabilis* (American dog tick). Salmon River virus causes a Colorado tick fever–like illness in rafters on the Salmon River in Idaho. Another similar virus, Eyach virus, has been implicated in neurologic illness in France and Germany and has been isolated from the deer ticks *Ixodes ricinus* and *Ixodes ventalloi*. Colorado tick fever has been reported rarely in mainland China.

PATHOBIOLOGY

Colorado tick fever virus replicates in $CD34^+$ stem cells in bone marrow and leads to mild to moderate leukopenia and thrombocytopenia. The virus also replicates in committed erythrocyte precursors and may be detected in circulating erythrocytes up to 4 weeks after infection.

CLINICAL MANIFESTATIONS

Patients report a tick bite or exposure in 90% of cases, but there is no notable local reaction to the tick bite. After a mean incubation of 3 to 4 days (range, 0 to 14 days), sudden-onset fever develops in association with malaise, chills, myalgia, weakness, headache, photophobia, retro-orbital pain, and cutaneous hyperesthesia. Conjunctival and oropharyngeal injection, palatal enanthem, lymphadenopathy, and splenomegaly may be present. The absence of prominent respiratory and gastrointestinal symptoms helps exclude other febrile illnesses. A petechial or maculopapular exanthem, found in 15% of patients, may be confused with the rash of Rocky Mountain spotted fever (Chapter 327). The illness has a "saddleback" fever pattern consisting of resolution of the initial fever within 1 week and recrudescence after a 2- to 3-day hiatus. A third fever episode may occur.

Leukopenia develops 5 to 6 days after onset of the illness. Mild thrombocytopenia and anemia may occur.

Myocarditis, pneumonitis, hepatitis, orchitis, and epididymitis may complicate adult infection, and aseptic meningitis or encephalitis may occur in up to 10% of childhood infections.

DIAGNOSIS

Clinical diagnosis is confirmed by demonstration of the Colorado tick fever viral genome or specific acute phase IgM antibody. The viral genome may be detected up to 6 weeks after infection by nucleic acid–based methods such as reverse transcription–polymerase chain reaction (RT-PCR) on blood or stored blood clots. Virions in circulating erythrocytes may be detected by immunofluorescent antibody labeling. Anti–Colorado tick fever virus IgM antibody is detected by antibody capture enzyme-linked immunosorbent assay (ELISA) or complement fixation. Neutralization assays using Vero or BHK-21 cells have been helpful.

Differentiating Colorado tick fever from Rocky Mountain spotted fever (Chapter 327) may be difficult before the appearance of the typical rash of the latter. However, Rocky Mountain spotted fever does not have a saddleback fever pattern and is 20 times less common than Colorado tick fever in the western endemic area.

TREATMENT

Rx

Treatment is supportive. Aspirin is contraindicated to avoid complicating thrombocytopenia.

PROGNOSIS

Extreme weakness and malaise may persist for weeks to months after final resolution of the fever. Older patients have a prolonged recovery. Seventy percent of patients older than 30 years may still have fatigue 3 weeks after the fever, whereas children and adolescents may recover completely within a week. Rare instances of maternal-fetal transmission have been reported. Full recovery eventually occurs, except when the disease course is complicated by neurologic insult. Patients should refrain from donating blood for 6 months.

DENGUE

DEFINITION

Dengue is an acute febrile illness characterized by severe muscle and joint pain, rash, malaise, and lymphadenopathy. The severity of the musculoskeletal complaints gave rise to the sobriquet *breakbone fever*. Dengue occurs in the tropical and subtropical climes of the Caribbean, Central and South America, Asia, and Africa. The mosquito range extends into the southeastern part of the United States, where dengue reemerged in the 1980s. After World War II, a spreading global pandemic has been associated with erosion of mosquito control programs, human population spread into rural settings, increased air travel, deterioration in public health infrastructure, and global warming. Each year, more than 200 million people worldwide are infected with dengue.[2]

The Pathogen

Dengue virus is a member of the Flaviviridae family, which consists of single-stranded RNA viruses with a lipid envelope approximately 50 nm in diameter. There are four serotypes of dengue: DEN-1, DEN-2, DEN-3, and DEN-4. No cross-protection is seen among the serotypes, so dengue can develop after infection with another serotype. Infection with a second serotype places the individual at risk for the development of hemorrhagic fever (Chapter 381).

EPIDEMIOLOGY

Dengue is transmitted to humans by the bite of female *Aedes aegypti* and *Aedes albopictus* mosquitoes. *A. albopictus* has become the dominant pest mosquito in many urban centers. Members of the two mosquito species acquire dengue virus by biting humans, typically during the day. The mosquitoes nest in stagnant water around human dwellings; they are not typically encountered in the forest. In the human host, dengue virus may reach a titer of greater than 10^8 median infectious doses per milliliter. The mosquito becomes infected when taking its meal from a viremic host. The virus continues replication in the midgut epithelium and salivary glands of female mosquitoes, which remain infectious for life. Within 8 to 12 days of the initial infection, the mosquito's salivary glands become infected, and virus is shed with saliva during the next blood meal. A given mosquito may infect multiple individuals, especially in view of its skittishness during feeding—slight movement of the host interrupts its meal, after which it returns to the original or another host. Zoonotic life cycles involving nonhuman primates (i.e., chimpanzees, gibbons, and macaques) and canopy-dwelling forest *Aedes* species have been demonstrated in western Africa and Malaysia.

The incubation period is typically 4 to 7 days but may range from 3 to 14 days. During outbreaks in the southeastern United States and Puerto Rico, the risk for infection may be as high as 79% in naïve hosts, and clinical disease may develop in up to 20%. Immunity against the infecting serotype is probably lifelong, but individuals remain susceptible to the remaining serotypes. Peak transmission occurs after increased rainfall, when rainwater collected in household containers allows expansion of mosquito populations. Epidemics tend to occur in 3- to 5-year cycles, but interepidemic cases occur regularly.

Dengue is a particular risk to visitors to the tropics and is a leading cause of pediatric morbidity and mortality in endemic areas.[3] Globalization and climate change have contributed to expansion of the geographic range. In one study of people who now live in the United States but who were born, lived

in, or traveled to dengue-endemic countries, 19% had IgG antibodies to dengue but 85% of them had no clinical history of dengue. Dengue accounts for approximately 2% of the febrile illnesses in travelers returning to the United States.

PATHOBIOLOGY

Dengue hemorrhagic fever (Chapter 381) and dengue shock syndrome are forms of dengue reinfection characterized by capillary leakage and hemorrhage. Previous infection with an alternative serotype allows antibody to the previously encountered serotype to combine with the newly infecting serotype. Although the first exposed serotype antibody is not neutralizing, it does allow enhanced antibody-mediated macrophage uptake, thereby leading to macrophage activation and increasing viral replication and viral load. Excretion of vasoactive inflammatory mediators by macrophages results in vascular leakage; severe vascular leak causes shock. Endothelial cell swelling and perivascular edema may occur. Rarely, dengue shock syndrome may occur with primary infection. Variation in a strain's ability to generate enhancing antibody, as well as differences in virulence, may account for differences in clinical behavior.

CLINICAL MANIFESTATIONS

Dengue infection is often subclinical. When it is symptomatic, dengue may be manifested as classic dengue, dengue hemorrhagic fever, or dengue shock syndrome. Patients may also have mild illness characterized by nonspecific fever, anorexia, and headache.

Classic dengue, which typically occurs in nonindigenous older children and adults, is characterized by sudden-onset fever, severe frontal headache, retro-orbital pain, myalgia, and, in many cases, nausea, vomiting, rash, lymphadenopathy, and arthralgia.[4] Patients may experience generalized weakness, altered taste, rigors, and cutaneous hyperesthesia. Classic dengue is self-limited, but some patients progress to dengue hemorrhagic fever or dengue shock syndrome, which is characterized by capillary leakage, hypotension, narrowed pulse pressure, and shock. Dengue in pregnancy may be severe.[5]

Physical examination demonstrates fever, relative bradycardia, scleral injection, ocular pressure tenderness, and pharyngeal injection. A transient macular rash appears on days 1 or 2 of illness. On days 2 and 3 of illness, fever and other symptoms may improve. The fever is typically but not consistently biphasic. After a hiatus of typically 2 days, fever and other symptoms recrudesce, although less severely. Generalized, nontender lymphadenopathy of the posterior cervical, epitrochlear, and inguinal regions may develop. Rash also recurs and appears as 2- to 5-mm speckles of pallor surrounded by erythema and occasionally accompanied by burning dysesthesia of the palms and soles. The rash may desquamate.

DIAGNOSIS

An adequate travel history and knowledge of occurrence of disease in the community can lead to consideration of dengue in the differential diagnosis. Viremia is of adequate intensity in infections with DEN-1, DEN-2, and DEN-3 to allow viral isolation. Viremia in DEN-4 infections is often less intense and more difficult to detect through inoculation of mosquito cells in vitro. Specific IgM antibody appears 3 to 5 days after infection. IgG antibody appears 9 to 10 days after infection. Cross-reactivity with other flaviviruses prevents serotype-specific diagnosis. Neutralization testing with hemagglutination inhibition is more specific, and complement fixation testing for IgG in paired sera is helpful. PCR-based assays are available.

Leukopenia develops by the second day of fever, falling to a low of 1000 to 2000 cells/mL by day 5 or 6, and is associated with granulocytopenia. In dengue hemorrhagic fever, thrombocytopenia of less than 100,000 cells/mL and a prolonged prothrombin time are characteristic.[6] Mild to moderate proteinuria and a few casts may be detected. Aspartate transaminase levels may be increased.

TREATMENT, PREVENTION, AND PROGNOSIS

Rx

Classic dengue resolves abruptly in 5 to 7 days, but fatigue and depression may linger for weeks; survival is uniform. The prognosis of patients with dengue hemorrhagic fever (Chapter 381) and dengue shock syndrome depends on early diagnosis and the introduction of supportive measures. Treatment is supportive and consists of antipyretics and analgesics. Initial resuscitation of patients with shock syndrome (Chapter 106) with crystalloid and colloidal solutions is indicated in those with moderately severe dengue shock syndrome. Fresh-frozen plasma and blood products are used as necessary.

A new tetravalent dengue vaccine can prevent about two-thirds of cases of dengue and about 95% of severe cases. Most infected patients recover fully, but the overall mortality rate is about 1% because of the poorer outcome of dengue shock and hemorrhagic fever (Chapter 381).

WEST NILE FEVER VIRUS

DEFINITION

West Nile fever is an acute febrile illness associated with malaise, rash, headache, myalgia, and lymphadenopathy. Infection involves a bird-mosquito-human cycle.[7] Viremia develops in all varieties of birds. Bats, cats, chipmunks, domestic rabbits, horses, skunks, squirrels, dogs, sheep, llamas, and alpacas may be infected.

The Pathogen

West Nile fever virus, which is the most widely distributed flavivirus, is transmitted by a variety of mosquito species. The mosquito vector varies: *Culex univittatus*, *Culex pipiens*, and *Culex molestus* in the Middle East and Africa; *Mansonia metallicus* in Uganda; and *Culex tritaeniorhynchus* in Asia. After introduction into the New York City area in 1999, *C. pipiens* became the most important vector in the United States. Other mosquito species may carry the virus.

EPIDEMIOLOGY

Viral transmission involves mosquitoes and wild birds, with mammals, including humans, being incidental end-stage hosts. In endemic areas, more than 60% of young adults have antibodies, thus suggesting a high prevalence of inapparent or undifferentiated febrile illness in children. There is no gender predominance. Between 0.5 and 1% of infected individuals experience a more severe illness. Incubation is typically 3 to 15 days but may be as short as 1 day. West Nile virus emerged in the United States in New York and has spread throughout the continental United States, Canada, Mexico, the Caribbean, and Central and South America. In America, birds of the Corvidae family (e.g., crows, jackdaws, ravens) are often infected, and recognition of increased death in crow populations continues to serve as a sentinel for the presence of West Nile virus. In addition to mosquito transmission, the virus has been transmitted by a transplanted organ, through blood transfusion, transplacentally, and in the laboratory. From 1999 to 2004, almost 17,000 neuroinvasive cases were reported to the Centers for Disease Control and Prevention.

Zika virus, a flavivirus related to dengue and West Nile virus, has emerged in Africa, Asia, and the Yap Islands of the southwestern Pacific. It causes outbreaks of fever, rash, arthralgia, and conjunctivitis.

PATHOBIOLOGY

West Nile virus grows in a variety of cells in vitro and produces cytopathic effects in *A. albopictus* cells. Individuals in whom encephalitis develops show evidence of diffuse brain inflammation and neuronal degeneration, with virus detected early in multiple sites. The virus initially replicates in keratinocytes and skin-resident dendritic cells, which then migrate to local lymph nodes, where replication generates viremia and organ dissemination. Either immune cell trafficking or disruption of the blood-brain barrier allows neuroinvasion.

CLINICAL MANIFESTATIONS

Approximately 80% of infections are asymptomatic, and most of the remaining infections are mild and manifested by fever, malaise, headache, nausea, anorexia, generalized lymphadenopathy, and myalgia.[8] Aseptic meningitis or encephalitis (Chapter 383) may occur in the elderly and, less commonly, in the very young. Severe neurologic disease, including meningitis, myelitis, encephalitis, and flaccid paralysis of the limbs and respiratory muscles, can develop. Like Colorado tick fever and dengue, West Nile fever may be biphasic. Nonpruritic, maculopapular, or roseolar rash occurs on the chest, back, and arms in half the patients, beginning during or with resolution of the fever. The rash persists for up to 1 week and then resolves with desquamation. Patients may experience vomiting, diarrhea, abdominal pain, and pharyngitis. Anterior myelitis or hepatitis may also occur. Disease is usually milder in children than in adults.

DIAGNOSIS

West Nile virus may be isolated from up to 77% of patients with West Nile fever on the first day of illness, but viral isolation is less common in patients with encephalitis (Chapter 383). Low-titer viremia may persist for the first 5 days of illness. Tests in acute and convalescent phase serum for virus-specific antibody using ELISA or immunofluorescence are diagnostic. However, virus-specific IgM may persist in the serum at 1 year after infection in a minority of patients. Detection of West Nile virus–specific IgM in cerebrospinal fluid is diagnostic of neuroinvasion because IgM normally does not cross the blood-brain barrier. Neutralization assays help distinguish cross-reactive antibodies to other flaviviruses. RT-PCR may detect viral RNA in human samples and in avian and insect specimens.

TREATMENT Rx

Treatment is supportive. The clinical value of antiviral agents is unknown. Infection is controlled by the endogenous development of neutralizing antibodies to protein E, the viral envelope protein.

PROGNOSIS

Illness generally persists 3 to 6 days before rapid recovery. The prognosis is excellent, but mortality rates of 10% or higher occur in patients with encephalitis (Chapter 383).

PHLEBOTOMUS FEVER

DEFINITION

Phlebotomus fever (i.e., sandfly fever, pappataci, and 3-day fever) is an acute, mild, self-limited febrile illness transmitted through the bite of *Phlebotomus* flies.

The Pathogen

Phlebotomus fever viruses are members of the genus Phlebovirus, family Bunyaviridae. The latter consists of a group of single-stranded RNA viruses that are 80 to 120 nm in diameter, possess a lipid envelope, and have three segments in the genome. A related virus, the Toscana virus hosted by *Phlebotomus perniciosus* and *Phlebotomus perfiliewi*, causes a similar illness in countries surrounding the northern Mediterranean basin and is emerging in western Europe. The related Punique, Granada, and sandfly fever Turkey viruses may cause similar febrile illnesses, acute meningitis, or meningoencephalitis.

EPIDEMIOLOGY

The virus's distribution parallels the distribution of *Phlebotomus* flies found throughout the Mediterranean basin, Middle East, and western India and Pakistan. In Central America, *Lutzomyia* fly species may transmit the virus. These tiny sandflies pass through mosquito netting to feed in the early evenings. Virus is maintained by transovarial and transstadial transmission. During outbreaks, humans may serve as a reservoir. Human infection is more common in rural areas during the summer months.[9] The incubation period is 2 to 6 days. Sandflies spread by hopping, thus limiting their travel range. Use of insect sprays locally is effective in decreasing risk.

CLINICAL MANIFESTATIONS

Sandfly fever virus causes an acute febrile illness associated with malaise, headache, photophobia, ocular pain, altered taste, myalgia, and arthralgia. The myalgias may be localized to specific regions (e.g., the chest) and simulate regional syndromes such as pleurodynia. A macular or urticarial rash may appear. Examination may show relative bradycardia after the first day, conjunctival injection, mild papilledema, or small palatal vesicles. Fever lasts 2 to 4 days and then subsides. Weakness and malaise may persist during convalescence. About 15% of patients experience recrudescence in 2 to 12 weeks. Aseptic meningitis may occur with mild cerebrospinal fluid pleocytosis. Peripheral leukopenia and lymphopenia may be present early in the illness. However, leukopenia may be delayed in some patients until the third day of illness, and a rebound relative lymphocytosis may be encountered.[10]

DIAGNOSIS

The diagnosis of phlebotomus fever is confirmed by isolation of virus after intracerebral inoculation of suckling mice, detection of the viral genome by RT-PCR, or detection of specific IgM antibody by ELISA.

TREATMENT AND PROGNOSIS Rx

Treatment is supportive, and recovery is complete. Ribavirin (Chapter 360) has been proposed as a therapeutic option.

RIFT VALLEY FEVER

Rift Valley fever, which is an acute-onset, febrile illness, is often associated with epizootic waves of spontaneous abortion in livestock.[11]

DEFINITION

The Rift Valley fever virus is a member of the family Bunyaviridae, genus Phlebovirus, but unlike other members of the genus, it is transmitted by *Aedes* mosquitoes.

EPIDEMIOLOGY

Rift Valley fever occurs throughout most of Africa, with the majority of epizootic outbreaks occurring in eastern and southern Africa, although outbreaks occurred in Saudi Arabia and Yemen in 2000 and Mauritania in 1998, 2003, 2010, and 2012. The principal initial vectors are probably the *Aedes* species associated with flooding, although *Stomoxys* flies have been shown to transmit virus as well. Shallow pools along rivers and streams play an important role as mosquito breeding sites. Feeding on nearby livestock allows a local epizootic outbreak and amplification of the virus in local mosquito populations, including *C. pipiens* in Egypt and *Culex theileri* in eastern Africa. Exposure to aborted livestock increases human risk of infection.

Hemorrhagic fever in humans is typically seen 1 to 2 weeks after a wave of abortion in livestock. Initial human cases usually occur in those who have close contact with livestock. The virus is highly transmissible through aerosolization. Although the risk for severe human infection is less than 1%, the extensive exposure associated with outbreaks can lead to significant morbidity and mortality. For example, in the 1977-1978 Egyptian outbreak associated with movement of camels from Sudan, an estimated 200,000 people were infected, with 600 deaths. Zinga virus, isolated in central Africa and Madagascar and shown to be responsible for mild human illness, is a strain of Rift Valley fever virus.

PATHOBIOLOGY

Rift Valley fever virus grows well in a variety of cell cultures and has cytopathic effects. After infection by a mosquito bite, virus is transported through the lymphatics to regional lymph nodes, where replication allows amplification of the input inoculum and development of viremia with systemic spread. Viral replication in liver, spleen, lymph node, adrenal, lung, and kidney tissues is highly cytopathic. In severe cases, hepatic necrosis and, rarely, focal brain necrosis may occur. Encephalitis is not associated with viremia, thus suggesting that this sequela is immune mediated rather than a direct viral effect. Inflammatory cell infiltration is associated with focal necrosis in the brain. Spontaneous abortion is common in livestock, but fetal loss in humans is not clearly correlated with viral infection.

CLINICAL MANIFESTATIONS

Most human infections are mild, with an abrupt onset of fever, chills, malaise, and arthralgia following a 2- to 6-day incubation. Despite the development of neutralizing antibodies, however, about 1 to 2% of infections progress to more severe disease, including a severe hemorrhagic fever associated with hepatic necrosis and disseminated intravascular coagulopathy. Recovery is complicated by retinal vasculitis or encephalitis, which occurs in less than 0.5% of patients 1 to 4 weeks after recovery and is associated with recurrent fever. In severe cases, focal brain necrosis and encephalitis may lead to hallucinations, stupor, coma, and death.

DIAGNOSIS

Intense viremia allows detection of virus by quantitative real-time RT-PCR. Specific IgM and IgG are detectable by ELISA applied to acute and convalescent (after 1 to 2 weeks) paired sera.

TREATMENT, PREVENTION, AND PROGNOSIS

Rx

Treatment is supportive. Ribavirin (see Table 381-5 and Chapter 360) has been proposed as a therapeutic option but exhibits limited penetration of the blood-brain barrier. Favipiravir, a selective inhibitor of RNA-dependent RNA polymerase, is also a candidate therapeutic under study. In endemic areas, vaccination of livestock is the most effective preventive measure.

The prognosis is good in the absence of retinitis or encephalitis. A high viral load at initial evaluation is a prognostic indicator for a poor outcome.

CHIKUNGUNYA FEVER

DEFINITION

Chikungunya fever is a febrile arthritis that occurs in sporadic cases and in epidemics.[12]

The Pathogen

Chikungunya, an enveloped, single-stranded RNA virus 60 to 70 nm in diameter, is a member of the family Togaviridae, genus Alphavirus. Chikungunya virus is transmitted by mosquitoes, principally *Aedes* species, but also by *Mansonia africana* and other genera. Known animal reservoirs are monkeys, baboons, and, in Senegal, *Scotophilus* bat species. During outbreaks, humans are the major reservoir.

EPIDEMIOLOGY

Chikungunya, which is endemic in sub-Sahara Africa, India, the Philippines, and Southeast Asia, spread in 2004-2005 to the Seychelles, Mauritius, and Mayotte islands with a genotype better adapted to *A. albopictus*. This genotype then spread to India, where the outbreak continues with millions being affected. The global emergence of this disease is exemplified by outbreaks in Réunion Island, Bhutan, Papua New Guinea, and Italy as well as by recently reported cases in several Caribbean islands.[13] Non-travel related cases have been reported in Florida, United States.[14] Outbreaks typically develop after heavy rains. In urban settings, outbreaks are explosive. In endemic areas, seroprevalence rates may be as high as 90%, thus suggesting that time required for loss of herd immunity is the reason for the prolonged absence of cases in a region after an outbreak. Globalization may contribute to increasing propensity for spread. After inoculation, the incubation period is typically 2 to 3 days but ranges from 1 to 12 days.

PATHOBIOLOGY

Intense viremia develops within 48 hours of the mosquito bite and wanes 2 to 3 days later. Onset of hemagglutination inhibition and neutralizing antibodies clears the viremia. Superficial capillaries in rash-involved skin demonstrate erythrocyte extravasation and perivascular cuffing. The virus adsorbs to human platelets and causes them to aggregate. Synovitis probably results from direct chikungunya viral infection of synovium.[15]

CLINICAL MANIFESTATIONS

Chikungunya fever is characterized by an explosive onset of fever and severe arthralgia. Constitutional symptoms, fever (temperature to 40° C), rigors, headache, photophobia, retro-orbital pain, conjunctival injection, pharyngitis, anorexia, nausea, vomiting, abdominal pain, tense lymphadenopathy, and myalgia are common. A maculopapular rash located on the torso, extremities, and occasionally the face, palms, and soles occurs in most patients 1 to 10 days after onset of the illness. Appearance of the rash is often associated temporally with initial defervescence; the rash may recur with fever and may be pruritic. Isolated petechiae and mucosal bleeding may occur, but significant hemorrhage is rare. Desquamation may take place when the rash resolves. The initial acute illness may last 2 to 3 days (range, 1 to 7 days). Fever may recrudesce after a 1- to 2-day hiatus. The polyarthralgia is migratory and predominantly affects the small joints of the hands, wrists, feet, and ankles, with less prominent involvement of the large joints.[16] Previously injured joints may be more severely affected. Stiffness and swelling may occur, but large effusions are uncommon. Synovial fluid shows decreased viscosity with poor mucin clot and 2000 to 5000 white blood cells per milliliter. Symptoms, including arthralgia, arthritis, and tenosynovitis, may persist for months to years. Maternal-fetal transmission may result in severe neonatal infection.

DIAGNOSIS

Chikungunya fever must be differentiated from dengue and o'nyong-nyong fever. Chikungunya virus may be isolated from blood during the initial 2 to 4 days of illness. In some patients, viral antigen may be detected in acute sera by hemagglutination assay as a result of the intensity of the viremia. Commercially available real-time RT-PCR assays on acute phase serum may be used to confirm the diagnosis. Specific IgM antibody may be detected for 6 months or longer. Hemagglutination inhibition and neutralization antibodies develop as the viremia is cleared. Complement fixation antibodies are positive by the third week and slowly decrease during the subsequent year.

TREATMENT AND PROGNOSIS

Rx

Treatment is supportive. Nonsteroidal anti-inflammatory agents are useful. During the acute arthritis, range of motion exercises lessen the stiffness. In most cases, mild joint symptoms may persist for months. Destructive arthropathy is rare and may be associated with low-titer rheumatoid factor, thus suggesting an unrelated, underlying inflammatory arthritis. Following the outbreak on Réunion Island in 2006, 70% of affected patients had episodic arthralgia, typically symmetric and incapacitating, with joint swelling in 63% at 3 years after infection.[17] In children, arthralgia and arthritis are milder and briefer in duration. A safe, immunogenic vaccine is currently under investigation.[18]

O'NYONG-NYONG FEVER

DEFINITION

O'nyong-nyong means "joint breaker" in the Acholi dialect of northwestern Uganda, where o'nyong-nyong fever first appeared in February 1959.

The Pathogen

O'nyong-nyong fever is clinically similar to chikungunya fever, and the viruses share antigenic similarity. O'nyong-nyong virus is also a member of the family Togaviridae, genus Alphavirus.

EPIDEMIOLOGY

Within 2 years of its appearance in 1959, the o'nyong-nyong fever virus spread through Uganda and eastern Africa and affected 2 million people. Serologically determined attack rates ranged from 50 to 60%, with case rates of 9 to 78%. Disease spread at a rate of 2 to 3 km daily. After the epidemic, the virus was not detected again until it was isolated from *Anopheles funestus* mosquitoes in Kenya in 1978. *Anopheles gambiae* also serves as a vector. Serologic surveys suggested that o'nyong-nyong virus is endogenous, but cases were not detected again until 1996-1997, during an outbreak in south central Uganda. An outbreak in western Côte d'Ivoire occurred in 2003. The nonhuman vertebrate reservoir for o'nyong-nyong virus is not known. The incubation period lasts at least 8 days. O'nyong-nyong fever virus vectors include *A. funestus*, *A. gambiae*, and other species.

Igbo-ora (meaning "the disease that breaks your wings") virus is a variant of o'nyong-nyong, with 98.5% homology between the two at the genomic level. Igbo-ora is serologically similar to the chikungunya and o'nyong-nyong viruses. In 1984, an epidemic of fever, rash, arthralgia, and myalgia occurred in four villages on the Ivory Coast. The virus was isolated from *A. funestus* and *A. gambiae* mosquitoes and from affected individuals.

PATHOBIOLOGY

Little is known about the pathobiology of o'nyong-nyong fever.

CLINICAL MANIFESTATIONS

Illness begins with a sudden onset of polyarthralgia and polyarthritis. Between 4 and 7 days later, rash begins with improvement in joint symptoms. The rash is uniform in nature, begins on the face, and then spreads to the torso and extremities and occasionally to the palms. The rash lasts 4 to 7 days before fading. Fever is not prominent, but postcervical lymphadenopathy may be marked. Arthralgia is incapacitating in most patients for up to a week, but residual joint pain may persist for months.

DIAGNOSIS

O'nyong-nyong fever is difficult to differentiate from chikungunya fever and may also be mistaken for measles. Specific hemagglutination inhibition and

complement fixation tests are available. Mouse antisera raised against chikungunya virus react equally well with o'nyong-nyong virus, but o'nyong-nyong antisera do not react well with chikungunya virus. O'nyong-nyong–specific RT-PCR is available in reference laboratories.

TREATMENT AND PROGNOSIS

Rx

Treatment is symptomatic. Although residual joint pain often persists, there do not appear to be any long-term sequelae.

MAYARO FEVER

DEFINITION

Mayaro fever is an acute febrile illness characterized by fever, rash, arthralgia, and arthritis. Mayaro virus was first recognized in Trinidad in 1954. It has caused recorded outbreaks in Bolivia and Brazil and is endemic in the rain forest region where Bolivia, Brazil, and Peru share borders. Mayaro virus has a monkey reservoir and is transmitted to humans by *Haemagogus* mosquitoes dwelling in the tropical rain forest canopy.

The Pathogen

Mayaro virus is a member of the family Togaviridae, genus Alphavirus.

EPIDEMIOLOGY

Mayaro virus was responsible for an outbreak in Belterra, Brazil, in 1988. Eight hundred of 4000 exposed latex gatherers became infected, with a clinical attack rate of 80%. Forest workers and hunters continue to be at greatest risk. Cases of imported Mayaro virus infection have been documented in the United States and in Europe after travel to the endemic Brazil-Bolivia-Peru interborder region.[19] The virus has been isolated from a bird in Louisiana, thus raising the specter of emergence in North America.

PATHOBIOLOGY

Viremia occurs during the first 1 to 2 days of illness.

CLINICAL MANIFESTATIONS

Illness is characterized by a sudden onset of fever, headache, dizziness, chills, and arthralgia in the small joints of the hands and feet. About 20% of patients have joint swelling. Unilateral inguinal lymphadenopathy is seen occasionally. Leukopenia is common. Fever resolves after 3 to 7 days, but a maculopapular rash then develops on the trunk and extremities of about two thirds of patients and lasts about 3 days.

DIAGNOSIS

Mayaro virus may be isolated from blood by growth in Vero or C6/36 cells. RT-PCR with ELISA is available. A specific IgM is also available as an antibody capture ELISA.

TREATMENT AND PROGNOSIS

Rx

Treatment is supportive. Recovery is complete, although some patients have persistent arthralgia 6 months later.

ROSS RIVER FEVER VIRUS (EPIDEMIC FEBRILE POLYARTHRITIS)

Ross River fever virus causes an acute-onset, febrile illness characterized by rash and arthralgia. Ross River virus is a member of the family Togaviridae, genus Alphavirus.

EPIDEMIOLOGY

Epidemics of fever and rash have been observed in Australia since 1928. Isolation of Ross River virus from mosquitoes, its serologic association with epidemic polyarthritis, and isolation of the virus from epidemic polyarthritis patients in Australia confirmed Ross River virus as the etiologic agent of epidemic polyarthritis. Seroprevalence has been observed in endogenous populations in Papua New Guinea, western New Guinea, the Bismarck Archipelago, Rossel Island, and the Solomon Islands. An outbreak in the Fiji Islands affected more than 40,000 individuals in 1979 to 1980. A similar epidemic occurred in the Cook Islands early in 1980. Antibodies to Ross River virus are not found in individuals west of Weber's line, a hypothetical line separating the Australian geographic zone from the Asiatic zone. Endemic cases and epidemics occur in tropical and temperate regions in Australia. Queensland and New South Wales have a particularly high annual incidence associated with higher rainfall. High rainfall usually precedes epidemic periods, with cases subsequently occurring from spring through fall. Seroprevalence may reach just 6 to 15% in temperate coastal zones but is 27 to 39% in the plains of the Murray Valley river system. In Queensland, annual rates of disease range from 31.5 to 288.3 per 100,000 person-years. From 1992 to 2006, 55,000 cases of Ross River virus infection were reported in Australia.

Aedes vigilax is the major vector on the eastern coast of Australia and *Aedes camptorhynchus* in the salt marshes of southern Australia. *Culex annulirostris* is a freshwater breeding vector. Other Australian *Aedes* species and *Mansonia uniformis* may also serve as vectors. In outbreaks on the Pacific islands, *Aedes polynesiensis*, *A. aegypti*, *A. vigilax*, and *C. annulirostris* may have contributed to transmission. Domestic animals, rodents, and marsupials may serve as intermediate hosts. Virus may persist in *Aedes* mosquitoes.

There is a predominance of women among infected individuals. Children have a case attack rate ratio lower than that of adults. The incubation period is 7 to 11 days.

Barmah Forest virus, another alphavirus found in Australia in 1986, may be manifested in a fashion similar to epidemic febrile polyarthritis. The number of cases reported annually has been increasing since its initial discovery.

PATHOBIOLOGY

Ross River viral antigen may be detected in monocytes and macrophages early in infection, but intact virus is not identifiable by electron microscopy or cell culture. Dermal vessels show mild perivascular mononuclear cell infiltrates, mostly T lymphocytic, in erythematous and purpuric areas. Vessels in purpuric areas also show erythrocyte extravasation. Antigen can be demonstrated in epithelial cells in erythematous or purpuric skin and in the perivascular zone in erythematous skin. However, viral antigens have not been found in normal skin. Synovium undergoes lining cell hypertrophy and sublining vascular proliferation and mononuclear cell infiltration. Viral RNA can be identified by RT-PCR. Synovial fluid cell counts range from 1500 to 13,800 cells/mL and consist of monocytes, vacuolated macrophages, and a few neutrophils. Animal models of infection indicate that Ross River virus targets bone, joint, and skeletal muscle and elicits an inflammatory response mediated by the innate immune system.

CLINICAL MANIFESTATIONS

Arthralgia typically occurs abruptly, followed in 1 to 2 days by a macular, papular, or maculopapular rash that may be pruritic.[20] Three fourths of patients have severe, incapacitating arthralgia in an asymmetrical and migratory distribution. Commonly affected joints are the metacarpophalangeal joints, finger interphalangeal joints, wrists, knees, and ankles. The shoulder, elbow, toe, spine, hip, and temporomandibular joints may also be affected. Arthralgias are worse in the morning and after periods of inactivity. A third of patients have synovitis. Polyarticular swelling and tenosynovitis are common. Up to a third of patients have paresthesias or palm or sole pain. Classic carpal tunnel syndrome may occur.

In some individuals, rash may precede or follow the joint symptoms by 11 or 15 days, respectively. On occasion, vesicles, papules, or petechiae are seen. The trunk and extremities are typically involved, but the palms, soles, and face may also be affected. The rash resolves by fading to a brownish discoloration or by desquamation. Fever tends to be mild to moderate and lasts 1 to 3 days. Headache, nausea, and myalgia are common. Mild photophobia, respiratory symptoms, and lymphadenopathy may occur.

DIAGNOSIS

In the Australian epidemics before 1979, patients were antibody positive at the time of initial evaluation. However, in the Pacific island epidemics of 1979 to 1980, patients remained viremic and serologically negative for up to a week after the onset of symptoms. Virus in serum is stable for up to a month at 0° to −10° C. Current testing in Australia is performed with an indirect ELISA. The presence of specific IgM or evidence of seroconversion to IgG positivity supports a recent infection.

TREATMENT AND PROGNOSIS

Rx

Treatment is supportive. Nonsteroidal anti-inflammatory drugs provide relief of joint pain. Half of all patients return to activities of daily living within 4 weeks despite residual polyarthralgia. Joint symptoms may recur, but episodes gradually resolve. In some patients, joint symptoms may persist for up to 3 years. Mild exercise tends to improve the joint symptoms.

SINDBIS

Sindbis virus causes a sudden-onset, febrile illness associated with arthralgia and rash. It is known as Ockelbo disease in Sweden, Pogosta disease in Finland, and Karelian fever in the Karelian Isthmus of Russia. *Aedes*, *Culex*, and *Culiseta* mosquitoes transmit the virus to humans, with birds serving as intermediate hosts.

EPIDEMIOLOGY

The virus was first isolated from *Culex* mosquitoes in the Egyptian village of Sindbis in 1952. Outbreaks frequently occur in the forested areas of Sweden, Finland, and the Karelian Isthmus, but sporadic cases and small outbreaks have occurred in Uganda, South Africa, Zimbabwe, central Africa, and Australia. Individuals involved in outdoor activities or occupations are at greatest risk. In northern Sweden, 2.9% of the population has Sindbis virus–specific serum IgG positivity indicative of prior infection.[21]

PATHOBIOLOGY

Skin lesions show perivascular hemorrhage, lymphocytic infiltrates, edema, and areas of necrosis. Virus has been isolated from skin lesions. Antiviral IgM may persist for years, thus raising the possibility that Sindbis virus arthritis is associated with viral persistence and a direct viral effect on the synovium. Autophagy, an evolutionarily conserved intracellular mechanism for recycling cytoplasmic material to lysosomes for degradation during times of stress, may be disrupted in neurons by Sindbis virus infection, thereby leading to programmed cell death or apoptosis.

CLINICAL MANIFESTATIONS

Arthralgia and rash are the initial symptoms, although one may precede the other by a few days. Arthralgia and arthritis involve the small joints of the hands and feet, wrists, elbows, ankles, and knees. On occasion, arthralgia affects the spine. Tendinitis is common and often involves the Achilles and hand extensor tendons. Fever, if present, tends to be mild to moderate. Constitutional symptoms, headache, fatigue, malaise, nausea, vomiting, pharyngitis, and paresthesias may be present but are not usually severe. Macular rash typically begins on the torso and then involves the arms, legs, palms, soles, and occasionally the head. Macules evolve to papules that have a tendency to vesiculate. Vesiculation is prominent on pressure points, including the palms and soles. As the eruption fades, a brownish discoloration is left. Vesicles on the palms and soles may become hemorrhagic. Rash may recur during convalescence.

DIAGNOSIS

Specific IgM detected by enzyme immunoassay supports a diagnosis of Sindbis virus infection. IgM titers may wane during a period of 3 to 4 years.

TREATMENT AND PROGNOSIS

Rx

Treatment is supportive. Nonerosive chronic arthropathy is common in Sweden and Finland, with up to half of all patients having joint symptoms 2.5 years after infection. In a few cases, symptoms may persist for up to 6 years.

Grade A

Grade A Reference

A1. Villar L, Dayan GH, Arredondo-Garcia JL, et al. Efficacy of a tetravalent dengue vaccine in children in Latin America. *N Engl J Med*. 2015;372:113-123.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Brackney MM, Marfin AA, Staples JE, et al. Epidemiology of Colorado tick fever in Montana, Utah, and Wyoming, 1995-2003. *Vector Borne Zoonotic Dis*. 2010;10:381-385.
2. Bäck AT, Lundkvist A. Dengue viruses—an overview. *Infect Ecol Epidemiol*. 2013;3:19839.
3. Murray NE, Quam MB, Wilder-Smith A. Epidemiology of dengue: past, present and future prospects. *Clin Epidemiol*. 2013;5:299-309.
4. Hasan SR, Riaz M, Jafri FA. Characteristics and outcome of dengue infection; clinical perspective from a secondary care hospital of Karachi. *Pak J Med Sci*. 2013;29:115-118.
5. Chawla P, Yadav A, Chawla V. Clinical implications and treatment of dengue. *Asian Pac J Trop Med*. 2014;7:169-178.
6. Pongpan S, Wisitwong A, Tawichasri C, et al. Development of dengue infection severity score. *ISRN Pediatr*. 2013;2013:845876.
7. Petersen LR, Brault AC, Nasci RS. West Nile virus: review of the literature. *JAMA*. 2013;310:308-315.
8. Racsa L, Gander R, Chung W, et al. Clinical features of West Nile virus epidemic in Dallas, Texas, 2012. *Diagn Microbiol Infect Dis*. 2014;78:132-136.
9. Bailey MS, Trinick TR, Dunbar JA, et al. Undifferentiated febrile illnesses amongst British troops in Helmand, Afghanistan. *J R Army Med Corps*. 2011;157:150-155.
10. Kocak Tufan Z, Weidmann M, Bulut C, et al. Clinical and laboratory findings of a sandfly fever Turkey Virus outbreak in Ankara. *J Infect*. 2011;63:375-381.
11. Sow A, Faye O, Ba Y, et al. Rift valley fever outbreak, southern Mauritania, 2012. *Emerg Infect Dis*. 2014;20:296-299.
12. Caglioti C, Lalle E, Castilletti C, et al. Chikungunya virus infection: an overview. *New Microbiol*. 2013;36:211-227.
13. Enserink M. Infectious diseases. Crippling virus set to conquer Western Hemisphere. *Science*. 2014;344:678-679.
14. Kendrick K, Stanek D, Blackmore C. Notes from the field: Transmission of chikungunya virus in the continental United States—Florida, 2014. *MMWR Morb Mortal Wkly Rep*. 2014;63:1137.
15. Assunção-Miranda I, Cruz-Oliveira C, Da Poian AT. Molecular mechanisms involved in the pathogenesis of alphavirus-induced arthritis. *Biomed Res Int*. 2013;2013:973516.
16. Waymouth HE, Zoutman DE, Towheed TE. Chikungunya-related arthritis: case report and review of the literature. *Semin Arthritis Rheum*. 2013;43:273-278.
17. Schilte C, Staikowsky F, Couderc T, et al. Chikungunya virus–associated long-term arthralgia: a 36-month prospective longitudinal study. *PLoS Negl Trop Dis*. 2013;7:e2137.
18. Chang LJ, Dowd KA, Mendoza FH, et al. Safety and tolerability of chikungunya virus-like particle vaccine in healthy adults: a phase 1 dose-escalation trial. *Lancet*. 2014;384:2046-2052.
19. Theilacker C, Held J, Allering L, et al. Prolonged polyarthralgia in a German traveller with Mayaro virus infection without inflammatory correlates. *BMC Infect Dis*. 2013;13:369.
20. Lau C, Weinstein P, Slaney D. Imported cases of Ross River virus disease in New Zealand—a travel medicine perspective. *Travel Med Infect Dis*. 2012;10:129-134.
21. Ahlm C, Eliasson M, Vapalahti O, et al. Seroprevalence of Sindbis virus and associated risk factors in northern Sweden. *Epidemiol Infect*. 2013;1-7.

REVIEW QUESTIONS

1. A 24-year-old woman presents with 3 weeks of arthralgia involving the fingers and toes. The illness began suddenly with fever and headache 2 weeks ago, about 1 week after returning from her 2-year tour in the Peace Corps in Peru teaching English to children of Amazon lumber workers. Fever and headache resolved after the first week of illness. During the second week of illness, she noticed a blotchy red rash on her torso and arms. She denies joint swelling. Examination was unremarkable. Laboratory testing excluded parvovirus B19, rubella, Epstein-Barr virus, hepatitis B virus, hepatitis C virus, and coxsackievirus infections. The most likely diagnosis is

A. Chikungunya virus infection
B. Dengue virus infection
C. Mayaro virus infection
D. Ross River virus infection
E. Sindbis virus infection

Answer: C The geographic exposure would eliminate chikungunya virus, Ross River virus, and Sindbis virus as causative agents. Mayaro virus and dengue have overlap in their geographic distribution. The history suggests a milder course with mild arthralgias, which would argue against dengue, also called breakbone fever. At this juncture, the diagnosis can be confirmed by finding a positive Mayaro virus–specific serum IgM.

2. A 67-year-old retired businessman from Greenwich, Connecticut, with a known history of insulin-dependent type 2 diabetes mellitus and chronic sinusitis was in fair health until August, when he had onset of fever, headache, nausea, and myalgia. He noticed a nonpruritic macular rash over his torso soon after the fever began; 4 days later, the involved skin began peeling. His headache then began to worsen. During the second week of illness, he developed nuchal rigidity and became obtunded, at which time he was hospitalized. A social history revealed a pigeon fancier hobby. Examination revealed somnolence with difficult arousal, nuchal rigidity, and flaccid paralysis. A differential diagnosis of West Nile virus infection was entertained. At this juncture, the best way to demonstrate that the neurologic presentation is due to West Nile virus infection is

A. Detect West Nile virus in peripheral blood by reverse transcription–polymerase chain reaction (RT-PCR)–based detection of viral RNA
B. Detect West Nile virus IgM antibody in serum
C. Detect West Nile virus IgM *and* IgG antibody in serum
D. Save a serum sample at this time and demonstrate a four-fold increase in West Nile virus IgG in a convalescent serum sample obtained 3 weeks later
E. Detect West Nile virus IgM antibody in cerebrospinal fluid

Answer: E At this time, during the second week of illness, detection of viremia is unlikely. Demonstrating West Nile virus IgM in the serum confirms a recent West Nile virus infection but does not definitively demonstrate that the neurologic involvement is caused by West Nile virus in this elderly man with diabetes and chronic sinusitis. Note that the IgM may persist for up to 1 year in a minority of patients; subclinical infection may have occurred more remotely in this man. Detecting a four-fold increase in West Nile virus IgG in convalescent serum confirms a recent infection but does not definitively demonstrate that the neurologic involvement is due to West Nile virus. IgM normally does not cross the blood-brain barrier, so detecting West Nile virus IgM in the cerebrospinal fluid demonstrates central nervous system infection with the virus, although leakiness in the blood-brain barrier from another cause with passive movement of serum IgM into the cerebrospinal fluid cannot be completely excluded.

3. A 20-year-old male premedical student from Chicago enjoyed his spring break rock climbing in the Rocky Mountains with his girlfriend. One week later, during his organic chemistry midterm, he developed headache and myalgia. He attributed his symptoms to the previous all-nighter studying for his examination. That evening he developed photophobia, eye pain, and weakness. The following morning, he noticed punctate red dots, which did not blanch on pressure, over his torso. In the Student Health Service, palatal petechiae, petechial rash on the torso, generalized lymphadenopathy, and mild splenomegaly were noted. Laboratory testing showed mild leukopenia and thrombocytopenia. A presumptive diagnosis of Rocky Mountain spotted fever was made by the Health Service Director, and the patient was started on doxycycline. However, you begin to doubt the diagnosis the following day when the patient fails to develop the classic spotted fever rash (see Chapter 327). Diagnostic tests are pending. The likely diagnosis is

A. Colorado tick fever
B. Dengue
C. Disseminated gonorrhea
D. Rocky Mountain spotted fever
E. None of the above

Answer: A The exposure and course are consistent with Colorado tick fever. Dengue is unlikely because the Rockies are outside of the geographic distribution for dengue, which is found in the southeast United States and other tropical and subtropical climates. Patients with disseminated gonorrhea present with an arthritis-dermatitis syndrome, in which generalized arthralgias and myalgias quickly settle into one or more locations with inflammatory arthritis or tenosynovitis; scattered 1- to 3-mm pustules may evolve necrotic centers. This patient lacked inflammatory arthritis or tenosynovitis; his exanthem consisted of multiple petechiae rather than pustules, and the lesions were more plentiful than the pustules typically seen in disseminated gonorrhea. The rash argues against Rocky Mountain spotted fever, in which the "spotted" rash of 1- to 5-mm macules classically becomes purpuric. The patient had petechiae that did not blanch on pressure. However, the Health Service Director was not incorrect in starting doxycycline. The classic rash can appear later or not at all in the course of Rocky Mountain spotted fever, and serologic diagnosis often requires acute and convalescent sera. Doxycycline can be life-saving when it is started early in suspected Rocky Mountain spotted fever.

4. An 18-year-old U.S. Marine stationed in Helmand Province, Afghanistan, flew home on emergency leave to see her sick father in Ottumwa, Iowa. Four days after arriving, she developed headache, stiff neck, eye pain, photophobia, arthralgia, and chest pain. Examination showed a macular rash measuring 2 × 4 cm with scattered papules within it on her left calf. Laboratory testing showed leukopenia and lymphopenia. Cerebrospinal fluid showed mild pleocytosis. The likely diagnosis is

A. Chikungunya fever
B. Dengue
C. O'nyong-nyong fever
D. Phlebotomus fever
E. Ross River fever

Answer: D In Afghanistan, the potential for exposure to the sandfly vector is significant. These small flies may pass through mosquito netting. Among British soldiers in Helmand Province, 52% of undifferentiated fever cases were attributed to phlebotomus fever. Myalgias may be sufficiently severe to be perceived as pleurodynia. Mild aseptic meningitis is common. Chikungunya fever is endemic in sub-Sahara Africa, India, the Philippines, and Southeast Asia and has spread to islands in the Indian Ocean and the Caribbean, but it is not found in the mountains of Afghanistan. Likewise, dengue is distributed in tropical and subtropical climates. Recent outbreaks of o'nyong-nyong fever have occurred in Uganda but not in Afghanistan. Ross River virus is limited to east of Weber's line, the virtual line that separates the Australian biogeographic zone from the Asian biogeographic zone. Furthermore, the rash in Ross River virus infection is a more widely distributed maculopapular rash with occasional vesicles, papules, or petechiae over the torso, extremities, face, palms, and soles; the rash also may desquamate. The rash described in this patient is more localized and is attributable to the sandfly bites.

5. A 26-year-old hipster born and living in Manhattan spends winters in Jamaica with friends. Shortly after returning to New York this year, she experienced sudden-onset fever, severe headache, eye pain, nausea, vomiting, myalgia, and arthralgia. By the next day, the joint and muscle pain became severe. Two days later, she sought medical attention. Examination showed an ill-appearing woman, alert but in apparent pain. Temperature was 100.5° F, pulse 104/minute, blood pressure 92/50 mm Hg, and respirations 28/minute. Multiple petechiae appeared on her left arm distal to where the blood pressure cuff was applied. Cervical and inguinal lymph nodes were palpable and nontender. Joint examination was unremarkable. Laboratory testing showed white cell count of 1500/μL and platelet count of 55,000/μL. Hemoglobin was 19 g/dL. Prothrombin time was 22 seconds. The pregnancy test was positive. A diagnosis of dengue virus infection was suspected. Dengue virus serology was positive for IgG antibody to serotype 1 as well as for IgM and IgG antibodies to serotype 2. The diagnosis is

A. Dengue virus infection, classic
B. Dengue virus infection, hemorrhagic fever
C. Dengue virus infection, shock syndrome
D. Dengue virus infection, classic with preeclampsia

Answer: B The patient began with classic dengue features but developed low-normal blood pressure, tachycardia, and hemoconcentration suggestive of volume loss. The positive tourniquet sign with the development of petechiae distal to the blood pressure cuff is a warning sign that hemorrhagic fever has developed. The presence of thrombocytopenia and prolonged prothrombin time provide laboratory support. The serologic evidence of a previous dengue virus infection with serotype 1 (IgM–, IgG+) supports the increased risk for development of hemorrhagic fever due to enhancement of immune response by preexisting antiviral antibodies. The patient has not yet developed hemorrhagic shock syndrome, but prompt intervention with supportive measures should improve prognosis. The patient is pregnant, which may worsen the clinical picture, but her low blood pressure argues against preeclampsia.

383

ARBOVIRUSES AFFECTING THE CENTRAL NERVOUS SYSTEM

THOMAS P. BLECK

Arboviruses, which are also termed *arthropod-borne viruses*, can affect the central nervous system (CNS). These viruses share a number of clinical and epidemiologic similarities and have an RNA genome, but they do not form a formal virologic taxonomic group. Arboviruses generally have avian or small mammalian reservoirs and are transmitted to humans and other large mammals incidentally when an infected mosquito or other arthropod obtains a blood meal.

EPIDEMIOLOGY

Most human disease is subclinical; a few patients have a brief febrile illness resembling influenza, and a small percentage, usually at the extremes of age, suffer meningitis or encephalitis. The diseases (Table 383-1) reflect the quotidian and seasonal characteristics of their insect vectors. Other viruses of the same genera cause hemorrhagic fever (Chapter 381), and other less frequently encountered arboviruses are also capable of producing encephalitis.

Many of these agents cause notifiable disease in the United States: St. Louis, West Nile, Powassan, eastern equine, western equine, and the California serogroup encephalitis viruses. Case definitions and additional information are available at http://www.cdc.gov/ncidod/dvbid/arbor/index.htm.

The diseases described here are zoonoses (Chapter 328), that is, illnesses caused by viruses transmitted from animals to humans. They are more prevalent in the tropics and subtropics and are usually localized because of ecologic restrictions on their transmission.

PATHOBIOLOGY

Two pathologic processes are common to the arboviral encephalitides: neuronal and glial damage mediated by intracellular viral infection; and migration of immunologically active cells into the perivascular space and brain parenchyma. Endothelial cell swelling and proliferation, destruction of myelin sheaths in deep white matter areas, and vasculitis are present in some arboviral encephalitides.

After a bite by an infected arthropod, viral replication occurs in local tissues and regional lymph nodes. Viremia, which seeds extraneural tissues, occurs and persists, depending on the extent of replication in extraneural sites, the rate of viral clearance by the reticuloendothelial system, and the appearance of humoral antibodies. The sites of extraneural infection vary among the viruses. Many alphaviruses and flaviviruses involve striated muscle and endothelium, whereas Venezuelan encephalitis virus is associated with myeloid and lymphoid tissue invasion. During viremia, the neural parenchyma may be invaded, but the mode of penetration of virus across the blood-brain barrier is not completely understood. Possible mechanisms include passive movement of virus across vascular membranes and viral replication in the cerebral capillary endothelium. Factors that increase vascular permeability or disrupt the blood-brain barrier promote invasion of the nervous system. Infected monocytes may also bring the virus into the CNS. In experimental animal infection, flaviviruses enter the CNS through the olfactory epithelium.

The immune response to flaviviruses starts with an innate interferon response to viral replication. Neurons then produce chemokines that recruit various components of the cellular immune response. The induced T-cell and monocyte trafficking is needed to clear the virus from the CNS, but it can also damage neurons.

The immature brain is more susceptible to damage by western equine, Venezuelan equine, and California serogroup encephalitis viruses (Table 383-2). St. Louis encephalitis and West Nile encephalitis principally affect the elderly, whereas Japanese encephalitis and eastern equine encephalitis have a bimodal incidence and strike both children and elderly persons. In endemic areas, immunity accumulated with increasing age may reduce the incidence of disease in older persons for some viruses; however, the reasons for the increased severity of illness with other viruses remain unknown.

TABLE 383-1 ARTHROPOD-BORNE VIRUSES ASSOCIATED WITH HUMAN ENCEPHALITIS

VIRUS	INSECT VECTOR	COMMON VERTEBRATE HOSTS	GEOGRAPHIC DISTRIBUTION
TOGAVIRIDAE			
Alphaviruses	Mosquitoes		
Eastern equine encephalitis	*Culiseta* spp, *Aedes* spp, *Coquillettidia* spp		Eastern United States and Gulf Coast, Caribbean region, South America
Western equine encephalitis	*Culiseta* spp, *Culex* spp		Western United States, Canada
Venezuelan equine encephalitis	*Aedes* spp, *Culex* spp, *Psorophora* spp, and *Mansonia* spp		South America, Central America, Florida and southwestern United States
FLAVIVIRIDAE			
Japanese serocluster	Mosquitoes		
Japanese encephalitis	*Culex* and *Aedes* spp		East and Southeast Asia, India, Australia
West Nile encephalitis	*Aedes* spp, *Culex* spp, and others		Africa, Middle East, North America
St. Louis encephalitis	*Culex* spp		Western Hemisphere
Murray Valley encephalitis	*Culex* spp		Australia
Tick-borne encephalitis complex			
Central European encephalitis	*Ixodes* spp	Goats, sheep	Europe, Russia
Russian spring-summer encephalitis	*Ixodes* spp		Europe, northern and central Asia
Kyasanur Forest disease	*Haemaphysalis spinigera*	Rodents, insectivores	India
Omsk hemorrhagic fever	*Dermacentor reticulatus*	Rodents	Central Asia
Powassan	*Ixodes* spp	Squirrels, groundhogs	North America, Russia
Louping ill	*Ixodes ricinus*	Small mammals, sheep, birds	British Isles
Langat	*Ixodes* spp	Rodents	Malaysia, Thailand, parts of former Soviet Union
BUNYAVIRIDAE			
California encephalitis	*Aedes melanimon, Aedes dorsalis*	Rodents, rabbits	California
La Crosse encephalitis	*Aedes triseriatus*	Chipmunks, squirrels	Eastern and Midwestern United States

TABLE 383-2 FEATURES OF ARBOVIRAL ENCEPHALITIDES IMPORTANT IN THE UNITED STATES

	EASTERN EQUINE ENCEPHALITIS	WESTERN EQUINE ENCEPHALITIS	VENEZUELAN EQUINE ENCEPHALITIS	WEST NILE ENCEPHALITIS	ST. LOUIS ENCEPHALITIS	CALIFORNIA SEROGROUP ENCEPHALITIS
Annual U.S. cases of symptomatic disease	10	0-2 cases, mostly infants and children	Rare, mostly children	Up to 3000, mostly >40 years	0-2000, mostly >50 years	10-50, mostly children
Time of year	Late summer, early fall	Early and mid summer	Summer	Summer, fall	Mid to late summer	July-September
Case-fatality rate	50-70%, highest in children <15 years and adults >55 years	3-5% in children	35% in children, <10% in older persons	14-19%, 30% in adults >70 years	9% overall; 0% <20 years, 30% >65 years	<1%
Residual damage	30-50%, especially in children	33% in infants	Frequent in children	50%, more frequent in elderly	Frequent in elderly	Probably rare
Cerebrospinal fluid findings (cells/μL)	500-2000 cells, predominantly neutrophils	<500 cells, predominantly lymphocytes	<500 cells, predominantly lymphocytes	<500 cells, predominantly lymphocytes	<500 cells, predominantly lymphocytes	<500 cells, predominantly lymphocytes

CLINICAL MANIFESTATIONS

Clinical symptoms and signs vary among the viral causes (see later), although all share common signs and symptoms of encephalitis (Chapter 414).

DIAGNOSIS

Diagnosis depends on a careful history that includes exposure to vertebrate animals and arthropod vectors, age, season, and travel, including the geographic site of exposure. Laboratory confirmation of infection is essential. The virus may be isolated from acute phase serum or whole blood in laboratory animals or in tissue culture. Neutralization, complement fixation (CF), hemagglutination inhibition (HI), fluorescent antibody, and enzyme-linked immunosorbent assay (ELISA) of acute and 3-week convalescent sera can also produce the correct diagnosis. Antigen detection and IgM capture ELISA often permit diagnosis on initial evaluation and within a week of the onset of illness in most cases. Sensitive nucleic amplification assays using reverse transcription–polymerase chain reaction (RT-PCR) are under development for a number of the arboviruses and may lead to earlier diagnosis.

Differential Diagnosis

The most important initial consideration is to differentiate arboviral encephalitides from other acute CNS infections, including infections other than encephalitis (Chapters 412 and 413), treatable causes of encephalitis (Chapter 414), and paraneoplastic and autoimmune encephalitis (Chapters 412 and 414). Anti–*N*-methyl-D-aspartate (NMDA) receptor encephalitis, which is as common as most viral causes of encephalitis in parts of the United States,[1] is an important part of the differential diagnosis, particularly in young women and especially because—unlike arboviral infections—it often responds to immunosuppressive treatment.

The early prodrome resembles influenza (Chapter 364). Bacterial meningitis (Chapter 412; especially early or partially treated), infective bacterial endocarditis (Chapter 76), brain abscess (Chapter 413), subdural empyema (Chapter 413), and cerebral thrombophlebitis may mimic viral encephalitis, and the cerebrospinal fluid (CSF) profile is sometimes similar. Other infections that occasionally cause meningoencephalitis that may resemble arthropod-borne viral encephalitis include tuberculosis (Chapter 324), cryptococcosis (Chapter 336), histoplasmosis (Chapter 332), coccidioidomycosis (Chapter 333), Rocky Mountain spotted fever (Chapter 327), leptospirosis (Chapter 323), falciparum malaria (Chapter 345), trichinosis (Chapter 357), *Naegleria* meningitis (Chapter 412), typhoid fever (Chapter 308), Lyme disease (Chapter 321), and *Mycoplasma* pneumonia (Chapter 317).

Acute meningoencephalitis may result from infections with other viruses, including herpesviruses (Chapter 374), human immunodeficiency virus (Chapter 386), mumps virus (Chapter 369), enteroviruses (Chapter 379), lymphocytic choriomeningitis virus (Chapter 412), rabies (Chapter 414), influenza (Chapter 364), and the exanthematous viral infections of childhood (Chapters 367 and 368). The exposure history, the presence of similar disease in the community, and the summer-fall occurrence are principal clues to an arboviral etiology. Enteroviruses (Chapter 379) also cause summer-fall outbreaks, but the predominant syndrome is aseptic meningitis, and the

concomitant occurrence of rash or pleurodynia is a helpful clue. Herpes simplex encephalitis (Chapter 414) presents an important diagnostic challenge because effective therapy is available and should be started quickly. The presence of localizing neurologic signs, localizing findings on computed tomography (CT) or magnetic resonance imaging (MRI), and detection of herpes simplex DNA in CSF by PCR help distinguish herpes simplex encephalitis from the arboviral encephalitides.

Noninfectious diseases of the CNS, such as stroke (Chapter 407), may rarely be confused with viral encephalitis. Subarachnoid hemorrhage (Chapter 408) produces meningismus, fever, headache, and neurologic signs that mimic an infectious etiology. Metabolic encephalopathies occasionally have features suggesting infectious encephalitis. Neoplastic or granulomatous diseases involving the CNS and a variety of diseases of uncertain etiology (Behçet disease [Chapter 270], Reye syndrome, acute multiple sclerosis [Chapter 411], and systemic lupus erythematosus [Chapter 266]) must be considered in the differential diagnosis as well.

PREVENTION

Control can be achieved by interruption of the cycle, including vaccination of reservoir animals, vector control, and education on vector avoidance. Practical measures include wearing long-sleeved clothing, using insect repellents, limiting outdoor activities during peak mosquito season, and eliminating standing pools of water. Vaccines are currently available for Japanese encephalitis.

TREATMENT Rx

Treatment is symptomatic and may include bedrest, antipyretics, and analgesics. Early empirical treatment of herpes simplex encephalitis (Chapter 374) may be appropriate while the diagnostic evaluation to document arboviral encephalitis proceeds. To date, no immunologic therapy has demonstrated a useful effect in humans.

EASTERN EQUINE ENCEPHALITIS

EPIDEMIOLOGY

Human disease is relatively rare, with fewer than 10 cases occurring each year in the Gulf Coast and Atlantic states, usually in association with an equine epizootic involving 100 to 300 animals. Outbreaks generally occur during the late summer and early fall. The occurrence of equine cases or outbreaks of fatal encephalitis in penned exotic birds precedes the appearance of human cases by several weeks or more. Epizootics of eastern equine encephalitis have been reported in the Caribbean (Hispaniola) and South America.[2]

In temperate areas, eastern equine encephalitis virus circulates between wild birds and *Culiseta melanura* mosquitoes in a freshwater swamp habitat. Equine epizootics and associated human cases result from extension of the transmission cycle to involve *Aedes* and *Coquillettidia* mosquitoes, which feed on horses and humans.[3]

PATHOBIOLOGY

The brain is grossly edematous and congested, and the inflammatory response is predominantly polymorphonuclear. The areas most affected are the basal ganglia, thalamus, hippocampus, and frontal and occipital cortices. Focal vasculitis, endothelial cell swelling, intravenous and arteriolar thrombus formation, demyelination, necrosis, neuronolysis, and neuronophagia are prominent. Eastern equine encephalitis virus appears to make use of host micro-RNAs to limit replication in myeloid cells, thereby restricting the host immune response and causing more severe neurologic damage.[4]

CLINICAL MANIFESTATIONS

Onset is abrupt, with high fever, vomiting, and somnolence. Stupor, coma, myoclonus, and generalized convulsions appear within 24 hours to as long as 10 days later. Autonomic disturbances (sialorrhea) may be prominent, and respiratory difficulty and cyanosis are frequent. In children, facial, periorbital, or generalized edema may be present.

A striking peripheral leukocytosis with immature neutrophils occurs frequently in patients with eastern equine encephalitis. CSF examination reveals 500 to 2000 white blood cells/μL (predominantly neutrophils). As the total cell count falls, neutrophils may persist as a significant fraction. Red blood cells may be present, protein concentration is elevated, and the glucose level is normal.

DIAGNOSIS

Brain CT and MRI are frequently abnormal and reveal lesions in the basal ganglia, thalami, and brain stem. The virus can rarely be isolated from blood or CSF. Serologic diagnosis by demonstration of a rise in antibody titer in appropriately timed paired sera is the most practical and available test. Because of the rapid course of the clinical disease, sera should be obtained at 2- to 3-day intervals during the acute phase of illness.

PREVENTION

An experimental formalin-inactivated chick embryo cell culture vaccine is used to protect laboratory and field workers. Reduction of mosquito populations by appropriate use of insecticides may be effective in threatened or established outbreaks.

TREATMENT Rx

Treatment is supportive. Control of fever, intracranial pressure, seizures, fluid and electrolyte disturbances, and the airway is critical. Although attempts at immunologic therapy have been reported, no controlled data are available.

PROGNOSIS

The case-fatality rate is 50 to 70%. Mortality, like incidence, is highest in children younger than 15 years and in persons older than 55 years, with no gender predilection. Death usually occurs during the first week; in surviving patients, recovery begins during the second week and may progress rapidly. Good functional recovery is associated with a long prodromal course and absence of coma. Residual damage, found in 30 to 50% of patients, is often severe, especially in children, and is characterized by mental retardation, spastic paralysis, and radiographic evidence of brain atrophy.

WESTERN EQUINE ENCEPHALITIS

EPIDEMIOLOGY

Few cases of western equine encephalitis have been reported in recent decades; the most recent epidemic occurred in Colorado in 1987. Epidemics occur in early or mid summer and may follow heavy snow melt or flooding, conditions favorable for breeding of mosquitoes. Cases of encephalitis in equines often precede the appearance of human disease. The illness principally affects residents of rural communities, and the incidence is higher in males than in females.

The ratio of inapparent to apparent infection is also age dependent and ranges from about 1 : 1 in infants younger than 1 year, to 58 : 1 in children 1 to 4 years old, to more than 1000 : 1 in persons older than 14 years. Western equine encephalitis virus also occurs in South America. Equine epizootics in Argentina have been associated with human cases.

Western equine encephalitis virus circulates between wild birds and *Culex tarsalis* mosquitoes. *Cx. tarsalis* is responsible for infection of humans and equines, which have low or undetectable viremia and do not perpetuate the chain of transmission. In temperate areas, transmission ceases during the winter months.

PATHOBIOLOGY

Pathologic examination of the brains of infants reveals massive parenchymal destruction; children dying months or years after the acute insult often have large cystic lesions in many areas of the brain. In older children and adults, acute western equine encephalitis is characterized by focal necrosis and perivascular cuffing, predominantly in the basal ganglia and thalami but also in deep cerebral white matter.

CLINICAL MANIFESTATIONS

The disease usually begins with an influenza-like illness consisting of fever, headache, malaise, and myalgia lasting 1 to 4 days. Somnolence, lethargy, photophobia, vomiting, and neck stiffness may follow; neurologic involvement may rapidly progress to stupor, coma, and seizures. Pareses, cranial nerve deficits, tremors, and abnormal reflexes may be present. In fatal cases,

patients die 1 to 2 days after coma develops. Congenital infections have been documented and result in severe and progressive neurologic deterioration.

Leukocytosis and a shift to the left are common. The CSF contains fewer than 500 white blood cells/μL (at first polymorphonuclear, then mononuclear) and an elevated protein concentration (usually 90 to 110 mg/dL).

DIAGNOSIS

Viral isolation from blood or CSF is almost never successful. Diagnosis is achieved by demonstration of a rise in HI, fluorescent antibody, CF, ELISA, or neutralizing antibody titers in appropriately timed (10 to 14 days apart) paired sera. Demonstration of IgM antibodies in serum or CSF by ELISA provides a presumptive diagnosis.

PREVENTION

An experimental formalin-inactivated vaccine grown in chick embryo cell culture has been used to protect laboratory workers but is not indicated for others. In threatened or ongoing epidemics, residents should be advised to use protective clothing, insect repellents, and window screens and to restrict outdoor activity in the early morning, late afternoon, and evening (times of greatest mosquito activity). Public health measures include spraying insecticides aimed at the adult *Cx. tarsalis* vector.

TREATMENT Rx

There is no specific therapy for western equine encephalitis. Supportive therapy is similar to that discussed earlier for eastern equine encephalitis.

PROGNOSIS

Western equine encephalitis is most severe in infants and young children. The case-fatality rate is between 3 and 5%. Survivors generally experience sudden and rapid recovery. However, about a third of surviving infants suffer mental retardation, cerebellar damage, choreoathetosis, and spastic paralysis. Children with protracted illnesses in whom convulsions develop during the acute stage are more likely to suffer long-term neurologic sequelae. Adults may have a prolonged convalescent syndrome, but objective residua are rare.

VENEZUELAN EQUINE ENCEPHALITIS

Six antigenic subtypes of Venezuelan equine encephalitis virus (I to VI) with several antigenic variants of subtypes I and III are recognized serologically. Subtypes IAB and IC are responsible for epidemics involving humans and equines. In Florida, subtype II is enzootic and produces sporadic human disease. Methods of transmitting Venezuelan equine encephalitis virus as a biologic warfare agent were developed in the 1960s; an epidemic of Venezuelan equine encephalitis, especially if humans and horses become ill simultaneously, could represent an attack rather than naturally occurring illness.

EPIDEMIOLOGY

Before 1973, large equine epizootics occurred at 5- to 10-year intervals in Venezuela, Colombia, Ecuador, and Peru and involved many thousands of animals with mortality rates as high as 40%. Associated human morbidity was also great (up to 32,000 clinical cases). The disease was quiescent for several years but has reemerged in the Gulf Coast region of Mexico in the past decade. The last major outbreak occurred in Venezuela and Colombia in 1995, with more than 85,000 human cases. Laboratory infections are common in unvaccinated persons who work with the virus or infected animals.

A large variety of mosquito vectors, including species of the genera *Aedes*, *Psorophora*, and *Mansonia*, transmit subtypes IAB and IC during epizootic epidemics. Equines are the principal viremic hosts. Virus may be present in the pharyngeal excretions of human patients; contact or aerosol person-to-person spread, although possible, is not epidemiologically important.

The other members of the Venezuelan equine encephalitis viral complex, including subtype II in Florida, have enzootic transmission cycles involving *Culex* mosquitoes and small forest rodents and marsupials. Human disease is sporadic and relatively uncommon.

PATHOBIOLOGY

Pathologic changes in the CNS include edema, congestion, meningeal and perivascular inflammation, intracerebral hemorrhage, neuronal degeneration, and vasculitis. In addition, hepatocellular degeneration and necrosis, widespread lymphoid depletion and follicular necrosis, and interstitial pneumonitis are frequent. Congenitally infected fetuses demonstrate massive and widespread necrosis of brain tissue, hemorrhage, and resorption of brain material resulting in hydranencephaly.

CLINICAL MANIFESTATIONS

The predominant syndrome is a self-limited influenza-like illness; encephalitis develops in only about 4% of infected persons, principally children younger than 15 years. Subclinical infections are rare.

After an incubation period of 2 to 5 days, there is a sudden onset of fever, chills, malaise, and headache, followed by myalgias, nausea, vomiting, and occasionally diarrhea. Physical examination reveals fever, tachycardia, conjunctival injection, and, in some cases, nonexudative pharyngitis. The acute illness generally subsides in 4 to 6 days, and convalescent symptoms may last up to 3 weeks. A biphasic course has sometimes been noted; acute symptoms can reappear after a brief remission, within a week after initial onset.

When it occurs, severe encephalitis is characterized by meningeal signs, seizures, tremor, stupor, coma, spastic paralysis, abnormal reflexes, cranial nerve palsies, and central respiratory failure. Residual neurologic damage occurs in severe cases. Infections of pregnant women acquired during the first and second trimester may result in fetal encephalitis and death.

The peripheral leukocyte count is often low, with a decrease in both lymphocytes and neutrophils, or normal with relative lymphopenia. In patients with CNS signs, the CSF contains up to 500 cells/μL, predominantly lymphocytes. Serum lactate dehydrogenase and aspartate aminotransferase concentrations may be elevated.

DIAGNOSIS

In contrast to the other arboviral encephalitides, Venezuelan equine encephalitis virus can be isolated from blood or from throat swabs or washings during the first 3 or 4 days of illness. Serodiagnosis is usually more practical and is achieved by testing appropriately timed paired sera by HI, CF, ELISA, neutralization, or IgM immunoassay.

PREVENTION

An experimental live attenuated vaccine made from subtype IAB is used for adult laboratory personnel. It provides solid immunity to subtypes IAB and IC but incomplete protection against heterologous Venezuelan equine encephalitis viruses. Epidemics and epizootics can be prevented by effective vaccination of equines. Spraying insecticides to reduce adult (infective) mosquito populations is the only means of immediate control in the face of an ongoing epidemic. Individual protection against mosquitoes is advised.

TREATMENT Rx

No specific therapy is available, and treatment of encephalitis cases is supportive.

PROGNOSIS

The case-fatality rate in children 5 years or younger with encephalitis is approximately 35%, but it is less than 10% in older persons.

JAPANESE ENCEPHALITIS

EPIDEMIOLOGY

Japanese encephalitis virus is a flavivirus that causes epizootics of clinical encephalitis in equines. The disease occurs throughout Asia, including Japan, China, the Korean peninsula, Taiwan, Okinawa, Vietnam, the Philippines, Burma, Malaysia, Bangladesh, east and south India, Sri Lanka, Thailand, and Indonesia. More than 30,000 clinical cases occur annually, about one third of which are fatal. Japanese encephalitis is a summertime disease in temperate areas but occurs sporadically year-round in the tropics. Several species of *Culex* mosquitos can transmit the virus, most notably *Cx. tritaeniorhynchus*.

Epidemics are most frequent at the northern fringe of the tropical zone, with a high incidence noted in southern China.[5] It is a predominantly rural disease, and the incidence in males is often higher than in females. In hyperendemic areas, more than 70% of adult populations surveyed have antibodies, and children younger than 15 years are principally affected by the disease.

In areas without a high prevalence of background immunity (e.g., northern India), however, all age groups are affected. In Japan, where schoolchildren have been protected by vaccination campaigns targeted at this age group, encephalitis has become prominent in the elderly. The ratio of clinically inapparent to apparent infection is higher than 500 : 1 in children and decreases with age; in Korea, the ratio in American servicemen was estimated at 25 : 1.

PATHOBIOLOGY

Neuropathologic changes and the distribution of lesions are similar to those described for St. Louis encephalitis (see later).

CLINICAL MANIFESTATIONS

Manifestations of Japanese encephalitis include abrupt fever, headache, and gastrointestinal symptoms. Meningeal irritation develops within 24 hours and is followed on the second or third day by the appearance of irritability, impaired consciousness, seizures (especially in children), muscle rigidity, parkinsonian findings, ataxia, coarse tremor, involuntary movements, cranial nerve deficits, paresis, hyperactive deep tendon reflexes, and pathologic reflexes. Weight loss and dehydration are often striking findings. In mild cases, fever subsides after the first week, and neurologic signs resolve by the end of the second week after onset. In severe cases, hyperpyrexia, progressive neurologic dysfunction, and coma result in death, usually between the seventh and tenth days. About 25% of patients undergo a prolonged recovery, with permanent sequelae often remaining. The occurrence of such sequelae correlates with severity of the acute stage of illness, and young children are most susceptible. Cardiorespiratory complications are frequent during the acute stage in these patients. A poor prognosis is associated with protracted high fever, frequent or prolonged seizures, high protein content in CSF, Babinski signs, and early respiratory depression. Fetal death from transplacental Japanese encephalitis infection has been reported.

DIAGNOSIS

Moderate peripheral leukocytosis and neutrophilia occur early in the disease. Pleocytosis (predominantly lymphocytic), protein elevation, and normal glucose concentration in CSF are usual findings.

MRI in Japanese encephalitis reveals edema in the basal ganglia, thalami, and focal areas of the cerebral cortex; evidence of hemorrhage in these areas may likewise be present. Enhancement may also be noted in the meninges, brain stem, and spinal cord.

Virus is rarely isolated from blood. Virus is also rarely recovered from the CSF of patients who live but may be recovered from the CSF of a third of patients who die. HI and neutralizing antibodies appear during the first week, and CF antibodies appear during the second week. Cross-reactions with other flaviviruses make serodiagnosis difficult. Specific IgM antibodies in serum or CSF are detectable by immunoassays in more than three fourths of patients at the time of hospital admission.

PREVENTION

Newer inactivated Vero cell culture–derived vaccines, including IC51 (marketed as Ixiaro and Jespect), CC-JEV (marketed as Encevac), and Jenvac, are available for use in persons older than 17 years and have recently been shown to be safe in children. They are more immunogenic than the older inactivated mouse brain–derived vaccine. A combined yellow fever–Japanese encephalitis vaccine (ChimeriVax-JE) is available in Australia and Thailand. A new chimeric vaccine, which is effective when it is administered concomitantly with measles-mumps-rubella vaccine, holds promise for children in endemic areas. In China, a live attenuated vaccine (SA14-14-2) is commonly used.

Ixiaro is licensed in the United States for individuals aged 2 months and older traveling to high-risk areas.[6] Generalized urticaria and angioedema may occur in 0.3%. Because two doses of the inactivated vaccine are used and approximately 1 month is required to confer protection, vaccination is not a practical measure in the event of an ongoing epidemic.[7]

Reduction of vector mosquito populations by the application of insecticides may help abort outbreaks. Immunization of swine is an ancillary control strategy.

TREATMENT

Rx

Treatment is supportive (see Eastern Equine Encephalitis). A randomized trial of interferon alfa showed no benefit.

PROGNOSIS

The case-fatality rate is probably about 25%. Sequelae such as mental impairment, emotional lability, choreoathetosis, tremor, parkinsonism, autonomic disturbances, paralysis, and psychiatric disturbances have been reported in up to 75% of patients.

WEST NILE FEVER AND ENCEPHALITIS

Before 1996, the predominant clinical manifestation of West Nile virus was a brief influenza-like illness, sometimes with a rash (Chapter 382) but infrequently with neurologic manifestations. Epidemics since then in Romania, Israel, and North America have added meningitis, meningoencephalitis, and myelitis to the list of disorders attributable to the virus.[8] Sequence analyses of various isolates of the virus indicate two lineages. The first includes viruses from North America, Europe, Israel, western Africa, India, Russia, and Australia; the second includes viruses from sub-Saharan Africa and Madagascar.

EPIDEMIOLOGY

See Chapter 382. The incidence of neuroinvasive West Nile disease varies in different parts of the United States but can be seen nearly everywhere (E-Fig. 383-1).[9] In a recent outbreak in Dallas, Texas, the incidence rate for West Nile encephalitis was 7.3 per 100,000 residents, with cases clustered in neighborhoods with high housing density.[10]

CLINICAL MANIFESTATIONS

The majority of infected people are asymptomatic. Fever develops in about 20%, and CNS manifestations are seen in less than 1%, although this percentage is higher in the elderly (Chapter 382). In those in whom clinical manifestations develop, an incubation period of 1 to 6 days is followed by the abrupt onset of symptoms, usually without a prodrome. The temperature rises quickly to 38.3° to 40.0° C, with rigors in a third of patients. Symptoms include drowsiness, severe frontal headache, ocular pain, myalgia, and pain in the abdomen and back. A small number of patients have dryness of the throat, anorexia, and nausea. Cough is common. Examination shows facial flushing, conjunctival injection, and coating of the tongue. Generalized lymphadenopathy had been a prominent feature in past epidemics but is no longer commonly reported. The spleen and liver are occasionally slightly enlarged. The temperature curve may be biphasic. A pale roseolar maculopapular rash, predominantly on the trunk and upper part of the arms, may appear from the second to fifth day but is now less common as well; it may be evanescent (several hours) or persist until defervescence, and it does not desquamate. Vesicular lesions occur rarely. The illness lasts 3 to 5 days in 80% of patients. Clinical manifestations appear to be more frequent in organ transplant recipients.

In the past decade, the incidence of CNS disease has increased in several epidemics, apparently because of a true increase in the invasiveness and neurovirulence of the virus. The virus causes a syndrome resembling poliomyelitis (Chapter 379), with prominent lower motor neuron dysfunction (acute flaccid paralysis with asymmetrical weakness and decreased deep tendon reflexes but preserved sensory function), which may be seen independently or with signs of meningoencephalitis. In some patients, prolonged, possibly permanent ventilatory failure requiring mechanical ventilation develops. About 95% of patients present with neuroinvasive disease and a significant neurologic defect; about 50% have weakness, 35% have tremor, and 15% have cranial neuropathies.[11] Some patients may have seizures, cranial nerve involvement, ataxia, tremors, or myoclonus. Acute inflammatory polyneuropathy (e.g., Guillain-Barré syndrome; Chapter 420) has also been reported. Other rare complications include myocarditis, pancreatitis, and hepatitis. Convalescence is often prolonged, lasting several weeks with prominent symptoms of fatigue. Lymph node enlargement requires several months to regress. Laboratory findings often include leukocytosis, whereas 10 to 15% of patients have leukopenia.

DIAGNOSIS

CSF examination may reveal a lymphocytic pleocytosis (<1800 cells/μL) with some increase in protein but a normal glucose concentration. Although West Nile virus may be isolated from the blood of three fourths of patients with West Nile fever on the first day of illness (Chapter 382), patients with West Nile encephalitis appear less likely to be viremic, and isolation of virus from CSF is infrequent. Viral RNA is detected in CSF by RT-PCR in about 50% of cases (Fig. 383-1). IgM antibody capture immunoassay, which is the test of choice, is more sensitive than RT-PCR. IgM antibodies may remain

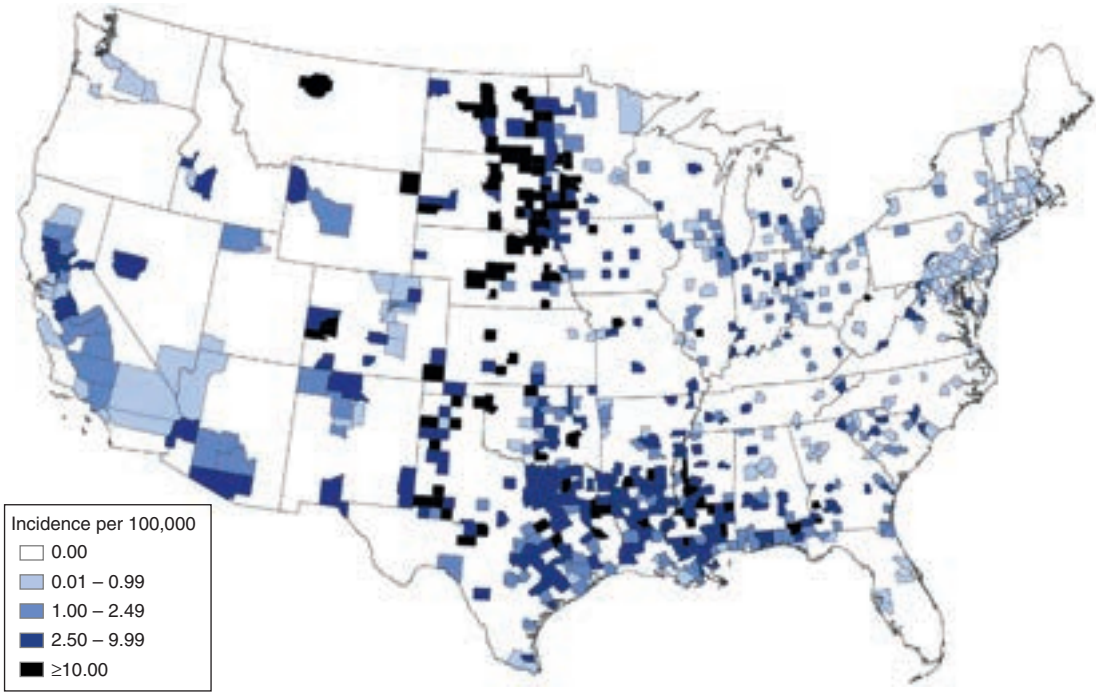

E-FIGURE 383-1. West Nile virus neuroinvasive disease incidence map presenting data reported by state and local health departments to the Centers for Disease Control and Prevention's ArboNET surveillance system. This map shows the incidence of human neuroinvasive disease (e.g., meningitis, encephalitis, or acute flaccid paralysis) by county for 2012 with shading ranging from 0.01-0.99 to 1.0-2.49, 2.50-9.99, and greater than 10.0 per 100,000 population. (From http://www.cdc.gov/westnile/statsMaps/finalMapsData/data/2012CountyIncidenceMap.pdf. Accessed March 14, 2015.)

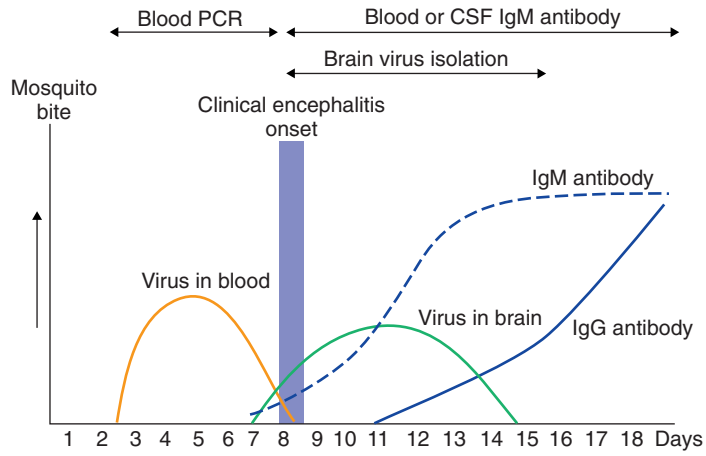

FIGURE 383-1. Typical time course for development of West Nile virus in blood and brain and occurrence of IgG and IgM antibodies in humans after West Nile virus infection. CSF = cerebrospinal fluid; PCR = polymerase chain reaction. (Reproduced from Davis LE, Beckham JD, Tyler KL. North American encephalitic arboviruses. *Neurol Clin.* 2008;26:727-757.)

detectable for up to 500 days after infection, so their presence should suggest acute infection only with compatible clinical manifestations. In some situations, cross-reactions with other flaviviruses, particularly St. Louis encephalitis virus, may complicate the interpretation of IgM immunoassays.

PREVENTION

Two phase II clinical trials of a live attenuated human chimeric vaccine (ChimeriVax-WN02) have demonstrated safety and both durable neutralizing antibody and cytotoxic T-cell responses.[12] A phase I trial of another chimeric vaccine (rWN/DEN4Δ30) also suggests promise. Because West Nile virus may be transmitted by organ transplantation, it is important to consider the possibility of this infection in organ donors.[13]

TREATMENT

Rx

Treatment is symptomatic and similar to that suggested for eastern equine encephalitis. Patients may require prolonged mechanical ventilation for either the polio-like syndrome or Guillain-Barré syndrome. The Guillain-Barré syndrome seen in West Nile patients does not appear to respond to plasma exchange. A trial of intravenous human immunoglobulin with high-titer anti–West Nile virus activity could not be completed because of the unpredictable annual geographic shifts in the activity of the disease. A recombinant humanized monoclonal antibody appeared safe in a phase I trial but has not been tested further. Ribavirin has activity against West Nile virus, but its clinical utility remains unknown.

PROGNOSIS

The overall case-fatality rate is 4 to 14%, but it is higher in the elderly. Risk factors for death include more severe weakness or coma, immunocompromise, and failure to produce virus-specific IgM. The prognosis for neurologic recovery is generally good; most U.S. patients regain physical and mental function by 1 year, but about half the patients in some series still report difficulties 1 year after the illness.

ST. LOUIS ENCEPHALITIS

ETIOLOGY

St. Louis encephalitis virus, a member of the family Flaviviridae,[14] shares close antigenic relationships with Japanese encephalitis, Murray Valley encephalitis, and West Nile viruses and is related to yellow fever (Chapter 381) and dengue (Chapter 382) viruses. Strains associated with *Culex pipiens*–borne epidemics in the northern United States are distinct from endemic strains transmitted by *Cx. tarsalis* in the western states.

EPIDEMIOLOGY

The virus is present in all parts of the Western Hemisphere, but epidemics occur only in North America and some Caribbean islands. During epidemic years, the virus has been responsible for up to 80% of all reported cases of encephalitis of known etiology in the United States. Epidemics of up to 2000 cases have taken place, mainly in urban-suburban localities of the Ohio-Mississippi River basin and in eastern and central Texas and Florida. Small outbreaks have also occurred in the western United States. Epidemics generally take place between July and September but may arise later in the year in warm areas such as Florida. Previous exposure and immunity to dengue may provide a degree of cross-protection against clinical St. Louis encephalitis. The ratio of inapparent to apparent infection is 800 : 1 in children up to 9 years of age, 400 : 1 in persons aged 10 to 49 years, and 85 : 1 in persons older than 60 years.

In most of the eastern United States, St. Louis encephalitis virus circulates between wild birds and *Cx. pipiens* mosquitoes, which breed in polluted water. In Florida and in parts of the Caribbean, *Cx. nigripalpus* is the principal vector. The cycle in the western United States also involves wild birds, but the vector is *Cx. tarsalis*, also that of western equine encephalitis. Because of the similar ecology of St. Louis encephalitis and western equine encephalitis viruses in the West, mixed outbreaks occur, mostly in rural, agricultural areas.

Above-average summer temperatures and conditions such as deficient rainfall, which create stagnant pools suitable for *Cx. pipiens* breeding, are associated with epidemics in the eastern United States. St. Louis encephalitis in the western states is favored by warm spring temperatures, heavy snow melt, and flooding.

PATHOBIOLOGY

Pathologic changes in fatal cases are limited to microscopic findings. Leptomeningitis is characterized by lymphocytic inflammation. Parenchymal changes consist of lymphocytic perivascular cuffing, cellular nodule formation, and neuronal degeneration.

CLINICAL MANIFESTATIONS

Three clinical syndromes are recognized: febrile headache, aseptic meningitis, and encephalitis. After an incubation period of 4 to 21 days, a variable period of nonspecific symptoms, including fever (temperature of 38° to 41° C), headache, malaise, drowsiness, myalgia, and sore throat, may be followed by an acute or subacute onset of meningeal or encephalitic signs, or both. Nausea, vomiting, and photophobia are common.

Neurologic abnormalities occur in up to 25% of patients. Extrapyramidal abnormalities and altered consciousness are the most significant findings. Other findings include meningismus, cranial nerve deficits (particularly the facial nerve), abnormal reflexes, tremors, myoclonic twitching, nystagmus, and ataxia. Motor abnormalities are infrequent, and sensory changes are extremely uncommon. Seizures occur in 10% of patients and are a poor prognostic sign, as is a persistent high temperature of 40° to 41° C. Signs of markedly increased intracranial pressure are unusual. A Guillain-Barré–like syndrome (Chapter 420) has occasionally been associated with St. Louis encephalitis, both as an acute manifestation and during the convalescent period.

In uncomplicated cases of St. Louis encephalitis, a moderate peripheral neutrophilic leukocytosis and shift to the left are noted. CSF pressure is elevated, protein level is mildly elevated, and glucose concentration is normal; a pleocytosis of up to 500 cells/μL is present, with an early neutrophilia predominance changing to lymphocytes within days. Serum creatine kinase, aspartate aminotransferase, and aldolase levels are frequently elevated. The electroencephalogram typically shows polymorphic delta activity, most prominently in the frontal and temporal regions; electrographic seizures are common. CT scans are normal, but MRI may show edema involving deep structures such as the substantia nigra. Hypo-osmolality, presumably as a result of the syndrome of inappropriate antidiuretic hormone secretion (Chapter 225), is noted in a third of patients.

Genitourinary tract symptoms (urgency, frequency, incontinence, and retention), microscopic hematuria, pyuria, proteinuria, and elevated blood urea nitrogen are frequent. St. Louis encephalitis viral antigen in cells of the urinary sediment has been detected by fluorescent techniques, and virus-like particles have been detected in urine by immunoelectron microscopy.

DIAGNOSIS

St. Louis encephalitis virus is rarely isolated from blood or CSF obtained during the acute phase of illness. Serologic diagnosis is achieved by demonstration of changing antibody titers; the HI, fluorescent, ELISA, and neutralizing tests demonstrate antibody within the first week after onset, and titers rise during the ensuing 2 weeks. CF antibodies appear 10 to 20 days after onset. Rapid, early diagnosis is possible by detection of IgM antibodies by ELISA in serum and CSF. Serologic cross-reactions may occur in persons

with previous exposure to dengue, West Nile, and other related flaviviruses. RT-PCR provides a more specific diagnosis, but its sensitivity is uncertain.

PREVENTION

No vaccine is available for St. Louis encephalitis. Surveillance of viral activity in vectors and avian hosts is used to define the risk for human infection and to initiate vector control efforts. In an established outbreak, avoidance of mosquito bites and spraying to reduce infected adult mosquitoes are the only effective means of control.

TREATMENT Rx

Treatment is supportive.

PROGNOSIS

A convalescent syndrome characterized by weakness, fatigue, nervousness, tremulousness, sleeplessness, irritability, depression, difficulty concentrating, and headaches occurs in 30 to 50% of older persons and clears in 80% of them within 3 years. The overall case-fatality rate is approximately 9%. Mortality is negligible in persons younger than 20 years but rises steeply after 55 years to approximately 30% in patients older than 65 years. Approximately 50% of the deaths occur during the first week, and 80% occur within 2 weeks after onset.

MURRAY VALLEY ENCEPHALITIS AND ROCIO ENCEPHALITIS

Murray Valley encephalitis and Rocio encephalitis, which are similar to Japanese encephalitis in their pathogenesis and clinical features, are caused by closely related flaviviruses. Murray Valley encephalitis has occurred in small epidemics in the Murray and Darling River valleys of Victoria and New South Wales, Australia. The virus is endemic in northern Australia and New Guinea, where it is maintained in a bird-mosquito cycle. Rocio encephalitis has caused epidemics of 1000 or more cases in São Paulo State, Brazil.

TICK-BORNE ENCEPHALITIS

PATHOGENS

A complex of six antigenically related tick-borne flaviviruses cause encephalitis: Powassan, tick-borne encephalitis, louping ill, Kyasanur Forest disease, Omsk hemorrhagic fever, and Langat viruses. The predominant syndrome is hemorrhagic fever (Chapter 381), but meningoencephalitis may be a component of the disease spectrum. Two subtypes of tick-borne encephalitis virus (central European encephalitis and Russian spring-summer encephalitis) are distinguished by serologic tests, are ecologically distinct, and differ in virulence for humans. Powassan and louping ill viruses are rare causes of encephalitis in North America and the British Isles, respectively. These viruses are easily distinguished serologically from mosquito-borne flaviviruses but induce cross-reactions within the complex.

EPIDEMIOLOGY

Tick-borne encephalitis occurs in Europe (including eastern Europe and Ukraine), southern Scandinavia, and far eastern Russia during the summer months, which corresponds to peak tick vector populations.[15] Several hundred to more than 2000 cases are reported annually, with morbidity rates of up to 20 per 100,000 inhabitants. Adults older than 20 years are mainly affected, and persons frequenting wooded areas that are heavily tick infested are at highest risk. In Europe, the disease is relatively mild (case-fatality rate of 1 to 2%), but in the Far East, it is severe (20 to 25%).

The vector of tick-borne encephalitis is *Ixodes ricinus* in Europe and *Ixodes persulcatus* in the Far East. The tick vector also serves as a reservoir for the virus. Larval ticks parasitize small rodents, which serve as amplifying viremic hosts during the spring and summer. Large vertebrates (goats, sheep, cattle) are hosts for nymphal and adult ticks. Outbreaks have occurred in families or groups of individuals ingesting unpasteurized milk or cheese from goats or sheep.

CLINICAL MANIFESTATIONS

Inapparent infections are common. Symptomatic tick-borne encephalitis in Europe typically (but not invariably) has a biphasic course beginning 7 to 14 days after exposure with an influenza-like illness that lasts 1 week, followed by a period of clinical remission for several days and then an abrupt onset of aseptic meningitis or meningoencephalitis. The meningoencephalitis is usually benign, although severe paralytic illness, myelitis, myeloradiculitis, and bulbar forms may occur.

In the Far East, tick-borne encephalitis begins suddenly with fever, headache, and gastrointestinal symptoms, followed rapidly by the appearance of depressed sensorium, coma, convulsions, and paralysis. Bulbar paralysis and cervical myelitis are frequent findings. In fatal cases, death occurs in the first week after onset. Aseptic meningitis and milder forms of encephalitis also occur. Chronic forms of tick-borne encephalitis have been described, with active clinical and pathologic abnormalities present a year or more after onset.

DIAGNOSIS

Brain MRI in patients with tick-borne encephalitis shows evidence of edema in the basal ganglia, thalami, and brain stem in about 20% of cases. MRI of the spinal cord may show anterior horn cell lesions corresponding to lower motor neuron weakness on examination.

Isolation of virus from blood is also possible during the early phase of illness. Serologic diagnosis is achieved by the HI, CF, neutralization, or ELISA techniques.

PREVENTION

In eastern Europe and the former Soviet Union, vaccines are used in high-risk groups (forestry and agricultural workers, military personnel). In Austria, immunization of the general population has resulted in a marked decline in incidence. Avoidance of tick exposure by wearing of protective clothing and use of repellents may be recommended in areas of high tick-borne encephalitis activity.

TREATMENT Rx

Treatment is supportive (see Eastern Equine Encephalitis).

PROGNOSIS

In European tick-borne encephalitis, convalescence is often prolonged, and residual paralysis may follow in severe cases. In the Far East, survivors frequently have residual paralysis, especially lower motor neuron paralysis of the upper extremities or shoulder girdle as a result of spinal cord involvement.

Louping Ill Encephalitis

Louping ill causes encephalitis in sheep (rarely in cattle, horses, and swine) in Scotland, northern England, and Ireland. Sporadic human cases have been recognized. Louping ill virus is maintained in nature by *I. ricinus* ticks and a variety of hosts, including small mammals, ground-dwelling birds (grouse), and probably sheep. The clinical features of louping ill resemble the European form of tick-borne encephalitis.

Powassan Virus Encephalitis

Powassan virus encephalitis has been documented in a small number of cases in the northeastern United States and eastern Canada. The virus is not associated with animal disease. The transmission cycle of Powassan virus involves *Ixodes cookei*, *Ixodes marxi*, and possibly other tick species along with mammals, particularly rodents and carnivores. Powassan encephalitis is characterized by fever and nonspecific symptoms, followed by encephalitic signs, which are frequently severe. Peripheral blood and CSF changes are similar to those described for other forms of flaviviral encephalitis. The case-fatality rate is about 50%, and residual paralysis may persist in survivors.

CALIFORNIA SEROGROUP ENCEPHALITIS

At least four members of the California serogroup of the Bunyaviridae family (Bunyavirus genus)—La Crosse, California encephalitis, Jamestown Canyon, and snowshoe hare viruses—cause encephalitis. California encephalitis virus occurs in the western United States (California, New Mexico, Utah, Texas) and has been implicated only rarely in human infections. In contrast, La Crosse virus, distributed more widely in the eastern half of the United States and southern Canada, is a major human pathogen. Jamestown Canyon and snowshoe hare viruses have also been implicated in sporadic cases of human encephalitis in the north central United States and Canada. California serogroup viruses have been implicated in human disease in China and the former Soviet Union.

EPIDEMIOLOGY

California serogroup encephalitis occurs as an endemic rather than an epidemic disease, with individual or small clusters of cases scattered across affected areas. Seventy to 120 cases are reported each year, generally between July and September, with a peak incidence in August. The virus primarily affects persons younger than 15 years living in rural and suburban areas characterized by deciduous hardwood forests. It is most prevalent in the north central states, where it is responsible for as many as 20% of cases of acute CNS infection in children. Focal "hot spots" (communities, even backyards) of recurrent summertime viral activity are recognized. The ratio of inapparent to apparent infection has been estimated variably at between 26 : 1 and 157 : 1.

The vector of La Crosse virus is *Aedes triseriatus*, which breeds both in forest tree holes and in artificial containers, notably discarded tires. This vector also serves as a reservoir for La Crosse virus. Wild rodents (squirrels, chipmunks) contribute to the cycle of transmission as viremic hosts. Humans acquire the disease by being bitten by an infected mosquito. *Aedes communis, Aedes stimulans, A. triseriatus,* and possibly anopheline mosquitoes are involved in transmitting Jamestown Canyon virus, and deer are the principal vertebrate hosts.

PATHOBIOLOGY

Histopathologic features in the CNS are qualitatively similar to those of other viral encephalitides. However, absence of inflammatory lesions in the cerebellum, medulla, and spinal cord may be a distinguishing feature of La Crosse infection.

CLINICAL MANIFESTATIONS

The clinical spectrum of California serogroup virus infection includes nonspecific febrile illness, aseptic meningitis, and meningoencephalitis. The disease begins with fever, headache, sore throat, and gastrointestinal symptoms. In mild cases, CNS signs appear on the third day after onset and subside within 7 to 8 days. In the more severe form, neurologic signs appear within 24 to 48 hours of onset, usually in the form of generalized seizures, elevated intracranial pressure, and altered consciousness, and persist longer. Encephalitis may be severe in the acute stage, but the disease is almost always self-limited; death is extremely rare.

The peripheral white blood cell count is elevated, with a predominance of polymorphonuclear cells and a shift to the left. CSF contains up to 500 lymphocytes/μL; the protein level is normal or mildly elevated, and the glucose concentration is normal. The electroencephalogram reveals generalized slowing in the delta and theta range; focal delta wave activity related to cortical destruction and focal seizures are also common findings.

DIAGNOSIS

In contrast to the other arboviral encephalitides, brain MRI in patients with California serogroup encephalitides may show lesions involving the temporal lobe in a pattern similar to that of herpes simplex encephalitis. The virus cannot be recovered from blood or CSF obtained during the acute phase. Diagnosis is best achieved by counterimmunoelectrophoresis, HI, CF, fluorescent, ELISA, and neutralization tests for antibody in paired acute and convalescent sera. The most practical, sensitive, and reliable methods are the HI test with La Crosse viral antigen and IgM antibody capture ELISA. Viral RNA can be detected in CSF or brain tissue by RT-PCR, although the sensitivity of the test remains to be determined.

PREVENTION

There is no vaccine for California encephalitis, although research involving DNA-based vaccines appears promising. Vector control methods are of uncertain usefulness in this disease. In defined hot spots of recurrent viral activity, breeding sites for *A. triseriatus* should be eliminated, particularly by draining or eliminating standing water (e.g., discarded tires or birdbaths) and filling holes in trees. Parents should protect children by limiting exposure and using mosquito repellents.

TREATMENT Rx

Treatment is supportive.

PROGNOSIS

The case-fatality rate is less than 1%. The risk for permanent neuropsychiatric sequelae is unclear, but hemiparesis and persistent seizure disorders have been reported.

Grade A References

A1. Feroldi E, Pancharoen C, Kosalaraksa P, et al. Single-dose, live-attenuated Japanese encephalitis vaccine in children aged 12-18 months: randomized, controlled phase 3 immunogenicity and safety trial. *Hum Vaccin Immunother*. 2012;8:929-937.

A2. Miyazaki C, Okada K, Ozaki T, et al. Phase III clinical trials comparing the immunogenicity and safety of the Vero cell–derived Japanese encephalitis vaccine Encevac with those of mouse brain–derived vaccine by using the Beijing-1 strain. *Clin Vaccine Immunol*. 2014;21:188-195.

A3. Huang LM, Lin TY, Chiu CH, et al. Concomitant administration of live attenuated Japanese encephalitis chimeric virus vaccine (JE-CV) and measles, mumps, rubella (MMR) vaccine: randomized study in toddlers in Taiwan. *Vaccine*. 2014;32:5363-5369.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Gable MS, Sheriff H, Dalmau J, et al. The frequency of autoimmune *N*-methyl-D-aspartate receptor encephalitis surpasses that of individual viral etiologies in young individuals enrolled in the California Encephalitis Project. *Clin Infect Dis*. 2012;54:899-904.
2. Carrera JP, Forrester N, Wang E, et al. Eastern equine encephalitis in Latin America. *N Engl J Med*. 2013;369:732-744.
3. Armstrong PM, Andreadis TG. Eastern equine encephalitis virus—old enemy, new threat. *N Engl J Med*. 2013;368:1670-1673.
4. Trobaugh DW, Gardner CL, Sun C, et al. RNA viruses can hijack vertebrate microRNAs to suppress innate immunity. *Nature*. 2014;506:245-248.
5. Feng Y, Fu S, Zhang H, et al. High incidence of Japanese encephalitis, southern China. *Emerg Infect Dis*. 2013;19:672-673.
6. Centers for Disease Control and Prevention (CDC). Use of Japanese encephalitis vaccine in children: recommendations of the Advisory Committee on Immunization Practices, 2013. *MMWR Morb Mortal Wkly Rep*. 2013;62:898-900.
7. Centers for Disease Control and Prevention (CDC). Japanese encephalitis surveillance and immunization—Asia and the Western Pacific, 2012. *MMWR Morb Mortal Wkly Rep*. 2013;62:658-662.
8. Petersen LR, Brault AC, Nasci RS. West Nile virus: review of the literature. *JAMA*. 2013;310:308-315.
9. Centers for Disease Control and Prevention (CDC). West Nile virus and other arboviral diseases—United States, 2012. *MMWR Morb Mortal Wkly Rep*. 2013;62:513-517.
10. Chung WM, Buseman CM, Joyner SN, et al. The 2012 West Nile encephalitis epidemic in Dallas, Texas. *JAMA*. 2013;310:297-307.
11. Hart J Jr, Tillman G, Kraut MA, et al. West Nile virus neuroinvasive disease: neurological manifestations and prospective longitudinal outcomes. *BMC Infect Dis*. 2014;14:248.
12. Dayan GH, Pugachev K, Bevilacqua J, et al. Preclinical and clinical development of a YFV 17 D-based chimeric vaccine against West Nile virus. *Viruses*. 2013;5:3048-3070.
13. Winston DJ, Vikram HR, Rabe IB, et al. Donor-derived West Nile virus infection in solid organ transplant recipients: report of four additional cases and review of clinical, diagnostic, and therapeutic features. *Transplantation*. 2014;97:881-889.
14. Turtle L, Griffiths MJ, Solomon T. Encephalitis caused by flaviviruses. *QJM*. 2012;105:219-223.
15. Beck C, Jimenez-Clavero MA, Leblond A, et al. Flaviviruses in Europe: complex circulation patterns and their consequences for the diagnosis and control of West Nile disease. *Int J Environ Res Public Health*. 2013;10:6049-6083.

384 EPIDEMIOLOGY AND DIAGNOSIS OF HUMAN IMMUNODEFICIENCY VIRUS INFECTION AND ACQUIRED IMMUNODEFICIENCY SYNDROME

THOMAS C. QUINN

GLOBAL STATISTICS

By 2014, approximately 80 million people had become infected with HIV since the beginning of the epidemic in 1981. Of these individuals, more than 45 million people had already died of AIDS, and it became ranked as one of the leading causes of death throughout the world.[1,2] According to estimates by the Joint United Nations Program on HIV/AIDS, 35 million people were living with HIV by 2013 (Fig. 384-1 and Table 384-1). In 2013 alone, 2.1 million people became newly infected, half of whom were young individuals between the ages of 15 and 24 years. The continuing rise in the population of people living with HIV infection reflects the combined effects of continued high rates of HIV infection and the beneficial impact of antiretroviral therapy (ART) resulting in fewer deaths[3] (Fig. 384-2).

The latest epidemiologic data indicate that globally, the spread of HIV appears to have peaked in 1996, when 3.5 million new HIV infections occurred.[4] In 2013, the estimated number of new HIV infections was approximately 35% lower than at the epidemic's peak about 15 years earlier. The epidemic appears to have stabilized in most regions, with new infections decreasing by 50% or more in 25 countries. Half of these reductions in the past 2 years have been among children. Despite these gains in some countries, prevalence continues to increase in Eastern Europe, Central Asia, and other parts of Asia because of the high rates of HIV infection. Sub-Saharan Africa remains the most heavily affected region and accounted for 72% of all new HIV infections in 2013.

Nearly 90% of all new infections occurred in developing countries; 50% occurred in women; and the major mode of transmission was heterosexual transmission, although infections continue to spread at high rates among men who have sex with men (MSM). Resurgence of the epidemic among MSM in high-income countries is increasingly well documented.[2] Differences are apparent in all regions, with some national epidemics continuing to expand even as the overall regional incidence of HIV infection stabilizes. Epidemic patterns have been changing over time. Perinatal transmission continues to occur in developing countries, where access to antiretroviral drugs to prevent mother-to-infant transmission is limited. In 2013, 240,000 children became newly infected, and 25 million children have been orphaned by the premature deaths of their parents from AIDS. In 2013, 1.5 million people died of AIDS, including 230,000 children.

DEMOGRAPHIC, SOCIAL, AND ECONOMIC IMPACT

For the first two decades of the epidemic, fatality rates from AIDS steadily increased and the average life expectancy in some countries in sub-Saharan Africa declined from 62 to 47 years of age. In Haiti, life expectancy was nearly 6 years less than it would have been in the absence of AIDS. Cambodia experienced a reduction in life expectancy of more than 4 years. However, the scaling up of antiretroviral therapy in low- and middle-income countries, as described later, has transformed national AIDS responses and generated broad-based health gains (Fig. 384-3). Since 1995, antiretroviral therapy has saved 14 million life-years in low- and middle-income countries, including 9 million in sub-Saharan Africa. As programmatic scale-up has continued, health gains have accelerated, with the number of life-years saved by antiretroviral therapy in Sub-Saharan Africa quadrupling in the last 4 years. Experience in the hyperendemic KwaZulu-Natal Province in South Africa illustrates the macroeconomic and household livelihood benefits of expanded treatment access, with employment prospects sharply increasing among individuals receiving antiretroviral therapy.[5,6]

THE GLOBAL RESPONSE

At a 2001 special session of the United Nations General Assembly on AIDS, 189 nations agreed that AIDS was a national and international security issue of the highest priority. The Global Fund for AIDS, Tuberculosis, and Malaria raised funds from private donations and industrialized countries to help support access to care and treatment in developing countries. This initiative was complemented by the U.S. government's Emergency Plan for AIDS Relief, now committed over a 10-year period through the Bush and Obama administrations to provide treatment and care for HIV-infected individuals, as well as additional resources for enhancing prevention efforts to prevent further transmission. As of 2014, approximately 13.6 million people in low- and middle-income countries were receiving antiretroviral therapy (ART)—a more than 25-fold increase since 2003 (see Fig. 384-3). The rapid expansion of antiretroviral therapy is one of the most remarkable achievements in recent public health history. ART coverage rose from 7% in 2003 to 54% in 2012 for individuals with CD4 counts less than 350 cells/mm^3, with especially high coverage achieved in Latin America (68%), the Caribbean (67%), Oceania (69%), and Eastern and Southern Africa (56%). Sixty-three percent of all people now receiving treatment in low- and middle-income countries are living in Sub-Saharan Africa, compared with 25% in late 2003. Coverage remains low in Eastern Europe and Central Asia (25%) and in the Middle East and North Africa (15%).

Unfortunately, not all have equal access to therapy, even within countries with middle to high prevalence. For example, an estimated 800,000 children younger than 15 years now require ART, but only 200,000 are receiving treatment. Children account for approximately 14% of AIDS deaths. Nearly 90% of children infected with HIV are African. The median proportion of HIV-infected children receiving treatment is just 28% in Sub-Saharan Africa. Similarly, less than 40% of HIV-infected pregnant women in low- and middle-income countries are benefiting from antiretroviral prophylaxis despite success in certain countries, such as Botswana, Brazil, and Thailand, and the virtual elimination of pediatric HIV disease in the industrialized world. Likewise, in Eastern Europe and Central Asia, injection drug use accounts for more than 70% of HIV-positive persons. However, only 25% of treatment recipients in this region are injection drug users. In years to come, additional resources will be required to reach the additional millions of HIV-infected people who require treatment with antiretroviral drugs, especially in light of recent recommendations to expand access to all HIV-infected adults who have a CD4+ count of less than 500/mm^3 based on the finding that early initiation of ART reduces mortality and incident tuberculosis.

REGIONAL EPIDEMICS

Sub-Saharan Africa

Sub-Saharan Africa represents the epicenter of the global HIV/AIDS pandemic (see Table 384-1). Studies in the late 1990s and early 2000s support the theory that HIV originated in Africa and that humans probably became infected sometime in the mid-20th century from a similar, related retrovirus in chimpanzees and sooty mangabey monkeys. For years, the infection remained limited to remote rural regions of Africa, but with urbanization, infected individuals migrated to major urban centers, where transmission was amplified and HIV spread to thousands of individuals within a relatively short period.

Seventy-five percent of all women infected with HIV live in Sub-Saharan Africa. In Côte d'Ivoire, home to the most serious epidemic in West Africa, HIV prevalence in females (6.4%) was more than twice as high than in males (2.9%). In Sub-Saharan Africa as a whole, women account for 60% of the estimated HIV infections. The risk for becoming infected is disproportionate for girls and young women. In Kenya, young women between 15 and 19 are three times more likely to be infected than their male counterparts, whereas 20- to 24-year-old women are 5.5 times more likely than men in their age cohort to be living with HIV infection. Among people aged 15 to 24 in Tanzania, females are four times more likely than males to be living with HIV. In the nine countries of Southern Africa most affected by HIV, the prevalence in young women 15 to 24 was on average three times higher than that in men of the same age.

Women's vulnerability to HIV in Sub-Saharan Africa stems not only from their greater physiologic susceptibility to heterosexual transmission but also from the severe social, legal, and economic disadvantages they often confront. HIV prevalence generally tends to peak at a younger age for women than for men. The very young are often at extremely high risk for infection through mother-to-child transmission.

HISTORICAL PERSPECTIVE

Three decades after initial recognition of acquired immunodeficiency syndrome (AIDS) in the United States, the disease has become epidemic in every country of the world. Initially reported as a disease primarily affecting homosexual men, AIDS was rapidly identified in many other risk groups, and it became evident that it was caused by an infectious agent transmitted through sexual activities, parentally through blood transfusions and injection drug use, and perinatally from mother to infant. Early investigations in the 1980s demonstrated that the etiologic agent of AIDS was the human immunodeficiency virus (HIV), which existed in two types, HIV-1 and HIV-2. After the development of diagnostic assays for HIV antibody, it became possible to track and monitor the escalating spread of HIV throughout the world, definitively define the modes and probabilities of transmission, and study the natural history of HIV infection. Within years of its recognition, HIV disseminated rapidly throughout the world, caused a massive epidemic, and became one of the leading causes of death worldwide. Efforts at prevention and treatment with antiretroviral drugs have tempered the spread and decreased the fatality rate in some countries, but in developing countries where the social, demographic, cultural, and economic impact of the AIDS epidemic has been the greatest, these gains were initially too limited and too slow to reverse the escalating trend of the epidemic. More recently, with the increased availability of antiretroviral drugs in many countries, the annual rate of increase in HIV infections may have peaked and mortality has declined.

From these epidemiologic trends it is projected that within the next decade, an additional 25 million people could become infected with HIV, unless the world succeeds in mounting a drastically expanded global prevention and treatment effort. It is anticipated that more than 40% of these new infections will occur in Asia, the Pacific, and Eastern Europe, although it is apparent that the epidemic will continue to devastate nearly all countries of the African continent. During the next decade, without full access to treatment and care in many countries, millions of individuals will join the ranks of the more than 45 million people who have already died of AIDS.

Globally an estimated 0.8% of adults 15 to 49 years of age are living with HIV, although the burden of the epidemic continues to vary considerably between countries and regions. AIDS is the leading cause of death in sub-Saharan Africa and the fourth largest killer worldwide. Sub-Saharan Africa remains the most severely affected region, with nearly 1 in every 20 adults (4.9%) living with HIV and accounting for more than two thirds of people living with HIV worldwide. Although the regional prevalence of HIV infection is nearly 25 times higher in Sub-Saharan Africa than in Asia, almost 5 million people are living with HIV in South, Southeast, and East Asia combined. After sub-Saharan Africa, the regions most heavily affected are the Caribbean, and Eastern Europe; and Central Asia, where 1.0% of adults were living with HIV in 2013.

Effective prevention of mother-to-child transmission depends on the simultaneous support of several strategies that work synergistically to reduce the odds that an infant will become infected as a result of exposure to the mother's virus. In ideal conditions, provision of antiretroviral prophylaxis and replacement feeding or continued antiretroviral treatment with breastfeeding can reduce transmission from an estimated 30 to 35% with no intervention to 1 to 2%. Unfortunately, most countries have not yet reached all pregnant women with these services, with estimates of only 30 to 60% of HIV-infected mothers receiving appropriate HIV care and prophylaxis for mother-to-child transmission (globally antiretroviral coverage for pregnant women reached 57% in 2013). Despite these limitations, it has been estimated that nearly 700,000 new HIV infections have been averted in the past decade because of provision of antiretroviral prophylaxis during pregnancy and to newborn infants.

Most HIV transmissions in Sub-Saharan Africa occur through sexual intercourse, with unsafe blood transfusions and unsafe injections accounting for a smaller fraction. Although sexual behavior is the most important factor influencing the spread of HIV in Africa, that behavior varies greatly across cultures, age groups, socioeconomic class, and gender. The interplay of multiple factors, biologic and behavioral, determines the spread of HIV. One study of four African cities (Cotonou, Kisumu, Ndola, and Yaounde) revealed that the most common behavioral and biologic factors in those cities with the highest HIV prevalence were young age at a woman's first sexual intercourse, young age at first marriage, age difference between spouses, the presence of herpes simplex type 2 infection and trichomoniasis, and lack of male circumcision. There is substantial evidence that sexually transmitted diseases (STDs) enhance the risk for sexual transmission of HIV, that the level of the HIV viral load in an individual enhances the probability of infectiousness, and that male circumcision is associated with a reduced risk for transmission. Although the complex interplay of factors makes it difficult to estimate the probable growth of the epidemic within the region, evidence from the past 30 years demonstrates that HIV can spread rapidly and widely from very low prevalence levels.

A little more than a 10th of the world's population lives in Sub-Saharan Africa, but it is home to 69% of all people infected with HIV in the world. In 2013, an estimated 1.5 million people living in Sub-Saharan Africa became newly infected with HIV, thus bringing the total number of people infected with HIV to 23.5 million. Sub-Saharan Africa accounted for 71% of all the world's AIDS-related deaths in 2013. Even though the rate of new HIV infections in sub-Saharan Africa has slowly declined, with the number of new infections in 2013 being approximately 25% lower than at the epidemic's peak in the region in 1995, the number of people infected with HIV has slightly increased in 2013, in part because of the increased longevity stemming from improved access to treatment.

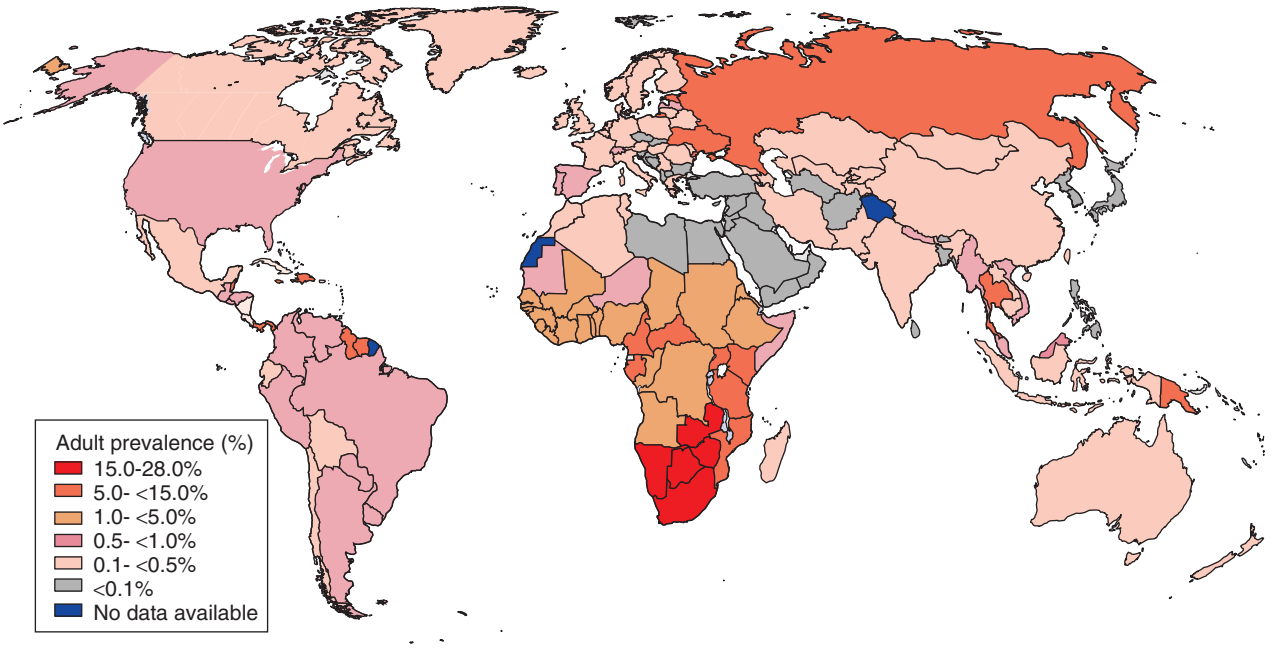

FIGURE 384-1. A total of 35.0 million people (33.2 to 37.2 million) were living with HIV infection in 2013. The adult prevalence of HIV infection is shown by country. (Data from UNAIDS GAP Report. UNAIDS; 2014.)

TABLE 384-1 REGIONAL HUMAN IMMUNODEFICIENCY VIRUS AND ACQUIRED IMMUNODEFICIENCY SYNDROME STATISTICS AND FEATURES AT THE END OF 2013

REGION	ADULTS AND CHILDREN LIVING WITH HIV/AIDS	ADULTS AND CHILDREN NEWLY INFECTED WITH HIV	ADULT PREVALENCE RATE* (%)	MAIN MODES OF TRANSMISSION FOR ADULTS LIVING WITH HIV/AIDS
Sub-Saharan Africa	24,700,000	1,500,000	4.9	Hetero
North Africa and Middle East	230,000	25,000	0.2	Hetero, IDU
Asia and the Pacific	4,800,000	350,000	0.3	Hetero, IDU
Latin America	1,600,000	94,000	0.4	MSM, IDU, Hetero
Caribbean	250,000	12,000	1.0	Hetero, MSM
Eastern Europe and Central Asia	1,100,000	110,000	1.0	IDU, Hetero, MSM
Western Europe	1,100,000	30,000	0.2	MSM, IDU, Hetero
North America	1,300,000	55,000	0.6	MSM, IDU, Hetero
Total	35,000,000	2,100,000	0.8	

*The proportion of adults (15 to 49 years of age) living with HIV infection or AIDS in 2013 using 2013 population numbers.
AIDS = acquired immunodeficiency syndrome; Hetero = heterosexual transmission; HIV = human immunodeficiency virus; IDU = transmission through injection drug use; MSM = sexual transmission among men who have sex with men.

Unfortunately, progress has been strikingly uneven in gaining access to ART in Sub-Saharan Africa, with coverage reaching or exceeding 50% in some countries (Botswana, Namibia, and Uganda) but remaining below 20% in most others.

Although some countries in Sub-Saharan Africa such as Kenya, Uganda, and Zimbabwe have shown recent declines in HIV prevalence, there is no evidence of any decline in Southern Africa, including the Republic of South Africa, Botswana, Namibia, and Swaziland, where exceptionally high infection levels continue. Southern Africa remains the area most heavily affected by the epidemic. The nine countries with the highest HIV prevalence worldwide are all located in this subregion, with each of these countries experiencing adult HIV prevalence greater than 10%; prevalence was as high as 26% in Swaziland, 24% in Botswana, and 23% in Lesotho. The Republic of South Africa is home to the world's largest population of people living with HIV (6.3 million), with a prevalence of 17.3%. Almost one in three pregnant women attending public antenatal clinics were infected with HIV. South Africa accounts for a large percentage of the treatment scale-up in sub-Saharan Africa, but as of 2013, antiretroviral therapy was only reaching 55% of South Africans eligible for treatment.

In Botswana, more than a third of pregnant women attending antenatal clinics and close to 50% of women 30 to 34 years of age were infected with HIV in 2013. Similarly, Lesotho has a national adult HIV prevalence of 23%, with 27% documented in women attending antenatal clinics. A third of pregnant women 25 to 34 years of age were infected. In Namibia, the prevalence of HIV infection is 13.4% in all adults, with an HIV prevalence of 42% in antenatal clinics in selected areas. In neighboring Mozambique, Malawi, and Zambia, HIV prevalence has been documented to be between 10% and 14%. There is wide geographic variation, however, with HIV infection rates in pregnant women ranging from less than 10% in some places to as high as 30% in others.

In the countries of Eastern Africa, HIV prevalence has either decreased or remained stable in the past several years. In Uganda, which saw a steep decline in HIV prevalence during the mid and late 1990s, adult HIV prevalence was estimated to be 6.7% in 2005. However, recent trends suggest that HIV prevalence may be increasing in selected areas (nationally at 7.2%), in part because of a decrease in mortality with access to ART but potentially also because of increasing incident rates as a result of decreasing condom use and an increased percentage of multiple partners. In neighboring Kenya, Eritrea, Tanzania,

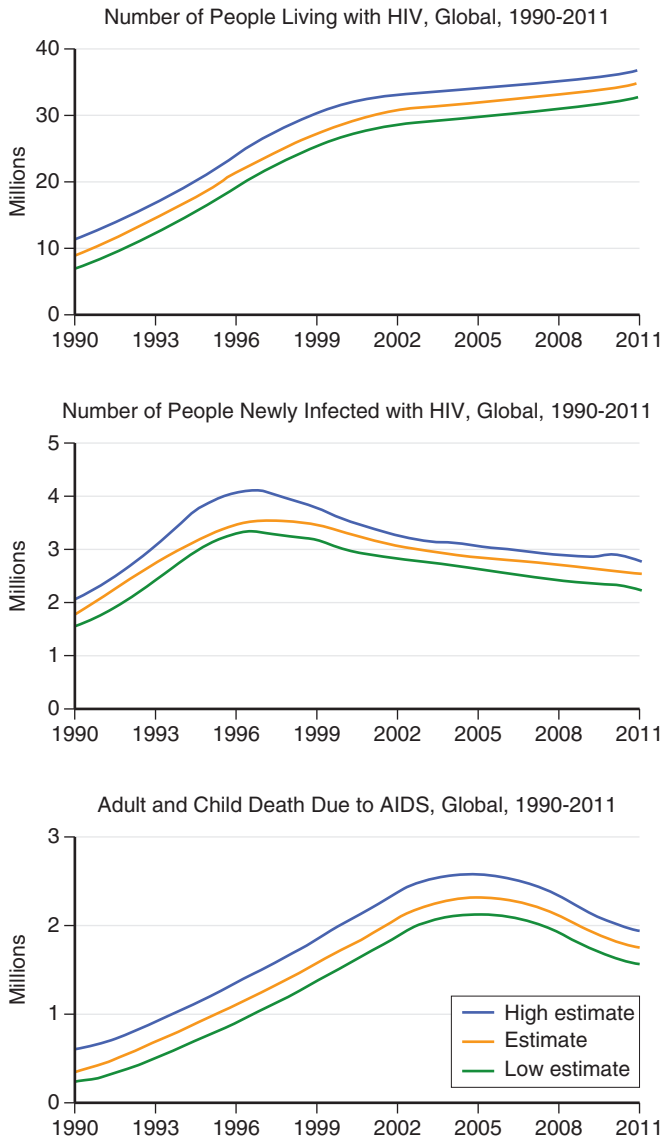

FIGURE 384-2. Global HIV trends in number of people living with HIV, new infections and fatalities, from 1990 to 2011. (Data from UNAIDS Global Report. *UNAIDS Report on the Global AIDS Epidemic 2012.* Geneva: UNAIDS; 2013.)

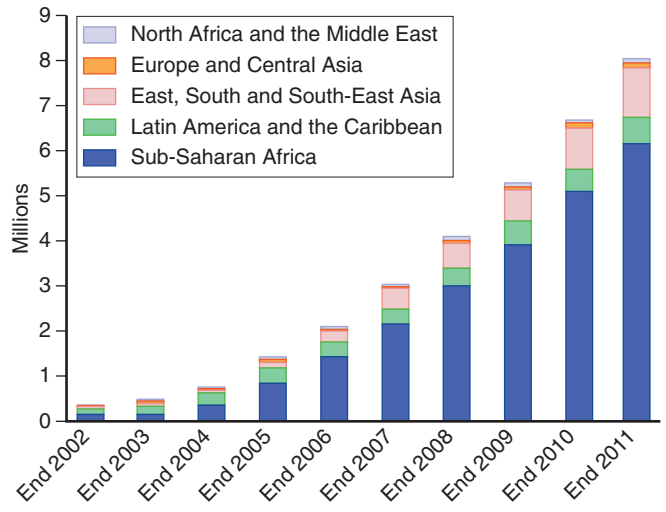

FIGURE 384-3. Estimated number of people receiving antiretroviral therapy in low- and middle-income countries from 2002 through 2011. (Data from UNAIDS Global Report. *UNAIDS Report on the Global AIDS Epidemic 2012.* Geneva: UNAIDS; 2013.)

Burundi, and Rwanda, the HIV epidemic has been stable in recent years, with prevalence rates ranging from 1% in Eritrea to 6% in Kenya, Tanzania, Burundi, and Rwanda. In Ethiopia, HIV prevalence (1.4%) has stabilized in urban areas but appears to be increasing in more distant rural areas, where access to treatment and care is more limited.

Western Africa is less severely affected than other parts of Sub-Saharan Africa, with national adult HIV prevalence rates of 2% in several countries. The highest rate in the region is in Côte d'Ivoire, at 3.0%. In Nigeria, infection levels vary radically across the country from 2.6% in the southwest to 6.1% in the north central zones (nationally 3.7%). HIV continues to spread rapidly among female sex workers and their clients, as well as in the general population. In some urban populations, more than 10% of adults are infected and the annual incidence is as high as 3%.

Despite the scale-up of treatment, AIDS is still one of the leading causes of death and years of productive life lost throughout the continent. Excess deaths attributable to HIV are highest in the 25- to 34-year-old group, usually a group with low mortality. Nearly 90% of deaths in this age group are in excess of background rates and were attributable to HIV. Because AIDS deaths are concentrated in childhood and young adult age groups, their effects are substantial, with life expectancy reduced markedly in several countries. HIV or AIDS cases will put an increasing strain on health care systems, which are already overburdened, and on individual households that are trying to manage with limited economic resources. Care and support for children orphaned by AIDS will become a growing concern throughout the region.

Asia and the Pacific

After Sub-Saharan Africa, Asia and the Pacific, home to 60% of the world's population, have the second largest number of HIV-infected individuals in the world, estimated at 4.8 million. In 2013, 350,000 adults and children became newly infected, in part because of growth of the HIV epidemic in China, India, and several other countries in Southeast Asia. With the exception of Thailand, national HIV prevalence levels remain comparatively low in most countries of Asia and the Pacific, with HIV prevalence being less than 1%, to some extent due to their large population base. Thus, because of the region's large population, Asia's comparatively low HIV prevalence translates into a substantial portion of the global HIV burden. India's national adult HIV prevalence rate of 0.3% offers little indication of the serious situation facing the country. An estimated 4.8 million people were living with HIV at the end of 2013—one of the highest figures in the world after South Africa.

Throughout the region, injection drug use remains one of the most prominent modes of transmission of HIV. More than 50% of injection drug users have already acquired HIV in Malaysia, Myanmar, Nepal, Thailand, Indonesia, Manipur, and Southern China. Very high rates of needle sharing have been documented among users in Bangladesh and Vietnam, along with evidence that a considerable proportion of sex workers in Vietnam also inject drugs. More recently, the epidemic in many parts of Asia is steadily expanding into lower risk populations through transmission to the sexual partners of those most at risk. In China, where the epidemic was previously driven by transmission through injection drug use, heterosexual transmission has become the predominant mode of HIV transmission.

China, with a fifth of the world's population, has also witnessed a dramatic escalation of the HIV epidemic in the past decade. A total of 780,000 Chinese individuals are estimated to be living with HIV. The HIV epidemic is particularly severe among injection drug users, who account for a quarter of all HIV infections. To compound the tragedy of the epidemic in China, reports from Henan province in Central China demonstrate that tens of thousands and possibly more rural villagers became infected by selling their blood to collecting centers that did not follow basic blood donation safety procedures. It has been estimated that 150,000 people have been infected through these practices. There are new signs of heterosexually transmitted HIV epidemics in at least three provinces—Guangdong, Guangxi, and Yunnan. Several other factors highlight the swift escalation of HIV infection in China. STDs quadrupled between 1997 and 2002, thus suggesting that unprotected sex with non-monogamous partners is increasing in China. There is massive population mobility. Approximately 100 million Chinese are temporarily or permanently away from their registered addresses, and increasing socioeconomic disparities add to the likelihood of the spread of HIV.

Indonesia, the world's fourth most populous country, is another example of how quickly the AIDS epidemic can emerge. After more than 10 years of negligible HIV prevalence, the infection rate in injection drug users, sex workers, and blood donors in some regions is rapidly increasing, with a 25% increase between 2001 and 2013. Papua New Guinea also has reported the

As of 2013, 48% of adults and children in need of ART in the region were estimated to be receiving such services. ART coverage is notably higher in eastern and southern Africa (56%) than in western and central Africa (35%). Treatment coverage for adults (50%) remains higher than that for children (38%). Children's access to ART is especially limited in Western and Central Africa.

AIDS-related deaths have fallen by 29% in Kenya, and a study in Uganda found that timely initiation of ART and cotrimoxazole prophylaxis reduced mortality by 95% and produced a 93% reduction in HIV-related orphans. In Botswana, where ART coverage exceeds 80%, the estimated annual number of AIDS-related deaths declined by more than half. More than 50% of all people in need of treatment are still not receiving such services.

highest HIV infection rates among the Pacific Island countries and territories. Even though the Philippines has maintained a low HIV prevalence, higher rates of STDs among Filipino sex workers, their clients, and MSM indicate low levels of condom use and the potential for rapid spread of HIV.

In some countries of Southeast Asia where HIV prevalence rose rapidly in the 1990s, strong prevention programs have limited the spread, most notably in Thailand, Cambodia, and Myanmar. Furthermore, the increased access to ART has coincided with a drastic drop in AIDS-related deaths. Despite these advances, AIDS is still a leading cause of death in Thailand and 1.2% of the country's population is infected with HIV. Although STDs and heterosexual transmission have declined as a result of the government's prevention programs, HIV continues to spread rapidly among injection drug users and MSM.

Treatment scale-up in Asia has been mixed. As of December 2013, 47% of those in South and Southeast Asia needing ART were receiving it, but only 18% were receiving it in East Asia, much lower than the global average (54%) for all low- and middle-income countries. In 2013, only one country in the Asia Pacific region (Cambodia) reached more than 80% coverage of antiretroviral therapy, whereas in Pakistan less than 20% of eligible infected persons received therapy. In Oceania, an estimated 69% received therapy.

Eastern Europe and Central Asia

The HIV epidemic has increased faster in Eastern Europe and Central Asia between 2000 and 2013 than in any other area of the world. In 2013 there were an estimated 1.1 million people living with HIV, a 20-fold increase in less than a decade. In recent years, the Russian Federation has experienced an exceptionally steep rise in reported HIV infections, 90% of which have been attributed to injection drug use. It is estimated that nearly 1% of the young people in Eastern Europe and Central Asia are injecting drugs, which places these individuals and their sex partners at high risk for becoming infected with HIV. In countries such as Azerbaijan, Georgia, Tajikistan, and Uzbekistan, HIV has experienced explosive growth. Similar explosive high rates of HIV are being documented in injection drug users and heterosexuals at risk for STDs in other countries of the Commonwealth of Independent States, in the Baltic States, and in Romania.

In Estonia, Latvia, and Lithuania, major HIV outbreaks are also occurring in selected populations, such as prison inmates. In one prison in Lithuania, 15% of the inmates were HIV positive, thus confirming the role of prisons in the spread of HIV in many countries of the region. The concentration of large numbers of young people in overcrowded prisons or juvenile justice facilities, often marked by an abundance of drugs but a scarcity of HIV information, clean needles, or condoms, provides fertile ground for the rapid spread of HIV among inmates and, on their eventual release, into the wider population.

Initially driven by injection drug use in young people, heterosexual transmission of HIV has become a prominent mode of spread in Belarus and Ukraine. With an estimated adult HIV prevalence rate of 0.8%, Ukraine is one of the most severely affected countries in the region. Three fourths of HIV infections in Ukraine are related to injection drug use, with a prevalence of 21.5% in injection drug users, and the proportion of sexually transmitted infection (STIs) is increasing, suggesting potential spread heterosexually and among MSM. Although many of these infections may occur in sex partners of injection drug users, the trend also may indicate spread into the wider population of these countries. In the Russian Federation and the Ukraine, up to 30% of female injection drug users are also involved in commercial sex work. In Odessa, 67% of sex workers who inject drugs were HIV positive. The public health efforts to stem the tide of the epidemic in these countries are limited and, in some cases, nonexistent. In contrast, HIV prevalence remains low in Poland, the Czech Republic, Hungary, and Slovenia, where well-designed national HIV/AIDS programs are in operation.

A number of countries in the region have expanded access to ART, although treatment coverage remains relatively low. By December 2013, only 25% of adults in need of therapy were receiving it—a level much less than the global average. Injection drug users, the population most at risk for HIV in Eastern Europe and Central Asia, are often least likely to receive ART when they are medically eligible. If effective interventions are not implemented in the more severely affected countries, it is likely that the situation will become dramatically worse over the next 5 years.

Latin America and the Caribbean

An estimated 1.6 million adults and children are living with HIV in Latin America and the Caribbean. Twelve countries in this region have an estimated HIV prevalence of 1% or greater in pregnant women. In several Caribbean countries, adult HIV prevalence rates are surpassed only by the rates experienced in sub-Saharan Africa, which makes this region the second most affected in the world. Haiti and the Bahamas remain the worst affected, with an estimated national prevalence higher than 1.8% in Haiti and a prevalence of 2.8% in the Bahamas. AIDS is the leading cause of death in some countries of the Caribbean basin. In Haiti, the Bahamas, and Guyana, the number of deaths in 15- to 34-year-olds is 2.5 times higher than it would have been in the absence of AIDS.

Homosexual and heterosexual transmission continues to be the major mode of transmission throughout the region, although there is evidence that spread of HIV is increasing through sharing of infected drug equipment. Population mobility, spurred by high rates of unemployment and poverty, is emerging as a significant factor in the epidemic's growth in this region. Central America's geographic position also makes it an important transit zone for people moving between the rest of the region and North American countries. Appropriately, protecting vulnerable populations on the move, including adolescent girls and young women, is now the focus of a regional prevention program in Central America. In Mexico, adult HIV prevalence in the wider population is still well under 1%, but prevalence rates are higher in specific population groups—6% in injection drug users and 15% in MSM. There is significant overlap between injection drug users and MSM, especially in Brazil and the southern Latin American countries, where injection drug use is a growing social phenomenon. Injection drug use is also a major route of HIV transmission in Argentina, Chile, and Uruguay.

Despite many constraints, the region has made progress in the provision of treatment and care. By reducing HIV-related morbidity through treatment, Brazil's treatment and care program is estimated to have avoided 234,000 hospitalizations in a 4-year period, thereby demonstrating a cost-effective approach to care. Argentina, Costa Rica, Uruguay, and Cuba now guarantee free and universal access to drugs through the public sector, and sharp reductions have recently been secured in Honduras and Panama. Treatment coverage has risen from 10% of those in need in 2004 to 68% in 2012. Similarly, pediatric antiretroviral coverage in the Caribbean was high (>60%) in comparison to the global average for children of 28%. Consistent with the evolving guidelines of HIV treatment, a growing number of people living with HIV in Latin America are starting treatment earlier at CD4+ cell counts of 350 or lower rather than waiting until the count drops below 200. Earlier initiation of therapy offers the possibility that medical outcomes in the region may improve further still and reduce the population-level viral load, which might result in lower transmission rates.

Western Europe

More than 1.1 million HIV-infected individuals reside in Western Europe, with trends similar to those witnessed in the United States, Australia, and New Zealand. Longer survival of people infected with HIV has led to a steady increase in the number of people living with the virus in high-income countries. In a multicountry study in Europe, Australia, and Canada, mortality rates in people living with HIV now approach those in the non–HIV-infected population, although excess mortality in HIV-infected people increases with the duration of infection. In the United Kingdom, half of all people living with HIV are receiving ART, with no appreciable increase in the number of patients with virologic failure or resistance to the drugs.

The HIV epidemic in Western Europe is a result of a multitude of epidemics that differ in their timing, scale, and effects on populations. A larger proportion of new HIV diagnoses in Western European countries occur through heterosexual intercourse. More than half of the new HIV infections in the United Kingdom in 2013 resulted from heterosexual sex, compared with 33% in 1998. In Ireland, a similar trend is visible, with numbers of heterosexually transmitted HIV infection increasing four-fold between 1998 and 2001. Unsafe sex between men remains an important factor for spread in most European countries, particularly in the United Kingdom, Germany, the Netherlands, and Spain. Injection drug use remains a major mode of transmission in Spain, France, and Portugal, but like in other countries in Europe, approximately a fourth of all HIV infections are now heterosexually transmitted. Most data from high-income countries demonstrate that the epidemic has shifted to the poor and marginalized sections of society. Underscoring the need for renewed prevention efforts, especially in young people, are findings of increases in high-risk behavior, less frequent condom use, and higher rates of STDs in several countries. In the United Kingdom, for example, rates of gonorrhea, syphilis, and chlamydial infections have more than doubled since 1995, and increases have been found in other Western European countries as well.

The United States

By 2013, more than 1.2 million people were living in the United States with HIV (see Table 384-1). Nationally, the adult HIV prevalence was estimated to be 0.6%. This increase reflects mixed results in the United States' effort to combat its epidemic. More people infected with HIV are living longer because of antiretroviral therapy, but unfortunately, the early gains made in prevention have not been sustained (Fig. 384-4). The number of newly recorded HIV cases in 46 states with confidential name–based reporting has varied only slightly since the late 1990s. In 2006, the annual incidence of HIV infection was estimated at 56,300, which was approximately 40% higher than previously estimated. The CDC estimates that approximately 50,000 people in the United States are newly infected with HIV each year. In 2011, there were an estimated 50,007 new HIV infections. Nearly two thirds of these new infections occurred in MSM; 18% were women who acquired HIV via heterosexual contact, 10% were men who also acquired HIV by heterosexual contact (Figs. 384-5 and 384-6), and 11% acquired HIV via injection drug use. The estimated rate of diagnosis of HIV infection for the United States was 16.1 in 100,000, with 31.4 in 100,000 in men and 8.0 in 100,000 in women.[7] In the United States, more than 620,000 people with AIDS have died since the epidemic began.

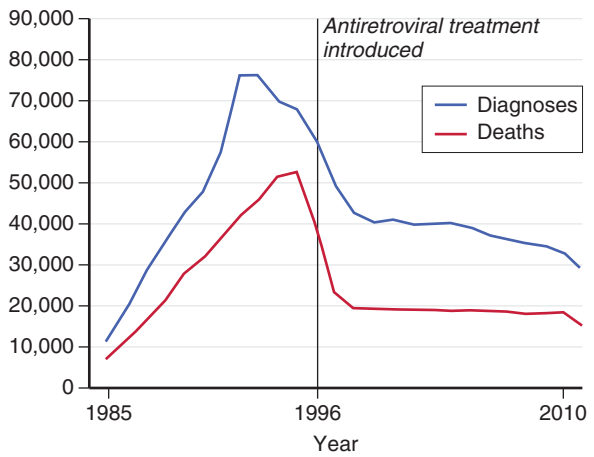

FIGURE 384-4. AIDS diagnosis and deaths in the United States from 1985 to 2010. (Data from Centers for Disease Control and Prevention. HIV Surveillance—United States, 1981-2008. *MMWR.* 2011;60:689-693.)

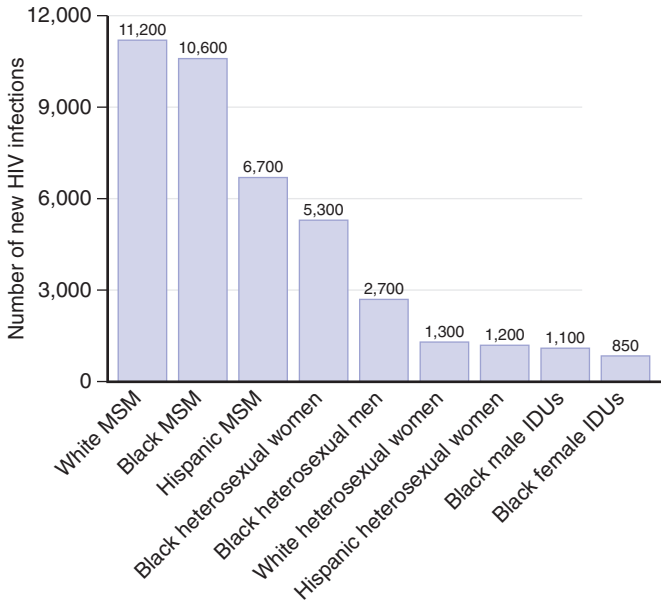

FIGURE 384-5. Estimated new HIV infections in the United States, 2010, for the most-affected subpopulations. IDUs = injection drug users; MSM = men who have sex with men. (Data from Centers for Disease Control and Prevention. Estimated HIV incidence in the United States, 2007-2010. *HIV Surveillance Suppl Rep.* 2012;17:4.)

More women are being infected with HIV through unprotected heterosexual exposure (86%) and injection drug use (14%).[8] The main risk factor for women who acquire HIV during sex is the risk behavior of their male partners, such as injection drug use, commercial sex, or sex with other men. As in Latin America, women living in impoverished and marginal circumstances appear to be at disproportionate risk for HIV infection. In North Carolina, HIV-positive women were considerably more likely to be unemployed, requiring public assistance, and exchanging sex for money and gifts.

As the U.S. epidemic evolves, it is becoming more an epidemic of African Americans and other minorities (see Fig. 384-5). African Americans make up 12% of the U.S. population but account for 47% of new HIV diagnoses. Hispanics represent 17% of the population but account for 21% of new HIV diagnoses. Among African Americans and Hispanics, most men infected with HIV were exposed during sex with other men (72% and 79%, respectively), whereas most women with HIV become infected heterosexually (89% and 86%, respectively). African American women are more than 10 times more likely to be infected with HIV than are white women. AIDS continues to be one of the leading causes of death in African American women aged 25 to 34 years and ranks in the top three causes of death in African American men aged 25 to 54 years. In the United States, the challenge of slowing the rate of new HIV infections overlaps with the need to provide diagnosis, treatment, and care services more equitably. The number of AIDS-related deaths was 69% lower than in 1994. However, further declines in mortality will require greater success in encouraging timely diagnosis of HIV infection. An estimated 21% of people living with HIV are unaware of their HIV status[9] (Fig. 384-7). Moreover, in 36% of people in whom HIV was diagnosed, AIDS was diagnosed within 12 months. With the aim of increasing the percentage of people who receive a timely diagnosis of HIV, the Centers for Disease Control and Prevention recommends routine voluntary HIV testing in all health care settings unless the patient "opts out," to not be tested.

With an epidemic that is nearly into its third decade, complacency has increased and prevention efforts have dwindled as a result of declining mortality. Multiple studies illustrate that prevention efforts are not reaching the large number of at-risk individuals who engage in unsafe sex. Increased rates of STIs among MSM have been documented in the United States, Australia, Great Britain, Canada, and other developed countries. Rates of gonorrhea, syphilis, and chlamydia have more than doubled in the past 5 years among MSM in selected U.S. and European cities. Renewed efforts to enhance prevention efforts, particularly in HIV care clinics, are being echoed throughout all these countries.

DIAGNOSIS

Screening for HIV Infection

The history and physical examination are of limited value in making the diagnosis of early HIV infection,[10] so laboratory testing is key to making the diagnosis. The U.S. Preventive Services Task Force on HIV screening recommends that clinicians screen for HIV infection in all adolescents and adults aged 15 to 65 years.[11] Younger adults and older adults who are at increased risk also should be screened.[12] In addition, they recommended that clinicians screen all pregnant women for HIV, including those who present in labor who are untested and whose HIV status is unknown. These updated recommendations were based on increasing evidence of the benefits of early antiretroviral therapy for HIV-infected persons and its effectiveness in preventing HIV transmission. This recommendation for increased screening was based on

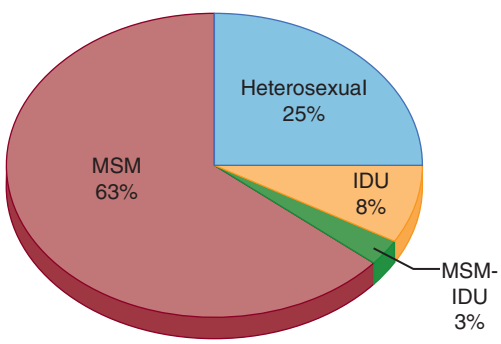

FIGURE 384-6. Estimated new HIV infections in 2010 in the United States by transmission category. IDUs = injection drug users; MSM = men who have sex with men. (Data from Centers for Disease Control and Prevention. Estimated HIV incidence in the United States, 2007-2010. *HIV Surveillance Suppl Rep.* 2012;17:4.)

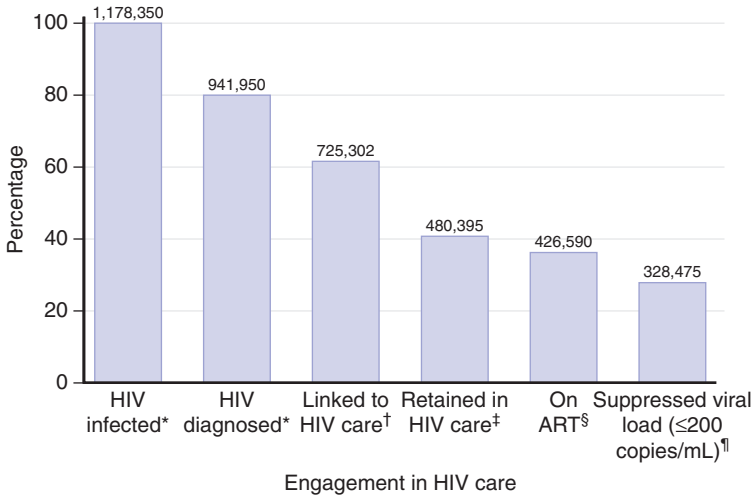

FIGURE 384-7. Number and percentage of HIV-infected persons engaged in selected stages of the continuum of HIV care in the United States. HIV = human immunodeficiency virus; ART = antiretroviral therapy. (Data from Centers for Disease Control and Prevention. Vital signs: HIV prevention through care and treatment—United States. *MMWR.* 2011;60:1618-1623.)
*HIV-infected, n = 1,178,350; HIV-diagnosed, n = 941,950.
†Calculated as estimated number diagnosed (941,950) × estimated percentage linked to care (77%); n = 725,302
‡Calculated as estimated number diagnosed (941,950) × estimated percentage retained in care (51%); n = 480,395.
§Calculated as estimated number retained in care (480,395) × percentage prescribed ART in the Medical Monitoring Project (MMP) (88.8%); n = 426,590.
¶Calculated as estimated number on ART (426,590) × percentage with suppressed viral load in MMP (77.0%); n = 328,475 (28% of the estimated 1,178,350 persons in the United States who are infected with HIV).

the fact that over 20% of HIV-infected individuals have never been tested and are unaware of their infection, and based on evidence that identification and treatment of HIV infection are associated with a markedly reduced risk for progression to AIDS, AIDS-related events, and deaths in individuals with immunologically advanced disease. One randomized trial (HPTN 052) clearly demonstrated that the use of antiretroviral therapy is associated with substantially decreased risk for transmission from HIV-positive persons to uninfected sexual partners. Furthermore, evidence also demonstrates that the identification and treatment of pregnant women dramatically reduces rates of mother-to-child transmission. The overall benefits of screening for HIV infection in adolescents, adults, and pregnant women are substantial.

On the basis of HIV prevalence data, MSM and active injection drug users are at very high risk for new HIV infection and would qualify for increased HIV screening. Behavioral risk factors for HIV infection include having unprotected vaginal or anal intercourse; having sexual partners who are HIV infected, bisexual, or injection drug users; or exchanging sex for drugs or money. Other persons considered at high risk include those who have acquired or request testing for other STIs. Patients may request HIV testing in the absence of reported risk factors. Individuals not at increased risk for HIV infection include persons who are not sexually active, those who are sexually active in exclusive monogamous relationships with uninfected partners, and those who do not fall into any of the previously mentioned categories. It is recognized that these risk categories are not mutually exclusive, the degree of sexual risk is a continuum, and individuals may not be aware of their sexual partner's risk factors for HIV infection. For patients younger than 15 years and older than 65 years, it would be reasonable for clinicians to consider HIV risk factors on an individual basis, especially those with new sex partners. However, clinicians should bear in mind that adolescent and adult patients may also be reluctant to disclose having HIV risk factors even when asked.

The evidence is insufficient to determine optimum time intervals for HIV screening. One reasonable approach would be a one-time screening of adolescents and adult patients to identify persons who are already HIV positive with repeated screening of those who are known to be at risk for HIV infection, those who are actively engaged in risky behaviors, and those who live or receive medical care in a high-prevalence setting (HIV seroprevalence > 1%). High-prevalence settings include STD clinics, correction facilities, homeless shelters, tuberculosis clinics, clinics serving MSM, and adolescent health clinics with high prevalence of STDs. Currently, a reasonable approach may be to re-screen groups at *very high* risk for new HIV infection at least annually and individuals at *increased risk* at slightly longer intervals (3 to 5 years). Women screened for HIV during previous pregnancies should be re-screened for HIV at all subsequent pregnancies. The CDC also recommends that all persons aged 13 to 65 years be screened for HIV in health care settings located in areas where the prevalence of undiagnosed HIV infection is greater than 0.1% and that persons with increased risk for HIV be re-tested at least annually.

Laboratory Assays

Diagnosis of HIV infection is usually based on serologic detection of immunoglobulin G (IgG) antibodies to HIV-specific proteins. A conventional serum test for diagnosing HIV infection is the repeatedly reactive immunoassay followed by a confirmatory Western blot or immunofluorescence assay. The combined tests are highly accurate with sensitivity and specificity greater than 99.5%. Results are available within 1 to 2 days for most commercial laboratories. New and improved assays are now available for early detection and confirmation of acute HIV infection, including combination tests (p24 antigen and HIV antibodies) and qualitative and quantitative HIV-1 RNA assays.

The diagnostic accuracy of HIV infection has improved with each generation of serologic assays. Whereas the first-generation tests were based on whole viral lysate and an indirect enzyme immunoassay, second-generation tests use synthetic and recombinant peptide antigens that have improved sensitivity and specificity. Third-generation assays have used "sandwich" assay formats that allow simultaneous detection of IgM and IgG antibodies. Now, fourth-generation assays combine antibody and antigen testing within the same diagnostic test format. With increasing sensitivity of these diagnostic assays, the "window period" wherein HIV antibodies may not be detected because of acute or very recent infection has gradually shortened from 6 weeks to less than 3 weeks.[11] This shortening of the window is particularly important when acute infection may not be suspected. In patients with symptoms and signs of acute HIV infection, direct testing with sensitive assays such as nucleic acid testing for HIV RNA may be used.

Rapid tests represent a major advance in HIV serologic testing. Rapid HIV testing may use either blood or oral fluid specimens and can provide results in 5 to 40 minutes. The sensitivity and specificity of the rapid tests are also greater than 99.5%; however, initially positive results require confirmation with conventional methods. Rapid testing can be offered on site in a variety of settings, including clinics, mobile vans, health fairs, and places of worship. Rapid testing is becoming the test of choice for all patients who request screening for immediate feedback and opportunities for quick intervention and counseling. Rapid tests are particularly important in management decisions of occupational or nonoccupational exposures, when patients are unlikely to return for results and seroprevalence rates are high, such as STD clinics or emergency departments, and when patients with an acute illness in which HIV-related complication is being considered and serostatus is not known.

In 2012, the U.S. Food and Drug Administration (FDA) approved the OraQuick in-home HIV test, the first self-administered HIV test kit to detect antibodies to both HIV-1 and HIV-2. The test is available for consumers in drugstores, and individuals may obtain test results within 20 to 40 minutes after collecting an oral fluid sample by swabbing the upper and lower gums inside the mouth and placing the sample into a developer vial provided as part of the kit. As with all rapid HIV assays, positive test results are preliminary and need to be confirmed with a standard HIV antibody test. In clinical trials, self-testing with this rapid HIV assay had a sensitivity of 92% and a specificity of 99.98% compared with the standard enzyme immunoassay (EIA) screening assay. An alternative to home testing is the home access HIV test system, which allows blood samples to be taken at home using a fingerstick test strip that is mailed to a laboratory for screening and confirmation. Results are obtained by phone using an individual identifier code supplied with the product.

Other methods to establish HIV infection include viral isolation or qualitative or quantitative detection of HIV nucleic acid through polymerase chain reaction techniques, branch-chain DNA testing, or nucleic acid sequence–based amplification. Limitations of these assays include cost, the requirement for venipuncture and more laboratory technology, and the time interval between sample collection and test results. None of these tests is considered superior to routine serologic testing. However, viral detection is useful in specific situations such as diagnosis of neonatal HIV infection when maternal antibody is passively transferred to the fetus, potentially providing a false-positive serologic result in uninfected infants and in patients with indeterminate serologic results or in those who may be in the window period before HIV seroconversion.[14]

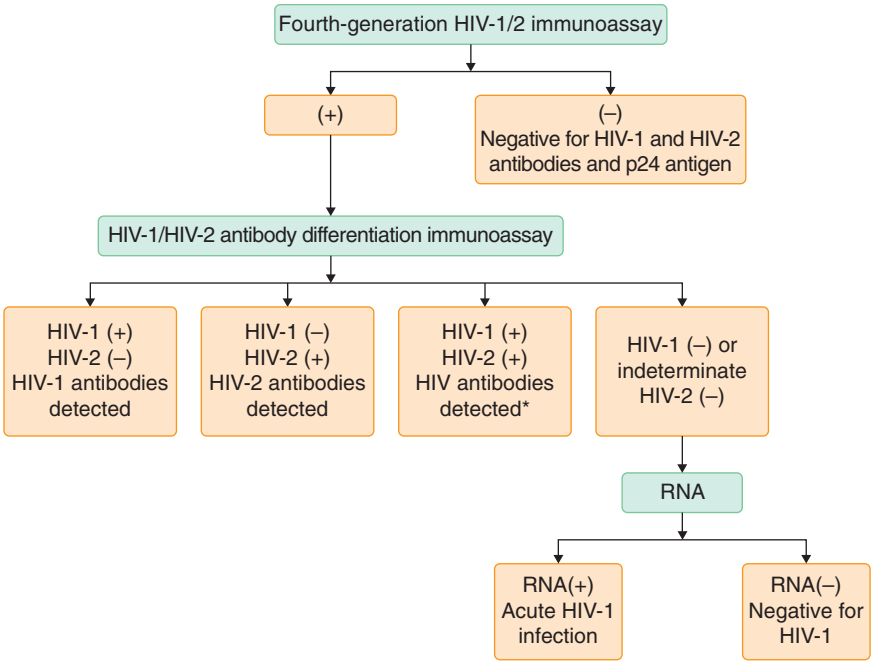

FIGURE 384-8. New HIV diagnostic testing algorithm evaluated in the United States, 2011-2013 HIV = human immunodeficiency virus. (From Centers for Disease Control and Prevention. Detection of acute HIV infection in two evaluations of a new HIV diagnostic testing algorithm—United States, 2011-2013. *MMWR.* 2013;62:489-494.
*Additional testing required to rule out dual infection with HIV-1 and HIV-2.

In 2013, the CDC evaluated and later offered an alternative-testing algorithm for the diagnosis of HIV infection[15] (Fig. 384-8). In this algorithm, all initial testing of serum is performed by an FDA-approved fourth-generation HIV-1/2 immunoassay. Specimens that are reactive on the fourth-generation assay should be re-tested/confirmed with an FDA-approved second-generation antibody assay that differentiates HIV-1 antibodies from other HIV antibodies, providing a definitive diagnosis of either HIV-1 or HIV-2. Seropositive individuals should initiate medical care that includes additional laboratory tests such as viral load, CD4 determination, and antiretroviral resistance assays to stage HIV disease and for the selection of initial antiretroviral drug regimens. Specimens that are reactive on the fourth-generation assay but negative on the HIV-1/HIV-2 antibody differentiation assay should be re-tested with an FDA-approved nucleic acid test for HIV-1 RNA. Under these circumstances a reactive nucleic acid test indicates the presence of acute HIV infection. A negative result would indicate the absence of HIV-1, and either a false-positive result on the initial fourth-generation assay or rarely recent HIV-2 infection. If HIV-2 infection is a possibility, a nucleic acid amplification test (NAAT) for HIV-2 DNA can be considered. However, HIV-2 infection is rare in the United States and there is no FDA NAAT for HIV-2. If a fourth-generation screening assay is not available, a third-generation HIV-1/2 immunoassay can be used as the initial test, followed by subsequent testing as specified in the algorithm. This alternative will miss some acute HIV infections in antibody-negative persons.

The previously described algorithm emphasizes high sensitivity during initial testing with the fourth-generation immunoassay, in which false-positive antibody-negative test results might occur, but these can be resolved during subsequent laboratory testing as recommended as part of initial clinical evaluation. The new diagnostic algorithm replaces the Western blot with an HIV-1/HIV-2 antibody differentiation assay as the supplemental test and includes an RNA test to resolve reactive immunoassays with negative supplemental test results. In retrospective studies, this algorithm performed better than Western blot at identifying HIV antibody-positive persons, detecting acute HIV infections, and diagnosing unsuspected HIV-2 infections.

Grade A References

A1. Severe P, Juste MA, Ambroise A, et al. Early versus standard antiretroviral therapy for HIV-infected adults in Haiti. *N Engl J Med*. 2010;363:257-265.

A2. Cohen MS, Chen YQ, McCauley M, et al. HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med*. 2011;365:493-505.

A3. De Vicenzi I. Kesho Bora Study Group. Triple antiretroviral compared with zidovudine and single-dose nevirapine prophylaxis during pregnant and breast-feeding for prevention of mother-to-child transmission of HIV-1 (Kesho Bora study): a randomized controlled trial. *Lancet Infect Dis*. 2011;11:171-180.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Piot P, Quinn TC. Response to the AIDS Pandemic: a global health model. *N Engl J Med*. 2013;368:2210-2218.
2. UNAIDS Global Report. *UNAIDS Report on the Global AIDS Epidemic 2012*. Geneva: UNAIDS; 2013; and UNAIDS GAP Report, Geneva, UNAIDS 2014.
3. World Health Organization, UNAIDS. *Global Update on HIV Treatment: Results, Impact and Opportunities*. Geneva: World Health Organization; 2013.
4. Maartens G, Celum C, Lewin SR. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet*. 2014;384:258-271.
5. Bor J, Herbst AJ, Newell ML. Bäminghausen T. Increases in adult life expectancy in rural South Africa: valuing the scale-up of HIV treatment. *Science*. 2013;339:961-963.
6. Tanser F, Bäminghausen T, Grapsa E, et al. High coverage of ART associated with decline in risk of HIV acquisition in rural KwaZulu-Natal, South Africa. *Science*. 2013;339:966-971.
7. Centers for Disease Control and Prevention. Estimated HIV incidence in the United States, 2007-2010. *HIV Surveillance Suppl Rep*. 2012;17:4.
8. Centers for Disease Control and Prevention. HIV Surveillance—United States, 1981-2008. *MMWR*. 2011;60:689-693.
9. Centers for Disease Control and Prevention. Vital signs: HIV prevention through care and treatment—United States. *MMWR*. 2011;60:1618-1623.
10. Wood E, Kerr T, Rowell G, et al. Does this adult patient have early HIV infection?: The Rational Clinical Examination systematic review. *JAMA*. 2014;312:278-285.
11. Moyer VA. Screening for HIV: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2013;159:51-60.
12. Chou R, Selph S, Dana T, et al. Screening for HIV: Systematic review to update the 2005 U.S. Preventive Services Task Force Recommendation. *Ann Intern Med*. 2012;157:706-718.
13. Cornett JK, Kim TJ. Laboratory diagnosis of HIV in adults: a review of current methods. *Clin Infect Dis*. 2013;57:712-718.
14. Burgard M, Blanche S, Jasseron C, et al. Performance of HIV-1 DNA or HIV-1 RNA tests for early diagnosis of perinatal HIV-1 infection during anti-retroviral prophylaxis. *J Pediatr*. 2012;160:60-66.
15. Centers for Disease Control and Prevention. Detection of acute HIV infection in two evaluations of a new HIV diagnostic testing algorithm—United States, 2011-2013. *MMWR*. 2013;62:489-494.

REVIEW QUESTIONS

1. Which of the following statements is *incorrect:*

A. In 2013, 2.1 million people became newly infected, half of whom were young individuals between the age of 15 and 24 years.
B. By the end of 2013, 35 million people were living with HIV globally.
C. The epidemic has stabilized in many regions, with new infections decreasing by 50% or more in 25 countries.
D. HIV prevalence and incidence had declined sharply in 2013 in Eastern Europe and Central Asia.

Answer: D Prevalence and incidence have increased sharply over the past decade in Eastern Europe and Central Asia, partially driven by injection drug use and sexual transmission. In addition, treatment for those infected has also lagged behind in these regions.

2. Which of the following regions has yet to attain over 50% coverage of antiretroviral therapy for eligible individuals?

A. Latin America
B. The Caribbean
C. Oceania
D. Eastern Europe and Central Asia
E. Eastern and southern Africa

Answer: D Eastern Europe and Central Asia have reached coverage of only 25% of antiretroviral therapy for HIV-infected individuals eligible for treatment. The Middle East and North Africa have attained only a 15% coverage rate, whereas Latin America, the Caribbean, Oceania, and Eastern and southern Africa have all attained coverage rates over 56 to 69%.

3. Which of the following statements regarding the HIV epidemic in the United States is *false*?

A. The incidence of new HIV infections in 2012 has remained relatively stable over the past decade at approximately 50,000 new cases.
B. The incidence of HIV infection has declined in men who have sex with men (MSM), but has increased among injecting drug users.
C. HIV occurs at disproportionately higher rates among African Americans and Hispanics compared to white and other ethnicity groups.
D. Most women acquire HIV through unprotected heterosexual exposure.

Answer: B The frequency of new HIV infections among MSM continues to be relatively stable or in some areas slightly increased. There have been no observations within the last decade of any significant declines in HIV among this population. In contrast, HIV incidence has decreased among injection drug users.

4. The U.S. Preventive Services Task Force recently recommended that clinicians should screen for HIV in all adolescents and adults aged 15 to 65 years of age. This recommendation was based on which of the following facts?

A. There is increasing evidence of clinical benefit through early antiretroviral therapy for HIV-infected persons.
B. A recent randomized, clinical trial demonstrated that antiretroviral therapy is effective in reducing HIV transmission.
C. Approximately 20% of HIV-infected individuals have never been tested and are unaware of their infection.
D. Use of antiretroviral therapy during pregnancy has been shown to be effective in decreasing mother-to-child transmission of HIV.
E. All of the above.

Answer: E All of the above supporting statements for early diagnosis and treatment providing benefit to HIV infected persons provided the evidence for the USPSTF recommendations for routine screening for HIV.

385

IMMUNOPATHOGENESIS OF HUMAN IMMUNODEFICIENCY VIRUS INFECTION

JOEL N. BLANKSON AND ROBERT F. SILICIANO

Human immunodeficiency virus type 1 (HIV-1) infection results in a progressive state of immune system dysregulation that ultimately leads to the acquired immunodeficiency syndrome (AIDS), which is characterized by profound depletion of CD4+ T lymphocytes and the inability to control infections by opportunistic pathogens that do not cause disease in individuals with a normal immune system. Despite decades of research on this condition, the basic pathogenic mechanisms are still incompletely understood. This chapter reviews current understanding of the mechanisms involved in HIV-1 immunopathogenesis.

PRIMARY INFECTION

The natural history of HIV-1 infection is illustrated in Figure 385-1. During acute HIV-1 infection, massive viral replication occurs in CD4+ T lymphocytes in the absence of an adaptive immune response. CD4+ T cells in the gut-associated lymphoid tissue and other mucosal sites express high levels of the HIV coreceptor CCR5 and are thus particularly prone to infection and depletion by the commonly transmitted R5 variants of HIV-1. In animal models, it has been shown that approximately 30% of memory CD4+ T cells are infected and depleted by 4 days after infection. This is in contrast to chronic infection, in which less than 1% of all CD4+ T cells are productively

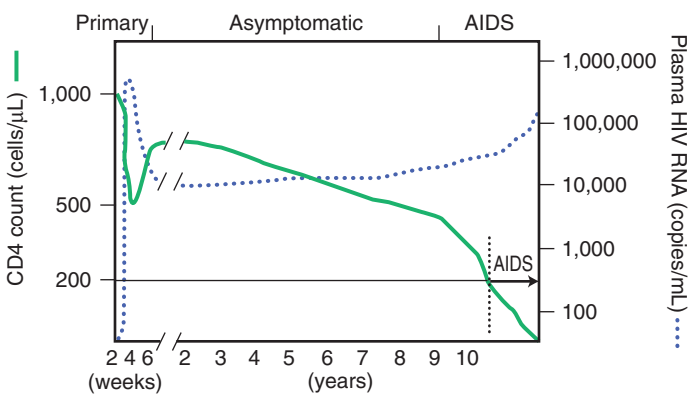

FIGURE 385-1. Natural history of HIV-1 infection. CD4 counts and viral load are shown in the three phases of infection. AIDS = acquired immunodeficiency syndrome.

infected at any given time. As a result of this massive early infection, plasma levels of virion-associated HIV-1 RNA of more than 1 million copies per milliliter are typically seen in plasma within 2 weeks of infection and patients tend to experience a constellation of signs and symptoms known as the *acute retroviral syndrome.* There can be significant declines in peripheral CD4+ T-cell counts in primary infection resulting in opportunistic infections. Within several weeks, the development of an effective HIV-1–specific cytolytic T-lymphocyte response results in the partial control of viral replication, and the plasma HIV-1 RNA level (commonly known as the viral load) falls and reaches a steady-state level known as the *set point.* The magnitude of the set point viral load during the second asymptomatic phase of the infection reflects a dynamic equilibrium between viral replication and the HIV-1–specific immune response. This set point determines the rate of progression to the final phase, clinical AIDS. The median set point is approximately 30,000 HIV-1 RNA copies per milliliter, and most patients who have this level of viremia will develop AIDS in a 5- to 10-year period if they are left untreated. Patients with much higher set point viral loads will tend to be rapid progressors who develop AIDS much more quickly; patients with much lower viral loads tend to be long-term nonprogressors.

SPECIFIC IMMUNITY FOR HUMAN IMMUNODEFICIENCY VIRUS

The host mounts a vigorous immune response to HIV-1 infection. The virus is thought to activate plasmacytoid dendritic cells through toll-like receptors, resulting in the secretion of type I interferons and other inflammatory cytokines. Whereas type I interferons have direct antiviral properties and enhance the HIV-1–specific immune response, excessive secretion may play a key role in pathogenic immune activation of CD4+ and CD8+ T cells (Fig. 385-2). Natural killer (NK) cells are important effector cells in the innate immune response that become activated when infection of target cells by HIV-1 or other viruses results in the downregulation of HLA molecules. Patients expressing certain NK receptor alleles are more likely to become long-term nonprogressors, suggesting that these cells may play a protective role possibly by controlling early HIV-1 replication, leading to the development of an effective adaptive immune response. Myeloid dendritic cells play a key role in the presentation of HIV-1 antigens to HIV-1–specific CD4+ and CD8+ T cells, which results in initiation of the adaptive immune response. They express CD4 molecules and have been shown to bind HIV-1. It is thought that in the process of presenting antigen, these cells may inadvertently transmit HIV-1 to clusters of activated CD4+ T cells.

The role of the humoral response in HIV-1 infection is not clear. HIV-1–specific antibodies, which are used to diagnose HIV-1 infection, do not develop until after peak viremia occurs. There is thus a window period in primary HIV-1 infection during which viremia is present in the absence of detectable antibodies. A subset of the antibodies that eventually appear are capable of preventing infection by blocking the interaction of the HIV-1 envelope protein gp120 with CD4 and coreceptor proteins on the surface of target cells. These so-called neutralizing antibodies are present at relatively low titers and have limited access to the critical regions of gp120. Recent studies suggest that the most effective neutralizing antibodies do not develop until 2 to 3 years after infection.[1] These broadly neutralizing antibodies may eventually form the basis of a vaccine by preventing new infections. However, although neutralizing antibodies in general can exert significant selective pressure on the virus, immunologic escape through rapid viral evolution is common, and the bulk of evidence suggests that these antibodies do not play a major role in the control of viral replication in most long-term nonprogressors.

The selective depletion of CD4+ T cells is the main reason that HIV-1 infection results in such profound immunosuppression; these so-called helper T cells play a major role in every facet of the adaptive immune response. The ability of HIV-1–specific CD4+ T cells to proliferate and to secrete key cytokines such as interleukin-2 (IL-2) is lost shortly after primary infection, setting up the entire HIV-1–specific response for failure.

CD8+ T cells contribute to the control of HIV-1 infection by the direct lysis of infected cells and by the secretion of soluble factors such as macrophage inflammatory protein 1β that bind to chemokine receptors, thereby preventing HIV-1 entry into target cells.[2] However, the HIV-1–specific CD8+ T-cell response that partially controls HIV-1 replication after peak viremia in primary infection does not achieve sterilizing immunity, partly because of a reservoir of latent virus in resting memory CD4+ T cells that develops shortly after infection. These quiescent cells probably do not make HIV-1 proteins and thus are not recognized by cytolytic T lymphocytes. Furthermore, the cytolytic T-lymphocyte response in patients with progressive disease is of poor quality with limited proliferative capacity. Most importantly, the low fidelity of HIV-1 reverse transcriptase results in the development of mutations with each round of replication. Mutations that lead to escape from cytolytic T-lymphocyte responses have a selective advantage and are thus selected for rapidly.

THE EFFECT OF HUMAN IMMUNODEFICIENCY VIRUS-1 REPLICATION ON THE IMMUNE SYSTEM

Whereas the HIV-1–specific immune response helps limit the rate of viral replication, sterilizing immunity is never achieved and ongoing viral replication has a negative impact on the immune system. Continuous viral replication results in chronic immune activation (see Fig. 385-2).[3] The mechanism is not understood. The chronic immune response to the virus may lead to nonspecific inflammation, and microbial translocation resulting from the depletion of CD4+ T cells in the gut-associated lymphoid tissue also may be important. Whatever the mechanism, immune activation appears to drive the depletion of CD4+ T cells. The level of immune activation markers on CD8+ T cells correlates better with the rate of CD4 decline than does the magnitude of the viral load in untreated patients.

Increased levels of activation markers are seen on NK cells, B cells, CD4+ T cells, and CD8+ T cells. Activation is accompanied by an increase in the turnover rate of these cells. The function of NK cells is compromised, which may predispose to the poor control of other viruses. B-cell defects result in hypergammaglobulinemia and the production of autoantibodies. Poor antibody responses to vaccines are also seen as CD4+ T cells decline.

Studies of viral dynamics make it clear that most productively infected cells live only a short time (~1 day) before succumbing to viral cytopathic effects or host cytolytic T lymphocytes or NK cells. Although the loss of infected CD4+ T cells contributes to CD4 depletion, there is marked depletion of $CD4^+$ T cells even though at any given time during chronic infection only 1% or less of these cells is productively infected. However, recent studies have suggested that non-productively infected CD4+ T cells are also susceptible to cell death by a pro-apoptotic and proinflammatory host response. Thus, it is currently thought that death of non-productively infected CD4+ T cells and chronic immune activation leading to the death of noninfected CD4+ T cells are the principle mechanisms for CD4 depletion.[4,5] Support for this idea comes from studies of the closely related simian immunodeficiency virus that replicates in natural simian hosts without causing immune activation or CD4 depletion. In addition to the quantitative loss of CD4+ T cells in HIV-1 infection, marked skewing of the CD4 T-cell repertoire is seen, and there is a diminished qualitative memory response to recall antigens long before the CD4+ T-cell count drops to 200 cells/μL. The chronic activation and high turnover rate of these cells eventually result in the progressive CD4 decline that is characteristic of HIV-1 infection.

There is evidence of immune exhaustion for both CD4+ and CD8+ T cells, and there is decline in the qualitative features of the CD8+ T-cell response to other chronic viruses such as cytomegalovirus and Epstein-Barr virus. This may be the result of anergy or depletion of the CD4+ T cells that are needed to sustain functional CD8+ T-cell responses.

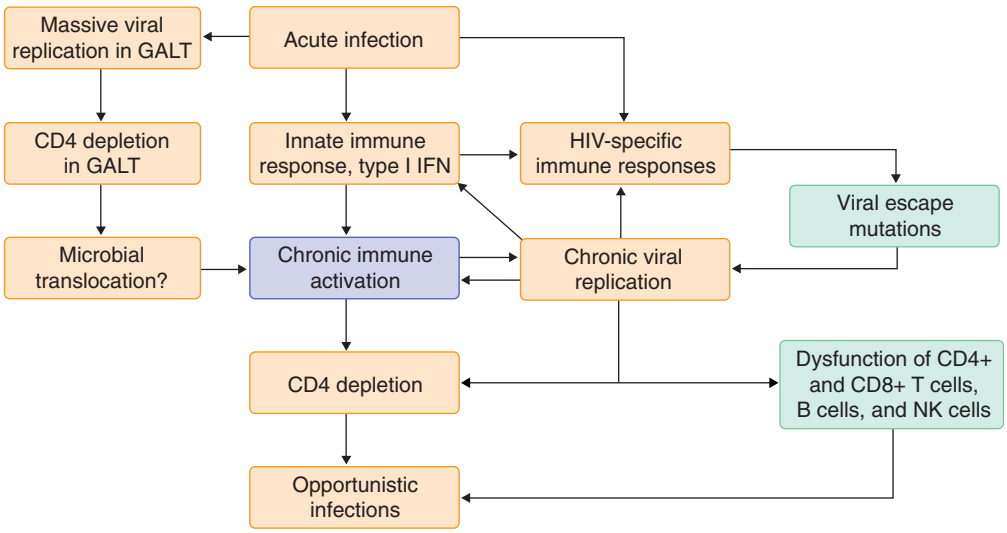

FIGURE 385-2. Parameters involved in chronic immune activation and CD4+ T-cell depletion. GALT = gut-associated lymphoid tissue; IFN = interferon; NK = natural killer.

CLINICAL CONSEQUENCES OF HUMAN IMMUNODEFICIENCY VIRUS INFECTION

Clinical immunodeficiency is associated with the late stages of HIV-1 infection, when profound CD4+ T-cell depletion has occurred. However, some level of immunodeficiency may be present shortly after infection because of qualitative changes in the immune response related to ongoing viral replication. As a result, patients are more susceptible to infections such as *Mycobacterium tuberculosis* infection before the CD4 count reaches the 200 cells/μL threshold that defines AIDS. Patients are also much more susceptible to malignant neoplasms such as non-Hodgkin's lymphoma at any CD4+ T-cell count. Other opportunistic infections arise at fairly predictable CD4+ T-cell counts. *Pneumocystis jiroveci* infections occur at CD4 counts of lower than 200 cells/μL, *Cryptococcus neoformans* and *Toxoplasma gondii* infections occur at CD4 counts of lower than 100 cells/μL, and *Mycobacterium avium* complex and cytomegalovirus infections occur at CD4 counts below 50 cells/μL. Whereas these infections are a consequence of diminished cellular immunity, there is also a marked increase in invasive pneumococcal infections in HIV-1–infected patients, possibly because of defects in humoral immunity.

The Response to Antiretroviral Therapy

Treatment with selected combinations of antiretroviral drugs, which is known as highly active antiretroviral therapy (HAART), suppresses viral replication to below the limits of detection of current commercial assays (50 copies of HIV-1 RNA per milliliter of plasma). Current evidence suggests that HAART produces a complete or nearly complete arrest in viral replication in adherent patients, but trace amounts of viremia persist because of stable viral reservoirs, including the latent reservoir in resting CD4+ T cells. This suppression of viral replication is usually accompanied by a substantial increase in CD4+ T-cell counts. The initial rise in CD4+ T-cell counts occurs mostly as a consequence of migration of cells from lymph nodes (where 98% of all CD4 T cells reside) to the peripheral blood as inflammation diminishes in lymphoid tissue. Subsequently, there is an increase in the production of memory CD4+ T cells in most individuals. Naïve T-cell production is also sometimes seen at lower levels. Clinical studies have shown that patients who experience significant immune reconstitution[6] can safely discontinue prophylactic therapy for opportunistic infections. However, it is not clear whether patients who maintain undetectable viral loads while receiving HAART yet do not achieve significant CD4 T-cell immune reconstitution are still at risk for opportunistic infections. Two large studies have shown that although the use of IL-2 treatment in conjunction with HAART will cause a significant increase in CD4+ T-cell counts in these patients, the enhanced immune reconstitution is not associated with any clinical benefit.

Just as a decline is seen in the functional CD4+ T-cell response shortly after primary infection, there is a qualitative improvement in $CD4^+$ T-cell function shortly after HAART is initiated. In some cases, there are exaggerated immune responses to opportunistic infections leading to the immune reconstitution inflammatory syndrome (IRIS; Chapter 395), particularly when there is rapid control of viral replication after the initiation of HAART. IRIS usually presents as a paradoxical worsening of a disease process a few weeks after HAART is started and can occur even before there are significant changes in the absolute CD4+ T-cell counts. IRIS has been reported for virtually all known opportunistic infections, and in some cases it can occur in response to previously unrecognized infections. There have not been clinical trials looking at treatment of this condition, but nonsteroidal anti-inflammatory drugs and corticosteroids have been routinely used with varying degrees of success (Chapter 395).

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Moir S, Fauci AS. Insights into B cells and HIV-specific B-cell responses in HIV-infected individuals. *Immunol Rev*. 2013;254:207-224.
2. Hersperger AR, Migueles SA, Betts MR, et al. Qualitative features of the HIV-specific CD8+ T-cell response associated with immunologic control. *Curr Opin HIV AIDS*. 2011;6:169-173.
3. Paiardini M, Müller-Trutwin M. HIV-associated chronic immune activation. *Immunol Rev*. 2013; 254:78-101.
4. Doitsh G, Galloway NL, Geng X, et al. Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature*. 2014;505:509-514.
5. Monroe KM, Yang Z, Johnson JR, et al. DNA sensor is required for death of lymphoid CD4 T cells abortively infected with HIV. *Science*. 2014;343:428-432.
6. Martin-Blondel G, Mars LT, Liblau RS. Pathogenesis of the immune reconstitution inflammatory syndrome in HIV-infected patients. *Curr Opin Infect Dis*. 2012;25:312-320.

REVIEW QUESTIONS

1. What is the most likely cause of CD4 depletion in HIV-infected patients?

A. Chronic immune activation
B. Direct killing of productively infected cells
C. Infection by opportunistic pathogens
D. Metabolic disturbances

Answer: A Chronic immune activation is thought to play the most significant role, although a recent study suggests that non-productively infected cells may die from a protective host response. Only 1% of cells are productively infected, and therefore it is felt that the death of these cells cannot explain the level of CD4 depletion seen.

2. At what CD4 count do patients become susceptible to *Pneumocystis jiroveci*?

A. 50 cells/μL
B. 100 cells/μL
C. 200 cells/μL
D. 500 cells/μL

Answer: C Clinical studies have shown that patients become susceptible to *Pneumocystis jiroveci* at a CD4 count of less than 200 cell//μL. *Cryptococcus neoformans* and *Toxoplasma gondii* infections occur at CD4 counts < 100 cells/μL; and *Mycobacterium avium* complex and CMV infections occur at CD4 counts < 50 cells/μL.

3. HIV-1–infected patients are highly susceptible to invasive pneumococcal disease because of:

A. Immune activation.
B. Dysfunction of B cells.
C. Apoptosis of CD8+ T cells.
D. Increased exposure to pneumococcus.
E. Hyposplenism.

Answer: B Although HIV-1 infection typically causes a defect of cellular immunity, humoral immunity is also affected because of a B-cell dysfunction and patients become susceptible to bacterial infections.

4. At what point after HIV infection do patients develop the acute retroviral syndrome?

A. 2 weeks
B. 6 weeks
C. 3 days
D. 3 months

Answer: A Patients typically develop the constellation of symptoms known as the acute retroviral syndrome 2 weeks after infection. At that point, as a result of the massive early infection, plasma levels of virion-associated HIV-1 RNA of more than 1 million copies/mL are typically seen in plasma.

5. What is the mechanism of the drop in HIV-1 RNA to the set point level in primary infection?

A. The development of an HIV-1–specific CD4+ T-cell immune response
B. The establishment of the latent reservoir
C. The development of neutralizing antibodies
D. The development of a an HIV-1–specific CD8+ T-cell immune response

Answer: D The development of the HIV-specific CD8+ T-cell response has been associated with partial control of viral replication in clinical studies and depletion of CD8+ T cells in an animal model abrogates this control of viremia.

6. Which of the following is the biggest barrier to the eradication of HIV-1?

A. The establishment of the latent reservoir
B. Chronic immune activation
C. B-cell dysfunction
D. Microbial translocation

Answer: A The quiescent memory CD4+ T cells of the latent reservoir probably do not express HIV antigens and thus are not targeted by the immune response. They are also not effectively targeted by antiretroviral therapy, thus preventing the eradication of HIV-1.

386

BIOLOGY OF HUMAN IMMUNODEFICIENCY VIRUSES

FRANK MALDARELLI

Human immunodeficiency virus (HIV) causes progressive immune deficiency and death from opportunistic infections or neoplastic diseases. Over 34 million individuals worldwide are currently infected with HIV, and over 30 million have died since the disease was first recognized in 1981. Developments in diagnosis, prevention, and treatment have reduced morbidity and mortality from HIV, but the epidemic remains substantial; in 2012, 2.3 million new HIV infections occurred and 1.6 million people died from HIV and acquired immunodeficiency syndrome (AIDS).[1] Improvements in HIV care grew from an understanding of the discovery, characterization, and elucidation of replication of HIV. Continued research led to discoveries of cellular processes, mechanisms of pathogenesis, and new concepts of host antiviral immunity. Here we summarize basic concepts of HIV biology. Our understanding of HIV remains incomplete, and efforts are ongoing to improve testing, characterize viral replication and pathogenesis, identify novel antiviral targets, and develop innovative strategies to eradicate HIV infection.

CLASSIFICATION AND ORIGIN

HIV belongs to the lentivirus genus of the Orthoretrovirinae subfamily of Retroviridae; all retroviruses are defined by the presence of a specific enzyme, reverse transcriptase, that catalyzes the synthesis of DNA from an RNA

template, the central and unique event in retrovirus replication permitting integration of the viral DNA into the host genome. Retroviridae is a large family of viruses infecting diverse vertebrate hosts, mostly mammals and birds, and to a lesser degree, reptiles and fish. Retroviruses are responsible for a spectrum of diseases, including immunodeficiencies and neoplastic, neurologic, hematologic, encephalitic, and inflammatory disorders. Members of the lentivirus genus cause chronic, recurrent, or progressive diseases, including immunodeficiencies in various mammal species.

The origin of HIV variants currently circulating in humans has been traced by analyzing nucleic acid sequences of HIV and closely related viruses in primates; these analyses strongly indicate HIV emerged from zoonotic transmissions from primates to humans during the period 1890 to 1930 in Central and West Africa.[2] Zoonotic transmission requires close contact of blood and body fluid; scratches, bites, and butchery of captured infected animals provide ready mechanisms for transmission. Such opportunities for zoonotic transmission have likely taken place for thousands of years, and the reasons why an epidemic spread did not occur until recently are unclear. A number of factors may have contributed to epidemic spread during the late 19th and 20th centuries, including profound increases in human population density, habitat destruction forcing more contact between humans and other primates, malnutrition contributing to underlying immunodeficiency, population shifts due to political upheaval, and the development of infrastructure, such as roads, facilitating human travel over long distances.

The viral etiology of AIDS was first identified after intensive investigation of patients identified with AIDS. Cell cultures inoculated with plasma or co-cultured with lymphocytes from individuals with AIDS resulted in marked cytopathic effect; cell-free material transmitted the infection to fresh uninfected cultures and reproduced cytopathology; cultures contained both reverse transcriptase enzymatic activity and virions with morphology characteristic of lentiviruses. The virus was initially named HTLV-III because of its apparent relationship to the other human retroviruses HTLV-I and II; subsequent study demonstrated the virus responsible for AIDS was only distantly related to HTLV, and the new virus was renamed HIV. Molecular clones were constructed that reproduced cellular cytopathology. Plasma from infected individuals contained substantial antibodies to the virus, permitting development of robust new enzyme-linked immunosorbent assay (ELISA) and Western blot detection assays useful for patient diagnosis, epidemiologic surveillance, and blood product protection.

Following the development of the first tools for laboratory diagnosis of HIV infection, additional analysis identified individuals with symptomatic AIDS but who did not have serologic responses characteristic of HIV, leading to the identification of a second distinct immunodeficiency virus, denoted HIV-2. HIV-2 had a distribution restricted largely to West Africa and to countries having close economic, political, or cultural ties to West Africa.

Further genetic analysis and sampling of nonhuman primate species shed new light on sources of the HIV epidemic. HIV-1 can be classified using nucleic acid sequence analysis into four distinct groups: a large group of viruses found throughout the world (M, for "main"), a relatively small group of viruses in central Africa (O, for "other"), and two small groups in individuals with West African origin, comprising only fewer than 20 total infections, denoted N (not M, not O) and P. Phylogenetic reconstructions demonstrate distinct lineages for HIV-1 groups, indicating each is the result of a distinct zoonotic event (Fig. 386-1), which in the case of HIV-1 M, has spread worldwide. M viruses are closely related to a similar virus, simian immunodeficiency virus, present in chimpanzee species (denoted SIV_{cpz}, simian immunodeficiency virus, cpz to indicate the specific animal reservoir). Group O is most closely related to the SIV present in gorillas and in chimpanzees. In contrast, HIV-2 is highly related to SIV_{smm}, a lentivirus commonly present in the sooty mangabey, which has a geographic range including West and Central Africa, where HIV-2 was identified.

Group M viruses represent the great majority of infections in humans and are quite diverse, with at least nine subtypes, denoted A to D, F to H, J, and K. Analysis of the relationships of the nucleic acid sequences reveals an overall "starlike" phylogeny (see Fig. 386-1), indicating that, in general, all the current variants emerged from a common ancestor. Studies of the earliest viral sequences identified revealed it is likely that HIV underwent diversification early after transmission from animal reservoirs. Recombination among these subtypes may occur (see Fig. 386-1, *A/D*; and see later), yielding new recombinant viruses that are classified as circulating recombinant forms. Dual HIV-1/HIV-2 infections are also possible in geographic regions where viruses co-circulate, typically from West Africa.

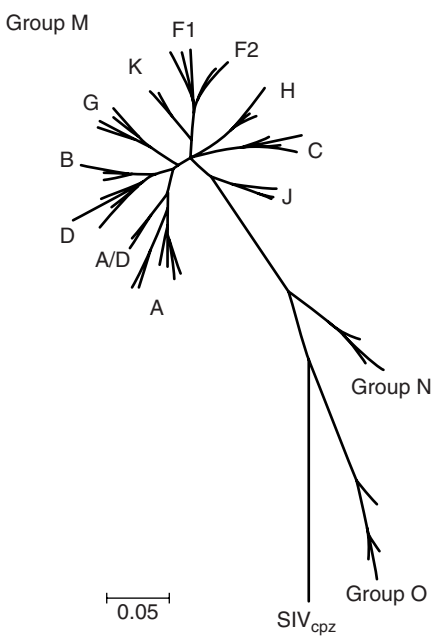

FIGURE 386-1. Phylogenetic relationships of HIV-1 groups and subtypes. Reference sequences of HIV RT were aligned in maximum likelihood phylogenetic trees constructed and rooted on the distant ancestor, SIV chimpanzee sequence (SIV_{cpz}). Group M viruses radiate in a starlike fashion, consistent with a common ancestor. An example of recombinant sequences (A/D) containing portions of parental subtypes A and D are identified as intermediate between the parental subtypes. Group O and N are distantly related to group M and likely represent independent zoonotic events. Marker = genetic distance depicting percent difference. HIV Sequence Database, http://www.hiv.lanl.gov/.

Once AIDS emerged, HIV spread rapidly throughout the world; while Africa maintained a highly diverse group of viruses, founder effects resulted in spread of one or a limited number of subtype viruses in countries outside Africa. The epidemic in the United States began with subtype B virus; although initially attributed to single case reports of individuals, careful and exhaustive analyses traced the HIV epidemic from Africa to the United States through Haiti. Importantly, current laboratory methods to detect HIV identify *all* HIV-1 and HIV-2 variants. In individuals from relevant geographic origin, especially West Africa, it is of paramount importance to characterize the infection precisely to target therapy appropriately; HIV-2 and group O HIV-1 are naturally resistant to non-nucleoside reverse transcriptase inhibitors and to fusion inhibitors.[3] As dual HIV-1/HIV-2 infections are possible, it is imperative that all individuals from endemic areas for these two viruses (especially West Africa) be tested for *both* HIV-1 and HIV-2.

STRUCTURE AND MOLECULAR BIOLOGY

Genome Structure and Organization

Like all retroviruses, HIV replicates via a DNA intermediate. The virion contains two copies of single-stranded (+) sense RNA (denoted viral RNA), and the stably infected cell contains double-stranded viral DNA integrated into the host genome (denoted the provirus). As shown in Figure 386-4 and E-Figure 386-1, viral RNA and the provirus have distinct genomic organization in untranslated 3′ and 5′ regions. Untranslated regions at each end of the genome contain a short terminal repeat, as well as unique sequences at the 5′ (U5) and 3′ (U3) regions that are duplicated during replication, generating a longer duplicated sequence termed the *long terminal repeat* at each end of the provirus.

HIV encodes nine genes whose products are required for structural, enzymatic, regulatory, and innate immune neutralization functions (see E-Fig. 386-1). By convention, HIV genes are denoted in lower italics (*gag, pol,* etc.) with names that broadly reflect their function, location in the virion, or a historical vestige of prior viral classification (e.g., *gag,* "group antigen"). Additional genes were characterized that function in regulation or replication (*tat, rev, vpr*) or in blocking immune responses to HIV (*vif, vpu, nef*); all are critical for HIV infection in vivo.

Virion Structure

The HIV virion contains viral gene products and cellular components essential to transmit infection and establish the proviral state. HIV virions are

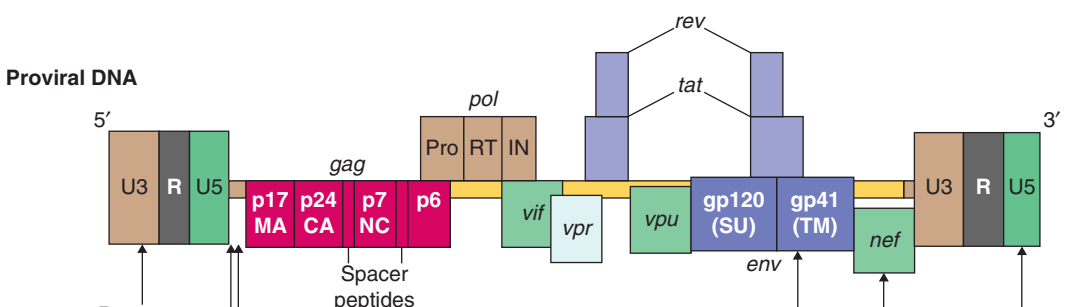

HIV gene	Function	Host binding partner	FDA approved inhibitor
***gag* ("group antigen")**			
p17 matrix MA	**Binds cellular membranes**	**Host membrane**	
p24 capsid CA	**Virion structure**	**TRIM5-alpha**	
pol ("polymerase")			
Protease	Processes gag/pol polyprotein		Yes
RT	Reverse transcription of viral RNA		Yes
RT-RNase H	Degrades RNA from RNA:DNA hybrid		
Integrase	Integrates linear HIV DNA into host genome		Yes
vpr ("viral protein R")	Uncertain		
vif ("viral infectivity factor")	Innate immune evasion	APOBEC family of proteins	
tat ("trans-activator of transcription")	Transcriptional regulation	Transcription factors	
rev ("regulator of expression of viral proteins")	Post-transcriptional RNA processing	CRM-1	
vpu ("viral protein U")	Adaptive immune evasion, facilitate Env trafficking	CD4	
	Innate and adaptive immune evasion	Tetherin	
***env* ("envelope")**			
gp120	**HIV attachment**	**CD4, coreceptor**	**Yes**
gp41	**HIV fusion**		**Yes**
nef ("negative regulatory factor")	Adaptive immune evasion	MHC1	
	Adaptive immune evasion	CD4	

E-FIGURE 386-1. HIV Genome organization and function. Proviral DNA with complete long terminal repeat (LTR) is depicted. HIV genes are lowercase *italics*. Gene products are capitalized Roman. Genes (not drawn to scale) are color coded, and the function of the corresponding gene product is summarized. Host components interacting with each gene product are noted, and gene products for which there are U.S. Food and Drug Administration–approved therapeutic drugs are indicated.

roughly spherical particles with a diameter of 80 to 120 nm and are composed of a viral core enveloped by a lipid membrane (Fig. 386-2).

The core of the mature virion is a conelike structure,[4] composed of the HIV p24 capsid (CA), which encapsidates components necessary for replication: two copies of HIV genomic RNA template complexed with HIV p6 nucleocapsid (NC); $tRNA^{lys}$ primer; HIV enzymes reverse transcriptase, protease, and integrase; and HIV Vif. The viral envelope, which is derived from the plasma membrane of the host cell as HIV undergoes budding, contains viral proteins Gp120 (SU) and Gp41 (TM), as well as a structural matrix protein, MA.

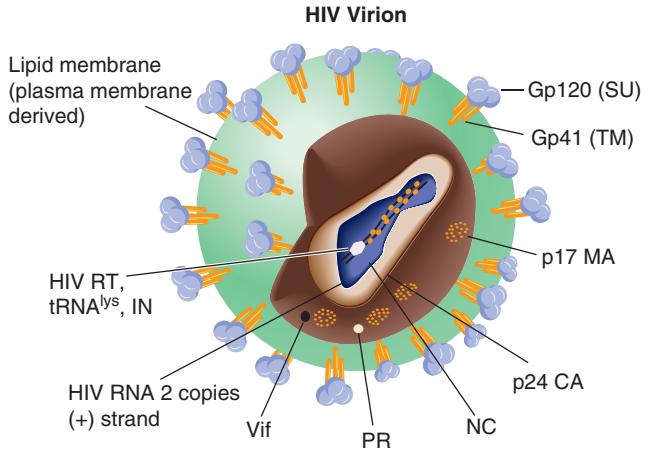

FIGURE 386-2. The HIV virion. HIV is an enveloped virus consisting of two (+) sense copies of viral RNA, enzymes required for replication contained in a viral core enveloped in a membrane derived by budding from the infected cell.

Replication Cycle

Early Events in Replication

Attachment and Fusion

Virus replication is initiated by direct contact of virions or infected cells with susceptible host cells (Fig. 386-3). Productive infection requires specific and essential interactions mediated by surface Env glycoprotein trimeric complexes consisting of HIV sp120 SU noncovalently bound to HIV gp41 TM.[5] The attachment phase is mediated exclusively by SU, which engages two distinct cell surface proteins for attachment, a receptor and a coreceptor (see Fig. 386-3). HIV initially binds CD4, resulting in conformational change in Env, which facilitates coreceptor binding. Typically, SU proteins use either the human chemokine receptor 5 (CCR5) or the human chemokine receptor 4 (CXCR4), to infect CD4+ T cells, but CCR5/CXCR4 dual tropic viruses circulate as well. Although virus infection can be propagated with either CCR5 or CXXCR4 as coreceptor, initial infection likely requires CCR5 tropic virus. Human populations have a significant population of individuals encoding a mutant CCR5 gene, who do not synthesize functional CCR5. HIV infection of individuals homozygous for CCR5 mutation is exceedingly rare, suggesting that initiation of infection virtually always requires interactions with CCR5. Inhibition of SU-coreceptor interactions has been achieved pharmacologically, and the U.S. Food and Drug Administration–approved coreceptor inhibitor maraviroc has potent anti-HIV activity.[6]

Engaging both receptor and coreceptor results in conformational change in Gp41, which reorganizes its structure and provides sufficient energy to drive membrane fusion. TM-mediated membrane fusion can be inhibited using specific peptide inhibitors that bind to Gp41; one such peptide fusion inhibitor, enfuvirtide, is in clinical practice.

Uncoating

Following entry into CD4 cells, HIV cores undergo uncoating (see Fig. 386-3) to release virion nucleic acid into the cytoplasm. Uncoating represents a critical checkpoint for innate antiviral activity that can strongly

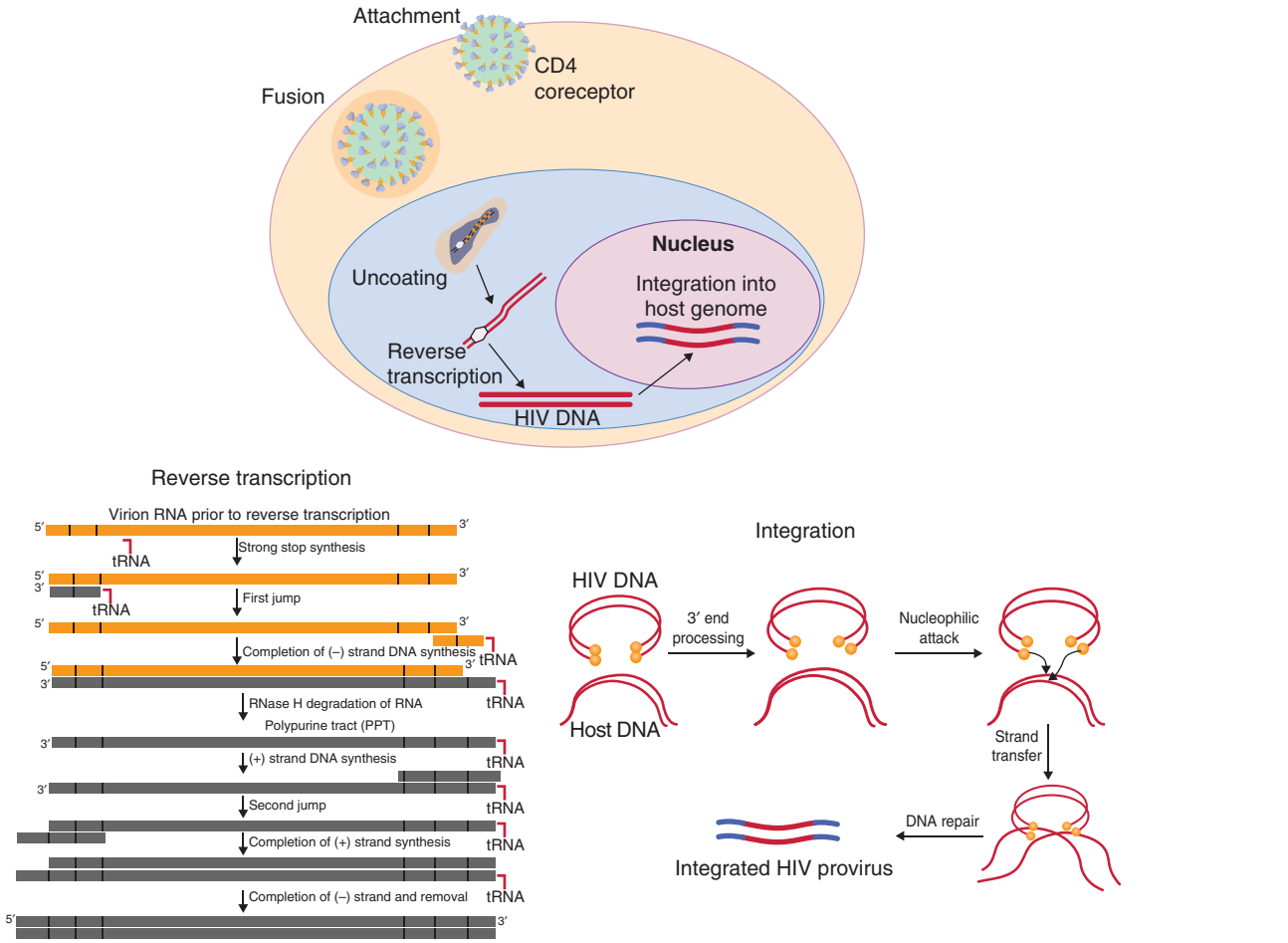

FIGURE 386-3. Early events in HIV replication include attachment, fusion, uncoating, reverse transcription, and integration into the host genome. Details of HIV reverse transcription and integration are depicted.

restrict infectivity.[7,8] An interferon-inducible host protein, TRIM5-alpha, can block uncoating of a number of viruses through interactions with the viral core; unfortunately, TRIM5-alpha cannot restrict either HIV-1 or HIV-2. TRIM5-alpha is under substantial genetic selection, with one of the fastest rates of positive selection of any human gene. Studies suggest that the current version of TRIM5-alpha present in human populations may have been selected during a previous epidemic to protect against an ancient retroviral infection. In this model, simply summarized as "generals are always fighting the last war, especially if they have won it," the current TRIM5-alpha was selected in the past to prevent a retroviral infection. Unfortunately, the current TRIM5-alpha is unable to restrict HIV and uncoating takes place unabated in human cells.

Reverse Transcription: Viral Genome Replication

Following uncoating, HIV has fresh access to nucleoside triphosphates, permitting reverse transcription to take place in the cytoplasm (see Fig. 386-3).[9] The (+)-strand template strands of HIV RNA are complexed with reverse transcriptase and a specific $tRNA^{lys}$ that functions as a primer located at a specific primer binding site (see E-Fig. 386-1) near the 5′ end of the template RNA and HIV RNA. Reverse transcription is a multistep process (see Fig. 386-3) that first synthesizes a DNA copy of the RNA genome and then excises the RNA from the RNA-DNA hybrid using an RNase H function of reverse transcriptase; RNA removal is incomplete, and residual RNA in a region denoted the polypurine tract (see Fig. 386-3) then serves to prime the next round of DNA synthesis. During reverse transcription a number of strand transfer events occur, permitting opportunities for frequent recombination. Reverse transcriptase is highly error prone, with only a rudimentary editing function. As a result, complete reverse transcription yields at least one mutation per virion synthesized per replication cycle. The replication cycle for HIV, estimated as 1 to 2 days, is relatively short, and the replicating population size is substantial. The combination of rapid and error-prone synthesis in a large replicating population results in a genetically diverse population that can respond rapidly to immune- or drug-selective pressure. Thus, rapid error-prone replication, combined with recombination, represents an important pathogenic determinant for HIV.

Reverse transcription was the first target for antiretroviral therapy, and a number of direct and allosteric inhibitors of reverse transcription have been developed. All reverse transcription inhibitors inhibit RNA-dependent DNA synthesis or DNA-dependent DNA synthesis. Nucleoside and nucleotide reverse transcriptase inhibitors are dideoxy analogues of deoxynucleotides that are incorporated as the template is copied and act as chain terminators, blocking additional nucleic acid synthesis. Non-nucleoside inhibitors of reverse transcription bind to reverse transcriptase in a hydrophobic domain proximal to the active site, deforming the enzyme structure and disrupting nucleic acid synthesis.

Nuclear Transport and Integration

Newly synthesized HIV DNA transports to the nucleus for integration into the host genome. A complex of the HIV protein integrase, newly synthesized HIV DNA, and associated proteins has been labeled an *intasome*. Structural studies have revealed a tetramer of integrase molecules bound to the ends of the retroviral DNA, bringing the DNA ends in close proximity, poised for integration into the genome in a multistep process (see Fig. 386-3). Integration is a highly successful target for antiretroviral therapy and inhibitors that block the transfer of HIV DNA strand into the host genome are now in clinical practice.

Early Evasion of Intracellular Immunity: Vif and Vpu

HIV infection triggers a complex set of immune responses, and several viral factors function to counteract innate and adaptive immunity. As described earlier, HIV is able to infect human cells because an otherwise effective restriction mechanism blocking uncoating by TRIM5-alpha is unable to detect the incoming virus. Infection results in activation of two additional interferon-induced genes, the ABOBEC family of nucleic acid editing enzymes and BST-2 (tetherin). HIV in turn encodes specific functions to counteract such antiviral mechanisms.

The interferon-induced APOBEC family (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) of proteins are nucleic acid editing enzymes capable of catalyzing the removal of amino groups from the cytosine portion of cytidine; on copying, these deaminated cytidines base pair with adenosine, not guanine, and the net result is the introduction of multiple G-to-A mutations, resulting in hypermutation and complete viral inactivation. APOBEC can, in the absence of viral factors, be incorporated into new virions. As a result, during the next round of infection, after HIV enters a new host cell, APOBEC proceeds to hypermutate the newly reverse transcribed HIV genome. To block APOBEC-mediated inactivation of HIV, a viral gene product, Vif (see E-Fig. 386-1), directly binds to APOBEC proteins, redirecting it to degradation by ubiquitination pathways. Redirecting APOBEC for elimination rather than virion incorporation effectively neutralizes a potent innate immune response.

A second product of interferon induction is the bone marrow stromal antigen 2 (BST-2), CD317, or tetherin. BST-2 is tethered to the plasma membrane at both its amino and carboxy terminus and is enriched at virion budding. Tetherin can block the budding process directly; newly budding virions may contain one end of the tetherin molecule in the virion membrane, while the other end of tetherin remains on the cellular membrane, thereby effectively blocking virion release. To counter this host mechanism, the HIV-encoded protein Vpu effectively blocks tetherin by a number of mechanisms, including direct binding and redirecting the protein to intracellular degradation.

Late Steps in Replication

Transcription and Translation: Exploiting Cellular Processes to Balance Production of Viral Gene Products

Once the proviral state is established, HIV produces viral RNA and proteins using viral factors in concert with cellular mechanisms of transcription and translation. Thus, HIV replicates not by dismantling cell functions but rather by employing specific interactions between host factors, which ensures a balanced abundance of viral gene products.

Transcription

The integrated provirus is expressed in the context of host chromatin. The U3 portion of the HIV long terminal repeat contains binding sites for transcription factors common in lymphocytes and macrophage-monocyte lineages, including activator protein 1 (AP-1), specificity protein 1 (SP-1), nuclear factor kappa B (NF-κB), and nuclear factor of activated T cells (NFAT) binding sites. The presence of the viral transcription factor transactivator of transcription (Tat)[10,11] markedly stimulates transcription through binding to a specific RNA enhancer, denoted the transactivating region (TAR), and recruiting additional transcription factors (Fig. 386-4).

Post-transcriptional Processing

Retroviruses transcribe all of their genes from a single promoter. Full-length HIV RNA is processed for expression of all nine gene products and encapsidated as the viral genome into virions. To provide sufficient mRNA species to express all HIV proteins, RNA processing is highly regulated by alternative splicing and differential RNA transport (see Fig. 386-4).

Full-length RNA consists of a primary capped mRNA of approximately 9.2 kb that undergoes alternative splicing to produce three broad classes of RNA species: 9.2-kb unspliced RNA, responsible for translation of *gag/pol;* a number of distinct 4.5-kb singly spliced RNA species, responsible for translation of *vif, vpr,* and v*pu/env;* and 1.8-kb multiply spliced species, responsible for *tat, rev,* and *nef.* Multiply spliced 1.8-kb RNAs are constitutively expressed, but unspliced and singly spliced HIV mRNAs require specific transport out of the nucleus. The mechanism of retention in the nucleus is unclear, but the presence of cis regulatory sequences present in *gag/pol* and *env* result in nuclear retention. To export unspliced and singly spliced HIV RNAs, the viral protein Rev binds to a specific region in the *env* portion of the unspliced and singly spliced RNA, denoted as the Rev responsive element (RRE); the RRE is a sequence of c.240 nt that folds into a specific structure to which Rev binds. The Rev-RNA complex then engages the nuclear host CRM-1 transport apparatus, which transports RNA out of the nucleus. Engaging appropriate host elements therefore results in a balance of 9.2-, 4.5-, and 1.8-kb mRNA species.[12]

Translation of mRNA

All HIV RNAs are translated by cellular ribosomes on either smooth or rough endoplasmic reticulae. Abundance of several HIV proteins is regulated by translational mechanisms and post-translational modifications mediated by host mechanisms that are critical for viral protein function.

Full-length HIV RNA serves as RNA for encapsidation into the virion, and for Gag/Pol synthesis. Gag is synthesized as a 55-kd polyprotein precursor in relatively abundant amounts; Gag protein is N-myristoylated, permitting the protein to bind cellular membranes. The *pol* gene products are translated

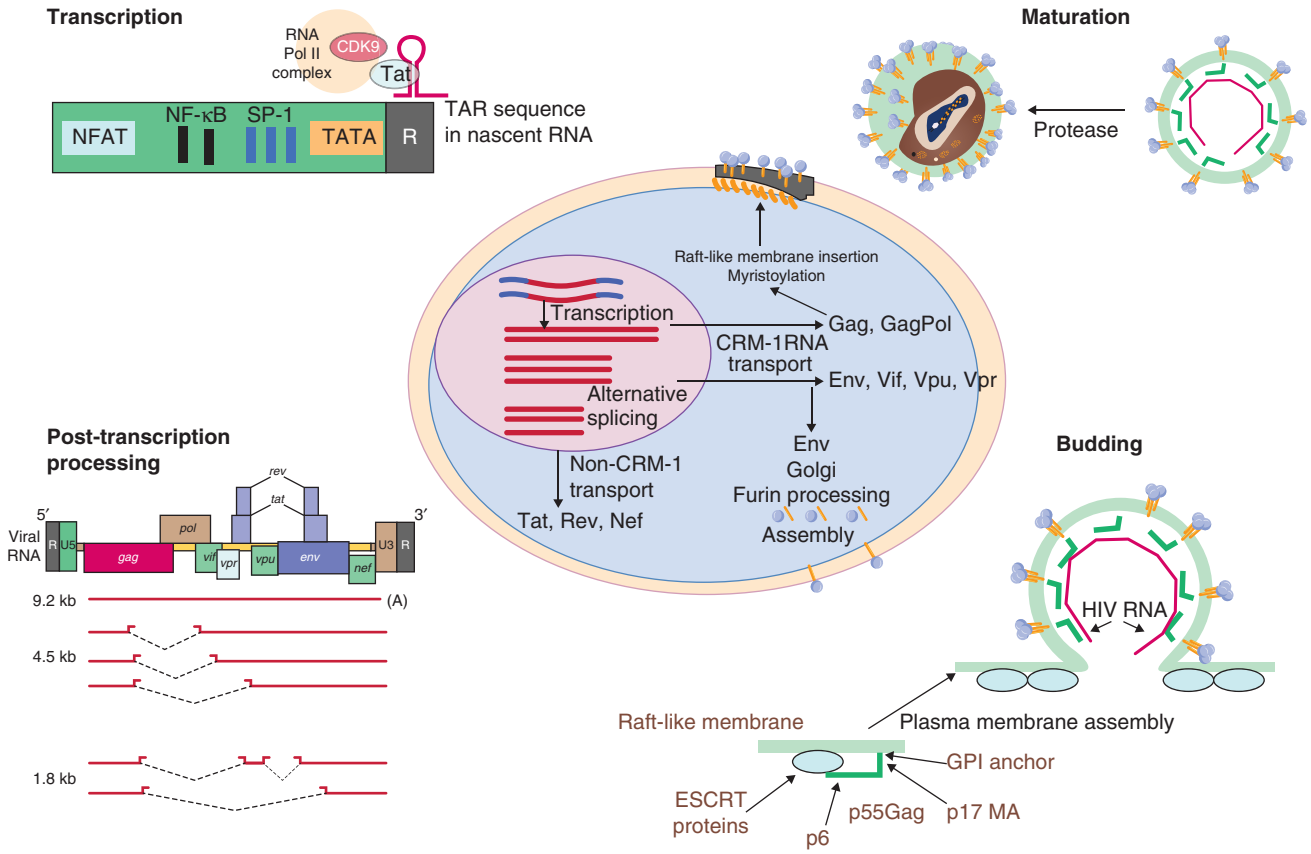

FIGURE 386-4. Late events in HIV replication. Following establishment of the provirus, transcription, post-transcriptional processing translation, virion assembly, budding, and virion maturation take place by co-opting cellular processes.

though frameshifting mechanisms.[13] Frameshifting is relatively inefficient, and the relative abundance of *pol* enzyme gene products is substantially lower than Gag proteins, effectively controlling the levels of enzymes in favor of an abundance of structural proteins.

HIV Vpu and Env proteins are translated from a 4.5-kb bicistronic mRNA, permitting effective production of both Vpu and Env proteins. As Vpu sequesters CD4 by direct protein-protein interactions on intracellular membranes (see later), it prevents CD4-Env interactions that would arrest Env within the cell. Thus, the coordinated expression of Vpu with Env ensures efficient expression of Env. Env is synthesized as a gp160 precursor of Env SU and TM, cotranslationally inserted into the lumen of rough endoplasmic reticulum membranes and glycoslylated predominantly at a number of canonical N glycosylation sites.

Multiply spliced mRNA species encoding *tat, rev,* and *nef* are translated earliest after infection; Tat and Rev are transported to the nucleus, where they activate transcription, and transport (see earlier). Nef serves a number of cytoplasmic functions in evasion of host immunity by binding MHC-1 and CD4, redirecting them away from plasma membranes. Thus, at an early time in the infectious cycle, immune molecules that help identify and target virus-infected cells are downregulated from the cell surface of infected cells, facilitating virus replication.

Transport and Assembly of Virion Components: An Elegant Dance

The late steps in virus replication are complex, but dissecting the individual steps has identified a number of critical interactions between host and virion components that represent active areas for development of useful therapeutics.

HIV virion proteins and virion RNA species traffic to the plasma membrane, where virion assembly takes place. Nascent Gag and GagPol polyprotein precursors undergo cotranslational modifications that target Gag to membranes.[14] Gag has self-assembly properties, but correct initial assembly of the virion core into a hexagonal lattice includes incorporation of virion RNA and additional factors. Virion RNA undergoes dimerization largely through direct RNA-RNA interactions that require specialized RNA sequences, denoted psi sequences, at the 5′ region of the genome favoring incorporation of only unspliced HIV RNA species. Gag polyprotein precursor can bind HIV RNA directly, providing a potential mechanism for transport to the cell membrane. Gag also accumulates in specialized membrane microdomains enriched with sphingomyelin-saturated phospholipids and cholesterol typical of lipid rafts and binds components of the endosomal sorting complexes required for transport (ESCRT) pathway (see Fig. 386-4). The ESCRT pathway is normally involved in intracellular membrane remodeling and scission events necessary for events such as organelle biogenesis, lysosome formation, and cytokinesis; by recruiting ESCRT to sites of HIV assembly on the plasma membrane, HIV engages a highly specialized pathway to execute budding. Elegant studies of HIV transmission have suggested these events may also participate in cell-cell transmission of HIV at "virologic synapses," specialized areas of cell-cell contact.

During the budding process, HIV undergoes several maturation events, including proteolytic processing of the Gag and GagPol precursor proteins by HIV protease (see Fig. 386-4). Protease is embedded as part of the GagPol precursor, cleaves itself from the precursor and then proceeds to process Gag and GagPol into component proteins. Processing is essential for virion infectivity, and protease inhibitors are highly effective agents in HIV therapy. During processing, the core lattice begins to bend by introducing pentamers of CA at critical points of the hexagonal lattice, effectively folding the structure into a fullerene-like cone that encapsidates dimeric HIV RNA-tRNA^{lys} and HIV enzymes. Proteolytic cleavage at the CA-SP1 site is critical; the development of maturation inhibitors that block cleavage by binding Gag instead of inhibiting protease represent a new antiviral target.

HIV Env precursor glycoprotein is processed by a cellular furin-like protease into mature products, gp120 (SU) and gp41, which traffic to plasma membranes through Golgi and post-Golgi membranes. Because the cellular receptor CD4 is also processed through similar mechanisms, intracellular Env-CD4 binding can effectively block Env from reaching the cell surface. The HIV Vpu protein, which directly binds and redirects CD4 to proteolytic degradation, thereby increases the proportion of Env reaching the cell surface and sites of HIV budding.

SUMMARY

By the completion of the infectious cycle, HIV has produced progeny for the next round of infection, obstructed adaptive and innate immune responses,

and established a proviral state in the infected cell. Despite multiple mechanisms of immune evasion, the majority of infected cells (>99.9%) die within 1 to 2 days of infection. Thus, the virus has a strikingly short period to engage critical cell pathways to complete replication, while the cells are undergoing destruction. A minority of cells survive infection and persist for prolonged periods. As a consequence, current antiretroviral therapy that targets active steps in HIV replication does not eradicate HIV infection. Mechanisms of persistence remain poorly understood, but likely include transcriptional and immunologic mechanisms.[15] Additional research will be essential to determine mechanisms of persistence and to identify new strategies to cure HIV infection.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. HIV/AIDS Data and Statistics. 2014. http://www.who.int/hiv/data/en/. Accessed January 15, 2015.
2. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med*. 2011;1:a006841.
3. Cortez KJ, Maldarelli F. Clinical management of HIV drug resistance. *Viruses*. 2011;3:347-378.
4. Sundquist WI, Krausslich HG. HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med*. 2012;2:a006924.
5. Wilen CB, Tilton JC, Doms RW. HIV: cell binding and entry. *Cold Spring Harb Perspect Med*. 2012;2:a006866.
6. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services. 2014 http://aidsinfo.nih.gov/contentfiles/lvguidelines/AdultandAdolescentGL.pdf; Accessed January 15, 2015.
7. Ambrose Z, Aiken C. HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. *Virology*. 2014;454-455:371-379.
8. Strebel K. HIV accessory proteins versus host restriction factors. *Curr Opin Virol*. 2013;3:692-699.
9. Hu WS, Hughes SH. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med*. 2012;2:a006882.
10. Van Lint C, Bouchat S, Marcello A. HIV-1 transcription and latency: an update. *Retrovirology*. 2013;10:67.
11. Karn J, Stoltzfus CM. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med*. 2012;2:a006916.
12. Leblanc J, Weil J, Beemon K. Posttranscriptional regulation of retroviral gene expression: primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA. *Wiley Interdiscip Rev RNA*. 2013;4:567-580.
13. Brakier-Gingras L, Charbonneau J, Butcher SE. Targeting frameshifting in the human immunodeficiency virus. *Expert Opin Ther Targets*. 2012;16:249-258.
14. Sundquist WI, Krausslich HG. HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med*. 2012;2:a006924.
15. Bullen CK, Laird GM, Duran CM, et al. New ex vivo approaches distinguish effective and ineffective single agents for reversing HIV-1 latency in vivo. *Nat Med*. 2014;20:425-429.

REVIEW QUESTIONS

1. Which of the following animals represents the *most likely* source of HIV-1?

A. Sooty mangabey
B. Chimpanzee
C. Gorilla
D. Macaques
E. Mandrill

Answer: B HIV-1 is responsible for over 30 million infections worldwide and is closely related to a simian immunodeficiency virus (SIV) present in chimpanzee populations. Although gorillas have an SIV, that virus is more distantly related to HIV-1 than the chimpanzee virus. Sooty mangabeys have SIV that is very closely related to HIV-2 but not to HIV-1. Macaques and mandrills are species of monkeys with SIV variants that are only distantly related to HIV.

2. Which of the following inhibits HIV infection by introducing numerous G-to-A mutations in the HIV genome?

A. Tetherin
B. APOBEC
C. TRIM5-alpha
D. Vif

Answer: B APOBEC is a cytidine deaminase and one part of the innate immune system that inactivates virus by introducing multiple G-to-A changes, which result in multiple mutations and stop codons. Tetherin is a cell surface protein that retains virions at the cell surface and is antagonized by VPU. TRIM5-alpha is involved in blocking uncoating before reverse transcription, and Vif is a viral protein that abrogates Apobec activity by excluding it from incorporation into the virion.

3. HIV-2 is naturally resistant to which of the following agents?

A. Non-nucleoside reverse transcriptase inhibitors (NNRTI)
B. Nucleoside reverse transcriptase inhibitors (NRTI)
C. Protease inhibitors (PI)
D. Integrase inhibitors (INSTI)

Answer: A HIV-2 is only distantly related to HIV-1 and encodes a reverse transcriptase that is naturally resistant to non-nucleoside reverse transcriptase inhibitors. HIV-2 remains susceptible to NRTI, PI, and integrase inhibitors. It is critical that individuals with HIV-2 or HIV-1/2 *not* be treated with NNRTI, because mutants resistant to the other components of the regimen will emerge.

4. HIV virion assembly takes place:

A. At the plasma membrane.
B. Within the cytoplasm.
C. At the nuclear membrane
D. Within the nucleus.

Answer: A HIV assembles at the plasma membrane; no virus-like particles are seen in the cytoplasm or at or within the nucleus.

5. HIV protease inhibitors inhibit which of the following steps in HIV replication?

A. Proteolytic processing of HIV Gag and Gag/pol precursors
B. Proteolytic processing of Env precursor
C. Proteolytic processing of the HIV core during virion uncoating
D. Proteolytic processing of the intasome before integration

Answer: A Protease inhibitors are a key component of combination antiretroviral therapy. They block a late step in virus replication after virions have assembled. The immature virus is not infectious until after protease acts on Gag and Gag/Pol.

387 PREVENTION OF HUMAN IMMUNODEFICIENCY VIRUS INFECTION

CARLOS DEL RIO AND MYRON S. COHEN

More than 30 years have passed since the first report of a case of human immunodeficiency virus (HIV) infection, and the pandemic has spread worldwide and infected more than 70,000,000 people, of whom approximately 35,000,000 have died as a consequence of acquired immunodeficiency virus (AIDS). HIV prevention efforts have been "front and center" since the virus was discovered as the cause of AIDS, as summarized in Figure 387-1. Behavioral interventions focused on HIV-negative persons have likely played a role in the falling population level incidence in some countries reported by the United Nations Program in HIV/AIDS (UNAIDS) in their 2012 report; however, approximately 2.7 million new infections still occur each year, and we have made little progress in reducing HIV incidence in the groups at highest risk. In the past few years, several promising new prevention strategies have demonstrated efficacy in clinical trials and are now being implemented,[1] leading President Obama to look toward an "AIDS-free generation." This chapter will provide a detailed view of HIV prevention approaches that are useful to the practicing clinician.

MODES OF TRANSMISSION AND PREVENTION

Sexual Transmission

The primary mode of HIV transmission throughout the world is sexual contact. However, the geographic distribution of cases attributable to homosexual or heterosexual transmission varies markedly. In the United States, most sexually transmitted cases of HIV are observed in men who have sex with men (MSM), and heterosexual transmission accounts for a smaller number of new infections except among women. However, heterosexual transmission is the leading mode of transmission worldwide and remains the primary mode of disease acquisition in Africa. Sexual transmission of HIV is relatively inefficient, but behavioral and biologic factors influence the likelihood of HIV transmission in a given sexual encounter. In particular, coinfection with classical sexually transmitted infections (STIs) (especially genital ulcerative diseases such as herpes simplex) greatly increases the infectiousness and the susceptibility of an individual. STIs increase the concentration of HIV in genital secretions, which increases the likelihood for transmission.

The risk for acquisition of HIV per coital act has been estimated to be 5 per 10,000 for insertive unprotected penile-vaginal intercourse to 50 per 10,000 for receptive unprotected anal intercourse. However, the risk is not stable and varies depending on the stage of infection and other amplifying cofactors. HIV transmission risk is highest in early HIV infection and in advanced infection (Fig. 387-2), demonstrating that the viral concentration in the genital secretions is the strongest predictor of the risk for transmission.

Prevention Strategies

Traditional strategies for the prevention of sexual transmission of HIV have focused on encouraging abstinence, reducing unsafe sexual behaviors (especially unprotected anal intercourse and concurrent relationships), encouraging proper condom use, and treating STIs. These interventions primarily focus on HIV-negative persons.

In situations in which a decision to engage in sexual activity has been made and the HIV status of the partner is positive, unknown, or in doubt, safe sexual practices ("safe sex") should be implemented. Consistent use of latex condoms has been shown to be effective in preventing HIV transmission at both individual and population levels. The condom should be made of latex and must be used properly. Natural skin condoms should not be used because they do not prevent transmission of HIV. Petroleum-based lubricants enhance the likelihood of rupture of latex condoms and should be avoided. If needed, water-based lubricants such as K-Y jelly should be used.

The effectiveness of condoms in preventing heterosexual transmission of HIV has been estimated to be 87%, but it may be as low as 60% or as high as 96%. The effectiveness of condoms during anal intercourse is probably lower because the frequency of condom breakage and slippage may be considerably higher than during vaginal intercourse. Circumcision represents an alternative strategy to protect men from HIV. Three randomized clinical trials have demonstrated a protective benefit of male circumcision, with the risk for acquisition of HIV infection through heterosexual intercourse decreasing by approximately 60%, with increasing reduction in HIV acquisition over time. However, this benefit has not been confirmed for MSM.

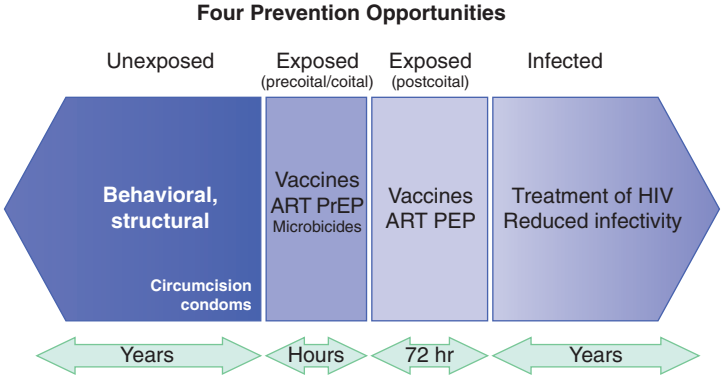

FIGURE 387-1. Opportunities for HIV prevention. ART = antiretroviral therapy; PEP = postexposure prophylaxis; PrEP = pre-exposure prophylaxis. (Modified from Cohen MS. Recent developments in HIV prevention. *AIDS* 2008, Abstract TUPL0102.)

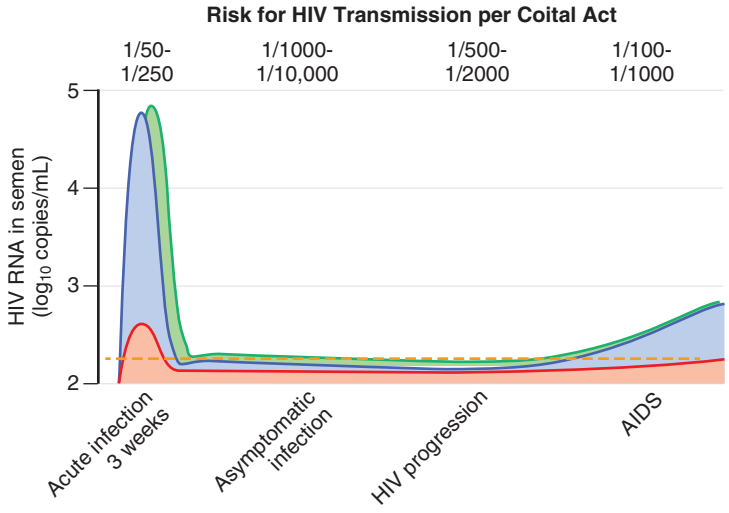

FIGURE 387-2. Prediction of the efficiency of HIV transmission according to HIV burden in the genital tract. Probability of male-to-female HIV transmission per coital act, as a function of HIV disease stage in the index case. Dashed line = a potential threshold for HIV transmission; orange = theoretical effect of a biologic intervention designed to reduce viral excretion; blue = expected distribution of viral burden in semen among men over time. (Modified from Cohen MS, Pilcher CD. Amplified HIV transmission and new approaches to HIV prevention. *J Infect Dis.* 2005;191:1391-1393.)

Antiretroviral therapy may influence infectivity and the subsequent risk for transmission through sexual contact. In the HPTN 052 study, early antiretroviral therapy of HIV-infected patients beginning at a CD4 count of 350 to 550 cells/mm^3 reduced sexual transmission of HIV-1 by 96%. Observational studies suggest that broader antiretroviral use at a population level can reduce HIV incidence. Antiretrovirals also are effective for the prevention of HIV when administered prophylactically to HIV-uninfected at-risk individuals as pre-exposure prophylaxis (PrEP). In a randomized trial, daily administration of co-formulated tenofovir plus emtricitabine as pre-exposure prophylaxis of MSM decreased the risk for HIV infection by 44%. Other studies using tenofovir with emtricitabine show a decrease of as much as two thirds in infections,[2] although findings have varied across studies dependings on adherence. Based on these studies, in 2013 the U.S. Food and Drug Administration approved the use of co-formulated tenofovir plus emtricitabine for pre-exposure prophylaxis to prevent sexual transmission of HIV. The Centers for Disease Control and Prevention (CDC) has issued guidelines for clinicians for pre-exposure prophylaxis with antiretroviral drugs for the prevention of HIV-infection in the United States.[3] Similarly, a tenofovir-containing vaginal gel led to a 39% reduction in the risk for HIV infection among women, but a tenofovir vaginal gel is not currently available commercially. In addition, antiretrovirals can be given after exposure to prevent HIV acquisition (postexposure prophylaxis [PEP]). The CDC and professional organizations have published guidelines for antiretroviral postexposure prophylaxis after sexual, injection drug use, and other nonoccupational exposures to HIV.[4] In these guidelines it is recommended that persons seek care 72 hours or sooner after nonoccupational exposure to the blood, genital secretions, or other potentially infected body fluids of a person known to have HIV infection be offered a 28-day course of antiretroviral therapy.

Transmission in Injection Drug Users

The primary mode of HIV transmission in injection drug users is sharing of contaminated needles and syringes. Sharing of injection paraphernalia ("works") is commonplace among injection drug users and is reinforced by the cultural, economic, and legal environment in that community. The risk for transmission of HIV is highest in injection drug users who share needles and use drugs that are injected more frequently, such as cocaine or methamphetamines.

Prevention Strategies

The primary mode of preventing HIV transmission in PWID is to stop the use of intravenous drugs. Education programs that are culturally sensitive and geared to young audiences have the best chance of preventing drug use. Access to treatment centers for injection drug users is the best approach. However, approximately 80% of active drug users in the United States are not in substance abuse treatment because of either choice or the unavailability of treatment centers. For injection drug users who do not wish to seek treatment or who are unable to gain access to treatment, the most effective way to prevent HIV infection is to avoid sharing needles and paraphernalia. Some communities have adopted programs that provide free needles and syringes for injection drug users, and there is strong evidence that these programs, when implemented properly, are effective in reducing HIV transmission and do not result in increased drug use among participants. Where supplies cannot be obtained, needles and syringes should be cleaned after each use, preferably with readily accessible virucidal cleansers such as chlorine bleach (diluted 1 : 10). As with sexual transmission, preliminary data from Vancouver, Canada, suggest that antiretroviral treatment of injection drug users decreases incidence of HIV among drug users. Most recently, a randomized controlled trial conducted in injection drug users in Thailand demonstrated a 49% reduction in HIV acquisition when tenofovir was given as PrEP. As a result the CDC issued guidance for the use of PrEP among injection drug users but has recommended the use of co-formulated tenofovir plus emtricitabine rather than tenofovir alone as the preferred PrEP regimen among injection drug users.[5] However, this is an off-label indication.

Transmission through Blood Products and Other Tissues

HIV has been transmitted through the transfusion of single-donor blood and blood products, including whole blood, fresh-frozen plasma, packed red blood cells, cryoprecipitate, clotting factors, and platelets. Confidential donor exclusion, as well as the institution of HIV antibody screening in 1985, followed by additional testing for antibodies to HIV-2 and p24 antigen in 1996 and nucleic acid testing in 2002, has reduced the risk for HIV infection through the transfusion of blood or blood products to approximately 1 in 2,135,000. Transmission of HIV by liver, heart, kidney, pancreas, bone, and possibly skin transplantation has been reported. In contrast, relatively avascular tissues such as corneas and processed tissues have not been associated with transmission.

Prevention Strategies

The institution of HIV antibody testing of donated blood and blood products in 1985 has had the most dramatic effect on lowering the incidence of transfusion-related transmission. When combined with voluntary self-deferral and nucleic acid testing, the blood supply in most countries has become virtually free of HIV. Heat inactivation processes for cryoprecipitate and clotting factor concentrates has eliminated transmission of HIV through use of these products. Other products, such as immunoglobulin preparations and hepatitis B vaccines, are produced by fractionation methods that remove HIV and have never been associated with transmission of HIV. Organ and tissue donors should be evaluated and serologically screened in a manner similar to blood donors. In addition, donations of semen and bone from a living donor may be quarantined until subsequent testing has definitively ruled out the possibility of delayed seroconversion in the donor.

Transmission to Health Care Workers

Detailed studies examining the risk associated with specific exposures to health care workers, such as needlestick injuries and mucous membrane exposure, have demonstrated low risk for acquisition of disease in the workplace. More than 3628 health care workers have been prospectively examined in carefully designed surveillance studies at 10 high-incidence medical centers. The overall risk for seroconversion after a percutaneous needlestick from a known HIV-infected source is 0.3% per exposure. A retrospective study conducted by the CDC found that the risk for transmission of HIV to health care workers is increased when the device causing the injury is visibly contaminated with blood, when the device has been used for insertion into a vein or artery, when the device causes a deep injury, or when the source patient dies within 2 months after the exposure. Exposure of mucous membranes to HIV-infected blood has resulted in seroconversion only rarely, and the risk for transmission is estimated to be 0.09% per exposure.

Prevention Strategies

In August 1987, the CDC published guidelines recommending that the principles of "universal precautions" be incorporated into health care settings to minimize exposure of health care workers to blood and body fluids that may be infected with blood-borne pathogens such as HIV. Universal precautions are based on the premise that any patient may be infected with blood-borne infectious agents and it may be difficult, if not impossible, to differentiate those with infection from their uninfected counterparts. All specimens containing blood or blood-tinged fluids obtained from *any* patient should be considered hazardous and handled as such. The use of universal precautions helps minimize the transmission of many transmissible diseases in addition to HIV.

Gowns, protective eyewear, and masks are not usually needed except in circumstances in which splattering or splashing of blood-containing fluids is likely to occur. Health care workers with denuded skin, open lesions, or active dermatitis should avoid direct patient contact and should not process contaminated equipment or materials. Handling of sharp instruments ("sharps") represents the greatest risk for transmission of HIV to health care workers. Although injuries from sharps cannot be eliminated entirely, the number of exposures can be reduced substantially by adhering to guidelines put forth in universal precautions. Before a sharp instrument is used, thought should be given to where the instrument will be disposed after use. Impervious containers should be readily available in all patient care areas and identified by the health care worker *before* the use of sharps. These containers should be checked frequently and should not be allowed to overfill. Used needles should never be manipulated, bent, broken, or recapped. Recapping of needles is the single most common activity that results in needlestick injuries. Technologic developments that do not rely on health care worker compliance, such as self-sheathing needles, also have been important in decreasing the risk for needlestick injury.

Antiretroviral agents are also used for PEP in health care workers. A case-control study suggested that risk for HIV seroconversion after occupational exposure was decreased by approximately 81% with the use of zidovudine alone. Subsequent recommendations have incorporated the newer antiretroviral drugs, as well as risk stratification for the type of exposure, in the management of occupational exposure to HIV. In 2013, the CDC issued revised

recommendations for the use of PEP after exposure to HIV among health care workers.[6] The essential elements of management of a health care worker after a needlestick or mucous membrane exposure include appropriately evaluating the donor (patient) and recipient (health care worker) at the time of exposure, counseling of the health care worker, and providing follow-up HIV testing. There is no longer the need to determine the severity of exposure to direct the number of antiretroviral drugs to be used; a regimen containing three or more drugs is now recommended with co-formulated tenofovir/emtricitabine plus raltegravir as the preferred regimen. Antiretroviral drugs for PEP should be initiated within 72 hours of exposure and continued for 4 weeks. The U.S. Public Health Service has established a National Clinicians' Post-Exposure Prophylaxis Hotline (PEPline) to provide expert consultation about the management of health care workers with potential HIV exposure. The PEPline can be accessed at 1-888-448-4911 or through the Internet at http://www.ucsf.edu/hivcntr/Hotlines/PEPline.

PREVENTION INTERVENTIONS FOR INFECTED INDIVIDUALS

Antiretroviral therapy administered to HIV-infected individuals decreases the risk for HIV transmission by over 96%, and thus suppression of HIV replication to undetectable levels is the most effective intervention to decrease HIV transmission from an HIV-infected person. As a result, the U.S. Department of Health and Human Services Antiretroviral treatment guidelines now recommend initiation of antiretroviral treatment regardless of the CD4 cell count and monitoring for viral suppression.

In addition, the CDC, Health Resources and Services Administration, National Institutes of Health, and Infectious Diseases Society of America have published joint recommendations for incorporating HIV prevention into the HIV medical care setting. These guidelines reflect four basic priorities: (1) screening for risky behavior and STIs; (2) providing general and tailored risk reduction messages to patients; (3) when indicated, referring patients for additional risk reduction services and other services that may affect risk reduction (e.g., substance abuse treatment); and (4) ensuring that patients are provided with partner counseling and referral services. HIV-infected persons should be screened and treated for STIs. The CDC's 2010 sexually transmitted disease treatment guidelines recommend that all patients with newly diagnosed HIV infection undergo screening for gonorrhea, chlamydial infection, hepatitis B and C virus infection, and syphilis. Screening for curable STIs (gonorrhea, chlamydial infection, and syphilis) should be performed at least annually in sexually active patients. More frequent screening for STI's might be appropriate depending on individual risk behaviors, the local epidemiology of STIs and whether incident STIs are detected by screening or by presence of symptoms.

PREVENTION IN DEVELOPMENT

As shown in Figure 387-1, many prevention strategies are available, and some are in development. Vaccine strategies are increasingly focused on the generation of different types of antibodies. Broad neutralizing antibodies offer almost compete protection in macaques from infection for several months, suggesting the feasibility of a protective vaccine.[7] Injectable antiretroviral agents in different classes may serve as an advance for PEP and PreP, and such agents are in phase 2 testing. Delivering of slow-release topical tenofovir from a cervical ring is being examined in two clinical trials.

Grade A References

A1. Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med*. 2011;365:493-505.

A2. Grant RM, Lama JR, Anderson PL, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *N Engl J Med*. 2010;363:2587-2599.

A3. Baeten JM, Donnell D, Ndase P, et al. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *N Engl J Med*. 2012;367:399-410.

A4. Thigpen MC, Kebaabetswe PM, Paxton LA, et al. Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana. *N Engl J Med*. 2012;367:423-434.

A5. Choopanya K, Martin M, Sutharasami P, et al. Antiretroviral prophylaxis for HIV infection among people who inject drugs in Bangkok, Thailand: a randomized, double-blind, placebo-controlled trial. *Lancet*. 2013;381:2083-2090.

A6. Abdool Karim Q, Abdoool Karim SS, Frohlich JA, et al. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science*. 2010;329:1168-1174.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Dieffenbach CW, Fauci AS. Universal voluntary testing and treatment for prevention of HIV transmission. *JAMA*. 2009;301:2380-2382.
2. Grant RM, Anderson PL, McMahan V, et al. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. *Lancet Infect Dis*. 2014;14:820-829.
3. Centers for Disease Control and Prevention. Pre-exposure prophylaxis for the prevention of HIV infection in the United States—2014: a clinical practice guideline. http://www.cdc.gov/hiv/pdf/guidelines/PrEPguidelines2014.pdf; Accessed March 14, 2015.
4. Marrazzo JM, del Rio C, Holtgrave DR, et al. HIV prevention in clinical care settings: 2014 recommendations of the International Antiviral Society-USA Panel. *JAMA*. 2014;312:390-409.
5. Centers for Disease Control and Prevention. Update to Interim guidance for preexposure prophylaxis (PrEP) for the prevention of HIV infection: PrEP for injection drug users. *MMWR Morb Mortal Wkly Rep*. 2013;62:463-465.
6. Kuhar DT, Henderson DK, Struble KA, et al. Updated US Public Health Service guidelines for the management of occupational exposures to human immunodeficiency virus and recommendations for postexposure prophylaxis. *Infect Control Hosp Epidemiol*. 2013;34:875-892.
7. Excler JL, Rubb ML, Kim JH. HIV-1 vaccines: challenges and new perspectives. *Hum Vaccin Immunother*. 2014;10:1734-1746.

REVIEW QUESTIONS

1. After occupational exposure by a health care worker to HIV through a needlestick injury, management should include which of the following?

A. The HIV status of the exposure source patient should be determined.
B. Postexposure prophylaxis should contain no more than two antiretrovirals.
C. Antretroviral therapy should be started within 72 hours of exposure and continued for 4 weeks
D. All of the above are correct.
E. Only a and c and correct.

Answer: E In managing health care personnel who experience occupational exposure to blood and/or body fluids that might contain HIV, the USPHS guidelines published in 2013 recommend that after occupational exposure to HIV occurs the HIV status of the exposure source be determined (if possible). Postexposure prophylaxis should be started as soon as possible (within 72 hours) and continued for 4 weeks and three or more antiretrovirals should be prescribed.

2. The most effective strategy for the prevention of sexual transmission of HIV from an infected to an uninfected partner is:

A. Voluntary HIV counseling and testing.
B. Antiretroviral therapy of the infected partner.
C. Antretroviral therapy of the uninfected partner.
D. Prompt treatment of sexually transmitted infections.
E. Avoidance of oral sex.

Answer: B Antiretroviral therapy to bring HIV replication below the limits of detection has proved to be an effective strategy to decrease by 96% the risk for HIV transmission among serodiscordant couples.

3. Confidential donor exclusion, HIV antibody screening, testing for antibodies to HIV-2 and p24 antigen, and nucleic acid testing has reduced the risk for HIV infection through the transfusion of blood or blood products to:

A. Approximately 1 in 20,000.
B. Approximately 1 in 200,000.
C. Approximately 1 in 2,000,000.
D. Approximately 1 in 20,000,000.

Answer: C HIV has been transmitted through the transfusion of single-donor blood and blood products, including whole blood, fresh-frozen plasma, packed red blood cells, cryoprecipitate, clotting factors, and platelets. Confidential donor exclusion, as well as the institution of HIV antibody screening in 1985, followed by additional testing for antibodies to HIV-2 and p24 antigen in 1996 and nucleic acid testing in 2002, has reduced the risk for HIV infection through the transfusion of blood or blood products to approximately 1 in 2,135,000.

4. The effectiveness of condoms in preventing heterosexual transmission of HIV has been estimated to be:

A. As low as 60% and as high as 96% (~87%).
B. Higher for anal than for vaginal sex.
C. Equal for latex and natural skin condoms.
D. Enhanced by petroleum-based lubricants.

Answer: A The effectiveness of condoms in preventing heterosexual transmission of HIV has been estimated to be 87%, but it may be as low as 60% or as high as 96%. Consistent use of latex condoms has been shown to be effective in preventing of HIV transmission at both individual and population levels. The condom should be made of latex and must be used properly. Natural skin condoms should not be used because they do not prevent transmission of HIV. Petroleum-based lubricants enhance the likelihood of rupture of latex condoms and should be avoided. If needed, water-based lubricants such as K-Y jelly should be used.

5. HIV transmission risk is highest in:

A. Patients taking antiretrovirals.
B. Patients with a sexually transmitted infection such as genital wart.
C. Early HIV infection and advanced infection.
D. An episode of a concomitant viral infection such as influenza.

Answer: C The risk for HIV transmission is highest in early HIV infection and in advanced infection, demonstrating that the HIV viral concentration in the genital secretions is the strongest predictor of the risk for transmission.

388

ANTIRETROVIRAL THERAPY OF HUMAN IMMUNODEFICIENCY VIRUS AND ACQUIRED IMMUNODEFICIENCY SYNDROME

ROY M. GULICK

The development of effective antiretroviral therapy (ART) for human immunodeficiency virus (HIV) infection is one of the most notable achievements in modern medicine. The first cases of acquired immunodeficiency syndrome (AIDS) were reported from Los Angeles in 1981. In the early to mid-1980s, without any available antiretroviral treatments, the life expectancy of an individual diagnosed with AIDS was only approximately 6 to 12 months. The first antiretroviral drug, zidovudine (azidothymidine, AZT) was approved by the U.S. Food and Drug Administration (FDA) in 1987 on the basis of a short-term survival benefit. Triple-drug therapy was first introduced in the mid-1990s and resulted in a two-thirds decrease in HIV-related deaths within 2 years in developed countries. Today, there are a total of 28 antiretroviral drugs that are approved by the FDA and three-drug combination regimens are the standard of care. The benefits of ART were extended to developing countries, and an estimated over 14 million people currently are taking ART worldwide. The life expectancy of an HIV-infected individual appropriately treated with ART is now estimated to be nearly that of the general population, both in developed[1] and developing countries.[2]

WHEN TO START ART?

The rationale for starting ART early in someone with HIV infection is based on several principles: (1) untreated infection causes progressive immunodeficiency resulting in opportunistic diseases and death; (2) current ART regimens decrease plasma HIV RNA (viral load) levels and the risk for the emergence of drug resistance, as well as increase CD4+ T-lymphocyte counts and general immune function; and (3) current treatment results in years of virologic suppression. In addition, it is now appreciated that ART reduces inflammation and immune activation that leads to end-organ diseases (cardiac, hepatic, neurologic, oncologic, and renal) and that ART reduces HIV transmission. The rationale for delaying ART traditionally centered on practical factors: medication adherence can be challenging; drug toxicities occur, and the long-term side effects of antiretroviral drugs are not known; and the risk for clinical progression is low in early disease. Additionally, although ART can prevent HIV transmission, drug-resistant virus can be transmitted in the community. In fact, the U.S. Centers for Disease Control and Prevention estimates at least 15% of Americans newly diagnosed with HIV are infected with drug-resistant virus.[3]

Whereas before 2008, U.S. treatment guidelines recommended waiting to start antiretroviral treatment, current guidelines from both the U.S. Department of Health and Human Services and the International Antiviral Society–USA recommend starting ART in *all* HIV-infected patients, regardless of CD4 cell count number, because of both clinical benefits to the patient and reduction in HIV transmission to others.[4,5] This recommendation is supported by the fact that current ART regimens are potent, convenient, and generally well-tolerated, as well as by supportive clinical cohort data[6] and randomized controlled clinical trials. There are no randomized clinical trial data to support starting ART in patients with CD4 counts greater than 550 cells/μL, and thus some controversy in the field remains, as evidenced by British, European, and World Health Organization treatment guidelines that do not routinely recommend ART in patients with CD4 cell counts greater than 500/μL (Table 388-1).

WHAT TO START?

Since 1987, the FDA has approved 28 antiretroviral drugs from six mechanistic classes (Table 388-2). The goal of ART is to suppress viral replication, prevent the emergence of drug-resistant viral strains, enhance immunologic responses, decrease clinical events, and prolong healthy life. Antiretroviral drugs interfere with individual steps in the HIV replication cycle (Fig. 388-1). The first step of the life cycle is HIV entry, a three-step process

TABLE 388-1 WHEN TO START ANTIRETROVIRAL THERAPY

ANTIRETROVIRAL TREATMENT GUIDELINE AND YEAR (REFERENCE)	AIDS/ SYMPTOMS	ASYMPTOMATIC			
		CD4 <200	CD4 200-349	CD4 350-499	CD4 ≥500
U.S. Department of Health and Human Services (DHHS) 2014 http://www.aidsinfo.nih.gov	Yes	Yes	Yes	Yes	Yes
International Antiviral Society–USA 2014 *JAMA*. 2014;312:390	Yes	Yes	Yes	Yes	Yes
British HIV Association (BHIVA) 2013 *HIV Med.* 2014;15(suppl 1):1-85 http://www.bhiva.org	Yes	Yes	Yes	Certain patients	Certain patients
European AIDS Clinical Society (EACS) 2014 http://www.europeanaidsclinicalsociety.org/	Yes	Yes	Yes	Certain patients	Certain patients
World Health Organization (WHO) 2013 http://www.who.int/hiv/pub/guidelines/arv2013/en/index.html	Yes	Yes	Yes	Yes	No

TABLE 388-2 ANTIRETROVIRAL DRUGS

MECHANISTIC DRUG CLASS	GENERIC NAME	ABBREVIATION(S)	TRADE NAME	YEAR OF U.S. FDA APPROVAL
HIV NUCLEOSIDE ANALOGUE REVERSE TRANSCRIPTASE INHIBITORS (NRTIs)				
	zidovudine	ZDV, AZT	Retrovir	1987
	didanosine	ddI	Videx	1991
	zalcitabine	ddC	Hivid	1992*
	stavudine	d4T	Zerit	1994
	lamivudine	3TC	Epivir	1995
	abacavir	ABC	Ziagen	1998
	tenofovir	TDF	Viread	2001
	emtricitabine	FTC	Emtriva	2003
HIV NON-NUCLEOSIDE ANALOGUE REVERSE TRANSCRIPTASE INHIBITORS (NNRTIs)				
	nevirapine	NVP	Viramune	1996
	delavirdine	DLV	Rescriptor	1997
	efavirenz	EFV	Sustiva	1998
	etravirine	ETR	Intelence	2008
	rilpivirine	RPV	Edurant	2010
HIV PROTEASE INHIBITORS (PIs)				
	saquinavir	SQV	Invirase	1995
	ritonavir	RTV	Norvir	1996
	indinavir	IDV	Crixivan	1996
	nelfinavir	NFV	Viracept	1997
	amprenavir	APV	Agenerase	1999*
	lopinavir/ritonavir	LPV/r	Kaletra	2000
	atazanavir	ATV	Reyataz	2003
	fosamprenavir	FPV	Lexiva	2003
	tipranavir	TPV	Aptivus	2005
	darunavir	DRV	Prezista	2006
HIV ENTRY INHIBITORS (EIs)				
Fusion inhibitor	enfuvirtide	ENF, T-20	Fuzeon	2003
CCR5 antagonist	maraviroc	MVC	Selzentry	2007
HIV INTEGRASE INHIBITORS (INSTIs)				
	raltegravir	RAL	Isentress	2007
	elvitegravir	EVG	Viteka	2011
	dolutegravir	DTG	Tivicay	2013

*Withdrawn.
U.S. FDA = U.S. Food and Drug Administration; HIV = human immunodeficiency virus.

starting with binding of the HIV external membrane glycoprotein (gp120) to the CD4 receptor on the surface of the T lymphocyte. This binding induces a conformational change in gp120 permitting the second step of HIV entry, binding to a second cellular receptor, the chemokine receptor, either CCR5 (bound by R5 viruses) or CXCR4 (bound by X4 viruses). Some HIV strains are dual-tropic and can bind to either the CCR5 or CXCR4 receptor, and some individuals are infected with a mixed population of R5 and X4 viral strains. Binding to the chemokine receptor induces an additional conformational change allowing the viral protein gp41 to pierce the target cell membrane and then fold in on itself in a coil-on-coil interaction, leading to fusion of the viral and cellular membranes and extrusion of the viral particle contents (viral RNA, viral proteins) into the cytoplasm of the cell.

Inside the cell, viral RNA is transcribed to viral DNA by a viral-specific enzyme called *HIV reverse transcriptase*. Following transcription, two strands of viral DNA form a double-stranded complex catalyzed by a second viral-specific enzyme called *HIV integrase,* which also promotes transport of the viral DNA complex into the nucleus of the cell and strand transfer—the random integration of the viral DNA into the cellular genome. At the point of viral DNA integration, the cell is infected for life. It may enter a latent period, with a cellular lifespan as long as 60 years, or may be activated to transcribe both cellular and viral DNA into RNA and then translate RNA into proteins, including viral proteins, which assemble at the surface of the cell and then bud off into new viral particles. After budding, a third viral-specific enzyme, *HIV protease,* cleaves viral precursor proteins, a step that is necessary

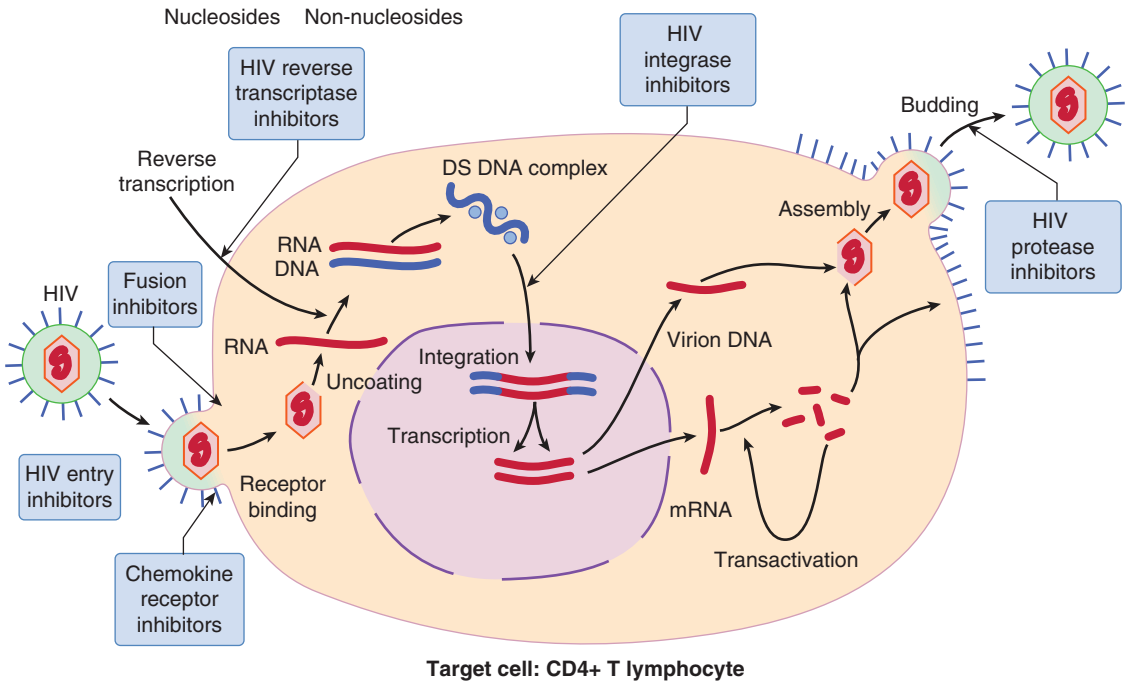

FIGURE 388-1. Life cycle of human immunodeficiency virus (*HIV*) and mechanisms of action of the six antiretroviral drug classes. See text (What to Start? section) for details.

for viral maturation and infectivity. One infected cell can produce hundreds to thousands of viral particles, and many will be capable of starting the process again on encountering another CD4+ T lymphocyte.

Antiretroviral drugs inhibit different steps of the HIV replication cycle (see Fig. 388-1). Two kinds of HIV entry inhibitors target the first step in the HIV life cycle, viral entry, by inhibiting either CCR5 chemokine receptor binding (CCR5 antagonists) or membrane fusion (fusion inhibitors). The nucleoside analogue reverse transcriptase inhibitors (NRTIs) target the viral-specific enzyme *HIV reverse transcriptase.* A second class of HIV reverse transcriptase inhibitors are the non-nucleoside analogue reverse transcriptase inhibitors (NNRTIs), which bind to a different part of the same enzyme. The integrase strand transfer inhibitors (INSTIs) inhibit the viral-specific enzyme *HIV integrase* by specifically targeting the transfer of viral DNA to the host cell genome. The HIV protease inhibitors (PIs) bind to the active site of the *HIV protease* enzyme and prevent precursor protein cleavage, viral maturation, and infectivity.

Initially, single nucleoside analogue therapy was studied for the treatment of HIV infection. However, virologic suppression, immune enhancement, and clinical benefits were only transient and drug-resistant virus emerged (Fig. 388-2). When more than one drug was available, dual nucleoside analogue therapy was studied and found to be better than single-drug therapy, but virologic, immunologic, and clinical benefits again were only temporary. A three-drug regimen was superior to two-drug therapy and led to essentially complete virologic suppression, preventing the emergence of drug-resistant viral strains and increasing CD4 cell counts, and resulted in durable effects. With potent virologic suppression and immunologic recovery came dramatic decreases in HIV-related illnesses and prolonged survival.

Despite their life-saving potential, early antiretroviral drug regimens had significant associated side effects and toxicities and were complicated to take. Some early drugs were associated with anemia and leukopenia (zidovudine), peripheral neuropathy (didanosine, zalcitabine, stavudine), pancreatitis (didanosine, stavudine), kidney stones (indinavir), nausea and vomiting (ritonavir), diarrhea (nelfinavir), and Stevens-Johnson syndrome (nevirapine). A common three-drug regimen in 1996 consisted of 20 pills, divided for dosing every 8 hours. Whereas the first decade of ART developed effective regimens that controlled viral replication, the second decade was about developing potent, well-tolerated, and convenient regimens to enable long-term use. Use of low-dose ritonavir or the newer drug cobicistat inhibits the cytochrome P450 3A4 isoenzyme, decreasing the metabolism of most protease inhibitors and the integrase inhibitor elvitegravir, allowing once- or twice-daily dosing; this strategy is known as pharmacokinetic boosting. Additional antiretroviral drugs were approved, and co-formulations (more than one

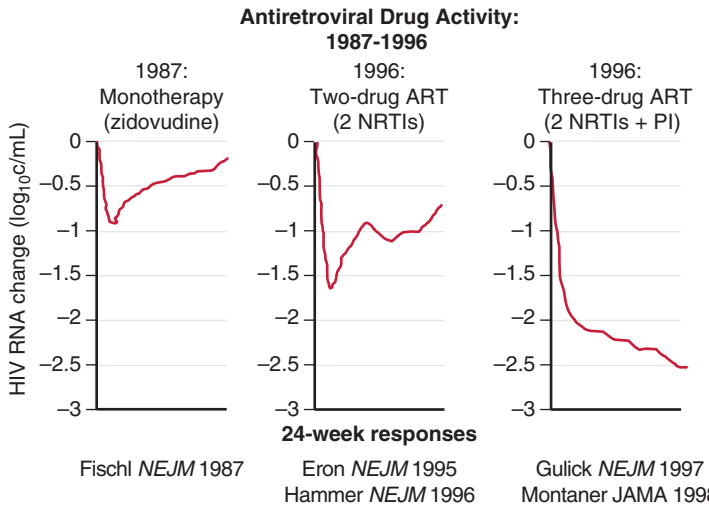

FIGURE 388-2. Antiretroviral drug (*ART*) activity: 1987 to 1996. Major clinical trials leading to finding of sustained efficacy of three-drug antiretroviral therapy. NRTI = nucleoside analogue reverse transcriptase inhibitors; PI = protease inhibitor. (1987: Fischl MA, et al. *N Engl J Med.* 1987;317:185-191; 1994: Eron JJ, et al. *N Engl J Med.* 1995;333:1662-1669 and Hammer SM, et al. *N Engl J Med.* 1996;335:1081-1090; 1996: Gulick RM, et al. *N Engl J Med*. 1997; 337:734-739 and Montaner JS, et al. 1998;279: 930-937.)

medicine in a single pill) were developed such that one-pill, once-daily HIV treatment was possible (Table 388-3).

Currently, there are four FDA-approved, one-pill, once-daily regimens; these are popular with providers and patients and lead to excellent adherence, with virologic suppression rates over 85% in some groups. Potency, convenience, and tolerability are essential qualities of current antiretroviral drug regimens and account for the durable clinical benefits. Current recommended initial regimens in the U.S. ART guidelines (Table 388-4) are combinations of two nucleoside analogues together with a third drug (non-nucleoside reverse transcriptase inhibitor, protease inhibitor, or integrase inhibitor). Globally, the most widely used antiretroviral regimen is two nucleoside analogues combined with a non-nucleoside analogue (NNRTI). A study from Denmark estimated that the life expectancy of someone with HIV infection who is appropriately treated and is free from comorbidities (hepatitis C infection, injection drug use) is now that of the age-matched general population.

TABLE 388-3 ANTIRETROVIRAL FIXED-DOSE COMBINATIONS

DRUG CLASS(ES)	GENERIC NAMES	ABBREVIATION(S)	TRADE NAMES	DOSING	YEAR OF U.S. FDA APPROVAL
2 NRTI	zidovudine + lamivudine	ZDV/3TC	Combivir	Twice daily	1997
3 NRTI	abacavir + zidovudine + lamivudine	ABC/ZDV/3TC	Trizivir	Twice daily	2000
Boosted PI	lopinavir + ritonavir	LPV/RTV	Kaletra	Once or twice daily	2000
2 NRTI	tenofovir + emtricitabine	TDF/FTC	Truvada	Once daily	2004
2 NRTI	abacavir + lamivudine	ABC/3TC	Epzicom	Once daily	2004
2 NRTI + NNRTI	tenofovir + emtricitabine + efavirenz	TDF/FTC/EFV	Atripla	Once daily	2006
2 NRTI + NNRTI	tenofovir + emtricitabine + rilpivirine	TDF/FTC/RPV	Complera	Once daily	2011
2 NRTI + boosted INSTI	tenofovir + emtricitabine + elvitegravir + cobicistat	TDF/FTC/EVG/c	Stribild	Once daily	2012
2 NRTI + INSTI	abacavir + lamivudine + dolutegravir	ABC/3TC/DTG	Triumeq	Once daily	2014

NNRTI = non-nucleoside reverse transcriptase inhibitor; NRTI = nucleoside analogue reverse transcriptase inhibitor; PI = protease inhibitor.

TABLE 388-4 RECOMMENDED INITIAL ANTIRETROVIRAL DRUG REGIMENS

REGIMEN*	DRUGS
Non-nucleoside (NNRTI)-based	tenofovir/emtricitabine/efavirenz (co-formulated) tenofovir/emtricitabine (co-formulated) + atazanavir/ritonavir[†] abacavir/emtricitabine + efavirenz[‡] tenofovir/emtricitabine/rilpivirine (co-formulated)[§]
Protease inhibitor (PI)-based	tenofovir/emtricitabine (co-formulated) + darunavir/ritonavir[†] abacavir/lamivudine + atazanavir/ritonavir[†]
Integrase strand transfer inhibitor (INSTI)-based	tenofovir/emtricitabine (co-formulated) + raltegravir tenofovir/emtricitabine/elvitegravir/cobicistat (co-formulated) tenofovir/emtricitabine + dolutegravir abacavir/lamivudine + dolutegravir

*From Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services. http://aidsinfo.nih.gov/contentfiles/lvguidelines/AdultandAdolescentGL.pdf. November 13, 2014.
[†]Ritonavir and cobicistat are each used as pharmacokinetic boosters.
[‡]Only for baseline HIV RNA <100,000 copies/ml.
[§]Only for baseline HIV RNA <100,000 copies/ml and CD4 >200 cells/uL.

WHEN TO CHANGE ANTIRETROVIRAL THERAPY?

In a stable patient on ART, HIV RNA should be monitored every 3 to 6 months. Even with virologic suppression rates of over 85%, some patients will experience antiretroviral treatment failure, most commonly presenting as virologic failure, with a once-suppressed viral load level now being detectable above the limit of the HIV RNA assay (20, 40, or 50 copies/mL). The reasons for regimen failure can include suboptimal adherence, baseline drug resistance or cross-resistance, prior use of ART, use of less potent antiretroviral drug regimens, drug levels and drug-drug interactions, penetration into tissue reservoirs (e.g., genital tract, central nervous system [CNS]), suboptimal provider experience, and other, unknown reasons. Also, more than one factor may play a role in an individual patient. One of the challenges with treatment failure is to try to determine the cause of failure and then select the next regimen that can address and overcome the reason for failure.

U.S. ART guidelines suggest focusing on virologic failure and changing antiretroviral regimens promptly when failure is confirmed. Virologic failure can be defined as repeated detection of HIV RNA levels in a drug-adherent patient. HIV RNA levels suppressed below the level of detection of the assay are unlikely to result in the emergence of drug-resistant viral strains. Levels greater than 200 copies/mL (and certainly 500 copies/mL) will lead to the selection of drug-resistant viral strains and result in treatment failure. More controversial are HIV RNA levels between the level of detection and 200 copies/mL that may represent a higher virologic set point rather than ongoing viral replication and may not necessitate treatment change.

HIV is prone to errors in gene replication, and thus an individual is infected not with just one virus but a "swarm" of related viral strains with distinct genetic mutational patterns. Drug resistance is conferred by viral strains with specific amino acid substitutions in viral proteins (HIV reverse transcriptase, HIV protease, HIV integrase) that are selected for in the presence of an antiretroviral drug. For example, with ongoing replication in the presence of the nucleoside analogue emtricitabine or lamivudine, a viral strain with the substitution of valine for methionine at amino acid position 184 (M184V) of HIV reverse transcriptase will be selected and confer complete resistance to these drugs. Drugs with which single substitutions confer resistance are considered to have a low barrier to resistance and include the nucleoside analogues emtricitabine and lamivudine, the NNRTIs efavirenz and nevirapine, and the integrase inhibitors elvitegravir and raltegravir. Drugs that require multiple substitutions are considered to have a high barrier to resistance, including the nucleoside analogue zidovudine and most of the HIV protease inhibitors. Drugs with overlapping resistance patterns lead to cross resistance. For example, a patient who develops resistance to emtricitabine with the M184V substitution will have complete cross resistance to lamivudine, even though the patient never took lamivudine. Drug resistance can be assessed by an HIV genotype (that identifies amino acid substitutions that must be correlated with drug resistance) or an HIV phenotype (that assesses viral growth in the presence of each of the drugs).[7]

Another common clinical conundrum is immunologic failure—a patient taking ART who achieves virologic suppression but fails to increase the CD4 cell count. The causes of immunologic failure are not clear, but associations with various factors, including CD4 count less than 200 cells/μL at the time of ART initiation, older age, coinfections, medications (e.g., zidovudine; the combination of didanosine and tenofovir), persistent immune activation, and loss of regenerative potential have been reported. There is no accepted treatment for immunologic failure. Neither changing nor adding antiretroviral drugs results in an improved CD4 cell response. Immune-based therapies have been studied but are not effective. In two large randomized clinical trials, interleukin-2 was associated with increased CD4 cell counts, but failed to demonstrate associated clinical benefits. Current management is to continue antiretroviral therapy, optimize opportunistic infection prophylaxes, and follow the patient closely.

WHAT ANTIRETROVIRAL THERAPPY TO CHANGE TO?

The current goal for all HIV-infected patients taking ART, regardless of treatment experience, is to maximally suppress the viral load level below the level of assay detection.[4] In a treatment-experienced patient, this is done by reviewing the ART history with a focus on adherence, tolerability, and possible drug-drug interactions; conduct drug-resistance testing (HIV genotype for first or second failure; HIV genotype and phenotype for more advanced failure); identify susceptible drugs and drug classes and, ultimately, design a subsequent regimen with at least two (and preferably three) fully active antiretroviral agents.[4]

In the last 10 years, a number of newer antiretroviral drugs have made this goal possible, including drugs in existing mechanistic classes with activity against class-resistant virus (the HIV protease inhibitors darunavir and tipranavir and the HIV NNRTI etravirine) and drugs with new mechanisms of action (the fusion inhibitor enfuvirtide; the CCR5 antagonist maraviroc; and the HIV integrase inhibitors raltegravir, elvitegravir, and dolutegravir). Several recent studies in treatment-experienced patients show that new active antiretroviral regimens result in virologic suppression rates that are nearly the same as those for treatment-naïve patients.[8] A recent study found that including nucleoside analogues in subsequent regimens is not necessary if the regimen contains more than two active antiretroviral drugs.

SIDE EFFECTS AND TOXICITY

Antiretroviral drugs (like all drugs) are associated with side effects and toxicity.[4] Probably the most common side effect of ART as a group is gastrointestinal (nausea, vomiting, diarrhea), although some drugs are more associated than others (e.g., zidovudine, ritonavir). Toxicities may be divided according to seriousness and drug classes. Life-threatening toxicities occur and include drug-related hepatitis associated with NNRTIs and protease inhibitors. Of these, the NNRTI nevirapine is unique in causing drug-related hepatitis more frequently in patients with higher CD4 counts (>250 cells/μL in women and > 400 cells/μL in men), likely due to an immunologic mechanism. A hypersensitivity reaction characterized by rash and constitutional symptoms is associated with the nucleoside analogue abacavir and the NNRTIs etravirine and nevirapine. An elegant study linked the abacavir-associated hypersensitivity reaction to a genetic locus that can be screened for in patients (HLA-B*5701, at a cost of approximately $50); if the drug is avoided in patients with the genetic marker, the risk for HSR is minimized. Lactic acidosis has been associated with the nucleoside analogue class (particularly stavudine and zalcitabine). The NNRTIs, although structurally unrelated, all are associated with rash; Stevens-Johnson syndrome is rarely described with etravirine or nevirapine. Teratogenicity is described with efavirenz (FDA pregnancy category D); however, newer U.S. recommendations allow continuing the drug in a pregnant woman with maximal virologic suppression, although the drug should not be started in a pregnant woman.[9]

Acute side effects can be troubling to the patient and can lead to suboptimal adherence. Providers should be in close contact with a patient starting a new antiretroviral regimen and have a low threshold to substitute offending drugs for side effects or toxicities. As noted, probably the most common side effect of antiretroviral drugs is gastrointestinal toxicity, though this often can be managed by taking ART with food. Zidovudine causes anemia, neutropenia, and fatigue. Efavirenz causes CNS side effects (e.g., vivid dreams, somnolence) in up to 50% of people and should be dosed at bedtime. Atazanavir causes increased indirect bilirubin by inhibiting uridine 5′-diphosphoglucuronosyltransferase, which can be associated with frank jaundice but is not associated with other liver test abnormalities.

With the current expectation that ART is life-long, chronic and cumulative toxicities also are important. Increased cardiovascular events are associated with some protease inhibitors[10] and controversially with abacavir. Indinavir and atazanavir cause renal stones. Metabolic changes, including hyperglycemia and frank diabetes, hyperlactatemia, and/or hyperlipidemia, are associated with stavudine and some protease inhibitors. Morphologic changes occur and can be very distressing for patients, including lipoatrophy (loss of fat in the face and extremities) associated with stavudine and zidovudine, and lipoaccumulation (gain of fat in the breasts, abdomen, and dorsocervical fat bad [buffalo hump]) associated with some protease inhibitors. Some NRTIs, including didanosine and stavudine, cause a progressive toxic peripheral neuropathy. Tenofovir is associated with proximal renal tubular dysfunction characterized by hypophosphatemia, proteinuria, glycosuria, and eventually, elevated creatinine. Tenofovir also has been associated with loss of bone mineral density over the first year of treatment that appears to stabilize thereafter. Newer antiretroviral drugs and investigational antiretroviral agents often are developed and selected for less toxicity.

SPECIAL POPULATIONS

Acute Infection

Previously debated, it is now recommended that an individual identified with acute HIV infection start three-drug ART. ART reduces signs and symptoms of acute HIV and also prevents ongoing HIV transmission. Because of the risk for acquiring drug-resistant HIV, ART should be started while awaiting the results of the HIV genotype. Many experts would start a protease inhibitor–containing regimen and then adjust when the genotypic results are available.

Acute Opportunistic Infection

A patient presenting with an acute opportunistic infection who was not previously known to have HIV infection prompts the question of the optimal time to start ART. Despite prior concerns of drug-drug interactions and precipitating the immune reconstitution inflammatory syndrome (IRIS) (Chapter 395), ART demonstrates benefits in patients with acute opportunistic infections. One study of patients with a treatable opportunistic infection diagnosed in the prior 2 weeks (the majority with *Pneumocystis* pneumonia) randomized them to start ART within 48 hours or to wait at least 4 weeks. There were significantly fewer clinical events (disease progression and death) in the group that started ART earlier. Additional studies of patients with tuberculosis also demonstrated clinical benefits to starting ART earlier, particularly in patients with CD4 counts less than 50 cells/μL. Starting ART within 2 weeks of an opportunistic infection is now considered the standard of care. An exception is CNS opportunistic infections (e.g., cryptococcal or tuberculous meningitis) in which studies demonstrated increased mortality in patients who started ART earlier.[11,12]

Coinfection with Hepatitis B

If treatment is started for either infection, both need to be treated optimally: two active drugs for hepatitis B and three active drugs for HIV. The antiretroviral drugs emtricitabine, lamivudine, and tenofovir have activity against both viruses; thus, a suitable regimen to treat both infections would be tenofovir, emtricitabine (or lamivudine), and a third antiretroviral drug. Stopping drugs with activity against hepatitis B may result in a serious hepatitis flare.

Coinfection with Hepatitis C

The optimal time to treat hepatitis C infection in a patient with HIV infection is not known. The hepatitis C drug ribavirin has significant interactions with the antiretroviral drugs didanosine, stavudine, and zidovudine, and these combinations should be avoided. The newer hepatitis C protease inhibitors are unrelated to the HIV protease inhibitors and have no activity against HIV but may have significant drug interactions with HIV NNRTIs and protease inhibitors, and this is an active area of research.

Pregnancy

The recently updated U.S. Perinatal Treatment Guidelines recommend ART for prevention of mother-to-child transmission of HIV for all pregnant women, regardless of CD4 cell count or HIV RNA level. Based on their safety records, recommended drugs in pregnancy include the NRTIs lamivudine and zidovudine; the NNRTI nevirapine; and the PIs atazanavir, lopinavir, and ritonavir. Other drugs are considered alternatives (e.g., the NRTIs abacavir and tenofovir and the PI darunavir), although insufficient data are available on newer drugs (the NNRTIs etravirine and rilpivirine, and the CCR5 antagonist maraviroc). Efavirenz is teratogenic and should not be started in pregnant women; however, updated guidelines recommend continuing efavirenz in a women found to be pregnant if she is maximally virologically suppressed.

ANTIRETROVIRAL THERAPY FOR PREVENTION

One of the most successful HIV prevention strategies is the use of ART, both in HIV-infected and HIV-uninfected people. The first example of this was the prevention of HIV-infected mother-to-child transmission by giving the mother ART. In a classic study, the use of a single drug, zidovudine, in an HIV-infected mother reduced the risk for transmission to her infant from 25 to 8%. Current standard of care is to treat the mother with three-drug ART, with a resultant reduction in the risk for transmission to less than 1%.

A recent study randomized HIV-infected individuals with CD4 counts of 350 to 550/μL who were members of a committed couple with an HIV-uninfected individual to start ART immediately or to wait until the CD4 count decreased to less than 250 cells/μL and followed the seronegative partners for HIV seroconversion. Of 28 linked cases, 27 occurred in the group not on ART versus only 1 in an individual on ART (who had only recently started); thus, treating the infected individual with ART was associated with a 96% reduction in transmission to the HIV-uninfected partner.

Giving ART to at-risk HIV-uninfected individuals to avoid infection also is a strategy that has been explored. Postexposure prophylaxis is recommended on the basis of an older case-control study of health care workers exposed to HIV in which taking zidovudine was associated with an 81% decrease in the risk for seroconversion compared to no prophylaxis. The U.S. Centers for Disease Control and Prevention (CDC) recommends administering three-drug ART for 4 weeks following significant exposure (occupational or nonoccupational) to HIV.[13,14]

Pre-exposure prophylaxis (PrEP)[15] is a newer strategy in which one- or two-drug ART is given to HIV-uninfected individuals who are at risk for acquiring HIV infection on the basis of published studies in men who have sex with men,[A12] serodiscordant heterosexual couples,[A13] sexually active heterosexuals,[A14] and, most recently, injection drug users.[A15] Notably, two other PrEP studies in African women failed to show a benefit, although this may have been due to suboptimal adherence.[A16] Current U.S. CDC guidance recommends PrEP for high-risk individuals with two-drug ART with tenofovir-emtricitabine (coformulated) following exclusion of acute or chronic HIV infection, ongoing monitoring of HIV status and renal function, screening for sexually transmitted diseases, and risk reduction counseling and condom distribution.

CURE

Although current ART is highly effective, it is not curative. Soon after infection, HIV establishes a latent-CD4+ T-lymphocyte cell reservoir that can persist for an estimated 60 years. Even with years of ART-induced prolonged virologic suppression, the latent-cell reservoir does not decrease significantly. Strategies of intensification of ART by changing or adding additional antiretroviral drugs have not been successful in decreasing the reservoir. In 2009, the first known cure of HIV infection was reported: a 40 year-old man with HIV infection well controlled on ART developed acute myelogenous leukemia and underwent radiation, cytotoxic chemotherapy, and ultimately a bone marrow transplant from a donor with a deletion in the gene that codes for the CCR5 receptor that is required for HIV entry.[16] The patient's course was complicated by transplant rejection, administration of antirejection medications, cessation of ART, and a second bone marrow transplant. Ultimately, his HIV RNA remained suppressed below detection in the plasma in the absence of ART, his CD4 cell count returned to normal levels, and additional intensive investigations failed to identify replication-competent virus. After more than 5 years off of ART, he is considered to be cured. This case report is intriguing in that the cause of the cure is not clear (radiotherapy, chemotherapy, the CCR5-negative transplants, graft rejection, antirejection medications, or a combination of these), but the fact that it occurred is stimulating an intense research agenda for HIV cure.

Grade A References

A1. Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med*. 2011;365:493-505.

A2. Severe P, Juste MA, Ambroise A, et al. Early versus standard antiretroviral therapy for HIV-infected adults in Haiti. *N Engl J Med*. 2010;363:257-265.

A3. Cohen CJ, Andrade-Villanueva J, Clotet B, et al. Rilpivirine versus efavirenz with two background nucleoside or nucleotide reverse transcriptase inhibitors in treatment-naïve adults infected with HIV-1 (THRIVE): a phase 3, randomised, non-inferiority trial. *Lancet*. 2011;378:229-237.

A4. Molina JM, Cahn P, Grinsztejn B, et al. Rilpivirine versus efavirenz with tenofovir and emtricitabine in treatment-naïve adults infected with HIV-1 (ECHO): a phase 3 randomised double-blind active-controlled trial. *Lancet*. 2011;378:238-246.

A5. DeJesus E, Rockstroh JK, Henry K, et al. Co-formulated elvitegravir, cobicistat, emtricitabine, and tenofovir disoproxil fumarate versus ritonavir-boosted atazanavir plus co-formulated emtricitabine and tenofovir disoproxil fumarate for initial treatment of HIV-1 infection: a randomised, double-blind, phase 3, non-inferiority trial. *Lancet*. 2012;379:2429-2438.

A6. Sax PE, DeJesus E, Mills A, et al. Co-formulated elvitegravir, cobicistat, emtricitabine, and tenofovir versus co-formulated efavirenz, emtricitabine, and tenofovir for initial treatment of HIV-1 infection: a randomised, double-blind, phase 3 trial, analysis of results after 48 weeks. *Lancet*. 2012;379:2439-2448.

A7. Mallal S, Phillips E, Carosi G, et al. HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med*. 2008;358:568-579.

A8. Zolopa A, Andersen J, Powderly W, et al. Early antiretroviral therapy reduces AIDS progression/death in individuals with acute opportunistic infections: a multicenter randomized strategy trial. *PLoS ONE*. 2009;4:e5575.

A9. Abdool Karim SS, Naidoo K, Grobler A, et al. Integration of antiretroviral therapy with tuberculosis treatment. *N Engl J Med*. 2011;365:1492-1501.

A10. Blanc FX, Sok T, Laureillard D, et al. Earlier versus later start of antiretroviral therapy in HIV-infected adults with tuberculosis. *N Engl J Med*. 2011;365:1471-1481.

A11. Havlir DV, Kendall MA, Ive P, et al. Timing of antiretroviral therapy for HIV-1 infection and tuberculosis. *N Engl J Med*. 2011;365:1482-1491.

A12. Grant RM, Lama JR, Anderson PL, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *N Engl J Med*. 2010;363:2587-2599.

A13. Baeten JM, Donnell D, Ndase P, et al. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *N Engl J Med*. 2012;367:399-410.

A14. Thigpen MC, Kebaabetswe PM, Paxton LA, et al. Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana. *N Engl J Med*. 2012;367:423-434.

A15. Choopanya K, Martin M, Suntharasamai P, et al. Antiretroviral prophylaxis for HIV infection in injecting drug users in Bangkok, Thailand (the Bangkok Tenofovir Study): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet*. 2013;381:2083-2090.

A16. Van Damme L, Corneli A, Ahmed K, et al. Preexposure prophylaxis for HIV infection among African women. *N Engl J Med*. 2012;367:411-422.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Lohse N, Hansen AB, Pedersen G, et al. Survival of persons with and without HIV infection in Denmark, 1995-2005. *Ann Intern Med*. 2007;146:87-95.
2. Johnson LF, Mossong J, Dorrington RE, et al. Life expectancies of South African adults starting antiretroviral treatment: collaborative analysis of cohort studies. *PLoS Med*. 2013;10:e1001418.
3. Wheeler WH, Ziebell RA, Zabina H, et al. Prevalence of transmitted drug resistance associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, U.S—2006. *AIDS*. 2010;24:1203-1212.
4. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services. http://aidsinfo.nih.gov/contentfiles/lvguidelines/AdultandAdolescentGL.pdf; Accessed February 10, 2015.
5. Günthard HF, Aberg JA, Eron JJ, et al. Antiretroviral treatment of adult HIV infection: 2014 recommendations of the International Antiviral Society-USA Panel. *JAMA*. 2014;312:410-425.
6. Kitahata MM, Gange SJ, Abraham AG, et al. Effect of early versus deferred antiretroviral therapy for HIV on survival. *N Engl J Med*. 2009;360:1815-1826.
7. Hirsch MS, Günthard HF, Schapiro JM, et al. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. *Clin Infect Dis*. 2008;47:266-285.
8. Yazdanpanah Y, Fagard C, Descamps D, et al. High rate of virologic suppression with raltegravir plus etravirine and darunavir/ritonavir among treatment-experienced patients infected with multidrug-resistant HIV: results of the ANRS 139 TRIO trial. *Clin Infect Dis*. 2009;49:1441-1449.
9. Panel on Treatment of HIV-Infected Pregnant Women and Prevention of Perinatal Transmission. Recommendations for use of antiretroviral drugs in pregnant HIV-1-infected women for maternal health and interventions to reduce perinatal HIV transmission in the United States. http://aidsinfo.nih.gov/contentfiles/lvguidelines/PerinatalGL.pdf; Accessed February 10, 2015.
10. DAD Study Group, Friis-Møller N, Reiss P, Sabin CA, et al. Class of antiretroviral drugs and the risk of myocardial infarction. *N Engl J Med*. 2007;356:1723-1735.
11. Makadzange AT, Ndhlovu CE, Takarinda K, et al. Early versus delayed initiation of antiretroviral therapy for concurrent HIV infection and cryptococcal meningitis in sub-saharan Africa. *Clin Infect Dis*. 2010;50:1532-1538.
12. Török ME, Yen NT, Chau TT, et al. Timing of initiation of antiretroviral therapy in human immunodeficiency virus (HIV)–associated tuberculous meningitis. *Clin Infect Dis*. 2011;52:1374-1383.
13. Kuhar DT, Henderson DK, Struble KA, et al. Updated US Public Health Service guidelines for the management of occupational exposures to human immunodeficiency virus and recommendations for postexposure prophylaxis. *Infect Control Hosp Epidemiol*. 2013;34:875-892.
14. Smith DK, Grohskopf LA, Black RJ, et al. Antiretroviral postexposure prophylaxis after sexual, injection-drug use, or other nonoccupational exposure to HIV in the United States: recommendations from the U.S. Department of Health and Human Services. *MMWR Recomm Rep*. 2005;54:1-20.
15. U.S. Public Health Service. Preexposure prophylaxis for the prevention of HIV infection in the United States—2014. http://www.cdc.gov/hiv/pdf/prepguidelines2014.pdf. Accessed February 10, 2015.
16. Hütter G, Nowak D, Mossner M, et al. Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N Engl J Med*. 2009;360:692-698.

REVIEW QUESTIONS

1. Which CD4 cell count should prompt starting antiretroviral therapy in an HIV-infected individual?

A. <200/μL.
B. <350/μL.
C. <500/μL.
D. Treatment should be started regardless of CD4 cell count.

Answer: D.

2. Which treatment strategy is *not* recommended for first-line antiretroviral therapy?

A. Two nucleoside analogue reverse transcriptase inhibitors (NRTIs) + a non-nucleoside reverse transcriptase inhibitor (NNRTI)
B. Three NRTIs
C. Two NRTIs + a protease inhibitor (PI)
D. Two NRTIs + an integrase strand transfer inhibitor (INSTI)

Answer: B.

3. When should you change an antiretroviral regimen for treatment failure?

A. HIV RNA repeatedly 50 to 200 copies/mL
B. HIV RNA repeatedly over 500 copies/mL
C. Any confirmed HIV RNA above the limit of detection
D. CD4 cell count < 200 cells/μL

Answer: B.

4. What is the best regimen to treat both HIV and HBV infections?

A. zidovudine/lamivudine + darunavir/ritonavir
B. abacavir/lamivudine + raltegravir
C. tenofovir/emtricitabine + zidovudine
D. tenofovir/emtricitabine/efavirenz

Answer: D.

5. HIV pre-exposure prophylaxis (PrEP) should be offered to all *except:*

A. An active heroin user.
B. An HIV-negative partner of an HIV-positive individual with an HIV RNA of 250,000 copies/mL.
C. A heterosexual woman with several sexual partners who "doesn't always use condoms."
D. A gay man in a monogamous relationship with an HIV-negative partner.

Answer: D.

389

INFECTIOUS AND METABOLIC COMPLICATIONS OF HUMAN IMMUNODEFICIENCY VIRUS AND ACQUIRED IMMUNODEFICIENCY SYNDROME

HENRY MASUR, LETHA M. HEALEY, AND COLLEEN HADIGAN

INFECTIOUS COMPLICATIONS

The initial cases of the acquired immunodeficiency syndrome (AIDS) were identified when unusual infectious diseases occurred in patients who had no prior diagnosis of an immunodeficiency. *Pneumocystis* pneumonia (PCP), toxoplasma encephalitis, cytomegalovirus (CMV) retinitis, and *Mycobacterium avium* bacteremia, as well as Kaposi sarcoma and central nervous system (CNS) lymphoma were so unusual in previously healthy patients that suspicion was quickly raised that the patients must have had some new form of immune deficit, especially when individual patients manifested several such infections either serially or concurrently, and when the number of such patients rapidly increased. The specific infectious syndromes were so characteristic of this new syndrome that their occurrence in previously healthy patients was soon considered as "AIDS defining" until the retroviral etiology of the syndrome was discovered and diagnostic tests for human immunodeficiency virus (HIV) became available for widespread clinical testing.

Despite the widespread availability of effective antiretroviral regimens in the United States since the late 1990s, AIDS-related opportunistic infections are still seen frequently at many health care facilities, especially those serving populations with poor access to health care. In the United States, opportunistic infections occur in two distinct populations: those who are not under effective care and those in whom effective antiretroviral therapy (ART) has recently been started.[1]

At least 25% of HIV-infected patients in the United States are unaware of their retroviral infection. This group often presents to health care facilities with opportunistic infections as their initial clue that they have HIV infection. A substantial number of patients, additionally, are aware of their HIV

The first few years of the AIDS epidemic were devoted to learning how to recognize and treat these infectious diseases and tumors. At the onset of the epidemic, clinicians were hampered by a dearth of available and accurate diagnostic tools for many of these pathogens. There was also a dearth of therapeutic and preventive therapies for some of the common causative organisms. As better diagnostic techniques were developed, and effective drug therapies were discovered and tested, patients survived their opportunistic infections with increasing frequency. The demonstration that specific chemoprophylaxis could prevent many of these infectious diseases also improved the quality and duration of patient survival. However, patient survival did not improve dramatically until antiretroviral regimens were developed that provided durable HIV viral suppression with immunologic reconstitution, which thus substantially reduced or eliminated the occurrence of subsequent opportunistic infections.

infection but are not in care as a result of economic, behavioral, or social factors. These patients also present with initial or serial opportunistic infections because they are not benefitting from stable and effective ART.

A second population who develop opportunistic infections includes patients who have access to care and are durably suppressed virologically. In this group, opportunistic infections develop when the initiation of ART unmasks the presence of opportunistic pathogens, creating symptomatic clinical syndromes. In addition, immune reconstitution inflammatory syndromes occur in the first few weeks or months after initiation of ART, representing exaggerated immune responses to viable organisms or to antigen (Chapter 395).

In addition to the relatively acute opportunistic infections that have been traditionally associated with AIDS, as patients live longer they are developing increasing morbidity and mortality due to certain latent viruses such as hepatitis C virus (HCV), and hepatitis B virus (HBV) (Chapter 149), as well as human papillomavirus (HPV) (Chapter 373).[2] HCV and HBV progression is accelerated in HIV-infected patients compared to HIV-uninfected patients, leading to earlier development of cirrhosis, liver failure, and hepatoma.[3,4] HPV is associated with cervical carcinoma and anal carcinoma, as well as oral cancers.

In addition to opportunistic infections and malignant neoplasms, an unexpected complication of chronic HIV disease in patients who have suppressed HIV viral loads has been a persistent inflammatory state that appears to be related to low-level HIV viremia. This viremia is usually below the level of detection of standard clinical assays for blood and may be related to viral replication in poorly understood reservoirs. The persistent inflammatory state appears to accelerate atherosclerotic cardiovascular and cerebrovascular disease, renal disease, and hepatic disease. This persistent inflammatory state interacts with the AIDS-related metabolic dysfunctions, including dysglycemias and dyslipidemias, which exacerbate and complicate the inflammation-induced accelerated atherosclerosis.[5,6]

PATHOBIOLOGY

HIV infection causes cellular immune dysfunction by multifaceted mechanisms that include reducing the number and function of CD4 lymphocytes. The immune defect in patients with HIV infection is unique: in no other population do PCP, toxoplasma encephalitis, CMV retinitis, cryptococcal meningitis, cryptosporidiosis, microsporidiosis, and Kaposi sarcoma occur so frequently and with such characteristic presentations. The opportunistic infections that occur in patients with HIV infection differ substantially in their prevalence and natural history compared with patients with other immunodeficiencies. They differ in subtle ways in terms of response to therapy and prevention from the syndromes and causative organisms that occur in patients with other immunologic disorders due to corticosteroids or calcineurin inhibitors, for instance, or from immunologic and inflammatory defects associated with neutropenia or antibody deficiency or complement disorders.

The number of circulating CD4 cells is an excellent indicator of patient prognosis and of susceptibility to opportunistic infection for patients with HIV/AIDS (Fig. 389-1). Monitoring CD4 counts prospectively has thus become a cornerstone of patient management. The blood HIV viral load is also an independent predictor of host susceptibility to opportunistic infection, but is not nearly as sensitive and specific for estimating survival or for assessing susceptibility to opportunistic infection as are CD4 counts.[7] Although CD4 cells are enormously important to host defenses in all persons, the circulating CD4 count is not nearly as specific and sensitive a predictor of opportunistic infection susceptibility in any other patient population as they are in patients with HIV infection.

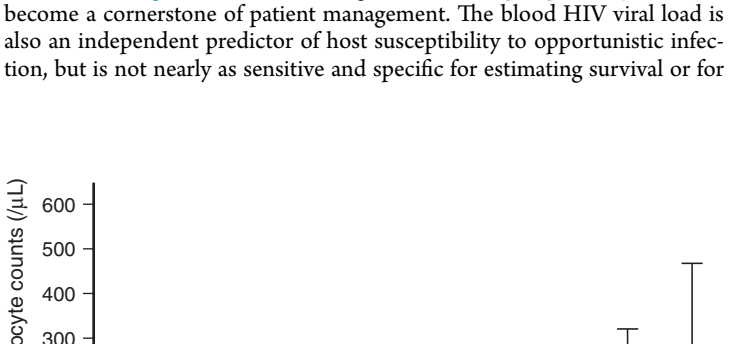

FIGURE 389-1. Distribution of CD4+ lymphocyte counts at diagnosis of opportunistic infection

The specific opportunistic infections that an HIV-infected patient develops are influenced not only by the patient's specific immunologic defects but also by environmental and behavioral factors. For instance, in areas of the world where exposure to *M. tuberculosis* is common, tuberculosis is a major cause of morbidity and mortality in patients with AIDS regardless of CD4 count, although the incidence of disease increases as the CD4 count declines. In contrast, however, in areas of the world such as the United States where tuberculosis exposure is relatively uncommon, tuberculosis is rarely seen except in immigrants and persons exposed to special populations, such as those in prisons or homeless shelters. Other mycobacteria, such as the environmental pathogen *Mycobacterium avium,* have become much more common than tuberculosis. Similarly, in many areas of the developing world, salmonellosis is a common complication of HIV infection and often causes diarrhea and life-threatening bacteremic disease. In the developed world, however, salmonella exposure is much less frequent and thus enteric disease is more likely to be caused by microsporidia, cryptosporidia, or nonopportunistic pathogens such as *Clostridium difficile* than by *Salmonella.*

Behavioral factors are also important determinants of which opportunistic infections occur. Patients with a history of intravenous drug abuse are more likely to be infected with HCV and HBV than matched patients without substance abuse histories. They are also more likely to develop nonopportunistic processes such as *Staphylococcal* sepsis because of their parenteral exposures. Men who have sex with men who develop proctitis or colitis are more likely to have lymphogranuloma venereum proctitis or gonococcal proctitis than patients with different risk factors.

The pathogens that cause active disease in patients with HIV infection may be organisms that were acquired recently or may represent reactivation of latent organisms acquired months or years previously. *Mycobacterium tuberculosis, Pneumocystis jiroveci, Trypanosoma cruzi, Leishmania donovani, Histoplasma capsulatum,* and *Coccidioides immitis* are examples of pathogens that can cause acute disease either soon after exposure or after many months or years of latency as assessed by molecular typing or clinical epidemiology. Thus, many pathogens need to be considered as possible etiologic agents despite exposure that may have occurred in the distant past.

CLINICAL MANIFESTATIONS

One of the early observations about clinical disease in patients with AIDS was that the clinical manifestations of opportunistic infections were not identical to the presentations in other immunosuppressed patients. In patients with AIDS, for instance, PCP is much more likely to manifest with subacute symptoms over weeks or months than in HIV-uninfected patients with cancer or transplant recipients who most often present acutely over a few days. When patients are diagnosed with PCP, those with AIDS are usually less hypoxemic and have less impressive radiographic infiltrates despite the longer duration of symptoms before diagnosis. The number of organisms found in sputum or bronchoalveolar lavage specimens is also greater in patients with AIDS than in other immunosuppressed individuals, such as patients with cancer or transplant recipients, despite the less severe symptoms. Patients with AIDS also are more likely to develop treatment-limiting toxicity associated with trimethoprim-sulfamethoxazole than patients with cancer or transplant recipients. Patients with AIDS are also more likely to have multiple recurrences if they are not treated with chemoprophylaxis than other immunosuppressed populations.

For infections due to *Toxoplasma gondii,* patients with HIV/AIDS characteristically develop toxoplasma encephalitis. Other immunosuppressed populations more often develop disseminated visceral disease involving the liver, spleen, or kidneys. Similarly, CMV in patients with HIV/AIDS causes retinitis and colitis. In patients with stem cell transplants, however, retinitis is relatively uncommon and pneumonia is frequent.

Some pathogens that have been recognized to cause frequent disease among patients with HIV/AIDS, such as *M. avium, Cryptosporidium, Microsporidium,* and *Bartonella,* were rarely recognized as causes of life-threatening human disease before the HIV/AIDS epidemic. Even as diagnostic studies have improved for these pathogens, they are far more often recognized among patients with HIV/AIDS than among other highly immunosuppressed patient populations. Conversely, some pathogens that were deemed likely to be AIDS associated based on the mechanisms for host

immune response, such as *Listeria monocytogenes* and disseminated *Strongyloides stercoralis,* are rarely seen in patients with HIV/AIDS. The reasons why some pathogens are unexpectedly frequent, or unexpectedly unusual despite similar environmental exposures, have not been fully elucidated.

DIAGNOSIS

For any infection in any patient population, management is likely to be more effective and to be associated with fewer complications if the specific cause is conclusively identified, the appropriate therapy is started quickly, and unnecessary drugs are avoided. For patients with HIV infection, such an approach is especially appropriate given the broad range of opportunistic and nonopportunistic infections that could cause a particular syndrome, as well as the noninfectious causes, including drug toxicities, that can masquerade as infections.

Although the diagnostic approach always should be individualized to the specific patient considering the current CD4 count, past and current exposures, prior infections, the history, physical findings, and routine laboratory tests of the current illness, certain tests are consistently useful. Blood cultures for routine bacteria and fungi, a serum cryptococcal antigen test, or a syphilis serologic test often provide useful information. If the patient has pulmonary dysfunction, Gram stain and routine culture of expectorated or induced sputum is usually useful. If a history of appropriate geographic exposure is present, serum and urine histoplasma antigen and Coccidioides antibodies also can be useful, as can a *Toxoplasma* IgG test.

The utility of specific tests needs to be validated in each patient population to determine their positive and negative predictive value. Certain tests that are useful in other patient populations or for research purposes are not necessarily useful to clinically diagnose opportunistic infections. For instance, serum polymerase chain reaction (PCR) for CMV is very useful for managing patients who have received stem cell transplants, because their positive and negative predictive values are high. However, serum CMV PCR in patients with HIV infections correlates mainly with the degree of immunosuppression and does not have sufficient positive and negative predictive value to be useful for assessing the cause of end organ disease. Serum PCR for Epstein-Barr virus (EBV), varicella zoster virus (VZV), or herpes simplex virus (HSV) also would not be useful in most circumstances for similar reasons. An increasing number of laboratories are offering other molecular tests for opportunistic pathogens, but care must be taken to be certain that the predictive value of these tests is proved. For instance, PCR for *Pneumocystis* in bronchoalveolar lavage may have excellent negative predictive value, but its positive predictive value is very low because many immunosuppressed patients appear to be colonized with *Pneumocystis* and thus a positive result does not prove with any confidence that *Pneumocystis* is the cause of the pulmonary dysfunction.

Imaging is an important aspect of patient evaluation. Patients with HIV infection may have pathologic processes despite a paucity of symptoms or normal readings on routine chest radiographs, for example. Thus, computerized tomography (CT) of the lungs may reveal unexpected pathologic findings such as diffuse interstitial infiltrates suggestive of PCP despite the absence of cough, shortness of breath, or oxygen desaturation. Such a finding could lead to an induced sputum or bronchoalveolar lavage diagnosis of the process at a time when disease is mild and the likelihood of successful treatment is high. CT of the abdomen also should be considered in patients with low CD4 counts even in the absence of abdominal symptoms, because such a study may reveal unexpected adenopathy or organ infiltration that could be readily biopsied more feasibly than other more obvious clinical manifestations. Positron emission tomography and nuclear scans have roles for identifying the etiology of infectious syndromes.

TREATMENT

Empirical Management

For the initial management of a presumed infectious syndrome in a patient with HIV infection or AIDS, clinicians need to determine the urgency of starting therapy before the specific causative process is definitively identified. Given the range of processes that can cause disease in patients with HIV infection or AIDS, the optimal approach is to establish the specific cause before starting therapy. However, some patients will be too sick or deteriorating too quickly to permit withholding therapy until a diagnosis can be established. Thus, for some patients, empirical therapy may be the best management strategy, with careful monitoring of the patient to determine if the therapy is effective.

When assessing an HIV-infected patient with any clinical syndrome, especially a syndrome associated with fever, opportunistic infections are immediate considerations. However, the likelihood of an opportunistic infection depends on the current CD4 count: if the patient's current CD4 count is greater than 200 to 300 cells/μL, the likelihood of most opportunistic infections (other than tuberculosis) is low (but not zero). For any patient, regardless of CD4 count, common community-acquired, nonopportunistic infections also must be considered because HIV-infected patients are equally susceptible to these as their HIV-uninfected counterparts. In addition, noninfectious syndromes must be considered, especially as the patient population ages and congestive heart failure, cerebrovascular disease, or chronic renal disease occur, potentially accelerated by the HIV-related inflammatory state. Patients also may have more than one process occurring concurrently. A familiar scenario, for instance, for a patient with documented PCP, would be the failure to recognize that the reason for pulmonary deterioration is not progressive PCP, but superimposed congestive heart failure, pulmonary hypertension, pulmonary emboli, or secondary lung infection due to methicillin-resistant *Staphylococcus aureus* (MRSA), *Streptococcus pneumoniae,* or *Cryptococcus neoformans*.

For some syndromes, empirical therapy is appropriate in most patients, with response to therapy providing a presumptive diagnosis. For instance, an empiric 2-week course of pyrimethamine plus sulfadiazine would be appropriate for a patient with AIDS with a CNS mass lesion, CD4 count less than 100 cells/μL, and positive antitoxoplasma serum immunoglobulin G (IgG), before a brain biopsy would be performed. Given the potential morbidity of a brain biopsy, and the high probability that a patient with AIDS with cerebral toxoplasmosis would demonstrate clinical and radiologic improvement within 14 days, such an approach has been considered preferred compared to immediate brain biopsy or even, in some circumstances, lumbar puncture. Similarly, for a patient with a CD4 count less than 200 cells/μL who presents with fever, cough, shortness of breath, severe hypoxemia, and diffuse bilateral interstitial pulmonary infiltrates, empirical therapy with ceftriaxone or vancomycin, plus azithromycin, plus trimethoprim-sulfamethoxazole to adequately provide coverage against common causes of community-acquired pneumonia as well as PCP would often be appropriate if the patient were too unstable to tolerate bronchoscopy without a high risk for intubation.

Definitive Therapy

The NIH-CDC-HIVMA Guidelines for the Prevention and Treatment of Opportunistic Infections in HIV-Infected Adults and Adolescents and other chapters in this text provide details on the diagnostic, therapeutic, and preventive approaches to specific syndromes. The drugs of choice are also listed in Table 389-1, adapted from the CDC-NIH-IDSA Guideline on Management of Opportunistic Infections in Adults and Adolescents, which is updated regularly online throughout the year (http://www.aidsinfo.nih.gov).

The treatments for opportunistic infections change as new drugs and new data become available.[8] Thus major AIDS guidelines are now updated promptly on-line (www.aidsinfo.nih.gov). For hepatitis-C, recommendations are changing so rapidly that the on-line site should be consulted before therapy is initiated unless the provider is very familiar with current data (www.hcvguidelines.org).

The institution of therapeutic or preventive drugs requires careful considerations of pharmacokinetics and drug-drug interactions. Patients with HIV infection often have organ dysfunction that may alter absorption or excretion of drugs. These patients are also often on multiple drugs (both AIDS related and AIDS unrelated) that can interact with clinically important consequences for drug effectiveness or toxicity. Such therapy thus requires considerable experience and consultation with up-to-date references for guidance.

When a patient with HIV/AIDS who has not been receiving ART develops an opportunistic infection, prospective studies demonstrate that the patient will have longer survival and fewer AIDS-defining complications if the patient is put promptly on ART. The definition of "prompt," that is, the decision as to how soon to start ART, is a complex analysis that must factor in the patient's willingness and ability to take ART, the evolution of the opportunistic infection if ART is not initiated, access to medical care and drugs after hospitalization, ability to absorb the drugs, potential interactions of ART with other drugs, including those used to treat the opportunistic infection, the patient's ability to tolerate potential drug toxicities, and the possible consequences if immune reconstitution inflammatory syndrome occurs.[9] The general principle is to start ART as soon as possible, but each patient will require individual assessment to determine the optimal interval between recognition of an opportunistic infection and initiation of ART.

When patients who are already receiving ART develop an opportunistic infection or any other complication not directly related to the drug itself, ART should be continued. The regimen should be reassessed to ensure that it is optimal in terms of antiviral activity, tolerability, potential toxicity and drug interactions, and the patient's ability to attain adequate serum levels given their ability to absorb oral drugs.

TABLE 389-1 PROPHYLAXIS TO PREVENT FIRST EPISODE OF OPPORTUNISTIC DISEASE

OPPORTUNISTIC INFECTIONS	INDICATION	PREFERRED	ALTERNATIVE
Streptococcus pneumoniae	For individuals who have not received any pneumococcal vaccine, regardless of CD4 count, followed by:	PCV13 0.5 mL IM × 1	PPV23 0.5 mL IM × 1
	• If CD4 count ≥ 200 cells/μL	PPV23 0.5 mL IM at least 8 wk after the PCV13 vaccine	
	• If CD4 count < 200 cells/μL	PPV23 can be offered at least 8 wk after receiving PCV13 or can wait until CD4 count increased to > 200 cells/μL	
	For individuals who have previously received PPV23	One dose of PCV13 should be given at least 1 yr after the last receipt of PPV23	
	Revaccination		
	• If age 19-64 yr and ≥ 5 yr since the first PPV23 dose	PPV23 0.5 mL IM × 1	
	• If age ≥ 65 yr and if ≥ 5 yr since the previous PPV23 dose	PPV23 0.5 mL IM × 1	
Influenza A and B virus	All HIV-infected patients	Inactivated influenza vaccine annually (per recommendation for the season) Live-attenuated influenza vaccine is contraindicated in HIV-infected patients	
Syphilis	For individuals exposed to a sex partner with a diagnosis of primary, secondary, or early latent syphilis within past 90 days *Or* For individuals exposed to a sex partner > 90 days before syphilis diagnosis in the partner, if serologic test results are not available immediately and the opportunity for follow-up is uncertain	Benzathine penicillin G 2.4 million U IM for 1 dose	*For penicillin-allergic patients:* • Doxycycline 100 mg PO q12h for 14 days *Or* • Ceftriaxone 1 g/day IM or IV for 8-10 days *Or* • Azithromycin 2 g PO for 1 dose; not recommended for MSM or pregnant women
Histoplasma capsulatum	CD4 count ≤ 150 cells/μL and at high risk because of occupational exposure or lives in a community with a hyperendemic rate of histoplasmosis (>10 cases/100 patient-years)	Itraconazole 200 mg/day PO	

(Guidelines for the Prevention and Treatment of Opportunistic Infections in HIV-Infected Adults and Adolescents. *http://aidsinfo.nih.gov/guidelines*.)
HIV = human immunodeficiency virus; IM = intramuscularly; IV = intravenously; MSM = men who have sex with men; PO = orally.

PREVENTION

Soon after the initial recognition of AIDS, before the era of ART or pathogen specific chemoprophylaxis, clinicians recognized that PCP ultimately occurred in 60 to 80% of patients in North America. Moreover, many of the patients who survived their first episode of PCP had one or more subsequent episodes. One of the first interventions demonstrated to prolong life was the institution of anti-*Pneumocystis* prophylaxis for patients who had oral thrush, oral hairy leukoplakia, a prior episode of PCP, or a CD4 count less than 200 cells/mm^3. Subsequently, the concept of primary and secondary chemoprophylaxis was extended to other pathogens such as *M. avium* complex (MAC) and toxoplasma.

These observations about *Pneumocystis,* MAC, and toxoplasma led to the development of a comprehensive preventive strategy to minimize the impact of opportunistic infections on those HIV-infected individuals who are immunologically vulnerable as measured by their CD4 count and viral load or by prior experience with an opportunistic infection (CDC-NIH-IDSA Guideline on Management of Opportunistic Infections in Adults and Adolescents, http://www.aidsinfo.nih.gov). Specific chemotherapy for primary prevention is indicated for the duration of immunosuppression, with the CD4 count thresholds depending on the pathogen. Chronic suppressive therapy should be continued for durations that depend on the pathogen and the patient's CD4 count. Recommendations for primary prophylaxis are summarized in Table 389-2; more complete information on primary and secondary prophylaxis is available in the CDC-NIH-IDSA Guideline on Management of Opportunistic Infections in Adults and Adolescents (http://www.aidsinfo.nih.gov).

Although immune reconstitution with ART is the most effective method to prevent opportunistic infections, many patients with low CD4 counts and uncontrolled viremia will continue to benefit from chemoprophylaxis. These efforts should be modified when patients have had a durable suppression of their viremia and increase in their circulating CD4 counts.

Preventive strategies focus not only on chemoprophylaxis but also on immunization and reducing exposure to opportunistic pathogens. Examples of exposure reduction interventions likely to be effective would be reducing exposure of patients to puppies and kittens from commercial breeders and reducing consumption of untreated surface water to prevent cryptosporidiosis. Other useful interventions would include precautions in travel to the Southwest United States to avoid coccidioidomycosis and avoiding exposure to outdoor cats and rare meat to reduce the likelihood of toxoplasmosis.

Immunizations such as pneumococcal vaccine, HPV vaccine, and hepatitis B vaccine also can be important, although host response to vaccines is reduced when patients are immunosuppressed. Live virus vaccines must be avoided in patients with low CD4 counts.

METABOLIC DISORDERS

With effective and well-tolerated ART, HIV-infected patients have a survival that is almost as long as HIV-uninfected persons.[10] It has become clear that patients with HIV infection experience increased incidence of cardiovascular and cerebrovascular disease; metabolic diseases, including diabetes mellitus; chronic kidney and liver disease; and perhaps accelerated neurocognitive decline, compared to HIV-uninfected patients. The proposed etiology for these disorders appears to be multifactorial and includes various ART-associated toxicities, immune dysregulation, and chronic inflammation related to chronic retroviral disease, even in the context of durable viral suppression.

Text continued on p. 2302

TABLE 389-2 TREATMENT OF ACQUIRED IMMUNODEFICIENCY VIRUS–ASSOCIATED OPPORTUNISTIC INFECTIONS (INCLUDES RECOMMENDATIONS FOR ACUTE TREATMENT AND SECONDARY PROPHYLAXIS/CHRONIC SUPPRESSIVE/MAINTENANCE THERAPY)

OPPORTUNISTIC INFECTION	PREFERRED THERAPY	ALTERNATIVE THERAPY	OTHER COMMENTS
Pneumocystis pneumonia (PCP) infection	Patients who develop PCP despite TMP-SMX prophylaxis usually can be treated with standard doses of TMP-SMX. Duration of PCP treatment: 21 days *For moderate-to-severe PCP:* • TMP-SMX: (TMP 15-20 mg and SMX 75-100 mg)/kg/day q6h or q8h IV; can switch to PO after clinical improvement *For mild-to-moderate PCP:* • TMP-SMX: (TMP 15-20 mg and SMX 75-100 mg)/kg/day, given PO in three divided doses *Or* • TMP-SMX: (160 mg/800 mg or DS) 2 tablets q8h PO *Secondary prophylaxis, after completion of PCP treatment:* • TMP-SMX DS: 1 tablet/day PO *Or* • TMP-SMX (80 mg/400 mg or SS): 1 tablet/day PO	*For moderate-to-severe PCP:* • Pentamidine 4 mg/kg/day IV infused over ≥ 60 min; can reduce dose to 3 mg/kg/day IV because of toxicities *Or* • Primaquine 30 mg (base)/day PO + (clindamycin 600 mg q6h IV or 900 mg q8h IV) or (clindamycin 300 mg q6h PO or 450 mg q8h PO) *For mild-to-moderate PCP:* • Dapsone 100 mg/day PO + TMP 5 mg/kg q8h PO *Or* • Primaquine 30 mg (base)/day PO + (clindamycin 300 mg q6h or 450 mg q8h PO) *Or* • Atovaquone 750 mg q12h PO with food *Secondary prophylaxis, after completion of PCP treatment:* • TMP-SMX DS: 1 tablet PO three times weekly *Or* • Dapsone 100 mg/day PO *Or* • Dapsone 50 mg/day PO + (pyrimethamine 50 mg + leucovorin 25 mg)/wk PO *Or* • (Dapsone 200 mg + pyrimethamine 75 mg + leucovorin 25 mg)/week PO *Or* • Aerosolized pentamidine 300 mg/mo via Respirgard II nebulizer *Or* • Atovaquone 1500 mg/day PO *Or* • (Atovaquone 1500 mg + pyrimethamine 25 mg + leucovorin 10 mg)/day PO	*Indications for adjunctive corticosteroids:* • PaO_2 < 70 mm Hg at room air *Or* • Alveolar-arterial O_2 gradient > 35 mm Hg *Prednisone doses (beginning as early as possible and within 72 hours of PCP therapy):* • Days 1-5: 40 mg q12h PO • Days 6-10: 40 mg/day PO • Days 11-21: 20 mg/day PO IV methylprednisolone can be administered as 75% of prednisone dose. Benefit of corticosteroid if started after 72 hr of treatment is unknown, but some clinicians will use it for moderate-to-severe PCP. Whenever possible, patients should be tested for G6PD before use of dapsone or primaquine. Alternative therapy should be used in patients found to have G6PD deficiency. Patients who are receiving pyrimethamine/sulfadiazine for treatment or suppression of toxoplasmosis do not require additional PCP prophylaxis. If TMP-SMX is discontinued because of a mild adverse reaction, reinstitution should be considered after the reaction resolves. The dose can be increased gradually (desensitization) or reduced or the frequency modified. TMP-SMX should be permanently discontinued in patients with possible or definite Stevens-Johnson syndrome or toxic epidermal necrosis.

Toxoplasma gondii encephalitis	*Treatment of acute infection:* • Pyrimethamine 200 mg PO 1 time, followed by weight-based therapy as follows: • If < 60 kg, pyrimethamine 50 mg/day PO + sulfadiazine 1000 mg q6h PO + leucovorin 10-25 mg/day PO • If > 60 kg, pyrimethamine 75 mg/day PO + sulfadiazine 1500 mg q6h PO + leucovorin 10-25 mg/day PO • Leucovorin dose can be increased to 50 mg q12h or q24h *Duration for acute therapy:* • At least 6 wk; longer duration if clinical or radiologic disease is extensive or response is incomplete at 6 wk *Chronic maintenance therapy:* • Pyrimethamine 25-50 mg/day PO + sulfadiazine 2000-4000 mg/day PO (in two to four divided doses) + leucovorin 10-25 mg/day PO	*Treatment of acute infection:* • Pyrimethamine (leucovorin)* + clindamycin 600 mg q6h IV or PO *Or* • TMP-SMX (TMP 5 mg/kg and SMX 25 mg/kg) q12h IV or PO *Or* • Atovaquone 1500 mg q12h PO with food + pyrimethamine (leucovorin)* *Or* • Atovaquone 1500 mg q12h PO with food + sulfadiazine 1000-1500 mg q6h PO (weight-based dosing, as in preferred therapy) *Or* • Atovaquone 1500 mg q12h PO with food *Or* • Pyrimethamine (leucovorin)* + azithromycin 900-1200 mg/day PO *Chronic maintenance therapy:* • Clindamycin 600 mg q8h PO + (pyrimethamine 25-50 mg + leucovorin) *Or* • TMP-SMX DS 1 tablet q12h *Or* • Atovaquone 750-1500 mg q12h PO + (pyrimethamine 25 mg + leucovorin 10 mg)/day PO *Or* • Atovaquone 750-1500 mg q12h PO + sulfadiazine 2000-4000 mg/day PO (in two to four divided doses) *Or* • Atovaquone 750-1500 mg q12h PO with food • Pyrimethamine and leucovorin doses are the same as for preferred therapy.	Adjunctive corticosteroids (e.g., dexamethasone) should be administered only when clinically indicated to treat mass effect associated with focal lesions or associated edema; discontinue as soon as clinically feasible. Anticonvulsants should be administered to patients with a history of seizures and continued through acute treatment, but should not be used as seizure prophylaxis. If clindamycin is used in place of sulfadiazine, additional therapy must be added to prevent PCP.
Mycobacterium tuberculosis (TB) disease	After collecting specimen for culture and molecular diagnostic tests, empirical TB treatment should be started in individuals with clinical and radiographic presentation suggestive of TB. Initial phase (2 mo, given daily, 5-7 times/wk by DOT): • INH + [RIF or RFB] + PZA + EMB, Continuation phase: • INH + (RIF or RFB) daily (5-7 times/wk) or three times weekly Total duration of therapy (for drug-susceptible TB): • Pulmonary TB: 6 mo • Pulmonary TB and culture- positive after 2 mo of TB treatment: 9 mo • Extrapulmonary TB with a CNS infection: 9-12 mo • Extrapulmonary TB with bone or joint involvement: 6-9 mo • Extrapulmonary TB in other sites: 6 mo Total duration of therapy should be based on number of doses received, not on calendar time.	*Treatment for Drug-Resistant TB* Resistant to INH: • (RIF or RFB) + EMB + PZA + (moxifloxacin or levofloxacin) for 2 mo; followed by (RIF or RFB) + EMB + (moxifloxacin or levofloxacin) for 7 mo Resistant to rifamycin + 1 other drug: • Regimen and duration of treatment should be individualized based on resistance pattern, clinical and microbiologic responses, and in close consultation with experienced specialists.	Adjunctive corticosteroid improves survival for TB meningitis and pericarditis. RIF is ***not recommended*** for patients receiving HIV PI because of its induction of PI metabolism. RFB is a less potent CYP3A4 inducer than RIF and is preferred in patients receiving PIs. Once weekly rifapentine can result in development of rifamycin resistance in HIV-infected patients and is ***not recommended***. Therapeutic drug monitoring should be considered in patients receiving rifamycin and interacting ART. Paradoxical IRIS that is not severe can be treated with NSAIDs without a change in TB or HIV therapy. For severe IRIS reaction, consider prednisone and taper over 4 wk based on clinical symptoms. For example: • If receiving RIF: prednisone 1.5 mg/kg/day for 2 wk, then 0.75 mg/kg/day for 2 wk • If receiving RFB: prednisone 1 mg/kg/day for 2 wk, then 0.5 mg/kg/day for 2 wk A more gradual tapering schedule over a few months may be necessary for some patients.

TABLE 389-2 TREATMENT OF ACQUIRED IMMUNODEFICIENCY VIRUS–ASSOCIATED OPPORTUNISTIC INFECTIONS (INCLUDES RECOMMENDATIONS FOR ACUTE TREATMENT AND SECONDARY PROPHYLAXIS/CHRONIC SUPPRESSIVE/MAINTENANCE THERAPY)—cont'd

OPPORTUNISTIC INFECTION	PREFERRED THERAPY	ALTERNATIVE THERAPY	OTHER COMMENTS
Disseminated *Mycobacterium avium* complex (MAC) disease	*At least two drugs as initial therapy with:* • Clarithromycin 500 mg q12h PO + ethambutol 15 mg/kg/day PO *Or* • Azithromycin 500-600 mg + ethambutol 15 mg/kg)/day PO if drug interaction or intolerance precludes the use of clarithromycin Duration: • At least 12 mo of therapy, can discontinue if no signs and symptoms of MAC disease and sustained (>6 mo) CD4 count>100 cells/μL in response to ART	Addition of a third or fourth drug should be considered for patients with advanced immunosuppression (CD4 counts < 50 cells/μL), with high mycobacterial loads (>2 log CFU/mL of blood), or in the absence of effective ART. *Third or fourth drug options may include:* • RFB 300 mg/day PO (dosage adjustment may be necessary based on drug interactions) • Amikacin 10-15 mg/kg/day IV *Or* • Streptomycin 1 g/day IV or IM *Or* • Moxifloxacin 400 mg/day PO or levofloxacin 500 mg/day PO	Testing of susceptibility to clarithromycin and azithromycin is recommended. NSAIDs can be used for patients who experience moderate-to-severe symptoms attributed to IRIS. If IRIS symptoms persist, short-term (4-8 wk) systemic corticosteroids (equivalent to 20-40 mg prednisone) can be used.
Salmonellosis	All HIV-infected patients with salmonellosis should be treated because of high risk for bacteremia. • Ciprofloxacin 500-750 mg q12h PO (or 400 mg q12h IV) if susceptible *Duration of therapy:* For gastroenteritis without bacteremia: • If CD4 count ≥ 200 cells/μL: 7-14 days • If CD4 count < 200 cells/μL: 2-6 wk For gastroenteritis with bacteremia: • If CD4 count ≥ 200/μL: 14 days; longer duration if bacteremia persists or if the infection is complicated (e.g., if metastatic foci of infection are present) • If CD4 count < 200 cells/μL: 2-6 wk *Secondary prophylaxis should be considered for:* • Patients with recurrent *Salmonella* gastroenteritis ± bacteremia *Or* • Patients with CD4 < 200 cells/μL with severe diarrhea	• Levofloxacin 750 mg q24h PO or IV *Or* • Moxifloxacin 400 mg q24h PO or IV *Or* • TMP, 160 mg, SMX 800 mg q12h (PO or IV) *Or* • Ceftriaxone 1 g q24h IV *Or* • Cefotaxime 1 g q8h IV	Oral or intravenous rehydration if indicated. Antimotility agents should be avoided. The role of long-term secondary prophylaxis in patients with recurrent *Salmonella* bacteremia is not well established. Must weigh benefit against risks for long-term antibiotic exposure. Effective ART may reduce the frequency, severity, and recurrence of *Salmonella* infections.
Mucocutaneous candidiasis	*For oropharyngeal candidiasis; initial episodes (for 7-14 days):* Oral therapy: • Fluconazole 100 mg/day PO *Or* Topical therapy: • Clotrimazole troches, 10 mg PO five times daily *Or* • Miconazole mucoadhesive buccal 50-mg tablet—apply to mucosal surface over the canine fossa once daily (do not swallow, chew, or crush) *For esophageal candidiasis (for 14-21 days):* • Fluconazole 100 mg (up to 400 mg)/day PO or IV *Or* • Itraconazole oral solution 200 mg/day PO *For uncomplicated vulvovaginal candidiasis:* • Oral fluconazole 150 mg for 1 dose *Or* • Topical azoles (clotrimazole, butoconazole, miconazole, tioconazole, or terconazole) for 3-7 days *For severe or recurrent vulvovaginal candidiasis:* • Fluconazole 100-200 mg/day PO for ≥ 7 days *Or* • Topical antifungal ≥ 7 days	*For oropharyngeal candidiasis; initial episodes (for 7-14 days):* Oral therapy: • Itraconazole oral solution 200 mg/day PO *Or* • Posaconazole oral solution 400 mg q12h PO for 1 day, then 400 mg/day Topical therapy: • Nystatin suspension 4-6 mL q6h 1-2 flavored pastilles four or five times daily *For esophageal candidiasis (for 14-21 days):* • Voriconazole 200 mg q12h PO or IV *Or* • Posaconazole 400 mg q12h PO *Or* • Anidulafungin 100 mg IV 1 time, then 50 mg/day IV *Or* • Caspofungin 50 mg/day IV *Or* • Micafungin 150 mg/day IV *Or* • Amphotericin B deoxycholate 0.6 mg/kg/day IV *Or* • Lipid formulation of amphotericin B 3-4 mg/kg/day IV *For uncomplicated vulvovaginal candidiasis:* • Itraconazole oral solution 200 mg/day PO for 3-7 days	Chronic or prolonged use of azoles may promote development of resistance. Higher relapse rate for esophageal candidiasis seen with echinocandins than with fluconazole use. Suppressive therapy usually not recommended unless patients have frequent or severe recurrences. *If decision is to use suppressive therapy:* Oropharyngeal candidiasis: • Fluconazole 100 mg/day PO or three times weekly • Itraconazole oral solution 200 mg/day PO Esophageal candidiasis: • Fluconazole 100-200 mg/day PO • Posaconazole 400 mg q12h PO Vulvovaginal candidiasis: • Fluconazole 150 mg PO once weekly

Cryptococcosis	*Cryptococcal meningitis:* Induction therapy (for at least 2 wk, followed by consolidation therapy): • Liposomal amphotericin B 3-4 mg/kg/day IV + flucytosine 25 mg/kg PO QID (*Note:* Flucytosine dose should be adjusted in patients with renal dysfunction.) Consolidation therapy (for at least 8 wk, followed by maintenance therapy): • Fluconazole 400 mg/day PO (or IV) Maintenance therapy Liposomal: • Fluconazole 200 mg/day PO for at least 12 mo *For non-CNS, extrapulmonary cryptococcosis and diffuse pulmonary disease:* • Treatment same as for cryptococcal meningitis *Non-CNS cryptococcosis with mild-to-moderate symptoms and focal pulmonary infiltrates:* • Fluconazole, 400 mg/day PO for 12 mo	*Cryptococcal meningitis:* Induction therapy (for at least 2 wk, followed by consolidation therapy): • Amphotericin B deoxycholate 0.7 mg/kg/day IV + flucytosine 25 mg/kg q6h PO *Or* • Amphotericin B lipid complex 5 mg/kg/day IV + flucytosine 25 mg/kg q6h PO *Or* • Liposomal amphotericin B 3-4 mg/kg/day IV + fluconazole 800 mg/day PO or IV *Or* • Amphotericin B deoxycholate 0.7 mg/kg/day IV + fluconazole 800 mg/day PO or IV *Or* • Fluconazole 400-800 mg/day PO or IV + flucytosine 25 mg/kg q6h PO *Or* • Fluconazole 1200 mg/day PO or IV *Consolidation therapy (for at least 8 wk, followed by maintenance therapy):* • Itraconazole 200 mg q12h PO for 8 wk—less effective than fluconazole *Maintenance therapy:* • No alternative therapy recommendation	Addition of flucytosine to amphotericin B has been associated with more rapid sterilization of CSF and decreased risk for subsequent relapse. Patients receiving flucytosine should have either blood levels monitored (peak level 2 hr after dose should be 30-80 µg/mL) or close monitoring of blood counts for development of cytopenia. Dosage should be adjusted in patients with renal insufficiency. Opening pressure should always be measured when an LP is performed. Repeated LPs or CSF shunting are essential to effectively manage increased ICP. Corticosteroids and mannitol are ineffective in reducing ICP and are **not recommended**. Some specialists recommend a brief course of corticosteroid for management of severe IRIS symptoms.
Histoplasmosis	*Moderately severe to severe disseminated disease:* Induction therapy (for at least 2 wk or until clinically improved): • Liposomal amphotericin B 3 mg/kg/day IV Maintenance therapy: • Itraconazole 200 mg q8h PO for 3 days, then 200 mg q12h PO *Less severe disseminated disease:* Induction and maintenance therapy: • Itraconazole 200 mg q8h PO for 3 days, then 200 mg q12h PO Duration of therapy: • At least 12 mo *Meningitis:* Induction therapy (4-6 wk): • Liposomal amphotericin B 5 mg/kg/day Maintenance therapy: • Itraconazole 200 mg q8h to q12h PO to for ≥ 1 yr and until resolution of abnormal CSF findings Long-term suppression therapy: For patients with severe disseminated or CNS infection after completion of at least 12 mo of therapy and those who relapse despite appropriate therapy: • Itraconazole 200 mg/day PO	*Moderately severe to severe disseminated disease:* Induction therapy (for at least 2 wk or until clinically improved): • Amphotericin B lipid complex 3 mg/kg/day IV *Or* • Amphotericin B cholesteryl sulfate complete 3 mg/kg/day IV Alternatives to itraconazole for maintenance therapy or treatment of less severe disease: • Voriconazole 400 mg q12h PO for 1 day, then 200 mg q12h *Or* • Posaconazole 400 mg q12h PO • Fluconazole 800 mg/day PO *Meningitis:* • No alternative therapy recommendation Long-term suppression therapy: • Fluconazole 400 mg/day PO	Itraconazole, posaconazole, and voriconazole may have significant interactions with certain ARV agents. These interactions are complex and can be bi-directional. Therapeutic drug monitoring and dosage adjustment may be necessary to ensure triazole antifungal and ARV efficacy and reduce concentration-related toxicities. Random serum concentration of itraconazole + hydroxy itraconazole should be > 1 µg/mL. Clinical experience with voriconazole or posaconazole in the treatment of histoplasmosis is limited. Acute pulmonary histoplasmosis in HIV-infected patients with CD4 counts > 300 cells/µL should be managed as in nonimmunocompromised host.

TABLE 389-2 TREATMENT OF ACQUIRED IMMUNODEFICIENCY VIRUS–ASSOCIATED OPPORTUNISTIC INFECTIONS (INCLUDES RECOMMENDATIONS FOR ACUTE TREATMENT AND SECONDARY PROPHYLAXIS/CHRONIC SUPPRESSIVE/MAINTENANCE THERAPY)—cont'd

OPPORTUNISTIC INFECTION	PREFERRED THERAPY	ALTERNATIVE THERAPY	OTHER COMMENTS
Cytomegalovirus (CMV) disease	*CMV retinitis:* Induction therapy: For immediate sight-threatening lesions (adjacent to the optic nerve or fovea): • Intravitreal injections of ganciclovir (2 mg) or foscarnet (2.4 mg) for one to four doses over a period of 7-10 days to achieve high intraocular concentration faster • Plus one of the listed preferred or alternative systemic therapy: Preferred systemic induction therapy: Valganciclovir 900 mg q12h PO for 14-21 days For peripheral lesions: Administer one of the preferred or alternative systemic therapy Chronic maintenance (secondary prophylaxis): Valganciclovir 900 mg/day PO *CMV esophagitis or colitis:* • Ganciclovir 5 mg/kg q12h IV; may switch to valganciclovir 900 mg q12h PO once the patient can tolerate oral therapy • Duration: 21-42 days or until symptoms have resolved • Maintenance therapy is usually not necessary, but should be considered after relapses *Well-documented. histologically confirmed CMV pneumonia:* • Experience for treating CMV pneumonitis in patients with HIV is limited. Use of IV ganciclovir or IV foscarnet is reasonable (doses same as for CMV retinitis) • The optimal duration of therapy and the role of oral valganciclovir have not been established *CMV neurologic disease:* (***Note:* Treatment should be initiated promptly.**) • Ganciclovir 5 mg/kg q12h IV + (foscarnet 90 mg/kg q12h IV or 60 mg/kg q8h IV) to stabilize disease and maximize response, continue until symptomatic improvement and resolution of neurologic symptoms • The optimal duration of therapy and the role of oral valganciclovir have not been established	*CMV retinitis:* Alternative systemic induction therapy: • Ganciclovir 5 mg/kg q12h IV for 14-21 days *Or* • Foscarnet 90 mg/kg q12h IV or 60 mg/kg q8h for 14-21 days *Or* • Cidofovir 5 mg/kg/wk IV for 2 wk; saline hydration before and after therapy and probenecid, 2 g PO 3 hr before dose, followed by 1 g PO 2 hr and 8 hours after the dose (total of 4 g). (*Note:* This regimen should be avoided in patients with sulfa allergy because of cross sensitivity with probenecid.) Chronic maintenance (secondary prophylaxis): • Ganciclovir 5 mg/kg IV 5-7 times weekly *Or* • Foscarnet 90-120 mg/kg IV once daily *Or* • Cidofovir 5 mg/kg IV every other week with saline hydration and probenecid as above *CMV esophagitis or colitis:* • Foscarnet 90 mg/ kg q12h IV or 60 mg/kg q8h for patients with treatment-limiting toxicities to ganciclovir or with ganciclovir resistance *Or* • Valganciclovir 900 mg q12h PO in milder disease and if able to tolerate PO therapy *Or* • For mild cases, if ART can be initiated without delay, consider withholding CMV therapy. • Duration: 21-42 days or until symptoms have resolved	The choice of therapy for CMV retinitis should be individualized, based on location and severity of the lesions, level of immunosuppression, and other factors (e.g., concomitant medications and ability to adhere to treatment). The ganciclovir ocular implant, which is effective for treatment of CMV retinitis is no longer available. For sight-threatening retinitis, intravitreal injections of ganciclovir or foscarnet can be given to achieve higher ocular concentration faster. The choice of chronic maintenance therapy (route of administration and drug choices) should be made in consultation with an ophthalmologist. Considerations should include the anatomic location of the retinal lesion, vision in the contralateral eye, the patients' immunologic and virologic status and response to ART. Patients with CMV retinitis who discontinue maintenance therapy should undergo regular eye examinations—normally every 3 mo—for early detection of relapse IRU, and then annually after immune reconstitution. IRU may develop in the setting of immune reconstitution. *Treatment of IRU:* • Periocular corticosteroid or short courses of systemic steroid Initial therapy in patients with CMV retinitis, esophagitis, colitis, and pneumonitis should include initiation or optimization of ART.

Herpes simplex virus (HSV) disease	*Orolabial lesions (for 5-10 days):* • Valacyclovir 1 g q12h PO *Or* • Famciclovir 500 mg q12h PO *Or* • Acyclovir 400 mg q8h PO *Initial or recurrent genital HSV (For 5-14 days):* • Valacyclovir 1 g q12h PO *Or* • Famciclovir 500 mg q12h PO *Or* • Acyclovir 400 mg q8h PO *Severe mucocutaneous HSV:* • Initial therapy acyclovir 5 mg/kg q8h IV • After lesions begin to regress, change to PO therapy, as above. Continue until lesions are completely healed. *Chronic suppressive therapy:* For patients with severe recurrences of genital herpes or patients who want to minimize frequency of recurrences: • Valacyclovir 500 mg q12h PO • Famciclovir 500 mg q12h PO • Acyclovir 400 mg q12h PO • Continue indefinitely regardless of CD4 cell count.	*For acyclovir-resistant HSV:* Preferred therapy: • Foscarnet 80-120 mg/kg/day IV in two or three divided doses until clinical response Alternative therapy: • IV cidofovir (dosage as in CMV retinitis) *Or* • Topical trifluridine *Or* • Topical cidofovir *Or* • Topical imiquimod • Duration of therapy: 21-28 days or longer	Patients with HSV infections can be treated with episodic therapy when symptomatic lesions occur or with daily suppressive therapy to prevent recurrences. Topical formulations of trifluridine and cidofovir are not commercially available. Extemporaneous compounding of topical products can be prepared using trifluridine ophthalmic solution and the IV formulation of cidofovir.
Varicella zoster virus (VZV) disease	*Primary varicella infection (chickenpox):* Uncomplicated cases (for 5-7 days): • Valacyclovir 1 g q8h PO *Or* • Famciclovir 500 mg q8h PO Severe or complicated cases: • Acyclovir 10-15 mg/kg q8h IV for 7-10 days • Can switch to oral valacyclovir, famciclovir, or acyclovir after defervescence if no evidence of visceral involvement. *Herpes zoster (shingles):* Acute localized dermatomal: • For 7-10 days; consider longer duration if lesions are slow to resolve • Valacyclovir 1 g q8h PO *Or* • Famciclovir 500 mg q8h Extensive cutaneous lesions or visceral involvement: • Acyclovir 10-15 q8h mg/kg IV until clinical improvement is evident • Can switch to PO therapy (valacyclovir, famciclovir, or acyclovir) after clinical improvement (i.e., when no new vesicle formation or improvement of signs and symptoms of visceral VZV), to complete a 10-14 day course. Progressive outer retinal necrosis (PORN): • (Ganciclovir 5 mg/kg ± foscarnet 90 mg/kg) q12h IV + (ganciclovir 2 mg/0.05 mL ± foscarnet 1.2 mg/0.05 mL) intravitreal injection twice weekly • Initiate or optimize ART Acute retinal necrosis (ARN): • (Acyclovir 10-15 mg/kg q8h IV) + (ganciclovir 2 mg/0.05 mL intravitreal injection twice weekly × 1-2 doses) for 10-14 days, followed by valacyclovir 1 g q8h PO for 6 wk	*Primary varicella infection (chickenpox):* Uncomplicated cases (for 5-7 days): • Acyclovir 800 mg PO 5 times per day *Herpes zoster (shingles)* Acute localized dermatomal: • For 7-10 days; consider longer duration if lesions are slow to resolve • Acyclovir 800 mg PO 5 times per day	In managing VZV retinitis, consultation with an ophthalmologist experienced in management of VZV retinitis is strongly recommended. Duration of therapy for VZV retinitis is not well defined, and should be determined based on clinical, virologic, and immunologic responses and ophthalmologic responses. Optimization of ART is recommended for serious and difficult-to-treat VZV infections (e.g., retinitis, encephalitis).

ART = antiretroviral therapy; ARV = antiretroviral; CNS = central nervous system; DOT = directly observed therapy; EMB = ethambutol; ICP = intracranial pressure; IRIS = immune reconstitution inflammatory syndrome; IRU = immune recovery uveitis; NSAIDs = nonsteroidal antiinflammatory drugs; PI = protease inhibitor; PZA = pyrazinamide; RFB = rifabutin; RIF = rifampicin; TMP-SMX = trimethoprim/sulfamethoxazole.

Numerous studies have now demonstrated an increased risk for coronary heart disease (CHD) among HIV-infected populations compared to uninfected contemporaries.[11] Although much of this increased risk is attributable to traditional CHD risk factors, such as diabetes, hypertension, dyslipidemia, and smoking, HIV-specific factors such as ART exposure, chronic inflammation, and immune dysfunction are also likely to be contributory. ART-associated increases in lipids as well as nonspecific ART toxicities account for a portion of the observed increased CHD in HIV. Some of the more compelling data on this topic were obtained from the Strategies for Management of Antiretroviral Therapy (SMART) study, which was designed to test whether CD4 T-cell guided reductions in exposure to ART would result in decreased cardiovascular events. Unexpectedly, there was an apparent increase in cardiovascular events observed with drug conservation. A subsequent nested case-control analysis of these data showed strong associations between elevated markers of inflammation at baseline, such as interleukin 6 and D-dimer and all-cause mortality,[12] highlighting the contribution of inflammation and immune activation in the context of chronic viral infection.

Careful identification and management of CHD risk factors is warranted in HIV care. Smoking, which is often enriched in populations with HIV infection, is an important modifiable risk factor that may be addressed to provide CHD risk reduction. As in the general population, smoking cessation in HIV-infected patients is accompanied by a reduction in incident CHD.[13] Dyslipidemia, a commonly identified condition among HIV-infected patients, caused in part by certain protease inhibitors and non-nucleoside reverse transcriptase inhibitors, is another target for CHD reduction. Selection of ART agents with limited effects on dyslipidemia is one potential approach to optimize lipid levels, but use of direct lipid-lowering therapy is often indicated. However, observational data suggests that standard therapy with 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors (statins) may be less effective for preventing cardiovascular disease[14] in the context of HIV despite their demonstrated ability to reduce markers of inflammation.[15] In addition, careful attention to potential drug-drug for preventing cardiovascular disease interactions is required when coadministering various protease inhibitors and statins due to effects on CYP3A4 activity that can lead to increased levels of statins.

Diabetes, insulin resistance, kidney injury, and neurocognitive dysfunction are also increased in HIV-infected populations. Similar to in the general population, the presence of chronic HCV infection is associated with an increased risk for diabetes and the development of chronic kidney disease[16] among individuals with HIV. However, immune dysfunction and toxicities associated with certain ART medications also increase the risk for diabetes and kidney injury in HIV.[17] Finally, the risk for developing HIV-associated neurocognitive dysfunction (HAND) has declined in the era of widespread ART use; however, longer survival and the increasing age of the HIV-infected population has led to a rise in the overall prevalence of HAND.[18] Indeed, with increased life expectancy, attentive care directed toward the recognition and management of metabolic, cardiovascular, and neurodegenerative conditions will become increasingly important in the health care approach to individuals living with HIV infection.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Gardner EM, McLees MP, Steiner JF, et al. The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. *Clin Infect Dis*. 2011;52:793-800.
2. Beachler DC, Sugar EA, Margolick JB, et al. Risk factors for acquisition and clearance of oral human papillomavirus infection among HIV-infected and HIV-uninfected adults. *Am J Epidemiol*. 2015;181:40-53.
3. Yaphe S, Bozinoff N, Kyle R, et al. Incidence of acute hepatitis C virus infection among men who have sex with men with and without HIV infection: a systematic review. *Sex Transm Infect*. 2012;88:558-564.
4. Thio CL, Smeaton L, Saulynas M, et al. Characterization of HIV-HBV coinfection in a multinational HIV-infected cohort. *AIDS*. 2013;27:191-201.
5. Shresta S, Irvin MR, Grunfeld C, et al. HIV, inflammation, and calcium in atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2014;34:244-250.
6. Sabin CA, Ryom L, De Wit S. Associations between immune depression and cardiovascular events in HIV infection. *AIDS*. 2013;27:2735-2748.
7. Mocroft A, Bannister WP, Kirk O, et al. The clinical benefits of antiretroviral therapy in severely immunocompromised HIV-1-infected patients with and without complete viral suppression. *Antivir Ther*. 2012;17:1291-1300.
8. Sulkowski MS, Naggie S, Lalezari J, et al. Sofosbuvir and ribavirin for hepatitis C in patients with HIV coinfection. *JAMA*. 2014;312:353-361.
9. Luetkemeyer AF, Kendall MA, Nyirenda M, et al. Tuberculosis immune reconstitution inflammatory syndrome in A5221 STRIDE: timing, severity, and implications for HIV-TB programs. *J Acquir Immune Defic Syndr*. 2014;65:423-428.
10. Losina E, Freedberg KA. Life expectancy in HIV. *BMJ*. 2011;343:d6015.
11. Freiberg MS, Chang CC, Kuller LH, et al. HIV infection and the risk of acute myocardial infarction. *JAMA Intern Med*. 2013;173:614-622.
12. Nordell AD, McKenna M, Borges ÁH, et al. Severity of cardiovascular disease outcomes among patients with HIV is related to markers of inflammation and coagulation. *J Am Heart Assoc*. 2014;3:e000844.
13. Petoumenos K, Worm S, Reiss P, et al. Rates of cardiovascular disease following smoking cessation in patients with HIV infection: results from the D:A:D study(*). *HIV Med*. 2011;12:412-421.
14. Rasmussen LD, Kronborg G, Larsen CS, et al. Statin therapy and mortality in HIV-infected individuals; a Danish nationwide population-based cohort study. *PLoS ONE*. 2013;8:e52828.
15. Funderburg NT, Jiang Y, Debanne SM, et al. Rosuvastatin reduces vascular inflammation and T cell and monocyte activation in HIV-infected subjects on antiretroviral therapy. *J Acquir Immune Defic Syndr*. 2014;[Epub ahead of print].
16. Lucas GM, Jing Y, Sulkowski M, et al. Hepatitis C viremia and the risk of chronic kidney disease in HIV-infected individuals. *J Infect Dis*. 2013;208:1240-1249.
17. Hadigan C, Edwards E, Rosenberg A, et al. Microalbuminuria in HIV disease. *Am J Nephrol*. 2013;37:443-451.
18. Heaton RK, Franklin DR, Ellis RJ, et al. HIV-associated neurocognitive disorders before and during the era of combination antiretroviral therapy: differences in rates, nature, and predictors. *J Neurovirol*. 2011;17:3-16.

REVIEW QUESTIONS

1. Which of the following opportunistic infections is most typical of HIV infection with a CD4 count less than 50 cells/mm^3?

A. Mucormycosis
B. BK virus nephropathy
C. Nocardiosis
D. *Mycobacterium avium* complex disease
E. *Listeria monocytogenes* meningitis

Answer: D Patients with HIV infection and low CD4 counts (<50 cells/mm^3) characteristically develop pneumocystosis, toxoplasmosis, CMV retinitis, tuberculosis, cryptococcal meningitis, cryptosporidiosis, and disseminated *Mycobacterium avium* complex infection. Mucormycosis is typically seen in patients with neutropenia or diabetes mellitus and is rare in patients immunosuppressed because of HIV infection. Listeria, surprisingly, does not appear to occur with enhanced frequency among patients with HIV infection although it is an intracellular pathogen. Nocardiosis can occur in patients with HIV infection, but is unusual. BK virus causes disease in transplant recipients, with nephropathy seen in renal transplant patients. BK virus is most unusual in patients with HIV infection.

2. Which of the following peripheral blood assays is the best indicator of susceptibility to opportunistic infections for patients with HIV infection?

A. CD4 lymphocyte count
B. CD8 lymphocyte count
C. Total lymphocyte count
D. Total neutrophil count
E. Total leukocyte count

Answer: A For HIV-infected patients, CD4 T cell counts in the peripheral blood are excellent indicators of susceptibility to opportunistic infections. There is a tight correlation between the counts and susceptibility to AIDS-defining illnesses. Counts under 200 cells/mm^{3}are a good estimate of susceptibility to pathogens such as *Pneumocystis* pneumonia. Counts under 100 cells/mm^3 are indicators of susceptibility to CMV, *Toxoplasma,* and *Mycobacterium avium.* The total lymphocyte count also can be used, but is not nearly so specific and sensitive. CD8 suppressor cells are related to immune response, but also are not tightly correlated with infectious susceptibility. Patients with HIV infection become neutropenic only with very-late-stage disease, when their marrow is involved by certain infections or tumors, or if they have marrow suppression resulting from a drug effect or a nutritional deficiency. Neutropenia in this population is not as ominous as in patients who have had chemotherapy and who have eroded mucosal barriers.

3. For HIV-infected patients on long-term antiretroviral therapy, and viral loads less than 50 copies and CD4 counts greater than 300 cells/mm^3, which of the following groups of problems are the major cause of HIV-related morbidity and mortality?

A. *Candida,* herpes zoster, and herpes simplex
B. Kaposi's sarcoma and multicentric Castleman's disease
C. Progressive multifocal leukoencephalopathy and central nervous system lymphoma
D. Accelerated liver, heart, and kidney disease
E. Sepsis and septic shock

Answer: D For patients with HIV infection well controlled on antiretroviral therapy, accelerated liver disease, coronary atherosclerosis, renal disease, and neurocognitive disorders are emerging as major causes of morbidity. At CD4 counts greater than 300, Kaposi's sarcoma occurs, but it is not common in the United States. Similarly, at such counts progressive multifocal leukoencephalopathy and CNS lymphoma are uncommon. Septic shock does not occur with substantially increased frequency in patients with HIV infection. They are susceptible at all CD4 counts to increased incidence of pneumococcal disease, but most cases of sepsis in this population come from predictable sources such as intravenous catheters or intravenous drug abuse.

4. For a patient with HIV infection whose CD4 count is 50 cells/mm^3 and who is going to start antiretroviral therapy in the next few weeks, which of the following prophylactic drugs would be most beneficial pending sustained HIV viral suppression?

A. Fluconazole
B. Valganciclovir
C. Valacyclovir
D. Trimethoprim-sulfamethoxazole
E. Caspofungin

Answer: D Patients with CD4 counts below 200 cells/mm^3 are susceptible to *Pneumocystis* pneumonia (PCP), and the susceptibility increases as the CD4 count decreases. In the pre-antiretroviral era, 70 to 80% of HIV-infected patients ultimately developed PCP if they did not take anti-pneumocystis prophylaxis. Azithromycin to prevent disseminated *Mycobacterium avium* complex infection would also be appropriate, although this option was not offered as an answer. Antiviral prophylaxis with valacyclovir would reduce the frequency of herpes simplex and herpes zoster, but this prophylaxis is not recommended in preference to treating episodes caused by these pathogens when they occur. Similarly, fluconazole could be used to reduce the likelihood of mucosal candidiasis, but the recommendation is not to give prophylaxis but to treat episodes when they occur. Caspofungin, a parental antifungal, would have no role for prophylaxis. Some drugs that are effective for prophylaxis, such as fluconazole and acyclovir, are not recommended for routine use because of potential drug interactions, drug toxicities, cost, and the fact that episodes of the targeted diseases are usually not life-threatening and can be treated when they occur acutely.

5. Which of the following vaccines would be appropriate to give to a patient with newly diagnosed HIV infection (CD4 count = 100 cells/mm^3, viral load = 1 million copies) who had childhood immunizations but none since then and who will be starting on antiretroviral therapy plus chemoprophylaxis with trimethoprim-sulfamethoxazole and azithromycin in 1 week?

A. Herpes zoster vaccine
B. Herpes simplex vaccine
C. Pneumococcal vaccine (conjugated)
D. Attenuated influenza vaccine
E. Measles-mumps-rubella vaccine

Answer: C For patients with low CD 4 counts (<200 cells/mm^3) it is not prudent to administer live virus vaccines such as the zoster vaccine or measles-mumps-rubella or the live attenuated influenza virus vaccine. There is no licensed vaccine for herpes simplex. Pneumococcal vaccine is important to administer. CDC guidelines and guidelines at http://www.aidsinfo.nih.gov provide information on the vaccine products to use. Pneumococcal vaccines are effective in reducing the morbidity of pneumococcal disease in this patient population. It may be wise to readminister this when the CD4 count has risen substantially if the initial immunization was administered when the CD4 count was below the 100 to 200 range.

390 GASTROINTESTINAL MANIFESTATIONS OF HIV AND AIDS

TAMSIN A. KNOX AND CHRISTINE WANKE

INTRODUCTION

Gastrointestinal (GI) diseases were among the leading causes of morbidity and mortality in persons infected with human immunodeficiency virus (HIV) before the advent of highly active antiretroviral therapy (HAART) in 1996. Profound wasting and chronic diarrhea are still manifestations of end-stage acquired immunodeficiency syndrome (AIDS) in persons who are not receiving or are resistant to antiretroviral therapy (ART). However, the common GI problems found today in persons receiving effective ART have shifted away from the enteric and hepatobiliary infections associated with low $CD4^+$ counts. Current GI issues are mainly due to side effects of antiretroviral medications, to nutritional and metabolic disorders associated with chronic HIV disease and therapy, and to liver disease from coinfection with hepatitis B virus (HBV) and hepatitis C virus (HCV).

The degree of immunosuppression, measured by the $CD4^+$ count, is the most important determinant of the likelihood of a GI illness, the type of disease, and its severity. The more profound the $CD4^+$ count suppression, the more likely a GI disease will be present.

The route of transmission of HIV may affect GI function; men who have sex with men (MSM) have higher rates of diarrhea and Kaposi sarcoma than do those who acquired HIV from intravenous drug use (IVDU). Alcohol abuse or IVDU predisposes to liver disease. Regional exposures may increase the risk for disease. In resource-limited settings, lack of ART, increased exposure to pathogens, lack of clean water, poverty, and hunger are associated with increased prevalence and severity of GI disease. In contrast, nutritional and metabolic diseases such as obesity, metabolic syndrome, and fatty liver are seen in HIV populations with high caloric and fat intake and lack of exercise. GI complications of HIV infection and recommendations for their treatment are summarized in Table 390-1.

GASTROINTESTINAL DISEASES FOUND WITH $CD4^+$ COUNTS GREATER THAN 200 TO 500 CELLS/μL

The GI diseases discussed in this section may also be seen with more advanced HIV infection.

Side Effects of Medications

Adverse effects of medications are frequently seen in the course of HIV treatment and may be due to ART or to the medications required for treatment or prevention of opportunistic infections. Gastric effects, including nausea, vomiting, loss of appetite, and dyspepsia, may occur as a symptom of the HIV infection itself or in response to ART regimens, particularly those with zidovudine, didanosine, ritonavir, amprenavir, or indinavir. Hypersensitivity reactions to ART, such as occurs in persons with the HLA-B5701 haplotype who take abacavir, may also be manifested as fever, abdominal pain, and rash. Didanosine, stavudine, and pentamidine may cause acute pancreatitis.

Lactic acidosis is a life-threatening condition caused by mitochondrial toxicity from nucleoside reverse transcriptase inhibitors (NRTIs), particularly stavudine, didanosine, and zidovudine. Early discontinuation of NRTIs and supportive care are key to recovery, but the syndrome may be difficult to diagnose because of nonspecific findings such as fatigue, nausea, muscle aches, weight loss, or abdominal pain. Blood lactate levels are usually greater than 5 mmol/mL. Nevirapine has been associated with liver toxicity when used in men and women with $CD4^+$ counts higher than 250 and 400 cells/μL, respectively. One of the most common adverse effects of atazanavir has been asymptomatic elevation in indirect bilirubin. Some ART agents have also been associated with the development of significant elevations in triglycerides (to levels > 1000 mg/dL) and the associated development of pancreatitis.

Although all antiretrovirals have been associated with diarrhea, nelfinavir is particularly known to cause a secretory form of diarrhea. Although antiretroviral therapy is now likely the most common cause of diarrhea in treated HIV-infected patients, intestinal pathogens should still be sought in stool examinations, despite a normal $CD4^+$ count or low HIV load.

Nutritional and Metabolic Disorders

With the advent of effective ART and the conversion of HIV to a chronic manageable disease, overnutrition and obesity have become frequent complications of HIV infection. Lifestyle factors and poor dietary choices may help promote the development of overweight/obesity, but some weight gain and increases in visceral fat may be related to the use of specific ART agents. HIV-associated lipodystrophy syndrome,[1] which was recognized after the introduction of effective triple ART regimens, is a complex of four distinct components: visceral fat accumulation, subcutaneous fat atrophy, atherogenic lipid profiles, and glucose intolerance. The pathogenesis remains unclear, but nutritional intake probably contributes to all components except fat atrophy. Chronic inflammation caused by long-term HIV infection and selected antiretrovirals has been implicated as well, particularly the protease

TABLE 390-1 RELATIONSHIP OF GASTROINTESTINAL DISEASES TO CD4⁺ COUNTS IN HIV-INFECTED PATIENTS

CD4⁺ COUNT	SYSTEM	SYMPTOMS AND DISEASES	TREATMENT
>200-500 cells/μL	GI and pancreas	**Nausea, vomiting, abdominal pain**	Changes in antiretroviral regimen
		Side effects of medications	Lipid-lowering agents, low-fat diet, no alcohol
		Pancreatitis	Treat specific pathogens, if present
		Diarrhea	Antimotility agents and fiber
	Nutritional and metabolic	**Lactic acidosis**	Stop nucleoside reverse transcriptase inhibitors
		Stavudine, didanosine, zidovudine	
		Changes in body shape	
		Obesity	Diet, exercise
		Metabolic syndrome	Diet, exercise, lipid-lowering agents, insulin-sensitizing agents
		Visceral obesity	hGH/GHRH in clinical trials
	Liver	**Liver function abnormalities, cirrhosis**	
		Chronic hepatitis with HCV	Combination of direct-acting agents against HCV
		Chronic HBV infection	Nucleosides/-tides for HBV infection
		Hepatotoxicity	Stop implicated antiretrovirals
	Oral lesions (≈400-500 cells/μL)	**Plaques, pain**	
		Oral thrush	Clotrimazole troche, nystatin, oral azoles
		Hairy leukoplakia	Acyclovir
		Herpes simplex-1	Acyclovir, famciclovir, valacyclovir
		Idiopathic ulcers	Corticosteroids, thalidomide (men)
		Gingivitis	Topical chlorhexidine, oral metronidazole
		Kaposi sarcoma	Radiation, intralesional injection, systemic chemotherapy
	Anorectal lesions	**Anorectal pain, mass, discharge, tenesmus**	
		Anal fissure, fistula, or perirectal abscess	Local care, surgery
		Foreign body in rectum	Endoscopic or surgical removal
		Sexually transmitted disease:	
		Gonorrhea, *Chlamydia trachomatis*, syphilis	Specific antibiotic therapy
		Herpes simplex infection	Acyclovir, famciclovir, valacyclovir
		Condyloma	Podophyllin, surgical excision, or ablation
		Anal cancer	Surgery, chemoradiation therapy
<200 cells/μL	Esophagus	**Dysphagia, odynophagia**	
	<200 cells/μL	*Candida*	Fluconazole, itraconazole, (voriconazole)
	<50 cells/μL	CMV	Ganciclovir
		Idiopathic ulcers	Oral corticosteroids, thalidomide (men)
		HSV	Acyclovir, famciclovir, valacyclovir
<100 cells/μL	Intestine	**Diarrhea**	
		Bacterial, viral, or protozoal infection	Treat specific pathogens
		AIDS enteropathy	Antimotility agents, nutritional support, ART
		Abdominal pain, fever	
		CMV	Treat specific pathogens
		Mycobacterium tuberculosis, MAC, *Histoplasma* *Cryptosporidium*	Surgery if perforation or dead bowel
	Malignancy	**GI bleeding, weight loss, or abdominal pain**	
		Kaposi sarcoma	As above
		Lymphoma	Systemic chemotherapy
<50 cells/μL	Wasting	**Weight loss**	
		Increased metabolic demands (infection, HIV)	Treat HIV and underlying infections
		Decreased oral intake	Appetite stimulants
		Malabsorption	Anabolic steroids, hGH
	Hepatobiliary	**Hepatomegaly, elevated liver function test results**	
		Opportunistic infections—MAC, fungal	Consider liver biopsy
		Lymphoma	
		Jaundice, RUQ pain, elevated liver function test results	
		AIDS cholangiopathy	ERCP and sphincterotomy
		Acalculous cholecystitis	Laparoscopic cholecystectomy

AIDS = acquired immunodeficiency syndrome; ART = antiretroviral therapy; CMV = cytomegalovirus; ERCP = endoscopic retrograde cholangiopancreatography; GHRH = growth hormone–releasing hormone; GI = gastrointestinal; HBV = hepatitis B virus; HCV = hepatitis C virus; hGH = human growth hormone; HIV = human immunodeficiency syndrome; HSV = herpes simplex virus; MAC = *Mycobacterium avium* complex; RUQ = right upper quadrant.

inhibitors. Host factors (e.g., genetics) also probably play a role in producing these syndromes. The co-occurrence of visceral fat accumulation, atherogenic lipid profiles, and glucose intolerance defines metabolic syndrome, the diagnosis of which requires three of the following factors: high triglyceride levels, low high-density lipoprotein levels, hypertension, elevated waist circumference, and glucose intolerance. The prevalence of metabolic syndrome has increased dramatically in HIV-infected individuals, and there is concern that it may contribute to the increased risk for cardiovascular disease in persons infected with HIV.[1,2]

Liver Disease

Liver disease is now the second leading cause of death in those with HIV infection, owing to the prevalence of coinfection with chronic viral hepatitis (Chapter 149). Coinfection with HCV is present in approximately 25%, depending on the HIV transmission category: 50 to 90% of IVDUs and hemophiliacs, 10 to 20% of heterosexuals, and 5 to 10% of MSM.[3] Chronic HBV infection, defined by the presence of hepatitis B surface antigen (HBsAg), is found in approximately 10% of persons with HIV, and some have more than one viral coinfection. Hepatic fibrosis progresses more rapidly in HIV-infected individuals and leads to cirrhosis, decompensated liver disease, or hepatic failure (Chapter 154). This may be further accelerated by alcohol use, steatosis, or fatty liver disease and by hepatotoxicity from ART. Coinfection with HCV or HBV increases the risk for death by four- to six-fold over HIV infection alone and increases the risk of hepatic decompensation.[4] Mortality from liver disease increases as CD4⁺ cells decline, more markedly with CD4⁺ counts lower than 100 cells/μL. Early diagnosis and treatment of HCV

and HBV may arrest or slow the progression of hepatic fibrosis and reduce the hepatotoxicity associated with antiretroviral treatment.

All persons with HIV infection should be screened for coinfection with HBV and HCV.[5] A positive serologic test for HBsAg indicates chronic HBV infection and should be followed by tests for HBV DNA and hepatitis B e antigen. Those without chronic HBV infection and with no protective antibody to HBV surface antigen (negative anti-HBs) should be vaccinated against HBV. If treatment of hepatitis B is to be initiated in the absence of ART, a dual regimen of telbivudine and adefovir is appropriate because these agents will not select for HIV resistance. A positive test for HCV antibody should prompt testing for HCV RNA, which determines whether chronic HCV infection is present. Noninvasive markers of hepatic fibrosis, serum biomarkers, and hepatic elastography (to measure hepatic stiffness) are being developed to replace liver biopsy. The prevalence of hepatocellular carcinoma (HCC) (Chapter 196) is increased in those with HIV coinfection; it is found at a younger age and is more aggressive in HIV-infected patients. Regular 6-month screening for HCC by serum α-fetoprotein levels and ultrasound imaging of the liver in persons with cirrhosis may help identify lesions earlier.

Randomized clinical trials have shown that anti-HCV therapy can induce a sustained virologic response (SVR)—that is, undetectable levels of HCV RNA measured 6 months after the end of treatment. However, SVR rates are approximately 60 to 80% of those achieved in persons with HCV monoinfection. Newer drug regimens in development with direct-acting antivirals hold promise for improved SVR rates without the side-effects of pegylated interferon.[6] To avoid drug interactions, neither didanosine (lactic acidosis) nor zidovudine (anemia) should be given with ribavirin.

Hepatotoxicity from ART occurs in up to 12% of patients started on a new ART regimen. Risk factors for hepatotoxicity include chronic HCV or HBV infection, female sex, and the use of ritonavir or nevirapine in the regimen. Effective ART has been associated with reduced hepatic inflammation, less progression of hepatic fibrosis, and decreased mortality from liver disease.

Oral Lesions

New oral lesions should raise suspicion for HIV disease. The lesions may be unsightly but asymptomatic or cause discomfort and difficulty eating. Thrush appears as white plaques, usually on the buccal mucosa. Oral hairy leukoplakia (Figure 425-5 in Chapter 425), which is associated with Epstein-Barr virus, consists of raised shaggy, dirty white patches on the sides of the tongue but is not premalignant.

Oral ulcers (Chapter 425) are caused by infection with herpes simplex virus type 1 (HSV-1) and tend to be more severe and persistent than ulcers in immunocompetent hosts. Idiopathic ulcers similar to those found in the esophagus are found in patients with advanced HIV disease. These deep necrotic ulcers are most commonly found on the buccal and pharyngeal mucosa. They respond to systemic corticosteroids or thalidomide. Gingivitis and periodontal disease appear as linear or diffuse erythema of the gums. With declining $CD4^+$ counts, they may progress to necrotizing gingivitis with pain and hemorrhage. There is increasing concern that these bacterial diseases may contribute to systemic inflammation and more rapid progression of HIV disease.

Kaposi sarcoma (Chapter 393), a multicentric malignancy of endothelial cells caused by human herpesvirus 8 infection, is more common in MSM with HIV infection in the United States but is also prevalent in Africa. Oral lesions are found in half of persons with other sites of Kaposi sarcoma and appear on the palate or gums as raised reddish or bluish nodules.

Anorectal Diseases

Anorectal disorders in HIV-infected individuals are more prevalent in MSM. Anal fissures, fistulas, abscesses, and foreign objects may be found on anal examination as a consequence of receptive anal intercourse or manipulation. Proctitis caused by gonorrhea *(Neisseria gonorrhoeae)* (Chapter 299) is characterized by mucopurulent discharge from the anus, tenesmus, and bleeding. *Chlamydia trachomatis* infection (Chapter 318) also causes proctitis and inguinal lymphadenopathy. Cultures and swabs of the discharge are diagnostic.

Primary syphilis (Chapter 319) is manifested as an anal chancre or ulcer that may be tender because of its location. Secondary syphilis appears 2 to 6 months later as condylomata lata, or warty masses around the anus. Perianal shallow, painful ulcers caused by HSV (Chapter 374) evolve as the initial vesicles rupture. These ulcerations may be cultured for HSV, or scrapings will show viral inclusions. Patients may have concurrent herpes proctitis.

Anal condylomata or warts are caused by infection with human papillomavirus (HPV) (Chapter 373) and appear as white, pink, or gray painless lesions around the anus and in the anal canal. HPV types 16 and 18 are associated with the development of dysplasia and squamous cell cancer of the anus, which is manifested as bleeding or a mass. The prevalence of anal dysplasia may be as high as 50% in HIV-positive MSM, and the incidence of anal cancer has increased more than six-fold in the past two decades. Screening via anal cytology is being pioneered for high-risk individuals.

GASTROINTESTINAL DISEASES MORE COMMONLY SEEN WITH $CD4^+$ COUNTS LESS THAN 200 CELLS/μL

GI illness increases in prevalence and severity as the $CD4^+$ count declines. With marked immunosuppression, there may be multiple concurrent infections. Although specific antimicrobial therapy may control the infection, HAART is the most effective therapy.

Esophageal Diseases

Candidal esophagitis (Chapter 338) is the most common esophageal infection in HIV-infected individuals. The combination of oral thrush and dysphagia has a positive predictive value of 90% for esophageal candidiasis. If the symptoms do not resolve with treatment in 7 days, endoscopy is indicated.[7] Cytomegalovirus (CMV) (Chapter 376) and idiopathic esophageal ulcerations may appear when the $CD4^+$ count falls below 50 cells/μL. Odynophagia, or severe pain on swallowing, is the initial symptom. Endoscopy shows large ulcerations with elevated margins, and CMV inclusions are seen in biopsy samples. HSV infections are much less common and appear as confluent shallow ulcerations in the esophagus. Idiopathic ulceration is diagnosed by the absence of viral inclusions on biopsy specimens.

Diarrhea

Diarrhea remains a common illness in HIV-infected individuals, even in the era of effective ART. Infections of the small intestine are accompanied by large-volume watery diarrhea. Colonic disease may be manifested as bloody, inflammatory, or small-volume diarrhea and tenesmus. Routine bacterial pathogens, including *Salmonella*, *Campylobacter*, and *Shigella*, have all been documented to cause diarrhea in HIV-infected patients. Enteroaggregative *Escherichia coli* has been identified as a cause of persistent diarrhea in HIV-infected individuals. The use of multiple antibiotics in HIV-infected patients has led to the frequent occurrence of *Clostridium difficile* colitis (Chapter 296). Mycobacterial infections with either *Mycobacterium tuberculosis* or atypical mycobacteria, most commonly *Mycobacterium avium* complex (MAC), may infiltrate the small bowel in HIV-infected patients. Although *M. tuberculosis* infection occurs with exposure in individuals with moderate $CD4^+$ cell depletion, disseminated MAC disease occurs only in individuals with $CD4^+$ counts lower than 50 cells/μL.

Most often, viral diarrheas are similar in manifestation to those in the general population, pathogens are not routinely identified in the clinical setting, and disease is self-limited. However, CMV may cause disease throughout the length of the GI tract in HIV-infected patients with $CD4^+$ counts lower than 50 cells/μL. In the small bowel, CMV may cause a watery diarrhea, and in the colon, a colitis with blood and signs of inflammation on endoscopy. Biopsy with identification of virus in tissue is required to make the diagnosis of CMV enteric disease.

Parasitic causes of diarrhea are common in HIV-infected individuals.[8] Opportunistic parasitic pathogens include *Cryptosporidium parvum*, the microsporidial organisms *Enterocytozoon bieneusi* and *Encephalitozoon intestinalis*, and *Cyclospora cayetanensis*. Persistent diarrhea caused by cryptosporidiosis (Chapter 350) occurs when the $CD4^+$ count is less than 200 cells/μL, and the most effective control of this pathogen is treatment of HIV.

Pathogens are not found in up to half of cases of diarrhea in HIV-infected patients, even after a full diagnostic evaluation. This condition has been called AIDS enteropathy and is associated with chronic diarrhea, malnutrition, and wasting.[9] It is possible that some of the chronic diarrhea is caused by pathogens that are not yet recognized or by noninfectious causes such as lymphoma. It is also likely that some of these pathogen-negative diarrheas are caused by HIV itself, which may infect the enterocyte as well as lymphoid tissue within the gut.

Malignancy

Kaposi sarcoma (Chapter 393) may develop at any location in the GI tract. Kaposi lesions appear endoscopically as raised reddish nodules that may bleed spontaneously. Non-Hodgkin lymphomas may involve any portion of

the GI tract, as well as the liver, and are manifested as bleeding, obstruction, weight loss, or abdominal pain.

Wasting

Wasting in HIV-infected individuals is defined by the World Health Organization as a body mass index of less than 18.5 kg/m^2. Weight loss and wasting are still seen in patients with treated HIV infection, and even modest weight loss of 3% or greater increases mortality.

In uncontrolled HIV infection, multiple factors contribute to wasting. Caloric requirements are increased by the metabolic demands of HIV replication, concurrent opportunistic infections, and fever. Oral intake may be poor because of nausea, anorexia, dysphagia, odynophagia, or chronic diarrhea or as a result of food insecurity, depression, or dementia.[10] Weight loss is exacerbated by malabsorption of nutrients from small intestinal or pancreatic disease. HIV-infected persons may not be able to increase caloric intake to overcome the loss of nutrients in stool and meet the metabolic demands of their illness. Targeted approaches to wasting and weight loss include treatment of HIV or opportunistic infections, frequent small meals, and use of appetite stimulants (megestrol acetate, dronabinol), anabolic steroids (testosterone, oxandrolone, nandrolone), or recombinant human growth hormone, as indicated. With control of HIV replication, nutritional status and weight generally improve, although they may not return to premorbid levels if the wasting was severe.

Hepatobiliary Disease

Hepatobiliary disease is associated with very low CD4$^+$ counts, often less than 20 cells/μL. Evaluation of abnormalities in liver function should initially exclude HBV and HCV infection and drug-induced hepatotoxicity, particularly that involving antiretroviral or sulfa-containing medications. The presence of hepatomegaly suggests an infiltrative process such as MAC or fungal infection or lymphoma. Alkaline phosphatase is disproportionately elevated. Liver imaging often does not show a localized lesion but determines whether biliary dilation is present, a finding suggestive of extrahepatic disease. Liver biopsy with special stains and cultures may be definitive.

In contrast, the presence of right upper quadrant pain with or without jaundice indicates biliary tract disease (Chapter 155). AIDS cholangiopathy[11] includes a sclerosing cholangitis–like process with diffuse biliary strictures and dilations and papillary stenosis with narrowing of the distal common bile duct at the entrance to the duodenum. In papillary stenosis, the common bile duct is dilated proximal to the stenosis. Acalculous cholecystitis may also be seen. These changes are caused by infection of the biliary tree with *Cryptosporidium*, microsporidia, *Isospora belli*, or CMV in the setting of profound immunosuppression. Endoscopic retrograde cholangiopancreatography can be performed to establish the diagnosis, obtain brushings for examination, and treat papillary stenosis by sphincterotomy. Treatment of the specific organism seldom eradicates infection because of the degree of immunosuppression. However, treatment with ART can result in marked improvements in biliary findings.

APPROACH TO GASTROINTESTINAL DISEASES

The GI tract is host to a myriad of infectious and infiltrative diseases in HIV-infected persons. Even so, routine GI diseases such as diverticulitis, choledocholithiasis, or peptic ulcer disease must be considered, particularly in those treated with ART. The clinician is guided by the symptom complex, as laid out in Table 390-1, to identify the probable organisms or tumors by clinical findings. The CD4$^+$ count influences the likelihood of the type of GI infection. It is important to remember that multiple illnesses may occur simultaneously with HIV infection and that finding one cause of a symptom or complaint may not be sufficient.

With diarrhea, careful stool evaluations should include cultures for *Salmonella*, *Shigella*, and *Campylobacter*; assay for *Clostridium difficile* toxin; and at least three examinations for parasites, including acid-fast stains for MAC and *Cryptosporidium*. If these measures are unrevealing, upper GI endoscopy with biopsy of the small intestine is indicated to diagnose protozoal infections with *Cryptosporidium*, *Isospora belli*, or microsporidia. Signs of colitis are best evaluated by flexible sigmoidoscopy or colonoscopy with biopsy and cultures. Examination and biopsy of the terminal ileum with acid-fast stain is necessary to confirm the diagnosis of pathogens such as *M. tuberculosis* or MAC.

All persons with HIV should be screened for coinfection with HBV and HCV and vaccinated against hepatitis A and B if not immune. Finally, close attention to serial measurement of weight and calculation of the body mass index is crucial to identify those who may have poor outcomes as a result of weight loss or excessive weight gain.

Grade A References

A1. Malan N, Su J, Mancini M, et al. Gastrointestinal tolerability and quality of life in antiretroviral-naive HIV-1-infected patients: data from the CASTLE study. *AIDS Care*. 2010;22:677-686.

A2. Sardar P, Jha A, Roy D, et al. Therapeutic effects of nandrolone and testosterone in adult male HIV patients with AIDS wasting syndrome (AWS): a randomized, double-blind, placebo-controlled trial. *HIV Clin Trials*. 2010;11:220-229.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Lake JE, Currier JS. Metabolic disease in HIV infection. *Lancet Infect Dis*. 2013;13:964-975.
2. Hemkens LG, Bucher HC. HIV infection and cardiovascular disease. *Eur Heart J*. 2014;35:1373-1381.
3. Kim AY, Onofrey S, Church DR. An epidemiologic update on hepatitis C infection in persons living with or at risk of HIV infection. *J Infect Dis*. 2013;207(suppl 1):S1-S6.
4. Re VL, Kallan MJ, Tate JP, et al. Hepatic decompensation in antiretroviaral-treated patients co-infected with HIV and hepatitis C virus compared with hepatitis C virus-monoinfected patients. *Ann Intern Med*. 2014;160:369-379.
5. Sulkowski MS, Cheever LW, Spach DH. A guide for evaluation and treatment of hepatitis C in adults coinfected with HIV. US Department of Health and Human Services, Health Resources and Services Administration (HRSA) (January 14, 2011). http://hab.hrsa.gov/deliverhivaidscare/files/hepccoinfectguide2011.pdf; Accessed February 9, 2015.
6. Cooper C, Klein M. HIV/hepatitis C virus coinfection management: changing guidelines and changing paradigms. *HIV Med*. 2014;15:621-624.
7. Cassone A, Cauda R. *Candida* and candidiasis in HIV-infected patients: where commensalism, opportunistic behavior and frank pathogenicity lose their borders. *AIDS*. 2012;26:1457-1472.
8. Huppmann AR, Orenstein JM. Opportunistic disorders of the gastrointestinal tract in the age of highly active antiretroviral therapy. *Hum Pathol*. 2010;41:1777-1787.
9. Cello JP, Day LW. Idiopathic AIDS enteropathy and treatment of gastrointestinal opportunistic pathogens. *Gastroenterology*. 2009;136:1952-1965.
10. Raiten DJ, Mulligan K, Papathakis P, et al. Executive summary–nutritional care of HIV-infected adolescents and adults, including pregnant and lactating women: what do we know, what can we do, and where do we go from here? *Am J Clin Nutr*. 2011;94:1667S-1676S.
11. Tonolini M, Bianco R. HIV-related/AIDS cholangiopathy: pictorial review with emphasis on MRCP findings and differential diagnosis. *Clin Imaging*. 2013;37:219-226.

REVIEW QUESTIONS

1. A 47-year-old man with poorly controlled HIV infection complains of pain on swallowing both liquids and solids. His HIV viral load is 157,468 copies/mL and his CD4 count is 152 cells/mm^3. He admits to poor adherence to antiretroviral therapy. On exam, he is wasted. There are white patches on his buccal mucosa. His abdominal exam is unremarkable. You recommend:

A. Upper gastrointestinal (GI) endoscopy with biopsies and brushings of the esophagus
B. Empirical treatment with fluconazole
C. Ophthalmic exam for cytomegalovirus (CMV) involvement of the retina
D. Empirical treatment with ganciclovir
E. Oral corticosteroids for idiopathic HIV related esophageal ulcers

Answer: B The most likely diagnosis is *Candida* esophagitis causing odynophagia. Esophageal infections do not cause obstruction, so there is no dysphagia, but the esophageal contractions with swallowing cause pain in an inflamed esophagus with either liquids or solids. Oral thrush and dysphagia are 90% predictive of esophageal candidiasis. An empirical trial of an antifungal medication directed at *Candida* is an appropriate treatment, reserving upper GI endoscopy for patients who do not quickly respond to treatment. CMV, herpes, and idiopathic ulcerations of the esophagus generally occur in patients with CD4 counts below 50 cells/mm^3. GI endoscopy with biopsies of the ulcers is the best way to make these diagnoses.

2. On initial laboratory reports, a 55-year-old man living with HIV/AIDS who is new to the clinic is found to have the following biochemical tests of liver function. He feels generally well. He is on a stable antiretroviral regimen with good control of HIV infection. He takes a multivitamin and Percocet for knee pain. His abdominal exam shows no hepatosplenomegaly. His CD4 count is 410 cells/mm^3, and his HIV viral load is undetectable.

Laboratory Measurement	Value	Normal Range
Bilirubin	1.0	0.2-1.3 mg/dL
Alkaline phosphatase	120	38-126 UL
AST	80	5-34 UL
ALT	90	0-55 UL
Total protein	7.8	6.3-8.2 g/dL
Albumin	4.0	3.5-5.0 g/dL

The most appropriate next step would be:

A. A liver biopsy to determine whether he has cirrhosis
B. Stop all medications and see if the biochemical tests return to normal.
C. Test for autoimmune hepatitis with a panel including antinuclear antibody and serum protein electrophoresis.
D. Weight reduction diet and exercise for fatty liver
E. Test for hepatitis B and C with hepatitis B surface antigen and anti–hepatitis C antibody.

Answer: E Biochemical tests of liver function are frequently abnormal in HIV/AIDS, with mild to moderate abnormalities in the aminotransferases (AST, ALT) being the most frequent abnormalities seen. The differential diagnosis includes drug-related hepatotoxicity, fatty liver, alcohol, and overuse of acetaminophen, but autoimmune hepatitis is uncommon in a male with HIV/AIDS. Over 25% of persons with HIV infection are coinfected with hepatitis C or hepatitis B virus. Chronic hepatitis B can be diagnosed by a positive hepatitis B surface antigen (which is also seen transiently in acute infection). Chronic hepatitis C infection is diagnosed by a positive anti–hepatitis C antibody (anti-HCV), and chronic infection is confirmed by testing for virus in the blood with an HCV-RNA. National recommendations (Health Resources and Services Administration) are to screen all persons with HIV/AIDS for chronic hepatitis B and C infection. Those who are negative should have repeated screening yearly.

3. A 51-year-old man with HIV/AIDS comes to clinic complaining of weight gain. He has a body mass index (BMI) of 31 kg/m^2. Control of his HIV infection is documented with a CD4 count of 520 cells/mm^3 and an undetectable HIV viral load. On exam, there is a fat deposit (buffalo hump) over his upper dorsocervical spine, abdominal obesity, but slender-appearing arms and legs. Clinical considerations should include:

A. Uncontrolled HIV infection causing wasting
B. Normal aging
C. Lipodystrophy with evidence of metabolic syndrome
D. An eating disorder
E. An inherited metabolic disorder

Answer: C HIV-associated lipodystrophy syndrome has four components: visceral fat accumulation (abdominal obesity), subcutaneous fat atrophy (slender arms and legs), atherogenic lipid profile, and glucose intolerance. The etiology of HIV-associated lipodystrophy has not been fully elucidated. However, dietary excess, chronic inflammation due to long-standing HIV infection, genetic predisposition, and the role of certain antiretrovirals may all play a role. The prevalence of metabolic syndrome has markedly increased in persons with HIV/AIDS and may predispose to accelerated cardiovascular disease. This patient should have fasting glucose and lipids checked and should be counseled on weight control and exercise.

4. A 42-year-old woman with HIV/AIDS presents with diarrhea. This has been going on for at least 2 months and is associated with mild abdominal cramping and weight loss because eating increases her cramps and diarrhea. She denies fevers or vomiting. What information is most helpful in planning her evaluation?

A. A family history of inflammatory bowel disease
B. The presence of blood in the stool
C. Obtaining a history of recent travel in the United States
D. Her most recent CD4 cell count
E. A history of lactose intolerance

Answer: D Diarrhea is very frequent among persons living with HIV/AIDS. The etiology of diarrhea is best predicted by a person's CD4 count. In those with a normal CD4 count, diarrhea is often related to medications, lactose intolerance, or irritable bowel syndrome. Usually, weight loss is not a significant component. When the CD4 count declines, opportunistic infections are more likely. At CD4 counts below 100 cells/mm^3, enteric pathogens are common, including bacterial pathogens (*Salmonella*, *Shigella*, *Campylobacter*, *Clostridium difficile*, and enteroaggregative *Escherichia coli*) and parasites (*Cryptosporidium*, microsporidia). At CD4 counts below 50 cells/mm^3, CMV, disseminated *Mycobacterium avium* complex (MAC), or AIDS enteropathy may cause diarrhea. The diagnosis can be made by stool cultures and smears for parasites, but both upper and lower GI endoscopy may be needed to obtain biopsies if stool examinations are negative.

5. A 52-year-old man with HIV/AIDS is seen in HIV clinic complaining of fatigue. When pressed, he admits to intermittent diarrhea and occasional fevers. He is on antiretrovirals. He is slender with a BMI of 20.3 kg/m^2. His examination is otherwise unremarkable except for a few raised cherry-colored nodules on his arms and chest. Which of the following could contribute to HIV-related wasting?

A. Excessive exercising
B. Inadequate dietary intake
C. An opportunistic infection or uncontrolled HIV infection increasing his energy requirements
D. AIDS enteropathy causing malabsorption
E. All of the above

Answer: E HIV-associated wasting is multifactorial. A careful dietary history and evaluation by a nutritionist is important to ensure adequate caloric intake. Dietary energy requirements are increased in persons who exercise regularly or who have increased metabolic demands due to an opportunistic infection and/or uncontrolled HIV infection from resistant virus. Malabsorption due to disease of the small intestine or pancreatic insufficiency also increases dietary caloric requirements because a percentage of nutrients are not being absorbed. Symptoms of nausea, vomiting, or anorexia may further decrease needed caloric intake.

391

PULMONARY MANIFESTATIONS OF HUMAN IMMUNODEFICIENCY VIRUS AND THE ACQUIRED IMMUNODEFICIENCY SYNDROME

KRISTINA CROTHERS AND ALISON MORRIS

INTRODUCTION

Pulmonary disease has historically been a leading cause of morbidity and mortality in patients with human immunodeficiency virus (HIV) and the acquired immunodeficiency syndrome (AIDS). Case reports of the previously rare *Pneumocystis* pneumonia (PCP) were the first harbingers of the AIDS epidemic in the 1980s, and in the early era of HIV, pulmonary infections such as bacterial pneumonia, PCP, and tuberculosis (TB) were frequently encountered. Pulmonary malignancies such as Kaposi sarcoma and lymphoma were also common. Prognosis for persons with HIV has changed dramatically with introduction of chemoprophylaxis for common infections such as PCP and with combination antiretroviral therapy (ART). The range of pulmonary diseases encountered in the HIV-infected patient has also changed over the course of the HIV epidemic and now includes fewer opportunistic infections with potential increases in diseases such as chronic obstructive pulmonary disease (COPD).

Pathobiology of HIV and Effects on Pulmonary Immunity

HIV infection in the absence of ART is characterized by immune dysfunction, dysregulation, and progressive immunodeficiency that results in a substantially increased risk for infections and other complications (Fig. 391-1). Early after initial infection, CD4 lymphocytes are depleted from mucosal-associated lymphoid tissue. During the chronic phase of untreated HIV, generalized immune activation and systemic CD4 lymphocyte depletion occurs, and remaining T cells may mount abnormal responses to antigens. Accompanying B-cell dysfunction results in abnormal polyclonal activation, hypergammaglobulinemia, and lack of specific antibody responses. Although ART decreases opportunistic infections and mortality in HIV-infected patients, persistent immune activation, dysfunction, and chronic low-level inflammation can persist and may contribute to the increased risk of several chronic comorbid diseases among HIV-infected individuals. The contribution of this chronic immune activation to pulmonary diseases such as COPD is currently unknown.

Within the lung parenchyma, HIV infection results in impaired innate and adaptive immune responses to pathogens.[1] Alveolar macrophages and lung CD4 T cells from HIV-infected individuals have deficiencies and impaired responses in pathogen recognition (e.g., influenza, *Mycobacterium tuberculosis*).[2] HIV can also lead to other abnormalities in host defense of the lung, including mucociliary function and soluble defense molecules within respiratory secretions. Among individuals who initiate ART, lung HIV viral levels and inflammation generally decrease, mirroring responses in the systemic

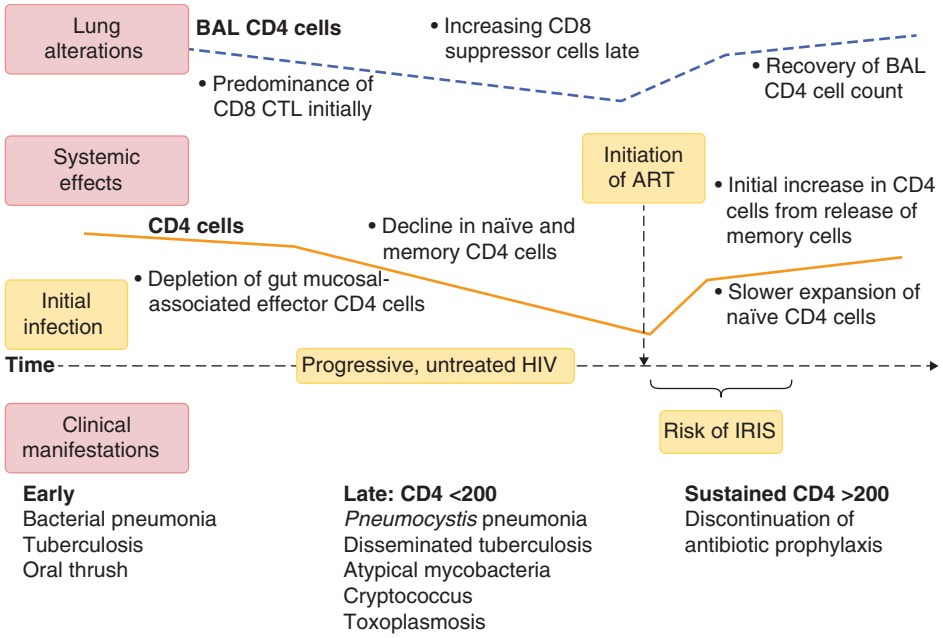

FIGURE 391-1. Systemic and lung alterations in T lymphocytes with HIV infection and initiation of antiretroviral therapy (ART). After initial HIV infection, CD4 lymphocytes of the effector memory type are depleted from mucosal-associated lymphoid tissue; the CD4 cells within the alveolar space appear to be spared but gradually decrease over time with progressive untreated HIV. During the chronic phase of HIV infection, there is progressive decline in the systemic CD4 cell count owing to decreased naïve and memory T cells. Within the alveolar space, HIV-specific cytotoxic CD8 T lymphocytes (CTL) predominate, although in late-stage disease these are replaced with CD8 suppressor lymphocytes. As discussed in the text, HIV infection is also associated with abnormal function of T cells, B-cell dysfunction, and within the lung, abnormalities in several other lines of host defense. With initiation of ART, CD4 cell counts increase systemically and in the lung. ART, antiretroviral therapy; BAL = bronchoalveolar lavage; IRIS, immune reconstitution inflammatory syndrome.

circulation. Nonetheless, HIV-infected individuals appear to have an increased risk of chronic lung diseases, although the mechanisms by which this increased risk occurs are not yet well understood.

EVALUATION OF THE HIV-INFECTED PATIENT WITH LUNG DISEASE

Respiratory disease is a common cause of both outpatient and inpatient visits in HIV-infected individuals. Pulmonary disease in this population can result from both infectious and noninfectious causes that are related to HIV infection or HIV medications, or that are unrelated to HIV. These pulmonary diseases often have characteristic clinical and radiographic manifestations, but there is also variability and overlap among them. In addition, HIV-infected patients often have more than one condition. For these reasons, definitive diagnosis of HIV-associated pulmonary diseases is encouraged when possible. The diagnostic evaluation should be guided by the constellation of the clinical signs and symptoms, laboratory testing, and radiographic appearance, as well as the severity of disease (Fig. 391-2A and 2B).

Clinical Findings

Often, the clinical history, physical examination, and laboratory testing can provide clues for a specific diagnosis (Table 391-1). For example, injection drug users have a greater risk of bacterial pneumonia and TB, whereas men who have sex with men have an increased risk of Kaposi sarcoma. As in the HIV-uninfected population, cigarette smoking is associated with bacterial pneumonia as well as COPD. In general, patients with a history of a previous opportunistic infection such as PCP are at increased risk of recurrence, although use of prophylaxis decreases disease incidence. Travel or place of residence also influences the risk of endemic fungal infections and TB. Certain extrapulmonary symptoms can also implicate specific diseases. For example, complaints of headache or altered mental status in a person with respiratory symptoms should prompt a search for *Cryptococcus* pneumonia and meningitis. Many diseases such as TB, malignancies, and fungal infections can cause extrapulmonary signs and symptoms such as lymphadenopathy, hepatic dysfunction, and bone marrow infiltration.

The CD4 cell count is one of the most critical pieces of information in determining the differential diagnosis of HIV-associated pulmonary disease (Table 391-2). Some diseases such as TB and bacterial pneumonia can occur at any CD4 cell count but are more common at lower counts, with a more atypical presentation of disease that is more likely to be disseminated. In contrast, disease such as pulmonary Kaposi sarcoma, *Toxoplasma gondii* pneumonia, and *Mycobacterium avium* complex are usually only seen if the CD4 cell count is below 100 cells/μL, and more often below 50 cells/μL. PCP is uncommon above a count greater than 200 cells/μL and less common if the CD4 cell count is between 100 and 200 cells/μL in a person receiving ART. If available, the most recent CD4 cell count *prior to* hospitalization is often more useful in HIV-infected inpatients to guide decision making, because the CD4 count can fall in the setting of acute illness.

Radiographic Studies

The radiographic picture also provides important clues to the diagnosis of HIV-associated pulmonary disease (Table 391-3). Certain radiographic findings are "classic" for certain diseases such as bilateral perihilar diffuse infiltrates in PCP (Fig. 391-3), but atypical presentations of pulmonary disease are not uncommon. The radiographic presentation also varies with the CD4 cell count. For example, TB presents with upper lung zone infiltrates that are often cavitary in patients with CD4 cell counts above 200 cells/μL (Fig. 391-4), but cavitation is uncommon with TB in patients with low CD4 cell counts, and lower lobe consolidation mimicking bacterial pneumonia can be seen (Fig. 391-5).

PULMONARY INFECTIONS

Pulmonary infections remain a major cause of morbidity and mortality in HIV-infected populations. Although pulmonary infections are more common in individuals without access to ART, a study of over 9000 HIV-infected and HIV-uninfected individuals found incident pulmonary infections remained more common in the era of combination ART.[3] The CD4 cell count has significant impact on the epidemiology of pneumonia, and HIV-infected individuals with advanced immunosuppression are at risk of a large spectrum of infectious causes of pneumonia (Fig. 391-6). Individuals with higher CD4 cell counts also remain at increased risk of bacterial pneumonia and TB when compared to HIV-uninfected populations. The general approach to the HIV-infected patient with pneumonia is described above, and we discuss the three most common HIV-associated pulmonary infections in further detail below.

Bacterial Pneumonia

EPIDEMIOLOGY

Recurrent bacterial pneumonia, defined as two or more episodes within 12 months, is an AIDS-defining illness. The risk for bacterial pneumonia is substantially increased among HIV-infected individuals, and in contrast to other opportunistic infections, this elevated risk persists in the combination ART era. For example, in a study of HIV-infected veterans in the ART era, the incidence rate of bacterial pneumonia in HIV-infected individuals was 28.0

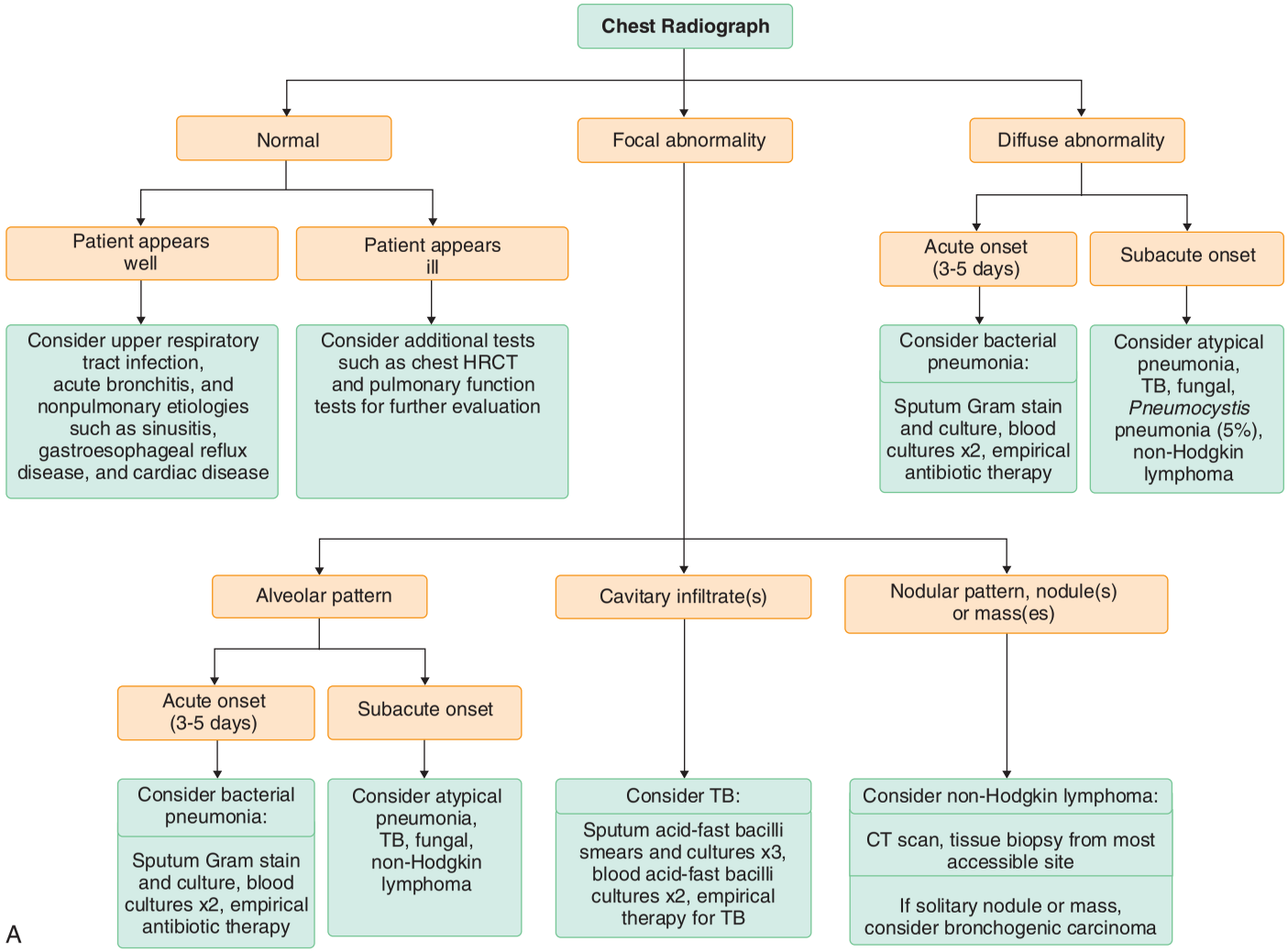

FIGURE 391-2. **Diagnostic approach to the HIV-infected patient with clinical manifestations of possible pulmonary disease and a CD4 cell count greater than 200 cells/μL (A) and a CD4 cell count below 200 cells/μL (B).** HRCT = high-resolution computed tomography; TB = tuberculosis.

Continued

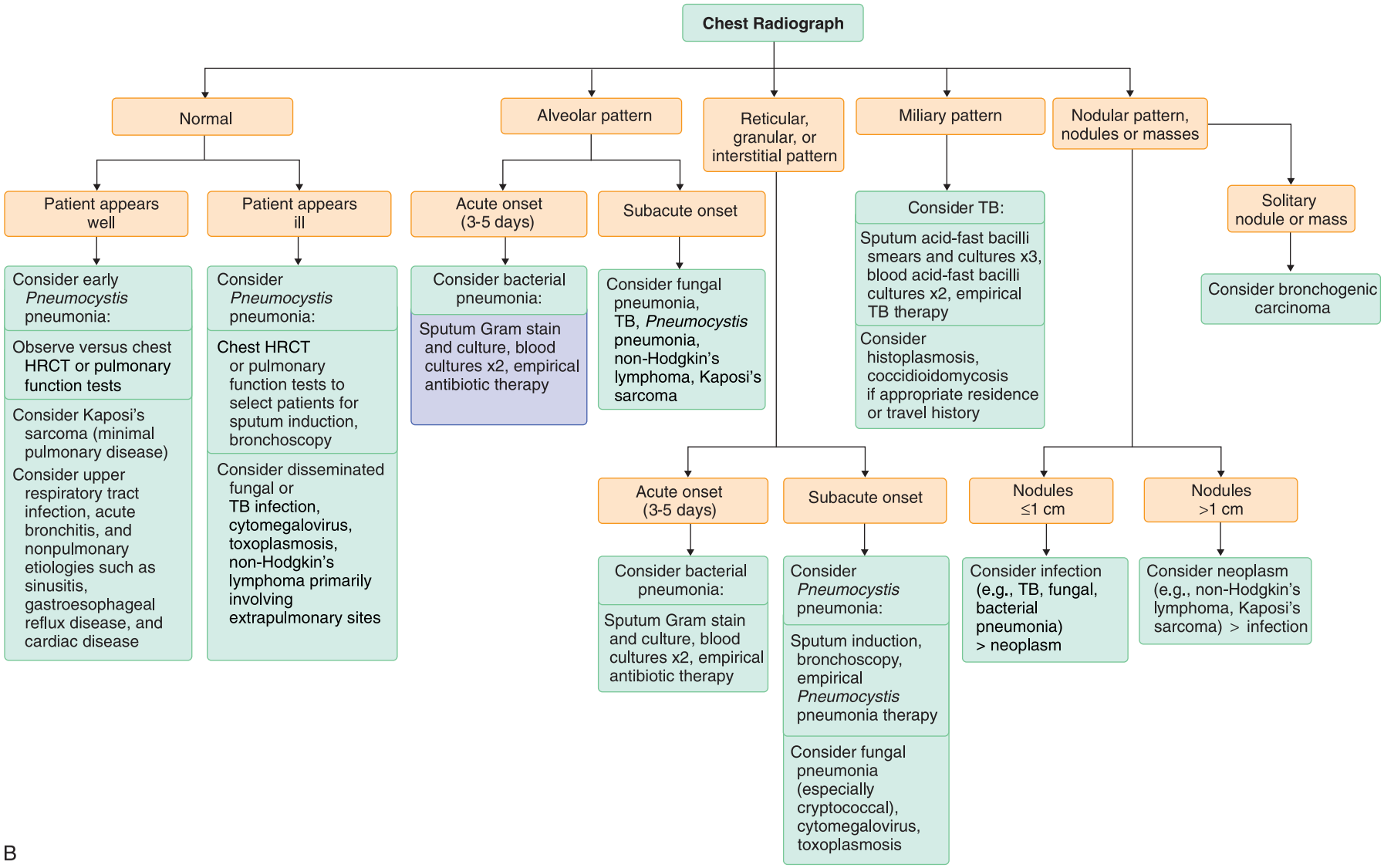

B

FIGURE 391-2, cont'd.

TABLE 391-1 DIAGNOSTIC CLUES TO ETIOLOGY OF HIV-ASSOCIATED PULMONARY DISEASES

CLINICAL SETTING

Ambulatory: URI > acute bronchitis > bacterial pneumonia > *Pneumocystis* pneumonia
Hospital: bacterial pneumonia > *Pneumocystis* pneumonia > TB > pulmonary KS
Intensive care unit: *Pneumocystis* pneumonia > bacterial pneumonia; non–AIDS associated conditions > AIDS-associated conditions

CD4 CELL COUNT (see Table 391-2)

PATIENT BACKGROUND

HIV transmission category: MSM—increased incidence of KS; IDU—increased incidence of bacterial pneumonia, TB
Habits: cigarettes—increased incidence of bacterial bronchitis, bacterial pneumonia, COPD, bronchogenic carcinoma
Travel and residence: assess risk for endemic fungal diseases, TB, NTM

MEDICAL BACKGROUND AND USE OF PROPHYLAXIS

Previous disease: increased incidence of recurrence of bacterial pneumonia, *Pneumocystis* pneumonia, fungal pneumonias
Prophylaxis/maintenance: decreased incidence of disease—*Pneumocystis* pneumonia, fungal pneumonias, TB (if PPD positive or positive interferon-γ release assay)

SYMPTOMS AND SIGNS

Respiratory symptoms: especially cough (productive or nonproductive) and symptom duration
Symptoms suggesting extrapulmonary or disseminated disease
Physical examination of the chest: focal or nonfocal findings
Signs suggesting extrapulmonary or disseminated disease

LABORATORY TESTS

WBC count: elevated or, if normal, elevated relative to baseline—bacterial pneumonia
Serum LDH: elevated—nonspecific but classically seen in *Pneumocystis* pneumonia
Arterial blood gas: nonspecific but useful for prognosis, management decisions (e.g., admission and whether corticosteroids are indicated for *Pneumocystis* pneumonia)

CHEST RADIOGRAPHY (see Table 391-3)

AIDS = acquired immunodeficiency syndrome; COPD = chronic obstructive pulmonary disease; HIV, human immunodeficiency virus; IDU = injection drug user; KS = Kaposi sarcoma; LDH = lactate dehydrogenase; MSM = men who have sex with men; NTM, nontuberculous mycobacteria; PPD = purified protein derivative (in an HIV-infected person, PPD is considered positive if ≥ 5-mm induration); TB = tuberculosis; URI = upper respiratory tract infection; WBC = white blood cell.

TABLE 391-2 CD4 CELL COUNT RANGES FOR HIV-ASSOCIATED PULMONARY DISEASES

ANY CD4 CELL COUNT

Bacterial pneumonia (most often *Streptococcus pneumoniae*, *Haemophilus* species)
Mycobacterium tuberculosis pneumonia
Influenza
Non-Hodgkin lymphoma
Lung cancer
Nonspecific interstitial pneumonitis
Lymphocytic interstitial pneumonitis
Pulmonary arterial hypertension
Chronic obstructive lung disease

CD4 CELL COUNT < 200 CELLS/μL

Pneumocystis pneumonia
Cryptococcus neoformans pneumonia

CD4 CELL COUNT < 100 CELLS/μL

Bacterial pneumonia caused by *Pseudomonas aeruginosa*
Toxoplasma gondii pneumonia
Pulmonary Kaposi sarcoma

CD4 CELL COUNT < 50 CELLS/μL

Mycobacterium avium complex—usually associated with disseminated disease
Histoplasma capsulatum—usually associated with disseminated disease
Coccidioides immitis—usually associated with disseminated disease
Aspergillus species (most often *Aspergillus fumigatus*) pneumonia
Cytomegalovirus pneumonia—usually associated with disseminated disease

TABLE 391-3 COMMON RADIOGRAPHIC FEATURES OF HIV-ASSOCIATED PULMONARY DISEASES

CONDITION	COMMON RADIOGRAPHIC FINDINGS
Pneumocystis pneumonia	Bilateral perihilar infiltrates Pneumatoceles, pneumothorax Ground-glass opacities Normal (chest radiograph) Upper lobe or asymmetrical infiltrates (less common)
Bacterial pneumonia	Focal infiltrates Pleural effusions Cavitary lesions
Mycobacterium tuberculosis	Focal infiltrates Normal (in extrapulmonary disease) Diffuse infiltrates/miliary pattern Nodules Cavitary lesions Intrathoracic adenopathy Pleural effusions Upper lobe involvement if high CD4 cell count
Mycobacterium kansasii	Consolidation Nodules Diffuse infiltrates Intrathoracic adenopathy Cavitary lesions Pleural effusions
Mycobacterium avium complex	Normal Intrathoracic adenopathy Focal pneumonia (rare)
Fungal pneumonia	Multifocal or diffuse infiltrates Focal infiltrate Normal Cystic lesions Nodules Intrathoracic adenopathy Pleural effusions (especially *Cryptococcus neoformans*)
Cytomegalovirus	Multifocal or diffuse infiltrates Reticular pattern Ground-glass opacities
Toxoplasma gondii	Bilateral infiltrates Pleural effusions
Non-Hodgkin lymphoma	Single or multiple nodules Focal infiltrates Diffuse interstitial infiltrates Pleural effusions Intrathoracic adenopathy
Kaposi sarcoma	Multifocal or diffuse infiltrates Bilateral central opacities Nodules Peribronchial cuffing, tram-track Kerley B lines Pleural effusions Intrathoracic adenopathy
Lung cancer	Masses Nodules Intrathoracic adenopathy Pleural effusions
COPD	Large lung volumes Flattened diaphragms Bullae Emphysema (chest CT)
Interstitial lung diseases	Normal (chest radiograph) Increased interstitial markings Small lung volumes Fibrosis Traction bronchiectasis Peribronchovascular nodules (sarcoidosis, LIP) Intrathoracic lymphadenopathy (sarcoidosis, rarely LIP)
Pulmonary hypertension	Normal Enlarged pulmonary arteries Vascular pruning Right heart enlargement

COPD = chronic obstructive pulmonary disease; CT = computed tomography; LIP = lymphocytic interstitial pneumonia

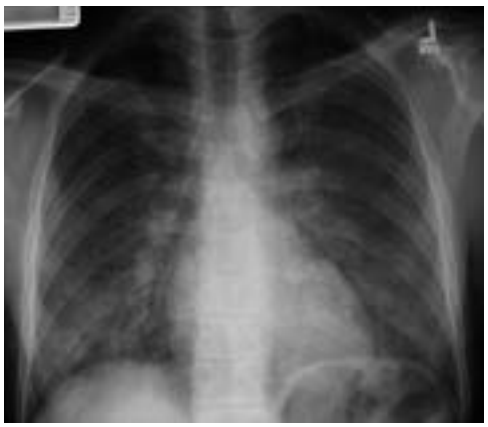

FIGURE 391-3. **Chest radiograph of an HIV-infected person, CD4 cell count less than 200 cells/μL, demonstrating characteristic bilateral reticular-granular opacities of *Pneumocystis* pneumonia.** Bronchoscopic alveolar lavage demonstrated *Pneumocystis* on Giemsa stain. Courtesy Laurence Huang, MD. Used with permission.

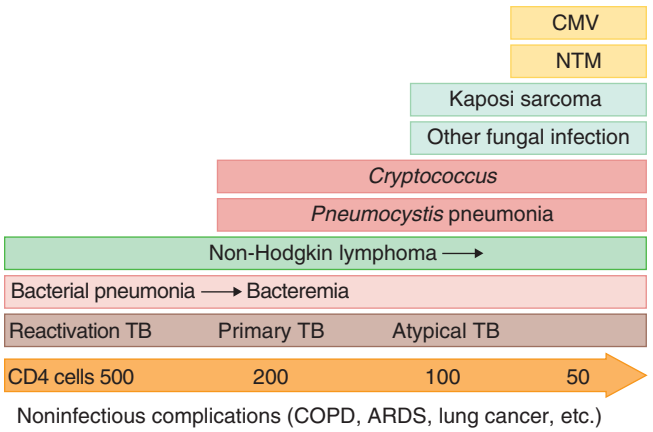

FIGURE 391-6. **Risk of pulmonary complications by CD4 cell count.** ARDS = acute respiratory distress syndrome; CMV = cytomegalovirus; COPD = chronic obstructive pulmonary disease; NTM = nontuberculous mycobacteria; TB = tuberculosis.

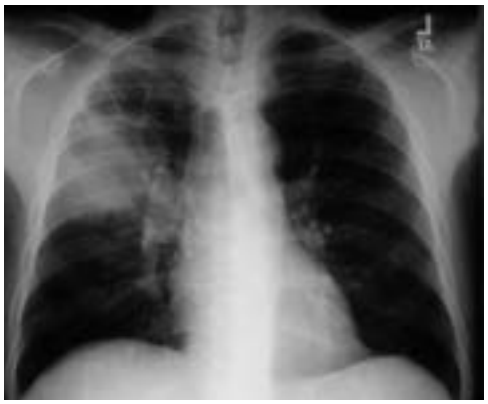

FIGURE 391-4. **Chest radiograph of an HIV-infected person, CD4 cell count greater than 200 cells/μL, demonstrating a right upper lobe infiltrate with areas of cavitation.** Sputum acid-fast bacillus stain was positive, and multiple sputum cultures grew *Mycobacterium tuberculosis*. Courtesy Laurence Huang, MD. Used with permission.

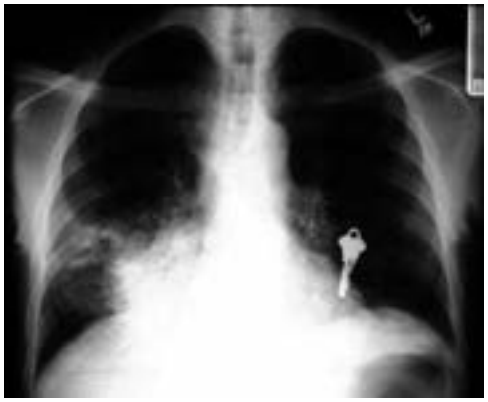

FIGURE 391-5. **Chest radiograph of an HIV-infected person, CD4 cell count less than 200 cells/μL, revealing right lower lung zone consolidation with air bronchograms.** Sputum culture grew *Mycobacterium tuberculosis* that was resistant to rifampin. Courtesy Laurence Huang, MD. Used with permission.

per 1000 person-years compared with 5.8 per 1000 person-years in the HIV-uninfected group.[4] Bacterial pneumonia can occur throughout the course of HIV infection and at any CD4 lymphocyte count, although the incidence increases as the CD4 lymphocyte count declines. Additional risk factors for bacterial pneumonia include injection drug use and cigarette smoking.

PATHOBIOLOGY AND PATHOGENS

Immunodeficiency as reflected by circulating CD4 T-cell counts, as well as local abnormalities within the host defenses of the lung, confer increased risk for bacterial pneumonia. Numerous bacteria may cause pneumonia in HIV. *Streptococcus pneumoniae* is the most commonly isolated cause of community-acquired bacterial pneumonia in HIV-infected populations and is often complicated by bacteremia and invasive disease. Other frequent causes of pneumonia in HIV-infected persons include *Haemophilus* species and *Staphylococcus aureus*, the latter of which can cause community-acquired pneumonia in patients both with and without a history of injection drug use. More unusual bacterial organisms that can cause pneumonia in HIV-infected persons include *Rhodococcus equi* and *Nocardia* species. Atypical causes of community-acquired bacterial pneumonia such as *Mycoplasma* or *Legionella* species do not seem to occur with increased frequency in HIV-infected compared to uninfected patients.

CLINICAL MANIFESTATIONS

Patients typically present with acute onset of symptoms over a few days, often with cough productive of purulent sputum, fever, systemic malaise, and leukocytosis with a neutrophilic predominance. Typical radiographic findings consist of segmental, lobar, or multilobar consolidation on chest radiograph, although more diffuse reticular or interstitial patterns can occur as well. Cavitary infiltrates may result from a number of different causes, including bacteria such as *S. aureus*, *Pseudomonas*, and *Nocardia* species. Numerous cavitary infiltrates, particularly from staphylococcal species, should prompt consideration of septic pulmonary emboli. The risk of invasive disease secondary to *S. pneumoniae* is increased in those who are not on ART and who have not received pneumococcal vaccine. Other risk factors for bacteremia secondary to *S. pneumoniae* include low CD4 cell count, alcohol abuse, current smoking, recent hospitalization, and other comorbid illnesses.

DIAGNOSIS

Blood cultures should be obtained, particularly in individuals with low CD4 cell counts, and sputum cultures considered. Thoracentesis should be considered for patients with pleural effusion, particularly if they are not responding appropriately to antimicrobial therapy.

TREATMENT

Rx

The treatment of HIV-infected patients with community-acquired bacterial pneumonia should follow published guidelines.[5] Initial empirical therapy should include coverage against frequently identified organisms (e.g., *S. pneumoniae* and *Haemophilus* species) and should target both typical and atypical causes of bacterial pneumonia. Local drug resistance patterns should be taken into account, and monotherapy with a macrolide should be avoided if patients are already on macrolide prophylaxis. Macrolides should also be avoided if TB is in the differential, because resistance can develop. For patients with CD4 lymphocyte counts below 100 cells/μL or with recent hospitalization, neutropenia, broad-spectrum antimicrobial use, or underlying structural lung disease, consideration should be given to including coverage against *Pseudomonas aeruginosa*. For patients with recent health care contact, consideration should be given to antimicrobial therapy directed against microorganisms

associated with health care–associated pneumonia (e.g., more resistant gram-positive and gram-negative organisms). Empirical coverage for methicillin-resistant *S. aureus* with vancomycin should also be considered in injection drug users and other high-risk populations.

PREVENTION

Pneumococcal vaccine should be given to HIV-infected patients. Current recommendations are for the 13-valent pneumococcal conjugate vaccine (PCV13) to be given first, followed by 23-valent pneumococcal polysaccharide vaccine (PPV23) at least 8 weeks or more later (Chapter 18). For those who have already received the PPV23, the PCV13 should be given at least 1 year later. Prevention of bacterial pneumonia also includes yearly administration of inactivated influenza vaccine to all HIV-infected individuals.

PROGNOSIS

Time to clinical stability and length of stay for HIV-infected patients with community-acquired bacterial pneumonia are generally similar to HIV-uninfected patients. Although 90-day mortality has overall improved among HIV-infected patients after an episode of pneumonia when comparing the early ART era to the current era, mortality rates remain substantially higher in patients during the 1 year following pneumonia. Risk factors for increased mortality following pneumonia include a CD4 cell count below 200 cells/μL, no ART use, older age, and greater burden of comorbid illness.

Mycobacterium Tuberculosis

EPIDEMIOLOGY

TB is the most common opportunistic infection seen in HIV infection worldwide (Chapter 324). A global explosion in TB cases has resulted from the AIDS epidemic, especially in low-income countries. In contrast to many other HIV-associated infections, TB can be transmitted from person to person via an airborne route, leading to disease in HIV-uninfected individuals as well. TB is often the first manifestation of HIV and is the leading cause of death. In the United States, although TB cases have declined, TB remains an important infectious complication in HIV-infected patients, particularly in high-risk groups such as those living in communal settings or who are incarcerated, foreign-born individuals, or those with a history of TB exposure. Multidrug-resistant (defined as resistance to at least isoniazid and rifampin) and extensively drug-resistant TB (defined as multidrug-resistant TB plus resistance to any fluoroquinolone and at least one second-line injectable drug) are increasing in the HIV population, and some studies find that HIV is an independent risk factor for drug resistance. Drug resistance is associated with decreased survival, especially in HIV-infected individuals with low CD4 cell counts or those not receiving ART. ART use significantly decreases the risk of TB.

CLINICAL MANIFESTATIONS

Although HIV-infected patients may present with characteristic symptoms of TB including fever, sweats, weight loss, and cough, individuals with more advanced immunosuppression (CD4 cell counts < 350 cells/μL) often have an atypical presentation with minimal symptoms. Patients with more advanced HIV are more likely to present with disseminated disease with risk increasing the lower the CD4 cell count. Other clinical history that is suggestive of TB is a history of recent or remote TB exposure or of latent TB without adequate treatment. On physical examination, patients may have signs of lung consolidation, lymphadenopathy, or hepatosplenomegaly. Laboratory testing is often nonspecific, but may reflect underlying organ involvement. The radiographic presentation depends on the degree of immunosuppression. Patients on ART with preserved CD4 cell counts usually have a chest radiographic pattern similar to that in the HIV-uninfected population with reactivation of TB, with upper lung zone infiltrates, often with cavitation (see Fig. 391-4). In patients with more advanced immunosuppression, the chest radiograph pattern is more consistent with primary TB (see Fig. 391-5). There are diffuse infiltrates, often in the middle and lower lung zones. Cavitation is uncommon in patients with a low CD4 cell count. A normal radiograph, miliary disease, or lymphadenopathy are also seen in HIV-infected patients with low CD4 cell counts. Pleural effusions may also be seen.

DIAGNOSIS

Definitive diagnosis by either culture of *M. tuberculosis* from an affected body site or identification by nucleic-acid amplification testing (NAA) is considered the gold standard. Visualization of acid-fast bacilli (AFB) on sputum smear may be the only feasible test in some resource-poor settings; however, HIV-infected individuals are more likely to have negative smears despite active disease. Current recommendations support obtaining three sputum specimens for AFB smear and culture with NAA testing of at least one sputum specimen. Use of NAA allows for rapid identification of the mycobacterial species in smear-positive specimens, allowing determination of TB versus nontuberculous mycobacteria. It is also able to identify *M. tuberculosis* in smear-negative patients more rapidly than culture. Because NAA testing does not provide information on drug sensitivity, it should always be paired with culture. Genotype testing can also be used and rapidly identifies drug resistance mutations, but it is not available in many settings. If three spontaneous sputum samples cannot be obtained, induced sputum or bronchoscopy should be performed. Specimens from other body sites such as cerebrospinal fluid, pleural fluid, and bone marrow aspirates can be evaluated. In patients with isolated pleural effusion where TB is suspected, pleural biopsy should be considered to obtain samples for culture and sensitivity. Mycobacterial blood cultures should be obtained, particularly in individuals with CD4 cell counts below 200 cells/μL.

TREATMENT Rx

To prevent transmission of TB, empirical treatment should be started in any HIV-infected person suspected of having TB, even if smears are negative (Chapter 324). Patients with AFB detected on smears or culture should also be started on therapy directed against TB while waiting for identification of the mycobacterial species. The recommended regimen for HIV-infected individuals with suspected TB is isoniazid, rifampin, pyrazinamide, and ethambutol with pyridoxine (Table 391-4). Consultation with an expert in TB treatment is recommended in cases of drug resistance. Directly observed therapy is recommended for HIV-infected individuals with TB.

ART is an important adjunctive therapy in HIV-infected individuals with TB. Concurrent ART treatment is shown to decrease mortality, prolong AIDS-free survival, and decrease time to sputum conversion.[A] Patients receiving ART should be continued on it during TB treatment. In patients with a CD4 cell count below 50 cells/μL, randomized trials have shown that starting ART within 2 weeks of TB therapy is of benefit.[B] In individuals with higher CD4 cell counts and severe clinical disease, ART should be initiated within 2 to 4 weeks. In all other individuals (except those with TB meningitis [who should have delayed ART]), ART should be started within 8 to 12 weeks. Initiation of ART can be difficult during TB treatment because there are a myriad of drug interactions. Immune reconstitution inflammatory syndrome (IRIS) can also be seen in about one third of individuals (Chapter 395). IRIS is a paradoxical reaction that presents with a temporary worsening of clinical symptoms or radiographic findings, resulting from increased immunity against TB (see later).

PREVENTION

Prevention of primary infection with TB consists of avoiding exposure. Prevention of transmission of drug-resistant TB is of particular importance. To decrease transmission to other patients or health care workers, HIV-infected patients with any suspicion of TB should be placed in isolation until three sputum smears are negative for AFB. Improved ventilation, rapid initiation of therapy and sensitivity testing, mask use, and shifts to outpatient therapy can help decrease nosocomial transmission.

HIV-infected individuals exposed to TB who develop latent infection are at high risk of reactivation, but this risk can be significantly decreased with appropriate prophylaxis. Prevention of reactivation disease involves testing for latent TB infection (LTBI) with either a tuberculin skin test or an interferon-γ release assay (IGRA). HIV-infected persons should be tested when first diagnosed and depending on the clinical situation (Table 391-5). Individuals testing positive (≥5 mm of induration on skin test or a positive IGRA) should receive a course of prophylaxis once active infection has been ruled out. Isoniazid (300 mg daily or 900 mg twice weekly) combined with pyridoxine should be given for 9 to 12 months.[A] A recent randomized study demonstrated similar efficacy of a 3-month regimen of directly observed rifapentine and INH, with better completion rates.[B] Other regimens can also be used (see Table 391-4).

Secondary prevention of TB after treatment of active infection is not generally recommended in the United States. Some studies have found that continuation of isoniazid after completion of 6 months of TB therapy decreases risk of recurrence in high-burden countries.

PATHOBIOLOGY

Mycobacterium tuberculosis is an acid-fast bacillus that often has a characteristic beaded appearance (E-Fig. 391-1). The organism enters the lung, where it is ingested by alveolar macrophages. Defects in alveolar macrophage function in HIV can increase TB susceptibility. After entry, the host may clear the organism, become latently infected, or progress to active disease. Bacilli can also disseminate via the lymphatics and blood stream, producing disease in other organs. Typical lung pathology consists of necrotic and caseating granulomas made up of macrophages and other inflammatory cells. In individuals with HIV, classic granulomas may not be seen, particularly in those who are more immunocompromised.

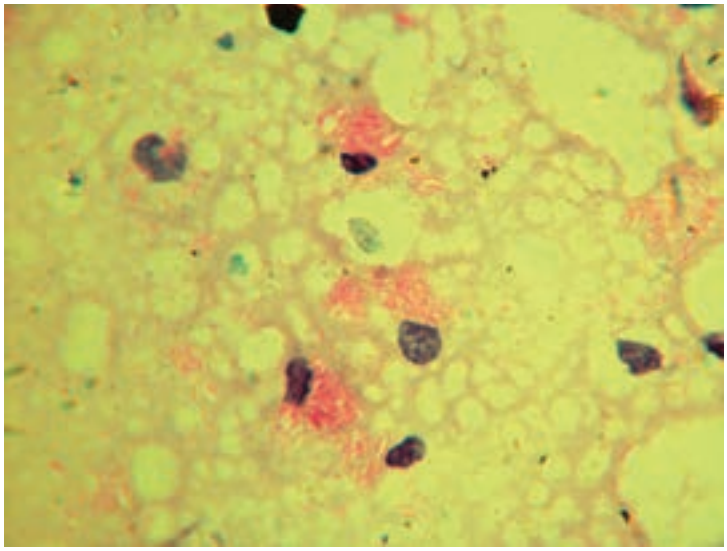

E-FIGURE 391-1. Stain of sputum sample demonstrating multiple acid-fast bacilli shown to be *Mycobacterium tuberculosis*. Courtesy Derek Armstrong, QBC Diagnostics, and Richard Chaisson.

TABLE 391-4 TREATMENT AND PREVENTION REGIMENS FOR TUBERCULOSIS IN THE HIV-INFECTED PATIENT

	DRUGS	TOTAL DURATION	COMMENTS
Drug-susceptible pulmonary tuberculosis (TB)	*Initial phase (2 months):* Isoniazid + Rifampin or rifabutin + Pyrazinamide + Ethambutol *Continuation phase (4-7 months):* Isoniazid Rifampin or rifabutin	• 6 months • 9 months if culture positive after 2 months of TB therapy	• Give isoniazid with supplemental pyridoxine • Rifabutin preferred if patient is receiving protease inhibitor • Directly observed therapy 5-7 days/week recommended
Isoniazid-resistant pulmonary TB	*Initial phase (2 months):* Rifampin or rifabutin + Pyrazinamide + Ethambutol + Moxifloxacin or levofloxacin *Continuation phase (7 months):* Rifampin or rifabutin + Ethambutol + Moxifloxacin or levofloxacin	• 9 months	• Rifabutin preferred if patient is receiving protease inhibitor • Directly observed therapy recommended
Rifamycin-resistant or other drug-resistant pulmonary TB	Regimen per resistance pattern, clinical and microbiological responses		Consultation with experts recommended
Latent TB: preferred regimens	Isoniazid 300 mg PO daily *Or* Isoniazid 900 mg PO twice weekly	• 9 months	Latent TB defined as: 1. Screening test positive with no active disease and no previous history of treatment for active or latent TB *Or* 2. Close contact of person with infectious TB and no evidence of active disease • Give isoniazid with supplemental pyridoxine • Twice-weekly isoniazid regimen should be given by directly observed therapy
Latent TB: alternative regimens	Rifampin 600 mg PO daily or rifabutin (dose adjusted based on ART agents)	• 4 months	• Rifabutin preferred if patient receiving protease inhibitor

For detailed dosing and additional information, please see Reference 5.
ART = antiretroviral therapy.

TABLE 391-5 INDICATIONS FOR TESTING FOR LATENT TUBERCULOSIS IN HIV INFECTION

INDICATION	COMMENTS
When HIV first identified	Can use either tuberculin skin test or interferon-γ release assay to diagnose latent tuberculosis
Annually if patient at continuing risk of exposure to tuberculosis	Active disease should be ruled out before starting therapy
When individuals with a negative test have an increase in CD4 cell count to > 200 cells/μL with antiretroviral therapy	

PROGNOSIS

Response to therapy and time to negative cultures depends on the setting and the immunologic state of the patient. In the United States, response to therapy and time to convert sputum cultures in HIV-infected and HIV-uninfected individuals are similar. In contrast, the risk of death and relapse in sub-Saharan Africa is higher in HIV-infected individuals than in HIV-uninfected persons. HIV-infected patients with drug-resistant TB have decreased survival, particularly those with extensively drug-resistant TB, whose mortality may be as high as 90%.

Pneumocystis Pneumonia

EPIDEMIOLOGY

PCP was the first opportunistic infection described in the AIDS epidemic and was responsible for almost two thirds of AIDS-defining diagnoses (Chapter 341). With the widespread use of PCP prophylaxis followed by the introduction of ART, PCP incidence has decreased dramatically, although it remains the most common serious opportunistic infection in the United States. Some areas of the developing world such as Asia and South America have a high incidence of PCP, but the frequency of PCP in Africa may be lower, potentially because of geographic variations in *Pneumocystis* distribution, decreased susceptibility to PCP, lack of resources to detect *Pneumocystis*, or from competing risk of death from other infections (e.g., TB, bacterial pneumonia) that occur before HIV-infected individuals reach a CD4 cell count where they are susceptible to PCP.

Most cases of PCP occur in HIV-infected individuals who are not on ART, do not adhere to PCP prophylaxis, or are unaware of their HIV status. The primary risk factor is a CD4 cell count below 200 cells/μL, with the risk increasing as the CD4 cell count falls to lower levels. Risk of PCP is also higher in individuals with oropharyngeal candidiasis or a previous history of PCP. This risk can be reversed with successful ART. Individuals on ART also have a decreased risk of primary PCP if they have a suppressed HIV viral level with a CD4 cell count between 101 and 200 cells/μL.[6]

CLINICAL MANIFESTATIONS

The classic presentation of PCP in an HIV-infected patient is fever, a nonproductive cough, and dyspnea on exertion. The course of PCP in HIV infection may be subacute with duration of symptoms of 2 weeks or more, in contrast to the often fulminant disease in patients immunocompromised for reasons other than HIV, or in contrast to the acute course of bacterial pneumonia. Pleuritic chest pain and purulent sputum are much less common in PCP than in bacterial pneumonia. Complaints of hemoptysis should prompt investigation of other diagnoses. Physical examination in PCP is usually nonspecific, and the lungs are often normal on auscultation. Crackles are the most common abnormality, and signs of focal consolidation are less often seen. Patients are often hypoxic or desaturate with exertion, but this finding is nonspecific. A CD4 cell count below 200 cells/μL is usually seen. Serum lactate dehydrogenase (LDH) is often increased in patients with PCP, but an elevated serum LDH is not sufficiently specific to establish the diagnosis of PCP, nor is a normal LDH sufficiently sensitive to rule it out. A room air arterial blood gas should be measured if feasible to determine severity of disease, and an increase in the alveolar-arterial (A-a) oxygen gradient is common. The classic radiographic presentation is bilateral reticular or granular infiltrates (see Fig. 391-3). The infiltrates start in the perihilar region and extend toward the periphery with progressive disease. The opacities are

PATHOBIOLOGY

Pneumocystis is a fungal organism that exists in trophic and cystic forms. After inhalation, the organism attaches to alveolar epithelial cells. Progression to acute pneumonia is marked by development of a foamy alveolar exudate, damage to the alveolar-capillary membrane, and inflammation. Macrophages, dendritic cells, neutrophils, and B lymphocytes all play a role in *Pneumocystis* clearance, but CD4 T lymphocytes are critical in clearance. Colonization, in which the organism or its DNA is detected in an individual without clinical pneumonia, can occur.

PCP was historically thought to result from reactivation of latent infection acquired during childhood, but it now seems likely that person-to-person transmission and de novo infection from environmental sources occur as well. Airborne transmission has been demonstrated in laboratory animals, and infection can be passed to immunosuppressed hosts via a nonimmunosuppressed carrier. The environmental niche for the organism is unknown, but its DNA has been detected in air, soil, and water samples.

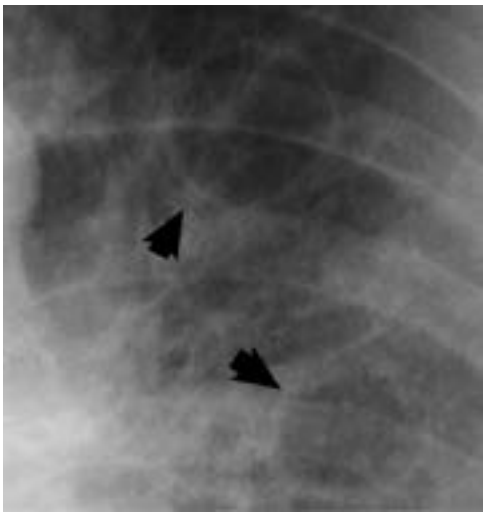

FIGURE 391-7. Close-up of the left lung from a chest radiograph of an HIV-infected person, CD4 cell count less than 200 cells/μL with *Pneumocystis* pneumonia. Two thin-walled cysts (*arrows*) are present. Courtesy Laurence Huang, MD. Used with permission.

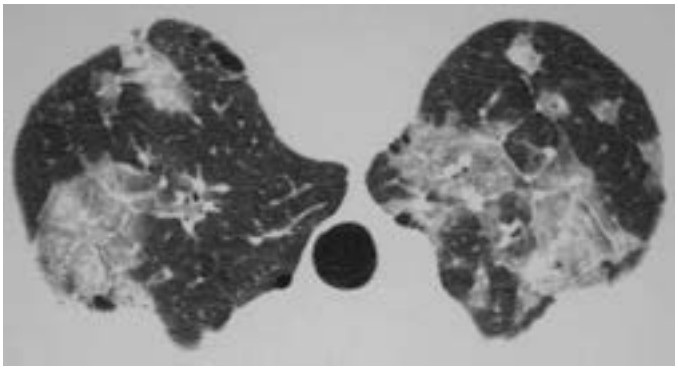

FIGURE 391-8. Chest high-resolution computed tomography (HRCT) in an HIV-infected person, CD4 cell count less than 200 cells/μL, whose chest radiograph showed normal findings. Because of clinical suspicion of *Pneumocystis* pneumonia (PCP), the patient underwent HRCT demonstrating characteristic patchy ground-glass opacities consistent with PCP. Courtesy Laurence Huang, MD. Used with permission.

typically symmetrical, but unilateral or asymmetrical disease can occur. Chest radiographs may also be normal, and atypical radiographic manifestations can occasionally be seen, including focal lobar consolidation, nodules, cavitary lesions, or miliary disease. In patients receiving aerosolized pentamidine for prophylaxis, upper lung infiltrates are more common owing to enhanced drug deposition in the lower lobes, and this presentation can be confused with TB. Pneumatoceles (thin-walled cysts) and pneumothoraces are common (Fig. 391-7). Pleural effusions and intrathoracic adenopathy are rare. Chest computed tomography (CT) can be a useful diagnostic study in patients who are suspected of having PCP, but present with a normal chest radiograph. In these patients, ground-glass opacities (Fig. 391-8) can suggest—but do not confirm—the diagnosis, because this finding is nonspecific; however, a normal high-resolution chest CT essentially rules out the diagnosis of PCP.

DIAGNOSIS

Empirical diagnosis is sometimes necessary in resource-limited settings, but definitive diagnosis with microscopic visualization of *Pneumocystis* in respiratory specimens is preferable. Because *Pneumocystis* culture cannot be performed, staining of respiratory samples is used to visualize organisms. Methenamine silver or toluidine blue O will stain the cyst wall. Giemsa and Diff-Quik staining will demonstrate both cystic and trophic forms. Immunofluorescence stains are also used and are more sensitive.

Spontaneously expectorated sputum cannot generally be used for PCP diagnosis, but induced sputum is often the first sample obtained in patients. Sensitivity of induced sputum varies by the clinical setting and experience of the institution. Repeated sputum samples are generally not obtained as they are in TB. Because a negative induced sputum cannot rule out PCP, bronchoscopy with bronchoalveolar lavage (BAL) should be performed if the initial induced sputum is negative (Chapter 341). BAL has excellent sensitivity (>90%) in the HIV-infected population, and transbronchial biopsy is rarely needed to diagnose PCP.

Newer diagnostic techniques are being investigated and may become more widespread. Polymerase chain reaction (PCR) of respiratory specimens can be used to detect *Pneumocystis* DNA in various respiratory samples, including oral washes. It is very sensitive and specific for *Pneumocystis*, but it may be difficult to distinguish between colonization and disease and is not widely available. Tests to detect elevated levels of serum 1→3, β-D-glucan (a component of fungal cell walls) can be used to support a diagnosis of PCP, but levels can also be elevated in other fungal infections.[7]

TABLE 391-6 TREATMENT REGIMENS FOR *PNEUMOCYSTIS* PNEUMONIA IN HIV INFECTION*

TREATMENT REGIMEN	DOSE(S), FREQUENCY	TOXICITIES
MILD PCP† (ARTERIAL Po_2 > 70 mm Hg *AND* ALVEOLAR-ARTERIAL O_2 DIFFERENCE < 35 mm Hg)		
Trimethoprim-sulfamethoxazole (TMP-SMX)	15-20 mg/kg (TMP component) daily (q6-8h or TMP-SMX 2 double strength tablets q8h)	Fever, dermatologic, gastrointestinal (GI), hematologic
TMP *plus* Dapsone	15 mg/kg daily (q8h) 100 mg once daily	Dermatologic, GI, hematologic
Clindamycin *plus* Primaquine	1800 mg daily (q6-8h) 30 mg (base) once daily	Dermatologic, GI, hematologic
Atovaquone	750 mg bid (with food)	Dermatologic, GI
MODERATE-SEVERE PCP‡ (ARTERIAL Po_2 < 70 mm Hg *OR* ALVEOLAR-ARTERIAL O_2 DIFFERENCE > 35 mm Hg)		
TMP-SMX	15-20 mg/kg (TMP component) daily (q6-8h)	Fever, dermatologic, GI, hematologic
Pentamidine	3-4 mg/kg once daily	Renal, pancreatic
Clindamycin *plus* Primaquine	1800-2400 mg (q6-8h) 30 mg (base) once daily	Dermatologic, GI, hematologic

*Recommended duration of therapy = 21 days.
†Oral route is preferred for patients with mild PCP who are treated as outpatients.
‡IV route is preferred (at least until clinical improvement) for patients with moderate-severe PCP. Adjunctive corticosteroids (prednisone 40 mg PO twice daily for 5 days, then 40 mg PO once daily for 5 days, then 20 mg PO once daily for 11 days or potency-equivalent methylprednisolone IV) should also be administered.
For detailed dosing and additional information, please see Reference 5.

TREATMENT

Anti-*Pneumocystis* Therapies

The first-line agent for treating PCP, regardless of severity, is trimethoprim-sulfamethoxazole (TMP-SMX) (Table 391-6). It is as effective as intravenous (IV) pentamidine but less toxic. Clinical resistance testing is not possible in PCP, but TMP-SMX can be used even if patients have been taking this agent for prophylaxis. Patients with mild PCP (Pao_2 > 70 mm Hg and A-a oxygen gradient < 35 mm Hg) who are able to tolerate oral medications can receive TMP-SMX by mouth. Patients who have moderate or severe PCP or who are unable to tolerate oral therapy should be given IV TMP-SMX. Therapy should be continued for a total duration of 21 days and may be switched to an oral regimen when the patient has clinically improved. Side effects of TMP-SMX are common in HIV-infected individuals. Frequent adverse effects include rash, nausea and vomiting, abnormal liver function tests, hyperkalemia, fever, and myelosuppression. Some toxicities of TMP-SMX can be life threatening, including Stevens-Johnson syndrome, toxic epidermal necrolysis, and a distributive shock syndrome that presents similarly to anaphylaxis, but these toxicities are rare. If side effects are mild, continuing treatment is recommended along with symptomatic treatment for adverse effects.

If patients cannot take TMP-SMX or if they develop a treatment-limiting side effect, several alternative regimens are available (see Table 391-6). IV pentamidine, clindamycin with primaquine, trimethoprim with dapsone, or atovaquone may be used. IV pentamidine is associated with a host of toxicities, including renal dysfunction, dysglycemias, pancreatitis, and torsades de pointes. Renal function and glucose levels need to be monitored during therapy; renal dysfunction and hypo- and hyperglycemia commonly occur. Clindamycin with primaquine has comparable efficacy to TMP-SMX or TMP-dapsone in mild to moderate disease. It may also be more effective than IV pentamidine in patients with severe disease. Because primaquine is administered orally, this regimen is less desirable if there is concern about effective

gastrointestinal absorption. Dapsone with trimethoprim is recommended only in mild or moderate PCP. Oral atovaquone should not be given in moderate or severe PCP and is less effective than other regimens for mild disease.

Adjunctive Corticosteroids

Randomized studies have demonstrated an increase in survival when corticosteroids are used as adjunctive therapy in PCP. Patients with an initial room air Po_2 below 70 mm Hg or an A-a gradient above 35 mm Hg should receive corticosteroids (see Table 391-6). Corticosteroids should be started at the time of PCP treatment, or at least within 72 hours of starting treatment, because they work to decrease the inflammation produced by killing of *Pneumocystis*, thus preventing deterioration in oxygenation. Corticosteroids should not be held during treatment while awaiting confirmation of the diagnosis of PCP.

Antiretroviral Therapy

Many patients with PCP are not receiving ART at the time of diagnosis. Early initiation of ART during treatment for acute PCP is supported by a randomized trial demonstrating a decrease in the end-point of progression to AIDS or mortality when ART was started within 2 weeks of diagnosis of an opportunistic infection.■ This study did not include PCP patients with respiratory failure who required mechanical ventilation. Given the risk of IRIS in these patients, optimal timing of ART is not well-defined (see "Intensive Care of the HIV-Infected Patient").

Treatment Failure

Patients with PCP often show an initial clinical worsening during treatment, with worsening oxygenation, increased radiographic infiltrates, and fever as the host reacts to dying organisms. If the patient fails to demonstrate improvement after 4 to 8 days of therapy, treatment failure should be considered. Other causes of clinical worsening, such as fluid overload, development of a pneumothorax, or presence of another infection, should be sought before deciding to change treatment.

PREVENTION

Evidence of airborne transmission and clusters of PCP outbreaks suggest that isolation of patients with PCP might prevent transmission of infection, but current guidelines do not find sufficient evidence to support recommending respiratory isolation for HIV-infected individuals with PCP. Primary prophylaxis is instituted in HIV-infected individuals with a CD4 cell count below 200 cells/μL or in those with oral candidiasis. TMP-SMX is the preferred agent for prophylaxis (Table 391-7). Dapsone, atovaquone suspension, or aerosolized pentamidine can also be used. Secondary prophylaxis should be instituted in individuals after they complete PCP treatment. In general, secondary prophylaxis of PCP is similar to primary prophylaxis, except that aerosolized pentamidine has been associated with more breakthrough PCP when used as secondary prophylaxis in individuals with a CD4 cell count below 100 cells/μL. Prophylaxis should be continued for life unless individuals experience immune reconstitution with ART. If the CD4 cell count is sustained above 200 cells/μL for at least 3 months, primary and secondary prophylaxis can be discontinued.

TABLE 391-7 PREVENTION REGIMENS FOR *PNEUMOCYSTIS* IN HIV INFECTION

PREVENTION REGIMEN	ALTERNATIVE DOSING	COMMENTS
Trimethoprim-sulfamethoxazole 1 double-strength (DS) tablet daily	1 single-strength tablet daily 1 DS tablet thrice weekly	Also effective prophylaxis against *Toxoplasma gondii* and many bacterial pathogens
Dapsone 100 mg daily		Combine with pyrimethamine and leucovorin in persons who are *T. gondii* immunoglobulin G antibody positive. Consider combining with pyrimethamine and leucovorin when used for secondary prophylaxis.
Atovaquone suspension 1500 mg daily		Improved bioavailability compared to tablets
Aerosolized pentamidine 300 mg monthly via RespirGard II nebulizer		May be associated with increased risk of extrapulmonary disease

PROGNOSIS

Prognosis for persons with PCP has improved dramatically from the beginning of the AIDS epidemic. Improvements resulted first from use of corticosteroids in more severe disease. Whether use of ART has improved short-term prognosis from PCP is debated. Changes in intensive care unit (ICU) care (e.g., use of low-tidal-volume ventilation) have also improved outcomes of severe PCP. Overall mortality from PCP is around 10%, with mortality in patients requiring mechanical ventilation approximately 30%.[8] Studies have found that survival correlates with degree of hypoxemia, underlying comorbidities, younger patient age, higher albumin, and first episode of PCP.

Other Infections

FUNGAL INFECTIONS

Although *Pneumocystis* is the most common fungal cause of pneumonia in HIV-infected patients, other endemic and nonendemic fungi may cause disease, particularly in those who are more severely immunosuppressed (e.g., CD4 cell counts < 200 cells/μL, and more often when < 100 cells/μL). Histoplasmosis can be encountered in patients who have been to the U.S. Midwest, and coccidioidomycosis in patients from the Southwest. *Cryptococcus neoformans* is found throughout the world. Of note, although invasive aspergillosis is a well-documented complication of various immunosuppressive disorders, it is uncommon in patients with HIV disease. Risk factors for the development of aspergillosis in HIV-infected individuals include use of corticosteroids, neutropenia, marijuana, and broad-spectrum antimicrobial drugs.

In more severely immunodeficient HIV-infected patients, fungal infection is often disseminated. For example, the most commonly encountered manifestation of cryptococcal disease is meningitis. Coccidioidomycosis in HIV-infected persons can present with focal or diffuse pneumonia, as well as cutaneous disease, meningitis, liver or lymph node involvement, or disseminated disease. Disseminated histoplasmosis most often presents as a febrile wasting illness in HIV-infected persons. Isolated pulmonary disease secondary to fungal infection tends to be more rare but can occur, particularly with histoplasmosis and coccidioidomycosis; isolated or focal pulmonary disease is more likely in those with CD4 cell counts above 250 to 300 cells/μL. The most commonly encountered chest radiographic findings of cryptococcal, coccidioidal, and histoplasmosis infection consist of diffuse bilateral interstitial infiltrates that are often reticular or reticulonodular. Focal consolidation, nodular opacities, cavitation, pleural effusion, and hilar adenopathy are less frequent, but may also be seen (Fig. 391-9). Radiographic findings can mimic other diseases, particularly PCP and atypical bacterial infections.

Diagnosis of pulmonary fungal infection is usually established by culture of sputum or BAL fluid and occasionally of pleural fluid; however, serologic and antigen testing can aid in the diagnosis of a variety of fungal infections. Cryptococcal antigen is a sensitive and specific test performed on serum, cerebrospinal fluid, urine, BAL fluid, or pleural fluid. The *Histoplasma* antigen test is also a sensitive method for rapid diagnosis of disease and can be obtained on urine, serum, cerebrospinal fluid, or BAL fluid. Serologic tests are useful in evaluation of suspected coccidioidomycosis, with an estimated 80 to 90% sensitivity of complement fixation and tube precipitin test for diagnosis.

VIRAL INFECTIONS

Viral pneumonias are not frequently encountered in HIV-infected patients. Although often isolated in the BAL, cytomegalovirus (CMV) is a less common cause of pneumonia (Chapter 376). Most cases of CMV disease occur in patients with a CD4 lymphocyte count less than 50 cells/μL. Symptoms of CMV pneumonia are nonspecific; chest radiograph findings include reticular or ground-glass opacities, alveolar infiltrates, and nodules or nodular opacities. The diagnosis of CMV pulmonary disease can be challenging because BAL fluid culture is not specific. Definitive diagnosis requires biopsy and demonstration of widespread specific cytopathic changes in the lungs, but biopsies may not be feasible. Patients suspected of having CMV pneumonitis should undergo a careful evaluation for evidence of disseminated disease, particularly ocular involvement. Treatment of CMV pneumonia is the same as currently recommended for disseminated disease.

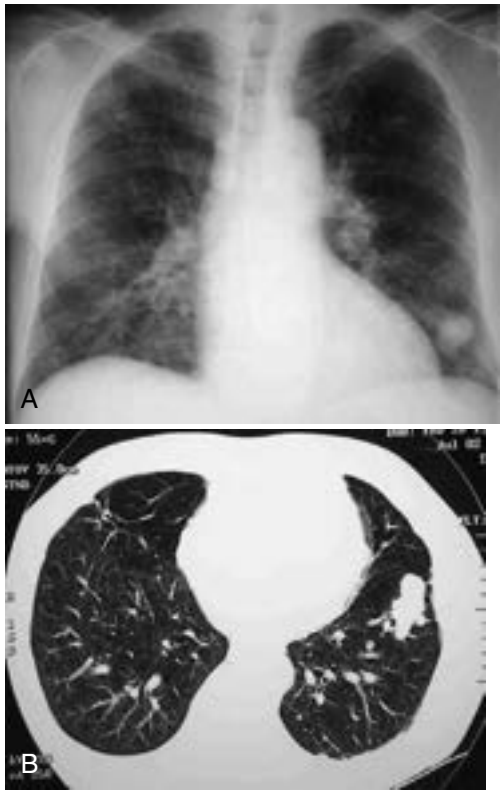

FIGURE 391-9. **Chest radiograph (A) and chest computed tomography (CT) scan (B) of an HIV-infected individual, CD4 cell count less than 200 cells/μL. A multilobulated noncalcified mass measuring 4.6 × 2.4 cm is present in the left lower lobe, with multiple adjacent satellite nodules. Culture from CT-guided transthoracic needle aspiration revealed *Cryptococcus neoformans*. Courtesy Laurence Huang, MD. Used with permission.**

Influenza generally presents similarly and has a comparable clinical course in HIV-infected as in HIV-uninfected adults (Chapter 364). HIV infection does not appear to increase the risk for influenza, although influenza-related mortality may be greater in those with AIDS compared to the general U.S. population. Prompt initiation of antiviral therapy directed against influenza is recommended. All HIV-infected persons should receive the inactivated influenza vaccine annually.

NONTUBERCULOUS MYCOBACTERIA

TB is the most common mycobacterial lung disease seen in HIV, but other nontuberculous mycobacterial pulmonary infections may occur. The two most common causes in HIV are *Mycobacterium avium* complex (MAC) and *Mycobacterium kansasii* (Chapter 325). Although the lungs may be a portal of entry for MAC, patients usually present with disseminated disease, and isolated pulmonary MAC in HIV is rare. When isolated pulmonary MAC is seen, it has been reported in patients receiving ART and thus may be a manifestation of immune reconstitution inflammatory syndrome. The chest radiograph is most often normal in MAC infection, even when cultures of respiratory specimens grow MAC. Focal infiltrates are rare. Endobronchial lesions containing MAC can sometimes be seen. MAC treatment consists of at least ethambutol and a macrolide. Additional agents can be considered in advanced immunosuppression, high mycobacterial burden, or inability to use ART. Primary MAC prophylaxis with azithromycin or clarithromycin is recommended in HIV-infected individuals with a CD4 cell count below 50 cells/μL. Secondary prophylaxis is recommended after a minimum of 12 months of treatment and can be discontinued when the CD4 cell count is above 100 cells/μL for at least 6 months.

M. kansasii is most common in the southern and central United States, with clusters of disease also reported in Europe, Asia, and Africa. It can occur at any CD4 cell count, but is most common in individuals with CD4 cell counts below 100 cells/μL. The clinical presentation and radiographic appearance of *M. kansasii* is similar to TB. Diagnostic work-up is also similar to that of TB, with the exception that identification of *M. kansasii* does not always indicate disease, because colonization can occur. Treatment consists of isoniazid, rifampin, and ethambutol with pyridoxine for a minimum of 12 months after cultures become negative. Rifabutin or a macrolide can be substituted for rifampin in patients taking a protease inhibitor or non-nucleoside reverse transcriptase inhibitor.

NONINFECTIOUS PULMONARY DISEASES

Chronic Obstructive Lung Disease

EPIDEMIOLOGY

HIV infection has been associated with several different manifestations of obstructive lung disease, including features of emphysema and chronic bronchitis, which together comprise COPD, as well as bronchial hyperresponsiveness, which characterizes asthma. Bronchiectasis causes an obstructive ventilatory defect and has been described in patients with HIV. COPD and bronchiectasis are known sequelae of severe or repeated opportunistic infection; however, evidence suggests that HIV infection is associated with an increase in COPD apart from the effects of opportunistic infections and after controlling for other traditional risk factors for COPD such as smoking and illicit drug use (Chapter 88).[9]

PATHOBIOLOGY

The pathogenesis of COPD in HIV infection is incompletely understood and likely involves multiple pathways. Respiratory tract infections, prior bacterial pneumonia or PCP, and colonization with microorganisms likely play an important role. The association of lung function with HIV specific markers suggests a potential pathogenic role for the virus itself. In comparison to HIV-uninfected persons, HIV-infected persons with poor viral control or greater degree of immunodeficiency have more severe diffusing capacity impairment.[10] Furthermore, a high viral load or low CD4 T-cell count (<100 cells/μL) have been associated with accelerated decline in airflow over time in a longitudinal cohort study of injection drug users.

CLINICAL MANIFESTATIONS

In general, the clinical presentation of chronic obstructive lung diseases is similar in HIV-infected as in uninfected persons. The majority of patients with COPD are cigarette smokers. HIV-infected individuals may be more likely to experience respiratory symptoms such as chronic cough, phlegm production, and dyspnea on exertion, for a given degree of impairment in pulmonary function, compared to HIV-uninfected individuals.

TREATMENT

Rx

In general, treatment of asthma and COPD in patients with HIV infection is similar to that for the HIV-uninfected population, although no studies have specifically examined these treatments in persons with HIV. Smoking cessation should be prioritized. Protease inhibitors, particularly ritonavir, have been reported to increase systemic levels of inhaled or intranasal fluticasone. The use of high-dose inhaled corticosteroids for COPD in patients with HIV also requires careful monitoring for oral candidiasis and pneumonia.

PROGNOSIS

No studies have specifically addressed whether the prognosis for COPD in HIV-infected patients is similar to that in HIV-uninfected patients. COPD is likely to be a major cause of morbidity and mortality as HIV-infected patients are aging.

Lung Cancer

EPIDEMIOLOGY

Whereas AIDS-defining cancers such as Kaposi sarcoma and non-Hodgkin lymphoma have decreased, non-AIDS-defining cancers in HIV-infected patients have increased, primarily among those aged 50 years and older. Lung cancer is now the most common infection-unrelated non-AIDS-defining cancer and is a leading cause of mortality in HIV-infected persons. The incidence of lung cancer is greater among HIV-infected persons than among HIV-uninfected persons.[11]

PATHOBIOLOGY

Although lung cancer can develop at any CD4 lymphocyte count, immunodeficiency is postulated as a mechanism for the enhanced risk of lung cancer associated with HIV. Prior lung disease and infections, particularly bacterial pneumonia and TB, are also risk factors in epidemiologic studies. HIV viral

load, CD4 cell count, and use of ART do not appear associated with lung cancer risk.

CLINICAL MANIFESTATIONS

Most HIV-infected patients who develop lung cancer are cigarette smokers. Although all pathologic types are seen, adenocarcinoma is the most frequent pathologic type reported; squamous cell carcinoma is the second most frequently observed pathologic type. The distribution of tumor stage at diagnosis and histologic type appear similar in HIV-infected as in uninfected patients. Radiographic appearance is also similar to that in the HIV-uninfected population (E-Fig. 391-2).

TREATMENT

Rx

Surgical resection should be considered for any patient who meets criteria based on stage of cancer and underlying medical condition. Chemotherapy and radiation may also be indicated. Significant decreases in CD4 cell counts and hematologic toxicity can occur with chemotherapy.

PROGNOSIS

Studies have been conflicting on the impact of HIV infection on survival for patients with lung cancer. Some studies suggest that survival is overall no different when stratified by stage of disease and propensity score adjustment, with an overall median survival for all stages of 7 months for HIV-uninfected controls versus 8 months for HIV-infected patients. Other studies have shown that when controlling for confounders and competing risks, HIV infection was associated with a greater risk of lung cancer–specific death, with an overall median survival of 6 months among non–small cell lung cancer patients with HIV compared to 20 months in patients without evidence of HIV. HIV-infected patients may less frequently receive lung cancer treatments, potentially accounting for survival differences. Whether lung cancer screening with an annual low-dose CT scan will have a beneficial effect on mortality in HIV-infected smokers, as it has in older HIV-uninfected heavy smokers, is not yet known.[12,13]

Other Thoracic Malignancies

The most common HIV-associated malignancy is Kaposi sarcoma (KS), although its incidence has decreased dramatically with combination ART. KS is an angioproliferative tumor and most commonly presents with mucocutaneous involvement. The lymph nodes, gastrointestinal tract, and lungs can also be involved. HIV-associated KS is substantially more likely in men who have sex with men than in other HIV risk groups. Human herpesvirus 8 (HHV-8) has been found in all forms of KS and appears to play a central role in pathogenesis.

Pulmonary KS presents in most patients when their CD4 T-cell counts are below 200 cells/μL. Respiratory symptoms typically are nonspecific and include nonproductive cough, dyspnea, and fever. Most but not all patients with pulmonary KS have concomitant mucocutaneous disease. Typical chest radiograph findings of pulmonary KS consist of bilateral perihilar or central opacities, as well as linear densities, nodular opacities, pleural effusions, and intrathoracic lymphadenopathy (Fig. 391-10A). The diagnosis of pulmonary KS can often be established by visualization of characteristic lesions on bronchoscopy. Endobronchial lesions from KS are flat or slightly raised, red or violaceous lesions (see Fig. 391-10B), and their presence is often sufficient to establish a presumptive diagnosis without requiring biopsy. HHV-8 may be detected in BAL. Tumors can regress in size and number in response to ART, and therefore all patients with KS should receive combination ART if no other contraindications exist. Treatment of more advanced systemic disease also includes chemotherapy. Of malignancies, KS is the most likely to be associated with IRIS in patients who initiate ART.

Almost all HIV-associated non-Hodgkin lymphomas (NHL) are of B-cell origin (Chapter 185). The majority are associated with Epstein-Barr virus (EBV) infection. As with KS, the incidence of NHL has declined dramatically in the era of combination ART. Most HIV-infected patients with NHL present with disseminated disease and extranodal involvement at diagnosis. Clinically apparent pulmonary involvement occurs in up to 30% of patients with AIDS-related NHL. Occasionally, the lung is the only site involved. Although NHL can present at a wide range of CD4 lymphocyte counts, most patients have advanced HIV infection, with median CD4 T-cell counts around 100 cells/μL.

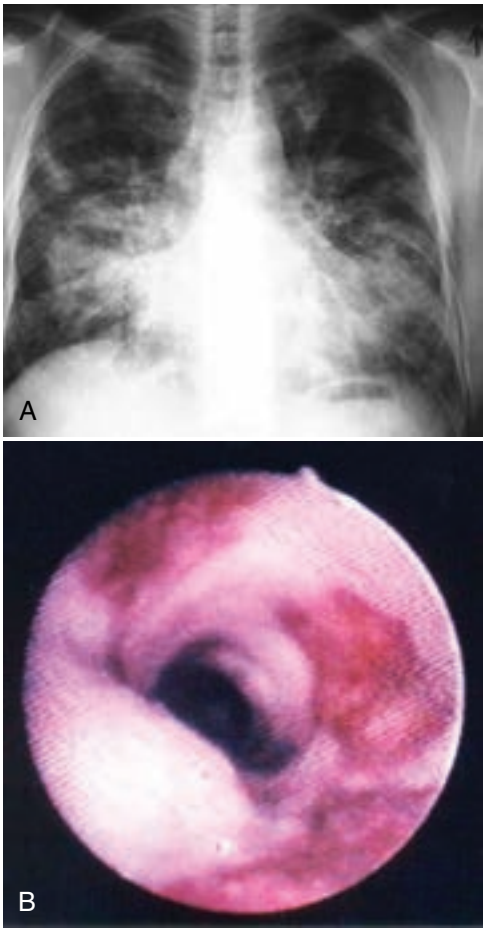

FIGURE 391-10. Chest radiograph (A) of an HIV-infected patient, CD4 cell count less than 100 cells/μL, with characteristic bilateral middle and lower lung zone, predominantly central, distribution of abnormalities of Kaposi sarcoma. The patient had no evidence of mucocutaneous disease, and the diagnosis was made by visualization of characteristic erythematous violaceous Kaposi sarcoma throughout the airways on bronchoscopy (**B**). **B**, courtesy Laurence Huang, MD. Used with permission.

Common presenting features of thoracic involvement with NHL are nonspecific and include cough and dyspnea. Systemic symptoms such as fever, sweats, and weight loss are common. Chest radiograph findings typically include single or multiple nodules, nodular opacities or masses, lobar infiltrates, and diffuse interstitial infiltrates; pleural effusions and intrathoracic lymphadenopathy are common accompanying findings. The diagnosis of NHL requires demonstration of malignant lymphocytes on cytology or biopsy specimens. Most often, the diagnosis is made by biopsy of an extrathoracic site. The yield of pleural cytology is significantly higher in HIV-associated pulmonary lymphoma as opposed to HIV-uninfected cases. Pulmonary involvement in NHL is treated as part of the systemic disease (Chapter 185).

Interstitial Lung Disease

EPIDEMIOLOGY

Interstitial lung disease is a less common cause of pulmonary disease than infections, malignancies, or obstructive lung diseases. HIV-infected individuals may develop a number of diagnoses seen in the general population such as nonspecific interstitial pneumonitis (NSIP), sarcoidosis, and hypersensitivity pneumonitis. Sarcoidosis is sometimes associated with initiation of ART, and it is debated whether it is a manifestation of IRIS. Lymphocytic interstitial pneumonia (LIP) is a disease entity that is more common in HIV-infected children than adults, and its incidence has decreased significantly since introduction of ART.

PATHOBIOLOGY

Causes of interstitial lung diseases are poorly understood. Pathology is similar to that seen in the HIV-uninfected population (Chapter 92). It is unclear whether LIP represents a unique pathologic disorder or results from multiple

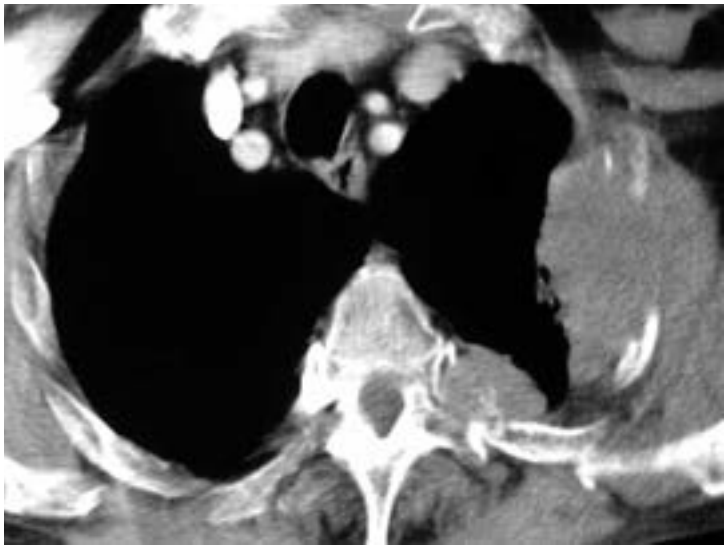

E-FIGURE 391-2. Chest computed tomography (CT) in an HIV-infected smoker with a CD4 cell count of 200 cells/μL and HIV viral level below 50 copies/mL, with chronic obstructive pulmonary disease. CT-guided biopsy demonstrated squamous cell cancer. Courtesy Laurence Huang, MD. Used with permission.

causes that lead to similar lung pathology. Lung biopsy in LIP demonstrates interstitial lymphocytes with spreading into the alveolar septae.

CLINICAL MANIFESTATIONS

Manifestations vary depending on the specific diagnosis, but symptoms often include progressive shortness of breath and nonproductive cough. Patients may be febrile. Physical examination may reveal crackles on lung examination. Systemic findings such as lymphadenopathy in sarcoidosis can provide clues to specific etiologies. Radiographic manifestations vary with the disease. Pulmonary function testing generally shows a pattern consistent with restrictive lung disease, and diffusing capacity for carbon monoxide is often impaired.

DIAGNOSIS

BAL is a useful initial procedure for ruling out infectious causes of lung disease, but definitive diagnosis of interstitial lung disease generally requires a lung biopsy. In most cases, a video-assisted thoracoscopic lung biopsy is preferred to transbronchial biopsy, although in the case of sarcoidosis, transbronchial biopsies are useful in establishing a diagnosis.

TREATMENT Rx

There have been no treatment trials of interstitial lung diseases in HIV-infected patients, and treatment should generally follow guidelines for the HIV-uninfected population (Chapter 92). In LIP, initiation of ART is recommended, with consideration of corticosteroids depending on the clinical severity.

PROGNOSIS

Prognosis is generally the same as that in the HIV-uninfected population. In a small number of cases, LIP may be a precursor to lymphoma, which carries a worse outcome, however LIP usually has a better prognosis in HIV infection because it may respond to initiation of ART.

Pulmonary Hypertension

EPIDEMIOLOGY

HIV is recognized as a cause of World Health Organization (WHO) Group 1 pulmonary arterial hypertension (PAH). PAH has been reported to occur with a prevalence of 0.5% in the HIV-infected population, compared to 1 to 2 cases per million in HIV-uninfected individuals.[14] HIV-infected individuals also may have additional risk factors for PAH, including IV drug use or liver disease.

PATHOBIOLOGY

The hallmark pathologic manifestation of HIV-PAH is the plexiform lesion. More subtle changes such as medial hypertrophy and intimal hyperplasia of pulmonary arterioles can also be seen. The exact causes of PAH in HIV are not well understood, but may involve genetic or environmental insults in addition to HIV.

CLINICAL MANIFESTATIONS

PAH in the HIV-infected patient presents similarly to PAH in the HIV-uninfected individual, although patients are often younger and have a better New York Heart Association functional class at presentation. Initial symptoms are often nonspecific, but patients eventually may report symptoms of dyspnea with exertion, syncope, or exertional chest pain. Physical examination can be normal or suggestive of right-sided heart failure.

DIAGNOSIS

The gold standard for diagnosis of PAH is right heart catheterization. Echocardiography is often performed as a screening test, but pulmonary arterial pressures on echocardiography may not correlate with pressures on catheterization. Other causes of pulmonary hypertension such as underlying lung disease, left-sided heart failure, or cirrhosis should be evaluated as well.

TREATMENT Rx

There have been few trials of standard PAH agents in HIV-infected populations, but in general, guidelines for PAH treatment in the HIV-uninfected population should be followed (Chapter 68). Epoprostenol may be associated with a higher risk of intravascular infections in HIV-infected individuals, especially those who are active IV drug users. Sildenafil levels can increase when used in combination with antiretrovirals, particularly with regimens that include ritonavir. Tadalafil appears to have less interaction with ritonavir. Treatment with warfarin can also be used in addition to PAH-specific therapies. Although results of studies are conflicting, there may be some benefit to ART in this disease, and patients should be started on ART if they are not already receiving it.

PROGNOSIS

One-year survival for HIV-infected individuals with PAH is 58 to 88% and varies with the cohort studied. Lower CD4 cell count and higher HIV viral levels are associated with worse survival.

Intensive Care of the HIV-Infected Patient

The epidemiology of ICU care for HIV-infected ICU patients has shifted multiple times over the course of the AIDS epidemic. A recent study of HIV-infected veterans with a first ICU admission between 2002 and 2010 found that 15% of HIV-infected individuals had a medical or cardiac ICU admission, compared to 10% of HIV-uninfected individuals.[15] HIV-infected patients were also younger, had a longer length of stay, and were more likely to require mechanical ventilation. Another study of over 2500 HIV-infected individuals found that 4.2% required ICU admission over a median of 2.2 years of follow-up from an AIDS diagnosis; a diagnosis of HIV after the CD4 cell count had fallen below 350 cells/μL and low CD4 cell count were associated with increased incidence of ICU admission.[16] ART use was protective for ICU admission.

HIV-infected patients may be admitted to the ICU with conditions related to their HIV, with side effects of ART, or for reasons unrelated to HIV. Currently, about half of ICU admission diagnoses in this population are directly related to HIV (e.g., opportunistic infections, malignancies). HIV-infected individuals may also be admitted with common conditions such as myocardial infarction, the risk for which may be increased in HIV-infected individuals on ART. Life-threatening complications of ART such as lactic acidosis or severe IRIS can also precipitate ICU admission. In recent series, respiratory failure is still the most common cause of ICU admission, with PCP, TB, and bacterial pneumonia causing varying degrees of disease depending on the population. The outcome of HIV-infected ICU patients with acute respiratory distress syndrome (Chapter 104) now is similar to HIV-negative patients.[17] In about 20 to 40% of cases, HIV infection is not known at the time of ICU admission, and providers need to consider HIV infection in their differential even if HIV infection has not been confirmed.

Antiretroviral Therapy Use

Among HIV-infected patients admitted to the ICU, clinicians need to be aware of numerous antiretroviral drug toxicities, drug-drug interactions, and concerns regarding absorption and metabolism of specific medications. Side effects of ART are numerous. Several toxicities secondary to ART that may require ICU admission include acute liver failure, lactic acidosis, pancreatitis, and hypersensitivity reactions.

In HIV-infected ICU patients, clinicians need to decide whether to continue or initiate ART. In general, expert opinion is that ART should be continued in the ICU for critically ill HIV-infected patients already on ART, if ART can be safely administered with minimal drug-drug interactions or risk of toxicities while receiving ICU medications. For patients admitted to the ICU for AIDS-related diseases, initiation of ART should be strongly considered in consultation with an infectious disease specialist. Survival is improved when ART is started concurrently with or in close proximity to treatment of opportunistic infection, but existing data do not include critically ill patients with respiratory failure that is severe enough to require mechanical ventilation. Nevertheless, HIV-infected patients admitted for non-AIDS-related conditions but with prolonged ICU hospitalizations or severely decreased CD4 T-cell count should also be considered for ART initiation.

HIV-infected patients who initiate ART, particularly in close proximity to treatment for an opportunistic infection, are at risk for development of IRIS. IRIS is a paradoxical worsening of clinical status typically related to recovery of the immune system after immunosuppression (Chapter 395). IRIS occurs as a result of host inflammatory responses to previously recognized or subclinical infections, or to cancer or self-antigens. Manifestations of IRIS that can result in critical illness include pneumonitis, meningitis, hepatitis, and pericarditis. Respiratory failure secondary to IRIS is most often associated

with TB and PCP. Other causes of clinical worsening such as a new infection, treatment failure, or drug reaction should be ruled out. IRIS is usually self-limited, and ART should be continued. Nonsteroidal anti-inflammatories can provide symptomatic relief.■

Grade A References

A1. Abdool Karim SS, Naidoo K, Grobler A, et al. Integration of antiretroviral therapy with tuberculosis treatment. *N Engl J Med*. 2011;365:1492-1501.

A2. Abdool Karim SS, Naidoo K, Grobler A, et al. Timing of initiation of antiretroviral drugs during tuberculosis therapy. *N Engl J Med*. 2010;362:697-706.

A3. Blanc FX, Sok T, Laureillard D, et al. Earlier versus later start of antiretroviral therapy in HIV-infected adults with tuberculosis. *N Engl J Med*. 2011;365:1471-1481.

A4. Havlir DV, Kendall MA, Ive P, et al. Timing of antiretroviral therapy for HIV-1 infection and tuberculosis. *N Engl J Med*. 2011;365:1482-1491.

A5. Rangaka MX, Wilkinson RJ, Boulle A, et al. Isoniazid plus antiretroviral therapy to prevent tuberculosis: a randomised double-blind, placebo-controlled trial. *Lancet*. 2014;384:682-690.

A6. Sterling TR, Villarino ME, Borisov AS, et al. Three months of rifapentine and isoniazid for latent tuberculosis infection. *N Engl J Med*. 2011;365:2155-2166.

A7. Zolopa A, Andersen J, Powderly W, et al. Early antiretroviral therapy reduces AIDS progression/death in individuals with acute opportunistic infections: a multicenter randomized strategy trial. *PLoS ONE*. 2009;4:e5575.

A8. Meintjes G, Wilkinson RJ, Morroni C, et al. Randomized placebo-controlled trial of prednisone for paradoxical tuberculosis-associated immune reconstitution inflammatory syndrome. *AIDS*. 2010;24:2381-2390.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

Prognosis for HIV-infected ICU patients has improved dramatically over the course of the AIDS epidemic. Many studies demonstrate mortality comparable to the HIV-uninfected population, although a study of HIV-infected veterans found that 30-day mortality was higher in those with HIV infection. Factors that predict increased mortality include an AIDS-associated diagnosis, need for mechanical ventilation, low serum albumin, and a diagnosis of PCP. In PCP patients, specific mortality risks include mechanical ventilation and development of a pneumothorax.

GENERAL REFERENCES

1. Beck JM. Abnormalities in host defense associated with HIV infection. *Clin Chest Med*. 2013;34:143-153.
2. Jambo KC, Banda DH, Afran L, et al. Asymptomatic HIV-infected individuals on antiretroviral therapy exhibit impaired lung CD4(+) T-cell responses to mycobacteria. *Am J Respir Crit Care Med*. 2014;190:938-947.
3. Gingo MR, Balasubramani GK, Kingsley L, et al. The impact of HAART on the respiratory complications of HIV infection: longitudinal trends in the MACS and WIHS cohorts. *PLoS ONE*. 2013;8:e58812.
4. Crothers K, Huang L, Goulet JL, et al. HIV infection and risk for incident pulmonary diseases in the combination antiretroviral therapy era. *Am J Respir Crit Care Med*. 2011;183:388-395.
5. Panel on Opportunistic Infections in HIV-Infected Adults and Adolescents. Guidelines for the prevention and treatment of opportunistic infections in HIV-infected adults and adolescents: recommendations from the Centers for Disease Control and Prevention, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. Available at http://aidsinfo.nih.gov/contentfiles/lvguidelines/adult_oi.pdf; Accessed January 28, 2015.
6. Mocroft A, Reiss P, Kirk O, et al. Is it safe to discontinue primary *Pneumocystis jiroveci* pneumonia prophylaxis in patients with virologically suppressed HIV infection and a CD4 cell count < 200 cells/microL? *Clin Infect Dis*. 2010;51:611-619.
7. Sax PE, Komarow L, Finkelman MA, et al. Blood (1->3)-beta-D-glucan as a diagnostic test for HIV-related *Pneumocystis jiroveci* pneumonia. *Clin Infect Dis*. 2011;53:197-202.
8. Miller RF, Huang L, Walzer PD. *Pneumocystis* pneumonia associated with human immunodeficiency virus. *Clin Chest Med*. 2013;34:229-241.
9. Drummond MB, Kirk GD. HIV-associated obstructive lung diseases: insights and implications for the clinician. *Lancet Respir Med*. 2014;2:583-592.
10. Fitzpatrick ME, Gingo MR, Kessinger C, et al. HIV infection is associated with diffusing capacity impairment in women. *J Acquir Immune Defic Syndr*. 2013;64:284-288.
11. Lambert AA, Merlo CA, Kirk GD. Human immunodeficiency virus-associated lung malignancies. *Clin Chest Med*. 2013;34:255-272.
12. Hulbert A, Hooker CM, Keruly JC, et al. Prospective CT screening for lung cancer in a high-risk population: HIV-positive smokers. *J Thorac Oncol*. 2014;9:752-759.
13. Sigel K, Wisnivesky J, Shahrir S, et al. Findings in asymptomatic HIV-infected patients undergoing chest computed tomography testing: implications for lung cancer screening. *AIDS*. 2014;28:1007-1014.
14. Barnett CF, Hsue PY. Human immunodeficiency virus-associated pulmonary arterial hypertension. *Clin Chest Med*. 2013;34:283-292.
15. Akgun KM, Tate JP, Pisani M, et al. Medical ICU admission diagnoses and outcomes in human immunodeficiency virus-infected and virus-uninfected veterans in the combination antiretroviral era. *Crit Care Med*. 2013;41:1458-1467.
16. Shrosbree J, Campbell LJ, Ibrahim F, et al. Late HIV diagnosis is a major risk factor for intensive care unit admission in HIV-positive patients: a single centre observational cohort study. *BMC Infect Dis*. 2013;13:23.
17. Nirappil FJ, Maheshwari A, Andrews J, et al. Characteristics and outcomes of HIV-1-infected patients with acute respiratory distress syndrome. *J Crit Care*. 2015;30:60-64.

REVIEW QUESTIONS

1. Which of the following pulmonary complications is most likely in an HIV-infected person with a CD4 lymphocyte count above 500 cells/μL?

A. *Streptococcus pneumoniae* pneumonia
B. Disseminated *Mycobacterium tuberculosis*
C. *Pneumocystis* pneumonia
D. Pulmonary Kaposi sarcoma
E. Lung cancer

Answer: A At a CD4 lymphocyte count above 500 cells/μL, HIV-associated complications such as *Pneumocystis* pneumonia (PCP) and pulmonary Kaposi sarcoma are unlikely. Among the remaining choices, *S. pneumoniae* pneumonia is more frequent than either disseminated *M. tuberculosis* (which less commonly presents with disseminated disease at a CD4 count above 500 cells/μL) or lung cancer.

2. *Pneumocystis* pneumonia may present with which of the following chest radiographic findings?

A. Normal chest radiograph
B. Focal consolidation
C. Bilateral reticular or granular opacities
D. Pneumatoceles
E. All of the above

Answer: E PCP can present with a wide range of chest radiographic findings, including all of the abnormal findings listed, as well as a normal radiograph.

3. A 40-year-old HIV-infected man presents to the emergency room with cough productive of purulent sputum and fevers for 2 days. He is homeless, smokes cigarettes, and has a history of alcohol abuse. He denies injection drug use and has never had pneumonia previously. He has not been on antiretroviral therapy (ART) nor antibiotic prophylaxis; his most recent CD4 T-cell count 3 months ago was 310 cells/μL. He has not been in health care contact since that time. He has no known drug allergies. On examination, he is alert and in no acute respiratory distress. His temperature is 38.4° C, blood pressure is 130/80, heart rate 95, respirations 15, and oxygen saturation of 96% on room air. His chest x-ray shows a right middle lobe consolidation. Which of the following should be recommended initially:

A. Obtain a chest computed tomography (CT) scan with IV contrast.
B. Initiate treatment with ceftriaxone and azithromycin.
C. Initiate treatment with vancomycin and piperacillin/tazobactam.
D. Initiate treatment with IV trimethoprim-sulfamethoxazole.
E. Initiate treatment with isoniazid, rifampin, ethambutol, and pyrazinamide.

Answer: B The acute onset of illness with a focal opacity on chest radiograph is consistent with bacterial pneumonia, and the patient's history of cigarette smoking and alcohol abuse in addition to HIV infection increase his risk for bacterial pneumonia. The results of a chest CT are unlikely to influence initial treatment decisions in this case. Although opportunistic infections are possible, they are less likely because his CD4 count recently was over 200, and etiologies such as PCP or tuberculosis (TB) are less likely, given the acuity of symptoms and radiographic findings. He does not have risk factors for health care–associated pneumonia or known underlying lung disease, thus coverage for community-acquired pneumonia with ceftriaxone and azithromycin would be an appropriate first regimen.

4. You are asked to evaluate a 58-year-old HIV-infected man who was initially admitted to the hospital 15 days ago with fevers, chills, night sweats, weight loss, and cough. He has a history of TB treated 30 years ago. On admission, he was not on ART and was found to have a CD4 T-cell count of 10 cells/μL and an HIV viral load of over 100,000 copies/mL. A sputum induction was negative for acid-fast bacilli but positive for *Pneumocystis*. He was treated with trimethoprim-sulfamethoxazole, and after 8 days was improving. He was not given corticosteroids. He was then started on highly active antiretroviral therapy (ART) and discharged. He now presents with recurrent fever and increased dyspnea. His chest x-ray on admission showed faint bilateral, hazy air space opacities. Now 15 days later at his second hospitalization, his chest x-ray reveals increased dense bilateral air space opacities. What is the next most appropriate step in this patient's management:

A. Start rifamycin, isoniazid, pyrazinamide, and ethambutol.
B. Switch from trimethoprim-sulfamethoxazole to pentamidine.
C. Stop HAART.
D. Start Solu-Medrol.
E. Bronchoscopy with bronchoalveolar lavage

Answer: E The differential diagnosis to consider in this patient who presents with worsening of respiratory symptoms and air space opacities includes inadequately treated or nonresponding PCP, a new opportunistic infection, a noninfectious process such as congestive heart failure, or immune reconstitution inflammatory syndrome (IRIS). Because his initial diagnosis of PCP was based on sputum induction, a bronchoscopy would be indicated to assess for other infections before diagnosing IRIS, which is ultimately what this patient was thought to have. He improved with the addition of a brief course of corticosteroids, completion of 21 days of trimethoprim-sulfamethoxazole, and continuation of his ART.

5. When should ART be started in an HIV-infected person with a new diagnosis of TB and a CD4 cell count below 50?

A. After completion of TB treatment
B. Within 2 weeks of starting TB treatment
C. After 1 month of TB treatment
D. When sputum cultures are no longer positive for TB
E. When patient is no longer febrile

Answer: B Although initiation of ART used to be delayed until after TB therapy, owing to drug interactions and the concern for the immune reconstitution syndrome, several randomized studies have shown decreased mortality, greater AIDS-free survival, and more rapid conversion of sputum smears and culture with early initiation of ART. Thus, all HIV-infected individuals with TB should receive ART. If the patient is naïve to ART, it is recommended that ART be initiated within 2 weeks of TB treatment initiation in individuals with CD4 cell counts below 50 cells/μL and within 8 to 12 weeks in all others. Individuals receiving ART should continue its use during TB therapy.

392

SKIN MANIFESTATIONS IN PATIENTS WITH HUMAN IMMUNODEFICIENCY VIRUS INFECTION

TOBY MAURER

From the beginning of the epidemic, skin disease has often been the initial feature of infection with human immunodeficiency virus (HIV). Cutaneous manifestations are often predictive of the stage of immunosuppression and long-term prognosis.[1] Morbidity from skin diseases, particularly from opportunistic infections, has decreased with the advent of antiretroviral therapy (ART), although there are still significant dermatologic problems for patients in the post-ART era (Table 392-1). Cutaneous manifestations seen in patients with HIV may be categorized as infectious, neoplastic, inflammatory, or related to ART.

INFECTIOUS MANIFESTATIONS

HIV Seroconversion Exanthem

Acute retroviral syndrome (ARS) is an infectious mononucleosis–like illness consisting of fever, lymphadenopathy, pharyngitis, and neurologic symptoms.[2] It can be the presenting sign of primary HIV infection in 40 to 90% of patients. The rash typically presents as a nonspecific morbilliform eruption affecting the upper part of the trunk and face, with relative sparing of the peripheries. It is often accompanied by aphthous and penile ulcerations. There is evidence that very early after infection, HIV goes to dendritic cells in skin and causes an inflammatory response, which may explain the rash. Patients with primary HIV infection may be highly infectious because of the presence of a high viral burden in blood and genital secretions. The HIV antibody test, however, is likely to be negative because ARS precedes seroconversion by 2 to 6 weeks. Confirmation of ARS can be made by finding positive plasma HIV RNA in the setting of a negative HIV antibody. The potential benefits to public and individual health may now justify treatment of patients with acute HIV infection, particularly those who are symptomatic.[3] Initiation of ART in this phase may decrease disease severity, alter the initial viral set point, and decrease viral replication so as to reduce the risk HIV transmission.

TABLE 392-1 SKIN CONDITIONS COMMON IN PATIENTS WITH CD4$^+$ COUNTS LESS THAN 200 CELLS/μL AND NOT RECEIVING ANTIRETROVIRAL THERAPY

Photodermatitis
Psoriasis, difficult to control or involving more than 50% of the body
Pruritic papular eruption associated with HIV infection
Prurigo nodularis
Oral hairy leukoplakia
Molluscum contagiosum
Eosinophilic folliculitis

Viruses

HERPES ZOSTER

Herpes zoster is 10 times more common in HIV patients than in the general population.[4] Because varicella virus (VZV) reactivation can be observed even at relatively normal CD4$^+$ counts, herpes zoster may be the first feature of HIV infection. Herpes zoster in the setting of HIV may affect more than one dermatome, involve unusual locations, or become disseminated. Cutaneous dissemination often presents with multiple and monomorphic vesicular skin lesions in a generalized distribution. These can affect a number of distinct dermatomes and often cross the midline. Complications such as blindness (when the cranial nerve V1 distribution is affected), post-herpetic neuralgia, myelitis, encephalitis, and visceral involvement can cause significant morbidity. HIV patients with limited disease usually respond well to oral antiviral therapy, and antiviral treatment should be started at any time during the disease course (not only within the first 48 hours, as with herpes zoster not associated with HIV). Intravenous acyclovir should be considered for disseminated disease or for lesions involving the eye. There have been a few reports of a verrucous chronic form of zoster occurring in severely immunosuppressed patients; this form of zoster is usually associated with acyclovir resistance. Zoster has been one of the more commonly reported diseases in the immune reconstitution inflammatory syndrome (IRIS) (Chapter 395) and typically appears during the second stage (≈3 months after initiation of ART). The zoster vaccine is not recommended in patients with CD4 counts below 200 cells/μL. Use of the vaccine in patients with CD4 counts above 200 cells/μL is controversial because the safety and efficacy of zoster vaccination in this population has not been demonstrated.[5]

HERPES SIMPLEX VIRUS

Infection with herpes simplex virus (HSV) should always be considered in HIV patients with mucocutaneous ulcerations (Fig. 392-1), particularly when localized to the anogenital area. The presentation is generally typical and characterized by multiple tiny vesicles, punched-out erosions, or crusts. Atypical or extensive ulcerative, vegetative, or tumor-like lesions may also occur. Patients with CD4$^+$ counts below 200 cells/μL may require higher doses of acyclovir to treat the infection. Patients who fail to respond may have HSV with acyclovir resistance, and alternative agents such as foscarnet or cidofovir may be required. In those unable to tolerate the toxicity profile of

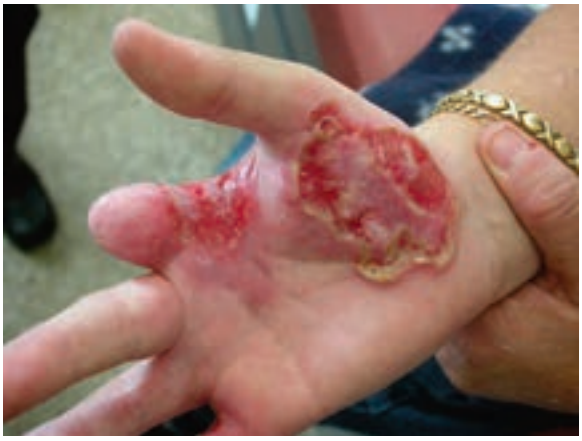

FIGURE 392-1. Chronic herpes simplex infection.

these agents, topical or intralesional cidofovir has been reported to be beneficial.[6]

HUMAN PAPILLOMAVIRUS

Human papillomavirus (HPV) infection (Chapter 373) is very common in all stages of HIV infection. Common, flat, genital, and plantar warts are all seen with increased frequency. Lesions are often more extensive and more difficult to treat than in immunocompetent patients. The warts themselves are rarely symptomatic unless they are on the soles of the feet or around the fingernails, where they may cause excruciating pain. Severely immunosuppressed patients may have extensive warts that are recalcitrant to standard treatment. Although the advent of ART has not reduced the number or severity of warts, most patients commencing on ART often notice considerable improvement or resolution of existing lesions. Some patients receiving ART will not clear their warts despite improvement in the $CD4^+$ count, suppression of viral load, and resolution of all other opportunistic infections. The reasons for this observation remain obscure. Treatment modalities include cryotherapy, surgery, laser, topical salicylic acid preparations, podophyllin, intralesional bleomycin, and contact sensitization immunotherapy. Immunomodulatory therapy in the form of imiquimod has had less promising results in HIV patients. Topical cidofovir has been used, but the high cost of formulation limits its use. Owing to the frequent presence of high-risk HPV types, there is a risk of malignant transformation in HIV-infected patients. Anal carcinoma and anal intraepithelial neoplasia (AIN) are frequent (men and women) and may develop more rapidly in this population. Regular screening for cervical neoplasia should be performed in HIV-infected women, and screening anal pap smears may be helpful in identifying anal dysplasia in men. HPV vaccination appears to be safe and immunogenic in the presence of HIV and is reasonable to consider in persons aged 9 to 26 years (as is suggested in the HIV-uninfected population). Investigations are ongoing as to whether HPV vaccination will demonstrate benefit in the broad group of HIV-infected patients older than 26.

MOLLUSCUM CONTAGIOSUM

Molluscum contagiosum is a poxvirus that causes a self-limited papular umbilicated eruption commonly seen in children. It is rarer to see in adults, and when on the face, persistent, or severe, it raises the possibility of HIV coinfection. It tends to occur when the $CD4^+$ count falls below 200. First-line treatment is ART because molluscum contagiosum is invariably cleared with rising $CD4^+$ counts. Ablative treatments such as cryotherapy or curettage will treat individual lesions more expeditiously; however, without ART the lesions are likely to recur.

EPSTEIN-BARR VIRUS (ORAL HAIRY LEUKOPLAKIA)

Oral hairy leukoplakia (OHL) is characterized by nonpainful white plaques with a feathered edge, particularly on the lateral border of the tongue. It is associated with Epstein-Barr virus and is very rare in immunocompetent hosts. OHL can occur at any $CD4^+$ count and may therefore be the initial feature of HIV infection. Table 392-2 lists diseases not affected by the $CD4^+$ count. The appearance of OHL was a poor prognostic indicator before ART. There is no specific treatment of OHL, although it tends to resolve when patients are taking ART. Superinfection with *Candida* species should be considered if the lesions are painful.

Bacteria

STAPHYLOCOCCUS AUREUS

Infections with *Staphylococcus aureus* are commonly seen in HIV disease. Staphylococcal folliculitis (Fig. 392-2) tends to be more severe or refractory to treatment than in patients who are HIV negative. Other cutaneous manifestations of staphylococcal infection include cellulitis, bullous impetigo, abscesses, ecthyma (necrotic plaques), or rarely, botryomycosis. HIV-infected persons are at a heightened risk of infection with methicillin-resistant *S. aureus* (MRSA) and appear to have increased susceptibility for recurrence.[7] Choice of treatment is dictated by location and severity of infection. The use of antiseptic washes such as benzoyl peroxide or chlorhexidine solution may have a role in the treatment of *S. aureus* folliculitis, but these agents can dry out the skin and lead to eczematous eruptions prone to secondary bacterial infections. Systemic antibiotics can be used, and susceptibility testing (or knowledge of local resistance patterns) should guide antibiotic selection. The use of topical mupirocin to reduce MRSA carriage is controversial because it does not appear to reduce infection rates. Abscesses should be incised and drained. Rifampin can be used in combination with other antibiotics to reduce carriage rates of *S. aureus*, but caution should be used in patients receiving ART, because rifampicin is a potent inducer of cytochrome P-450 and will therefore interfere with protease inhibitors. More severe infections may require intravenous antibiotics.

BARTONELLA

Bacillary angiomatosis (BA) is caused by *Bartonella* species (most commonly *Bartonella henselae* and *Bartonella quintana*). The cutaneous lesions appear as angiomatous lesions that can present in papular, nodular, or verrucous forms. Clinically, they may resemble lesions of Kaposi sarcoma (KS) and may also be confused with pyogenic granuloma or cutaneous lymphoma. Systemic involvement is common; visceral disease may present as osseous lesions, hepatic and splenic tumors, lymph node disease, pulmonary lesions, brain lesions, bone marrow infiltration, and widespread fatal systemic involvement. BA can present with unexplained fever, bacteremia, or endocarditis and should be considered in any HIV-infected patient with fever of unknown origin. Suspicious cutaneous lesions should always be biopsied and examined with hematoxylin, eosin, and Warthin-Starry silver staining. Blood and tissue cultures should be obtained, and indirect fluorescent antibody testing is useful when available. Polymerase chain reaction (PCR)-based tests play an important role and can be performed on biopsy and serum samples. Immunohistochemical staining for anti-HHV8 can be used to differentiate KS from BA. Bacillary angiomatosis responds to antibiotics such as erythromycin or doxycycline. Although cutaneous lesions resolve in 3 to 4 weeks, the treatment should be continued for at least 3 months. Severely ill patients should be treated with intravenous doxycycline combined with either gentamicin or rifampin for at least 3 months.

SYPHILIS

Syphilis (Chapter 319) is frequently seen in patients with HIV infection. Given the increasing number of reported syphilis cases, regular screening for syphilis, even in asymptomatic HIV-infected patients, is recommended. Its cutaneous manifestations are protean, and therefore syphilis should be

TABLE 392-2 SKIN CONDITIONS THAT CAN OCCUR AT ANY $CD4^+$ COUNT

Eczema
Xerosis
Tinea/onychomycosis
Kaposi sarcoma
Warts
Syphilis

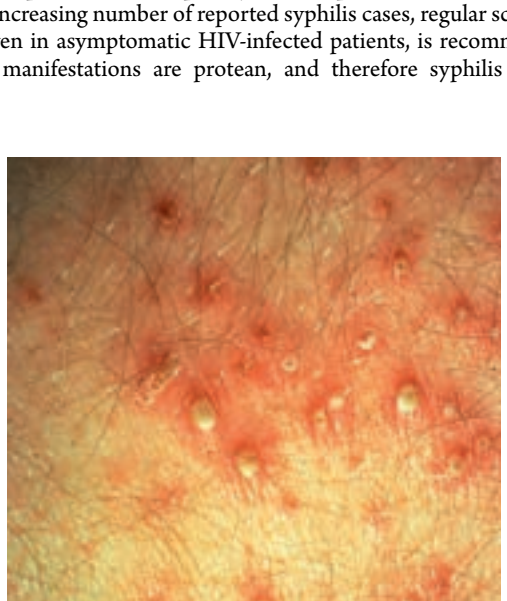

FIGURE 392-2. Staphylococcal folliculitis.

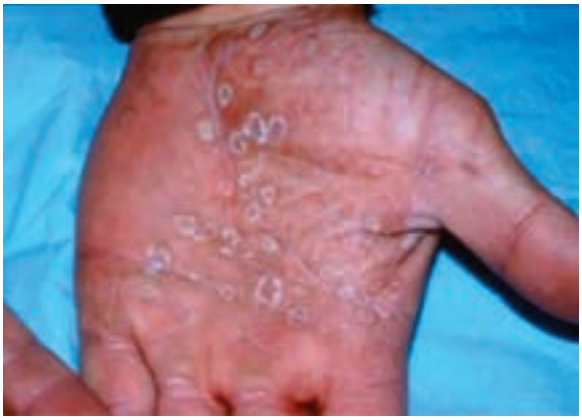

FIGURE 392-3. Secondary syphilis.

considered in any patient infected with HIV who has a new cutaneous eruption. An example of a secondary syphilis exanthem is shown in Figure 392-3. Primary, secondary, and tertiary forms of syphilis may be manifested clinically as they are in HIV-negative individuals, although atypical findings are not uncommon. Central nervous system (CNS) involvement may occur early in HIV patients, and relapse in the CNS may be more common after standard treatment. In most HIV patients, serologic tests are accurate and reliable for the diagnosis of syphilis and to monitor response to treatment. Atypical results can occur, and the use of other tests (biopsy and darkfield microscopy) should be considered when results do not match clinical findings.[8] First-line treatment is with benzathine penicillin G, 2.4 million units in a single dose. Azithromycin resistance is becoming more common. Close follow-up is essential, and titers of syphilis serology should be documented to ensure adequate treatment and CNS clearance. HIV-infected patients who meet criteria for treatment failure should be managed in the same manner as HIV-uninfected patients, because serologic clearance of syphilis appears largely unaffected by the use of ART. CSF examination and retreatment should be strongly considered for patients whose non-treponemal test titers do not decrease appropriately within 6 to 12 months of treatment.

Fungi

CANDIDA

Oral candidiasis is a common and localized infection caused by the yeast *Candida albicans*. It is commonly seen in patients with immunodeficiency, including HIV infection. It is characterized by white cheesy plaques and papules, loosely adherent to the tongue and oropharynx. Topical therapies include clotrimazole troches and nystatin oral suspension. Fluconazole is recommended for moderate to severe disease, but resistance has been noted in the setting of advanced HIV with extensive prior triazole use. Second-line treatment is with oral itraconazole, although it is unclear whether pulsed therapy is preferable.

DERMATOPHYTES

Cutaneous dermatophyte infections are common in HIV-infected patients, but it is not clear if these are any more frequent than in the non–HIV-infected population. Tinea infections follow a normal pattern and can involve the hands, feet, lower trunk, groin, and buttocks. The dermatophyte can spread to hair-bearing areas and present with plaques of folliculitis known as Majocchi granuloma. In those with advanced HIV, tinea infections can be atypical, diffuse, and severe. Tinea of the palms, soles, and other areas can be treated with topical imidazoles or terbinafine. If extensive areas are involved, oral antifungals are usually a more effective alternative and should be continued for approximately 1 month in cases of Majocchi granuloma. Treatment of onychomycosis with oral antifungals is standard but should be used selectively after risks and benefits of treatment are considered.

OTHER FUNGI

Cryptococcus neoformans is an encapsulated yeast that is a well-recognized opportunistic infection in advanced HIV ($CD4^+$ count < 200 cells/μL). Clinical manifestations most commonly involve the lungs and CNS, but cutaneous dissemination can be seen in approximately 10% of patients.[9] Skin lesions can occur anywhere on the body and present as pearly 2- to 5-mm

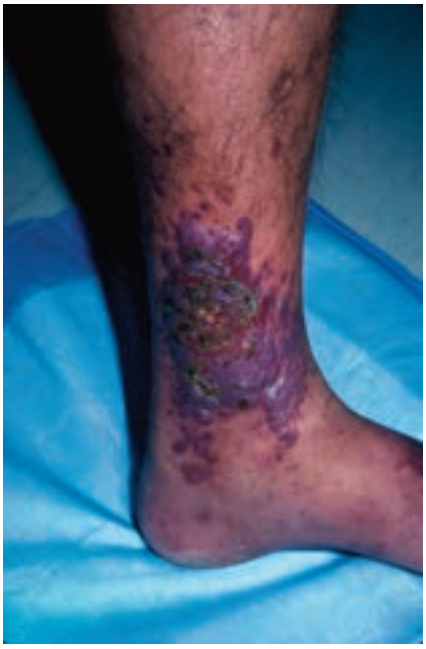

FIGURE 392-4. Kaposi sarcoma.

molluscum-appearing papules. Unlike molluscum, however, they develop over a short period of time. Large gelatinous plaques with umbilicated areas may also occur. Diagnosis is established by skin biopsy and culture. *Cryptococcus* on the skin in HIV is always associated with systemic cryptococcal infection and a serum cryptococcal antigen will be positive. *Histoplasma capsulatum* and *Penicillium marneffei* may morphologically resemble molluscum. Patients who have had these diseases should be evaluated for long-term suppressive doses of antifungal medications.

Infestations

SCABIES

Scabies is caused by infestation with the mite *Sarcoptes scabiei* and is commonly seen in HIV infection. It can be manifested as the classic rash of scabies, with burrows affecting the finger and toe webs and widespread excoriated papules with a predilection for the axillae, nipples, and genitalia. With advancing immunosuppression, the infestation can become more widespread and refractory to treatment. Crusted scabies can develop with advanced HIV and presents with thick crusts that are teeming with mites. Combination treatment with a topical agent such as permethrin or benzoyl benzoate and systemic treatment with ivermectin tablets is recommended for crusted scabies. Topical agents may be sufficient for classic scabies. Lindane lotion is often used to treat scabies, but there have been reports of lindane-resistant scabies and neurologic side effects attributed to this treatment.

TUMORS

Kaposi Sarcoma

KS is the most common HIV-associated malignancy, but since the introduction of ART, there has been a significant decrease in its incidence. It is characterized by dusky purple nodules and plaques with a predilection for the extremities and oral mucosa (Fig. 392-4). Poor prognostic factors include lymphedema and internal organ involvement. It is caused by HHV8, also known as Kaposi sarcoma–associated virus. The virus may be transmitted vertically, sexually, or casually, and its presence is necessary for the development of KS. Immune reconstitution with ART is the first-line treatment for early disease. Chemotherapy can be added for aggressive or unresponsive KS.

Malignant Melanoma

Malignant melanoma occurs with increased incidence in HIV-positive patients. Melanoma behavior can be more aggressive because HIV patients are more likely to present with metastases and have worse outcomes. High-risk patients should be screened annually, and those with a history of malignant melanoma should be followed closely for recurrence or early evidence of metastasis.[10]

Nonmelanoma Skin Cancer

Persons with AIDS have a three- to five-fold increased risk of developing a nonmelanoma skin cancer. The clinical manifestations of squamous cell carcinoma (SCC) and basal cell carcinoma (BCC) are identical to those seen in the uninfected population. Cutaneous SCC may be dangerous in the context of HIV infection, because lesions can present at a younger age and are associated with a high risk for local recurrence, metastasis, and increased mortality. Management of SCC is surgical excision, whereas curettage and electrodessication may be appropriate for BCC lesions on the extremities or trunk. Surgical excision should be the treatment of choice for BCC on the face.

INFLAMMATORY MANIFESTATIONS

Immune Reconstitution Inflammatory Syndrome

Most patients benefit substantially after the commencement of ART with a large reduction in morbidity and mortality. However, a subset of patients experience unmasking of new skin disease or paradoxical worsening of existing dermatologic conditions, attributable to IRIS (Chapter 395). Risk factors for developing IRIS include starting ART with a low CD4 nadir (<200 cells/μL) and the presence of subclinical opportunistic or other infections at the time of ART initiation. The most common skin manifestations of early IRIS include varicella-zoster reactivation and eosinophilic folliculitis. KS can present with organ involvement during IRIS. Other cutaneous manifestations include HPV (presenting as genital, flat, or common warts), reactivation of HSV or cytomegalovirus, cutaneous mycobacterial infection, or molluscum contagiosum. Leprosy, fungal infections, and parasitic infections such as leishmaniasis have also been reported. IRIS is most effectively treated by identifying and treating any underlying infection.

Seborrheic Dermatitis

Seborrheic dermatitis is a common dermatosis in the general population, but it has a strikingly increased prevalence in HIV-infected patients. Although more common in those with advanced HIV, the condition can be seen in all stages of immunosuppression. It appears as orange-red scaly patches affecting the scalp, eyebrows, nose, and cheeks. Seborrheic dermatitis can also affect the central portion of the chest and genitalia. ART appears to make seborrheic dermatitis more responsive to the standard therapy consisting of combinations of topical antifungals (econazole or ketoconazole) and low-potency corticosteroids.

Atopic Dermatitis

Atopic dermatitis (AD) is common in HIV-infected patients. The condition is characterized by pruritic scaly plaques that are classically localized to flexural areas. The distribution, however, can vary with ethnicity; AD in black Africans often involves the extensor surfaces. Treatment for AD is the same in the non–HIV-infected population. Frequent application with emollients should be emphasized. Topical steroids and sedating antihistamines are useful for flares. Topical pimecrolimus and tacrolimus are licensed for the treatment of atopic dermatitis, but their safety has not been demonstrated in HIV infection. As with other dermatoses, AD improves with ART, but in patients who had a $CD4^+$ nadir less than 200 cells/μL it tends to be more persistent even with ART.

Psoriasis

Psoriasis may appear early in HIV infection. The severity ranges from mild to severe, with more severe disease correlating with worsening immunosuppression. This observation is paradoxical because psoriasis is thought to be a T-cell-mediated disease but may be explained by immune dysregulation. All psoriasis subtypes occur in the setting of HIV, but inverse (flexural), guttate, and erythrodermic forms are seen most often. Psoriatic arthritis is more common and severe than in the non-HIV population. Treatment should be tailored to disease severity. First-line therapy for mild to moderate psoriasis includes topical steroids, calcipotriol, and retinoids (tazarotene). In those with moderate to severe disease, ultraviolet therapy (UVB, PUVA) and ART are the first-line treatments. Oral retinoids are an appropriate second-line agent. Immunosuppressive agents such as methotrexate and cyclosporine can be used for severe or refractory disease. Patients on these agents should be followed closely, and concomitant prophylaxis for opportunistic infections should be considered. Given the limited evidence and risks associated with their use (particularly in the HIV-infected population), tumor necrosis factor (TNF)-α inhibitors should be reserved for patients with very refractory psoriasis and those with debilitating arthritis. ART invariably reduces the severity of psoriasis and is considered a first-line treatment in those with moderate to severe disease.

Papular Pruritic Eruption of HIV

Papular pruritic eruption (PPE) is commonly seen in Africa and Asia in association with HIV. The condition presents with excoriated papules that initially appear on the extensor extremities but subsequently extend to involve the trunk and face. Pruritus can be severe. Because its presentation is nondescript, other itchy skin diseases such as eczema or eosinophilic folliculitis should be considered in the differential diagnosis. Skin biopsies have been considered to be important to distinguish between PPE and other itchy papular eruptions.[11] When present, PPE is highly predictive of HIV infection and is often the presenting sign. It has been hypothesized that this condition represents a hyperactive immune response to arthropod bites. Treatment is difficult, but potent topical corticosteroids and UVB may be of some benefit. Because PPE presents at CD4 counts of 350 cells/μL or less, it may be an indication for ART initiation. The condition has been reported to improve dramatically with effective ART.

Eosinophilic Folliculitis

Eosinophilic folliculitis is characterized by urticarial follicular-based papules and pustules located primarily on the scalp, face, neck, and upper chest (Fig. 392-5). These lesions are intensely itchy, and skin biopsy is helpful in ruling out other causes of folliculitis. Eosinophilic folliculitis tends to be seen when $CD4^+$ counts are less than 200 cells/μL. It can also be seen as part of the immune reconstitution syndrome within the first 16 weeks of starting ART. Treatments include topical corticosteroids, antihistamines, itraconazole, metronidazole, oral retinoids, and UV light therapy. If immune reconstitution is implicated, these agents can be given for an 8- to 12-week period while immune restoration stabilizes.

Prurigo Nodularis

Prurigo nodularis is characterized by pruritic dome-shaped nodules that initially develop on photoexposed areas of the extremities but eventually extend to involve the trunk. Development of prurigo nodularis is associated with background skin pigment and CD4 counts below 100 cells/mm^2. Standard treatment options include emollients, potent topical steroids, and sedating antihistamines such as chlorpheniramine or hydroxyzine. Thalidomide is often effective in recalcitrant cases but requires careful monitoring for peripheral neuropathy, and women of childbearing potential require effective contraception. Immune reconstitution and reduction of viremia with ART are helpful. Regimens that include raltegravir may be useful in refractory cases.[12]

Photodermatitis

HIV-infected patients have a tendency toward photosensitivity, a phenomenon that is poorly understood. Patients are commonly on photosensitizing medications like trimethoprim-sulfa and dapsone for prophylaxis and treatment of concomitant infections. But even in the absence of classic photosensitizing drugs, HIV itself and pigmented skin are risk factors for photodermatitis. Photodermatitis presents with pruritic plaques or indurated plaques in photoexposed areas of the skin like the cheeks, ears, lower lip, tip of the nose, and dorsal hands. Sunscreen and sun avoidance are the treatments.

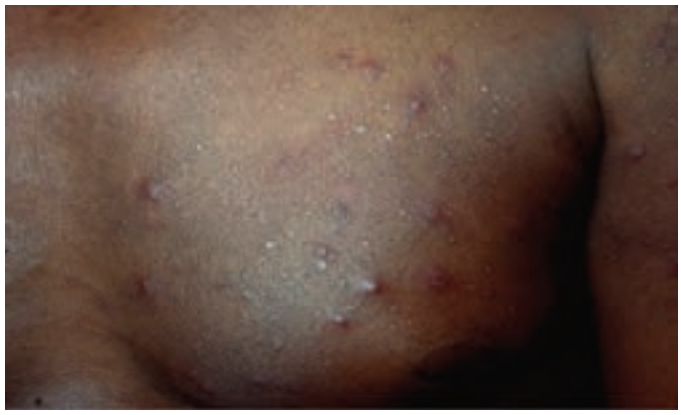

FIGURE 392-5. Eosinophilic folliculitis.

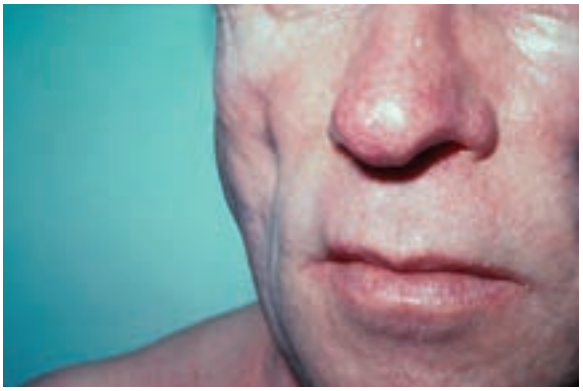

FIGURE 392-6. HIV-associated facial lipoatrophy.

METABOLIC MANIFESTATIONS

HIV Lipodystrophy Syndrome

Lipodystrophy syndrome was first described in 1998 and consists of peripheral lipoatrophy, central lipohypertrophy, lipid abnormalities, and insulin resistance. It is likely that it is caused by a number of factors, including ART, although it is unlikely that a single agent is responsible. Cutaneous facial lipoatrophy tends to be the most problematic dermatologic aspect, with significant associated psychological morbidity (Fig. 392-6). Currently, no treatments are effective for lipodystrophy syndrome, although surgical correction with dermal fillers may improve the cosmetic appearance.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Rane SR, Agarwal PB, Kadgi NV, et al. Histopathological study of cutaneous manifestations in HIV and AIDS patients. *Int J Dermatol*. 2014;53:746-751.
2. Wood E, Kerr T, Rowell G, et al. Does this adult patient have early HIV infection?: The Rational Clinical Examination systematic review. *JAMA*. 2014;312:278-285.
3. Cohen MS, Shaw GM, McMichael AJ, et al. Acute HIV-1 infection. *N Engl J Med*. 2011;364: 1943-1954.
4. Blank LJ, Polydefkis MJ, Moore RD, et al. Herpes zoster among persons living with HIV in the current antiretroviral therapy era. *J Acquir Immune Defic Syndr*. 2012;61:203-207.
5. Weinberg A, Levin MJ, Macgregor RR. Safety and immunogenicity of a live attenuated varicella vaccine in VZV-seropositive HIV-infected adults. *Hum Vaccin*. 2010;6:318-321.
6. Castelo-Soccio L, Bernardin R, Stern J, et al. Successful treatment of acyclovir-resistant herpes simplex virus with intralesional cidofovir. *Arch Dermatol*. 2010;146:124-126.
7. Shadyab AH, Crum-Cianflone NF. Methicillin-resistant *Staphylococcus aureus* (MRSA) infections among HIV-infected persons in the era of highly active antiretroviral therapy: a review of the literature. *HIV Med*. 2012;13:319-332.
8. Workowski KA, Berman S. Sexually transmitted diseases treatment guidelines, 2010. *MMWR Recomm Rep*. 2010;59:1-110.
9. Warkentien T, Crum-Cianflone NF. An update on *Cryptococcus* among HIV-infected patients. *Int J STD AIDS*. 2010;21:679-684.
10. Olson CM, Knight LL, Green AC. Risk of melanoma in people with HIV/AIDS in the pre- and post-HAART eras: a systematic review and meta-analysis of cohort studies. *PLoS ONE*. 2014;9:e95096.
11. Chua SL, Amerson EH, Leslie KS, et al. Factors associated with pruritic papular eruption of human immunodeficiency virus infection in the antiretroviral therapy era. *Br J Dermatol*. 2014;170: 832-839.
12. Unemori P, Leslie KS, Maurer T. Persistent prurigo nodularis responsive to initiation of combination therapy with raltegravir. *Arch Dermatol*. 2010;146:682-683.

REVIEW QUESTIONS

1. All of the following skin conditions are increased in frequency and/or severity in patients with HIV/AIDS. Which one of them is clearly exacerbated (or unmasked) by initiation of antiretroviral therapy (ART) as an immune reconstitution manifestation?

A. Seborrheic dermatitis
B. Papular pruritic eruption (PPE) of HIV
C. Malignant melanoma
D. Atopic dermatitis
E. Herpes zoster

Answer: E Herpes zoster has been one of the more commonly encountered diseases in the immune reconstitution inflammatory syndrome (IRIS) and typically appears during the second stage (≈3 months after initiation of ART). Other skin disorders in HIV-positive patients that manifest with worsening that is attributed to immune reconstitution include: human papillomavirus, reactivation of herpes simplex and cytomegalovirus, cutaneous mycobacterial infection, and molluscum contagiosum. The other choices listed are not known to be significantly exacerbated by initiation of ART; in fact, some of them show marked improvement with it.

2. Which of the following pruritic eruptions is ***not*** a skin disorder that is known to be increased in frequency and/or severity in AIDS patients?

A. Secondary syphilis
B. Prurigo nodularis
C. Pityriasis rosea
D. Seborrheic keratosis
E. PPE

Answer: C Pityriasis rosea is an acute self-healing exanthem of unknown etiology for which many infectious agents have been incriminated. It is not known to be definitively associated with HIV/AIDS. Syphilis is frequently seen in patients with HIV infection: primary, secondary, and tertiary forms of syphilis may be manifested clinically as they are in HIV-negative individuals, but atypical findings are not uncommon. Prurigo nodularis is characterized by pruritic dome-shaped nodules that particularly develop in persons with background skin pigment and CD4 counts below 100 cells/mm^2. Seborrheic dermatitis and keratosis have a strikingly increased prevalence in HIV-infected patients. PPE is commonly seen in Africa and Asia in association with HIV. Skin biopsies are indicated to distinguish between PPE and other itchy papular eruptions in patients with HIV/AIDS. When present, PPE is highly predictive of HIV infection and is often the presenting sign.

3. Oral hairy leukoplakia is associated with which of the following infectious organisms?

A. Epstein-Barr virus (EBV)
B. *Candida*
C. Herpes simplex
D. Cytomegalovirus (CMV)
E. *Cryptococcus*

Answer: A Oral hairy leukoplakia is characterized by nonpainful white plaques with a feathered edge, especially on the lateral borders of the tongue. It is associated with EBV and is very rare in immunocompetent hosts. There is no specific treatment for it, but it tends to resolve when patients are taking ART. Superinfection with *Candida* species should be considered if the lesions are painful.

393

HEMATOLOGY AND ONCOLOGY IN PATIENTS WITH HUMAN IMMUNODEFICIENCY VIRUS INFECTION

THOMAS S. ULDRICK AND ROBERT YARCHOAN

Patients with human immunodeficiency virus (HIV) infection or acquired immunodeficiency syndrome (AIDS) have a substantially increased risk of developing a number of cancers. Three of these cancers, Kaposi sarcoma (KS), certain non-Hodgkin lymphomas (NHLs), and cervical cancer, confer a diagnosis of AIDS when they arise in an HIV-infected patient and are referred to as AIDS-defining malignancies (ADMs). HIV increases the risk for a number of other tumors, including classic Hodgkin lymphoma (cHL), lung cancer, anal cancer, oropharyngeal cancer, hepatocellular carcinoma, and nonmelanoma skin cancer. These other HIV-associated tumors are referred to as non–AIDS-defining malignancies (NADMs) (Table 393-1).[1] Cytopenias and coagulation abnormalities are also common in patients with AIDS.

Effective combination antiretroviral therapy (cART) for HIV became broadly available in 1996. Its widespread use dramatically reduced opportunistic infections (OIs) and increased the longevity of patients with HIV infection. In part for these reasons, the number of persons living with HIV in the United States has approximately doubled since 1996. The use of cART also decreased the incidence of KS and NHL in the HIV-infected population. Nonetheless, these tumors remain important causes of morbidity and mortality. However, as the HIV-infected population ages, NADM, as well as other cancers not associated with HIV, are becoming increasingly common, and cancer is now a leading cause of death in HIV-infected persons. Cytopenias are also less common with cART, and management of neutropenia and anemia has shifted from growth factor support to effective treatment of HIV. Still, HIV increases the risk of hematologic abnormalities, which continue to cause diagnostic and therapeutic challenges. With more than 1 million people in the United States and 30 million people globally infected with HIV, management of HIV-infected patients with cancers and blood disorders is increasingly important.

CANCERS IN HIV-INFECTED PATIENTS

Care of the HIV-infected patient with cancer requires integration of oncologic and infectious disease expertise, including evaluation of comorbidities, planning of timing and selection of cART, and OI prophylaxis. Although patients with advanced AIDS may tolerate chemotherapeutic regimens poorly, patients whose HIV is controlled with cART can often receive full-dose regimens and do as well as their HIV-negative counterparts. There are 30-plus approved antiretroviral agents. Consideration of potential pharmacokinetic interactions between antiviral drugs and cancer therapy is required.

EPIDEMIOLOGY

The standardized incidence ratios of KS and certain aggressive B-cell NHLs are markedly increased in patients with HIV, especially those with low CD4 counts. KS, certain NHLs, cervical cancer, and a number of other NADMs are directly or indirectly caused by oncogenic viruses, especially Kaposi sarcoma herpesvirus (KSHV, also called human herpesvirus 8 [HHV-8]), Epstein-Barr virus (EBV), human papillomavirus (HPV), hepatitis C virus (HCV), and hepatitis B virus (HBV) (see Table 393-1). Many of these viruses are prevalent in various HIV-infected populations and are poorly controlled in the setting of immunosuppression. HIV-infected individuals also often have increased exposure to other cancer risk factors, such as cigarette smoke.

Substantial changes in the patterns of cancer in the U.S. AIDS epidemic have been noted since the introduction of cART. The incidence of KS and NHL have decreased relative to their peak, while the burden of NADMs, most of which develop over a longer period of time, is rising. Whereas most cancers developing in HIV patients were ADMs early in the epidemic, these cancers are now divided approximately evenly between ADM and other cancers (Fig. 393-1).[2] Overall, cancer is one of the leading causes of death in HIV-infected patients in countries where cART is widely available, in part because fewer patients are dying from other causes, such as OIs associated with low CD4 counts or AIDS itself. Also, HIV-associated cancers, especially ADMs, are a major public health concern in many resource-limited regions, such as sub-Saharan Africa.

PREVENTION AND SCREENING

Several chemopreventive strategies and behavioral interventions should be used to prevent HIV-associated malignancies. The risk of KS and NHL in patients with HIV is decreased with cART. Some studies show that improved immunity with cART also can reduce the prevalence of premalignant HPV-associated cervical[3] and anal squamous intraepithelial neoplasia and lesions. Vaccination against HBV and antiviral therapy against HBV and HCV decrease the risk of hepatocellular carcinoma in HIV-uninfected populations and are warranted. Smoking is prevalent in many HIV-infected populations, and smoking cessation interventions are advised. Nonmelanoma skin cancer is increased with HIV, and reduction of exposure to ultraviolet radiation is prudent.

HPV vaccines can reduce the risk of HPV-associated cancers and premalignant conditions. They should optimally be administered in early adolescence before the onset of sexual activity and exposure to HPV infection. A concern in this area is that the rate of uptake of the HPV vaccine is now low in the United States, especially in boys. Recent studies suggest that there may also be value in vaccination after the onset of sexual activity, in part to prevent reinfection after clearance of high-risk strains. The HPV vaccine has been shown to be safe in HIV-infected children and adults, and those with relatively preserved immune function can develop antibodies to the vaccine; however, it has not been shown whether the vaccine is protective in this population. Women with HIV infection, squamous intraepithelial lesions (SIL), and/or poor control of oncogenic HPV subtypes need more frequent gynecologic evaluation than low-risk women (see section on cervical cancer). Given the markedly increased risk of anal cancer in women and men with HIV (see Table 393-1), programs have been developed employing cytologic examination of anal mucosa to screen for and treat high-grade squamous intraepithelial lesions (HSIL).[4] A National Cancer Institute (NCI)-funded prospective study is planned to evaluate whether this approach is effective in preventing anal cancer.

TABLE 393-1 AIDS-DEFINING MALIGNANCIES AND COMMON NON–AIDS-DEFINING MALIGNANCIES IN PERSONS WITH HIV IN THE COMBINED ANTIRETROVIRAL THERAPY ERA

MALIGNANCIES	STANDARD INCIDENCE RATIO (HIV ONLY / AIDS)	ESTIMATED % OF ALL CANCERS 2004-2007 IN HIV/AIDS IN UNITED STATES	VIRAL ASSOCIATIONS	OTHER IMPORTANT RISK FACTORS
AIDS-DEFINING MALIGNANCIES				
Non-Hodgkin lymphoma				
Systemic	10-15 / 30-60	25.9%	EBV, KSHV*	
Primary CNS lymphoma	250 / 1020	3%	EBV	
Kaposi sarcoma	1300 / 3640	18.5%	KSHV	
Cervical cancer	2.9 / 5.3	2.4%	HPV	Smoking
NON–AIDS-DEFINING MALIGNANCIES				
Lung cancer	2.6 / 2.6	10%	—	Smoking
Anal cancer	9.2 / 20	5.7%	HPV	Smoking
Classic Hodgkin lymphoma	5.6 / 14	4.4%	EBV	
Oropharyngeal carcinoma	1.7 / 2.1	2.5%	HPV	Smoking, alcohol
Hepatocellular carcinoma	2.7 / 3.3	2.3%	HBV, HCV	Alcohol, aflatoxins, tobacco

*EBV associations vary with different AIDS-related lymphomas; approximately 30% Burkitt lymphomas, approximately 30-60% diffuse large B-cell lymphomas, and 100% of plasmablastic lymphomas. KSHV is associated with primary effusion lymphoma (≈80% coinfected with EBV) and large cell lymphoma arising in the setting of KSHV-associated multicentric Castleman disease (EBV negative).
CNS = central nervous system; EBV = Epstein-Barr virus; HBV = hepatitis B virus; HCV = hepatitis C virus; HPV = human papillomavirus; KSHV = Kaposi sarcoma herpesvirus.
From HIV/AIDS Cancer Match Study: standardized incidence rates in people with HIV but not AIDS at baseline versus AIDS at baseline, each compared to the general U.S. population. HIV estimates from Engels EA, et al. *Int J Cancer.* 2008;123:187-194. AIDS estimates from Engels EA, et al. *AIDS.* 2006;20:1645-1654. Percentage of total cancers of based on 34 states included in HACM, from Shiels MS, Pfeiffer RM, Gail MH, et al. Cancer burden in the HIV-infected population in the United States. *J Natl Cancer Inst.* 2011;103:753-762.

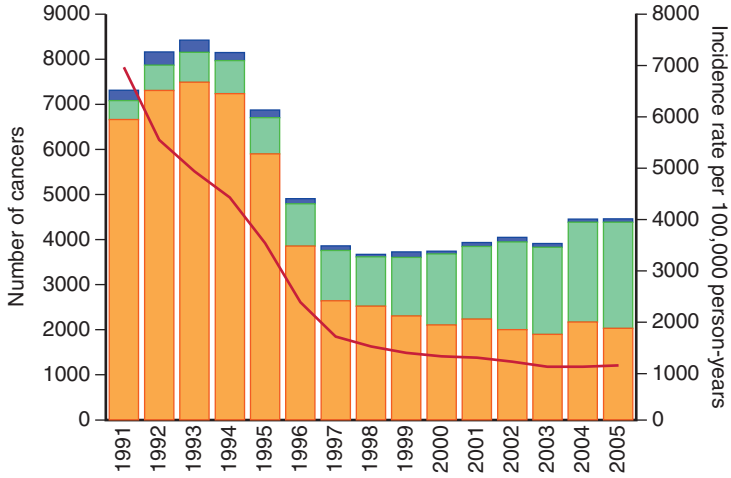

FIGURE 393-1. Trends in AIDS-defining malignancies (*orange bars*) and other malignancies (*green bars*) in persons with HIV/AIDS in the United States in the years 1991-2005. Blue = poorly classified cancers; line represents incidence rate. From Shiels MS, Pfeiffer RM, Gail MH, et al. Cancer burden in the HIV-infected population in the United States. *J Natl Cancer Inst.* 2011;103:753-762.

Although they have not been specifically evaluated in patients with HIV, routine recommended cancer screening, such as stool guaiac examinations or colonoscopy, mammography, and perhaps low-dose chest computed tomography (CT)[5] should be offered based on indications for the general population. In addition to these preventive measures, physicians caring for HIV-infected patients should be aware of cancers associated with HIV infection (see Table 393-1) and be alert for their development.

Non-Hodgkin Lymphoma and Classic Hodgkin Lymphoma in HIV

The Centers for Disease Control and Prevention (CDC) 1993 definition of AIDS considers Burkitt (or equivalent), immunoblastic (or equivalent), or primary brain lymphoma as AIDS defining. This classification is outdated, however. Using the current World Health Organization classification, patients with HIV are at substantially increased risk for eight distinct lymphomas or lymphoproliferative disorders. In this chapter, we use the term AIDS-related lymphoma (ARL) to collectively describe the NHLs with substantially increased standardized incidence ratios in patients with HIV/AIDS.[6]

EPIDEMIOLOGY

ARLs are aggressive mature B-cell lymphomas, including diffuse large B-cell lymphoma (DLBCL) with centroblastic (germinal center) or immunoblastic (activated B-cell) phenotypes, Burkitt lymphoma (BL), plasmablastic lymphoma (PBL), primary diffuse large B-cell lymphoma of the central nervous system (PCNSL), primary effusion lymphoma, and large cell lymphoma arising in the setting of KSHV-associated multicentric Castleman disease (KSHV-MCD). Classic Hodgkin lympoma (cHL) and KSHV-MCD also occur with markedly increased frequency in HIV-infected patients, although these are not NHLs and thus are not usually categorized as ARL. HIV testing should be considered in a patient with any of these tumors. ARL risk increases with immunosuppression and HIV viremia. With cART, ARL incidence decreased in the United States by 50%, mainly because of decreases in subtypes occurring at low $CD4^+$ counts. Still, ARL risk is greatly elevated. The estimated U.S. incidence is 97 per 100,000 person-years. ARL is a major public health problem in sub-Saharan Africa.

PATHOBIOLOGY

Many aggressive mature B-cell lymphomas (ARLs) are caused by the gammaherpesviruses EBV or KSHV, and for these, poor viral immune control related to global and in some cases virus-specific immune defects contributes to lymphomagenesis (see Table 393-1). EBV and KSHV encode for a number of genes, including mimics of human immune genes that are important for viral survival and transmission but are also implicated in lymphomagenesis. Examples include EBV-encoded CD40-like latent membrane protein (LMP-1) and KSHV-encoded viral FLICE-like inhibitory protein (vFLIP). Furthermore, chronic B-cell immune activation, especially in the setting of uncontrolled HIV, appears important, especially for NHL subtypes that occur at relatively preserved $CD4^+$ counts. Activation-induced cytidine deaminase (AICDA), an enzyme required for germinal center class switch recombination and somatic hypermutation, also induces mutations and pathogenic translocations. Importantly, *c-myc/Ig* translocations are seen in Burkitt lymphoma (BL), a majority of EBV-positive plasmablastic lymphoma (PBLs), and 24% of DLBCLs, likely accounting for the high proliferative rate and aggressive nature of these tumors.

CLINICAL MANIFESTATIONS

ARL and classic Hodgkin lympoma (cHL) often present with adenopathy or "B" symptoms (fever, night sweats, or weight loss in excess of 10% body weight). Involvement of bone marrow, liver, gastrointestinal (GI) tract, and CNS are common. HIV patients with cHL often have EBV-associated disease with either mixed cellularity or lymphocyte-depleted subtypes, present with high-risk disease, and also may present with extranodal involvement. PBL classically presents as an oral-cavity mass; however, other nodal or extranodal presentations are common. Primary effusion lymphoma often presents with pleural, peritoneal, or pericardial effusions but may have extracavitary presentations. Patients with primary effusion lymphoma commonly have concurrent KS or KSHV-MCD. Edema, hypoalbuminemia, and cytopenias are also common in primary effusion lymphoma and largely related to the underlying KSHV-associated malignancy rather than HIV.

DIAGNOSIS

Staging Evaluation of Systemic Lymphoma and Classic Hodgkin Lymphoma

Diagnosis of aggressive mature B-cell lymphoma (ARL) or cHL requires biopsy of a lymph node or involved extranodal lesion. Immunohistochemical markers are required for classification and include a combination of B-cell lineage markers and KSHV-encoded latency-associated nuclear antigen (LANA) and *in situ* hybridization for EBV-encoded small RNAs (EBER). Evaluation for *c-myc*/*Ig* translocations is warranted. Viral markers are particularly important for the CD20-negative large cell lymphomas plasmablastic lymphoma (PBL), and primary effusion lymphoma.

Staging includes CT of the neck, chest, abdomen, and pelvis; (^{18}F)-fluorodeoxyglucose positron emission tomography (^{18}FDG-PET) when available; and bone marrow biopsy. For ARL, CNS imaging by magnetic resonance imaging (MRI) with gadolinium (preferred) or CT with contrast and lumbar puncture with cerebrospinal fluid (CSF) cytology and flow cytometry are required. CNS imaging and CSF evaluation are not standard for cHL, although in HIV-associated cases, the CNS can be involved.

Baseline laboratory tests include HIV viral load, CD4 count, lactate dehydrogenase (LDH), complete blood cell count with differential, evaluation of renal and liver function, HBV surface and core antibody and surface antigen, and HCV serology or HCV RNA viral load. Patients with detectable HBV core antibody or surface antigen require a quantitative HBV viral load. Evaluation of cardiac function is recommended.

TREATMENT

Rx

General Approach to Supportive Care

Supportive care and combination antiretroviral therapy (cART) deserve special consideration. The survival of patients with ARL and cHL has improved in the cART era, and use of cART during ARL therapy may improve long-term outcomes. However, cART is not always essential during chemotherapy; 85 to 90% long-term survival rates have been obtained for the germinal center phenotype of DLBCL and BL with regimens that involved withholding cART.[7,8] Pharmacokinetic and toxicity interactions between cART and chemotherapeutic agents must be considered. Protease inhibitors (especially ritonavir) and cobicistat, a pharmacologic booster used in some cART formulations, inhibit CYP3A4, decrease metabolism of many chemotherapeutic agents, potentially increase toxicity, and should be avoided if possible. Zidovudine has overlapping hematotoxicity, and tenofovir requires renal monitoring. Often, cART is continued in patients on effective regimens when feasible, preferably with a protease inhibitor–sparing regimen; a new cART regimen need not necessarily be introduced before initiating lymphoma therapy. In patients with toxicities due to drug-drug interactions during curative-intent therapy, temporary suspension of nonessential drugs, including cART, is important.

OI prophylaxis is important, especially in patients with CD4 counts of less than 100 cells/mm^3. Regardless of baseline CD4 count, patients should receive *Pneumocystis jiroveci* pneumonia (PCP) prophylaxis, preferably trimethoprim-sulfamethoxazole (TMP-SMX). Patients with less than 100 CD4 cells/mm^3 also require prophylaxis against atypical mycobacterial infections with azithromycin 1200 mg weekly. Prophylaxis against herpes simplex virus and varicella zoster reactivation using valacyclovir should be strongly considered. Patients with active HBV require anti-HBV therapy. Prophylactic antifungals should be considered to prevent oral candidiasis, but azoles should be avoided during chemotherapy administration. Monitoring for OIs in patients with AIDS is important. Patients with HIV undergoing lymphoma therapy remain immunosuppressed beyond the end of therapy. Extension of these OI prophylaxis recommendations for at least 6 months beyond the end of therapy, and until CD4 cells adequately recover, is reasonable.

PERIPHERAL LYMPHOMA IN HIV

TREATMENT AND PROGNOSIS

Rx

Diffuse Large B-cell Lymphoma

In the cART era, lymphoma-specific characteristics and treatment-related factors are important in determining disease-free survival. The age-adjusted International Prognostic Index (based on stage, performance status, and LDH), CD4 cut-off of 100 cells/mm^3, and HCV status provide additional prognostic information for overall survival. Most patients, even with low CD4 counts and poor performance status, should be approached with curative intent. For CD20-positive tumors, the anti-CD20 monoclonal antibody rituximab substantially improves long-term cancer outcomes, but patients with CD4$^+$ counts less than 50 cells/mm^3 remain at high risk of infectious complications[9] and do no better overall. The best-studied regimens for DLBCL are DA-EPOCH-R (rituximab; continuous-infusion etoposide, doxorubicin, and vincristine; oral prednisone; bolus cyclophosphamide; and filgrastim), short-course DA-EPOCH with dose-dense rituximab (SC-EPOCH-RR), and R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone).[10] Pooled data suggest that EPOCH provides a better backbone regimen than CHOP for HIV-associated DLBCL as long as the CD4$^+$ count is above 50 cells/mm^3.

Burkitt Lymphoma

For BL, CHOP is inadequate. Preliminary results from a study of modified CODOX-M/IVAC combined with rituximab in HIV-associated BL demonstrated 1-year overall survival of 83% but 9% treatment-related mortality. A recent study reported 90% long-term overall survival for BL in a small number of HIV-infected patients treated with SC-EPOCH-RR with intrathecal methotrexate and deferred cART. EPOCH has a better toxicity profile than CODOX-M/IVAC. During the first cycle, BL patients require allopurinol and adequate hydration to avoid tumor lysis syndrome.

Plasmablastic Lymphoma

PBL is a recently recognized CD20-negative aggressive large B-cell lymphoma. Retrospective data suggest that outcomes are poor with CHOP or localized disease treated with radiation, with median survival less than 2 years. Recently, *c-myc/Ig* translocations have been noted in a majority of cases. Regimens effective in NHL with high proliferative rates, such as DA-EPOCH, are recommended.

KSHV-Associated Lymphomas, Including Primary Effusion Lymphoma

There is no established therapy for KSHV-associated NHL. Anthracycline-based regimens may be curative in some cases. Concurrent KSHV-MCD requires specific treatment.

CNS Prophylaxis and Treatment

CNS involvement in ARL is common and confers a poor prognosis. Routine CNS prophylaxis is required for BL and is commonly included for other ARLs. Treatment includes intrathecal methotrexate and/or cytosine arabinoside. Intensive intraventricular or intrathecal therapy is required for patients with leptomeningeal disease at the time of diagnosis.

Resource-Limited Settings

Physicians treating ARL in resource-limited settings face many challenges, often including a shortage of pathologists and laboratory support required to make an accurate diagnosis. Lower-cost regimens that are easy to administer and do not require extensive supportive care are often used. CHOP is a commonly used regimen for DLBCL and other ARLs in resource-limited settings but may be challenging to administer. HIV-associated BL in Africa is often treated with a cyclophosphamide-based combination regimen.

Classic Hodgkin Lymphoma

Despite some biological differences in cHL between HIV-infected and uninfected patients, outcomes appear comparable in the cART-era. ABVD (doxorubicin, bleomycin, vinblastine, and dacarbazine) or risk-adapted therapy are often used. In a retrospective study of ABVD, 5-year progression-free survival and overall survival were 59 and 81%, respectively. In a prospective study of risk-adapted therapy, 2-year progression-free survival and overall survival were 91 and 92%, respectively. However, treatment-related mortality was 6% with the more intensive regimen BEACOPP (bleomycin, etoposide, cyclophosphamide, vincristine, procarbazine, and prednisone).

PRIMARY CENTRAL NERVOUS SYSTEM LYMPHOMA (PCNSL)

DIAGNOSIS

AIDS-related PCNSL almost always arises in patients with CD4 counts of less than 50 cells/mm^3, many of whom are not aware they have AIDS or for other reasons have not been controlled with cART. Patients with AIDS-related PCNSL usually present with focal neurologic symptoms, seizures, or headaches related to the CNS mass. Outcomes remain poor in the cART era, with a 2-year overall survival of less than 25%. This is in part due to difficulty in diagnosis, neurologic comorbidities, and delayed or inadequate lymphoma treatment.[11] Patients with AIDS and CNS masses require urgent evaluation to minimize lag time to definitive therapy. Brain MRI with gadolinium demonstrates single or multiple contrast-enhancing masses in PCNSL that are not reliably distinguishable from toxoplasmosis. Prior to effective treatment for HIV, a course of empirical antibiotics for toxoplasmosis was sometimes an initial step in managing patients with AIDS with ring-enhancing CNS masses; however, empirical therapy for toxoplasmosis without concomitant CNS evaluation should not be considered an initial diagnostic maneuver for

patients in the cART era. Nuclear imaging with ^{201}Tl-SPECT or ^{18}FDG-PET generally differentiates infections from malignancies. Imaging of the neck, chest, abdomen, and pelvis is required to exclude systemic malignancy. Lumbar puncture should be undertaken when safe to evaluate CSF cell count, glucose, protein, cytopathology, flow cytometry, EBV, JC virus, toxoplasmosis, and cryptococcal antigen. Definitive diagnosis requires biopsy. However, in patients with less than 50 CD4 cells/mm^3, a combination of a ring-enhancing brain mass on MRI, positive ^{201}Tl-SPECT or ^{18}FDG-PET, and a high CSF EBV viral load is adequate to institute therapy for presumptive AIDS-related PCNSL if a biopsy is not feasible. Staging requires ophthalmologic evaluation.

TREATMENT Rx

Whole-brain radiation has often been used but can have severe neurologic toxicities. With cART availability, radiation-sparing therapy may be preferable, with promising preliminary results demonstrated in one clinical trial to date.

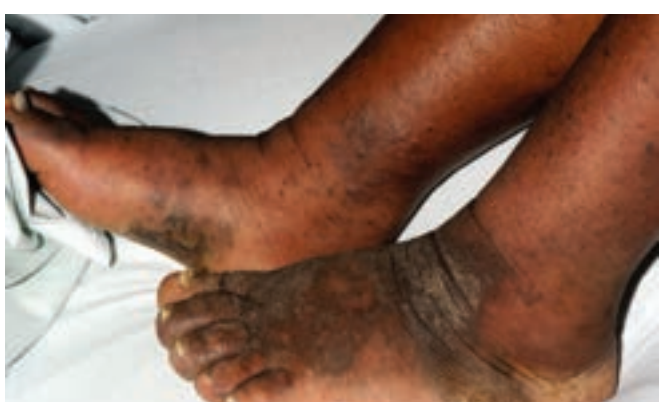

FIGURE 393-2. Kaposi sarcoma involving the feet.

Kaposi Sarcoma

EPIDEMIOLOGY

There are four epidemiologic categories of KS: (1) classic, commonly seen in elderly Mediterranean men (2) endemic, seen in African men, women, and children (3) iatrogenic, seen in transplant or other patients on chronic immunosuppressive medications, and (4) epidemic AIDS related. KSHV is the causative agent for all epidemiologic forms, but in the absence of immunosuppression, relatively few KSHV-infected patients develop KS. HIV coinfection or other immunosuppression dramatically enhances the risk of KS in KSHV-infected persons. Transmission of KSHV appears to be largely due to saliva exchange, and transmission may be increased in the setting of uncontrolled HIV or other infections such as malaria. KSHV seroprevalence is high in gay men in the United States but lower in injection drug users and less than 10% in the general population. In sub-Saharan Africa, seroprevalence ranges from 40 to 80%.

In the United States, KS incidence initially decreased by 84% with introduction of cART. Further declines have been more modest, and estimated incidence has recently stabilized at about 62 cases per 100,000 person-years. KS is now the second most common tumor in persons with HIV in the United States. In parts of sub-Saharan Africa, KS is the most common tumor in men overall, and second most common in women.

PATHOBIOLOGY

KS is an angioproliferative tumor with abnormal vascularity and infiltrating inflammatory cells. The pathognomonic KS spindle cells are generally poly- or oligoclonal and KSHV infected. KSHV encodes several micro-RNA and viral homologues of human genes that effect immune signaling and angiogenesis. Most KSHV-infected spindle cells express a limited set of latent viral proteins, such as LANA, which provide a proliferative advantage and suppress apoptosis. A minority (2 to 3%) of cells also express lytic KSHV-proteins, such as KSHV-encoded viral interleukin-6 (vIL-6) and viral G-protein–coupled receptor (vGPCR), which amplify production of angiogenic and immune modulatory factors that affect the tumor microenvironment and appear important in KS pathogenesis. Defective immune surveillance of KSHV-infected cells in HIV-infected patients is permissive for the development of KS. HIV may promote KS by other mechanisms as well. For example, the Tat protein of HIV can enhance infection of target cells by KSHV.

CLINICAL MANIFESTATIONS

KS generally presents as multiple painless, cutaneous, purplish or brown lesions that are initially flat and can become nodular and congruent (Fig. 440-19). KS often presents in the feet (Fig. 393-2) but can involve other areas of the skin as well. Advanced KS can have associated edema, ulceration, pain, and superinfection. KS frequently involves the oral palate. It can involve lymph nodes, the GI system, lungs, and may have associated pleural effusions. Pulmonary KS can be life threatening. GI disease is often asymptomatic, but occult blood loss is common. Involvement of the liver, bones, or soft tissues has been described. Visceral-only disease may be seen. KS may wax and wane with immune status. Progressive disease has been described after starting cART, and an immune reconstitution inflammatory syndrome (IRIS) (Chapter 395) has been proposed. KS may also progress in the setting of uncontrolled KSHV-associated multicentric Castleman disease (KSHV-MCD) or primary effusion lymphoma.

DIAGNOSIS AND STAGING

KS is diagnosed by biopsy. An estimate of the extent of cutaneous disease and evaluation for complications of advanced cutaneous disease is required. Evaluation for visceral disease includes fecal occult blood testing and chest imaging. Abnormalities require endoscopic evaluation. Pulmonary KS is diagnosed based on visualization of typical endobronchial lesions, with endobronchial biopsy generally deferred because of bleeding risk. KS in a lymph node biopsy often requires evaluation for concurrent lymphoma or KSHV-MCD.

KS is staged by the AIDS Clinical Trials Group (ACTG) TIS system, which is based on tumor burden ($T_{0\text{ or }1}$), immune status ($I_{0\text{ or }1}$), and presence of any systemic illness ($S_{0\text{ or }1}$). Poor risk (subscript 1) is defined by extensive oral disease, tumor-associated edema or ulceration, or non-nodal visceral disease (T_1); CD4 below 150 cells/mm^3 (I_1); and the presence of OIs, constitutional symptoms, or poor performance status (S_1).

TREATMENT Rx

cART is fundamental in treating HIV-associated KS. Immunosuppressive medications such as steroids, cyclosporine, and rituximab should be avoided when possible, and surgery has little role except for purpose of biopsy. Limited (T_0) HIV-associated KS is often managed with cART alone, which can induce KS regression over several months. KS-specific therapy is generally reserved for cases causing morbidity or other patient distress. Local therapies, such as topical alitretinoin gel, intralesional injection of vinblastine, laser therapy, or cryotherapy can be effective but are associated with local toxicities. They are occasionally used for very localized disease. Most patients who require KS-specific therapy are now given systemic therapy. Chemotherapy may be required urgently for symptomatic pulmonary KS. Progressive KS in patients who have recently initiated cART is often an indication for specific KS therapy, and chemotherapy plus cART is better than cART alone. Generally, KS is treated until remission or response plateau is attained. The number of cycles required is variable. The main objective of therapy is long-term remission or durable control. Because KSHV cannot be eradicated, KS is not considered curable, and relapses can occur.

Liposomal anthracyclines (doxorubicin and daunorubicin) and paclitaxel are approved by the U.S. Food and Drug Administration (FDA) for use in KS. Liposomal doxorubicin 20 mg/m^2 every 3 weeks is the most commonly used therapy. In the cART era, response rates of 45 to 80% may be anticipated. The FDA warns against cumulative lifetime doses exceeding 550 mg/m^2, although risk of cardiotoxicity is believed to be lower with liposomal formulations than with bolus non-liposomal anthracyclines. Paclitaxel should be considered in patients who have inadequate KS regression, cannot tolerate liposomal anthracyclines, or have reached their cumulative lifetime dose. A randomized trial comparing liposomal doxorubicin 20 mg/m^2 every 3 weeks with paclitaxel 100 mg/m^2 every 2 weeks in persons with advanced KS showed comparable efficacy, but increased toxicity in the paclitaxel arm.

Interferon-α can be effective in patients with preserved CD4$^+$ counts but is generally not used owing to poor tolerability of side effects. Several available targeted therapies have recently been evaluated in KS. In renal transplant–associated KS, modification of immunosuppression from cyclosporine to the mTOR inhibitor sirolimus leads to tumor regression. However, in a study of

sirolimus in AIDS-related KS, responses were limited, and drug interactions with ritonavir complicated dosing. Bevacizumab and imatinib have shown activity but alone appear inferior to cytotoxic agents. Thalidomide is active at high doses that are associated with toxicities. Studies of the thalidomide analogs (IMiDs) lenalidomide and pomalidomide are underway.

Low-cost agents that are easy to administer are required for KS treatment in sub-Saharan Africa. Vincristine, doxorubicin, bleomycin, and etoposide all are used in this setting. Oral etoposide was associated with an overall response rate of 36% in previously treated patients, the majority of whom were not on cART. In a randomized study of ABV (doxorubicin, bleomycin, vincristine) combined with cART versus cART alone in treatment-naïve patients, most with advanced KS, early addition of chemotherapy was feasible and effective in this setting, with an increase of 1-year overall response rate from 39 to 66%. To reduce risk of cardiac toxicity, some practitioners exclude doxorubicin and use BV in combination with cART.

PROGNOSIS

The prognosis for patients with AIDS-related KS has improved dramatically with cART. There are two main risk groups: good risk (T_0S_0, T_1S_0, or T_0S_1) versus poor risk (T_1S_1). Pulmonary involvement increases risk of death, and women with have a worse prognosis independent of other TIS risk factors. Poor CD4 immune reconstitution is often associated with need for recurrent therapy.

Multicentric Castleman Disease

MCD describes a group of lymphoproliferative disorders associated with dysregulated IL-6. The plasmablastic variant is often caused by KSHV, and nearly all MCD in patients with HIV is KSHV-MCD.

PATHOBIOLOGY

KSHV-MCD is a polyclonal lymphoproliferative disorder of KSHV-infected B cells with plasmablastic morphology. Symptoms are associated with KSHV-lytic activation and vIL-6 upregulation. Marked elevations in human IL-6 and IL-10 are noted, and along with KSHV are strongly implicated in disease pathogenesis.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

KSHV-MCD is characterized by intermittent severe inflammatory symptoms, including fevers, night sweats, fatigue and cachexia, and edema, as well as lymphadenopathy and splenomegaly. GI and respiratory symptoms are common; rheumatologic, neurologic, and dermatologic manifestations may also be present. Laboratory abnormalities include anemia, thrombocytopenia, hyponatremia, hypoalbuminemia, and elevated KSHV viral load. Most patients requiring therapy have an elevated C-reactive protein, and this can be a useful test to include in the initial evaluation of patients with suspected KSHV-MCD. Diagnosis requires pathologic confirmation, usually from a lymph node. Patients should be evaluated for concurrent KS and lymphoma. Although KSHV-MCD has been considered rare, its diagnosis is often missed and may thus be underreported. Indeed, there is evidence that the incidence of KSHV-MCD is increasing in the cART era. KSHV-MCD should be considered in HIV patients with unexplained fever, anemia, or other manifestations, especially if they have KS or are at risk for KSHV infection. Severe inflammatory symptoms similar to KSHV-MCD that are also related to excess cytokines, especially IL-6, have been described in KSHV-infected patients without KSHV-MCD.[12] Many of these patients also have KS or primary effusion lymphoma. The term KSHV inflammatory cytokine syndrome (KICS) has been proposed for such patients.

TREATMENT Rx

The best-studied treatment for KSHV-MCD is rituximab 375 mg/m^2 weekly for 4 weeks, which leads to resolution of symptoms in most patients. However, it may be insufficient in advanced disease and is associated with worsening KS. Approximately 30% of patients will relapse within 1 year. Rituximab combined with liposomal doxorubicin appears promising for patients with concurrent KS or severe symptoms.[13] Corticosteroids are sometimes administered transiently to help control acute life-threatening symptoms, but they do not appear to treat the underlying disease and can be associated with worsening KS and increased risk of infections, so should be avoided when possible. High-dose zidovudine with valganciclovir is active, although not as effective as rituximab for severe inflammatory symptoms and cytopenias. Chemotherapy regimens used in NHL and/or splenectomy have been used but are largely replaced by targeted approaches. HIV-infected patients require cART. Length of KSHV-MCD treatment, role of maintenance therapy, and evaluation and management of concurrent malignancies remain an active area of investigation. Patients are treated to resolution of symptoms and improvement in laboratory abnormalities.

PROGNOSIS

Without treatment, the prognosis of KSHV-MCD prognosis is poor, with patients succumbing to inflammatory manifestations, infections, or lymphoma. With effective therapy, 1-year survival is greater than 85%, and long-term remissions are possible.

Cervical Cancer

HIV-infected patients are at increased risk of chronic infection with high-risk genotypes of HPV, cervical HSIL, and development of cervical cancer compared to HIV-negative women. In the United States, cervical cancer screening guidelines for HIV-negative women are currently based on periodic cervical cytology, or Papanicolaou (Pap) testing, with the option of co-testing for high risk-HPV in women aged 30 and older. Because of the increased risk and persistence of HPV in HIV-infected women, the current U.S. Department of Health and Human Services (DHHS) guidelines for cervical cancer screening include only cytology and not high-risk HPV testing.[14] These guidelines recommend that HIV-infected women should be screened for cervical cancer with Pap testing at the time of HIV diagnosis, and Pap testing should be repeated in 6 months. Provided both Pap tests are normal, cervical cancer screening should be continued with annual Pap tests. In general, a Pap test with cytologic abnormalities should prompt referral for a colposcopic evaluation of the cervix as well as the vagina and vulva with directed biopsies. However, for a Pap test showing atypical squamous cells of uncertain significance (ASC-US), options include either immediate referral to colposcopy or repeat cytology in 6 to 12 months. If a cervical biopsy shows HSIL, then the patient should undergo treatment to prevent malignant transformation. Treatment options are similar to those for HIV-negative women, involving either ablative (with cryotherapy or laser) or excisional therapies (with conization by loop electrosurgical excisional procedure (LEEP) or cold-knife). Because HIV-infected women, especially those with low CD4 counts and those not on cART, have high rates of persistent and recurrent SIL post treatment, these women must be followed closely with cytology and colposcopy, as indicated, and treated for recurrences. Cervical cancer screening guidelines in the general population have been updated frequently in the past two decades, and physicians should be alert for changes in screening recommendations for HIV-infected women.

In resource-limited settings, cervical cancer is a leading cause of cancer-related mortality in women. "Screen-and-treat" approaches based on cervical visual inspection with acetic acid (VIA) or rapid HPV testing with immediate treatment are being evaluated. Treatment of cervical cancer in HIV-infected patients is similar to its treatment in the general population. Patients whose HIV is well controlled on cART generally tolerate therapy and have similar outcomes to HIV-uninfected patients.

Non–AIDS-Defining Malignancies

The term non–AIDS-defining malignancy (NADM) is used to categorize cancers with increased standardized incidence ratios in HIV-infected populations that are not considered AIDS-defining by the CDC. Like AIDS-defining malignancies, many are caused by oncogenic viruses. The commonest NADMs are lung cancer, anal cancer (Fig. 393-3A) (caused by HPV), classic Hodgkin lymphoma (Fig. 393-3B) (caused by EBV), oropharyngeal cancers (caused by HPV), and liver cancer (often caused by HBV or HCV). Less common NADMs include vulvar and penile cancer (caused by HPV), Merkel cell carcinoma (caused by Merkel cell polyomavirus), and conjunctival carcinoma, which is seen especially in sub-Saharan Africa. Increased prevalence of smoking contributes to lung cancer risk, as well as risk for some other NADMs. However, HIV remains a risk factor for lung cancer independent of smoking, perhaps as a result of chronic inflammation or recurrent lung infections. HIV viremia is an independent risk factor for at least some virus-associated NADMs, including anal cancer and cHL.[15] Increased longevity in HIV patients with chronic liver disease from HBV, HCV, or other causes appears to account for an increasing number of hepatocellular carcinomas in HIV patients. In addition to NADM, incidental cancers whose risk is not increased by HIV, such as breast or colon cancer, are also increasing in number as the number of HIV-infected persons increases and this population ages.

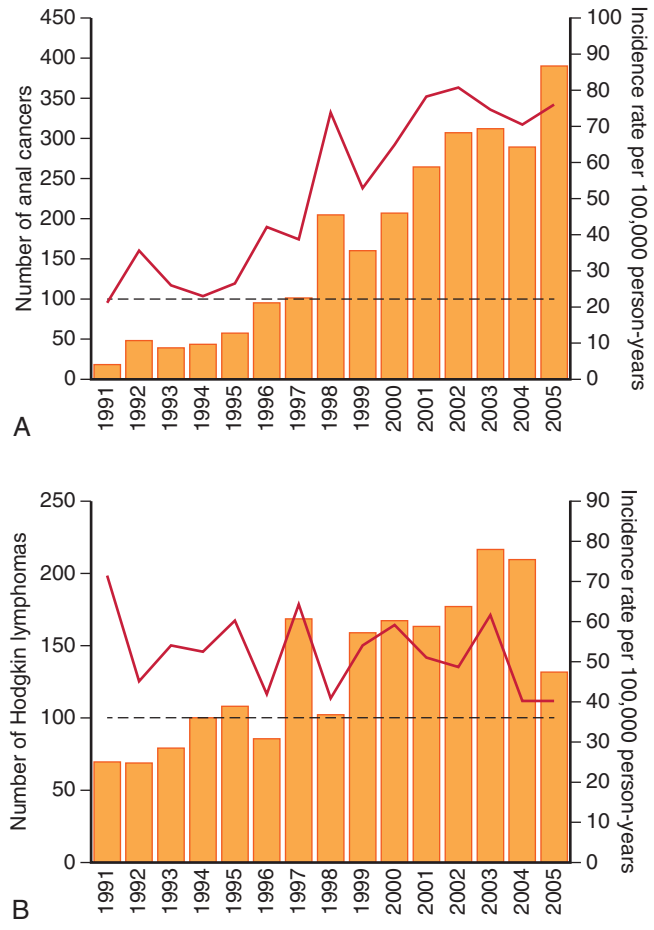

FIGURE 393-3. Trends in (A) anal cancer and (B) classic Hodgkin lymphoma in persons with HIV/AIDS in the United States in the years 1991-2005. Bars represent number of cases, line represents incidence rate. From Shiels MS, Pfeiffer RM, Gail MH, et al. Cancer burden in the HIV-infected population in the United States. *J Natl Cancer Inst*. 2011;103:753-762.

The approach to management of NADM in patients with HIV is generally comparable to HIV-negative patients and should be based on performance status and stage in patients without serious HIV-associated comorbidities. With cART, many HIV-infected patients tolerate standard cancer therapy, including chemoradiation with intensity-modulated radiation therapy for localized cervical cancer or anal cancer, liver transplant for hepatocellular carcinoma, and allogeneic stem cell transplantation for hematologic malignancies. Attention to potential pharmacokinetic interactions between cART and chemotherapy is required. $CD4^+$ counts and HIV viral load should be monitored and appropriate OI prophylaxis administered.

HEMATOLOGIC ABNORMALITIES IN HIV-INFECTED PATIENTS

Red Blood Cell Disorders and Anemia

Anemia is common in patients with HIV and has a broad differential diagnosis (Table 393-2). Many patients have anemia of chronic inflammation.

EPIDEMIOLOGY

HIV-associated anemia is associated with decreased $CD4^+$ count, complications of AIDS, and HIV itself. Although less common in patients treated with cART, mild anemia is still seen in over 25% of persons with HIV and over 70% with advanced AIDS. More severe anemia (hemoglobin < 10 mg/dL) is seen in 5 to 10% of patients on cART and is also related to low CD4 count. Anemia may be a sign of comorbidities such as poor nutrition or undiagnosed cancer, as well as a marker of poor HIV control. Anemia is an important independent predictor of mortality.[16]

PATHOBIOLOGY

Anemia may be related to multiple factors in patients with HIV (see Table 393-2). Anemia of chronic inflammation is common (Chapter 158). Hepcidin, a mediator of the anemia of inflammation, is inversely correlated with $CD4^+$ count. HIV may also decrease the number of erythroid progenitor cells and affect erythropoietin signaling through a direct effect of inflammatory cytokines such as IL-1 and IL-6.

TABLE 393-2 DIFFERENTIAL DIAGNOSIS OF ANEMIA IN HIV-INFECTED PATIENTS

CAUSED BY PREDOMINANTLY DECREASED RED BLOOD CELL PRODUCTION
Anemia of chronic inflammation
Malignancies
Non-Hodgkin lymphoma
Classic Hodgkin lymphoma
Kaposi sarcoma herpesvirus-associated multicentric Castleman disease
Infections
Parvovirus B-19
Mycobacterial infections
Fungal infections (*Cryptococcus neoformans*, histoplasmosis)
Bartonella infections
Nutritional deficiencies
B_{12} deficiency
Folate deficiency
Iron deficiency
Drug effects
Zidovudine
Trimethoprim-sulfamethoxazole
Cytotoxic chemotherapy
Hepatitis C therapy (interferon-α/ribavirin)
CAUSED BY PREDOMINANTLY INCREASED RED BLOOD CELL DESTRUCTION OR BLOOD LOSS
Hemophagocytic syndrome, generally associated with infectious agents
Epstein-Barr–virus associated diseases (e.g., classic Hodgkin lymphoma)
Kaposi sarcoma herpesvirus-associated diseases
Cytomegalovirus
Mycobacterium tuberculosis
Histoplasmosis
Autoimmune anemia
Kaposi sarcoma herpesvirus-associated multicentric Castleman disease
Lymphomas
Idiopathic
Hemolysis
Inherited hemoglobinopathy
Glucose-6-phosphate dehydrogenase deficiency and exposure to offending drugs
Malaria
Microangiopathic hemolytic anemia
Thrombotic thrombocytopenia purpura with low ADAMTS13 activity
Idiopathic
Gastrointestinal blood loss

DIAGNOSIS

Patients with anemia despite effective cART should be evaluated for underlying causes; the etiology may be multifactorial. Medications should be reviewed for bone marrow suppressive agents, such as chemotherapy or TMP-SMX. Zidovudine causes macrocytic anemia, but newer cART regimens rarely cause anemia. Complete blood cell counts, reticulocyte count, and mean corpuscular volume (MCV) will direct additional evaluation. Iron studies including serum ferritin help differentiate anemia of inflammation from iron deficiency. Deficiencies of vitamin B_{12} or folic acid should be considered. Patients with HIV often have a positive direct Coombs test, but this is not generally associated with hemolysis. KSHV-MCD should be considered, especially in patients with KS or at risk for KSHV infection. Patients with hemolysis require evaluation for an etiology such as medications (e.g., dapsone in glucose-6-phosphate dehydrogenase deficiency) or lymphoproliferative disorders. Occult GI bleeding (e.g., caused by KS or lymphoma of the GI tract) should be ruled out by testing for stool blood and GI endoscopy as indicated. Severe anemia of unclear etiology requires bone marrow biopsy with evaluation for malignancy, hemophagocytic syndrome, or infection.

TREATMENT

Rx

Treatment of anemic HIV patients requires cART, which alone can be corrective. Additional therapy is based on the management of other underlying causes. Recombinant erythropoietin is rarely indicated.

White Blood Cell Disorders

A hallmark of HIV infection is a decrease in the $CD4^+$ T-lymphocyte count, which occurs largely in the setting of uncontrolled HIV. $CD8^+$ cells are often relatively preserved. Neutropenia is common in AIDS. Especially in patients with CD4 counts less than 200 cells/mm^3 not on cART, absolute neutrophil counts less than 1000×10^6/L increases risk for bacterial infections. However, patients of African descent may have benign ethnic neutropenia associated with a polymorphism in Duffy antigen/receptor chemokine gene (*DARC*), which does not appear to increase infection risk. Neutropenic patients should be evaluated for HIV viremia, nutritional deficiencies, and hematotoxic medications. Neutropenia often improves with cART and immune reconstitution. HIV-associated and ethnic neutropenia responds to filgrastim if needed to support delivery of chemotherapy.

Thrombocytopenia

Thrombocytopenia is common with uncontrolled HIV and often related to platelet destruction due to platelet activation, immune thrombocytopenia, or (rarely) thrombotic thrombocytopenia purpura (TTP). Patients with immune thrombocytopenia should be evaluated for underlying causes. TTP with decreased ADAMTS13 activity and anti-ADAMTS13 antibodies, as well as idiopathic HIV-associated thrombotic microangiopathies have been described. The differential diagnosis of thrombocytopenia in HIV also includes medications, OIs, malignancy, KSHV-MCD, HCV infection, liver disease, disseminated intravascular coagulation, and heparin-induced thrombocytopenia.

HIV-associated thrombocytopenia usually improves with cART, whereas that from other causes may require specific therapies. Severe immune thrombocytopenia responds to steroids, rituximab, intravenous immunoglobulin, or anti-D globulin. Splenectomy should generally be avoided because of risk of infections in splenectomized HIV-infected patients. TTP with low ADAMTS13 activity usually responds to cART and plasma exchange.

Thrombosis

Estimated venous thromboembolism incidence in HIV is 2.6 to 5.7 per 1000 person-years. The incidence is increased in hospitalized patients. It is associated with lower $CD4^+$ counts and is commonly related to surgery, catheters, OIs, underlying malignancy (including KS or PCNSL), or medications. However, correcting for known venous thromboembolism risk factors, HIV itself further increases risk by about 25%. Management of thrombosis in patients with HIV should generally follow standard guidelines (Chapter 38). However, interactions between specific antiretroviral agents and warfarin have been reported with cART and may be related to CYP2C9 inhibition. Also, HIV may specifically increase the risk of heparin-induced thrombocytopenia.

FUTURE DIRECTIONS

With increased use of cART worldwide, continued aging of the HIV-infected population, and changes in the prevalence of other cancer risk factors, the epidemiology of cancer and hematologic disorders in the setting of HIV will continue to evolve. In addition to treating HIV, there may be opportunities to reduce the risk of some HIV-associated cancers by preventing or treating specific associated oncogenic viruses. HBV and HPV vaccines hold the potential to dramatically reduce the incidence of hepatocellular, anogenital, and oropharyngeal cancers. An ongoing trial should clarify the role of screening for and treatment of anal high-grade squamous intraepithelial lesions in preventing anal cancer. Prevention strategies are particularly important for sub-Saharan Africa.

In the last decade, AIDS-defining malignancies have continued to be an important cause of morbidity and mortality, while non-AIDS defining malignancies and incidental tumors are increasingly common. Although outcomes for patients with HIV and cancer can be similar to the general population, they are generally more complex to treat. Looking forward, improved understanding of the safety and tolerability of specific cancer drugs and immunotherapies in patients with HIV is required. Targeted therapies for less common subtypes of lymphoma as well as KS may improve outcomes and decrease toxicity. There is an NCI initiative to encourage enrollment of HIV-infected patients in NCI-funded trials whenever possible. Advances in treatment of AIDS-defining malignancies in resource-limited settings will require improved diagnostics, training capacity, and infrastructure. Low-cost regimens and preferably oral agents for common cancers are urgently needed. Evaluation of the effect of HIV on hematopoietic progenitor cells is an area of active research, especially in relation to allogeneic stem cell transplant and other cell therapies[17] that are being evaluated for a potential HIV cure.[18]

Grade A References

A1. Silverberg MJ, Neuhaus J, Bower M, et al. Risk of cancers during interrupted antiretroviral therapy in the SMART study. *AIDS*. 2007;21:1957-1963.

A2. Kaplan LD, Lee JY, Ambinder RF, et al. Rituximab does not improve clinical outcome in a randomized phase III trial of CHOP with or without rituximab in patients with HIV-associated non-Hodgkin's lymphoma: AIDS-malignancies consortium trial 010. *Blood*. 2005;106:1538-1543.

A3. Sparano JA, Lee JY, Kaplan LD, et al. Rituximab plus concurrent infusional EPOCH chemotherapy is highly effective in HIV-associated B-cell non-Hodgkin lymphoma. *Blood*. 2010;115:3008-3016.

A4. Gbabe OF, Okwundu CI, Dedicoat M, et al. Treatment of severe or progressive Kaposi's sarcoma in HIV-infected adults. *Cochrane Database Syst Rev*. 2014;8:CD003256.

A5. Cianfrocca M, Lee S, Von Roenn J, et al. Randomized trial of paclitaxel versus pegylated liposomal doxorubicin for advanced human immunodeficiency virus–associated Kaposi sarcoma: evidence of symptom palliation from chemotherapy. *Cancer*. 2010;116:3969-3977.

A6. Mosam A, Shaik F, Uldrick TS, et al. A randomized controlled trial of highly active antiretroviral therapy versus highly active antiretroviral therapy and chemotherapy in therapy-naive patients with HIV-associated Kaposi sarcoma in South Africa. *J Acquir Immune Defic Syndr*. 2012;60:150-157.

A7. Kuhn L, Wang C, Tsai WY, et al. Efficacy of human papillomavirus–based screen-and-treat for cervical cancer prevention among HIV-infected women. *AIDS*. 2010;24:2553-2561.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Cabone A, Vaccher E, Gloghini A, et al. Diagnosis and management of lymphomas and other cancers in HIV-infected patients. *Nat Rev Clin Oncol*. 2014;11:223-238.
2. Shiels MS, Pfeiffer RM, Gail MH, et al. Cancer burden in the HIV-infected population in the United States. *J Natl Cancer Inst*. 2011;103:753-762.
3. Minkoff H, Zhong Y, Burk RD, et al. Influence of adherent and effective antiretroviral therapy use on human papillomavirus infection and squamous intraepithelial lesions in human immunodeficiency virus-positive women. *J Infect Dis*. 2010;201:681-690.
4. Wells JS, Holstad MM, Thomas T, et al. An integrative review of guidelines for anal cancer screening in HIV-infected persons. *AIDS Patient Care STDS*. 2014;28:350-357.
5. Sigel K, Wisnivesky J, Shahrir S, et al. Findings in asymptomatic HIV-infected patients undergoing chest computed tomography testing: implications for lung cancer screening. *AIDS*. 2014;28:1007-1014.
6. Gobert A, Mounier N, Lavole A, et al. HIV-related malignancies: state of art. *Bull Cancer*. 2014;101:1020-1029.
7. Dunleavy K, Little RF, Pittaluga S, et al. The role of tumor histogenesis, FDG-PET, and short course EPOCH with dose-dense rituximab (SC-EPOCH-RR) in HIV-associated diffuse large B-cell lymphoma. *Blood*. 2010;115:3017-3024.
8. Dunleavy K, Pittaluga S, Shovlin M, et al. Low-intensity therapy in adults with Burkitt's lymphoma. *N Engl J Med*. 2013;369:1915-1925.
9. Barta SK, Xue X, Wang D, et al. Treatment factors affecting outcomes in HIV-associated non-Hodgkin lymphomas: a pooled analysis of 1546 patients. *Blood*. 2013;122:3251-3262.
10. Hentrich M, Hoffmann C, Mosthaf F, et al. Therapy of HIV-associated lymphoma-recommendations of the oncology working group of the German Study Group of Physicians in Private Practice Treating HIV-Infected Patients (DAGNA), in cooperation with the German AIDS Society (DAIG). *Ann Hematol*. 2014;93:913-921.
11. Uldrick TS, Pipkin S, Scheer S, Hessol NA. Factors associated with survival among patients with AIDS-related primary central nervous system lymphoma. *AIDS*. 2013;28:397-405.
12. Uldrick TS, Wang V, O'Mahony D, et al. An interleukin-6-related systemic inflammatory syndrome in patients co-infected with Kaposi sarcoma-associated herpesvirus and HIV but without Multicentric Castleman disease. *Clin Infect Dis*. 2010;51:350-358.
13. Uldrick TS, Polizzotto MN, Aleman K, et al. Rituximab plus liposomal doxorubicin in HIV-infected patients with KSHV-associated multicentric Castleman disease. *Blood*. 2014;124:3544-3552.
14. Guidelines for the Prevention and Treatment of Opportunistic Infections in HIV-Infected Adults and Adolescents: Human Papillomavirus Disease. U.S. Department of Health and Human Services, 2013. at http://aidsinfo.nih.gov/guidelines/html/4/adult-and-adolescent-oi-prevention-and-treatment-guidelines/343/hpv; Accessed January 29, 2015.
15. Kowalkowski MA, Day RS, Du XL, et al. Cumulative HIV viremia and non-AIDS-defining malignancies among a sample of HIV-infected male veterans. *J Acquir Immune Defic Syndr*. 2014;67:204-211.
16. Justice AC, Freiberg MS, Tracy R, et al. Does an index composed of clinical data reflect effects of inflammation, coagulation, and monocyte activation on mortality among those aging with HIV? *Clin Infect Dis*. 2012;54:984-994.
17. Tebas P, Stein D, Tang WW, et al. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med*. 2014;370:901-910.
18. Zou S, Glynn S, Kuritzkes D, et al. Hematopoietic cell transplantation and HIV cure: where we are and what next? *Blood*. 2013;122:3111-3115.

REVIEW QUESTIONS

1. A 47-year-old man with a history of sex with men presents to a clinic complaining of a purplish 1.3-cm lesion on his left calf, fatigue, and intermittent sweats. Initial evaluation reveals a temperature of 38.3° C, positive serology for HIV, 246 CD4 cells/mm^3, 1540 neutrophils/mm^3, and hemoglobin of 8.6 g/dL. A skin punch biopsy of the calf lesion reveals Kaposi sarcoma (KS). Evaluation and care includes all of the following ***except:***

A. cART with emtricitabine, tenofovir, ritonavir, and darunovir
B. A reticulocyte count
C. A blood test for C-reactive protein
D. A stool guaiac for occult blood
E. Recombinant erythropoietin to a target hemoglobin of greater than 10 g/dL.

Answer: E This patient presents with KS and symptoms suggestive of Kaposi sarcoma herpesvirus–associated Castleman disease (KSHV-MCD), including fever and severe anemia despite a relatively preserved CD4 count. Appropriate evaluation of anemia is indicated. This includes a reticulocyte count and evaluation for gastrointestinal blood loss. For patients in whom KSHV-MCD is suspected, C-reactive protein is an appropriate first test. Additionally, patients with HIV and KS should be started on combination antiretroviral therapy (cART). Erythropoiesis-stimulating agents were approved for use with zidovudine but not other cART. Given the risk of serious side effects from erythropoietin, especially when targeted to achieve hemoglobin levels that are close to normal, the decreased occurrence of anemia with newer cART, effective treatment of HIV-associated anemia with cART itself, and in this case the likelihood of an additional treatable cause of anemia (KSHV-MCD), erythropoietin would not be used.

2. A 35-year-old man presents with nausea, vomiting, fever, and weight loss. Computed tomography scan shows small bowel thickening with obstruction, adenopathy, and liver nodules. Biopsy from a small section of resected bowel shows Burkitt lymphoma. An HIV screening enzyme-linked immunosorbent assay (ELISA) is positive. CD4 count is 350 cells/mm^3. Which medication or medications should ***not*** be administered as an initial part of therapy?

A. Allopurinol
B. cART with emtricitabine, tenofovir, ritonavir, and darunovir
C. Rituximab
D. Full-dose cyclophosphamide
E. Trimethoprim-sulfamethoxazole

Answer: B Patients with Burkitt lymphoma require full-dose chemotherapy including rituximab. Important supportive care includes prevention of tumor lysis syndrome with allopurinol, and trimethoprim-sulfamethoxazole for *Pneumocystis jiroveci* pneumonia (PCP) prophylaxis.[8] Initiation of a ritonavir-based HIV regimen is not advised as an initial part of therapy owing to the CYP3A4 inhibitory activity of ritonavir and its effect on the pharmacokinetics of chemotherapy drugs needed for the treatment of Burkitt lymphoma.

3. A 47-year-old woman with HIV is found to have atypical squamous cells of undetermined significance on a Pap smear. CD4 count is 275 cells/mm^3. Colposcopy reveals no abnormalities. What established intervention is appropriate and most likely to reduce the risk of development of squamous intraepithelial lesions (SIL).

A. A 3-month course of interferon-alpha
B. cART with emtricitabine, tenofovir, ritonavir, darunovir
C. PAP smears every 6 months for 3 years
D. HBV vaccine
E. Valacyclovir 1000 mg daily

Answer: B cART is associated with improved immune control of HPV, the etiologic agent of cervical cancer, and reduced risk of squamous intraepithelial lesions (SIL).[3] Interferon-alpha is not indicated for CIN prevention. HBV vaccine prevents against hepatitis B and associated hepatocellular carcinoma. HIV-infected women are recommended to have Pap smears every 6 months for 1 year, then yearly thereafter if normal, *not* every 6 months for 3 years.

4. A 27-year-old African American man presents to an emergency room with right arm weakness. Physical exam is also remarkable for cachexia, oral candidiasis, and right lower extremity swelling. Brain magnetic resonance imaging shows three ring-enhancing masses with mass effect. A rapid HIV screen is positive. CD4 cell count is 5 cells/mm^3. Which of the following are ***not*** indicated as part of the initial diagnostic evaluation?

A. Lower extremity ultrasound
B. Brain biopsy if it can be done safely
C. 2 weeks of empirical pyrimethamine combined with sulfadiazine and leucovorin for possible toxoplasmosis, withholding definitive diagnostic tests during this period
D. Lumbar puncture with evaluation of cerebrospinal fluid
E. (^{18}F)-fluorodeoxyglucose positron emission tomography

Answer: C AIDS patients with ring-enhancing brain masses need urgent evaluation to establish a correct diagnosis. Appropriate evaluation includes nuclear imaging of the brain, evaluation of cerebrospinal fluid, and biopsy if it can be done safely. In the cART era, empirical treatment of toxoplasmosis as a diagnostic tool is no longer indicated, although the use of pyrimethamine combined with sulfadiazine and leucovorin for possible toxoplasmosis while awaiting biopsy results is reasonable in some cases, especially in patients with positive toxoplasmosis serology. Cancer increases risk for thrombosis, and this patient also has evidence of a possible right leg deep vein thrombosis that requires evaluation with a lower extremity ultrasound.

5. A 55-year-old man with HIV infection who has had sex with men and has a past history of intravenous drug use is being evaluated by a primary care physician. He currently smokes about 10 to 15 cigarettes per day. His CD4 count is 175 cells/mm^3. What is the most important intervention to reduce the risk of lung cancer?

A. Smoking cession intervention
B. Quadrivalent HPV vaccine
C. Empirical valganciclovir therapy for presumptive cytomegalovirus (CMV) infection
D. Screening for HCV infection and treatment if found to have chronic hepatitis C
E. Empirical trimethoprim-sulfamethoxazole and prednisone for *Pneumocystis jiroveci* pneumonia (PCP)

Answer: A Smoking cessation interventions are indicated for reducing the risk of lung cancer. Other common malignancies observed in patients with HIV that are also associated with smoking include oropharyngeal, cervical, anal, and liver cancer. Lung cancer is not associated with HPV or HCV. Lung cancer may be associated with chronic lung infections; however, empirical therapy for CMV or PCP is not established for the prevention of lung cancer.

394

NEUROLOGIC COMPLICATIONS OF HUMAN IMMUNODEFICIENCY VIRUS INFECTION

JOSEPH R. BERGER AND AVINDRA NATH

The neurologic complications of human immunodeficiency virus (HIV) infection can affect any portion of the neuraxis. They can be broadly divided into two large groups: those that are the consequence of HIV infection and those that are secondary in nature and occur chiefly as a result of the associated immunosuppression (Table 394-1). With respect to the former group, some of these disorders, such as acute HIV meningitis, are relatively rare, whereas others, such as HIV-associated neurocognitive disorders (HAND) and HIV-associated peripheral neuropathy, are common. HIV meningitis generally occurs at the time of seroconversion, but other disorders, such as HAND, HIV myelopathy, and HIV peripheral neuropathy, are typically not observed until advanced stages of immunosuppression.

The most common neurologic complications occurring as a secondary consequence of the virus are opportunistic infections. The most frequent of such infections are central nervous system (CNS) toxoplasmosis (Chapter 349), cryptococcosis (Chapter 336), tuberculosis (Chapter 324), cytomegalovirus encephalitis (Chapter 376), JC virus infection that results in progressive multifocal leukoencephalopathy (Chapter 370), and varicella-zoster virus encephalitis or myelitis (Chapter 375). However, many noninfectious

TABLE 394-1 CLASSIFICATION OF THE NEUROLOGIC COMPLICATIONS OF HIV INFECTION

DIRECT (HIV ASSOCIATED)	INDIRECT
Acute meningitis	Opportunistic infections
Chronic meningitis	Neoplasms
Encephalopathy	Toxic/metabolic
Myelopathy	Drug effects
Peripheral neuropathy	Cerebrovascular disease
Myositis/myopathy	
Neurocognitive disorders	

complications occur as well. The incidence of primary CNS lymphoma is remarkably high. Unlike non–AIDS-associated primary CNS lymphoma, in which Epstein-Barr virus can be recovered from only 50% of tumors (Chapter 185), all AIDS-related primary CNS lymphomas have been associated with Epstein-Barr virus infection (Chapter 393). A wide variety of toxic and metabolic complications have been observed in HIV-infected patients, including Wernicke encephalopathy and vitamin B_{12} deficiency. Drugs, particularly antiretroviral agents, have also been associated with neurologic complications. Certain nucleoside analogues (e.g., the "d" drugs didanosine [ddI], dideoxycytidine [ddC], and stavudine [d4T]; see Table 394-5) are frequently associated with peripheral neuropathy. Finally, a wide variety of cerebrovascular disorders have been associated with HIV infection. Thrombotic ischemic stroke in this population has been attributed to an as yet undefined procoagulant tendency. Certain infections seen with increased frequency in patients infected with HIV, such as syphilis, tuberculosis, and cryptococcosis, may be associated with stroke. In addition, CNS vasculitis has occasionally been observed with HIV infection, either from HIV or from an associated varicella-zoster infection.

In recent years, it is being increasingly recognized that in a few weeks or months after initiation of combination antiretroviral therapy (ART), some patients will clinically deteriorate even when there is a robust drop in plasma viral load and recovery of CD4 cell counts. In these patients, the recovery of immune function may result in an inflammatory syndrome termed *immune reconstitution inflammatory syndrome* (IRIS; Chapter 395). This most often occurs when a patient has an underlying opportunistic infection, in which case the inflammation is targeted to the site of the infection. On occasion, IRIS may occur in the brain without any identifiable opportunistic infection and is attributable to a restoration of a robust immune response to HIV itself.

Few HIV-infected individuals will escape experiencing one or more of these complications during their lifetime.[1] They are responsible for a significant amount of morbidity and mortality. Not infrequently, more than one neurologic condition coexists in the same patient, which is often a source of confusion because one of the illnesses may be appropriately diagnosed and treated, yet the patient continues to deteriorate. Furthermore, one must also consider that any neurologic abnormalities observed are the consequence of more pedestrian disorders, such as radiculopathy secondary to disc herniation, and are unrelated to the HIV infection. In patients not responding to appropriate therapy for a proven disorder, careful clinical reevaluation is essential.

HIV-ASSOCIATED NEUROCOGNITIVE DISORDERS

DEFINITION

HIV dementia has been referred to by a number of names, including subacute encephalitis, multinucleated giant cell encephalitis, AIDS dementia complex, and HIV-associated neurocognitive disorders (HAND).

A spectrum of increasing cognitive impairment has been described in association with HIV infection. An acquired impairment in cognitive function involving two ability domains, with a performance at least 1.0 standard deviations (SD) below the mean for norms on standardized neuropsychological tests but no cognitive impairment in everyday function, has been referred to as asymptomatic neurocognitive impairment. Patients with mild neurocognitive disorder have a similar disorder on neuropsychological testing but exhibit problems with daily functions. Those with HIV dementia have marked impairment (≥2 SD) on at least two domains in neuropsychological testing and marked interference in day-to-day activities because of cognitive impairment. By definition, the diagnosis of HAND requires that there be no delirium and no other etiology to explain the deficit.

EPIDEMIOLOGY

In the pre-ART era, HIV dementia was observed more commonly. Early studies suggested that it was evident in more than 50% of preterminal patients. The Multicenter AIDS Cohort Study found an incidence of 4% coincident with the diagnosis of AIDS, and dementia developed in 7% within 1 year of the development of AIDS and in 14% within 2 years. Other studies have found slightly higher rates. After the introduction of combination ART in 1996, the incidence rate of HIV dementia declined substantially, but the prevalence of the disorder increased; this finding may be the result of prolonged survival of affected patients.[2] Studies suggest that the cumulative incidence of HAND may be climbing to involve as many as 50% of patients with HIV infection, but most patients have milder forms of cognitive impairment.[3] Risk factors for HAND include low CD4 cell counts, concurrent anemia, extremes of age, and history of substance abuse or addiction. Depression is also a common comorbidity. Congenitally or perinatally infected children may exhibit delayed development. Genetic factors such as *APOE4* and polymorphisms in chemokine monocyte chemotactic protein-1 (CCL-2) and tumor necrosis factor receptor have been identified as risk factors for the development of HAND.

PATHOBIOLOGY

HIV enters the brain early after the initial infection. In all likelihood, the virus enters the brain in infected mononuclear cells, although it is possible that it also enters as cell-free virus. HIV can be demonstrated in the brain as early as 2 weeks after infection. The viral infection predominates in invading macrophages in the brain, microglial cells, and multinucleated giant cells, usually in perivascular areas. HIV can also infect astrocytes, and studies suggest that in patients with cognitive impairment, 16 to 19% of astrocytes may be infected. HIV may remain latent for extended periods. Infection of other cell types is a rare event. The brain may thus be an important reservoir for HIV and a potential source for the emergence of drug-resistant viruses. Certain proteins of the virus, Tat and gp120, have been demonstrated to have a significant pernicious effect on neuronal function and viability either through direct toxic effects on neurons or by stimulation of glial cells, which in turn produce neurotoxic metabolites—cytokines and chemokines—or induce oxidative stress. Similarly, the perivascular HIV-infected inflammatory cells produce a wide variety of chemokines and cytokines that also have similar deleterious effects. Therefore, there appear to be parallel paths, direct viral neurotoxicity and toxicity from inflammatory byproducts, that lead to the development of HAND.[4] Drugs of abuse, such as opiates, cocaine, and methamphetamine, can stimulate HIV replication in glial cells and synergize with the HIV proteins to cause neurotoxicity and glial cell activation.

On gross examination, the brain typically exhibits cortical atrophy, sulcal widening, and ventricular dilation. Histopathologic examination shows multinucleated giant cells, microglial nodules, white matter pallor, astrogliosis, and perivascular inflammation (Fig. 394-1). The multinucleated giant cell, which is a syncytium of macrophages, has been regarded as the hallmark of infection, but it is not uniformly present in patients with HAND. The pathologic changes are most prominent in the basal ganglia, prefrontal cortex, and hippocampus. Concomitant vacuolar myelopathy is seen in as many as 40% of patients, although it has become less common since the introduction of ART.

CLINICAL MANIFESTATIONS

Most commonly, patients describe impaired memory, poor concentration, impairment of executive function, and difficulty reading. They often find it difficult to recall a passage recently read. Other frequent symptoms include gait problems, depression, and tremors, and patients may appear apathetic and slow or become socially withdrawn. Headaches, fatigue, and sexual dysfunction are frequent. The most commonly observed physical findings are psychomotor slowing with bradyphrenia and impaired rapid repetitive and alternating movements, hyperreflexia, increased tone, facial masking, frontal release signs (snout, glabellar, involuntary grasp), and abnormal ocular motility with breakdown of smooth pursuit. The dementia has the features of a subcortical dementia not inconsistent with the relatively marked disease burden observed in dopaminergic basal ganglia. Screening tools for HAND have been proposed (Table 394-2) and a battery of specific neuropsychological tests used (Table 394-3). Staging of the disease along a continuum has been widely adopted (Table 394-4).

DIAGNOSIS

Computed tomography of the head typically shows brain atrophy, but findings may be normal. Central atrophy is generally more pronounced than cortical atrophy. In children, calcification of the basal ganglia is often observed. Similar findings are seen on magnetic resonance imaging (MRI). Basal ganglia hyperintensity is seen rarely on T2-weighted images and may resolve during the course of ART. White matter hyperintensities are also commonly detected on T2-weighted images and fluid-attenuated inversion recovery (FLAIR) sequences (Fig. 394-2). These lesions may be discrete and focal or large and confluent. They are generally symmetrically distributed and may be confused with lesions of progressive multifocal leukoencephalopathy. Unlike progressive multifocal leukoencephalopathy lesions, they are typically not hypointense on T1-weighted sequences.

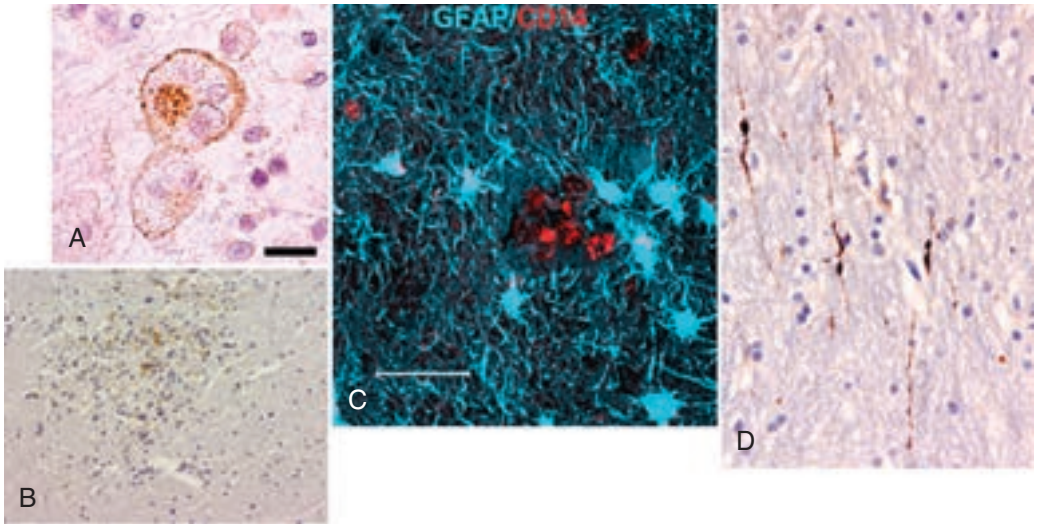

FIGURE 394-1. Histopathology of HIV dementia. A, Multinucleated giant cell immunostaining for HIV antigen. **B,** A microglia nodule shows a collection of inflammatory cells. The HIV-infected cells are immunostained for HIV antigen and are brown. **C,** A confocal photomicrograph shows astrocytosis. The astrocytes are stained blue with antibody to glial acidic fibrillary protein (GAFP). The macrophages are immunostained with antibody to CD14 (*red*). **D,** In neurons immunostained with antibody to amyloid precursor protein, beading of the neuritis suggests interruption of axonal flow. (**C,** Courtesy Carlos Pardo, Johns Hopkins University, Baltimore, MD; **D,** Courtesy Chris Zink, Johns Hopkins University, Baltimore, MD.)

TABLE 394-2 SCREENING TEST FOR HIV-ASSOCIATED NEUROCOGNITIVE DISORDERS

MAXIMUM SCORE	SCORE	
		MEMORY REGISTRATION
		Give 4 words to recall (dog, hat, green, peach), 1 second to say each. Then ask the patient all 4 after you have said them.
4	( )	ATTENTION
		Antisaccadic eye movements: 20 commands _____ errors of 20 trials ≤3 errors = 4; 4 errors = 3; 5 errors = 2; 6 errors = 1; >6 errors = 0
6	( )	PSYCHOMOTOR SPEED
		Ask patient to write the alphabet in uppercase letters horizontally across the page (use the back of this form) and record the time: _____sec ≤21 sec = 6; 21.1-24 sec = 5; 24.1-27 sec = 4; 27.1-30 sec = 3; 30.1-33 sec = 2; 33.1-36 sec = 1; >36 sec = 0
4	( )	MEMORY RECALL
		Ask for 4 words from Registration above. Give 1 point for each correct. For words not recalled, prompt with a "semantic" clue, as follows: animal (dog); piece of clothing (hat), color (green), fruit (peach). Give ½ point for each correct after prompting.
2	( )	CONSTRUCTION
		Copy a cube; record the time: _____sec <25 sec = 2; 25-35 sec = 1; 35 sec = 0
TOTAL SCORE: _____/16		

From Power C, Selnes OA, Grim JA, McArthur JC. HIV dementia scale: a rapid screening test. *J Acquir Immune Defic Syndr Hum Retrovirol.* 1995;8:273-278.

TABLE 394-3 NEUROPSYCHOLOGICAL BATTERY PROPOSED FOR EVALUATION OF HIV-ASSOCIATED NEUROCOGNITIVE DISORDERS

Fine motor control Grooved pegboard Finger tapping
Rapid sequential problem solving Trail making A and B Digit symbol
Visuospatial problem solving Block design
Spontaneity Verbal fluency
Visual memory Visual reproduction

TABLE 394-4 MEMORIAL SLOAN-KETTERING SCALE FOR HIV-ASSOCIATED NEUROCOGNITIVE DISORDERS

Stage 0 (normal)	Normal mental and motor function
Stage 0.5 (equivocal or subclinical)	Absent, minimal, or equivocal symptoms without impairment of work or capacity to perform activities of daily living
Stage 1 (mild)	Able to perform all but the more demanding aspects of work or activities of daily living
Stage 2 (moderate)	Able to perform all the basic activities of daily self-care but cannot work or maintain the more demanding aspects of daily life
Stage 3 (severe)	Major intellectual incapacity or motor disability with slowing
Stage 4 (end stage)	Nearly vegetative

From Price RW, Brew BJ. The AIDS dementia complex. *J Infect Dis.* 1988;158:1079-1083.

TREATMENT

Rx

Whether ART needs to cross the blood-brain barrier for a better neurocognitive outcome remains uncertain. Higher cerebrospinal fluid (CSF) penetrant ability is associated with lower levels of HIV RNA in the CSF.[5] Prudence dictates that highly CNS-penetrant ARTs would best be used, although other factors should be considered (e.g., the drug's pharmacokinetics, protein binding, and IC_{95}). Brain HIV RNA levels have been demonstrated to be significantly higher in patients with HIV encephalopathy.[6] Newer CCR5 antagonists, maraviroc and vicriviroc, demonstrate good CSF penetration and may be uniquely suited for the treatment of HAND. Table 394-5 provides a list of ARTs with their respective CSF-to-plasma ratios, which may provide a proxy for CNS penetration. A variety of other therapies have been proposed for the treatment of HAND that chiefly addresses the proposed inflammatory pathway of the pathogenesis of the disorder. To date, none has been established to be effective. The flow chart in Figure 394-3 provides an algorithm for the management of HAND. Careful attention to treatment of comorbidities is also essential. Special consideration needs to be given to the choice of anticonvulsants, since drug-drug interactions are common owing to induction of hepatic enzymes or serum albumin binding properties that may alter drug levels of ARTs.

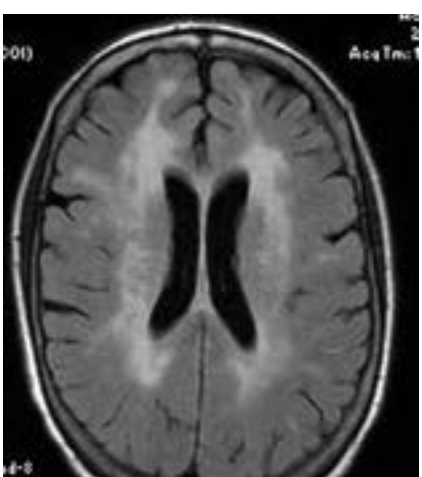

FIGURE 394-2. Magnetic resonance image of the brain in HIV dementia. The FLAIR image shows confluent areas of signal hyperintensity adjacent to ventricles.

TABLE 394-5 CEREBROSPINAL FLUID–TO-PLASMA RATIOS FOR ANTIRETROVIRAL THERAPIES

	CEREBROSPINAL FLUID–TO-PLASMA RATIO
NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS	
Zidovudine (AZT)	0.3-1.35
Stavudine (d4T)	0.16-0.97
Abacavir (ABC)	0.3-0.42
Didanosine (ddI)	0.16-0.19
Lamivudine (3TC)	0.11
Zalcitabine (ddC)	0.09-0.37
Emtricitabine	0.04
NUCLEOTIDE REVERSE TRANSCRIPTASE INHIBITORS	
Tenofovir	<0.05
FUSION INHIBITORS	
Enfuvirtide	NA
NON-NUCLEOSIDE REVERSE TRANSCRIPTASE INHIBITORS	
Nevirapine (NVP)	0.28-0.45
Delavirdine	0.02
Efavirenz	0.01
PROTEASE INHIBITORS	
Indinavir	0.02-0.76
Saquinavir	<0.05
Nelfinavir	<0.05
Ritonavir	<0.05
Amprenavir	<0.05
Lopinavir	<0.05
Atazanavir	0.0021-0.0226
Fosamprenavir	<0.05

Adapted from McArthur JC, Haughey N, Gartner S, et al. Human immunodeficiency virus–associated dementia: an evolving disease. *J Neurovirol.* 2003;9:205-221.

PROGNOSIS

In the pre-ART era, the life expectancy of patients with HIV dementia was less than 6 months. The institution of effective ART has considerably increased the life expectancy of affected patients, and partial recovery of neurocognitive deficits may be seen.

HIV MYELOPATHY

A broad spectrum of myelopathies occurs with HIV infection (Table 394-6). HIV causes a distinct myelopathy referred to as HIV vacuolar myelopathy. Evident in 20 to 55% of autopsies, this myelopathy is under-recognized clinically. It is characterized by an insidious onset of leg weakness and gait abnormality that usually occurs during the course of advanced HIV infection. Sensory complaints include vague leg discomfort and distal paresthesias. Bowel and bladder dysfunction is frequent. Physical examination reveals spastic paraparesis with lower extremity hyperreflexia, a spastic-ataxic gait,

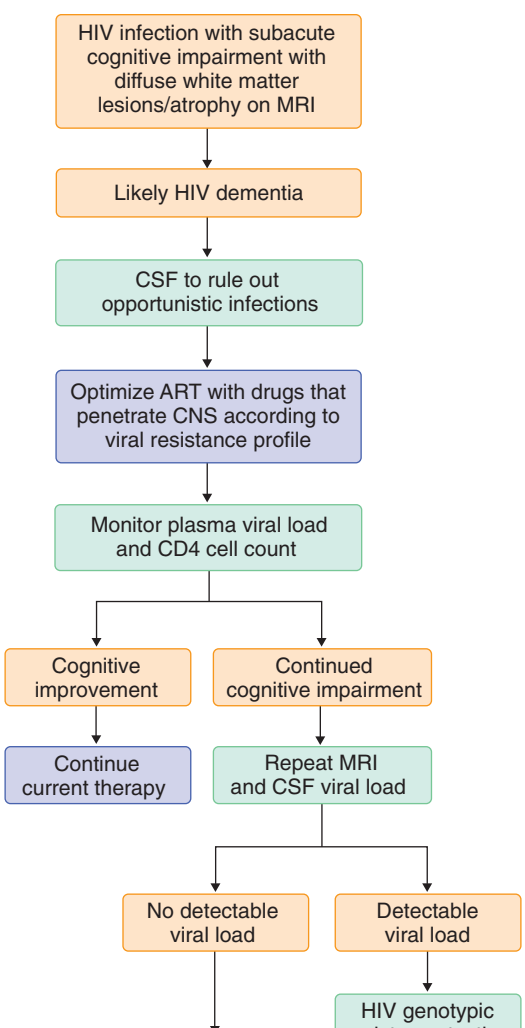

FIGURE 394-3. Flow chart for the treatment of HIV dementia. ART = antiretroviral therapy; CNS = central nervous system; CSF = cerebrospinal fluid; HIV = human immunodeficiency virus; MRI = magnetic resonance imaging.

and impaired sensation with vibratory and position perception disproportionately affected. The sensory loss may be asymmetrical, but the presence of a discrete sensory level should suggest an alternative cause of the myelopathy. MRI of the spinal cord is usually normal but may show cord atrophy or hyperintense signal abnormalities within the cord on T2-weighted images. The value of spinal MRI lies in exclusion of other diagnostic possibilities, particularly structural abnormalities of the cord. Somatosensory evoked potentials generally demonstrate delayed conduction. Although findings on gross examination of the cord and dura are usually normal, histopathologic examination typically reveals loss of myelin with spongy degeneration, axonal preservation, microglial nodules, and multinucleated giant cells involving the lateral and posterior columns of the spinal cord. The pathogenesis of this disorder remains obscure. The clinical and pathologic features are suggestive of vitamin B_{12} deficiency (Chapter 164), but no consistent abnormalities are found in serum levels of vitamin B_{12} or its metabolites, and replacement therapy with cyanocobalamin or *S*-methyl-L-methionine has not proved to be beneficial. The most effective management appears to be ART coupled with physical therapy.

HIV PERIPHERAL NEUROPATHY

Although a large number of peripheral neuropathies have been described in association with HIV infection, the one most commonly observed is a distal symmetrical sensorimotor peripheral neuropathy that ultimately affects at least a third of all HIV-infected persons. This neuropathy is seen more often in advanced disease. The patient typically complains of "burning feet." Distal paresthesias and numbness are reported, and the pain may be debilitating.

TABLE 394-6 SPECTRUM OF MYELOPATHIES IN HIV INFECTION

HIV ASSOCIATED
Vacuolar myelopathy
Acute myelitis
Relapsing remitting encephalomyelitis
Spinal myoclonus
VIRAL
Cytomegalovirus
HTLV-1 and HTLV-2
Varicella-zoster virus
Measles virus
Progressive multifocal leukoencephalopathy
BACTERIAL
Mycobacterium tuberculosis
Treponema pallidum
Pseudomonas cepacia
FUNGAL
Cryptococcus immitis
Aspergillus spp.
Nocardia
PARASITIC
Toxoplasma gondii
Schistosoma
MALIGNANCY
Primary CNS lymphoma
Metastatic CNS lymphoma
Other tumors (e.g., glioma)
VASCULAR
Necrotizing vasculitis
Disseminated intravascular coagulation
TOXIC/METABOLIC
Vitamin B_{12} deficiency
Protease inhibitor epidural lipomatosis

CNS = central nervous system; HTLV = human T-lymphotropic virus.

Distal vibratory, pinprick, and temperature sensory perceptions are diminished, and the Romberg sign is often abnormal. Ankle jerks are depressed or absent, and mild weakness of the toes with associated atrophy of the intrinsic muscles of the foot is frequently noted. Nerve conduction studies show reduced sensory nerve action potentials and conduction velocities. Nerve conduction amplitudes are reduced disproportionately to the reduction in conduction velocity. Electromyography may demonstrate features consistent with either acute or chronic denervation. Skin biopsy shows a reduction in cutaneous nerve fibers.

A clinically indistinguishable neuropathy may be seen with some of the antiretroviral agents, particularly with ddI, ddC, and d4T. The neuropathy associated with these agents appears to be dose dependent. Other drugs used in treatment of AIDS or its complications, such as hydroxyurea, isoniazid, vincristine, and thalidomide, may result in a similar peripheral neuropathy. Discontinuation of the use of these drugs should result in improvement in the neuropathic features; however, a phenomenon referred to as coasting, in which continued worsening of the neuropathy occurs after discontinuation of the offending agent, may be observed for a period of several months.

Shortly after seroconversion, a demyelinating peripheral neuropathy may be observed that is identical to Guillain-Barré syndrome or chronic inflammatory polyradiculoneuropathy (Chapter 420) if it progresses for a duration of 3 weeks or more. An autoimmune process is the probable pathogenesis, and these neuropathies respond to therapy routinely used in the treatment of these conditions. In rare instances, mononeuritis multiplex (Chapter 420) may be seen with HIV infection.

HIV MYOPATHY

Several forms of muscle disease occur with HIV infection and, like other neurologic disorders, may be due to HIV or be secondary to other processes. Among the latter are myopathies associated with drugs such as zidovudine and cholesterol-lowering agents, opportunistic infections, neoplastic infiltrates, and vasculopathies. HIV myopathy mirrors the clinical and laboratory findings of classic polymyositis (Chapter 269). It is characterized by a progressive symmetrical weakness of the limb girdle and neck flexor muscles. Fatigue, myalgias, and wasting are observed in up to 50% of affected persons. Muscle enzymes, including creatine kinase and aldolase, are elevated, and the electromyogram shows short, brief, polyphasic unit action potentials. Muscle biopsy reveals myofibrillar necrosis, phagocytosis, variation in fiber size, and regeneration and degeneration, typically with endomysial inflammatory infiltrates. This myopathy is rare, with an incidence of less than 1%, and has been observed in any stage of HIV infection. All patients with weakness who are taking a nucleoside reverse transcriptase inhibitor (NRTI) should have blood lactate levels measured because of an association between a recently described neuromuscular weakness syndrome and lactic acidosis in patients taking NRTIs. The contribution of zidovudine to the genesis of HIV-associated myopathy remains controversial. As in non-HIV polymyositis, corticosteroid therapy is beneficial and can be administered with tolerable side effects.

HIV-ASSOCIATED CENTRAL NERVOUS SYSTEM IRIS

Some individuals may develop a subacute encephalopathy following the introduction of ARTs. Symptoms include impairment in neurocognitive function, seizures, and ultimately impairment of consciousness and coma; if untreated, the syndrome may result in death.[7] Risk factors include a low CD4 lymphocyte nadir (usually < 50 cells/mm^3) or a rapid decline in viral load following initiation of ART. Although the exact mechanism is unclear, the syndrome is associated with immune restoration (Chapter 395) and infiltration of CD8 cytotoxic T cells into the brain. MRI may show diffuse multifocal white matter hyperintensities on FLAIR and T2-weighted images suggestive of edema. CSF shows a mild lymphocytosis. In some of these patients, the CSF HIV viral load is greater than that in the plasma. It thus appears that the target of the immune restoration may be the viral reservoir in the brain. A second type of CNS IRIS, with HIV infection manifesting as a fulminant focal encephalitis associated with demyelination, has also been described. This resembles the Marburg variant of multiple sclerosis. Successful treatment with corticosteroids has been described for both forms of IRIS (Chapter 395), although no clinical trials have been performed.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bilgrami M, O'Keefe P. Neurologic diseases in HIV-infected patients. *Handb Clin Neurol*. 2014;121:1321-1344.
2. Heaton RK, Franklin DR Jr, Deutsch R, et al. Neurocognitive Change in the Era of HIV Combination Antiretroviral Therapy: The Longitudinal CHARTER Study. *Clin Infect Dis*. 2015;60:473-480.
3. Spudich S. HIV and neurocognitive dysfunction. *Curr HIV/AIDS Rep*. 2013;10:235-243.
4. Gannon P, Khan MZ, Kolson DL. Current understanding of HIV-associated neurocognitive disorders pathogenesis. *Curr Opin Neurol*. 2011;24:275-283.
5. Letendre S. Central nervous system complications in HIV disease: HIV-associated neurocognitive disorder. *Top Antivir Med*. 2011;19:137-142.
6. Gelman BB, Lisinicchia JG, Morgello S, et al. Neurovirological correlation with HIV-associated neurocognitive disorders and encephalitis in a HAART-era cohort. *J Acquir Immune Defic Syndr*. 2013;62:487-495.
7. Johnson T, Nath A. Immune reconstitution inflammatory syndrome and the central nervous system. *Curr Opin Neurol*. 2011;24:284-290.

REVIEW QUESTIONS

1. The diagnosis of HIV-associated neurocognitive disorders relies on:

A. Neuropsychological testing demonstrating an impairment in two ability domains at least 1.0 standard deviation below the mean for norms but no cognitive impairment in everyday function
B. Defects in cognitive function resulting in even mild impairment of daily function
C. Neuropsychological testing demonstrating more than 2.0 standard deviations in at least two cognitive domains
D. SufficFactors demonstrate that impaired memory makes reading difficult.
E. Abnormalities on brain neuroimaging, with head CT and T2-weighted images and fluid-attenuation inversion recovery (FLAIR) sequences by MRI as determined by standardized criteria for the diagnosis

Answer: A Although HAND may be suspected by the patient's complaints, the cognitive impairment is typically subtle because it does not affect day-to-day function. Therefore, HAND is best diagnosed by employing formal neuropsychological testing. Imaging studies are important adjunctive diagnostic tests and also useful for follow-up but by themselves do not make the diagnosis.

2. Factors that may contribute to the development of HIV dementia include ***all but one*** of the following:

A. *APOE4* heterozygosity
B. Age
C. Profound immunosuppression
D. Prior substance or alcohol abuse
E. Race

Answer: E A wide variety of factors have been associated with HIV dementia. It is more likely in the setting of *APOE4* heterozygosity, advanced age, profound immunosuppression, and prior substance and alcohol abuse. However, there is no evidence that any particular race is more or less likely to be affected when one controls for other factors.

3. A common early complaint of individuals affected by HIV-dementia is:

A. Headache
B. Seizures
C. Difficulty reading
D. Paresthesias
E. Getting lost in familiar environments

Answer: C The inability to recall what one has just read is a common complaint among persons with early HIV dementia. This complaint is likely the result of impaired concentration and poor memory. Although headaches are reported, they correlate poorly with HIV dementia. Seizures occur rarely and usually, like the complaint of getting lost in familiar environments, in more advanced disease. Paresthesias are generally the consequence of an associated peripheral neuropathy.

4. A feature that distinguishes HIV myelopathy from many of the other causes of HIV-associated myelopathy (e.g., lymphoma, varicella-zoster myelitis) is:

A. Spasticity of the lower extremities
B. Imbalance and gait disturbance
C. Sphincter dysfunction
D. The absence of a discrete sensory level over the trunk to pinprick
E. Babinski signs

Answer: D Unlike many of the other myelopathies that occur with HIV infection, HIV vacuolar myelopathy does not exhibit a discrete sensory level over the trunk. The presence of such a level should suggest another etiology of the spinal cord disorder.

5. ***All but one*** of the following statements regarding peripheral neuropathy associated with AIDS is true:

A. Many of the medications used in the treatment of AIDS can result in a clinically identical peripheral neuropathy.
B. A demyelinating peripheral neuropathy resembling Guillain-Barré syndrome or chronic inflammatory peripheral neuropathy may be observed shortly after initial infection.
C. Peripheral neuropathy is the most common neurologic abnormality observed in patients with HIV infection.
D. Skin biopsy with examination of the cutaneous nerves may be helpful in establishing the diagnosis.
E. Recovery from HIV-associated peripheral neuropathy often follows the use of effective antiretroviral therapy.

Answer: E Although symptomatic therapy is often helpful for the pain and paresthesias that occur with HIV-associated peripheral neuropathy, reversal of the disorder is virtually unheard of. In addition to many antiretroviral therapies, particularly, the "d" drugs, other therapies frequently employed in patients with AIDS (e.g., isoniazid, thalidomide) may cause a clinically similar peripheral neuropathy.

395

IMMUNE RECONSTITUTION INFLAMMATORY SYNDROME IN HIV/AIDS

ROBERT COLEBUNDERS AND MARTYN A. FRENCH

DEFINITION

Treatment of human immunodeficiency virus (HIV) infection with combination antiretroviral therapy (ART) results in the restoration of protective pathogen-specific immune responses and the regression or prevention of opportunistic infections (OIs) and cancers in most individuals (Chapter 389). However, restoration of an immune response against a pathogen may also result in immunopathology at body sites infected by the pathogen. This has been referred to as immune restoration disease (IRD) to differentiate it from immunodeficiency disease, but it is now commonly known as immune reconstitution inflammatory syndrome (IRIS), because an inflammatory illness is the most common clinical feature.[1]

Essentially any pathogen that causes an infection as a result of HIV-induced cellular immunodeficiency may be associated with IRIS after ART is commenced. However, the clinical characteristics and severity of IRIS associated with each type of pathogen vary greatly (Table 395-1). For example, IRIS associated with *Mycobacterium tuberculosis*, cryptococcal, or JC polyomavirus (JCV) infection is manifested differently from an OI caused by these

TABLE 395-1 EXAMPLES OF IMMUNE RECONSTITUTION INFLAMMATORY SYNDROME

PATHOGEN	NOMENCLATURE	TYPICAL CHARACTERISTICS OF THE DISEASE
Mycobacterium tuberculosis	TB-IRIS	Paradoxical exacerbation of TB
Nontuberculous mycobacteria (NTM)	NTM immune reconstitution syndrome	Mainly lymphadenitis, also pulmonary disease and hepatitis
Bacille Calmette-Guérin (BCG)	BCG-IRIS	Necrotizing regional lymphadenitis
Mycobacterium leprae	Leprosy-related IRIS	Borderline and type 1 reactional state
Cryptococcus neoformans	Cryptococcal-IRIS	Mainly meningitis, also lymphadenitis
Pneumocystis jiroveci	*Pneumocystis*-IRIS	Paradoxical exacerbation of pneumonitis
Cytomegalovirus (CMV)	CMV retinitis after ART or immune recovery uveitis	Acute retinitis after commencing ART or uveitis
JC polyomavirus	Inflammatory PML	Multifocal leukoencephalopathy with inflammatory features
Human herpesvirus 8	KS-IRIS	Rapid progression of existing and/or new KS lesions
Hepatitis B or C virus	Some cases of ART-associated hepatotoxicity	Hepatitis flare and/or liver enzyme elevation
Herpes zoster virus		Mainly dermatomal or multidermatomal zoster, occasionally myelitis
Molluscum contagiosum virus	Inflammatory molluscum contagiosum	Inflamed molluscum lesions
Malassezia spp.	Inflammatory seborrheic dermatitis	Abnormally inflamed seborrheic dermatitis

ART = antiretroviral therapy; IRIS = immune reconstitution inflammatory syndrome; KS = Kaposi's sarcoma; PML = progressive multifocal leukoencephalopathy; TB = tuberculosis.

pathogens, and the resulting illness is often severe and may result in death. In contrast, herpes zoster after ART is usually indistinguishable from that occurring before ART, and it is only the timing of onset that suggests that it results from IRIS.

IRIS develops mainly during the first 3 months of ART but occasionally later. Two patterns are recognized. Paradoxical IRIS refers to the worsening or atypical manifestation (or both) of an established OI after ART is commenced. In most cases the infection had been treated before ART was initiated, and the immune response appears to be against residual antigens of the pathogen. Unmasking IRIS refers to disease that occurs for the first time after ART is commenced and appears to result from an immune response against a subclinical infection by an opportunistic pathogen.[2]

EPIDEMIOLOGY

The reported incidence of IRIS has varied from 8% to over 40% in different studies. To some extent, the large variation reflects the lack of universally accepted diagnostic criteria. It also probably reflects differences in risk factors in the populations of patients studied. The most important risk factors for IRIS are a low $CD4^+$ T-cell count when ART is initiated and, in patients in whom paradoxical IRIS develops, disseminated infection and a short time between treatment of the infection and commencement of ART.

PATHOBIOLOGY

Information about the pathogenesis of IRIS has been obtained mostly by studying patients who experience disease associated with a mycobacterial infection.[3] Both clinicopathologic and immunologic studies have shown an association with a T_H1 cellular immune response against mycobacterial antigens, which has been demonstrated by measuring delayed-type hypersensitivity skin test responses or the frequency of circulating antigen-specific T cells that produce interferon (IFN)-γ. However, there is increasing evidence that innate immune responses by myeloid cells (monocytes, macrophages, and neutrophils) and their mediators also contribute to the immunopathology, particularly in paradoxical tuberculosis (TB)-IRIS. The immunopathogenesis of IRIS associated with other pathogens is less well understood and appears to vary depending on the provoking pathogen. For example, IRIS associated with JCV infection (inflammatory progressive multifocal leukoencephalopathy [PML]) is characterized by an inflammatory cell infiltrate dominated by $CD8^+$ T cells in affected areas of the brain.

CLINICAL MANIFESTATIONS

The clinical manifestations of IRIS are different for each associated pathogen and will therefore be described for each pathogen. Only disease that presents a significant patient management problem will be discussed.

Mycobacterium tuberculosis

M. tuberculosis is the most common pathogen involved in IRIS, with estimates of incidence ranging from 7 to 43% of patients with HIV infection and treated TB. Most TB-IRIS develops within the first 3 months after the initiation of ART. Patients in whom paradoxical TB-IRIS develops typically give a history of having improved with treatment of TB before initiation of ART. After starting ART, recurrent, worsening, or new clinical or radiologic manifestations of TB develop. Common manifestations include fever, enlargement of lymph nodes, and worsening radiographic pulmonary infiltrates. Tracheal compression by intrathoracic lymph nodes or massive pleural effusions can cause life-threatening dyspnea. Respiratory failure as a result of worsening pulmonary infiltrates and acute respiratory distress syndrome have occasionally been reported. In a large prospective cohort of HIV-TB coinfected patients in Mozambique, TB-IRIS occurrence within 12 weeks of starting ART was independently associated with the mortality of HIV-TB coinfected patients at 48 weeks post ART initiation.[3] In a prospective case series from South Africa, neurologic TB-IRIS accounted for 12% of cases of paradoxical TB-IRIS. Meningitis, tuberculoma, or both were the most common manifestations. TB-IRIS should be considered as a potential cause of hepatitis, especially in patients with disseminated TB who are being treated with a combination of anti-TB therapy and ART. Peritonitis secondary to bowel perforation and splenic rupture are other unusual findings. Though usually negative, mycobacterial cultures may be positive, particularly if IRIS occurs early during anti-TB therapy and in patients with multidrug-resistant TB. Histologic examination often reveals necrotizing granulomas.

High rates of TB have been reported during ART, especially in the initial months of treatment in ART programs in resource-limited settings. This type of TB has been referred to as ART-associated TB because the mechanisms underlying the manifestations of TB after initiating ART are likely to be heterogeneous. Diagnoses of active TB before initiation of ART may be missed because of the inherent insensitivity of TB diagnostics in this patient group and only later be diagnosed during ART. Because ART-induced immune recovery is a time-dependent process and some patients fail to respond immunologically, a proportion of cases may develop as a result of persisting immunodeficiency. Other patients may have active subclinical disease at the time of ART initiation, and progression to symptomatic disease may be accelerated by ART-induced restoration of a cellular immune response against *M. tuberculosis* antigens. Of patients in this latter group, some have exuberant inflammatory clinical features that are consistent with a diagnosis of TB-IRIS.

Nontuberculous Mycobacteria

Atypical manifestations of *Mycobacterium avium* complex (MAC) disease in patients who had commenced zidovudine monotherapy were the first indication that IRIS may be a complication of ART. MAC and other nontuberculous mycobacteria have been associated with IRIS in 3 to 4% of patients who commence combination ART with a $CD4^+$ T-cell count lower than 100/μL. Disease is usually localized, as opposed to the disseminated nontuberculous mycobacterial disease of patients with acquired immunodeficiency syndrome (AIDS), and is most commonly manifested as fever, night sweats, and lymphadenitis. Unmasking disease is most common. Peripheral lymphadenitis may suppurate and sometimes cause chronically discharging fistulas to the skin. Abdominal disease frequently causes pain, which is usually associated with lymphadenitis and occasionally with omental masses, hepatitis, and inflammation of the spleen (Fig. 395-1). Pulmonary and thoracic disease usually

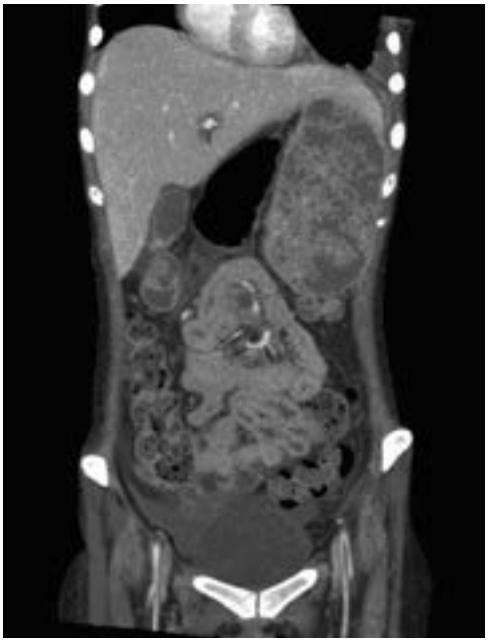

FIGURE 395-1. ***Mycobacterium avium*** **complex–associated immune reconstitution inflammatory syndrome manifested as necrotizing inflammation in the spleen and abdominal lymph nodes.**

causes cough that is sometimes associated with chest pain. Bronchoscopy may reveal endobronchial lesions. Microscopic examination of biopsy material or aspirates from affected tissues often reveals mycobacteria, but these may not be cultured.

In HIV-seropositive children vaccinated with bacille Calmette-Guérin (BCG), a BCG-associated lymphadenitis and abscesses may develop after starting highly active antiretroviral therapy (HAART) (Fig. 395-2).

Leprosy-related IRIS is usually manifested as unmasking of previous subclinical *Mycobacterium leprae* infection, with a borderline and a type I reactional state.

Cryptococci

The proportion of patients with HIV infection and treated cryptococcosis in whom cryptococcosis-IRIS develops ranges from 8 to 42%. The majority represent a recurrence of previously treated cryptococcal meningitis. Inflammatory reactions to quiescent meningeal infection during the first few weeks of ART have also been reported. The time at onset of cryptococcosis-IRIS varies from 4 days after initiation of ART to around 3 years. Central nervous system (CNS) features of cryptococcosis-IRIS include intracranial cryptococcoma or abscesses, spinal cord abscesses, recalcitrant raised intracranial pressure, optic disc swelling, cranial nerve lesions, dysarthria, hemiparesis, and paraparesis. Extracranial manifestations of cryptococcosis-IRIS include lymphadenitis, eye disease, suppurating soft tissue lesions, and pulmonary disease that may include cavitating or nodular lesions.

At the diagnosis of cryptococcal meningitis, cerebrospinal fluid (CSF) white blood cell counts of 25 cells/μL or less and protein levels of 50 mg/dL or less are associated with the development of IRIS. On the other hand, CSF profiles at the moment of paradoxical cryptococcosis-IRIS may show an increased white blood cell count and an increased opening pressure of greater than 25 cm H_2O, but these features overlap significantly with those observed in patients with non–IRIS-related relapses of cryptococcal meningitis. The value of determining serum and CSF cryptococcal antigen titers in predicting or diagnosing cryptococcosis-IRIS is unclear, but a positive CSF cryptococcal culture prior to commencing ART is a predictor. Cultures of CSF or tissue samples obtained at the time of paradoxical cryptococcosis-IRIS may be negative even when cryptococci can be seen on microscopy.

Pneumocystis jiroveci

Patients who have been treated for *P. jiroveci* pneumonitis (PJP) may experience pulmonary inflammation after ART is commenced. It is usually characterized by fever, cough, dyspnea, chest discomfort, and patchy alveolar infiltrates on the chest radiograph (Fig. 395-3). In some patients, organizing pneumonia develops.

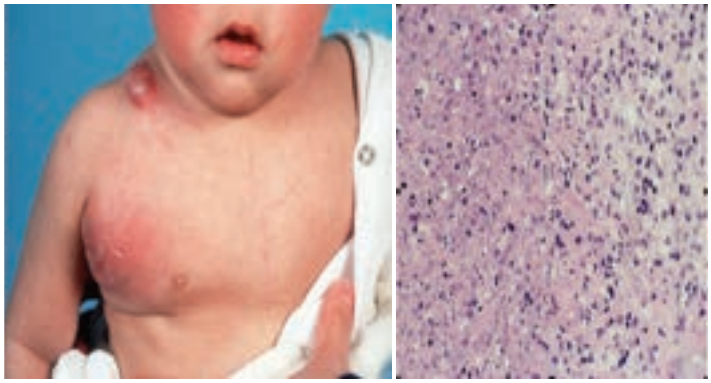

FIGURE 395-2. **Bacille Calmette-Guérin (BCG)-associated immune reconstitution inflammatory syndrome after antiretroviral therapy for HIV infection in a child who received BCG vaccination shortly after birth. A biopsy specimen from the larger lesion demonstrated necrotizing granulomatous inflammation.**

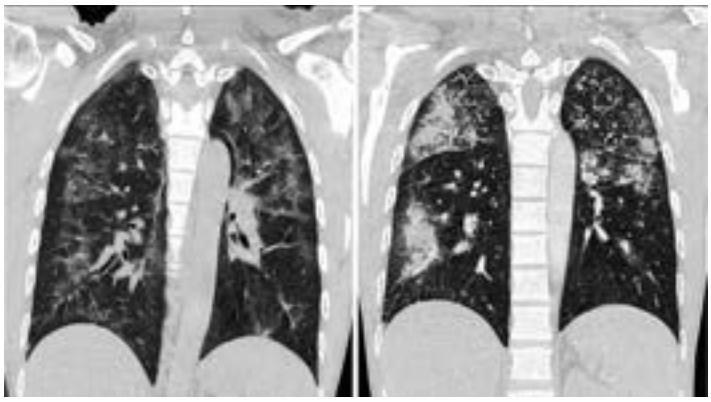

FIGURE 395-3. ***Pneumocystis jiroveci*****–associated immune reconstitution inflammatory syndrome.** ***Left*****, Before treatment of the** ***P. jiroveci*** **infection.** ***Right*****, After treatment of the** ***P. jiroveci*** **infection and commencing antiretroviral therapy.**

JC Polyomavirus

Progressive PML of the brain occurs when cellular immune responses fail to control JCV infection of oligodendrocytes and astrocytes. It is characterized by a paucity of inflammatory cells in brain lesions. ART is the only effective therapy, presumably because it enhances cellular immune responses against JCV antigens. However, ART may also result in a paradoxical worsening of established PML or in unmasking of subclinical JCV infection and appearance of PML for the first time. These manifestations of PML are often atypical in that imaging studies of the brain demonstrate changes associated with inflammation, and brain biopsy specimens demonstrate inflammatory cell infiltrates with a prominence of $CD8^+$ T cells. Between 19 and 23% of cases of PML in HIV-infected patients are due to paradoxical or unmasking IRIS. The median time at onset is 7 weeks, and most cases occur within the first 3 months of ART but very occasionally as late as 26 months after commencing ART. Predictors of PML-IRIS have not been identified.

Human Herpesvirus 8

A prospective study of Kaposi's sarcoma (KS) in patients from Mozambique commencing ART found that paradoxical KS-IRIS developed in 31% of patients with pre-ART KS, and that unmasking KS-IRIS developed in 7% of patients without pre-ART KS. Clinical manifestations included an increased number of preexisting skin lesions that sometimes exhibited increased nodularity and ulceration, new skin or mucosal lesions, and lymphedema. Independent risk factors for the development of KS-IRIS were KS before ART, human herpesvirus 8 DNA detectable in plasma, a hematocrit of less than 30%, and a plasma HIV RNA level greater than 5 $\log_{10}$ copies/mL.

Some cases of KS-IRIS will resolve without treatment, but chemotherapy is usually necessary. Doxorubicin has been used most often.

Cytomegalovirus

Eye disease is the most common manifestation of IRIS associated with cytomegalovirus (CMV) infection. Retinitis usually develops during the first few

weeks of ART as a "paradoxical" worsening of treated retinitis or as a new manifestation of CMV retinitis. Previously treated CMV infection is the most common cause of immune recovery uveitis (IRU), which presumably results from the restoration of an immune response against residual CMV antigens in the eye. The risk for development of CMV-associated IRU is greatest in patients who had a large proportion of the retina affected by CMV infection. It may develop up to 21 months after ART is commenced, and the clinical manifestations vary in severity from a transient vitreitis to persistent uveitis, papillitis, cystoid macular edema, and epiretinal membranes.

Hepatitis B and C Viruses

Hepatotoxicity manifested as elevated serum liver enzyme levels occurs in up to 18% of patients after ART is initiated. Several causes have been defined, but the most important risk factor is concomitant infection with hepatitis B virus (HBV) or hepatitis C virus (HCV). Prospective studies of patients with HIV infection who are coinfected with HBV, HCV, or both, who commenced ART demonstrated that 22 to 24% of patients with HBV coinfection, 13.5% of patients with HCV coinfection, and 50% of patients with both HBV and HCV coinfection experienced a "flare" of hepatitis. Flares of HBV hepatitis were associated with increased plasma levels of several immune mediators, thus suggesting that it was a manifestation of IRIS in the liver. Patients who experienced flares of HBV hepatitis had higher plasma HBV DNA levels and serum alanine transaminase levels before ART was commenced. Severe hepatitis after ART in patients with HIV infection and coinfection with HBV or HCV is uncommon but can occasionally result in liver decompensation and death.

Herpes Simplex Virus

Exacerbation of mucocutaneous herpes simplex virus (HSV) disease may occur after ART is initiated. Sometimes, lesions become hemorrhagic and exhibit tissue necrosis. Rarely, HSV infection of the brain may be unmasked by commencing ART and be manifested as encephalitis.

DIAGNOSIS

Immunologic tests for diagnosing IRIS are currently not available for routine use. In the absence of diagnostic tests, IRIS may be established with diagnostic criteria that take into consideration the timing, clinical characteristics, and pathology of the disease, as well as the virologic response to the ART.

TREATMENT

Rx

The general approach to the treatment of IRIS is to continue ART and provide antimicrobial therapy for the provoking infection when it is active (usually in unmasking IRIS).[4] Cessation of ART should be considered only in patients with life-threatening disease when all other measures have failed. Anti-inflammatory therapy should not be given routinely but be reserved for patients with severe inflammation, particularly when it is life-threatening. Corticosteroid therapy is used most often, but its effectiveness may vary from one type of IRIS to another. Thus, a randomized controlled trial in South Africa demonstrated that corticosteroids (prednisone 1.5 mg/kg/day for 2 weeks, then 0.75 mg/kg/day for 2 weeks) are a safe and effective treatment option for paradoxical TB-IRIS.[A4] In contrast, in an analysis of data from previously reported cases of PML-IRIS, it was suggested that corticosteroid therapy is not effective, although it was indicated that only early use of corticosteroid therapy is likely to be effective. There is anecdotal evidence suggesting that corticosteroid therapy can be effective in other types of IRIS, but there are potential risks to using corticosteroid therapy in HIV patients who are very immunodeficient, and it should only be used after weighing all considerations and should be based on evidence from clinical trials if available. Corticosteroid therapy for IRIS affecting the eye should be supervised by an ophthalmologist.

PREVENTION

Given that a low $CD4^+$ T-cell count is a major risk factor for the development of IRIS, commencing ART at a $CD4^+$ T-cell count higher than 350/μL, as recommended by many treatment guidelines, will prevent most cases. However, this is not possible in patients who are seen for the first time with an OI or in many patients from countries with limited resources. Other strategies to prevent paradoxical IRIS are therefore under investigation.

Several observations indicate that a high pathogen load is an important risk factor for IRIS, including the association with disseminated and drug-resistant TB, a shorter duration of treatment of TB or cryptococcal meningitis, and positive CSF cultures for cryptococcal infection prior to commencing ART. Therefore, delaying the introduction of ART so that the OI can be fully treated might be beneficial.[5,6] However, doing so may increase the risk for development of other OIs or cancers. The results of the AIDS Clinical Trial Group study A5164 provided evidence supporting the introduction of ART at the same time as antimicrobial therapy, particularly in patients with PJP.[A2] In addition, the results of three randomized controlled trials (SAPiT,[A3] CAMELIA,[A4] and STRIDE[A5]) demonstrated that for HIV/TB-coinfected patients with advanced immunosuppression, the survival benefit of starting ART within the first 2 weeks of TB therapy might outweigh the risk for IRIS and other adverse events. Nevertheless, a more recent clinical trial of the effect of timing of ART initiation on TB treatment outcomes for HIV-positive patients with CD4 counts of 220 cells/μL or more showed that ART can be delayed until after completion of 6 months of TB treatment in this population.[A6] In contrast, commencing ART at the same time as treatment of cryptococcal meningitis has been shown to reduce long-term survival when compared with delaying ART until the meningitis has been treated.[A7] It seems probable that IRIS affecting the CNS is more likely than other types of IRIS to result in morbidity and mortality. Therefore, a single approach to this issue may not be possible, and a strategy for commencing antimicrobial therapy and ART may have to be determined for each pathogen or for infections of the CNS.

PROGNOSIS

The prognosis for patients in whom IRIS develops is highly variable because of differences in the extent of the infection by the provoking pathogen, the characteristics of the immunopathology caused by the restored immune response, and the body site affected. Most cases of IRIS are self-limited, and outcomes are usually good. However, mortality rates of up to 66% have been reported for cryptococcosis-IRIS. The mortality rate for TB-IRIS is much lower, but hospital admissions are common. Mortality and hospitalization rates are particularly high when TB- or cryptococcosis-IRIS affects the CNS. Indeed, involvement of the CNS by any type of IRIS may result in death or permanent neurologic disability.[7-9] For example, mortality rates of 53% have been reported for paradoxical PML-IRIS and 31% for unmasking PML-IRIS. Furthermore, patients who survive PML-IRIS may have neurologic sequelae such as hemiparesis or seizures. Patients with lymphadenitis resulting from TB-IRIS and those with meningitis resulting from cryptococcosis-IRIS occasionally undergo recurrent relapses and can become steroid dependent.

Autoimmune Disease and Sarcoidosis

Patients with HIV infection who are receiving ART have an increased susceptibility to some autoimmune diseases, mainly Graves disease, and sarcoidosis. Although sometimes referred to as types of IRIS, they appear to have a different immunopathogenesis.

Grade A References

A1. Meintjes G, Wilkinson RJ, Morroni C, et al. Randomized placebo-controlled trial of prednisone for paradoxical tuberculosis-associated immune reconstitution inflammatory syndrome. *AIDS*. 2010; 24:2381-2390.

A2. Zolopa A, Andersen J, Powderly W, et al. Early antiretroviral therapy reduces AIDS progression/death in individuals with acute opportunistic infections: a multicenter randomized strategy trial. *PLoS ONE*. 2009;4:e5575.

A3. Abdool Karim SS, Naidoo K, Grobler A, et al. Timing of initiation of antiretroviral drugs during tuberculosis therapy. *N Engl J Med*. 2010;362:697-706.

A4. Blanc FX, Sok T, Laureillard D, et al; CAMELIA (ANRS 1295–CIPRA KH001) Study Team. Earlier versus later start of antiretroviral therapy in HIV-infected adults with tuberculosis. *N Engl J Med*. 2011;365:1471-1481.

A5. Havlir DV, Kendall MA, Ive P, et al; AIDS Clinical Trials Group Study A5221. Timing of antiretroviral therapy for HIV-1 infection and tuberculosis. *N Engl J Med*. 2011;365:1482-1491.

A6. Mfinanga SG, Kirenga BJ, Chanda DM, et al. Early versus delayed initiation of highly active antiretroviral therapy for HIV-positive adults with newly diagnosed pulmonary tuberculosis (TB-HAART): a prospective, international, randomized, placebo-controlled trial. *Lancet Infect Dis*. 2014;14: 563-571.

A7. Boulware DR, Meya DB, Muzoora C, et al. Timing of antiretroviral therapy after diagnosis of cryptococcal meningitis. *N Engl J Med*. 2014;370:2487-2498.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1 Chang CC, Sheikh V, Sereti I, et al. Immune reconstitution disorders in patients with HIV infection: from pathogenesis to prevention and treatment. *Curr HIV/AIDS Rep*. 2014;11:223-232.

2. French MA. Immune reconstitution inflammatory syndrome: a reappraisal. *Clin Infect Dis*. 2009; 48:101-107.
3. Bonnet M, Baudin E, Jani IV, et al. Incidence of paradoxical tuberculosis-associated immune reconstitution inflammatory syndrome and impact on patient outcome. *PLoS ONE*. 2013;8:e84585.
4. Meintjes G, Scriven J, Marais S. Management of the immune reconstitution inflammatory syndrome. *Curr HIV/AIDS Rep*. 2012;9:238-250.
5. Müller M, Wandel S, Colebunders R, et al. Immune reconstitution inflammatory syndrome in patients starting antiretroviral therapy for HIV infection: a systematic review and meta-analysis. *Lancet Infect Dis*. 2010;10:251-261.
6. Sereti I, Rodger AJ, French MA. Biomarkers in immune reconstitution inflammatory syndrome: signals from pathogenesis. *Curr Opin HIV AIDS*. 2010;5:504-510.
7. Post MJ, Thurnher MM, Clifford DB, et al. CNS-immune reconstitution inflammatory syndrome in the setting of HIV infection, part 1: overview and discussion of progressive multifocal leukoencephalopathy-immune reconstitution inflammatory syndrome and cryptococcal-immune reconstitution inflammatory syndrome. *AJNR Am J Neuroradiol*. 2013;34:1297-1307.
8. Post MJ, Thurnher MM, Clifford DB, et al. CNS-immune reconstitution inflammatory syndrome in the setting of HIV infection, part 2: discussion of neuroimmune reconstitution inflammatory syndrome with and without other pathogens. *AJNR Am J Neuroradiol*. 2013;34:1308-1318.
9. Bahr N, Boulware DR, Marais S, et al. Central nervous system immune reconstitution inflammatory syndrome. *Curr Infect Dis Rep*. 2013;15:583-593.

REVIEW QUESTIONS

1. A major risk factor for developing paradoxical tuberculosis immune reconstitution inflammatory syndrome (TB-IRIS) is:

A. Starting antiretroviral therapy (ART) in a patient with a low $CD4^+$ T-cell count
B. Starting ART in a patient with a high $CD4^+$ T-cell count
C. Starting ART in a patient with pulmonary TB
D. Starting a protease inhibitor–containing regimen
E. Starting ART in a patient with rifampin-resistant TB

Answer: A TB-IRIS has been associated with a low $CD4^+$ T-cell count (usually < 50/μL) in several studies, probably because it is a marker of a high pathogen load and an increased susceptibility to restoration of a pathogen-specific immune response that causes immunopathology. All ART regimens may cause TB-IRIS, and TB-IRIS may develop in patients with all types of TB, including patients with multidrug-resistant TB.

2. Which of the the following statements is correct?

A. In an HIV-infected patient with cryptococcal meningitis, ART should be started within 2 weeks after the start of the cryptococcal treatment.
B. In an HIV-infected patient with cryptococcal meningitis, ART should be started at least 4 weeks after the start of the cryptococcal treatment.
C. In an HIV-infected patient with cryptococcal meningitis and a $CD4^+$ T-cell count below 50 cells/μL, ART should be started within 2 weeks after the start of the cryptococcal treatment.
D. In an HIV-infected patient with cryptococcal meningitis and very few leucocytes in the cerebrospinal fluid (CSF), ART can be started together with the start of the cryptococcal treatment.

Answer: B Commencing ART within the first 2 weeks after starting cryptococcal meningitis treatment is associated with increased mortality compared with starting ART at least 4 weeks after the start of the cryptococcal treatment. The presence of very few leukocytes in the CSF is a further risk factor for developing cryptococcal IRIS.

3. Which of the the following statements is correct?

A. All patients with IRIS should be treated with prednisone.
B. Patients with IRIS should never be treated with prednisone.
C. In patients with IRIS, ART should be stopped.
D. In patients with TB-IRIS, prednisone may be beneficial.

Answer: D The effectiveness of prednisone as treatment for IRIS varies from one type of IRIS to another. A randomized controlled trial in South Africa demonstrated that prednisolone therapy is a safe and effective treatment option for paradoxical TB-IRIS. For progressive multifocal leukoencephalopathy (PML)-IRIS, only early use of corticosteroid therapy may be effective. For Kaposi's sarcoma (KS)-IRIS, the benefits of corticosteroids are not established. Potential risks of using corticosteroid therapy in patients with HIV infection who are very immunodeficient should always be considered.

4. IRIS can be prevented by which one of the following?

A. Early ART at a high $CD4^+$ T-cell count
B. Prophylaxis for opportunistic infections
C. Early diagnosis and treatment of opportunistic infections
D. All of the above

Answer: D Commencing ART at a high $CD4^+$ T-cell count, as well as prophylaxis for opportunistic infections, will prevent the occurrence of opportunistic infections and therefore IRIS. Earlier diagnosis and treatment of opportunistic infections will reduce the pathogen load and therefore the risk of developing IRIS.

396

APPROACH TO THE PATIENT WITH NEUROLOGIC DISEASE

ROBERT C. GRIGGS, RALPH F. JÓZEFOWICZ, AND MICHAEL J. AMINOFF

CLINICAL MANIFESTATIONS

Many symptoms of nervous system diseases are a part of everyday experience for most normal people. Slips of the tongue, headaches, backache and other pains, dizziness, lightheadedness, numbness, muscle twitches, jerks, cramps, and tremors all occur in totally healthy persons. Mood swings with feelings of elation and depression, paranoia, and displays of temper are equally a part of the behavior of completely normal people. The rapid increase in information about neurologic diseases, coupled with the intense interest of people in all walks of life in medical matters, has focused public attention on both common and rare neurologic conditions.

Most older people are concerned that they or their spouse have or are developing Alzheimer disease or stroke. The almost ubiquitous tremor of the elderly prompts concern about Parkinson disease. Many younger patients are concerned about multiple sclerosis or brain tumor, and few normal people lack one or more symptoms suggesting the diagnosis of a serious neurologic disease. For most of these and other common diagnoses, the results of imaging and other tests are typically normal when symptoms first appear, and such tests should not be performed to reassure the patient or physician. Moreover, the widespread availability of neurodiagnostic imaging and electrophysiologic, biochemical, and genetic testing has led to the detection of "abnormalities" in many young and most elderly persons. In evaluating a patient's symptoms, it is imperative that a clinical diagnosis be reached without reference to a neurodiagnostic laboratory finding. Patients with disorders such as headache, anxiety, and depression do not usually have abnormal laboratory results. Abnormalities noted on various neurodiagnostic studies are often incidental findings whose treatment may be justified and necessary, but such treatment will not improve the patient's symptoms. Abnormalities detected incidentally that are not accompanied by signs or symptoms may, as for disorders such as hypertension, require aggressive evaluation and treatment, but in general, the adage that it is difficult to improve an asymptomatic patient should be kept in mind. Thus, in elderly patients, few imaging or electrophysiologic studies are interpreted as "normal," but in the absence of specific complaints consistent with the findings, treatment and even further evaluation should reflect an estimate of the specificity and sensitivity of the test as well as the likelihood that the patient will require and benefit from treatment. It is a good rule of thumb that one should never perform (or refer to the result of) a neurodiagnostic procedure without a specific diagnosis or at least a differential diagnosis in mind.

It is important to allow patients to describe any symptoms in their own words. Direct questions are often necessary to fully characterize the problem, but suggested terms or descriptors for symptoms are frequently grasped by a patient unfamiliar with medical terminology and then parroted to subsequent interviewers. The patient's terms should always be used in recording symptoms. Terms such as *lameness*, *weakness*, *numbness*, *heaviness*, *cramps*, and *tiredness* may each mean pain, weakness, or alteration of sensation to some patients.

DIAGNOSIS

History

In neurologic diagnosis, the history usually indicates the nature of the disease or the diagnosis, whereas the neurologic examination localizes it and quantitates its severity. For many diseases, the history is almost the only avenue to explore. Examples of such disorders include headaches, seizures, developmental disorders, memory disorders, and behavioral diseases. In arriving at a diagnosis, the following points are useful. Consider the entire medical history of the patient. Early life events or long-standing processes such as head or spine trauma, unilateral hearing or visual loss, poor prowess in sports, poor performance in school, spinal curvature, and bone anomalies are easily overlooked but may point to the underlying disease process.

Consider the tempo and duration of the symptoms. Have the symptoms been progressive without remission, or have there been plateaus or periods of return to normal? Cerebral mass lesions (tumor, subdural) tend to have a progressive but fluctuating course; seizures and migraine, an episodic course; and strokes, an abrupt ictal onset with worsening for 3 to 5 days, followed by partial or complete recovery.

Can one disease account for all of the symptoms and signs? The clinician should formulate a diagnostic opinion in anatomic terms. Is the history suggestive of a single (e.g., stroke or tumor) focus or multiple sites of nervous system involvement (e.g., multiple sclerosis), or is the process a disease of a system (vitamin B_{12} deficiency, myopathy, or polyneuropathy)?

The neurologic history is the most important component of neurologic diagnosis. A careful history frequently determines the cause and allows one to begin localizing the lesions, which aids in establishing whether the disease is diffuse or focal. Symptoms of acute onset suggest a vascular cause or seizure; symptoms that are subacute in onset suggest a mass lesion such as a tumor or abscess; symptoms that have a waxing and waning course with exacerbations and remissions suggest a demyelinating cause; and symptoms that are chronic and progressive suggest a degenerative disorder.

The history is often the only way of diagnosing neurologic illnesses that typically have normal or nonfocal findings on neurologic examination. These illnesses include many seizure disorders, narcolepsy, migraine and most other headache syndromes, the various causes of dizziness, and most types of dementia. The neurologic history may often provide the first clues that a symptom is psychological in origin. The following are points to consider in obtaining a neurologic history.

- *Carefully identify the chief complaint or problem.* Not only is the chief complaint important in providing the first clue to the physician about the differential diagnosis, but it is also the reason the patient is seeking medical advice and treatment. If the chief complaint is not properly identified and addressed, the proper diagnosis may be missed and an inappropriate diagnostic work-up may be undertaken. Establishing a diagnosis that does not incorporate the chief complaint frequently focuses attention on a coincidental process irrelevant to the patient's concerns.
- *Listen carefully to the patient for as long as necessary.* A good rule of thumb is to listen initially for at least 5 minutes without interrupting the patient. The patient often volunteers the most important information at the start of the history. During this time, the examiner can also assess mental status, including speech, language, fund of knowledge, and affect, and observe the patient for facial asymmetry, abnormalities in ocular movements, and an increase or a paucity of spontaneous movements as seen with movement disorders.
- *Steer the patient away from discussions of previous diagnostic test results and the opinions of previous caregivers.* Abnormal results of laboratory studies may be incidental to the patient's primary problem or may simply represent a normal variant.
- *Take a careful medical history, medication history, psychiatric history, family history, and social and occupational history.* Many neurologic illnesses are complications of underlying medical disorders or are due to adverse effects of drugs. For example, parkinsonism is a frequent complication of the use of metoclopramide and most neuroleptic agents. A large number of neurologic disorders are hereditary, and a positive family history may establish the diagnosis in many instances. Occupation plays a major role in various neurologic disorders such as carpal tunnel syndrome (in machine operators and people who use computer keyboards) and peripheral neuropathy (caused by exposure to lead or other toxins).
- *Interview surrogate historians.* Because patients with dementia or altered mental status are generally unable to provide exact details of the history, a family member may need to provide the key details required to make an accurate diagnosis. This situation is especially common with patients who have dementia and certain right hemispheric lesions with various agnosias (lack of awareness of disease) that may interfere with their ability to provide a cogent history. Surrogate historians also provide missing historical details for patients with episodic loss of consciousness, such as syncope and epilepsy.
- *Summarize the history for the patient.* Summarizing the history is an effective way to ensure that all details were covered sufficiently for a tentative diagnosis to be made. Summarizing also allows the physician to fill in historical gaps that may not have been apparent when the history was initially taken. In addition, the patient or surrogate may correct any historical misinformation at this time.

- *End by asking what the patient thinks is wrong.* This question allows the physician to evaluate the patient's concerns about and insight into the condition. Some patients have a specific diagnosis in mind that spurs them to seek medical attention. Multiple sclerosis, amyotrophic lateral sclerosis, Alzheimer disease, and brain tumors are diseases that patients often suspect may be the cause of their neurologic symptoms. This discussion will also help guide how to discuss prognosis, especially in patients with advanced neurologic disease.[1]

Diagnostic Challenges

Two common situations provide special challenges to the diagnostic skills of the physician.

Physical Abuse as a Cause of Neurologic Symptoms

Traumatic injury inflicted by family members or others is usually difficult to detect by the medical history and examination. Physically battered babies, abused children, battered women, and traumatized seniors are often unable or unwilling to complain of this cause or contribution to symptoms. The only method to prevent overlooking this frequent cause of common problems is systematic consideration of the possibility in every patient and awareness of the often subtle signs that suggest physical trauma: ecchymoses or fractures (often attributed to a logical cause), denial of expected symptoms, failure to keep appointments, and unexplained intensification of neurologic symptoms (headache, dizziness, ringing in the ears, blackouts).

Alcoholism and Drug Abuse

See Chapters 33 and 34. A host of neurologic disorders can be the result of intentional ingestion of toxins (Chapter 110). Patients do not give an accurate account of their use of these agents. Consequently, physical signs and laboratory screening test results that give evidence of drug-related hepatic and other metabolic abnormalities may point to a major underlying problem.

ACUTE NEUROLOGIC DISORDERS REQUIRING IMMEDIATE DIAGNOSIS AND TREATMENT

Most neurologic diagnoses are arrived at by a careful, thorough history and an appropriately complete examination. However, the tempo of illness and the availability of life-saving treatment that is effective only if it is administered within minutes of first evaluating a patient dictate rapid action in several specific circumstances.[2] Coma (Chapter 404), repetitive seizures (Chapter 403), acute stroke (Chapters 407 and 408), suspected meningitis and encephalitis (Chapters 412 and 414), head and spine trauma (Chapter 399), and acute spinal cord compression are diagnosed by clinical and laboratory assessment, and urgent treatment must be instituted as soon as ventilation and cardiac status are stabilized.

NEUROLOGIC EXAMINATION

The neurologic examination is always tailored to the clinical setting of the patient. The approach to an ambulatory office patient is very different than the approach to a critically ill patient.[3] A complete neurologic examination of a child is much different from that of an elderly adult, and the examination of a patient with specific complaints focuses on findings pertinent to that patient. Thus, more detailed testing of cognition is indicated in patients with behavioral or memory disturbance, and more detailed testing of sensation should be performed in patients with complaints of pain, numbness, or weakness.

However, many tests of neurologic function are routinely indicated in all patients because they provide a baseline for future examination and are frequently helpful in detecting unsuspected neurologic disease in apparently normal persons or in patients whose symptoms initially suggest disease outside the nervous system. It is particularly important to perform all routine tests in patients with abnormalities in one sphere of neurologic dysfunction; otherwise, erroneous localization of a lesion or disease process is likely. For deviations from normal to be recognized and quantitated, it is essential for a physician to have extensive experience in the routine assessment of normal persons.

The General Examination

Specific neurologic symptoms or signs should prompt attention to the assessment of general findings. Head circumference should be measured in patients with central nervous system (CNS) or spinal cord disease (normally 55 ± 5 cm in adults). Head enlargement is occasionally a normal, often hereditary variant but should suggest a long-standing anomaly of the brain or spinal cord. The skin should be inspected for café au lait maculae, adenoma sebaceum, vascular malformations, lipomas, neurofibromas, and other lesions (Chapter 417). Neck range of motion, straight leg raising, and spinal curvature (scoliosis) should be assessed. Carotid auscultation for bruits is indicated in all older adults; carotid palpation is seldom informative. In patients with bladder, bowel, or leg symptoms, a rectal sphincter examination for tone and ability to contract voluntarily is usually indicated. Limitation of joint range of motion or painless swelling of joints is often a sign of an unsuspected neurologic lesion.

Neurologic Examination

The various aspects of the detailed neurologic examination are considered in specific symptom and disease sections noted later. The five major divisions of the examination should be assessed in all patients. During a careful medical history, mental status is often adequately assessed: level of consciousness, orientation, memory, language function, affect, and judgment. If any of these functions are abnormal, more detailed testing is needed. Cranial nerve function that should be tested in all patients includes visual acuity (with and without correction); optic fundi; visual fields; pupils (size and reactivity to direct and consensual light); ocular motility; jaw, facial, palatal, neck, and tongue movement; and hearing.

Examination of the motor system (Chapter 421) is essential in all patients because incipient weakness is generally overlooked by the patient. Muscle tone (flaccid, spastic, or rigid), muscle size (atrophy or hypertrophy), and muscle strength can be assessed rapidly. Muscle strength testing should always assess specific functional activities, including the ability to walk on heel and toe, to sit up from a supine position, to rise from a deep knee bend or deep chair, to lift the arms over the head, and to make a tight fist. Gait, stance, and coordination are assessed. The patient should be observed for tremor and other abnormal movements and the muscles inspected for fasciculations.

Sensory testing (Chapter 420) need not be detailed unless there are sensory symptoms. However, vibration perception in the toes and the normality of perception of pain, temperature, and light touch in the hands and feet should be assessed.

Muscle stretch reflexes and plantar responses should always be assessed by evaluating right-left symmetry and disparity between proximal and distal reflexes or arm and leg reflexes. Biceps, triceps, brachioradialis, quadriceps, and ankle reflexes should be quantitated from 1 to 4 (4 = clonus; 3 = spread; 2 = brisk; 1 = hypoactive).

The Comatose Patient

The rapid examination required for a patient with an altered state of consciousness is much different from that of an alert, aware individual (Chapter 404). Many aspects of the neurologic examination cannot be tested: cognitive function, subtleties of sensory perception, specific motor functions, coordination, gait, and stance. Moreover, the muscle stretch reflexes are likely to fluctuate from one moment to the next, and minor asymmetries are much less important than in an awake patient. Instead, attention should focus on examination of the level of consciousness, respiratory pattern, eyelid position and eye movements, pupils, corneal reflexes, optic fundi, and motor responses. Particular elements of the general examination must also be assessed quickly: evidence of cranial and spinal trauma, tenderness of the skull to percussion, nuchal rigidity (but not in patients with head or neck trauma), and evidence of physical abuse.

COMMON COMPLAINTS OF POSSIBLE NEUROLOGIC ORIGIN

Weakness

It is axiomatic that patients typically have motor signs before motor symptoms and, conversely, sensory symptoms before sensory signs. Thus, patients with even severe weakness may not report symptoms of weakness. Somewhat paradoxically, patients who complain of "weakness" often do not have confirmatory findings on examination that document the presence of weakness.

Weakness, when it is actually a symptom of neurologic disease, is frequently caused by diseases of the motor unit (Chapters 419, 421, and 422) and is usually reported by a patient in terms of loss of specific functions—for example, difficulty with tasks such as climbing stairs, rising from a chair, sitting up, lifting objects onto a high shelf, or opening jars. Symptoms may also reflect the consequences of weakness, such as frequent falls or tripping. Such symptoms can be remarkably quantitative. A patient with leg muscle

TABLE 396-1 DISORDERS COMMONLY ACCOMPANIED BY WEAKNESS

Disorders of the motor unit
Upper motor neuron lesions—spasticity
Basal ganglia disorders—rigidity
General medical conditions
Heart failure
Respiratory insufficiency
Renal, hepatic, and other metabolic disease
Alcoholism and other toxin-related disease
Psychiatric and behavioral disorders
Depression
Malingering

weakness who is falling even as infrequently as once a month almost invariably has severe weakness of the knee extensor muscles and can be shown on examination to have a knee extension lag, an inability to lift the leg fully against gravity and lock the knee.

The symptom of weakness without findings of weakness on examination is not generally the result of neuromuscular disease but can be a sign of neurologic disease outside the motor unit or, more commonly, a symptom of disease outside the nervous system altogether (Table 396-1).

Episodic and Intermittent Weakness

The complaint of attacks of severe weakness or paralysis occurring in a patient with baseline normal strength is an uncommon symptom. It is typical of the periodic paralyses and may also be seen with episodic ataxias and myotonic disorders (Chapter 421). All these disorders are ion channelopathies. These channelopathies (e.g., the calcium channelopathy hypokalemic periodic paralysis) are rare but treatable disorders (Chapter 421). Episodic weakness is also seen in patients with neuromuscular junction disorders such as myasthenia gravis and the myasthenic syndrome (Chapter 422). On occasion, patients with narcolepsy complain of intermittent paralysis as a reflection of sleep paralysis (Chapter 405).

Fatigue

Complaints of fatigue, tiredness, and lack of energy are even less likely than the symptom of weakness to reflect definable neurologic disease. With the exception of neuromuscular junction disorders such as myasthenia gravis, fatigue is rarely a complaint of diseases of the motor unit. Fatigue can be a sign of upper motor neuron disease (corticospinal pathways) and is a common complaint of established multiple sclerosis and other multifocal CNS disease. Similarly, any process that produces bilateral corticospinal tract or extrapyramidal disease can cause fatigue. Examples include motor neuron disease (Chapter 419), spinal cord disease in the cervical cord region (Chapter 400), and Parkinson disease (Chapter 409). In addition, disorders that impair sleep (Chapter 405) may include fatigue as a complaint.

Fatigue, like weakness, is much more often than not a sign of disease outside the central and peripheral nervous system. Depression and other psychiatric and behavioral disorders (Chapter 397) as well as the medical illnesses associated with a complaint of weakness are all frequent causes of fatigue.

Chronic fatigue syndrome and many cases of fibromyalgia (Chapter 274) have fatigue as a dominant disabling symptom. These disorders are defined in part by the absence of consistent neurologic findings and lack of demonstrable disease in the nervous system.

Spontaneous Movements

Muscle tremors, jerks, twitches, cramps, and spasms (Chapter 410) are frequent symptoms. The cause of spontaneous movements can reside at any level of the nervous system. In general, movements that occur in an entire limb or in more than one muscle group concurrently are caused by CNS disease. Movements confined to a single muscle are likely to be a reflection of disease of the motor unit (including the motor neurons of the brain stem and spinal cord). When spontaneous movements of a muscle are associated with severe pain, patients often use the term *cramp.* Cramp is a medically defined disorder that reflects the intense contraction of a large group of motor units. Leg cramps are occasionally a sign of an underlying disease of the anterior horn cell, nerve roots, or peripheral nerve; however, cramps are frequent in normal persons and particularly common in older patients, and they are usually benign. When they are severe, cramps can produce such intense muscle contraction that muscle injury is caused and muscle enzyme (e.g., creatine kinase) levels are elevated in blood.

The rare muscle diseases in which an enzyme deficiency interferes with substrate use as fuel for exercise (e.g., McArdle disease) are often associated with severe exercise-provoked muscle contractures. These contractures are electrically silent on electromyography, in contrast to the intense motor unit activity seen with cramps. Such contractures should not be confused with the limitation of joint range of motion resulting from long-standing joint disease or long-standing weakness, also termed contractures.

The intense muscle contractions of tetany are frequently painful. Although tetany is usually a reflection of hypocalcemia (Chapter 245), it can occasionally be seen without demonstrable electrolyte disturbance. Tetany results from hyperexcitability of peripheral nerves. Similarly, in the syndrome of tetanus produced by a clostridial toxin (Chapter 296), intensely painful life-threatening muscle contractions arise from hyperexcitable peripheral nerves. A number of toxic disorders, such as strychnine poisoning and black widow spider envenomation, produce similar neurogenic spasms.

Muscle Pain

Acute muscle pain in the absence of abnormal muscle contractions is an extremely common symptom. When such pain occurs after strenuous exercise or in the context of an acute viral illness (e.g., influenza), it probably reflects muscle injury. In such patients, the serum creatine kinase level is often raised. It is uncommon for this frequent and essentially normal sign of muscle injury to be associated with weakness or demonstrable ongoing muscle disease. Chronic muscle pain is a common symptom but is seldom related to a definable disease of muscle.

Loss of Balance

Unsteadiness of gait is a common symptom. When it is associated with complaints of dizziness or vertigo (Chapter 428), disease of the labyrinth, vestibular nerve, brain stem, or cerebellum is a probable cause. When unsteadiness and loss of balance are unassociated with dizziness, particularly if the unsteadiness appears to be out of proportion to other symptoms of the patient, a widespread disorder of sensation or motor function is likely.

Abnormal Gait and Posture

The ability to stand and walk in a well-coordinated, effortless fashion requires integrity of the entire nervous system.[4] Relatively subtle deficits localized to one part of the central or peripheral nervous system produce characteristic abnormalities (Table 396-2).

Sensory Symptoms

Sensory symptoms can be negative or positive. Negative symptoms represent a loss of sensation, such as a feeling of numbness. Positive symptoms, by contrast, consist of sensory phenomena that occur without normal stimulation of receptors and include paresthesias and dysesthesias. Paresthesias may include a feeling of tingling, crawling, itching, compression, tightness, cold, or heat and are sometimes associated with a feeling of heaviness. The term *dysesthesias* is used correctly to refer to abnormal sensations—often tingling, painful, or uncomfortable—that occur after innocuous stimuli, whereas *allodynia* refers to painful perception from a stimulus that is not normally painful. For some patients, it may be difficult to distinguish paresthesias and dysesthesias from pain. *Hypesthesia* denotes a loss or impairment of touch, whereas *hypalgesia* denotes a loss of pain sensibility. By comparison, *hyperesthesia* and *hyperalgesia* indicate a lowered threshold to tactile or painful stimuli such that there is increased sensitivity to such stimuli.[5]

With the use of a wisp of cotton, a single-use pin, and a tuning fork, the trunk and extremities are examined for regions of abnormal or absent sensation. Special instruments are available for quantifying sensory function, such as the computer-assisted sensory examination, which is based on the detection of touch, pressure, vibratory, and thermal sensation thresholds.

Alterations in pain and tactile sensibility can generally be detected by clinical examination. It is important to localize the distribution of any such sensory loss to distinguish between nerve, root, and central dysfunction. Similarly, abnormalities in proprioception can be detected by clinical examination when patients are unable to detect the direction in which a joint is

TABLE 396-2 CHARACTERISTIC GAIT DISORDERS

SPECIFIC DISORDER	LOCATION OF LESION	CHARACTERISTICS
Spastic gait	Bilateral corticospinal pathways within the thoracic or cervical cord or in the brain	Legs stiff, feet turning inward, "scissoring"
Hemiparetic gait	Unilateral central nervous system, cervical cord, or brain	Affected leg circumducted, foot extended, arm flexed
Sensory ataxia	Posterior columns of the spinal cord or peripheral nerve	Wide-based, high steps; Romberg sign present
Cerebellar ataxia	Brain stem or cerebellum	Wide-based steps; Romberg sign absent
Parkinsonian gait	Basal ganglia	Shuffling, small steps
Dystonic gait	Basal ganglia; also corticospinal pathways	Abnormal posture of the arms, head, neck
Gait disorder of the elderly	Multifactorial: bihemispheric disease, spinal cord disease, impaired proprioception, muscle weakness	Stooped posture, wide-based steps; often retropulsion
Steppage gait	Distal muscle weakness	High steps ("steppage")
Waddling gait	Proximal muscle weakness	Both legs circumducted to allow locking of the knees
Antalgic gait	Non-neurologic; reflects disease of joints, bones, or soft tissue	Minimizes pain in the hip, spine, leg
Hysterical gait	Psychiatric or behavioral disorder	Reeling side to side, associated astasia-abasia, bizarre arm and trunk movements

moved. With severe loss of proprioception, patients may develop pseudoathetoid movements of the outstretched hands, sensory ataxia, or postural and action tremors.

Disorders of peripheral nerves commonly lead to sensory disturbances that depend on the population of affected nerve fibers (Chapter 420). Some neuropathies are predominantly large-fiber neuropathies. Appreciation of movement and position is impaired, and paresthesias are common. Examination reveals that vibration and position sensations are impaired, and movement becomes clumsy and ataxic. Pain and temperature appreciation is relatively preserved. The tendon reflexes are lost early. In other neuropathies, small fibers are selectively affected: spontaneous pain is common and may be burning, lancinating, or aching in quality. Pain and temperature appreciation is disproportionately affected in these neuropathies, and autonomic dysfunction may be present. Examples of small-fiber neuropathies include diabetes (Chapter 229) and alcoholism (Chapter 33). Most sensory neuropathies are characterized by a distal distribution of sensory loss, whereas sensory neuronopathies are characterized by sensory loss that may also involve the trunk and face and tends to be particularly severe. Sensory changes in a radiculopathy conform to a root territory; in cauda equina syndromes, sensory deficits involve multiple roots and may lead to saddle anesthesia and loss of the normal sensation associated with the passage of urine or feces.

Lesions of the posterolateral columns of the cord, such as occur in multiple sclerosis (Chapter 411), vitamin B_{12} deficiency (Chapter 416), and cervical spondylosis (Chapter 400), often lead to a feeling of compression in the affected region and to a Lhermitte sign (paresthesias radiating down the back and legs on neck flexion). Examination reveals ipsilateral impairment of vibration and joint position senses, with preservation of pain and temperature appreciation. Conversely, lesions of the anterolateral region of the cord (as by cordotomy) or central lesions interrupting fibers crossing to join the spinothalamic pathways (as in syringomyelia; Chapter 417) lead to impairment in pain and temperature appreciation with relative preservation of vibration, joint position sense, and light touch. Motor deficits may also be present and help localize the lesion. Upper motor neuron dysfunction

FIGURE 396-1. **Cerebrospinal fluid (CSF) examination. A,** Normal crystal-clear CSF. **B,** Blood in the CSF, which could result from a traumatic (bloody) tap or from subarachnoid hemorrhage. In a traumatic tap, subsequent tubes of CSF are usually less bloody. **C,** Centrifuged CSF in a traumatic tap. The supernatant is nearly clear. **D,** CSF from a patient with subarachnoid hemorrhage. There is blood at the bottom of the tube and the supernatant is yellow (xanthochromic) as a result of breakdown of blood cells in the CSF before the lumbar puncture. (From Forbes CD, Jackson WD. *Color Atlas and Text of Clinical Medicine.* 3rd ed. London: Mosby; 2003, with permission.)

(Chapter 400) from cervical lesions leads to quadriplegia, whereas more caudal lesions lead to paraplegia; lesions below the level of the first lumbar vertebra may simply compress the cauda equina and result in lower motor neuron deficits from a polyradiculopathy as well as impairment of sphincter and sexual function.

NEUROLOGIC DIAGNOSTIC PROCEDURES

Lumbar Puncture

Sampling of cerebrospinal fluid (CSF) by lumbar puncture is crucial for accurate diagnosis of meningeal infections and carcinomatosis (Fig. 396-1). Ultrasound imaging can reduce the risk of an unsuccessful or traumatic lumbar puncture. CSF analysis is also helpful in evaluating patients with central or peripheral nervous system demyelinating disorders and with intracranial hemorrhage, particularly when imaging studies are inconclusive.

The CSF formula often provides an important clue to the pathologic process involved (Table 396-3). An elevated white blood cell count is seen with infections and other inflammatory diseases as well as with carcinomatosis. The differential white blood cell count may point to a specific class of pathogen; polymorphonuclear leukocytes suggest a bacterial process, whereas mononuclear cells suggest a viral, fungal, or immunologic cause. The CSF glucose concentration is typically reduced in bacterial and fungal infections as well as with certain viral infections (e.g., mumps virus) and sarcoidosis. The CSF protein concentration is elevated in a variety of disorders, including most infections and demyelinating neuropathies.

Specialized tests that can be performed on CSF include oligoclonal bands, a pathologic pattern of bands on CSF electrophoresis that is seen in up to 90% of patients with multiple sclerosis. The bands, which represent monoclonal immunoglobulins that are locally synthesized in the CNS, are not specific for multiple sclerosis and may be seen with other inflammatory and noninflammatory conditions, including systemic lupus erythematosus, human immunodeficiency virus infection, and stroke.

CSF polymerase chain reaction is a rapid, sensitive, and specific test for the diagnosis of herpes simplex encephalitis (Chapter 414), for which it has replaced brain biopsy as the diagnostic procedure of choice. The CSF VDRL (Venereal Disease Research Laboratory) assay is a specific although insensitive test for neurosyphilis (Chapter 319).

A lumbar puncture should not be performed in patients who have an obstructive noncommunicating hydrocephalus or a focal CNS mass lesion causing raised intracranial pressure, because reducing CSF pressure acutely in these settings by lumbar puncture may result in cerebral or cerebellar herniation. Lumbar puncture may be safely performed in patients with a communicating hydrocephalus, such as with idiopathic intracranial hypertension (pseudotumor cerebri), and it may even be an effective treatment in selected patients with this condition.

Electroencephalography

Electroencephalography is the recording and measurement of scalp electrical potentials to evaluate baseline brain functioning and paroxysmal brain electrical activity suggestive of a seizure disorder.

Electroencephalography is performed by securing 20 electrodes to the scalp at predetermined locations based on an international system that uses

TABLE 396-3 CHARACTERISTIC CEREBROSPINAL FLUID FORMULAS

	TURBIDITY AND COLOR	OPENING PRESSURE	WBC COUNT	DIFFERENTIAL CELLS	RBC COUNT	PROTEIN	GLUCOSE
Normal	Clear, colorless	70-180 mm H_2O	0-5 cells/μL	Mononuclear	0	<60 mg/dL	>⅔ serum
Bacterial meningitis	Cloudy, straw colored	↑	↑↑	PMNs	0	↑↑	↓
Viral meningitis	Clear or cloudy, colorless	↑	↑	Lymphocytes	0	↑	Normal
Fungal and tuberculous meningitis	Cloudy, straw colored	↑	↑	Lymphocytes	0	↑↑	↓↓
Viral encephalitis	Clear or cloudy, straw colored	Normal to ↑	↑	Lymphocytes	0 (herpes ↑)	Normal to ↑	Normal
Subarachnoid hemorrhage	Cloudy, pink	↑	↑	PMNs and lymphocytes	↑↑	↑	Normal (early); ↓ (late)
Guillain-Barré syndrome	Clear, yellow	Normal to ↑	0-5 cells/μL	Mononuclear	0	↑	Normal

PMN = polymorphonuclear leukocyte; RBC = red blood cell; WBC = white blood cell.

standardized percentages of the head circumference, the "10-20 system." Each electrode is labeled with a letter and a number, the letter identifying the skull region (Fp = frontopolar; F = frontal; P = parietal; C = central; T = temporal; O = occipital) and the number identifying the specific location, with odd numbers representing left-sided electrodes, even numbers representing right-sided electrodes, and zero representing midline placements. These electrodes are then connected in various combinations of pairs to generate voltage potential differences, and the potentials are displayed on a computer screen.

To delineate the spatial distribution of the changing electrical field for an electroencephalogram (EEG), an orderly arrangement of electrode pairs is used, and each specific arrangement is known as a montage. Montages are generally of two types: referential, in which each electrode is connected to a single reference electrode such as the ear; and bipolar, in which electrodes are connected sequentially to one another to form a chain. A standard EEG generally records about 30 minutes of brain activity, both in the awake state and in the first two stages of sleep. Various activating procedures are used during the recording of an EEG, including hyperventilation and photic stimulation. These activating procedures may precipitate seizure discharges in some patients with seizure disorders, thereby increasing the sensitivity of the test.

The amplitudes of scalp electrical potentials are low, averaging 30 to 100 μV. They represent a summation of excitatory postsynaptic potentials and inhibitory postsynaptic potentials that are largely generated by the pyramidal cells in layer 4 of the cerebral cortex. Action potentials are of too brief a duration to have an effect on the EEG.

The EEG is analyzed with respect to symmetry between each hemisphere, wave frequency and amplitude, and the presence of spikes (20 to 70 milliseconds) and sharp waves (70 to 200 milliseconds), which may indicate a seizure focus. Electroencephalographic frequencies are divided into four categories as follows: delta (<4 Hz), theta (4-7 Hz), alpha (8-13 Hz), and beta (>13 Hz).

The normal waking EEG (Fig. 396-2A) in a patient with eyes closed contains rhythms of alpha frequency in the occipital leads and beta frequency in the frontal leads. Normal sleep causes a generalized slowing of EEG frequencies and an increase in amplitude in each stage of sleep, such that stage N3 sleep consists of more than 50% large-amplitude delta rhythms. EEG abnormalities are of two types: abnormalities in background rhythm and abnormalities of a paroxysmal nature (Table 396-4).[6]

The major usefulness of the EEG is for diagnosis and categorization of a seizure disorder (see Fig. 396-2B).[7] EEGs are neither highly sensitive nor completely specific for a diagnosis of seizures. Because seizures are paroxysmal events, it is not unusual for an EEG to be normal—or only minimally abnormal—in a patient with epilepsy if it is recorded during an interictal phase (the period between seizures). Only about 50% of patients with seizures show epileptiform activity on the first EEG. Repeating the EEG with provocative maneuvers such as sleep deprivation, hyperventilation, and photic stimulation may increase this percentage to 90%. Conversely, about 1% of adults and 3.5% of children who are neurologically normal and who never had a seizure have epileptiform activity on an EEG.

The EEG may provide clues to the diagnosis of certain neurologic conditions, including viral encephalitis, prion disorders, and some forms of coma. In each of these situations, the EEG can have specific patterns that suggest a specific neurologic diagnosis. In herpes simplex encephalitis, periodic lateralizing epileptiform discharges emanating from the temporal lobes are frequently present. Triphasic slow waves are common in hepatic encephalopathy (see Fig. 396-2C) but are a nonspecific finding. Creutzfeldt-Jakob disease is characterized by the presence of bilateral synchronous repetitive sharp waves. The EEG is also helpful in evaluating comatose patients, in confirming brain death when an apnea test cannot be performed because of cardiac instability, and for staging sleep in polysomnography.

In the past, the EEG was often used to localize neurologic lesions such as stroke, brain tumor, and abscess. With the advent of neuroimaging, EEG is almost never used for these purposes in developed countries.

Nerve Conduction Study

A nerve conduction study (NCS) is the recording and measurement of the compound nerve and muscle action potentials elicited in response to an electrical stimulus. To perform a motor NCS, a surface (active) recording electrode is placed over the belly of a distal muscle that is innervated by the nerve in question. A reference electrode is placed distally over the tendon. The nerve is then supramaximally stimulated at a predetermined distance proximal to the active electrode, and the resultant compound motor action potential (CMAP) is recorded. The terminal latency, amplitude, and duration of the evoked potential are measured directly, and the conduction velocity is calculated from the latencies of the evoked potentials with stimulation at two different points; the distance between the two points (conduction distance) is divided by the difference between the corresponding latencies (conduction time) to derive a calculated velocity (conduction velocity = distance ÷ time).

To perform a sensory NCS, the active recording electrode is placed over the portion of the skin innervated by the nerve in question, and a sensory nerve action potential is recorded after electrical stimulation of the nerve, similar to that noted for a motor NCS. NCS abnormalities include reduced amplitudes, prolonged terminal latencies, conduction block, and slowed conduction velocities (Table 396-5).

The NCS is helpful in documenting the existence of a neuropathy, quantifying its severity, and noting its distribution (i.e., whether it is distal, proximal, or diffuse). In addition, the NCS can provide information on the modality involved (i.e., motor versus sensory) and can suggest whether the lesion is axonal or demyelinating. The NCS is also helpful in diagnosis of compressive mononeuropathies, such as carpal tunnel syndrome, ulnar palsy, peroneal nerve palsy, and tarsal tunnel syndrome.

F Wave and H Reflex

The F wave and H reflex are ways of looking at the conduction characteristics for proximal portions of nerves, including the nerve roots. The F wave is a late CMAP evoked intermittently from a muscle by a supramaximal electrical stimulus to the nerve, and it is due to antidromic activation (backfiring) of alpha motor neurons. F waves can be elicited from practically all distal motor nerves. The H reflex is a late CMAP that is evoked regularly from a muscle by a submaximal stimulus to a nerve, and it is due to stimulation of Ia afferent fibers (a spinal reflex). The H reflex can be routinely obtained from calf muscles only with stimulation of the tibial nerve in the popliteal fossa.

F waves are helpful in diagnosis of Guillain-Barré syndrome, in which demyelination is often confined to the proximal portions of nerves early in the course of the disease. The H reflex is often absent in patients with acute S1 radiculopathy.

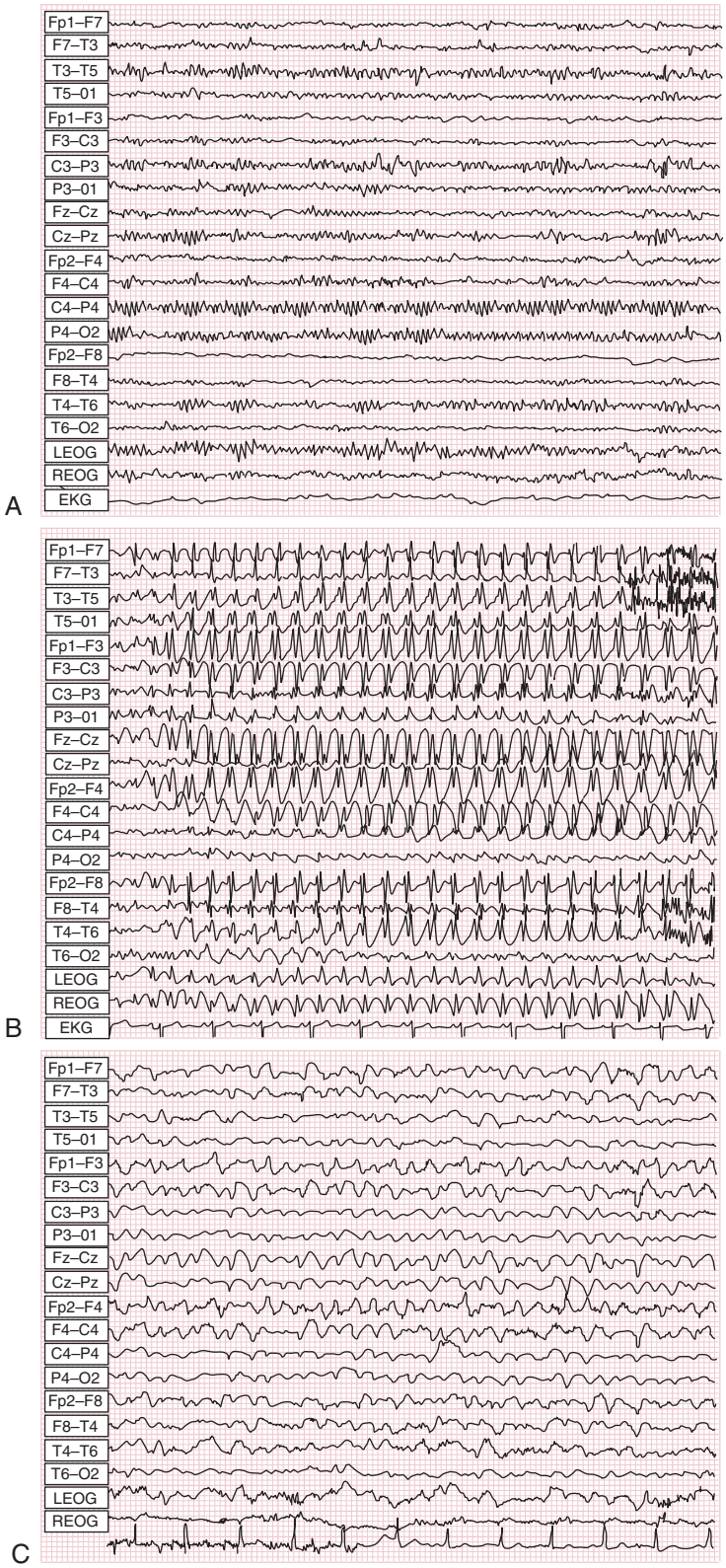

FIGURE 396-2. **Normal and abnormal electroencephalograms. A, The EEG of a normal awake adult. B, A 3-Hz spike and wave activity, a pattern seen in absence epilepsy. In each record, channels 1 through 8 and 11 through 18 represent left- and right-sided bipolar electrode placements, respectively. Channels 9 and 10 represent midline bipolar electrode placements, and channels 19 and 20 represent the left and right electro-oculograms (eye movements). Each major horizontal division represents 1 second. C, Triphasic slow waves, a pattern seen in hepatic or other metabolic encephalopathies.**

TABLE 396-4 ELECTROENCEPHALOGRAPHIC ABNORMALITIES

ELECTROENCEPHALOGRAPHIC ABNORMALITY	CLINICAL CORRELATE
BACKGROUND RHYTHM ABNORMALITIES	
Generalized slowing	Most metabolic encephalopathies
Triphasic waves	Hepatic, renal, and other metabolic encephalopathies
Focal slowing	Large mass lesions (tumor, large stroke)
Electrocerebral inactivity with lack of response to all stimuli	Neocortical death, hypothermia, drug overdose
PAROXYSMAL ABNORMALITIES	
3-Hz spike and wave, augmented by hyperventilation	Absence epilepsy
3- to 4-Hz spike and wave in light sleep or with photic stimulation	Primary generalized epilepsy
Central to midtemporal spikes	Benign rolandic epilepsy, other partial epilepsies
Anterior temporal spikes or sharp waves	Simple or complex focal (partial) seizures of mesial temporal origin
Hypsarrhythmia (high-voltage chaotic slowing with multifocal spikes)	Infantile spasms (West syndrome)
Burst suppression	Severe anoxic brain injury, barbiturate coma

TABLE 396-5 NERVE CONDUCTION STUDY ABNORMALITIES

ABNORMALITY	CLINICAL CORRELATE
Reduced CMAP amplitude	Axonal neuropathy
Prolonged terminal latency	Demyelinating neuropathy Distal compressive neuropathy
Conduction block	Severe focal compressive neuropathy Severe demyelinating neuropathy
Slowed conduction velocity	Demyelinating neuropathy

CMAP = compound muscle action potential.

Repetitive Stimulation Study

A repetitive stimulation study is a method of measuring electrical conduction properties at the neuromuscular junction. To perform a repetitive stimulation study, a surface recording electrode is placed over a muscle belly, and the nerve innervating that muscle is electrically stimulated with a supramaximal stimulus at a certain frequency. A series of electrical potentials are then recorded whose amplitude is roughly proportional to the number of muscle fibers being activated.

A repetitive stimulation study is helpful in diagnosis of neuromuscular junction disorders such as myasthenia gravis and myasthenic syndrome (Lambert-Eaton syndrome). In myasthenia gravis, the amplitudes of evoked potentials become progressively smaller with repetitive stimulation in clinically involved muscles. Clinically uninvolved muscles often do not demonstrate this decrement. In myasthenic syndrome, an increment is seen in the amplitudes of evoked potentials with rapid repetitive electrical stimulation.

Electromyography

Electromyography (EMG) is the recording and study of insertional, spontaneous, and voluntary electrical activity of muscle.[8] It allows physiologic evaluation of the motor unit, including the anterior horn cell, peripheral nerve, and muscle.

EMG is performed by insertion of a needle electrode into the muscle in question and evaluation of the motor unit action potentials both visually (on a computer screen) and aurally (over a loudspeaker). Muscles are typically studied at rest and during voluntary contraction. During EMG, the electrical activity of muscle is studied in four settings (Table 396-6): insertional activity (occurring within the first second of needle insertion), spontaneous activity (electrical activity at rest), voluntary activity (electrical activity with muscle

contraction), and recruitment pattern (change in electrical activity with maximal contraction).

EMG is helpful in evaluation of patients with weakness in that it can help determine whether the weakness is due to anterior horn cell disease, nerve root disease, peripheral neuropathy, or an intrinsic disease of muscle itself (myopathy). EMG can differentiate acute denervation from chronic denervation and may thus give an indication about the time course of the lesion causing the neuropathy. In addition, on the basis of which muscles have an abnormal EMG pattern, it is possible to determine whether the neuropathy is due to a lesion of a nerve root (radiculopathy), the brachial or lumbosacral plexus (plexopathy), an individual peripheral nerve (mononeuropathy), or multiple peripheral nerves (polyneuropathy).

EMG is also helpful in differentiating active (inflammatory) myopathies from chronic myopathies. Active myopathies include dermatomyositis, polymyositis, inclusion body myositis, and some forms of muscular dystrophy (e.g., Duchenne dystrophy). Chronic myopathies include the other muscular dystrophies, the congenital myopathies, and some metabolic myopathies.

TABLE 396-6 ELECTROMYOGRAPHIC ABNORMALITIES

ABNORMALITY	CLINICAL CORRELATE
INSERTIONAL ACTIVITY	
Prolonged	Acute denervation Active (usually inflammatory) myopathy
SPONTANEOUS ACTIVITY	
Fibrillations and positive waves	Acute denervation Active (usually inflammatory) myopathy
Fasciculations	Chronic neuropathies Motor neuron disease (rare fasciculations may be normal)
Myotonic discharges	Myotonic disorders Acid maltase deficiency
VOLUNTARY ACTIVITY	
Neuropathic potentials: large-amplitude, long-duration, polyphasic potentials	Chronic neuropathies and anterior horn cell diseases
Myopathic potentials: small-amplitude, short-duration, polyphasic potentials	Chronic myopathies Neuromuscular junction disorders
RECRUITMENT	
Reduced	Chronic neuropathic disorders
Rapid	Chronic myopathies

Myotonic dystrophy and myotonia congenita produce characteristic myotonic discharges.

It may take several weeks for a muscle to develop EMG signs of acute denervation after nerve transection. For this reason, EMG performed in the acute setting after nerve injury should be interpreted with caution, and it may need to be repeated at a later date.

Evoked Potentials

Evoked potentials are ways of measuring conduction velocities for sensory pathways in the CNS by means of computerized averaging techniques. Three types of evoked potentials are routinely performed: visual, brain stem auditory, and somatosensory.

Pattern Reversal Visual Evoked Potentials

The pattern reversal visual evoked potential (PVEP) assesses the function of central visual pathways, in particular the optic nerves. To perform this test, EEG electrodes are placed over the occipital regions of the scalp, and the patient is asked to look at the center of a black-and-white checkerboard screen with one eye patched. The color of the checks alternates about twice per second, a process known as pattern reversal. The scalp potentials elicited by approximately 100 such pattern reversals are then recorded and signal averaged by a computer. This signal averaging cancels the random EEG activity and differentially amplifies the evoked potential, which consists of a major positivity with a latency of about 100 milliseconds (the so-called P100 response). This response is recorded for each eye, and its latency is measured. A prolonged P100 latency in one eye, in the absence of ocular disease, implies slowed conduction velocity in the optic nerve and suggests demyelination of that nerve. PVEP testing is helpful when multiple sclerosis is suspected clinically and it is necessary to document the presence of a second demyelinating lesion in the CNS that may not be clinically evident (Fig. 396-3).

Brain Stem Auditory Evoked Potentials

The brain stem auditory evoked potentials (BAEP) assess function in the central auditory pathways in the brain stem. EEG electrodes are placed over the vertex and mastoid process, and a series of clicks at a frequency of 5 Hz are delivered to each ear separately for 3 minutes. The scalp potentials elicited by the clicks are then recorded and signal averaged by a computer. This signal averaging cancels the random EEG activity and differentially amplifies the evoked potential. A series of five waves are recorded for each ear, and each wave corresponds to a different point in the central auditory pathway (Table 396-7). The wave latencies for the right and left ears are compared, and a delay in any of the latencies suggests a lesion at that point in the central brain stem auditory pathway. BAEP testing is helpful in diagnosis of acoustic schwannoma and other tumors in the cerebellopontine angle.

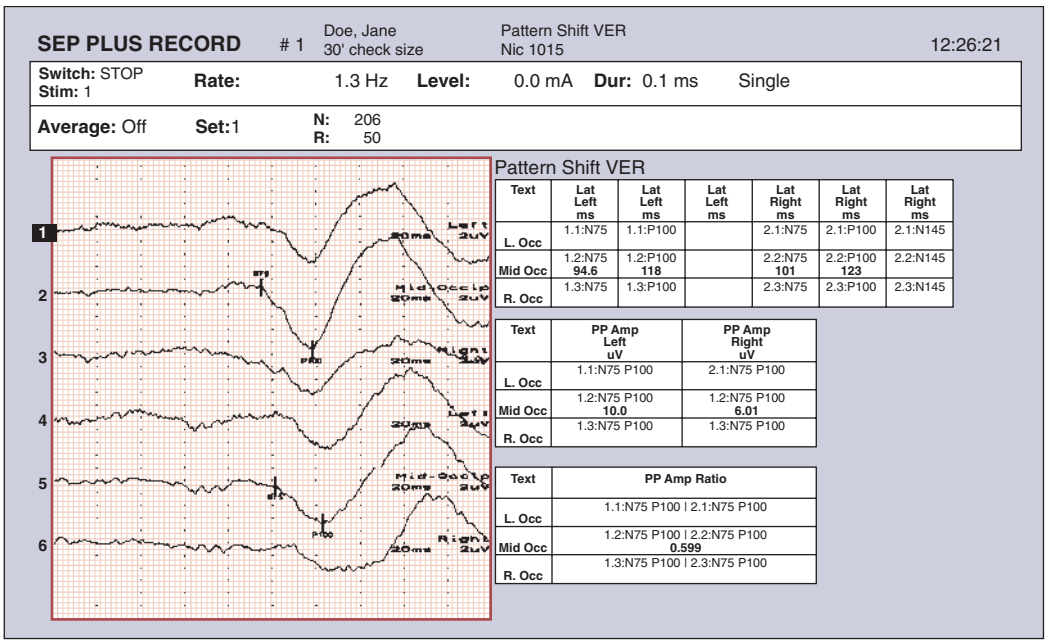

FIGURE 396-3. Abnormal pattern reversal visual evoked potential in a patient with multiple sclerosis. The prolonged P100 wave latency with left eye stimulation suggests a conduction defect in the left optic nerve. The top three channels represent right eye stimulation, and the bottom three channels represent left eye stimulation. Each horizontal division represents 20 milliseconds.

Somatosensory Evoked Potentials

The somatosensory evoked potential (SEP) assesses conduction in the central somatosensory pathways in the posterior columns of the spinal cord, brain stem, thalamus, and primary sensory cortex in the parietal lobes. To perform SEP testing, recording electrodes are placed over Erb point and the cervical spine (for medial or ulnar nerve stimulation), over the popliteal fossa and lumbar spine (for peroneal or tibial nerve stimulation), and over the scalp. A series of 1000 to 2000 electric shocks at a frequency of 5 Hz are delivered to the median or ulnar nerve (for an upper extremity SEP) or to the fibular (peroneal) or tibial nerve (for a lower extremity SEP). The scalp potentials elicited by the electric shocks are then recorded and signal averaged by a computer. This signal averaging cancels the random EEG activity and differentially amplifies the evoked potential. A series of waves are recorded for each nerve stimulated, with each wave corresponding to a different point in the somatosensory pathways in the spinal cord, brain stem, and cerebral cortex. The wave latencies for the right and left limbs are compared, and a delay in any of the latencies suggests a lesion at that point in the somatosensory pathways.

TABLE 396-7 BRAIN STEM AUDITORY EVOKED POTENTIAL WAVE GENERATORS

WAVE	LOCATION
I	Auditory nerve
II	Cochlear nucleus
III	Superior olivary nucleus
IV	Lateral lemniscus
V	Inferior colliculus

SEP testing, like PVEP, is helpful when multiple sclerosis is suspected clinically and it is necessary to document the presence of a second demyelinating lesion in the CNS that may not be clinically evident. SEP testing is also useful for prognostication in comatose patients and for monitoring of spinal cord function intraoperatively in patients undergoing spinal surgery.

Electronystagmography

Electronystagmography accurately records eye movements and nystagmus after certain provocative maneuvers. To perform this test, disc electrodes are placed over the bridge of the nose and lateral to each outer canthus, and the electrical leads from these discs are connected to an oscilloscope. Because the cornea is electropositive and the retina is electronegative, these electrodes accurately record lateral eye movements. The patient is first observed for spontaneous nystagmus with the eyes open and closed and then for nystagmus evoked with lateral gaze, for nystagmus induced by hot and cold air instilled in the outer ears (caloric induced), and for positional nystagmus. The last is performed by rotating the patient in a specialized chair. Spontaneous nystagmus suggests a vestibular pathologic lesion, as does an imbalance in the nystagmus evoked by these maneuvers in the right and left ears.

Imaging

On the basis of the relative advantages and disadvantages of computed tomography (CT), magnetic resonance imaging (MRI), and other neuroimaging modalities, different clinical entities can and should be assessed differently (Table 396-8). In acute ischemic stroke (Chapter 407) without bleeding, CT abnormalities typically appear within 4 to 12 hours and are seen even earlier with larger infarctions and embolic infarctions. CT detects hemorrhagic stroke (Chapter 408) acutely and can estimate its age. CT is also the preferred initial imaging modality for detection of intraparenchymal hemorrhage and subarachnoid hemorrhage, and it often suggests whether an aneurysm is the likely cause. Either CT angiography or magnetic resonance angiography can display the three-dimensional anatomy of aneurysms with sufficient detail for therapy to be planned, but surgical treatment generally

TABLE 396-8 STRENGTHS AND WEAKNESSES OF SELECTED IMAGING MODALITIES

MODALITY	STRENGTHS	WEAKNESSES
Computed tomography (CT)	Fast; best test for acute intraparenchymal or subarachnoid hemorrhage and calcification; easy to monitor patients; excellent for bones	Less sensitive to parenchymal lesions than MRI; potential for significant reaction to contrast material; radiation exposure
Conventional angiography	Best imaging modality for aneurysms, vascular malformations, and vasculitis	Invasive and often lengthy; risk of stroke and other complications
Conventional myelography	Good images of nerve roots and small osteophytic lesions; accurate for bony stenosis; useful in patients with contraindications to MRI	Invasive, with risk of complications from lumbar puncture and instillation of contrast material; does not image intramedullary lesions well
CT myelography	Excellent for imaging nerve roots and detecting root compression from degenerative processes	Invasive, with risk of complications from lumbar puncture and instillation of contrast material
Magnetic resonance imaging (MRI)	Noninvasive; no radiation; multiplanar; extremely sensitive, safe contrast agent	Less sensitive than CT for detection of subarachnoid hemorrhage and calcification; less sensitive for bony skull fractures; contraindicated in patients with implanted metallic devices or foreign bodies; the patient must be able to cooperate and tolerate confined space; time-consuming relative to CT
Magnetic resonance angiography (MRA)	Noninvasive; good for screening for extracranial and intracranial vascular disease; may be performed with or without contrast agent	Need cooperative patient; technically demanding; may overestimate the degree of vascular stenosis (noncontrast MRA); cannot image distal vessels optimally without contrast agent; may miss small lesions (e.g., aneurysms)
Positron emission tomography (PET)	Limited role in helping to distinguish radiation necrosis from tumor; sometimes helpful in the diagnosis of Alzheimer disease and epilepsy	Requires a cyclotron to generate radioisotopes with a short half-life; lower resolution and less available than MRI or CT
Single-photon emission computed tomography (SPECT)	Occasionally useful in epilepsy; sensitive for diffuse pathologic processes; easier to use than PET	Lower resolution than PET, MRI, or CT
Proton magnetic resonance spectroscopy	Localization of seizure focus; may help diagnose and classify dementias such as Alzheimer disease; may distinguish brain tumors from other mass lesions; may distinguish radiation necrosis from recurrent tumor	Specificity not yet determined; not routinely available; lower resolution; time-consuming
Ultrasonography	Fast; easy to use; can be performed at the bedside to assess vessel patency	Does not assess the vertebral arteries; less sensitive and specific than MRA; cannot visualize vessels in the upper neck and cranial base
Transcranial Doppler (TCD)	Fast; easy to use; assesses vascular velocities quantitatively; can assess cerebral vasospasm and occluded vessels	Does not provide images of vessels

Reproduced and modified from Hackney D. Radiologic imaging procedures. In: Goldman L, Ausiello D, eds. *Cecil Medicine*, 23rd ed. Philadelphia: Saunders Elsevier; 2008:2623-2627.

requires preprocedure catheter arteriography. CT is the first-line method for evaluation of brain trauma and diagnosis of a subdural or epidural hematoma (Chapter 399), usually without requiring intravenous contrast material. However, MRI is better than CT to delineate the anatomy of a subdural hematoma and to estimate the age of the lesion. Many brain tumors are initially recognized on CT scans, but MRI is the preferred modality for detection and characterization of all brain tumors (Chapter 189), including those that might be the cause of new-onset seizures in adults.[9]

Grade A Reference

A1. Shaikh F, Brzezinski J, Alexander S, et al. Ultrasound imaging for lumbar punctures and epidural catheterisations: systematic review and meta-analysis. *BMJ*. 2013;346:f1720.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Holloway RG, Gramling R, Kelly AG. Estimating and communicating prognosis in advanced neurologic disease. *Neurology*. 2013;80:764-772.
2. Pringsheim T, Fiest K, Jette N. The international incidence and prevalence of neurologic conditions: how common are they? *Neurology*. 2014;83:1661-1664.
3. Sharshar T, Citerio G, Andrews PJ, et al. Neurological examination of critically ill patients: a pragmatic approach. Report of an ESICM expert panel. *Intensive Care Med*. 2014;40:484-495.
4. Marshall FJ. Approach to the elderly patient with gait disturbance. *Neurol Clin Pract*. 2012;2:103-111.
5. De Luigi AJ, Fitzpatrick KF. Physical examination in radiculopathy. *Phys Med Rehabil Clin N Am*. 2011;22:7-40.
6. Maganti RK, Rutecki P. EEG and epilepsy monitoring. *Continuum (Minneap Minn)*. 2013;19:598-622.
7. Cascino GD, Palmini A, Kwan P, et al. What is the standard approach to assessment of an unprovoked seizure in an adult? *Neurol Clin Pract*. 2012;2:294-300.
8. Pitt M. Update in electromyography. *Curr Opin Pediatr*. 2013;25:676-681.
9. Ho K, Lawn N, Bynevelt M, et al. Neuroimaging of first-ever seizure: contribution of MRI if CT is normal. *Neurol Clin Pract*. 2013;3:398-403.

REVIEW QUESTIONS

1. A 55-year-old woman has had gradually progressive weakness and numbness in her lower limbs. On examination, her right lower limb has spastic tone and 4–/5 strength diffusely. The left lower limb has normal tone and full power. Vibration and position sense are markedly impaired in the right lower limb, and pain and temperature perception are markedly impaired in the left lower limb. The remainder of the neurologic examination is normal. Where would you best localize her lesion?

A. Brain stem
B. Cerebral hemispheres
C. Diencephalon
D. Peripheral nervous system
E. Spinal cord

Answer: E Neurologic examination findings confined to the lower limbs are highly suggestive of a spinal cord localization. The dissociated sensory abnormalities in this patient (impairment of vibration and position sense in the right lower limb and impairment of pain and temperature perception in the left lower limb), as well as the weakness and spastic tone in the right lower limb, all point to a hemisection of the right spinal cord at a midthoracic level, a so-called Brown-Séquard syndrome. Recall that there are two major somatosensory systems in the spinal cord: (1) the small-fiber pain and temperature system that courses in the contralateral anterolateral funiculus of the spinal cord and (2) the large-fiber vibration and position sense system that courses in the ipsilateral dorsal funiculus. The descending corticospinal tract courses in the ipsilateral lateral funiculus of the spinal cord. Lesions in the other structures listed are unlikely to produce this clinical picture.

2. A 70-year-old woman with hypertension and diabetes mellitus fell while getting out of bed and was brought to the emergency department by her daughter. CT scan of the brain shows a hemorrhage in the cerebellar vermis. Which of the following findings on neurologic examination is she most likely to have?

A. An involuntary movement disorder
B. Internuclear ophthalmoplegia
C. Limb ataxia
D. Lower limb spasticity
E. Truncal ataxia

Answer: E The cerebellum is the major brain structure involved in coordinating limb and trunk movements. It does this by comparing the present position of the limbs and trunk with the intended movement. The cerebellar hemispheres coordinate the limbs, and the cerebellar midline vermis coordinates the trunk. Hence, a hemorrhage in the cerebellar vermis, likely hypertensive in origin, is most likely to produce truncal ataxia (inability to stand upright with eyes open). Involuntary movement disorders usually arise from basal ganglia lesions. Internuclear ophthalmoplegia (inability to adduct one eye with horizontal gaze) is seen with brain stem lesions that involve the medial longitudinal fasciculus. Lower limb spasticity is seen with lesions of the corticospinal tracts and is most likely due to a spinal cord lesion.

3. A 24-year-old woman abruptly became unresponsive and stared for about a minute. About 10 seconds later, she started rubbing her thighs rhythmically with her right hand for about a minute. She was confused afterwards for about 10 minutes, following which she was tired and had a mild headache for the rest of the day. What is the most likely diagnosis?

A. Absence seizure
B. Focal seizure with dyscognitive features
C. Migraine with aura
D. Myoclonic seizure
E. Syncope

Answer: B A focal seizure with dyscognitive features (formerly classified as a complex partial seizure) typically originates in the temporal lobe and involves limbic circuits. The amygdala and hippocampus, which are the major components of the limbic system, are located in that lobe. During a focal seizure with dyscognitive features, consciousness is impaired and the patient manifests automatisms (e.g., rubbing the thighs rhythmically). The confusion following the spell represents the postictal state, which is often seen following partial seizures. Absence seizures are seen in children and consist of several seconds of unresponsiveness and eye blinking without postictal confusion. Migraine with aura rarely causes unresponsiveness, and rhythmic limb movements would be most unlikely. Myoclonus is a brief irregular jerk of a limb and can be seen with juvenile myoclonic epilepsy; the movement described in this vignette is too prolonged for myoclonus.

4. A 60-year-old man has noted difficulty walking for the past 4 days. He recently lost 20 pounds from dysphagia as a complication of radiation treatment for tongue cancer. On examination, he has weakness of right ankle dorsiflexion and eversion. Right ankle inversion is normal. There is a patch of numbness on the dorsum of the right foot over the first web space. These findings suggest involvement of which of the following?

A. Femoral nerve
B. Lateral femoral cutaneous nerve
C. Obturator nerve
D. Paroneal nerve
E. Tibial nerve

Answer: D The peroneal nerve innervates the anterior tibialis muscle, which causes ankle dorsiflexion, and the peroneal muscle, which causes ankle eversion. It also supplies sensation to the first web space of the foot. This nerve is commonly compressed at the fibular head, and rapid weight loss can facilitate compression at this site. The femoral nerve innervates the knee extensor muscles and supplies sensation to the anterior thigh. The lateral femoral cutaneous nerve supplies sensation to the anterolateral thigh and has no motor function. The obturator nerve innervates the thigh adductor muscles and provides sensation to the medial thigh. The tibial nerve innervates the plantar flexor muscles and supplies sensation to the sole of the foot.

5. A 60-year-old man sees his physician because of slowly progressive weakness in his limbs for the past 5 years. On examination, he has atrophy of the intrinsic muscles in his hands and feet. He is unable to walk on his heels and cannot make a tight grip with either hand. He has flaccid muscle tone, absent muscle stretch reflexes, and flexor plantar responses. He most likely has a lesion involving which of the following structures?

A. Brain stem
B. Cerebellum
C. Cerebral hemispheres
D. Peripheral nerves
E. Spinal cord

Answer: D Muscle atrophy (especially of distal muscles), flaccid muscle tone, and areflexia are all highly suggestive of a lesion of peripheral nerves. A lesion to the brain stem usually results in cranial nerve abnormalities; cerebellar lesions typically cause ataxia of the trunk and/or limbs. Lesions to the cerebral hemispheres often produce visual field deficits and loss of higher cortical functions, such as aphasia (defect in production or understanding of language), apraxia (defect in performance of a complex motor task), and agnosia (defect in recognizing a complex sensory stimulus). A spinal cord lesion typically produces loss of motor and sensory function below the level of the lesion. Extensor plantar responses (Babinski sign) would be expected with lesions to the brain stem, cerebral hemispheres, and spinal cord.

397

PSYCHIATRIC DISORDERS IN MEDICAL PRACTICE

JEFFREY M. LYNESS

OVERVIEW

Disorders in Psychiatry

Psychiatric disorders, also known as mental illnesses, are extraordinarily common and have a profound impact on well-being and functional status. Collectively, psychiatric disorders account for more aggregate disability than do those involving any other organ system, with depression alone being second only to cardiovascular disorders.

Psychiatric disorders are defined as disorders of the psyche—that is, conditions that affect thoughts, feelings, or behaviors. By definition, such mental disturbances must be sufficient to produce significant distress in the patient or impairment in role or other functioning. Because the pathogeneses of most psychiatric disorders are incompletely understood, classification is based on clinical syndromes that are defined by diagnostic criteria with high inter-rater reliability because they emphasize discrete reportable or observable symptoms and signs. Interestingly, however, underlying genetic variations and pathophysiologic mechanisms seem to cut across these descriptive diagnostic categories,[1] although such underlying findings do not yet aid in predicting the clinical course of disease or therapeutic decision making.

Specific Syndromes

Because many psychiatric disorders result from the direct influence of neurologic conditions, systemic diseases, or drugs on brain functioning, assessment of any new or worsened psychiatric condition must include evaluation for their potential contributions (Table 397-1). Delirium (Chapter 28) and dementia (Chapter 402), which are neurocognitive disorders defined by impairment in intellectual functions such as attention, memory, or language, are always the result of neurologic abnormalities, systemic illnesses, or drugs. Although intellectual impairment is the hallmark of neurocognitive disorders, these conditions also may manifest as alterations in other aspects of mental status, including mood, thought content, thought process, and behavior. If a noncognitive psychiatric syndrome is caused by an identifiable underlying condition, it is known as a secondary psychiatric disorder (e.g., depression secondary to hypothyroidism).

The major nonsecondary, noncognitive psychiatric syndromes (Table 397-2)[2] can coexist with multiple syndromes; for example, a patient suffering major depression with psychotic features may have depressive, anxiety, and psychotic syndromes simultaneously. Addictive disorders are considered in Chapters 33 and 34.

TABLE 397-1 IMPORTANT CAUSES OF PSYCHIATRIC SYNDROMES

CENTRAL NERVOUS SYSTEM DISEASES
Trauma
Tumor
Toxins
Seizures
Vascular
Infections
Genetic/congenital malformations
Demyelinating diseases
Neurodegenerative diseases
Hydrocephalus
SYSTEMIC DISEASES
Cardiovascular
Pulmonary
Endocrine
Metabolic
Nutritional
Infections
Cancer
DRUGS (e.g., recreational, prescription, or over-the-counter drugs)
Drug intoxication
Drug withdrawal

Comorbid Conditions in Psychiatry

It is common for persons who suffer from mental disorders to meet the diagnostic criteria for more than one condition. Although such comorbidity may reflect the limitations of current approaches to diagnosis, psychiatric comorbidity influences the choices or sequence of indicated treatments and may worsen the overall prognosis. Comorbidity with other medical conditions also is common, probably reflecting complex bidirectional causal relationships between physical and mental illnesses, and such comorbidity also often worsens the prognosis for both conditions.

Treatments in Psychiatry

Treatments in psychiatry are intended to reduce or eliminate symptoms, thereby improving the patient's distress and dysfunction and averting suicidal behavior. Pharmacotherapy remains an evidence-based mainstay of the treatment of many psychiatric conditions. The evidence for a number of forms of psychotherapy administered in individual, group, or family modalities supports its use as primary treatment or co-treatment of many conditions. Other psychosocial interventions, ranging from self-help groups to the use of structured treatment or residential programs, are often important adjuncts to treatment. Nonpharmacologic evidence-based somatic therapies include electroconvulsive therapy, light therapy, and vagal nerve stimulation for particular forms of major depression. Encouraging data are emerging to support deep brain stimulation for selected cases of severe depressive or obsessive-compulsive disorders.

Mood Disorders

Mood disorders are categorized as either depressive (also termed *unipolar*), characterized by depressive episodes only, or bipolar, characterized by manic or hypomanic episodes, typically with depressive episodes as well.

MAJOR DEPRESSIVE DISORDER

DEFINITION

Major depressive disorder is characterized by one or more episodes of idiopathic major depressive syndrome (Table 397-3).

EPIDEMIOLOGY

In the United States, major depression has a 12-month prevalence of approximately 7%, and it is at least 1.5 times more common in females than males. Lifetime prevalence is up to 10% in males and 20 to 25% in females. New depressive episodes have an annual incidence of approximately 3%. Depression accounts for more than twice as much disability in midlife as any other medical condition, and its overall cumulative burden is greater than that from all but cardiovascular disorders. The economic impact is also enormous, with U.S. estimates of annual costs for depression exceeding $12 billion for

TABLE 397-2 IMPORTANT PSYCHIATRIC SYNDROMES AND DISORDERS

SYNDROME	MAIN SYMPTOMS AND SIGNS	MAY OCCUR AS PART OF THESE DISORDERS
Neurocognitive	Deficits in intellectual functions (e.g., level of consciousness, orientation, attention, memory, language, praxis, visuospatial, executive functions)	Neurocognitive disorders Intellectual disability (if onset in childhood)
Mood: depressive	Lowered mood, anhedonia, negativistic thoughts, neurovegetative symptoms	Neurocognitive disorders Mood disorders (bipolar or depressive) (primary or secondary) Psychotic disorders (schizoaffective disorder)
Mood: manic	Elevated or irritable mood, grandiosity, goal-directed hyperactivity with increased energy, pressured speech, decreased sleep need	Neurocognitive disorders Bipolar disorder (primary or secondary) Psychotic disorders (schizoaffective disorder)
Anxiety	All include anxious mood and associated physiologic symptoms (e.g., palpitations, tremors, diaphoresis). May include various types of dysfunctional thoughts (e.g., catastrophic fears, obsessions, flashbacks) and behavior (e.g., compulsions, avoidance behavior).	Neurocognitive disorders Mood disorders (bipolar or depressive) (primary or secondary) Psychotic disorders (primary or secondary) Anxiety disorders (primary or secondary) Obsessive-compulsive and related disorders
Psychotic	Impairments in reality testing: delusions, hallucinations, thought process derailments	Neurocognitive disorders Mood disorders (bipolar or depressive) (primary or secondary) Psychotic disorders
Somatic symptom syndromes	Somatic symptoms with associated distressing thoughts, feelings, or behaviors	Mood disorders (bipolar or depressive) (primary or secondary) Anxiety disorders (primary or secondary) Obsessive-compulsive and related disorders Trauma-related disorders Somatic symptom disorders
Personality pathology	Enduring patterns of dysfunctional emotional regulation, thought patterns, interpersonal behavior, impulse regulation	Neurocognitive disorders (dementia) Personality change due to another medical condition Personality disorders

Author summary based on categories and criteria from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* 5th ed. (DSM-5) Washington, DC: American Psychiatric Association; 2013.

treatment, $8 billion for associated morbidity, and $33 billion for lost earnings and work productivity.

PATHOBIOLOGY

Major depression is probably not a single disease entity but rather a heterogeneous group of conditions with multiple pathogenic mechanisms. It is both multifactorial and polygenic: genetic factors account for approximately 40% of the risk for depression, but multiple gene loci, most of which are currently unknown, are probably involved in a complex interplay with developmental and environmental influences. Alterations in the brain's noradrenergic and serotonergic systems are likely related to the efficacy of current antidepressant medications. The hypothalamic-pituitary-adrenal axis is hyperactive in depression, as evidenced by a nonsuppressed response to the dexamethasone suppression test, although this test is too insensitive and nonspecific for clinical use as a diagnostic tool. Neuroimaging studies in subjects with depression show an array of findings, including smaller hippocampal volumes that may be the result of exposure to chronically elevated cortisol levels, and altered cerebral metabolic activity in regions including frontal-striatal circuitry and the anterior cingulate cortex. Cognitive psychology studies have demonstrated dysfunctional patterns of negative thinking, with distorted thoughts about self, the future, and the environment. Poor quality or absence of social relationships, and stressful life events, particularly events such as deaths, separations, or functional impairment, are powerfully associated with depression as well.

CLINICAL MANIFESTATIONS

The symptoms of depression (see Table 397-3) may be conceptually grouped as alterations in mood, ideation (i.e., thought content), and somatic/neurovegetative functioning. Importantly, patients with depressive illness may be seen without a depressed mood, albeit by definition they then must have loss of interest or pleasure in their usually desired activities. They may also exhibit prominent anxiety, irritability, or somatization. Although mild forms of major depression in the community often remit spontaneously within a few months without medical care, patients may have persistent symptoms for months or years before seeking treatment.

DIAGNOSIS

The diagnosis is made clinically by elicitation of findings from the history and mental status examination to determine the presence of major depressive syndrome. The differential diagnosis includes other idiopathic disorders with episodes of major depression, such as bipolar disorder (distinguished by a history of manic episodes) and schizoaffective disorder (distinguished by a history of psychotic episodes in the absence of depression). Major depression may accompany delirium or dementia, and secondary depression also commonly accompanies serious medical illnesses in the elderly;[3] these comorbid conditions require careful, well-coordinated care. Screening instruments (see Table 27-3 in Chapter 27) can help identify cases of depression. For example, using the two-item version of the Patient Health Questionnaire, the screener asks the patient: Over the past 2 weeks, how often have you (1) Had little interest or pleasure in doing things, or (2) Been feeling down, depressed, or hopeless. Responses for each question are scored as: 0 = not at all, 1 = several days, 2 = more than half the days, 3 = nearly every day. A score of 3 points or higher on the two-item screen is associated with 75% probability of having a depressive disorder.

TREATMENT

Rx

The three phases of treatment include (1) acute, in which treatment is provided to resolve the major depressive episode; (2) continuation, in which the acute treatment is continued for at least 4 to 8 months to prevent relapse; and (3) maintenance, for those with two to three or more episodes of recurrent depression, for whom treatment is maintained indefinitely to reduce the frequency and severity of future recurrences.[4] Combinations of psychotherapy and medication are used for more complex or severe clinical conditions.

Acute treatment of depression includes focused psychotherapies (Table 397-4), which are more efficacious than usual care and equivalent to medications when used for patients in primary care settings. Based on the patient's preference, psychotherapy rather than medication may be the initial treatment of mild to moderate major depression with prominent psychosocial stressors. Involvement of family members for education, support, and sometimes formal family therapy may be an important adjunctive or primary therapeutic approach. These therapies may be administered with decreased frequency during the continuation or maintenance phases of treatment. However, psychotherapies alone are insufficient for more severe forms of depression, including major depression with psychotic features. Meta-analyses suggest that the combination of medication with psychotherapy is more effective than medication alone in the initial treatment of mild to moderate major depression.

Medications should be used as initial treatment for most patients with more severe forms of major depression. Antidepressant medications (Table 397-5) are also effective for acute, continuation, and maintenance therapy. A meta-analysis found that sertraline and escitalopram have the best profiles of efficacy and tolerability, whereas mirtazapine and venlafaxine also have strong efficacy in head-to-head comparisons with other antidepressants. Overall,

TABLE 397-3 SYMPTOMS/SIGNS OF AN EPISODE OF MAJOR DEPRESSIVE SYNDROME

DIAGNOSTIC CRITERIA ***(must be present for a minimum of 2 consecutive weeks)***
Depressed mood (may be irritable mood in children and adolescents) most of the day, nearly every day, *OR*
Markedly diminished interest or pleasure most of the day, nearly every day *AND*
Weight loss or gain, or change in appetite (decrease or increase) nearly every day
Change in sleep (insomnia or hypersomnia) nearly every day
Psychomotor agitation or retardation nearly every day
Fatigue or loss of energy nearly every day
Feeling of worthlessness or guilt nearly every day
Diminished concentration or indecisiveness nearly every day
Recurrent thoughts of death or suicidal ideation, or a suicide attempt, or a specific suicide plan
MNEMONIC TO AID RECALL OF DIAGNOSTIC CRITERIA: ***"SIG: E CAPS" (i.e., prescribe energy capsules)*** **for depressed mood**
Sleep change
Interests decreased
Guilt
Energy decreased
Concentration decreased
Appetite/weight disturbance
Psychomotor changes
Suicide thoughts
DEPRESSIVE SYMPTOMS/SIGNS GROUPED CONCEPTUALLY, WITH ADDITIONAL COMMON PHENOMENA
Emotional
Depressed mood, sadness, tearfulness
Irritability (seen in all ages, perhaps most commonly in children/adolescents and the elderly)
Anxiety
Loss of interests or pleasure (anhedonia)
Ideational
Worthlessness/lowered self-esteem
Guilt
Hopelessness/nihilism
Helplessness
Thoughts of death, dying, suicide
Somatic/Neurovegetative
Change in appetite/weight
Change in sleep
Anergia
Decreased libido
Trouble concentrating
Diurnal variation in symptoms *(mornings—worst pattern is most characteristic)*
Other
Ruminative thinking *(tendency to dwell on one [negativistic] theme)*
Somatoform symptoms or somatic worry
Psychotic symptoms *(negativistic delusions most characteristic)*—defines the subtype "Major Depression with Psychotic Features"

Based on criteria from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* 5th ed. (DSM-5) Washington, DC: American Psychiatric Association; 2013.

TABLE 397-4 TREATMENTS FOR DEPRESSION

NAME OF PSYCHOTHERAPY	APPROACH
Cognitive psychotherapy	Identify and correct negativistic patterns of thinking
Interpersonal psychotherapy	Identify and work through role transitions or interpersonal losses, conflicts, or deficits
Problem-solving therapy	Identify and prioritize situational problems; plan and implement strategies to deal with top-priority problems
Psychodynamic psychotherapy	Use therapeutic relationship to maximize use of the healthiest defense mechanisms and coping strategies

however, data suggest that no second-generation agent is predictably better than others,[■] although agents targeting noradrenergic as well as serotonergic systems may be more efficacious in more severe depression. Because antidepressant medications typically do not begin to improve symptoms for at least 1 to 2 weeks, with maximal benefit accruing up to at least 6 to 8 weeks, it is crucial to see patients regularly (every 1 to 2 weeks initially) to monitor their clinical status, provide support and education, and foster adherence. Antidepressant medications appear to increase the relative risk for suicidal behavior in adolescents and young adults, so such patients require careful benefit/risk assessments and close monitoring. By comparison, the relative risk for suicidal behavior is not increased by drug treatment in individuals older than age 25 and is substantially lowered in older adults. For patients with a psychotic depression, the addition of an antipsychotic medication (see later) to an antidepressant may be more efficacious than either alone.[■] A single intravenous dose of ketamine may rapidly reduce severe depressive symptoms within 24 hours, although its use must be considered investigational at the present time.[■] Electroconvulsive therapy is preferred for the most severe forms of major depression, including major depression with psychotic features, and is also used for depression refractory to other forms of treatment. Deep brain stimulation is an investigational therapy for otherwise refractory depression.

Optimal care for depression in primary care and other treatment settings may be enhanced by the use of collaborative chronic care models. However, despite considerable evidence supporting such models,[■] lack of reimbursement mechanisms has limited their implementation in most communities and clinical settings.

PROGNOSIS

Optimal guideline-based treatment of major depression results in full remission in up to 80% of patients, and the expectation is that patients with major depression will return to baseline functioning after resolution of the depressive episodes. However, at least 50 to 70% of patients will suffer recurrent episodes, up to 20% may experience chronic major depression, and many more will achieve only partial remission with persistent lower-level symptoms because of a variety of factors, including limited access to care, nonadherence, or insufficiently assertive treatments.

BIPOLAR DISORDER

DEFINITION AND EPIDEMIOLOGY

Bipolar disorder is characterized by recurrent episodes of idiopathic mania. Most persons with bipolar disorder also have recurrent episodes of major depression.[5]

The 12-month prevalence of bipolar disorder is approximately 0.6%. Males are affected slightly more often than females. The average age at first onset is late adolescence or early adulthood. Childhood onset is possible, but diagnosis may be difficult because of symptomatic overlap with other conditions of childhood, including attention-deficit/hyperactivity disorder. Onset in midlife to late life is also possible, although most late-onset mania is secondary to other medical conditions or drugs rather than idiopathic bipolar disorder.

PATHOBIOLOGY

Even though the pathogenesis of bipolar disorder remains unclear, genetic factors play a greater role than in unipolar depressive conditions. Heritability has been traced to several specific loci in rare families, but genetic screening is not yet clinically useful, and the gene associations have to date revealed no unifying pathophysiologic themes. Most cases of bipolar disorder are polygenic and multifactorial, with genetic factors accounting for approximately 50% of the risk for the disorder. Dysregulation of the frontostriatal systems is probably involved in the manifestations of the illness. Though not specific enough to be diagnostic, structural neuroimaging studies show increased ventricular-brain ratios suggestive of parenchymal atrophy. Phase advance of central circadian rhythms can precipitate episodes of mania, so the decreased sleep need of persons with incipient mania may produce a vicious cycle in which phase-advanced circadian cycles lead to a further decreased need for sleep, thereby resulting in further phase advancement. Psychosocial stressors also often play a role in precipitating episodes of both mania and depression.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The symptoms of mania include a distinct period of abnormally and persistently elevated (euphoric) or irritable mood; goal-directed hyperactivity, often for pleasurable activities, with poor judgment that leads to long-lasting adverse financial, psychosocial, or medical consequences (e.g., sprees of

TABLE 397-5 COMMONLY USED ANTIDEPRESSANT MEDICATIONS*

NAME OF CLASS/ SPECIFIC MEDICATION	IMMEDIATE MECHANISM OF ACTION	INITIAL ADULT DOSE	TARGET ADULT DOSE RANGE†	SIDE EFFECTS	COMMENTS
SSRIs (selective serotonin reuptake inhibitors)	Inhibit presynaptic reuptake of serotonin			Nausea, diarrhea, sexual dysfunction, serotonin syndrome	
Citalopram		20 mg daily	20-40 mg daily (maximum 20 mg daily in patients age > 60 yr)	Risk of QTc prolongation/ torsade de pointes in at-risk patients	Few drug-drug interactions
Escitalopram		10 mg daily	10-20 mg daily		Enantiomer of citalopram
Fluoxetine		20 mg daily	20-40 mg daily (depression), up to 80 mg daily (OCD)		Long half-life; tends to be activating
Paroxetine		20 mg daily	20-50 mg daily	Anticholinergic effects	Tends to be sedating
Sertraline		25-50 mg daily	50-200 mg daily		Few drug-drug interactions
SNRIs (serotonin and norepinephrine reuptake inhibitors)	Inhibit presynaptic reuptake of serotonin and norepinephrine			Nausea, diarrhea, serotonin syndrome, sinus tachycardia, mild elevation in blood pressure, tremor	
Duloxetine		30-60 mg daily	30-60 mg daily on a twice-daily schedule, maximum of 120 mg/day		
Venlafaxine		37.5 mg bid	150-375 mg/day on bid schedule		XR form allows once-daily dosing
Desvenlafaxine		50 mg daily	50 mg daily, maximum of 100 mg ER daily		Metabolite of venlafaxine
TCAs (tricyclic antidepressants)	Inhibit presynaptic reuptake of serotonin and norepinephrine (in varying proportions depending on the specific TCA)			Anticholinergic effects, sedation, orthostatic hypotension, tremor, cardiac conduction delays, ventricular arrhythmias	
Amitriptyline		25-75 mg qhs	150-300 mg qhs		Strongly anticholinergic and sedating; aim for combined amitriptyline/nortriptyline blood level of 120-250 ng/mL
Desipramine		25-75 mg daily	150-300 mg daily		Aim for blood level of 115-250 ng/mL
Doxepin		25-75 mg qhs	150-300 mg qhs		Strongly sedating
Imipramine		25-75 mg daily	150-300 mg daily		Strongly anticholinergic; aim for combined imipramine/ desipramine blood level of 180-350 ng/mL
Nortriptyline		25-50 mg qhs	50-150 mg qhs		Aim for blood level of 50-150 ng/mL; least anticholinergic of the TCAs
MAOIs (monoamine oxidase inhibitors)	Inhibit monoamine oxidase, the enzyme that catalyzes oxidative metabolism of monoamine neurotransmitters			Need for tyramine-free diet to avoid sympathomimetic (hypertensive) crisis; sedation, anticholinergic effects, tremor, orthostatic hypotension	
Isocarboxazid		10 mg bid	20-60 mg/day in bid-qid dosing		
Phenelzine		15 mg tid	45-90 mg/day in tid or qid dosing		
Selegiline	(selective MAO-B inhibitor)	5 mg bid	5 mg bid	Tyramine-free diet not required	Take with meals
Tranylcypromine		10 mg tid	30-60 mg/day in tid dosing		
Other					
Bupropion	Unknown, although it is a weak inhibitor of presynaptic reuptake of norepinephrine and dopamine	75-150 mg/ day	300-450 mg/day	Activating; risk for seizures reduced by divided dosing and careful dosage titration	Divided dosing required unless using SR or XL forms
Mirtazapine	Antagonist at α_2 and 5-HT_2 receptors	15 mg qhs	30-45 mg qhs; maximum of 45 mg qhs	Sedation, hyperphagia	Becomes more stimulating at higher doses
Trazodone	Inhibits presynaptic reuptake of serotonin; antagonist at 5-HT_2 and 5-HT_3 receptors	25-50 mg qhs	300-600 mg qhs for depression, 25-100 mg qhs for insomnia	Sedation, priapism	Few sexual side effects
Vilazodone	Inhibits presynaptic reuptake of serotonin; agonist at 5-HT_{1A} receptors	10 mg daily	40 mg daily	Nausea, diarrhea, sexual side effects	Dosage must be increased slowly

*Patients on any of these medications must be monitored for suicidal thoughts.
†Target doses in the elderly may be lower.
ER = extended release; 5-HT_2 = 5-hydroxytryptamine; OCD = obsessive-compulsive disorder; qhs = at bedtime; SR = sustained release; XR = extended release.

TABLE 397-6 SYMPTOMS/SIGNS OF AN EPISODE OF MANIA

DIAGNOSTIC CRITERIA
A distinct period of abnormally; persistently elevated, expansive, or irritable mood; and abnormally and persistently increased goal-directed activity or energy lasting ≥ 1 week and present most of the day, nearly every day, *AND*
3 or more of the following symptoms/signs (4 or more if the mood abnormality is only irritability):
Inflated self-esteem/grandiosity
Decreased need for sleep
More talkative or pressure to keep talking
Subjective experience of racing thoughts or flight of ideas observed on examination
Distractibility
Increase in goal-directed activity or psychomotor agitation
Excessive involvement in activities with a high potential for painful consequences
MANIC SYMPTOMS/SIGNS GROUPED CONCEPTUALLY, WITH ADDITIONAL COMMON PHENOMENA
Emotional
Euphoria
Irritability
Labile affect
Ideational
Grandiosity
Somatic/Neurovegetative
Increased energy
Psychomotor agitation
Decreased need for sleep
Distractibility
Other
Goal-directed hyperactivity
Pressured speech
Impaired judgment
Flight of ideas
Psychotic symptoms (may include delusions, hallucinations, or derailment of thought processes such as loose associations)—defines the subtype "mania with psychotic features"

From *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. (DSM-5) Washington, DC: American Psychiatric Association, 2013, with permission.

spending, sexual activity, or gambling); increased energy; decreased need for sleep; pressured speech; and distractibility.[6]

As with major depression, the diagnosis is based on findings from the history and examination revealing a pattern of recurrent manic episodes (Table 397-6) that are usually interspersed with major depressive episodes and cannot be explained by other medical conditions, medications, or other substances. Although persons with bipolar disorder may become psychotic while in manic or depressed states, a history of psychotic symptoms in the absence of mania or depression indicates a diagnosis other than bipolar disorder. Manic and depressive episodes may also be seen in the course of delirium (Chapter 28) and dementia (Chapter 402), in which case the psychiatric symptoms are accompanied by the neurocognitive impairment that is the hallmark of the latter conditions.

TREATMENT

Rx

The mainstay of treatment for bipolar disorder is mood stabilizer medications to reduce the frequency and severity of recurrent manic and depressive episodes.[7] Traditional mood stabilizers with substantial evidence base to support their use include lithium (typical dose of 600 to 1500 mg/day or higher given in two or three divided doses as needed to achieve plasma levels of 0.6 to 1.2 mEq/L [up to 1.4 mEq/L in acute mania]), valproic acid (typical dose of 500 to 1500 mg/day or higher as tolerated to achieve plasma levels of 50 to 100 µg/mL), and carbamazepine (typical dose of 400 to 1200 mg/day as tolerated to achieve plasma levels of 4 to 12 µg/mL). The combination of lithium plus valproate is superior to valproate alone for prevention of relapses.[8] A number of other anticonvulsants have been tried but generally with less empirical support for their use, although lamotrigine (starting at 25 mg/day, maximum dose of 200 mg/day, titrated slowly to minimize the risk for Stevens-Johnson syndrome) can be used for prophylaxis against depressive episodes. Even though several second-generation antipsychotic medications have received approval by the U.S. Food and Drug Administration (FDA) for their mood-stabilizing properties, their potential to precipitate metabolic syndrome (and to a lesser extent, tardive dyskinesia) should limit their use as maintenance medications to patients for whom other mood stabilizers are inefficacious or poorly tolerated. For acute episodes of mania, second- or first-generation antipsychotics are more rapidly efficacious than mood stabilizers, with doses similar to their use for acute psychosis (see Table 397-12). For acute treatment of depressive episodes, antidepressants may be required, but they may precipitate mania. Therefore, patients should receive therapeutic doses of a mood stabilizer first, and exposure to antidepressant medication should be for the minimum dose and duration required. Electroconvulsive therapy is useful for refractory mania or depression and for patients with relative contraindications to medications, such as pregnant women. Standard psychotherapies for unipolar depression also may be used for bipolar depression. Ongoing psychotherapy may be important to encourage compliance with maintenance treatments and help patients manage psychosocial stressors, thereby minimizing their impact on precipitating mania or depression.

PROGNOSIS

Although the classically described course of bipolar disorder includes return to baseline functioning between episodes, some patients may experience frequent debilitating episodes (known as "rapid cycling," defined as four or more episodes per year), and others may experience deterioration in overall functioning over time.

OTHER MOOD DISORDERS

Although the diagnosis of chronic major depression should be made in patients with long-lasting major depressive episodes, others may have chronic (≥2 years) lower-level depressive symptoms known as persistent depressive disorder (dysthymia), a significant minority of whom will improve with a combination of antidepressant medication and psychotherapy. Other patients may have "less than major depression" of shorter duration, often referred to as "subsyndromal" or "subthreshold" depression. Growing evidence suggests that broad psychosocial interventions (e.g., bibliotherapy, social activation) may prevent progression to full-fledged major depression in such patients. Premenstrual dysphoric disorder manifests as cyclical depressive and anxiety symptoms that resolve in the week post menses and recur in the week before the onset of menses; this is the only mood disorder that may respond to brief cyclical administration of antidepressant medication.

Less severe bipolar-related disorders include bipolar II disorder, which is characterized by episodes of hypomania (i.e., low-level manic symptoms without substantial functional impairment and without psychosis) and episodes of major depression. Such patients typically seek care during depressive episodes rather than during hypomania, but antidepressant medication may worsen the manic symptoms. It is therefore imperative to ask about a history of manic or hypomanic symptoms in the evaluation of all patients with depression. Cyclothymic disorder, which includes episodes of hypomania and low-level depressive episodes, may be difficult to distinguish from the mood instability seen in "cluster B" personality disorders (see later).

ANXIETY DISORDERS

DEFINITION

The anxiety disorders (Table 397-7) are a group of conditions whose hallmark is idiopathic anxiety, typically accompanied by psychological (i.e., thought content) and somatic symptoms. Anxiety is a common accompanying symptom in many other psychiatric disorders, but the primary anxiety disorders lack the neurocognitive deficits, depressive or manic symptoms, or psychosis seen in the other disorders. Trauma-related and obsessive-compulsive disorders are now classified separately from the anxiety disorders.

EPIDEMIOLOGY

Anxiety disorders are a worldwide problem.[8] Panic disorder has a 12-month prevalence of 2 to 3%. Generalized anxiety disorder has a 12-month prevalence of approximately 3%, and the phobias collectively have a prevalence of 10 to 15% in the adult population. Clear data on incidence rates are not available. Most primary anxiety disorders have an age at first onset in adolescence through the mid-30s, with generalized anxiety disorder toward the older end of that range. Most anxiety symptoms with new onset in later life are due to mood or neurocognitive disorders or are secondary to medical illnesses or

TABLE 397-7 TYPES OF ANXIETY DISORDERS

ANXIETY DISORDER	MAJOR CLINICAL CHARACTERISTICS
Panic disorder	Recurrent unexpected panic attacks, typically with anticipatory anxiety and avoidance behavior
Generalized anxiety disorder	Excessive anxiety and worry, not meeting the criteria for other anxiety disorders, lasting ≥ 6 months
Phobias:	
Agoraphobia	Anxiety about or avoidance of places or situations from which escape might be difficult or embarrassing or in which help might not be available in the event of panic symptoms
Social phobia (social anxiety disorder)	Anxiety provoked by exposure to social situations, typically with ensuing avoidance behavior; may be generalized (i.e., in response to many interpersonal situations) or specific in response to a particular social situation (e.g., using a public restroom, public speaking)
Specific phobia	Anxiety provoked by exposure to a specific feared object or (nonsocial) situation, typically with ensuing avoidance behavior

Author summary based on categories and criteria from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. (DSM-5) Washington, DC: American Psychiatric Association; 2013.

TABLE 397-8 COMMON SOMATIC MANIFESTATIONS OF ANXIETY

CARDIORESPIRATORY
Palpitations
Chest pain
Dyspnea or sensation of being smothered
GASTROINTESTINAL
Sensation of choking
Dyspepsia
Nausea
Diarrhea
Abdominal bloating or pain
GENITOURINARY
Urinary frequency or urgency
NEUROLOGIC/AUTONOMIC
Diaphoresis
Warm flushes or chills
Dizziness or presyncope
Paresthesias
Tremor
Headache

drugs; true late-onset primary anxiety disorders are often triggered by traumatic or other stressful life events.

PATHOBIOLOGY

Anxiety may be understood in part as inappropriate triggering of the stress response system, commonly referred to as the "fight-or-flight" response. However, it is important to recognize that the responses involve a wide range of cognitive, motor, neuroendocrine, and autonomic systems and thus are not limited to manifestations of sympathetic nervous system activity. The central nucleus of the amygdala is believed to play a crucial role in coordinating the anxiety response. The amygdala receives excitatory glutamatergic input from several cortical areas and from the thalamus, thereby allowing it to respond to a wide variety of stimuli, including sensory input from the external world, as well as stressors that are processed and recognized by cortical association areas. The amygdala in turn projects to the many brain regions that subserve the clinical manifestations of the anxiety response, in part through its direct projections to the important centers of monoaminergic systems: dopaminergic neurons of the ventral tegmental area in the midbrain, noradrenergic neurons in the locus caeruleus, and serotonergic neurons in the raphe nuclei.

From a cognitive psychology perspective, the pathogenesis of many anxiety disorders, particularly panic, may be understood as catastrophic misinterpretations of normal somatic sensations. A vulnerable individual may become aware of a normal or minimally abnormal body sensation, which is interpreted as something concerning, thereby leading to sympathetic and other autonomic arousal, which in turn leads to further somatic sensations (e.g., tachycardia, sweating) in what becomes a vicious cycle of thoughts and somatic symptoms.

CLINICAL MANIFESTATIONS

Most individuals experience one or more somatic symptoms (Table 397-8) that accompany psychic anxiety, regardless of whether the anxiety is normal or part of a pathologic condition. Such somatic symptoms may be referable to virtually every body organ system.

Many anxiety disorders include acute, discrete periods of symptoms known as panic attacks. In a panic attack, the patient experiences an abrupt surge in anxiety, fear-related thoughts, and somatic symptoms in the space of a few minutes ("crescendo onset"). The acute symptoms resolve quickly, typically within an hour or less.

Panic Disorder

Panic disorder consists of recurrent panic attacks. Although some panic attacks may be precipitated by situations known to be stressful, at least some attacks must be unexpected ("out of the blue"). Patients also exhibit anticipatory anxiety in which they experience ongoing psychic distress by worrying about their next panic attack or the attack's effects (e.g., humiliation if the attack were to happen in public view). In addition, patients manifest avoidance behavior by staying away from known triggers or from situations in which having a panic attack might be dangerous (e.g., driving) or particularly distressing (e.g., in public spaces). For many patients, the anticipatory anxiety and avoidance behavior may be more disabling than the panic attacks themselves. Avoidance behavior may overlap with agoraphobia, which is defined as a distressing and disabling fear of places or situations from which escape might be difficult or embarrassing or from which help might not be available in the event of panic-like symptoms. Common agoraphobic foci include being outside one's home alone, being on bridges or in tunnels, traveling by vehicle, or being in crowds or lines. A third or more of patients with panic disorder have comorbid agoraphobia, whereas others have agoraphobia alone or comorbid with other conditions.

Generalized Anxiety Disorder

This more heterogeneous condition is defined by the presence of clinically significant anxiety and associated somatic symptoms for 6 or more months. Generalized anxiety disorder is often "trumped" in the diagnostic hierarchy by other conditions that produce anxiety.

Phobias

The phobias are a group of conditions defined by the consistent ability of a specific environmental stimulus to elicit a pathologic anxiety response. Exposure to such a stimulus nearly always produces this response, so the patient avoids the stimulus whenever possible or endures the stimulus with considerable distress. In addition to agoraphobia, the other main types of phobias are social phobia (social anxiety disorder) and specific phobias (see Table 397-7).

DIAGNOSIS

Diagnosis of anxiety disorders must rest on consideration of both syndromic and etiologic perspectives.[9] From a syndromic perspective, a careful history and mental status examination are required to determine the pattern of anxiety and associated symptoms and to determine whether the phenomenology fits the pattern for any of the anxiety disorders as described earlier. The history and mental status examination must also assess for the presence of any other psychiatric disorder that might truly be comorbid with the anxiety disorder but might also supersede the anxiety disorder in the diagnostic hierarchy. For example, generalized anxiety may be seen as part of neurocognitive disorders (delirium or dementia), depressive or bipolar disorders, and psychotic disorders.

From an etiologic perspective, it is important to determine whether the anxiety disorder is primary (idiopathic) or secondary to a systemic or neurologic condition (see Table 397-1), drug intoxication, or withdrawal state. The evaluation should include laboratory tests (e.g., toxic drug screen) as guided by the differential diagnosis generated from the clinical evaluation.

TABLE 397-9 SELECTED ANTIANXIETY AND HYPNOTIC DRUGS*

DRUG	TRADE NAME	INITIAL DOSE	TARGET DOSE RANGE†	SIDE EFFECTS	COMMENTS
Benzodiazepines				Sedation, ataxia, risk for falls	Potential for abuse/dependence
Lorazepam	Ativan	0.5 mg bid-qid	2-6 mg/day, tid-qid dosing		Reliable IM absorption
Diazepam	Valium	2-5 mg bid-tid	10-40 mg/day, bid-tid dosing		Long half-life of drug and active metabolites
Triazolam	Halcion	0.125 mg qhs	0.125-0.25 mg qhs	Rebound insomnia	Used as hypnotic
Chlordiazepoxide	Librium	5 mg bid-tid	10-40 mg/day, bid-tid dosing		Long half-life of drug and active metabolites
Temazepam	Restoril	7.5 mg qhs	7.5-30 mg qhs		Used as hypnotic
Alprazolam	Xanax	0.25 mg tid-qid	2-8 mg/day, tid-qid dosing	Possibly greater addictive potential	
Clorazepate	Tranxene	7.5-15 mg bid-tid	15-60 mg/day, bid-tid dosing		
Flurazepam	Dalmane	15-30 mg qhs	15-30 mg qhs	Daytime somnolence	Used as hypnotic
Oxazepam	Serax	10-15 mg tid-qid	10-30 mg tid-qid		
Clonazepam	Klonopin	0.5 mg bid-tid	0.5-5 mg bid-tid		Long duration of action
Zaleplon	Sonata	5-10 mg qhs	5-20 mg qhs		"Nonbenzodiazepine" hypnotic
Zolpidem	Ambien	5-10 mg qhs	5-10 mg qhs		"Nonbenzodiazepine" hypnotic
Eszopiclone	Lunesta	1-2 mg qhs	1-3 mg qhs		"Nonbenzodiazepine" hypnotic
β-Blockers					
Propranolol	Inderal	20 mg bid	Individualize, 40-120 mg/day	Bradycardia, hypotension, potential for mental slowing	Only helps with sympathetically mediated somatic symptoms of anxiety

*Antidepressants (see Table 397-5) are often first-line agents of choice for primary anxiety disorders.
†Target doses in the elderly may be lower.
qhs = at bedtime.

TREATMENT

Rx

Empirical evidence from controlled trials demonstrates the efficacy of cognitive-behavioral psychotherapies for most of the anxiety disorders. Such therapies, which use the principles of learning theory to extinguish unhelpful behavior and positively reinforce more functional behavior, help the patient learn to identify and correct the dysfunctional patterns of thinking ("automatic thoughts") that underlie or trigger the cognitive-physiologic cascade of pathologic anxiety responses. Cognitive behavioral therapy may be used as sole therapy, particularly for specific phobias, or in combination with pharmacotherapy. Frequently, cognitive behavioral therapy may be administered as part of family therapy (e.g., to help family members avoid behavior that inadvertently reinforces the patient's symptoms) or in group therapy settings.

Although anxiolytic drugs such as the benzodiazepines (Table 397-9) will usually relieve acute anxiety symptoms, concerns about their long-term efficacy and side effects (e.g., risk for abuse, risk for neurocognitive impairment or falls) make antidepressant medications the more attractive pharmacologic agents for most anxiety disorders (see Table 397-5). Most antidepressants, with the probable exception of bupropion, are helpful for panic disorder, generalized anxiety disorder, and social phobia.

PROGNOSIS

In general, most persons with ongoing anxiety disorders tend to have a chronic course of waxing and waning symptoms. Maintenance therapies should often be used for patients with more chronic anxiety disorders, although evidence to support long-term therapies is not as robust as for mood and psychotic disorders.

Obsessive-Compulsive Disorder

The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) has created a new category of "Obsessive-Compulsive and Related Disorders" in recognition that obsessive-compulsive disorder (OCD) has a distinct pathogenesis from other anxiety disorders. OCD is likely related to other conditions such as body dysmorphic disorder, hoarding disorder, trichotillomania (hair-pulling), and excoriation (skin-picking) disorder.

Patients with OCD have recurrent obsessions or compulsions (Table 397-10), and most patients have both.[10] OCD should not be confused with obsessive-compulsive personality traits or disorder, described later under "Personality Disorders." Obsessions, not to be confused with obsessing (ruminating) on a topic, are recurrent, persistent, and typically distressing thoughts that at some point during the course of the disorder are experienced as intrusive and unwanted. The latter quality may be described in language such as "I don't know where this thought comes from" or "I don't know why I have this thought, I would never actually do such a thing!" Compulsions are repetitive behaviors or mental acts the individual feels driven to perform in response to an obsession or according to rigid rules. For example, compulsive handwashing may relate to obsessional thoughts about germs or contamination. Patients with OCD typically attempt to ignore, suppress, or neutralize their obsessions, but doing so causes great psychic distress. OCD patients may spend many hours per day related to their obsessions and compulsions.

TABLE 397-10 COMMON TYPES OF OBSESSIONS AND COMPULSIONS IN OBSESSIVE-COMPULSIVE DISORDER

OBSESSIONS
Aggressive (fears of harming self or others, of blurting out obscenities, or of other unwanted aggressive acts; unwanted violent or horrific images)
Contamination (concerns about dirt, germs, body waste or secretions, environmental contaminants, or animals/insects)
Sexual (concerns about unwanted sexual images or impulses)
Hoarding/saving
Religious (scrupulosity) (excessive concerns about sacrilege, blasphemy, right/wrong, morality)
Need for symmetry/exactness
Somatic (excessive concern about illness, body part, or appearance)
COMPULSIONS
Cleaning/washing (excessive or ritualized handwashing, showering, or other grooming)
Checking (checking locks, stove, appliances; checking body in relation to somatic obsessions; checking that did not or will not harm self or others)
Repeating rituals (rereading or rewriting; routine activities such as going through a door or arising from a chair)
Counting
Ordering/arranging
Hoarding/saving

Adapted from Goodman WK, Price LH, Rasmussen SA, et al. The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Arch Gen Psychiatry.* 1989;46:1006-1011.

The 12-month prevalence of OCD is approximately 1%, with onset typically in childhood, adolescence, or young adulthood. Remission rates are low in adults, with most persons experiencing a chronic waxing and waning course. Pathogenesis probably involves altered functioning of the striatofrontal systems, as well as a prominent role for central serotonergic systems. Obsessions and compulsions may represent inappropriate triggering of neural "scripts" involving thoughts and behaviors that have been analogized to the scripts involved in animal grooming and other complex but stereotypical behaviors.

The only efficacious antidepressants in OCD are those with strong activity on serotonergic systems, such as the selective serotonin reuptake inhibitors and the tricyclic compound clomipramine. Cognitive-behavioral therapies

also have well-demonstrated efficacy. Deep brain stimulation (DBS)[11] targeting the ventral capsule/ventral striatum is FDA approved (as a humanitarian device exemption) for severe treatment-refractory OCD. Although early studies are relatively encouraging, the role of DBS in clinical practice remains to be defined.

Acute Stress Disorder and Post-traumatic Stress Disorder

Acute stress disorder and post-traumatic stress disorder (PTSD) are specific manifestations of symptoms referable to an extremely traumatic event. The event by definition must involve exposure to actual or threatened death, serious injury, or sexual violence, as reported directly by the patient or by family members or friends. Patients suffer from repeated or extreme exposure to aversive details of the event. It is important to recognize that acute stress disorder or PTSD does not develop in all individuals exposed to a common traumatic event (e.g., a natural or man-made disaster). Some individuals may instead develop other anxiety disorders, major depression, mania, or psychosis, and diagnosable psychopathology may never develop at all in some or many others.

PTSD symptoms by definition persist for more than 1 month after the traumatic event and include the following types of clinical phenomena: (1) intrusion, such as intrusive memories, dreams, flashbacks, or intensely distressing psychological or physiologic responses to reminders of the trauma; (2) avoidance of distressing memories or external reminders of the trauma; (3) negative cognitions and mood, such as amnesia for aspects of the event, negativistic thoughts about oneself in general or blame related to the event, persistent negative emotions, diminished interests or activities, or feelings of detachment; and (4) alterations in arousal and reactivity. Acute stress disorder by definition resolves in less than 1 month, with symptoms of intrusion, avoidance, or arousal as well as negative mood or dissociative symptoms (e.g., "in a daze").

The 12-month prevalence of PTSD in the United States is about 3%, with projected lifetime risk approaching 9%. About half of adults with PTSD have complete recovery within 3 months, but PTSD may persist for many months or years. Both cognitive-behavioral and psychodynamic psychology perspectives are useful in informing psychotherapeutic treatments.[12] Antidepressants also have demonstrated efficacy in PTSD.

PSYCHOTIC DISORDERS

Psychotic symptoms, defined as a loss of reality testing, include delusions (fixed false beliefs), hallucinations (false sensory perceptions), and major derailments in thought processes (e.g., loose associations). Psychotic symptoms may be seen in the course of neurocognitive, secondary, and mood disorders. The psychotic disorders are defined by the presence of psychotic symptoms in the absence of prominent mood disturbance, or of neurocognitive deficits at the level seen in delirium or dementia. In general, the diagnosis and treatment of patients with psychotic disorders should be conducted in mental health specialty settings, but primary care settings are common points of entry to care.

Schizophrenia

DEFINITION AND EPIDEMIOLOGY

Schizophrenia, the prototypical psychotic disorder, necessarily includes symptoms of psychosis ("positive" symptoms) and also often includes "negative symptoms" such as affective flattening, abulia, apathy, and social withdrawal. The level of functioning is impaired in one or more realms (e.g., occupational, interpersonal, or self-care). The lifetime prevalence of schizophrenia is slightly less than 1%, and its chronic debilitating course takes a considerable toll on patients, families, and society. Peak onset is in late adolescence to young adulthood, slightly younger for males than females. The annual incidence is approximately 15 per 100,000, but with marked variability across study samples and populations. When narrowly defined as above, the condition is slightly more common in males than in females.

PATHOBIOLOGY

The pathogenesis of schizophrenia remains unknown. Twin studies show that the disease is multifactorial. Genetic factors account for up to 50% of the risk, and multiple gene loci appear to be involved. Studies of postmortem brains indicate a nongliotic neuropathologic process with subtle disruptions of cortical cytoarchitecture. It is likely that psychosocial factors and neurodevelopment interact with a nonlocalizable brain "lesion" that is either present at birth or acquired early in life. The dopaminergic mesocortical and mesolimbic pathways are important in the production of psychotic symptoms.

TABLE 397-11 SYMPTOMS AND SIGNS OF MAJOR PSYCHOTIC DISORDERS

SCHIZOPHRENIA
Delusions Hallucinations Disorganized speech (i.e., thought process derailments) Grossly disorganized or catatonic behavior Negative symptoms: affective flattening, alogia, avolition Major impairment in social or occupational functioning Duration of at least 6 months
SCHIZOAFFECTIVE DISORDER
During the course of illness, at least one episode of schizophrenia-like psychotic symptoms *PLUS* a mood syndrome (either major depression or mania) *AND* During the course of illness, at least 2 weeks of schizophrenia-like psychotic symptoms *in the absence of* a mood syndrome
DELUSIONAL DISORDER
One or more delusions for at least 1 month, most often nonbizarre (i.e., potentially plausible, such as delusions of being followed, poisoned, infected, loved at a distance, deceived by a spouse or lover, or having a disease) *Not* meeting full criteria for an acute episode of schizophrenia Functioning *not* markedly impaired other than as related to the impact of the delusion(s) and its ramifications

Based on criteria from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* 5th ed. (DSM-5) Washington, DC: American Psychiatric Association; 2013.

DIAGNOSIS

The diagnosis of schizophrenia is based on the presence of delusions, hallucinations, and disorganized speech and behavior, often accompanied by apathy and social withdrawal and resulting in major impairment in functioning for at least 6 months (Table 397-11). In patients with single schizophrenia-like psychotic episodes of briefer duration, with subsequent return to asymptomatic baseline functioning, brief psychotic disorder (<1 month) or schizophreniform disorder (1 to 6 months) may be diagnosed.

TREATMENT

Rx

Antipsychotic medications (Table 397-12), often with adjunctive benzodiazepines, are used to treat acute psychotic episodes, commonly in acute inpatient settings so that the patient can be managed safely until the acute symptoms improve.[13] Although maintenance antipsychotic medications help reduce the severity and frequency of acute psychotic episodes, comprehensive psychosocial rehabilitation programs are required to help patients manage interpersonal and other stressors and to improve overall clinical outcomes. Second-generation ("atypical") antipsychotic medications have replaced first-generation antipsychotics in common U.S. practice because of their lower rates of extrapyramidal side effects, including tardive dyskinesia, although their efficacy is better than that of first-generation drugs. However, second-generation drugs contribute to the increase in obesity and metabolic syndrome in patients with chronic schizophrenia (Chapter 434). One trial found that oral long-chain ω-3 polyunsaturated fatty acids reduced the rate of onset of psychosis by more than four fifths during a 1-year follow-up of adolescents and young adults with especially high-risk profiles for incipient psychosis. Unfortunately, programs that include compulsory supervision to ensure adherence to outpatient medication regimens have not reduced subsequent readmission rates for psychotic patients.

PROGNOSIS

The prognosis of individuals with schizophrenia is often poor, with recurrent episodes of psychotic exacerbations superimposed on progressively deteriorating baseline functioning. However, antipsychotic drugs significantly reduce relapse rates at 1 year from 64 to 27%. Some patients have a more favorable course, and a small number of individuals may recover completely. Male sex, prominent negative symptoms, younger age at first onset, and enduring psychosocial stressors and family discord all predict poorer outcomes. Although many patients with schizophrenia survive into later life, overall life expectancy is shortened by at least 10 to 15 years because of poor health behaviors, higher rates of other medical disorders including metabolic syndrome, and a lifetime suicide risk of approximately 5 to 6%.

TABLE 397-12 COMMONLY USED ANTIPSYCHOTIC MEDICATIONS

DRUG NAME	INITIAL DOSE FOR PSYCHOSIS IN SCHIZOPHRENIA*	TARGET DOSE FOR PSYCHOSIS IN SCHIZOPHRENIA†	SIDE EFFECTS	CHLORPROMAZINE DOSAGE EQUIVALENCE (FIRST-GENERATION DRUGS ONLY)/ OTHER COMMENTS
First-generation drugs			Low-potency drugs: anticholinergic effects, orthostatic hypotension, prolongation of QT interval, cholestatic jaundice High-potency drugs: extrapyramidal side effects (dystonias, akathisia, parkinsonism, neuroleptic malignant syndrome), hyperprolactinemia with galactorrhea	
Chlorpromazine	100 mg daily	300-1000 mg/day, daily-bid dosing		100 mg
Thioridazine	50-100 mg daily	300-800 mg/day, daily-bid dosing	Pigmentary retinopathy at higher doses	100 mg
Thiothixene	2-5 mg daily	5-60 mg/day, daily-bid dosing		5 mg
Trifluoperazine	2-5 mg daily	5-40 mg/day, daily-bid dosing		5 mg
Perphenazine	4-8 mg daily	8-64 mg/day, daily-tid dosing		8 mg
Haloperidol	0.5-2 mg daily	2-10 mg/day (up to 40 mg/day or higher in refractory cases), daily-bid dosing		2 mg; available in depot IM form
Fluphenazine	1-2.5 mg daily	2.5-10 mg/day (up to 40 mg/day in refractory cases), daily-bid dosing		2 mg; available in depot IM form
Second-generation drugs			Metabolic syndrome, risk for stroke and mortality in older patients with dementia, QT prolongation	
Risperidone	0.5-1 mg daily-bid	2-4 mg/day, daily-bid dosing	Extrapyramidal side effects at higher doses	Available in depot IM form
Olanzapine	5 mg daily	5-10 mg daily (up to 20 mg/day in refractory cases)		
Ziprasidone	20 mg bid	20-80 mg bid		
Quetiapine	25-50 mg bid-tid	300-800 mg/day, bid-tid dosing		Extended-release form for daily dosing
Asenapine	5 mg bid	5-10 mg bid		Sublingual form only
Paliperidone	3-6 mg daily	6-12 mg daily		
Iloperidone	1 mg bid	2-12 mg bid		
Lurasidone	40 mg daily	40-160 mg daily		
Aripiprazole	10-15 mg daily	10-30 mg daily		Partial agonist/antagonist at D_2 receptors
Clozapine	12.5 mg daily-bid	300-900 mg/day, daily-bid (titrate dose slowly by 25-50 mg/day every 3-7 days)	Risk for agranulocytosis, requires ongoing monitoring of complete blood count	Efficacy superior to that of other antipsychotics, but hematologic risks and need for monitoring limit its use

*Doses for other indications, such as agitation in delirium or dementia, may be much lower.
†Target doses in the elderly may be lower.

Schizoaffective Disorder

Schizoaffective disorder is a chronic recurrent disorder with a lifetime prevalence of approximately 0.3%. It is characterized by episodes of psychosis in the absence of mania or depression, and also by mood episodes (manic or depressed) with psychotic features. As a result, the diagnosis of schizoaffective disorder requires knowledge of the patient's course over time and cannot be based on the patient's clinical findings at any one point in time. Treatment is symptomatic and involves the use of antipsychotic medications (see Table 397-12), mood stabilizers (see the Treatment box for bipolar disorders), and antidepressant medications (see Table 397-5) to target specific psychotic and mood symptoms. The outcomes of schizoaffective disorder are heterogeneous but on average intermediate between those of schizophrenia and mood disorders.

Delusional Disorder

Delusional disorders are characterized by one or more delusions in the absence of a thought process disorder, prominent hallucinations, or the negative symptoms seen in schizophrenia. The most characteristic types of delusions are potentially plausible ("nonbizarre"), such as unfounded beliefs of a partner's infidelity. Delusional disorder has a lifetime prevalence of approximately 0.2%. The pathogenesis of delusional disorder remains largely unknown. It is often only partially responsive to antipsychotic medications (see Table 397-12), but patients' functioning may be largely unimpaired if they are able, with the aid of antipsychotics and psychotherapy, to avoid acting on their delusions.

SOMATIC SYMPTOM AND RELATED DISORDERS

Formerly termed *somatoform disorders*, the somatic symptom disorders include both somatic symptoms and associated thoughts, feelings, or behaviors that are distressing and disabling (Table 397-13). Although identifiable physical disease is insufficient to explain the patient's presentation fully, in all these conditions (other than factitious disorder) the patient's distress and dysfunction are *not* consciously produced and thus are just as distressing and baffling to patients as would be similar symptoms produced by physical disease. Malingering is the conscious feigning of illness for conscious gain and is therefore not in this category; indeed, malingering is not considered to be a mental disorder at all.

TREATMENT

Rx

Management of patients with somatic symptom disorders is often difficult because physicians must simultaneously maintain an appropriate level of vigilance for undiagnosed physical illness while avoiding unnecessary tools and therapies. Keys to ongoing care include maintaining an ongoing therapeutic alliance, setting regular office visits, conveying empathy for the patient's very real distress without colluding with the patient's belief in an identifiable physical disorder, and assertively treating depression, anxiety, or other comorbid psychopathology. Antidepressant medications may benefit selected patients (e.g., some chronic pain syndromes), even in the absence of comorbid psychiatric disorders.

TABLE 397-13 SOMATIC SYMPTOM & RELATED DISORDERS

TYPE	MAIN CLINICAL MANIFESTATIONS
Somatic symptom disorder	One or more distressing somatic symptoms, together with excessive thoughts, feelings, or behaviors related to these symptoms; subsumes most of the former terms somatization disorder, pain disorder, undifferentiated somatoform disorder, and many with the former diagnosis of hypochondriasis.
Illness anxiety disorder	Illness preoccupation and excessive health-related behaviors, in the absence of or disproportionate to somatic symptoms; subsumes some patients with the former diagnosis of hypochondriasis.
Conversion disorder (functional neurologic symptom disorder)	Neurologic somatoform symptoms (other than pain) with clinical evidence incompatible with recognized neurologic or general medical conditions (e.g., paralysis, blindness, dyscoordination, convulsion-like phenomena, memory or other neurocognitive complaints)
Psychological factors affecting other medical conditions	Psychological factors adversely affecting a (non-mental disorder) medical symptom or condition by worsening the course, interfering with treatment, adding to known health risks, or influencing underlying pathophysiology
Factitious disorder (commonly called Munchausen)	Falsification of physical or psychological signs or symptoms, with health- or help-seeking behaviors, in the absence of clear external rewards

Author summary based on criteria from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. (DSM-5) Washington, DC: American Psychiatric Association; 2013.

TABLE 397-14 PERSONALITY DISORDERS

TYPE OF PERSONALITY DISORDER	MAIN IDENTIFYING CHARACTERISTICS
CLUSTER A: ODD/ECCENTRIC	
Schizoid personality disorder	Detachment from social relationships, restricted emotional expression
Schizotypal personality disorder	Discomfort with close relationships, cognitive or perceptual distortions, eccentric behavior
Paranoid personality disorder	Pervasive distrust and suspiciousness of others' motives as malevolent
CLUSTER B: DRAMATIC/EMOTIONAL/ERRATIC	
Borderline personality disorder	Instability of interpersonal relationships, self-image, and affects, and marked impulsivity
Narcissistic personality disorder	Grandiosity, need for admiration, and lack of empathy
Antisocial personality disorder	Pervasive disregard for and violation of the rights of others, lack of true remorse ("conscience")
Histrionic personality disorder	Pervasive excessive emotionality (theatricality) and attention seeking
CLUSTER C: ANXIOUS/FEARFUL	
Avoidant personality disorder	Social inhibition, feelings of inadequacy, and sensitivity to negative views from others
Dependent personality disorder	Pervasive and excessive need to be taken care of, resulting in submissive and clinging behavior and fears of separation
Obsessive-compulsive personality disorder	Pervasive preoccupation with orderliness, perfectionism, and mental and interpersonal control

Author summary based on criteria from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. (DSM-5) Washington, DC: American Psychiatric Association; 2013.

PERSONALITY DISORDERS

Personality is defined as the repertoire of enduring patterns of inner mental experience and behavior, including affect and impulse regulation, defense and coping mechanisms, and interpersonal relatedness.[14] Dimensional models of personality (i.e., using multiple continuous measures of constructs such as neuroticism, extraversion, and openness to experience) likely are a more accurate representation of the spectrum of human personality, but categorical diagnostic categories (i.e., personality disorders) are more useful for clinicians to determine prognosis and treatments. Personality and personality disorders are the result of complex interactions among genetic, environmental, and developmental factors. The cumulative point prevalence of all personality disorders in the general adult population is approximately 10 to 15%, with rates as high as 50% in patients receiving care in psychiatric treatment settings.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

A personality disorder is diagnosed when enduring personality traits lead to pervasive (if variable) distress or dysfunction in a broad range of personal and social situations (Table 397-14). In diagnosing personality disorders, care must be taken to distinguish personality *traits*, which by definition are enduring, from time-limited *states*. Most persons can regress to more primitive personality styles not characteristic of their baseline personality traits under the influence of substantial psychosocial stressors.

TREATMENT

Rx

In most circumstances, the goal is not to alter fundamental personality structure but rather to help the patient maximize use of their personality strengths (e.g., optimal defense and coping mechanisms) while minimizing the harmful effects of emotional dysregulation, dysfunctional defenses, and destructive behavior. Dialectic behavior therapy is an evidence-based, focused psychotherapy that is based on specific cognitive-behavioral techniques and has been demonstrated to reduce self-injurious behavior and suicidality in patients with borderline personality disorder.

Although pharmacotherapy is not the mainstay of treatment of most personality disorders, drugs can be useful in selected patients. Antipsychotic drugs may be used to target escalating paranoia in paranoid personality disorder or for short-term reduction in emotional and impulse regulation with a wide range of (often cluster B, see Table 397-14) personality disorders in times of crisis. For longer-term treatment of emotional dysregulation in borderline and other cluster B personality disorders, mood stabilizers or antidepressants may be used.

SUICIDE AND EVALUATION OF SUICIDALITY

Suicide is a leading cause of death worldwide. Suicide rates in the United States average approximately 11 per 100,000 per year, with considerable variability geographically and demographically. Of all age-, gender-, and race-based demographic groups, the highest U.S. suicide rates occur in older white men, while suicide is the third leading cause of death in adolescents and young adults and the tenth leading cause of death in the population overall. Suicide attempts, which outnumber completed suicides by a factor of up to 11 : 1, lead to considerable morbidity and utilization of health care resources. Persons who attempt suicide represent an overlapping but distinct population from those who die by suicide. Nonetheless, a previous history of a suicide attempt is a powerful risk for subsequent death by suicide. Suicide attempts and verbal threats should always be evaluated carefully and never dismissed as "gestures" or "attention-seeking" behavior.

Suicide is a potentially preventable cause of death, but despite considerable research on risks for suicidal behavior, specific predictions about an individual's behavior cannot be made with certainty. Nonetheless, the linchpin of clinical evaluation is a methodical assessment of risks for suicide (Table 397-15), together with direct questioning of the patient regarding thoughts of death, dying, and suicide; specific plans (in ideation or action) for suicide; and the details of any attempts. Patients at significantly increased risk for suicide should immediately be referred for psychiatric evaluation, with emergency referral if the risk is deemed to be imminent or increasing.

WHEN TO REFER A PATIENT FOR PSYCHIATRIC EVALUATION

Clinical decisions to refer a patient for specialty psychiatric evaluation must be made on an individual basis by taking into account the patient's clinical findings, including any previous history and immediate needs, and the

TABLE 397-15 SOME IMPORTANT RISKS FOR SUICIDE AND SUICIDE ATTEMPTS

Mental disorder, particularly depression, bipolar, substance use, psychotic, and personality disorders
Other symptoms of acute psychic distress, particularly hopelessness and panic attacks
Previous history of suicide attempt
Family history of suicide or suicide attempt (and, to a lesser degree, of any mental disorder)
Family violence, including physical or sexual abuse
Access to firearms or other lethal methods
Incarceration
Exposure to suicidal behavior of others (family, peers, public figures)
Social isolation
Interpersonal discord or other psychosocial stressors
Demographic factors, including male, non-Hispanic white or American Indian/Alaska Native race, older age

TABLE 397-16 GENERAL CONSIDERATIONS IN DECIDING TO REFER A PATIENT FOR PSYCHIATRIC SPECIALTY CARE

Diagnosis or ongoing care of severe/chronic mental disorders, including bipolar disorder, psychotic disorders such as schizophrenia, and psychotic symptoms in other disorders
Management of more severe forms of other mental disorders and those refractory to treatment, including depression, anxiety disorders, and substance use disorders
Need for safety evaluation or management, including suicidality, homicidality or other aggressivity, or inability to care for self
Diagnostic uncertainty
Psychiatric comorbid conditions complicating diagnosis or treatment, including personality and substance use disorders coexisting with other psychiatric disorders
Psychiatric-medical comorbid conditions complicating diagnosis or treatment, including management of psychiatric disorders during pregnancy
Need for expertise in psychopharmacologic treatment
Need for expertise in other somatic therapies (e.g., electroconvulsive therapy, light therapy)
Need for expertise in psychotherapy or other psychosocial interventions

clinician's own experience and expertise in assessing and managing the disorder (Table 397-16).

Grade A References

A1. Bortolotti B, Menchetti M, Bellini F, et al. Psychological interventions for major depression in primary care: a meta-analytic review of randomized controlled trials. *Gen Hosp Psychiatry*. 2008;30:293-302.

A2. Cuijpers P, Sijbrandij M, Koole SL, et al. Adding psychotherapy to antidepressant medication in depression and anxiety disorders: a meta-analysis. *World Psychiatry*. 2014;13:56-67.

A3. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet*. 2009;373:746-758.

A4. Gartlehner G, Hansen RA, Morgan LC, et al. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. *Ann Intern Med*. 2011;155:772-785.

A5. Wijkstra J, Lijmer J, Burger H, et al. Pharmacological treatment for psychotic depression. *Cochrane Database Syst Rev*. 2013;11:CD004044.

A6. Murrough JW, Iosifescu DV, Chang LC, et al. Antidepressant efficacy of ketamine in treatment-resistant major depression: a two-site randomized controlled trial. *Am J Psychiatry*. 2013;170:1134-1142.

A7. Woltmann E, Grogan-Kaylor A, Perron B, et al. Comparative effectiveness of collaborative chronic care models for mental health conditions across primary, specialty, and behavioral health care settings: systematic review and meta-analysis. *Am J Psychiatry*. 2012;169:790-804.

A8. BALANCE investigators and collaborators. Lithium plus valproate combination therapy versus monotherapy for relapse in bipolar I disorder (BALANCE): a randomized open-label trial. *Lancet*. 2010;375:385-395.

A9. Cuijpers P, Sijbrandij M, Koole S, et al. Psychological treatment of generalized anxiety disorder: a meta-analysis. *Clin Psychol Rev*. 2014;34:130-140.

A10. Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet*. 2013;382:951-962.

A11. Burns T, Rugkasa J, Molodynski A, et al. Community treatment orders for patients with psychosis (OCTET): a randomised controlled trial. *Lancet*. 2013;381:1627-1633.

A12. Leucht S, Tardy M, Komossa K, et al. Antipsychotic drugs versus placebo for relapse prevention in schizophrenia: a systematic review and meta-analysis. *Lancet*. 2012;379:2063-2071.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*. 2013;381:1371-1379.
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Association; 2013.
3. Taylor WD. Clinical practice. Depression in the elderly. *N Engl J Med*. 2014;371:1228-1236.
4. Kupfer DJ, Frank E, Phillips ML. Major depressive disorder: new clinical, neurobiological, and treatment perspectives. *Lancet*. 2012;379:1045-1055.
5. Kendall T, Morriss R, Mayo-Wilson E, et al. Assessment and management of bipolar disorder: summary of updated NICE guidance. *BMJ*. 2014;349:g5673.
6. Phillips ML, Kupfer DJ. Bipolar disorder diagnosis: challenges and future directions. *Lancet*. 2013;381:1663-1671.
7. Geddes JR, Miklowitz DJ. Treatment of bipolar disorder. *Lancet*. 2013;381:1672-1682.
8. Baxter AJ, Vos T, Scott KM, et al. The global burden of anxiety disorders in 2010. *Psychol Med*. 2014;44:2363-2374.
9. Herr NR, Williams JW Jr, Benjamin S, et al. Does this patient have generalized anxiety or panic disorder?: The Rational Clinical Examination systematic review. *JAMA*. 2014;312:78-84.
10. Grant JE. Clinical practice: Obsessive-compulsive disorder. *N Engl J Med*. 2014;371:646-653.
11. Lipsman N, Giacobbe P, Lozano AM. Deep brain stimulation in obsessive-compulsive disorder: neurocircuitry and clinical experience. *Handb Clin Neurol*. 2013;116:245-250.
12. Tol WA, Barbui C, van Ommeren M. Management of acute stress, PTSD, and bereavement: WHO recommendations. *JAMA*. 2013;310:477-478.
13. Kuipers E, Yesufu-Udechuku A, Taylor C, et al. Management of psychosis and schizophrenia in adults: summary of updated NICE guidance. *BMJ*. 2014;348:g1173.
14. Gask L, Evans M, Kessler D. Clinical Review. Personality disorder. *BMJ*. 2013;347:f5276.

REVIEW QUESTIONS

1. Which of the following types of disorders may manifest with a combination of substantial intellectual deficits, mood symptoms, psychotic symptoms, and anxiety symptoms?

A. Anxiety disorders
B. Bipolar disorders
C. Depressive disorders
D. Neurocognitive disorders
E. Psychotic disorders

Answer: D See Table 397-2. Although the hallmark of neurocognitive disorders is intellectual deficits, they often manifest with symptoms affecting other parts of the mental status examination, including mood, psychotic, and anxiety symptoms. Intellectual deficits consistent with delirium or dementia are not characteristic of anxiety, bipolar, depressive, or psychotic disorders.

2. After successful resolution of a first episode of major depression with antidepressant medication therapy, how long should the medication be continued?

A. Not at all; discontinue as soon as the episode has resolved
B. 1 week
C. 1 month
D. At least 4 to 8 months
E. Indefinitely for lifelong maintenance therapy

Answer: D As noted in the section on treatment of major depression, continuation of treatment is required to prevent relapse into the depressive episode. The continuation phase should be at least 4 to 8 months, and some experts recommend up to 1 year. However, assuming the patient's symptoms remain in full remission at the end of the continuation phase, lifelong maintenance therapy is indicated only for patients with recurrent episodes of depression (and, some believe, for others at particularly high risk for recurrence or for the destructive effects of another depressive episode, e.g., late-onset depression, highly suicidal depression).

3. For patients who present with a major depressive episode, it is important to inquire about any prior history of hypomania, because such a history has which of the following implications for treatment?

A. Need to treat all first-degree relatives prophylactically
B. Need to obtain informed consent from the nearest relative
C. More likely to respond to antidepressant medication
D. More likely to respond to psychotherapy
E. May require mood stabilizer treatment prior to antidepressant therapy

Answer: E Please see text regarding bipolar II disorder under "Other Mood Disorders." Antidepressant therapy in this condition may precipitate hypomania (or mania). At the least, extremely careful monitoring is required, and many experts recommend not starting antidepressant therapy at all until reevaluating the patient after institution of treatment with a mood stabilizer such as lithium or valproate.

4. Which of the following disorders has a pathophysiology that is probably mediated primarily by striatofrontal systems?

A. Generalized anxiety disorder
B. Obsessive-compulsive disorder
C. Panic disorder
D. Posttraumatic stress disorder
E. Social anxiety disorder

Answer: B As discussed in the text under anxiety disorders and other disorders with prominent anxiety, obsessive-compulsive disorder has been reclassified separate from the anxiety disorders because of its differing pathophysiology. Post-traumatic stress disorder also has been classified separately (under trauma- and stress-related disorders), but its neurobiology appears to be more closely related to the anxiety disorders than to obsessive-compulsive disorder.

5. Which of the following predicts better longer-term outcome in schizophrenia?

A. Absence of negative symptoms such as apathy and social withdrawal
B. Enduring family discord
C. Being male
D. Ongoing psychosocial stressors
E. Younger age of onset

Answer: A As discussed in the section on schizophrenia, each of factors B through E predicts a poorer outcome. Prominent negative symptoms predict a poorer course, as reflected in the poor response of negative symptoms to antipsychotic medications (hence the need for psychosocial rehabilitation programs in this condition). The absence of negative symptoms suggests higher functioning if the positive (psychotic) symptoms of the disorder can be controlled with medication.

398 HEADACHES AND OTHER HEAD PAIN

KATHLEEN B. DIGRE

DEFINITION

Headache, which is a very common symptom, can be secondary to a serious underlying abnormality but is usually a primary headache disorder such as migraine headache, tension-type headache, cluster headache, and paroxysmal hemicrania.[1]

EPIDEMIOLOGY

About 90% of all adults experience headache at some time in their lives, and over 75% of children have complained of headaches by the age of 15 years. In the United States, the direct and indirect costs associated with migraine are over $20 billion annually. Patients at most risk for lost days of employment are those with transformed migraine and daily headache.

In large population-based studies, the relative risk of having migraine, tension-type headaches, or cluster headaches increases up to four times if a first-degree relative has the same kind of headaches. Studies of twins, especially identical twins, also show a similar susceptibility.

PATHOBIOLOGY

Headache pain is initiated by primary trigeminal afferents that innervate the blood vessels, mucosa, muscles, and tissues. Fibers from these sources coalesce in the trigeminal ganglion, especially the first division. The trigeminal afferents terminate in the primary sensory nucleus of cranial nerve V and its spinal nucleus, which has several small subnuclei, the most important of which is the subnucleus caudalis. This subnucleus receives afferents from meningeal vessels, dura-sensitive neurons, and even the upper cervical cord and then projects them to the lateral and medial thalamus by way of the spinothalamic tract and to diencephalic and brain stem regions that are involved in the regulation of autonomic functions. Thalamic nociceptive information ascends to the sensory cortex, as well as to other areas of the brain.

Although secondary headaches may stimulate the pathway by way of processes such as inflammation and compression, primary headache disorders occur spontaneously by means of chemical mediators. The sequence of events commences with peripheral activation caused by neurogenic plasma extravasation activated spontaneously or by cortical spreading depression. The trigeminocervical complex, especially the nucleus caudalis, is then activated, and patients can experience allodynia, a condition in which a nonnoxious stimulus is sensed as painful.

Aura is defined as a focal visual, sensory, or motor neurologic disturbance that may occur with or without headache. Aura is thought to occur when cortical spreading depression causes depolarization of membranes. Both neurons and glia can cause both constriction and dilation of blood vessels. Migraine headache clearly has a genetic component. Familial hemiplegic migraine can be caused by mutations in the *CACNA1A* gene, which is located on chromosome 19p13.2-p13.1 and encodes for voltage-gated neuronal calcium channels. Mutations in the *CACNA1A* gene also cause episodic ataxia and epilepsy. Another mutation is in *ATP1A2*, also called the familial hemiplegia migraine 2 (*FHM2*) gene, which is located on chromosome 1q21-q23 and encodes for the sodium-potassium adenosine triphosphatase (Na^+,K^+-ATPase) transport protein. A third genetic locus is the *SCN1A* gene on chromosome 2q24.3, which is a voltage-gated sodium channel. In addition, many single nucleotide polymorphisms have been associated with migraine. Although there are linkages to many genetic loci for more common forms of migraine, migraine and other headaches probably have multiple gene interactions with environmental factors, and it is clear that the genetic contributions are complex.

CLINICAL MANIFESTATIONS

Patients with headache may describe the pain as throbbing, bandlike, or aching. The pain is frequently unilateral but can be bilateral. Migraine headache is often associated with nausea, vomiting, photophobia, and phonophobia. It is invariably moderate to severe and interferes with activities. Other autonomic manifestations that can accompany migraine, cluster, and other headache variants include ptosis, conjunctival injection, tearing, rhinorrhea, Horner syndrome, and facial edema. Secondary headaches sometimes may appear to be similar to tension-type or migraine headaches, but "red flags" may suggest a secondary rather than a primary headache disorder (Table 398-1). Particular attention should be paid to the sudden onset of severe headaches, which frequently have an underlying secondary cause.[2]

DIAGNOSIS

Evaluation of an individual's headache is five elements of the history. The family history helps determine whether a person has a genetic predisposition to headache. The life history of headache determines whether the headache is new or has evolved over the course of a lifetime. The attack history provides the clinical features of the headache or headaches. The medical and psychiatric history determines whether there are comorbid conditions that can cause or worsen the headache. The medication and drug history determines whether the headache could be caused by or worsened by medications or drugs the person has ingested.

Diagnosis of the type of headache is based on the type of pain, the duration of headache, and accompanying features (Table 398-2). Secondary headaches are usually due to an underlying condition such as a brain tumor (Chapter 189), increased or low intracranial pressure, sinus disease (Chapter 426), or a vascular malformation (Chapter 408); on removing the cause, the headache generally improves. Headaches that occur at a frequency of less than 15 days a month are called episodic, whereas headaches that occur more than 15 days a month are considered chronic.

The diagnostic evaluation for headache depends on the clinical findings. If there is a typical history without any reason for further diagnostic evaluation and if the findings on neurologic examination are completely normal, no further evaluation is needed. The features of the history that are most likely to predict migraine headache without a secondary disorder include a pulsating quality, duration of 4 to 72 hours, unilateral location, nausea, and disabling nature. However, if there are atypical features of the history or any abnormality on neurologic examination, further evaluation is indicated. Patients with cluster headache types and headaches of undetermined cause need imaging to exclude secondary causes.[3]

In patients with acute headache, computed tomography (CT) is best for assessing acute hemorrhage as the cause of the headache, whereas magnetic resonance imaging (MRI) is best for assessing most persistent headaches to look for mass lesions, evidence of intracranial hypertension or hypotension, hemosiderin (old hemorrhage), and congenital abnormalities (e.g., Chiari malformation). In individuals older than 60 years with an unexplained new or unusual headache, the erythrocyte sedimentation rate (ESR) and/or C-reactive protein (CRP) level should be measured to evaluate for giant cell arteritis (Chapter 78). Cerebrospinal fluid (CSF) analysis, including opening pressure, protein, glucose, cells, culture, and cytology, is indicated in patients with suspected intracranial hypertension or meningitis.

TABLE 398-1 REASONS FOR FURTHER EVALUATION TO LOOK FOR SECONDARY HEADACHES

Beginning of headaches at an older age, without a previous history or a positive family history
Unexplainable and abnormal worsening of previously existing migraines
Dramatic or unusual change in character of the prodrome or the headache previously present
Headaches awakening the patient in the middle of the night (except for a cluster headache)
Headaches much worse when recumbent or with coughing, sneezing, or the Valsalva maneuver
Unusually severe headache of sudden onset ("worst headache of my life")
Focal deficits that do not disappear after the headache is over
Any abnormal neurologic or new psychiatric finding on examination
A new headache in a patient with human immunodeficiency virus infection, malignancy, or pregnancy

TREATMENT

Rx

Treatment of acute headache depends on the type and severity of the headache. For mild headaches, simple analgesics such as acetaminophen (500 to 1000 mg), acetaminophen with caffeine, aspirin (250 to 1000 mg), and nonsteroidal anti-inflammatory drugs (NSAIDs, e.g., ibuprofen, 400 to 800 mg; naproxen sodium, 220 to 500 mg) will suffice.

PREVENTION

Preventive medications are recommended when headaches are frequent or severe enough to interfere with quality of life. The choice of medications should be based on the type of headache (migraine, tension type), their side-effect profiles, and the patient's comorbid conditions (Table 398-3).

PROGNOSIS

The natural history of headache depends on many factors, including the type of headache, comorbid conditions that accompany the headache, and success of treatment. Risk factors for chronic headache include female sex, migraine-type headaches, frequent headaches, obesity, low education and socioeconomic level, overuse of medication, depression, anxiety, stressful life events, and sleep apnea.

TABLE 398-2 DIFFERENTIAL DIAGNOSIS OF HEADACHE

HEADACHE TYPE	GENETICS	EPIDEMIOLOGY	CHARACTERISTIC FEATURES	LENGTH	ACCOMPANYING SYMPTOMS
Migraine headache	Complex genetics but usually a family history	More frequent in women	Unilateral, bilateral; throbbing; moderate to severe; worsens with activity	Hours to days	Photophobia, phonophobia, nausea and/or vomiting
Tension-type headache	Usually a family history	Equal frequency in men and women	Tight band–like pain; bilateral; pain may be mild to moderate; improves with activity	Hours to days	No nausea or vomiting; small amount of light or sound sensitivity, but not both
Cluster headache	May have a family history	More frequent in men	Unilateral severe pain in the face	Minutes to hours	Ipsilateral ptosis, miosis, rhinorrhea, eyelid edema, tearing
Paroxysmal hemicrania	Usually no family history	More frequent in women	Unilateral pain in the face	Minutes	Ipsilateral ptosis, miosis, rhinorrhea, eyelid edema, tearing; responds to indomethacin
Short unilateral headache with conjunctival injection, tearing	No family history	More frequent in men	Unilateral eye pain; orbit pain	Seconds to 240 seconds	Conjunctival injection, tearing
Hemicrania continua	No family history	More frequent in women	Unilateral continuous headache with episodic stabbing pains	Continuous	Ipsilateral autonomic features: ptosis, miosis, rhinorrhea, eyelid edema, tearing

TABLE 398-3 PREVENTIVE MEDICATIONS FOR HEADACHE

DRUG	RATIONAL USE	DOSAGE	SIDE EFFECTS	CONTRAINDICATIONS/ CAUTION
β-Blockers (e.g., propranolol, nadolol, timolol)	Migraine, anyone with elevated blood pressure	20-80 mg; may increase	Lethargy, depression	Asthma, low blood pressure
Calcium-channel antagonists: verapamil, amlodipine	Cluster headache, elevated blood pressure	Verapamil, 120 to 480 mg/day	Low blood pressure	
Nonsteroidal anti-inflammatory drugs:				
naproxen, ibuprofen	Migraine, tension-type, and menstrual migraine	Naproxen, 200-600 mg/day; ibuprofen, 600-800 mg bid-tid	Gastrointestinal	Ulcers, sensitivity, allergy
Indomethacin	Paroxysmal hemicrania, hemicrania continua	25 mg tid	Gastrointestinal	Ulcers
Tricyclic antidepressants: amitriptyline, nortriptyline, imipramine	Migraine, tension-type headache, anyone with poor sleep	10-25 mg qhs; may increase	Dry mouth, orthostatic hypotension, weight gain	Sensitivity
Anticonvulsants: topiramate valproate	Migraine, cluster headache	Topiramate, 25-50 mg bid; valproate, 250-500 mg bid	Topiramate: weight loss, kidney stones, intra-ocular hypertension Valproate: weight gain	Pregnancy—both U.S. FDA classification D

MIGRAINE HEADACHE

DEFINITION

Migraine is an inherited headache disorder that is typically unilateral but sometimes bilateral, moderate to severe, worsens by routine physical activity, associated with nausea and/or vomiting, and accompanied by photophobia and phonophobia. The headache occurs anytime and persists from 4 to 72 hours. It may occur with or without an aura (a focal neurologic symptom that may be visual, sensory, or motor). Visual auras may have positive (photopsias) and negative (scotomas) features.

EPIDEMIOLOGY

The prevalence of migraine is 15 to 20% in women and 4 to 7% in men. In children the prevalence may be as high as 17% and is equal in boys and girls. At puberty, the prevalence rises in girls and remains higher throughout their lifespan. The highest prevalence occurs between the ages of 25 and 55. Migraine with aura affects 5% of the adult population, and 90% of auras are visual. Migraine is more prevalent in white persons and in those with a lower socioeconomic status or income.

Comorbid conditions that may be associated with migraine headache include epilepsy, stroke, depression, anxiety, myocardial infarction, patent foramen ovale, Raynaud phenomenon, irritable bowel syndrome, and pain disorders such as fibromyalgia. Menstruation and ovulation may increase the frequency of headache.

PATHOBIOLOGY

The aura of a migraine headache is thought to be due in part to cortical spreading depression, which is associated with a brief reduction in blood flow followed by hyperemia. These changes do not seem to correlate with the phase of the headache. Pain occurs when trigeminal afferents of the dura are stimulated.

CLINICAL MANIFESTATIONS

Migraine headache often begins with a prodrome that may persist for hours to days, when patients note difficulty concentrating or fatigue without headache. An aura may or may not occur but is generally present before the headache begins. The headache may be unilateral or bilateral, throbbing, moderate to severe, and worsened with activity. Accompanying clinical features include nausea, vomiting, and sensitivity to light and sound. Other clinical features include neck pain, occasionally dizziness, osmophobia (sensitivity to odors), and difficulty thinking clearly.

The migraine aura is generally visual but can be sensory or include aphasia or vertigo. Migraine aura without headache begins with a neurologic disturbance (e.g., a visual phenomenon), but without a subsequent headache. Although an aura is traditionally thought to precede the headache, it can be present during the headache phase.

DIAGNOSIS

The diagnosis of migraine is based on the history. The differential diagnosis includes tension-type headache, but most moderate to severe headaches are migraine. In patients with a history suggestive of a secondary headache, further evaluation with MRI should be considered (see Table 398-2). However, if the headache is typical of migraine and the findings on neurologic examination are normal, no further studies are needed.

TREATMENT

Rx

Treatment of migraine is divided into treatment of the acute headache and prevention of subsequent migraine attacks. Acute treatment is most effectively accomplished with migraine-specific care: a nonspecific analgesic agent or combination analgesic therapy for milder migraine, and most frequently aggressive migraine-specific therapy for migraine (Table 398-4). For example, mild attacks can generally be treated successfully with over-the-counter analgesics such as acetaminophen (suggested dose, 650 to 1000 mg) or NSAIDs (aspirin, 900 to 1000 mg; ibuprofen, 1000 to 1200 mg; naproxen, 500 to 825 mg; or ketoprofen, 75 mg). If the migraine headaches are moderate to severe, patients benefit from migraine-specific therapies (see Table 398-4) such as triptans (sumatriptan, zolmitriptan, rizatriptan, almotriptan, naratriptan, frovatriptan, and eletriptan), ergotamine (dihydroergotamine, ergotamine tartrate), or isometheptene,[■] and the combination of naproxen plus sumatriptan may be better than either alone[■] (see Table 398-4).

During pregnancy, mild to moderate attacks can be treated with acetaminophen. Moderate headaches may respond to the combination of acetaminophen, isometheptene mucate (a mild vasoconstrictor, 65 mg), and dichloralphenazone (a mild sedative, 100 mg). Antinausea agents include prochlorperazine (10 to 25 mg) and metoclopramide (2.5 to 10 mg).

Stratification of care, including tailoring the treatment according to the type of headache, results in fewer days of disability and use of medications.[■] Which migraine-specific drug will work for any individual patient depends on the patient. It is important to avoid overuse of analgesic and other medications (especially opiates) because overuse can cause chronic daily headache in susceptible individuals. Prompt treatment improves the outcome of headache when compared with late treatment. Contraindications to use of triptans (see Table 398-4) include uncontrolled hypertension, clinical evidence of ischemic heart disease, and Prinzmetal angina.

Opioids such as *N*-acetyl-*p*-aminophenol (APAP) with codeine, or butorphanol, benefit some patients, but meperidine is not effective. Oral opiates should not be used for chronic recurrent, primary headaches, although sometimes opiates (e.g., acetaminophen, 325 mg, with codeine, 30 mg) are often the only option during pregnancy or in patients with severe vascular disease. When opiates are used, caution is required, and the associated risks of rebound headache and dependency must be recognized by both the patient and physician. Barbiturates (with caffeine and aspirin) have not been efficacious in controlled trials but may be helpful in individual patients in whom other migraine-specific drugs cannot be used.

For moderate to severe attacks, options include dihydroergotamine (1 to 2 mg intranasally)[■]; oral, intranasal, or subcutaneous administration of

TABLE 398-4 SPECIFIC TRIPTAN MEDICATIONS FOR THE TREATMENT OF ACUTE MIGRAINE

	SUMATRIPTAN	ZOLMITRIPTAN	NARATRIPTAN	RIZATRIPTAN	ALMOTRIPTAN	FROVATRIPTAN	ELETRIPTAN
Trade name	Imitrex	Zomig	Amerge	Maxalt	Axert	Frova	Relpax
Forms	SC, nasal (NS), oral	Oral: tablet/ZMT, NS	Oral	Oral: tablet/MLT	Oral	Oral	Oral
Dose	Oral: 50-100 mg (200 mg/24 hr max.) SC: 4-6 mg (12 mg/24 hr max.) NS: 5-20 mg (40 mg/24 hr max.)	2.5-5 mg (10 mg/24 hr max.) 5 mg (10 mg/24 hr max.)	1-2.5 mg (5 mg/24 hr max.)	5-10 mg (30 mg/24 hr max.)	6.25-12.5 mg (25 mg/24 hr max.)	2.5 mg (7.5 mg/24 hr max.)	20-40 mg (80 mg/24 hr max.)
Half-life	2-3 hr	3-4 hr	6-8 hr	2-3 hr	3-4 hr	26 hr	4-6 hr
Crosses blood-brain barrier	–	+	+	+	+	+	+
Use with monoamine oxidase inhibitor (MAOI)	–	–	+	–	+	+	–
Good for recurrences	–	–	+	–	–	+	–
Rapid response	SC, 10-15 min NS, 15-20 min Oral, 30 min	30 min	1-4 hr	30 min	60 min	1-4 hr	20-30 min
Menstrual migraine	+	+	+	+	+	+	+
Other	Now in combination with naproxen (Treximet)			Decrease dose by half with propranolol			Do not use with CYP3A4 drugs (ketoconazole and some macrolide antibiotics)

NS = nasal spray; SC = subcutaneous.

sumatriptan (25 to 100 mg orally, 20 mg intranasally, or 4 to 6 mg subcutaneously); or other triptans (e.g., naratriptan, 2.5 mg; zolmitriptan, 5 mg; rizatriptan, 10 mg; eletriptan, 40 mg; frovatriptan, 2.5 mg; or almotriptan, 12.5 mg). Ergotamine (2 mg sublingually or 1 to 2 mg orally), when given early in the migraine attack, can be effective if the associated nausea and peripheral vasoconstriction are tolerable.

For very severe attacks, dihydroergotamine (1 mg subcutaneously or 0.5 to 1 mg intravenously) is usually effective but generally requires an antiemetic (e.g., promethazine, 25 mg) before IV use.[4,5] Ketorolac (60 mg IM or 30 mg IV), prochlorperazine (10 to 25 mg IM or 10 mg IV delivered over a 5-minute period), or metoclopramide (10 mg IV) are useful for patients who are nonresponsive or have contraindications to vasoactive abortive agents.

PREVENTION

Preventive treatment (see Table 398-3) is often recommended when the headaches interfere with activities on 3 or more days per month, the headaches are severe or prolonged, or migraine is complicated by events such as cerebral infarction. Prophylactic options include β-adrenergic blockers, calcium-channel antagonists, NSAIDs, tricyclic antidepressants, valproate, and topiramate. Topiramate, divalproex, timolol, propranolol, metoprolol, atenolol, nadolol, acebutolol, captopril, lisinopril, and candesartan reduce migraine frequency by 50% or more compared with placebo, with no statistically significant differences among them. Other alternatives include the serotonergic drug cyproheptadine (4 to 20 mg) or the monoamine oxidase inhibitor phenelzine (30 to 60 mg). Acupuncture and biofeedback have been used successfully. OnabotulinumtoxinA injection is also effective for prophylaxis of chronic migraine.

PROGNOSIS

The prognosis for patients with migraine is variable. In many patients, headaches decrease in severity with age, but migraine aura without headache becomes more frequent with older age. Modification of inciting factors such as avoiding dietary triggers (tyramine, phenylethylamine, ethanol), ameliorating or preventing insomnia, and averting environmental triggers (light, sound, odor) may improve outcome. Migraines may become chronic, defined as more than 15 days per month, especially when associated with obesity, snoring, depression, and low socioeconomic status.[6]

TENSION-TYPE HEADACHE

DEFINITION

Tension-type headache is defined as a mild or moderate holocranial headache without nausea or vomiting. Patients may have either photophobia or phonophobia but not both, and the headache does not worsen with activity.

EPIDEMIOLOGY

The 1-year prevalence is 14 to 93 per 100,000 individuals for episodic tension-type headache and 8.1 per 100,000 for chronic tension-type headache. Tension-type headaches are more common in women than in men, regardless of age, race, and educational level. Tension-type headaches are more common in Western countries and less frequent in Asian countries, and they are more common in white persons than in African Americans.

PATHOBIOLOGY

The pathophysiology of tension-type headache is less well understood than that of the other types of headache. Myofascial tenderness is increased, especially in chronic tension-type headache. Genetic factors are uncertain. Migraine and tension-type headache often coexist. Although tension-type headaches are not due to emotion or muscle contraction, triggers of a tension-type headache are similar to those associated with migraine: stress, fatigue, and lack of sleep. Comorbid conditions in patients with tension-type headache include depression and anxiety in more than 50% of individuals.

CLINICAL MANIFESTATIONS

Tension-type headaches are usually mild to moderate in severity, and most individuals do not seek care. Tension-type headache can be episodic (occurring < 15 days per month) or chronic (occurring > 15 days per month). In many patients headaches remain episodic, but about 25% progress to chronic

headache. Of the patients with chronic tension-type headache, about a quarter to a third continue as chronic, half can improve to episodic, and in about a quarter medication overuse headache can develop. Episodic tension-type headaches can last minutes, hours, or days.

DIAGNOSIS

Headaches that can be misdiagnosed as tension-type headache include migraine, hemicrania continua, new daily persistent headache, and headaches caused by brain tumors, elevated or low intracranial pressure, or giant cell arteritis. A careful history is the best way to distinguish other types of headaches.

TREATMENT

Rx

Episodic tension-type headaches are generally treated successfully[7] with acetaminophen (650 to 1000 mg) or NSAIDs (aspirin, 900 to 1000 mg; naproxen, 250 to 500 mg; ibuprofen, 200 to 800 mg; or ketoprofen, 12.5 to 75 mg). However, analgesic use for more than 3 days per week can worsen headaches and lead to medication-induced headache.

PREVENTION

Chronic tension-type headaches may benefit from prophylactic treatment with amitriptyline (starting with 10 mg at bedtime and increased slowly up to 100 mg until the patient improves or intolerable side effects develop), nortriptyline (25 to 100 mg each evening), doxepin (25 to 75 mg/day), maprotiline (10 to 25/mg/day), or fluoxetine (10 to 20 mg/day). Tricyclics are generally more efficacious than serotonin reuptake inhibitors. Muscle relaxants, physical therapy, localized botulinum toxin injection, and acupuncture can be useful.

PROGNOSIS

Tension-type headache has a variable prognosis. Adolescents with tension-type headache and two or more psychiatric factors (e.g., depression and anxiety) have a worse prognosis.

CLUSTER HEADACHE AND OTHER TRIGEMINAL AUTONOMIC CEPHALALGIAS

DEFINITION

Trigeminal autonomic cephalalgias, including cluster headaches, are unilateral headaches associated with ipsilateral autonomic features. Other trigeminal autonomic cephalalgias include paroxysmal hemicrania, which is characterized by bouts of headache that persist for 5 to 30 minutes, is generally unilateral, and usually occurs in women; they typically respond to indomethacin. Hemicrania continua, another indomethacin-responsive headache seen in both men and women, is characterized by continuous unilateral pain and mild associated autonomic features; it frequently coexists with a form of chronic daily headache. Short unilateral neuralgiform headache with conjunctival injection and tearing is a rare trigeminal autonomic cephalalgia that occurs in men; individual headaches persist for only a short time (seconds to 2 minutes).

EPIDEMIOLOGY

Cluster headache occurs in 56 to 401 per 100,000 persons and is more frequent in men (3 : 1 to 7 : 1). Attacks usually begin between 20 and 30 years of age. Paroxysmal hemicrania occurs in 56 to 381 per 100,000 persons; it affects women more often (2 : 1) and can begin at any age but usually commences at 34 to 41 years. Short unilateral neuralgiform headache with conjunctival injection and tearing is rare, with a slight male preponderance (2 : 1).

PATHOBIOLOGY

Cluster headache may have a genetic predisposition. Imaging studies such as positron emission tomography and functional MRI show inferior posterior hypothalamic activation at the onset of cluster headache and other trigeminal autonomic cephalalgias. In addition, the trigeminovascular complex and the cranial autonomic system are activated. The pathophysiology of hemicrania continua is unknown, and there is debate whether it is associated with hypothalamic involvement or whether it resembles migraine.

CLINICAL MANIFESTATIONS

Cluster headache is almost always unilateral, rarely bilateral, and has characteristic ipsilateral autonomic features, commonly including lacrimation and conjunctival injection and occasionally nasal congestion, rhinorrhea, ptosis, miosis, flushing, and eyelid edema (Table 398-5). The location of the pain is usually behind or above the eye or in the temple but can include the forehead, cheek, teeth, or jaw. The pain reaches its maximum intensity in about 9 minutes and tends to end abruptly. Attacks occur one to eight times a day and are usually described as "boring" or "stabbing" excruciating pain that persists for 15 minutes to 2 hours. Migraine symptoms may coexist, including unilateral photophobia, phonophobia, and rarely, an aura. Unlike migraine patients, who usually try to rest, patients with cluster headaches pace and are unable to sit or lie down. Cluster headaches, often precipitated by alcohol, histamine, or nitroglycerin, have a daily periodicity and may also have a seasonal periodicity. For example, episodic cluster headache may occur annually or every 2 years, often in the same season each time. Chronic cluster headache occurs without a remission.

Paroxysmal hemicrania is pain of short duration, usually 2 to 30 minutes, and occurs unilaterally around the eye, temple, or maxillary region, sometimes precipitated by head movements. Autonomic features similar to cluster headache can occur. The usual attack rate is up to 40 episodes each day. Bouts of pain may be episodic, separated by a remission, but most patients have daily chronic paroxysmal hemicrania without a remission.

Short unilateral neuralgiform headache with conjunctival injection and tearing attacks are unilateral and consistently on the same side. Although the

TABLE 398-5 DISTINGUISHING CHARACTERISTICS OF THE TRIGEMINAL AUTONOMIC CEPHALALGIAS

CHARACTERISTIC	CLUSTER	PAROXYSMAL HEMICRANIA	HEMICRANIA CONTINUA	SHORT UNILATERAL NEURALGIFORM HEADACHE WITH CONJUNCTIVAL INJECTION AND TEARING
Sex—F:M	1 : 3-7	2 : 1	2 : 1	1 : 2
Unilateral	+	+	+	+
Attack frequency	1-8/day	1-40/day		3-200/day
Attack duration	15-80 min	2-30 min	Continuous with episodic exacerbations	5-240 sec
Autonomic features	+	+	+ with exacerbations	+
Indomethacin effect	–	+++	+++	–
Acute treatment at onset	Oxygen, sumatriptan SC, DHE nasal spray; sumatriptan or zolmitriptan nasal spray (A-level evidence)	None	None	None
Preventive medications	Verapamil, lithium, corticosteroids, anticonvulsants (A level)	Indomethacin (A level)	Indomethacin (A level)	Lamotrigine, topiramate, gabapentin (B level)

DHE = dihydroergotamine; SC = subcutaneous.

pain is excruciating, the attack is brief, usually seconds; most patients are free of pain between attacks, although a dull ache can be present. Associated autonomic features include ipsilateral conjunctival injection and tearing.

DIAGNOSIS

The diagnostic criteria for cluster headache include severe unilateral orbital, supraorbital, or temporal pain persisting for 15 to 180 minutes with at least one of the following: ipsilateral conjunctival injection or lacrimation, nasal congestion or rhinorrhea, eyelid edema, forehead and facial sweating, miosis with or without ptosis, and restlessness or agitation. Attacks occur between once and as often as eight times each day. There is no other cause of the disorder.

Paroxysmal hemicrania is defined by unilateral pain persisting for 2 to 30 minutes, about 5 times each day, with one or more autonomic features such as conjunctival injection, nasal congestion, eyelid edema, forehead and facial sweating, and miosis or ptosis (or both). Complete prevention may be achieved with indomethacin.

Hemicrania continua is a unilateral headache that occurs daily and continuously without pain-free periods; its intensity is moderate, with exacerbations of severe pain. During the exacerbations, at least one ipsilateral autonomic feature is present: conjunctival redness, lacrimation, nasal congestion, ptosis, or miosis. It responds to indomethacin.

Short unilateral neuralgiform headache with conjunctival injection and tearing is diagnosed by unilateral orbital, supraorbital, temporal stabbing pain persisting for 5 to 240 seconds at a frequency of 3 to 200 per day. It is associated with conjunctival injection and tearing.

An imaging procedure such as MRI is indicated for all patients at the onset of cluster headaches or other trigeminal autonomic cephalalgias, because they can be the result of infection (Chapters 412 to 414), vascular malformation (Chapter 408), or neoplasm, especially a pituitary tumor (Chapter 189). Other possibilities in the differential diagnosis include migraine, hypnic headache (rare short-lasting headaches exclusively during sleep in the elderly), and trigeminal neuralgia.

TREATMENT Rx

Because the course of the headache is brief, oral medications take too long to work to be effective.[8] The use of 100% oxygen at 7 to 10 L/min for 15 to 30 minutes benefits some patients. Sumatriptan or zolmitriptan nasal spray or sumatriptan subcutaneously (4 to 6 mg) can be helpful. Dihydroergotamine can be helpful when given nasally, intramuscularly, or even intravenously. Refractory cases may respond to occipital nerve stimulation. Chronic paroxysmal hemicranias and hemicrania continua are characterized by a response to indomethacin, 25 to 50 mg three times daily. Short unilateral neuralgiform headache with conjunctival injection and tearing attacks is so brief that there are no medications to treat it acutely.

PREVENTION

Preventive medications should be started at the beginning of a cluster bout. Verapamil, 240 to 480 mg, is the drug of choice. Lithium (300 mg twice daily) is an alternative. Corticosteroids (e.g., prednisone, 40 mg/day, or dexamethasone, 4 mg twice daily for 2 weeks) act rapidly as a bridge to prevent cluster headache while other preventive medications are started. Valproic acid (500 to 1500 mg/day in divided doses), topiramate (50 to 100 mg/day), melatonin (4 mg at bedtime), and gabapentin (300 mg three times daily) are sometimes beneficial. Surgical approaches, including suboccipital steroid injections, occipital nerve stimulators, sphenopalatine ganglion stimulation, hypothalamic stimulation, and destructive procedures, are sometimes necessary for this disabling headache.

Paroxysmal hemicrania and hemicrania continua respond to daily indomethacin (25 to 50 mg three times daily). If the patient cannot tolerate indomethacin, calcium-channel blockers (e.g., verapamil, 240 to 480 mg/day) or melatonin may be helpful. Preventive treatment of short unilateral neuralgiform headache with conjunctival injection and tearing includes lamotrigine (100 to 400 mg/day), topiramate (50 to 100 mg), gabapentin (300 to 900 mg), or IV lidocaine (starting at 2 mg/minute with cardiac monitoring).[9]

Short unilateral neuralgiform headache with conjunctival injection and tearing is regarded as a more difficult headache to prevent. Lamotrigine and topiramate may be helpful.

PROGNOSIS

Cluster headache is often a lifelong problem, but remissions may persist for longer periods as the patient ages. The other trigeminal autonomic cephalalgias are probably lifelong; nevertheless, symptomatic treatment combined with preventive medications is helpful.

CHRONIC DAILY HEADACHE

DEFINITION

Though not a specific disorder, chronic daily headache, defined as a headache that is present on more than 15 days per month, is challenging for both patients and physicians. These headaches may be chronic migraine, chronic tension-type headache, new daily persistent headache, or chronic cluster headache, with or without overuse of medications.

EPIDEMIOLOGY

Up to 5% of the population suffers from chronic daily headache, most commonly chronic tension type or chronic migraine. Trigger factors such as a previous infection, mild head injury, or stressful life event are present in 40 to 60% of patients with new daily persistent headache. Risk factors for chronic daily headache include medication overuse, history of migraine headache, frequent headache, depression, female sex, obesity, snoring, stressful life events, and low educational level.

PATHOBIOLOGY

Chronic daily headache is probably related to migraine, with both central and peripheral abnormalities. Once migraine has been prolonged and headache occurs on a daily basis, allodynia, a sense that a usually nonpainful stimulus is becoming painful, often develops. Use of an opiate for more than 8 days per month, especially in men, use of barbiturates for more than 5 days per month, especially in women, or use of triptans for more than 10 to 14 days per month can often lead to chronic migraine headache or at least worsening of headaches.

CLINICAL MANIFESTATIONS

New daily persistent headache is characterized by daily occurrence, onset at specific time, and an unrelenting course. It is generally bilateral, nonpulsating, mild to moderate, and associated with features of migraine, photophobia, phonophobia, or nausea. Severe nausea or vomiting is rare. New daily persistent headache can be disabling and is difficult to treat. Chronic daily headache is often associated with profound psychiatric comorbidity, especially depression and anxiety; such psychiatric comorbidity predicts intractability.

DIAGNOSIS

Diagnosis of chronic daily headache is based on the history. It is important to identify the underlying type of primary chronic daily headache: chronic migraine, chronic tension-type headache, new daily persistent headache, or hemicrania continua. Headaches of less than 4 hours' duration can also be chronic and daily: cluster headache, paroxysmal hemicrania, hypnic headaches occurring every night (usually in the elderly), and episodic stabbing headache. It is most important to exclude secondary headaches (including post-traumatic headache), headaches associated with vascular disorders (e.g., giant cell arteritis, arteriovenous malformations, carotid and vertebral artery dissections), and headaches associated with nonvascular disorders (e.g., intracranial hypertension, intracranial hypotension, infections). MRI and laboratory studies (e.g., ESR in an elderly individual) are commonly recommended. Lumbar puncture (LP) to assess intracranial pressure may also be indicated in selected patients.

TREATMENT Rx

The most common cause of chronic daily headache is overuse of medications, so patients must be weaned off the overused symptomatic medication. Treatment of underlying depression, anxiety, and pain may also be helpful. Occasionally, hospital admission is necessary to break the headache cycle. Acute migraine-specific treatments (see earlier), especially IV dihydroergotamine (0.5 to 2 mg), are helpful in terminating migrainous attacks.

PREVENTION

Medications that are helpful in preventing chronic daily headache include tricyclic antidepressants, selective serotonin reuptake inhibitors if patients are depressed, anticonvulsants, β-blockers, and calcium-channel blockers (see Tables 398-3 and 398-4). For hemicrania continua, indomethacin (25 to 50 mg three times daily) is the preferred treatment.

PROGNOSIS

The prognosis depends on the underlying headache diagnosis. If medication overuse is the cause and the patient is successfully detoxified, about 75% of patients improve when treated with preventive medications. Treatment may fail if the diagnosis is incorrect or because of continued overuse of medications, overuse of caffeine, lack of sleep, dietary or other life triggers, hormonal factors, or psychiatric factors. Explaining medication overuse headache to the patient, in-patient and out-patient detoxification, and multidisciplinary care treatments have been found helpful.

SECONDARY CAUSES OF HEADACHES

Sinus Headache

Rhinosinusitis (Chapter 426) is characterized by inflammation or infection of the nasal mucosa and sinuses. The sinuses themselves are relatively insensate, but ducts, turbinates, blood vessels, and ostia are the painful structures.[10]

Headaches attributed to rhinosinusitis are frontal headaches with pain in the face, ears, or teeth. The onset of pain is simultaneous with the rhinosinusitis, and the headache and face pain resolve within 7 days after successful treatment. The diagnosis requires imaging and clinical evidence that support the diagnosis of acute rhinosinusitis. Many acute and most chronic headaches that are initially thought to result from sinus disease are found to be migraine or tension-type headache.

The headache should resolve with treatment of acute sinusitis (Chapter 426). If it does not, an underlying primary headache disorder is likely.

Temporal (Giant Cell) Arteritis

Temporal arteritis (Chapter 271) is an inflammatory process seen almost exclusively in elderly individuals. Headache, especially pain in the jaw when chewing, is one of the most common features. Its incidence is approximately 12 per 100,000 and increases with age to 51 per 100,000 in individuals older than 80. It affects women more often than men (3 : 1) and is more common in white individuals, especially those of Scandinavian and British descent. It is associated with polymyalgia rheumatica.

The headache has no specific feature, but the pain is usually continuous, generalized, and occasionally throbbing. The temples are generally painful, and patients complain of pain when performing certain activities of daily living, such as chewing food or combing their hair. Transient monocular blindness, permanent blindness, and diplopia can occur.

Elevation of the ESR and C-reactive protein occurs almost invariably. The diagnosis is made by finding giant cells in a temporal artery biopsy specimen. Immediate treatment with corticosteroids, sometimes before the biopsy result is available, is necessary in doses between 40 and 80 mg daily, with the dose then titrated downward while monitoring the ESR or CRP. Used early enough, corticosteroids (Chapter 271) generally prevent the complications of temporal arteritis, including blindness. The disorder can be long lasting.

Intracranial Hypertension and Pseudotumor Cerebri

Intracranial hypertension can be primary and idiopathic or secondary to cerebral venous thrombosis (Chapter 407), a mass in the brain (Chapter 189), hydrocephalus, or other intracranial processes. *Pseudotumor cerebri* is an all-encompassing term referring to increased intracranial pressure without obvious mass lesions.[11] Primary idiopathic intracranial hypertension occurs in obese women of childbearing age. Secondary pseudotumor cerebri causes a similar syndrome but is due to an offending agent such as medications (e.g., tetracycline, minocycline, lithium, vitamin A–related medications, growth hormone), endocrine disorders (e.g., parathyroid dysfunction), and sleep apnea.

Idiopathic increased intracranial pressure occurs in 1 to 2 per 100,000 individuals but in 19 to 20 per 100,000 individuals (15 to 55 years of age) who are obese. Women are affected more frequently than men (6-8 : 1). Onset is usually in young adulthood.

The cause of the increased pressure is either poor CSF absorption, as is thought to be the problem in idiopathic intracranial hypertension; venous hypertension, as is seen in venous thrombosis; or a mass that causes an increase in pressure. A genetic component is also likely because there are reports of the condition occurring in families.

CLINICAL MANIFESTATIONS

Idiopathic intracranial hypertension is characterized by headache in more than 90% of individuals, about 90% of whom are obese. The headache may be pulsatile and is frequently felt behind the eyes. Patients often report neck pain, upper back pain, or even radicular pain. The intensity of headache does not correlate with the height of the intracranial pressure. Pulse-synchronous tinnitus is a frequent accompaniment, as are transient visual obscurations and diplopia.

On examination, papilledema (Chapter 423, Fig. 423-27) may be found. The remainder of the general and neurologic examination is usually normal in patients with idiopathic intracranial hypertension, but abnormalities on examination may point to a secondary cause, such as underlying venous sinus thrombosis (Chapter 407), ischemic stroke, central nervous system infection (Chapters 412 and 413), or brain tumor (Chapter 189). Although idiopathic intracranial hypertension often persists for years, the condition can be self-limited. In about a third of patients, there are permanent visual sequelae related to the effect of papilledema.

DIAGNOSIS

The diagnosis of intracranial pressure is made by the symptoms and signs such as papilledema (Chapter 423, Fig. 423-27). MRI is necessary to exclude secondary causes of increased intracranial pressure. MR or CT venography is often needed to exclude venous sinus thrombosis (Chapter 407). LP must be performed unless patients have a contraindication such as an intracranial mass lesion, and CSF pressure should be measured. The diagnosis can be made if the pressure is elevated (CSF > 250 mm H_2O) and the fluid itself is normal in terms of its protein level, glucose level, and cell count. Visual fields must be examined formally because visual acuity is not affected until late in the course of the disorder.

TREATMENT Rx

Acetazolamide (doses ranging from 500 to 4000 mg daily) combined with a weight loss program is more efficacious for individuals with idiopathic intracranial hypertension and mild to moderate visual loss than is placebo. Any underlying secondary cause should also be treated (e.g., stopping an offending medication, treatment of sleep apnea [Chapter 100]). Weight loss is beneficial in obese subjects. If visual loss progresses, surgical procedures should be considered. Optic nerve sheath fenestration allows CSF to escape through slits or windows in the orbit; sometimes the treatment of one side decreases the optic disc swelling on the other side as well. Complications include visual loss or diplopia, so visual fields must be followed carefully to anticipate and prevent visual loss. Lumbar or ventricular peritoneal diversion procedures also reduce intracranial pressure, but their complications include infection and shunt obstruction.

PROGNOSIS

The prognosis of patients with idiopathic intracranial hypertension is good with treatment, but up to a third of inadequately treated patients can experience permanent defects of visual fields or loss of visual acuity. Individuals are susceptible to recurrence if they suddenly gain weight.

Intracranial Hypotension

Intracranial hypotension (or CSF hypovolemia) causes a headache that is characteristically better when the patient is supine and worse when the patient is upright. It can be primary (spontaneous) or secondary to another underlying cause, most commonly a previous LP.[12]

Intracranial hypotension was once considered rare, but modern imaging techniques suggest an incidence of about 5 per 100,000 per year; it is slightly more common in women than men. The onset is usually at about 40 years of age, but it can occur in children and the elderly. Post-LP headaches occur more commonly but only infrequently persist.

PATHOBIOLOGY

The cause of primary intracranial hypotension is thought to be a small leak or tear in the dura, usually in the lumbar region around cystic structures called Tarlov cysts. The cause of intracranial hypotension may not be the tear itself but rather the low CSF volume and low epidural venous pressure that assists in development of the lower pressure and hence the leak. The leaks frequently occur in the thoracic and cervicothoracic junction spine. Previous trauma history is reported in only one third of cases. Genetic and connective tissue disorders (e.g., Ehlers-Danlos syndrome, Marfan syndrome [Chapter 260]) may predispose individuals to have these leaks.

CLINICAL MANIFESTATIONS

Intracranial hypotension is characterized clinically by a positional headache. The location of the pain is variable, and the most constant characteristic is the orthostatic change in the pain. If the leak is untreated for a long time, the headache may lose the orthostatic characteristic. Posterior neck pain can also occur. Changes in hearing, taste, and balance, as well as blurred vision and diplopia, can develop if hindbrain herniation occurs. If very severe hindbrain herniation occurs, changes in consciousness, subdural hygromas, ataxia, a pseudo-frontotemporal dementia can occur.

DIAGNOSIS

The diagnosis of intracranial hypotension is made by MRI showing pachymeningeal enhancement, venous engorgement, dural thickening, pituitary fossa enlargement, and herniation of the hindbrain (Fig. 398-1). Hindbrain herniation appears as a downward descent of the posterior fossa along with loss of the prechiasmatic cistern, flattening of the pons against the clivus, and descent of the cerebellar tonsils, which is often misconstrued as a Chiari I malformation. LP may also show low (<50 mm H_2O) CSF pressure, but it also may be normal. The diagnosis is most commonly made by clinical characteristics and imaging, so the decision as to whether or not to do an LP should be made on a case-by-case basis because there is at least a theoretical risk of more hindbrain herniation. The differential diagnosis includes new daily persistent headache, chronic migraine, or another secondary headache. The diagnosis is confirmed if a CSF leak is demonstrated by isotope studies, CT myelography, or MR myelography.[13]

TREATMENT AND PROGNOSIS

Rx

For spontaneous intracranial hypotension, the recommended treatment is bed rest and an epidural blood patch (blind or directed). Treatment of CSF leaks includes bed rest, caffeine (200 to 300 mg two to three times daily), an abdominal binder wrapped around the abdomen to increase central pressure, and generous intake of oral fluids.[14] For post–dural puncture headache and most spontaneous episodes, an epidural blood patch usually improves the symptoms within days. Surgical repair is rarely required. With treatment, the symptoms and MRI findings should resolve completely. Recurrence is infrequent.

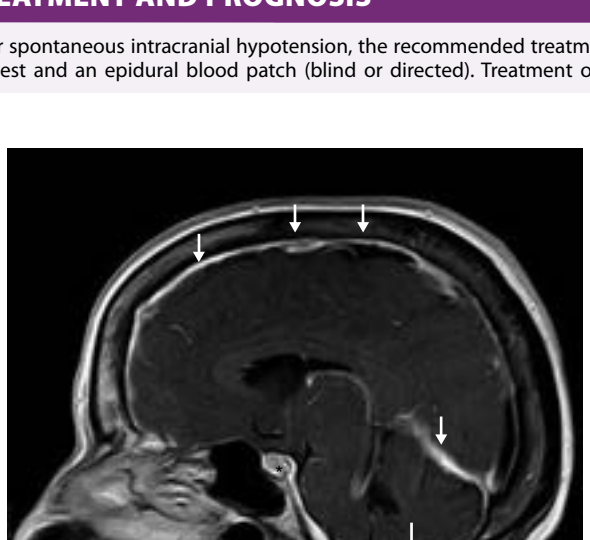

FIGURE 398-1. Intracranial hypotension. This 56-year-old woman had headaches that initially were positional. Gadolinium-enhanced magnetic resonance imaging shows characteristic findings of engorgement of the pituitary (*), slumping of the posterior fossa with tonsillar herniation (*large arrow*), and meningeal enhancement (*smaller arrows*).

Trigeminal Neuralgia

Trigeminal neuralgia is a distinct, excruciatingly painful condition provoked by sensory stimuli in the distribution of the trigeminal nerve.[15] Trigeminal neuralgia occurs in 4 per 100,000 individuals, most commonly in persons between 50 and 70 years of age and in women slightly more than in men (1.5 : 1).

In younger individuals, multiple sclerosis (Chapter 411) can be associated with the condition. In older individuals, an ectatic artery in the vertebrobasilar system often causes the syndrome. The trigeminal nerve root entry zone is thought to be the site of pathology. Either demyelination or compression of this region increases the firing of trigeminal afferents. When a specific cause can be defined, the term *symptomatic trigeminal neuralgia* is often used.

CLINICAL MANIFESTATIONS

Trigeminal neuralgia pain is characteristically sharp, lancinating (shooting), and electric shock–like in the distribution of the trigeminal nerve: cheek (V2), chin or lower teeth (V3), and around the eye (V1). A combination of V2 and V3 is the most common. The paroxysms are brief—usually seconds but up to 2 minutes. Some patients have a dull and continuous interictal pain, whereas most have only staccato-like volleys of pain. Pain is usually triggered by stimuli such as touching the face, brushing the teeth, air moving across the face, or masticating food. Once a volley of pain is triggered, there is usually a refractory period in which pain will not occur.

DIAGNOSIS

Diagnostic criteria include paroxysmal attacks of pain persisting for a second to 2 minutes and affecting one or more divisions of the trigeminal nerve. To make the diagnosis, the pain must be intensely sharp, stabbing, or precipitated by a trigger. Each attack is stereotypical, and there are usually no other neurologic defects. Idiopathic trigeminal neuralgia by definition has no causative lesion, whereas symptomatic trigeminal neuralgia has a cause such as vascular compression of the trigeminal nerve root exit zone. The differential diagnosis includes trigeminal autonomic cephalgia, which has autonomic accompaniments that are not associated with trigeminal neuralgia. Atypical facial pain, idiopathic stabbing headache, and Tolosa-Hunt syndrome, an inflammatory syndrome of the anterior cavernous sinus, are also included in the differential. MRI is recommended to evaluate possible secondary causes of trigeminal neuralgia, such as demyelination, tumors, and vascular loops on the trigeminal nerve exit zone.

TREATMENT

Rx

Trigeminal neuralgia is treated with medications or surgery. Carbamazepine (400 to 1200 mg) is considered the first-line agent for the neuralgia. Phenytoin (200 to 300 mg), baclofen (40 to 80 mg), clonazepam (2 to 6 mg), valproic acid (500 to 1500 mg), lamotrigine (100 to 400 mg), gabapentin (900 to 1800 mg), oxcarbazepine (300 to 1800 mg), levetiracetam 2 to 4 g), and topiramate (50 to 200 mg) are also used. Botulinum toxin may be a novel treatment for this disorder. Surgical treatments include microvascular decompression, which may alleviate the symptoms and preserve sensory function. Other treatments include partial destruction of the trigeminal nerve with heat (radio frequency lesions) or with glycerol (chemical destruction).[16]

PROGNOSIS

Patients with trigeminal neuralgia can have spontaneous or medication-induced remissions. Microvascular decompression is often curative. In patients whose pain is triggered by mastication, weight loss and inanition may develop; prompt treatment is essential.

Glossopharyngeal Neuralgia

Less common than trigeminal neuralgia, glossopharyngeal neuralgia is unilateral pain in the distribution of the glossopharyngeal and vagal nerves in the ear, jaw, throat, and base of the tongue.[17] This neuralgia is rare, with a prevalence of less than 1/100,000. The cause is thought to be compression of the glossopharyngeal nerve by blood vessels, tumor, or aneurysm and demyelination or infection.

The pains are paroxysmal and persist for less than seconds to 2 minutes, but patients can experience 30 to 40 attacks in a day. Like trigeminal neuralgia, the pain is triggered by chewing, swallowing, or talking.

The diagnosis is made clinically. MRI should be done to evaluate the glossopharyngeal nerve to exclude a tumor or vascular abnormality. The differential diagnosis includes trigeminal neuralgia, geniculate neuralgia, and atypical pain syndrome.

Pharmacologic therapy is similar to that for trigeminal neuralgia, and carbamazepine 200 to 800 mg) is usually the drug of choice. Surgical therapy and microvascular decompression or radio frequency ablation should be considered in patients whose weight loss does not respond promptly to medication.[18]

Grade A References

A1. Prior MJ, Codispoti JR, Fu M. A randomized, placebo-controlled trial of acetaminophen for treatment of migraine headache. *Headache*. 2010;50:819-833.

A2. Pini LA, Guerzoni S, Cainazzo M, et al. Comparison of tolerability and efficacy of a combination of paracetamol + caffeine and sumatriptan in the treatment of migraine attack: a randomized, double-blind, double-dummy, cross-over study. *J Headache Pain*. 2012;13:669-675.

A3. Kirthi V, Derry S, Moore RA. Aspirin with or without an antiemetic for acute migraine headaches in adults. *Cochrane Database Syst Rev*. 2013;4:CD008041.

A4. Evers S, Afra J, Frese A, et al. EFNS guideline on the drug treatment of migraine—revised report of an EFNS task force. *Eur J Neurol*. 2009;16:968-981.

A5. Law S, Derry S, Moore RA. Sumatriptan plus naproxen for acute migraine attacks in adults. *Cochrane Database Syst Rev*. 2013;10:CD008541.

A6. Lipton RB, Stewart WF, Stone AM, et al. Disability in Strategies of Care Study Group. Stratified care vs step care strategies for migraine: the Disability in Strategies of Care (DISC) Study: a randomized trial. *JAMA*. 2000;284:2599-2605.

A7. Aurora SK, Silberstein SD, Kori SH, et al. MAP0004, orally inhaled DHE: a randomized, controlled study in the acute treatment of migraine. *Headache*. 2011;51:507-517.

A8. Thorlund K, Mills EJ, Wu P, et al. Comparative efficacy of triptans for the abortive treatment of migraine: a multiple treatment comparison meta-analysis. *Cephalalgia*. 2014;34:258-267.

A9. Kostic MA, Gutierrez FJ, Rieg TS, et al. A prospective, randomized trial of intravenous prochlorperazine versus subcutaneous sumatriptan in acute migraine therapy in the emergency department. *Ann Emerg Med*. 2010;56:1-6.

A10. Friedman BW, Garber L, Yoon A, et al. Randomized trial of IV valproate vs metoclopramide vs ketorolac for acute migraine. *Neurology*. 2014;82:976-983.

A11. Silberstein SD, Holland S, Freitag F, et al. Evidence-based guideline update: pharmacologic treatment for episodic migraine prevention in adults: report of the Quality Standards Subcommittee of the American Academy of Neurology and the American Headache Society. *Neurology*. 2012;78:1337-1345.

A12. Linde M, Mulleners WM, Chronicle EP, et al. Valproate (valproic acid or sodium valproate or a combination of the two) for the prophylaxis of episodic migraine in adults. *Cochrane Database Syst Rev*. 2013;6:CD010611.

A13. Linde M, Mulleners WM, Chronicle EP, et al. Topiramate for the prophylaxis of episodic migraine in adults. *Cochrane Database Syst Rev*. 2013;6:CD010610.

A14. Shamliyan TA, Choi JY, Ramakrishnan R, et al. Preventive pharmacologic treatments for episodic migraine in adults. *J Gen Intern Med*. 2013;28:1225-1237.

A15. Aurora SK, Dodick DW, Diener HC, et al. OnabotulinumtoxinA for chronic migraine: efficacy, safety, and tolerability in patients who received all five treatment cycles in the PREEMPT clinical program. *Acta Neurol Scand*. 2014;129:61-70.

A16. Jackson JL, Shimeall W, Sessums L, et al. Tricyclic antidepressants and headaches: systematic review and meta-analysis. *BMJ*. 2010;341:c5222.

A17. Schiapparelli P, Allais G, Rolando S, et al. Acupuncture in primary headache treatment. *Neurol Sci*. 2011;32(suppl 1):S15-S18.

A18. Cohen AS, Burns B, Goadsby PJ. High-flow oxygen for treatment of cluster headache: a randomized trial. *JAMA*. 2009;302:2451-2457.

A19. Law S, Derry S, Moore RA. Triptans for acute cluster headache. *Cochrane Database Syst Rev*. 2013;7:CD008042.

A20. Leroux E, Valade D, Taifas I, et al. Suboccipital steroid injections for transitional treatment of patients with more than two cluster headache attacks per day: a randomised, double-blind, placebo-controlled trial. *Lancet Neurol*. 2011;10:891-897.

A21. Wall M, McDermott MP, Kieburtz KD, et al. Effect of acetazolamide on visual function in patients with idiopathic intracranial hypertension and mild visual loss: the idiopathic intracranial hypertension treatment trial. *JAMA*. 2014;311:1641-1651.

A22. Bradbury CL, Singh SI, Badder SR, et al. Prevention of post-dural puncture headache in parturients: a systematic review and meta-analysis. *Acta Anaesthesiol Scand*. 2013;57:417-430.

A23. Gronseth G, Cruccu G, Alksne J, et al. Practice parameter: the diagnostic evaluation and treatment of trigeminal neuralgia (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology and the European Federation of Neurological Societies. *Neurology*. 2008;71:1183-1190.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Society HCCotIH. *The International Classification of Headache Disorders*. 3rd ed. (beta version). *Cephalalgia*. 2013;33:629-808.
2. Ju YE, Schwedt TJ. Abrupt-onset severe headaches. *Semin Neurol*. 2010;30:192-200.
3. Carville S, Padhi S, Reason T, et al. Diagnosis and management of headaches in young people and adults: summary of NICE guidance. *BMJ*. 2012;345:e5765.
4. Silberstein SD, Kori SH. Dihydroergotamine: a review of formulation approaches for the acute treatment of migraine. *CNS Drugs*. 2013;27:385-394.
5. Sumamo Schellenberg E, Dryden DM, Pasichnyk D, et al. AHRQ Comparative Effectiveness Reviews. Acute Migraine Treatment in Emergency Settings. Rockville, Md: Agency for Healthcare Research and Quality (US); 2012.
6. Halker RB, Hastriter EV, Dodick DW. Chronic daily headache: an evidence-based and systematic approach to a challenging problem. *Neurology*. 2011;76:S37-S43.
7. Freitag F. Managing and treating tension-type headache. *Med Clin North Am*. 2013;97:281-292.
8. Francis GJ, Becker WJ, Pringsheim TM. Acute and preventive pharmacologic treatment of cluster headache. *Neurology*. 2010;75:463-473.
9. Goadsby PJ. Trigeminal autonomic cephalalgias. *Continuum (Minneap Minn)*. 2012;18:883-895.
10. Kaymakci M, Cikriklar HI, Pay G. The aetiology underlying sinus headaches. *J Int Med Res*. 2013;41:218-223.
11. Friedman DI, Liu GT, Digre KB. Revised diagnostic criteria for the pseudotumor cerebri syndrome in adults and children. *Neurology*. 2013;81:1159-1165.
12. Mokri B. Spontaneous low pressure, low CSF volume headaches: spontaneous CSF leaks. *Headache*. 2013;53:1034-1053.
13. Schievink WI, Dodick DW, Mokri B, et al. Diagnostic criteria for headache due to spontaneous intracranial hypotension: a perspective. *Headache*. 2011;51:1442-1444.
14. Amoozegar F, Guglielmin D, Hu W, et al. Spontaneous intracranial hypotension: recommendations for management. *Can J Neurol Sci*. 2013;40:144-157.
15. Zakrzewska JM, Linskey ME. Trigeminal neuralgia. *BMJ*. 2014;348:g474.
16. Cheng JS, Lim DA, Chang EF, et al. A review of percutaneous treatments for trigeminal neuralgia. *Neurosurgery*. 2014;10(suppl 1):25-33.
17. Reddy GD, Viswanathan A. Trigeminal and glossopharyngeal neuralgia. *Neurol Clin*. 2014;32:539-552.
18. Singh PM, Dehran M, Mohan VK, et al. Analgesic efficacy and safety of medical therapy alone vs combined medical therapy and extraoral glossopharyngeal nerve block in glossopharyngeal neuralgia. *Pain Med*. 2013;14:93-102.

REVIEW QUESTIONS

1. A 30-year-old woman presents with frequent bilateral, non-throbbing headaches that cause photophobia, phonophobia, and nausea. The headaches worsen with activity. They occur 4 times monthly and last 24 to 48 hours. The most likely diagnosis is:

A. Migraine
B. Tension-type headache
C. Hemicrania continua
D. Atypical facial pain

Answer: A The most likely diagnosis is migraine. If a headache is associated with both photophobia/phonophobia and nausea, the diagnosis cannot be tension-type headache. Hemicrania continua is usually unilateral, and although it can last for longer periods of time, it may have autonomic symptoms. Atypical facial pain is really not a diagnosis in this case, because there are migrainous features. (See Society HCCotIH. *The International Classification of Headache Disorders*. 3rd ed. [beta version]. *Cephalalgia.* 2013;33: 629-808.)

2. A 35-year-old woman presents with a unilateral headache over her right eye. She has associated tearing and redness of the eye with the pain. The headaches last 40 minutes and occur daily, usually in the evening. The correct diagnosis is:

A. Paroxysmal hemicrania
B. Cluster headache
C. Hemicrania continua
D. Acute glaucoma

Answer: B She has unilateral brief pain with two autonomic features that are diagnostic of cluster headaches. Although men are more likely to have cluster headache, women suffer from this type of headache as well. Paroxysmal hemicrania would be more brief, usually lasting less than 30 minutes. Hemicrania continua is longer lasting, and acute glaucoma would not be intermittent or occur only at a specific time.

3. A 25-year-old man has increasingly frequent headaches characterized by bilateral holocranial headache, photophobia, nausea, and rare vomiting. The headaches occur 25 days each month, and he is now using almost daily sumatriptan. The most likely diagnosis is:

A. Chronic migraine
B. Chronic migraine with medication overuse
C. Chronic tension-type headache
D. Chronic tension-type headache with medication overuse

Answer: B The most likely diagnosis is chronic migraine with medication overuse, because he has a headache more than 15 days per month and uses sumatriptan more than 3 days in a week. The headache may revert to episodic or chronic migraine when he reduces his daily acute rescue treatment. The headache is associated with nausea and rare vomiting, so it is not a tension-type headache.

4. A 30-year-old woman presents with chronic daily headache, pulsatile tinnitus, and episodic blurring of her vision. On examination she has papilledema. What evaluation would you perform?

A. Magnetic resonance imaging (MRI) and lumbar puncture (LP)
B. MRI, LP, and visual fields
C. Computed tomography (CT)
D. LP alone
E. None of the above

Answer: B The MRI is required to look for any mass lesions and can often show venous sinus thrombosis. The LP must be done to not only be sure that the opening pressure is elevated but also that the cerebrospinal fluid itself is normal. Visual fields must be performed and monitored to prevent blindness; visual acuity testing alone is insufficient. CT scans are good for excluding mass lesions but are not adequate alone. LP alone without imaging is contraindicated. LP also may be contraindicated if the patient has an associated significant Chiari I malformation.

399

TRAUMATIC BRAIN INJURY AND SPINAL CORD INJURY

GEOFFREY S.F. LING

EPIDEMIOLOGY

Traumatic brain injury and traumatic spinal cord injury are common preventable diseases. Approximately 1.4 million cases of traumatic brain injury are reported annually in the United States, but this number is widely recognized as a gross underestimation. Concussions and milder brain injuries occur in many millions of individuals each year. In the U.S. military alone, over 25,000 service members suffered traumatic brain injuries in 2013, of which about 85% were mild concussions. Moderate to severe traumatic brain injury results directly in about 52,000 deaths in the United States annually—almost a third of all injury-related deaths—and is the single leading cause of traumatic death and disability (Chapter 111). The majority of traumatic brain injuries are due to falls (Chapter 25), motor vehicle accidents, and assaults. An additional approximately 11,000 cases of severe spinal cord injury occur each year in the United States, resulting from of motor vehicle accidents, falls, sports-related injuries, and work-related accidents (Chapter 111). The majority of patients with traumatic brain and spinal cord injuries are young adult males.

Over the past 20 years, overall mortality associated with traumatic brain and spinal cord injuries has decreased because of prompt neurosurgical intervention, improved care in intensive care units (ICUs), and prevention of complications such as deep vein thrombosis and decubitus ulcers. The almost 5.5 million survivors of traumatic brain and spinal cord injuries in the United States often require extended rehabilitation.[1] Because the majority of these patients are young and otherwise in good physical health at the time of injury, many need chronic care for decades. Even relatively minor injury can lead to major disability. If untreated, many patients with mild to moderate traumatic brain injury continue to have residual symptoms months later, and many are unable to return to gainful employment.[2]

PATHOBIOLOGY

Traumatic injury to the central nervous system has two phases. The first is neuronal injury and occurs as a direct result of the initiating traumatic event. The second or late phase, caused by multiple neuropathologic processes, can continue for days to weeks after the initial injury.

Primary Injury Phase

The primary injury phase is immediate, and its damage, which can cause death almost instantaneously, is often complete by the time medical care can be instituted. In closed compartment injury to the head or spine, the direct impact of neuronal tissue against the bony vault and shearing of neurovascular structures result in brain damage. Because brain neuronal structures reside in a fluid-filled compartment, these structures can lag behind the bony structure as it moves during sudden stopping of the body in motion. Thus, the structures will strike both anteriorly and posteriorly against the inner bony table, and a coup-contrecoup lesion will result. If a rotational component is present, the structures will torque, twist, and shear, thereby causing diffuse axonal injury. Motor vehicle accidents are particularly injurious because of the sudden deceleration. In penetrating lesions, the moving projectile tears neural, vascular, and support structures as it traverses through the brain or spinal cord. A projectile moving at high velocity (e.g., bullet) creates a vacuum that can cause tissue cavitation in its wake. The temporary cavity, which will ultimately collapse, may be many-fold larger than that of the projectile itself. The transient expansion of surrounding tissue can cause substantial irreversible damage.

Secondary Injury Phase

The delayed secondary phase of injury, which begins immediately after the primary phase and can continue for a prolonged period, involves both neurons and glia. Most neurologic injury may be related to this secondary injury, when "neuron suicide" is caused by processes such as hypoxia, ischemia, inflammation, and the effects of free radicals, excitatory amino acids, and certain ions (e.g., calcium). The injured brain is more susceptible to hypoxic-ischemic states. The most commonly affected areas are the hippocampus and "watershed" areas. It has been hypothesized that much of the delayed neurologic compromise can be attributed to delayed ischemia.

Diffuse microvascular damage is due to early loss of cerebral vascular autoregulation and loss of integrity of the blood-brain barrier, with resulting endothelial changes such as the formation of intraluminal microvilli. Although the clinical significance of this injury is uncertain, it may play a role in the development of cerebral edema.

Diffuse axonal injury, which consists of shearing of axons in cerebral white matter, causes neurologic deficits such as nonfocal encephalopathy. The consequences of this type of injury can be delayed for up to 12 hours after the initial trauma.

Following a single concussion, these same processes probably occur but not to the degree that they cause detectable permanent damage. If multiple concussions are sustained during a patient's lifetime, however, chronic traumatic encephalopathy may develop and result in dementia and other neurodegenerative disorders. One potential mechanism is an accumulation of tau proteins.

CLINICAL MANIFESTATIONS

Traumatic Brain Injury

The signs and symptoms of traumatic brain injury vary with its severity. Patients suffering from mild traumatic brain injury, oftentimes called a *mild concussion*, often experience transient loss of consciousness, headache, difficulty concentrating, anxiety, and disrupted sleep. Chronic traumatic encephalopathy can present with alterations in mood and behavior or with cognitive impairment.[3]

The clinical examination is typically normal, but detailed neuropsychological testing may reveal mild cognitive abnormalities. With moderate traumatic brain injury, patients may have an abnormal sensorium, motor and sensory involvement, and impaired language; the results of neurologic examination will be abnormal. In severe traumatic brain injury, patients are comatose; at best, they may exhibit some eye opening and decorticate or decerebrate posturing to stimulation. Traumatic brain injury may be accompanied by a transient increase in systemic arterial pressure, and some patients may become apneic.

Focal injuries cause neurologic deficits related to the site of impact. The orbitofrontal and anterior temporal lobes are most commonly affected. Extreme vigilance is needed to recognize the development of delayed hematomas and edema, which can be manifested days later.

Traumatic Spinal Cord Injury

Spinal Cord Syndromes

There are three main spinal cord syndromes: Brown-Séquard, central cord, and anterior cord syndromes. In Brown-Séquard syndrome, the deficits are referable to a lesion of a lateral half of the cord; findings consist of loss of ipsilateral motor, touch, proprioception, and vibration sensation, as well as contralateral loss of pain and temperature sensation. Central cord syndrome is manifested as bilateral loss of motor function involving the upper extremities but sparing the lower extremities and is sometimes referred to as "man in a barrel syndrome." Proximal weakness is greater than distal weakness. Pain and temperature sensation is reduced, whereas proprioception and vibration are usually spared. Anterior cord syndrome is manifested by deficits referable to bilateral anterior and lateral spinal cord columns or funiculi. There is loss of touch, pain, and temperature sensation and motor function below the level of the lesion, but the posterior column functions of proprioception and vibratory sensation remain intact.

Spinal Shock

After acute traumatic spinal cord injury, patients may suffer from spinal shock or temporary loss of spinal reflexes below the level of injury, including loss of muscle stretch reflexes, the bulbocavernosus reflex, and the anal wink. In high cervical injuries, the lower reflexes (bulbocavernosus and anal wink) may be preserved. Some patients demonstrate the Schiff-Sherrington phenomenon, in which reflexes are affected above the level of injury. Patients with spinal shock also may lose autonomic reflexes, thereby leading to neurogenic hypotension, ileus, and urinary retention.

DIAGNOSIS

Traumatic Brain Injury

Point-of-injury standardized clinical tools, such as the Standardized Assessment of Concussion (Table 399-1), can help first responders identify patients at risk of having suffered a mild traumatic brain injury or concussion. The diagnosis of traumatic brain injury is still made clinically by a physician.

For more severe injury, the Glasgow Coma Scale score (Table 399-2) should be calculated promptly, and a detailed neurologic examination should be performed to determine the extent of injury and the severity of impairment. Important clinical signs of occult injury may be revealed on a general physical examination. For example, a scalp laceration should be palpated for evidence of an underlying skull fracture. Periorbital ecchymosis ("raccoon

TABLE 399-1 ASSESSING BRAIN INJURY

Neurologic screening: loss of consciousness, incoordination, memory loss, blank look, facial injury			
Neck examination: describe range of motion, tenderness, upper and lower limb sensation and strength			
Balance examination: double leg stance, single leg stance, tandem stance			
Coordination examination: finger to nose			
Orientation:			
What month is it?			
What is the date today?			
What is the day of the week?			
What year is it?			
What time is it right now? (within 1 hour)			
Immediate memory			
Word list:	*Alternate lists:*		
elbow	candle	baby	finger
apple	paper	monkey	penny
carpet	sugar	perfume	blanket
saddle	sandwich	sunset	lemon
bubble	wagon	iron	insect
Concentration: digits backward			
4-9-3	6-2-9	5-2-6	4-1-5
3-8-1-4	3-2-7-9	1-7-9-5	4-9-6-8
6-2-9-7-1	1-5-2-8-6	3-8-5-2-7	6-1-8-4-3
7-1-8-4-6-2	5-3-9-1-4-8	8-3-1-9-6-4	7-2-4-8-5-6
Concentration: months of the year in reverse order			
Delayed recall: memory (administered after physical examinations, same word list as before)			

Table modified from McCrea M. Standardized mental status assessment of sports concussion. *Clin J Sport Med.* 2001;11:176-181; and SCAT3. *Br J Sports Med.* 2013;47:259.

TABLE 399-2 GLASGOW COMA SCALE SCORE

BEST EYE RESPONSE	BEST VERBAL RESPONSE	BEST MOTOR RESPONSE
1 = No eye opening	1 = No verbal response	1 = No motor response
2 = Eye opening to pain	2 = Incomprehensible sounds	2 = Extension to pain
3 = Eye opening to verbal command	3 = Inappropriate words	3 = Flexion to pain
4 = Eyes open spontaneously	4 = Confused 5 = Oriented	4 = Withdrawal from pain 5 = Localizing pain 6 = Obeys commands

To calculate the score, sum the numbers from each of the three columns.

TABLE 399-3 DECISION RULES FOR DETERMINING INDICATIONS FOR CT SCAN IN PATIENTS WITH MINOR HEAD INJURY

STUDY	POPULATION OF PATIENTS	INDICATIONS FOR CT SCAN	REPORTED ACCURACY (%)* SENSITIVITY	SPECIFICITY
Canadian CT Head Rule[†]	GCS score of 13-15, loss of consciousness, no neurologic deficit, age ≥ 16 yr	High-risk patients: GCS score <15 at 2 hr after injury, suspected skull fracture, any sign of basal skull fracture, vomiting (≥2 times), age ≥ 65 yr[‡]	100	69
		Medium-risk patients: retrograde amnesia > 30 min, dangerous mechanism (pedestrian vs. motor vehicle, ejection from motor vehicle, fall from height > 1 m or 5 stairs)[§]	98	50
New Orleans Criteria[¶]	GCS score of 15, loss of consciousness, no neurologic deficit, no seizure, no anticoagulation, age ≥ 3 yr	Headache, vomiting, seizure, intoxication, short-term memory deficit, age > 60 yr, or injury above the clavicles	100	25

*Validity for identifying patients with traumatic CT findings.
[†]Stiell IG, Wells GA, Vandemheen K, et al. The Canadian CT Head Rule for patients with minor head injury. *Lancet.* 2001;357:1391-1396.
[‡]High-risk patients in whom a CT scan is mandatory.
[§]Medium-risk patients in whom a CT scan is recommended but close clinical observation is an alternative.
[¶]Haydel MJ, Preston CA, Mills TJ, et al. Indications for computed tomography in patients with minor head injury. *N Engl J Med.* 2000;343:100-105.
CT = computed tomography; GCS = Glasgow Coma Scale.

TABLE 399-4 AMERICAN ACADEMY OF NEUROLOGY: DIAGNOSIS AND MANAGEMENT OF CONCUSSION

CRITICAL STEPS	SUPPORTING INFORMATION
Immediately remove from play or work	Adherence to state concussion laws
First responders should use a validated clinical tool to determine risk of concussion	(See Table 399-1)
Diagnosis and clinical care is made by LHCP	Clinical practice guidelines for treatment of symptoms
Return to play or work only after detailed evaluation and written authorization by LHCP	Graded physical activity that does not exacerbate symptoms Cognitive restructuring through education, reassurance, reattribution of symptoms
Retirement from activity counseling by LHCP	Retirement counseling for patients with history of multiple concussions with subjective neurobehavioral symptoms should begin a discussion with the patient about retirement

LHCP = Licensed health care provider, an individual who has acquired knowledge and skills relevant to evaluation and management of sports concussions and is practicing within the scope of his or her training and experience.
Adapted from Giza CC, Kutcher JS, Ashwal S, et al. Summary of evidence-based guideline update: evaluation and management of concussion in sports: report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology.* 2013;80:2250-2257.

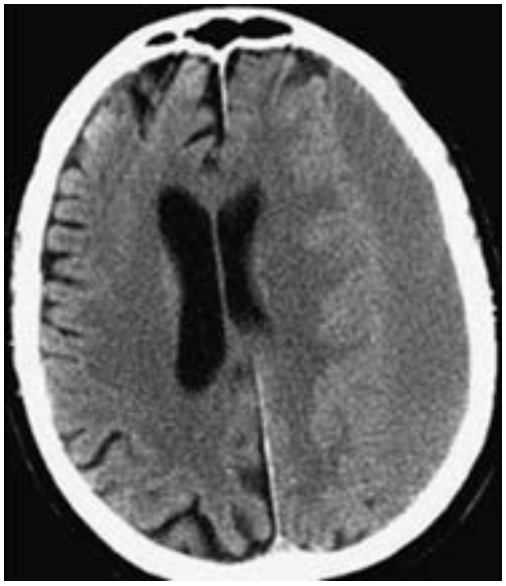

FIGURE 399-1. Subdural hematoma.

eyes") and postauricular ecchymosis ("Battle sign") suggest a basal skull fracture. A clear or blood-tinged watery discharge from the nose or ear may be a cerebrospinal fluid leak.

Intracranial bleeding caused by traumatic brain injury includes subdural hematoma, epidural hematoma, intraparenchymal hemorrhage, contusion, and traumatic subarachnoid hemorrhage (Chapter 408). The most common is subdural hematoma, which is the basis of approximately 50% of admissions for head injury. Epidural hematoma accounts for about 3%. An associated skull fracture, especially at the temporoparietal junction, increases the incidence of epidural hematoma, usually by disruption of the middle meningeal artery.

Imaging

A computed tomography (CT) scan without contrast should be obtained as soon as possible after the initial clinical assessment. The need for neuroimaging is best determined by using the Glasgow Coma Scale score and a validated clinical prediction instrument such as the Canadian CT Head Rule (Table 399-3). In any patient suspected of having suffered a head injury, the severity of the concussion should be assessed (Table 399-4). A subdural hematoma (Fig. 399-1) is blood that accumulates above the brain but below the dura; on CT imaging it appears as a crescentic or concave opacity overlying the brain. An epidural hematoma (Fig. 399-2) is blood that accumulates below the skull but above the dura; it appears as a convex or lenticular opacity on CT imaging. Skull fractures are best diagnosed with the use of CT bone windows.

Traumatic Spinal Cord Injury

A detailed neurologic examination is needed to identify the level of the injury and the severity of any deficits, as well as to document the degree of neurologic dysfunction at the earliest time possible. The level of the injury is the lowest spinal cord segment with intact motor and sensory function. Normal neurologic findings in patients with a clear sensorium obviate the need for imaging studies. However, any complaints of pain over the spine, numbness, tingling, or weakness should raise suspicion of spinal cord injury. In particular, a complaint of "burning hands" suggests traumatic spinal cord injury.

The time of injury should be recorded as accurately as possible. The prognosis for neurologic improvement is better if the lesion is incomplete as opposed to complete. During the acute period, serial examinations must be performed frequently.

If spinal cord injury is suspected, the patient should be appropriately immobilized, such as with a rigid collar and back board. In patients who are able to cooperate with a neurologic examination, are not intoxicated, and do not have painful distracting injuries (e.g., femoral fracture, which would interfere with the leg motor and sensory examination), normal neurologic findings effectively rule out cervical spine disease.

Imaging

In patients who are alert and stable, the Canadian C-Spine Rule (Fig. 399-3) can be used to reduce unnecessary spinal imaging without any adverse effect on patients' outcomes.■ In other patients, the radiologic evaluation should begin with plain radiographs of the bony spine, with further neuroimaging of any abnormalities that are found. Bony vertebrae should be examined with CT, whereas the spinal cord and intervertebral and paravertebral soft tissue are best studied with MRI. A chest radiograph is usually indicated to provide images of the lower cervical and thoracic vertebrae; the presence of a pleural effusion in the setting of a possible thoracic spine injury suggests a hemothorax.

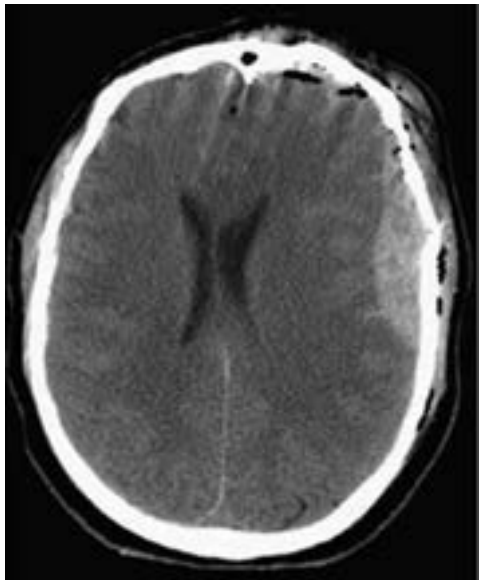

FIGURE 399-2. Epidural hematoma.

Ligamentous Injury versus Spinal Cord Injury

If plain radiographs of the cervical spine are normal but the patient still complains of neck pain, a ligamentous injury should be considered. Ligamentous injury can be evaluated by flexion-extension radiographs of the cervical spine. If pain prevents an adequate study, patients should be kept in a rigid cervical collar for 3 to 5 days until the pain and muscle spasm resolve. If studies at that time are normal, the patient will no longer require the collar. Conversely, abnormal results warrant surgical evaluation to determine whether further immobilization or surgical correction is necessary.

TREATMENT

Rx

The immediate goals of therapy are to arrest ongoing injury, preserve and if possible restore neurologic function, and avoid secondary medical complications. To achieve this goal, an organized team approach is essential. Despite major research efforts, current clinical treatment is largely confined to supportive measures: maintaining perfusion pressure, minimizing intracompartment hypertension (e.g., increased intracranial pressure [ICP]), and indirectly treating edema.

Traumatic Brain Injury

Initial Management

It is crucial that prehospital providers optimize perfusion and oxygenation; the duration and severity of hypoxia and hypotension in this critical early period have dramatic consequences on clinical outcome. Treatment begins with immediate attention to airway and cardiopulmonary function, early

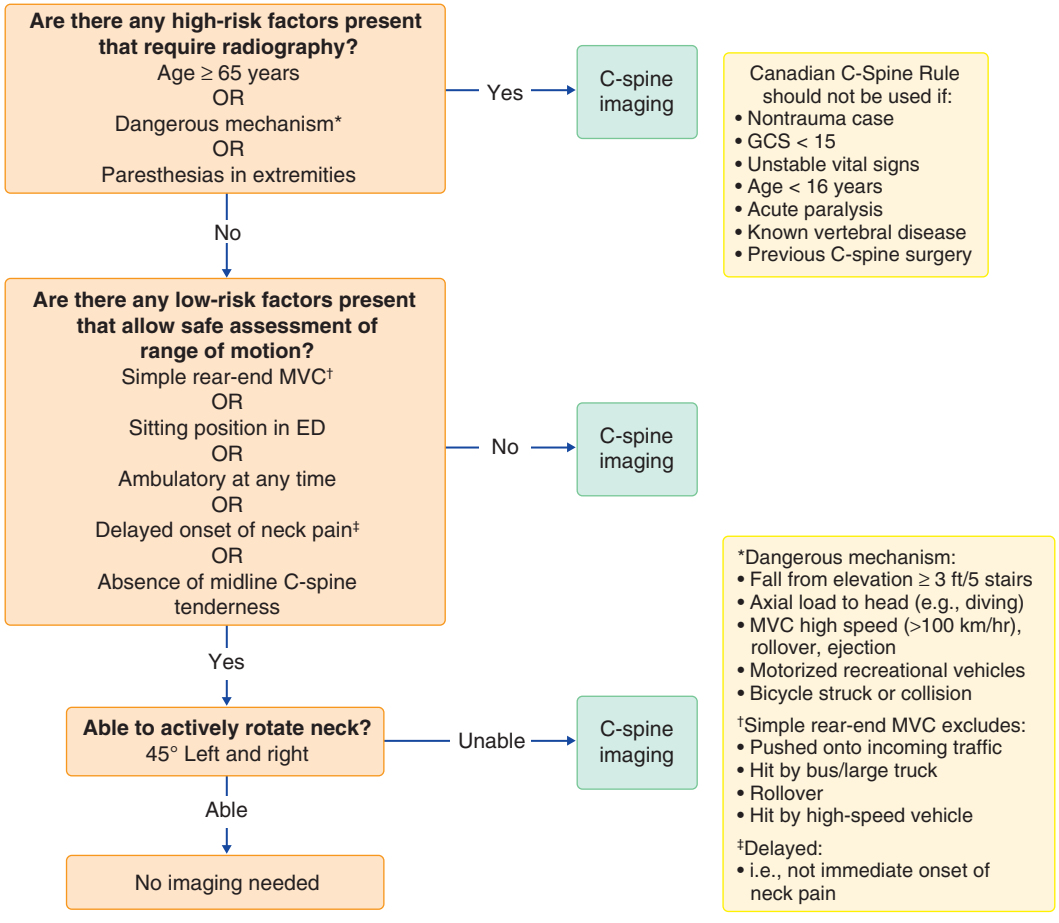

FIGURE 399-3. Canadian C-Spine Rule. For alert (Glasgow Coma Scale ≥ 15) and stable trauma patients in whom cervical spine injury is a concern. ED = emergency department; GCS = Glasgow Coma Scale (see Table 399-2); MVC = motor vehicle collision. (Modified from Stiell IG, Clement CM, McKnight RD, et al. Comparative validation of the Canadian C-Spine Rule and the NEXUS low-risk criteria in alert and stable patients. *N Engl J Med.* 2003;349:2510-2518; and Stiell IG, Wells GA, Vandemheen KL, et al. The Canadian C-Spine Rule for radiography in alert and stable trauma patients. *JAMA.* 2001;286:1841-1848).

identification of the potential for traumatic brain injury, and minimization of secondary insults such as hypoxia and ischemia.[4]

Individuals with a suspected concussion immediately should be removed from play or work[5] and by law should not return until a detailed evaluation with written authorization can be made by an appropriate and experienced physician expert.[6] Patients with mild or moderate traumatic brain injury often have returned to normal or are rapidly recovering by the time they reach advanced medical care. The critical element is the duration of altered mental status, amnesia, or loss of consciousness (see Table 399-4). Longer periods of abnormal sensorium are associated with higher grades of concussion, and higher grades of concussion necessitate longer periods of convalescence.

Severe Traumatic Brain Injury

Patients with Glasgow Coma Scale scores of 8 or less are considered to have severe traumatic brain injury. With this level of impaired consciousness, even with an intact gag reflex, patients are unable to protect their airway adequately. Intubation should be performed with either an endotracheal or a nasotracheal tube, depending on clinical circumstances. The patient should be in a rigid neck collar with the head elevated 30 degrees. The neck collar is used not only to protect the cervical spine until appropriate imaging can be performed but also to keep the head midline to avoid compromising venous drainage.

Certain lesions require prompt surgical intervention, whereas others do not. Penetrating wounds, intracerebral hemorrhage with a mass effect (including subdural and epidural blood), and bone injury (e.g., displaced fracture and vertebral subluxation) require emergency surgical evaluation for intervention. However, focal hypoxic-anoxic, diffuse axonal, and diffuse microvascular injuries do not warrant surgical intervention; treatment remains primarily with the critical care clinician. Skull fractures and intracranial hemorrhages require neurosurgical evaluation. In general, if a fracture is displaced more than the thickness of the skull, it needs to be elevated.

If a surgical lesion is not identified, the patient should be admitted to an ICU. When intracranial hypertension is suspected, a 30-mL intravenous (IV) dose of 23% hypertonic saline through a central venous catheter may be better than mannitol (at a dose of 0.5 to 1 g/kg IV) to reduce it. IV steroids are of no benefit acutely and increase mortality at 2 weeks after the injury. Continuous infusion of 3% hypertonic saline through a central venous catheter may be started at a rate of 75 to 100 mL/hour, with the goal of a serum sodium level of 150 to 155 mM/L to maintain ICP below 20 mm Hg. Hyperventilation also may be tried but has a potential to exacerbate ischemia; if used, the goal should be hyperventilation to a P_{CO_2} of 34 to 36 mm Hg. Induced hypothermia for traumatic brain injury remains controversial, but high-quality randomized trials show no benefit and perhaps even deleterious effects. For closed head injury, a transfusion threshold of 7 g/dL is preferable to a threshold of 10 g/dL.

In addition to ICP control, cerebral perfusion must be maintained. The goal is to maintain cerebral perfusion pressure, which is the difference between mean arterial pressure and ICP, higher than 60 mm Hg. Volume resuscitation is the first therapeutic intervention, with the aim of achieving euvolemia or only slight hypervolemia to a central venous pressure (CVP) goal of 4 to 6 mm Hg. For fluid resuscitation, saline is preferred over albumin. If a cerebral perfusion pressure above 60 mm Hg cannot be achieved with IV fluids alone, vasoactive pharmacologic agents such as norepinephrine (beginning at 2 µg/minute by continuous IV infusion) and phenylephrine (100 µg/minute) may be required. Invasive hemodynamic monitoring with an arterial pressure line and CVP catheter may be needed.

Regular neurologic examinations and appropriate brain imaging are useful to guide ongoing therapy. In one randomized trial, ICP monitoring to maintain a pressure of 20 mm Hg or less was no better than care based on imaging and clinical examination. If the patient is still at a Glasgow Coma Scale score of 8 or less, however, U.S. guidelines recommend that an ICP-monitoring device be used. An intraventricular catheter provides the most reliable data. It is also a treatment option because it allows drainage of cerebrospinal fluid. However, a subdural bolt and fiberoptic catheter are less invasive alternatives.

Pharmacologic Coma and Surgical Decompression

If ICP remains poorly controlled after the aforementioned efforts, pharmacologic coma or surgical decompression is considered. The postulated effect of pharmacologic coma on ICP is through reduction of cerebral metabolism. If the decision to use pharmacologic coma is made, pentobarbital can be administered at a loading dose of 5 mg/kg IV, followed by an infusion of 1 to 3 mg/kg/hour. Another option is propofol (loading dose of 2 mg/kg IV, followed by an infusion of up to 5 mg/kg/hour). Continuous electroencephalographic monitoring is helpful because the target response is burst suppression. Barbiturates and propofol are myocardial depressants, so aggressive cardiovascular management is often necessary to achieve the desired cerebral perfusion pressure.

Recalcitrant elevated ICP despite these interventions is an ominous sign. In such cases, bifrontotemporoparietal craniectomy can reduce ICP and the length of ICU stay but has not been shown to improve outcomes.

Complications

If the patient is agitated, an evaluation should be made to determine whether the patient is in pain or poorly tolerating mechanical ventilation. If pain is a concern, a narcotic analgesic such as fentanyl (50 to 100 µg IV) or morphine (1 to 2 mg IV) should be administered. Because these agents are easily reversed by naloxone, periodic reassessment of neurologic status can be performed. If agitation alone is the issue, haloperidol (0.5 to 2 mg IV), a nonsedating agent that still maintains the ability to perform a neurologic examination, should be considered.

The P_{O_2} level should be maintained at approximately 100 mm Hg. Phenytoin (loading dose of 1000 mg IV, followed by a maintenance dose of 300 mg/day IV) reduces seizures during the first week after traumatic brain injury, but its later usefulness is less clear. Fever greatly increases cerebral metabolism; antipyretic interventions such as acetaminophen and cooling blankets should be used as needed. Gastric stress ulcers may be prevented with H_2 antagonists such as ranitidine (50 mg IV three times daily) or proton pump inhibitors such as omeprazole (20 mg/day orally [PO]). Low-dose heparin (5000 units subcutaneously [SC] twice daily) or a low-molecular-weight heparin such as enoxaparin (40 mg/day SC) and pneumatic stockings should be instituted to avoid deep vein thrombosis. A nasogastric or orogastric tube should be placed for nutrition. Feeding should be initiated as soon as practical, usually on the second day after injury. Because cerebral edema is a concern, hyperosmotic feeding should be instituted. If ileus is present, total parenteral nutrition (TPN; Chapter 217) should be given.

After the first 6 to 12 hours, effort should be made to reduce hyperventilation. Otherwise, the metabolic compensation to chronic hyperventilation negates the ameliorative effects of the respiratory alkalosis.

Regular neurologic examinations and monitoring of ICP and cerebral perfusion pressure are useful to guide ongoing therapy. Generally, the peak period of cerebral edema is from 48 to 96 hours after traumatic brain injury. Thereafter, cerebral edema spontaneously resolves, often associated with clinical improvement.

Recovery

Recovering patients may experience "postconcussive syndrome," which is primarily manifested as headache. Other symptoms may include difficulty concentrating, changes in appetite, sleep abnormalities, and irritability. In general, postconcussive syndrome lasts a few weeks after injury, but it can persist beyond a year or more. Amantadine (100 mg twice daily and increasing up to 200 mg twice daily) can accelerate early recovery after severe traumatic brain injury but may not improve the ultimate outcome.

Therapies are based on the patient's symptoms.[7] For headache, nonsteroidal anti-inflammatory agents (e.g., ibuprofen, 400 to 600 mg PO), migraine drugs (e.g., sumatriptan, 25 to 50 mg PO), and biofeedback may be considered. For cognitive dysfunction, neuropsychological testing may be helpful in determining appropriate intervention. Amantadine (100 mg twice daily) is effective for reducing irritability and aggression in post-head trauma patients with normal renal function.

Traumatic Spinal Cord Injury

Initial Management

Emergency management of traumatic injury to the spinal cord begins with the basics of airway, breathing, and circulation.[8] A secure airway is essential. For patients suffering from high cervical lesions, spontaneous ventilation will be lost. Cervical lesions below C5 may also be associated with impaired ventilatory capability. If there is any concern that the airway or ventilatory effort is compromised, emergency intubation is required. In a patient in whom the cervical spine has not been imaged, the preferred method is nasotracheal intubation under fiberoptic guidance. Other approaches are nasotracheal (blind) or orotracheal intubation, provided in-line traction is applied.

Other immediate concerns are bleeding and circulation. Hypotension may be due to either neurogenic shock or hypovolemia. For neurogenic shock, vasopressive pharmacologic agents such as phenylephrine (beginning as a continuous IV infusion at 100 µg/minute with titration to clinical effect) may be needed. If tachycardia is present, hypovolemia is more likely, so fluid resuscitation would be more appropriate.

Targeted Therapy

Methylprednisolone is no longer advocated for the treatment of acute spinal cord injury. The decision for surgical intervention should be based on the stability of the anterior, middle, and posterior vertebral columns. The anterior column consists of the anterior half of the vertebral body and the vertebral disc. The middle column is the posterior half of the body and the disc. The posterior column is composed of the arch, facets, and ligaments. In general, if two of the three columns are damaged, surgical stabilization is needed. If immediate surgery is not indicated, the patient should be admitted to the ICU for further management.

Acute and Subacute Management

Patients with severe spinal cord injuries require close cardiovascular and ventilatory care, supportive care for bladder and bowel function, approaches to avoid pressure ulcers (Chapter 25), and general measures similar to those used for patients with traumatic brain injury.

Neurogenic Shock and Dysautonomia

After traumatic spinal cord injury, patients are at risk for neurogenic shock and dysautonomia. Lesions of the cervical and thoracic spine disrupt the descending sympathetic pathways to the intermediolateral cell column of the thoracolumbar spinal cord, thereby leading to peripheral vasodilation and hypotension. If the lesion is at T3 or above, sympathetic tone to the heart is compromised. In this setting, hypotension is accompanied by bradycardia, thus producing the neurogenic shock triad of bradycardia, hypotension, and peripheral vasodilation.

Initial therapy for dysautonomia should be fluid administration to restore an adequate circulating volume with a target CVP of 4 to 6 mm Hg. A hematocrit of 30 is optimal for perfusion of the central nervous system, so blood can be used if the patient is anemic. If blood is not required, either colloid (e.g., albumin solutions) or crystalloid (e.g., normal saline) may be used. If there is a suspicion of cardiac or pulmonary disease, a pulmonary artery catheter may be needed briefly to assess fluid status and the relationship between pulmonary pressure and CVP.

Once adequate circulating volume has been achieved, hypotension should be managed with vasopressive agents such as phenylephrine (see earlier), norepinephrine (see earlier), or dopamine (beginning at 1 µg/kg/minute by continuous IV infusion) (Chapter 106), with the goal of a mean arterial pressure of 85 mm Hg or greater. Symptomatic bradycardia can be treated with atropine (1 mg IV).

Ventilatory Compromise

An injury at C5 or higher results in diaphragmatic denervation and requires complete ventilatory assistance.[9] Proper management requires endotracheal or nasotracheal intubation and mechanical ventilation, with an appropriate tidal volume (6 to 10 mL/kg), an FIO_2 to achieve a PO_2 between 80 and 100 mm Hg, and a rate to give a PCO_2 of 40 mm Hg. Positive end-expiratory pressure should also be given to minimize atelectasis (Chapter 90). If the patient does not show signs of ventilatory recovery within 2 weeks of intubation, a tracheostomy should be considered.

Lesions below C5 may also be associated with inadequate spontaneous ventilation. Midcervical lesions may be associated with intact but compromised diaphragm function. If suspected, a "sniff" test under fluoroscopy can be performed to determine whether both hemidiaphragms are functioning properly. If not, intubation/tracheostomy with volume-controlled ventilation may be needed. If intact, pressure support ventilation may be sufficient (Chapter 105) to achieve an appropriate tidal volume.

Cervical lesions at C6 and below spare the phrenic nerves but may disrupt innervation of the intercostal muscles. The primary finding is decreased cough and an inability to increase ventilation when needed, thereby leading to atelectasis and pneumonia; assisted elimination of tracheal secretions is essential.

Thromboembolic Disease

Thromboembolic disease (Chapters 81 and 98) is a leading cause of morbidity and mortality after traumatic spinal cord injury. Prolonged immobility of the lower extremities leads to deep venous thrombosis in up to 70% of spinal cord–injured patients. Patients should receive prophylaxis with low-molecular-weight heparin (e.g., enoxaparin 30 mg twice daily SC) within 72 hours of injury. Anticoagulation can be held on the day of surgery but should be resumed 24 hours after surgery. A less effective alternative is intermittent compression devices (e.g., pneumatic stockings) with low-dose unfractionated heparin. An inferior vena cava filter may be placed if anticoagulation therapy is contraindicated.

Visceral Function

The abdominal wall musculature is innervated by T7 to T12. The stomach, small bowel, liver, pancreas, and proximal two thirds of the colon receive innervation from T5 to L2. Spinal cord injury at these levels or above may impair visceral function. For ileus, a nasogastric tube should be placed to decompress the stomach. Parental nutrition should be started as soon as possible. Enteral feeding should be delayed until gastrointestinal motility returns, usually within 2 to 3 weeks. In comparison with conservative bowel management, transanal irrigation improves constipation, fecal incontinence, and symptom-related quality of life in patients with spinal cord-injuries.

Stress-induced peptic ulcer disease occurs in nearly a third of patients without prophylaxis. H_2-receptor antagonists such as ranitidine (50 mg IV three times daily) or a proton pump inhibitor such as omeprazole (20 mg/day PO) reduce the incidence of ulcers.

Bladder tone may be lost because of spinal shock. A Foley catheter should be placed for a minimum of 5 to 7 days to drain the bladder and evaluate volume and renal status. After spinal shock has resolved, autonomic dysreflexia may occur as a result of bladder distention. Clinical signs such as sweating, skin flushing, and hypertension may be present. Clinical examination with palpation and percussion will reveal a distended bladder, which can be treated by bladder training or intermittent catheterization.

Nutrition

Until enteral feeding can begin, parenteral nutrition should be used. Ideally, TPN should be started. However, if TPN is not possible, peripheral parenteral nutrition should be used until TPN (Chapter 217) can begin. Energy expenditures of 19 kcal/kg/day for high cervical injuries to 35.8 kcal/kg/day for injuries at T10 and below have been reported. A caloric level of 80% of the Harris-Benedict prediction should be used for quadriplegic patients. The full Harris-Benedict predicted amount should be used in patients with thoracic spine injuries and below. Indirect calorimetry should be used to determine the caloric needs of each patient so as to optimize nutritional support.

Other Therapy

In a randomized trial, pregabalin, 150 to 600 mg/day, was effective in reducing central neuropathic pain after spinal cord injury. Patients with traumatic spinal cord injury have a propensity for the development of decubitus ulcers and pressure sores (Chapter 25). Mechanical kinetic beds, regular log rolling (every 2 hours), and padded orthotics are all useful in minimizing this complication. Orthotics, physical therapy, and occupational therapy (for cervical cord injury) are also important to minimize contractures and begin the rehabilitation process.

PROGNOSIS

Traumatic Brain Injury

The most useful prognostic indicator after traumatic brain injury is the neurologic examination at initial evaluation. For patients with severe traumatic brain injury, the initial Glasgow Coma Scale score is the most reliable prognostic indicator. The lower the initial Glasgow Coma Scale score, the less likely a patient will have meaningful neurologic or functional recovery. After traumatic brain injury, 40% of patients with a score of 8 have a good recovery versus only 7% when the score is 3. Furthermore, only 27% of patients with a score of 3 survive versus 88% of patients with a score of 8. Patients in whom the Glasgow Coma Scale score remains the same or worsens over a period of 6 hours do worse clinically than those whose score improves. Further prognostic stratification at 24 hours can be based on pupillary responses, motor responses, and age (Chapter 404). Substantial increases of CSF α-synuclein may indicate widespread neurodegeneration and reflect secondary neuropathologic events after severe traumatic brain injury.[10]

A subsequent head injury before full recovery from even a mild traumatic brain injury may occasionally result in "second impact syndrome," which can worsen the clinical outcome. When seen (mostly in children and adolescents), coma develops rapidly after the second injury, often within minutes. There is decreased autoregulation, diffuse cerebral edema, and intracranial hypertension. Second impact syndrome is associated with high mortality.

Traumatic Spinal Cord Injury

For traumatic spinal cord injury, the completeness of the injury is the most useful predictor (Table 399-5). A grade "A" or complete motor and sensory deficit below the lesion has a poor prognosis. If such a lesion persists for 24 hours, there is little likelihood of meaningful recovery. On the other hand, even severe partial injuries have a higher probability of recovery.

TABLE 399-5 AMERICAN SPINAL INJURY ASSOCIATION IMPAIRMENT SCALE

GRADE	INJURY TYPE	DEFINITION	LIKELIHOOD OF RECOVERY*
A	Complete	No motor or sensory function below the lesion	15.5% (cervical) and 7% (thoracic)
B	Incomplete	Sensory but no motor function	47%
C	Incomplete	Some motor strength (<3)	84%
D	Incomplete	Motor strength > 3	84%
E	None	Sensory and motor function normal	100%

*Data from Coleman WP, Geisler FH. Injury severity as primary predictor of outcome in acute spinal cord injury: retrospective results from a large multicenter clinical trial. *Spine J.* 2004;4:373-378.

Grade A References

A1. Stiell IG, Clement CM, Grimshaw J, et al. Implementation of the Canadian C-Spine Rule: prospective 12 centre cluster randomised trial. *BMJ*. 2009;339:b4146.

A2. Wakai A, McCabe A, Roberts I, et al. Mannitol for acute traumatic brain injury. *Cochrane Database Syst Rev*. 2013;8:CD001049.

A3. Edwards P, Arango M, Balica L, et al. Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury–outcomes at 6 months. *Lancet*. 2005;365:1957-1959.

A4. Georgiou AP, Manara AR. Role of therapeutic hypothermia in improving outcome after traumatic brain injury: a systematic review. *Br J Anaesth*. 2013;110:357-367.

A5. Robertson CS, Hannay HJ, Yamal JM, et al. Effect of erythropoietin and transfusion threshold on neurological recovery after traumatic brain injury: a randomized clinical trial. *JAMA*. 2014;312:36-47.

A6. Myburgh J, Cooper DJ, Finfer S, et al. Saline or albumin for fluid resuscitation in patients with traumatic brain injury. *N Engl J Med*. 2007;357:874-884.

A7. Chesnut RM, Temkin N, Carney N, et al. A trial of intracranial-pressure monitoring in traumatic brain injury. *N Engl J Med*. 2012;367:2471-2481.

A8. Timmons SD, Ullman JS, Eisenberg HM. Craniectomy in diffuse traumatic brain injury. *N Engl J Med*. 2011;365:373.

A9. Cooper DJ, Rosenfeld JV, Murray L, et al. Decompressive craniectomy in diffuse traumatic brain injury. *N Engl J Med*. 2011;364:1493-1502.

A10. Giacino JT, Whyte J, Bagiella E, et al. Placebo-controlled trial of amantadine for severe traumatic brain injury. *N Engl J Med*. 2012;366:819-826.

A11. Hammond FM, Bickett AK, Norton JH, et al. Effectiveness of amantadine hydrochloride in the reduction of chronic traumatic brain injury irritability and aggression. *J Head Trauma Rehabil*. 2014;29:391-399.

A12. Christensen P, Bazzocchi G, Coggrave M, et al. A randomized, controlled trial of transanal irrigation versus conservative bowel management in spinal cord-injured patients. *Gastroenterology*. 2006;131:738-747.

A13. Cardenas DD, Nieshoff EC, Suda K, et al. A randomized trial of pregabalin in patients with neuropathic pain due to spinal cord injury. *Neurology*. 2013;80:533-539.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Selvarajah S, Hammond ER, Haider AH, et al. The burden of acute traumatic spinal cord injury among adults in the United States: an update. *J Neurotrauma*. 2014;31:228-238.
2. Scholten AC, Haagsma JA, Panneman MJ, et al. Traumatic brain injury in the Netherlands: incidence, costs and disability-adjusted life years. *PLoS ONE*. 2014;9:e110905.
3. Stern RA, Daneshvar DH, Baugh CM, et al. Clinical presentation of chronic traumatic encephalopathy. *Neurology*. 2013;81:1122-1129.
4. Hodgkinson S, Pollit V, Sharpin C, et al. Early management of head injury: summary of updated NICE guidance. *BMJ*. 2014;348:g104.
5. Putukian M, Raftery M, Guskiewicz K, et al. Onfield assessment of concussion in the adult athlete. *Br J Sports Med*. 2013;47:285-288.
6. Giza CC, Kutcher JS, Ashwal S, et al. Summary of evidence-based guideline update: evaluation and management of concussion in sports: report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology*. 2013;80:2250-2257.
7. Marshall S, Bayley M, McCullagh S, et al. Clinical practice guidelines for mild traumatic brain injury and persistent symptoms. *Can Fam Physician*. 2012;58:257-267.
8. Joint Committee of the AANS and CNS. Guidelines for the management of acute cervical spine and spinal cord injuries. *Neurosurgery*. 2013;72(suppl 2):1-259.
9. Tester NJ, Fuller DD, Fromm JS, et al. Long-term facilitation of ventilation in humans with chronic spinal cord injury. *Am J Respir Crit Care Med*. 2014;189:57-65.
10. Mondello S, Buki A, Italiano D, et al. alpha-Synuclein in CSF of patients with severe traumatic brain injury. *Neurology*. 2013;80:1662-1668.

400 MECHANICAL AND OTHER LESIONS OF THE SPINE, NERVE ROOTS, AND SPINAL CORD

RICHARD L. BARBANO

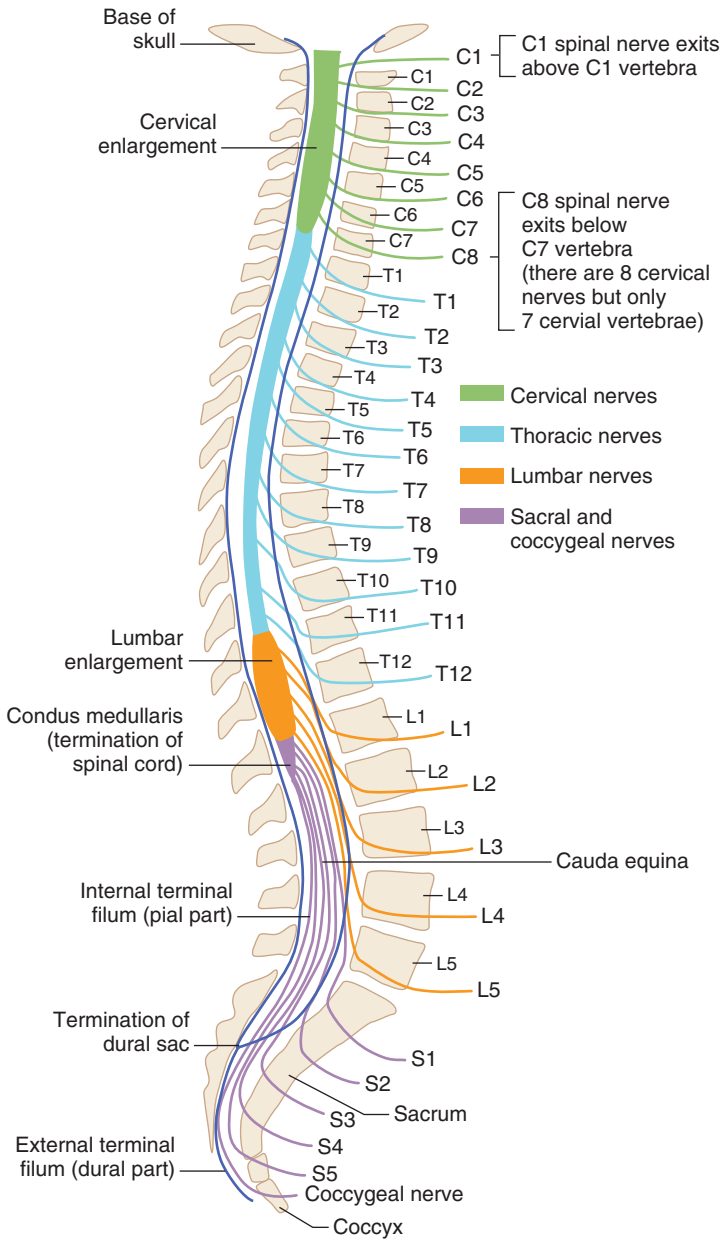

FIGURE 400-1. Anatomy of the spinal cord.

DEFINITION

Disorders of the spine, nerve roots, and spinal cord are frequent reasons for a patient to visit a physician. Many of these disorders either initially or eventually involve more than one element of the vertebra–spinal cord–nerve root unit, so there is much overlap in the pathobiology and clinical manifestations of these diseases.

The spine consists of 30 vertebrae: 7 cervical (C), 12 thoracic (T), 5 lumbar (L), 5 sacral (S), and the coccyx (Fig. 400-1). The ring shape of the bony vertebrae forms a protective circle around the spinal cord while leaving ample room to allow the cord to move within this canal during flexion and extension of the spine. The vertebral bodies help bear the compressive weight of the body and provide the surface area to support the intervertebral discs, which act to cushion the axial force along the spine. The overlapping facet joints and multiple sets of longitudinal ligaments give the spine stability during its many ranges of motion. The posteriorly placed foramina allow the exit of spinal nerves.

The spinal cord consists of 31 spinal segments, with one more cervical cord segment (8) than vertebrae; each gives rise to a bilateral pair of spinal nerves. Spinal nerves C1 to C7 exit the canal above their corresponding vertebral body, the C8 nerve exits below the C7 vertebra, and subsequent inferior nerves also exit below the numbered vertebrae. The spinal segments of the cord itself, however, lie progressively superior to the vertebrae, so that the end of the spinal cord, the conus medullaris, in adults is approximately adjacent to the L1 vertebra. The more caudal spinal nerves travel as the cauda equina in the subarachnoid space within the spinal canal before exiting their respective foramina. The spinal cord does not have a uniform diameter; the cervical and lumbar segments are wider compared with the thoracic and lower sacral areas because the increased motor and sensory neurons supplying the arms and legs enlarge the cord.

Spinal nerves are formed by the joining of the anterior and posterior spinal roots, which directly exit and enter the spinal cord. The anterior root derives from axons of the anterior horn cells and lateral columns, and it serves motor and autonomic efferent pathways; the posterior root mostly derives from the axons from the dorsal root ganglion and carries afferent sensory signals (Fig. 400-2). The sensory root is twice the thickness of the motor root and lies in a more anterior and inferior location as it crosses the foramen.

CLINICAL MANIFESTATIONS

Disorders of the spinal nerve root produce signs and symptoms referable to the corresponding dermatome or myotome. By far the most frequent complaint is localized neck or back pain, but compromise of the nerve roots or spinal cord will cause symptoms such as abnormal or painful sensations (paresthesias or dysesthesias), loss of sensation, weakness, and autonomic dysfunction (most commonly bladder or bowel incontinence).

When it affects a myotome (the group of muscles served by motor neurons of a spinal cord segment; Fig. 400-3), the motor deficit associated with a spinal root disorder is of the lower motor neuron type. Typical findings are weakness, hypotonia, depressed or absent reflexes, and, if the syndrome has

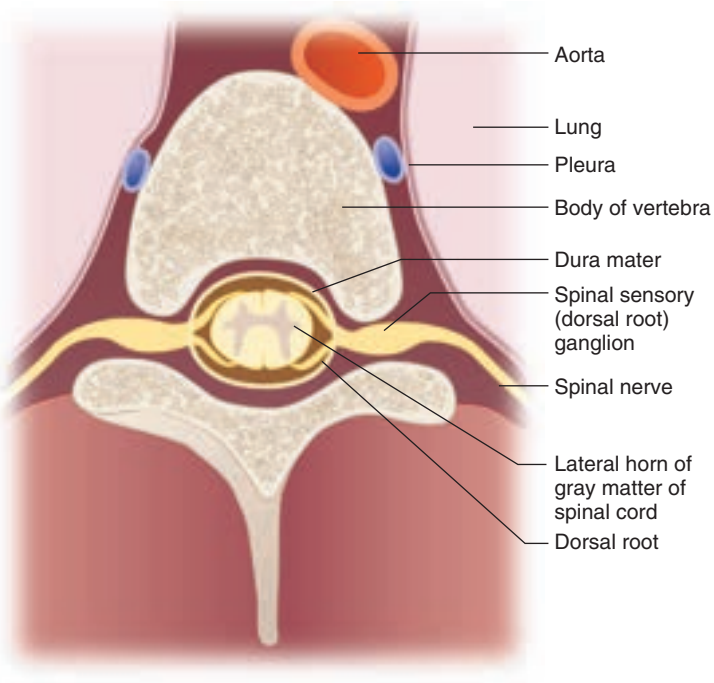

FIGURE 400-2. **Anatomy of the spinal cord: section through a thoracic vertebra.**

persisted for at least several weeks, atrophy with or without fasciculations. Sensation at the root level is diminished or absent for all modalities, but sensation below the affected root level is intact.

Conversely, disorders of the spinal cord produce a "level" below which sensation is abnormal and motor deficits are of the upper motor neuron type, with weakness without atrophy (unless complicated by disuse), hypertonia, and increased reflexes. At the level of a spinal cord lesion, the motor deficits can be of the lower motor neuron type as the anterior horn cell bodies or exiting fibers are affected; below this level, an upper motor neuron syndrome will predominate. With strokes (Chapter 406) and other central nervous system (CNS) disorders, the full upper motor neuron syndrome may not be present in the acute phase of cord injury and can take time to appear.

DIAGNOSIS

The clinical history can help localize the patient's symptoms, especially complaints of pain and sensory alterations that may exist in the absence of objective sensory loss on probing with light touch, pinprick, and vibration stimuli. The neurologic examination should include evaluation of the sensory, motor (Table 400-1), and reflex (Table 400-2) functions. Careful side-to-side comparisons can help assess subtle deficits. For example, elderly patients often have decreased or absent ankle jerks, so contralateral comparison is necessary. However, all muscles receive innervation from more than one root/spinal cord level, and all roots send fibers to multiple muscles. The clinical implication of this anatomic pattern is that individual muscles are rarely profoundly weak, and patients rarely report isolated muscle weakness in single root involvement syndromes. Likewise, overlap of the sensory dermatomes (see Fig. 400-3) explains why sharp demarcations in the sensory examination rarely occur.

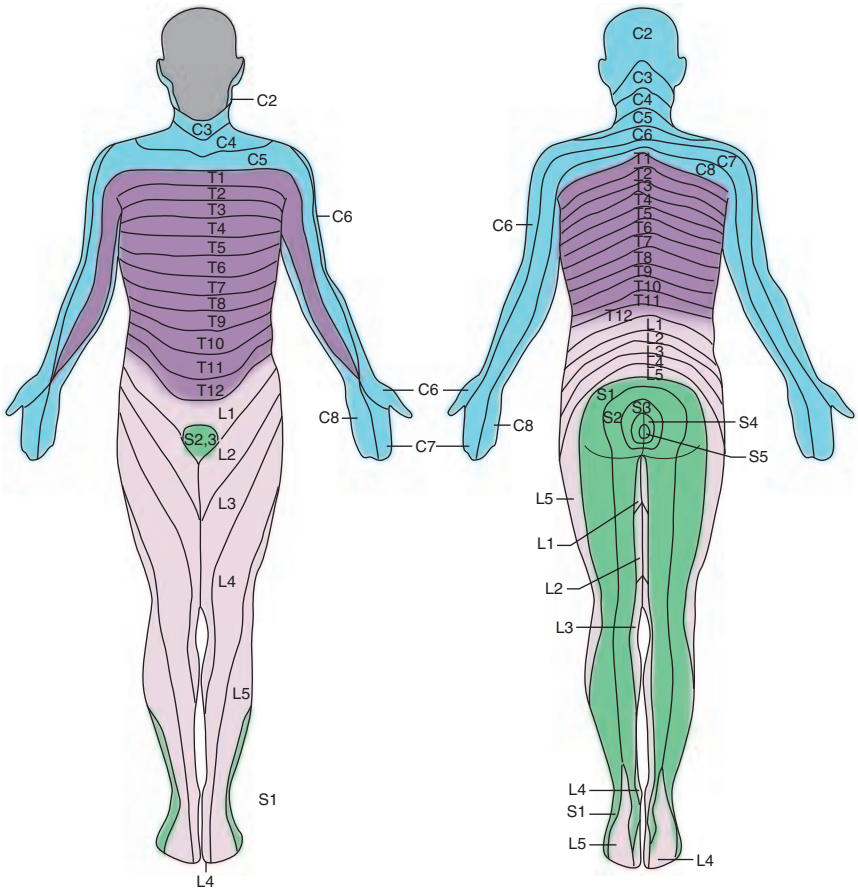

Levels of principal dermatomes

C5	Clavicles
C5,6,7	Lateral parts of upper limbs
C8, T1	Medial sides of upper limbs
C6	Thumb
C6,7,8	Hand
C8	Ring and little fingers
T4	Level of nipples
T10	Level of umbilicus
T12	Inguinal or groin regions
L1,2,3,4	Anterior and inner surfaces of lower limbs
L4,5 S1	Foot
L4	Medial side of great toe
S1,2, L5	Posterior and outer surfaces of lower limbs
S1	Lateral margin of foot and little toe
S2,3,4	Perineum

FIGURE 400-3. **Schematic demarcation of levels of principal dermatomes shown as distinct segments.** There is actually considerable overlap between any two adjacent dermatomes.

TABLE 400-1 ESSENTIAL MUSCLE TESTING

ROOT	MUSCLES	ACTION/TESTING
ARM		
C5	Deltoid Infraspinatus	Abducts arm Externally rotates arm with elbow flexed
C6	Brachioradialis	Flexes elbow (along with biceps [C5-6])
C7	Triceps Extensor digitorum	Extends elbow Extends fingers
C8	Flexor digitorum Flexor pollicis longus	Flexes fingers (both superficialis and profundus) Flexes distal phalanx of thumb
T1	Interossei	Spread fingers
LEG		
L1-2	Iliacus	Flexes hip
L2-3	Adductor magnus	Adducts thigh (as part of adductor group)
L3-4	Quadriceps	Extends knee
L4	Tibialis anterior	Dorsiflexes foot
L5	Extensor hallucis longus Extensor digitorum longus	Great toe extension Toe extension
S1	Hamstrings Flexor hallucis longus	Flex knee Flexes great toe (along with S2)

TABLE 400-2 ESSENTIAL REFLEX EXAMINATION

REFLEX	EFFECT	ROOT/IMPLICATION
Jaw jerk	Jaw closes with tap on slightly opened jaw	Cranial nerve V; implies lesion above cervical cord
Biceps	Tap tendon: elbow flexion	C5-6; musculocutaneous nerve
Brachioradialis	Tap tendon over distal radius with elbow flexed and mid pronation: elbow flexion	C5-6; radial nerve
Triceps	Tap tendon: elbow extension	C6 < C7
Finger flexion	Tap partially flexed fingertips: finger flexion	C6-T1; when hyperactive, may imply lesion above midcervical spinal cord
Patella	Tap patella: quadriceps contraction (knee extension)	L2-4
Achilles	Tap Achilles tendon: foot plantar flexion	S1-2
Babinski	Scratch sole: great toe flexion	Great toe extension implies lesion of cord (or brain) above L4
Anal wink	Scratch perineum: external anal sphincter contraction	Absence of contraction implies S2-4 lesion

DISORDERS OF THE SPINE

Neck and Back Pain

DEFINITION

Most neck and back pain is mechanical, that is, theoretically emanating from the spine's structural elements, which are the vertebrae, discs, ligaments, tendons, and muscles. The location of pain is either axial, which means it is located along the spine itself, or referred. The term *perceived pain* is sometimes used when the pain from a spinal lesion is felt elsewhere by the patient, whereas *referred pain* is often used to describe pain that is experienced by the patient in the spinal area but caused by nonspinal structures. If the pain follows a dermatomal (nerve root) distribution, it is referred to as *radicular pain*, which is likely to involve the nerve root.

EPIDEMIOLOGY

Neck and back pain is a frequent reason for visits to a primary care physician. Low back pain is more common than neck pain, but both are common. The thoracic spine, possibly because of rib attachments and limited range of motion, is an uncommon location for back pain. An exception to this general rule is the condition of diffuse idiopathic skeletal hyperostosis (DISH; Chapter 273), which is a noninflammatory age-related condition of unknown etiology, characterized by ossification of paravertebral ligaments and peripheral entheses. It is more common in men, with a prevalence of 30% in men by age 65. Pain in the thoracic spine region occurs in up to 80% of patients and is accompanied by notable decreased range of motion.

In the general population, the incidence of self-reported neck pain is 213 per 1000; the 12-month prevalence of any pain is typically between 30 and 50%, and pain severe enough to limit activity is between 1.7 and 11.5%. The prevalence is higher among women. Risk factors for neck pain include inherited factors, poor psychological health, and tobacco use; the presence of disc degeneration is not a significant factor.

More than 70% of people will experience low back pain significant enough to inhibit their participation in daily activities at some time in their life. The highest prevalence is in the 45- to 64-year age group. There is less of a gender difference than with neck pain, although tobacco use is an associated risk factor. Physical work–related factors (e.g., heavy lifting, prolonged sitting, repetitive twisting) increase risk; prospective studies show that psychosocial issues such as work monotony and job dissatisfaction also are major predisposing factors.

PATHOBIOLOGY

Although the popular impression is that the disc is the source of most spine pain, it is estimated that disc disease such as protrusion accounts for only 5% of all low back problems. Degenerative changes are a much more common cause of both acute and chronic spine pain. There is a genetic predisposition to intervertebral disc degeneration, with heritability estimates in the range of 34 to 61%.

Degenerative changes can result in spondylosis, a condition that includes degenerative disc disease with bulging and occasionally herniation. The condition is often accompanied by the formation of osteophytes, ligamentous hypertrophy, and sometimes facet fracture and vertebral subluxation. Spondylosis, which is a consequence of age-related disc disease, is exemplified by the fact that almost everybody has at least anterior osteophytes by the age of 40 years, and it is not necessarily painful. Spondylosis probably starts with age-related disc desiccation and loss of elasticity of the annulus fibrosus. Tension of the longitudinal ligaments results in the formation of hypertrophic osteophytes. Compromise of microvascular supply may also contribute. Eventually, the facet joints can ride over one another, thereby leading to instability, formation of more osteophytes, and inflammation of the synovial joints. If there is fracture of the pars interarticularis, the term *spondylolysis* is used. Further instability between the intervertebral segments leads to spondylolisthesis, in which one vertebral body shifts sagittally in relation to its adjacent vertebra. Spondylolisthesis is graded by the amount of shift as measured with flexion and extension lateral spine films.

DISH is characterized by calcification and ossification of the anterior longitudinal spinal ligament and less frequently the posterior longitudinal ligament. The latter can compromise spinal cord roots as well as the spinal cord itself, especially in the cervical region.[1]

Whiplash, an acute flexion-extension injury of the cervical spine, is common after motor vehicle accidents and other situations of rapid deceleration. Although specific acute and chronic manifestations are controversial, the acute syndrome is generally accepted to be a result of mechanical irritation of pain-sensitive structures in the cervical spine, with or without nerve root injury. More severe trauma can cause fracture and vertebral instability, both of which require rapid surgical evaluation.

CLINICAL MANIFESTATIONS

Acute neck pain and low back pain are commonly limited to the axial region, although radicular signs and symptoms can occur in the presence of nerve root irritation. The most common radicular pain occurs in the distribution of a dermatome. Other radicular signs and symptoms can include dysesthesias or sensory loss in the affected dermatome, decreased strength in muscles of the affected myotome, and decreased reflex. Cranial nerve findings, diffuse weakness throughout a limb or in more than one limb, hemisensory symptoms, autonomic symptoms, and increased reflexes are not manifestations of spine disease and should prompt more extensive evaluation for other conditions that affect the brain or brain stem. Bowel and bladder symptoms should prompt urgent evaluation of a cauda equina or myelopathy syndrome (see later).

Acute spine problems can also cause referred or perceived pain at sites other than their anatomic source. For example, mechanical low back pain may

include aching in the buttocks or thigh, most often posteriorly and occasionally the hip region but rarely below the knee. More commonly, however, the term *referred pain* denotes the situation in which other structures, usually internal organs, refer pain to the spine or back. Areas of referred pain usually share the same embryologic origin and, during development, the same sensory pathways. Differentiation of referred pain from localized back pain depends on the history and examination. Mechanical pain is often exacerbated by movement such as twisting, bending, extension, or flexion, whereas referred pain tends to be independent of such activities.

Chronic spine disorders lead to chronic back pain directly and as a secondary complication. For example, chronic degenerative arthropathy can lead to degenerative lumbar scoliosis with secondary involvement of neural structures. Back pain and radiculopathy are the most prominent symptoms, present in upward of 80% of patients, but symptoms of neurogenic claudication also develop in about 50% of patients.

DIAGNOSIS

History and Clinical Examination

The history and examination are essential for the initial evaluation and triage of patients with neck and back pain. Patients with so-called red flags (Table 400-3) merit special attention, as does any patient who awakens from sleep because of pain or has pain that is constant and unchanged by position, is unremitting and progressive, or is accompanied by any systemic signs or symptoms.

As part of the history in the setting of acute neck trauma, well-established screening protocols such as the Canadian C-Spine Rule and the NEXUS Low-Risk Criteria (Fig. 400-4) are validated ways to detect cervical spine fracture and direct appropriate radiographic evaluation. In such a setting, a computed tomographic (CT) scan of the cervical spine would be the imaging test of choice.

On clinical examination, inspection should assess evidence of trauma, muscle wasting, fasciculations, erythema, rashes, and scars. Palpation is directed to areas of point tenderness during evaluation for more diffusely tender regions, muscle spasm, and masses. If light percussion of the spinous process evokes significant pain, a focal process, such as fracture, malignant neoplasm (Chapter 189), or infection (Chapter 413), should be considered because such a finding is unusual in typical mechanical spine pain. Finally, the active and passive range of motion for flexion, extension, rotation, and tilt should be noted. Many provocative tests have been described for the evaluation of neck and back pain, but few have undergone formal evaluation of their diagnostic accuracy. For neck pain, contralateral rotation of the neck with extension of the arm and fingers (Video 400-1) suggests cervical root involvement, particularly in combination with other provocative tests, such as the Spurling maneuver, in which the patient's head is rotated 45 degrees to the contralateral side, with the neck in slight extension to minimize the foraminal opening. Downward pressure on the top of the head by the examiner will reproduce arm dysesthesias (Video 400-2). Provocative tests also can diminish symptoms. For example, in the cervical distraction test, the examiner's hands are placed under the jaw and occiput; gentle upward pulling of the head will temporarily reduce or alleviate the symptoms (Video 400-3).

For low back pain, the straight leg raise (Video 400-4) has sensitivity of 0.85 to 0.91 but a specificity of only 0.26 to 0.52 for the diagnosis of sciatica due to a herniated disc. The crossed straight leg raise test (Video 400-5) has a lower sensitivity of 0.23 to 0.34 but a much higher specificity of 0.86 to 0.90. The seated straight leg raise (Video 400-6) can be used for confirmation of root irritation as the spine-leg angle is increased to 90 degrees. A negative result of the seated straight leg raise in the setting of a positive result of the straight leg raise suggests the possibility of a nonorganic component, although a mechanical alteration of the root exit zone in this position should also be considered.

Ancillary Testing

For neck pain (Fig. 400-5), plain radiography and CT scanning, which are the mainstays of cervical spine imaging, allow adequate view of the bony structures. Magnetic resonance imaging (MRI) has largely replaced myelography, which is still used occasionally to provide information about the spinal cord and nerve roots. However, MR abnormalities are common and can have a high false-positive rate; for example, 12 to 17% of patients younger than 30 years and 86 to 89% of patients aged 60 have disc degeneration as evidenced by loss of signal intensity, disc protrusion, narrowing of the disc space, or foraminal stenosis. Cervical discography also has a high false-positive rate and cannot be recommended as a diagnostic test in the assessment of neck pain.

Uncomplicated acute low back pain, with or without radiculopathy, is generally self-limited, and imaging studies are unnecessary unless any of the red flags (see Table 400-3) are present.[1] The American College of Physicians recommends MRI only in patients who have major or progressive neurologic deficits, in whom a serious underlying condition is expected, or in whom surgery or epidural steroids are being considered.[2] For trauma, osteoporosis, or patients older than 70 years, plain radiography may suffice if the results are normal and no other abnormalities are present. Otherwise, with few exceptions, MRI is the test of choice given its superiority in evaluating soft tissue structures and its lack of radiation exposure. Care must be taken, however, to ensure correlation with the clinical syndrome, because 28% of asymptomatic volunteers with a mean age of 42 have herniated discs, 52% have bulging discs, and 14% have annular tears. The percentage of imaging abnormalities increases even more in asymptomatic volunteers older than 60 years; 57% have significantly abnormal scans, with 36% showing herniated discs and close to 98% showing disc degeneration. Abnormalities of the Modic end plate, anterolisthesis, and disc extrusion are more strongly associated with

TABLE 400-3 "RED FLAGS" IN THE EVALUATION OF SPINE PAIN

Recent significant trauma or minor trauma at age > 50 years
Unexplained weight loss
Unexplained fever
Immunosuppression
History of cancer
History of prior local surgery
Systemic disorder, bone or arthritic disorder
Intravenous drug use
Prolonged use of corticosteroids or osteoporosis
Age > 70 years
Focal neurologic deficit with progressive symptoms
Duration > 6 weeks
Thoracic spine pain

Modified from Davis PC, Wippold FJ, Brunberg JA, et al. ACR Appropriateness Criteria on low back pain. *J Am Coll Radiol.* 2009;6:401-407.

The NEXUS Low Risk Criteria (NLC) Algorithm for screening of neck injuries

- No posterior midline cervical spine tenderness—Midline posterior bony cervical-spine tenderness is present if the patient reports pain on palpation of the posterior midline neck from the nuchal ridge to the prominence of the first thoracic vertebrae, or if the patient evinces pain with direct palpation of any cervical spinous process.
- No evidence of intoxication—Patients should be considered intoxicated if they have either of the following: a recent history provided by the patient, or an observer of intoxication or intoxicating ingestion, or evidence of intoxication on physical examination such as an odor of alcohol, slurred speech, ataxia, dysmetria, or other cerebellar findings, or any behavior consistent with intoxication. Patients may also be considered to be intoxicated if tests of bodily secretions are positive for alcohol or drugs that affect level of alertness.
- A normal level of alertness—An altered level of alertness can include the following: a Glasgow Coma Scale score of 14 or less; disorientation to person, place, time, or events; an inability to remember three objects at five minutes; a delayed or inappropriate response to external stimuli; or other findings.
- No focal neurological deficit—A focal neurological deficit is any focal neurological finding on motor or sensory examination.
- No painful distracting injuries—No precise definition of painful distracting injury is possible. This category includes any condition thought by the clinician to be producing pain sufficient to distract the patient from a second (neck) injury. Such injuries may include, but are not limited to, any long-bone fracture; a visceral injury requiring surgical consultation; a large laceration, degloving injury, or crush injury; large burns; or any other injury causing acute functional impairment. Physicians may also classify any injury as distracting if it is thought to have the potential to impair the patient's ability to appreciate other injuries.

FIGURE 400-4. The NEXUS Low-Risk Criteria (NLC) algorithm for screening of neck injuries. (Reproduced from Hoffman JR, Mower WR, Wolfson AB, et al. Validity of a set of clinical criteria to rule out injury to the cervical spine in patients with blunt trauma. National Emergency X-Radiography Utilization Study Group. *N Engl J Med.* 2000;343:94-99.)

VIDEO 400-1. Cervical Rotation as provocative maneuver to elicit radiculopathic symptoms

VIDEO 400-2. The Spurling Maneuver to elicit radiculopathic symptoms

VIDEO 400-3. Cervical distraction maneuver:relief of radiculopathic symptoms

VIDEO 400-4. Straight Leg Raise: provocative maneuver for lumbar radiculopathy

VIDEO 400-5. Crossed Straight Leg Raise: provocative maneuver for lumbar radiculopathy

VIDEO 400-6. Seated Straight Leg Raise: Provocative maneuver for lumbar radiculopathy

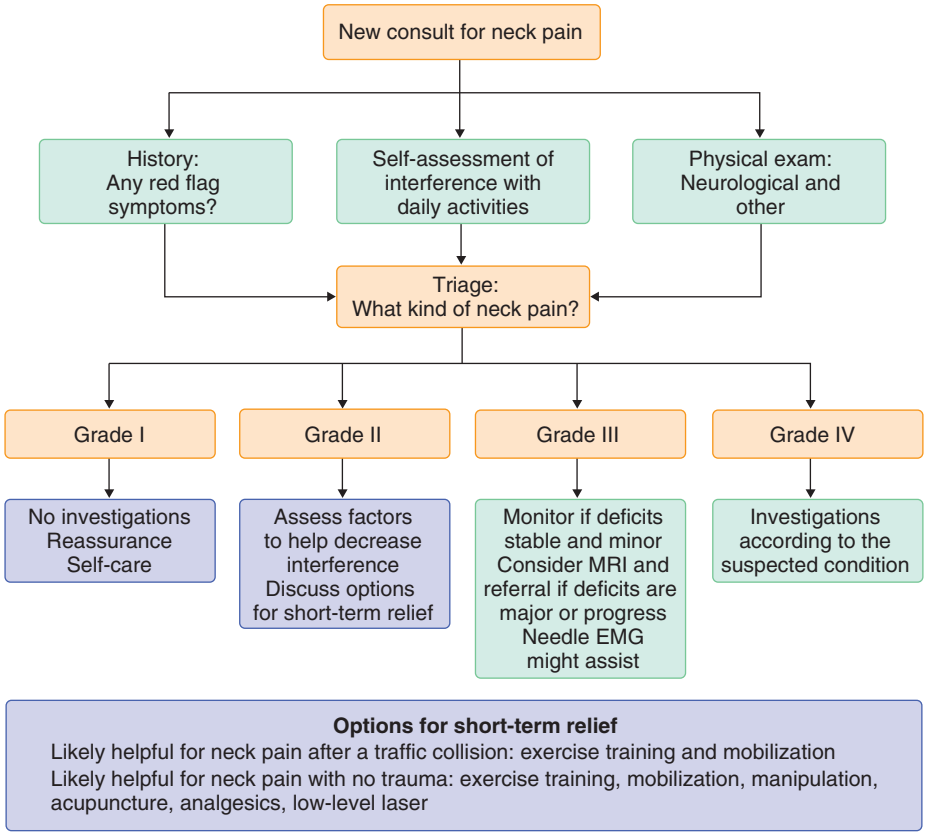

FIGURE 400-5. Approach to new-onset neck pain. EMG = electromyography; MRI = magnetic resonance imaging. (Modified from Guzman J, Haldeman S, Carroll LJ, et al. Clinical practice implications of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. From concepts and findings to recommendations. *Spine.* 2008;33:S199-S213.)

TABLE 400-4 MECHANICAL NECK PAIN

	NECK STRAIN	HERNIATED NUCLEUS PULPOSUS	OSTEOARTHRITIS	MYELOPATHY	WHIPLASH
Age (yr)	20-40	30-50	>50	>60	30-40
Pain location	Neck	Arm	Neck	Arm/leg	Neck
Onset	Acute	Acute	Insidious	Insidious	Acute
Flexion	+	+	–	–	+
Extension	–	+/–	+	+	+
Plain radiography	–	–	+	+	–

+ = present; – = absent.
From Borenstein DG, Wiesel SW, Boden SD. *Neck Pain: Medical Diagnosis and Comprehensive Management.* Philadelphia: WB Saunders; 1996.

low back pain than is disc degeneration without end plate changes. Situations in which alternative imaging should be considered include spondylosis and stress fracture, for which bone scintigraphy with single-photon emission CT (SPECT) is more sensitive than MRI. CT can also be useful when MRI is contraindicated or to evaluate scoliosis, bone graft integrity, surgical fusion, and instrumentation. For back and neck pain that persists for 6 weeks, electrodiagnostic testing can demonstrate compromise of spinal root function but is not usually helpful in axial spine pain without neurologic symptoms.

Differential Diagnosis

Mechanical or idiopathic pain explains up to 97% of cases of neck pain (Table 400-4) and low back pain (Table 400-5); the remaining 3% is nonmechanical in origin and includes referred pain and other conditions. Acute mechanical neck pain is most often caused by a neck strain, a herniated nucleus pulposus, or whiplash; for pain of insidious onset, osteoarthritis and myelopathy are the leading causes. For back pain, muscle strain and a herniated nucleus pulposus are acute causes; insidious causes include osteoarthritis, spinal stenosis, spondylolisthesis, and scoliosis. Queries regarding the red flags will identify serious and nonmechanical causes of neck and back pain (Fig. 400-6).

Abdominal and pelvic structures can refer pain to the low back (referred pain). Abdominal aortic aneurysms (Chapter 78) can present with a mid- to low back ache that may radiate to the hips or anterior thighs. Cholecystitis (Chapter 155) can cause pain in the midthoracic area; pancreatic disease (Chapter 144) can cause pain in the L1 region; and diverticulitis (Chapter 142) in the left lower quadrant can cause diffuse low back pain. Genitourinary disorders (Chapter 123) can cause colicky referred pain to the flanks and costovertebral angle. Bladder disorders (Chapter 123) may occasionally refer pain to the sacral area, as can prostate problems (Chapter 129). Pelvic disorders in women that can cause referred low back pain include endometriosis (Chapter 236), ectopic pregnancy, and pelvic inflammatory disease (Chapters 299 and 318). Most of these disorders have additional signs and symptoms to aid in the diagnosis.

Myocardial ischemia (Chapters 71 to 73) can be associated with anterior neck pain, although less commonly than with left arm or jaw pain. Arterial dissections (Chapter 78) are more commonly associated with neck pain; for example, up to 20% of patients with carotid dissections complain of anterolateral pain, and about 80% of patients with vertebral dissections have posterior or occipital pain. Patients with arterial dissections frequently but not necessarily have signs and symptoms of stroke (Chapter 407). Disorders of the esophagus (Chapter 138) and mass lesions of the throat (Chapters 190 and 429) can also present as neck pain.

Acute spine pain can precede the rash in herpes zoster (Chapter 375) or can be seen in the vaso-occlusive crisis of sickle cell anemia (Chapter 163). Infections of the disc (Chapter 413) cause sharp back pain worsened by

TABLE 400-5 MECHANICAL LOW BACK PAIN

	MUSCLE STRAIN	HERNIATED NUCLEUS PULPOSUS	OSTEOARTHRITIS	SPINAL STENOSIS	SPONDYLOLISTHESIS	SCOLIOSIS
Age (yr)	20-40	30-50	>50	>60	20	30
Pain pattern location	Back (unilateral)	Back (unilateral)	Back (unilateral)	Leg (bilateral)	Back	Back
Onset	Acute	Acute (prior episodes)	Insidious	Insidious	Insidious	Insidious
Standing	↑	↓	↑	↑	↑	↑
Sitting	↓	↑	↓	↓	↓	↓
Bending	↑	↑	↓	↓	↑	↑
Straight leg	–	+	–	+ (stress)	–	–
Plain radiography	–	–	+	+	+	+

From Borenstein DG, Wiesel SW, Boden SD. *Low Back Pain: Medical Diagnosis and Comprehensive Management.* 2nd ed. Philadelphia: WB Saunders; 1995.

movement. Arachnoiditis (Chapter 412), an inflammatory process of the arachnoid space, can cause diffuse chronic back pain, often after the introduction of foreign substances or manipulation of the intrathecal space. Finally, 20 to 50% of patients with depression (Chapter 397) will complain of back pain that often is diffuse and described in emotionally laden terms. Complaints of low back pain are also common in malingering patients.

TREATMENT

Rx

Treatment options vary according to the severity of pain, presence of radicular signs or symptoms, and any underlying disease. Acute nontraumatic neck pain is common and usually benign. Beneficial treatments include nonsteroidal anti-inflammatory drugs (NSAIDs, e.g., ibuprofen 600-800 mg three times daily for 2 weeks), exercise, and physical therapy.[1] The low risk of myocardial infarction or upper gastrointestinal bleed from NSAIDs must be weighed against pain relief and the relative benignity of the condition. Chiropractic manipulation is of benefit but has the rare complication of posterior circulation stroke from arterial dissection, especially in patients younger than 45 years. Low-level laser therapy[1] and acupuncture[1] are also of short-term benefit. Cervical collars and traction are not of established benefit. Even in the absence of radicular pain, surgical intervention for neck pain is indicated in cases of vertebral instability, such as caused by fracture or dislocation. Surgery, including fusion, may be needed for stabilization after surgery for lesions such as tumors, infections, or hemorrhages. For neck pain that is accompanied by signs or symptoms of radiculopathy, surgery should be considered but is not usually the initial therapy (see later). If a cervical spine lesion is causing spinal cord compression, emergent evaluation for possible surgery is indicated (see later). For chronic neck pain, yoga therapy yields significant pain relief with the possible added benefits of improving quality of life and psychological well-being.[1]

Because acute low back pain is generally benign, invasive therapy should be avoided in the first 3 months. Conservative treatment includes NSAIDs (e.g., ibuprofen 600-800 mg three times daily as needed) and controlled physical activity; strict bed rest for longer than 2 days is no better than restricted physical activity.[1] Acetaminophen is no better than placebo.[1] Other options include local heat and massage. Spinal manipulation, exercise therapy, massage, and cognitive-behavioral therapy are moderately effective.[1] True acupuncture appears no better than sham acupuncture.[1]

For chronic nonradicular low back pain, exercise and cognitive-behavioral therapy are recommended; yoga should also be considered.[1] Prolotherapy, in which a mild irritant is injected into a tendon or ligament to increase blood flow and promote healing, and facet joint injections can be helpful, but intradiscal steroid injections and percutaneous intradiscal radiofrequency thermocoagulation are not effective.[1] Transcutaneous electrical neurostimulation (TENS) does not appear to provide any functional improvement in patients with chronic lumbar pain.[1] Oral analgesic medications are frequently offered and include NSAIDs; duloxetine (30 to 60 mg daily) may be a reasonable alternative. In one randomized trial, tanezumab (10 to 20 mg intravenously, repeated 8 weeks later), a humanized monoclonal antibody that specifically inhibits nerve growth factor, reduced pain better than naproxen or placebo.[1]

When persistent nonradicular low back pain is accompanied by associated degenerative spine changes, surgical fusion is of benefit but not superior to interdisciplinary rehabilitation.[1] If degenerative changes lead to lumbar scoliosis with or without neurologic signs, however, decompressive surgery with or without fusion appears effective for at least 5 years for the majority of patients. Outcomes after instrumented lumbar spinal fusion are improved if rehabilitation is delayed for 12 weeks after the operation. In more focal disorders, surgery to remove disc material pressure from pain-sensitive structures can be considered; techniques include open, laser, and microdiskectomy approaches, with little evidence to favor one option over the other.

PROGNOSIS

Between 50 and 85% of patients who have neck pain that persists for more than 1 day report recurrence of symptoms in 1- and 5-year follow-up. Slightly more than 50% of patients recover within 3 months, and those who remain symptomatic generally have relatively little pain and disability.[3] Patients younger than 45 years have less recurrence, and patients aged 45 to 59 have the highest risk. Prior neck injury or pain, coexistent low back pain, and self-perceived poor general health are risk factors for symptoms persisting past 3 months or recurrence.

Mechanical spine pain, even with radicular symptoms, resolves without specific intervention within 30 days in many patients and within 3 months in 90% of patients. Recurrence is frequent, however, especially in patients with spondylosis, because the underlying process persists and further degeneration of the spinal elements can be expected.

Long-term disability is more common with obesity, low education level, tobacco use, high levels of pain at the onset, tendency to somatization, job dissatisfaction, lack of availability of light-duty employment, and need to perform significant lifting at work. The strongest factors affecting outcome are psychological, especially worrying, fear avoidance, anger, and frustration.[4] Genetic variability, such as in polymorphisms of catechol *O*-methyltransferase, also may play a role in the development of chronic pain.

Spinal Stenosis

DEFINITION

Spinal stenosis, which is a narrowing of the spinal canal, results in compression of neural structures in the cervical and lumbar regions, where the diameter of the spinal cord is largest. Signs and symptoms of spinal stenosis are referable to these levels. In the lumbar region, L4-5 is the most common level of stenosis, followed by L3-4 and L5-S1.

EPIDEMIOLOGY

The annual incidence of spinal stenosis in the United States is about 1 to 2 per l00,000 for the cervical region and 5 per 100,000 for the lumbar region.[5] Spinal stenosis often coexists in the cervical and lumbar regions, and the incidence is higher in patients with more complex degenerative anatomy.

PATHOBIOLOGY

Primary spinal stenosis is due to a congenital narrowing of the spinal canal. Causes of secondary stenosis include chronic degenerative conditions such as spondylosis and thickening of the ligamenta flava and longitudinale posterius neoplasia, osteomyelitis, and rheumatoid arthritis. In patients receiving corticosteroids long term, epidural lipomatosis is a cause predominantly at the thoracic level. The underlying cause of symptoms may be multifactorial, with direct nerve pressure, duration of the pressure, capillary restriction, venous congestion, and reliance on the anastomosis of cerebrospinal fluid for metabolic homeostasis all potentially playing a role.

In spinal stenosis, myelopathy results from compression of the veins that drain the canal, thereby leading to capillary stasis and edema, which result in

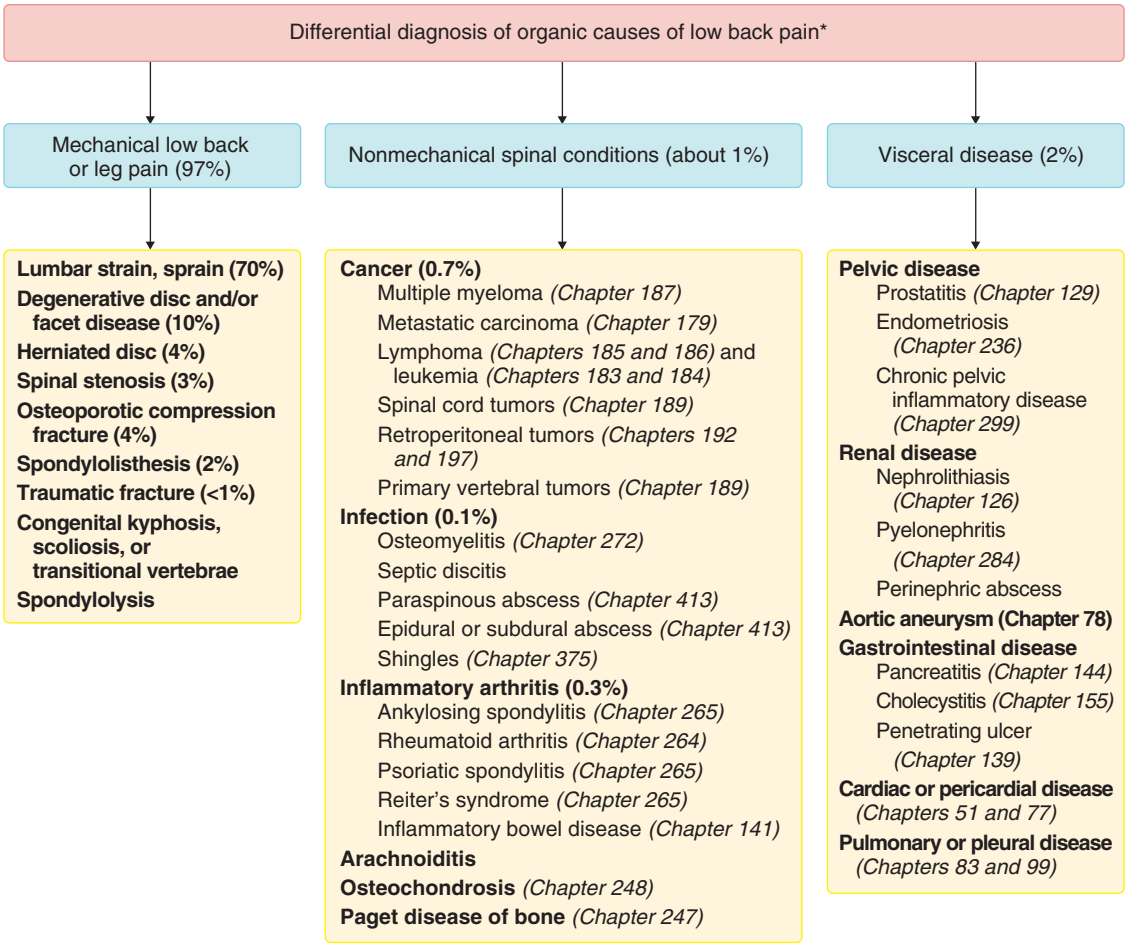

FIGURE 400-6. Differential diagnosis of organic causes of low back pain. * Percentages are approximations and may vary substantially in different practices. (Data from Deyo RA, Weinstein JN. Low back pain. *N Engl J Med.* 2001;344:363-370.)

further compromise of the spinal cord. If the anterior spinal artery is also occluded intermittently or chronically, ischemic gliosis develops. The cord itself can be compressing, and dynamic injury can occur during flexion and extension, especially in chronic degenerative conditions associated with thickening and buckling of the ligamentum flavum. Superimposed trauma, such as severe flexion-extension from whiplash injury or falls, can cause the central cord syndrome (see later, Disorders of the Spinal Cord).

CLINICAL MANIFESTATIONS

Cervical stenosis, which usually results from spondylosis (cervical spondylotic myelopathy) and degenerative spine changes superimposed on congenital stenosis, can be manifested in a number of ways. At the C5-8 level, both the cord and roots are involved, and patients develop lower motor neuron findings in the arms and upper motor neuron signs in the legs; the result is the slow onset of painless atrophy of the weakened hand muscles, accompanied by gait abnormalities. Above the C5 level, the arms may become numb and clumsy, but there is little atrophy. Gait problems are frequently described as leg stiffness, heaviness, or incoordination, and patients may complain of falls or fear of falling, difficulty walking on uneven surfaces, and needing to watch where they are walking. In general, neck pain is not as prominent a complaint as is low back pain in lumbar spinal stenosis, but radicular pain occurs in about one third of patients. On examination, a Lhermitte's sign, which is a shock-like tingling traveling down the spine with neck flexion or less commonly with extension, may be elicited. Sensation to pinprick examination will often be decreased in the hands, but this finding may be patchy and not well demarcated. Increased muscle tone will predominate in the legs, and reflexes will likely be increased; the Babinski sign is often present. The stance and gait may be wide based, with slow, short steps. Romberg's sign may be present.

Lumbar spinal stenosis can present with intermittent or persistent signs and symptoms referable to both cord and roots; axial neck or back pain is almost always present. However, the correlation between symptoms and the diameter of the canal is poor. The pain can also radiate in a pseudoradicular pattern from the lumbar region into the gluteal region, groin, and one or both upper legs, but pain rarely radiates distal to the knee unless spinal root involvement occurs. Lumbar neurogenic claudication is perhaps the most specific; symptoms are precipitated by compressive effects on the spinal cord and exiting roots and are characterized by increasing leg paresthesias and weakness as the patient remains standing and gravity pulls the lumbar cord enlargement into the narrowed spinal segment. As opposed to vascular claudication (Chapter 79), symptoms occur even at rest and can be relieved only by bending at the waist or sitting; the patient will characteristically report improved ability to walk if bent at the waist. Other features that suggest neurogenic claudication rather than vascular claudication include back pain, paresthesias, numbness, weakness, and prolonged presence of symptoms after changing position (seconds to 1 to 2 minutes in vascular claudication but 2 to 20 minutes in neurogenic claudication).

DIAGNOSIS

Although MRI can localize and quantify the severity of the stenosis, the degree of stenosis does not correlate well with the degree of symptoms. The normal diameter of the lumbar canal is approximately 22 to 25 mm, relative lumbar stenosis is 10 to 12 mm diameter, and a diameter of less than 10 mm is considered absolute stenosis. MRI also can identify any comorbid spinal disease, such as degenerative changes and spondylolisthesis. However, CT myelography, which provides the details needed for surgical treatment, is another alternative. Electromyography (EMG) is not generally useful in the diagnosis of spinal stenosis except in assessment of comorbid conditions (e.g., localized radiculopathy) that might be important for surgical planning, as well as for detection of other potentially significant disorders, such as polyneuropathy.

The differential diagnosis of lumbar spinal stenosis includes vascular claudication from peripheral vascular disease (Chapter 79) and abdominal aortic

aneurysm (Chapter 78).[6] The symptoms of vascular claudication are activity dependent, unlike the positional dependency of the spinal stenosis. Other conditions that can mimic lumbar stenosis include sacroiliac joint dysfunction, radiculopathy or disc prolapse, polyneuropathy, tethered cord, and spina bifida. Hip pathology, including avascular necrosis, should also be considered. Because cervical stenosis can cause both arm and leg symptoms, the differential diagnosis also includes multiple sclerosis (Chapter 411), syringomyelia (Chapter 417), Arnold-Chiari malformation (Chapter 417), cerebrovascular disease (Chapters 407 and 408), normal-pressure hydrocephalus (Chapter 189), and motor neuron disorders (Chapter 419).

TREATMENT

Rx

Spinal stenosis is not necessarily an indication for surgery. Conservative therapy with NSAIDs (e.g., ibuprofen 400-800 mg four times daily (max 3.2 gm/d) or acetaminophen 650 mg every 4 to 6 hrs (max 3.25 gm/d), muscle relaxants, physical therapy, and a hard cervical collar can help many patients. Epidural injection of lidocaine is as effective as injection of lidocaine plus steroids in patients whose pain is disabling.[■] For spinal stenosis, epidural injection with etanercept, a tumor necrosis factor (TNF)-α inhibitor, has an analgesic effect that may be superior to steroids.[■] However, surgery should be considered in patients with severe symptoms and signs of progressive myelopathy. Decompressive laminectomy, with or without fusion, is superior to nonsurgical therapy for improvement of pain and function for at least a year in patients with lumbar spinal stenosis.[■]

PROGNOSIS

Because spinal stenosis is usually degenerative, it is often a slowly progressive condition without rapid deterioration. Nevertheless, up to 50% of patients are stable in the long term. After symptoms develop and the condition is affecting quality of life, decompressive surgery should be considered. Because the underlying condition itself is progressive, however, surgery is not a permanent cure, and symptoms often recur.

DISORDERS OF THE NERVE ROOTS

DEFINITION

Disorders of the nerve root, termed *radiculopathy*, lead to symptoms referable to a dermatome or myotome. Either the ventral (anterior, motor) or posterior (dorsal, sensory) root can be involved independently or after they join to form the spinal nerve. Symptoms will follow the anatomic location. Because the root must traverse the vertebral foramen, it is prone to disorders of the spine at these locations. *Sciatica*, which is a commonly used but poorly defined term, often connotes low back pain with radiation into the ipsilateral leg, thereby implying pain radiating along the sciatic nerve, which anatomically contains fibers originating in the L4-S2 roots. The cauda equina syndrome results from disease involving the roots of the lower lumbar and sacral spinal cord levels as they traverse inside the spinal canal on their way to exit below their respective vertebral bodies.

PATHOBIOLOGY

Irritation of the spinal sensory nerve root or dorsal root ganglion causes symptoms referable to that dermatome. Spontaneous dysesthesias are hypothesized to result from ectopic discharges from the injured nerve; sensitization of the injured nerve leads to tactile evoked dysesthesias in the same distribution. Mechanical compression of the nerve contributes to the syndrome. In addition, inflammatory cytokines leak from the nucleus pulposus into the epidural space, where they result in endoneurial edema and pain. The pro-inflammatory cytokine TNF-α is a likely main contributor. Rupture of the nucleus pulposus releases phospholipase A_2, which also plays an important role in the inflammatory process. The inflammatory process itself can cause pain even in the absence of frank root compression.

The nerve root exits through the intervertebral foramen, where it is subject to compression and injury. The proximal portion of the nerve root has a small region of decreased vascular supply, where it is especially prone to edema, which can exacerbate the effect of the original injury. Treatment of this edema is one of the potential therapeutic effects of corticosteroid injections. The S1 root is the most susceptible to injury at the foramen because it is the largest-diameter spinal nerve and exits through the narrowest lumbar foramen. In addition, because it is traveling inferiorly from the S1 spinal cord level to pass out of the foramen under the S1 vertebral body, it passes through the superior, most narrow part of the foramen. The superior location of the sensory root on these nerves may account for the early predominance of sensory symptoms versus motor symptoms.

CLINICAL MANIFESTATIONS

The symptoms of radiculopathy depend on the affected root. Root involvement is likely if pain radiates beyond the shoulder or the knee. In the thoracic region, root involvement often produces symptoms that "wrap around" the trunk. Radicular pain is often worsened by activities that increase intraspinal pressure, such as coughing, sneezing, straining, and other Valsalva maneuvers. The characteristics of the pain vary, but when they are exacerbated by such provocation, they are often described as sharp, shooting, electrical, and tingling. When reporting the symptoms, patients may point to or rub the distal dermatome where they are experiencing the discomfort (perceived pain). Patients also may report specific positions that increase or decrease pain; for example, sitting will often worsen the pain of acute lumbar disc herniation, and neck extension can produce radiating pain in cervical disc herniation or other processes that narrow the foramen.

It is uncommon for patients to spontaneously note anesthesia, but they often note dysesthesias in radiculopathy, even in the absence of spine pain. The localization of these dysesthesias often follows the dermatome but also may be described diffusely by the patient. Likewise, complaints of weakness may be difficult to isolate to a particular muscle; however, exceptions exist, such as when the patient complains of a weak grip or a foot drop.

On examination, side-by-side strength testing of specific muscles (Chapter 421, Table 421-3) can help identify slight weakness. Slight weakness may also be identified by evaluating for pronator drift in the arms or asking patients to walk on their toes, walk on their heels, and do shallow knee bends on each leg independently. Sensory examination should test all potential root distributions; pinprick is often sufficient, and it is helpful to ask the patient to report any abnormalities, not just frank hypesthesia. Hyperreflexia in the arms is not expected in a spinal nerve disorder and, if unexplained, should be further investigated for an injury to the spinal cord or brain (Chapter 399). Patients who are older than 65 years or who have a peripheral neuropathy (Chapter 420) might have reduced or even absent ankle jerks.

The cauda equina syndrome is manifested as unilateral or bilateral leg weakness, saddle anesthesia, urinary dysfunction with hesitancy or retention, and, less commonly, bowel dysfunction. Depending on the cause, it is frequently accompanied by low back pain. The syndrome can be accompanied by severe sciatica, which can be unilateral or bilateral but also involve perineal pain. The weakness of the legs, which may be asymmetrical, is of the lower motor neuron type. The major causes of the cauda equina syndrome include lumbar disc herniation, neoplasm, and lumbar spinal stenosis.

DIAGNOSIS

The history and physical examination are similar to the evaluation of neck and back pain (see earlier), with special emphasis on finding evidence of nerve root involvement.

The patient should be queried about bowel and bladder dysfunction. Frank incontinence needs to be investigated for either the cauda equina syndrome or a myelopathy. If the patient has any loss of perineal sensation, such as might be noted during or after voiding or bowel movement, the examination should test perianal sensation, anal sphincter tone, and anal wink reflex. Because the cauda equina syndrome involves nerve roots, reflexes should be normal or decreased; hyperactive reflexes or a Babinski sign would indicate a myelopathy (see later).

Ancillary Testing

Imaging to evaluate a potential radiculopathy is similar to the evaluation for neck and back pain (see above), but MRIs must be interpreted judiciously. For example, MRI performed 1 year after disc herniation with sciatica cannot determine which patients continue to be symptomatic and which are symptom free.[7] Alternatively, EMG can localize radicular abnormalities and their severity as well as assess possible comorbid neurologic diseases, such as diffuse peripheral or entrapment neuropathies. EMG electrodiagnostic localization can be extremely helpful in determining whether a finding on MRI is truly associated with neurologic impairment, and it has a high sensitivity and specificity for identification of acute and chronic denervation when the motor (anterior) aspect of the root is involved. However, EMG is less sensitive (about 30 to 70%) if only the sensory (posterior) limb of the root is involved by the lesion. MRI with its high sensitivity and EMG with its high specificity should be considered complementary tests.

Differential Diagnosis

Many mechanical processes can injure the spinal nerve root (see earlier, Neck and Back Pain), and most of the causes of spine pain can cause root disorders. Spinal cord compression may accompany radiculopathy.

In addition to conditions that can affect the nerve root as it leaves the spinal column, intracolumn abnormalities below the level of the conus medullaris can affect the lumbar and sacral roots before they exit, thereby resulting in the cauda equina syndrome. Most commonly, the cauda equina syndrome is caused by extrinsic compression of the caudal sac by a mass, such as a large and centrally herniated lumbar disc, metastatic tumor, abscess, or epidural hematoma, but arachnoiditis or chronic meningitis must be considered.

Disorders of the brachial or lumbosacral plexus can cause pain radiating down a limb in a radicular or polyradicular pattern. Painful peripheral neuropathies (Chapter 420) can also resemble a radiculopathy. Non-neurologic disorders, such as fibromyalgia (Chapter 274) and polymyalgia rheumatica (Chapter 271), can cause axial pain that mimics a radiculopathy. Cervical radiculopathy also can be mimicked by acromioclavicular joint arthropathy, shoulder bursitis, and the shoulder impingement syndrome; lumbar radiculopathy can be mimicked by hip arthritis, trochanteric bursitis, iliotibial band syndrome, and hamstring tendinitis (Chapter 263).

TREATMENT

Rx

Acute neck and back pain, even with radicular symptoms, is usually self-limited and resolves.[8] If radiculopathy is associated with an underlying non-structural lesion, such as infection or tumor, treatment should be directed toward that underlying lesion (see also later, Metastatic Spinal Cord Compression). If symptoms or neurologic dysfunction progress or persist for more than 6 weeks, interventional options should be considered. Surgery is indicated for spinal instability, progressive neurologic deficits, or severe radicular pain that persists for more than 3 months despite conservative therapy, especially in patients with spinal stenosis and herniated discs.

In cervical radiculopathy, either foraminal or epidural corticosteroid injection may be of benefit. A series of injections, usually between one and three, provides short-term relief of radicular symptoms, with an acceptable adverse event profile. Although minor events, such as increased neck pain, headache, and vasovagal reactions, are relatively frequent (5 to 20%), serious events, such as epidural hematoma or abscess, are uncommon (<1%). In cervical spondylopathy with radiculopathy, surgery provides more rapid pain relief than does physiotherapy but little additional benefit in the long term. Anterior cervical discectomy, with or without fusion, is beneficial. Cervical anterior discectomy and arthroplasty result in a lower risk of postoperative dysphagia compared with discectomy and fusion, but more complicated surgeries are not of established benefit.

In lumbar radiculopathic pain, epidural steroid injections may provide relatively minor short-term symptom relief for 2 to 6 weeks, but they do not improve function or relieve pain beyond 3 months. They usually will not delay or avoid surgery and therefore are not routinely recommended. Chemonucleolysis of a lumbar disc is moderately superior to placebo but inferior to surgery.

For symptomatic radiculopathy associated with a herniated disc, either open surgical discectomy or microdiskectomy is superior to nonsurgical therapy for at least 3 months.[9] If the condition is isolated to a single disc and no significant degenerative changes are present, longer periods of benefit are more likely. Patients who derive a greater benefit from surgery include those with radiculopathy in which the MRI shows a herniated disc with a resultant compression of the thecal sac of one third or more, or those with nerve root compression.[10] Systematic reviews suggest that conservative discectomy allows a quicker return to work and has less long-term back pain than does more aggressive surgery,[11,12] but with the downside of an increased risk of recurrent disc herniation. Video 400-7 demonstrates a right L4-5 hemilaminectomy with removal of a herniated disc fragment. The presurgical MRIs show the position and extent of the disc herniation.

DISH can cause symptomatic radiculopathy.[13] When progressive or associated with myelopathy, surgical decompression of the ossified posterior longitudinal ligament should be considered. No particular surgical approach has been shown to be superior, and each case must be approached based on individual factors.[14]

DISORDERS OF THE SPINAL CORD

DEFINITION AND GENERAL OVERVIEW

A disorder of the spinal cord itself is termed a *myelopathy*. A myelopathy can be intramedullary, as the result of a disorder intrinsic to the cord, or extramedullary, as the result of an abnormality that is extrinsic to the cord but compressing it.

TABLE 400-6 CAUSES OF MYELOPATHY

Trauma/compression	Neoplastic
Direct ± vertebral spine disease	Metastatic cord compression
Spondylotic myelopathy/stenosis	Spinal tumors
Post-traumatic syrinx	Paraneoplastic
Arachnoid cyst	Infectious
Vascular	Epidural abscess
Cord infarction	Syphilis
Dural arteriovenous malformation	Lyme disease
Inflammatory/autoimmune	Tuberculosis
Multiple sclerosis	HIV infection
Devic disease	Tropical spastic paraparesis
Acute disseminating encephalomyelitis	Herpes zoster
Adrenomyeloneuropathy	Toxic/metabolic
Systemic lupus erythematosus	Post-radiation myelopathy
Sjögren syndrome	Vitamin B_{12} deficiency
Mixed connective tissue disease	Vitamin E deficiency
Postinfectious/postvaccination	Heroin
Arachnoiditis	Epidural lipomatosis
	Congenital/hereditary
	Chiari malformation
	Syringomyelia

EPIDEMIOLOGY AND PATHOBIOLOGY

Spinal cord disorders can be caused by a wide range of conditions (Table 400-6).

The functional elements of the spinal cord (Fig. 400-7) include descending tracts largely to motor and autonomic neurons, motor neurons, autonomic neurons, and sensory ascending tracts. The anterior horn cell motor neuron is the cell body for the axon that will become the anterior nerve root and continue directly to innervate the muscle. The cell bodies for the primary sensory neurons reside in the dorsal root ganglion outside the spinal cord itself.

CLINICAL MANIFESTATIONS

The clinical manifestations of myelopathy result from the spinal level of the lesion. The majority of signs will be bilateral, but asymmetry, or even unilaterality, does not exclude a spinal cord lesion.

In general, the three major functions affected are motor, sensory, and autonomic, especially bowel, bladder, and erectile function. If anterior horn cells are involved at the lesion level, the corresponding myotome will exhibit lower motor neuron function (hypotonic weakness), and reflexes may be decreased at that level. Below the lesion, however, patients will have hypertonic weakness that can progress to spastic paralysis, hyperreflexion, and Babinski sign. Sensation will be decreased from the level of the lesion and distally. Because increased tone and spasticity often develop over time, they may not be dramatic at the initial clinical presentation. If the posterior columns are compromised, patients may lose joint position sense and develop ataxia, especially of gait. If the posterior columns of the cervical cord are impaired, patients may have pseudoathetosis of the fingers, manifested as unconscious athetotic movements of the fingers of the outstretched arm when the eyes are closed.

The anterior cord syndrome is manifested as lower motor neuron weakness at the level of the lesion (anterior horn); upper motor neuron weakness and spasticity below the lesion (corticospinal tracts); autonomic dysfunction below the level of the lesion (lateral horn), most often bowel and bladder dysfunction; and loss of pain and temperature sensation below the level of the lesion (spinothalamic tract). Vibration and joint position sense remain intact (posterior columns). The major causes of this syndrome are vascular, such as an anterior spinal artery, or an anteriorly impinging mass lesion, such as disc or vertebral body mass.

The central cord syndrome is manifested as lower motor neuron signs and symptoms at the level of the lesion (anterior horn cells) and upper motor neuron signs and symptoms below the lesion (corticospinal tracts), urinary retention, and a band of loss of temperature and pain sensation at the level of the lesion (anterior white commissure decussation of these fibers). In the midcervical level, this syndrome is typical of syringomyelia (Chapter 417). Other major causes include intramedullary tumors (Chapter 189) and post-traumatic cervical injury (Chapter 399) in patients with disc herniation or preexisting cervical spondylosis.

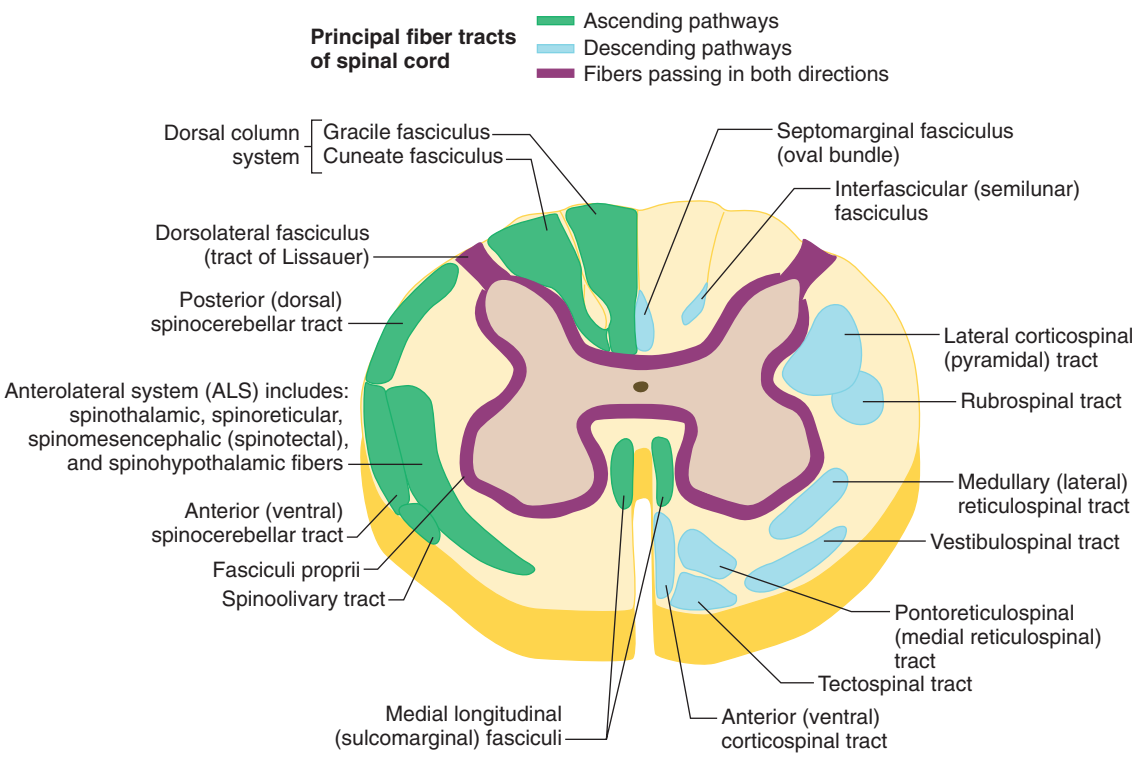

FIGURE 400-7. **Principal fiber tracts of spinal cord.**

The posterior cord syndrome is manifested as complaints of imbalance, especially in the dark or with eyes closed, and an examination notable for ataxia, presence of Romberg's sign, and loss of vibration sense and proprioception below the level of the lesion (posterior columns), with preservation of pain and temperature sensation. Patients are seldom weak. Posterior compression may be caused by spondylotic disease, but other major causes include deficiencies of vitamin B_{12} or vitamin E (Chapters 218 and 416), syphilis (Chapter 319), AIDS-associated vacuolar myelopathy (Chapter 394), and nitrous oxide inhalation (Chapter 432).

The Brown-Séquard (cord hemisection) syndrome combines features of these syndromes. At the level of the lesion, patients exhibit ipsilateral lower motor weakness (anterior horn) and loss of all sensation (posterior root entry zone). Below the level of the lesion, patients have ipsilateral upper motor weakness and spasticity (corticospinal tract) and ipsilateral loss of vibration sense and proprioception (posterior columns), with contralateral loss of pain and temperature (spinothalamic tract, the fibers of which have crossed from the opposite side through the anterior white commissure). The Brown-Séquard syndrome is often caused by trauma (Chapter 399) or eccentric compression.

The conus medullaris syndrome refers to dysfunction of the distal-most tapered portion of the spinal cord, which anatomically lies at approximately the T12-L1 vertebral spine level. The arms are normal; weakness in the legs is variable but often symmetrical when it is present. The main signs and symptoms are sexual dysfunction, loss of bowel and bladder control, and perianal anesthesia with loss of the anal wink reflex. The major causes are disc herniation, lumbar stenosis, and neoplasm.

DIAGNOSIS

Symptoms of bilateral involvement of the arms or legs suggest a myelopathy, although bilateral leg involvement can be seen in lumbar spinal stenosis and the cauda equina syndrome. Complaints of leg stiffness or incoordination suggest spasticity from myelopathy. Other symptoms include a recent change in bowel or bladder function, erectile dysfunction, imbalance (especially in the dark or with eyes closed), and catching of feet when walking. These central symptoms can also reflect lesions of the brain stem and higher, so the patient should be queried about cortical and brain stem symptoms (e.g., cognitive function, vision, facial strength, sensation, and swallowing). Focal back pain supports a myelopathy.

Examination for a potential myelopathy must include an evaluation of anal tone and perineal sensation. The patient should be examined in a gown to allow inspection of the spine as well as the overlying skin. A sensory level, which represents a point where distal (inferior) sensation is altered, should also be sought, but a spinal cord lesion rarely causes a sharp line of sensory demarcation. Joint position sense can be diminished if the posterior columns are involved. A tandem gait will assess possible gait ataxia. Finally, the patient should be examined for distal hypertonicity, if not frank spasticity, by assessment for hyperreflexia and Babinski signs. In acute spinal cord lesions, however, a state of "spinal shock" can cause hyporeflexia or even a flaccid paralysis.

Ancillary Testing

MRI is the test of choice because it provides anatomic detail of the spine and subarachnoid space as well as the spinal cord. MRI may also show evidence of demyelinating or metastatic disease. Plain films are not adequate to evaluate the spinal cord. If the MRI is normal, lumbar puncture is sometimes useful for evaluation of conditions that resemble myelopathy: Guillain-Barré syndrome (Chapter 420), infectious or carcinomatous meningitis (Chapter 412), arachnoiditis, and transverse myelitis.

Differential Diagnosis

The many causes of myelopathy (see Table 400-6) can typically occur at any spinal level, but certain conditions predominate at specific spinal levels.

Any lesion that has an upper cervical location must be evaluated for disorders of the craniocervical junction, especially disorders that can produce atlantoaxial instability, such as rheumatoid arthritis (Chapter 264); after trauma, cervical or odontoid fracture must be excluded. Disorders at the base of the skull, such as Chiari I and other congenital malformations (Chapter 417), can sometimes affect the upper cervical cord. Syringomyelia (Chapter 417), which may or may not be associated with Chiari malformation, also has a predilection for the cervical cord.

The thoracic cord is relatively protected from all but direct trauma, but it is the most common site for metastatic cord compression. Transverse myelitis is most commonly thoracic, and the thoracic cord also is particularly vulnerable to a watershed ischemic myelopathy due to severe hypotension. Epidural lipomatosis often is most symptomatic at the thoracic level. DISH also has a predilection for the thoracic cord and can produce spinal cord compression by ossification of the posterior longitudinal ligament.

The lumbar-conus region is the most common site for disc herniations. In addition, ependymomas are relatively more common in this region, as are metastases from more caudal locations and compression from arachnoiditis.

The rapidity of onset helps in diagnosis. Acute or relatively acute myelopathy suggests vascular causes, trauma, demyelinating lesions, or sudden

decompensation of a preexisting lesion, such as a pathologic fracture. In a young person with no other comorbid illnesses, a demyelinating illness, such as multiple sclerosis (Chapter 411) or acute disseminated encephalomyelitis (Chapter 414), is suggested; other CNS lesions separated in space with white matter signal abnormalities on MRI would increase the likelihood of this diagnosis. In older persons or patients with known vascular risk factors, hypotension, or an onset in the immediate postoperative period, a spinal cord infarction is possible. Sudden sharp back pain suggests mechanical disorders (e.g., pathologic fracture, sudden worsening spondylolisthesis) or spinal cord infarction, whereas demyelinating lesions are often painless.

Myelopathy developing subacutely, especially accompanied by back pain, can be caused by metastatic disease (Chapter 189) and abscesses (Chapter 413). Both of these conditions must be evaluated and treated as true emergencies to prevent permanent paralysis. Subacute or chronic myelopathies include vitamin B_{12} deficiency (Chapter 416), although nitrous oxide inhalation may cause an acute expression of the disorder; syringomyelia (Chapter 417); and more slowly growing tumors, such as meningiomas (Chapter 189), lipomas, and neurofibromas (Chapter 417).

In the setting of known malignant disease or unexplained weight loss, metastatic cord compression (Chapter 189) must be considered. Weight loss, back pain, and fever can be seen in infection (Chapters 412 and 413) and occasionally spondyloarthropathies (Chapter 265). Infectious causes also include tropical spastic paraparesis (human T-lymphotropic virus type 1 [HTLV-1; Chapter 378]). Syphilis (Chapter 319) is the cause of tabes dorsalis; patients may have other signs and symptoms, such as lancinating pains, ataxia, depressed leg reflexes, and Argyll Robertson pupils. Myelopathies that follow an infectious illness include acute disseminated encephalomyelitis and progressive necrotizing myelopathy. Transverse myelitis can follow viral infections, such as herpes zoster (Chapter 375). Myelopathy accompanied by evidence of multifocal cortical dysfunction would likely be multiple sclerosis or acute disseminated encephalomyelitis (Chapter 414). An accompanied peripheral neuropathy is seen in vitamin B_{12} deficiency myelopathy (Chapter 416), which tends to cause gait ataxia. Rheumatoid arthritis (Chapter 264) causes progressive loss of cartilage and bone destruction, potentially leading to atlantoaxial instability and cervical subluxation. Other systemic illnesses, such as systemic lupus erythematosus (Chapter 266), Behçet syndrome (Chapter 270), and sarcoidosis (Chapter 95), can also cause myelopathies. In patients with exogenous or endogenous hypercortisolemia (Chapter 227), epidural deposition of unencapsulated fat can cause epidural lipomatosis that compresses the spinal column. Patients with a distant history of trauma might be evaluated for post-traumatic syringomyelia. Patients with a history of lumbar puncture, surgery, or intrathecal injections can develop arachnoiditis.

Younger patients are more likely to be symptomatic from congenital disorders, ankylosing spondylitis, or multiple sclerosis. In patients older than 55 years, cervical spondylotic myelopathy is the most common cause of myelopathic symptoms.

TREATMENT AND PROGNOSIS

Rx

High-dose steroids should be used in cord compression from metastatic tumors (Chapter 189). Steroids are also frequently used in transverse myelitis, although no controlled trials have been performed. Steroids are no longer recommended for acute spinal cord trauma (Chapter 399).

Patients with spinal cord lesions must be assessed emergently for any potential complications. With high cervical spine lesions, paresis or paralysis of the diaphragm and respiratory depression can occur. Although not emergent, higher cervical or medulla extension of inflammatory myelopathies can cause distressing hiccups and nausea. In lesions at the thoracic level or above, interruption of the lateral column autonomic pathways can lead to autonomic instability, including altered blood pressure responses. At almost any level, but especially at the conus and cauda equina, acute urinary retention may require catheterization. Long-term complications of spinal cord injury include osteoporosis (Chapter 243), orthostatic hypotension (Chapters 51 and 62), and chronic neuropathic pain (Chapter 420), all of which may require specific therapy. Other aspects of treatment and prognosis depend on the specific cause.

Specific Causes of Myelopathy

VASCULAR MYELOPATHIES

The CNS tissue of the spinal cord is as intolerant of ischemia as the brain is. Vascular myelopathy occurs when there is loss of blood flow to the spinal cord, whether it is acute or chronic and whether the cause is ischemic or hemorrhagic.

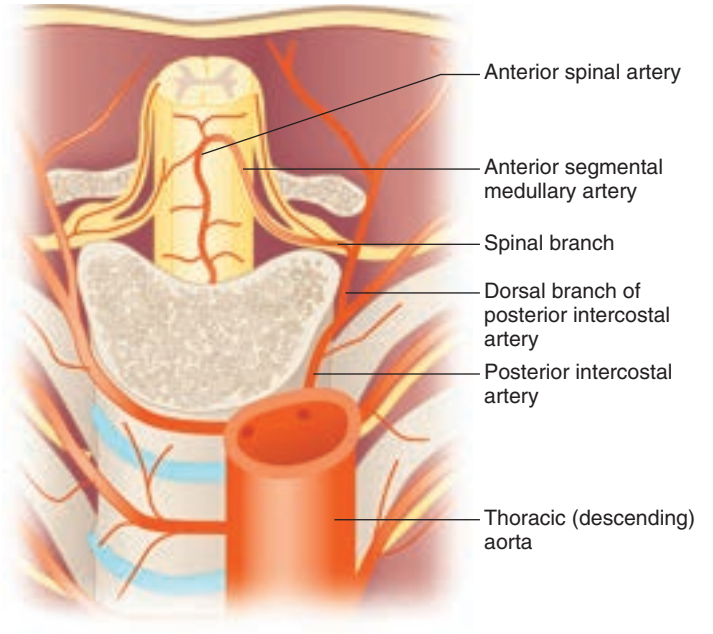

FIGURE 400-8. Blood supply of the spinal cord: section through thoracic level, anterosuperior view.

The blood supply to the spinal cord comes from the anterior and posterior spinal arteries that run longitudinally along the length of the spinal cord (Fig. 400-8). The paired posterior spinal arteries are derived in their most rostral origin as branches of the vertebral arteries at the level of the medulla and then run inferiorly along the posterolateral surface of the spinal cord. Along their course, they are fed by a series of small arteries that enter the spinal canal through the intervertebral foramina. Inside the canal, they anastomose extensively to provide redundancy. The anterior spinal artery is formed superiorly when branches of the vertebral artery join to form a single anterior spinal artery, which then runs down the midline of the anterior surface of the spinal cord. It also receives feeders along its length, but not to the same extent as the posterior segmental branches do. A large anterior radicular artery at C5-6 supplies the cervical enlargement. The main caudal anterior blood supply, however, is from the large artery of Adamkiewicz that enters the spinal canal between the cord levels of T9 and L2 and serves as the main blood supply to the anterior spinal artery, which supplies the lumbar enlargement, the lower thoracic cord, and the conus medullaris.

Compromise of the microvascular supply to the cord underlies the gliotic changes in many slowly progressive myelopathies such as spondylotic myelopathy. Ischemic causes include general hypotension, atherosclerotic disease, embolic events, vasculitis, and vascular steal; hemorrhagic events usually result from rupture of abnormal vascular malformations.

Severe global hypotension or aortic dissection or surgery can cause ischemic myelopathy, especially in watershed areas of the spinal cord, notably the thoracic region. Atherosclerosis, especially of the artery of Adamkiewicz, can lead to ischemic infarction of the cord by either decreased perfusion or thromboembolic events. Ischemia has also been reported to be a result of compression of the anterior spinal artery by a centrally herniated T12-L1 disc.

Embolic events and local thrombosis can occur in pregnancy and sickle cell disease (Chapter 163). During decompression sickness (Chapter 94), nitrogen bubbles cause microvascular emboli and ischemia. Vasculitis affecting the spinal cord is rare, but granulomatous angiitis of the CNS and polyarteritis nodosa (Chapter 270) can lead to infarction.

Spinal dural arteriovenous fistulas are the most common type of vascular malformation of the spinal cord. Arteriovenous malformations are most common in the thoracic cord, especially in patients older than 30 years. Vascular malformations can cause myelopathy by being mass lesions that compress local structures, by interfering with normal venous drainage, by diverting blood as part of a vascular steal with exercise of muscles that compete for blood flow, or by hemorrhage.

Spinal cord hemorrhage, which is rare, can occur intramedullary in the cord itself or in subarachnoid, subdural, or epidural locations. Intramedullary hemorrhage is most often caused by trauma, although bleeding can also occur into a tumor or from an intramedullary vascular malformation. Bleeding from a malformation that enters the subarachnoid space can cause back pain and headache. Epidural hematomas, which cause extramedullary cord compression, can occur as a complication of surgery, myelography, or lumbar puncture, particularly in patients with bleeding diatheses.

CLINICAL MANIFESTATIONS

Occlusion of the artery of Adamkiewicz usually presents with signs of thoracic watershed ischemia—paraplegia with relative sparing of the sacral roots. Infarction in the anterior spinal artery distribution results in dysfunction of the anterior two thirds of the cord, including the anterior horns, spinothalamic tracts, and corticospinal tracts; patients usually present with acute paraparesis and impaired bowel and bladder function. Sharp and sometimes circumferential pain at the level of the infarct is often described. Below the level of the lesion, temperature and pain sensation are lost, but vibration and position sense (posterior columns) are preserved.

Infarction of the posterior arteries is less common because of their better collateral circulation. Clinical manifestations, which are less dramatic, include loss of vibratory and position sense, ataxia, gait coordination problems, and Romberg sign; reflexes may be depressed at the level of the infarction.

The central cord vasculature syndrome is similar clinically to a traumatic central cord syndrome (Chapter 399). It may occur as a watershed lesion between the territories of the anterior and posterior spinal circulation and is most common in the cervical cord in older patients with preexisting cervical spondylotic disease.

Vascular malformations most often have a chronic progressive clinical course. Pain is common. Arteriovenous fistulas, commonly in the thoracic cord, present as progressive paraplegia. Patients may have exacerbations of symptoms with exercise and certain postures, and sudden worsening usually indicates hemorrhage.

DIAGNOSIS

Vascular malformations are initially evaluated by MRI, which also can assess the health of the surrounding tissue. When the vascular malformation is connected to the dura (arteriovenous fistula), myelography can occasionally detect lesions that are not seen or are poorly defined by MRI. If embolization or surgery is being considered, spinal angiography is needed to identify feeding and draining vessels, although the test carries a small risk of infarction. Intramedullary arteriovenous malformations are more commonly found in the cervical and thoracic levels and may require angiography to be visualized. When imaging is equivocal, lumbar puncture can be considered; an elevated leukocyte count (>10 cells/μL) suggests an inflammatory myelopathy rather than a vascular lesion.

TREATMENT AND PROGNOSIS

Rx

Treatment options are limited and include reversal of the cause of ischemia, such as by correcting hypotension (Chapter 106) or treating for emergent sickle crisis (Chapter 163). Prognosis for most cases of spinal cord infarct is poor unless blood flow is restored rapidly. In one study, for example, 3-year mortality was 23%, 42% of survivors needed wheelchairs, but 40% of those in wheelchairs at hospital discharge were able to walk at a 3-year follow-up.[15]

Arteriovenous fistulas are treated by occluding the shunt with embolization or surgery. Successful treatment may arrest and occasionally improve symptoms. Patients with suspected epidural hematomas (Chapter 399) require emergency treatment and surgery if there is progressive neurologic dysfunction.

INFLAMMATORY AND METABOLIC MYELOPATHIES

Transverse myelitis, multiple sclerosis, and other demyelinating diseases are considered in Chapter 411. Metabolic myelopathies can be caused by vitamin B_{12}, vitamin E, and copper deficiencies (Chapter 218).

Acute disseminated encephalomyelitis is mostly a monophasic disorder of demyelination of the spinal cord and brain. If it is isolated to the spinal cord, it might best be termed *transverse myelitis*. Parainfectious or postvaccination causes account for at least 75% of cases. Postvaccination acute disseminated encephalomyelitis is associated with measles-mumps-rubella vaccinations and diphtheria-tetanus-polio vaccinations, as well as with vaccinations for influenza, hepatitis B, pertussis, and Japanese B encephalitis.

Connective tissue diseases can infrequently be a cause of myelopathy. Systemic lupus erythematosus (Chapter 266), with or without antiphospholipid antibody, can include myelitis in 1 to 3% of patients. Sjögren syndrome (Chapter 268), Behçet syndrome (Chapter 270), sarcoidosis (Chapter 95), ankylosing spondylitis (Chapter 265), mixed connective tissue disease (Chapter 270), and systemic sclerosis (Chapter 267) can be associated with inflammatory myelitis.

Human T-lymphotropic virus type 1 (HTLV-1)–associated myelopathy/tropical spastic paraparesis (Chapter 378) is a chronic progressive myelopathy that causes leg weakness, spasticity, loss of vibratory sense, and bladder dysfunction. More than 90% of infected persons remain asymptomatic, with transformation to a symptomatic condition thought to be largely related to the host's inflammatory response. Other neurologic dysfunctions associated with HTLV-1 infection include mild cognitive impairment, sensory neuropathy, and erectile dysfunction. There is no effective therapy.

DIAGNOSIS

In general, diagnosis of the inflammatory myelopathies is based on the clinical examination. MRI often shows a high T2 signal focal enlargement of the cord.

TREATMENT AND SECONDARY PREVENTION

Rx

Intravenous corticosteroid infusions (e.g., methylprednisolone, 1 g intravenously daily for 5 days) are usually the mainstay for treatment of acute attacks of inflammatory myelopathies.

In patients who do not respond to corticosteroids, plasma exchange is effective in the acute treatment of CNS demyelinating disorders; about 60% of patients show improvement at 6 months. Factors predicting improvement are initiation of treatment within 15 days of the onset of symptoms and evidence of early improvement.

METASTATIC SPINAL CORD COMPRESSION

When metastatic cancer invades the spine or epidural space, the resultant destruction and growth compress the spinal cord and lead to a myelopathy. The prevalence of metastatic spinal cord compression may be as high as 5% in patients with cancer, depending on the type of malignant neoplasm and its tendency to metastasize to bone. Prostate (Chapter 201), breast (Chapter 198), and lung (Chapter 191) cancers each account for approximately 15 to 20% of cases, and non-Hodgkin lymphoma (Chapter 185), renal cell cancer (Chapter 197), and multiple myeloma (Chapter 187) account for about 5 to 10% each.

Most metastatic disease causes compression as a result of an extradural lesion, although a smaller number of metastatic lesions can be intradural-extramedullary disease. Intramedullary metastases are rare. Symptoms can be caused by direct compression of the cord and roots in the epidural space as the result of direct extension from hematogenous metastasis to the vertebral body. However, some tumors (e.g., lymphomas) may grow through the intervertebral foramen without causing significant bone destruction; the accompanying edema can compromise the local vasculature and cause ischemic damage in addition to direct compression. Vertebral destruction can make the spine unstable and cause pathologic fractures that can lead to cord and root damage.

CLINICAL MANIFESTATIONS

About 90% of patients present with pain that is classically worse on lying down and increases with the Valsalva maneuver. If the nerve root is involved, the pain will have a radicular component; if there is bone collapse, pain can be made worse by movement. Muscle weakness is present in 35 to 75% of patients at the time of diagnosis, sensory deficits in 50 to 70% of patients, and autonomic dysfunction in 50 to 60% of patients. The range of signs will depend on the level of compression.

DIAGNOSIS

Spinal cord compression, which must be suspected when any patient with cancer complains of spine pain even in the absence of neurologic signs or symptoms, is a neurologic emergency. MRI is the test of choice because plain films, which might recognize bone metastases and vertebral collapse, will

miss soft tissue tumors that are in the epidural space and yield no information about the spinal cord itself. Conventional myelography should be used if MRI cannot be performed because of availability or the presence of metallic implants in the patient. Because up to 35% of patients have more than one site of metastasis, care should be taken to image the entire spine by MRI, myelography, or isotope bone scanning.

Differential Diagnosis

For extradural lesions, the differential diagnosis includes lipomas, fibromas, meningiomas, and chordomas as well as vascular malformations and abscesses. Intradural-extramedullary lesions include neurofibromas (Chapter 417), neurinomas, meningiomas, vascular malformations, and (less often) metastases. Arachnoid cysts, although benign, can cause compression through pressure effect. Finally, intramedullary lesions that can present as myelopathy and must be considered in the differential of metastatic cord compression include intramedullary vascular malformations, ependymomas, astrocytomas, and syringomyelia.

TREATMENT

Rx

Prompt initiation of corticosteroids (e.g., dexamethasone, loading dose of 10 to 16 mg followed by tapering over 10 to 14 days) and radiation therapy are the mainstays of initial therapy. Surgical decompressive surgery plus radiation therapy is better than radiation therapy alone for maintaining ambulation in patients who have a radioinsensitive tumor, have displacement of the spinal cord on MRI, have a single site of cord compression, and have not been totally paraplegic for more than 48 hours.

PROGNOSIS

Metastatic spinal cord compression usually occurs in the setting of metastases to multiple locations, and the expected survival prognosis is generally less than 6 months. Prognosis is improved in patients with malignant neoplasms that are sensitive to steroid therapy (especially lymphoma and leukemia) or are radiosensitive (e.g., multiple myeloma, small cell lung cancer). Patients who are ambulatory at the time of diagnosis, have a single site of compression, and had a less rapid onset of symptoms also generally have a better prognosis.

Grade A References

A1. Chou R, Fu R, Carrino JA, et al. Imaging strategies for low-back pain: systemic review and meta-analysis. *Lancet*. 2009;373:463-472.
A2. Gross A, Miller J, D'Sylva J, et al. Manipulation or mobilization for neck pain. *Cochrane Database Syst Rev*. 2010;5:CD004249.
A3. Chow RT, Johnson MI, Lopes-Martins RA, et al. Efficacy of low-level laser therapy in the management of neck pain: a systematic review and meta-analysis of randomised placebo or active-treatment controlled trials. *Lancet*. 2009;374:1897-1908.
A4. Fu LM, Li JT, Wu WS. Randomized controlled trials of acupuncture for neck pain: systematic review and meta-analysis. *J Altern Complement Med*. 2009;15:133-145.
A5. Michalsen A, Traitteur H, Ludtke R, et al. Yoga for chronic neck pain: a pilot randomized controlled clinical trial. *J Pain*. 2012;13:1122-1130.
A6. Dahm KT, Brurberg KG, Jamtvedt G, et al. Advice to rest in bed versus advice to stay active for acute low-back pain and sciatica. *Cochrane Database Syst Rev*. 2010;6:CD007612.
A7. Williams CM, Maher CG, Latimer J, et al. Efficacy of paracetamol for acute low-back pain: a double-blind, randomised controlled trial. *Lancet*. 2014;384:1586-1596.
A8. Chou R, Huffman LH, American Pain Society, American College of Physicians. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med*. 2007;147:492-504.
A9. Cherkin DC, Sherman KJ, Kahn J, et al. A comparison of the effects of 2 types of massage and usual care on chronic low back pain: a randomized, controlled trial. *Ann Intern Med*. 2011;155:1-9.
A10. Bronfort G, Evans R, Anderson AV, et al. Spinal manipulation, medication, or home exercise with advice for acute and subacute neck pain: a randomized trial. *Ann Intern Med*. 2012;156:1-10.
A11. Vas J, Aranda JM, Modesto M, et al. Acupuncture in patients with acute low back pain: a multicentre randomised controlled clinical trial. *Pain*. 2012;153:1883-1889.
A12. Tilbrook HE, Cox H, Hewitt CE, et al. Yoga for chronic low back pain: a randomized trial. *Ann Intern Med*. 2011;155:569-578.
A13. Chou R, Atlas SJ, Stanos SP, et al. Nonsurgical interventional therapies for low back pain. A review of the evidence for an American Pain Society Clinical Practice Guideline. *Spine*. 2009;34:1078-1093.
A14. Buchmuller A, Navez M, Milletre-Bernardin M, et al. Value of TENS for relief of chronic low back pain with or without radicular pain. *Eur J Pain*. 2012;16:656-665.
A15. Kivitz AJ, Gimbel JS, Bramson C, et al. Efficacy and safety of tanezumab versus naproxen in the treatment of chronic low back pain. *Pain*. 2013;154:1009-1021.
A16. Wang X, Wanyan P, Tian JH, et al. Meta-analysis of randomized trials comparing fusion surgery to non-surgical treatment for discogenic chronic low back pain. *J Back Musculoskelet Rehabil*. 2014; [Epub ahead of print].
A17. Friedly JL, Comstock BA, Turner JA, et al. A randomized trial of epidural glucocorticoid injections for spinal stenosis. *N Engl J Med*. 2014;371:11-21.
A18. Ohtori S, Miyagi M, Eguchi Y, et al. Epidural administration of spinal nerves with the tumor necrosis factor-alpha inhibitor, etanercept, compared with dexamethasone for treatment of sciatica in patients with lumbar spinal stenosis: a prospective randomized study. *Spine (Phila Pa 1976)*. 2012;37:439-444.
A19. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical versus nonsurgical therapy for lumbar spinal stenosis. *N Engl J Med*. 2008;358:794-810.
A20. Nikolaidis I, Fouyas IP, Sandercock PA, et al. Surgery for cervical radiculopathy or myelopathy. *Cochrane Database Syst Rev*. 2010;1:CD001466.
A21. Bicket MC, Horowitz JM, Benzon HT, et al. Epidural injections in prevention of surgery for spinal pain: systematic review and meta-analysis of randomized controlled trials. *Spine J*. 2015;15:348-362.
A22. Patchell RA, Tibbs PA, Regine WF, et al. Direct decompressive surgical resection in the treatment of spinal cord compression caused by metastatic cancer: a randomized trial. *Lancet*. 2005;366:643-648.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Mader R, Verlaan JJ, Buskila D. Diffuse idiopathic skeletal hyperostosis: clinical features and pathogenic mechanisms. *Nat Rev Rheumatol*. 2013;9:741-750.
2. Chou R, Qaseem A, Owens DK, for the Clinical Guidelines Committee of the American College of Physicians, et al. Diagnostic imaging for low back pain: advice for high-value health care from the American College of Physicians. *Ann Intern Med*. 2011;154:181-189.
3. Leaver AM, Maher CG, McAuley JH, et al. People seeking treatment for a new episode of neck pain typically have rapid improvement in symptoms: an observational study. *J Physiother*. 2013;59:31-37.
4. Campbell P, Foster NE, Thomas E, et al. Prognostic indicators of low back pain in primary care: five-year prospective study. *J Pain*. 2013;14:873-883.
5. Deyo RA, Mirza SK, Martin BI, et al. Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA*. 2010;303:1259-1265.
6. Suri P, Rainville J, Kalichman L, et al. Does this older adult with lower extremity pain have the clinical syndrome of lumbar spinal stenosis? *JAMA*. 2010;304:2628-2636.
7. el Barzouhi A, Vleggeert-Lankamp CL, Lycklama a Nijeholt GJ, et al. Magnetic resonance imaging in follow-up assessment of sciatica. *N Engl J Med*. 2013;368:999-1007.
8. Wong JJ, Cote P, Quesnele JJ, et al. The course and prognostic factors of symptomatic cervical disc herniation with radiculopathy: a systematic review of the literature. *Spine J*. 2014;14:1781-1789.
9. Kreiner DS, Hwang SW, Easa JE, et al. An evidence-based clinical guideline for the diagnosis and treatment of lumbar disc herniation with radiculopathy. *Spine J*. 2014;14:180-191.
10. Lurie JD, Moses RA, Tosteson AN, et al. Magnetic resonance imaging predictors of surgical outcome in patients with lumbar intervertebral disc herniation. *Spine (Phila Pa 1976)*. 2013;38:1216-1225.
11. Johans SJ, Amin BY, Mummaneni PV. Minimally invasive lumbar decompression for lumbar stenosis: review of clinical outcomes and cost effectiveness. *J Neurosurg Sci*. 2015;59:37-45.
12. Wang JC, Dailey AT, Mummaneni PV, et al. Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 8: lumbar fusion for disc herniation and radiculopathy. *J Neurosurg Spine*. 2014;21:48-53.
13. Nascimento FA, Gatto LA, Lages RO, et al. Diffuse idiopathic skeletal hyperostosis: A review. *Surg Neurol Int*. 2014;5:S122-S125.
14. McClendon J, Sugrue PA, Ganju A, et al. Management of ossification of the posterior longitudinal ligament of the thoracic spine. *Neurosurg Focus*. 2011;30:E16.
15. Robertson CE, Brown RD Jr, Wijdicks EF, et al. Recovery after spinal cord infarcts: Long-term outcome in 115 patients. *Neurology*. 2012;78:114-121.

REVIEW QUESTIONS

1. A 34-year-old otherwise healthy construction worker presents to your office with a 3-week history of low back pain. He does not recall any specific trauma. He denies pain or weakness in his legs. His examination is notable for a decreased right ankle tendon reflex compared with the left. The most appropriate next step is to:

A. Order magnetic resonance imaging (MRI) of the lumbar spine
B. Order an MRI of the lumbar and thoracic spine
C. Order an electromyogram (EMG) of the right leg
D. Prescribe 3 days of bed rest
E. Prescribe PRN analgesics and physical therapy

Answer: E Uncomplicated acute low back pain, with or without radiculopathy, is generally self-limited. Imaging studies are unnecessary unless any of the red flags (see Table 400-3) are present. The American College of Physicians recommends MRI only in patients who have serious or progressive neurologic deficits, in whom a serious underlying condition is expected, or in whom surgery or epidural steroids are being considered.

2. Long-term disability owing to mechanical neck pain is associated with all of the following ***except***:

A. Tobacco use
B. Age younger than 45 years
C. More severe pain at the onset
D. Tendency to somatization
E. Employment requiring heavy lifting

Answer: B Patients younger than 45 years have less recurrence. Long-term disability is more common with obesity, low education level, tobacco use, high levels of pain at the onset, tendency toward somatization, job dissatisfaction, lack of availability of light-duty employment, and need to perform significant lifting at work.

3. All of the following are true statements regarding spinal stenosis ***except***:

A. The symptoms may be intermittent.
B. There is a good correlation between canal diameter and symptoms.
C. Pain is almost always present.
D. MRI may be useful in localizing the site of stenosis.
E. EMG is generally not useful except to evaluate comorbid conditions.

Answer: B Lumbar spinal stenosis can present with intermittent or persistent signs and symptoms referable to both cord and roots; axial neck or back pain is almost always present. There is, however, poor correlation between symptoms and the diameter of the canal. MRI can localize and quantify the severity of the stenosis. Computed tomography (CT) myelography, which provides the details needed for surgical treatment, is another alternative. EMG generally is not useful in the diagnosis of spinal stenosis except to assess possible comorbid conditions.

4. An 81-year-old man presents to your office with 2 months of moderately severe mid-back pain. In certain positions, the pain can radiate to his right lateral ribcage and even anteriorly toward his umbilicus. All of the following would support a diagnosis of myelopathy ***except***:

A. Sensation of tingling in his left leg
B. Weakness of both legs
C. Difficulty with urinary incontinence
D. Depressed reflexes in his right leg
E. Positive Babinski reflex in his left leg

Answer: D Although there may be decreased reflexes in an acute myelopathy, reflexes would usually be increased in a chronic myelopathy of 2 months' duration. In this case, the presence of radicular pain helps with localizing a level but does not eliminate the possibility of a comorbid myelopathy.

5. Metabolic myelopathies have been associated with all of the following ***except***:

A. Vitamin A deficiency
B. Vitamin B_{12} deficiency
C. Vitamin E deficiency
D. Copper deficiency
E. None of the above

Answer: A Metabolic myelopathies can be caused by a deficiency in vitamin B_{12}, vitamin E, and copper. Vitamin A deficiency is most commonly associated with blindness, xerophthalmia, poor bone growth, dermatologic problems, and impaired immune function.

401 REGIONAL CEREBRAL DYSFUNCTION: HIGHER MENTAL FUNCTIONS

DAVID S. KNOPMAN

DEFINITION

Higher mental function is at the core of what defines competent, independent individuals. Impairment of higher mental function can be broadly classified into four categories. Intellectual developmental disorder is a form of cognitive impairment that is present from infancy. Acquired forms of cognitive impairment are delirium, dementia, and focal cognitive disorders. Delirium (Chapter 28) is defined by its acute or subacute onset and coexistent alterations in alertness. Dementia (Chapter 402) represents an acquired cognitive impairment that is usually gradual in onset and not associated with alterations in alertness. Focal cognitive disorders involve only one aspect of cognition: memory, language, visuospatial cognition, or executive cognitive functioning, each of which is supported by a different cerebral region.

For the majority of patients in a non-neurology practice, a global description such as "normal mental function" or "cognitively impaired" will suffice. Cognitive impairment then becomes a diagnosis that subsumes all forms of altered higher mental function regardless of which domains are affected or how severely they are affected.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

An informal conversation with a patient lacks sensitivity for detecting cognitive impairment. If cognitive impairment is suspected from the patient's history, formal assessments should be performed. Bedside evaluations of orientation, memory, language, reasoning, and visuospatial function can be used to derive an overall view of cognitive function but do not automatically translate into diagnoses, because alertness, cooperation, education, native language, sensorimotor function, and mood must be taken into account. Although scores on bedside mental status examinations correlate strongly with severity and prognosis, they provide only rough guides to cognitive ability and cannot localize a cognitive deficit anatomically in the brain. The Mini-Cog Test (Chapter 27, Table 27-5) is among the most brief of available bedside examinations. If cognitive dysfunction is discovered in the course of the bedside examination, further exploration of individual cognitive domains must be undertaken.

MEMORY FUNCTION AND AMNESIC DISORDERS

DEFINITION

Human memory operates over a wide time range, from seconds to decades, and with quantities of information ranging from a single word to a lifetime's experience. Each neural system that achieves this monumental dynamic range has its own brain localization[1,2] (Table 401-1).

TABLE 401-1 DESCRIPTION OF MEMORY SYSTEMS

TYPE OF MEMORY FUNCTION	REGIONAL LOCALIZATION	LEARNING EFFICIENCY	TIME SPAN UNTIL EFFECTIVE RETRIEVAL	CAPACITY	CLINICAL TESTING TECHNIQUES	EXAMPLES IN DAILY LIFE
Declarative episodic memory	Hippocampus, medial thalamus	Single exposure	Decades	Very large, with rehearsal and elaboration	Recall of 3-4 words after 5 minutes	Recall of recent events and conversations
Declarative semantic memory	Temporal-parietal association cortices	Capable of single exposure; enhanced with repetition	Decades	Very large, perhaps limitless	Confrontation naming, general knowledge	Vocabulary, knowledge of life events from remote past
Attention span, "immediate memory"	Primary auditory or visual cortex	Single exposure only	Seconds	Very small: 7 ± 2 digits (auditory)	Digit span	Dialing a telephone number after hearing it or reading it
Working memory	Lateral frontal cortex	Single exposure only	Seconds	Small	Digits backward	Supporting many mental activities, such as mental arithmetic, abstract reasoning
Procedural memory	Basal ganglia, probably association neocortices	Requires extensive training	Decades	Moderate	Experimental laboratory methods only	Retention of motor skills (e.g., riding a bicycle, typing)

Declarative memory describes the type of learning and retrieval of facts and information that occur with conscious attention and intent; examples include remembering conversations, events, and intentions. Declarative memory has semantic and episodic components. Semantic memory refers to the brain's storehouse of knowledge, words, and facts. Episodic memory refers to learning and recall of specific events. Retention of information for more than a few seconds in the face of exposure to additional facts, details, or events requires declarative episodic memory to store and organize the information suitable for later recall. It is this declarative episodic memory system that is assessed as "memory" in the clinical setting. Anterograde amnesia is the clinical manifestation of disturbances in declarative episodic memory. Anterograde refers to failure to learn, and hence recall, new information on an ongoing basis. Most disorders of memory also exhibit retrograde amnesia, a disturbance of the ability to retrieve information from the past.

Immediate recall of information with zero delay and zero intervening information is a very short-term declarative memory function. Immediate memory is capable of storing an image of an auditory message in exact form, but only a small amount and for a short period. The fidelity of immediate memory recall accuracy drops off dramatically over seconds, particularly if intervening sensory stimuli attract attention. A comparable system exists in the visual modality in that the memory acts like a photograph that fades rapidly. From a clinical perspective, immediate memory is separate from declarative episodic memory. Immediate recall is generally used as a marker of attention and alertness and not memory per se.

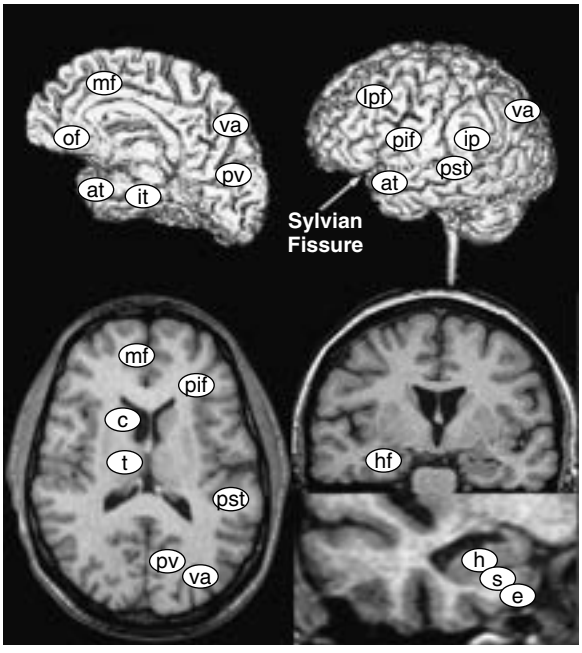

FIGURE 401-1. Magnetic resonance images of a normal brain. *Upper left,* Midsagittal view; *upper right,* left lateral view; *lower left,* axial view through the head of the caudate and body of the thalamus; *lower right,* coronal view through the mammillary bodies, with a magnified view of the medial temporal lobe. at = anterior temporal; c = caudate nucleus; e = entorhinal cortex; h = hippocampus; hf = hippocampal formation; ip = inferior parietal cortex; it = inferior temporal; lpf = lateral prefrontal cortex; mf = medial frontal cortex; of = orbital frontal cortex; pif = posterior inferior frontal cortex (Broca's area); pst = posterior superior temporal; pv = primary visual cortex (area 17); s = subiculum; t = temporal; va = visual association cortex (areas 18 and 19). (Courtesy Maria Shiung and Clifford Jack, MD.)

PATHOBIOLOGY

The hippocampal formations are the anatomic structures of importance for the declarative episodic memory system. The hippocampal formations are imaged well with magnetic resonance imaging (MRI) (Fig. 401-1). The principal input to the hippocampus comes through the entorhinal cortex from multimodal association areas in the frontal, parietal, and temporal neocortices. A second important input is a cholinergic pathway that originates in the septum of the medial-orbital frontal lobe. There are two principal output circuits of the hippocampal formations. One is via the subiculum back to multimodal association areas. The other hippocampal efferent pathway projects via the fornix to the mammillary bodies. The projection from the mammillary bodies passes through the medial thalamus to the ventral anterior nucleus of the thalamus, then to the posterior cingulate, and then back to the entorhinal cortex. The hippocampal circuit is believed to facilitate the formation of memory in association neocortices. The hippocampus does not store a particular learned fact, but rather it enables the appropriate region in a multimodal association cortical region to do so.

Lesions in one hippocampal formation will not generally have as devastating an impact on episodic memory as bilateral lesions will. However, in older persons who may have subclinical bilateral hippocampal pathology, a unilateral lesion, particularly in the dominant hemisphere, may produce a dense anterograde amnesia. Lesions in the columns of the fornix, mammillary bodies, and medial thalamus have also been linked to anterograde amnesia.

CLINICAL MANIFESTATIONS

Patients with anterograde amnesia have poor or no recollection of events, conversations, or observations. Family members report that patients repeat themselves in conversation or re-ask the same questions over the course of a few minutes to hours. Patients will generally forget important events and conversations, even when they were fully engaged in them. They will lose track of the date and time of day. They will forget appointments, even with reminders. Generally, patients with anterograde amnesia will fail to encode most events and happenings around them. The consequences of such memory

failure are usually more evident to the family and acquaintances of patients with the disorder than they are to the patients themselves. Anosognosia (lack of awareness) for the deficit of anterograde amnesia is very common, though not universal. Patients who most vehemently complain of memory loss are often suffering from depression rather than focal cognitive dysfunction.

Because some degree of forgetting is ubiquitous in human experience, it is challenging to distinguish between "everyday" forgetting and forgetting that is pathologic. All adults occasionally misplace important items, overlook an appointment, or forget some part of a conversation. In cognitively normal individuals, distraction, preoccupation, inattention, exhaustion, sleep deprivation, or other major life stressors inevitably produce some instances of excess forgetting. Pathologic forgetting as a result of a brain disorder produces a much greater degree of forgetting than occurs in the course of normal daily life, but there is no formulaic description of the boundary at which normal forgetting ends and pathologic forgetting begins.

DIAGNOSIS

The diagnosis of anterograde amnesia begins with a complaint of memory impairment from the patient or someone close to the patient. Testing of memory can be performed at the bedside in alert patients. The patient is asked to learn three or four words and recall them after 1 or 2 minutes. A patient with severe anterograde amnesia will recall none or at most one of the words, whereas individuals with normal memory can recall all of the words or all but one.

In patients with questionable memory difficulties, assessment by an experienced neuropsychologist is often a necessary part of the evaluation. Standardized tests of memory have greater precision and reliability and involve the use of lengthier material to be remembered and a longer delay between learning and recall.

Determining the Cause

Alzheimer's disease is the most common disorder in which anterograde amnesia occurs (Chapter 402). In Alzheimer's disease, anterograde amnesia is usually the dominant cognitive symptom, particularly early in the illness. Hippocampal atrophy is common (Chapter 402, Fig. 402-3). Anterograde amnesia also occurs in other dementing illnesses, such as vascular dementia and dementia with Lewy bodies.

Strokes can damage regions involved in episodic memory. Occlusion of the medial temporal branch of the posterior cerebral artery causes infarction of the hippocampus. Infarction in the territory of penetrating branches of the tip of the basilar artery causes bilateral medial thalamic infarcts.

Anterograde amnesia may be a major residual deficit after herpes simplex encephalitis (Chapter 374). Herpes simplex encephalitis has a predilection for damaging structures at the base of the cerebral hemispheres; frequently, the temporal lobes are severely damaged. Korsakoff's syndrome, the residual of the encephalopathy of thiamine deficiency (Chapter 416), is characterized by profound anterograde amnesia. Hemorrhagic necrosis of the mammillary bodies occurs in Korsakoff's syndrome. Survivors of closed head injuries (Chapter 399) may have anterograde amnesia because the medial temporal lobes are vulnerable to trauma as a result of their close proximity to the temporal bone. Survivors of an episode of anoxic-ischemic encephalopathy may also have dense anterograde amnesia. The pyramidal neurons of the CA1 region of the hippocampus are particularly vulnerable to hypoxic injury.

The syndrome of transient global amnesia involves anterograde amnesia, but the duration of the amnesia is a matter of 6 to 12 hours rather than the weeks or months seen in post-traumatic amnesia or the permanent deficits in patients with Alzheimer's disease or Korsakoff's syndrome. Patients with transient global amnesia remain alert though inattentive; the key element of the syndrome is that they lay down no new memories during the event. As a consequence, they are amnestic for the several hours of the episode. Transient global amnesia generally affects middle-aged or elderly individuals. Its cause is not known, although it is not usually due to typical cerebrovascular disease or epilepsy. Electroencephalography is typically not specifically abnormal, but diffusion-weighed MRI often shows distinctive abnormalities of the hippocampus a day or more after the onset of transient global amnesia.

THE APHASIAS

DEFINITION

Aphasia is a disorder of language at the conceptual level. Aphasics may have difficulty producing language, comprehending language, or both.

PATHOBIOLOGY

In more than 99% of right-handed individuals, language is localized to the left hemisphere. In left-handed individuals, language is also predominantly localized to the left hemisphere, although varying degrees of bilateral or rarely right hemispheric dominance may be seen. The hemisphere involved in language is referred to as the dominant hemisphere. Anatomic differences in the temporal and parietal lobes of the dominant hemisphere versus the other hemisphere also reflect its specialization for language.

Different aspects of language processing can be localized to specific regions within the dominant hemisphere. Experimental studies with positron emission tomography and functional magnetic resonance imaging can provide a rather detailed perspective on localization of various language subfunctions,[3] but the clinical neuroanatomy of language is less precise. Conceptualizing language functions as receptive or expressive, there are a few major clinical-anatomic relationships. Lesions in the dominant hemisphere's auditory association areas cause receptive language dysfunction. The critical regions are located in the superior temporal lobes adjacent to the primary auditory cortex and in the adjacent supramarginal and angular gyri of the inferior parietal lobule, an area known as Wernicke's area. Lesions in the dominant hemisphere's lateral inferior posterior frontal lobes, often referred to as Broca's area, result in expressive language deficits. Loss of access to one's vocabulary for either understanding spoken language or expressing oneself results from lesions in any portion of the region of the dominant hemisphere around the sylvian fissure, including the lateral posterior inferior frontal lobe, the inferior parietal lobule, and the superior and middle temporal gyri. Coronal and axial MRI scans give a detailed view of the critical language regions (see Fig. 401-1).

In clinical practice, aphasia may be caused by cerebrovascular or neurodegenerative diseases, especially frontotemporal lobar degenerations and Alzheimer's disease[4,5] (Chapter 402). Less commonly, space-occupying lesions such as brain tumors (Chapter 189) or brain abscesses (Chapter 413) can cause aphasic syndromes.

CLINICAL MANIFESTATIONS

The language comprehension difficulties in persons with aphasia must be distinguished from hearing disorders (Chapter 428), and the motor speech dysfunction in aphasia must be distinguished from dysarthria. Errors of articulation in persons with aphasia reflect altered conceptual selection of what is to be said. In aphasia, mispronunciation of a sound within one word may be followed by perfect pronunciation of the same sound in a different word. In dysarthria, by comparison, the errors in articulation or phonation are consistent.

Aphasia has three principal components: impaired verbal comprehension, disordered verbal expression, and impaired naming. Disorders of reading, writing, and sentence repetition are additional elements of the aphasia syndrome. The disordered verbal comprehension may range from profound to mild. When profound, patients are unable to grasp the meaning of single words. In milder forms of disordered comprehension, patients may be able to follow one-step but not two- or three-step commands. Usually, the comprehension difficulty involves both spoken and written language, but each can be affected separately. Anomia, which is an inability to produce names of people or objects, is common in almost all aphasic syndromes.

In expressive aphasic syndromes, written material and spoken speech are most often affected in parallel. Speech is labored in the expressive aphasias, and it lacks the normal melody and variation in intonation that characterize normal speaking. Melody and intonation are referred to as the prosody of speech. Speech is often grammatically impoverished. The number of words per utterance is greatly reduced, thus giving the speech a choppy, staccato character. These features are referred to as speech apraxia. *Nonfluency* is a related term that describes the reduced number of words and the terseness of verbal output. In some aphasic syndromes, speech is often degraded by anomia and paraphasic errors (word or syllable substitutions), even when fluency, melody, and intonation are preserved.

Specific Aphasic Syndromes

Specific common aphasic syndromes exhibit various combinations of receptive and expressive difficulty (Table 401-2).

WERNICKE'S APHASIA

In Wernicke's aphasia, verbal comprehension of both written and verbal language is severely impaired. Patients with Wernicke's aphasia have difficulty

TABLE 401-2 MAJOR APHASIC SYNDROMES

APHASIA SYNDROME	REGIONAL LOCALIZATION	SPONTANEOUS SPEECH ABNORMALITIES	AUDITORY COMPREHENSION	CONFRONTATION NAMING	SENTENCE REPETITION
Broca's aphasia	Lateral inferior frontal lobe	Nonfluent, labored, agrammatic	Preserved	Poor	Poor
Wernicke's aphasia	Posterior superior temporal-parietal supramarginal gyrus	Fluent, many paraphasic errors, very little information content	Very impaired	Poor	Poor
Global aphasia	Major portions of the frontoparietal operculum and superior temporal lobe	Nonfluent or virtually absent	Very impaired	Poor	Poor
Anomic aphasia	Small lesion somewhere in the perisylvian region	Fluent, may contain some paraphasias	Normal or mildly impaired	Poor to moderately impaired	Preserved or impaired

understanding the meaning of individual words and may not be able to follow any command consisting of greater than one step. Their speech is fluent but marred by paraphasia and anomia. Wernicke's aphasics tend to lack awareness of the extent of their communicative difficulties and are often unaware that the words they are uttering are fundamentally incorrect. Embolic strokes are the most common cause of Wernicke's aphasia. The location that typically causes Wernicke's aphasia is the dominant posterior superior temporal lobe or inferior supramarginal gyrus (see Fig. 401-1).

SEMANTIC VARIANT OF PRIMARY PROGRESSIVE APHASIA

The aphasic disturbance of semantic variant of primary progressive aphasia is characterized by a loss of access to the meaning of words. Spontaneous speech melody, intonation, and grammatical integrity are preserved, but patients have marked difficulties with production of nouns and verbs. This condition is usually caused by left anterior temporal lobe degeneration owing to one of the frontotemporal lobar degenerations (Chapter 402).

BROCA'S APHASIA

Broca's aphasia is a syndrome in which expressive language is prominently affected. Patients with Broca's aphasia have nonfluent labored speech. The location of the lesion that typically causes Broca's aphasia is the dominant posterior inferior frontal lobe (see Fig. 401-1). The typical syndrome is usually due to embolic strokes. Patients with Broca's aphasia have largely preserved comprehension and as a result are acutely aware of their difficulties and become frustrated with them. Depression is common in Broca's aphasics.

NONFLUENT/AGRAMMATIC VARIANT OF PRIMARY PROGRESSIVE APHASIA

The nonfluent/agrammatic variant of primary progressive aphasia is characterized by the gradual onset of labored, hesitant, sparse speech that is often grammatically impoverished. Comprehension of spoken speech is typically preserved. This syndrome is usually caused by one of the frontotemporal lobar degenerations (Chapter 402).

GLOBAL APHASIA

Global aphasia occurs when both expressive and receptive problems are present. Global aphasia often appears acutely after a major infarction, hemorrhage, or traumatic brain injury involving the dominant hemisphere. Global aphasia may also be present in the context of severe dementia.

ANOMIA

Anomia is at the milder end of the spectrum of language disorders. Some anomic aphasics also have difficulty with sentence repetition, even in the presence of relatively preserved comprehension and verbal expressive abilities. There is some controversy whether this latter syndrome, called conduction aphasia, represents a disconnection between the perisylvian centers for comprehension and expression or whether it represents a lesion in the cortical auditory areas involved in immediate auditory memory.

IDEOMOTOR APRAXIA

Ideomotor apraxia is a disorder at the interface between comprehension and execution of facial or limb motor actions. Patients with ideomotor apraxia have no paresis of the face or limb musculature and are able to carry out simple tasks, but they are unable to execute more complex tasks or commands. For example, in a woman who is able to name a comb and use her right hand to point to parts of her body, ideomotor apraxia can be demonstrated if she is unable to indicate through her actions how she would use the comb.

STUTTERING

The left pars opercularis is a locus where the intrinsic functional architecture of speech-language processes is altered in patients with persistent developmental stuttering.[6]

DIAGNOSIS

The diagnosis of aphasia is made by listening to the patient speak and by examining comprehension, naming ability, reading, and writing in a standardized fashion. Frequently the diagnosis of aphasia is made during attempts to obtain a history from the patient. It is helpful to prompt patients to speak about a neutral topic such as what they had for their last meal or what they did the previous day. Listening to their spontaneous speech allows the examiner to characterize its fluency, grammatical form, articulation, melody, and intonation, as well as difficulty finding words, the presence of paraphasias, and the overall information content.

Comprehension should be examined formally by asking the patient to perform tasks that range from one to at least three steps. Naming can be tested by asking the patient to name a series of common objects, such as the parts of the hand and arm (e.g., thumb, palm, knuckles, wrist, elbow). In general, the more commonly a word is used in the language, the easier it will be to name, whereas infrequent words are harder for aphasics. Reading and writing should also be tested.

Portions of the dominant perisylvian cerebral cortex may be damaged by infarction (Chapters 407 and 408), hemorrhage, and other space-occupying brain lesions such as neoplasms (Chapter 189) and abscesses (Chapter 413). Aphasia secondary to stroke has an abrupt onset, usually with some subsequent improvement. Recovery from aphasia after a stroke may occur as ischemic zones around an infarction eventually regain function. Regions remote from the infarction may also be synaptically depressed acutely after a stroke (diaschisis) but eventually regain function. Finally, regions in the nondominant hemisphere may become more active over the course of recovery. Aphasia that has a gradual and slowly progressive onset occurs in the degenerative dementia syndromes of progressive aphasia and semantic dementia (Chapter 402).

TREATMENT

Rx

Speech therapy may be helpful for patients in the first few months after a brain injury that causes aphasia.

CORTICAL DISORDERS OF VISUAL FUNCTION AND HEMISPATIAL NEGLECT

DEFINITIONS

Cortical disorders of vision and spatial cognition are caused by lesions in the occipitoinferotemporal or occipitoposteroparietal lobes. The principal disorders of cortical visual functioning are alexia (impaired reading), object agnosia (impaired recognition of visual forms), and prosopagnosia (impaired face recognition). The principal disorders of spatial cognition are simultanagnosia (impaired integration of complex visual scenes), dressing apraxia, and visual hemispatial neglect (lack of awareness of the personal or extrapersonal hemispace). Diagnosis of a cortical visual disorder requires integrity of primary visual function from the cornea to the lateral geniculate nuclei.

PATHOBIOLOGY

Higher visual function is localized to a network centered in the occipital lobe and includes the inferior temporal and posterior parietal lobes[7] (see Fig. 401-1). From area 17, processing of visual information passes to visual association areas 18 and 19. From there it proceeds in several directions. Disorders of higher visual function can be related to a ventral or dorsal pathway. The ventral pathway from the visual centers to the medial temporal lobe links visual information to meaning ("What is the object?"). The dorsal visual processing pathway has several target regions. One links the visual centers to the parietal lobes and is concerned with locating objects in space and determining spatial relationships among objects in order to grasp a complete visual scene ("Where is the object?"). Another integral part of the dorsal visual processing stream is the cortical control of the extraocular muscles in the parietal and prefrontal regions, whereby the eyes are directed to various elements of a visual scene so that the individual elements are synthesized into a coherent ensemble. Yet a third part of the dorsal visual pathway leads to premotor areas that, in conjunction with eye movement control, facilitate visually guided limb motor actions.

Alexia occurs as a result of lesions in the ventral pathway of the dominant hemisphere. Object agnosia may also occur with lesions, usually bilateral, in the ventral pathway. Alexia and object agnosia occur with neurodegenerative diseases that affect the parieto-occipital cortex (Fig. 401-2). Simultanagnosia, dressing apraxia, and hemispatial neglect are syndromes caused by lesions in the dorsal pathway. Limb apraxia and impaired visuomotor activities may result from disruption of the premotor pathways that interact with the dorsal visual system. Simultanagnosia usually requires bilateral posterior parietal lesions. Dressing apraxia and hemispatial neglect arise from unilateral lesions, most often in the nondominant hemisphere. Cortical blindness is a consequence of bilateral occipitoparietal pathology.

CLINICAL MANIFESTATIONS

Alexia may occur as an isolated deficit, or it may occur in the context of other evidence of aphasia. Patients may be able to recognize individual letters but are unable to recognize a string of letters as a word. In pure alexia, auditory comprehension of words and sentences is preserved. Patients with object agnosia may be unable to identify objects visually, but they will be able to recognize the object based on its characteristic sound or how it feels to touch. In simultanagnosia, patients may be able to identify small objects easily if they happen to appear within the narrow viewing area at the center of their visual field. At the same time, such patients will fail to grasp the bigger visual picture. They may appear to be functionally blind. This clinical picture is referred to as Balint's syndrome.

Patients with cortical disorders usually have difficulty with visuoconstructional tasks such as copying figures or drawing simple objects such as a flower, house, or clock. Dressing apraxia represents a deficit of practical significance in which patients are unable to comprehend the orientation of articles such as a shirt or a blouse and to manipulate them.

The most severe form of a cortical disorder of visuospatial processing is cortical blindness. In this condition, in which the anterior visual pathways can be reasonably believed to be intact, patients appear functionally blind. On occasion, they also exhibit anosognosia for the blindness and claim they can see. This latter condition is referred to as Anton's syndrome.

Hemispatial neglect occurs in the setting of acute strokes involving the nondominant perisylvian region. Even when there is no hemianopia as measured by single visual stimuli, presentation of double simultaneous stimuli to the patient reveals unawareness in the nondominant field. Hemispatial neglect can be demonstrated at the bedside with a task such as drawing a clock. A patient with hemispatial neglect will fail to place the numbers on the nondominant side (i.e., the left side in a right-handed person). Patients with hemispatial neglect may sometimes deny that their paretic limb belongs to them.

DIAGNOSIS

Information about visual functioning can be obtained from the history. The patient or the patient's informant may report that the patient cannot read, cannot read a clock, or cannot find objects when asked to get something off a table or out of a cupboard. There is often a history of motor vehicle accidents in which the patient failed to see another vehicle, the curb, or the side of a garage. Patients may report difficulty recognizing people's faces even though they are able to recognize them by their voices or by other cues.

Bedside tests that screen for visuospatial deficits include either copying a simple geometric design or drawing an object. Intersecting pentagons and a cube are objects used clinically. Clock drawing is a brief but informative exercise. Reading of words or commands and naming of objects can be done at the bedside as well. Face recognition is more difficult to assess at the bedside. Formal testing of visuospatial function in the neuropsychology laboratory involves the use of specially designed instruments to characterize visual processing.

The etiology of lesions that cause deficits in cortical vision and spatial cognition ranges from focal cerebrovascular disease, neoplasms, infectious processes, and brain trauma to neurodegenerative disorders. When a stroke causes a disorder of cortical visuospatial processing or hemispatial neglect, it is usually abrupt in onset. Space-occupying brain lesions such as neoplasms or brain abscesses that cause cortical visual disorders do so on a subacute basis. Disordered visuospatial function may also appear insidiously when caused by the degenerative disorder posterior cortical atrophy. Patients with posterior cortical atrophy, which is usually due to Alzheimer's disease, show marked atrophy of the occipital lobe (see Fig. 401-2).

EXECUTIVE COGNITIVE DYSFUNCTION AND CONTROL OF PERSONAL BEHAVIOR

DEFINITIONS

Integrative abilities that are broadly referred to as executive cognitive function include mental agility, abstract reasoning, and problem solving. Executive cognitive function represents processes that support mental flexibility, adaptability, focus, and tenacity. Control of personal actions and regulation of interpersonal relationships are also closely related to executive cognitive dysfunction. The term *comportment* denotes how a person behaves, particularly toward other people.

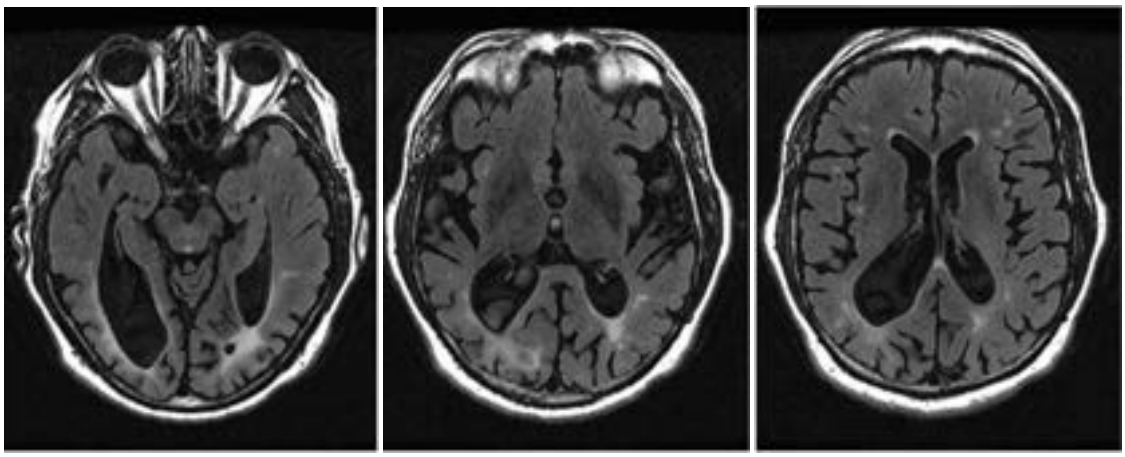

FIGURE 401-2. Magnetic resonance (MR) scans of a patient with the syndrome of posterior cortical atrophy caused by Alzheimer's disease. The MR scan shows marked atrophy of the primary visual areas and parieto-occipital association areas. The right hemisphere is more affected than the left.

PATHOBIOLOGY

The anatomic basis of executive cognitive function and comportment is a network of brain regions anchored by the prefrontal and anterior temporal lobe neocortex[8,9] (see Fig. 401-1). These regions receive input from multiple cortical and subcortical regions. The caudate nucleus is the site of a major frontal lobe efferent pathway. The medial thalamus is a major afferent source to the frontal lobes. The anterior temporal lobes are also part of the same integrative circuitry as the prefrontal regions. Lesions in the lateral prefrontal regions are associated with slowing of cognitive processing, difficulty with set shifting (switching from one idea or task to another), difficulty initiating tasks, and loss of mental flexibility. The frontal and anterior temporal lobes are involved in the modulation of personal behavior and interpersonal relationships. Patients with lesions in the medial prefrontal lobes are often apathetic and lack initiative. Patients with lesions in the orbital frontal or right anterior temporal lobe may exhibit disinhibition, impulsivity, and a striking loss of ability to interpret or predict the feelings of others.

Traumatic brain injury (Chapter 399) is a common cause of frontal lobe and anterior temporal damage. The orbital frontal, frontal polar, and anterior temporal regions are particularly vulnerable to contusions because of their proximity to the skull (Fig. 401-3). Patients with traumatic brain injuries may also suffer diffuse white matter damage as a result of shear injuries. Disconnection of the frontal and anterior temporal lobes from other parts of the brain can produce executive cognitive dysfunction and altered control of personal behavior.

CLINICAL MANIFESTATIONS

Executive cognitive functioning and control and regulation of behavior are usually affected concurrently. Patients with executive dysfunction are deficient in goal-oriented behavior; they lose the ability to predict the consequences of their actions or words. Patients with executive dysfunction also exhibit poor mental agility and inflexibility in their thinking and control of their actions. They are easily distracted and exhibit a tendency to perseverate, in which the answer to a prior question is repeated in response to subsequent questions. They are disinhibited; as a consequence, when asked to recall a specific event, they may glibly answer with a fabrication, a phenomenon referred to as confabulation.

Patients with lateral prefrontal pathology exhibit poor performance on tests of abstract reasoning and mental agility. In a test such as verbal similarities, they tend to be very concrete and narrowly focused. They become easily distracted and are slow in performing tasks that require sustained attention. Because of their mental rigidity and difficulty in set shifting, they do poorly on tests that require the ability to vary their response strategies, such as verbal fluency tests.

Patients with medial frontal lesions are often profoundly apathetic and lack initiative and motivation. They may be laconic and completely unable to express emotion, whether it be anger, sadness, or elation. They tend to be indifferent to their surroundings, a state referred to as abulia. The majority of patients with substantial prefrontal or anterior temporal lobe pathology lack insight into the extent of their inappropriate behavior.

In contrast, other patients with altered comportment exhibit different manifestations of dysregulation of personal actions and interpersonal behavior. These alterations may include difficulty controlling impulsivity, poor social graces (manifested as rude behavior or caustic comments), a disregard for the feelings of others (loss of empathy), and a general failure to understand what constitutes acceptable behavior in a particular social context. If the underlying disease is progressive, gross alterations in table manners and loss of interest in maintaining personal hygiene may appear. Inappropriate sexual behavior may occur. Patients with prominent disease of the frontal lobes may also exhibit hyperorality, which is a compulsion to put nonfood objects into their mouths. Hyperorality can be life-threatening, depending on the substance ingested.

DIAGNOSIS

The clinical history is essential for documenting the characteristic changes in personality, behavior, and interpersonal relationships. The history must almost always be obtained from an informant who knows the patient well, because the patient may assert that there are no problems.

Mental status examination is an integral part of the diagnosis of an executive cognitive disorder. Simply interacting with the patient may be quite revealing. The patient may exhibit abulia, disinhibition, socially inappropriate behavior, or easy distractibility. Tests of executive cognitive function that are suitable for bedside use include verbal similarities and differences, the digits backward test, reciting the months of the year backward, or spelling a word backward. Verbal fluency, which is a very useful test of mental flexibility and set shifting, is tested by asking patients to produce as many words as they can that begin with a particular letter of the alphabet in 60 seconds. Frequently, a patient with a frontal lesion will quickly produce two or three words and then stop.

Bedside testing of executive cognitive dysfunction provides only a superficial view of the cognitive domain. Assessment in the neuropsychology laboratory gives a more refined estimate of the degree of executive dysfunction.

Space-occupying lesions of the frontal lobes (e.g., neoplasms, brain abscesses) can lead to the cognitive and behavioral syndromes of frontal lobe dysfunction. With these diseases, executive cognitive dysfunction and alteration of control of personal behavior develop over a period of weeks.

In patients with acute brain trauma (Chapter 399), brain imaging at the time of initial medical and surgical evaluation will reveal whether the brain suffered acute traumatic lesions. Chronically, traumatic brain injury may later lead to encephalomalacia of the frontal lobes (see Fig. 401-3).

Neurodegenerative diseases such as frontotemporal lobar degeneration (Chapter 402) are associated with dysfunction and brain loss in the prefrontal (Chapter 402, Fig. 402-7) and anterior temporal lobes (Chapter 402, Fig. 402-8). These disorders may produce the entire spectrum of executive cognitive dysfunction and altered control of personal behavior over a period of a year or longer.

Some diseases that do not directly damage the frontal or anterior temporal neocortex may cause executive cognitive dysfunction and alteration of control of personal behavior because of the interconnectedness of the frontal and anterior temporal lobes with other cortical and subcortical regions. Multiple sclerosis (Chapter 411), a disorder of white matter pathways, may cause abnormalities in cognition and behavior of the frontal type. Similarly, Huntington's disease (Chapter 410) and progressive supranuclear palsy, which affect the caudate nuclei, may also resemble a frontal cognitive and behavioral

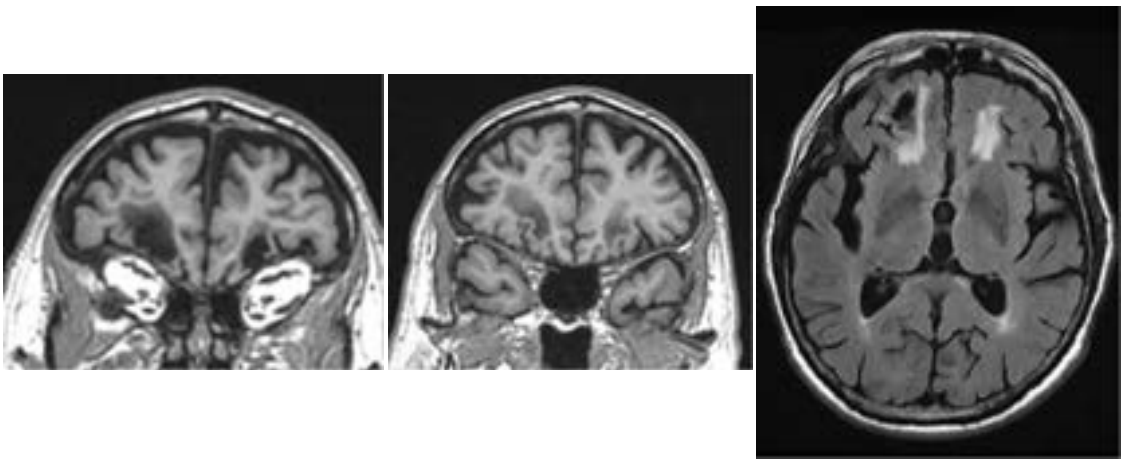

FIGURE 401-3. Coronal (*left and middle*) and axial magnetic resonance images of frontal brain trauma. This patient who suffered a closed head injury several years ago now has encephalomalacia in the orbital frontal cortices bilaterally.

syndrome and result in executive cognitive dysfunction and alterations in comportment.

TREATMENT

Rx

Cognitive-behavioral therapies offer modest but definite benefit for patients with aphasia and for those with mild attention deficits and mild memory deficits caused by brain injury.

FUTURE DIRECTIONS

The assessment of cognition is being supplemented by new imaging techniques. Functional magnetic resonance imaging can provide an unprecedented view into cortical connectivity patterns, which are influenced by aging and disease.[10]

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Eichenbaum H. Time cells in the hippocampus: a new dimension for mapping memories. *Nat Rev Neurosci*. 2014;15:732-744.
2. Henke K. A model for memory systems based on processing modes rather than consciousness. *Nat Rev Neurosci*. 2010;11:523-532.
3. Price CJ. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*. 2012;62:816-847.
4. Tippett DC, Niparko JK, Hillis AE. Aphasia: Current concepts in theory and practice. *J Neurol Transl Neurosci*. 2014;2:1042.
5. Mesulam MM. Primary progressive aphasia and the language network: the 2013 H. Houston Merritt Lecture. *Neurology*. 2013;81:456-462.
6. Lu C, Chen C, Peng D, et al. Neural anomaly and reorganization in speakers who stutter: a short-term intervention study. *Neurology*. 2012;79:625-632.
7. Kravitz DJ, Saleem KS, Baker CI, et al. A new neural framework for visuospatial processing. *Nat Rev Neurosci*. 2011;12:217-230.
8. Robinson H, Calamia M, Glascher J, et al. Neuroanatomical correlates of executive functions: a neuropsychological approach using the EXAMINER battery. *J Int Neuropsychol Soc*. 2014;20:52-63.
9. Possin KL, LaMarre AK, Wood KA, et al. Ecological validity and neuroanatomical correlates of the NIH EXAMINER executive composite score. *J Int Neuropsychol Soc*. 2014;20:20-28.
10. Barkhof F, Haller S, Rombouts SA. Resting-state functional MR imaging: a new window to the brain. *Radiology*. 2014;272:29-49.

REVIEW QUESTIONS

1. Episodic declarative memory is a cognitive function most closely associated with what brain region?

A. Hippocampal formations
B. Primary visual cortex
C. Supplementary motor areas
D. Putamen
E. Hypothalamus

Answer: A The hippocampal formations are the anatomic locus of encoding new information, the necessary first step in creating new memories for facts or events. The other regions play much more peripheral roles in episodic declarative memory.

2. All of the following are commonly seen in the syndrome of transient global amnesia ***except***:

A. Impaired recall of events
B. Brief loss of consciousness
C. Preserved ability to speak
D. Preserved ability to walk

Answer: B Loss of consciousness is not a feature of transient global amnesia. If a spell of impaired recall occurred in conjunction with loss of consciousness, other diagnoses such as a seizure disorder or syncope would be more likely.

3. Which brain region is most closely associated with comprehension of spoken language?

A. Dominant prefrontal cortex
B. Dominant inferior frontal lobe
C. Nondominant inferior parietal lobe
D. Dominant inferior parietal lobe
E. Dominant hippocampal formation

Answer: D The dominant inferior parietal lobe is the region most closely linked to auditory comprehension.

4. The syndrome of hemispatial neglect refers to:

A. Lack of awareness of one's own distorted speech
B. Lack of awareness of objects in the hemispace contralateral to a perisylvian (frontotemporal-parietal) lesion
C. Lack of awareness of objects in the hemispace ipsilateral to a perisylvian (frontotemporal-parietal) lesion
D. Weakness in limbs contralateral to a perisylvian (frontotemporal-parietal) lesion

Answer: B Hemispatial neglect occurs in the visual hemispace contralateral to perisylvian lesions. Lack of awareness of impaired speech, such as occurs in Wernicke's aphasia, is a form of loss of awareness of one's deficits, but not of hemispatial neglect. Weakness per se is not a manifestation of hemispatial neglect, although persons with nondominant hemisphere perisylvian lesions often experience anosognosia for their weakness.

5. All of the following may be seen with injury to the frontal lobes ***except***:

A. Apathy (abulia)
B. Lack of energy
C. Impulsiveness
D. Lack of regard for the feelings of others (loss of empathy)
E. Cortical blindness

Answer: E Cortical blindness is a manifestation of lesions in the occipital lobe. All of the other symptoms are ones seen in lesions in the frontal lobes.

402

ALZHEIMER DISEASE AND OTHER DEMENTIAS

DAVID S. KNOPMAN

DEMENTIA

DEFINITION

Dementia, which is a disorder of cognition, interferes with daily functioning and results in loss of independence (Table 402-1). The majority of dementias are of gradual onset, are progressive in course, and occur in persons with previously normal cognition. However, none of these features are necessary aspects of the definition of dementia. Some dementias, such as those caused by an acute neurologic illness secondary to stroke, encephalitis, or head trauma, may begin abruptly and then remain static for long periods. Conversely, a small subset of dementias, such as Creutzfeldt-Jakob disease (Chapter 415), have a rapid onset and a course that can run for less than a year. Dementia may also occur in persons with developmental disabilities and long-standing cognitive deficits.

TABLE 402-1 DEFINITION OF DEMENTIA

Dementia is cognitive impairment that interferes with the ability to function at work or at usual activities; *and* It represents a decline from prior levels of functioning and performing; *and* The cognitive impairment and impaired functioning are not explained by delirium or major psychiatric disorder.
The cognitive impairment of dementia is detected and diagnosed through a combination of: (a) History-taking from the patient and a knowledgeable informant; *and* (b) Objective cognitive assessment, either a "bedside" mental status examination or neuropsychological testing.
The cognitive or behavioral impairment of dementia involves **at least two** of the following domains: Impaired ability to acquire and remember new information Impaired reasoning and handling of complex tasks; poor judgment Impaired visuospatial abilities Impaired language functions (speaking, reading, writing) Changes in personality, behavior, or comportment

Adapted from McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011;7:263-269.

EPIDEMIOLOGY

The prevalence and incidence of dementia increase with advancing age. Dementia is uncommon before 50 years of age.[1] In individuals older than 65 years, the prevalence of dementia of all types is about 7%. In the age range of 65 to 69, the prevalence of dementia is only 1 to 2%, but it increases to 20 to 25% in the 85- to 89-year age range and continues to rise steadily thereafter. The incidence of new cases of dementia is about 1 per 100 per year at the age of 70 and rises to about 2 to 3 new cases per 100 per year by about the age of 80. Incidence rates continue to rise into the ninth and tenth decades of life. With the dramatic increase in longevity in North America, the societal burden of dementia has risen substantially.

In absolute numbers, far more women than men have dementia, because women live longer. However, men and women have an equal age-adjusted risk for the development of dementia. There are no racial or ethnic differences in the risk for dementia.

PATHOBIOLOGY

Dementia is the culmination of dysfunction in the cerebral hemispheres, especially the association cortices, hippocampal formations, their supporting subcortical nuclear structures (e.g., caudate nuclei, thalamus), and their white matter interconnections (see Fig. 401-1). Specific diseases that cause dementia do so by affecting particular parts of the cerebral cortex, subcortical nuclei, or the underlying white matter pathways linking different cortical regions.

CLINICAL MANIFESTATIONS

Any of the major domains of cognition—declarative episodic memory, executive cognitive functioning, visuospatial function, or language—may be affected in dementia (Chapter 401). Because Alzheimer disease is the most common dementia, anterograde amnesia is typically present first and most intensely in the majority of dementia patients. In other dementing illnesses, deficits in the other cognitive domains may be dominant. A pervasive and nearly invariant aspect of dementia is a loss of insight (anosognosia) into the extent of one's cognitive and functional losses.

Neuropsychiatric symptoms are also common in dementia. Apathy and loss of initiative are almost always present. Depression and anxiety are frequent, as are irritability, paranoia, delusional thinking, and hallucinations. Daily functioning of patients with dementia is compromised. In early dementia, difficulty is likely to be present in management of finances and medications, independent travel, preparation of meals, and keeping of appointments. In more advanced disease, difficulty becomes evident in basic activities of daily living such as bathing, dressing, toileting, and feeding oneself. Dementias secondary to cerebrovascular or Lewy body disease are often associated with specific abnormalities in strength, coordination, gait, or balance. Alzheimer disease, the most common dementia, typically has no associated motor abnormalities.

DIAGNOSIS

Clinical Examination

Dementia is strictly a clinical diagnosis based on evidence of cognitive dysfunction in both the history and the mental status examination.[2] The key elements of the history flow from the definition of dementia: What is the evidence for impairment in one or more domains of cognition? What is the evidence that daily functioning is affected? The mental status examination is necessary to establish that alertness is preserved (i.e., the patient does not have delirium [Chapter 28]) and to determine what specific areas of cognition exhibit directly observable impairment. For diagnosis of the syndrome of dementia, no laboratory test supersedes the clinical history and mental status examination. Laboratory testing is critical, however, to determine the cause of the dementia.

Bedside testing of mental status is based on the principles of cognitive neurology (Chapter 401). For moderate or severe dementia to be distinguished from normal cognitive states, a bedside mental status examination such as the Mini-Cog test (Chapter 27, Table 27-5) is accurate. However, for mild dementia, bedside mental status examinations lack sensitivity (i.e., they fail to diagnose some cases of mild dementia). For patients with suspected mild dementia,[3] neuropsychometric testing is a useful adjunct to the bedside examination. The neurologic examination is also important for evaluation of signs of specific causes of dementia, including signs of cerebrovascular disease (e.g., hemiparesis [Chapter 406]) and signs of extrapyramidal disease (e.g., rigidity, bradykinesia, resting tremor [Chapter 409]).

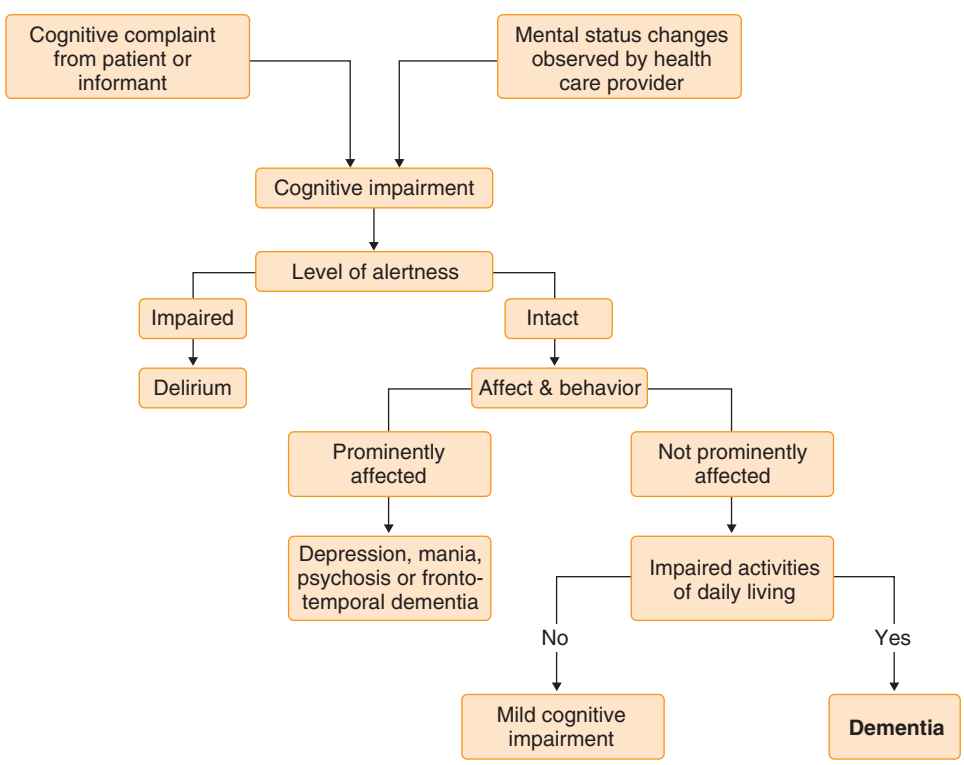

FIGURE 402-1. Flow diagram to establish the diagnosis of dementia.

Differential Diagnosis

Dementia must be distinguished from other disorders of cognition (Fig. 402-1). Delirium (Chapter 28) also affects cognition directly; key features distinguishing it from dementia include impaired arousal and attention. Delirium is almost always of sudden onset, whereas the majority of cases of dementia are of gradual onset.

Primary psychiatric diseases (Chapter 397) such as major depression, bipolar disorder, and schizophrenia may also impair cognition. In dementia, however, the impairment in cognition is typically equivalent to or more pervasive than the changes in mood and behavior.

The principal diseases that cause dementia are three neurodegenerative diseases—Alzheimer disease, Lewy body disease, and frontotemporal lobar degeneration—and cerebrovascular disease (Fig. 402-2). The neurodegenerative diseases that cause dementia are typically slow and insidious in onset and inexorably progressive. Dementia secondary to cerebrovascular disease may be of either sudden or gradual onset.

Many much less common secondary causes account for less than 2% of all dementias. Drug intoxication (Chapters 34 and 416), metabolic disorders (Chapter 205), central nervous system infections (Chapters 412 to 414), and brain structural lesions (Chapter 401) are typically subacute in onset; if they are diagnosed and treated early, the cognitive deficits improve or resolve completely. A number of medications such as sedatives, pain medications, corticosteroids, digoxin, and others cause mental confusion, particularly but not always at toxic levels (Chapter 110). Metabolic disorders that may also cause subacute confusion and produce a cognitive disorder include hypothyroidism or hyperthyroidism (Chapter 226), vitamin B_{12} deficiency (Chapter 416), chronic liver disease (Chapter 153), chronic renal failure (Chapter 130), and hypocalcemia or hypercalcemia (Chapter 245). Chronic viral infections of the brain, especially human immunodeficiency virus infection, frequently cause dementia (Chapter 394). Chronic meningitides in the differential diagnosis of dementia include cryptococcal meningitis (Chapter 336), tuberculous meningitis (Chapter 324), and tertiary syphilis (Chapter 319). Finally, structural lesions of the brain, including primary and metastatic tumors (Chapter 189), chronic subdural hematomas (Chapter 399), and normal-pressure hydrocephalus (Chapter 189), can cause a syndrome resembling dementia that consists of a subacute or slowly progressive decline in cognition with few or no other neurologic symptoms or signs.

PROGNOSIS

Except for the secondary causes of dementia and the rare dementing illnesses caused by single episodes of brain injury (e.g., severe head trauma, anoxic encephalopathy), dementia is a condition that invariably leads to worsening of cognition and function. Almost all dementia patients progress from mild stages to severe dementia during the course of several years if they do not die prematurely. The rate of cognitive decline is variable among individuals and, of course, also varies with the specific disease. In general, dementia can be said to decrease life expectancy by half compared with the life expectancy of nondemented individuals.

End-of-Life Care

The terminal stage and end-of-life care issues (Chapter 3) associated with the common dementias are usually similar. Dementia itself does not directly cause death, but it is strongly linked to reduced survival. Patients with dementia typically die of the same illnesses that affect debilitated individuals, such as sepsis, pneumonia, pulmonary embolism, or heart disease.

Most patients with dementia experience their terminal illnesses in hospitals or extended care facilities. Given the inexorably progressive nature of most dementing illnesses and their likelihood of producing severe and completely disabling cognitive and functional impairment, it is widely accepted that patients with end-stage dementia should receive conservative care. Feeding tubes and ventilatory support should not generally be considered.

MILD COGNITIVE IMPAIRMENT

DEFINITION

Mild cognitive impairment represents the transition between the state of normal cognition and dementia.[4] Patients with mild cognitive impairment have abnormalities in a specific aspect of cognition to such an extent that it is clearly different from normal performance but does not interfere with daily functioning to any appreciable degree. More than one domain of cognition may be affected. The amnesic form of mild cognitive impairment, in which declarative episodic memory is impaired, is the most common. Alterations in attention, concentration, and mental agility may also be seen (Table 402-2). The term *cognitively impaired, not demented* includes patients whose mild cognitive impairment may progress to dementia but also encompasses anyone who is neither cognitively normal nor demented, such as individuals with stable lifelong cognitive impairment.

EPIDEMIOLOGY

The prevalence and incidence of mild cognitive impairment are about the same as those of dementia. Both increase with advancing age.

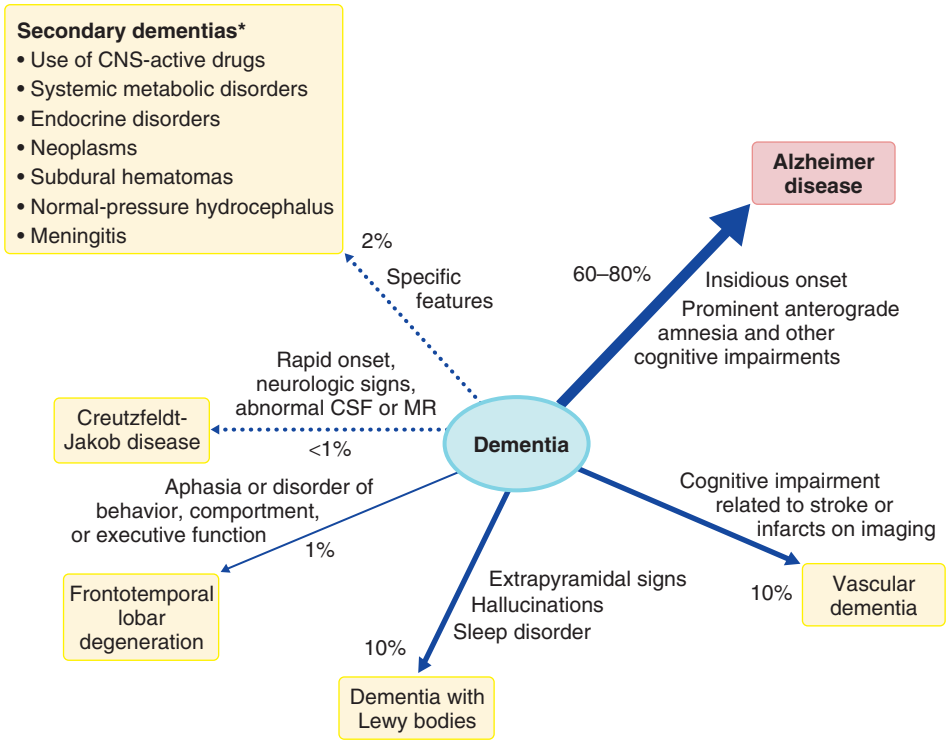

FIGURE 402-2. Flow diagram for the differential diagnosis of dementia. The percentage contributions of various diagnoses are approximate. *The list of secondary causes of dementia is not exhaustive. CNS = central nervous system; CSF = cerebrospinal fluid; MR = magnetic resonance imaging.

TABLE 402-2 DIAGNOSTIC CRITERIA FOR AMNESIC MILD COGNITIVE IMPAIRMENT

The presence of a new memory complaint, preferably corroborated by an informant
Objective evidence of an impairment in episodic declarative memory (for age)
Normal general cognitive functions
No substantial interference with work, usual social activities, or other activities of daily living
No dementia

Adapted from Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011;7:270-279.

PATHOBIOLOGY

Mild cognitive impairment is a risk state for the subsequent development of dementia. Alzheimer disease, followed by cerebrovascular disease and Lewy body disease, is the most common underlying cause.

CLINICAL MANIFESTATIONS

Patients with mild cognitive impairment may have more insight into their emerging cognitive difficulties than do patients with dementia. Hence, some patients with mild cognitive impairment may themselves seek medical consultation because of concern about their memory or their thinking. They or their family members may report a milder extent of many of the symptoms of dementia. Patients with mild cognitive impairment forget recent events and conversations or have trouble with mental flexibility, multitasking, problem solving, or completing mentally challenging activities at the speed they once did. Mental status testing will sometimes corroborate the complaints, but neuropsychometric testing may be needed to document impairment. Other patients with mild cognitive impairment may have virtually no insight into their memory loss, but the impairment is diagnosed after family members force the patient to undergo an evaluation.

DIAGNOSIS

When mild cognitive impairment is suspected, the principal alternative diagnosis is that the person is cognitively intact or that the person has dementia. The diagnosis of normal function should be straightforward when the patient and family have no complaints of cognitive impairment and the patient scores normally on bedside cognitive testing. However, a number of circumstances may cloud the issue, including very low or very high levels of prior educational and occupational achievement, instances in which English (or whatever the dominant language is) was a second language, severe hearing loss or blindness, and major alterations in mood or major motor disabilities that interfere with daily functioning. In such circumstances, the history of cognitive difficulty and the examination of cognitive status may be so confounded by these other phenomena that distinguishing between normal cognitive function and mild cognitive impairment is challenging. Conversely, distinguishing between mild cognitive impairment and dementia may be straightforward when daily functioning is obviously impaired. In other circumstances, it may be difficult to ascertain whether a person is functioning fully independently or not. For example, many older adults reside in assisted living facilities that provide services such as cooking and housekeeping. In such individuals with few daily responsibilities, it is difficult to determine whether they are functionally impaired.

TREATMENT

Rx

As of 2014, no treatments have been approved for mild cognitive impairment, although a randomized trial demonstrated that donepezil therapy, 10 mg/day, significantly reduced the rate of development of dementia secondary to Alzheimer disease at 1 year but not at 3 years. Physical activity appears to provide a modest slowing of cognitive decline.

PROGNOSIS

Mild cognitive impairment should be viewed as a risk state for the subsequent development of dementia. With use of the definition of amnesic mild cognitive impairment in Table 402-2, the rate of evolution from mild cognitive impairment to dementia is 15% per year. Mild (relative risk, 1.2) and moderate (relative risk, 1.4) cognitive impairment are also associated with increased all-cause mortality.

ALZHEIMER DISEASE

DEFINITION

Alzheimer disease, which is a pathophysiologic process involving β-amyloidosis and limbic and isocortical neurodegeneration, produces a

TABLE 402-3 DIAGNOSTIC CRITERIA FOR PROBABLE ALZHEIMER DISEASE DEMENTIA

The clinical diagnosis of probable Alzheimer disease dementia is made when:

Criteria for dementia met (see Table 402-1), and the illness has the following characteristics:
- Insidious onset: symptoms have a gradual onset over months to years; *and*
- Clear-cut history of worsening of cognition by report or observation; *and*
- The initial and most prominent cognitive deficits are evident on history and examination consistent with an amnestic disorder (most common) or a nonamnestic cognitive disorder (less common) (aphasia, visuospatial disorder, or behavioral/dysexecutive disorder).

The diagnosis of probable Alzheimer disease dementia **should not** be applied when there is substantial evidence for another neurodegenerative disease, extensive cerebrovascular disease, or a non-neurologic medical comorbidity or medication use that could have a substantial impact on cognition.

Research criteria for probable Alzheimer disease dementia with higher certainty when imaging or cerebrospinal fluid biomarkers are available:

Probable Alzheimer disease dementia is present based on above clinical criteria, and the following neuroimaging and cerebrospinal fluid (CSF) biomarker profiles are present:
- Highest-probability biomarker profile: β-amyloid marker (CSF or imaging) "positive" *and* neuronal injury marker (CSF tau, FDG-PET or structural MR) "positive"
- Intermediate probability biomarker profile: β-amyloid marker "positive" *and* neuronal injury marker "negative" or indeterminate
- Highest probability: probable Alzheimer disease dementia is present based on above clinical criteria, and a known pathogenic mutation is present in *APP*, *PSEN1*, or *PSEN2*

FDG-PET = ^{18}fluoro-deoxyglucose positron emission tomography; MR = magnetic resonance.
Adapted from McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011;7:263-269.

dementing illness in which anterograde amnesia is a dominant symptom (Table 402-3). The clinical diagnosis implies that the causative pathologic process is of the Alzheimer type, whereas the pathologic diagnosis rests on the findings of characteristic histopathologic features.

EPIDEMIOLOGY

Between 60 and 80% of all dementing illness is due to Alzheimer disease. Among all individuals older than 65 years, the prevalence of Alzheimer disease is estimated to be about 5%. As with dementia in general, the prevalence doubles in every 5-year interval after age 65, and the incidence continues to rise into the 10th and 11th decades of life. Men and women may be equally affected, although on an absolute basis, far more women have prevalent Alzheimer disease because women live longer than men. There are no ethnic or racial differences in the predilection for Alzheimer disease.

Risk Factors

Established risk factors for Alzheimer disease include advancing age and a family history. Putative risk factors include diabetes mellitus, hypertension, cardiovascular disease, and head trauma. Evidence for and against each of these four conditions is inconclusive, but the consensus is that at least diabetes and hypertension may play a role in the pathogenesis of Alzheimer disease. Low educational achievement is also a consistent risk factor, but most experts believe educational level is a proxy for some other factor, such as socioeconomic status or the early childhood medical and psychosocial environment. Protective factors have also been proposed, but their status is much debated.

PATHOBIOLOGY

The histopathologic diagnosis of Alzheimer disease is based on the joint presence of a substantial cerebral burden of neuritic plaques and neurofibrillary tangles.[5] Neuritic plaques consist of a core of aggregated β-amyloid peptide surrounded by degenerating neurites, which are fragments of axons and dendrites. β-Amyloid contains 39 to 42 amino acids and is proteolytically derived from a larger protein, the amyloid precursor protein. Neurofibrillary tangles are intracellular aggregations of an excessively phosphorylated form of the microtubule-associated protein tau. The altered tau protein self-aggregates and forms neurofibrillary tangles. In a low-powered microscopic section of frontal, temporal, or parietal cortex, at least six neuritic plaques and neurofibrillary tangles should be visible for the diagnosis of Alzheimer disease to be made.

Pathophysiology

The progression of changes of β-amyloidosis follows a roughly predictable pattern in Alzheimer disease. Positron emission tomographic (PET) imaging with ligands that bind to β-amyloid shows that β-amyloid begins to accumulate in the neocortex as long as 20 years before dementia occurs. Soluble aggregates of β-amyloid in oligomeric (consisting of a small number of monomers) forms may be the key pathogenic molecules that eventually induce or accelerate neuronal injury. By the time clinical dementia due to Alzheimer disease is present, large numbers of β-amyloid peptide–containing deposits invariably are found in neuritic plaques in the neocortex. Neuritic plaques represent the end stage of the Alzheimer process. Because β-amyloidosis begins well before clinical symptoms appear and probably reaches a plateau in terms of abundance, the amount of β-amyloidosis does not closely mirror the severity of dementia in Alzheimer disease.

The regional extent of neurofibrillary tangles in Alzheimer disease grows as the disease progresses. Neurofibrillary tangles appear in the medial temporal lobe and brain stem in cognitively normal persons by the fourth decade of life. In persons destined to develop Alzheimer disease, a critical part of the pathophysiology involves transsynaptic spread of neurofibrillary tangle pathology to cortical association areas. At the time clinical symptoms develop, neurofibrillary tangles are found in association neocortices of the frontal, parietal, and temporal lobes. It is only in the most severe and final stages that neurofibrillary tangles are found in the occipital lobes and primary motor and sensory cortices. The location of neurofibrillary tangles corresponds faithfully to the clinical evolution of specific symptoms and severity of Alzheimer disease. In mild cognitive impairment, the earliest clinical manifestation of Alzheimer disease, the most intense burden of neurofibrillary tangles is in the entorhinal cortex and hippocampi, precisely the regions involved in declarative episodic memory. Hippocampal atrophy is characteristic, and reductions in hippocampal volumes may be observed on magnetic resonance imaging (MRI; Fig. 402-3). Involvement of the association neocortices with neurofibrillary tangles represents the histopathologic correlate of the progression to dementia. Quantitative MRI in patients with mild cognitive impairment who later progress to dementia shows increasing atrophy of key cortical association areas, such as the lateral temporal lobes, inferior parietal lobes, posterior cingulate cortex, and lateral frontal lobes. Reflecting the spread to association neocortex, language functions, visuospatial functions, and executive cognitive functions typically become impaired some time after declarative episodic memory dysfunction occurs.

The most consistent neurotransmitter deficit in Alzheimer disease is in cholinergic neurotransmission. The cells of origin of hippocampal and neocortical cholinergic projections are located in the septum, diagonal band, and nucleus basalis. Neurofibrillary tangles accumulate in the neurons in these regions as Alzheimer disease develops, but there is also neurochemical evidence that these neurons are stressed much earlier in the disease.

Genetics

The overwhelming majority of Alzheimer disease is due to sporadic (not genetic) disease. However, in a very small number of instances, Alzheimer disease occurs as an autosomal dominant disease. The three known genes involved in autosomal dominant Alzheimer disease all are directly involved in the production of β-amyloid peptide. The first is the amyloid precursor protein (*APP*) gene, located on chromosome 21q21.3. Eighteen known mutations in this gene lead to excess production of β-amyloid and are reliably associated with a very early onset (20 to 50 years of age) of Alzheimer disease. Another line of evidence implicating the *APP* gene in Alzheimer disease is the invariable appearance of the pathologic process of Alzheimer disease in individuals with Down syndrome (trisomy 21 [Chapter 41]), who have an extra copy of the *APP* gene as a result of the trisomy.

The other two genes associated with autosomal dominant Alzheimer disease are the presenilin 1 and 2 genes, located on chromosomes 14q24.3 and 1q31.42. A large number of presenilin 1 mutations account for the majority of autosomal dominant Alzheimer disease. Both genes code for a similar protein known as presenilin. Presenilin is involved in degradation of the APP molecule at the gamma cleavage site. It is believed that the Alzheimer disease–causing mutations in presenilin 1 and 2 lead to a "toxic gain of function" that produces excess β-amyloid peptide. The presenilin mutations are also associated with early-onset (age 40 to 60) Alzheimer disease.

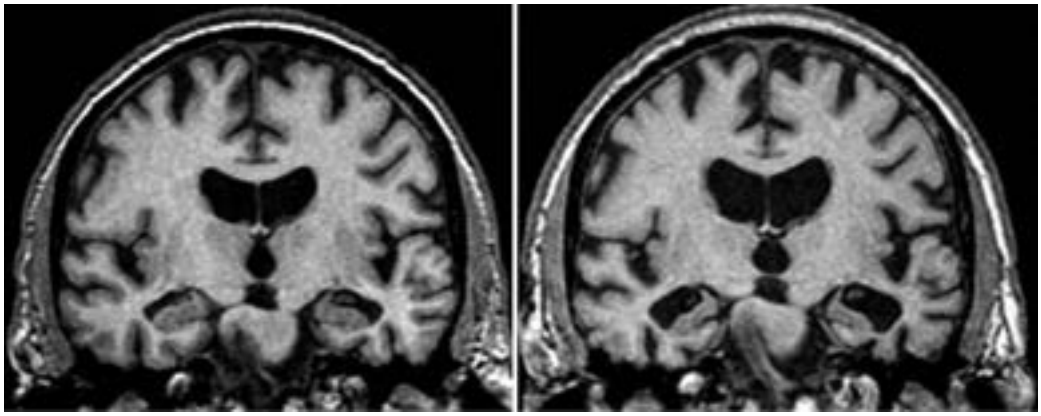

FIGURE 402-3. Serial coronal images from magnetic resonance imaging of a patient with Alzheimer disease. The scan on the left was performed when the patient was clinically normal. The scan on the right was performed 11 years later when the patient was demented. Hippocampal atrophy has increased dramatically from the first to the subsequent scan. (Courtesy Maria Shiung and Clifford Jack.)

Studies of the familial aggregation of Alzheimer disease have shown that later-onset disease also displays genetic risks, but only a few genes have been definitively linked to later-onset Alzheimer disease. The most prominent gene related to later-onset Alzheimer disease, located on chromosome 19q13.2, encodes apolipoprotein E (apo E), a protein involved in lipid transport. In humans, three allelic variants of apolipoprotein gene (*APOE*) are determined by differences in the amino acids cysteine and arginine at positions 112 and 158 of the 299–amino acid protein. One of the allelic variants, with arginine at both positions, designated the ε4 variant, is strongly associated with a 14-fold increased risk for Alzheimer disease in homozygotes and a three-fold increase in heterozygotes. In many series, almost 50% of Alzheimer disease patients but only about 25% of nondemented controls have at least one copy of the *APOE* ε4 allele. The presence of an *APOE* ε4 allele does not always cause Alzheimer disease in that the disease never develops in some carriers of the genotype. The mechanism by which the *APOE* ε4 allele predisposes to Alzheimer disease is not established, but the tertiary structure of the APOE protein with arginine at positions 112 and 158 may lead to impaired binding to β-amyloid, which in turn reduces the clearance of β-amyloid from cells.

A rare missense mutation in the *TREM2* gene also increases the risk of Alzheimer disease.[6] The pathophysiology appears to be impaired containment of inflammatory processes rather than a direct effect on neurologic function.

CLINICAL MANIFESTATIONS

The early course of Alzheimer disease is dominated by difficulties with anterograde amnesia. Some of the usual complaints include forgetting recent events and conversations, misplacing items, problems with keeping track of the date, getting lost in familiar surroundings, and problems with remembering to complete tasks. The frequency and severity of the memory lapses progress from occasional difficulty to more pervasive and consistent failure.

In mild Alzheimer disease, declarative episodic memory function may be lost. Familiarity and access to previous knowledge may allow patients to function in their usual daily routines as long as nothing out of the ordinary is required of them. They may still retain the ability to prepare simple meals and take walks in their neighborhood without getting lost. However, even in mild Alzheimer disease, medication-taking errors and difficulty managing money or balancing a checkbook are likely to occur. Traveling to unfamiliar places often accentuates confusion. Changes in personality commonly accompany the cognitive losses. Apathy, loss of initiative, and loss of interest in previous hobbies and pastimes are ubiquitous in early Alzheimer disease.

As the disease progresses, the ability to perform necessary daily tasks becomes more and more difficult to the point that the patient will need assistance preparing meals, paying bills, taking transportation, and keeping house. As the disease moves into the severe stages, assistance and supervision in basic activities such as bathing, dressing, toileting, and eating become necessary.

In the terminal stages of the disease, all communicative abilities may be lost. Mobility may still be preserved until late in the disease. Alzheimer disease patients commonly die of illnesses that strike other debilitated elderly individuals, such as sepsis, pneumonia, and congestive heart failure.

The duration of the course of clinical Alzheimer disease is long but variable. The time from mild dementia to death may be as short as 2 to 3 years or may be well over a decade. For patients in whom mild dementia is diagnosed, about 10% per year reach the stage of severe dementia.

Rarely, Alzheimer disease is associated with prominent symptoms in cognitive domains other than memory. The most common of the atypical syndromes is one in which profound visuospatial deficits occur without the typical severe anterograde amnesia. This syndrome is referred to as posterior cortical atrophy.

DIAGNOSIS

The diagnosis of Alzheimer disease, like that of dementia itself, is largely a clinical one based on the history and examination. The key elements in the history are a gradual onset and insidious progression of cognitive impairment, especially anterograde amnesia. The mental status examination should demonstrate impairment in short-term memory and other cognitive deficits. Alzheimer disease should be thought of as a diagnosis of inclusion: if the history and examination are compatible with Alzheimer disease and if certain exclusions can be verified, the diagnosis can be made with confidence.

The pathophysiology of Alzheimer disease in patients with mild cognitive impairment or dementia can be assessed with cerebrospinal fluid (CSF) protein markers (β-amyloid and tau) (see Table 402-3)[7] and with brain imaging (structural MRI, ^{18}fluoro-deoxyglucose PET,[8] and amyloid positron emission tomography[9]) (Figs. 402-4 and 402-5). PET imaging of abnormal deposition of tau protein, now available for research purposes, may enhance the accurate diagnosis of Alzheimer disease.[10] Efforts are now underway to diagnose Alzheimer disease while individuals are still asymptomatic.

Differential Diagnosis

A number of other conditions that bear similarity to Alzheimer disease must be excluded on clinical or laboratory grounds (see Fig. 402-2). One is dementia with Lewy bodies, which is suggested by the presence of parkinsonism, prominent visual hallucinations, and a specific sleep disorder. At autopsy, the pathologic processes of Lewy body disease and Alzheimer disease often coexist, thus suggesting that the diagnoses overlap. Frontotemporal lobar degeneration is suggested by prominent behavioral and personality changes or by prominent language difficulties early in the course. Hippocampal sclerosis has unique neuropathologic findings but is virtually impossible to distinguish from Alzheimer disease by clinical features. Other neurodegenerative conditions in the differential diagnosis of Alzheimer disease include Huntington disease (Chapter 410), progressive supranuclear palsy (Chapter 410), corticobasal degeneration (Chapter 410), amyotrophic lateral sclerosis (Chapter 419), and Wilson disease (Chapter 211); however, these diseases invariably have prominent motor manifestations early in their course. Normal-pressure hydrocephalus (Chapter 189; see later) is a rare cause of dementia associated with a gait disorder.

It is particularly challenging to distinguish dementia caused by cerebrovascular disease from Alzheimer disease (see later). The fact that Alzheimer disease and cerebrovascular disease often coexist requires clinicians to consider both simultaneously.

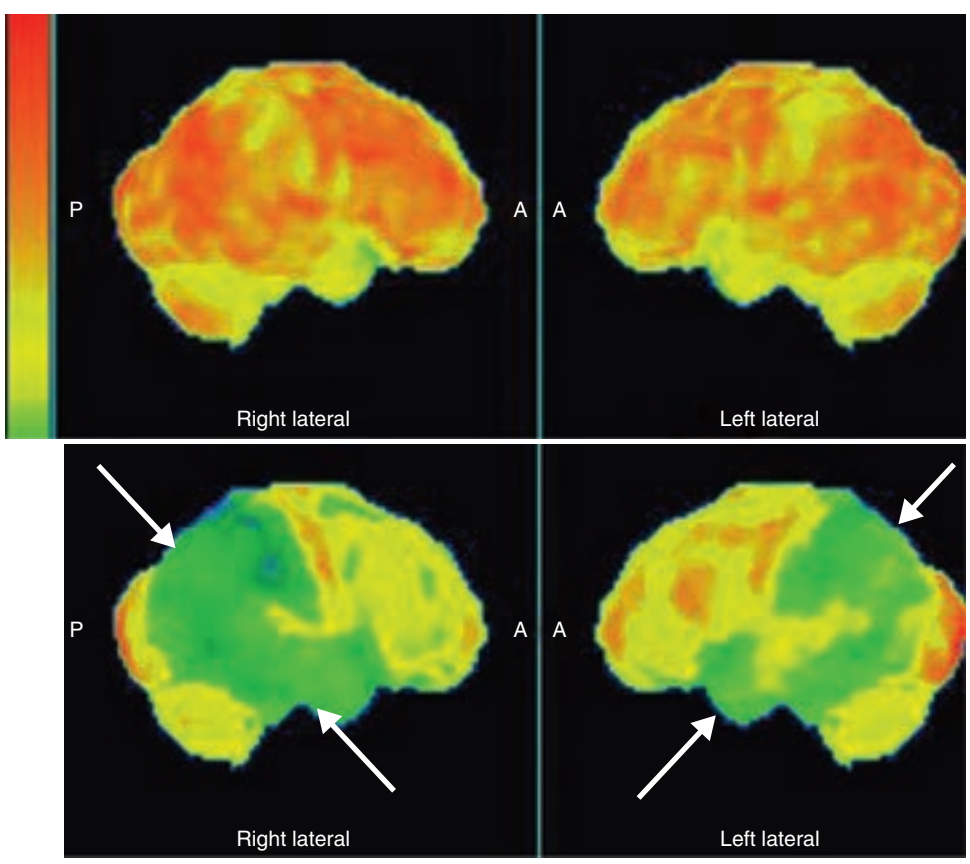

FIGURE 402-4. **^{18}Fluoro-deoxyglucose positron emission tomographic scan of a patient with Alzheimer disease dementia.** Computerized reconstructions of the regional glucose uptake ratio (using pons as reference) of the cortical surface show hotter colors (*yellow and orange*) in areas of normal glucose uptake, whereas cooler colors (*green and blue*) indicate hypometabolism. The scan on the top is from a normal individual of the same age. The scan on the bottom is from a patient with typical Alzheimer disease dementia, and it shows hypometabolism in the temporal and parietal cortical regions (*arrows*).

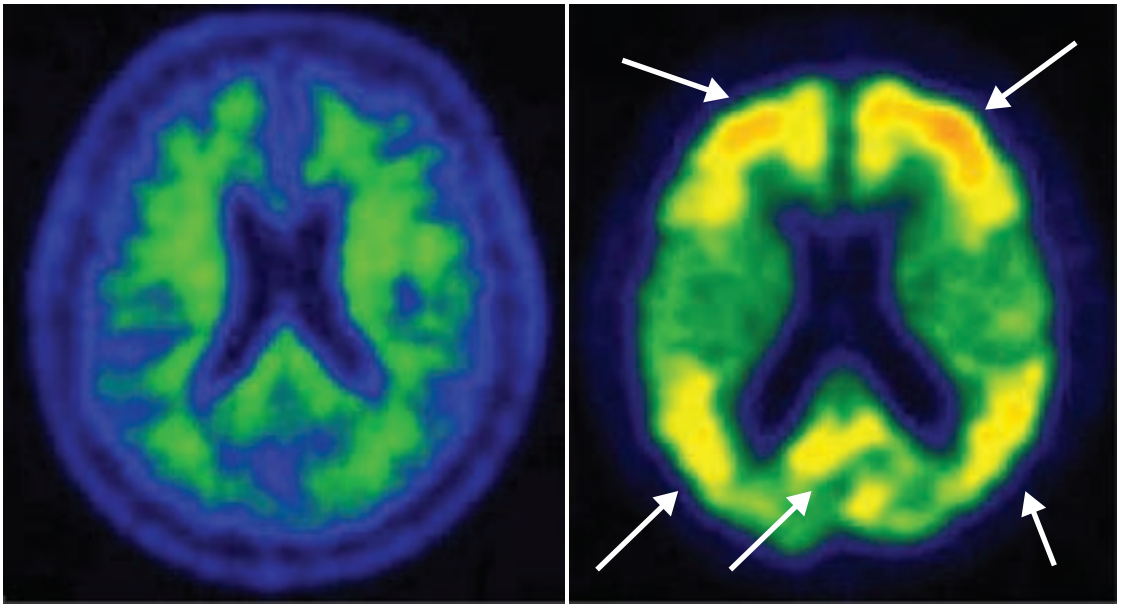

FIGURE 402-5. **Amyloid positron emission tomographic (PET) scan of a patient with Alzheimer disease dementia.** On this axial image of a ^{11}C Pittsburgh compound B PET scan, the left scan is from an individual with no cortical amyloid retention. The green signal represents low levels of nonspecific white matter binding. On the right, a scan of a patient with Alzheimer disease dementia shows the prominent retention of the amyloid imaging agent in the frontal, parietal, and posterior cingulate cortices (*arrows*).

PREVENTION AND TREATMENT

Rx

There are no established preventive therapies. Although a healthy diet, physical exercise, and stimulating cognitive leisure activities are sensible, they have not been shown to protect against Alzheimer disease. Except in patients who have folate or vitamin B_{12} deficiency, vitamin B supplementation is not effective in slowing cognitive decline.■ Treatment of diabetes (Chapter 229) and hypertension (Chapter 67) is beneficial for other reasons, but it is not clear that such treatment alters the course of Alzheimer disease. Antidepressants generally are ineffective and increase adverse events when used to treat patients with Alzheimer disease.■

Once Alzheimer disease becomes symptomatic, support for family caregivers is a critical intervention that cannot be overemphasized. Support groups

through the Alzheimer Association (available at www.alz.org) can benefit families coping with the disease.

Important safety issues include supervision of medications, supervision of finances, and close scrutiny of motor vehicle operation. Operation of other potentially dangerous tools, firearms, appliances, and equipment should also be carefully monitored or avoided. Patients with Alzheimer disease often wander and can become lost long distances from home. Identification of patients can prevent tragic occurrences.

Evidence-Based Treatments

Two classes of drugs are approved for the treatment of Alzheimer disease: cholinesterase inhibitors and memantine, a glutamate receptor antagonist.[11] The rationale for use of cholinomimetic drugs (donepezil, 5 or 10 mg/day; galantamine, 16 or 24 mg/day; or rivastigmine, (6-12 mg/day orally or 4.5-9 mg/day by skin patch) is the reduced levels of cholinergic markers in the neocortex of patients dying of Alzheimer disease. All three agents delay the progression of symptoms to a statistically significant but clinically marginal extent at 6 to 12 months in patients with mild to moderate Alzheimer disease. In community-dwelling patients who had been treated with donepezil for at least 3 months and who had moderate or severe Alzheimer disease, those who continued donepezil at 10 mg daily had clinically important functional benefits over the next 12 months compared with those who discontinued it. There is no compelling evidence that these agents alter its biological progression. Individual patients often do not show any clear benefits of treatment. Memantine, which is a low- to moderate-affinity uncompetitive *N*-methyl-D-aspartate receptor antagonist that acts on glutamate neurotransmission, appears to delay the progression of functional decline in patients with moderate to severe Alzheimer disease at a dose of 10 mg twice daily.

One study of patients with moderately severe Alzheimer disease showed that vitamin E was effective in delaying progression, and recent studies in patients with mild to moderate Alzheimer disease also suggest a benefit. Multivitamins are not efficacious, nor are humanized monoclonal antibodies that bind soluble forms of amyloid or inhibit the formation of amyloid plaques.

PROGNOSIS

Alzheimer disease is inevitably progressive, and severe cognitive impairment and complete dependence on others develop in virtually all patients unless they die prematurely. Alzheimer disease also contributes to premature death; the mortality rate in patients with Alzheimer disease is about 10% per year. In patients with advanced dementia, the 6-month mortality is about 55%; pneumonia, fever, and eating problems are associated with poor prognosis.

VASCULAR DEMENTIA

DEFINITION

Vascular dementia is a dementing illness in which the underlying cause is cerebral infarction. For a cognitive disorder to be attributed to cerebrovascular disease from a neuropathologic perspective, there must be sufficient cerebral infarction in locations known to be responsible for the cognitive deficits in the absence of other neurodegenerative neuropathologic changes (Table 402-4). When cerebrovascular disease produces cognitive impairment that is not severe enough to meet the criteria for dementia, it is referred to as vascular cognitive impairment.

TABLE 402-4 DIAGNOSTIC CRITERIA FOR THE SYNDROME OF DEMENTIA CAUSED BY CEREBROVASCULAR DISEASE (VASCULAR DEMENTIA)

Dementia as defined in Table 402-1
Clinically important cerebrovascular disease is demonstrable by *either* of the following:
Onset of the cognitive disturbance or dramatic worsening of an existing disturbance that occurred within 3 months of a stroke, where stroke is defined as a focal neurologic deficit of acute onset in which the symptoms and signs persist for more than 24 hours
Neuroimaging evidence of bilateral brain infarctions rostral to but including the thalamus

EPIDEMIOLOGY

In clinical studies, as many as 20% of dementia patients have cerebrovascular disease. Like Alzheimer disease, it is less common in patients younger than 65 years and increases steadily thereafter. In neuropathologic studies, about 25% of all cases of dementia have some vascular component. Roughly half that number are relatively pure vascular dementia; the remainder consists of vascular disease mixed with Alzheimer disease. Men and women are equally affected.

Risk Factors

Risk factors for vascular dementia include cardiovascular disease, atrial fibrillation,[12] higher glucose levels,[13] diabetes, and hypertension. There are no known protective factors other than treatment of these risk factors. Populations with high rates of generalized vascular disease should have higher rates of vascular dementia, but competing mortality from cardiovascular disease may obscure part of the relationship. Microinfarcts contribute to brain atrophy and cognitive impairment, particularly before dementia is clinically evident. In the first year after a stroke, the risk for development of dementia is about nine-fold higher than the rate in persons without a stroke; the risk remains about two-fold higher in subsequent years.

PATHOBIOLOGY

The majority of vascular disease causing cognitive impairment is due to atherosclerosis. One mechanism is through large infarctions, such as those secondary to occlusive disease in major cerebral vessels, including the carotid arteries and the anterior, middle, and posterior cerebral arteries (Chapter 407). A second mechanism of infarction is at the arteriolar level, with lacunar infarctions in the thalamus, basal ganglia, and subcortical white matter. Both these processes can be detected by brain MRI. Infarcts in the hippocampal formations, medial thalamus, caudate nuclei, and parietal association areas are highly likely to produce cognitive impairment but not necessarily dementia. Microinfarcts, which are small zones of infarction that are not visible to the naked eye but can be observed with light microscopy, may also contribute to the dementia. The simultaneous presence of Alzheimer disease is common in vascular dementia.

There are other uncommon causes of vascular dementia. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) is a very rare inherited disease that usually becomes clinically evident between the ages of 30 and 50 years and causes severe white matter disease, headaches, and dementia. The cause of CADASIL is mutations in the *notch3* gene on chromosome 19q12. Cerebral amyloid angiopathy, a β-amyloidosis in which the β-amyloid peptide accumulates in the media of small to medium-sized arteries in the leptomeninges and superficial cortex, causes cerebral hemorrhages that may lead to dementia if it occurs in sufficient number and in critical locations. Cerebral amyloid angiopathy is also seen in Alzheimer disease, but its hemorrhagic manifestations may occur in individuals with little evidence of Alzheimer disease clinically and modest evidence pathologically. Cerebral vasculitis (Chapter 270) is a very rare cause of dementia.

CLINICAL MANIFESTATIONS

The spectrum of cognitive changes in patients with cerebrovascular disease is broad. The more common cognitive syndromes in cerebrovascular disease include mild cognitive impairment, a dementia with prominent anterograde amnesia, and a dementia with prominent changes in personality and executive function. Some patients with vascular cognitive impairment without dementia may have deficits in only one domain (Chapter 401). A number of aphasia syndromes are a result of cerebral infarction or hemorrhage in the perisylvian regions of the dominant hemisphere. Infarction or hemorrhage in the occipitotemporal or occipitoparietal regions may produce one of the disorders of visual cognition, such as alexia or visual agnosia. Infarcts in the caudate nuclei, particularly if they are bilateral, may produce a cognitive syndrome that includes both amnesia and disordered executive function, thus mimicking dementia. Large infarcts in the right parietal lobe can also produce dementia. Infarcts in the medial thalami or in the hippocampal formations can produce isolated amnesia.

The evolution of symptoms in vascular dementia does not follow a stereotypical pattern. In some, the dementia syndrome may remain static. In others, new strokes may lead to substantial declines in cognition and function. Some patients with vascular dementia may experience a gradually declining illness. Patients with vascular cognitive impairment without dementia or vascular

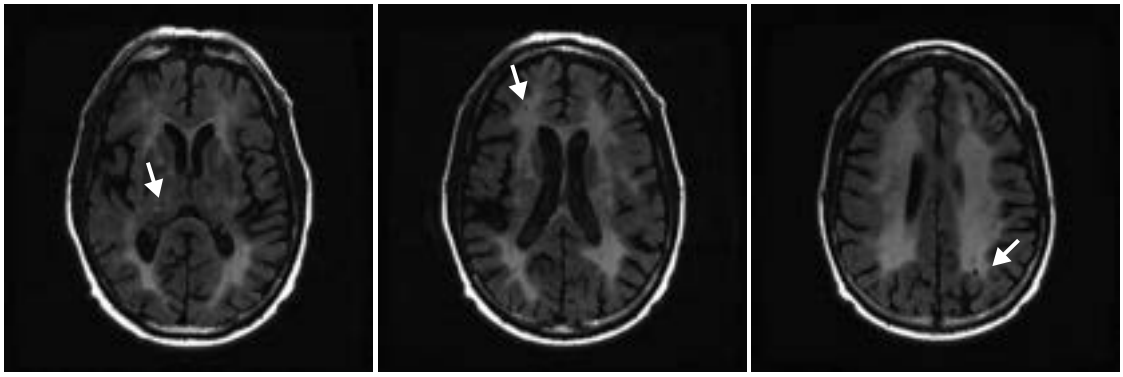

FIGURE 402-6. Axial magnetic resonance images of a patient with extensive cerebrovascular disease. The images show extensive white matter hyperintensities bilaterally. There are also lacunar infarcts (*arrows*).

dementia may also have other neurologic signs typical of patients with cerebrovascular disease, such as hemiparesis, hemianopia, hemisensory changes, or cranial nerve abnormalities.

DIAGNOSIS

The diagnosis of vascular dementia is based on the neurologic history and examination. Brain imaging, preferably with MRI, is essential to establish the presence of infarcts. The cardinal diagnostic features of vascular dementia are that (1) the cognitive disorder should have begun within 3 months of a clinical stroke event and (2) there should be multiple bilateral infarcts in the cerebral hemispheres visible on brain imaging studies (Fig. 402-6). A temporal link between the onset or worsening of cognitive impairment and a stroke is important in demonstrating that cerebrovascular disease is etiologically relevant to the cognitive impairment. Brain imaging of infarcts in the cerebral cortex, basal ganglia, thalamus, and cerebral white matter has obvious value for establishment of cerebrovascular disease. In contrast to actual infarcts on imaging, the presence of white matter hyperintensities without infarcts on brain MRI is much less specific.

The accuracy of the clinical diagnosis of vascular dementia is generally lower than that of Alzheimer disease. The combination of (1) a temporal relationship between dementia and a stroke and (2) imaging evidence of bilateral infarcts is diagnostically specific for vascular dementia but is insensitive. Broader diagnostic criteria (see Table 402-4) are more sensitive but less specific. The usual alternative diagnosis is Alzheimer disease, and there is typically no way to be certain whether and how much Alzheimer disease is simultaneously present.

PREVENTION AND TREATMENT

Rx

Some cases of vascular dementia should be preventable. With early lifelong aggressive treatment of diabetes (Chapter 229), hypertension (Chapter 67), and hyperlipidemia (Chapter 206), the number of cerebral infarcts should be reduced, with a corresponding reduction in the number of cases of vascular dementia. Evidence for this link comes from large-scale studies in which the treatment of hypertension reduced the frequency of strokes and incident dementia. Once vascular dementia develops, cholinesterase inhibitors have shown some benefit,[1] but the major goal is to prevent future strokes.

PROGNOSIS

Patients with vascular dementia can often be expected to have severe cardiovascular disease and a greater likelihood of future strokes and cardiac ischemic events. Survival of patients with vascular dementia is poorer than that of patients with Alzheimer disease.

DEMENTIA WITH LEWY BODIES

DEFINITION

Dementia with Lewy bodies is a multifaceted dementing disorder in which the underlying pathologic process includes Lewy bodies in limbic and cortical structures (Table 402-5). Some clinicians make a distinction between patients in whom parkinsonism preceded the cognitive disorder and those in whom the cognitive disorder occurred either simultaneously with or before the movement disorder. This distinction may be somewhat useful in clinical practice, but there are few clinical or neuropathologic differences based on different sequences of signs and symptoms. The diagnosis of dementia with Lewy bodies is similar in principle to diagnosis of both dementia and Parkinson disease (Chapter 409) in the same individual, but *dementia with Lewy bodies* is a term with broader connotations because of other features (hallucinations, fluctuations, and sleep disorder) that may be more apparent than the movement disorder.

TABLE 402-5 DIAGNOSTIC CRITERIA FOR THE DEMENTIA SYNDROME ASSOCIATED WITH LEWY BODY PATHOLOGY

Dementia as defined in Table 402-1
The cognitive disturbance is of insidious onset and is progressive, based on evidence from the history or serial cognitive examination.
The presence of **at least two** of the following: Parkinsonism (rigidity, resting tremor, bradykinesia, postural instability, parkinsonian gait disorder) Prominent, fully formed visual hallucinations Substantial fluctuations in alertness or cognition Rapid eye movement sleep behavior disorder (Chapter 405) Severe worsening of parkinsonism by antipsychotic drugs
The disturbance is not better accounted for by a systemic disease or another brain disease.

Adapted from McKeith IG, Dickson DW, Lowe J, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology.* 2005;65:1863-1872.

EPIDEMIOLOGY

Dementia with Lewy bodies is about a quarter as common as Alzheimer disease.[14] Lewy body disease becomes more common with advancing age, and the prevalence of dementia with Lewy bodies increases with advancing age as well. As with the other dementias, there are no known ethnic or racial differences, but dementia with Lewy bodies may be more common in men. There are no known risk factors for dementia with Lewy bodies. Dementia develops in up to 30% of patients with Parkinson disease, and advancing age is the major risk factor.

PATHOBIOLOGY

The pathology of dementia with Lewy bodies is a mixture of Lewy body disease and Alzheimer disease. In general, the more intense the Lewy body disease, the less abundant the Alzheimer disease. Lewy bodies, which are intraneuronal inclusions that contain α-synuclein, are found in the nucleus basalis, pars compacta of the substantia nigra, locus caeruleus, other brain stem structures, amygdala, cingulate gyrus, and neocortex. The earliest locations of Lewy bodies are the brain stem, where they affect nuclei involved in sleep and arousal, and the substantia nigra, the locus caeruleus, and cranial nerve nuclei IX and X. Typically, the nucleus basalis, transentorhinal cortex, cingulate gyrus, and neocortex become involved later.

In Lewy body disease, the α-synuclein protein becomes misfolded and aggregates intraneuronally. Mutations in the α-synuclein gene have been seen

in a few families with autosomal dominant Parkinson disease, but most cases of dementia with Lewy bodies are sporadic.

CLINICAL MANIFESTATIONS

The clinical manifestations of dementia with Lewy bodies include four major abnormalities: the cognitive disorder, the neuropsychiatric disorder, the motor disorder, and the disorder of sleep and wakefulness. The cognitive disorder may differ from Alzheimer disease, although there is considerable overlap.[15] In a typical patient with dementia with Lewy bodies, visuospatial deficits, impaired concentration, and impaired attention dominate the picture. In some patients, the deficits in executive functions may be similar to what is seen in frontotemporal lobar degeneration. Anterograde amnesia is usually present but milder than in Alzheimer disease. Language deficits are not prominent. The neuropsychiatric manifestations of dementia with Lewy bodies, including prominent apathy, loss of initiative, and depression, may be more disabling than the cognitive symptoms. The motor manifestations include bradykinesia, gait disturbances, postural disturbances, and rigidity. Rest tremor is less common in dementia with Lewy bodies in patients in whom the cognitive disorder appears before the parkinsonism. Visual hallucinations, fluctuations in alertness, and rapid eye movement (REM) sleep disorders are part of a broader disorder of the regulation of sleep and wakefulness. Visual hallucinations are often graphic, detailed, and bizarre, perhaps because the sleep phenomenon of dreaming intrudes into wakefulness. Patients with dementia with Lewy bodies have large fluctuations in their alertness and arousal from day to day.

REM sleep behavior disorder (Chapter 405) is a parasomnia in which patients exhibit dream enactment behavior, often with violent, threatening overtones. Patients typically relate that they feel as though they are being chased by something or someone. Their behavior while they are asleep consists of excessive talking, calling out or shouting, and thrashing about, often to the point of striking a bed partner or falling out of bed. The REM sleep behavior disorder may precede the development of Parkinson disease and dementia with Lewy bodies by years.

DIAGNOSIS

The diagnosis of dementia with Lewy bodies is based on clinical information that corroborates the presence of abnormalities in cognition, motor function, neuropsychiatric behavior, and regulation of sleep and wakefulness. Formal neuropsychological testing is often helpful in evaluating memory, executive function, and visuospatial function in a detailed manner. Neuroimaging has only a limited role in the diagnosis of dementia with Lewy bodies.

Differential Diagnosis

Other disorders that must be considered in patients with dementia and a movement disorder include progressive supranuclear palsy (Chapter 410), which can resemble dementia with Lewy bodies in terms of both the dementia and the motor disorder. In progressive supranuclear palsy, patients are much less likely to have disorders of arousal and typically have other distinctive signs and symptoms, including the characteristic supranuclear gaze palsy and other brain stem findings. The corticobasal degenerations, which are members of the family of frontotemporal lobar dementias (see later), may also produce a movement disorder and dementia. Huntington disease (Chapter 410) is associated with dementia and a movement disorder, but the movement disorder of Huntington disease includes prominent chorea and athetosis, neither of which is present in dementia with Lewy bodies.

Normal-pressure hydrocephalus (Chapter 189) is a rare disorder typically characterized by the triad of a gait disorder, dementia, and urinary incontinence. Altered dynamics of CSF flow in the ventricular system appear to reduce periventricular metabolism and also induce damage in periventricular axons.[16] Normal-pressure hydrocephalus can be suspected when computed tomography or MRI shows ventricular enlargement that is out of proportion to the amount of sulcal widening. Predicting a favorable response to ventriculoperitoneal shunting in suspected normal-pressure hydrocephalus has proved to be difficult. Imaging studies that measure CSF flow through the aqueduct of Sylvius or that measure flow of radiolabeled CSF with radionuclide cisternography have not been useful. Clinical response to the removal of a high volume (e.g., 30 mL) of CSF through lumbar puncture is sometimes used to select patients for surgery, although its positive and negative predictive value is unclear. Normal-pressure hydrocephalus is very rare relative to dementia with Lewy bodies and Alzheimer disease.

TREATMENT

Rx

Management of patients with dementia with Lewy bodies is challenging because of the simultaneous appearance of a cognitive disorder, a neuropsychiatric disorder, a motor disorder, and a sleep and wakefulness disorder. Treatment of the motor disorder is accomplished with antiparkinsonian drugs such as levodopa and dopaminergic agonists (Chapter 409). Treatment with these agents should be instituted for dementia with Lewy bodies if there are prominent gait or balance problems that threaten safety and interfere with independence. These medications may worsen hallucinations and exacerbate confusional states, but this concern should not preclude a treatment trial if the motor symptoms pose safety risks or interfere with independence.

Cholinesterase inhibitors, which do not exacerbate parkinsonian symptoms, have a beneficial effect on neuropsychiatric symptoms and perhaps on the cognitive disorder. Autonomic disturbances such as urinary incontinence can be challenging to treat in persons with dementia with Lewy bodies, because the usually prescribed medications have anticholinergic pharmacologic profiles. Anticholinergic drugs have a definite risk of increasing confusion.

Hallucinations and agitation impair quality of life for the patient and family and often require treatment. Some antipsychotic agents that might otherwise control these symptoms dramatically exacerbate the parkinsonism in dementia with Lewy bodies. Atypical antipsychotics are usually recommended, but there is insufficient experience from controlled clinical trials. Many movement disorder specialists prefer to use quetiapine in doses of 25 to 200 mg/day or clozapine at 6.25 to 50 mg/day because these agents appear to have the lowest rate of extrapyramidal side effects. However, it is not possible to make any strong statements about the relative efficacy of atypical antipsychotics in treating the hallucinations in dementia with Lewy bodies, especially in view of the possibility that atypical antipsychotic agents may be associated with higher-than-expected mortality.

REM sleep behavior disorder (Chapter 405) can be disabling, but there are no controlled clinical trials to inform treatment. Some sleep disorder specialists typically use either melatonin, 3 to 12 mg, or clonazepam, 0.5 to 2 mg, at bedtime.

Treatment of depressive symptoms may substantially improve a patient's functioning. Use of one of the newer-generation antidepressants, such as sertraline (25 to 100 mg/day) or citalopram (10 to 20 mg/day), may be beneficial and does not necessarily interfere with management of the other symptoms (Chapter 397).

PROGNOSIS

As opposed to patients with Alzheimer disease, some studies show that patients with dementia with Lewy bodies have a more rapidly progressive course and poorer survival. As a result of the combination of manifestations, patients with dementia with Lewy bodies may become disabled sooner in their course.

FRONTOTEMPORAL LOBAR DEGENERATION

DEFINITION

The frontotemporal lobar degenerations are a group of neurodegenerative disorders with distinctive clinical manifestations and a predilection for the prefrontal and anterior temporal neocortices. The most common clinical syndrome is a disorder of behavior and personal relationships (comportment) with a loss of executive functions (Table 402-6). This syndrome is referred to as behavior-variant frontotemporal dementia. Other syndromes in the clinical spectrum of frontotemporal lobar degeneration involve different aspects of language or motor dysfunction of the limbs.

EPIDEMIOLOGY

Unlike Alzheimer disease, the frontotemporal lobar degenerations have a peak age at onset in the 50- to 70-year range, and the incidence declines after the age of 70. In patients with dementia who are younger than 70 years, frontotemporal lobar degeneration makes up 10 to 20% of cases.[17] However, across the entire age spectrum, the frontotemporal lobar degenerations are much less common than Alzheimer disease, dementia with Lewy bodies, or vascular dementia. Both men and women are affected equally. There are no known risk factors for the frontotemporal lobar degenerations except a family history.

PATHOBIOLOGY

The clinical syndrome in frontotemporal lobar degeneration is determined by the lobar location of the pathologic process. Right prefrontal or anterior

TABLE 402-6 DIAGNOSTIC CRITERIA FOR BEHAVIOR-VARIANT FRONTOTEMPORAL DEMENTIA

The following symptom must be present to meet criteria for behavior-variant frontotemporal dementia:
- Shows progressive deterioration of behavior and/or cognition by observation or history (as provided by a knowledgeable informant)

Three of the following behavioral/cognitive symptoms that are persistent or recurrent must be present **within 3 years of onset** to meet criteria for **possible** behavior-variant frontotemporal dementia:
- Early behavioral disinhibition such as socially inappropriate behavior, loss of manners or decorum, or impulsive, rash, or careless actions
- Early apathy or inertia
- Early loss of sympathy or empathy
- Early perseverative, stereotyped, or compulsive/ritualistic behavior
- Hyperorality and dietary changes such as altered food preferences, binge eating, increased consumption of alcohol or cigarettes, or oral exploration or consumption of inedible objects
- Neuropsychological profile exhibits executive/generation deficits with relative sparing of memory and visuospatial functions

Probable behavior-variant frontotemporal dementia is diagnosed when **all of the following** are present:
- Criteria met for possible behavior-variant frontotemporal dementia
- Significant functional decline present by caregiver report
- Imaging results that demonstrate frontal and/or anterior temporal atrophy on MRI or CT, or frontal hypoperfusion or hypometabolism on PET or SPECT

The diagnosis of behavior-variant frontotemporal dementia **should not** be applied when the pattern of deficits is better explained by a psychiatric diagnosis, other non-degenerative nervous system disorders, or medical disorders.

CT = computed tomography; MRI = magnetic resonance imaging; PET = positron emission tomography; SPECT = single-photon emission computed tomography.
Adapted from Rascovsky K, Hodges JR, Knopman D, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain.* 2011;134:2456-2477.

temporal disease and brain atrophy produce behavioral syndromes like frontotemporal dementia. Left frontal involvement tends to produce progressive nonfluent aphasia. Predominant left anterior temporal lobe involvement may produce semantic dementia.

On histopathologic grounds, patients with frontotemporal lobar degeneration can be divided into three groups: those whose inclusions contain the microtubule-associated protein tau, those whose inclusions contain the TAR DNA-binding protein 43 (TDP-43), and those whose inclusions contain the fused in sarcoma (FUS) protein, another ribonucleic acid–binding protein. The latter is much less common than the first two. Each type includes both genetically determined and sporadic forms.

Among the tau-positive varieties are Pick disease, in which intracellular tau-positive inclusions known as Pick bodies are seen. Several other pathologic tau-positive subtypes occur, including progressive supranuclear palsy, corticobasal degeneration, and the disorder associated with mutations in the tau gene. Nearly 50 mutations in the *MAPT* gene on chromosome 17q21 are associated with autosomal dominant frontotemporal lobar degeneration syndromes, each with a slightly different clinical and neuropathologic phenotype.[18] The most common is a proline-to-leucine mutation at codon 301, located in exon 10. The tau gene undergoes alternative splicing, resulting in six isoforms of the tau protein. Pathologic mutations appear to disrupt splicing of alternative isoforms of tau protein, which in turn adversely affects the binding of tau to microtubules in neurons. Reduced binding of tau to microtubules is deleterious to microtubule function and neuronal integrity.

The TDP-43–positive frontotemporal lobar degenerations are almost equally common. Immunostaining shows that there are distinctive TDP-43–containing inclusions. Mutations in the granulin (*GRN*) gene, also located on chromosome 17q21, cause familial autosomal dominant forms of frontotemporal lobar degeneration with TDP-43–positive inclusions. Nearly 70 different mutations in the granulin gene are linked to frontotemporal lobar degenerations. All of the mutations lead to premature degradation of the messenger RNA, a process termed *haploinsufficiency*. Granulin mutation carriers have an abnormally low amount of the protein progranulin. The normal function of granulin in the brain is unclear, and the pathophysiologic basis for dementia in persons with granulin gene mutations is unknown. The link between alterations in TDP-43 and *GRN* mutations is also unknown at this time.

A third important gene mutation involved in frontotemporal lobar degenerations associated with TDP-43 inclusions is the hexanucleotide repeat expansion in *C9ORF7* gene located on chromosome 9p21. This latter mutation is the most common of the mutations causing frontotemporal lobar degeneration.[19]

Frontotemporal lobar degenerations that are FUS positive are much less common. At this time, all FUS-positive cases have had the behavioral variant of frontotemporal dementia.

CLINICAL MANIFESTATIONS

The clinical manifestations of the syndrome of frontotemporal dementia begin insidiously. Apathy, loss of initiative, and flattening of affect are common early symptoms. As the disease progresses, the entire spectrum of behavioral changes associated with dysfunction of the frontal and anterior temporal lobes appears.[20] On cognitive assessments, patients may have preserved memory functions, but they typically have difficulty with tests of executive cognitive function. When frontotemporal dementia progresses to moderate or severe stages, the behavioral changes remain prominent, but the disease becomes more difficult to distinguish from other dementias such as Alzheimer disease. The neuropathology of behavior-variant frontotemporal dementia may be either tau positive or TDP-43 positive.

In some patients with frontotemporal lobar degeneration, signs and symptoms of motor neuron disease (Chapter 419) develop, such as weakness, atrophy, and fasciculation in the limbs or the bulbar musculature. In other patients with frontotemporal lobar degeneration, asymmetrical limb apraxia develops that is part of the corticobasal syndrome. Features of progressive supranuclear palsy may also appear in patients with behavior-variant frontotemporal dementia.

Aphasic disturbances are often the presenting manifestation of patients with frontotemporal lobar degeneration. The two most characteristic syndromes are a nonfluent/agrammatic variant of primary progressive aphasia or the semantic variant of primary progressive aphasia. The nonfluent/agrammatic primary progressive aphasia variant is seen in patients who exhibit hesitancy in selecting words in their speech, a problem that may be difficult for others to appreciate at first. Anomia is an early sign. Gradually, the patient's speech becomes laconic and labored. Eventually, a nonfluent, apractic, agrammatic speech develops. In other cognitive domains, patients often have no deficits. Other nonfluent/agrammatic primary progressive aphasic patients may eventually become virtually mute even though they may appear to have preserved memory and visuospatial functions. Patients with nonfluent/agrammatic primary progressive aphasia often have tau-positive neuropathologic findings.

The semantic variant of primary progressive aphasia, previously known as semantic dementia, is a disorder that involves dissolution of the meaning of words or objects. A patient with semantic variant primary progressive aphasia may also become unable to access knowledge about objects (object agnosia) and people's faces (prosopagnosia). The most striking demonstration of the deficit in semantic variant primary progressive aphasia is when a patient can produce the name of an object—a watch, for example—but then cannot say what a watch is for when asked. Often, patients with semantic variant primary progressive aphasia have preservation of the ability to learn a list of words, even if their knowledge of the meaning of the words is diminished. Patients with semantic variant primary progressive aphasia usually have TDP-43–positive neuropathologic findings.

Not all patients with primary progressive aphasia fit neatly into a well-delineated syndrome. Although the semantic and nonfluent/agrammatic variants of primary progressive aphasia are almost always due to frontotemporal lobar degenerations, other variants, especially one in which word-finding problems predominate (the logopenic variant of primary progressive aphasia), may be due to Alzheimer disease.

DIAGNOSIS

Frontotemporal lobar degeneration must first be suspected on clinical grounds, based on the appearance of one of the distinctive clinical syndromes such as frontotemporal dementia (see Table 402-6) or one of the aphasic subtypes.[21,22] Neuropsychological testing can also aid in the diagnosis by detecting abnormalities in executive function and verifying that memory function is preserved, as it often is. For all frontotemporal lobar degeneration syndromes, MRI showing focal atrophy of the frontal (Fig. 402-7) or temporal lobes (Fig. 402-8) is highly likely to be diagnostic. Imaging with fluorodeoxyglucose-enhanced PET can also be useful when the clinical diagnosis is uncertain and MRI is nondiagnostic.

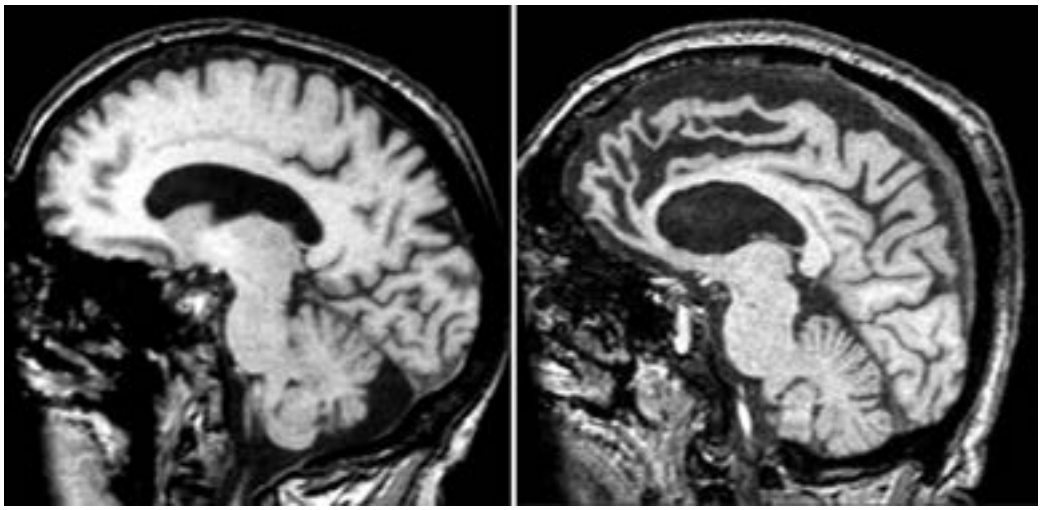

FIGURE 402-7. Parasagittal image from magnetic resonance imaging of a patient with frontotemporal dementia (*left*). Atrophy of the frontal lobes is dramatic compared with the brain of a normal individual (*right*). (Courtesy Maria Shiung and Clifford Jack.)

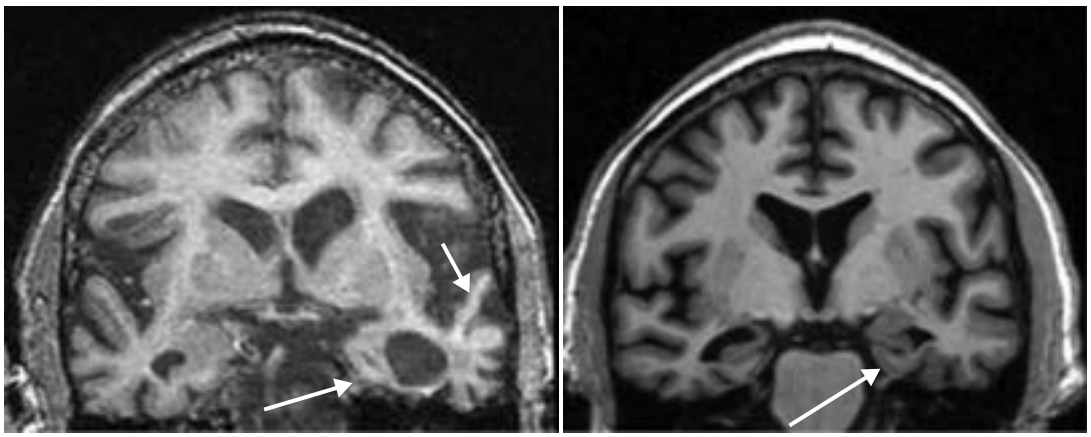

FIGURE 402-8. Coronal images from magnetic resonance imaging of a patient with semantic variant of primary progressive aphasia (*left*). There is prominent asymmetrical atrophy of the left anterior temporal lobe involving the amygdala, head of the hippocampus, and the lateral temporal lobe neocortex. By comparison, on the scan of a patient with Alzheimer disease dementia (*right*), the neocortex is preserved even though there is atrophy involving the amygdala and head of the hippocampus (*arrows*).

TREATMENT

Rx

There is no symptomatic therapy specifically for frontotemporal lobar degeneration. In patients with agitation, paranoia, delusions, or obsessive behavior, atypical antipsychotics (e.g., quetiapine, 25 to 200 mg/day) are used, but no controlled clinical trials are available. There are no preventive or disease-modifying treatments of frontotemporal lobar degeneration.

PROGNOSIS

Specific frontotemporal lobar degeneration syndromes have dramatic differences in their clinical course and outcome. In patients with motor neuron signs and symptoms, the prognosis is usually poor, with survival of only about 2 years from the time of diagnosis. Patients with semantic variant and nonfluent variant primary progressive aphasia have much more protracted and gradual trajectories; survival for more than 10 years is not uncommon. Behavior variant frontotemporal dementia itself can also exhibit a more protracted course.

Grade A References

A1. Petersen RC, Thomas RG, Grundman M, et al. Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med*. 2005;352:2379-2388.

A2. Lautenschlager NT, Cox KL, Flicker L, et al. Effect of physical activity on cognitive function in older adults at risk for Alzheimer disease: a randomized trial. *JAMA*. 2008;300:1027-1037.

A3. Aisen PS, Schneider LS, Sano M, et al. High-dose B vitamin supplementation and cognitive decline in Alzheimer disease: a randomized controlled trial. *JAMA*. 2008;300:1774-1783.

A4. Banerjee S, Hellier J, Dewey M, et al. Sertraline or mirtazapine for depression in dementia (HTA-SADD): a randomised, multicentre, double-blind, placebo-controlled trial. *Lancet*. 2011;378:408-411.

A5. Courtney C, Farrell D, Gray R, et al. AD2000 Collaborative Group. Long-term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): randomised double-blind trial. *Lancet*. 2004;363:2105-2115.

A6. Raina P, Santaguida P, Ismaila A, et al. Effectiveness of cholinesterase inhibitors and memantine for treating dementia: evidence review for a clinical practice guideline. *Ann Intern Med*. 2008;148:379-397.

A7. Howard R, McShane R, Lindesay J, et al. Donepezil and memantine for moderate-to-severe Alzheimer's disease. *N Engl J Med*. 2012;366:893-903.

A8. Winblad B, Kilander L, Eriksson S, et al. Donepezil in patients with severe Alzheimer's disease: double-blind, parallel-group, placebo-controlled study. *Lancet*. 2006;367:1057-1065.

A9. Reisberg B, Doody R, Stoffler A, et al. Memantine in moderate-to-severe Alzheimer's disease. *N Engl J Med*. 2003;348:1333-1341.

A10. Dysken MW, Sano M, Asthana S, et al. Effect of vitamin E and memantine on functional decline in Alzheimer disease: the TEAM-AD VA cooperative randomized trial. *JAMA*. 2014;311:33-44.

A11. Grodstein F, O'Brien J, Kang JH, et al. Long-term multivitamin supplementation and cognitive function in men: a randomized trial. *Ann Intern Med*. 2013;159:806-814.

A12. Doody RS, Thomas RG, Farlow M, et al. Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N Engl J Med*. 2014;370:311-321.

A13. Salloway S, Sperling R, Fox NC, et al. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. *N Engl J Med*. 2014;370:322-333.

A14. Doody RS, Raman R, Farlow M, et al. A phase 3 trial of semagacestat for treatment of Alzheimer's disease. *N Engl J Med*. 2013;369:341-350.

A15. Erkinjuntti T, Kurz A, Gauthier S, et al. Efficacy of galantamine in probable vascular dementia and Alzheimer's disease combined with cerebrovascular disease: a randomised trial. *Lancet*. 2002;359:1283-1290.

A16. Rolinski M, Fox C, Maidment I, et al. Cholinesterase inhibitors for dementia with Lewy bodies, Parkinson's disease dementia and cognitive impairment in Parkinson's disease. *Cochrane Database Syst Rev*. 2012;3:CD006504.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Hebert LE, Weuve J, Scherr PA, et al. Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology*. 2013;80:1778-1783.
2. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:263-269.
3. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:270-279.
4. Langa KM, Levine DA. The diagnosis and management of mild cognitive impairment: a clinical review. *JAMA*. 2014;312:2551-2561.
5. Hyman BT, Phelps CH, Beach TG, et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement*. 2012;8:1-13.
6. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med*. 2013;368:107-116.
7. Bateman RJ, Xiong C, Benzinger TL, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med*. 2012;367:795-804.
8. Marcus C, Mena E, Subramaniam RM. Brain PET in the diagnosis of Alzheimer's disease. *Clin Nucl Med*. 2014;39:e413-e422.
9. Johnson KA, Minoshima S, Bohnen NI, et al. Appropriate use criteria for amyloid PET: a report of the Amyloid Imaging Task Force, the Society of Nuclear Medicine and Molecular Imaging, and the Alzheimer's Association. *J Nucl Med*. 2013;54:476-490.
10. Okamura N, Furumoto S, Fodero-Tavoletti MT, et al. Non-invasive assessment of Alzheimer's disease neurofibrillary pathology using 18F-THK5105 PET. *Brain*. 2014;137:1762-1771.
11. Ihl R, Bunevicius R, Frolich L, et al. World Federation of Societies of Biological Psychiatry guidelines for the pharmacological treatment of dementias in primary care. *Int J Psychiatry Clin Pract*. 2014;1-6.
12. Thacker EL, McKnight B, Psaty BM, et al. Atrial fibrillation and cognitive decline: a longitudinal cohort study. *Neurology*. 2013;81:119-125.
13. Crane PK, Walker R, Hubbard RA, et al. Glucose levels and risk of dementia. *N Engl J Med*. 2013;369:540-548.
14. Vann Jones SA, O'Brien JT. The prevalence and incidence of dementia with Lewy bodies: a systematic review of population and clinical studies. *Psychol Med*. 2014;44:673-683.
15. Goldman JG, Williams-Gray C, Barker RA, et al. The spectrum of cognitive impairment in Lewy body diseases. *Mov Disord*. 2014;29:608-621.
16. Jeppsson A, Zetterberg H, Blennow K, et al. Idiopathic normal-pressure hydrocephalus: pathophysiology and diagnosis by CSF biomarkers. *Neurology*. 2013;80:1385-1392.
17. Knopman DS, Roberts RO. Estimating the number of persons with frontotemporal lobar degeneration in the US population. *J Mol Neurosci*. 2011;45:330-335.
18. Sieben A, Van Langenhove T, Engelborghs S, et al. The genetics and neuropathology of frontotemporal lobar degeneration. *Acta Neuropathol*. 2012;124:353-372.
19. Boeve BF, Boylan KB, Graff-Radford NR, et al. Characterization of frontotemporal dementia and/or amyotrophic lateral sclerosis associated with the GGGGCC repeat expansion in C9ORF72. *Brain*. 2012;135:765-783.
20. Karageorgiou E, Miller BL. Frontotemporal lobar degeneration: a clinical approach. *Semin Neurol*. 2014;34:189-201.
21. Rascovsky K, Hodges JR, Knopman D, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*. 2011;134:2456-2477.
22. Harris JM, Gall C, Thompson JC, et al. Sensitivity and specificity of FTDC criteria for behavioral variant frontotemporal dementia. *Neurology*. 2013;80:1881-1887.

REVIEW QUESTIONS

1. Which of the following features distinguishes the syndrome of mild cognitive impairment from dementia?

A. Degree of impairment in daily activities
B. Extent of anomia
C. Age
D. Degree of hippocampal atrophy

Answer: A By definition, persons with cognitive impairment are diagnosed with mild cognitive impairment if they are largely independent in daily affairs, whereas persons whose cognitive impairment makes them dependent on others are diagnosed with dementia. Although the severity of cognitive impairment is clearly correlated with the impairment in daily affairs, the level of cognitive impairment per se should not be used to distinguish mild cognitive impairment from dementia.

2. All of the following are true statements about the epidemiology of dementia due to Alzheimer disease ***except***:

A. The prevalence doubles every 5 years after age 65.
B. Family history of dementia is a risk factor for Alzheimer disease dementia.
C. Women are more than 3 times more likely to be affected than men.
D. There are no geographic regions with exceptionally high prevalence.

Answer: C Although there are more women than men with dementia, owing to lower survival in men after the age of 65 years, there is no difference in the age-adjusted prevalence between men and women.

3. Which of the following would be the most specific and sensitive indicator that cerebrovascular disease was the cause of a patient's cognitive impairment?

A. A diagnosis of hypertension
B. Bilateral infarcts in thalamus on magnetic resonance scan
C. A single infarct in the cerebellum
D. Family history of dementia
E. Sparing of learning and memory

Answer: B Bilateral infarcts in the thalamus are highly likely to cause cognitive impairment, including aphasia or amnesia. The other features have no relationship or only a weak relationship to cerebrovascular disease.

4. Which of the following is ***not*** a diagnostic feature of dementia with Lewy bodies?

A. Parkinsonism
B. Rapid eye movement (REM) sleep behavior disorder
C. Visual hallucinations
D. Unexplained frequent falls
E. Cerebellar ataxia

Answer: E Cerebellar ataxia is not a feature of dementia with Lewy bodies.

5. You are asked to see an 80-year-old man who has parkinsonism, dementia, prominent visual hallucinations, dream enactment behavior, and urinary incontinence. Which of the following would be likely to occur with treatment of the urinary incontinence with an anticholinergic medication?

A. Worsening of parkinsonism
B. Exacerbation of nocturnal dream enactment behavior
C. Increased confusion
D. Diarrhea

Answer: C Anticholinergic medications, even if they claim not to cross the blood-brain barrier, can cause increased confusion. The other symptoms are unlikely to be a result of anticholinergic medications.

403 THE EPILEPSIES

SAMUEL WIEBE

DEFINITION

A seizure is defined by transient focal or generalized signs or symptoms due to abnormal excessive or synchronous neuronal activity in the brain. Focal seizures, which originate within neuronal networks limited to one cerebral hemisphere, produce signs and symptoms corresponding to the specific region of the brain affected by the seizure. Generalized seizures rapidly affect extensive neuronal networks on both cerebral hemispheres, and their signs and symptoms are consistent with substantial involvement of both sides of the brain.

Seizures are not synonymous with epilepsy. The epilepsies should be distinguished from situations in which acute brain insults (e.g., infections, trauma, intoxication, metabolic disturbances) cause one or more seizures without a resulting chronic seizure tendency. Acute symptomatic seizures, or provoked seizures, constitute about 40% of all incident cases of nonfebrile seizures, typically respond to treatment of the provoking factor, and do not require long-term treatment with antiepileptic drugs.

The epilepsies are a group of conditions in which an underlying neurologic disorder results in a chronic tendency to have recurrent unprovoked seizures. Under these circumstances, the diagnosis of epilepsy is established if (1) two or more unprovoked seizures occur, or (2) one seizure occurs in a person whose risk of recurrence is at least 60%, or (3) one or more seizures occur in the context of a known epilepsy syndrome.[1] The causes, types, and clinical expression of the epilepsies are numerous and varied. However, some of the epilepsies conform into identifiable epileptic syndromes, which consist of clusters of clinical and electroencephalographic (EEG) features that have specific causes, respond to particular treatments, and may have specific prognostic implications.

EPIDEMIOLOGY

Incidence and Prevalence

Seizures are common in the general population, and about 1 in 10 people will experience a seizure in their lifetime. Most of these seizures are provoked by acute events and are not related to epilepsy. The overall annual incidence of acute symptomatic seizures, excluding febrile seizures, in developed countries is about 39 per 100,000 people. The incidence is higher in men and follows a bimodal age distribution. Incidence is at its highest peak in the first year of life (up to 300 per 100,000), reaches a nadir of 15 per 100,000 in the third and fourth decades of life, and rises again to 123 per 100,000 after 75 years of age. These differences are attributable to the high incidence of acute symptomatic seizures associated with metabolic, infectious, and encephalopathic causes during the neonatal period, and of cerebrovascular and degenerative diseases in elderly persons.

The epilepsies are common and affect humans of any age. After headache, the epilepsies are the most frequent chronic neurologic condition seen in general practice worldwide. In developed countries, the prevalence of active epilepsy ranges from 5 to 7 per 1000 persons, and the median annual incidence is 45 per 100,000 (range, 30 to 67), varying by age and socioeconomic status.[2] One in 26 people will develop epilepsy during their lifetime (1 in 21 males and 1 in 28 females).[3] The incidence of epilepsy peaks in children younger than 5 years at 60 to 70 per 100,000, decreases throughout adolescence to 30 per 100,000 in early adulthood, and rises again after the sixth decade, reaching a peak of 150 to 200 per 100,000 persons older than 75 years. Overall, the incidence and prevalence of the epilepsies are higher in developing countries, largely owing to a higher frequency of perinatal insults, trauma, and infectious disorders of the brain and to suboptimal treatment. In these countries, the median prevalence of active epilepsy is 12.5 per 1000 (range, 5 to 57 per 1000), and the annual incidence ranges from 78 to 190 per 100,000. Furthermore, the patterns of age-specific incidence are quite different in developing countries, where incidence peaks in young adults, not in elderly persons.

Risk Factors

Among all age groups, the top five risk factors for developing acute symptomatic seizures are head trauma (16%), stroke (16%), infectious disorders (15%), toxic-metabolic disorders (15%), and drug and alcohol withdrawal (14%) (Table 403-1).

TABLE 403-1 COMMON CAUSES OF ACUTE SYMPTOMATIC (PROVOKED) SEIZURES

METABOLIC
Hypernatremia, hyponatremia, hypocalcemia, hypoxia, hypoglycemia, nonketotic hyperosmolar hyperglycemia, renal failure
DRUG INDUCED
Theophylline, meperidine, tricyclic antidepressants, ephedra, gingko, phenothiazines, quinolones, β-lactams, isoniazid, antihistamines, cyclosporine, interferons, tacrolimus, cocaine, lithium, amphetamines
DRUG WITHDRAWAL
Alcohol, benzodiazepines, barbiturates
ENDOCRINE
Hyperthyroidism, hypothyroidism, peripartum
OTHER SYSTEMIC CONDITIONS
Sickle cell crisis, hypertensive encephalopathy, systemic lupus erythematosus, polyarteritis, eclampsia, high fever
CENTRAL NERVOUS SYSTEM DISORDERS
Trauma, stroke, intracerebral hemorrhage, encephalitis, abscess, bacterial meningitis

The risk factors for developing epilepsy differ in adults and children. In childhood, excluding inherited epilepsies, the risk is increased by febrile seizures, head trauma, infections of the brain, mental retardation, cerebral palsy, and attention-deficit/hyperactivity disorder. Perinatal insults do not carry an increased risk for epilepsy unless they are accompanied by mental retardation or cerebral palsy.

In adults, risk factors for developing epilepsy can be identified in only one third of patients, in whom head trauma, brain infections, stroke, and Alzheimer disease are the most common. The risk of developing epilepsy is increased more than 500-fold by a history of a military head injury, 30-fold by a severe civilian head injury (Chapter 399), 20-fold each by stroke (Chapter 407) and brain infections (Chapters 412 to 414),[4] and 10-fold each for Alzheimer disease (Chapter 402), migraine headache (Chapter 398), and hypertension. In Latin America, the most frequently identified risk factor is brain infection. In endemic areas, neurocysticercosis (Chapter 354) accounts for about 10% of all newly diagnosed cases of epilepsy.

Pathobiology

Pathogenesis

The pathologic substrates and mechanisms underpinning initiation and propagation differ for focal and generalized seizures. In focal seizures, an aggregate of cortical or subcortical neurons develop high-frequency bursts of sodium-dependent action potentials caused by a shift in calcium conductance, thereby resulting in the typical EEG spike discharge (Fig. 403-1). Spread of bursting activity to other neurons is normally prevented by surrounding inhibitory mechanisms, such as hyperpolarization and inhibitory interneurons. When a sufficient number of neurons are engaged in sustained bursting, further excitatory phenomena ensue, including the increased release of excitatory neurotransmitters owing to presynaptic accumulation of Ca^{2+}, depolarization of surrounding neurons owing to increased extracellular K^{+}, and further neuronal activation caused by depolarization-induced activation of *N*-methyl-D-aspartate (NMDA) receptors. As excitation increases and inhibition decreases, additional neurons are recruited regionally and distantly, thereby resulting in seizure propagation. The mechanisms by which neurons develop a tendency toward anomalous bursting activity include alterations in neurotransmitters, membrane receptors, ion channels, second-messenger systems, and gene expression of various proteins.

Considerably less is known about the basic mechanisms underlying generalized seizures, which depend prominently on thalamocortical circuits. In absence seizures, the classic generalized spike-and-wave discharges seen on EEG (Video 403-1) are related to alterations in oscillatory rhythms generated by circuits that connect the thalamus and cortex and that involve T-type Ca^{2+} channels, which are located in the reticular nucleus of the thalamus. In generalized convulsive seizures, cortical neurons exhibit prolonged depolarization during the tonic phase, followed by rhythmic depolarization and

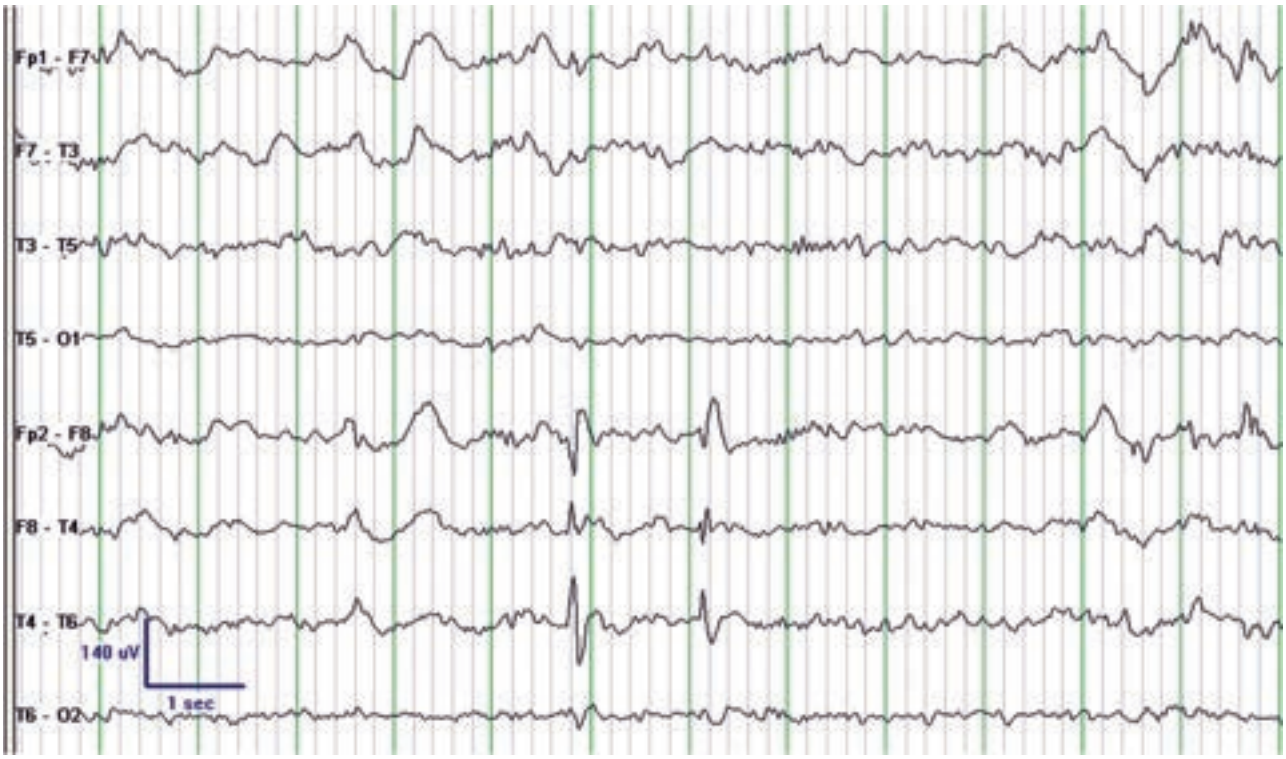

FIGURE 403-1. Selected electroencephalogram channels showing a typical right anterior temporal spike, the archetypal interictal footprint of temporal lobe epilepsy. The patient had right hippocampal sclerosis.

repolarization during the clonic phase. Activation of NMDA receptors increases calcium Ca^{2+} influx, thereby leading to further neuronal excitation. The initiation and modulation of generalized convulsive seizures involve cholinergic, noradrenergic, serotonergic, and histaminergic afferents from the brain stem and basal forebrain structures, which modulate excitability of hemispheric motor mechanisms.

Genetics

Only 15% of patients have one or more first-degree relatives who also suffer from epilepsy, and of those, about 75% have just one affected relative. However, the risk is still higher in first-degree relatives of patients with epilepsy than in the general population. In a large population-based study, the cumulative incidence of epilepsy to age 20 years was 2.5-fold higher in siblings and 3.4-fold higher in offspring of patients with epilepsy.

The genetics of epilepsy is evolving rapidly and can be categorized into three large groups[5]:

1. Conditions in which epilepsy forms part of a mendelian disorder include over 200 rare conditions that encompass neurocutaneous disorders (Chapter 417), neurodegenerative disorders, inherited malformations of cortical development (Chapter 417), and inherited metabolic disorders. For example, genes have been identified in progressive myoclonic epilepsies (e.g., Unverricht-Lundborg disease, Lafora disease, and the neuronal ceroid lipofuscinosis), X-linked myoclonic epilepsy with mental retardation, and cortical malformation syndromes (e.g., polymicrogyria, pachygyria, periventricular nodular heterotopia).
2. Epilepsies that can be directly explained by single gene mutations are rare and account for only about 1% of all epilepsy cases. Over 30 genes have been identified involving the following 15 epilepsy syndromes: genetic epilepsy with febrile seizures plus, severe myoclonic epilepsy of infancy and related syndromes, benign familial neonatal convulsions, benign familial neonatal-infantile seizures, benign familial infantile seizures, juvenile myoclonic epilepsy, childhood absence epilepsy, West syndrome, early infantile epileptic encephalopathy with suppression burst, malignant migrating partial seizures of infancy, autosomal dominant nocturnal frontal lobe epilepsy, familial infantile myoclonic epilepsy, epilepsy + paroxysmal exercise-induced dyskinesia, familial lateral temporal lobe epilepsy, and familial focal epilepsy with variable foci. Genetic mutations may affect neuronal excitability, neuronal metabolism, synaptic function, or network development. Although most of these gene mutations affect ion channels (*SCN1A, SCN1B, SCN2A, KCNQ2, KCNQ 3, KCNT1, KCTD7*), other cellular functions affected include neurotransmitter release (*STXBP1*), neurotransmitter receptors (*CHRNA, CHRNB, GABRD, GABRG2, GRIN2A, GRIN2B*), synaptic function (*SYN1*), glucose transport (*SLC2A1*), glutamate transport (*SLC25A22*), gene regulation and transcription (*ARX*), cell adhesion (*PCDH19*), cell membrane function (*PRRT2, TBC1D24, DEPDC5*), protein kinase and cell energy function (*CDKL5, BCKDK, ATP1A2*), and neuronal signaling (*EFHC1, LGI1, PLCB1*). An increased genetic predisposition for epilepsy is associated with specific genotypes (*MTHFR C677T*) in patients who develop post-traumatic epilepsy. Transcranial magnetic stimulation shows increased cortical excitability in siblings of patients with epilepsy, even when these epilepsies are acquired.
3. In some patients, the epilepsy is associated with "complex" disease genes. In this large group, which constitutes about 50% of all patients with epilepsy, multiple genes with individually small but additive effects act in combination with environmental factors to produce an increased risk for epilepsy.

CLINICAL MANIFESTATIONS

The clinical expression of seizures varies widely depending on the type of seizure and the areas of the brain involved by the epileptic activity. Accurate identification of the specific types of seizures determines the syndrome and dictates the type of drug the patient should receive.

Focal Seizures

Focal seizures originate within neuronal networks limited to one area of one cerebral hemisphere and produce signs and symptoms corresponding to the function subserved by the area of cerebral cortex engaged by the seizure (Table 403-2). Focal seizures are now subclassified according to their clinical expression; if consciousness or awareness is predominantly impaired, they are referred to as dyscognitive seizures. For example, patients who formerly were classified as having simple partial seizures now are classified as having focal seizures with preserved consciousness.

An aura consists of sensory, autonomic, or psychic symptoms that are experienced at the start of an observable seizure. The aura is a focal seizure itself, and it is often missed because patients and clinicians focus on the more dramatic dyscognitive or convulsive seizure that follows. Careful inquiry about the occurrence of an aura is of crucial importance for three reasons. First, it points to a focal as opposed to a generalized onset, thereby implying an underlying focal structural or functional brain abnormality (e.g., a tumor) that requires further investigation. Second, focal seizures have important implications for therapy and for prognosis (see later). Third, the nature of the

TABLE 403-2 CLINICAL MANIFESTATIONS OF DIFFERENT TYPES OF FOCAL SEIZURES AND AREAS OF THE BRAIN INVOLVED

SEIZURE TYPE	AREAS OF BRAIN INVOLVED	CLINICAL EXPRESSION
Somatosensory	Postcentral rolandic; parietal	Contralateral intermittent or prolonged tingling, numbness, sense of movement, desire to move, heat, cold, electric shock. Sensation may spread to other body segments.
	Parietal	Contralateral agnosia of a limb, phantom limb, distortion of size or position of body part
	Second sensory; supplementary sensory-motor	Ipsilateral or bilateral facial, truncal or limb tingling, numbness, or pain. Often involve lips, tongue, fingertips, feet
Motor	Precentral rolandic	Contralateral regional clonic jerking, usually rhythmic, may spread to other body segments in jacksonian motor march. Often accompanied by sensory symptoms in same area
	Supplementary sensory-motor	Bilateral tonic contraction of limbs causing postural changes, may exhibit classic fencing posture, may have speech arrest or vocalization
	Frontal	Contralateral head and eye version, salivation, speech arrest or vocalization; may be combined with other motor signs (as above) depending on seizure spread
Auditory	Heschl gyrus—auditory cortex in superior temporal lobe	Bilateral or contralateral buzzing, drumming, single tones, muffled sounds
Olfactory	Orbitofrontal; mesial temporal cortex	Often described as unpleasant odor
Gustatory	Parietal; rolandic operculum; insula; temporal lobe	Often unpleasant taste, acidic, metallic, salty, sweet, smoky
Vertiginous	Occipitotemporal-parietal junction; frontal lobe	Sensation of body displacement in various directions
Visual	Occipital	Contralateral static, moving, or flashing colored or uncolored lights, shapes, or spots. Contralateral or bilateral, partial or complete loss of vision.
	Temporal; occipitotemporal-parietal junction	Formed visual scenes, faces, people, objects, animals
Limbic	Limbic structures: amygdala, hippocampus, cingulum, olfactory cortex, hypothalamus	Autonomic: abdominal rising sensation, nausea, borborygmi, flushing, pallor, piloerection, perspiration, heart rate changes, chest pain, shortness of breath, cephalic sensation, lightheadedness, genital sensation, orgasm Psychic: déjà vu, jamais vu, depersonalization, derealization, dreamlike state, forced memory or forced thinking, fear, elation, sadness, sexual pleasure, hallucinations or illusions of visual, auditory, or olfactory nature
Dyscognitive	Usually bilateral involvement of limbic structures (see above)	Previously known as "complex partial seizures", characterized by a predominant alteration of consciousness or awareness. The current definition requires involvement of at least two of five components of cognition: perception, attention, emotion, memory, and executive function.

NOTE: Focal seizures may evolve into bilateral convulsive seizures.

TABLE 403-3 GENERALIZED SEIZURES: CLASSIFICATION AND CLINICAL EXPRESSION

SEIZURE TYPE	SUBTYPE	CLINICAL EXPRESSION
Absence	Typical	Abrupt cessation of activities, motionless, blank stare and loss of awareness lasting about 10 seconds. Attack ends suddenly, and patient resumes normal activities immediately.
	Atypical	Longer duration than typical absence, often accompanied by myoclonic, tonic, atonic, and autonomic features as well as automatisms
	With myoclonus	Absence with myoclonic components of variable intensity
Myoclonic	Myoclonic	Sudden, brief (<100 msec), shock-like, involuntary, single or multiple contractions of muscle groups of various locations
	Myoclonic atonic	A sequence consisting of a myoclonic followed by an atonic phase
	Myoclonic tonic	A sequence consisting of a myoclonic followed by a tonic phase
Tonic		Sustained increase in muscle contraction lasting a few seconds to minutes
Clonic		Prolonged regularly repetitive contractions involving the same muscle groups at a rate of 2-3 cycles per second
Atonic		Sudden loss or diminution of muscle tone lasting 1-2 seconds, involving head, trunk, jaw or limb musculature
Tonic-clonic		A sequence consisting of a tonic followed by a clonic phase

symptoms points to the area of the brain that gives rise to the seizure and that could be a target for surgical treatment.

The neuronal discharge causing the focal seizure may remain confined to the region where it began (as an aura or more objective focal event), or it may spread to involve additional brain areas. Thus, a focal seizure originating in the cortical area that represents sensation of the hand (rolandic area) may begin with contralateral hand tingling and then progress to involve additional cortical regions ipsilaterally, producing more extensive sensory symptoms as well as clonic motor signs. Seizures of rolandic origin in particular exhibit a peculiar type of propagation, in which the seizure activity "marches" from hand to arm to leg area ipsilaterally, a process referred to as a jacksonian march. After the clonic motor activity ends, patients are often weak; a postictal or Todd paralysis may last hours or even a day or 2, with gradual resolution (Video 403-2). The seizure may also propagate to distant ipsilateral or contralateral regions along known anatomic pathways.

In dyscognitive seizures, seizure propagation sufficiently involves limbic and bilateral structures to cause alteration of consciousness (Videos 403-3 and 403-4). Focal seizures originating from any region can become dyscognitive seizures, and unilateral focal seizures can progress to involve bilateral brain areas and cause a convulsive seizure (Video 403-5). Such convulsive seizures usually take the form of generalized tonic-clonic events rather than another type of generalized seizure (Table 403-3).

The evolution of the focal clinical seizure reflects the evolution of the EEG changes, which in turn reflects the pathophysiology of the process. A simultaneous rhythmic, localized discharge (often in the 4- to 7-Hz range) becomes higher in amplitude and lower in frequency as the seizure continues (see Video 403-5). Some seizures that begin in the association cortex (e.g., frontal or parietal lobes) have bizarre or extremely brief clinical manifestations without postictal deficits and create diagnostic challenges (Videos 403-6, 403-7, and 403-8). The stereotyped nature of the clinical events, with identification of EEG changes if present, may be the only way to make an appropriate diagnosis. The diagnosis can be even more challenging if the seizure spreads to different cortical regions during different seizure episodes, thereby producing variable constellations of clinical findings at different times.

Focal seizures with or without dyscognitive features can also occur as a series of single events without intervening normal behavior, thereby resulting in focal status epilepticus. Focal status epilepticus with dyscognitive seizures is characterized by prolonged confused behavior. EEG findings may be normal in a focal seizure without altered awareness, even in patients with status epilepticus, but the diagnosis is usually evident from the clinical

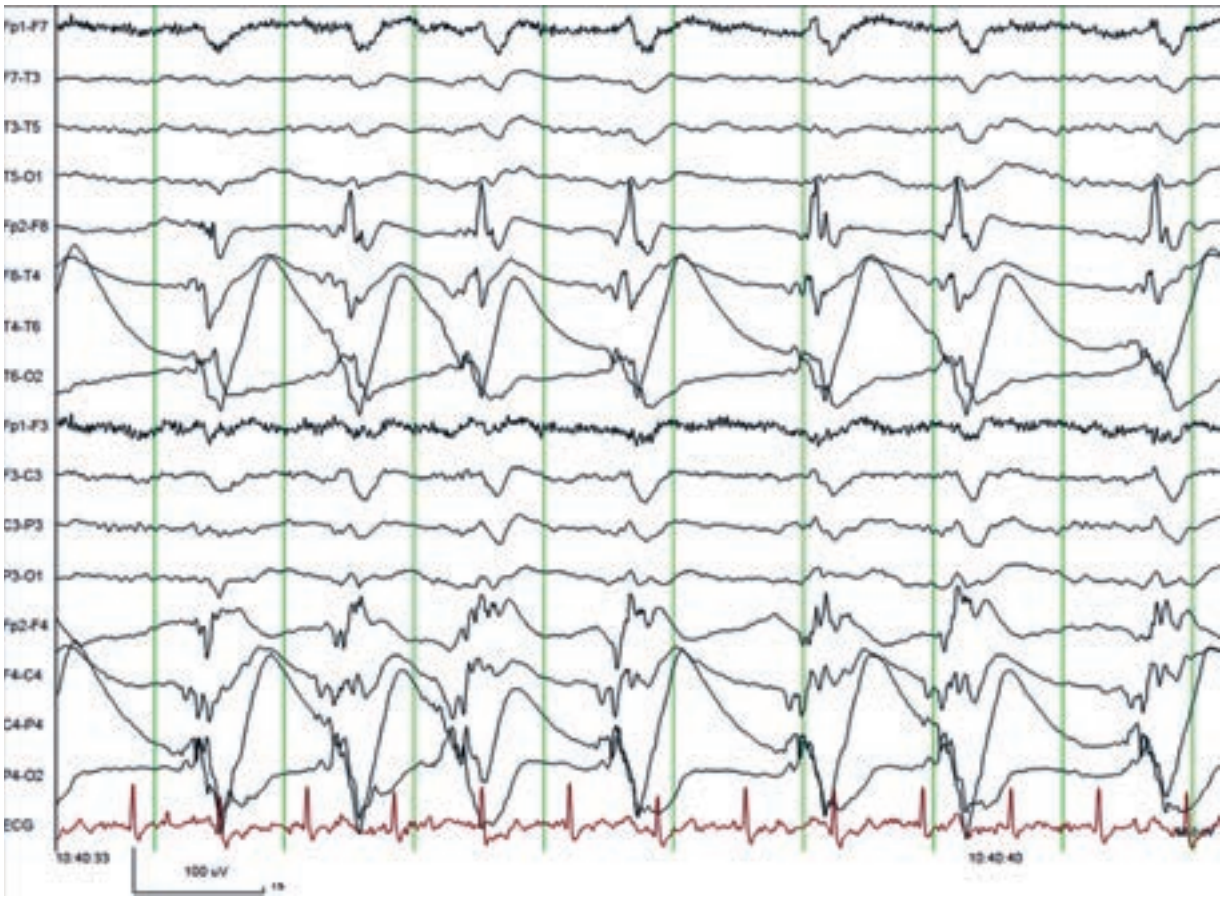

FIGURE 403-2. **Focal right hemisphere nonconvulsive status epilepticus in a comatose patient with a large right hemisphere infarct.**

features. In status epilepticus of focal dyscognitive seizures, EEG recordings show continuous abnormalities that are not of the same nature as seen in single seizures in that individual. The most common are a slow background with superimposed rhythmic high-amplitude sharp waves or repetitive rhythmic seizure discharges (Fig. 403-2). This type of status epilepticus is most frequent with frontal lobe seizures but can occur in temporal lobe seizures as well. The factors that precipitate status epilepticus are not well defined, nor are the implications for treatment or prognosis.

Nonconvulsive status epilepticus consists of a state of confusion or impaired mental status in patients with various neurologic diagnoses (i.e., trauma, stroke) in the acute intensive care unit setting (Video 403-9). It also denotes a condition that can occur de novo in older adults without a precipitating cause and that is characterized by prolonged confusional episodes, which are caused by generalized slow spike-and-wave status epilepticus. Clinical suspicion should prompt an EEG study, which is essential for diagnosis.

Generalized Seizures

Generalized seizures rapidly affect both cerebral hemispheres, and their clinical expression is consistent with substantial involvement of both sides of the brain (see Table 403-3). Convulsive seizures, which are also referred to as grand mal seizures, consist of excessive abnormal muscle contractions that may be sustained or interrupted and usually are a combination of tonic and clonic phases (generalized tonic-clonic seizures). This type of seizure may involve both hemispheres at the onset or may result from propagation of a focal seizure. These dramatic seizures often frighten witnesses and cause severe disruption of social interaction and development. They may begin with a "cry" as a result of abrupt air movement across the glottis from sudden tonic muscle contraction. The patient becomes diffusely stiff, usually with limb and body extension (Video 403-10). Breathing is suspended, cyanosis occurs, and urinary incontinence is common. After 15 to 45 seconds, the tonic activity gives way to clonic, rhythmic, sometimes asymmetrical jerking of all four extremities (Video 403-11). The rhythmic contractions gradually become slower in frequency until the event stops; the patient is apneic, comatose, and diaphoretic, but breathing with stridor and gasping begins within 60 seconds. Patients who have generalized tonic-clonic seizures in public often prompt bystanders to initiate resuscitation efforts, although such patients begin spontaneous respiration within 1 minute or so. Postictal stupor persists for a variable length of time. The patient generally sleeps for 2 to 8 hours and then complains of severe headache, sore muscles, a bitten tongue, and the inability to concentrate for a day or more. After generalized tonic-clonic seizures, some individuals have severe memory loss that gradually improves, sometimes over a period of weeks. Generalized tonic-clonic seizures also are a common expression of many metabolic, toxic, traumatic, or ischemic insults (see Table 403-1), but these provoked seizures do not qualify for the diagnosis of epilepsy.

Absence seizures, or petit mal seizures, are the second most common type of generalized seizure. Patients experience an abrupt onset and termination of a momentary lapse of awareness. Patients have no perception of any aspect of the event and may or may not realize that some time was lost, although individuals often lose their train of thought. Because consciousness is abruptly lost and immediately regained, there is neither an aura nor residual postictal symptoms. These seizures begin in childhood, and school teachers are often the first to notice them. In absence seizures, patients stop abruptly, stare vacantly, may have brief eye blinking or myoclonic movements (see Table 403-3), particularly if the event extends beyond 10 seconds (as judged by EEG), and regain function instantly (see Video 403-1). These seizures can occur many times a day but are not associated with progressive neurologic disease. They can also occur in a more continuous form as nonconvulsive status epilepticus with resultant confusion.

Some patients with extensive bilateral brain disease have a variation of absence seizures known as atypical absence. The event is similar in terms of loss of contact, but there is more motor, autonomic, or automatic activity, and the EEG demonstrates discharges that are slower than the 3-Hz spike and wave of typical absence seizures.

Myoclonic seizures consist of brief episodes of sudden motor contraction (see Table 403-3) that can be focal (Video 403-12), with one arm involved, or bilateral and massive, with involvement of the face, both upper extremities, and the trunk. Consciousness is preserved but can be difficult to evaluate because of the brevity of these seizures. Myoclonic seizures form part of three main clinical constellations: juvenile myoclonic epilepsy, which starts in childhood or adolescence and often persists into adulthood; epilepsy with various combinations of absence and myoclonic seizures; and epilepsy in

TABLE 403-4 DISORDERS RESEMBLING SEIZURES

VASCULAR AND PERFUSION DISORDERS
Migraine, syncope, transient ischemic attack, transient global amnesia, arrhythmia/hypoperfusion
PSYCHIATRIC DISORDERS
Psychogenic nonepileptic seizures, panic disorder, dissociative disorder
MOVEMENT DISORDERS
Tics, paroxysmal dystonia, paroxysmal choreoathetosis, paroxysmal ataxia
SLEEP DISORDERS
Night terrors, sleep walking, sleep myoclonus, narcolepsy/cataplexy, rapid eye movement sleep intrusions
METABOLIC DISTURBANCES
Alcoholic blackouts, delirium tremens, hypoglycemia, hallucinogenic drugs
OTHER
Breath-holding spells, paroxysmal vertigo, migraine with recurrent abdominal pain and cyclic vomiting

TABLE 403-5 CLINICAL FEATURES THAT HELP DISTINGUISH A GENERALIZED TONIC CLONIC SEIZURE FROM SYNCOPE

	SEIZURE	SYNCOPE
Clinical context and circumstances	Neurologic or systemic conditions that predispose to seizures, family history of seizures. Mental fatigue, sleep deprivation, alcohol use or withdrawal, systemic illness	Cardiovascular disorders, dehydration, anemia. Family history of syncope
Triggers	Usually none (unless reflex epilepsy)	Orthostatic hypotension, venipuncture, painful and noxious stimuli, emotional stress, micturition, Valsalva maneuver
Clinical features		
Onset	No warning unless there is an aura. Abrupt loss of consciousness, generalized stiffening, and fall. Occurs in any position.	Tiredness, nausea, diaphoresis, tunneling of vision. Loss of consciousness over few seconds and fall. Occurs usually standing.
Course	Prominent tonic phase then clonic movements lasting about 1 minute, cyanosis, labored breathing, may bite tongue or cheeks, sometimes urinary incontinence	Usually loss of tone, pallor, multifocal myoclonic jerks lasting < 15 seconds, sometimes urinary incontinence, usually no tongue or cheek biting
Offset	Postictal sleepiness and confusion lasting up to hours, headache, myalgia	Rapid recovery over seconds to less than few minutes, no confusion, headache or myalgia. May have fatigue.

the setting of degenerative or inherited syndromes with bilateral cerebral involvement and abnormal cerebral function. Myoclonic seizures most commonly occur in the morning after awakening and often increase in frequency to culminate in a generalized tonic-clonic seizure.

Atonic and tonic seizures are brief but extremely disabling motor events that are characterized by a sudden increase or decrease in muscle tone. The result is falls and injuries with variable impairment of awareness. Such seizures frequently begin in children with diffuse central nervous system (CNS) disease and multiple types of seizures, but they persist during adulthood (Videos 403-13 and 403-14).

DIAGNOSIS

The basic diagnosis of seizures is established by the clinical history. Although EEG, imaging, and laboratory studies are commonly required to determine the type of epilepsy, epilepsy syndrome, site of origin of focal seizures, and occurrence of nonepileptic seizures, the answer to the basic question of whether the patient's episodes are seizures or not rests almost entirely on a careful clinical history. The diagnosis of epilepsy can also be established by history, because epilepsy is defined as the occurrence of two unprovoked seizures or one unprovoked seizure in the context of a high underlying risk of recurrence or an epileptic syndrome.

Differential Diagnosis

The first question facing clinicians is whether the episodes under consideration are indeed seizures. The diverse clinical expression of seizures entails a large differential diagnosis among conditions that produce episodic neurologic dysfunction (Table 403-4). Common conditions resembling seizures include syncope (Chapters 51 and 62), transient ischemic attacks (Chapter 407), migraine (Chapter 398), movement disorders (Chapter 410), and psychogenic nonepileptic seizures (see Table 403-4).

A number of historical elements dramatically change the likelihood of this diagnosis. Three essential elements help determine whether an episode is a seizure (Table 403-5) and distinguish seizures from other causes of temporary loss of consciousness, especially syncope (Chapters 51 and 62):

1. The clinical context, including medical and family history and circumstances under which the episode occurred
2. Specific triggers or provoking factors
3. A detailed clinical description of the event, including four key components:
   - What is the first symptom or sign (presence and type of aura, evidence of focal seizure at onset)?
   - How does it evolve after onset (what happens during the seizure proper, what are the signs or symptoms, how long does it last)?
   - How does it end (gradually or abruptly)?
   - Are there any neurologic deficits after the seizure ends?

Because patients have limited or no recall, the history from others is crucial. Observers can contribute important information about the patient's activity, responses, and appearance, including changes in color, diaphoresis, respirations, vocalization, and muscle tone. This information is often essential to characterize the type of seizure and to distinguish seizures from conditions that resemble seizures.

Migraine (Chapter 398) and focal seizures not only resemble each other but also coexist as comorbid conditions and share genetic susceptibility loci. Features that favor a diagnosis of seizures over classic migraine include an inconsistent occurrence of headache during the event, a brief duration, and the occurrence of more severe seizures. Myoclonus (Chapter 410) occurs in a variety of settings (e.g., metabolic encephalopathies) without any association with epilepsy or the EEG changes seen in myoclonic epilepsy.

Frontal lobe seizures arise predominantly during sleep and can have dramatic motor expression. They can be confused with nonepileptic psychogenic seizures, sleep disorders (Chapter 405), or movement disorders (Chapters 409 and 410). Video EEG monitoring may be necessary for diagnosis (see Videos 403-7 and 403-8).

Patients with panic attacks (Chapter 397) can experience events that mimic focal seizures with autonomic and psychic features. However, panic attacks usually have a longer duration, do not progress to more severe seizures, and can be linked to specific circumstances. Nevertheless, focal seizures with limbic symptoms are often misdiagnosed as panic attacks.

Psychogenic nonepileptic seizures are behaviors that resemble seizures and are often part of a conversion reaction (Chapter 397) precipitated by underlying psychological distress. Psychogenic seizures can be difficult to diagnose because they can mimic almost any type of seizure, and they often coexist with epilepsy in the same patient. An erroneous diagnosis of nonepileptic seizures poses a risk for inappropriate discontinuation of medication, with resulting status epilepticus. Conversely, an erroneous diagnosis of seizures can result in iatrogenic illness owing to unnecessary therapy, excessive sedation, and cardiorespiratory depression. Features suggesting nonepileptic seizures include variable clinical manifestations across episodes, frequent and prolonged episodes, lack of response to antiseizure medication, out-of-phase upper and lower body movements, prominent pelvic thrusting, and lack of rigidity. Secondary gain is usually evident, and there is often a history of sexual abuse. Nevertheless, the peculiarities of these attacks may require continuous video EEG monitoring for diagnosis.

Diagnostic Investigations

A detailed history, EEG recordings, and magnetic resonance imaging (MRI) can lead to a definitive diagnosis of epilepsy and its cause in up to 50% of

patients. In other patients, the information is insufficient or inconsistent, but the physiologic and CNS abnormalities surrounding the actual event allow it to be placed provisionally into a specific diagnostic category in about another 30% of patients. Continuous video EEG monitoring in an inpatient epilepsy unit can increase diagnostic sensitivity and specificity.

Single Seizures

Acute symptomatic seizures (see Table 403-1) are the known consequence of an acute condition, and investigations should be directed at the possible cause of these seizures. When no known cause is readily apparent, the seizures are considered to be unprovoked. Evaluation of patients who present with a first unprovoked seizure includes either brain computed tomography (CT) or MRI, which reveals a possible cause in about 10% of patients. An EEG obtained after the seizure will demonstrate abnormalities with prognostic significance in 20 to 25% of these patients. Blood tests (including levels of serum electrolytes, glucose, calcium, and magnesium; tests of liver and kidney function; a complete blood cell count; and screening for suspected toxins) will reveal abnormalities in up to 15% of these patients but are often nonspecific. Lumbar puncture is indicated if CNS infections are suspected and in all patients infected with human immunodeficiency virus (HIV), even in the absence of clinical findings suggestive of infection.

Epilepsy

Electroencephalogram

The EEG is the keystone investigation in all patients with seizures and epilepsy. Between seizures, the EEG can assess overall brain function and the type, location, and amount of epileptiform (spike) discharges. The EEG is crucial in determining the epilepsy syndrome and choosing appropriate antiepileptic drugs. In focal epilepsies, the EEG often demonstrates focal slowing and spike discharges in the area of abnormality.

The EEG can establish the definitive diagnosis of epilepsy if electrical changes consistent with a seizure are recorded during a clinical seizure. However, the EEG may fail to demonstrate electrical changes during a typical clinical seizure if the seizure focus is too small (at least 6 cm^2 of cortical involvement is needed to create an EEG epileptiform change), if the seizure focus is deep or in the mesial or inferior surfaces of the brain, or if the event in question is not an epileptic seizure. The EEG is always abnormal during generalized convulsive and absence seizures.

The initial EEG is normal in up to 60% of people with known epilepsy. However, epileptiform abnormalities occur in more than 80% of individuals with focal epilepsy if three or more interictal EEG studies are performed. In generalized epilepsies, interictal epileptiform discharges are more common and are easier to capture in the EEG (see Video 403-1).

The type of abnormality points to the epileptic syndrome. For example, the EEG can show hypsarrhythmia in West syndrome (see later) or the classic 3-Hz generalized spike wave in generalized epilepsies with absence seizures (see Table 403-3 and Video 403-1).

In some circumstances, it is imperative to record seizures, such as in the evaluation of patients for epilepsy surgery and when the diagnosis of seizures is in question (Video 403-15; also see Video 403-7). Continuous video EEG monitoring for prolonged periods has made it possible to capture these events. Continuous EEG is also used in comatose patients in the intensive care unit setting when nonconvulsive seizure or status epilepticus is suspected.

Magnetoencephalography

Magnetoencephalography measures the small magnetic fields that are generated by electrical activity in the brain and approximates their location using mathematical models. Its use is largely restricted to the evaluation of patients for epilepsy surgery, in whom it is used for mapping interictal discharges and the localization of brain function when superimposed on brain MRI.

Imaging Studies

Brain MRI, which can demonstrate lesions in most patients whose epilepsy is associated with a structural cause, should be performed in essentially all patients with new-onset seizures. The most common lesions in adults with new-onset focal seizures are post-stroke or post-traumatic gliosis or encephalomalacia (50%), tumors (15%), vascular abnormalities (15%), developmental abnormalities (15%), and mesial temporal sclerosis (9%).[6] The use of fluid-attenuated inversion recovery (FLAIR) (Fig. 403-3A) sequences increases the sensitivity to detect abnormalities of cortical development as well as hippocampal sclerosis, which point to the need for chronic anticonvulsant therapy or possible surgical treatment. Functional imaging procedures such as positron emission tomography (PET) for analysis of metabolism and single-photon emission computed tomography (SPECT) (see Fig. 403-3B) for determination of blood flow are also used to help localize areas of the brain to be targeted with epilepsy surgery.

Genetic Testing

Based on genetic test accuracy, implications for diagnosis and management, and ability to offer genetic counseling, an international consensus panel has identified eight epilepsy syndromes of genetic origin for which genetic testing of patients is most useful: Ohtahara syndrome, early-onset infantile spasms, X-linked infantile spasms, Dravet syndrome, epilepsy and mental retardation limited to females, early-onset absence epilepsy, autosomal dominant nocturnal frontal lobe epilepsy, and epilepsy with paroxysmal exercised-induced dyskinesia. As with other conditions, the ethical aspects and potential harms and benefits of genetic testing must be carefully considered.

Epileptic Syndromes and Constellations

Epileptic syndromes include 27 age-related syndromes, of which all but 6 begin or occur in infancy and childhood (Table 403-6). In addition, specific clinical constellations represent diagnostically meaningful forms of epilepsy, with specific implications for treatment, especially surgery, and also categorize for structural-metabolic causes, epilepsies of unknown cause, and conditions characterized by seizures that are not a form of epilepsy (e.g., febrile seizures). The diagnosis of epileptic syndromes and constellations is based on the types of seizures, the setting in which seizures occur, the patient's neurologic and cognitive status, age at onset, family history, and results of diagnostic studies, including EEG and MRI. The selection of specific drug and surgical treatment depends on the types of seizures present (Table 403-7). The need for lifelong treatment, the risk for genetic transmission, the likelihood of concurrent neurologic diseases, the risk for comorbid conditions, and the long-term prognosis are critical factors that can be addressed only with knowledge of the specific epileptic syndrome or constellation.

Some Specific Seizure Syndromes and Constellations

Neonatal and Infantile Epilepsy Syndromes

Benign neonatal convulsions occur in previously healthy newborns on about day 5 as focal or generalized tonic seizures. Mutations in two potassium channel genes (*KCNQ2, KCNQ3*) have been associated with this syndrome. Potassium channel regulation may be age dependent and therefore account for the age-related appearance of the seizures. The EEG shows rhythmic slow-wave activity or spiking with seizures. The seizures are refractory to treatment, are recurrent over a brief interval, and disappear within a month. About 90% of such infants subsequently have normal development, whereas 10 to 20% have subsequent seizures.[7]

Genetic epilepsy with febrile seizures plus (GEFS+) is a syndrome that consists of febrile seizures in combination with other nonfebrile types of seizures, including myoclonic, absence, atonic, tonic-clonic, and focal seizures. Mutations in at least three genes for voltage-gated ion sodium channels (*SCN1A, SCN1B, SCN2A*), two for GABA receptors (*GABRD, GABRG2*), and one for cell adhesion function (*PCDH19*) have been identified.

Dravet syndrome (severe myoclonic epilepsy of infancy) starts in the first year of life with myoclonic seizures plus other seizure types, including absence, atonic, and focal. In this devastating syndrome, the seizures are resistant to treatment and are accompanied by developmental and cognitive decline. Mutations in the *SCN1A* sodium channel have been identified.

West syndrome comprises a triad of epileptic spasms, developmental arrest, and an EEG pattern called hypsarrhythmia (a markedly abnormal EEG pattern with high-amplitude slowing and superimposed multifocal spikes, polyspikes, and spike and slow-wave complexes). It appears before the age of 12 months and ceases by 5 years of age, often to be replaced by other epilepsy syndromes such as Lennox-Gastaut syndrome. Tuberous sclerosis (Chapter 417) and hypoxia are among the common causes, but a cause may not be found. Associated abnormalities often include developmental delay, porencephaly, atrophic lesions, calcifications, and agenesis of the corpus callosum. West syndrome and early infantile epileptic encephalopathy have been associated with mutations in genes involved in a number of neurotransmitter and cellular functions (*ARX, CDKL5, STXBP1*).

Childhood Epilepsy Syndromes

Childhood absence epilepsy begins before age 12 years, and its onset peaks at age 5 to 7 years, with a strong genetic tendency. It is more common in girls

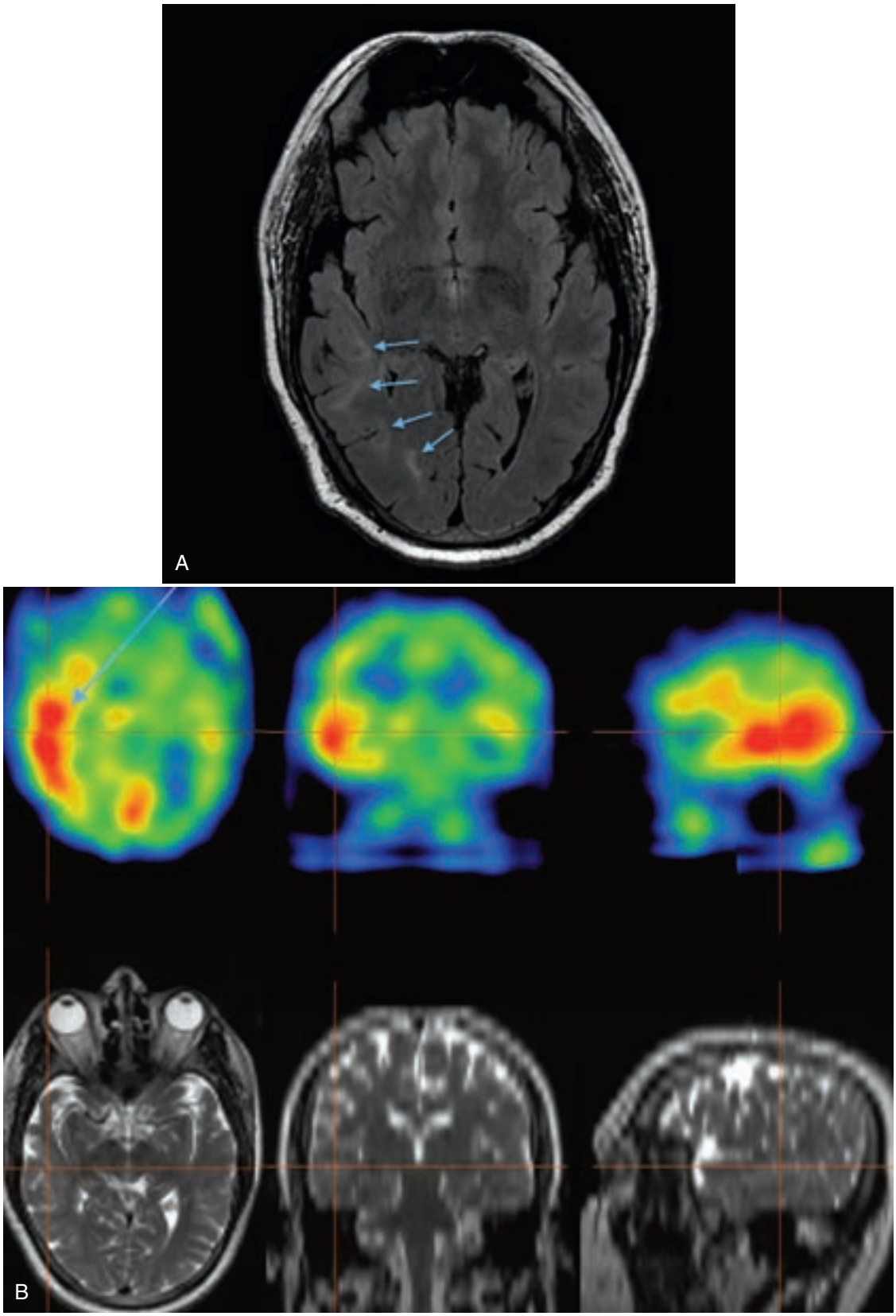

FIGURE 403-3. **Imaging studies from a patient with dramatic motor seizures that were initially attributed incorrectly to psychogenic nonepileptic events (see Video 403-7). A,** Fluid-attenuated inversion recovery (FLAIR) axial magnetic resonance image (MRI) demonstrating a large developmental cortical abnormality involving the mid-posterior right temporal lobe. **B,** Ictal SPECT demonstrating an area of hyperperfusion during a seizure that corresponds to the abnormality seen on the MRI and confirms the area of seizure origin.

than boys and is characterized by very frequent daily absence seizures (up to hundreds per day), rarely with other types of generalized seizures. It occurs in the setting of otherwise normal brain structure and function, and it is self-limited in about 40% of cases. The seizures are accompanied by a characteristic 3-Hz spike-and-wave EEG discharge, which appears in short bursts between seizures and in continuous runs during seizures. Remission usually occurs before the age of 12 years, but generalized tonic-clonic seizures occasionally may develop in adolescence. In early-onset absence epilepsy, mutations have been found in genes related to GABA receptors (*GABRA1*, *GABRG2*) and glucose transport (*SLC2A1*).

TABLE 403-6 EPILEPTIC SYNDROMES AND CONSTELLATIONS ACCORDING TO THE NEW INTERNATIONAL CLASSIFICATION

BY AGE AT ONSET
Neonatal Period
Benign familial neonatal epilepsy
Early myoclonic encephalopathy
Ohtahara syndrome
Infancy
Epilepsy of infancy with migrating partial seizures
West syndrome
Myoclonic epilepsy in infancy
Benign infantile epilepsy
Benign familial infantile epilepsy
Dravet syndrome
Myoclonic encephalopathy in nonprogressive disorders
Childhood
Febrile seizures plus (can start in infancy)
Panayiotopoulos syndrome
Epilepsy with myoclonic atonic (previously astatic) seizures
Benign epilepsy with centrotemporal spikes
Autosomal dominant nocturnal frontal lobe epilepsy
Late-onset childhood occipital epilepsy
Epilepsy with myoclonic absences
Lennox-Gastaut syndrome
Epileptic encephalopathy with continuous spike and wave during sleep
Landau-Kleffner syndrome
Childhood absence epilepsy
Adolescence-Adult
Juvenile absence epilepsy
Juvenile myoclonic epilepsy
Epilepsy with generalized tonic-clonic seizures alone
Progressive myoclonus epilepsies
Autosomal dominant partial epilepsy with auditory features
Other familial temporal lobe epilepsies
LESS SPECIFIC AGE RELATIONSHIP
Familial focal epilepsy with variable foci (childhood to adult)
Reflex epilepsies
DISTINCTIVE CONSTELLATIONS
Mesial temporal lobe epilepsy with hippocampal sclerosis
Rasmussen syndrome
Gelastic seizures with hypothalamic hamartoma
Hemiconvulsion-hemiplegia-epilepsy

Lennox-Gastaut syndrome is one of the most severe childhood epilepsies. It starts before age 8 years (peak from 3 to 5 years) and is characterized by a triad of mental retardation, multiple types of generalized seizures (atypical absence, generalized tonic-clonic, tonic, atonic), focal seizures that are highly resistant to treatment, and a typical EEG pattern of slow spike and wave (slower than the typical 3 Hz associated with absence seizures) and bursts of fast rhythms at 10 to 12 Hz during sleep. It often follows the resolution of West syndrome.

Benign epilepsy with centrotemporal spikes (benign rolandic epilepsy) starts between 3 and 13 years of age and is characterized by almost exclusively nocturnal focal motor or sensory seizures that have a facial or oral onset and often evolve to convulsive seizures. Nearly 50% of cases have a family history of epilepsy, but most patients have no known brain abnormality. The EEG shows spiking in the centrotemporal region. In some cases, the disorder may not require treatment because it usually remits spontaneously.

Adolescence and Adult Epilepsy Syndromes and Constellations

Juvenile myoclonic epilepsy usually starts in the second decade with generalized tonic-clonic and myoclonic seizures. Mutations in γ-aminobutyric acid (GABA) receptors (*GABRG1*) and in genes related to neuronal signaling (*EFHC1*) can be found. Seizures typically occur in the morning immediately after awakening. The seizures are especially linked to sleep deprivation and tend to appear in college students. A proportion of these patients have had absence seizures as well. The EEG typically shows fast (4 to 6 Hz) generalized spike and wave. Lifetime treatment is generally needed.

Mesial temporal lobe epilepsy with hippocampal sclerosis is the most common epilepsy to produce focal dyscognitive seizures in adults. It is characterized by recurrent focal limbic seizures (see Table 403-2), with and without impaired awareness, that originate in mesial temporal and limbic structures. Up to 70% of patients have a risk factor such as lengthy and complicated seizures before the age of 4 years, frequently associated with fever or encephalitis, meningitis, or trauma. However, the characteristic seizures generally begin some years later. Although most cases are sporadic, familial forms of mesial temporal lobe epilepsy have been associated with a novel susceptibility locus on chromosome 18(P11.31).

Various components of the mesial temporal limbic network (including the hippocampus, entorhinal cortex, amygdala, neocortical areas of the frontal and temporal lobes, and dorsal medial thalamus) are probably involved in the pathogenesis of these seizures. Mesial temporal sclerosis, also called hippocampal sclerosis, is characterized by neuronal loss and gliosis, mostly in the CA1 and CA3 regions of the hippocampus, with mossy fiber reorganization that is seen as sprouting of neuropeptide Y and dynorphin interneurons into the inner third of the dentate molecular layer. Whether hippocampal sclerosis is the cause or the result of seizures (or both) is not known. However, up to 12% of children with febrile status epilepticus have MR evidence of hippocampal injury, thereby suggesting a causal association.[8] The seizures of mesial temporal lobe epilepsy often begin at 5 to 15 years of age. Seizures are typically dyscognitive with limbic symptoms; they begin with an aura of a rising epigastric sensation or a feeling of déjà vu, followed by oral and alimentary automatisms and later by contralateral arm dystonia and ipsilateral arm automatisms. The seizures are lengthy (lasting several minutes), rarely generalize, and typically occur three to five times a month. Auras without subsequent seizures are common. Hippocampal atrophy and increased hippocampal signal are best seen on T2-weighted and FLAIR coronal MRI sequences, and widespread interictal hypometabolism is seen in the temporal lobe on PET. Material-specific (verbal or visual) memory impairment corresponds to primary involvement of the dominant or nondominant hippocampus. EEG recordings show temporal lobe spikes interictally as well as rhythmic 4- to 7-Hz discharges over the appropriate temporal lobe during seizures.

Seizures with Less Specific Age Relationship

Reflex seizures are triggered reliably by specific simple (e.g., flashing lights, sound) (see Video 403-15) or elaborate (e.g., reading) stimuli. The mechanisms are diverse and may involve cortical and brain stem pathways, cortical dysregulation of extracellular calcium concentrations, and an imbalance between excitatory and inhibitory neurotransmitters. Visual-sensitive seizures (triggered by light or visual patterns) are the most common type of reflex seizures. They occur most commonly in females, and their incidence peaks around puberty, when they represent up to 10% of all new cases of epilepsy. Other triggers of reflex seizures include specific thoughts, actions, reading, tactile stimuli, adopting certain positions, eating, listening to music, startle, and contact with hot water. The triggered seizures can be myoclonic, convulsive, atonic, or focal, depending on the triggering stimulus. Avoiding the offending stimulus is crucial to avoid seizures, emphasizing the importance of careful questioning about seizure triggers in patients with epilepsy.

TREATMENT

Rx

The treatment of seizures and epilepsy is guided by accurate knowledge of the type of seizure and epileptic syndrome, the probability of recurrent seizures, the likelihood and severity of psychosocial or physical consequences with further seizures, and whether the benefit from treatment substantially outweighs the risks for side effects. It is important to identify and correct any environmental, physiologic, or lifestyle factors, such as sleep deprivation and irregular sleep habits, and alcohol abuse, which can lower the seizure threshold and trigger seizures in patients with epilepsy.

Single Unprovoked Seizures

The decision to treat single unprovoked seizures depends on the likelihood of recurrence according to prognostic variables (see Prognosis) and on the individual patient's profile and preference. A meta-analysis demonstrated that antiepileptic drug treatment after a first seizure reduces the absolute risk of having a second seizure in the short term by 33%, corresponding to a number needed to treat (NNT) of 3. However, at least two randomized trials have shown that treatment of the first seizure with antiepileptic drugs does not prevent the development of epilepsy in the long term.[■] Therefore, the decision to treat the first seizure should be individualized based on the patient's

TABLE 403-7 ANTIEPILEPTIC DRUG SELECTION BY SEIZURE TYPE

SEIZURE TYPE	COMMONLY USED (ALPHABETICAL ORDER)	LESS COMMONLY USED (ALPHABETICAL ORDER)	*EFFECTIVENESS (GRADE A RECOMMENDATION)*	
			NEW-ONSET SEIZURES	REFRACTORY SEIZURES
Focal seizures with or without dyscognitive features or evolution to convulsions	Carbamazepine (CBZ) Gabapentin (GBP) Lamotrigine (LTG) Levetiracetam (LEV) Oxcarbazepine (OXC) Phenytoin (PHT) Tiagabine (TIAG) Topiramate (TPM) Valproate (VPA) Zonisamide (ZNS) Lacosamide (LAC)	Acetazolamide (ACZ) Clonazepam (CLN) Clorazepate (CLZ) Phenobarbital (PB) Primidone (PRM) Felbamate (FBM)	CBZ*§ GBP*§ LEV§ LTG*§ OXC*§ PB* PHT*§ TPM* VPA* ZNS§	CBZ† GBP* LAM* LEV* OXC* PB† PHT† TIAG* TPM* VPA† ZNS*
Generalized convulsive seizures (clonic, tonic or tonic-clonic seizures)	Carbamazepine Lamotrigine Levetiracetam Oxcarbazepine Phenytoin Topiramate Valproate Zonisamide	Acetazolamide Clonazepam Clorazepate Felbamate Phenobarbital Primidone	CBZ† LEV* LTG* PHT† VPA*	CBZ† LAM* LEV* PHT† TPM* VPA*
Absence seizures	Ethosuximide (ESM) Lamotrigine Valproate Topiramate	Acetazolamide Clonazepam Phenobarbital Primidone	ESM*§ LTG* VPA*§	
Myoclonic seizures	Clonazepam Levetiracetam Valproate Zonisamide	Phenobarbital	VPA†	

*Supported by class I evidence, American Academy of Neurology.
§Supported by class I evidence for initial monotherapy, International League Against Epilepsy.
†Often the "standard" of comparison, without evidence of effectiveness by randomized controlled trials.

preference, the risk for and impact of recurrent seizures (e.g., driving and employment), and the risk for medication side effects.

Acute Symptomatic (Provoked) Seizures

Seizures that are provoked by specific exposures are usually self-limited and not associated with an enduring seizure tendency, so the primary therapeutic consideration should be identification and treatment of the underlying disorder (see Table 403-1). However, the risk of developing epilepsy after febrile seizures is about 10 times that of the general population. If antiepileptic drugs are needed to treat seizures acutely, they usually can be discontinued after the patient has recovered from the primary illness. Some acute conditions like stroke (Chapter 407), brain infections (Chapters 412 to 414), and trauma (Chapter 399) can produce both acute symptomatic seizures and an enduring seizure tendency, so it would seem logical to use long-term antiepileptic drug treatment. To date, however, randomized controlled trials have not been able to demonstrate that antiepileptic drugs prevent the development of epilepsy in these conditions, so long-term therapy is not recommended unless epilepsy develops.

Epilepsy Syndromes with a Favorable Course

In syndromes such as benign epilepsy of childhood with centrotemporal spikes and some types of childhood occipital epilepsy, seizures are mild, infrequent, or exclusively nocturnal, and they remit spontaneously, thereby making treatment generally unnecessary. In selected cases, treatment may be desirable to prevent recurrences and to help alleviate parental concerns. In such cases, drug treatment is usually limited to 1 to 2 years regardless of interictal EEG abnormalities, which can persist long after seizures have remitted. The recommended antiepileptic drugs are those used in focal epilepsy in children, including oxcarbazepine, carbamazepine, valproate, gabapentin, lamotrigine, and topiramate (Table 403-8). Some patients with reflex seizures may require antiseizure medication, which should be chosen according to seizure type (see Table 403-7).

Choice of Antiepileptic Drugs

The ultimate goal of treatment is to obtain complete freedom from seizures without side effects. Some of the newer antiepileptic drugs (see Table 403-8) are better tolerated and have better pharmacokinetics than older drugs, but there is no robust evidence to support superior efficacy of one drug over another. The choice of medication depends on the type of seizure and epilepsy syndrome (thereby making a correct diagnosis crucial) and the medication's side effects, cost, and ease of use. Specific drugs are effective for specific types of seizures, and some drugs can worsen other types of seizures. Knowledge of individual drugs as they relate to age, sex, comorbid conditions, drug interactions, sedation, tolerance, mood, and withdrawal is critical in the drug selection process (see Table 403-7). For example, ethosuximide and valproic acid are more effective than lamotrigine for the treatment of childhood absence epilepsy.

Drugs that cause enzyme induction (e.g., carbamazepine, phenytoin, phenobarbital, oxcarbazepine, topiramate) or inhibition (e.g., valproic acid) can be difficult to manage when additional medications, such as oral contraceptives, are used for independent conditions. For these clinical settings and in elderly patients, gabapentin and levetiracetam are particularly useful because they have no appreciable drug interactions.

In patients with newly diagnosed focal epilepsy, the underlying cause influences the response to antiepileptic drugs. The likelihood of achieving seizure freedom is higher for patients with vascular malformations, stroke, and tumors (63 to 78%), and lower for patients with hippocampal sclerosis and malformations of cortical development (40 to 50%). Among patients presenting with a new diagnosis of epilepsy, about 65% achieve seizure remission on antiepileptic drug treatment. Of these patients, about 45 to 50% achieve seizure remission with the first antiepileptic drug, 10 to 15% with the second, 1% with the third, and 3% with a combination of two or more antiepileptic drugs. Because the likelihood of achieving subsequent seizure remission is small if two drug trials fail, the 35% or so of patients who fail adequate trials of two antiepileptic drugs are considered to be drug resistant. In these patients, other forms of treatment, including surgery, should be considered. The first consideration in managing apparently drug-resistant patients is to ensure that the diagnosis is correct and the antiepileptic drug is appropriate. Other common causes of a poor response to drugs include poor adherence to antiepileptic drugs, sleep deprivation, alcohol use, fatigue, emotional stress, systemic illnesses, use of concurrent medications, and nonepileptic seizures. After addressing these factors, patients who remain drug resistant should be considered potential candidates for surgical therapy.[9]

Surgical Treatment

Surgical treatment entails resection or disconnection of the cerebral region that contains the seizure focus. Removal of an epileptogenic region requires accurate identification of the region as well as documentation of a lack of functional consequences after its removal. Video EEG monitoring with seizure recording from scalp electrodes, MRI protocols with special attention to areas

TABLE 403-8 CHARACTERISTICS OF MAJOR ANTIEPILEPTIC DRUGS

NAME	TOTAL MILLIGRAMS PER DAY (USUAL SCHEDULE)	THERAPEUTIC RANGE (μg/mL)	PROMINENT SIDE EFFECTS	OTHER EFFECTS	OTHER ISSUES
Carbamazepine	400-1600 (bid)	4-12	Diplopia, fatigue, hyponatremia	Mood stabilizer	Enzyme inducer
Ethosuximide	750-1250 (daily, bid)	40-100	Ataxia, lethargy	Rash, bone marrow suppression	
Gabapentin	600-6000 (tid, qid)	2-12	Fatigue	Treatment of pain	No drug interactions
Lamotrigine	100-600 (bid)	4-18	Insomnia, headache, tremor, anxiety	Mood stabilizer	Risk for Stevens-Johnson syndrome; slow start-up
Lacosamide	200-400 (bid)	Not well established	Dizziness, diplopia, tremor,	Minor prolongation of PR interval	Low risk of drug interaction
Levetiracetam	500-3000 (bid)	3-63	Mood change, irritability, lethargy		No drug interactions
Oxcarbazepine	300-2400 (tid)	6-40	Diplopia, hyponatremia, sedation	Mood stabilizer	
Phenobarbital	60-240 (at bedtime)	15-40	Fatigue, depression, sedation	Joint pain	Enzyme inducer
Phenytoin	200-600 (bid)	10-20	Fatigue, hirsutism, gingival hypertrophy	Treatment of some pain	Enzyme inducer
Topiramate	50-600 (bid)	2-12	Anorexia, weight loss, kidney stones, speech disturbance, distal paresthesias	Headache prophylaxis, mood stabilizer	Enzyme inducer
Valproate	4000 (bid or tid)	50-100	Weight gain, hair loss, tremor	Headache prophylaxis, mood stabilizer	Enzyme inhibitor, parkinsonian effects in elderly patients
Zonisamide	100-600 (at bedtime)	10-40	Anorexia, kidney stones, dizziness, distal paresthesias	Mood stabilizer	

commonly associated with refractory seizures (e.g., the medial temporal and frontal lobes), and functional neuroimaging, including PET and SPECT, are used to make the assessment. In temporal lobe epilepsy, neuropsychological evaluation is essential to localize dysfunction and establish the level of function in the region considered for resection. EEG localization of the region of seizure onset and mapping of brain function may require the surgical implantation of intracranial electrodes for recording and for stimulating cortical tissue. These procedures are performed by multidisciplinary teams in specialized epilepsy centers.

Epilepsy surgery interventions that have been subjected to rigorous randomized trials include temporal lobe resection compared with medical therapy for mesial temporal lobe epilepsy, comparison of different amounts of temporal lobe resection, different intensities of vagus nerve stimulation, and thalamic stimulation compared with medical therapy. The most dramatic surgical effect is seen for temporal lobe resection compared with medical therapy. In one randomized trial, 58% of surgical patients and only 8% of medical patients became seizure free at 1 year. Among these patients, clinically meaningful improvement in quality of life was achieved in 56% of patients treated surgically compared with only 11% of patients treated medically.[9] In another small randomized trial of patients with drug-resistant temporal lobe epilepsy, surgery plus continued antiepileptic medications was successful in eliminating seizures in 11 of 15 patients at 2 years, whereas all medically treated patients continued to have seizures at 2 years.[9] As a result, patients with drug-resistant temporal lobe epilepsy should be evaluated for epilepsy surgery. Nonrandomized studies demonstrate enduring freedom from seizures at 10 years or more after hemispheric disconnection (61%), temporal lobe resection (64%), parieto-occipital resection (46%), and frontal lobe resection (27%). Palliative surgical procedures such as callosotomy and multiple subpial transections have lower success rates and are used when surgical resection of the seizure focus is not possible. In the long term, about 65% of patients undergoing surgery achieve sustained seizure freedom (40 to 50% immediately after surgery, and 15% after a period of initial seizures), 16% have a fluctuating course of relapsing-remitting seizures, and 18% never become seizure free.[10,11] Promising surgical therapies for epilepsy include radiosurgery and various types of electrical stimulation of the brain.[12]

Status Epilepticus

Status epilepticus is a medical emergency in which seizures occur continuously or repeatedly without intervening resumption of consciousness for 30 minutes. However, even 5 minutes of generalized tonic-clonic seizures cause hypoxia, lactic acidosis, muscle breakdown, and neuronal damage. Most episodes of status epilepticus are caused by an acute brain insult in persons without underlying epilepsy, so a cause should be sought promptly. After securing the airway and stabilizing cardiovascular function, immediate intervention with parenteral agents is needed to stop the seizures. In a randomized trial of adults in status epilepticus treated before arriving at the hospital, 10 mg of intramuscular midazolam was more effective and at least as safe as 4 mg of intravenous (IV) lorazepam for stopping seizures.[9] In the emergency department, options include IV lorazepam (0.1 mg/kg given at 2 mg/minute),[9] a continuous IV midazolam (0.1 to 2 mg/kg/hour) followed by IV phenytoin (15 mg/kg at a rate of 50 mg/minute) or fosphenytoin (15-20 mg/kg at a rate of 150 mg/minute) to provide a more long-lasting effect. If seizures continue for 10 to 15 minutes, options include phenobarbital (20 mg/kg IV) or continuous IV midazolam (0.1 to 2 mg/kg/hour), pentobarbital (0.5 to 3 mg/kg/hour), or propofol (2 to 4 mg/kg/hour), most appropriately in an intensive care setting. In refractory cases, general anesthesia for 24 hours is used. In children, convulsive status epilepticus can be controlled within 10 minutes in 70 to 75% of patients treated with either IV diazepam (0.2 mg/kg) or IV lorazepam (0.1 mg/kg).[9]

Considerations in Women

Changes in hormone levels during the menstrual cycle may aggravate seizures perimenstrually in some women (i.e., catamenial epilepsy). The administration of oral contraceptives (Chapter 238), Depo-Provera, acetazolamide (250 to 500 mg/day), or clobazam (10 to 20 mg/day) may reduce perimenstrual seizures. Enzyme-inducing antiepileptic drugs (see Table 403-8) that reduce estrogen levels by enhancing its metabolism require patients to be treated with higher doses of estrogen or alternative methods of contraception.

Pregnancy poses challenges with regard to seizure control, teratogenesis, and outcomes of pregnancy. Nevertheless, pregnancy itself has no consistent effect on the frequency of seizures, and more than 90% of pregnancies in women with epilepsy are safe and successful. Freedom from seizures for at least 9 months preceding pregnancy is associated with a high probability of freedom from seizures during the pregnancy. Serum levels of lamotrigine, phenytoin, carbamazepine, levetiracetam, and oxcarbazepine may change during pregnancy and should be monitored. Valproate carries a higher risk for major congenital malformations and an enduring reduction in cognitive abilities in children exposed to this medication in utero; therefore its use should be avoided during pregnancy if seizure control permits it. Similarly, polytherapy and high doses of antiepileptic drugs should be avoided if possible, but antiepileptic drugs should not be discontinued. There is no increased risk for cesarean section or premature contractions, and epilepsy itself does not increase the risk for cognitive impairment in the child. Supplementation with at least 0.4 mg of folic acid daily should be given before conception and during pregnancy to reduce the risk for neural tube defects.[13]

Discontinuing Antiepileptic Drugs

About 60% of patients have seizures that are easy to control with antiepileptic drugs. Medications may be slowly tapered over 4 to 6 months in patients who have remained free of seizures for 2 years or longer, have had few seizures before treatment started, and who have a normal neurologic examination and EEG. However, the increased absolute risk for recurrent seizures after withdrawal of medication is about 20% (number needed to harm of 5). The consequences of a recurrent seizure, the costs and side effects of drugs, and aspects such as personal preferences influence the decision to withdraw antiepileptic drugs in patients who have been free of seizures.

PROGNOSIS

The prognosis is favorable in the majority of patients who experience either unprovoked seizures or one of the epilepsies.

Prognosis after Febrile Seizures

Febrile seizures are common and usually consist of generalized tonic-clonic seizures. They are provoked by fever and therefore are not considered epilepsy. The seizures begin after 6 months of age and stop before 6 years of age. Usually, febrile seizures are left untreated because the prognosis is benign.[14] When febrile seizures occur in the setting of a neurologic abnormality or are prolonged or complicated, the risk for later epilepsy is increased.

Prognosis after a Single Unprovoked Seizure

The risk of experiencing recurrent seizures after a first unprovoked seizure ranges from 21 to 69% at 2 years and from 34 to 70% at 5 years. The risk is lower in the general population than in hospital-based studies (36% at 1 year and 45% at 2 years). The probability of a relapse decreases with time; about 50% of recurrences occur within 6 months of the initial seizure, and 76 to 96% occur within 2 years. The two most consistent predictors of recurrence are the presence of a neurologic cause for the seizure, which is often uncovered on brain MRI or by the neurologic examination and history, and an epileptiform or slow EEG. The 2-year risk for recurrence is lowest for patients without an identified neurologic cause and with a normal EEG (about 25%), intermediate for patients with an identified neurologic cause or without a cause but with an abnormal EEG (48%), and highest for those with a neurologic cause and an abnormal EEG (about 65%). The risk rises dramatically if more than one seizure has occurred; after a second unprovoked seizure, the risk for a third seizure is 73%, and after a third seizure, the risk for a fourth seizure is 76%.

Prognosis of Epilepsy

The natural history of untreated epilepsy, mostly in developing countries, shows that 30 to 40% of patients obtain 5- to 10-year remissions without treatment. In developed countries, where treatment is generally started after two unprovoked seizures have occurred, the likelihood of 5-year remission is about 60% when patients are followed for 10 years, and about 70% when patients are followed for 20 years. The rate of 5-year remission in children is about 75%. In the long term, sustained freedom from seizures is achieved in about 60% of patients (early remission in about 35 to 40% of patients, and late remission in about 20 to 25%), about 16% of patients fluctuate between relapses and remissions, and about 25% never achieve seizure remission.[15] Epilepsy is considered to be resolved in patients who had an age-dependent epilepsy syndrome and are now past the applicable age, or in patients who have been seizure free for at least 10 years, with no seizure medications for the last 5 years.

Conversely, the duration of active epilepsy before achieving control is one of the most powerful predictors of remission. If seizures remain uncontrolled during the first year after diagnosis, the chance of ever achieving control is only 60%. If the period of uncontrolled seizures extends to 4 years, the chance of ever achieving control is only 10%. The presence of multiple seizure types and frequent generalized tonic-clonic seizures is associated with a lower likelihood of remission. Less than 40% of patients with newly diagnosed mesial temporal lobe epilepsy will be controlled with medications, although familial cases are more easily managed medically.

Children whose seizures remain uncontrolled are at risk of developing cognitive impairment, especially at a younger age, thereby emphasizing the importance of prompt seizure control. In children with absence epilepsy, the 12-month probability of seizure control and remaining on medication is about 35 to 40% overall, but it is higher for ethosuximide (45%) and valproic acid (44%) than for lamotrigine (21%).[A7] In longitudinal population studies of children with newly diagnosed epilepsy, quality of life improves over time in about 50%, remains stable in 30%, and deteriorates in 20%.

Patients with epilepsy are at risk for poor psychosocial outcomes, depression, and increased mortality.[16] The risk for death is two to three times higher in epilepsy than in the general population, and it can be up to five times higher in patients with frequent generalized convulsions and drug-resistant epilepsy.[17] The major causes of death are underlying conditions such as stroke and pneumonia. Sudden unexpected death in epilepsy occurs in 1 per 1000 patient-years and is particularly devastating because it affects young individuals with frequent uncontrolled seizures.

Grade A References

A1. Glauser T, Ben-Menachem E, Bourgeois B, et al. Updated ILAE evidence review of antiepileptic drug efficacy and effectiveness as initial monotherapy for epileptic seizures and syndromes. *Epilepsia*. 2013;54:551-563.

A2. Fiest KM, Sajobi TT, Wiebe S. Epilepsy surgery and meaningful improvements in quality of life: Results from a randomized controlled trial. *Epilepsia*. 2014;55:886-892.

A3. Engel J Jr, McDermott MP, Wiebe S, et al. Early surgical therapy for drug-resistant temporal lobe epilepsy: a randomized trial. *JAMA*. 2012;307:922-930.

A4. Silbergleit R, Durkalski V, Lowenstein D, et al. Intramuscular versus intravenous therapy for prehospital status epilepticus. *N Engl J Med*. 2012;366:591-600.

A5. Prasad M, Krishnan PR, Sequeira R, et al. Anticonvulsant therapy for status epilepticus. *Cochrane Database Syst Rev*. 2014;9:CD003723.

A6. Chamberlain JM, Okada P, Holsti M, et al. Lorazepam vs diazepam for pediatric status epilepticus: a randomized clinical trial. *JAMA*. 2014;311:1652-1660.

A7. Glauser TA, Cnaan A, Shinnar S, et al. Ethosuximide, valproic acid, and lamotrigine in childhood absence epilepsy: initial monotherapy outcomes at 12 months. *Epilepsia*. 2013;54:141-155.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Fisher RS, Acevedo C, Arzimanoglou A, et al. ILAE official report: a practical clinical definition of epilepsy. *Epilepsia*. 2014;55:475-482.
2. Ngugi AK, Kariuki SM, Bottomley C, et al. Incidence of epilepsy: a systematic review and meta-analysis. *Neurology*. 2011;77:1005-1012.
3. Hesdorffer DC, Logroscino G, Benn EKT, et al. Estimating risk for developing epilepsy: a population-based study in Rochester, Minnesota. *Neurology*. 2011;76:23-27.
4. Singh TD, Fugate JE, Hocker SE, et al. Postencephalitic epilepsy: Clinical characteristics and predictors. *Epilepsia*. 2015;56:133-138.
5. Pandolfo M. Pediatric epilepsy genetics. *Curr Opin Neurol*. 2013;26:137-145.
6. Hakami T, McIntosh A, Todaro M, et al. MRI-identified pathology in adults with new-onset seizures. *Neurology*. 2013;81:920-927.
7. Pisani F, Facini C, Pavlidis E, et al. Epilepsy after neonatal seizures: literature review. *Eur J Paediatr Neurol*. 2015;19:6-14.
8. Shinnar S, Bello JA, Chan S, et al. MRI abnormalities following febrile status epilepticus in children: the FEBSTAT study. *Neurology*. 2012;79:871-877.
9. Jette N, Reid AY, Wiebe S. Surgical management of epilepsy. *CMAJ*. 2014;186:997-1004.
10. Englot DJ, Chang EF. Rates and predictors of seizure freedom in resective epilepsy surgery: an update. *Neurosurg Rev*. 2014;37:389-404.
11. Edelvik A, Rydenhag B, Olsson I, et al. Long-term outcomes of epilepsy surgery in Sweden: a national prospective and longitudinal study. *Neurology*. 2013;81:1244-1251.
12. Chambers A, Bowen JM. Electrical stimulation for drug-resistant epilepsy: an evidence-based analysis. *Ont Health Technol Assess Ser*. 2013;13:1-37.
13. Meador KJ, Baker GA, Browning N, et al. Fetal antiepileptic drug exposure and cognitive outcomes at age 6 years (NEAD study): a prospective observational study. *Lancet Neurol*. 2013;12:244-252.
14. Neligan A, Bell GS, Giavasi C, et al. Long-term risk of developing epilepsy after febrile seizures: a prospective cohort study. *Neurology*. 2012;78:1166-1170.
15. Brodie MJ, Barry SJE, Bamagous GA, et al. Patterns of treatment response in newly diagnosed epilepsy. *Neurology*. 2012;78:1548-1554.
16. Fazel S, Wolf A, Langstrom N, et al. Premature mortality in epilepsy and the role of psychiatric comorbidity: a total population study. *Lancet*. 2013;382:1646-1654.
17. Nevalainen O, Ansakorpi H, Simola M, et al. Epilepsy-related clinical characteristics and mortality: a systematic review and meta-analysis. *Neurology*. 2014;83:1968-1977.

REVIEW QUESTIONS

1. Which of the following is correct about the definition of seizures and epilepsy?

A. Seizures and epilepsy are synonymous, therefore any person who has seizures can be diagnosed as having epilepsy.
B. People who develop seizures owing to acute factors such as alcohol withdrawal or metabolic derangements can be diagnosed as having epilepsy if they occur on more than one occasion.
C. Epilepsy is characterized by two or more unprovoked seizures (without apparent cause) occurring more than 24 hours apart.
D. Only specific types of seizures, such as generalized tonic-clonic or absence seizures, qualify for establishing the diagnosis of epilepsy.
E. All seizures should be treated with antiepileptic drugs.

Answer: C Epilepsy is characterized by an enduring predisposition of the brain to generate seizures, and it is defined clinically by two or more unprovoked epileptic seizures of any type occurring more than 24 hours apart. Seizures and epilepsy are not synonymous. Many acute brain insults can produce seizures (known as acute provoked seizures) without denoting the presence of epilepsy. These acute provoked seizures can recur if the acute provoking factor is repeated, they usually respond to treatment of the underlying cause, and they do not require antiepileptic drug treatment.

2. Which of the following is correct about the incidence and prevalence of epilepsy?

A. About 1 in 26 people in developed countries can be expected to develop epilepsy during their lives, and the risk is higher in men than in women.
B. Epilepsy is equally frequent in persons of any socioeconomic status and living in any region of the world.
C. Risk factors for epilepsy can be identified in the majority of people who develop epilepsy.
D. Most people who develop epilepsy have developmental abnormalities of the brain.
E. About 1 in 200 people develop a seizure at some point in their lives.

Answer: A Epilepsy is one of the most common neurological conditions seen in general practice. Population studies in developed countries show that the lifetime risk of developing epilepsy is about 1 in 21 men and 1 in 28 women (1 in 26 overall). The risk is higher in developing countries and for individuals of lower socioeconomic status. Risk factors for epilepsy can be identified in only about 30% of persons with epilepsy, and the most common causes in adults are head injury, infections, stroke, and dementia. About 1 in 10 persons can be expected to have a seizure at some point in their lives.

3. Which of the following is ***not*** correct about the diagnosis of epilepsy?

A. The diagnosis of epilepsy is made by a careful clinical history to determine whether the events in question are seizures and whether 2 or more have occurred more than 24 hours apart.
B. Although an increasing number of genes have been linked to epilepsy, genetic testing of patients with epilepsy is indicated only in selected clinical conditions.
C. Investigations such as electroencephalogram (EEG) and magnetic resonance imaging (MRI) are often necessary to determine the type of epilepsy syndrome and guide treatment.
D. The absence of epileptiform discharges on a routine EEG casts serious doubts on the diagnosis of epilepsy.
E. Long-term video EEG monitoring is helpful to determine the type and number of seizures a patient is experiencing, as well as identify the focus of the origin of the seizure and whether it is amenable to surgical resection.

Answer: D The initial EEG is normal in up to 60% of people with epilepsy, so a normal EEG does not detract from the diagnosis of epilepsy. The diagnosis is established by a careful history comprising three elements: the clinical context, provoking factors, and a detailed description of the event. The EEG and MRI are essential to establish the type of epilepsy and epilepsy syndrome, which in turn guide treatment decisions and prognosis. Only selected conditions merit genetic testing for the purpose of management and counseling. Long-term video EEG is necessary to establish the diagnosis in situations in which the clinical history is ambiguous, to determine the type and number of seizures the patient is having, and especially to identify the seizure focus in the brain and whether it can be safely resected surgically.

4. Which of the following is ***not*** correct about the treatment of epilepsy?

A. About two thirds of patients with epilepsy are well controlled with one antiepileptic drug, but those who fail two antiepileptic drugs have a low chance of being controlled with further drug trials and are considered pharmacoresistant.
B. The type of seizure and epilepsy syndrome determine which antiepileptic drug should be used to treat an individual patient.
C. Antiepileptic drugs with a broad spectrum of efficacy against many types of seizures include valproate and levetiracetam, but some drugs like carbamazepine can exacerbate absence seizures.
D. Valproate carries a high risk of major congenital malformations and poor intellectual development in children exposed to this medication in utero, so it should be avoided during pregnancy if possible.
E. Newer antiepileptic drugs are more effective than older drugs.

Answer: E Although at least a dozen new antiepileptic drugs have been developed in recent years, they are no more effective than older drugs such as phenytoin, valproate, or carbamazepine. For example, in children with absence epilepsy, ethosuximide and valproate are more effective than the newer drug lamotrigine. Valproate, levetiracetam, lamotrigine, and topiramate have a broad spectrum of efficacy against focal and generalized seizures. Fortunately, most patients are well controlled with the first antiepileptic medication if it is appropriately chosen. However, because a lack of response to the first two antiepileptic drugs entails a poor prognosis, these patients are considered drug resistant. Although several drugs have risks of fetal malformations, valproate has the highest risk and also produces enduring deleterious effects on the intellectual development of children exposed in utero. Valproate should be avoided during pregnancy if at all possible. If it must be used, the lowest possible dose is recommended because the teratogenic effect is dose dependent.

5. Which of the following is ***not*** correct about the surgical treatment of epilepsy?

A. Brain surgery for epilepsy is safe and superior to medical therapy in patients with temporal lobe epilepsy.
B. The benefits of surgery are enduring; over 50% of patients can be expected to remain entirely seizure free at 10 years.
C. Only patients who have failed all first-line drugs and still have frequent seizures are eligible for surgery.
D. Before considering surgery, clinicians should ensure that common causes of poor response to antiepileptic drugs, such as lifestyle issues, are addressed and that the patient is adherent to appropriate antiseizure medications.
E. Various types of electrical stimulation of the nervous system, including deep-brain stimulation, may be beneficial for selected patients with drug-resistant epilepsy.

Answer: C Patients who have failed two adequate trials of antiepileptic drugs should be referred for an evaluation of epilepsy surgery. Early surgery, performed within 2 years of developing drug resistance, is vastly superior to medical therapy in patients with temporal lobe epilepsy, with a number needed to treat with surgery, as compared to medical therapy, to achieve seizure freedom of 2. The benefits of surgery persist at 5 and 10 years. About 65% of patients are free of seizures in the long term, as compared with only 8% with medical therapy. Surgery remains underused despite clear evidence of its safety and efficacy. Pseudoresistance to antiepileptic drugs is common and should be investigated thoroughly. The most common causes are an incorrect diagnosis of epilepsy, using the wrong drug, poor adherence, and lifestyle issues that increase the chance of seizures, such as sleep deprivation and excessive alcohol use. Novel surgical strategies include radiosurgery and electrical stimulation of the brain (e.g., thalamus, hippocampus, cerebral cortex) and the peripheral nervous system (e.g., vagus nerve, trigeminal nerve).

404

COMA, VEGETATIVE STATE, AND BRAIN DEATH

JAMES L. BERNAT AND EELCO F.M. WIJDICKS

The assessment and treatment of a comatose patient are among the most challenging activities in clinical medicine. Physicians must systematically and rapidly identify the cause of coma while simultaneously supporting vital systems and taking action to reverse the pathologic process. If coma is caused by a major medical illness, the damage to the brain may be irreversible. For example, resuscitation from cardiac arrest (Chapter 63) is successful only if the brain has not been irreversibly damaged by the hypoxic-ischemic injury. Many patients in acute coma have a hemispheric lesion that causes a mass effect. In such situations the mass effect may need to be reduced medically, or the mass may need to be removed to avoid permanent secondary brain stem injury.

Consciousness is usually considered a global brain function. Focal cerebral hemispheric lesions (Chapters 396 and 406) that alter fragments of consciousness may produce cognitive disturbances such as aphasia, apraxia, or agnosia (Chapter 401). Although language, praxis, and gnosis are elements of normal consciousness, their selective loss does not usually result in a diminution of the quantity of consciousness, so these focal syndromes are not classified as disorders of consciousness.

Disorders of consciousness (Table 404-1) must be distinguished from brain death and locked-in syndrome or other causes of unresponsiveness such as catatonia or psychogenic stupor (Chapter 397). Human consciousness has two measurable clinical dimensions that correspond to two distinct brain neuronal systems: (1) wakefulness, which is the organism's arousal and readiness to respond to internal or external stimuli and which is provided by the reticular system of the rostral brain stem and its thalamic and forebrain ascending projections; and (2) awareness of self and environment, which is provided by a diffuse parallel network of thalamocortical and corticocortical circuits. Wakefulness is a prerequisite for awareness, but as exemplified by patients in a vegetative state, awareness may be lost despite maintained wakefulness.

COMA

Coma is a pathologic state of eyes-closed unresponsiveness in which the patient has neither awareness nor wakefulness and from which the patient cannot be aroused to awareness or wakefulness by vigorous stimuli. Stupor is a similar disorder in which stimuli can temporarily arouse the patient to limited responsiveness, but in the absence of stimuli the patient returns to an unresponsive state. Sleep, by contrast, is a normal state of active cyclic unconsciousness from which subjects can be fully and persistently aroused to full normal consciousness.

TABLE 404-1 COMPARISON OF DISORDERS OF CONSCIOUSNESS*

	AWARENESS	WAKEFULNESS	BRAIN STEM/ RESPIRATORY	MOTOR	EEG	EVOKED POTENTIALS	PET/fMRI	PROGNOSIS
Brain death	Absent	Absent	Absent	Absent	ECS	Absent	Absent cortical metabolism	The person has died
Coma	Absent	Absent	Depressed, variable	Reflex or posturing	Polymorphic delta, burst suppression	BAER variable; cortical ERPs often absent	Resting < 50%	Variable
Vegetative state	Absent	Present, intact sleep-wake cycles	Intact	Reflex, nonpurposeful	Delta, theta, or ECS	BAER preserved; cortical ERPs variable	Resting < 50%; primary areas stimulatable	Poor, when chronic
Minimally conscious state	Intact but poorly responsive	Intact	Intact	Variable with purposeful movements	Nonspecific slowing	BAER preserved; cortical ERPs often preserved	Reduced; secondary areas also stimulatable	Variable
Locked-in syndrome	Intact but communication difficult	Intact	Intact breathing; often brain stem signs	Quadriplegia, pseudobulbar palsy	Usually normal	BAER variable; cortical ERPs normal	Normal or nearly normal	Poor

*The table lists typical findings, which are not necessarily present in all patients. Locked-in syndrome may be mistaken for a disorder of consciousness.
BAER = brain stem auditory evoked response; ECS = electrocerebral silence; EEG = encephalography; ERP = event-related potential; fMRI = functional magnetic resonance imaging; PET = positron emission tomography.
From Bernat JL. *Ethical Issues in Neurology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008:292.

Coma is not a univocal state; it has levels of depth depending on the degree of reflex response to stimulation. Disorders of consciousness comprise a continuum from the mildest state of lethargy to the deepest stage of coma.

EPIDEMIOLOGY

The frequencies of the various causes of coma vary widely depending on the setting. In most settings, however, post-traumatic, metabolic, anoxic, and toxic causes are the most common (Table 404-2).

PATHOBIOLOGY

Wakefulness is provided by a network of neurons and their connections in the central tegmentum of the pons and midbrain (reticular system) that receives input at each level as it ascends into the central basal forebrain, thalamus, and cerebral cortex. Damage to this neuronal network by trauma, ischemia, hypoxia, edema, or metabolic or toxic insults leads to coma because the ascending arousal mechanism is disturbed.

Awareness of self and environment requires not only wakefulness but also normal functioning of massive parallel reverberating neuronal circuits between the thalamus and multiple cortical regions to provide an integrated and unified experience. These structures and their connections can be damaged by the same pathologic conditions that affect the arousal system, but thalamic and cortical neurons are more susceptible to damage because of their higher metabolic demands. A given global brain insult, such as systemic hypoxia and ischemia suffered during cardiac arrest, can selectively damage the cortical and thalamic neurons necessary for awareness while largely sparing the phylogenetically older and less metabolically demanding neurons of the arousal network of the reticular system. This selective damage can result in the vegetative state, which is characterized by wakefulness without awareness.

Coma can be caused by (1) structural damage as a result of brain trauma, edema, inflammation, ischemia, or mass lesions or (2) diffuse metabolic and toxic effects on brain neurons. Structural lesions can affect the arousal neuronal network of the brain stem and basal forebrain directly through local neuronal damage or indirectly by downward or lateral pressure or displacement that causes local ischemia. Metabolic and toxic encephalopathies diffusely affect all brain neurons, particularly the metabolically sensitive cortical and thalamic neurons. However, acute metabolic derangements or toxicities also can cause structural brain injury by altering blood pressure or oxygenation (e.g., opioid toxicity [Chapter 34]), brain edema (e.g., acute liver failure [Chapter 154]), or acute demyelination (e.g., from too rapid correction of chronic hyponatremia [Chapter 116]).

Structural lesions that cause coma typically produce clinically recognizable syndromes of cerebral "herniation" in which intracranial pressure shifts produce caudal displacement and ischemia of the midbrain and medial temporal lobe through the tentorial incisura that induces dysfunction of cranial nerves, breathing, and motor systems. Central transtentorial herniation from slowly expanding axial lesions is uncommon; more common is uncal herniation from rapidly expanding and laterally placed lesions that trap the ipsilateral oculomotor nerve against the uncus of the temporal lobe. Lateral displacement of brain structures can supplement or exceed downward displacement. The ascending arousal system also can be damaged directly by primary brain stem catastrophes such as pontomesencephalic hemorrhage and infarction, or indirectly by downward-directed pressure waves produced by hemispheric mass lesions such as from brain trauma (Chapter 399) or supratentorial neoplasms (Chapter 189), abscesses (Chapter 413), hemorrhages (Chapter 408), or large infarctions (Chapter 407).

TABLE 404-2 CAUSES OF STUPOR AND COMA

- Traumatic brain injury*
  - Contusion
  - Intracerebral, epidural, subdural, or subarachnoid hemorrhage
  - Raised intracranial pressure
- Neoplasms and other mass lesions
- Infections
  - Meningitis
  - Encephalitis
  - Brain abscess or empyema
  - Sepsis or other infection, especially in the elderly or a demented patient*
- Cerebrovascular disease
  - Subarachnoid hemorrhage
  - Infarction in the brain stem or cerebellum or large hemispheric infarction
  - Hemorrhage in the brain stem or cerebellum or large hemispheric hemorrhage
  - Vasculitis, disseminated intravascular coagulation, thrombotic thrombocytopenic purpura
- Seizures
  - Status epilepticus
  - Spike-wave stupor
  - Postictal state
- Metabolic encephalopathies*
  - Hypoglycemia, hyperglycemia
  - Hypercalcemia
  - Hyponatremia, hypernatremia
  - Hypoxemia, including anoxia after cardiac arrest
  - Acidosis
  - Organ system failure: hepatic, renal, pulmonary, cardiac
  - Endocrinopathy (e.g., myxedema coma)
- Toxic encephalopathies
  - Drug intoxications*: alcohol, barbiturates, benzodiazepines, opioids, stimulants, salicylates, anticonvulsants, anticholinergics, psychotropic drugs, or others
  - Poisoning: carbon monoxide, industrial toxins
- Other encephalopathies
  - Hypertensive encephalopathy
  - Acute hydrocephalus
  - Pituitary apoplexy
- Other
  - Conversion, malingering, catatonia

*Most common causes.

Metabolic encephalopathies disturb the neuronal microenvironment by altering the precise metabolic conditions necessary for normal neuronal excitability. Disturbances in the neuronal milieu can be caused by alterations in blood flow, oxygen delivery, glucose concentration, temperature, electrolyte concentrations, and intracranial pressure, as well as by meningitis, seizures, and organ failure. The depth of the resulting alteration of consciousness depends on the severity of the metabolic disturbance: mild metabolic encephalopathies can cause slowness or lethargy, whereas severe metabolic encephalopathies can produce deep coma. The rapidity of onset is of particular importance. A sudden drop in the serum sodium concentration (Chapter 116) may result in coma and seizures, whereas a slow decline to an equivalent level may not. Toxic encephalopathies can be caused by poisoning with exogenous agents such as depressant drugs or by endogenous toxins resulting, for example, from renal or hepatic failure and produce the same continuum of severity. Acute meningeal inflammation, caused most commonly by bacterial meningitis, induces coma by a combination of inflammatory and vascular changes.

CLINICAL MANIFESTATIONS

A comatose patient is unresponsive and cannot be aroused to awareness or wakefulness. The level of consciousness can be assessed by loudly speaking the patient's name directly in the ear. Patients should be asked to look up and down to test for locked-in syndrome, in which vertical eye movements may be the only remaining voluntary movement. Noxious stimuli can be used to elicit motor responses. Stimulation of nasal hair and the nasal septum with a cotton-tipped swab may elicit airway protective reflexes. Acceptable examples of nontraumatic noxious stimuli that can elicit a rapid response if present include compression of the supraorbital nerve, temporomandibular joints, or nail beds, or a sternal rub with fingers or knuckles. Motor responses to these stimuli usually can be graded as localization, withdrawal, reflex extensor posturing, and none.

DIAGNOSIS

The diagnosis of coma requires a detailed history, physical, and neurologic examination (Table 404-3), laboratory tests, and neuroimaging studies.[1] Immediate attention should be paid to whether the patient has signs of meningitis (e.g., fever, nuchal rigidity) or head trauma (Chapter 399) or has focal findings suggestive of a mass, lesion, bleeding, or ischemic injury.

TABLE 404-3 SOME INITIAL CLINICAL CLUES TO THE DIAGNOSIS OF STUPOR AND COMA

STRUCTURAL CAUSES
History
Abrupt onset of unconsciousness
Sudden headache
Vomiting
Examination
Focal neurologic signs (hemiparesis, posturing, asymmetrical reflexes)
Abnormal pupillary light reflexes
METABOLIC OR TOXIC CAUSES
History
Gradual onset of unconsciousness
Preceding confusional state
Seizures
Known cognitive impairment
Taking insulin or street drugs
Examination
Absence of focal neurologic signs
Presence of frontal release signs
Intact pupillary light reflexes
Tremor, asterixis, or multifocal myoclonus
Evidence of systemic infection
Needle tracks
MENINGITIS
History
Worsening headache
Neck stiffness and pain
Fever, chills
Progressive stupor and coma
Examination
Fever, rigors
Nuchal rigidity and signs of meningeal inflammation

Coma assessment scales are useful to describe the depth of coma, serially assess changes, and estimate prognosis. The widely used Glasgow Coma Scale (Chapter 399, Table 399-1) was devised to assess patients with traumatic brain injury and is a combination of three responses. The FOUR Score (Table 404-4) is more useful for all causes of coma, because it more accurately assesses brain stem function and quantifies awareness.

The relevant history includes eyewitness accounts of any preceding headache, vomiting, confusional state, prescription and street drug use, alcohol consumption, diabetes, fever, head trauma, seizure activity, and medical illnesses, especially atrial fibrillation. The rate at which neurologic function declined can be especially helpful.

After determining the level of consciousness, the neurologic examination focuses on four systems whose careful assessment can distinguish structural from metabolic causes of coma and delineate the functional brain level caused by the pathologic process: (1) the respiratory rate and pattern; (2) the pupils' size, shape, and reactivity; (3) spontaneous eye movements and elicited vestibuloocular reflexes; and (4) motor responses to stimuli (Table 404-5).

The respiratory rate and pattern should be observed. Cheyne-Stokes respiration is a periodic form of breathing whose amplitude forms a sine wave, with 5- to 45-second periods of apnea punctuating periods of hyperpnea; it is seen in metabolic encephalopathies, especially those caused by heart failure, and during sleep. Central neurogenic hyperventilation, which is continuous hyperpnea and tachypnea that produces a pure respiratory alkalosis, occurs with lesions of the rostral brain stem tegmentum at the midbrain level; rapid deep breathing (Kussmaul) that is compensating for a severe metabolic acidosis (Chapter 118) looks similar. Irregular breathing patterns with apneic periods may indicate severe brain stem involvement and can be agonal.

Pupillary size and reactivity to a bright light stimulus can be assessed to evaluate the integrity of the optic and oculomotor nerves, midbrain, and sympathetic nerves. The reactivity of the pupils to light is an important sign that discriminates structural coma from metabolic-toxic coma. The pupils usually remain reactive through varying depths of metabolic-toxic coma, often until apnea ensues, whereas pupillary reflexes are lost earlier in structural coma caused by transtentorial herniation. The pupils are small, equal, and reactive in patients with metabolic encephalopathies. When the oculomotor nerve or the midbrain is involved, the ipsilateral pupil becomes unreactive to light because of damage to the parasympathetic pupilloconstrictors and dilates because of the unopposed sympathetic pupillodilators. When herniation proceeds further, the brain stem sympathetic tracts are also damaged so the affected pupil(s) returns to midposition and becomes unreactive to light or dark. Lesions that affect only the pons and not the midbrain (e.g., pontine hemorrhage, infarction) can cause pinpoint pupils whose intact

TABLE 404-4 FOUR SCORE COMA ASSESSMENT SCALE*

EYE RESPONSE
E4 = Eyelids open or unopened, tracking or blinking to command
E3 = Eyelids open but not tracking
E2 = Eyelids closed but open to pain
E1 = Eyelids remain closed with pain stimuli
MOTOR RESPONSE
M4 = Thumbs up, fist, or peace sign
M3 = Localizing to pain
M2 = Flexion response to pain
M1 = Extension response to pain
M0 = No response to pain or generalized myoclonic status epilepticus
BRAIN STEM REFLEXES
B4 = Pupillary and corneal reflexes present
B3 = One pupil dilated and unreactive to light
B2 = Pupillary or corneal reflexes absent
B1 = Pupillary and corneal reflexes absent
B0 = Absent pupillary, corneal, or cough reflexes
RESPIRATION
R4 = Regular breathing pattern
R3 = Cheyne-Stokes breathing pattern
R2 = Irregular breathing pattern
R1 = Triggers or breathes above the ventilator rate
R0 = Apnea or breathes at the ventilator rate

*For nontraumatic coma and other disorders of consciousness.
From Wijdicks EFM. *The Comatose Patient.* 2ed. New York: Oxford University Press; 2014.

TABLE 404-5 BRAIN FUNCTIONAL LEVELS DETERMINED BY FINDINGS IN CLINICAL SYSTEMS

FUNCTIONAL LEVEL	CONSCIOUSNESS	RESPIRATION	PUPILS	VESTIBULO-OCULAR REFLEXES	MOTOR RESPONSES
CENTRAL TRANSTENTORIAL HERNIATION					
High diencephalic	Light stupor	Eupnea, yawning, post-hyperventilation apnea	Small, reactive	Loss of checking component	Paratonia, grasp
Low diencephalic	Deep stupor	Cheyne-Stokes	Small, reactive	Loss of checking component	Decorticate posturing
Midbrain	Coma	Central neurogenic hyperventilation	Midposition, fixed	Loss of medial rectus function	Decerebrate posturing
Upper pons	Coma	Central neurogenic hyperventilation	Midposition, fixed	Loss of medial rectus function	Decerebrate posturing
Lower pons	Coma	Ataxic	Midposition, fixed	Absent	Flaccid
Medulla	Coma	Apnea	Midposition, fixed	Absent	Flaccid
UNCAL TRANSTENTORIAL HERNIATION					
Early third nerve	Unreliable	Normal	Ipsilateral dilated, fixed	Normal	Contralateral hemiparesis
Late third nerve	Coma	Cheyne-Stokes or central neurogenic hyperventilation	Ipsilateral dilated, fixed; contralateral dilated, fixed	Medial rectus dysfunction	Ipsilateral hemiparesis and contralateral decerebrate posturing
Midbrain-pons	Coma	Central neurogenic hyperventilation or ataxic	Midposition, fixed	Absent	Bilateral decerebrate posturing

reaction to light can be seen with a magnifying glass. Preexisting disease (e.g., diabetes) or locally applied eye medications can also impair pupillary reflexes.

Spontaneous eye movements may have localizing value. Conjugate horizontal eye deviation points to the side of brain lesions rostral to the brain stem (usually in the cerebral hemispheres) but to the side opposite brain stem lesions. Tonic downward eye deviation suggests acute lesions of the thalamus or dorsal midbrain. Tonic upward eye deviation is unusual but is seen in patients with hypoxic-ischemic lesions. Ocular bobbing with a rapid downward movement followed by a slow return upward suggests a pontine lesion. Reverse ocular bobbing with a slow downward and rapid upward movement ("ocular dipping") has poor localizing value but may be seen after hypoxic-ischemic insults and metabolic disorders. "Ping-pong" gaze with alternating conjugate horizontal eye movements is nonspecific, but a slower and otherwise similar disorder called periodic alternating gaze is seen in patients with portosystemic encephalopathy (Chapter 153). Ocular skew deviation, in which one eye is higher than the other on primary gaze, suggests a brain stem lesion.

The vestibuloocular reflex assesses brain stem and cerebral hemispheric function by reflexively inducing eye movements. First, the external auditory canal should be inspected to exclude perforation of the tympanic membrane and obstruction by cerumen. Ice water is then injected into the canal (10 mL for usual assessment but 50 mL for assessment of brain death), and the induced reflex eye movements are observed (see Table 404-5). In patients with normal consciousness, such as in psychogenic coma, marked horizontal nystagmus is produced. In patients with stupor at a diencephalic level, such as from metabolic encephalopathy, the fast component of nystagmus is suppressed, so the patient responds with full tonic conjugate eye movements toward the injected ear. With lesions of the oculomotor or abducens nerves or lesions of the midbrain or pons, ophthalmoplegia of localizing value is observed. In brain death or total brain stem failure, no response is observed. The vestibuloocular reflexes may be ablated after treatment with ototoxic antibiotics.

Motor responses are observed after noxious stimulation. At lower functional levels of structural coma, limb posturing is often observed. Limb posturing is a unilateral or bilateral, stereotyped, tonic brain stem reflex movement induced by stimulation, especially noxious stimuli. Decorticate posturing, in which the arm is flexed and the ipsilateral leg is extended, suggests a midbrain functional level. Decerebrate posturing, in which both the arm and the ipsilateral leg are extended, suggests a pontine functional level. When the entire brain stem is destroyed, as in brain death, all limbs remain flaccid during stimulation. Metabolic-toxic encephalopathies usually produce symmetrical motor signs, whereas structural causes of coma frequently produce asymmetrical motor signs. Hypoglycemia and acute hyponatremia are exceptions in which aphasia, gaze paresis, and hemiparesis may be seen. Myoclonic seizures with continuous or intermittent rhythmic clonic movements frequently develop in patients who have suffered hypoxic-ischemic neuronal damage during cardiopulmonary arrest. Myoclonic status epilepticus is a poor prognostic sign.

On the general physical examination, assessment of vital signs, otoscopy, optic funduscopy, and inspection for head trauma, nuchal rigidity, and needle tracks can provide key findings. Emergency laboratory testing should generally include a complete blood cell count, serum electrolytes, a blood glucose level, tests of renal and liver function, coagulation tests, thyroid function tests, arterial blood gas analysis, a blood alcohol concentration, a urine drug screen, and an electrocardiogram. If intoxication is likely, particularly in the absence of ketones, uremia, or an abnormal lactate level, the anion gap (Chapter 118) and osmolar gap should be measured. Routine toxicologic testing rarely changes acute management and usually adds little to the diagnostic evaluation, except in patients with a normal CT scan and no lateralizing signs or physical examination.

If signs of meningitis are present, blood cultures and a lumbar puncture should be performed without the delay of obtaining brain imaging if no marked focal signs (e.g., hemiparesis) are present. Brain computed tomography (CT) should be performed urgently in nearly every patient in coma. When a CT scan is normal in a comatose patient with no clear lateralizing neurologic signs, the coma is usually the result of intoxication. Pesticides (Chapter 110), ethanol, atypical alcohols, opioids (including heroin [Chapter 34]), and benzodiazepine intoxication (Chapter 397) should be considered.

After the neurologic examination, screening laboratory tests, brain CT, and lumbar puncture have been accomplished, a tentative diagnosis can be made in most patients. If focal signs are present despite normal findings on CT, consideration should be given to an acute stroke involving the posterior circulation (Chapters 407 and 408). Brain CT angiography or magnetic resonance imaging (MRI) can clarify the diagnosis. If lateralizing signs are absent, metabolic and toxic causes are most likely. An electroencephalogram (EEG) is useful in patients in whom nonconvulsive seizure activity may be causing stupor or coma.

TREATMENT

Rx

Management of coma requires simultaneous diagnostic, supportive, and treatment measures (Table 404-6). Specific treatments, which depend on the causative diagnosis, include urgent attention to any head trauma (Chapter 399). Emergency stabilization of respiration and circulation and control of seizures are critical for all patients. In patients without focal findings or obvious meningitis, 50% dextrose (25 g intravenously [IV]), thiamine (50 mg IV), naloxone (0.4 to 2 mg IV), and flumazenil (0.2 mg IV) can be administered during the diagnostic assessment. If fever, nuchal rigidity, or leukocytosis is present, the patient should be treated presumptively for bacterial meningitis (Chapter 412) with IV antibiotics before neuroimaging and lumbar puncture.

Elevated intracranial pressure must be lowered urgently. Treatments include hyperventilation by bag or ventilator, IV hyperosmolar agents such as mannitol, and IV glucocorticoid drugs for patients with vasogenic edema from brain tumors (Chapter 189), abscesses (Chapter 413), or bacterial meningitis (Chapter 412). Therapeutically induced mild hypothermia for several days improves outcomes in patients who are in coma after diffuse hypoxic-ischemic neuronal

TABLE 404-6 EMERGENCY MANAGEMENT OF COMATOSE PATIENTS

1. Ensure oxygenation
2. Maintain the circulation
3. Administer 50% dextrose, 25 g IV, and control glucose
4. Lower raised intracranial pressure
5. Stop seizures with lorazepam, 1-2 mg IV
6. Search for and treat infections
7. Restore acid-base and electrolyte balance
8. Normalize body temperature
9. Administer thiamine, 50 mg IV, and multivitamins
10. Consider administration of opioid antagonists (naloxone, 0.4-2 mg IV)
11. Consider administration of benzodiazepine antagonists (flumazenil, 0.2 mg IV)
12. Control agitation
13. Protect the eyes
14. Consider inducing therapeutic hypothermia for diffuse hypoxic-ischemic causes

Modified from Posner JB, Saper CB, Schiff ND, et al. *Plum and Posner's Diagnosis of Stupor and Coma*. 4th ed. New York: Oxford University Press; 2007:311.

TABLE 404-7 DIAGNOSIS OF THE VEGETATIVE STATE

I. Absence of:
- Awareness of self or environment
- Purposeful or voluntary behavioral response to all stimuli
- Language comprehension or expression

II. Presence of:
- Intermittent wakefulness manifested by the presence of sleep-wake cycles
- Autonomic functions
- Cranial nerve and spinal reflexes

III. Potential behavioral repertoire:
- Breathe spontaneously
- Spontaneous roving eye movements
- Utter sounds but no words
- Grimace to pain, make facial expressions
- Yawn, make chewing jaw movements, swallow saliva
- Move limbs nonpurposefully, arch back, decorticate limb posturing
- Flexion withdrawal from noxious stimuli
- Move head or eyes briefly toward sound or movement
- Auditory startle

damage caused by cardiac arrest,[1] and a temperature target of 36° C is as good as 33° C.[1] Neurosurgical removal of expanding mass lesions and unroofing the skull (hemicraniectomy) can be life-saving in selected patients.[1]

PROGNOSIS

The prognosis of coma is highly variable[2] and depends on the cause, stage, degree of structural brain damage, and potential reversibility. The majority of surviving patients who undergo therapeutic hypothermia after cardiac arrest have preserved cognitive function and are able to return to work.[3] Prediction rules for recovery apply only to a specific cause. The prognosis after traumatic brain injury can be predicted by the Glasgow Coma Score (Chapter 399, Table 399-1). In patients who have survived cardiopulmonary arrest and resuscitation and in whom toxic and metabolic factors (e.g., sedation, neuromuscular blockade, hypothermia, organ failure, and shock) are not present, the likelihood for recovery of awareness is less than 1% if the following signs are present:

Day 1: presence of myoclonic status epilepticus
Days 1 to 3: bilateral absence of the N20 response of the somatosensory evoked potential
Days 1 to 3: serum neuron-specific enolase concentration higher than 33 µg/L
Day 3: absent pupillary or corneal reflexes; extensor or absent motor responses

If the patient remains comatose on day 3 without these findings and without a contribution from a potentially reversible metabolic or toxic encephalopathy, the probability of recovery of awareness is below 10% if there is no withdrawal response to painful stimuli and below 40% if the patient withdraws to painful stimuli but lacks spontaneous eye opening.

Some patients may appear to recover completely within days after an episode of impaired cerebral oxygenation but then regress days to weeks later with the syndrome of delayed posthypoxic leukoencephalopathy, which is caused by demyelination. Neuroimaging typically shows diffuse hemispheric demyelination that spares the cerebellum and brain stem. The cerebrospinal fluid may show an elevated myelin basic protein level, and biopsy can confirm leukoencephalopathy. Some patients slowly recover over 3 to 12 months, but usually with substantial neurologic sequelae.[4] Others may linger in a vegetative or minimally conscious state (see below).

THE VEGETATIVE STATE

The vegetative state is a disorder of consciousness in which wakefulness is retained but awareness of self and environment is entirely absent to the extent that it can be tested clinically.[5] The vegetative state may be a transient stage during spontaneous recovery from coma to awareness, or it may be a chronic unchanging state. Adjectives such as "persistent" or "permanent" should be avoided because they generate confusion by confounding the diagnosis and prognosis.

EPIDEMIOLOGY AND PATHOBIOLOGY

The vegetative state is caused by diffuse or multifocal brain lesions that disconnect the polymodal cerebral cortices from the thalami but spare the brain stem and hypothalamus. The prevalence of a transient vegetative state after brain injury is unknown. The prevalence of a chronic stable vegetative state is 19 per million.

Causative lesions can be located bilaterally in the thalami, diffusely in the cerebral cortex, or diffusely in the white matter that connects the thalami to the cortex. Two clinical disorders are most commonly responsible: diffuse hypoxic-ischemic neuronal damage to the thalami and cortex suffered during cardiopulmonary arrest and diffuse axonal injury from a traumatic injury caused by a torque force. These two disorders have different pathologies: hypoxic-ischemic injury affects cortical, thalamic, and cerebellar neurons, whereas diffuse axonal injury shears and disconnects the axons at the gray matter–white matter junction diffusely or multifocally in the cortex.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The clinical features of the vegetative state (Table 404-7) are dominated by what patients do. A careful neurologic examination must be performed to search for any evidence of awareness, because up to 40% of patients in whom a vegetative state is initially diagnosed are actually in a minimally conscious state (see later).

The vegetative state is a clinical syndrome with a spectrum of severity. The typical patient has diffuse slow-wave activity on the EEG, but the most severely affected patients have isoelectric EEGs. However, some patients who are entirely unresponsive generate appropriate EEG responses to distinct commands, thereby suggesting they have residual cognitive function and conscious awareness. As determined by functional neuroimaging studies, a subset of patients with a clinically diagnosed vegetative state possess awareness as evidenced by their ability to perform ideational tasks on command.[6] Such functional neuroimaging tests are not yet part of routine clinical practice but may become standard clinical assessment tools for patients who are suspected to be in a vegetative state.

TREATMENT AND PROGNOSIS

Rx

No treatments reverse or improve a long-standing stable vegetative state. The aggressiveness of medical treatment of patients in the vegetative state should ideally be guided by their previously stated wishes. Patients require the same medical and nursing care, physical therapy, and nutritional needs as patients in coma. Patients should be referred to specialized neurorehabilitation units when possible.

People who recover after a prolonged anoxic vegetative state usually have an initially preserved pupillary light reflex and nociceptive response, paroxysmal sympathetic hyperactivity, and median nerve somatosensory evoked potentials.[7] Patients who are in a vegetative state from nontraumatic causes and who do not regain awareness within 3 months of the insult have less than a 1% chance of experiencing significant neurologic recovery and often raise serious ethical and moral dilemmas for their families and caregivers.[8] After traumatic brain injury, the prognosis cannot be estimated with a similar degree of certainty until after 1 year, although functional neuroimaging is a promising approach to identify patients who are destined to recover awareness.

TABLE 404-8 DIAGNOSIS OF THE MINIMALLY CONSCIOUS STATE

Globally impaired responsiveness
Limited but discernable evidence of awareness of self and environment as demonstrated by the presence of one or more of the following behaviors that occur in a contingent relationship to relevant environmental stimuli and are not simply reflexive movements: Follow simple commands Gesture yes/no answers Make intelligible vocalizations or gestures in direct response to a question's linguistic content Reach for objects that demonstrates a clear relationship between object location and direction of reach Touch and hold objects in a manner that accommodates the size and shape of the object Sustain visual pursuit to moving stimuli Smile or cry appropriately to linguistic or visual content of emotional but not to affectively neutral topics or stimuli

THE MINIMALLY CONSCIOUS STATE

The minimally conscious state (Table 404-8) is a disorder of altered consciousness characterized by a profound lack of responsiveness but with partial or intermittent evidence of awareness of self and environment. Patients typically may have suffered less severe injuries than patients in the vegetative state. The minimally conscious state is much more common than the vegetative state, from which it must be distinguished. When compared with patients in the vegetative state, patients in the minimally conscious state are more likely to respond to environmental and sensory stimuli and to stimulant medications such as levodopa or dopamine agonists (which stimulate thalamic dopaminergic neurons and are prescribed in the same dose ranges as for the treatment of Parkinson disease [Chapter 409]). In patients who are minimally conscious or in a vegetative state at 4 to 16 weeks after traumatic brain injury, amantadine (starting at 100 mg twice daily and increasing up to 200 mg twice daily) can accelerate the early pace of rehabilitative recovery but may not improve ultimate recovery.[A4] Patients in the minimally conscious state require the same specialized neurorehabilitation services as those in the vegetative state. There are no good prognostic data for the minimally conscious state other than for recovery of the subset of patients after traumatic brain injury (Chapter 399).

THE LOCKED-IN SYNDROME

The locked-in syndrome, a state of profound paralysis, is not a disorder of consciousness but may be mistaken for one. In its classic form, it is produced when a large infarction or hemorrhage in the pontine tegmentum and base produces quadriplegia, pseudobulbar palsy, and paralysis of horizontal eye movements. Once the acute encephalopathy resolves, locked-in patients usually remain awake and alert, breathe spontaneously, and have normal consciousness and cognition, to the extent that they can be tested accurately. Inexperienced examiners may incorrectly diagnose locked-in patients as being comatose because of their profound paralysis, pinpoint pupils, and seeming unresponsiveness. A similar state of profound global paralysis with intact cognition can be produced by advanced amyotrophic lateral sclerosis (Chapter 419), Guillain-Barré syndrome (Chapter 420), or critical illness polyneuropathy (Chapter 420).

Patients can be taught to communicate with voluntary vertical eye movements and eyelid movements, which are typically their only retained volitional movements, because they are controlled rostral to the pons. Most affected patients, particularly older patients with comorbid illnesses, die within a few months, but some otherwise healthy young patients who have become locked in as a result of basilar artery occlusion have survived for many years. Occasional patients may recover function to become independent. Computerized systems targeting remaining voluntary eye movements can help patients communicate.

BRAIN DEATH

Brain death is the term popularly applied to the determination of human death based on tests that show irreversible cessation of all clinical brain functions. Once illness or injury has destroyed the brain or rendered its clinical functions irreversibly lost, a human being is dead.[9] Brain death is a medically and legally accepted determination of human death throughout North America, Europe, Australia, most of the developed world, and much of the developing world.[10] Brain-dead patients serve as ideal multiorgan donors.

TABLE 404-9 TESTS FOR BRAIN DEATH IN ADULTS

I. Preconditions showing irreversibility: all necessary • Presence of a structural brain lesion sufficient to produce all the clinical signs • Absence of reversible significant toxic or metabolic encephalopathy: No depressant drug intoxication No neuromuscular blockade (use electroneurography if uncertain) No severe hypothermia No severe hypotension • Sequential repeated testing or one test followed by a confirmatory blood flow test
II. Signs showing complete cessation of all clinical brain functions: all necessary • Coma: no spontaneous movements, no response to any stimuli, and no reflex movements integrated by the brain • Apnea: no breathing or respiratory effort when the $P_{ACO_2} \geq 60$ mm Hg while protecting the P_{AO_2} • Brain stem areflexia: all necessary Absent pupillary light and dark reflexes Absent corneal touch reflexes Absent facial movement to noxious stimuli Absent vestibulo-ocular reflexes tested by caloric irrigation of the external auditory canal with 50 mL of ice water Absent pharyngeal and tracheal reflexes to endotracheal tube suctioning
III. Confirmatory tests: optional but desirable; neuroimaging preferred • Neuroimaging that shows complete absence of intracranial blood flow; one test Intravenous radionuclide angiography Transcranial Doppler ultrasound Computed tomographic angiography Magnetic resonance angiography or diffusion Single-photon emission computed tomography • Electrophysiologic testing (use only if intracranial pressure is not elevated) Electroencephalography + brain stem auditory evoked responses + somatosensory evoked responses: all isoelectric

EPIDEMIOLOGY AND PATHOBIOLOGY

Most cases of brain death result from massive traumatic brain injury (Chapter 399), intracranial hemorrhage (Chapter 408), meningitis (Chapter 412), or diffuse hypoxic-ischemic neuronal damage as a result of cardiac arrest (Chapter 63) or asphyxia. Marked cerebral edema from the primary injury or illness produces severe intracranial hypertension. When intracranial pressure exceeds mean arterial blood pressure (or systolic blood pressure in some cases), intracranial blood flow ceases and widespread ischemic death of brain neurons ensues.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Brain-dead patients have no brain functions measurable at the bedside. The diagnosis should be suspected in any patient who is deeply comatose, is unresponsive to stimuli, has absent pupillary light reflexes, and is apneic and completely ventilator dependent. The diagnosis requires a comprehensive evaluation (Table 404-9). In the United States, one examination typically suffices, but diagnostic criteria in several states and many countries require two examinations, and some require two physicians to certify the findings.

TREATMENT

Rx

Once the diagnosis has been made, the patient is declared dead. If the family has agreed to allow the deceased patient to serve as an organ donor, the ventilator is reattached following the apnea test, and the patient is moved to the surgical suite. If the patient is not an organ donor, the ventilator is not reattached, and all lines and monitors are discontinued. Physicians should be knowledgeable about local laws that may restrict making the diagnosis in patients belonging to certain religious groups that do not accept brain death as human death.

Grade A References

A1. Arrich J, Holzer M, Havel C, et al. Hypothermia for neuroprotection in adults after cardiopulmonary resuscitation. *Cochrane Database Syst Rev*. 2012;9:CD004128.

A2. Nielsen N, Wetterslev J, Cronberg T, et al. Targeted temperature management at 33 degrees C versus 36 degrees C after cardiac arrest. *N Engl J Med*. 2013;369:2197-2206.

A3. Jüttler E, Unterberg A, Woitzik J, et al. Hemicraniectomy in older patients with extensive middle-cerebral-artery stroke. *N Engl J Med*. 2014;370:1091-1100.

A4. Giacino JT, Whyte J, Bagiella E, et al. Placebo-controlled trial of amantadine for severe traumatic brain injury. *N Engl J Med*. 2012;366:819-826.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Edlow JA, Rabinstein A, Traub SJ, et al. Diagnosis of reversible causes of coma. *Lancet*. 2014;384:2064-2076.
2. Luce JM. Chronic disorders of consciousness following coma: Part one: medical issues. *Chest*. 2013;144:1381-1387.
3. Fugate JE, Moore SA, Knopman DS, et al. Cognitive outcomes of patients undergoing therapeutic hypothermia after cardiac arrest. *Neurology*. 2013;81:40-45.
4. Shprecher D, Mehta L. The syndrome of delayed post-hypoxic leukoencephalopathy. *Neurorehabilitation*. 2010;26:65-72.
5. Giacino JT, Fins JJ, Laureys S, et al. Disorders of consciousness after acquired brain injury: the state of the science. *Nat Rev Neurol*. 2014;10:99-114.
6. Stender J, Gosseries O, Bruno MA, et al. Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. *Lancet*. 2014;384:514-522.
7. Estraneo A, Moretta P, Loreto V, et al. Predictors of recovery of responsiveness in prolonged anoxic vegetative state. *Neurology*. 2013;80:464-470.
8. Luce JM. Chronic disorders of consciousness following coma: Part two: ethical, legal, and social issues. *Chest*. 2013;144:1388-1393.
9. Bernat JL. Controversies in defining and determining death in critical care. *Nat Rev Neurol*. 2013;9:164-173.
10. Wijdicks EFM, Varelas PN, Gronset GS, et al. Practice parameter update: determining brain death in adults: report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*. 2010;74:1911-1918.

REVIEW QUESTIONS

1. What clinical sign differentiates between a minimally conscious state and a vegetative state?

A: Blinking to threat
B: Marked dysautonomia
C: Abnormal brain stem reflexes
D: Myoclonus status
E: Tracking an examiner's hand

Answer: E The only sign that, when found consistently, indicates awareness is the ability to track an object. All other signs can be seen in both conditions.

2. A 70-year-old woman was found snoring and unresponsive by her husband. On arrival to the emergency department, she had marked anisocoria, ocular bobbing, extensor posturing, and irregular breathing with pooling secretions requiring intubation. The vital signs were normal, and she was afebrile. Her computed tomographic (CT) scan was normal. Routine laboratory tests including glucose and blood gases are unrevealing. What is a likely diagnosis?

A. Embolus to the basilar artery
B. Intoxication/poisoning
C. Status epilepticus
D. Meningoencephalitis
E. Anoxic-ischemic encephalopathy

Answer: A Ocular bobbing, anisocoria, and extensor posturing points to a brain stem lesion. The computed tomographic (CT) scan can be normal in the early hours after presentation. An embolus to the basilar artery may be seen as a hyperdense CT lesion, but a CT angiogram is a better study. Intoxication is unlikely due to localizing findings. Anoxic-ischemic encephalopathy usually spares the brain stem. Meningoencephalitis would be highly unlikely in a nonfebrile, acutely comatose patient with new brain stem findings.

3. What observation precludes brain death examination?

A. Core temperature of 36° C
B. Systolic blood pressure of 100 mm Hg
C. Preserved tendon reflexes and Babinski signs
D. Leg withdrawal after noxious stimulus
E. Extensor posturing

Answer: E Extensor posturing indicates preserved brain stem function. Generally, blood pressures above 100 mm Hg and core temperatures above 35° C do not influence the neurologic examination. Tendon reflexes, Babinski signs, and leg withdrawal (triple flexion) response can be generated at a spinal level without cortical input and can be preserved in brain dead individuals.

4. A locked-in syndrome is characterized by all the following signs ***except***:

A. Absent horizontal eye movements
B. Present vertical eye movements
C. Blinking on command
D. Deafness
E. Quadriplegia

Answer: D Locked-in syndrome damages the ventral portion of the pons, sparing structures above and below it. In most patients with a locked-in syndrome, consciousness is spared because the ascending reticular formation, which produces wakefulness, is located in the dorsal portion of the pons.

5. A patient presents to the emergency department in coma (intact brain stem findings and localization only to noxious stimuli; moaning). Vital signs are normal. The CT scan is normal. Arterial blood gases show a metabolic acidosis. Lactate and glucose levels are normal, as are measures of renal function. What additional laboratory test is needed?

A. Serum anion and osmolar gap
B. Serum ammonia level
C. Serum thyroid level
D. Serum cortisone level
E. Blood cultures

Answer: A Acute lactate metabolic acidosis points to ingestion of salicylates or ethylene glycol. The other disorders generally do not cause a metabolic acidosis.

405 DISORDERS OF SLEEP

BRADLEY V. VAUGHN

DEFINITION

Sleep is essential to good health and a sense of well-being. This normal state of decreased responsiveness promotes good function of bodily processes, restores the properties of alertness, and promotes memory storage and learning.[1] Conversely, the disruption of sleep is associated with a variety of complaints and physiologic consequences. Second in frequency only to pain, sleep-wake complaints lead over one in three individuals to seek medical attention. Sleep disruption frequently manifests as intrusion of components of the sleep state into periods of wakefulness. Untreated sleep disorders cause various health issues, impair job performance, and affect psychosocial interactions. Sleep disruption may also exacerbate symptoms of other diseases by worsening a preexisting disorder or impairing the ability to cope with the symptoms of the original disease. It is often difficult to recognize that such signs are related to dysfunctional sleep.

PATHOBIOLOGY

Complex humoral, neurochemical, and neuronal networks affect the sleep-wake state. Dynamic in organization, sleep is composed of non–rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep. NREM is divided into three stages (N1, N2, and N3) and REM is denoted as stage R. Each stage has a distinct physiologic regulation. Each of these sleep stages may have specific contributions to health, and the stages also interact.

Wakefulness involves activation of the monoaminergic neuronal groups, basal forebrain cholinergic neurons, and the brain stem reticular activating system (E-Fig. 405-1). These areas work in concert to promote the brain's ability to respond to stimuli. The reticular activating system promotes the relay of sensory information to the cerebral hemispheres, while the forebrain cholinergic and monoaminergic neurons promote attention of the hemispheric networks to sensory information.

With the onset of sleep, information from two major drives (homeostatic and circadian) influences the ventral lateral preoptic area to suppress the networks of wakefulness and allow the initiation of NREM sleep (E-Fig. 405-2). NREM sleep typically starts as stage N1, with mild slowing of the electroencephalogram (EEG) and slow eye movements (Table 405-1). Stage N1 is associated with the feeling of drowsiness. Although minimal sensory processing can occur, memory is not stored. Blood pressure may decrease slightly, and breathing becomes more periodic. Stage N1 represents about 5% of the night.

The hallmark of stage N2 sleep is the presence of characteristic sleep spindles and K complexes on the EEG. Although stage N2 is considered light sleep, this stage is associated with less responsiveness to stimuli than stage N1 and less responsiveness to elevated CO_2 and low oxygen. Stage N2 typically represents about 50% of a night's sleep.

Stage N3 is characterized by slow waves (0.5 to 2.0 Hz, >75uV) on the EEG. These slow waves are more prominent in brain areas that are more heavily used during the preceding waking period and may be related to reorganization of neuronal synapses. In stage N3, which constitutes 20% of the night, the individual is difficult to arouse and has rhythmic breathing that is slightly less responsive to elevated CO_2 and low oxygen than during stage N2.

REM sleep (stage R) is primarily generated by cholinergic neurons in the subcaeruleus nucleus in the brain stem, which then activate other neuronal groups to produce the rapid eye movements, active theta and alpha frequency EEG waveforms, loss of muscle tone, sensory processing, and reduced temperature regulation associated with stage R. Most vivid dreaming occurs in stage R, but dreams can occur in other stages. Despite the vivid dreams during stage R, most muscles are paralyzed. Ventilation is solely dependent upon the diaphragm, yet this stage is associated with the least amount of responsiveness to low oxygen and elevated CO_2. Stage R typically encompasses about 20% of the night and is ended by activation of norepinephrine and serotonergic neurons (E-Fig. 405-3). Age influences sleep. REM sleep occupies 50% of sleep at birth and then gradually declines to 20 to 25% by age 3 years. Slow wave sleep is prominent in children and declines in men in their late 20s and in women by age 40 to 50 years.

Sleep stages graphed through the night demonstrate the dynamic interplay of the various stages. As seen in a hypnogram (E-Fig. 405-4), sleep has repeating cycles of approximately 90 minutes. These cycles show a predominance of stage N3 in the first two cycles, and a gradual lengthening of the periods of stage R sleep in the latter half of the sleep period. The reason for this progression is unknown, but these features suggest other complex drivers are at work.

Many models that consider the array of neurochemical pathways that influence sleep can theoretically explain its physiologic regulation. The most accepted two-driver model uses the homeostatic and circadian drivers to

TABLE 405-1 SLEEP STAGE PARAMETERS*

STAGE	EEG FINDINGS	EYE MOVEMENTS (EOG)	SUBMENTAL EMG	ASSOCIATED PHYSIOLOGY
Wakefulness (W)	**More than 50% of an epoch has alpha rhythm over occipital region (posterior dominant rhythm).**	**Rapid eye movements** to slow movements. **Blinking** may be present.	**Normal to high muscle tone**	Memory registration, voluntary control over breathing
Stage N1	**Attenuation of the posterior dominant rhythm for > 50% of the epoch, replaced with mixed theta frequency low-amplitude activity.** *Vertex sharp waves.* N1 continues until beginning of N2 or arousal.	*Slow rolling eye movements*	*Variable but less than wake*	Automatic behavior can occur, mild cognitive processing, periodic breathing
Stage N2	**K-complexes and/or sleep spindles.** *Low-amplitude mixed-frequency EEG.* N2 persists until transition to N3, R, or an arousal.	*No eye movements, but slow eye movements may persist.*	*Variable amplitude, typically lower than W and higher than R*	No memory, decreased arousal to stimuli, less response to elevated CO_2 and low oxygen
Stage N3	**Slow wave activity (0.5-2 Hz, >75 μV) for > 20% of an epoch.** Sleep spindles may persist. N3 persists until transition to N2, R, or an arousal.	*No eye movements seen*	*Variable amplitude, typically lower than N2 and can be as low as R*	No memory, least responsive to arousing stimuli, less response to elevated CO_2 and low oxygen, but monotonous breathing pattern
Stage R (REM sleep)	**Low-amplitude mixed-frequency EEG.** *Sawtooth waves.* R persists until transition to N1, transition to N2, between K complexes without eye movements, or an arousal.	**Rapid eye movements**	**Low muscle tone**	Similar response to stimuli as light sleep, irregular breathing pattern, least response to elevated CO_2 and low oxygen

*Sleep staging requirements. **Boldfaced** items are requirements for staging. *Italicized* items are non-required associated findings that may be present in that stage.
EEG = electroencephalogram; EMG = electromyogram; EOG = electro-oculogram.
Adapted from the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events*. 2nd ed. version 2.1. Westchester, IL: American Academy of Sleep Medicine; 2014.

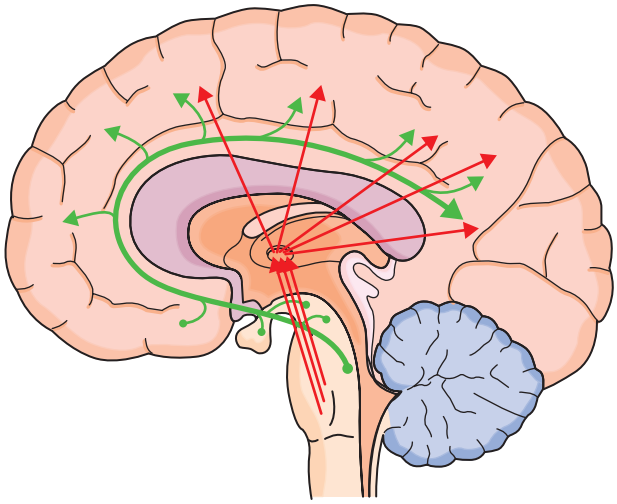

E-FIGURE 405-1. **The two components that produce wakefulness are the reticular activating system (*red*), which permits sensory input to be relayed by the thalamus, and the monoaminergic nuclei and basal forebrain cholinergic system, which promote responsiveness of the cerebral hemispheres.** (Adapted from Saper CB, Scammell TE, Lu J. Hypothalamic regulation of sleep and circadian rhythms. *Nature.* 2005;437:1257-1263.)

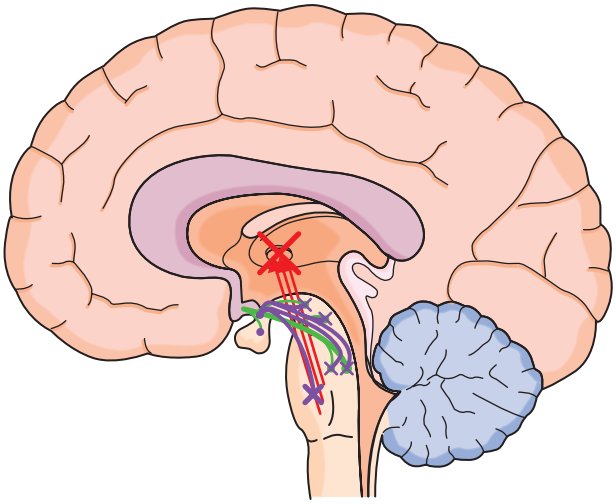

E-FIGURE 405-2. **Activation of the ventral lateral preoptic nucleus subsequently inhibits (*purple arrows*) the reticular activating system (*red*) and the monoaminergic nuclei (*green*), thereby producing non–rapid eye movement sleep and blocking sensory input to the hemispheres.** (Adapted from Saper CB, Scammell TE, Lu J. Hypothalamic regulation of sleep and circadian rhythms. *Nature.* 2005;437:1257-1263.)

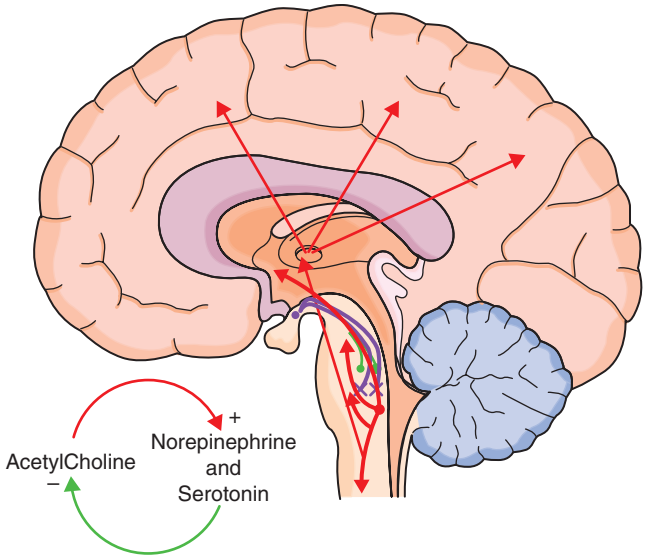

E-FIGURE 405-3. **Rapid eye movement sleep (REM) sleep is activated by activation of the subcaeruleus nucleus, which uses acetylcholine (*red arrows*), and further deactivation (*purple arrows*) of norepinephrine- and serotonin-producing neurons (*green*).** The reciprocal relationship of REM on (cholinergic neurons) and REM off (norepinephrine and serotonin neurons) is depicted in the lower corner. (Adapted from Saper CB, Scammell TE, Lu J. Hypothalamic regulation of sleep and circadian rhythms. *Nature.* 2005;437:1257-1263.)

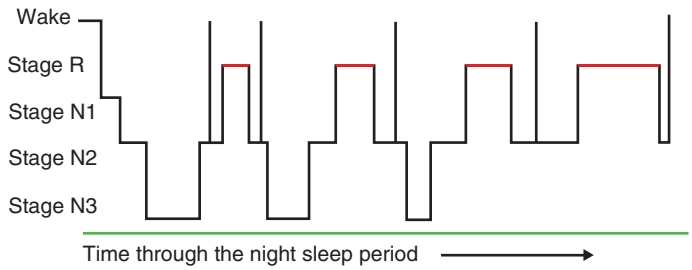

E-FIGURE 405-4. **Hypnogram showing sleep stages across a sleep period.**

explain sleep-wake state.[2] Other issues such as psychological status also play a role. In the two-driver model, the homeostatic drive is the accumulation of substances that promote sleepiness while the person is awake. These substances are metabolized during sleep. Mental and physical activities increase this drive by producing neuronal byproducts (e.g., adenosine), whereas caffeine blunts this drive by blocking adenosine. In contrast, the circadian rhythm drive promotes wakefulness and, through its predictable cycle, prepares the body for anticipated activities. The circadian rhythm is a naturally occurring rhythm that is slightly longer than 24 hours but is readjusted each day to maintain alignment with the natural day-night cycle. The circadian rhythm is primarily adjusted by bright light and to a lesser extent by other factors such as exercise, food, and social interactions. The hormone melatonin, which is released in response to darkness, can also influence the phase of the circadian rhythm. Throughout the 24-hour period, the homeostatic and circadian drives maintain balance between the sleep and wake states. When the circadian rhythm is stronger than the homeostatic drive, the person is awake, and when the homeostatic drive is stronger than the circadian rhythm, the person is sleepy (Fig. 405-1). This theoretical model helps explain aspects of sleep-wake regulation, such as the periods of post-lunch sleepiness or evening wakefulness.

CLINICAL MANIFESTATIONS

Most patients who seek medical help for sleep issues present with one of three complaints: (1) excessive sleepiness, (2) difficulty attaining or sustaining sleep, (3) or unusual events associated with sleep. Excessive sleepiness may be confused with fatigue or lack of energy. Common symptoms include morning headaches, lapses of attention, or diffuse muscle aches. Difficulty with sleep at night may be a clue to daytime issues, and nocturnal events may be a clue to brain issues.

DIAGNOSIS

Both subjective information and objective tests are used to investigate sleep complaints. Questionnaires such as the Pittsburgh Sleep Quality Index (Fig. 405-2) can provide a broad overview of sleep symptoms, including bedtime, wake time, activities, medications, and other substances that could influence sleep.

Objective testing of sleep includes actigraphy, polysomnography, multiple sleep latency testing, and maintenance of wakefulness testing. Actigraphy monitors movement, typically of a nondominant extremity, over 7 to 28 days (E-Fig. 405-5). When combined with a sleep diary, actigraphy estimates total sleep time and assesses the sleep-wake schedule. Polysomnography (Chapter 100, Fig. 100-1) assesses both sleep stage and associated physiology. Sleep stage is determined by EEG, electro-oculogram, and submental electromyogram activity. Measures assessing physiology include respiratory function (flow, effort, and gas exchange), limb muscle activity, electrocardiogram, and sometimes esophageal pH or core body temperature. Polysomnography is most useful for sleep disruption such as sleep apnea (Chapter 100), excessive movements, parasomnias, or for unexplained excessive sleepiness[3] (Table 405-2). More limited overnight recordings may focus on strictly respiratory

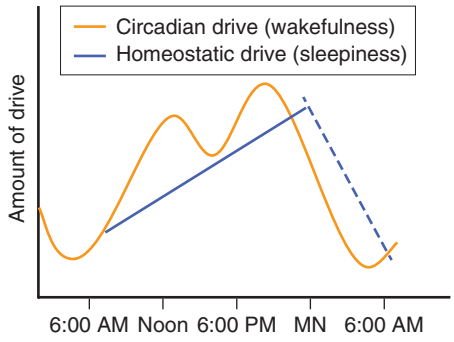

FIGURE 405-1. Idealized graph depicting the two-process model. The circadian driver promotes wakefulness (*orange*), and the homeostatic drive promotes sleep (*blue*). The dynamic interaction of the two drives is shown.

TWO WEEK SLEEP DIARY

INSTRUCTIONS:
1. Write the date, day of the week, and type of day: Work, School, Day Off, or Vacation.
2. Put the letter "C" in the box when you have coffee, cola or tea. Put "M" when you take any medicine. Put "A" when you drink alcohol. Put "E" when you exercise.
3. Put a line (I) to show when you go to bed. Shade in the box that shows when you think you fell asleep.
4. Shade in all the boxes that show when you are asleep at night or when you take a nap during the day.
5. Leave boxes unshaded to show when you wake up at night and when you are awake during the day.

SAMPLE ENTRY BELOW: On a Monday when I worked, I jogged on my lunch break at 1 PM, had a glass of wine with dinner at 6 PM, fell asleep watching TV from 7 to 8 PM, went to bed at 10:30 PM, fell asleep around Midnight, woke up and couldn't got back to sleep at about 4 AM, went back to sleep from 5 to 7 AM, and had coffee and medicine at 7:00 in the morning.

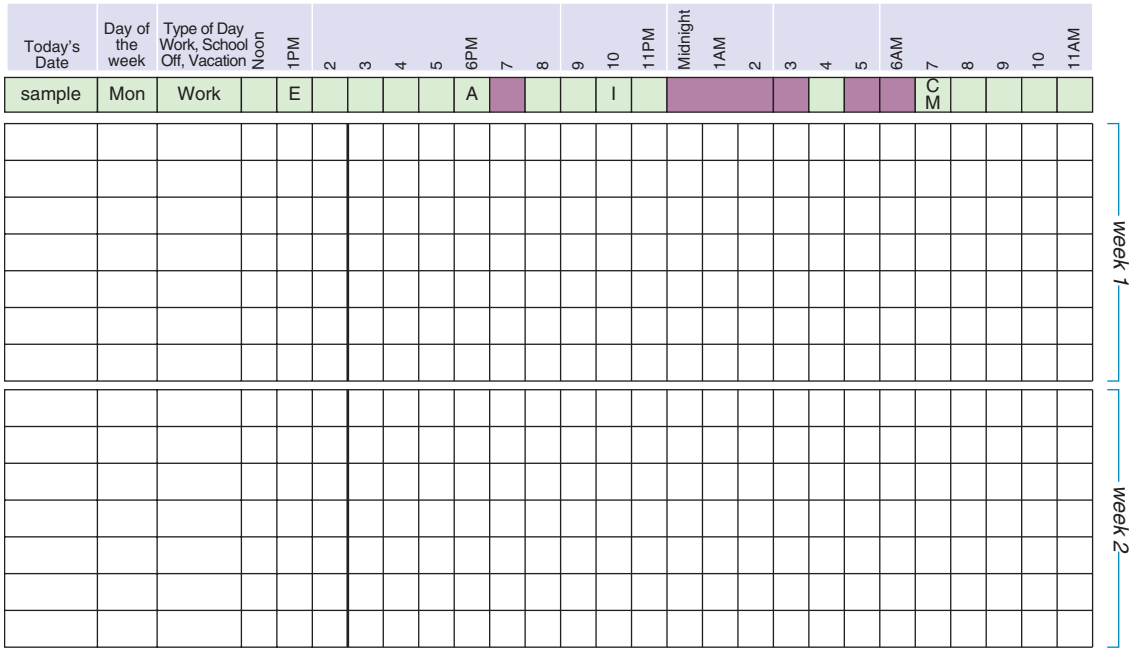

Used with permission from the American Academy of Sleep Medicine, Darien, Illinois.

FIGURE 405-2. Example of a sleep diary. Patients record their daily schedule, work, and medications.

E-FIGURE 405-5. Actigraphy report. The shaded areas are scored as rest time; in this instance, the bedtime is 11 PM to 12 midnight, and wake time about 6 AM. One nap is observed on day 3 in the middle of the afternoon.

TABLE 405-2 INDICATIONS FOR POLYSOMNOGRAPHY

POLYSOMNOGRAPHY IS ROUTINELY INDICATED FOR:
Diagnosis of sleep-related breathing disorders (SRBD), including suspected obstructive sleep apnea (OSA) in patients with coronary heart disease, history of stroke or transient ischemic attacks, or significant tachyarrhythmias or bradyarrhythmias
Positive airway pressure (PAP) titration in patients with sleep-related breathing disorders
A preoperative clinical evaluation to evaluate for the presence of OSA before upper airway surgery for snoring or OSA
Patients with heart failure if they have nocturnal symptoms suggestive of sleep-related breathing disorders (disturbed sleep, nocturnal dyspnea, snoring) or if they remain symptomatic despite optimal medical management
Patients with neuromuscular disorders and sleep-related symptoms
Patients suspected of periodic limb movement disorder
Polysomnography and multiple sleep latency test on the ensuing day for patients suspected of having narcolepsy
Follow-up polysomnography:
After titration of oral appliance treatment in patients with moderate to severe OSA
Following surgical treatment of patients with moderate to severe OSA
After surgical or dental treatment of patients with SRBDs whose symptoms return
Substantial weight gain or loss in patients on PAP for SRBD
Insufficient clinical response to PAP therapy
Assessment of oral appliance after final fitting (guideline)
Evaluation of patients with sleep behaviors suggestive of unusual or atypical parasomnias or in which specific motor patterns are in question
POLYSOMNOGRAPHY IS OPTIONAL FOR:
Evaluation of sleep behaviors suggestive of potentially injurious parasomnias
POLYSOMNOGRAPHY IS *NOT* ROUTINELY INDICATED FOR:
Patients whose symptoms resolve with continuous positive airway pressure (CPAP) treatment
Diagnosis of chronic lung disease
Diagnosis of typical, uncomplicated, and noninjurious parasomnias when the diagnosis is clearly delineated
Patients with a seizure disorder who have no specific complaints consistent with a sleep disorder
Diagnosis or treatment of restless legs syndrome, except where diagnostic uncertainty exists
Establishing the diagnosis of depression
Diagnosis of circadian rhythm sleep disorders

measurements. Two tests quantify the ability to fall asleep and stay awake: the multiple sleep latency test and maintenance of wakefulness test. The multiple sleep latency test quantifies objective sleepiness based upon the time to onset of sleep across five daytime naps. The multiple sleep latency test is useful for narcolepsy, but there is overlap between normal individuals and patients with sleep disruption. The maintenance of wakefulness test quantifies the propensity to stay awake across four 40-minute epochs, and it can provide objective evidence of the daytime efficacy of stimulant therapy.

HYPERSOMNIA

Sleepiness is normal just prior to a typical sleep period or after prolonged wakefulness. In 5 to 20% of adults, sleep is excessive because it occurs in inappropriate settings. When mild, sleepiness may have a minor effect on quality of life. When severe, however, sleepiness intrudes on activities such as driving, conversation, or eating, and it may cause lapses of attention or diminished cognitive abilities, such as missing an exit on the highway. The perception of sleepiness is reduced with prolonged sleep deprivation, so that chronically sleep-deprived individuals become accustomed to their impairment and fail to recognize their degree of sleepiness.

DIAGNOSIS

Clinicians should question hypersomnic patients for clues about sleep debt, dyssomnia, brain issues, or medical or psychiatric causes (Fig. 405-3). Patients should be queried regarding their schedule during the week and weekends. Information regarding sleep habits and environment may disclose important factors contributing to the sleepiness. Patients with sleep apnea (Chapter 100), narcolepsy, excessive periodic limb movements, circadian rhythm disorders, and parasomnias may have excessive daytime sleepiness as their main complaint. A history of snoring, observed apnea, morning headaches, cataplexy, sleep paralysis, hypnogogic hallucinations, or altered sleep schedule suggests contributions of a specific sleep disorder. Excessive sleepiness can also result from many medical disorders and medications. Patients with heart (Chapter 58), kidney (Chapter 131), or liver failure (Chapter 153), rheumatologic disease, or endocrinologic disorders such as hypothyroidism (Chapter 226) and diabetes (Chapter 229) may note sleepiness and fatigue. Neurologic disorders such as stroke (Chapters 407 and 408), tumor (Chapter 189), demyelinating disease (Chapter 411), and head trauma (Chapter 399) can cause excessive sleepiness.

Sleepiness can be quantified subjectively by questionnaires or by physiologic measures such as a multiple sleep latency test. The Epworth Sleepiness

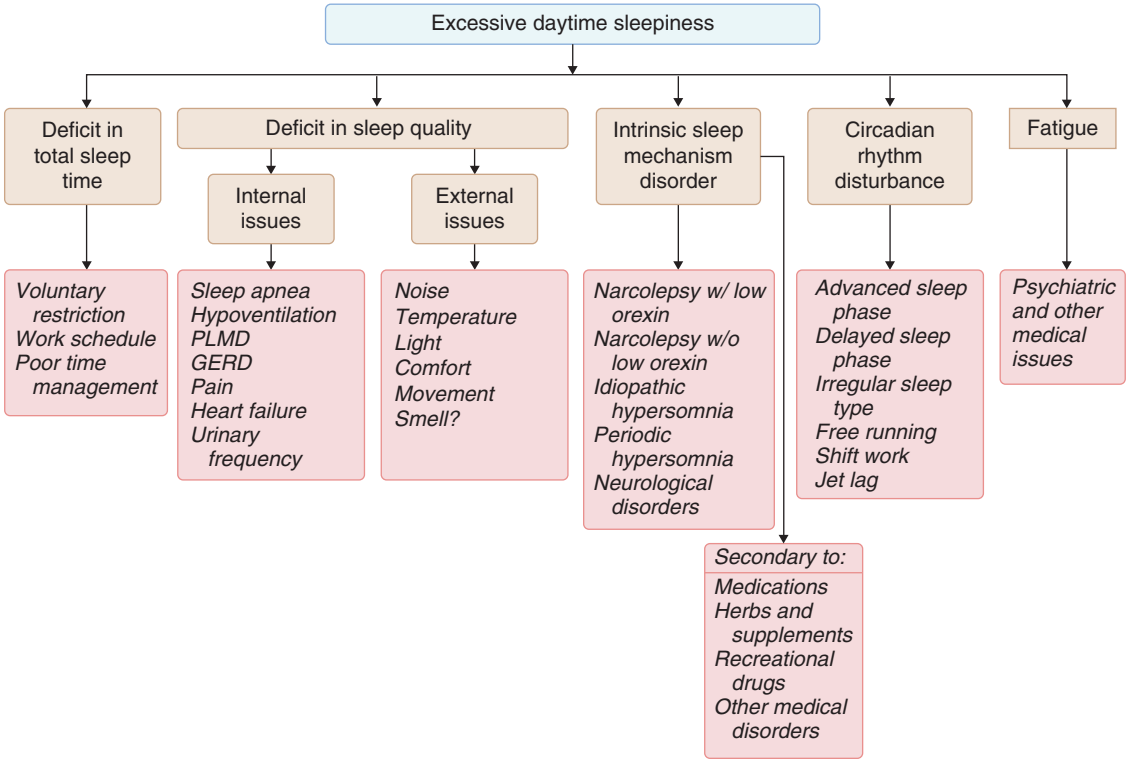

FIGURE 405-3. Differential diagnosis of excessive daytime sleepiness. GERD = gastroesophageal reflux disease; PLMD = periodic limb movement disorder.

TABLE 405-3 EPWORTH SLEEPINESS SCALE

How likely are you to doze off or fall asleep in the following situations, in contrast to just feeling tired? This refers to your usual way of life in recent time. Even if you have not done some of these things recently, try to work out how they would have affected you. Use the following scale to choose the most appropriate number for each situation.

0 = would never doze
1 = slight chance of dozing
2 = moderate chance of dozing
3 = high chance of dozing

SITUATION	CHANCE OF DOZING
Sitting and reading	
Watching TV	
Sitting and inactive in a public place (theater or meeting)	
As a passenger in a car for an hour without a break	
Lying down to rest in the afternoon when circumstances permit	
Sitting and talking to someone	
Sitting quietly after lunch (without alcohol)	
In a car, while stopped for a few minutes in traffic	
Total	

Adapted from Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep.* 1991;14:540-545.

Scale quantifies sleepiness by asking the subject to rate on a scale of 0 to 3 (0, no chance; 3, high likelihood) the chance of dozing in eight situations (Table 405-3). A score of 7 is considered average, whereas a score of 10 or more is consistent with subjective sleepiness. This score has a modest correlation with physiologic measures of sleepiness but a better correlation with the respiratory disturbance index in patients with obstructive sleep apnea (Chapter 100). Daytime studies, the multiple sleep latency test, and the maintenance of wakefulness test, may be used to assess sleepiness or wakefulness across a series of trial "naps." The multiple sleep latency test is validated as a test for narcolepsy, whereas the maintenance of wakefulness test gives a snapshot of the patient's ability to stay awake.

Other Hypersomnias

Idiopathic hypersomnia is a disorder in which hypersomnia cannot be explained by another disorder, is characterized by unrelenting hypersomnia, and is only minimally improved with therapy. Patients find that their symptoms persist despite long sleeping periods. These patients have average sleep latencies of less than 8 minutes, typically do not display REM sleep, but may have stage N3 on their multiple sleep latency test studies. Fluctuating symptoms of hypersomnia can also occur in other disorders such as in Kleine-Levin syndrome (a unique syndrome of periodic hypersomnia, hyperphagia, and hypersexuality) and in perimenstrual hypersomnia.

TREATMENT Rx

Treatment of sleepiness should focus on correcting the underlying cause of sleepiness.[4] Stimulants such as modafinil (200 to 400 mg) should be used only in individuals who are impaired by the symptoms and in whom other therapies have failed to correct the hypersomnia. Some patients with Kleine-Levin syndrome respond to lithium (Chapter 397).

Narcolepsy

DEFINITION

Narcolepsy includes a tetrad of excessive sleepiness, cataplexy, sleep paralysis, and hypnogogic hallucinations. In the past, narcolepsy was divided into patients with cataplexy (type 1) and patients without it (type 2), but the subtypes of narcolepsy are also defined based upon the presence or absence of the neurotransmitter hypocretin-1.

EPIDEMIOLOGY

Narcolepsy with cataplexy (narcolepsy type 1) affects1 in 2000 to 6000 individuals; 40 to 80% have the complete tetrad, and approximately 50% complain of sleep disruption. Over 90% of individuals in the United States with cataplexy have the HLA-DQB1*0602 gene,[5] and a similar percentage have low CSF hypocretin-1 levels. Narcolepsy without cataplexy (narcolepsy type 2) occurs in about 2 per 1000 individuals; approximately 40% have the HLA-DQB1*0602 gene, and fewer exhibit low CSF hypocretin-1 levels. Despite the connection to a gene, the risk to first-degree relatives is only 1 to 2%, or about a 10- to 50-fold increased risk compared with the general population.

PATHOBIOLOGY

Narcolepsy with cataplexy (narcolepsy type 1) reflects the loss of hypocretin-producing neurons in the lateral hypothalamus. This neurotransmitter is important for stabilizing the sleep-wake state and for motor control. Thus the manifestations of the disease are related to frequent stage shifts and intrusion of REM sleep atonia into wakefulness. Why these neurons are lost is not known, but immune mechanisms are postulated.[6]

CLINICAL MANIFESTATIONS

The tetrad of excessive sleepiness, cataplexy, hypnogogic hallucinations, and sleep paralysis are the major clinical manifestations. Cataplexy is abrupt loss of muscle tone triggered by strong emotional stimuli such as laughter, surprise, or anger. Patients are aware of their surroundings but lose muscle control, first in the face and neck, followed by the arms and then the trunk and legs. Hypnogogic (sleep-onset) and hypnopompic (sleep-offset) hallucinations are vivid and often frightening visual or auditory events. Sleep paralysis is an inability to move or speak, typically during the transition out of sleep when individuals have complete or partial awareness of their surroundings. Patients may describe a strong feeling of impending doom, being chased, or having to escape imminent danger. Patients with narcolepsy are often considered perpetually sleepy, but most have normal sleep duration over a 24-hour period. However, their sleep is fragmented, with sleep intruding into daily activities and interrupted at night with wakefulness. Sleep paralysis and hypnogogic hallucinations can occur in normal individuals, especially after sleep deprivation, but cataplexy is virtually pathognomonic for narcolepsy.

DIAGNOSIS

The diagnosis of narcolepsy type 1 and type 2 is based upon a mean sleep latency of less than 8 minutes and the presence of REM sleep on at least two of the five naps of a multiple sleep latency test. The multiple sleep latency test is predicated on the documentation of at least 6 hours of sleep prior to the study. The previous night's polysomnography must also not show other sleep pathologies. A low cerebrospinal fluid hypocretin level in the setting of excessive sleepiness can also confirm of the diagnosis of narcolepsy type 1, but this finding is not seen in type 2.

TREATMENT Rx

Treatment of narcolepsy focuses on improving symptoms of excessive sleepiness, cataplexy, and REM sleep intrusion into wakefulness (Table 405-4). Sleepiness requires a three-pronged approach of improving the quality and quantity of nighttime sleep, scheduling naps, and prescribing stimulants. Nighttime sleep may be improved with sodium oxybate (20 to 40 mg/kg in divided nighttime doses), which improves daytime alertness and reduces cataplexy. Stimulants such as modafinil (100 to 600 mg/day), armodafinil (50 to 250 mg/day), methylphenidate (5 to 60 mg/day), and dextroamphetamine (5 to 60 mg/day) improve daytime function but do not return the patient to a normal level.[7] Patients should not use stimulants in the evening and night hours. Selective serotonin reuptake inhibitors (SSRIs) (Chapter 397, Table 397-5) and combined serotonin-norepinephrine reuptake inhibitors (SNRIs) (Chapter 397, Table 397-5) also reduce cataplexy as well as sleep paralysis and hallucinations.

PROGNOSIS

Narcolepsy is lifelong disorder. Patients who present in adolescence or young adulthood may progress to more severe symptoms, but the disorder does not affect longevity.

SLEEP-RELATED BREATHING DISORDERS

Sleep Apnea

Sleep apnea (Chapter 100) is defined by repetitive breathing pauses that may interrupt sleep hundreds of times per night. The patient's bed partner may note that the patient's breathing has stopped or that the individual "holds their breath," but most patients are unaware of the events. In the sleep laboratory, apneas are defined by cessation of breathing for more than 10

TABLE 405-4 THERAPIES FOR NARCOLEPSY

MODALITY	STARTING DOSE	HIGHEST DOSE	DOSE AT
Scheduled naps	10-15 minutes	15 minutes	Just prior to time needing to be awake
OVER-THE-COUNTER STIMULANT			
Caffeine	25 mg	300 mg	AM
STIMULANTS			
Modafinil	100-200 mg	600 mg	AM and noon
Armodafinil	50-150 mg	250 mg	AM
Methylphenidate	5-10 mg	120 mg	AM and noon
Methylphenidate ER	10-20 mg	120 mg	AM
Dextroamphetamine	5-10 mg	60 mg	AM and noon
Combination dextroamphetamine/amphetamine	5-10 mg	60 mg	AM and noon
IMPROVE NIGHTTIME SLEEP, DAYTIME ALERTNESS, AND CATAPLEXY			
Sodium oxybate	225 mg	900 mg	Bedtime and 4 hr into sleep
THERAPIES FOR CATAPLEXY (NOT FDA APPROVED)			
Fluoxetine	10-20 mg	40 mg	AM
Venlafaxine	75 mg	225 mg	AM at lower dose or divided
Protriptyline	5 mg	40 mg	AM at lower dose or divided

seconds and are usually associated with oxygen desaturation and arousal occurring more frequently than five events per hour of sleep. Sleep apnea is classified as two major forms: obstructive and central. Obstructive apnea is defined as the loss of flow due to obstruction, typically in the upper airway, whereas central apnea is the absence of airflow due to the absence of effort.

Obstructive apnea is the most common form of sleep apnea. Approximately 50% of patients with sleep apnea have daytime sleepiness, but other symptoms such as insomnia and parasomnia events may be clues to underlying obstructive sleep apnea. Standardized questionnaires (Table 405-5) can help select patients for definitive polysomnography (Chapter 100, Fig. 100-1). Treatment typically involves the use of continuous positive airway pressure (CPAP), an oral appliance, or surgery. Neither supplemental oxygen■ nor medication■ provide substantial benefit.

Central apnea is the absence of ventilation without an effort to breath (E-Fig. 405-6) (Chapter 86). These patients have respiratory pauses that are associated with oxygen desaturation and arousals. Central apneas can be caused by cardiac disease, narcotics, or neurologic abnormalities that result in dysregulation of respiration. Cheyne-Stokes breathing, which may have features of both central and obstructive apnea, often occurs only during sleep. The classical Cheyne-Stokes pattern of crescendo-decrescendo breathing with central apnea can be seen in individuals with heart failure, neurologic lesions, and metabolic or toxic encephalopathies. Central apneas can be diagnosed by overnight polysomnography. The addition of a CO_2 monitor can distinguish between apnea related to normal or low CO_2 versus high CO_2 and help direct treatment. Low CO_2 levels in the presence of apnea may suggest a high CO_2 apnea threshold that may respond to increasing CO_2 levels, whereas apnea in the setting of an elevated CO_2 level would suggest failure of the respiratory control mechanism, as seen in patients who are taking narcotics. Therapy depends upon the etiology but can include reduction or elimination of respiratory suppressants (narcotics), CPAP, bilevel PAP, nocturnal ventilation, and respiratory stimulants.

TABLE 405-5 STOP-BANG QUESTIONNAIRE FOR OBSTRUCTIVE SLEEP APNEA

Snoring		
Do you snore loudly? (louder than talking or loud enough to be heard through closed doors)	Yes	No
Tired		
Do you often feel tired, fatigued, or sleepy during the daytime?	Yes	No
Observed		
Has anyone observed you stop breathing during your sleep?	Yes	No
Blood pressure		
Do you have or are you being treated for high blood pressure?	Yes	No
Body Mass Index (BMI)		
BMI more than 35 kg/m^2?	Yes	No
Age		
Age older than 50 yr?	Yes	No
Neck circumference		
Neck circumference greater than 40 cm?	Yes	No
Gender		
Gender male?	Yes	No
Elevated risk of OSA: answering yes to three or more items		
Low risk of OSA: answering yes to less than three items		

Adapted from Chung F, Yegneswaran B, Liao P, et al. STOP questionnaire: a tool to screen patients for obstructive sleep apnea. *Anesthesiology.* 2008;108:812-821.

HYPOVENTILATION

Hypoventilation, as defined by elevated CO_2 levels (Chapter 86), may occur solely during sleep. Patients may note daytime sleepiness, fatigue, morning headache, or unrefreshing sleep. Although the prevalence is unknown, hypoventilation is common in individuals with central obesity, neuromuscular disease, pulmonary disease, and narcotic use. Although typically worse in REM sleep, the elevation of CO_2 and commonly coexisting drop in oxygen saturation is more prolonged than the pattern seen with sleep apnea. Hypoventilation syndrome is treated with positive airway pressure or noninvasive ventilation.

Insomnia

DEFINITION

Insomnia is the complaint of difficulty initiating or maintaining sleep, or of unrefreshing sleep that results in daytime symptoms of excessive fatigue or impairment of performance.[8] Daytime sequelae differentiate individuals with a limited need for sleep from individuals with insomnia. Chronic insomnia is defined by symptoms that persist more than 1 month.

EPIDEMIOLOGY

Most individuals have occasional nights with difficulty falling asleep or maintaining sleep, often provoked by psychological challenges or sudden changes in their environment. Approximately 35% of individuals complain of intermittent difficulty with sleep, and approximately 10% have chronic insomnia. Women, older individuals, and patients with psychiatric or chronic medical illness are predisposed to develop insomnia. Insomnia is also more common in individuals with lower socioeconomic status and poor education. Patients with behavioral traits such as obsessive-compulsive tendency, frequent rumination, or poor coping strategies are also at greater risk for insomnia. Lack of "good-quality" sleep disrupts life and may lead to other symptoms.

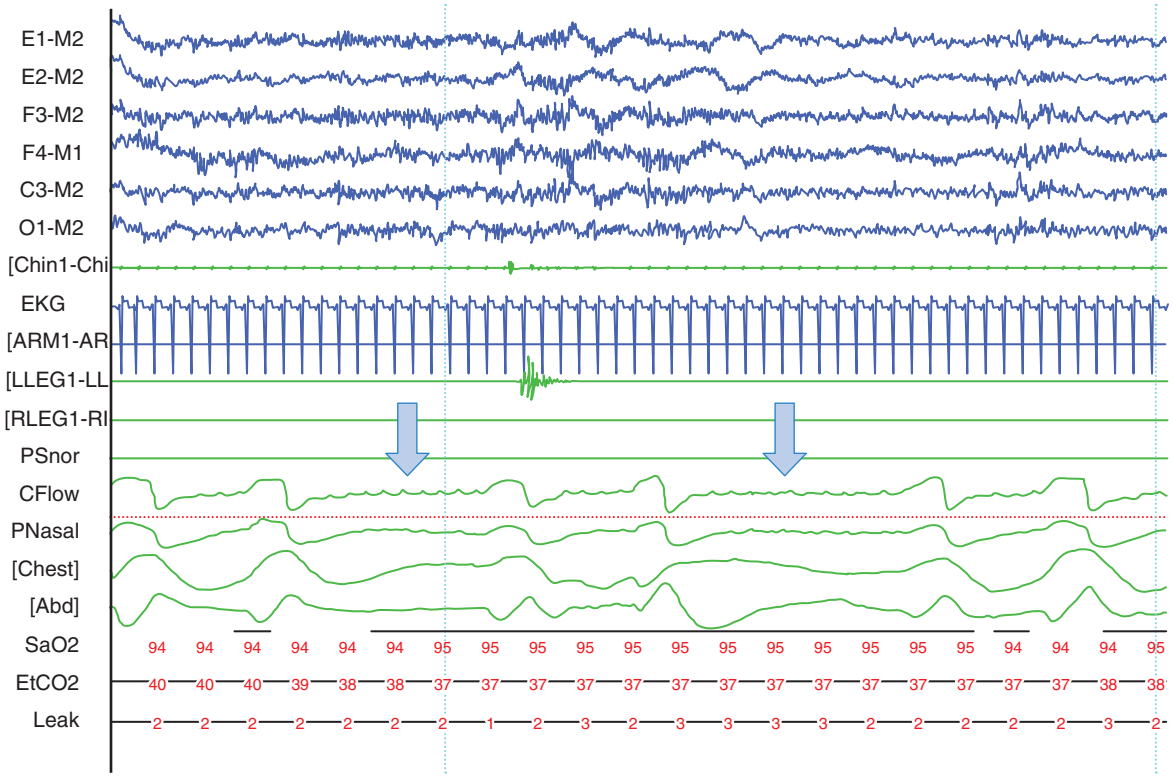

E-FIGURE 405-6. **This epoch from a polysomnographic recording shows repetitive central sleep apnea events (*blue arrows*). The patient has no chest movement, indicating this event is central in nature.**

PATHOBIOLOGY

Patients with insomnia frequently give historical clues directed toward the mechanisms behind their insomnia. Studies on patients with insomnia show these individuals are in a state of hyperarousal. Increased brain metabolic rates during NREM sleep may provide a neurophysiologic basis for chronic insomnia. Most patients have multiple factors that contribute to the insomnia, including features that predispose them to insomnia, events that precipitated the insomnia, and behaviors that perpetuate the insomnia. Effective treatment requires identifying these contributing factors. Many patients have a coincident psychiatric disorder (Chapter 397) or psychological or medical issues. Patients with depression or anxiety may have insomnia for years prior to the presentation of the affective disorder. Patients with heart (Chapter 58), liver (Chapter 153), or renal (Chapter 131) failure or disturbances of the gastrointestinal or respiratory systems commonly complain of insomnia. Patients with heart failure may note difficulty remaining in bed owing to breathing issues. Restless legs syndrome frequently presents as insomnia. Pain of any origin can interrupt sleep, and patients with limited mobility, such as muscular dystrophy (Chapter 421) or Parkinson disease (Chapter 409), may have pressure points that awaken them. Sleep schedules may be influenced by disease (e.g., patients with dementia [Chapter 402] in whom circadian rhythm abnormalities promote nighttime awakenings).

CLINICAL MANIFESTATIONS

The patient's symptom complex may give clues to a poor sleep environment, maladaptive behaviors, psychological stress, psychiatric or neurologic disease, primary sleep disorder, or other medical issues. Insomnia may be initiated by events that shift schedules or by a change in medications. Initiating events may play little role in long-term insomnia but give important clues to preventing further recurrence of the insomnia. If insomnia persists, many patients adopt behaviors that perpetuate the insomnia. Maladaptive habits that may occur during the day or night include heavy daytime caffeine or alcohol use, watching television or playing video games while in bed, or eating or exercising during the usual sleep period. A subgroup of patients may develop sleep phobias or have anxiety about the oncoming sleep period. This expectation of poor sleep promotes apprehension about sleep and may perpetuate counterproductive sleep rituals. These maladaptive behaviors become the predominant feature of the subtype of psychophysiologic insomnia (Table 405-6). Some patients exaggerate their symptoms, whereas other patients may not perceive that they are asleep. Individuals with paradoxical insomnia have normal physiologic sleep but do not recognize that they have been asleep. Other patients may have the unrealistic expectation that sleep should not be interrupted by any arousals or that they must sleep a set number of hours. Rarely, idiopathic insomnia starts in childhood and continues as a lifelong difficulty of sleep. These patients may have defective sleep mechanisms. Noting the timing of the insomnia during the sleep period may also be helpful. Difficulty with the onset of sleep suggests an underlying delayed sleep phase, and insomnia with early morning arousal suggests underlying depression or advanced sleep phase. Documentation of schedule changes (e.g., from jet lag or shift work) can be useful in determining links to circadian rhythm issues.

DIAGNOSIS

The diagnosis of insomnia is based upon the patient's history that difficulty with sleep results in daytime sequelae (Fig. 405-4). Frequently, more than one subtype occurs in the same patient, and there is little evidence that subtypes direct therapy.

TABLE 405-6 CLASSIFICATION OF ADULT INSOMNIA

INSOMNIA
Subtypes:
Psychophysiologic insomnia—maladaptive behaviors conditioned in response to associating the bed environment or thoughts of bedtime with heightened arousal; patients typically sleep better in a different environment, such as away on vacation.
Idiopathic insomnia—insomnia beginning in infancy or childhood, with a persistent unremitting course and no improvement with change in environment
Paradoxical insomnia (sleep state misperception)—insomnia characterized by a marked mismatch between the patient's description of sleep duration and objective polysomnographic findings
INSOMNIA ASSOCIATED WITH
Adjustment insomnia—associated with an acute or active psychosocial stressor
Inadequate sleep hygiene—associated with lifestyle habits that impair the ability to sleep
Insomnia comorbid with a psychiatric disorder—associated with an active psychiatric disorder such as anxiety or depression
Insomnia comorbid with a medical condition—associated with a condition such as renal failure, hepatic failure, chronic pain, nocturnal cough or dyspnea, or hot flashes
Insomnia caused by a drug or substance—secondary to consumption or discontinuation of medications, drugs of abuse, alcohol, or caffeine

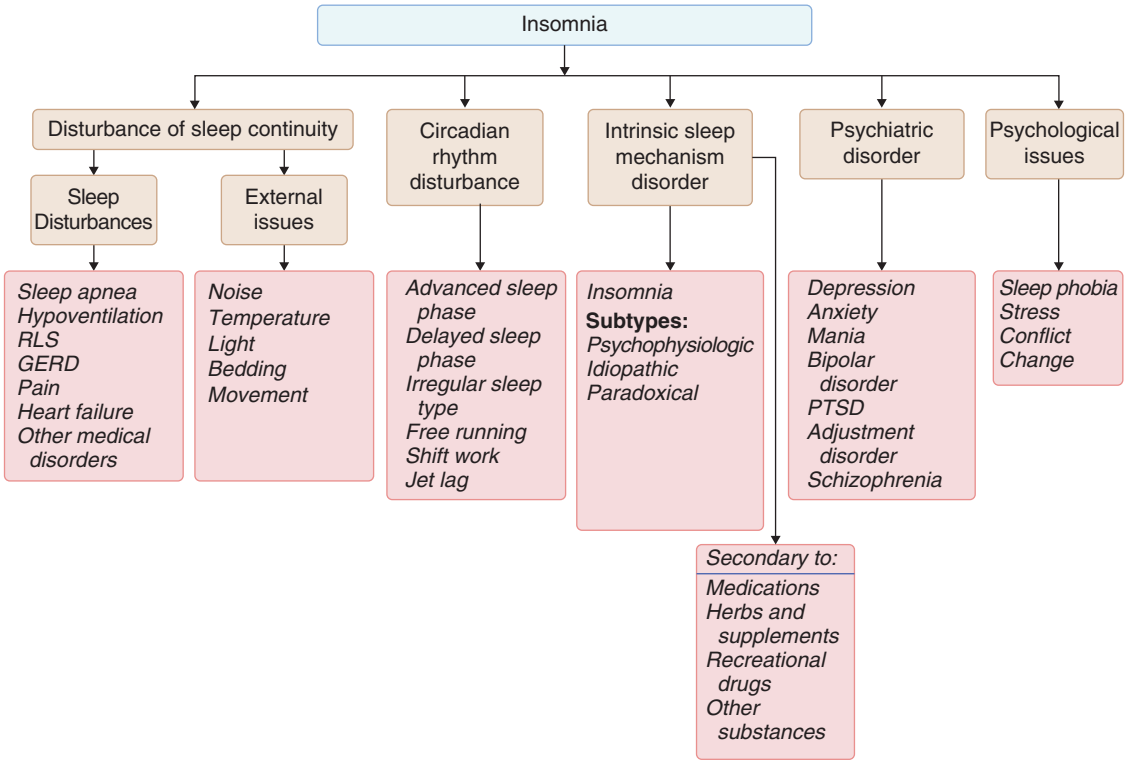

FIGURE 405-4. Differential diagnosis of insomnia. GERD = gastroesophageal reflux disease; PTSD = post-traumatic stress disorder; RLS = restless leg syndrome.

The history should include a review of the patient's 24-hour schedule, meals, caffeine, tobacco and medicine intake, sleep environment, attitudes about sleep, and the sleep experience. In addition, a thorough history from the bed partner may disclose features the patient is unaware of, such as snoring, limb movements, and sleep habits. Patients should be asked to keep a diary of their daily events for at least 3 weeks. This diary will often show specific patterns or clues the patient is unaware of. Actigraphy can also help determine the patient's sleep-wake schedule. Polysomnography should be considered only if the patient has symptoms of sleep apnea or has failed multiple therapeutic trials.

TREATMENT

Rx

Insomnia is generally treated as a single disorder even though it has been divided into subtypes (see Table 405-6), but effective treatment requires identifying factors that contribute to insomnia. Treatment is multipronged and includes improving behaviors that promote sleep, addressing the perpetuating factors, and deciding if hypnotic medication is appropriate. Every patient with insomnia needs to develop good sleep hygiene habits that should be reviewed with the patient and bed partner. Cognitive behavioral therapy provides long-term success for insomnia (Table 405-7), but a combination of cognitive behavioral therapy with hypnotics outperforms either alone. The cognitive portion focuses on restructuring beliefs about sleep, whereas behavioral therapies focus on actions that may mitigate maladaptive behaviors and promote better sleep behaviors: progressive relaxation techniques, stimulus control, sleep or time-in-bed restriction, and factors that accentuate homeostatic and circadian drives.

Hypnotics are best used for short-term treatment while starting cognitive behavioral therapy (Table 405-8). The benzodiazepine receptor agonists zolpidem (5 to 10 mg [CR form, 6.25 to 12.5 mg]), zaleplon (5 to 10 mg), and longer-acting eszopiclone (1 to 3 mg) are the usual initial therapies. Agents with rapid onset and a short half-life are used for difficulties initiating sleep. Agents with a longer half-life or continuous-release agents are used for sleep maintenance. Melatonin or a melatonin receptor agonist (ramelteon 8 mg) improves the initiation and maintenance of sleep. Antidepressant medications (Chapter 397, Table 397-5) are also used for insomnia. Low-dose doxepin (3 and 6 mg) promotes sleep by blocking the effects of central nervous system histamine. In patients with insomnia and an underlying affective disorder, the combination of the short-term use of a hypnotic, such as a benzodiazepine receptor agonist, and long-term antidepressant or anxiolytic is better for both the insomnia and affective disorder than either therapy alone.

PROGNOSIS

Most patients will improve, although some will relapse. Intractable insomnia often heralds an affective disorder (Chapter 397). Except for very rare patients with the prion-induced fatal familial insomnia (Chapter 415), patients with insomnia have only a slightly lower life expectancy, primarily from cardiovascular disease.[9]

Circadian Rhythm Disorders

DEFINITION

Circadian rhythm disorders cause misalignment of the person's sleep-wake cycle and the naturally occurring day-night cycle. Pathologic symptoms must be persistent or recurrent, and the patient must incur some social, occupational, or additional impairment. Individuals may note insomnia, excessive daytime sleepiness, or both. Circadian rhythm sleep-wake disorder is typically classified by comparing the patient's rhythm to the naturally occurring day. Circadian rhythm sleep-wake disorders are subtyped into advanced phase type (early to bed and rise), delayed phase type (late to bed and rise), irregular type (no clear pattern), and free-running type (no entrainment to the environment). Additionally, the inciting situation is included (e.g., jet lag type or shift work type). Another disorder related to circadian rhythm involves the gastrointestinal cycle: night eating syndrome, in which individuals consume over half of their caloric intake after 9 PM.

TABLE 405-7 NONPHARMACOLOGIC THERAPIES FOR INSOMNIA

COGNITIVE BEHAVIORAL THERAPY with or without relaxation therapy (Standard)
The combination of multiple modalities noted below
STIMULUS-CONTROL THERAPY (Standard)
Go to bed only when sleepy.
Use the bedroom only for sleeping and sex.
Go to another room when unable to sleep in 15 to 20 minutes, read or engage in other quiet activities, and return to bed only when sleepy; repeat if necessary.
Have a regular wake time regardless of the duration of sleep.
Avoid daytime napping.
SLEEP-RESTRICTION THERAPY (Guideline)
Reduce time in bed to the estimated total sleep time (minimum, 5 hr).
Increase time in bed by 15 minutes every week when the patient estimates the sleep efficiency is at least 85% (ratio of time asleep to time in bed).
RELAXATION THERAPY (Standard)
Physical component: progressive muscle relaxation, autogenic training
Mental component: reducing intrusive thoughts through imagery training, meditation, or hypnosis
PARADOXICAL INTENTION (Guideline for sleep-onset difficulties)
Instruct the patient to remain passively awake in bed and avoid any effort to fall asleep.
COGNITIVE THERAPY (Insufficient evidence as a single therapy)
Education to alter maladaptive or unrealistic beliefs and attitudes about sleep, such as that a minimum of 8 hours of sleep per night is required for health.
SLEEP HYGIENE EDUCATION (Insufficient evidence as a single therapy)
Correction of extrinsic factors and behaviors that affect sleep, such as environmental disruption (pets, music, or television); bedroom temperature; fixation on the bedside clock; use of alcohol, nicotine, or caffeine; lack of exercise or exercise too close to bedtime.

TABLE 405-8 MEDICATIONS FOR INSOMNIA

NAME	DOSE	TIME OF DOSE	FDA INDICATION	COMMON SIDE EFFECTS	HALF-LIFE	MECHANISM
Melatonin	0.25-6 mg	Evening 1-3 hr before bed	No	Grogginess, headache	30-50 min	Melatonin receptor agonist (dark)
Tryptophan	1-15 g	Evening	No	Drowsiness, headaches, dizziness	1-3 hr	Modulates serotonin
Zolpidem SL	1.75-10 mg	Bedtime	Yes	Sleepiness, amnesia, falls parasomnias	1-2 hr	Benzodiazepine receptor agonist (BZRA)
Zolpidem reg	5-10 mg	Bedtime	Yes	Sleepiness amnesia, falls, parasomnias	1-2 hr	BZRA
Zolpidem CR	6.25-12.5 mg	Bedtime	Yes	Sleepiness, amnesia, falls	1-2 hr but continued release	BZRA
Zaleplon	5-20 mg	Bedtime	Yes	Sleepiness, dizziness, parasomnias	1 hr	BZRA
Eszopiclone	1-3 mg	Bedtime	Yes	Sleepiness, dizziness	4-8 hr	BZRA
Doxepin	3-6 mg	Bedtime	Yes	Drowsiness, dizziness, nausea	17 hr	Histamine receptor antagonist
Mirtazapine	7.5-15 mg	Bedtime	No	Drowsiness, dizziness, weight gain	20 hr	Histamine receptor antagonist
Ramelteon	4-8 mg	Bedtime	Yes	Sleepiness, headache	1-2 hr	Melatonin receptor agonist

FDA = U.S. Food and Drug Administration.

EPIDEMIOLOGY

The prevalence of circadian rhythm disorders is not known. Some patterns of sleep are inherent in specific age groups. Advanced sleep phase issues are more common in the elderly, and delayed sleep phase issues are more common in adolescents, but these stereotypes may not indicate the true prevalence of the disorder. Purposeful shifting of the circadian rhythm, such as with shift work or jet lag, is common, but although 28% of the U.S. workforce works nights or rotating shifts, only one third of these individuals have this disorder. A free-running circadian rhythm is more common among blind persons, of whom about 25% have the disorder.

PATHOBIOLOGY

Circadian rhythm sleep-wake disorders may be more prevalent in today's "24-hour society," which offers constant stimuli to remain awake. Teens are more vulnerable to the phase delaying effects of light in the evening. The human "master clock" resides in the suprachiasmatic nucleus of the hypothalamus, but peripheral tissues also generate a self-sustained circadian rhythm based upon clock gene expression. About 2 to 10% of genes are expressed with circadian rhythmicity. Abnormalities in genetic clock genes may contribute to circadian rhythm disorders.[10] Variations in the *Clock*, *Per2*, and *Per3* genes appear to influence the morning/evening preference. Advance sleep phase type has been associated with the *Per2* S662G mutation and the *Ck1d* T44A mutation, whereas delayed sleep phase type is associated with the *Per3* V647G and *Ck1e* S408N mutations. The latter mutation is also associated with free-running type. Free-running type in individuals who are blind appears to be related to the loss of the photoreceptive ganglion cell input to the hypothalamus and not into the retina itself.

CLINICAL MANIFESTATIONS

Circadian rhythm disorders cause insomnia and/or excessive sleepiness. Patients may incur sleep deprivation by trying to maintain schedules that are not consistent with their inherent clocks. Some individuals "catch up" on the weekends by sleeping during their preferred times. Once asleep, the patient has sound sleep. Circadian rhythm sleep-wake disorders are associated with an increased risk of accidents and impair quality of life.

Delayed phase–type patients typically have trouble falling asleep and may not fall sleep for over 2 hours later than the conventional bedtime (E-Fig. 405-7A). They then have trouble arousing in the morning, preferring late wake times. Advanced phase–type patients fall asleep early in the evening and awaken several hours earlier than the conventional morning awakening (see E-Fig. 405-7B). Patients complain of early morning awakening and the inability to maintain wakefulness during evening activities. Free-running type individuals have a circadian rhythm that continues to run on the 24.3- to 25-hour cycle. In this disorder, also known as non-24-hour sleep-wake rhythm disorder, patients have alternating episodes of insomnia and excessive sleepiness, depending upon the phase of the endogenous sleep-wake cycle. This disorder can be easily confused with a periodic hypersomnia. Irregular-type patients have both excessive sleepiness and insomnia, with deceased functioning during the waking period.

Jet lag–type patients are affected by temporary changes in the environment owing to travel across time zones. Most find travel west easier than travel east because of the inherent nature of the clock being longer than 24 hours.

Shift work–type patients suffer from circadian misalignment because most shift workers try to resume a typical diurnal pattern on their days off despite nocturnal waking on their work days. This schedule predisposes the worker to poor adaptation.

DIAGNOSIS

The diagnosis of a circadian rhythm disorder is made by history and a 14-day sleep diary or actigraphy recording.[11] Normal individuals have a tendency toward "morningness" or "eveningness," so the diagnosis requires documentation of a negative impact of circadian rhythm on quality of life.

TREATMENT

Rx

Most therapy is directed toward aligning the circadian rhythm with the desired sleep-wake schedule, commonly by gradually shifting the sleep-wake schedule and then maintaining the schedule in the correct phase. Shifting the schedule, known as chronotherapy, can be accomplished by allowing a gradual delay or free running of the inherent schedule (promoted with the use of time clues) to move the phase of the circadian rhythm to the desired time.

Time clues may also be used to fix a circadian rhythm into a specific phase, but the circadian rhythm is susceptible to such clues only if they are given at an appropriate time of the endogenous circadian rhythm. For example, bright light given to a normal individual in the evening delays the cycle, whereas bright light in the morning may advance the cycle. A reliable point of reference is the nadir of the temperature cycle, which typically occurs approximately 2 hours prior to the natural wake up time. Typically, bright light, exercise, food, and social interactions delivered prior to the temperature nadir will delay the cycle, whereas these stimuli delivered after the temperature nadir will advance the cycle. Melatonin has an opposite effect and typically advances the cycle if given 2 to 4 hours prior to the onset of sleep and may delay the cycle if used after the temperature nadir.

After chronotherapy realigns the circadian rhythm, patients with delayed phase type may benefit from morning bright light or from melatonin in the evening. These individuals are subject to relapses, and the schedule should be strictly maintained. Patients with advanced phase type may be misdiagnosed with depression,[12] but they may benefit from evening bright light to delay the onset of sleep. Jet lag can be improved by appropriate use of time clues in reference to the endogenous cycle. Although short-term hypnotics and stimulants are commonly used to help adjustment, these do not realign the circadian rhythm any faster than the usual one time zone per day.

For shift-work disordered sleep, modafinil 100 to 200 mg at the start of the shift may help maintain alertness. Patients can wear sunglasses to minimize the phase shifting effects of the morning light. Melatonin once they return home and short-term use of a hypnotic may increase sleep duration. Alternatively, some shift workers may find a modified shift in their sleep schedule more appealing—sleeping 8 AM to 4 PM on their work days and 4 AM to noon on their days off. With this schedule, the circadian rhythm shifts by only 4 hours, so patients experience fewer total of hours shifting in any period.

Parasomnia

DEFINITION

Parasomnias are undesirable behavioral events or experiential phenomena occurring during entry into, within, or as part of arousal from sleep. These events include abnormal movements, behaviors, emotions, perceptions, dreaming, and activities of the autonomic nervous system. Parasomnias are typically subdivided into disorders of arousal from NREM sleep, REM sleep-related parasomnias, and other parasomnias. The disorders of arousal include disorders of sleepwalking, sleep terrors, and confusional arousal. REM-related parasomnias include nightmare disorder, REM sleep behavior disorder, and recurrent sleep paralysis. Other parasomnias include sleep-related eating, catathrenia (repetitive nocturnal groaning), or exploding head syndrome.

EPIDEMIOLOGY

Approximately 3% of adults and 15% of children have a sleep-related behavior. Although some parasomnias, such as disorders of arousal from NREM sleep (sleepwalking, sleep terrors, and confusional arousals) are more common in children, others have no age predilection or are more common in older individuals. REM-related parasomnias, such as nightmare disorder, are common among all ages and especially individuals with post-traumatic stress disorder (Chapter 397). REM sleep behavior disorder, which is another REM sleep–related parasomnia, is more common in the elderly, but the exact prevalence is unknown.

CLINICAL MANIFESTATIONS

Individuals and bed partners may complain of frequent movement during sleep. This complaint may be more concerning to the bed partner than the patient. Some individuals will complain of being active sleepers.

DIAGNOSIS

The history is the mainstay of the diagnosis of most parasomnias. Key features include age of onset, time of night of the events, memory for the events, and family history (Table 405-9). Stereotypical behavior, the same behavior with each event, can also help in categorizing the events. Events such as periodic limb movements, rhythmical movement disorder, or epileptic seizures are stereotypical, whereas sleepwalking, sleep or night terrors, and dream enactment have different behavior with each event. Although historical features can help distinguish among these disorders, many patients may require polysomnography to delineate the cause (see Table 405-9).

DISORDERS OF AROUSAL FROM NREM SLEEP

These disorders include a spectrum of behaviors that occur as a partial arousal from deep NREM sleep: sleep walking, sleep terrors, and confusional arousals. Although probably a continuum, the individual symptoms distinguish the

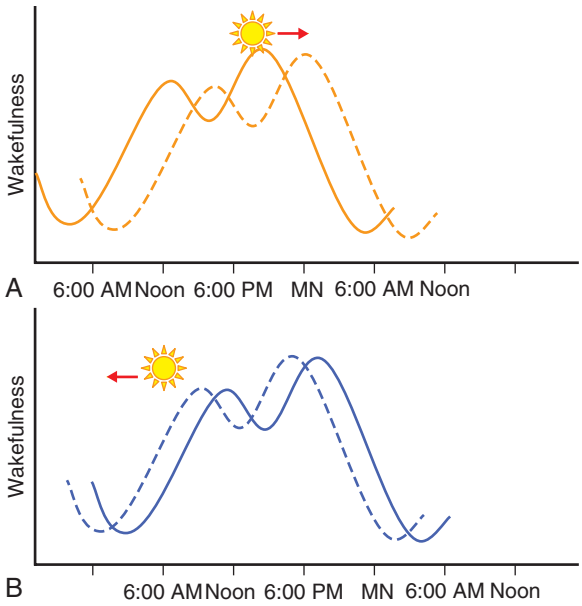

E-FIGURE 405-7. **A, An idealized graph of the circadian rhythm (*solid orange curve*), shows the phase shift delay (*dotted curve*) caused by bright light given prior to the temperature nadir. B,** This idealized graph of the circadian rhythm (*solid blue curve*), shows the phase shift advance (*dotted curve*) caused by bright light given after the temperature nadir.

TABLE 405-9 KEY FEATURES OF NOCTURNAL EVENTS

DISORDER	SYMPTOMS	TIME OF NIGHT	DURATION	FREQUENCY	STEREOTYPICAL	MEMORY	POLYSOMNOGRAPHIC FINDINGS
Sleepwalking	Slow, deliberate, complex behaviors	First half of sleep period	Seconds to minutes	Less than one per night to fewer	No	No or partial vague memory	Arousal from slow wave sleep
Sleep terrors	Piercing scream, followed by fight or flight response	First half	Seconds to minutes	Less than one per night or fewer	No	No or partial vague memory	Arousal from slow wave sleep
Confusional arousals	Variety of unusual behaviors upon sudden awakening	Anytime	Seconds to minutes	Less than one per night or fewer	No	No or partial vague memory	Arousal from slow wave sleep
Sleep-related eating	Eating of high-calorie or strange foods in a messy manner	First half	Minutes	May occur nightly	No	No or partial vague memory	Arousal typically from NREM sleep
Nightmares	Frightening dreams associated with anxiety	Latter half	Seconds to minutes	Variable	No, but may have a common theme	Yes	Events occur in REM sleep
REM sleep behavior disorder	Dream enactment, may be violent	Latter half	Seconds	Multiple times per night	No	Yes	Excessive EMG activity in REM sleep
Rhythmic movement disorder	Rocking, head banging	Near sleep onset but may be throughout the night	Minutes to hours	Multiple times per night	Yes	Yes	Rhythmic movement in transition from waking to sleeping
Catathrenia	Nocturnal prolonged moaning	Intermittent throughout the night	Minutes to hours	Multiple	Yes	No	Prolonged expiratory moans and groans, with slowed respiratory rate
Exploding head syndrome	Loud painless sound of explosion inside the head	Near the onset of sleep	Seconds	Rare, typically infrequent	Yes	Yes	Typically events are close to sleep onset

EMG = electromyogram; NREM = non–rapid eye movement; REM = rapid eye movement.

disorders. Individuals with sleepwalking have ambulation as part of their episodes, whereas sleep terrors are accompanied by a piercing scream or cry and expression of intense fear. Confusional arousals are characterized by disorientation, slow speech, and mentation or inappropriate behavior such as eating, fighting, or sexual intimacy. Pathobiologically, these individuals have NREM sleep simultaneously with the awake state. These events are more common in the first third of the night, are associated with no or little memory for the event, and are not stereotypical. Events are more likely to occur after sleep deprivation, alcohol ingestion, sleeping in strange environments, or coincidental conditions such as sleep apnea that evoke arousals. Patients are neurologically and psychiatrically normal during wakefulness. Polysomnography shows the episodes occur during slow wave sleep, with some features of wakefulness. Therapy includes ensuring safety for those who may injure themselves or others (e.g., placing the bed on the floor, blocking windows, or moving the patient's bedroom to the ground floor), decreasing factors that may cause arousals, and avoiding inciting factors such as sleep deprivation or alcohol. There are no established medications, but treatment of sleep apnea (Chapter 100) appears to reduce events. Other treatments, such as clonazepam 0.5 to 2 mg and tricyclic antidepressants (Chapter 397, Table 397-5), have been tried with varying success.

REM SLEEP BEHAVIOR DISORDER

In REM sleep behavior disorder, patients lose the characteristic sleep and muscle atonia of REM sleep and act out during their dreams.[13] REM sleep behavior disorder can be violent, with patients injuring themselves or bed partners. This elaborate motor activity is often associated with vivid recall of a dream that correlates with the witnessed behavior. Patients can have single or multiple events, commonly in the latter half of the night. REM sleep behavior disorder usually begins in late adulthood, but it can occur in children. This behavior disorder can be induced by medications such as tricyclic antidepressants, monoamine oxidase inhibitors, and serotonin reuptake inhibitors. Chronic REM sleep behavior disorder has been linked to alpha-synucleinopathies and the subsequent development of disorders such as Parkinson disease (Chapter 409), multiple system atrophy (Chapter 409), and Lewy body dementia (Chapter 402); over two thirds of adult patients with REM sleep behavior disorder eventually develop a neurodegenerative disease. The diagnosis is based upon the presence of excessive electromyographic activity during REM sleep and the history of dream enactment. Patients should be evaluated for signs of degenerative disorders (Chapters 402 and 409), strokes (Chapters 407 and 408), posterior fossa tumors (Chapter 189), or demyelinating disease (Chapter 411). Most patients respond well to clonazepam (0.25 to 3 mg) or melatonin (1 to 9 mg).

NIGHTMARES

Nightmares or recurrent disturbing dreams can be a presenting symptom of a sleep disturbance. Nightmares are emotionally intense dreaming associated with fear, anxiety, anger, sadness, or other negative emotions. Individuals awaken from stage R or light NREM sleep to full alertness and usually recall the event immediately. Nightmares are most commonly associated with a psychologically disturbing event, but they also may occur as a result of antihypertensive medications, antidepressants, or dopamine agonists. If related to medication, treatment starts with removal of the provocative substance. Prazosin (5 to 20 mg) and imagery rehearsal may be effective.

OTHER PARASOMNIAS

Individuals with sleep-related eating disorder consume high-calorie, sometimes bizarre, foods during sleep and have no or little memory for the consumption. They have morning anorexia and unexplained weight gain. Catathrenia is a rare disorder characterized by repetitive nocturnal groaning. Bed partners usually express concern because the patient has long expiratory groans that sound mournful. These patients respond to CPAP. Rhythmic movement disorder includes a variety of stereotyped movements, usually involving large muscles, that are sustained into light sleep. Movements may include head banging, body rocking, leg rolling, humming, and chanting. Patients are unaware of the movement or describe the movement as a compulsion prior to sleep. This behavior is difficult to treat but diminishes with age. Exploding head syndrome is an abrupt sensation or perceived loud sound of an explosion near the onset of sleep. It is painless, and the events are not a harbinger of other underlying pathology.

RESTLESS LEGS SYNDROME

Restless legs syndrome (Chapter 420) is characterized by four essential features: discomfort or urge to move, worse with rest, better with movement, and worse in the evening.[14] Patients may complain of an unpleasant crawling

or deep unusual sensation in the legs or arms, with improvement after moving the extremities. Patients with restless legs syndrome may relay that the discomfort can be debilitating and cause them to walk or continuously move their legs until the early morning hours. Some patients note that their legs will move or dance on their own, thereby indicating periodic limb movements in wakefulness. About 85 to 90% of restless legs syndrome patients will have periodic limb movements in sleep, but only a minority of patients with periodic limb movements in sleep will meet the clinical criteria of restless legs syndrome.

FDA-approved therapies for restless legs syndrome are dopamine agonists (pramipexole 0.125 to 1.5 mg or ropinirole 0.25 to 3 mg), transdermal rotigotine (1 to 3 patch/24 hours), and gabapentinoid medications (e.g., gabapentin-encarbil 600 to 1800 mg).[15] Augmentation, in which a dopamine agonist increases the severity of symptoms, is treated by carefully substituting another agent, such as pregabalin (300 mg daily), for the dopamine agonist. In some patients, restless legs syndrome has linked to low iron in the central nervous system. Some patients improve with oral iron therapy (e.g., 325 mg ferrous sulfate two to three times per day for 3 to 4 months until ferritin levels exceed 50 mg/L and iron saturations exceed 20%), and some require more aggressive intravenous iron therapy. More intractable patients may require chronic narcotics (Chapter 30, Table 30-4).

Grade A References

A1. Philip P, Chaufton C, Taillard J, et al. Modafinil improves real driving performance in patients with hypersomnia: a randomized double-blind placebo-controlled crossover clinical trial. *Sleep*. 2014;37:483-487.

A2. Alshaikh MK, Tricco AC, Tashkandi M, et al. Sodium oxybate for narcolepsy with cataplexy: systematic review and meta-analysis. *J Clin Sleep Med*. 2012;8:451-458.

A3. Gottlieb DJ, Punjabi NM, Mehra R, et al. CPAP versus oxygen in obstructive sleep apnea. *N Engl J Med*. 2014;370:2276-2285.

A4. Mason M, Welsh EJ, Smith I. Drug therapy for obstructive sleep apnoea in adults. *Cochrane Database Syst Rev*. 2013;5:CD003002.

A5. Ferracioli-Oda E, Qawasmi A, Bloch MH. Meta-analysis: melatonin for the treatment of primary sleep disorders. *PLoS ONE*. 2013;8:e63773.

A6. Liira J, Verbeek JH, Costa G, et al. Pharmacological interventions for sleepiness and sleep disturbances caused by shift work. *Cochrane Database Syst Rev*. 2014;8:CD009776.

A7. Allen RP, Chen C, Garcia-Borreguero D, et al. Comparison of pregabalin with pramipexole for restless legs syndrome. *N Engl J Med*. 2014;370:621-631.

A8. Trenkwalder C, Benes H, Grote L, et al. Prolonged release oxycodone-naloxone for treatment of severe restless legs syndrome after failure of previous treatment: a double-blind, randomised, placebo-controlled trial with an open-label extension. *Lancet Neurol*. 2013;12:1141-1150.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Yang G, Lai CS, Cichon J, et al. Sleep promotes branch-specific formation of dendritic spines after learning. *Science*. 2014;344:1173-1178.
2. Striz M, O'Hara BF. Clock genes and sleep homeostasis: a fundamental link within the two-process model? *Sleep*. 2013;36:301-302.
3. Qaseem A, Dallas P, Owens DK, et al. Diagnosis of obstructive sleep apnea in adults: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. 2014;161:210-220.
4. Mignot EJ. A practical guide to the therapy of narcolepsy and hypersomnia syndromes. *Neurother*. 2012;9:739-752.
5. Tafti M, Hor H, Dauvilliers Y, et al. DQB1 locus alone explains most of the risk and protection in narcolepsy with cataplexy in Europe. *Sleep*. 2014;37:19-25.
6. De la Herran-Arita AK, Garcia-Garcia F. Narcolepsy as an immune-mediated disease. *Sleep Disord*. 2014;2014:792687.
7. Thorpy MJ, Dauvilliers Y. Clinical and practical considerations in the pharmacologic management of narcolepsy. *Sleep Med*. 2015;16:9-18.
8. Buysse DJ. Insomnia. *JAMA*. 2013;309:706-716.
9. Li Y, Zhang X, Winkelman JW, et al. Association between insomnia symptoms and mortality: a prospective study of U.S. men. *Circulation*. 2014;129:737-746.
10. Ebisawa T. Analysis of the molecular pathophysiology of sleep disorders relevant to a disturbed biological clock. *Mol Genet Genomics*. 2013;288:185-193.
11. Nesbitt AD, Dijk DJ. Out of synch with society: an update on delayed sleep phase disorder. *Curr Opin Pulm Med*. 2014;20:581-587.
12. Campos Costa I, Nogueira Carvalho H, Fernandes L. Aging, circadian rhythms and depressive disorders: a review. *Am J Neurodegener Dis*. 2013;2:228-246.
13. St Louis EK. Key sleep neurologic disorders: narcolepsy, restless legs syndrome/Willis-Ekbom disease, and REM sleep behavior disorder. *Neurol Clin Pract*. 2014;4:16-25.
14. Leschziner G, Gringras P. Restless legs syndrome. *BMJ*. 2012;344:e3056.
15. Garcia-Borreguero D, Kohnen R, Silber MH, et al. The long-term treatment of restless legs syndrome/Willis-Ekbom disease: evidence-based guidelines and clinical consensus best practice guidance: a report from the International Restless Legs Syndrome Study Group. *Sleep Med*. 2013;14:675-684.

REVIEW QUESTIONS

1. Which of the following symptoms are clues to an underlying sleep disorder?

A. Brief lapses in memory
B. Excessive daytime sleepiness
C. Insomnia
D. Morning headache
E. All of the above

Answer: E Sleep disruption can cause all of these symptoms, and all are common among individuals with disturbed nighttime sleep. An individual might be attuned to specific aspects of their lives and perceive the disruption of sleep in different ways. Thus, the clinician can use these symptoms as an opportunity to further question aspects of sleep.

2. Hypersomnia can be caused by which of the following?

A. Sleep deprivation
B. Sleep apnea
C. Medication
D. Liver failure
E. All of the above

Answer: E The symptom of hypersomnia, or excessive ability to fall asleep, may be created by a wide range of issues. The clinician can think of sleepiness in general categories of sleep deprivation, sleep disruption (external [toxin] or internal [brain issue]) disturbing the mechanism for wake or sleep, a disturbance of circadian timing, or misperception of fatigue.

3. The diagnosis of insomnia requires which of the following?

A. Actigraphy
B. History of difficulty of sleeping at night
C. History of difficulty with sleep at night and daytime sequelae
D. History of somnogenic substance use and daytime sequelae
E. Polysomnography

Answer: C Insomnia is a clinical diagnosis best made by a detailed history that demonstrates difficulty with sleep and a resulting symptom of daytime impairment. No other laboratory data are needed to make the diagnosis, but other information may be helpful in determining the best therapy in selected cases. Most patients with insomnia do not need polysomnography, although actigraphy can be helpful in determining the patient's bedtime and wake schedule.

4. The diagnosis of a circadian rhythm disorder is best made by:

A. Sleep diary
B. Polysomnography
C. Multiple sleep latency test
D. Maintenance of wakefulness test
E. None of the above

Answer: A The diagnosis of a circadian rhythm disorder is dependent upon the complaint of insomnia or excessive sleepiness that is persistent for more than 3 months and can be confirmed by a sleep log or actigraphy for at least 14 days. Unfortunately, polysomnography and the daytime studies do not help in making this diagnosis.

5. Rapid eye movement (REM) sleep behavior disorder, characterized by excessive muscle activity in REM sleep and dream enactment, is most likely a harbinger of:

A. Alzheimer disease
B. Conversion disorder
C. Heart failure
D. Parkinson disease
E. Renal failure

Answer: D REM sleep behavior disorder has been linked to degenerative disorders known as synucleinopathies (Lewy body dementia, Parkinson disease, and multiple system atrophy). The symptoms of dream enactment may present more than 30 years prior to other symptoms of these degenerative disorders, and up to 80% of individuals with REM sleep behavior disorder develop one of the three. REM sleep behavior disorder also has been associated with other disorders, such as narcolepsy, and with brain stem strokes, lesions, and tumors. However, it is not associated with the other disorders listed as answers.

406

APPROACH TO CEREBROVASCULAR DISEASES

LARRY B. GOLDSTEIN

DEFINITION

The term *cerebrovascular disease* refers to a group of conditions in which injury to the brain or spinal cord occurs from a vascular cause. The onset is generally abrupt, but it also can be insidious. Clinical manifestations depend on the location and extent of damage to neural structures. Although risk factors and treatments may overlap, cerebrovascular diseases are pathophysiologically divided into those in which an insufficiency in the blood supply causes ischemic injury and those in which bleeding, either into the parenchyma (intracerebral or much more rarely intraspinal hemorrhage) or into the space between the pial and arachnoid coverings over the brain or spinal cord (subarachnoid hemorrhage), causes direct neural injury, leads to secondary ischemic injury, or acts as a space-occupying lesion. Cerebrovascular disease is often both preventable and treatable.

EPIDEMIOLOGY

Nearly 800,000 Americans have a stroke each year, and about 75% are first strokes. Measured in terms of disease-attributed healthy years of life lost, cerebrovascular disease ranks second in the United States and third worldwide.[1] Stroke, which is a generic term for cerebrovascular disease, has fallen from the third to the fourth leading cause of death in the United States (behind heart disease, cancer, and lung and respiratory diseases), primarily because of a dramatic reduction in stroke-related mortality. From 2000 to 2010, the annual stroke death rate in the United States fell by about 36%, with the actual number of stroke-related deaths falling by about 23%.[2] Stroke is the underlying cause of death of about 130,000 Americans each year, corresponding to approximately 1 in 19 deaths in the country. About 60% of stroke deaths occur in women, but the rates are actually highest in African American men. It is estimated that someone in the United States has a stroke about once every 40 seconds.

The overall prevalence of stroke is estimated at 2.8%, with 6.8 million American adults having had a stroke.[2] Even though the incidence of stroke has been declining substantially, largely because of better prevention, the declining case-fatality rate has kept the population prevalence reasonably stable.

The risk of stroke generally increases with age, and it doubles for every decade after the age of 55 years. In addition, blacks, people with lower levels of education, individuals who reside in the southeastern portion of the country (the "Stroke Belt"), and individuals with a first-degree relative who had a stroke before the age of 65 years have a higher risk of stroke and of stroke-related mortality. Poor diet, lack of exercise (Chapter 16), cigarette smoking (Chapter 32), exposure to environmental tobacco smoke, obesity (Chapter 220), and excess alcohol consumption (Chapter 33) are lifestyle factors that greatly increase the risk of stroke. Of the medical conditions that increase the risk of stroke, hypertension (Chapter 67) has the highest population-attributable risk. Other stroke risk factors include atrial fibrillation (Chapter 64), diabetes (Chapter 229), dyslipidemia (Chapter 206), inflammatory states, elevated homocysteine levels, high lipoprotein (a), carotid artery stenosis, patent foramen ovale (Chapter 69), other congenital heart defects, and sleep apnea (Chapter 100).[3] Coagulation disorders (Chapter 176), oral contraceptive agents (Chapter 238), and migraine headache with aura (Chapter 398) also may contribute to the risk. Mendelian diseases associated with stroke include sickle cell disease (Chapter 163); mitochondrial encephalopathy, lactic acidosis, and stroke (MELAS); cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL; Chapter 402); Fabry disease (Chapters 208 and 275); and Marfan syndrome (Chapter 260). In addition, autosomal dominant polycystic kidney disease (Chapter 127) is associated with intracranial aneurysms and fibromuscular dysplasia. Ehlers-Danlos type IV (Chapter 260) is also associated with intracranial aneurysms as well as cervical arterial dissection. Several genetic polymorphisms also have been associated with stroke (e.g., variants on chromosome 9p21 and 4q25), although these genetic markers are not yet clinically relevant.

PATHOBIOLOGY

Anatomy

An understanding of vascular anatomy and its normal variants as well as their relationships to functional neuroanatomy can provide important clues for identifying the cause of cerebrovascular symptoms and signs in individual patients and can also help guide treatment.

Aortic Arch

Paired carotid and vertebral arteries normally supply the brain (Fig. 406-1). The right common carotid artery arises from the brachiocephalic trunk (innominate artery), which then gives rise to the right subclavian artery. The right vertebral artery generally arises from the proximal portion of the right subclavian artery. The left common carotid artery usually arises directly from the aortic arch; but in some individuals, it may arise from the proximal portion of the brachiocephalic trunk ("bovine" anatomy). The left subclavian artery originates from the aortic arch distal to the left common carotid artery and also supplies the left vertebral artery.

Internal Carotid Arteries

The common carotid arteries bifurcate into the internal carotid artery and external carotid artery in the neck, generally at the level of the thyroid cartilage. The bifurcation may less commonly occur above the lower level of the

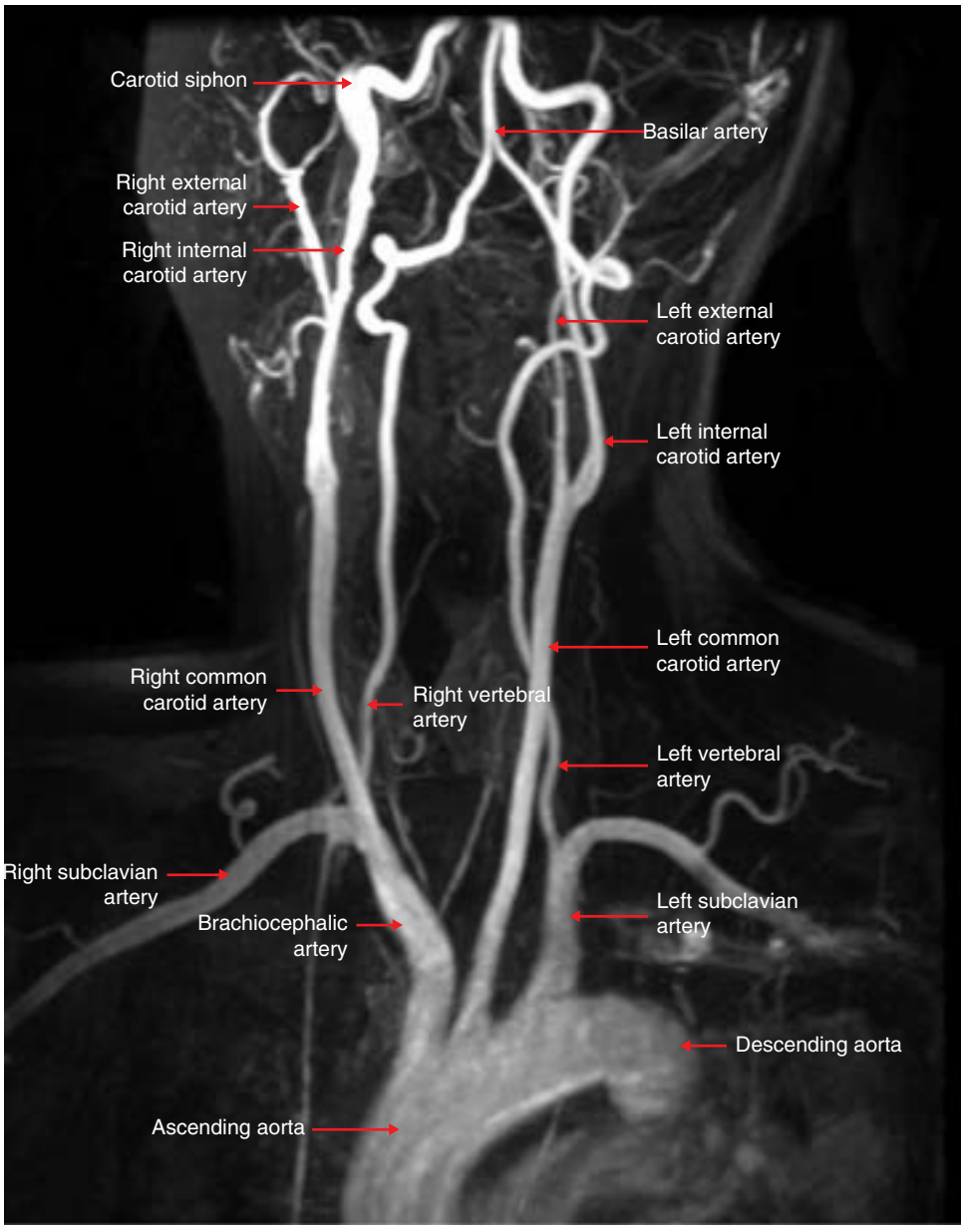

FIGURE 406-1. **Magnetic resonance angiogram of normally configured aortic arch.**

mandible or lower in the neck. The internal carotid artery enters the skull through the foramen lacerum and travels through the petrous bone adjacent to the inner ear. It then enters the cavernous sinus, ascends in an S shape (carotid siphon), penetrates the dura, and finally divides into the anterior cerebral artery and middle cerebral artery (Fig. 406-2). The ophthalmic artery can originate from the internal carotid artery in the carotid siphon, but it more commonly arises from the supraclinoid internal carotid artery, followed by the posterior communicating and anterior choroidal arteries.

External Carotid Arteries

In contrast to the internal carotid arteries, the external carotid arteries have extracranial branches. The superficial temporal arteries (palpable anterior to the ears) and facial arteries can anastomose with the intracranial circulation through branches of the ophthalmic artery and can be clinically important in the setting of a proximal internal carotid artery occlusion.

Vertebral Arteries

Although the vertebral arteries generally arise from the subclavian arteries, they can also originate from the aortic arch or thyrocervical trunk. They most commonly enter the C6 transverse process but may also enter at the C4, C5, or C7 levels. They exit the transverse processes at C1, turn posteriorly behind the atlantoaxial joint, and then pass through the dura at the foramen magnum. Intracranially, they typically join at the pontomedullary junction to form the single basilar artery, although the vertebral artery can end in the posterior inferior cerebellar artery in some individuals (Fig. 406-3). The portion of the vertebral artery between its origin and its entry into the transverse process is referred to as the V1 segment. The V2 segment refers to the portion of the artery traveling through the transverse foramina; the V3 segment, the portion between where the artery exits the transverse foramina and penetrates the dura; and the V4 segment, the intracranial portion of the artery. One vertebral artery may be hypoplastic (E-Fig. 406-1). Clues are that the ipsilateral transverse foramina are generally smaller on the side of the hypoplastic artery and that the proximal portion of the basilar artery can be displaced ipsilateral to the hypoplastic artery. The V3 segment is particularly vulnerable to mechanical injury that can lead to dissection. The vertebral arteries have medial branches that unite to form the anterior spinal artery and lateral branches that supply the dorsolateral medulla and inferior portion of the cerebellum, which also supplies the vestibular nuclei (Fig. 406-4). Other medial branches of the vertebral artery supply the medullary pyramid, inferior olivary nucleus, medial lemniscus, and hypoglossal nerve fibers. Longer circumferential branches from the vertebral arteries and posterior cerebral arteries supply the spinothalamic tracts and sympathetic fibers as they traverse the medulla, the sensory nuclei, and the descending tracts from cranial nerve V as well as emerging fibers from the vagus and glossopharyngeal nerves.

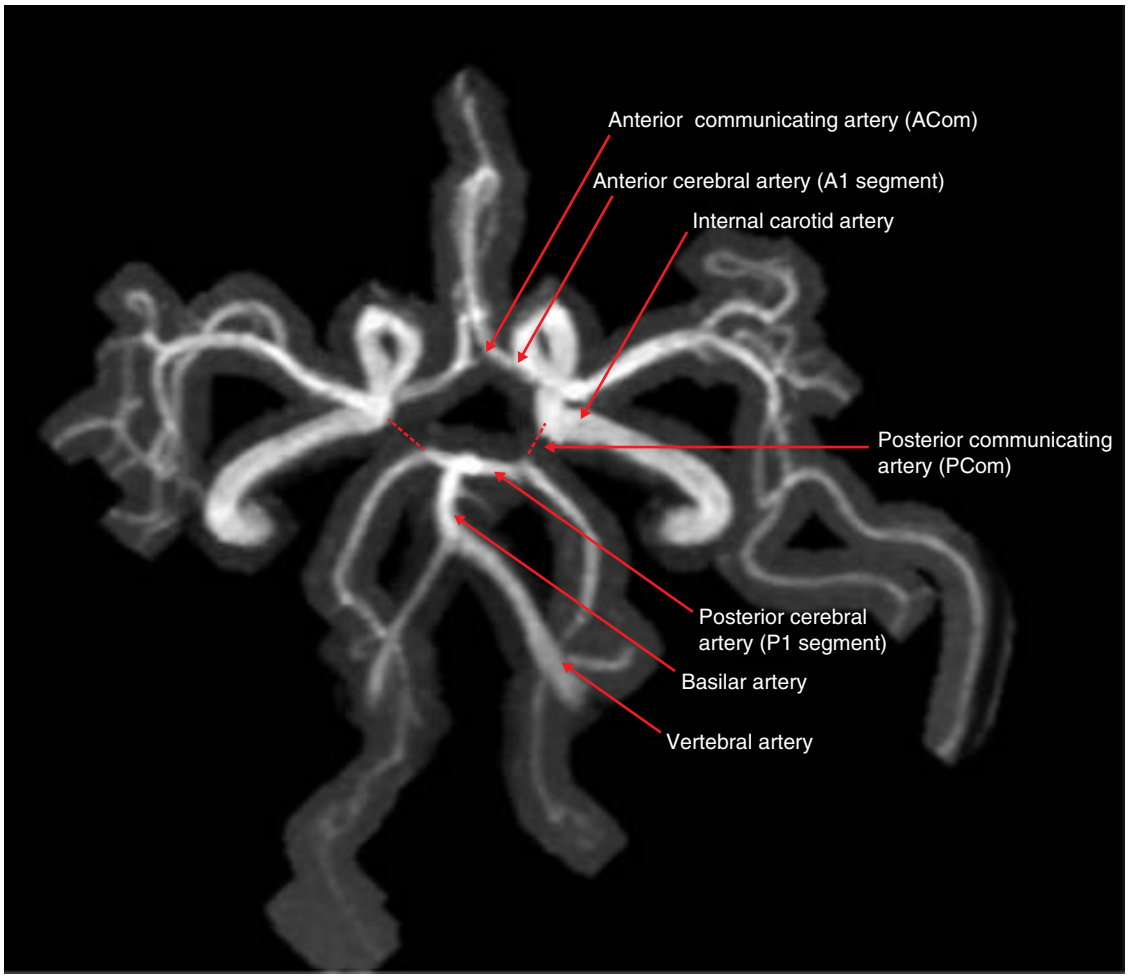

E-FIGURE 406-1. Magnetic resonance angiogram of the circle of Willis. The approximate positions of the posterior communicating arteries are shown with *dotted lines*.

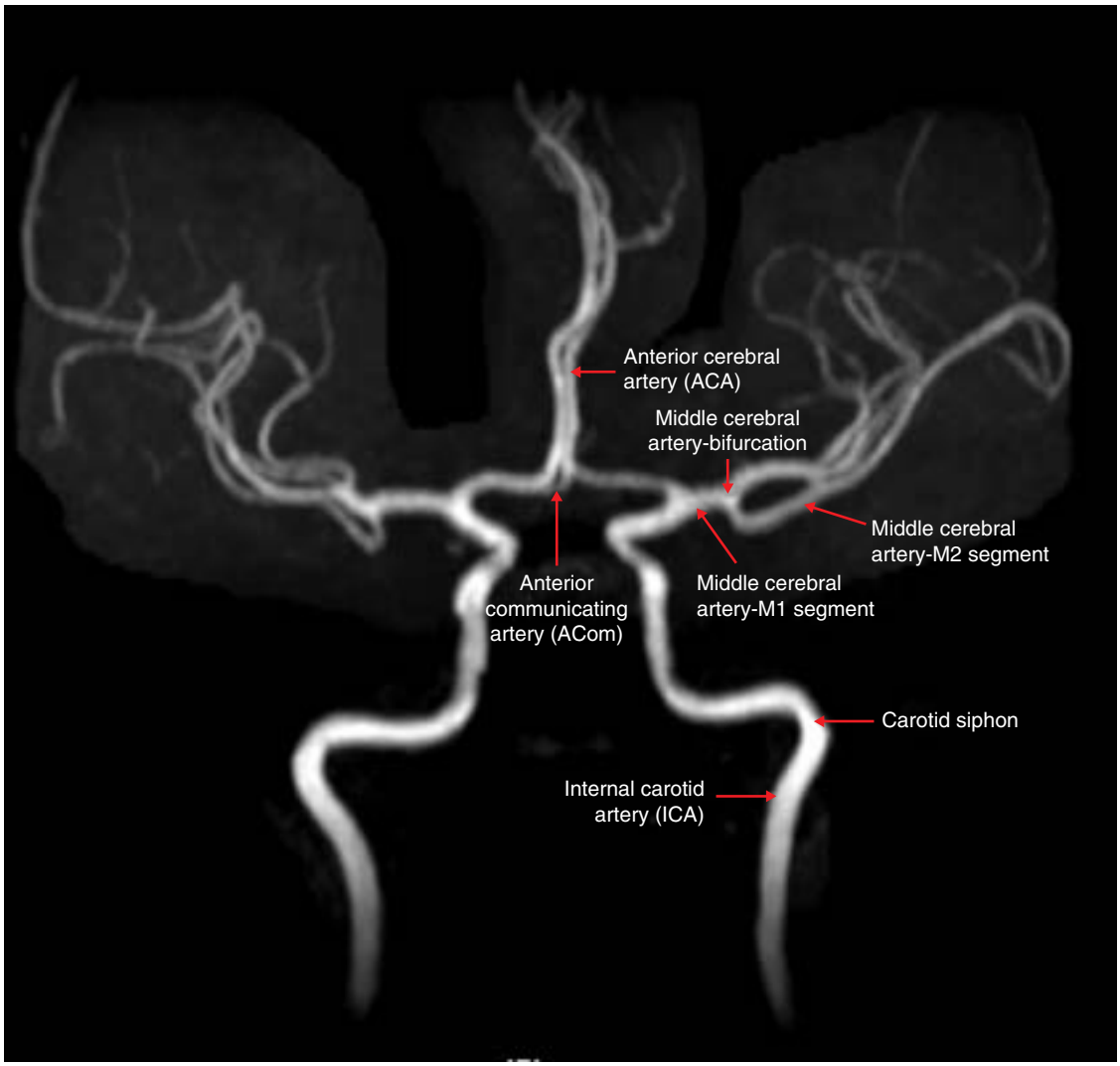

FIGURE 406-2. **Magnetic resonance angiogram of the intracranial portion of the internal carotid artery and its main branches.**

Basilar Artery

The basilar artery has small penetrating branches supplying the dorsal portions of the pons and midbrain (see Figs. 406-3 and 406-4). The anterior inferior cerebellar arteries originate from the mid–basilar artery. They supply portions of the cerebellar hemispheres in addition to the lateral pons; cranial nerves V, VII, and VIII; and pontine portions of the spinothalamic tracts and sympathetic fibers. The two superior cerebellar arteries arise from the distal basilar artery at the level of the midbrain proximal to the common origin of the two posterior cerebral arteries. The oculomotor nerve exits the midbrain between the superior cerebellar artery and posterior cerebral artery. The superior cerebellar arteries give branches supplying the dorsal midbrain, including the colliculi and the superior portions of the cerebellar hemispheres and vermis. The long circumferential vessels also supply the dorsolateral brain stem.

In addition to the anterior inferior cerebellar artery and superior cerebellar artery, the basilar artery has paramedian vessels supplying the middle portion of the basis pontis and midline pontine structures, including the corticospinal tracts, medial longitudinal fasciculus, and pontine reticular nuclei. At the midbrain level, paramedian branches of the basilar artery supply the cerebral peduncles, cranial nerve III nuclei and fibers, and medial portions of the red nucleus and medial lemniscus. Short circumferential branches supply the ventrolateral pons and midbrain.

Circle of Willis

The anastomosis at the base of the brain is termed the circle of Willis (see E-Fig. 406-1). The two anterior cerebral arteries are connected by the anterior communicating artery. The posterior communicating arteries connect the supraclinoid internal carotid arteries with the proximal posterior cerebral arteries. In persons with an intact circle of Willis, the entire intracranial circulation can be supplied by a single patent internal carotid artery or vertebral artery. The majority of individuals, however, have an incomplete circle of Willis (see Fig. 406-2). One common variant is for the portion of the anterior cerebral artery between the internal carotid artery and the anterior communicating artery (A1 segment) to be hypoplastic or absent. In this case, both anterior cerebral arteries can be supplied from a single internal carotid artery. Another common variant is for the portion of the posterior cerebral artery between its normal origin from the basilar artery and the posterior communicating artery (P1 segment) to be absent or hypoplastic (termed a "fetal" posterior cerebral artery). In these individuals, the distal posterior cerebral artery territory is supplied by the carotid rather than by the vertebrobasilar arteries.

Anterior Cerebral Arteries

The anterior cerebral arteries travel anteriorly and then turn posteriorly with leptomeningeal branches supplying the medial portions of the frontal and parietal lobes (Figs. 406-5 to 406-7; see also Fig. 406-2). In about half of people, the anterior cerebral artery divides into pericallosal and callosal marginal branches. Terminal portions of the latter artery supply the medial cortex between the parietal and occipital lobes. Damage to this area can be confused with "watershed" hypoperfusion injury. A series of small lenticulostriate arteries originate from the A1 and A2 (between the anterior communicating artery and corpus callosum) segments of the anterior cerebral artery. The recurrent artery of Heubner is a large, important medial striate artery that provides blood supply to the anterior and inferior portions of the anterior limb of the internal capsule, anterior and inferior portions of the caudate nucleus, anterior globus pallidus, putamen, hypothalamus, olfactory bulbs and tracts, and uncinate fasciculus. It can be inadvertently damaged during surgical clipping of an anterior communicating artery aneurysm.

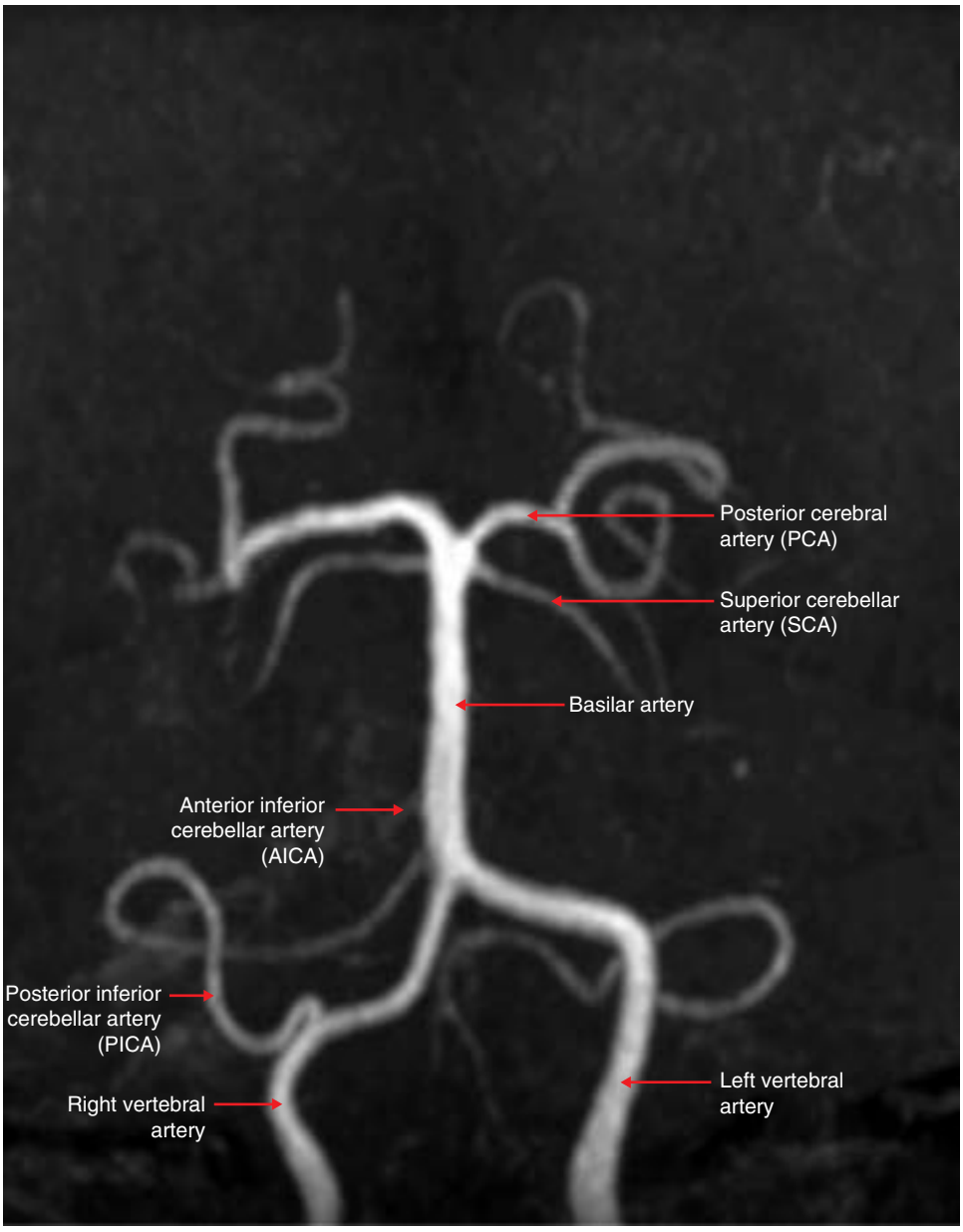

FIGURE 406-3. **Magnetic resonance angiogram of the intracranial portion of the vertebrobasilar system.**

Anterior Choroidal Artery

The anterior choroidal artery (medial striate artery) commonly arises from the supraclinoid internal carotid artery distal to the posterior communicating artery. It travels posteriorly over the medial optic tract and enters the brain at the choroidal fissure. It gives branches to the optic tract, anterior hippocampus, amygdala, tail of the caudate nucleus, geniculate body, and inferior portion of the posterior limb of the internal capsule (see Fig. 406-7). Ischemic lesions in this area can be confused with lesions arising from the middle cerebral artery.

Middle Cerebral Artery

The middle cerebral artery supplies the bulk of the frontal, parietal, and lateral portions of the temporal lobes (Figs. 406-8 and 406-9; see also Figs. 406-6 and 406-7). The M1 segment refers to the portion of the middle cerebral artery between its origin from the supraclinoid internal carotid artery and its distal branches (see Fig. 406-2). The middle cerebral artery bifurcates in the sylvian fissure in 20 to 30% of individuals and trifurcates in about 70% of individuals. The superior division supplies the frontal and parietal lobes, and the inferior division supplies the lateral portion of the temporal lobe. The M1 segment gives rise to some medial and all of the lateral lenticulostriate arteries. These arteries supply the head and body of the caudate nucleus, the putamen, and the globus pallidus as well as the anterior limb, genu, and superior portions of the posterior limb of the internal capsule (see Fig. 406-7).

Posterior Cerebral Artery

The distal portion of the posterior cerebral artery divides into an anterior and a posterior division (see Fig. 406-3). The anterior division supplies the inferior and medial portions of the temporal lobe into the middle cranial fossa with distal branches anastomosing with those of the middle cerebral artery (see E-Fig. 406-1). The posterior division supplies the occipital lobe, including the calcarine cortex, with terminal branches anastomosing with those of the middle cerebral artery and anterior cerebral artery. The proximal portions of both the posterior cerebral artery and the posterior communicating artery give off small penetrating arteries to the thalamus (thalamoperforators). In some individuals, a single common artery arising from the P1 segment (artery of Percheron) can supply both thalami. Unless the posterior cerebral artery has a fetal-type origin from the internal carotid artery, thalamic strokes are generally related to the vertebrobasilar circulation. Two posterior choroidal arteries arise separately from the posterior cerebral artery and supply the choroid plexus, posterior thalamus, fornix, and midbrain tectum. Posterior cerebral artery perforators also supply the medial portions of the cerebral peduncles, substantia nigra, red nuclei, hippocampus, and posterior hypothalamus.

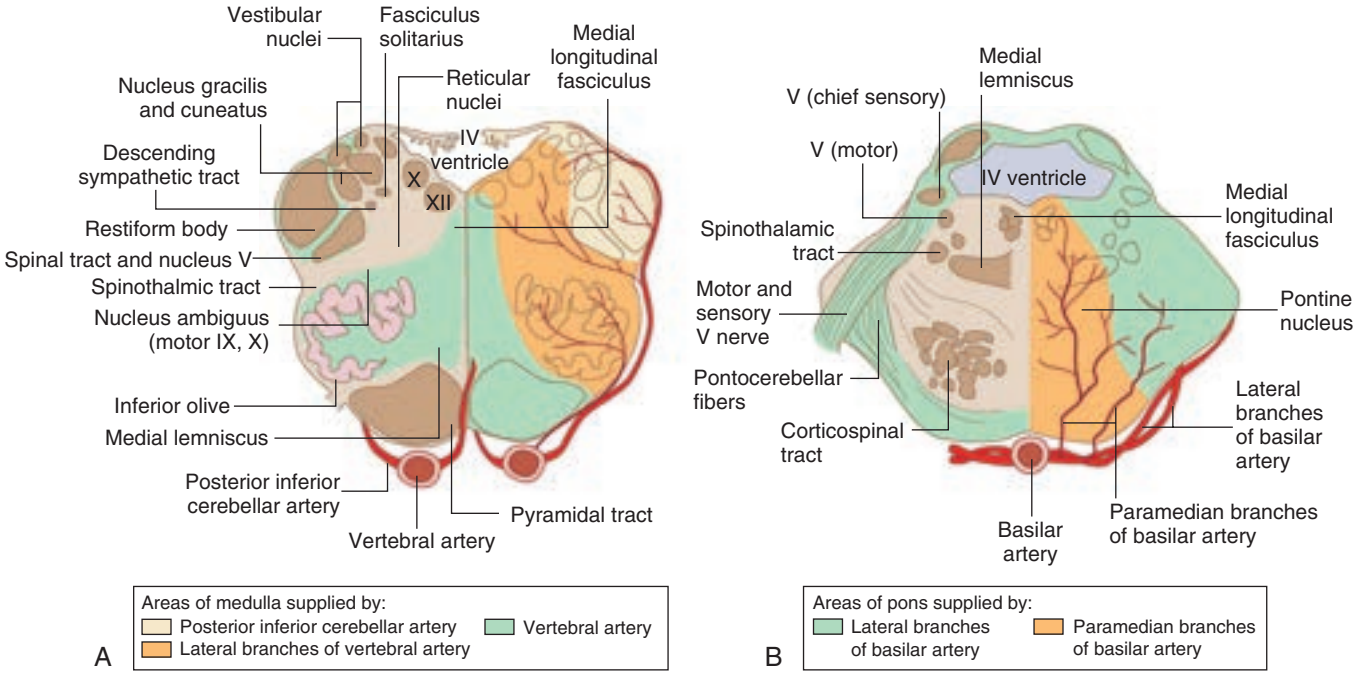

FIGURE 406-4. **Brain stem blood supply. A,** Cross section of the medulla oblongata at the level of the hypoglossal nuclei (cranial nerve XII). Short branches of the vertebral and anterior spinal arteries supply the medulla. Longer circumferential branches, including the posterior inferior cerebellar artery, supply the lateral portions of the medulla. **B,** Cross section of the midpons region. The medial portion receives blood supply from short, perforating basilar artery branches. More laterally, the blood supply comes from lateral basilar artery branches. (From Zivin JA. Approach to cerebrovascular diseases. In: Goldman L, Schafer AI. *Goldman's Cecil Medicine*. 24th ed. Philadelphia: Elsevier Saunders; 2012.)

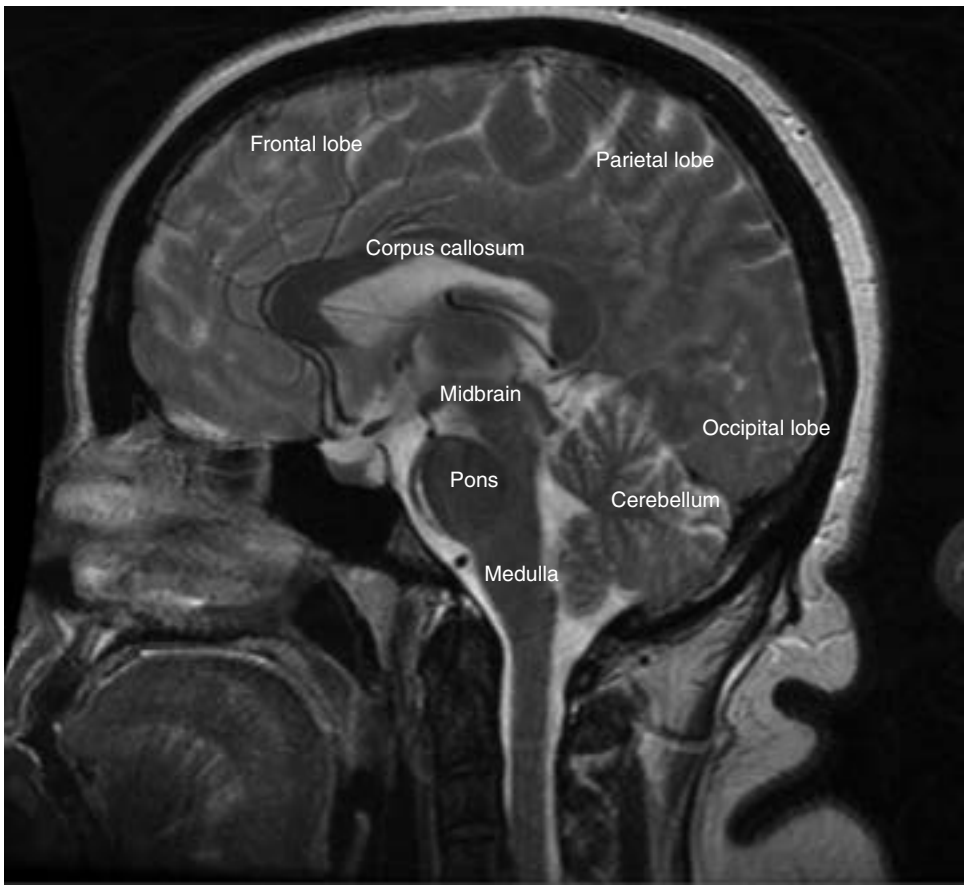

FIGURE 406-5. **Parasagittal T2-weighted magnetic resonance image showing midline structures.**

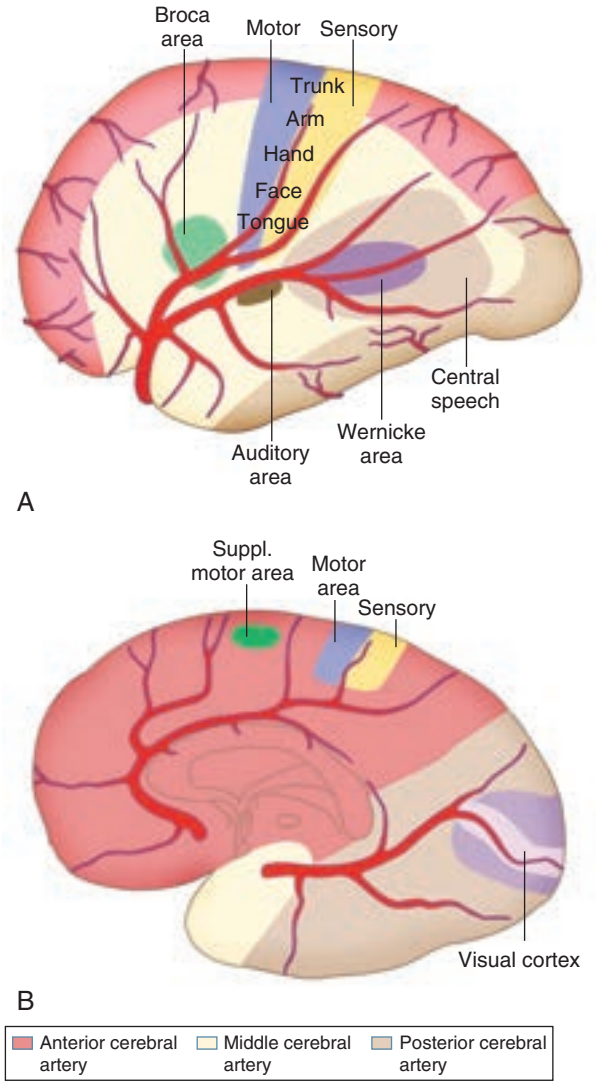

FIGURE 406-6. Surface cerebral arterial anatomy. Lateral (A) and medial (B) views of the cerebral hemisphere show the surface distributions of the anterior, middle, and posterior cerebral arteries. (From Zivin JA. Approach to cerebrovascular diseases. In: Goldman L, Schafer AI. *Goldman's Cecil Medicine*. 24th ed. Philadelphia: Elsevier Saunders; 2012.)

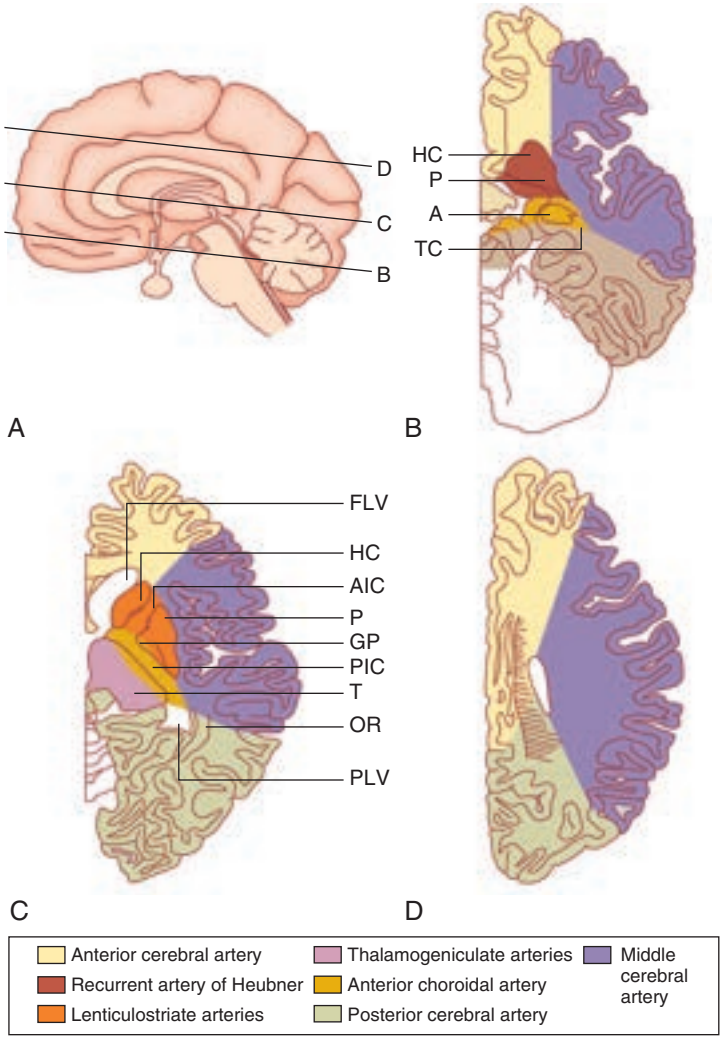

FIGURE 406-7. Arterial supply of the deep brain structures. A, Sagittal view of the brain showing the computed tomographic (CT) planes through which views **B, C,** and **D** were taken. **B,** CT plane through the head of the caudate nucleus (HC), putamen (P), amygdala (A), tail of the caudate nucleus (TC), hypothalamus, temporal lobe, midbrain, and cerebellum. **C,** CT plane through the frontal horn of the lateral ventricle (FLV), head of the caudate nucleus (HC), anterior and posterior limbs of the internal capsule (AIC, PIC), putamen (P), globus pallidus (GP), thalamus (T), optic radiations (OR), and posterior horn of the lateral ventricle (PLV). **D,** CT plane through the centrum semiovale. (Modified from De Armond S, Fusco MM, Dewey MM. *Structure of the Human Brain, a Photographic Atlas*. 3rd ed. New York: Oxford University Press; 1989, with permission.)

Venous System

The venous drainage of the brain is divided into superficial and deep systems (Figs. 406-10 and 406-11). Deep structures drain into the inferior sagittal sinus and vein of Galen that join to form the straight sinus, which runs along the tentorium to join the superior sagittal sinus at the torculum. The cerebral veins drain into the sagittal sinus. The two transverse sinuses extend laterally from the torculum into the sigmoid sinus, which then forms the jugular vein. Oftentimes, one hypoplastic transverse sinus can cause confusion if a sinus thrombosis is suspected. In these cases, the jugular notch in the occipital bone and jugular foramen may be smaller on the side of the hypoplastic transverse sinus. Each cavernous sinus surrounds the ipsilateral internal carotid artery. Fibers from cranial nerve VI run within the cavernous sinus inferior to the carotid artery, with fibers from cranial nerves III, IV, V1, and V2 running in its lateral wall. The two cavernous sinuses connect to each other and drain into the petrosal sinus and then the sagittal sinus.

Physiology

Cerebral Blood Flow

The brain, which is among the body's most metabolically active tissues, receives about 14% of resting cardiac output. Normal resting metabolism of brain tissue requires 140 µmol of oxygen and 24 µmol of glucose per 100 g of tissue per minute. Although total blood flow to the brain remains constant in normal conditions, regional flow changes with mental activity, often manifested by changes in synaptic activity, and provides the basis for functional magnetic resonance imaging or positron emission tomography imaging studies. Approximately 80% of glucose is used to generate energy, with the remainder metabolized to lactate or used for synthetic activities. Little glucose is stored in the brain, and the brain's high metabolic demand makes it particularly vulnerable to reductions in oxygen and blood supply. Cerebral blood flow at rest averages 50 to 100 mL per 100 g of brain tissue per minute. If blood flow falls below this level, normal neuronal function is suppressed (i.e., neurons become electrically quiescent). If the deficit persists, irrevocable neural injury can result.

Cerebral blood flow is regulated though a variety of mechanisms in addition to mental activity. Constant, overall cerebral blood flow is maintained through autoregulation. This autoregulatory relationship is reflected in the equation cerebral blood flow = cerebrovascular resistance/mean arterial pressure. If the mean arterial pressure is decreased, there is a compensatory decrease in cerebrovascular resistance (through dilation of cerebral arterioles) to maintain cerebral blood flow constant. If the mean arterial pressure is increased, there is a compensatory increase in cerebrovascular resistance (through constriction of cerebral arterioles). There are, however, limits to cerebral autoregulation. At mean arterial pressures greater than about 150 mm Hg, cerebral arterioles are maximally constricted, and cerebral blood flow rises. At mean arterial pressures below about 50 mm Hg, cerebral arterioles are maximally dilated, and cerebral blood flow falls. In the setting of

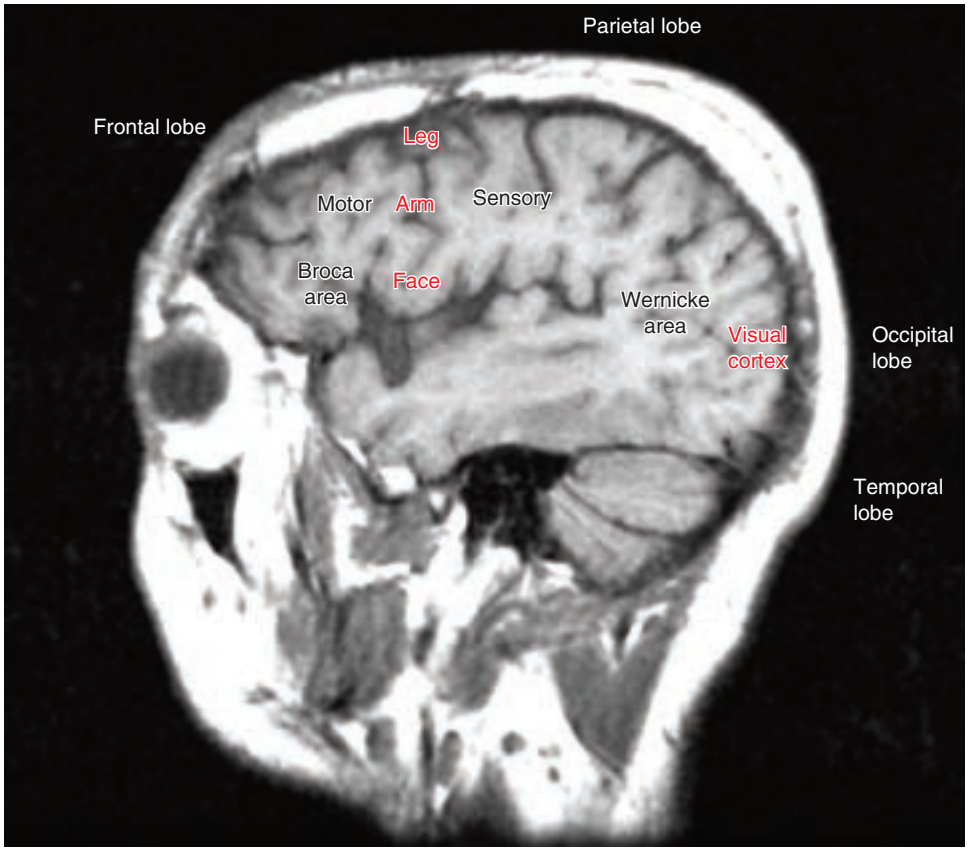

FIGURE 406-8. Sagittal, lateral T1-weighted magnetic resonance image showing cortical motor, sensory, visual, and language areas.

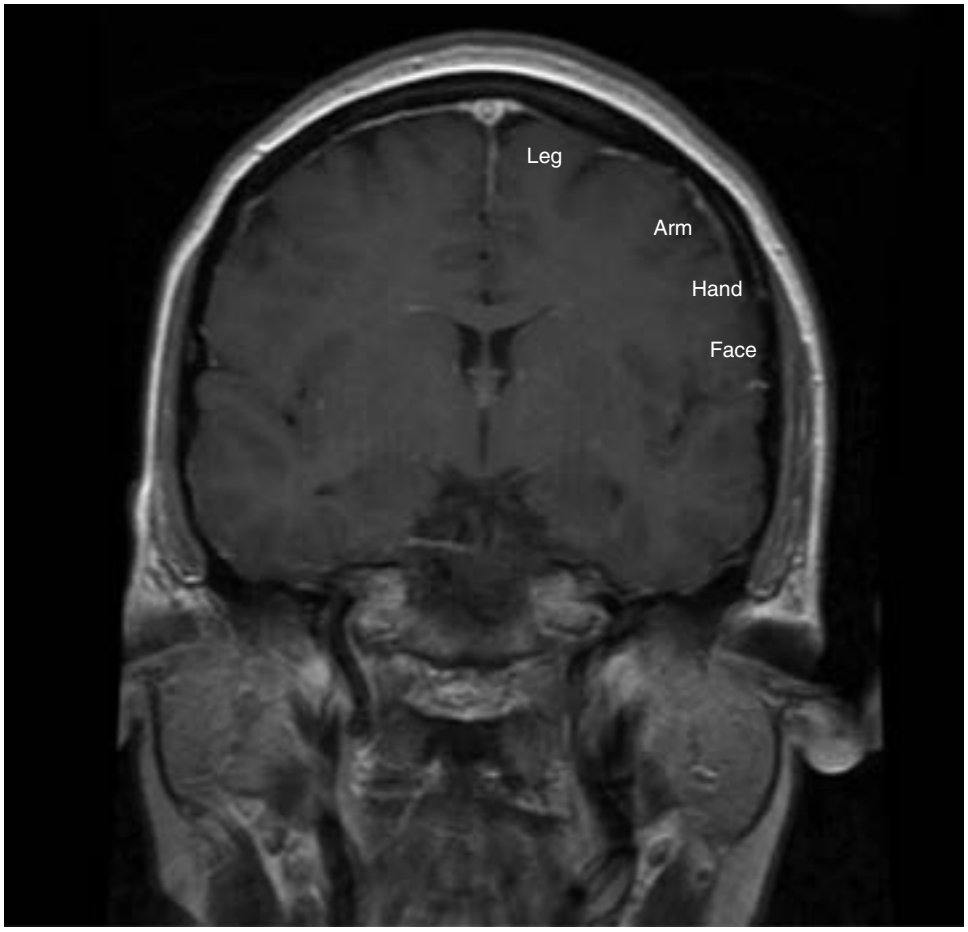

FIGURE 406-9. Coronal T1-weighted magnetic resonance image showing cortical areas for the leg, arm, hand, and face.

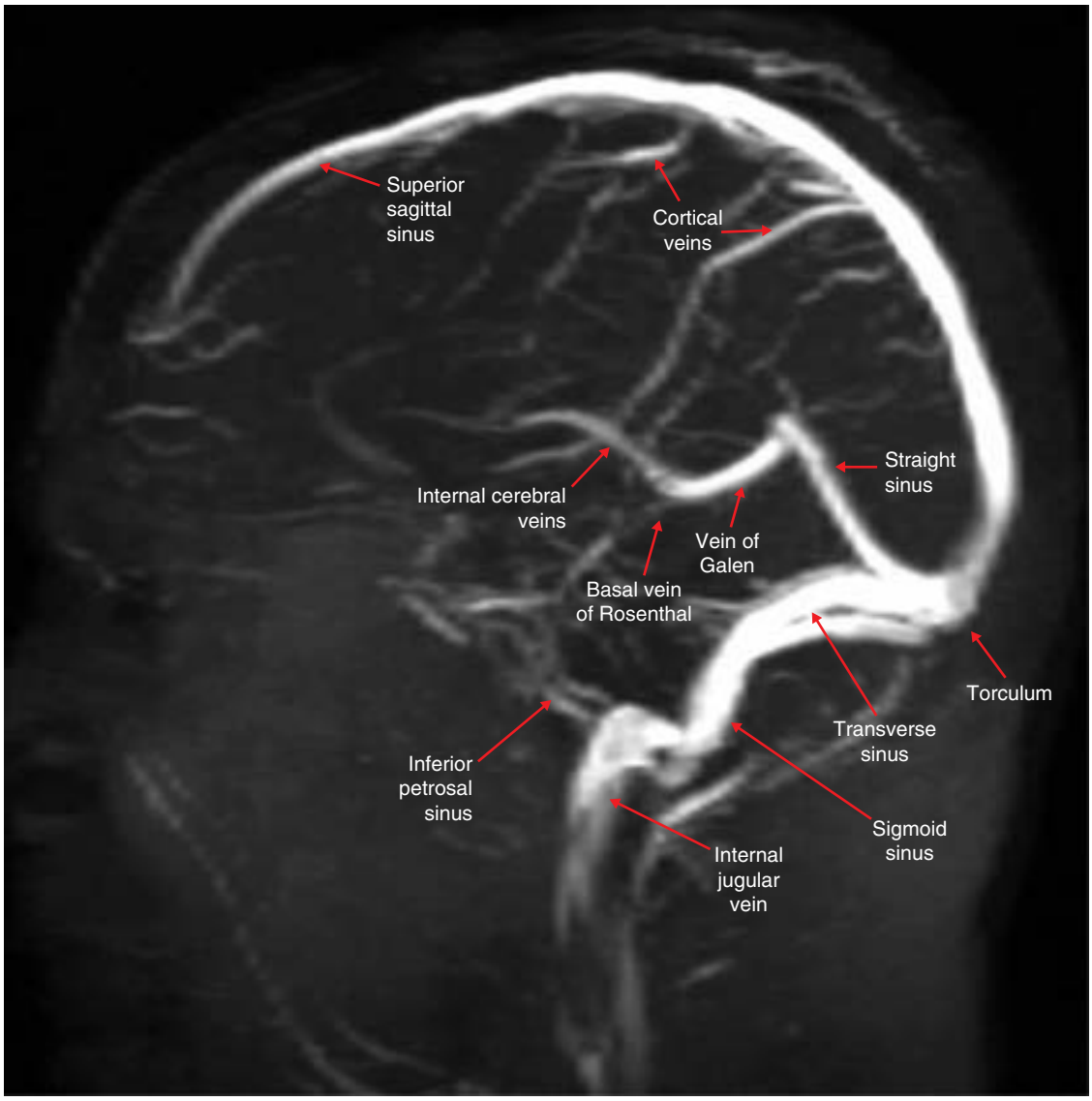

FIGURE 406-10. **Parasagittal magnetic resonance venogram showing venous structures.**

chronic hypertension, the autoregulatory relationship between cerebrovascular resistance and mean arterial pressure is shifted to higher critical mean arterial pressures (i.e., cerebral blood flow falls at a higher mean arterial pressure).

Metabolic factors can also affect cerebral blood flow. Hypercapnia causes cerebral vasodilation, and hypocapnia causes cerebral vasoconstriction that is mediated by changes in the pH of the brain's extracellular fluid. Cerebral blood flow declines by approximately 2% for every 1 mm Hg decline in P_{CO_2}. In patients who have increased intracranial pressure and threatened herniation, a short period of hyperventilation (target arterial P_{CO_2} of 30 to 35 mm Hg) can be used as a temporary measure until more definitive treatment can be instituted. The response is only transient because of compensation by the choroid plexus, and a rebound increase in P_{aCO_2} can lead to a rise in intracranial pressure when hyperventilation is discontinued.

Blood-Brain Barrier

The triggering of a neuronal action potential depends on the relative concentrations of Na^+, K^+, and Ca^{2+}, and it is also modulated by Mg^{2+} and a variety of neurotransmitters. The blood-brain barrier is critical for maintaining the environment necessary for normal neuronal function.[4] The blood-brain barrier consists anatomically of the capillary endothelial cells, a basement membrane with pericytes, and astrocytic perivascular footplates. The brain's vascular endothelial cells, which are the principal component of the blood-brain barrier, are joined by tight junctions and generally lack the transport channels found elsewhere in the body. As a result, the blood-brain barrier prevents hydrophilic polar and large molecules in the blood from entering the brain. By comparison, oxygen and carbon dioxide rapidly cross the blood-brain barrier. Nutrients, toxins, and drugs can cross the blood-brain barrier by simple diffusion, by transport through carrier molecules based on concentration gradients (facilitated transport), or by energy-dependent mechanisms (active transport). Glucose is the brain's sole source of energy. Glucose transport into the brain is through non–energy-dependent facilitated transport (glucose transporter isotype 1, Glut1). In the setting of ischemia, endothelial cell function can be compromised, and the blood-brain barrier can fail.

The Neurovascular Unit

The concept of the neurovascular unit has become important for understanding the complex relationships between anatomic structures and the integrity of brain function. The term reflects the physiologic interrelatedness of the brain's various components, including endothelial cells, vascular smooth muscle, adventitial cells, glia, and neurons. The concept reflects the observation that local pH as well as neural activity can affect local cerebral blood flow. In addition to linking neural activity with blood flow and maintaining the blood-brain barrier, the neurovascular unit can secrete a variety of immunologic and neurotrophic factors that further affect both normal function and the brain's response to injury.

CEREBRAL ISCHEMIA

Because of its high metabolic demands, brain function is completely dependent on its supply of blood and oxygen. Clinical symptoms ensue when global or regional blood supply falls below the critical 50 mL per 100 g per minute. Permanent neural injury does not occur if the supply of blood and oxygen is

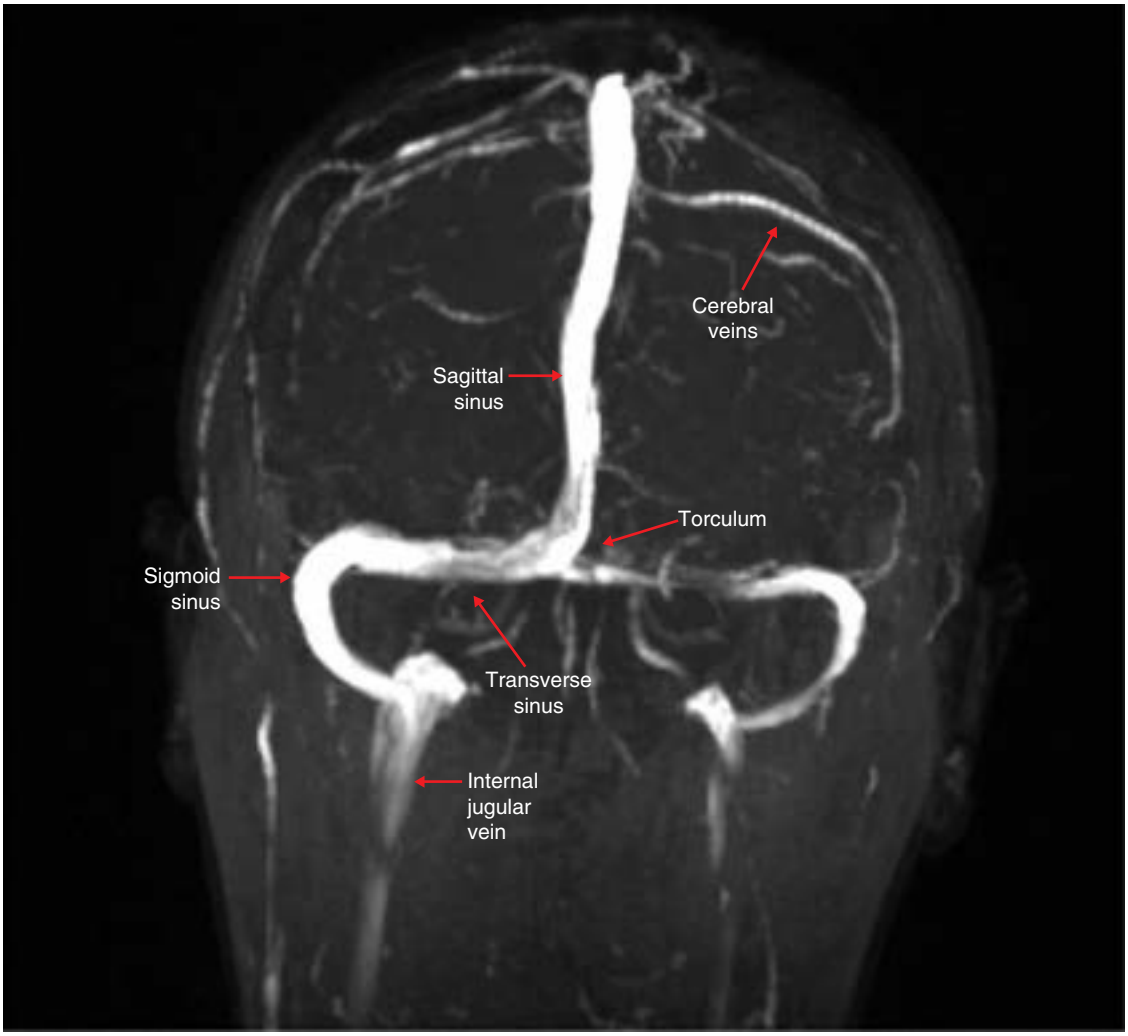

FIGURE 406-11. Anteroposterior magnetic resonance venogram showing venous structures.

quickly restored, such as with a faint (Chapter 62) in the setting of a global reduction in the supply of blood or oxygen or a transient ischemic attack (Chapter 407) with brief, local reductions in cerebral blood flow. Certain groups of neurons may be particularly vulnerable to hypoxic-ischemic injury (i.e., regions of the hippocampus, cerebellar Purkinje cells, and neocortical layers III, V, and possibly VI). Hypoxic-ischemic injury can be global, diffuse, or focal.

Global Ischemic Injury

Global ischemic injury occurs in the setting of complete cardiovascular collapse, such as with ventricular fibrillation, electromechanical dissociation, and asystole (Chapter 63). Some neurons are particularly vulnerable to ischemic injury and will be selectively damaged, whereas neurons only millimeters away may be spared.[5] In the setting of hypotension, areas of brain between the territories of major arteries (i.e., between the anterior cerebral artery and middle cerebral artery in the frontal cortex and adjacent subcortical white matter), between the middle cerebral artery and posterior cerebral artery (in the parieto-occipital cortex and adjacent subcortical white matter), and between penetrating arteries from distal branches of the middle cerebral artery and lenticulostriate arteries (deep hemispheric white matter, centrum semiovale) are especially vulnerable and are termed watershed areas.

The duration of anoxia, the duration of cardiopulmonary resuscitation (CPR), and the cause of cardiac arrest are related to poor outcome after CPR (Chapters 63 and 404), but none of these factors accurately discriminate between poor and favorable outcomes. Prognosis also cannot be based on the circumstances of CPR or on elevated body temperature alone. Myoclonus or status epilepticus within the first day after cardiac arrest implies a poor prognosis, as does the absence of pupillary or corneal reflexes or extensor motor responses 3 days after cardiac arrest in patients who remain comatose. Bilateral absence of cortical somatosensory evoked responses within 1 to 3 days also portends a poor prognosis.

Out-of-hospital cardiac arrest carries a poor prognosis if effective CPR is not rapidly instituted. A period of therapeutic hypothermia may improve neurologic outcome after resuscitated cardiac arrest if it can be instituted rapidly (Chapter 63). If the cerebral cortex is irreversibly damaged but the relatively resistant brain stem control of respiration and cardiovascular regulation is preserved, the patient can enter a persistent vegetative state (Chapter 404).

Diffuse Hypoxic Injury

Diffuse hypoxia can alter cognition, cause confusion, impair consciousness, and lead to coma, which can be irreversible. Causes include travel to high altitudes, severe anemia, and pulmonary disease. Symptoms are generally present when the Pao_2 abruptly falls to less than 40 mm Hg. Increases in cerebral blood flow can partially compensate for slow declines in Pao_2, which may still cause symptoms with further or rapid reductions.

Focal Ischemic Injury

Focal ischemic injury is caused by occlusion of a cervical or intracranial artery that supplies the brain. Although this injury can occur from many causes (including infection, inflammation, metabolic disorders, trauma, and hematologic disorders), the majority of strokes are related to thrombotic or embolic occlusion (Fig. 406-12). If flow is not restored within minutes, a core area of irreversible brain injury is commonly produced. A surrounding area of variable size, depending on the artery involved and the integrity of collaterals in which blood flow is reduced, will suffer injury that is not irreversible. The brain in this area, termed the penumbra, is electrically quiescent and contributes to the resulting neurologic deficit. Because the pH of the extracellular fluid in the penumbral zone is low, vessels are maximally dilated and the cerebral autoregulatory response is inoperative. Because cerebrovascular resistance in the penumbral zone is fixed, any decline in mean arterial pressure can further reduce its cerebral blood flow, thereby extending the volume

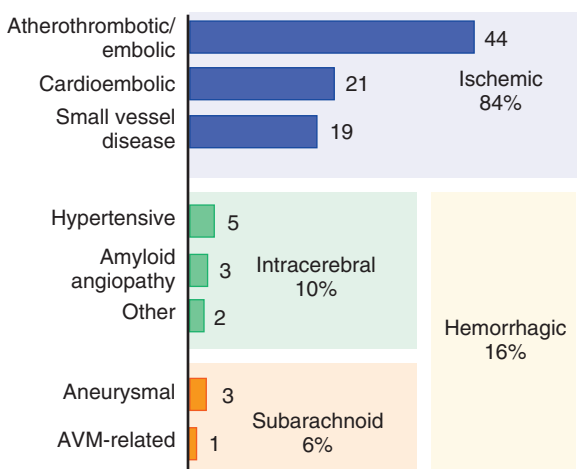

FIGURE 406-12. Classification of cerebrovascular disease by cause. AVM = arteriovenous malformation. (From Zivin JA. Approach to cerebrovascular diseases. In: Goldman L, Schafer AI. *Goldman's Cecil Medicine*. 24th ed. Philadelphia: Elsevier Saunders; 2012.)

of infarcted brain tissue. A variety of neuroimaging techniques can help distinguish penumbral from infarcted brain tissue (i.e., magnetic resonance diffusion-perfusion mismatch, computed tomographic perfusion imaging) but have not been standardized and have not yet proved useful for clinical decisions regarding the use of reperfusion therapy. Many putative neuroprotective strategies aimed at preserving ischemic brain tissue until it can be reperfused through collateral flow have failed in clinical trials.

PATHOLOGY

Permanent occlusion of a cerebral artery results in necrosis of its supplied neurons, glia, and endothelial cells (pan-necrosis). In gross appearance, the area of infarcted brain may be pale or hemorrhagic if secondary bleeding occurred. Over time, the lesion becomes cavitary (encephalomalacia). On microscopic examination, ischemic neurons initially appear small and angular. The cytoplasm becomes homogeneously eosinophilic, and the nucleus becomes dark and pyknotic. As endothelial cells die, associated areas of petechial hemorrhage may appear. An initial inflammatory reaction may lead to microvascular occlusions, such that flow to ischemic tissue may not be restored even if a proximal thrombus is removed (no-reflow phenomenon). Leukocytes that infiltrate ischemic tissue can also release interleukins and cytokines, which can contribute to cytotoxic injury. Blood macrophages begin to reach the infarcted tissue, and neovascularization peaks after about 2 weeks. Macrophage-mediated removal of cellular debris peaks at about 3 to 4 weeks after the infarct. Astrocytes then form a glial scar around the area of infarction.

PATHOPHYSIOLOGY

Because the brain has no reserve energy supply, energy-dependent neuronal and glial processes stop soon after acute deprivation of blood and oxygen. Calcium ions enter depolarized neurons and glia, where they activate second messengers, including lipases and proteases, thereby releasing free fatty acids and generating free radicals that degrade cellular organelles and membranes. Depolarized neurons also release high levels of excitatory neurotransmitters, such as glutamate into synapses, which leads to further neuronal depolarization and calcium entry. Once this cascade has been initiated, neurons may still degenerate over time by apoptosis (programmed cell death) even if blood flow is restored. Although promising in the laboratory, all attempts to block the ischemic cascade pharmacologically have failed in clinical trials to date.

CEREBRAL HEMORRHAGE

Subarachnoid hemorrhage, which is bleeding between the pial and arachnoid coverings over the brain, is most commonly related to a ruptured aneurysm (Chapter 408).[6] Cerebral aneurysms may occur spontaneously or be acquired as a result of infection or trauma. They are more common in first-degree relatives of patients who have a cerebral aneurysm and with certain conditions, such as autosomal dominant polycystic kidney disease (Chapter 127) and type IV Ehlers-Danlos syndrome (Chapter 260). Noninfectious aneurysms are typically situated at branch points of major cerebral arteries: anterior cerebral artery–anterior communicating artery, internal carotid artery–posterior communicating artery, middle cerebral artery bifurcation, basilar artery tip. Initial brain injury can be caused by an acute increase in intracranial pressure, with delayed ischemic injury related to the development of vasospasm after 7 to 10 days. Interference with the absorption of cerebrospinal fluid though the arachnoid granulations can lead to communicating hydrocephalus. Clot within the third or fourth ventricle or cerebral aqueducts can cause obstructive hydrocephalus.

The most common causes of intracerebral parenchymal brain hemorrhages are hypertension (Chapter 67) and cerebral amyloid angiopathy. Myriad other potential vascular and nonvascular causes, including vascular malformations, vasculitis (Chapter 270), venous sinus thrombosis, and coagulopathies (Chapters 172, 173, and 174), are less common. Some tumors (e.g., melanoma [Chapter 203] and renal cell carcinoma [Chapter 197]) can be manifested as an intracerebral hemorrhage. Hypertension-related intracerebral hemorrhage occurs in typical areas of the brain (i.e., basal ganglia, thalamus, basis pontis, and cerebellum). In contrast, intracerebral hemorrhage related to cerebral amyloid angiopathy is typically lobar and located closer to the cortical surface. Without sequential neuroimaging studies showing an initial area of ischemic injury, lobar hemorrhages may be difficult to distinguish from a hemorrhagic infarction. Susceptibility-weighted brain magnetic resonance imaging sequences may reveal prior microhemorrhages at the gray-white junction in patients with cerebral amyloid angiopathy.

CEREBRAL EDEMA

When neurons and glia are injured by ischemia, energy metabolism fails and the cells can no longer maintain normal ion gradients between the intracellular and extracellular compartments. The result is cytotoxic edema, in which cells swell soon after the injury. Neurons, glia, and endothelial cells can be affected. Vasogenic edema, which may occur as a result of disruption of the blood-brain barrier due to injury to the endothelium, allows large molecules to pass through the blood-brain barrier and to gain access to the brain. Edema generally peaks between 48 and 72 hours after the onset of ischemic injury. In patients with ischemic stroke, the development of cytotoxic edema can lead to an increase in intracranial pressure and, when severe, herniation. In selected patients, craniotomy can be considered to relieve the pressure until the edema subsides.

Neurons, glia, and endothelial cells are also damaged in the setting of intracerebral hemorrhage. The hemorrhage itself is a space-occupying lesion that can also be associated with both cytotoxic and vasogenic edema. Mass effect from cerebellar hemorrhages can compress the fourth ventricle (thereby leading to obstructive hydrocephalus), compress the brain stem (thereby compromising the reticular activating system and impairing consciousness), or cause herniation. Surgical evacuation of intracerebral hemorrhage is not of proven value (Chapter 408), but emergent evacuation of cerebellar hemorrhages can be life-saving and leave surviving patients with little or no long-term functional impairment.

Grade A Reference

A1. Arrich J, Holzer M, Havel C, et al. Hypothermia for neuroprotection in adults after cardiopulmonary resuscitation. *Cochrane Database Syst Rev*. 2012;9:CD004128.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Feigin VL, Forouzanfar MH, Krishnamurthi R, et al. Global and regional burden of stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. *Lancet*. 2014;383:245-254.
2. Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics—2014 update: a report from the American Heart Association. *Circulation*. 2014;129:e28-e292.
3. Goldstein LB, Bushnell CD, Adams RJ, et al. Guidelines for the primary prevention of stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42:517-584.
4. Engelhardt B, Liebner S. Novel insights into the development and maintenance of the blood-brain barrier. *Cell Tissue Res*. 2014;355:687-699.
5. Baron JC, Yamauchi H, Fujioka M, et al. Selective neuronal loss in ischemic stroke and cerebrovascular disease. *J Cereb Blood Flow Metab*. 2014;34:2-18.
6. Chalouhi N, Hoh BL, Hasan D. Review of cerebral aneurysm formation, growth, and rupture. *Stroke*. 2013;44:3613-3622.

REVIEW QUESTIONS

1. Which of the following conditions is associated with the highest population-attributable risk of stroke?

A. Cigarette smoking
B. Diabetes
C. Hypertension
D. Poor diet
E. Sedentary lifestyle

Answer: C Each of the listed conditions and lifestyle factors is associated with an increase in the risk of stroke. Of these, hypertension has the highest population-attributable risk for stroke.

2. The left common carotid artery most commonly arises from which of the following structures?

A. Right subclavian artery
B. Innominate artery
C. Aortic arch
D. Left subclavian artery
E. Thyrocervical trunk

Answer: C The left common carotid artery most commonly arises directly from the aortic arch. In some individuals, it may arise from the proximal portion of the brachiocephalic trunk ("bovine" anatomy).

3. Which of the following arteries is commonly the first intradural branch of the internal carotid artery?

A. Posterior communicating artery
B. Ophthalmic artery
C. Anterior choroidal artery
D. Anterior cerebral artery
E. Recurrent cerebral artery (Heubner)

Answer: B The ophthalmic artery may rarely arise from the carotid siphon, but it is generally the first branch after the internal carotid artery pierces the dura, followed by the posterior communicating artery and the anterior choroidal artery. The recurrent cerebral artery is generally a branch of the anterior cerebral artery.

4. Which of the following equations describes the normal cerebral autoregulatory relationship (CBF, cerebral blood flow; CVR, cerebral vascular resistance; MAP, mean arterial pressure)?

A. CVR = MAP − CBF
B. MAP = CBF − CVR
C. CBF = MAP/CVR
D. MAP = CVR − CBF
E. CBF = CVR/MAP

Answer: E If MAP is decreased, there is a compensatory decrease in CVR (through dilation of cerebral arterioles) to maintain CBF constant. If MAP is increased, there is a compensatory increase in CVR (through constriction of cerebral arterioles). In the setting of ischemia, cerebral arterioles are maximally dilated; decreased MAP leads to a reduction in CBF to ischemic brain.

407 ISCHEMIC CEREBROVASCULAR DISEASE

LARRY B. GOLDSTEIN

DEFINITION

Ischemic cerebrovascular disease is caused by an impairment of blood supply to the brain. The injury may be focal (related to occlusion of a single artery), multifocal (related to occlusion of several arteries), or diffuse. Although certain clinical features (e.g., severe hypertension, headache, impaired consciousness) may suggest brain hemorrhage (Chapter 408) rather than ischemia, it is not possible to differentiate the two sets of conditions without a brain imaging study. In the absence of an inflammatory disease such as vasculitis or other rare conditions, simultaneous involvement of more than one vascular distribution suggests a proximal source of embolism (i.e., a cardiogenic or a proximal arterial source). Involvement of a single vascular territory may be due to either local steno-occlusive disease (e.g., atherosclerosis) or a proximal source of embolism. Involvement in the distribution of a single penetrating artery suggests small-vessel type intracranial disease, but ischemic strokes in this distribution may also be caused by proximal arterial steno-occlusive disease or embolism.

The definition of ischemic stroke is brain, spinal cord, or retinal cell death attributable to ischemia with neuropathologic, neuroimaging, or clinical evidence of permanent injury.[1] Overall, approximately 85% of strokes are related to ischemic disease, with 44% attributable to atherosclerosis (Chapter 70), 21% to cardiogenic embolism, and 20% to small-vessel disease.

Transient ischemic attack (TIA) is defined as a brief episode of neurologic dysfunction resulting from focal cerebral ischemia with no evidence of corresponding tissue injury. Symptoms are similar to those of ischemic stroke. Previously characterized as a transient deficit with symptoms persisting for less than 24 hours, evidence of corresponding tissue injury can be seen on brain magnetic resonance imaging (MRI) in 30 to 40% of patients who otherwise fulfill the clinical definition of TIA.

EPIDEMIOLOGY

Stroke (ischemic and hemorrhagic) is the second leading cause of death worldwide and the fourth leading cause of death in the United States.[2] It is also one of the leading causes of adult disability. In addition to age, race or ethnicity, and family history, a variety of lifestyle factors and medical conditions increase the risk of stroke (Chapter 406; Table 407-1). Of these, hypertension is the single most important (Chapter 67; Table 407-1), and the risk of stroke increases with increasing blood pressure with no threshold effect. Diabetes (Chapter 229) is associated with an approximate doubling of the risk of stroke (Table 407-1). Atrial fibrillation (Chapter 64; Table 407-1) is associated with up to 25% of ischemic strokes, with the absolute risk varying by concomitant risk factors.

Extracranial carotid artery stenosis is found in up to 5 to 10% of individuals older than 65 years and is associated with about 10% of all ischemic strokes. Untreated asymptomatic carotid stenosis carries only about a 1 to 2% annual risk of stroke, and the risk may now be much lower, perhaps as low as 0.5% annually, with standard medical therapy. Stroke is also a complication of sickle cell disease (Chapter 163), with risk dramatically reduced with transfusion therapy in high-risk children. Unlike with coronary heart disease, the overall association between high cholesterol concentration and the risk of stroke is less certain. Ischemic stroke risk is associated with higher levels of total cholesterol, whereas the risk of hemorrhagic stroke is increased with lower cholesterol levels.

Other factors associated with the risk of stroke include migraine headaches with aura (Chapter 398), particularly in women who smoke and are receiving oral contraceptives; elevated homocysteine level; high lipoprotein (a) level; postmenopausal hormone replacement therapy (Chapter 240); coagulation disorders (Chapter 176); systemic infection (Chapter 76); renal impairment (Chapter 130); low vitamin D levels (Chapters 218 and 244); and a variety of environmental factors, including high levels of air pollution.

TABLE 407-1 COMMON STROKE RISK FACTORS

FACTOR	POPULATION-ATTRIBUTABLE RISK	RISK REDUCTION WITH TREATMENT
LIFESTYLE		
Cigarette smoking	12-14%	50% within 1 year of quitting
Physical inactivity	30%	?
Excess alcohol consumption	7%	?
MEDICAL		
Hypertension	>90%	32%
Diabetes	5-27%	—
Atrial fibrillation	2-24%	64%
Carotid stenosis	2-7%	50%
Sickle cell disease	—	91% with transfusion therapy in children

Data from Goldstein LB, Bushnell CD, Adams RJ, et al. Guidelines for the primary prevention of stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* 2011;42:517-584.

PATHOBIOLOGY

For patients who have a TIA and who are by definition at increased risk of having an ischemic stroke during the next few days or weeks or who have an ischemic stroke, distinguishing among the major pathophysiologic causes (i.e., atherothrombotic, cardioembolic, small vessel) is critical to guide secondary prevention. Atherothrombosis due to atherosclerosis (Chapter 70) is the most common cause of a TIA or stroke that is related to steno-occlusive disease in a single artery.[3] The ischemia may be caused when progressive stenosis at the site of an atherosclerotic plaque leads to hemodynamic compromise affecting distal brain tissue. Sometimes bleeding into the plaque can lead to abrupt arterial occlusion, and sometimes a thrombus that has formed on an ulcerated plaque may embolize and occlude a distal artery. Occlusion of a cerebral artery, however, does not necessarily lead to ischemic brain injury. Blood may still reach the supplied territory through collaterals, either through the circle of Willis or from extracranial-intracranial anastomoses (see E-Fig. 406-1).

Arterial dissection, previously thought to occur only rarely and usually to result in major stroke, is now recognized more frequently on the basis of noninvasive vascular imaging such as MR angiography or computed tomography (CT) angiography. Other arteriopathies, such as fibromuscular dysplasia (Chapters 67, 80, and 125), may also lead to single, large-vessel distribution, ischemic stroke. Atherosclerosis of the ascending aorta or aortic arch can lead to the formation of thrombus, which can then embolize to a cerebral artery.

Atrial fibrillation is the single most common cause of cardioembolic stroke, with annual risks of 3 to 5% if it is not treated with anticoagulation but declining to about one fourth of that risk with anticoagulation (Chapter 64). The use of extended cardiac rhythm monitoring (i.e., 30-day event-triggered loop monitoring; Chapter 62) reveals occult atrial fibrillation in up to 25% of patients with an otherwise cryptogenic stroke. Other cardiac causes of cerebral embolism include clots or vegetations in patients with valvular heart disease (Chapter 75), such as mechanical prosthetic heart valves (Chapter 75), infectious endocarditis (Chapter 76), and nonbacterial endocarditis (Chapter 76); and mural thrombi in patients with a cardiomyopathy (Chapter 60) or myocardial infarction (MI), particularly anteroseptal MI (Chapter 73). Paradoxical embolism of a venous clot across a congenital heart defect, such as a patent foramen ovale or an atrial septal defect (Chapter 69), is another potential cause of embolic stroke.[4]

Small-vessel intracranial disease may result in ischemic stroke in the distribution of a single penetrating vessel. These strokes commonly affect deep structures (e.g., centrum semiovale, basal ganglia, thalamus, internal capsule, pons) and occur more frequently in patients with hypertension and diabetes. Classically, small-vessel strokes are caused by lipohyalinosis, which is a thickening of the vessel wall resulting in a diminished luminal area, but they also can be caused by atherothrombotic embolism.

Symptoms of ischemic stroke may worsen during the first hours or days through various mechanisms. For example, decreases in systemic blood pressure may decrease cerebral blood flow to marginally perfused, ischemic brain. In the setting of atherothrombotic disease, a partially occluded artery may progress to complete occlusion. Recurrent embolism may occur from a proximal arterial or cardiac source. Cerebral edema may develop during the first few days after an ischemic stroke, and the resulting mass effect can lead to

TABLE 407-2 CLINICAL MANIFESTATIONS OF ISCHEMIC CEREBROVASCULAR DISEASE

OCCLUDED ARTERY	TYPICAL MAJOR CLINICAL MANIFESTATIONS*
Internal carotid artery	Ipsilateral visual loss Ipsilateral middle cerebral artery syndrome
Anterior choroidal artery	Contralateral hemiparesis Contralateral sensory impairment Contralateral visual field defect
Anterior cerebral artery	Contralateral leg > arm paresis Contralateral leg > arm sensory deficit
Middle cerebral artery	Contralateral hemiparesis affecting face and arm > leg Contralateral sensory deficit affecting face and arm > leg Contralateral visual field defect Aphasia (dominant hemisphere) Contralateral hemispatial neglect (nondominant or dominant hemisphere)
Posterior cerebral artery	Contralateral homonymous hemianopia (or homonymous superior or inferior quadrantanopia) Contralateral sensory deficits (thalamic involvement)
Basilar artery tip	Bilateral central visual loss Confusion
Basilar artery	Ipsilateral cranial nerve deficit Contralateral hemiparesis Contralateral sensory impairment affecting arm and/or leg Coordination deficit
Vertebral artery, posterior inferior cerebellar artery	Ipsilateral sensory impairment over the face Dysphagia Ipsilateral Horner syndrome Ataxia
Superior cerebellar artery	Gait ataxia Ipsilateral limb ataxia Variable contralateral limb weakness

*Note: not all may be present.

clinical deterioration (Chapter 406). Secondary bleeding can occur in an area that was primarily the site of an ischemic injury when reperfusion, either through collateral vessels or as the result of a therapeutic intervention, restores blood flow into vessels in which the endothelium was damaged by the original ischemic insult.

CLINICAL MANIFESTATIONS

Neurologic deficits that occur in the setting of ischemic stroke depend on the involved vascular territory (Table 407-2) and underlying cause. Embolic stroke is generally characterized by the presence of a maximal deficit at onset, whereas the onset may be more gradual or stuttering in the setting of an atherothrombotic stroke. The distinction, however, is not of great use for diagnosis in individual patients. Transient symptoms in the same distribution can be caused by TIA if there is no permanent tissue injury.

Internal Carotid Artery

The bifurcation of the common carotid artery into the internal and external carotid arteries in the neck is a common site of atherosclerotic disease (see Fig. 406-1). With occlusion of the internal carotid artery, patients who have an incomplete circle of Willis can suffer profound contralateral loss of motor and sensory function affecting the face, arm, and leg. In patients with an intact anterior communicating artery that can supply the ipsilateral anterior cerebral artery (see E-Fig. 406-1), the leg may be relatively spared, and an internal carotid artery occlusion may be clinically indistinguishable from a middle cerebral artery occlusion. If the anterior communicating artery is absent on the side opposite an internal carotid artery occlusion, the ipsilateral leg may also be affected, and the presentation may be confused with a cardioembolic cause because both hemispheres are involved. Occlusion of the ipsilateral ophthalmic artery can lead to blindness in that eye. Transient symptoms of retinal ischemia, classically described by patients as a "shade coming down over my vision," indicate amaurosis fugax. Other common symptoms include a darkening or blurring of vision in the affected eye. Transient hypoperfusion ipsilateral to a high-grade internal carotid artery stenosis can cause limb-shaking TIAs that can be confused with seizures. Systemic hypotension in the setting of a high-grade carotid stenosis can lead to ischemic injury in watershed zones between the major intracranial arteries and in the border zone between the distal territories of cortical and lenticulostriate penetrating vessels.

Anterior Choroidal Artery

The anterior choroidal artery generally arises from the supraclinoid portion of the internal carotid artery (see Fig. 406-7). Causes of occlusion of the anterior choroidal artery are similar to those of occlusion of the small intracranial arteries. Symptoms can include contralateral motor and sensory deficits and contralateral visual field deficits, the latter of which can occur in isolation.

Cerebral Arteries

About 2% of strokes are related to isolated occlusion of the anterior cerebral artery (see Figs. 406-6 and 406-7). Occlusion of the A1 segment in patients in whom the contralateral A1 segment is hypoplastic or absent can lead to bilateral leg involvement, abulia, and urinary incontinence because of infarction of both frontal lobes.

The middle cerebral artery is the most common artery involved in occlusions related to cardiogenic embolism. It supplies the lateral portions of the frontal, parietal, and temporal lobes as well as the basal ganglia and the anterior limb and genu of the internal capsule. Middle cerebral artery occlusions are characterized by involvement of the contralateral face and arm to a greater extent than of the leg (see Figs. 406-6, 406-8, and 406-9), often accompanied by a contralateral hemispatial neglect. When the dominant hemisphere is involved, the patient may have an aphasia. With frontal lobe involvement, patients often have an ipsilateral, conjugate deviation of the eyes, which can be forced past the midline with vigorous encouragement, oculocephalic maneuvers, or caloric stimulation.

Branch middle cerebral artery occlusions can result in partial syndromes. For example, a branch middle cerebral artery occlusion with intact collaterals can cause a global aphasia without an accompanying motor deficit. Anterior branch, dominant hemisphere middle cerebral artery occlusions can cause an expressive, cortical-type motor (Broca) aphasia with sparing of comprehension. Occlusion of the angular branch of the middle cerebral artery can cause receptive, cortical-type (Wernicke) aphasia. Borderzone infarcts can result in transcortical aphasias, characterized by relatively preserved repetitions.

Both posterior cerebral arteries arise from the basilar artery in about 75% of people (see E-Fig. 406-1). In the other 25%, one or both P1 segments are hypoplastic or absent, with the posterior cerebral arteries arising from the ipsilateral internal carotid artery (so-called fetal circulation). Without vascular imaging, it is not possible to determine if a posterior cerebral artery distribution infarct (see Figs. 406-6 to 406-8) is related to carotid or vertebrobasilar circulation disease. The posterior cerebral artery and posterior communicating arteries supply the thalamus. Thalamic infarctions can result in contralateral hemianesthesia and ataxia. Contralateral hemiballismus can result if the subthalamic nucleus is damaged. Infarction of the ipsilateral occipital lobe causes a contralateral homonymous hemianopia that can be partial, depending on the extent of injury. The visual field deficit tends to become more congruous in the two eyes as the area of injury becomes more posterior (i.e., the closer to the occipital pole).

Vertebral and Basilar Arteries

Occlusion of the basilar artery (see Figs. 406-3 and 406-4B) can lead to "locked-in syndrome" (Chapter 404) in which the patient is awake and alert, because the periaqueductal gray can receive a separate blood supply, but unable to move or to communicate except for vertical eye movements, because of sparing of the collicular nuclei in the midbrain. The tip of the basilar artery is a common location for embolic occlusion. Symptoms can include visual field defects due to unilateral or bilateral occipital injury and confusional states due to thalamic involvement.

Occlusions of penetrating and circumferential branches of the basilar artery and vertebral artery can produce a variety of symptoms (see Table 407-2), depending on the portion of the artery involved, several of which constitute eponymic midbrain (E-Fig. 407-1), pontine (E-Fig. 407-2), or medullary (E-Fig. 407-3) syndromes. Occlusion of the superior cerebellar artery can cause truncal ataxia because of infarction of the cerebellar vermis, with or without ataxia of the ipsilateral limbs, which can be caused by infarction of the ipsilateral cerebellar hemisphere.

Small Vessels

Occlusion of a small penetrating intracranial vessel can result in one of the classic lacunar syndromes (Table 407-3). These syndromes are not otherwise

Claude Syndrome

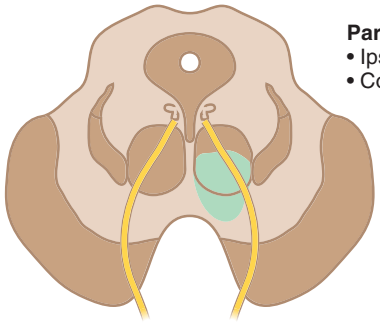

Paramedian infarction
- Ipsilateral 3rd nerve
- Contralateral ataxia

Benedikt Syndrome

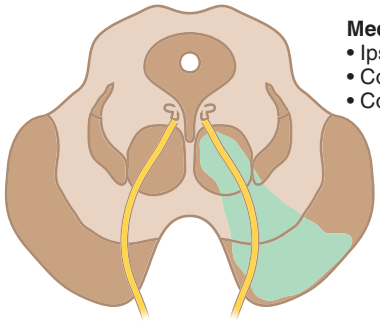

Medial infarction
- Ipsilateral 3rd nerve
- Contralateral ataxia
- Contralateral hemiparesis

Weber Syndrome

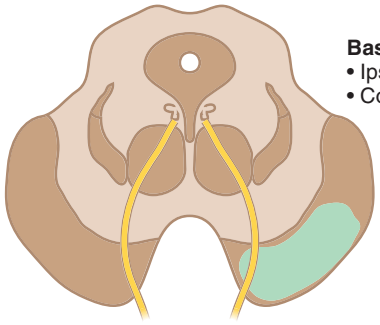

Basal infarction
- Ipsilateral 3rd nerve
- Contralateral hemiparesis

E-FIGURE 407-1. **Midbrain stroke syndromes.**

Raymond Syndrome

Mediolateral infarction
- Ipsilateral internuclear ophthalmoplegia
- Contralateral hemiparesis
- Contralateral UMN 7th nerve
- Contralateral light touch, proprioception
- +/– contralateral pain/temperature (with ipsilateral Horner)

Foville Syndrome

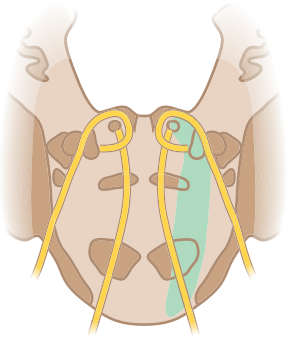

Mediolateral infarction
- Ipsilateral LMN 7th nerve
- Ipsilateral conjugate gaze palsy
- Contralateral hemiparesis

Millard-Gubler Syndrome

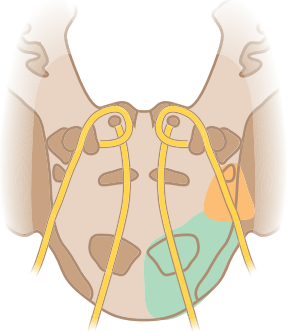

Basal infarction
- Ipsilateral LMN 7th nerve
- Ipsilateral 6th nerve
- Contralateral hemiparesis
- +/– contralateral pain/temperature

Locked-in Syndrome

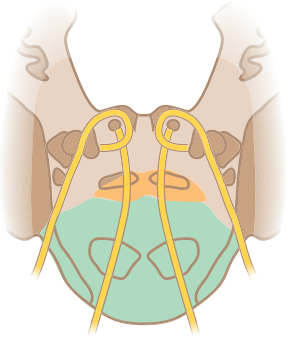

Bilateral basis pontis infarction
- Bilateral LMN 7th nerve
- Bilateral 6th nerve
- Quadriplegia
- +/– bilateral light touch/proprioception

E-FIGURE 407-2. **Pontomedullary and pontine stroke syndromes.** LMN = lower motor neuron; UMN = upper motor neuron.

Dejerine Syndrome

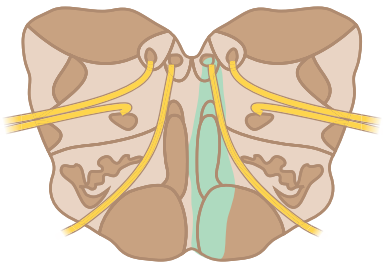

Medial infarction
- Ipsilateral XII nerve
- Contralateral hemiparesis
- Contralateral light touch, proprioception

Jackson Syndrome

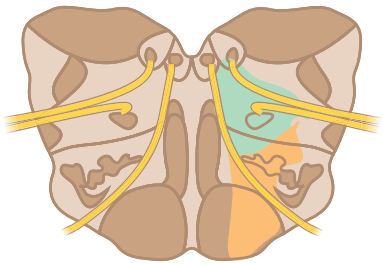

Lateral median infarction
- Ipsilateral IX, X nerve
- Ipsilateral XII nerve
- Variable contralateral hemiparesis

Avellis Syndrome

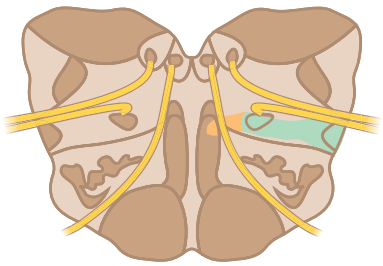

Lateral infarction
- Ipsilateral IX, X nerve
- Contralateral pain, temperature (ipsilateral Horner)
- +/– contralateral hemiparesis

Wallenberg Syndrome

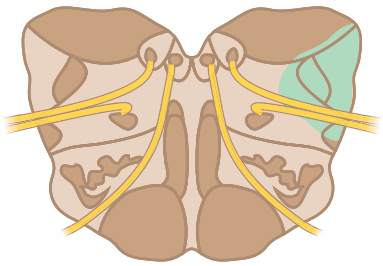

Dorsolateral infarction
- Ipsilateral facial anesthesia
- Contralateral pain, temperature
- Ipsilateral Horner
- Ipsilateral IX, X
- Ipsilateral ataxia (olivocerebellar tract)

E-FIGURE 407-3. Medullary stroke syndromes.

localizing and can occur with occlusions of small penetrating vessels in either the anterior or vertebrobasilar circulations. Lacunar syndromes are not pathognomonic of small-vessel intracranial disease and can be caused by a variety of other conditions, including emboli from a more proximal arterial or cardioembolic source or brain hemorrhage (Chapter 408).

DIAGNOSIS

The diagnosis of ischemic stroke depends on acquiring an accurate history, eliciting key findings on general and neurologic examinations, and obtaining supporting data from selected laboratory studies (Fig. 407-1). An initial anatomic and pathophysiologic differential diagnosis is usually established on the basis of the patient's history. Findings on physical and neurologic examinations can support or refute initial conclusions based on the history and can further refine the differential diagnosis.

History

The abrupt onset of a focal neurologic deficit in the distribution of a specific vascular territory is the hallmark of acute ischemic stroke. The differential and most likely diagnosis can often be determined on the basis of history alone. For example, a patient with a history of atrial fibrillation who abruptly develops word-finding difficulties associated with a right hemiparesis and sensory impairment most likely had a cardiogenic embolus to the left middle cerebral artery. A patient with the acute onset of diplopia, vertigo, and a hemiparesis most likely has a lesion in the brain stem.

Goals of the immediate history include determining the exact time when symptoms began or the last time the patient was known to be well, concomitant medical illnesses, risk factors, medications, allergies, and other potential causes for symptoms that might mimic acute ischemic stroke. Because a stroke may affect a patient's ability to communicate, the history may require input from a witness. Additional details of the patient's past medical, family, and social history may need to be deferred in the emergent setting, but these issues can be explored if the information is important for acute treatment decisions.

Physical Examination

Severely elevated blood pressures in the setting of neurologic deficits referable to the basal ganglia, thalamus, pons, or cerebellum increase the likelihood of a brain hemorrhage (Chapter 408). In a patient with transient vertigo associated with left arm movement, a reduced blood pressure in that arm suggests subclavian steal syndrome. Detection of an anterior cervical bruit contralateral to symptoms and signs indicative of a middle cerebral artery distribution infarct increase the likelihood of symptomatic carotid stenosis. An irregularly irregular heart rhythm with or without a cardiac murmur may indicate atrial fibrillation and a cardioembolic etiology. Finding a cholesterol embolus on funduscopic examination can be consistent with a proximal source of atheroembolism. Funduscopy can also show evidence of a small-vessel disease related to diabetes or hypertension (see Figs. 423-24 and 423-26).

A general neurologic examination (Chapter 396) including evaluations of cognition, language, spatial neglect, cranial nerves, motor function, sensation, coordination, gait, and reflexes is important both for documenting stroke-related deficits and for providing information critical for determining the area of the brain affected by the stroke and the severity of the injury. The use of a standardized graded neurologic impairment assessment provides a tool for measuring the severity of the stroke, determining the risks and benefits of treatment interventions, assessing prognosis, and observing patients objectively over time. The National Institutes of Health Stroke Scale (Table 407-4), which is the most commonly used approach, is both reliable and well validated. The individual items are summed to provide a total score.

TABLE 407-3 LACUNAR SYNDROMES

Pure motor stroke
Pure sensory stroke
Ataxic hemiparesis
Clumsy hand–dysarthria

Initial Laboratory Tests

Laboratory testing can help exclude conditions that may mimic, complicate, or lead to an acute ischemic stroke (Table 407-5).[5] Tests that should be obtained in all patients with suspected ischemic stroke include a complete blood count and platelet count, prothrombin time/international normalized ratio (INR), activated partial thromboplastin time, blood glucose level, serum electrolytes, tests of renal function, troponin level, and oxygen saturation. An electrocardiogram should be obtained urgently, and the patient should be sent for a CT brain scan or MRI as soon as he or she is stable enough. Additional tests are indicated in selected patients. For example, women of childbearing age should have a pregnancy test. A toxicology screen and blood alcohol levels should be obtained if drug or alcohol abuse is suspected. In patients who may be receiving a direct thrombin inhibitor or a factor Xa inhibitor, a thrombin time or ecarin clot time may be helpful in determining whether the patient is anticoagulated. An elevated erythrocyte sedimentation rate may point to an inflammatory cause or systemic infection.

The complete blood count can provide information about both the potential cause of the stroke and possible therapeutic interventions. An elevated white blood cell count may indicate an infectious cause of stroke, such as infective endocarditis (Chapter 76). Systemic infection may also cause a recrudescence of prior stroke symptoms in a patient who had previously

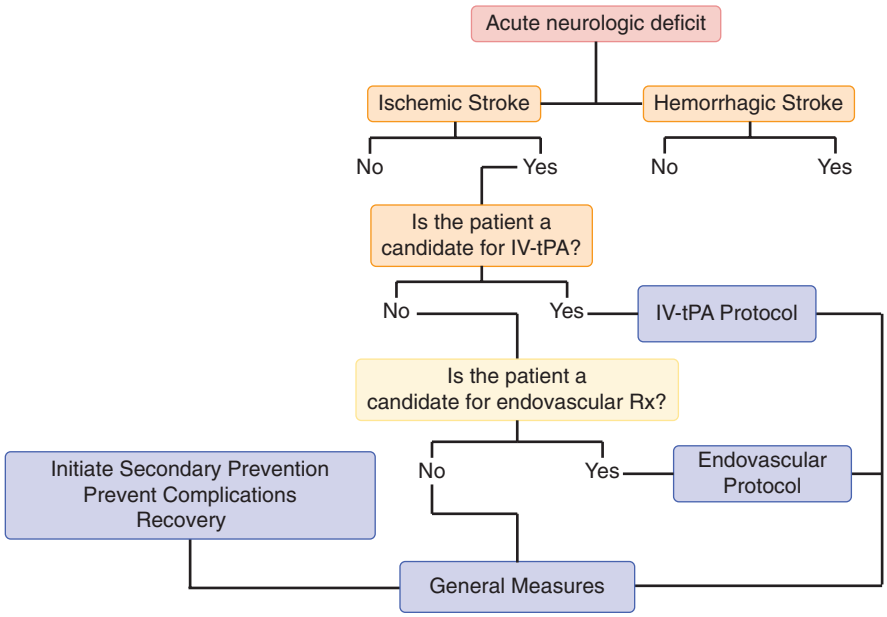

FIGURE 407-1. Approach to ischemic stroke. IV-tPA = intravenous tissue plasminogen activator; Rx = therapy. (From Goldstein LB. Modern medical management of acute ischemic stroke. *Methodist DeBakey Cardiovasc J.* 2014;10:99-104.)

TABLE 407-4 NATIONAL INSTITUTES OF HEALTH STROKE SCALE

Administer stroke scale items in the order listed. Record performance in each category after each subscale exam. Do not go back and change scores. Follow directions provided for each exam technique. Scores should reflect what the patient does, not what the clinician thinks the patient can do. The clinician should record answers while administering the exam and work quickly. Except where indicated, the patient should not be coached (i.e., repeated requests to patient to make a special effort).

INSTRUCTIONS	SCALE DEFINITION	SCORE
1a. **Level of Consciousness:** The investigator must choose a response if a full evaluation is prevented by such obstacles as an endotracheal tube, language barrier, orotracheal trauma/bandages. A 3 is scored only if the patient makes no movement (other than reflexive posturing) in response to noxious stimulation.	0 = **Alert**; keenly responsive. 1 = **Not alert**; but arousable by minor stimulation to obey, answer, or respond. 2 = **Not alert**; requires repeated stimulation to attend, or is obtunded and requires strong or painful stimulation to make movements (not stereotyped). 3 = Responds only with reflex motor or autonomic effects or totally unresponsive, flaccid, and areflexic.	_____
1b. **LOC Questions**: The patient is asked the month and his/her age. The answer must be correct—there is no partial credit for being close. Aphasic and stuporous patients who do not comprehend the questions will score 2. Patients unable to speak because of endotracheal intubation, orotracheal trauma, severe dysarthria from any cause, language barrier, or any other problem not secondary to aphasia are given a 1. It is important that only the initial answer be graded and that the examiner not "help" the patient with verbal or nonverbal cues.	0 = **Answers** both questions correctly. 1 = **Answers** one question correctly. 2 = **Answers** neither question correctly.	_____
1c. **LOC Commands:** The patient is asked to open and close the eyes and then to grip and release the nonparetic hand. Substitute another one-step command if the hands cannot be used. Credit is given if an unequivocal attempt is made but not completed due to weakness. If the patient does not respond to command, the task should be demonstrated to him or her (pantomime), and the result scored (i.e., follows none, one, or two commands). Patients with trauma, amputation, or other physical impediments should be given suitable one-step commands. Only the first attempt is scored.	0 = **Performs** both tasks correctly. 1 = **Performs** one task correctly. 2 = **Performs** neither task correctly.	_____
2. **Best Gaze:** Only horizontal eye movements will be tested. Voluntary or reflexive (oculocephalic) eye movements will be scored, but caloric testing is not done. If the patient has a conjugate deviation of the eyes that can be overcome by voluntary or reflexive activity, the score will be 1. If a patient has an isolated peripheral nerve paresis (CN III, IV, or VI), score a 1. Gaze is testable in all aphasic patients. Patients with ocular trauma, bandages, preexisting blindness, or other disorder of visual acuity or fields should be tested with reflexive movements, and a choice made by the investigator. Establishing eye contact and then moving about the patient from side to side will occasionally clarify the presence of a partial gaze palsy.	0 = **Normal**. 1 = **Partial gaze palsy**; gaze is abnormal in one or both eyes, but forced deviation or total gaze paresis is not present. 2 = **Forced deviation**, or total gaze paresis not overcome by the oculocephalic maneuver.	_____
3. **Visual:** Visual fields (upper and lower quadrants) are tested by confrontation, using finger counting or visual threat, as appropriate. Patients may be encouraged, but if they look at the side of the moving fingers appropriately, this can be scored as normal. If there is unilateral blindness or enucleation, visual fields in the remaining eye are scored. Score 1 only if a clear-cut asymmetry, including quadrantanopia, is found. If patient is blind from any cause, score 3. Double simultaneous stimulation is performed at this point. If there is extinction, patient receives a 1, and the results are used to respond to item 11.	0 = **No visual loss**. 1 = **Partial hemianopia**. 2 = **Complete hemianopia**. 3 = **Bilateral hemianopia** (blind including cortical blindness).	_____
4. **Facial Palsy:** Ask—or use pantomime to encourage—the patient to show teeth or raise eyebrows and close eyes. Score symmetry of grimace in response to noxious stimuli in the poorly responsive or noncomprehending patient. If facial trauma/bandages, orotracheal tube, tape, or other physical barriers obscure the face, these should be removed to the extent possible.	0 = **Normal** symmetrical movements. 1 = **Minor paralysis** (flattened nasolabial fold, asymmetry on smiling). 2 = **Partial paralysis** (total or near-total paralysis of lower face). 3 = **Complete paralysis** of one or both sides (absence of facial movement in the upper and lower face).	_____
5. **Motor Arm**: The limb is placed in the appropriate position: extend the arms (palms down) 90 degrees (if sitting) or 45 degrees (if supine). Drift is scored if the arm falls before 10 seconds. The aphasic patient is encouraged using urgency in the voice and pantomime, but not noxious stimulation. Each limb is tested in turn, beginning with the nonparetic arm. Only in the case of amputation or joint fusion at the shoulder, the examiner should record the score as untestable (UN) and clearly write the explanation for this choice.	0 = **No drift**; limb holds 90 (or 45) degrees for full 10 seconds. 1 = **Drift**; limb holds 90 (or 45) degrees, but drifts down before full 10 seconds; does not hit bed or other support. 2 = **Some effort against gravity**; limb cannot get to or maintain (if cued) 90 (or 45) degrees, drifts down to bed, but has some effort against gravity. 3 = **No effort against gravity**; limb falls. 4 = **No movement**. UN = **Amputation** or joint fusion, explain: _____ 5a. **Left Arm** 5b. **Right Arm**	_____
6. **Motor Leg:** The limb is placed in the appropriate position: hold the leg at 30 degrees (always tested supine). Drift is scored if the leg falls before 5 seconds. The aphasic patient is encouraged using urgency in the voice and pantomime, but not noxious stimulation. Each limb is tested in turn, beginning with the nonparetic leg. Only in the case of amputation or joint fusion at the hip, the examiner should record the score as untestable (UN) and clearly write the explanation for this choice.	0 = **No drift**; leg holds 30-degree position for full 5 seconds. 1 = **Drift**; leg falls by the end of the 5-second period but does not hit bed. 2 = **Some effort against gravity**; leg falls to bed by 5 seconds, but has some effort against gravity. 3 = **No effort against gravity**; leg falls to bed immediately. 4 = **No movement**. UN = **Amputation** or joint fusion, explain: _____ 6a. **Left Leg** 6b. **Right Leg**	_____

TABLE 407-4 NATIONAL INSTITUTES OF HEALTH STROKE SCALE—cont'd

INSTRUCTIONS	SCALE DEFINITION	SCORE
7. **Limb Ataxia**: This item is aimed at finding evidence of a unilateral cerebellar lesion. Test with eyes open. In case of visual defect, ensure testing is done in intact visual field. The finger-nose-finger and heel-shin tests are performed on both sides, and ataxia is scored only if present out of proportion to weakness. Ataxia is absent in the patient who cannot understand or is paralyzed. Only in the case of amputation or joint fusion, the examiner should record the score as untestable (UN) and clearly write the explanation for this choice. In case of blindness, test by having the patient touch nose from extended arm position.	0 = **Absent**. 1 = **Present in one limb**. 2 = **Present in two limbs**. UN = **Amputation** or joint fusion, explain: _____	_____
8. **Sensory**: Sensation or grimace to pinprick when tested, or withdrawal from noxious stimulus in the obtunded or aphasic patient. Only sensory loss attributed to stroke is scored as abnormal, and the examiner should test as many body areas (arms [not hands], legs, trunk, face) as needed to accurately check for hemisensory loss. A score of 2, "severe or total sensory loss," should only be given when a severe or total loss of sensation can be clearly demonstrated. Stuporous and aphasic patients will, therefore, probably score 1 or 0. The patient with brain stem stroke who has bilateral loss of sensation is scored 2. If the patient does not respond and is quadriplegic, score 2. Patients in a coma (item 1a = 3) are automatically given a 2 on this item.	0 = **Normal**; no sensory loss. 1 = **Mild-to-moderate sensory loss**; patient feels pinprick is less sharp or is dull on the affected side; or there is a loss of superficial pain with pinprick, but patient is aware of being touched. 2 = **Severe to total sensory loss**; patient is not aware of being touched in the face, arm, and leg.	_____
9. **Best Language**: A great deal of information about comprehension will be obtained during the preceding sections of the examination. For this scale item, the patient is asked to describe what is happening in the attached picture, to name the items on the attached naming sheet, and to read from the attached list of sentences. Comprehension is judged from responses here, as well as to all of the commands in the preceding general neurological exam. If visual loss interferes with the tests, ask the patient to identify objects placed in the hand, repeat, and produce speech. The intubated patient should be asked to write. The patient in a coma (item 1a = 3) will automatically score 3 on this item. The examiner must choose a score for the patient with stupor or limited cooperation, but a score of 3 should be used only if the patient is mute and follows no one-step commands.	0 = **No aphasia**; normal. 1 = **Mild-to-moderate aphasia**; some obvious loss of fluency or facility of comprehension, without significant limitation on ideas expressed or form of expression. Reduction of speech and/or comprehension, however, makes conversation about provided materials difficult or impossible. For example, in conversation about provided materials, examiner can identify picture or naming card content from patient's response. 2 = **Severe aphasia**; all communication is through fragmentary expression; great need for inference, questioning, and guessing by the listener. Range of information that can be exchanged is limited; listener carries burden of communication. Examiner cannot identify materials provided from patient response. 3 = **Mute, global aphasia**; no usable speech or auditory comprehension.	_____
10. **Dysarthria**: If patient is thought to be normal, an adequate sample of speech must be obtained by asking patient to read or repeat words from the attached list. If the patient has severe aphasia, the clarity of articulation of spontaneous speech can be rated. Only if the patient is intubated or has other physical barriers to producing speech, the examiner should record the score as untestable (UN) and clearly write an explanation for this choice. Do not tell the patient why he or she is being tested.	0 = **Normal**. 1 = **Mild-to-moderate dysarthria**; patient slurs at least some words and, at worst, can be understood with some difficulty. 2 = **Severe dysarthria**; patient's speech is so slurred as to be unintelligible in the absence of or out of proportion to any dysphasia, or is mute/anarthric. UN = **Intubated** or other physical barrier, explain:_____	_____
11. **Extinction and Inattention (formerly Neglect):** Sufficient information to identify neglect may be obtained during the prior testing. If the patient has a severe visual loss preventing visual double simultaneous stimulation, and the cutaneous stimuli are normal, the score is normal. If the patient has aphasia but does appear to attend to both sides, the score is normal. The presence of visual spatial neglect or anosognosia may also be taken as evidence of abnormality. Since the abnormality is scored only if present, the item is never untestable.	0 = **No abnormality**. 1 = **Visual, tactile, auditory, spatial, or personal inattention** or extinction to bilateral simultaneous stimulation in one of the sensory modalities. 2 = **Profound hemi-inattention or extinction to more than one modality**; does not recognize own hand or orients to only one side of space.	_____

From http://www.ninds.nih.gov/doctors/nih_stroke_scale.pdf. Accessed February 26, 2015.

recovered or in whom a stroke had not been previously recognized. Polycythemia (Chapter 166) can cause hyperviscosity that leads to occlusion of small intracranial vessels. Thrombocytopenia, either primary or secondary, can lead to platelet thrombi. The prothrombin time/INR and activated thromboplastin time provide indices that may reveal an underlying coagulation disorder, and thrombocytopenia and coagulation disorders may preclude treatment with intravenous recombinant tissue plasminogen activator (rtPA).

Both hypoglycemia (Chapter 230) and hyperglycemia (Chapter 229) may cause strokelike symptoms. Impaired renal function (Chapter 130) is a risk factor for ischemic stroke and may increase the risks of using thrombolytic and anticoagulant medications. Abnormalities of other serum electrolytes (e.g., hyponatremia; Chapter 116) can also cause neurologic symptoms.

The electrocardiogram may reveal changes suggestive of acute myocardial ischemia as well as atrial fibrillation, the most common cause of embolic stroke. Stroke may also cause a variety of cardiac arrhythmias. Acute MI, especially anteroseptal MI, is associated with a higher risk of cardiogenic embolism, and an acute stroke may also precipitate an MI. A troponin level is usually adequate for this purpose, especially because it remains elevated for several days after the MI when embolism from a mural thrombus is most likely to occur. Patients with acute stroke should be placed on telemetry monitoring. Urgent echocardiography is used selectively.

Brain Imaging

CT or MRI brain imaging is an essential part of the evaluation of all patients with suspected ischemic stroke. Imaging can locate the area of damage, distinguish a brain hemorrhage from an ischemic stroke, and identify mass lesions such as tumor (Chapter 189), abscess (Chapter 413), or subdural hematoma that can present acutely and mimic a stroke. Brain CT is widely and rapidly available and provides the information necessary for the treatment of most patients with acute stroke. Brain MRI can detect areas of acute ischemic injury not apparent on CT brain imaging (Fig. 407-2), but it cannot be performed in patients with metal implants and devices such as cardiac pacemakers and is a challenge to perform in unstable patients.

The changes on brain CT, such as loss of gray-white distinction, loss of the insular ribbon, and blurring of the borders of the basal ganglia, can be subtle. The area of ischemic injury on brain CT scan appears as a relative hypodensity (Fig. 407-3), in contrast to brain hemorrhage, which appears hyperdense

compared with the surrounding brain parenchyma (see Fig. 408-2). CT can also show acute hemorrhage in the subarachnoid space, which can be indicative of aneurysmal rupture (see Fig. 408-1). The dense middle cerebral artery sign or the dot sign, in which an artery in the sylvian fissure may appear dense, can indicate thrombus in these vessels.

The findings on CT are often normal in the acute phase of ischemic stroke, and MRI is more sensitive for detecting acute ischemic injury (Fig. 407-4). Because brain CT imaging of posterior fossa structures is often obscured by beam hardening artifact from the petrous bones, MRI is also more sensitive for visualizing the brain stem and cerebellum. MRI signal patterns also can distinguish acute from subacute and remote ischemic injury, distinguish acute and remote brain hemorrhage, and identify other nonvascular conditions. However, MRI is not required before treatment with intravenous rtPA because CT can reliably exclude parenchymal brain hemorrhage and can detect other common conditions that may mimic a stroke, such as a mass lesion.

Lumbar Puncture

Lumbar puncture is rarely necessary in the evaluation of patients with acute stroke. In occasional patients, meningitis, especially septic meningitis from

TABLE 407-5 IMMEDIATE DIAGNOSTIC STUDIES: EVALUATION OF A PATIENT WITH SUSPECTED ACUTE ISCHEMIC STROKE

ALL PATIENTS
Noncontrast brain CT or brain MRI
Blood glucose
Oxygen saturation
Serum electrolytes/renal function tests*
Complete blood count, including platelet count*
Markers of cardiac ischemia*
Prothrombin time/INR*
Partial thromboplastin time*
ECG*
SELECTED PATIENTS
Thrombin time and/or ecarin clotting time if it is suspected the patient is taking direct thrombin inhibitors or direct factor Xa inhibitors
Hepatic function tests
Toxicology screen
Blood alcohol level
Pregnancy test
Arterial blood gas tests (if hypoxia is suspected)
Chest radiography (if lung disease is suspected)
Lumbar puncture (if meningitis is suspected or subarachnoid hemorrhage is suspected but the CT scan is negative for blood)
Electroencephalogram (if seizures are suspected)

*Although it is desirable to know the results of these tests before giving intravenous recombinant tissue-type plasminogen activator, fibrinolytic therapy should not be delayed while awaiting the results unless (1) there is clinical suspicion of a bleeding abnormality or thrombocytopenia, (2) the patient has received heparin or warfarin, or (3) the patient has received other anticoagulants (direct thrombin inhibitors or direct factor Xa inhibitors).
CT = computed tomography; ECG = electrocardiogram; INR = international normalized ratio; MRI = magnetic resonance imaging.
From Jauch EC, Saver JL, Adams HP Jr, et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* 2013;44:870-947.

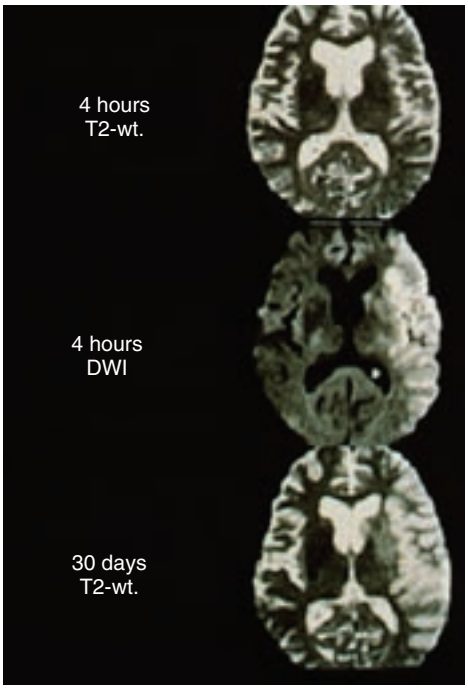

FIGURE 407-2. Magnetic resonance imaging (MRI) showing possible advantages of diffusion-weighted imaging (DWI) relative to conventional MRI at early times after vascular occlusion. ***Top,*** **Conventional T2-weighted MRI 4 hours after symptom onset that appears normal.** ***Middle,*** **At the same time, a DWI scan shows abnormalities in the left hemisphere.** ***Bottom,*** **Repeated T2-weighted MRI 1 month later showed an infarction in the same location as the initial DWI scan. (Courtesy Gregory W. Albers, Stanford University, Stanford, Calif.)**

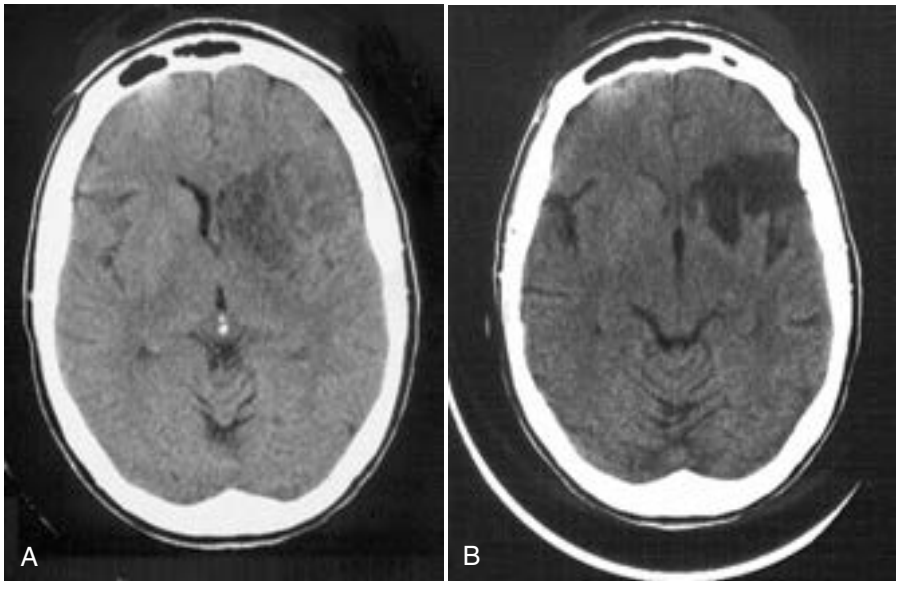

FIGURE 407-3. Computed tomographic imaging. A, A computed tomography (CT) scan of a patient with a left hemisphere infarction 6 to 24 hours after the onset of symptoms shows a hypodense area in the basal ganglia region and compression of the frontal horn of the lateral ventricle. **B,** A CT scan shows the chronic infarction 1 year later; atrophy and loss of tissue volume are visible. (Courtesy Gregory W. Albers, Stanford University, Stanford, Calif.)

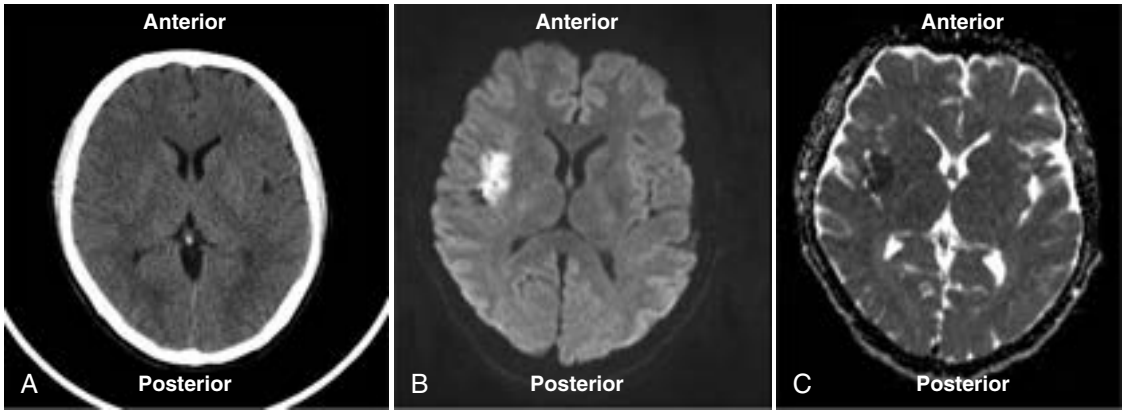

FIGURE 407-4. **A, Computed tomographic imaging. B, Magnetic resonance imaging, diffusion sequence. C, Magnetic resonance imaging, apparent diffusion coefficient (ADC) map. Computed tomography shows no evidence of ischemic injury. There is an obvious area of restricted diffusion in the right frontotemporal cortex that is dark on the ADC map, consistent with an area of acute ischemic injury.**

cardiogenic embolism in a patient with infective endocarditis, may cause stroke or strokelike symptoms and be an indicator for an urgent lumbar puncture. Brain CT usually demonstrates blood in the subarachnoid space in patients presenting with symptoms and signs of subarachnoid hemorrhage (Chapter 408), such as headache and meningismus. If, however, brain CT fails to visualize a subarachnoid hemorrhage in a patient in whom the clinical suspicion of subarachnoid hemorrhage is high, lumbar puncture should still be obtained.

Other Imaging

Carotid duplex ultrasonography, which combines B-mode vascular imaging with measures of blood flow velocity, is commonly used to screen for extracranial carotid artery stenosis but is rarely indicated in the acute setting. Both CT and MR angiography provide noninvasive vascular imaging of the extracranial and intracranial cerebral circulation, and either study can be obtained in conjunction with parenchymal imaging in the acute setting. CT or MR angiography may be helpful in identifying cervical artery dissections in a patient with headache, neck pain, and symptoms and signs consistent with ipsilateral ischemic injury. Venous sinuses should be imaged if a sinus thrombosis is being considered. In patients who are not candidates for intravenous rtPA but who may be considered for acute endovascular therapy, these studies can be used to identify a proximal arterial occlusion. Transcranial Doppler ultrasonography is an alternative for evaluating the proximal cerebral vessels, but it cannot be obtained in some patients because of inadequate sonographic windows.

Diagnostic catheter angiography carries about a 0.5 to 1.5% risk of causing a stroke and has largely been supplanted in the acute setting by noninvasive vascular imaging. Catheter angiography is, however, superior to CT or MR angiography for visualizing smaller intracranial vessels and for detecting intracranial vasculopathies such as vasculitis (Chapter 270).

TABLE 407-6 TIME GOALS FOR EVALUATION AND TREATMENT OF PATIENTS WITH ACUTE ISCHEMIC STROKE

TIME AFTER EMERGENCY DEPARTMENT ARRIVAL	GOALS
10 minutes	Assess ABCs, vital signs Provide oxygen if hypoxemic Obtain intravenous access Obtain laboratory studies CBC, coagulation, electrolytes Check glucose level, treat if indicated Perform screening neurologic assessment Activate stroke team Order "stroke code" brain CT or MRI Obtain 12-lead ECG
25 minutes	Review history Establish time at onset or last known normal Perform neurologic examination NIH Stroke Scale
45 minutes	Review laboratory studies Review brain CT or MRI results Evaluate inclusion and exclusion criteria (see Tables 407-8 and 407-9)
60 minutes	Review risks and benefits Obtain consent Begin infusion

ABCs = airway, breathing, circulation; CBC = complete blood count; CT = computed tomography; ECG = electrocardiogram; MRI = magnetic resonance imaging; NIH = National Institutes of Health.

Differential Diagnosis

The hallmark of acute ischemic stroke is the abrupt onset of a focal neurologic deficit, frequently attributable to an area of brain supplied by a specific artery or arteries. In some patients, however, the onset of ischemic stroke may be stuttering, and the stroke symptoms may have been heralded by a prior TIA. The detection of an ipsilateral cervical artery bruit may also support the diagnosis. Embolic stroke typically has its maximal severity at onset, but it can involve multiple vascular territories. The diagnosis of an embolic stroke may be further suggested by finding a cardiac murmur, an irregularly irregular heart rhythm, or signs of emboli in other vascular territories.

A variety of other neurologic conditions may also be manifested acutely. Migraine with aura (Chapter 398) can be associated with focal neurologic deficits, including speech impairment, visual changes, vertigo, weakness, numbness, and imbalance. Partial seizures (Chapter 403) may have negative symptoms, including aphasia and paresis, and a patient with a postictal Todd paralysis may appear to have had a stroke. As a further challenge in diagnosis, seizures may occur in patients who are having an acute stroke. In a patient without a previous diagnosis, the first episode of multiple sclerosis (Chapter 411) can mimic a stroke. Mass lesions such as neoplasms (Chapter 189) and abscesses (Chapter 413) are generally associated with a slowly progressive worsening of neurologic symptoms but may occasionally be manifested acutely. Metabolic disorders such as hypoglycemia (Chapter 230) or hyperglycemia (Chapter 229), toxin exposures (Chapters 22 and 110), and drug intoxications (Chapter 34) can cause focal symptoms similar to stroke. Strokelike symptoms may also be a manifestation of malingering, a conversion disorder, or other psychiatric illness.

TREATMENT

Intravenous rtPA

After initial respiratory and hemodynamic stabilization, the management of patients with acute ischemic stroke is directed at determining expeditiously (Table 407-6) whether treatment with intravenous rtPA is appropriate (Table 407-7; see also Fig. 407-1). Intravenous rtPA, administered within 4.5 hours of the onset of symptoms, does not reduce mortality but results in a higher odds of a better neurologic outcome at 3 months compared with placebo. The benefit of rtPA declines over time within this 4.5-hour treatment window (E-Fig. 407-4), with the odds ratio for a favorable 3-month outcome declining from 2.55 for treatment within 0 to 90 minutes, to 1.64 for 91 to 180 minutes, to 1.26 for 181 to 270 minutes, and to no statistical benefit for treatment

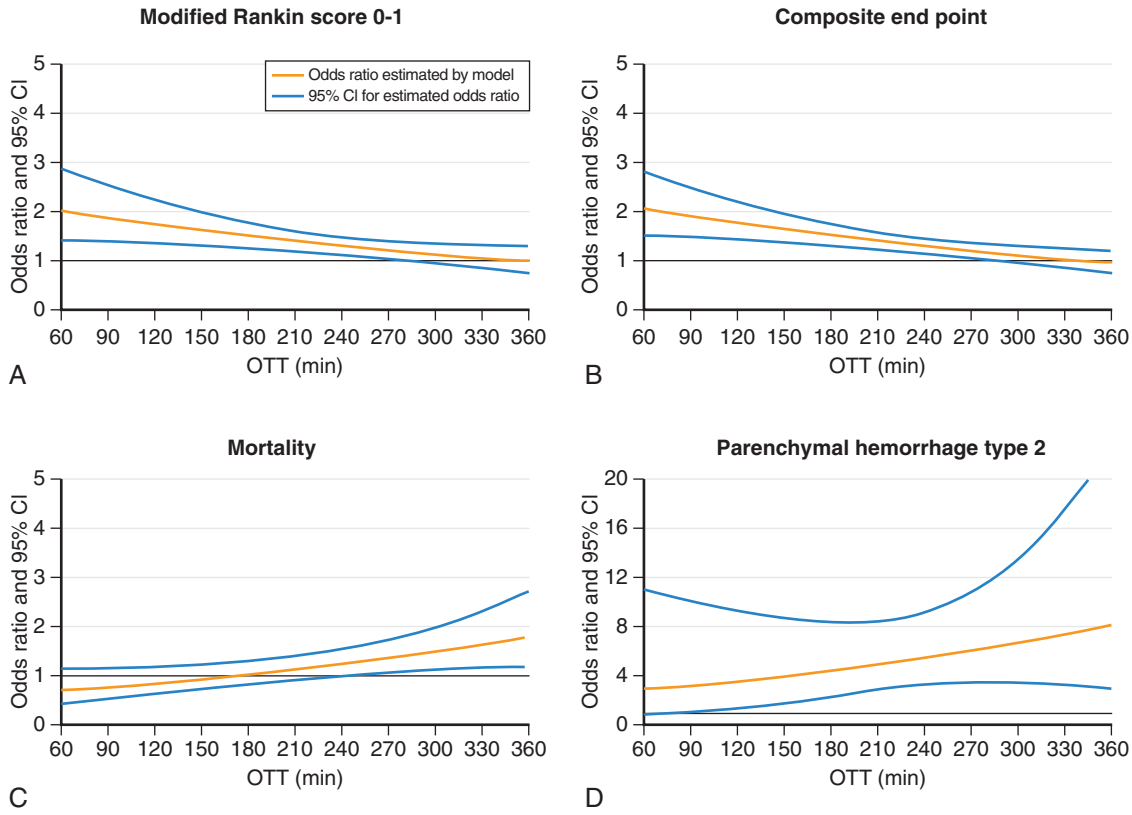

E-FIGURE 407-4. Relation of onset to treatment delay with treatment effect with intravenous rtPA. Relation of stroke onset to start of treatment (OTT) with treatment effect after adjustment for prognostic variables assessed by (**A**) day 90 modified Rankin score 0-1 versus 2-6 (interaction *P* = 0269, n = 3530 [excluding EPITHET7 data *P* = 0116, n = 3431]); (**B**) global test that incorporates modified Rankin score 0-1 versus 2-6, Barthel Index score 95-100 versus 90 or lower, and NIHSS score 0-1 versus 2 or more (interaction *P* = 0111, n = 3535 [excluding EPITHET7 data *P* = 0049, n = 3436]); (**C**) mortality (interaction *P* = 0444, n = 3530 [excluding EPITHET7 data *P* = 0582, n = 3431]); and (**D**) parenchymal hemorrhage type 2 (interaction *P* = 4140, n = 3531 [excluding EPITHET7 data *P* = 4578, n = 3431]). Thus, for parenchymal hemorrhage type 2, the fitted line is not statistically distinguishable from a horizontal line. For each graph, the adjusted odds ratio is shown with the 95% confidence interval (CI). CIs from the models will differ from those shown in the tables because the model uses data from all patients treated within 0 to 360 minutes, whereas the categorized analyses in the tables are based on subsets of patients; the modeled CIs are deemed to be more reliable. (From Lees KR, Bluhmki E, von Kummer R, et al. Time to treatment with intravenous alteplase and outcome in stroke: an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials. *Lancet*. 2010;375:1695-1703.)

TABLE 407-7 ADMINISTRATION OF rtPA FOR ACUTE ISCHEMIC STROKE

Infuse 0.9 mg/kg (maximum dose 90 mg) over 60 minutes, with 10% of the dose given as a bolus over 1 minute.
Admit the patient to an intensive care or stroke unit for monitoring.
If the patient develops severe headache, acute hypertension, nausea, or vomiting or has a worsening neurological examination, discontinue the infusion (if IV rtPA is being administered) and obtain emergent CT scan.
Measure blood pressure and perform neurological assessments every 15 minutes during and after IV rtPA infusion for 2 hours, then every 30 minutes for 6 hours, then hourly until 24 hours after IV rtPA treatment.
Increase the frequency of blood pressure measurements if systolic blood pressure is >180 mm Hg or if diastolic blood pressure is >105 mm Hg; administer antihypertensive medications to maintain blood pressure at or below these levels (Table 407-10).
Delay placement of nasogastric tubes, indwelling bladder catheters, or intra- arterial pressure catheters if the patient can be safely managed without them.
Obtain a follow-up CT or MRI scan at 24 hours after IV rtPA before starting anticoagulants or antiplatelet agents.

CT = computed tomography; IV = intravenous; MRI = magnetic resonance imaging; rtPA = recombinant tissue plasminogen activator.
From Jauch EC, Saver JL, Adams HP Jr, et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* 2013;44:870-947.

beyond 4.5 hours. Registry studies support a benefit in routine clinical practice similar to that in randomized trials.[6] As a result, current guidelines recommend that treatment with intravenous rtPA (Food and Drug Administration approved up to 3 hours after symptom onset) not be given if more than 4.5 hours have elapsed since the onset of symptoms. Treatment with rtPA is efficacious and safe among patients who are chronically treated with warfarin, provided their INR is 1.7 or lower, and is contraindicated with an INR higher than 1.7. Treatment increases the risk of intracranial hemorrhage, but the overall benefit includes these adverse events, which do not significantly increase in frequency during the 4.5-hour treatment window. Some patients have absolute exclusion criteria against treatment with intravenous rtPA (Table 407-8), with additional relative contraindications for treatment between 3 and 4.5 hours (Table 407-9). In patients without contraindications, treatment should begin as soon as possible in either treatment window.

Endovascular Therapy

Although strokes caused by large proximal occlusions tend to benefit less from treatment with intravenous rtPA compared with more distal or small-vessel obstructions, no randomized trials show additional benefit of infusing rtPA directly into the thrombus or of other endovascular treatments compared with intravenous rtPA alone, even though several devices are Food and Drug Administration approved to remove clots from brain blood vessels. As a result, current guidelines recommend treatment with intravenous rtPA even if intra-arterial treatments are available. Endovascular therapy can be considered in selected patients who cannot be treated with intravenous rtPA, such as patients who had a recent surgical procedure and who present up to 6 hours after a middle cerebral artery occlusion and perhaps longer after basilar artery occlusion.

Other Treatments

Regardless of whether the patient received intravenous rtPA or endovascular therapy, care in a comprehensive specialized stroke unit that incorporates rehabilitation is associated with better patient outcomes. Urgent anticoagulation to prevent recurrent stroke, to prevent worsening, or to improve functional outcome of patients with acute ischemic stroke is not recommended. Aspirin should not be started within 24 hours of treatment with intravenous rtPA, but aspirin should be started at 325 mg daily within 24 to 48 hours after the onset of stroke. Hemicraniectomy can increase survival in patients with extensive middle cerebral artery strokes, but most survivors will require assistance with their body needs.

Antihypertensive medications to reduce blood pressure acutely by 10 to 25% in the first 24 hours with a goal of blood pressure to below 140/90 mm Hg by 1 week does not improve outcomes compared with discontinuation of all antihypertensive medications. Current guidelines recommend that antihypertensive medications not be given unless the blood pressure rises to more than 220/120 mm Hg or higher in the absence of other indications. An exception is that blood pressure can be lowered in patients who are otherwise candidates for intravenous rtPA with a goal of maintaining blood pressures below 180/105 mm Hg after treatment (Fig. 407-5).

Several potential complications of acute stroke can often be avoided. Patients with stroke in any vascular distribution are at risk of aspiration pneumonia (Chapter 97). Stroke patients should not receive oral medications or nutrition until their ability to swallow safely has been assessed. Urinary tract infections (Chapter 284) are a potential complication; the routine placement of indwelling bladder catheters should be avoided, and patients who require an indwelling bladder catheter should have it removed as soon as feasible. Any infectious complications should be treated aggressively, and antipyretics should be used to maintain euthermia because fever is associated with more ischemic injury and poorer outcomes. Immobilized patients should receive deep venous thrombosis prophylaxis with subcutaneous unfractionated heparin or low-molecular-weight heparin (see Table 38-2) if it is not contraindicated, with mechanical intermittent pneumatic compression if anticoagulation is contraindicated, or with both.

TABLE 407-8 INCLUSION AND EXCLUSION CHARACTERISTICS OF PATIENTS WITH ISCHEMIC STROKE WHO COULD BE TREATED WITH IV rtPA WITHIN 3 HOURS FROM SYMPTOM ONSET

INCLUSION CRITERIA
Diagnosis of ischemic stroke causing measurable neurological deficit
Onset of symptoms <3 hours before beginning treatment
Aged ≥18 years
EXCLUSION CRITERIA
Significant head trauma or prior stroke in previous 3 months
Symptoms suggest subarachnoid hemorrhage
Arterial puncture at noncompressible site in previous 7 days
History of previous intracranial hemorrhage
Intracranial neoplasm, arteriovenous malformation, or aneurysm
Recent intracranial or intraspinal surgery
Elevated blood pressure (systolic >185 mm Hg or diastolic >110 mm Hg)
Active internal bleeding
Acute bleeding diathesis, including but not limited to
Platelet count <100,000/mm^3
Heparin received within 48 hours, resulting in aPTT greater than the upper limit of normal
Current use of anticoagulant with INR >1.7 or PT >15 seconds
Current use of direct thrombin inhibitors or direct factor Xa inhibitors with elevated sensitive laboratory tests (such as aPTT, INR, platelet count, and ECT; TT; or appropriate factor Xa activity assays)
Blood glucose concentration <50 mg/dL (2.7 mmol/L)
CT demonstrates multilobar infarction (hypodensity >⅓ cerebral hemisphere)
RELATIVE EXCLUSION CRITERIA
Recent experience suggests that under some circumstances—with careful consideration and weighting of risk to benefit—patients may receive fibrinolytic therapy despite 1 or more relative contraindications. Consider risk to benefit of IV rtPA administration carefully if any of these relative contraindications are present:
Only minor or rapidly improving stroke symptoms (clearing spontaneously)
Pregnancy
Seizure at onset with postictal residual neurological impairments
Major surgery or serious trauma within previous 14 days
Recent gastrointestinal or urinary tract hemorrhage (within previous 21 days)
Recent acute myocardial infarction (within previous 3 months)
NOTES:
• The checklist includes some FDA-approved indications and contraindications for administration of IV rtPA for acute ischemic stroke. Recent guideline revisions have modified the original FDA-approved indications. A physician with expertise in acute stroke care may modify this list.
• Onset time is defined as either the witnessed onset of symptoms or the time last known normal if symptom onset was not witnessed.
• In patients without recent use of oral anticoagulants or heparin, treatment with IV rtPA can be initiated before availability of coagulation test results but should be discontinued if INR is >1.7 or PT is abnormally elevated by local laboratory standards.
• In patients without history of thrombocytopenia, treatment with IV rtPA can be initiated before availability of platelet count but should be discontinued if platelet count is <100,000/mm^3.

aPTT = activated partial thromboplastin time; CT = computed tomography; ECT = ecarin clotting time; FDA = Food and Drug Administration; INR = international normalized ratio; IV = intravenous; PT = partial thromboplastin time; rtPA = recombinant tissue plasminogen activator; TT = thrombin time.
From Jauch EC, Saver JL, Adams HP Jr, et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* 2013;44:870-947.

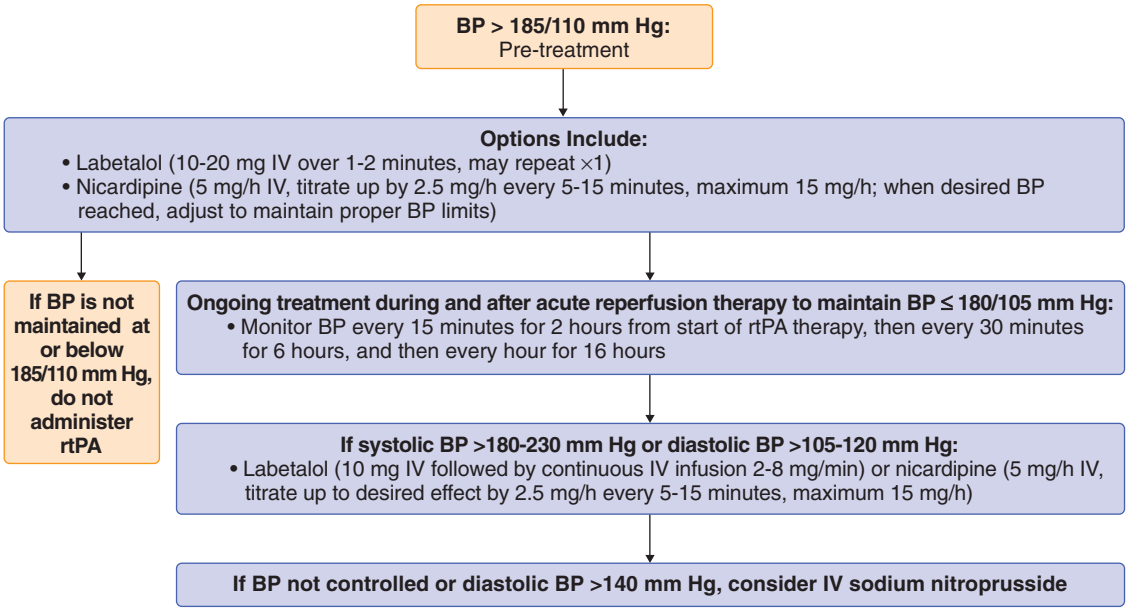

Abbreviations: BP, blood pressure; IV, intravenously; and rtPA, recombinant tissue plasminogen activator.

Adapted from Jauch EC, Saver JL, Adams HP, Jr., et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/ American Stroke Association. *Stroke.* 2013;44:870–947.

FIGURE 407-5. Potential Approaches to Arterial Hypertension in Acute Ischemic Stroke Patients Who are Candidates for Acute Reperfusion Therapy.

TABLE 407-9 RELATIVE CONTRAINDICATIONS TO IV rtPA IN PATIENTS WITHIN 3 TO 4.5 HOURS AFTER ONSET OF SYMPTOMS OF ACUTE ISCHEMIC STROKE

National Institutes of Health Stroke Scale >25 (see Table 407-4)
Age > 80 years old
Taking an oral anticoagulant regardless of INR
History of diabetes *and* a prior ischemic stroke

INR, international normalized ratio; IV, intravenous; and rtPA, recombinant tissue plasminogen activator.

Adapted from Jauch EC, Saver JL, Adams HP, Jr., et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* 2013;44:870-947.

UNUSUAL CAUSES OF STROKE

Ischemic strokes may be caused by a variety of rarer conditions. Specific treatments, many of which are not supported by extensive clinical trial data, vary accordingly (Table 407-10).

Cerebral Venous Thrombosis

Thrombosis of a cerebral venous sinus can cause headache, focal strokelike manifestations, seizures, altered mental status, and papilledema.[7] With superior sagittal sinus obstruction (see Fig. 406-10), patients can develop bilateral leg weakness and sensory changes. Obstruction of a transverse sinus or one of the major veins over the cerebral convexity (see Fig. 406-11) can also produce symptoms, depending on the area of the brain that is injured. Cerebral venous sinus thrombosis is an uncommon condition that is usually seen in patients with coagulopathies, disseminated cancer, or a prior inner ear infection. It can also occur in the peripartum period. Venous obstruction can mimic an ischemic arterial stroke, but symptoms and signs are often more diffuse and resemble encephalitis (Chapter 414) or meningitis (Chapter 412). The diagnosis can be suspected on routine CT or MRI and confirmed by CT or MR venography (Fig. 407-6). Initial acute treatment options include either body weight–adjusted subcutaneous low-molecular-weight heparin (see Table 38-2) or dose-adjusted intravenous heparin (see Table 81-4), even if patients have some degree of hemorrhage,[8] and one small randomized trial found that treatment with low-molecular-weight heparin was associated with lower mortality.[■] Oral warfarin anticoagulation should be started and continued for at least 3 months, with an INR target of 2.0 to 3.0. Longer periods of anticoagulation may be considered, depending on the cause of the sinus thrombosis.

Cervical Artery Dissection

A cervical artery dissection or cerebral artery dissection, each of which is caused by the formation and subsequent longitudinal extension of an intramural hematoma, can narrow or obstruct the arterial lumen. These dissections can be spontaneous, or they can be associated with major neck injury, relatively minor trauma (such as a chiropractic neck manipulation or neck hyperextension), or otherwise innocuous activities, such as coughing, sneezing, or lifting. Patients may have underlying fibromuscular dysplasia (Chapters 67 and 80); inherited conditions, such as Marfan syndrome (Chapter 260), Ehlers-Danlos syndrome, or tuberous sclerosis (Chapter 417); an elevated blood homocysteine level; or no identified underlying cause. The diagnosis can be challenging but should be considered especially in an otherwise healthy young patient who has neck or facial pain in conjunction with a stroke. MR angiography may show a hyperintense mass adjacent to a flow void, and MR angiography or catheter angiography can show a tapered lumen leading to an obstruction or even a double lumen. Treatment may include thrombolysis, anticoagulation, or endovascular or surgical repair, depending on individual circumstances, but no randomized trials are available to guide such decisions.

Vasculitis (Chapters 266, 270, and 271) can cause focal or multifocal cerebral ischemia due to local inflammation, stenosis, and even necrosis of extracranial or intracranial blood vessels.[9] Patients can have preexisting or concurrent headaches, cognitive changes, and seizures. Because vasculitis often involves multiple arteries, multiple foci of ischemic injury on neuroimaging studies may mimic multiple emboli. Cerebral angiography classically shows multiple areas of beadlike segmental narrowing, but findings may be normal. Similar findings may occur with other causes of intracranial vasculopathy, and the angiographic appearance is not specific. The diagnosis may require leptomeningeal/cortical biopsy, which may be negative because the inflammatory process can be multifocal rather than diffuse. Examples of vasculitides that can cause strokelike symptoms include primary central nervous system vasculitis, systemic lupus erythematosus (Chapter 266), rheumatoid vasculitis (Chapter 264), Behçet disease (Chapter 270), Takayasu arteritis (Chapters 78 and 270), temporal arteritis (Chapter 271), fibromuscular dysplasia (Chapters 67 and 80), granulomatosis with angiitis (Chapter 270), sarcoidosis (Chapter 95), meningovascular syphilis (Chapter 319), and lymphomatoid angioendotheliomatosis.

TABLE 407-10 USUAL CAUSES OF ISCHEMIC STROKE

CAUSE	SETTING	NOTES	POSSIBLE TREATMENT
Vasculitis (Chapter 270)	Patients commonly but not always have a prior known vasculitic condition	Vasculitis can be manifested with headache, cognitive impairment, multiple areas of infarction, or hemorrhage. Vasculitis affecting the cerebral vasculature can occur in the setting of systemic vasculitis or be confined to the central nervous system (primary vasculitis of the central nervous system). Diagnosis is supported by evidence of an inflammatory response in the spinal fluid, meningeal enhancement on MRI, and a typical pattern of focal stenoses on cerebral angiography. Brain/meningeal biopsy is often necessary to exclude other causes.	Steroids, immunosuppression
Sickle cell disease (Chapter 163)	Patients with known sickle cell disease; persons of African, Indian, and Mediterranean descent	Sickle cell disease can cause stroke by occlusion of small brain vessels or by intimal fibrosis leading to large-vessel occlusion. Patients with sickle cell disease can be monitored with transcranial Doppler ultrasound.	Transfusions to reduce hemoglobin S to <30% to 50% of total hemoglobin, hydroxyurea
Atrial myxoma (Chapter 60)	Usually not previously diagnosed; may be suspected on physical examination but usually otherwise asymptomatic and discovered on echocardiography	Atrial myxoma is the most common primary cardiac tumor and can lead to emboli.	Surgical removal of the tumor
Coagulation disorders (Chapter 176)	Young people with stroke of unknown cause or history suggesting coagulopathy, such as prior venous thrombosis, pulmonary embolism, multiple (particularly late) miscarriages	Prothrombotic coagulopathies can be inherited or acquired. These are more commonly associated with venous thrombosis and can lead to stroke in patients with a right to left shunt as paradoxical emboli. Late miscarriages in women or unprovoked deep venous thrombosis may be clues to an underlying coagulopathy. In addition to abnormalities of fibrinogen occurring in patients with cancer, anticardiolipin/antiphospholipid antibodies, and lupus anticoagulants are most commonly associated with ischemic stroke. Coagulopathies need to be considered in all patients with a cerebral venous sinus thrombosis.	Platelet antiaggregants or anticoagulation
Hyperviscosity	Usually caused by polycythemia vera, macroglobulinemia, or multiple myeloma	Hyperviscosity can cause ischemic stroke through occlusion of small intracranial vessels.	Treatment of underlying hematologic disorder
Cervical artery dissection	Often in otherwise healthy younger individuals	Dissections can be caused by trauma or occur spontaneously. Underlying causes include fibromuscular dysplasia (which can also affect renal arteries), Marfan syndrome, Ehlers-Danlos type IV, and tuberous sclerosis.	Dissections often heal spontaneously with no evidence of residual vascular injury. A period of anticoagulation is commonly used in patients with stroke related to dissection; however, data showing the benefit of the approach are lacking.
Aortic dissection (Chapter 78)	Marfan syndrome, Ehlers-Danlos syndrome	Aortic dissection can be manifested with chest pain radiating to the back.	Emergency surgical repair
Moyamoya	Usually discovered with intracranial vascular imaging, such as CT angiography, MR angiography, or catheter angiography	Moyamoya refers to neovascularization in patients with occlusion of the distal intracranial internal carotid arteries or proximal middle cerebral arteries. Moyamoya can be the result of other conditions (moyamoya syndrome) or occur without identifiable cause (moyamoya disease). Moyamoya can be associated with both ischemic stroke and brain hemorrhage.	Extracranial-intracranial bypass
Fabry disease (Chapter 208)	Stroke in the setting of typical historic features and physical examination findings; may be considered in young persons with stroke of unknown cause	Fabry disease is an X-linked disorder causing reduction in α-galactosidase. In addition to stroke, Fabry disease can cause angiokeratomas, acroparesthesias, hypohidrosis, corneal opacities, and renal and cardiac disease.	Enzyme replacement therapy is available but has not been shown to reduce stroke risk.

CT = computed tomography; MR = magnetic resonance.

Strokes occur in about 8 to 17% of patients with sickle cell disease and in about 2% of individuals with sickle cell trait (Chapter 163). Ischemic strokes are more common in children, whereas hemorrhagic strokes are more common in adults. Transfusion therapy can markedly reduce the risk of a first or recurrent stroke.

Drug-Related Causes of Stroke

A variety of legal and illicit drugs (Chapter 34) can precipitate an ischemic stroke. Intravenous drug users are more likely to develop bacterial endocarditis (Chapter 76), which can cause embolic stroke. Solid adulterants in injected material can reach the brain through an existing shunt, such as a patent foramen ovale, or they can cause local pulmonary arteriolitis that damages the endothelium and results in arteriovenous shunts through which microemboli can reach the brain. Potent vasoconstricting drugs (e.g., cocaine, ephedrine, phenylpropanolamine, and fenoxazoline) and dietary supplements (e.g., ephedra) may precipitate cerebral vasospasm and ischemic stroke, although hemorrhagic strokes are more common (Chapter 408). These drugs have been used in high doses as appetite suppressants, and case reports suggest that stroke may occur even after the first use of these products.

Rare Genetic Causes

A number of relatively rare genetic diseases can cause ischemic stroke. *Cerebral autosomal dominant arteriopathy with small subcortical infarcts and leukoencephalopathy* (CADASIL) can cause multiple deep infarcts and dementia in patients without other risk factors for stroke. A mutation in the Notch3 receptor gene on the short arm of chromosome 19 leads to an accumulation of Notch3 protein in vascular smooth muscle cells. The mean age at onset is about 40 years, although migraine with aura often antedates strokes by several years. Dementia usually develops within 10 to 15 years. Antenatal diagnosis is recommended in affected families. Treatment is symptomatic, with CADASIL-associated headaches potentially responding to acetazolamide (125 to 500 mg daily).

X-linked *Fabry disease* (angiokeratoma corporis diffusum) (Chapter 208) frequently includes cerebrovascular occlusion due to the accumulation of

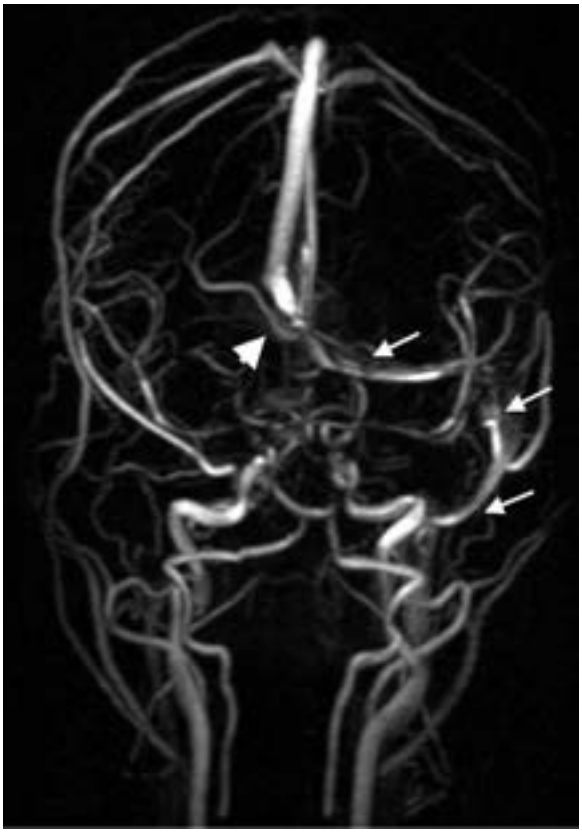

FIGURE 407-6. **Magnetic resonance venogram showing absent flow in the right transverse sinus (*arrowhead*) and sigmoid sinus and intact flow in the left transverse sinus and sigmoid sinus (*arrows*).**

glycolipids in small and medium-sized arteries. Enzyme replacement therapy is recommended, although it is not proven to reduce the risk of stroke. *Neurofibromatosis* (Chapter 417) can occlude the internal carotid arteries or the proximal part of the anterior cerebral circulation. Marfan syndrome (Chapter 260) can cause ischemic stroke due to dissection of the carotid arteries or related valvular heart disease.

Fat Embolism

Fat embolism (Chapter 98) after trauma to the long bones (Chapter 111), orthopedic procedures, and even severe trauma to large fat deposits can cause a stroke, usually several days later. Diffuse embolization can produce encephalopathy or seizures, but more focal emboli can be manifested as an ischemic stroke.

Cryptogenic Stroke

An echocardiogram may reveal an undiagnosed patent foramen ovale (Chapter 68) as a potential cause of a cryptogenic stroke. Despite a comprehensive evaluation, however, no definitive cause of stroke is found in 15 to 40% of patients with strokes. Randomized trials confirm that prolonged ECG monitoring with an event-triggered recorder or an insertable monitor increases the detection of atrial fibrillation to 9 to 16% compared with 1 to 3% with just 24-hour monitoring.[■■] Other initially cryptogenic strokes may be due to embolism from either a cardiac or other proximal arterial source.[10] Clinical trials are needed to determine the most appropriate antithrombotic management of these patients.

RECOVERY/REHABILITATION

The process of recovery begins even before the sequelae of acute brain injury have resolved. Multidisciplinary physiotherapy should include assessments by speech pathologists, physical therapists, and occupational therapists. Organized inpatient multidisciplinary rehabilitation is associated with a 34% lower odds of death, a 30% lower odds of death or institutionalization, and a 35% lower odds of death or dependency for patients with deficits warranting these services. All patients with stroke-related deficits should be assessed for rehabilitative interventions. Because depression can complicate stroke and affect recovery, all patients should be screened for depression.

PREVENTION Rx

Primary Prevention

Because more than 75% of strokes are first events, primary prevention of stroke is of paramount importance.[11,12] Following a healthy lifestyle (not smoking, following a diet low in sodium and rich in fruits and vegetables, getting at least 30 minutes of moderate or vigorous physical activity daily, having a body mass index below 25 kg/m^2, and consumption of no more than one alcoholic drink per day for women and one or two for men) is associated with an 80% lower risk of a first stroke compared with people who do not follow these lifestyles. The effect is graded, with increasing benefit depending on the number of healthy lifestyles an individual follows. There is no evidence that prophylactic treatment with aspirin or other antiplatelet drugs reduces the risk of stroke in low-risk individuals.

Risk factors that are amenable to treatment (see Table 407-1) include hypertension, diabetes, atrial fibrillation, and carotid stenosis. Treatment of hypertension dramatically reduces the risk of stroke.[13] Regular blood pressure screening and treatment of hypertension (see Tables 67-7 and 67-9) with a goal below 140/90 mm Hg are recommended. Blood pressure treatment and the use of a statin in patients with diabetes (Chapter 229) are recommended to lower the risk of a first stroke. Statins (Chapter 206) are also recommended to prevent a first ischemic stroke in patients with coronary heart disease. Although microvascular complications of diabetes are reduced with adequate glycemic control (target glycated hemoglobin level <7%), there is no evidence that tight control reduces the risk of stroke or coronary heart events. Patients who have atrial fibrillation are at increased risk of embolism and benefit from treatment with warfarin or a novel oral anticoagulant (Chapter 64).[14] Anticoagulation with warfarin is indicated for stroke prevention in patients with a mechanical prosthetic heart valve (Chapter 75). The U.S. Preventive Services Task Force recommends against screening for asymptomatic carotid artery stenosis.[15]

Prevention of Stroke in the Patient with Asymptomatic Carotid Stenosis

The benefit of carotid endarterectomy for patients with asymptomatic carotid stenosis is currently uncertain because of advances in medical therapy.[16] The risk of ipsilateral stroke associated with an asymptomatic carotid stenosis may be considerably less than 1% per year on the basis of older observational studies and clinical trials,[17] and the reported benefit of carotid endarterectomy depends on surgical success and complication rates that may not be widely achievable outside of randomized trials. Clinical trials are in progress comparing carotid revascularization in asymptomatic patients with current best medical therapy. Population screening for asymptomatic carotid stenosis is not recommended.

Secondary Prevention after a Transient Ischemic Attack or Stroke

Although clinical trials demonstrating the efficacy of lifestyle interventions for secondary stroke prevention are generally lacking, the same lifestyle behaviors associated with a reduced risk of a first stroke are an essential part of secondary stroke prevention.[18] Patients should routinely be prescribed an antiplatelet agent unless there are contraindications. Short-term dual antiplatelet therapy with aspirin and clopidogrel is more efficacious than single-agent therapy,[■] but long-term therapy is not and may cause more serious bleeding.[■] The choice of agent needs to be individualized, but aspirin (50 to 325 mg daily), clopidogrel (75 mg daily), or aspirin plus sustained-release dipyridamole (25/200 mg twice daily) are proven options.[19] The exception is the patient who has a specific indication for treatment with an anticoagulant, such as atrial fibrillation or a prosthetic heart valve, or in whom antithrombotic therapy is contraindicated.

Blood pressure reduction is recommended to lower the risk of recurrent stroke and other vascular events.[20] The precise timing of initiation of antihypertensive therapy after ischemic stroke is not established, but it can begin once the patient is stabilized after the acute period, generally after at least 24 hours. An average reduction of 10/5 mm Hg is associated with about a 25% reduction in the risk of recurrent stroke. In a randomized trial of patients with MRI-defined symptomatic lacunar infarctions, reducing systolic blood pressure to a target of less than 130 mm Hg starting 2 weeks later significantly reduced the rate of intracerebral hemorrhage and insignificantly reduced all subsequent stroke compared with a target of 130 to 149 mm Hg.[■] The choice of a specific antihypertensive regimen for secondary prevention should be individualized (see Tables 67-7 and 67-8); some data suggest that increased blood pressure variability, which is more common with β-blockers, blunts the benefit of blood pressure reduction for preventing recurrent stroke.

Patients with a prior stroke or TIA and known atherosclerotic disease, diabetes, or hyperlipidemia meeting criteria for statin therapy should be treated with a high-potency statin (e.g., 40 to 80 mg of atorvastatin or 20 to 40 mg of rosuvastatin daily; see also Table 206-5), unless contraindicated, to reduce the risk of recurrent stroke and of other cardiovascular events.[21] Stopping a statin

in the setting of an acute ischemic stroke is associated with increased morbidity and mortality.

In addition to these general measures, additional specific treatment for secondary stroke prevention depends on the cause of the stroke. Atrial fibrillation–related stroke is associated with a high risk of recurrence (i.e., 6 to 10% annually). Patients with atrial fibrillation–related stroke should be treated with warfarin or a novel oral anticoagulant (e.g., direct thrombin inhibitor, factor Xa inhibitor). Anticoagulation with warfarin is indicated in patients with stroke related to acute MI.

Procedures and Devices

The rate of recurrent stroke after stroke or TIA related to a high-grade (70 to 99%) extracranial carotid artery stenosis may be as high as 25% during the next 2 years, with the highest risk in the first weeks after the index event. Because this risk of recurrence is decreased by 50% with successful carotid revascularization, selected patients with stroke associated with 70 to 99% extracranial carotid artery stenosis within the prior 6 months benefit from carotid revascularization, provided the procedure can be performed with less than 6% morbidity. Depending on characteristics such as age, gender, comorbid conditions, and increasing time since the index event, patients with a 50 to 69% extracranial carotid stenosis may also benefit from revascularization. Patients with less than 50% carotid stenosis do not benefit from carotid revascularization. Extracranial-intracranial bypass does not reduce the risk of recurrent strokes in patients with complete occlusion of an extracranial carotid artery.

Stroke related to a high-grade (i.e., 50 to 99%) intracranial stenosis is associated with a high risk of recurrence. Aspirin is preferred to warfarin in these patients because it is as efficacious and is associated with fewer complications. Intracranial angioplasty/stenting is no better than medical treatment (e.g., aggressive lifestyle modification, platelet antiaggregant therapy, and a high-potency statin) for the prevention of a recurrent stroke in this setting.

Stenting is generally not as efficacious as endarterectomy for secondary stroke prevention in patients with carotid stenosis, although younger patients may have a lower combined risk of stroke, MI, and death with stenting and older patients may do better with endarterectomy.[22]

Intention-to-treat analysis of data from three prospective randomized trials of transcatheter closure of a patent foramen ovale in patients with cryptogenic stroke found a nonsignificant reduction in the combined risk of stroke and TIA and in the risk of stroke with closure compared with medical therapy alone. There remains no Food and Drug Administration–approved device for this purpose, and additional clinical trials are in progress.

PROGNOSIS

TIA is a major risk factor for stroke and requires urgent evaluation to detect specific causes that may require immediate treatment. Overall, approximately 10% of patients who have a TIA will have a stroke within 90 days, with almost half occurring within 2 days. The strokes that occur are frequently disabling or fatal. Factors associated with higher risk include age older than 60 years, diabetes, impaired speech or weakness, symptoms lasting more than 10 minutes, and evidence of ischemic injury on brain MRI. After the acute period, about 20% of patients who had a TIA will have a stroke during the next 10 years.

Stroke-related mortality varies by age. The 30-day stroke mortality rate is estimated to be 9% for patients aged 65 to 74 years, 13% for patients aged 74 to 84 years, and 23% for patients older than 85 years. About 30% of patients who have had a stroke will have a recurrent stroke within 5 years. Stroke is also a leading cause of disability. Among stroke survivors, approximately 45% have cognitive deficits, 30% are unable to walk without assistance, 25% are institutionalized, and 25% are dependent in activities of daily living after 6 months.

Grade A References

A1. Emberson J, Lees KR, Lyden P, et al. Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomised trials. *Lancet*. 2014;384:1929-1935.

A2. Xian Y, Liang L, Smith EE, et al. Risks of intracranial hemorrhage among patients with acute ischemic stroke receiving warfarin and treated with intravenous tissue plasminogen activator. *JAMA*. 2012;307:2600-2608.

A3. Sandercock PA, Counsell C, Tseng MC, et al. Oral antiplatelet therapy for acute ischaemic stroke. *Cochrane Database Syst Rev*. 2014;3:CD000029.

A4. Jüttler E, Unterberg A, Woitzik J, et al. Hemicraniectomy in older patients with extensive middle-cerebral-artery stroke. *N Engl J Med*. 2014;370:1091-1100.

A5. He J, Zhang Y, Xu T, et al. Effects of immediate blood pressure reduction on death and major disability in patients with acute ischemic stroke: the CATIS randomized clinical trial. *JAMA*. 2014;311:479-489.

A6. Dennis M, Sandercock P, Reid J, et al. Effectiveness of intermittent pneumatic compression in reduction of risk of deep vein thrombosis in patients who have had a stroke (CLOTS 3): a multicentre randomised controlled trial. *Lancet*. 2013;382:516-524.

A7. Misra UK, Kalita J, Chandra S, et al. Low molecular weight heparin versus unfractionated heparin in cerebral venous sinus thrombosis: a randomized controlled trial. *Eur J Neurol*. 2012;19:1030-1036.

A8. Wang WC, Dwan K. Blood transfusion for preventing primary and secondary stroke in people with sickle cell disease. *Cochrane Database Syst Rev*. 2013;11:CD003146.

A9. Gladstone DJ, Spring M, Dorian P, et al. Atrial fibrillation in patients with cryptogenic stroke. *N Engl J Med*. 2014;370:2467-2477.

A10. Sanna T, Diener HC, Passman RS, et al. Cryptogenic stroke and underlying atrial fibrillation. *N Engl J Med*. 2014;370:2478-2486.

A11. Lee M, Saver JL, Hong KS, et al. Risk-benefit profile of long-term dual- versus single-antiplatelet therapy among patients with ischemic stroke: a systematic review and meta-analysis. *Ann Intern Med*. 2013;159:463-470.

A12. Wang Y, Wang Y, Zhao X, et al. Clopidogrel with aspirin in acute minor stroke or transient ischemic attack. *N Engl J Med*. 2013;369:11-19.

A13. Benavente OR, Coffey CS, Conwit R, et al. Blood-pressure targets in patients with recent lacunar stroke: the SPS3 randomised trial. *Lancet*. 2013;382:507-515.

A14. Powers WJ, Clarke WR, Grubb RL Jr, et al. Extracranial-intracranial bypass surgery for stroke prevention in hemodynamic cerebral ischemia: the Carotid Occlusion Surgery Study randomized trial. *JAMA*. 2011;306:1983-1992.

A15. Derdeyn CP, Chimowitz MI, Lynn MJ, et al. Aggressive medical treatment with or without stenting in high-risk patients with intracranial artery stenosis (SAMMPRIS): the final results of a randomised trial. *Lancet*. 2014;383:333-341.

A16. Liu ZJ, Fu WG, Guo ZY, et al. Updated systematic review and meta-analysis of randomized clinical trials comparing carotid artery stenting and carotid endarterectomy in the treatment of carotid stenosis. *Ann Vasc Surg*. 2012;26:576-590.

A17. Pineda AM, Nascimento FO, Yang SC, et al. A meta-analysis of transcatheter closure of patent foramen ovale versus medical therapy for prevention of recurrent thromboembolic events in patients with cryptogenic cerebrovascular events. *Catheter Cardiovasc Interv*. 2013;82:968-975.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Sacco RL, Kasner SE, Broderick JP, et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2013;44:2064-2089.
2. Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics—2013 update: a report from the American Heart Association. *Circulation*. 2013;127:e6-e245.
3. Qureshi AI, Caplan LR. Intracranial atherosclerosis. *Lancet*. 2014;383:984-998.
4. Windecker S, Stortecky S, Meier B. Paradoxical embolism. *J Am Coll Cardiol*. 2014;64: 403-415.
5. Jauch EC, Saver JL, Adams HP Jr, et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2013;44:870-947.
6. Saver JL, Fonarow GC, Smith EE, et al. Time to treatment with intravenous tissue plasminogen activator and outcome from acute ischemic stroke. *JAMA*. 2013;309:2480-2488.
7. Weimar C. Diagnosis and treatment of cerebral venous and sinus thrombosis. *Curr Neurol Neurosci Rep*. 2014;14:417.
8. Einhaupl K, Stam J, Bousser MG, et al. EFNS guideline on the treatment of cerebral venous and sinus thrombosis in adult patients. *Eur J Neurol*. 2010;17:1229-1235.
9. Hajj-Ali RA, Calabrese LH. Diagnosis and classification of central nervous system vasculitis. *J Autoimmun*. 2014;48-49:149-152.
10. Hart RG, Diener HC, Coutts SB, et al. Embolic strokes of undetermined source: the case for a new clinical construct. *Lancet Neurol*. 2014;13:429-438.
11. Goldstein LB, Bushnell CD, Adams RJ, et al. Guidelines for the primary prevention of stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42:517-584.
12. Bushnell C, McCullough L. Stroke prevention in women: synopsis of the 2014 American Heart Association/American Stroke Association guideline. *Ann Intern Med*. 2014;160:853-857.
13. James PA, Oparil S, Carter BL, et al. 2014 Evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA*. 2014;311:507-520.
14. Culebras A, Messe SR, Chaturvedi S, et al. Summary of evidence-based guideline update: prevention of stroke in nonvalvular atrial fibrillation: report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology*. 2014;82:716-724.
15. LeFevre ML. Screening for asymptomatic carotid artery stenosis: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;161:356-362.
16. Jonas DE, Feltner C, Amick HR, et al. Screening for asymptomatic carotid artery stenosis: a systematic review and meta-analysis for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014;161:336-346.
17. Grotta JC. Clinical practice. Carotid stenosis. *N Engl J Med*. 2013;369:1143-1150.
18. Eckel RH, Jakicic JM, Ard JD, et al. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129:S76-S99.
19. Furie KL, Kasner SE, Adams RJ, et al. Guidelines for the prevention of stroke in patients with stroke or transient ischemic attack: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42:227-276.
20. Go AS, Bauman MA, Coleman King SM, et al. An effective approach to high blood pressure control: a science advisory from the American Heart Association, the American College of Cardiology, and the Centers for Disease Control and Prevention. *Hypertension*. 2014;63:878-885.
21. Stone NJ, Robinson J, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129(suppl 2):S1-S45.
22. Voeks JH, Howard G, Roubin GS, et al. Age and outcomes after carotid stenting and endarterectomy: the carotid revascularization endarterectomy versus stenting trial. *Stroke*. 2011;42: 3484-3490.

REVIEW QUESTIONS

1. A patient complains of impaired vision in his left eye and has a left arm weakness. Examination shows a superior temporal visual defect in his left eye, and a cholesterol plaque is seen in his left eye on ophthalmoscopy. He has a left lower facial paresis and weakness in his left arm. Which of the following is the most likely cause of the stroke?

A. Stenosis of the left carotid artery
B. Stenosis of the left anterior choroidal artery
C. Right basal ganglia hemorrhage
D. Stenosis of the left vertebral artery
E. Embolism from the ascending aorta

Answer: E Two vascular territories are involved simultaneously: the left carotid artery supplies the left ophthalmic artery, but the left-sided weakness would not be explained by a stroke in the left carotid distribution.

2. Which of the following is a contraindication to treatment with intravenous tPA (tissue plasminogen activator) in a patient with an acute ischemic stroke?

A. Ischemic stroke 1 year ago
B. Blood pressure 170/85 mm Hg
C. Intracerebral hemorrhage 4 years ago
D. Prior treatment with intravenous tPA for acute stroke 1 year ago
E. Pure motor stroke consistent with a "lacunar" syndrome

Answer: C A prior ischemic stroke within 3 months is a contraindication to tPA, but not a stroke a year ago. Patients cannot be treated with a systolic blood pressure above 185 mm Hg or diastolic blood pressure above 110 mm Hg that does not respond to antihypertensive treatment. Patients who received intravenous tPA for a prior stroke can be treated, as long as they are beyond the 3-month period. Patients with ischemic stroke in any vascular distribution can be treated. However, intravenous tPA is contraindicated in patients with a prior history of intracerebral hemorrhage.

408

HEMORRHAGIC CEREBROVASCULAR DISEASE

STEPHAN A. MAYER

Approximately 20% of all strokes are due to spontaneous intracranial hemorrhage, of which about three quarters are caused by intracerebral hemorrhage and one quarter by subarachnoid hemorrhage. Intracerebral hemorrhage, which is most frequently caused by the rupture of small penetrating arteries, results in a focal collection of clot within the brain parenchyma. By comparison, subarachnoid hemorrhage is caused by rupture of vessels on the brain's surface, most often due to a congenital aneurysm, and results in diffusion of blood throughout the cerebrospinal fluid (CSF) spaces. In about 40% of either form of hemorrhagic stroke, blood extends into the brain's ventricles, a devastating complication known as intraventricular hemorrhage. Both types of hemorrhagic stroke have high mortality rates, but recovery and survival have improved in recent decades owing to advances in neurocritical care.

SUBARACHNOID HEMORRHAGE

EPIDEMIOLOGY AND PATHOBIOLOGY

About 30,000 new cases of spontaneous subarachnoid hemorrhage occur each year in the United States, predominantly involving young adults. Subarachnoid hemorrhage accounts for 5% of all strokes, with an incidence of approximately 1 in 10,000 individuals annually. Women are affected more than men are, and the rate is twice as high in African Americans as in whites. About 10% of patients with subarachnoid hemorrhage have a first-degree relative who has also had a subarachnoid hemorrhage, even in the absence of an identifiable genetic predisposition such as polycystic kidney disease (Chapter 127), Marfan syndrome (Chapter 260), or Ehlers-Danlos syndrome (Chapter 260). Modifiable risk factors for subarachnoid hemorrhage include cigarette smoking, heavy alcohol use, chronic and poorly controlled hypertension, and use of sympathomimetic agents such as cocaine and phenylpropanolamine. The use of warfarin, especially if the international

TABLE 408-1 NONANEURYSMAL CAUSES OF SUBARACHNOID HEMORRHAGE (IN APPROXIMATE ORDER OF FREQUENCY)

Trauma
Idiopathic perimesencephalic subarachnoid hemorrhage
Arteriovenous malformation
Intracranial arterial dissection (Chapter 407)
Cocaine and amphetamine use (Chapter 34)
Mycotic aneurysm (Chapter 76)
Pituitary apoplexy (Chapter 224)
Moyamoya disease (Chapter 407)
Central nervous system vasculitis (Chapter 270)
Sickle cell disease (Chapter 163)
Coagulation disorders (Chapters 172, 174, and 175)
Primary or metastatic neoplasm (Chapter 189)

normalized ratio (INR) is higher than 3, but not of aspirin is associated with an increased risk of intracerebral hemorrhage compared with no therapy.[1]

In 80% of cases, subarachnoid hemorrhage is caused by a rupture of an intracranial saccular or berry aneurysm. In about half of subarachnoid hemorrhages in which an aneurysm is not identified, the blood has a focal perimesencephalic distribution around the midbrain or anterior to the pons. In these cases, the source of bleeding is probably venous. In the remainder of cases of nonaneurysmal subarachnoid hemorrhage, the bleeding source is usually a thin-walled arterial "blister." Other causes include arteriovenous malformations; mycotic aneurysms (Chapter 76); vasculitis (Chapter 270); tumors (Chapter 189); and severe coagulation disorders, such as hemophilia (Chapter 174), marked thrombocytopenia (Chapter 172), and disseminated intravascular coagulation (Chapter 175) (Table 408-1).

CLINICAL MANIFESTATIONS

The classic symptom of subarachnoid hemorrhage is a rapidly developing, severe "thunderclap" headache, which the patient typically refers to as the "worst headache of my life." The headache is usually generalized, but focal pain may refer to the site of aneurysmal rupture (e.g., periorbital pain related to an ophthalmic artery aneurysm). Commonly associated symptoms include stiff neck, loss of consciousness, nausea, vomiting, back or leg pain, and photophobia. In patients who lose consciousness, tonic posturing may occur and may be difficult to differentiate from a seizure. Although aneurysmal rupture often occurs during periods of exercise or physical stress, subarachnoid hemorrhage can occur at any time, including sleep. More than one third of patients give a history of a premonitory, "sentinel headache" in the days to weeks before presenting with subarachnoid hemorrhage. These prodromal symptoms are usually caused by minor leaks of blood from the aneurysm, but they also may be caused by acute thrombosis or expansion of an aneurysm.

About 5% of patients with subarachnoid hemorrhage present with signs related to local compression of the cranial nerves or brain stem from a large aneurysm and have focal neurologic signs that may point to the site of bleeding and clot formation. The most common syndrome is a third cranial nerve palsy, which is manifested as ptosis, exodeviation of the eye, and anisocoria due to compression from a large posterior communicating artery aneurysm. Large fusiform dolichoectatic aneurysms most frequently occur in the basilar artery and can be manifested with signs and symptoms related to brain stem or cranial nerve compression. Hemiparesis or aphasia suggests a middle cerebral artery aneurysm, and paraparesis or abulia suggests an aneurysm of the proximal anterior cerebral artery.

DIAGNOSIS

In about 15% of patients, subarachnoid hemorrhage is initially misdiagnosed as a migraine headache or viral syndrome, especially in patients with milder symptoms. Neck stiffness, seizures, diastolic blood pressure above 110 mm Hg, vomiting, and headache increase the likelihood of a hemorrhagic stroke rather than an ischemic stroke, but neuroimaging is required for reliable diagnosis.[2] Approximately 40% of misdiagnosed patients experience subsequent neurologic deterioration due to rebleeding, hydrocephalus, or vasospasm before returning to medical attention. A high degree of vigilance is required to establish the diagnosis of subarachnoid hemorrhage, by either computed tomography (CT) or lumbar puncture if the initial CT scan is normal.

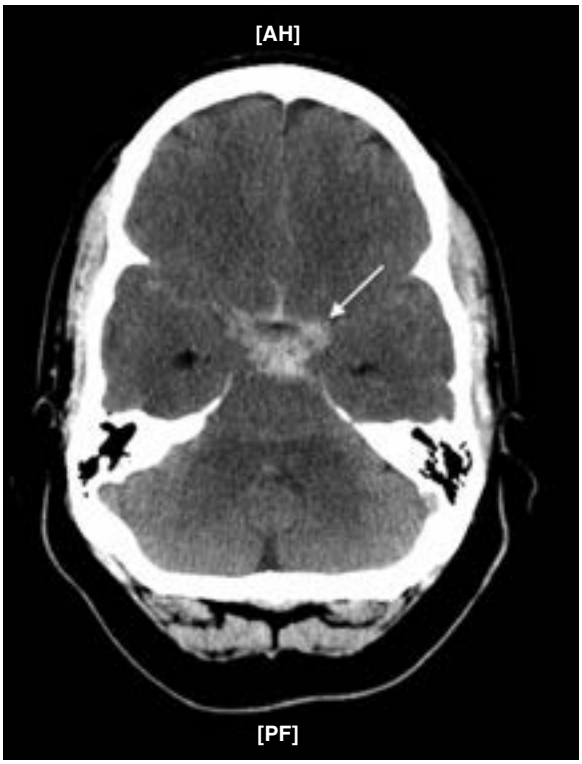

FIGURE 408-1. Computed tomographic imaging. Subarachnoid hemorrhage (note hyperintensity in the suprasellar cistern, *arrow*). (Courtesy Dr. Larry B. Goldstein.)

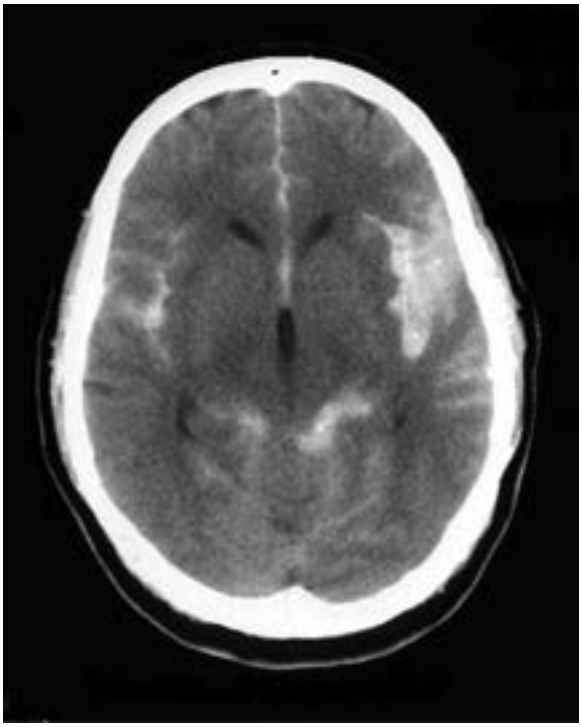

FIGURE 408-2. Computed tomographic scan showing a diffuse, thick subarachnoid hemorrhage.

Computed Tomography

Any patient with suspected subarachnoid hemorrhage should then be sent immediately for an emergency CT scan, which will almost always reveal blood within the basal cisterns if it is performed within 24 hours of the onset of symptoms (Fig. 408-1). More severe hemorrhage can extend into the interhemispheric and bilateral sylvian fissures (Fig. 408-2). The sensitivity of CT declines, however, as time passes from the clinical onset of bleeding, and

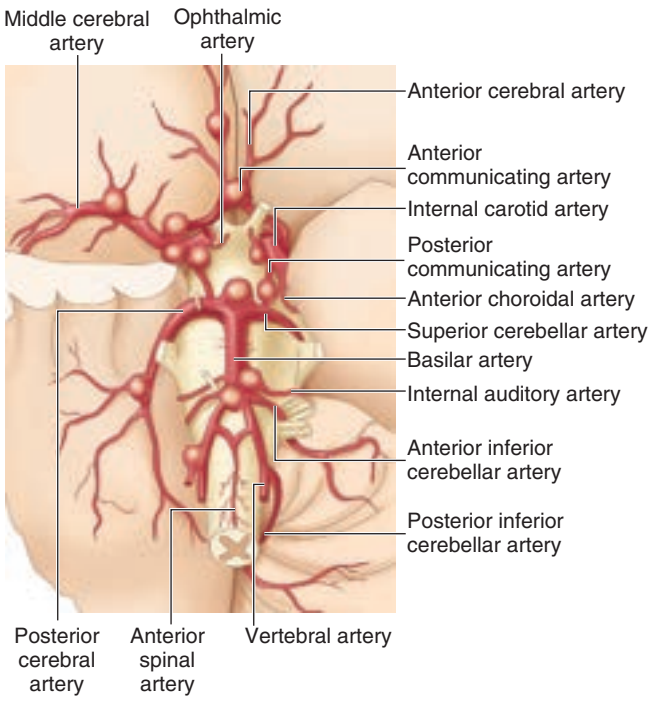

FIGURE 408-3. Saccular aneurysms. Saccular or berry aneurysms typically develop at the bifurcations of arteries on the undersurface of the brain. (Courtesy Dr. Justin Zivin.)

the sensitivity of CT is only about 75% by 48 hours after the onset of symptoms. As a result, lumbar puncture is mandatory if the CT scan is normal but the index of clinical suspicion remains high.

Lumbar Puncture

The CSF is usually grossly bloody. Subarachnoid hemorrhage can be differentiated from a traumatic tap by the xanthochromic (yellow-tinged) appearance of the supernatant fluid after centrifugation. However, xanthochromia may not be present until up to 12 hours after the subarachnoid hemorrhage. The CSF pressure is nearly always high, and the protein level is elevated. Initially, the proportion of CSF leukocytes to erythrocytes is that of peripheral blood, with a usual ratio of 1 : 700; after several days, however, a sterile chemical meningitis caused by blood in the CSF may induce a reactive leukocytosis with a low CSF glucose level. Red blood cells and xanthochromia disappear in about 2 weeks unless hemorrhage recurs.

Angiography

Cerebral angiography is the definitive diagnostic procedure to detect intracranial aneurysms and to define their anatomy (Fig. 408-3). Although the increasing availability and image quality of CT and magnetic resonance angiography have allowed some centers to use these tests to make the initial diagnosis, four-vessel (bilateral internal carotid and vertebral artery injections) angiography is mandatory when results of those tests are negative. In approximately 20% of cases of subarachnoid hemorrhage, the initial angiogram is normal. Vasospasm, local thrombosis, or poor technique can lead to a false-negative angiogram. For this reason, patients with an initially normal angiogram should have a follow-up study 1 to 2 weeks later; an aneurysm will be demonstrated in about 5% of these cases. The exception to this rule is found in patients with perimesencephalic subarachnoid hemorrhage, who do not require follow-up angiography.

Magnetic Resonance Imaging

Conventional magnetic resonance imaging (MRI) sequences (T1- or T2-weighted scans) are generally less sensitive than CT scans for detecting blood. Susceptibility-weighted imaging may be useful for documenting a completely thrombosed aneurysm in selected patients who have subarachnoid hemorrhage but a normal angiogram.

Laboratory Testing

In addition to routine admission laboratory tests, care should be taken to check an INR, partial thromboplastin time, and platelet count to diagnose a coagulopathy; an electrocardiogram and serum troponin level to diagnose sympathetically mediated cardiac injury; and a chest radiograph to look for neurogenic pulmonary edema or aspiration pneumonitis.

TABLE 408-2 MORTALITY ACCORDING TO THE HUNT-HESS GRADING SCALE FOR ANEURYSMAL SUBARACHNOID HEMORRHAGE

GRADE	CLINICAL FINDINGS	HOSPITAL MORTALITY (%)
I	Asymptomatic or mild headache	5
II	Moderate to severe headache, or oculomotor palsy	5
III	Confused, drowsy, or mild focal signs	10
IV	Stupor (localizes to pain)	34
V	Coma (posturing or no motor response to pain)	52
Total		20

Data are from patients treated at Columbia University Medical Center.

TREATMENT

The initial goals of treatment are to minimize acute primary brain injury in poor-grade patients with a depressed level of consciousness (Hunt-Hess grades III to V; Table 408-2), to manage secondary complications, and to prevent rebleeding.[3] Mortality after subarachnoid hemorrhage is substantially lower when patients are treated at high-volume regional centers with access to skilled interventionalists and specialized neurocritical care.

Severe primary brain injury related to the acute effects of hemorrhage is the leading cause of death and disability after subarachnoid hemorrhage, and comprehensive management can reduce both morbidity and mortality (Table 408-3). In poor-grade patients, the immediate concern in the emergency department is reducing intracranial pressure (ICP) and preventing secondary cerebral hypoxic-ischemic injury. Patients with an impaired ability to protect the airway should be intubated with supplemental oxygen given as needed. Hypotension should be treated aggressively with fluids and vasopressors to maintain a mean arterial pressure of 90 mm Hg (Chapter 106). Stuporous or comatose patients with extensive subarachnoid blood, intraventricular hemorrhage, acute obstructive hydrocephalus, or global cerebral edema should be empirically treated for intracranial hypertension with 1.0 g/kg of 20% mannitol before emergent placement of an external ventricular drain. Repeated doses of mannitol or 0.5 to 2.0 mL of 23.4% hypertonic saline can be repeated hourly as needed to reduce ICP to less than 20 mm Hg or to reverse signs of transtentorial herniation, with careful monitoring of serum sodium, osmolality, and volume status.

Cerebral Aneurysms

Recurrent bleeding is a devastating complication of subarachnoid hemorrhage. Preventive strategies depend on the cause, which is usually a cerebral aneurysm, and the site of the initial bleed.

Saccular or berry aneurysms most often occur at the circle of Willis or its major branches, especially at bifurcations. They arise where the arterial elastic lamina and tunica media are defective, tend to enlarge with age, and can become paper-thin. Because saccular aneurysms are rarely detected in children and the incidence of subarachnoid hemorrhage increases with age, it seems clear that congenital wall defects develop into aneurysms only over time. The site of rupture is usually through the dome of the aneurysm. Approximately 15% of patients who present with subarachnoid hemorrhage from an identifiable aneurysm also harbor another unruptured intracranial aneurysm.

Intracranial aneurysms are seen in about 2% of adults, thereby suggesting that approximately 2 to 3 million Americans have an aneurysm. However, more than 90% of these aneurysms are small (<10 mm) and remain asymptomatic throughout life. The annual risk of rupture of an asymptomatic intracranial aneurysm is approximately 0.7%. Important risk factors for the initial rupture of an intracranial aneurysm include increasing size, prior subarachnoid hemorrhage from another aneurysm, active cigarette smoking, and location in the basilar apex and posterior communicating artery. CT angiography is a sensitive (97%) and specific (98%) test for the diagnosis of cerebral aneurysms.[4]

If an aneurysm rebleeds, approximately 50% of patients die immediately, and another 30% suffer incremental brain injury. The risk of rebleeding is highest within the first 24 hours after the initial aneurysmal rupture (4%) and remains elevated (approximately 1% to 2% per day) for the next 4 weeks (Fig. 408-4). The cumulative risk of rebleeding in untreated patients is 20% at 2

TABLE 408-3 MANAGEMENT PROTOCOL FOR ACUTE SUBARACHNOID HEMORRHAGE

Blood pressure	• Control elevated blood pressure during the preoperative phase (systolic blood pressure <160 mm Hg) with IV labetalol or nicardipine (see Table 67-14) to prevent rebleeding.
Rebleeding prophylaxis	• ε-Aminocaproic acid 4 g IV on diagnosis followed by 1 g/hr until aneurysm repair, for a maximum of up to 72 hours after ictus
Intravenous hydration	• Normal (0.9%) saline at 1.0-1.5 mL/kg/hr
Laboratory testing	• Periodically check complete blood count and electrolytes. • Obtain serial ECGs and check admission cardiac troponin level to evaluate for cardiac injury; perform echocardiography in patients with abnormal ECG findings or elevated troponin levels.
Seizure prophylaxis	• Fosphenytoin or phenytoin IV load (15-20 mg/kg); discontinue on postoperative day 1 unless patient has seized, is poor grade, or has focal cortical disease or is otherwise unstable. Starting dose is 300 mg daily IV adjusted to maintain therapeutic levels of 10-20 mg/dL.
Vasospasm prophylaxis	• Nimodipine 60 mg orally every 4 hours until day SAH 21 or discharge
Physiologic homeostasis	• Cooling blankets to maintain temperature ≤37.5° C • Insulin drip to maintain glucose 100-120 mg/dL • Transfuse to maintain hemoglobin >7.0 g/dL (in the absence or active cerebral or cardiac ischemia)
Cerebral edema	• Mannitol 1.0 g/kg IV or 30 mL of 23.4% hypertonic saline solution as needed to maintain ICP <20 mm Hg
Hydrocephalus	• Emergent external ventricular drain placement in all stuporous/comatose patients (Hunt-Hess IV/V) as well as lethargic patients with hydrocephalus
Vasospasm diagnosis	• Transcranial Doppler sonography every 1 or 2 days until the tenth day after SAH • Computed tomography angiography and perfusion on day 4-8 after SAH or for neuroworsening
Therapy for symptomatic vasospasm	• Place patient in Trendelenburg (head down) position. • Infuse 1 liter normal saline during 30 minutes. • If the deficit persists, raise the systolic blood pressure with phenylephrine or norepinephrine until the deficit resolves (target 180-220 mm Hg). • If refractory, monitor cardiac output and add dobutamine or milrinone to maintain cardiac index ≥4.0 L/min/m^2. • Transfuse to maintain hemoglobin >10.0 g/dL. • Emergency angiography for intra-arterial verapamil or cerebral angioplasty unless the patient responds well to the above measures.

ECG = electrocardiogram; ICP = intracranial pressure; SAH = subarachnoid hemorrhage.

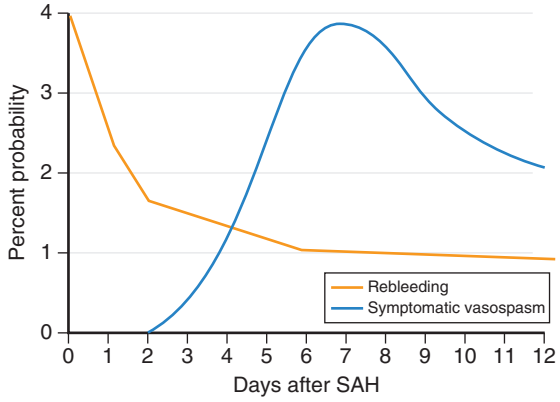

FIGURE 408-4. The daily percentage probability for the development of symptomatic vasospasm or rebleeding after subarachnoid hemorrhage (SAH). Day 0 denotes day of onset of subarachnoid hemorrhage.

weeks and 30% at 1 month. Poor clinical grade and larger aneurysms are the strongest risk factors for in-hospital rebleeding.

Prolonged infusion (i.e., 2 weeks) of antifibrinolytic agents such as ε-aminocaproic acid or tranexamic acid reduces the risk of rebleeding but does not improve outcomes because of an increased risk of cerebral ischemia.[■] However, very short-term therapy (e.g., ε-aminocaproic acid 4 g followed by 1 g/hour until 4 hours before angiography for a maximum of 72 hours after the onset of hemorrhage) appears to be beneficial.[■] Control of arterial hypertension (maintain systolic blood pressure <160 mm Hg) and administration of an anticonvulsant to minimize the risk of acute seizure activity (phenytoin 20 mg/kg intravenously is most commonly used) may also reduce acute rebleeding, although trials to support this practice are lacking.

Complete obliteration of a ruptured saccular aneurysm by either endovascular coiling or surgical clipping is the definitive treatment for the prevention of rebleeding and should be performed as an emergency procedure. For small to medium-sized anterior circulation aneurysms in good-grade patients, endovascular coil embolization results in better 6-month outcomes than does surgical clipping,[■] with a 23% lower likelihood of death or disability at 1 year. The only exception to this rule is Hunt-Hess grade V patients, who have an extremely poor neurologic prognosis. In addition to preventing rebleeding, early aneurysm repair permits treatment of symptomatic vasospasm with hypertensive hypervolemic therapy, which is not feasible with an unprotected aneurysm.

Endovascular coil embolization involves packing of the ruptured aneurysm with soft, thrombogenic detachable platinum coils. This procedure leads to complete obliteration of small to medium-sized aneurysms (<10 mm in diameter) in 80 to 90% of cases, with an acceptable complication rate of approximately 10%. About 5% of coiled patients develop recurrent dilation at the neck of the original aneurysm and require repeated coil embolization or delayed surgical clipping.

Surgical clipping requires a craniotomy to expose the aneurysm, after which the neck of the aneurysm is clipped under an operating microscope to exclude it entirely from the parent artery. Surgical repair carries a 5 to 15% risk of major morbidity or mortality, especially a stroke due to inadvertent occlusion of an adjacent vessel or intraoperative rebleeding. The risk of surgery increases with larger aneurysms and is lower in the hands of more experienced operators. Clipping is preferred and sometimes the only option for treatment of wide-necked aneurysms that are unable to safely contain a coil mass without migration onto the parent artery.

Fusiform Aneurysms

Fusiform aneurysms are elongated, atherosclerotic ectasias of large arteries. They are usually in the basilar artery but can be seen in the internal, middle, and anterior cerebral arteries. As fusiform aneurysms progressively dilate, they compress surrounding structures and cause focal neurologic dysfunction, such as facial pain (cranial nerve V), hemifacial spasm (cranial nerve VII), and hearing loss with vertigo (cranial nerve VIII). Fusiform aneurysms even can mimic pituitary (Chapter 224) and suprasellar mass lesions or cerebellopontine angle tumors (Chapter 189). Fortunately, fusiform aneurysms rarely rupture; but if they do, total occlusion is usually required because their stiff walls and shape make surgical clipping difficult.

Mycotic Aneurysms

An infected embolism, usually from infectious endocarditis (Chapter 76), may lodge in a distal branch of a cerebral artery, where it causes microinfarction or microabscesses. The artery may rupture acutely, or focal arteritis and mycotic aneurysms may develop. Up to 10% of these aneurysms, which are often multiple and in distal cerebral arteries, may eventually rupture, but treatment other than as for the endocarditis itself is uncertain. As a result, diagnostic imaging usually is undertaken only after symptoms appear, and the potential value of serial imaging is controversial. Anticoagulation is contraindicated in the setting of acute septic emboli to the brain because of the high risk of hemorrhagic complications.

Other Causes of Subarachnoid Hemorrhage

In patients who suffer subarachnoid hemorrhage from other causes, treatment is aimed at the underlying condition. In patients with an idiopathic perimesencephalic venous subarachnoid hemorrhage, rebleeding is rare, symptomatic vasospasm does not occur, and no specific treatment is

TABLE 408-4 MODIFIED FISHER COMPUTED TOMOGRAPHY RATING SCALE FOR THE PREDICTION OF SYMPTOMATIC VASOSPASM

GRADE	CRITERIA	PERCENTAGE OF AFFECTED PATIENTS	FREQUENCY OF	
			*Delayed Cerebral Ischemia**	*Infarction*
0	No SAH or IVH	5%	0%	0%
1	Minimal/thin SAH, no biventricular IVH	30%	12%	6%
2	Minimal/thin SAH, *with* biventricular IVH	5%	21%	14%
3	Thick SAH,† no biventricular IVH	43%	19%	12%
4	Thick SAH, *with* biventricular IVH	17%	40%	28%
	All patients	100%	20%	12%

IVH = intraventricular hemorrhage; SAH = subarachnoid hemorrhage.
*Delayed cerebral ischemia is defined as symptomatic deterioration, cerebral infarction, or both resulting from vasospasm.
†Thick SAH is defined as completely filling at least one cistern or fissure.
From Claassen J, Bernardini GL, Kreiter K, et al. Effect of cisternal and ventricular blood on risk of delayed cerebral ischemia after subarachnoid hemorrhage: the Fisher scale revisited. *Stroke.* 2001;32:2012-2020, with permission.

indicated. Coagulation and platelet disorders require prompt treatment (Chapters 173 and 174) to prevent further bleeding. Arteriovenous malformations, which more commonly cause intracerebral rather than subarachnoid hemorrhage, are discussed later.

Prevention, Diagnosis, and Treatment of Vasospasm

Delayed cerebral ischemia from vasospasm accounts for a large proportion of morbidity and mortality after subarachnoid hemorrhage.[5] Progressive arterial narrowing develops in approximately 70% of patients, but delayed ischemic deficits develop in only 20 to 30%. The process begins 3 to 5 days after the hemorrhage, becomes maximal at 5 to 14 days, and gradually resolves during 2 to 4 weeks (see Fig. 408-4). The most important risk factor for symptomatic vasospasm is thick cisternal or intraventricular clot, which can be graded by the modified Fisher scale (Table 408-4).

Blood pressure control can be liberalized once the aneurysm has been repaired and brain perfusion becomes the dominant consideration. The calcium-channel blocker nimodipine (60 mg orally every 4 hours) reduces the frequency of delayed ischemic deterioration and infarction by about 30%. Patients should receive isotonic fluid resuscitation (i.e., 1 mL/kg/hour of 0.9% saline) to maintain a euvolemic state guided by total fluid balance, a central venous pressure above 5 mm Hg, and other measures of volume status, such as inferior vena cava ultrasound and cardiac output monitoring.

Symptomatic vasospasm usually involves a decrease in the level of consciousness, hemiparesis, or both. Transcranial Doppler ultrasonography is widely used to diagnose vasospasm of the larger cerebral arteries after subarachnoid hemorrhage but has important limitations. CT angiography and perfusion are rapidly gaining acceptance as more useful tests to diagnose large-vessel spasm and reductions in tissue blood flow.

Treatment of acute symptomatic vasospasm relies on antispasmodic medication and on increasing blood volume, blood pressure, and cardiac output in an attempt to improve cerebral blood flow through arteries that have lost the capacity to autoregulate. Pressors such as norepinephrine (starting at 2.5 µg/kg/minute) and phenylephrine (starting at 10 µg/minute) should be titrated as needed to elevate systolic blood pressure to levels as high as 180 to 220 mm Hg. Short-term clinical improvement occurs in about 70% of patients. Cerebral angioplasty can lead to dramatic improvement in patients who have severe deficits that are refractory to hemodynamic augmentation.

Cerebral Edema

Brain edema after subarachnoid hemorrhage may be focal and related to a space-occupying hematoma, or it can be global, which is an ominous pattern that implies severe primary brain injury and a poor prognosis. Treatment, which should be guided by continuous ICP monitoring targeted at a goal of less than 20 mm Hg, incorporates CSF drainage, sedation, cerebral perfusion pressure optimization (target of 60 to 90 mm Hg), bolus osmotherapy with mannitol or hypertonic saline, hyperventilation, and, in the most severe cases, induced hypothermia (see Table 408-3). There is no evidence supporting the use of dexamethasone or other corticosteroids for the treatment of brain swelling after subarachnoid hemorrhage.

Hydrocephalus

Patients with subarachnoid hemorrhage may present acutely with obstructive hydrocephalus, which can lead to dangerous elevations in ICP and precipitate transtentorial herniation. Later in the disease course, the majority of patients with acute hydrocephalus transition to a treatable form of normal-pressure hydrocephalus. These patients have persistent psychomotor slowing, confusion, and gait instability that respond to permanent ventriculoperitoneal shunting.

Seizures

Generalized tonic-clonic seizures occur in about 10% of patients after subarachnoid hemorrhage: about 5% at onset and about 5% during the hospitalization. Prehospital seizures and focal pathologic changes on CT (i.e., subdural hematoma or cerebral infarction) are risk factors for in-hospital seizures. Antiepileptic therapy (see Table 408-3) is typically started at the time of diagnosis to minimize the risk of rebleeding caused by surges in arterial blood pressure and cerebral blood flow but can be safely discontinued in good-grade patients on postoperative day 1. Prophylactic antiepileptic therapy until discharge from the intensive care unit is a treatment option for comatose patients who remain at risk for nonconvulsive seizures, which occur in 15% of comatose patients monitored with continuous electroencephalography.

Medical Complications

Subarachnoid hemorrhage places patients at risk for a variety of common medical complications that occur as a consequence of homeostatic derangements. The most common are fever, anemia, hyperglycemia, and hyponatremia. The extent and severity of these derangements are independently correlated with poor outcome and should be actively managed according to an established protocol (see Table 408-3). Many poor-grade patients develop acute cardiopulmonary dysfunction due to massive sympathetic outflow at the time of bleeding. Electrocardiographic QT segment prolongation with T-wave inversion and minor elevations in troponin levels signal the possibility of cardiac injury. The most common important clinical manifestations are pulmonary edema (Chapter 59) and left ventricular neurogenic stunning (Chapters 59 and 107) that resolve during the first week. Treatment is supportive.

PREVENTION

Approximately 15% of patients who suffer a subarachnoid hemorrhage have two or more aneurysms. Secondary prevention of subarachnoid hemorrhage requires surgical or endovascular repair of any unruptured aneurysms because they are at high risk to bleed in the future. Other important measures to reduce the risk of aneurysm formation or bleeding from an unruptured aneurysm include blood pressure control (Chapter 67), cessation of cigarette smoking, and abstaining from alcohol.

Patients will sometimes present with an incidental and asymptomatic unruptured intracranial aneurysm discovered by neuroimaging. The risk-benefit tradeoff of treatment versus observation in these cases is complex and warrants referral to an experienced neurologist or neurosurgeon. Large size is the most important risk factor for subsequent aneurysm rupture, followed by symptoms related to expansion or compression, cigarette smoking, midline location, family history of subarachnoid hemorrhage, and postmenopausal status.[6] Small unruptured aneurysms of the internal carotid artery, by contrast, are at extremely low risk of bleeding and should be managed conservatively.

PROGNOSIS

Approximately 20% of patients with subarachnoid hemorrhage treated at high-volume centers do not survive to discharge. The most important determinant of outcome after subarachnoid hemorrhage is the patient's neurologic condition on arrival to the hospital. On the modified Hunt-Hess grading scale (see Table 408-2), patients who are classified as grade I or grade II have a

relatively good prognosis; grade III carries an intermediate prognosis, and grade IV and grade V have a poor prognosis. The severity of the clinical grade on admission generally correlates with the overall extent of bleeding and obstructive hydrocephalus. Other risk factors for mortality include advanced age, large aneurysm size, aneurysm rebleeding, cerebral infarction from vasospasm, and global cerebral edema.[7]

About 50% of survivors remain disabled by a neurocognitive syndrome that includes prominent memory loss, fatigue, inability to concentrate, depression, and anxiety. Cognitive and physical rehabilitation are essential for maximizing recovery in severely affected patients.

The rate of rupture among previously unruptured cerebral aneurysms is about 1% per year overall but ranges from 0.34% per year for aneurysms smaller than 5 mm to about 3% per year for aneurysms 10 to 24 mm and about 20% per year for aneurysms 25 mm or larger.[8] By comparison, patients who have an idiopathic perimesencephalic subarachnoid hemorrhage usually recover fully. For other less common causes, prognosis depends on the underlying condition. One-year survivors of subarachnoid hemorrhage have a two-fold higher rate of death as they age compared with the matched general population.[9]

INTRACEREBRAL HEMORRHAGE

Intracerebral hemorrhage is defined as acute spontaneous bleeding into the brain parenchyma. Primary intracerebral hemorrhage results from microscopic small-artery degeneration in the brain, caused by either chronic, poorly controlled hypertension (80% of cases) or amyloid angiopathy (20% of cases). Secondary intracerebral hemorrhage refers to intraparenchymal bleeding from a diagnosable anatomic vascular lesion or coagulopathy (Table 408-5).

EPIDEMIOLOGY

Intracerebral hemorrhage is responsible for 10 to 15% of all strokes in Western countries but up to 20 to 30% of strokes among Asian populations. The incidence of intracerebral hemorrhage in the United States is approximately 60,000 per year. By far the most important risk factor for intracerebral hemorrhage is hypertension, particularly when it is poorly controlled. The risk of intracerebral hemorrhage is about 40% higher in blacks than in whites. Worldwide, the incidence of intracerebral hemorrhage ranges from 10 to 40 per 1 million people, with the rate in Japan being at the top end of this range. Age-adjusted rates for men are about 50% higher than those for women. Other risk factors for intracerebral hemorrhage include heavy alcohol use, coagulopathy, and low serum cholesterol levels.

PATHOBIOLOGY

Primary intracerebral hemorrhage typically consists of a large, space-occupying confluent area of blood that has clotted within the brain parenchyma (Fig. 408-5). Abrupt arterial rupture leads to rapid accumulation of blood within the brain parenchyma, thereby causing increased local tissue pressure, physical distortion, and displacement of the brain. After the bleeding has stopped, the blood clots. Plasma that is rich in thrombin and other clotting factors then seeps into the surrounding brain tissue, where it triggers a cascade of secondary brain injury that evolves during days to weeks. This unique form of *neurohemoinflammation* causes local brain edema, programmed neuronal and glial apoptotic cell death, and breakdown of the brain-blood barrier.

The arterial disease that results in primary intracerebral hemorrhage is microscopic. Poorly controlled chronic hypertension (Chapter 67) causes a small-vessel vasculopathy characterized by fragmentation, degeneration, and the eventual rupture of penetrating arteries within the brain. The most commonly affected structures are the basal ganglia and thalamus (50%), followed by the lobar regions (33%) and the brain stem and cerebellum (17%) (Fig. 408-6). In 40% of cases, blood also ruptures into the ventricular system, thereby causing intraventricular hemorrhage.

Cerebral *amyloid angiopathy*, which is a distinctive cause of nonhypertensive lobar intracerebral hemorrhage in the elderly, is characterized by the deposition of β-amyloid protein in small to medium-sized blood vessels of the brain and leptomeninges. Amyloid can be demonstrated when microscopic examination of brain tissue demonstrates birefringence after application of Congo red stain. Amyloid angiopathy usually occurs as a sporadic disorder, and it is unrelated to systemic amyloidosis (Chapter 188). In addition to lobar intracerebral hemorrhage, patients may present with dementia, gait disturbance, complex partial seizures due to multiple microbleeds, or small multifocal white matter demyelinating lesions that are thought to

TABLE 408-5 CAUSES OF SECONDARY INTRACEREBRAL HEMORRHAGE

Trauma
Arteriovenous malformation
Intracranial aneurysm
Coagulopathy
Hemorrhagic conversion of cerebral infarct
Dural sinus thrombosis
Intracranial neoplasm
Cavernous angioma
Dural arteriovenous fistula
Venous angioma
Cocaine or sympathomimetic drug exposure
Central nervous system vasculitis

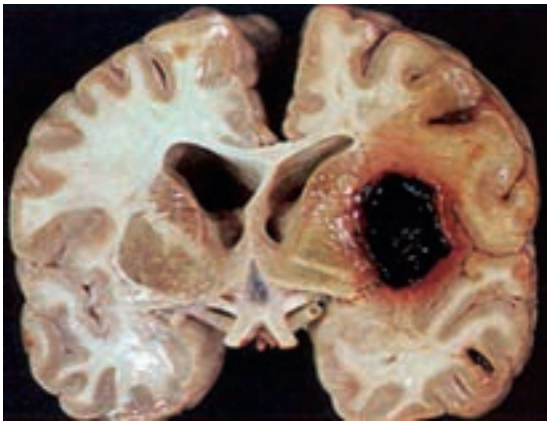

FIGURE 408-5. Pathology specimen showing a large basal ganglia parenchymal hemorrhage in the left hemisphere. (Courtesy Gregory W. Albers, Stanford University, Stanford, Calif.)

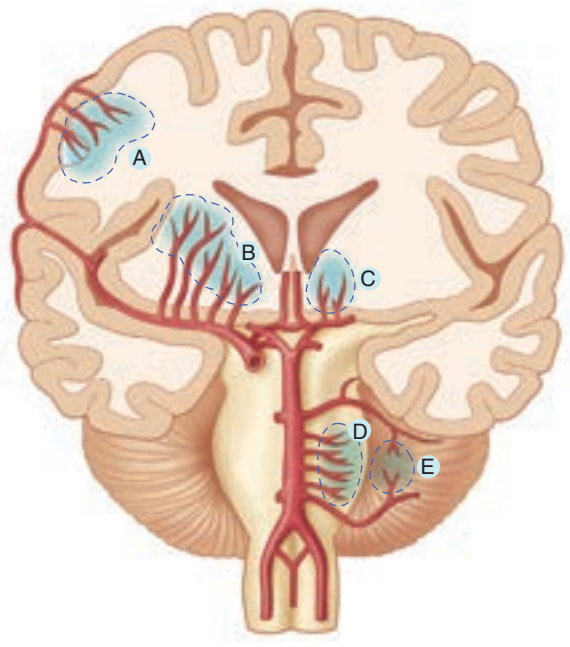

FIGURE 408-6. Typical sites and sources of intracerebral hemorrhage. Intracerebral hemorrhages most commonly involve the cerebral lobes and originate from penetrating cortical branches of the anterior, middle, or posterior cerebral arteries (*A*); the basal ganglia and originate from ascending lenticulostriate branches of the middle cerebral artery (*B*); the thalamus and originate from ascending thalamogeniculate branches of the posterior cerebral artery (*C*); the pons and originate from paramedian branches of the basilar artery (*D*); and the cerebellum and originate from penetrating branches of the posterior inferior, anterior inferior, and superior cerebellar arteries (*E*). (From Qureshi AI, Tuhrim S, Broderick JP, et al. Spontaneous intracerebral hemorrhage. *N Engl J Med*. 2001;344:1450-1460.)

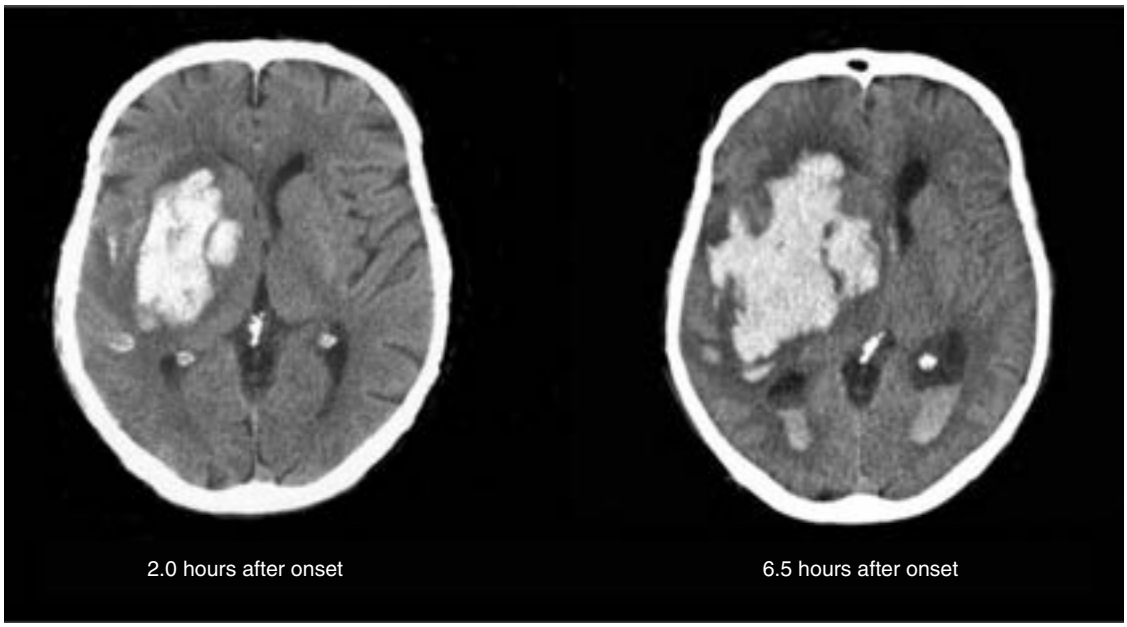

FIGURE 408-7. Early hematoma growth in a 48-year-old chronically hypertensive woman. *Left,* The baseline computed tomography scan shows a moderate-sized intracerebral hemorrhage in the right putamen. At this point, she is stuporous with a left hemiparesis. *Right,* A follow-up computed tomography scan performed after she deteriorated to coma with bilateral decerebrate posturing shows massive expansion of the hematoma as well as new intraventricular hemorrhage and obstructive hydrocephalus. Within 24 hours, she was declared brain dead. (From Mayer SA, Rincon F. Treatment of intracerebral haemorrhage. *Lancet Neurol.* 2005;4:662-672).

represent an autoimmune response to β-amyloid and that may engender an inflammatory response.

CLINICAL MANIFESTATIONS

Primary intracerebral hemorrhage usually is manifested as an acute focal neurologic deficit. It is clinically indistinguishable from ischemic stroke (Chapter 407), except that the onset and evolution of the deficit tend to be more violent. Unlike an aneurysmal subarachnoid hemorrhage, which often causes a dramatic surge in ICP with sudden loss of consciousness at its onset, intracerebral hemorrhage tends to produce progressive headache, vomiting, and a depressed level of consciousness during several hours. In fulminant cases, however, catastrophic bleeding can lead to a massive hematoma and brain death within 6 hours of onset.

The *putamen* is the site most frequently affected. When the expanding hematoma involves the adjacent internal capsule, patients develop dense contralateral hemiparesis, usually with hemianesthesia and hemianopia. Larger hemorrhages progressively affect the overlying cortex, thereby resulting in aphasia, hemispatial neglect, and contralateral gaze paresis. When the hemorrhage arises in the *thalamus*, hemianesthesia can initially precede the hemiparesis. The completed syndrome is usually characterized by a dense contralateral sensorimotor deficit that may be accompanied by a contralateral visual field deficit, impaired upward gaze, or both.

Lobar hemorrhages, which usually originate at the junctions between gray and white matter in the cerebral hemispheres, may result from either hypertension or amyloid angiopathy. The clinical manifestation depends on the location of the hemorrhage.

In 40% of cases, deep parenchymal cerebral bleeding ruptures into the ventricular system, thereby causing intraventricular hemorrhage. Blood in the third or fourth ventricle blocks the normal anterograde flow of CSF through the ventricular system, thereby resulting in acute hydrocephalus and intracranial hypertension. Left untreated, massive intraventricular hemorrhage results in rapid descent into coma with motor posturing and rostrocaudal loss of brain stem reflexes.

Pontine hemorrhage typically causes coma with quadriparesis, grossly disconjugate ocular motility disorders, and miotic pupils, although small hemorrhages may mimic syndromes of infarction. *Cerebellar hemorrhage* usually begins abruptly with vomiting and ataxia that usually is severe enough to prevent standing and walking. It is occasionally accompanied by dysarthria, adjacent cranial nerve (mostly sixth and seventh) dysfunction, and paralysis of conjugate ipsilateral gaze.

Seizures complicate the course of intracerebral hemorrhage in about 12% of patients. Although the risk is higher when the cortex is the primary site of bleeding, seizures can complicate deep intracerebral hemorrhage as well.

Expansion of the hematoma due to active, ongoing bleeding is an important cause of early neurologic deterioration after an intracerebral hemorrhage (Fig. 408-7). When the initial CT scan is obtained within 3 hours of the onset of symptoms, follow-up imaging reveals obvious enlargement of the hematoma in nearly 40% of patients, even in the absence of a coagulopathy. The likelihood of progression is even higher if the first CT scan is obtained earlier but minimal when the initial CT scan is obtained more than 6 hours after the onset of the bleed. Enlargement of the mass usually does not change the clinical picture until there is enough brain stem compression to precipitate coma, which can happen abruptly.

DIAGNOSIS

Intracerebral hemorrhage cannot be distinguished from ischemic stroke (Chapter 407) on the basis of clinical findings alone. Nonenhanced CT imaging of the brain is the method of choice for making the emergency diagnosis of intracerebral hemorrhage (see Fig. 408-7). CT readily demonstrates the size and location of the hematoma, any extension into the ventricular system, the degree of surrounding edema, and tissue displacement, such as a midline shift, due to mass effect. CT angiography may reveal secondary intracerebral hemorrhage due to an aneurysm or arteriovenous malformation or active extravasation of contrast material into the clot ("spot sign"), which implies an increased risk of early growth of the hematoma when it is identified soon after the onset of symptoms.

MRI techniques such as gradient-echo are highly sensitive for the diagnosis of intracerebral hemorrhage as well. The diagnosis of probable amyloid angiopathy is made clinically in patients with the appropriate clinical picture of lobar hemorrhages when gradient MRI reveals multiple cortical microbleeds. Conventional diagnostic cerebral angiography should be reserved for patients in whom secondary causes of intracerebral hemorrhage, such as aneurysms, arteriovenous malformations, cortical vein or dural sinus thrombosis, or vasculitis, are suspected.

TREATMENT

Rx

Treatment in an intensive care unit or stroke unit is strongly recommended for at least the first 24 hours after the onset of an intracerebral hemorrhage because the risk of neurologic deterioration is highest during this time. The most urgent treatment consideration is whether to proceed emergently with surgical evacuation or the placement of a ventricular drain.

Surgical Management

Although intracerebral hemorrhage has traditionally been considered a neurosurgical problem, randomized controlled trials have shown that

craniotomy and surgical evacuation of the hematoma within 24 hours do not improve outcome compared with initial medical management,[9] even with larger bleeds within 1 cm of the cortical surface.[9] However, these trials did not enroll patients if their physician thought that emergency surgery was a life-saving intervention, and many experts still believe that urgent craniotomy can improve the outcome of younger patients with large lobar hemorrhages and a deteriorating course due to mass effect.

In contrast to supratentorial intracerebral hemorrhage, it is widely accepted that patients with cerebellar hemorrhages exceeding 3 cm in diameter benefit from emergent surgical evacuation. Because abrupt and dramatic deterioration to coma can occur within the first 24 hours of the onset of symptoms in these patients, surgery should be performed promptly before further clinical deterioration occurs.

Ventricular Drainage

External ventricular drainage is indicated in all stuporous or comatose patients with intraventricular hemorrhage and ventricular enlargement in whom aggressive support is indicated. This life-saving procedure, which can be performed at the bedside, decompresses the intracranial vault and arrests the process of downward brain stem herniation by allowing drainage of bloody CSF into a drainage receptacle. Connecting the drainage system to a pressure transducer also allows measurement of ICP.

Reverse of Anticoagulation

Fifteen percent of intracerebral hemorrhages are associated with the use of oral anticoagulants, and these patients face a high risk of progressive bleeding during a prolonged time window of many hours. For warfarin-associated intracerebral hemorrhage, failure to correct the INR promptly to below 1.4 further increases the risk of progressive bleeding and is associated with increased mortality. Patients who are taking warfarin and who present with an intracerebral hemorrhage should be treated immediately with a four-factor prothrombin complex concentrate and intravenous vitamin K (Table 408-6), a regimen that results in a substantially faster correction of the INR in a greater proportion of patients.[9] A single dose of recombinant activated factor VIIa (3 to 6 mg by intravenous push) normalizes the INR within minutes, promotes hemostasis, and is an attractive option for expediting life-saving neurosurgical intervention when minutes count, but at the cost of a 5% risk of a thromboembolic complication such as myocardial infarction or stroke. Patients with intracerebral hemorrhage anticoagulated with unfractionated or low-molecular weight heparin should be reversed with protamine sulfate (Chapters 38 and 174). Patients with thrombocytopenia or platelet dysfunction can be treated with a single dose of desmopressin, platelet transfusions, or both, but evidence for efficacy is lacking. Treatment options for intracerebral hemorrhage associated with novel oral anticoagulants (e.g., rivaroxaban, dabigatran, apixaban, and edoxaban) are extremely limited (Chapter 38). If anticoagulation must be restarted after recovery, the risk of recurrent intracranial hemorrhage is about 2.5 per 100 patient-years.[10]

TABLE 408-6 MEDICAL MANAGEMENT PROTOCOL FOR ACUTE INTRACEREBRAL HEMORRHAGE

Blood pressure	• Maintain mean arterial pressure <140 mm Hg with continuous infusion of labetalol (2-10 mg/min) or nicardipine (5-15 mg/hr). • If stuporous or comatose, measure ICP and maintain CPP >70 mm Hg.
Reversal of anticoagulation	• For elevated INR: vitamin K 10 mg IVP *and* 4F-PCC INR 2 to <4: 25 units/kg; not to exceed 2500 units INR 4-6: 35 units/kg; not to exceed 3500 units INR >6: 50 units/kg; not to exceed 5000 units • For heparin: protamine sulfate 10 to 50 mg slow IV push (1 mg reverses approximately 100 units of heparin) • For thrombocytopenia or platelet dysfunction: desmopressin 0.3 µg/kg IVP and/or transfuse 6 units of platelets • Expedited INR reversal for life-saving neurosurgical intervention: recombinant activated factor VIIa 40-80 µg/kg (approximately 3.0-6.0 mg) IVP
Intracranial hypertension	• Elevate head of bed to 30 degrees • Mannitol 1.0-1.5 g IV • Hyperventilate to P_{CO_2} of 30 mm Hg
Fluids and nutrition	• Normal (0.9%) saline at 1.0 mL/kg/hr • Begin enteral feeding through nasoduodenal tube within 24 hours
Seizure prophylaxis	• For coma with intracranial hypertension or acute seizures: fosphenytoin or phenytoin IV load (15-20 mg/kg); 300 mg IV daily for 7 days
Physiologic homeostasis	• Cooling blankets to maintain temperature ≤37.5° C • Insulin drip to maintain glucose 120-180 mg/dL

4F-PCC = four-factor prothrombin complex concentrate containing factors II, VII, IX, and X; CPP = cerebral perfusion pressure; ICP = intracranial pressure; INR = international normalized ratio; IVP = intravenous push.

Blood Pressure Control

Acute intracerebral hemorrhage often leads to extreme arterial hypertension. Aggressive blood pressure reductions in the setting of impaired autoregulation can result in exacerbation of ischemic injury, whereas lack of blood pressure control can theoretically exacerbate the early growth of the hematoma and the risk of vasogenic edema. Although current guidelines[11] recommend a systolic blood pressure target of less than 180 mm Hg and a mean blood pressure target of 160 mm Hg,[9] lowering of systolic blood pressure to less than 140 mm Hg within 6 hours of onset of symptoms does not reduce mortality and provides only borderline improvement in the extent of disability among survivors compared with a goal of less than 180 mm Hg.[9] Given the need to precisely control blood pressure levels in the setting of impaired autoregulation, use of fast-acting continuous infusion agents with intra-arterial monitoring is recommended. Agents of choice are labetalol, a combined α- and β-blocker, and nicardipine, which is a calcium-channel blocker (see Table 67-14). Sodium nitroprusside should be avoided because of its lack of a reliable dose-response effect and its capacity to directly increase ICP.

Cerebral Edema

Brain swelling can progress for many days after the onset of intracerebral hemorrhage, but it most often causes neurologic deterioration within the first 72 hours in patients with hemorrhages exceeding 30 mL in volume. Management of cerebral edema should be guided by measurement of the ICP with an external ventricular drain or parenchymal monitor, with efforts directed at maintaining ICP below 20 mm Hg and cerebral perfusion pressure above 70 mm Hg. As is the case with subarachnoid hemorrhage, therapy should be directed by an approach that includes sedation, blood pressure optimization, bolus osmotherapy, controlled hyperventilation, mild hypothermia, and, as a last resort, salvage hemicraniectomy in selected younger patients (Chapter 399). Dexamethasone and other corticosteroids do not effectively treat intracerebral hemorrhage–related brain edema and are contraindicated. Hypotonic intravenous fluids (i.e., 0.45% saline or 5% dextrose in water) should be strictly avoided because the free water in these solutions can aggravate brain edema.

Medical and Neurologic Complications

To combat malnutrition and muscle wasting, early enteral feeding (Chapter 216) should be initiated through a nasoduodenal feeding tube in patients who lack the capacity to swallow. Body temperature and the blood glucose level should be maintained in the normal to slightly elevated range with surface cooling and continuous insulin infusion. About 12% of patients with intracerebral hemorrhage experience convulsive seizures during their hospitalization, and risk is increased with lobar location. Prophylactic anticonvulsant therapy with phenytoin (20 mg/kg intravenously) or a similar agent is reasonable for high-risk stuporous or comatose patients. If seizures have not occurred, anticonvulsants should be discontinued at discharge because they can hamper neurologic recovery during rehabilitation. Even with anticonvulsant therapy, continuous electroencephalographic monitoring reveals electrographic seizure activity in 20% of comatose patients. It is unclear whether midazolam infusion or other aggressive measures to eliminate these seizures (Chapter 403) can improve outcome.

PREVENTION

Blood pressure reduction (Chapter 67), which significantly decreases the risk of intracerebral hemorrhage and other forms of stroke, is by far the most effective method for preventing recurrent intracerebral hemorrhage. Angiotensin-converting enzyme inhibitors are particularly effective (see Table 67-7). Antiplatelet agents and anticoagulants of all types should be meticulously avoided in patients with multiple lobar microbleeds caused by amyloid angiopathy.

PROGNOSIS

Factors that consistently predict death or functional disability at 30 days include a large volume of intracerebral hemorrhage, depressed level of consciousness, intraventricular hemorrhage, infratentorial location, and older age. The volume of the hematoma can be easily calculated from CT scan images by use of the "ABC ÷ 2 method," which involves multiplying the

diameter of the hematoma in three dimensions and dividing by two. A simple clinical grading scale (E-Table 408-1) incorporates these variables and can give a reliable prediction of mortality risk at 30 days.

Except in the most severe cases, however, caution is warranted in communicating a hopeless prognosis before aggressive efforts have been made to resuscitate victims of intracerebral hemorrhage. Physicians tend to underestimate the chances of a good outcome, and many poor outcomes result from self-fulfilling prophesies of doom. Mortality after intracerebral hemorrhage is lower among patients in specialty neurologic intensive care units, presumably because of adherence to best medical practices, early transition to rehabilitation, and cautious optimism when setbacks occur.

BRAIN VASCULAR MALFORMATIONS

Brain vascular malformations are space-occupying congenital anomalies that can often exist for a lifetime without symptoms. The most feared and dangerous complication is rupture, which can be manifested as intracerebral hemorrhage, as intraventricular hemorrhage, or less often as subarachnoid hemorrhage.

EPIDEMIOLOGY

About 10% of intracerebral hemorrhages but only about 1% of strokes are caused by vascular malformations. The prevalence of an arteriovenous malformation is about 0.5%, and the annual incidence of hemorrhage is between 1 and 3 cases per 100,000 people. Hemorrhage from an arteriovenous malformation is most common during the second through fourth decades. The risk of rebleeding is about 7% acutely. For the next 5 years, the risk of bleeding is about 2% per year, and it then falls to about 1 to 2% annually thereafter. Over a lifetime, a young person therefore has a 50 to 60% probability of another hemorrhage, each of which carries a 10 to 15% risk of acute death. Unlike with some causes of cerebral hemorrhage, preexisting hypertension does not seem to be a risk factor.

PATHOBIOLOGY

Cerebrovascular malformations are characterized on the basis of their histologic appearance and the intervening neural parenchyma. The most frequent type of vascular malformation is an arteriovenous malformation, in which a core or nidus of dysplastic vessels is fed by arteries and drained by veins without intervening capillaries. The result is a low-resistance, high-flow shunt that leads to progressive arterial dilation and venous wall thickening. The nidus usually does not contain any intervening neural tissue. Bleeding from a feeding artery aneurysm usually results in subarachnoid hemorrhage, bleeding from the nidus itself usually results in intracerebral hemorrhage, and bleeding from a draining vein usually is manifested as intraventricular hemorrhage.

The next most common vascular malformations are cavernous angiomas or hemangiomas. These malformations, which also do not contain neural tissue, are composed of small-caliber sinusoidal vascular channels that are commonly thrombosed.

Dural arteriovenous fistulas are typically acquired lesions that result from the formation of small arteriovenous shunts in the wall of a cavernous sinus as a consequence of dural sinus thrombosis. Over time, flow through the fistula increases, leading to pulsatile expansion of regional veins and subsequent rupture. Rare familial cases have been described.

CLINICAL MANIFESTATIONS

About 50% of arteriovenous malformations are manifested with intracranial hemorrhage, about 30% initially are manifested as seizures, and about 20% may be manifested with progressive neurologic disability. An increasing proportion, however, are now detected by brain imaging as part of the evaluation of headaches (Chapter 398), to which arteriovenous malformations may or may not be causally related.

Because an arteriovenous malformation can bleed into the subarachnoid space, the brain parenchyma, or the ventricular system, symptoms and signs depend on the location and severity of bleed. Post-bleeding cerebral vasospasm, which is less common than aneurysmal bleeding, occurs in less than 5% of cases and is typically linked to thick cisternal clot or extensive intraventricular hemorrhage.

Patients who develop seizures as a result of these arteriovenous malformations often have focal seizures (Chapter 403). Even without seizures, patients can develop focal neurologic deficits due to vascular thrombosis or the shunting of blood through the malformation rather than allowing it to perfuse normal brain tissue.

DIAGNOSIS

A noncontrast CT scan may show bleeding, sometimes in a location that is unusual for a primary intracerebral hemorrhage or a ruptured aneurysm. Contrast-enhanced CT may show marked enhancement of the feeding arteries and draining veins. Another option is MRI with signal void on T1- or T2-weighted images.[12] However, angiography is the definitive test to identify an arteriovenous malformation and to delineate its size, gross morphology, feeding arteries, and draining veins. Even if an arteriovenous malformation is found by unilateral carotid injection, four-vessel angiography is indicated because malformations can be multiple and can be associated with saccular aneurysms.

When cavernous angiomas or hemangiomas hemorrhage, they tend to produce minor focal syndromes that appear on MRI as a classic target lesion that results from multiple previous minor bleeding events. The low flow rate through these lesions makes them difficult to detect by angiography.

TREATMENT

Rx

In a patient who survives the initial hemorrhage, the two therapeutic goals are to avoid neurologic deterioration and to remove the arteriovenous malformation completely. General medical treatment measures for intracranial hemorrhage related to arteriovenous malformation are the same as for intracerebral hemorrhage (see Table 408-6). Removal of the arteriovenous malformation can be curative, but surgery is challenging for malformations in critical neurologic areas. Options include selective embolization of the feeding arteries, surgical resection, and radiation-induced thrombosis, alone or sometimes in combination. Selective embolization can reduce the size of the malformation and blood flow through it but rarely can obliterate it completely. Stereotactic radiosurgery is used only for small lesions, and its therapeutic effect depends on the gradual shrinkage of abnormal vessels after the procedure.

Microsurgical removal of an arteriovenous malformation is often performed in stages until a postoperative angiogram shows no residual malformation. However, recanalization and recurrent hemorrhage can occur, and long-term success rates are unknown.

PREVENTION AND PROGNOSIS

The prognosis of an unruptured arteriovenous malformation varies according to its location, size, and morphology. In a randomized trial, medical management emphasizing control of hypertension, avoidance of anticoagulants, and use of anticonvulsants to control seizures was superior to multimodality intervention with surgery, embolization, or radiotherapy, with a 10% rate of death or stroke at 33 months compared with a 30% risk in the intervention group. Until further data are available, routine interventional treatment of unruptured arteriovenous malformations is not justified.[13]

Grade A References

A1. Baharoglu MI, Germans MR, Rinkel GJ, et al. Antifibrinolytic therapy for aneurysmal subarachnoid haemorrhage. *Cochrane Database Syst Rev*. 2013;8:CD001245.

A2. Gaberel T, Magheru C, Emery E, et al. Antifibrinolytic therapy in the management of aneurismal subarachnoid hemorrhage revisited. A meta-analysis. *Acta Neurochir (Wien)*. 2012;154:1-9.

A3. Molyneux AJ, Kerr RS, Yu LM, et al. International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised comparison of effects on survival, dependency, seizures, rebleeding, subgroups, and aneurysm occlusion. *Lancet*. 2005;366:809-817.

A4. Feigin VL, Rinkel GJ, Algra A, et al. Calcium antagonists in patients with aneurysmal subarachnoid hemorrhage: a systematic review. *Neurology*. 1998;50:876-883.

A5. Mendelow AD, Gregson BA, Fernandes HM, et al. Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the International Surgical Trial in Intracerebral Haemorrhage (STICH): a randomised trial. *Lancet*. 2005;365:387-397.

A6. Mendelow AD, Gregson BA, Rowan EN, et al. Early surgery versus initial conservative treatment in patients with spontaneous supratentorial lobar intracerebral haematomas (STICH II): a randomised trial. *Lancet*. 2013;382:397-408.

A7. Sarode R, Milling TJ Jr, Refaai MA, et al. Efficacy and safety of a 4-factor prothrombin complex concentrate in patients on vitamin K antagonists presenting with major bleeding: a randomized, plasma-controlled, phase IIIb study. *Circulation*. 2013;128:1234-1243.

A8. Tsivgoulis G, Katsanos AH, Butcher KS, et al. Intensive blood pressure reduction in acute intracerebral hemorrhage: a meta-analysis. *Neurology*. 2014;83:1523-1529.

A9. Anderson CS, Heeley E, Huang Y, et al. Rapid blood-pressure lowering in patients with acute intracerebral hemorrhage. *N Engl J Med*. 2013;368:2355-2365.

A10. Mohr JP, Parides MK, Stapf C, et al. Medical management with or without interventional therapy for unruptured brain arteriovenous malformations (ARUBA): a multicentre, non-blinded, randomised trial. *Lancet*. 2014;383:614-621.

E-TABLE 408-1 THE ICH SCORE

COMPONENT	SCORE POINTS
GCS Score	
3-4	2
5-12	1
13-15	0
ICH Volume, cm^3	
≥30	1
<30	0
IVH	
Yes	1
No	0
Infratentorial ICH	
Yes	1
No	0
Age, y	
≥80	1
<80	0
Estimated 30-day Mortality:	
TOTAL SCORE	**MORTALITY**
0	0%
1	13%
2	26%
3	72%
4	97%
5+	100%

GCS = Glasgow coma score; ICH = intracerebral hemorrhage
From: Hemphill JC, 3rd, Bonovich DC, Besmertis L, et al. The ICH score: a simple, reliable grading scale for intracerebral hemorrhage. *Stroke.* 2001;32:891-897.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Garcia-Rodriguez LA, Gaist D, Morton J, et al. Antithrombotic drugs and risk of hemorrhagic stroke in the general population. *Neurology*. 2013;81:566-574.
2. Runchey S, McGee S. Does this patient have a hemorrhagic stroke? Clinical findings distinguishing hemorrhagic stroke from ischemic stroke. *JAMA*. 2010;303:2280-2286.
3. Connolly ES Jr, Rabinstein AA, Carhuapoma JR, et al. Guidelines for the management of aneurysmal subarachnoid hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2012;43:1711-1737.
4. Menke J, Larsen J, Kallenberg K. Diagnosing cerebral aneurysms by computed tomographic angiography: meta-analysis. *Ann Neurol*. 2011;69:646-654.
5. Dusick JR, Gonzalez NR. Management of arterial vasospasm following aneurysmal subarachnoid hemorrhage. *Semin Neurol*. 2013;33:488-497.
6. Backes D, Vergouwen MD, Velthuis BK, et al. Difference in aneurysm characteristics between ruptured and unruptured aneurysms in patients with multiple intracranial aneurysms. *Stroke*. 2014; 45:1299-1303.
7. Chen G, Arima H, Wu G, et al. Subarachnoid extension of intracerebral hemorrhage and 90-day outcomes in INTERACT2. *Stroke*. 2014;45:258-260.
8. Morita A, Kirino T, Hashi K, et al. The natural course of unruptured cerebral aneurysms in a Japanese cohort. *N Engl J Med*. 2012;366:2474-2482.
9. Korja M, Silventoinen K, Laatikainen T, et al. Cause-specific mortality of 1-year survivors of subarachnoid hemorrhage. *Neurology*. 2013;80:481-486.
10. Poli D, Antonucci E, Dentali F, et al. Recurrence of ICH after resumption of anticoagulation with VK antagonists: CHIRONE study. *Neurology*. 2014;82:1020-1026.
11. Morgenstern LB, Hemphill JC 3rd, Anderson C, et al. Guidelines for the management of spontaneous intracerebral hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2010;41:2108-2129.
12. Asif K, Leschke J, Lazzaro MA. Cerebral arteriovenous malformation diagnosis and management. *Semin Neurol*. 2013;33:468-475.
13. Al-Shahi Salman R, White PM, Counsell CE, et al. Outcome after conservative management or intervention for unruptured brain arteriovenous malformations. *JAMA*. 2014;311:1661-1669.

409

PARKINSONISM

ANTHONY E. LANG

Parkinsonism is a clinical syndrome that consists of four cardinal signs: tremor, rigidity, akinesia, and postural disturbances (TRAP). Parkinson disease is a common cause of the TRAP syndrome, but there are numerous other causes (Table 409-1).

TABLE 409-1 DIFFERENTIAL DIAGNOSIS OF PARKINSONISM

PARKINSON DISEASE
Sporadic
Genetic
Autosomal dominant (e.g., α-synuclein gene mutations, duplications, triplications; *LRRK2* mutations)
Autosomal recessive (e.g., *parkin, DJ1, PINK1*)
SECONDARY PARKINSONISM
Neurodegenerative diseases (sporadic or genetic)
Progressive supranuclear palsy* (Videos 409-3 through 409-6)
Multiple system atrophy* (Videos 409-7 through 409-9)
Corticobasal degeneration* (Videos 409-10 and 409-11)
Dementia with Lewy bodies*
Alzheimer disease*
ALS-parkinsonism-dementia complex of Guam
Huntington disease
Rapid-onset dystonia-parkinsonism
Pallidopyramidal degeneration (including PARK9 and PARK15)
Neuroacanthocytosis
Spinocerebellar ataxias (e.g., SCA-3, SCA-2)
Wilson disease
Pantothenate kinase–associated neurodegeneration (Hallervorden-Spatz syndrome)
Neuroferritinopathy
Calcification of the basal ganglia (Fahr disease)
Dopa-responsive dystonia (not a degenerative disorder)
Drugs*
Neuroleptics, metoclopramide, prochlorperazine, tetrabenazine, reserpine, cinnarizine, flunarizine, α-methyldopa, lithium
Toxic
MPTP, manganese (including illicit use of ephedrone), carbon monoxide, mercury
Infectious
Encephalitis lethargica
Other encephalitis, including HIV associated
Subacute sclerosing panencephalitis
Creutzfeldt-Jakob disease
Vascular*
Atherosclerosis
Amyloid angiopathy
Neoplastic
Brain tumor
Other mass lesions
Normal-pressure hydrocephalus*
Head trauma
Multiple sclerosis

*See Table 409-4 for additional details.
ALS = amyotrophic lateral sclerosis; HIV = human immunodeficiency virus; MPTP = 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine.
Modified from Cloutier M, Lang AE. Movement disorders: an overview. In: Factor SA, Lang AE, Weiner WJ, eds. *Drug Induced Movement Disorders*. Malden, Mass: Blackwell; 2005:3-19.

PARKINSON DISEASE

EPIDEMIOLOGY

Parkinson disease, which is the second most common neurodegenerative disorder after Alzheimer disease, occurs in approximately 1 in 1000 in the general population and in 1% of persons older than 65 years. Men are affected slightly more often than women (3 : 2).

PATHOBIOLOGY

The cause of Parkinson disease is believed to be a variable combination of poorly understood genetic[1] and environmental factors.[2] Both autosomal dominant and recessive genes can cause classic Parkinson disease. The protein α-synuclein, which is the chief constituent of the hallmark cytoplasmic inclusion, the Lewy body (Chapter 402), is critical in the pathogenesis of Parkinson disease. Abnormal aggregation of the protein, either from mutations in the α-synuclein gene or as a result of excessive production of the normal protein because of gene duplications or triplications, is associated with varying disease phenotypes. Other defined genetic abnormalities may be associated with classic later-onset Parkinson disease, including *LRRK2*, which is currently the most common cause of autosomal dominantly inherited Parkinson disease, or with early-onset parkinsonism, typically found in the autosomal recessive forms associated with *parkin*, *DJ1*, and *PINK1*. Other genes in which mutations may increase the risk for development of Parkinson disease include the glucocerebrosidase gene (*GBA*).

Strong support for the "environmental hypothesis" of sporadic Parkinson disease relates to the observation that the selective neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) causes acute parkinsonism due to loss of dopamine neurons in the substantia nigra pars compacta (SNc). MPTP is oxidized to the active toxin MPP^+, which is a selective inhibitor of complex I of the mitochondrial electron transport chain. This knowledge, combined with recognition of the importance of dopamine (see later), has implicated oxidative stress in the pathogenesis of Parkinson disease. Other proposed pathogenetic factors include mitochondrial dysfunction, protein misfolding or aggregation, excitotoxicity, inflammation, apoptotic cell death, and loss of trophic support.

Pathology

Many of the features of Parkinson disease are due to loss of dopamine in the neostriatum (especially the putamen) secondary to loss of pigmented dopaminergic neurons in the SNc of the midbrain. Approximately 60% of these dopaminergic neurons will have degenerated before clinical features of the disease develop.[3]

In addition to the prominent degenerative changes in the SNc (cell loss, gliosis, abnormal deposition of aggregated α-synuclein as Lewy bodies and Lewy neurites), pathologic changes are also evident in other brain stem nuclei, in cortical regions, and in peripheral autonomic neurons. Indeed, it has been suggested that Parkinson disease may begin in the lower brain stem and the olfactory system, where it causes early loss of the sense of smell and only later involves the substantia nigra. Independent of the order of involvement, it is likely that the widespread extranigral neurodegenerative changes account for the many symptoms that do not respond to dopamine replacement and that become increasingly problematic as the disease progresses. How the disease progresses and spreads in the nervous system is unknown. Studies suggest the possibility of cell-to-cell transmission of a form of α-synuclein that may then induce abnormal folding and aggregation of the normal protein in a "permissive templating" fashion similar to prion diseases (Chapter 415).[4]

CLINICAL MANIFESTATIONS

Typically, the symptoms begin in one limb. This asymmetry often persists into later stages of the disease.

Motor Symptoms

Tremor

The classic "resting tremor" of Parkinson disease has characteristic clinical features.[5] The tremor has a frequency of 4 to 6 cycles per second, typically with a "pill-rolling" character when it involves the hand. It is generally present with the limb in complete repose and typically subsides when the limb moves and takes up a new position, although the tremor may reemerge ("reemergent tremor") within a short time after maintaining the new position (Video 409-1). Because resting tremor diminishes or subsides with action, it may not be disabling but can be embarrassing and may be associated with aching or

fatigue of the affected limb. Resting tremor is usually accentuated by stress (e.g., by asking the patient to perform mental calculations). It is also characteristically present in the upper limbs while walking. A higher-frequency (e.g., 7 to 10 Hz) postural and kinetic tremor is also common in patients with various causes of parkinsonism.

Rigidity

Rigidity is a form of increased muscle tone appreciated best on slow passive movements. It may be characterized as "cogwheel" when a tremor is superimposed or as "lead pipe" when it is not. Rigidity is "activated" or accentuated on examination by asking the patient to move the limb opposite the one being tested. Patients may complain of stiffness, but the rigidity is not usually disabling.

Akinesia

Akinesia or bradykinesia comprises a variety of disturbances in movement, including slowness, reduced amplitude, fatiguing, and interruptions in ongoing movement. This disabling aspect of parkinsonism interferes with all voluntary activities and accounts for many of the well-known features of parkinsonism: lack of facial expression with reduced blinking (hypomimia or masked facies—the "reptilian stare"), soft monotonous speech (hypophonia), impaired swallowing resulting in drooling (sialorrhea), small handwriting (micrographia), reduced arm swing while walking, shortened stride and shuffling gait, difficulty arising from a low chair, and problems turning over in bed. Arrest in ongoing movement ("motor block") can interfere with a variety of activities, but it is best appreciated as freezing of gait (Video 409-2). Bradykinesia is evident on inspection and elicited by testing rapid repetitive and alternating movements: finger tapping, opening and closing the fist, pronating and supinating the wrist, and toe and heel tapping.

Postural Disturbances

Postural disturbances include a flexed posture in the limbs and trunk (stooped, simian posture) as well as postural instability resulting in imbalance and falls. Patients may complain of being unable to stop themselves from going forward (propulsion) or backward (retropulsion). Clinical assessment of postural instability includes the "pull test," in which the examiner abruptly pulls the patient off balance while being ready to catch the patient in the event of a fall.

Other Symptoms

In addition to the motor features of parkinsonism, a variety of non–motor-related features are extremely common. These include pain and other sensory disturbances; dysautonomic complaints, such as urinary urgency and frequency; orthostatic faintness; constipation; male erectile dysfunction; sleep abnormalities, including rapid eye movement behavioral disorder (Chapter 405); anxiety; fatigue; depression; and cognitive disturbances, including dementia.[6] As the disease progresses, more resistant features develop, including "axial" motor disturbances (speech and swallowing abnormalities, freezing, and postural instability) as well as neurobehavioral and cognitive dysfunction.

Complications

In addition to the manifestations of the disease itself, complications of drug therapy include motor- and non–motor-related fluctuations and psychiatric or behavioral disturbances.[7] Thus, in the later stages of the disease, the clinical picture often fluctuates from hour to hour and even from minute to minute. Accordingly, patients exhibit a mixture of the classic features of parkinsonism, which may improve considerably in response to medication; symptoms that persist despite the peak benefit of medication; and symptoms that occur as a complication of dopaminergic medication (Table 409-2).

TABLE 409-2 PROBLEMS IN LATE-STAGE PARKINSON DISEASE

PROBLEM	SYMPTOMS
LATER TREATMENT-RESISTANT SYMPTOMS	
Motor	Dysarthria Dysphagia Freezing of gait (on-period freezing) Postural instability with falls
Non–motor	Dysautonomia, weight loss Sensory symptoms, including pain (some may be responsive to levodopa) Changes in mood or behavior (depression, anxiety), sleep disturbances (excessive daytime sleepiness often caused by or aggravated by dopaminergic medication) Rapid eye movement sleep behavior disorder (may develop before parkinsonism) Fatigue Cognitive dysfunction and dementia
RELATED TO TREATMENT AND DISEASE	
Motor fluctuations	Wearing off of drug effect (predictable end-of-dose deterioration, morning akinesia), increased latency to benefit ("delayed-on"), dose failures ("no-on") On-off phenomenon, more rapid and unpredictable fluctuations Concomitant fluctuations of non–motor-related symptoms ("non-motor fluctuations") that may be as disabling as motor symptoms (or more so)
Dyskinesias (abnormal involuntary movements)	Peak-dose dyskinesias: chorea, athetosis, and, less often, more prolonged dystonia, typically worse on the initially affected side (Video 409-12) Diphasic dyskinesia ("beginning-of-dose" and "end-of-dose" dyskinesias): mixtures of choreoathetosis, ballism, dystonia, alternating movements (especially in the legs) Off-period dystonia: most often involving the legs and feet (including morning foot dystonia)
Psychiatric disturbances	Vivid dreams and nightmares Visual hallucinations with a clear sensorium Hallucinations with confusion Mania, impulse control disorders (e.g., hypersexuality, problem gambling), dopaminergic drug addiction Paranoid psychosis

Modified from Lang AE, Lozano AM. Parkinson's disease—second of two parts. *N Engl J Med.* 1998;339:1130-1143.

TABLE 409-3 CLINICAL CLUES TO AN ALTERNATIVE (NON–PARKINSON DISEASE) CAUSE OF PARKINSONIAN SIGNS AND SYMPTOMS

Extraocular movements—e.g., nystagmus, limitation of vertical gaze, especially with slowing of downward saccadic eye movements (Video 409-4)
Early and prominent dysarthria or dysphagia
Prominent or early abnormal neck postures: flexion or extension (Video 409-8)
Ataxia—limb, gait (impaired tandem gait)
Lower body distribution with relative sparing of upper limb function
Early postural instability, falls, or freezing (Video 409-3)
Dysautonomia (early and prominent), prominent hypotensive response to dopaminergic medication
Pyramidal tract signs—very brisk reflexes, clonus, extensor plantar responses
Peripheral nerve dysfunction—loss of reflexes, distal sensory loss, weakness
Apraxia and cortical sensory changes
Early severe dementia
Poor response to levodopa

DIAGNOSIS

Testing for monogenetic forms of Parkinson disease (e.g., *parkin*) is becoming available, but guidelines for its use have not yet been developed. Given the classic clinical manifestations, the diagnostic evaluation focuses largely on ways to exclude other causes of parkinsonism (Table 409-3).[8] Young-onset patients should have Wilson disease excluded by determination of 24-hour urine copper and serum ceruloplasmin and by slit lamp examination (Chapter 211). Findings on magnetic resonance imaging are generally normal in Parkinson disease, but it is indicated to exclude other diagnoses (Table 409-4).[9] Positron emission tomography, which can assess the presynaptic and postsynaptic sides of the nigrostriatal dopamine system, is useful for research, but the most common ligand, [^{18}F]fluorodopa, does not reliably distinguish Parkinson disease from many other neurodegenerative diseases that mimic it. The same limitations apply to evaluation of the dopamine transporter by single-photon emission computed tomography, which is available for clinical use. The finding of increased echogenicity in the SNc of the midbrain on transcranial ultrasound may be more specific in diagnosing Parkinson disease, but the data are not conclusive.

TABLE 409-4 DISEASES THAT MUST BE DISTINGUISHED FROM PARKINSON DISEASE

DIAGNOSIS	IMPORTANT DISTINGUISHING CLINICAL FEATURES	RESPONSE TO LEVODOPA/COMMENTS (INCLUDING IMAGING)
Multiple system atrophy (MSA) (includes older terms: striatonigral degeneration, sporadic olivopontocerebellar atrophy, and Shy-Drager syndrome) (a "synucleinopathy")	Early dysautonomia (including orthostatic hypotension and sexual impotence) and bladder dysfunction (with autonomic and nonautonomic components) Cerebellar dysfunction	Good response initially evident in 20% and sustained partial response in ≈15% Dyskinesias or motor fluctuations possible; cranial dystonia may be prominent (Video 409-7) Patient is wheelchair bound despite response to levodopa (early loss of postural reflexes, with or without ataxia)
MSA-P, a predominant parkinsonian manifestation	Pyramidal tract signs Stimulus-sensitive myoclonus of the hands and face	MRI (including diffusion-weighted imaging and gradient-echo sequences) often shows diagnostic changes in the striatum in MSA-P and "hot cross bun sign" in the pons and hyperintensity in middle cerebellar peduncles in MSA-C
MSA-C, a predominant cerebellar manifestation (mixed features are common)	Extreme forward neck flexion (anterocollis) Mottled, cold hands Inspiratory stridor (Video 409-9) Prominent dysarthria	
Progressive supranuclear palsy (a "tauopathy")	Supranuclear vertical ophthalmoplegia (Video 409-4) Other oculomotor and eyelid disturbances (Video 409-6) Axial rigidity greater than limb rigidity Early falls, speech and swallowing disturbances Nuchal extension Cognitive or behavioral changes Progressive nonfluent aphasia Possibly a higher incidence of hypertension than in Parkinson disease and other neurodegenerative causes of parkinsonism	Good response rarely evident; benefit only for classic parkinsonian features, such as limb rigidity, classic bradykinesia with fatiguing of amplitude of repetitive movements, and rare examples of tremor at rest MRI often demonstrates profound midbrain atrophy ("hummingbird sign" on a midline sagittal view, "morning glory sign" on axial view)
Corticobasal (cortical-basal ganglionic) degeneration (a "tauopathy")	Apraxia, cortical sensory loss, alien limb phenomenon (Video 409-10) Pronounced asymmetrical rigidity Limb dystonia Stimulus-sensitive myoclonus (Video 409-11) Aphasia (progressive nonfluent aphasia) Cognitive dysfunction (frontotemporal dementia)	Usually negligible MRI may show pronounced asymmetrical cortical atrophy
Vascular parkinsonism	"Lower-half" parkinsonism with gait disturbances predominating, often with minimal or much milder upper body involvement Additional neurologic deficits (e.g., pyramidal tract signs, pseudobulbar palsy)	Usually poor, but some respond well Imaging demonstrates multiple infarcts involving the basal ganglia and subcortical white matter
Dementia with Lewy bodies (a "synucleinopathy")	Early dementia (cognitive profile somewhat different from that of Alzheimer disease) Spontaneous hallucinations, fluctuating cognitive status, falls, orthostatic hypotension, RBD Pronounced sensitivity to the extrapyramidal side effects of neuroleptic drugs Parkinsonism may be similar to typical Parkinson disease, although rigidity may be more prominent than bradykinesia or tremor	Motor features may respond well; psychiatric side effects of dopaminergic drugs are typically dose limiting
Alzheimer disease	Early dementia (memory loss, apraxia, aphasia) Tremor uncommon Spontaneous hallucinations less common than in dementia with Lewy bodies	Poor
Normal-pressure hydrocephalus	"Lower-half" parkinsonism ("gait apraxia") Urinary complaints (frequency, urgency, incontinence) Cognitive disturbances	Generally poor Imaging demonstrates ventriculomegaly out of proportion to cortical atrophy
Drug-induced parkinsonism	All the classic features of parkinsonism (tremor may be less common than in Parkinson disease) Usually symmetrical signs and symptoms Other drug-induced movement disorders (e.g., tardive dyskinesia with neuroleptics)	Usually poor because of ongoing dopamine receptor blockade; may aggravate movements of tardive dyskinesia

MRI = magnetic resonance imaging; RBD = rapid eye movement sleep behavior disorder.

TREATMENT

Rx

Treatment of Parkinson disease is directed at slowing its progression ("neuroprotective" or "disease-modifying" treatments); improving symptoms, typically by restoring dopaminergic tone medically or by correcting basal ganglia neurophysiology surgically ("symptomatic"); or attempting to restore or to regenerate the damaged neurons ("neurorestorative" or "neuroregenerative" therapy).

Exercises may provide benefit. In one randomized trial, for example, tai chi training reduced balance impairments and lowered the incidence of falls in patients with mild to moderate Parkinson disease better than did stretching training.[■] Another trial of three different forms of exercise showed variable benefit in all groups, possibly more with treadmill and resistance exercises.[■]

Medical Treatment

To date, no medical treatment (Table 409-5) has been proved to modify the progressive course of Parkinson disease. The selective monoamine oxidase B inhibitors selegiline and rasagiline[■] may exert disease-modifying effects, and the potential effects of agents such as the calcium-channel blocker isradipine, the pro-urate agent inosine, the peroxisome proliferator-activated receptor γ agonist pioglitazone, and the nicotine patch (encouraged by the lower incidence of Parkinson disease in smokers) are currently under study.

Early treatment in a patient with little or no disability may entail only education, psychological support, encouragement to remain active and to become involved in an exercise program, and ongoing follow-up. There is some evidence that early treatment, even when patients are only mildly symptomatic, may preserve quality of life. A treatment philosophy that still requires evidence-based support involves the early initiation of symptomatic therapy to bolster the brain's compensatory mechanisms that have begun to fail as the physical symptoms of parkinsonism become manifested.

When symptoms begin to interfere with function, mildly effective drugs such as a monoamine oxidase B inhibitor, amantadine, and anticholinergics (the last predominantly for tremor in younger patients) may provide adequate benefit (see Table 409-5).[10] When symptoms are more pronounced or inadequately controlled with these approaches, dopaminergic therapy should be introduced. In patients younger than 65 years who are cognitively intact and

TABLE 409-5 DRUGS FOR PARKINSON DISEASE

CLASS	DRUG	USUAL STARTING DOSE	USUAL FINAL DOSAGE	IMPORTANT ADVERSE EFFECTS	COMMENTS	INDICATIONS
Anticholinergic	Many (e.g., benztropine, trihexyphenidyl)	Benztropine or trihexyphenidyl, 1-2 mg 2-3 times per day	Varied	Peripheral effects, e.g., dry mouth, blurred vision, constipation, difficulty with urination Central effects, e.g., confusion, memory problems, hallucinations	Relatively contraindicated in the elderly and contraindicated in patients with cognitive disturbances	Early treatment of tremor
Miscellaneous	Amantadine	100 mg once per day	100 mg 2 or 3 times per day	Confusion, visual hallucinations; livedo reticularis, swelling of the ankles; dose reduction or drug withdrawal necessary in patients with renal failure	Previously considered a dopaminergic drug, now thought to act primarily through NMDA antagonist effects	Early treatment; later for dyskinesias
	Memantine	5 mg once daily	10 mg twice daily	Confusion, fatigue, dizziness, headache	NMDA antagonist	Possibly effective for cognitive dysfunction in PDD
Dopamine precursor	Levodopa given with peripheral dopa decarboxylase inhibitor (DDCI) (carbidopa [in 4:1 and 10:1 ratios] or benserazide [4:1]*)	50 (levodopa)/12.5 (DDCI) mg (4:1 preparation) 3 times per day (with meals to reduce nausea and vomiting)	Varied; begin with 3-times-daily schedule (controlled-release levodopa-carbidopa may be given twice daily at first); late in the disease, patients may require multiple doses per day (sometimes >2 g/day) Initially give with meals to reduce GI upset; later avoid meals to improve absorption and reliability of response	Peripheral and central dopaminergic side effects Peripheral: nausea, vomiting, and orthostatic hypotension Central: motor fluctuations, dyskinesias, psychiatric disturbances	Peripheral side effects often controlled by additional carbidopa or the peripheral dopamine receptor blocker domperidone* Controlled-release formulations often less bioavailable with less reliable absorption (more "dose failures" later on)	Formulations: immediate-release—for early and later treatment Controlled-release (with carbidopa [4:1] or benserazide [4:1]*)—for predictable motor fluctuations (wearing off) and nighttime akinesia Stalevo (with carbidopa and entacapone)—for wearing off Parcopa (orally disintegrating tablets for faster absorption)—for patients with problematic long latency to benefit with individual doses Melevodopa* (methyl ester of levodopa; effervescent prodrug with much higher water solubility than tablets of levodopa; available in Italy) Duodopa* (used with a pump for duodenal infusions)—for problematic motor fluctuations

TABLE 409-5 DRUGS FOR PARKINSON DISEASE—cont'd

CLASS	DRUG	USUAL STARTING DOSE	USUAL FINAL DOSAGE	IMPORTANT ADVERSE EFFECTS	COMMENTS	INDICATIONS
Dopamine agonists Ergot derived	Bromocriptine	1.25 mg 3 times per day with meals	30-40 mg/day	Peripheral and central dopaminergic side effects; pedal edema, excessive daytime sleepiness Pleuropulmonary reaction, retroperitoneal fibrosis, erythromelalgia Impulse control disorders probably equally common with all dopamine agonists	Peripheral side effects often well controlled with domperidone* Rare pulmonary, retroperitoneal, and skin effects possibly caused by ergot derivation (drug withdrawal usually required)	Early and adjunctive therapy
	Pergolide	0.05 mg once per day × 2 days, increasing slowly thereafter	3-5 mg/day	As for bromocriptine; cardiac valvulopathy	As for bromocriptine	Not the first agonist because it causes restrictive cardiac valve disease
	Cabergoline*	0.5-1 mg once per day	2-6 mg/day	As for pergolide	As for pergolide Long half-life allows once-daily dosage	As for pergolide, although advantage of a long half-life may outweigh this concern
	Lisuride*	0.1-0.2 mg 1-3 times per day	2-5 mg/day	As for bromocriptine	As for bromocriptine	Uncertain whether cardiac valve abnormalities occur Parenteral formulations allow chronic infusion (pump) therapy
Non–ergot derived	Ropinirole	0.25 mg 3 times per day	Up to 24 mg/day in 3 divided doses Once-daily extended/prolonged-release formulation available	Peripheral and central dopaminergic side effects similar to those of ergot-derived dopamine agonists, with the probable exceptions of pleuropulmonary reaction, retroperitoneal fibrosis, erythromelalgia, and cardiac valvulopathy	Effective as first-line and adjunctive therapy; dopamine D_3 agonist effects may contribute to efficacy Some patients withdrawing from the drug (especially those with impulse control disorders) experience symptoms similar to an addictive drug withdrawal ("dopamine agonist withdrawal syndrome")	De novo therapy shown to be associated with fewer motor complications than with levodopa Implications of less progressive loss of dopamine terminal function on imaging uncertain
	Pramipexole	0.125 mg 3 times per day	Up to 4.5 mg/day in 3 divided doses Once-daily extended/prolonged-release formulation available	As for ropinirole	As for ropinirole, possibly greater "D_3-preferring" effects—may account for antidepressant effect	As for ropinirole
	Rotigotine	Nominal dose: 2.0 mg/day (10 cm^2 containing 4.5 mg)	Transdermal patch nominal dose 4.0-16 mg/day (patch content 9-36 mg; 20-80 cm^2)	As for ropinirole Additional adverse effects related to skin patch application (dermatitis)	May be effective for both first-line and adjunctive therapy	
	Piribedil*	50 mg once/day	150-250 mg/day (in 3-5 doses per day)	As for ropinirole	As for ropinirole	
	Apomorphine	3-5 mg SC injection	Parenteral agent given as needed or as continuous infusion	Peripheral and central dopaminergic side effects Local skin reactions, including nodule formation	Concomitant antiemetic (e.g., domperidone,* trimethobenzamide) needed	Late-stage problematic motor fluctuations Long-term use of infusions may reduce dyskinesias as well as motor fluctuations

Monoamine oxidase B inhibitors	Selegiline	5 mg once per day	5 mg 2 times per day	Dopaminergic effects of other drugs possibly accentuated, insomnia, confusion	Last dose given at midday to avoid insomnia	Early mild disease Some controversial evidence suggesting disease-modifying effects Predictable motor fluctuations (wearing off)
	Zydis selegiline	1.25 mg once per day	1.25 or 2.5 mg/day (wafer formulation)	As for selegiline	As for selegiline Absorbed from the buccal mucosa, thereby avoiding first-pass hepatic metabolism and methamphetamine metabolite of selegiline	As for selegiline
	Rasagiline	1 mg once per day	1-2 mg once per day	As for selegiline		Possible disease-modifying effects As for selegiline
Catechol *O*-methyltransferase (COMT) inhibitors	Tolcapone	100 mg 3 times per day	100 or 200 mg 3 times per day (at 6-hour intervals)	Effects of levodopa accentuated Diarrhea in approximately 5% of patients Hepatotoxicity Urine discoloration	Dose of levodopa may have to be reduced by as much as 25%; diarrhea (sometimes explosive) typically forces discontinuation Ongoing monitoring of liver function tests required (second-line COMT inhibitor)	Motor fluctuations, especially wearing off (probably more effective than entacapone)
	Entacapone	200 mg with each dose of levodopa	200 mg 4-10 times per day (given with doses of levodopa)	Effects of levodopa accentuated 10% note brown/orange urine discoloration	As for tolcapone; diarrhea possibly less frequent Liver function monitoring unnecessary	As for tolcapone Available in a combination tablet with levodopa/carbidopa (Stalevo)
A_{2A} antagonist	Istradefylline* (Japan only)	20 mg once per day	20 mg once per day	Increased dyskinesias		
Atypical neuroleptics	Clozapine	12.5 mg hs	Wide range (6.25-150 mg/day), usually <75 mg/day	Agranulocytosis, sedation, hypotension, sialorrhea	Very low risk of worsening parkinsonism; agranulocytosis rare (<1%) and reversible if discovered early (requires regular monitoring of complete blood count)	Drug-induced psychosis Other "off-label" indications include drug-resistant tremor and possibly levodopa-induced dyskinesias
	Quetiapine	12.5-25 mg hs	25-150 mg/day	Sedation May worsen parkinsonism	Probably less effective than clozapine	Drug-induced psychosis
Acetylcholinesterase inhibitors	Donepezil	5 mg once per day	5-10 mg/day	Peripheral cholinergic side effects: nausea, vomiting, diarrhea, syncope, bradycardia Increased tremor, worsening of other Parkinson features		Dementia Possibly effective for psychotic symptoms, especially hallucinations
	Rivastigmine	1.5 mg twice per day	3-12 mg/day	As for donepezil	Patch formulation available for transdermal administration—tolerability may be improved over oral formulation	As for donepezil

*Unavailable in the United States.
GI = gastrointestinal; NMDA = *N*-methyl-D-aspartate; PDD = Parkinson disease dementia.
For evidence-based treatment recommendations, see Suchowersky O, Gronseth G, Perlmutter J, et al. Practice parameter: neuroprotective strategies and alternative therapies for Parkinson disease (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*. 2006;66:976-982; Pahwa R, Factor SA, Lyons KE, et al. Practice parameter: treatment of Parkinson disease with motor fluctuations and dyskinesia (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*. 2006;66:983-995; and Miyasaki JM, Shannon K, Voon V, et al. Practice parameter: evaluation and treatment of depression, psychosis, and dementia in Parkinson disease (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*. 2006;66:996-1002.

lack other major medical problems, initial therapy with a dopamine agonist may delay the development of motor complications. However, these drugs result in more excessive sleepiness, leg edema, "impulse control disorders" (such as pathologic gambling, hypersexuality, binge eating, and shopping), and hallucinations than levodopa does. If a full dose of a dopamine agonist does not provide adequate clinical benefit or has intolerable side effects, levodopa should be initiated. In older patients, in those with cognitive dysfunction (more prone to hallucinations with dopamine agonists), and in circumstances that require more rapid improvement of pronounced disability, levodopa should be the initial drug used.

Alleviating Symptoms

Levodopa is the most effective treatment of Parkinson disease, but it is associated with a variety of side effects (see Table 409-2). For the first year or more, the benefit of levodopa lasts throughout the day with little symptomatic variability. However, in time, the duration of benefit declines, with worsening of symptoms the first thing in the morning (morning akinesia) and for a variable time before scheduled daytime doses (wearing-off/end-of-dose akinesia). Within 2 to 5 years of initiation of treatment, up to 50% of patients may also experience involuntary movements (chorea, athetosis, dystonia), most often at the peak action of the medication. These complications, which are generally more prominent and occur earlier in patients with an onset of disease at a younger age, reflect the short half-life of levodopa combined with the underlying progressive loss of presynaptic dopamine neurons and result in nonphysiologic "pulsatile" stimulation of striatal dopamine receptors, which then induces "neuroplastic" changes in postsynaptic striatal neurons. Initially, these complications rarely cause major disability.

Although initiation of therapy with a dopamine agonist rather than with levodopa may be associated with a delay in the onset of these motor problems, the clinical benefit is generally less than with levodopa,[A1] and all patients eventually require the addition of levodopa to control symptoms. No data support delaying treatment with levodopa, and some data suggest that levodopa could have a neuroprotective effect. Even as Parkinson disease progresses, most of the classic features continue to respond after 20 years or more of treatment. It is not clear that delaying motor complications in the first 5 years of treatment by the initial use of a dopamine agonist improves long-term outcome or quality of life; indeed, clinical status, including the incidence of motor complications, may be no different after 10 years of treatment in those initiating therapy with a dopamine agonist and those starting with levodopa.

There is no clear advantage to starting initial treatment with a controlled-release rather than with an immediate-release preparation of levodopa or combining levodopa with a catechol *O*-methyltransferase inhibitor. When motor fluctuations develop during levodopa therapy, however, they can be managed by a number of approaches (see Table 409-5), including increasing the frequency of the dose, using a controlled-release preparation, prolonging the action by blocking metabolism (monoamine oxidase B or catechol *O*-methyltransferase inhibition), or adding a dopamine agonist. For example, adding rasagiline or entacapone to levodopa provides significant incremental benefits.

Newer levodopa formulations that provide more reliable, sustained plasma levels are in active development.[A2] A formulation that provides continuous infusion into the duodenum (Duodopa) can significantly improve symptoms during "off" time without increasing dyskinesias compared with immediate-release levodopa[A5]; this formulation is available in most European countries and was recently approved in Canada and the United States for patients with problematic motor fluctuations. Dyskinesias improve when doses of dopaminergic medications are reduced, but the parkinsonism often increases to an intolerable level. Amantadine may improve the dyskinesias without worsening the parkinsonism. Newer agents under study include the α_2-adrenergic receptor antagonist fipamezole[A7] and antagonists of the metabotropic glutamate receptor mGluR5, such as AFQ056,[A6] but these medications are not yet approved for use. Antagonists of the adenosine A_{2A} receptor are being actively studied but with variable results. One of these, istradefylline, is marked in Japan for the treatment of wearing-off phenomenon.

Medical management of Parkinson disease often includes a variety of other agents, including medications directed at the treatment of orthostatic hypotension (Chapter 62), depression (Chapter 397), anxiety (Chapter 397), urinary frequency and urgency (Chapters 26 and 129), and male erectile dysfunction (Chapter 234).[11] Management of late-stage Parkinson disease requires skill in polypharmacy and an understanding of the complicated benefit-risk ratios of the many drugs needed.

Surgical Treatment

Bilateral deep brain stimulation of the subthalamic nucleus or globus pallidus improves the symptoms of Parkinson disease, often permits lower doses of antiparkinson medications to be used, improves self-reported quality of life, and is about twice as effective as medical therapy despite the adverse effects associated with the procedure.[A3] Early use of subthalamic nucleus deep brain stimulation, at a time when patients are just beginning to develop motor complications (mean duration of disease, 7.5 years), provides significantly greater benefit than best medical therapy. Thalamic deep brain stimulation is of limited utility because it is effective only for tremor. The best predictor of a good response to deep brain stimulation of the subthalamic nucleus is the patient's ongoing clinical response to levodopa.[12] Apart from tremor, which may be resistant to the highest tolerable dose of levodopa but generally responds well to surgery, symptoms that are resistant to the peak effect of levodopa (e.g., dysarthria, postural instability with falls) also fail to respond to deep brain stimulation. The typical good candidate for deep brain stimulation of the subthalamic nucleus is an otherwise healthy, relatively young, cognitively intact, and psychiatrically stable patient who still responds well to levodopa (apart from tremor) but is suffering from disabling motor fluctuations and dyskinesias.

Double-blind randomized trials of transplantation of fetal substantia nigra into the striatum have failed to show significant efficacy and also have been associated with the side effect of transplant-induced off-medication dyskinesias. Postmortem assessments in a small number of patients surviving for more than 10 years after fetal transplants have shown Lewy bodies in the transplanted dopamine neurons, thus suggesting that the pathologic process can be "transmitted" to neurons placed in the diseased host.[13] Nevertheless, some patients treated with fetal transplants have responded well for many years, and research in this field is ongoing.

A sham-controlled study demonstrated significant benefit with gene transfer of glutamic acid decarboxylase (AAV-GAD) into the subthalamic nucleus in patients with advanced Parkinson disease.[A4] By comparison, double-blind randomized trials of bilateral intraputamenal infusion of glial-derived neurotrophic factor and gene therapy with AAV-neurturin (administered into the putamen and the substantia nigra) failed to confirm the benefits that were suggested by unblinded studies.[14]

PROGNOSIS

Parkinson disease progresses inexorably during a period of many years; the speed and course of progression vary considerably from patient to patient. So far, genotypic information has not helped predict outcomes.[15] Some patients maintain an excellent response to treatment and seem to change very little during prolonged follow-up, but most note increasing disability, with the development of many symptoms that are poorly responsive to medications. Factors such as poor postural stability, falls, dysarthria, dysphagia, dysautonomia, excessive daytime sleepiness, and dementia contribute to the disability and increased mortality.

FUTURE DIRECTIONS

Gene therapies directed at either modifying neurotransmitter function or inducing neuroregeneration and other cell-based therapies are under development. Future treatments must also address the widespread, multisystemic nature of the disease, especially symptoms that are unrelated to nigrostriatal dopamine deficiency and that fail to respond to current therapies.

OTHER CAUSES OF PARKINSONISM

The numerous causes of parkinsonism (see Table 409-1) are sometimes termed akinetic-rigid syndrome, Parkinson syndrome, atypical parkinsonism, or even Parkinson-plus syndrome to emphasize that these patients commonly demonstrate additional clinical features indicative of the more widespread and particularly more severe pathologic involvement of areas beyond the dopaminergic SNc. These other parkinsonism conditions are generally associated with "postsynaptic" changes that result in a poor or unsustained response to levodopa, and this unresponsiveness serves as one of the most important of several clues that the parkinsonism features are caused by conditions other than Parkinson disease (see Table 409-4) (i.e., "parkinsonism minus" a levodopa response; see Table 409-3).

Grade A References

A1. Li F, Harmer P, Fitzgerald K, et al. Tai chi and postural stability in patients with Parkinson's disease. *N Engl J Med*. 2012;366:S11-519.

A2. Shulman LM, Katzel LI, Ivey FM, et al. Randomized clinical trial of 3 types of physical exercise for patients with Parkinson disease. *JAMA Neurol*. 2013;7:183-190.

A3. Olanow CW, Rascol O, Hauser R, et al. A double-blind, delayed start trial of rasagiline in Parkinson's disease. *N Engl J Med*. 2009;361:1268-1278.

A4. Gray R, Ives N, Rick C, et al. Long-term effectiveness of dopamine agonists and monoamine oxidase B inhibitors compared with levodopa as initial treatment for Parkinson's disease (PD MED): a large, open-label, pragmatic randomised trial. *Lancet*. 2014;384:1196-1205.

A5. Hauser RA, Hsu A, Kell S, et al. Extended-release carbidopa-levodopa (IPX066) compared with immediate-release carbidopa-levodopa in patients with Parkinson's disease and motor fluctuations: a phase 3 randomised, double-blind trial. *Lancet Neurol*. 2013;12:346-356.

A6. Olanow CW, Kieburtz K, Odin P, et al. Continuous intrajejunal infusion of levodopa-carbidopa intestinal gel for patients with advanced Parkinson's disease: a randomised, controlled, double-blind, double-dummy study. *Lancet Neurol*. 2014;13:141-149.

A7. Lewitt PA, Hauser RA, Lu M, et al. Randomized clinical trial of fipamezole for dyskinesia in Parkinson disease (FJORD study). *Neurology*. 2012;79:163-169.

A8. Stocchi F, Rascol O, Destee A, et al. AFQ056 in Parkinson patients with levodopa-induced dyskinesia: 13-week, randomized, dose-finding study. *Mov Disord*. 2013;28:1838-1846.
A9. Deuschl G, Schade-Brittinger C, Krack P, et al. A randomized trial of deep-brain stimulation for Parkinson's disease. *N Engl J Med*. 2006;355:896-908.
A10. Schuepbach WM, Rau J, Knudsen K, et al. Neurostimulation for Parkinson's disease with early motor complications. *N Engl J Med*. 2013;368:610-622.
A11. Weaver FM, Follett KA, Stern M, et al. Randomized trial of deep brain stimulation for Parkinson disease: thirty-six-month outcomes. *Neurology*. 2012;79:55-65.
A12. LeWitt PA, Rezai AR, Leehey MA, et al. AAV2-GAD gene therapy for advanced Parkinson's disease: a double-blind, sham-surgery controlled, randomised trial. *Lancet Neurol*. 2011;10:309-319.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Trinh J, Farrer M. Advances in the genetics of Parkinson disease. *Nat Rev Neurol*. 2013;9:445-454.
2. Kieburtz K, Wunderle KB. Parkinson's disease: evidence for environmental risk factors. *Mov Disord*. 2013;28:8-13.
3. Jellinger KA. Neuropathology of sporadic Parkinson's disease: evaluation and changes of concepts. *Mov Disord*. 2012;27:8-30.
4. Visanji NP, Brooks PL, Hazrati LN, et al. The prion hypothesis in Parkinson's disease: Braak to the future. *Acta Neuropathol Commun*. 2013;1:2.
5. Elias WJ, Shah BB. Tremor. *JAMA*. 2014;311:948-954.
6. Emre M, Ford PJ, Bilgic B, et al. Cognitive impairment and dementia in Parkinson's disease: practical issues and management. *Mov Disord*. 2014;29:663-672.
7. Aarsland D, Taylor JP, Weintraub D. Psychiatric issues in cognitive impairment. *Mov Disord*. 2014;29:651-662.
8. Berardelli A, Wenning GK, Antonini A, et al. EFNS/MDS-ES/ENS recommendations for the diagnosis of Parkinson's disease. *Eur J Neurol*. 2013;20:16-34.
9. Seibyl J, Russell D, Jennings D, et al. Neuroimaging over the course of Parkinson's disease: from early detection of the at-risk patient to improving pharmacotherapy of later-stage disease. *Semin Nucl Med*. 2012;42:406-414.
10. Connolly BS, Lang AE. Pharmacological treatment of Parkinson disease: a review. *JAMA*. 2014; 311:1670-1683.
11. Seppi K, Weintraub D, Coelho M, et al. The Movement Disorder Society Evidence-Based Medicine Review Update: treatments for the non-motor symptoms of Parkinson's disease. *Mov Disord*. 2011;26(suppl 3):S42-S80.
12. Okun MS. Deep-brain stimulation for Parkinson's disease. *N Engl J Med*. 2012;367:1529-1538.
13. Barker RA, Barrett J, Mason SL, et al. Fetal dopaminergic transplantation trials and the future of neural grafting in Parkinson's disease. *Lancet Neurol*. 2013;12:84-91.
14. Kordower JH, Bjorklund A. Trophic factor gene therapy for Parkinson's disease. *Mov Disord*. 2013;28:96-109.
15. Chung SJ, Biernacka JM, Armasu SM, et al. Alpha-synuclein repeat variants and survival in Parkinson's disease. *Mov Disord*. 2014;29:1053-1057.

REVIEW QUESTIONS

1. You are observing a 55-year-old man who was diagnosed with Parkinson disease 3 years ago. On examination, he has symmetrical slowing of his limb movements and gait with bradykinesia, rigidity, postural tremor, and postural instability. Cognitive function is preserved, and eye movements are normal. He has been receiving gradually increasing doses of a levodopa preparation and is currently taking levodopa/carbidopa 800/200 mg per day. He has noted clear benefit, but he has facial dystonic movements and additional wearing off between doses. At peak benefit, he has a soft voice that is somewhat difficult to understand; he has a great deal of difficulty with a variety of activities of daily living and has had to start using a cane because of the tendency to fall. Which of the following options should be considered at this time?

A. Add an anticholinergic drug such as trihexyphenidyl 2 mg tid.
B. Add a dopamine agonist such as pramipexole increasing to 1.5 mg tid.
C. Add a cholinesterase inhibitor such as donepezil 5 mg bid.
D. Obtain a magnetic resonance imaging (MRI) study including T2* or GRE sequences.
E. Refer him to a surgical program for consideration of subthalamic nucleus deep brain stimulation.

Answer: D This patient's disability has progressed too rapidly for typical Parkinson disease. He has no resting tremor, no asymmetry of symptoms, and early postural instability. His response to levodopa is relatively poor, and levodopa-induced dyskinesias are limited to facial dystonia. These features are strongly suggestive of an atypical parkinsonism, such as multiple system atrophy. In support of multiple system atrophy, it would be extremely important to assess for evidence of autonomic dysfunction on his history and physical examination, including orthostatic hypotension, erectile dysfunction, and bladder incontinence. None of the treatment options outlined would be appropriate. Anticholinergics are largely effective for resting tremor. There is no evidence in this situation that a dopamine agonist would provide greater benefit than levodopa, and it may be more poorly tolerated, especially if he has problematic orthostatic hypotension. Cholinesterase inhibition is largely used for cognitive dysfunction in Parkinson disease; although it has been suggested that gait disturbances and postural instability may benefit from this treatment, such benefit has not been shown in the atypical forms of parkinsonism. Deep brain stimulation would not be expected to benefit symptoms that are not responding to levodopa. MRI scanning may provide support for the diagnosis of multiple system atrophy, including putamenal atrophy and evidence for iron deposition in putamen on T2* or GRE sequences.

2. A 55-year-old lawyer presents with right arm resting tremor, mild slowing and fatiguing on the performance of rapid repetitive movements, increased tone in his right arm with activation of the left, and reduced right arm swing but normal stride and postural stability. His writing is a bit smaller than before, and he is having mild difficulty using his computer mouse and performing tasks requiring fine dexterity in the right hand. His left side is entirely normal. Which therapy (with titration as necessary) would not be considered appropriate at this time?

A. Levodopa/carbidopa 100/25 mg tid
B. Rasagiline 1 mg/day
C. Ropinirole 5 mg tid
D. Stalevo 100 mg tid
E. Pramipexole prolonged release 3 mg/day

Answer: D The patient's symptoms are mild but bothersome and justify initiation of therapy. Rasagiline is the least effective of all the agents listed, but it may provide mild benefit and is easy to use and well tolerated. Dopamine agonists are associated with fewer motor complications than levodopa, particularly dyskinesias, but they are less efficacious and have important side effects that must be monitored carefully, including excessive daytime sleepiness and impulse control disorders. Levodopa is the most effective treatment for Parkinson disease, and it is a reasonable choice even in the earliest stages if symptoms are interfering with quality of life. Stalevo is a combination preparation of levodopa/carbidopa/entacapone. In the early stages of disease, this treatment does not provide an advantage over levodopa and may be associated with more dyskinesias. The main indication for Stalevo is predictable wearing off of benefit with levodopa.

3. A 46-year-old woman has had Parkinson disease for the past 5 years. She is taking levodopa/carbidopa 150/37.5 mg every 3 hours as well as pramipexole 1 mg three times per day. At the peak benefit of levodopa, she functions near-normally but still has generalized dyskinesias that are occasionally bothersome to her. About 15 to 20 minutes before each dose of levodopa, she begins to notice a return of her tremor and hand clumsiness as well as occasional shuffling of gait and rare freezing. It takes approximately 30 minutes for the benefit of the next dose to "kick in," and occasional doses of levodopa do not seem to work. Which of the following treatment options would be most appropriate?

A. Switch the immediate-release pramipexole to the extended-release formulation in the same dosage.
B. Switch the immediate-release to the controlled-release levodopa/carbidopa preparation.
C. Switch the levodopa/carbidopa preparation to Stalevo 175 mg every 3 hours.
D. Add selegiline 5 mg/day.
E. Refer to a surgical program for consideration of deep brain stimulation.

Answer: E This young patient who retains an excellent response to levodopa but with problematic motor fluctuations is an ideal candidate for deep brain stimulation. Extended-release formulations of dopamine agonists usually are no more effective for motor fluctuations than are the immediate-release preparations. Controlled-release formulations of levodopa are less bioavailable than immediate-release preparations and would likely worsen some of the unpredictability that she has in her responses. Additional entacapone, particularly with higher doses of levodopa, would likely worsen the dyskinesias, although the amount of off-time might improve. Selegiline would likely provide only minimal or mild benefit, particularly at this lower dose.

4. A 75-year-old woman has had Parkinson disease for 15 years. She continues to obtain benefit from levodopa, but even at the peak response she requires assistance for many tasks, uses a walker for ambulation, and has bothersome dyskinesias. She takes a levodopa preparation every 4 hours while awake. Each dose lasts approximately 3 hours, at which time her tremor can become bothersome and she has more difficulty in walking. Her short-term memory has declined, and she is occasionally confused. She comments on seeing people that she does not recognize in the dining room or children in the branches of the tree outside her kitchen. Which of the following treatments might be considered in the management of some of her problems?

A. Amantadine
B. Clozapine
C. Pramipexole
D. Risperidone
E. Trihexyphenidyl

Answer: B Clozapine may be beneficial for several of this woman's symptoms, including hallucinations, bothersome off-period tremor, and dyskinesias. Amantadine, pramipexole, and trihexyphenidyl would be contraindicated, given her cognitive difficulties and the presence of hallucinations. Risperidone is also contraindicated in Parkinson disease because it can worsen the underlying parkinsonism.

5. A 65-year-old man has been diagnosed with Parkinson disease for 8 years. Which of the following clinical features would not be compatible with this diagnosis?

A. A drop in systolic blood pressure of 30 mm Hg and diastolic blood pressure of 15 mm Hg from lying to standing
B. A prominent postural and action tremor but no resting tremor
C. Inability to obtain an erection for the past 3 years
D. Frequent falls, particularly triggered by freezing of gait
E. A wide-based gait and complete inability to perform tandem walking

Answer: E Patients with Parkinson disease can develop disabling autonomic dysfunction, including orthostatic hypotension and erectile dysfunction. These symptoms generally are not present or problematic early in the disease. Falls become a problem after several years of disease, but they also are not an early manifestation. Many patients with Parkinson disease do not have a resting tremor, but most of these will have a postural and action tremor. Even when patients with Parkinson disease have postural instability, they tend to walk on a narrow base and can perform tandem walking unless they are extremely unstable. This latter feature suggests additional cerebellar disease and may be supportive of a diagnosis of multiple system atrophy.

410 OTHER MOVEMENT DISORDERS

ANTHONY E. LANG

DEFINITION

Movement disorders are first divided into hypokinetic and hyperkinetic categories. *Hypokinetic disorders,* which are characterized by akinesia, bradykinesia, and rigidity, are parkinsonian syndromes and are discussed elsewhere (Chapter 409). The common *hyperkinetic movement* disorders (Table 410-1) are defined by their specific clinical phenomena.

CLINICAL MANIFESTATIONS AND DIAGNOSTIC APPROACH

The traditional approach to a neurologic symptom is first to address localization within the nervous system (i.e., "Where is the lesion?"), followed by an evaluation of the origin ("What is the lesion?").[1] The neurologic examination is critical in determining the localization of the lesion, and generally the history, including the nature of onset and the progression of the symptoms, determines the most likely diagnosis. However, when a movement disorder is the predominant problem, the approach is somewhat different. The pathophysiology of most movement disorders is complex and often poorly understood. Many of these disorders are the result of dysfunction of different circuits in the brain, and it is often impossible to ascertain a specific anatomic localization. Instead, an accurate appreciation of the clinical phenomena is the first important step in evaluating these patients. The clinician must observe and examine the patient to define the type of movement disorder that best describes the clinical picture. This accurate characterization then allows the generation of a differential diagnosis for the specific movement disorder. The age and nature of onset, the distribution, the progression of symptoms, a family history of similar or related symptoms, and the presence of other neurologic and systemic signs then help to determine the most likely explanation for that movement disorder.

TREMOR

Tremor, which is a rhythmic, sinusoidal movement of a body part, is caused by regular, either synchronous or alternating, contractions of reciprocally innervated muscles.[2] Tremors are classified based on whether they occur at rest (weight fully supported against gravity) or in action. Resting tremors are typically seen in Parkinson disease and other parkinsonism syndromes (see Table 409-1). Action tremors are further divided into postural, kinetic, or intention tremors. A postural tremor is seen with the maintenance of a posture against gravity (e.g., when the arms are outstretched in front of the body). A *kinetic tremor* is seen with a voluntary movement of the limb (e.g., a tremor in an upper limb when performing the finger-to-nose test). An intention tremor increases in amplitude on approaching a target.

TABLE 410-1 HYPERKINETIC MOVEMENT DISORDERS

Tremor
Chorea
Ballism
Dystonia
Athetosis
Tics
Myoclonus
Startle
Stereotypies
Miscellaneous

CLINICAL MANIFESTATIONS

Most action tremors (Table 410-2) combine postural and kinetic components. All tremors worsen with stress, including performing an affected activity in public. Initially, a tremor may be evident only when one attempts fine, dexterous tasks such as threading a needle, soldering, or using a screwdriver. More severe tremors interfere with activities such as handwriting, fastening buttons, shaving, eating soup with a spoon, or drinking from a cup. Patients

TABLE 410-2 DIFFERENTIAL DIAGNOSIS OF TREMOR AND RHYTHMIC MOVEMENT DISORDERS

ENHANCED PHYSIOLOGIC TREMOR
Metabolic disorders
Hyperthyroidism
Hyperparathyroidism
Hypoglycemia
Pheochromocytoma
Drugs
Caffeine
Theophylline
Amphetamines
Lithium
Valproic acid
Antidepressants
Amiodarone
β-Agonists
Others
Withdrawal of drugs
Benzodiazepines
Alcohol
Others
Fever, sepsis
Anxiety, stress, fatigue
PRIMARY OR IDIOPATHIC TREMOR
Essential tremor
Task-specific tremor
Orthostatic tremor
Idiopathic palatal tremor
TREMOR ASSOCIATED WITH CENTRAL NERVOUS SYSTEM DISEASES
Tremor with parkinsonian syndromes
Idiopathic Parkinson disease
Multiple system atrophy
Progressive supranuclear palsy
Corticobasal degeneration
Neuroleptic-induced parkinsonism
Wilson disease
Multiple sclerosis
Fragile X premutation-tremor/ataxia syndrome
Stroke
Arteriovenous malformation
Tumor
Head trauma
Midbrain tremor (Holmes tremor)
TREMOR ASSOCIATED WITH PERIPHERAL NEUROPATHIES
PSYCHOGENIC TREMOR
OTHER RHYTHMIC MOVEMENT DISORDERS
Rhythmic movements in dystonia (dystonic tremor)
Rhythmic myoclonus (including myoclonic tremor)
Asterixis
Clonus
Epilepsia partialis continua
Hereditary chin quivering
Spasmus nutans
Head bobbing with hydrocephalus
Nystagmus

Modified from Cloutier M, Lang AE. Movement disorders: an overview. In: Factor SA, Lang AE, Weiner WJ, eds. *Drug Induced Movement Disorders*. Malden, Mass: Blackwell; 2005:3-19.

often adapt or use compensatory measures, such as switching an activity to a less affected hand (e.g., shaving with the nondominant hand), using two hands to drink, drinking only from an incompletely filled glass or cup, or completely avoiding more challenging feeding activities in public. Severe action and intention tremors can cause handwriting to become completely illegible and can result in dependence on others for care.

Head tremors, which may be side to side, up and down, or mixed, are rarely disabling but are often a source of embarrassment. Tremor of the larynx, which causes the voice to quaver, is best appreciated by asking the patient to sustain a note. Action tremor of the lower limbs is assessed by having the patient hold the foot up to a target (e.g., the examiner's hand) and then perform a heel-knee-shin test.

Most upper limb action tremors affect many activities to a similar extent. Less commonly, tremors can affect a single task in isolation (task-specific tremors), the most common being a primary writing tremor. Orthostatic tremor is apparent in the legs and in antigravity muscles only when the patient is standing in one spot and subsides during walking or leaning against a wall; these patients commonly complain of a tremendous sense of insecurity while standing and a fear of falling. Electrophysiologic assessment demonstrates a very characteristic high-frequency tremor (14 to 16 Hz).

Enhanced Physiologic Tremor

A 7- to 12-Hz tremor is detectable in everyone with electrophysiologic recording. This physiologic tremor is enhanced and may become symptomatic in a variety of circumstances, including fatigue, anxiety, and excitement. This same tremor may be accentuated by drugs and systemic processes.

Essential Tremor

Essential tremor affects up to 5% of the general population after the age of 60 years. Essential tremor is often inherited in an autosomal dominant fashion, with the phenotype showing genetic heterogeneity from at least three different genes, most recently *LINGO1*,[3] as well as environmental influences. Recent pathology studies have variably demonstrated microscopic abnormalities of cerebellar Purkinje cells. The age of onset may be as early as the first or second decade of life, but senile tremor may be delayed until the mid-60s. Patients first become aware of a mild postural and action tremor in the hands, which is indistinguishable from an enhanced physiologic tremor and may result in little functional impairment for many years until it gradually interferes with activities. Older patients with large-amplitude, lower frequency tremors can have a resting component that is often misdiagnosed as Parkinson disease (see Table 409-1) (Video 410-1, Essential Tremor).

TREATMENT

Rx

Treatment of essential tremor does not influence the course of the illness and therefore is justified only when the tremor interferes with function. At least 50% of patients note improvement or complete amelioration of tremor following the ingestion of a small amount of ethanol.

First-line drug treatment includes trials of a noncardioselective β-adrenergic blocker (e.g., propranolol, ≤ 320 mg/day), primidone (starting in a low dose of 25 to 62.5 mg at night and increasing to 500 to 750 mg/day), or topiramate (≤400 mg/day). Other drugs that have been shown probably to be effective in double-blind crossover trials include gabapentin (1200 to 1800 mg/day), atenolol (50 to 150 mg/day), alprazolam (0.125 to 3 mg/day), and sotalol (75 to 200 mg/day). However, sotalol is associated with ventricular arrhythmias and dose-related QT interval prolongation, so it is not routinely considered as treatment of essential tremor. The medications that have been shown to be of possible benefit include nadolol (120 to 240 mg/day), nimodipine (120 mg/day), and clonazepam (0.5 to 6 mg/day), but many patients remain resistant to all drugs. Botulinum toxin may be effective, but it also may result in dose-dependent weakness and pain at the injection site. If disability is substantial, thalamic deep brain stimulation or thalamotomy can be of major benefit, with 60 to 90% reductions following bilateral treatment.[4] However, a few patients suffer permanent neurologic sequelae, such as speech problems, owing to intracranial hemorrhage or other postoperative complications, even with unilateral procedures, and even more suffer such problems after bilateral procedures. Newer approaches to thalamotomy include gamma knife and focused ultrasound.[5]

CHOREA

Chorea (Table 410-3) consists of irregular, random, brief, flowing movements that often flit from one body part to another in an unpredictable and purposeless sequence. Patients may incorporate choreiform movements into a voluntary movement to mask them. The severity varies from the appearance of being slightly fidgety or restless, to striking, continuous movements involving the whole body. Many patients with chorea seem unaware of their movements, whereas others can be very troubled and disabled.

TABLE 410-3 DIFFERENTIAL DIAGNOSIS OF CHOREA

GENETIC DISORDERS
Benign hereditary chorea
Huntington disease
Huntington-like conditions
Neuroferritinopathy
Neuroacanthocytosis, including McLeod syndrome
Dentatorubropallidoluysian atrophy
Wilson disease
Neurodegeneration with brain iron accumulation 1 (NBIA 1) (previously Hallervorden Spatz disease)
Spinocerebellar ataxias
Ataxia-telangiectasia
Ataxia-oculomotor apraxia type 1
Tuberous sclerosis
INFECTIONS/PARAINFECTIOUS CAUSES
Sydenham chorea
Acquired immunodeficiency syndrome (including complications)
Encephalitis and postencephalitic disorders
Creutzfeldt-Jakob disease
DRUGS
Levodopa
Dopaminergic agonists used for Parkinson disease
Amphetamines
Anticholinergics
Anticonvulsants (especially phenytoin)
Neuroleptics
Tricyclic antidepressants
Selective serotonin reuptake inhibitors (occasionally)
Oral contraceptives (typically in patients with a prior history of Sydenham chorea)
Antihistaminics
ENDOCRINOLOGIC/METABOLIC CONDITIONS
Hyperthyroidism
Hypoparathyroidism
Chorea gravidarum
Acquired hepatolenticular degeneration
IMMUNOLOGIC DISORDERS
Systemic lupus erythematosus
Antiphospholipid syndrome
Henoch-Schönlein purpura
VASCULAR DISORDERS
Stroke
Hemorrhage
Arteriovenous malformation
Polycythemia rubra vera
OTHER CONDITIONS
Cerebral palsy
Kernicterus
Head trauma
Cardiopulmonary bypass with hypothermia
Neoplastic and paraneoplastic syndromes
Paroxysmal dyskinesias

Modified from Cloutier M, Lang AE. Movement disorders: an overview. In: Factor SA, Lang AE, Weiner WJ, eds. *Drug Induced Movement Disorders*. Malden, Mass: Blackwell; 2005:3-19.

Huntington Disease

DEFINITION AND EPIDEMIOLOGY

Huntington disease is a fully penetrant autosomal dominant neurodegenerative disorder caused by an expanded trinucleotide (CAG) repeat in the gene for the protein huntingtin. The worldwide 2.71 per 100,000 prevalence ranges from 5.7 per 100,000 for individuals of European descent to 0.4 per 100,000 for Asians.[6] The age at diagnosis is driven by the longest expanded allele and as yet unidentified genetic or environmental factors.

PATHOBIOLOGY

Huntington disease is characterized neuropathologically by neuronal loss accompanied by intraneuronal inclusions and gliosis, especially in the caudate

nucleus and putamen (the striatum) and the cerebral cortex. Understanding how these changes result from the expanded polyglutamine tract in the mutated huntingtin protein is the goal of current research.

CLINICAL MANIFESTATIONS

Symptoms typically begin between the ages of 30 and 55 years, but 5 to 10% of patients have an onset before the age of 20 years (juvenile Huntington disease) and a few patients begin to have symptoms quite late in life. Symptoms include a combination of a movement disorder, psychiatric disturbances, and cognitive dysfunction. Early on, the movement disorder is predominantly chorea, but parkinsonism and dystonia develop later (Video 410-2, Huntington Disease). Some patients, especially those with juvenile onset, have a more rapidly progressive akinetic-rigid and dystonic form (the Westphal variant). Psychiatric manifestations, which are universal but widely variable, include personality changes, impulsiveness, aggressive behavior, depression, and paranoid psychosis. These psychiatric symptoms may precede the motor manifestations, and psychotropic drug therapy may be incorrectly blamed for the subsequent development of the movement disorder. Cognitive changes result in progressive subcortical dementia with disturbed attention, concentration, judgment, and problem-solving that differs from the typical cortical dementia of Alzheimer disease. Oculomotor dysfunction, most often manifested by difficulties with refixating the gaze and a resulting tendency to use blinks and head thrusts, is another common feature.

DIAGNOSIS

The diagnosis is confirmed by genetic testing. Normal alleles of the *IT15* gene have fewer than 30 CAG repeats, whereas 40 or more repeats invariably result in clinical illness. An earlier age of onset correlates with larger numbers of CAG repeats. Patients with intermediate alleles (27 to 35) have more behavioral abnormalities, such as apathy and suicidal ideation, than unaffected individuals.

TREATMENT AND PROGNOSIS

Rx

Current care for patients with Huntington disease involves a multidisciplinary team of clinical geneticists, neurologists, psychiatrists, psychologists, social workers, occupational and physical therapists, speech therapists, nutritionists, and nurses. Genetic counseling for patients and family members is critical. Chorea may be extremely responsive to drugs that reduce central dopamine activity, especially tetrabenazine, starting at 12.5 mg two or three times daily and gradually increasing to up to 100 to 200 mg/day. Amantadine (300 to 400 mg/day) and possibly riluzole (200 mg/day) may reduce chorea. Other potential agents that work by blocking dopamine receptors include haloperidol (3 to 30 mg/day), pimozide (0.5 to 10 mg/day), fluphenazine (0.5 to 20 mg/day), and reserpine (0.75 to 5 mg/day). These agents should be reserved for patients with disabling chorea because they may be associated with increased parkinsonism, postural instability, depression, sedation, and other adverse effects. Unfortunately, physical function may not improve significantly even when the chorea is controlled. Psychiatric symptoms (e.g., anxiety, psychosis, depression) can be managed effectively with the same strategies as in other psychiatric diseases (Chapter 397). Disease-modifying strategies are under active development.[7]

Progression can be monitored by following changes in gray matter volumes in both premanifest and early-stage patients.[8] Patients inexorably decline at a relatively constant rate,[9] and the disease progresses to institutionalization and death over the course of approximately 15 years.

Other Choreas

Most of the non-neurodegenerative causes of chorea (see Table 410-3) can be excluded by a careful history (including a detailed drug history) and a focused set of investigations, including, in appropriate circumstances, wet preparation of peripheral blood for acanthocytes (which are associated with neurodegeneration), immunologic studies (including anticardiolipin antibodies), endocrine assessment (hyperthyroidism, pregnancy), and neuroimaging. Among 36 adult cases of autoimmune chorea seen at one institution in 5 years, 50% had a coexisting autoimmune disorder, especially systemic lupus erythematosus (Chapter 266), and most of the remainder had a paraneoplastic cause, especially small cell carcinoma of the lung and adenocarcinoma.[10]

Sydenham chorea, which is a late component of rheumatic fever (Chapter 290), is presumably the result of immunologic cross-reactivity between the causative group A β-hemolytic streptococcus and the basal ganglia. This disorder is infrequently seen in North America but is more common in developing countries. Sydenham chorea usually affects children and young adults, and it is more common in girls before puberty. Adults with a history of Sydenham chorea in childhood may develop chorea during pregnancy or in response to taking oral contraceptive agents or estrogen preparations. They also may have a higher rate of subsequent psychiatric disturbances and impaired executive neurologic function even when in remission.[11] Drugs that can cause chorea should be withdrawn if possible.

TABLE 410-4 DIFFERENTIAL DIAGNOSIS OF BALLISM

Focal lesions in basal ganglia
Vascular: Stroke (including infarction and hemorrhage), cavernous angioma, postsurgical complications
Neoplastic: Metastases, primary central nervous system tumors
Infections: Cryptococcosis, toxoplasmosis, tuberculoma
Inflammatory: Multiple sclerosis
Iatrogenic: Subthalamotomy, thalamotomy
Immunologic: Systemic lupus erythematosus, scleroderma; Behçet's disease
Nonketotic hyperglycemia (high-intensity lesions in striatum on T1 MRI)
Hypoglycemia
Sydenham's chorea
Head injury
Drugs
Anticonvulsants
Oral contraceptives
Levodopa

Modified from Cloutier M, Lang AE. Movement disorders: an overview. In: Factor SA, Lang AE, Weiner WJ, eds. *Drug Induced Movement Disorders*. Malden, Mass: Blackwell; 2005:3-19.
MRI = magnetic resonance imaging.

Ballism

Ballism, which is considered an extreme form of chorea, involves large-amplitude, random, often violent flinging movements of the proximal limbs (Table 410-4). It is most often a consequence of an acute cerebral insult, such as a stroke, and it usually involves one side of the body, particularly the arm, hence the term *hemiballism* (Video 410-3, Hemiballism). When a causative lesion can be demonstrated, it typically involves the region of the subthalamic nucleus or the striatum. When the condition is caused by a stroke, movements usually subside spontaneously over days to weeks, although they may persist indefinitely in some patients. Treatment often requires the use of medication that antagonizes the effects of dopamine in the brain, including dopamine receptor blockers (neuroleptics such as haloperidol, 3 to 30 mg/day) or dopamine depleters (e.g., tetrabenazine, 50 to 200 mg/day). Functional neurosurgery (e.g., pallidotomy, deep brain stimulation) can be considered in patients with refractory, persistent symptoms.

DYSTONIA

DEFINITION AND PATHOBIOLOGY

In dystonia, sustained muscle contractions, often initiated or worsened by voluntary action, result in repetitive twisting and sometimes tremulous movements and abnormal postures. Dystonia can be classified as primary dystonia, dystonia-plus, secondary dystonia, and heredogenerative dystonias (Table 410-5). Recently, a new classification uses five descriptors to specify the clinical characteristics: age at onset, body distribution, temporal pattern, whether dystonia occurs in isolation (or only accompanied by tremor; "pure dystonia") or coexists with other movement disorders (typically parkinsonism and myoclonus).[12] *Etiology* is defined as the presence or absence of degenerative or structural nervous system pathologic process and by whether the mode of inheritance is autosomal dominant, autosomal recessive, X-linked recessive, mitochondrial, or acquired. A commonly used classification scheme for the genetic dystonias involves applying the "DYT" prefix followed by a number (e.g., 1 to 25); however, several shortcomings have encouraged an active reevaluation of this approach. Acquired causes include drugs, toxins, infections, vascular disease, neoplasia, trauma, and psychogenic.

CLINICAL MANIFESTATIONS

Common forms of dystonia include eyelid closure (blepharospasm), jaw opening or closing (oromandibular dystonia), pulling or turning of the neck in any one or combination of directions (cervical dystonia: rotatory torticollis, laterocollis, retrocollis, anterocollis), hyperadduction and less often excessive abduction of the vocal cords (laryngeal dystonia or spasmodic

TABLE 410-5 CLASSIFICATION AND CAUSES OF DYSTONIA

PRIMARY DYSTONIAS (PRIMARY TORSION DYSTONIA)
Familial (several genetic causes and types)
Sporadic, usually adult onset, focal, or segmental
DYSTONIA-PLUS
Dystonia with parkinsonism
Dopa-responsive dystonia
Dopamine agonist–responsive dystonia (e.g., aromatic acid decarboxylase deficiency)
Myoclonus dystonia
SECONDARY DYSTONIAS
Perinatal cerebral injury
Athetoid cerebral palsy
Delayed-onset dystonia
Pachygyria
Kernicterus
Encephalitis
Reye syndrome
Subacute sclerosing leukoencephalopathy
Wasp sting
Creutzfeldt-Jakob disease
Human immunodeficiency virus infection
Head trauma
Thalamotomy
Brain stem lesion
Primary antiphospholipid syndrome
Stroke
Arteriovenous malformation
Hypoxia
Brain tumor
Multiple sclerosis
Central pontine myelinolysis
Cervical cord injury
Peripheral injury
Drugs
Toxins
Hypoparathyroidism
Psychogenic conditions

HEREDODEGENERATIVE DYSTONIAS
X-linked
Lubag disease
Deafness-dystonia-optic atrophy (Mohr-Tranebjaerg) syndrome
Pelizaeus-Merzbacher disease
Lesch-Nyhan syndrome
Autosomal dominant
Rapid-onset dystonia-parkinsonism
Juvenile parkinsonism (e.g., from mutations in the *parkin* gene)
Huntington disease
Machado-Joseph disease (SCA3) and other SCAs
Dentatorubropallidoluysian atrophy
Autosomal recessive
Wilson disease
Niemann-Pick disease type C
GM_1 gangliosidosis
GM_2 gangliosidosis
Metachromatic leukodystrophy
Homocystinuria
Glutaric acidemia
Triose-phosphate isomerase deficiency
Hartnup disease
Ataxia-telangiectasia
Neurodegeneration with brain iron accumulation (NBIA 1) (previously Hallervorden Spatz disease)
Juvenile neuronal ceroid lipofuscinosis
Neuroacanthocytosis
Intranuclear hyaline inclusion disease
Hereditary spastic paraplegia with dystonia
Probably autosomal recessive
Familial basal ganglia calcifications (also dominantly inherited)
Progressive pallidal degeneration
Rett syndrome
Mitochondrial
Leigh disease
Leber disease
Other mitochondrial cytopathies
Sporadic, with parkinsonism
Parkinson disease
Progressive supranuclear palsy
Multiple system atrophy
Corticobasal degeneration

Modified from Cloutier M, Lang AE. Movement disorders: an overview. In: Factor SA, Lang AE, Weiner WJ, eds. *Drug Induced Movement Disorders*. Malden, Mass: Blackwell; 2005:3-19.

dysphonia), abnormal posturing and tightness of the hand while writing or using the hand for other tasks (writer's cramp, manual dystonia), abnormal posturing of the trunk or pelvis (axial dystonia), or abnormal posturing of the lower limb, including plantar flexion and inversion of the foot (Videos 410-4, Blepharospam; 410-5, Oromandibular Dystonia; 410-6, Cervical Dystonia; 410-7, Writer's Cramp). The movements are often slow and sustained, although they also may be rapid (dystonic spasms). Slower, sinuous writhing dystonic movements, particularly present in the distal limbs, are referred to as *athetosis*. Dystonia is often made worse by activity (action dystonia), and a unique aspect of dystonia is that only selected acts may be affected, with complete sparing of all other activities in the same limb (task-specific dystonia, including writer's cramp and musician's cramp) (Video 410-8, Embouchure Dystonia). In some patients, dystonia remains isolated and action specific over many years; in others, it progresses to involve adjacent muscles (overflow dystonia) and may eventually be present at rest, in which case joint contractures may result. Another common feature of dystonia is its transient improvement with the use of a sensory trick (geste antagoniste), such as lightly touching the chin to relieve severe cervical dystonia or the lid to relieve disabling blepharospasm (Video 410-9, Sensory Trick in Cervical Dystonia). Patients with dystonia, independent of cause, often have additional postural and action tremors, phenotypically similar to those in essential tremor. Some patients also demonstrate more irregular, coarse, lower frequency rhythmic movements called *dystonic tremor*.

Dystonia is often classified according to the site of involvement: focal, only one body part (e.g., blepharospasm, cervical dystonia, writer's cramp); segmental, two or more contiguous body parts; multifocal, two or more noncontiguous body parts; generalized, trunk and at least two other sites (with or without leg involvement); and hemidystonia, unilateral (generally a causative focal brain lesion is found most often involving the putamen).

DIAGNOSIS AND PROGNOSIS

For diagnostic and prognostic purposes, dystonia also may be distinguished by age of onset as childhood-onset, adolescent-onset, or adult-onset dystonia. The younger the age of onset, the more likely a cause can be defined. Conversely, isolated dystonia beginning in adult life is most often an idiopathic disorder; further investigations are typically unrewarding and are usually not indicated. Likewise, independent of the cause, dystonia beginning in childhood commonly progresses to segmental or generalized involvement whereas adult-onset dystonia usually remains focal or segmental.

Specific Dystonias

PRIMARY (IDIOPATHIC) OR ISOLATED DYSTONIAS

Primary dystonia accounts for up to 90% of patients with a pure dystonic syndrome, in which dystonia either is the only motor feature or is accompanied only by tremor. To date, no consistent neuropathologic changes have been found in the small numbers of brains affected by primary dystonia that have been studied.

When symptoms begin in childhood, a definable genetic cause is often identified, the most common being DYT1, usually resulting from the autosomal dominant inheritance of a GAG deletion in the *torsin A* gene (Oppenheim dystonia). This disorder is more common in persons of Ashkenazi Jewish descent. The dystonia often begins in the first decade of life and can progress to severe disability, although the spectrum of disease, even within the same family, can be quite varied and penetrance is relatively low (~40%) (Video 410-10, Generalized Dystonia). Other genetic forms of dystonia include *THAP1* mutations for DYT6 and *TUBB4A* mutations for DYT4 or "whispering dystonia." Genetic testing is available but in the case of DYT1 is recommended only when the age of onset in the patient or another affected family member is less than 26 years.

ADULT-ONSET IDIOPATHIC DYSTONIA

Adult-onset idiopathic dystonia is the most common type of dystonia seen in general neurologic practice. The dystonia typically begins in the face, neck, or arm and may remain focal and nonprogressive or spread only to contiguous muscles after many years.[13] The cause of this disorder is not known, although a positive family history may be noted if multiple family members can be examined. Genetic forms of adult-onset focal or segmental dystonia include *ANO3* and *GNAL* mutations for craniocervical dystonia and possibly *CIZ1* mutations for cervical dystonia.

DYSTONIA-PLUS

The term *dystonia-plus* refers to a small number of disorders characterized by dystonia with other neurologic signs that result from a known or presumed genetic defect without an underlying progressive neurodegenerative process. In the newer classification, these conditions are included in the group of disorders with dystonia combined with other neurologic features.

Dopa-responsive dystonia, which usually results in dystonia beginning in the first decade of life, most often in the lower limbs, sometimes can be mistaken for hereditary spastic paraplegia or cerebral palsy. Most patients with dopa-responsive dystonia have a mutation in the *GCH1* gene, which results in reduced production of dopamine. Approximately 75% of patients have notable worsening of dystonia as the day progresses (diurnal variation). Exercise often aggravates the dystonia. Patients commonly demonstrate some degree of bradykinesia (especially in the legs) and postural instability. Rare adult-onset disease may result in a pure parkinsonian phenotype. Dopa-responsive dystonia should be considered in all children with dystonia. Symptoms are exquisitely sensitive to low doses of levodopa (typically as little as 50 mg/day of levodopa), and this treatment allows patients to live a normal life without the usual complications seen in Parkinson disease (Chapter 409).

Myoclonus dystonia, which usually begins within the first decade of life, combines dystonia with separate multifocal myoclonic jerks. Myoclonus dystonia is genetically heterogeneous; the most common definable cause is a mutation in the *e-sarcoglycan* gene. The dystonia in these patients most often involves the neck or upper limbs, is mild, and is often overlooked. The disorder also can include psychopathology, such as obsessive-compulsive behavior. A characteristic feature of this disorder is the marked ameliorative effect of ethanol on both the myoclonus and the dystonia, a feature that sometimes results in alcohol abuse.

OTHER DYSTONIAS

Dystonia may be a symptom of many diseases. The nature and extent of the investigations undertaken depend on such factors as age at onset, clues provided on the history, and additional neurologic or systemic features on examination. Wilson disease (Chapter 211) is an important consideration in the diagnosis of dystonia beginning in children and young adults. Another potentially treatable form of dystonia caused by a mutation in the *SLC30A10* gene, which codes for a manganese transporter.[14] The phenotype is similar to that of Wilson disease, with generalized dystonia, cirrhosis, and hyperintensities in the basal ganglia on T1 MRI scans.

Some patients with dystonia, chorea, or a mixture of the two (choreoathetosis) have intermittent symptoms (paroxysmal dyskinesias) and may be normal between episodes. The duration of symptoms can be as brief as a few seconds to a few minutes or persist for several hours. Symptoms triggered by sudden movement, which are termed *kinesigenic,* are typically brief; prolonged episodes are commonly triggered by exercise, stress, fatigue, caffeine, or alcohol. Paroxysmal dyskinesias may be genetically determined, idiopathic, the manifestation of another disorder (e.g., head injury, brain tumor, or stroke), or even psychologically based. A mutation of the *PRRT2* gene has been described in a large proportion of genetically determined paroxysmal kinesigenic dystonia.

TREATMENT

Rx

Ideally, treatment is directed at the underlying cause, such as dopa-responsive dystonia, which is treated with levodopa (usually up to 300 mg/day) or Wilson disease (Chapter 211). Patients with mutations in the manganese transporter gene may benefit from chelation with ethylenediaminetetraacetic acid (EDTA) (see Table 22-1). Unfortunately, cause-specific treatment usually is not possible, so a variety of symptomatic treatments may be tried, often unsuccessfully, in an attempt to reduce disability.

Focal injections of botulinum toxin are now usually the first choice for treatment of focal and segmental dystonias. This approach can improve symptoms of patients with cranial (blepharospasm, oromandibular dystonia) and cervical dystonia. Patients with task-specific limb dystonias (e.g., writer's cramp) often benefit less because weakness of the treated muscles, which is the most common side effect of this therapy, can impair other important upper limb functions.

Young patients in particular can tolerate and benefit from high doses of anticholinergic drugs such as trihexyphenidyl (6 to 40 mg/day, but sometimes as much as 100 mg/day). Muscle relaxants, including benzodiazepines (diazepam, 5 to as much as 100 mg/day) and baclofen (40 to 120 mg/day), may provide some benefit. Dopamine-depleting (e.g., tetrabenazine, 50 to 200 mg/day) and dopamine-blocking (e.g., haloperidol, 3 to 30 mg/day) agents are occasionally helpful (more often effective in tardive dystonia than in other types). Paroxysmal kinesigenic dyskinesia usually responds well to anticonvulsant drugs. Paroxysmal exercise-induced dyskinesias are associated with mutations in the *SLC2A1* gene, which encodes the glucose transporter GLUT1, and may respond to a ketogenic diet. Neurosurgical treatments, particularly deep brain stimulation of the internal segment of the globus pallidus, can be considered in medically refractory, disabling dystonia, especially in patients with idiopathic dystonia (e.g., DYT1, adult-onset cervical dystonia).[15]

TICS

EPIDEMIOLOGY AND PATHOBIOLOGY

Tics are repetitive, stereotyped movements (motor tics) or vocalizations (vocal tics). Transient tics are extremely common in childhood, and simple tics may begin in childhood and persist throughout adult life.[16] Most tics (Table 410-6) are primary or idiopathic and have no identifiable cause. Secondary tics are caused by a defined underlying brain disease or environmental factor.

CLINICAL MANIFESTATIONS

Tics vary in terms of complexity, from abrupt, brief, meaningless movements or sounds (simple motor tics such as eye blinking, nose wrinkling, or head jerking; simple vocal-phonic tics such as sniffing, throat clearing, or grunting) to more sustained, more deliberate, almost meaningful gestures or utterances (complex motor tics such as touching, hand shaking, and jumping; complex vocal tics such as echolalia [repeating others], palilalia [repeating oneself], and coprolalia [uttering profanities]). The frequency of the tics in an individual patient varies markedly over minutes, hours, days, weeks, and years.

DIAGNOSIS

Various characteristics help to differentiate tics from other abnormal movements. Tics are often described by patients as being "semivoluntary" in response to an inner, irresistible urge. Premonitory sensory symptoms occasionally precede the tic, usually in the same general anatomic area as the tic itself. Relief is often associated with the production of the tic. Tics can be partially or completely voluntarily suppressed for variable periods, but often at the expense of mounting inner tension and psychological discomfort. Performing the tic or sometimes even substituting another more acceptable behavior for the socially inappropriate tic alleviates the tension. Many patients report that some tics occur in response to a typical urge, whereas the same or different tics may be unexpected and totally involuntary.

Tourette Syndrome

EPIDEMIOLOGY AND PATHOBIOLOGY

The exact relationship between childhood tics and Gilles de la Tourette syndrome remains uncertain. Tourette's syndrome is a common disorder, with an overall prevalence of 7.7 per 1000 children. There is a male preponderance of 3 : 1 for the classic syndrome, but female patients manifest obsessive-compulsive features more often than tics. A functional mutation in the *HDC* gene encoding L-histidine decarboxylase can be a rare cause of Tourette syndrome, thereby suggesting a role for histaminic neurotransmission in its pathogenesis (Video 410-11, Tics).

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The criteria for this disorder include the presence of multiple motor and at least one vocal tic beginning before the age of 21 years (typically between ages 2 and 10 years) and lasting for more than 1 year, waxing and waning symptoms over time (new tics replacing old ones; previous tics sometimes recurring years after they had originally resolved), and the absence of other

TABLE 410-6 ETIOLOGIC CLASSIFICATION OF TICS

PRIMARY OR IDIOPATHIC TICS
Transient motor or phonic tics
Chronic motor or phonic tics
Adult-onset tics
Tourette syndrome
SECONDARY TICS
Genetic disorders
Neuroacanthocytosis
Huntington disease
Neurodegeneration with brain iron accumulation 1 (NBIA 1) (previously Hallervorden-Spatz disease)
Idiopathic dystonia*
Tuberous sclerosis*
Chromosomal disorders
Infections
Sydenham chorea
PANDAS†
Encephalitis and postencephalitic disorders
Creutzfeldt-Jakob disease
Neurosyphilis
Drugs
Methylphenidate
Amphetamines
Cocaine
Levodopa
Carbamazepine
Phenytoin
Phenobarbital
Lamotrigine
Neuroleptics
Developmental disorders
Mental retardation
Pervasive developmental disorders/autism
Other causes
Head trauma
Stroke
Carbon monoxide poisoning
Cardiopulmonary bypass with hypothermia
RELATED DISORDERS
Mannerisms, stereotypies
Compulsions
Self-injurious behavior

Modified from Cloutier M, Lang AE. Movement disorders: an overview. In: Factor SA, Lang AE, Weiner WJ, eds. *Drug Induced Movement Disorders*. Malden, Mass: Blackwell; 2005:3-19.
*Tics have been described with these conditions but may simply be coincidental.
†Pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections. The existence of this disorder remains somewhat controversial.

explanatory medical conditions. Involuntary swearing (coprolalia), a highly publicized feature of the syndrome, is present in fewer than 10% of patients and is usually manifested by aborted forms such as "fu" and "shi." Patients commonly exhibit a variety of comorbid disorders including obsessive-compulsive disorder, attention-deficit disorder (with or without hyperactivity), impulse control problems, and other behavioral disturbances.

TREATMENT

Most patients who fulfill diagnostic criteria for Tourette syndrome have mild symptoms that do not require drug treatment; education, reassurance, behavioral therapy, and follow-up are often sufficient.[17] When tics (isolated or as part of Tourette syndrome) interfere with social and physical function, low-dose clonazepam (0.5 to 4 mg/day) may be effective. Clonidine (0.05 to 0.5 mg/day) is variably effective in controlling tics and may be useful for impulse control and symptoms of attention-deficit/hyperactivity disorder (ADHD); alternatively, guanfacine (0.5 to 4 mg/day) can be used. The most effective treatments for disabling tics are the dopamine receptor blockers such as risperidone (0.5 to 16 mg/day), haloperidol (0.5 to 20 mg/day), pimozide (0.5 to 10 mg/day), fluphenazine (0.5 to 20 mg/day), and aripiprazole (5 to 15 mg/day), but caution is required in view of the potential for important side effects, including tardive dyskinesia, with long-term use. An alternative without this complication is the dopamine depleter tetrabenazine (50 to 200 mg/day). Injected botulinum toxin may be effective for simple motor tics of the face and neck and may also reduce the urge to perform the tic. More aggressive use of botulinum toxin in neck muscles should be considered in patients with very forceful neck tics, which have rarely been associated with complications such as noncompressive myelopathy and vertebral artery dissection. Comorbid ADHD can be treated safely with stimulant therapy (e.g., methylphenidate, 2.5 to 60 mg/day) without increasing the severity of tics. Obsessive-compulsive symptoms may respond well to selective serotonin reuptake inhibitors (e.g., clomipramine, 25 to 250 mg/day; paroxetine, 10 to 60 mg/day; or citalopram, 10 to 40 mg/day). Behavioral disorders, which remain a major therapeutic challenge, may require a variety of psychotherapeutic or behavioral modification approaches. Even in the absence of behavioral disturbances, comprehensive behavioral intervention, which incorporates habit reversal training, can be very effective as first-line therapy for tic disorders. For this approach to be generalized, however, more trained providers will be needed. Promising preliminary reports of deep brain stimulation require confirmation in controlled clinical trials.

PROGNOSIS

The natural history of Tourette syndrome is to stabilize and often improve in adolescence. Approximately 50% of patients have a complete or partial remission at this time.

MYOCLONUS

DEFINITION

Myoclonus (or myoclonic jerks) consists of sudden, brief, shocklike, involuntary movements that result from both active muscle contraction (positive myoclonic jerks) and brief inhibition of ongoing muscle activity (negative myoclonic jerks). The most common form of negative myoclonic jerk is asterixis.

PATHOBIOLOGY

Myoclonus generally arises in the central nervous system, although rare peripheral causes are described, and it is distinct from abnormal muscle activity associated with peripheral nervous system diseases, such as fasciculations or myokymia. Myoclonus can be classified according to origin (Table 410-7), including physiologic, essential, epileptic, and symptomatic forms. Physiologic myoclonus, such as hypnic (sleep) jerks and hiccups, occurs in normal healthy subjects. Patients with essential myoclonus, which may be sporadic or inherited, often have additional postural tremor or dystonia, and this disorder is probably the same as what is now referred to as myoclonus dystonia (see Dystonias, earlier). Epileptic myoclonus arises in the context of seizures (Chapter 403), including many inherited generalized epileptic syndromes and the progressive myoclonic epilepsies. Symptomatic myoclonus occurs in association with a large number of encephalopathic states.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Myoclonic jerks are very short, typically lasting less than 150 msec. Myoclonus can be spontaneous, action induced, reflex (induced by various sensory stimuli), or a combination. Spontaneous myoclonus occurs at rest, without any provocation. Action myoclonus occurs during purposeful movement and is often very disabling owing to its interference with volitional activity. Reflex myoclonus can be triggered by visual, auditory, or somesthetic stimuli. The distribution of myoclonus may be focal, segmental, multifocal, or generalized. When myoclonus involves more than one body area, the movements may be synchronous or asynchronous. Myoclonus can be intermittent or repetitive, and it sometimes is rhythmic (e.g., usually originating in the brain stem or spinal cord). Palatal myoclonus, now referred to as palatal tremor, is a rhythmic movement disorder originating in the brain stem and involving the soft palate as well as the eyes, facial muscles, neck, and limbs; it is commonly the result of a focal lesion (e.g., stroke, demyelination) in the connections between the dentate nucleus of the cerebellum and the inferior olives of the medulla (symptomatic palatal tremor).

DIAGNOSIS

Myoclonus can be classified according to the anatomic site of origin, usually with the assistance of detailed electrophysiologic assessments.[18] These sites may be cortical, subcortical (e.g., thalamus; lower brain stem [reticular myoclonus]), or spinal (two types: spinal segmental and propriospinal).

TREATMENT

Management of myoclonus, when possible, should be directed specifically at the underlying cause. Drug treatment includes a variety of anticonvulsant

TABLE 410-7 CLASSIFICATION AND CAUSES OF MYOCLONUS

PHYSIOLOGIC MYOCLONUS

Sleep myoclonus
Anxiety-induced myoclonus
Exercise-induced myoclonus
Hiccups
Benign infantile myoclonus during feeding

ESSENTIAL MYOCLONUS

Essential myoclonus*
  Hereditary
  Sporadic
Myoclonus dystonia*

EPILEPTIC MYOCLONUS

Fragments of epilepsy
  Isolated epileptic myoclonic jerks
  Photosensitive myoclonus
  Myoclonic absences
  Epilepsia partialis continua
  Idiopathic stimulus-sensitive myoclonus
Childhood myoclonic epilepsies
  Infantile spasms
  Lennox-Gastaut syndrome
  Cryptogenic myoclonus epilepsy
  Juvenile myoclonic epilepsy of Janz
Benign familial myoclonic epilepsy
Baltic myoclonus (Unverricht-Lundborg)

SYMPTOMATIC MYOCLONUS

Storage disease
  Lafora body disease
  Lipidoses
  Neuronal ceroid lipofuscinosis
  Sialidosis
Spinocerebellar degeneration
  Friedreich ataxia
  Ataxia-telangiectasia
  Other spinocerebellar degenerations
Basal ganglia degenerations
  Wilson disease
  Idiopathic torsion dystonia
  Neurodegeneration with brain iron accumulation 1 (NBIA 1) (Hallervorden-Spatz disease)
  Progressive supranuclear palsy
  Corticobasal degeneration
  Parkinson disease
  Multiple system atrophy
  Huntington disease
  Dentatorubropallidoluysian atrophy
Mitochondrial cytopathies
Dementias
  Alzheimer disease
  Creutzfeldt-Jakob disease
  Dementia with Lewy bodies
  Frontotemproal lobar degeneration with TDP-43 positive inclusions
Viral encephalopathies
  Subacute sclerosing panencephalitis
  Encephalitis lethargica
  Herpes simplex encephalitis
  Arbovirus encephalitis
  Human immunodeficiency virus infection
  Postinfectious encephalitis
Metabolic disorders
  Hepatic failure
  Renal failure
  Dialysis dysequilibrium syndrome
  Hyponatremia
  Hypoglycemia
  Nonketotic hyperglycemia
  Infantile myoclonic encephalopathy
  Multiple carboxylase deficiency
  Biotin deficiency
Toxins
  Bismuth
  Heavy-metal poisoning
  Methylbromide, dichlorodiphenyltrichloroethane (DDT)
  Drugs (multiple)
Physical encephalopathies
  Posthypoxic myoclonus (Lance-Adams)
  Post-traumatic status
  Heat stroke
  Electric shock
  Decompression injury
Focal central nervous system damage
  Stroke
  Post-thalamotomy status
  Tumor
  Trauma
  Spinal cord lesions
Peripheral myoclonus (lesions of peripheral nerve, plexus, or nerve root)
Whipple disease
Paraneoplastic syndromes
Psychogenic myoclonus

Modified from Cloutier M, Lang AE. Movement disorders: an overview. In: Factor SA, Lang AE, Weiner WJ, eds. *Drug Induced Movement Disorders*. Malden, Mass: Blackwell; 2005:3-19.
*Probably represents the same entity.

medications, most notably clonazepam (1.5 to 15 mg/day), valproic acid (10 to 15 mg/kg/day), carbamazepine (600 to 1200 mg/day), and levetiracetam (1000 to 4000 mg/day). Lacosamide (200 to 400 mg/day) is also effective in select patients. Postanoxic action myoclonus (the Lance-Adams syndrome) in some patients who survive severe cerebral anoxia also may respond to 5-hydroxytryptophan (400 to 2800 mg/day) given with carbidopa (75 to 300 mg/day). Acetazolamide (250 to 1000 mg/day) may be useful for patients with action myoclonus.

HYPEREXPLEXIA

Hyperexplexia, which is a disorder related to myoclonus, manifests as an excessive startle response to tactile, visual, and/or auditory stimulations. Genetic causes are mainly abnormalities in synaptic transmission of the inhibitory neurotransmitter glycine, including glycine receptor α_1 gene (*GLRA1*), glycine receptor subunit gene (*GLRB*), and the presynaptic glycine transporter 2 gene *SLC6A5*. Some patients demonstrate only generalized body jerking or an exaggerated startle response that habituates poorly after repeated stimuli. By comparison, other patients experience disabling stiffness in response to sudden unexpected stimuli, such as loud sound. The disorder typically responds well to clonazepam (1.5 to 15 mg/day) therapy. Additional medications that have been tried with mixed results include clobazam, levetiracetam, valproic acid, and phenobarbital.

OTHER MOVEMENT DISORDERS

Drug-Induced Movement Disorders

All the movements listed in Table 410-1 can be induced by medications. Neuroleptic drugs, which block postsynaptic dopamine receptors, particularly the D2 subtype, can result in a variety of movement disorder syndromes, including acute dystonic reactions, akathisia, drug-induced parkinsonism (including "the rabbit syndrome" with perinasal and perioral rest tremor), the neuroleptic malignant syndrome, and a variety of later-onset, often persistent, movements referred to as tardive dyskinesia.

ACUTE DYSTONIC REACTIONS

Acute dystonic reactions (Chapter 434) are most often seen in young patients who are receiving potent antipsychotic agents (e.g., young male patients receiving high doses of haloperidol for acute psychosis), but they also occur in patients receiving dopamine receptor blockers, including metoclopramide as antiemetic therapy. Symptoms range from overt dystonic postures of the face and neck, to involuntary prolonged deviation of the eyes (oculogyric crises), to simple slurring of speech and difficulty coordinating the tongue.

Symptoms often vary from moment to moment and can increase with anxiety and improve with relaxation or reassurance. Acute dystonic reactions are self-limited and respond rapidly to a parenteral injection of an anticholinergic drug such as benztropine (2 mg intravenously [IV] followed by 2 mg three orally [PO] times daily for a variable duration depending on neuroleptic use) or an antihistaminic such as diphenhydramine (50 mg IV followed by oral benztropine).

AKATHISIA

Akathisia refers to a sense of restlessness and a need to move. Typically, the patient performs a variety of purposeful or semipurposeful, often complex, movements in response to an uncomfortable subjective restlessness, including pacing when standing, marching in place, rocking, shifting weight, moving legs when sitting, picking at clothing or hair, rubbing body parts with hands, and other similar movements. Akathisia is most often a side effect of medications, especially neuroleptic drugs and selective serotonin reuptake inhibitors (Chapter 397). Symptoms occur in a dose-related fashion and usually resolve on drug withdrawal. Akathisia is a common reason for psychiatric patients to comply poorly with their medications; management includes adjustment of the dose or type of antipsychotic agent and trials of β-blockers (e.g., propranolol, 80 mg/day) or antiparkinson agents, such as anticholinergics (e.g., benztropine (6 mg/day) or amantadine (200 to 300 mg/day). Rare patients experience a very disabling and persistent form referred to as tardive akathisia. Akathisia is also sometimes seen in patients with Parkinson disease.

NEUROLEPTIC MALIGNANT SYNDROME

The neuroleptic malignant syndrome (Chapters 432 and 434) is an uncommon but severe, sometimes fatal, complication of neuroleptic therapy. Patients usually manifest a combination of features including fever, marked rigidity, changes in level of arousal, and autonomic instability. Laboratory abnormalities include a marked increase in the serum creatine kinase level and the blood leukocyte count. Management involves early recognition, withdrawal of the causative agent, systemic supportive therapy, a dopamine agonist (most experience has been with the older agent bromocriptine, ≤ 60 mg/day), and, when necessary, dantrolene sodium (50 to 600 mg/day PO or ≤ 10 mg/kg/day IV) to reduce muscle contraction.

TARDIVE DYSKINESIA

Epidemiology and Pathobiology

The term *tardive dyskinesia* encompasses a wide variety of abnormal movements caused by chronic neuroleptic therapy (Chapter 434). The cumulative 5-year incidence rate in patients taking classic neuroleptics is approximately 25%, and the incidence may continue to increase almost linearly beyond that point. The annualized risk is estimated to be 5% in haloperidol-treated patients compared with 2% in patients treated with atypical neuroleptics (Video 410-12, Tardive Dyskinesia). The pathophysiology commonly has been attributed to hypersensitivity or upregulation of dopamine D2 receptors induced by chronic blockade. However, this explanation is generally felt to be inadequate, especially for more persistent symptoms, and other proposed mechanisms include oxidative stress from increased dopamine turnover and a maladaptive synaptic plasticity.

CLINICAL MANIFESTATIONS

Tardive dyskinesia generally begins after a minimum of 6 weeks of treatment. One of the most common forms involves the lower facial muscles and has been given a variety of names, including orobuccolinguomasticatory dyskinesia. The movements generally include repetitive chewing and smacking movements with the tongue either protruding between the lips (fly-catching movements) or pushing into the cheek (bonbon sign). Although the movements are somewhat choreic, they are not as random as true chorea. The more stereotypical, repetitive nature of the movements, involving not only face but also the limbs (e.g., piano playing movements of the fingers, rocking or thrusting of the pelvis), has encouraged the more recent term tardive stereotypies. This term, however, fails to fulfill the definition of stereotypy owing to the lack of distractibility and the unpredictability of the sequence of movements. Many patients with classic orofacial tardive dyskinesia seem unaware of the presence of the movements and are not disabled by them, but others are embarrassed or otherwise impaired.

Tardive akathisia and tardive dystonia are less common but particularly disabling subtypes of tardive dyskinesia. Rarer forms include tardive tics (tourettism), tardive tremor, tardive myoclonus, and even tardive oral or genital pain.

TREATMENT AND PROGNOSIS

Rx

Treatment is often unsatisfactory, but the dopamine depleter tetrabenazine (50 to 200 mg/day) can be very effective. Other drugs that possibly provide benefit include amantadine, propranolol, zolpidem, ginkgo biloba, and clonazepam.[19] Consideration should be given to discontinuing concurrent anticholinergic medications. Prevention is the most important consideration. The physician must regularly reassess the need for ongoing neuroleptic therapy, consider switching to an atypical agent when possible (particularly quetiapine and clozapine; Chapter 397), and routinely evaluate the patient for the presence of early subtle clinical features, such as mild pursing of the lips or rolling movements of the tongue in the mouth. Unfortunately, tardive dyskinesia may persist for many years despite withdrawal of neuroleptic treatment in up to 50% of patients. Several atypical neuroleptics, such as risperidone and olanzapine, nevertheless block dopamine D2 receptors sufficiently to cause drug-induced parkinsonism and tardive dyskinesias.

Restless Legs Syndrome

EPIDEMIOLOGY

Restless legs syndrome (Chapter 405) is now recognized as an extremely common disorder affecting between 3 and 29% of the general population. Women are affected more frequently than men. Although the incidence increases with age, it also can affect children, in whom it may be confused with "growing pains" or ADHD.

PATHOBIOLOGY

Restless legs syndrome is most often primary or idiopathic, in which case it is frequently inherited in an autosomal dominant fashion. Eight genetic loci associated with restless legs syndrome include variants in *MEIS1, BTBD9, MAP2K5/LBXCOR1, PTPRD,* and *PCDHA3,* as well as loci on 2p14 and 16q12.1. Restless legs syndrome also may be secondary to other causes, including peripheral neuropathy, uremia, pregnancy, and iron deficiency, and it may occur more commonly than by chance in some neurodegenerative disorders such as Parkinson disease. The pathophysiology of restless legs syndrome is uncertain, but central iron dysregulation may somehow alter central dopamine. Serum ferritin levels are often low, even in the presence of normal values of hemoglobin, hematocrit, iron, and iron-binding capacity.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

In restless legs syndrome, as in akathisia, movements occur because of the subjective need to move. However, unlike in akathisia, the patient typically complains of a variety of sensory disturbances in the legs, including pins and needles, creeping or crawling sensations, aching, itching, stabbing, heaviness, tension, burning, or coldness. Occasionally, similar symptoms are appreciated in the upper limbs or other areas of the body. These symptoms are usually experienced during periods of prolonged inactivity, especially with recumbency in the evening, and are often associated with insomnia (Chapter 405). The discomfort appears particularly during the transition from wake to sleep in the evening and often follows a circadian pattern, peaking between midnight and 4 AM. Symptoms are typically relieved only by movement or stimulation of the legs; although these maneuvers are effective while they are being performed, the discomfort usually returns as soon as the individual becomes inactive or returns to bed to try to sleep. Patients often have significant problems with immobility during long automobile drives or plane flights.

In approximately 80% of patients, this condition is associated with another movement disorder, periodic leg movements in sleep, sometimes inappropriately called nocturnal myoclonus. These periodic, slow, sustained (1 to 2 seconds) movements range from synchronous or asynchronous dorsiflexion of the toes and feet to triple flexion of one or both legs. In 15% of patients, more rapid myoclonic movements or slower, prolonged dystonic-like movements of the feet and legs are present while patients are awake. In the absence of evidence of a secondary cause of restless legs syndrome, the only useful routine test is a serum ferritin level.

TREATMENT

Rx

Dopamine agonists (e.g., pramipexole, 0.125 to 1.5 mg at bedtime), ropinirole (0.25 to 3 mg at bedtime), and transdermal rotigotine (1 to 3 mg/24 hours) are the treatments of choice in moderate-to-severe restless legs syndrome and can be very effective. Levodopa preparations (100 to 300 mg of levodopa at

bedtime; consider controlled-release preparation) are also effective but are more often associated with disabling rebound symptoms early in the morning or during the day (augmentation). Gabapentin enacarbil (a gabapentin prodrug at 600 mg to 1200 mg/day) is also effective and approved in the United States at the 600-mg dose. Patients with milder symptoms may respond to gabapentin (300 to 2400 mg/day). Opiate agonists (e.g., oxycodone, 5 mg at bedtime; codeine, 30 mg at bedtime; propoxyphene, 65 mg or N-100 mg at bedtime) and less often benzodiazepines (e.g., clonazepam, 0.5 to 2 mg at bedtime) also may be effective. Tolerance or loss of original benefit may occur with all these treatments. Iron replacement is indicated in patients with reduced serum ferritin levels (325 mg ferrous sulfate two or three times per day for 3 to 4 months until ferritin levels exceed 50 mg/L and iron saturations exceed 20%).

Painful Legs and Moving Toes

Another uncommon but well-defined movement disorder of the lower limbs has been termed *painful legs and moving toes.* Patients typically complain of a deep pulling or searing pain in the lower limbs, associated with continuous involuntary wriggling or writhing of the toes. Occasionally, the ankle and less commonly more proximal muscles of the legs are involved. Rarely, a similar problem is seen in the upper limbs as well. Although a peripheral nerve trigger, such as a radiculopathy, may be evident, the pain and movements probably are generated centrally in the spinal cord or brain stem. Various treatments have been tried without much benefit to the pain, which is typically the major concern of the patient.

Other Abnormal Movements

Numerous abnormal movements are caused by dysfunction of the peripheral nerves (e.g., fasciculations, myokymia); these movements are usually easily separated from the movement disorders described earlier. *Hemifacial spasm* is a common disorder in which irregular clonic and tonic movements involve the muscles innervated by the facial nerve, usually owing to compression of the seventh nerve as it exits the brain stem, most often by a normal small artery or vein and less often by a mass lesion or inflammatory process. Eyelid twitching is usually the first symptom, followed at variable intervals by lower facial muscle involvement. Magnetic resonance imaging (MRI) with careful assessment of the posterior fossa is necessary to exclude secondary causes. Treatment usually involves injections of botulinum toxin into selected facial muscles, although surgical decompression can be curative (Video 410-13, Hemifacial Spasm).

Cerebellar Ataxias and Spastic Paraplegias

There are an extremely large number of causes of cerebellar ataxia (Table 410-8). Many are hereditary, with the full spectrum of possible inheritance patterns. Sporadic or noninherited ataxias are common; in many cases, a cause can be defined and treatment may be effective in halting or even reversing the process. However, a large proportion of ataxias in adults are progressive, presumably owing to a degenerative cause, many of which remain to be determined.

TABLE 410-8 DIFFERENTIAL DIAGNOSIS OF ADULT-ONSET ATAXIA

Inherited
Autosomal dominant, including the spinocerebellar ataxias (SCAs)
Autosomal recessive, including Friedreich ataxia
X-linked, including fragile X tremor ataxia syndrome (FXTAS)
Mitochondrial
Episodic ataxias
Autoimmune (e.g., paraneoplastic, anti-GAD antibodies, postinfectious)
Degenerative (e.g., multiple system atrophy [MSA-C])
Demyelinating (e.g., multiple sclerosis)
Infectious
Metabolic (e.g., hypothyroidism, vitamin E deficiency)
Stroke
Trauma (e.g., closed head injury)
Toxic (e.g., alcoholic cerebellar degeneration, lithium)
Tumor: Primary and secondary brain tumors

GAD = glutamate decarboxylase.

HEREDITARY CEREBELLAR ATAXIAS

The hereditary cerebellar ataxias, which may begin in childhood or adulthood, can progress at widely varying rates. These ataxias are divided into early-onset ataxias, which are usually inherited as autosomal recessive disorders,[20] and adult-onset ataxias, which are usually autosomal dominant. A small number are X-linked. Because most of these ataxias are untreatable, it is important to recognize the rare causes of treatable or preventable progressive ataxias.

Friedreich Ataxia

EPIDEMIOLOGY AND PATHOBIOLOGY

The most common progressive inherited ataxia in children is Friedreich ataxia. Friedreich ataxia is a trinucleotide-repeat disorder that affects the central and peripheral nervous systems, the heart, and many other organs. Friedreich ataxia is an autosomal recessive disorder with no anticipation. It has an estimated carrier frequency in the population of approximately 1 in 100 and a resulting disease prevalence of approximately 1 per 50,000.

The normal length of the GAA repeat on the long arm of chromosome 9 (9q13-q21) is 10 to 21 copies, but expansion in individuals with Friedreich ataxia results in 200 to 900 copies and disrupts the expression of the protein frataxin. GAA unstable expansion, which occurs on an intron, leads to gene silencing rather than to the production of an abnormal protein. Higher numbers of copies correlate with more severe neurologic deficits. Frataxin appears to be critical for iron export and mitochondrial function. Because accumulation of mitochondrial iron affects the production of oxygen radicals, loss of frataxin may lead to oxidative mitochondrial damage. The pathology of Friedreich ataxia includes spinal cord atrophy, which often is evident on MRI, with loss of neurons in Clarke columns and the dorsal root ganglia. Degeneration occurs in spinocerebellar tracts, pyramidal tracts, dorsal column tracts, and peripheral nerves, with minor cell loss in the brain stem and cerebellum. Cardiomyopathy is associated with ventricular hypertrophy and chronic interstitial myocardial fibrosis.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Typical Friedreich ataxia first manifests clinically during puberty with progressive ataxia, loss of lower extremity deep tendon reflexes, and extensor plantar responses (i.e., Babinski signs). Other common clinical features include nystagmus, dysarthria, stocking-glove sensory loss, and weakness in the lower extremities. Patients frequently have kyphosis, scoliosis, and pes cavus. Interstitial myocardial disease may cause a typical hypertrophic cardiomyopathy (Chapter 60). A small number of patients have a later onset and less severely progressive course, sometimes with retained or even brisk reflexes.

The diagnosis is made by genetic testing for the trinucleotide repeat expansion, which usually is present on at least one allele. Point mutations are sometimes present in the other allele and are more difficult to detect. Potentially treatable conditions with similar clinical manifestations include vitamin B_{12} deficiency (Chapter 218), abetalipoproteinemia (Chapter 140), and a selective defect in vitamin E absorption (Chapter 218).

TREATMENT AND PROGNOSIS

Rx

No effective disease-modifying treatments are available, so treatment consists of supportive measures. Intensive inpatient rehabilitation can improve overall function. Nicotinamide can increase frataxin concentrations,[21] but whether it alters the clinical course of the disease is unproven. Future treatments may include histone deacetylase inhibitors that may increase frataxin gene expression. The disorder is progressive, and patients usually are wheelchair bound by their mid-20s. The average age at death is 37 years, and the major cause of death is hypertrophic cardiomyopathy (Chapter 60).

Other Spinocerebellar Ataxias

The hereditary spinocerebellar ataxias are routinely classified by their specific molecular diagnosis. At least 20 autosomal recessive and more than 35 autosomal dominant cerebellar ataxias have been identified. Clinical features, ethnic origin, and family history may suggest an autosomal recessive, autosomal dominant, or X-linked inheritance and often narrow the search for the genetic mutation. As the molecular pathogenesis of many of the hereditary ataxias is unraveled, the current numerical classification, which largely reflects

the chronology of the identification of causative mutations, is likely to be replaced by a gene-specific or pathophysiologic approach. Spinocerebellar ataxias 1, 2, 3, 6, 7, and 17 are caused by trinucleotide expansions in or adjacent to a protein-coding region of a gene. These expansions result in polyglutamine expansions in the protein product, which likely results in a toxic gain of function in a manner analogous to the pathogenesis of Huntington disease.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The predominant clinical features of the spinocerebellar ataxias are ataxia and dysarthria. Other cerebellar signs include titubation, dysdiadochokinesia, and dysmetria. With increasing ataxia, patients can become wheelchair bound. Additional clinical signs include ophthalmoplegia, dementia, optic atrophy, retinal pigmentary degeneration, deafness, dysphagia, and peripheral neuropathy. Extrapyramidal features include masked facies, cogwheel rigidity, dystonia, athetosis, and chorea. Levodopa-responsive parkinsonism (Chapter 409) may be seen in some patients, particularly in spinocerebellar ataxias 2 and 3. Pyramidal dysfunction includes spastic limbs, especially legs; hyperreflexia; and Babinski response. Diagnosis is based on genetic testing.

TREATMENT AND PROGNOSIS

Rx

In a small randomized trial, varenicline (a partial agonist of 24β2 neuronal nicotinic acetylcholine receptors, at 1 mg/day for 2 weeks then 2 mg/day) improved gait, stance, and timed 25-foot walk but did not improve appendicular function, except for rapid alternating movements, in adults with genetically confirmed SCA3.[A8] However, these results have not been reproduced, and the drug is often poorly tolerated. No treatment is currently available for the other spinocerebellar ataxias, although preliminary data indicate that physiotherapy may improve gait and balance. The spinocerebellar ataxias are progressive, with worsening gait, hand coordination, speech, and eye movements, but with preserved mental function in most forms. Pneumonia is a common cause of death.

HEREDITARY SPASTIC PARAPLEGIAS

Hereditary spastic paraplegias, also known as Strümpell disease, are a group of clinically and genetically heterogeneous monogenic neurodegenerative disorders.[22] The prevalence is approximately 1 per 10,000 in the population. Over 70 different genetic loci have been identified; approximately 20 autosomal dominant, over 45 autosomal recessive, 5 X-linked, and one a maternal trait of inheritance. The most common forms of hereditary spastic paraplegia are autosomal dominant mutations in one of four proteins: spastin (SPG4), atlastin-1 (SPG3A), *REEP1* (SPG31), and reticulon-2 (SPG12). These proteins are involved in the endoplasmic reticulum network, whose morphology and distribution in neurons has a special importance for their normal function. Defects of ganglioside biosyntheses and defects in glucocerebrosidase functions are present in some forms. At autopsy, patients with hereditary spastic paraplegia have axonal degeneration of the pyramidal tracts and dorsal column tracts with lesser involvement of the spinocerebellar tracts. The neurons of origin are intact. The peripheral nervous system is unaffected.

CLINICAL MANIFESTATIONS

Patients with hereditary spastic paraplegia have a progressive gait disturbance with spasticity of lower extremities, hyperreflexia, clonus, and extensor plantar responses. Cranial nerves, speech, swallowing, and upper extremities remain normal. Although patients can experience weakness of their lower extremities, spasticity is usually the disabling component. The progressively increased leg spasticity results in tripping and an inability to run. Pain is infrequent, and sensation is normal. Other clinical features include pes cavus (30 to 50%), decreased vibratory sensation, and urinary frequency, urgency, and hesitancy. Pure hereditary spastic paraplegia is limited to symptoms and signs of spasticity, whereas complex or complicated hereditary spastic paraplegia can include cognitive impairment, dementia, epilepsy, extrapyramidal disturbances, cerebellar involvement, retinopathy, optic atrophy, deafness, polyneuropathy, or skin lesions.

DIAGNOSIS

Hereditary spastic paraplegia is diagnosed when patients meet clinical criteria and when other causes of spasticity are excluded. MRI may show spinal cord atrophy, but cerebrospinal fluid analysis and nerve conduction studies are normal. The differential diagnosis of spastic paraplegia includes other genetic conditions, spinal cord disease from structural lesions, multiple sclerosis, and vitamin deficiencies or retroviral infections (Table 410-9). Even a positive family history does not obviate the need to exclude potentially treatable alternative diagnoses.

TABLE 410-9 DIFFERENTIAL DIAGNOSIS OF SPASTIC PARAPLEGIAS

- Hereditary
  - Dopa-responsive dystonia
  - Spinocerebellar ataxias
  - Adult-onset adrenoleukodystrophy (Chapters 227 and 411)
- Structural lesions of the spinal cord (Chapter 400)
- Cervical spondylosis (Chapter 400)
- Tumor (Chapter 179)
- Arteriovenous malformation (Chapter 408)
- Syringomyelia (Chapter 417)
- Multiple sclerosis (Chapter 411)
- Primary lateral sclerosis (Chapter 419)
- Vitamin B_{12} deficiency (Chapter 416)
- Copper deficiency (Chapter 416)
- Infections
  - Human immunodeficiency virus (Chapter 394)
  - Human T-lymphotropic virus type 1 (Chapter 378)
  - Tertiary syphilis (Chapter 319)

TREATMENT AND PROGNOSIS

Rx

No specific treatment is available. Symptomatic therapy is aimed at decreasing disability and preventing complications, such as contractures. Antispastic agents, such as oral baclofen (usually 10 to 20 mg three times daily), improve spasticity but should be used with caution because they may worsen weakness. Some reports suggested an improved therapeutic response to intrathecal baclofen, but no controlled trials have addressed this issue. Preliminary data also raise the possible utility of injected botulinum neurotoxin type A injections for improving spasticity. Most patients become nonambulatory between 60 and 70 years of age. Patients with complicated hereditary spastic paraplegia often have other disabling features. Some patients with parkinsonism may benefit from dopaminergic therapies such as levodopa.

Grade A References

A1. Zappia M, Albanese A, Bruno E, et al. Treatment of essential tremor: a systematic review of evidence and recommendations from the Italian Movement Disorders Association. *J Neurol*. 2013;260:714-740.

A2. Armstrong MJ, Miyasaki JM. Evidence-based guideline: pharmacologic treatment of chorea in Huntington disease: report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology*. 2012;79:597-603.

A3. Hallett M, Albanese A, Dressler D, et al. Evidence-based review and assessment of botulinum neurotoxin for the treatment of movement disorders. *Toxicon*. 2013;67:94-114.

A4. Volkmann J, Wolters A, Kupsch A, et al. Pallidal deep brain stimulation in patients with primary generalised or segmental dystonia: 5-year follow-up of a randomised trial. *Lancet Neurol*. 2012;11:1029-1038.

A5. Piacentini J, Woods DW, Scahill L, et al. Behavior therapy for children with Tourette disorder: a randomized controlled trial. *JAMA*. 2010;303:1929-1937.

A6. Wilt TJ, MacDonald R, Ouellette J, et al. Pharmacologic therapy for primary restless legs syndrome: a systematic review and meta-analysis. *JAMA Intern Med*. 2013;173:496-505.

A7. Kume A. Gabapentin enacarbil for the treatment of moderate to severe primary restless legs syndrome (Willis-Ekbom disease): 600 or 1,200 mg dose? *Neuropsychiatr Dis Treat*. 2014;10:249-262.

A8. Zesiewicz TA, Greenstein PE, Sullivan KL, et al. A randomized trial of varenicline (Chantix) for the treatment of spinocerebellar ataxia type 3. *Neurology*. 2012;78:545-550.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Abdo WF, van de Warrenburg BP, Burn DJ, et al. The clinical approach to movement disorders. *Nat Rev Neurol*. 2010;6:29-37.
2. Elias WJ, Shah BB. Tremor. *JAMA*. 2014;311:948-954.
3. Kuhlenbaumer G, Hopfner F, Deuschl G. Genetics of essential tremor: meta-analysis and review. *Neurology*. 2014;82:1000-1007.
4. Zesiewicz TA, Shaw JD, Allison KG, et al. Update on treatment of essential tremor. *Curr Treat Options Neurol*. 2013;15:410-423.
5. Elias WJ, Huss D, Voss T, et al. A pilot study of focused ultrasound thalamotomy for essential tremor. *N Engl J Med*. 2013;369:640-648.
6. Lee JM, Ramos EM, Lee JH, et al. CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. *Neurology*. 2012;78:690-695.
7. Schapira AHV, Olanow CW, Greenamyre JT, et al. Slowing of neurodegeneration in Parkinson's disease and Huntington's disease: future therapeutic perspectives. *Lancet*. 2014;384:545-555.
8. Tabrizi SJ, Scahill RI, Owen G, et al. Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *Lancet Neurol*. 2013;12:637-649.
9. Dorsey ER, Beck CA, Darwin K, et al. Natural history of Huntington disease. *JAMA Neurol*. 2013;70:1520-1530.
10. O'Toole O, Lennon VA, Ahlskog JE, et al. Autoimmune chorea in adults. *Neurology*. 2013; 80:1133-1144.
11. Beato R, Maia DP, Teixeira AL Jr, et al. Executive functioning in adult patients with Sydenham's chorea. *Mov Disord*. 2010;25:853-857.
12. Albanese A, Bhatia K, Bressman SB, et al. Phenomenology and classification of dystonia: a consensus update. *Mov Disord*. 2013;28:863-873.
13. Jinnah HA, Berardelli A, Comella C, et al. The focal dystonias: current views and challenges for future research. *Mov Disord*. 2013;28:926-943.
14. Tuschl K, Clayton PT, Gospe SM Jr, et al. Syndrome of hepatic cirrhosis, dystonia, polycythemia, and hypermanganesemia caused by mutations in SLC30A10, a manganese transporter in man. *Am J Hum Genet*. 2012;90:457-466.
15. Vidailhet M, Jutras MF, Grabli D, et al. Deep brain stimulation for dystonia. *J Neurol Neurosurg Psychiatry*. 2013;84:1029-1042.
16. Knight T, Steeves T, Day L, et al. Prevalence of tic disorders: a systematic review and meta-analysis. *Pediatr Neurol*. 2012;47:77-90.
17. Scahill L, Woods DW, Himle MB, et al. Current controversies on the role of behavior therapy in Tourette syndrome. *Mov Disord*. 2013;28:1179-1183.
18. Popa T, Milani P, Richard A, et al. The neurophysiological features of myoclonus-dystonia and differentiation from other dystonias. *JAMA Neurol*. 2014;71:612-619.
19. Bhidayasiri R, Fahn S, Weiner WJ, et al. Evidence-based guideline: treatment of tardive syndromes. Report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology*. 2013;81:463-469.
20. Anheim M, Tranchant C, Koenig M. The autosomal recessive cerebellar ataxias. *N Engl J Med*. 2012;366:636-646.
21. Libri V, Yandim C, Athanasopoulos S, et al. Epigenetic and neurological effects and safety of high-dose nicotinamide in patients with Friedreich's ataxia: an exploratory, open-label, dose-escalation study. *Lancet*. 2014;384:504-513.
22. Fink JK. Hereditary spastic paraplegia: clinico-pathologic features and emerging molecular mechanisms. *Acta Neuropathol*. 2013;126:307-328.

REVIEW QUESTIONS

1. A 49-year-old woman presents with a 5-year history of involuntary pulling of her head toward the right shoulder. At times the movement has a tremulous or jerky component. The involuntary movement is associated with pain in the muscles in the back right side of the neck. The neck movements are interfering with her daily activities, particularly attending meetings, reading, watching television (she prefers to lie down for the latter), and driving. Initially she found that lightly touching the left side of her face or chin would allow her to keep her head in the mid-position, but this maneuver has become less effective recently. She has no other neurologic complaints. Past and family histories are unremarkable, and she denies use of any prescription medications. On examination, she has rotation and tilting of her head to the right shoulder. She is unable to maintain her head in the midline for more than 15 seconds, and when asked to allow the head to take on its preferred position it rests on the shoulder with near 90-degree rotation and significant tilting. She also demonstrates a mild postural and action tremor in her hands. The remainder of the examination it is unremarkable. Which of the following statements is false?

A. The history and clinical examination are typical of idiopathic adult-onset cervical dystonia.
B. Genetic testing, including assessment of mutations in the *torsin A* and *THAP1* genes, would be negative and are generally not indicated.
C. Deep brain stimulation would not be considered appropriate given the very focal nature of the dystonia.
D. Botulinum toxin A injections would be considered appropriate as the first line of therapy.
E. Drug history is important to elicit in such patients given the potential for tardive dystonia to manifest this phenotype.

Answer: C This patient has typical idiopathic cervical dystonia. Genetic testing is generally unremarkable and largely unnecessary unless there are other clinical clues to an alternative diagnosis. Botulinum toxin A remains the mainstay of therapy and is generally the treatment of choice unless the neck movements are extremely variable and complex and do not lend themselves to the benefit of weakening selected muscles, in which case medical therapy might be chosen initially. Tardive dystonia due to dopamine receptor blocking drugs (antipsychotics, some antiemetics) can result in a phenotype identical to that of idiopathic cervical dystonia, although a retrocollis pattern may be more common and other movements (e.g., orofacial dyskinesia), are not uncommonly also present. Deep brain stimulation of the globus pallidus has been shown in several studies to be very effective for medically refractory cervical dystonia.

2. You are seeing a 25-year-old student who has a long-standing history of jerky movements. These began when he was age 7 with excessive eye blinking. Symptoms have waxed and waned over the years and have included a variety of facial contractions such as nose wrinkling, jaw opening, head jerking, shoulder shrugging, hand shaking, contraction of abdominal and pelvic muscles, and leg kicking. When he was younger, he variably made sniffing or throat clearing sounds but never had spoken words or other vocalizations. His father has long-standing facial twitching and sniffing, and his mother has a tendency to obsessive-compulsive behavior. The patient has no history of behavioral or learning disabilities. He is an honors student and is considering applying for law school. The movements are having a negative impact on his social life, and he is concerned that they may interfere with his ability to interview successfully for law school. Which of the following statements is true?

A. Behavioral therapy has a role to play in managing patients with this disorder.
B. The description of the patient's movements is most compatible with multifocal myoclonus, probably due to "essential myoclonus."
C. This syndrome constitutes a relatively rare disorder that is largely limited to the pediatric age group.
D. Most patients with this disorder should be considered for early drug treatment to avoid the problems currently experienced by this patient.
E. The absence of coprolalia or involuntary swearing makes it unlikely that his symptoms fulfill criteria for the diagnosis of Tourette syndrome.

Answer: A This patient fulfills the diagnostic criteria for Tourette syndrome. This common neuropsychiatric disorder typically begins in childhood and may remit in adolescence. However, it often persists to a greater or lesser extent in adulthood, and a small proportion of patients have symptoms beginning in adult life. Coprolalia is a relatively uncommon symptom. A large proportion of patients can be managed without drug therapy, which should be reserved for patients in whom tics are causing disability or impairing quality of life. Comorbid symptoms such as attention-deficit/hyperactivity disorder, obsessive compulsive disorder, and learning disabilities often are more disabling than the tics themselves. Behavioral intervention, including habit reversal training, is a proven treatment.

3. Which of the following statements related to the neuroleptic drug-induced movement disorders is true?

A. Acute dystonic reactions most often occur in elderly patients who receive atypical neuroleptics.
B. Acute akathisia is an idiosyncratic side effect that generally resolves as the dose of the neuroleptic is increased.
C. Neuroleptic malignant syndrome remains a severe and sometimes fatal complication of older typical neuroleptics, but newer atypical agents are largely devoid of this side effect.
D. The introduction of atypical neuroleptics has clearly been associated with a reduced incidence of tardive dyskinesia.
E. The first-line therapy of classic orobuccolinguomasticatory tardive dyskinesia is either trihexyphenidyl or benztropine 2 mg three times daily followed by reduction and if possible discontinuation of the causative neuroleptic.

Answer: D There has been a considerable reduction in the occurrence of tardive dyskinesia with the introduction of atypical neuroleptics, although these drugs still cause the problem to a lesser extent. Acute dystonic reactions occur more frequently in young patients who receive potent neuroleptic agents. Acute akathisia is often an early side effect that worsens in a dose-related fashion and resolves on withdrawal of the drug. Neuroleptic malignant syndrome remains an important complication of the atypical neuroleptics. Orobuccolinguomasticatory tardive dyskinesia typically worsens in response to anticholinergic drugs, which should be withdrawn in patients with problematic classic tardive dyskinesia. Conversely, these drugs are appropriate for the treatment of tardive dystonia, even though some patients develop more typical oral dyskinesias in response to them.

411

MULTIPLE SCLEROSIS AND DEMYELINATING CONDITIONS OF THE CENTRAL NERVOUS SYSTEM

PETER A. CALABRESI

The disorders of myelin encompass a wide range of diseases in which either myelin is not formed in a normal fashion (dysmyelinating disease) or normally formed myelin is destroyed or not maintained appropriately (demyelinating disease) (Table 411-1). *Dysmyelinating* diseases are uncommon and include an array of leukodystrophies that have a genetic basis. *Demyelinating* diseases are much more common and include multiple sclerosis (MS), which represents more than 95% of all types of disorders of central nervous system (CNS) myelin.

Some disorders of myelin have a distinct pathogenesis in which the disruption of myelin is secondary. Further, in many of the diseases of myelin, the axon degenerates as a result of decreased trophic support from loss of myelin, impaired health of the oligodendrocyte, or increased susceptibility to injury in the absence of myelin. This observation led to the recent hypothesis that axonal loss is the underlying substrate for permanent disability in MS, adrenoleukodystrophy, and perhaps other diseases of myelin.[1]

MULTIPLE SCLEROSIS

DEFINITION

MS is a disease characterized by multifocal areas of demyelination in the brain and spinal cord, with associated inflammatory cell infiltrates, reactive gliosis, and axonal degeneration. It typically manifests in young adults with episodic neurologic dysfunction. Although the exact origin of MS remains enigmatic, evidence suggests that it is an immune-mediated attack on myelin, with secondary disruption of axons leading to progressive disability over time in most afflicted patients.

TABLE 411-1 DISEASES OF MYELIN

IDIOPATHIC
Recurrent or chronic progressive demyelination (multiple sclerosis and its variants)
Monophasic demyelination (may be the first clinical episode of multiple sclerosis)
Optic neuritis
Acute transverse myelitis
Acute disseminated encephalomyelitis; acute hemorrhagic leukoencephalopathy
VIRAL INFECTIONS
Progressive multifocal leukoencephalopathy
Subacute sclerosing panencephalitis (Chapter 370)
NUTRITIONAL AND METABOLIC DISORDERS (Chapter 416)
Combined systems disease (vitamin B_{12} deficiency)
Copper deficiency (dorsal columns and subacute optic neuropathy)
Demyelination of the corpus callosum (Marchiafava-Bignami disease)
Central pontine myelinolysis
ANOXIC-ISCHEMIC SEQUELAE (Chapter 404)
Delayed postanoxic cerebral demyelination
Progressive subcortical ischemic encephalopathy
LEUKODYSTROPHIES PRIMARILY AFFECTING CENTRAL NERVOUS SYSTEM MYELIN
Adrenoleukodystrophy (Schilder disease)
Pelizaeus-Merzbacher disease (sudanophilic leukodystrophies)
Spongy degeneration
Vanishing white matter disease
Others (Alexander disease, Canavan disease)
Leukodystrophies of the central and peripheral nervous system
Metachromatic leukodystrophy
Globoid cell leukodystrophy (Krabbe disease)

EPIDEMIOLOGY

The annual incidence of MS varies by location and ranges between 1.5 and 11 per 100,000 people. MS is second only to trauma as the most common cause of neurologic disability in young adults. Recent studies suggest that the incidence rate has increased, in part because of recognition of more cases at an earlier stage, but probably also because of a truly rising incidence, especially in women. The prevalence is estimated at 350,000 to 400,000 in the United States and more than 1,000,000 worldwide, but these numbers may be underestimates owing to incomplete recognition of the disease, even in developed countries, and the increased incidence since these estimates were made.

MS occurs 2- to 2.5-fold more frequently in women than in men, a sex predilection that is common in autoimmune diseases. The disease most often manifests in the third to fourth decades of life, but with an incidence age range from postpubertal teenagers to persons in their 50s. Rare cases occur in infants or in patients in their 60s, but extreme caution is warranted in these situations to exclude alternative processes. In many of the late-onset MS cases, symptoms were present in younger years and were attributed to other causes.

MS is most common in people of Northern European descent. In many areas of the world, MS is more prevalent in temperate latitudes (approaching 1 in 500 in some locations) and becomes less common toward the equator (1 in 20,000 or rare case reports only in some locations), perhaps explained in part by migration patterns of people with the same gene pools. However, the absence of complete genetic penetrance in monozygotic twin studies and recent increases in incidence in genetically stable populations strongly suggest an environmental component to the disease. Indeed, an outbreak of MS was documented on the Faroe Islands following World War II, and numerous other clusters have been reported, although a single environmental trigger has not been identified.[2]

Several studies have linked cigarette smoking with risk for MS. High levels of vitamin D and early exposure to excessive sunlight (sunburns) have been linked with lower risk for MS, possibly related to the beneficial effects of cholecalciferol on regulating immune cell responses.

PATHOBIOLOGY AND GENETICS

Monozygotic twins with MS show a concordance rate of between 15 and 50%, compared with only 3 to 5% concordance in dizygotic twins, consistent with a strong but incomplete role for genes in causing MS. The lifetime risk for MS is increased to 2 to 4% in individuals with a first-degree relative with MS, compared with the general population risk of 0.1%. In addition, between 10 and 20% of patients with MS have a first-degree relative with another autoimmune disease, commonly rheumatoid arthritis, systemic lupus erythematosus, or autoimmune thyroid disease. Psoriasis (Chapter 438) and inflammatory bowel disease (Chapter 141) also may be more common in patients with MS. Genetic modeling of the disease strongly argues against a single MS gene and suggests that many different genes predispose to MS and account for its many phenotypes and its overlap with other autoimmune diseases.[3] Linkage and association studies have identified the human leukocyte antigen (HLA) or major histocompatibility complex (MHC) region on chromosome 6p21 as one genetic determinant for MS. The MHC class II region, involved in presentation of antigen to CD4+ T cells, is the most strongly associated locus. The HLA-DR2 allele and, more specifically, the molecular haplotype HLA-DRB*1501 allele have repeatedly been implicated. Multiple single-nucleotide polymorphisms (SNPs) in the interleukin-2 (IL-2) receptor-α gene and the IL-7 receptor-α gene also appear to be associated with a higher risk for MS. Over 100 other gene SNPs have been identified, most of which are related to immune function. Although patterns are emerging to suggest dysregulation of differing immune cell subsets, the associations to date are not strong enough to have clinical predictive value.

PATHOLOGY

Most cases are characterized by multifocal areas of demyelination and gross gliotic scar in the brain and spinal cord. Classic locations of these lesions, called *plaques,* are the optic nerves, periventricular white matter, deep white matter, juxtacortical white matter, corpus callosum, cerebellar peduncles, and dorsolateral spinal cord. However, there is a bias toward recognition of lesions in white matter because of the relative ease of detecting demyelination and inflammation in white compared with gray matter. Indeed, more recent pathologic studies have confirmed demyelination, neuritic damage, and atrophy in the cerebral cortex (pial surface and intracortical) and deep gray

matter structures. At the microscopic level, one usually sees multiple areas of perivenular inflammatory cell infiltrates with extravasation into the surrounding tissue parenchyma. In the acute active plaque, CD4 helper T (T_H) cells are prominent in the perivenular areas. Proinflammatory cytokines released from T_H1 (interferon-γ [IFN-γ]) and T_H17 (IL-17, TNF, and granulocyte-macrophage colony-stimulating factor [GM-CSF]) cells are thought to mediate damage. Increasingly, large numbers of CD8 cytotoxic T cells have been documented in brain tissue, especially in the parenchyma, and these cells may mediate direct damage to axons and oligodendrocytes through release of proteases such as granzyme B. Most parenchymal inflammatory cells, especially in chronic plaques, are CD68+ macrophages and microglia. In addition to the influx of circulating immune cells, prominent astroglial activation and in some cases oligodendrocyte precursor cell differentiation occurs in response to injury. Over time, the inflammation becomes less prominent in the center of the plaque, but a chronic active rim of inflammation with microglial activation exists at a well-demarcated border between abnormal and normal unharmed myelin. This characteristic of MS is seldom seen in other disorders of myelin. Although oligodendrocytes may survive, proliferate, and result in partial remyelination (shadow plaques) in some early cases, this process is hardly ever complete in MS. Over time, remyelination is less successful, and oligodendrocyte precursor cells appear unable to differentiate into mature myelinating oligodendrocytes.[4]

The number of damaged axons correlates with the extent of inflammation. Further, axonal damage and even neuronal apoptosis and loss are seen in the cortex and retina. Atrophy of both the brain and spinal cord, which occurs more rapidly in MS than in normal aging, reflects loss of both myelin and axons.

No consistent microbial cause has been discerned from careful examination of MS tissues for known infectious pathogens. Differential expression of human herpesvirus type 6, which is acquired by most people in childhood, has been noted in oligodendrocytes of patients with MS, but whether this virus is a cofactor in demyelination or just a bystander remains unclear. Evidence suggests the possibility that the earliest event in MS may be an insult to the oligodendrocytes, with subsequent activation of resident immune cells and secondary recruitment of other immune cells only at later stages.

Some pathologists believe that four distinct subtypes of MS can be discerned, in which the pathologic characteristics are consistent in every lesion, thereby allowing classification of patients with differing pathologic categories rather than just describing evolution of lesions over time. Type I lesions are characterized by typical perivenular inflammatory infiltrates consisting mainly of T cells, with early preservation of oligodendrocytes. Type II lesions are similar to type I but have an additional humoral component with immunoglobulin G (IgG) deposition and complement activation. Type III lesions are distinguished by not being based around venules and by prominent loss of myelin-associated glycoprotein, with evidence for oligodendrocyte apoptosis. Type IV lesions have inflammatory infiltrates more similar to those in types I and II but also have oligodendrocyte loss as in type III. These varying pathologic features may begin to explain clinical subtypes of the disease.

PATHOGENESIS

It remains possible that the autoimmune hypothesis is wrong and that the inflammation observed in MS is secondary to an as yet uncharacterized primary degenerative process. Proponents of this theory cite evidence from pathologic features of hyperacute cases, in which the oligodendrocytes appear to die before any systemic immune response occurs, as well as recent data revealing neuronal and axonal death or demyelination in the absence of inflammation.

Macrophages and microglia, which make up the majority of cells within the parenchymal infiltrate in chronic MS plaques, are potent antigen-presenting cells and express HLA and costimulatory molecules. Activated macrophages and microglia also have effector functions, including release of cytokines that are partly (IL-6, tumor necrosis factor-α) or completely distinct from the T cells (IL-1β, IL-12, and IL-23). In high concentrations, these cytokines may damage oligodendrocytes and neurons and activate T cells.

CLINICAL MANIFESTATIONS

Presenting Symptoms

MS, which can manifest in many ways across a broad age range, may initially masquerade as a variety of different illnesses (Table 411-2; see Table 411-1). In a classic presentation, a young white person, more often a woman, will have the acute to subacute onset of impaired vision or sensation. Fatigue, depression, bladder urgency, weakness, impaired balance, and impaired coordination also are common symptoms. The often remarkably mild nature of the first symptoms often dissuades the patient from seeking medical attention or is insufficiently impressive to stimulate the physician to order diagnostic tests. Furthermore, patients may initially have few objective neurologic findings, especially between attacks.

TABLE 411-2 CONDITIONS THAT CAN BE MISTAKEN FOR MULTIPLE SCLEROSIS AND OTHER DISEASES OF MYELIN

VASCULAR DISEASE
Small-vessel cerebrovascular disease
Vasculitides
Arteriovenous malformation
CADASIL
Antiphospholipid antibody syndrome
STRUCTURAL LESIONS
Craniocervical junction, posterior fossa, or spinal tumors
Cervical spondylosis or disc herniation
Chiari malformation or syrinx
DEGENERATIVE DISEASES
Hereditary myelopathy
Spinocerebellar degeneration
INFECTIONS
HTLV-1 infection
HIV myelopathy or HIV-related cerebritis
Neuroborreliosis (e.g., Lyme disease)
JC virus/progressive multifocal leukoencephalopathy
Neurosyphilis
OTHER INFLAMMATORY CONDITIONS
Systemic lupus erythematosus
Sjögren syndrome
Sarcoidosis
MONOFOCAL OR MONOPHASIC DEMYELINATING SYNDROMES
Transverse myelitis
Optic neuritis
Neuromyelitis optica/Devic disease
Acute disseminated encephalomyelitis
OTHER CONDITIONS
Hashimoto thyroiditis with or without encephalopathy
Nonspecific MRI abnormalities related to migraine, aging, or trauma

CADASIL = cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy; HIV = human immunodeficiency virus; HTLV = human T-cell lymphotropic virus; MRI = magnetic resonance imaging.

Paresthesias of a limb that are circumferential and do not follow a dermatome suggest a spinal cord lesion; these symptoms often manifest distally and then ascend to involve more proximal parts of the limb, spread to the contralateral limb, or progress from a leg to an arm. Similarly, bandlike sensations around a limb or the torso also suggest a myelopathic process.

Incomplete transverse myelitis is a focal (partial) spinal cord syndrome that is usually inflammatory and does not follow vascular territories. It is a common presentation of MS.

Lhermitte sign, an electrical sensation moving down the spine into the limbs on flexion of the neck, is characteristic of cervical myelitis from any cause, including MS. Frank loss of sensation is less common as an early symptom or sign but is seen in more advanced cases. Burning, electrical, or deep aching sensations are also common in MS.

Sensory Abnormalities

On examination, the most common sensory findings are loss of vibration perception, most prominent in the feet, and incomplete spinal cord levels to pinprick or vibration, which are often more notable in a graded fashion rather than at a distinct level. Such sensory levels may be asymmetrical and differ by sensory modality because of isolated demyelination in the dorsal columns compared with the spinothalamic tracts. Patchy or seemingly nonanatomic focal areas of impaired sensation can occur, and some patients describe bizarre sensations such as water dripping or bugs crawling on an area of the body.

Visual Effects

Optic neuritis (Chapter 424) is a classic manifesting syndrome, typically with visual symptoms in one eye. In optic neuritis, patients often complain of pain over the temporal eyebrow and worsening on lateral eye movement. The visual impairment may be described as looking through frosted glass or a veil. The scotoma or area of greatest loss often can be mapped in a centrocecal distribution (central focal point to the blind spot laterally), which in mild cases may be evident only as desaturation to red color using the head of a pin. More severe cases may result in total loss of light perception. In most acute cases of optic neuritis, the inflammation is retrobulbar (behind the disc), so no immediate changes are visible on the optic disc, thereby leading to the aphorism "the patient sees nothing, and the doctor sees nothing." However, there should be a relative afferent papillary defect (Marcus-Gunn pupil; Chapter 424) with paradoxical dilation of the affected eye to direct light on swinging a flashlight from the unaffected eye in which consensual constriction was induced. In cases of bilateral optic neuritis (new or old), this abnormality may not be seen. Patients usually spontaneously recover substantial vision after weeks to months. Later, the optic disc may become pale, especially in the temporal region, a finding reflecting damage to the axons following inflammation and demyelination, even with recovery of normal visual acuity. Patients often have more subtle chronic visual impairment for colors, low contrast visual acuity, and contrast sensitivity. Visual testing using low contrast letter acuity charts commonly reveals substantial visual loss after clinical optic neuritis.

Visual impairment from impaired tracking of eye movements owing to brain stem or cerebellar disease most commonly occurs in the setting of an acute lesion affecting the medial longitudinal fasciculus, which is the neurologic pathway that yokes the eyes together on lateral saccades. Patients may experience frank diplopia or just blurred vision, especially when they look off to one side rapidly, such as when looking over one's shoulder while driving. The neurologic sign of this problem is called *internuclear ophthalmoplegia* (Chapter 424) and manifests as slowed or absent adduction of one eye with abducting nystagmus of the other eye. It may occur bilaterally or may exist in milder forms, such that the adduction lag is imperceptible to the human observer. Blurred vision from cerebellar damage with nystagmus is very common in MS and is often worse on extreme lateral or vertical gaze. *Oscillopsia,* the sensation that the environment is moving when it actually is not, is another symptom of impaired cerebellar coordination of the eyes. Saccadic eye movement or loss of smooth pursuit is common in MS and also can be seen in numerous neurologic conditions or with aging.

Motor Symptoms

The most common motor symptoms of MS are weakness and impaired coordination in a leg, with ascending involvement from distal to proximal and commonly spreading to the contralateral leg or ipsilateral arm. The lesion causing these symptoms is more commonly in the cervical spinal cord rather than the thoracic spinal cord, even when the first sign is partial footdrop. It is likely that axons that must conduct impulses over the longest distance (entire length of the spinal cord) from a site of inflammatory demyelination will become symptomatic before axons delivering signals to closer synapses (adjacent anterior horn cells in the cervical cord). Clinically, the weakness may be severe and may result in an obvious paralysis or be so subtle as to be undetectable. Heat-induced fatigue and weakness, as manifested by focal symptoms (slapping of a foot or dragging a leg) occurring after 15 to 20 minutes of exercise and resolving with rest, are characteristic of early demyelinating disease. The early absence of associated hyperreflexia and plantar extensor responses (Babinski sign) may make it difficult to document corticospinal tract involvement. Later, in more established MS, classic corticospinal tract signs are often evident and manifest clinically as spastic gait (either hemiparetic or paraparetic), muscle cramps, and clonus (sustained reflex loop), sometimes occurring with positional changes and mistaken for signs of a cerebellar tremor.

Ataxia may occur as a result of impaired delivery of sensory information up the spinal cord or from demyelination of cerebellar pathways in the brain stem or cerebellum. Often, the two are mixed and may be confounded further by visual loss and impaired ability to compensate by fixing on the environment; this combination commonly causes dizziness in crowds, in which fixation may be further obscured. Appendicular dysmetria resulting in tremor on reaching for an object is a common cause of impaired coordination and dexterity. Lower extremity and truncal ataxia may result in a wide-based (drunk) gait. Other movement disorders, such as postural tremor and titubation (head tremor), are much less common in MS. *Myokymia* (wormlike muscle movements) under the skin, especially around the face, however, is fairly common. Pseudoathetosis and parkinsonism can be seen in severe cases.

Cognitive and Behavioral Symptoms

Over 50% of patients with MS experience bouts of moderate-to-severe depression (Chapter 397). There is also increased incidence of bipolar disease, which may manifest after treatment of depression or treatment with corticosteroids. Pseudobulbar affect, either pathologic laughing or crying, is seen in patients with more advanced disease. Numerous cognitive symptoms, including short-term memory loss, word-finding difficulty, trouble with multitasking, and cognitive fatigue, may be mistaken for depression but are well-recognized primary symptoms of MS pathology. Most patients do not progress to dementia (Chapter 402), but cognitive and behavioral impairments are major causes of losing employment and marital discord.

Organ Dysfunction

Bladder symptoms are extremely common, but often are not volunteered, so specific questions must be asked concerning urinary frequency, urgency, incontinence, or retention. Careful discrimination of a spastic bladder (detrusor muscle spasm) causing incontinence from an atonic bladder or spasm of the external sphincter (the latter two causing retention) leading to overflow incontinence is critical to designing treatment (Chapter 26). Urinary tract infections (Chapter 284) owing to bladder dysfunction may aggravate symptoms of MS.

Bowel dysfunction commonly manifests as constipation (Chapter 136), which may be primary (related to spinal cord involvement) or secondary (related to self-induced dehydration to manage urinary frequency or to side effects of anticholinergic drugs). Bowel incontinence secondary to an incompetent anal sphincter is less common and most often occurs as an isolated episode of fecal urgency, sometimes related to dietary change or diarrheal illness.

Sexual dysfunction is common and underdiscussed in MS. In men, erectile dysfunction is frequent. In women and men, loss of libido and inability to achieve orgasm can occur as a result of medication, loss of sensation, heat-induced worsening of symptoms, physical barriers to intercourse (impaired mucosal moisture, spasticity, and pain), depression, or disorders of body image.

Systemic Symptoms

Fatigue is common in MS. It may be linked to depression but often occurs independently and can be the most disabling symptom of the disease. A sleep history is important to exclude daytime fatigue resulting from disrupted sleep secondary to pain, cramps, bladder frequency, sleep apnea, periodic limb movements, depression, or disrupted sleep-wake cycles. Daytime fatigue even after a good night of sleep may occur in mid-afternoon and may be described as being "unplugged" or completely drained. Many patients obtain benefit from a short daytime nap.

Sensitivity to heat, which is a classic symptom of MS, occurs only in some patients. Even minor elevations of the body temperature can dramatically worsen symptoms (Uhthoff phenomenon). Some patients complain of worsened symptoms in cold weather, likely related to increased dysfunction of already stiff muscles or signal blockade consistent with the known physiology of nerve conduction, which has an inverted -shaped temperature versus conduction curve.

Pregnancy

Women with MS may have children, and the activity of MS lessens during the course of pregnancy, especially by the third trimester, when the frequency of exacerbations is reduced by approximately two thirds.[5] Relapses are more frequent in the first 6 postpartum months, but no evidence indicates that pregnancy changes the natural history of the disease. Whether breast-feeding alters the course of MS is unclear, but it is contraindicated for patients who resume disease-modifying drugs following delivery.

Types of Multiple Sclerosis

The three major clinical types of MS are relapsing remitting, secondary progressive, and primary progressive. Approximately 85 to 90% of patients present with relapsing-remitting MS, characterized by acute or subacute episodes of new or worsening old neurologic symptoms that increase in severity, plateau, and then partly or completely remit. Patients may have no detectable residual deficit, or they may accumulate significant permanent disability from

an attack. Most patients with relapsing-remitting MS convert to secondary progressive MS after 20 to 40 years. This stage of the disease, which is characterized by at least 6 months of progressive worsening without evidence of a relapse, can be diagnosed with confidence only retrospectively. Some patients with secondary progressive MS also have interposed relapses distinct from their periods of progressive worsening, although these episodes become less frequent with time. Primary progressive MS, which is characterized by progressive deterioration from the onset for at least 1 year without a history of distinct relapses, occurs in approximately 10 to 15% of patients. It is more common in middle-aged men and typically has more involvement of the spinal cord and fewer inflammatory brain lesions.

Other uncommon types of MS also are described. Progressive relapsing MS refers to a fairly uncommon variant of MS (6%), in which a relapse ensues after an initially primary progressive course. Acute progressive MS (Marburg disease) causes acute or subacute progressive neurologic deterioration leading to severe disability within days to a month in a patient with no prior history of MS. This rare form of the disease may progress to a quadriplegic, obtunded state with death as a result of intercurrent infection, aspiration, or respiratory failure from brain stem involvement.

DIAGNOSIS

The diagnosis of MS rests on demonstrating evidence of at least two inflammatory demyelinating lesions referable to different locations within the CNS, occurring at different times (usually ≥ 1 month apart), and for which no better explanation exists.[6] Diagnostic criteria allow for the diagnosis to be made on clinical grounds alone as long as appropriate exclusionary testing is performed (Table 411-3). Clinical evidence of a lesion requires objective findings on examination, not just a symptom. Further, repeated episodes of neurologic dysfunction that could be explained based on one lesion (e.g., a cervicomedullary junction lesion causing brain stem, cerebellar, and corticospinal tract dysfunction) is not enough evidence to diagnose MS.

Magnetic Resonance Imaging

No definitive diagnostic laboratory test exists for MS, but magnetic resonance imaging (MRI) of the brain is extremely useful and should be performed in all patients in whom MS is a diagnostic consideration.[7] More than 95% of patients with clinically definite MS have an abnormal brain MRI, and the presence of high-signal, bright lesions is so characteristic of MS that a normal brain MRI should suggest an alternative diagnosis. Brain MRI is also useful in predicting future MS at the time of a clinically isolated demyelinating syndrome. Specific MRI findings allow for confirmation of disease disseminated in time and space (different parts of the brain or spinal cord) and fulfilling evidence for dissemination in time (Table 411-4). MS plaques typically appear as high-signal (white) areas on fluid attenuation inversion recovery (FLAIR) T2-weighted images, which allow for the best discrimination of the supratentorial lesions by suppressing high signal from cerebrospinal fluid (CSF) in the ventricles (Fig. 411-1). Lesions generally range in size from 2 mm to 2 cm; larger plaques occasionally resemble a tumor. Features of an MRI lesion suggesting MS include an elliptical shape, discrete borders, lack of mass effect, and gadolinium enhancement. Typical locations include the periventricular area (perpendicular to or abutting the walls of the ventricles) (Fig. 411-2), the corpus callosum, the cerebellar peduncles, the brain stem, the juxtacortical area, and the dorsolateral spinal cord (Fig. 411-3). Cortical

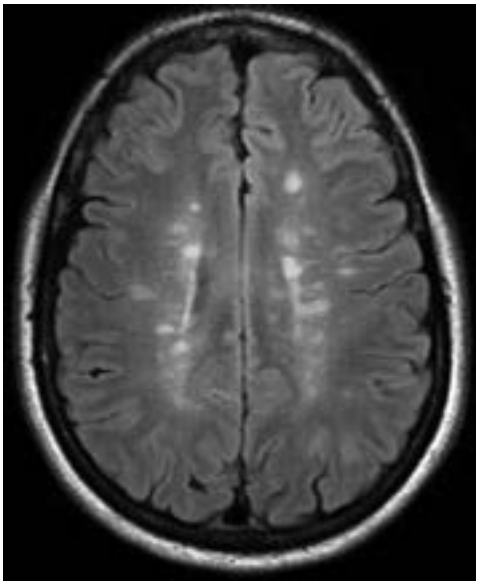

FIGURE 411-1. Axial fluid attenuation inversion recovery image of the brain from a patient with multiple sclerosis revealing classic multiple periventricular and deep white matter high signal lesions.

TABLE 411-3 2010 REVISIONS TO THE MCDONALD DIAGNOSTIC CRITERIA FOR MULTIPLE SCLEROSIS

CLINICAL PRESENTATION	ADDITIONAL DATA NEEDED FOR DIAGNOSIS OF MULTIPLE SCLEROSIS
Two or more attacks; objective clinical evidence of two or more lesions; or one lesion with a prior attack	None*
Two or more attacks; objective clinical evidence of one lesion	Dissemination in space, demonstrated by: 1. MRI (see Table 411-4), or 2. Two or more MRI-detected lesions consistent with MS plus positive CSF, or 3. Await further clinical attack implicating a different site
One attack; objective clinical evidence of two or more lesions	Dissemination in time, demonstrated by: 1. MRI (see Table 411-4), or 2. Second clinical attack
One attack; objective clinical evidence of one lesion (monosymptomatic presentation; clinically isolated syndrome)	1. Dissemination in space, demonstrated by: a. MRI (see Table 411-4), or b. Two or more MRI-detected lesions consistent with MS plus positive CSF, and 2. Dissemination in time, demonstrated by: a. MRI (see Table 411-4), or b. Second clinical attack

Modified from Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol.* 2011;69:292-302.
*Must rule out other causes (e.g., see Table 411-2).
CSF = cerebrospinal fluid; MRI = magnetic resonance imaging; MS = multiple sclerosis.

TABLE 411-4 MAGNETIC RESONANCE IMAGING CRITERIA IN MULTIPLE SCLEROSIS (INTERNATIONAL PANEL RECOMMENDATIONS: 2010)

DISSEMINATION IN TIME
Detection of a new T2 or gadolinium-enhancing lesion if it appears at any time compared with a reference scan*
Simultaneous presence of asymptomatic gadolinium-enhancing and nonenhancing lesions at any time
DISSEMINATION IN SPACE
At least one gadolinium-enhancing lesion in at least two of four areas: Periventricular Juxtacortical Infratentorial Spinal cord
DIAGNOSIS OF PRIMARY PROGRESSIVE MULTIPLE SCLEROSIS
One year of disease progression (retrospectively or prospectively determined) plus two of the following: a. Positive brain magnetic resonance imaging (≥1 T2 lesion in at least one characteristic area: periventricular, juxtacortical, or infratentorial) b. Positive spinal cord magnetic resonance imaging (two focal T2 lesions) c. Positive cerebrospinal fluid (isoelectric focusing evidence of oligoclonal immunoglobulin G bands or increased immunoglobulin G index, or both)

From Polman Ch, Reingold SC, Banwell G, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol.* 2011;49:292-302.
*CAUTION: Determination that a T2 lesion is indeed new can be challenging. A new T2 lesion must be of sufficient size and location to reflect one that could not have been missed previously for technical reasons of slice orientation, thickness or spacing, tissue contract, patient motion, or other artifacts. This judgment requires standardized scanning procedures, with emphasis on careful repositioning, as well as input from qualified evaluators experienced in multiple sclerosis imaging.

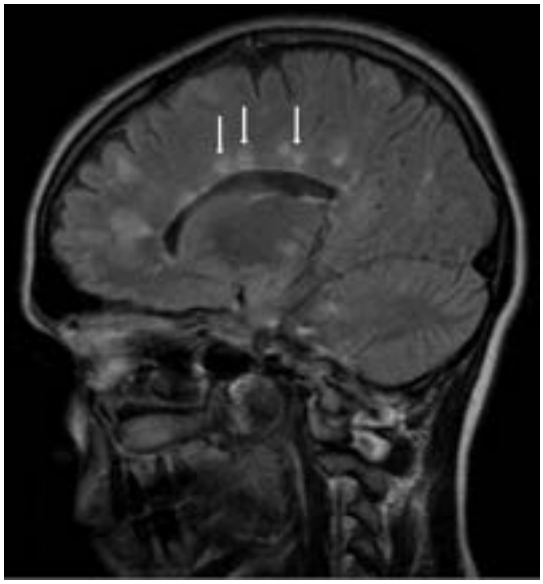

FIGURE 411-2. Sagittal fluid attenuation inversion recovery image of the brain from a patient with multiple sclerosis revealing classic periventricular lesions radiating outward from the ventricles (*arrows*).

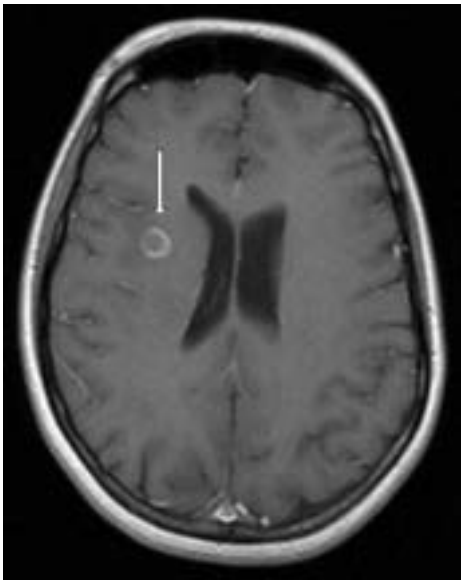

FIGURE 411-4. Axial T1-weighted image after gadolinium contrast showing an actively inflamed ring-enhancing lesion (*arrow*) in a patient with multiple sclerosis.

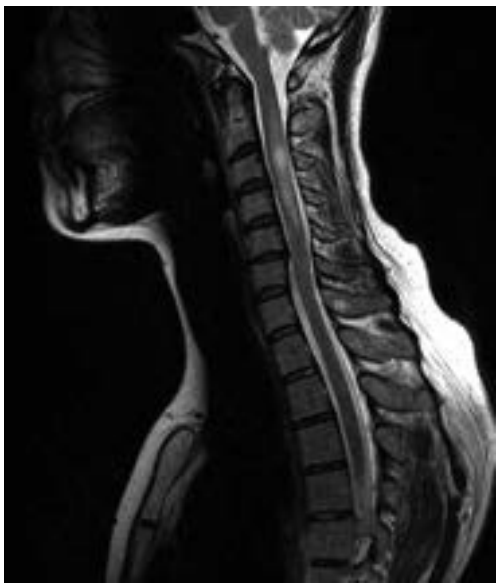

FIGURE 411-3. Sagittal T2-weighted image of the brain and cervical spine from a patient with multiple sclerosis. The image shows a high-signal plaque from C3-C5 in the spinal cord.

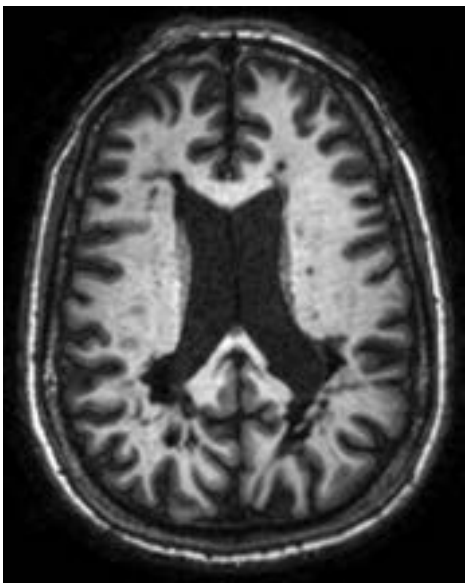

FIGURE 411-5. Axial T1-weighted image showing numerous areas of T1 low signal ("black holes"), ventricular enlargement, and diffuse atrophy.

and deep gray matter lesions also occur but are less clearly seen on conventional MRI. Gadolinium enhancement, which suggests permeability of the blood-brain barrier, is correlated with new or active inflammation in lesions (Fig. 411-4). Lesions that enhance on a T1-weighted sequence usually have a concomitant lesion in the same location on a T2-weighted image. However, T2-weighted lesions may form without evident enhancement. Gadolinium enhancement typically persists for 2 to 8 weeks and thus may be missed on intermittent scans. Persistent areas of low signal on T1-weighted images before contrast ("black holes") correlate with pathologic evidence of axonal loss and atrophy (Fig. 411-5).

Cerebrospinal Fluid

Examination of the CSF is useful in many cases but is not mandatory in patients with a typical clinical presentation and MRI evidence of disseminated disease. CSF evaluation includes cell counts, total protein, glucose, oligoclonal bands with a paired serum sample, and an IgG index. The presence of myelin basic protein is not specific for MS because it can be elevated secondary to any disruption of CNS tissue. Oligoclonal IgG bands in the CSF or an elevated IgG index provides evidence for intrathecal production of immunoglobulins. However, although oligoclonal bands are common in MS, they also can occur with infection or other immune-mediated processes. As a result, the test lacks specificity for MS and has a sensitivity of only approximately 85 to 90% of patients with clinically definite MS. In clinically isolated demyelinating syndromes (see later), the sensitivity is even lower (~50%). Further, the sensitivity depends on local laboratory techniques.

CSF evaluation is generally recommended if an alternative diagnosis is considered, especially if one suspects an infectious or neoplastic process (e.g., fever, sweats, unusual travel history, tick bite, or rash). CSF analysis also may be useful if clinical or MRI criteria are incomplete to provide confirmation of the diagnosis.

Evoked Potential Tests

Evoked potentials (Chapter 396) may be useful in some situations to document objective evidence of slowed conduction owing to demyelination in locations different from those recognized clinically. However, visual evoked potentials (VEPs), brain stem auditory evoked potentials, and somatosensory evoked potentials are less sensitive and less specific for MS than is high-resolution MRI. Multifocal VEPs may be more sensitive than global VEPs in revealing focal areas of abnormal conduction along the optic nerve.

Optical Coherence Tomography

Optical coherence tomography is performed with an office-based device that uses the reflection of infrared light (from an exogenous source directed

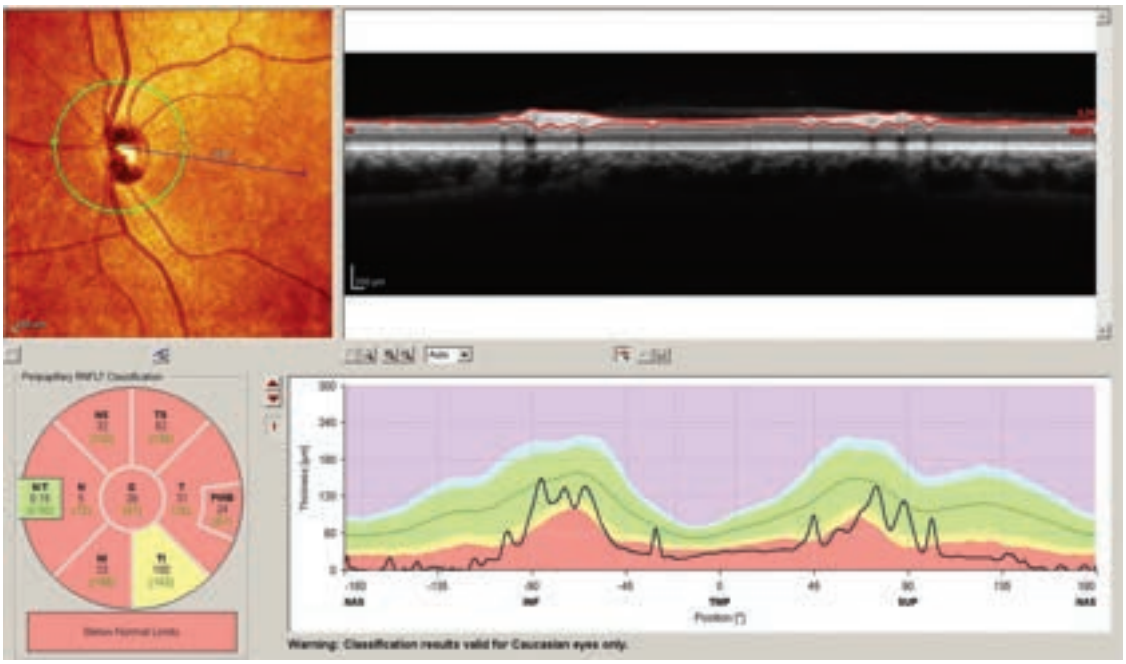

FIGURE 411-6. High-resolution spectral domain optical coherence tomography scan. The scan is from the retina of a patient with multiple sclerosis and history of optic neuritis in the scanned eye. The *upper left* is a fundus photo; the *upper right* is the tomogram map of the peripapillary retinal nerve fiber layer. *Lower panels* show regional quantitative data and color maps based on percentile compared to age- and sex-matched healthy controls (*green* is 5th to 95th percentile, *yellow* is 5th percentile, and *red* is first percentile).

through the pupil) off the back of the eye to quantify the thickness of retinal tissues, including the peripapillary retinal nerve fiber layer and macular layers. This test, which has been widely used in glaucoma, can monitor axonal and retinal ganglion cell damage, both in the setting of acute optic neuritis and in detecting subclinical neuroaxonal damage (Fig. 411-6). Retinal nerve fiber layer thinning correlates with brain atrophy and may be useful as a surrogate marker of more global neurodegeneration in MS.

Differential Diagnosis

The diagnosis of MS may be so clear that it is recognized by the patient and is readily confirmed by the primary physician or so obscure that even experienced specialists disagree.[8] Many processes (see Table 411-2) can mimic the clinical, radiologic, and CSF findings associated with MS, and there is no "gold standard" diagnostic test that is 100% sensitive and specific for the disease.

Processes that mimic MS include structural lesions, especially of the base of the brain and of the spinal cord, in which one lesion can cause symptoms referable to many different tracts and at different perceived locations in the body. Chiari malformations with or without syrinx (Chapter 417), disc herniation (Chapter 400), cervical spondylosis, and low-grade tumors (Chapter 189) can produce symptoms of MS both in newly presenting patients and in patients who truly have MS but who also have a second process.

Various infectious diseases can mimic MS. Examples include human T-cell lymphotropic virus types I and II (virally associated myelopathy or tropical spasticparaparesis;Chapter378),humanimmunodeficiencyvirus(neuropathy, myelopathy, cognitive impairment, CNS white matter changes; Chapter 394), neuroborreliosis (Lyme disease; Chapter 321), neurosyphilis (Chapter 319), Epstein-Barr virus (Chapter 377), cytomegalovirus (Chapter 376), herpes simplex virus (Chapter 374), varicella-zoster virus myelitis (Chapter 375), and JC virus (progressive multifocal leukoencephalopathy; Chapter 370).

Inflammatory diseases that usually involve other parts of the body can concomitantly affect or, rarely, manifest in the CNS. Examples include sarcoidosis (Chapter 95), systemic lupus erythematosus (Chapter 266), Sjögren syndrome (Chapter 268), and vasculitides (Chapter 270). Ischemic vascular disease secondary to any cause also can resemble MS. Metabolic and nutritional disorders that can mimic MS include vitamin B_{12} deficiency and methylmalonic acidemia (in some cases distinct from cyanocobalamin deficiency). Rarely, central pontine myelinolysis (Chapters 116 and 416) is mistaken for MS. Thyroid disease (Chapter 226) may mimic the fatigue of MS and may cause dysesthesias and disorders of the optic nerve and muscles. Nutritional deficiency (Chapter 215) and malabsorption have been associated with demyelination and may mimic MS. Copper deficiency can cause dorsal column pathology, neuropathy, anemia, and optic neuropathy. Vitamin D deficiency (Chapter 244), which is becoming increasingly common, can cause proximal weakness, fatigue, asthenia, bone loss, and impaired immune function. Vitamin A deficiency, although not common in industrialized countries, can cause night blindness and immune dysfunction.

Monophasic demyelinating syndromes with or without multiple other lesions often, but not always, progress to become MS (see later). Spinocerebellar atrophy and hereditary myelopathy cause slowly progressive disease but do not cause sensory and visual abnormalities.

Hereditary diseases are increasingly recognized as mimicking MS. Spinocerebellar atrophy may manifest as progressive myelopathy and ataxia. A variety of genetic neuropathies (type 2 CMT mitofusinopathies, adult-onset polyglucosan body disorder; Chapter 420), ataxias (Friedreich, ataxia telangiectasia; Chapter 410), mitochondrial diseases (progressive optic atrophy, Leber, MELAS, MERRF; Chapter 421), and metabolic diseases (urea cycle disorders; Chapter 205) can have CNS manifestations that could lead to misdiagnosis.

TREATMENT

Rx

The treatment of MS can be divided into drugs designed to relieve symptoms, drugs designed to modify the course of the disease, and nondrug measures.[9] Numerous drugs target specific aspects of MS: depression, fatigue, muscle spasticity, pain, insomnia, and bladder, bowel, and sexual dysfunction. Before considering a symptomatic therapy, the patient should be educated about the purpose of the drug and its side-effect profile. On learning that these drugs have no long-term impact on disease activity, patients may elect not to use them for relief of symptoms alone. Symptomatic therapies are best started at low doses and frequently require titration to obtain the optimal balance between efficacy and side effects.

Treatment of Specific Symptoms

Depression and emotional lability are common symptoms of MS. In addition to appropriate supportive care and counseling, antidepressant therapy with one of the "activating" serotonergic or noradrenergic drugs (fluoxetine, sertraline, citalopram, escitalopram, venlafaxine, or bupropion) can be of benefit (see Table 397-5). If anxiety and panic symptoms predominate, a less activating drug such as paroxetine may be preferable. Patients with pain or insomnia may benefit more from a sedating antidepressant (amitriptyline, nortriptyline, or trazodone) given at bedtime, which may have the added anticholinergic benefits on urinary bladder urgency.

Spasticity can be managed by physical therapy, stretching, and institution of either baclofen (5 to 160 mg in divided doses) or tizanidine (2 to 32 mg in divided doses). Either drug should be started as a single agent at a low dose at bedtime, gradually increasing to three to four times daily, with a larger dose

at bedtime to target nocturnal symptoms. Decreasing muscle tone can result in weakness. Baclofen should never be discontinued abruptly because of the potential for a severe withdrawal reaction.

Bladder urgency resulting from detrusor muscle spasm can be managed effectively with anticholinergics such as oxybutynin (5 to 20 mg in divided doses) or tolterodine (1 to 4 mg) or focal injections of botulinum toxin, but these agents can cause temporary urinary hesitancy or retention. Bladder ultrasonography permits accurate bedside assessment of postvoid residual volume to determine whether a patient is retaining excessive amounts of urine. Urinary retention may be improved by removing drugs known to induce it (e.g., anticholinergics and opioids). Primary urinary retention is difficult to treat with drugs, but external sphincter spasm can be treated with α_{1a}-adrenergic receptor blockers such as tamsulosin (0.4 to 0.8 mg) and doxazosin (1 to 8 mg). Bethanechol (10 to 150 mg in divided doses) may be tried for an atonic bladder, but intermittent catheterization is often required. Alternative causes of bladder symptoms such as urinary tract infections, prostatic enlargement, or anatomic changes following pregnancy should be considered and managed separately. Prolonged urinary retention predisposes to infections, structural damage to the bladder and kidneys, and malignancy. Persistent postvoiding residual volumes greater than 300 cc should be treated medically, with intermittent straight catheterization recommended for patients with large volume retention refractory to medical therapy.

Painful dysesthesias and paroxysmal dystonic spasms may be managed effectively with antiepileptic drugs (gabapentin, 300 to 5400 mg/day in divided doses; pregabalin, 75 to 600 mg/day in divided doses; or carbamazepine, 100 to 2400 mg/day in divided doses) or tricyclic antidepressants (amitriptyline, 10 to 150 mg; or nortriptyline, 10 to 50 mg). Patients with trigeminal neuralgia (Chapter 398) may respond to these drugs or to baclofen, misoprostol, botulinum toxin, or decompression surgery.

Sexual dysfunction in MS is often multifactorial. Patients with erectile dysfunction usually respond well to the phosphodiesterase inhibitors, which enhance penile vasodilation (Chapter 234). Education regarding the use of lubrication, alternative sensory stimulation, and the adverse effect of heat, can improve sexual function.

Symptoms related to heat sensitivity may improve on cooling. Cooling devices can prevent this phenomenon, but there is no persistent benefit of inducing hypothermia.

Systemic Treatments

Corticosteroids (e.g., methylprednisolone, 1 g/day IV for 3 to 5 days) shorten the duration and severity of symptoms from an acute exacerbation but have no proved effect on long-term disability. Oral corticosteroids used in equivalent dosage are probably equally efficacious and safe. Intravenous immunoglobulin and plasma exchange may occasionally to benefit steroid-refractory patients, but large randomized placebo-controlled trials in relapsing MS have failed to show consistent benefits, perhaps because only patients with type II disease (humoral component) are likely to respond.

Approved Disease-Modifying Treatments

Twelve disease-modifying agents have been approved by the U.S. Food and Drug Administration (FDA): IFN-β1b (Betaseron and Extavia), IFN-β1a (Avonex), IFN-β1a (Rebif), pegylated interferon (Plegridy), glatiramer acetate (Copaxone), natalizumab (Tysabri), alemtuzumab (Lemtrada), mitoxantrone (Novantrone), fingolimod (Gilenya), teriflunomide (Aubagio), and dimethyl fumarate (Tecfidera). All of these agents are approved for relapsing-remitting MS, and mitoxantrone is indicated for worsening forms of MS and for secondary progressive MS.

The four IFN-β drugs reduce the relapse rate by approximately one third. IFN-β1b (8 million international units [IU], subcutaneously [SC] every other day [Betaseron and Extavia]) and IFN-β1a (30 μg intramuscularly [IM] weekly [Avonex] or 22 to 44 μg SC three times weekly [Rebif]) appear to have a more rapid onset of action, perhaps based on their dosing regimen, compared with weekly IFN-β1a (30 μg IM weekly). However, weekly IFN-β1a Avonex is less immunogenic and results in only a 3% incidence of neutralizing antibodies, which reduce efficacy, compared with 20 to 30% for the other IFN-β preparations. The major side effects of IFN-β are a flulike reaction (low-grade fever, chills, and myalgias 6 to 24 hours after the injection), local reactions at the injection site (pain, erythema, and rarely necrosis), and elevated aminotransferase levels (rarely severe hepatitis). These side effects can be managed by initiating the drug slowly and by prophylaxis with acetaminophen and nonsteroidal anti-inflammatory agents, and they improve in most patients after 3 to 6 months. A long-acting pegylated version of IFN-β1a dosed at 125 μg SC every 2 weeks reduced the annualized relapse rate by 38% and gadolinium-enhancing MRI lesions by 82% versus placebo in a phase 3 trial. Side effects remained typical of the interferon-β drugs and were not prolonged.

Glatiramer acetate is a copolymer of four amino acids designed to mimic myelin basic protein; given as 20 mg/day SC or as 40 mg SC three times a week, it also reduces relapses by about one third and is well tolerated by most patients. Major side effects are local reactions at the injection site, (swelling, hives, and delayed lipoatrophy), and a rare, self-limited (15 to 20 minutes) systemic reaction consisting of chest pain, palpitations, and anxiety. No monitoring of blood tests is required for this medication. The effect of glatiramer acetate on MRI T2-weighted and gadolinium-enhancing lesions is less dramatic than for the interferons (30% reduction), perhaps because its primary effect is not at the blood-brain barrier.

Natalizumab is a monoclonal antibody directed against the α_4-integrin chain of the leukocyte adhesion molecule VLA-4. In a large phase 3 trial, this drug, at a dose of 300 mg intravenously (IV) every 4 weeks, reduced relapses by 68% compared with placebo and reduced gadolinium-enhancing lesions by 92%. However, approximately 1 in 500 patients develop JC virus brain infection (Chapter 370) after 24 months of exposure, which causes progressive multifocal leukoencephalopathy (PML). The risk for PML appears to be in higher in patients with JC virus serum antibody titers greater than 0.9 compared with low-titer or seronegative patients, although the results may change over time so patients require repeated testing.

Alemtuzumab, a monoclonal antibody that targets CD52 on lymphocytes and monocytes, reduces annualized relapse rates by approximately 50% and also can reduce the progression of disability compared with IFN-β1a. The drug is given as yearly courses at 12 or 24 mg daily for 5 consecutive days in year 1 and for 3 days in years 2 and 3. Serious side effects associated with alemtuzumab include a 20 to 25% risk for developing autoimmune thyroid disease and rare cases of immune thrombocytopenic purpura (Chapter 172), autoimmune hemolytic anemia (Chapter 160), autoimmune neutropenia (Chapter 167), and Goodpasture syndrome (Chapter 121).

Mitoxantrone, which is an anthracenedione antineoplastic agent with potent immunosuppressive activity, is approved to slow the progression of neurologic disability and reduce the relapse rate in patients with relapsing-remitting MS and secondary progressive MS. The recommended dose is 5 to 10 mg/m^2 intravenous infusion every 3 months, and the lifetime use of this drug is limited to 2 to 3 years (or a cumulative dose of 120 to 140 mg/m^2) because of its cardiotoxicity.

Fingolimod is a sphingosine-1 phosphate receptor modulator that is given orally once daily at 0.5 mg. Fingolimod reduces relapse rates and the progression of disease compared with placebo and compared with interferon therapy. However, it generally is not considered first-line therapy because its risks, including type 2 heart block and herpes encephalitis, are concerning, especially in young, otherwise healthy people with MS. Fingolimod side effects include first-dose bradycardia, macular edema, and respiratory infections.

Teriflunomide is approved as an oral agent for MS based on two phase 3 trials in which the annualized relapse rate was reduced by 31% compared with placebo, and disability also was reduced. It is dosed at 7 or 14 mg/day PO and requires monitoring of blood tests for rare liver and kidney toxicities.

Oral dimethyl fumarate (240 mg twice daily) reduces the annualized relapse rate (49 to 55%) and disease progression in addition to suppressing active MRI lesions. A differently formulated drug that is a combination of monomethyl fumarate and dimethyl fumarate has been marketed in Germany for psoriasis since 1994. Four cases of PML have been reported in patients who had received the German formulation and were either exposed to other drugs associated with PML or had alternative risk factors such as preexisting or drug-induced leukopenia. Dimethyl fumarate results in severe leukopenia in 3% of patients.

No specific treatment algorithm can be recommended because the disease is heterogeneous, there are few head-to-head studies between medications, and reported effect sizes of the approved drugs versus placebo depend on the varying characteristics of patients in different trials. As a result, the treatment decision is best made in conjunction with the patient based on the patient's disease, the side-effect and safety profiles of the various drugs, and the physician's assessment of severity and prognosis.[10] Older drugs such as IFN-β or glatiramer acetate have provided variable responses, with some patients doing well for many years. For the typical newly diagnosed patient with relapsing MS but no early signs of poor prognosis (e.g., high relapse rate with early accrual of disability, African American ancestry, high lesion load, T1 black holes on MRI, multiple spinal cord lesions), one reasonable strategy is to start with one of the older, relatively safe drugs such as IFN-β1b or glatiramer acetate and then switch to the other if the patient experiences a severe relapse, multiple small relapses, or new MRI lesions, before escalating to one of the newer, more potent drugs associated with more potential risk.

Other Therapies

Rituximab (1000 mg IV 2 weeks apart repeated every 6 months) is a monoclonal antibody that depletes B-cell lymphocytes and can significantly reduce inflammatory brain lesions and relapse by approximately 50% for up to 48 weeks in patients with relapsing-remitting MS. However, it has not been effective in primary progressive MS. Newer, fully humanized anti-CD20 monoclonal antibodies (ocrelizumab and ofatumumab) have shown equal or better efficacy in phase 2 trials of relapsing MS.

Daclizumab (an anti-CD52 [IL-2 receptor-α] monclonal antibody dosed at 150 to 300 mg SC every 4 weeks) reduced annualized relapse rate by 50 to 60% and MRI activity in two phase 2 trials A. Cladribine (2-chlorodeoxyadenosine, 3.5 or 5.25 mg/kg/day) given as a short course once per year reduced the annualized relapse rate by 55% and disease progression

by one third compared with placebo in a phase 3 trial.[■] Both these drugs may have a role in treating MS, but they also have global immunosuppressive effects that increase the risk for serious infections and possibly other systemic complications.

In a phase 2 trial, treatment with 0.6 mg/day of laquinimod, a novel oral immunomodulatory agent, resulted in a 40% decrease in MRI lesions in relapsing-remitting multiple sclerosis at 36 weeks.[■] In a subsequent large randomized trial of patients with relapsing-remitting multiple sclerosis, laquinimod (0.6 mg/day PO) reduced the relapse rate, the progression of disability, and the MRI lesions.[■] However, this drug has not been approved in the United States because of concerns about its long-term side effects.

In a randomized trial, sustained-release dalfampridine, a potassium-channel blocker at 10 mg twice daily, improved walking in 35% of patients compared with only 8% of patients receiving placebo, leading to its approval by the FDA as a symptomatic therapy for MS.[■] Although only a subset of patients appear to benefit, the effect is fairly rapid and likely extends beyond just ambulation in those who are responders. Dalfampridine is contraindicated in patients with a history of seizures and may cause dizziness, insomnia, and an increase in paresthesias.

Other forms of immunosuppression, including methotrexate, azathioprine, mycophenolate mofetil, and cyclophosphamide, may have some efficacy in MS, although either no definitive clinical trials have been done with these agents or safety profiles have outweighed risks, and none is as yet approved for MS by the FDA.

Other Approaches to Well-Being

Patients with MS are at high risk for developing osteopenia or osteoporosis (Chapter 243), so prophylaxis with vitamin D and calcium and treatment with bisphosphonates or other proved approaches should be considered. Patients with suboptimal 25-OH vitamin D levels (<30 ng/mL) on standard 1000-IU cholecalciferol replacement therapy should consider increasing to 4000 to 5000 IU/day or 50,000 IU every other week or, in some cases, weekly, with appropriate monitoring of vitamin D and serum plus urine calcium levels. If osteoporosis has already been diagnosed, bisphosphonate therapy, such as alendronate (10 mg/day or 70 mg/week) or a similar drug, is generally indicated.

Nonmedical treatment of MS is a critical part of managing the disease. Patients derive benefit from a health care team approach consisting of an experienced MS physician, nurse, social worker, therapist, and counselor, with appropriate referral to other subspecialties as needed. Alternative and complementary therapies (Chapter 39) are commonly used by patients with MS, and the risks and benefits of these approaches must be discussed with the patient.

PROGNOSIS

The average lifespan of patients with MS is approximately 8 years less than normal, a finding reflecting a bimodal distribution in which many patients live a normal lifespan and a few die earlier owing to aggressive disease, severe disability, infection, or suicide.[11] Most patients presenting with relapsing-remitting MS convert to secondary progressive MS after 20 to 40 years. Only one third of patients will require use of a wheelchair, but 50% may need assistive devices and nearly two thirds will have disability that prevents them from working. African Americans and men of all races tend to have a more aggressive course and are more likely to become disabled. Immunomodulating therapy early in the course of the disease appears to slow progression of disability, but long-term follow-up data are open-label and uncontrolled, so it is difficult to quantify the extent of this benefit.

OTHER DISEASES OF MYELIN

Monofocal and Monophasic Demyelinating Processes

OPTIC NEURITIS AND TRANSVERSE MYELITIS

Optic neuritis (Chapter 424) and transverse myelitis are inflammatory processes that can occur as entities distinct from MS or as part of MS (see earlier).[12] In addition, optic neuritis and transverse myelitis can occur together in the syndrome called *neuromyelitis optica* (Devic disease).

Optic Neuritis

Optic neuritis (Chapter 424) is an inflammatory disease that usually involves the retrobulbar portion of the optic nerve and sometimes parts of the optic chiasm. Although optic neuritis is most often associated with MS (50 to 75%), it also can be seen as an isolated idiopathic disorder (25 to 50%), as part of neuromyelitis optica, or associated with other inflammatory and infectious diseases such as chronic relapsing inflammatory optic neuropathy, systemic lupus erythematosus, Sjögren syndrome, sarcoidosis, Lyme disease, syphilis, and human immunodeficiency virus infection. The pathobiologic features are thought to be similar to those of MS and are characterized by idiopathic inflammatory demyelination followed by secondary axonal injury. Hereditary optic neuropathies may be unmasked during periods of stress and manifest as an acute monocular visual loss.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The clinical presentation, which typically is monocular visual loss with pain over the brow that worsens with lateral eye movement, is similar regardless of whether it manifests as part of MS (see the earlier discussion of the visual effects of MS) or not. When it involves the optic nerve head, it is called *papillitis* and, in bilateral cases, can be impossible to differentiate from papilledema. Optic neuritis also can be mimicked by anterior segment, choroidal, or retinal diseases. Optic neuritis is distinguished from optic neuropathy, which is a chronic, generally noninflammatory condition of the optic nerve caused by tobacco or nutritional amblyopia, ischemia, Leber disease, Charcot Marie Tooth type 2a (mitofusinopathy; Chapter 420), or a number of other rare hereditary diseases (Chapter 424). Subclinical optic neuropathy in the absence of painful monocular visual loss may result in retinal nerve fiber layer thinning over time.

TREATMENT

Rx

Among patients with optic neuritis, the 15-year risk for developing MS is 25% in patients without lesions on their baseline brain MRI but 72% in patients with one or more baseline MRI lesions. Treatment with intravenous methylprednisolone as in MS may shorten the duration and severity of the attack, but no definitive evidence indicates that it changes the long-term outcome. Oral prednisone alone, without prior treatment with intravenous methylprednisolone, may increase the risk for recurrent optic neuritis and should be avoided. Data support the use of IFN-β drugs and glatiramer acetate in patients whose optic neuritis is at high risk for conversion to MS (one or more typical brain MRI lesions).

Transverse Myelitis

Transverse myelitis is a rare (~1 in 100,000 people) monophasic inflammatory process of the spinal cord that is usually distinct from MS in that it either involves the entire cross section or is longitudinally extensive along three vertebral body segments rostrocaudally.[13] Transverse myelitis or myelopathy may be idiopathic or associated with inflammatory diseases (systemic lupus erythematosus, Sjögren syndrome, vasculitis, or MS), infectious diseases, or vascular diseases (antiphospholipid antibody syndrome or dural venous fistula).

CLINICAL MANIFESTATIONS AND DIAGNOSIS

In its fulminant form, transverse myelitis causes complete loss of motor and sensory function below the affected level of the spinal cord and causes concomitant bowel, bladder, and sexual dysfunction. Autonomic involvement can be seen in cervical and high thoracic spine cases. Transverse myelitis also may manifest in an incomplete or partial form, which is more commonly associated with MS. In older patients, patients with vascular risk factors, or patients with central cord edema pattern on MRI, spinal angiography should be considered to exclude spinal cord ischemia or infarction (Chapter 400).

TREATMENT AND PROGNOSIS

Rx

Treatment of the inflammatory process is usually with methylprednisolone (1000 mg IV for 3 to 5 days), followed by specific treatment of any identifiable underlying disease process. The prognosis is worse than in MS in that significant recovery is seen in fewer than 50% of patients and many patients remain completely paralyzed after the initial attack. Plasma exchange or cyclophosphamide may be considered in steroid-refractory cases.

NEUROMYELITIS OPTICA

Neuromyelitis optica (NMO) is now recognized as an entity distinct from MS and is characterized by an optic neuritis, often bilateral and temporally associated with a fulminant multilevel transverse myelitis. A specific serum IgG (NMO-IgG) directed against aquaporin 4 strongly predicts this process. Brain lesions may be seen on MRI and have a predilection for the brain stem. Neuromyelitis optica may be similar to what is called *opticospinal MS* in Japan, although the latter overlaps with MS. There is no proved effective treatment, but patients are usually given anti-inflammatory and immunosuppressive medications (e.g., azathioprine 2 to 3 mg/kg or prednisone 1 mg/

kg). Therapies directed against B cells (anti–CD-20 or anti-CD19 monoclonal antibodies), humoral factors (complement), or nonpathogenic antibody blockers of aquaporin 4–IgG binding also have shown efficacy, but no placebo-controlled trials have been completed as yet. The prognosis is generally poor; if not treated, most patients develop sustained disabling visual loss and weakness.

ACUTE DISSEMINATED ENCEPHALOMYELITIS

Acute disseminated encephalomyelitis and its hyperacute form, acute necrotizing hemorrhagic encephalopathy, are thought to be forms of monophasic immune-mediated inflammatory demyelination. They differ from MS in that they are typically monophasic, whereas MS is by definition multiphasic or chronically progressive. However, no reliable clinical or pathologic criteria are available to differentiate the two processes, which may represent a continuum. Patients may present with fever, headache, meningeal signs, and altered consciousness, which are exceedingly rare in MS. There is no known effective treatment. Large numbers of patients, especially children, make remarkable recoveries, but the necrotizing form can be severely disabling or fatal. Relapsing forms of the disease in children are more likely to become MS.

Leukodystrophies

The leukodystrophies represent a variety of diseases formerly characterized by their common clinical and pathologic characteristics of white matter and, presumably, myelin. Many of these diseases now have a defined biochemical and genetic basis, and some (e.g., Alexander disease) are no longer considered dysmyelinating diseases.[14]

ADRENOLEUKODYSTROPHY AND ADRENOMYELONEUROPATHY

Adrenoleukodystrophy and adrenomyeloneuropathy, which are caused by impaired ability of the peroxisomes to metabolize very-long-chain fatty acids, represent different phenotypes resulting from the same X-linked, incompletely recessive genetic defect. Impaired oxidation of very-long-chain fatty acids results from deficient function of the enzyme lignoceroyl–coenzyme A ligase. The defective gene maps to Xq28 and codes for a peroxisomal membrane protein (ALDP), which is a member of a large family of proteins referred to as the adenosine triphosphate–binding cassette (ABC) transporters, specifically *ABCD1*.

Childhood cerebral adrenoleukodystrophy, which is the most common form of the disorder, represents 45% of all cases; it is seen only in male patients, with an onset at ages 4 to 11 years. Adolescent (5%) and adult (3%) cerebral forms progress at a similar or slower rate than the childhood form.

CLINICAL MANIFESTATIONS

Adrenomyeloneuropathy begins in young men as slowly progressive paraparesis with hypogonadism, impotence, sphincter disturbances, variable adrenal insufficiency, and axonal neuropathy affecting mainly the lower extremities. A rare acute inflammatory form with rapid progression and dementia may occur. A similar, but usually milder, disorder can be seen in up to 20% of women who are hemizygous for the disease.

DIAGNOSIS

Diagnosis is established in male patients by finding elevated very-long-chain fatty acids in the plasma. DNA-based diagnosis in carriers is reliable and is recommended in women because of false-negative results using the plasma assay.

TREATMENT Rx

Treatment is unsatisfactory. A 4:1 mixture of glyceryl trioleate and glyceryl trierucate (i.e., "Lorenzo oil") normalizes plasma very-long-chain fatty acids within 4 weeks and has few side effects. Although clinical trials suggested that treatment in presymptomatic patients delayed or prevented the onset of disease, this treatment is ineffective after symptoms have begun and the disease progresses relentlessly.

Pelizaeus-Merzbacher Disease

Pelizaeus-Merzbacher disease is a rare, chronic, familial leukodystrophy usually caused by a genetic defect in the X-linked myelin proteolipid protein (PLP) gene. In classic Pelizaeus-Merzbacher disease, age at onset varies between 3 months and 9 years, and the age at death varies between 6 years and 25 years. However, milder forms of spastic paraplegia 2 are now well recognized in adults. The disease manifests as a slowly progressive myelopathy, often with cerebellar and cognitive involvement, and the diagnosis is established by genetic testing for mutations in the *PLP* gene. A variety of different types of PLP mutations account for the variability in clinical phenotypes. An autosomal recessive disease called Pelizaeus-Merzbacher–like disease 1 and the less-severe spastic paraplegia 44, caused by mutations of the gap junction protein gamma-2 gene *(GJC2),* are recognized variants. No specific treatment exists beyond supportive therapy.

Metachromatic Leukodystrophy

Metachromatic leukodystrophy usually results from a recessively inherited defect in the lysosomal enzyme arylsulfatase A. Absence of arylsulfatase A results in the accumulation of sulfatide in both central and peripheral myelin and myelin-forming cells; instability of the myelin membranes results in the breakdown of myelin. Metachromatic leukodystrophy is generally divided into four subtypes: congenital, late infantile (most common), juvenile, and adult. It appears in all ethnic groups and has an overall frequency of 1 in 40,000.

The clinical manifestations are variable and may include progressive spastic paraparesis, extrapyramidal signs, seizures, and peripheral neuropathy. Brain MRI usually shows large confluent symmetrical high-signal areas in the cerebral white matter, brain stem, and cerebellum, but a more patchy appearance resembling MS is occasionally seen in adult cases. At present, no satisfactory treatment exists. Some evidence suggests that bone marrow transplantation delays the onset in presymptomatic patients and may slow progression of the disease.

Globoid Cell Leukodystrophy

Globoid cell leukodystrophy (Krabbe disease; Chapter 208) is characterized biochemically by accumulation of galactocerebroside in cerebral white matter as a result of deficient galactocerebroside β-galactosidase activity. The disease is transmitted as an autosomal recessive trait and affects infants in the first 2 to 3 months of life, initially manifesting with behavioral changes and failure to achieve developmental milestones. Rare late-onset cases manifest with progressive motor impairment and, less frequently, visual failure. Neuropathologic examination reveals marked loss of myelin throughout the brain, with the presence of round or oval macrophages and large, irregular, multinucleated cells, termed *globoid cells,* that are filled with galactocerebroside. Accumulation of galactosylsphingosine (psychosine) is thought to cause destruction of oligodendrocytes and marked reduction of myelin formation.

Canavan Disease

Canavan disease is a fatal, progressive leukodystrophy with an autosomal recessive inheritance, caused by mutations in the gene for aspartoacylase, an enzyme that hydrolyzes *N*-acetylaspartate into L-aspartate and acetate. Aspartoacylase deficiency results in elevated levels of its substrate molecule, *N*-acetylaspartate, brain edema, and dysmyelination. Clinically, the disease manifests with retardation, seizures, and diffuse, symmetrical white matter degeneration in the subcortical areas, with involvement of the globus pallidum on MRI. No treatment is available.

Vanishing White Matter Disease

Vanishing white matter disease is an increasingly recognized autosomal recessive disorder with a broad range of clinical manifestations from rapidly progressive presentations in infants to slowly progressive disease in adults. The disease is caused by mutations in the eukaryotic translation initiation factor 2B (*eIF2B*) genes 1 to 5, which code for proteins involved in the integrated stress response of cells. Pathologic characteristics include vacuolated myelin with cystic appearance on MRI. No specific therapy, other than avoidance of stress, is known.

Grade A References

A1. Filippini G, Del Giovane C, Vacchi L, et al. Immunomodulators and immunosuppressants for multiple sclerosis: a network meta-analysis. *Cochrane Database Syst Rev.* 2013;6:CD008933.

A2. Calabresi PA, Kieseier BC, Arnold DL, et al. Pegylated interferon beta-1a for relapsing-remitting multiple sclerosis (ADVANCE): a randomised, phase 3, double-blind study. *Lancet Neurol.* 2014;13:657-665.

A3. Johnson KP, Brooks BR, Cohen JA, et al. Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. *Neurology.* 1995;45:1268-1276.

A4. Polman C, O'Connor PW, Havrdovra E, et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med.* 2006;354:899-910.

A5. Cohen JA, Coles AJ, Arnold DL, et al. Alemtuzumab versus interferon beta 1a as first-line treatment for patients with relapsing-remitting multiple sclerosis: a randomised controlled phase 3 trial. *Lancet*. 2012;380:1819-1828.
A6. Coles AJ, Twyman CL, Arnold DL, et al. Alemtuzumab for patients with relapsing multiple sclerosis after disease-modifying therapy: a randomised controlled phase 3 trial. *Lancet*. 2012; 380:1829-1839.
A7. Hartung H-P, Gonsette R, Konig N, et al. Mitoxantrone in progressive multiple sclerosis: a placebo-controlled, double-blind, randomized, multicentre trial. *Lancet*. 2002;360:2018-2025.
A8. Kappos L, Radue EW, O'Connor P, et al. A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. *N Engl J Med*. 2010;362:387-401.
A9. Calabresi PA, Radue EW, Goodin D, et al. Safety and efficacy of fingolimod in patients with relapsing-remitting multiple sclerosis (FREEDOMS II): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Neurol*. 2014;13:545-556.
A10. O'Connor P, Wolinsky JS, Confavreux C, et al. Randomized trial of oral teriflunomide for relapsing multiple sclerosis. *N Engl J Med*. 2011;365:1293-1303.
A11. Fox RJ, Miller DH, Phillips JT, et al. Placebo-controlled phase 3 study of oral BG-12 or glatiramer in multiple sclerosis. *N Engl J Med*. 2012;367:1087-1097.
A12. Gold R, Kappos L, Arnold DL, et al. Placebo-controlled phase 3 study of oral BG-12 for relapsing multiple sclerosis. *N Engl J Med*. 2012;367:1098-1107.
A13. Castillo-Trivino T, Braithwaite D, Bacchetti P, et al. Rituximab in relapsing and progressive forms of multiple sclerosis: a systematic review. *PLoS ONE*. 2013;8:e66308.
A14. Gold R, Giovannoni G, Selmaj K, et al. Daclizumab high-yield process in relapsing-remitting multiple sclerosis (SELECT): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2013;381: 2167-2175.
A15. Giovannoni G, Comi G, Cook S, et al. A placebo-controlled trial of oral cladribine for relapsing multiple sclerosis. *N Engl J Med*. 2010;362:416-426.
A16. Comi G, Jeffery D, Kappos L, et al. Placebo-controlled trial of oral laquinimod for multiple sclerosis. *N Engl J Med*. 2012;366:1000-1009.
A17. Vollmer TL, Sorensen PS, Selmaj K, et al. A randomized placebo-controlled phase III trial of oral laquinimod for multiple sclerosis. *J Neurol*. 2014;261:773-783.
A18. Goodman AD, Brown TR, Krupp LG, et al. Sustained-release oral fampridine in multiple sclerosis: a randomised, double-blind, controlled trial. *Lancet*. 2009;373:732-738.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Saab AS, Tzvetanova ID, Nave KA. The role of myelin and oligodendrocytes in axonal energy metabolism. *Curr Opin Neurobiol*. 2013;23:1065-1072.
2. Kurtzke JF. Epidemiology in multiple sclerosis: a pilgrim's progress. *Brain*. 2013;136:2904-2917.
3. Beecham AH, Patsopoulos NA, Xifara DK, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet*. 2013;45:1353-1360.
4. Lassmann H. Multiple sclerosis: Lessons from molecular neuropathology. *Exp Neurol*. 2014; 262:2-7.
5. Houtchens M. Multiple sclerosis and pregnancy. *Clin Obstet Gynecol*. 2013;56:342-349.
6. Deangelis TM, Miller A. Diagnosis of multiple sclerosis. *Handb Clin Neurol*. 2014;122:317-342.
7. Tillema JM, Pirko I. Neuroradiological evaluation of demyelinating disease. *Ther Adv Neurol Disord*. 2013;6:249-268.
8. Gelfand JM. Multiple sclerosis: diagnosis, differential diagnosis, and clinical presentation. *Handb Clin Neurol*. 2014;122:269-290.
9. Perry M, Swain S, Kemmis-Betty S, et al. Multiple sclerosis: summary of NICE guidance. *BMJ*. 2014;349:g5701.
10. Wingerchuk DM, Carter JL. Multiple sclerosis: current and emerging disease-modifying therapies and treatment strategies. *Mayo Clin Proc*. 2014;89:225-240.
11. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology*. 2014;83:278-286.
12. Petzold A, Plant GT. Diagnosis and classification of autoimmune optic neuropathy. *Autoimmun Rev*. 2014;13:539-545.
13. Cree BA. Acute inflammatory myelopathies. *Handb Clin Neurol*. 2014;122:613-667.
14. Perlman SJ, Mar S. Leukodystrophies. *Adv Exp Med Biol*. 2012;724:154-171.

REVIEW QUESTIONS

1. A definite diagnosis of multiple sclerosis can be made in which of the following clinically isolated syndromes in which no better explanation can be found:

A. In a case of optic neuritis in a young woman with fatigue.
B. In a case of transverse myelitis in a young woman with CSF oligoclonal bands.
C. In a case of a 50-year-old woman with paresthesias in her feet and visual symptoms, if the MRI shows more than three white matter lesions none of which enhances after gadolinium.
D. In a case of 25-year-old patient with an internuclear ophthalmoplegia and an MRI that shows nine periventricular and juxtacortical white matter lesions with and without gadolinium enhancement.
E. In the case of a young woman with depression, fatigue, and CSF pleocytosis

Answer: D The diagnosis of multiple sclerosis can be made in a patient with a clinically isolated syndrome and no other better explanation only if the MRI fulfills criteria for dissemination in space and time (enhancing and nonenhancing lesions are now interpreted as dissemination in time).

2. A patient with established multiple sclerosis reports urinary frequency and urgency resulting in incontinence twice a month. The best advice is to:

A. Check a urinalysis for infection, exclude urinary retention by checking a postvoid residual, and consider oxybutynin or other bladder medication to suppress detrusor muscle spasm.
B. Check a urinalysis for infection, exclude urinary retention by checking a post-void residual, and consider tamsulosin or similar bladder medication to relax external sphincter muscle spasm.
C. Recommend botulinum toxin injections without further invasive testing.
D. Recommend that this patient learn how to perform intermittent straight catheterization to empty the bladder.
E. Treat for urinary tract infection with antibiotics and then reassess.

Answer: A A urinary tract infection and component of urinary bladder retention first need to be excluded, and then the patient should be considered for treatment with drugs that relax the detrusor muscle contraction that is likely mediating her urgency and incontinence. Botox is a reasonable alternative but only after retention and infection are excluded. Tamsulosin might relieve bladder retention but will not help detrusor muscle spasm.

412

MENINGITIS: BACTERIAL, VIRAL, AND OTHER

AVINDRA NATH

BACTERIAL MENINGITIS

DEFINITION

Meningitis is an inflammation of the arachnoid membrane, the pia mater, and the intervening cerebrospinal fluid (CSF). The inflammatory process extends throughout the subarachnoid space around the brain and spinal cord and involves the ventricles. Pyogenic meningitis is usually an acute bacterial infection that evokes a polymorphonuclear response in CSF. By comparison, tuberculous meningitis (Chapter 324) is often subacute and characterized initially by a modest polymorphonuclear pleocytosis that rapidly evolves to lymphocytic predominance.

EPIDEMIOLOGY

The incidence of bacterial meningitis has dropped dramatically in developed countries since the introduction of vaccines against bacterial pathogens such as *Haemophilus influenzae* type b (Chapter 300), *Streptococcus pneumoniae* (Chapter 289), and *Neisseria meningitidis* (Chapter 298).[1] Since the advent of the *Haemophilus* vaccine, the incidence of bacterial meningitis in the United States has decreased by approximately 30%, *S. pneumoniae* has become the most common pathogen, and the disease is now more common in older adults than children; mortality rates (~15%) have not changed.[2] Worldwide, however, bacterial meningitis remains a major cause of mortality and morbidity. Although all human microbes have the potential to cause meningitis, only a few organisms account for most cases of bacterial meningitis.

The clinical setting in which meningitis develops may provide a clue to the specific bacterial cause. *H. influenzae* (Chapter 300) affects primarily children, whereas *S. pneumoniae* (Chapter 289) causes meningitis in adults, especially those older than 50 years with comorbid conditions. Meningococcal meningitis (Chapter 298) most often occurs in outbreaks. In developed countries, *Listeria monocytogenes* (Chapter 293) is emerging as the most common cause of bacterial meningitis, with peak frequencies in the neonatal period and in persons 60 years of age and older. Simultaneous mixed bacterial meningitis is rare but occurs in the setting of neurosurgical procedures, penetrating head injury, head trauma with fracture of the cribriform plate, erosion of the skull or vertebrae by adjacent neoplasm, extension of osteomyelitis, or intraventricular rupture of a cerebral abscess; isolation of anaerobes should strongly suggest the latter two of these situations. Meningitis involving anaerobes also may occur very rarely as a result of an intestinal-meningeal fistula following surgery and radiation therapy for colorectal cancer. In approximately 10% of patients with pyogenic meningitis, the bacterial cause cannot be defined.

Over the past several decades, gram-negative bacillary meningitis has doubled in frequency in adults, a change reflecting more frequent and extensive neurosurgical procedures, as well as other nosocomial factors. *L. monocytogenes* has increased 8- to 10-fold as a cause of bacterial meningitis in large urban general hospitals. *Listeria* infections are most often food-borne via dairy products, processed meats, uncooked vegetables, and precut salads. Although *Listeria* meningitis may occur in immunocompetent individuals, it occurs mostly in organ transplant recipients, patients undergoing hemodialysis, patients receiving corticosteroids or cytotoxic drugs for treatment of cancer or autoimmune diseases, patients with liver disease, alcoholic patients, pregnant women, and neonates. Meningitis caused by coagulase-negative staphylococci, which represents approximately 3% of cases in large urban hospitals, occurs as a complication of neurosurgical procedures and is often caused by methicillin-resistant strains. Viridans streptococci, *Pseudomonas,* and other gram-negative bacteria are the agents most often associated with meningitis that complicates diagnostic myelography and percutaneous trigeminal rhizotomy.

In large tertiary care hospitals, approximately 40% of cases of bacterial meningitis in adults are of nosocomial origin.[3] The leading causes are gram-negative bacilli (primarily *Escherichia coli* and *Klebsiella*), which account for approximately 40% of nosocomial episodes, as well as various streptococci, *Staphylococcus aureus,* and coagulase-negative staphylococci, each responsible for approximately 10% of nosocomial cases.[4]

Meningococcal disease, including meningitis, may occur sporadically and in cyclic outbreaks. High-risk groups include individuals who live in close quarters such as crowded classrooms, college dormitories, military barracks, or jails. In children the greatest risk is in the first year of life. In industrialized countries, serogroups B and C account for the majority of infections. In developing countries, serogroups A and, to a lesser extent, C are dominant. In sub-Saharan Africa, the so-called meningitis belt, recurrent yearly waves of serogroup A meningococcal infections can occur. The incidence of meningococcal meningitis was probably underestimated historically when the diagnosis was based on isolation of the organism. Polymerase chain reaction (PCR) testing suggests twice the number of cases.

Predisposing factors for the development of pneumococcal meningitis include acute otitis media (Chapters 289 and 426), with or without mastoiditis, which is seen in approximately 20% of adult patients. Pneumonia is present in approximately 15% of patients with pneumococcal meningitis, a much higher frequency than in meningitis caused by *H. influenzae* or *N. meningitidis*. Acute pneumococcal sinusitis (Chapter 426) is occasionally the initial focus from which infection spreads to the meninges. A recent or remote major head injury (Chapter 399) precedes approximately 10% of episodes of pneumococcal meningitis, and CSF rhinorrhea (usually caused by a defect or fracture in the cribriform plate) is present in approximately 5% of patients. Cochlear implants, particularly those that include a positioner, have been implicated in cases of childhood bacterial meningitis, especially episodes resulting from *S. pneumoniae.* Occasionally, meningitis caused by *S. pneumoniae* develops in patients with central nervous system (CNS) shunts. Splenectomy or splenic dysfunction, as in sickle cell anemia (Chapter 163) cirrhosis (Chapter 153) with portal hypertension, or defects in humoral immunity also predispose patients to pneumococcal meningitis. Alcoholism (Chapter 33) is an underlying risk factor in 10 to 25% of adults with pneumococcal meningitis in urban hospitals. The estimated annual incidence of bacterial meningitis (primarily pneumococcal) in patients infected with human immunodeficiency virus (HIV) is 150-fold higher than in the general population.

S. aureus meningitis is seen most commonly as a complication of a neurosurgical procedure, after penetrating skull trauma, or occasionally secondary to staphylococcal bacteremia and endocarditis. Meningitis attributable to

gram-negative bacilli takes one of three forms: neonatal meningitis, meningitis after trauma or neurosurgery, or spontaneous meningitis in adults (e.g., bacteremic *Klebsiella* meningitis in a patient with diabetes mellitus). The most common causes of gram-negative bacillary meningitis in adults are *E. coli* (≈30%) and *Klebsiella-Enterobacter* (≈40%). Meningitis caused by group A streptococci is uncommon but occasionally occurs after acute otitis media, more often in children than in adults. *H. influenzae* type b meningitis in an adult suggests an underlying anatomic or immunologic defect.

Patients with defects in cell-mediated immunity are susceptible to the development of CNS infections with intracellular organisms such as *L. monocytogenes*. Patients with defective humoral immunity and an inadequate antibody response are particularly vulnerable to meningitis with *S. pneumoniae* and *H. influenzae*. Patients with neutropenia are at higher risk for meningitis with *Pseudomonas aeruginosa* and members of the Enterobacteriaceae family.

PATHOBIOLOGY

Pathology

On gross examination, purulent exudate in the subarachnoid space is most abundant in the cisterns at the base of the brain and over the convexities of the rolandic and sylvian sulci, which are expansions of the subarachnoid space. Although neither the infecting organism nor the inflammatory exudate directly invades cerebral tissue, the subjacent brain becomes congested and edematous. The effectiveness of the pial barrier generally prevents bacterial meningitis from causing a cerebral abscess; when these two processes coexist, the sequence is usually that an initial abscess leaks its contents into the ventricular system and produces secondary ventriculitis and meningitis.

The inflammatory exudate can extend around the perivascular spaces to adjacent structures, especially the arteries and veins that carry a layer of pia mater and arachnoid membrane as they enter the brain from the cortical surface. *Cortical thrombophlebitis* results from venous stasis and adjacent meningeal inflammation. Infarction of cerebral tissue may follow. *Involvement of cortical and pial arteries* by peripheral aneurysm formation and vascular occlusion or narrowing (related to spasm, arteritis, or both) of the supraclinoid portion of the internal carotid artery at the base of the brain occurs in approximately 15% of patients with meningitis. The anterior and middle cerebral arteries may have markedly increased intracerebral blood flow velocity (an index of stenosis or arterial spasm) on transcranial Doppler ultrasonography, a finding corresponding to focal cerebral signs. In fulminating cases, particularly meningococcal meningitis, *cerebral edema* may be marked even though the pleocytosis is only moderate. Rarely, temporal lobe herniation through the tentorium develops in such patients and compresses the midbrain, thereby leading to ipsilateral third nerve palsy and contralateral hemiparesis or cerebellar herniation through the foramen magnum with compression of the medulla, which results in apnea, hemodynamic instability, and coma. *Damage to cranial nerves* occurs in areas where dense exudate accumulates around the nerves; the third and sixth cranial nerves are also vulnerable to damage by increased intracranial pressure. *Ventriculitis* accompanies most cases of bacterial meningitis and may rarely progress to *ventricular empyema*. As the exudates continue to accumulate, obstruction of the flow of CSF may result in *hydrocephalus*. Obstruction of the foramina of Magendie and Luschka at the base of the fourth ventricle results in noncommunicating or obstructive hydrocephalus, whereas obstruction at the level of the arachnoid granulations in the venous sinuses results in communicating hydrocephalus. *Subdural effusions* are sterile transudates that develop over the cerebral cortex and can be demonstrated readily by computed tomography (CT) as low-density areas about the cerebrum; rarely, such effusions become infected and produce subdural empyema.

Pathogenesis

Bacteria may gain access to the meninges by several routes: (1) hematogenous spread from a distant site, (2) direct ingress from the upper respiratory tract or skin through an anatomic defect (e.g., skull fracture, meningocele, sequela of surgery), (3) passage intracranially through venules in the nasopharynx, or (4) spread from a contiguous focus of infection (infection of the paranasal sinuses, leakage of a brain abscess). Bacteremic spread of *H. influenzae, N. meningitidis,* and *S. pneumoniae* is probably the most frequent path of infection. Bacteremia is usually initiated by pharyngeal adhesion and colonization by an infecting strain. Adhesion of such strains, as well as of *S. pneumoniae,* to mucosal surfaces is abetted by their capacity to produce proteases that cleave immunoglobulin A, thus inactivating this local antibody defense. Adhesion of *N. meningitidis* to nasopharyngeal cells is affected by fimbriae or pili and promoted by previous damage to ciliated cells such as from smoking or viral infections. Meningococci invade the nasopharyngeal mucosal cells by means of endocytosis and are transported to the abluminal side in membrane-bound vacuoles. *H. influenzae,* in contrast, invades intercellularly by causing separation of the apical tight junctions between columnar epithelial cells. When these meningeal pathogens gain access to the blood stream, their intravascular survival is aided by the presence of polysaccharide capsules that inhibit phagocytosis and confer resistance to complement-mediated bactericidal activity.

Bacteria also may travel along nerve tracts to invade the brain. For example, *L. monocytogenes* invades the intestine, and animal models suggest that these bacteria can travel along the vagus nerve to the brain stem, from where they also may invade the meninges in the posterior fossa.

The mechanism by which bacteria gain access to the subarachnoid spaces from blood appears to be related to specific adhesion molecules on brain endothelial cells. Once established in any part of the meninges, infection quickly extends throughout the subarachnoid space. Bacterial replication proceeds relatively unhindered because the low CSF levels of immunoglobulin and complement early in meningeal inflammation result in minimal or no opsonic or bactericidal activity and because surface phagocytosis of unopsonized organisms is meager in such a fluid environment. During meningitis, the concentrations of immunoglobulins in CSF increase but still remain relatively low. Secondary bacteremia may follow meningeal infection and may itself contribute to continuing further inoculation of CSF.

Bacterial meningitis following head trauma occurs because of a dural fistula from the nasal cavity, paranasal sinuses, or middle ear to the subarachnoid space. The most frequent site is at the cribriform plate, where the bone is thin and the dura is tightly adherent to the bone. Leakage of CSF results in CSF rhinorrhea and loss of smell.

Bacterial components (e.g., pneumococcal cell walls or lipoteichoic acid, *H. influenzae* lipo-oligosaccharide) are major elicitors of meningeal inflammation by causing release into the subarachnoid space of various pro-inflammatory cytokines such as interleukin-1 and tumor necrosis factor (TNF) from endothelial and meningeal cells, macrophages, and microglia. Cytokines appear to enhance the passage of leukocytes by inducing several families of adhesion molecules that interact with the corresponding receptors on leukocytes. Cytokines also can increase the binding affinity of a leukocyte selectin, leukocyte adhesion molecule, for its endothelial cell receptor and may thereby further contribute to trafficking of neutrophils into the subarachnoid space.

In bacterial meningitis, neutrophils move into the subarachnoid space but are not able to control the bacterial infection because their phagocytic properties are inefficient as a result of a lack of opsonic and bactericidal activity. Within the subarachnoid space, neutrophils release prostaglandins, matrix metalloproteinases, and free radicals that disrupt the endothelial intercellular tight junctions and the subendothelial basal lamina. The increased local vascular permeability of the blood-brain barrier may cause cerebral edema, which also can be caused by increased CSF pressure as a result of obstruction of CSF outflow because of interstitial inflammation at the level of the arachnoidal villi.

Cerebral blood flow, which depends on mean arterial pressure, appears to be increased in the early stages of meningitis, but it subsequently decreases, substantially in some patients, and this decreased blood flow itself may cause ischemic neurologic injury. Localized regions of marked hypoperfusion, attributable to focal vascular inflammation or thrombosis, can occur in patients with normal blood flow. Impairment of cerebral blood flow autoregulation, as measured by transcranial Doppler ultrasonography of the middle cerebral artery, occurs early in acute bacterial meningitis and causes cerebral blood flow to correspond directly to mean arterial blood pressure, with attendant hyperperfusion or hypoperfusion of the brain. On recovery, the ability of the cerebral vasculature to maintain a constant level of perfusion despite variations in mean arterial pressure is restored.

CLINICAL MANIFESTATIONS

History

Acute-onset fever, generalized headache, vomiting, and stiff neck are common to many types of meningitis (Table 412-1). Most patients with community-acquired pyogenic meningitis have had an antecedent or accompanying upper respiratory tract infection or nonspecific febrile illness, acute otitis (or mastoiditis), or pneumonia. Myalgia, particularly in patients with meningococcal disease, backache, and generalized weakness are common symptoms. The illness usually progresses rapidly, with the development of confusion,

TABLE 412-1 SYMPTOMS AND SIGNS OF BACTERIAL MENINGITIS*

CHARACTERISTIC	EPISODES OF MENINGITIS
Duration of symptoms < 24 hr	48%
Predisposing conditions	
Otitis or sinusitis	25%
Pneumonia	12%
Immunocompromise†	16%
Symptoms at initial evaluation	
Headache	87%
Nausea	74%
Neck stiffness	83%
Triad of fever, neck stiffness, and change in mental status	44%
Focal neurologic deficits	33%
Aphasia	23%
Hemiparesis	7%
Indices of CSF inflammation	
Opening pressure (mm H_2O)‡	370 ± 130
White cell count§	
Mean (cells/μL)	7753 ± 14,736
<100/μL	7%
100-999/μL	14%
>999/μL	78%
Protein (g/L)	4.9 ± 4.5
CSF/blood glucose ratio	0.2 ± 0.2
Positive blood culture‖	66%
Blood tests	
ESR (mm/hr)¶	46 ± 37
C-reactive protein (g/L)**	225 ± 132
Platelet count (platelets/μL)††	198,000 ± 100,000

*Data from 696 cases reported in van de Beek D, de Gans J, Spanjaard L, et al. Clinical features and prognostic factors in adults with bacterial meningitis. *N Engl J Med.* 2004;351:1849-1859.
The study included 671 patients who had a total of 696 episodes of community-acquired meningitis. Plus-minus values are means ± standard deviation.
†Immunocompromise was defined by the use of immunosuppressive drugs, a history of splenectomy, or the presence of diabetes mellitus or alcoholism, as well as patients infected with human immunodeficiency virus.
‡CSF pressure was measured in 216 patients.
§The CSF leukocyte count was determined in 659 patients; CSF specimens from 14 patients had too many leukocytes for an exact count to be performed.
‖Blood culture was performed in 611 patients.
¶The ESR was determined in 549 patients.
**C-reactive protein levels were determined in 394 patients.
††The thrombocyte count was determined in 653 patients.
CSF = cerebrospinal fluid; ESR = erythrocyte sedimentation rate.

obtundation, and loss of consciousness. Occasionally, the onset may be less acute, with meningeal signs being present for several days to a week.

General Physical Findings

Evidence of meningeal irritation is usually present, as evidenced by a stiff neck, Kernig sign (inability to straighten the leg when the hip is flexed to 90 degrees), and Brudzinski sign (involuntary flexion of the hip and knee when the neck is passively flexed). Neck stiffness, Kernig sign, and Brudzinski sign each have sensitivities of approximately 30% or lower for diagnosing acute bacterial meningitis in adults.[5] Although the classic triad of fever, stiff neck, and change in mental status is initially present in only 44% of episodes, a combination of two of four symptoms (headache, fever, stiff neck, and altered mental status) is found in 95% of patients. The findings of meningitis may be easily overlooked in infants, obtunded patients, elderly patients with heart failure or pneumonia, or immunosuppressed individuals, who may have meningitis without prominent meningeal signs; in such patients, lethargy should be investigated carefully, meningeal signs should be sought, and examination of CSF is indicated if any doubt exists. In elderly patients, neck stiffness may be difficult to evaluate because of osteoarthritis in the neck or stiffness of neck muscles secondary to basal ganglia disorders. When neck stiffness is caused by meningitis, the neck resists flexion but can be rotated passively from side to side; with cervical spine disease, however, resistance is present in all directions of neck movement. Neck stiffness disappears during coma.

The presence of a petechial or ecchymotic rash (see Fig. 298-3) in a patient with meningeal findings almost always indicates meningococcal infection and requires prompt treatment because of the rapidity with which this infection can progress (Chapter 298). Rarely, extensive petechial and ecchymotic lesions occur in meningitis caused by *S. pneumoniae, H. influenzae,* or echovirus type 9. Very rarely, skin lesions almost indistinguishable from those of meningococcal bacteremia occur in patients who have acute *S. aureus* endocarditis (see Fig. 76-1) and who also have meningeal signs and pleocytosis (secondary to either staphylococcal meningitis or embolic cerebral infarction). Usually, one or two of the lesions in such a patient represent purulent purpura; aspiration of material reveals staphylococci on Gram staining. In the summer, viral aseptic meningitis may produce meningeal signs, macular and petechial skin lesions, and a pleocytosis of several hundred cells, sometimes with neutrophils predominating initially.

Fulminant meningococcal septicemia may cause hemorrhages within the adrenal glands and result in Waterhouse-Friderichsen syndrome (Chapter 227), a condition characterized by the sudden onset of a febrile illness, large petechial hemorrhages in the mucous membranes and skin, cardiovascular collapse, and disseminated intravascular coagulation. In contrast, hyponatremia and the syndrome of inappropriate secretion of antidiuretic hormone may develop in patients with meningitis attributable to *H. influenzae*. A concurrent respiratory tract infection or acute otitis media may be present with either *H. influenzae* or *S. pneumoniae*.

In patients with a basilar skull fracture, the potential for development of a dural fistula and bacterial meningitis is indicated by the presence of CSF rhinorrhea, periorbital ecchymoses, bruising behind the ear (Battle sign), hemotympanum, or blood in the external auditory canal. Meningitis complicating neurosurgical procedures may be insidious in onset and difficult to distinguish from the altered consciousness and signs of meningeal irritation that are expected in the postoperative period. However, fever or prolonged obtundation is an indication for evaluation of CSF.

Neurologic Findings and Complications

Neurologic complications in patients with inadequately treated bacterial meningitis can be severe and disabling. Cranial nerve abnormalities, involving principally the third, fourth, sixth, or seventh nerve, occur in 5 to 10% of adults with community-acquired meningitis and usually disappear shortly after recovery. Persistent sensorineural hearing loss occurs in 10% of children with bacterial meningitis, and another 16% have transient conductive hearing loss. The most likely sites of involvement in patients with persistent sensorineural deafness appear to be the inner ear (infection or toxic products possibly spreading from the subarachnoid space along the cochlear aqueduct) and the acoustic nerve. In children, permanent hearing impairment is more common after meningitis caused by *S. pneumoniae* than by *H. influenzae* or *N. meningitidis*.

Seizures (focal or generalized; Chapter 403) occur in 20 to 30% of patients and may result from reversible causes (high fever or hypoglycemia in infants, penicillin neurotoxicity when large doses are administered intravenously (IV) to patients with renal failure) or, more commonly, from focal cerebral injury related to arterial hypoperfusion and infarction, cortical venous thrombosis, or focal edema and cerebritis. Seizures can occur during the first few days or can appear with associated focal neurologic deficits caused by vascular inflammation some days after onset of the meningitis. In adults with seizures accompanying meningitis, *S. pneumoniae* is more commonly the cause, but alcohol withdrawal is a confounding factor.

Increased CSF pressure, which can be caused by brain swelling or hydrocephalus, is associated with seizures, vomiting, sixth and third nerve dysfunction, abnormal reflexes, reduced consciousness or coma, dilated and poorly reactive pupils, and the Cushing response of decerebrate posturing, hypertension, bradycardia, and irregular respirations. In approximately a fourth of fatal cases of community-acquired meningitis in adults, cerebral edema accompanied by temporal lobe herniation is observed at autopsy.

Papilledema (see Fig. 423-27) occurs in less than 1% of patients with bacterial meningitis, even with high CSF pressure, probably because the patient is seen early in the process before changes in the nerve head have occurred. The presence of this sign should indicate the possibility of another associated or independent suppurative intracranial process, such as subdural empyema or brain abscess. Marked central hyperpnea sometimes occurs in patients with severe bacterial meningitis; CSF acidosis, which is principally due to increased lactic acid levels, provides much of the respiratory stimulus.

Focal cerebral signs (principally hemiparesis, dysphasia, visual field defects, and gaze preference) occur in approximately a third of adults with community-acquired bacterial meningitis. These signs may develop because

of arterial or venous occlusion. In addition, cerebral blood flow velocity may be decreased in patients with increased intracranial pressure and may lead to temporary or lasting neurologic dysfunction. It is important to distinguish these vascular effects from postictal changes (Todd paralysis), which usually persist for less than a day. Meningitis may cause the syndrome of inappropriate secretion of antidiuretic hormone.

DIAGNOSIS

Bacterial meningitis is a medical emergency that requires immediate diagnosis and rapid institution of antimicrobial therapy. Delay in treatment is the most critical factor in determining the morbidity and mortality of patients with bacterial meningitis. The diagnosis of bacterial meningitis is not difficult in a febrile patient with meningeal symptoms and signs developing in the setting of a predisposing illness. The diagnosis may be less obvious in an elderly, obtunded patient with pneumonia or a confused alcoholic patient in impending delirium tremens.

When the diagnosis of bacterial meningitis is entertained, blood cultures should be performed, CSF examined and cultured, and antimicrobial therapy instituted promptly. If a mass lesion (cerebral abscess, subdural empyema) is suspected from the history, clinical setting, or physical findings (papilledema, focal cerebral signs), CT with or without contrast enhancement or magnetic resonance imaging (MRI) should be performed because of the danger of brain herniation with lumbar puncture. Antibiotics can and commonly should be started immediately, even before performing lumbar puncture, because it takes approximately 2 hours for antibiotics to affect CSF cultures. Diagnostic lumbar puncture should not be delayed to perform CT or MRI except in patients who have focal neurologic findings suggestive of a parameningeal collection or other intracranial mass lesions; in such patients, it is critical to initiate antimicrobial therapy for meningitis of unknown origin or brain abscess before CT or MRI is performed. Patients with community-acquired meningitis rarely have important abnormalities detected on CT in the absence of focal neurologic findings.

Laboratory Findings

Cerebrospinal Fluid Examination

Initial CSF pressure is usually moderately elevated (200 to 300 mm H_2O in adults). Striking elevations (≥450 mm H_2O) occur in occasional patients with acute brain swelling complicating meningitis in the absence of an associated mass lesion. Findings on CSF analysis are strikingly abnormal in patients with meningitis, and such findings help suggest the cause even before the results of culture are available (Table 412-2). In patients with skull fractures, CSF rhinorrhea can be distinguished from nasal secretions by the presence of glucose.

Gram-Stained Smear

By the time of hospitalization, most patients with pyogenic meningitis have large numbers (≥10^5/mL) of bacteria in their CSF. Careful examination of the Gram-stained smear of the spun sediment of CSF reveals the etiologic agent in 60 to 80% of cases. In most instances in which gram-positive diplococci (or short-chain cocci) are observed on a stained CSF smear, they are pneumococci. *Enterococcus,* an occasional cause of nosocomial meningitis, is detected by latex particle agglutination. Rarely, three species may morphologically mimic *Neisseria* in CSF or may suggest a mixed infection with short gram-negative rods and meningococci: *Acinetobacter baumannii, Moraxella* sp, and *Pasteurella multocida*. Culture of CSF reveals the etiologic agent in 80 to 90% of patients with bacterial meningitis if CSF is obtained before or within 1 to 2 hours of the initiation of antibiotics.

Special Testing Procedures

Broad-range PCR, which can be performed on CSF within 1.5 hours, can diagnose bacterial meningitis in patients in whom antimicrobial therapy was begun before lumbar puncture or when cultures are negative and a bacterial origin is still suspected. However, the sensitivity of PCR for bacterial infections ranges from approximately 80 to 95%, so a negative result does not exclude the diagnosis. In resource-poor countries, a reagent strip that detects cells, proteins, and glucose has a sensitivity and specificity above 96% for diagnosing bacterial meningitis.[6] In many parts of the world, it may be especially useful for distinguishing bacterial meningitis from CNS malaria (Chapter 345).

Cell Count

Cells counts should be determined promptly because the cells will begin to lyse after 90 minutes. The normal CSF white blood cell count is less than 5/μL (all mononuclear). The cell count in untreated meningitis usually ranges between 100 and 10,000/μL, with polymorphonuclear leukocytes predominating initially (>80%) and lymphocytes appearing subsequently.

The cell count in *L. monocytogenes* meningitis tends to be lower (median, 585/μL) than in other types of community-acquired pyogenic meningitis. Extremely high cell counts (>50,000/μL) should raise the possibility of intraventricular rupture of a cerebral abscess. Cell counts as low as 10 to 20/μL may be observed early in bacterial meningitis, particularly that caused by *N. meningitidis* and *H. influenzae*. Occasionally, in granulocytopenic patients or in elderly persons with overwhelming pneumococcal meningitis, CSF may contain very few leukocytes and yet may appear grossly turbid because of the presence of a myriad of organisms and an elevated protein level. Meningitis caused by several bacterial species (*Mycobacterium tuberculosis, Borrelia burgdorferi, Treponema pallidum, Leptospira* sp, *Francisella tularensis, Brucella* sp) is characteristically associated with a lymphocytic pleocytosis. With *L. monocytogenes* meningitis in an adult, there is usually a polymorphonuclear response but lymphocytes may predominate in rare instances.

Glucose

CSF glucose is reduced to values of 40 mg/dL or less (or < 50% of the simultaneous blood level) in 50% of patients with bacterial meningitis; this finding helps distinguish bacterial meningitis from most viral meningitides or parameningeal infections. However, a normal CSF glucose value does not exclude the diagnosis of bacterial meningitis. The blood glucose level should be determined simultaneously because patients with diabetes mellitus (or those who are receiving intravenous glucose infusions) have an elevated CSF glucose level that can be appreciated only by comparison with the simultaneous blood level; however, it may take 90 to 120 minutes for equilibration to occur after major shifts in the level of glucose in the circulation. The hypoglycorrhachia characteristic of pyogenic meningitis appears to result from interference with normal carrier-facilitated diffusion of glucose and increased utilization of glucose by host cells.

Protein

The level of protein in lumbar CSF is usually elevated to greater than 100 mg/dL, and higher values are more commonly observed in pneumococcal meningitis. Extreme elevations, 1000 mg/dL or greater, may indicate subarachnoid block with obstruction of CSF flow. Values higher than 15 mg/dL in ventricular CSF are considered abnormal. If the lumbar puncture is traumatic, the CSF protein level is corrected by subtracting 1 mg/dL for every 1000 red blood cells.

TABLE 412-2 COMMON CEREBROSPINAL FLUID FINDINGS IN PATIENTS WITH MENINGITIS

MICROORGANISM	CSF OPENING PRESSURE (cm H_2O)	CELL COUNT (CELLS/μL)	PROTEIN (mg/dL)	GLUCOSE (mg/dL)
Bacteria*	>20	>1000	>100	<10
Mycobacterium tuberculosis	>20	100-500	>100	10-45
Borrelia burgdorferi	<20	100-500	50-150	10-45
Treponema pallidum	<20	5-500	50-150	10-45
Fungi	<20	5-500	>100	10-45
Viruses	<20	5-500	50-150	Normal

Modified from Kim KS. Acute bacterial meningitis in infants and children. *Lancet Infect Dis.* 2010;10:32-42.
*Group B streptococci, *Escherichia coli, Listeria monocytogenes, Streptococcus pneumoniae, Neisseria meningitidis, and Haemophilus influenzae* type b.
CSF = cerebrospinal fluid.

Other Abnormalities

Elevated levels of lactic acid occur in pyogenic meningitis. The diagnostic accuracy of a CSF lactate level is at least as good as a cell count for differentiating bacterial from aseptic meningitis,[7] and a value above 3.0 mmol/L has a sensitivity and specificity of 94 to 95% for bacterial meningitis.[8] However, the CSF lactate level is less useful in patients who have received antibiotics and it also may be increased in other conditions such as cerebral ischemia, stroke, and head trauma. Although lactate dehydrogenase levels are higher in patients with bacterial meningitis than in patients with viral infections of the CNS, these alterations are not helpful in determining the specific etiologic agent.

Blood and Respiratory Tract Cultures

Bacteremia is demonstrable in approximately 80% of patients with *H. influenzae* meningitis, 50% of patients with pneumococcal meningitis, and 30 to 40% of patients with meningococcal meningitis. Hence, blood cultures should be performed routinely in patients suspected of having bacterial meningitis. Cultures of the upper respiratory tract are not helpful in establishing an etiologic diagnosis.

Determination of serum creatinine and electrolyte levels is important in view of the gravity of the illness, the occurrence of specific abnormalities secondary to the meningitis (syndrome of inappropriate secretion of antidiuretic hormone), and problems with therapy in patients with renal dysfunction (seizures and hyperkalemia with high-dose penicillin therapy). In patients with extensive petechial and purpuric skin lesions, evaluation for coagulopathy is indicated. Elevated serum procalcitonin levels have been used to distinguish bacterial meningitis from that of viral origin, but CSF examination (Gram stain, white blood cell count, glucose, culture) usually provides more direct and specific information.

Radiologic Studies

Because of the frequency with which pyogenic meningitis is associated with primary foci of infection in the chest, nasal sinuses, or mastoid, radiographs of these areas should be taken when clinically indicated at the appropriate time after antimicrobial therapy is begun. Initial head CT or MRI is not indicated in most patients with bacterial meningitis. For example, in patients who undergo head CT or MRI before lumbar puncture for suspected meningitis, only approximately 5% have a mass effect identified on CT. Baseline clinical features associated with abnormal findings on CT include age older than 60 years, history of CNS disease, seizure within the previous week, abnormal level of consciousness, abnormal visual fields, limb drift, and aphasia. In patients without any of these clinical findings, only approximately 1% have a mass effect identified on CT or MRI that would raise concern regarding lumbar puncture.

Specific changes that may be observed on CT or MRI during meningitis include cerebral edema and enlargement of the subarachnoid spaces, contrast enhancement of the leptomeninges and the ependyma, or patchy areas of diminished density as a result of associated cerebritis and necrosis. In patients with meningitis whose clinical status deteriorates or fails to improve, CT or MRI may help demonstrate suspected complications—that is, sterile subdural collections or empyema; ventricular enlargement secondary to communicating or obstructive hydrocephalus; prominent persisting basilar meningitis; extensive areas of cerebral infarction resulting from occlusion of major cerebral arteries, veins, or venous sinuses; or marked ventricular wall enhancement suggesting ventriculitis or ventricular empyema. MRI is superior to CT for visualizing these abnormalities. Rarely, cerebral hemorrhage identifiable on CT may complicate acute bacterial meningitis in adults. In approximately 10% of adults with bacterial meningitis, findings on cranial CT (mastoid or sinus wall defect, eroding retrobulbar mass, pneumocephalus) are indicative of disruption of the dural barrier.

Rarely, paraparesis or tetraparesis resulting from myelitis may complicate bacterial meningitis. In this situation, T2-weighted or short tau inversion recovery (STIR) sequences on MRI can be helpful to exclude spinal cord compression by an extramedullary mass.

Differential Diagnosis

Headache, fever, stiff neck, confusion, vomiting, and pleocytosis are features of meningeal inflammation and are common to many types of meningitis (e.g., bacterial, fungal, viral, chemical) and also some parameningeal processes. The CSF findings are most helpful in distinguishing among these processes (Chapters 413 and 414).[9] Although a lymphocyte-predominant pleocytosis without hypoglycorrhachia is characteristic of viral (usually enteroviral or herpes simplex virus type 2 [HSV-2]) meningitis or meningoencephalitis (HSV-1), the initial CSF finding may be a polymorphonuclear response (of ≤ 60%) that quickly becomes mononuclear. HSV-1 encephalitis is suggested by neurologic findings (dysphasia, hemiparesis, olfactory hallucinations, other temporal lobe signs, seizures), abnormalities in the orbitofrontal and medial temporal lobes on MRI, and distinctive electroencephalographic changes in the temporal lobe or lobes. The rash, fever, and headache of Rocky Mountain spotted fever (Chapter 327) may suggest meningococcal infection, but the geographic and seasonal predilections of the former can provide clues. Approximately 10% of patients hospitalized with Rocky Mountain spotted fever have CSF cell counts higher than 100/μL (>70% polymorphonuclear), and thus the condition initially may be confused with bacterial meningitis. The rash associated with enteroviral infections typically consists of erythematous macules and papules on the face, neck, and trunk. Acute subarachnoid hemorrhage (Chapter 408) may be confused with bacterial meningitis because of headache, stiff neck, and vomiting. However, subarachnoid hemorrhage usually has a more abrupt onset without a prodromal fever but with evidence of subarachnoid blood on CT or CSF examination. In patients with neuroleptic malignant syndrome (Chapters 410 and 418), fever, generalized rigidity, and a fluctuating level of consciousness with autonomic instability and leukocytosis may develop. The most specific laboratory abnormality in these patients is a markedly elevated creatine kinase level.

In a patient with meningitis but whose CSF does not reveal the etiologic agent on a Gram-stained smear, particularly when the CSF glucose level is normal and the polymorphonuclear pleocytosis is atypical, certain treatable processes that can mimic bacterial meningitis should be considered in the differential diagnosis:

1. *Parameningeal infections.* The presence of infections (chronic ear or nasal accessory sinus infections, lung abscess) predisposing to brain abscess, epidural (cerebral or spinal) abscess, subdural empyema, or pyogenic venous sinus phlebitis should be sought (Chapter 413). Neurologic symptoms may appear in the course of primary bacterial meningitis, but their presence should alert the physician to the need for close scrutiny for the presence of a space-occupying infectious process in the CNS. Neurologic symptoms or findings antedating the onset of meningeal symptoms should suggest the possibility of a parameningeal infection. Isolation of an anaerobic organism should suggest the possibility of intraventricular leakage of a cerebral abscess.
2. *Bacterial endocarditis.* Bacterial meningitis may occur during bacterial endocarditis (Chapter 76) caused by pyogenic organisms such as *S. aureus* and enterococci. In subacute bacterial endocarditis, sterile embolic infarctions of the brain may produce meningeal signs and a pleocytosis consisting of several hundred cells, including polymorphonuclear leukocytes. A history of dental manipulation, fever, and anorexia antedating the meningitis should be sought; careful examination for heart murmurs and peripheral stigmata of endocarditis is indicated.
3. *"Chemical" meningitis.* The clinical and CSF findings (polymorphonuclear pleocytosis and even reduced glucose level) of bacterial meningitis may be produced by chemically induced inflammation. Acute meningitis after diagnostic lumbar puncture or spinal anesthesia may result from bacterial or chemical contamination of equipment or anesthetic agent. Chemical meningitis, characterized by polymorphonuclear pleocytosis, hypoglycorrhachia, and a latent period of 3 to 24 hours, occurs after 1% of metrizamide myelograms. Endogenous chemical meningitis resulting from material from an epidermoid tumor or a craniopharyngioma leaking into the subarachnoid space, a glioblastoma invading the ventricles (Chapter 189), or carcinomatous mengingitis (see later) can produce polymorphonuclear pleocytosis and hypoglycorrhachia.

Complications

Non-neurologic Complications

Shock

When shock occurs in patients with pyogenic meningitis, it is usually a manifestation of the accompanying intense bacteremia, as in fulminant meningococcemia, rather than a manifestation of the meningitis itself. Management is guided by the principles of septic shock therapy (Chapter 108), with appropriate modifications in patients with heart failure (Chapter 59).

Coagulation Disorders

Coagulopathies (Chapter 174) are frequently associated with the intense bacteremia (usually meningococcal, occasionally pneumococcal) and hypotension that can accompany meningitis. The changes may be mild, such as

thrombocytopenia (with or without prolongation of the prothrombin and partial thromboplastin times), or more marked, with clinical evidence of disseminated intravascular coagulation (Chapter 175).

Septic Complications

Endocarditis

In patients with pneumococcal meningitis, particularly those with concomitant bacteremia and pneumonia, acute endocarditis (Chapter 76) can develop, most commonly on the aortic valve. In such patients, febrile relapse and a new cardiac murmur may appear shortly after the completion of antimicrobial therapy for meningitis.

Pyogenic Arthritis

Septic arthritis may result from the bacteremia associated with meningitis caused by *S. pneumoniae, N. meningitidis,* or *H. influenzae*.

Prolonged Fever

With appropriate antimicrobial treatment of community-acquired bacterial meningitis, patients become afebrile within 2 to 5 days. Sometimes, however, the fever persists or recurs after an afebrile period. In a patient with persisting headache, obtundation, and cerebral findings, inadequate drug therapy or neurologic sequelae (cortical venous thrombophlebitis, ventriculitis, subdural collections) are important considerations. Re-evaluation of CSF, particularly Gram-stained smear and culture, is essential in these circumstances. Drug-induced fever (Chapters 254 and 280) should be suspected in patients who continue to show clinical improvement in all other respects. Metastatic infection (septic arthritis, purulent pericarditis, thoracic empyema, endocarditis) may be the cause of continuing or recurrent fever. A syndrome, probably immunologic, consisting of fever, arthritis, and pericarditis 3 to 6 days after the initiation of effective antimicrobial therapy for meningococcal meningitis occurs in approximately 10% of patients (Chapter 298).

Recurrent Meningitis

Repeated episodes of bacterial meningitis generally indicate a host defect, either in local anatomy or in antibacterial and immunologic defenses (e.g., recurrent *N. meningitidis* infections in patients with congenital or acquired deficiencies of complement, particularly the late-acting components). Approximately 10% of episodes of pneumococcal meningitis in adults are recurrent meningitis, but only 0.5% of patients with community-acquired meningitis caused by other microorganisms have recurrent attacks. *S. pneumoniae* is the cause of a third of episodes of community-acquired recurrent meningitis; various streptococci, *H. influenzae,* and *N. meningitidis* are the cause of another third of episodes. In contrast, in nosocomial recurrent meningitis, gram-negative bacilli and *S. aureus* are the cause of approximately 60% of episodes. A history of head trauma is frequent in patients with recurrent meningitis. Organisms may enter the subarachnoid space directly, through a defect in the cribriform plate (the most common site), in association with the empty sella syndrome, by means of a basilar skull fracture, through an erosive sequestrum of the mastoid, through congenital dermal defects along the craniospinal axis (usually evident before adult life), or as a consequence of penetrating cranial trauma or neurosurgical procedures. The anatomic defect may produce a frank CSF leak (rhinorrhea or, less commonly, otorrhea) or may entrap a vascular cuff of meninges that may subsequently serve as a direct route for organisms to reach the meninges. CSF rhinorrhea may be intermittent, and meningitis may occur months or years after head injury.

Any patient with bacterial meningitis, particularly if the meningitis is recurrent, should be evaluated carefully for congenital or post-traumatic defects. The presence of CSF rhinorrhea should be sought at admission and subsequently (rhinorrhea may clear during active meningitis only to recur when the inflammation has resolved). Clinical clues suggesting the presence of a CSF fistula through the cribriform plate, pericranial air sinuses, or temporal bone include (1) a salty taste in the throat, (2) positionally dependent rhinorrhea (rhinorrhea only in the lateral recumbent or prone position suggests an otic or sphenoid origin), (3) anosmia (cribriform plate leak), and (4) hearing loss or full feeling in the ear, often with a finding of fluid or bubbles behind the tympanic membrane (leakage into the middle ear). Quantitative determination of the glucose and chloride content of nasal secretions and detection of a transferrin band unique to CSF by protein electrophoresis can definitively establish the presence of CSF rhinorrhea.

Recurrent pneumococcal meningitis may develop without apparent predisposing circumstances, and cryptic CSF leaks should be sought actively in such patients by CT of the frontal and mastoid regions and by radioisotope techniques. Radioiodine-labeled albumin is introduced intrathecally, and pledgets of cotton placed in the nares are subsequently examined for the radionuclide. Intrathecal introduction of fluorescein as a visual tracer (under ultraviolet light) can similarly be used to detect active leaks. Surgical closure of CSF fistulas should be performed to prevent further episodes of meningitis. Extracranial approaches through the ethmoidal sinuses can be used to repair cribriform plate or sphenoidal sinus dural defects and avoid the higher morbidity associated with craniotomy.

In most patients with CSF otorrhea and rhinorrhea after an acute head injury, the leak ceases in 1 or 2 weeks. *Persistent rhinorrhea for more than 4 to 6 weeks is an indication for surgical repair.* Prolonged administration of penicillin does not prevent pneumococcal meningitis and may encourage infection with more drug-resistant species.

TREATMENT

Antimicrobial Agents

Antimicrobial therapy should be initiated promptly in this life-threatening emergency. Subsequent management should be undertaken with close monitoring, often in an intensive care unit. Treatment should be aimed at the most likely causes based on clinical clues, such as the age of the patient, the presence of a petechial or purpuric rash, a recent neurosurgical procedure, and CSF rhinorrhea. However, it is difficult to distinguish among the various causes of bacterial meningitis on clinical grounds alone, although patients with pneumococcal meningitis frequently have altered mental status and progress rapidly to coma, often with recurrent seizures and the rapid development of focal neurologic deficits. If the infecting organism is observed on examination of a Gram-stained smear of the CSF sediment, specific therapy is initiated. If the etiologic agent is not seen on a smear from a patient with suspected bacterial meningitis or if lumbar puncture is delayed because head CT is needed, empirical antimicrobial therapy should be initiated (Table 412-3).

Adequate CSF bactericidal activity, which is critical to cure the meningitis, depends on the ability of the antibiotic to penetrate CSF and maintain its activity in the purulent exudate, as well as on its metabolism and rate of clearance from CSF. The ability of the antibiotic to penetrate CSF depends on its lipid solubility, protein binding in serum, molecular size, and the status of the blood-CSF barrier. For example, chloramphenicol has very high lipid solubility, whereas β-lactam antibiotics have poor solubility. With the exception of rifampin and chloramphenicol, the commonly used antimicrobial agents do not readily penetrate the normal blood-brain barrier, but the passage of penicillin and other antimicrobial agents is enhanced in the presence of meningeal inflammation (Table 412-4). Antimicrobial drugs should be administered IV throughout the treatment period; the dose should not be reduced as the patient improves because normalization of the blood-brain barrier during recovery reduces the achievable CSF drug levels. Bactericidal drugs (penicillin, ampicillin, third-generation cephalosporins) are preferred whenever possible, and CSF levels of antibiotics at least 10 to 20 times the minimal bactericidal concentration are needed for optimal therapy. Some antibiotics are removed from CSF by active transport into blood via the epithelium of the choroid plexus; by comparison, third-generation cephalosporin antibiotics persist in CSF for longer periods. First- or second-generation cephalosporins and clindamycin do not provide effective levels in CSF and should not be used.

Empirical Treatment

Initial treatment of presumed bacterial meningitis when the etiologic agent cannot be identified on a Gram-stained smear of CSF is based on the available clinical clues.[10] In older children and adults, therapy with vancomycin and a third-generation cephalosporin (cefotaxime or ceftriaxone) is recommended (see Table 412-3). In adults older than 50 years and in high-risk groups, ampicillin is also added because of the possibility of the presence of *L. monocytogenes,* which is susceptible to ampicillin or amoxicillin but not to third-generation cephalosporins. In a penicillin-allergic individual, trimethoprim-sulfamethoxazole is a suitable alternative for *Listeria* meningitis. In special settings, such as nosocomial meningitis associated with neurosurgical procedures or penetrating head trauma, more resistant species such as methicillin-resistant *S. aureus* (MRSA), coagulase-negative staphylococci, and *P. aeruginosa* may be responsible; in these situations, vancomycin in addition to cefepime is indicated as initial therapy.

Meningitis of Specific Bacterial Cause

Pneumococcal Meningitis

The treatment of choice for pneumococcal meningitis in adults has historically been penicillin, with vancomycin (or chloramphenicol) being a reasonable alternative in patients allergic to penicillin (see later). However, penicillin-resistant pneumococcal strains are found worldwide, including 25% of clinical isolates in the United States. Thus, antimicrobial susceptibilities should be determined for all pneumococcal isolates from CSF, blood, or sterile body fluids (see Table 412-4). Approximately 9% of pneumococcal isolates

TABLE 412-3 INITIAL EMPIRICAL THERAPY FOR COMMUNITY-ACQUIRED AND NOSOCOMIAL PURULENT MENINGITIS BASED ON AGE AND CLINICAL SETTING (see Table 412-7 for dosing schedules)

PREDISPOSITIONS	LIKELY PATHOGENS	PREFERRED ANTIMICROBIALS	ALTERNATIVE ANTIMICROBIALS
Age			
<1 mo	Group B streptococcus, *Escherichia coli, Listeria monocytogenes*	Amoxicillin/ampicillin plus cefotaxime	Amoxicillin/ampicillin plus aminoglycoside
1-23 mo	*Streptococcus pneumoniae, Neisseria meningitidis*, group B streptococci, *Haemophilus influenzae, E. coli*	Vancomycin* plus ceftriaxone or cefotaxime	Meropenem (? plus vancomycin*)
2-50 yr	*N. meningitidis, S. pneumoniae*	Vancomycin* plus ceftriaxone or cefotaxime	Meropenem (? plus vancomycin*)
>50 yr	*S. pneumoniae, N. meningitidis, L. monocytogenes*	Vancomycin* plus ceftriaxone or cefotaxime plus ampicillin	Vancomycin* plus ceftriaxone or cefotaxime plus trimethoprim-sulfamethoxazole
Impaired immunity	*L. monocytogenes*, gram-negative bacilli, *S. pneumoniae*	Ampicillin plus ceftazidime or meropenem plus vancomycin*	Trimethoprim-sulfamethoxazole plus meropenem
Cerebrospinal fluid leak or basilar skull fracture	*S. pneumoniae*, various streptococci, *H. influenzae*	Vancomycin* plus cefotaxime or ceftriaxone	Vancomycin* plus meropenem
After neurosurgery or penetrating trauma	*S. aureus*, coagulase-negative staphylococci, aerobic gram-negative bacilli (including *P. aeruginosa*)	Vancomycin* plus cefepime	Vancomycin* plus ceftazidime or vancomycin* plus meropenem
Cerebrospinal fluid shunts (external or internal)	Coagulase-negative staphylococci, *S. aureus*, aerobic gram-negative bacilli (including *P. aeruginosa*), *Propionibacterium acnes*	Vancomycin* plus cefepime	Vancomycin* plus ceftazidime or vancomycin* plus meropenem

Modified from Tunkel AR, Hartman BJ, Kaplan SL, et al. Practice guidelines for the management of bacterial meningitis. *Clin Infect Dis.* 2004;39:1267-1284.
*If dexamethasone is also administered, consideration should be given to the addition of rifampin.

TABLE 412-4 PERMEABILITY OF ANTIBIOTICS INTO CEREBROSPINAL FLUID

GOOD CONCENTRATIONS IN CSF WITH AND WITHOUT MENINGITIS	ADEQUATE CONCENTRATIONS IN CSF IN MENINGITIS	FAIR TO POOR CONCENTRATIONS IN CSF IN MENINGITIS
Chloramphenicol	Penicillin	Early cephalosporins
Sulfonamides	Ampicillin	Cephalothin
Cephalosporins	Methicillin	Cefoxitin
Cefotaxime	Oxacillin	Aminoglycosides
Ceftriaxone	Nafcillin	Gentamicin
Ceftazidime	Carbenicillin	Tobramycin
Moxalactam	Ticarcillin	Amikacin
Cefepime	Tetracycline	Clindamycin
Metronidazole	Erythromycin	Benzathine penicillin
Trimethoprim-sulfamethoxazole	Ethambutol	
Isoniazid	Rifampin	
	Vancomycin	
	Meropenem	

Courtesy Allen Aksamit, Mayo Clinic, Rochester, Minn.
CSF = cerebrospinal fluid.

from patients with meningitis in the United States are resistant to third-generation cephalosporins, with a minimal inhibitory concentration of 2 µg/mL or greater. If the minimal inhibitory concentration for cefotaxime or ceftriaxone (≤1.0 µg/mL) indicates a susceptible isolate, cefotaxime or ceftriaxone would be the drug of choice. If the isolate is highly penicillin resistant or is resistant to 1 µg/mL ceftriaxone or cefotaxime, alternative therapy (vancomycin with or without rifampin IV) is indicated. Because of the increasingly wide distribution of highly resistant strains, initial therapy (pending susceptibility testing) with cefotaxime (or ceftriaxone) in addition to vancomycin IV is recommended. When initial adjunctive therapy with dexamethasone is used (see later) along with vancomycin, it should be borne in mind that vancomycin levels in CSF may be reduced by concomitant corticosteroid use.

Although resistance to chloramphenicol is unusual in pneumococcal isolates from the United States, chloramphenicol has poor bactericidal activity against penicillin-resistant isolates from children with meningitis in South Africa. The relative chloramphenicol resistance of such strains may not be discerned on usual laboratory testing, but it is revealed when the minimum bactericidal concentration is determined. For this reason, vancomycin is preferred over chloramphenicol for the initial treatment of pneumococcal meningitis in a highly penicillin-allergic patient.

The β-lactam antibiotic meropenem is as effective as cefotaxime for meningitis caused by *S. pneumoniae, N. meningitidis,* and *H. influenzae* in adults and in children. Cefepime is also similar to ceftriaxone and cefotaxime for infection with *S. pneumoniae, N. meningitidis,* and *H. influenzae,* and it has greater activity than these antibiotics against *Enterobacter* sp and *P. aeruginosa* (Table 412-5).

Meningococcal Meningitis

Intravenous administration of penicillin G and ampicillin, in doses used to treat meningitis caused by penicillin-susceptible pneumococci, successfully treats *N. meningitidis* meningitis resulting from susceptible strains. Meningococci resistant to penicillin have occasionally been isolated in Spain (≤50% of strains), South Africa, and Canada but rarely in the United States. Most of these isolates have been only intermediately resistant to penicillin (minimal inhibitory concentration of 0.1 to 1.0 µg/mL), although rare strains have had high-level resistance related to β-lactamase production and require third-generation cephalosporins such as ceftriaxone, which is as effective as the potentially more toxic chloramphenicol. Nevertheless, "meningitis doses" of penicillin or ampicillin may provide CSF levels that are sufficient for infections with some strains of intermediately penicillin-resistant *N. meningitidis*. Usually, a 7-day course of antibiotics is sufficient.

Haemophilus influenzae Meningitis

At present, 25 to 35% of isolates of *H. influenzae* type b in the United States are β-lactamase producers and are ampicillin resistant; cefotaxime or ceftriaxone is the initial therapy of choice (see Table 412-5). Alternatives include cefepime or the combination of chloramphenicol and ampicillin; if the isolate proves susceptible to ampicillin, chloramphenicol may be discontinued. Although more than 50% of isolates are chloramphenicol resistant in some areas of Spain, less than 1% of isolates have been found to be resistant in the United States. A 10-day course of antibiotics is usually sufficient.

Staphylococcal Meningitis

For the treatment of adult meningitis caused by MRSA or in a penicillin-allergic patient, vancomycin is the alternative of choice (Tables 412-6 and 412-7). Because penetration of vancomycin into CSF is limited, adjunctive intrathecal (or intraventricular) therapy with vancomycin (without preservative) is occasionally used when CSF cultures have remained positive after 48 hours of intravenous therapy alone and CSF levels can be monitored. For adult meningitis caused by MRSA, intravenous vancomycin (with adjunctive intra-

TABLE 412-5 ANTIMICROBIAL THERAPY FOR COMMUNITY-ACQUIRED BACTERIAL MENINGITIS OF KNOWN CAUSE IN ADULTS OR CHILDREN (See Table 412-7 for Dosing Schedules)

ORGANISM	PREFERRED ANTIMICROBIAL THERAPY	ALTERNATIVE ANTIMICROBIAL THERAPY
Streptococcus pneumoniae		
Penicillin MIC < 0.1 μg/mL	Penicillin G or ampicillin	Cefotaxime, or ceftriaxone, or vancomycin, or chloramphenicol
Penicillin MIC 0.1-1 μg/mL	Ceftriaxone or cefotaxime	Vancomycin,* or meropenem, or cefepime
Penicillin MIC ≥ 2.0 μg/mL	Vancomycin* (plus cefotaxime or ceftriaxone)	Moxifloxacin or gatifloxacin
Cefotaxime or ceftriaxone MIC ≥ 1.0 μg/mL	Vancomycin* (plus cefotaxime or ceftriaxone)	Moxifloxacin or gatifloxacin
Neisseria meningitidis		
Penicillin MIC < 0.1 μg/mL	Penicillin G or ampicillin	Ceftriaxone, or cefotaxime, or chloramphenicol
Penicillin MIC 0.1-1.0 μg/mL	Ceftriaxone or cefotaxime	Chloramphenicol, or meropenem, or gatifloxacin, or moxifloxacin
Haemophilus influenzae		
β-Lactamase negative	Ampicillin	Ceftriaxone, or cefotaxime, or cefepime, or chloramphenicol
β-Lactamase positive	Ceftriaxone or cefotaxime	Cefepime, or chloramphenicol, or gatifloxacin, or moxifloxacin
Listeria monocytogenes	Ampicillin† or penicillin G†	Trimethoprim-sulfamethoxazole or meropenem
Streptococcus agalactiae (group B streptococci)	Ampicillin† or penicillin G†	Cefotaxime or ceftriaxone

*Addition of rifampin should be considered. Consider intrathecal (or intraventricular vancomycin, 5 to 20 mg/day) if not responding to intravenous therapy.
†Addition of intravenous gentamicin should be considered.
MIC = minimal inhibitory concentration.

TABLE 412-6 THERAPY FOR NOSOCOMIAL MENINGITIS OF KNOWN BACTERIAL CAUSE IN ADULTS

ORGANISM	THERAPY OF CHOICE	ALTERNATIVE THERAPY
Staphylococcus aureus		
Methicillin susceptible	Nafcillin or oxacillin; in difficult cases may add rifampin	Vancomycin or meropenem
Methicillin resistant	Vancomycin; in difficult cases may add rifampin	Daptomycin, ceftaroline, or trimethoprim-sulfamethoxazole
Coagulase negative	Vancomycin; may consider addition of rifampin	Daptomycin
Enterococcus sp		
Ampicillin susceptible	Ampicillin plus gentamicin	Vancomycin plus gentamicin
Ampicillin resistant	Vancomycin plus gentamicin	Daptomycin
Ampicillin and vancomycin resistant	Daptomycin	
Escherichia coli and other	Cefotaxime, ceftriaxone, or cefepime	Meropenem or aztreonam or ampicillin or trimethoprim-sulfamethoxazole
Enterobacteriaceae*		
*Pseudomonas aeruginosa**	Cefepime or ceftazidime	Meropenem or aztreonam or ciprofloxacin

Modified from van de Beek D, Drake JM, Tunkel AR. Nosocomial bacterial meningitis. *N Engl J Med.* 2010;362:146-154.
*Selection of specific antimicrobial drug should be based on in vitro susceptibility results, with consideration given to the addition of an aminoglycoside (e.g., tobramycin, gentamicin, or amikacin).

thecal vancomycin as needed) is the treatment of choice. In severe or refractory cases, the addition of rifampin is warranted.

Listeria Meningitis

Ampicillin is the drug of choice for *Listeria* meningitis. When combined with gentamicin, it can have a synergistic bactericidal effect. Third-generation cephalosporins and vancomycin are not effective. In patients allergic to ampicillin, intravenous trimethoprim-sulfamethoxazole may be used, followed by oral trimethoprim alone.

Gram-Negative Bacillary Meningitis

Cefotaxime or ceftriaxone (see Tables 412-6 and 412-7) is used to treat meningitis known to be caused by susceptible gram-negative bacilli (e.g., *E. coli, Klebsiella, Proteus*), but they should not be used to treat meningitis caused by less susceptible species such as *P. aeruginosa* and *Acinetobacter*. After identifying the specific pathogen and determining its drug susceptibilities, alterations in antimicrobial therapy may be indicated. Although experience is limited, fluoroquinolones can also be used. If the organism is *P. aeruginosa,* ceftazidime or cefepime is recommended and may be combined with vancomycin (see Tables 412-6 and 412-7).

Zoonotic Meningitis

Brucella meningitis (Chapter 310) is a subacute or chronic process that is often accompanied by other manifestations of neurobrucellosis (encephalitis, polyradiculitis, myelitis). Infection is transmitted to humans in endemic areas (Central and South America, Mediterranean littoral, Arabian peninsula) from the ingestion of unpasteurized milk or cheese or direct contact with domestic animals. Neurobrucellosis occurs in 2 to 5% of patients with brucellosis. CSF findings consist of a lymphocytic pleocytosis (<500 cells/μL), hypoglycorrhachia, and an elevated protein level, findings that could mistakenly suggest tuberculous meningitis. The diagnosis is based on demonstration of antibody in serum and CSF or by isolation of *Brucella* from blood; the microorganism is isolated from CSF in only a minority of cases. Treatment of adults involves the three-drug combination of doxycycline (200 mg/day), rifampin (600 mg/day), and trimethoprim-sulfamethoxazole (20 mg/kg/day IV, based on trimethoprim component, in 6-hour aliquots) for several months, depending on the clinical and CSF responses.

Streptococcus suis is an uncommon cause of meningitis seen in pig breeders, butchers, and abattoir workers in Europe, Canada, and China. *S. suis* meningitis, which is an acute illness with a brisk neutrophilic pleocytosis, is often initially mistaken for pneumococcal meningitis on the basis of Gram stain of CSF. Treatment of adults consists of penicillin (12 to 24 million units [U]/day in 4-hour aliquots) or ampicillin (12 g/day in 4-hour aliquots) IV for 10 to 14 days.

Bacillus anthracis (Chapter 294) is a rare cause of meningitis that most often develops as a complication of inhalation anthrax following exposure to aerosols of anthrax spores in the setting of large-scale processing of wool and hides or a bioterrorism attack (Chapter 21). Anthrax meningitis is an acute process characterized by hemorrhagic or serohemorrhagic CSF with a neutrophilic predominance (several thousand cells per cubic millimeter), hypoglycorrhachia, an elevated protein level, and prominent large gram-positive bacilli on stained smear. Treatment of adults initially includes ciprofloxacin (400 mg at 12-hour intervals) in addition to penicillin (24 million U/day in 4-hour aliquots) and chloramphenicol (4 g/day in 6-hour aliquots) IV. Alternatively, treatment could substitute levo- or moxifloxacin for ciprofloxacin, meropenem for penicillin, and linezolid or chloramphenicol (Chapter 278). Whether all drugs are continued (or treatment is narrowed to one or two antimicrobials) and the duration of treatment depend on whether the meningitis is of suspected bioterrorist origin (Chapter 21) or caused by cutaneous anthrax resulting from animal (or animal product) exposure (Chapter 294). Consultation with infectious disease and public health authorities should be sought.

Duration of Therapy

The frequency of CSF examination depends on the clinical course, but examination should be repeated in 24 to 48 hours if there has not been satisfactory improvement or if the causative microorganism is a more resistant gram-negative bacillus or a highly penicillin-resistant (or cephalosporin-resistant) *S. pneumoniae* strain, especially in patients who are receiving adjunctive dexamethasone therapy. Routine "end-of-treatment" CSF examination is unnecessary in most patients with the common types of community-acquired bacterial meningitis. Although 5 days of ceftriaxone treatment is as good as

TABLE 412-7 DOSES OF ANTIMICROBIAL DRUGS FOR TREATMENT OF BACTERIAL MENINGITIS*

ANTIMICROBIAL DRUG	ADULTS (24-HR DOSE)	INFANTS AND CHILDREN (24-HR DOSE)
β-LACTAMS		
Penicillin G	24 million U, q4h aliquots	300,000 U/kg, q4h aliquots
Ampicillin	12 g, q4h aliquots	300 mg/kg, q4h aliquots
Nafcillin	10-12 g, q4h aliquots	200 mg/kg, q4h aliquots
Oxacillin	10-12 g, q4h aliquots	200 mg/kg, q4h aliquots
Aztreonam (a monobactam)	6-8 g, q6-8h aliquots	
Meropenem (a carbapenem†)	6 g, q8h aliquots	120 mg/kg, q8h aliquots
CEPHALOSPORINS		
Cefotaxime	12 g, q4h aliquots	200-300 mg/kg, q6h aliquots
Ceftriaxone‡	4 g, q12h aliquots	80-100 mg/kg, q12h aliquots
Ceftazidime	6 g, q8h aliquots	150 mg/kg, q8h aliquots
Cefepime	6 g, q6-8h aliquots	150 mg/kg, q8h aliquots
Ceftaroline	600 mg, q12h aliquots	Safety not established in children
AMINOGLYCOSIDES		
Gentamicin§	5 mg/kg, q8h aliquots	7.5 mg/kg, q8h aliquots
Tobramycin§	5 mg/kg, q8h aliquots	7.5 mg/kg, q8h aliquots
Amikacin§	15 mg/kg, q8h aliquots	20-25 mg/kg, q8h aliquots
FLUOROQUINOLONES		
Ciprofloxacin	800-1200 mg, q8-12h aliquots	—
Gatifloxacin‖	400 mg, q24h dosing	—
Moxifloxacin‖	400 mg, q24h dosing	—
OTHERS		
Chloramphenicol	4-6 g, q6h aliquots	75-100 mg/kg, q6h aliquots
Vancomycin¶	2-3 g, q6-8h aliquots	50-60 mg/kg, q6h aliquots
Rifampin	600 mg, q24h dosing	10-20 mg/kg, q12-24h aliquots
Trimethoprim-sulfamethoxazole**	20 mg/kg, q6h aliquots	20 mg/kg, q6h aliquots
Daptomycin	6 mg/kg, q24h dosing	6 mg/kg, q24h dosing

*Dosages are intravenous and for patients with normal renal and hepatic function.
†Use may be associated with seizures, but much less so than with imipenem.
‡Four-gram maximum daily dose.
§Peak and trough serum levels should be monitored.
‖No data are available on the optimal dosage required for bacterial meningitis.
¶Monitoring of trough serum levels is advisable; they should be maintained at concentrations of 15 to 20 μg/mL. If the patient is not responding well, one may need to monitor cerebrospinal fluid levels and, if low, temporarily increase the daily dose accordingly or add adjuvant intrathecal vancomycin (5 to 20 mg), as for the treatment of methicillin-resistant *Staphylococcus aureus* meningitis.
**Dosage based on the trimethoprim component of the combination.

10 days in children who are stable at 5 days,[] longer courses are still recommended in adults. Meningococci are rapidly eliminated from the circulation and CSF with appropriate antimicrobial therapy, which should be continued for 4 to 7 days after the patient becomes afebrile. If the patient has responded well, a follow-up lumbar puncture is not necessary. *H. influenzae* meningitis should be treated for 7 to 10 days. Follow-up CSF examination may be omitted in patients who have responded with rapid clinical resolution of the meningitis. In pneumococcal meningitis, antimicrobial treatment should be continued for 10 to 14 days and follow-up examination of CSF should be performed, particularly when the patient has coexistent mastoiditis. More prolonged therapy is indicated with concomitant parameningeal infection. Meningitis caused by *L. monocytogenes* should be treated for 21 days. Treatment of gram-negative bacillary meningitis with parenteral antimicrobials is prolonged, usually for a minimum of 3 weeks (particularly in patients after a recent neurosurgical procedure) to prevent relapse. Repeated examinations of CSF are necessary both during and at the conclusion of treatment to determine whether bacteriologic cure has been achieved.

Other Aspects of Treatment

Adjunctive Corticosteroids

In children, the routine use of dexamethasone administered IV (either 0.15 mg/kg every 6 hours for 4 days or 0.4 mg/kg every 12 hours for 2 days), either at the time of or 10 to 20 minutes before initiating antimicrobial therapy (third-generation cephalosporin), has no effect on mortality but reduces the incidence of neurologic sequelae (primarily bilateral sensorineural hearing loss). In adults with community-acquired bacterial meningitis, adjunctive dexamethasone therapy (10 mg every 6 hours IV or 4 days) significantly reduced the proportion of patients with an unfavorable neurologic outcome from 25 to 15% or a fatal outcome from 15 to 7%.[] Adverse events were not increased in those receiving dexamethasone. Notably, the risk for gastrointestinal bleeding was not increased in the dexamethasone-treated group. The beneficial effect of dexamethasone was most evident in the subgroup of patients with pneumococcal meningitis, in whom the rate of unfavorable outcomes was reduced from 52 to 26% and deaths from 34 to 14%. In a study of adolescents and adults with bacterial meningitis in Vietnam, dexamethasone significantly reduced death and disability by approximately 54% at 6 months in patients with confirmed disease but not in those with suspected disease.[] By comparison, adjunctive corticosteroids were not effective in treating bacterial meningitis in a large trial of predominantly HIV-positive patients in sub-Saharan Africa.[] Based on these data, adjunctive dexamethasone (0.15 mg/kg every 6 hours for 2 to 4 days, with the initial dose given 10 to 20 minutes before or simultaneously with the initial dose of antimicrobial therapy) is recommended in adults with suspected or demonstrated pneumococcal meningitis and perhaps routinely in all cases of bacterial meningitis, at least in non–HIV-infected patients in high-income countries,[] and its benefits extend for at least 13 years after the event.[] When vancomycin is used for treatment of meningitis resulting from highly cephalosporin-resistant *S. pneumoniae,* as is recommended in the United States, the addition of rifampin should be considered because dexamethasone may reduce the CSF concentration of vancomycin (see Table 412-4).

Elevated Cerebrospinal Fluid Pressure (Brain Swelling)

Occasional patients with acute bacterial meningitis experience marked brain swelling (CSF pressure > 450 mm H_2O), which may lead to temporal lobe or cerebellar herniation after lumbar puncture. To decrease the possibility of this complication when the pressure is found to be this high, only a small amount of CSF should be removed for analysis (the amount present in the manometer), and a 20% solution of mannitol (0.25 to 0.5 g/kg IV) should be infused over a period of 20 to 30 minutes while monitoring (if possible) for a decline in CSF pressure to a lower level before the spinal needle is removed. Continued control of increased intracranial pressure, if needed thereafter, may be effected with additional mannitol; dexamethasone (10 mg IV, followed by 0.15 mg/kg every 6 hours) should be used in patients with brain swelling regardless of the suspected bacteriologic cause of meningitis.

In a stuporous patient or one with respiratory insufficiency and markedly increased intracranial pressure, use of a ventilator to reduce the arterial carbon dioxide pressure to between 25 and 32 mm Hg is reasonable, and the patient's head should be elevated 30 to 45 degrees. Intubation should be performed with minimal stimulation to avoid an appreciable further rise in pressure; pharmacologic aids to intubation are recommended, such as succinylcholine and opioids, with the possible use of adjunctive intravenous lidocaine. Subsequently, transient increases in intracranial pressure associated with hyperactive airway reflexes can be mitigated by intratracheal instillation of lidocaine before vigorous suctioning. With continued marked and fluctuating elevations in intracranial pressure, use of a continuous intracranial monitoring device may be warranted. Induced hypothermia is not beneficial and may be harmful.[]

Hypotension

Initial hypovolemia or hypotension, if present, should be treated with fluid to prevent significantly decreased cerebral blood flow. Over the next 24 to 48 hours, inappropriate secretion of antidiuretic hormone may contribute to further brain swelling; in such cases, fluid should be restricted to 1200 to 1500 mL daily in adults if possible, although a study in children suggests that routine fluid restriction does not improve outcome and that the resulting decrease in extracellular water may increase the likelihood of hypovolemia and an adverse outcome.

Supportive Care

Patients with acute bacterial meningitis should receive constant nursing attention in an intensive care unit to ensure prompt recognition of seizures and to prevent aspiration. If seizures occur, they should be treated acutely in adults with diazepam (administered slowly IV at a dose of 5 to 10 mg) or lorazepam (4 to 8 mg). Maintenance anticonvulsant therapy can be continued thereafter with intravenous phenytoin (Chapter 403) until the medication can be administered orally (PO). Sedation should be avoided because of the danger of respiratory depression and aspiration.

Surgery

Surgical treatment of an accompanying pyogenic focus such as mastoiditis should be undertaken when recovery from the meningitis is as complete as possible but under continuing antibiotic administration. Rarely, the mastoid infection (e.g., Bezold abscess) is so hyperacute that early drainage may be required after 48 hours or so of antibiotic therapy when the acute meningeal process has subsided somewhat.

PROGNOSIS

Prompt treatment of bacterial meningitis usually results in rapid recovery of neurologic function. Persistent or late-onset obtundation and coma without focal findings suggest brain swelling, subdural effusion, hydrocephalus, loculated ventriculitis, cortical thrombophlebitis, or sagittal sinus thrombosis. The last three conditions are commonly associated with fever and continuing pleocytosis.

The mortality rate for community-acquired bacterial meningitis in adults varies with the etiologic agent and clinical circumstances. With current antimicrobial therapy, the mortality rate for *H. influenzae* meningitis is less than 5% and that for meningococcal meningitis is approximately 10%. The highest mortality is seen with pneumococcal (20%) and *L. monocytogenes* (20 to 30%) meningitis.

The mortality rate for gram-negative bacillary meningitis, commonly nosocomial in origin, has been 20 to 30% in adults, but may be decreasing. The mortality rate for recurrent community-acquired meningitis in adults (≈5%) is lower than the 20% rate for nonrecurrent episodes. Poor prognostic factors include advanced age, the presence of other foci of infection, underlying diseases (leukemia, alcoholism), obtundation, seizures within the first 24 hours, and delay in instituting appropriate therapy.

Residual neurologic damage is seen in 10 to 20% of patients who recover from bacterial meningitis. Approximately 25% of adults considered clinically well recovered (expected to function independently and resume activities of daily life, including work) from pneumococcal meningitis show neuropsychological abnormalities, mainly loss of cognitive speed, when they are examined 6 to 24 months after hospital discharge. Developmental delay and speech defects are each observed in approximately 5% of children, and bacterial meningitis is associated with lower subsequent educational achievement and economic self-sufficiency in adulthood.[11]

PREVENTION

Vaccination

The meningococcal vaccine is approximately 85% protective against only four of the strains that cause illness: A, C, Y, and W-135 (Chapters 18 and 298). Making a universal vaccine against the B strains has been challenging because there are many types that cause illness in different parts of the world. All 11- to 12-year-olds should be vaccinated with meningococcal conjugate vaccine, and a booster dose should be given at age 16 years (Chapter 18). For adolescents who receive the first dose at age 13 through 15 years, a one-time booster dose should be administered, preferably at age 16 through 18 years, before the peak in increased risk. Adolescents who receive their first dose of quadrivalent meningococcal conjugate vaccine at or after age 16 years do not need a booster dose. Effective vaccines are now available for many subtypes of *H. influenzae* type b (Chapter 300). Adherence to recommended vaccination (Chapter 18) substantially reduces meningitis from each of these organisms.

Chemoprophylaxis

Prompt prophylaxis of close contacts (individuals who frequently slept and ate in the same household with the patient, girlfriend, or boyfriend) is warranted because up to a third of secondary cases of meningococcal disease develop within 2 to 5 days of illness in the initial case. Only hospital personnel who were in close contact with a patient (mouth-to-mouth resuscitation, initial examination before institution of respiratory precautions) are at special risk. Commonly, oral rifampin is used for prophylaxis: for adults (other than pregnant women), 600 mg twice daily for 2 days; for children, 10 mg/kg twice daily for 2 days. Alternatively, for adults, ciprofloxacin (500 mg), ofloxacin (400 mg), or azithromycin (500 mg), each given PO as a single dose, may be used. Another choice is ceftriaxone intramuscularly as a single dose in adults (250 mg) or children (125 mg).

Widespread use of *H. influenzae* type b polysaccharide protein-conjugate vaccine in developed countries has largely eliminated the need for chemoprophylaxis of close childhood contacts of patients with *H. influenzae* meningitis or invasive infection. However, prophylaxis would be indicated for unimmunized close household contacts of an index patient (e.g., recent immigrant) younger than 6 years. If two or more cases of invasive *H. influenzae* type b disease occur in children at a daycare center, prophylaxis of other unimmunized attendees is warranted with rifampin (20 mg/kg/day PO) for 4 days.

VIRAL MENINGITIS

DEFINITION

The nonspecific term *aseptic meningitis* describes an inflammatory process involving the meninges, usually accompanied by a mononuclear pleocytosis, without evidence of pyogenic bacterial infection on Gram stain or culture. The definition encompasses various processes that produce similar clinical pictures and inflammatory responses: viral meningitis, atypical and nonpyogenic bacterial and fungal meningitis, chemically induced meningitis, drug-induced meningitis, neoplastic meningitis, meningeal inflammation caused by adjacent pyogenic infections, and meningitis associated with autoimmune hypersensitivity diseases. Aseptic meningitis, which is usually an acute or subacute process, can be further divided into types by the duration of illness (chronic versus chronic-intermittent) and distinctive cellular responses in CSF (e.g., eosinophilic meningitis).

Many of the viruses causing meningitis also may cause infection of the brain parenchyma (encephalitis; Chapter 414) or spinal cord. Sometimes, parenchymatous involvement and meningeal involvement occur simultaneously in the same patient and are referred to as meningoencephalitis and meningomyelitis.

EPIDEMIOLOGY

Most cases of community-acquired aseptic meningitis are the result of viruses, principally enteroviruses, which account for more than 60% of viral meningitides and for 90% of those for which an etiologic agent is identified (Table 412-8). Enteroviruses are members of the Picornaviridae (small RNA) family, which consists of more than 60 serotypes: 28 echoviruses, 23 group A and 6 group B coxsackieviruses, 4 numbered enteroviruses (68 to 71), and 3 polioviruses. The most common serotypes implicated in viral meningitis from year to year have been echoviruses 4, 6, 9, 11, 16, and 30 (most recently 13 and 33) and coxsackie B serotypes 2 to 5. Currently, poliovirus infections (Chapter 379) are limited to parts of Asia and Africa, although rare cases occur secondary to attenuated vaccine strains.

Many viruses that produce the clinical picture of aseptic meningitis, such as arthropod-borne viruses, HSV-1, enterovirus 71, lymphocytic choriomeningitis virus, mumps virus, HIV-1, cytomegalovirus, and Epstein-Barr virus, also can produce the clinical picture of meningoencephalitis and encephalitis (Chapter 414). In addition, some viruses involve the spinal cord, including the anterior horn cells (poliovirus, West Nile virus) or the dorsal root ganglia (HSV-2).

Enterovirus

An estimated 10 to 15 million clinical enteroviral infections (Chapter 379) occur annually in the United States, and these include an estimated 50,000 to 75,000 cases of enteroviral meningitis. In temperate climates, enteroviral meningitis peaks during the summer and fall, especially in children. Serotypes tend to cycle with varying periodicity, and outbreaks are related to lack

TABLE 412-8 AGENTS OF VIRAL MENINGITIS

COMMON
Nonarthropod Viruses
Picornavirus (RNA)
Enterovirus
Echovirus
Coxsackie A
Coxsackie B
Enterovirus 70, 71
Poliovirus
Herpes simplex virus type 2 (HSV-2) (DNA)
Arthropod-Borne Viruses (Arboviruses)
Togavirus (alphavirus, RNA)
Eastern equine encephalitis (EEE)
Western equine encephalitis (WEE)
Venezuelan equine encephalitis (VEE)
Flavivirus (RNA)
St. Louis encephalitis (SLE)
West Nile virus (WNV)
Bunyavirus (RNA)
California encephalitis
UNCOMMON
Arenavirus (RNA)
Lymphocytic choriomeningitis (LCM)
Paramyxovirus (RNA)
Mumps
Retrovirus (RNA)
Human immunodeficiency virus (HIV-1)
RARE
Herpesvirus (DNA)
Herpes simplex virus type 1 (HSV-1)
Epstein-Barr virus (EBV)
Cytomegalovirus (CMV)
Varicella-zoster virus (VZV)
Human herpesvirus type 6 (HHV-6)
Adenovirus (DNA)
Coltivirus (RNA)
Colorado tick fever
Bunyavirus (RNA)
Toscana virus (a Phlebovirus)

of previous exposure to a particular serotype. Serotype-specific protective antibodies develop following infection, so subsequent episodes of enteroviral meningitis are uncommon and are caused by a different serotype.

Humans are the only known reservoir of enteroviruses. Enteroviral infection is spread predominantly by the fecal-oral route and occasionally by the respiratory route.

Herpes Simplex Virus

HSV (Chapter 374) accounts for 1 to 3% of all episodes of aseptic meningitis and occurs most commonly in sexually active adults or adolescents. In individuals with primary genital herpes (HSV-2) infection, up to 36% of women and 13% of men have symptoms of aseptic meningitis. Recurrences of genital herpes are common and are sometimes accompanied by aseptic meningitis. More than 80% of cases of benign recurrent aseptic meningitis are caused by HSV-2. In contrast, HSV-1 CNS infection almost always manifests as encephalitis rather than aseptic meningitis. Herpesviruses also may be reactivated in patients taking immunomodulatory drugs, which are often used to treat autoimmune diseases.

Arboviruses

Although the most common form of CNS infection caused by arboviruses (Chapters 383 and 414) is encephalitis, aseptic meningitis also may occur. These vector-borne viruses are introduced subcutaneously by a mosquito (e.g., West Nile virus, Japanese B encephalitis), tick (e.g., Colorado tick fever), or sandfly (e.g., Toscana virus). Birds, which are vectors of mosquito-borne arboviruses, may not be obviously sick, although West Nile virus may cause prominent die-offs of corvine species, especially crows and blue jays, which can provide clues to an outbreak affecting humans.

The geographic spread of alphavirus infections (Eastern equine encephalitis, Western equine encephalitis, Venezuela equine encephalitis) in the United States is determined by the range of their individual mosquito vectors. Eastern equine encephalitis occurs sporadically or as focal outbreaks in the summer in the eastern and Gulf coasts, most frequently in children and elderly persons. Western equine encephalitis occurs predominantly in the western states, and Venezuela equine encephalitis is found in Florida. St. Louis encephalitis infections were originally recognized in the Midwest, but sporadic cases and outbreaks have occurred more recently in most parts of the United States; it is the most common arbovirus causing aseptic meningitis in the United States. West Nile virus infections first appeared in the United States in 1999 and now account for approximately 3000 cases of meningitis and another 3000 cases of encephalitis annually.

Mumps

Mumps virus (Chapter 369) was the leading identifiable cause of viral meningitis before widespread immunization in the 1960s. Episodes occurred most frequently in the winter and spring. It is now an uncommon cause of viral meningitis in the United States.

Lymphocytic Choriomeningitis

Lymphocytic choriomeningitis virus is transmitted to humans by rodents through direct contact, through ingestion of animal-contaminated food, or via aerosol or an animal bite. Cases tend to occur in early winter when mice seek shelter in homes. Outbreaks have occurred following exposure to pet or laboratory hamsters. Currently, lymphocytic choriomeningitis virus is infrequently a cause of aseptic meningitis.

PATHOBIOLOGY

The two basic routes for virus to gain access to the CNS are hematogenous (enteroviral infection) or neuronal (HSV infection). Enteroviruses pass through the stomach, where they resist the acid pH, and proceed to the lower gastrointestinal tract. Some virus also undergoes replication in the nasopharynx and spreads to regional lymphatics. After presumably binding to specific enterocyte receptors, the virus breaches the epithelial lining and undergoes primary replication in a permissive cell. From there, the virus progresses to Peyer patches, where further replication occurs. A minor enterovirus viremia then seeds the CNS, heart, liver, and reticuloendothelial system. Following extensive replication at the latter sites, a major viremia ensues, often accompanying the onset of clinical illness. The mechanism by which enterovirus enters the CNS is presumed to involve crossing the blood-CSF barrier's tight endothelial junctions and then entering CSF, probably at the choroid plexus.

In contrast, HSV infections may reach the CNS via the neuronal route: in HSV-1 encephalitis, from oral sites via the trigeminal and olfactory nerve; in HSV-2 (and the rare HSV-1) aseptic meningitis, by spread from a primary genital lesion and ascent along the sacral nerve roots to the meninges. After subsidence of the primary infection, HSV-1 may remain dormant in the trigeminal or olfactory root ganglia only to reactivate at a later date, enter the temporal lobe, and produce encephalitis. Similarly, HSV-2 may remain latent in the sacral root ganglia until subsequent reactivation causes later episodes of aseptic meningitis.

CLINICAL MANIFESTATIONS

Enteroviral Meningitis

The clinical features of enteroviral meningitis (Chapter 379) in older children and adults often begin abruptly with headache (85 to 100%), fever (80 to 100%), and stiff neck (50 to 80%). In some patients the course is biphasic, with the initial prodromal phase being characterized by low-grade fever and nonspecific symptoms (malaise, sore throat, diarrhea), followed by a second phase at which time the meninges are seeded, with the development of higher fever, nausea, vomiting, myalgia, photophobia, and stiff neck. Other enteroviral syndromes may coexist, particularly pleurodynia or pericarditis resulting from coxsackieviruses. Rash may be a manifestation of infections caused by echoviruses, particularly echovirus type 9, coxsackieviruses A9 and A16, and enterovirus 71; the latter three cause hand-foot-and-mouth disease, which may occur alone or accompany aseptic meningitis. Echovirus 9 epidemics often produce syndromes of exanthem, enanthem (small, grayish white lesions resembling Koplik spots on the buccal mucosa), and aseptic meningitis, either alone or in combination; a macular and petechial rash in the presence of a meningitic syndrome must be differentiated from meningococcal meningitis.

Neurologic abnormalities affecting the cerebrum are rarely observed because such cases would be defined as encephalitis or meningoencephalitis rather than enteroviral meningitis. In agammaglobulinemic individuals

in whom enteroviral CNS infection develops, meningitis may progress to a chronic meningoencephalitis with multiple neurologic features, including headache, seizures, ataxia, weakness, hearing loss, obtundation, and coma.

The clinical course of enteroviral meningitis is benign, even in the minority of patients in whom the onset is acute and even fulminant. Symptoms subside within a week in children but may continue for several weeks in adults.

Herpes Simplex Virus Type 2 Meningitis

Aseptic meningitis is a common complication of primary genital HSV-2 infection (Chapter 374); up to 36% of women and 13% of men have headache (developing over 2 to 3 days), stiff neck, and photophobia. Clinical features of meningitis occur 3 to 12 days after the appearance of genital lesions and usually last for 4 to 7 days. Neurologic complications occur in up to 37% of patients and include dysesthesia or paresthesia in the perineum or sacral area, urinary retention, and constipation; evidence of transverse myelitis with motor weakness in the lower extremities, hyporeflexia, and paraparesis occasionally ensues. Recurrent episodes of HSV-2 meningitis may occur at intervals of months or years in 20% of patients. In recurrent HSV-2 meningitis, fever may develop but is not as prominent as in bacterial or acute enteroviral meningitis. Recurrent vesicular lesions, paresthesia, or dysesthesia in areas of previous genital herpes may or may not precede individual recurrences of meningitis. Between recurrences, CSF findings and clinical manifestations return to normal. In patients who have had neurologic complications with a first episode of HSV-2 meningitis, the findings subside within 6 months.

Mumps Meningitis

Symptomatic CNS disease, principally meningitis or meningoencephalitis, occurs in 1 to 10% of patients with mumps parotitis (Chapter 369), but pleocytosis occurs in more than 50% of patients with mumps, most of whom lack CNS symptoms. When meningitis occurs in patients with mumps, it usually follows parotitis by 4 to 10 days, but it may precede parotitis by up to 1 week. The typical features of viral meningitis (headache, fever, vomiting) are each present in 50 to 100% of patients. Stiff neck (40 to 90%) is common, and abdominal pain (perhaps complicating pancreatitis or oophoritis) or orchitis (in ≤ 20% of men with mumps) may be present. Other complications of mumps may involve the nervous system (eighth nerve damage, transient facial nerve paralysis, and rarely, fifth nerve palsy) but are usually independent of mumps meningitis or meningoencephalitis. The incubation period for mumps is 18 to 21 days. When mumps meningitis occurs in the absence of clinical parotitis, it is difficult to distinguish it from other forms of viral meningitis.

When meningitis complicates mumps, fever, which had been low grade, rises to 103° F or higher and persists at this level for 3 or 4 days. Most cases are uncomplicated, with approximately a 10-day duration of illness and then complete recovery. However, symptomatic mumps meningitis may persist for more than 14 days in some patients.

Meningitis Caused by Lymphocytic Choriomeningitis Virus

Lymphocytic choriomeningitis virus infections are uncommon, and clinical illness occurs after an incubation period of 1 to 3 weeks. Illness begins with a grippe-like syndrome of fever, rigors, malaise, myalgia, anorexia, and photophobia. Sore throat and arthralgia or arthritis of the digits are noted by some patients. Orchitis or parotitis occurs rarely. This grippe-like illness lasts 1 to 3 weeks in humans, but 15% of patients have a biphasic illness consisting of transient improvement and then recrudescence, 1 to 2 days later, of fever, photophobia, and more prominent headache. Meningeal signs are observed during the second phase. The duration of meningitis caused by lymphocytic choriomeningitis virus, like that of mumps meningitis, tends to be longer than the 7 to 10 days for enteroviral meningitis.

Meningitis Caused by Human Immunodeficiency Virus

Initial infection with HIV-1 (Chapter 384) is symptomatic in 40 to 90% of patients but is frequently overlooked. The interval between exposure and onset of symptoms is 2 to 4 weeks. This acute illness resembles mononucleosis, with fever, malaise, lymphadenopathy, arthralgia, myalgia, anorexia, nausea, headache, and morbilliform rash. A few patients with this initial syndrome have manifestations of aseptic meningitis (headache, photophobia, nausea, vomiting, and stiff neck). Occasionally, encephalopathy or cranial nerve palsies (seventh, eighth, and fifth) develop. Symptoms of the initial HIV-1 aseptic meningitis syndrome last several weeks and then subside. Occasionally, manifestations similar to those of the initial infection may appear later in the course of untreated infection.

DIAGNOSIS

Cerebrospinal Fluid Examination

CSF findings in all types of viral meningitis are similar and consist of a predominantly lymphocytic pleocytosis, usually 50 to 1000/μL but occasionally up to several thousand per cubic millimeter, a normal glucose concentration, and a mildly elevated protein level, usually less than 150 mg/dL. During the first 24 to 48 hours of enteroviral meningitis, a predominance of neutrophils (55 to ≤ 90%) is observed in approximately 50% of patients; subsequently, the principal cells in CSF change to lymphocytes. Occasionally, no pleocytosis is noted in patients proved by culture or PCR to have early enteroviral meningitis. Rarely, hypoglycorrhachia occurs in meningitis resulting from mumps or lymphocytic choriomeningitis virus or in infants with enterovirus.

Polymerase Chain Reaction versus Culture or Antibody Detection

The recent development of reverse-transcription PCR for enteroviruses can reduce detection time to as little as 5 hours, thereby shortening hospital stay and minimizing the unnecessary use of antimicrobial agents. Its sensitivity in CSF is 85 to 100%, with a specificity of 90 to 100%, depending on the laboratory.[12] By comparison, viral culture of enterovirus from CSF has a sensitivity of only 65 to 75% and takes 4 to 8 days.

HSV-2 can be cultured from CSF in approximately 75% of patients with aseptic meningitis during an initial episode of genital HSV-2 infection, but it is rarely isolated from CSF during meningitis associated with recurrent genital herpes. PCR for HSV-2 DNA is usually positive in the CSF of patients with initial episodes of meningitis and is positive in approximately 80% of patients with benign recurrent meningitis caused by lymphocytic choriomeningitis virus.

The diagnosis can be made retrospectively by demonstrating seroconversion in antibody to gG-2 antigen in HSV-2 meningitis. A four-fold rise in titer to mumps or lymphocytic choriomeningitis virus between acute and convalescent sera is also diagnostic. Serodiagnosis is not practical for sporadic enteroviral meningitis because of the lack of specificity of antibodies to individual serotypes.

Differential Diagnosis

The most important process to distinguish from viral meningitis is bacterial meningitis. A predominance of CSF neutrophils, hypoglycorrhachia, and bacteria on Gram-stained smear or culture indicate bacterial meningitis. An early neutrophilic predominance in CSF combined with a macular and petechial rash in enteroviral meningitis may mimic meningococcemia with meningitis. Occasional bacteria and fungi cause meningitis with a predominantly lymphocytic pleocytosis similar to that of most viral meningitides (Table 412-9). Epidemiologic considerations and clinical findings aid in distinguish-

TABLE 412-9 NONVIRAL INFECTIOUS CAUSES OF ASEPTIC MENINGITIS

UNCOMMON	RARE
BACTERIAL	
Leptospira interrogans serovars	*Mycoplasma pneumoniae*
Borrelia burgdorferi	*Ehrlichia chaffeensis*
Treponema pallidum	*Listeria monocytogenes*
Mycobacterium tuberculosis	*Borrelia recurrentis* and *Borrelia hermsii*
Brucella sp	*Chlamydia psittaci*
Parameningeal infections	Staphylococcal enterotoxin or TSST-1
Subacute bacterial endocarditis	*Rickettsia rickettsii* and *Rickettsia prowazekii*
Partially treated bacterial (pyogenic) meningitis	
FUNGAL	
Cryptococcus neoformans	*Blastomyces dermatitidis*
Coccidioides immitis	*Sporothrix schenckii*
Histoplasma capsulatum	*Candida* sp
PROTOZOAN	
	Trypanosoma brucei sp
	Toxoplasma gondii
	Acanthamoeba sp

TSST-1 = toxic shock syndrome toxin 1.

ing leptospiral, Lyme *Borrelia,* and syphilitic meningitis, whereas hypoglycorrhachia suggests tuberculous and cryptococcal meningitis.

PREVENTION AND TREATMENT

Rx

The introduction of live attenuated mumps vaccine in the United States reduced mumps from the leading cause of aseptic meningitis and meningoencephalitis to the point at which it occurs only rarely. Chronic enteroviral meningitis and meningoencephalitis in agammaglobulinemic patients have been controlled by parenteral (even intrathecal) administration of immune globulin.

No approved antiviral chemotherapy is available for enteroviral meningitis. Pleconaril, a drug that prevents attachment of virus to host cells, can produce clinical improvement in agammaglobulinemic patients with chronic enteroviral meningoencephalitis.

Intravenous acyclovir (5 to 10 mg/kg three times daily) is used to treat hospitalized, symptomatic patients with HSV-2 meningitis, particularly when the disease is associated with primary genital herpes, although it has not been shown in clinical trials to alter the course of illness. In patients with frequent recurrences of HSV meningitis, it is reasonable to attempt prophylaxis with oral antivirals: valacyclovir (500 mg/day), famciclovir (250 mg twice daily), or acyclovir (400 mg twice daily).

PROGNOSIS

The course and outcome in patients with enteroviral meningitis are almost always benign, although approximately 1% of patients have subsequent abnormalities, probably reflecting a meningoencephalitic process. Most viral meningitides are self-limited, but some cause chronic or recurrent illness. Persistent meningitis or meningoencephalitis, sometimes fatal, can occur in individuals with hereditary (usually X-linked agammaglobulinemia or common variable immunodeficiency) deficiencies in B-lymphocyte function. HIV-1 may produce a prolonged meningeal inflammation. HSV-2 infection is the most common viral cause of recurrent episodes of aseptic meningitis.

OTHER MENINGITIDES

Nonviral Infectious Causes of Aseptic Meningitis

Categories of aseptic meningitis other than the viral meningitides include nonviral infectious processes (see Table 412-9), noninfectious processes (Table 412-10), chronic meningitides (Table 412-11), recurrent meningitis (Table 412-12), and eosinophilic meningitis (Table 412-13). Nonviral infectious causes are uncommon or rare in comparison to viral or acute suppurative meningitis. Some of the bacterial causes (e.g., *Leptospira* serovars, *B. burgdorferi, Brucella* sp, *T. pallidum*) produce a lymphocytic pleocytosis; others (partially treated bacterial meningitis, subacute bacterial endocarditis with embolic cerebral infarcts) produce a mixed neutrophilic-mononuclear pleocytosis; and *M. tuberculosis,* though producing a lymphocytic response with developing hypoglycorrhachia, may show a predominantly neutrophilic response in a minority of patients early in the disease. Although patients with *L. monocytogenes* infection usually have neutrophilic pleocytosis, this

TABLE 412-10 NONINFECTIOUS CAUSES OF ASEPTIC MENINGITIS

Drug hypersensitivity
Systemic disease
Systemic lupus erythematosus
Familial Mediterranean fever
Behçet syndrome
Granulomatosis with polyangiitis (formerly Wegener)
Cogan syndrome
Sarcoidosis
Still disease
Kawasaki disease
Lead poisoning
Neoplastic disease
Metastatic carcinomatous meningitis
Central nervous system tumors (meningeal gliomatosis, dysgerminomas, ependymomas)
Tumors that leak inflammatory material into cerebrospinal fluid (squamous cells in epidermoid tumors of the posterior fossa, cholesteatomas)
Inflammatory processes involving central nervous system structures primarily
Chemical meningitis following myelography (water-soluble nonionic contrast material)
Continuous spinal and epidural anesthesia, inflammation after neurosurgery
Granulomatous cerebral vasculitis
Vogt-Koyanagi-Harada syndrome

TABLE 412-11 INFECTIOUS CAUSES OF CHRONIC (PERSISTENT) LYMPHOCYTIC MENINGITIS

CAUSATIVE CONDITIONS	OTHER CSF FINDINGS
BACTERIAL	
Mycobacterium tuberculosis	Usually < 500 white blood cells/μL, low glucose, high protein
Borrelia burgdorferi (Lyme disease)	Normal glucose, elevated protein
Treponema pallidum (secondary syphilitic meningitis, tertiary meningovascular syphilis)	Elevated protein; Venereal Disease Research Laboratory positive in CSF and serum
Brucella sp (uncommon)	Often low glucose; elevated protein
Tropheryma whippelii (rare)	Cells positive for periodic acid–Schiff on meningeal biopsy
Partially treated bacterial meningitis	Mixture of PMNs and lymphocytes, bacteria on Gram stain and culture
Parameningeal infections	Lymphocytes or mixed lymphocytic-PMN response, normal glucose
FUNGAL	
Cryptococcus neoformans	Low glucose, elevated protein, budding yeast on fungal wet mount, antigen detectable
Coccidioides immitis	Often low glucose, may have 10-20% eosinophils, elevated protein, presence of complement-fixing antibody
Histoplasma capsulatum	Low glucose; complement-fixing antibodies in CSF; antigen detectable in urine, CSF, serum
Blastomyces dermatitidis	Low glucose
Candida sp	Low glucose, may have PMN or lymphocyte predominance, fungal stain may be positive
Aspergillus sp	Lymphocytes or PMNs predominate
Sporothrix schenckii (sporotrichosis)	Low glucose; protein, 200-800 mg/dL
PROTOZOAL	
Toxoplasma gondii	Usually, picture is that of an encephalitis; often in patients with AIDS; pleocytosis is mild (<60 cells/μL) and protein is mildly elevated
Trypanosoma gambiense or *Trypanosoma rhodesiense*	Meningoencephalitis is stage II of disease, elevated protein and immunoglobulin M, trypanosomes on Giemsa-stained smear
VIRAL	
Mumps	Rarely, low glucose
Lymphocytic choriomeningitis	Rarely, low glucose
Echovirus (in patients with congenital agammaglobulinemia)	Occasionally, low glucose
HIV-1	Cell counts lower (10-20/μL) than in acute self-limited meningitis at clinical onset of HIV infection or may develop during course of AIDS

AIDS = acquired immunodeficiency syndrome; CSF = cerebrospinal fluid; HIV = human immunodeficiency virus; PMN = polymorphonuclear leukocyte.

TABLE 412-12 CAUSES OF CHRONIC (RECURRENT) MENINGITIS

Infections
Herpes simplex virus type 2
Leakage of contents from central nervous system tumors (chemical meningitis)
Epidermoid tumors
Craniopharyngiomas
Cholesteatomas
Drug hypersensitivity with repeated use of agent
Inflammatory processes
Behçet syndrome
Systemic lupus erythematosus
Mollaret meningitis
Vogt-Koyanagi-Harada syndrome

TABLE 412-13 CAUSES OF EOSINOPHILIC MENINGITIS*

CAUSATIVE CONDITIONS	SOURCE
PARASITIC DISEASE	
Angiostrongylus cantonensis	Ingestion of raw shellfish; Pacific
Taenia solium (cysticercosis)	Fecal-oral transmission of *T. solium* eggs
Gnathostoma spinigerum	Ingestion of raw fish; Japan, Southeast Asia
Baylisascaris procyonis	Accidental ingestion of *B. procyonis* eggs from raccoon feces
Trichinella spiralis (trichinosis)	Ingestion of poorly cooked pork
Schistosoma sp	Exposure of skin to fresh water; Africa, Middle East
Echinococcus granulosus	Contact with infected dogs passing eggs in feces
Toxoplasma gondii	Ingestion of meat containing cysts or food contaminated with oocysts from cat feces
Toxocara canis (visceral larva migrans)	Ingestion of infective eggs from dog feces
FUNGAL INFECTIONS	
Coccidioides immitis	Southwestern United States
NEOPLASTIC DISEASE	
Lymphoma, leukemia, metastatic carcinoma	
Hypereosinophilic syndrome (myeloproliferative disorder)	
INFLAMMATORY PROCESSES	
Sarcoidosis	
Drug hypersensitivity	
Presence of foreign body in the central nervous system	

*The percentage of eosinophils varies from as little as 6% to the majority of cells.

infection may suggest aseptic meningitis because of its sometimes indolent onset and, occasionally, an early predominantly lymphocytic response in young children. Fungal (e.g., *Cryptococcus neoformans, Coccidioides immitis, Histoplasma capsulatum*) meningitides are associated with a predominantly mononuclear response, sometimes with a small percentage of eosinophils, particularly in coccidioidal meningitis (Chapter 333). Patients with Rocky Mountain spotted fever (Chapter 327), an acute disease with a macular and petechial rash, may exhibit confusion. When examined, the CSF in approximately 20% of such patients shows a pleocytosis of 10 to 100 or more cells/μL, with either a neutrophilic or lymphocytic predominance. The clinical picture may suggest either enteroviral or meningococcal disease.

Epidemiologic factors are important in raising suspicion for nonviral aseptic meningitis. Leptospirosis (Chapter 323) may be suggested by a history of recent direct or indirect exposure to animals (e.g., dogs, rodents, dairy cattle) and their urine. Neurobrucellosis (Chapter 310) is suggested by the recent ingestion of unpasteurized cheese from the Mediterranean littoral, Middle East, or Mexico or by work as a veterinarian or in an abattoir. Specific endemic mycoses may be a consideration with residence in the southwestern United States (coccidioidomycosis; Chapter 333) and the Mississippi River valley (histoplasmosis; Chapter 332). The setting of immunosuppression by drugs or illness such as acquired immunodeficiency syndrome would raise the possibility of *C. neoformans* (Chapter 336) or *L. monocytogenes* (Chapter 293). Sexual promiscuity and the macular rash of secondary syphilis could suggest *T. pallidum* (Chapter 319) as the cause in a patient with lymphocytic meningitis.

Noninfectious Causes of Aseptic Meningitis

Noninfectious causes fall into four principal categories (see Table 412-10): drug hypersensitivity; systemic processes such as systemic lupus erythematosus and other collagen-vascular diseases; neoplastic disease, primary or metastatic, infiltrating the leptomeninges; and inflammatory processes primarily involving the CNS. Although a mononuclear cell predominance is found in the CSF in most noninfectious aseptic meningitides, there are several important exceptions. Drug hypersensitivity meningitis usually causes a neutrophilic response, although occasionally mononuclear cells or eosinophils predominate. In systemic lupus erythematosus (Chapter 266), the pleocytosis may be predominantly lymphocytic or neutrophilic (sometimes several thousand per cubic millimeter) with a normal CSF glucose level. Hypoglycorrhachia is a feature of few noninfectious aseptic meningitides and suggests malignant disease or sarcoidosis. Various drugs, most commonly the nonsteroidal anti-inflammatory drugs, have also been implicated in aseptic meningitis.

Chronic (Persistent) Meningitis

Chronic meningitis is defined by the clinical syndrome of headache, stiff neck, altered mental status, nausea and vomiting, evidence of myelopathy or radiculopathy with or without cranial nerve palsies (e.g., III, IV, VI, VII, VIII), and an inflammatory response in the CSF for 4 weeks or longer. Obstruction of CSF flow may produce hydrocephalus and papilledema.

Infectious Causes

Among the more common bacterial causes of chronic meningitis, *M. tuberculosis* (Chapter 324) is the most important to identify because if untreated, it is almost always fatal within 4 to 8 weeks (see Table 412-11).[13] Similarly, parameningeal infections (Chapter 413) must be recognized and treated promptly because surgery often is necessary to provide a specific bacteriologic diagnosis and prevent neurologic residua. Tuberculosis should be suspected in patients with a previous history of a tuberculous illness, a history of recent exposure, HIV infection or another immunosuppressed state, particularly the use of drugs and biologics that block TNF-α and that are often used to treat autoimmune diseases. Clinical manifestations include fever and night sweats, sixth cranial nerve palsies, stroke related to arteritis, or lesions on the chest radiograph. The purified protein derivative skin test may be negative in patients who are severely immunosuppressed or who have recently acquired or overwhelming disease. Acid-fast smear and culture of concentrated CSF can provide the diagnosis, and PCR can be very helpful despite its low sensitivity of only about 70%.[14] As a result, it may be difficult to establish an early diagnosis. When clinical and CSF findings suggest the diagnosis, treatment (Chapter 324) should be initiated while awaiting the culture results. Drug resistance and coinfection with HIV infection can be major impedements to adequate treatment. Rifampicin resistance can be easily detected by PCR, because almost all the mutations that confer rifampicin resistance are contained within a well-defined segment of the *rpoB* gene. Resistance to other drugs is less easily detected by these methods.

Parameningeal infections (Chapter 413) should be suspected when chronic meningitis with focal neurologic signs develops in the setting of chronic otitis media or sinusitis, pleuropulmonary infection, or right-to-left cardiopulmonary shunting. Contrast-enhanced CT or MRI of the head is important to delineate brain abscess, sinus infection, and epidural or subdural infections.

Meningitis may accompany the skin, mucous membrane, and lymph node features of secondary syphilis (Chapter 319), or it may occur alone. Individual cranial nerves (II to VII) may be involved; visual abnormalities, hearing loss, and facial palsy are most frequent. The fluorescent treponema antibody absorption test or microhemagglutination *T. pallidum* serologic studies are helpful in distinguishing the process from biologic false-positive Venereal Disease Research Laboratory (or rapid plasma reagent) results in serum.

Lyme disease meningitis (Chapter 321) should be suspected on the basis of epidemiologic grounds (geographic location, season, tick exposure) and associated clinical features (erythema migrans rash, Bell palsy, radiculopathy). The diagnosis is made by enzyme-linked immunosorbent assay with Western blot confirmation.

A variety of fungal infections can cause a chronic meningitis. Cryptococcal meningitis (Chapter 336) is common in immunosuppressed individuals and can be diagnosed by detection of cryptococcal antigen in the CSF. Histoplasmosis (Chapter 332) should be suspected in endemic regions. Aspergillosis (Chapter 339) is angiocentric and can cause associated cerebral infarcts. Mucormycocis (Chaper 340) is common in patients with poorly controlled diabetes mellitus. Flucytosine is superior to fluconazole when used with amphotericin B for treatment of cryptococcal meningitis (Chapter 336).■

Noninfectious Causes

Noninfectious causes of meningitis include malignant disease, chemical meningitis, and primary inflammatory conditions (Table 412-14). Malignant disease may be diagnosed by cytologic examination of large volumes of CSF. Contrast-enhanced MRI may disclose thickening of the meninges and nerve roots, but meningeal biopsy may be required for diagnosis. Chemical meningitis from previous subarachnoid injection may persist, with xanthochromia noted in CSF; meningeal inflammation may be identified on contrast-enhanced CT or MRI.

Meningeal or CNS sarcoid (Chapter 95) may be isolated or occur with other organ involvement, such as pulmonary granulomas, lymphadenopathy, or myopathy. Neurologic findings can include diabetes insipidus and cranial

TABLE 412-14 NONINFECTIOUS CAUSES OF CHRONIC (PERSISTENT) LYMPHOCYTIC MENINGITIS

CAUSATIVE CONDITIONS	OTHER CSF FINDINGS
NEOPLASMS	
Metastatic: Lung, breast, stomach, pancreas, lymphoma, melanoma, leukemia	Low glucose; elevated protein, cytologic examination; polarizing microscopy; clonal lymphocyte markers
Central nervous system: Meningeal gliomatosis, meningeal sarcoma, cerebral dysgerminoma; epidermoid tumors/cysts	
CHEMICAL INFLAMMATION	
Endogenous: Epidermoid tumor, craniopharyngioma	Low glucose, elevated protein
Exogenous: Recent injection into the subarachnoid space	Low glucose, elevated protein
PRIMARY INFLAMMATORY PROCESSES	
Central nervous system sarcoid	Often low glucose, elevated protein, elevated angiotensin-converting enzyme levels in CSF (and serum)
Granulomatosis with polyangiitis (formerly Wegener)	Elevated protein
Behçet syndrome	Elevated protein
Isolated granulomatous angiitis of the central nervous system	Elevated protein
Systemic lupus erythematosus	Elevated protein
?Chronic idiopathic benign meningitis	Elevated protein

CSF = cerebrospinal fluid.

TABLE 412-15 CAUSES OF CHRONIC (PERSISTENT) MENINGITIS WITH NEUTROPHIL PREDOMINANCE

UNCOMMON	OTHER CSF FINDINGS
BACTERIAL	
Nocardia asteroides	Low glucose, markedly elevated protein, culture positive
Actinomyces israelii	Low glucose, elevated protein, anaerobic culture positive
Arachnia propionica	Low glucose, elevated protein, anaerobic culture positive
FUNGAL	
Candida sp	Low glucose, elevated protein, culture positive
Aspergillus sp	Low glucose, elevated protein, enzyme immunoassay or enzyme-linked immunosorbent assay for *Aspergillus* galactomannan
Zygomycetes	Low glucose, elevated protein
Dematiaceous fungi	Low glucose, protein may be markedly elevated
NONINFECTIOUS	
Systemic lupus erythematosus	Low glucose, elevated protein
Chemical meningitis	Low glucose, protein may be markedly elevated
VERY RARE	
Bacterial	
Brucella sp	Low glucose, elevated protein
Mycobacterium tuberculosis	Low glucose, elevated protein, polymerase chain reaction positive for *M. tuberculosis* DNA
Fungal	
Pseudoallescheria boydii	Low glucose, protein may be markedly elevated
Coccidioides immitis	Low glucose, elevated protein, presence of complement-fixing antibody
Blastomyces dermatitidis	Low glucose, protein elevated, antigen detection possible in CSF and urine
Histoplasma capsulatum	Low glucose; protein mildly elevated; complement-fixing antibodies in CSF; antigen detectable in CSF, urine, serum

CSF = cerebrospinal fluid.

nerve palsies. Granulomatosis with polyangiitis (Chapter 270) may produce meningeal inflammation and cranial nerve palsies, often in association with air sinus disease. The diagnosis is suggested by lesions on the chest radiograph, microscopic hematuria, skin lesions, peripheral neuropathy, and serum antineutrophil cytoplasmic antibodies. Aseptic meningitis associated with systemic lupus erythematosus (Chapter 266) may be accompanied by other neurologic manifestations (seizures, encephalopathy, stroke, transverse myelopathy), systemic manifestations (rash, arthritis), and antinuclear and anti-DNA antibodies.

Chronic (Intermittent) Meningitis

In chronic intermittent meningitis, all clinical and CSF abnormalities resolve completely between episodes without antimicrobial therapy (see Table 412-11). Uncommonly, a patient may have several episodes resulting from different viral agents. The major causes of recurrent aseptic meningitis are infections (almost always viral and resulting from HSV-2), endogenous chemical meningitis, drug hypersensitivity (including the use of intravenous immunoglobulins) with meningitis following each use, and inflammatory and autoimmune diseases.

In HSV-2 recurrent meningitis, lymphocytes predominate, with the cell numbers being approximately 40% higher in the initial episode than in recurrences. Leakage of material from intracranial epidermoid cysts produces 1000 to 5000 cells/μL (≈80% polymorphonuclear leukocytes) initially, with a subsequent mononuclear cell predominance. Occasionally, polarizing microscopy may demonstrate keratin and cholesterol crystals in the CSF of patients with endogenous chemical meningitis. In Behçet syndrome (Chapter 270), the CSF may have predominantly mononuclear cells or polymorphonuclear leukocytes. Mollaret meningitis, a syndrome of benign recurrent meningitis usually caused by HSV-2, is initially associated with neutrophils and monocytes in the CSF without hypoglycorrhachia but subsequently transitions to a predominantly lymphocytic pleocytosis. However, prolonged treatment with valcyclovir 1 g/day does not prevent recurrences of HSV-2–associated meningitis. Vogt-Koyanagi-Harada syndrome, a rare uveomeningoencephalitis, consists of recurrent meningitis/meningoencephalitis and anterior or posterior uveitis, followed by vitiligo, poliosis, alopecia, and dysacousia; the CSF cellular response is mononuclear, and an autoimmune origin, directed against a melanocyte antigen, has been suggested.

Chronic Meningitis with Predominantly Neutrophilic Pleocytosis

Chronic persistent neutrophilic meningitis (Table 412-15) is defined by the following combination: (1) clinical features consistent with meningitis; (2) initial CSF examination showing greater than 50% neutrophils, hypoglycorrhachia, and elevated protein concentration; (3) antimicrobial therapy that would be appropriate for the usual causes of bacterial meningitis; (4) negative smears and cultures for bacteria on the initial CSF specimen; and (5) repeated CSF examination 7 days or more after initial analysis showing 50% or greater neutrophils, hypoglycorrhachia, and elevated protein concentration.

Among the bacterial causes (see Table 412-15) are organisms (*Actinomyces israelii* and *Arachnia propionica* [Chapter 329]) that can be isolated by culture only under anaerobic conditions. Coexisting pulmonary lesions may suggest *Nocardia* (Chapter 330) or *M. tuberculosis* (Chapter 324) as the cause, although the initial polymorphonuclear pleocytosis present in some cases uncommonly persists much beyond a week before changing to a lymphocytic predominance. *Brucella* (Chapter 310) and endemic invasive mycotic infections would be suggested by epidemiologic considerations. Other fungal causes may be diagnosed, particularly in immunocompromised patients, by antigen testing with enzyme-linked immunosorbent assay (*Aspergillus* sp galactomannan; Chapter 339), or meningeal biopsy may be required.

Occasionally, exogenous chemical meningitis secondary to intrathecal injection of antimicrobials, chemotherapeutic agents, or contrast media may produce persisting pleocytosis and hypoglycorrhachia resulting from sclerosing arachnoiditis well after the inciting medication has been withdrawn. Systemic lupus erythematosus (Chapter 266) can produce a variety of meningitides, including acute lymphocytic or neutrophilic aseptic meningitis, as well as chronic persistent lymphocytic or neutrophilic CSF responses.

Eosinophilic Meningitis

The presence of 5% or greater eosinophils in CSF is uncommon and suggests parasitic disease, certain fungal infections such as coccidioidal or candidal meningitis, neoplastic diseases, or a few inflammatory processes (see Table 412-13).[15] In most cases, eosinophils are mixed with lymphocytes,

which predominate; the highest percentage of eosinophils is seen with meningitis caused by migrating larvae of the raccoon ascarid *Baylisascaris procyonis* (Chapter 357) and the rat lung worm *Angiostrongylus cantonensis* (Chapter 357). In fungal meningitides, particularly those resulting from *C. immitis* (Chapter 333), the CSF response is primarily mononuclear with 6 to 20% eosinophils; hypoglycorrhachia may be a feature of *C. immitis* and *Candida* meningitis (Chapter 338) and of neoplastic processes and sarcoid.

Most patients with eosinophilic meningitis, except those with cases resulting from trichinosis (Chapter 357) or drug hypersensitivity, have prolonged symptoms suggesting chronic meningitis. Most patients with meningitis of parasitic or neoplastic origin have evidence of cerebral involvement as well.

Grade A References

A1. Molyneux E, Nizami SQ, Saha S, et al. 5 versus 10 days of treatment with ceftriaxone for bacterial meningitis in children: a double-blind randomised equivalence study. *Lancet*. 2011;377:1837-1845.

A2. van de Beek D, Farrar JJ, de Gans J, et al. Adjunctive dexamethasone in bacterial meningitis: a meta-analysis of individual patient data. *Lancet Neurol*. 2010;9:254-263.

A3. Nguyen TH, Tran TH, Thwaites G, et al. Dexamethasone in Vietnamese adolescents and adults with bacterial meningitis. *N Engl J Med*. 2007;357:2431-2440.

A4. Scarborough M, Gordon SB, Whitty CJ, et al. Corticosteroids for bacterial meningitis in adults in sub-Saharan Africa. *N Engl J Med*. 2007;357:2441-2450.

A5. Brouwer MC, McIntyre P, Prasad K, et al. Corticosteroids for acute bacterial meningitis. *Cochrane Database Syst Rev*. 2013;6:CD004405.

A6. Fritz D, Brouwer MC, van de Beek D. Dexamethasone and long-term survival in bacterial meningitis. *Neurology*. 2012;79:2177-2179.

A7. Mourvillier B, Tubach F, van de Beek D, et al. Induced hypothermia in severe bacterial meningitis: a randomized clinical trial. *JAMA*. 2013;310:2174-2183.

A8. Aurelius E, Franzen-Rohl E, Glimaker M, et al. Long-term valacyclovir suppressive treatment after herpes simplex virus type 2 meningitis: a double-blind, randomized controlled trial. *Clin Infect Dis*. 2012;54:1304-1313.

A9. Day JN, Chau TT, Wolbers M, et al. Combination antifungal therapy for cryptococcal meningitis. *N Engl J Med*. 2013;368:1291-1302.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Martin NG, Sadarangani M, Pollard AJ, et al. Hospital admission rates for meningitis and septicaemia caused by *Haemophilus influenzae, Neisseria meningitidis,* and *Streptococcus pneumoniae* in children in England over five decades: a population-based observational study. *Lancet Infect Dis*. 2014;14:397-405.
2. Thigpen MC, Whitney CG, Messonnier NE, et al. Bacterial meningitis in the United States, 1998-2007. *N Engl J Med*. 2011;364:2016-2025.
3. Bardak-Ozcem S, Sipahi OR. An updated approach to healthcare-associated meningitis. *Expert Rev Anti Infect Ther*. 2014;12:333-342.
4. van de Beek D, Drake JM, Tunkel AR. Nosocomial bacterial meningitis. *N Engl J Med*. 2010; 362:146-154.
5. Brouwer MC, Thwaites GE, Tunkel AR, et al. Dilemmas in the diagnosis of acute community-acquired bacterial meningitis. *Lancet*. 2012;380:1684-1692.
6. Kumar A, Debata PK, Ranjan A, et al. The role and reliability of rapid bedside diagnostic test in early diagnosis and treatment of bacterial meningitis. *Indian J Pediatr*. 2015;82:311-314.
7. Sakushima K, Hayashino Y, Kawaguchi T, et al. Diagnostic accuracy of cerebrospinal fluid lactate for differentiating bacterial meningitis from aseptic meningitis: a meta-analysis. *J Infect*. 2011; 62:255-262.
8. Mekitarian Filho E, Horita SM, Gilio AE, et al. Cerebrospinal fluid lactate level as a diagnostic biomarker for bacterial meningitis in children. *Int J Emerg Med*. 2014;7:14.
9. Brouwer MC, Thwaites GE, Tunkel AR, et al. Dilemmas in the diagnosis of acute community-acquired bacterial meningitis. *Lancet*. 2012;380:1684-1692.
10. van de Beek D, Brouwer MC, Thwaites GE, et al. Advances in treatment of bacterial meningitis. *Lancet*. 2012;380:1693-1702.
11. Roed C, Omland LH, Skinhoj P, et al. Educational achievement and economic self-sufficiency in adults after childhood bacterial meningitis. *JAMA*. 2013;309:1714-1721.
12. Steiner I, Schmutzhard E, Sellner J, et al. EFNS-ENS guidelines for the use of PCR technology for the diagnosis of infections of the nervous system. *Eur J Neurol*. 2012;19:1278-1291.
13. Thwaites GE, van Toorn R, Schoeman J. Tuberculous meningitis: more questions, still too few answers. *Lancet Neurol*. 2013;12:999-1010.
14. Solomons RS, van Elsland SL, Visser DH, et al. Commercial nucleic acid amplification tests in tuberculous meningitis-a meta-analysis. *Diagn Microbiol Infect Dis*. 2014;78:398-403.
15. Sawanyawisuth K, Chotmongkol V. Eosinophilic meningitis. *Handb Clin Neurol*. 2013;114: 207-215.

REVIEW QUESTIONS

1. A 16-year-old girl develops fever, photophobia, stiff neck, hemorrhagic lesions in the skin, and drop in blood pressure. Symptoms first started 6 hours ago. Appropriate management would be:

A. Investigate with an MRI brain scan and CSF analysis, and then treat for viral or bacterial meningitis depending on results.
B. Maintain blood pressure, then initiate investigations with MRI of brain, blood culture, and CSF analyses.
C. Precious time may be lost in obtaining an MRI scan; once blood pressure is maintained, obtain CSF analyses to make diagnosis.
D. Obtain blood cultures, treat with intravenous antibiotics, maintain blood pressure, and then initate investigations

Answer: D The patient most likely has meningococcal meningitis, which is fulminant and can be fatal if not treated immediately. Enteroviruses can cause a skin rash but not hemorrhagic lesions. This clinical presentation is also termed *Waterhouse-Friderichsen syndrome,* and the drop in blood pressure can be due to bleeding in the adrenal glands.

2. A 76-year-old man with neutropenia following treatment for colon cancer develops dysphagia, ataxia, and left-sided weakness. MRI of the brain shows a contrast-enhancing lesion in the brain stem and extending into the cerebellum, with enhancement of the meninges in the posterior fossa. The most likely diagnosis is:

A. Metastatic colon cancer.
B. Metastatic cancer from another primary source such as lung.
C. Brain abscess and meningitis due to *Mycobacterium tuberculosis.*
D. Viral encephalitis.
E. *Listeria* brain abscess and meningitis.

Answer: E Colon cancer rarely goes to the brain. Some metastatic cancers seed the meninges and cause carcinomatous meningitis at the base of the skull; these metastases can result in multiple cranial nerve palsies, but they do not invade the brain stem. Neutropenia predisposes a person to bacterial infections, and *Listeria* commonly invades the brain stem and posterior fossa and can occur in the absence of intestinal symptoms. The term *rhombencephalitis* is often used to describe it. Animal models show that the bacteria can travel from the intestine to the brain stem via the vagus nerve.

3. Point(s) to consider for the use of corticosteroids as an adjunctive therapy for bacterial meningitis is/are:

A. It should be initiated before or simultaneously with the use of antibiotics.
B. Its use has been associated with a decrease in morbidity and mortality from bacterial meningitis if used early in the course of the illness.
C. It acts by decreasing injury that otherwise would be caused by inflammatory mediators.
D. All of the above.
E. A and B only.

Answer: D Two landmark studies provide strong evidence in support of the use steroids and an adjunctive treatement for bacterial meningitis.

4. Recurrent aseptic meningitis with history of genital lesions is usually due to:

A. Herpes simplex virus type 2.
B. Syphilis.
C. Behçet disease.
D. All of the above.

Answer: D Herpes simplex virus type-2 can be diagnosed by PCR on the CSF, syphilis by testing for VDRL in the CSF, and Behçet disease is usually associated with orogenital ulcers and evidence of demylination on the MRI scan of the brain.

413 BRAIN ABSCESS AND PARAMENINGEAL INFECTIONS

AVINDRA NATH AND JOSEPH BERGER

Brain abscess affects the brain's parenchyma directly, whereas parameningeal infections produce suppuration in potential spaces covering the brain and spinal cord (epidural abscess and subdural empyema) or produce occlusion of the contiguous venous sinuses and cerebral veins (cerebral venous sinus thrombosis).

BRAIN ABSCESS

EPIDEMIOLOGY

The frequency of various causes of brain abscess (Table 413-1) in the population has been difficult to ascertain because of wide variations among case series, in part as a result of referral patterns. In addition, children with brain abscesses often have cyanotic congenital heart disease or otogenic infection. Cryptogenic abscesses account for a greater percentage of cases in more recent series, perhaps related to the presence of a patent foramen ovale. On average, 90% of brain abscesses occur as a consequence of a focus of suppuration elsewhere in the body, with the remainder due to introduction of the infection from head wounds or neurosurgical procedures. Males predominate in virtually all series of brain abscess. Terminally ill patients in whom medical care is withdrawn may be found to have abscesses at autopsy, but these abscesses are of little clinical importance.

PATHOBIOLOGY

Brain abscesses are collections of purulent material (neutrophils and necrotic tissue) caused by infection with a variety of bacterial, fungal, and parasitic organisms. Infection arising from other sites typically seeds the brain hematogenously. When contiguous to the brain, infection enters the brain by direct extension or by traveling along veins with associated thrombophlebitis of pial veins and sinuses. Within the brain, the infection begins as a cerebritis with perivascular infiltrates and infiltration of neutrophils into the brain parenchyma. With time, the developing abscess is characterized by a purulent exudate that includes necrotic brain tissue as well as viable and necrotic neutrophils. Granulation tissue develops at the interface between necrotic and viable tissue, and eventually, the abscess is walled off by a fibrous capsule. Formation of the capsule depends on the virulence of the organism and the immune status of the individual. More virulent organisms are associated with larger lesions, more necrosis, earlier ependymitis, and a greater degree of inflammation outside the collagen capsule.

CLINICAL MANIFESTATIONS

The clinical picture reflects a triad of the infectious nature of the lesion, focal brain involvement, and an increasing intracranial mass effect (Table 413-2).[1] One or two elements may be absent in a given case, particularly early in the course. Among infectious symptoms, fever is present at onset or early in the course in only about 60% of cases. Neck stiffness is an infrequent complaint, and meningeal signs are elicited in about 30% of cases. The absence of classical signs may delay diagnosis.[2]

Focal neurologic deficits depend on the site and size of the lesion, which in turn will be determined by the causal or predisposing condition. In some patients, seizures precede the diagnosis. The early deficits in patients with temporal lobe lesions, which are typically caused by spread of an otogenic

TABLE 413-1 CONDITIONS THAT PREDISPOSE TO THE DEVELOPMENT OF BRAIN ABSCESS

Otogenic
Otitis media
Mastoiditis
Dental
Cardiac
Cyanotic heart disease
Tetralogy of Fallot
Patent foramen ovale
Infective endocarditis
Pulmonary
Pulmonary arteriovenous fistula
Lung infection
Lung abscess
Bronchiectasis
Esophageal strictures
Cerebral infarcts and tumors
Penetrating and nonpenetrating head injury
Postoperative neurosurgical procedure (trauma and nontrauma related)
Dermal sinus tracts
Sepsis
Immunosuppression
Unknown mechanism

TABLE 413-2 BRAIN ABSCESS: INITIAL FEATURES IN 123 CASES

Headache	55%
Disturbed consciousness	48%
Fever	58%
Nuchal rigidity	29%
Nausea, vomiting	32%
Seizures	19%
Visual disturbance	15%
Dysarthria	20%
Hemiparesis	48%
Sepsis	17%

abscess, are contralateral homonymous superior quadrantic visual field defects and, if in the dominant hemisphere, aphasia. Motor deficits eventually occur in 40 to 50% of supratentorial abscesses. Cerebellar abscesses, which are often caused by aural-mastoid infections, are characterized by ipsilateral limb ataxia; there may also be abnormal head positioning (forward and away from the side of the lesion) and nystagmus that is slow and coarse on gaze to the side of the abscess and rapid in the opposite direction. Patients with multiple brain abscesses may have multifocal signs or encephalopathy. Patients with *Toxoplasma* species (Chapter 349) brain abscesses often have movement disorders because these abscesses frequently localize to the basal ganglia. In fact, nearly all patients with human immunodeficiency virus (HIV) infection in whom hemiballism or hemichorea is present have *Toxoplasma* species brain abscesses.

Headache is an important initial symptom in 80 to 90% of patients with bacterial abscess but is less frequent (≈20%) in patients with fungal abscesses. Symptoms of increased intracranial pressure, such as nausea, depressed level of consciousness, and papilledema, occur less often. The development of headache in a patient with a known chronic anaerobic infection, such as aural-mastoid, paranasal sinus, or pulmonary suppuration, suggests the possibility of brain abscess. Similarly, the development of headache in a child with cyanotic congenital heart disease is often related to a brain abscess. Tetralogy of Fallot (Chapter 69) is the most common congenital heart anomaly associated with brain abscess.

DIAGNOSIS

Examination of the cranium, ears, paranasal sinuses, oral cavity, heart, and lungs may provide important clues to the etiology, as may overt signs of infection at other sites. Cultures of blood and sputum may identify the organism and its antimicrobial sensitivity. In patients with signs of raised intracranial pressure, lumbar puncture may be contraindicated because of the risk of herniation.

Magnetic resonance imaging (MRI) can detect early changes such as brain edema and is preferable to computed tomography (CT).[3] In the early cerebritis stage, T2-weighted MRI shows abnormally high signal intensity corresponding to low signal intensity on the T1-weighted images. The fluid-attenuated inversion recovery (FLAIR) sequence provides superior visualization of brain edema. On T1-weighted images, the area of cerebritis that is seen initially as a low-signal-intensity, ill-defined area later progresses to a central cavity with slightly higher signal intensity than cerebrospinal fluid (CSF), surrounded by edema that is slightly hypointense in comparison to brain parenchyma. Later stages of infection show central necrosis and formation of a rim of slightly high signal intensity on T1-weighted images (Fig. 413-1). With gadolinium administration, there is a ring-enhancing lesion. Diffusion-weighted imaging helps differentiate abscesses from brain tumors; an abscess cavity demonstrates high signal with decreased apparent diffusion coefficient values, whereas necrotic tumor cavities demonstrate the opposite.

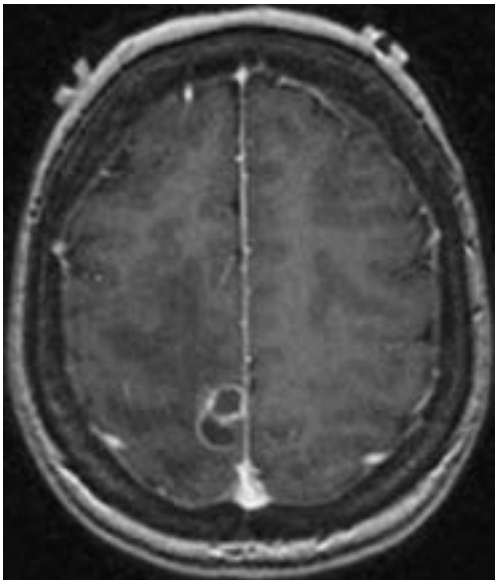

FIGURE 413-1. Brain abscess. Magnetic resonance imaging with gadolinium shows a multiloculated ring-enhancing lesion caused by *Nocardia* species infection.

Surgical aspiration or excision of the lesion may be necessary to establish a microbial diagnosis. Gram stain and culture from abscess fluid, with proper handling, have high yield, with or without previous antibiotic therapy. If immediate surgery is planned, antibiotics can be deferred until culture material has been acquired. Multiplex polymerase chain reaction testing is being developed for rapid identification of bacterial organisms and detection of antibiotic resistance genes.

TREATMENT

Rx

Brain abscess requires urgent intervention.[4] Because of the risk for cerebral herniation with large lesions, treatment of cerebral edema (intravenous (IV) dexamethasone, 16 to 24 mg/day in four divided doses) may be needed even while initiating surgical intervention. Corticosteroids often decrease edema within 8 hours but may retard the formation of a capsule around the brain abscess, suppress the immune response to the infection, and decrease penetration of antibiotics. Hence, they should be used for short periods, usually only until surgical decompression by needle drainage or surgical removal is possible. Empirical antibiotic therapy (Table 413-3) is recommended prior to surgery, based on the likely source of infection.

Successful antibiotic management of brain abscess is based on knowledge of proved or suspected pathogens as well as familiarity with a drug's spectrum of activity and penetration into the central nervous system. When surgery cannot be performed, empirical antibiotic therapy must be initiated. A trial of nonsurgical treatment may be considered in patients with (1) small lesion size, (2) an already identified pathogen, (3) no symptoms or signs of increased intracranial pressure requiring neurosurgical intervention, (4) a deep or inaccessible lesion, (5) multiple abscesses, (6) a contraindication to surgery (e.g., a bleeding diathesis), (7) a short duration of symptoms, which suggests that the lesion is in the cerebritis stage, and (8) availability of monitoring with MRI.[5]

In patients who are suspected of having a brain stem abscess, the possibility of listerial infection (Chapter 293) should be considered (Fig. 413-2), even in the absence of a clear immunodeficiency. Empirical parenteral antibiotics to cover *Listeria* species should be started (Chapter 293).

Brain abscesses caused by *Toxoplasma* species (Chapter 349) usually occur in immunocompromised patients (e.g., patients with HIV infection), are not accompanied by capsule formation, and hence respond well to antibiotic therapy alone. As a result, patients with acquired immunodeficiency syndrome and suspected cerebral toxoplasmosis (Chapter 349) should receive antimicrobial therapy initially.

TABLE 413-3 COMMON PATHOGENS AND EMPIRICAL THERAPY FOR BRAIN ABSCESS

PREDISPOSING CONDITION	COMMON PATHOGENS	ANTIMICROBIAL AGENTS*
Dental abscess	Streptococci, *Bacteroides fragilis*	Penicillin + metronidazole
Chronic otitis	*Bacteroides fragilis; Pseudomonas, Proteus, Klebsiella* species	Cefotaxime or ceftriaxone + metronidazole; ceftazidime or cefepime for *Pseudomonas* species
Sinusitis	Streptococci; *Haemophilus, Staphylococcus* species	Cefotaxime, ceftriaxone, or nafcillin + metronidazole
Penetrating trauma or postsurgical	*Staphylococcus, Pseudomonas, Enterobacter* species; streptococci	Nafcillin or vancomycin + ceftriaxone or cefotaxime + metronidazole
Bacterial endocarditis or drug use	Mixed flora, streptococci, *Staphylococcus* species	Nafcillin or vancomycin + ceftriaxone or cefotaxime + metronidazole
Congenital heart disease	Streptococci	Cefotaxime or ceftriaxone
Pulmonary infection	*Nocardia* species, *Bacteroides fragilis*, streptococci, mixed flora	Penicillin + metronidazole + trimethoprim-sulfamethoxazole
HIV infection	*Toxoplasma gondii*	Pyrimethamine + sulfadiazine + folinic acid

*See Table 287-4 in Chapter 287 for dosing schedules.

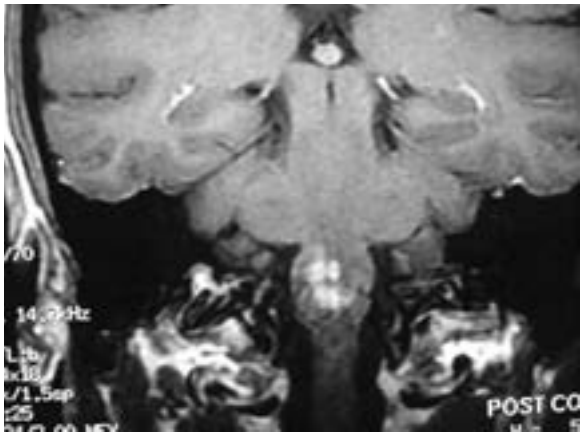

FIGURE 413-2. Brain stem abscess. Magnetic resonance imaging with gadolinium shows an enhancing lesion in the brain stem caused by *Listeria* species infection.

TABLE 413-4 INITIAL CHARACTERISTICS OF 915 PATIENTS WITH SPINAL EPIDURAL ABSCESS

STAGE 1	
Back pain	71%
Fever	66%
STAGE 2	
Radicular pain	20%
STAGE 3	
Muscle weakness	26%
Sphincter incontinence	24%
Sensory deficits	13%
STAGE 4	
Paralysis	31%
Quadriplegia	3%

From Reihsaus E, Waldbaur H, Seeling W. Spinal epidural abscess: a meta-analysis of 915 patients. *Neurosurg Rev.* 2000;23:175-204.

PROGNOSIS

Before the CT scan era, the mortality of brain abscesses ranged from 40 to 60%, and even with the drastic reduction in mortality in the era of modern neuroimaging, the mortality rate remains about 10%.[6] About 70% of patients recover fully. In post-transplantation patients and those with deep hemispheric or brain stem abscesses, mortality rates may exceed 80%. Other factors associated with a poor prognosis include extremes of age, multiple abscesses, and diagnostic delay in the absence of systemic signs of infection. Impaired level of consciousness is a poor prognostic sign even with early hospitalization and rapid diagnosis. Anaerobic and gram-negative organisms and culture-negative cases also have a poor prognosis. Seizures (Chapter 403) develop in up to 50% of patients, sometimes after latencies as long as 5 years.

SPINAL EPIDURAL ABSCESS

DEFINITION

Infection within the epidural space around the spinal cord is an uncommon but often readily treatable potential cause of paralysis and death. The epidural space surrounds the dural sac and is limited by the posterior longitudinal ligament anteriorly, the ligamenta flava and the periosteum of the laminae posteriorly, and the pedicles of the spinal column and the intervertebral foramina containing their neural elements laterally. The space communicates with the paravertebral space through the intervertebral foramina. Superiorly, the space is closed at the foramen magnum. Caudally, the space is closed by the sacrococcygeal ligament. The epidural space contains loose areolar connective tissue, semiliquid fat, lymphatics, arteries, an extensive plexus of veins, and the spinal nerve roots.

EPIDEMIOLOGY

Spinal epidural abscesses can result from hematogenous spread of infection; risk factors include IV drug use, organ transplantation, chronic steroid use, malignancy, and diabetes. Local infection after acupuncture for back pain or epidural analgesia can also cause epidural abscesses. Cutaneous sites of infection are the most common remote sources, especially in IV drug users. Abdominal, respiratory tract, and urinary sources are also common. Osteomyelitis may be a cause of either direct extension or hematogenous spread, particularly when associated with sepsis. Contiguous spread may occur from epidurally placed catheters, psoas abscesses, decubitus ulceration, perinephric and retropharyngeal abscesses, or surgical sites. Minor back trauma has been implicated in causing a paraspinal hematoma, which may subsequently be seeded hematogenously. *Staphylococcus aureus* is the most common organism isolated from spinal epidural abscesses. In 2012, an outbreak of fungal paraspinal infections was attributed to epidural injections of contaminated methylprednisolone.[7]

PATHOBIOLOGY

Because the dura mater around the cord is adherent to the vertebral column anteriorly, more epidural abscesses lie posteriorly, and because no anatomic barriers separate the spinal segments in the posterior epidural space, such abscesses usually extend over several vertebral segments. Spinal cord dysfunction probably reflects toxic processes secondary to inflammation, as well as venous thrombosis, thrombophlebitis, ischemia secondary to compression of the spinal arteries, and edema.

CLINICAL MANIFESTATIONS

The presence of a risk factor (>80% of patients) in the setting of neurologic deficits or back or radicular pain should suggest a spinal epidural abscess. The clinical manifestations can be divided into four stages (Table 413-4). Back pain (71%), fever (66%), tenderness of the spine with focal percussion (17%), spinal irritation (20%), and headache (3%) are common.[8] Radicular pain can be mistaken for sciatica, a visceral abdominal process, chest wall pain, or cervical disc disease. Clinical signs are often substantially greater than would be predicted from the anatomic extent of pus or granulation tissue.

Unfortunately, the diagnosis often is missed initially. If the condition goes unrecognized at an early stage, the symptoms can evolve over a period of hours to days to paralysis below the spinal level of infection.

DIAGNOSIS

The differential diagnosis includes compressive and inflammatory processes involving the spinal cord: transverse myelitis (Chapter 411), herniation of an intervertebral disc (Chapter 400), epidural hemorrhage (Chapter 400), or metastatic tumor (Chapter 189), none of which are associated with evidence of systemic infection. Blood leukocytosis may not be present, but the sedimentation rate is often elevated. Other infectious processes that may produce back or neck pain or tenderness must be excluded: bacterial meningitis (Chapter 412), perinephric abscess, disc space infection, and bacterial endocarditis (Chapter 76).

Lumbar puncture should be avoided in patients suspected of having a spinal epidural abscess, for fear of spreading the infection to the subarachnoid space and causing meningitis. Gadolinium-enhanced MRI (Fig. 413-3) is the method of choice for diagnosis,[9] but MRI findings in patients undergoing epidural analgesia can resemble those of epidural spinal abscess even when no infection is present.

TREATMENT

Rx

Patients with a progressing neurologic deficit should undergo urgent surgical drainage; CT-guided aspiration may be useful, and antibiotics plus percutaneously guided needle aspiration may be as therapeutically effective as antibiotics plus surgery. Unless culture results and sensitivities dictate otherwise, empirical therapy should cover *S. aureus* (nafcillin, 2 g every 6 hours; vancomycin, 1 g every 12 hours for methicillin-resistant strains). Additional gram-negative coverage with a third-generation cephalosporin (e.g., cefotaxime, 2 g every 6 hours, or ceftriaxone, 2 g every 12 hours) or a quinolone (e.g., ciprofloxacin, 400 mg every 12 hours) should be considered for severe disease. Rifampin (300 mg every 12 hours) may be added because of its ability to penetrate the abscess cavity. IV therapy should be continued for 3 to 4 weeks except in the presence of osteomyelitis (6 to 8 weeks).

PROGNOSIS

The mortality rate associated with spinal epidural abscess is about 15%. Approximately 50% of survivors have residual neurologic deficits. More severe preoperative neurologic deficits and deficits of longer duration are associated with a worse prognosis. In general, patients who develop paralysis that persists for longer than 36 hours do not recover function.

SUBDURAL EMPYEMA

Subdural empyema is an infection in the space between the dura and the arachnoid. It usually results from infected paranasal sinuses and rarely from infected mastoid sinuses by extension of thrombophlebitis from the sinuses into the subdural space. The infection is most commonly unilateral because bilateral spread is prevented by the falx. The empyema may evolve to cause cortical vein thrombosis, cerebral abscesses, or purulent meningitis.

CLINICAL FEATURES AND DIAGNOSIS

The most common symptoms are headache, fever, a neurologic deficit, and a stiff neck. However, subdural empyema may progress and cause signs of raised intracranial pressure, such as vomiting, altered level of consciousness, seizures, and papilledema. A high degree of suspicion is needed to establish the diagnosis early in the course of the illness. In patients with sinusitis (Chapter 426), the symptoms of subdural empyema may be incorrectly attributed to the sinusitis.

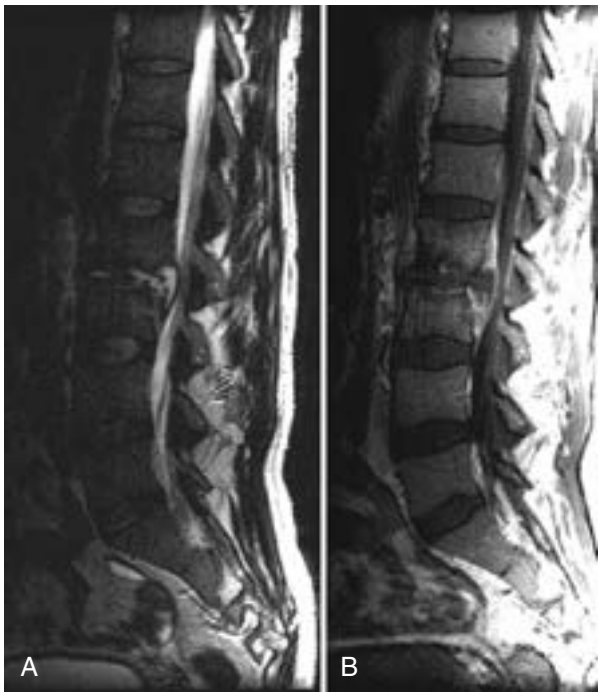

FIGURE 413-3. Spinal epidural abscess. A and B, Magnetic resonance images of the lumbosacral spine show a lesion in the epidural space compressing the thecal sac.

MRI with gadolinium enhancement and diffusion-weighted images is particularly useful in visualizing the subdural infection as a crescent-shaped mass with an enhancing rim over the cerebral convexities and below the inner table of the skull (Fig. 413-4). CSF evaluation is useful only if there is accompanying meningitis. In a patient with signs of raised intracranial pressure, lumbar puncture should be avoided because of the risk for herniation.

TREATMENT

Rx

Surgical drainage of the empyema is mandatory. IV antibiotic therapy is also necessary and is based on the organisms isolated at the time of craniotomy.

PROGNOSIS

Mortality rates in most series are about 25%, with severe neurologic sequelae remaining in 20% of survivors. Accompanying venous sinus thrombosis or brain abscess carries a poor prognosis.

VENOUS SINUS THROMBOSIS SECONDARY TO INFECTION

The venous sinus system (Fig. 413-5) lacks valves, thereby permitting retrograde propagation of clots or infections that emanate from structures located in the central portion of the face or the middle ear.[10]

Septic Cavernous Sinus Thrombosis

DEFINITION

The cavernous sinuses are the most caudal dural venous chambers at the base of the skull. The paired structures lie on either side of the pituitary fossa immediately above the midline sphenoid sinus. The cavernous sinus encloses the "cavernous portion" of the internal carotid artery as well as the third, fourth, and sixth cranial nerves en route to the apex of the orbit.

EPIDEMIOLOGY AND PATHOBIOLOGY

The infection usually spreads from the paranasal sinuses, dental abscesses, or other infections affecting the orbit or middle third of the face. *S. aureus* is the most common organism. Streptococci, pneumococci, and gram-negative bacilli are less common; anaerobic infection has also been reported. Many cases of idiopathic intracranial hypertension (Chapters 189 and 398) are due to thrombosis in the lateral sinuses.

CLINICAL MANIFESTATIONS

Cavernous sinus thrombosis may be manifested as an acute fulminant disease or have an indolent subacute manifestation. Fever and other systemic symptoms from sepsis may be present. Clinical symptoms and signs are related to anatomic structures within the cavernous sinuses or drained by them: unilateral periorbital edema, headache, photophobia, proptosis, ophthalmoplegia, pupillary dilation, decreased corneal reflex, and periorbital sensory loss. Obstruction of venous drainage from the retina can result in papilledema, retinal hemorrhages, and visual loss. The infection can spread rapidly (24 to 48 hours) through the intercavernous sinuses to the contralateral cavernous sinus. Thrombus can extend to other dural venous sinuses, adjacent vascular structures, or the brain parenchyma.

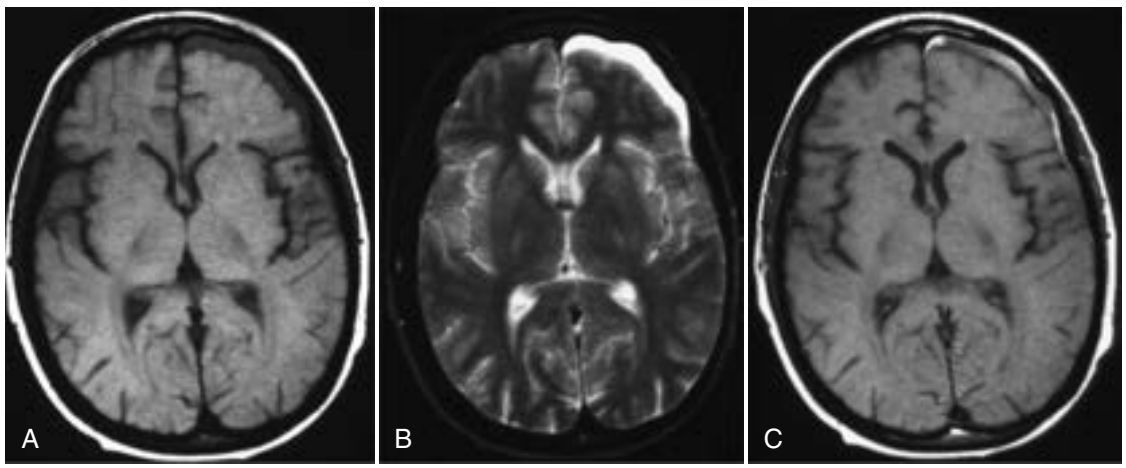

FIGURE 413-4. Subdural abscess. A, T1-weighted MRI shows a hypodense area in the left frontal region. B, T2-weighted image shows increased signal intensity in the same region. C, A contrast scan shows enhancement in the same region.

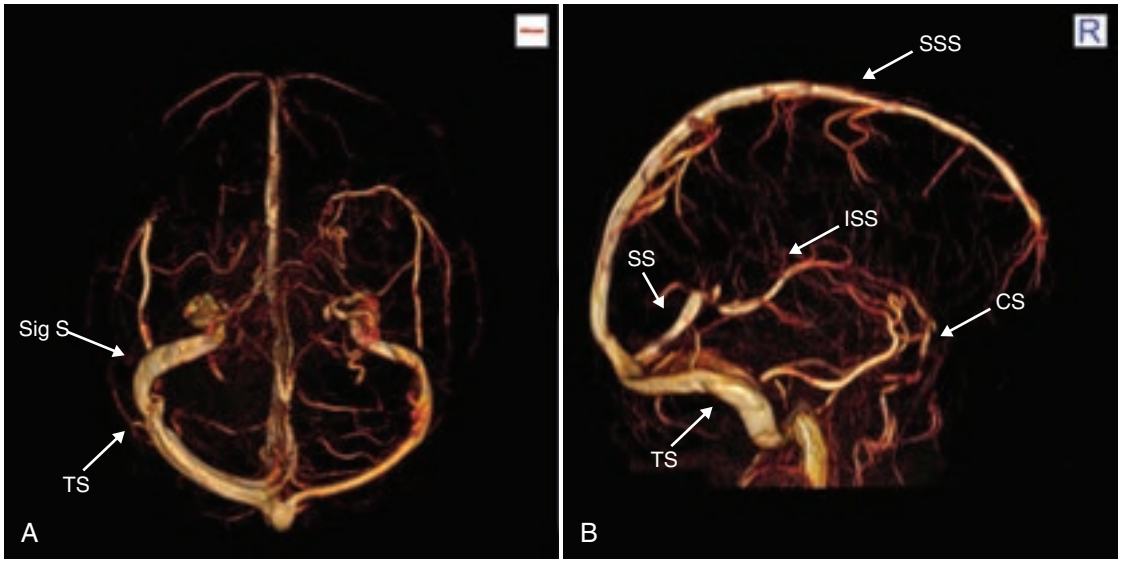

FIGURE 413-5. Anatomy of major venous sinuses. Magnetic resonance venography of the brain shows the normal venous sinuses. **A** shows sigmoid sinus (Sig S) and transverse sinus (TS). **B** shows superior sagittal sinus (SSS), inferior sagittal sinus (ISS), straight sinus (SS), transverse sinus (TS), and cavernous sinus (CS).

DIAGNOSIS

The diagnosis is made on clinical findings and confirmed by radiographic studies. Radiologic evaluation includes sinus imaging, particularly the sphenoid and ethmoid sinuses. MRI using flow parameters and MR venogram is sensitive and may reveal deformity of the cavernous portion of the internal carotid artery, a heterogeneous signal from the abnormal cavernous sinus, and an obvious hyperintense signal of thrombosed vascular sinuses. MRI with IV gadolinium can demonstrate venous thrombosis by illustrating a lack of the normal "flow void" within vascular structures. Cranial CT scans are less helpful but may show a subtle increase in the size and enhancement of the thrombosed sinus. MR angiography may demonstrate extrinsic narrowing of the intracavernous portion of the internal carotid artery.

TREATMENT AND PROGNOSIS

Rx

Blood cultures are often negative, so delays in diagnosis are common. Even when the diagnosis is established, empirical antimicrobial treatment may not provide full coverage.

Treatment consists of prompt drainage of infected paranasal sinuses or other identifiable source of infection, as well as specific antistaphylococcal agents (Chapter 288). Heparin anticoagulation without a loading dose is sometimes initiated to reduce morbidity from associated brain ischemia, but experience in septic venous thrombosis is limited compared with the more frequent use of anticoagulation in nonseptic venous thromboses. Hemorrhage caused by anticoagulation is rare in this setting. Despite modern therapy, mortality rates remain as high as 44%.

Lateral Sinus Thrombosis

Septic thrombosis of the lateral sinus results from acute or chronic infections of the middle ear.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Symptoms consist of ear pain and fever followed by headache, nausea, vomiting, loss of hearing, and vertigo, usually evolving over a period of several weeks. Symptoms or signs suggestive of otitis media (Chapter 426), including mastoid swelling, may be seen. Sixth cranial nerve palsies can occur, but other focal neurologic signs are rare. In some patients with nonseptic lateral sinus thrombosis, headache may be the only symptom. Papilledema occurs in 50% of cases, and elevated CSF pressure is present in most, especially with occlusion of the right lateral sinus, which is the major venous conduit from the superior sagittal sinus (Fig. 413-6).

CSF is usually normal, although a parameningeal inflammatory profile (mild pleocytosis, slight elevation in protein level, and a normal glucose level) may be seen. The diagnosis is confirmed by MR venography.

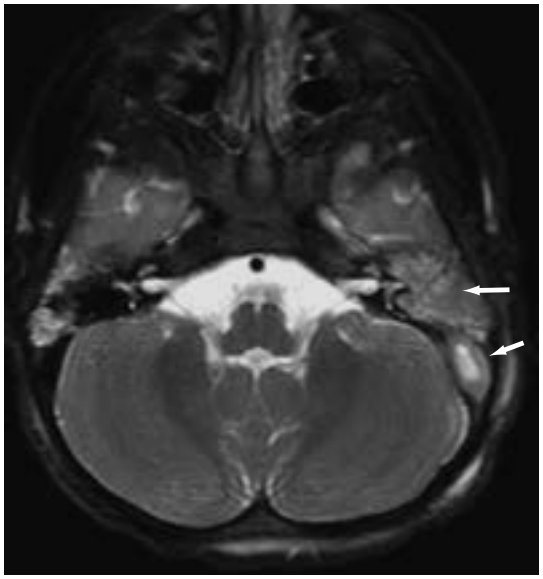

FIGURE 413-6. Lateral sinus thrombosis. Magnetic resonance imaging shows a thrombus in the lateral sinus (*short arrow*) with accompanying mastoiditis (*long arrow*).

TREATMENT

Rx

Treatment includes IV antibiotics to cover staphylococci, anaerobes, and gram-negative bacilli such as *Proteus* species and *Escherichia coli* (nafcillin, 2 g every 6 hours, or vancomycin, 1 g every 12 hours; plus cefotaxime, 2 g every 6 hours, or ceftriaxone, 2 g every 12 hours; plus metronidazole, 7.5 mg/kg every 6 hours, or clindamycin, 300 mg every 6 hours; plus ciprofloxacin, 400 mg every 12 hours). Surgical drainage (mastoidectomy or tympanoplasty) may be required to eradicate the nidus of infection and determine the antibiotic susceptibility of the organism. If the sinus contains pus, it must be opened so the septic thrombus can be removed. Unless vision is compromised, increased intracranial pressure seldom requires specific treatment such as drainage or placement of a shunt.

PROGNOSIS

Broad IV antibiotic coverage and eradication of the perisinus infection, which may require surgical drainage, early in the course of the illness lead to a good prognosis. Neurologic sequelae may include a sixth nerve palsy, ataxia, and hearing loss.

Septic Sagittal Sinus Thrombosis

Although superior sagittal sinus thrombosis is the most common form of venous sinus thrombosis and is frequently associated with the use of oral contraceptives, septic sagittal sinus thrombosis is an uncommon condition that occurs as a consequence of purulent meningitis, infections of the ethmoid or maxillary sinuses spreading through venous channels, compound infected skull fractures, or (rarely) neurosurgical wound infections.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Symptoms are primarily related to the elevated intracranial pressure and can evolve rapidly to stupor and coma. Seizures and hemiparesis may result from cortical infarction. Early recognition and treatment are necessary because septic sagittal sinus thrombosis carries a high mortality rate. The rate of progression, severity of symptoms, and prognosis are all related to the location of thrombosis. Obstruction of the anterior third of the sinus produces less intense symptoms and evolves more slowly.

CSF abnormalities are frequent, including enough red blood cells that the CSF can sometimes be mistaken for a subarachnoid hemorrhage; the opening pressure is increased in proportion to the extent of sagittal sinus involvement. A septic sagittal sinus is best visualized during the venous phase of cerebral angiography or MR venography. The diagnosis can also be made by MRI, which demonstrates an abnormal increase in signal intensity (absent flow void) within the affected venous sinus. Contrast-enhanced CT scanning may reveal a contrast void lying at the junction of the transverse and sagittal sinuses (the region of the torcular); this so-called delta sign is an intraluminal clot surrounded by contrast material.

TREATMENT Rx

IV antibiotics should be directed at organisms recovered from the meningeal process or the meningeal site. *S. aureus* (Chapter 288), β-hemolytic streptococci (Chapter 290), pneumococci (Chapter 289), and gram-negative aerobes such as *Klebsiella* species (Chapter 306) are the most common organisms. Associated paranasal sinusitis should be drained surgically.

PROGNOSIS

If the thrombosis progresses to involve the middle and posterior thirds of the sinus, deterioration progresses rapidly. The prognosis is poor, with a mortality rate of nearly 30%.

NEUROLOGIC COMPLICATIONS OF INFECTIOUS ENDOCARDITIS

Neurologic complications develop in nearly one third of patients with infective endocarditis (Chapter 76), and neurologic manifestations are the initial symptom in 20% of patients with infective endocarditis. In nearly 30% of patients, the neurologic complications occur within 2 weeks after the initiation of treatment. Stroke is the most common manifestation; most strokes are due to cerebral emboli, and others are due to intracerebral hemorrhage. Infective endocarditis should always be considered in a patient with a fever and stroke.

PATHOBIOLOGY

Cerebral embolization occurs as a result of dislodgement or disruption of the cardiac vegetations and frequently causes occlusion of cerebral blood vessels. Emboli occurring before the initiation or completion of treatment with antibiotics may contain microorganisms capable of causing metastatic infections such as abscesses, arteritis, meningitis, or mycotic aneurysms. Most cerebral emboli involve small or moderate-sized blood vessels, and multiple cerebral emboli are common. Intracranial hemorrhage is usually due to rupture of a mycotic aneurysm (Chapter 408), septic erosion of the arterial wall without the formation of an aneurysm, or hemorrhagic transformation of a large cerebral infarct. Mycotic aneurysms are observed in approximately 2 to 3% of patients with infective endocarditis. About 20% of patients with mycotic aneurysms have multiple aneurysms; involvement of the middle cerebral artery and its branches occurs in more than 75% of patients, unlike congenital aneurysms, which occur predominantly in the circle of Willis. Mycotic aneurysms develop as a result of either septic embolization into the vasa vasorum or direct penetration of the microorganism into the wall of the artery. Streptococci and staphylococci account for nearly 90% of all mycotic aneurysms.

CLINICAL MANIFESTATIONS

The nature of the clinical manifestations depends on the underlying pathophysiology.[11] Embolic stroke typically causes the acute onset of a focal neurologic deficit. Seizures may also occur. Multiple microemboli result in an altered or fluctuating level of consciousness not adequately explained by other abnormalities.

Most patients with mycotic aneurysms have a sudden, often fatal, subarachnoid or intracerebral hemorrhage without warning signs. Warning signs, if present, include severe localized headache, ischemic events, seizures, and cranial nerve abnormalities. In some patients, mycotic aneurysms may be asymptomatic and resolve with antibiotic therapy. Some patients develop micro- or macroabscesses, septic or aseptic meningitis (Chapter 412), or a generalized toxic metabolic encephalopathy.

DIAGNOSIS

MRI is the modality of choice for the diagnosis of cerebral infarcts and brain abscesses related to endocarditis. Gradient echo sequences on the MRI may be more sensitive than a CT scan for detecting intracranial hemorrhage and also can detect microbleeds. MRI should also include diffusion-weighted sequences for detection of infarcts. An MR angiogram is preferred for diagnosing an aneurysm. CSF evaluation is useful if accompanying meningitis or a slow leak from an aneurysm is suspected but not visualized with these imaging tests.

TREATMENT Rx

Treatment of patients with infective endocarditis and cerebral emboli requires prevention of embolization with appropriate antibiotic therapy and sometimes cardiac surgery (Chapter 76). Anticoagulation is contraindicated in patients with cerebral infarcts and septic emboli because of the high risk for complications from intracerebral bleeding.

Patients with unruptured aneurysms smaller than 7 mm in diameter, proximal aneurysms, multiple aneurysms, ruptured aneurysms without an intracerebral hematoma, and aneurysms for which excision is likely to cause a neurologic deficit can be monitored conservatively with serial MRI and MR angiography. All other aneurysms require surgical excision of the aneurysm and the adjacent septic vessel wall. Patients who cannot undergo surgery may be candidates for endovascular embolization of the aneurysmal vessel.

PROGNOSIS

Mortality rates in patients with infective endocarditis and cerebral emboli range from 30 to 80%.[12] Mortality is high if there is hemorrhagic transformation of the infarct. Mortality in patients with ruptured mycotic aneurysms is 80%, and even patients with unruptured aneurysms have a mortality rate of 30%.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Helweg-Larsen J, Astradsson A, Richhall H, et al. Pyogenic brain abscess, a 15 year survey. *BMC Infect Dis*. 2012;12:332.
2. Brouwer MC, Tunkel AR, McKhann GM 2nd, et al. Brain abscess. *N Engl J Med*. 2014;371:447-456.
3. Rath TJ, Hughes M, Arabi M, et al. Imaging of cerebritis, encephalitis, and brain abscess. *Neuroimaging Clin N Am*. 2012;22:585-607.
4. Ratnaike TE, Das S, Gregson BA, et al. A review of brain abscess surgical treatment–78 years: aspiration versus excision. *World Neurosurg*. 2011;76:431-436.
5. Arlotti M, Grossi P, Pea F, et al. Consensus document on controversial issues for the treatment of infections of the central nervous system: bacterial brain abscesses. *Int J Infect Dis*. 2010;14:S79-S92.
6. Brouwer MC, Coutinho JM, van de Beek D. Clinical characteristics and outcome of brain abscess: systematic review and meta-analysis. *Neurology*. 2014;82:806-813.
7. Smith RM, Schaefer MK, Kainer MA, et al. Fungal infections associated with contaminated methylprednisolone injections. *N Engl J Med*. 2013;369:1598-1609.
8. Patel AR, Alton TB, Bransford RJ, et al. Spinal epidural abscesses: risk factors, medical versus surgical management, a retrospective review of 128 cases. *Spine J*. 2014;14:326-330.
9. Malani AN, Vandenberg DM, Singal B, et al. Magnetic resonance imaging screening to identify spinal and paraspinal infections associated with injections of contaminated methylprednisolone acetate. *JAMA*. 2013;309:2465-2472.
10. Weimar C. Diagnosis and treatment of cerebral venous and sinus thrombosis. *Curr Neurol Neurosci Rep*. 2014;14:417.
11. Pruitt AA. Neurologic complications of infective endocarditis. *Curr Treat Options Neurol*. 2013;15:465-476.
12. Garcia-Cabrera E, Fernandez-Hidalgo N, Almirante B, et al. Neurological complications of infective endocarditis: risk factors, outcome, and impact of cardiac surgery: a multicenter observational study. *Circulation*. 2013;127:2272-2284.

REVIEW QUESTIONS

1. A patient with a solitary large ring-enhancing lesion with a necrotic center on magnetic resonance imaging (MRI) should be treated by:

A. Surgical drainage
B. Empirical antibiotics
C. An investigation for peripheral source of infection and treatment with appropriate antibiotics
D. Steroids and then biopsy of the lesion
E. Repeated MRI scans with intervention only if the lesion grows further

Answer: A All large lesions will require drainage. Antibiotics cannot penetrate the capsule, and even if they did, there would be a dose gradient with low concentrations of the antibiotic in the center of the lesion. Aggressive tumors rarely have a necrotic center, so some tissue should also be obtained from the edge of the lesion.

2. An HIV-infected immunocompromised patient has multiple ring-enhancing lesions on brain MRI. The appropriate management would be:

A. Empirical antituberculosis therapy
B. Brain biopsy to determine if it is a tumor, such as central nervous system lymphoma, or infection followed by biopsy-guided treatment
C. Empirical antitoxoplasmosis therapy
D. Cerebrospinal fluid analysis should be performed for detection of microorganisms and for detection of tumor cells by flow cytometry, followed by treatment based on the results.
E. Positron emission tomography should be performed to see if there are any peripheral lesions that may be easier to biopsy.

Answer: C Multiple ring-enhancing lesions in an HIV-infected patient with a low CD4 cell count are most often due to toxoplasmosis. Owing to immunosuppression, the capsule is not well formed, and the lesions often are not large; as a result, antibiotic penetration is good. If the patient does not respond to treatment in 2 weeks as determined by a repeat MRI scan, brain biopsy should be considered.

3. A 35-year-old man with sinusitis for 2 weeks develops blurry vision with a left-sided frontal headache. Examination shows mild proptosis and conjunctival edema with retinal venous hemorrhages. Appropriate management would be:

A. Computed tomography (CT) scan of the orbit to localize the infection or abscess in the orbital bone, because it may extend to the orbit by eroding the sinuses
B. Drain the sinuses to diagnose the organism, and treat immediately with appropriate antibiotics.
C. Obtain an echocardiogram to exclude infective endocarditis as a source of the infection.
D. Treat empirically with antibiotics and corticosteroids.
E. Obtain MRI and MR venogram to look for occlusion of venous sinuses.

Answer: E The clinical presentation is suggestive of a cavernous sinus thrombosis, which may occur because of extension of infection from sinusitis. A combination of antibiotics and anticoagulation may be necessary. The cavernous sinus thrombosis can spread to other sinuses and cause venous infarcts in the brain. Systemic infection may also require emergent treatment.

4. Which of the following is true regarding mycotic aneurysms?

A. They are caused by mycoses or fungal infections.
B. They represent metastatic infections in which the microorganisms get lodged in the blood vessels and cause the aneurysm.
C. As far as possible, all cerebral mycotic aneurysms should be treated with surgical excision or endovascular coils/embolization.
D. The most common clinical manifestation is when occlusion of the aneurysm leads to either an infarct or abscess.
E. They can occur in the brain without any systemic infection.

Answer: B Mycotic aneurysms may be due to a variety of different microorganisms. The primary infection is usually in the heart, from where the organism may embolize to the brain or other organs. For example, splinter hemorrhages may be seen in the nail beds, and retinal hemorrhages may be seen owing to occlusion of the distal small blood vessels. The most common manifestation of a mycotic aneurysm is a subarachnoid hemorrhage, which is often fatal. Surgical or endovascular intervention should be considered in a nonruptured aneurysm that is 1 cm or greater in diameter.

5. Over a period of 2 to 3 weeks, a 50-year-old woman developed pain in the right ear, with a low-grade fever followed by headache, nausea, vomiting, loss of hearing, and vertigo. The most likely diagnosis is:

A. Thrombosis of the lateral venous sinuses
B. Middle ear infection
C. Meniere disease
D. Brain stem abscess
E. Any of the above

Answer: A Middle ear infections are not associated with vertigo. Meniere disease does not cause earache and fever. Brain stem abscess is unlikely to cause earache or loss of hearing. Hence, the most likely cause of this constellation of symptoms is a lateral venous sinus thrombosis.

414

ACUTE VIRAL ENCEPHALITIS

ALLEN J. AKSAMIT JR.

DEFINITION

Encephalitis is a diffuse or focal inflammation of the parenchyma of the brain. The term *encephalitis* indicates that the predominant clinical syndrome arises from infection and inflammation in the parenchyma of the brain rather than in the leptomeninges. When both the leptomeninges and brain parenchyma are involved, the term *meningoencephalitis* is used.

EPIDEMIOLOGY

Viral encephalitis has an estimated incidence of 7 per 100,000 per year. In general, a specific cause is identified in less than 50% of patients in the United

TABLE 414-1 COMMON CAUSES OF ENCEPHALITIS IN THE UNITED STATES

I. Causes of viral encephalitis
- A. Nonseasonal
  - Herpes simplex virus type 1 (herpes simplex encephalitis)
  - Herpes simplex virus type 2 (neonatal encephalitis or adult meningoencephalitis)
- B. Seasonal—summer and fall—arboviruses (arthropod borne)
  - West Nile virus
  - St. Louis encephalitis virus
  - Eastern equine encephalitis virus
  - Western equine encephalitis virus
  - La Crosse/California encephalitis virus
- C. Seasonal—non–arthropod borne
  - Summer and fall: enteroviruses (including coxsackieviruses, echoviruses, polioviruses, and enterovirus 71)
  - Winter: influenza virus
- D. Immunosuppressed patients
  - Human immunodeficiency virus (chronic HIV encephalitis)
  - Varicella-zoster virus (subacute encephalitis)
  - JC virus (progressive multifocal leukoencephalopathy)
  - Cytomegalovirus (ventriculitis or encephalitis)
  - Human herpesvirus 6 (subacute encephalitis)
  - Epstein-Barr virus (subacute encephalitis)

II. Uncommon causes in the United States
- Powassan fever encephalitis virus
- Lymphotropic choriomeningitis virus
- Rabies
- Measles (subacute sclerosing panencephalitis)
- Mumps
- Adenovirus
- Herpes B virus (of monkeys)
- Rubella (progressive rubella panencephalitis)

III. Causes outside the United States
- Tick-borne encephalitis virus (Russia, Asia)
- Japanese encephalitis virus (Japan, Southeast Asia, Malaysia)
- Venezuelan equine encephalitis virus (Central and South America)
- Dengue virus (Southern Asia, Africa, South America)
- Rift Valley fever virus (east central Africa)
- Murray Valley encephalitis virus (Australia)
- Powassan fever encephalitis virus (Canada)
- Nipah virus (Malaysia and Bangladesh)

States. Many viruses (Table 414-1) are implicated, and testing by serologic or nucleic acid identification (by polymerase chain reaction [PCR]) is required to identify the specific virus.[1] The epidemiology of each virus responsible for central nervous system infection (see Table 414-1) is distinct in terms of the patients who are at highest risk, geographic distribution, and seasonal occurrence, especially the arboviruses (Chapter 383) and enteroviruses (Chapter 379), which are covered in separate chapters.

In the United States, the most common cause of nonepidemic encephalitis is herpes simplex encephalitis, which is caused by herpes simplex virus type 1 (Chapter 374). The most common epidemic virus in the United States is now West Nile virus (Chapter 383), which is a mosquito-transmitted Flavivirus related to St. Louis encephalitis virus and its Asian counterpart Japanese encephalitis virus. There is serologic cross-reactivity between St. Louis encephalitis, Japanese encephalitis, and West Nile viruses.

PATHOBIOLOGY

In general, gross pathologic inspection of an encephalitic brain does not reveal purulence visible to the naked eye. If focal purulence is present, *cerebritis* is the more correct term. If frank necrosis and purulence are present, the correct pathologic term is *brain abscess* (Chapter 413). Encephalitis, however, can be associated with substantial necrosis, and patients with severe acute viral encephalitis frequently have microscopic evidence of necrosis. Certain viral encephalitides, such as herpes simplex encephalitis, can be both focal and hemorrhagic. Viruses that cause acute encephalitis may often also cause meningitis (Chapter 412). Indeed, patients with encephalitis virtually always have some microscopic inflammatory changes in the leptomeninges. Conversely, patients with viral meningitis will inevitably have some component of microscopic encephalitis. The degree of inflammatory change present in the brain is determined by the individual viral pathogen and by host immune factors, which are responsible for the reaction to the invading virus.[2]

CLINICAL MANIFESTATIONS

The clinical findings in patients with acute viral encephalitis start with a prodrome of fever, headache, malaise, myalgia, and nonspecific symptoms.[3] Nausea, vomiting, diarrhea, cough, sore throat, and rash can precede the neurologic symptoms as part of the systemic initial manifestations of the infection. Invasion of the nervous system is typically accompanied by headache, photophobia, and altered consciousness, with symptoms progressing over a period of several days. Seizures are also a common heralding symptom. Signs of meningeal irritation may be present and are an unreliable finding in encephalitis.

Focal brain dysfunction is seen with some viruses. For example, West Nile virus (Chapter 383) can cause a brain stem encephalitis with an early onset of coma. Herpes simplex virus (Chapter 374) tends to cause focal cortical neurologic deficits, including hemiparesis, aphasia, and seizures. Limbic parts of the brain commonly involved by herpes simplex encephalitis or rabies can lead to prominent behavioral changes at the beginning of the illness before the patient's level of consciousness is depressed. Focal or generalized seizures are particularly common when encephalitis affects the hippocampus and limbic system. Rabies is typically associated with brain stem–mediated laryngospasm, hydrophobia, and depressed consciousness. Because of spinal cord anterior horn cell involvement, West Nile virus, St. Louis encephalitis virus, poliovirus, and rabies virus infections can cause focal or asymmetrical weakness with areflexia.

DIAGNOSIS

In patients with coma or focal deficits, computed tomography (CT) of the head should usually be performed before spinal fluid analysis to exclude a substantial mass effect and to avoid the risk of herniation during lumbar puncture.[4] In patients without focal findings, however, lumbar puncture should be performed immediately to establish the diagnosis and allow early empirical treatment. Opening pressures should be measured because increased intracranial pressure can occur with all forms of viral encephalitis and may need additional treatment.[5]

Spinal fluid analysis typically reveals an elevated protein level, which usually is less than 120 mg/dL. The cerebrospinal fluid (CSF) glucose level is typically normal and greater than 40% of the coincident serum value, but rare patients may have a low CSF glucose level suggestive of a bacterial infection (Chapter 412). The CSF white blood cell count is typically elevated, usually in the range of 10 to 500 cells/μL. The cell type is usually a lymphocytic predominance. However, a polymorphonuclear predominance is seen in some cases of West Nile encephalitis and cytomegalovirus ventriculitis.

Serologic or PCR testing on spinal fluid is helpful (Table 414-2). PCR testing has the added advantage of proving direct viral infection within the central nervous system, but serologic testing is more appropriate for some infections like West Nile virus encephalitis, which is best confirmed by an IgM antibody response in spinal fluid.

Magnetic resonance imaging (MRI) of the brain is the most sensitive technique for defining abnormalities in patients with viral encephalitis.[6] However, frank viral encephalitis can occur with normal findings on MRI. MRI findings also can suggest the responsible virus. For example, herpes simplex encephalitis has a characteristic pattern involving the mesiotemporal, inferofrontal, and insular cortices, usually unilateral or asymmetrically bilateral.

Differential Diagnosis

A number of nonviral pathogens can cause encephalitis that is clinically and pathologically indistinguishable from viral encephalitis.[7] Examples include *Rickettsia* (Chapter 327), *Borrelia* (Chapter 322), Whipple disease (Chapters 140 and 275), *Toxoplasma* (Chapter 349), *Mycoplasma* (Chapter 317) and *Acanthamoeba* (Chapter 352). Other forms of infectious non-viral causes mimicking viral encephalitis include bacterial cerebritis, meningovascular syphilis, and cerebral cysticercosis.

Additionally, autoimmune encephalitides can mimic viral encephalitis, including limbic paraneoplastic encephalitis, especially associated with antibodies against the voltage-gated potassium channel complex, Hashimoto encephalopathy associated with autoimmune thyroiditis (Chapter 226), and encephalitis associated with anti–*N*-methyl-D-aspartate (NMDA) receptor antibodies.[8] In parainfectious encephalitis, a systemic viral infection is associated with a febrile encephalopathy, sometimes with inflammatory spinal fluid but without direct evidence of brain invasion by the organism. Examples of parainfectious encephalitis include infection and encephalopathy associated with influenza virus (Chapter 364), varicella virus (Chapter 375), and

TABLE 414-2 SELECTED TESTS FOR VIRAL ENCEPHALITIS

ORGANISM/ SYNDROME	TEST	COMMENT
WEST NILE VIRUS		
West Nile encephalitis	IgM in CSF	Diagnostic of CNS invasive disease including encephalitis or acute flaccid paralysis
HERPES SIMPLEX VIRUS TYPE 1		
Herpes simplex encephalitis	PCR in CSF	Sensitive and specific in the acute phase
HERPES SIMPLEX VIRUS TYPE 2		
Neonatal encephalitis Relapsing meningitis	PCR in CSF PCR in CSF	Confirmatory, high sensitivity Sensitive and specific in first 3 days of illness
HUMAN HERPESVIRUS 6		
Limbic encephalitis	PCR in CSF	Confirmatory, sensitivity unknown
VARICELLA-ZOSTER VIRUS		
Meningoencephalitis	PCR in CSF	Confirmatory when used with clinical and spinal fluid findings; sensitivity unclear
EPSTEIN-BARR VIRUS		
EBV encephalitis	PCR in CSF	Suggests CNS invasion by virus
JC VIRUS		
Progressive multifocal leukoencephalopathy	PCR in CSF	Diagnostic but incompletely (70%) sensitive
CYTOMEGALOVIRUS		
CMV ventriculitis	PCR in CSF	Sensitive and specific

CNS = central nervous system; CSF = cerebrospinal fluid; IgM = immunoglobulin M; PCR = polymerase chain reaction.

Epstein-Barr virus (Chapter 377). Furthermore, primary demyelinating disease (Chapter 411), particularly in the form of acute disseminated encephalomyelitis, overlaps clinically with viral encephalitis.

SELECTED SPECIFIC VIRUSES

Herpes Simplex Encephalitis

EPIDEMIOLOGY

Herpes simplex (Chapter 374) encephalitis, which is second only to West Nile encephalitis (Chapter 383) as the most common form of encephalitis in the United States, has an annual incidence of two to four cases per million people per year. There is no seasonal or gender predisposition. The encephalitis can strike older children but is most commonly a disease of adults.

PATHOBIOLOGY

Herpes simplex encephalitis usually occurs in immunocompetent patients, but immunosuppressed patients may also be affected. Patients who are deficient in toll-like receptor 3 in the immune system may be selectively vulnerable to herpes simplex encephalitis.

Herpes simplex virus type 1 infects and establishes latency in the majority of the population. Whether herpes simplex encephalitis arises from reactivation of a latent viral infection in the trigeminal ganglion or is a primary nasopharyngeal infection that ascends into the olfactory nervous system is uncertain.

The pathology of herpes simplex encephalitis is a necrotizing hemorrhagic inflammatory encephalitis in a characteristic pattern affecting the mesiotemporal, inferofrontal, and insular cortices, with gray matter predominance. Even if the brain is affected bilaterally, the pathologic features are usually asymmetrical, a pattern that helps distinguish herpes simplex from other forms of limbic encephalitis.

CLINICAL MANIFESTATIONS

The clinical manifestations of herpes simplex encephalitis usually begin with a nonspecific febrile prodrome that is followed within hours to days by the symptoms of headache, malaise, nausea, and vomiting. A reduced level of consciousness may occur early. Seizures may be the first manifestation of this encephalitis. Focal neurologic deficits, such as hemiparesis or aphasia, appear early and can be mistaken for stroke. More specific manifestations of herpes simplex encephalitis are symptoms of limbic system–associated behavioral changes, such as behavioral or emotional lability and inappropriateness. Memory is affected early if consciousness is preserved. As the encephalitis progresses, symptoms of increased intracranial pressure, lethargy, and coma are usual. Focal findings alone in the context of clinical encephalitis are not sufficient to confirm a diagnosis of herpes simplex encephalitis.

DIAGNOSIS

Spinal fluid analysis is necessary in the diagnosis of herpes simplex encephalitis. In a patient with focal encephalitis or coma, however, CT of the brain should be performed before spinal fluid analysis to avoid the risk of herniation. Elevation of the CSF protein level and the white blood cell count, with a predominance of lymphocytes, is the most frequent pattern; red blood cells are also commonly seen. The CSF glucose level is usually normal but is less than 50% of the blood glucose level in about 5% of patients.

The best and most accurate test for proof of herpes simplex encephalitis is the presence of herpes simplex virus type 1 DNA amplified by PCR in the spinal fluid. Herpes simplex virus type 1 can be distinguished from herpes simplex virus type 2 by specific primer amplification, applied as part of the PCR analysis. Because herpes simplex type 2 can cause encephalitis in neonates and meningoencephalitis in adults, this distinction may guide therapy.

MRI typically shows characteristic focal involvement with increased T2 and fluid-attenuated inversion recovery (FLAIR) signal in the mesiotemporal lobes (including the amygdala, hippocampus, and uncus), the inferofrontal lobes (cingulate gyrus and orbital frontal cortex), and the insular cortex (Fig. 414-1). MRI abnormalities are often unilateral but can be bilateral and asymmetrical. Focal MRI abnormalities must be distinguished from brain abscess (Chapter 413), cerebral infarction (Chapter 407), cerebral hemorrhage (Chapter 408), brain tumors (Chapter 189), and paraneoplastic limbic encephalitis. Radiographically detected involvement of the mesiotemporal rather than the lateral temporal areas and involvement of the gray matter rather than the white matter suggest herpes simplex encephalitis as the diagnosis. Early gadolinium contrast enhancement may occur but is not universal.

CT of the head is less sensitive than MRI for detecting mild cases of herpes encephalitis. However, because herpes simplex encephalitis can be hemorrhagic, CT may sometimes identify the hemorrhage more accurately than MRI can.

Electroencephalography is an adjunctive test that can show periodic lateralized epileptiform discharges ipsilateral to the involved temporal lobe. However, the findings are not specific for herpes simplex encephalitis and commonly occur in patients with cerebral infarction (Chapter 407) and occasionally other forms of viral encephalitis.

TREATMENT Rx

Multicenter prospective trials emphasize that early treatment affects outcome. When suspicion for herpes simplex encephalitis is raised in the acute setting by the presence of focal signs or symptoms, early empirical treatment is recommended even while the diagnostic evaluation is proceeding.

Intravenous acyclovir (10 mg/kg every 8 hours for 14 to 21 days) is the therapy of choice. No prospective data support a longer duration of therapy or higher doses of acyclovir to improve neurologic outcomes.

Rabies

Human rabies is an encephalitic illness caused by the rabies virus, usually transmitted by an animal bite. It produces a fatal encephalitis, although the latency between animal bite exposure and occurrence of neurologic symptoms may sometimes obscure the diagnosis.

EPIDEMIOLOGY AND PATHOBIOLOGY

Rabies is a rare illness in the United States and developed world.[9] However, initially unsuspected cases have been transmitted via trivial bites by infected bats, which are widely distributed in every state in the United States except Hawaii. Rabies virus variants in bats are now responsible for the majority of recent human cases in the United States and Canada. Raccoon rabies has extended from Florida into Georgia, Alabama, and South Carolina.

Canine rabies is still endemic in much of the developing world, including Africa, Latin America, Eastern Europe, and Asia, and the vast majority of

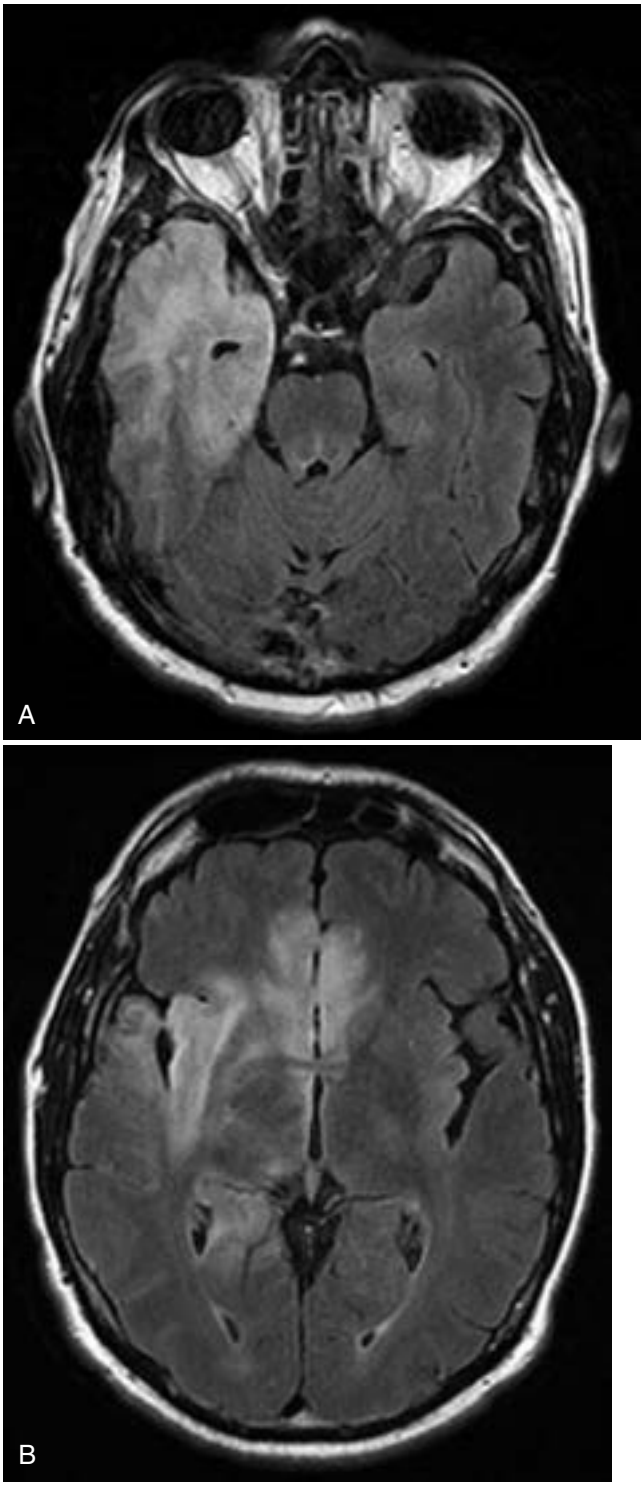

FIGURE 414-1. Magnetic resonance imaging (MRI) in herpes simplex encephalitis. Fluid-attenuated inversion recovery (FLAIR) MRI scan of the brain, showing increased signal (**A**) in the right mesiotemporal lobe (including the amygdala, hippocampus, and uncus), and (**B**) in the bilateral inferofrontal lobes (cingulate gyrus and orbital frontal cortex) and the right insular cortex.

human rabies cases occur as a result of untreated dog bites from endemic areas. Dog rabies came under control in the United States during the 1950s and was associated with a marked reduction in the number of human cases transmitted by dogs. Much of the dog-related clinical rabies seen in the United States is the result of dog bites that occurred in developing countries, before the patient migrated to the United States. Rare cases of transmission of rabies to transplant organ recipients have occurred in the United States.

The presence of Negri bodies (intracytoplasmic viral inclusions) in neurons of the brain stem, cerebellum (especially the Purkinje cells), or hippocampus defines rabies pathologically. These inclusions are often not present, but detection of antigen by immunohistochemical means can aid in the pathologic diagnosis. Because the disorder tends to be a brain stem encephalitis, the bulbar cardiovascular and respiratory centers are affected.

CLINICAL MANIFESTATIONS

Human rabies usually develops 20 to 90 days after a bite, although rarely disease develops after only a few days or after a year or more following bite exposure. Multiple bites and facial bites are associated with shorter incubation times.

Nonspecific prodromal symptoms include fever, chills, malaise, fatigue, insomnia, anorexia, headache, and irritability. In the majority of patients, pain or paresthesias will develop in the limb that was affected by the bite. Following the prodromal illness, an encephalitic form develops in about 80% of patients and causes behaviors ranging from episodes of agitated arousal to quiet lethargy. Fever is a common accompaniment but not universal at this phase. Disinhibition of brain stem reflexes leads to hydrophobia with laryngospasm and an inability to deal with salivation, swallowing of water, or other oral intake. When the brain stem encephalitis affects the bulbar, cardiovascular, and respiratory centers, autonomic dysfunction, cardiopulmonary complications, and respiratory failure may occur.

Another form of rabies that affects up to a third of patients is known as paralytic rabies. This form of rabies is manifested as acute flaccid paralysis, which may be multifocal and affect both the limbs and the bulbar musculature, thereby resembling poliomyelitis (Chapter 379) because of its multifocality. It also can be confused with Guillain-Barré syndrome (Chapter 420). Paralytic rabies typically occurs in conjunction with febrile encephalitis.

DIAGNOSIS

Findings on spinal fluid analysis may be abnormal in human rabies. A lymphocytic pleocytosis, usually less than 100 white cells/μL, is found in more than 50% of patients in the first week of illness. The CSF protein concentration usually is mildly elevated, and the glucose level is usually normal.

Imaging of patients with rabies is sometimes useful. MRI may show gray matter involvement, particularly involvement of the brain stem, with increased T2 signal, commonly without enhancement. Spinal cord MRI in patients with paralytic rabies may show multifocal increased T2 signal mimicking acute disseminated encephalomyelitis. Involvement of brain gray matter, including the hippocampus and basal ganglia structures, indicates the gray matter predilection and often bilateral involvement of supratentorial structures. However, MRI cannot be relied on to exclude rabies.

Serum antibodies against rabies virus are not usually present in unimmunized patients until the second week of illness, and patients can die before having a detectable serum antibody level. Serum antibodies may also be present in spinal fluid, but their absence is unreliable in excluding the diagnosis. Classically, staining a skin biopsy sample taken from an area near the nape of the neck for rabies antigen in the sensory nerves can confirm the diagnosis of rabies. Alternatively, small amounts of rabies RNA can be detected by PCR testing. Typical specimens to detect virus include saliva, brain tissue, or spinal fluid. A positive result confirms the diagnosis, but the exclusionary value of negative results is unknown.

TREATMENT

Rx

After an animal bite, local treatment with antirabies immunoglobulin and systemic treatment with vaccination are typically offered. Rabies postexposure prophylaxis includes local wound cleansing, passive immunization with immunoglobulin, and active immunization with rabies vaccine.[10] Inactivated cell culture rabies vaccines are used for active immunization, and the risk for vaccination-induced acute disseminated encephalomyelitis has been markedly reduced by the use of these vaccines. However, once rabies encephalitis is manifested, it is unclear whether vaccination, though regularly used, has any role in improving outcome. Antiviral therapy and a variety of immunotherapies, including ribavirin and interferon alfa, have been tried in the treatment of rabies, usually without success. Although one patient has survived with the use of therapeutic coma without vaccination, subsequent reports of patients treated in similar fashion have been associated with a fatal outcome. Treatment is otherwise supportive, and outcome is essentially always fatal.[11]

Rare Causes of Encephalitis

Lymphocytic choriomeningitis (Chapter 412) virus is a human infection acquired from mice. Typically, humans acquire the infection by contact with food or dust that is contaminated by excreta of the common house mouse. Most commonly, human disease occurs in winter, when the natural host

tends to move indoors. It can also be acquired as a consequence of laboratory exposure by human caretakers.

Mumps virus (Chapter 369) is typically acquired by the respiratory route. Infection can occur throughout the year, but the incidence is higher during the spring. Although mumps virus infects both sexes equally, meningoencephalitis develops in males three times more frequently than in females. Vaccination programs in the United States have made mumps encephalitis rare.

TREATMENT

Rx

Effective antiviral therapy does not exist for most forms of viral encephalitis, except for herpes simplex encephalitis.[12] However, because of the usual delay in establishing or excluding the diagnosis of herpes simplex encephalitis, patients suspected of having encephalitis should start acyclovir therapy (10 mg/kg intravenously every 8 hours for 2 weeks) even while specific serologic and spinal fluid analyses are being performed to make a specific diagnosis.

Supportive measures for patients with encephalitis typically include intensive care unit monitoring and treatment in the initial phases of the illness. Seizures are common and frequently refractory to antiepileptic drugs; however, the seizures themselves can increase morbidity and mortality, so vigorous treatment attempts are required (Chapter 403).

In patients who are immunosuppressed (see Table 414-1), the spectrum of possible infections is broader and potentially more treatable. Examples include varicella-zoster virus (Chapter 375), with acyclovir administered at doses similar to those used for herpes simplex virus, and cytomegalovirus (Chapters 370 and 376), with ganciclovir administered at 5 mg/kg intravenously every 12 hours for 2 weeks or cidofovir administered at 5 mg/kg intravenously weekly for 2 weeks, although some patients require long-term oral valganciclovir (900 mg every 24 hours) or intravenous cidofovir (5 mg/kg every 2 weeks). HIV encephalitis (Chapter 394) responds in variable degree to triple antiretroviral therapy. By comparison, no specific treatments are currently effective for Epstein-Barr virus (Chapters 370 and 377) and JC virus (progressive multifocal leukoencephalopathy [Chapter 370]). Variable success has been reported for treatment of HHV-6 encephalitis in hematopoietic stem cell transplant recipients using ganciclovir, foscarnet, or valganciclovir alone or in combination (Chapter 360, Table 360-4).[13]

PROGNOSIS

The prognosis of encephalitis is dependent on the cause, with an overall mortality rate of about 6% in the U.S.[14] Herpes simplex encephalitis, even with adequate treatment, has a 20% mortality, and the likelihood of major persistent morbidity with seizures or defects in memory and behavior is 35 to 40%. Each of the arboviruses has a different mortality rate, with eastern equine encephalitis virus associated with the highest mortality. La Crosse encephalitis virus has the lowest mortality and is the most benign. Some forms of encephalitis have specific sequelae, such as sensorineural deafness or hydrocephalus associated with mumps encephalitis.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Granerod J, Ambrose HE, Davies NW, et al. Causes of encephalitis and differences in their clinical presentations in England: a multicentre, population-based prospective study. *Lancet Infect Dis*. 2010;10:835-844.
2. Greenlee JE. Encephalitis and postinfectious encephalitis. *Continuum (Minneap Minn)*. 2012; 18:1271-1289.
3. Roos KL. Encephalitis. *Handb Clin Neurol*. 2014;121:1377-1381.
4. Steiner I, Budka H, Chaudhuri A, et al. Viral meningoencephalitis: a review of diagnostic methods and guidelines for management. *Eur J Neurol*. 2010;17:999-1009.
5. Gibani MM, Brown RL, Davies NW. Demystifying encephalitis: guidelines for an emergency not to miss. *Br J Hosp Med (Lond)*. 2014;75:12-15.
6. Moritani T, Capizzano A, Kirby P, et al. Viral infections and white matter lesions. *Radiol Clin North Am*. 2014;52:355-382.
7. Venkatesan A, Tunkel AR, Bloch KC, et al. Case definitions, diagnostic algorithms, and priorities in encephalitis: consensus statement of the international encephalitis consortium. *Clin Infect Dis*. 2013;57:1114-1128.
8. Gable MS, Sheriff H, Dalmau J, et al. The frequency of autoimmune *N*-methyl-D-aspartate receptor encephalitis surpasses that of individual viral etiologies in young individuals enrolled in the California Encephalitis Project. *Clin Infect Dis*. 2012;54:899-904.
9. Fooks AR, Banyard AC, Horton DL, et al. Current status of rabies and prospects for elimination. *Lancet*. 2014;384:1389-1399.
10. Rupprecht CE, Briggs D, Brown CM, et al. Use of a reduced (4-dose) vaccine schedule for postexposure prophylaxis to prevent human rabies: recommendations of the advisory committee on immunization practices. *MMWR Recomm Rep*. 2010;59:1-9.
11. de Souza A, Madhusudana SN. Survival from rabies encephalitis. *J Neurol Sci*. 2014;339:8-14.
12. Solomon T, Michael BD, Smith PE, et al. Management of suspected viral encephalitis in adults–Association of British Neurologists and British Infection Association national guidelines. *J Infect*. 2012;64:347-373.
13. Bhanushali MJ, Kranick SM, Freeman AF, et al. Human herpes 6 virus encephalitis complicating allogeneic hematopoietic stem cell transplantation. *Neurology*. 2013;80:1494-1500.
14. Vora NM, Holman RC, Mehal JM, et al. Burden of encephalitis-associated hospitalizations in the United States, 1998-2010. *Neurology*. 2014;82:443-451.

REVIEW QUESTIONS

1. A 57-year-old man presents with 3 days of fever and headache, which were followed today by lethargy and right hemiparesis. The best test to confirm a suspected diagnosis of herpes simplex encephalitis is:

A. Computed tomographic (CT) scan of the head
B. Magnetic resonance imaging (MRI) of the head
C. electroencephalogram (EEG)
D. Spinal fluid herpes simplex virus culture
E. Spinal fluid polymerase chain reaction (PCR) for herpes simplex virus

Answer: E Herpes simplex virus spinal fluid PCR is the most sensitive and specific test to confirm herpes simplex encephalitis (HSE). MRI of the brain frequently will show highly suggestive findings of HSE, but the same findings can be mimicked by autoimmune limbic encephalitis or viral encephalitis associated with HHV-6. CT scanning is usually normal unless there is a hemorrhagic lesion. EEG in HSE shows periodic lateralized epileptiform discharges, but so does infarction of the temporal lobe. Herpes simplex virus culture from spinal fluid is insensitive for making the diagnosis.

2. Empirical treatment to be started at the time of presentation for a patient with suspected herpes simplex encephalitis should include antibiotics (vancomycin and ceftriaxone) and:

A. Acyclovir
B. Cidofovir
C. Dexamethasone
D. Phenytoin
E. Mannitol

Answer: A Acyclovir is the treatment of choice for herpes simplex encephalitis, and a successful outcome depends on its early initiation before diagnostic microbiological studies are completed. Cidofovir is principally used for CMV infection. Empirical dexamethasone therapy is principally used in bacterial meningitis. Phenytoin anticonvulsant therapy is generally not routinely recommended for encephalitis unless clinical seizures are manifest. Mannitol is not indicated in early herpes simplex encephalitis.

3. A 48-year-old man presents with low-grade fever, malaise, and a change in mental status 3 weeks after a bite by a bat. Other clinical features that would support a clinical diagnosis of rabies encephalitis include all of the following ***except***:

A. Laryngospasm when trying to drink water
B. Meningismus with neck flexion
C. Ascending paresthesias in the upper limb on which he was bitten
D. Flaccid paralysis of all four limbs
E. An agitated mental state

Answer: B Meningismus is only rarely associated with rabies meningoencephalitis. Laryngospasm in response to oral fluids (hydrophobia) is characteristic of rabies encephalitis because of brain stem involvement with disinhibition of bulbar reflexes. Paresthesias in the bitten limb are common owing to sensory nerve involvement and axonal transport of the virus. One form of rabies involves the spinal cord and anterior horn cell and causes acute flaccid paralysis, which sometimes mimics Guillain-Barré syndrome. Rabies encephalitis can cause either agitation or a depressed level of consciousness.

415

PRION DISEASES

PATRICK J. BOSQUE

DEFINITION

Prion diseases are a group of closely related neurodegenerative conditions of humans and other mammals. They are caused by an accumulation of abnormally aggregated forms of the prion protein (PrP), a protein that is normally produced in the central nervous system (CNS). This abnormal form of the protein can act as an infectious agent called a prion and transmit disease to another host. The prion is thus an infectious protein conformation that contains no specific nucleic acid. The name *Creutzfeldt-Jakob disease* (CJD) is applied to most human forms of prion disease, although other names are used for some forms.

EPIDEMIOLOGY

Prion diseases occur worldwide, with an incidence of about one case per million annually. These conditions can be acquired sporadically, genetically, or infectiously, but sporadic disease accounts for about 90% of cases, and genetic forms account for almost all the remainder. Both dietary and iatrogenic exposure have transmitted prion disease to humans, but these infectiously acquired forms represent less than 1% of cases in most human populations.

Two human outbreaks of prion disease, kuru and variant CJD, were caused by dietary exposure.[1] Kuru was epidemic in tribes of the Fore language group in the highlands of New Guinea. It probably arose as a case of sporadic prion disease, and then was spread by the practice of ritual cannibalism. The last exposures are thought to have occurred in the late 1950s, but new clinical cases have occurred as recently as 2009, thereby indicating an incubation period of more than 50 years.

Variant CJD is caused by eating meat from cattle infected with the prion disease known as bovine spongiform encephalopathy (BSE). Variant CJD first arose in Great Britain in 1994, about 10 years after an outbreak of a massive BSE epidemic there. Despite the exposure of millions of people to meat contaminated with BSE prions, fewer than 200 people worldwide have contracted variant CJD. The incidence of variant CJD has decreased in recent years as the BSE outbreak in cattle has been contained and the entry of contaminated meat into the food supply restricted.[2]

Contaminated cadaveric dura mater allografts and pituitary-derived growth hormone injections have each caused more than 200 iatrogenic cases of CJD.[3] Most dura mater–associated cases have been from a single product, Lyodura, manufactured before May 1987. All growth hormone cases involve product derived from cadaveric pituitaries before recombinant growth hormone became available in the 1980s. Blood and blood products derived from donors with variant CJD have transmitted the illness, but perhaps surprisingly, sporadic CJD seems not to have been transmitted through blood products. Other modes of iatrogenic spread of CJD are quite rare. Contaminated surgical instruments are persuasively documented to have transmitted CJD on six occasions, and corneal transplants have been documented to transmit CJD only twice.

PATHOBIOLOGY

PrP is a cell surface glycoprotein normally produced in the brain and several other tissues. Its function is unknown, but it may play a role in copper metabolism. An abnormally aggregated form of PrP termed PrP^{Sc} accumulates in the brain in prion disease. Remarkably, PrP^{Sc} is able to recruit the normal form of PrP into the pathologic aggregate. The precise structure of PrP^{Sc} aggregates and the mechanism of prion propagation are incompletely understood, but a basic conceptual model proposes that the normally α-helical regions of PrP directly interact with the β sheets of PrP^{Sc}, lose their normal α-helical structure, and then join the aggregate. At some point, the growing aggregate fractures, thereby creating additional aggregate particles. In this way, an aggregate of PrP^{Sc} can propagate as an infectious agent. A curious feature of prion diseases is that more than one aggregated structure of PrP^{Sc} can be stably propagated, and these various "strains" of prions can give rise to distinct clinical manifestations.

Predominantly β-sheet aggregates of other proteins are implicated as the causes of more common neurodegenerative diseases (e.g., the β-amyloid protein in Alzheimer disease [Chapter 402] and synuclein in Parkinson disease [Chapter 409]).[4] Although these other diseases are not infectiously transmitted, recent studies suggest that a mechanism of self-propagation like that described above for prion diseases may play a role in their pathogenesis.

What precisely initiates sporadic or genetic prion diseases is not known. In infectious forms of prion disease that are transmitted by the alimentary route, prions first replicate in the enteric lymphatic system, including Peyer patches. From the lymphatic system, prions spread to the CNS via sympathetic nerves in lymphatic tissue. Once in the CNS, prions appear to spread trans-synaptically. As with other neurodegenerative diseases associated with accumulations of aggregated proteins, the mechanism by which the PrP aggregates cause neuronal dysfunction and death is unknown.

Pathology

Traditionally, prion diseases are recognized by a combination of vacuolization (status spongiosus) of the gray matter, astrocytic gliosis, and loss of

neurons. In modern practice, they are diagnosed by demonstrating the presence of PrP^{Sc}, using techniques that exploit the enhanced resistance to degradation displayed by these PrP aggregates. In the biochemical method, homogenized brain tissue is treated with a protease that dissolves the normal form of PrP but leaves a resistant core of PrP^{Sc} intact. This protease-resistant core can then be identified by Western blotting. The histologic technique involves hydrolyzing proteins on tissue sections. Normal PrP is dissolved, but PrP^{Sc} can still be identified immunohistochemically. Certain forms of prion disease have a distinct and characteristic histochemical appearance. For example, variant CJD produces a peculiar type of amyloid plaque surrounded by vacuoles, the so-called florid plaque.

Genetics

All inherited forms of prion disease are caused by mutations in the PrP coding sequence of the gene *PRNP*.[5] Mutations associated with familial forms of prion disease include more than 20 missense mutations, two premature stop mutations, and a series of insertions in a region of a repeated eight–amino acid sequence. Genetic forms of prion disease are transmitted in an autosomal dominant pattern, usually with high but incomplete penetrance. Three distinctive forms of prion disease are associated with certain *PRNP* mutations. First, the Gerstmann-Straüssler-Scheinker syndrome is caused by any of several mutations in *PRNP*, the most common of which codes for a substitution of leucine for proline at codon 102 (P102L). Pathologically, there are accumulations of plaques of PrP amyloid in the brain, especially in the cerebellum. Second, fatal familial insomnia is caused by a D178N mutation on the same allele as a methionine at the polymorphic codon 129 of *PRNP*. Pathologically, there is neuronal loss and accumulation of PrP^{Sc} in the thalamus. In contrast, the D178N mutation on an allele with valine at codon 129 causes a disease that is indistinguishable from sporadic CJD. Third, some *PRNP* mutations cause slowly progressive dementia. The most common of these mutations are large expansions of the octapeptide repeat region.

Certain common genotypes affect susceptibility to prion disease. Codon 129 of *PRNP* is polymorphic, with alleles coding for either valine or methionine. Persons who are homozygous (129VV or 129MM) at this allele are overrepresented among victims of sporadic CJD, and all victims of variant CJD carry 129M on both *PRNP* alleles.

CLINICAL MANIFESTATIONS

Sporadic CJD is the most common form of human prion disease. It typically begins in later midlife at an average of about 60 years of age, although onset as young as 17 years and as old as 80 years has been reported. In about 25% of cases, patients or their families report a prodrome of a psychiatric disturbance such as anxiety, depression, or altered sleep.[6] Cognitive dysfunction is usually the most prominent neurologic sign. Unlike Alzheimer disease, however, prion disease typically causes motor signs (e.g., ataxia, bradykinesia, spasticity), vague somatic sensory disturbances, or alterations in visual perception. Myoclonus is a characteristic but not pathognomonic sign. Perhaps the most distinctive feature of prion disease is the pace of its progression. Typically, clear decrements in neurologic function can be observed over a period of weeks.

Variations on this typical presentation can occur in sporadic cases and are more common in genetic and infectiously transmitted disease. At least some of these variations are probably caused by the propagation of prion strains that are different from the strain usually associated with sporadic CJD. Recently, a sporadic form of prion disease termed *variable protease encephalopathy* has been characterized; behavioral disorders related to frontal lobe dysfunction are common, and the PrP^{Sc} found in the brain of victims is generally more sensitive to protease digestion than is typical of PrP^{Sc}.[6,7] Slowly progressive CJD may have clinical manifestations similar to those of familial Alzheimer disease (Chapter 402) or resemble Huntington disease (Chapter 410).

In Gerstmann-Straüssler-Scheinker syndrome, the clinical signs are prominent ataxia, a slower rate of progression than occurs with sporadic CJD (typically 5 to 6 years from onset to death), and late dementia. Fatal familial insomnia begins with anxiety, depression, and sleep disturbance. Ataxia or other motor signs may also develop early in the disease course. Dementia occurs relatively late in the condition. Very rarely, a sporadic form of CJD will be manifested as a clinical syndrome of fatal insomnia. A family with a *PRNP* Y163X truncation has been reported to have prion protein amyloid throughout their peripheral organs, including bowel and peripheral nerves. Patients developed diarrhea, a severe autoimmune neuropathy, cortical amyloid plaques, and cerebral amyloid angiopathy.[8]

Variant CJD acquired by exposure to BSE prions is distinguished clinically from sporadic CJD by a much younger mean age at onset (mean, 26 years; range, 12 to 74 years), the prominence of psychiatric and sensory signs early in the disease, and the later emergence of dementia and motor signs, typically more than 6 months after the first symptoms. Iatrogenic CJD usually resembles sporadic CJD, but a subset of patients may have an ataxic form that clinically and pathologically shares some features with Gerstmann-Straüssler-Scheinker syndrome. Kuru begins with limb pain followed by cerebellar ataxia and tremor ("kuru" means "shiver" in the Fore language). Overt dementia occurs late in the disease course.

DIAGNOSIS

The diagnosis of prion disease should be considered in patients with relatively rapidly progressive dementia, but certain treatable structural, inflammatory, metabolic, endocrine, and nutritional causes of rapidly progressive dementia entities can mimic prion disease (Table 415-1).[9] In particular, any signs of inflammation in the cerebrospinal fluid (CSF) should prompt consideration of a diagnosis other than prion disease. The clinician may also need to consider special tests to search for some rare but treatable conditions (Table 415-2).

If the initial evaluation fails to yield an alternative diagnosis, a number of available tests further support the diagnosis of prion disease. However, prion diseases are rare, and ancillary tests have less-than-perfect specificity and sensitivity; if they are applied indiscriminately, these tests will frequently yield false-positive results. The presence of either elevated CSF levels of the proteins 14-3-3[10] or CSF[11] or blood levels of tau[12] is relatively specific for CJD if inflammatory and ischemic causes of dementia are excluded. An unusual hyperintensity of the deep gray matter (basal ganglia and thalamus)

TABLE 415-1 DIFFERENTIAL DIAGNOSIS OF RAPIDLY PROGRESSIVE DEMENTIA

Neurodegenerative diseases that may mimic CJD	Alzheimer disease (Chapter 402), diffuse Lewy body disease (Chapter 402), frontotemporal dementia (Chapter 402), corticobasal ganglion degeneration (Chapter 402), progressive supranuclear palsy (Chapter 409)
Some less common treatable diseases that mimic CJD	**Autoimmune:** CNS vasculitis (Chapter 270), limbic encephalitis (Chapter 414), Hashimoto encephalopathy (Chapter 414), anti–voltage-gated potassium channel encephalopathy, sarcoidosis (Chapter 95), steroid-responsive autoimmune encephalopathy (Chapter 414) **Infections:** viral encephalitis (Chapter 414), chronic meningitis (Chapter 412), Whipple disease (Chapter 275) **Neoplasms:** primary CNS lymphoma (Chapter 185), intravascular lymphoma (Chapter 185) **Nutritional:** Wernicke encephalopathy (Chapter 416) **Toxicities:** lithium, bismuth, methotrexate **Structural:** Normal-pressure hydrocephalus (Chapter 189)

CJD = Creutzfeldt-Jakob disease; CNS = central nervous system.

TABLE 415-2 EVALUATION OF RAPIDLY PROGRESSIVE DEMENTIA

Initial screening	**Serum tests:** glucose, sodium, calcium, blood urea nitrogen, creatinine, hepatic aminotransferases, albumin, prothrombin time, TSH, antinuclear antigen, vitamin B_{12}, HIV and syphilis serology **Imaging:** brain MRI **CSF:** glucose, protein, cell counts, VDRL
Further tests to consider	**Serum:** antibodies against thyroglobulin, thyroid peroxidase, voltage-gated potassium channel, Hu (ANNA-1) **CSF:** cytology, flow cytometry **Brain biopsy**
Test findings supporting a diagnosis of prion disease	**MRI:** T2 hyperintensity in the basal ganglia, sometimes in the cortex **CSF 14-3-3 protein:** elevated levels fairly specific for CJD **EEG:** Periodic sharp wave complexes

CJD = Creutzfeldt-Jakob disease; CSF = cerebrospinal fluid; EEG = electroencephalography; HIV = human immunodeficiency virus; MRI = magnetic resonance imaging; TSH = thyroid-stimulating hormone; VDRL = Venereal Disease Research Laboratory test.

and sometimes the cortical gray matter on certain magnetic resonance imaging sequences (T2-weighted, fluid-attenuated inversion recovery, and diffusion-weighted weighting) occurs in about two thirds of CJD cases. The electroencephalogram in patients with CJD may show a pattern of periodic large-amplitude triphasic complexes. New methods for amplifying prions *in vitro* may enable clinicians to diagnose prion disease from urine,[13] nasal brushings,[14] and other tissues.

The definite diagnosis of prion disease can be made by brain biopsy. National or regional specialized prion disease centers, such as the National Prion Disorders Pathology Service Center (available at http://www.cjdsurveillance.com) in the United States, can assist pathologists in tissue analysis. In patients with a family history of neurodegenerative disease consistent with prion disease, determining the sequence of the protein-coding region of the prion protein gene can be diagnostic if a mutation is found.

TREATMENT AND PROGNOSIS

Rx

Prion diseases are incurable, and no treatment significantly improves the course of disease. Most patients with sporadic CJD die within a year of the onset of symptoms. Patients with Gerstmann-Straüssler-Scheinker syndrome and certain other genetic or variant forms of prion disease may live longer. Excellent animal models of prion disease exist, and a number of novel therapeutic approaches are under active investigation. It is worthwhile to consider enrollment in experimental clinical trials if they are available (http://clinicaltrials.gov).

PREVENTION

Most cases of prion disease occur sporadically and cannot be prevented. Genetic cases can potentially be prevented through genetic counseling and prenatal testing, although whether such measures are warranted to prevent a disease that may not manifest until midlife or later is an ethically complex question. Infectiously transmitted cases are currently amenable to preventive measures, including avoidance of surgical transmission as a result of contaminated instruments or tissue grafts and protection of the human food supply from meat products contaminated with BSE or other ruminant prions. Chronic wasting disease is epidemic among deer and elk in certain regions of the United States, and scrapie, which affects sheep and goats, is endemic at low levels in the United States and many other countries. Neither of these prion diseases has been convincingly linked to human illness, but prudence dictates that humans should avoid eating any prion-infected animal.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Liberski PP, Sikorska B, Lindenbaum S, et al. Kuru: genes, cannibals and neuropathology. *J Neuropathol Exp Neurol*. 2012;71:92-103.
2. Latest NCJDRSU CJD Monthly Statistics. *Natl CJD Res Surveill Unit Data Reports*. Available at: http://www.cjd.ed.ac.uk/data.html. Accessed March 14, 2015.
3. Brown P, Brandel JP, Sato T, et al. Iatrogenic Creutzfeldt-Jakob disease, final assessment. *Emerg Infect Dis*. 2012;18:901-907.
4. Prusiner SB. Cell biology. A unifying role for prions in neurodegenerative diseases. *Science*. 2012;336:1511-1513.
5. Mastrianni JA. The genetics of prion diseases. *Genet Med*. 2010;12:187-195.
6. Thompson A, MacKay A, Rudge P, et al. Behavioral and psychiatric symptoms in prion disease. *Am J Psychiatry*. 2014;171:265-274.
7. Zou WQ, Puoti G, Xiao X, et al. Variably protease-sensitive prionopathy: a new sporadic disease of the prion protein. *Ann Neurol*. 2010;68:162-172.
8. Mead S, Gandhi S, Beck J, et al. A novel prion disease associated with diarrhea and autonomic neuropathy. *N Engl J Med*. 2013;369:1904-1914.
9. Paterson RW, Torres-Chae CC, Kuo AL, et al. Differential diagnosis of Jakob-Creutzfeldt disease. *Arch Neurol*. 2012;69:1578-1582.
10. Muayqil T, Gronseth G, Camicioli R. Evidence-based guideline: diagnostic accuracy of CSF 14-3-3 protein in sporadic Creutzfeldt-Jakob disease: report of the guideline development subcommittee of the American Academy of Neurology. *Neurology*. 2012;79:1499-1506.
11. Skillback T, Rosen C, Asztely F, et al. Diagnostic performance of cerebrospinal fluid total tau and phosphorylated tau in Creutzfeldt-Jakob disease: results from the Swedish mortality registry. *JAMA Neurol*. 2014;71:476-483.
12. Jackson GS, Burk-Rafel J, Edgeworth JA, et al. Population screening for variant Creutzfeldt-Jakob disease using a novel blood test: diagnostic accuracy and feasibility study. *JAMA Neurol*. 2014; 71:421-428.
13. Moda F, Gambetti P, Notari S, et al. Prions in the urine of patients with variant Creutzfeldt-Jakob disease. *N Engl J* Med. 2014;371:530-539.
14. Orrú CD, Bongianni M, Tonoli G, et al. A test for Creutzfeldt-Jakob disease using nasal brushings. *N Engl J Med*. 2014;371:519-529.

REVIEW QUESTIONS

1. A 65-year-old man from the U.S. state of Kansas develops a rapidly progressive dementia and undergoes a brain biopsy. He is found to have Creutzfeldt-Jakob disease. He is a rancher and is a fan of steak tartare (raw beef), which he has eaten several times each month for many decades. There is no family history of dementia. In order of probability, the most likely cause of his disease is:

A. Sporadic, genetic, dietary exposure
B. Dietary exposure, sporadic, genetic
C. Genetic, dietary exposure, sporadic
D. Sporadic, dietary exposure, genetic
E. *Mycobacterium bovis*, dietary exposure, sporadic

Answer: A Sporadic, genetic, dietary exposure. Virtually all cases of Creutzfeldt-Jakob disease are either sporadic (90%) or genetic (10%). Consumption of beef can cause CJD if the beef is contaminated with prions, but epidemic bovine spongiform encephalopathy (BSE) has not occurred in the United States. Interestingly, it appears that very rare sporadic cases of BSE do occur in U.S. cattle, but it is highly unlikely that the described patient would have acquired CJD through his diet for the following reasons: (1) In the BSE/vCJD epidemic in Great Britain, there was only about 1 human vCJD case per 1000 diagnosed BSE cases in cattle; (2) the rate of sporadic BSE in cattle is probably on the order of 1 per million or less; (3) consumption of raw beef is not a particularly risky behavior with regard to prion disease, because cooking does not significantly inactivate prions if they are present in meat. *M. bovis* can cause an encephalitis, but the histopathologic diagnosis of prion disease is highly specific, so a misdiagnosis on brain biopsy would not be likely.

2. A 45-year-old man develops progressive confusion and gait ataxia over a period of 8 weeks. He was previously entirely well and has no history of substance abuse. On examination, he is alert but oriented only to name and year. He scores 10/30 on the Folstein mini–mental status exam. There is myoclonus of the limbs, diffuse hyperreflexia, and an unstable gait. Cerebrospinal fluid (CSF) is abnormal with a protein of 85 mg/mL. An assay for the 14-3-3 protein in CSF is positive. Which of the following further diagnostic tests would be most appropriate to obtain next?

A. Brain biopsy
B. Levels of serum antibodies against thyroid peroxidase and thyroglobulin
C. An electroencephalogram (EEG)
D. Positron emission tomography (PET) of the brain using fluorodeoxyglucose

Answer: B The patient presents with a rapidly progressive dementia. All the tests listed might have some role in the evaluation of such a patient, but the best approach is to look for treatable causes of the condition. Hashimoto encephalopathy is an autoimmune condition associated with high levels of serum antibodies to thyroglobulin or thyroid peroxidase antigens. Importantly, it can mimic several features of CJD, including causing myoclonus, elevated levels of 14-3-3 protein in the CSF, and periodic sharp waves on the EEG. Brain biopsy should be considered, but the procedure is relatively morbid and a specific diagnosis is not often obtained, so it would not usually be the next test performed. PET scan is not particularly useful in the evaluation of most cases of rapidly progressive dementia.

3. A 50-year-old woman develops signs of depression, followed 2 months later by memory problems. Her family history is notable for her father dying at age 60 of a rapidly progressive dementia that was found at autopsy to be Creutzfeldt-Jakob disease. Her paternal grandfather also died of a rapidly progressive dementia at age 43. On examination, she appears anxious and depressed. She has modest cognitive impairment, scoring 25/30 on the mini–mental status examination. The remainder of her neurologic exam is normal. Which of the following diagnostic approaches would be most appropriate?

A. Magnetic resonance imaging (MRI) of the brain, looking for increased FLAIR signal in the thalamus
B. EEG
C. Neuropsychiatric testing
D. Genetic testing to determine the sequence of the *PRNP* open reading frame
E. A lumbar puncture, with CSF assays for cell count, protein, glucose, and 14-3-3 levels

Answer: D All of the listed tests might be useful in the evaluation of this patient. However, the family history strongly points to inherited prion disease. Because the patient's father was affected, she has a 50% chance of carrying a mutation in *PRNP* and a high likelihood (generally 90% or greater) of developing disease if she carries a mutation. Her clinical findings are consistent with the initial manifestations of prion disease, although other diagnoses, such as depression and pseudodementia, should be considered. Finding a *PRNP* mutation greatly increases the likelihood that she has early prion disease, whereas negative testing almost excludes this possibility. (Outside of cases cannibalism with Kuru, prion disease is never transmitted infectiously in families). MRI, EEG, or CSF 14-3-3 testing are neither sufficiently sensitive nor specific to diagnose or exclude prion disease, and the sensitivity of these tests is particularly poor in genetic cases.

4. Which of the following increase the risk of developing prion disease?

A. Receiving a blood transfusion from a donor who developed variant CJD 1 year after donating
B. Being treated with growth hormone between 1995 and 1998
C. Eating sheep eyes
D. Caring for a hospitalized patient with Creutzfeldt-Jakob disease
E. Receiving a blood transfusion from a donor who developed sporadic CJD 1 year after donating

Answer: A Blood transfusions have been reported to transmit variant CJD, the form of human prion disease caused by eating meat from cattle infected with BSE prions, to three recipients. In contrast, there is no evidence the more common sporadic form of prion disease can be transmitted by blood or blood products. Growth hormone derived from cadaveric human pituitary is implicated in more than 200 cases of CJD in recipients, but the recombinant form of the protein that has been used since the late 1980s has never transmitted prion disease. It used to be thought that the high rate of CJD among a group of Jews originating from Libya was caused by eating sheep eyes, but it is now known that a genetic form of prion disease caused by a mutation changing glutamate to lysine at codon 200 of *PRNPI* is prevalent in this population. There is no evidence for the transmission of prion disease from patients to health care providers.

416

NUTRITIONAL AND ALCOHOL-RELATED NEUROLOGIC DISORDERS

BARBARA S. KOPPEL

An adequate supply of vitamins and minerals is necessary for embryonic and early development as well as subsequent maintenance of metabolic function of both the central and peripheral nervous systems. Deficiencies of vitamins and minerals can cause a variety of neurologic syndromes (Table 416-1), each with well-described constellations of symptoms that are dependent on the location of the resulting pathologic changes within the nervous system.[1]

Vitamin deficiency (Chapter 218) can be caused by either malnutrition (Chapter 215) or malabsorption (Chapter 140). In addition to the insufficient intake of needed elements, functional deficiency can also result from increased demand owing to sepsis, chronic inflammatory conditions, dialysis, and problems incorporating the element or delivering it to its site of action. Exposure to alcohol or other neurotoxins in the setting of certain vitamin deficiencies (usually B_1) synergistically contributes to neuropathology. In malnourished patients, multiple simultaneous vitamin deficiencies can result in complex symptoms owing to overlapping areas of nervous system involvement. Malnutrition is the most common cause of vitamin deficiency in economically disadvantaged countries. Overdependence on single food sources that may be neurotoxic (e.g., cassava, grass peas, spoiled grains) may complicate the presentation. Another example of this phenomenon is the alcoholic (Chapter 33) who obtains calories from alcohol, which is neurotoxic, and fails to take in micronutrients and vitamins. Even when adequate food supplies are readily available, malnutrition may be caused by inadequate consumption due to mechanical obstruction from cancer of the mouth or gastrointestinal tract, unbalanced ("fad") diets, fasting, anorexia, chronic nausea, or recurrent or persistent vomiting. Rarely, deficiencies arise in patients with genetic diseases that block intestinal absorption, transport across the blood-brain barrier, or uptake into mitochondria, neurons, and glia.

Iatrogenic causes include failure to feed patients who are comatose, not self-sufficient (from dementia, brain injury, psychiatric illness), or dysphagic, as can occur following a stroke or spinal cord injury. Vitamin deficiency can also result from failure to include adequate amounts of vitamin and mineral supplements in parenteral or liquid enteral diets.

As bariatric surgery becomes more common in the treatment of morbid obesity (Chapter 220), it has become a common cause of neurologic disorders associated with many vitamin deficiencies. Patients who have undergone restrictive ("lap band," gastric stapling) or especially bypass procedures require lifelong, not just perioperative, supplementation and monitoring of vitamin levels.

DEFICIENCY OF WATER-SOLUBLE VITAMINS

Thiamine (Vitamin B_1) Deficiency

Thiamine is converted to thiamine pyrophosphate, which serves as a coenzyme in glucose and lipid metabolism and in the synthesis of neurotransmitters from branched-chain amino acids (Chapter 218). To avoid deficiency, thiamine must be consumed regularly in adequate amounts, at least 0.33 mg per 1000 calories, or about 1 mg/day (more during pregnancy or lactation). Food sources include whole grains, legumes, meat, and fortified bread or cereals.

BERI-BERI

In developing countries, the most common manifestation of thiamine deficiency is beri-beri, which is characterized by a peripheral sensorimotor axonal neuropathy with numbness, paresthesias, or burning pain, occasionally accompanied by heart failure ("wet beri-beri"). Other causes of thiamine deficiency include reliance on foods in which the vitamin has been inactivated by processing (e.g., polished rice), overcooking, or eating foods that contain thiaminase-producing bacteria (e.g., raw fish).

WERNICKE ENCEPHALOPATHY

Even short-term (<3 weeks) thiamine deficiency can result in Wernicke encephalopathy, a syndrome characterized by insidious development and progression (over days to weeks) of confusion or delirium, abnormal eye movements, and ataxia.[2] Fewer than one third of patients develop all three elements of this triad. Wernicke encephalopathy occurs most often in the setting of poor nutrition and prolonged vomiting in patients with chronic alcohol abuse. Based on pathologic changes discovered at autopsy, about 12% of heavy alcohol users and 60% of patients dying of alcohol-related causes have Wernicke encephalopathy. The symptoms and signs of Wernicke encephalopathy reflect the preferential dysfunction of brain regions that have a high demand for thiamine, a cofactor in energy-producing cycles. These areas include the blood-brain barrier, anterior and centromedian thalamus, mammillary bodies, periaqueductal gray matter, superior and inferior colliculi, and floor of the fourth ventricle. The most common pathologic changes in these regions include neuronal swelling and microscopic hemorrhages, followed by gliosis. Rarely, the cerebral cortex and hypothalamus may be involved as well. Deficiency of α-ketoglutamate dehydrogenase activity in astrocytes leads to microglial activation and glutamatergic toxicity.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The full triad of mental status change, abnormal eye movements, and ataxia occurs in only about one third of cases. Acute symptoms may be provoked if intravenous (IV) glucose or food is given before thiamine has been replaced. Because a medical history may be unobtainable until the patient's confusion clears, physical signs of chronic alcoholism (e.g., gynecomastia, skin angiomata, pulmonary erythema, ascites, jaundice) (Chapters 146 and 152) must be sought.

TABLE 416-1 SUMMARY OF VITAMIN AND MINERAL DEFICIENCIES

VITAMIN AND MINERAL DEFICIENCIES	NEUROLOGIC SYNDROME(S)	SUPPORTING TESTS	TREATMENT	CAUSES (OTHER THAN MALNUTRITION)
A (Retinol)	Blindness from retinal or corneal damage	Visual fields, visual acuity Serum level < 30-65 µg/dL	30,000 IU vitamin A daily × 1 wk	Hypothyroidism, diabetes, renal or liver failure
B_1 (thiamine)	Wernicke encephalopathy: ataxia, nystagmus, ophthalmoparesis, confusion, delirium Korsakoff syndrome: amnesia, confabulation Beri-beri: axonal neuropathy	MRI: symmetrical lesions of midbrain (periaqueductal area), pons, hypothalamus, thalamus, cerebellum MRI: necrosis of mamillary bodies, dorsomedial and anterior thalamus Nerve conduction tests: decreased amplitude Serum thiamine level < 20 ng/dL Decreased erythrocyte transketolase	Prevent by 100 mg PO daily before and 1 year after bariatric surgery, 100 mg IV before glucose administration or refeeding after starvation Treat Wernicke encephalopathy with 5 days of thiamine, 100-500 mg IV or IM daily, until improvement stabilizes, then PO 100 mg daily Antioxidants (*N*-acetylcysteine)	Alcoholism, bariatric or other major GI surgery, prolonged vomiting, hemodialysis, diuretic treatment of heart failure, cachexia, 5-fluorouracil, other blockers of thiamine phosphate production
B_3 (niacin)	Pellagra: confusion, dementia, weakness, ataxia, spasticity, myoclonus, glossitis, dermatitis, photosensitivity	Erythrocyte NAD, plasma niacin, urinary N1-methylnicotinamide	Nicotinic acid, 50 mg PO tid or 25 mg IV tid; nicotinamide, 50-100 mg IM or PO tid	Alcoholism, corn- or cereal-based diet, Hartnup syndrome, carcinoid syndrome
B_5 (pantothenic acid)	Dysesthesias, foot paresthesias	Deficient coenzyme A	5 mg PO daily	Severe malnutrition
B_6 (pyridoxine)	Neuropathy, sensory ataxia, depression Infantile pyridoxine-deficient epilepsy	Plasma PLP < 27 nmol/L; urinary 4-pyridoxic acid, < 3 nmol ↑ Homocysteine after methionine loading challenge ↑ α-AASA in urine, plasma, CSF	50-100 mg PO daily for neuropathy (preventive use if taking B_6 antagonist) 100-200 mg daily for adult epilepsy	Diverticulosis, isoniazid, cycloserine, other antagonists Genetic defects in antiquitin (aldehyde dehydrogenase), pyridoxal synthesis
B_{12} (cobalamin)	Myelopathy with spastic paraparesis and sensory ataxia, peripheral neuropathy, optic neuropathy, memory loss, dementia; indirect contributor to stroke	Blood level < 200 pg/mL ↑ Methylmalonic acid > 145 nmol/L Intrinsic factor antibodies Schilling test, megaloblastic anemia Delayed somatosensory evoked potentials ↑ Homocysteine, total > 12.5 µmol/L	IM B_{12}, 1000 µg daily for 1 week, then weekly for 1 month, then monthly; or oral B_{12}, 1000 µg daily; or nasal B_{12}, 500 µg weekly for lifetime if abnormal absorption, 50-100 µg daily if normal absorption	Achlorhydria, gastric or ileal resection, blind loop syndrome, sprue, HIV infection, nitrous oxide anesthesia (especially abuse), fish tapeworm, vegan diet
D (calciferol)	Proximal myopathy, often painful; cognitive impairment Secondary compression of spinal cord, plexus, or peripheral nerves from rickets or osteomalacia	25-(OH) vitamin D_3 level < 10 ng/mL in urine Serum calcium ↑ PTH > 54 pg/mL Osteopenia/porosis on bone densitometry	Daily supplementation with 400 IU, >50,000 IU 3 times per wk if malabsorption; use blood level or urine calcium excretion to guide (should be > 100 mg/day)	Lack of exposure to sunlight, including sunblock protection; chronic antiepileptic drug use
E (tocopherol)	Spinal and cerebellar ataxia, Babinski sign, ophthalmoplegia, peripheral neuropathy, retinitis pigmentosa	Vitamin E level < 2.5 mg/L (normal, 6-15 with normal lipid level) ↑ A-β-lipoprotein levels, antigliadin antibodies Genetic analysis to rule out other spinocerebellar ataxias such as Friedreich ataxia	Supplement with 6-800 IU, 5-10 mg/kg twice daily, for ataxia of genetic causes, water-soluble 200 mg/kg/day or IM α-tocopherol for malabsorption	Biliary atresia, celiac sprue, Genetic: ↓ α-tocopherol transport protein (8q13), microsomal triglyceride transfer protein
Folate	Dementia, B_{12} deficiency, stroke	↑ Homocysteine, plasma level < 2.5 µg/L	1 mg 3 times daily until normal level, then maintenance of 1 mg/day Pregnancy: additional 0.4 mg/day if taking a folate antagonist	Malabsorption or use of antagonist (methotrexate) or antiepileptic medication
K (phytonadione)	Intracranial hemorrhage	INR or PT elevation	IM phytonadione at birth, maternal vitamin K for last month of pregnancy	Medication use that increases metabolism (e.g., phenytoin)
Copper	Myelopathy, neuropathy	Serum Cu < 75 µg/dL, ↓ urinary Cu, ceruloplasmin < 23 mg/dL MRI: ↑ T2 signal in cervical cord, dorsal column Mutation in *ATP7A* gene (Menkes disease)	Elemental Cu, 8 mg/day PO week 1, 6 mg/day week 2, 4 mg/day week 3, 2 mg/day ongoing malabsorption Menkes disease: 250 mg SC bid	Wilson disease, Menkes disease, alcoholism, malabsorption, gastric bypass, zinc toxicity
Magnesium	Seizures, encephalopathy	Serum magnesium < 1.5 mg/dL, correct for low albumin	Magnesium sulfate IV or PO Avoid magnesium-wasting drugs	Alcoholism, especially beer
Potassium	Muscle weakness, chronic, acute	Serum potassium < 3.5 mEq/L, ECG	IV or PO KCl until normalized	Diuretic use, bulimia

AASA = aminoadipic semialdehyde; CSF = cerebrospinal fluid; ECG, electrocardiography; GI = gastrointestinal; INR = international normalized ratio; MRI = magnetic resonance imaging; NAD = nicotinamide adenine dinucleotide; PLP = pyridoxal-5-phosphate (active coenzyme of pyridoxine); PT = prothrombin time; PTH = parathyroid hormone.

Mental status changes range from mild memory impairment or inattention to delirium, often with apathy or abulia. Eye movement abnormalities include nystagmus, dysconjugate gaze, and gaze palsies[3] (Video 416-1). Ataxia can affect the limbs (legs more than arms), trunk, and gait. Patients with Wernicke encephalopathy can also have autonomic and hypothalamic dysfunction, with bradycardia and hypothermia as well as papilledema, optic neuropathy, seizures, and myoclonus.

In symptomatic patients, T2-weighted magnetic resonance imaging (MRI) can be normal but often demonstrates symmetrical increased signal due to edema or hemorrhage in affected areas, most often periventricular thalamus,

VIDEO 416-1. Wernicke encephalopathy eye movements. A 49-year-old heavy alcohol user admitted in confusional state with nystagmus and lateral rectus (Video 1A), which resolved 5 days later after receiving 100 mg intravenous thiamine daily (Video 1B).

periaqueductal regions in the floor of the fourth ventricle or cerebellum, and in the mamillary bodies. Low thiamine levels (<50 mg/mL) are common, although levels may be normal in about 10% of cases. Because thiamine deficiency disrupts carbohydrate metabolism, serum levels of lactate and pyruvate can be elevated.

KORSAKOFF SYNDROME

Korsakoff syndrome becomes apparent in up to 80% of patients who survive Wernicke encephalopathy. It is more likely to follow Wernicke encephalopathy in the setting of alcoholism than in nutritional deficiency alone, thereby implying a synergistic mechanism that may be due to repeated episodes of alcohol withdrawal with associated glutamate neurotoxicity, compounded by lack of thiamine. The primary pathologic findings occur in the limbic system, especially the mamillary bodies, amygdala, and dorsomedial and anterior thalamus. Cortical involvement may be related to alcohol neurotoxicity rather than thiamine deficiency.

As confusion and delirium of Wernicke syndrome improve, an amnestic state in which patients are often unaware of their memory impairment becomes apparent. Korsakoff syndrome can be reliably identified only following resolution of acute delirium and global confusional states. It is characterized by disproportionate retrograde and anterograde episodic amnesia, transient confabulation, and hallucinations. Occasionally, Korsakoff psychosis is present clinically or pathologically without documented episodes of Wernicke encephalopathy, perhaps because Wernicke encephalopathy was subclinical or was not recognized acutely.

The memory deficit, which precludes learning new information or acquisition of new memories, is disproportionately severe in relation to other aspects of cognitive function. For example, alertness, attention, social interactions, and motor learning (procedural memory) are generally well preserved. There may be mild disorientation with respect to time and place, and sometimes apathy and other emotional changes are present. Confabulation, in which the intrusion of errors in response to questions leads to fabrication of answers without the intention to deceive, is sometimes present spontaneously in the first weeks after Wernicke encephalopathy. Confabulation may be a compensatory mechanism, and it usually lessens over time. Neuropsychological testing frequently demonstrates emotional changes and mild problems in executive function, which are indicative of frontal lobe involvement.

TREATMENT AND PROGNOSIS

Rx

Untreated, Wernicke encephalopathy is fatal in 90% of cases. However, timely thiamine replacement (see Table 416-1) can prevent or treat Wernicke encephalopathy as well as beri-beri. In the acute setting, high-dose IV or IM thiamine, which is recommended to circumvent any problems with swallowing or absorption, will lead rapidly—often within hours—to complete resolution of nystagmus and oculomotor paresis, followed by resolution of the ataxia and eventually of the mental status changes attributable to thiamine deficiency. However, many alcoholic patients may have residual ataxia and cognitive impairment, including memory dysfunction, owing to the toxic effects of alcohol itself (see later).[4] Because Korsakoff syndrome does not respond to thiamine replacement, prevention by timely recognition and treatment of Wernicke encephalopathy is essential. Magnesium (Chapter 119) must also be replaced if deficient.

Cobalamin (Vitamin B_{12}) Deficiency

Cobalamin is involved in methionine pathways that regulate myelination during development and maintain myelin throughout life (Chapter 218). Deficiency results in combined system disease (peripheral neuropathy and spinal cord degeneration) or subacute combined degeneration of the dorsal (sensory) and lateral (motor) tracts (i.e., myelopathy).[5] The spinal cord tracts that are dysfunctional result in impaired position and vibratory sensation and spastic paraparesis.

Cobalamin deficiency (Chapters 164 and 218) is most common in individuals older than 60, because the incidence of atrophic gastritis (Chapter 139) and achlorhydria rises in older individuals, and acid is required for B_{12} processing. Long-term use of proton pump inhibitors may play a role in development of B_{12} deficiency, as does a lack of gastric intrinsic factor needed for absorption of vitamin B_{12}. Cobalamin deficiency is rarely due to inadequate dietary intake (e.g., a vegan diet for several years), because it is stored in fat. A more common cause in recent years is bypass surgery for weight loss. Nitrous oxide ("laughing gas") toxicity, usually from illicit use rather than administration as an anesthetic, can cause cobalamin deficiency by inactivating the cobalamin-dependent enzyme methionine synthase. Long-term treatment of diabetes with metformin also can lower B_{12} levels. Low vitamin B_{12} levels have been associated with increased homocysteine levels, but a relationship to vascular disease or vascular dementia has not been established.

CLINICAL MANIFESTATIONS

Demyelination of the dorsal columns causes proprioceptive loss that can result in sensory ataxia owing to loss of position sense in the feet.[6] Romberg sign (failure to maintain balance with the eyes closed) distinguishes sensory from cerebellar ataxia. An axonal peripheral neuropathy with numbness and tingling in the hands and feet is almost always present. Motor function eventually becomes impaired as well. The optic nerve can be involved, and vagal neuropathy has been reported. Signs of cerebral involvement include memory loss, personality changes, and occasionally hallucinations and psychosis. Encephalopathy and dementia may be present, but B_{12} deficiency may be a secondary phenomenon in a patient with another cause of memory impairment, or both conditions may coexist without a causative relationship. Neurologic abnormalities may be present without anemia, especially because the anemia can be corrected by high-dose folate replacement. Symptoms generally progress slowly, but they can appear rapidly after exposure to nitrous oxide anesthesia in individuals with preexisting subclinical cobalamin deficiency.

DIAGNOSIS

Serum vitamin B_{12} levels are usually low (<300 pg/mL) but can rarely be normal in symptomatic patients. In such cases, serum levels of methylmalonic acid and homocysteine are useful ancillary tests because these levels are increased as a result of impaired cobalamin-dependent reactions. Pernicious anemia (Chapter 164) is severe in about 20% of patients. However, both the hematocrit and mean corpuscular volume are sometimes normal because the hematologic effects of cobalamin deficiency can be partially masked by folate supplementation.

Low levels of cobalamin are sometimes present in normal people, especially the elderly, in which dementia, peripheral polyneuropathy, and myelopathy may be due to a myriad of causes. Therefore, a low cobalamin level may reflect poor nutrition or absorption rather than being the cause of these conditions. Causality is definitively confirmed by clinical improvement after cobalamin replacement, which usually begins after several weeks and may continue for up to a year.

TREATMENT

Rx

Treatment usually begins with a subcutaneous or intramuscular (IM) injection of 500 to 1000 µg of cobalamin daily for 1 week and then weekly for 1 month. After that time, oral supplementation with 50 to 100 µg daily of cyanocobalamin usually suffices in patients with achlorhydria or other causes of malabsorption; 1000 µg daily should be used in patients with intrinsic factor antibodies. Sublingual, transdermal patch, and nasal gel forms (500 µg weekly) can also be effective.

PROGNOSIS

Neurologic symptoms, especially paresthesias, typically improve to some extent within 3 months of achieving adequate B_{12} serum levels. Numbness and areflexia often persist, especially if treatment is delayed. If there is no improvement whatsoever, vitamin B_{12} deficiency is unlikely to be the cause of the condition. For example, human immunodeficiency virus–associated myelopathy, which can have a similar clinical manifestation, does not reverse with supplemental vitamin B_{12} because it is caused by disruption of transmethylation pathways, not low cobalamin levels (see Table 416-1).

Folate Deficiency

Folate is an important coenzyme in the metabolism of nucleic and amino acids (Chapter 218). Folate deficiency is an important risk factor for neural tube defects in utero. In Cuban adults, especially alcoholic men, a crop failure led to semistarvation and an epidemic of blindness that resolved with folate and B vitamin supplementation. Folate deficiency also results in elevated levels of homocysteine, which is associated with an increased risk for ischemic heart disease and stroke. Folate and B vitamin supplementation in patients with elevated homocysteine levels but without classic homocysteinemia did not reduce adverse vascular events in clinical trials. Patients with

genetic folate deficiency due to lack of methylenetetrahydrofolate reductase, which converts ingested folate to the active metabolic cofactor, have an increased risk of intracerebral hemorrhage.

The supplementation of flour with folate has greatly reduced the risk of folate deficiency, and all pregnant women are now prescribed supplemental folate. Women of child-bearing age should be treated even before pregnancy if they have any conditions that predispose to folate deficiency.

Folate deficiency also leads to megaloblastic anemia (Chapter 164). Before correction of megaloblastic anemia with folate alone, vitamin B_{12} levels should be checked to avoid ongoing neurologic injury resulting from unrecognized cobalamin deficiency. Folate deficiency is treated with 1 mg three times daily for 1 month, followed by 1 mg daily.

Pyridoxine (Vitamin B_6) Deficiency

Pyridoxine is a coenzyme in multiple reactions that involve gluconeogenesis, biosynthesis of neurotransmitters, and the metabolism of amino acids, nucleic acids, and lipids. Pyridoxine deficiency can be caused by genetic defects, such as defective antiquitin, that lead to increased utilization of pyridoxine. In adults, low serum levels of pyridoxine are well tolerated, so symptomatic deficiency is rare. However, symptomatic deficiency can occur in the setting of renal failure (Chapter 130), dialysis (Chapter 131), or cirrhosis (Chapter 153), or with medications such as isoniazid for antitubercular therapy (Chapter 324) or hydralazine for heart failure (Chapter 59) if patients do not receive concurrent supplementation. Deficiency is also seen with extreme malnutrition, especially diets consisting predominantly of white rice.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Prolonged pyridoxine deficiency causes a painful peripheral axonal neuropathy that leads to weakness and sensory ataxia. Some patients have skin thickening, seborrheic dermatitis, or glossitis, which can suggest pellagra. Serum levels of the active form of pyridoxine, pyridoxal 5′-phosphate, and urine levels of the metabolite 4-pyridoxic acid are low. Ancillary tests include nerve conduction studies, which show significantly reduced amplitudes in sensory and motor action potentials with normal conduction velocity times, typical of an axonal neuropathy.

In epilepsy caused by pyridoxine deficiency, seizures begin in the neonatal period and may persist, along with intellectual disability. Electroencephalography shows a highly disorganized pattern with excessive slow frequency activity and abundant multifocal and generalized spikes resembling hypsarrhythmia. Pyridoxine deficiency in pregnancy can be caused by hyperemesis gravidarum and may rarely produce neurologic disease in offspring, but routine supplementation is not recommended for all pregnancies.

Toxicity due to excess pyridoxine intake (>100 mg daily) leads to a ganglioneuropathy manifested by pure sensory symptoms, including sensory loss, ataxia, areflexia, and the presence of Romberg sign. However, this syndrome, which can result from an overdosing of vitamin supplements, is less common than deficiency-related neurologic disease.

TREATMENT Rx

Patients in whom symptoms of pyridoxine deficiency develop or who take pyridoxine antagonists should receive supplemental pyridoxine (50 to 100 mg daily). Children with pyridoxine-dependent epilepsy require immediate and lifelong supplementation with 100 mg of pyridoxine daily.

For pyridoxine toxicity, simply stopping excess oral vitamin intake will eventually completely reverse the damage. The only exception is if a very large IV dose was administered, in which case neuropathy is not reversible.

DEFICIENCY OF FAT-SOLUBLE VITAMINS

Vitamin E (Tocopherol) Deficiency

Although vitamin E is composed of several tocopherols, it is the α form that is biologically active in humans and contained in most foods. Because it is so widely available, deficiency is almost never due to inadequate dietary consumption. Rather, vitamin E deficiency is almost always the result of malabsorption because of such conditions as biliary and pancreatic disease (Chapters 155 and 144), cystic fibrosis (Chapter 89), celiac disease (Chapter 140), Crohn disease (Chapter 141), extensive small bowel resection, and blind loop syndrome (Chapter 140). In addition, vitamin E deficiency is associated with genetic defects in the α-tocopherol transfer protein, hypolipoproteinemia and abetalipoproteinemia, chylomicron retention disease, and ataxia with vitamin E deficiency.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Neurologic manifestations of vitamin E deficiency include a spinocerebellar syndrome with ataxia, loss of vibration and position senses, hyporeflexia, extensor plantar responses, and rarely myopathy. Other findings can include ptosis, abducens paresis, nystagmus, and retinopathy.

Serum levels of vitamin E can vary with the serum lipid levels. In extreme hyperlipidemia, such as cholestasis, the ratio of vitamin E to cholesterol will be more reliable than the absolute vitamin E level, which can be normal despite a low ratio of vitamin E to cholesterol.

Although generally not toxic, excessive intake of vitamin E causes bleeding, including hemorrhagic infarcts, in adults, probably owing to its effects on platelet function. During pregnancy, high doses of vitamin E can interfere with oxidation in the fetus and cause growth retardation.

TREATMENT Rx

The amount of vitamin E replacement required depends on the cause of the deficiency. Malabsorption syndromes require 1000 to 2000 mg daily for infants and 10 to 20 g daily for adults. Genetic causes can be treated with 5 to 10 g/day. After bariatric surgery, supplementation with 300 mg daily is recommended. For vitamin E, 1 mg is equivalent to 1.49 IU. Although the mechanism is unknown, 2000 IU/day of α-tocopherol may slow functional decline in patients with moderate to severe Alzheimer disease.

Vitamin D (Calciferol) Deficiency

Vitamin D deficiency (Chapter 218) results from inadequate exposure to sunlight, dietary insufficiency, or malabsorption caused by celiac disease, inflammatory bowel disease, or extensive small bowel resection. The average multivitamin contains 400 IU of combined D_2 (ergocalciferol) and D_3 (cholecalciferol), whereas 20 minutes of full-body summer sun exposure provides 10,000 IU of D_3, the form used in the body. Vitamin D deficiency causes rickets in children and osteomalacia (Chapter 244) in adults. Bone remodeling may lead to compression of the spinal cord or roots (Chapter 400) owing to changes in vertebral bodies and foramina. Abrupt lack of vitamin D causes hypocalcemia with secondary hyperparathyroidism (Chapter 245). Hypocalcemia, in turn, can cause tetany, encephalopathy and generalized seizures.[7] Deficiency also causes proximal myopathy, worse in the legs than the arms, which can lead to peculiar patterns of gait as a result of weakness and fear of falling. It is associated with sleep disorders and fibromyalgia. Vitamin D plays a role in immune modulation of regulatory T cells, which may explain the relationship of vitamin D deficiency to a later risk of developing multiple sclerosis[8] (Chapter 411) and narcolepsy. Vitamin D deficiency is also associated with an increased incidence of Alzheimer and all-cause dementia.[9]

Individuals whose dietary absorption is normal require 400 to 600 IU (10 to 15 μg) of vitamin D_3 daily, but persons with malabsorption may need more than twice that amount. Reliance on sunlight is not recommended owing to risks of skin cancer (Chapter 203) as well as variability in personal environmental exposure. The eventual dose of supplementation, either weekly with 50,000 units (1.25 mg) of ergocalciferol (D_2) or daily with 400 to 800 IU of cholecalciferol (D_3), should be titrated to produce a serum 25-hydroxyvitamin D level of at least 20 ng/mL.

Toxicity from excess absorption of vitamin D, as is seen in sarcoidosis (Chapter 95) or granulomatous conditions or from excessive intake, is rare. Like deficiency states, vitamin D toxicity causes muscle and bone pain (Chapter 245).

Vitamin A Deficiency

Vitamin A deficiency, which is most often associated with retinal dysfunction ("night blindness" [Chapter 423]), can also impair the ability to taste and has rarely been associated with raised intracranial pressure in children. Malabsorption seldom leads to vitamin A deficiency, because it is stored in the body for long periods. Vitamin A toxicity, which is occasionally a complication of using isotretinoin for acne or excessive dietary consumption of liver (especially from fish-eating mammals such as polar bear, seal, and walrus), can cause idiopathic intracranial hypertension (Chapters 189 and 398) with headache and papilledema that leads to diminished vision or even blindness in severe cases.

Vitamin K Deficiency

Vitamin K deficiency, a rare consequence of malabsorption, is most often seen in patients who have compromised liver synthesis or who are taking the antagonist warfarin (Chapter 218). Vitamin K deficiency causes excessive bleeding and increases the risk of intracerebral hemorrhage (Chapter 408), especially in newborns whose mothers are taking antagonists such as phenytoin. Now that newborns are routinely given injections of vitamin K, maternal supplementation using 5 mg daily for the last month of pregnancy is reserved for women on vitamin K antagonists.

DEFICIENCY OF MISCELLANEOUS ELEMENTS AND NUTRIENTS

Copper Deficiency

Acquired copper deficiency (Chapter 218) is rare and can be difficult to recognize. It occurs most often in premature or malnourished infants and in patients with malabsorption due to celiac disease, cystic fibrosis (Chapter 89), Crohn disease (Chapter 141), or intestinal blind loops after surgery (e.g., a Whipple procedure for pancreatic cancer or other malignancy, or bypass surgery for weight loss). It can also occur in patients with nephrotic syndrome (Chapter 121) and intestinal bacterial overgrowth (Chapter 140). Copper deficiency is also a well-recognized consequence of excessive intake of zinc, which upregulates copper chelation. Excessive zinc exposure usually results from overuse of zinc-containing products, such as denture cream or herbal preparations to treat rhinitis and sinusitis, or as a consequence of parenteral overload during hemodialysis (Chapter 131).

The most common neurologic complication of copper deficiency is a myelopathy that is clinically very similar to that seen with cobalamin deficiency.[10] The most prominent features are spastic paraparesis and sensory ataxia. A peripheral polyneuropathy of the axonal type is usually present as well. Optic neuropathy, wrist drop, and foot drop have also been reported. Copper levels, including excreted urinary copper, should be measured in patients who are suspected of having vitamin B_{12} deficiency but fail to respond to cyanocobalamin replacement. Treatment consists of oral copper supplementation, 8 mg/day tapering weekly over 3 weeks, followed by maintenance of 2 mg/day for life.

Wilson disease (Chapter 211) is an autosomal recessive disorder caused by mutation in the *ATP7B* gene that encodes ceruloplasmin, a copper-transporting protein. Symptoms result from excessive copper accumulation, primarily in the liver and brain, as a result of failed transport and impaired copper excretion. Psychiatric features such as personality change, disinhibition, depression, and psychosis sometimes overshadow neurologic manifestations such as dementia, dysarthria, chorea, tremor, and dystonia. Deposition of copper in Descemet membrane causes the characteristic Kayser-Fleischer ring, which is seen in the iris in over 95% of patients (Chapter 211, Fig. 211-2). Serum ceruloplasmin levels are low. Accumulation of copper in the liver leads to chronic liver failure. Wilson disease is treated by copper chelation (Chapter 211) and by minimizing dietary intake.

Menkes disease, an X-linked recessive copper deficiency caused by mutations in the *ATP7A* gene needed for absorption, is characterized by severe intellectual disability and kinky hair. The diagnosis is made by finding low serum copper levels or changes in the ratio of dopamine to norepinephrine. Large doses of copper histidine (250 mg twice daily until age 1, then daily until age 3) must be delivered subcutaneously, but improvement is variable. Patients survive to adulthood only if copper injections are begun in the neonatal period.

Other Nutritional Disorders

Biotin deficiency is caused by lack of protein in the diet, unsupplemented total parenteral nutrition, and an autosomal recessive disorder affecting biotinidase, which prevents biotin from being accessible for use. Genetic causes result in developmental delay, seizures, ataxia, and deafness if supplementation is not begun in the newborn period. Dietary deficiencies lead to lethargy, myalgias, and paresthesias, along with rash.

Some epidemics of peripheral neuropathy or optic neuropathy, such as Strachan Jamaican neuropathy and Cuban tobacco-alcohol amblyopia, which have occurred in the setting of malnutrition or overdependence on one source of food, have responded to replacement of B vitamins or folate and are therefore presumed to be due to deficiencies of these nutrients. Cyanide poisoning (Chapter 110) from smoking or consumption of sugarcane may contribute to toxicity. Iodine deficiency leads to hypothyroidism (Chapter 226), which causes varying degrees of cretinism.

Overreliance on one hardy food source (grass peas, chick peas) may lead to spastic paraparesis from lathyrism owing to oxalyldiaminoproprionic acid, which is a neurotoxic glutamate agonist. In older Nigerian men, cassava tuberis (konzo), which can potentiate neurotoxicity from cyanate and glucoside (Chapter 110), causes sensory neuropathy, ataxia, optic atrophy, and sensorineural deafness. In contrast, women and children develop spasticity, presumably from the same toxin. Amyotrophic lateral sclerosis and Parkinson-dementia complex in Guam are probably caused by cycad toxins in flour.

ALCOHOL-RELATED DISORDERS

Alcohol (Chapter 33) is responsible for a wide spectrum of neurologic disorders. At one extreme, it can cause irreversible dementia, cerebellar degeneration, optic neuropathy, and peripheral polyneuropathy. After chronic overuse, withdrawal from alcohol causes transient neurologic overexcitation syndromes such as seizures, tremors, hallucinosis, and delirium tremens. Acute intoxication can range from mild euphoria to vestibular and cerebellar dysfunction to coma and death. It can also contribute to falls, accidents, behavioral dyscontrol with violence, and subsequent trauma. When alcohol is the main source of calories, it contributes to nutritional deficiency syndromes such as Wernicke-Korsakoff syndrome, beri-beri, and pellagra. Alcohol-induced liver toxicity results in hepatic encephalopathy and non-Wilsonian hepatolenticular degeneration. Coagulopathy caused by liver disease or suppressed platelet production raises the risk for subdural or intracranial hematoma (Fig. 416-1). Fetal alcohol syndrome reflects the vulnerability of the developing nervous system to the toxic effects of alcohol.

Signs of intoxication correlate with blood alcohol levels: 50 mg/dL for personality changes; 150 mg/dL for ataxia, vestibular dysfunction, and nystagmus; 300 mg/dL for stupor; 400 mg/dL for coma; and up to 500 mg/dL for respiratory depression or apnea. However, the effects vary greatly depending on the chronicity of intake and the rate at which high levels develop. Intoxication alone should never be assumed to be the sole cause of a depressed mental state, because alcoholics are at increased risk for other causes of coma (Chapter 404).

Specific Clinical Syndromes

Seizures and status epilepticus (Chapter 403) can be a direct consequence of intoxication, withdrawal, hyponatremia (Chapter 116), and hypomagnesemia (Chapter 119), or they can result from epileptogenic foci owing to previous head trauma or stroke. Even in patients with delirium tremens, other causes of seizures should be investigated and treated appropriately.

Hepatic encephalopathy is most often seen in patients with alcoholic cirrhosis (Chapter 153), especially patients who have bleeding esophageal varices. It is characterized by irritability alternating with depressed mental status, seizures, tremor, and asterixis. Although temporary reversal of encephalopathy with flumazenil (2 mg IV) can confirm the diagnosis, treatment focuses on trying to reduce the serum ammonia level, usually with lactulose (15 to 30 mL orally twice daily) and nonabsorbable antibiotics such as rifaximin (550 mg twice daily), neomycin (500 mg to 1 g three times daily), or metronidazole (250 mg two to four times daily).[11] Closure of spontaneous portosystemic shunts by embolization may be effective for reducing intractable encephalopathy in some patients.[12]

Dementia develops even when nutrition is well maintained because of alcohol's direct neurotoxic effects, although the absolute amounts of alcohol necessary to produce dementia are unclear.[13] In younger patients with dementia, 10 to 25% of cases are attributed to alcohol. In the elderly, excessive alcohol consumption is associated with faster cognitive decline compared with light to moderate alcohol consumption.[14] Frontal lobe dysfunction results in executive dysfunction (planning, abstract reasoning) rather than the amnesia that is prominent in Wernicke-Korsakoff syndrome. Contributing factors include head injury, status epilepticus, and cerebrovascular disease. The appearance of cerebral atrophy is further evidence of alcohol's deleterious effect involving both gray and white matter. The dementia is probably not reversible, although case reports credit the NMDA receptor antagonist memantine (28 mg XR daily) with improving scores on the mini–mental status examination, and small studies have reported improved cognition in patients taking rivastigmine (306 mg twice daily), an acetylcholinesterase inhibitor.

Wernicke encephalopathy and Korsakoff syndrome are described in the section on thiamine.

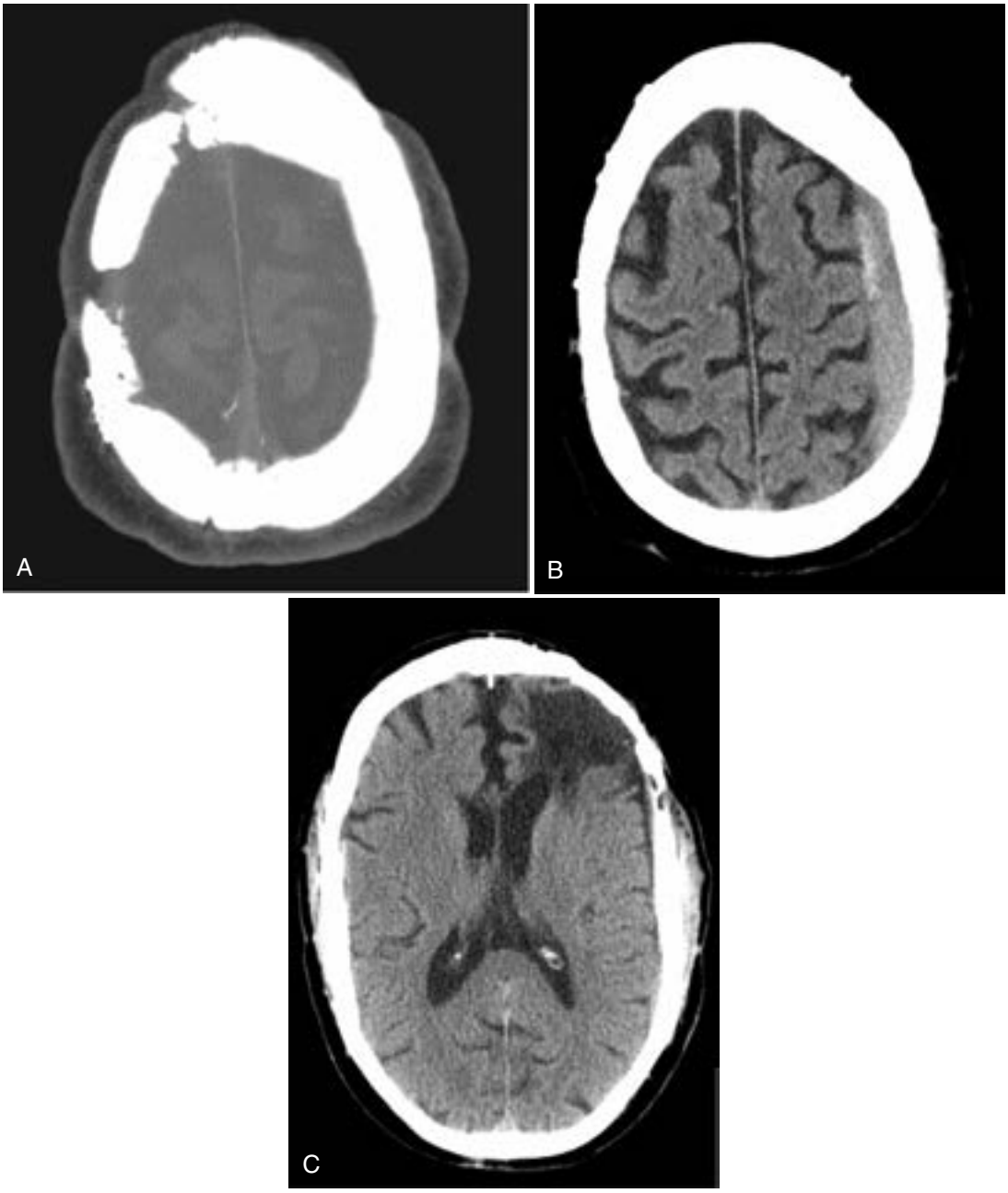

FIGURE 416-1. Alcoholic man admitted after a seizure, with no clinical signs of head trauma, lateralized weakness, or aphasia, but drowsy several hours after receiving lorazepam 2 mg intravenously. Computerized tomography axial brain scan shows (**A**) craniotomy defect on right, (**B**) mixed chronic and acute subdural hematoma on left, and (**C**) left frontal encephalomalacia from prior trauma.

Marchiafava-Bignami syndrome was first described in postmortem studies of Italian chianti drinkers but can occur in persons who consume any type of alcohol. Acute signs include coma, seizures, hemiparesis, rigidity, and sometimes death. The most severe pathology involves demyelination and necrosis of the corpus callosum. Findings on MRI include increased T2- and diffusion-weighted signals in the corpus callosum, especially the splenium.

Cerebellar degeneration and ataxia result from alcohol-induced loss of Purkinje cells, mainly in the anterior superior part of the cerebellar vermis; the cerebellar hemispheres are less affected. As a result, the clinical picture is one of mainly truncal and gait ataxia, with a wide-based unsteady gait and inability to walk tandem. The arms are much less involved if affected at all. Intention tremor, nystagmus, and dysarthria are rare. Findings are exacerbated by concurrent thiamine deficiency, leading to Wernicke encephalopathy. Alcohol may activate antibodies against Purkinje cells in individuals with gluten intolerance.

Optic neuropathy, which occurs with severe chronic alcohol abuse, is manifested as progressive painless visual loss as a result of damage to the optic nerve fibers. The macular region is most affected. Similar findings, first described in Cuban men who were heavy cigar smokers, have also been attributed to tobacco. However, neither alcohol nor tobacco is apparently directly responsible for the optic nerve damage, so it is more likely the syndrome results from malnutrition and deficiencies of multiple vitamins, including vitamins A and B.

Peripheral neuropathy (Chapter 420) ("alcoholic neuropathy") is the most common neurologic complication of chronic alcoholism. It is an axonal sensorimotor neuropathy that causes dysfunction of small nerve fibers, thereby leading to distal painful sensory symptoms such as burning dysesthesias and paresthesias of the soles of the feet. The typical numbness develops in a stocking-glove distribution, with loss of ankle reflexes. Mild distal weakness eventually occurs in some patients. Involvement of the autonomic nervous system frequently causes impotence as well as urinary or bowel complaints. Although vitamin supplements, especially thiamine and pyridoxine, may lead to some improvement, especially in the painful paresthesias, complete resolution is rare. If abstinence from alcohol is not also achieved, the symptoms persist, thereby implying that a direct toxic effect of alcohol is likely.

Compressive neuropathies, especially of the radial nerve ("Saturday night palsy") and peroneal nerve, can result after prolonged pressure on a nerve while the patient is obtunded from heavy alcohol consumption. Recovery takes many weeks but is generally complete.

Myopathy occurs in binge drinkers, in whom severe muscle injury with rhabdomyolysis (Chapter 113) can develop, especially in the setting of fasting and prolonged absence of movement. Myoglobinuria can result in kidney damage. Heavy alcohol consumption is also associated with cardiomyopathy (Chapter 60), which can lead to arrhythmias even in the absence of hypokalemia. Chronic alcohol abuse causes a symmetrical proximal weakness (Chapter 421), which is not usually severe enough to prevent walking or standing. It can be detected in up to 50% of heavy users, but only with a careful examination.

Fetal alcohol syndrome is recognized at birth in infants whose mothers consumed significant amounts of alcohol in the early stages of pregnancy, but the quantity required to place a fetus at risk has not been definitively determined. The characteristic findings are growth retardation, microcephaly, hypotonia, skeletal and cardiac anomalies, and characteristic facial features (micrognathia, small palpebral fissures). Recently, migration defects have been demonstrated by diffusion-weighted MRI and tractography. Exposure of the developing brain to alcohol can also lead to subtle or severe neurocognitive defects and attention deficit disorder, which may not be detected until later in childhood. Although malnutrition and excess alcohol have synergistic deleterious effects, the teratogenic effects of alcohol are not prevented by adequate amounts of thiamine, folate, and other vitamins.

Grade A Reference

A1. Andres E, Fothergill H, Mecili M. Efficacy of oral cobalamin (vitamin B_{12}) therapy. *Expert Opin Pharmacother*. 2010;11:249-256.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Kumar N. Acute and subacute encephalopathies: deficiency states (nutritional). *Semin Neurol*. 2011;31:169-183.
2. Osiezagha K, Ali S, Freeman C, et al. Thiamine deficiency and delirium. *Innov Clin Neurosci*. 2013;10:26-32.
3. Cerejo R, Newey C, Stillman M. Teaching NeuroImages: Wernicke encephalopathy: diagnostically deceptive but treatable. *Neurology*. 2013;80:e92.
4. Kuhn AL, Hertel F, Boulanger T, et al. Vitamin B_1 in the treatment of Wernicke's encephalopathy due to hyperemesis after gastroplasty. *J Clin Neurosci*. 2012;19:1303-1305.
5. Schwendimann RN. Metabolic, nutritional, and toxic myelopathies. *Neurol Clin*. 2013;31:207-218.
6. Stabler SP. Clinical practice. Vitamin B_{12} deficiency. *N Engl J Med*. 2013;368:149-160.
7. Al Shahrani F, Al Johani N. Vitamin D: deficiency, sufficiency and toxicity. *Nutrients*. 2013;5:3605-3613.
8. Ascherio A, Munger KL, White R, et al. Vitamin D as an early predictor of multiple sclerosis activity and progression. *JAMA Neurol*. 2014;71:306-314.
9. Littlejohns TJ, Henley WE, Lang IA, et al. Vitamin D and the risk of dementia and Alzheimer disease. *Neurology*. 2014;83:920-928.
10. Verma R, Praharaj HN, Khanna VK, et al. Study of micronutrients (copper, zinc and vitamin B_{12}) in posterolateral myelopathies. *J Neurol Sci*. 2013;329:11-16.
11. Leise MD, Poterucha JJ, Kamath PS, et al. Management of hepatic encephalopathy in the hospital. *Mayo Clin Proc*. 2014;89:241-253.
12. Laleman W, Simon-Talero M, Maleux G, et al. Embolization of large spontaneous portosystemic shunts for refractory hepatic encephalopathy: a multicenter survey on safety and efficacy. *Hepatology*. 2013;57:2448-2457.
13. Ridley NJ, Draper B, Withall A. Alcohol-related dementia: an update of the evidence. *Alzheimers Res Ther*. 2013;5:3.
14. Sabia S, Elbaz A, Britton A, et al. Alcohol consumption and cognitive decline in early old age. *Neurology*. 2014;82:332-339.

REVIEW QUESTIONS

1. A 12-year-old Muslim boy participated in the observance of Ramadan for the first time. After 10 days of daytime fasting, during which he frequently ended his fast by consuming candy, he began vomiting. Three days later he complained of blurred and double vision and trouble walking. He gradually became lethargic and disoriented. On examination, he was apathetic, had lateral rectus palsy, and nystagmus in horizontal gaze, and he had a broad-based gait with inability to tandem walk. Dysmetria was present in the legs but not the arms. Computed tomography and cerebrospinal fluid examinations were normal. Intravenous (IV) hydration did not improve his oculomotor findings, which resolved with IV thiamine. His mental status changes improved after 2 weeks. The diagnosis is:

A. Dehydration
B. Copper deficiency
C. B_1 (thiamine) deficiency
D. Meningococcal meningitis
E. Pellagra

Answer: C Thiamine (B_1) deficiency can cause Wernicke encephalopathy even in children, especially when glucose is consumed without preceding thiamine. Thiamine supplies last about 2 to 3 weeks in normal states, less in situations requiring high energy. B_{12} deficiency occurs in vegans and older patients with pernicious anemia, but a period of at least 12 months is required to deplete B_{12} stores. Copper deficiency resembles B_{12} clinically and similarly takes several months to develop, especially after gastric bypass surgery. Meningococcal meningitis was a problem in pilgrimages to Mecca, but now a vaccine must be demonstrated by the Saudi government prior to entering the country.

2. A 40-year-old woman presents with very profound proximal weakness and areflexia 8 weeks after "lap band" bariatric surgery, with which her stomach size was severely restricted. Since the operation, she has suffered from unremitting vomiting and has consumed only liquids, with no vitamin supplements. Quadriceps biopsy shows necrosis and fat replacement of muscle. The diagnosis is:

A. Alcoholic proximal myopathy
B. Magnesium deficiency
C. Vitamin D deficiency
D. Carnitine deficiency
E. Thiamine deficiency

Answer: E Thiamine deficiency is most often found in patients who have excessive recurrent vomiting, and is famously associated with Wernicke encephalopathy followed by Korsakoff syndrome, or peripheral neuropathy (beri-beri). Alcohol use and vitamin D deficiency cause a milder myopathy, and carnitine is associated with central nervous system dysfunction.

3. A 35-year-old Indian woman in her fifth pregnancy presents with leg weakness and spasticity, pain in her feet, and fear of delivering a baby with deformities as she had in her last pregnancy. She has used prenatal vitamins and extra folate because she knew that her vegetarian diet may be lacking some nutrients. Her deficiency is due to:

A. Folate
B. Thiamine
C. B_{12}
D. Iron
E. B_6

Answer: C B_{12} is present in meat and some nuts, but prolonged reliance on a vegetarian diet (>18 months) can use up all stores from the liver, especially in the face of multiple pregnancies. B_{12} deficiency, like folate deficiency, is associated with an increased risk of neural tube defects. Clinically it causes a combined spinal cord degeneration of posterior columns, which leads to proprioception deficits, and lateral columns, which leads to spasticity. It also is frequently accompanied by a peripheral neuropathy that may be painful. Although up to 25% of Indian women are folate deficient, it is correctable with vitamin supplements, which she is using. Other deficiencies produce anemia, neuropathy, and mental status changes but not myelopathy.

4. A famine relief agency flies to Nigeria in the face of a drought that has compromised the already marginal nutritional status of the residents, who rely almost exclusively on flour from the cassava plant for their food. Because there is limited water, the tuber is not always detoxified. The population has many older men who have trouble walking because of a spastic gait and inability to locate their feet. Some are blind or deaf as well. Some women and children with more severe permanent weakness and spasticity cannot walk at all. The cause is:

A. Niacin deficiency
B. Lead poisoning from contaminated well water
C. B_1 deficiency with pellagra
D. Lathyrism
E. Cassavism

Answer: E Cassavism is caused by a failure to remove toxins from the cassava plant and causes variable syndromes depending on patients' age and probable simultaneous vitamin deficiencies. Niacin deficiency causes peripheral neuropathy but not spasticity. Lead poisoning can cause isolated nerve palsies, especially the radial nerve, and static encephalopathy in children but not spasticity. Thiamine deficiency can cause peripheral neuropathy but with no spinal cord involvement. Lathyrism causes spasticity but is caused by reliance on chick peas as a single food source.

417

CONGENITAL, DEVELOPMENTAL, AND NEUROCUTANEOUS DISORDERS

JONATHAN W. MINK

CONGENITAL DISORDERS

Malformations of Cerebral Cortex

Developmental malformations of the cerebral cortex arise from a wide variety of etiologies, including genetic mutations, intrauterine infections, intrauterine ischemia, and toxic exposures.[1] These malformations are heterogeneous and can result from disrupted neuronal proliferation, migration, or cortical organization. In general, disorders that arise early in development are more severe than those that arise after the basic architecture of the brain has developed. When small areas of the brain are involved, the patient may have minor impairment of neurologic function. When larger areas of the brain are involved, patients often have cognitive deficits and more severe neurologic dysfunction. Epilepsy (Chapter 403), which is the most common manifestation of abnormal cortical development, may occur with or without other neurologic signs or symptoms.

Disorders of Neuronal Proliferation

Neuronal proliferation can be abnormally increased or decreased owing to a variety of mechanisms. These disorders can manifest with megalencephaly or microcephaly, or head size can be normal. Abnormal proliferation can involve specific cell types, thereby resulting in focal or multifocal areas of dysplasia or in the formation of hamartomas (see Tuberous Sclerosis, later).

FOCAL CORTICAL DYSPLASIA WITH BALLOON CELLS

Focal cortical dysplasia is caused by abnormal proliferation of both neurons and glia. Its neuropathology is characterized by the presence of giant dysmorphic neurons and "balloon cells" associated with altered cortical lamination, but some lesions have abnormal cortical layering with ectopic neurons in white matter. Affected patients typically present with partial seizures that are often intractable to medical therapy. These seizures can begin at any age but most commonly present during childhood or adolescence. The type of seizure depends on the anatomic location of the dysplasia. Other neurologic manifestations such as sensory, motor, or cognitive impairments depend on the extent of the dysplasia and whether multiple brain regions are affected. The diagnosis of focal cortical dysplasia is usually made with brain magnetic resonance imaging (MRI), which demonstrates focal thickening of a gyrus or alteration of the gray–white matter junction. Management includes medical treatment of seizures, but surgical resection of the epileptic focus may be required for complete remission (Chapter 403).

Disorders of Neuronal Migration

Disorders of neuronal migration typically result in disruption of the normal laminar organization of the cerebral cortex. Defects include impaired initiation of neuronal migration, impaired orderly migration, and impaired termination of migration. All result in abnormal cortical organization and function.

LISSENCEPHALY AND BAND HETEROTOPIA

The lissencephalies (smooth brain) are a group of disorders that are caused by arrested migration of neurons to the cerebral cortex. Lissencephaly, which is typically diagnosed in infancy or early childhood, is usually accompanied by microcephaly, severe global developmental delay, cerebral palsy,[2] and intractable epilepsy. Lissencephaly genes include *PAFAH1B1* (also known as *LIS1* on chromosome 17p13.3), *DCX* (on chromosome Xq22), and *TUBA1A* (on chromosome 12), all of which are involved in the regulation of microtubule organization and function. Individuals with mutations of *LIS1* typically have severe malformations that are most prominent in the posterior cerebrum. More extensive mutations in the region of *LIS1* result in Miller-Dieker syndrome, a condition characterized by lissencephaly and distinctive facial features that include a prominent forehead, midface hypoplasia, low-set and abnormally shaped ears, and a small jaw. Males with *DCX* mutations typically have a severe lissencephaly that is most prominent in the anterior cerebrum. Individuals with *TUBA1A* mutations may have isolated lissencephaly or may have lissencephaly with cerebellar hypoplasia. Diagnosis of lissencephaly is made by brain MRI that shows a smooth cortex with minimal sulcation. Genetic testing for *LIS1*, *DCX*, and *TUBA1A* is available. Management consists of seizure control, genetic counseling, and supportive care.

Band heterotopia (double cortex) is a less severe form of lissencephaly that is usually seen in women with *DCX* mutation. Clinical manifestations of band heterotopia range from mild to severe and include seizures, intellectual disability, and developmental delay. Women with a *DCX* mutation are at risk of having male children with severe lissencephaly. Brain MRI demonstrates a band of gray matter underlying a nearly normal-appearing cerebral cortex. Management consists of seizure control and genetic counseling.

NODULAR HETEROTOPIA

Nodular heterotopias are characterized by nodular ectopic collections of neurons and glia in the subependyma or in the subcortical white matter. The most important form is subependymal nodular heterotopia, a condition characterized by multiple gray matter nodules in the walls of the lateral ventricles bilaterally. This X-linked condition is due to a mutation in *FLNA* (chromosome Xq28), which codes for filamin A, an actin-cross-linking phosphoprotein that is critical for the initiation of migration. As a result of this mutation, many neurons do not migrate out of the subventricular zone. Most affected individuals are heterozygous females. Males are severely affected and often die in infancy. Most affected females present with seizures during childhood or adolescence. Females may be intellectually normal or have mild disability. Individuals with subependymal nodular heterotopia appear to be at increased risk for aortic or carotid dissection and for cardiac valvular abnormalities.

The diagnosis is based on brain MRI, which shows gray matter nodules along the walls of the lateral ventricles. Genetic testing for *FLNA* is available. Management consists of seizure control and genetic counseling.

Disorders of Cortical Organization

Disorders of cortical organization include conditions such as polymicrogyria and schizencephaly. These disorders are not due to abnormal numbers of neurons or impaired migration but instead include abnormalities of gyration, sulcation, connectivity, or synaptogenesis. The best-understood of these disorders are polymicrogyria and schizencephaly.

POLYMICROGYRIA

Polymicrogyria is characterized by regions of complex cortical convolutions with miniature gyri that are fused and superimposed together. Polymicrogyria is caused by failure of cortical organization as a result of in utero injury or genetic mutation; it has been associated with prenatal infections (e.g., cytomegalovirus) and possible vascular abnormalities, but often it is idiopathic. A single gene, *GPR56* (chromosome 16q13), has been associated with bilateral frontoparietal polymicrogyria. *GPR56* codes for a G protein–coupled receptor that appears to be important for human cerebral cortical development. Clinical manifestations include epilepsy, developmental delay, cerebral palsy, and intellectual disability, depending on the location and extent of the abnormality. The diagnosis of polymicrogyria is made by brain MRI. Clinical management consists of seizure management and supportive therapies.

SCHIZENCEPHALY

Schizencephaly is characterized by infolding of cortical gray matter along a hemispheric cleft near the primary cerebral fissures. It is thought to represent a more extensive injury than what leads to polymicrogyria. In most cases, the cause cannot be determined, but it has been associated with in utero insult. A rare familial form has been described, but no gene has been identified. Clinical features include developmental delay, cerebral palsy, dysarthria, and epilepsy. The clinical abnormalities are more severe with large open-lip schizencephaly and with bilateral lesions than with small unilateral closed-lip schizencephaly. Diagnosis is made by brain MRI. Management consists of seizure control and supportive therapies when indicated.

Malformations of Cerebellum and Brain Stem

Developmental abnormalities of the hindbrain are less well understood than are abnormalities of cerebral cortical development.[3] Two of the better known and important syndromes are Joubert syndrome and Dandy-Walker malformation.

JOUBERT SYNDROME

Joubert syndrome is characterized by a distinctive pattern of cerebellar and brain stem developmental malformation. Four causative genes (*NPHP1*, *CEP290*, *AHI1*, and *TMEM67* [*MKS3*]) together account for approximately 30% of cases. Clinical features include hypotonia, truncal ataxia, developmental delay, abnormal eye movements, and disordered breathing. The combination of signs and severity can be variable. Some individuals with Joubert syndrome also have retinal dystrophy, renal disease, ocular colobomas, occipital encephalocele, or hepatic fibrosis. No formal diagnostic criteria exist. The diagnosis is usually based on the combination of hypotonia in infancy with later development of ataxia, intellectual impairment, and abnormal breathing pattern, or abnormal eye movements in combination with a characteristic MRI finding known as the molar tooth sign. The molar tooth sign results from hypoplasia of the cerebellar vermis and accompanying brain stem abnormalities on axial imaging through the junction of the midbrain and pons. Genetic testing is available for the four identified genes. Management is supportive. Caffeine can be helpful for periodic hypoventilation, but some patients require tracheostomy.

DANDY-WALKER MALFORMATION

Dandy-Walker malformation is characterized by cerebellar vermis hypoplasia and cystic dilation of the fourth ventricle. Rare familial cases have been reported, but a genetic basis has not been identified. This heterogeneous disorder is usually accompanied by hypotonia, delayed motor development, and ataxia. Intellectual disability is present in about 50% of affected individuals. In some cases, hydrocephalus requires shunting. Diagnosis is based on characteristic findings on brain MRI. Treatment is supportive, with cerebrospinal fluid (CSF) shunting when indicated.

CHIARI MALFORMATIONS

Four types of Chiari malformation have been described. The most common of these are Chiari types I and II. Chiari I malformations are most often diagnosed in adulthood, whereas Chiari II malformations are associated with spina bifida and are usually diagnosed in childhood.

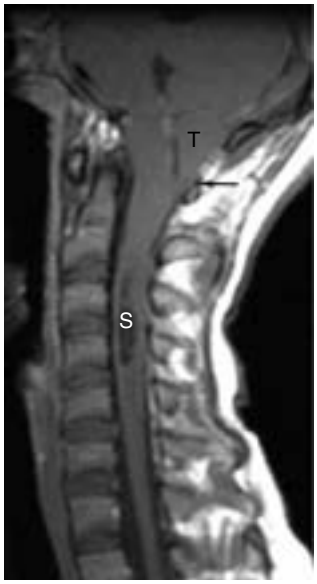

FIGURE 417-1. Chiari I malformation. A sagittal magnetic resonance image shows low, pointed cerebellar tonsils (i.e., Chiari I malformation, T) that extend to the level of C1 (*arrow*) and a dilated central canal of the spinal cord (i.e., syringohydromyelia [S]). (From Barkovich AJ, Kuzniecky RI. Congenital, developmental, and neurocutaneous disorders. In Goldman L, Ausiello D, eds. *Cecil Textbook of Medicine*. 23rd ed. Philadelphia: Saunders Elsevier; 2008:2790.)

Chiari I malformations are characterized by downward displacement of the cerebellar tonsils through the foramen magnum, often first accompanied by compression of the tonsils. Chiari I is a developmental abnormality that is thought to be congenital in most cases, even though symptoms may not present until adulthood, typically in the third or fourth decade of life. The abnormality is often asymptomatic and discovered only as an incidental finding. However, clinical manifestations can result from compression of neural structures at the cranial-cervical junction or obstruction of CSF flow. Signs and symptoms include headaches that worsen with straining or coughing, lower cranial nerve findings, downbeat nystagmus, ataxia, or long-tract signs. Chiari I malformations are accompanied by syringomyelia (see later) in up to 80% of cases. Diagnosis is made with brain MRI, which shows the cerebellar tonsils extending through the foramen magnum 5 mm or more (Fig. 417-1). Surgical treatment with craniocervical decompression is recommended for symptomatic patients but usually not for asymptomatic individuals or patients whose only symptom is headache.[4]

Chiari II malformations, commonly called Arnold-Chiari malformations, are characterized by descent of the cerebellar tonsils, the inferior vermis, and portions of the cerebellar hemispheres into the spinal canal, along with elongation and displacement of the brain stem and fourth ventricle. Chiari II malformations are almost always associated with meningomyelocele and spina bifida. Hydrocephalus requiring shunting occurs in most cases. Brain stem dysfunction may result from intrinsic malformation or from compression of neural structures at the craniocervical junction. Treatment is surgical repair of the myelomeningocele, relief of hydrocephalus, and occasionally cervical bone decompression. The prognosis depends on the level and extent of the myelomeningocele and on the severity of brain anomalies.

Malformations of Spinal Cord

TETHERED SPINAL CORD

Tethered spinal cord syndrome is a disorder caused by an anomalous filum terminale that restricts the normal ascent of the conus medullaris and limits the movement of the spinal cord within the spinal column. The result is an abnormal stretching of the spinal cord, with neurologic symptoms referable to the lower spinal cord. Tethering may also develop after spinal cord injury. Associated spinal anomalies are common and may include diastematomyelia, spinal lipomas, dermal sinuses, and fibrolipomas of the filum terminale. Symptoms can occur at any age but usually develop during periods of rapid growth in childhood or adolescence. However, tethered spinal cord syndrome may go undiagnosed until adulthood, when sensory and motor problems and loss of bowel and bladder control emerge. Erectile dysfunction may

occur in males. Symptoms are typically progressive. Diagnosis is made with MRI, which shows a low conus medullaris (i.e., below the bottom of the L2 vertebral body) or a thickened or fat-containing filum terminale. Diminished pulsations of the spinal cord may also be seen. Treatment consists of surgical release of the tethered cord. With successful surgery, symptoms typically do not progress and may improve.

SYRINGOHYDROMYELIA

Syringohydromyelia is a condition in which the central canal of the spinal cord (hydromyelia) or the substance of the spinal cord (syringomyelia) is expanded by the accumulation of CSF. In many cases, both hydromyelia and syringomyelia are present (syringohydromyelia). The proximate cause of syringes probably is altered flow of CSF, with variations in pressure in different parts of the subarachnoid space. The pressure variations create forces that drive CSF into the spinal cord. Possible causes include narrowing of the foramen magnum, Chiari I and II malformations, intramedullary and extramedullary spinal cord tumors, and subarachnoid scarring. Subsequent extension of the cyst may result from rapid changes in intraspinal pressure owing to such events as coughing or sneezing. Symptoms of syringohydromyelia most commonly begin in late adolescence or early adulthood and progress irregularly, with long periods of stability. The classic presentation is asymmetrical weakness and atrophy in the upper extremities, loss of upper limb deep tendon reflexes, and loss of pain and temperature sensation (with preservation of vibration and proprioception) in the neck, arms, and upper part of the trunk. With progression, spasticity and hyperreflexia develop in the lower extremities. Progressive ascending and descending levels of weakness and sensory impairment typically occur over time. The diagnosis is made by spinal MRI (see Fig. 417-1). If syringohydromyelia is identified, it is important to perform a brain MRI to look for associated abnormalities of the craniocervical junction. Occasionally, mild central canal dilation is discovered incidentally in patients without spinal cord symptoms or signs. If no associated cause is found, the prognosis of such incidentally discovered anomalies is generally good. Treatment is directed at the cause if one can be identified. Syringopleural or syringoperitoneal shunting is sometimes performed, with variable benefit.

DEVELOPMENTAL DISORDERS

Disorders that result from impaired postnatal neurodevelopmental function range from specific disorders such as fragile X syndrome and Rett syndrome, to complex syndromes such as autism, to nonspecific developmental delay and learning disabilities.

Fragile X Syndrome

Fragile X syndrome is an X-linked trinucleotide repeat disorder that is characterized by nonsyndromic intellectual disability in most affected males. It is the most common genetic cause of intellectual disability, affecting 1/4000 males and 1/8000 females. The pathogenesis of fragile X syndrome is not well understood. The classic disorder is seen in males with full mutations (>200 repeats) in the *FMR1* gene.[5] Fragile X syndrome may present as only moderate to severe intellectual disability, but it is often associated with a prominent forehead, large ears, prominent jaw, and macro-orchidism. Postpubertal males often have poor impulse control, perseveration, and poor eye contact. Up to 25% of affected males have autism. Heterozygous females may be asymptomatic or may have a syndrome similar to what is seen in males, depending on repeat size and random X-inactivation.

Other disorders associated with *FMR1* include the fragile X ataxia syndrome, which is characterized by the late onset, usually after age 50 years, of progressive cerebellar ataxia and intention tremor in individuals who have an *FMR1* premutation (60 to 200 repeats). It occurs equally in males and females. Diagnosis of *FMR1* disorders is by molecular genetic testing. Cytogenetic testing for fragile sites is no longer recommended because it is less sensitive and more expensive than molecular testing. Treatment is symptomatic and supportive. Genetic counseling is recommended for affected individuals and their families.

Rett Syndrome

Rett syndrome is a neurodevelopmental disorder that occurs classically in females with mutations in the *MECP2* gene. *MECP2* mutations are generally lethal in male embryos, but Rett syndrome has been reported in males with XXY karyotype or with somatic mosaicism. *MECP2* is thought to mediate transcriptional silencing of methylated DNA. Most mutations are probably de novo or may reflect germline mosaicism; 99% of cases represent a single occurrence within a family. Affected girls are usually normal at birth and have apparently normal development for the first 6 to 18 months of life. Brain growth decelerates, and development stagnates, followed by rapid regression of language and motor skills. A classic feature of Rett syndrome is the loss of purposeful hand use and the development of repetitive stereotyped hand movements that usually have the appearance of wringing or clapping. Other features present to variable degree are bruxism, episodic apnea and hyperpnea, seizures, gait disorders, and tremor.[6] Non-neurologic features include growth failure and wasting, bowel dysmotility, scoliosis, osteopenia, and vasomotor changes in the limbs. Diagnosis is by clinical criteria followed by molecular genetic testing. Treatment is symptomatic.

Autism

Autism or autism spectrum disorder is characterized by impaired social communication and interactions as well as restricted and repetitive behaviors. Autism is associated with many different causes and is often idiopathic.[7] Fragile X syndrome and tuberous sclerosis are two important entities in which an autistic phenotype can occur and in which autism may be the most prominent feature.

Symptoms typically present before 3 years of age and persist into adulthood. Autism is a spectrum ranging from severe, with impairment in all domains, to mild with normal intellect and language but with impaired social interactions and repetitive behaviors or restricted interests. Autism has many causes but in most cases is idiopathic. Epilepsy is common in autism. Diagnosis is based on careful diagnostic interview and examination (Table 417-1). When epilepsy is present, treatment with antiepileptic medications is indicated. Behavioral therapy can help individuals learn rules for social interaction and can improve communication. It can also help with problematic behavior. Educational support is important. Medications such as atypical antipsychotics, selective serotonin reuptake inhibitors, and anxiolytics (Chapter 397) can help with aggressive behavior, repetitive behaviors, and anxiety.

NEUROCUTANEOUS DISORDERS

Neurocutaneous disorders are congenital syndromes characterized by dysplastic and neoplastic lesions primarily involving the nervous system and skin. The more than 40 described syndromes include neurofibromatosis, tuberous sclerosis, Sturge-Weber syndrome, and von Hippel-Lindau disease.

Neurofibromatosis

Neurofibromatosis encompasses a spectrum of syndromes with distinctive neural and cutaneous lesions. The two major forms of neurofibromatosis are genetically and clinically distinct.

NEUROFIBROMATOSIS TYPE 1

Neurofibromatosis type 1, which is the classic disorder described by von Recklinghausen, is an autosomal dominant condition with an incidence of

TABLE 417-1 DIAGNOSTIC CRITERIA FOR AUTISM SPECTRUM DISORDER

1. Deficits in Social Communication/Interaction (must have all three criteria):
   a. Problems reciprocating social or emotional interaction, including difficulty establishing or maintaining back-and-forth conversations and interactions, inability to initiate an interaction, and problems with shared attention or sharing of emotions and interests with others
   b. Severe problems maintaining relationships—ranges from lack of interest in other people to difficulties in pretend play and engaging in age-appropriate social activities, and problems adjusting to different social expectations
   c. Nonverbal communication problems such as abnormal eye contact, posture, facial expressions, tone of voice and gestures, as well as an inability to understand these
2. Restricted and Repetitive Behavior (at least 2 criteria must be met):
   a. Stereotyped or repetitive speech, motor movements or use of objects
   b. Excessive adherence to routines, ritualized patterns of verbal or nonverbal behavior, or excessive resistance to change
   c. Highly restricted interests that are abnormal in intensity or focus
   d. Hyper- or hyporeactivity to sensory input or unusual interest in sensory aspects of the environment

Symptoms must be present in early childhood but may not become fully manifest until social demands exceed capacities. Symptoms need to be *functionally impairing* and not better described by another DSM-5 diagnosis.

From American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Publishing, 2013.

1 per 2500 to 3000 births.[8] Although it is an autosomal dominant disease, approximately 50% of cases are due to new mutations. Most mutations in *NF1* occur in the parental germline. The *NF1* gene, which is located on chromosome 17q11.2, codes a protein called neurofibromin, which is thought to function as a tumor suppressor by acting as a negative regulator of the Ras signaling pathway. Neurofibromatosis type 1 is characterized by multiple café au lait spots, axillary and inguinal freckling, multiple discrete cutaneous neurofibromas (Fig. 417-2), and Lisch nodules (Table 417-2). Subcutaneous neurofibromas may be painful or disfiguring. Learning disabilities are present in at least 50% of individuals. Other manifestations include plexiform neurofibromas, optic nerve and other central nervous system (CNS) gliomas, malignant peripheral nerve sheath tumors, tibial dysplasia, and vasculopathy.

Management of patients depends on the specific manifestations and often requires multidisciplinary collaboration. Most patients with neurofibromatosis type 1 do not require treatment, but all require surveillance (Table 417-3). Subcutaneous, intraspinal, and intracranial tumors can be treated surgically. Optic nerve gliomas may be treated with chemotherapy; both cisplatin and temozolomide have shown some benefit. Radiation is not recommended. Genetic counseling should be provided to all patients and their families.

NEUROFIBROMATOSIS TYPE 2

Neurofibromatosis type 2, which is often referred to as central neurofibromatosis, is an autosomal dominant condition with an incidence of approximately 1 in 25,000 individuals.[9] The *NF2* gene is located on chromosome 22q12.2. Its gene product merlin is a cytoskeletal protein thought to act as a membrane-stabilizing protein. The specific function of merlin is unknown. Neurofibromatosis type 2 is characterized by bilateral vestibular schwannomas, which usually present with symptoms of tinnitus, hearing loss, and imbalance. The age at onset is usually in young adulthood, but some individuals may develop posterior subcapsular lens opacities or mononeuropathy in childhood. Almost all affected individuals develop bilateral vestibular schwannomas by age 30 (Table 417-4). Affected individuals may also develop schwannomas of other cranial and peripheral nerves, meningiomas, and (rarely) ependymomas or astrocytomas. Posterior subcapsular lens opacities are the most common ocular abnormality.

Management is dependent on the specific manifestations and complications. In individuals who either have tested positive for known *NF2* mutations or have a family history of neurofibromatosis type 2 and whose genetic status cannot be determined with genetic testing, annual brain MRI is recommended starting between ages 10 and 12 years and continuing until at least age 40 years. Hearing evaluations may be useful in detecting changes in auditory nerve function before changes can be visualized by MRI. Routine complete eye examinations should be part of the care of all individuals.

Bevacizumab, a vascular endothelial growth factor inhibitor (5 mg/kg intravenously every 2 weeks), can improve hearing in some patients with neurofibromatosis type 2 and vestibular schwannomas. Surgical treatment of schwannomas and meningiomas may be indicated to preserve function or to relieve compression of adjacent structures, especially in patients with intramedullary spinal tumors. Genetic counseling should be provided to affected individuals and their families.

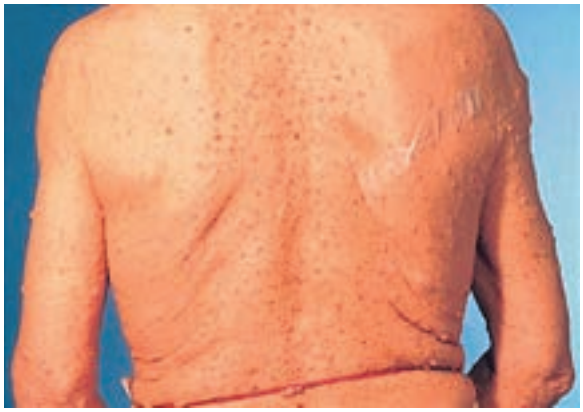

FIGURE 417-2. Multiple neurofibromas covering the back of a patient with neurofibromatosis type 1.

TABLE 417-2 DIAGNOSTIC CRITERIA FOR NEUROFIBROMATOSIS TYPE 1

Two or more of the following clinical features signify the presence of neurofibromatosis type 1:
- Six or more café au lait macules (>0.5 cm at largest diameter in prepubertal individuals or >1.5 cm in individuals past puberty)
- Axillary freckling or freckling in inguinal regions
- Two or more neurofibromas of any type or ≥ 1 plexiform neurofibroma
- Two or more Lisch nodules (iris hamartomas)
- A distinctive osseous lesion
- A first-degree relative with neurofibromatosis type 1 diagnosed by using the above-listed criteria

TABLE 417-3 RECOMMENDED SURVEILLANCE IN PATIENTS WITH NEUROFIBROMATOSIS TYPE 1

Annual physical examination by a physician who is familiar with the individual and with the disease
Annual ophthalmologic examination in early childhood, less frequent examination in older children and adults
Regular developmental assessment by screening questionnaire (in childhood)
Regular blood pressure monitoring
Other studies only as indicated on the basis of clinically apparent signs or symptoms
Monitoring of those who have abnormalities of the central nervous system, skeletal system, or cardiovascular system by an appropriate specialist

Tuberous Sclerosis

Tuberous sclerosis complex is characterized by abnormalities of the brain, kidney, and heart.[10] Tuberous sclerosis may occur as an autosomal dominant syndrome or result from spontaneous mutation. Two tuberous sclerosis genes have been identified. *TSC1* (chromosome 9q34) codes for a protein called hamartin, a protein that interacts with the product of the *TSC2* gene to inhibit the mammalian target of rapamycin (mTOR). *TSC2* (chromosome 16p13) codes for tuberin, which interacts with hamartin. *TSC2* mutations account for about 60% of individuals with clinical tuberous sclerosis.[11]

Specific findings vary across individuals, and severity ranges from minimal to severe. Skin lesions are seen in almost 100% of affected individuals, but CNS lesions are the leading cause of morbidity and mortality. Epilepsy is seen in as many as 80% of patients with CNS lesions. Intellectual impairment and developmental delay are common, and up to 40% of patients have an autism spectrum disorder. Giant cell astrocytoma is the leading cause of death. Up to 80% of children with tuberous sclerosis have an identifiable renal lesion (Chapter 197) by 10.5 years of age, and renal disease is the second leading cause of early death in individuals with tuberous sclerosis. Cardiac rhabdomyomas, which can occur in up to 50% of patients, are usually present at birth and typically regress over time. Diagnosis of tuberous sclerosis (Table 417-5) is usually clinical and confirmed by identification of calcified or uncalcified hamartomas on imaging studies (Fig. 417-3).

Treatment is directed at complications of the disease, particularly epilepsy (Chapter 403). Neurosurgical intervention may sometimes be indicated for epilepsy and for symptomatic treatment of complications such as hydrocephalus, which results from midline giant cell tumors. In a randomized

TABLE 417-4 DIAGNOSTIC CRITERIA FOR NEUROFIBROMATOSIS TYPE 2

Presence of one or more of the following makes the diagnosis of neurofibromatosis type 2:
- Bilateral vestibular schwannomas
- A first-degree relative with neurofibromatosis type 2, *and*
  - Unilateral vestibular schwannoma, *or*
  - Any two of: meningioma, schwannoma, glioma, neurofibroma, posterior subcapsular lenticular opacities*
- Unilateral vestibular schwannoma, *and*
  - Any two of: meningioma, schwannoma, glioma, neurofibroma, posterior subcapsular lenticular opacities*
- Multiple meningiomas, *and*
  - Unilateral vestibular schwannoma, *or*
  - Any two of: schwannoma, glioma, neurofibroma, cataract*

*"Any two of" refers to two individual tumors or cataracts.

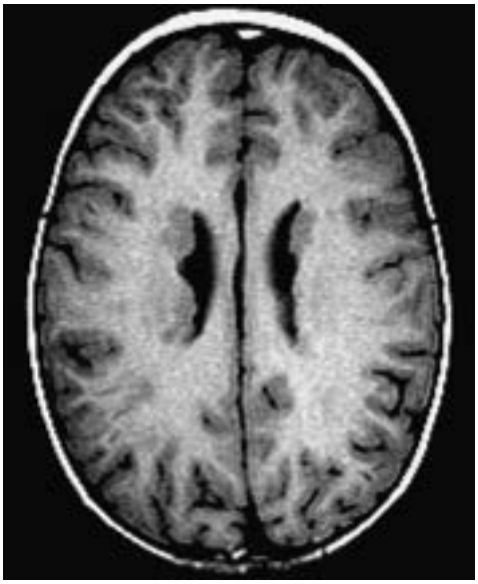

FIGURE 417-3. **Subependymal nodules and multiple cortical tubers in a patient with tuberous sclerosis.**

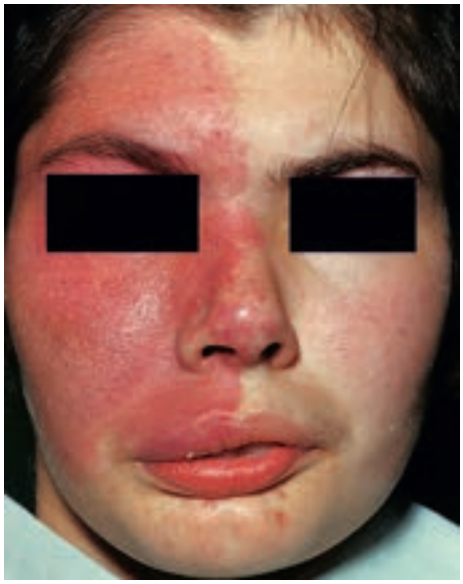

FIGURE 417-4. **Sturge-Weber syndrome.** This patient has a classic diffuse capillary hemangioma in the distribution of the ophthalmic, nasociliary, and maxillary branches of the trigeminal nerve. The lesion extends backward over the anterior two thirds of the crown of the head. (From Forbes CD, Jackson WD. *Color Atlas and Text of Clinical Medicine*. 2nd ed. London: Mosby; 1996.)

TABLE 417-5 DIAGNOSTIC CRITERIA FOR TUBEROUS SCLEROSIS COMPLEX

Definite—Two major features or one major feature plus two minor features
Probable—One major feature plus one minor feature
Possible—One major feature or two or more minor features

MAJOR FEATURES
Facial angiofibromas or forehead plaque
Nontraumatic ungual or periungual fibromas
More than three hypomelanotic macules (ash leaf spots)
Shagreen patch (connective tissue nevus)
Multiple retinal nodular hamartomas
Cortical tuber
Subependymal nodule
Subependymal giant cell astrocytoma
Cardiac rhabdomyoma, single or multiple
Lymphangiomyomatosis
Renal angiomyolipoma
MINOR FEATURES
Multiple dental enamel pits
Hamartomatous rectal polyps
Bone cysts
Cerebral white matter radial migration lines
Gingival fibromas
Nonrenal hamartoma
Retinal achromic patch
"Confetti" skin lesions
Multiple renal cysts

controlled trial, treatment with everolimus (10 mg/day) reduced the size of the angiomyolipomas in 42% percent of participants receiving active drug as compared with 0% of participants receiving placebo.■ In another double-blind placebo controlled trial, everolimus titrated to a concentration of 5 to 15 ng/mL was effective in reducing the size subependymal giant cell astrocytomas by at least 50% in 35% of participants receiving active drug as compared with 0% of participants receiving placebo.■ At similar doses, everolimus also can change fractional anisotropy and radial diffusivity, suggesting that the genetic defect of tuberous sclerosis complex in the brain may be modified pharmacologically. Serial brain MRI and renal ultrasound screening may be indicated in some patients because benign tumors of these organs may enlarge rapidly. Genetic counseling is an important part of management.

Sturge-Weber Syndrome

Sturge-Weber syndrome is a sporadic disorder characterized by facial vascular nevi, epilepsy, cognitive impairment, and sometimes hemiparesis, hemianopsia, or glaucoma. It is most commonly due to somatic mutation in *GNAQ* (chromosome 9q21).[12] The characteristic CNS feature of this disorder is capillary angiomatosis of the pia mater. Cerebral cortical calcifications are generally seen in a pericapillary distribution and are progressive. Most patients with Sturge-Weber syndrome have epilepsy. The diagnosis is usually based on the presence of a facial nevus (Fig. 417-4), which is manifested as a typical port-wine stain, and confirmatory imaging on a contrast brain MRI showing leptomeningeal enhancement.

Regular ophthalmologic examination is warranted because of the risk for glaucoma. Treatment is usually aimed at the epilepsy, which can be medically intractable. In patients with intractable epilepsy and infantile-onset hemiplegia, hemispherectomy can improve the seizures and the neurodevelopmental outcome.

Von Hippel-Lindau Disease

Von Hippel-Lindau disease (i.e., CNS angiomatosis) is an autosomal dominant disorder caused by a defective tumor suppressor gene (*VHL*) at chromosome 3p25-p26.[13] It is characterized by retinal angiomas, brain (usually cerebellar) and spinal cord hemangioblastomas, renal cell carcinomas, endolymphatic sac tumors, pheochromocytomas, papillary cystadenomas of the epididymis, angiomas of the liver and kidney, and cysts of the pancreas, kidney, liver, and epididymis. Both sexes are affected equally.

Symptoms typically begin during the third or fourth decade. Retinal inflammation with exudate, hemorrhage, and retinal detachment from the retinal angiomas typically precedes the cerebellar complaints, but the order is not constant. The ocular findings are nonspecific, and the retinal detachment may mask the underlying lesion. Headache, vertigo, and vomiting result from cerebellar tumors. Cerebellar signs such as ataxia, dysdiadochokinesis, and dysmetria are common. Rare patients present with symptoms of spinal cord or visceral lesions, or may have hearing loss from tumors of the endolymphatic sac.

Clinical diagnosis is established if the patient has more than one CNS hemangioblastoma, one hemangioblastoma with a visceral manifestation of the disease, or one manifestation of the disease and a known family history. Molecular genetic testing detects mutations in the *VHL* gene in nearly 100% of affected individuals.

For patients with von Hippel-Lindau disease and for those with a disease-causing VHL mutation, surveillance is recommended with annual ophthalmologic examination, annual blood pressure monitoring, measurement of urinary catecholamine metabolites beginning at age 5 years in families with pheochromocytoma, and annual abdominal ultrasound examination beginning at age 16 years, with evaluation of suspicious lesions by computed tomography or MRI. Treatment is symptomatic. Retinal detachments and tumors are treated by laser therapy. Large brain tumors (Chapter 189), renal cell carcinomas (Chapter 197), pheochromocytomas (Chapter 228), epididymal tumors (Chapter 200), and endolymphatic sac tumors are treated surgically; smaller CNS tumors may be treated by gamma knife.

Grade A References

A1. Plotkin SR, Stemmer-Rachamimov AO, Barker FG 2nd, et al. Hearing improvement after bevacizumab in patients with neurofibromatosis type 2. *N Engl J Med*. 2009;361:358-367.

A2. Bissler JJ, Kingswood JC, Radzikowska E, et al. Everolimus for angiomyolipoma associated with tuberous sclerosis complex or sporadic lymphangioleiomyomatosis (EXIST-2): a multicentre, randomised, double-blind, placebo-controlled trial. *Lancet*. 2013;381:817-824.

A3. Franz DN, Belousova E, Sparagana S, et al. Efficacy and safety of everolimus for subependymal giant cell astrocytomas associated with tuberous sclerosis complex (EXIST-1): a multicentre, randomised, placebo-controlled phase 3 trial. *Lancet*. 2013;381:125-132.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Jamuar SS, Lam AN, Kircher M, et al. Somatic mutations in cerebral cortical malformations. *N Engl J Med*. 2014;371:733-743.
2. Colver A, Fairhurst C, Pharoah PO. Cerebral palsy. *Lancet*. 2014;383:1240-1249.
3. Doherty D, Millen KJ, Barkovich AJ. Midbrain and hindbrain malformations: advances in clinical diagnosis, imaging, and genetics. *Lancet Neurol*. 2013;12:381-393.
4. Klekamp J. Surgical treatment of Chiari I malformation—analysis of intraoperative findings, complications, and outcome for 371 foramen magnum decompressions. *Neurosurgery*. 2012;71: 365-380.
5. Visootsak J, Hipp H, Clark H, et al. Climbing the branches of a family tree: diagnosis of fragile X syndrome. *J Pediatr*. 2014;164:1292-1295.
6. Neul JL, Kaufmann WE, Glaze DG, et al. Rett syndrome: revised diagnostic criteria and nomenclature. *Ann Neurol*. 2010;68:944-950.
7. Lai MC, Lombardo MV, Baron-Cohen S. Autism. *Lancet*. 2014;383:896-910.
8. Gutmann DH, Parada LF, Silva AJ, et al. Neurofibromatosis type 1: modeling CNS dysfunction. *J Neurosci*. 2012;32:14087-14093.
9. Ferner RE. The neurofibromatoses. *Pract Neurol*. 2010;10:82-93.
10. Kothare SV, Singh K, Hochman T, et al. Genotype/phenotype in tuberous sclerosis complex: associations with clinical and radiologic manifestations. *Epilepsia*. 2014;55:1025-1029.
11. Kwiatkowski DJ, Manning BD. Molecular basis of giant cells in tuberous sclerosis complex. *N Engl J Med*. 2014;371:778-780.
12. Shirley MD, Tang H, Gallione CJ, et al. Sturge-Weber syndrome and port-wine stains caused by somatic mutation in GNAQ. *N Engl J Med*. 2013;368:1971-1979.
13. Richard S, Gardie B, Couve S, et al. Von Hippel-Lindau: how a rare disease illuminates cancer biology. *Semin Cancer Biol*. 2013;23:26-37.

REVIEW QUESTIONS

1. An 18-year-old man with more than 10 café au lait spots, axillary freckling, and a family history of neurofibromatosis type 1 has recently come into your practice for primary care. He is otherwise normal. Which of the following should be done as part of routine surveillance for complications of neurofibromatosis type 1?

A. Annual brain MRI scan
B. Annual physical examination with blood pressure
C. Annual electrocardiogram
D. Annual orthopedic evaluation
E. Annual ophthalmologic examination

Answer: B Current guidelines recommend annual physical examination and blood pressure check in adults with neurofibromatosis type 1. Other evaluations should be based on presence of symptoms or history of nervous system, skeletal or cardiovascular abnormalities. This patient has only cutaneous manifestations.

2. A 21-year-old with tuberous sclerosis has been diagnosed recently with a subependymal giant cell astrocytoma. Which of the following treatments should be considered to reduce the growth of this tumor?

A. Everolimus
B. Bevacizumab
C. Rituximab
D. Cyclophosphamide
E. Ventriculoperitoneal shunting

Answer: A Everolimus has been shown in a controlled phase 3 randomized trial to reduce the size of subependymal giant cell astrocytomas by more than 50% in 35% of study participants. Bevacizumab has been used in patients with neurofibromatosis type 2 and vestibular schwannomas to improve hearing. The other treatments have not been shown to reduce the growth of subependymal giant cell astrocytomas.

418

AUTONOMIC DISORDERS AND THEIR MANAGEMENT

WILLIAM P. CHESHIRE, JR.

The peripheral autonomic nervous system and the central integration of autonomic reflexes maintain homeostasis and modulate the complex physiologic responses at the interface between the internal milieu and the external world. Autonomic activity, which generally occurs below the level of conscious control, regulates cardiovascular, thermal, metabolic, gastrointestinal (GI), urinary, and reproductive functions and coordinates the adaptive response to stress.

Many diseases can involve the autonomic nervous system, which in turn can involve all organ systems. Examples include brain lesions that affect any part of the central autonomic network, disorders that damage peripheral nerve function, and systemic illnesses that impair autonomic responses.

EPIDEMIOLOGY

The most frequently disabling manifestation of autonomic failure is orthostatic hypotension, which increases in prevalence with aging and is associated with a two-fold increased risk for falls, fractures, syncope, transient ischemic attacks, and decreased functional capacity in elderly individuals. The prevalence of orthostatic hypotension is 5 to 20% among all elderly persons, but it rises to 30% in persons older than 75 years and to greater than 50% in frail individuals who live in nursing homes. Neurally mediated syncope (vasodepressor, vasovagal), as well as various situational syncopes that occur in response to emotional distress, micturition, defecation, coughing, carotid sinus stimulation, and other factors, accounts for 1 to 3% of all emergency room visits. The lifetime prevalence of syncope is about 20%. Another common autonomic syndrome is postprandial hypotension, which occurs in 20 to 60% of elderly individuals and is associated with an increased risk for mortality.

Diabetes mellitus (Chapter 229) is the most common cause of autonomic neuropathy in the developed world. Within 10 to 15 years of the onset of diabetes, laboratory evidence of autonomic neuropathy can be detected in about 30% of patients, and symptomatic autonomic failure with orthostatic hypotension can be detected in about 5% of patients.[1] Other autonomic symptoms in diabetics include constipation in 40 to 60% of patients, gastroparesis in 20 to 40%, bladder dysfunction in 30 to 80%, erectile impotence in more than 30% of males, and occasionally intermittent diarrhea.

Hyperhidrosis involving the palms and soles represents the most common form of essential hyperhidrosis and affects about 1% of the population. Although excessive sweating is usually symptomatic, anhidrosis may go unnoticed unless it interferes with the thermoregulatory response to heat stress. Impaired thermoregulation can increase mortality during heat waves or in times of heat stress (Chapter 109).

PATHOBIOLOGY

The peripheral autonomic nervous system comprises three main divisions: (1) the sympathetic outflow from the thoracolumbar segments of the spinal cord, (2) the parasympathetic outflow from cranial nerves III, VII, IX, and X and the sacral spinal segments, and (3) the enteric ganglionated plexuses intrinsic to the wall of the gut. Disorders of the autonomic nervous system may occur suddenly or evolve gradually. They may affect specific or multiple autonomic pathways, depending on their pathogenesis and localization. Orthostatic hypotension, a hallmark of autonomic disorders, results from sympathetic vasomotor denervation, which renders a standing patient unable to constrict the splanchnic and other peripheral vascular beds in response to the pooling of blood volume (300 to 800 mL) owing to gravity.

Physiologic effects of sympathetic activation include pupillary dilation, increased heart rate and contractility, increased peripheral vascular resistance, bronchodilation, increased glandular secretions, decreased GI motility, increased sweating, decreased function of reproductive organs, and mobilization of energy substrates. The effects of parasympathetic activation include pupillary constriction, lacrimal and salivary secretion, decreased heart rate and contractility, bronchoconstriction, increased GI motility, and contraction of the detrusor muscle of the bladder. Sympathetic and parasympathetic responses, though generally antagonistic, are not always equally counterbalanced among organ systems.

Sympathetic preganglionic neurons, which use acetylcholine as their primary neurotransmitter, originate in the segmentally organized intermediolateral column of the spinal cord and exit via the ventral roots to pass through the white rami communicans and reach the paravertebral sympathetic chain ganglia, which innervate all organs and tissues except those of the abdomen and pelvis. The superior cervical ganglion, for example, innervates cranial structures, and the stellate ganglion innervates the upper limb. Sympathetic preganglionic axons also form the splanchnic nerves, which innervate the celiac, superior mesenteric, and hypogastric ganglia, as well as the adrenal medulla. With the exception of neurons that innervate the sweat glands, which are cholinergic, all other sympathetic postganglionic neurons are adrenergic neurons with norepinephrine as their primary transmitter.

Preganglionic innervation of parasympathetic neurons is also cholinergic; however, in contrast to their sympathetic counterparts, the major postganglionic parasympathetic neurotransmitter is acetylcholine. Parasympathetic preganglionic fibers originate in the Edinger-Westphal, salivatory, and vagal dorsal motor nuclei in the brain stem. The ciliary, sphenopalatine, otic, submandibular, sublingual, and pelvic ganglia send postganglionic parasympathetic fibers to their target organs. Cranial nerves IX and X, which constitute the afferent limbs of the baroreceptor reflex, relay beat-to-beat information about systemic arterial pressure to the nucleus of the solitary tract.

The GI tract contains neural plexuses, the most prominent of which are the myenteric (Auerbach) plexus found between the two layers of the muscularis externa and the submucosal (Meissner) plexus. Disorders of the enteric nervous system mainly affect GI motility or sphincter control rather than absorptive or secretory functions.

BRAIN DISORDERS

Important disease targets involved in central autonomic regulation include the interrelated neuronal cell groups of the hypothalamus, as well as the ventrolateral medulla, nucleus of the solitary tract, parabrachial nucleus, periaqueductal gray matter, amygdala, and insular and prefrontal cortices. Together, these relay systems and integrative centers compose the central autonomic network, which when impaired, fails to activate or modulate sympathetic and parasympathetic tone and neurohumoral responses.

A number of neurodegenerative disorders disrupt the central pathways of autonomic regulation. The most severe form is multiple system atrophy, a sporadic adult-onset disease in which severe autonomic failure accompanies and may precede parkinsonism (Shy-Drager syndrome; Chapter 409) or cerebellar ataxia (Chapter 410). In multiple system atrophy, degeneration of the striatum, pigmented nuclei, pontine nuclei, inferior olives, cerebellar Purkinje cells, and dorsal vagal and vestibular nuclei is associated with oligodendroglial cytoplasmic inclusions composed of α-synuclein filamentous aggregates. Autonomic dysfunction also occurs in dementia with Lewy bodies (Chapter 402) and to a lesser degree in Parkinson disease,[2] both of which also involve abnormal neuronal accumulation of α-synuclein.

Essential to the integration of behavioral, autonomic, and neuroendocrine responses is the tightly packed interwoven group of cells that constitute the hypothalamus. Lesions of the anterior hypothalamus may alter thirst perception and sodium regulation. Dysfunction of the magnocellular arginine vasopressin neurons of the supraoptic and supraventricular nuclei can be manifested as decreased secretion of antidiuretic hormone, thereby resulting in diabetes insipidus (Chapter 225) with hypovolemia, or as inappropriately increased secretion of antidiuretic hormone, thereby resulting in hyponatremia. Medial preoptic–anterior hypothalamic dysfunction associated with

dysgenesis of the corpus callosum (Shapiro syndrome) is characterized by episodic hyperhidrosis and hypothermia. Lesions of the posterior hypothalamus can result from hypothermia, Wernicke-Korsakoff syndrome (Chapter 416), acute traumatic brain injury (Chapter 399), multiple sclerosis (Chapter 411), mesodiencephalic hematoma, and toluene toxicity.

Catastrophic neurologic conditions such as subarachnoid hemorrhage (Chapter 408), head trauma (Chapter 399), status epilepticus (Chapter 403), and acute hydrocephalus with increased intracranial pressure can profoundly stimulate sympathetic responses with cardiovascular consequence. Release of the hypothalamus as a result of cortical inhibition is the presumed mechanism. These paroxysmal sympathetic storms (diencephalic syndrome) are characterized by episodic sympathetic hyperactivity with hypertension, tachycardia, hyperventilation, pupillary dilation, flushing, and diaphoresis. Communicating hydrocephalus, structural lesions affecting the medial frontal cortex, and degenerative conditions affecting the frontobasal ganglia can cause urinary incontinence with uninhibited bladder contractions.

Autonomic responses are closely linked to emotional states. Portions of the insular and anterior cingulate cortices mediate the autonomic responses to emotional stress. Because cardioregulatory function is represented within the insular cortex, insular strokes have been associated with destabilization of sympathoregulatory balance and occasional adverse cardiac events. Correlations of electroencephalography and electrocardiography have shown that seizures arising from the mesial temporal lobe may induce ictal tachycardia or, more rarely, bradycardia or asystole.

Lesions of the brain stem may also be manifested as autonomic dysfunction. Damage to the medulla oblongata may give rise to hypertension, orthostatic hypotension, or syncope. Medullary ischemia or compression can cause acute neurogenic hypertension (Cushing response). Lateral medullary infarction (Wallenberg syndrome) typically produces ipsilateral Horner syndrome and occasionally more extensive dysautonomia, including bradycardia, acute hypertension, supine hypotension, or central hypoventilation.

SPINAL CORD DISORDERS

Lesions of the spinal cord (Chapter 400), whether compressive, demyelinating, vascular, or neoplastic, commonly result in an overactive bladder (Chapter 26) with symptoms of frequency, urgency, and sometimes incontinence. Lesions that involve the sacral cord segments or cauda equina result in an underactive bladder with incomplete emptying, overflow incontinence, sphincter atonia, and sexual dysfunction.

After spinal cord injuries above the level of splanchnic sympathetic outflow at T5, sprouting of afferent fibers in the thoracolumbar dorsal horns and necrosis of the descending white matter connections to sympathetic preganglionic neurons result in autonomic dysreflexia. In these patients, strong peripheral sensory stimuli such as bladder or bowel distention can induce a reversible state of sympathetic hyperresponsiveness that may present with hypertension, diaphoresis, flushing, or headache.

PERIPHERAL GANGLIONOPATHIES AND NEUROPATHIES

Autonomic dysfunction can arise at the level of the autonomic ganglia or peripheral nerves (Table 418-1). Peripheral autonomic nerves are generally small in caliber and unmyelinated or thinly myelinated. Peripheral neuropathies that selectively involve small nerve fibers can cause various combinations of sensory, sympathetic, or parasympathetic signs and symptoms.

Among the peripheral dysautonomias is the syndrome of pure autonomic failure, which is defined as the insidious onset of severe generalized autonomic failure as the sole clinical feature in the absence of any signs of extrapyramidal, cerebellar, or sensory or motor peripheral nerve dysfunction. Pathologic studies have found Lewy body accumulation in autonomic ganglia, peripheral autonomic nerves, the substantia nigra, and the locus ceruleus.

Diabetic neuropathy involves autonomic nerves early in its course, and about 20% of patients advance to a clinically consequential cardiovascular autonomic neuropathy, which is a marker for and increased risk of cardiovascular mortality and morbidity.[3] The incidence increases with the duration of diabetes and advancing age. In diabetic patients, microvascular ischemia causes progressive peripheral nerve damage, although an autoimmune mechanism may also play a role in a subset of patients. Glycemic burden appears to represent a continuum of risk for autonomic neuropathy, because some patients with impaired glucose regulation or newly diagnosed diabetes already have evidence of small fiber neuropathy.

Other syndromes of neurologic autoimmunity (Table 418-2) include Guillain-Barré syndrome, in which antiganglioside antibodies mediate an acute inflammatory demyelinating polyradiculoneuropathy that may be associated with tachycardia, blood pressure lability, and pupillomotor, sudomotor, and vasomotor disturbances (Chapter 420). The syndrome of acute pandysautonomia typically develops dramatically over a period of days to weeks as combined sympathetic and parasympathetic failure with GI dysmotility and, in contrast to Guillain-Barré syndrome, sparing of the somatic nerves. An antecedent (presumably viral) infection is reported in about 50%

TABLE 418-1 SOME CAUSES OF PERIPHERAL AUTONOMIC NEUROPATHY

Metabolic	Diabetes mellitus Alcohol Acute intermittent porphyria Uremia
Autoimmune	Autoimmune autonomic ganglionopathy Guillain-Barré syndrome Morvan syndrome Lambert-Eaton myasthenic syndrome Chronic inflammatory demyelinating polyradiculoneuropathy Sjögren syndrome Systemic lupus erythematosus Mixed connective tissue diseases
Paraproteinemic	Amyloidosis
Nutritional	Cyanocobalamin deficiency Thiamine deficiency Gluten-sensitive neuropathy
Toxic	Heavy metals Organic solvents Organophosphates Vacor Acrylamide
Drug induced	Cisplatin Vincristine Amiodarone Metronidazole Perhexiline Paclitaxel
Infectious	Human immunodeficiency virus Leprosy Chagas disease Botulism Diphtheria Lyme disease
Genetic	Hereditary sensory and autonomic neuropathies Types I and II Type III (familial dysautonomia) Type IV (congenital insensitivity to pain) Type V Fabry disease
Idiopathic	Adie syndrome Ross syndrome Acute cholinergic neuropathy Chronic idiopathic anhidrosis Amyotrophic lateral sclerosis

TABLE 418-2 AUTOIMMUNE AUTONOMIC NEUROPATHIES

CLINICAL SYNDROME	ASSOCIATED AUTOANTIBODY
Autoimmune autonomic ganglionopathy	Anti-GAChR
Guillain-Barré syndrome	Anti-GM1, anti-GM3
Paraneoplastic autonomic neuropathy	ANNA-1 (anti-Hu), PCA-2, CRMP-5
Lambert-Eaton myasthenic syndrome	Anti-VGCC
Morvan syndrome	Anti-VGKC
Sjögren syndrome	SSA (anti-Ro), SSB (anti-La)
Intestinal dysmotility syndromes	Anti-GnRH

ANNA-1 = antineuronal nuclear antibody; anti-GAChR = nicotinic ganglionic acetylcholine receptor antibody; anti-GM1, anti-GM3 = antiganglioside antibody; anti-GnRH = gonadotropin-releasing hormone antibody; anti-VGCC = P/Q-type voltage-gated calcium-channel antibody; anti-VGKC = voltage-gated potassium-channel antibody; CRMP-5 = collapsin response mediator protein 5; PCA-2 = Purkinje cell cytoplasmic antibody type 2.

of cases. The finding of antibodies against the nicotinic acetylcholine receptor in the autonomic ganglia in many of these patients has established autoimmune autonomic ganglionopathy as a definable disease entity.[4] Ganglionic acetylcholine receptor antibodies occur in some patients with lung cancer or thymoma. Low levels of these antibodies have also been found in a subset of patients with isolated GI dysmotility.

Autoimmune neuromyotonia is characterized by peripheral nerve hyperexcitability, insomnia, fluctuating delirium, and prominent dysautonomia with hyperhidrosis and orthostatic intolerance. Most patients have antibodies to voltage-gated potassium channels (Chapter 422).

Paraneoplastic autonomic neuropathies, which can predate the diagnosis of malignancy, are a rare epiphenomenon of malignancy, most frequently small cell lung carcinoma (Chapter 191), and can also occur in association with ovarian carcinoma (Chapter 199), breast carcinoma (Chapter 198), thymoma, lymphoma (Chapter 185), and other cancers. The most commonly encountered paraneoplastic antibody is antineuronal nuclear antibody type 1 (ANNA-1 or anti-Hu), which binds to a 35- to 40-kD family of neuronal nuclear RNA-binding proteins, including those in autonomic and enteric ganglia. Antibodies against collapsing response mediator proteins (CRMP-5 or anti-CV2) have also been associated with paraneoplastic autonomic neuropathy. Dysautonomia occurs in approximately 10 to 30% of patients with ANNA-1 and in 30% of patients with CRMP-5 seropositivity. Small cell lung cancer (Chapter 191) has been found in more than 80% of patients seropositive for ANNA-1 or CRMP-5.

Hereditary sensory and autonomic neuropathy (HSAN) type I, which is due to mutations of the gene for serine palmitoyltransferase, is inherited in an autosomal dominant pattern and is characterized by distal anhidrosis with loss of nociceptive and thermal perception. Mutations in *HSN2* have been linked to HSAN type II, an autosomal recessive neuropathy that causes distal anhidrosis and sensory loss. Sural nerve biopsy specimens disclose virtual absence of myelinated fibers and decreased numbers of unmyelinated fibers. Autonomic involvement is most prominent in HSAN type III, commonly known as familial dysautonomia or Riley-Day syndrome. This autosomal recessive disorder, which affects 1 in 3600 live births in parents of Ashkenazi Jewish descent, has been linked to mutations in the I-κB kinase–associated protein gene (*IKBKAP*). Pathologic studies disclose severely depleted sympathetic preganglionic and postganglionic neuronal populations with preservation of parasympathetic neurons. Affected children cry without tears, feed poorly, lack lingual fungiform papillae, have depressed patellar reflexes, and are subject to orthostatic hypotension and autonomic storms owing to impaired baroreflex afferent neurons.

HSAN type IV, or congenital insensitivity to pain with anhidrosis, is an autosomal recessive disorder that is associated with mental retardation and repeated episodes of fever. HSAN type IV results from mutations in the gene for neurotrophic tyrosine kinase receptor type 1 (*NTRK1*), which is important for inducing neurite outgrowth in embryonic sensory and sympathetic neurons. Peripheral nerve biopsy discloses virtual absence of unmyelinated fibers.

HSAN type V is an autosomal recessive disorder characterized by loss of pain and thermal sensation. It is caused by mutations in the nerve growth factor beta subunit (*NGFB*) gene.

The age-related decline in baroreflex sensitivity, adrenergic responses to sympathetic activation, parasympathetic control of heart rate, esophageal and GI motility, and efficient thermoregulation predisposes the elderly to orthostatic, postprandial, and drug-induced hypotension. Pheochromocytomas (Chapter 228) cause paroxysmal autonomic symptoms, including blood pressure lability, and carcinoid syndrome (Chapter 232) causes flushing.

CLINICAL MANIFESTATIONS

Several distinct patterns of autonomic dysfunction help in identifying underlying causes.

Generalized Autonomic Failure

Early symptoms of adrenergic failure typically include lightheadedness on arising in the morning or following a warm shower, physical exercise, or a large meal. Other common symptoms include male erectile dysfunction, decreased sweating, dry mouth, constipation, and bladder dysfunction. Severe orthostatic hypotension without pulse acceleration (Chapter 62), which is the hallmark of severe generalized autonomic failure, occurs in at least 50% of patients and may be accompanied by supine and nocturnal hypertension, in which the normal diurnal decrease in blood pressure during sleep is reversed.

Acute Autonomic Syndromes

Acute or subacute manifestations of autonomic dysfunction may result from rapidly developing disease or from decompensation of chronic autonomic disease. An abrupt onset of new focal autonomic signs, particularly when accompanied by headache or motor or sensory deficits, should prompt a detailed neurologic assessment for an acute cerebral or spinal syndrome, which may be caused by vascular, traumatic, inflammatory, neoplastic, or infectious diseases.

Chronic Autonomic Syndromes

The clinical spectrum of chronic autonomic neuropathies includes distal small fiber neuropathies with a stocking-and-glove distribution of anhidrosis, often combined with loss of pain and temperature sensibility. Other symptoms include orthostatic hypotension and impaired exercise tolerance, with either an increased resting heart rate due to a parasympathetic cardiovagal neuropathy or a fixed heart rate that does not increase adequately in response to physiologic demands as a result of the involvement of sympathetic fibers. Some patients with distal sudomotor neuropathy will complain of spontaneous or gustatory proximal hyperhidrosis with episodic sweating involving the face, head, and upper part of the trunk.

Gastroparesis, which is a common feature of autonomic neuropathies, is characterized by delayed gastric emptying, which may be manifested as early satiety, nausea, anorexia, bloating, and sometimes pain and weight loss. Intestinal dysmotility may cause severe constipation (Chapter 136). In advanced peripheral neuropathies (Chapter 420), in ganglionopathies in which involvement is not length dependent, and in central degenerative disorders, the loss of sweating may extend to proximal body regions or globally. Widespread anhidrosis may result in impaired thermoregulatory sweating and the potential for hyperthermia in conditions of heat stress (Chapter 109).

Paroxysmal Dysautonomias

Paroxysmal and episodic autonomic symptoms include postprandial hypotension, which is a reduction in systolic blood pressure of at least 20 mm Hg within 2 hours of the start of a meal, typically one high in carbohydrate content. Postprandial hypotension may occur in elderly individuals with or without orthostatic hypotension.

Dysfunction of the afferent limb of the baroreflex system leads to volatile blood pressure in patients with acute inflammatory demyelinating polyneuropathy and syndromes of arterial baroreflex failure. Additional causes of aberrant activation of autonomic reflexes include epilepsy, diencephalic syndrome, subarachnoid hemorrhage, acute head trauma, pheochromocytoma, intoxication, drug withdrawal, neurally mediated syncope, and panic disorder.

Selective Autonomic Syndromes

Regional autonomic disorders are characterized by focal or system-selective autonomic dysfunction. An example is harlequin syndrome, in which heat stress, exercise, or sudden emotion in a patient with hemifacial cutaneous sympathetic denervation evokes a dramatic facial division in which the denervated half remains pale and dry and the intact half flushes red. Oculosympathetic paresis (Horner syndrome) may also be present. Harlequin syndrome may occur in patients with Holmes-Adie syndrome, which consists of tonic pupils with asymmetrical or absent tendon reflexes, and has been described in patients with Ross syndrome, a partial dysautonomia consisting of the clinical triad of unilateral or bilateral tonic pupils, tendon hyporeflexia, and segmental body anhidrosis.

Other Specific Syndromes

Baroreflex failure is commonly the result of damage to the carotid sinus baroreceptors or the glossopharyngeal nerves. Similarly, interruption of vagal input from the aortic arch baroreceptors to the nucleus of the solitary tract can impair baroreflex responses. Baroreflex failure also occurs in patients who have undergone surgery or irradiation of the neck. These patients will exhibit severe and labile hypertension with concomitant tachycardia, palpitations, headache, diaphoresis, and emotional lability.[5] Takotsubo cardiomyopathy (Chapter 60) is a stress-induced syndrome of typically reversible left ventricular myocardial dysfunction brought on by a catecholamine surge.

Serotonin syndrome develops within hours or days of the addition of a new serotonergic agent to a drug regimen that already enhances serotonergic neurotransmission, on overdose with a selective serotonin reuptake inhibitor, or from abuse of psychostimulants such as amphetamine, methamphetamine,

and 3,4-methylenedioxymetamphetamine (MDMA or "ecstasy"). Manifestations include agitation, hypervigilance, confusion, hyperthermia, increased sweating, fluctuating blood pressure, hyperreflexia, and myoclonus.

Neuroleptic malignant syndrome (Chapter 432) is a potentially life-threatening hypermetabolic condition that develops within days to weeks in 0.2% of patients who receive drugs that block dopamine 2 receptors. The clinical findings consist of hyperthermia, profuse sweating, muscle rigidity, bradykinesia, and delirium. If the offending medication is not withheld, the syndrome may progress to tachycardia, tachypnea, labile blood pressure, myoclonus, obtundation, and catatonia.

Among the infectious neuropathies, tetanus infection (Chapter 296) causes sympathetic overactivity in a third of patients because of the exotoxin tetanospasmin, which is taken up by peripheral nerve terminals and transported across synaptic junctions to reach the central nervous system. There it binds to gangliosides at presynaptic junctions to disinhibit preganglionic neurons and damages autonomic brain stem nuclei. Sympathetic hyperactivity results in labile or persistent hypertension or hypotension, tachyarrhythmias, peripheral vasoconstriction, fever, and profuse sweating. Diphtheritic neuropathy (Chapter 292) causes bulbar weakness and may be associated with cardiovagal impairment but not usually with orthostatic hypotension.

The acute cholinergic neuropathy of botulism (Chapter 296) occurs along with bulbar and generalized neuromuscular paralysis 12 to 36 hours after the ingestion of food contaminated with the gram-positive anaerobic bacterium *Clostridium botulinum*. Botulinum toxin binds with high affinity to presynaptic receptors of cholinergic nerve terminals and inhibits the release of acetylcholine, thereby blocking neuromuscular and cholinergic autonomic transmission. Autonomic manifestations include anhidrosis, dry eyes, dry mouth, paralytic ileus, gastric dilation, urinary retention, and sometimes orthostatic hypotension with fluctuating blood pressure and vasomotor tone.

Human immunodeficiency virus infection commonly causes autonomic disturbances, particularly in its advanced stages. Manifestations can include orthostatic hypotension, tachycardia, urinary dysfunction, impotency, diarrhea, and cardiac conduction defects. Perivascular mononuclear inflammatory infiltrates and neuronal degeneration in biopsy specimens of sympathetic ganglia suggest an autoimmune pathogenesis.

Chagas disease (Chapter 347) causes a predominantly parasympathetic neuropathy characterized by megaesophagus, megaduodenum, and megacolon, as well as sympathetic cardiovascular failure with cardiomegaly and conduction defects. The autonomic neuropathy has an autoimmune basis and develops years to decades following primary infection with *Trypanosoma cruzi*.

Leprosy (Chapter 326), one of the most common causes of neuropathy worldwide, frequently causes peripheral autonomic neuropathy as a result of an immune reaction against *Mycobacterium leprae*. Focal anhidrosis occurs in areas of hypopigmented and hypoesthetic skin. Cardiac denervation and orthostatic hypotension have been described.

Toxins known to cause autonomic neuropathy include the rodenticide Vacor, as well as thallium, arsenic, mercury, acrylamide, and organic solvents such as carbon disulfide and hexacarbon. Organophosphate poisoning (Chapter 110) induces miosis and copious secretions. Ergot poisoning from rye contaminated with the fungus *Claviceps purpurea* results in intense vasoconstriction, paresthesia, seizures, and diarrhea. Poisoning with muscarine, which is present in certain poisonous mushrooms (Chapter 110), results in increased salivation, sweating, and lacrimation followed by nausea, abdominal pain, and diarrhea.

Medicinal drugs that may induce an autonomic peripheral neuropathy include cisplatin, vincristine, amiodarone, metronidazole, perhexiline maleate, and paclitaxel. Many drugs are capable of increasing or decreasing sweating (Table 418-3).

Nutritional deficiencies that lead to autonomic dysfunction include alcoholic neuropathy, which is a dying-back neuropathy identical to that of beriberi that is caused by thiamine deficiency (Chapter 416). Distal parts of the vagus nerve are affected early, and orthostatic hypotension may occur in more advanced stages. Subacute combined degeneration from vitamin B_{12} deficiency (Chapter 218) results in axonal degeneration and is occasionally manifested as orthostatic hypotension. Autonomic neuropathy has been described in some cases of celiac disease (Chapter 140).

Amyloidosis (Chapter 188) results from the focal deposition of insoluble fibrillary proteins arranged in β-pleated sheet configurations within the extracellular space of various tissues, which may include the vasculature of peripheral autonomic nerves and sympathetic ganglia. Amyloid neuropathy is typically manifested as a painful distal small fiber sensory and severe autonomic neuropathy. autonomic dysfunction frequently occurs in primary AL (amyloid light-chain), immunoglobulin light chain–associated disease, and hereditary amyloidosis, but only rarely in reactive or AA (amyloid A) amyloidosis.

TABLE 418-3 SOME COMMONLY PRESCRIBED DRUGS THAT AFFECT SWEATING

Drugs that increase sweating
Opioids
Serotonin reuptake inhibitors
Anticholinesterases
Cholinergic agonists
Drugs that decrease sweating
M_3 anticholinergics
Carbonic anhydrase inhibitors
Tricyclic antidepressants
Neuroleptics
Antihistamines
Central α-adrenergic agonists
Botulinum toxin

Functional Dysautonomias

A functional autonomic disorder is a medical condition that impairs normal autonomic function in some way but in the absence of a known structural neurologic deficit. Examples include neurally mediated syncope (Chapter 62), irritable bowel syndrome (Chapter 137), and some forms of orthostatic intolerance and pain, the molecular basis of which await discovery. Syndromic features can help to disambiguate functional dysautonomias from psychosomatic disorders (Chapter 397), which also may manifest autonomic symptoms just as they may manifest sensory or motor symptoms.

DIAGNOSIS

Clinical evaluation of autonomic dysfunction begins with a careful history. It is important to distinguish chronic and stable conditions from progressive and episodic phenomena and to recognize the circumstances that provoke or modify symptoms. Orthostatic hypotension, for example, is typically worse in the morning and may be aggravated by dehydration, deconditioning, prolonged standing, physical exertion, heat, carbohydrate ingestion, or menstruation (Table 418-4).[6]

Bedside Evaluation

The skin examination should assess turgor, pallor, flushing, and acral cyanosis, as well as any asymmetry of sweating, which may be more palpable than visible. Signs of pupillary asymmetry, ptosis, mucosal dryness, distal sensory or reflex changes, bradykinesia, or rigidity should be noted.

Blood pressure and heart rate should be measured with the patient supine and again after standing for 1 to 3 minutes and correlated with symptoms. Orthostatic hypotension is defined as a reduction in systolic blood pressure of at least 20 mm Hg or a reduction in diastolic blood pressure of at least 10 mm Hg, with or without symptoms, within 1 to 3 minutes of assuming an erect posture.[7] Neurogenic orthostatic hypotension is typically sustained with continued standing. Measurements taken immediately on standing can be misleading because some healthy young persons without orthostatic hypotension will exhibit transient hypotension within 30 seconds of standing but then recover. Except in patients treated with β-blockers, orthostatic hypotension without reflex tachycardia is evidence of generalized adrenergic failure. If reflex tachycardia occurs, dehydration or excessive venous pooling should be considered.

Some patients with orthostatic intolerance experience an abnormal increase in heart rate rather than a drop in blood pressure on standing. Postural tachycardia syndrome[8] is defined as an increase in heart rate by more than 30 beats per minute in adults (40 beats per minute in adolescents) and to consistently greater than 120 beats per minute when standing.

Laboratory Evaluation

Appropriate laboratory testing depends on the type and distribution of autonomic dysfunction. Investigations may include a complete blood cell count, fasting glucose, electrolytes, morning cortisol, thyroid function testing, vitamin B_{12} level, and when indicated, autoimmune markers. Creatine kinase should be checked in patients with hyperthermia.

In a patient with an autonomic neuropathy, seropositivity for any of the characterized paraneoplastic autoantibodies should prompt a careful search for an underlying malignancy, even if the results of routine imaging studies

TABLE 418-4 GRADING OF ORTHOSTATIC INTOLERANCE

	SYMPTOM FREQUENCY	ACTIVITIES OF DAILY LIVING IN THE UPRIGHT POSTURE	STANDING TIME (ON MOST OCCASIONS)	ORTHOSTATIC BLOOD PRESSURE
Grade I	Infrequent orthostatic symptoms developing only under conditions of increased stress*	Unrestricted	>15 min	May or may not be abnormal
Grade II	Intermittent orthostatic symptoms occurring at least weekly	Some limitation	>5 min	Some changes in cardiovascular indices, e.g., oscillations, or decrease in pulse pressure by > 50%
Grade III	Frequent orthostatic symptoms occurring on most occasions	Marked limitation	>1 min	Orthostatic hypotension is present > 50% of the time, recorded on different days
Grade IV	Orthostatic symptoms are consistently present	Incapacitated and unable to stand without presyncope or syncope developing	<1 min	Orthostatic hypotension is severe and consistently present

*Conditions that increase orthostatic stress include dehydration, deconditioning from prolonged bedrest, physical exertion, heat stress, and medications that lower blood pressure or impair adrenergic function.
Adapted from Low PA, Singer W. Update on management of neurogenic orthostatic hypotension. *Lancet Neurol.* 2008;7:451-458.

are normal. Positron emission tomography is more sensitive than computed tomography in detecting small tumor foci. In patients with suspected pheochromocytoma, the most sensitive screening test is the free metanephrine level (Chapter 228).

Pupillary responses to the instillation of dilute pilocarpine, epinephrine, and cocaine can assist in the localization of oculosympathetic and oculoparasympathetic deficits. Lacrimal secretion may be quantified by the Schirmer and rose bengal tests (Chapter 268). Sialography may be useful to quantify salivary flow. Salivary gland biopsy may be necessary to diagnose Sjögren syndrome, particularly the seronegative form.

Manometry and scintigraphic studies are useful in the diagnosis of disorders of GI motility. Postvoiding residual volumes or urodynamic studies can elaborate patterns of urinary bladder dysfunction (Chapter 26). Suspected amyloid may require biopsy (Chapter 188).

Ambulatory blood pressure testing (Chapter 67), usually over a period of 24 hours, is useful to detect patterns of nocturnal hypertension, postprandial hypotension, and the labile hypertension of baroreflex failure. Adrenergic function can be assessed noninvasively by tilt table testing (Chapter 62). Adrenergic failure can also be defined by deficient recovery and overshoot of arterial pressure following 15 seconds of expiration at 40 mm Hg. The Valsalva ratio, defined as the maximum heart rate generated by the Valsalva maneuver divided by the lowest heart rate within 30 seconds of the peak, is a measure of parasympathetic cardiovagal function. A sensitive index of cardiovagal function is the heart rate response to sinusoidal deep breathing, which quantifies respiratory sinus arrhythmia.

Noninvasive tests of sudomotor function, such as the quantitative sudomotor axon reflex test, the Silastic sweat imprint, or the thermatoregulatory sweat test, can assess sweating function. These tests as well as skin biopsies to examine intraepidermal nerve fiber density can be useful in detecting autonomic involvement in small fiber neuropathies.[9,10]

TREATMENT

Rx

Treatment begins with educating patients about the underlying physiology, helping them avoid exacerbations, and managing their symptoms. In cases of mild dysautonomia, medications may not be needed. Elderly patients may be able to compensate for some of the age-associated decline in autonomic function through regular exercise.

Efforts at treating the underlying cause of an autonomic neuropathy should be pursued. Good control of glucose in patients with diabetes mellitus (Chapter 229) reduces the rate of complications, including neuropathy. Meticulous foot care can prevent cutaneous and joint trauma, ulceration, and infection of desensitized skin.

Orthostatic Intolerance

The goals of treatment are to increase the time the patient is able to stand without orthostatic symptoms developing, while simultaneously avoiding excessive recumbent hypertension. Mild orthostatic hypotension and syncope may respond to conservative measures such as increasing oral hydration (2 to 2.5 L/day), drinking sports beverages, and adding dietary salt or sodium tablets to increase daily salt intake to 10 to 20 g. Prolonged bedrest and medications that could potentially exacerbate orthostatic hypotension should be avoided if possible. Elevating the head of the bed by inserting 4- to 6-inch blocks under the head posts can improve orthostatic tolerance in some patients by reducing nocturnal natriuresis and stimulating release of renin.

Water bolus treatment (drinking 16 oz of water) can increase systolic blood pressure by 20 mm Hg for about 2 hours by a sympathetic reflex.[■■] The addition of either glucose or salt is counterproductive in that both attenuate the systemic vasoconstriction produced by pure water. Lower extremity resistance strength training combined with education about physical countermaneuvers (leg crossing, squatting, bending forward, or placing one foot on a chair) can help patients increase venous return to the heart and improve orthostatic tolerance by activating leg muscles.[■] Compressive stockings are effective if they are tightly fitting, especially if they provide abdominal compression in addition to leg compression.[■] However, they may be poorly tolerated in warm climates and are cumbersome to apply.

Pharmacologic measures are of variable benefit, and patients commonly stop taking them soon after they are prescribed.[11] Examples include midodrine (5 to 10 mg three times daily) to constrict capacitance vessels,[■] fludrocortisone (0.1 to 0.4 mg/day) to expand plasma volume and sensitize peripheral vascular α-adrenergic receptors,[■] and pyridostigmine (30 to 60 mg two or three times daily) to enhance ganglionic transmission during orthostatic stress.[■] Yohimbine (5.4 mg three times daily) improves orthostatic hypotension by engaging residual sympathetic tone.[■] Pyridostigmine has a more modest pressor effect but has the advantage of inducing less supine hypertension. Droxidopa (100 to 600 mg three times daily), an orally active synthetic precursor of norepinephrine, has recently been approved in the U.S. for the treatment of symptoms of neurogenic orthostatic hypotension.[■] Aldose reductase inhibitors (e.g., epalrestat [50 mg three times daily]) may provide some benefit for patients with diabetic cardiovascular autonomic neuropathy,[12] but no such drugs are currently approved by the U.S. Food and Drug Administration.

Postprandial hypotension may be managed by dividing meals to avoid large carbohydrate loads. Caffeine or midodrine with breakfast may be helpful.

Sweating Disorders

Initial therapy for palmar hyperhidrosis begins with tap water iontophoresis, in which a low-level electric current applied to the skin surface blocks sweat ducts at the level of the stratum corneum; for severe cases, endoscopic thoracic sympathotomy is effective. Focal hyperhidrosis, such as gustatory sweating caused by aberrant innervation of the facial sweat glands by regenerating parasympathetic fibers of the facial nerve, responds well to botulinum toxin injections. Generalized hyperhidrosis can be suppressed by oral anticholinergic drugs (e.g., glycopyrrolate, 1 to 2 mg one to three times daily), which tends to be better tolerated than other anticholinergic agents because very little crosses the blood-brain barrier); dry mouth is an invariable side effect. Aluminum chloride hexahydrate (20%) in anhydrous ethyl alcohol applied topically to dry skin at bedtime), oral belladonna (0.2 mg 1 to 2 times daily), propantheline (15 mg three times daily), topiramate (beginning at 25 mg twice daily), and clonidine (0.1 mg three times daily orally or by transdermal patch) may also be helpful for hyperhidrosis. Botulinum toxin types A and B, injected intradermally, are effective for axillary hyperhidrosis.

In a patient who is unable to sweat, hyperthermia cannot be prevented by drinking more water. Seeking shade, avoidance of exertion in hot weather, moistening the skin with a wet washcloth, and the use of portable fans can be effective. Carbonic anhydrase inhibitors, such as topiramate and zonisamide, and anticholinergic medications can inhibit thermoregulatory sweating and should be avoided if patients experience heat intolerance.

Hyperadrenergic Disorders

Nocturnal hypertension may be minimized by avoiding pressor agents within several hours of bedtime, elevating the head of the bed, or having a

nighttime snack. In severe cases, bedtime hydralazine (25 mg), nifedipine (10 mg), amlodipine (2.5 to 5 mg), or a nitroglycerin patch (0.1 mg/hour) may be needed. Paroxysmal and labile hypertension in patients with arterial baroreflex denervation may respond to clonidine (0.1 mg three times daily orally or by transdermal patch).

PROGNOSIS

The prognosis depends on the nature of the autonomic disorder. In general, the development of orthostatic hypotension worsens the prognosis.

A diagnosis of multiple system atrophy carries an estimated life expectancy of 7 to 9 years.[13] A patient with pure autonomic failure may have a prolonged and stable clinical course, but this syndrome may progress years later to a phenotype of multiple system atrophy or dementia with Lewy bodies. Amyloid neuropathy portends a median survival of less than 1 year if orthostatic hypotension is present. Diabetic autonomic neuropathy is associated with an approximately two-fold increased risk for silent myocardial ischemia and overall mortality.

Regional Sympathetic Dysfunction

Regional sympathetic dysfunction may accompany the pain that sometimes follows peripheral nerve injuries. For example, sympathetic activation can occur as a normal physiologic response to any painful state.

Complex regional pain syndrome is characterized by severe ongoing neuropathic pain that is disproportionate in intensity, duration, and distribution to the expected sequela of limb trauma.[14] In this syndrome, allodynia (pain in response to normally nonpainful stimuli, such as light touch or cold) or hyperalgesia (increased sensitivity to painful stimuli) accompany cutaneous vasomotor or sudomotor abnormalities. The vasomotor changes are manifested as vasodilation with a warm, red, or swollen limb, or alternatively as vasoconstriction with a cold pale limb. The sudomotor findings range from regional hyperhidrosis to anhidrosis. Regional dystrophic changes, such as dry atrophic skin, sparse or coarse hair, brittle nails, and osteopenia, may also develop. Although sympathetic dysfunction may be pronounced, it does not appear to cause the pain. The mechanisms of pain and sympathetic dysfunction in this condition are incompletely understood and may result from crosstalk among aberrantly regenerated peripheral nerve fibers, expression of new α-adrenergic receptors on sensory nerve fibers and sweat glands, release of substance P and pro-inflammatory peptides at the site of injury, and sensitization of pain-mediating structures at multiple levels within the central nervous system.

Mobilization of the affected limb is of paramount importance in the early treatment of complex regional pain syndrome.[15] A primary goal of analgesic medication or regional anesthesia in early treatment is to facilitate participation in physical therapy. No therapy is supported by consistent data from properly sized and conducted randomized trails. Small trials report improvement with the use of bisphosphonates (e.g., alendronate, 40 mg orally or 7.5 mg intravenously daily), steroids (e.g., prednisone, 40 mg daily, or methylprednisolone, 8 mg four times daily initially and then tapered), dimethyl sulfoxide (50% cream one to four times daily), epidural clonidine (300 to 700 µg daily), intrathecal baclofen (25 to 75 µg daily), and epidural spinal cord stimulation. Intravenous immunoglobulin (0.5 g/kg) has also been used with variable results.

Grade A References

A1. Shannon JR, Diedrich A, Biaggioni I, et al. Water drinking as a treatment for orthostatic syndromes. *Am J Med*. 2002;112:355-360.

A2. Young TM, Mathias CJ. The effects of water ingestion on orthostatic hypotension in two groups of chronic autonomic failure: multiple system atrophy and pure autonomic failure. *J Neurol Neurosurg Psychiatry*. 2004;75:1737-1741.

A3. Winker R, Barth A, Bidmon D, et al. Endurance exercise training in orthostatic intolerance: a randomized, controlled trial. *Hypertension*. 2005;45:291-298.

A4. Podoleanu C, Maggi R, Brignole M, et al. Lower limb and abdominal compression bandages prevent progressive orthostatic hypotension in elderly persons: a randomized single-blind controlled study. *J Am Coll Cardiol*. 2006;58:1425-1432.

A5. Izcovich A, González Malla C, Manzotti M, et al. Midodrine for orthostatic hypotension and recurrent reflex syncope: a systematic review. *Neurology*. 2014;83:1170-1177.

A6. Ong AC, Myint PK, Shepstone L, et al. A systematic review of the pharmacological management of orthostatic hypotension. *Int J Clin Pract*. 2013;67:633-646.

A7. Logan IC, Witham MD. Efficacy of treatments for orthostatic hypotension: a systematic review. *Age Ageing*. 2012;41:587-594.

A8. Shibao C, Okamoto LE, Gamboa A, et al. Comparative efficacy of yohimbine against pyridostigmine for the treatment of orthostatic hypotension in autonomic failure. *Hypertension*. 2010;56:847-851.

A9. Kaufmann H, Freeman R, Biaggioni I, et al. Droxidopa for neurogenic orthostatic hypotension: a randomized, placebo-controlled, phase 3 trial. *Neurology*. 2014;83:328-335.

A10. O'Connell NE, Wand BM, McAuley J, et al. Interventions for treating pain and disability in adults with complex regional pain syndrome. *Cochrane Database Syst Rev*. 2013;4:CD009416.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Tarvainen MP, Laitinen TP, Lipponen JA, et al. Cardiac autonomic dysfunction in type 2 diabetes—effect of hyperglycemia and disease duration. *Front Endocrinol (Lausanne)*. 2014;5:130.
2. Kim JB, Kim BJ, Koh SB, et al. Autonomic dysfunction according to disease progression in Parkinson's disease. *Parkinsonism Relat Disord*. 2014;20:303-307.
3. Dimitropoulos G, Tahrani AA, Stevens MJ. Cardiac autonomic neuropathy in patients with diabetes mellitus. *World J Diabetes*. 2014;5:17-39.
4. Koike H, Watanabe H, Sobue G. The spectrum of immune-mediated autonomic neuropathies: insights from the clinicopathological features. *J Neurol Neurosurg Psychiatry*. 2013;84:98-106.
5. Briasoulis A, Silver A, Yano Y, et al. Orthostatic hypotension associated with baroreceptor dysfunction: treatment approaches. *J Clin Hypertens (Greenwich)*. 2014;16:141-148.
6. Feldstein C, Weder AB. Orthostatic hypotension: a common, serious and underrecognized problem in hospitalized patients. *J Am Soc Hypertens*. 2012;6:27-39.
7. Freeman R, Wieling W, Axelrod FB, et al. Consensus statement on the definition of orthostatic hypotension, neurally mediated syncope and the postural tachycardia syndrome. *Auton Neurosci*. 2011;161:46-48.
8. Benarroch EE. Postural tachycardia syndrome: a heterogeneous and multifactorial disorder. *Mayo Clin Proc*. 2012;87:1214-1225.
9. Myers MI, Peltier AC. Uses of skin biopsy for sensory and autonomic nerve assessment. *Curr Neurol Neurosci Rep*. 2013;13:323.
10. Haensch CA, Tosch M, Katona I, et al. Small-fiber neuropathy with cardiac denervation in postural tachycardia syndrome. *Muscle Nerve*. 2014;50:956-961.
11. Shibao C, Grijalva CG, Lipsitz LA, et al. Early discontinuation of treatment in patients with orthostatic hypotension. *Auton Neurosci*. 2013;177:291-296.
12. Hu X, Li S, Yang G, et al. Efficacy and safety of aldose reductase inhibitor for the treatment of diabetic cardiovascular autonomic neuropathy: systematic review and meta-analysis. *PLoS OnE*. 2014;9:e87096.
13. Wenning GK, Geser F, Krismer F, et al. The natural history of multiple system atrophy: a prospective European cohort study. *Lancet Neurol*. 2013;12:264-274.
14. Littlejohn G. Complex regional pain syndrome. *Rheumatology (Oxf)*. 2014;53:1157-1158.
15. Harden RN, Oaklander AL, Burton AW, et al. Complex regional pain syndrome: practical diagnostic and treatment guidelines, 4th edition. *Pain Med*. 2013;14:180-229.

REVIEW QUESTIONS

1. A 38-year-old woman seeks evaluation for impaired concentration, severe insomnia, excessive sweating, fluctuating blood pressure, and painful muscle cramps. Polysomnography shows complete absence of sleep during the recording. Electromyography shows hyperexcitability of peripheral nerves. Computed tomography of the chest shows an anterior mediastinal mass. Autoantibodies to which of the following receptors are most likely to be found in this patient?

A. Nicotinic ganglionic acetylcholine receptor
B. Nicotinic acetylcholine receptor
C. Voltage-gated sodium channel
D. Voltage-gated calcium channel
E. Voltage-gated potassium channel

Answer: E This patient has typical clinical features of Morvan syndrome, which is an autoimmune disorder characterized by central, autonomic, and peripheral nerve hyperactivity. As many as 50% of cases have a thymoma. The characteristic antibody is to the voltage-gated potassium channel. By contrast, antibodies to the voltage-gated calcium channel are associated with Lambert-Eaton myasthenic syndrome. Antibodies to the nicotinic acetylcholine receptor are associated with myasthenia gravis, and antibodies to the ganglionic receptor are associated with autoimmune autonomic ganglionopathy. Antibodies to voltage-gated sodium channels are associated (depending on the specific protein) with epilepsy, hyperkalemic periodic paralysis, paramyotonia congenita, long QT syndrome, Brugada syndrome, and idiopathic ventricular fibrillation.

2. A 48-year-old previously healthy man taking no medications presents with a 1-year history of progressive fatigue following exercise, dizziness when showering, and inability to sustain an erection. His blood pressure is 136/68 mm Hg supine, 130/62 mm Hg seated, and 90/54 mm Hg standing, and his heart rate is 70 beats per minute in all positions. Which of the following laboratory tests would be useful in evaluating this patient's adrenergic failure?

A. Plasma level of free metanephrine
B. Blood pressure response to carotid sinus massage
C. Heart rate response to periodic deep breathing
D. Blood pressure response to forced expiration
E. Heart rate response to forced expiration

Answer: D This patient has orthostatic hypotension, which is likely neurogenic because it is severe and without a compensatory tachycardia. Patients with autonomic failure experience orthostatic hypotension because the sympathetic nervous system is unable to increase peripheral vascular resistance, which is regulated by postganglionic adrenergic neurons, in response to the downward pooling of venous blood when standing. In addition to upright tilt table testing, blood pressure recovery and overshoot in response to the Valsalva maneuver (which consists of forced expiration of 40 mm Hg for 15 seconds) is useful in evaluating and quantifying adrenergic function. The heart rate response to the Valsalva maneuver and to periodic deep breathing assess cardiovagal rather than adrenergic function. Carotid sinus massage may be useful, not in assessing the integrity of adrenergic function but rather in evaluating reflex syncope with bradycardia. Plasma (or urine) free metanephrine levels are useful in detecting pheochromocytoma, which is a neuroendocrine tumor of the adrenal medulla.

3. A 62-year-old man with a 10-year history of diabetes mellitus has recently developed lightheadedness upon standing when he gets out of bed each morning. He has not had vertiginous sensations of motion or rotation. His only medication is metformin. Glucose by fingerstick during symptoms was 166 mg/dL. Neurologic examination finds absent Achilles tendon reflexes and decreased sensation to pinprick below the mid-calves. Blood pressure measured supine is 170/88 mm Hg, with heart rate 74 beats per minute. After 2 minutes of standing, blood pressure is 132/72 mm Hg without symptoms, and heart rate is 76 beats per minute. Which of the following treatment recommendations is most appropriate?

A. β-Blocker metoprolol 50 mg twice daily
B. Tamsulosin 0.4 mg each morning
C. Meclizine
D. A glass of orange juice, 8 oz. each morning
E. Bolus water drinking, 16 oz. each morning

Answer: E This patient's postural lightheadedness is most probably caused by orthostatic hypotension, which is defined as a reduction in systolic blood pressure of at least 20 mm Hg or a reduction in diastolic blood pressure of at least 10 mm Hg, with or without symptoms, within 1 to 3 minutes of assuming a standing posture. Symptoms such as lightheadedness can be variably present. In this case, the most likely cause of orthostatic hypotension is an autonomic neuropathy, because signs of sensory and motor peripheral neuropathy are also present. An established treatment for orthostatic hypotension is bolus water drinking, which induces an osmopressor reflex mediated through the sympathetic nervous system. Water drinking is effective only if it is hypo-osmotic (unlike fruit juices which are hyperosmotic). Although micturition presyncope might be a consideration in this patient if his symptoms were to occur while voiding and if his prostate were enlarged, tamsulosin, which is an α-adrenergic antagonist used to treat benign prostatic hyperplasia can worsen orthostatic hypotension. β-Blockers, which are useful in treating hypertension, can also worsen orthostatic hypotension. The antihistamine meclizine can be useful in treating benign paroxysmal positional vertigo, which is a different form of dizziness characterized by a rotational sensation following head movement and is typically not associated with orthostatic hypotension.

4. A 27-year-old man who is paraplegic following traumatic injury to his spine at the level of T2 complains of recurrent headaches. The headaches occur nearly daily, persist for 30 to 60 minutes, and are accompanied by facial sweating, flushing, and elevations in blood pressure as high as 180/110 mm Hg, but no visual symptoms. The patient rates the intensity of the headaches as 9 on a scale of 0 to 10. Ibuprofen and hydrocodone have provided no relief. Computed tomography of the brain is normal. Which of the following is the most appropriate recommendation?

A. Suprapubic ultrasound to assess for a full bladder
B. Begin hydrochlorothiazide to treat hypertension
C. Begin midodrine to treat orthostatic hypotension
D. Urine 5-hydroxyindoleacetic acid to screen for carcinoid
E. Urine metanephrine to screen for pheochromocytoma

Answer: A This patient has autonomic dysreflexia, which is a paroxysmal disorder of excessive sympathetic outflow. Autonomic dysreflexia, which can follow spinal cord injuries above the level of T5, is triggered by any strong visceral or peripheral sensory stimulus, which the patient may or may not be able to perceive. The most appropriate treatment is identification and resolution of the sensory stimulus, such as a distended bladder, distended bowel, or skin irritation. Pharmacologic treatment of the hypertension without removing its source would be unlikely to succeed and could cause hypotension in between hypertensive episodes. Midodrine, which is useful in treating orthostatic hypotension, can worsen, not alleviate, hypertension. Pheochromocytoma also produces paroxysmal sympathetic surges but is quite rare, whereas autonomic dysreflexia is common in patients with spinal cord injury. Carcinoid syndrome also produces flushing, typically with diarrhea, but not hypertension.

5. A 42-year-old premenopausal woman reports frequent, diffuse, and embarrassing excessive sweating involving her face, neck, chest, and back. The sweating occurs not only in response to exercise and in hot weather but also at rest and in cool environments. Her medications are timolol eye drops for glaucoma, bupropion for anxiety, and oxycodone for chronic low back pain. Her blood glucose level, thyroid function, and complete blood count are normal. Which of the following is the most appropriate treatment recommendation for the patient's complaint?

A. Oral anticholinergic agent
B. Change or eliminate current medications
C. Replace fluid lost by drinking more water
D. Botulinum toxin subcutaneous injections
E. Tap water iontophoresis

Answer: B This patient has either essential hyperhidrosis or drug-induced hyperhidrosis or both. As the known side effects of two of her current medications—the serotonin reuptake inhibitor bupropion and the opioid oxycodone—include increased sweating, reassessment of the need for continuation of those medications or consideration of switching to alternative agents would be the most direct approach to resolving her hyperhidrosis. Although an oral anticholinergic agent could be used to reduce sweating (acetylcholine is the neurotransmitter at the eccrine neuroeffector junction), anticholinergic agents are contraindicated in glaucoma because they increase intraocular pressure. Drinking more water, while appropriate to replace fluid lost if the sweating is profuse or frequent, would not improve the patient's symptom. Botulinum toxin injections and tap water iontophoresis are appropriate treatments for some regional forms of hyperhidrosis, but they are not indicated for generalized hyperhidrosis; it would be impracticable to treat such a wide area of skin.

419 AMYOTROPHIC LATERAL SCLEROSIS AND OTHER MOTOR NEURON DISEASES

PAMELA J. SHAW

DEFINITION

The motor neuron diseases (Table 419-1) are a heterogeneous group of disorders in which selective loss of function of upper motor neurons, lower motor neurons, or both results in impairment of the nervous system's control of voluntary movement. The most common acquired motor neuron disease, amyotrophic lateral sclerosis (ALS), is a combined upper and lower motor neuron disorder. The features of lower motor neuron involvement are muscle wasting, fasciculations, and flaccid weakness, with normal or depressed tendon reflexes. Upper motor neuron dysfunction may cause increased muscle tone, clonus, weakness in a pyramidal distribution, and extensor plantar responses. Recent advances in the molecular genetics of hereditary motor neuron diseases have improved their classification and enhanced the careful diagnosis that is essential for genetic counseling, guidance, treatment, and advising patients about prognosis.

AMYOTROPHIC LATERAL SCLEROSIS

EPIDEMIOLOGY

ALS is a neurodegenerative disorder that causes progressive injury and cell death of lower motor neurons in the brain stem and spinal cord, as well as upper motor neurons in the motor cortex. ALS has an incidence of about 2 per 100,000 and a prevalence of 6 to 8 per 100,000. The global incidence is fairly uniform, with the exception of a few high-incidence foci such as the Western Pacific island of Guam. The disease affects predominantly middle-aged and elderly individuals, with a mean age at onset of 55 to 60 years, although younger individuals can also be affected. Increasing age, male sex (male/female ratio ≈ 1.6 : 1), and genetic susceptibility are the only proven risk factors, although ongoing research is assessing the effects of athleticism/physical exercise and other potential environmental risk factors. Approximately 90% of cases of ALS occur sporadically, but 5 to 10% are familial, usually with an autosomal dominant mode of inheritance.

PATHOBIOLOGY

The process of neuronal degeneration in ALS is complex. The subtype of disease caused by *SOD1* mutations accounts for 20% of familial ALS cases and 2% of ALS overall. Mutant *SOD1* appears to trigger a complex interplay of multiple pathogenic processes, including oxidative stress, protein aggregation, mitochondrial dysfunction, excitotoxicity, and impaired axonal transport.[1] Non-neuronal cells in the vicinity of motor neurons may contribute importantly to neuronal injury. Genetically engineered mouse models of *SOD1*-related ALS have shown that normal astrocytes can protect motor neurons expressing mutant *SOD1* and that removing the expression of mutant *SOD1* from microglia or astrocytes slows the progression of disease in these murine models. Astrocytes expressing mutant *SOD1* exert toxic effects on neighboring motor neurons through as yet undefined mechanisms.[2]

Familial motor neuron disease has been linked to mutations in multiple genes, including *SOD1,* alsin, senataxin, angiogenin, *VAPB*, dynactin, *TARDBP*, and *FUS/TLS*. Accumulating evidence suggests that defective RNA processing likely plays a key role in the pathogenesis of ALS. A very important recent discovery is the finding of GGGGCC hexanucleotide

TABLE 419-1 CLASSIFICATION OF MOTOR NEURON DISORDERS

COMBINED UPPER AND LOWER MOTOR NEURON DISORDERS
Amyotrophic lateral sclerosis
Familial adult onset
Familial juvenile onset
Sporadic
ALS-plus syndromes
ALS with frontotemporal dementia
Western Pacific ALS–parkinsonism-dementia complex
UPPER MOTOR NEURON DISORDERS
Primary lateral sclerosis
Hereditary spastic paraplegias
Neurolathyrism
Konzo
LOWER MOTOR NEURON DISORDERS
Hereditary
Spinal muscular atrophies (SMAs)
Proximal autosomal recessive SMA (associated with *SMN* mutations) types I to IV
Other forms of SMA not associated with *SMN* mutations
Distal spinal muscular atrophies/hereditary motor neuronopathies
Kennedy disease (X-linked spinobulbar neuronopathy)
Hexosaminidase deficiency (GM2 gangliosidosis)
Acquired
Monomelic focal and segmental spinal muscular atrophies
Multifocal motor neuropathies
Acute motor axonal neuropathy (AMAN)
Postpolio syndrome
Postirradiation syndrome
Infective disorders
Acute poliomyelitis
West Nile fever
Other viral infections (e.g., enterovirus 71 and rabies virus)
Human immunodeficiency virus–associated motor neuron disorder
Lyme disease
Creutzfeldt-Jakob disease (amyotrophic forms)
DISORDERS OF THE BULBAR MOTOR SYSTEM
Kennedy disease (X-linked bulbospinal neuronopathy)
Brown-Vialetto-Van Laere syndrome
Fazio-Londe disease
TOXIC DISORDERS OF THE MOTOR NEURON
Neurolathyrism
Konzo
Heavy metal toxicity (lead, mercury)
Western Pacific ALS–parkinsonism-dementia complex
Postirradiation motor neuron injury
DISORDERS OF MOTOR NEURON OVERACTIVITY
Neuromyotonia
Stiff person syndrome
MISCELLANEOUS MOTOR NEURON DISORDERS
Endocrinopathies (e.g., hyperthyroidism, hyperparathyroidism, hypoglycemia)
Copper deficiency syndrome
Benign cramp-fasciculation syndrome

intronic expansions in the chromosome 9 *C9ORF72* gene. Such changes account for up to 40% of familial ALS cases, and up to 7% of sporadic ALS.[3] *SMN1* gene duplications are associated with a two-fold increased risk of sporadic ALS. In the sporadic disease, associations have been reported with alterations in at least eight other genes.

Pathology

At autopsy, the gross pathologic features of ALS consist of atrophy of the cerebral precentral gyrus, as well as sclerosis and pallor of the corticospinal tracts of the spinal cord. Thinning of the hypoglossal nerves and ventral spinal roots may be observed, and muscle atrophy is obvious. Microscopically, ALS patients will typically have lost at least 50% of their spinal motor neurons and have diffuse astrocytic gliosis in the spinal gray matter. By comparison, motor neurons in Onuf nucleus in the sacral spinal cord (which innervate the pelvic floor muscles) and the motor nuclei of cranial nerves III, IV, and VI (which control eye movements) are relatively preserved. A cardinal feature in residual motor neurons is the presence of ubiquitinated proteinaceous inclusions, which may be compact or skein-like. TDP-43 has been recognized as a major protein constituent of these aggregates. In the motor cortex, there is variable loss of upper motor neurons and astrocytic gliosis. In the descending corticospinal tracts, axonal loss, myelin pallor, and gliosis are seen. The atrophied skeletal muscle shows clusters of angular atrophic fibers and fiber-type grouping that results from serial denervation and reinnervation. The selectivity of the disease process for the motor system is now recognized to be relative, and involvement of extramotor parts of the central nervous system can be found, especially in the sensory and spinocerebellar pathways, substantia nigra neurons, and dentate granule cells in the hippocampus. In the newly described ALS variant caused by C9ORF72 expansions, the characteristic extramotor system pathology demonstrates cerebellar and hippocampal inclusions that are P62+ and TDP-43-negative by immunostaining. Some of these inclusions comprise dipeptide proteins generated by aberrant translation of the G4C2 repeats.[4]

CLINICAL MANIFESTATIONS

ALS is characterized by a combination of upper and lower motor neuron degeneration. Lower motor neuron degeneration causes weakness, atrophy, and fasciculation of the limb and bulbar musculature. Features of upper motor neuron dysfunction include the incongruous presence of active or brisk tendon reflexes in a wasted limb, increased muscle tone, and sometimes the presence of Babinski sign. Upper motor neuron bulbar disease causes pseudobulbar palsy, with emotional lability, a brisk jaw jerk, slowing of repetitive tongue movements, and strained effortful speech. Fatigue and weight loss are also common symptoms. With end-stage disease, most patients will have features of upper and lower motor neuron dysfunction affecting all four limbs and the bulbar musculature.

In approximately 75% of patients, the disease starts distally, focally, and asymmetrically in an upper or lower limb (Video 419-1), followed by progressive spread of injury in an anatomically logical progression to contiguous groups of motor neurons. Affected individuals may notice weakness, wasting or clumsiness of one hand, or unilateral footdrop. Muscle cramps may precede other clinical features, and fasciculations are most noticeable in the large proximal limb muscles. In the upper limbs, the thenar and intrinsic hand muscles tend to be severely affected, whereas the triceps and finger flexors are relatively spared until late in the disease. In the lower limbs, the pattern of weakness is often in a pyramidal distribution (flexors weaker than extensors), with early weakness of hip flexion and ankle dorsiflexion and severe involvement of the distal muscles.

Bulbar symptoms, which are the initial feature in approximately 25% of patients, are especially common in elderly women with ALS (Video 419-2). The first problem is usually slurring of speech, initially apparent only when the individual is tired. Patients often have a mixed spastic/flaccid dysarthria in which speech develops a tight strangled quality because of the upper motor neuron component, with a superimposed nasal quality as a result of the flaccid lower motor neuron weakness of the palate and nasopharynx. In patients with bulbar disease, examination often reveals weakness of the facial muscles; a spastic, weak, wasted, and fasciculating tongue; and a brisk jaw jerk. Dysphagia, initially more pronounced for liquids than for solids, usually follows the dysarthria within a few weeks or months (Video 419-3). Complications include weight loss and prolonged and arduous meal times with frequent episodes of coughing, drooling of saliva, and aspiration pneumonia.

Respiratory muscle weakness is rarely the initial feature of ALS. More commonly, respiratory muscle weakness develops insidiously and causes dyspnea and orthopnea. Diaphragmatic weakness may be apparent from the paradoxical movement of the abdominal wall during inspiration and a marked decline in forced vital capacity in the supine position. Symptoms of nocturnal carbon dioxide retention may develop, including interrupted sleep, morning headaches, anorexia, and daytime somnolence.

Neck muscle weakness, which is common later in the course of disease, causes difficulty holding the head upright (dropped head syndrome). Eye movements tend to be spared even in advanced disease, thereby permitting limited communication by movements of the eyes. Similarly, the strength of the pelvic floor muscles is relatively preserved, so patients with ALS usually remain continent throughout the course of the disease.

Overt features of frontotemporal dementia (Chapter 402), with progressive deterioration in personality and behavior, will develop in approximately 5% of patients with ALS. Cognitive dysfunction may precede, follow, or coincide with the features of motor dysfunction. Up to 50% of ALS patients

VIDEO 419-1. **Limb symptoms and signs.**

VIDEO 419-2. **Bulbar symptoms and signs.**

VIDEO 419-3. **Normal swallowing.**

without overt dementia may show more subtle features of frontal lobe dysfunction.[5] The C9ORF72-ALS variant causes both ALS and/or frontotemporal dementia (Chapter 402), and patients with this subtype of ALS are more likely to have cognitive disturbances as well as a family history of dementia or psychosis.[6]

About 5 to 10% of ALS patients have the progressive muscular atrophy variant with clinical features reflecting only degeneration of lower motor neuron groups in the spinal cord. In primary lateral sclerosis, patients have pure upper motor neuron degeneration. Although severe spastic spinobulbar paresis ultimately develops in these patients, the duration of survival is commonly 10 to 15 years after the onset of symptoms. The progressive bulbar palsy variant usually progresses to involve the limbs, although limb signs may not be present initially.

Several ALS variants follow a more segmental pattern than is typical in ALS. Up to 10% of patients with ALS have flail arm syndrome, which is more common in men and is associated with a longer median survival than seen in those with typical ALS. A similar focal manifestation in the lower limbs, flail leg syndrome, is another recognized segmental variant.

DIAGNOSIS

The diagnosis of ALS is essentially clinical, and there is no specific diagnostic test. Diagnosis requires evidence of lower motor neuron degeneration by clinical, electrophysiologic (Chapter 396), or neuropathologic examination; upper motor neuron degeneration by clinical examination; and progressive spread of symptoms or signs within a region or to other regions, as determined by the history or examination. The diagnosis also requires the absence of other disease processes as determined by electrophysiologic testing, neuroimaging, and (if performed) biopsy. Generally accepted criteria (Table 419-2) classify patients as having definite, probable, or possible ALS. However, a number of other conditions may mimic ALS (Table 419-3), and

TABLE 419-2 AWAJI-SHIMA CONSENSUS CRITERIA FOR DIAGNOSING AMYOTROPHIC LATERAL SCLEROSIS

The diagnosis of ALS requires:
1. Evidence of LMN loss (reduced interferential pattern on full contraction and increased firing rate)
2. Evidence of reinnervation (motor units of large amplitude and longer duration)
3. Fibrillation and sharp waves or fasciculation potentials (fibrillation and sharp waves are required in weak limb muscles)

Number of muscles affected by region:
- Cervical and lumbar-sacral region: a minimum of 2 muscles innervated by different roots and nerves
- Bulbar and thoracic region: a minimum of 1 muscle

Diagnostic classification: Awaji-Shima Consensus Recommendations and the Revised El Escorial Criteria
- Clinically definite ALS:
  - clinical or electrophysiologic evidence of the presence of LMN as well as UMN signs in the bulbar region and at least 2 spinal regions or the presence of LMN and UMN signs in 3 spinal regions.
- Clinically probable ALS:
  - clinical or electrophysiologic evidence of LMN and UMN signs in at least 2 regions, with some UMN signs necessarily rostral to (above) the LMN signs.
- Clinically possible ALS:
  - clinical or electrophysiologic signs of UMN and LMN dysfunction are found in only 1 region, or UMN signs are found alone in ≥ 2 regions, or LMN signs are found rostral to UMN signs.

ALS = amyotrophic lateral sclerosis; LMN = lower motor neuron; UMN = upper motor neuron.

UMN signs: clonus, Babinski sign, absent abdominal reflexes, hypertonia, loss of dexterity.
LMN signs: atrophy, weakness. If only fasciculation, search with EMG for active denervation.
Regions reflect segmental motor neuron pools: bulbar, cervical, thoracic, and lumbosacral.

Adapted from Costa J, Swash M, de Carvalho M. Awaji criteria for the diagnosis of amyotrophic lateral sclerosis: a systematic review. *Arch Neurol.* 2012;69:1410-1416.

TABLE 419-3 DISORDERS THAT CAN MIMIC AMYOTROPHIC LATERAL SCLEROSIS/MOTOR NEURON DISEASE

FORM OF MOTOR NEURON DISEASE	MIMIC SYNDROMES	CLINICAL CLUES
Progressive muscular atrophy (PMA)/LMN-predominant phenotype	Multifocal motor neuropathy	Weakness out of proportion to wasting. Neurophysiology identifies conduction block. Anti-GM1 antibodies may be raised.
	Kennedy disease	Gynecomastia, distal sensory features, perioral fasciculation, indolent progression
	Spinal muscular atrophy	SMA can be adult onset. Pure LMN syndrome. Slower progression than PMA. Probably no family history.
	Chronic idiopathic demyelinating polyneuropathy	Electrophysiology identifies peripheral nerve demyelination.
	Benign cramp-fasciculation syndrome	Predominantly middle-aged men. Largely calf involvement. Failure to progress. No active denervation on EMG.
	Postpolio syndrome	Pure LMN syndrome. Past history of an illness compatible with poliomyelitis. Indolent progression.
	Lead poisoning	Extramotor clinical features, e.g., constipation, nail and buccal signs
	Acute motor axonal neuropathy (AMAN—a Guillain-Barré syndrome variant)	Acute onset, with progression ceasing after a few weeks. Nerve conduction studies show features of motor axonopathy.
	Hereditary motor neuropathies	Pure LMN syndrome. Family history, clinical signs indicating chronicity, slower rate of progression.
	Porphyria	Extramotor clinical features, family history, episodic exacerbations
	Compressive focal motor neuropathies	Pure motor disorders can result from compression of the deep palmar branch of the ulnar nerve and posterior interosseous branch of the radial nerve. Failure to extend beyond territory of one nerve. Electrophysiology with or without imaging helpful.
Amyotrophic lateral sclerosis	Multilevel spinal cord and root compression by discs, osteophytes, or tumor	Sensory symptoms and pain are common. UMN signs often caudal to LMN signs
	Thyrotoxicosis	Systemic symptoms and signs
	Combined peripheral neuropathy and cervical myelopathy	MRI of the spine and electrophysiology will differentiate
	Inclusion body myositis	Rarer than ALS. Characteristic pattern of weakness with early involvement of the long finger flexors and quadriceps.
	Paraneoplastic syndromes, especially lymphoma	History of malignancy or systemic features
	Sjögren syndrome	Non–motor-related symptoms
	Radiation myelopathy	History of radiotherapy
	Structural lesions of the bulbar region (e.g., tumor of the tongue base)	Pain, failure of features to extend outside the bulbar territory
Primary lateral sclerosis	Hereditary spastic paraplegia	Family history. Symptoms rarely extend beyond the lower limb territory. Prominent bladder dysfunction.
	Multiple sclerosis	Non–motor-related symptoms and signs (e.g., eye, bladder, cerebellar, and sensory involvement)
	Spinal cord compression by disc or tumor	Pain and sensory involvement usually present

ALS = amyotrophic lateral sclerosis; EMG = electromyography; LMN = lower motor neuron; MRI = magnetic resonance imaging; SMA = spinal muscular atrophy; UMN = upper motor neuron.

about 8% of patients in whom ALS is initially diagnosed have other lower motor neuron syndromes, such as multifocal motor neuropathy with conduction block, Kennedy disease, or mixed spinal cord and root compression (Chapter 400).[7] Conversely, 10 to 15% of patients in whom ALS is ultimately diagnosed may first undergo inappropriate surgery for presumed spinal cord or root compression abnormalities.

Blood tests that may be helpful in distinguishing ALS from mimic syndromes (see Table 419-3) include a complete blood count and serum calcium level, thyroid function tests, serum protein electrophoresis, Venereal Disease Research Laboratory test, creatine kinase level, inflammatory markers (erythrocyte sedimentation rate and C-reactive protein), and levels of anti-GM1 ganglioside and anti–myelin-associated glycoprotein (MAG) antibodies. Further testing, which is guided by the patient's clinical findings, might include acetylcholine receptor antibody; mutation screening in patients with familial disease, suspected Kennedy disease, or spinal muscular atrophy (SMA); heavy metal screening; urinary porphyrins; serum hexosaminidase A and B levels; *Borrelia* titers; and testing for human immunodeficiency virus.

Typical features of ALS on electromyography (EMG) include evidence of active denervation (i.e., positive sharp waves, fibrillation, and fasciculation potentials) and chronic denervation, as evidenced by large motor unit potentials that cannot be explained by a single nerve, root, or plexus lesion. Neuroimaging of the brain and spinal cord is usually needed to exclude structural pathology.

Baseline respiratory function tests should be performed on all patients. Muscle biopsy is indicated only in atypical cases when diagnostic uncertainty persists.

TREATMENT

ALS is best managed in specialized centers that offer multidisciplinary care. Teams typically include a neurologist, nurse specialist, occupational therapist, physical therapist, speech and language therapist, and dietitian. During the course of the disease, patients frequently require referral for placement of a gastrostomy tube and to provide respiratory support.

No therapy currently has a dramatic effect in slowing the progression of ALS. Riluzole, a sodium channel blocker whose primary mechanism of action is to reduce excitotoxicity through inhibition of presynaptic glutamate release, prolongs survival by approximately 3 months when given at 50 mg twice daily. It may cause fatigue, nausea, and dizziness, but these effects are frequently transient. Liver function tests should be performed at baseline and monthly for the first 3 months of therapy. Other trials of potential neuroprotective therapies have so far proved negative. New experimental approaches include the use of gene therapy and antisense oligonucleotide technology to reduce the expression of disease-causing genes,[8] and cell-based therapy aimed primarily at providing a supportive environment to prolong the survival of endogenous motor neurons.[9]

Good clinical care must focus on symptoms and preservation of independence and quality of life. In patients with progressive bulbar problems, optimal positioning, attention to food and fluid consistency, and protective swallowing techniques are helpful. If weight loss continues, high-calorie nutritional supplements are added between meals. In ALS patients with dysphagia due to upper motor neuron impairment of the upper esophageal sphincter, local injection of botulinum toxin type A can significantly improve the dysphagia and may represent an alternative to percutaneous endoscopic gastrostomy. Placement of a gastrostomy tube via endoscopy or under radiologic guidance is recommended in patients in whom dehydration, weight loss of 10 to 15%, frequent distressing choking episodes, prolonged and tiring mealtimes, or aspiration pneumonia develop. Tube placement is higher risk in patients with respiratory insufficiency. Tubes should ideally be placed before the patient's forced vital capacity falls below 50% of expected. Some evidence suggests that radiologically guided gastrostomy insertion may be safer in frail patients in the late stages of ALS.[10]

Respiratory muscle weakness, which can develop insidiously during the course of ALS, causes breathlessness, orthopnea, daytime somnolence, morning headaches, and interrupted sleep. Management must emphasize detection and prevention of aspiration pneumonia, assistance in clearing of secretions by agents to reduce saliva production (e.g., anticholinergic drugs such as glycopyrrolate, 1 to 2 mg 3 to 4 times daily, or intrasalivary botulinum toxin injections), providing a suction machine, use of a mucolytic agent such as carbocisteine (in a dose of up to 750 mg three times daily), adoption of a semi-upright position for sleep, and aggressive antibiotic therapy for chest infection (Chapters 96 and 97). A small dose of sublingual lorazepam (0.5 to 1 mg) may be useful if the dyspnea is accompanied by extreme anxiety; opiate therapy (e.g., morphine, diamorphine, fentanyl [Chapter 30, Table 30-4]) may be given orally, transdermally, or by subcutaneous infusion to relieve respiratory distress during the later stages of the disease.

As respiratory function worsens, noninvasive ventilation can alleviate symptoms of chronic hypoventilation, significantly improve quality of life, and prolong survival, especially in patients with orthopnea, daytime hypercapnia, and nocturnal oxygen desaturation. Full 24-hour ventilation via a tracheostomy is an option that is chosen uncommonly by fully informed patients. Ongoing clinical research is evaluating the value of cough assist devices and diaphragm pacing, as well as the optimal way to manage respiratory symptoms at the end of life. Palliative care teams and hospices can contribute substantially to the care of ALS patients in the later stages of the disease. In the absence of ventilatory support, ALS patients will almost always die in their sleep from hypercapnic coma. In the terminal phases (Chapter 3), the aim of treatment is to ensure comfort by prescribing opiates and anxiolytic medications as required to alleviate discomfort or distress.

PROGNOSIS

Clinical features associated with a worse prognosis include older age at onset of symptoms, early compromise of respiratory function, bulbar symptoms, and more rapid presentation to medical attention. The mean duration from the onset of symptoms to death in patients with sporadic ALS ranges from 27 to 43 months. The average 5-year survival rate is 25%, and approximately 5% of patients will survive for more than 10 years. The usual cause of death is respiratory failure, which may be accompanied by bronchopneumonia.

SPINAL MUSCULAR ATROPHIES

DEFINITION

The term *spinal muscular atrophy* encompasses a group of pure lower motor neuron disorders that cause progressive symmetrical muscle weakness and wasting. Because the bulbar musculature may be affected, an alternative term, *hereditary motor neuronopathy*, has been proposed. The time of onset is variable and ranges from in utero to adult life.

EPIDEMIOLOGY AND PATHOBIOLOGY

The most common type of SMA is caused by mutations in the survival motor neuron (*SMN*) gene and is inherited as an autosomal recessive disorder. The estimated carrier frequency of an *SMN* mutation is 1 in 50. Type 1 SMA (Werdnig-Hoffmann disease) has an incidence of 1 in 8000 births. SMA is divided into subtypes I to IV according to age at onset and severity of the phenotype.

The human *SMN* gene on chromosome 5q13 exists in two forms, with 5–base pair differences between *SMN1* and its centromeric homologue *SMN2*. A change in exon 7 of *SMN2* leads to skipping of exon 7, and as a result, 80% of the protein encoded by *SMN2* is truncated and nonfunctional rather than full length. The majority of patients with SMA have homozygous absence of *SMN1* exon 7, but *SMN1* may be replaced by a copy of *SMN2* during DNA replication by a process known as gene conversion. An individual may have one to four copies of *SMN2*, with a proportional increase in the amount of full-length SMN protein. A molecular basis for the wide variation in the phenotypic severity of SMA, which can range from in utero onset (SMA type I) to adult onset (SMA type IV), is the number of copies of *SMN2* and the SMN protein levels, although other disease-modifying factors have also been implicated.

The SMN protein oligomerizes and associates with other proteins to form the SMN complex, which in turn has an important role in the assembly of spliceosomal small nuclear ribonucleoproteins that have a function in pre-mRNA splicing in the nucleus. These cellular processes are ubiquitous, so either the clinical features of SMA may be caused by a particular susceptibility of lower motor neurons to defects in RNA processing or SMN may have functions that are specific to the motor neuron, including axonal transport of mRNA molecules essential for the health of the distal axon. Recent work has highlighted the role of dysregulation of ubiquitin homeostasis and β-catenin signaling, as well as genes involved in motor neuron synaptogenesis in the pathophysiology of SMA.[11]

At autopsy, patients with SMA have atrophic spinal cords with loss of α-motor neurons and evidence of motor neuron degeneration and gliosis. The ventral roots are atrophic, and muscle atrophy is apparent with microscopic evidence of denervation and reinnervation.

CLINICAL MANIFESTATIONS

Type I SMA (Werdnig-Hoffman disease) is characterized by severe generalized muscle weakness and hypotonia at birth or by the age of 6 months;

affected children never sit or walk. Type II is an intermediate form with an onset of muscle weakness before the age of 18 months; patients can sit but are never able to walk unaided. Type III SMA (Wohlfart-Kugelberg-Welander disease) appears after 18 months of age; patients acquire the ability to stand and walk but often become wheelchair dependent in adolescence or adult life, although life expectancy is normal. Patients with type IV SMA have an onset of muscle weakness in adult life.

DIAGNOSIS

The diagnosis of SMA caused by changes in *SMN* can be made by genetic testing in a patient with appropriate clinical signs and symptoms; 95% of affected individuals have *SMN* deletions. Prenatal diagnosis is available. Electrophysiology and muscle biopsy reveal evidence of denervation.

Other disorders can present in infancy or childhood as hypotonia, and a pattern of weakness similar to *SMN*-related SMA can be distinguished by associated clinical features such as early respiratory distress or vocal cord paralysis, or an atypical distribution of motor features such as upper limb– or lower limb–predominant or scapuloperoneal involvement. The etiologic relationship of these disorders to classic SMA can be clarified by testing for *SMN* mutations.

It is important to distinguish SMA type I from infantile botulism, which can have a similar initial clinical picture. EMG with high-frequency repetitive nerve stimulation shows a decrement in botulism, and testing for the presence of botulinum toxin can confirm the diagnosis (Chapter 296). SMA II and SMA III can be distinguished from chronic inflammatory demyelinating polyneuropathy (Chapter 420) by the presence of normal cerebrospinal fluid protein and normal nerve conduction studies in SMA. Patients with SMA type III can have clinical features that are similar to those of the hereditary motor and sensory neuropathies, but it can be distinguished by neurophysiologic assessment and genetic testing.

TREATMENT AND PROGNOSIS

Rx

No disease-modifying treatment of SMA is currently available,[A1] although experimental approaches to upregulate expression of the SMN protein, enhance SMN2 exon 7 inclusion, or replace the deficient SMN protein using gene therapy are being actively explored.[12] In a randomized trial of children with SMA II and SMA III, initial results have indicated that oral olesoxime (10 mg/kg/day), which preserves mitochondrial function in stressed cells, slows the progressive loss of motor function. Another experimental approach is to alter splicing of the *SMN2* pre-mRNA to produce a functional SMN protein.[13] Such children may benefit from passive and active physical therapy, lightweight braces, surgical correction of scoliosis, and respiratory support measures.

Patients with type I SMA usually die by the age of 18 months, patients with type II typically survive into adolescence, and patients with type III and type IV have a normal life expectancy.

SPINOBULBAR MUSCULAR ATROPHY/KENNEDY DISEASE

EPIDEMIOLOGY

Kennedy disease, or spinobulbar muscular atrophy (SBMA), is an X-linked degenerative disorder of the lower motor neurons. Though rare, it is important not to miss the diagnosis because of the genetic implications for the family and a more benign course than occurs with ALS. The diagnosis should be considered in any male patient with a pure lower motor neuron disorder, particularly when the disease course is relatively indolent, gynecomastia is present, or there is evidence of a mild accompanying sensory neuropathy.

PATHOBIOLOGY

SBMA is a trinucleotide repeat disorder in which a CAG expansion encodes for a polyglutamine tract in the first exon of the androgen receptor gene on chromosome Xq11-12. The androgen receptor, which contains three functional domains, is transported to the nucleus, where it binds to DNA and acts as a transcription factor. Expansion of the polyglutamine tract results in reduced target gene transactivation, and neurodegeneration occurs when the polyglutamine tract reaches a critical length of approximately 40 repeats. The neurodegeneration in patients with SBMA is considered to result from a ligand-dependent toxic gain of function of the mutant androgen receptor protein. Complete loss of its function, as seen in testicular feminization syndrome (Chapter 233), does not lead to motor neuron degeneration. The toxicity has not been fully characterized, but protein aggregation, impairment of protein degradation pathways, disruption of gene transcription, impairment of axonal transport, and neurotrophic factor signaling may all contribute.

Pathologic examination reveals mild spinal cord atrophy with ventral horn gliosis and loss of α-motor neurons. Misfolding of the polyglutamine (Q)-expanded protein leads to the formation of nuclear inclusions that contain the amino-terminal epitopes of the mutant androgen receptor within motor neurons and certain non-neuronal tissues.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

The mean age at onset of SBMA is 30 years, with a range of 15 to 60 years, and the severity of the disease and its age at onset correlate with the size of the repeat expansion. Initial symptoms consist of hand tremors, fasciculations, and muscle cramps, followed by progressive weakness and atrophy of the limb and bulbar muscles. Limb muscle weakness tends to be proximal and predominantly involves the lower limbs. There are no clinical signs of upper motor neuron dysfunction. Weakness of the lower facial and tongue muscles causes dysarthria, and jaw weakness may cause the mouth to hang open. The presence of perioral fasciculations with quivering of the chin is a characteristic feature. Pharyngeal involvement can cause dysphagia, and respiratory muscle weakness causes breathlessness. Mild distal sensory loss is frequently present in the lower limbs. Features of mild androgen insensitivity are frequent: gynecomastia, testicular atrophy, and erectile dysfunction. Heterozygous female carriers of SBMA may show mild clinical manifestations of the disease.

EMG and muscle biopsy, which are often performed because the creatine kinase level tends to be elevated, reveal evidence of chronic denervation. Genetic screening for the CAG repeat expansion in exon 1 of the androgen receptor gene is diagnostic.

TREATMENT AND PROGNOSIS

Rx

Because there are no established disease-modifying therapies for SBMA, current therapy consists of supportive care to prevent complications.[14] Recent human trials of the luteinizing hormone–releasing analogue leuprorelin and clenbuterol have not produced conclusive evidence of improvement. The course of the disease is slowly progressive in comparison to ALS and is compatible with normal life expectancy, although a proportion of patients may die of respiratory failure. Patients may become wheelchair dependent over a period of 2 to 3 decades, but some remain ambulatory until late in life.

Grade A References

A1. Miller RG, Mitchell JD, Moore DH. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *Cochrane Database Syst Rev*. 2012;3:CD001447.

A2. Restivo DA, Casabona A, Nicotra A, et al. ALS dysphagia pathophysiology: differential botulinum toxin response. *Neurology*. 2013;80:616-620.

A3. Bourke SC, Tomlinson M, Williams TL, et al. Effects of non-invasive ventilation on survival and quality of life in patients with amyotrophic lateral sclerosis: a randomized controlled trial. *Lancet Neurol*. 2006;5:140-147.

A4. Wadman RI, Bosboom WM, van der Pol WL, et al. Drug treatment for spinal muscular atrophy type I. *Cochrane Database Syst Rev*. 2012;4:CD006281.

A5. Wadman RI, Bosboom WM, van der Pol WL, et al. Drug treatment for spinal muscular atrophy types II and III. *Cochrane Database Syst Rev*. 2012;4:CD006282.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Ferraiuolo L, Kirby J, Grierson AJ, et al. Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nat Rev Neurol*. 2011;7:616-630.
2. Haidet-Phillips AM, Hester ME, Miranda CJ, et al. Astrocytes from familial and sporadic ALS patients are toxic to motor neurons. *Nat Biotechnol*. 2011;29:824-828.
3. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*. 2011;72:245-256.
4. Mori K, Weng SM, Arzberger T, et al. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science*. 2013;339:1335-1338.
5. Phukan J, Pender NP, Hardiman O. Cognitive impairment in amyotrophic lateral sclerosis. *Lancet Neurol*. 2007;6:994-1003.
6. Devenney E, Hornberger M, Irish M, et al. Frontotemporal dementia associated with the C9ORF72 mutation: a unique clinical profile. *JAMA Neurol*. 2014;71:331-339.
7. Hardiman O, van den Berg LH, Kiernan MC. Clinical diagnosis and management of amyotrophic lateral sclerosis. *Nat Rev Neurol*. 2011;7:639-649.
8. Miller TM, Pestronk A, David W, et al. An antisense oligonucleotide against SOD1 delivered intrathecally for patients with SOD1 familial amyotrophic lateral sclerosis: a phase 1, randomised, first-in-man study. *Lancet Neurol*. 2013;12:435-442.
9. Feldman EL, Boulis NM, Hur J, et al. Intraspinal neural stem cell transplantation in amyotrophic lateral sclerosis: phase 1 trial outcomes. *Ann Neurol*. 2014;75:363-373.
10. Stavroulakis T, Walsh T, Shaw PJ, et al. Gastrostomy use in motor neurone disease (MND): a review, meta-analysis and survey of current practice. *Amyotroph Lateral Scler Frontotemporal Degener*. 2013;14:96-104.
11. Wishart TM, Mutsaers CA, Riessland M, et al. Dysregulation of ubiquitin homeostasis and beta-catenin signaling promote spinal muscular atrophy. *J Clin Invest*. 2014;124:1821-1834.
12. Zanetta C, Nizzardo M, Simone C, et al. Molecular therapeutic strategies for spinal muscular atrophies: current and future clinical trials. *Clin Ther*. 2014;36:128-140.
13. Naryshkin NA, Weetall M, Dakka A, et al. Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science*. 2014;345:688-693.
14. Grunseich C, Rinaldi C, Fischbeck KH. Spinal and bulbar muscular atrophy: pathogenesis and clinical management. *Oral Dis*. 2014;20:6-9.

REVIEW QUESTIONS

1. A 53-year-old woman presents with a 12-month history of progressive wasting and weakness of the right hand. She had been aware of muscle cramps and muscle twitching affecting the proximal upper limb muscles. In the preceding 3 months, she had developed mild slurring of speech. Neurologic examination revealed a combination of upper and lower motor neuron signs affecting the bulbar and upper limb regions. Neurophysiologic examination revealed normal nerve conduction but evidence of active denervation in the bulbar, upper limb, and lower limb territories, as well as the thoracic paraspinal muscle. A diagnosis of upper limb onset amyotrophic lateral sclerosis was made. Which of the following additional clinical features may be present?

A. Emotional lability
B. Distal sensory loss in the limbs
C. Restriction of conjugate eye movements
D. Urinary incontinence
E. Pain in the right arm

Answer: A Emotional lability is commonly present in amyotrophic lateral sclerosis (ALS), especially in patients with upper motor neuron and bulbar signs. Sensory loss and pain are not expected to be present in patients with ALS and should be considered "red flags" to prompt consideration of alternative diagnoses. Pain may arise later in the disease course (e.g., development of a frozen shoulder secondary to immobility). Abnormalities of eye movements and pelvic floor muscles are rarely seen in patients with ALS, because the motor neurons within the cranial nerve motor nuclei (controlling eye movements) and Onuf nucleus in the sacral spinal cord are less vulnerable to the disease process compared to other groups of motor neurons.

2. A 62-year-old man presented with a 6-month history of progressive lower limb weakness, starting initially with right-sided footdrop. There was a family history of neurologic disease in that his older brother had died from ALS 5 years earlier after a 2-year disease course. The patient's mother had developed dementia at the age of 58, and the family had observed a marked personality change and behavioral disturbances during the 4-year disease course. Neurologic examination and testing were consistent with the diagnosis of ALS. Genetic testing showed the presence of a pathologic hexanucleotide expansion in intron 1 of the *C9ORF72* gene. Which of the following statements are correct?

A. The pathologic changes in the central nervous system of this patient are likely to be identical to those present in most cases of sporadic ALS.
B. The patient's condition is likely to be linked to the dementia disorder that developed in his mother.
C. The pattern of inheritance of this disorder is likely to be autosomal recessive.
D. Riluzole is not effective as a disease-modifying therapy in this subtype of ALS.
E. *C9ORF72* mutations are a less common cause of ALS than *SOD1* mutations.

Answer: B *C9ORF72*-related ALS is an autosomal dominant disorder (with incomplete penetrance). This gene change can also cause frontotemporal dementia, the features of which were present in the patient's mother. Patients with C9ORF72-related ALS are more likely to develop cognitive disturbance and have a family history of dementia or psychosis. It is unknown whether riluzole has a disease modifying effect in this subgroup of patients. The pathology of C9ORF-related ALS is characteristic, with protein inclusions that stain for P62, but which are negative for TDP-43 in neurons outside the motor system (e.g., cerebellum, hippocampus).

3. A 47-year-old man was diagnosed with lower limb–onset ALS 2 years ago. He now reports that he has not been sleeping well in the previous 2 months and that he has experienced loss of appetite and weight loss, though his speech and swallowing function remained normal. In the preceding 2 weeks, he noticed that he was tending to wake up with a bifrontal morning headache that would resolve approximately 30 minutes after getting out of bed. Respiratory assessment showed that his forced vital capacity (FVC) was 70% of predicted for his age and height. Which of the following management plans would be most appropriate?

A. Advise the patient to lie flat in bed at night, using only one pillow.
B. Gastrostomy tube insertion because of the weight loss
C. Nighttime sedation to improve the quality of sleep
D. Overnight oximetry followed by a trial of noninvasive ventilation
E. Administration of analgesia for the morning headaches and referral for repeat brain imaging to exclude the presence of an intracranial space-occupying lesion

Answer: D This patient has classic symptoms of neuromuscular respiratory failure. Initially the weakness of the respiratory muscles tends to cause problems during the night, when REM sleep may further compromise the function of weak respiratory muscles and the lying position may impair the function of a weak diaphragm. Intermittent hypoxia causes nocturnal restlessness and interruption of sleep, and CO_2 retention causes loss of appetite and morning headaches. Diaphragmatic weakness is *more* symptomatic when the patient attempts to lie flat, causing orthopnea. The patient should be advised to sleep in a more upright position. Overnight oximetry will be helpful in confirming the presence of nocturnal derangement of the blood gases. Nighttime sedation should *not* be used, because it may cause suppression of the respiratory drive and further compromise of respiratory function. If the presence of abnormal blood gases is confirmed, the patient may well benefit from the use of noninvasive ventilation, both in terms of quality of life and life expectancy. Gastrostomy tube insertion can be a helpful symptomatic measure in the presence of dysphagia but would not normally be considered for anorexia in the absence of difficulty with swallowing.

4. A 40-year-old man presents with a 2-year history of tremor of his hands, as well as muscle twitching, cramps, and slowly progressive weakness of the lower limb muscles. He also has mild slurring of his speech. Clinical examination reveals perioral and thigh muscle fasciculation and symmetrical proximal lower limb weakness with normal muscle tone and reflexes and flexor plantar responses. He also has mild distal sensory loss to light touch and pinprick in the lower limbs and a degree of gynecomastia. The neurophysiologic assessment shows evidence of neurogenic changes, predominantly in the lower limb and tongue muscles. Which additional investigation is likely to prove most useful in establishing the diagnosis in this patient?

A. Muscle biopsy
B. Genetic analysis of the *C9ORF72* gene
C. MRI of the lumbosacral spine
D. Cerebrospinal fluid examination
E. Genetic analysis of the androgen receptor gene

Answer: E This male patient had evidence of a pure lower motor neuron disorder that predominantly involved the lower limb and bulbar muscles at the time of presentation. The clinical clues (or red flags) that Kennedy disease might be the correct diagnosis include: the relatively indolent course of disease progression, the presence of distal sensory loss in the lower limbs that would not be commonly found in other motor neuron disorders such as ALS, and the presence of gynecomastia (a manifestation of androgen insensitivity). The diagnosis is established by detecting the presence of a CAG expansion of more than 40 repeats in the first exon of the androgen receptor gene located on chromosome Xq11-12.

420 PERIPHERAL NEUROPATHIES

MICHAEL E. SHY

APPROACH TO PERIPHERAL NEUROPATHIES

DEFINITION AND PATHOBIOLOGY

Peripheral neuropathy is a general term for disorders affecting peripheral nerves. The peripheral nervous system consists of motor, sensory, and autonomic neurons that extend outside the central nervous system (CNS) and are ensheathed by Schwann cells or ganglionic satellite cells. The peripheral nervous system includes the dorsal and ventral spinal roots, spinal and cranial nerves, sensory and motor terminals, and part of the autonomic nervous system. Motor neurons extend from their cell body in the ventral horn of the spinal cord to the neuromuscular junctions at the muscle that they innervate. The cell bodies of primary sensory neurons lie outside the spinal cord in the dorsal root ganglia, where they extend peripherally to specialized sensory end organs, including nociceptors, thermoreceptors, and mechanoreceptors. Central projections from dorsal root ganglia enter the spinal cord through the dorsal roots. At each spinal segment, the ventral roots, which carry motor axons, and the dorsal roots, which carry sensory axons, join to form mixed sensorimotor nerves. In the cervical, brachial, and lumbosacral areas, the mixed spinal nerves form plexuses from which arise the major anatomically defined limb nerves. Each mixed nerve is composed of large numbers of myelinated and nonmyelinated nerve fibers of varying diameter. The large myelinated axons include motor neurons and large fiber sensory nerves that mediate position and vibration sense. Small, thinly myelinated and nonmyelinated axons primarily provide nociception and autonomic functions. Preganglionic sympathetic autonomic fibers begin in the intermediolateral column of the spinal cord and synapse in ganglia of the sympathetic trunk. Preganglionic parasympathetic fibers travel long distances from their cell bodies in the brain stem or sacral spinal cord to reach terminal ganglia near the organs that the parasympathetic fibers innervate.

CLINICAL MANIFESTATIONS

Symptoms of peripheral neuropathy include weakness, sensory loss, abnormal balance, and autonomic dysfunction. Weakness is often distal and more severe in the legs than the arms. Deep and superficial muscles that are innervated by the peroneal nerve, such as the tibialis anterior and peroneus brevis and longus muscles, often cause more symptoms than do the plantar flexion muscles innervated by the tibial nerve, such as the gastrocnemius. As a result, tripping on a carpet or curb and ankle sprains are frequent symptoms. In the hands, symptoms typically involve fine movements, such as using buttons or zippers and inserting and turning keys in locks. Cramps, the painful knotting of a muscle, frequently occur with motor or sensorimotor neuropathies.

The sensory symptoms of neuropathy reflect disease of small, thinly myelinated or nonmyelinated fibers subserving pain and temperature, as well as large myelinated fibers subserving position sense. Common symptoms of small fiber sensory neuropathy include feeling as though the feet are "walking on pebbles" or "ice cold" and difficulty determining whether bath water is hot or cold with the foot. Painful dysesthesias, such as feeling as though the feet are "on fire," "on hot coals," or "stuck with pins," are also associated with small fiber abnormalities. Similar symptoms occur less frequently in the hands because most neuropathies are dependent on the length of the nerves; as a general rule, sensory symptoms appear in the hands after sensory symptoms in the legs have progressed up to the knee. An exception is when the patient also has carpal tunnel syndrome, which causes pain and tingling in the hands and can awaken patients from sleep. Large fiber sensory loss usually impairs balance, which may be worse at night when vision cannot overcome the loss of proprioception. Loss of proprioception is also frequently length dependent, so a patient may improve balance by lightly touching a wall with the hand to improve proprioceptive input to the brain.

Autonomic symptoms are frequent in neuropathies associated with diabetes or amyloidosis and include urinary retention or incontinence, abnormalities of sweating, constipation alternating with diarrhea, and lightheadedness when standing. Impotence is frequent.

DIAGNOSIS

Systematic Approach to Patients with Peripheral Neuropathy

Evaluation begins with the history and physical examination to demonstrate peripheral nerve disease and proceeds to neurophysiologic testing to characterize whether the process is demyelinating or axonal. Other specific tests are then ordered (Fig. 420-1). Peripheral neuropathies usually affect both motor and sensory nerves, causing both weakness and sensory loss. However, certain neuropathies are predominantly sensory, such as diabetes, or motor, such as multifocal motor neuropathy (Tables 420-1 and 420-2). Most neuropathies are symmetrical and length dependent. Pronounced asymmetries in symptoms suggest specific disorders, such as mononeuritis multiplex or hereditary neuropathy with liability to pressure palsies. It is also useful to know whether symptoms are acute (<1 month), subacute (<6 months), or

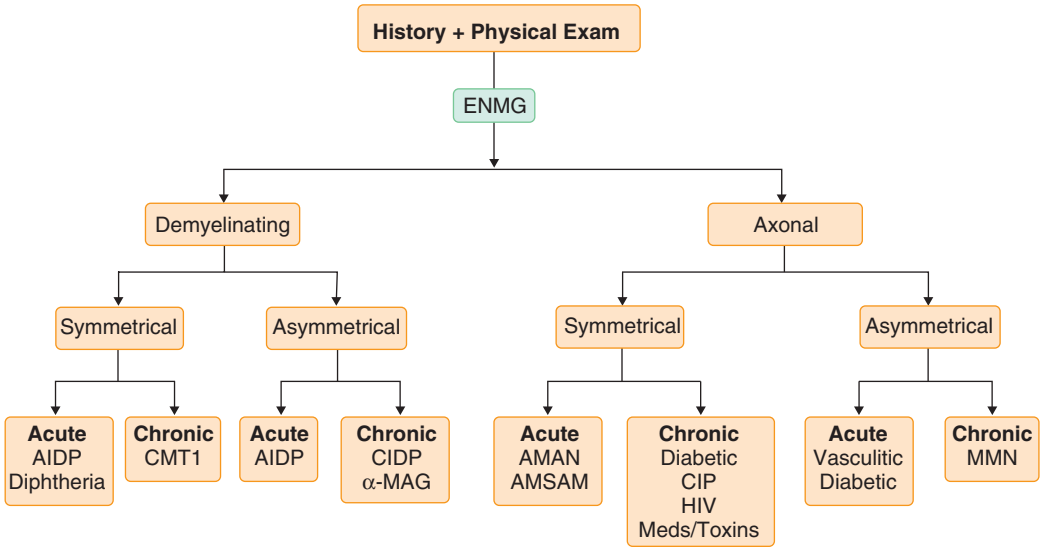

FIGURE 420-1. A systematic approach to evaluating neuropathy. The diseases listed are examples of neuropathies associated with specific neurophysiologic and clinical findings. Diabetic distal, predominantly sensory neuropathies are manifested as chronic axonal neuropathies; acute asymmetrical neuropathies can also occur with diabetes. Most neuropathies caused by toxins or by side effects of medication are chronic, symmetrical axonal neuropathies. AIDP, AMAN, and AMSAN are subtypes of Guillain-Barré syndrome. These and other examples are discussed in more detail in the text. AIDP = acute inflammatory demyelinating polyradiculoneuropathy; AMAN = acute motor axonal neuropathy; AMSAN = acute motor and sensory axonal neuropathy; CIDP = chronic inflammatory polyradiculoneuropathy; CIP = chronic illness polyneuropathy; CMT1 = Charcot-Marie-Tooth disease type 1, a genetic disorder; HIV = human immunodeficiency virus–related neuropathy; α-MAG = anti–myelin-associated glycoprotein; MMN = multifocal motor neuropathy.

TABLE 420-1 PREDOMINANTLY SENSORY NEUROPATHIES

CLASSIFICATION	SUBGROUP	TYPE	FIBERS
Genetic		HSAN	Large, small
Inflammatory/immune	Monoclonal gammopathy	Anti-MAG (early on)	Mainly large
	Vasculitis	Sjögren syndrome	Large, small
	Paraneoplastic	Anti-Hu	Large
Metabolic	Diabetic	Distal symmetrical polyneuropathy	Large, small
Infectious	HIV	HIV neuropathy	Small
	Herpes zoster	Focal radiculoneuropathy	Small
	Leprosy	Tuberculoid	Small
Toxic/deficiency	Medications	Vincristine	Small
		Paclitaxel	Large
		Cisplatin	Large
		Thalidomide	Large
		NRTIs	Small
	Toxins	Thallium	Small
		Acrylamide	Large
		Pyridoxine (B_6)	Large, small
	Deficiency states	Vitamin B_1	Large, small
		Vitamin B_{12}	Large
		Vitamin E	Large
Idiopathic	Frequent		Large, small

HIV = human immunodeficiency virus; HSAN = hereditary sensory and autonomic neuropathy; MAG = myelin-associated glycoprotein; NRTIs = nucleoside reverse transcriptase inhibitors.

TABLE 420-2 PREDOMINANTLY MOTOR NEUROPATHIES

CLASSIFICATION	SUBGROUP	TYPE
Genetic		HMN
Inflammatory/immune	Guillain-Barré	AMAN
	Multifocal motor neuropathy	MMN
	Critical illness myopathy	CIM
Toxic/deficiency	Medications	Dapsone
	Toxins	Lead (adults)

AMAN = acute motor axonal neuropathy; HMN = hereditary motor neuropathy.

chronic (>6 months). For example, Guillain-Barré syndrome develops over a period of days to weeks, whereas chronic inflammatory demyelinating polyneuropathy (CIDP) evolves over months and inherited neuropathies may develop over years.

Neurologic Examination

Wasting of muscle is prominent in many sensorimotor or motor neuropathies, regardless of whether they are primary axonal or primary demyelinating disorders, because even demyelinating neuropathies are associated with secondary axonal degeneration. Atrophy frequently occurs in muscles of dorsiflexion, such as the tibialis anterior, and in intrinsic hand muscles, such as the first dorsal interosseus. Fasciculations, which appear as small twitches of the muscle, are sometimes present, particularly in axonal neuropathies.

Weakness is often most pronounced in foot dorsiflexion and eversion and in the intrinsic hand muscles. In the lower extremities, weakness usually progresses to the muscles of plantar flexion before more proximal muscles become involved.

Sensory loss is usually in a stocking-glove distribution in both large and small fiber neuropathies. Cold, erythematous, or bluish discolored feet suggest loss of small fiber function. Large fiber sensory loss, or "sensory ataxia," in the upper extremities can often be detected by an inability of the patient to locate the thumb accurately with the opposite index finger while the eyes are closed or by the presence of a characteristic irregular tremor (pseudoathetosis) of the outstretched fingers.

The sensory examination should include vibration, position, and light touch, as well as pain and temperature. It is important to determine the degree and extent of sensory loss, in addition to the pattern of deficits (symmetrical or asymmetrical; distal or generalized; focal, multifocal, or diffuse).

The complete absence of reflexes early in the course of a neuropathy suggests a demyelinating neuropathy (Chapter 411); for example, the absence of reflexes in early childhood is often the first detectable abnormality in children with inherited demyelinating neuropathies. Alternatively, the absence of ankle reflexes but the presence of normal patellar or upper extremity reflexes is common in "dying back" (length-dependent) axonal neuropathies, both acquired and inherited. Reflexes may be present in small fiber neuropathies.

On gait testing, subtle weakness in the feet can be detected by an inability of the patient to heel-walk. Sensory ataxia can be appreciated by a wide-based gait or inability to tandem-walk.

Neurologic Testing

Neurophysiology

Electromyography (EMG) and nerve conduction studies can determine whether a neuropathy is primarily demyelinating or axonal and can confirm whether the process is symmetrical or asymmetrical (Chapter 396).

Motor nerve conduction velocities measure conduction over the main body of nerves but not their proximal or distal portion. Distal motor latencies and F wave latencies measure velocities over the distal and proximal portions of the nerves. When slowing is roughly the same over the proximal, distal, and main portion of the nerve, the slowing is said to be uniform. When the slowing is multifocal or asymmetrical, either along the same nerve or between different nerves, the slowing is said to be nonuniform. Slowed conduction velocities (to less than 70% of normal) suggest that the neuropathy is primarily demyelinating.

The sensory nerve action potential is a summation of action potentials from individual large-diameter sensory axons. In axonal neuropathies, amplitudes of the compound muscle action potential or sensory nerve action potential are reduced. When there has been a loss of individual sensory axons, amplitudes of the sensory nerve action potential are reduced.

The presence of spontaneous activity on EMG, such as fibrillations or positive sharp waves, suggests that an acute or active process is damaging axons and denervating muscle. The presence of large, polyphasic motor units suggests partial reinnervation of muscle by regenerating axons (i.e., a more chronic process). Recruitment of motor units is also reduced in patients with demyelinating and axonal neuropathies.

Quantitative Sensory Testing

Quantitative sensory testing can assess and quantify vibratory, thermal, or painful sensory function in patients with peripheral neuropathies or other sensory disorders. Although the stimulus is an objective physical event, the response represents a subjective report and requires cooperation from the patient; as a result, this test by itself cannot diagnose sensory neuropathies or sensory loss.

Nerve and Skin Biopsy

Nerve biopsy is occasionally indicated to address specific questions, such as whether vasculitis, tumor, or another infiltrative or metabolic disorder is present. Biopsy of sural nerves is performed just above the ankle. After biopsy, patients lose sensation over the region on the lateral aspect of the foot that is innervated by the sural nerve, and transient painful dysesthesias may develop around the biopsy site. Teased sural nerve fiber analysis can demonstrate segmental demyelination or remyelination, and electron microscopy can demonstrate features of nerve regeneration and identify specific pathologic processes.

Epidermal skin biopsies with quantification of a loss of small epidermal nerve fibers may aid in the diagnosis of sensory neuropathies, particularly in neuropathies such as diabetes mellitus, human immunodeficiency virus (HIV) infection, or chemotherapeutic drugs. Skin biopsies can be used to evaluate myelinated sensory nerves, although these techniques are largely for research purposes rather than clinical management.

Laboratory Findings

Evaluation of all patients with suspected neuropathy should include blood glucose and creatinine levels, as well as a complete blood count (including red blood cell indices to detect possible macrocytosis). If the history and EMG are consistent with exposure to a toxin or a vitamin deficiency state, specific testing is indicated. Most patients should also have a vitamin B_{12} level measurement (Chapter 164), a test for syphilis (Chapter 319), and serum immunofixation electrophoresis for possible monoclonal gammopathy (Chapter 187). In unexplained sensory neuropathies, HIV testing should be considered

(Chapter 394). In selected patients, electrodiagnostic studies will suggest the need to test for specific antibodies, such as antibodies reacting to ganglioside GM_1 or myelin-associated glycoprotein (MAG). Genetic testing is most cost-effective when selection of candidate genes is based on the patient's nerve conduction studies, inheritance pattern, and clinical findings.

INHERITED NEUROPATHIES

DEFINITION

Inherited neuropathies are frequently called Charcot-Marie-Tooth (CMT) disease based on the three physicians who initially characterized the disorders in the late 19th century. They are also referred to as hereditary motor and sensory neuropathies, hereditary motor neuropathies, or hereditary sensory and autonomic neuropathies, depending on their clinical manifestation. Autosomal dominant forms are subdivided into demyelinating (CMT1) and axonal (CMT2) forms based on electrophysiologic and neuropathologic criteria. X-linked (CMTX) and autosomal recessive (CMT4) forms are also seen.[1] Each type of CMT is subdivided by the specific genetic cause of the neuropathy. For example, the most common form of CMT1, termed CMT1A, is caused by a duplication of a fragment of chromosome 17 containing the peripheral myelin protein 22-kD *(PMP22)* gene (see later). Mutations in more than 70 genes have been identified as causes of inherited neuropathies (E-Table 420-1).

EPIDEMIOLOGY

The prevalence of CMT is about 1 in 2500, without ethnic predisposition. The 17p11.2 duplication causing CMT1A accounts for 60 to 70% of CMT1 patients, CMT1X for approximately 10 to 20% of CMT cases, CMT1B for less than 5% of patients, and CMT2 for about 20% of cases. The prevalence of hereditary neuropathy with liability to pressure palsies is not known, but about 85% of patients with clinical evidence of this syndrome have a chromosome 17p11.2 deletion.

PATHOBIOLOGY

Dysmyelination, demyelination, remyelination, and axonal loss are characteristic features of the demyelinating forms of CMT1. In Dejerine-Sottas neuropathy, myelin may never have formed normally. In CMT1, onion bulbs of concentric Schwann cell lamellae are usually present on nerve biopsies, with loss of both small- and large-diameter myelinated fibers and sometimes axons. Focal, sausage-like thickenings of the myelin sheath (tomacula) are characteristic of hereditary neuropathy with liability to pressure palsies (HNPP) but may also be found in other forms of CMT1, particularly CMT1B. In CMT1, disability typically correlates better with secondary axonal degeneration than with demyelination itself, thereby demonstrating the importance of Schwann cell–axonal interactions in demyelinating disease (E-Fig. 420-1).

CLINICAL MANIFESTATIONS

Despite phenotypic variability, the typical clinical course of CMT1 and CMT2 patients includes normal development before weakness, and sensory loss appearing gradually within the first 2 decades of life. Affected children are often slow runners and have difficulty with activities that require balance (e.g., skating, walking across a log). Ankle-foot orthoses are frequently required by the third decade. Fine movements of the hands for activities such as turning a key or using buttons and zippers may be impaired, but the hands are rarely as affected as the feet. Most patients remain ambulatory throughout life and have a normal lifespan.

A minority of CMT patients have a more severe phenotype with delayed motor milestones and onset in infancy, termed *Dejerine-Sottas neuropathy.* Especially severe cases are classified as congenital hypomyelination if myelination appears to be disrupted during embryologic development. Many patients have de novo autosomal dominant disorders, and the term Dejerine-Sottas neuropathy is currently used primarily to denote severe early-onset clinical phenotypes regardless of the inheritance pattern. Patients with hereditary motor neuropathies sometimes have mild sensory abnormalities, and patients with hereditary sensory and autonomic neuropathies usually have some weakness. The same mutations in the same gene *(GARS)* cause both CMT2D and hereditary motor neuropathy type V (Video 420-1).

DIAGNOSIS

Molecular testing, performed after the family history, neurologic examination, and neurophysiologic testing have suggested the probable candidate genes (GeneClinics—available at www.geneclinics.org), is the "gold standard" for the diagnosis of inherited neuropathies.[2] Nerve conduction velocity testing can distinguish between demyelinating and axonal neuropathies. Most CMT1 patients, particularly those with CMT1A, have a uniformly slow nerve conduction velocity of about 20 m/second. Asymmetrical slowing, which is characteristic of hereditary neuropathy with liability to pressure palsies, may be found in patients with missense mutations in *PMP22, MPZ, EGR2,* and *GJB1.* Although CMT2 is characterized by axonal loss and reduced compound muscle action potential or sensory nerve action potential amplitudes, virtually all forms of CMT1 have axonal loss as well as demyelination. Genetic testing should focus on specific possibilities.

Differential Diagnosis

Inherited neuropathies must be distinguished from acquired neuropathies (see later). Other genetic disorders of the CNS, such as hereditary spastic paraplegia (Chapter 410) or leukodystrophies (Chapter 411), may mimic inherited neuropathies by causing length-dependent weakness, sensory loss, and foot deformities such as pes cavus; these patients often have upper motor neuron signs, such as increased reflexes or Babinski signs, and do not have neurophysiologic evidence of neuropathy.

TREATMENT

Rx

There is no specific therapy for the inherited neuropathies, but clinical and genetic counseling and symptomatic and rehabilitative treatment are important. A detailed family history and often examination of family members are required for prognosis and genetic counseling.

Ankle-foot orthoses to correct footdrop may return gait and balance to normal for years. Foot surgery is sometimes offered to correct inverted feet, pes cavus, and hammertoes. Surgery may improve walking, alleviate pain over pressure points, and prevent plantar ulcers. Foot surgery, however, is generally unnecessary and does not improve weakness and sensory loss. Ascorbic acid, progesterone antagonists, and subcutaneous injections of neurotrophin 3 have improved animal models of CMT1A but have not improved patients with CMT1A or other forms of CMT.

INFLAMMATORY AND IMMUNOLOGIC NEUROPATHIES

Guillain-Barré Syndrome

DEFINITION

Guillain-Barré syndrome refers to acquired, inflammatory peripheral neuropathies that have (1) an acute onset, (2) elevated cerebrospinal fluid (CSF) protein levels with low CSF cell counts (cyto-albumologic dissociation), and (3) a monophasic illness with at least partial recovery.[3] Guillain-Barré syndrome is subdivided into acute inflammatory demyelinating polyneuropathy, acute motor and sensory axonal neuropathy, acute motor axonal neuropathy, and Miller-Fisher syndrome (Table 420-3).

EPIDEMIOLOGY

Acute inflammatory demyelinating polyneuropathy accounts for up to 97% of cases of Guillain-Barré syndrome in North America and Europe. It is a sporadic disorder with an incidence of 0.6 to 1.9 cases per 100,000 in North America and Europe. Men are more likely to be affected than women (1.4 : 1). In 60% of cases, acute inflammatory demyelinating polyneuropathy is preceded by a respiratory tract infection (e.g., cytomegalovirus [Chapter 376], Epstein-Barr virus [Chapter 377], or gastroenteritis [*Campylobacter jejuni*; Chapter 303]). In the Netherlands, 5% of cases have been attributed to a preceding hepatitis E infection.[4] Acute motor axonal neuropathy and acute motor and sensory axonal neuropathy are rare in North American and Europe but more frequent in China, Japan, Mexico, Korea, and India.[5]

PATHOBIOLOGY

All forms of Guillain-Barré syndrome probably result from postinfectious molecular mimicry in which nerve antigens are attacked by the immune system because they resemble antigens presented by microbes, in particular, *C. jejuni.* For example, the HS/0 : 19 serotype of *C. jejuni* is common in northern Chinese patients with Guillain-Barré syndrome and in other countries. Assays with antiganglioside antibodies, bacterial toxins, and lectins have characterized potential immunogenic regions of diarrhea-associated *C. jejuni* strains. However, it is not clear that molecular mimicry causes acute

E-TABLE 420-1 CLASSIFICATION OF CHARCOT-MARIE-TOOTH DISEASE

TYPE	GENE/LOCUS	SPECIFIC PHENOTYPE
AUTOSOMAL DOMINANT (AD) CMT1		
CMT1A	Dup 17p *(PMP22)*	Classic CMT1
	PMP22 (point mutation)	Classic CMT1/DSS/CHN/HNPPs
CMT1B	*MPZ*	CMT1/DSS/CHN/intermediate/CMT2
CMT1C	*LITAF*	Classic CMT1
CMT1D	*EGR2*	Classic CMT1/DSS/CHN
CMT1E	*NEFL*	CMT2 but can have slow MNCVs in CMT1 range ± early-onset severe disease
HNPP		
HNPP	Del 17p *(PMP22)*	Typical HNPP
	PMP22 (point mutation)	Typical HNPP
X-linked CMT1 (CMT1X)		
CMT1X	*GJB1*	Intermediate ± patchy MNCVs/male MNCVs < female MNCVs
AUTOSOMAL RECESSIVE (AR) DEMYELINATING CMT (CMT4)		
CMT4A	*GDAP1*	Demyelinating or axonal, usually early onset and severe/vocal cord and diaphragm paralysis described/rare AD CMT2 families described
CMT4B1	*MTMR2*	Severe CMT1/facial/bulbar/focally folded myelin
CMT4B2	*SBF2*	Severe CMT1/glaucoma/focally folded myelin
CMT4C	*SH3TC2*	Severe CMT1/scoliosis/cytoplasmic expansions
CMT4D (HMSNL)	*NDRG1*	Severe CMT1/gypsy/deafness/tongue atrophy
CMT4E	*EGR2*	Classic CMT1/DSS/CHN
CMT4F	*PRX*	CMT1/more sensory/focally folded myelin
CMT4H	*FGD4*	CMT1
CMT4J	*FIG4*	CMT1
CCFDN	*CTDP1*	CMT1/gypsy/cataracts/dysmorphic features
HMSN-Russe	10q22-q23	CMT1
CMT1	*PMP22* (point mutation)	Classic CMT1/DSS/CHN/HNPPs
CMT1	*MPZ*	CMT1/DSS/CHN/intermediate/CMT2
AUTOSOMAL DOMINANT (AD) CMT 2		
CMT2A	*MFN2*	CMT2/usually severe/optic atrophy
CMT2B	*RAB7A*	CMT2 with predominant sensory involvement and sensory complications
CMT2C	12q23-q24	CMT2 with vocal cord and respiratory involvement
CMT2D	*GARS*	CMT2 with predominant hand wasting/weakness or dHMN V
CMT2E	*NEFL*	CMT2 but can have slow MNCVs in CMT1 range ± early-onset severe disease
CMT2F	*HSPB1 (HSP27)*	Classic CMT2 or dHMN II
CMT2G	12q12-q13.3	Classic CMT2
CMT2L	*HSPB8 (HSP22)*	Classic CMT2 or dHMN II
CMT2	*MPZ*	CMT1/DSS/CHN/intermediate/CMT2
CMT2 (HMSNP)	3q13.1	CMT2 with proximal involvement
AUTOSOMAL RECESSIVE (AR) CMT2 (ALSO CALLED CMT4)		
AR CMT2A	*LMNA*	CMT2 proximal involvement and rapid progression described/also causes muscular dystrophy/cardiomyopathy/lipodystrophy
AR CMT2B	19q13.1-13.3	Typical CMT2
AR CMT2	*GDAP1*	CMT1 or CMT2 usually early onset and severe/vocal cord and diaphragm paralysis described/rare AD CMT2 families described
DOMINANT INTERMEDIATE CMT (DI-CMT)		
DI-CMTA	10q24.1-25.1	Typical CMT
DI-CMTB	*DNM2*	Typical CMT
DI-CMTC	*YARS*	Typical CMT, with motor predominance
DI-CMTE	*IFN2*	Syndromic CMT
DI-CMTF	*GNB4*	Typical CMT
HEREDITARY NEURALGIC AMYOTROPHY (HNA)		
HNA	*SEPT9*	Recurrent neuralgic amyotrophy

AD, autosomal dominant; AR, autosomal recessive; CHN, congenital hypomyelinating neuropathy; CMT, Charcot-Marie-Tooth; CTDP1, CTD phosphatase subunit 1; Del, deletion; DMN2, dynamin 2; DSN, Dejerine-Sottas neuropathy; Dup, duplication; EGR2, early growth response 2; FGD4, FYVE, RhoGEF and PH domain containing 4; FIG4, FIG 4 homologue; GARS, glycyl tRNA synthetase; GDAP1, ganglioside-induced differentiation associated protein 1; GJB1, gap junction protein β1; HNPP, hereditary neuropathy with liability to pressure palsies; HSP22, heat shock 22-kD protein 8; HSP27, heat shock 27-kD protein 1; KIF1Bb, kinesin family member 1B-b; LITAF, lipopolysaccharide-induced tumor necrosis factor; LMNA, lamin A/C; MCV, motor conduction velocity ; MFN2, mitofusin 2; MPZ, myelin protein zero; MTMR2, myotubularin-related protein 2; MTMR13, myotubularin-related protein 13; NDRG1, N-myc downstream regulated gene 1; NEFL, neurofilament, light polypeptide 68 kD; PMP22, peripheral myelin protein 22; PRX, periaxin; RAB7, RAB7, member *RAS* oncogene family; SEPT9, septin 9; SH3TC2, SH3 domain and tetratricopeptide repeats 2; YARS, tyrosyl tRNA synthetase.

Reprinted and updated from Reilly MM, Shy ME. Diagnosis and new treatments in genetic neuropathies. *J Neurol Neurosurg Psychiatry.* 2009;80:1304-1314.

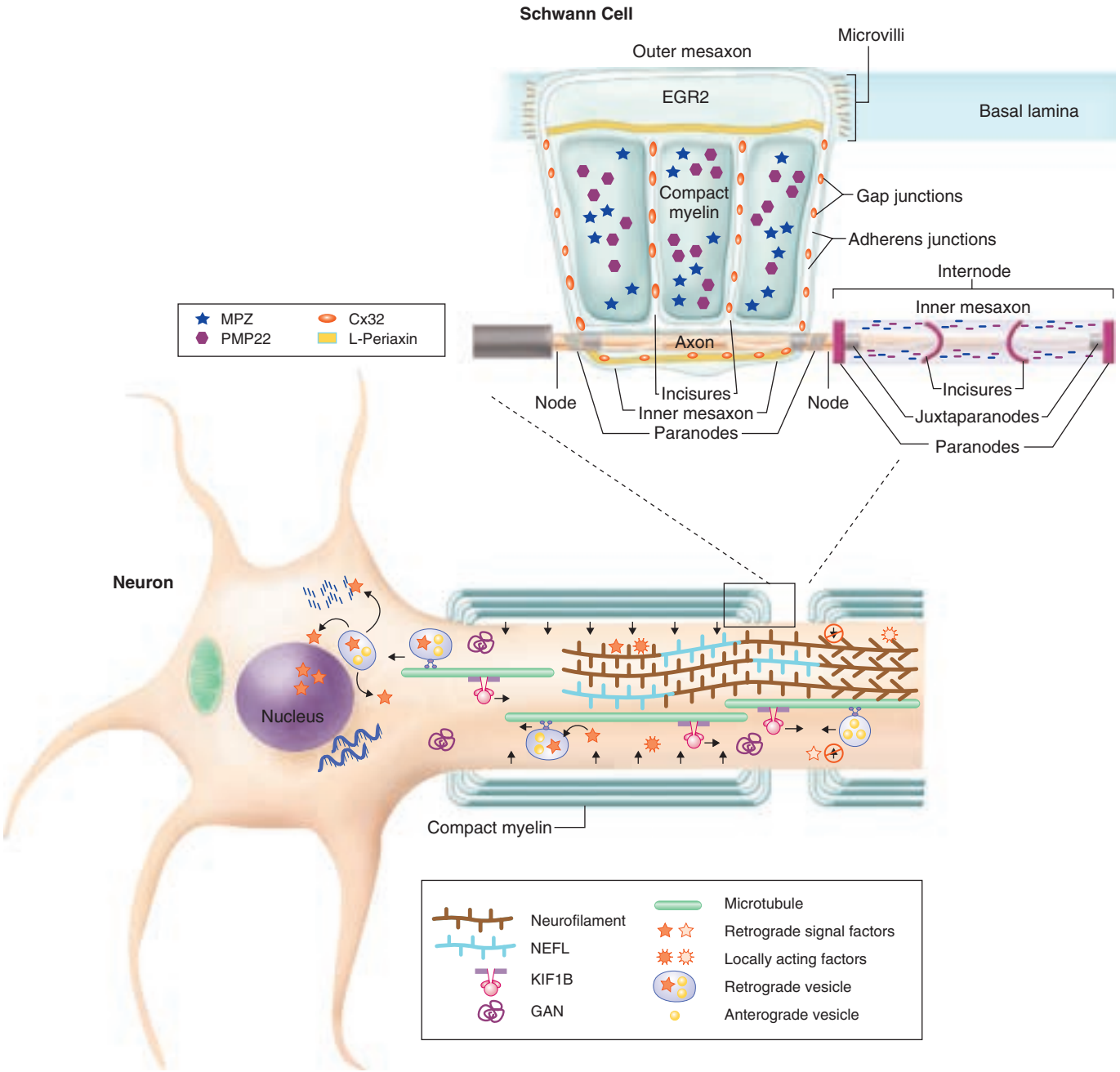

E-FIGURE 420-1. Schematic view of an axon and its myelinating Schwann cell. Proteins mutated in inherited peripheral neuropathies are shown in color at their cellular location. In the upper part of the panel, the myelinating Schwann cell has been unraveled to show the nucleus and regions of both compact myelin and noncompact myelin. Cytoskeletal elements within the axon are also illustrated. (From Shy ME, Garbern JY, Kamholz J. Hereditary motor and sensory neuropathies: A biological perspective. *Lancet Neurol.* 2002;1:110-118.)

VIDEO 420-1. Charcot-Marie-Tooth (CMT) Disease Exam and Walk.

TABLE 420-3 INFLAMMATORY AND IMMUNE-RELATED NEUROPATHIES

DISORDER	TYPE	CLINICAL TRAITS	PATHOBIOLOGY	TREATMENT*
Guillain-Barré syndrome	AIDP	Acute flaccid weakness, sensory loss	Demyelination, lymphocyte infiltration	Plasma exchange, IVIG
	AMAN/AMSAN	Acute flaccid weakness, no sensory loss in AMAN	Molecular mimicry, association with *Campylobacter jejuni*	? Plasma exchange or IVIG
	Fisher syndrome	Acute ataxia, ophthalmoparesis and areflexia	Anti-GQ1b antibodies	? Plasma exchange or IVIG
CIDP		Slower onset of weakness, sensory loss	Inflammatory/immune-mediated demyelination	Corticosteroids, plasma exchange, IVIG
Monoclonal gammopathy	IgM	Sensory > motor	Particularly anti-MAG	Immune suppression (Chapter 187)
	IgG	Sensorimotor	Many probably chance associations, osteosclerotic myeloma, solitary plasmacytoma; POEMS may be immune mediated	Treatment of myeloma
Multifocal motor neuropathy		Pure motor	Focal demyelination, antibodies to GM_1 frequent	IVIG

*Typical regimens: IVIG, 0.4 g/kg/day × 5 days for a total of 2 g/kg; may be repeated monthly as needed. Corticosteroids: prednisone, 60 to 80 mg/day for up to 3 months, followed by gradual tapering, depending on the clinical response, with a goal to about 20 mg on an alternate-day regimen.
AIDP = acute idiopathic demyelinating polyneuropathy; AMAN = acute motor axonal neuropathy; AMSAN = acute motor and sensory axonal neuropathy; CIDP = chronic inflammatory polyradiculoneuropathy; Ig = immunoglobulin; IVIG = intravenous immune globulin; MAG = myelin-associated glycoprotein; POEMS = polyneuropathy, organomegaly, M protein, skin changes.

inflammatory demyelinating polyneuropathy, which is the most common form in the United States and Europe.

CLINICAL MANIFESTATIONS

Weakness, the most common initial symptom in both acute inflammatory demyelinating polyneuropathy and acute motor and sensory axonal neuropathy, can be mild, such as difficulty walking, or severe, such as total quadriplegia and respiratory failure. Bilateral weakness of facial muscles (facial diplegia) occurs in about 50% of cases. The most common manifestation is leg weakness that progresses into the arms. Although Guillain-Barré syndrome has been described as an "ascending paralysis," proximal weakness is common, and 5% of cases have isolated cranial nerve involvement that subsequently descends into the limbs. Slight sensory loss occurs in most patients. The autonomic nervous system is involved in about 65% of patients.

Length-dependent weakness without sensory loss develops in patients with acute motor axonal neuropathy, including cranial nerve involvement in about 25%. Miller-Fisher syndrome consists of the triad of ophthalmoplegia, ataxia, and areflexia. Facial weakness, ptosis, and pupillary abnormalities may be present. Nerve conduction velocities in Miller-Fisher syndrome are generally normal, unlike the case with acute inflammatory demyelinating polyneuropathy.

DIAGNOSIS

The diagnosis of acute inflammatory demyelinating polyneuropathy and acute motor and sensory axonal neuropathy is based on the history, physical examination, and CSF evaluation. Deep tendon reflexes are decreased or absent, and the CSF is abnormal, with a high protein level but a paucity of white blood cells. The weakness is symmetrical. The presence of CNS abnormalities should cast doubt on the diagnosis. Acute inflammatory demyelinating polyneuropathy is distinguished from acute motor and sensory axonal neuropathy by nerve conduction studies. In both acute inflammatory demyelinating polyneuropathy and acute motor and sensory axonal neuropathy, the CSF should have fewer than five white blood cells (WBCs)/mL. If the CSF cell count is greater than 50 WBCs/mL, another diagnosis, such as HIV infection (Chapter 394) or Lyme disease (Chapter 321), should be considered. Elevated CSF protein may not be apparent in the first 7 to 10 days of the illness; in up to 10% of cases, CSF protein levels remain normal. Approximately 5% of patients with Guillain-Barré syndrome have Miller-Fisher syndrome, and more than 85% of these patients have polyclonal antibodies that react with the ganglioside GQ_{1b}.

Differential Diagnosis

The differential diagnosis varies in different parts of the world. Historically, poliomyelitis (Chapter 379) was the major cause of acute flaccid quadriparesis. In North America, polio has been eradicated, but other viral illnesses may induce polio-like syndromes: ECHO 70 (Chapter 379), coxsackievirus (Chapter 379), West Nile virus (Chapter 383), and rarely, rabies (Chapter 414). Because these diseases are not demyelinating disorders, they can be distinguished from acute inflammatory demyelinating polyneuropathy by their normal nerve conduction velocity. However, the results of electrodiagnostic studies are similar in both acute motor axonal neuropathy and the polio-like syndromes, thus making distinction between acute motor axonal neuropathy and these viral syndromes difficult.

Tick paralysis (Chapter 359), caused by a toxin within the tick, can mimic Guillain-Barré syndrome, particularly in children. Usually, removal of the tick is associated with improvement within hours, although progression can occur. Progression is particularly likely in Australia, where the toxin differs from that found in North America.

Botulism (Chapters 296 and 422) rapidly produces a flaccid paralysis. Patients have ophthalmoplegia, bulbar weakness, dry mouth, constipation, and orthostatic hypotension, but sensory symptoms do not develop. Other entities that can mimic Guillain-Barré syndrome are acute spinal cord compression (Chapter 400), acute transverse myelitis (Chapter 411), and vascular myelopathies, all of which are characterized by decreased reflexes before the development of upper motor neuron signs such as increased reflexes. Carcinomatous or lymphomatous meningitis can also cause a rapidly developing quadriparesis, but both are associated with elevated CSF WBC counts.

TREATMENT

Rx

Patients with Guillain-Barré syndrome require hospitalization because of the potential for respiratory compromise. Pulmonary function tests should be performed frequently; a vital capacity of less than 1 L or a negative inspiratory force of less than −70 suggests the need for ventilator support in an intensive care unit. Autonomic instability and difficulty swallowing should be monitored. Therapies directed at modulating the immune system are effective in Guillain-Barré syndrome. Although intravenous immunoglobulin (IVIG; 2 g/kg divided over 2 to 5 days within the first 2 weeks) and plasma exchange are equally effective, at least in the first 2 weeks, IVIG is preferred because of its convenience unless there are contraindications, such as low serum immunoglobulin A (IgA) levels, renal failure, or severe hypertension. Plasma exchange, usually four exchanges of 1.5 L of plasma spread over a 10-day period, is also effective. Two plasma exchanges may be sufficient in mild cases, and six exchanges are not superior to four in severely affected patients. This therapy should be administered within the first 2 weeks and not later than 4 weeks after the onset of clinical disease. Intravenous corticosteroids alone are not beneficial. Methylprednisolone (500 mg/day for 5 days) in association with IVIG has a slight initial advantage over IVIG alone but no benefit in terms of long-term disability.

Ten percent of patients with Guillain-Barré syndrome relapse after initially responding to plasma exchange or IVIG. These patients usually respond to a second cycle of the previously effective treatment. The combined use of plasma exchange followed by IVIG does not improve the prognosis.

PROGNOSIS

Fifty percent of patients progress to their nadir, or maximum disability, within 2 weeks, 75% within 3 weeks, and more than 90% within 4 weeks of the onset

of symptoms. With supportive care, mortality in Guillain-Barré is 3% at 6 months, primarily in the elderly and severely affected patients, and especially during the recovery phase.[6] After a brief period of stabilization, slow spontaneous recovery occurs over a period of weeks or months. Most patients recover completely or are left with minor sequelae; about 20% have a persistent disability. The long-term prognosis depends at least in part on the extent of axonal loss.[7] Patients with low compound muscle action potential amplitudes in the upper extremities are more likely to have a poor prognosis.

Patients with acute motor axonal neuropathy will recover after approximately 2 months, but the extent of recovery may be less than in Guillain-Barré syndrome. In general, the prognosis in Miller-Fisher syndrome is excellent.

Chronic Inflammatory Demyelinating Polyradiculoneuropathy

DEFINITION

CIDP is a chronic acquired demyelinating sensorimotor neuropathy that may be monophasic, relapsing, or progressive. By definition, CIDP develops over at least a 2-month period, which is slower than for acute inflammatory demyelinating polyneuropathy, which it otherwise resembles.

EPIDEMIOLOGY

CIDP occurs in all age groups, with a mean age range of 30 to 50 years. Women are more likely to be affected than men. Antecedent events are less common than in Guillain-Barré syndrome; they occur in about 30% of patients and include upper respiratory infections, gastrointestinal infections, vaccinations, surgery, and trauma.

PATHOBIOLOGY

CIDP is considered an autoimmune disorder based on pathologic findings in nerve biopsy samples from patients and on animal models, such as experimental allergic neuritis, in which a similar disorder follows immunization with peripheral nervous system myelin components and Freund adjuvant. Nerve biopsy shows macrophage-mediated segmental demyelination, occasional endoneurial lymphocytic T-cell infiltrates, and endoneurial edema. The major histocompatibility complex class I and II antigens are upregulated, and there are often deposits of immunoglobulins and complement split products on the outer Schwann cell membranes or myelin sheaths. CIDP can be passively transferred to animals by patient sera, but no clear autoantigen has been identified.

CLINICAL MANIFESTATIONS

Weakness and sensory loss begin insidiously and progress over a period of months to years.[8] Weakness is commonly proximal as well as distal. Patients can become bedridden. Loss of proprioception from damage to large-diameter sensory nerves may affect balance and result in an action tremor. Deep tendon reflexes are usually absent or markedly decreased. Facial weakness (15%), ptosis or ophthalmoparesis (5%), and papilledema occur occasionally. Variant forms include pure motor, pure sensory, and multifocal disease.

DIAGNOSIS

Diagnosis is based on clinical symptoms and signs, electrodiagnostic studies, and CSF examination. Nonuniform, asymmetrical slowing on nerve conduction studies is characteristic. One portion of a nerve may have different conduction than another. For example, if damage is primarily in the spinal roots, proximal conduction velocities and F wave latencies may be most affected. Compound muscle action potentials are generally reduced because of the concomitant axonal degeneration that occurs with demyelinating neuropathies. However, temporal dispersion and conduction block may also reduce the amplitude of muscle action potential in any demyelinating neuropathy. Sensory nerve conduction is also slow in CIDP, but because sensory nerve action potentials are often not detectable, sensory conduction velocity may be unmeasurable.

CSF results resemble those of acute inflammatory demyelinating polyneuropathy: WBC counts are usually less than 10 cells/μL, and protein levels are higher than 60 mg/dL. CSF cell counts greater than 50/μL suggest another diagnosis, such as HIV infection or hematologic malignancy. CSF protein levels may be normal early in the course of CIDP.

Differential Diagnosis

CIDP is distinguished from acute inflammatory demyelinating polyneuropathy by its time course. CIDP can also occur in diabetes, lymphoma, monoclonal gammopathies (see later), and asymmetrical inherited neuropathies (see earlier).

TREATMENT Rx

Corticosteroids, IVIG (as used for the Guillain-Barré syndrome except that it often needs to be repeated on an approximately monthly basis for up to and sometimes exceeding 6 months), and plasma exchange appear to be equally effective[■] treatments for CIDP, with about two thirds of patients responding. For steroids, a standard approach is oral prednisone (1 mg/kg/day) for 6 to 8 weeks, followed by slow tapering over a 3- to 12-month period to a maintenance level of about 0.1 mg/kg/day. A response to prednisone may take months to occur, and occasional patients may worsen before they respond. Other alternatives are pulsed dexamethasone (6 cycles of 40 mg/day orally for 4 days) or short-term prednisolone (60 mg/day for 5 weeks, then tapering to zero), which is equally effective for CIDP.[■] Plasma exchange is also effective. Because of the side effects or inadequate response to long-term corticosteroids, azathioprine, cyclosporine, cyclophosphamide, methotrexate, mycophenolate mofetil, rituximab, and interferon-α or -β have been used with variable success in uncontrolled reports.[9]

Neuropathy Associated with Monoclonal Gammopathy

DEFINITION

Monoclonal gammopathy refers to the presence in the β-γ region of serum protein electrophoresis of an abnormal spike (variably termed a paraprotein, monoclonal protein, or M protein) consisting of immunoglobulins of the same isotype, all produced by a single clone of abnormally proliferating lymphocyte/plasma cells. In some cases, the M protein is part of a malignant lymphoproliferative disease such as multiple myeloma, solitary plasmacytoma (IgG and IgA), Waldenström IgM macroglobulinemia (Chapter 187), lymphoma (Chapter 185), chronic lymphocytic leukemia (Chapter 184), primary amyloidosis (Chapter 188), or cryoglobulinemia (Chapter 187). In most instances, however, monoclonal gammopathy is not initially associated with any of these disorders and is classified as a monoclonal gammopathy of uncertain significance (MGUS), although in patients with MGUS, the gammopathy may evolve into a malignant form (Chapter 187).

EPIDEMIOLOGY

Monoclonal gammopathy occurs in up to 8% of patients with peripheral neuropathy of unknown etiology. However, MGUS is frequent, being found in 1% of the population older than 50 years and in 3% older than 70 years, and most subjects with MGUS do not have neuropathy. In some cases, the co-occurrence of neuropathy and M protein may be a coincidence, but in other cases, the M protein is clearly related to the neuropathy.

The prevalence of neuropathy is higher in patients with IgM versus IgG or IgA M proteins. The prevalence of symptomatic neuropathy associated with IgM monoclonal gammopathy in patients older than 50 years is approximately 20 per 100,000. In half of such patients, the M protein reacts with either the HNK1 carbohydrate moiety of MAG or with other glycoproteins (MPZ, PMP22) and glycolipids (sulfoglucuronylparagloboside [SGPG] and lactosaminylparagloboside [SGLPG]). IgM M proteins associated with neuropathy may also bind to other neural antigens.

In patients with IgG monoclonal gammopathy and neuropathy, the relationship is less clear than with IgM. Although about 10% of patients with multiple myeloma have neuropathy, in most cases, the M protein does not react with a neural antigen, and patients do not improve with immunotherapy (see later). Conversely, approximately 50% of patients with the osteosclerotic form of myeloma have neuropathy, often associated with the non-neurologic manifestations of organomegaly, endocrine abnormalities, and brown, tannish discoloration of the skin). Collectively, the *M* protein, *p*olyneuropathy and other features are referred to by the acronym POEMS (Chapter 187).[10] Similarly, about 50% of patients with light chain amyloidosis have neuropathy.

PATHOBIOLOGY

In patients with IgM M proteins that immunoreact with MAG, nerve biopsies demonstrate segmental demyelination with deposits of M protein and complement. The myelin lamellae are often widened on sural nerve biopsies, but a biopsy is not necessary for diagnosis. High titers (>1 : 10,000) of anti-MAG IgM antibodies are associated with neuropathy, and intraneural or systemic

injection of anti-MAG IgM M proteins causes complement-mediated demyelination of nerves in animals.

CLINICAL MANIFESTATIONS

Most patients with anti-MAG neuropathies are initially seen in their sixth to seventh decade of life with dysesthesias and paresthesias in their legs and unsteadiness while walking because of loss of proprioception. Physical examination shows a length-dependent large fiber sensory neuropathy. Weakness may develop later.

DIAGNOSIS

Nerve conduction velocities are slow (about 25 m/second) with pronounced delays in distal motor latencies, thus prompting the designation *distal acquired demyelinating symmetrical neuropathy* to distinguish the disorder from CIDP.

TREATMENT Rx

Treatment of neuropathies associated with monoclonal gammopathy is similar to that for CIDP (see earlier). However, patients with anti-MAG-related neuropathies do not respond as well to treatment as do patients with CIDP. Anecdotal data support the benefit of rituximab (375 mg/m^2 weekly for 4 weeks) in some patients.

PROGNOSIS

Progression of the neuropathy of monoclonal gammopathy disables about 25% of patients after 10 years and 50% after 15 years. The course of patients with osteosclerotic myeloma and neuropathy depends on the response to treatment of the myeloma. In patients whose myeloma responds to treatment, more than 50% have improvement in neuropathy. In other patients with sensorimotor neuropathies associated with plasma cell dyscrasias, the course is variable, and the M protein may not be related to their neuropathy.

Multifocal Motor Neuropathy and Lewis-Sumner Syndrome

DEFINITION

Multifocal motor neuropathy is characterized by progressive, distal more than proximal, asymmetrical limb weakness, mostly affecting the upper limbs with minimal or no sensory impairment.

EPIDEMIOLOGY

The prevalence of multifocal motor neuropathy is estimated at 2 per 100,000. Men are more frequently affected than women (2.6 : 1). Initial symptoms develop in 80% between the ages of 20 and 50 years, with a mean age at onset of 40 years. Lewis-Sumner syndrome occurs less frequently than multifocal motor neuropathy.

PATHOBIOLOGY

Multifocal motor neuropathy is considered to be an autoimmune neuropathy based on its clinical improvement with immunologically based therapies and because of a frequent association with antiglycolipid antibodies. Patients with MMN often have serum antibodies that react with ganglioside GM_1, and these titers decrease during effective treatment. GM_1 is highly represented in neural membranes at the nodes of Ranvier, compact myelin, and the motor end plate at the neuromuscular junction. A blocking effect on mouse distal motor nerve conduction has been induced in vitro by sera from multifocal motor neuropathy patients with and without high anti-GM_1 antibody titers. These data support the presence of serum factors responsible for conduction block in the sera of patients with multifocal motor neuropathy, although these factors are not invariably related to anti-GM_1 antibodies. Lewis-Sumner syndrome, however, is not associated with antiganglioside antibodies.

CLINICAL MANIFESTATIONS

The usual pattern is progressive, distal, asymmetrical arm weakness, often in the distribution of a single nerve. In a minority of patients, weakness may start proximally or in the legs. The disease will frequently affect other nerves, occasionally with a crossed distribution (i.e., one arm and the contralateral leg). Asymmetry and predominance of arm weakness may become less evident as the disease progresses. Localized muscle atrophy may be mild or absent in the early stage of the disease but can become prominent later as a result of axonal degeneration.

Fasciculations, cramps, and myokymia occur in patients with multifocal motor neuropathy and those with amyotrophic lateral sclerosis (Chapter 419), thereby making distinction between the two disorders difficult. Marked asymmetry in the degree of clinical findings and electrophysiologic abnormalities between contiguous nerves is suggestive of multifocal motor neuropathy rather than amyotrophic lateral sclerosis. Cranial nerve involvement or respiratory failure as a result of unilateral or bilateral phrenic nerve palsy rarely occurs in multifocal motor neuropathy. The presence of sensory loss suggests Lewis-Sumner syndrome.

DIAGNOSIS

The diagnosis is established by the presence of multifocal, persistent partial conduction blocks on motor but not sensory nerve conduction studies. Lewis-Sumner syndrome has sensory loss as well as weakness, with conduction block in both sensory and motor nerves.

TREATMENT Rx

IVIG (2 g/kg) is the initial treatment for multifocal motor neuropathy, and almost 80% of patients respond within a week. However, improvement is typically brief (3 to 6 weeks), so repeated treatments are required indefinitely.[11] Clinical improvement is often accompanied by a reduction or resolution of the motor conduction block in some nerves, but it does not consistently correlate with a reduction in antiganglioside antibody titers. Patients may eventually become refractory to IVIG, and another agent may be needed, such as rituximab (e.g., 375 mg/m^2 weekly for 4 weeks) or azathioprine (2 to 3 mg/kg/day). Plasma exchange and corticosteroids are generally ineffective and have been associated with worsening neuropathy in some patients. Lewis-Sumner syndrome, which is a multifocal variant of CIDP, responds to the same treatments as CIDP.

PARANEOPLASTIC NEUROPATHIES

DEFINITION

Paraneoplastic neuropathies (Chapter 179) are a "remote effect of cancer" not caused by metastatic invasion of neural tissue; radiation therapy or chemotherapy; metabolic, vascular, or hormonal disturbances; or opportunistic infections. It is hypothesized that they are the result of host immune responses to a tumor antigen or antigens that are also present in neural tissues.

EPIDEMIOLOGY

Paraneoplastic syndromes occur in less than 1% of patients with cancer; peripheral neuropathy is only one of the paraneoplastic syndromes. Although more than 25% of patients with cancer have evident neuropathy on neurologic examination, the relationship to malignancy is unclear in most. Paraneoplastic neuropathy may develop before, during, or after the tumor is diagnosed. In certain tumors, neuropathies are distinctive and should prompt a thorough investigation for cancer. Small cell carcinoma of the lung (Chapter 191) is by far the most common underlying neoplasm, followed by carcinoma of the stomach, breast, colon, rectum, ovary, and prostate.

PATHOBIOLOGY

Subacute sensory neuropathy, the most characteristic paraneoplastic neuropathy, results from an immune-mediated ganglionitis that destroys sensory neurons in the dorsal root ganglia. Mononuclear inflammatory infiltrates composed of $CD4^+$ and prominent $CD8^+$ T cells, along with plasma cells, are found in the stroma surrounding the dorsal root ganglion neurons. Other findings include atrophy of the dorsal roots; loss of sensory neurons, which appear to be replaced by a proliferation of satellite cells (Nageotte nodule); axonal degeneration; and secondary degeneration of the dorsal column of the spinal cord. Inflammatory infiltrates can also be found in peripheral nerves or muscle. Sural nerve biopsies typically reveal only loss of myelinated nerve fibers and are not useful for diagnosis.

CLINICAL MANIFESTATIONS

Subacute sensory neuropathy is characterized by subacute, progressive impairment of all sensory modalities and is associated with severe sensory ataxia and areflexia.[12] Subacute sensory neuropathy may precede the diagnosis of tumor by months or even years. At onset, patients may have shooting

pain and burning sensations. Other symptoms include numbness, tingling, and a progressive sensory loss that may be asymmetrical. Symptoms usually progress rapidly to involve all four limbs, the trunk, and face. Findings may then stabilize, although by this time the patient is often totally disabled. Occasional patients have an indolent course.

Neurologic examination reveals loss of deep tendon reflexes and involvement of all modalities of sensation; large fiber modalities such as vibration and joint position sense are most severely affected. The loss of position sense may lead to severe sensory ataxia with pseudoathetoid movements of the hands and an inability to walk despite normal strength. Cranial nerve involvement includes sensorineural deafness, loss of taste, and facial numbness. The asymmetrical pattern of symptoms sometimes suggests a radiculopathy or plexopathy.

A paraneoplastic encephalomyelitis characterized by patchy, multifocal neuronal loss in regions of the cerebral hemispheres, the limbic system, the cerebellum, the brain stem, the spinal cord, and autonomic ganglia often develops in patients with subacute sensory neuropathy. Autonomic symptoms include impotence, dry mouth, and constipation.

DIAGNOSIS

The diagnosis is based on recognizing the typical neuropathy in the setting of malignancy.[13] The results of routine laboratory studies are generally normal. The diagnosis is supported by finding serum polyclonal IgG anti-Hu antibodies, also called antineuronal antibodies type 1, or by indirect immunofluorescence or immunohistochemistry and confirmed by Western blot analysis.

Subacute painful, asymmetrical neuropathy or neuronopathy in an elderly patient should prompt a search for carcinoma of the lung because small cell lung cancer (Chapter 191) accounts for more than 80% of the associated tumors. Subacute sensory neuropathy has also been reported in patients with adenocarcinoma of the lung, breast, ovary, stomach, colon, rectum, and prostate, as well as Hodgkin and non-Hodgkin lymphoma. In patients with no evidence of cancer, detection of anti-Hu antibodies should prompt a computed tomography study of the chest with special attention to the mediastinal lymph nodes. The use of whole body positron emission tomography with fluorodeoxyglucose has been advocated for early diagnosis in patients with anti-Hu antibodies or clinical suspicion of subacute sensory neuropathy because it may reveal neoplastic adenopathy months before computed tomography or magnetic resonance imaging.

TREATMENT Rx

Subacute sensory neuropathy responds poorly to plasma exchange, IVIG, or immunosuppressant medications, even when such treatment is started early in the course of the disease. Successful treatment of the tumor rarely induces remission of subacute sensory neuropathy but may stabilize symptoms.

Other Neuropathies Possibly Associated with Cancer

SENSORIMOTOR NEUROPATHY

Sensorimotor neuropathy occurs in approximately 25% of patients with all types of tumors. The neuropathy can have an acute or subacute onset, with a progressive or relapsing-remitting course. Because no antineuronal antibody has been specifically associated with these neuropathies, their paraneoplastic nature is not established. Severe or relapsing neuropathies often precede the diagnosis of cancer, but the search for malignancy is generally limited to a chest radiograph, stool samples for blood, and routine blood tests. There are no specific treatments for these neuropathies, and their progression does not necessarily correlate with that of the malignancy.

PARANEOPLASTIC VASCULITIS OF NERVES

A nonsystemic vasculitic neuropathy, which may also involve muscle, occurs with various types of tumor, including small cell lung cancer, lymphoma, and carcinoma of the kidney, stomach, and prostate. Neurologic symptoms may develop either before or after the tumor is diagnosed. The neuropathy is subacute and progressive and usually affects older men. Like many paraneoplastic disorders, these neuropathies often respond poorly to treatment.

VASCULITIC NEUROPATHIES

DEFINITION

Vasculitic neuropathies (Table 420-4) typically present as painful acute or semiacute axonal mononeuritis multiplex. There is acute motor and sensory loss in multiple nerve territories. The number of nerves involved may be extensive enough to make the distinction between a multifocal and diffuse neuropathy difficult. Occasionally, vasculitic neuropathy can present as sensory neuropathy, trigeminal neuropathy, compressive neuropathy, or autonomic neuropathy. Neuropathy can occur in systemic vasculitis associated with other organ systems, as well as in nonsystemic vasculitis affecting just nerve and muscle.

EPIDEMIOLOGY

Systemic vasculitic neuropathy is more common than nonsystemic vasculitic neuropathy. Peak ages at onset of both are the fifth to eighth decades, but vasculitis can occur at any age. Neuropathy, particularly mononeuritis multiplex, is common in several forms of systemic vasculitis. Rheumatoid arthritis (Chapter 264) evolves into systemic rheumatoid vasculitis in 5 to 15% of patients, and vasculitic neuropathy will develop in about 50% of these patients. More than 50% of patients with Churg-Strauss syndrome (Chapter 270), 40 to 50% with granulomatosis with polyangiitis (Chapter 270), 35 to 75% with polyarteritis nodosa (Chapter 270), and a majority of patients with mixed cryoglobulinemia (Chapter 187) have neuropathy. Patients with Sjögren syndrome (Chapter 268) are often initially found to have sensory neuropathies. Neuropathies are uncommon in systemic lupus erythematosus.

PATHOBIOLOGY

In patients with mononeuritis multiplex, axonal degeneration develops as a result of nerve ischemia caused by the vasculitic process. Immune-mediated inflammation and necrosis of blood vessel walls occlude the vessel's lumen, thereby resulting in ischemic damage. Small arteries or arterioles (50 to 300 μm) are most commonly affected, particularly those that occur in watershed areas between the distribution of the major nutrient arteries of proximal nerves. True nerve infarcts are rare.

TABLE 420-4 SYSTEMIC VASCULITIS AND NEUROPATHY

TYPE	SEROLOGY FEATURES	ASSOCIATED FEATURES	USUAL NEUROPATHY TYPE	NEUROPATHY PREVALENCE
Rheumatoid arthritis	RF 80–90%	Arthralgias, arthritis frequent; multiple organs	Mononeuritis multiplex and sensorimotor neuropathy	50% of patients with vasculitis
Churg-Strauss	c-ANCA <30% p-ANCA <50%	Eosinophilia Asthma	Mononeuritis multiplex	20% of patients
Granulomatosis with polyangiitis	c-ANCA 75-90% p-ANCA <20%	Pulmonary and renal	Mononeuritis multiplex	15% of patients
Polyarteritis nodosa	c- and p-ANCA rare	Multiple organs, 30% hepatitis B	Mononeuritis multiplex	60% of patients
Mixed cryoglobulinemia	No	Hepatitis C, purpura frequent	Mononeuritis multiplex	20-90% of patients
Sjögren syndrome	α-Ro/SS-A 60%, α-La/SS-B 50%	Dry eyes, dry mouth; women 90%	Sensory	25% of patients
Systemic lupus erythematosus	ANA screen >90%	Multiorgan	Sensorimotor neuropathy	5-20% of patients

ANA = antinuclear antibodies; p- and c-ANCA = perinuclear and cytoplasmic antineutrophil cytoplasmic antibodies; RF = rheumatoid factor; α-Ro/SS-A and α-La/SS-B = antibodies to the Ro/SS-A and La/SS-B antigens.

The immune-mediated inflammation is associated with antibody-antigen complexes that are deposited in the wall of the blood vessel. Antibodies also bind directly to endothelial cell antigens. In both circumstances, complement is activated, as evidenced by deposition of membrane attack complex. Chemotactic factors then recruit neutrophils, which release proteolytic enzymes and generate toxic oxygen free radicals.

The sensory neuropathy of Sjögren syndrome probably results from the infiltration of dorsal root ganglia by cytotoxic T cells. Some patients with systemic vasculitis have symmetrical neuropathies rather than mononeuritis. The pathogenesis of such cases is not defined.

CLINICAL MANIFESTATIONS

Patients typically have a relatively sudden onset of painful, focal or multifocal weakness, or sensory loss. These symptoms reflect ischemia anywhere along the length of the nerve or nerves, generally in the lower extremities.

DIAGNOSIS

Nerve biopsy of clinically affected sensory nerves (sural, superficial peroneal, or superficial radial) is often necessary because therapy may be aggressive and long-term. Superficial peroneal nerve biopsy may be combined with muscle biopsy from the same incision. Pathologic features diagnostic of vasculitis occur in 60% of patients, and less specific features such as multifocal loss of fibers occur in others. Findings diagnostic of vasculitis include destruction of the vessel and inflammation within the vessel wall. Fibrinoid necrosis, vessel wall scarring, recanalization, neovascularization, and hemosiderin are common but not essential histopathologic features of vasculitis.

Although nerve biopsy is the gold standard for diagnosis, clinical, serologic, and electrophysiologic findings can suggest the diagnosis. For example, EMG and nerve conduction velocity studies can distinguish between mononeuritis multiplex and a symmetrical neuropathy. It is essential to confirm nerve conduction velocity abnormalities in a nerve before biopsy. An acute or subacute onset of asymmetrical weakness or sensory loss in the distribution of individual nerves suggests mononeuritis multiplex, particularly in the setting of a known connective tissue disorder. Systemic symptoms, such as unexplained weight loss and purpura, or constitutional symptoms, such as fever, myalgias, arthralgias, pulmonary disease, abdominal complaints, rashes, or night sweats, suggest systemic vasculitis in a patient with mononeuritis multiplex.

The erythrocyte sedimentation rate is usually elevated in the systemic vasculitides but is normal in nonsystemic cases. Perinuclear and cytoplasmic antineutrophil cytoplasmic antibody (p-ANCA and c-ANCA) suggests granulomatosis with polyangiitis (Chapter 270) or Churg-Strauss syndrome (Chapter 270). Hepatitis C (Chapter 149) is usually associated with the presence of cryoglobulins. Serum complement levels, extractable nuclear antigen, angiotensin-converting enzyme levels, serum protein electrophoresis, and HIV serology are generally indicated. CSF analysis is not usually helpful in cases of vasculitic neuropathy but may be needed to exclude infectious (e.g., Lyme disease [Chapter 321]) or other inflammatory causes.

Differential Diagnosis

Acute or subacute mononeuritis multiplex may also result from diabetes, sarcoidosis, Lyme disease, and malignant infiltration of nerves. Multifocal motor neuropathy with conduction block and Lewis-Sumner syndrome can resemble vasculitic mononeuritis multiplex. Sensory neuropathies similar to those in Sjögren syndrome may occur in patients with diabetes, paraneoplastic syndromes associated with anti-Hu antibodies, and pyridoxine deficiency.

TREATMENT

Systemic Vasculitis

For most vasculitic neuropathies (Chapter 270), oral prednisone (1 mg/kg) is appropriate in relatively mild cases, but intravenous methylprednisolone (1000 mg/day for 3 to 5 days) may be indicated as initial treatment in severe cases. Daily dosing is commonly used for the first 2 months or longer if the disease remains active. Subsequently, the dose is gradually tapered, with a transition to alternate-day dosing and discontinuation depending on the clinical picture and associated systemic features.

Corticosteroid treatment may be adequate for Churg-Strauss syndrome, but additional medication is generally needed in other forms of systemic vasculitic neuropathy. In most cases of granulomatosis with polyangiitis, combined therapy with glucocorticoids and oral cyclophosphamide (2 mg/kg/day) or weekly oral methotrexate (7.5 mg/week) is used. Azathioprine (2 to 3 mg/kg/day) is an alternative.

Because patients with nonsystemic vasculitic neuropathy may recover spontaneously or have a relatively benign course, low-dose or alternate-day oral prednisone (60 to 80 mg/day) is often adequate therapy. Azathioprine or weekly methotrexate can be used as a glucocorticoid-sparing agent. Doses such as 60 mg of prednisone on alternate days, 2 to 3 mg/kg/day of azathioprine, or 7.5 to 15 mg/week of methotrexate are reasonable starting doses that can ultimately be decreased if the treatments prove effective.

PROGNOSIS

Most systemic and nonsystemic vasculitis cases respond at least partially to treatment. The prognosis of patients with nonsystemic vasculitis is better than that of patients with systemic vasculitis, with fewer episodes of nerve damage; the disease may be monophasic or relapsing-remitting over a period of years. Most patients recover the ability to walk.

CRITICAL ILLNESS NEUROPATHY

DEFINITION

Critical illness polyneuropathy is an acute or subacute axonal length–dependent neuropathy that occurs in critically ill patients, not as a direct consequence of their underlying illness. The neuropathy is monophasic and recovers, at least in part, if the patient survives the underlying illness.[14]

EPIDEMIOLOGY

The incidence of critical illness polyneuropathy is uncertain because of variable diagnostic criteria. Moreover, it frequently accompanies critical illness myopathy (Chapter 421), which may be indistinguishable from it. Critical illness polyneuropathy frequently occurs in patients with systemic inflammatory response syndrome (Chapters 106 and 108), a generalized inflammatory host response to severe illness; up to 70% of patients with sepsis develop length-dependent axonal neuropathy.

PATHOBIOLOGY

Nerve biopsies have identified perivascular lymphocytic infiltration, macrophages, and cytokines such as interleukin-1β, interferon-γ, and interleukin-12. Ischemia caused by a sepsis-induced abnormal distribution of capillary blood flow, nutritional deprivation, and hypoglycemia has also been implicated in critical illness polyneuropathy.

CLINICAL MANIFESTATIONS

The typical finding is rapid development of profound limb weakness days to weeks after acquiring a severe illness that necessitated intensive care unit admission and ventilator support. Respiratory muscles are often involved, thereby resulting in inability to wean from the ventilator. Elicitable deep tendon reflexes distinguish critical illness polyneuropathy from acute inflammatory demyelinating polyneuropathy. Sensory testing is difficult to perform in severely ill patients but is usually normal.

DIAGNOSIS

Laboratory studies are rarely helpful. CSF protein is normal, unlike acute inflammatory demyelinating polyneuropathy. A lack of cells excludes infectious or inflammatory disorders. Creatine kinase levels are normal, unlike critical illness myopathy, in which they may be elevated.

Motor conduction velocities are normal with reduced or absent compound muscle action potential amplitudes. Critical illness polyneuropathy is predominantly a motor disorder, so sensory conduction velocities and sensory nerve action potential amplitudes are normal. Abnormal spontaneous activity reflecting axonal damage may occur within 1 to 3 weeks of onset. The presence of neuropathic and myopathic abnormalities suggests that critical illness polyneuropathy and critical illness myopathy coexist.

Differential Diagnosis

Distinguishing between critical illness myopathy and critical illness polyneuropathy can be difficult. Glucocorticoids and neuromuscular blocking agents predispose to critical illness myopathy (Chapter 421), which occurs in up to 5% of critically ill patients and is also manifested as rapidly progressive weakness of the limbs and diaphragm. Muscle biopsy and special electrical techniques aid in the diagnosis of critical illness myopathy. Acute inflammatory

demyelinating polyneuropathy can mimic critical illness polyneuropathy but can be distinguished by slow nerve conduction velocity and abnormal CSF. Acute motor axonal neuropathy or acute motor and sensory axonal neuropathy may be more difficult to distinguish, particularly if sepsis or another underlying disorder induces an abnormal CSF; nevertheless, the presence of antiganglioside antibodies may help distinguish these disorders from critical illness polyneuropathy. Myasthenia gravis (Chapter 422), botulinum toxin (Chapter 296), and other toxins can cause a similar clinical picture.

TREATMENT

Rx

Treatment is directed at the underlying disease or diseases, such as sepsis (Chapter 108). Glucocorticoids and neuromuscular blocking agents should be avoided if possible because both have been associated with critical illness myopathy.

PROGNOSIS

In-hospital mortality has been reported to be as high as 84% in patients with critical illness polyneuropathy compared with 50% in similarly ill patients without it. Although most patients improve if they survive their underlying illness, up to 10% have persistent severe limb weakness and are dependent on a ventilator. Most patients have some weakness 2 years after discharge.

DIABETIC AND OTHER METABOLIC NEUROPATHIES

DEFINITION

Diabetic peripheral neuropathies can be separated into two large groups: (1) symmetrical, predominantly sensory or autonomic neuropathies (or both); and (2) asymmetrical mononeuropathies or plexopathies.

EPIDEMIOLOGY

Diabetes (Chapter 229) is the most common cause of neuropathy in the Western world. Diabetic neuropathy occurs in 8 to 70% of patients with diabetes, depending on the criteria used to diagnose neuropathy, and patients with retinopathy or overt albuminuria are more than twice as likely to have neuropathy. Distal symmetrical polyneuropathies are the most common diabetic neuropathy, but distal autonomic neuropathy is also common.[15] For example, impotence develops in 20 to 60% of diabetic men, but widespread autonomic dysfunction develops in less than 5% of diabetic patients.

PATHOBIOLOGY

Distal Symmetrical Polyneuropathy and Autonomic Neuropathy

The pathogenesis of distal symmetrical polyneuropathy and autonomic neuropathy involves both microvascular and metabolic abnormalities, with a causal link between increased blood glucose levels and the development and progression of diabetic neuropathy. The mechanisms by which hyperglycemia causes nerve dysfunction may include activation of the polyol pathway, extensive glycation, altered diacylglycerol/protein kinase activity, and oxidative stress. Evidence from animal models suggests a role for neurotrophic factors, in particular, nerve growth factor, which selectively supports small fiber sensory and sympathetic neurons.

Acute Asymmetrical Neuropathies

The focal nature of these diabetic neuropathies is presumed to result from occlusion of endoneurial arterioles with resultant ischemic damage to the nerve. Changes suggestive of vasculitis are observed in epineurial and perineurial blood vessels in about 50% of cases, and perivascular lymphocytic infiltrates are common.

CLINICAL MANIFESTATIONS

Distal Sensory Polyneuropathy and Autonomic Neuropathy

Distal symmetrical polyneuropathy is typically manifested as insidious symmetrical sensory loss of small (pain and temperature) and large (proprioception) fiber modalities. Paresthesias or painful dysesthesias (e.g., burning or tingling feet) are common although not invariable. An unsteady gait may be the initial finding. Weakness is usually minimal, even in the distal foot muscles. Ankle reflexes are generally absent, although patellar reflexes may be present. Feet and distal calves are often cold and erythematous. Slow distal proximal progression of sensory symptoms and signs is the rule. By the time that symptoms reach the knees, abnormalities often begin in the hands.

When sensory changes reach the level of the knees, symptoms of autonomic neuropathy often begin: gastroparesis, constipation that may alternate with diarrhea, orthostatic hypotension, anhidrosis, cardiac arrhythmias, and impotence. Autonomic abnormalities can be the most disabling component of diabetic neuropathy.

Acute Asymmetrical Neuropathies

Asymmetrical, acute neuropathies cause focal or multifocal symptoms, depending on the peripheral nerve or nerves affected. They are usually accompanied by acute pain in the afflicted region. The pain may be deep and aching or throbbing and lancinating. Most cases of acute focal or multifocal diabetic neuropathy eventually resolve, at least partially. Pain may resolve within a few months, whereas weakness may take a year or more to recover and may persist. Characteristic manifestations include the following:

Diabetic lumbosacral radiculoplexus neuropathy. Patients are frequently elderly with type 2 diabetes (Chapter 229). Asymmetrical pain in the upper part of the thigh is followed by progressive weakness and atrophy of the proximal leg muscles. Progression to the other leg frequently occurs. In about 50% of affected patients, autonomic symptoms (Chapter 418), including orthostatic, hypotension, and gastrointestinal and sexual dysfunction, also develop. Weakness may progress, and about 50% of patients require a wheelchair for ambulation. Many patients require opiates for pain. After the nadir, the patient will usually stabilize for several months, followed by progressive improvement. As many as 50% of patients do not regain full ambulation.

Truncal radiculopathy. An acute, focal onset of pain and sensory loss develop over a region of the trunk. In extreme cases, the abdominal wall muscles may become weak, resembling a hernia. As with diabetic lumbosacral radiculoplexus neuropathy, at least partial improvement will occurs after a period of months, but the pain is difficult to control.

Cranial neuropathies. The classic manifestation is an acute oculomotor nerve palsy in which retro-orbital pain is followed by diplopia and ptosis. Pupillary fibers are often spared, thereby distinguishing the disorder from lesions that compress the oculomotor nerve and cause a dilated pupil. Similar findings may occur with the trochlear or abducens nerves. Bell palsy is more frequent in diabetic patients and is less likely to involve taste than in patients without diabetes.

Compressive mononeuropathies. Compressive neuropathies, such as carpal tunnel syndrome, occur more frequently in diabetic patients for unclear reasons. It is not known whether the response of carpal tunnel syndrome to treatment is as effective as when these mononeuropathies occur independently of diabetes.

DIAGNOSIS

Distal Sensory Polyneuropathy

The diagnosis of distal symmetrical polyneuropathy is based on identification of a predominantly sensory length–dependent neuropathy in the presence of either type 1 or type 2 diabetes. Neuropathy can develop independently of good control of blood sugar. Clinically similar neuropathies occur in patients with glucose abnormalities that are detectable only by oral glucose tolerance testing. Nerve conduction studies usually show low-amplitude or nondetectable sensory nerve action potential amplitudes; when detectable, sensory conduction may be slightly slow. Compound muscle action potential amplitudes are often reduced. Motor conduction studies are slightly slowed even if there is only minimal motor involvement clinically. Needle EMG in distal muscles typically demonstrates changes characteristic of chronic denervation. Occasional fibrillations and positive sharp waves may also be present.

Acute Asymmetrical Neuropathy

The focal neuropathies tend to occur in older patients with type 2 diabetes. The characteristic syndromes are diagnosed on the basis of their clinical manifestations and association with diabetes. EMG may demonstrate pronounced denervation in affected muscles. Concomitant evidence of distal symmetrical polyneuropathy is often present clinically and by electrophysiologic studies.

TREATMENT

Distal Symmetrical Polyneuropathy

An important treatment goal in distal symmetrical polyneuropathy is prevention of osteomyelitis (Chapter 272) and the resultant amputation of toes and feet. Because patients often do not sense injuries to their feet, diligence in foot care is important. Specific treatments to reverse or halt progression of the distal symmetrical polyneuropathy in diabetic patients are not yet available. Current therapy is based on control of hyperglycemia, management of symptoms, and foot care. Careful control of blood glucose (Chapter 229) remains the only treatment proved to delay the onset and slow progression of distal symmetrical polyneuropathy. There is, however, no HbA_{1c} threshold below which patients avoid risk for neuropathy. For the treatment of pain (Chapter 30), combination therapy with gabapentin (beginning at 300 mg three times a day but increasing up to more than 600 mg three times a day) and low doses of sustained-release morphine (e.g., 15 mg twice daily) is effective when less aggressive treatments fail. Pregabalin is effective at doses of 300 mg to 600 mg daily. Duloxetine (60 to 120 mg daily) and tricyclic antidepressant medications (e.g., amitriptyline 10 to 100 mg or nortriptyline 10 to 100 mg) may also be useful. Because of potential sedative side effects, these medications are usually begun with a low dose given at bedtime and then gradually increased based on benefits and toxicity.

Acute Asymmetrical Neuropathy

Although acute asymmetrical neuropathy generally improves spontaneously, improvement may take months and remain incomplete. Intravenous corticosteroids, IVIG, or plasma exchange may improve the speed and extent of recovery, but these treatments are not of established benefit. Pain management is similar to that for distal symmetrical polyneuropathy. Because the pain in patients with diabetic lumbosacral radiculoplexus neuropathy and truncal radiculopathy is focal, topical therapy with capsaicin may be effective.

INFECTIOUS NEUROPATHIES

Neuropathies Associated with HIV Infection

The peripheral nervous system may be involved in all phases of HIV infection (Chapter 394). The most common peripheral neuropathy is a distal, painful, sensory axonal polyneuropathy that is similar to the toxic neuropathy caused by nucleoside reverse transcriptase inhibitors (NRTIs), including zidovudine, zalcitabine, didanosine, stavudine, and lamivudine. When an iatrogenic neuropathy is suspected, discontinuation of NRTIs may improve symptoms. Conversely, a neuropathy caused by HIV is likely to stabilize or improve with antiretroviral treatment.

Inflammatory neuropathies such as chronic or acute inflammatory demyelinating polyneuropathy can also occur in the early stages of HIV infection; the CSF cyto-albumin dissociation usually seen with these conditions may not be evident in these patients because of a mild CSF mononuclear pleocytosis. The response of these neuropathies to plasma exchange or IVIG is generally good. In later stages of HIV infection, cytomegalovirus (Chapter 376) may cause either an acute lumbosacral polyradiculopathy as a result of direct invasion of nerve roots or a mononeuritis multiplex through a vasculitic mechanism.

Neuropathies Associated with Herpes Zoster

Varicella-zoster virus (Chapter 375) usually remains latent in cranial or spinal ganglia after resolution of a systemic infection. Reactivation, which is more frequent in elderly and immunocompromised patients, causes a vesicular skin eruption accompanied by pruritus and dysesthesias. Herpes zoster resolves spontaneously but is frequently followed by post-herpetic neuralgia, which is characterized by severe pain persisting for more than 6 weeks after the rash appears. Early treatment with oral acyclovir (800 mg, five times daily for 7 days) may reduce both the duration of the acute phase and the risk for post-herpetic neuralgia. The use of concomitant corticosteroids in addition to acyclovir improves acute pain without exacerbating viral spread but does not reduce the incidence or severity of post-herpetic neuralgia.

Neuropathy Associated with Lyme Disease

Borrelia burgdorferi causes a disease with three stages (Chapter 321). In the first stage, shortly after and in the same area of a tick bite, a nonpruritic rash (erythema migrans) appears and spontaneously disappears after a few weeks. The second stage is frequently associated with neurologic complications such as lymphocytic meningitis and focal and multifocal peripheral and cranial neuropathies; characteristic manifestations are unilateral or bilateral facial palsy and radiculitis. The third stage is associated with severe neurologic complications, including encephalopathy, encephalomyelitis, and a predominantly sensory axonal polyneuropathy. A lymphocytic pleocytosis in CSF and demonstration of *B. burgdorferi* infection in serum or CSF are the main laboratory findings. Treatment is discussed in Chapter 321.

Neuropathy Associated with Leprosy

Leprosy (Chapter 326) is infrequent in the United States but is the most common cause of peripheral neuropathy in some developing countries. Leprosy presents in different forms, depending on the host's immune system. Patients with normal cell-mediated immunity are more likely to have a tuberculoid form characterized by hypopigmented skin lesions associated with decreased sensation. In patients with abnormal cell-mediated immunity, the more severe lepromatous form with large disfiguring lesions may develop. A mononeuritis multiplex pattern with prominent superficial sensory loss is the most typical clinical manifestation of leprosy. If treated early, neuropathies in leprosy improve. World Health Organization recommendations call for combination therapy that includes dapsone (50 to 100 mg/day or 200 to 250 mg/week), rifampicin (600 mg monthly), and clofazimine (100 mg/day) (Chapter 326).

Neuropathy Associated with Diphtheria

Vaccination has made diphtheria (Chapter 292) rare in developed countries, but it is an important cause of subacute neuropathy in developing countries. Some strains of *Corynebacterium diphtheriae* produce a potent neurotoxin that causes palatal weakness, lens accommodation deficits, and extraocular palsies. These acute manifestations are followed by limb paralysis that resembles acute inflammatory demyelinating polyneuropathy (see earlier). The neuropathy caused by the neurotoxin usually resolves with resolution of the infection. The diphtheria organism can be eradicated by therapy with antibiotics such as erythromycin (2 g/day intravenously divided twice daily for adults) or penicillin (procaine penicillin G, 1.2 million U/day intramuscularly divided twice daily for 14 days). However, the neuropathy, as with other manifestations of the disease, generally requires treatment with diphtheria antitoxin, a hyperimmune antiserum produced in horses. Depending on the severity of the disease, antitoxin is administered intramuscularly or intravenously (80,000 to 120,000 units for extensive disease for 3 or more days; Chapter 292).

TOXIC AND DEFICIENCY SYNDROMES

In Western countries, toxic neuropathies are frequently the side effects of medications[16] rather than a result of environmental exposure. In most cases, iatrogenic neuropathy is manifested as a length-dependent or "dying-back" axonal neuropathy. Treatment requires a correct diagnosis (Table 420-5) and discontinuation of the drug. Improvement often takes many months. In a randomized trial of patients with painful chemotherapy-induced peripheral neuropathy, duloxetine (30 mg orally once daily for 1 week, then 60 mg daily for 4 weeks) resulted in a greater reduction in pain compared with placebo.

Compressive Neuropathies

Peripheral nerves are vulnerable to chronic compression in many sites: median nerve compression at the wrist within the carpal tunnel *(carpal tunnel syndrome)*, median nerve compression in the upper part of the forearm, ulnar nerve compression in the hand *(cubital tunnel syndrome)*, ulnar nerve compression at the elbow or wrist, tibial nerve compression behind the medial malleolus *(tarsal tunnel syndrome)*, and peroneal nerve compression over the lateral fibular head.

TABLE 420-5 TOXIC AND DEFICIENCY NEUROPATHIES

Associated with antineoplastic agents: vincristine, paclitaxel (Taxol), cisplatin, suramin, thalidomide
Associated with antimicrobials: chloroquine, dapsone, isoniazid, metronidazole, nitrofurantoin
Associated with cardiac medications: amiodarone, perhexiline, hydralazine
Associated with other medications: colchicine, tacrolimus, gold salts, phenytoin, disulfiram (Antabuse), pyridoxine (vitamin B_6)
Associated with heavy metals: lead, arsenic, mercury, thallium
Associated with chemical compounds: acrylamide, carbon disulfide, ethylene glycol, hexacarbons, organophosphate esters, vacor
Deficiency neuropathies: vitamin B_1 deficiency, vitamin B_{12} deficiency, vitamin E deficiency

CARPAL TUNNEL SYNDROME

Entrapment of the median nerve at the wrist reflects the limited space available for the median nerve because of the surrounding bone, joint, ligaments, tendons, and synovium. Repetitive motion of the fingers is an exacerbating element. Other precipitating factors include trauma, osteoarthritis, synovial cysts, myxedema, and amyloid deposition. Symptoms typically include paresthesias of the first three fingers, often at night, and are relieved by shaking or elevating the hand. In severe disease, objective sensory loss in the median nerve distribution, weakness of median-innervated muscles such as the abductor pollicis brevis, and prolongation of nerve conduction across the carpal tunnel (prolonged distal latency) are characteristic. The diagnosis is supported by identification of *Tinel sign,* in which tapping the carpal tunnel elicits paresthesias in the median nerve distribution, and by paresthesias produced by sustained flexion of the wrist *(Phalen sign)*. Treatment begins with splinting of the wrist in slight dorsiflexion during sleep. Injection of corticosteroids into the carpal tunnel provides temporary benefit. Severe carpal tunnel syndrome is treated surgically by release of the carpal ligament.

Bell Palsy

Unilateral facial paralysis of acute onset frequently occurs on an idiopathic basis (Bell palsy). The diagnosis is one of exclusion. Facial nerve palsies also occur in the setting of *herpes zoster oticus* and are associated with otalgia and varicelliform lesions affecting the external ear, ear canal, or tympanic membrane. Facial paralysis of a lower motor neuron type can be caused by carcinomatous meningitis (Chapters 195 and 412), sarcoidosis (Chapter 95), Lyme disease (Chapter 321), and HIV infection (Chapter 394).

Primary tumors of the facial nerve can cause rapidly developing facial paralysis. Facial paralysis can also occur in *CNS disease* affecting the pontomedullary junction, such as stroke or multiple sclerosis (Chapter 411).

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Most cases of facial paralysis are idiopathic. Patients typically notice facial paralysis on inspection in the mirror in the morning. Facial paralysis may be heralded or accompanied by pain behind the ear. The severity of paralysis varies widely.

TREATMENT

Rx

In a randomized trial, 10 days of oral corticosteroids (prednisolone 25 mg twice daily for 10 days) administered early in the course increased the return of facial function from 63 to 83% at 3 months in patients with idiopathic Bell palsy, but acyclovir was of no benefit. In severe cases, protection of the cornea from drying and injury is essential.

PROGNOSIS

Most patients improve, but about 10% of patients have little recovery. Aberrant regeneration of the facial nerve can cause synkinesias, such as "jaw winking" (when the eye is closed) or tearing accompanying salivation ("syndrome of crocodile tears").

Trigeminal Neuralgia (Tic Douloureux)

Trigeminal neuralgia and other painful cranial neuralgias are discussed in Chapter 398.

Grade A References

A1. Hughes RA, Swan AV, van Doorn PA. Intravenous immunoglobulin for Guillain-Barré syndrome. *Cochrane Database Syst Rev*. 2014;9:CD002063.

A2. Raphael JC, Chevret S, Hughes RA, et al. Plasma exchange for Guillain-Barré syndrome. *Cochrane Database Syst Rev*. 2012;7:CD001798.

A3. El-Bayoumi MA, El-Refaey AM, Abdelkader AM, et al. Comparison of intravenous immunoglobulin and plasma exchange in treatment of mechanically ventilated children with Guillain Barré syndrome: a randomized study. *Crit Care*. 2011;15:R164.

A4. Hughes RA, van Doorn PA. Corticosteroids for Guillain-Barré syndrome. *Cochrane Database Syst Rev*. 2012;8:CD001446.

A5. van Koningsveld R, Schmitz PI, Meche FG, et al. Effect of methylprednisolone when added to standard treatment with intravenous immunoglobulin for Guillain-Barré syndrome: randomised trial. *Lancet*. 2004;363:192-196.

A6. Eftimov F, Winer JB, Vermeulen M, et al. Intravenous immunoglobulin for chronic inflammatory demyelinating polyradiculoneuropathy. *Cochrane Database Syst Rev*. 2013;12:CD001797.

A7. Eftimov F, Vermeulen M, van Doorn PA, et al. Long-term remission of CIDP after pulsed dexamethasone or short-term prednisolone treatment. *Neurology*. 2012;78:1079-1084.

A8. Callaghan BC, Little AA, Feldman EL, et al. Enhanced glucose control for preventing and treating diabetic neuropathy. *Cochrane Database Syst Rev*. 2012;6:CD007543.

A9. Gilron I, Bailey JM, Tu D, et al. Morphine, gabapentin, or their combination for neuropathic pain. *N Engl J Med*. 2005;352:1324-1334.

A10. Moore RA, Straube S, Wiffen PJ, et al. Pregabalin for acute and chronic pain in adults. *Cochrane Database Syst Rev*. 2009;3:CD007076.

A11. Lunn MP, Hughes RA, Wiffen PJ. Duloxetine for treating painful neuropathy, chronic pain or fibromyalgia. *Cochrane Database Syst Rev*. 2014;1:CD007115.

A12. Griebeler ML, Morey-Vargas OL, Brito JP, et al. Pharmacologic interventions for painful diabetic neuropathy: an umbrella systematic review and comparative effectiveness network meta-analysis. *Ann Intern Med*. 2014;161:639-649.

A13. Smith EM, Pang H, Cirrincione C, et al. Effect of duloxetine on pain, function, and quality of life among patients with chemotherapy-induced painful peripheral neuropathy: a randomized clinical trial. *JAMA*. 2013;309:1359-1367.

A14. Jarvik JG, Comstock BA, Kliot M, et al. Surgery versus non-surgical therapy for carpal tunnel syndrome: a randomised parallel-group trial. *Lancet*. 2009;374:1074-1081.

A15. Gronseth GS, Paduga R. Evidence-based guideline update: steroids and antivirals for Bell palsy: report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology*. 2012;79:2209-2213.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Tazir M, Hamadouche T, Nouioua S, et al. Hereditary motor and sensory neuropathies or Charcot-Marie-Tooth diseases: an update. *J Neurol Sci*. 2014;347:14-22.
2. Rossor AM, Polke JM, Houlden H, et al. Clinical implications of genetic advances in Charcot-Marie-Tooth disease. *Nat Rev Neurol*. 2013;9:562-571.
3. Winer JB. An update in Guillain-Barré syndrome. *Autoimmune Dis*. 2014;2014:793024.
4. van den Berg B, van der Eijk AA, Pas SD, et al. Guillain-Barré syndrome associated with preceding hepatitis E virus infection. *Neurology*. 2014;82:491-497.
5. Kuwabara S, Yuki N. Axonal Guillain-Barré syndrome: concepts and controversies. *Lancet Neurol*. 2013;12:1180-1188.
6. van den Berg B, Bunschoten C, van Doorn PA, et al. Mortality in Guillain-Barré syndrome. *Neurology*. 2013;80:1650-1654.
7. Drenthen J, Jacobs BC, Maathuis EM, et al. Residual fatigue in Guillain-Barré syndrome is related to axonal loss. *Neurology*. 2013;81:1827-1831.
8. Eftimov F, van Schaik I. Chronic inflammatory demyelinating polyradiculoneuropathy: update on clinical features, phenotypes and treatment options. *Curr Opin Neurol*. 2013;26:496-502.
9. Bright RJ, Wilkinson J, Coventry BJ. Therapeutic options for chronic inflammatory demyelinating polyradiculoneuropathy: a systematic review. *BMC Neurol*. 2014;14:26.
10. Dispenzieri A. POEMS syndrome: 2014 update on diagnosis, risk-stratification, and management. *Am J Hematol*. 2014;89:214-223.
11. Bayas A, Gold R, Naumann M. Long-term treatment of Lewis-Sumner syndrome with subcutaneous immunoglobulin infusions. *J Neurol Sci*. 2013;324:53-56.
12. Muppidi S, Vernino S. Paraneoplastic neuropathies. *Continuum (Minneap Minn)*. 2014;20:1359-1372.
13. Koike H, Sobue G. Paraneoplastic neuropathy. *Handb Clin Neurol*. 2013;115:713-726.
14. Latronico N, Bolton CF. Critical illness polyneuropathy and myopathy: a major cause of muscle weakness and paralysis. *Lancet Neurol*. 2011;10:931-941.
15. Dixit S, Maiya A. Diabetic peripheral neuropathy and its evaluation in a clinical scenario: a review. *J Postgrad Med*. 2014;60:33-40.
16. Park SB, Goldstein D, Krishnan AV, et al. Chemotherapy-induced peripheral neurotoxicity: a critical analysis. *CA Cancer J Clin*. 2013;63:419-437.

421

MUSCLE DISEASES

DUYGU SELCEN

DEFINITION

Muscle diseases, which are also called myopathies, are disorders of skeletal muscle structure or function. Myopathies can be primary and occur in isolation, or they can be part of a multisystem disorder.

EPIDEMIOLOGY

Many muscle diseases (Table 421-1) are inherited as autosomal dominant, autosomal recessive, X-linked, or maternal (mitochondrial) conditions. Environmental factors that may precipitate myopathies include recent infection, foreign travel, exposure to medications such as statins, and alcohol abuse (Chapter 33). Exercise commonly precipitates symptoms in patients with metabolic myopathies, whereas exposure to cold and high carbohydrate or potassium-rich food can precipitate weakness in muscle channelopathies.

The prevalence of muscle disease is estimated to be about 1 per 1000 people, including acute and transient disorders (e.g., myositis owing to infectious or toxic causes) and chronic inflammatory or genetic disorders that cause substantial morbidity over decades or a lifetime. Myopathies can cause premature death owing to neuromuscular weakness and secondary respiratory infections or to involvement of other organs in multisystem diseases. Myocardial involvement, which is particularly common in some muscle diseases, can cause heart failure or life-threatening arrhythmias.

TABLE 421-1 CLASSIFICATION OF MYOPATHIES

HEREDITARY
Muscular dystrophies
Congenital myopathies
Myotonia and channelopathies
Metabolic myopathies
Mitochondrial myopathies
ACQUIRED
Inflammatory myopathies
Endocrine myopathies
Myopathies associated with systemic illness
Drug-induced/toxic myopathies

Adapted from Goldman L, Schafer AI, eds. *Cecil Medicine*. 24th ed. Philadelphia: Elsevier; 2012.

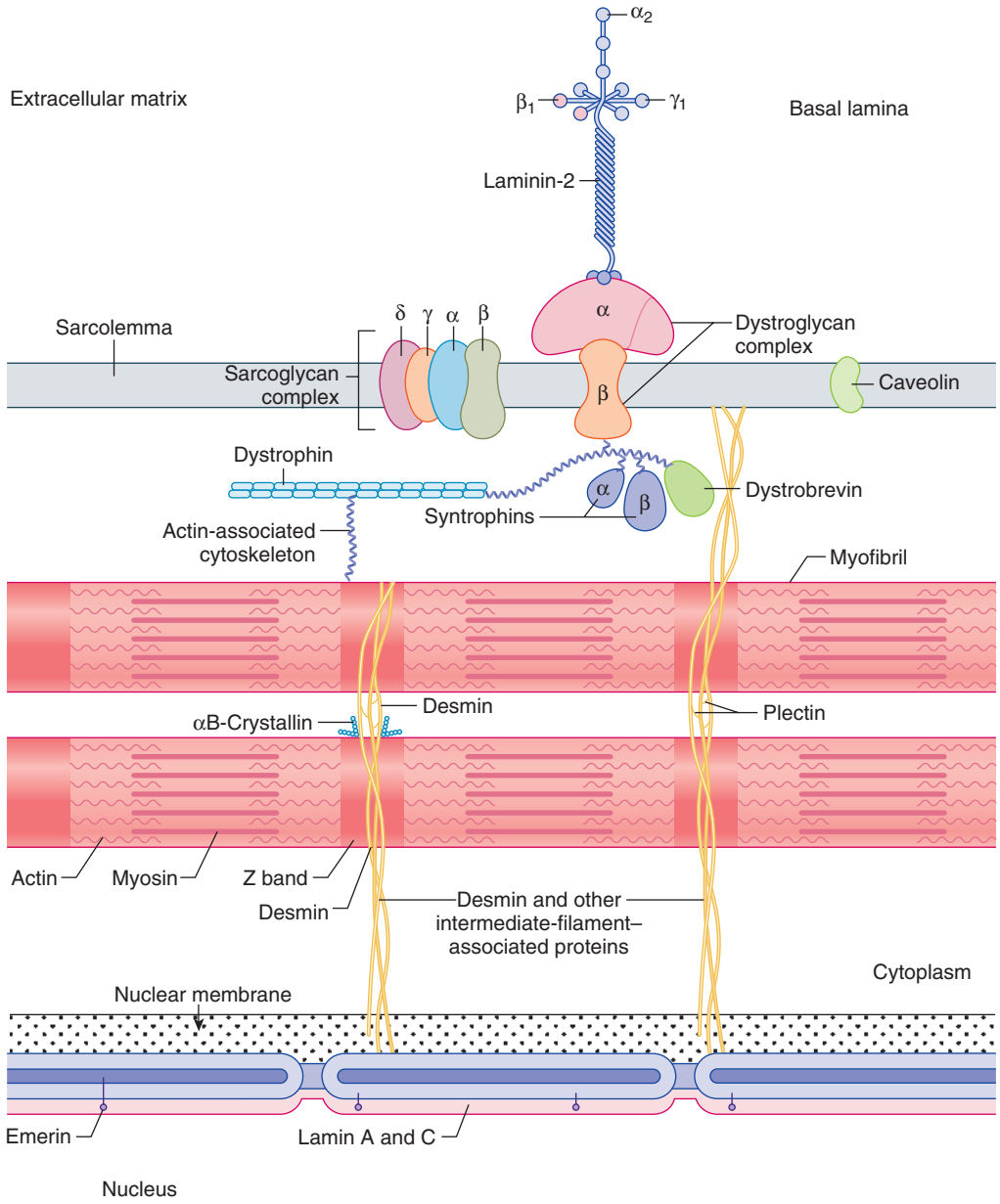

FIGURE 421-1. Muscle structure. (From Goldman L, Schafer AI, eds. *Cecil Medicine*. 24th ed. Philadelphia: Elsevier; 2012.)

PATHOBIOLOGY

Muscle disease can result from a perturbation in the anatomy or any of the physiologic processes required for muscle contraction or the genes that control them. Skeletal muscle is part of a motor unit, which is defined as the anterior horn cell body, its axon, the neuromuscular junction, and the skeletal muscle fibers innervated by the one axon. The motor unit is coordinated in a manner that allows efficient muscle contraction and function. The number of muscle fibers innervated by each motor unit varies from a few (e.g., in muscles controlling very precise movements, such as extraocular muscles) to more than 1000 (e.g., large and powerful but less precise muscles, such as the quadriceps).

Skeletal muscle is composed of myriad muscle fibers. Muscle fibers, which are multinucleated cells formed by fusion of myoblasts during development, are surrounded by a plasma membrane, the sarcolemma, which is surrounded by a basal lamina and endomysial connective tissue. Groups of muscle fibers compose the fascicles, which are surrounded by perimysium, and the groups of fascicles in turn are surrounded by epimysium. Nerve branches, blood vessels, muscle spindles, and fat cells lie within the connective tissue of the muscle.

Each muscle fiber is composed of myofibrils, which themselves are composed of repeat units a few microns long called sarcomeres. The sarcomere consists of a highly organized protein network that gives the muscle fiber its characteristic striated appearance. Each sarcomere is flanked by two Z discs. Z discs are composed of multiple proteins, including α-actinin. Emanating from the Z disc are thin filaments, composed of actin, troponin, and tropomyosin. Thick filaments consist of myosin. Other structures comprise subcellular organelles, including the mitochondria, which are the principal energy source, the endoplasmic reticulum, and the transverse tubules that communicate with the extracellular space.

Muscle function (Fig. 421-1) is dependent on chemical energy from adenosine triphosphate (ATP). In the first 30 minutes of sustained activity, ATP is produced by the breakdown of glycogen (glycolysis), and after 30 minutes ATP is produced by fatty acid β-oxidation and oxidative phosphorylation within the mitochondria. The process that leads to muscle contraction begins with the generation of the muscle fiber action potential (Chapter 53), which initiates muscle contraction after it is propagated into the interior of muscle fiber through the transverse tubular system. The release of calcium from the endoplasmic reticulum triggers a coordinated series of events that lead to the coupling of excitation to contraction. Calcium binds to troponin, which interacts with tropomyosin and results in actin-myosin binding. The repeated formation and cleavage of actin-tropomyosin cross-bridges, in an ATP-dependent process, results in sliding of thick and thin filaments and shortening of the sarcomere.

The structural integrity of the muscle fiber surface membrane is maintained by a network of proteins within the muscle. Dystrophin is a key component of the subsarcolemmal cytoskeleton. In combination with several glycoproteins called sarcoglycans (α, β, δ, γ), dystroglycans (α, β), and

syntrophins (α, β1, β2), which form the dystrophin-sarcoglycan complex, it anchors the contractile elements of the muscle fiber to the sarcolemma and to the extracellular basal lamina. The basal lamina contains several important proteins, such as collagen, fibronectin, and laminin, which includes merosin and related proteins. The intermediate proteins, including desmin, connect the Z disc and other organelles to the subsarcolemmal cytoskeleton.

CLINICAL MANIFESTATIONS

Muscle diseases often present with localized or diffuse muscle weakness, reduced exercise tolerance, resting or exercise-induced muscle pain, muscle enlargement or atrophy, cramps, delayed relaxation, or rarely, myoglobinuria. These symptoms and signs can be masked by other neurologic or systemic features in patients with multisystem diseases.

History

The assessment of patients with neuromuscular diseases begins with a careful history, general physical examination, and detailed neurologic examination. The age of onset, the rate of progression, and whether the process is episodic, static, or progressive can provide important clues. Congenital and childhood onset myopathies can be associated with reduced fetal movements, breech delivery, weak cry or suck, and the delayed acquisition of motor milestones. Weakness is the most common presenting symptom, but other symptoms of muscle disease include muscle pain, reduced exercise intolerance, change in the muscle bulk (hypertrophy or atrophy), abnormal spontaneous muscle activity, delayed relaxation, fatigue, or myoglobinuria. Weakness may be relatively static as in some congenital myopathies, progressive as in muscular dystrophies, intermittent as in periodic paralysis, fluctuating as in neuromuscular junction disorders (Chapter 422), or exercise related as in metabolic myopathies. The most common distribution of weakness is proximal or limb-girdle weakness, which results in difficulties in getting out of low chairs, a bathtub, or a car seat; climbing up and down stairs; arising from squat; or getting off the floor. Proximal arm weakness manifests as difficulty reaching to shelves, washing or brushing hair, or raising arms to put on a shirt. Distal leg weakness can lead to difficulty walking on uneven surfaces, tripping over curbs, difficulty standing on the toes, or slapping feet owing to footdrop. Distal upper limb weakness results in difficulty opening jars, typing at a keyboard, writing, or buttoning clothes. Bilateral facial weakness can result in difficulty whistling, blowing up balloons, or drinking through a straw. Predominant involvement of ocular muscles can produce ptosis and diplopia. Weakness of the bulbar muscles manifests as difficulties with speech and swallowing, neck weakness that can lead to a dropped head, and respiratory muscle weakness that can lead to symptoms suggestive of nocturnal hypoventilation or respiratory failure. The early recognition of progressive respiratory failure is essential because it is treatable with noninvasive positive-pressure ventilation.

Fatigue and exercise intolerance can be presenting symptoms of muscle diseases, but they can also be multifactorial and nonspecific. In isolation, these symptoms usually do not indicate a primary muscle disease.

Muscle pain is another nonspecific symptom that can arise from many systemic and psychiatric conditions. Sometimes patients describe aching, stiffness, numbness, or burning as pain. Muscle diseases rarely cause diffuse, generalized, or persistent muscle pain. Muscle pain without muscle weakness is often a feature of fibromyalgia (Chapter 274) or chronic fatigue syndrome. Diffuse myalgia can occur in inflammatory muscle disease such as polymyositis or dermatomyositis, vasculitis, or viral or parasitic myositis. Muscle pain precipitated by exercise usually suggests a metabolic myopathy.

Muscle cramps are involuntary painful contractions that may occur in healthy individuals. Dehydration, renal failure (Chapter 130), and electrolyte imbalances (Chapters 116, 117, 119, and 245) can also produce muscle cramps. Muscle stiffness can occur in inflammatory, metabolic, and ion-channel diseases as well as in conditions such as multiple sclerosis (Chapter 411), polymyalgia rheumatica (Chapter 271), and connective tissue diseases (Chapter 256).

Fasciculations are caused by spontaneous firing of muscle fibers that are innervated by a single motor unit. Fasciculations may occur in normal persons, in whom they are usually exacerbated by stress and increased caffeine intake. Fasciculations in association with muscle weakness suggest anterior horn cell disease. Myotonia, often described as muscle stiffness, is characterized by prolonged contraction and delayed relaxation of muscle.

Myotonia can affect limb, facial, or bulbar muscles, and it can lead to persistent limb muscle contraction, eyelid closure, or dysphagia. Myotonic dystrophy is the most common muscle disease associated with myotonia, but patients usually complain more of weakness than the myotonia. Conversely, the myotonia associated with sodium and chloride channelopathies can be disabling. Patients who describe locking of their hands but do not have objective myotonia rarely have a physical explanation for their symptoms.

Tetany is the most severe form of sustained muscle contraction. Tetany occurs in patients with hypocalcemia and hypomagnesemia (Chapter 119), and it is aggravated by metabolic or respiratory alkalosis (Chapter 118).

Severe acute muscle damage, termed *rhabdomyolysis* (Chapters 113), results in myoglobinuria that presents as dark brown or red urine. Such discoloration must be distinguished from other causes of pigmenturia (Chapter 114) such as hemolysis or porphyria.

The detailed family history should include questions about muscle disease, including specific questions about the use of canes, braces, or wheelchairs. It also should assess whether family members have had a cardiomyopathy, unexpected sudden death, diabetes, or cataracts.

Physical Examination

A full physical examination must look for signs that may suggest any of the systemic diseases that are associated with myopathies. The skin examination can give clues to systemic illness, such as the heliotrope rash of dermatomyositis (Chapter 269).

A comprehensive neurologic examination should be performed in each patient to exclude possible central or peripheral nervous system disorders (Table 421-2). The examination begins as soon as the patient enters the examination room. Proximal leg muscle weakness may be evident if patients push themselves up on their thighs or have a waddling gait. Patients should be examined for possible facial muscle weakness or wasting, ptosis, or characteristic dysmorphic features, such as with myotonic dystrophy (Fig. 421-2), that can lead to an immediate clinical diagnosis.

Patients should be asked to rise from a squatting position and walk on their toes to assess possible calf weakness and on their heels to assess ankle dorsiflexion weakness. Patients should be asked to stand to assess posture and any evidence of rigidity or scoliosis. Joints should be moved passively to assess for contractures.

All muscle groups should be inspected for evidence of involuntary movements, atrophy, or hypertrophy. Muscles should be palpated for tenderness

TABLE 421-2 CLINICAL FINDINGS DIFFERENTIATING MUSCLE FROM NERVE DISEASE

FINDING	MYOPATHY	ANTERIOR HORN CELL DISEASE	PERIPHERAL NEUROPATHY	NEUROMUSCULAR JUNCTION DISEASE
Distribution	Usually proximal and symmetrical but can be distal or asymmetrical at onset	Distal, asymmetrical, and bulbar	Distal, symmetrical	Extraocular, bulbar, proximal limb, but sometimes distal
Atrophy	Slight early, marked late	Marked early	Moderate	Absent
Fasciculations	Absent	Frequent	Sometimes present	Absent
Reflexes	Lost late	Variable, can be hyperreflexic	Lost early	Normal or hyporeflexic
Pain	Variable	Absent	Variable, distal when present	Absent
Cramps	Rare	Frequent	Occasional	Absent
Sensory loss	Absent	Absent	Usually present	Absent
Serum creatine kinase	Usually elevated	Occasionally mildly elevated	Normal	Normal

Adapted with revision from Goldman L, Schafer AI, eds. *Cecil Medicine*. 24th ed. Philadelphia: Elsevier; 2012.

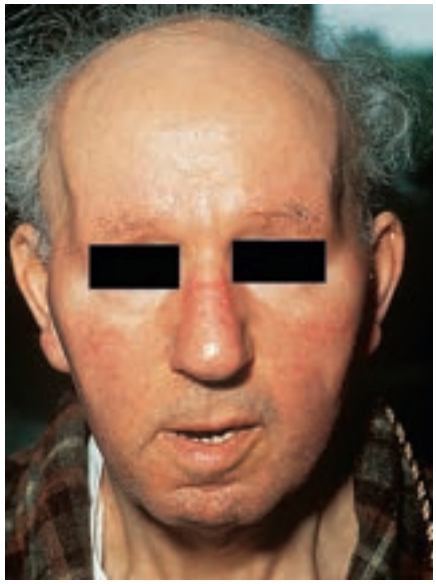

FIGURE 421-2. Myotonic dystrophy in a 50-year-old man. His appearance is typical with facial weakness, atrophy of the temporal muscles and sternocleidomastoids, and frontal baldness, which gives a monklike appearance. (From Goldman L, Schafer AI, eds. *Cecil Medicine*. 24th ed. Philadelphia: Elsevier; 2012.)

TABLE 421-3 MEDICAL RESEARCH COUNCIL SCALE OF MUSCLE STRENGTH

GRADE	DEGREE OF STRENGTH
5	Normal power
4	Active movement against gravity and resistance (often subdivided into 4–, 4, and 4+)
3	Active movement against gravity, but not against resistance
2	Active movement, with gravity eliminated
1	Observable muscle contraction, but not capable of initiating movement
0	No contraction

Adapted from Goldman L, Schafer AI, eds. *Cecil Medicine*. 24th ed. Philadelphia: Elsevier; 2012.

TABLE 421-4 GUIDING PRINCIPLES FOR ASSESSING MUSCLE DISEASES

1. History
- Age of onset (most inherited and acquired disorders have a characteristic age of onset)
- Rate of progression (acute suggests an acquired, often inflammatory cause)
- Fluctuating weakness (may indicate a neuromuscular junction disorder, metabolic myopathy, or a channelopathy)
- Relationship to exercise (may indicate metabolic myopathy or channelopathy)
- Muscle pain (may indicate inflammatory or metabolic myopathy)
- Relevant multisystem involvement (may indicate a mitochondrial cytopathy or myotonic dystrophy)
- Family history (may indicate a genetically determined chronic muscle disease)

2. Pattern of weakness
- Limb-girdle weakness is relatively nonspecific.
- Inherited disorders often have specific patterns of muscle involvement.
- Fluctuating weakness may indicate a neuromuscular junction disorder, metabolic myopathy, or a channelopathy.

3. Testing
- The serum creatine kinase can be normal.
- Electromyography is often normal in metabolic myopathies.
- All patients with muscular dystrophy should have an electrocardiogram and echocardiogram to look for cardiomyopathy and/or conduction defect.
- Muscle magnetic resonance imaging is increasingly being used to guide muscle biopsy and may sometimes reveal diagnostic patterns of muscle involvement.
- Some inherited muscle diseases can be diagnosed clinically and confirmed by genetic testing without requiring other investigations (e.g., myotonic dystrophy; Duchenne, Becker, and facioscapulohumeral dystrophy).
- Muscle biopsy will reveal the cause in most cases; it identifies specific inflammatory myopathies and muscular dystrophies and often provides diagnostic clues for other muscle and neurogenetic disorders.

Adapted with revision from Goldman L, Schafer AI, eds. *Cecil Medicine*. 24th ed. Philadelphia: Elsevier; 2012.

or unusual texture. Myotonia can be assessed by the inability to relax the muscle belly after percussion with a reflex hammer or the inability to relax the fingers from a firm grip.

Strength should be graded (Table 421-3) in each muscle group. Observing children and infants when they play with toys and how they stand up and walk usually reveals more than formal manual muscle strength testing. The pattern of muscle involvement can provide clues for the diagnosis of a specific myopathy.

DIAGNOSIS

Neurophysiologic testing, measurement of serum creatine kinase (CK), muscle biopsy, and genetic testing help guide the diagnosis of muscle diseases (Table 421-4).

A complete blood count and serum levels of alanine aminotransferase, aspartate aminotransferase, and creatinine can assess possible systemic involvement. An elevated erythrocyte sedimentation rate or C-reactive protein level is found in some inflammatory myopathies and is typical of connective tissue disorders (Chapters 256 and 257). Additional tests to evaluate patients suspected of having inflammatory myopathy or connective tissue disease can include antinuclear antibodies, extractable nuclear antigens, rheumatoid factor, antineutrophilic cytoplasmic antibodies, anti-Jo-1, anti-Mi2, anti-MDA5, anti-3-hydroxy-3-methylglutaryl-coenzyme A reductase, or anti–signal recognition particle antibodies (Chapter 257).

The muscle isoform (MM) of CK is frequently elevated in patients with muscle disease, although a normal level is seen in metabolic myopathies and in some chronic myopathies. A mild to moderate increase in the serum CK level can occur in patients with peripheral neuropathy, radiculopathy, and anterior horn cell diseases. The serum CK level is markedly increased in dystrophinopathies, dysferlinopathy, some of the sarcoglycanopathies α-dystroglycanopathies, and during rhabdomyolysis (Chapter 113). However, it can decline later in muscular dystrophies as the disease progresses. When the serum CK level exceeds the upper limit of normal by about 10-fold, levels of alanine aminotransferase, aspartate aminotransferase, and lactate dehydrogenase also can be elevated, and some patients can be initially misdiagnosed as having hepatitis (Chapters 147 and 148) before the serum CK level is measured. The blood lactate level can be increased in patients with a mitochondrial myopathy, but a normal value does not exclude this diagnosis. In patients with acute muscle pain and/or weakness, electrolytes and thyroid function tests should be checked.

In patients who are diagnosed with dermatomyositis (Chapter 269), the evaluation should include a search for an underlying malignancy. In patients with suspected mitochondrial cytopathy, a serum and/or spinal fluid lactate level should be obtained. In patients with suspected fatty acid β-oxidation defects, blood acylcarnitine profile may be useful for the diagnosis.

Electromyography

Electromyography (EMG) consists of nerve conduction studies, repetitive nerve stimulation, and needle examination of muscles (Chapter 396). Nerve conduction studies (Chapter 420) are normal in patients with myopathies. In myopathies, the needle examination typically shows complex, polyphasic, low-amplitude motor unit potentials. Myotonia is caused by recurrent depolarization of the muscle fiber surface membrane and has characteristic waxing and waning rhythmical discharges or fibrillation potentials during the needle examination on an EMG. Contractures are electrically silent on EMGs, but muscle cramps are associated with high-amplitude, high-frequency bursts of motor unit activity. The EMG can be normal in some focal myopathies, such as inflammatory myositis, in metabolic myopathies, and in some congenital myopathies.

Genetic Testing

The widespread availability of molecular genetic testing has revolutionized the approach to patients who are suspected of having hereditary muscle disease. Dystrophin genetic testing can identify the defects in dystrophin in 90 to 95% of the patients with Duchene and Becker muscular dystrophy. Commercial test panels are available for muscle diseases such as limb-girdle

muscular dystrophies, congenital muscular dystrophies, myofibrillar myopathies, and congenital myopathies; however, the test panels have only about an 85% sensitivity because they do not cover the entire coding region (unlike whole exome sequencing) or genome (unlike whole genome sequencing). The sensitivity is higher if multiple affected and some unaffected family members and more than one family are analyzed at the same time. Novel genes causing muscular dystrophies are still being discovered using whole exome sequencing.

Muscle Biopsy

Despite advances in genetics and molecular biology, muscle biopsy remains a key component in the diagnosis of most muscle diseases. The biopsy site should be carefully chosen from a clinically affected but not too severely involved muscle. Fresh-frozen sections of the specimens should be used for histochemical studies because even marked morphologic alterations may be undetected in paraffin-embedded tissue. Immunocytochemical localization of specific proteins is useful and diagnostic in some forms of muscular dystrophies. Specific enzyme histochemistry and biochemistry can be used for metabolic myopathies. Genetic analysis of the muscle specimen can be more informative than blood specimens for mitochondrial myopathies.

Imaging

Muscle computed tomography (CT) and magnetic resonance imaging (MRI) are of limited utility for evaluating muscle diseases but can be very useful in excluding spinal cord abnormalities that may cause weakness. Some inherited myopathies are associated with patterns of atrophy and the replacement of muscle with fat. In patchy or focal inflammatory myopathy, MRI can guide muscle biopsy. Functional MRI is sometimes useful in patients with suspected mitochondrial disorders.

SPECIFIC MUSCLE DISEASES

Inherited Muscle Diseases

The four main categories of inherited muscle diseases include muscular dystrophies (Table 421-5), congenital myopathies (Table 421-6), muscle ion-channel disorders (Table 421-7), and metabolic myopathies (Table 421-8). Some gene defects cause specific phenotypes that are instantly recognizable at the bedside by an experienced clinician, but less specific phenotypes can be caused by defects in more than one gene. A systematic approach is critical for an efficient and successful investigation of these disorders.

Muscular Dystrophies

The term *muscular dystrophy* refers to the primary degeneration of the muscle fiber, usually associated with an increase in fatty and fibrous connective tissue. The common clinical presentation is progressive muscle weakness.[1]

DYSTROPHINOPATHIES

Duchenne and Becker muscular dystrophies are caused by mutations in the dystrophin gene, which is located on the X chromosome. Female carriers can develop variable phenotypes, including a severe Duchenne-like presentation, mild adult-onset limb-girdle weakness, asymptomatic CK elevation, and cardiomyopathy.

Duchenne Muscular Dystrophy

Duchenne muscular dystrophy is the most common inherited muscle disease, with an incidence of about 1 in 5000 male births. About one third of patients carry a de novo mutation without a family history. In most patients, a frameshift mutation in the dystrophin gene results in a complete absence of the dystrophin protein. This absence of dystrophin disrupts the mechanical link between the sarcomere and the sarcolemma, thereby causing a calcium leak that leads to necrosis of muscle fibers.

CLINICAL MANIFESTATIONS

Duchenne dystrophy typically presents in young boys who are between 2 and 5 years of age with delayed motor milestones, difficulty running, increasing falls, and enlarged calves. The disorder is relentlessly progressive and can cause a cardiomyopathy (Chapter 60) that leads to heart failure and fatal arrhythmias. Intellectual disability, learning disorders, autism, and attention deficit hyperactivity disorder can be associated features. By 12 years of age, most affected individuals can no longer ambulate. By the age of 20 years, most patients develop joint contractures and kyphoscoliosis that lead to further respiratory compromise.

TABLE 421-5 MUSCULAR DYSTROPHIES

X-LINKED
Dystrophinopathies (Duchenne/Becker muscular dystrophy)
Emery-Dreifuss (Emerin)
FHL1-related (FHL1)
AUTOSOMAL DOMINANT
Facioscapulohumeral dystrophy (D4Z4 repeat deletions in 4q35 subtelomeric region in 95%; toxic gain of function in *DUX4* in 5%)
Myotonic dystrophy, type 1 (ctg repeats in *DMPK*)
Myotonic dystrophy, type 2 (cctg repeats in *ZNF9*)
Oculopharyngeal muscular dystrophy (*PABPN1*)
Myotilinopathy* (*MYOT*)
Laminopathy (*LMNA*)
Caveolinopathy (*CAV3*)
DNAJB6opathy* (*DNAJB6*)
Desminopathy* (*DES*)
Zaspopathy* (*ZASP*)
Bag3opathy* (*BAG3*)
Filaminopathy* (*FLNC*)
α-B-crystallinopathy* (*CRYAB*)
Titinopathy* (*TTN*)
VCPopathy† [valosin-containing protein] (*VCP*)
MYH7-myopathy†,‡ [Laing myopathy] (*MYH7*)
MYH2-myopathy† (*MYH2*)
Nesprinopathy (*SYNE1, SYNE2*)
TIA1opathy [Welander distal myopathy] (*TIA1*)
KLHL9opathy [kelch-like homologue-9] (*KLHL9*)
AUTOSOMAL RECESSIVE
Calpainopathy (*CAPN3*)
Dysferlinopathy (*DYSF*)
Anoctaminopathy (*ANO5*)
α-sarcoglycanopathy (*SGCA*)
β-sarcoglycanopathy (*SGCB*)
δ-sarcoglycanopathy (*SGCD*)
γ-sarcoglycanopathy (*SGCG*)
Telethoninopathy (*TCAP*)
TRIM32opathy (*TRIM32*)
α-Dystroglycanopathies§ (*FKRP, POMT1, POMT2, FKTN, POMGNT1, LARGE, DAG1, DPM2, DPM3*)
Titinopathy (*TTN*)
Laminopathy (*LMNA*)
Plectinopathy (*PLEC*)
Merosin deficiency§ (LAMA2)
Selenoproteinopathy‡,§ (*SEPN1*)
CollagenVIopathy§ (*COL6A1, COL6A2, COL6A3*)
With generalized lipodystrophy (*PTRF*)
Integrinα7opathy§ (*ITGA7*)
Integrinα9opathy§ (*ITGA9*)
With mitochondrial structural abnormalities§ (*CHKB*)
GNE-opathy† [glucosamine (UDP-*N*-acetyl)-2-epimerase/*N*-acetylmannosamine kinase] (*GNE*)
Matrin3opathy (*MATR3*)

*Associated with pathologic features of myofibrillar myopathy.
†Associated with pathologic features of inclusion body myopathy.
‡Can also cause congenital myopathy.
§Can also cause congenital muscular dystrophy.

TABLE 421-6 CONGENITAL MYOPATHIES

Central core, multi-minicore disease (*RYR1, SEPN1, MYH7, ACTA1, LMNA*)
Centronuclear myopathy (*MTM1, DNM2, RYR1, BIN1*)
Nemaline rod myopathy (*NEB, ACTA1, TPM3, TPM2, TNNT1, CFL2, KLHL40, KBTBD13*)
Congenital fiber-type disproportion (*ACTA1, SEPN1, RYR1, MYH7*)
Myosin storage myopathy (*MYH7*)
Other structural myopathies: Cap myopathy, zebra body myopathy, sarcotubular myopathy, spheroid body myopathy, fingerprint body myopathy, trilaminar myopathy, cylindrical spiral myopathy, myopathy with muscle spindle excess, myopathy with tubular aggregates.

DIAGNOSIS

Affected individuals have a 20- to 100-fold elevation of their serum CK level. Confirmation of diagnosis requires DNA analysis of the dystrophin gene. If genetic testing is negative, then a muscle biopsy is indicated. The pathologic features are typical of a chronic myopathy. Immunostaining shows an absence of dystrophin except in revertant fibers that express dystrophin (Fig. 421-3).

TREATMENT

Rx

Management requires a multidisciplinary team approach,[2] including physical therapy to prevent contractures and the timely provision of appropriate devices and wheelchairs. The cardiomyopathy is typically managed by using β-blockers and angiotensin-converting enzyme inhibitors (Chapter 59). An orthopedic surgeon should help with monitoring of scoliosis and spinal fusion if indicated. A pulmonologist should assess and follow respiratory function, including the initiation and monitoring of noninvasive ventilation. An endocrinologist can be helpful for managing osteoporosis and adrenal suppression during chronic steroid use. Ophthalmologic evaluation is needed to address cataract formation. Prednisone (0.75 mg/kg/day or weekend 10 mg/kg/day) prolongs the ability to ambulate despite its significant side effects (Chapter 35). Prednisone may also help respiratory function and slow the progression of scoliosis. Agents used for exon skipping and for read-through of premature stop codons are being tested clinical trials.[3]

TABLE 421-7 CHANNELOPATHIES AND RELATED DISORDERS

DISORDER	PATTERN OF CLINICAL FEATURES
Thomsen disease	Myotonia
Becker disease*	Myotonia and weakness
Paramyotonia congenita	Paramyotonia
Hyperkalemic periodic paralysis	Periodic paralysis, sometimes with myotonia and paramyotonia
Hypokalemic periodic paralysis	Periodic paralysis
Andersen-Tawil syndrome	Periodic paralysis, cardiac arrhythmias, skeletal abnormalities
Myotonia fluctuans	Myotonia
Myotonia permanens	Myotonia
Acetazolamide-responsive myotonia	Myotonia
Schwartz-Jampel syndrome* (Chondrodystrophic myotonia)	Pseudomyotonia, dysmorphism, blepharospasm
Rippling muscle disease	Muscle mounding/stiffness
Brody disease*	Delayed relaxation, no myotonia on electromyogram
Malignant hyperthermia	Anesthetic-induced excessive calcium release by sarcoplasmic reticulum

*Autosomal recessive; all other listed diseases are autosomal dominant.
Adapted with revision from Goldman L, Schafer AI, eds. *Cecil Medicine*. 24th ed. Philadelphia: Elsevier; 2012.

PROGNOSIS

With ventilatory support, patients often live well into the third or even fourth decades. Chronic respiratory failure is the primary cause of death in the late 20s or early 30s, with most patients succumbing to pneumonia, heart failure, or arrythmias.

Becker Muscular Dystrophy

Becker dystrophy is a milder form of Duchenne dystrophy caused by an in-frame mutation in the dystrophin gene.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Becker dystrophy can present in boys older than age 5 years, teenagers, or even in adults. Typical findings are symmetrical proximal weakness and prominent calf hypertrophy. Heart failure is common and may be the initial manifestation in some patients. The CK is elevated, although not to the same degree as seen in Duchenne dystrophy. Genetic testing for dystrophin gene is positive in about 90 to 95% of patients. If genetic testing is negative in a

TABLE 421-8 METABOLIC AND MITOCHONDRIAL MYOPATHIES

DISORDERS OF GLYCOGEN METABOLISM
Type II: α-1,4-Glucosidase (acid maltase) (*GAA*)
Type III: Debrancher (*AGL*)
Type IV: Branching (*GBE1*)
Type V: Myophosphorylase (*PYGM*)
Type VII: Phosphofructokinase (*PFKM*)
Type VIII: Phosphorylase b kinase (*PHBK*)
Type IX: Phosphoglycerate kinase (*PGK*)
Type X: Phosphoglycerate mutase (*PGAM-M*)
Type XI: Lactate dehydrogenase (*LDHA*)
Type XII: Aldolase A, (*ALDOA*)
Type XIII: β-Enolase (*ENO3*)
Type XIV: Phosphoglucomutase 1 (*PGM1*)
DISORDERS OF LIPID METABOLISM
Carnitine palmitoyltransferase II (*CPT2*)
Primary systemic carnitine deficiency (*SCL22A5*)
Secondary carnitine deficiency
Very long-chain acyl coenzyme A dehydrogenase deficiency (*ACADVL*)
Long-chain acyl coenzyme A dehydrogenase deficiency (*ACADL*)
Medium-chain acyl coenzyme A dehydrogenase deficiency (*ACADM*)
Short-chain acyl coenzyme A dehydrogenase deficiency (*ACAD5*)
Long chain hydroxyl/acyl coenzyme A dehydrogenase deficiency (*LCHAD*)
Multiple acyl coenzyme A dehydrogenase deficiency (*ETFA*)
Medications (valproic acid)
MITOCHONDRIAL MYOPATHIES
Chronic progressive external ophthalmoplegia
Kearns-Sayre syndrome
Mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes (MELAS)
Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE)
Myoclonic epilepsy with ragged red fibers (MERRF)
Infantile myopathy and lactic acidosis
Cytochrome c oxidase deficiency

Adapted with revision from Goldman L, Schafer AI, eds. *Cecil Medicine*. 24th ed. Philadelphia: Elsevier; 2012.

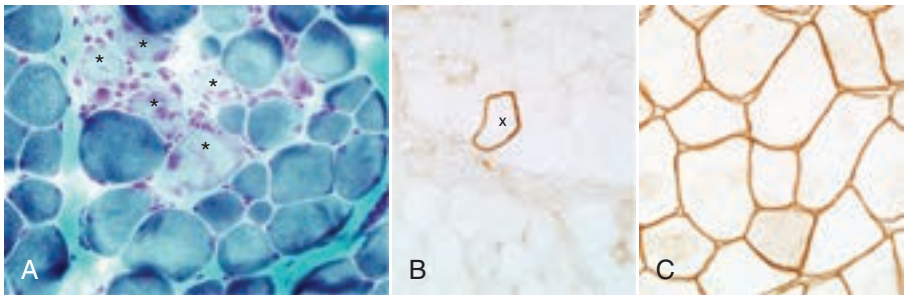

FIGURE 421-3. Duchenne muscular dystrophy. A, Trichromatically stained section of a specimen from a patient with Duchenne dystrophy displaying necrotic fibers (*) and increased endomysial and perimysial connective tissue. **B,** Dystrophin immunostain shows absent dystrophin reactivity in all fibers except for a revertant fiber (x). **C,** Normal sarcolemmal dystrophin reactivity in a control section.

patient suspected of having Becker dystrophy, a muscle biopsy is indicated. Muscle biopsy findings are similar to those of Duchenne dystrophy but less severe: immunohistochemistry shows decreased dystrophin expression, and immunoblotting reveals decreased expression and/or a lower molecular weight dystrophin protein.

TREATMENT AND PROGNOSIS

Management is largely supportive. Corticosteroids are rarely used. Similar to Duchenne dystrophy, screening for respiratory function and cardiac monitoring is indicated. Heart transplantation (Chapter 82) has been performed in patients with severe restrictive cardiomyopathy (Chapter 60).

Many patients have a normal lifespan, although some develop respiratory failure and have a shortened lifespan owing to respiratory complications. Heart failure and arrhythmias occur late in the course of disease.

Female Carriers of Duchenne or Becker Dystrophy

Female carriers of a dystrophin gene mutation are usually totally asymptomatic. However, about 2.5 to 10% of carriers can develop symptoms, including myalgias, proximal muscle weakness, and cardiomyopathy. Rarely, they present with a Duchenne phenotype owing to an XO karyotype (Turner syndrome) or skewed X-chromosome inactivation. If a specific mutation is identified in the family, targeted DNA analysis can confirm the diagnosis. Immunostaining of the muscle specimens shows a mosaic pattern, in which some fibers express dystrophin normally and others show decreased or even absent expression. Management of symptomatic carriers is similar to the management of Duchenne and Becker dystrophy patients with similar disease severities.

Facioscapulohumeral Muscular Dystrophy

Facioscapulohumeral muscular dystrophy is an autosomal dominant disorder with variable penetrance. It is the third most common dystrophy after the dystrophinopathies and myotonic muscular dystrophy, with a prevalence of about 1 in 15,000. About 95% of patients have a truncated D4Z4 tandem repeat region in chromosome 4q35.[4] The other 5% have hypomethylation of the D4Z4 region along with a mutation in the *SMCHD1* gene, which is critical for the structural maintenance of chromosome-flexible hinge-domain-containing protein 1.[5]

CLINICAL MANIFESTATIONS

Facioscapulohumeral muscular dystrophy has a highly variable penetrance within the same family. Severe proximal facial weakness can present in infancy, or mild and almost asymptomatic distal weakness can present in late adulthood. Some gene carriers never present with clinical symptoms or signs. The muscle weakness initially affects the face, where it causes difficulty smiling or whistling. Patients then develop scapular, humeral, truncal, and lower limb weakness leading to foot drop. Scapular winging is a typical feature. Muscle involvement is often asymmetrical.

Associated symptoms can include high-frequency hearing loss and retinal telangiectasia. Rare patients with retinal vascular abnormalities can develop retinal exudation leading to retinal detachment (Chapter 423). Infants with profound facial diplegia can also have intellectual disability and intractable epilepsy.

DIAGNOSIS

The CK level ranges from normal to mildly elevated. EMG shows typical myopathic features. Muscle biopsy shows chronic myopathic changes sometimes with an inflammatory exudate. Definite diagnosis is based on genetic testing.

TREATMENT AND PROGNOSIS

Management is supportive, and corticosteroids are of no benefit. Patients do benefit, however, from physical therapy for motion exercises for the shoulder girdle, molded ankle-foot orthoses for the footdrop, hearing aids for patients with hearing loss, and scapular fixation surgery to improve shoulder range of motion. Cardiac features are not prominent, and respiratory muscle weakness is a late feature. The prognosis is highly variable, depending on the severity and age of onset. Many patients have a normal lifespan.

Myotonic Dystrophies

Myotonic dystrophies, which are the second most common inherited muscle disease, affect about 1 in 8000 of the population. The two types, DM1 and DM2, are inherited in an autosomal dominant manner. Both cause multisystem disease and can be difficult to distinguish from each other. DM1 is caused by an abnormal expansion of CTG nucleotide repeats in an untranslated region of the dystrophia myotonica protein kinase (*DMPK*) gene on chromosome 19q. DM2 is caused by an abnormal expansion of CCTG nucleotide repeats in intron 1 of the zinc finger protein 9 (*ZNF9*) gene on chromosome 3q.

CLINICAL MANIFESTATIONS

Patients typically have frontal balding, ptosis, and temporal and masseter muscle wasting. Speech is nasal in quality, and patients display a high-steppage gait owing to their distal myopathy. On neurologic examination, myotonia is seen with percussion (inability to relax the muscle after percussion with a reflex hammer), after a grip (inability to relax the fingers after a firm grip), and in the eyelids (inability to open forcibly closed eyelids). Weakness in DM1 predominantly affects facial, oropharyngeal, forearm flexor, and foot dorsiflexor muscles. In DM2, weakness is predominantly proximal, although deep finger flexors are frequently affected. Muscle pain and stiffness are common in DM2 but can be seen in DM1 as well. Systemic features include premature subcapsular lens cataracts, testicular atrophy, intellectual disability, impotence, and hypersomnolence mediated by both central and neuromuscular mechanisms. Endocrine dysfunction is common, including diabetes mellitus and thyroid abnormalities. Dysphagia and constipation are common. Progressive cardiac conduction defects can lead to sudden death. Patients may require ventilation support after general anesthesia. Women who transmit DM1 are at high risk for having a child with a severe congenital form, including hypotonia at birth, respiratory failure, failure to thrive, and globally delayed developmental milestones with mild to severe intellectual disability.

DIAGNOSIS

Classical myotonic dystrophy usually can be diagnosed clinically by a patient's essentially pathognomonic facial features (see Fig. 421-2). The CK may be normal or mildly elevated. EMG, which is useful when the diagnosis is unsuspected or unclear, reveals myopathic features and myotonic discharges. Molecular genetic analysis of the nucleotide repeats confirms the diagnosis.[6]

TREATMENT

All patients should have an annual electrocardiogram and a Holter monitor to detect conduction system abnormalities (Chapter 62). An echocardiogram should be performed at diagnosis and repeated about every 2 to 4 years. Respiratory function testing also is usually recommended about every 2 to 4 years, and a sleep study is useful to detect nocturnal hypoventilation for symptomatic patients.

Mexilitene (150 to 200 mL three times daily) is well tolerated and can improve muscle relaxation. Hypersomnolence (Chapter 405) can be treated with overnight positive-pressure ventilation. Methylphenidate (200 mg daily) may be preferable to modafinil (300 mg daily) for excessive day-time sleepiness, but neither provides dramatic results.[7] Cardiac pacing, which is frequently required, can reduce the incidence of paroxysomal atrial fibrillation. Physical therapy can help to prevent contractures.

OTHER MUSCULAR DYSTROPHIES

Limb-girdle muscular dystrophies are a diverse group of myopathies caused by gene defects or deficiencies of muscle proteins that are critical for the normal function of the muscle cell membrane and especially the dystrophin-sarcoglycan complex.[8] Inheritance can be autosomal dominant or recessive. Although most patients have classical limb-girdle muscle weakness at the onset, some can present with distal leg muscle involvement that may initially be misdiagnosed as sensorimotor neuropathy. Some causes can be identified clinically by an experienced clinician. However, EMG can help to differentiate these conditions from neuropathies, and muscle biopsy, immunohistochemistry studies, and genetic analysis are often required to make a precise diagnosis.[9]

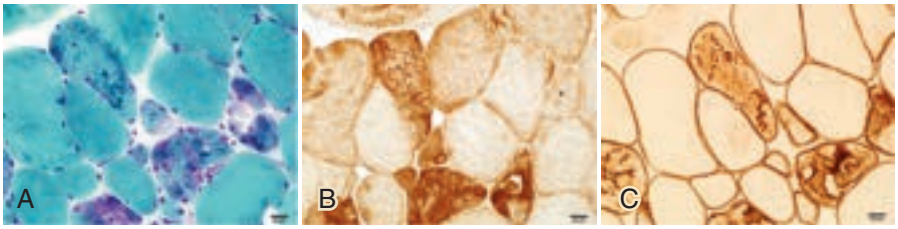

FIGURE 421-4. Myofibrillar myopathy. A, Trichromatically stained section of a specimen from a patient with myofibrillar myopathy displaying abnormal fibers with hyaline deposits or amorphous material and vacuoles. Abnormal fibers show abnormal and ectopic expression for desmin **(B)** and dystrophin **(C)**.

Myofibrillar myopathies present with progressive distal or proximal muscle weakness and characteristic morphologic features on muscle biopsy (Fig. 421-4). Cardiomyopathy and neuropathy can be associated features.[10]

Emery-Dreifuss muscular dystrophy[11] was originally X-linked and was initially shown to be caused by mutations in the emerin gene, a nuclear membrane protein. However, mutations in five other genes can cause a similar phenotype. Patients have a distinctive phenotype, including progressive joint contractures, scapuloperoneal distribution weakness, and cardiomyopathy with a progressive cardiac conduction disorder. The CK is often elevated but can be normal. Cardiac function needs to be monitored periodically. As with the dystrophinopathies, female carriers of the X-linked forms may develop weakness and cardiac disease.

Oculopharyngeal muscular dystrophy is an autosomal dominant myopathy caused by a trinucleotide repeat expansion of the poly(A)-binding protein, nuclear 1 (*PABPN1*) gene. Onset typically occurs in the fifth or sixth decade of life with dysphagia and marked ptosis. Marked distal and proximal weakness occurs later in the disease course. Surgical correction of ptosis often yields excellent results, but the dysphagia can be more difficult to manage. Patients often have a normal lifespan.

CONGENITAL MUSCULAR DYSTROPHIES

Congenital muscular dystrophies (see Table 421-5) are a rare group of autosomal recessive muscle diseases that present in infancy or childhood with hypotonia and muscle weakness.[12,13] The main differential diagnosis is spinal muscular atrophy (Chapter 419) and congenital myasthenia (Chapter 422). Affected infants can have joint contractures, which can be severe at birth. Some have a pure muscle phenotype and survive into adulthood. Others have severe central nervous system and eye involvement, which are associated with hypoglycosylation of α-dystroglycan that may be fatal in early childhood.

Congenital Myopathies

Congenital myopathies are rare inherited muscle diseases that are generally less severe than congenital muscular dystrophies. Patients present at birth with hypotonia, myopathic facies, and delayed motor milestones. The weakness is usually slowly progressive.[14] Respiratory muscle weakness can occur, and patients can present with ventilator failure at birth or insidiously in adult life. The most severely affected patients present in utero with reduced fetal movements and polyhydramnios. Muscle biopsy is diagnostic.

Central core myopathy (E-Fig. 421-1) is usually due to mutations in the ryanodine receptor gene (*RYR1*) and is strongly associated with malignant hyperthermia (Chapter 432). *Nemaline myopathies* are caused by different gene defects, including nebulin (*NEB*), skeletal α-actin (*ACTA1*), -tropomyosin$_{slow}$ (*TPM3*), β-tropomyosin (*TPM2*), troponin T$_{slow}$ (*TNNT1*), cofilin (*CFL2*), kelch-like family member 40 (*KLHL40*), and kelch-repeat- and BTB-[POZ]-domain-containing 13 (*KBTBD13*). Characteristic nemaline rods are seen on muscle biopsy (E-Fig. 421-2). *Centronuclear myopathy* can be X-linked owing to mutations in myotubularin (*MTM1*), dominant owing to mutations in dynamin 2 (*DNM2*), or recessive owing to mutations in *RYR1* or amphiphysin 2 (*BIN1*). Muscle biopsy is diagnostic (E-Fig. 421-3).

No specific treatments are available. Management requires a multidisciplinary team approach similar to muscular dystrophies.[15]

Ion Channelopathies

Ion channelopathies (see Table 421-7) are genetically determined disorders in which the muscle membrane functions abnormally. The combined prevalence of the various skeletal muscle channelopathies is about 1.1 per 100,000.[16] Each has a specific molecular cause, but the phenotypes overlap.

CHLORIDE CHANNELOPATHIES

Mutations in the muscle chloride-channel gene *CLCN1* cause autosomal dominant (Thomson) and autosomal recessive (Becker) myotonia congenita. Patients present with painless myotonia, muscle stiffness that may be slightly worse in the cold and improves with exercise (the warm-up phenomenon), muscle hypertrophy, and grip and percussion myotonia. EMG shows myotonic discharges. The myotonia responds to mexiletine (150 to 200 mg up to three times a day). Patients have a normal lifespan.

SODIUM CHANNELOPATHIES

Mutations in *SCN4A*, which is the voltage-gated sodium-channel gene, cause a range of autosomal dominant phenotypes, including hyperkalemic periodic paralysis, paramyotonia congenita, and potassium aggravated myotonia. Periodic paralysis is typically precipitated by sustained exercise that leads to weakness during the rest period, by potassium-rich food, or sometimes by emotional stress or cold. The attacks can persist for hours, during which the patient can be quadriplegic with depressed tendon reflexes but normal sensation, eye movements, and respiration. The serum potassium level may be high or normal during the attack. The physical examination is usually normal between attacks, but some patients develop fixed proximal weakness later in the disease. In patients with paramyotonia, the muscle stiffness paradoxically increases with exercise and is often painful. Cold sensitivity is typically more extreme than that seen in myotonia congenita, and cold can precipitate the muscle weakness. The diagnosis is made clinically, although patients with fixed proximal muscle weakness, which can be confused with other myopathies, have vacuolar changes and sometimes tubular aggregates on muscle biopsy. For both phenotypes, management involves avoidance of precipitating factors such as cold or strenuous exercise. If the myotonia is disabling, medications that act on sodium channels such as dichlorphenamide (50 to 450 mg daily), acetazolamide (250 mg twice daily), mexiletine (150 mg three times a day), phenytoin (300 to 600 mg daily), and carbamazepine (400 to 800 mg daily) can be considered. Life expectancy is normal.

CALCIUM CHANNELOPATHIES

Hypokalemic periodic paralysis is usually caused by autosomal dominant mutations in the voltage-dependent calcium channel gene *CACNA1S* but in about 10% of cases is caused by dominant mutations in *SCN4A*. The attacks of weakness, which are usually more severe and prolonged than in hyperkalemic period paralysis, generally persist for hours to days before gradually resolving. Attacks occur spontaneously or during prolonged rest after vigorous exercise and also can be precipitated by a high carbohydrate meal. The serum potassium level is reduced or low-normal during the attack. Avoidance of high carbohydrate loads and treatment with acetazolamide (125 to 1000 mg/day) or dichlorphenamide (50 to 400 mg/day) is effective. Patients may develop permanent muscle weakness if they have frequent attacks.

OTHER FORMS OF PERIODIC PARALYSIS AND MUSCLE STIFFNESS

Periodic paralysis can occur in a wide range of metabolic and electrolyte disorders (Table 421-9). Mutations in the *KCNJ2* gene, which encodes the inward rectifier potassium channel Kir2.1, cause *Andersen-Tawil syndrome*, an autosomal dominant, usually hypokalemic, periodic paralysis that is associated with distinctive facial features, including hypertelorism and low-set ears, as well as a propensity for cardiac arrhythmias. Treatment includes acetazolamide (250 mg twice daily) and dichlorphenamide. For patients with sulfa allergy, potassium-sparing diuretics such as spironolactone (25 to 100 mg/day) or triamterene (25 to 100 mg/day) can be used. The patients should be

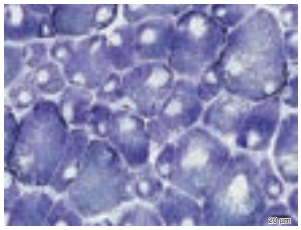

E-FIGURE 421-1. Core myopathy. NADH dehydrogenase reacted section displaying cores in almost all fibers in a patient with central core disease.

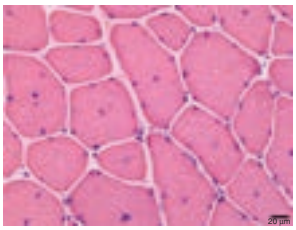

E-FIGURE 421-2. Centronuclear myopathy. Hematoxylin and eosin stained section displaying central nuclei in almost all fibers in a patient with centronuclear myopathy.

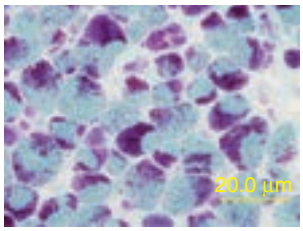

E-FIGURE 421-3. Nemaline rod myopathy. Trichromatically stained section showing myriad small nemaline rods in a patient with nemaline rod myopathy.

TABLE 421-9 SECONDARY CAUSES OF PERIODIC PARALYSIS

HYPOKALEMIC
Thyrotoxicosis
Primary hyperaldosteronism (Conn syndrome)
Renal tubular acidosis (e.g., Fanconi syndrome)
Juxtaglomerular apparatus hyperplasia (Bartter syndrome)
Gastrointestinal potassium wastage
Villous adenoma
Pancreatic non-insulin-secreting tumors with diarrhea
Nontropical sprue
Barium intoxication
Potassium-depleting diuretics
Amphotericin B
Licorice
Corticosteroids
Toluene toxicity
p-Aminosalicylic acid
Carbenoxolone
HYPERKALEMIC
Addison disease
Hypoaldosteronism
Excessive potassium supplementation
Potassium-sparing diuretics
Chronic renal failure

From Goldman L, Ausiello DA, eds. *Cecil Textbook of Medicine*. 23rd ed. Philadelphia: Elsevier; 2008.

treated as indicated for the cardiac arrhythmia and prolonged QT interval (Chapter 65). *Brody disease*, an autosomal recessive disorder caused by mutations in the SR calcium ATPase gene (*ATP2A1*), is characterized by exercise-induced muscle stiffness that is electrically silent on EMG. There are case reports of treatment with dantrolene, verapamil, or nifedipine with varying success.

Neuromyotonia (Isaac syndrome) is an autoimmune disorder associated with peripheral nerve hyperexcitability. It is caused by voltage-gated potassium-channel antibodies and is part of a spectrum of disorders, including limbic encephalitis. Oral immunomodulatory therapy, intravenous immunoglobulin, plasmapheresis, and symptomatic therapy with carbamazepine or phenytoin have been used with varying response. *Rippling muscle syndrome*, which can be caused by mutations in the *CAV3* gene, is characterized by rippling muscles triggered by exercise or percussion. Rippling muscle syndrome and neuromyotonia can be paraneoplastic phenomena, and a search for malignancy should be considered in these patients.

Metabolic Myopathies

Metabolic myopathies (see Table 421-8) are caused by enzyme defects that affect the three principal stages of muscle metabolism: (1) carbohydrate disorders due to a defect of glucose-glycogen metabolism; (2) disorders of fatty acid oxidation; and (3) disorders of mitochondrial oxidative phosphorylation. Muscle dysfunction can be acute, recurrent, and reversible, but the exercise intolerance can cause progressive weakness or even rhabdomyolysis (Chapter 113).

DISORDERS OF CARBOHYDRATE METABOLISM

Because glucose and glycogen are the primary energy sources for muscle contraction, any defects of glucose-glycogen metabolism cause muscle pain, cramps, contracture, and weakness within the first 30 minutes of exercise. The most common form is myophosphorylase deficiency (Chapter 207), and other forms are extremely rare. Most of these disorders are inherited autosomal recessively, although phosphoglycerate kinase deficiency is X-linked. Patients also note exercise intolerance and become deconditioned. Severe episodes are associated with very high CK levels, rhabdomyolysis, and myoglobinuria.

Phosphorylase Deficiency

Phosphorylase deficiency (McArdle disease, type IV glycogenosis) typically presents with muscle pain or cramps after short bursts of exercise. Some patients present with recurrent rhabdomyolysis. Persistent exercise beyond 30 minutes leads to the "second-wind" phenomenon, when fatty acids become the primary source of muscle energy. Clinical examination and the CK can be normal between the attacks, although some patients develop fixed proximal muscle weakness with myopathic features on EMG. Histochemical and enzyme analysis of muscle confirms the diagnosis. High protein diet and 37 g oral sucrose shortly before exercise and graded exercise may improve symptoms, but patients are at risk for developing contractures. The life expectancy is normal.

Acid Maltase Deficiency

Acid maltase deficiency (type II glycogenosis, α-1,4-glucosidase), also called Pompe disease, may present in infancy as a very severe generalized muscle disease that is fatal before age 2 years, as a juvenile variant that causes muscle weakness and death by the second or third decade owing to respiratory failure, or as an adult-onset form that presents with limb-girdle muscle weakness or sometimes with respiratory failure (Chapter 208). In each type, EMG reveals myotonic discharges. Abnormal glycogen storage and acid phosphatase–positive vacuoles are seen on muscle biopsy. The disorder can be diagnosed by measuring the enzyme activity in leukocytes or in muscle, but the use of dried blood samples to measure enzyme activity is increasingly becoming the standard practice. Mutation analysis is also clinically available. Enzyme replacement therapy appears promising in children and in the late-onset form.

DISORDERS OF FATTY ACID METABOLISM

After about 30 minutes of exercise, when the muscle glycogen reserves become exhausted, fatty acids become the principal source of muscle energy. Fatty acid metabolism involves the transport of fatty acids from the serum into the muscle and mitochondria, where both carnitine and carnitine palmitoyltransferase are key components of the β-oxidation pathway.

Fatty acid oxidation disorders can present with a proximal myopathy, exercise intolerance, muscle pain, rhabdomyolysis, and cardiomyopathy. Other features can include neuropathy, pigmentary retinopathy, recurrent hypoketotic hypoglycema, seizures, and intellectual disability. There may be a family history of sudden unexpected death syndrome. The most common fatty acid oxidation defect is *medium-chain acyl-CoA dehydrogenase (MCAD) deficiency*. *Carnitine palmitoyltransferase I deficiency* presents in childhood with an encephalopathy and liver failure associated with hypoglycemia and a high blood ammonia during metabolic crises. *Carnitine palmitoyltransferase II deficiency* can present as a fatal infantile onset form or more commonly between first to sixth decade of life with muscle pain, exercise intolerance, and myoglobinuria, typically after a long period of fasting or sustained exercise.

Carnitine deficiency can be primary or secondary. *Primary carnitine deficiency* causes myopathy, cardiomyopathy, and encephalopathy in association with hypoketotic hypoglycemia, although pure muscle presentations have been described. The diagnosis can be made by finding a low blood level, although prominent fat deposition in muscle is another clue. Other metabolic myopathies that can cause a secondary carnitine deficiency include disorders of β-oxidation and mitochondrial oxidative phosphorylation and may cause similar symptoms as the primary carnitine deficiency. Analysis of serum acylcarnitines, urine organic acids, and urine acylglycines and specific enzyme assays in fibroblasts can help to pinpoint the enzyme defect.

TREATMENT Rx

General treatment approach is avoidance of precipitating factors, such as prolonged fasting or prolonged exercise. Carbohydrate intake is advised before exercise, and patients should be prescribed a high carbohydrate, low fat diet with frequent feedings. Both primary and secondary carnitine deficiencies respond well to oral carnitine replacement (200 to 400 mg/kg/day in divided doses). Some patients have a multiple acyl-coenzyme A dehydrogenase deficiency (also called trifunctional enzyme deficiency, or glutaric aciduria type II), which responds well to riboflavin (100 mg daily).

DISORDERS OF MITOCHONDRIAL OXIDATIVE PHOSPHORYLATION

Disorders of mitochondrial oxidative phosphorylation (see Table 421-8), which are among the most common causes of inherited metabolic diseases, can present with isolated myopathy but often are multisystemic with cardiac involvement, diabetes mellitus, and both central and peripheral neurologic features.[17] Abnormal fatigability or exercise intolerance is a frequent complaint. Common CNS manifestations include epilepsy, migraine, strokelike episodes, myoclonus, ataxia, neuropathy, pigmentary retinopathy, dementia, and psychomotor regression.

Mitochondrial oxidative phosphorylation requires five respiratory chain complexes that are located on the inner mitochondrial membrane.[18]

Mitochondrial dysfunction results in energy deficits, which can lead to organ failure. Mitochondrial proteins can be coded by mitochondrial DNA (mtDNA), which is maternally inherited, and nuclear DNA, which can be inherited in an autosomal dominant, recessive, or X-linked manner. Phenotypic presentation of mtDNA defects depends on heteroplasmy, which is the amount and tissue distribution of the mutant mtDNA. If the amount of heteroplasmy exceeds a certain threshold, symptoms become apparent. Mitochondrial DNA disorders affect the structure or amount of the respiratory chain proteins, whereas nuclear DNA disorders can affect the proteins, the assembly of the respiratory chain, or the maintenance of mtDNA.

CLINICAL MANIFESTATIONS

Mitochondrial diseases should be considered in all patients who have a complex multisystemic myopathy, especially patients with neuromuscular, ocular, and endocrine involvement.

Mitochondrial encephalomyopathy with lactic acidosis and strokelike episodes (MELAS) is most frequently caused by a point mutation of mtDNA (m.3243A>G). Patients can have myopathy, cardiomyopathy, strokelike attacks, and encephalopathy. Some patients have one or only a few of these characteristics, some only have diabetes and deafness, and some only have cardiomyopathy.

Myoclonic epilepsy with ragged-red fibers (MERRF) is usually caused by a point mutation of mtDNA (m.8344A>G). It presents with a proximal myopathy associated with slowly progressive ataxia, epilepsy, peripheral neuropathy, and myoclonus.

Leber hereditary optic neuropathy (Chapter 424) predominantly affects young adult men, more than 95% of whom have mtDNA point mutations in m.3460G>A, m.11778G>A, or m.14484T>C. Patients develop subacute bilateral visual failure in both eyes within 2 to 3 months.

Chronic progressive external ophthalmoplegia with ptosis and gradual limitation of eye movements is seen in up to 20% of mitochondrial disorders. About 95% of patients have sporadic mtDNA point mutations or deletions, but the disease can be inherited as either an autosomal dominant or recessive trait. Mutations in *POLG* gene, which encodes the mitochondrial polymerase γ, are the most common causes of autosomal dominant or recessive progressive external ophthalmoplegia. *Kearns-Sayre syndrome* is characterized by the triad of external ophthalmoplegia, retinitis pigmentosa, and onset before the age of 20 years plus at least one of the following: heart block, cerebellar ataxia, or cerebrospinal fluid protein greater than 100 mg/dL. Kearns-Sayre syndrome is usually sporadic and caused by a single deletion of mtDNA.

Mitochondrial DNA depletion syndromes can present in neonatal period or infancy with subacute necrotizing encephalomyopathy (*Leigh syndrome*), hepatorenal failure, cardiomyopathy, and severe lactic acidosis. Children with *Pearson syndrome*, which is caused by accumulation of mtDNA deletions, typically present with pancytopenia, sideroblastic anemia, and exocrine pancreatic failure. *Primary coenzyme Q10 (ubiquinone) deficiency* is a rare autosomal recessive disorder that can present with encephalopathy, lipid storage myopathy, myoglobinuria, seizures, and cerebellar ataxia, or as an isolated nephrotic syndrome or an isolated myopathy.

DIAGNOSIS, TREATMENT, AND PROGNOSIS

The investigation of suspected mitochondrial disorders involves a systematic screen for multisystem complications, especially diabetes and cardiomyopathy; muscle biopsy to look for ragged red fibers, cytochrome c oxidase deficiency or biochemical evidence of respiratory chain dysfunction; search for mitochondrial deletion or depletion in muscle; and molecular genetic tests. Some primary mtDNA defects are not detectable in blood, so skeletal muscle is often required for the biochemical and genetic tests. For example, diagnosis of primary coenzyme Q10 (CoQ10) deficiency is made by measuring CoQ10 in muscle but not in blood.

Patients with primary CoQ10 deficiency can respond dramatically to CoQ10 supplementation (30 mg/kg/day in children and up to 2400 mg/day in adults in three divided doses). Vitamins and cofactors, including thiamine, riboflavin, and CoQ10, have shown varying degrees of benefit in different mitochondrial diseases. Management is largely supportive with monitoring and treatment of complications. Prognosis varies depending on the phenotype, ranging from the relatively normal life expectancy with chronic external ophthalmoplegia to a relatively rapid demise with Leigh syndrome.

OTHER METABOLIC AND TOXIC MYOPATHIES

Myopathy can complicate many metabolic disorders, including hypothyroidism (Chapter 226), Addison disease (Chapter 227), hyperaldosteronism (Chapter 227), hyperparathyroidism (Chapter 245), vitamin D deficiency (Chapter 244), and liver and renal failure (Chapters 130 and 153). The myopathy is often subtle, the CK level and EMG are often normal, and the muscle biopsy may be nonspecifically abnormal.

Many drugs cause myopathy (Table 421-10) with proximal muscle weakness, muscle pain, and exercise intolerance. The CK and EMG can be normal, and muscle biopsy findings may be nonspecific. The diagnosis may depend on the resolution of symptoms after the toxic agent is removed. Perhaps the most commonly incriminated medications are statins, which can cause muscle pain, an increased CK level, and rarely, myoglobinuria.

TABLE 421-10 TOXIC MYOPATHIES

INFLAMMATORY
Cimetidine
D-Penicillamine
Procainamide
L-Tryptophan
L-Dopa
NONINFLAMMATORY NECROTIZING OR VACUOLAR
Statins
Chloroquine
Colchicine
Emetine
ε-Aminocaproic acid
Labetalol
Cyclosporine
Tacrolimus
Isoretinoic acid (vitamin A analogue)
Vincristine
Alcohol
RHABDOMYOLYSIS AND MYOGLOBINURIA
Statins
Alcohol
Heroin
Amphetamine
Toluene
Cocaine
ε-Aminocaproic acid
Pentazocine
Phencyclidine
MALIGNANT HYPERTHERMIA
Halothane
Ethylene
Diethyl ether
Methoxyflurane
Ethyl chloride
Trichloroethylene
Gallamine
Succinylcholine
MITOCHONDRIAL
Zidovudine
MYOTONIA
2,4-D-Chlorophenoxylacetic acid
Anthracene-9-carboxylic acid
Cholesterol-lowering agents
Chloroquine
Cyclosporine
MYOSIN LOSS
Nondepolarizing neuromuscular blocking agents*
Intravenous glucocorticosteroids*

*In the setting of critical illness.
Adapted with revisions from Goldman L, Ausiello DA, eds. Cecil Textbook of Medicine, 23rd ed. Philadelphia: Elsevier; 2008.

INFLAMMATORY MUSCLE DISEASES

Inflammatory myopathies are a heterogeneous group of acquired muscle diseases (Table 421-11) that usually present with muscle weakness and exercise intolerance, with or without pain. Most patients have an elevated CK level and an abnormal EMG. Muscle biopsy shows an inflammatory infiltrate. However, the inflammatory process can be patchy and missed on the EMG

TABLE 421-11 CLASSIFICATION OF INFLAMMATORY MYOPATHIES

IDIOPATHIC
Polymyositis
Dermatomyositis
Inclusion body myositis
Overlap syndromes with other connective tissue disease (scleroderma, systemic lupus erythematosus, mixed connective tissue disease, Sjögren syndrome, rheumatoid arthritis, polyarteritis nodosa)
Sarcoidosis and other granulomatous myositis
Behçet disease
Inflammatory myopathies and eosinophilia
Eosinophilic polymyositis
Diffuse fasciitis with eosinophilia
Focal myositis
Myositis ossificans
INFECTIOUS
Bacterial: *Staphylococcus aureus,* streptococci, *Escherichia coli, Yersinia* sp., *Legionella* sp., gas gangrene *(Clostridium welchii),* leprous myositis, Lyme disease *(Borrelia burgdorferi)*
Viral: acute myositis after influenza or other viral infections (adenovirus, coxsackievirus, echovirus, parainfluenza virus, Epstein-Barr virus, arbovirus, cytomegalovirus), retrovirus-related myopathies (HIV, HTLV-1), hepatitis B and C
Parasitic: trichinosis *(Trichinella spiralis),* toxoplasmosis *(Toxoplasma gondii),* cysticercosis, sarcosporidiosis, trypanosomiasis *(Taenia solium)*
Fungal: *Candida* sp., *Cryptococcus* sp., sporotrichosis, actinomycosis, histoplasmosis

HIV = human immunodeficiency virus; HTML-1 = human T-lymphotrophic virus 1.
From Goldman L, Ausiello DA, eds. *Cecil Textbook of Medicine*. 23rd ed. Philadelphia: Elsevier; 2008.

or muscle biopsy, especially if the specimen is small or if a clinically unaffected muscle is biopsied. Similarly, a short period of corticosteroid therapy can mask the findings. MRI guidance can help to identify high-yield locations for muscle biopsy.

Systemic diseases associated with an inflammatory myopathy include polymyositis, dermatomyositis, inclusion body myositis (Chapter 269), systemic lupus erythematosus (Chapter 266), mixed connective tissue disease (Chapter 77), Sjögren syndrome (Chapter 268), rheumatoid arthritis (Chapter 264), and sarcoidosis (Chapter 95). Systemic viral illnesses and other infectious microorganisms (see Table 421-11) frequently cause muscle pain and an elevated CK, which rarely are major clinical problems.

Sarcopenia and Muscle Wasting

Muscle wasting is a common problem in elderly people (Chapter 25), partly related to hormonal changes and largely related to underuse. Critically ill patients rapidly lose muscle owing to inactivity and reduced protein synthesis,[19,20] with some potential to slow this process with protein-calorie nutrition (Chapter 111). Sarcopenia is also a prominent feature of many cancers, end-stage heart failure (Chapter 58) and renal failure (Chapter 131), and eating disorders (Chapter 219).

Grade A References

A1. Manzur AY, Kuntzer T, Pike M, et al. Glucocorticoid corticosteroids for Duchenne muscular dystrophy. *Cochrane Database Syst Rev*. 2008;1:CD003725.
A2. Logigian EL, Martens WB, Moxley RT IV, et al. Mexiletine is an effective antimyotonia treatment in myotonic dystrophy type 1. *Neurology*. 2010;74:1441-1448.
A3. Puymirat J, Bouchard JP, Mathieu J. Efficacy and tolerability of a 20-mg dose of methylphenidate for the treatment of daytime sleepiness in adult patients with myotonic dystrophy type 1: a 2-center, randomized, double-blind, placebo-controlled, 3-week crossover trial. *Clin Ther*. 2012;34:1103-1111.
A4. Russo V, Rago A, Politano L, et al. The effect of atrial preference pacing on paroxysmal atrial fibrillation incidence in myotonic dystrophy type 1 patients: a prospective, randomized, single-bind cross-over study. *Europace*. 2012;14:486-489.
A5. Statland JM, Bundy BN, Wang Y, et al. Mexiletine for symptoms and signs of myotonia in nondystrophic myotonia: a randomized controlled trial. *JAMA*. 2012;308:1357-1365.
A6. Sansone V, Meola G, Links TP, et al. Treatment for periodic paralysis. *Cochrane Database Syst Rev*. 2008;1:CD005045.
A7. Matthews E, Portaro S, Ke Q, et al. Acetazolamide efficacy in hypokalemic periodic paralysis and the predictive role of genotype. *Neurology*. 2011;77:1960-1964.
A8. Andersen ST, Haller RG, Vissing J. Effect of oral sucrose shortly before exercise on work capacity in McArdle disease. *Arch Neurol*. 2008;65:786-789.
A9. van der Ploeg AT, Clemens PR, Corzo D, et al. A randomized study of alglucosidase alfa in late-onset Pompe's disease. *N Engl J Med*. 2010;362:1396-1406.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Mercuri E, Muntoni F. Muscular dystrophies. *Lancet*. 2013;381:845-860.
2. Goemans N, Buyse G. Current treatment and management of dystrophinopathies. *Curr Treat Options Neurol*. 2014;16:287.
3. Jarmin S, Kymalainen H, Popplewell L, et al. New developments in the use of gene therapy to treat Duchenne muscular dystrophy. *Expert Opin Biol Ther*. 2014;14:209-230.
4. Statland JM, Tawil R. Facioscapulohumeral muscular dystrophy: molecular pathological advances and future directions. *Curr Opin Neurol*. 2011;24:423-428.
5. Lemmers RJ, Tawil R, Petek LM, et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet*. 2012;44:1370-1374.
6. Hilbert JE, Ashizawa T, Day JW, et al. Diagnostic odyssey of patients with myotonic dystrophy. *J Neurol*. 2013;260:2497-2504.
7. Laberge L, Gagnon C, Dauvilliers Y. Daytime sleepiness and myotonic dystrophy. *Curr Neurol Neurosci Rep*. 2013;13:340.
8. Wicklund MP. The muscular dystrophies. *Continuum (Minneap Minn)*. 2013;19:1535-1570.
9. Narayanaswami P, Weiss M, Selcen D, et al. Evidence-based guideline summary: diagnosis and treatment of limb-girdle and distal dystrophies: report of the guideline development subcommittee of the American Academy of Neurology and the practice issues review panel of the American Association of Neuromuscular & Electrodiagnostic Medicine. *Neurology*. 2014;83:1453-1463.
10. Selcen D. Myofibrillar myopathies. *Neuromuscul Disord*. 2011;21:161-171.
11. Puckelwartz M, McNally EM. Emery-Dreifuss muscular dystrophy. *Handb Clin Neurol*. 2011; 101:155-166.
12. Mercuri E, Muntoni F. The ever-expanding spectrum of congenital muscular dystrophies. *Ann Neurol*. 2012;72:9-17.
13. Bönnemann CG, Wang CH, Quijano-Roy S, et al. Diagnostic approach to the congenital muscular dystrophies. *Neuromuscul Disord*. 2014;24:289-311.
14. Colombo I, Scoto M, Manzur AY, et al. Congenital myopathies: natural history of a large pediatric cohort. *Neurology*. 2015;84:28-35.
15. Wang CH, Dowling JJ, North K, et al. Consensus statement on standard of care for congenital myopathies. *J Child Neurol*. 2012;27:363-382.
16. Horga A, Raja Rayan DL, Matthews E, et al. Prevalence study of genetically defined skeletal muscle channelopathies in England. *Neurology*. 2013;80:1472-1475.
17. Schapira AH. Mitochondrial diseases. *Lancet*. 2012;379:1825-1834.
18. Archer SL. Mitochondrial dynamics: mitochondrial fission and fusion in human diseases. *N Engl J Med*. 2013;369:2236-2251.
19. Puthucheary ZA, Rawal J, McPhail M, et al. Acute skeletal muscle wasting in critical illness. *JAMA*. 2013;310:1591-1600.
20. Fan E, Cheek F, Chlan L, et al. An official American Thoracic Society Clinical Practice guideline: the diagnosis of intensive care unit-acquired weakness in adults. *Am J Respir Crit Care Med*. 2014;190:1437-1446.

422

DISORDERS OF NEUROMUSCULAR TRANSMISSION

AMELIA EVOLI AND ANGELA VINCENT

DEFINITION

Neuromuscular transmission depends on the release of acetylcholine from synaptic vesicles that are stored in the terminal boutons of the motor nerve axon (Fig. 422-1). Invasion of the motor nerve terminal by the action potential opens voltage-gated calcium channels, resulting in the Ca^{2+}-dependent release of the vesicular contents into the synaptic space. Acetylcholine binds to the acetylcholine-gated ion channels (acetylcholine receptors [AChRs]) on the postsynaptic membrane, thereby leading to the opening of these channels and a local depolarization, the end- plate potential. If the end-plate potential exceeds the critical firing threshold, voltage-gated sodium channels (sited at the bottom of the postsynaptic folds) open to generate the muscle action potential that propagates along the muscle fiber and activates muscle contraction. The action of acetylcholine is terminated by its dissociation from the AChRs, which close spontaneously after 1 to 4 milliseconds; hydrolysis of acetylcholine by acetylcholinesterase; and acetylcholine diffusion from the synaptic cleft. Meanwhile, in the motor nerve terminal, the voltage-gated calcium channels close spontaneously, and the resting membrane potential is restored through the transient opening of voltage-gated potassium channels.

The extent to which the amplitude of the end-plate potential exceeds the threshold for activation of the voltage-gated sodium channels is called the safety factor. In healthy individuals, the amplitude decreases during repeated activity but does not fall below this threshold; thus, neuromuscular transmission is not compromised. However, if there is an abnormally low end-plate potential amplitude, failure of neuromuscular transmission may occur. Causes include defects in the release of acetylcholine, the postsynaptic response to acetylcholine, or the number or sensitivity of the voltage-gated sodium channels. Morphologic changes to the presynaptic or postsynaptic components or to the basal lamina between them may also influence the efficacy of transmission. Although myasthenia gravis and some neurotoxic envenomations (Chapter 112) are the most common disorders of neuromuscular transmission, a number of conditions have been implicated (Table 422-1).

AUTOIMMUNE DISEASES

Myasthenia Gravis

EPIDEMIOLOGY

Myasthenia is the most common disorder of neuromuscular transmission, with a prevalence of about 15 per 100,000 in Western countries.[1] All races can be affected, and it can occur at any age from year 1 onward. There is a small peak in the incidence rate in women in the third decade and a larger peak, the majority males, at later ages. The annual incidence rises to about 5 per 100,000 after age 70 years. It is especially important to differentiate myasthenia gravis from other causes of limb or bulbar muscle weakness in elderly people.

Myasthenia gravis itself is heterogeneous and can be divided into different subtypes; the relative frequency of these different forms is not known, but relatively mild childhood forms are frequent in Asian countries. Neonatal myasthenia gravis, due to the placental transfer of maternal antibodies to the AChR or to muscle-specific kinase (MuSK), affects up to one in eight babies born to mothers with myasthenia gravis. Autoimmune myasthenia gravis must be distinguished from congenital myasthenic syndromes, which are caused by gene mutations.

PATHOBIOLOGY

Pathophysiology

Myasthenia gravis is the result of a defect in neuromuscular transmission. The postsynaptic response to acetylcholine, the end-plate potential, is reduced so

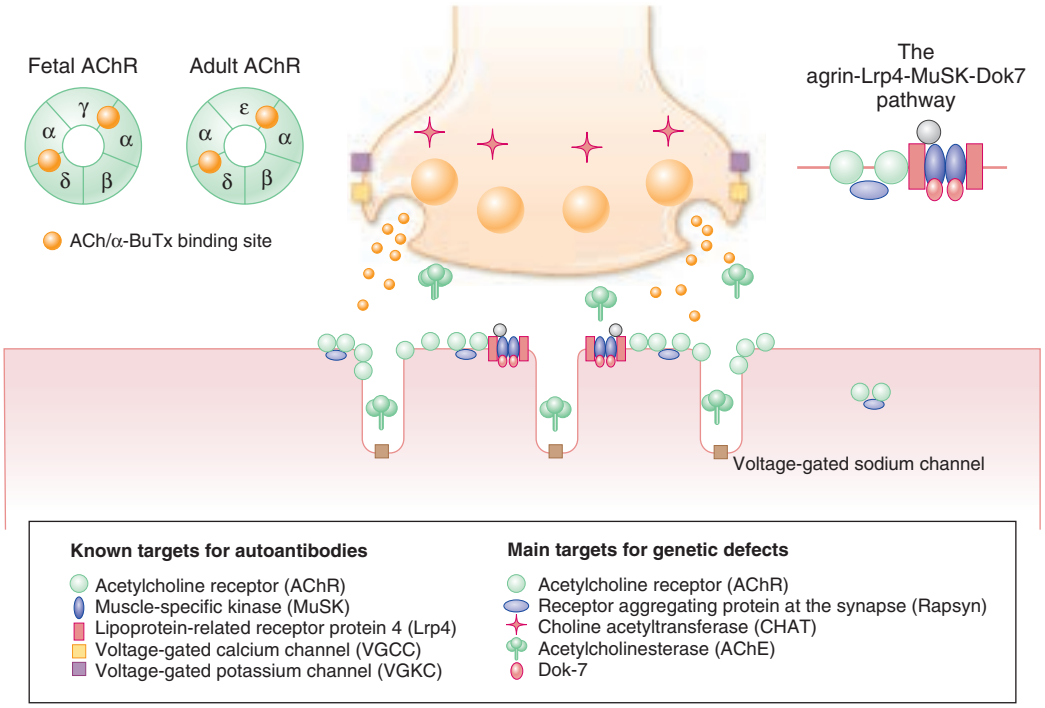

FIGURE 422-1. Diagrammatic representation of the neuromuscular junction, indicating the ion channels, receptors, enzymes, and associated proteins that are the most frequent targets for autoimmune diseases *(left)* or mutations in genetic diseases *(right)*. The acetylcholine receptor exists in fetal and adult isoforms as illustrated at *top left*. The replacement of the fetal form by the adult form takes place toward the end of gestation in humans. The agrin-Lrp4-MuSK-DOK7 pathway associated with AChRs on the postsynaptic membrane is illustrated at *top right*. α-BuTx, α-bungarotoxin, the snake toxin that binds with high specificity to the two Ach binding sites on the AChRs.

TABLE 422-1 DISORDERS OF NEUROMUSCULAR TRANSMISSION

DISEASE	TARGET	PATHOBIOLOGY
AUTOIMMUNE		
Myasthenia gravis	AChRs	Antibodies to AChR in 85% reduce AChR numbers and EPP amplitude
	MuSK	Antibodies to MuSK in 5-10%; mechanism not clear
	Lrp4	Antibodies to Lrp4 in a proportion of patients; mechanism not clear
Transient neonatal myasthenia	AChRs, MuSK	Maternal antibodies cause transient disease in neonate; not seen commonly if mother receiving treatment
Arthrogryposis	Fetal AChR	Maternal antibodies that inhibit fetal AChR function resulting in paralyses in the fetus in utero, leading to joint contractures and rarely arthrogryposis
Lambert-Eaton myasthenic syndrome	VGCCs	Antibodies to VGCC in 90% reduce VGCC numbers and ACh release and EPP amplitude
Acquired neuromyotonia	VGKCs	Antibodies to VGKC-complex in 40% lead to increased and spontaneous ACh release
GENETIC		
Acetylcholine receptor deficiency	AChR	Recessive mutations in AChR-subunit genes cause reduced AChR expression
Acetylcholine receptor deficiency		Recessive mutations in *RAPSYN* cause reduced anchoring of AChR on the postsynaptic membrane, or in *DOK-7* cause a synaptopathy
AChR kinetic abnormalities	AChR	Dominant or recessive mutations in AChR-subunit genes cause kinetic defects—"slow" and "fast" channel syndromes
Choline acetyltransferase deficiency	Choline acetyltransferase	Recessive mutations in the gene for choline acetyltransferase (*CHAT*) cause reduced ACh release
Acetylcholine esterase deficiency	AChE	Recessive mutations in the collagen tail (*COLQ*) that anchors AChE at the neuromuscular junction cause absence of AChE
Arthrogryposis, multiple pterygium, Escobar's syndrome	Can occur with rapsyn, δ- or γ-subunit AChR mutations	Fetal akinesia
NEUROTOXIC		
Botulism	Presynaptic ACh release	Botulinum toxin gains entry into the presynaptic motor nerve and cleaves proteins involved in ACh release mechanism
Envenomation following bites from snakes, spiders, scorpions, etc.	Varied sites of action	Neurotoxins specific for VGCCs, VGKCs, AChE, AChRs, voltage-gated sodium channels, and other targets are frequent in many animal venoms and generally inhibit function
Drugs and insecticides	Varied sites of action	Muscle relaxants and other drugs
		Many antibiotics and quinine-related drugs can alter neuromuscular transmission at high dose
		Organophosphates block AChE and have complicated acute and chronic actions

AChE = acetylcholinesterase; AChR = acetylcholine receptor; EPP = end-plate potential; MuSK = muscle-specific kinase; Lrp4 = low-density lipoprotein-related receptor protein 4; VGCC = voltage-gated calcium channel; VGKC-complex = voltage-gated potassium channel and associated proteins.

that the threshold for activation of the muscle action potential is not reached. At a severely affected end plate, this deficiency can occur at the initiation of contraction, but it is most common during repetitive activity when the end-plate potential naturally declines, despite a compensatory rise in the release of acetylcholine. This phenomenon, occurring across many end plates within a muscle, is responsible for the decrement in the amplitude of the compound muscle action potential on repetitive nerve stimulation, a finding that is diagnostic of a disorder of neuromuscular transmission.

In myasthenia gravis, the reduced end-plate potentials result from loss of functional AChRs on the postsynaptic membrane and also from simplification of the postsynaptic folds, which contain the voltage-gated sodium channels. In most patients, these changes are caused by antibodies against the

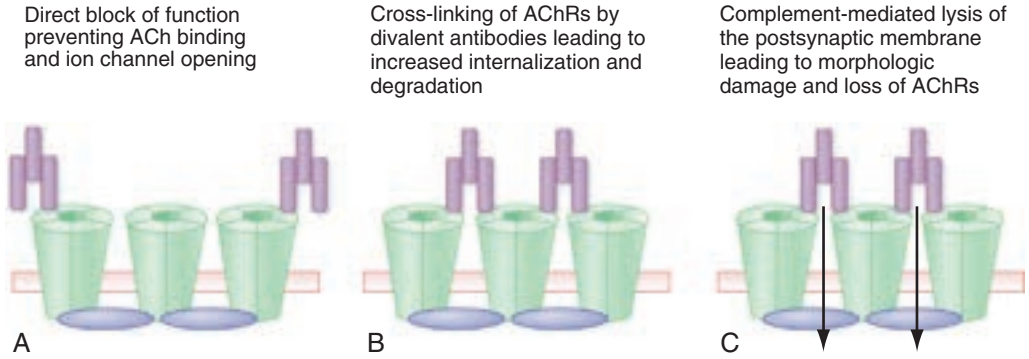

FIGURE 422-2. Mechanisms of loss of the acetylcholine receptor (AChR) at the neuromuscular junction. Antibodies can act (**A**) by directly blocking ACh binding or ion channel function; (**B**) by cross-linking the AChRs in the membrane, thereby leading to increased internalization and degradation; or (**C**) by complement-dependent lysis of the AChR-containing postsynaptic membrane. In myasthenia gravis, complement-dependent lysis is likely to be the most important mechanism overall. Interestingly, there is no evidence of complement-dependent mechanisms in either Lambert-Eaton myasthenic syndrome or acquired neuromyotonia, in which cross-linking of the respective ion channels with increased internalization seems to be the main mechanism.

AChRs. The pathophysiology in patients with antibodies to other postsynaptic proteins, such as MuSK and the low-density lipoprotein-related receptor protein 4 (Lrp4), involves different mechanisms that are not yet fully understood. Like most synapses, the neuromuscular junction is highly regulated. If the nerve is cut, leading to loss of neuromuscular transmission, the muscle responds by upregulating the expression of AChRs that revert to a fetal phenotype (see Fig. 422-1). Alternatively, if the activity of the postsynaptic muscle decreases, the motor nerve attempts to compensate. Consequently, the synthesis of AChRs in the muscle fiber and the release of acetylcholine from the motor nerve are increased in myasthenia gravis.

Pathogenesis

Myasthenia gravis is an antibody-mediated disease that is associated with other autoimmune disorders, especially thyroid disease (Chapter 226). Younger AChR antibody-positive patients have an increased prevalence of the human leukocyte antigen (HLA)-B8 and -DR3 haplotypes that are also frequently associated with autoimmunity. AChR antibodies are immunoglobulin G (IgG), have high affinity, are highly specific for the native human AChR, and act by three main mechanisms (Fig. 422-2). First, a few antibodies directly inhibit the binding of acetylcholine to the AChR, thereby causing a pharmacologic-like blockade of function.[2] Second, because of their divalence, antibodies can bind simultaneously to two adjacent AChRs, through the α-subunits that are present in duplicate in each receptor, to form AChR-antibody complexes that are internalized and degraded by the muscle fiber, thereby leading to loss of AChRs. Third, most of the antibodies are IgG1 subclass, which binds and activates complement. The result is activation of the membrane attack complex with destruction of the postsynaptic membrane and morphologic damage. All these effects are strictly limited to the neuromuscular junction; the remainder of the muscle fiber is essentially normal.

Specific antibody production requires helper T cells that can recognize AChR epitopes. The thymus gland, which is often abnormal in myasthenia, is thought to play a role in the immune response against AChR.[3] In patients with early-onset disease, the thymus is often the site of follicular hyperplasia, with T- and B-cell lymphocytic infiltrates in an expanded medulla. These infiltrates, which are very similar to the germinal centers found in lymph nodes, contain B cells that express surface immunoglobulin specific for AChRs and plasma cells that synthesize AChR antibodies. In the thymic medulla, muscle-like "myoid" cells have AChRs on their surface in both normal and myasthenic individuals; these cells may be an early target of complement and antibodies, thereby providing the antigenic stimulus responsible for chronic germinal center formation and AChR antibody production.

In late-onset myasthenia gravis and in patients with MuSK antibodies, the thymus is mostly normal for age. However, some patients without conventional AChR or MuSK antibodies have typical thymic hyperplasia and antibodies that bind to tightly clustered AChR on transfected cells.

Thymomas, which are epithelial cell tumors, occur in 10 to 15% of myasthenic patients and nearly always are associated with AChR antibodies. Thymomas associated with myasthenia gravis correspond mainly to the World Health Organization types B1 and B2 and are characterized by active thymopoiesis (i.e., the capacity to promote T-cell maturation and export). Thymoma epithelial cells express muscle antigens and AChR subunits, and they are thought to be responsible for defective negative selection, with the export to the periphery of autoreactive T lymphocytes. Rarely, myasthenia gravis arises after removal of a thymoma.

About 40% of AChR antibody-negative patients have antibodies to MuSK, a muscle tyrosine kinase that is specifically expressed at the neuromuscular junction and plays a crucial role in the formation and maintenance of the postsynaptic membrane.[4] MuSK is activated by nerve-secreted agrin, through its coreceptor Lrp4; MuSK phosphorylation and dimerization induce an intracellular signaling cascade that leads to AChR clustering. In animal models, MuSK antibodies cause loss of AChR and reduced postsynaptic folds, as well as a lack of the normal compensatory presynaptic increase in the release of acetylcholine. However, MuSK antibodies are predominantly IgG4 and do not activate complement, so it is not yet clear how these antibodies cause the neuromuscular pathophysiology.

A small number of AChR and MuSK antibody-negative patients have serum antibodies to Lrp4.[5] Lrp4 antibodies interfere with Lrp4-agrin binding and reduce AChR expression in vitro, and their pathogenicity has been recently demonstrated in an animal model.

CLINICAL MANIFESTATIONS

Myasthenia gravis presents clinically with painless muscle weakness that increases with muscle use and improves after rest. In many patients, the weakness starts in the eye muscles, where it results in double vision and ptosis (drooping eyelids). In others, it may first affect bulbar muscles or limb muscles (Fig. 422-3). Virtually any skeletal muscle may be involved as the illness progresses. Typically, the weakness varies in distribution and severity from day to day or from week to week, and it is often worse in the evening. It may first appear following an infection. Established weakness can increase with anxiety, with infection, or with the menstrual period.

Ptosis, which is often asymmetrical, and diplopia initially can be transient and first noticed while driving, for example. Severity can range from mild unilateral ptosis or minimal diplopia to profound bilateral ptosis combined with almost complete ophthalmoplegia. Bulbar symptoms include weakness of facial muscles with difficulties in closing eyes and a "snarling" smile, difficulty in chewing, nasal or slurred speech that can noticeably deteriorate as speech continues, impaired swallowing sometimes associated with nasal regurgitation of fluids, reduced tongue movements, and head droop related to neck weakness.

Limb muscle involvement is common, and proximal muscles are usually more involved than distal. Weakness of the legs can lead to collapse when walking and can be misinterpreted as a functional (psychogenic) disorder. Weakness of elbow extension and of finger abduction may be prominent. By contrast, ankle dorsiflexion is rarely affected except in severe disease. Respiratory dysfunction is less common but can be life-threatening, especially if associated with dysphagia. Selective involvement of the diaphragm can cause severe breathlessness in the supine posture. Wasting is uncommon but can affect the facial muscles and tongue, for example, in long-standing disease. Tendon reflexes are typically brisk. Bladder disturbances are rare, and sensory symptoms do not occur.

Subtypes of Myasthenia Gravis

Several subgroups can be distinguished on the basis of clinical and pathologic criteria and can help to inform treatment.

Ocular Myasthenia Gravis

Ocular myasthenia gravis is confined to extraocular muscles; if it remains localized for at least 2 years, subsequent generalization is unlikely. AChR antibody levels are generally low and are undetectable in about 50% of patients. This subgroup rarely is associated with a thymoma. The neuromuscular junction of ocular muscles shows structural and physiologic differences from limb muscles. Ocular weakness is often the presenting symptom not only in myasthenia gravis but also in neurotoxin poisoning, for example, botulism (Chapter 296). Thus, physiologic factors or accessibility of the neuromuscular junctions of ocular muscles to circulating factors may make them particularly vulnerable to antibodies in myasthenia gravis.

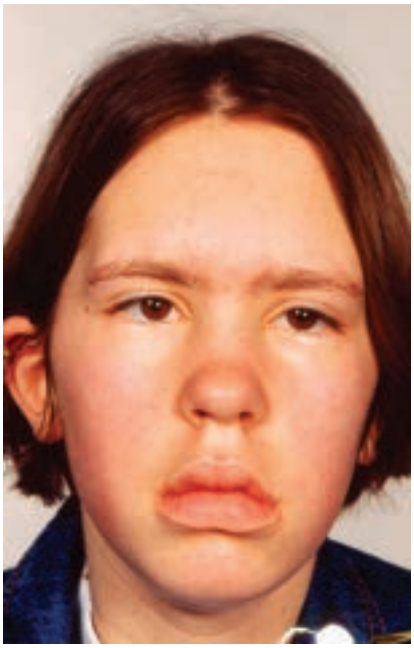

FIGURE 422-3. Marked ocular and facial muscle weakness in a young female with myasthenia gravis.

Generalized Myasthenia Gravis with Acetylcholine Receptor Antibodies

Among patients with generalized disease and AChR antibodies, there are three clinical subgroups. Early-onset myasthenia gravis is more frequent in females and associates strongly with HLA-A1, -B8, and -DR3. The thymus is generally hyperplastic. AChR antibody titers are usually high and decline to varying degrees after successful treatments, including thymectomy.

Late-onset myasthenia gravis is becoming increasingly common with the aging of the population and, when associated with bulbar weakness, may be mistaken for amyotrophic lateral sclerosis (Chapter 419) or brain stem cerebrovascular disease. Among older patients, males are more frequently affected.

Thymoma-associated myasthenia gravis is an important distinction because thymectomy or other specific tumor therapy is required. Most patients with thymomas and myasthenia gravis present between the ages of 30 and 60 years.

Myasthenia Gravis with Muscle-Specific Kinase Antibodies

About 15% of myasthenic patients with generalized symptoms do not have detectable AChR antibodies. Up to 40% of these patients have antibodies to MuSK.[6] MuSK antibodies are absent or very infrequent in patients with AChR antibodies, patients with persistent ocular symptoms, and patients with thymoma. Compared with typical myasthenia gravis, MuSK antibody-positive disease is characterized by high prevalence in younger females; predominant bulbar, neck, and respiratory muscle weakness; and an increased rate of facial and tongue muscle atrophy.

Myasthenia Gravis with Neither Acetylcholine Receptor nor Muscle-Specific Kinase Antibodies

Some patients with thymus hyperplasia and good response to treatment, including thymectomy, may have antibodies that bind only to clustered AChRs on AChR-expressing cells. A variable proportion of AChR and MuSK antibody-negative patients have Lrp4 antibodies. These antibodies have not been found associated with thymoma.

DIAGNOSIS

Diagnosis is based on the clinical features, serologic testing for specific antibodies, electromyography (EMG), and, if doubt still remains or specialized facilities are not available, the clinical response to anticholinesterase medication (Table 422-2).[7] Mediastinal imaging is needed to exclude a thymoma especially in patients with AChR antibodies.

TABLE 422-2 DIAGNOSTIC EVALUATION (EXCLUDES NEUROMYOTONIA)

	AChR MG	MuSK MG	Lrp4-MG SN-MG	NEONATAL MG	LEMS	CMS	BoTx	MM
Onset birth, recovery of muscle strength within 2 mo	–	–	–	+	–	AChR γ-subunit mutations, variable severity	–	–
Onset birth plus arthrogryposis	–	–	–	+	–	Rapsyn or AChR δ-subunit mutations	–	–
Onset at <1 yr and persistent	–	–	–	–	–	Any CMS Dok-7, rapsyn deficiency, and SCS may present later	+	+/–
Infantile apneas	–	–	–	+/–	–	Fast channel syndrome, rapsyn, or ChAT mutation		
AChR Ab positive	+	–	–	+/–	–	–	–	–
MuSK Ab positive	–	+	–	+/–	–	–	–	–
VGCC Ab positive	–	–	–	–	+	–	–	–
EMG decrement >10%	+	+/–	+/–	+	+	+	+/–	–
EMG jitter increased	+	+ Especially face muscles	+	+	+	+	+	+/–
Post-tetanic potentiation	–	–	–	–	+	–	+	–
AChE inhibitor response	+	Poor or deterioration	+	+	Often weak	Except *SCS, COLQ,* or *DOK7* mutations	+/–	–
Thymoma	+/–	–	–	–	–	–	–	–

AChE = acetylcholinesterase; AChR = acetylcholine receptor; BoTx = botulism; ChAT = choline acetyltransferase; CMS = congenital myasthenic syndromes; Dok-7 = downstream of kinase 7; LEMS = Lambert-Eaton myasthenic syndrome; Lrp4 = low-density lipoprotein-related receptor protein 4; MG = myasthenia gravis; MM = mitochondrial myopathy; MuSK = muscle-specific kinase; SCS = slow-channel syndrome; SN = seronegative for AChR and MuSK antibodies.

If AChR antibodies are absent, especially in patients with generalized symptoms, testing for MuSK antibodies is recommended. Both AChR and MuSK antibodies are very specific, and their detection in symptomatic patients confirms the diagnosis. It is not yet clear whether testing for the rarer Lrp4 or clustered AChR antibodies, which is available in specialist centers, will prove helpful in patients with inconclusive clinical and EMG findings.

The electrophysiologic abnormality is an abnormally large decrement (>10%) in the amplitude of the compound muscle action potential on low-rate (3-Hz) repetitive nerve stimulation or increased jitter on single-fiber EMG. In patients with MuSK antibodies, EMG abnormalities may be detectable only in facial muscles. These EMG changes are not specific for myasthenia gravis but can occur in any disorder that interferes with neuromuscular transmission.

Intravenous administration of edrophonium (Tensilon), a short-acting cholinesterase inhibitor, transiently improves myasthenic weakness but requires an appropriate medical setting, including resuscitative facilities and the availability of atropine, because of the risk for adverse events and severe cholinergic reactions, including syncope. A test dose of 2 mg is given intravenously, followed 30 seconds later by 6 to 8 mg if no adverse event has occurred. The equivalent doses in children are a 20-µg/kg test dose followed by 60 to 80 µg/kg. Some patients improve sufficiently with the test dose, so it is not necessary to give the full dose. An alternative pharmacologic test in adults is a single dose of subcutaneous or intramuscular neostigmine (1 to 2.5 mg) or of oral pyridostigmine (60 mg).

Differential Diagnosis

Congenital myasthenic syndromes (see later) should be considered in patients who have clinical and EMG evidence of myasthenia but are seronegative on antibody assays. Lambert-Eaton myasthenic syndrome almost always begins with difficulty in walking; ocular symptoms are rare, and specific laboratory tests are available (see later). The ocular muscle involvement that characterizes Miller-Fisher syndrome is more rapid in onset than is usual in myasthenia gravis and is associated with GQ1b antibodies (Chapter 420). Mitochondrial myopathy may show signs that are similar to those of myasthenia gravis (e.g., asymmetrical ptosis and limitation of eye movements), and there may be increased jitter on single-fiber EMG, but this condition and oculopharyngeal dystrophy can be distinguished from myasthenia gravis by the nonfluctuating weakness and by muscle biopsy (Chapter 421). In neurasthenia and chronic fatigue syndrome (Chapter 274), the laboratory tests for myasthenia gravis are negative.

TREATMENT

Rx

Most patients with AChR antibodies respond to oral pyridostigmine, 30 to 60 mg four or five times daily; in patients with mild disease, this dose may adequately control symptoms.[8] Doses in excess of 90 mg are likely to cause gastrointestinal side effects, abdominal cramps and diarrhea, which can be controlled with oral propantheline bromide, 15 mg, or loperamide, 2 mg. Patients with MuSK antibodies generally have an unsatisfactory response.[9] In some of these patients, pyridostigmine, even at low doses, can increase weakness and cause nicotinic side effects (muscle cramps and diffuse fasciculations).

Neonatal Myasthenia Gravis

Pyridostigmine, 3 to 5 mg, can be given every 4 hours to about an hour before a feeding. Close monitoring and respiratory support in a special unit may be required.

Ocular Myasthenia

Diplopia can sometimes be helped by the use of prisms. Ocular symptoms that respond incompletely to pyridostigmine are generally improved by low-dose prednisone therapy (e.g., 5 mg every other day), increasing by 5 mg at weekly intervals either until symptoms are completely controlled or until a ceiling dose (e.g., 1 mg/kg) is reached.[10] When remission is established, the dose can be slowly reduced (e.g., by 5 mg at 2-weekly intervals) until symptoms recur and then adjusted upward to define the effective minimal dose. Full withdrawal of prednisone is usually followed by a symptomatic relapse. Thymectomy is not considered beneficial for nonthymomatous ocular myasthenia gravis. In patients who fail to respond adequately to prednisone or who are intolerant of the medication, the addition of azathioprine (2 to 2.5 mg/kg body weight) or ocular muscle surgery is an option. However, the diagnosis should be questioned in patients who show no improvement with high-dose prednisone treatment.

Thymoma

Thymoma is usually an indication for surgery, but removal of the tumor seldom improves muscle weakness. If the tumor is locally invasive, postoperative radiotherapy is indicated. If tumor spread is more extensive, chemotherapy with cisplatin-containing regimens can be considered. Thymoma-associated myasthenia gravis is generally severe, and most patients need long-term treatment with steroids and immunosuppressants (see below).[11]

Generalized Nonthymomatous Myasthenia Gravis

When generalized symptoms are inadequately controlled by pyridostigmine, thymectomy is often recommended even for patients without a thymoma, especially patients younger than age 50 years. On the basis of uncontrolled studies, thymectomy in early-onset patients with AChR antibodies appears to be associated with an increased rate of remission, and a randomized trial evaluating the efficacy of thymectomy for nonthymomatous myasthenia gravis is in progress. By comparison, thymectomy is not beneficial in patients who have MuSK antibodies and in whom the thymus is generally devoid of hyperplastic changes. As a general rule, thymectomy, even in the presence of a thymoma, should never be an emergency treatment but rather should be postponed until a stable control of myasthenic symptoms is achieved.

Immunosuppressive therapy with prednisone is usually administered in the initial phases of treatment owing to its short-latency effect.[■] Most patients respond to alternate-day prednisone, started at a low dose (e.g., 10 mg every other day) and increasing by 5 to 10 mg per dose to 1.0 to 1.5 mg/kg. Because starting prednisone can temporarily exacerbate the disease, patients are usually best managed in the hospital, especially if they have bulbar or respiratory muscle involvement. When remission is established, the dose can be reduced by 5 to 10 mg every 2 weeks (or more slowly) to the effective minimal dose. Prophylactic treatment for osteoporosis (Chapter 243) and careful follow-up for other side effects is mandatory in all patients.

For long-term treatment, other immunosuppressive medication is required in patients who do not respond satisfactorily to prednisone or who need high maintenance doses. Because these agents have a long latency of effect, they are generally combined with prednisone (see earlier) during initial treatment and then used as monotherapy if steroids can be withdrawn or are contraindicated. Azathioprine (2.5 mg/kg/day) is the preferred treatment; compared with prednisone alone, combination treatment is better tolerated and associated with fewer relapses.[■] Cyclosporine (3 to 5 mg/kg daily) is effective as monotherapy or combined with steroids[■] and is frequently used as the second-choice immunosuppressant. Although the efficacy of mycophenolate mofetil in association with prednisone is questioned, this agent at the standard dose of 2000 mg/day is used in patients who are unresponsive to or intolerant of azathioprine. Tacrolimus is considered a third-line immunosuppressant.[■] Methotrexate (5 to 15 mg weekly) has similar efficacy and safety as azathioprine as a steroid-sparing agent.[■] When remission has been achieved, doses of these agents can be reduced slowly and cautiously; full withdrawal is likely to be followed by relapse.

Immunoglobulin infusion and plasmapheresis are equally efficacious[■] for providing short-term improvement, typically persisting 4 to 6 weeks, and can be used in preparation for thymectomy, to cover the initiation of prednisone therapy, or to control disease exacerbations. An immunoglobulin infusion of 1 g/kg given on day 1 only is as effective as 1 g/kg given on day 1 and again on day 2. Because of the short-lived benefits of these therapies, they must be accompanied by additional immunosuppressive therapy (see earlier).

Inhibition of the production of acetylcholinesterase using a short antisense oligonucleotide was both effective and safe in a phase Ib study on AChR antibody-positive patients (a phase II study is underway). High-dose cyclophosphamide[■] and two monoclonal antibodies, rituximab (which markedly reduces circulating B cells) and eculizumab (which binds C5 to preventing complement activation),[■] have been used successfully in patients with refractory disease.

PROGNOSIS

The increasing use of immunosuppressive therapies, coupled with advances in critical care, has greatly improved the prognosis of myasthenia gravis. Patients with myasthenic crisis are at high risk for recurrences,[12] but many patients achieve optimal control of symptoms with a normal life expectancy. The prognosis is not as good, however, in patients with invasive thymoma, who have a 5-year survival rate of about 80%, or with invasive thymic carcinomas, who have a 5-year survival rate of only about 40%.

Lambert-Eaton Myasthenic Syndrome

DEFINITION AND EPIDEMIOLOGY

The Lambert-Eaton myasthenic syndrome, which is a rare disorder that affects all races, can occur in paraneoplastic and nonparaneoplastic forms. The incidence of the paraneoplastic form is higher, but its shorter survival

results in a similar prevalence of the two types. The associated tumor is usually a small cell lung cancer (about 2% of patients with small cell lung cancer develop Lambert-Eaton myasthenic syndrome), and more rarely a lymphoma. The nonparaneoplastic form associates with HLA-A1, -B8, and -DR3, as in early-onset myasthenia gravis.

PATHOBIOLOGY

Lambert-Eaton myasthenic syndrome is an antibody-mediated presynaptic disorder characterized by a reduced number of acetylcholine quanta (vesicles) released by each nerve impulse. End-plate potentials recorded from intercostal muscle biopsies are consequently much reduced in amplitude. During high-frequency repetitive nerve stimulation, the end-plate potential amplitude increases, probably because build-up of calcium in the motor nerve terminal leads to increased release of acetylcholine. Freeze-fracture electron microscopic studies of motor nerve terminals show that the "active zone" particles, which correspond to voltage-gated calcium channels, are reduced in number and disorganized. The antibodies in Lambert-Eaton myasthenic syndrome bind to the presynaptic nerve terminal at the sites of acetylcholine release and appear to act principally by cross-linking the voltage-gated calcium channels, thereby leading to their clustering and internalization. The antibodies also interfere with transmitter release from postganglionic parasympathetic and sympathetic neurons in injected mice, providing an explanation for the autonomic dysfunction observed in many patients.

CLINICAL MANIFESTATIONS

Almost all patients present with difficulty in walking, which exhibits a rolling characteristic.[13] Weakness in ocular, bulbar, and respiratory muscles is less common than in myasthenia gravis. Weakness predominantly affects proximal muscles, which may show augmentation of strength during the first few seconds of a maximal contraction. Reflexes are absent or depressed but can increase after 10 seconds of maximal contraction of the muscle (post-tetanic potentiation). Autonomic symptoms such as dry mouth, constipation, and erectile dysfunction are present in most patients. Cerebellar ataxia may be present in association with small cell lung cancer. Patients with nonparaneoplastic Lambert-Eaton myasthenic syndrome may have other autoimmune diseases, notably vitiligo.

DIAGNOSIS

Diagnosis is based on the clinical features, on a positive serum voltage-gated calcium-channel antibody test, and on the characteristic EMG findings (see Table 422-2). Antibodies specific for the α1A (P/Q) subtype of voltage-gated calcium channels are found in 90% of patients, both with and without small cell lung cancer. Patients may not respond convincingly to intravenous edrophonium. On EMG, the amplitude of the resting compound muscle action potential is reduced. It decreases further during low-rate repetitive nerve stimulation but increases by more than 100% immediately after 10 seconds of voluntary contraction of the muscle or during high-frequency (40-Hz) nerve stimulation. Single-fiber EMG is less specific because an increased jitter does not distinguish between myasthenia gravis and Lambert-Eaton myasthenic syndrome.

On diagnosis, an extensive search for malignancy is necessary. All patients should undergo thoracic computed tomography scanning and fluorodeoxyglucose positron emission tomography (FDG-PET). If tumor screening is negative, it should be repeated periodically every 3 to 6 months for at least 2 years after the onset of neurologic symptoms.

Differential Diagnosis

Botulinum poisoning (Chapter 296) causes blockade of presynaptic transmitter release at the neuromuscular junction as well as EMG changes similar to those in the Lambert-Eaton myasthenic syndrome. Botulism is detected by finding the toxin in serum or the *Clostridium botulinum* bacteria in the wound or feces. Myopathies (Chapters 269 and 421) can mimic Lambert-Eaton myasthenic syndrome clinically, but autonomic changes do not occur, EMG findings are different, and muscle biopsy is abnormal.

TREATMENT Rx

Plasmapheresis leads to clinical improvement within a few days in acutely ill patients, and most patients respond to immunosuppressive drugs or intravenous immunoglobulin therapy. Intravenous immunoglobulin therapy (1 g/kg for 2 days) improves strength, with an associated decline in specific antibody.[14] Specific tumor treatment often leads to improvement of the neurologic disorder. Most patients respond to symptomatic treatment with 3,4-diaminopyridine (10 to 20 mg four times daily). However, 3,4-diaminopyridine has not yet been approved by the U.S. Food and Drug Administration, but a phosphate version of the drug has been licensed in Europe. Immunosuppressive treatment with prednisone, azathioprine, or cyclosporine may be required in patients with severe weakness, using doses similar to those prescribed for myasthenia gravis. Rituximab has been used in few patients with severe weakness.

PROGNOSIS

Prognosis mainly depends on that of the associated malignancy. Patients with paraneoplastic Lambert-Eaton myasthenic syndrome tend to have a progressive disease and a less satisfactory response to treatment.

ACQUIRED NEUROMYOTONIA

DEFINITION AND EPIDEMIOLOGY

Neuromyotonia, or Isaacs' syndrome, is a rare disorder primarily characterized by myokymia (spontaneous undulating muscle contractions) that can be intermittent or continuous and may be present during sleep or general anesthesia. It results from the hyperexcitability of motor nerves. A milder variant, the cramp-fasciculation syndrome, is more common.

PATHOBIOLOGY

Neuromyotonia may be associated with other autoimmune diseases or other autoantibodies, and cerebrospinal fluid analysis may show oligoclonal bands. In about 15% of patients, it is paraneoplastic, usually associated with thymoma and more rarely with lung cancer. Occasionally, neuromyotonia follows infection or allergic reactions, and it may improve spontaneously within weeks to months in these cases.

In neuromyotonia, peripheral nerve hyperexcitability is caused by dysfunction of voltage-gated potassium channel (Kv1), whose activation within milliseconds of nerve depolarization limits the depolarizing afterpotential and prevents the generation of repetitive discharges. In autoimmune neuromyotonia, pathogenic antibodies may induce loss of voltage-gated potassium channels and are often directed against contactin-associated protein-2, which is required for clustering Kv1 channels at the juxtaparanodal regions.

CLINICAL MANIFESTATIONS

The clinical presentation is variable but can include muscle stiffness, cramps, myokymia, fasciculations, pseudomyotonia (e.g., failure to relax after fist clenching), and weakness. Increased sweating is common. In the cramp-fasciculation syndrome, symptoms are milder and mostly induced by exertion. Some patients have sensory symptoms, including neuropathic-type pain, or less severe paresthesias, dysesthesia, and numbness, and a few have autonomic and central nervous system features of an encephalopathy, with insomnia, hallucinations, delusions, and mood change (Morvan's syndrome).

DIAGNOSIS

EMG shows spontaneous motor unit discharges: distinctive doublet, triplet, or multiplet bursts with high intraburst frequency (40 to 300 per second), longer continuous bursts, and postactivation contraction. The abnormal muscle activity may be generated at different sites throughout the length of the nerve but is usually distal. Many patients have serum antibodies to the voltage-gated potassium channel-complex (Kv1 channel and associated proteins), predominantly to contactin-associated protein-2.[15] The differential diagnosis includes neuromyotonia caused by acquired and inherited neuropathies and neuromyotonia—usually associated with episodic ataxia—caused by voltage-gated potassium-channel gene mutations (Kv1.1).

TREATMENT AND PROGNOSIS Rx

Neuromyotonia can be improved by anticonvulsant drugs, such as carbamazepine (up to 800 to 1000 mg daily), phenytoin (up to 300 mg daily), or lamotrigine (up to 100 mg daily), that depress sodium channel function and reduce the hyperexcitability of nerves. Plasmapheresis and intravenous immunoglobulins, using the same regimen as for myasthenia gravis, may be followed by short-term improvement. Immunosuppressive medications (as for

myasthenia gravis) are effective in some patients. Neuromyotonia is often a monophasic disease that can be successfully managed with symptomatic and immunodulating treatment. When it is associated with myasthenia gravis, the administration of pyridostigmine can increase symptoms of motor nerve hyperexcitability. Prognosis is less favorable in cases with central nervous system involvement, which may be associated with invasive thymomas.

GENETIC MYASTHENIC SYNDROMES

Congenital myasthenic syndromes (see Table 422-1) are inherited disorders that result from mutations in genes encoding key proteins at the neuromuscular junction. In the United Kingdom, their prevalence is at least 6 per 1 million population.

PATHOBIOLOGY

Congenital myasthenic syndromes are classified by the site of the mutated protein: presynaptic, synaptic, or postsynaptic. Postsynaptic disorders are more frequent and most commonly involve the AChR ε-subunit gene, in which single nucleotide missense substitutions or frameshift mutations result in complete loss of function of the AChR ε-subunit. Because this subunit replaces the AChR γ-subunit around the time of birth, infants are normal in development but show weakness during late pregnancy and in the neonatal period. Survival depends on the continued expression of the γ-subunit. By contrast, null mutations in non-epsilon subunits are probably lethal. AChR deficiency can also result from defects in the gene for rapsyn, a cytoplasmic protein required for the clustering of the AChRs at the neuromuscular junction. Single-nucleotide changes in genes for any of the AChR subunits can affect affinity for acetylcholine and gating efficiency, thereby leading to kinetic defects. In the fast-channel syndrome (recessive), AChR openings are abnormally brief, whereas the opposite occurs in the slow-channel syndrome (dominant), in which the channel opens for prolonged periods, thereby resulting in subsynaptic accumulation of cations and degenerative changes with loss of AChR.

Mutations in the *COLQ* gene, which gives rise to the collagen tail that anchors acetylcholinesterase in the synaptic cleft, are less common. The absence of acetylcholinesterase is responsible for reduced quantal release and for continuous exposure of the postsynaptic membrane to acetylcholine, thereby leading to cation overload and junctional fold degeneration. Mutations in choline acetyltransferase, the enzyme responsible for the synthesis of acetylcholine, do not always lead to dysfunction at rest; during repetitive activity, however, the amount of acetylcholine in each packet decreases, with consequent failure of neuromuscular transmission. Mutations in *DOK7* cause a "synaptopathy" with small, simplified neuromuscular junctions. Dok-7 binds MuSK, and the mutations are thought to impair the signaling that maintains the synaptic structure. Other gene mutations are less common (see Table 422-1).

CLINICAL MANIFESTATIONS

Clinical manifestations may vary from death in utero in severe cases to mild symptoms that present in adulthood.[16] Although most cases present in infancy with ptosis, hypotonia, and difficulties with feeding and breathing, the slightly different patterns of muscle weakness provide clues that point to which gene is involved. Arthrogryposis multiplex congenita, indicative of fetal akinesia, occurs with rapsyn mutations. Life-threatening episodic apnea can occur with mutations in choline acetyltransferase or rapsyn or in fast-channel syndromes. Severe ophthalmoplegia occurs in end-plate acetylcholinesterase deficiency, AChR deficiency due to AChR subunit mutations, and fast-channel syndromes, but is rarely seen in the other genetic syndromes. Motor symptoms with *DOK7* mutations usually appear at about 2 years of age after the child first learns to walk and are characterized by a limb-girdle weakness associated with ptosis and by facial and bulbar muscle involvement.

DIAGNOSIS

A congenital myasthenic syndrome should be considered when symptoms are evident at birth or during early infancy and other relatives are affected. However, a negative family history does not exclude the diagnosis, and the onset can be later in the slow-channel syndrome, rapsyn, and *DOK7* mutations.

Impaired neuromuscular transmission can be detected by a decremental response on repetitive nerve stimulation and increased jitter on single-fiber EMG. In the slow-channel and acetylcholinesterase deficiency syndromes, the prolonged end-plate potential outlasts the refractory period of the muscle fiber, and a single nerve stimulus can be followed by a repetitive compound muscle action potential (double response) (see Table 422-2). Genetic analysis is essential to confirm the diagnosis and help in treatment, prognosis, and counseling, although the faulty gene has not been identified in many families.

The principal differential diagnoses are spinal muscular atrophy, infant botulism, hereditary neuropathies, and congenital myopathies or muscular dystrophies. Onset in early childhood, adolescence, or adulthood may lead to the incorrect diagnosis of seronegative myasthenia gravis.

TREATMENT AND PROGNOSIS

Rx

Many of the congenital myasthenic syndromes respond to acetylcholinesterase inhibitors, as used for myasthenia gravis, and to 3,4-diaminopyridine (1 mg/kg/day in four divided doses). Patients with the slow-channel syndrome respond to quinidine (at doses corresponding to serum levels of 1 to 2.5 mg/L) or to fluoxetine (60 to 100 mg/day in adults though some patients may respond to doses as low as 20 mg), but the use of fluoxetine in children or adolescents requires psychiatric supervision. For syndromes in which the neuromuscular junction is destabilized or there are degenerative changes, such as for Dok-7 or end-plate acetylcholinesterase deficiency, treatment with ephedrine (45 to 100 mg/day in adults, 3 mg/kg/day in children) or salbutamol (0.5 to 4 mg, three times a day) can be remarkably effective. The beneficial effects of this treatment are not seen immediately but build up over a period of 6 months or more.

Although congenital disorders can be fatal during infancy, usually because of apneic episodes during infections, most tend to be nonprogressive or even may improve during adolescence or adult life. The exceptions are the slow-channel syndrome and acetylcholinesterase deficiency, which, owing to the excess AChR activations, can be associated with end-plate progressive degenerative changes, although this risk is largely mitigated with treatment.

Grade A References

A1. Schneider-Gold C, Gajdos P, Toyka KV, et al. Corticosteroids for myasthenia gravis. *Cochrane Database Syst Rev*. 2005;2:CD002828.

A2. Palace J, Newsom-Davis J, Lecky B. A randomized double-blind trial of prednisolone alone or with azathioprine in myasthenia gravis. Myasthenia Gravis Study Group. *Neurology*. 1998;50:1778-1783.

A3. Hart IK, Sathasivam S, Sharshar T. Immunosuppressive agents for myasthenia gravis. *Cochrane Database Syst Rev*. 2007;4:CD005224.

A4. Nagane Y, Utsugisawa K, Obara D, et al. Efficacy of low-dose FK506 in the treatment of myasthenia gravis: a randomized pilot study. *Eur Neurol*. 2005;53:146-150.

A5. Heckmann JM, Rawoot A, Bateman K, et al. A single-blinded trial of methotrexate versus azathioprine as steroid-sparing agents in generalized myasthenia gravis. *BMC Neurol*. 2011;11:97.

A6. Gajdos P, Chevret S, Toyka KV. Intravenous immunoglobulin for myasthenia gravis. *Cochrane Database Syst Rev*. 2012;12:CD002277.

A7. De Feo LG, Schottlender J, Martelli NA, et al. Use of intravenous pulsed cyclophosphamide in severe, generalized myasthenia gravis. *Muscle Nerve*. 2002;26:31-36.

A8. Howard JF Jr, Barohn RJ, Cutter GR, et al. A randomized, double-blind, placebo-controlled phase II study of eculizumab in patients with refractory generalized myasthenia gravis. *Muscle Nerve*. 2013;48:76-84.

A9. Keogh M, Sedehizadeh S, Maddison P. Treatment for Lambert-Eaton myasthenic syndrome. *Cochrane Database Syst Rev*. 2011;2:CD003279.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Andersen JB, Heldal AT, Engeland A, et al. Myasthenia gravis epidemiology in a national cohort; combining multiple disease registries. *Acta Neurol Scand Suppl*. 2014;26-31.
2. Berrih-Aknin S, Le Panse R. Myasthenia gravis: a comprehensive review of immune dysregulation and etiological mechanisms. *J Autoimmun*. 2014;52:90-100.
3. Marx A, Pfister F, Schalke B, et al. The different roles of the thymus in the pathogenesis of the various myasthenia gravis subtypes. *Autoimmun Rev*. 2013;12:875-884.
4. Koneczny I, Cossins J, Vincent A. The role of muscle-specific tyrosine kinase (MuSK) and mystery of MuSK myasthenia gravis. *J Anat*. 2014;224:29-35.
5. Zisimopoulou P, Evangelakou P, Tzartos J, et al. A comprehensive analysis of the epidemiology and clinical characteristics of anti-LRP4 in myasthenia gravis. *J Autoimmun*. 2014;52:139-145.
6. El-Salem K, Yassin A, Al-Hayk K, et al. Treatment of MuSK-associated myasthenia gravis. *Curr Treat Options Neurol*. 2014;16:283.
7. Berrih-Aknin S, Frenkian-Cuvelier M, Eymard B. Diagnostic and clinical classification of autoimmune myasthenia gravis. *J Autoimmun*. 2014;48-49:143-148.
8. Skeie GO, Apostolski S, Evoli A, et al. Guidelines for treatment of autoimmune neuromuscular transmission disorders. *Eur J Neurol*. 2010;17:893-902.
9. Reddel SW, Morsch M, Phillips WD. Clinical and scientific aspects of muscle-specific tyrosine kinase-related myasthenia gravis. *Curr Opin Neurol*. 2014;27:558-565.
10. Kerty E, Elsais A, Argov Z, et al. EFNS/ENS guidelines for the treatment of ocular myasthenia. *Eur J Neurol*. 2014;21:687-693.
11. Ettinger DS, Riely GJ, Akerley W, et al. Thymomas and thymic carcinomas: clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2013;11:562-576.
12. Kalita J, Kohat AK, Misra UK. Predictors of outcome of myasthenic crisis. *Neurol Sci*. 2014;35:1109-1114.
13. Titulaer MJ, Lang B, Verschuuren JJ. Lambert-Eaton myasthenic syndrome: from clinical characteristics to therapeutic strategies. *Lancet Neurol*. 2011;10:1098-1107.
14. Evoli A, Liguori R, Romani A, et al. Italian recommendations for Lambert-Eaton myasthenic syndrome (LEMS) management. *Neurol Sci*. 2014;35:515-520.
15. Irani SR, Alexander S, Waters P, et al. Antibodies to Kv1 potassium channel-complex proteins leucine-rich, glioma inactivated 1 protein and contactin-associated protein-2 in limbic encephalitis, Morvan's syndrome and acquired neuromyotonia. *Brain*. 2010;133:2734-2748.
16. Rodríguez Cruz PM, Palace J, Beeson D. Inherited disorders of the neuromuscular junction: an update. *J Neurol*. 2014;261:2234-2243.

REVIEW QUESTIONS

1. A 54-year-old man has been recently diagnosed with anti-AChR-positive myasthenia gravis. The disease has had an acute onset and rapid progression, leading in few weeks to severe generalized weakness with marked bulbar symptoms, including dysphagia, dysarthria, neck weakness, and impaired respiratory function. A thymoma is detected on thoracic computed tomography scan. Symptoms are not satisfactorily relieved by full-dose pyridostigmine treatment. Which of the following should be considered as first therapeutic option in this case?

A. Immediate thymectomy
B. Plasmapheresis (or immunoglobulin infusion) and prednisone
C. Plasmapheresis
D. Azathioprine
E. Rituximab

Answer: B Thymectomy is indicated and should be performed as soon as possible because of the presence of a thymoma. The patient has severe and poorly controlled symptoms. In this situation, surgery should be deferred until the myasthenia gravis improves. Plasmapheresis (or immunoglobulin infusion) provides a rapid onset of improvement, which is, however, short-lived unless immunosuppressive treatment is added. Prednisone is the first-choice immunosuppressant because of its short-latency effect; the association of plasmapheresis (or immunoglobulin infusion) can enhance the therapeutic effect and minimize the temporary exacerbation that can occur at the start of prednisone administration. Because of its long-latency effect, azathioprine is not the first option in this case. Rituximab is reserved for refractory disease.

2. A 70-year-old woman presents with asymmetrical ptosis, which is more pronounced on her right eye, and intermittent diplopia. Her symptoms fluctuate daily, and they improve with rest but worsen in the evening. Her symptoms have been present for more than 1 year, with periods of spontaneous remission. On the whole, the symptoms do not interfere with her daily activities. On clinical examination, mild fatigability of neck extensors and arm proximal muscles is also evident. Myasthenia gravis is diagnosed by detecting anti-AChR antibodies, increased jitter on SF-EMG, and a positive response to neostigmine. Thoracic computed tomography scan is negative. The patient also has diabetes controlled by diet. Which of the following is the correct recommendation as first treatment in this patient?

A. Pyridostigmine
B. Prednisone and azathioprine
C. Thymectomy
D. Plasmapheresis
E. Fluoxetine

Answer: A This patient has mild generalized myasthenia gravis with predominantly ocular symptoms. The disease has a slow course and is not disabling. In this context, treatment with pyridostigmine can be sufficient to control patient's symptoms and should be tried first. Immunosuppressive therapy (prednisone and azathioprine) and plasmapheresis should be considered if the disease worsens. Thymectomy is not indicated because the patient has late-onset nonthymomatous disease. Fluoxetine blocks AChR channels; it is contraindicated in myasthenia gravis, and its use is limited to the treatment of the slow-channel myasthenic syndrome.

423

DISEASES OF THE VISUAL SYSTEM

MYRON YANOFF AND J. DOUGLAS CAMERON

The eye is a compact, complicated structure (Fig. 423-1) that is remarkably stable throughout life. Once the growth of the eye is complete, at approximately age 3 years, the structure of the eye changes very little for the next 60 to 80 years.

The eyelids physically protect the eye. The pathway of light through the eye, termed the *visual axis,* is transparent and contains no opaque structures such as blood vessels. Light passes through the tear film, cornea, intraocular aqueous, crystalline lens, vitreous, and retina, all of which, except for the crystalline lens, remain essentially transparent throughout life. Delicate intraocular structures are protected by a tough collagenous "eye wall" composed of the cornea and sclera. The optic nerve, which is composed of axons from the retinal ganglion cells, is supported by dura, arachnoid, and pia mater, which are contiguous with the brain. The optic nerve is long enough to allow free excursions of the eye through a 100-degree arc under the influence of six coordinated and critically placed rectus muscles. All these functional components are housed in a bony cavity, the orbit, which protects the eye from external injury.

The eyelid skin, only loosely connected to underlying structures, is among the thinnest of the body. The eyelid is unique because it contains the highest density of sebaceous glands in the body. These meibomian glands produce a sebaceous (lipid) material that is the principal evaporation retardant for the tear film. Malorientation of the eyelid margin or malorientation of the cilia (trichiasis) may cause extensive scarring of the anterior surface of the cornea, even to the point of blindness. The eyelid is opened by contraction of the levator muscle. The tendon of the levator tends to degenerate over time to produce mechanical ptosis. The soft tissue of the eyelid is separated from the soft tissue of the orbit by the orbital septum, a major collagenous barrier that protects intraorbital soft tissue from extension of preseptal eyelid inflammation. Extension of inflammation from preseptal cellulitis or ethmoiditis may cause septic optic neuropathy or cavernous sinus thrombosis. The elastic tissue supporting the skin of the anterior eyelid is reduced over time, thereby causing dermatochalasis ("baggy eyelids"). Redundant tissue may be sufficient in quantity to restrict the visual field, particularly superiorly.

The conjunctiva is a mucous membrane covered by stratified, nonkeratinizing squamous epithelium containing goblet cells. The epithelium is supported by delicate fibrovascular tissue that contains lymphatic channels. Squamous carcinoma or malignant melanoma originating in the conjunctiva may extend through these channels to regional lymph nodes or beyond. The conjunctival epithelium contains melanocytes. Immune processing cells are present in the epithelium (Langerhans cells) and in the stroma as collections of non-nodal B and T lymphocytes. Non-nodal primary lymphomas, which tend to have an indolent course in this location, may arise from this tissue. The aqueous portion of tears is formed constantly by accessory lacrimal glands in the conjunctiva and eyelid soft tissue as well as by reflex action from the lacrimal gland. Symptoms of itching and burning, as well as periodic disturbance of vision, may result from inadequacies of the tear film layer.

Tears drain through puncta at the nasal eyelid margin through the nasolacrimal duct to exits in the nasal cavity inferior to the inferior turbinate. The epithelium of the nasolacrimal duct also contains melanocytes and is supported by a resting lymphocyte population. Neoplasms including lymphoma, concretions (dacryoliths), and tissue injury from trauma may occlude the puncta in adults.

The cornea is avascular and lined both anteriorly and posteriorly by surface cells. Lack of an adequate tear film (dry eye syndrome) may seriously alter the ability of the cornea to transmit light, thereby affecting visual acuity. The posterior cellular lining of the cornea is a single layer of highly modified corneal endothelial cells that maintain tissue dehydration. Lack of effective pumping by the endothelial cells will allow excess hydration of the corneal stroma, that is, corneal edema. The corneal stroma is particularly sensitive to proteolysis from collagenases found with certain inflammatory conditions, such as herpes simplex keratitis. The cumulative effect of multiple episodes may be corneal thinning and possible perforation of the cornea.

Intraocular pressure is measured by applanation tomography. The amount of pressure necessary to flatten the central cornea is proportional to the intraocular pressure.

The anterior chamber is bounded by the posterior surface of the cornea, the anterior surface of the iris, and the anterior surface of the crystalline lens within the pupillary space. Aqueous material normally flows from the posterior chamber into the anterior chamber through the pupil and exits into the general circulation through the trabecular meshwork. Most causes of pathologically increased intraocular pressure and optic nerve damage (i.e. glaucoma) are due to abnormalities of filtration through the trabecular meshwork. The posterior chamber is bounded by the posterior surface of the iris, the ciliary body circumferentially, and the anterior surface of the vitreous. The crystalline lens is located entirely in the posterior chamber.

The anterior segment is composed of the cornea and the anterior and posterior chambers. Most of the anterior segment is derived from the skin and neural crest tissue. The posterior segment is the remainder of the eye. Most of the posterior segment structures are derived from the central nervous system and neural crest tissue.

When first formed, the crystalline lens is a totally cellular structure bounded by a true basement membrane. Throughout life, the new cells that are continuously added from the outer layer epithelial cells compress the central cells, thereby resulting in cell degeneration in the central core (nucleus). The lens doubles in volume from birth to age 70 years at the cost of both pliability (presbyopia) and clarity (cataract). The lens is suspended in the posterior chamber by fibers (zonules) attached to the ciliary body.

The ciliary body is the posterior extent of the iris. Its surface cells produce aqueous, and its muscles function in accommodation.

The vitreous is composed primarily of water and type II collagen. The vitreous makes up the majority of the volume and weight of the eye. It functions as a biochemical sink as well as to maintain neural retinal attachment. With time, the vitreous shrinks and separates from the retina (posterior vitreous detachment). Condensed and displaced vitreous casts a shadow on the retina, which is perceived by the patient as "floaters."

The retina is the site of photochemical conversion of light to electrical energy. Ganglion cells and their axons in the internal retina aggregate at the optic disc to form the optic nerve. Only the inner half of the retina is supplied by intraretinal vessels that are seen by ophthalmoscopy. The outer half of the retina is supplied by large-caliber capillary vessels in the choroid (the choriocapillaris). Only a 500-μm area of the posterior retina, the central macula (about 3 to 5% of the total retina), has the ability to resolve images to 20/20. The remainder of the retina has much less sensitive image resolution. Extensive biochemical support and control of stray light are performed by the retinal pigment epithelium located between the choriocapillaris and the photoreceptor outer segments. The blood-retinal barrier, which protects the biochemical integrity of the retina, is composed of anatomic attachments between neighboring retinal pigment epithelial cells, as well as attachments between vascular endothelial cells of the retinal circulation. The retina is held in place by physiologic forces that may be compromised by holes in the retina (rhegmatogenous retinal detachment) or by fluid accumulating in the subretinal space (serous retinal detachment).

The optic nerve is composed of approximately 1 million axons from retinal ganglion cells. Axons are separated into bundles by pial septa, which are in turn enclosed in an arachnoid layer. The dura is contiguous with the posterior sclera and the periosteum of the optic canal. Delicate vessels extending from the dura across the arachnoid to the pial septa supply the optic nerve. The central retinal artery is present in the axial layer of the optic nerve near the eye but does not supply blood to the optic nerve itself. The optic nerve axons travel through a collagenous sieve in the plane of the posterior sclera, the lamina cribrosa. The choroid is that portion of the uveal tract external to the retina. This layer is composed of various calibers of blood vessels that ultimately supply blood for the choriocapillaris. There are no lymphatic channels in the choroid.

The sclera is composed of dense, relatively disorganized collagen. It is opaque because of the nonhomogeneous structure of the collagen and the degree of hydration relative to the cornea. There are multiple scleral ostia for arteries, veins, and nerves, both posteriorly and anteriorly.

The orbit is composed of bones of the facial skeleton. Sutures between major bones exist in the superior nasal and superior temporal quadrants. Multiple vessels and nerves extend through the thin ethmoid bone from nasal sinus tissue medially. The orbital floor is poorly supported over the maxillary sinus and may rupture with increased intraorbital pressure. The nasolacrimal duct travels through a portion of the lacrimal bone. Portions of the sphenoid

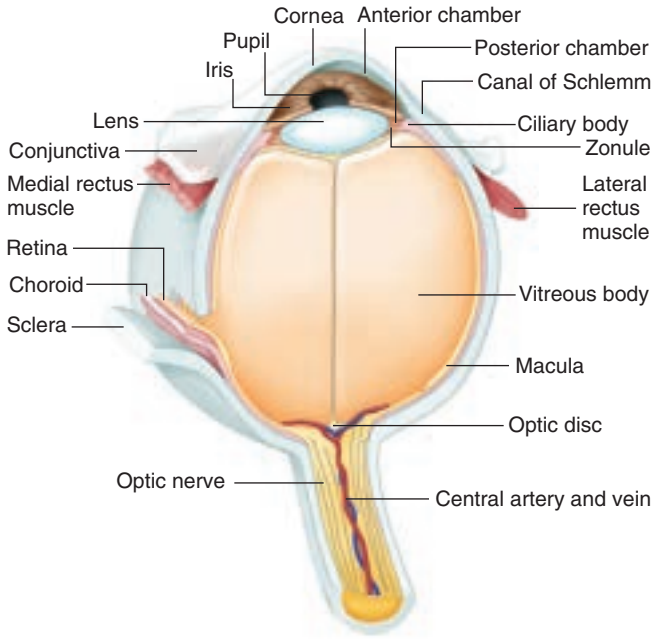

FIGURE 423-1. Anatomy of the eye.

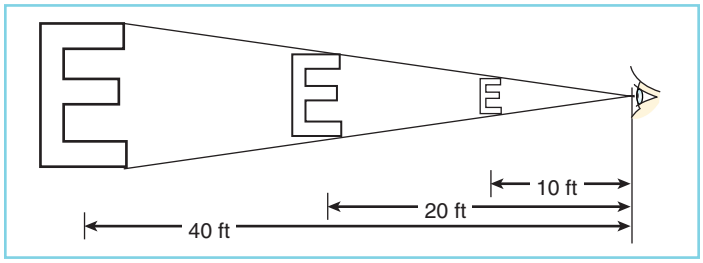

FIGURE 423-2. Snellen visual acuity. The most common test for visual acuity measures the eye's ability to resolve linear images at a test distance of 20 feet, approximating infinity (parallel rays of light). A 20/20 E subtends 5 minutes of arc at a distance of 20 feet, with each segment of the E subtending 1 minute of arc. The larger letters (e.g., 20/30, 20/40) are determined by the distance at which they subtend an angle of 5 minutes. Thus, an E that subtends 5 minutes at 40 feet, if viewed clearly at 20 feet, indicates 20/40 visual acuity.

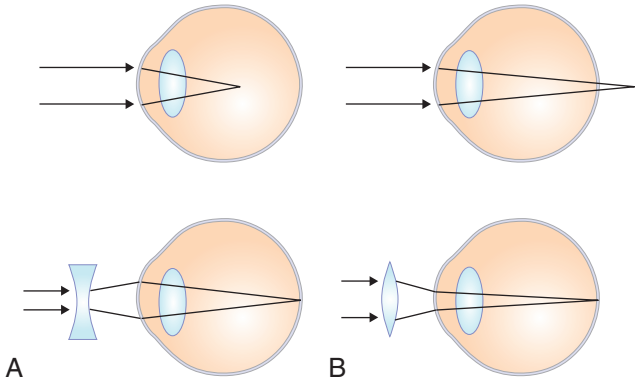

FIGURE 423-3. Myopia/hyperopia. A, In the myopic eye, parallel rays of light are focused anterior to the retina. A divergent lens can be used to compensate for the mismatch between refracting power and axial length. **B,** The hyperopic eye requires the additional power of a convergent lens to bring images into focus on the retina.

bone protect the optic nerve. Major cranial nerves travel through the adjacent superior orbital fissure, also a portion of the sphenoid bone. There is no normal lymphoid tissue in the orbit outside of the lacrimal gland. The rectus muscle may be enlarged by inflammation in thyroid eye disease, but the tendinous insertion into the sclera is not inflamed early in the course of the disease.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Patients may present with complaints of diminished vision, eye pain, red eyes, or pain around the eye. The causes may be primarily ophthalmic (e.g., cataract) or systemic (e.g., diabetic retinopathy). A comprehensive ophthalmologic examination also should evaluate for possible asymptomatic local (e.g., choroidal melanoma) or systemic (e.g., hypertensive retinopathy) abnormalities in patients with normal acuity and no subjective complaints.

Functional Evaluation

The most objective and common measure of ocular function is line letter acuity, with normal vision (Table 423-1) defined as the ability to see at 20 feet what a normal person sees at 20 feet (Fig. 423-2). Less than 20/20 vision can be caused by an abnormality anywhere from the tear film to the visual cortex of the occipital lobe (see Fig. 423-1). However, normal vision also includes other functions, such as the perception of color, motion, contrast, brightness, field, and depth, for which there is greater variation among individuals and no universally adopted, standardized scales. Normal visual acuity is potentially achievable in essentially all individuals, either naturally or with visual correction.

Correction of vision is based on the refraction of light (Fig. 423-3). The diopter (D) is the unit of measurement of the ability of an optical system to refract (bend) light. The normal human eye has the refractive capacity of approximately 60 D. If the eye is too short, light will be focused behind the eye (hyperopia). If the eye is too long, light will be focused in the vitreous in front the retina (myopia). Normally, a person can voluntarily control the crystalline lens, alternating between near and distant tasks. At approximately age 45 years, the lens becomes stiff, and the eye becomes set for distance (presbyopia). Refraction is the method of determining the amount of optical correction (strength of glasses) needed to establish 20/20 (6/6) vision.

Examination of pupillary response assesses whether neural function is intact (see Figs. 424-2 and 424-4). Confrontational visual field testing (see Fig. 424-1 in Chapter 424) should be performed in each eye to detect gross quadrantic defects. Extraocular motility should be assessed to exclude nerve or muscle abnormalities (see Table 424-4). Color vision testing plates are a sensitive indicator of optic nerve function.

Diagnostic testing during a routine eye examination also includes external examination of the lids and adnexa, applanation tonometry to determine intraocular pressure, biomicroscopy (slit lamp examination) of the anterior segment, and ophthalmoscopic examination of the ocular fundus. Other office tests when indicated include exophthalmometry (measurement of proptosis), visual field and electrophysiologic testing, and vascular imaging (fluorescein angiography, mainly in diabetic patients; ocular computed tomography [OCT] to investigate retinal macular disease); and corneal topography.

The conjunctiva, cornea, lens, and anterior chamber are evaluated using a slit lamp. The slit lamp is composed of a binocular microscope with variable magnification (40× and 80×) in conjunction with adjustable light sources. An increased concentration of protein can be detected in the anterior chamber because of the Tyndall (flare) effect, indicating vascular incompetence associated with either inflammation or ischemia. Even individual inflammatory cells can be resolved with the slit lamp. A cobalt blue filter can be used to detect fluorescein dye that accumulates in regions of abnormal epithelium (dendrite of herpes simplex keratitis or a corneal abrasion). The slit beam is used to examine the crystalline lens to determine the depth of the anterior chamber and the degree of opacification from a cataract. By using a green filter with the 90-D lens, the retinal vessels and retinal vascular

TABLE 423-1 VISUAL ACUITIES REQUIRED FOR COMMON DAILY TASKS

20/20	Physiologic vision
20/30-20/100	Driver's license, varies by state
20/50	Newspaper print
20/70	Large-print *Reader's Digest*
20/100	Write a check
20/200	Legally blind
20/400	Paper currency

abnormalities such as microaneurysm can be seen at relatively high magnification.

COMMON CLINICAL CONDITIONS

Chronic Abnormal Vision

MYOPIA

Nearsightedness (myopia; see Fig. 423-3) is usually discovered during childhood when children cannot perform distant tasks during school (reading the blackboard) or during school screening. Myopia usually progresses until the eye is fully developed, typically by age 20 to 25 years. Rapidly progressing myopia during childhood or at any time after age 25 years requires evaluation for juvenile glaucoma, diabetes mellitus (reversible metabolic changes in the crystalline lens), trauma (development of a cataract), or use of corticosteroids (development of a cataract). Myopia is usually fully correctable with glasses. Laser in situ keratomileusis (LASIK) is one of the refractive surgical procedures that may be used in adults to correct myopia and other refractive errors, with 95% of patients achieving visual acuity of 20/40 or better. Complications of LASIK include glare symptoms, dry eye, and undercorrection or overcorrection. Rare but serious complications include epithelial ingrowth, diffuse keratitis, and flap dislocation. In patients who undergo subsequent cataract surgery, special attention is needed to calculate the parameters for the intraocular lens.

Pathologic myopia is a heritable condition causing progressive weakening of the posterior sclera and resulting increases in the axial length of the eye. The posterior radius of curvature of the eye increases (posterior staphyloma). The refractive error is usually above −8 D and may be as high as −20 D in severe cases. Abnormal physical forces in pathologic myopia may lead to retinal hole formation, retinal detachment, or intraocular hemorrhage. The primary abnormality leading to loss of vision is choroidal neovascularization. Associated systemic conditions include trisomy 21, Cornelia de Lange's syndrome, Stickler's syndrome, and Marfan syndrome. Clinical surveillance must be maintained to watch for treatable complications such as retinal detachment. Pathologic myopia is usually treated with spectacles or contact lenses. Refractive procedures are less successful in pathologic myopia because of the severity of the refractive errors and the presence of posterior segment abnormalities. Surgical and laser procedures may be required to treat retinal and choroidal lesions (e.g., subretinal neovascularization).

HYPEROPIA

In hyperopia (farsightedness; see Fig. 423-3), in contrast to myopia, the eye tends to have a shorter than average axial length. Compensatory mechanisms of the crystalline lens may functionally correct small degrees of hyperopia until age 40 years, when the crystalline lens loses its pliability. The 40-year-old hyperopic changes from latent to manifest hyperopia simultaneously. The initial pair of glasses may have to correct for both distance and near tasks (bifocals). Refractive surgical procedures can correct up to 5 D of hyperopia.

ASTIGMATISM

Astigmatism usually is the result of irregularities in the radius of curvature of the cornea. Astigmatism is not a pathologic state, but rather a variation in anatomy; most people have some degree of astigmatism. Trauma to the cornea can cause alteration of structure, leading to irregular astigmatism. Symptoms are predominantly blurry vision and difficulty seeing fine detail. Regular astigmatism can be corrected with spectacle lenses or with rigid contact lenses, whereas irregular astigmatism requires rigid contact lenses. Some forms of astigmatism can also be corrected with laser ablation of the cornea.[1]

KERATOCONUS

Keratoconus is an acquired irregularity of corneal curvature; its cause is controversial but probably at least partly genetic. Onset generally is during adolescence, and the process often evolves over 5 years or so. The prevalence in the United States appears to be about 55 cases per 100,000 people but is probably greater when subclinical cases diagnosed by computerized videokeratography are considered.

Patients with keratoconus typically have astigmatism, which may be severe. The condition is usually bilateral but not symmetrical. Mild degrees of keratoconus can be corrected with glasses or contact lenses. Corneal collagen cross-linking can provide a sustained improvement in vision. For severe cases, corneal transplantation (graft) often is very successful.[2]

STRABISMUS

The control of simultaneous orientation of the two eyes to ensure that the visual axes of both eyes are aligned is not complete until several years following birth. Misalignments of the two eyes (strabismus) may be the result of abnormalities in the central nuclei of the brain, malfunction of one or several peripheral nerves, or intrinsic abnormalities of the rectus or oblique muscles (see Fig. 424-6). If the eyes are not simultaneously stimulated with the images of the same degree of clarity or complexity, only one eye will develop normally. The image processing ability of the blurred eye will not develop (amblyopia). In most cases, only the central vision is affected. The peripheral vision in both eyes is likely to be equal and normal.

Amblyopia also may be caused by a marked difference in refractive error between the two eyes (anisometropic amblyopia) or eyelid ptosis (deprivational amblyopia). Ptosis may be neurogenic or mechanical (e.g., congenital eyelid hemangioma). To avoid amblyopia, it is extremely important to refer a child with strabismus to an ophthalmologist as soon as the strabismus is noted.

Treatment options include patching or atropine eye drops, which are equally effective in providing good vision if patients are treated before age 7 years.[3] The outcome of treatment is more favorable if amblyopia is detected before age 2 to 3 years but may be successful into the teenage years.

Esotropia is deviation of one or both eyes inward. Esotropia may not be clinically evident until the child is 3 to 4 months of age. Delay in facial maturation (underdeveloped nasal bridge) may give the appearance of esotropia, even though the visual axes are correctly aligned. In true strabismus, the light reflex will be in the center of one cornea and decentered in the other.

Exotropia is deviation of one or both eyes outward. Exotropia tends to be intermittent and less likely to result in amblyopia.

Both esotropia and exotropia may be treated by using appropriate corrections of refractive error with glasses (occasionally bifocals). Occasionally extraocular muscle surgery is necessary to correct alignment. To allow the second eye to develop to its full potential, amblyopia is corrected by occluding the stronger eye with a patch or by pharmacologically occluding (blurring) the better eye with atropine. Success is directly related to compliance.

DIPLOPIA (DOUBLE VISION)

Acute onset of diplopia is an ominous sign suggestive of a third nerve palsy (Chapter 424). Diplopia of any kind is an intolerable symptom, but vertical diplopia is less tolerated than horizontal diplopia.

COLOR VISION CHANGE

Most cases of congenital color blindness go undetected for many years. Acquired color deficiency at any age may be caused by a cataract or optic nerve disease.

CHANGE IN VISION

If only one eye has a change in vision, the problem, such as a cataract or retinal detachment, is most likely in that eye. If both eyes have a change in vision, the problem generally is outside of the eye, such as homonymous hemianopia (Chapter 424). Improvement of near vision in middle age may be a sign of cataract ("second sight") or hyperglycemia. Transient complete unilateral or bilateral loss of vision may be caused by vascular abnormalities inside or outside of the eye (Table 423-2).

Acute Eye Abnormalities

PAIN

The most severe eye pain (Table 423-3), typically associated with a red eye, is caused by acute angle-closure glaucoma. Sharp, intermittent pain is usually caused by ocular surface abnormalities (e.g., corneal foreign body). Burning pain that clears with blinking generally relates to tear film abnormalities (dry eyes). Deep boring pain most often is associated with an ocular abnormality (e.g., uveitis).

RED EYE

A red or inflamed eye can be caused by conjunctivitis, iritis (anterior uveitis), acute glaucoma, corneal trauma, or infection (Table 423-4). Of these causes, all are typically painful, with the occasional exception of conjunctivitis.

DISTORTED VISION

Distorted vision (metamorphopsia), which is the perception that straight lines are distorted or bowed, results from macular dysfunction. Causes

include fluid under the retina; exudative macular degeneration, which tends to elevate the retina; and an epiretinal membrane, which tends to contract the retina.

NIGHT BLINDNESS

Retinitis pigmentosa, vitamin A deficiency, and systemic medications such as phenothiazines can cause true night blindness, in which patients have difficulty seeing any stars in the sky on a clear night and may be unable to ambulate without assistance in a dark environment. Patients with cataracts may have difficulty driving at night because of excessive glare and visual distortion.

SENSATION OF FLASHING LIGHTS

The sudden onset of flashes in the peripheral visual field suggests a posterior vitreous detachment with resulting traction of the vitreous on the peripheral retina, sometimes with a resulting retinal tear. The flashes, which may be more pronounced in the dark and with rapid eye movement, may be associated with the sudden onset of floaters, which can indicate debris or blood in the vitreous cavity. Because a tear in the retina can lead to a retinal detachment, urgent consultation with an ophthalmologist is required.

TABLE 423-2 DIFFERENTIAL DIAGNOSIS OF SUDDEN VISUAL LOSS

UNILATERAL	BILATERAL
Amaurosis fugax (carotid artery stenosis)	Eclampsia
Central retinal artery occlusion	Vertebrobasilar infarct
Occipital lobe infarct	Trauma
Temporal arteritis	
Nonarteritic anterior ischemic optic neuropathy	
Hemorrhage	
Preretinal (high altitude, Valsalva)	
Vitreous	
Aqueous (hyphema)	
Trauma	

TABLE 423-3 CAUSES OF EYE PAIN

Blepharitis	Glaucoma
Blocked tear duct	Hordeolum (stye)
Chalazion	Iritis
Conjunctivitis	Keratoconus
Corneal abrasion	Optic neuritis
Dry eyes	Scleritis
Ectropion	Trauma
Entropion	Uveitis
Foreign object	

Flashing light with a migraine (Chapter 398) is described as scintillations or zigzagging lights that march across the visual field for a few minutes or as long as 30 minutes, sometimes associated with transient visual field loss. Headache is not universal.

FLOATERS

Floaters, which are caused by small aggregates in the vitreous cavity, result from the normal aging of the vitreous. The acute onset of vitreous floaters may be associated with uveitis or with the sudden onset of bleeding in the vitreous cavity owing to diabetes or sickle cell anemia. Acute floaters, however, particularly if associated with flashing lights, may indicate a posterior vitreous detachment or a retinal tear with an impending retinal detachment. Urgent ophthalmic referral is essential.

PHOTOPHOBIA

Photophobia, particularly if associated with eye pain, redness, and decreased vision, is a symptom of uveitis or traumatic iritis. Photophobia is also typical of acute migraine and meningeal irritation. Prompt ophthalmologic referral is prudent.

HALOS AROUND LIGHTS

Patients with cataracts commonly see halos around lights, particularly when driving at night. Episodic decreased vision, redness, and halos around lights may be symptoms of impending angle-closure glaucoma. Halos also can occur as a complication of refractive eye surgery.

FOREIGN BODY SENSATION

A foreign body sensation is commonly caused by dry eyes. An entropion (Fig. 423-4) or misdirected lashes (trichiasis) also can cause a foreign body sensation. Most corneal abrasions cause severe pain, but minor corneal abrasions may be associated with a foreign body sensation rather than the severe pain that usually accompanies a more severe abrasion. An arc welder burn causes a punctate corneal keratopathy, and foreign body sensation may be a prominent symptom. A true conjunctival or corneal foreign body also may be present.

EXCESSIVE TEARING

Impairment of tear drainage can occur with an ectropion or any obstructions of the nasolacrimal drainage system. An entropion or abnormal lashes rubbing on the cornea (trichiasis) stimulate tear production.

EYELID TWITCHING

Any irritation of the conjunctiva or cornea can cause the eyelids to twitch. Occasional twitching of the lids usually is associated with stress or adrenergic stimulation. Benign essential blepharospasm is severe spasm of the lids leading to functional impairment. Multiple sclerosis (Chapter 411) also can cause lid spasm.

CONJUNCTIVITIS

Any ocular inflammation, including corneal ulcers, angle-closure glaucoma, endophthalmitis, and uveitis, can be associated with secondary

TABLE 423-4 DIFFERENTIAL DIAGNOSIS OF COMMON CAUSES OF INFLAMED EYE*

FEATURE	ACUTE CONJUNCTIVITIS	ACUTE IRITIS†	ACUTE GLAUCOMA‡	CORNEAL TRAUMA OR INFECTION
Incidence	Extremely common	Common	Uncommon	Common
Discharge	Moderate to copious	None	None	Watery or purulent
Vision	No effect on vision	Slightly blurred	Markedly blurred	Usually blurred
Pain	None	Moderate	Severe	Moderate to severe
Conjunctival injection	Diffuse: more toward fornices	Mainly circumcorneal	Mainly circumcorneal	Mainly circumcorneal
Cornea	Clear	Usually clear	Steamy	Change in clarity related to cause
Pupil size	Normal	Small	Moderately dilated and fixed	Normal or small
Pupillary light response	Normal	Poor	None	Normal
Intraocular pressure	Normal	Normal	Elevated	Normal
Smear	Causative organisms	No organisms	No organisms	Organisms found only in corneal ulcers related to infection

*Other less common causes of red eyes include endophthalmitis, foreign body, episcleritis, and scleritis.
†Acute anterior uveitis.
‡Angle-closure glaucoma.

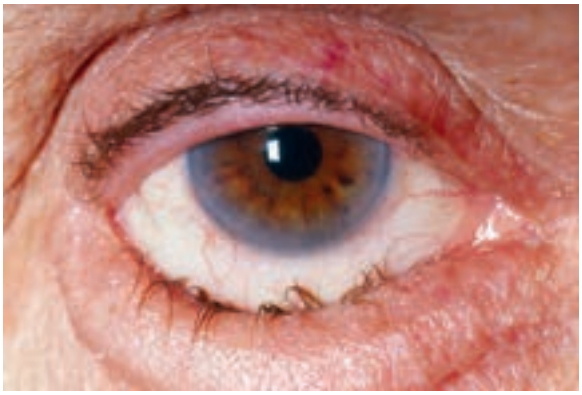

FIGURE 423-4. Involutional entropion. (From Palay DA, Krachmer JH. *Primary Care Ophthalmology,* 2nd ed. Philadelphia: Elsevier Mosby; 2005.)

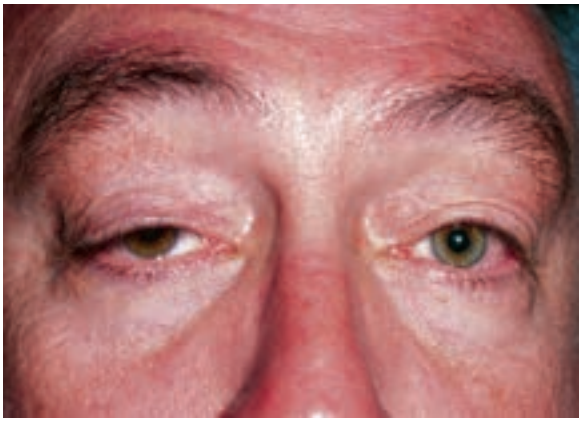

FIGURE 423-5. Ptosis of the right upper lid. (From Palay DA, Krachmer JH. *Primary Care Ophthalmology,* 2nd ed. Philadelphia: Elsevier Mosby; 2005.)

conjunctivitis. Conjunctivitis usually involves the entire conjunctiva, is associated with a discharge, and usually is not associated with pain (see Table 423-4).

PTOSIS (DROOPY EYELID)

Ptosis (Fig. 423-5) can be caused by a third nerve palsy, which usually is associated with diplopia and a reduction in elevation, depression, and medial movement of the pupil. Horner's syndrome (see Fig. 424-5) is associated with a small pupil. With myasthenia gravis (Chapter 422), other typical features of muscle weakness are usually present or can be elicited. Some patients have mild mechanical ptosis, especially after eye surgery.

PROPTOSIS (EXOPHTHALMOS)

Proptosis, or a prominent globe, can be a manifestation of thyroid abnormalities, especially Graves' disease (Chapter 226), in which proptosis is subacute and bilateral but sometimes asymmetrical (Fig. 423-6). An orbital pseudotumor can cause acute, usually unilateral proptosis, with severe pain, particularly with eye movement, and often with decreased vision. An optic nerve tumor causes chronic, unilateral proptosis associated with a slow onset of visual field loss. Acute cellulitis can be associated with unilateral proptosis, severe redness, and moderate to severe pain, commonly with sinusitis and an elevated white blood cell count.

SMALL PUPIL

A unilateral small pupil is best detected in dark conditions. Causes include Horner's syndrome, associated with ptosis on the same side; the bilaterally small, poorly reacting pupils of tertiary syphilis (Argyll Robertson pupils), which accommodate with normal constriction to a near object; miotic drops (e.g., pilocarpine); traumatic iritis; uveitis; and recent eye surgery.

LARGE PUPIL

Any α-adrenergic or anticholinergic agent placed into the eye can cause a large pupil. With eye trauma, the iris sphincter muscle can be damaged, and an abnormally large pupil can result. Tears in the iris sphincter can sometimes be appreciated on slit lamp examination. Third nerve palsy may cause a dilated pupil associated with ptosis and decreased elevation, depression, and medial eye movement. Adie's pupil (see Fig. 424-3) is an idiopathic, unilateral large pupil that is hypersensitive to weak cholinergic stimulation. Recent eye surgery, uveitis, closed-angle glaucoma, and traumatic iritis can cause a large pupil.

LEUKOKORIA

Leukokoria (white pupil) in a child is often a sign of retinoblastoma. However, any condition that changes transmission of ambient light to reflection of ambient light through the pupil may cause this sign. Some of the more common nonretinoblastoma conditions presenting with leukokoria (Table 423-5) include cataracts, retinal detachment, persistent hyperplastic primary vitreous (a developmental anomaly of the vitreous resulting in intraocular fibrosis and retinal detachment), Coats' disease (a developmental vascular malformation of the retina leading to retinal detachment), and ocular toxocariasis (a parasitic intraocular nematode infection, which leads to intraocular scarring and retinal detachment).

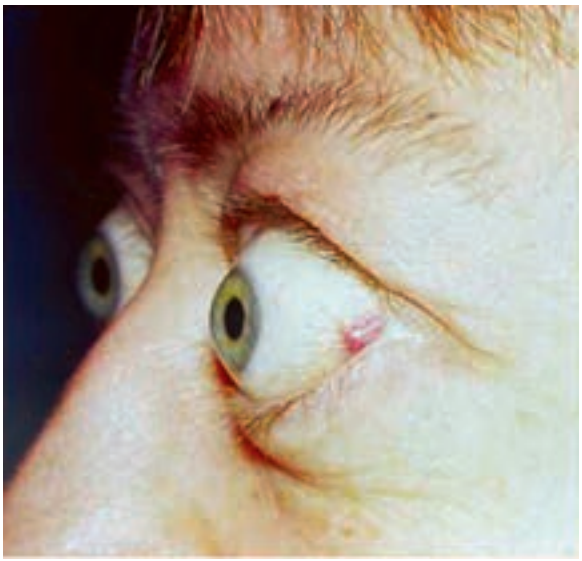

FIGURE 423-6. Graves' ophthalmopathy with characteristic exophthalmos and eyelid retraction.

TABLE 423-5 DIFFERENTIAL DIAGNOSIS OF LEUKOKORIA

Retinoblastoma
Cataract
Persistent hyperplastic primary vitreous
Retinopathy of prematurity (retrolental fibroplasia)
Coats' disease (retinal telangiectasia)
Retinal detachment
Toxocariasis
Familial exudative vitreoretinopathy (FEVR)

Eyelid Abnormalities

ECTROPION AND ENTROPION

An ectropion is an out-turning of the lower lid (Fig. 423-7), typically with the inner part of the lower end visible between the eye globe and the lid. Causes include aging, scarring, a mass on the lower lid, and seventh nerve palsy. Common symptoms include burning, itching, tearing, and the sense of a foreign body. Treatment is symptomatic, unless the underlying cause can be surgically corrected.

An entropion, which is an in-turning of the lower lid (see Fig. 423-4), is usually age-related and associated with irritation, burning, and a foreign body sensation. If it leads to trichiasis, in which the eyelashes rub or abrade the cornea, the lashes can be removed with forceps or surgery.

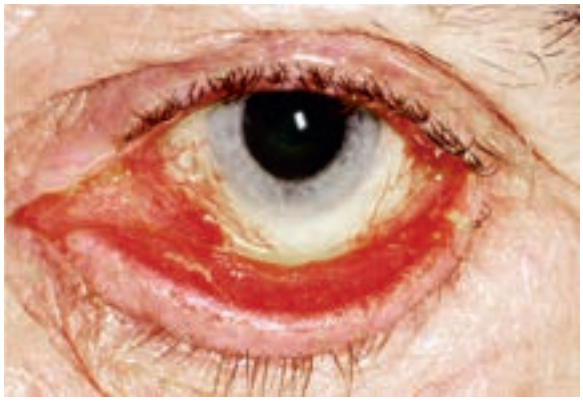

FIGURE 423-7. Involutional ectropion. (From Palay DA, Krachmer JH. *Primary Care Ophthalmology,* 2nd ed. Elsevier Mosby, Philadelphia, 2005.)

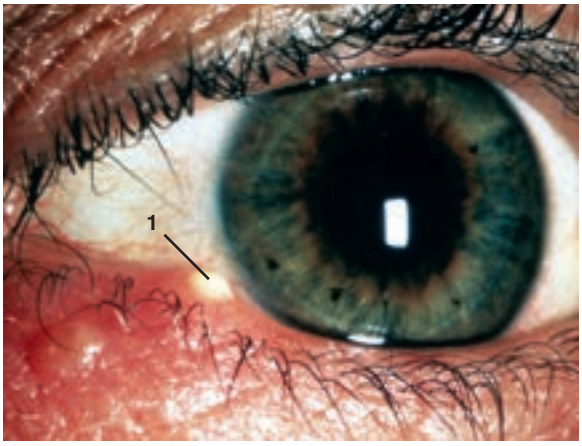

FIGURE 423-9. A lower lid stye *(1).* (From Palay DA, Krachmer JH. *Primary Care Ophthalmology,* 2nd ed. Philadelphia: Elsevier Mosby; 2005.)

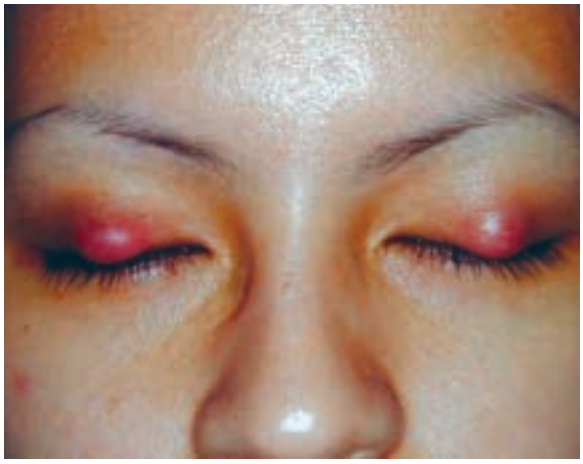

FIGURE 423-8. Bilateral chalazion in the upper eyelids.

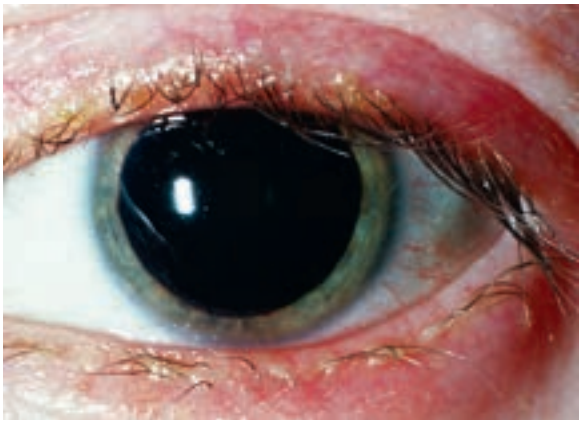

FIGURE 423-10. Staphylococcal blepharitis. The lid margins are very red and under high magnification demonstrate tiny ulcerations. (From Palay DA, Krachmer JH. *Primary Care Ophthalmology,* 2nd ed. Philadelphia: Elsevier Mosby; 2005.)

CHALAZION

A chalazion (Fig. 423-8) is a localized lipogranulomatous inflammation that results when the eyelid reacts to the contents of a ruptured sebaceous (meibomian) gland. The retained, lipid-rich sebaceous material acts as a foreign material that stimulates a lipogranulomatous foreign body inflammatory reaction. A painless or slightly tender, poorly demarcated, nonmobile nodule forms under the eyelid skin. Most lesions resolve over days to weeks with warm compresses or without specific treatment. Occasionally, an ophthalmologist may inject steroids to reduce inflammation or debulk the foreign material by incision and drainage through the tarsal conjunctiva. Some individuals may have recurrent chalazion.

HORDEOLUM (STYE)

A hordeolum (stye) (Fig. 423-9) is an extremely painful abscess in a hair or eyelash follicle or in a sebaceous gland. Styes are usually self-limited infections that respond to warm compresses and topical antibiotics (e.g., bacitracin or erythromycin ointment or moxifloxacin or gatifloxacin drops). An ophthalmologist may perform incision and drainage if symptoms do not improve within 48 hours.

BLEPHARITIS

Blepharitis (Fig. 423-10), which is a nonspecific inflammation of the eyelid skin, is common, particularly in men. The condition is usually bilateral and symmetrical. Rosacea (Chapter 439) is the most common associated cutaneous condition, and *Staphylococcus aureus* is the most common infectious agent. If untreated, blepharitis becomes chronic and may lead to corneal and conjunctival inflammation (blepharoconjunctivitis). Ophthalmic antibiotic ointment (e.g., bacitracin or erythromycin) is more efficacious than eye drops, but systemic antibiotics (e.g., minocycline, 50 to 100 mg, or doxycycline, 100 mg, once daily; tetracycline, 250 mg twice daily; or erythromycin, 250 mg three times daily) are recommended if there is any evidence of inflammation of the cornea or conjunctiva.

In seborrheic blepharitis, exfoliated keratinous debris accumulates along the eyelid margin, particularly at the follicles of the eyelashes, and irritates the conjunctiva. Treatment of this chronic condition is directed at mechanically removing the keratinous debris by scrubbing the eyelid and eyelashes daily with a mild detergent ("baby shampoo") in warm water applied with a soft cloth.

BENIGN EYELID NEOPLASMS

Skin tags, also known as squamous papillomas, are the most common benign skin lesions. Other skin lesions include seborrheic keratitis, actinic keratitis, inverted follicular keratitis, and benign lesions of the eccrine and apocrine systems. Most of these benign lesions are cured by simple excision.

SEBACEOUS CARCINOMA

Sebaceous carcinoma originates from sebaceous glands either in the tarsal plate (meibomian gland) or associated with eyelashes (glands of Zeis) and is capable of producing widespread metastasis resulting in death.[4] Muir-Torre syndrome is a syndrome of sebaceous tumors associated with visceral malignancy. Except for chronic, unilateral blepharitis, owing to the peculiar manner of spread of this tumor in the plane of the skin epithelium (pagetoid spread) without causing the formation of nodules, few symptoms occur early in the course of the disease. The tumor may progress to involve the tarsal conjunctiva, the bulbar conjunctiva, and even the corneal epithelium. A characteristic sign is regional loss of eyelashes. When the mass thickens, it may have the appearance of a chalazion, and a history of multiple chalazia in the same region of the eyelid is suggestive of sebaceous carcinoma.

Treatment is surgical removal. Surgical margins are difficult to estimate because of the intraepithelial extension of the tumor. Topical mitomycin C has been suggested as treatment for pagetoid invasion of the conjunctiva. In

advanced cases, removal of the eyelids, eye, and orbital contents (exenteration) may be necessary.

BASAL CELL CARCINOMA

Basal cell carcinoma (Fig. 423-11), which originates from the basal cell layer of the epithelium, is a common cutaneous malignancy (Chapter 203). The lesion, which usually is asymptomatic, is often a well-demarcated, elevated nodule that may have a central region of ulceration and fine cutaneous vascular channels (telangiectasias). A common benign cutaneous lesion, sometimes confused clinically with basal cell carcinoma, is seborrheic keratosis (Chapter 440), which tends to be soft and appear hyperpigmented; the most common site is the lower eyelid, especially in the nasal quadrant. Basal cell carcinoma, particularly near the medial canthus, may extend posteriorly into the soft tissues of the orbit. Imaging before surgical excision for medial canthal lesions may be necessary to determine the true extent of the tumor. Basal cell carcinoma is treated by surgical excision, using Mohs' technique with intraoperative histologic evaluation to determine adequate margins of excision, if possible.[5]

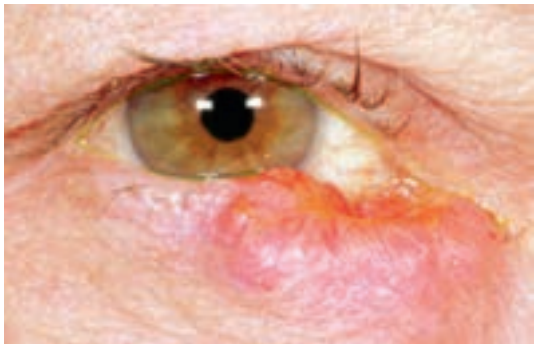

FIGURE 423-11. A typical nodular basal cell carcinoma. (From Palay DA, Krachmer JH. *Primary Care Ophthalmology,* 2nd ed. Philadelphia: Elsevier Mosby; 2005.)

Metastasis is extremely rare. With early detection and adequate excision of the local lesion, the prognosis is excellent.

EYELID SQUAMOUS CELL CARCINOMA

Eyelid squamous carcinoma, which is much less common than basal cell carcinoma, arises from the surface squamous epithelium. Ultraviolet light exposure is the major risk factor. In contrast to basal cell carcinoma, squamous cell carcinoma can metastasize, most often to regional lymph nodes. Treatment is surgical excision. Except in rare circumstances, such as in immunosuppressed patients or patients with xeroderma pigmentosa, the prognosis is excellent.

Ocular Surface Abnormalities

DRY EYES

Even minor disturbances in the tear film can cause itching, burning, pulling, and transient changes in vision. Dry eyes that cause conjunctival hyperemia without purulent discharge particularly disturb some patients. Paradoxically, decreased tearing can result in irritation and secondary increased (reflux) tearing.

Most daily tear production is not by the lacrimal gland but by small collections of lacrimal glands, mucus-producing glands, and sebaceous glands located throughout the conjunctiva, eyelid, and anterior orbital soft tissue. Over time, particularly in women, production of tear film diminishes. Because tear film production is lower during sleep, patients often note symptoms on awakening followed by slow resolution over minutes or hours. Wind and low humidity environments, such as in commercial airliners, can exacerbate symptoms. The reduction in aqueous components of tears is often associated with a compensatory increase in mucus production, which tends to blur vision until the patient blinks or uses supplemental tears. These symptoms are particularly prominent in persons who have rheumatoid arthritis (Chapter 264), Sjögren's syndrome (Chapter 268), Stevens-Johnson syndrome (Chapter 440), and ocular cicatricial pemphigoid (Fig. 423-12).

Treatment is not definitive and is rarely satisfactory. No medication increases the production of tears. Low-viscosity artificial tears (e.g., polyethylene glycol 400 0.4%), which do not tend to blur vision but have a short duration of action, are best used during visually important tasks.

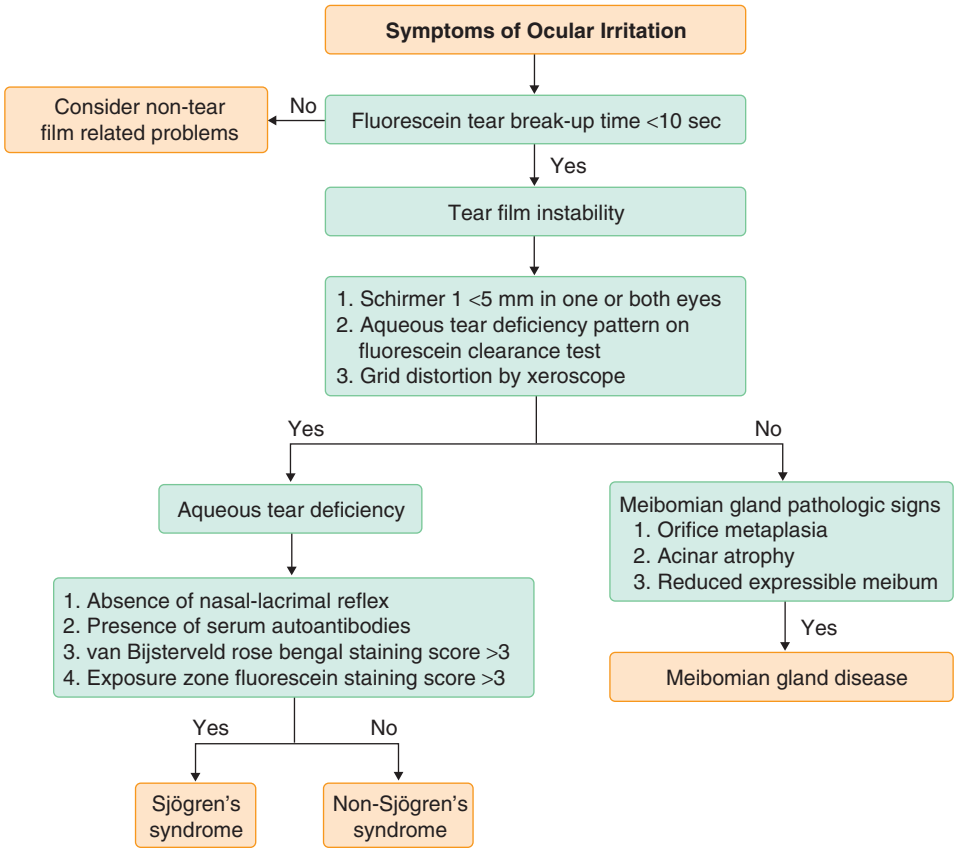

FIGURE 423-12. Diagnostic algorithm for ocular irritation. (Modified from Pflugfelder SC, Tseng SC, Sanabria O, et al. Evaluation of subjective assessments and objective diagnostic tests for diagnosing tear-film disorders known to cause ocular irritation. *Cornea.* 1998;17:38-56.)

High-viscosity tears (e.g., carboxymethylcellulose sodium) have a longer duration of action but tend to blur vision; they are best used at bedtime to maintain lubrication of the ocular surface during sleep. When artificial tears do not control symptoms, occlusion of the nasolacrimal duct with synthetic plugs or permanent surgical occlusion tends to retain the tears that are produced. Anti-inflammatory drugs (e.g., cyclosporine 0.05% drops, every 12 hours indefinitely) can preserve glandular tissue that may be affected by local inflammation. For patients with systemic disease associated with dry eyes, effective treatment of the systemic disease sometimes improves the eye abnormalities.

PINGUECULA AND PTERYGIUM

A pinguecula (Fig. 423-13) consists of a limbal (at junction of cornea and sclera) and bulbar conjunctival degenerative process caused by ultraviolet light damage to the subepithelial tissue. It is very common and rarely causes symptoms. If the supportive tissue degeneration extends into the cornea, it becomes a pterygium (Fig. 423-14), which may cause visual changes and require surgical excision. About 2 to 10% with a pterygium have a coexisting squamous carcinoma, which often is clinically unsuspected and diagnosed only by histopathologic examination.[6]

RECURRENT EROSION

Recurrent erosion of the cornea usually is a delayed reaction to minor traumatic corneal abrasion. The abrasion heals abnormally, resulting in a weakness of the epithelial attachment to underlying tissue. Weeks to months to years later, the patient is awakened in the middle of the night with extreme ocular pain on opening the eyelids. The epithelium has become "stuck" to the overlying upper lid and is mechanically abraded. The condition is treated with hyperosmotic drops and ointment. There is a tendency to recurrence.

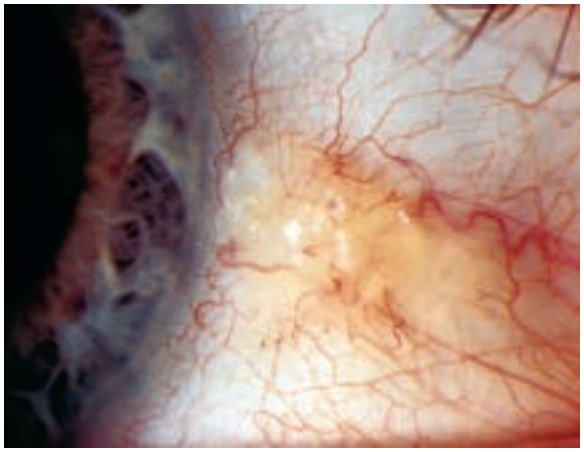

FIGURE 423-13. Pinguecula. These lesions are found at the 3-o'clock and 9-o'clock positions and are extremely common, especially in older patients. (From Palay DA, Krachmer JH. *Primary Care Ophthalmology,* 2nd ed. Philadelphia: Elsevier Mosby; 2005.)

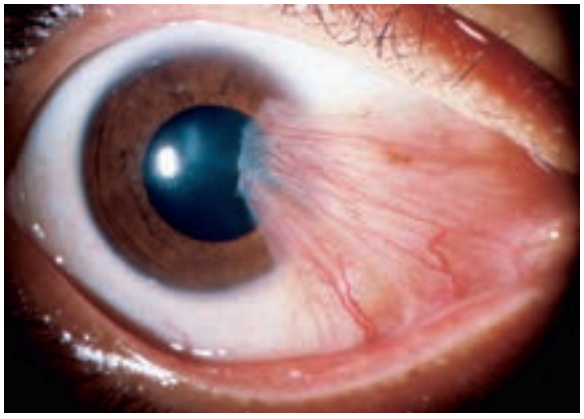

FIGURE 423-14. Pterygium. These lesions are found in the horizontal meridian, most common nasally. (From Palay DA, Krachmer JH. *Primary Care Ophthalmology,* 2nd ed. Philadelphia: Elsevier Mosby; 2005.)

ACCIDENTAL TRAUMA

With ocular trauma, many tissues of the eye can be easily disrupted, and the effects of trauma may not be manifest for months or even years after the episode of trauma. If the traumatic episode disrupts the eye wall (cornea and sclera), surgical repair is necessary, usually urgently. If the eye wall is intact, surgical treatment is often not necessary, at least initially.

CORNEAL ABRASION

Shearing trauma or hypoxia associated with overwearing of contact lenses may cause corneal abrasion, one of the most common forms of ocular injury. Symptoms are often intense and intolerable. Healing (i.e., re-epithelialization) of the cornea occurs within 24 to 48 hours. Rust from metallic fragments is toxic to the epithelium and should be removed. Fungal keratitis may complicate injuries from fingernails or vegetable matter, such as tree branches. Treatment usually consists of a topical antibiotic (e.g., erythromycin ointment, four times daily for 10 days) to prevent bacterial keratitis. Subsequent scarring usually does not occur unless deeper structures, such as Bowman's membrane, are affected. Topical anesthetics never should be prescribed to control pain because they increase the risk for microbial keratitis and scarring and may delay healing.

MAJOR OCULAR TRAUMA

Hyphema (Fig. 423-15) is hemorrhage into the anterior chamber caused by blunt trauma. If the patient is in an intensive care unit and supine, the blood will distribute uniformly over the iris to cause the appearance of increased pigmentation of the iris (heterochromia iridis). If the patient has been sitting, the blood may settle by gravity to form an aqueous-blood interface with the blood in the dependent portion of the anterior chamber. Hyphema, which is a sign of serious ocular damage, may lead to secondary glaucoma and blood staining of the cornea. It requires prompt evaluation by an ophthalmologist.

The most common site of rupture of a globe is at the limbus (junction of cornea and sclera), where a pigmented mass may be noted. The mass may be either a blood clot or an anteriorly displaced uveal tract (usually iris). Any manipulation of the globe may force the remaining intraocular tissue through the wound and may make the injury irreparable. Surgical repair is usually indicated.

Cataract and retinal detachment are not common except in severe accidental trauma. A unilateral cataract or unilateral glaucoma may occur decades after the injury, even when an injury is too minor to be recalled. Traumatic cataract and traumatic glaucoma are treated in the same manner as other forms of these conditions.

INFLAMMATORY EYE DISORDERS

Uveitis

Inflammation of any part or parts of the uveal tract (iris, ciliary body, and choroid) may be called anterior or posterior uveitis, iritis, iridocyclitis, or choroiditis. Symptoms include a red eye (see Table 423-4), decreased vision, and photophobia. The inflammation is chronic, and a cause is rarely found. However, uveitis accompanies many autoimmune diseases, often without correlation with the activity of the systemic inflammation. Anterior

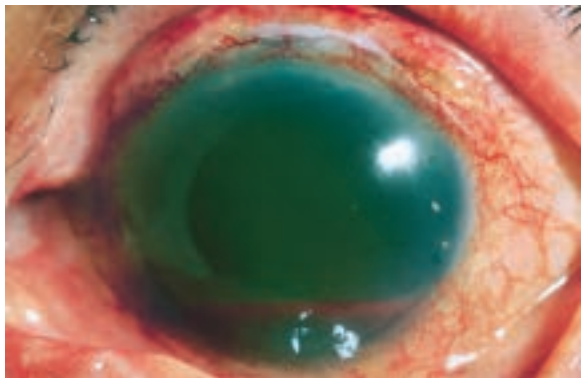

FIGURE 423-15. Hyphema following cataract surgery. (Courtesy of Dr. Myron Yanoff.)

uveitis or conjunctivitis is nearly universal in patients with reactive arthritis (Chapter 265). About 25% of patients with ankylosing spondylitis (Chapter 265) develop acute, recurrent anterior uveitis. Two to 12% of patients with inflammatory bowel disease (Chapter 141) develop anterior uveitis, which is also common with psoriatic arthritis but not with psoriasis alone (Chapters 265 and 438). Treatment with topical corticosteroids (e.g., prednisolone acetate 1%, one drop in the affected eye or eyes every 1 to 6 hours while awake) is usually sufficient to control the ocular disease.

Endophthalmitis

Endophthalmitis is extensive inflammation within the eye from any cause. Most cases of endophthalmitis involve a breach in the eye wall (cornea and sclera), associated with either accidental trauma (incidence of approximately 5%) or surgical procedures (incidence of approximately <0.03%).[7] The initial symptom is usually decreased vision followed by dull ocular pain. The initial sign is often evidence of inflammatory cells either within the aqueous (anterior uveitis) or within the vitreous (vitreitis). The cells can be seen only by slit lamp biomicroscopy. Common microbial organisms include toxin-producing gram-positive species and gram-negative species that are often associated with rapidly destructive course. Other organism of relatively low virulence, *Propionibacterium acnes* and *Staphylococcus epidermidis*, follow a more indolent course with less potential destruction. Metastatic endophthalmitis infection from a primary source outside of the eye is an unusual cause.

Diagnosis is established by sampling anterior chamber fluid or preferably vitreous fluid (vitreous tap) and evaluation of that fluid by Gram stain and culture. Prophylaxis against endophthalmitis includes preoperative topical instillation of povidone-iodine and intracameral antibiotic injection at the end of cataract surgery.

Systemic antibiotics are not effective, and patients benefit from vitrectomy only if the initial vision of the affected eye is light perception or worse. In severe cases, a vitrectomy through a pars plana incision reduces the microbial and inflammatory debris burden. Initial treatment is with topical antibiotics (e.g. commonly used but not limited to, vancomycin 1 mg/0.1 mL and ceftazidime 2.25 mg/0.1mL) and corticosteroids. Under certain circumstances, simultaneous intraocular antibiotics (vancomycin 1 mg and ceftazidime, 2.25 mg/0.1mL) minimize the destructive effects of retinal inflammation.

Allergic Conjunctivitis

Allergic conjunctivitis (Table 423-6) is commonly associated with atopy, hay fever, and allergic rhinitis (Chapter 251). Itching, a foreign body sensation, and a watery discharge are common. Treatment includes cool compresses and topical vasoconstrictors or antihistamines (e.g., naphazoline drops, four times daily during the allergic season, or levocabastine drops, four times daily). Long-term treatment with mast cell stabilizers (e.g., pemirolast drops, four times daily during the allergic season) or the combination of an antihistamine plus a mast cell stabilizer (e.g., olopatadine drops, two times daily during the allergic season) can be extremely effective in treating chronic symptoms.

INFECTIOUS EYE DISORDERS

Cellulitis

Preseptal cellulitis (Fig. 423-16) is soft tissue inflammation of the eyelid anterior to the orbital septum. The orbital septum divides the soft tissues of the eyelid from the soft tissues of the orbit. Orbital tissue is more susceptible to damage by the inflammation than is the preseptal tissue.

The clinical signs of preseptal cellulitis include soft tissue swelling, hyperemia, and conjunctival chemosis (edema). Movement of the eye is not restricted. Extension of inflammation posterior to the orbital septum is indicated by proptosis of the globe and ophthalmoplegia (restricted motion). Treatment of preseptal cellulitis includes oral antibiotics (e.g. commonly used but not limited to, amoxicillin-clavulanate, 500 mg orally every 8 hours for 10 days, or Bactrim 500 mg/PO, BID for cases of suspected Methicillin-resistant Staphylococcus aureus (MRSA)). Treatment of orbital cellulitis, which can lead to septic optic neuritis, intracranial spread, and cavernous sinus thrombosis, may require intravenous antibiotics and surgical drainage of a paraorbital abscess.

Adenoviral Conjunctivitis

Viral conjunctivitis (Fig. 423-17) is common (see Table 423-6), and adenoviral conjunctivitis (especially subtypes 7, 11, and 18) is the most common type.[8] The condition is highly contagious through direct contact or inhalation of respiratory particles. After an incubation period of 5 to 15 days, the patient presents with very red eyes (see Table 423-4), itching, burning, a foreign body sensation, and often a discharge and ocular discomfort (see Table 423-3), which persist for 5 to 15 days. Preauricular lymphadenopathy may be present,

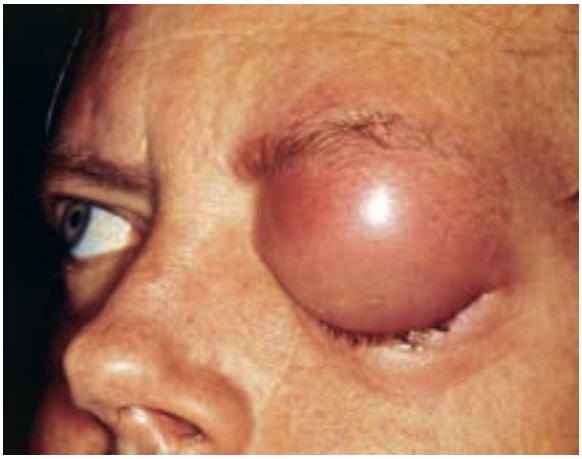

FIGURE 423-16. Eyelid abscess. Preseptal cellulitis, commonly resulting from minor penetrating trauma, may evolve into an abscess. Treatment requires incision and drainage followed by systemic antibiotics.

TABLE 423-6 OPHTHALMIC DISORDERS ASSOCIATED WITH CONJUNCTIVITIS

DISORDER	ACUTE OR CHRONIC	UNILATERAL OR BILATERAL	KEY SYMPTOMS	DEGREE OF INJECTION	DISCHARGE TYPE	OTHER FEATURES
Viral conjunctivitis	Acute	Bilateral, possibly asymmetrical	Itching, burning, soreness	4+	Watery	Preauricular lymphadenopathy
Bacterial conjunctivitis	Acute	Unilateral or bilateral	Burning	3+	Heavy, mucopurulent	Lids possibly adherent
Chlamydial conjunctivitis	Subacute, chronic	Usually unilateral	Burning, irritation	2+	Scant, mucopurulent	Usual occurrence in young, sexually active adults
Herpes simplex conjunctivitis	Acute	Unilateral	Photophobia, irritation	1-2+	None	Dendritic ulcer on the cornea or vesicles on the lid possible
Allergic conjunctivitis	Chronic	Bilateral	Itching	2+	Stringy, mucoid	Usual occurrence in atopic persons, possible seasonal symptoms
Blepharitis	Chronic	Bilateral	Itching, burning, foreign body sensation	1-2+	Usually none	Inflammation and crusting of lid margins
Dry eye	Chronic	Bilateral	Foreign body sensation	1+	Mucoid in severe cases	Punctate fluorescein staining of the cornea

Adapted from Palay DA, Krachmer JH. *Primary Care Ophthalmology,* 2nd ed. Philadelphia: Elsevier Mosby; 2005.

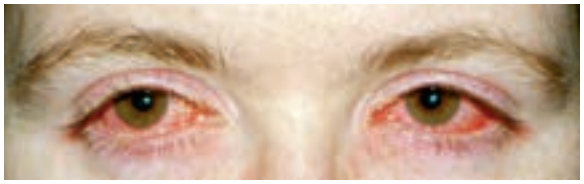

FIGURE 423-17. Diffuse injection of the conjunctiva with a watery discharge is evident in this case of viral conjunctivitis. (From Palay DA, Krachmer JH. *Primary Care Ophthalmology,* 2nd ed. Philadelphia: Elsevier Mosby; 2005.)

and a history of upper respiratory tract infection is common. The disease is self-limited, and treatment is aimed at patients' comfort. Cool compresses are often soothing. Patients are advised to wash their hands frequently. Topical antibiotics are not required, and topical corticosteroids are contraindicated.

Bacterial Conjunctivitis

Fewer than 5% of cases of conjunctivitis are caused by bacteria, mostly *Staphylococcus, Haemophilus,* or *Streptococcus* species. Patients have a mucoid or purulent discharge (Fig. 423-18), often with crusting and edema of the conjunctiva (chemosis) and lids. Bacterial conjunctivitis responds to broad-spectrum antibiotic solutions or ointments (e.g., topical erythromycin ointment three times daily for 2 weeks) (Table 423-7).

Chlamydial Conjunctivitis

Adult inclusion conjunctivitis is a chronic conjunctivitis caused by sexual transmission of *Chlamydia trachomatis* (Chapter 318). Patients often have preauricular lymphadenopathy. Oral erythromycin (500 mg orally, four times daily for 7 days) or azithromycin (1 g orally twice daily for 7 days) is required. Trachoma, which is a chronic cicatricial conjunctivitis after repeated chlamydial infection (Chapter 318), is the world's leading cause of corneal blindness. It causes an entropion, inversion of the eyelashes (trichiasis), corneal vascularization, and opacification. Topical erythromycin or tetracycline, twice daily for 3 to 4 weeks, can be effective, but surgical epilation or eyelid reconstruction may be required.

Herpes Simplex Keratitis

Herpes simplex keratitis is the most common cause of central corneal ulcer (Fig. 423-19).[9] Herpes simplex virus also can cause vesicular eyelid dermatitis. Initially, the main signs of primary herpes simplex keratitis are a red eye and a corneal epithelial dendritic ulcer. With appropriate antiviral therapy (e.g., ganciclovir 0.15% ophthalmic gel five times daily for at least 1 week or until healed), the keratitis usually heals without scarring. Recurrent herpes simplex keratitis may be precipitated by fever, menses, sunlight, irradiation, or stress. With recurrence, the disease may extend into the corneal stroma and cause a red eye, ocular discomfort, blurred vision, and corneal scarring. The treatment of stromal involvement is multifactorial and may not be successful. Corneal transplantation may be needed.

Herpes Zoster Ophthalmicus

Herpes zoster ophthalmicus (shingles, Chapter 375) has a propensity to involve one or more branches of the trigeminal nerve. The virus also can affect the uveal tract and, in immunosuppressed patients, the retina (i.e., acute retinal necrosis). When the trigeminal nerve is involved, spread to the inside of the eye (uveitis) is most likely if vesicles are present in the inner corner of the eyelids or on the nose, especially the tip of the nose. If the uvea is not involved, the skin lesions heal with some scarring but no long-term effects. In patients with moderate to severe skin involvement, treatment can be started with oral acyclovir (800 mg orally five times per day for 7 to 10 days). If uveitis develops, the treatment (e.g., prednisolone acetate drops 1% four times daily and atropine 1.0% once or twice daily) can be extended and difficult.

Pseudomonal and Gonococcal Keratitis

Keratitis, which is inflammation of the corneal stroma, can be caused by spread of pathogens internally from a corneal ulcer. *Pseudomonas aeruginosa* (Chapter 306), which causes a particularly virulent keratitis, is the most common gram-negative pathogen and is especially common in wearers of contact lenses. To avoid internal spread, urgent treatment is necessary (e.g., fortified tobramycin [9 mg/mL] or gentamicin [1.5 mg/mL], every hour, alternating with fortified cefazolin, 50 mg/mL every hour, so a treatment is given each one-half hour around the clock). The dosage and duration of the treatment depend on the response.

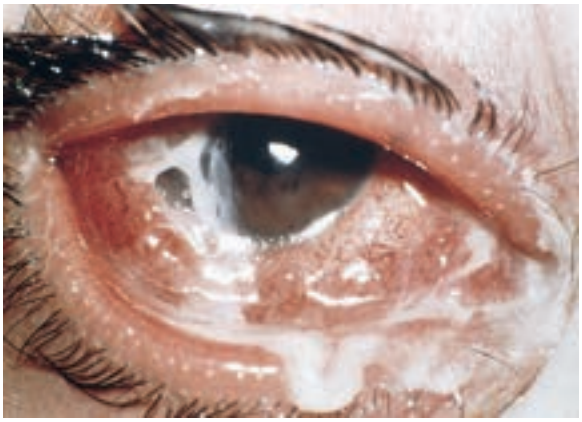

FIGURE 423-18. Bacterial conjunctivitis. Purulent discharge and conjunctival hyperemia suggest bacterial conjunctivitis. Viral conjunctivitis produces watery discharge, foreign body sensation, preauricular lymphadenopathy, and conjunctival follicles seen on slit lamp examination. (Reproduced with permission from the American Academy of Ophthalmology.)

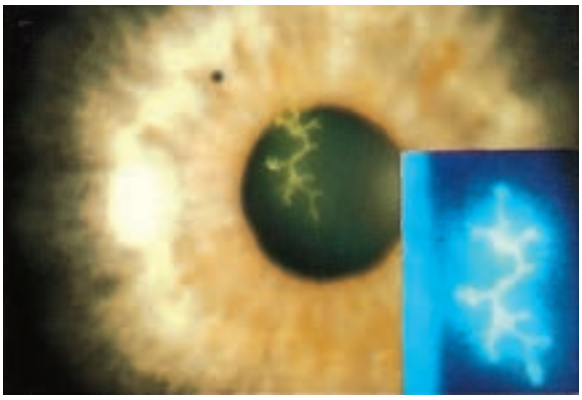

FIGURE 423-19. Herpes simplex corneal epithelial keratitis in diffuse light and in light passed through a cobalt blue filter after fluorescein staining *(inset)*. **Note the dendritic staining pattern characteristic of herpes simplex.**

TABLE 423-7 TOPICAL ANTIBIOTICS FOR EYE INFECTIONS

DRUG	TYPE	CONCENTRATION	DOSE
Moxifloxacin	Drops	0.5%	1 drop BID × 7 days
Gatifloxacin	Drops	0.5%	1 drop q2H × 24 hours; then QID × 6 days
Ciprofloxacin	Drops	0.3%	1 drop q2H × 48 hours; then q4H × 5 days
Gentamicin	Drops	0.3%	1 drop 4 times daily
Ofloxacin	Drops	0.3%	1 drop 2H × 48H; then q2H × 5 days
Bacitracin	Ointment	500 U/g	Put in eye, for several days
Tobramycin	Ointment	0.3%	Put in eye, for several days
Erythromycin	Ointment	0.5%	Put in eye, for several days

Another gram-negative cause of a virulent keratitis is *Neisseria gonorrhoeae* (Chapter 298). Corneal infection is accompanied by copious tearing and a characteristic hyperpurulent discharge. Prompt treatment is essential in preventing corneal perforation.

Cytomegalovirus Retinitis

Cytomegalovirus retinitis (Chapter 376) is unusual except in immunosuppressed patients, especially patients with human immunodeficiency virus infection. Clinically, a central retinochoroiditis is seen. The presumptive diagnosis is made on the characteristic intense, retinal, wedge-shaped reaction,

with considerable exudates and hemorrhages, giving the terms "pizza pie retinitis" and "hemorrhagic cottage cheese retinitis" to the entity. Treatment is antiviral drugs: ganciclovir (5 mg/kg intravenously twice daily, two to three times per week), foscarnet (90 mg/kg intravenously twice daily, twice per week), or cidofovir (5 mg/kg intravenously, weekly for 3 weeks) with follow-up maintenance. Intravitreal injections using an appropriately reduced dose is also commonly used in selected cases.

Acanthamoeba Keratitis

Acanthamoeba species (Chapter 352) can cause a severe, blinding keratitis. Contact lens wearing is a major risk factor. A characteristic stromal ring infiltrate develops, and uveitis may occur. Using a confocal microscope, the acanthamebic parasite can be observed clinically as a pear-shaped cyst (11 to 15 μm). Numerous treatment protocols exist (e.g., polyhexamethyl biguanide 0.02% drops every hour).[10] The duration and dose depend on the response, but the ideal treatment is not yet in hand. Corneal transplantation may be necessary in cases of severe corneal scarring.

Toxoplasmic Retinitis

Toxoplasma gondii (Chapter 349) causes both a congenital and acquired retinochoroiditis, which is more common in immunosuppressed patients. The lesions begin as an acute retinitis that atrophies centrally and pigments peripherally as it heals. The protozoa are found both in free and encysted forms within the retina. The condition may be self-limited and diagnosed as a healed incidental finding that does not need treatment. Standard treatment of vision-threatening toxoplasmosis remains controversial.[11] When active lesions are in the macula or a severe vitreitis causes at least a two-line decrease in vision, 4 to 6 weeks of quadruple therapy (pyrimethamine, 200 mg oral loading dose then 25 mg orally daily; folinic acid, 10 mg orally every other day; sulfadiazine, 2 g oral loading dose then 1 g four times daily; and oral corticosteroids, e.g., prednisone, 20 to 60 mg orally daily beginning at least 24 hours after antibiotic therapy is started and tapered 10 days before stopping antibiotics) usually produces good results. Alternative regimens may include clindamycin (150 to 450 mg orally three to four times daily) or atovaquone (1 g oral loading dose then 500 mg daily).

Fungal Endophthalmitis

Fungal endophthalmitis is infrequent (7% of microbial endothalmitis)[12] but is a potentially disastrous infection of the inside of the eye, often leading to blindness. The primary organisms are *Candida, Coccidioides,* and *Aspergillus* species, which can gain access inside the eye either by traumatic introduction or through hematogenous spread. The patient presents with a red eye, ocular pain, and decreased vision. Multiple vitreous abscesses tend to be caused by fungi, whereas a solitary abscess is more likely caused by bacteria. Usually a vitreous "tap" (biopsy) is performed for culture and to guide therapy. Vitrectomy and intraocular fungal agents are indicated in advanced cases.

Tuberculosis

About 1% of patients with pulmonary tuberculosis (Chapter 324) have uveal involvement, usually as iridocyclitis or diffuse choroiditis. Painless progressive visual loss is the most common symptom. Small yellow choroidal lesions may be seen, and retinal periphlebitis may occur secondarily. Treatment is as for the primary disease.

Syphilis

About 5% of patients with secondary syphilis (Chapter 319) develop anterior uveitis or neuroretinitis. In tertiary syphilis, the miotic Argyll Robertson pupil reacts poorly to light but briskly to accommodation. Treatment is as for the systemic disease.

STRUCTURAL AND AGE-RELATED DISORDERS

Cataract

A cataract is an opacification of the crystalline lens. The lens doubles in volume between birth and age 70 years as new lens "fiber cells" are laid down on the external aspect of the lens cortex, beneath the lens capsule. The older fibers in the center of the lens cannot be desquamated into the surrounding aqueous and thus are compressed into the center of the lens. At birth, the lens is pliable and totally transparent. By age 45 years, the lens loses its pliability, which compromises near vision. As the process progresses, the lens loses its transparency, beginning at the center of the lens (nuclear sclerosis). The concurrent change in density of the lens nucleus may alter the optical characteristics of the eye to cause acquired nearsightedness ("second sight"). Ultimately, the cataract may become so dense that cataract surgery is necessary to restore vision.

Symptoms are typically loss of vision, especially at night, and glare. Cataract surgery, performed as an outpatient, is elective and depends on how much the decreased vision interferes with the normal lifestyle of the patient. A synthetic intraocular lens implant is inserted into the eye during surgery. Prognosis for restoration of vision is excellent, depending on the function of the retina. In general, cataracts develop asymmetrically. The worst eye (vision-wise) should have surgery first. As the second eye's cataract worsens, decreasing vision and monocularity are indications for cataract surgery in the second eye.

Glaucoma

Glaucoma results from an imbalance between the production of aqueous fluid and the drainage of aqueous fluid. The aqueous is produced by the nonpigmented ciliary epithelium of the pars plicata of the ciliary body. Aqueous fluid leaves the eye through the trabecular meshwork into the venous circulation. If the drainage function does not match the production potential, the intraocular pressure increases. If the elevated intraocular pressure is high enough or is present long enough, ganglion cells in the retina are damaged, causing loss of their axons. Loss of axons can best be appreciated clinically at their normal exit from the eye, the optic disc. Bulk loss of axons will lead to enlargement of the optic cup, which is recorded as increase in the cup-to-disc ratio.

In open-angle glaucoma, there is apparent free anatomic access to the trabecular meshwork. In closed-angle glaucoma, there is a relative or absolute anatomic barrier to the flow of aqueous.

CHRONIC OPEN-ANGLE GLAUCOMA

The most common type of glaucoma in elderly people is chronic open-angle glaucoma. The first symptom is loss of peripheral visual field with retention of central visual function. The visual field may be reduced considerably before the patient notes loss of function. Most cases of chronic open-angle glaucoma are identified during routine eye examinations, either by discovery of abnormally high intraocular pressure[13] or by the presence of a high cup-to-disc ratio (Fig. 423-20). Average intraocular pressure is generally at or below 21 mm Hg, but exceptions exist depending on corneal thickness (causing artifacts of measurement in patients with excessively thin or thick corneas) and genetic disposition. The diagnosis of glaucoma is confirmed by characteristic visual field loss as determined by automated perimetry.

The treatment goal is to reduce intraocular pressure, initially with pharmacologic agents: β-blockers (e.g., betaxolol drops 0.5% twice daily), carbonic anhydrase inhibitors (e.g., dorzolamide drops twice daily), α-agonists (e.g., apraclonidine drops twice daily), and antiprostaglandins (e.g., latanoprost drops twice daily).[14] Generally, the drops are taken for a lifetime. Applying energy to the structures of the trabecular meshwork with a laser (laser trabeculoplasty) often results in years of control of intraocular pressure. In resistant cases, mechanical filtration is accomplished surgically by bypassing the trabecular meshwork either by creating a fistula (trabeculectomy) between the anterior chamber and the episcleral tissue or by implanting a synthetic filtration device (a tube-shunt) from the anterior chamber through the sclera into a collection reservoir located at the equator of the eye in the soft tissues of the orbit.

PSEUDOEXFOLIATIVE GLAUCOMA

Pseudoexfoliative glaucoma syndrome is a genetically determined biochemical abnormality of the basement membrane protein, fibrillin. The syndrome occurs among people throughout the world but is especially prominent in Scandinavians and Saudi Arabians. Affected individuals are identified by accumulation of abnormal fibrillin (exfoliative material) on the surface of the crystalline lens, most easily seen in the pupillary space. Pseudoexfoliative glaucoma increases the risk for developing open-angle glaucoma five-fold, to a lifetime risk of about 10%. Treatment is as for open-angle glaucoma.

ANGLE-CLOSURE GLAUCOMA

Angle-closure glaucoma (Fig. 423-21) may occur over a short period of time and cause extreme debilitating symptoms.[15] Alternatively, symptoms may develop over a long period of time with few specific symptoms.

The risk factors for angle-closure glaucoma are based on the anatomic configuration of the components of anterior chamber. Persons who are

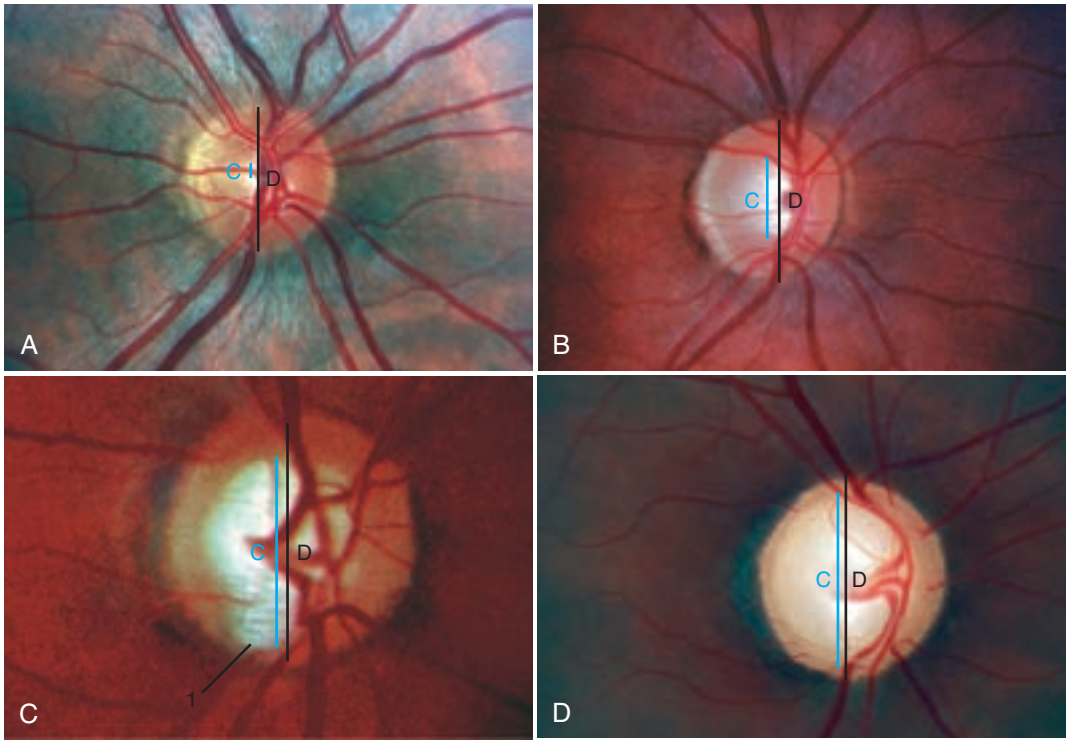

FIGURE 423-20. Cup-to-disc ratios. A, Normal cup-to-disc (C/D) ratio of 0.1. **B,** Likely normal C/D ratio of 0.5. **C,** C/D ratio of 0.8 vertically with inferior notching (1) of the nerve (glaucomatous change). **D,** C/D ratio of 0.90 vertically (glaucomatous change). C = cup; D = disc. (From Palay DA, Krachmer JH. *Primary Care Ophthalmology,* 2nd ed. Philadelphia: Elsevier Mosby; 2005.)

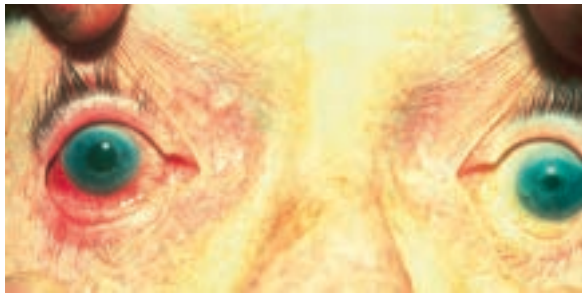

FIGURE 423-21. Acute angle-closure glaucoma. The left eye is normal. The red right eye has a nonreactive pupil. (Courtesy of Dr. Myron Yanoff.)

farsighted (hyperopia) have a shortened anterior-to-posterior axis of the eye, indicated clinically by a shallow anterior chamber. As the crystalline lens increases in volume with time, the iris is displaced anteriorly. At some point, the posterior surface of the iris may come in relatively tight contact with the anterior surface of the lens. Aqueous flow is restricted, and fluid accumulates in the posterior chamber, where it displaces the diaphanous peripheral iris anteriorly. When the peripheral iris comes in contact with the posterior cornea, the anterior chamber angle is suddenly occluded. Angle closure may be precipitated by pharmacologic dilation of the pupil. Patients who are farsighted (hyperopia) or have cataracts should be dilated with caution. The intraocular pressure may increase from 21 mm Hg to 50 to 70 mm Hg (nearly equaling diastolic arterial pressure). The symptoms of acute angle closure may include extreme pain, which may be poorly localized to the eye, nausea, and vomiting. Persistent vomiting may cause abdominal pain, simulating an acute abdomen.

Initial treatment is with topical (e.g., timolol 0.5% in one dose) and systemic pressure-lowering agents (e.g., carbonic acetazolamide, 250 to 500 mg intravenously or two 250-mg tablets orally in one dose if intravenous access or drug is not available), followed by creation of a fistula in the peripheral iris with a laser (YAG laser iridectomy) between the posterior chamber and the anterior chamber to bypass the obstruction. Most patients require a laser iridectomy prophylactically in the second eye to prevent angle-closure glaucoma.

Secondary glaucoma may also occur after intraocular hemorrhage, intraocular trauma, and intraocular inflammation. Some developmentally related secondary glaucomas, such as the iridocorneal endothelial syndrome, may not become evident until adulthood.

Retinal Detachment

A retinal detachment is a separation of the neural (sensory) retina from the retinal pigment epithelium. The two main types are rhegmatogenous, caused by a retinal hole, as may occur with a posterior vitreous detachment, and nonrhegmatogenous, caused by traction, such as in proliferative diabetic retinopathy or by fluid accumulation under neural retina in conditions such as malignant hypertension or eclampsia of pregnancy.

The classic symptoms are a sensation of flashes of light, floaters in the field of the involved eye owing to the causative vitreous detachment, and a shadow. Urgent treatment usually is needed. Laser photocoagulation is used to treat small posterior detachments, whereas cryotherapy is used for more peripheral tears. Intravitreal injection of 125 µg ocriplasmin (a recombinant protease with activity against fibronectin and laminin, which are components of the vitreoretinal interface) can significantly improve outcomes in patients with vitreomacular traction and macular holes. More extensive detachments may be treated with the injection of air. The air bubble will locate in the superior portion of the globe and tamponade the hole while it heals; the patient's head must be in the correct position for the duration of treatment. Large rhegamatogenous and traction retinal detachments are treated by more involved surgical procedures (scleral buckle surgery or vitreous surgery), which in most cases require general anesthesia and are associated with a longer postoperative course. Serous retinal detachments generally resolve without direct intervention when the underlying cause is successfully treated.

Age-Related Macular Degeneration

Age-related macular degeneration is a neurodegenerative disease that affects the junction between the neural retina and the retinal pigment epithelium, mainly in the sixth to ninth decades of life. About 8.5% of the world's blindness is caused by age-related macular degeneration, mainly in industrialized countries. In the United States, age-related macular degeneration affects more than 1.75 million persons, and its prevalence increases with each decade after

the age of 55 years. The genetic influence of age-related macular degeneration has not yet been determined, but abnormalities of complement factor H may play a role. Environmental factors such as smoking are known to accelerate this degenerative process.

Two types predominate: the "dry" or geographic type, and the "wet" or neovascular type. The wet type causes the most profound vision loss but is less common than the dry type.

The first sign of age-related macular degeneration is variation of the character and density of the retinal pigment epithelial pigment ("pigment dropout") in the posterior region of the retina, the macula. The pathologic process causes abnormal protein (lipofuscin) to accumulate between the retinal pigment epithelial cells and its basement membrane complex (Bruch's membrane), designated as *drusen.* The drusen of macular degeneration (soft drusen) are small (30 μm), hypopigmented, poorly defined areas in the deep retina. This stage, called dry age-related macular degeneration, may antedate subjective alteration in vision by several decades. The anticipated rate of progression is difficult to determine in individual cases.

The first symptom of age-related macular degeneration is loss of vision. Loss of vision corresponds with loss of photoreceptor outer segments. The effect is in the exact center of the most sensitive portion of the retina, the fovea. The peripheral retina is not involved. As the process progresses, individuals with advanced disease will be able to walk down a street without apparent difficulty (a peripheral retinal function) but will not be able to recognize facial features of people whom they meet (a macular retinal function). Visual aids and other devices, such as special glasses and television aids, may allow patients to continue with daily functions and to continue to live independently. This phase of the process, the "dry phase," advances at a slow rate (months to decades), during which patients develop retinal features of well-defined pigment loss (geographic retinal atrophy; Fig. 423-22). No current treatment exists for the dry phase of the disease, except for a possible positive influence of dietary supplements (antioxidants). Vitamin supplementation with vitamins C and E, β-carotene, zinc, and copper may retard the progression of moderate age-related macular degeneration to severe age-related macular degeneration.■ β-Carotene (a vitamin A precursor) of this quantity is not recommended for cigarette smokers because of an increased risk for lung cancer. Cessation of smoking, control of blood sugar, control of blood lipid levels, and control of systemic blood pressure are particularly important behavioral modifications.

In the "wet phase" of age-related macular degeneration (Fig. 423-23), frail, neovascular channels originating from the established vascular system of the choroid may extend through a breach in Bruch's membrane into the subretinal space (subretinal neovascularization). Spontaneous hemorrhage of the vessels adds to photoreceptor loss. Hemorrhage is accompanied by acute, and usually permanent, loss of central visual acuity (i.e., the wet phase of age-related macular degeneration). Both eyes are usually affected to a similar degree. Neovascularization can be identified by fluorescein angiography and OCT. Intraocular injections of antivascular endothelial growth factors (e.g., ranibizumab or bevacizumab) reduce the risk for visual loss in patients with neovascular age-related macular degeneration and can result in gains in vision.[16] Other treatments include argon laser photocoagulation or, in advanced cases, vitrectomy to remove subretinal neovascular membranes.

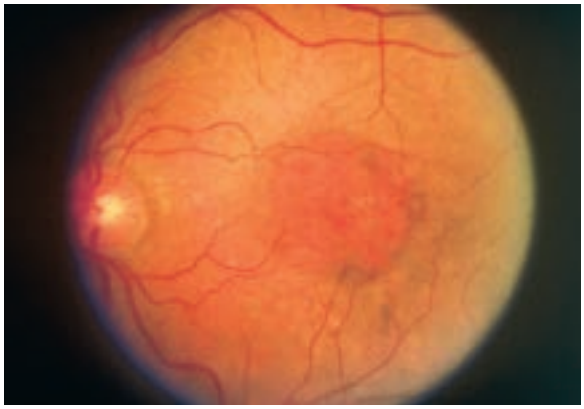

FIGURE 423-22. Dry age-related macular degeneration. Drusen are present in the posterior pole around a large area of geographic atrophy of the retinal pigment epithelium.

SYSTEMIC DISEASES WITH OCULAR SYMPTOMS DURING ADULTHOOD

Diabetes Mellitus

Diabetic retinopathy is one of the leading causes of blindness in the United States. More than 75% of the blind are women. Background diabetic retinopathy, with microaneurysms, hemorrhages, exudates (Fig. 423-24), and macular edema, accounts for most cases of decreased vision but rarely causes profound vision loss. Most diabetic patients never develop the more severe proliferative diabetic retinopathy (Fig. 423-25), which generally occurs only after 15 years or more of diabetes and causes a profound loss of vision.

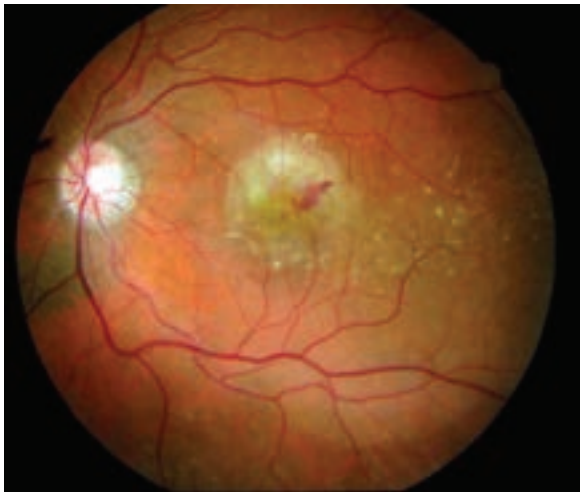

FIGURE 423-23. Wet age-related macular degeneration. A "dirty gray" neovascular membrane is present under the central macular area.

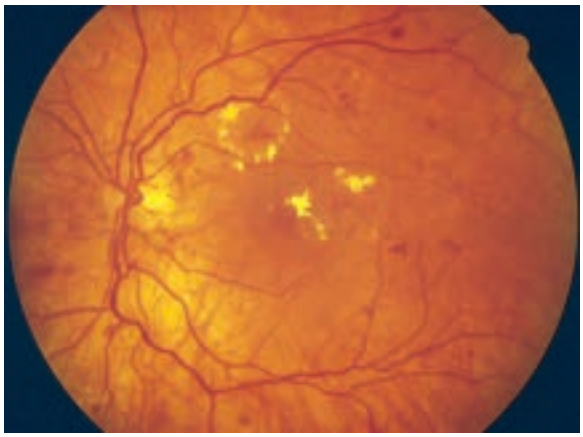

FIGURE 423-24. Background diabetic retinopathy. Exudates, microaneurysms, and small hemorrhages are seen in the posterior pole (right eye).

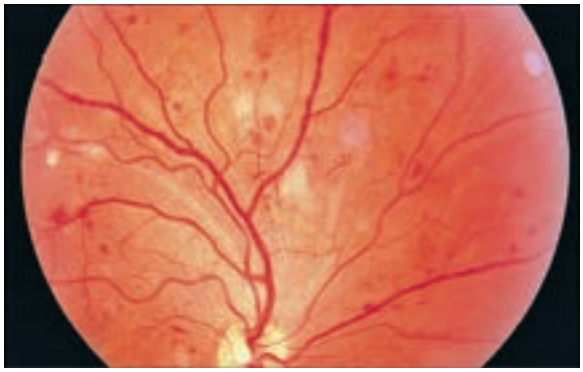

FIGURE 423-25. Severe proliferative diabetic retinopathy with cotton-wool spots, intraretinal microvascular abnormalities, and venous bleeding. (From Yanoff M, Duker JS, eds. *Ophthalmology.* Philadelphia: Mosby Elsevier; 2009.)

Diabetic retinopathy is closely correlated with the duration of diabetes mellitus (Chapter 229). The prevalence of diabetic retinopathy is about 27% among patients who have had type 1 diabetes for 5 to 10 years, 70 to 90% among patients who have had diabetes for more than 10 years, and 95% among patients who have had diabetes for 20 to 30 years. In patients with type 2 diabetes (Chapter 229), the prevalence of diabetic retinopathy is about 23% after 12 years and 60% after 16 years. Tight control of blood glucose greatly reduces the risk for the development of diabetic retinopathy.[17]

The treatment of diabetic retinopathy includes control of the diabetes and any hyperlipidemia, laser therapy for background diabetic retinopathy, and laser or surgical therapy to treat proliferative diabetic retinopathy. Antiangiogenic therapy (e.g., ranibizumab) may be superior to laser therapy for diabetic macular edema; intravitreal steroid injection is also beneficial. With acute hyperglycemia, accumulation of sorbitol may lead to lenticular swelling; secondary refractive errors may persist for 6 to 8 weeks.

Hypertension

In chronic hypertension (Chapter 67), the characteristic retinal vascular findings can assess the severity of hypertension. As the severity increases, patients develop arterial narrowing, arteriovenous nicking (Fig. 423-26), nerve fiber layer infarcts, and intraretinal hemorrhages. Moderately sclerosed arterioles have a "copper wiring" appearance, whereas severely sclerosed vessels demonstrate "silver wiring." Acute hypertension may cause optic nerve edema ("papilledema"; Fig. 423-27) and serous retinal detachments that usually resolve without significant sequelae if blood pressure is controlled.

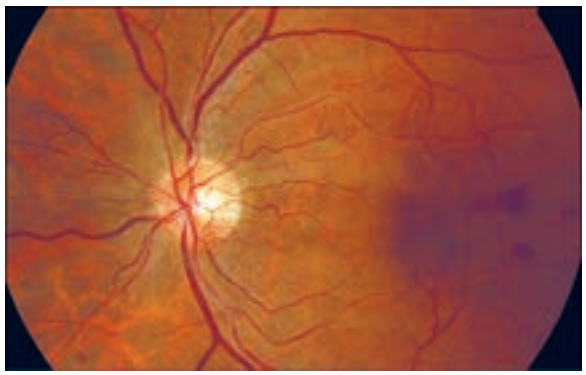

FIGURE 423-26. Hypertension retinopathy with narrowed arterioles whose sclerosed walls create the appearance of "nicking" when the arterioles cross venules. (From Yanoff M, Duker JS, eds. *Ophthalmology.* Philadelphia: Mosby Elsevier; 2009.)

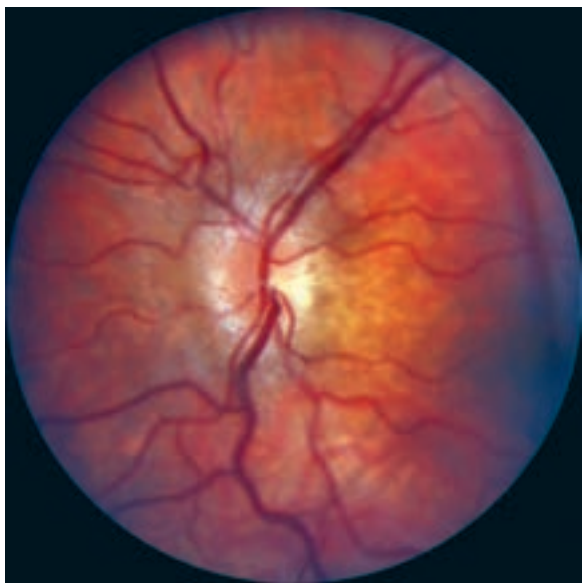

FIGURE 423-27. Papilledema. (Courtesy of Dr. Kathleen Digre.)

Other Systemic Diseases

In bacterial endocarditis (Chapter 76), emboli may cause conjunctival hemorrhages or the characteristic Roth's spot (Fig. 423-28). Accumulation of copper in the posterior cornea may aid in the diagnosis of Wilson's disease (Chapter 211), although its clinical diagnosis usually precedes the characteristic Kayser-Fleischer ring (see Fig. 211-2), which fades after treatment. Tay-Sachs and Niemann-Pick diseases (Chapter 208) are associated with a foveal cherry-red spot owing to the accumulation of gangliosides within perifoveal ganglion cells. Pseudoxanthoma elasticum (Chapter 260) is often associated with characteristic angioid streaks of the retina.

VASCULAR ABNORMALITIES OF THE EYE

The major vessels of the retina enter the eye at a point of relative constriction in the tissues of the lamina cribrosa of the optic disc. In persons who have generalized vascular disease, particularly systemic hypertension, occlusion of either the artery or vein may lead to a sudden loss of vision. Partial occlusion of either the artery or vein is associated with less visual loss but still increases the risk for developing neovascular glaucoma.

Arterial occlusion of the central retinal artery (Fig. 423-29) presents as painless acute loss of vision. The ischemic posterior retina generally has a pale gray appearance except at the fovea, where the normal color is preserved (cherry-red spot). The clinical appearance of edema resolves over time, but vision generally does not recover.

Venous occlusion of the central retinal vein (Fig. 423-30) presents as painless loss of vision. The retina, however, is characterized by extreme, generalized hemorrhages, which resolve very slowly and are often associated with opaque areas of focal retinal ischemia (cotton-wool spots). This process destroys the full thickness of the retina. Usually, no vision returns without

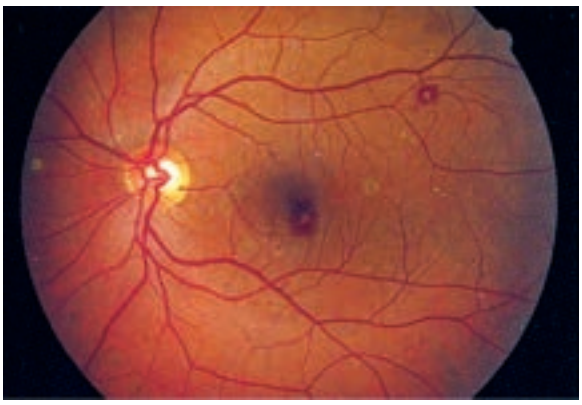

FIGURE 423-28. Roth's spots. Multiple white-centered hemorrhages in a man with recurrent subacute bacterial endocarditis. White-centered hemorrhages are also seen with leukemia and diabetes. The small white scars are probably the residua of previous episodes.

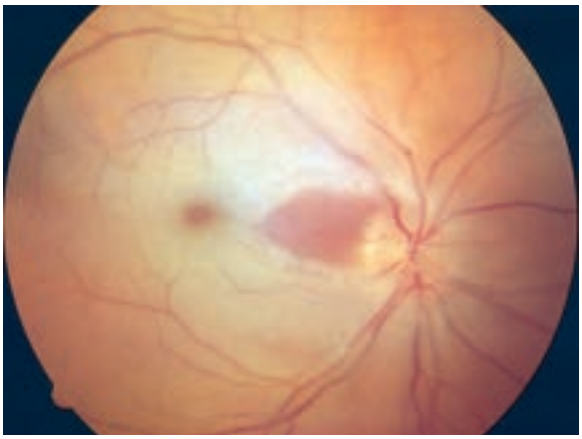

FIGURE 423-29. Central retinal artery occlusion. Fundus photograph shows diffuse retinal edema. The heavily pigmented fovea with its uniquely thin inner retina produces a cherry-red spot against the dusky macula. In this case, a small area of retina adjacent to the optic disc is spared, owing to the presence of a cilioretinal artery.

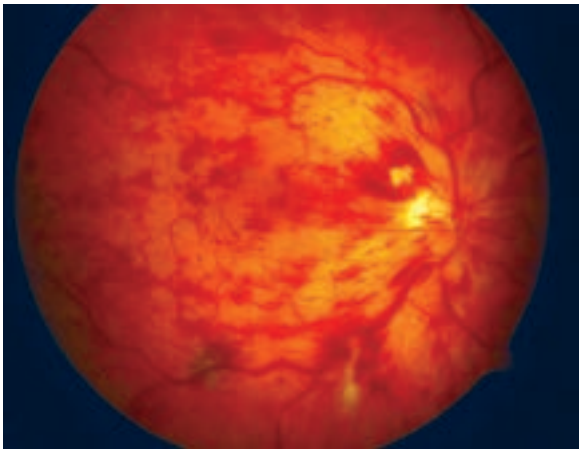

FIGURE 423-30. Central retinal vein occlusion with diffuse intraretinal hemorrhages in all four quadrants.

treatment, but intravitreal ranibizumab can provide significant and sustained visual improvement. Central vein occlusion is associated with a major risk for developing secondary neovascular glaucoma.

Giant cell arteritis (temporal arteritis [Chapter 271]) is occlusion of the blood supply of the optic disc by inflammation of the short posterior ciliary arteries. Occlusion causes acute, painless loss of vision. Presenting symptoms before visual loss often include a vague sensation of fatigue, scalp tenderness, or jaw claudication. Diagnosis is suspected by classic characteristics and by an elevated erythrocyte sedimentation rate or elevated C-reactive protein level. Histologic confirmation is provided by a biopsy showing granulomatous inflammation in the region of the internal elastic lamina of the temporal artery. Emergent treatment, often protracted, with systemic steroids (Chapter 271) can prevent vascular occlusion and visual loss, but lost vision is not usually restored by treatment. There is a significant risk for the process developing in the second eye.

Nonarteritic optic neuropathy is caused by occlusion of the posterior ciliary arteries and infarction of the optic disc, resulting in acute, usually unilateral, painless loss of vision. There often are no antecedent symptoms except those systemic signs and symptoms associated with systemic nonarteritic vascular disease such as systemic hypertension. Occlusion is thought to be due to atherosclerosis or some other lumen-compromising mechanism. No treatment restores vision. There is a risk for the same process affecting the second eye.

IDIOPATHIC INFLAMMATORY AND AUTOIMMUNE DISORDERS

Ocular or periocular tissues may be the primary focus of isolated idiopathic or autoimmune inflammation. Pain is common, and changes in vision may occur.

Keratoconjunctivitis Sicca

Keratoconjunctivitis sicca, or the dry eye syndrome, results from deficiency of any of the tear film layers. Symptoms include gritty, foreign body sensations, burning, photophobia, and decreased visual acuity. Idiopathic inflammation in keratoconjunctivitis sicca and xerostomia represents Sjögren's syndrome (Chapter 268). Recurrent corneal erosion, keratitis, and corneal opacification can occur. Many medications can also cause dry eyes (Table 423-8).

Artificial tears, up to four times daily, and lubricating ointments are helpful. Corticosteroids (e.g., loteprednol 0.5% eye-drops four times a day) are also effective as an initial treatment. Cyclosporine eye-drops (0.05%, one drop in each eye every 12 hours) are useful when other measures fail.

Scleritis

Episcleritis, which is an inflammation immediately underlying the conjunctiva, is distinguished from conjunctivitis because its radially oriented vessels do not move with the conjunctiva. Mild pain may be present. Episcleritis is self-limited. Instillation of 2.5% phenylephrine is helpful in making the diagnosis because it causes blanching in episcleritis but not in scleritis. Oral or topical nonsteroidal anti-inflammatory medications such as flurbiprofen or diclofenac may hasten resolution.

TABLE 423-8 PARTIAL LIST OF SYSTEMIC MEDICATIONS CAUSING DRY EYE

MEDICATION	CLASS
Ibuprofen	Nonsteroidal anti-inflammatory
Diphenhydramine	Antihistamine
Triprolidine	Antihistamine
Chlorpheniramine	Antihistamine
Atenolol	β-Blocker
Metoprolol	β-Blocker
Propranolol	β-Blocker
Clonidine	α-Agonist
Scopolamine	Anticholinergic
Amiodarone	Antiarrhythmic
Thiabendazole	Antinematode
Isotretinoin*	Retinoid

*Severe, long-term dry eye with onset up to several years after treatment. All others tend to abate with cessation.
From Goldman L, Ausiello DA, eds. *Cecil Textbook of Medicine*, 23rd ed. Philadelphia: Saunders Elsevier; 2008:2852.

Scleritis, which presents as severe pain and redness, is associated with infectious or autoimmune connective tissue disease in about 50% of cases. Vision may be reduced if the posterior sclera is involved. Diffuse or sectoral hyperemia is nonmobile and does not blanch with instillation of phenylephrine. Secondary uveitis and keratitis may occur. Diagnostic evaluation includes ultrasonography or magnetic resonance imaging and laboratory tests to identify potential underlying conditions. Treatment may require topical or oral nonsteroidal anti-inflammatory medications or corticosteroids.

Mooren's Ulcer

Mooren's ulcer is idiopathic, progressive, peripheral corneal thinning, likely autoimmune. It can be unilateral or bilateral, and pain is common. Topical corticosteroids, mucolytics, and cytotoxic agents have been used. Bandage contact lenses and conjunctival recession or advancement have also been used with variable success.

Orbital Pseudotumor

Nonspecific, idiopathic orbital inflammation involving the lacrimal gland (dacryoadenitis), extraocular muscles (myositis), orbital fat, sclera, or optic nerve sheath (optic perineuritis) can be caused by orbital pseudotumor. Some of the cases of idiopathic orbital inflammation have been recently associated with immunoglobulin G4 autoimmune disease. Pain is frequent. Patients may present with proptosis, limited ocular movements, or decreased acuity. OCT or magnetic resonance imaging excludes a mass lesion. Patients respond dramatically to systemic corticosteroids within 24 hours, but the steroids must be tapered slowly over months to prevent recurrence.

Iritis

Iritis presents with pain, photophobia, and blurred vision, with about 50% of cases related to systemic disease. Slit lamp examination shows inflammatory cells and protein exudate in the anterior chamber. Symptomatic treatment is with prednisolone acetate 1% suspension four times a day and cycloplegic drugs (cyclopentolate 1 or 2% twice daily) is usually effective, but repeated episodes require evaluation for autoimmune and infectious causes.

Rheumatoid Arthritis

Juvenile rheumatoid arthritis (Chapter 264) is the most common specific childhood entity associated with uveitis. In adults, the ocular manifestations of rheumatoid arthritis mainly affect the anterior part of the eye, cornea, and sclera. Nearly 50% of patients who have peripheral ulcerative keratitis of the cornea have an associated systemic disease, mainly collagen vascular disease, and especially rheumatoid arthritis. Similarly, almost half of patients who have scleritis have an associated systemic disease, and about 15% of these are connective tissue diseases. Scleromalacia perforans, which is aseptic necrosis of the sclera, is associated with rheumatoid arthritis about 46% of the time.

Systemic Lupus Erythematosus

Systemic lupus erythematosus (Chapter 266) causes eye manifestations from both the primary disease and its treatment with derivatives of chloroquine. The retina may show a retinal vasculitis, and the optic nerve an optic neuritis, ischemic or nonischemic. Chloroquine therapy can cause a toxic retinal degeneration, but this complication is rare if the daily dose of chloroquine does not exceed 250 mg (and 6.5 mg/kg of body weight for hydroxychloroquine).

Sarcoidosis

About 25% of patients with sarcoidosis (Chapter 95) develop chronic uveitis. Sarcoid also can involve the lids, conjunctiva, optic nerve, cranial nerves, and lacrimal glands. Anterior uveitis is treated topically with prednisolone acetate in decreasing doses, depending on degree of inflammation, and with daily cycloplegics (cyclopentolate 2%, atropine 1%). Posterior uveitis, dacryoadenitis, and neurologic manifestations require systemic corticosteroids, but the doses have not been standardized.

Sympathetic Ophthalmia

Sympathetic ophthalmia is an autoimmune disease characterized by bilateral, granulomatous uveitis following trauma to one eye. The condition is very rare, occurring in less than 1 per 10,000 cases of ocular surgical procedures and 1 per 1000 cases of accidental trauma.[18]

The identified antigen within the eye is thought to be located in the outer retina. The disease is recognized clinically by signs of inflammation in the uninjured eye, generally 2 weeks or longer after the injury. Generally, removal of the injured eye within these 2 weeks will protect against the development of sympathetic ophthalmia in the uninjured eye, but once the uninjured eye is involved, removal of the originally injured eye is not likely to influence the process. Left untreated, inflammation may destroy the function of both eyes. When sympathetic ophthalmia is established, the patient will require anti-inflammation treatment (e.g., prednisolone, 1.0 to 1.5 mg/kg orally per day), most likely for an extended period. Most patients retain useful vision if treated at an early stage.

GENETICALLY DETERMINED DISEASES THAT MAY BECOME SYMPTOMATIC DURING ADULTHOOD

Corneal Stromal Dystrophy

Most corneal dystrophies are autosomal dominant and bilateral, progress slowly, and primarily affect one layer of an otherwise normal cornea. Common types of dystrophies are anterior basement membrane, macular, granular, lattice, and Fuchs' endothelial. Some result from mutations within the same gene. For example, the *BIGH3* on 5q31 is associated with granular and lattice dystrophy and corneal dystrophy of Bowman. The main symptom, caused by opaque corneal deposits, is blurred vision. If the decreased vision interferes with activities of normal living, a corneal transplantation can be performed.

Choroidal Dystrophy

Choroidal dystrophies are progressive, inherited disorders characterized by atrophy of the retinal pigment epithelium and choroid. The main entities are central areolar choroidal sclerosis (autosomal dominant or recessive), gyrate atrophy (deficiency of the mitochondrial matrix enzyme ornithine-δ-aminotransferase), and choroideremia (deficiency of component A of Rab geranylgeranyl transferase). No treatment exists for central areolar choroidal sclerosis or choroideremia. An arginine-restricted diet may be helpful in treating gyrate atrophy.

Retinitis Pigmentosa

Retinitis pigmentosa (Fig. 423-31) is bilateral and symmetrical, starts in early adult life, and is progressive. Retinitis pigmentosa can be an autosomal dominant or recessive disease, X-linked, digenic, mitochondrial, or sporadic. The primary defect, apoptotic in nature, appears to be in the neural retinal receptors. The main findings consist of the tetrad of bone-corpuscular retinal pigmentation; a pale, waxy optic nerve; attenuation of retinal arterioles; and a posterior subcapsular cataract. Night blindness is the primary symptom. The electroretinogram usually shows no electrical evidence of retinal function. Vitamin A palmitate supplementation (15,000 IU daily) may slow the rate of progression. Gene therapy is under investigation.

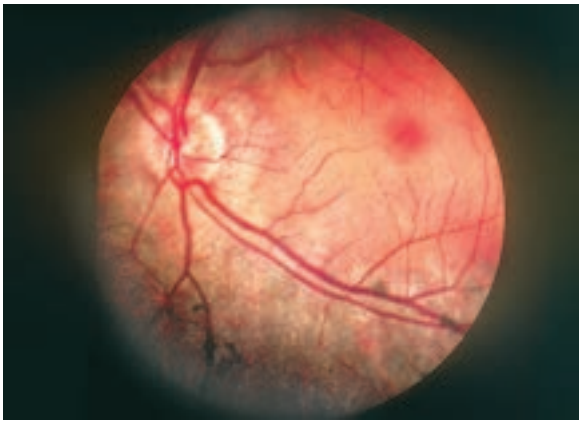

FIGURE 423-31. Retinitis pigmentosa. Fundus photograph shows "bone spicule" pigmentation of the midperipheral fundus, waxy pallor of the optic disc, and attenuated retinal vessels, the most consistent finding in retinitis pigmentosa. (Courtesy of Dr. John I. Loewenstein.)

COMMON PEDIATRIC OR ADOLESCENT DISEASES THAT MAY PERSIST INTO ADULTHOOD

Retinopathy of Prematurity

The peripheral retina, particularly the temporal retina, is not vascularized in premature infants (36 weeks or younger). In the premature infant, vessels of the immature retina may leave the plane of the retina and grow into the adjacent vitreous body, where tractional forces can cause total, irreversible retinal detachment. The detached retina forms a mass of fibrovascular tissue posterior to the crystalline lens, explaining the original name of this condition, *retrolental fibroplasia.* The primary risk factor for retinopathy of prematurity is the degree of prematurity of the infant and the use of supplemental oxygen. Early recognition and treatment of retinal neovascularization are essential for reestablishing normal retinal vascular development.

Hemangioma of the Eyelid

Hemangioma of the eyelid is a hamartomatous (tissue normally found in the area) malformation of vessels in the soft tissues of the eyelid. The abnormal vascular channel, which may not be clinically evident until several weeks after birth, can expand to enlarge the eyelid and cause mechanical ptosis. Eyelid hemangioma may involute over time, usually in 4 to 7 years, but in the interval there is risk for developing amblyopia. Traditional treatment includes intralesional steroid injection, laser photocoagulation, and, rarely, surgical debulking. Oral propanolol can result in complete regression of eyelid capillary hemangioma in 4 months[19] in patients who are younger than 1 year, which is when the risk for developing amblyopia is highest.

Congenital Cataract

Opacification of the crystalline lens occurs in many developmental and biochemical abnormalities. In congenital rubella (Chapter 368), the opacity is relatively limited to the fetal nucleus and has a pearl-like density. Cataract extraction is essential for dense cataracts of any type if amblyopia is to be prevented, but liberation of live rubella virus into the eye during cataract surgery may lead to necrotizing endophthalmitis and loss of the eye.

The cataract of galactosemia (Chapter 205) is potentially reversible with dietary restriction of galactose. Most other congenital cataracts do not progress. The degree of vision, which may be surprisingly adequate, cannot be accurately predicted by the clinical appearance of lens opacity.

TUMORS OF THE EYE

Retinoblastoma

Retinoblastoma, the most common intraocular malignant tumor of childhood, results from uncontrolled proliferation of retinoblasts, which are pluripotential neuroectodermal cells that will differentiate into the various components of mature retina. A genetic deletion of the *Rb* (retinoblastoma) gene occurs in the chromosomal region 13q14. The tumor initially proliferates in the plane of the retina but is capable of involving all structures within

the eye. Retinoblastoma may spread to the central nervous system through the optic nerve and through blood vessels to tissues at distant sites.

Approximately 40% of cases of retinoblastoma are heritable and generally diagnosed before age 12 months, much earlier than the nonheritable form. In 80% of heritable retinoblastomas, retinal tumors are multiple in each eye and are bilateral. These patients have a significant risk for developing a secondary primary malignant tumor (e.g., osteogenic sarcoma).

The nonheritable, sporadic form arises spontaneously and represents about 60% of cases. The average age at presentation is 24 months. Generally only a single tumor occurs in one eye, unlike the heritable form, in which multiple tumors in one eye and bilateral tumors are common. No detectable chromosomal abnormality is present, thereby making the risk for retinoblastoma in succeeding generations low. These children have the same risk for second primary tumors as the general population.

Retinoblastoma is often discovered by parents or relatives who notice a light reflex in one eye relative to the other either (white or "cat's eye" reflex; leukokoria). Children with retinoblastoma also may present with strabismus, iris neovascularization, dilated fixed pupil, secondary glaucoma, or tumor accumulation in the anterior chamber (neoplastic hypopyon). More advanced cases may present with signs of intraocular inflammation or ruptured globe with orbital extension. In some of cases of regressed retinoblastoma, the sole clinical sign may be a small, calcified tumor in the plane of the retina with surrounding retinal pigment epithelial scarring. Magnetic resonance imaging is the diagnostic test of choice.[20]

Most children are treated with chemotherapy, sometimes also with intraocular laser therapy or radiation, or both. Enucleation is used for advanced cases with no visual potential.

Malignant Melanoma

Malignant melanoma (Chapter 203) of the conjunctiva is rare. The individuals at risk are middle aged and lightly pigmented. Melanoma of the conjunctiva may arise from a preexisting nevus or de novo, but most arise from unilateral, acquired, variably pigmented regional, flat, or slightly elevated lesions in which the degree and distribution of pigmentation tend to vary over time. The regions of greatest risk are in the conjunctiva at the limbus (junction of cornea and sclera), in the conjunctival fornix (deep peripheral recesses of the conjunctiva), and in the caruncle (elevated nodule between the nasal lid margins). Extensions onto the corneal surface or formation of a nodule or loss of pigmentation are indications for biopsy. Proliferation and hyperpigmentation of melanocytes without nuclear or cellular atypia is not associated with progression to melanoma, whereas associated nuclear or cellular atypia, particularly coupled with mitotic activity, is highly linked with progression to melanoma. Frank melanoma with a thickness of more than 0.8 mm is a risk factor for metastatic melanoma. Treatment is surgical excision, often supplemented with cryoablation. The long-term outcome is less favorable than for cutaneous melanoma because the tumor may metastasize early when the primary tumor is very small (e.g., 2 mm).

Malignant melanoma of the uveal tract is the most common primary intraocular malignancy of adults, but its incidence is only 2 to 6 per 1 million per year in high-risk populations (blue-eyed white people). The tumor, which arises from preexisting nevi or dendritic melanocytes anywhere in the uveal tract, is almost always unilateral, unicentric, and nodular and is usually diagnosed at an asymptomatic phase during routine screening examination of the dilated fundus. Symptomatic tumors arise near sensitive portions of the retina (e.g., the macula) or cause retinal detachment or cystoid macular edema. Iris tumors (Fig. 423-32) are generally pigmented and elevated above the surrounding contour, where they are recognized early in their course. Posterior tumors may be completely amelanotic and occasionally may have a bilobed appearance. Accuracy in diagnosing melanoma of the uveal tract by clinical means alone is greater than 98%. Treatment is controversial; options include enucleation of the eye, external beam and plaque (iodine-125) radiation, and en bloc resection. The overall survival rate is approximately 50% at 15 years. Risk factors for metastasis, most commonly to the liver, include tumor size, cell type, angiogenic mimicry, the presence of monosomy 3, and other genetic markers.

Orbital Tumors

Primary tumors in the orbit of adults include cavernous hemangioma, schwannomas, and various proliferations of fibrous tissue (solitary fibrous tumor). Rhabdomyosarcoma may arise from ectopic rests of mesenchyme rather than from mature rectus muscle. Orbital tumors usually are diagnosed by imaging techniques. The treatment is orbital exploration and surgical removal.

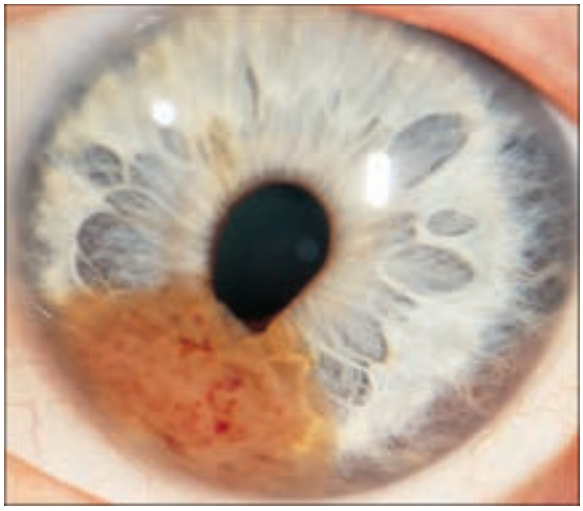

FIGURE 423-32. Iris melanoma that has prominent intrinsic blood vessels. Note peaking of the pupil toward the tumor. (From Yanoff M, Duker JS, eds. *Ophthalmology.* Philadelphia: Mosby Elsevier; 2009.)

Lymphoma

Orbital and conjunctival lymphomas are usually small B-cell mucosa-associated lymphoid tissue tumors (Chapter 185). Approximately 50% of cases ultimately include systemic disease that may, however, be delayed by decades. External beam radiation is used to treat isolated periocular disease, whereas systemic chemotherapy is required for systemic involvement.

Large-cell B-cell lymphoma (Chapter 185) may present in the eye as a form of vitreitis (cells suspended in the vitreous) or a subretinal or intraretinal infiltrate before it is discovered in the central nervous system. The diagnosis may be verified by cytologic examination of vitrectomy specimens. Treatment is generally systemic, but the outcome is usually poor.

The eye may infrequently be involved in multiple myeloma (Chapter 187). Retinal hemorrhages and vitreous opacification may occur. Periorbital osteolytic lesions of bone may be present.

Lacrimal Gland Tumors

The lacrimal gland contains a resting population of non-nodal lymphocytes and is a second common site for lymphomas. Epithelial neoplasms may arise from components of the acini and ducts of the lacrimal gland. Malignant epithelial tumors (adenoid cystic carcinoma) may metastasize at an early stage through perineural spaces of large peripheral nerves to adjacent bone. Most epithelial tumors are treated with total surgical removal of the lacrimal gland because of the risk for recurrence and malignant transformation of residual tumor. The prognosis for malignant lacrimal gland tumors is generally poor.

Ocular Metastasis

Metastasis to the orbit in adults is very uncommon because of its relatively small vascular volume. Metastasis to the rectus muscles present with adult-onset strabismus. Metastasis to the orbit is more common in childhood leukemia than adult leukemia.

Metastasis to the uveal tract is common, especially from primary breast and lung tumors. Metastatic lesions often grow rapidly and disturb visual function; serous retinal detachment is common.

OCULAR EFFECTS OF SYSTEMIC MEDICATIONS

Innumerable medications may cause ocular side effects (Table 423-9). Therefore, patients taking systemic medications often require periodic surveillance to identify ocular toxicity.

The most common cause of drug-induced glaucoma is topical corticosteroid therapy of more than 4 to 6 weeks' duration in the 5 to 6% of the population that is genetically predisposed. Nonsteroidal drugs usually cause narrow-angle glaucoma. Sulfa-containing drugs may induce glaucoma as an idiosyncratic reaction in persons with either narrow or open anterior chamber configuration. Treatment is the same as for non-drug-induced glaucoma.[21]

TABLE 423-9 SYSTEMIC MEDICATIONS WITH OCULAR EFFECTS

AGENT	EFFECT
Chloroquine	Dyschromatopsia, visual field defects
Hydroxychloroquine	Dyschromatopsia, visual field defects
Thioridazine	Blurred vision
Chlorpromazine	Blurred vision
Digoxin	Yellow vision
Ethambutol	Optic neuritis
Amiodarone	Corneal whorls, pigmentary retinopathy
Corticosteroids	Glaucoma, cataract
Plaquenil	Pigmentary maculopathy
Tamoxifen	Retinopathy
Neuroleptics	Nystagmus
Compazine	Oculogyric crisis
Vitamin A	Pseudotumor cerebri
5-Fluorouracil	Canalicular stenosis (tearing)
Isotretinoin	Severe dry eye (long-term effect)

From Goldman L, Ausiello DA, eds. *Cecil Textbook of Medicine*, 23rd ed. Philadelphia: Saunders Elsevier; 2008:2852.

Chloroquine and *hydroxychloroquine* may cause decreased color vision and visual field defects at high doses. Chloroquine toxicity is thought to occur after a cumulative dose of 300 g, whereas hydroxychloroquine may cause symptoms after long-term maintenance of 750 mg/day. The macular portion of the fundus develops a typical bull's-eye pattern of pigment disturbance. Corneal whorls composed of epithelial intracellular pigment may be seen. Vision loss from retinal toxicity is not reversible and tends to progress even after cessation of hydroxychloroquine treatment. Annual fundus examination with color testing, macular function tests, and automated visual field test may be indicated.

Ethambutol is often used for chronic pulmonary infection. The dose of drug is dependent on body weight. Ocular complications, toxic optic neuropathy, is rare but unpredictable in incidence and outcome. Careful follow up with close ophthalmic clinical surveillance is required.

Any of the commonly used antituberculous medications may cause optic neuropathy, although *ethambutol* carries the greatest risk. Pupillary response, color vision, acuity, and visual fields are the clinical parameters used to assess optic nerve function.

Cornea verticillata may be seen in patients taking *amiodarone* because of lysosomal accumulations within the epithelial basement membrane. *Fabry's disease* produces similar changes, as can other medications. Corneal whorls are usually reversible when caused by drug toxicity, and they rarely interfere with vision.

Grade A References

A1. Shortt AJ, Allan BD, Evans JR. Laser-assisted in-situ keratomileusis (LASIK) versus photorefractive keratectomy (PRK) for myopia. *Cochrane Database Syst Rev*. 2013;1:CD005135.

A2. Wittig-Silva C, Chan E, Islam FM, et al. A randomized, controlled trial of corneal collagen cross-linking in progressive keratoconus: three-year results. *Ophthalmology*. 2014;121:812-821.

A3. Kaufman HE, Haw WH. Ganciclovir ophthalmic gel 0.15%: safety and efficacy of a new treatment for herpes simplex keratitis. *Curr Eye Res*. 2012;37:654-660.

A4. Stalmans P, Benz MS, Gandorfer A, et al. Enzymatic vitreolysis with ocriplasmin for vitreomacular traction and macular holes. *N Engl J Med*. 2012;367:606-615.

A5. Evans JR, Lawrenson JG. Antioxidant vitamin and mineral supplements for slowing the progression of age-related macular degeneration. *Cochrane Database Syst Rev*. 2012;11:CD000254.

A6. Chakravarthy U, Harding SP, Rogers CA, et al. Alternative treatments to inhibit VEGF in age-related choroidal neovascularisation: 2-year findings of the IVAN randomised controlled trial. *Lancet*. 2013;382:1258-1267.

A7. Chew EY, Ambrosius WT, Davis MD, et al., for the ACCORD Eye Study Group. Effects of medical therapies on retinopathy progression in type 2 diabetes. *N Engl J Med*. 2010;363:233-244.

A8. Hoerle S, Kroll P. Evidence-based therapy of diabetic retinopathy. *Ophthalmologica*. 2007;221:132-141.

A9. Régnier S, Malcolm W, Allen F, et al. Efficacy of anti-VEGF and laser photocoagulation in the treatment of visual impairment due to diabetic macular edema: a systematic review and network meta-analysis. *PLoS OnE*. 2014;9:e102309.

A10. Zhang Y, Ma J, Meng N, et al. Comparison of intravitreal triamcinolone acetonide with intravitreal bevacizumab for treatment of diabetic macular edema: a meta-analysis. *Curr Eye Res*. 2013;38:578-587.

A11. Tan MH, McAllister IL, Gillies ME, et al. Randomized controlled trial of intravitreal ranibizumab versus standard grid laser for macular edema following branch retinal vein occlusion. *Am J Ophthalmol*. 2014;157:237-247.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Pasquali T, Krueger R. Topography-guided laser refractive surgery. *Curr Opin Ophthalmol*. 2012;23:264-268.
2. Jhanji V, Sharma N, Vajpayee RB. Management of keratoconus: current scenario. *Br J Ophthalmol*. 2011;95:1044-1050.
3. Repka MX, Kraker RT, Holmes JM, et al. Atropine vs patching for treatment of moderate amblyopia: follow-up at 15 years of age of a randomized clinical trial. *JAMA Ophthalmol*. 2014;132:799-805.
4. Esmaeli B, Nasser QJ, Cruz H, et al. American Joint Committee on Cancer T category for eyelid sebaceous carcinoma correlates with nodal metastasis and survival. *Ophthalmology*. 2012;119:1078-1082.
5. Ho SF, Brown L, Bamford M, et al. 5 years review of periocular basal cell carcinoma and proposed follow-up protocol. *Eye (Lond)*. 2013;27:78-83.
6. Oellers P, Karp CL, Sheth A, et al. Prevalence, treatment, and outcomes of coexistent ocular surface squamous neoplasia and pterygium. *Ophthalmology*. 2013;120:445-450.
7. Friling E, Lundstrom M, Stenevi U, et al. Six-year incidence of endophthalmitis after cataract surgery: Swedish national study. *J Cataract Refract Surg*. 2013;39:15-21.
8. Azari AA, Barney NP. Conjunctivitis: a systematic review of diagnosis and treatment. *JAMA*. 2013;310:1721-1729.
9. Hill GM, Ku ES, Dwarakanathan S. Herpes simplex keratitis. *Dis Mon*. 2014;60:239-246.
10. Baig AM, Zuberi H, Khan NA. Recommendations for the management of *Acanthamoeba* keratitis. *J Med Microbiol*. 2014;63:770-771.
11. Maenz M, Schluter D, Liesenfeld O, et al. Ocular toxoplasmosis past, present and new aspects of an old disease. *Prog Retin Eye Res*. 2014;39:77-106.
12. Bhoomibunchoo C, Ratanapakorn T, Sinawat S, et al. Infectious endophthalmitis: review of 420 cases. *Clin Ophthalmol*. 2013;7:247-252.
13. Moyer VA. Screening for glaucoma: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2013;159:484-489.
14. King A, Azuara-Blanco A, Tuulonen A. Glaucoma. *BMJ*. 2013;346:f3518.
15. Patel K, Patel S. Angle-closure glaucoma. *Dis Mon*. 2014;60:254-262.
16. Finger RP, Wickremasinghe SS, Baird PN, et al. Predictors of anti-VEGF treatment response in neovascular age-related macular degeneration. *Surv Ophthalmol*. 2014;59:1-18.
17. Arevalo JF. Diabetic macular edema: changing treatment paradigms. *Curr Opin Ophthalmol*. 2014;25:502-507.
18. Chang GC, Young LH. Sympathetic ophthalmia. *Semin Ophthalmol*. 2011;26:316-320.
19. Vassallo P, Forte R, Di Mezza A, et al. Treatment of infantile capillary hemangioma of the eyelid with systemic propranolol. *Am J Ophthalmol*. 2013;155:165-170.
20. de Jong MC, de Graaf P, Noij DP, et al. Diagnostic performance of magnetic resonance imaging and computed tomography for advanced retinoblastoma: a systematic review and meta-analysis. *Ophthalmology*. 2014;121:1109-1118.
21. Razeghinejad MR, Myers JS, Katz LJ. Iatrogenic glaucoma secondary to medications. *Am J Med*. 2011;124:20-25.

REVIEW QUESTIONS

1. A 45-year-old woman is having difficulty with near vision tasks in both eyes. What is the likely cause of this problem?

A. A decreased range of accommodation
B. Increased opacification of the crystalline lens
C. A separation of the retina from the retinal pigment epithelium
D. Scarring on the internal surface of the retina
E. Neovascularization and scarring of the external retina.

Answer: A At age 45 years, the amount of available accommodative reserve decreases to the point at which reading and other near tasks become difficult. Although cataracts can occur at age 45 years, the incidence is very low, and such early cataracts are usually unilateral. A serous detachment of the retina usually occurs in one eye and causes acquired asymmetrical vision. Scarring of the internal surface of the retina (epiretinal membrane formation) is also usually a unilateral abnormality. Neovascularization of the external surface of the retina (subretinal neovascularization) is also usually a unilateral acquired abnormality.

2. A 35-year-old man developed mild focal upper eyelid pain and swelling associated with a tender mass. What is the likely cause of the mass?

A. An abscess of the eyelash follicle
B. A rupture of a meibomian gland
C. Proliferation of the basal layers of the eyelid epidermis
D. A blockage of an eccrine duct of the eyelid
E. Proliferation of non-nodal lymphoid tissue of the conjunctiva.

Answer: B Unilateral eyelid pain and swelling is the usual presentation of a chalazion (rupture of a meibomian gland). A stye, which is an abscess of a pilosebaceous unit, presents with extreme pain and the risk for anterior or posterior cellulites. Basal cell carcinoma is a painless mass. Blockage of an eccrine duct (ductal cyst) is also painless and usually occurs in the conjunctiva. Proliferation of non-nodal lymphoid tissue (lymphoma) is also painless and usually found in the upper or lower fornix of the conjunctiva.

3. A 43-year-old near-sighted man notes the sensation of flashing lights and an increased number of small floaters in the right eye. This afternoon during his annual physical examination, he notes loss of the inferior visual field of that eye. What is the most appropriate course of action?

A. Chest imaging for a primary neoplasm
B. Carotid artery imaging
C. Evaluation for systemic hypertension
D. Immediate referral for intraocular evaluation
E. Schedule for next available eye examination

Answer: D Recent onset of the sensation of floaters and flashing lights is the common presentation of retinal detachment. Visual outcome is dependent on early diagnosis and treatment. The other systemic conditions present much differently.

4. A 71-year-old man has noted halos around oncoming car headlights and double vision in only one eye. What is the most likely cause of his vision change?

A. Intimal thickening of the ipsilateral carotid artery
B. Cerebral aneurysm
C. Brain stem vascular insufficiency
D. Serous retinal detachment
E. Increase of refractive index of the lens

Answer: E Monocular diplopia (diplopia with one eye open) is a common presentation of a cataract, although corneal surface abnormalities may present similarly. The other conditions present with bilateral diplopia (diplopia disappears with closure of one eye).

5. The notation in a drug insert states that the drug should not be used if the patient has glaucoma. Which type of glaucoma puts the patient at risk for vision loss with this drug?

A. Chronic open-angle glaucoma
B. Exfoliative glaucoma
C. Narrow angle glaucoma
D. Neovascular glaucoma
E. Trauma-associated glaucoma

Answer: C The most common cause of nonsteroidal-induced glaucoma is that associated with narrow angle anterior chamber configuration. The other types of glaucoma are not risk factors for drug-induced glaucoma.

6. A 50-year-old man with a hemoglobin A1c level of 10% has noted over the past week that he can read comfortably without his bifocals. What is the significance of this observation?

A. Induced myopia due to hyperglycemia
B. Macular edema
C. Occlusion of a branch retinal artery
D. Late-stage background diabetic retinopathy
E. Early-stage proliferative diabetic retinopathy

Answer: A This person has uncontrolled diabetes associated with induced myopia. The induced myopia has corrected his near vision but made his distance vision difficult. All of the other conditions will decrease visual acuity at near.

424

NEURO-OPHTHALMOLOGY

ROBERT W. BALOH AND JOANNA C. JEN

A mechanistic understanding of vision impairment along with disturbances in pupillary and oculomotor control lies close to the heart of diagnosing neurologic disorders.

VISION

One of the most difficult diagnostic problems is vision loss that cannot be explained by obvious abnormalities of the eye. To evaluate such a patient properly, the examining physician must be familiar with the anatomy and physiology of the afferent visual system. The afferent visual pathways cross the major ascending sensory and descending motor systems of the cerebral hemispheres and in their anterior portion are intimately related to the vascular and bony structures at the base of the brain. Not surprisingly, localization of lesions within the afferent visual pathways has great value in neurologic diagnosis.

Anatomy of the Visual Pathways

Light entering the eye falls on the retinal rods and cones, which transduce the stimulus into neural impulses to be transmitted to the brain. The distribution of visual function across the retina takes a pattern of concentric zones increasing in sensitivity toward the center, the fovea. The fovea consists of a "rod-free" central grouping of approximately 100,000 slender cones. The ganglion cells subserving these cones send their axons directly to the temporal aspect of the optic disc, where they form the papillomacular bundle. Axons originating from ganglion cells in the temporal retina curve above and below the papillomacular bundle and form dense arcuate bands.

The arteries supplying the optic nerve and retina derive from branches of the ophthalmic artery. The central retinal artery approaches the eye along each optic nerve and pierces the inferior aspect of the dural sheath about 1 cm behind the globe to enter the center of the nerve. The artery emerges in the fundus at the center of the nerve head, from which it nourishes the inner two thirds of the retina by superior and inferior branches. Anastomotic branches derived from the choroidal and posterior ciliary arteries, the ciliary system, supply the choroid, optic nerve head, and outer retinal layers, including the photoreceptors. In about 10% of the population, the macula is supplied by a retinociliary artery, a branch of the ciliary system. Venous drainage from the retina and nerve head flows primarily through the central retinal vein, whose course of exit from the eye parallels that of entry of the artery.

What each eye "sees" is termed its *visual field* (Fig. 424-1). The nasal side of the left retina and the temporal side of the right see the left side of the world, and the upper half of each retina sees the lower half of the world. Behind the eyes, the optic nerves pass through the optic canal to form the optic chiasm. In the chiasm, nerves from the nasal half of each retina decussate and join the fibers from the temporal half of the contralateral retina. From the chiasm, the optic tracts pass around the cerebral peduncles to reach the lateral geniculate ganglia. The orientation of the visual field is rotated 90 degrees in the lateral geniculate such that images from the inferior visual field project to the medial half, whereas images from the superior visual field project to the lateral half. The geniculocalcarine radiation initially fans out into superolateral and inferolateral projections, the latter passing around the lateral ventricle and for a short distance into the temporal lobe (Meyer's loop) before turning posteriorly to reach the striate cortex of the occipital lobe. In the occipital lobe, the striate cortex (area 17) lies along the superior and inferior bands of the calcarine fissure, with macular fibers projecting most

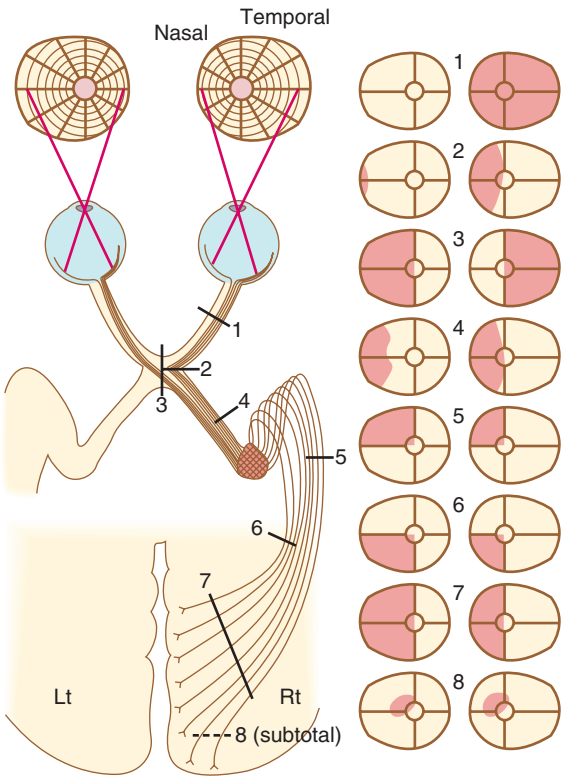

FIGURE 424-1. **Visual fields that accompany damage to the visual pathways.** *1,* Optic nerve: unilateral amaurosis. *2,* Lateral optic chiasm: grossly incongruous, incomplete (contralateral) homonymous hemianopia. *3,* Central optic chiasm: bitemporal hemianopia. *4,* Optic tract: incongruous, incomplete homonymous hemianopia. *5,* Temporal (Meyer's) loop of the optic radiation: congruous partial or complete (contralateral) homonymous superior quadrantanopia. *6,* Parietal (superior) projection of the optic radiation: congruous partial or complete homonymous inferior quadrantanopia. *7,* Complete parieto-occipital interruption of the optic radiation: complete congruous homonymous hemianopia with psychophysical shift of the foveal point, often sparing central vision and resulting in "macular sparing." *8,* Incomplete damage to the visual cortex: congruous homonymous scotomas, usually encroaching at least acutely on central vision.

TABLE 424-1 COMMON CAUSES OF TRANSIENT MONOCULAR VISION LOSS

CATEGORY (TYPICAL DURATION)	CAUSES	DIFFERENTIAL FEATURES
Thromboembolism (1-5 min)	Atherosclerosis	Other atherosclerotic vascular disease, associated crossed hemiparesis, angiography (carotid atheroma)
	Cardiac	Valvular disease, mural thrombi, atrial fibrillation, recent myocardial infarction
	Blood dyscrasia	Blood tests positive for sickle cell anemia, macroglobulinemia, multiple myeloma, polycythemia, other
Vasospasm (5-30 min)	Migraine	Ipsilateral headache, other classic aura, family history
Vascular compression (few seconds)	Increased intracranial pressure	Precipitated by position change, Valsalva maneuver, or pressure waves
	Tumor	Associated slowly progressive monocular visual loss
Vasculitis (1-5 min)	Temporal arteritis	Associated headache, polymyalgia rheumatica, palpable temporal artery, elevated sedimentation rate

posteriorly to the occipital pole and more peripheral retinal projections lying more anteriorly.

Localization of Lesions within Visual Pathways

Monocular vision loss is due to a lesion in one eye or optic nerve. Binocular visual loss, on the other hand, can result from disease located anywhere in the visual pathways from the corneas to the occipital poles. Lesions involving the optic chiasm produce nonhomonymous visual abnormalities (e.g., the bitemporal hemianopia illustrated by lesion 3 in Fig. 424-1). Optic tract abnormalities are comparatively rare but produce characteristic visual changes. The fibers serving identical points in the homonymous half fields do not fully commingle in the optic tract, so lesions damaging this structure produce incongruous homonymous hemianopia. Lesions of the geniculate nuclei, optic radiations, or visual cortex produce congruent hemianopic field defects that may go unrecognized unless the hemianopia intrudes on macular vision. Postgeniculate visual loss can be differentiated from pregeniculate visual loss by (1) a normal funduscopic appearance, (2) intact pupillary light reactions, and (3) appropriate lesions on brain imaging.

Examination of the Afferent Visual System

Visual function is most commonly assessed by "best-corrected visual acuity" (Chapter 423). If visual acuity is not normal, it must be determined whether acuity can be improved with lenses or at least with the use of a pinhole. The normal reference is recognition of letters at an idealized 20 feet, and acuity charts are designed with even larger letters that are normally recognized at proportionally greater distances. Thus, if one reads letters at 20 feet no better than those normally perceived at 40 feet, vision is recorded as 20/40. Small visual charts that are easily carried in the physician's case permit quick and fairly accurate bedside appraisal of acuity.

Visual fields can be tested at the bedside by confrontation, and rough estimates of their integrity can be made even in patients with reduced alertness. The fields should be tested individually for each eye because the pattern of visual field defects can provide important localizing information. A quick screen of the visual fields can be made by having the patient fixate on the examiner's nose and identify the number of fingers flashed in each of the four visual field quadrants. With practice and a cooperative subject, accurate confrontation fields can be obtained that outline even scotomas. Ophthalmoscopic examination permits direct visualization of the retina and optic disc. Corneal, lenticular, or vitreous opacities severe enough to produce visual symptoms can almost always be detected with the ophthalmoscope.

Common Causes of Visual Loss

Eye

The cause of monocular vision loss secondary to ocular and retinal lesions can often be detected by ophthalmoscopic examination or by measurement of intraocular pressure (Chapter 423). *Glaucoma* caused by impaired absorption of aqueous humor results in a high intraocular pressure that usually produces gradual loss of peripheral vision, "halos" seen around lights, and occasionally, pain and redness in the affected eye. The diagnosis is made by tonometric measurement of high intraocular pressure and may be suspected by palpating an abnormally firm globe and observing a deep, pale optic cup and attenuated blood vessels. *Retinal tears* and *detachments* give rise to unilateral distortions of the visual image seen as sudden angulations or curves of objects containing straight lines (metamorphopsia). *Hemorrhages* into the vitreous humor or infections or inflammatory lesions of the retina can produce scotomas that resemble those resulting from primary disease of the central visual pathway.

Binocular vision loss secondary to retinal disease in younger subjects is often due to *heredodegenerative conditions.* Vascular diseases, diabetes (Chapter 229), and age-related macular degeneration are causes in older patients. In most cases of *pigmentary retinal degeneration,* visual loss begins peripherally and slowly proceeds centrally. By contrast, *macular degeneration* (see Figs. 423-22 and 423-23) impairs central vision early in its course. A common variant in the complement factor H (*CFH*) gene is associated with a markedly increased risk for the development of age-related macular degeneration. Ranibizumab and bevacizumab are about equally effective treatment for neovascular age-related macular degeneration.

Optic Nerve

Acute or subacute monocular vision loss (Table 424-1) as a result of optic nerve disease is most commonly produced by demyelinating disorders, vascular obstruction, neoplasm, or hereditary optic neuropathy. Demyelinating disease of the nerve head (*optic neuritis* or *papillitis*) produces disc edema along with loss of central vision in the affected eye only; subjectively

unrecognized scotomas may sometimes be found in the other eye. Demyelination of the optic nerve behind the point where the retinal vein emerges (*retrobulbar neuritis*) initially leaves a normal-looking disc but a central or paracentral scotoma. With chronic demyelinating disorders, the optic disc becomes pale and atrophic. The clinical course and therapeutic response of optic neuritis depend on the underlying inflammatory mechanism. In more than 50% of patients initially seen with optic neuritis, typical symptoms and signs of multiple sclerosis eventually develop (Chapter 411). Optic neuritis caused by multiple sclerosis is not responsive to steroids, but optic neuritis related to systemic lupus erythematosus (Chapter 266), vasculitis (Chapter 270), or sarcoidosis (Chapter 95) may be steroid responsive.[1]

Optic neuritis with an associated transverse myelitis is the clinical hallmark of neuromyelitis optica, which is a severe demyelinating disease often mistaken for multiple sclerosis but now recognized to be caused by anti–aquaporin 4 autoantibodies. The recommended treatment options include rituximab (1 g infusions at an interval of 2 weeks) or azathioprine (3 mg/kg/day orally).[2] Doses should be adjusted based on response and immune suppression.

Intraocular arterial occlusion may produce either central visual loss or an altitudinal field defect (*ischemic optic neuropathy*). Tumors (Chapter 189) invading the optic nerve or space-occupying lesions compressing it anywhere between the orbit and chiasm cause gradually decreasing central vision or a sector defect of the peripheral visual field. With such chronic lesions, the affected optic nerve becomes visibly atrophic.

Acute binocular vision loss resulting from bilateral optic nerve disease is most often caused by demyelinating disease or by toxic (methanol, tobacco, isoniazid) or nutritional factors (B vitamin deficiency, particularly of thiamine) (Chapter 416). In younger persons and those lacking a clear history of toxic exposure, demyelinating lesions overwhelmingly predominate. Symptoms are of abrupt or subacute onset with visual blurring, which may progress rapidly to blindness within hours or days. There may be pain about the eyes, particularly with movement. *Leber's optic neuropathy,* caused by a mutation in mitochondrial DNA, typically begins painlessly and centrally in one eye, with the second eye affected weeks to months later.

Papilledema is disc edema secondary to increased intracranial pressure (Table 424-2).[3] Vision is normal except under one of two circumstances: (1) acute transient episodes of amaurosis lasting a few seconds and attributable to acute increases in intracranial pressure (plateau waves) and (2) progressive loss of peripheral vision with long-standing, severe papilledema caused by compression of the optic nerve head. Idiopathic intracranial hypertension (Chapter 189) is commonly seen in overweight women of childbearing age. Subacute or chronic binocular vision loss secondary to optic nerve disease can result from *toxic* and *nutritional* causes or from *inherited optic atrophy.* The latter sometimes accompanies spinocerebellar degeneration but may selectively affect the optic nerve. With either cause, visual loss is painless and primarily affects central vision; ophthalmoscopy shows optic atrophy.

Chiasm and Optic Tract

Patients with lesions of the optic chiasm or optic tract are often unaware of visual impairment until the deficit encroaches on central vision in one or both eyes. Intrinsic or extrinsic neoplasms and parachiasmal arterial aneurysms are the most common lesions in this location. Gliomas that arise within the chiasm or optic tract are rare in adulthood. Extrinsic lesions compressing the chiasm or tract include *pituitary adenomas* (Chapter 224), *dysgerminomas, craniopharyngiomas, meningiomas* (Chapter 189), and large *aneurysms* of the carotid or basilar artery (Chapter 408). The diagnosis rests on finding the characteristic visual field abnormalities (bitemporal hemianopia for chiasm and incongruous homonymous hemianopia for optic tract lesions) and identifying the lesion with computed tomography or magnetic resonance imaging. Pituitary apoplexy secondary to acute hemorrhage into the gland (Chapter 224) can result in sudden vision loss; prompt neurosurgical intervention under steroid coverage is required for most patients.

TABLE 424-2 DIFFERENTIATION OF OPTIC NEURITIS FROM PAPILLEDEMA

	OPTIC NEURITIS	PAPILLEDEMA
Central-cecocentral vision loss	Present	Absent
Distribution	Usually unilateral	Usually bilateral
Ocular pain on movement	Present	Absent
Direct light reflex	±Reduced	Intact
CT and MRI of head	White matter plaques	Tumor, venous occlusion, etc.
Visual evoked responses	Abnormal	Normal
Lumbar puncture pressure	Normal	Elevated

CT = computed tomography; MRI = magnetic resonance imaging.

Visual Radiations and Occipital Cortex

Lesions involving the postgeniculate visual pathways most often result from *vascular damage, traumatic injuries, neoplasms,* or rarely, *inflammatory* or *degenerative disorders* involving the cerebral white matter. Their localization can be deduced by the resulting visual field defects. Vascular disease of the occipital lobes is the most common cause of homonymous visual field defects in middle-aged and elderly people. *Anton's syndrome* refers to cerebral visual loss with denial of a visual defect. Affected patients not only deny that they are blind but also confabulate details of their visual environment from memory. Anton's syndrome results from bilateral lesions involving the parieto-occipital lobes or in the setting of metabolic encephalopathy. The reversible *posterior leukoencephalopathy syndrome,* which is characterized by headache, seizures, confusion, and cortical visual loss, is associated with an abrupt increase in blood pressure, such as may be seen with eclampsia and with immunosuppressive therapy after transplantation.

PUPILLARY CONTROL

The neuromechanisms that control pupil size and reactivity are complex, yet they can be evaluated by simple clinical procedures. The diameter of the pupil is determined by the antagonistic actions of the iris sphincter and dilator muscles, with the latter playing a minor role. If the sphincter muscle is severed or ruptured, it does not retract toward one quadrant but rather continues to function, except in the altered segment. Therefore, pupillary response can be evaluated even in the presence of significant damage to the iris.

Anatomy and Localization of Lesions within Pupillary Pathways

The size of the pupil is governed by tonic balance between sympathetic and parasympathetic innervation of the muscles of the iris.[4] Sympathetic stimulation dilates the pupil, whereas parasympathetic stimulation constricts it. In the normal resting state, light entering the eye provides the major stimulus governing the size of the pupil (Fig. 424-2). Light activates the retinal rods and cones, with maximal sensitivity in the macular area. The optic nerve fibers follow the crossed and uncrossed visual pathways to the pregeniculate portion of the optic tracts, where the receptor fibers for light diverge to the pretectal nucleus located at the midbrain-diencephalic junction. Interneurons project from this nucleus to the Edinger-Westphal nuclei atop the midbrain third nerve nuclear complex of either side. From that point, paired parasympathetic efferents leave the midbrain in the third nerves, travel in the interpeduncular space across the petroclinoid ligament and edge of the tentorium, traverse the cavernous sinus, and then enter the orbit through the superior orbital fissure. In the orbit, the parasympathetic efferents synapse in the ciliary ganglion, from which ciliary nerves enter the eye to reach the pupillary muscles.

The principal sympathetic control of the pupil originates in the ventral lateral hypothalamus (first-order neuron), from which fibers descend ipsilaterally through the brain stem tegmentum and thence to the cervical cord, where they synapse with preganglionic neurons in the intermedial lateral column of the upper three thoracic segments. Preganglionic fibers (second-order neurons) emerge with the ventral roots of C8, T1, and T2 and ascend in the neck to synapse in the superior cervical ganglion adjacent to the base of the skull. Postganglionic (third-order neurons) pupillary fibers accompany the internal carotid artery through the skull and then leave it to follow the ophthalmic branch of the trigeminal nerve to reach the pupillodilator muscle of the eye.

Examination of the Pupil

The pupillary response to light should be examined in a dimly lighted room, where the pupils are naturally dilated. First, the size and symmetry of the pupils are assessed by shining a dim light onto the face from below so that both pupils are seen simultaneously in the indirect illumination. To test light reactivity, gaze is directed at a distant object (so that constriction secondary to convergence is minimal), and first one and then the other pupil is illuminated with a bright light source. If a pupil reacts poorly to direct light, it is

observed as the opposite eye is illuminated (consensual response). Pupils that react poorly to light should be tested for reactivity to the near reflex by first having the patient gaze at a distant object and then quickly fixate on an object just in front of his or her nose. *Light-near dissociation* refers to a pupil that does *not* react to light but does accommodate by constricting to a near target.

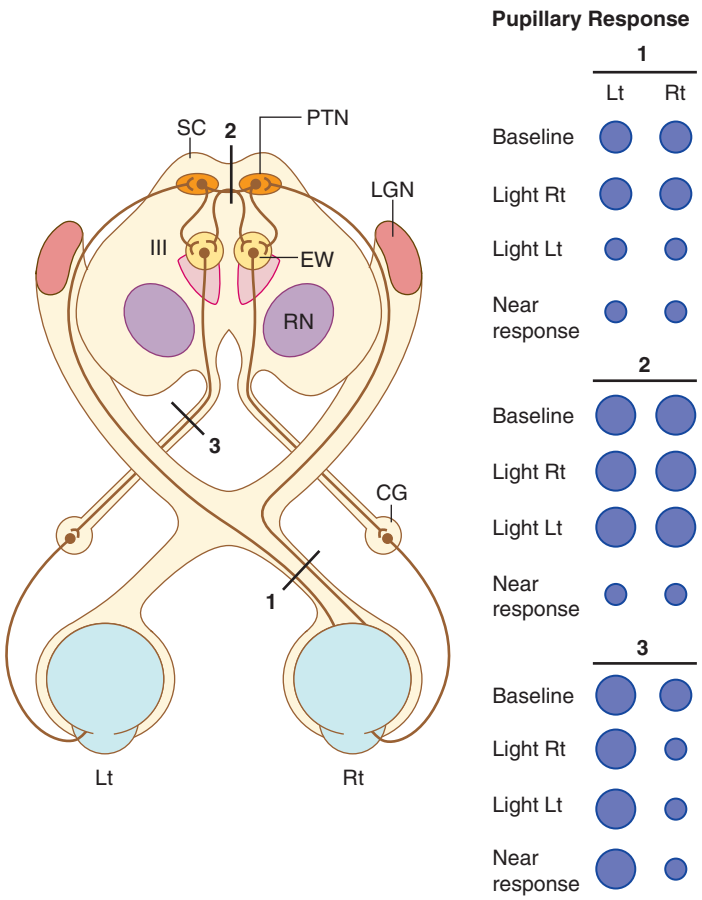

FIGURE 424-2. **Pupillary responses associated with lesions of the *(1)* optic nerve, *(2)* pretectum, and *(3)* oculomotor nerve.** Baseline is obtained with fixation on a distant target and the near response with a target in front of the nose. CG = ciliary ganglion; EW = Edinger-Westphal nucleus; LGN = lateral geniculate nucleus; PTN = pretectal nucleus; RN = red nucleus; SC = superior colliculus.

Common Causes of Pupillary Abnormalities

With so-called benign pupillary dilation or *physiologic anisocoria,* there is a long-standing difference in the size of the two pupils with normal reflex reactions; the disparity remains constant during constriction and dilation. Lesions compressing or damaging the pretectal region interrupt the afferent light reflex bilaterally to produce dilated and light-fixed pupils (e.g., lesion 2; see Fig. 424-2).[5] Pupillary constriction to the near response is preserved until late stages. Tumors of the pineal gland (e.g., dysgerminomas) and *localized infarctions* are the most common lesions in this location. *Adie's tonic pupil* (Fig. 424-3) is a medium to large (3 to 6 mm) pupil that constricts little or not at all to light and very slowly to accommodation but constricts with the instillation of dilute (0.125%) pilocarpine (Fig. 424-4). The condition usually affects one eye (occasionally both), is more common in women 25 to 45 years of age, and carries no serious implications. It most likely results from postviral denervation of the pupillary muscles. Unexplained unilateral or bilateral dilated pupils as an isolated finding can result from *accidental or intentional instillation of mydriatic drugs.* Transdermal scopolamine is a common cause. Failure of the pupil to constrict promptly with pilocarpine (1%) gives the diagnosis if the history is unclear. Interruption of the emerging third nerve in the ventral midbrain or along the proximal part of its course produces a dilated pupil 6 to 7 mm in diameter. Important causes of compression of the third nerve in this region are *aneurysms* (Chapter 408), *neoplasia* (Chapter 189), and *brain herniation* (Chapter 189) as a result of increased intracranial pressure. In nearly all cases, the pupillary involvement is associated with other signs of third nerve involvement (see later text).

Sympathetic paralysis of the eye with ptosis, anhidrosis, and miosis (Horner's syndrome; Fig. 424-5) can result from lesions anywhere along the pathway of the sympathetic innervation to the eye (Table 424-3). The diagnosis can sometimes be made by identifying associated signs in the

FIGURE 424-3. **Adie's tonic pupil in the right eye of a young woman.** The affected pupil is "tonic"; that is, it responds slowly to light and accommodation but on rapid testing appears unresponsive. The site of the lesion is usually obscure, but the condition is benign. There may be associated areflexia. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine,* 3rd ed. London: Mosby; 2003.)

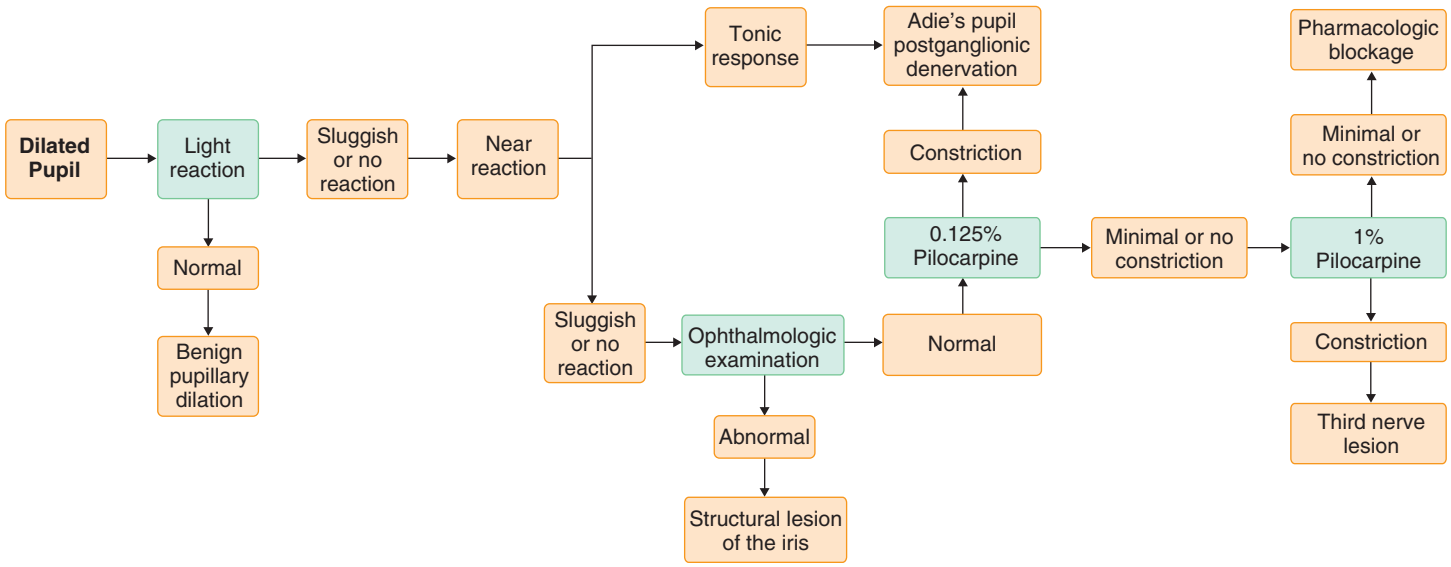

FIGURE 424-4. **Use of pilocarpine to help differentiate between different causes of a dilated pupil.**

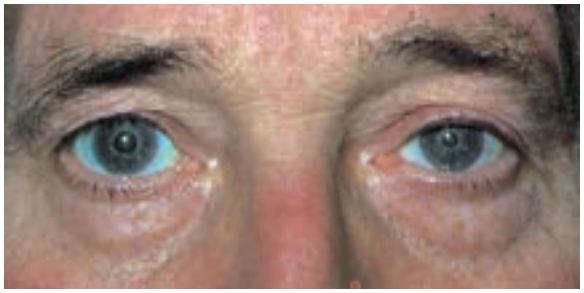

FIGURE 424-5. Horner's syndrome. Note the characteristic ptosis of the left eye associated with constriction of the pupil (miosis). This patient had syringomyelia, but Horner's syndrome has many possible causes. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine,* 3rd ed. London: Mosby; 2003.)

TABLE 424-3 HORNER'S SYNDROME RESULTS FROM LESIONS IN MULTIPLE LOCATIONS

LOCATION OF LESION	NEURON INVOLVED	TYPE OF LESION	ASSOCIATED SYMPTOMS AND SIGNS
Lateral brain stem	1st order	Infarction, glioma	Vertigo, nystagmus, imbalance, numbness, weakness
Apex of lung	2nd order	Lung cancer, trauma	Often none
Neck	3rd order	Carotid dissection or inflammation	Pain, monocular visual loss, hemiparesis

brain stem or neck or along the carotid artery. *Argyll Robertson pupils* are small (1 to 2 mm), unequal, irregular, and fixed to light; they constrict minimally to accommodation. Their principal cause is tertiary neurosyphilis (Chapter 319).

OCULOMOTOR CONTROL

Abnormal eye movements can result from disturbances at several levels. Disconjugate eye movements result from lesions in the individual ocular muscles, the myoneural junctions, the oculomotor nerves and their three paired nuclei in the brain stem, and the internuclear medial longitudinal fasciculus (MLF), which yokes the eyes in horizontal movements. Supranuclear lesions typically produce disorders of conjugate gaze (gaze palsies).

Anatomy and Localization of Lesions within the Oculomotor Pathways

Nuclear and Internuclear Pathways

The abducens (sixth) nerve supplies the lateral rectus muscle. Selective involvement of the abducens nerve anywhere along its pathway leads to isolated weakness of abduction of the affected eye. Destruction of the abducens nucleus in the brain stem results in a conjugate gaze paralysis (ipsilateral) because, in addition to oculomotor neurons, the nucleus contains interneurons destined for the contralateral medial rectus nucleus. The trochlear (fourth) nerve supplies the contralateral superior oblique muscle, which turns in and depresses the eye. Patients with superior oblique weakness note an increase in diplopia with head tilt toward the side of weakness and often tilt their head in the opposite direction. At rest, there is slight upward deviation of the involved eye, and downward movement is impaired when the affected eye is turned in. Patients typically complain of diplopia when reading or going down stairs. The third (oculomotor) cranial nerve supplies the remaining ocular muscles. Involvement of the third nerve nucleus in the midbrain always produces at least some bilateral oculomotor weakness; the superior rectus division of the nucleus supplies the contralateral superior rectus muscle (all other divisions supply ipsilateral muscles). Peripheral third nerve paralysis can result from lesions damaging the structure anywhere from its course within the ventral midbrain to where it enters the orbit through the superior orbital fissure. When complete, third nerve palsy produces a widely dilated pupil, severe ptosis, and an externally deviated eye held in position by unopposed contraction of the lateral rectus muscle. In such conditions, the continued trochlear action reveals itself by intorsion of the eye when the subject attempts to look down.

TABLE 424-4 SUPRANUCLEAR OCULOMOTOR CONTROL SYSTEMS

SYSTEM	DESCRIPTION	FUNCTION	KEY ANATOMIC STRUCTURES
Saccade	Fast conjugate voluntary movements	Move the fovea to a new target of interest	Frontal eye field, superior colliculus, pretectum—vertical, pontine—horizontal cerebellar dentate nucleus
Smooth pursuit/ optokinetic	Slow conjugate tracking	Match eye velocity to target velocity	Occipital, parietal, pons, cerebellar flocculus
Vestibulo-ocular	Slow conjugate compensatory eye movements	Keep eyes stable when head moves	Pontomedullary region
Vergence	Slow disconjugate tracking	Focus on near and far targets	Pretectum, tectum

The MLF interconnects the abducens nucleus in the pons with the contralateral oculomotor nuclear complex in the midbrain. It terminates cephalad in the interstitial nucleus in the rostral midbrain and can be traced as far caudad as the thoracocervical region of the spinal cord (coordinating nuchal-ocular control). Lesions involving the MLF characteristically produce internuclear ophthalmoplegia, in which the eyes are conjugate in the primary position but disconjugate on lateral gaze. With fully developed internuclear ophthalmoplegia on lateral gaze away from the side of the lesion, the contralateral eye abducts and shows nystagmus, whereas the ipsilateral adducting eye does not move nasally because of failure of ascending impulses to reach the medial rectus division of the third nerve nucleus. Adduction for convergence is usually relatively maintained.

Supranuclear Pathways

Pathways descending from the frontal eye fields in the frontal lobe through the superior colliculi to the contralateral brain stem regulate rapid voluntary eye movements (*saccades*) (Table 424-4).[6] Pathways descending from the parieto-occipital and frontal regions to the ipsilateral brain stem subserve slow visual tracking (smooth pursuit—foveal target; optokinetic—full-field target). For the vestibulo-ocular reflex, primary afferent neurons in the inner ear synapse with neurons in the vestibular nuclei, which in turn synapse with appropriate oculomotor neurons to produce compensatory eye movements. The *convergence* center is located in the rostral-dorsal midbrain near the vertical gaze center.

Examination of Eye Movements

Fixation and gaze holding are tested by having the patient look center, right, left, up, and down. Each position should be held steady and unwavering with the observer carefully documenting abnormal movements or ocular disconjugacies. Each supranuclear oculomotor control system is examined separately. *Saccades* are tested by having the patient alternately fixate on two targets such as the examiner's finger and nose; the speed and accuracy are noted. *Smooth pursuit* is tested by slowly moving a target back and forth and up and down and observing the patient's ability to produce smooth tracking movements. If the target velocity is low, normal subjects should be able to pursue without requiring catch-up saccades. The *vestibulo-ocular reflex* is evaluated with the head-thrust test (Chapter 428). *Convergence* is tested by having the patient follow a target moving from far to near. The degree of convergence depends to some extent on the cooperation of the patient. A clear sign that the patient is attempting to converge is simultaneous pupillary constriction.

Common Causes of Abnormal Oculomotor Control

Strabismus (Ocular Misalignment)

A comitant (same in all directions of gaze) strabismus present since childhood is usually a benign *congenital disorder.* Latent congenital strabismus can become manifested in adulthood in association with a systemic illness. An acquired skew deviation (vertical displacement of the ocular axes) indicates a lesion within the otolith-ocular pathways (generally the brain stem). Incomitant strabismus can result from restrictive disease of the orbit or from

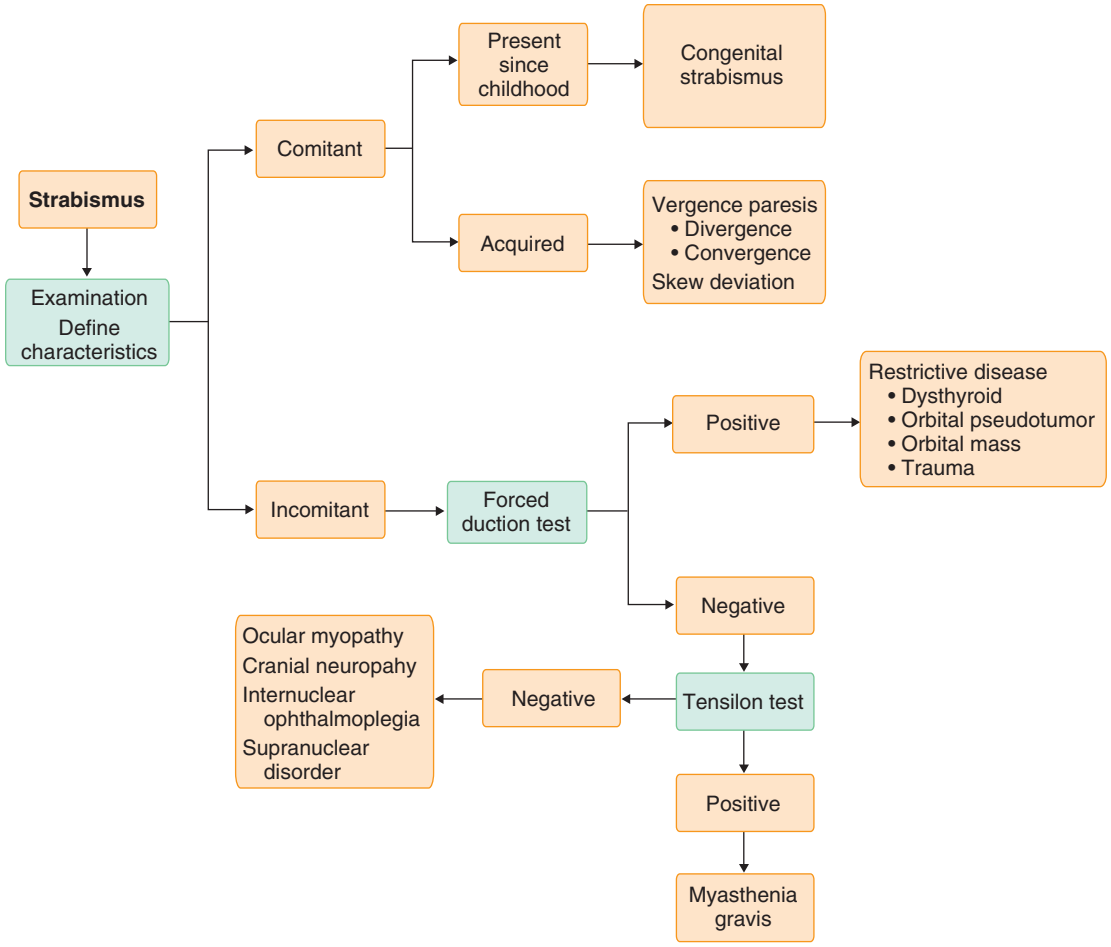

FIGURE 424-6. Diagnostic tests that help differentiate among common causes of strabismus.

abnormal muscle or oculomotor nerve function. The presence of mechanical restriction is confirmed by the use of forced duction testing (Fig. 424-6) (After a topical anesthetic is applied to the eye, the ophthalmologist grasps the muscle insertion with large blunt-toothed forceps. Failure of the eye to deviate fully in the pulled direction implies restriction.) Common causes of *orbital restrictive disease* include dysthyroid ophthalmopathy (Chapter 226), orbital pseudotumor, trauma, and orbital mass lesions (Chapter 423). Variable strabismus that increases with fatigue suggests *myasthenia gravis* (Chapter 422). A Tensilon test can usually confirm the diagnosis (see Fig. 424-6). If both restrictive disease and myasthenia gravis (Chapter 422) have been excluded, most patients with incomitant strabismus have processes affecting the oculomotor nuclei, their fascicles, or the cranial nerves themselves. Common causes of an *isolated third nerve palsy* in an adult include aneurysm (Chapter 408), small-vessel occlusive disease (including diabetes mellitus (Chapter 229), trauma (Chapter 399), and neoplasm. Typically, third nerve lesions secondary to vascular disease spare the pupil. Vascular disease and trauma are by far the most common causes of *isolated trochlear nerve palsy.* The abducens nerve is particularly vulnerable to isolated traumatic involvement because of its long pathway outside the brain stem. Lesions that produce increased intracranial pressure (Chapter 189) can lead to abducens nerve dysfunction regardless of the location and produce a "false localizing sign." Other common causes of *isolated sixth nerve palsy* are vascular disease (Chapter 407), trauma (Chapter 399), and neoplasm. About one fourth of cases of cranial nerve palsy (third, fourth, or sixth nerves) remain undiagnosed.

Internuclear Ophthalmoplegia

Internuclear ophthalmoplegia (Fig. 424-7) may be unilateral or bilateral, partial or complete, depending on the location of the lesion and the degree of damage to the MLF. *Demyelinating* and small *vascular lesions* are the most common causes of unilateral internuclear ophthalmoplegia unaccompanied by other ocular palsies or brain stem signs. Larger brain stem lesions that damage one or more oculomotor nuclei plus the MLF often produce

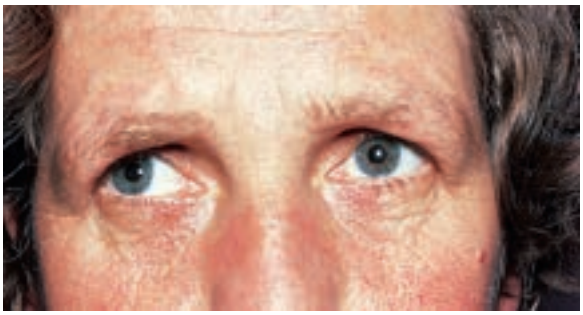

FIGURE 424-7. Internuclear ophthalmoplegia may be an initial feature of brain stem involvement in multiple sclerosis. On lateral gaze to the right, adduction of the left eye is incomplete. On convergence, eye movement was normal. The lesion is in the left medial longitudinal bundle, between the nucleus in the pons and the third nerve nucleus on the opposite side. (From Forbes CD, Jackson WF. *Color Atlas and Text of Clinical Medicine,* 3rd ed. London: Mosby; 2003.)

combinations of disconjugate eye movements coupled with nuclear oculomotor palsies. Myasthenia gravis (Chapter 422) can produce an ophthalmoparesis resembling internuclear ophthalmoplegia as a result of greater involvement of the medial rectus than the lateral rectus. Demyelinating diseases (Chapter 411) are the most common causes of bilateral internuclear ophthalmoplegia involvement.

Disorders of Conjugate Gaze

Acute lesions involving a frontal eye field (e.g., hemorrhage or infarction [Chapters 407 and 408]) result in a transient inability to direct the eyes contralaterally. Vertical eye movements are not affected by unilateral lesions. Bilateral damage to the frontal eye fields or their descending pathways may produce an inability to move the eyes voluntarily (horizontal or vertical)

despite preserved reflex eye movements, a condition called *oculomotor apraxia.* Lesions involving the horizontal gaze center in the pons produce an ipsilateral paralysis of conjugate gaze and tonic deviation of the eyes to the contralateral side (Chapter 423). Lesions of the pretectum selectively impair vertical gaze, with the vertical upgaze center being slightly rostral and dorsal to the vertical downgaze center. Patients with the *dorsal midbrain syndrome* (Parinaud's syndrome) have a conjugate upgaze paresis. When they attempt to make upward saccades, convergence retraction nystagmus develops. As noted earlier, impaired convergence and light-near dissociation of the pupillary reflexes are also part of the syndrome. The most common causes of the dorsal midbrain syndrome include tumors of the pineal gland (Chapter 223) (dysgerminomas), hydrocephalus (Chapter 189), and localized infarction. With the so-called locked-in syndrome (secondary to basilar artery thrombosis [Chapter 404]), voluntary horizontal eye movements are absent; the patient's only remaining motor functions are vertical eye and lid movements.

Nystagmus

Spontaneous nystagmus can be congenital or acquired. *Congenital nystagmus* typically has a high frequency and variable waveform (usually pendular) and is highly fixation dependent. It generally remains horizontal in all positions of gaze. The lifelong history and lack of symptoms confirm the diagnosis. Spontaneous nystagmus resulting from a *peripheral vestibular* lesion (i.e., in the labyrinth or vestibular nerve) usually has combined horizontal and torsional components (Table 424-5). The nystagmus resolves within a few days of the acute lesion. Acquired persistent spontaneous nystagmus indicates a lesion in the brain stem or cerebellum, or both. The latter is often purely vertical, horizontal, or torsional. Spontaneous *downbeat nystagmus* is commonly seen with lesions of the cerebellum or cervicomedullary junction (e.g., Arnold-Chiari malformation [Chapter 417]).

Gaze-evoked nystagmus is always in the direction of gaze and is usually present with and without fixation. It is most commonly produced by the ingestion of *drugs* such as phenobarbital, phenytoin, alcohol, and diazepam (Chapter 110). It can also occur in patients with such varied conditions as myasthenia gravis (Chapter 422), multiple sclerosis (Chapter 411), and cerebellar atrophy. Asymmetrical horizontal gaze-evoked nystagmus is caused by a structural brain stem or cerebellar lesion (particularly at the cerebellopontine angle), with the lesion generally being on the side of the larger amplitude nystagmus (Bruns' nystagmus). *Rebound nystagmus* is a type of gaze-evoked nystagmus that either disappears or reverses direction as the eccentric gaze position is held. When the eyes are returned to the primary position, nystagmus occurs in the direction of the return saccade. Rebound nystagmus occurs in patients with cerebellar atrophy and focal structural lesions of the cerebellum; it is the only variety of nystagmus thought to be specific for cerebellar involvement. *Disconjugate gaze–evoked nystagmus* most commonly results from lesions of the MLF (see earlier discussion), but it can also occur with other lesions of the brain stem involving the oculomotor nuclei. Positional nystagmus is discussed in Chapter 428.

Other Ocular Oscillations

Ocular bobbing consists of a fast conjugate downward eye movement followed by a slow return to the primary position. The phenomenon accompanies severe displacement or destruction of the pons or, less often, metabolic central nervous system depression. *Ocular myoclonus* consists of continuous rhythmic pendular oscillations, most often vertical, at a rate of 1 to 3 beats per second; it often accompanies palatal myoclonus and has a similar pathogenesis. *Square-wave jerks* and *ocular flutter* consist of brief, intermittent, horizontal oscillations (back-to-back saccades) arising from the primary gaze position. These types of ocular oscillation are most commonly seen with cerebellar disease but can also accompany more diffuse central nervous system disorders. *Opsoclonus* consists of rapid, chaotic, conjugate, repetitive saccadic eye movements (dancing eyes). Opsoclonus accompanies the cerebellar dysfunction, with the most chaotic varieties associated with brain stem encephalitis or the remote effects of systemic neoplasm, especially neuroblastoma in children. *Ocular dysmetria* refers to overshooting and undershooting of saccadic eye movements, often followed by multiple attempts at refixation. It reflects cerebellar dysfunction.

TABLE 424-5 KEY DISTINGUISHING FEATURES OF PERIPHERAL AND CENTRAL TYPES OF SPONTANEOUS AND POSITIONAL NYSTAGMUS

TYPE OF NYSTAGMUS	PERIPHERAL (END ORGAN AND NERVE)	CENTRAL (BRAIN STEM AND CEREBELLUM)
Spontaneous	Unidirectional, fast phase away from the lesion, combined horizontal torsional, inhibited with fixation	Bidirectional or unidirectional; often pure horizontal, vertical, or torsional; *not* inhibited with fixation
Static positional	Fixed or changing direction, inhibited with fixation	Fixed or changing direction, *not* inhibited with fixation
Paroxysmal positional	Vertical-torsional, occasionally horizontal-torsional, vertigo prominent, fatigability, latency	Often pure vertical, vertigo less prominent, no latency, nonfatigable

Grade A Reference

A1. Chakravarthy U, Harding SP, Rogers CA, et al. Ranibizumab versus bevacizumab to treat neovascular age-related macular degeneration: one-year findings from the IVAN randomized trial. *Ophthalmology*. 2012;119:1399-1411.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bennett JL, Nickerson M, Costello F, et al. Re-evaluating the treatment of acute optic neuritis. *J Neurol Neurosurg Psychiatry*. 2014;[Epub ahead of print].
2. Trebst C, Jarius S, Berthele A, et al. Update on the diagnosis and treatment of neuromyelitis optica: recommendations of the Neuromyelitis Optica Study Group (NEMOS). *J Neurol*. 2014; 261:1-16.
3. Lee AG, Wall M. Papilledema: are we any nearer to a consensus on pathogenesis and treatment? *Curr Neurol Neurosci Rep*. 2012;12:334-339.
4. Weber KP, Straumann D. Neuro-ophthalmology update. *J Neurol*. 2014;261:1251-1256.
5. Caglayan HZ, Colpak IA, Kansu T. A diagnostic challenge: dilated pupil. *Curr Opin Ophthalmol*. 2013;24:550-557.
6. Shaikh AG, Ghasia FF. Physiology and pathology of saccades and gaze holding. *Neurorehabilitation*. 2013;32:493-505.

REVIEW QUESTIONS

1. Lesions in which of the following structures could not produce a hemianoptic field defect?

A. Parietal lobe
B. Occipital lobe
C. Temporal lobe
D. Geniculate ganglion
E. All of the above

Answer: E Lesions in all of these structures can involve the optic radiations and produce a hemianoptic defect.

2. Which of the following would be atypical for retrobulbar neuritis?

A. Loss of visual acuity
B. Swelling of the optic disc
C. Atrophy of the optic disc
D. Color saturation
E. Later development of multiple sclerosis

Answer: B The optic disc looks normal with acute retrobulbar neuritis, but the disc can later become pale and atrophic. Monocular loss of visual acuity and color saturation are common, and more than 50% of patients will eventually be diagnosed as having multiple sclerosis.

3. Which of the following would be atypical for papilledema?

A. Enlargement of the blind spot
B. Loss of visual acuity
C. Gradual loss of peripheral vision
D. Brief episodes of amaurosis
E. None of the above

Answer: B Visual acuity is nearly always maintained with papilledema. A, C, and D all occur with papilledema.

4. Which of the following is not associated with Horner's syndrome?

A. Lateral medullary infarction
B. Lung cancer
C. Carotid artery dissection
D. Anterior communication artery aneurism
E. None of the above

Answer: D Anterior communicating artery aneurysms can cause an enlarged pupil by compressing the parasympathetic innervation of the pupil, which is carried in the third nerve. The other three conditions all can damage the sympathetic innervation of the pupil, thereby causing Horner's syndrome.

5. A 35-year-old woman with known multiple sclerosis presents with new-onset double vision. Which of the following findings would be unlikely on neurologic examination?

A. Nystagmus of the abducting eye on lateral gaze
B. Impaired adduction on lateral gaze
C. Atrophy of the left optic disc
D. Enlarged pupil on the left side
E. Loss of visual acuity on the left side

Answer: D A and B are features of internuclear ophthalmoplegia, which is a common cause of double vision in a patient with multiple sclerosis. Unilateral optic neuritis with visual loss and development of optic atrophy would be common additional findings in patients with known multiple sclerosis. The pupils are symmetrical in patients with optic neuritis because the light reflex is bilateral. Involvement of the parasympathetic control of the pupil would be unusual with multiple sclerosis.

425

DISEASES OF THE MOUTH AND SALIVARY GLANDS

TROY E. DANIELS AND RICHARD C. JORDAN

More than 200 primary lesions or diseases occur in the oral mucosa, gingiva, teeth, jaws, and minor or major salivary glands.[1] In addition, secondary abnormalities of the oral mucosa or salivary glands can be caused by systemic diseases or drugs. The most common or important of these diseases may be observed during physical examination and may be part of a systemic process.

ORAL MUCOSAL DISEASES

Acute Ulcerations

Painful short-term ulcerations can be caused by mechanical trauma, immunologic mechanisms, or bacterial or viral infections (Table 425-1). Soon after formation, oral mucosal ulcers become covered by a white to gray pseudomembrane, analogous to scabs on dry epidermis. Pseudomembrane-covered ulcers are distinguished from white hyperkeratotic lesions by their clinical features of pain, a flat surface, and an erythematous periphery. Traumatic ulcers are characteristically located on the tongue or inside the cheeks or lips, are close to the chewing surfaces of the teeth, and have irregular borders.

APHTHOUS STOMATITIS ("CANKER SORES")

These idiopathic recurrent ulcers, which afflict up to 20% of the population, are found on all nonkeratinized areas of the oral mucosa except the hard palate, gingiva, and vermilion, which are keratinized (Fig. 425-1).[2] They form well-defined circular lesions that may be single or multiple. There are three clinical forms: (1) minor, which are flat and less than 1 cm in diameter and last 5 to 10 days; (2) major, which have raised borders, are greater than 1 cm, and often last for weeks or months; and (3) herpetiform, which are usually clusters of very small ulcers that resemble recurrent herpetic lesions but are not preceded by vesicles and do not occur on keratinized mucosa. A viral or bacterial pathogenesis has not been established for any of these forms. Lesions clinically identical to minor aphthae occur in Behçet's syndrome (Chapter 270). Aphthae are occasionally associated with anemias or gluten-sensitive enteropathy and may become more frequent and severe in association with human immunodeficiency virus (HIV) infection (Table 425-2).

TABLE 425-1 ORAL MUCOSAL ULCERS

TYPE/DISEASE	CLINICAL FEATURES
INSIDIOUS ONSET, CHRONIC	
Multiple or Bilateral	Shallow ulcers on mucosa, skin, or both
Pemphigus vulgaris	Begin as short-duration blisters
Mucous membrane pemphigoid	Begin as short-duration blisters
Lichen planus	Bilaterally symmetrical lesions (associated with hyperkeratosis and/or erythema)
Lupus erythematosus	Asymmetrical lesions, with or without systemic lupus (associated with hyperkeratoses and/or erythema)
Drug reaction	Variable lesions; appropriate history of drug use (e.g., penicillamine, gold)
Epidermolysis bullosa	Begin as blisters; lifelong history
Solitary	Indurated or cratered ulcers
Squamous cell carcinoma	Most common on tongue, oropharynx, lip, mouth floor
Adenocarcinomas, various	Most commonly on palate, cheeks, mouth floor
Tuberculosis	Usually painful
Actinomycosis	Often associated with draining sinus
Deep mycoses (particularly histoplasmosis, coccidioidomycosis)	Associated with systemic infection
Midline granuloma	Associated with necrosis, may perforate palate
Underlying osteonecrosis	Associated with prior cancer radiation therapy or bisphosphonate use
ACUTE ONSET, OFTEN SELF-LIMITING	
Clusters	Usually small and shallow ulcers; history of blisters
Primary herpes simplex	Any oral mucosal site, associated with fever, malaise
Recurrent herpes simplex	Only on gingiva, hard palate, or lip (keratinized mucosa)
Varicella-zoster	Unilateral lesions along neural distribution
Herpangina	Usually on oropharynx
Measles (rubeola)	Precede skin rash; associated with fever, malaise
Solitary or Multiple (without Clustering)	Variable, usually without history of blisters
Traumatic ulcers	Usually solitary; history of trauma
Aphthous stomatitis (canker sores)	Circular, often multiple, only on nonkeratinized mucosa
Behçet's syndrome	Oral lesions similar to recurrent aphthae
Erythema multiforme	Multiple lesions, often involve lower labial mucosa; can be recurrent or chronic
Drug reaction	Appropriate history of drug use
Necrotizing sialometaplasia	Deep ulcers, usually on palate
Primary syphilis	Solitary, indurated, painless, any site
Gonorrhea	Painful, surrounded by erythema, any site

TABLE 425-2 ORAL LESIONS ASSOCIATED WITH HUMAN IMMUNODEFICIENCY VIRUS INFECTION

Kaposi's sarcoma (human herpesvirus type 8)
Candidiasis (pseudomembranous, hyperplastic and/or erythematous lesions)
Other opportunistic fungal infections (e.g., histoplasmosis or coccidioidomycosis)
Aphthous ulcers (increased frequency, duration, or size)
Virus-associated epithelial hyperplasias Hairy leukoplakia (Epstein-Barr virus) Oral wart (human papillomavirus type 11 and other types) Focal epithelial hyperplasia (Heck's disease) (human papillomavirus types 13 and 32) Condyloma acuminatum (human papillomavirus types 6 and 11)
Herpes zoster (varicella-zoster virus)
Exaggerated forms of gingivitis and inflammatory periodontal disease
Decreased salivary gland function
Parotid gland enlargement (lymphoepithelial lesion)
Non-Hodgkin's lymphoma (e.g., plasmablastic lymphoma)

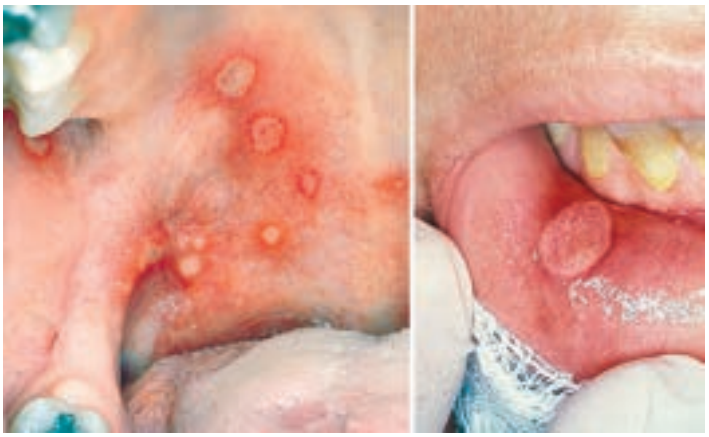

FIGURE 425-1. Aphthous ulcers. *Left,* A cluster of minor aphthae on the soft palate and buccal mucosa, present about 1 week. *Right,* A major aphthous ulcer on the labial mucosa, present about 3 weeks.

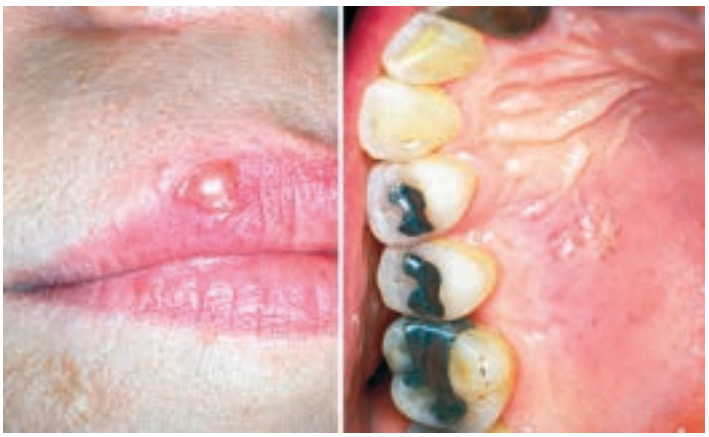

FIGURE 425-2. Clusters of recurrent herpes simplex vesicles. *Left,* on the lip; *right,* on the hard palate, both present 2 to 3 days, in different patients.

Minor or herpetiform aphthous ulcers may not require treatment unless they occur frequently. Topical steroids, such as fluocinonide gel or ointment, can reduce the severity and duration of the lesions only if applied with prodromal symptoms or earliest signs. A suspension of tetracycline or doxycycline in water used as a mouth rinse at the onset of symptoms also reduces the severity and duration of disease. None of these treatments prevent future lesions. Major aphthae usually require treatment with prednisone (e.g., 40 mg daily for 3 days); failure to respond significantly should prompt incisional biopsy to exclude neoplasia. Unfortunately, no treatment will cure a patient of recurrent aphthous stomatitis.

VIRAL ULCERS

Several types of virus (most commonly herpes simplex type 1; Chapter 374) cause multiple oral mucosal vesicles that last only a few hours or days and then become irregular shallow ulcers. In the initial infection by herpes simplex virus, usually in children, numerous vesicles may appear on any oral mucosal site (primary herpetic gingivostomatitis), accompanied by malaise, headache, fever, and cervical lymphadenopathy. Patients previously exposed to this virus may develop recurrent (secondary) lesions as clusters of small vesicles, most commonly on the lips (herpes labialis) and less commonly on the keratinized mucosa of the gingiva or hard palate (Fig. 425-2). Such lesions contain live virus and tend to recur at the same site but less frequently with increasing age.

Although widespread vaccination has reduced the incidence, similar mucosal vesicles may also accompany the initial infection by the varicella-zoster virus in children with chickenpox (Chapter 375), and unilateral lesions may occur if herpes zoster (Chapter 375) affects branches of the trigeminal nerve. Uncommonly, oral mucosal ulcers may be caused by different types of coxsackievirus (Chapter 379), appearing on any oral site in hand-foot-and-mouth disease or on the soft palate or pharynx in herpangina. After infection by the measles (rubeola) virus, small ulcers (Koplik's spots; see Fig. 367-1)

may form on the inside of the cheeks 1 to 2 days before development of the skin rash (Chapter 367).

ERYTHEMA MULTIFORME

In this potentially recurrent disease, painful oral mucosal ulcerations develop rapidly, with or without target-like skin lesions. It may be associated with a previous viral infection or hypersensitivity to a food or drug. The affected patients, usually young adults with minimal or no systemic symptoms, have irregularly shaped ulcers that can be small and few or involve large areas of the mucosa; the most common sites are the lower labial mucosa and vermilion. On the vermillion, hemorrhagic crusting is a characteristic finding. These lesions can be distinguished from those of primary herpes by the absence of oral vesicles and systemic symptoms or by the presence of characteristic skin lesions (Chapter 439). A major variant of this disease is Stevens-Johnson syndrome, in which ocular, genital, and other lesions may accompany the oral lesions.

VENEREAL INFECTIONS

Primary syphilis may arise as a solitary, indurated, painless ulcer on the oral mucosa that resolves spontaneously in 4 to 6 weeks (Chapter 319). Uncommonly, *Neisseria gonorrhoeae* (Chapter 299) may cause oral ulcers, usually in the pharynx, which may be confused with oral ulcers of other causes.

Oral Squamous Cell Carcinoma

About 4% of all cancers occur in the mouth, commonly as squamous cell carcinomas of the mucosal epithelium (Chapter 190). Oral carcinoma occurs usually in the fifth decade or beyond, in men twice as frequently as in women, and is associated with long-term use of tobacco in more than 80% of cases.

Oral carcinoma usually arises as a chronic, indurated, cratered ulcer, but white (leukoplakia) and especially red (erythroplakia) macular lesions (Table 425-3; Fig. 425-3) frequently exhibit premalignant dysplasia or early carcinoma. Oral carcinomas spread to cervical lymph nodes. The overall 5-year survival rate is about 40%, but early treatment of small, localized lesions can lead to survival rates as high as 90%. Nevertheless, current guidelines find insufficient evidence to recommend for or against screening in asymptomatic adults.[3]

Over the past decade, there has been a rapid increase in a type of head and neck cancer associated with the human papillomavirus (HPV) type 16 (Chapter 373).[4] Occurring primarily at the base of tongue and tonsillar region of the oropharynx, this form of nonkeratinizing squamous cell carcinoma is seen in younger patients who typically lack the history of smoking and alcohol that is usually associated with more traditional forms of oral cancer. Other features of the disease include advanced stage at presentation and its good response to radiation and chemotherapy (Chapter 190). In the absence of concurrent tobacco use, the 5-year survival rate is 70 to 80%.

TABLE 425-3 WHITE AND RED/BLUE ORAL MUCOSAL LESIONS

WHITE LESIONS (PLAQUES)
Squamous cell carcinoma (early)
Frictional keratosis (buccal mucosa at dental occlusal line)
Leukoplakia (with or without dysplasia)
Smokeless tobacco–associated lesions
Nicotine stomatitis (palate)
Lichen planus (reticular and plaque types)
Pseudomembranous candidiasis (thrush)
Hyperplastic candidiasis (candidal leukoplakia)
Hairy leukoplakia (HIV associated; usually on lateral tongue)
Geographic tongue
Mucous patch or condyloma latum of secondary syphilis
Pseudomembrane-covered ulcers (see Table 425-1)
RED OR BLUE LESIONS (MACULAR, MACULOPAPULAR)
Squamous cell carcinoma (early)
Erythroplakia (epithelial dysplasia)
Erythematous (atrophic) candidiasis
Median rhomboid glossitis
Mucocutaneous diseases (see Table 425-1)
Angular cheilitis
Telangiectasias and purpuras (red to blue)
Kaposi's sarcoma (blue to purple)

HIV = human immunodeficiency virus.

Other Chronic Ulcerations

Prescription drugs that can be responsible for chronic oral mucosal ulcerations include barbiturates, β-blockers, nonsteroidal anti-inflammatory drugs, and many others.[5] Several mucocutaneous diseases can cause chronic multifocal oral mucosal lesions composed of ill-defined areas of erythema and ulceration. They are among the most difficult oral lesions to diagnose and are discussed later with the red lesions (see Table 425-3). Several microbial infections or underlying osteonecrosis (e.g., associated with bisphosphonate and other medication use) can lead to indurated, chronic oral mucosal ulcerations with moderate symptoms (see Table 425-1).

White Lesions

White plaques are commonly found in the mouth but, like ulcerations, have a wide variety of causes and outcomes (see Table 425-3). The clinical descriptor term *leukoplakia* applies to a white plaque that does not rub off and whose appearance does not indicate another disease. Leukoplakia can occur in any area of the mouth and usually exhibits benign hyperkeratosis on biopsy (see Fig. 190-1). On long-term follow-up, 2 to 6% of these lesions undergo malignant transformation into squamous cell carcinoma. Areas of leukoplakia with a corrugated surface or mixed with areas of erythema are often found in the lower labial or buccal vestibule of patients who use smokeless tobacco.

Frictional keratoses are often found posterior to the lower molar teeth as irregular white plaques and on the buccal mucosa as white lines adjacent to the dental occlusion. Unlike leukoplakia, these lesions rarely become malignant.

LICHEN PLANUS

Oral lesions of lichen planus (Chapter 438) occur in about 1% of the population, usually as multiple, bilaterally symmetrical reticular white plaques, with or without adjacent areas of erythema (atrophy or erosion) or irregular ulcers[6] (Fig. 425-4). The presence of mucosal atrophy, erosion, or ulceration usually causes pain and sensitivity to certain foods. Most lesions can be

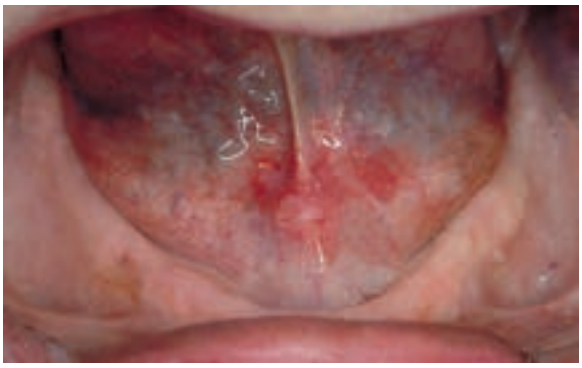

FIGURE 425-3. Squamous cell carcinoma. Biopsy of this area of erythroplakia with slight induration in the anterior mouth floor exhibited squamous cell carcinoma.

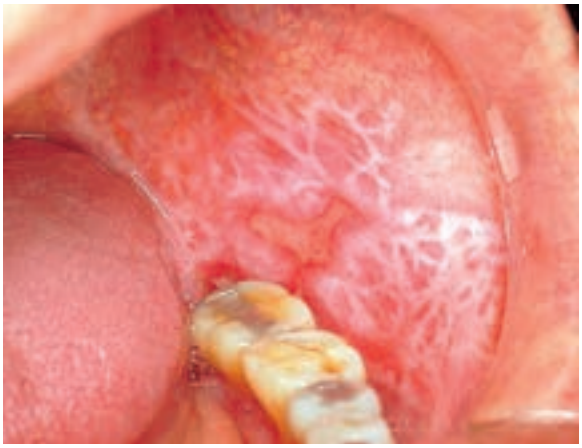

FIGURE 425-4. Lichen planus. A similar-appearing lesion is also present on the right buccal mucosa. Note central pseudomembrane-covered ulceration.

adequately controlled by topical application of fluocinonide or clobetasol gel or ointment (0.05%, three times a day) for periods of several weeks, although recurrence is common.

ORAL CANDIDIASIS

This common fungal disease (Chapter 338) has three clinical forms: pseudomembranous (thrush), erythematous (atrophic), and hyperplastic (candidal leukoplakia). Pseudomembranous candidiasis, usually of relatively short duration, occurs on any site and consists of white fungal plaques that can be rubbed off, leaving a red or bleeding base. Lesions of hyperplastic candidiasis are white, have fungal hyphae within the surface layers of hyperkeratotic epithelium, do not rub off, and are most often found on the anterior buccal mucosa or on the tongue. Erythematous candidiasis is discussed under Red Lesions. All forms of oral candidiasis represent overgrowth or superficial infection by *Candida* species from the oral flora, induced by a variety of causes, including suppression of bacterial flora by systemic antibiotics, chronic salivary dysfunction, uncontrolled diabetes mellitus or anemia, and immunosuppression (especially in HIV-infected patients). The condition can be managed with topical or systemic antifungal agents, although acquired resistance to fluconazole (200 mg on the first day, then 100 mg every day for 2 to 4 weeks) therapy can occur.

HAIRY LEUKOPLAKIA

The lesion of hairy leukoplakia, which is caused by Epstein-Barr virus, is a white plaque occurring most frequently on the lateral surfaces of the tongue bilaterally in immunosuppressed persons, usually HIV infected (Fig. 425-5). *Candida* may be present in the surface layers, but the lesion is not eliminated by effective antifungal therapy. The diagnosis of hairy leukoplakia by biopsy should raise suspicion of HIV infection or other forms of systemic or local immunosuppression.

GEOGRAPHIC TONGUE

Also called *benign migratory glossitis,* this benign idiopathic condition affects the dorsal tongue of about 2% of the population. It is characterized by well-defined areas of atrophied filiform papillae bordered by arcs of normal or hyperplastic filiform papillae and by gradual changes in the location of these lesions over time (Fig. 425-6). Treatment is usually not necessary.

SECONDARY SYPHILIS

Secondary syphilis may manifest as a well-defined white plaque on the labial or palatal mucosa, called *condyloma latum* (or "split papule," because of its lobulated periphery), or as a mucous patch.

Red Lesions

Solitary red macules or plaques (*erythroplakia*) are less common in the mouth than white lesions but should be viewed with concern because they may exhibit premalignant dysplasia, carcinoma in situ, or carcinoma (see Table 425-3 and Fig. 425-3). One exception is a red macule occurring in the midline of the posterior dorsal tongue, classified as *median rhomboid glossitis,* which is an idiopathic but uniformly benign condition that is often associated with localized overgrowth of *Candida* species.

ERYTHEMATOUS (ATROPHIC) ORAL CANDIDIASIS

Erythematous (atrophic) oral candidiasis is a chronic condition characterized by erythema and atrophy of the filiform papillae on the dorsal tongue or by patchy, ill-defined erythema on the palate, tongue, or buccal mucosa (Fig. 425-7). It is usually accompanied by symptoms of oral mucosal burning and sensitivity to spicy foods. It occurs most commonly in patients with chronic salivary hypofunction (e.g., Sjögren's syndrome or anticholinergic drug effects), but it also occurs in patients who wear removable dentures infected with *Candida*, in whom mucosal erythema is confined to the denture-bearing area.

For acute or chronic oral candidal infections, systemic or topical antifungal drugs are necessary to resolve the associated lesions. In patients who have clinically apparent salivary production, fluconazole (200 mg on the first day, then 100 mg every day for 2 to 4 weeks) is the drug of choice. However, systemic antifungal drugs may not be effective in patients who have severe salivary hypofunction and insufficient saliva to convey the drug from the bloodstream to oral mucosa. In such patients, with remaining natural teeth, *oral* antifungal preparations (troches or pastilles), all of which contain cariogenic amounts of sucrose or glucose, *must not be used* to avoid enhancing dental caries. Instead, slow oral dissolution (15 to 20 minutes for 2 weeks to 2 months) of vaginal nystatin tablets (twice daily) or miconazole tablets (50 mg daily), which contain little or no caries-supporting carbohydrates, is safe and effective; patients usually need frequent sips of water to aid in dissolving the tablets. Effective topical or systemic treatment significantly improves oral symptoms. Treatment of denture-associated candidiasis requires concurrent treatment of the denture.

The treatment end point is reached when mucosal burning symptoms cease, the patient can again tolerate acidic or spicy foods, and filiform papillae on the dorsal tongue have returned to normal; this recovery takes 2 to 12

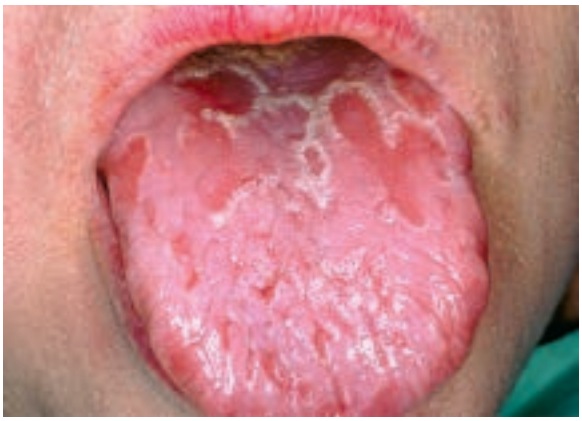

FIGURE 425-6. Geographic tongue. The distribution of these changes on the dorsal tongue may change over time, but they are asymptomatic and diagnosed by their characteristic appearance.

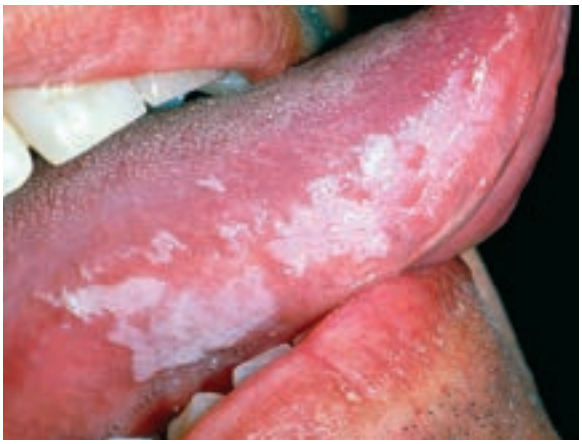

FIGURE 425-5. Hairy leukoplakia. These white plaques were the first visible sign of human immunodeficiency virus infection.

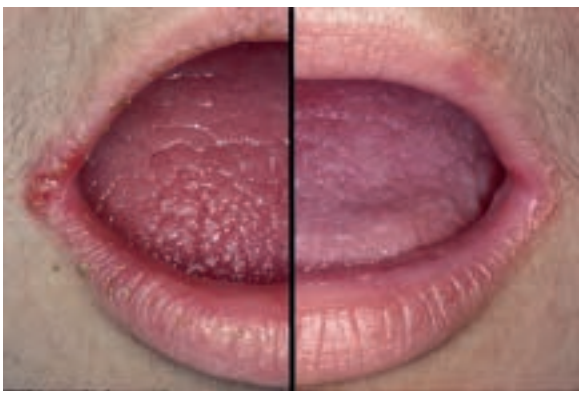

FIGURE 425-7. Erythematous oral candidiasis. *Left,* Erythematous candidiasis in a 26-year-old woman with primary Sjögren's syndrome, exhibiting symptomatic angular cheilitis, atrophic mucositis, and lingual papillary atrophy. *Right,* Asymptomatic and normal-appearing mucosa after treatment with appropriate topical antifungal drugs (see text).

weeks, depending on patients' salivary production and treatment compliance. Recurrence is common in patients with chronic salivary hypofunction or immunosuppression, which necessitates recurring or long-term treatment using a noncariogenic topical antifungal drug that provides a sufficient duration of oral mucosal contact (e.g., nystatin or miconazole tablets).

ANGULAR CHEILITIS

Erythema or crusting of the labial angles is usually caused by *Candida* species (see Fig. 425-7) and is usually associated with intraoral candidiasis. In such cases, topical treatment of the angular cheilitis with clotrimazole (1% cream) must be accompanied by intraoral or systemic antifungal treatment, as described previously.

MUCOCUTANEOUS DISEASES

The mucocutaneous diseases of pemphigus vulgaris, mucous membrane pemphigoid, atrophic or erosive lichen planus, and lupus erythematosus can cause similar-appearing oral lesions. Their diagnosis requires examination of a biopsy specimen by routine histopathology and direct immunofluorescence to identify characteristic deposits of various inflammatory proteins.

The first lesions of pemphigus vulgaris are usually oral mucosal vesicles that rapidly rupture, leaving painful erosions or ulcerations. These are followed by development of skin lesions. Rarely, the lesions remain confined to the mouth (Chapter 439).

Lesions of mucous membrane (cicatricial) pemphigoid are usually confined to the oral mucosa or conjunctivae and occur in patients older than 50 years. They begin as vesicles that quickly rupture, leaving ulcers that are chronic but only moderately symptomatic. Use of topical fluocinonide or clobetasol (0.05% gel or ointment, three times a day, 4 to 12 weeks) for several months, as described for lichen planus, is sometimes sufficient to treat the oral lesions, but some patients also need systemic treatment (Chapter 439).

Oral mucosal lesions of lupus may occur in patients who have systemic lupus erythematosus (SLE), in patients who do not have SLE but later develop that disease, or in patients who do not develop SLE (Chapter 266). In this latter group, the lesions of mucosal lupus may be analogous to the skin lesions of chronic discoid lupus. They take the form of reticular hyperkeratotic figures associated with erythema, often resembling lichen planus, but unlike lichen planus are usually solitary or bilaterally asymmetrical. They can be controlled by topical fluocinonide (0.05%, three times a day, 2 to 4 weeks) or intralesional triamcinolone suspension (5 mg/mL), or respond to systemic treatment of SLE.

Pigmentations

Brown or gray-black macules on the oral mucosa are relatively common and range from benign to highly malignant. They may be caused by localized increase in melanin production, proliferation of melanin-producing cells, or deposition of local or systemically distributed pigmented substances (Table 425-4). Mucosal pigmentation may occur after long-term administration of hydroxychloroquine, minocycline, ketoconazole, methyldopa, or cyclophosphamide. Malignant melanomas can occur at any oral mucosal site, but about 85% develop on the hard palatal mucosa or gingiva, or both. Diagnosis of any of these conditions is usually established by biopsy and knowledge of relevant underlying conditions.

Lesions of Kaposi's sarcoma associated with HIV infection often appear first on the oral mucosa, especially the palate. They begin as macules with a blue or purple color, at which time they must be distinguished from purpura. Later, they spread radially and expand vertically (Chapter 393).

TABLE 425-4 PIGMENTATIONS OF THE ORAL MUCOSA (BROWN OR GRAY-BLACK IN COLOR)

INCREASED MELANIN PRODUCTION (FLAT LESIONS)
Oral melanotic macule
Ephelis (vermilion border)
Systemic diseases: Addison's disease, von Recklinghausen's disease of skin, Albright's syndrome, Peutz-Jeghers syndrome
PROLIFERATION OF MELANIN-PRODUCING CELLS (FLAT OR RAISED LESIONS)
Pigmented cellular nevi (benign and premalignant types)
Atypical melanocytic hyperplasia, melanoma in situ, radial growth phase of melanoma
Malignant melanoma
NONMELANIN PIGMENTATION
Amalgam tattoo
Focal deposition of systemically distributed metal (lead, bismuth, mercury, others) usually at sites of chronic inflammation
Systemically administered drugs (chloroquine, minocycline, ketoconazole, cyclophosphamide)

ORAL SOFT TISSUE TUMORS

A range of oral benign soft tissue tumors should be treated by excisional biopsy.

Connective Tissue Hyperplasias

The most common oral soft tissue tumors are small, pedunculated masses of hyperplastic fibrous connective tissue covered by normal-appearing mucosa (Table 425-5). Solitary lesions are usually found on the inside of the cheeks or lips. Similar lesions may be present at the border of an ill-fitting denture or may occur in clusters on the hard palate under an ill-fitting denture (palatal papillomatosis).

Generalized or multifocal enlargement of the gingiva (gingival hyperplasia) may be caused by chronic administration of phenytoin, cyclosporine, and many of the calcium-channel blocking drugs (e.g., diltiazem, verapamil, or nifedipine; Fig. 425-8). It can also be associated with a hereditary defect or be caused by an infiltration of atypical white blood cells in some types of leukemia (particularly acute monocytic leukemia; Chapter 183) or by uncontrolled diabetes mellitus (Chapter 229).

Reactive Hyperplasias

Small masses with surfaces that are ulcerated or only partially covered by normal-appearing mucosa usually represent reactive lesions in the form of pyogenic granulomas (whose frequency increases during pregnancy), peripheral giant cell granulomas, or lymphoid hyperplasia of the lingual or other tonsillar tissue. The granulomas are most often located on the gingiva. Rarely, such lesions may represent a metastatic neoplasm.

Epithelial Proliferations

Small, white, wartlike epithelial masses are common and can occur in any area of the oral mucosa (Fig. 425-9). They are occasionally classified as epithelial neoplasms, but most do not continue to grow. Human papillomavirus types 2, 6, 11, 13, 32, and 57 have been identified in these wartlike

TABLE 425-5 ORAL SOFT TISSUE TUMORS

CONNECTIVE TISSUE HYPERPLASIA (NORMAL-APPEARING OVERLYING MUCOSA)
Irritation fibroma
Denture-associated hyperplasia
Palatal papillomatosis
Generalized gingival hyperplasia
Drug-induced (phenytoin, nifedipine, cyclosporine)
Hereditary
REACTIVE HYPERPLASIA (ERYTHEMATOUS OVERLYING MUCOSA)
Pyogenic granuloma/pregnancy tumor
Peripheral giant cell granuloma
Inflammatory gingival hyperplasia
Hyperplastic lingual tonsil
EPITHELIAL MASSES (USUALLY IRREGULAR WHITE SURFACE)
Papilloma/oral wart
Squamous cell carcinoma
Verrucous carcinoma
Focal epithelial hyperplasia (Heck's disease)
Condyloma acuminatum (venereal wart)
Keratoacanthoma (on lips)
SALIVARY DUCT OBSTRUCTION (MINOR SALIVARY GLANDS)
Mucocele/ranula (usually fluctuant)
Salivary stone (sialolith)
SUBEPITHELIAL NEOPLASMS
Primary connective tissue or salivary gland tumors
Metastatic lesions (especially in the mandible)
Lymphoma (especially in the palate or posterior mandible)
Focal or generalized leukemic infiltrates in the gingiva (especially with acute monocytic leukemia)

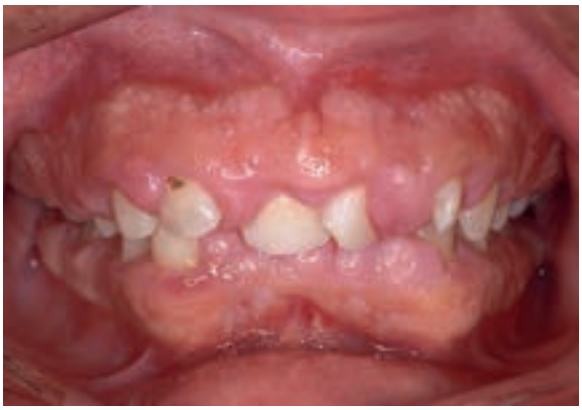

FIGURE 425-8. Drug-induced gingival hyperplasia. Similar clinical lesions may occur with prolonged use of various drugs or as a hereditary condition (see text).

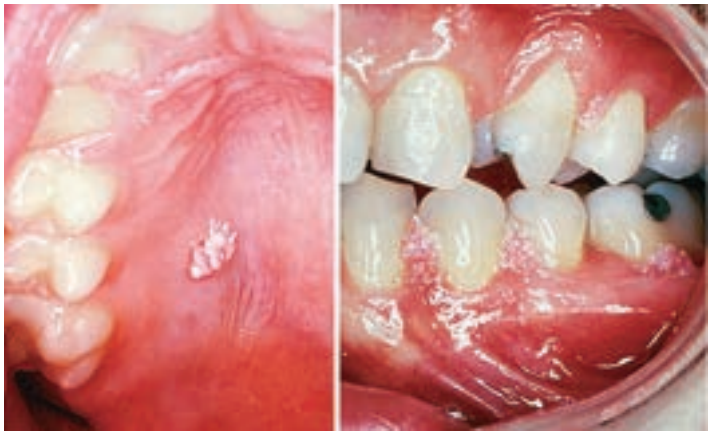

FIGURE 425-9. Papillary epithelial tumors. *Left,* A solitary squamous papilloma. *Right,* multiple gingival papillomas, occurring in all quadrants, from condyloma acuminatum, associated with papilloma virus subtype 6 or 11.

lesions, which are usually classified generically as squamous papillomas. A large wartlike lesion on the oral mucosa should raise suspicion of verrucous carcinoma.

Mucous Retention Lesions (Mucoceles)

Mucoceles are small, chronic, or recurring vesicles or bullae that occur commonly on the inside of the cheeks and lips, the posterior palate, and the mouth floor. They are caused by injury to one of the many submucosal minor salivary glands, resulting in extravasation of mucus, which causes granulomatous inflammation or blockage of the excretory duct, leading to cyst formation. Both types of lesions require conservative surgical excision because simple incision and drainage are usually followed by recurrence.

SALIVARY GLAND DISEASES

Primary Diseases of Salivary Glands

Unilateral major salivary gland enlargement that is markedly painful or tender to palpation and has a purulent exudate or nothing expressible from the duct suggests bacterial sialadenitis. Any exudate should be cultured, and initial treatment should be with an oral penicillinase-resistant antibiotic, such as cloxacillin or dicloxacillin, 500 mg, every 6 hours.

More than 20 types of benign or malignant neoplasms appear as firm and nontender unilateral or bilateral enlargement of a major gland or as a firm submucosal nodule on the palate or the labial or buccal mucosa (Table 425-6). Their causes are unknown, except for Warthin's tumor (adenolymphoma), which has a strong association with cigarette smoking. Uncommonly, unilateral major gland enlargement may be reactive—for example, chronic sialadenitis from duct obstruction or inadequately treated bacterial sialadenitis.

Salivary gland tumors are relatively uncommon and usually present as a swelling in one of the major paired salivary glands or in one of the minor glands of the mouth. Most occur in the major glands, with approximately 90% developing in the parotid. The most common benign salivary gland tumor is the pleomorphic adenoma. Most salivary gland tumors in the parotid gland are benign, in contrast to the sublingual gland, where more than 90% are malignant. Approximately half of tumors in the submandibular and minor glands are malignant. Benign tumors are generally slowly growing, not fixed to the skin, and do not show ulceration. Malignancies generally grow more quickly, are often fixed to the skin or adjacent normal structures, and tend to show ulceration. Adenoid cystic carcinoma has a characteristic local infiltration by perineural spread. Detection of any salivary gland mass lesions should be followed by appropriate imaging, cytology, and biopsy.[7] Both benign and malignant tumors are generally treated by surgery.[8]

TABLE 425-6 CAUSES OF SALIVARY GLAND ENLARGEMENT

USUALLY UNILATERAL
Benign or malignant salivary gland neoplasms (more than 20 different histopathologic types)
Bacterial infection
Chronic sialadenitis (single gland)
USUALLY BILATERAL AND ASSOCIATED WITH SALIVARY HYPOFUNCTION
Viral infection (mumps, cytomegalovirus, influenza, Coxsackie A)
Sjögren's syndrome (benign lymphoepithelial lesion)
Chronic granulomatous diseases (sarcoidosis, tuberculosis, leprosy)
Recurrent parotitis of childhood
Human immunodeficiency virus infection/acquired immunodeficiency syndrome
BILATERALLY SYMMETRICAL, SOFT, NONTENDER, PAROTID ONLY
Sialadenosis (asymptomatic parotid enlargement), idiopathic or associated with:
Diabetes mellitus
Hyperlipoproteinemia
Hepatic cirrhosis
Anorexia/bulimia
Chronic pancreatitis
Acromegaly
Gonadal hypofunction
Phenylbutazone use

Secondary Diseases of Salivary Glands

BILATERAL SALIVARY GLAND ENLARGEMENT AND DECREASED SALIVARY SECRETION ASSOCIATED WITH SYSTEMIC DISEASES

The best-known cause of bilateral salivary gland enlargement is infection by the mumps virus (Chapter 369) in children. However, the prevalence of mumps decreased in the United States by more than 98% after the introduction of an effective vaccine in 1967, and now there are only a few hundred to a few thousand cases per year. Uncommonly, a less acute, mumps-like illness may occur in adults in association with cytomegalovirus (Chapter 376), influenza (Chapter 364), or Coxsackie A (Chapter 379) virus infection.

About 15% of patients who meet the American College of Rheumatology classification criteria for Sjögren's syndrome[9] (Chapter 268) may gradually develop chronic bilateral enlargement of major salivary glands, which feel firm and nontender or only slightly tender to palpation. Histologically, the tumors begin as a benign lymphoepithelial lesion (myoepithelial sialadenitis), but after years of chronicity, some transform into an extranodal marginal zone lymphoma (Chapter 185).

In most patients with Sjögren's syndrome, gradually progressive salivary hypofunction can impair speech and swallowing and cause a characteristic pattern of progressive dental caries that leads to excessive tooth loss if not actively prevented. In severe cases, the oral mucosa becomes dry and sticky, saliva is not expressible from the major ducts, and about one third of patients have signs and symptoms of chronic erythematous candidiasis (see earlier discussion and Fig. 425-7).

The salivary component of Sjögren's syndrome is diagnosed from a labial salivary gland biopsy specimen that contains three to five minor glands and exhibits focal lymphocytic sialadenitis in the absence of nonspecific chronic sialadenitis or another disease, such as noncaseating granuloma.[10] A patient's symptoms of oral dryness (xerostomia) can be caused by a wide variety of conditions (Table 425-7).

Several chronic granulomatous diseases, such as sarcoidosis (Chapter 95), tuberculosis (Chapter 324), and leprosy (Chapter 326), can cause bilateral enlargement and decreased function of salivary glands. The clinical and serologic features of sarcoidosis may occasionally closely mimic those of Sjögren's syndrome, and the distinction is best made by minor salivary gland biopsy.

TABLE 425-7 CAUSES OF DECREASED SALIVARY SECRETION

TEMPORARY
Effects of short-term drug use (e.g., antihistamines)
Virus infections (e.g., mumps)
Dehydration
Psychogenic conditions (e.g., anxiety)
CHRONIC
Effects of chronically administered drugs (particularly antidepressants, monoamine oxidase inhibitors, neuroleptics, parasympatholytics, some combinations of drugs for treating hypertension)
Chronic diseases
Sjögren's syndrome
Sarcoidosis
Human immunodeficiency virus or hepatitis C infection
Depression
Diabetes mellitus (uncontrolled)
Amyloidosis (primary or secondary)
Central nervous system diseases
Other effects of treatment
Therapeutic radiation to the head and neck
Graft-versus-host disease
Absent or malformed glands (rare)

A few adult patients with HIV infection and most children who are infected in utero develop major salivary gland enlargement and reduced salivary secretion that are caused by lymphocytic infiltration. Parotid gland enlargement usually represents a solid or cystic lymphoepithelial lesion (see Table 425-2).

Recurrent parotitis of childhood includes episodes of unilateral or bilateral parotid enlargement. During flares of this illness, salivary secretion may be reduced, but usually without prominent secondary symptoms or signs. This condition, of unknown cause, usually subsides after puberty.

ASYMPTOMATIC PAROTID ENLARGEMENT (SIALADENOSIS, SIALOSIS)

Parotid glands can develop bilateral, symmetrical enlargement that is soft and nontender to palpation and associated with normal salivary function (see Table 425-6). Diagnosis is established by this clinical presentation and the presence of one of the systemic diseases known to be associated with it: diabetes mellitus (Chapter 229), hyperlipoproteinemia (Chapter 206), hepatic cirrhosis (Chapter 153), anorexia or bulimia (Chapter 219), chronic pancreatitis (Chapter 144), acromegaly (Chapter 224), and gonadal hypofunction. It can also result from use of phenylbutazone or be a reaction to iodine-containing contrast media. Biopsy of the affected glands is not indicated for diagnosis.

Impaired Salivary Secretion without Gland Enlargement

The common symptom of dry mouth (xerostomia) is most often a side effect of chronically administered drugs. Many classes of drugs reduce unstimulated salivary secretion through anticholinergic or other mechanisms (see Table 425-7). Patients experience these symptoms soon after beginning to use the drug but produce enough saliva during a meal for normal chewing and swallowing. However, the symptoms and associated dental caries are dose dependent and gradually increase with prolonged use of the drug. The classes of drugs producing the most profound effects are most tricyclic antidepressants, most neuroleptics, monoamine oxidase inhibitors, and all anticholinergics. A combination of drugs for treatment of hypertension may also cause symptoms of dry mouth.

Several systemic diseases affect salivary secretion. As noted earlier, most patients with Sjögren's syndrome, some with sarcoidosis, and a few patients with HIV infection experience symptoms of dry mouth to various degrees, with or without salivary gland enlargement. In addition, patients who have primary or secondary *amyloidosis* with salivary gland amyloid deposits may develop impaired secretion. The symptom of xerostomia is more prevalent in individuals who exhibit symptoms of depression, even in those not taking drugs for its treatment. Studies done before the availability of antidepressant drugs showed that symptoms of depression were associated with decreased salivary secretion.

Irradiation of the head and neck region to treat a malignant tumor usually produces profound dry mouth during therapy. Secretory capacity recovers only slightly in the months after treatment for patients with solid tumors but recovers significantly for those with multifocal tumors (e.g., Hodgkin's disease).

TREATMENT

Rx

Significant chronic salivary hypofunction from any cause produces a risk for dental caries (decay) in approximate proportion to the secretory impairment, but caries can largely be prevented if appropriate measures are taken as soon as the hypofunction begins. Remaining teeth should be protected by a comprehensive dental caries prevention program, monitored by a dentist and including frequent application of appropriate topical fluorides, removal of dental plaque, counseling on control of cariogenic dietary carbohydrates, and placement of appropriate dental restorations as necessary.

Symptomatic treatment of mild to moderately severe salivary hypofunction can include sialagogues such as sugar-free hard candies or chewing gum, regular sips of water, and use of saliva substitutes at night, but no topical therapy is reliably helpful.[A1] Symptoms of severe hypofunction can be improved by prescribing pilocarpine (5 mg four times a day), if not contraindicated, but such treatment alone will not prevent dental caries.

Chronic erythematous oral candidiasis is a frequent sequela of chronic salivary hypofunction, and its treatment and retreatment, as noted earlier, substantially improve the patient's oral symptoms.

PERIODONTAL DISEASE

Periodontal diseases are a group of oral infections that affect the periodontium, which are the tissues that support and maintain teeth in the jaws. A worldwide problem, it is the most common causes of tooth loss. Periodontal disease in its most prevalent form is associated with excessive build-up of plaque on teeth and roots. Most cases of periodontal disease begin with inflammation of the gingiva—termed *gingivitis*—that may progress to loss of the supporting bone around the roots of the teeth. Other subtypes of periodontal disease are recognized with differing risk factors and natural histories. The mainstay of treatment is the removal of subgingival calculus and biofilm deposits using mechanical methods (tooth brushing, flossing, scaling, and root planing). Referral for appropriate dental care is indicated. Despite ongoing concerns, there is no current evidence that periodontal disease is an independent risk factor for coronary artery disease.[I1]

Grade A References

A1. Femiano F, Buonaiuto C, Gombos F, et al. Pilot study on recurrent aphthous stomatitis (RAS): a randomized placebo-controlled trial for the comparative therapeutic effects of systemic prednisone and systemic montelukast in subjects unresponsive to topical therapy. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2010;109:402-407.

A2. Brocklehurst P, Tickle M, Glenny AM, et al. Systemic interventions for recurrent aphthous stomatitis (mouth ulcers). *Cochrane Database Syst Rev*. 2012;9:CD005411.

A3. Davari P, Hsiao HH, Fazel N. Mucosal lichen planus: an evidence-based treatment update. *Am J Clin Dermatol*. 2014;15:181-195.

A4. Furness S, Worthington HV, Bryan G, et al. Interventions for the management of dry mouth: topical therapies. *Cochrane Database Syst Rev*. 2011;12:CD008934.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Carlson ER, Ghali GE, Herb-Brower KE. Diagnosis and management of pathological conditions. *J Oral Maxillofac Surg*. 2012;70:e232-e271.
2. Akintoye SO, Greenberg MS. Recurrent aphthous stomatitis. *Dent Clin North Am*. 2014;58: 281-297.
3. Moyer VA. Screening for oral cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160:55-60.
4. Chung CH, Zhang Q, Kong CS, et al. p16 protein expression and human papillomavirus status as prognostic biomarkers of nonoropharyngeal head and neck squamous cell carcinoma. *J Clin Oncol*. 2014;32:3930-3938.
5. Kalmar J. Oral manifestations of drug reactions. http://emedicine.medscape.com/article/1080772-overview; Accessed March 14, 2015.
6. Le Cleach L, Chosidow O. Clinical practice. Lichen planus. *N Engl J Med*. 2012;366:723-732.
7. Colella G, Cannavale R, Flamminio F, et al. Fine-needle aspiration cytology of salivary gland lesions: a systematic review. *J Oral Maxillofac Surg*. 2010;68:2146-2153.
8. Feinstein TM, Lai SY, Lenzner D, et al. Prognostic factors in patients with high-risk locally advanced salivary gland cancers treated with surgery and postoperative radiotherapy. *Head Neck*. 2011;33: 318-323.
9. Shiboski SC, Shiboski CH, Criswell L, et al. American College of Rheumatology classification criteria for Sjogren's syndrome: a data-driven, expert consensus approach in the Sjogren's International Collaborative Clinical Alliance cohort. *Arthritis Care Res (Hoboken)*. 2012;64:475-487.
10. Daniels TE, Cox D, Shiboski CH, et al. Associations between salivary gland histopathologic diagnoses and phenotypic features of Sjögren's syndrome among 1,726 registry participants. *Arthritis Rheum*. 2011;63:2021-2030.
11. Lockhart PB, Bolger AF, Papapanou PN, et al. Periodontal disease and atherosclerotic vascular disease: does the evidence support an independent association? A scientific statement from the American Heart Association. *Circulation*. 2012;125:2520-2544.

REVIEW QUESTIONS

1. A new 62-year-old female patient complains of various bilateral joint pains. In the course of examination, you observe symmetrical parotid gland enlargement, which is soft and nontender to palpation. Intraoral examination is within normal limits. Which of the following is the most appropriate recommendation?

A. Request labial salivary gland biopsy to evaluate possible Sjögren's syndrome.
B. Request parotid gland biopsy to exclude multifocal neoplasia, such as lymphoma.
C. Request a chest radiograph to evaluate possible sarcoidosis.
D. Request serum anti-SS-A/-B, rheumatoid factor and antinuclear antibody to seek serologic evidence of Sjögren's syndrome.
E. Request a blood glucose level to consider diabetes.

Answer: E This patient has sialadenosis, distinguished by its bilaterally symmetrical enlargement and the absence of induration, which would indicate an absence of intraglandular inflammation or neoplasia. This condition is most commonly associated with diabetes mellitus or alcoholic cirrhosis, but it may also occur in patients with gonadal hypofunction, acromegly, chronic pancreatitis, hyperlipoprotenemia, or anorexia nervosa-bulimia. This type of parotid enlargement is self-limited and does not require treatment. Biopsy, which in this condition would show only parenchymal hyperplasia, is not indicated.

2. A 35-year-old female patient who is taking prescription drugs to manage depression complains of a progressive burning sensation on her tongue during the last several months. She smokes one pack of cigarettes per day. On physical examination, the dorsal surface of her tongue exhibits diffuse erythema and generalized atrophy of the filiform papillae. Her oral mucosa is moist, but little saliva accumulates in the floor of her mouth during your examination. Which of the following is the most appropriate recommendation?

A. Request biopsy of the affected area of her tongue to exclude mucosal dysplasia or neoplasia.
B. Change her current combination of medications.
C. Prescribe a course of topical steroid application to reduce the mucosal inflammation.
D. Prescribe a course of fluconazole, lasting at least 2 weeks or until the erythema resolves and filiform papillae return.
E. Recommend that she stop smoking.

Answer: D Although smoking cessation is certainly to be recommended for any patient, this patient's lingual mucosal changes are characteristic of mucosal candidiasis, possibly as a result of decreased salivary function from medications used to manage her depression. In patients with little or no residual salivary function (e.g., in advanced Sjögren's syndrome or after head and neck radiation therapy), treatment of oral mucosal candidiasis may require the use of topical antifungal therapy.

3. A 38-year-old man complains weight loss and night sweats. His history includes smoking one pack of cigarettes per day for about 20 years. On physical examination, you observe corrugated, well-defined white patches on the ventrolateral tongue, bilaterally. Which of the following is the most appropriate recommendation?

A. Refer the patient for biopsy of the lingual lesions to exclude dysplasia or carcinoma.
B. Advise the patient to discontinue smoking.
C. Request a complete blood count and HIV test, then have the lesions biopsied.
D. Have him return in 1 month to see if the lesions resolve spontaneously.
E. Search for causes of chronic trauma.

Answer: C Although discontinuance of smoking is always good advice, many oral mucosal lesions are not associated with tobacco consumption. This case may represent oral hairy leukoplakia, an Epstein-Barr virus (which can be identified in the biopsy specimen) lesion and a manifestation of systemic HIV infection. These are uncommon lesions and are associated with a high HIV viral load; clinical suspicion should encourage further search for other HIV-related symptoms or signs.

4. A 52-year-old woman complains of dry mouth of long duration. She is taking five prescription medications, two of which have anticholinergic effects, and her affect suggests clinical depression. On examination, palpation of the four major salivary glands exhibits slight to mild induration. Which of the following is the most appropriate recommendation?

A. Perform or refer the patient for a labial salivary gland biopsy to consider the salivary component of Sjögren's syndrome.
B. Replace the two prescription drugs with alternatives that have less or no anticholinergic effects.
C. Refer the patient for specialty management of her depression.
D. Include rheumatoid factor (RF) and antinuclear antibody (ANA) titer in your laboratory request to consider the serologic component of Sjögren's syndrome.
E. Perform an unanesthetized Schirmer tear test to consider the ocular component of Sjögren's syndrome.

Answer: A Although the patient's symptoms may be caused by the side effects of anticholinergic drugs, palpable salivary induration raises the suspicion of inflammation within them. Sjögren's syndrome is high on the differential diagnosis and is an indication for labial salivary gland biopsy. Positive RF and ANA are usually seen in patients with Sjögren's syndrome, but they lack specificity, even if both are positive. Positive anti-SS-A/SS-B serology can represent the serologic component of Sjögren's syndrome but is insufficient alone. A positive Schirmer test is currently insufficient to identify the ocular component of Sjögren's syndrome.

5. Examination of an asymptomatic 27-year-old male nonsmoker reveals numerous circular white lesions surrounding red atrophic patches on the dorsum of the tongue. The lesions cannot be rubbed off with a gauze square. The patient most likely has which of the following?

A. Fissured tongue
B. Speckled leukoplakia
C. Benign migratory glossitis (geographic tongue)
D. Squamous cell carcinoma
E. Lichen planus

Answer: C The patient has geographic tongue, a harmless and common condition that results in transient depapillation of the tongue with a surrounding white halo. Over the course of several weeks, the depapillated area returns to normal. However, other areas may become affected, thereby giving the patient the perception that the lesion is migrating over the tongue. Despite its superficial resemblance to psoriasis, microscopically there is no relation. For the symptomatic patient, bland mouth rinses (baking soda in warm water, salt water) may provide symptomatic relief. Antifungal treatments are not recommended. There is no association with oral cancer.

426

APPROACH TO THE PATIENT WITH NOSE, SINUS, AND EAR DISORDERS

ANDREW H. MURR

Patients with nose, sinus, and ear disorders may have a variety of chief complaints. Nasal symptoms most commonly relate to rhinorrhea or congestion, both of which may be due to allergic, infectious, inflammatory, neoplastic, or structural causes. Sinus disorders, which commonly arise as a feeling of stuffiness or congestion but are sometimes also manifested as pain or even

headache (Chapter 398), have a similar set of causes. Common ear complaints include pain, tinnitus, loss of hearing (Chapter 428), and vestibular symptoms (Chapter 428), commonly described by the patient as dizziness but recognized by the physician as vertigo that is different from the lightheadedness that characterizes presyncope and syncope (Chapters 51 and 62). Epistaxis, which is bleeding from the nose, is usually easy to distinguish from hemoptysis from the bronchial tree (Chapter 83) or hematemesis from the gastrointestinal tract (Chapter 135).

Loss of hearing (Chapter 428) and vestibular symptoms (Chapter 428) are discussed elsewhere, as are smell, the related sensation of taste (Chapter 427), and the details of head and neck tumors (Chapter 190). This chapter focuses on the approach to patients with other common nose, sinus, and ear complaints.

NASAL AND SINUS COMPLAINTS

Rhinitis and Sinusitis

DEFINITION

Rhinitis is generally defined as any inflammatory process in the nose, with the common result being a sensation of excess mucous or nasal congestion. The patient may have a sensation of fluid dripping from the nose, either coming from the nose anteriorly or coming from the nose posteriorly. Anterior nasal drainage may be perceived by the patient as being accompanied by an activity such as eating (gustatory rhinitis) and may be visible to an observer. Posterior nasal drainage is more nebulous and subjective, but it is very common and is referred to as postnasal drip.

In general, acute rhinitis and sinusitis describe inflammatory conditions of the nose and sinuses that last less than 4 weeks. Chronic rhinitis and sinusitis persist for more than 3 months despite treatment. Recurrent acute rhinitis and sinusitis are defined by exacerbations that occur four or more times per year and last 7 to 10 days per episode. Subacute rhinitis and sinusitis define symptoms that persist between 4 and 12 weeks and resolve completely with treatment.

EPIDEMIOLOGY

The most common reason for a patient to seek the advice of a physician in the United States concerns problems relating to rhinitis and sinusitis. More than 20 million visits by patients per year are devoted to this complaint, and billions of dollars are spent on medications that are expected to improve the condition.

PATHOBIOLOGY

Humans normally produce about 2 L of mucus per day from their nasal lining. The nose functions primarily as a humidification and filtration system, with a clean and refreshed nasal mucous blanket serving to trap particulate matter and organisms. The nasal and sinus lining consists of ciliated respiratory epithelium; the cilia function in a highly organized and orderly fashion under normal circumstances to transport particulate matter trapped in the mucous blanket in a consistent fashion so that the mucus can be swallowed, thereby avoiding deposition in the bronchi. The nose also serves as the organ of olfaction (Chapter 427) to allow patients to discern tastes and avoid spoiled foods that could cause illness.

The parasympathetic nervous system controls both vascular tone and mucus production in the nose. Inflammatory conditions, such as the common cold, can cause the nasal and sinus lining to swell, thus highlighting the nasal cycle governed by parasympathetic neural control. In a normal state, one side of the nose is relatively decongested and one side is relatively congested because of vascular engorgement. This vascular dilation allows humidification and warming of inspired air and can also affect the ability to discern odors in the process of olfaction. During rhinitis, the inflammation exaggerates the normal relative comparison between the decongested and congested sides of the nose and can be perceived as an uncomfortable nasal stuffiness that shifts from side to side over a period of several hours.

Sinusitis differs from rhinitis in that the term implies an infectious cause rather than physiologic dysfunction. Nevertheless, many different mechanisms of inflammation besides infection may give rise to what is currently generally termed sinusitis.

CLINICAL MANIFESTATIONS

When normal nasal mucosal function is lost, patients often complain of nasal crusting or obstruction, hypersecretion or postnasal drip, coughing, facial pressure, and fatigue. Nasal obstruction that shifts from side to side during the day is common in many types of rhinitis and may be considered an exaggeration of normal physiology.

TABLE 426-1 MAJOR AND MINOR SINUSITIS FACTORS

MAJOR FACTORS	MINOR FACTORS
Facial pain or pressure (in conjunction with other nasal symptoms)	Headache
Facial fullness	Halitosis
Nasal obstruction	Fatigue
Nasal discharge or purulence	Dental pain
Fever (in acute rhinosinusitis)	Fever (in nonacute rhinosinusitis)
	Cough
	Ear pressure or fullness

Major symptoms of sinusitis (Table 426-1) include facial pressure, facial congestion or fullness, nasal obstruction, nasal discharge, and anosmia. Minor symptoms include headache, halitosis, fatigue, dental pain, cough, and ear pressure. Major signs include purulence in the nose noted on examination and, with acute sinusitis, fever. Pain is a frequent complaint with acute sinusitis but infrequent with chronic sinusitis. Patients with chronic sinusitis often note a dull facial pressure that seems to worsen with dependency. Patients with acute sinusitis may have discrete facial pain or dental pain but also have obvious purulent nasal discharge, often with a frank fever. It is important to note that facial pain is not a symptom of chronic sinusitis in the absence of other nasal signs and symptoms. Generally, sinusitis is thought to be present on the basis of at least two major factors, one major factor and two minor factors, or purulence on nasal examination.

DIAGNOSIS

History

A thorough history should probe whether patients have tried over-the-counter or prescription medications, including antihistamines, decongestants, mucolytics, analgesics, mast cell stabilizers, and even steroids, and whether they have helped improve the condition. In addition, other prescription medications have side effects that affect nasal physiology, including birth control pills, antihypertensive medications that cause systemic vasodilation, aspirin, steroids, and antibiotics. Specific questions regarding allergies are important, including seasonality or environmental triggers, the presence or absence of pets, food sensitivities, recent changes in environment, and living conditions, with a focus on old or new carpets, mattresses, furnace filters, or freshly painted interior walls. A patient should be questioned about past allergy skin testing or other testing.

A recent history of other family members or coworkers being ill suggests an infectious process. An astute physician often suspects an infectious process by noting the similarity and time course of symptoms in other patients; this information can be related to patients so that they know what to expect in terms of time course and recovery. A careful past medical history should allow one to determine whether relevant conditions such as previous nasal surgery or trauma, granulomatous diseases, cystic fibrosis (Chapter 89), rheumatologic conditions, immune deficiencies (Chapter 250), or other problems may be contributing factors. Unilateral nasal congestion raises concern for either an anatomic abnormality, such as septal deviation, perhaps related to previous trauma, a polyp or other neoplastic mass, or perhaps even a foreign body.

Physical Examination

The nose should be inspected with a nasal speculum to assess nasal septal anatomy (Fig. 426-1), the most caudal aspect of the inferior turbinates (Fig. 426-2), and the possibility of large nasal polyps (Fig. 426-3) or other masses. In patients with allergic rhinitis, the physical examination may reveal pale and swollen inferior turbinates, whereas copious nasal secretions are more apparent with viral infections. By spraying the nose with a topical decongestant such as phenylephrine (Neo-Synephrine) or oxymetazoline, the middle meatus, which is the air space between the middle turbinate and lateral nasal wall, can often be visualized to assess for nasal polyps or purulent discharge. Examination of the mouth and oropharynx, including the posterior pharyngeal wall, with a tongue blade if necessary, can sometimes identify a stream of postnasal discharge or pus. Sinus palpation and transillumination, although part of the art of medicine, are not sufficiently reliable for diagnosis. The patient's ability to open the mouth without limitation helps exclude trismus, which can sometimes be caused by a deep neck infection.

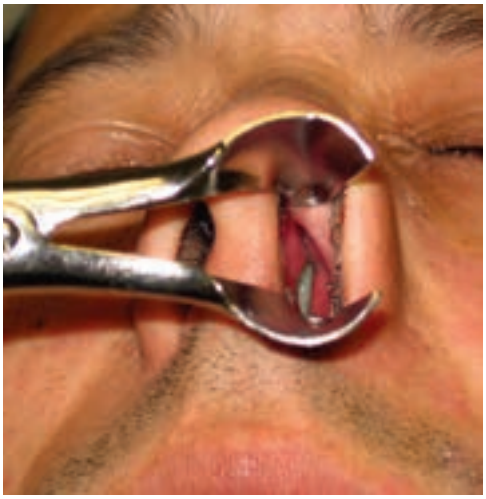

FIGURE 426-1. Purulent drainage from the middle meatus seen on anterior rhinoscopy.

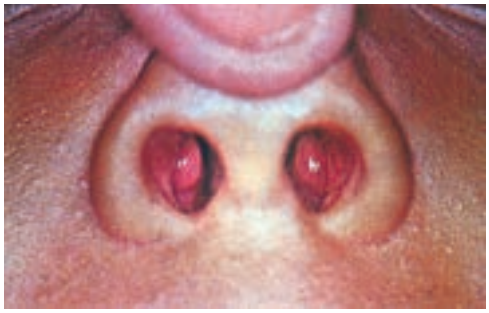

FIGURE 426-2. Edematous inferior turbinates narrowing the nasal airway in a patient with hay fever. Physical examination. (From Dhillon RS, East CA, eds. *Ear, Nose and Throat and Head and Neck Surgery,* 2nd ed. Edinburgh: Churchill Livingstone; 1994:34.)

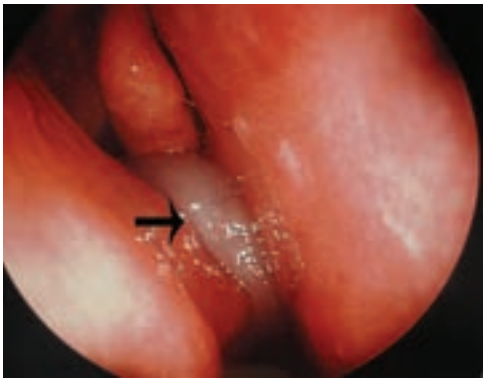

FIGURE 426-3. Nasal polyp in the right nasal cavity. A polyp is seen on the right side just inferior to the middle turbinate, next to the nasal septum. It is paler than the surrounding tissue.

A complete examination of the head and neck should be performed to look for signs of recent or old trauma such as ecchymosis under the eyelids, swelling of the soft tissue of the face, or deviation of the nasal dorsum. The neck should be palpated for adenopathy (Chapter 168) or other masses.

A basic eye examination should be performed to assess pupillary function, extraocular movements, and possible nystagmus (Chapter 424). An ear examination should be conducted to assess the tympanic membranes bilaterally. In patients with an abnormality of the tympanic membrane or concomitant complaints of hearing loss or disequilibrium (Chapter 428), pneumatoscopy using an air bulb attached to the otoscope can be used to insufflate the ear canal and assess for mobility of the tympanic membrane; decreased mobility suggests a middle ear effusion. Weber and Rinne testing using a 512-Hz tuning fork screens for conductive hearing loss, especially unilateral loss.

Endoscopic examination of the nose, almost always performed by a specialist, is the "gold standard" for evaluating rhinitis and sinusitis. A flexible or rigid fiberoptic scope can allow fine inspection of the septum, turbinates, middle meatus, and sphenoethmoid recess, as well as direct inspection of the nasopharynx, orifice of the eustachian tube, and fossa of Rosenmüller, which is just rostral to the eustachian tube in the nasopharynx and is often the site of origin of nasopharyngeal carcinoma (Chapter 190). Flexible endoscopy can be used to further inspect the oropharynx, larynx, and most of the hypopharynx (Chapter 429).

Laboratory Findings

Cultures

Cultures of the nostril or lower nasal cavity are not typically useful and are not recommended. An endoscopically guided culture of the middle meatus by a specialist may help guide treatment for acutely ill immunocompromised patients, for patients suspected of having acute bacterial rhinosinusitis, for patients with refractory chronic rhinosinusitis, or for those whose sinusitis is suspected of causing secondary meningitis, epidural or subdural abscess, brain abscess, orbital involvement, or cavernous sinus thrombosis.

Other Tests

A nasal smear can reveal eosinophils, which is consistent with allergic rhinitis (Chapter 251). Likewise, skin testing or radioallergosorbent testing can help pinpoint allergic triggers (Chapter 249). In patients with acute sinusitis, a white blood cell count with differential may be useful. In patients with chronic sinusitis, serum immunoglobulin levels can be helpful: highly elevated immunoglobulin E (IgE) levels can raise suspicion for allergic fungal sinusitis, whereas low levels of IgG and other subclasses suggest immunodeficiency (Chapter 250). If the patient has chronic nasal crusting as a primary complaint, screening serologic tests for sarcoid (Chapter 95), granulomatosis with polyangiitis (formerly called Wegener's granulomatosis) (Chapter 270), T-cell lymphomas (Chapter 185), syphilis (Chapter 319), tuberculosis (Chapter 324), Sjögren's syndrome (Chapter 268), and other chronic inflammatory diseases can be considered. Relatively rare infections such as rhinoscleroma may also be present, so biopsy and cultures may be indicated to help reveal a diagnosis. Use of illicit substances should be considered because cocaine and other illicit drugs may cause chronic nasal crusting. In a patient with a lifelong history of sinusitis since childhood, cystic fibrosis should also be considered (Chapter 89).

Imaging

Non-contrast-enhanced computed tomography (CT) is indicated for patients with known or suspected rhinitis and sinusitis. CT is generally performed to document the presence of disease or the effects of treatment to improve the disease. A CT with contrast is used to evaluate complications of sinusitis. Finally, CT is critical before any surgical treatment of the sinuses because of the anatomic information that it provides the surgeon. Opacification or other findings on CT (Fig. 426-4) can sometimes differentiate among the various causes of sinusitis. Plain films have little utility and are not generally recommended. Magnetic resonance imaging (MRI) is occasionally helpful, especially when evaluating tumors or processes that erode bone and are proximate to the brain or eye.

Differential Diagnosis

A rapid onset of sinus-related symptoms suggests a viral upper respiratory infection, especially if the patient also has typical systemic symptoms, such as arthralgia, myalgia, fever, chills, gastrointestinal symptoms, and cough in addition to nasal congestion, postnasal drip, and headache. By comparison, acute bacterial rhinosinusitis causes facial pressure and purulent postnasal discharge. Viral disease can progress to a secondary bacterial infection, which can become chronic. An acute onset of inhalant allergy is often seasonal or can be traced to a particular precipitant (Chapter 251). Allergic rhinitis typically responds to an empirical trial of antihistamines, whereas viral or bacterial rhinitis does not.

Chronic sinusitis must be differentiated from rhinitis, which is not accompanied by the same degree of incessant inflammation. Types of rhinitis include gustatory rhinitis associated with eating, rhinitis of pregnancy, rhinitis related to abuse of topical vasoconstrictors (rhinitis medicamentosa),

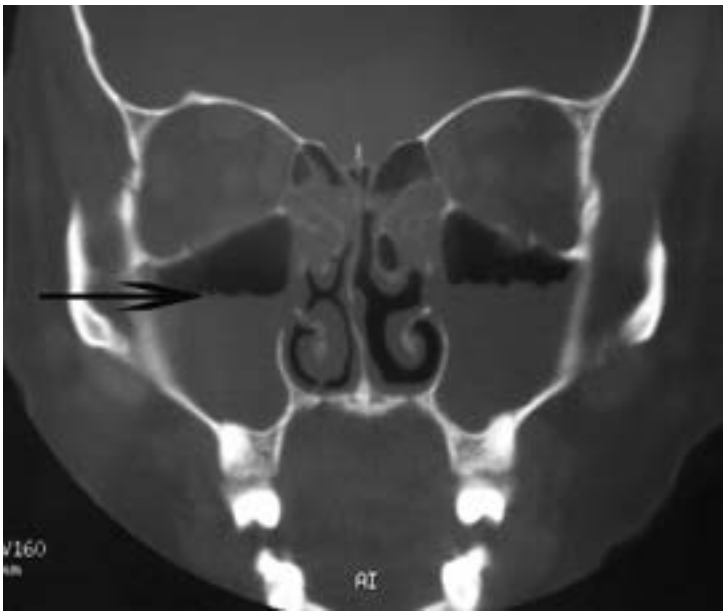

FIGURE 426-4. **Coronal computed tomography showing bilateral acute pansinusitis. There are bilateral fluid levels in the maxillary sinuses that, if aspirated, can be sent for microbiology.**

rhinitis associated with illicit drug use (e.g., cocaine or methamphetamine), rhinitis of aging, vasomotor rhinitis presumably related to a hypersecretory state mediated by the parasympathetic nervous system, and perennial allergic rhinitis, whose hallmark is a lack of seasonality.

Chronic sinusitis may be caused by chronic viral infection, chronic bacterial infection, chronic fungal infection, and chronic allergy. It is often difficult to pinpoint a specific cause, but the common underlying factor is often inflammatory in nature. Although maxillary antral punctures were used for diagnosis and treatment in the pre-CT era, endoscopically guided culture techniques combined with CT are now the standard of care, except for acute bacterial maxillary sinusitis, for which surgical decompression is desirable, or for some cases of refractory sinusitis in immunocompromised patients or patients in the intensive care unit, where direct culture can guide antibiotic therapy.

CT can reveal mucoceles, which are blocked individual sinuses that continue to secrete mucus and can slowly erode bone, expand to involve the eye and brain, or become acutely infected. A mycetoma, which is an isolated "fungus ball" in a sinus, has a characteristic hyperdensity within a sinus opacification. Mycetomas (Chapter 342) are noninvasive but may erode bone through pressure necrosis over a long period.

Mucus retention cysts, often present in the maxillary sinus, are manifested as a spherical opacification; an estimated 10% of the population has a mucus retention cyst, which is usually asymptomatic.

TREATMENT

Rx

Medical Therapy

Infectious Rhinitis

Viral rhinitis is treated with supportive care, including fluid replacement and treatment of the febrile component of the syndrome with acetaminophen or nonsteroidal anti-inflammatory medications. Steam has a mild decongestant effect, and vitamin C and good nutrition may help hasten the resolution of symptoms. Oral decongestants (e.g., pseudoephedrine, 120 mg every 12 hours for several days), mucolytics (e.g., guaifenesin, 200 to 400 mg every 4 to 6 hours for several days), and ipratropium bromide (0.03 or 0.06%, two sprays on each side of the nose every 12 hours for several days) are of potential benefit.

For clinically diagnosed acute purulent rhinitis or acute rhinosinusitis of less than 10 days' duration, antibiotics are of little benefit because a diagnosis of bacterial rhinosinusitis based on the history and physical examination is quite inaccurate.[1] For example, a 10-day course of amoxicillin does not reduce symptoms at day 3 or 10 compared with placebo among patients with acute rhinosinusitis, and it only slightly improves symptoms at day 7.[1]

Because the potential side effects of antibiotics are not trivial, they should be reserved for patients with a high probability of bacterial infection. The best clinical predictors of the presence of *acute* bacterial rhinosinusitis rather than viral rhinosinusitis include persistent symptoms for 10 or more days without evidence of clinical improvement; high fever (>39° C or 102° F) with purulent nasal discharge or facial pain for at least 3 to 4 consecutive days; or the onset of worsening symptoms more than 5 days after the onset of an apparent viral upper respiratory tract infection. In patients who meet one or more of those three criteria, the Infectious Diseases Society of America[1] recommends empirical antibiotic therapy, preferably with amoxicillin-clavulanate (875 mg/125 mg orally twice daily, increasing to 2000 mg/125 mg orally twice daily in patients with fever greater than 39° C or 102° F, immunocompromise, or recent antibiotic use). In patients who are allergic to penicillin, the best alternatives are doxycycline (100 mg orally twice daily) or a fluoroquinolone (e.g., levofloxacin 500 mg orally daily or moxifloxacin 400 mg orally daily); by comparison, macrolides, trimethoprim-sulfamethoxazole, and second- and third-generation oral cephalosporins are not recommended because of high levels of resistance. The usual course of therapy is 5 to 7 days, regardless of the medication chosen. Intranasal saline irrigations, using either physiologic or hypertonic saline, may be a useful adjunct in patients with acute bacterial rhinosinusitis,[1] but neither topical decongestants nor antihistamines are useful. If patients worsen despite 72 hours of treatment or do not improve after 5 to 7 days, further evaluation should include CT to localize the infection, and cultures—either by direct sinus aspiration or endoscopically guided cultures of the middle meatus; other cultures are unreliable.

Chronic sinusitis is a term that encompasses multiple pathophysiologic mechanisms and implies a prolonged course of sinus symptoms that have been refractory to symptomatic treatment over a period of at least 3 months. Chronic sinusitis presents with nasal congestion, nasal drainage, facial pressure, and sometimes anosmia. Unlike acute sinusitis, patients with chronic sinusitis do not typically have fever or severe headache. Corticosteroids, either in a topical spray (e.g., triamcinolone acetonide, two 55-µg sprays to each side of the nose every day; mometasone furoate, two 50-µg sprays to each naris every day; fluticasone propionate, two 50-µg sprays to each naris every day; or budesonide, two 32-µg sprays to each naris every day) for 6 weeks or delivered in an oral tapering dose (prednisone 40 mg per day for 5 days, followed by 30 mg per day for 5 days, followed by 20 mg per day for 5 days, followed by 10 mg per day for 5 days; or methylprednisolone 4 mg tablets beginning with 24 mg the first day and tapering by 4 mg each subsequent day for 6 days) are the mainstay of treating the symptoms of chronic rhinosinusitis. Endoscopically obtained cultures of the middle meatus can help define which patients may improve with culture-guided antibiotic treatment. Antifungal agents, including itraconazole in an oral or aerosolized form and amphotericin B in an aerosolized form, do not appear beneficial in the treatment of typical chronic sinusitis.[1]

Allergic Rhinitis

Allergic rhinitis (Chapter 251) responds to various antihistamines, such as diphenhydramine hydrochloride (25 to 50 mg every 4 to 6 hours), loratadine (5 mg twice a day or 10 mg a day), cetirizine hydrochloride (10 mg a day), fexofenadine hydrochloride (60 mg twice a day or 120 mg/day), and topical nasal steroids, including triamcinolone acetonide (two sprays [55 µg] to each side of the nose every day), mometasone furoate (two sprays [50 µg] to each naris every day), fluticasone propionate (two sprays [50 µg] to each naris every day), and budesonide (two sprays [32 µg] to each naris every day). Oral steroids such as prednisone and methylprednisolone in various doses are sometimes useful as well. Allergic desensitization is sometimes recommended when a discrete allergen elicits a strong reaction in a patient. Allergic desensitization, through injections or the sublingual route, may specifically be beneficial for some inflammatory disorders, such as allergic rhinitis.

Topical nasal steroids may be of benefit when acute rhinosinusitis is confirmed by either radiography or nasal endoscopy.[1] However, topical intranasal steroids are also not of proven benefit for treating sinusitis diagnosed by primary care physicians on purely clinical grounds[1] or for treating the common cold.[1] Topical nasal steroids are commonly used with good clinical effect in patients with chronic rhinosinusitis.

Surgical Therapy

Patients who have severe symptoms, who fail to respond to therapy, or who have unusual, or resistant, or recurrent infections should be sent to a specialist for further evaluation and treatment (Table 426-2). Surgery is recommended in patients with benign neoplasms, mucoceles, juvenile nasopharyngeal angiofibroma, and some types of malignancies. Surgery can correct septal deviations and anatomically related nasal obstruction. Surgery on the inferior turbinates may be beneficial for refractory rhinitis. Functional endoscopic surgery, which is designed to preserve mucociliary function and is performed with endoscopes through the nostril without skin incisions, can be useful for recurrent acute sinusitis and chronic rhinosinusitis.[1]

Nasal Polyps

During the evaluation of symptoms of rhinitis or sinusitis, the physical examination may reveal nasal polyps. Nasal polyps often present with symptoms of

TABLE 426-2 WHEN TO REFER A PATIENT WITH PRESUMED BACTERIAL RHINOSINUSITIS TO AN ENT SPECIALIST

- Temperature >39°C (>102° F); orbital edema; severe headache, visual disturbance, altered mental status, meningeal signs
- Failure to respond to more than two courses of antimicrobial therapy
- Nosocomial infection, anatomic abnormalities
- Immunocompromise or multiple comorbidities
- Unusual or resistant pathogens
- Fungal sinusitis or granulomatous disease
- Recurrent episodes suggesting chronic sinusitis

Adapted from Chow AW, Benninger MS, Brook I, et al. IDSA clinical practice guideline for acute bacterial rhinosinusitis in children and adults. *Clin Infect Dis.* 2012;54:e72-e112.

nasal blockage and anosmia along with typical symptoms of rhinitis. When polyps are present, the nasal congestion is often unrelenting. Sometimes, patients with prolonged symptoms will present with a visible mass in their nostril. Rarely, facial asymmetry or orbital involvement will be the presenting sign of long-ignored nasal polyps. Patient with nasal polyps may be more likely to complain of facial or ear pain than patients with rhinitis without polyps.[2]

Nasal polyps typically begin near the ethmoid sinuses in the middle meatus and extend into the nose, where they block the nasal airway and/or the sinuses. Nasal polyps may be caused by chronic inflammation and also often occur as part of a rare metabolic disorder of arachidonic acid metabolism triggered by exogenous aspirin intake—known as aspirin-exacerbated respiratory disease. Also known as Samter's triad, patients with this syndrome have asthma that is exacerbated by aspirin ingestion, a skin rash precipitated by aspirin, and often have difficult-to-control chronic nasal polyposis. This constellation of symptoms is thought to be caused by inflammation elicited by leukotrienes, which are upregulated by the prostaglandin blockade caused by aspirin and sometimes by other nonsteroidal anti-inflammatory drugs. Human papillomavirus (Chapter 373) may cause an inverted papilloma, which presents as a polyp causing unilateral nasal obstruction. This initially benign neoplasm responds to surgical excision but can transform to frank malignancy. Polyps are also seen in patients with cystic fibrosis, especially patients with the delta F508 mutation (Chapter 89). They are also seen in allergic fungal sinusitis, which is manifested by an elevated IgE level, positive fungal cultures (usually for aspergillosis), Charcot-Leyden crystals on histopathology, characteristic densities on CT, and nasal polyposis that is often, but not always, unilateral. Antral choanal polyps may extend into the nasal cavity or nasopharynx and cause obstruction.

Nasal polyps will be visible in a careful examination (see Fig. 426-3), and their extent can be shown on a CT scan (Fig. 426-5). Unilateral nasal polyposis is suggestive of antral choanal polyps, malignancy, inverted papilloma, or allergic fungal sinusitis; early biopsy is recommended.

Benign inflammatory nasal polyps frequently respond to oral steroids, either in a tapered burst dose or, in rare cases, in small amounts of titrated daily oral steroids such as prednisone (40 mg per day for 5 days, followed by 30 mg per day for 5 days, followed by 20 mg per day for 5 days, followed by 10 mg per day for 5 days) or methylprednisolone (beginning with 24 mg the first day and tapering by 4 mg each subsequent day for 6 days). Topical steroids are also efficacious for treating nasal polyps.

Surgery for benign nasal polyposis can improve symptomatic control and reduce the need for oral steroids. Surgery is always recommended for inverted papillomas, antral choanal polyps, and mucoceles, and surgery is likely to be helpful if acute sinusitis has caused central nervous system complications such as brain abscess (Chapter 413), meningitis (Chapter 412), epidural abscess, subdural abscess, or orbital abscess. Occasionally, surgery will be required when an untreated and aggressively growing polyp causes orbital or skull base erosion. Allergic fungal sinusitis is often treated with a combination of surgery, corticosteroids, and sometimes immunotherapy.

Epistaxis

For a patient with epistaxis, it is first critical to determine the severity of the blood loss. Persistent bleeding may result from warfarin, antiplatelet agents, or any underlying platelet (Chapters 172 and 173) or clotting deficiency (Chapter 174). Physical examination should focus on inspection of the anterior septum, which is the most frequent point of origin for epistaxis. Frequently, dilated blood vessels on the caudal septum can be seen with anterior

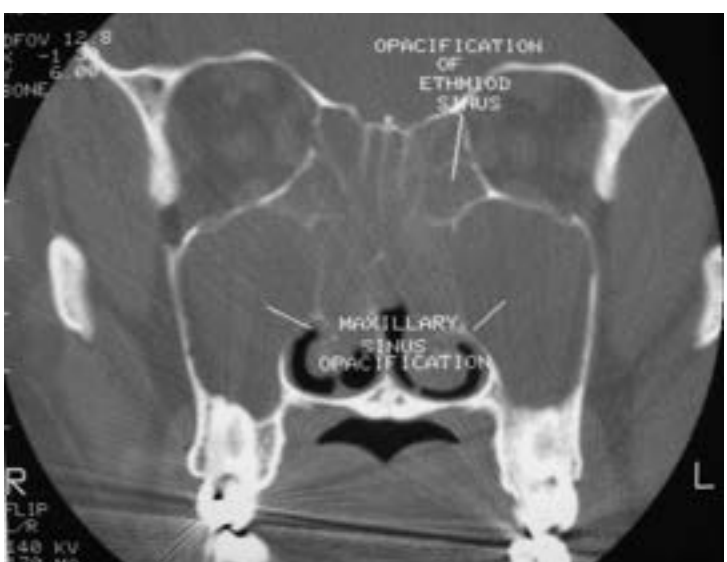

FIGURE 426-5. Computed tomography showing bilateral nasal polyposis of a chronic nature.

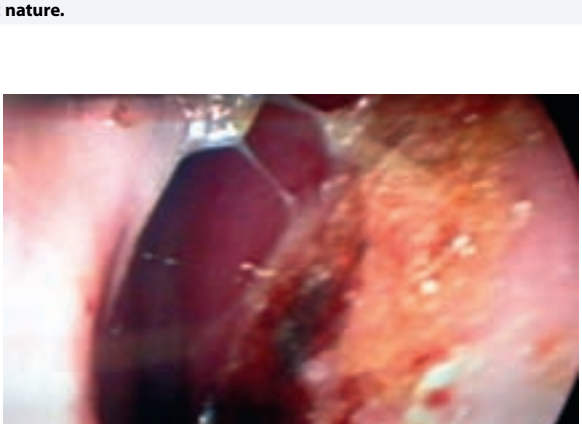

FIGURE 426-6. Dilated nasal vessels and crusting typical of a patient with epistaxis.

rhinoscopy (Fig. 426-6). The combination of unilateral otitis media, epistaxis, nasal congestion, and a neck mass would be concerning for nasopharyngeal carcinoma. Rare tumors that can arise with bleeding include juvenile nasopharyngeal angiofibromas in male patients.

Epistaxis can be treated by local pressure, packing (using nasal sponges, balloons, or by ½-inch by 72-inch gauze impregnated with petroleum jelly), humidification, and hydration. Hospitalization and transfusion are rarely required. Offending medications should be reduced in dose or discontinued temporarily if possible. Topical vasoconstrictive medication such as oxymetazoline spray, two sprays on each side of the nose every 12 hours for 3 days, can help prevent persistent epistaxis. Occasionally, lasers or other types of cautery are used to improve the problem. At times, surgical arterial clipping or interventional neuroradiologic arterial occlusion can address a specific bleeding area.

EAR PAIN

DEFINITION

Ear pain (Table 426-3) is discomfort perceived by a patient in the area of the temporal bone. Although the discomfort can often be localized by the patient, at times the cause of the discomfort may in fact be distant from the site where the pain is felt. This referred pain can be due to problems in the oral cavity, oropharynx, hypopharynx, or larynx.

TABLE 426-3 CAUSES OF OTALGIA

CAUSES OF OTALGIA	EXTERNAL EAR	MIDDLE EAR	UPPER AERODIGESTIVE TRACT
Likely	Otitis externa Herpes zoster oticus Chondritis Foreign body	Acute otitis media Acute eardrum perforation Barotrauma Chronic otitis media with impending complication	Tonsillitis Tonsil abscess Deep neck abscess Tumor (especially base of the tongue, tonsil, hypopharynx, larynx, nasopharynx)
Unlikely	Malignant otitis externa Tumor	Tumor	

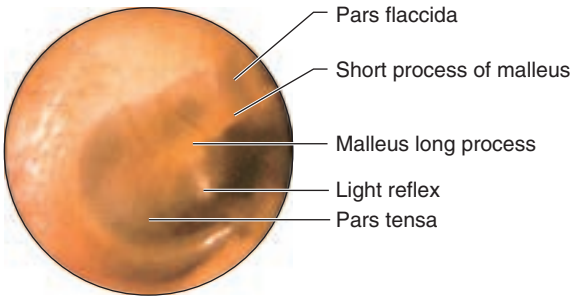

FIGURE 426-7. A normal tympanic membrane. (From Dhillon RS, East CA, eds. *Ear, Nose and Throat and Head and Neck Surgery,* 2nd ed. Edinburgh: Churchill Livingstone; 1994:2.)

PATHOBIOLOGY

The ear is well supplied with sensory nerves and is positioned on the side of the skull. The ear is divided into the outer ear, or pinna, and the ear canal; the middle ear, which encompasses the tympanic membrane and ossicles (Fig. 426-7); and the inner ear, which consists of the cochlea and the vestibular canals, including the utricle and saccule. In general, otalgia is due to problems in the outer or middle ear. The trigeminal nerve innervates the anterior-superior quadrant of the pinna, whereas the C2 and C3 cervical cutaneous nerves innervate the rest of the majority of the outer ear. However, there are contributions by the 9th and 10th nerves in the ear canal and even a small patch of sensory innervation by the 7th nerve in the posterior superior ear canal. It is the overlap in distribution of the 9th and 10th cranial nerves that establishes the anatomic basis for referred otalgia in diseases of the oral cavity, oropharynx, and larynx. Therefore, ear pain may be due to inflammatory conditions of the skin of the outer ear, the ear canal, or the middle ear, or it may be due to disease processes unrelated to the ear itself.

CLINICAL MANIFESTATIONS

Patients with ear pain often have complaints referable directly to the ear itself. In cases of otitis externa, frankly obvious erythema and swelling of the skin of the ear canal may be present. Even minute physical manipulation of the ear may be excruciating. In chondritis of the pinna, which may be related to rheumatologic disorders, infection, or trauma, the entire pinna may be swollen and painful. Hearing loss accompanying otalgia may indicate middle ear disease, especially otitis media. Patients sometimes complain of pain in the ear after air travel or driving from a mountainous region. Quick changes in pressure, such as encountered in scuba diving, may indicate barotrauma (Chapter 94), in which the eustachian tube is unable to compensate rapidly enough for the changes in pressure that are encountered. Pain may also be a post-traumatic symptom from relatively minor percussion injury, more severe head trauma, or percussion injury related to a blast. Pain related to noise exposure may also indicate damage to the middle ear or even the inner ear. Deep-seated boring pain over the temporal area accompanied by retro-orbital pain can be due to petrous apex disease, including petrous apicitis.

DIAGNOSIS

History

A patient with ear pain should be asked to reveal the location of the discomfort, the duration of the symptoms, and any activities related to onset of the condition. As an example, recent swimming would make otitis externa ("swimmer's ear") more likely, whereas a recent upper respiratory infection with hearing loss would suggest otitis media. Questions should address possible hearing loss, vertigo, otorrhea, hoarseness, voice change, dysphagia, odynophagia, dyspnea, hemoptysis, hematemesis, and weight loss. A social history with specific concentration on tobacco and alcohol use should be obtained. A possible family history of upper aerodigestive tract and nasopharyngeal carcinoma should be sought. A past surgical history can reveal distant ear or throat surgery.

Physical Examination

A complete head and neck examination, including general assessment for trauma and a basic eye examination, is required. The outer ear and pinna should be examined first. The ear canal should first be palpated and then

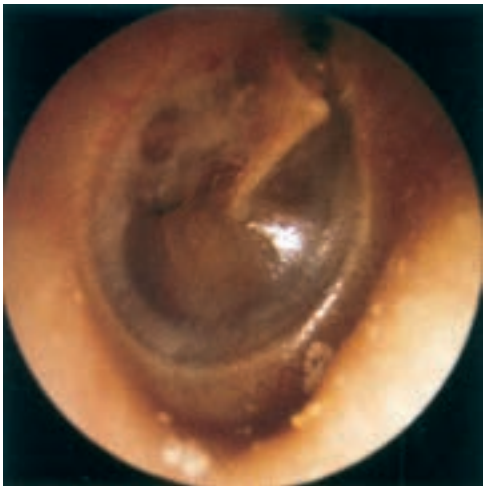

FIGURE 426-8. Otoscopic appearance in otitis media with effusion. The handle and short process of the malleus are brought into relief by retraction of the eardrum. There is a slightly yellow appearance of the eardrum related to the middle ear effusion. (From Dhillon RS, East CA, eds. *Ear, Nose and Throat and Head and Neck Surgery,* 2nd ed. Edinburgh: Churchill Livingstone; 1994:7.)

inspected. An otoscope with a pneumatic bulb attachment is critical to establish the presence or absence of a middle ear effusion. Inspection of the tympanic membrane should be accomplished with notations made about patency and perforation, translucency of the eardrum, position and definition of the malleus, and the eardrum's mobility with the ear canal sealed and a puff of air delivered by the pneumatic bulb. Abnormalities may be caused by infection (Fig. 426-8) or barotrauma (Fig. 426-9). Examination with a 512-Hz tuning fork should be performed to determine lateralization of the sound (Weber test) and whether air conduction is superior to bone conduction (Rinne test). Facial nerve function should be assessed (Chapter 396) by determining whether the patient can raise the eyebrows, close the eyes, wrinkle the nose, and purse the lips. The presence or absence of nystagmus should be recorded. Inspection of the nose, oral cavity, oropharynx, and neck should be accompanied by cranial nerve examination (Chapter 396). Palpation of the tongue and tonsils is especially important if the ear pain is intense and persistent. A careful neck examination should be performed to look for masses. Oral cavity infections (Chapter 425), such as a peritonsillar abscess or severe tonsillitis, may arise as ear pain, and the physical examination should reveal trismus, erythema, mass effect, and other common signs of pharyngitis.

Laboratory

An audiogram can assess hearing loss (Chapter 428). A tympanogram measures compliance of the middle ear system and is an accurate method for diagnosis of otitis media. Cultures are rarely performed because they require tympanocentesis, and cultures of the external ear can reveal a vast variety of organisms that are often treated empirically with antibiotics. If a fever and middle ear effusion are present and neck stiffness is found on physical examination, lumbar puncture may rarely be recommended.

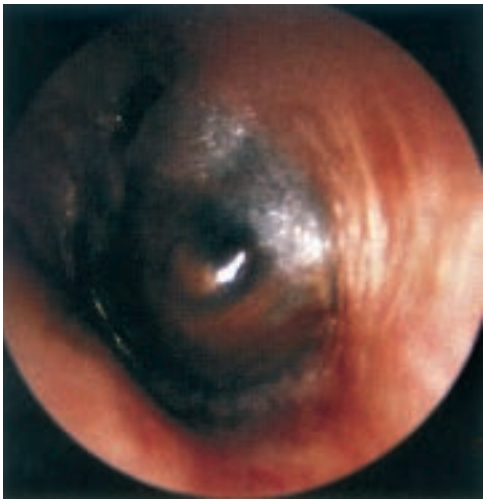

FIGURE 426-9. Blood in the middle ear (hemotympanum). Causes include otitic barotrauma, secretory otitis media, and a high jugular bulb. (From Dhillon RS, East CA, eds. *Ear, Nose and Throat and Head and Neck Surgery,* 2nd ed. Edinburgh: Churchill Livingstone; 1994:26.)

Imaging

In general, imaging is indicated if complications of acute or chronic otitis media are suspected or to look for occult causes of otalgia in the upper aerodigestive tract. If a patient is suspected of having meningitis, epidural or subdural abscess, brain abscess, or sagittal sinus thrombosis, imaging is mandatory. Imaging is also useful for operative planning in patients with chronic otitis media or (rarely) to evaluate for the presence of tumors in the middle or external ear.

Differential Diagnosis

Otitis externa, which is an infection of the skin of the ear canal, is often due to manipulating the ear after swimming or trying to scratch an ear canal that itches because of skin irritation. Patients exhibit erythema of the canal skin and extreme pain on manipulation of the ear canal. In the presence of concomitant cranial neuropathies, especially in diabetic or otherwise immunocompromised patients, malignant otitis externa with osteomyelitis should be suspected. Inspection of the tympanic membrane may reveal fluid consistent with otitis media; the tuning fork examination should support the presence of conductive hearing loss. Vesicles on the conchal portion of the pinna, especially when accompanied by facial nerve paralysis, strongly suggest herpes zoster oticus with Ramsay Hunt syndrome (Chapter 375). Perforation of the eardrum suggests either acute or chronic otitis media, traumatic perforation, or possibly cholesteatoma (Chapter 428) if the perforation is in the posterior-superior quadrant. Chronic draining otorrhea of long standing with a deep boring pain and perforation of the tympanic membrane suggests a complication of otitis media.

If findings on ear and cranial nerve examination are negative but the patient's complaints of otalgia are persistent, special effort needs to be made to visualize the upper aerodigestive tract, including the nasopharynx, oral cavity, oropharynx, larynx, and hypopharynx, to be sure that infection or tumor is not present in these hard-to-examine areas. MRI can be very useful in these cases.

TREATMENT

Rx

Otitis externa is often treated with office suctioning of debris under a microscope and the application of antibiotic drops (ciprofloxacin, tobramycin, neomycin, polymyxin B), with or without hydrocortisone in various combinations.[3] Frequently, a small wick or sponge is placed in the ear canal to help maintain patency of the canal and allow facile application of the medications (Fig. 426-10). For otitis media, oral antibiotic treatment is directed at eradicating *Haemophilus influenzae, Moraxella catarrhalis, Streptococcus pneumoniae,* and *Staphylococcus aureus* with amoxicillin or erythromycin as for sinusitis. The benefit is notable for children 2 years or younger with bilateral otitis media and for older children with otitis plus otorrhea, whereas other patients can be observed without antibiotics. In general, antibiotics provide somewhat better short-term outcomes but at the expense of significantly more diarrhea and rashes.[A1,A2] Interestingly, the natural history of acute otitis media is acute perforation of the eardrum, which often results in otorrhea and relief of pain. Most middle ear effusions clear spontaneously within 3 months regardless of whether they are treated. Most perforations of an eardrum caused by trauma heal without surgical intervention, but if an eardrum perforation persists for more than about 3 months, surgical closure and the use of tympanoplasty with or without mastoidectomy can be contemplated. Chronic draining perforations, especially if located in the posterior-superior quadrant of the tympanic membrane, may portend the presence of cholesteatoma and may require tympanomastoid surgery.

In patients in whom herpes zoster is suspected, acyclovir can be started at 800 mg by mouth five times per day for 7 days, with or without prednisone (Chapter 375). Intracranial complications of otitis media often need to be addressed surgically.

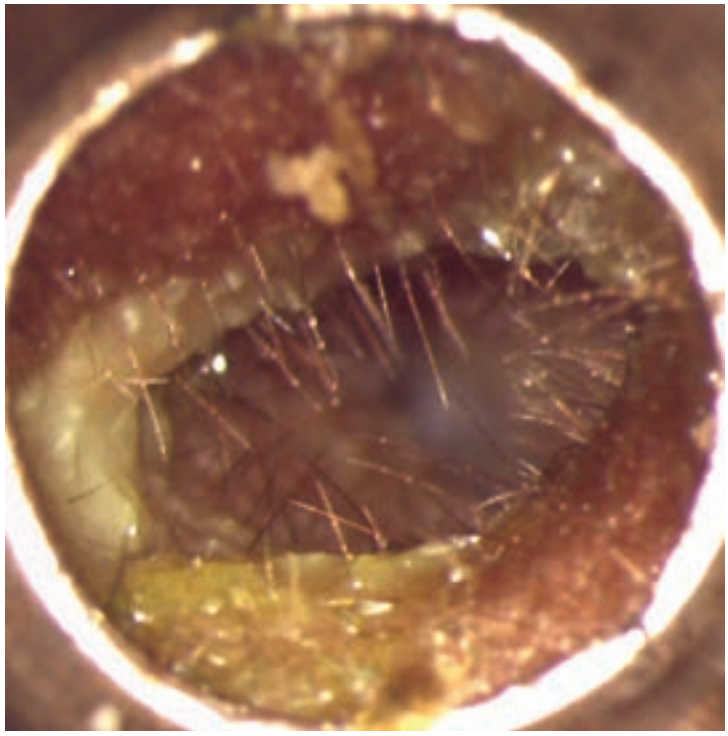

FIGURE 426-10. Otitis externa. Otitis externa in a patient's left ear with the tympanic membrane in the distance. Exudate and erythema are present. The ear canal is quite painful, and a wick may be necessary to maintain patency of the external auditory canal.

Grade A References

A1. Kenealy T, Arroll B. Antibiotics for the common cold and acute purulent rhinitis. *Cochrane Database Syst Rev*. 2013;6:CD000247.

A2. Lemiengre MB, van Driel ML, Merenstein D, et al. Antibiotics for clinically diagnosed acute rhinosinusitis in adults. *Cochrane Database Syst Rev*. 2012;10:CD006089.

A3. Garbutt JM, Banister C, Spitznagel E, et al. Amoxicillin for acute rhinosinusitis: a randomized controlled trial. *JAMA*. 2012;307:685-692.

A4. Wei CC, Adappa ND, Cohen NA. Use of topical nasal therapies in the management of chronic rhinosinusitis. *Laryngoscope*. 2013;123:2347-2359.

A5. Orlandi RR, Smith TL, Marple BF, et al. Update on evidence-based reviews with recommendations in adult chronic rhinosinusitis. *Int Forum Allergy Rhinol*. 2014;4(suppl 1):S1-S15.

A6. Zalmanovici TA, Yaphe J. Intranasal steroids for acute sinusitis. *Cochrane Database Syst Rev*. 2009;4:CD005149.

A7. Williamson IG, Rumsby K, Benge S, et al. Antibiotics and topical nasal steroid for treatment of acute maxillary sinusitis: a randomized controlled trial. *JAMA*. 2007;298:2487-2496.

A8. Hayward G, Thompson MJ, Perera R, et al. Corticosteroids for the common cold. *Cochrane Database Syst Rev*. 2012;8:CD008116.

A9. Smith TL, Kern R, Palmer JN, et al. Medical therapy vs surgery for chronic rhinosinusitis: a prospective, multi-institutional study with 1-year follow-up. *Int Forum Allergy Rhinol*. 2013;3:4-9.

A10. Poetker DM, Jakubowski LA, Lal D, et al. Oral corticosteroids in the management of adult chronic rhinosinusitis with and without nasal polyps: an evidence-based review with recommendations. *Int Forum Allergy Rhinol*. 2013;3:104-120.

A11. Rudmik L, Schlosser RJ, Smith TL, et al. Impact of topical nasal steroid therapy on symptoms of nasal polyposis: a meta-analysis. *Laryngoscope*. 2012;122:1431-1437.

A12. Tähtinen PA, Laine MK, Huovinen P, et al. A placebo-controlled trial of antimicrobial treatment for acute otitis media. *N Engl J Med*. 2011;364:116-126.

A13. Venekamp RP, Sanders S, Glasziou PP, et al. Antibiotics for acute otitis media in children. *Cochrane Database Syst Rev*. 2013;1:CD000219.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Chow AW, Benninger MS, Brook I, et al. IDSA clinical practice guideline for acute bacterial rhinosinusitis in children and adults. *Clin Infect Dis*. 2012;54:1041-1045.
2. Dietz de Loos DA, Hopkins C, Fokkens WJ. Symptoms in chronic rhinosinusitis with and without nasal polyps. *Laryngoscope*. 2013;123:57-63.
3. Rosenfeld RM, Schwartz SR, Cannon CR, et al. Clinical practice guideline: acute otitis externa. *Otolaryngol Head Neck Surg*. 2014;150(1 suppl):S1-S24.

REVIEW QUESTIONS

1. A patient without other systemic disease or immunocompromise presents with otitis externa confined to the external auditory canal. Which of the following interventions would be recommended?

A. The physician should prescribe an oral quinolone antibiotic.
B. The physician should order a computed tomographic scan with contrast of the temporal bone.
C. The physician should prescribe topical antibiotics.
D. The physician should obtain a neurosurgical consult.
E. The physician should recommend oral steroid medication.

Answer: C Otitis externa or "swimmer's ear" usually responds to topical antibiotic treatment. Oral or systemic antibiotics are specifically not recommended unless the patient is immunocompromised or the infection is spreading to the pinna cartilage of the external ear outside of the external auditory canal.

2. A patient presents to your office with recurrent epistaxis. A crust on the anterior septum is identified on the side of greatest bleeding. There is no active bleeding during the visit. The best course of action is which of the following choices?

A. The patient should be admitted to the hospital for interventional embolization of the internal maxillary artery.
B. The patient should be electively scheduled for a computed tomographic angiogram.
C. The patient should be given a prescription for a β-blocker.
D. The patient should be queried about aspirin use and be asked to take an "aspirin holiday."
E. A bone marrow biopsy should be obtained.

Answer: D Patients with epistaxis are often on antiplatelet drugs for a variety of reasons. On intake history, use of nonsteroidal anti-inflammatory drugs should be specifically sought because curtailing antiplatelet medication will often eliminate troublesome epistaxis.

3. An adult patient presents with left ear pain but no hearing loss, beginning 4 weeks ago and partially controlled with narcotic pain medication. On physical examination, there is no evidence of otitis externa or otitis media. Which of the following choices is the best next step?

A. A smoking history should be obtained, a neck examination should be performed, and the patient should be referred for upper airway endoscopic examination.
B. An empirical trial of oral antibiotics should be prescribed for presumptive otitis media.
C. An empirical trial of topical antibiotics should be prescribed for presumptive otitis externa.
D. A magnetic resonance image of the brain and temporal bone, with contrast, should be ordered.
E. The patient should be sent for a hearing test.

Answer: A Unexplained unilateral otalgia may be a referred pain from an upper aerodigestive tract lesion. Physical examination of the oral cavity, oropharynx, larynx, and hypopharynx should be performed. Tumors of the upper aerodigestive tract often present with an easily palpable, metastatic lymph node in the neck. Smoking, alcohol use, and human papillomavirus infection are risk factors for head and neck cancer.

4. A patient presents with nasal polyps visible on anterior rhinoscopy and also notes a history of asthma exacerbated by nonsteroidal anti-inflammatory drug use. What is the best treatment recommendation?

A. The patient should be scheduled for nasal surgery.
B. The patient should be scheduled for computed tomographic imaging of the sinuses.
C. The patient should be given a prescription for ciprofloxacin.
D. The patient should be scheduled for a sweat chloride test.
E. The patient should be given an oral prednisone taper.

Answer: E This patient likely has aspirin exacerbated respiratory disease, also called Samter's triad or "aspirin allergy." Oral corticosteroids can shrink nasal polyps and help control symptoms in patients with this disorder.

5. A patient presents to your office with a 4-day history of an upper respiratory infection that has prevented attendance at school. The physician should take which of the following steps?

A. The history suggests chronic infection, and the patient should be given oral antibiotics.
B. The history suggests acute infection, and the patient should be given oral antibiotics.
C. The history is consistent with a viral upper respiratory infection, and symptomatic treatment should be recommended.
D. The patient should be given a prescription for topical antifungal medication.
E. None of the above

Answer: C Several meta-analyses have shown that oral antibiotics or topical antibiotics provide no meaningful improvement in sinusitis during the first 7 to 10 days of treatment. Supportive care is the best option; it should include decongestants, hydration, fever control with antipyretics, and patience in allowing the symptoms to resolve.

427

SMELL AND TASTE

ROBERT W. BALOH AND JOANNA C. JEN

Millions of people suffer from disorders of taste and smell, but these disorders are often neglected because they are not fatal and, unlike abnormalities of vision and hearing, are not considered serious handicaps. Chemosensory disorders, however, often reduce the enjoyment and quality of life and are important to patients who suffer from them.

DEFINITION

The sensory receptor for taste, the taste bud, is made up of 50 to 150 cells arranged to form a pear-shaped organ. The lifespan of these cells is 10 to 14 days, and they are constantly being renewed from dividing epithelial cells surrounding the bud. Taste buds are located on the tongue, soft palate, pharynx, larynx, epiglottis, uvula, and upper third of the esophagus. The taste buds located on the anterior two thirds of the tongue and on the palate are innervated by the chorda tympani branch of the seventh cranial nerve. The ninth cranial nerve innervates the posterior third of the tongue. The ninth and tenth nerves innervate taste buds in the pharynx and larynx. Afferent signals from the taste buds project to the nucleus of the solitary tract in the medulla and then through a series of relays to the thalamus and postcentral somatosensory cerebral cortex (primary ipsilateral). Free nerve endings of the fifth cranial nerve are found on the tongue and in the oral cavity, and lesions involving these pathways can also alter taste perception.

Olfactory receptors lie in a roughly dime-sized area of specialized pigmented epithelium that arches along the superior aspect of each side of the nasal mucosa. Specialized bipolar sensory cells in this region thrust short receptor hairs into the overlying mucosa to detect aromatic molecules as they dissolve. Like taste buds, the specialized receptor portion of the bipolar neuron undergoes continuous renewal, with turnover occurring approximately every 30 days. Thin axons of the bipolar neurons course through small holes in the cribriform plate of the ethmoid bone to form connections in the overlying olfactory bulb on the ventral surface of the frontal lobe. From there, second- and third-order neurons project directly and indirectly to the prepiriform cortex and parts of the amygdaloid complex of both sides of the brain, which represents the primary olfactory cortex.

PATHOBIOLOGY

Pathology

Disorders of taste interfere with digestion because taste stimulants alter salivary and pancreatic flow, gastric contractions, and intestinal motility. Smell also contributes to the anticipation and ingestion of food because much of what is tasted is derived from olfactory stimulation during ingestion and chewing. An inability to detect noxious tastes and odors can result in food or gas poisoning, particularly in elderly subjects. In the extreme, chemosensory disorders can lead to overwhelming stress, anorexia, and depression. Genes that encode chemoreceptor proteins belong to the G protein–coupled receptor superfamily, which accounts for up to 1% of mammalian genomes.[1] Sequence diversity in these genes encodes unique structural motifs that bind to different ligands signaling different odors and tastes. Distinct and dedicated taste receptor cells express unique receptors to detect each of the five basic tastes: sweet (sensed by the heterodimers T1R1 and T1R3), umami (detected by the heterodimers T1R2 and T1R3), bitter (sensed by an estimated 30 T2Rs), sour (sensed by PKD2L1, with membrane-tethered carbonic anhydrase IV sensing carbonation), and salty (epithelial sodium channel). The taste receptor cells transform and transmit information to primary afferents through multiple cranial nerves (VII, IX, and X) that project to the solitary tract nucleus in the brain stem, with relay in the thalamus, and then onward to the primary cortex (Fig. 427-1).

Pathophysiology

Disorders of taste and smell can be divided into local, systemic, and neurologic categories (Table 427-1). The taste buds and the specialized receptor portion of the bipolar olfactory cells are constantly being renewed, and the

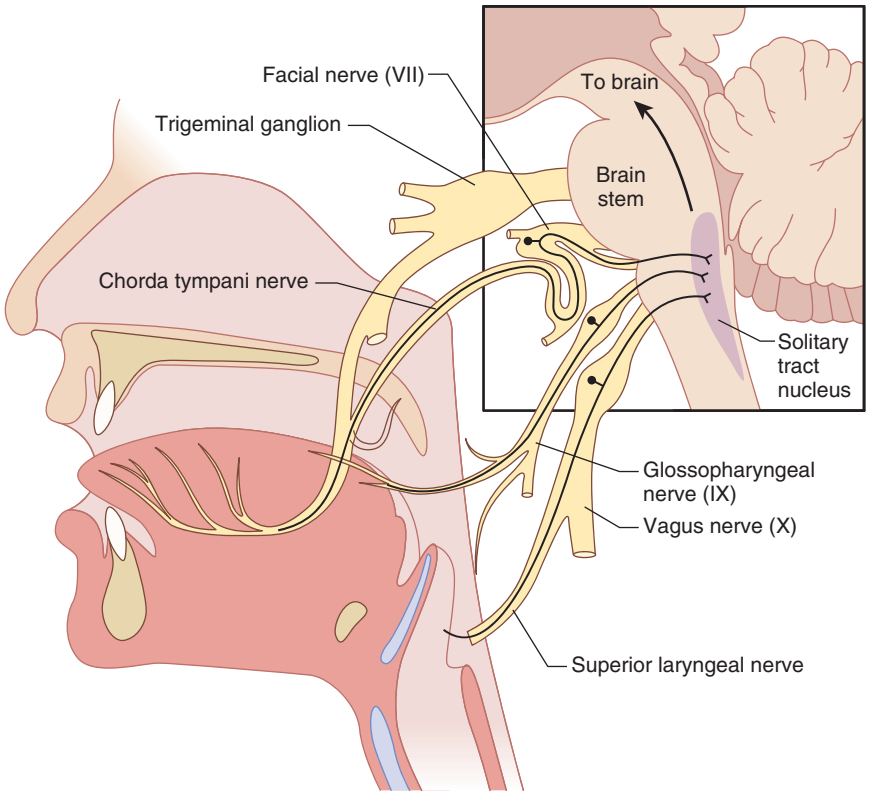

FIGURE 427-1. Anatomy of peripheral taste pathways. Taste information is transmitted from the mouth and the pharynx through multiple cranial nerves that project to the solitary tract nucleus in the brain stem, with relay in the thalamus before reaching the cortex. (Copyright 1999-2000 David Klemm. Reproduced from Bromley SM. Smell and taste disorders: a primary care approach. *Am Fam Physician.* 2000;61:427-436, 438.)

TABLE 427-1 COMMON CAUSES OF LOSS OF TASTE AND SMELL

	TASTE	SMELL
Local	Radiation therapy, oral infections, dentures, dental procedures	Allergic rhinitis, sinusitis, nasal polyposis, upper respiratory infection
Systemic	Cancer, renal failure, hepatic failure, nutritional deficiency (vitamin B_3, zinc), Cushing's syndrome, hypothyroidism, diabetes mellitus, infection (viral), drugs (antirheumatic and antiproliferative, e.g., corticosteroids, cisplatin, carboplatin, cyclophosphamide, doxorubicin, and methotrexate)	Renal failure, hepatic failure, nutritional deficiency (vitamin B_{12}), Cushing's syndrome, hypothyroidism, diabetes mellitus, infection (viral hepatitis, influenza), drugs (nasal sprays, antihistamines, decongestants, antibiotics, and antirheumatic and antiproliferative drugs that affect taste)
Neurologic	Bell's palsy, familial dysautonomia, multiple sclerosis	Head trauma, multiple sclerosis, Parkinson's disease, Alzheimer's disease, frontal tumor

process of renewal can be affected by nutritional, metabolic, and hormonal states as well as by therapeutic radiation, drugs, and age. For example, with interruption of mitosis by antiproliferative agents, return of normal taste function takes a minimum of 10 days, whereas return to normal olfactory function takes more than 30 days. Diuretics can block apical ion channels on a taste bud, and antifungal drugs inhibit cytochrome P-450–dependent enzymes at the level of the receptors. Numerous local conditions, such as colds and allergies, chronic sinusitis, and nasal polyposis, can influence the sense of smell by restricting airway patency. Accidental blows to the head can shear the fine axons of the bipolar olfactory neurons and result in loss of smell. Lesions of the fifth, seventh (chorda tympani), and ninth nerves can lead to disordered taste sensation. Olfactory and gustatory disturbances can serve as important diagnostic signs for focal neurologic lesions (e.g., frontal lobe tumors). Hallucinations of smell and taste occur in persons with epileptogenic lesions affecting the mesial temporal lobe and insular region, respectively. Finally, olfactory disturbances and hallucinations occur with a number of psychiatric illnesses (particularly depressive illness and schizophrenia).

CLINICAL MANIFESTATIONS

The most frequently encountered causes of loss of smell are local obstructive disease, viral infections, head injuries (Chapter 399), and normal aging (Chapter 25).[2,3] Patients can lose their sense of smell not only from chronic allergies and sinusitis (Chapter 426) but also from the nasal sprays and drops that they use to treat these conditions. The most common causes of loss of the sense of taste are viral infections and drug ingestion, particularly antirheumatic and antiproliferative drugs (see Table 427-1). Many of the systemic disorders listed in Table 427-1 probably produce their effect by decreasing the rate of turnover of sensory receptors on the tongue and olfactory epithelia. Disturbances of smell and taste in malnourished patients may be due to specific deficiencies in vitamins and minerals, such as zinc. Viral illnesses, such as influenza (Chapter 364), viral hepatitis (Chapter 148), and allergic rhinitis (Chapter 251), are the most common causes of loss of both taste and smell.[4] Multifocal neurologic disorders such as multiple sclerosis (Chapter 411) and traumatic head injuries (Chapter 399) can affect the central olfactory and gustatory pathways at multiple levels; as a result, abnormalities in taste and smell are common in such patients. Loss of smell is increasingly being recognized in the early stages of many neurodegenerative disorders, including Parkinson's disease, Alzheimer's dementia, motor neuron disease, and Huntington's disease.[5] An irritative lesion from a neoplastic, inflammatory, or demyelinating process may lead to a persistent disturbance rather than to a loss of taste.

DIAGNOSIS

Olfaction can be tested grossly at the bedside with a few easily recognized odors, such as coffee, chocolate, and the roselike aroma of the compound phenylethyl alcohol. Nasal irritants should be avoided. Each nostril is tested separately to determine whether the problem is unilateral or bilateral. Gustatory sensation is typically tested with weak solutions of sugar, salt, and acetic acid or vinegar. The patient must keep the tongue protruded and respond to questions by nodding the head or by pointing to names of the tastes written on cards. The anterior two thirds and posterior third of the tongue should be tested separately.

TREATMENT

Rx

Treatment of olfactory dysfunction secondary to nasal disease is aimed at opening the air passageways while preserving the olfactory epithelium (Chapter 426). Intranasal steroids for rhinosinusitis (Chapter 426), antibiotics as needed for sinusitis, and therapies for seasonal allergies (Chapter 251) are useful in selected cases. Drugs known to affect taste or smell (see Table 427-1) should be discontinued for a trial. Vitamin and mineral therapies are of unproven benefit.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bachmanov AA, Bosak NP, Lin C, et al. Genetics of taste receptors. *Curr Pharm Des*. 2014;20:2669-2683.
2. Schubert CR, Cruickshanks KJ, Klein BE, et al. Olfactory impairment in older adults: five-year incidence and risk factors. *Laryngoscope*. 2011;121:873-878.
3. Doty RL, Kamath V. The influences of age on olfaction. *Front Psychol*. 2014;5:20.
4. Henkin RI, Levy LM, Fordyce A. Taste and smell function in chronic disease: a review of clinical and biochemical evaluations of taste and smell dysfunction in over 5000 patients at The Taste and Smell Clinic in Washington, DC. *Am J Otolaryngol*. 2013;34:477-489.
5. Doty RL. Olfaction in Parkinson's disease and related disorders. *Neurobiol Dis*. 2012;46:527-552.

REVIEW QUESTIONS

1. A 53-year-old woman being treated for metastatic breast cancer complains of loss of taste. Which of the following is least likely to explain the taste loss?

A. Radiation therapy
B. Poor nutrition
C. Chemotherapy
D. Age
E. Hormonal changes

Answer: D Although age can lead to minor deficits in taste because of decreased turnover in taste bud cells, age would not be likely to explain taste loss in this relatively young woman. A, B, C, and E can cause severe loss of taste, particularly in this clinical circumstance.

2. A 39-year-old woman presents with subacute onset of loss of smell. What is the most likely cause?

A. Head trauma
B. Viral infection
C. Vitamin deficiency
D. Multiple sclerosis
E. Sinusitis

Answer: B Viral infections are the most common cause of loss of smell. The other conditions also can cause loss of smell but are less likely.

3. A 42-year-old man presents with subacute onset of hemifacial paralysis and persistent taste of sweetness. What clinical feature is not consistent with Bell's palsy (VII cranial nerve palsy)?

A. Hemifacial paralysis
B. Loss of taste in the anterior two thirds of the tongue
C. Persistent taste of sweetness
D. Intact taste in the posterior third of the tongue
E. Drooling or dry mouth with difficulty in eating

Answer: C A major branch of the mastoid segment of cranial nerve VII is the chorda tympani, which contains afferent taste fibers from the anterior two thirds of the tongue and is the anatomic basis for the loss of taste in the anterior two thirds of the tongue in idiopathic cranial nerve VII palsy. In contrast with loss of taste, a persistent sense of sweetness is a worrisome symptom for an irritative lesion, such as a neoplastic, inflammatory, or demyelinating lesion involving cranial nerve VII.

4. What is the molecular basis that allows us to taste carbonation in sodas, sparkling wines, and other carbonated drinks?

A. G protein–coupled glutamate receptor for umami
B. Carbonic anhydrase IV expressed on the surface of sour-sensing cells
C. TAS1R heterodimers for sweetness
D. TAS2R proteins for bitter taste
E. Epithelial sodium channel receptor for salty taste

Answer: B Mountaineers who take acetazolamide (a carbonic anhydrase inhibitor) to prevent altitude sickness have long reported that carbonated beverages lacked pleasurable taste or tingle sensation. Patients treated with acetazolamide and other medications with carbonic anhydrase–inhibiting activities also experience taste disturbances.

428 HEARING AND EQUILIBRIUM

ROBERT W. BALOH AND JOANNA C. JEN

DEFINITION

The neural pathways subserving hearing and those most important for equilibrium and spatial orientation are anatomically proximate in much of their course from their end organs in the inner ear to their termination in the superior portion of the temporal lobe. Because of the close anatomic linkage, disorders that affect hearing often affect equilibrium, and vice versa. For this reason, they are considered together here.

PATHOBIOLOGY

Despite their anatomic propinquity, however, substantial pathophysiologic differences make clinical examination of the two systems different. The auditory system is physiologically relatively isolated, so that its function and dysfunction can be tested independently of other neural systems. The vestibular system, in contrast, has many close physiologic links with other neural systems (particularly the visual-oculomotor, somatosensory, and autonomic systems) and can be difficult to test in isolation of these other systems.

DIAGNOSIS

Abnormalities of the auditory system lead to only a few well-defined and unique symptoms (i.e., hearing loss or tinnitus). Abnormalities of the vestibular system can cause symptoms that mimic disorders of other neural structures. Such symptoms include dizziness, visual distortion (oscillopsia), imbalance, nausea, vomiting, and even syncope.

DISORDERS OF THE AUDITORY SYSTEM

DEFINITION

Anatomy and Physiology of Hearing

In normal hearing, sound waves are transmitted from the tympanic membrane through the three ossicles of the air-filled middle ear (air conduction) to the oval window and the basilar membrane of the fluid-sealed cochlea. The ossicles increase the gain from the tympanum to oval window about 18-fold, compensating for the loss that sound waves moving from air to fluid would otherwise suffer. In the absence of this system, sound may reach the cochlea by vibration of the temporal bone (bone conduction) but with much less efficiency (approximately 60-dB loss). Hair cells, tonotopically organized along the cochlear basilar membrane, detect the vibratory movement of that membrane and transduce vibration into nerve impulses. The nerve impulses are relayed by nerve cells that synapse at the base of hair cells and have their

bodies in the spiral ganglion to the cochlear nucleus of the ipsilateral pontine tegmentum. The spiral cochlea mechanically analyzes the frequency content of sound. For high-frequency tones, only sensory cells in the basilar region are activated, whereas for low-frequency tones, all or nearly all sensory cells are activated. Therefore, with lesions of the cochlea and its afferent nerve, the hearing levels for different frequencies are usually unequal, typically resulting in better hearing sensitivity for low-frequency than for high-frequency tones. Within the brain stem, auditory signals ascend from the ventral and dorsal cochlear nuclei to reach the superior olivary nuclei of both sides. Thus, nervous system lesions central to the cochlear nucleus do not cause monaural hearing loss, and conversely, unilateral central lesions do not cause deafness. From these structures, the pathway projects by way of the lateral lemnisci to the inferior colliculi. Each inferior colliculus transmits to the other and to its ipsilateral medial geniculate body, which in turn sends the final projection to the transverse auditory gyrus lying in the superior portion of the ipsilateral temporal lobe.

The normal ear can detect sound frequencies ranging between 20 and 20,000 Hz; the upper range drops off fairly rapidly with advancing age. The ear is most sensitive between 500 and 4000 Hz, which roughly corresponds to the frequency range most important for understanding speech. The hearing level in this range has several practical implications in terms of the degree of handicap and the potential for useful correction with amplification. A 30- to 40-dB hearing level in the speech range would impair normal conversation, whereas an 80-dB hearing level would make everyday auditory communication almost impossible (the social definition of deafness).

EPIDEMIOLOGY

About 5% of the world population suffers from disabling hearing loss (defined by the World Health Organization as greater than 40 dB in the better hearing ear in adults and greater than 30 dB in the better hearing ear in children). The prevalence of disabling hearing loss is twice as high in poorer countries compared with richer countries. The prevalence increases with every age decade,[1] and it is higher in men than in women across all age decades. Hearing loss is independently associated with accelerated cognitive decline and incident cognitive impairment in community-dwelling older adults.

PATHOBIOLOGY

Localization of Lesions within the Auditory Pathways

Conductive hearing loss results from lesions involving the external or middle ear. It is typically characterized by an approximately equal loss of hearing at all frequencies and by well-preserved speech discrimination once the threshold for hearing is exceeded. Patients with conductive hearing loss can hear speech in a noisy background better than in a quiet background because they can understand loud speech as well as anyone.

Sensorineural hearing loss results from lesions of the cochlea or auditory division of the eighth cranial nerve, or both. With sensorineural hearing loss, the hearing levels for different frequencies are usually unequal, typically resulting in better hearing for low- than for high-frequency tones. Patients with sensorineural hearing loss often have difficulty in hearing speech that is mixed with background noise and may be annoyed by loud speech. Three important manifestations of sensorineural lesions are diplacusis, recruitment, and tone decay. Diplacusis and recruitment are common with cochlear lesions; tone decay usually accompanies eighth nerve involvement.

Central hearing disorders result from lesions of the central auditory pathways. As a rule, patients with central lesions do not have impaired hearing for pure tones, and they can understand speech as long as it is clearly spoken in a quiet environment. If the listener's task is made more difficult with the introduction of background noise or competing messages, performance deteriorates more markedly in patients with central lesions than in normal subjects.

DIAGNOSIS

Evaluation

Bedside Test

A quick test for hearing loss in the speech range is to observe the response to spoken commands at different intensities (whisper, conversation, shouting). Tuning fork tests permit a rough assessment of the hearing level for pure tones of known frequency. The clinician can use his or her own hearing level as a reference standard. In the Rinne test, nerve conduction is compared with bone conduction by holding a tuning fork (preferably 512 Hz) against the mastoid process until the sound can no longer be heard. It is then placed 1 inch from the ear and, in normal subjects, can be heard about twice as long by air as by bone. If bone conduction is better than air conduction, the hearing loss is conductive, but care must be taken to ensure that the bone conduction is not heard in the normal ear. In the Weber test, the tuning fork is placed on the patient's forehead or upper teeth. Normally, this sound is referred to the center of the head. If it is referred to the side of unilateral hearing loss, the hearing loss is conductive; if it is referred away from the side of unilateral hearing loss, the loss is sensorineural.

Audiometry

Pure tone testing is the cornerstone of most auditory examinations. Pure tones at selected frequencies are presented through either earphones (air conduction) or a vibrator pressed against the mastoid portion of the temporal bone (bone conduction), and the minimal level that the subject can hear (threshold) is determined for each frequency. Two speech tests are routinely used. The *speech reception threshold* is the intensity at which the patient can correctly repeat 50% of the words presented. The speech reception threshold is a test of hearing sensitivity for speech and should reflect the hearing level for pure tones in the speech range. The *speech discrimination test* is a measure of the patient's ability to understand speech when it is presented at a level that is easily heard. In patients with eighth nerve lesions, speech discrimination scores can be severely reduced, even when pure tone thresholds are normal or nearly normal; by comparison, in patients with cochlear lesions, discrimination tends to be proportional to the magnitude of hearing loss.

Brain stem auditory evoked responses can be recorded from scalp electrodes at 0 to 10 msec (early), 10 to 50 msec (middle), and 50 to 500 msec (late) following a click (a high-frequency stimulus). The early potentials reflect electrical activity at the cochlea, eighth cranial nerve, and brain stem; the later potentials reflect cortical activity. Computer averaging of the responses to 1000 to 2000 clicks separates the evoked potential from background noise. Early evoked responses may be used to estimate the magnitude of hearing loss and to differentiate among cochlea, eighth nerve, and brain stem lesions.

Differential Diagnosis

Conductive Hearing Loss

The history, examination, and audiometry usually provide the key differential features for identifying common causes of hearing loss (Fig. 428-1). The most common cause of conductive hearing loss is *impacted cerumen* in the external canal. This benign condition is usually first noticed after bathing or swimming when a droplet of water closes the remaining tiny passageway. The most common serious cause of conductive hearing loss is inflammation of the middle ear, *otitis media*, either infective (suppurative; see Fig. 426-8) or noninfective (serous). Fluid accumulates in the middle ear, impairing the conduction of airborne sound to the cochlea. Because the air cavity of the middle ear is in direct connection with the mastoid air cells, infection can spread through the mastoid bone and, occasionally, into the intracranial cavity. Chronic otitis media with perforation of the tympanic membrane can result in an invasion of the middle ear and other pneumatized areas of the temporal bone by keratinizing squamous epithelium (*cholesteatoma*). Cholesteatomas can produce erosion of the ossicles and bony labyrinth, resulting in a mixed conductive and sensorineural hearing loss. Barotrauma to the middle ear arises with otalgia and hearing loss and can be associated with serous effusion or hematotympanum (see Fig. 426-9). *Otosclerosis* commonly produces progressive conductive hearing loss by immobilizing the stapes with new bone growth in front of and below the oval window. The hearing loss is typically conductive, although in some persons the cochlea may be invaded by foci of otosclerotic bone, producing an additional sensorineural hearing loss. Otosclerosis usually stabilizes when the hearing level reaches 50 to 60 dB and rarely progresses to deafness. Other common causes of conductive hearing loss include trauma, congenital malformations of the external and middle ear, and glomus body tumors.

Sensorineural Hearing Loss

Hereditary Deafness

Genetically determined deafness, usually from hair cell aplasia or deterioration, may be present at birth or may develop in adulthood. The diagnosis of *hereditary deafness* rests on the finding of a positive family history. Mutations in connexin 26, a key component of gap junctions in the inner ear, account for most cases of recessively inherited deafness. *Intrauterine factors* resulting in congenital hearing loss include infection (especially rubella); toxic, metabolic, and endocrine disorders; and anoxia associated with Rh incompatibility and difficult deliveries.

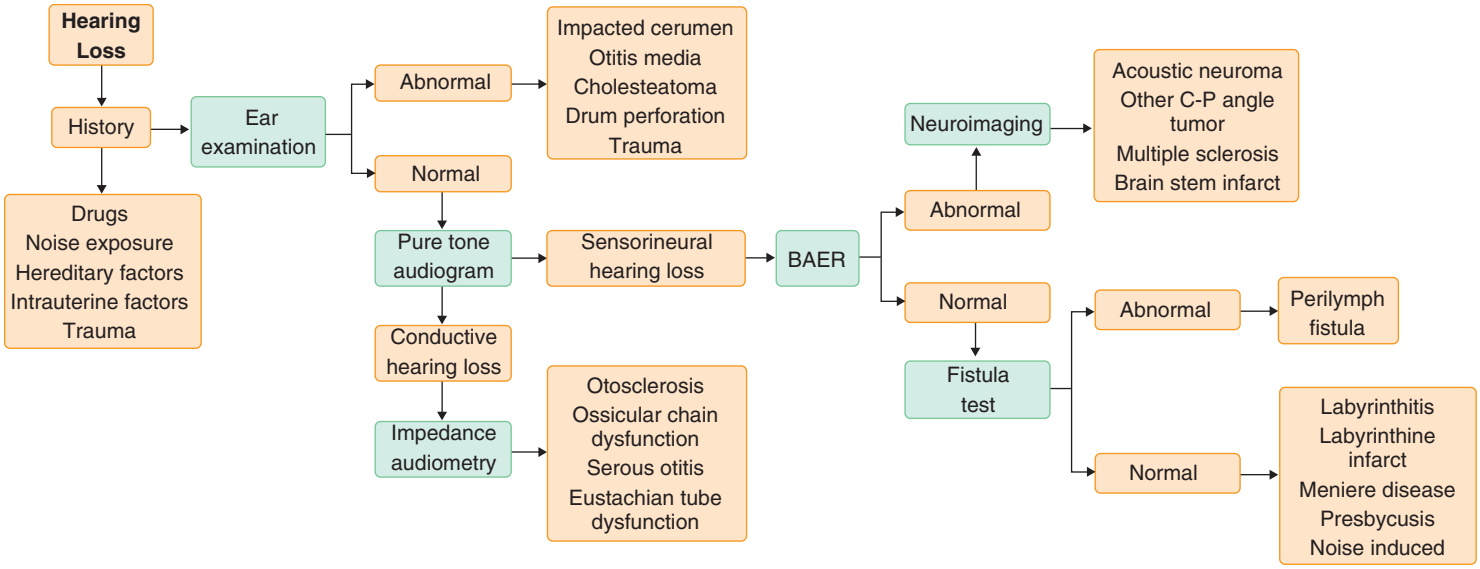

FIGURE 428-1. Evaluation of hearing loss. BAER = brain stem auditory evoked response; C-P = cerebellopontine.

Cochlear Damage

Acute unilateral deafness usually has a cochlear basis. *Bacterial or viral infections* of the labyrinth, *head trauma* with fracture or hemorrhage into the cochlea, or *vascular occlusion* of a terminal branch of the anterior inferior cerebellar artery can extensively damage the cochlea and the vestibular labyrinth. An isolated sudden unilateral sensorineural hearing loss is presumed to reflect a viral infection of the cochlea and auditory nerve terminals.[2] High-dose steroids followed by a rapid taper are recommended (see Treatment).

Sudden unilateral hearing loss often associated with vertigo and tinnitus can result from a *perilymphatic fistula.* Such fistulas may be congenital or may follow stapes surgery or head trauma.

Drugs

Drugs cause acute and subacute bilateral hearing impairment. Salicylates, furosemide, and ethacrynic acid have the potential to produce transient deafness when they are taken in high doses. More toxic to the cochlea are aminoglycoside antibiotics (gentamicin, tobramycin, amikacin, kanamycin, streptomycin, and neomycin). These agents can destroy cochlear hair cells in direct relation to their serum concentrations. Some antineoplastic chemotherapeutic agents, particularly cisplatin, cause severe ototoxicity.

Meniere Disease

Subacute relapsing cochlear deafness occurs with *Meniere disease*, a condition associated with fluctuating hearing loss and tinnitus, recurrent episodes of abrupt and often severe vertigo, and a sensation of fullness or pressure in the ear. Recurrent endolymphatic hypertension (hydrops) is believed to cause the episodes. On pathologic examination, the endolymphatic sac is dilated, and the hair cells become atrophic. The resulting deafness is subtle and reversible in the early stages but subsequently becomes permanent and is characterized by diplacusis and loudness recruitment. The disorder is usually unilateral, but in about 20 to 40% of patients, bilateral involvement occurs.

Presbycusis

The gradual, progressive, bilateral hearing loss commonly associated with advancing age is called presbycusis.[3] Presbycusis is not a distinct disease entity but rather represents multiple effects of aging on the auditory system. It may include conductive and central dysfunction, although the most consistent effect of aging is on the sensory cells and neurons of the cochlea. The typical audiogram of presbycusis is a symmetrical high-frequency hearing loss gradually sloping downward with increasing frequency. The most consistent pathologic finding associated with presbycusis is degeneration of sensory cells and nerve fibers at the base of the cochlea.

Noise

The recurrent trauma of *noise-induced hearing loss* affects approximately the same region at the base of the cochlea and is also common, particularly among those with exposure to loud explosive or industrial noises. Loud, blaring, modern music has become a recent offender. The loss almost always begins at 4000 Hz and does not affect speech discrimination until late in the disease process. With only brief exposure to loud noise (hours to days), there may be only a temporary threshold shift, but with continued exposure, permanent injury begins. The duration and intensity of exposure determine the degree of permanent injury.

Acoustic Neuroma

Progressive unilateral hearing loss, which arises insidiously, initially in the high frequencies, and worsens by almost imperceptible degrees, is characteristic of benign neoplasms of the cerebellopontine angle, most commonly *acoustic neuromas.* In about 10% of cases, the hearing loss can be acute, apparently due to either hemorrhage into the tumor or compression of the labyrinthine vasculature. Magnetic resonance imaging (MRI) with contrast enhancement reliably identifies small acoustic neuromas.

Central Hearing Loss

Central hearing loss is unilateral only if it results from damage to the pontine cochlear nuclei on one side of the brain stem from conditions such as *ischemic infarction* of the lateral brain stem (e.g., occlusion of the anterior inferior cerebellar artery [Chapter 407]), a plaque of *multiple sclerosis* (Chapter 411), or, rarely, invasion or compression of the lateral pons by a *neoplasm* or *hematoma* (Chapters 189 and 399). Bilateral *degeneration* of the cochlear nuclei accompanies some of the rare recessive inherited disorders of childhood. As noted, clinically important unilateral hearing loss never results from neurologic disease arising rostral to the cochlear nucleus. Although bilateral hearing loss could, in theory, result from bilateral destruction of central hearing pathways, in practice this is rare because involvement of neighboring structures in the brain stem or hemisphere would usually produce overwhelming neurologic disability.

TREATMENT

If an underlying disorder has not yet destroyed the auditory system and can be ameliorated medically or surgically, hearing may be improved or preserved. Most patients with otosclerosis respond to stapedectomy. Closure of a perilymph fistula may improve hearing. Antibiotic and decongestive treatment of otitis media (Chapter 426) should prevent permanent hearing loss.

A brief course of high-dose steroids is commonly used for patients with idiopathic sudden unilateral sensorineural deafness, but the evidence to support this approach is limited.■ Intratympanic corticosteroid treatment (four doses of 40 mg/mL of methylprednisolone during 2 weeks) is not inferior to oral treatment (60 mg/day of oral prednisone followed by a 5-day taper) for idiopathic sudden sensorineural hearing loss,■ and combination oral and intratympanic therapy may be better than either alone.■ A low-salt diet and diuretics are effective in selected cases of Meniere disease. Folic acid

supplementation appears to reduce the rate of hearing loss in the elderly. Hearing aids amplify sound, usually with the goal of making speech intelligible. Patients with conductive hearing loss require simple amplification, but those with sensorineural hearing loss often need frequency-selective amplification to make hearing aids useful. Cochlear implants can help patients with profound hearing loss if they have some intact auditory nerve fibers.[4] Intense postoperative speech recognition training is required.

Tinnitus

DIAGNOSIS

The evaluation of common causes of tinnitus (Fig. 428-2) begins with a careful history to identify common offending drugs.[5]

Objective Tinnitus

With objective tinnitus, the patient hears a sound arising external to the auditory system, a sound that can usually be heard by the examiner with a stethoscope. Objective tinnitus usually has benign causes, such as noise from temporomandibular joints, opening of eustachian tubes, or repetitive muscle contractions. Sometimes, in a quiet room, the patient can hear the pulsatile flow in the carotid artery or a continuous hum of normal venous outflow through the jugular vein. The latter can be obliterated by compression of the jugular vein or extreme lateral rotation of the neck. Pathologic objective tinnitus occurs when patients hear turbulent flow in vascular anomalies or tumors (e.g., glomus jugulare tumor). Objective tinnitus may also be an early sign of increased intracranial pressure. Such tinnitus, which probably arises from turbulent flow through compressed venous structures at the base of the brain, is usually overshadowed by other neurologic abnormalities.

Subjective Tinnitus

Subjective tinnitus can arise from sites anywhere in the auditory system. The sounds most frequently reported are metallic ringing, buzzing, blowing, roaring, or, less often, bizarre clanging, popping, or nonrhythmic beating. Tinnitus heard as a faint, moderately high pitched, metallic ring can be observed by almost anyone who concentrates attention on auditory events in a quiet room. Sustained louder tinnitus accompanied by audiometric evidence of deafness occurs in association with both conductive and sensorineural hearing loss. Tinnitus observed with otosclerosis tends to have a roaring or hissing quality, and that associated with Meniere disease often produces sounds that vary widely in intensity with time and quality, sometimes including roaring or clanging. Tinnitus with auditory nerve lesions tends to be higher pitched and ringing in quality. Audiometric and brain stem evoked response testing can help distinguish between lesions involving the conducting apparatus, the cochlea, and the auditory nerve. Tinnitus without observable deafness appears sporadically and for variable lengths of time in many persons without other evidence of an ongoing pathologic process.

TREATMENT Rx

Most patients with tinnitus can be helped by a careful evaluation to exclude serious underlying conditions and by subsequent reassurance when appropriate. Often, exacerbating factors such as chronic anxiety and depression can be treated. In patients with hearing loss and tinnitus, a hearing aid may improve tinnitus because the amplification of ambient sound may effectively mask the tinnitus. This mechanism probably explains the frequent observation that removal of cerumen from the external auditory canal to improve ambient hearing also improves tinnitus. Also, when cerumen is attached to the tympanic membrane, tinnitus may result from local mechanical effects on the conductive system. For patients who find their tinnitus most obtrusive when trying to sleep, recorded masking sounds (e.g., white noise, rainfall, mountain stream) can be helpful. A careful drug history should be taken (see Fig. 428-2), and a drug-free trial period should be considered when possible.

No medications are approved for the treatment of tinnitus in the United States or Europe. Benzodiazepines (e.g., diazepam, 2 to 5 mg every 8 hours) or tricyclic amines (e.g., amitriptyline, 25 to 75 mg at bedtime) may provide temporary symptomatic relief of tinnitus, but cognitive-behavioral therapy is a more effective long-term approach that can significantly decrease tinnitus and improve health-related quality of life. In patients with concomitant profound bilateral sensorineural hearing loss, cochlear implants can improve hearing and often decrease tinnitus.

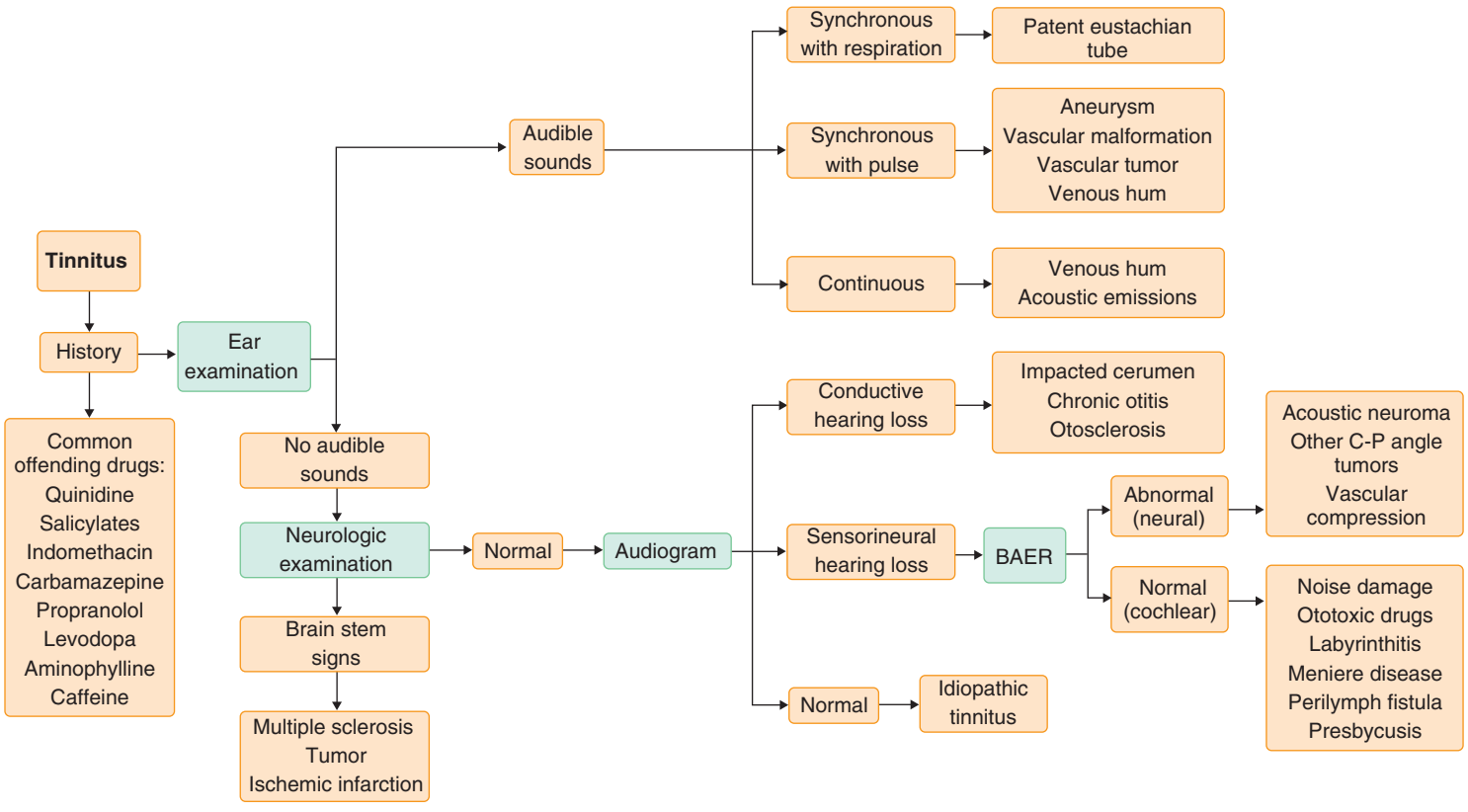

FIGURE 428-2. Evaluation of tinnitus. BAER = brain stem auditory evoked response; C-P = cerebellopontine.

EQUILIBRIUM–VESTIBULAR SYSTEM

PATHOBIOLOGY

Anatomy and Physiology of the Vestibular System

The paired vestibular end organs lie within the temporal bones next to the cochlea. Each organ consists of three semicircular canals that detect angular acceleration and two otolith structures, the utricle and saccule, that detect linear acceleration (including gravitational). Like the cochlea, these organs possess hair cells that act as force transducers, converting the forces associated with head acceleration into afferent nerve impulses. The hair cells of the three semicircular canals, each of which is oriented at right angles to the others, are located in the crista, where their cilia are embedded in a gelatinous mass called the *cupula*. Movement of the head causes the endolymph to flow either toward or away from the cupula, bending the cilia and, depending on the direction of endolymphatic movements, either exciting or inhibiting the afferent nerves at the base of the hair cells. Because the afferent nerves are tonically active, the baseline activity can be increased or decreased, depending on the direction of hair cell bending. Furthermore, the two sets of semicircular canals are approximately mirror images of each other, so that rotational movement of the head that excites one canal inhibits the analogous canal on the opposite side. The hair cells of the utricle and saccule are located in an area called the *macule*. The macule of the utricle lies approximately in the plane of the horizontal canal, and the macule of the saccule is approximately in the plane of the anterior canal. The hair cell cilia are embedded in a membrane that contains calcium carbonate crystals or otoliths; the density of otoliths is considerably greater than that of the endolymph. Linear accelerations of the head combine with the linear acceleration of gravity to distort the otolith membrane, thereby bending the cilia of the hair cells and modulating the activity of the afferent nerve terminals at the base of the hair cells.

The afferent vestibular nerves have their cell bodies in Scarpa's ganglion. The nerve fibers travel in the vestibular portion of the eighth cranial nerve contiguous to the acoustic portion. Fibers from different receptor organs terminate in different vestibular nuclei at the pontomedullary junction. There are also direct connections with many portions of the cerebellum, the greatest representation being in the flocculonodular lobe, the so-called vestibular cerebellum. Efferent fibers from the brain stem travel through the vestibular nerve to reach hair cells of the semicircular canals and macules, where they modulate afferent activities. From the vestibular nuclei, second-order neurons make important connections to the vestibular nuclei of the other side, to the cerebellum, to motor neurons of the spinal cord, to autonomic nuclei in the brain stem, and, most important for the examining clinician, to the nuclei of the oculomotor system. Fibers from the vestibular nuclei also ascend through the brain stem and thalamus to reach the cerebral cortex bilaterally.

DIAGNOSIS

Evaluation

History

Most vestibular problems presented to the physician are episodic, and often there are neither symptoms nor signs when the physician examines the patient. The history, therefore, can become paramount for identifying vestibular dysfunction. The history should attempt to distinguish vertigo (the illusion of movement in space) from lightheadedness (presyncope), ataxia (disequilibrium of the body without true movement in space), and psychogenic symptoms (the feeling of dissociation or, sometimes, disequilibrium).

About 12% of patients with vertigo have a central cause, and about 88% have a problem with the peripheral vestibular apparatus.[6] In general, peripheral vertigo is more severe, is more likely to be associated with hearing loss and tinnitus, and often leads to nausea and vomiting. Nystagmus associated with peripheral vertigo is usually inhibited by visual fixation. Central vertigo is generally less severe than peripheral vertigo and is often associated with other signs of central nervous system disease. The nystagmus of central vertigo is not inhibited by visual fixation and frequently is prominent when vertigo is mild or absent.

Common Causes of Vertigo

Physiologic Vertigo

Physiologic vertigo includes common disorders that occur in healthy people, such as *motion sickness, space sickness,* and *height vertigo* (Fig. 428-3). In these conditions, vertigo (defined as an illusion of movement) is minimal while autonomic symptoms predominate. With height vertigo, patients may experience acute anxiety and panic reaction. Individuals with motion sickness and space sickness typically develop perspiration, nausea, vomiting, increased salivation, yawning, and generalized malaise. Gastric motility is reduced and digestion impaired. Even the sight or smell of food is distressing. Hyperventilation is a common sign, and the resulting hypocapnia leads to changes in blood volume, with pooling in the lower parts of the body predisposing to postural hypotension and syncope. An unusual variant of motion-induced dizziness occurs when the subject returns to stationary conditions after prolonged exposure to motion (*mal de débarquement syndrome*). Typically, affected patients report that they feel the persistent rocking sensation of a boat long after returning to solid ground. Rarely, the syndrome can last for months to years after exposure to motion and can even be incapacitating. The cause is unknown.

Physiologic vertigo can often be suppressed by supplying sensory cues that help to match the signals originating from different sensory systems. Thus, motion sickness, which is caused by a mismatch of visual and vestibular signals, is exacerbated by sitting in a closed space or reading (giving the visual system the miscue that the environment is stationary). It may be improved by looking out at the horizon. Height vertigo, caused by a mismatch between

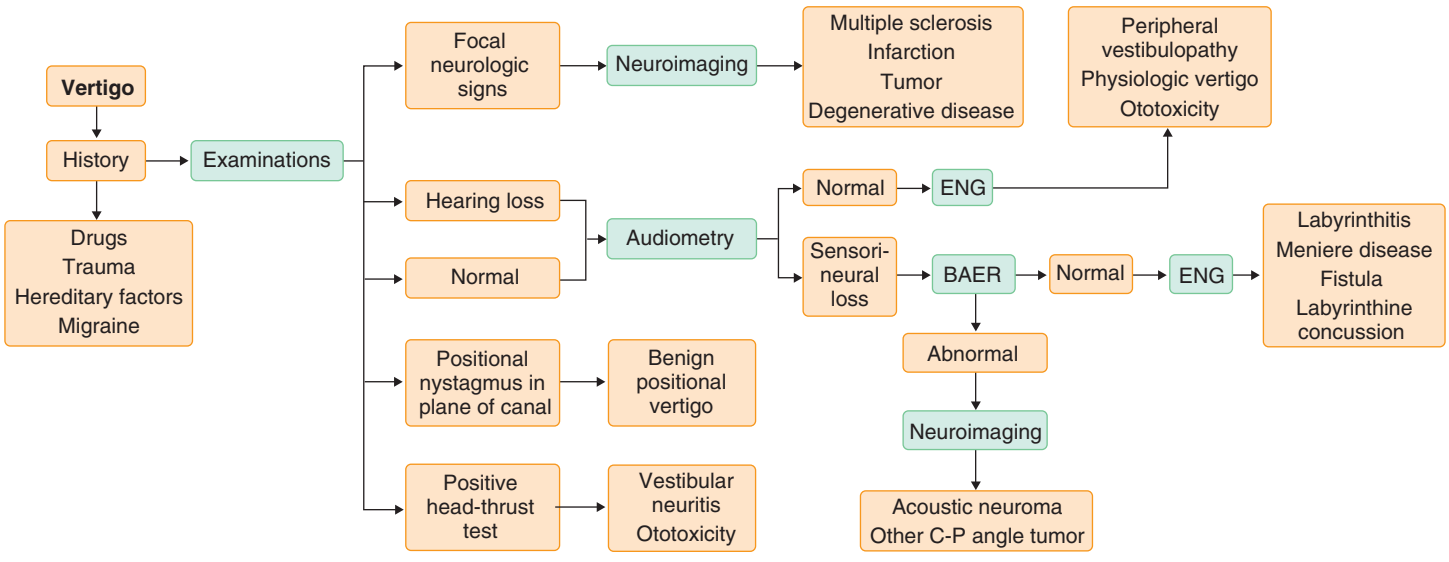

FIGURE 428-3. Evaluation of vertigo. BAER = brain stem auditory evoked response; C-P = cerebellopontine; ENG = electronystagmography.

sensation of normal body sway and lack of its visual detection, can often be relieved either by sitting or by visually fixating a nearby stationary object.

Benign Paroxysmal Positional Vertigo (Canalithiasis)

Benign paroxysmal positional vertigo is by far the most common cause of vertigo.[7] Patients with this condition develop brief episodes of vertigo (less than 1 minute) with position change, typically when turning over in bed, getting in and out of bed, bending over and straightening up, or extending the neck to look up (so-called top-shelf vertigo). Benign paroxysmal positional vertigo results when otolith debris inadvertently enters one of the semicircular canals. It can occur after head trauma or inner ear infection but most commonly occurs spontaneously in older people. The diagnosis rests on finding characteristic positional nystagmus in the plane of the affected canal (see later). It is important to recognize this syndrome because, in most patients, it can be cured by simple bedside maneuvers (Fig. 428-4). If the history or findings are atypical, the condition must be distinguished from other causes of positional vertigo that may occur with tumors or infarcts of the posterior fossa.

Acute Peripheral Vestibulopathy (Vestibular Neuritis)

One of the most common clinical neurologic syndromes at any age is the acute onset of vertigo, nausea, and vomiting lasting for several days and not associated with auditory or neurologic symptoms.[8] A viral origin is suspected, but attempts to isolate an agent have been unsuccessful, except for occasional findings of a herpes zoster infection. Pathologic studies showing atrophy of one or more vestibular nerve trunks, with or without atrophy of their associated sense organs, are evidence of a vestibular nerve site and, probably, viral cause for many patients with this syndrome. Many patients report an upper respiratory tract illness 1 to 2 weeks before the onset of vertigo. This syndrome occasionally occurs in epidemics (epidemic vertigo), may affect several members of the same family, and more often erupts in the spring and early summer. Most affected patients gradually improve during 1 to 2 weeks, but residual dizziness and imbalance can persist for months.

Meniere Disease

Meniere disease (see earlier) accounts for about 10% of all patients with vertigo.[9] The diagnosis is based on documenting episodic severe attacks accompanied by fluctuating hearing levels on audiometric testing beginning in the low frequencies.

Migraine

Vertigo is a common symptom with migraine (Chapter 398). It can occur with headaches or in separate isolated episodes, and it can predate the onset of headache. So-called benign paroxysmal vertigo of childhood is often the

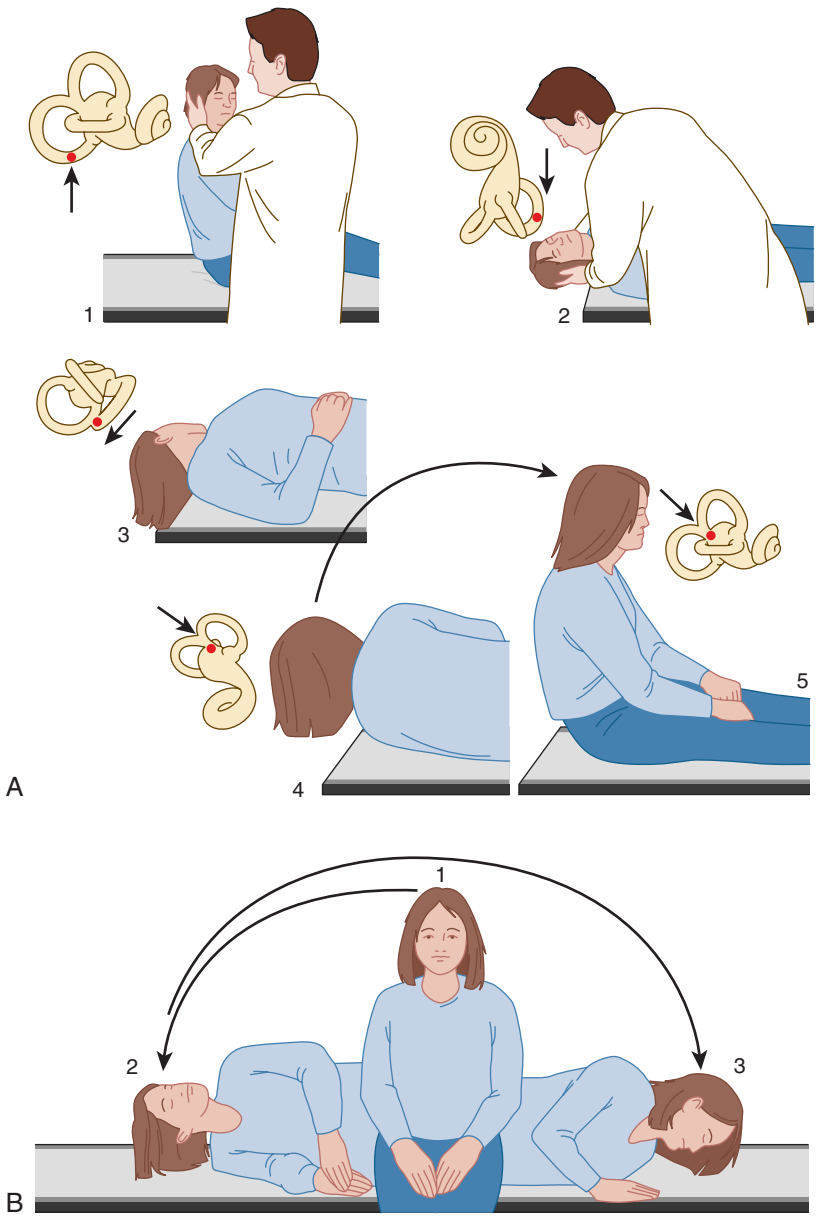

FIGURE 428-4. Modified Epley's (**A**) and Semont's (**B**) maneuvers for benign positional vertigo affecting the right posterior semicircular canal. The procedure is reversed to treat the left posterior semicircular canal. The entire sequence should be repeated until no nystagmus is elicited. (From Fife TD, Iverson DJ, Lempert T, et al. Practice parameter: therapies for benign paroxysmal positional vertigo [an evidence-based review]: report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology.* 2008;70:2067-2074.)

first symptom of migraine. The mechanism of vertigo with migraine is not clear, but both peripheral and central types of nystagmus can occur with attacks. A few develop typical features of Meniere disease.

Post-traumatic Vertigo

Vertigo, hearing loss, and tinnitus often follow a blow to the head (Chapter 399) that does not result in temporal bone fracture, termed *labyrinthine concussion.* Although they are protected by a bone capsule, the delicate labyrinthine membranes are susceptible to blunt trauma. Blows to the occipital or mastoid region are particularly likely to produce labyrinthine damage. *Transverse fractures* of the temporal bone typically pass through the vestibule of the inner ear, tearing the membranous labyrinth and lacerating the vestibular and cochlear nerves. Complete loss of vestibular and cochlear function is the usual sequela, and the facial nerve is interrupted in approximately 50% of cases. Examination of the ear often reveals hemotympanum (see Fig. 426-9), but bleeding from the ear seldom occurs because the tympanic membrane usually remains intact. As noted earlier, *benign paroxysmal positional vertigo* is also a common sequela of head trauma. *Fistulas* of the oval and round windows can result from impact noise, deep-water diving, severe physical exertion, or blunt head injury without skull fracture. The mechanism of the rupture is a sudden negative or positive pressure change in the middle ear or a sudden increase in cerebrospinal fluid pressure transmitted to the inner ear through the cochlear aqueduct and internal auditory canal. Clinically, the rupture leads to the sudden onset of vertigo or hearing loss, or both. Surgical exploration of the middle ear is warranted when there is a clear relationship between the onset of vertigo or hearing loss, or both, and the onset of severe exertion, barometric change, head injury, or impact noise.

Postconcussion Syndrome

The so-called postconcussion syndrome refers to a vague dizziness (rarely vertigo) associated with anxiety, difficulty in concentrating, headache, and photophobia induced by a head injury resulting in concussion (Chapter 399). On occasion, similar but less pronounced symptoms are associated with mild head injury judged to be trivial at the time. The cause is unknown, but animal studies indicate that small multifocal brain lesions (petechiae) commonly occur after concussive brain injury.

Other Peripheral Causes of Vertigo

Vertigo can be associated with *chronic bacterial otomastoiditis*, either from direct invasion of the inner ear by the bacteria or by erosion of the labyrinth by a cholesteatoma. Radiographic studies of the temporal bone readily identify these disorders. *Autoimmune inner ear disease* typically arises with episodic vertigo and fluctuating hearing levels similar to Meniere disease, but it is more fulminant with early bilateral involvement. It can occur in isolation or with other systemic features of autoimmune disease. About two thirds of patients have antibodies directed against heat shock protein 70. The aminoglycosides streptomycin and gentamicin are remarkably selective for vestibular ototoxicity. The patient may suffer acute vertigo if the toxic effect is asymmetrical. More often, there is a progressive symmetrical loss of vestibular function leading to imbalance but not vertigo. Unfortunately, many patients being treated with ototoxic drugs are initially bedridden and unaware of the vestibular impairment until they recover from their acute illness and try to walk. They then discover that they are unsteady on their feet and that the environment tends to jiggle in front of their eyes (*oscillopsia*). The diagnosis can be made at the bedside with a head-thrust test (bilateral corrective saccades; see later). Caloric and rotational testing can document the degree of vestibular loss. The best treatment is prevention. If the drug is discontinued early during the course of symptoms, the disorder may stabilize or improve.

Vascular Insufficiency

Vertebrobasilar insufficiency is a common cause of vertigo in older people. Whether the vertigo originates from ischemia of the labyrinth, brain stem, or both structures is not always clear because the blood supplies to the labyrinth, eighth cranial nerve, and vestibular nuclei originate from the same source, the basilar vertebral circulation (Chapter 406). Vertigo with *vertebrobasilar insufficiency* is abrupt in onset, usually lasting several minutes, and is frequently associated with nausea and vomiting. Associated symptoms resulting from ischemia in the remaining territory supplied by the posterior circulation include visual illusions and hallucinations, drop attacks and weakness, visceral sensations, visual field defects, diplopia, and headache. These symptoms occur in episodes either in combination with the vertigo or alone. Vertigo may be an isolated initial symptom of vertebrobasilar ischemia, but repeated episodes of vertigo without other symptoms should suggest another diagnosis. Vertebrobasilar insufficiency is usually caused by atherosclerosis of the subclavian, vertebral, and basilar arteries. On occasion, episodes of vertebrobasilar insufficiency are precipitated by postural hypotension, Stokes-Adams attacks, or mechanical compression from cervical spondylosis. MRI of the brain is usually normal because the vascular insufficiency is transient and function returns to normal between episodes. Magnetic resonance angiography can identify occlusive vascular disease most commonly involving the vertebral-basilar junction.

Vertigo is a common symptom with *infarction of the lateral brain stem* or *cerebellum* (Chapter 407), or both. The diagnosis is usually clear, based on the characteristic acute history and pattern of associated symptoms and neurologic findings. On occasion, cerebellar infarction or hemorrhage arises with severe vertigo, vomiting, and ataxia without associated brain stem symptoms and signs that might suggest the erroneous diagnosis of an acute peripheral vestibular disorder. The key differential is the finding of clear cerebellar signs (extremity and gait ataxia) and of direction-changing, gaze-evoked nystagmus. Such patients must be watched carefully for several days because they may develop progressive brain stem dysfunction due to compression by a swollen cerebellum.

Cerebellopontine Angle Tumors

Most tumors growing in the cerebellopontine angle (e.g., *acoustic neuroma, meningioma, epidermal cyst*) grow slowly, allowing the vestibular system to accommodate so that they produce only a vague sensation of disequilibrium rather than acute vertigo (Chapter 189). On occasion, however, episodic vertigo or positional vertigo heralds the presence of a cerebellopontine angle tumor. In virtually all patients, retrocochlear hearing loss is present, best identified by audiometric testing. MRI with contrast enhancement is the most sensitive diagnostic study for identifying a cerebellopontine angle tumor.

Other Central Causes of Vertigo

Acute vertigo may be the first symptom of *multiple sclerosis* (Chapter 411), although only a small percentage of young patients with acute vertigo eventually develop multiple sclerosis. Vertigo in multiple sclerosis is usually transient and often associated with other neurologic signs of brain stem disease, in particular, internuclear ophthalmoplegia or cerebellar dysfunction. Vertigo may also be a symptom of *parainfectious encephalomyelitis* or, rarely, *parainfectious cranial polyneuritis*. In this instance, the accompanying neurologic signs establish the diagnosis. The *Ramsay Hunt syndrome* (geniculate ganglion herpes) is characterized by vertigo and hearing loss associated with facial paralysis and, sometimes, pain in the ear. The typical lesions of herpes zoster (Chapter 375), which may follow the appearance of neurologic signs, are found in the external auditory canal and over the palate in some patients. Rarely is herpes zoster responsible for vertigo in the absence of the full-blown syndrome. *Granulomatous meningitis* (Chapter 412) or *leptomeningeal metastasis* and cerebral or systemic *vasculitis* (Chapter 270) may involve the eighth nerve, producing vertigo as an early symptom. In these disorders, cerebrospinal fluid analysis usually suggests the diagnosis (Chapter 396). Patients suffering from *temporal lobe epilepsy* (Chapter 403) occasionally experience vertigo as the aura. Vertigo in the absence of other neurologic signs or symptoms is never caused by epilepsy or other diseases of the cerebral hemispheres.

Bedside Tests

Hyperventilation

If the history is not clear, bedside provocative tests to mimic the symptom may assist in making a pathophysiologic diagnosis.[10] Hyperventilation, which lowers the arterial partial pressure of carbon dioxide ($Paco_2$) and decreases cerebral blood flow, causes a lightheaded sensation associated with syncope. Patients with compressive lesions of the vestibular nerve, such as with an acoustic neuroma or cholesteatoma, or with demyelination of the vestibular nerve root entry zone may develop vertigo and nystagmus after hyperventilation. Presumably, metabolic changes associated with hyperventilation trigger the partially damaged nerve to fire inappropriately.

Vestibulospinal Function

Bedside tests of vestibulospinal function are often insensitive because most patients can use vision and proprioceptive signals to compensate for any vestibular loss. Patients with acute unilateral peripheral vestibular lesions may past-point or fall toward the side of the lesion, but within a few days, balance returns to normal. Patients with bilateral peripheral vestibular loss have more

TABLE 428-1 DESCRIPTION, MECHANISM, AND FOCUS OF DIAGNOSTIC WORK-UP FOR COMMON TYPES OF DIZZINESS

TYPE OF DIZZINESS	DESCRIPTION	MECHANISM	FOCUS OF DIAGNOSTIC EVALUATION
Vertigo	Spinning (environment moves), tilt, drunkenness	Imbalance in tonic vestibular activity	Auditory and vestibular systems
Near-faint	Lightheaded, swimming	Decreased blood flow to entire brain	Cardiovascular system
Psychophysiologic	Dissociated from body, spinning inside (environment still)	Impaired central integration of sensory signals	Psychiatric assessment
Disequilibrium	Off balance, unsteady on feet	Loss of vestibulospinal, proprioceptive, cerebellar, or motor function	Neurologic assessment

difficulty compensating and usually show some imbalance on the Romberg and tandem walking tests (Chapter 396), particularly with eyes closed.

Doll's-Eye and Head-Thrust Tests

The vestibulo-ocular reflex can be tested at the bedside with the doll's-eye and head-thrust tests. In an alert human, rotating the head back and forth in the horizontal plane induces compensatory horizontal eye movements that are dependent on both the visual and vestibular systems. The doll's-eye test is a test of vestibular function in a comatose patient (Chapter 404) because such patients cannot generate pursuit or corrective fast components. In this setting, conjugate compensatory eye movements indicate normally functioning vestibulo-ocular pathways. Because the vestibulo-ocular reflex has a much higher frequency range than the smooth pursuit system, a qualitative bedside test of vestibular function can be made with the *head-thrust test*. It is performed by grasping the patient's head and applying brief, small-amplitude, high-acceleration head thrusts first to one side and then the other. The patient fixates on the examiner's nose and the examiner watches for corrective saccades, which are a sign of an inappropriate compensatory slow phase.

Caloric Test

The caloric test induces endolymphatic flow in the horizontal semicircular canal and horizontal nystagmus by creating a temperature gradient from one side of the canal to the other. With a cold caloric stimulus, the column of endolymph nearest the middle ear falls because of its increased density. This causes the cupula to deviate away from the utricle (ampullofugal flow) and produces horizontal nystagmus with the fast phase directed away from the stimulated ear. A warm stimulus produces the opposite effect, causing ampullopetal endolymph flow and nystagmus directed toward the stimulated ear (a mnemonic is COWS, meaning cold opposite, warm same). Because of its ready availability, ice water (approximately 0° C) can be used for bedside caloric testing. To bring the horizontal canal into the vertical plane, the patient lies in the supine position with head tilted 30 degrees forward. Infusion of 1 to 3 mL of ice water induces a burst of nystagmus usually lasting about a minute. Greater than a 20% asymmetry in nystagmus duration suggests a lesion on the side of the decreased response. The ice water caloric test is a useful way to test the integrity of the oculomotor pathways in a comatose patient. In this case, ice water induces only a slow tonic deviation toward the side of stimulation.

Positional Tests

Examination for pathologic vestibular nystagmus should include a search for spontaneous and positional nystagmus (see Table 424-5). Because vestibular nystagmus secondary to peripheral vestibular lesions is inhibited with fixation, the yield is increased by impairing fixation with +30 lenses (Frenzel glasses) or infrared video recordings. Two types of positional testing are typically performed: moving the patient from the sitting to head-hanging-right and head-hanging-left positions (Dix-Hallpike test) and turning the head to the right and left while the patient lies supine. Induced positional nystagmus may be paroxysmal or persistent, and it may be in the same direction in all positions or change directions in different positions. The most common cause of positional nystagmus is otolith debris in the semicircular canals, either free floating (paroxysmal) or attached to the cupula (persistent). This type of nystagmus always occurs in the plane of the affected canal—vertical torsional for the vertical canals and horizontal torsional for the horizontal canal. By contrast, central positional nystagmus is often pure vertical or horizontal and cannot be explained by stimulating a single semicircular canal.

Nystagmography

Nystagmography tests oculomotor control by inducing and recording eye movements. A standard test battery includes (1) tests of visual ocular control (saccades, smooth pursuit, and optokinetic nystagmus), (2) a careful search for pathologic nystagmus with fixation and with eyes open in darkness, and (3) the measurement of induced vestibular nystagmus (caloric and rotational). Nystagmography can be helpful in identifying a vestibular lesion and localizing it within the peripheral and central pathways.

Evaluating the "Dizzy" Patient

The history is key because it determines the type of dizziness (vertigo, near-faint, psychophysiologic disequilibrium), associated symptoms (neurologic, audiologic, cardiac, psychiatric), precipitating factors (position change, trauma, stress, drug ingestion), and predisposing illness (systemic viral infection, cardiac disease, cerebrovascular disease). The history provides direction for both the examination and the diagnostic evaluation (Table 428-1). When focal neurologic signs are found, neuroimaging usually leads to a specific diagnosis. When vertigo is present without focal neurologic symptoms or signs, head-thrust and positional testing are key to localizing the lesion to the labyrinth or eighth nerve. Audiometry and nystagmography are useful if the cause of vertigo is not clear after the history and examination. Patients with psychophysiologic dizziness should be identified early so that needless tests are not obtained. A detailed cardiac evaluation (including loop monitoring) often identifies the cause of episodic near-fainting (Chapters 51 and 62).

TREATMENT

Rx

Treatment of vertigo can be divided into three general categories: specific, symptomatic, and rehabilitative. When possible, treatment should be directed at the underlying disorder (Table 428-2). Specific therapies include particle repositioning maneuvers (the Epley and Semont maneuvers; see Fig. 428-4) for benign paroxysmal positional vertigo. For vestibular neuritis, steroids (e.g., methylprednisolone, 1 mg/kg/day for 5 days, then tapered during the next 15 days) are effective, at least for the short term, but antiviral agents are not. For Meniere disease, a low-salt diet and diuretics (e.g., 25 mg hydrochlorothiazide and 50 mg triamterene daily) are effective. Intratympanic gentamicin can significantly reduce vertigo in patients with unilateral Meniere disease who do not respond to medical treatment. Endolymphatic duct blockage is a potential option for medically refractory Meniere disease.

In many cases, however, symptomatic treatment either is combined with specific therapy or is the only treatment available. Many different classes of drugs have been found to have antivertiginous properties, and in most instances, the exact mechanism of action is uncertain. All these agents produce potentially unpleasant side effects, and the decision concerning which drug or combination to use is based on their known complications and on the severity and duration of the vertigo. An episode of prolonged, severe vertigo is one of the most distressing symptoms that a patient can experience. Affected patients prefer to lie still with eyes closed in a quiet, dark room. Antivertiginous drugs with sedation, such as promethazine HCl (25 mg) or diazepam (5 mg), may be helpful. Prochlorperazine suppositories (25 mg) may stop vomiting.

In more chronic vertiginous disorders, when the patient is trying to carry on normal activity, less sedating antivertiginous medications, such as meclizine (25 mg) or transdermal scopolamine (0.5 mg every 3 days), may provide relief. Chronic use of these drugs should be avoided.

Vestibular rehabilitation exercises are designed to help the patient compensate for permanent loss of vestibular function. As the acute stage of nausea and vomiting subsides, the patient should attempt to focus the eyes and to move and hold them in the direction that provokes the most dizziness. A useful exercise involves staring at a visual target while oscillating the head from side to side or up and down, slow at first and then fast. The patient should try to stand and walk, at first in contact with a wall or with an assistant, and make slow supported turns. As improvement occurs, head movements should be added while standing and walking.

TABLE 428-2 TREATMENT OF COMMON VERTIGO SYNDROMES

SYNDROME	TREATMENT
Benign positional vertigo	
Posterior canal variant	Epley's maneuver (see Fig. 428-4)
Horizontal canal variant	Barbecue roll toward normal side (side with less nystagmus), sleep with normal ear down
Vestibular neuritis	Methylprednisolone, 100 mg × 3 days, gradual taper during 22 days (must start within 3 days of onset)
Meniere disease	
Medical	Low salt (1-2 g salt/day) *and* either hydrochlorothiazide 25-50 mg/day *or* hydrochlorothiazide 25 mg/day plus triamterene 50 mg/day
Surgical	Intratympanic gentamicin, vestibular nerve section

Grade A References

A1. Wei BP, Stathopoulos D, O'Leary S. Steroids for idiopathic sudden sensorineural hearing loss. *Cochrane Database Syst Rev*. 2013;7:CD003998.

A2. Rauch SD, Halpin CF, Antonelli PJ, et al. Oral vs intratympanic corticosteroid therapy for idiopathic sudden sensorineural hearing loss: a randomized trial. *JAMA*. 2011;305:2071-2079.

A3. Gundogan O, Pinar E, Imre A, et al. Therapeutic efficacy of the combination of intratympanic methylprednisolone and oral steroid for idiopathic sudden deafness. *Otolaryngol Head Neck Surg*. 2013;149:753-758.

A4. Durga J, Verhoef P, Anteunis LJ, et al. Effects of folic acid supplementation on hearing in older adults: a randomized, controlled trial. *Lancet*. 2007;369:208-216.

A5. Cima RF, Maes IH, Joore MA, et al. Specialised treatment based on cognitive behaviour therapy versus usual care for tinnitus: a randomised controlled trial. *Lancet*. 2012;379:1951-1959.

A6. Hunt WT, Zimmermann EF, Hilton MP. Modifications of the Epley (canalith repositioning) manoeuvre for posterior canal benign paroxysmal positional vertigo (BPPV). *Cochrane Database Syst Rev*. 2012;4:CD008675.

A7. Kim JS, Oh SY, Lee SH, et al. Randomized clinical trial for apogeotropic horizontal canal benign paroxysmal positional vertigo. *Neurology*. 2012;78:159-166.

A8. Kim JS, Oh SY, Lee SH, et al. Randomized clinical trial for geotropic horizontal canal benign paroxysmal positional vertigo. *Neurology*. 2012;79:700-707.

A9. Fishman JM, Burgess C, Waddell A. Corticosteroids for the treatment of idiopathic acute vestibular dysfunction (vestibular neuritis). *Cochrane Database Syst Rev*. 2011;5:CD008607.

A10. Pullens B, van Benthem PP. Intratympanic gentamicin for Meniere's disease or syndrome. *Cochrane Database Syst Rev*. 2011;3:CD008234.

A11. Saliba I, Gabra N, Alzahrani M, et al. Endolymphatic Duct Blockage: A Randomized Controlled Trial of a Novel Surgical Technique for Ménière's Disease Treatment. *Otolaryngol Head Neck Surg*. 2015;152:122-129.

A12. Hillier SL, McDonnell M. Vestibular rehabilitation for unilateral peripheral vestibular dysfunction. *Cochrane Database Syst Rev*. 2011;2:CD005397.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Nash SD, Cruickshanks KJ, Klein R, et al. The prevalence of hearing impairment and associated risk factors: the Beaver Dam Offspring Study. *Arch Otolaryngol Head Neck Surg*. 2011;137:432-439.
2. Schreiber BE, Agrup C, Haskard DO, et al. Sudden sensorineural hearing loss. *Lancet*. 2010; 375:1203-1211.
3. Pacala JT, Yueh B. Hearing deficits in the older patient: "I didn't notice anything". *JAMA*. 2012; 307:1185-1194.
4. Gaylor JM, Raman G, Chung M, et al. Cochlear implantation in adults: a systematic review and meta-analysis. *JAMA Otolaryngol Head Neck Surg*. 2013;139:265-272.
5. Baguley D, McFerran D, Hall D. Tinnitus. *Lancet*. 2013;382:1600-1607.
6. Ozono Y, Kitahara T, Fukushima M, et al. Differential diagnosis of vertigo and dizziness in the emergency department. *Acta Otolaryngol*. 2014;134:140-145.
7. Kim JS, Zee DS. Clinical practice. Benign paroxysmal positional vertigo. *N Engl J Med*. 2014;370: 1138-1147.
8. Jeong SH, Kim HJ, Kim JS. Vestibular neuritis. *Semin Neurol*. 2013;33:185-194.
9. Harcourt J, Barraclough K, Bronstein AM. Meniere's disease. *BMJ*. 2014;349:g6544.
10. Huh YE, Kim JS. Bedside evaluation of dizzy patients. *J Clin Neurol*. 2013;9:203-213.

REVIEW QUESTIONS

1. Which of the following would not be typical of a conductive hearing loss?

A. Marked improvement with amplification
B. Ability to hear speech in a noisy background
C. Relatively maintained speech recognition
D. Loud speech is annoying
E. Bone greater than air conduction

Answer: D Patients with conductive hearing loss prefer loud speech, which they can understand as well as normal people can. A, B, C, and E are typical for conductive hearing loss.

2. Which of the following diagnoses would be least likely to be manifested with a unilateral hearing loss?

A. Acoustic neuroma
B. Meniere disease
C. Otosclerosis
D. Brain stem glioma
E. Otitis media

Answer: D Brain stem lesions rarely cause unilateral hearing loss unless they involve the root entry zone of the cochlear nerve. Otosclerosis is usually bilateral but often begins unilaterally. The other conditions are typically unilateral.

3. Which of the following is least likely to be manifested with vertigo?

A. Acoustic neuroma
B. Meniere disease
C. Lateral medullary infarction
D. Migraine
E. Vestibular neuritis

Answer: A Acoustic neuroma typically is manifested with unilateral hearing loss or tinnitus. It compresses the vestibular nerve slowly, so central compensation can occur. It rarely is manifested with vertigo. B, C, D, and E commonly are manifested with vertigo.

4. Which of the following would be an unusual precipitant for benign paroxysmal positional vertigo?

A. Turning in bed
B. Yoga class
C. Driving
D. Reaching for something on a high shelf
E. Working under an automobile

Answer: C Benign paroxysmal positional vertigo is typically triggered by movement in the vertical plane (plane of the posterior semicircular canal). It would be unusual to make such a movement while driving. The other maneuvers are common precipitating circumstances.

5. Which of the following would most likely be associated with a positive head-thrust test result?

A. Meniere disease
B. Lateral medullary infarction
C. Cerebellar infarction
D. Gentamicin ototoxicity
E. Otitis media

Answer: D Gentamicin is remarkably selective for the vestibular system, and the head-thrust test is useful for identifying toxicity at the bedside. The head-thrust test result is rarely positive with Meniere disease and would be negative with lateral medullary infarction, cerebellar infarction, and otitis media.

429

THROAT DISORDERS

PAUL W. FLINT

Nearly every systemic and infectious disease results in head and neck manifestations, with the majority affecting the upper aerodigestive tract. Diseases of the upper aerodigestive system include infection (acute and chronic; viral, bacterial, and fungal), systemic disease, and neoplasm (Chapter 190), some of which require urgent care or referral to an otolaryngologist.

Abnormalities of swallowing, respiratory function, voice, and speech are influenced by the anatomic site involved, the host's immune status and inflammatory response, the severity of the disease process, and the presence or absence of neurologic involvement.

In a patient with hoarseness, current clinical practice guidelines recommend visualization of the larynx for symptoms that persist for 3 months or longer; however, warning signs of a potentially serious or emergent throat condition warrant referral regardless of duration. These conditions include persistent throat pain with or without trismus, difficulty in swallowing, difficulty in breathing, hemoptysis, and ear pain with normal ear examination findings. Diagnostic testing options, which include fiberoptic examination, imaging, pulmonary function studies, and laboratory testing, are directed by history, symptoms, and physical findings.

ANATOMY OF THE UPPER AERODIGESTIVE TRACT

The pharynx is divided into three anatomic regions (Fig. 429-1). The *nasopharynx* is the region above the soft palate and uvula. Its anatomic components include the adenoids, the openings of the eustachian tubes, Rosenmüller's fossa at the junction of the posterior and lateral walls, and the posterior aspect of the inferior turbinates of the nasal cavity. Diseases of the nasopharynx typically produce few symptoms until the process is well advanced and causes nasal obstruction (Chapter 426), epistaxis (Chapter 426), ear pain (Chapter 426), headache (Chapter 398), or cranial nerve abnormalities due to extension to the skull base. The *oropharynx* begins at the level of the soft palate and extends inferiorly to the tip of the epiglottis. This region includes the faucial tonsils, the base of the tongue, the lingual tonsils, the soft palate, the uvula, and part of the posterior pharyngeal wall. The *hypopharynx*, which extends from the tip of the epiglottis to the upper esophagus (the cricopharyngeus muscle) below, includes the larynx (epiglottis, arytenoids, glottis or true vocal cords), the piriform sinuses (pharyngeal folds lateral to the larynx), and the posterior pharyngeal wall. The tip of the epiglottis can be visualized by an experienced examiner with use of a laryngeal mirror or sometimes even on a routine oral examination with just a flashlight and tongue blade. The nasopharynx and hypopharynx are best visualized with a flexible fiberoptic nasopharyngoscope.

INFECTIOUS DISEASES OF THE UPPER AERODIGESTIVE SYSTEM

Infectious disorders of the upper aerodigestive tract typically are manifested as sore throat (pharyngitis), changes in voice (laryngitis), or both. The clinical evaluation must differentiate among bacterial (usually streptococcus [Chapter 290]), viral, and other infections and systemic causes (Table 429-1). Clinical differentiation of sore throat is critical to primary care and emergency management of the airway.

Pharyngitis

Bacterial infection accounts for approximately 5 to 10% of pharyngitis in adults compared with 30 to 40% in children. Unfortunately, as many as two thirds of adults with a sore throat are prescribed antibiotics.

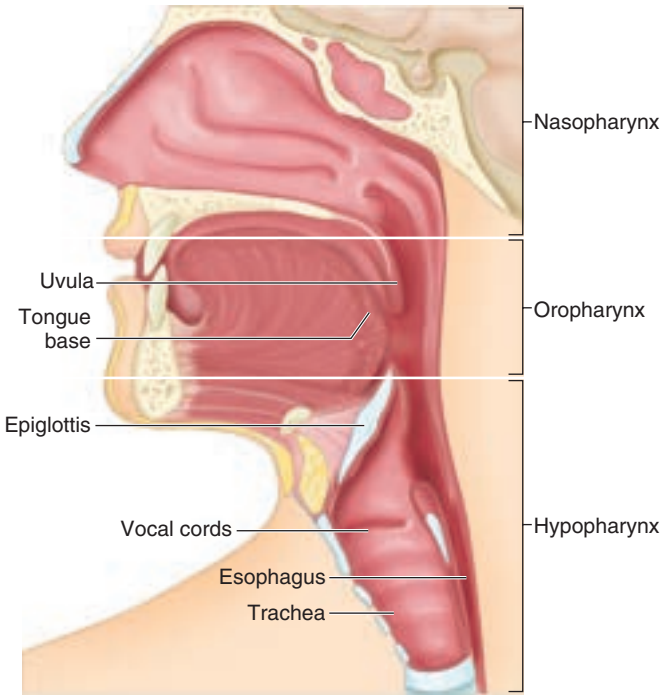

FIGURE 429-1. The pharynx (throat) is typically divided into three distinct anatomic regions (nasopharynx, oropharynx, and hypopharynx). (Courtesy Thomas A. Tami, MD.)

TABLE 429-1 CLINICAL DIFFERENTIATION OF COMMON CONDITIONS ARISING AS SORE THROAT

FEATURE	VIRAL PHARYNGITIS	BACTERIAL TONSILLITIS	PERITONSILLAR ABSCESS	EPIGLOTTITIS
Tonsillar enlargement	Usual	Rare	None	None
Tonsillar exudates	Occasional (infectious mononucleosis)	Usual	Often	None
Tonsillar asymmetry	None	None	Usual	None
Trismus (inability to open jaw)	None	None	Usual	None
Cervical adenopathy	Occasional	Usual (tender)	Usual (tender)	None
Tender larynx	Rare	None	None	Usual

From Tami TA. Throat disorders. In: Goldman L, Schafer AI, eds. *Goldman's Cecil Medicine.* 24th ed. Philadelphia: Elsevier Saunders; 2012.

STREPTOCOCCAL INFECTIONS

Group A beta-hemolytic *Streptococcus pyogenes* (Chapter 290) is the most common cause of bacterial pharyngitis in adults, although it accounts for only 10% of all pharyngitis in adults. Infection is manifested with the rapid onset of sore throat, often accompanied by pain with swallowing, fever, chills, malaise, headache, mild neck stiffness, and anorexia. Hypertrophic tonsils with exudates, foul breath, and tender cervical adenopathy are hallmark findings.[1] Some patients have palatal petechiae or a scarlatiniform rash. Rhinorrhea, hoarseness, cough, conjunctivitis, diarrhea, and ulcerative oral lesions are less common.

Untreated group A beta-hemolytic *S. pyogenes* pharyngitis usually resolves within 3 to 7 days. The administration of antibiotics within 24 to 48 hours reduces pain by approximately 1 day,■ whereas both immediate and delayed antibiotics reduce the risk of suppurative complications.[2] Antibiotics also reduce the contagious period from 2 weeks to 24 hours after administration. For prevention of rheumatic fever (Chapter 290), antibiotic therapy must be started within 10 days after the onset of symptoms. The risk of acute poststreptococcal glomerulonephritis (Chapter 121), however, is not affected by antibiotics.

To minimize the potential side effects and costs of unnecessary antibiotics, antibiotic therapy should be based on the presence of fever, tender anterior cervical adenopathy, tonsillar swelling or exudates, age, and the absence of cough (Table 429-2). If three or four of these criteria are present, the likelihood of group A beta-hemolytic streptococcal infection is 40 to 60%, whereas fewer criteria are associated with progressively lower probabilities. A rapid antigen test may be obtained because of the residual risk of false-positive or false-negative diagnoses, but its incremental diagnostic value is low, except in borderline cases.[3] If the rapid antigen test result is negative but the clinical suspicion remains high, a throat culture specimen should be obtained for confirmation.

Antibiotic options (Chapter 290) include penicillin (penicillin VK, 250 mg three times a day or 500 mg twice a day for 5 to 10 days), which is usually chosen to treat acute bacterial pharyngitis,■ although cefuroxime axetil (250 mg twice a day for 5 to 10 days) is even more effective for primary treatment and can be effective for persistent infection. In patients with proven recurrent infections, clindamycin (300 mg orally three times a day for 10 days) or amoxicillin–clavulanic acid (875 mg orally twice a day, or 500 mg three times a day for 10 days) is recommended. For patients who are allergic to penicillin, azithromycin (500 mg/day for 3 days or a single dose of 2 g)■ is another alternative. Evidence also suggests that a single dose of oral or intramuscular corticosteroids given at the start of treatment will reduce the pain of severe pharyngitis, especially in children.■ In patients with recurrent symptomatic episodes despite appropriate antimicrobial therapy, tonsillectomy can decrease future throat infections compared with continued observation.■

Non–group A beta-hemolytic streptococcal infections (Chapter 290), including groups B, C, and G, can cause acute pharyngitis with a clinical picture that mirrors that of group A beta-hemolytic streptococcal pharyngitis. Glomerulonephritis is a known sequela, whereas rheumatic fever is not. Penicillin or clindamycin, as prescribed for group A beta-hemolytic streptococcus, provides adequate coverage.

TABLE 429-2 GUIDELINES FOR THE MANAGEMENT OF PHARYNGITIS*

CENTOR SCORE†	PERCENTAGE POSITIVE FOR *STREPTOCOCCUS* INFECTION	ACP/CDC GUIDELINES
0	7%	Do not test, do not treat
1	12%	Do not test, do not treat
2	21%	Treat if rapid test result positive
3	38%	Option 1: treat if rapid test result positive Option 2: treat empirically
4	57%	Treat empirically

*As recommended by the American College of Physicians (ACP) and the Centers for Disease Control and Prevention (CDC). See *Arch Intern Med.* 2012;172:847-852.
†Calculated as follows: 1 point each for temperature >38° C, absence of cough, presence of swollen and tender anterior cervical nodes, tonsillar swelling or exudate, or age 3 to 14 years; and −1 point for age ≥45 years.

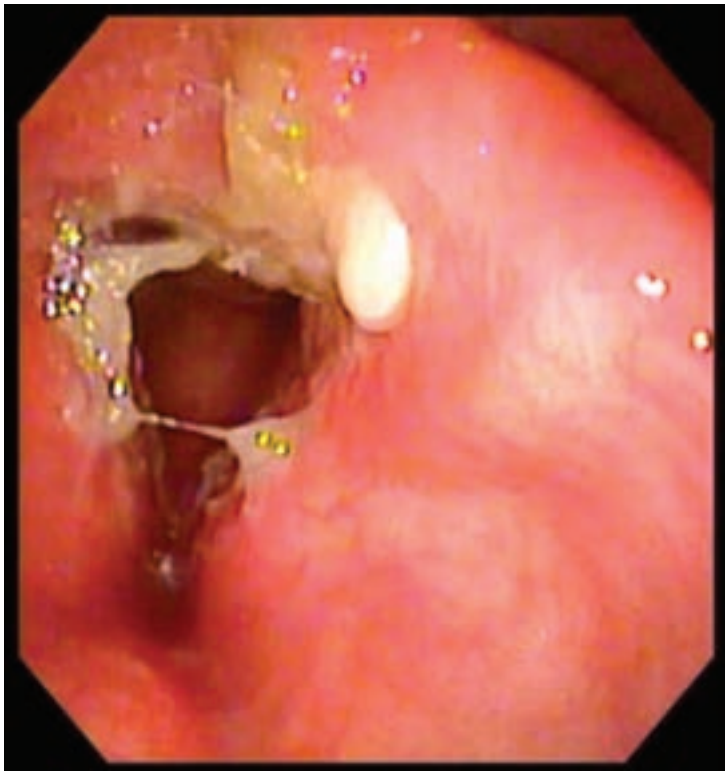

FIGURE 429-2. Fiberoptic laryngoscopy demonstrating chronic laryngitis secondary to methicillin-sensitive *S. aureus* infection. Diagnosis is based on culture and biopsy.

NONSTREPTOCOCCAL BACTERIAL PHARYNGITIS

A variety of bacteria other than streptococci can infect the throat. *Staphylococcus aureus* (Chapter 288) infections, whether caused by methicillin-resistant (MRSA) or methicillin-sensitive (MSSA) strains, usually are manifested with chronic hoarseness. Laryngoscopy typically reveals thickened erythematous vocal folds with edema, whitish debris, and crusting, which may resemble leukoplakia (Fig. 429-2). Trimethoprim-sulfamethoxazole (160 mg/800 mg twice daily for 2 to 4 weeks) has been shown to be effective,[4] although treatment should be guided by antibiotic sensitivities.

Bordetella pertussis (Chapter 313) infections have become more common in adults because of their gradual loss of immune protection after vaccination. Adults typically present with cough, often but not always accompanied by

nonspecific upper respiratory symptoms, fever, and leukocytosis. A *B. pertussis* serum immunoglobulin G titer higher than 27 IU/mL is highly predictive of recent infection.[5] Erythromycin (500 mg four times a day for 14 days) or azithromycin (500 mg single dose orally on day 1, then 250 mg daily on days 2 through 5) is effective treatment.

Neisseria gonorrhoeae (Chapter 299) can cause sexually transmitted gingivitis, stomatitis, glossitis, and pharyngitis, especially in men who have sex with men. Treatment is the same as for urogenital disease and should include treatment for chlamydia. Options include ceftriaxone, 250 mg intramuscularly in a single dose; azithromycin, 1 g orally in a single dose; and doxycycline, 100 mg orally twice daily for 7 days. *Treponema pallidum* (Chapter 319) can cause oral and oropharyngeal ulcerations that involve the lips, tongue, and tonsil. Treatment is the same as that recommended for systemic disease. *Chlamydia* (Chapter 318), which is commonly associated with pneumonia and bronchitis, also can cause pharyngitis and hoarseness, sometimes as its presenting symptoms. *Mycoplasma pneumoniae* (Chapter 317), which frequently accounts for 15 to 20% of cases of community-acquired pneumonias, also can cause sore throat, nasal congestion, and coryza. Treatment of chlamydia and mycoplasma infections with tetracyclines, macrolides, and quinolones is the same as for the pneumonias they cause (Chapters 318 and 317).

Francisella tularensis (Chapter 311) is a gram-negative bacillus that causes tularemia. Oropharyngeal involvement is associated with fever, pharyngeal erythema, exudative tonsillitis, and tender lymphadenopathy. A false-positive monospot test result and atypical lymphocytosis can mimic infectious mononucleosis. The organism is sensitive to macrolides, fluoroquinolones, and tetracyclines.

Corynebacterium diphtheriae infection (Chapter 292) involves the mucosal surfaces of the upper respiratory tract, where it causes a patchy gray-black pseudomembrane in the nasopharynx, oropharynx, larynx, and trachea. About 75% of patients complain of throat pain. Airway obstruction and severe dysphagia are life-threatening sequelae. Antitoxin is given in combination with penicillin, erythromycin, tetracycline, clindamycin, or rifampin.

Arcanobacterium haemolyticum (Chapter 292) is a gram-positive bacillus that can cause pneumonia, meningitis, osteomyelitis, brain abscess, and peritonsillar abscess in both normal and immunocompromised patients. Uncomplicated pharyngitis may be treated with erythromycin (500 mg four times a day for 10 days). Complicated infections require intravenous dosing with vancomycin, clindamycin, or cephalexin, with or without gentamicin (see Table 287-4).

PERITONSILLAR ABSCESS AND DEEP SPACE INFECTIONS

Only about 1% of patients with acute bacterial pharyngitis develop serious suppurative complications.[6] The best predictors are severe tonsillar inflammation and severe earache, but most complications occur in patients with neither of these findings. Peritonsillar abscesses usually can be diagnosed on physical examination (Fig. 429-3) and managed by an otolaryngologist in an outpatient setting by surgical or needle drainage. Oral antibiotics, such as amoxicillin (500 mg orally twice a day for 10 days) plus metronidazole (500 mg orally three times a day for 10 days), clindamycin (300 to 600 mg three times a day for 14 days), or amoxicillin-clavulanate (875 mg orally twice a day for 10 days), are recommended.

By comparison, patients with deep neck abscesses often have swelling of the external neck, trismus, torticollis, and even a compromised airway due to infection that has spread to the fascial planes of the neck and chest. These infections require urgent evaluation, usually with a computed tomography scan with contrast enhancement. Aggressive management includes incision and drainage as well as broad-spectrum intravenous antibiotics that cover aerobic and anaerobic bacteria (e.g., clindamycin, 600 mg intravenously every 8 hours; ampicillin-sulbactam, 3 g every 6 hours; or penicillin G, 2 million units every 4 hours plus metronidazole 500 mg every 6 hours).[7]

Severe infections of the pharynx can cause septic thrombophlebitis of the internal jugular vein (Lemierre's syndrome), an uncommon but serious complication. Patients typically present with persistent fever, difficulty in swallowing, neck pain, and a neck mass due to an underlying peritonsillar, retropharyngeal, or parapharyngeal abscess.[8] Diagnosis is best established with a contrast-enhanced computed tomography scan of the neck. This potentially life-threatening condition is almost always associated with anaerobic infection, especially with *Fusobacterium necrophorum* or *A. haemolyticum* (Chapter 297). Infection can extend into the intrathoracic vasculature, and patients can develop bacteremia and septic pulmonary emboli. Treatment should be directed toward anaerobic coverage (e.g., clindamycin, 600 mg every 8 hours, or metronidazole, 500 mg every 6 hours). Anticoagulation with heparin is controversial and generally reserved for persistent septic emboli. Surgical intervention includes drainage of the abscess. Ligation or excision of the jugular vein may be indicated for persistent septic emboli unresponsive to medical management. Mortality rates can be as high as 5%.

VIRAL INFECTIONS

In adults, the common cold (Chapter 361) causes 30 to 60% of cases of pharyngitis, with rhinovirus (Chapter 361) accounting for the majority of cases, followed by coronavirus (Chapter 366) and parainfluenza (Chapter 363) (see Table 429-2). U.S. adults experience an average of 2.5 episodes per year of noninfluenza upper respiratory infections, each with a 7.4-day average duration of symptoms. For the entire U.S. population, these 500 million episodes cost an estimated $40 billion annually, in part because of associated systemic symptoms such as fever, cough, and sinusitis and in part because of associated exacerbations of allergies, asthma, and chronic obstructive pulmonary disease.

The viral infection that is most likely to be confused with a bacterial infection is mononucleosis. Mononucleosis is caused by Epstein-Barr virus (Chapter 377), which has a seroprevalence of 67% in U.S. children and adolescents aged 6 to 19 years.[9] After an incubation period of 3 to 7 weeks, patients present with initial malaise, fever, and chills followed by sore throat, fever, and anorexia. Some patients have associated abdominal discomfort due to splenomegaly or hepatomegaly, headache, stiff neck, and rash. On physical examination, patients have erythematous pharyngitis with exudative tonsillar hypertrophy (Fig. 429-4), prominent lingual tonsils, and adenoid hypertrophy (Waldeyer's ring). Aphthous-type ulcerations and petechiae may be seen, especially at the junction of the hard and soft palate. Impressive cervical adenopathy is typical, and 50% of patients have splenomegaly. Lymphoid hyperplasia can cause some degree of upper airway obstruction in about 5% of patients.

A blood count will show lymphocytosis, usually with more 10% atypical lymphocytes. The heterophil antibody test result is often positive, and Epstein-Barr virus–specific antibody tests are diagnostic. β-Lactam

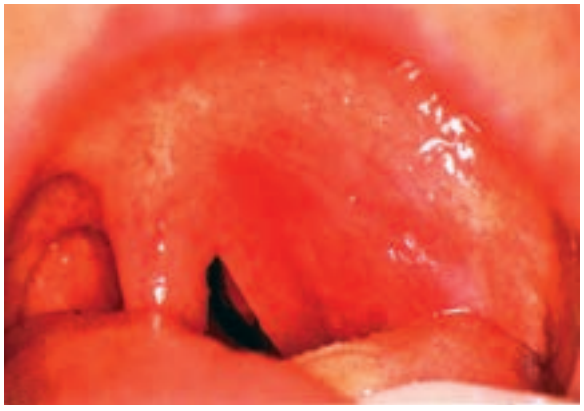

FIGURE 429-3. Left peritonsillar abscess identified by bulging anterior pillar and soft palate with midline shift. (Courtesy Thomas A. Tami, MD.)

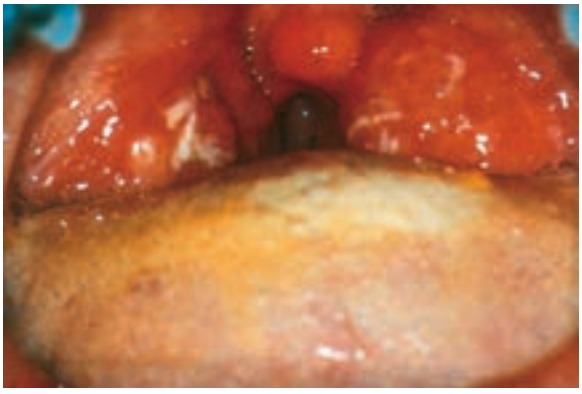

FIGURE 429-4. Mononucleosis with symmetrical exudative tonsillitis. (Courtesy Thomas A. Tami, MD.)

antibiotics, which may mistakenly be prescribed, will cause a maculopapular rash in 95% of patients. Laryngeal obstruction may require hospitalization and intravenous corticosteroids (e.g., dexamethasone, 8 to 10 mg intravenously three times a day).

Influenza virus infection can include nonexudative pharyngitis, but the predominant symptoms are tracheobronchial, usually accompanied by fever, headache, rhinorrhea, cough, and myalgia, without lymphadenopathy.[10] Adenovirus (Chapter 365) can cause pharyngitis associated with fever, nonproductive cough, nasal congestion, myalgia, headache, nausea, vomiting, and diarrhea, especially in outbreaks, such as among military recruits, or in immunocompromised patients.

Primary herpes simplex virus (Chapter 374) infection is characterized by pharyngitis with or without gingival stomatitis. Symptoms include sore throat, fever, and malaise; physical findings include erythema and hypertrophy of the tonsils with exudates, often with enlarged, tender cervical nodes. It may be difficult to distinguish clinically from group A beta-hemolytic streptococcal pharyngitis unless patients have herpes-like lesions of the oral cavity or oropharynx.

Human immunodeficiency virus infection (Chapter 384) can be manifested as an acute retroviral syndrome that mimics infectious mononucleosis and resolves in 1 to 2 weeks. The diagnosis must be considered in febrile patients with known risk factors. Once infection is established, oral and oropharyngeal infectious ulcerations may be due to herpes simplex virus, cytomegalovirus, syphilis, cryptococcus, histoplasmosis, or mycobacteria. Large, painful, noninfectious aphthous ulcers also can involve the tonsillar fossa, floor of mouth, hypopharynx, and epiglottis (Chapter 425).

FUNGAL INFECTIONS

By far, the most common fungal infection of the oropharynx and larynx is candidiasis. *Candida* (Chapter 338) is a normal commensal organism of the oral cavity and oropharynx, but it can be an opportunistic infectious agent in immunocompromised patients, patients who have received prior head and neck irradiation, patients with xerostomia or diabetes, and immunocompetent patients who have been treated with antibiotics or with systemic or inhaled steroids. *Candida* infection is manifested with sore throat, burning of the mouth and tongue, dysgeusia, dysphagia, and hoarseness. White pseudomembranes characteristic of thrush (Fig. 429-5) may involve the oral cavity, oropharynx, hypopharynx, larynx, and esophagus. Treatment includes oral hygiene, probiotics, and topical antifungals (Chapter 338). Fluconazole (200 mg orally once a day for 14 to 21 days) is indicated for laryngeal and esophageal involvement and for recurrent disease. Itraconazole (200 mg orally once a day for 14 to 21 days) should be considered for nonresponders.

Other fungal infections may affect the oropharynx and larynx in isolation or as part of a systemic infection. *Blastomycosis* (Chapter 334) involves the larynx in less than 5% of cases, in which it produces pseudoepitheliomatous hyperplasia and clinically resembles squamous cell carcinoma. *Histoplasmosis* (Chapter 332), which is endemic to the Ohio and Mississippi river valleys, can involve the oral cavity, oropharynx, and larynx in immunocompromised patients. *Cryptococcosis* (Chapter 336) can involve the larynx and occurs most often in an immunocompromised host, in whom pseudoepitheliomatous hyperplasia may occur.

Paracoccidioidomycosis (Chapter 335) is the leading cause of fungal laryngitis in South America, especially among farmers. It causes pseudoepitheliomatous hyperplasia that can be misdiagnosed initially as squamous cell carcinoma. *Coccidioidomycosis* (Chapter 333), which is endemic to southwestern United States, can involve the larynx, in which it results in hoarseness and throat pain and can progress to airway obstruction. For each of these fungal diseases, treatment is similar to what is recommended for the disseminated disease (Chapter 331) that usually accompanies pharyngeal and laryngeal disease.

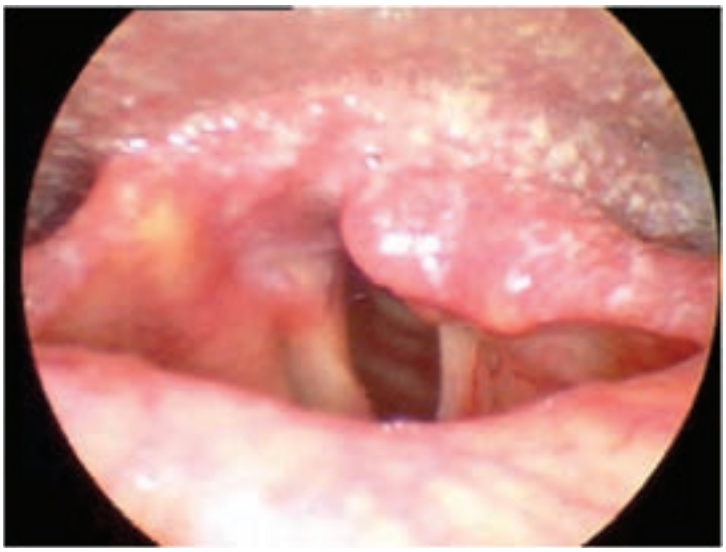

FIGURE 429-5. Fiberoptic laryngoscopy demonstrating characteristic erythema and white pseudomembrane secondary to *Candida* infection.

MYCOBACTERIAL INFECTIONS

The risk that tuberculosis (Chapter 324) infection will involve the oropharynx and larynx is low, with only about 1 to 1.5% of infected tuberculosis patients having involvement of the tonsils or larynx. Tonsillar involvement, which most often occurs in the presence of systemic disease, is manifested as sore throat with exudative tonsillitis and cervical adenopathy. Other diseases that can have a similar clinical presentation include lymphoma (Chapter 185), squamous cell carcinoma (Chapter 190), and sarcoidosis (Chapter 95). Tuberculosis in the larynx is most likely to involve the vocal folds and supraglottis (false vocal folds). It is manifested with hoarseness and odynophagia, and its lesions mimic squamous cell carcinoma. Approximately 50% of cases occur in the presence of active disseminated disease, about one third occur with inactive disease, and the other 15% occur as primary laryngeal disease.[11]

In contrast, *Mycobacterium leprae* infection (Chapter 326) is manifested with laryngeal disease in one third of patients with systemic disease. The clinical picture is indistinguishable from that of laryngeal tuberculosis. Atypical mycobacteria rarely involve the larynx, and infection most often is manifested as cervical adenopathy.

CHRONIC TONSILLITIS

Patients may develop deep tonsillar crypts that accumulate debris, such as food or sloughed mucosa, thereby providing an ideal environment for the growth of bacteria, especially anaerobes. Such patients commonly complain of whitish or yellow pieces of semisolid debris on or emanating from their tonsils. These tonsilliths often have a foul taste and odor, and they can cause halitosis. Some patients have chronic sore throats because of the persistent infection. Treatment includes frequent gargling of a hydrogen peroxide mouthwash and occasionally expressing this debris from the tonsil manually. Long-term amoxicillin (500 mg three times a day for 21 days) or clindamycin (300 mg orally three times a day for 21 days) may be effective; however, the presence of *Actinomyces*, a commensal organism of the oral cavity and oropharynx, is indicative of chronic infection that requires tonsillectomy because even long-term antibiotics are unlikely to be effective.[12]

Epiglottitis

Epiglottitis (supraglottitis) is an uncommon problem in adults and has become even less common in children because of routine *Haemophilus influenzae* vaccination in children (Chapters 18 and 300). In adults, *Streptococcus pneumoniae* (Chapter 289) is now the most common organism,[13] and adults present with a severe sore throat, odynophagia, fever, and "hot potato" voice. Airway obstruction occurs less frequently than in children, although it should be considered a possibility. Palpation or movement of the larynx causes significant pain. On transnasal fiberoptic laryngoscopy, the larynx typically reveals swelling, erythema, and occasionally exudates of the epiglottis and other supraglottic structures. Patients with a confirmed diagnosis of epiglottitis require intravenous antibiotics (e.g., cefotaxime, 2 g every 6 hours, or ceftriaxone, 1 to 2 g/day, intravenously), and they should be observed in an intensive care unit setting until symptoms improve because of the risk of rapidly progressive airway obstruction.

NONINFECTIOUS PHARYNGITIS

Laryngopharyngeal Reflux

Patients with gastroesophageal reflux disease (Chapter 138) can develop laryngopharyngeal reflux with intermittent hoarseness, nighttime or chronic cough, postnasal drip, "globus" sensation, reactive airway disease, halitosis,

and brackish or acid taste in the back of the mouth and throat. Findings on laryngoscopy, although nonspecific, may include posterior laryngitis, with swollen and erythematous arytenoid cartilages, thickening of the vocal folds, interarytenoid edema, and thickening of the mucosa. In severe cases, spasm or thickening of the cricopharyngeus muscle, also known as the upper esophageal sphincter (Chapter 138), can cause dysphagia as a result of poor pharyngeal emptying or even spillage of secretions into the larynx with aspiration.[14]

Treatment should address dietary change, behavioral modification, elevation of the head of the bed at night, and a therapeutic trial of a proton pump inhibitor for up to 3 months (Chapter 138). Empirical treatment without laryngoscopy is reasonable in patients with classic symptoms, but patients who do not respond within 3 months or patients with warning signs (e.g., ear pain, trismus, or odynophagia) require laryngoscopy to exclude more serious causes of hoarseness. In patients with persistent symptoms and a positive pH probe or with evidence of Barrett's esophagitis, an antireflux procedure (Chapter 138) should be strongly considered.[15]

SYSTEMIC DISEASES OF THE THROAT AND LARYNGITIS

About 80% of patients with *pemphigus* (Chapter 439) will have symptoms affecting the nasal cavity, oral cavity, and oropharynx, and half of these patients will have laryngeal involvement.[16] Shallow ulcerations with fibrinous material and surrounding erythema are characteristic. Bullous lesions are less likely to be observed because the epithelial layer sloughs during swallowing. Laryngeal involvement may result in stenosis with airway obstruction due to scarring. Upper aerodigestive tract involvement occurs in 35% of patients with *pemphigoid* (Chapter 439), and 50% of these patients will have laryngeal involvement. Treatment in both disorders consists of high-dose steroids (prednisone, 75 to 100 mg orally per day until remission) during the attack phase and then tapered to a maintenance dose (25 to 50 mg orally every other day). Other immunosuppressive medications, such as azathioprine, cyclophosphamide, or cyclosporine (Chapter 35), may be required in the maintenance phase. Perilesional or intralesional triamcinolone acetonide injections are recommended during the maintenance phase for new lesions.[17]

Granulomatosis with polyangiitis (Chapter 270) includes laryngeal involvement in 20% of patients, with a predilection for the subglottis (E-Fig. 429-1). Presenting symptoms may include hoarseness, cough, dyspnea, wheezing, and stridor. Flow-volume loops are useful and demonstrate flattening of both inspiratory and expiratory loops characteristic of a fixed extrathoracic obstruction. Patients presenting with airway obstruction require surgical intervention. Active disease with granulation tissue requiring airway management is treated with endoscopic dilation and steroid injection. Systemic immunosuppression is not effective for treatment of laryngeal involvement, so open resection may be required as the disease becomes chronic with the deposition of fibrous tissue.[18]

Relapsing polychondritis involves cartilage in the ear, nose, and upper and lower airway as well as in the articular joints and costal cartilage. About 50% of patients will develop dyspnea, cough, hoarseness, stridor, or wheezing due to the destruction of cartilage and the resulting loss of structural support of the airway (E-Fig. 429-2).[19] Airway obstruction may necessitate stenting or tracheostomy.

About 25 to 30% of patients with rheumatoid arthritis (Chapter 264) develop hoarseness, globus symptoms, and difficulty in swallowing. Hoarseness may result from acute inflammation or chronic nodular formation. Bilateral arytenoid joint involvement may impair vocal fold motion and cause airway obstruction with stridor. Surgical intervention may be necessary to open the airway. The role of steroids (systemic or injectable) for airway stenosis in rheumatoid arthritis has not been established.

About 1 to 5% of all patients with *sarcoidosis* (Chapter 95) develop laryngeal involvement, usually manifested as hoarseness. As the disease progresses, however, it can cause a conical stenosis due to thickening of the soft tissue (E-Fig. 429-3). Laryngeal paralysis can be caused by a mass effect or by adenopathy with peripheral nerve compression along the course of the vagus or recurrent laryngeal nerve. Vocal fold involvement responds to intralesional steroid injection, but endoscopic laser excision is recommended in patients with airway symptoms.

Amyloidosis (Chapter 188) may deposit anywhere in the upper aerodigestive tract. Hoarseness and cough are the most common symptoms in patients with laryngeal involvement (E-Fig. 429-4), but pharyngeal involvement may be associated with pain. Airway obstruction is rare. Surgical debulking with or without external beam radiation can relieve symptoms and even resolve the lesions.

NEUROLOGIC DISORDERS AFFECTING THE THROAT

Neurologic disorders of the oropharynx, hypopharynx, and larynx may be due to focal diseases or be a local manifestation of generalized neurologic disease. Head and neck manifestations of neurologic and neuromuscular disorders are classified as hyperfunctional and hypofunctional disorders. Hyperfunctional disorders include muscle tension dysphonia, dystonia (Chapter 410), essential tremor (Chapter 410), myoclonus, and stuttering. Hypofunctional neurologic disorders include Parkinson's disease (Chapter 409), multiple sclerosis (Chapter 411), neuromuscular disorders (Chapters 419 and 422), postpolio syndrome (Chapter 379), myopathies (Chapter 421), medullary disorders, and laryngeal paralysis. Given the variety of disorders with head and neck manifestations, it is critical to identify and to classify the physical findings and associated symptoms. For example, a patient who complains about the sound of the voice may in fact have a normal voice but actually have severe dysarthria and hypernasality secondary to amyotrophic lateral sclerosis. The site of the lesion associated with the neurologic disorder will result in characteristic physical findings and facilitate establishment of a correct diagnosis (Table 429-3).

Hyperfunctional Neurologic Disorders

In *dystonia* (Chapter 410), spasmodic dysphonia fluctuates moment to moment and day to day. Adductor spasms of the vocal folds produce strained, strangled vocal quality with pitch breaks. Abductor spasms produce breathy hypophonic word breaks. Nonspeech sounds (laughter) and singing voice may be normal. This condition responds well to botulinum toxin injections.[20]

Muscle tension dysphonia, which may be difficult to distinguish from dystonia, can occur secondary to underlying weakness (paresis, aging) with compensatory hyperfunction. Speech has a rough strained quality, perhaps with pitch breaks, and does not fluctuate moment to moment. It generally responds to voice therapy.

Vocal tremor is seen in 30% of patients with *essential tremor* (Chapter 410) and may be associated with spastic dystonia. Patients have a tremulous, quavering vocal quality, with or without head or hand tremor. Because of involvement of pharyngeal and laryngeal muscles, botulinum toxin may not be effective.

Myoclonus (Chapter 410) causes rhythmic contraction of the palate, pharynx, or larynx at a rate of one or two per second. Patients may have audible clicking from the eustachian tube or larynx. Voice may or may not be affected. Palate and vocal folds may be treated with botulinum toxin.

Pseudobulbar palsy (Chapters 404 and 419) causes spasticity and hyperreflexia of the bulbar muscles (pharynx, palate, lips tongue, and larynx).

TABLE 429-3 CORRELATING SITE OF LESION WITH PHYSICAL FINDINGS IN NEUROLOGICAL DISORDERS AFFECTING THE THROAT

SITE OF LESION	SIGNS
Cortex	Aphasia Aphonia Dysarthria Dysphonia Stridor
Extrapyramidal system	Vocal strain and pitch breaks Tremor Hypophonia/tachyphemia Spasmodic movements Focal, regional, or generalized dystonia
Cerebellum	Ataxia Dysmetria Tremor Incoordination
Brain stem	Flaccid paralysis Associated dense sensory deficit
Peripheral	Focal paresis or paralysis Other cranial nerve ± Palate involvement defines location

Modified from Blitzer A, Alexander RE, Grant NN. Neurologic disorders of the larynx. In: Flint PW, Haughey BH, Lund VJ, et al, eds. *Cummings: Otolaryngology–Head and Neck Surgery*. 6th ed. Philadelphia: Mosby Elsevier; 2015.

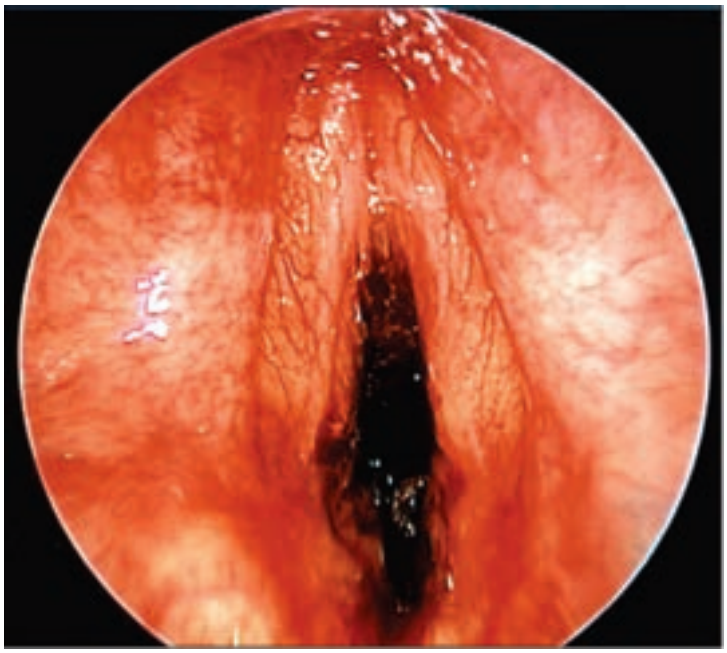

E-FIGURE 429-1. **Fiberoptic examination of a patient with granulomatosis with angiitis demonstrating laryngeal involvement at the level of the vocal folds and subglottis.**

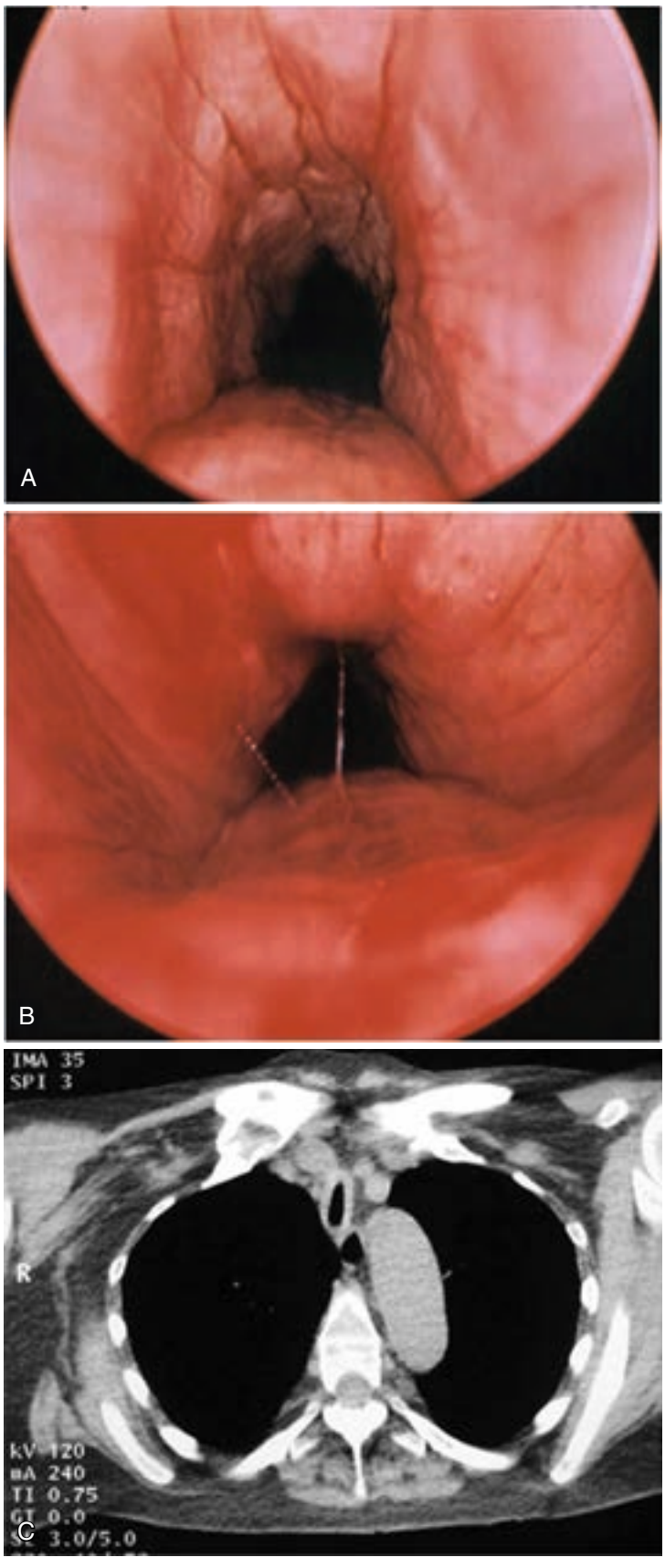

E-FIGURE 429-2. **Tracheoscopy in a patient with relapsing polychondritis with tracheal involvement. A,** On inspiration, negative intrathoracic pressure maintains a patent airway. **B,** During expiration, positive intrathoracic pressure collapses the airway because of loss of tracheal ring support. **C,** Computed tomography scan demonstrating thickening of the tracheal wall.

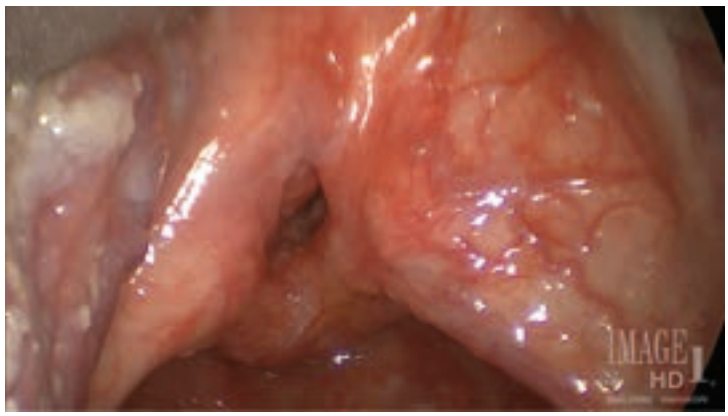

E-FIGURE 429-3. Supraglottic stenosis secondary to sarcoid demonstrating granulomatous destruction of epiglottis and aryepiglottic folds.

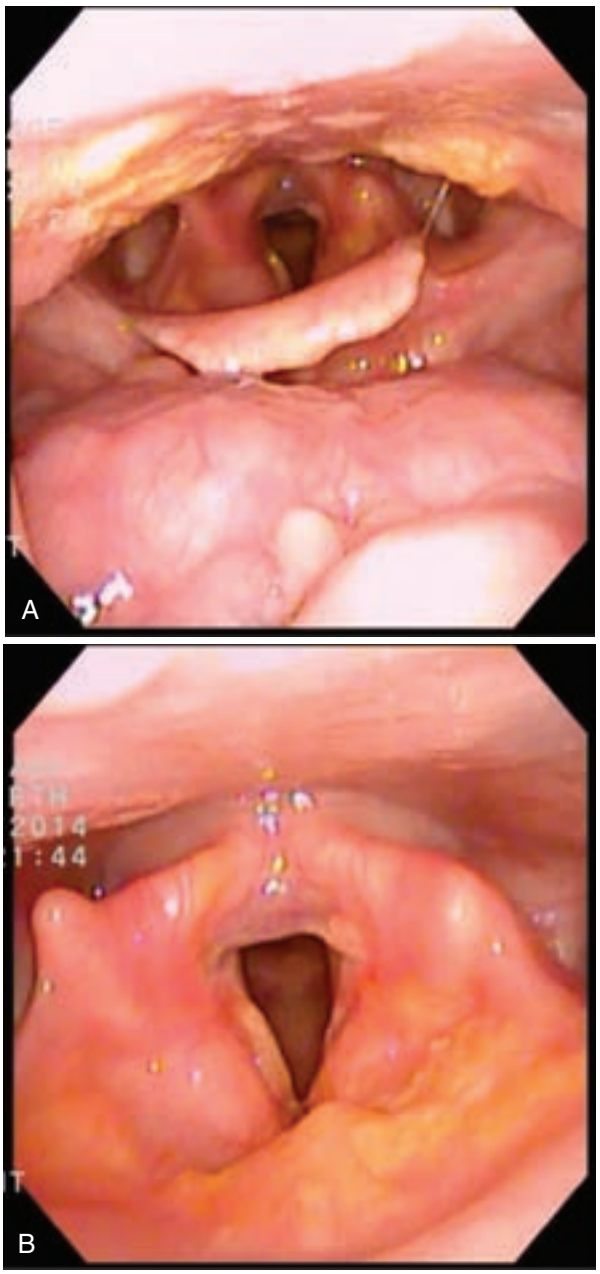

E-FIGURE 429-4. Amyloid deposition in the hypopharynx (A) and supraglottis and glottis (B).

Patients develop dysarthria, hypernasality, and a harsh strident strained vocal quality, which is more spastic than spasmodic.

Hypofunctional Neurologic Disorders

Parkinson's disease affects speech and swallowing in more than 80% of patients. Patients have dysarthria, prosody of speech, hypophonia, tachyphemia, monotonous pitch, and absence of vocal tremor.

Laryngeal examination shows bilateral vocal fold bowing with incomplete glottic closure. In advanced disease, vocal fold motion becomes hypokinetic. Pooling of secretions occurs as swallowing dysfunction progresses. Patients benefit from early intervention by speech-language therapists to address both voice and swallowing symptoms. In *progressive supranuclear palsy*, bulbar symptoms progress more rapidly, with pronounced speech and swallowing difficulty (Chapter 409). In *multiple system atrophy* (Chapter 409), progressive airway obstruction from bilateral vocal fold motion impairment may necessitate tracheostomy. In *myasthenia gravis* (Chapter 422), hypernasal speech, palatal weakness, and hypophonia may be accompanied by difficulties with swallowing and respiration. Some patients with *amyotrophic lateral sclerosis* (Chapter 419) present with bulbar symptoms that result from buccal-labial-lingual weakness, which produces speech and swallowing dysfunction, before the definitive diagnosis is made. In *multiple sclerosis* (Chapter 411), dysphonia and dysarthria are common.

LUMP IN THE THROAT

The sensation of a lump in the throat,[21] called *globus pharyngeus*, should prompt a careful history and physical examination by an otolaryngologist to exclude a serious diagnosis. When no underlying anatomic disease is found, possible causes include gastroesophageal reflux and spasm of the cricopharyngeus. Cognitive-behavioral therapy may be helpful.

HOARSENESS

Hoarseness, also called dysphonia, is characterized by altered voice quality, pitch, loudness, or vocal effort. Relevant history includes smoking status, occupation, and recent procedures involving the neck or affecting the recurrent laryngeal nerve. Symptomatic treatment should be considered in patients who have evidence suggesting a recent bacterial infectious process or gastroesophageal reflux without any historical features, such as ear pain, dysphagia, or odynophagia, or physical findings such as adenopathy or oral lesions suggestive of tumor. Additionally, there is no evidence to support the use of oral corticosteroids in patients with hoarseness. Hoarseness does not resolve within 3 months or the history or physical examination suggests a serious cause, prompt expert laryngoscopy is indicated.[22]

The symptom of hoarseness invariably points to the larynx as the site of disease. Benign lesions are most common, including vocal nodules (screamer's nodules), vocal cord cysts, vocal cord granulomas (usually resulting from intubation trauma or laryngeal hyperfunction), and vocal cord papillomas. Malignant neoplasms must be suspected (Chapter 190), especially in patients with a strong smoking history.

Of the malignant tumors that can occur in the hypopharynx and larynx, squamous cell carcinoma is the most common and is usually associated with tobacco and ethanol use. However, the incidence of human papillomavirus–related oropharyngeal squamous cell carcinoma is increasing and should be considered in the nonsmoking population. Squamous cell carcinoma (Chapter 190) can occur on essentially any mucosal surface in the head and neck. Symptoms can range from mild sore throat to hoarseness, severe dysphagia, and odynophagia. Pain is often referred to the jaw or ear. Associated cervical lymph node enlargement is also common in advanced disease. Successful management depends on early detection by a careful examination of the entire upper aerodigestive tract, biopsy with histopathologic examination, and aggressive treatment based on the clinical stage and site of the lesion.

LARYNGEAL PARALYSIS

Laryngeal paralysis most often is manifested as a unilateral paralysis as a result of a mediastinal tumor; surgical trauma during thyroid, carotid, or anterior cervical spine surgery; blunt or penetrating trauma; aortic aneurysm; progressive neurologic disease; or viral or idiopathic causes. The severity of impairment can be determined from subjective criteria based on the patient's symptoms, such as breathiness, aspiration, and exertional intolerance.

Unilateral vocal fold paralysis with a favorable prognosis occurs after blunt trauma, endotracheal intubation, idiopathic vocal fold paralysis, and paralysis associated with viral pathogens (Ramsay Hunt syndrome). In this setting, the severity of aspiration, dysphonia, and electromyographic findings can be used to determine the choice of procedure and timing of intervention. Temporary medialization with collagen injection is warranted in patients with unilateral paralysis and a good prognosis. Patients who have poor prognosis for recovery include those with injury after complete section of the nerve during a surgical resection of tumor, invasion of cranial nerves by a tumor, paralysis associated with thoracic aneurysm, or paralysis due to progressive neurologic disorders. In patients with a low likelihood of recovery, permanent medialization of the paralyzed vocal fold is warranted.

Bilateral vocal fold motion impairment, which is less common, has the same causes. Its management is most often directed toward improving the airway because the predominant symptom is airway obstruction.

Grade A References

A1. Spinks A, Glasziou PP, Del Mar CB. Antibiotics for sore throat. *Cochrane Database Syst Rev*. 2013;11:CD000023.

A2. van Driel ML, De Sutter AI, Keber N, et al. Different antibiotic treatments for group A streptococcal pharyngitis. *Cochrane Database Syst Rev*. 2013;4:CD004406.

A3. Jorgensen DM. Single-dose extended-release oral azithromycin vs. 3-day azithromycin for the treatment of group A beta-haemolytic streptococcal pharyngitis/tonsillitis in adults and adolescents: a double-blind, double-dummy study. *Clin Microbiol Infect*. 2009;15:1103-1110.

A4. Hayward G, Thompson MJ, Perera R, et al. Corticosteroids as stand alone or add-on treatment for sore throat. *Cochrane Database Syst Rev*. 2012;10:CD008268.

A5. Alho OP, Koivunen P, Penna T, et al. Tonsillectomy versus watchful waiting in recurrent streptococcal pharyngitis in adults: randomised controlled trial. *BMJ*. 2007;334:939.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Shulman ST, Bisno AL, Clegg HW, et al. Clinical practice guideline for the diagnosis and management of group A streptococcal pharyngitis: 2012 update by the Infectious Diseases Society of America. *Clin Infect Dis*. 2012;55:e86-e102.
2. Little P, Stuart B, Hobbs FD, et al. Antibiotic prescription strategies for acute sore throat: a prospective observational cohort study. *Lancet Infect Dis*. 2014;14:213-219.
3. Little P, Hobbs FD, Moore M, et al. Clinical score and rapid antigen detection test to guide antibiotic use for sore throats: randomised controlled trial of PRISM (primary care streptococcal management). *BMJ*. 2013;347:f5806.
4. Shah MD, Klein AM. Methicillin-resistant and methicillin-sensitive *Staphylococcus aureus* laryngitis. *Laryngoscope*. 2012;122:2497-2502.
5. Bock JM, Burtis CC, Poetker DM, et al. Serum immunoglobulin G analysis to establish a delayed diagnosis of chronic cough due to *Bordetella pertussis*. *Otolaryngol Head Neck Surg*. 2012;146:63-67.
6. Little P, Stuart B, Hobbs FD, et al. Predictors of suppurative complications for acute sore throat in primary care: prospective clinical cohort study. *BMJ*. 2013;347:f6867.
7. Saluja S, Brietzke SE, Egan KK, et al. A prospective study of 113 deep neck infections managed using a clinical practice guideline. *Laryngoscope*. 2013;123:3211-3218.
8. Gupta N, Kralovic SM, McGraw D. Lemierre syndrome: not so forgotten! *Am J Crit Care*. 2014;23:176-179.
9. Dowd JB, Palermo T, Brite J, et al. Seroprevalence of Epstein-Barr virus infection in U.S. children ages 6-19, 2003-2010. *PLoS OnE*. 2013;8:e64921.
10. Yun HC, Fugate WH, Murray CK, et al. Pandemic influenza virus 2009 H1N1 and adenovirus in a high risk population of young adults: epidemiology, comparison of clinical presentations, and coinfection. *PLoS OnE*. 2014;9:e85094.
11. Benwill JL, Sarria JC. Laryngeal tuberculosis in the United States of America: a forgotten disease. *Scand J Infect Dis*. 2014;46:241-249.
12. Kutluhan A, Salviz M, Yalciner G, et al. The role of the actinomyces in obstructive tonsillar hypertrophy and recurrent tonsillitis in pediatric population. *Int J Pediatr Otorhinolaryngol*. 2011;75:391-394.
13. Isakson M, Hugosson S. Acute epiglottitis: epidemiology and *Streptococcus pneumoniae* serotype distribution in adults. *J Laryngol Otol*. 2011;125:390-393.
14. Patcharatrakul T, Gonlachanvit S. Gastroesophageal reflux symptoms in typical and atypical GERD: roles of gastroesophageal acid refluxes and esophageal motility. *J Gastroenterol Hepatol*. 2014;29:284-290.
15. Hoppo T, Komatsu Y, Jobe BA. Antireflux surgery in patients with chronic cough and abnormal proximal exposure as measured by hypopharyngeal multichannel intraluminal impedance. *JAMA Surg*. 2013;148:608-615.
16. Robati RM, Rahmati-Roodsari M, Dabir-Moghaddam P, et al. Mucosal manifestations of pemphigus vulgaris in ear, nose, and throat; before and after treatment. *J Am Acad Dermatol*. 2012;67:e249-e252.
17. Fortuna G, Mignogna MD. Clinical guidelines for the use of adjuvant triamcinolone acetonide injections in oro-pharyngeal pemphigus vulgaris: the oral medicine point of view. *J Oral Pathol Med*. 2011;40:359-360.
18. Taylor SC, Clayburgh DR, Rosenbaum JT, et al. Progression and management of Wegener's granulomatosis in the head and neck. *Laryngoscope*. 2012;122:1695-1700.
19. Hong G, Kim H. Clinical characteristics and treatment outcomes of patients with relapsing polychondritis with airway involvement. *Clin Rheumatol*. 2013;32:1329-1335.
20. Sinclair CF, Gurey LE, Blitzer A. Palatal myoclonus: algorithm for management with botulinum toxin based on clinical disease characteristics. *Laryngoscope*. 2014;124:1164-1169.
21. Foden N, Ellis M, Shepherd K, et al. A feeling of a lump in the throat. *BMJ*. 2014;348:f7195.
22. Schwartz SR, Cohen SM, Dailey SH, et al. Clinical practice guideline: hoarseness (dysphonia). *Otolaryngol Head Neck Surg*. 2009;141:S1-S31.

430 PRINCIPLES OF MEDICAL CONSULTATION

GERALD W. SMETANA

APPROACH TO MEDICAL CONSULTATION

A general medical or subspecialty medical physician may receive a request to perform a consultation for a variety of purposes. In some settings, a single consultative encounter will be requested, or the consultant will determine that only one visit, either in the inpatient or in the outpatient setting, is necessary. More commonly, the approach will include one or more follow-up visits to meet the goals of the consultation from the perspectives of the requesting physician, patient, and consultant. The consultant generally assumes one of four roles: cognitive consultant, procedural consultant, comanager with shared care, or comanager with principal care. The consultant as a comanager continues to care for a component of the patient's needs in an ongoing fashion while being careful to coordinate this comanagement with the continuing role of the requesting physician. Finally, in some situations, it may be most appropriate for the physician who initially requested the consultation no longer to play an active role in the care of the patient but rather to transfer ongoing care exclusively to the consultant.

From a practical perspective, consulting medical physicians, whether they are generalists or subspecialists, enter into the consultative mode in a relatively limited number of ways. Surgeons may request a preoperative medical consultation to assess operative risk and obtain recommendations regarding perioperative care (Chapter 431) or, after surgery, to seek help in managing specific postoperative complications or assisting in the patient's long-term management. Both general medical physicians and medical subspecialists appropriately seek help from a subspecialist with particular knowledge in problems outside their own area of expertise to reduce uncertainty. Sometimes these requests are for specific medical procedures, but requests often seek cognitive guidance as well. Finally, noninternists may seek medical consultation for reasons other than perioperative care. For example, a psychiatrist may request a consultation to help determine whether the somatic symptoms of a particular patient represent important medical conditions (Chapter 434). In the peripartum setting, specific complications of pregnancy may require sophisticated medical consultation (Chapter 239).

Each of these settings raises different challenges for the medical consultant. In all settings, however, a number of general principles apply and can improve the effectiveness of consultations. Communication between physicians and other team members is critical to the process of consultation in all settings. The burgeoning patient safety movement has developed best practices for "hand-offs" that can be applied to consultations.

SETTING-SPECIFIC CONSULTATIVE ISSUES

An effective consultant must recognize the setting in which the consultation is requested and possess the required content knowledge. Distressingly, a number of studies have demonstrated that the requesting physician and the consultant often have different views on the reasons that a consultation was requested, and this initial disconnection, if present, will doom any medical consultation.

Preoperative Surgical Consultation

In the preoperative setting, the medical consultant should not "clear" a patient for surgery and must avoid the temptation to do so even if asked. Clearance may incorrectly imply that the procedure has no risk or that the medical consultant will take responsibility for having misled the patient and surgeon. Instead, the medical consultant should help determine the inherent risk associated with the proposed procedure for the particular patient, whether the patient is in the best possible condition for surgery, and whether any generic or patient-specific interventions would reduce the risk (Chapter 431). Effective consultation also requires a specialized knowledge base, whether the consultation is focusing on a particular organ system or on overall perioperative risk (Chapter 431). A poor consultant-patient interaction can have a substantial negative effect on a patient's confidence in the planned therapy.[1]

Postoperative Surgical Consultation

Surgeons typically request a postoperative medical consultation when a complication has developed that is beyond their area of expertise (Chapter 433). These problems are commonly urgent, so the goal is expeditious consultation and prompt intervention. Management of these problems does not usually differ from management in nonoperative settings. Another reason for a postoperative consultation is to obtain assistance in post-hospitalization care or to facilitate seamless discharge planning. Consultants should aid in the transition to the outpatient or long-term care setting by taking primary or consultative roles as appropriate.

Medical-Medical Consultations

Cross-consultations between medical subspecialists or between a subspecialist and a generalist, in either direction, are quintessential examples of collaboration. For subspecialty consultations, the key is to provide the requested expertise without overstepping into the domain of the expertise of the requesting physician. A growing example of medical-medical consultation is the situation in which a hospitalist assumes principal responsibility for an inpatient admission and then returns the patient to the primary care physician after discharge from the hospital. In this consultative interaction, close communication is critical because primary responsibility for the patient's care has shifted from the outpatient medical physician to the inpatient physician. This transfer of responsibility is not dissimilar to that occurring when a patient is submitted to the care of a subspecialist for a procedure such as cardiac catheterization or gastrointestinal endoscopy. Similar issues also arise when a critically ill patient with a condition such as a complicated myocardial infarction (Chapter 73) or shock (Chapters 106, 107, and 108) is managed principally by a critical care medical specialist and is then expected to return to the care of the primary physician after hospital discharge. However, a key difference is that the inpatient hospitalist physician, unlike the consulting subspecialist, will not typically have an ongoing comanagement role. Because of the higher risk for discontinuity, effective communication at the time of hospitalization, whenever key issues arise during hospitalization, and at the time of discharge, is even more important in the inpatient hospitalist model than in the other settings in which subspecialists may take on more of a comanagement role. Effective and comprehensive handoffs at the time of hospital discharge improve continuity and reduce the potential for errors and medicolegal liability.

Rapid Response Teams

A recent quality improvement initiative has been the development of rapid response teams. These teams aim to reduce the "failure to rescue," which often precedes an unplanned intensive care unit (ICU) transfer or non-ICU cardiac arrest. In this model, a prespecified consultative team urgently sees sick, hospitalized patients when a "trigger" abnormality indicates potential impending serious complications. The activation triggers for rapid response teams have varied somewhat among institutions, but there is substantial agreement on what constitutes an appropriate trigger (Table 430-1). Rapid response teams differ from traditional "code blue teams" in several important aspects, the most important of which is their goal to rescue patients before a crisis situation occurs (Table 430-2).

Although implementation of these teams has not consistently reduced hospital mortality, most trials have shown a reduction in unplanned ICU transfers and in hospital length of stay. For example, in a trial of 5391 patients at a single academic medical center, patients randomly assigned to a rapid response team had a shorter length of stay, fewer unplanned ICU transfers, and lower mortality in patients with unplanned ICU transfers.[■] Other studies have shown a greater impact on unplanned ICU transfers and a reduction in hospital cardiac arrest rates.[2] In a systematic review, rapid response teams significantly reduced non-ICU cardiac arrests and total in-hospital adult mortality.[3]

This consultative strategy crosses specialties and may include general internal medicine, hospital medicine, and critical care physicians, as well as respiratory therapists and ICU nurses. When an emergency consultation is generated by a trigger event, the consultant's relationship is primarily with the patient rather than with the referring physician.

Consultations for Special Populations

When consulting for psychiatrists or in the peripartum period, the consultant requires special expertise to understand the different expressions of signs and symptoms in specific populations, as well as how and when to modify typical

TABLE 430-1 TRIGGERS: CRITERIA FOR MOBILIZING AN IN-HOSPITAL RAPID RESPONSE SERVICE

VITAL SIGNS
Heart rate • Heart rate <40 beats/min, especially if symptoms • Heart rate >140 beats/min Blood pressure • Systolic blood pressure <90 mm Hg or >30-40 mm Hg below patient's usual stable blood pressure • Systolic blood pressure >200 mm Hg for >30 min • Diastolic blood pressure >110 mm Hg with symptoms Respiratory rate • Respiratory rate <8 breaths/min or >35 breaths/min • New onset of marked dyspnea, compromised airway, cyanosis Oxygenation • O_2 saturation <85% for >5 min (except patients with chronic severe hypoxemia) • Need for 100% supplemental O_2 or a non-rebreathing O_2 mask Temperature • Body temperature >39° C or associated with acute decompensation
NEUROLOGIC STATUS
• Acute change in mental status • New focal findings • Prolonged or repeated seizures • ≥2 Point decline in Glasgow coma scale score
GENERAL STATUS
• Uncontrollable bleeding • Decreased urine output to <50 mL over 4 hr

Based on criteria of The Joint Commission and on other sources.

TABLE 430-2 FEATURES OF RAPID RESPONSE TEAMS AND TRADITIONAL CODE TEAMS

FEATURE	TRADITIONAL CODE TEAM	RAPID RESPONSE TEAM
Criteria for calling team	No pulse, blood pressure, or respiratory effort; unresponsive	Low blood pressure, rapid heart rate, respiratory distress, change in mental status
Typical conditions	Cardiac arrest, respiratory arrest, airway obstruction	Sepsis, pulmonary edema, arrhythmias, respiratory failure
Typical team composition	Anesthesia fellow, ICU fellow, internal medicine house staff, ICU nurse	ICU fellow, ICU nurse, respiratory therapist, internal medicine house staff
Typical call rate (number per 1000 admissions)	0.5-5	20-40
In-hospital mortality	70-90%	0-20%

ICU = intensive care unit.
Adapted from Jones DA, DeVita MA, Bellomo R. Rapid-response teams. *N Engl J Med.* 2011;365:139-146.

medical recommendations because of special circumstances (Chapters 239 and 434). Although these consultations occasionally result in long-term comanagement, more commonly they revolve around the resolution of an isolated problem. In addition, the medical consultant should not make assumptions about the requesting physician's knowledge base regarding the specific medical issues that generate the consultation. The written report and verbal communications should be comprehensive and commonly include more basic medical content and specific recommendations than may be typical of a consultation for a medical subspecialist.

Subspecialty consultations requested by primary care physicians may be for advice, a technical procedure, or ongoing comanagement (Table 430-3). Consultations for general practitioners or family physicians generally follow

TABLE 430-3 REASONS TO CONSIDER A SUBSPECIALTY CONSULTATION

To provide ongoing comanagement: become a partner in inpatient care • For an acute, unstable problem • For a chronic condition
To perform a procedure (or advise on whether to perform it) • Diagnostic • Therapeutic
To provide one-time or periodic advice • Diagnostic guidance • Therapeutic guidance • Reassurance • Medicolegal issues

the same guidelines as those for general internists. However, because both the training and ongoing practices of such physicians may or may not include the same spectrum and complexity of disease encountered by typical general internists, the subspecialty consultant must use judgment regarding both the initial consultation and the advisability of ongoing comanagement.

STRATEGIES FOR EFFECTIVE CONSULTATION

It is critical that requesting and consulting physicians agree about the reason for the consultative request. For example, the nature of a preoperative medical consultant's evaluation would differ substantially if the request were (1) routine, (2) for advice on perioperative insulin management in a patient with type 1 diabetes, or (3) to assist in determining the risks and benefits of proceeding to vascular surgery in a high-risk patient with coronary artery disease and a previous stroke. Effective communication at the time of the request will improve the value of the consultation and clarify the question. When there is doubt regarding the reason for the consultation, the consultant should speak directly to the referring physician before completing the evaluation of the patient.

Common reasons for medical consultation include assessment of perioperative risk; interpretation of a laboratory abnormality; help in performing a procedure, in obtaining advice, or in selecting therapy; and aid in providing long-term care. Prophylactic strategies (venous thromboembolism, endocarditis, and surgical site infection) infrequently generate requests for medical consultation because they are commonly standardized to conform to practice guidelines at individual institutions.

Determining the question will also help narrow the scope of the consultant's advice and minimize the number of recommendations. Adherence by the requesting physician to any of the recommendations made by the consultant is higher for consultations with fewer recommendations. Another strategy to minimize the number of recommendations is to restrict advice to pertinent issues at the time, preferably in order of their importance. For example, a consultant who is asked to aid in the care of a critically ill pregnant patient with HELLP syndrome (hemolysis, elevated liver enzymes, and low platelet count) (Chapters 160 and 239) should not also make recommendations regarding the value of cigarette cessation (Chapter 32), even though this issue will be pertinent after the mother and child have survived the acute event.

Requesting physicians are also more likely to adhere to recommendations when the patient is sicker, when the consultation is performed promptly, when advice is given to institute specific therapy rather than perform more diagnostic testing, when the consultant writes frequent follow-up notes, and when computerized order entry support systems are used to convey recommendations. When giving advice about medications, consultants should indicate specific doses and duration of treatment. If a recommendation is likely to be controversial (e.g., postpone surgery), it is always preferable to speak directly to the requesting physician before writing a consultation note in the hope that direct conversation will provide an opportunity to develop a consensus that may then be reflected in the formal consultation note. In one study of 323 physicians, the three elements of greatest perceived importance for a high-quality consultation request were to frame the question, to indicate clearly whom to call with the response, and to establish urgency.[4]

Consultants should be careful to restrict their advice to their particular area of expertise. For example, unless a particular antipsychotic medication is contraindicated because of a medical issue (Chapter 434), it is wisest to defer decision making regarding psychiatric management to the psychiatrist. A

TABLE 430-4 GUIDELINES TO HELP MAKE CONSULTATIONS EFFECTIVE

Determine whether the goal is to obtain advice (and, if so, specifically for what) or to aid in ongoing comanagement (see Table 430-2).
Understand the urgency of the consultation so that the consultation will be performed on a timely basis that meets the needs of the requesting physician and patient.
Perform a focused but careful history and physical examination—do not rely on information gathered by others.
Do not rehash information in an overly detailed note—emphasize the key issues your evaluation reinforced or discovered.
Be sure your recommendations are clearly listed and appropriately detailed—for example, indicate specific drugs, doses, and durations.
Limit the number of total number of recommendations to improve adherence.
Indicate how to monitor the effectiveness of your recommendations, as well as how to contact you urgently if problems arise.
Adjust your involvement (advice vs. comanagement) as appropriate, in concert with the requesting physician.
Serve as a peer—teach as appropriate but also seek to learn.
Remember that personal contact with the requesting physician, even if very brief, may be far more helpful than the best of written notes.
Do not disappear—follow the patient as frequently as appropriate, in coordination with the requesting physician.

For a more complete discussion of this topic, see Goldman L, Lee T, Rudd P. Ten commandments for effective consultations. *Arch Intern Med*. 1983;143:1753-1755 and Salerno SM, Hurst FP, Halvorson S, et al. Principles of effective consultation: an update for the 21st-century consultant. *Arch Intern Med*. 2007;167:271-275.

strongly worded consultation note that advises against a particular strategy will put another physician in a difficult medicolegal position if it differs from the physician's usual or recommended practice. The most important attributes of a consultative role are simple, concise recommendations and a clearly stated rationale for decision making. Detailed differential diagnosis is less important, and literature support is usually unnecessary.

Timely consultation reports improve physician satisfaction and patient care. For example, primary care physicians note that failure to receive timely reports from consultants limits their ability to provide high-quality care.[5]

Salerno and colleagues have proposed modifications of Goldman's original "Ten Commandments" for effective consultations (Table 430-4). These recommendations serve as a helpful guide for consultants to improve adherence to their advice and, as a result, improve patient outcomes. Specific interactive training on the principles of consultation can increase the effectiveness of consultative communication.

SPECIAL CONSULTATIVE SITUATIONS

Curbside Consultation

Informal consultations are commonly called "curbside" consultations. In the current era, ready access to online medical references may potentially reduce curbside requests. Consultants in "cognitive" specialties such as infectious diseases, rheumatology, and endocrinology provide a disproportionate amount of informal (compared with formal) consultations. Informal consultations may occur by telephone, by e-mail, or in person. Curbside consultation is ingrained into the fabric of medical care.

Both generalists and specialists participate in an average of three to four curbside consultations per week. These consultations commonly involve questions about diagnostic tests, treatment plans, and the potential value of a formal consultation (which follows about one third of curbside consultations) in the future. To be effective, curbside consultations should be brief, involve a single question, and require no direct examination of the patient or medical records.

An important limitation of curbside consultations is that the consultant must rely on limited, secondhand information from the requesting physician rather than primary data from direct evaluation of the patient. Consultants report that such indirect information is inaccurate in as many as half of curbside consultations.[6] This reality, along with lack of financial compensation, leads to a greater degree of dissatisfaction with the curbside process by consultants than by requesting physicians. Salaried consultants may view a request for a curbside consultation more favorably than might those whose livelihood depends on a fee-for-service model of care.

Although requesting physicians may perceive reduced medicolegal risk when they obtain and even document a curbside consultation, consultants may fear the risk for malpractice liability themselves when offering such advice. However, courts have consistently found that curbside consultants have no liability because no direct physician-patient relationship exists. Rather, the relationship is only between the requesting and consulting physicians.

Electronic Consultations

Increasingly, physicians are using electronic communication, either as part of a shared electronic medical record or by e-mail, to supplement formal consultation requests.[7] An obvious advantage of this approach is that the requesting and consulting physician can communicate on their own schedules. Potential benefits include saving time, reducing costs, and improving continuity and access to specialty care. However, the lack of reimbursement for such activities may be a barrier to e-consultations.

Mandatory Consultations

In some situations, mandatory consultations may help enforce a standard of care. For example, mandatory inpatient infectious disease consultations, as part of an antimicrobial stewardship program, can improve the rational use of antimicrobial agents in the hospital and after discharge. Similarly, instituting routine consultation for patients with certain sentinel conditions, such as diabetes, may improve outcomes.

Comanagement

A consultation that begins with an initial encounter or a limited number of follow-up visits may evolve into ongoing comanagement. In such an arrangement, the physician initially serving in a consultative role becomes at least coequal to the requesting physician in the provision of ongoing care. In some situations, the consultant may actually become the primary physician. This arrangement is obvious in situations in which the consultation was requested specifically for the provision of ongoing care. Other examples include situations in which an oncologist assumes principal care of a patient with a malignancy or a nephrologist assumes principal care of a patient with end-stage renal disease who is being maintained on dialysis. In some of these situations, a general internist who initially requested the consultation may now become the consultant and provide advice on preventive care and occasional help with intercurrent medical problems.

Comanagement is also increasingly common in the hospital setting. Medical consultants may become comanagers of postoperative surgical patients, with a potential for improving outcomes. In an early randomized trial of perioperative patients, formal comanagement, in which a medical physician took responsibility for managing medical problems rather than acting in a consultative role, reduced postoperative complications; in addition, both nurses and surgeons preferred this model. However, not all studies have shown that comanagement improves patient outcomes. In one study of comanagement on a neurosurgical service, for example, nursing staff perceived substantial improvement in the quality of patient care and costs were reduced, but there were no differences in mortality, readmission rates, or length of stay.[8] In some institutions, comanagement for orthopedic or other surgical patients has become routine and does not require a specific request for consultation. Internists currently comanage more than one third of surgical inpatients in some hospitals.

In some settings, the outpatient primary care physician may retain a comanagement role even as a cardiologist cares for a patient with acute myocardial infarction, a pulmonologist cares for a patient in the ICU, a hospitalist cares for a patient in the general medical setting, or a noninternist addresses a specific problem.

RESPONSIBILITIES OF THE CONSULTANT

The consultant is responsible to two parties: the patient and the referring physician. A paternalistic approach is not desirable. The consultant should not limit communication to the referring physician and must not withhold discussion and recommendations from the patient. However, the consultant should not usurp the role of the referring physician, who remains responsible for assembling information and advice from varied sources, as well as for developing an integrated plan with the patient. For example, in the preoperative setting, the consultant should not express a final opinion regarding the suitability of proceeding to surgery without first discussing all relevant considerations with the referring surgeon. In the trilateral deliberative model, the patient, referring physician, and consultant each have responsibilities and constraints inherent in their relationships (Fig. 430-1). Above all, the role of the consultant is to improve patient care and outcomes.

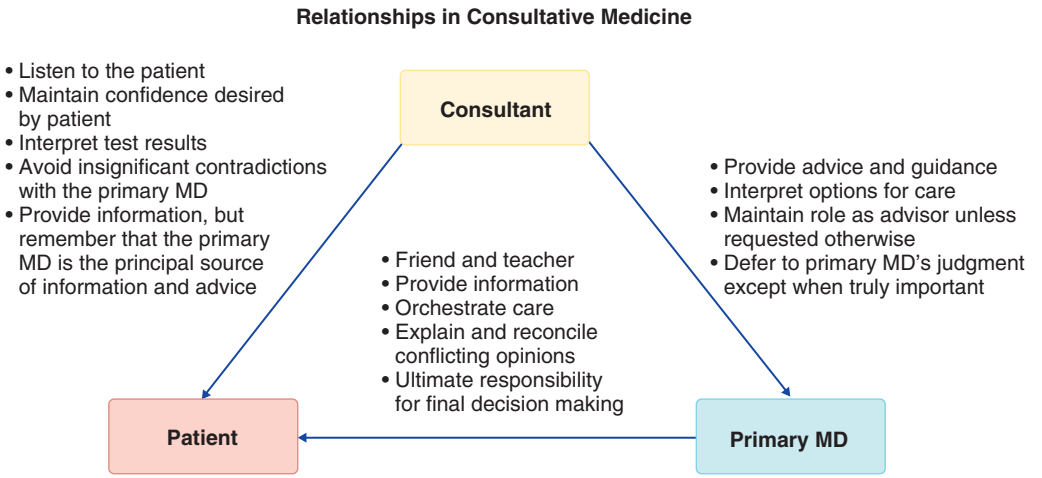

FIGURE 430-1. Consultative relationships. The consultant must execute his or her duty to the patient while acknowledging and hopefully conforming to the collegial professional model of advising the primary physician. If, however, the patient's safety or welfare is at stake, the consultant's duty is to be sure that the patient's needs always take precedence. (Based on Emanuel LL, Richter J. The consultant and the patient-physician relationship: a trilateral deliberative model. *Arch Intern Med.* 1994;154:1785-1790, and other sources.)

IMPACT OF CONSULTATIONS ON PATIENT OUTCOME

Few controlled trials have investigated the impact of medical consultations on outcomes. In one small study in the 1990s, an outpatient preoperative medical consultation reduced unnecessary admissions (those that did not result in surgery) compared with usual care (inpatient consultation at the discretion of the surgeon). Other reports indicate that anywhere from 5 to 50% of preoperative medical consultations result in changes in patient management. However, any potential benefit of preoperative medical consultation for improving important outcomes remains unproved.[9]

Grade A References

A1. Rothschild JM, Woolf W, Finn KM, et al. A controlled trial of a rapid response system in an academic medical center. *Jt Comm J Qual Patient Saf*. 2008;34:417-425.

A2. Kessler CS, Afshar Y, Sardar G, et al. A prospective, randomized, controlled study demonstrating a novel, effective model of transfer of care between physicians: the 5 Cs of consultation. *Acad Emerg Med*. 2012;19:968-974.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Greville-Harris M, Dieppe P. Bad is more powerful than good: the nocebo response in medical consultations. *Am J Med*. 2015;128:126-129.
2. Salvatierra G, Bindler RC, Corbett C, et al. Rapid response team implementation and in-hospital mortality*. *Crit Care Med*. 2014;42:2001-2006.
3. Winters BD, Weaver SJ, Pfoh ER, et al. Rapid-response systems as a patient safety strategy: a systemic review. *Ann Intern Med*. 2013;158:417-425.
4. Boulware DR, Dekarske AS, Filice GA. Physician preferences for elements of effective consultations. *J Gen Intern Med*. 2010;25:25-30.
5. O'Malley AS, Reschovsky JD. Referral and consultation communication between primary care and specialist physicians: finding common ground. *Arch Intern Med*. 2011;171:56-65.
6. Burden M, Sarcone E, Keniston A, et al. Prospective comparison of curbside versus formal consultations. *J Hosp Med*. 2013;8:31-35.
7. Horner K, Wagner E, Tufano J. Electronic consultations between primary and specialty care clinicians: early insights. *Issue Brief (Commonw Fund)*. 2011;23:1-14.
8. Auerbach AD, Wachter RM, Cheng HQ, et al. Comanagement of surgical patients between neurosurgeons and hospitalists. *Arch Intern Med*. 2010;170:2004-2010.
9. Wijeysundera DN, Austin PC, Beattie WS, et al. Outcomes and processes of care related to preoperative medical consultation. *Arch Intern Med*. 2010;170:1365-1374.

REVIEW QUESTIONS

1. Your hospital seeks your advice on the development of a comanagement model between internal medicine and the inpatient orthopedic service. Which of the following strategies will provide the most optimal comanagement?

A. Internists and orthopedists round together in the morning along with nursing staff.
B. Orthopedists write all orders.
C. Internists consult only when a postoperative medical complication occurs.
D. Primary care physician determines whether comanagement is necessary.

Answer: A Comanagement is a strategy in which all patients on a particular service, usually a surgical service such as orthopedics, receive a mandatory internal medicine consult. In most instances, hospitalists provide this service. The intent is to provide optimal postoperative care in order to prevent, rather than simply to treat, complications. Both the consulting internist and the surgeon may write orders. Optimally, both services round together in the morning to identify patient management issues and to establish the daily plan along with the assistance of the nursing staff.

2. Rapid response teams exist to prevent "failure to rescue" on inpatient non–intensive care unit (ICU) services. Which of the following represents an appropriate trigger to call a rapid response team?

A. Systolic blood pressure of 110 mm Hg
B. Oxygen saturation of 82% on room air
C. Temperature of 101.5° F
D. More than 5 ectopic ventricular beats per minute

Answer: B Rapid response teams are an example of a consultation team. Their purpose is to prevent "failure to rescue," reduce unplanned ICU transfers, and hopefully reduce in-hospital mortality. Certain triggers automatically initiate a request for consultation. The most important of these are critical vital sign values and respiratory distress. In this example, oxygen desaturation is a trigger. Fever alone warrants clinical evaluation but is not a trigger in the absence of other features.

3. Which of the following strategies does *not* improve adherence to a consultant's recommendations?

A. Limit recommendations to less than five
B. Provide contingency plans for potential clinical scenarios
C. Make a direct call to requesting physician to discuss care plans
D. Provide literature references
E. Determine the question

Answer: D Substantial literature provides guidance on strategies to improve adherence to a consultant's recommendations and thus to improve patient care. Consultations that are concise, with a limited number of key recommendations, are most likely to be followed. Contingency plans outline steps that the requesting doctor should take in the event of certain changes in clinical condition. It is critical to understand the question. If the requesting and consulting physician do not agree on or understand the question, the consult is doomed to be unhelpful. Requesting physicians are generally uninterested in literature references to support the recommendations of the consultant.

4. When performing a preoperative medical consultation, the consultant identifies risk factors for postoperative medical complications. The consultant should do which of the following?

A. Clear the patient for surgery
B. Advise the patient to forgo surgery
C. Determine the impact of the risk factor and propose strategies to reduce the risk
D. Request another consultation from a subspecialist

Answer: C A consultant should never "clear" a patient before surgery. Such a message may imply zero risk and fails to consider the impact of particular risk factors. Rather, the appropriate approach is to determine what factors may increase risk, estimate overall risk, and then propose strategies to reduce risk. The goal is for the patient to be in the best possible condition before surgery. The consulting medical physician's primary responsibility is to the requesting physician. The consultant should not directly advise the patient to forgo surgery. Rather, the consultant should share any concerns with the requesting surgeon and develop a collaborative plan. If the consultant determines that a subspecialty consultation is warranted, he or she should suggest that to the surgeon but should not directly obtain another consultation without first obtaining agreement from the surgeon.

5. Your primary care colleague asks you for a curbside consultation in infectious disease. You have not met the patient. Your consultation will achieve which of the following?

A. Improve patient care
B. Expose you to unnecessary medicolegal risk
C. Be based on an accurate and complete patient history
D. Generate income as an evaluation and management (E&M) activity

Answer: A Curbside consultations are a common and important activity for consultants, particularly in "cognitive" specialties in which the consultant does not commonly perform procedures. The consultant has no doctor-patient relationship. Therefore, most courts have held that there is no medicolegal risk to the consultant. When a formal consultation follows, it is often true that additional elements emerge from the history and physical examination. Despite this limitation, curbside consultations improve both access to subspecialty care and patient care. Unfortunately, curbside consultations are not generally billable activities; however, approximately one third of curbside consultations result in a request for a formal, billable consultation.

431

PREOPERATIVE EVALUATION

STEVEN L. COHN

Each year in the United States, more than 25 million inpatient surgical procedures and an additional 25 million outpatient procedures are performed. Although more than one third of these surgical patients are older than 65 years, overall morbidity and mortality are relatively low, in part because of modern anesthetic and surgical techniques. A crucial aspect of safety is careful preoperative evaluation of the patient not only by the surgeon and anesthesiologist but also, in many instances, by a general medical consultant or medical subspecialist.

OPERATIVE RISK ASSESSMENT

The components of perioperative risk include those related to the patient, procedure, provider, and anesthesia. Anesthetic risk is low, with mortality less than 0.03% in a normal healthy patient—American Society of Anesthesiology (ASA) class 1—but increasing to 0.2% in ASA class 2 (mild systemic disease), 1.2% in class 3 (severe systemic disease), 8% in class 4 (severe systemic disease that is a constant threat to life), and 34% in class 5 (a moribund patient not expected to survive for 24 hours without surgery). Meta-analysis suggests that when feasible, neuraxial (spinal or epidural) anesthesia may reduce postoperative complications compared with general anesthesia (Chapter 432), but decisions regarding the anesthetic technique should be the responsibility of the anesthesiologist and not be part of the preoperative medical consultation. With respect to the provider, data support a "learning curve," with better outcomes when procedures are performed by more experienced, higher-volume surgeons.

GENERAL RISK ASSESSMENT

History and Physical Examination

The medical history and physical examination are the most important components in assessing a patient's risk for surgery. The consultation should focus on pertinent medical problems, particularly cardiopulmonary symptoms and diseases that are associated with risk and are likely to influence perioperative management (Chapter 430). The importance of the past surgical history is to determine whether the patient was able to undergo major surgery in the recent past or had any perioperative medical or anesthetic-related complications that could occur again. The social history should assess and quantify the amount, duration, and last use of tobacco, alcohol, or illicit substances. It is important to document allergies to medications, foods, and latex as well as to obtain an accurate list of the patient's current prescription and over-the-counter medications, including doses and adherence. The family history is relevant primarily for any genetically associated complications such as malignant hyperthermia or a bleeding disorder. The review of systems should include the presence or absence of chest pain and dyspnea and the patient's exercise capacity. The physical examination must include the vital signs, assessment of the airway and respiratory status, cardiovascular examination, and documentation of any neurologic deficit.

Preoperative Tests

Screening preoperative test results in otherwise healthy individuals are usually normal and, even when abnormal, rarely affect management (generally <1%) (Table 431-1).[1] Most significant abnormalities can be predicted from the clinical information obtained, which then guides selective testing based on the history, the physical findings, and the planned type of surgery and anesthesia. Most patients undergoing low-risk surgery with local anesthesia require no preoperative testing. Repeat testing should be avoided if recent (within 3 months) results were normal, unless the patient's condition or medications have changed.

Perioperative Medications

Decisions regarding whether to continue a medication perioperatively should consider the drug's pharmacokinetics (Chapter 29) as well as its effects on the primary disease and perioperative risk, including potential interactions with anesthetic agents. Some medications are essential to continue (e.g., cardiac medications and corticosteroids), whereas others must be

TABLE 431-1 RECOMMENDATIONS FOR PREOPERATIVE LABORATORY TESTING

TEST	% ABNORMAL	% INFLUENCING MANAGEMENT	INDICATIONS
Hemoglobin	1.8	0.1	Expected major blood loss, symptoms of anemia, chronic kidney disease
White blood count	0.7	0.0	Suspected infection, myeloproliferative disorder, myelotoxic medications
Platelet count	0.9	0.02	Bleeding diathesis, myeloproliferative disorder, myelotoxic medications
Prothrombin time/INR	0.3	0.0	Bleeding diathesis, liver disease, malnutrition, antibiotic use, anticoagulants
Partial thromboplastin time	6.5	0.1	Bleeding diathesis, anticoagulant use
Electrolytes	12.7	1.8	Renal disease, medications affecting electrolytes (e.g., diuretics, digoxin, ACE inhibitor, ARBs)
Glucose	9.3	0.5	Known DM, steroids, morbid obesity
Renal function	8.2	2.6	Renal disease, DM, HTN, major surgery, older age, medications affecting renal function
Liver function tests	0.4	0.1	Known liver disease, albumin level if at risk for needing postoperative parenteral nutrition
Urinalysis	19.1	1.4	No indication unless GU symptoms or instrumentation planned (although often requested before joint replacement or spine surgery)
Electrocardiogram (<50 years old)	29.6 (19.7)	2.6	Vascular surgery; intermediate risk surgery with at least 1 RCRI risk factor; not indicated in asymptomatic patients undergoing low-risk procedures or solely based on age
Chest radiograph (<50 years old)	21.2 (4.9)	3.0	Acute cardiopulmonary disease suspected based on history and physical examination; history of stable chronic cardiopulmonary disease in patient older than 70 years without a chest radiograph in the past 6 months.

ACE = angiotensin-converting enzyme; ARB = angiotensin receptor blocker; DM = diabetes mellitus; GU = genitourinary; HTN = hypertension; INR = international normalized ratio; RCRI = Revised Cardiac Risk Index (1 point each for coronary artery disease; heart failure; prior cerebrovascular accident or transient ischemic attack; diabetes mellitus on insulin; creatinine >2.0; and abdominal, thoracic, or suprainguinal vascular surgery) (see Table 431-3).
Modified from Smetana GW, Macpherson DS. The case against routine preoperative laboratory testing. *Med Clin North Am.* 2003;87:7-40.

discontinued (e.g., oral hypoglycemic agents) or have their dose altered (e.g., insulin and anticoagulants). Still other medications should be started prophylactically to minimize perioperative risk (e.g., anticoagulants for prophylaxis against venous thromboembolism [Chapter 81] and antibiotic prophylaxis for surgical site infection or endocarditis [Chapter 76]). Data are often lacking or conflicting. Table 431-2 briefly summarizes "consensus" perioperative recommendations for the major classes of drugs.

CARDIAC RISK ASSESSMENT

A significant proportion of patients who undergo surgery have either known coronary artery disease or risk factors for it, and postoperative cardiac complications are second only to direct surgical complications as a cause of perioperative mortality. The goal is to risk-stratify patients clinically and to determine whether additional testing, new medications, or cardiac interventions will be beneficial.

History and Physical Examination

Important information includes any history of previous cardiac disease (myocardial infarction [MI], angina, heart failure, arrhythmias, valvular disease), cardiac interventions (e.g., coronary artery bypass grafting; percutaneous coronary intervention, including date and type of stent placed), cardiac evaluation (noninvasive testing, angiography), risk factors (hypertension, diabetes mellitus, dyslipidemia, cigarette smoking), and associated diseases (peripheral arterial disease, stroke, chronic kidney disease, and chronic obstructive pulmonary disease [COPD]). Current status regarding chest pain or dyspnea, functional capacity, and medications should be assessed. The physical examination serves to confirm findings in the history as well as to assess severity and control of the disease (e.g., heart failure, hypertension, valvular disease). The preoperative electrocardiogram rarely changes management unless it demonstrates evidence of a recent or silent MI, but it can be useful as a baseline against which to compare postoperative tracings.

Cardiac Risk Indices

Over the years, a number of risk indices have been proposed to assist in preoperative cardiac evaluation. The most widely used, the revised cardiac risk index (RCRI; Table 431-3), was derived during the evaluation of several thousand patients, was validated in thousands more, and has been incorporated into the consensus guidelines developed by the American College of Cardiology and the American Heart Association. These guidelines, which are updated periodically, use a stepwise strategy based on clinical risk factors, surgery-specific risk, and exercise capacity, combined with a systematic approach to perioperative testing and treatment in patients with known or suspected cardiac disease (Fig. 431-1). Because the RCRI may underestimate risk in major vascular surgery and did not include many types of lower-risk surgeries (Table 431-4), newer risk indices have been developed. Although these alternatives have not been widely validated, current American[2] and European[3] guidelines recommend either the RCRI or the National Surgical Quality Improvement (NSQIP) myocardial infarction/cardiac arrest risk calculator[4] (E-Table 431-1). Preoperative levels of troponin[5] and of brain natriuretic peptide (BNP)[6] are independent predictors of postoperative cardiac complications, although it is unclear how to use these biomarkers and whether any intervention based on these numbers will improve outcome.

Noninvasive Tests

One emphasis of current guidelines is to minimize cardiac testing unless the results are likely to alter management. A resting echocardiogram (Chapter 55) is indicated to evaluate valvular heart disease in patients with clinically suspicious murmurs and to evaluate left ventricular function in patients with heart failure. Other than for the assessment of aortic stenosis (see later), resting echocardiography is not a reliable predictor of perioperative cardiac events.

Exercise testing (with or without imaging) is preferred to pharmacologic stress testing because it assesses functional capacity (see Table 51-3), but its use is often limited by a patient's inability to achieve the target heart rate. Furthermore, patients with adequate exercise capacity by history rarely require preoperative stress testing.

Pharmacologic Stress Testing

Pharmacologic stress testing (either dipyridamole or adenosine with nuclear imaging [Chapters 56 and 71] or dobutamine echocardiography [Chapters 55 and 71]) is indicated when a patient who needs a stress test cannot perform adequate exercise (see Table 71-6 and Fig. 71-3 in Chapter 71). Both tests have a similar sensitivity for predicting perioperative ischemic complications, whereas stress echocardiography has fewer false-positive results. Nevertheless, local expertise commonly influences which test is selected. Dipyridamole and adenosine can cause bronchospasm and are best avoided in patients with symptomatic or severe asthma or obstructive lung disease, but they are preferred in patients with left bundle branch block, in whom exercise or stress echocardiography is more likely to give false-positive results. Quantitatively, the number and extent of reperfusion defects or wall motion abnormalities correlate with the severity of disease, likelihood of complications, and need for further evaluation by angiography.

Patients whose clinical condition would warrant stress testing independent of planned surgery should have such testing before elective operations.

E-TABLE 431-1 THE NATIONAL SURGICAL QUALITY IMPROVEMENT (NSQIP) MYOCARDIAL INFARCTION/CARDIAC ARREST RISK CALCULATOR

Age
Type of surgery
Functional status
Serum creatinine >1.5 mg/dL
American Society of Anesthesiologists (ASA) Class
I = completely healthy
II = mild systemic disease
III = severe systemic disease that is not incapacitating
IV = incapacitating disease that is a consistent threat to life
V = moribund patient not expected to live for 24 hours with or without the surgery

Adapted from Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg.* 2013;217:833-842. Calculator available at http://www.riskcalculator.facs.org/ (Accessed February 3, 2015).

TABLE 431-2 PERIOPERATIVE MANAGEMENT OF MEDICATIONS

MEDICATION CLASS	RECOMMENDATION
Anticoagulants (heparins, warfarin NOACs [novel oral anticoagulants])*	Continue for minor surgery Discontinue at an appropriate interval before major surgery Consider bridging anticoagulation for patients at high risk for interim thrombosis (Chapter 38)
Antiplatelet drugs	Continue for minor surgery Discontinue clopidogrel and ticagrelor at least 5 days before surgery and prasugrel at least 7 days before surgery, except in patients with recent coronary stenting If discontinuing aspirin, do so 3-7 days before surgery
Cardiovascular medications	Continue most agents Consider starting β-blockers in patients at high risk for perioperative cardiac morbidity (vascular or high-risk surgery) Withhold diuretics on the morning of surgery, especially if signs of volume depletion are present Consider stopping ACE inhibitors or ARBs at least 12 hours before surgery unless patient has heart failure or uncontrolled hypertension Stop tamsulosin before cataract surgery (floppy iris syndrome)
Lipid-lowering agents	Continue "statins" Discontinue other agents
Pulmonary agents	Continue
Gastrointestinal agents	Continue
Diabetic agents (see text)	Withhold oral hypoglycemic agents on the morning of surgery; restart when the patient resumes eating For type 1 diabetes, continue some form of insulin (long acting or intravenous) at all times For type 2 diabetes, decrease the dose of morning intermediate insulin; continue basal insulin
Thyroid agents (hypothyroidism and hyperthyroidism) (see text)	Continue thyroid replacement Continue antithyroid medication and postpone surgery until the hyperthyroidism is controlled
Oral contraceptives, hormone replacement, and SERMs	May discontinue 3 weeks before surgery only in patients at high risk for perioperative venous thromboembolism; otherwise continue
Corticosteroids (see text)	Continue chronic corticosteroids; increase the dosage to account for surgical stress
Psychotropic agents	Continue SSRIs but consider withholding them several weeks before CNS surgery Continue tricyclic antidepressants, benzodiazepines, lithium, and antipsychotics Usually discontinue MAOIs 10-14 days before surgery
Chronic opioids	Continue; substitute equianalgesic or higher doses for surgical pain
Rheumatologic agents	Continue methotrexate Discontinue other DMARDs and anticytokines about 2 weeks before surgery Continue hypouricemic agents
Neurologic agents	Continue antiseizure medications Consider withholding antiparkinsonian agents briefly Continue agents for myasthenia gravis
Herbal agents	Discontinue all agents

*See also Tables 431-6 and 431-7 for more detail.
ARB = angiotensin receptor blocker; ACE = angiotensin-converting enzyme; CNS = central nervous system; DMARD = disease-modifying antirheumatic drug; MAOI = monoamine oxidase inhibitor; SERM = selective estrogen receptor modulator; SSRI = selective serotonin reuptake inhibitor.
Adapted from Cohn SL, Macpherson DS. Perioperative medication management. In: Cohn SL, Smetana GW, Weed HG, eds. *Perioperative Medicine: Just the Facts.* New York: McGraw-Hill; 2006.

TABLE 431-3 CLINICAL FACTORS IMPORTANT IN ASSESSING PERIOPERATIVE CARDIAC RISK

REVISED CARDIAC RISK INDEX CRITERIA*
Ischemic heart disease defined as history of myocardial infarction, positive exercise test, current complaint of chest pain considered secondary to myocardial ischemia, use of nitrate therapy, or pathological Q waves on the electrocardiogram *Or* at least two of the following:
Heart failure defined as S_3 or bilateral rales on physical examination or pulmonary edema on chest radiograph
Cerebrovascular disease defined as history of transient ischemic attack or history of cerebrovascular accident
Insulin-dependent diabetes mellitus
Chronic renal insufficiency defined as baseline creatinine of 2.0 mg/dL or greater
High-risk surgery defined as intrathoracic, intra-abdominal, or suprainguinal vascular surgery

*Lee TH, Marcantonio ER, Mangione CM, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation.* 1999;100:1043-1049.

TABLE 431-4 RISKS OF VARIOUS SURGICAL PROCEDURES

HIGH (VERY ELEVATED) RISK (CARDIAC RISK >5%)
Major vascular surgery Emergent major operations Prolonged procedures with large fluid shifts or significant blood loss
INTERMEDIATE (BUT ELEVATED) RISK (CARDIAC RISK 1-5%)
Intraperitoneal or intrathoracic procedures Carotid endarterectomy Endovascular aortic aneurysm repair Head and neck surgery Orthopedic procedures Prostate surgery
LOW RISK (CARDIAC RISK <1%)
Superficial operations Cataract surgery Breast surgery Ambulatory surgery

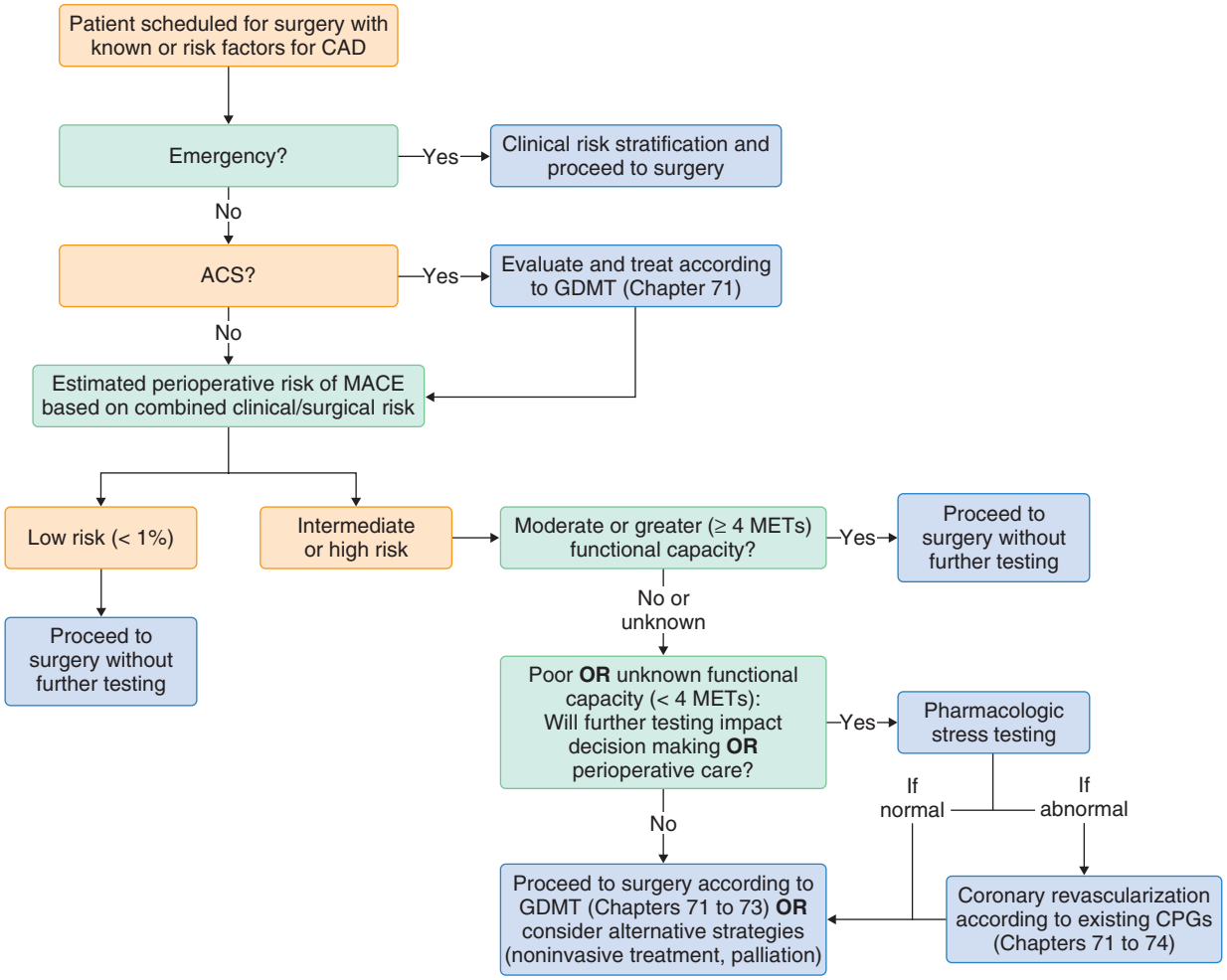

FIGURE 431-1. Stepwise approach to perioperative cardiac assessment for CAD. ACS = acute coronary syndrome (Chapter 72); CAD = coronary artery disease; CPGs = clinical practice guidelines; GDMT = guideline-directed medical therapy; MACE = major adverse cardiac event; METs = metabolic equivalents. Adapted from Fleisher LA, Fleischmann KE, Auerbach AD, et al. 2014 ACC/AHA Guideline on Perioperative Cardiovascular Evaluation and Management of Patients Undergoing Noncardiac Surgery: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2014;64:2373-2405.

Otherwise, stress testing is recommended only in patients at elevated risk for noncardiac surgery and with poor functional capacity (defined as the inability to walk two to four blocks at 3-4 mph on level ground or to climb one flight of stairs; see Table 51-5) if the results will change management.

Risk Reduction Strategies for Ischemic Heart Disease

Medical Therapy

In by far the largest trial of prophylactic perioperative β-blockers, a high dose of extended-release metoprolol, started hours before surgery, reduced perioperative MI, but at the expense of increasing strokes and overall mortality, in part owing to more hypotension and bradycardia.[■] Meta-analyses that exclude trials whose veracity has been called into question by ongoing institutional investigations of the primary author have reinforced that β-blockers can reduce perioperative MI but with the side effect of increasing stroke. The net result is no evidence for a reduction in overall mortality and probably increased risk in patients who do not have a very high risk for perioperative MI.[■■] It may be reasonable to start β-blockers before surgery in patients with intermediate-high risk ischemia on stress testing or with 3 or more RCRI risk factors, but validated data to support this option are insufficient to make definitive recommendations.[7] Bisoprolol (5 mg daily) or atenolol (25 mg daily) may be preferable to metoprolol, and any benefits are more likely to be seen when β-blockers are started at least 1 week before surgery at a low dose and titrated to a heart rate of 55 to 70 beats per minute.[8] Until more evidence is available, it seems prudent to avoid starting β-blockers immediately before surgery and to avoid them in the settings of emergency surgery, prior cerebrovascular disease, or sepsis.

Neither clonidine[■] nor aspirin[■] is beneficial for reducing perioperative cardiac events. Limited data on prophylactic calcium antagonists or nitrates have not shown major benefits in preventing complications after noncardiac surgery. Statins (Chapter 206) reduce endovascular inflammation and stabilize endothelial plaque. Current data suggest that they should be continued perioperatively and also begun preoperatively in patients who meet criteria for their ongoing use (Chapter 206).[■] Aggressive fluid management to optimize cardiac output is controversial, and recent studies show no clear advantage when added to standard medical therapy.[■■]

Invasive Therapies

Prophylactic coronary revascularization in patients who have stable cardiac symptoms and no aortic stenosis and do not meet standard criteria for the procedure (Chapter 71) does not reduce perioperative myocardial infarction, death within 30 days, or long-term mortality at an average of 2.7 years in patients who receive appropriate medical therapy.[■] Preoperative coronary revascularization is indicated only if the patient meets the criteria for coronary angiography or revascularization independent of the need for surgery.

For bare metal stents, data suggest that elective surgery should be delayed for 4 to 6 weeks after stenting because of the risk for in-stent thrombosis when dual antiplatelet therapy with aspirin and clopidogrel is discontinued early or because of the alternative risk for bleeding if surgery is performed in patients receiving such antiplatelet therapy (Chapter 74). For drug-eluting stents, elective surgery should be delayed for at least 6 months and preferably 12 months if possible[9] so that patients can complete an uninterrupted course of dual antiplatelet therapy. For balloon angioplasty without stenting, a delay of 2 weeks is generally recommended. If antiplatelet therapy has to be discontinued, clopidogrel is usually discontinued 5 to 7 days before the noncardiac procedure, prasugrel is stopped 7 days before, and ticagrelor is stopped

5 days before, whereas aspirin is continued, if possible.[10] If aspirin also must be discontinued, it is usually stopped approximately 5 to 7 days before surgery, but shorter durations are being evaluated.

Other Cardiovascular Diseases

Heart Failure

Heart failure, which is a major risk factor for surgery, requires treatment and optimization before surgery (Chapter 59). Routine use of *pulmonary artery catheters* does not reduce morbidity or mortality in patients undergoing elective noncardiac surgery. Although an elevated BNP level is a risk factor, there is no evidence that treatment to lower the BNP level or use of β-blockers will reduce postoperative complications in patients with heart failure.

Valvular Heart Disease

Patients with *symptomatic aortic stenosis* who meet the criteria for valve replacement (Chapter 75) independent of their need for surgery should undergo the valve replacement before the noncardiac surgery. However, patients usually survive noncardiac surgery with intensified care if they refuse valve replacement or time does not permit it. An asymptomatic aortic valve area of 1.0 to 1.5 cm^2 carries an increased risk for perioperative complications[11] but is an indication for more careful monitoring rather than valve surgery. Patients with severe mitral regurgitation also have higher risks for postoperative cardiovascular complications.[12] Endocarditis prophylaxis (Chapter 76) is appropriate for patients with mechanical heart valves, previous endocarditis, complex congenital heart disease, or valvular disease in a heart transplant recipient undergoing invasive dental or upper respiratory procedures (Chapter 76).

Hypertension

Hypertension with blood pressure lower than 110 mm Hg diastolic or 180 mm Hg systolic without significant target organ damage does not increase the risk for major perioperative cardiac complications. Even when the preoperative diastolic blood pressure is higher, limited data suggest that surgery is safe after additional antihypertensive therapy.

Arrhythmias

Although patients with arrhythmias have increased perioperative risk, the risk is increased because the arrhythmias are usually markers of more serious heart disease or cause hemodynamic problems. Patients with hemodynamically significant tachyarrhythmias and bradyarrhythmias should generally be treated as in the nonoperative setting (Chapters 64 and 65), except for the special circumstance of anticoagulation in the perioperative setting (Chapter 38).

PULMONARY RISK ASSESSMENT

Postoperative pulmonary complications are as common as cardiac complications and are associated with significant morbidity and mortality. Major complications include respiratory failure (e.g., reintubation, prolonged mechanical ventilation), pneumonia, atelectasis requiring bronchoscopy, and to a lesser degree, bronchospasm or an exacerbation of COPD requiring treatment and prolonged length of stay. Many postoperative pulmonary complications are due to exaggerations of the usual postoperative changes in pulmonary function: decreased lung volumes, diaphragmatic dysfunction, ventilation-perfusion mismatches and shunting, hypoventilation, hypoxemia, and impaired defense mechanisms. Pulmonary risk factors can be divided into patient-related and procedure-related factors, the latter of which include the type of surgery, anesthesia, and related factors.[13]

Patient-Related Factors

COPD (Chapter 88) increases the risk for postoperative pulmonary complications approximately twofold, depending on its severity, whereas well-controlled *asthma* (Chapter 87) does not increase risk. Active cigarette smokers are at increased risk, mainly related to the number of pack years smoked; smoking cessation at least 4 to 8 weeks before surgery may reduce the risk. *Obstructive sleep apnea* (Chapter 100), typically associated with obesity, confers an increased risk for hypercapnia and hypoxemia, and obese patients are at increased risk for atelectasis.[14] Advanced age, poor functional status, pulmonary hypertension, an altered mental state, and suppressed immune status from chronic steroid use, alcohol use, or diabetes may also increase the risk for postoperative pulmonary complications. Pulmonary hypertension has also been associated with increased risk.

Procedure-Related Factors

The most important predictors of postoperative pulmonary complications are the type of surgery and proximity of the surgical incision to the diaphragm. Pulmonary function decreases by approximately 50% after intrathoracic surgery, upper abdominal procedures, and abdominal aortic aneurysm repair and does not fully return to normal for several weeks. Lower abdominal surgery is associated with a 25% decrease in pulmonary function. Laparoscopic procedures are usually associated with lower rates of postoperative pulmonary complications and shorter hospital stays than open procedures. Neuraxial anesthesia (epidural or spinal) may be associated with decreased risk when compared with general anesthesia, but the decision about which type of anesthesia to use is best left to the anesthesiologist. Emergency surgery, prolonged duration of anesthesia or surgery (>2 to 6 hours), and routine postoperative nasogastric tube use increase the risk for postoperative pulmonary complications.

Pulmonary Function Tests

In general, pulmonary function tests (Chapter 85) are no more predictive of pulmonary complications than is clinical risk assessment alone. Such testing may be more helpful in assessing risk for lung resection surgery when it can predict the function of the remaining lung mass. However, even a postoperative predicted forced expiratory volume in 1 second (FEV_1) of less than 800 mL for lung resection, which is thought to portend a very high risk for death or prolonged mechanical ventilation, is not an absolute contraindication to surgery. Preoperative arterial blood gas evaluation is also of little benefit in predicting postoperative pulmonary complications. Cardiopulmonary exercise testing for maximal oxygen consumption is useful for evaluating high-risk patients before lung resection surgery.

Risk Reduction Strategies

Unfortunately, many of the risk factors for postoperative pulmonary complications cannot be modified. Inhaled bronchodilators (β-agonists and anticholinergics) and steroids can optimize the respiratory status of patients with COPD and asthma. Broad-spectrum antibiotics should be used to treat exacerbations caused by bacterial infection. Chest physiotherapy may be helpful, particularly for thoracic surgery. Smoking should be stopped at least 8 weeks before surgery, if possible.

Lung expansion maneuvers (either incentive spirometry or deep-breathing exercises) can significantly improve pulmonary function, minimize atelectasis, and reduce risk, especially for thoracic and upper abdominal surgery. Pain control (Chapter 30) improves pulmonary function by allowing deeper breathing. Epidural analgesia and patient-controlled intravenous analgesia reduce postoperative pulmonary complications and, when possible, are preferable to parenteral narcotics. Long-acting neuromuscular blockers should be avoided, and the selective rather than the routine use of a nasogastric tube may also decrease risk.

ENDOCRINE CONDITIONS

Diabetes Mellitus

The major risks associated with surgery in diabetic patients are cardiac complications and wound infections. Complications are probably related more to associated diseases and end-organ involvement (coronary artery disease, chronic kidney disease, and autonomic neuropathy) than to the glucose level itself. Significantly, elevated glucose levels may impair wound healing and interfere with leukocyte defense mechanisms. However, current recommendations suggest a glucose target level of 140 to 180 mg/dL rather than tight perioperative control.[15]

Patients whose diabetes is controlled by diet require only perioperative glucose monitoring (finger sticks) with short-acting insulin coverage on an as-needed basis. Patients taking oral hypoglycemic agents (Chapter 229) should not take them on the morning of surgery (chlorpropamide should be stopped 2 to 3 days before and metformin preferably 12 to 24 hours before major surgery) and should be monitored with sliding-scale insulin coverage as needed. Patients taking insulin are most often given one half to two thirds of their usual intermediate-acting insulin on the morning of surgery and are then given short-acting insulin on a sliding scale and correction dose based on finger stick monitoring (Chapter 229). Continuous intravenous insulin, which provides tighter glucose control but is associated with more episodes of hypoglycemia and requires a monitored setting, is typically used in patients

undergoing cardiac surgery and in critically ill patients. There are no validated guidelines regarding the management of patients taking long-acting basal insulin (glargine). In general, it should be continued, but its dose may be reduced for ambulatory surgery or in patients with tight control or with chronic kidney disease. Regardless of the mode of treatment, frequent monitoring of the glucose level is critical.

Exogenous Corticosteroids and Adrenal Insufficiency

The stress of surgery activates the hypothalamic-pituitary-adrenal (HPA) axis, which in turn stimulates release of adrenocorticotropic hormone (ACTH) and subsequent secretion of cortisol (Chapter 227), but a patient who is taking exogenous corticosteroids may have suppression of the HPA axis and not be able to respond to this stress adequately. As a result, hypotension and shock may occur.

In general, a daily dose equivalent to 5 mg or less of prednisone (Chapter 227), alternate-day short-acting therapy, or corticosteroids given for less than 3 weeks do not cause clinically significant HPA suppression, so no supplemental therapy is indicated. Conversely, doses greater than 20 mg/day of prednisone for longer than 3 weeks usually suppress the HPA axis and warrant perioperative supplemental corticosteroids. In patients who are taking intermediate dosing regimens or who took large doses in the past year but are not taking corticosteroids or are taking lower doses now, the options are to perform an ACTH (cosyntropin) stimulation test, if time permits, and treat only patients with an inadequate response (Chapter 227) or to prescribe supplemental corticosteroids empirically.

Despite a lack of definitive evidence that supplemental steroids are required, when supplemental corticosteroids are felt to be appropriate, short-term therapy tailored to the level of expected stress can provide protection without adverse effects on wound healing and with only short-term problems with glucose intolerance and fluid retention. For minor procedures or local anesthesia, the recommended approach is to give the patient's usual dose before surgery without further supplementation. For moderate surgical stress (e.g., open cholecystectomy, lower extremity vascular surgery), a reasonable approach is 50 mg of hydrocortisone intravenously before surgery, followed by 25 mg every 8 hours for 1 to 2 days, and then the patient's usual dose. For major surgical stress, patients are typically given 75 to 100 mg of hydrocortisone intravenously before induction of anesthesia, followed by 50 mg every 8 hours for 1 to 3 days until the stressful period resolves, and then their usual dose.

Thyroid Disease

An inadequately treated or undiagnosed hyperthyroid patient is potentially at risk for thyroid storm postoperatively. Elective surgery should be postponed in patients who are symptomatic or have resting tachycardia until they are euthyroid. Treatment of a thyrotoxic patient undergoing urgent or emergency surgery includes a combination of β-blockers, antithyroid agents, and iodine to control the resting heart rate to less than 90 beats per minute, as well as prophylactic corticosteroid supplementation, as used for thyroid storm (Chapter 226).

Conversely, patients with mild to moderate hypothyroidism tolerate surgery reasonably well. Patients with markedly symptomatic hypothyroidism should be treated with oral levothyroxine (T_4) for several weeks before elective surgery. For emergency surgery, intravenous liothyronine (T_3) or T_4 (200 to 300 μg intravenously, then 50 to 100 μg/day) and supplemental corticosteroids (hydrocortisone, 100 mg intravenously, then 25 to 50 mg every 6 hours) should be given. Myxedema coma is a rare complication of surgery.

LIVER DISEASE

Routine preoperative testing of liver function is not recommended, but elective surgery should be avoided in patients with acute viral, alcoholic, or drug-induced hepatitis. Patients with stable mild chronic hepatitis tolerate surgery well.

Patients with alcoholic liver disease or cirrhosis are at risk for postoperative complications, including bleeding, infection, poor wound healing, and delirium.[16] The severity of disease as assessed by Child-Turcotte-Pugh criteria and the MELD (Model for End-Stage Liver Disease) score (Chapter 154) can be used to estimate risk; the MELD score is thought to be more predictive of outcome. Child's C class and MELD score greater than 15 portend very high risk, and elective surgery is usually contraindicated. Aggressive treatment of coagulopathy, ascites, and encephalopathy is indicated before surgery.

HEMATOLOGIC PROBLEMS

Preoperative anemia, even to a mild degree, is independently associated with an increased risk for 30-day morbidity and mortality in patients undergoing major noncardiac surgery,[17] but operative patients generally tolerate hemoglobin levels as low as 7 g/dL. Preoperative transfusion should not be triggered solely by the hemoglobin level but should also consider the expected blood loss from the surgical procedure and the patient's comorbid conditions. For patients with cardiopulmonary disease, however, a goal of 10 g/dL has typically been recommended for major surgery, although a trigger of 8 to 9 g/dL may be appropriate.[18]

Patients without a personal or family history of abnormal bleeding require no preoperative testing of coagulative function, but those with such a history should be evaluated. Ideally, the prothrombin time should be within 3 seconds of control (international normalized ratio <1.5), the partial thromboplastin time within 10 seconds of control, and the platelet count above a minimum of 50,000, depending on the type of surgery.

The approach to perioperative anticoagulation, both in terms of prevention of venous thromboembolism and for the management of a patient taking warfarin, aspirin, or other antithrombotic medications, is described elsewhere (Chapter 38). For patients already on anticoagulants, perioperative recommendations depend on the short-term risks for thromboembolism and bleeding (Table 431-5)[19] and provide additional guidelines regarding perioperative aspirin use in cardiac and noncardiac surgery. However, the need for bridging therapy has been questioned because the risk for bleeding with early postoperative anticoagulation may outweigh any potential benefit.[20] Use of the new oral anticoagulants (dabigatran, rivaroxaban, and apixaban) obviates the need for bridging therapy because of their shorter half-lives and more rapid onset of action compared with warfarin (Table 431-6). They can be stopped closer to the time of surgery but should be started postoperatively only after adequate hemostasis has been ensured, usually 48 to 72 hours after major surgery.

TABLE 431-5 SUGGESTED APPROACH TO ANTICOAGULATION IN THE PERIOPERATIVE PATIENT

Low thromboembolic risk/low bleeding risk • Continue anticoagulant therapy with INR in therapeutic range.
Low thromboembolic risk/high bleeding risk • Discontinue anticoagulant therapy 5 days before the procedure. • Start LMWH prophylaxis once daily or UFH IV 1 day after acenocoumarol interruption and 2 days after warfarin interruption. Administer the last dose of LMWH at least 24 hr before the procedure or give UFH IV up to 4-6 hr before surgery. • Resume LMWH or UFH at the preprocedural dose 1-2 days (at least 12 hr) after the procedure according to hemostatic status. Resume anticoagulant therapy 1 to 2 days after surgery at the preprocedural dose + 50% boost dose for 2 consecutive days according to the hemostatic status. • LMWH or UFH is continued until the INR has returned to therapeutic levels.
High thromboembolic risk • Discontinue anticoagulant therapy 5 days before the procedure. • Start therapeutic LMWH twice daily or UFH IV 1 day after acenocoumarol interruption and 2 days after warfarin interruption. Administer the last dose of LMWH at least 24 hr before the procedure or give UFH IV up to 4-6 hr before surgery. • Resume LMWH or UFH at the preprocedural dose 24-72 hr after the procedure depending on the risk of bleeding. Resume anticoagulant therapy 12-24 hr after surgery according to hemostatic status. • LMWH or UFH is continued until the INR has returned to therapeutic levels.

LMWH = low-molecular-weight heparin; INR = international normalized ratio; IV = intravenous; UFH = unfractionated heparin.
Adapted from Poldermans D, Bax JJ, Boersma E, et al. Guidelines for pre-operative cardiac risk assessment and perioperative cardiac management in non-cardiac surgery. The Task Force for Preoperative Cardiac Risk Assessment and Perioperative Cardiac Management in Non-cardiac Surgery of the European Society of Cardiology (ESC) and endorsed by the European Society of Anaesthesiology (ESA). *Eur Heart J.* 2009;30:2769-2812 and Douketis JD, Spyropoulos AC, Spencer FA, et al. Perioperative management of antithrombotic therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest.* 2012;141:e326S-350S.

TABLE 431-6 PREOPERATIVE MANAGEMENT OF THE NEW ORAL ANTICOAGULANTS

DRUG	CREATININE CLEARANCE (mL/min)	HALF-LIFE (hr)	TIMING OF LAST DOSE BEFORE SURGERY: *Standard Bleeding Risk Surgery*	TIMING OF LAST DOSE BEFORE SURGERY: *High Bleeding Risk Surgery**
Dabigatran	>50	13-15	1 day	2 days
	31-50	18	2 days	3-4 days
	≤30	27	3-4 days	4-5 days
Rivaroxaban	>30	7-11	1 day	2 days
	≤30	?	2 days	3-4 days
Apixaban	>30	8-14	1 day	2 days
	≤30	?	2 days	3-4 days

*Examples include neurosurgery, spine surgery, cardiac, major abdominal, and vascular surgery.

RENAL DISORDERS

Chronic kidney disease is an independent risk factor for postoperative cardiovascular events and death.[21] Patients with chronic kidney disease typically have other comorbid diseases and may also have fluid and electrolyte abnormalities, anemia, and bleeding diatheses, which should be treated and optimized before surgery. Patients maintained on dialysis should ideally undergo dialysis the day before surgery to optimize their volume status, prevent hyperkalemia, and minimize acute shifts in acid-base balance.

NEUROLOGIC AND GERIATRIC PROBLEMS

The risk for a postoperative stroke in unselected patients after general surgery is less than 0.5%, but patients with a history of stroke, older patients, and those undergoing vascular surgery, especially carotid and cardiac surgery, have higher risk. Patients with symptomatic carotid bruits require further investigation and possible intervention before elective surgery (Chapter 407). Patients who newly receive perioperative β-blockers, particularly high doses of metoprolol, appear to be at increased risk for stroke, and the risk must be balanced against the protective effect of β-blockers for perioperative MI. However, there is no evidence that continuation of chronic β-blockade increases risk for postoperative stroke. There is no evidence to support preoperative intervention in patients with asymptomatic bruits before noncardiac surgery. The general recommendation is to delay elective surgery for at least 4 weeks after a stroke, although some data suggest waiting up to 9 months.[22]

The elderly are at higher risks for a variety of poor postoperative outcomes. Cognitive impairment (Chapters 27 and 28), frailty (Chapter 24), malnutrition, and prior institutionalization all are associated with a poorer prognosis.[23]

Grade A References

A1. Devereaux PJ, Yang H, Yusuf S, et al. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. *Lancet*. 2008;371:1839-1847.

A2. Bangalore S, Wetterslev J, Pranesh S, et al. Perioperative beta blockers in patients having non-cardiac surgery: a meta-analysis. *Lancet*. 2008;372:1962-1976.

A3. Nowbar AN, Cole GD, Shun-Shin MJ, et al. International RCT-based guidelines for use of preoperative stress testing and perioperative beta-blockers and statins in non-cardiac surgery. *Int J Cardiol*. 2014;172:138-143.

A4. Devereaux PJ, Sessler DI, Leslie K, et al. Clonidine in patients undergoing noncardiac surgery. *N Engl J Med*. 2014;370:1504-1513.

A5. Devereaux PJ, Mrkobrada M, Sessler DI, et al. Aspirin in patients undergoing noncardiac surgery. *N Engl J Med*. 2014;370:1494-1503.

A6. Sanders RD, Nicholson A, Lewis SR, et al. Perioperative statin therapy for improving outcomes during and after noncardiac vascular surgery. *Cochrane Database Syst Rev*. 2013;7:CD009971.

A7. Yealy DM, Kellum JA, Huang DT, et al. A randomized trial of protocol-based care for early septic shock. *N Engl J Med*. 2014;370:1683-1693.

A8. Pearse RM, Harrison DA, MacDonald N, et al. Effect of a perioperative, cardiac output-guided hemodynamic therapy algorithm on outcomes following major gastrointestinal surgery: a randomized clinical trial and systematic review. *JAMA*. 2014;311:2181-2190.

A9. McFalls EO, Ward HB, Moritz TE, et al. Coronary-artery revascularization before elective major vascular surgery. *N Engl J Med*. 2004;351:2795-2804.

A10. Sandham JD, Hull RD, Brant RF, et al. Canadian Critical Care Clinical Trials Group. A randomized, controlled trial of the use of pulmonary-artery catheters in high-risk surgical patients. *N Engl J Med*. 2003;348:5-14.

A11. Yong SL, Coulthard P, Wrzosek A. Supplemental perioperative steroids for surgical patients with adrenal insufficiency. *Cochrane Database Syst Rev*. 2012;12:CD005367.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Apfelbaum JL, Connis RT, Nickinovich DG, et al. Practice advisory for preanesthesia evaluation: an updated report by the American Society of Anesthesiologists Task Force on Preanesthesia Evaluation. *Anesthesiology*. 2012;116:522-538.
2. Fleisher LA, Fleischmann KE, Auerbach AD, et al. 2014 ACC/AHA Guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;130:2215-2245.
3. Kristensen SD, Knuuti J, Saraste A, et al. 2014 ESC/ESA Guidelines on non-cardiac surgery: cardiovascular assessment and management: The Joint Task Force on non-cardiac surgery: cardiovascular assessment and management of the European Society of Cardiology (ESC) and the European Society of Anaesthesiology (ESA). *Eur Heart J*. 2014;35:2383-2431.
4. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg*. 2013;217:833-842.
5. Weber M, Luchner A, Seeberger M, et al. Incremental value of high-sensitive troponin T in addition to the revised cardiac index for peri-operative risk stratification in non-cardiac surgery. *Eur Heart J*. 2013;34:853-862.
6. Rodseth RN, Biccard BM, Chu R, et al. Postoperative B-type natriuretic peptide for prediction of major cardiac events in patients undergoing noncardiac surgery: systematic review and individual patient meta-analysis. *Anesthesiology*. 2013;119:270-283.
7. Wijeysundera DN, Duncan D, Nkonde-Price C, et al. Perioperative beta blockade in noncardiac surgery: a systematic review for the 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *J Am Coll Cardiol*. 2014;64:2406-2425.
8. Dai N, Xu D, Zhang J, et al. Different beta-blockers and initiation time in patients undergoing noncardiac surgery: a meta-analysis. *Am J Med Sci*. 2014;347:235-244.
9. Hawn MT, Graham LA, Richman JS, et al. Risk of major adverse cardiac events following noncardiac surgery in patients with coronary stents. *JAMA*. 2013;310:1462-1472.
10. Darvish-Kazem S, Gandhi M, Marcucci M, et al. Perioperative management of antiplatelet therapy in patients with a coronary stent who need noncardiac surgery: a systematic review of clinical practice guidelines. *Chest*. 2013;144:1848-1856.
11. Agarwal S, Rajamanickam A, Bajaj NS, et al. Impact of aortic stenosis on postoperative outcomes after noncardiac surgeries. *Circ Cardiovasc Qual Outcomes*. 2013;6:193-200.
12. Bajaj NS, Agarwal S, Rajamanickam A, et al. Impact of severe mitral regurgitation on postoperative outcomes after noncardiac surgery. *Am J Med*. 2013;126:529-535.
13. Sabaté S, Mazo V, Canet J. Predicting postoperative pulmonary complications: implications for outcomes and costs. *Curr Opin Anaesthesiol*. 2014;27:201-209.
14. Practice guidelines for the perioperative management of patients with obstructive sleep apnea: an updated report by the American Society of Anesthesiologists Task Force on Perioperative Management of patients with obstructive sleep apnea. *Anesthesiology*. 2014;120:268-286.
15. Umpierrez GE, Hellman R, Korytkowski MT, et al. Management of hyperglycemia in hospitalized patients in non-critical care setting: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2012;97:16-38.
16. Bhangui P, Laurent A, Amathieu R, et al. Assessment of risk for non-hepatic surgery in cirrhotic patients. *J Hepatol*. 2012;57:874-884.
17. Musallam KM, Tamim HM, Richards T, et al. Preoperative anaemia and postoperative outcomes in non-cardiac surgery: a retrospective cohort study. *Lancet*. 2011;378:1396-1407.
18. Carson JL, Grossman BJ, Kleinman S, et al. Red blood cell transfusion: a clinical practice guideline from the AABB*. *Ann Intern Med*. 2012;157:49-58.
19. Douketis JD, Spyropoulos AC, Spencer FA, et al. Perioperative management of antithrombotic therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e326S-350S.
20. Siegal D, Yudin J, Kaatz S, et al. Periprocedural heparin bridging in patients receiving vitamin K antagonists: systematic review and meta-analysis of bleeding and thromboembolic rates. *Circulation*. 2012;126:1630-1639.
21. Mooney JF, Chow CK, Hillis GS. Perioperative renal function and surgical outcome. *Curr Opin Anaesthesiol*. 2014;27:195-200.
22. Jørgensen ME, Torp-Pedersen C, Gislason GH, et al. Time elapsed after ischemic stroke and risk of adverse cardiovascular events and mortality following elective noncardiac surgery. *JAMA*. 2014;312:269-277.
23. Oresanya LB, Lyons WL, Finlayson E. Preoperative assessment of the older patient: a narrative review. *JAMA*. 2014;311:2110-2120.

REVIEW QUESTIONS

1. A 70-year-old man is scheduled for cataract surgery. Past medical history includes hypertension, hyperlipidemia, and coronary artery disease with a myocardial infarction 10 years ago. He has been asymptomatic since then but has a sedentary lifestyle. He is a former smoker. Medications include aspirin, metoprolol, and atorvastatin. He denies chest pain or dyspnea. Vital signs and physical examination are within normal limits. His electrocardiogram at the time of his last visit 3 months ago showed normal sinus rhythm with left ventricular hypertrophy and evidence of an old inferior wall myocardial infarction. Which of the following would you recommend preoperatively?

A. Basic metabolic panel (electrolytes, renal function, glucose)
B. Electrocardiogram
C. Hemoglobin
D. Stress test
E. No further testing

Answer: E Despite having multiple risk factors, the patient is undergoing a low-risk procedure, and no further testing would be helpful. His risk for major cardiac complications is less than 1%, and the procedure is not associated with any significant hemodynamic changes or blood loss. Therefore, even an abnormal result (anemia, chronic kidney disease, hyperglycemia) would be unlikely to change management. (Keay L, Lindsley K, Tielsch J, et al. Routine preoperative medical testing for cataract surgery. *Cochrane Database Syst Rev.* 2012;3:CD007293.)

2. A 60-year-old man is scheduled for colonoscopy and polypectomy. His medical history is pertinent for hypertension, diabetes mellitus, and atrial fibrillation, and he is taking warfarin, metformin, and amlodipine. Which of the following would you recommend regarding perioperative anticoagulation management?

A. Continue warfarin
B. Continue warfarin but give fresh-frozen plasma 2 hours before surgery
C. Stop warfarin 5 days before surgery
D. Stop warfarin 5 days before surgery and "bridge" with low-molecular-weight heparin preoperatively
E. Stop warfarin 5 days before surgery and "bridge" with low-molecular-weight heparin preoperatively and postoperatively

Answer: C In general, warfarin needs to be stopped before a polypectomy. This patient has a $CHADS_2$ score of 2, in which case the risk for thromboembolism is low if temporarily stopping warfarin. Furthermore, if full-dose anticoagulation is started too early after the procedure, there is an increased risk for bleeding. The American College of Chest Physicians guidelines recommend stopping warfarin 5 days before the procedure (to allow the international normalized ratio to drop below 1.5) without the need for bridging therapy in this case. (Douketis JD, Spyropoulos AC, Spencer FA, et al; American College of Chest Physicians. Perioperative management of antithrombotic therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest.* 2012;141[2 Suppl]:e326S-50S. Erratum in: *Chest.* 2012; 141:1129.)

3. A 60-year-old woman is seen preoperatively before a vaginal hysterectomy. She has hypertension, diabetes, chronic kidney disease, and coronary artery disease but no history of myocardial infarction, recent chest pain, or dyspnea. She walks 1 mile 3 times a week and can climb 2 flights of stairs without symptoms. She states she had a negative stress test 5 years ago. She is taking aspirin, nitrates, metoprolol, lisinopril, furosemide, and insulin. Her physical examination is unremarkable, and her preoperative test results are: glucose 150 mg/dL, blood urea nitrogen 40/creatinine 2.1. Her electrocardiogram shows normal sinus rhythm with no ischemic changes. Which of the following would you recommend preoperatively?

A. Brain natriuretic peptide
B. Two-dimensional echocardiogram
C. Exercise electrocardiogram
D. Pharmacologic stress test (nuclear or echo)
E. No further testing

Answer: E Although she has a revised cardiac risk index score of 3, she has no active cardiac conditions and has a functional capacity of >4 MET, so no further cardiac testing is necessary. Although an elevated preoperative (or postoperative) brain natriuretic peptide level may be associated with an increased risk for complications, the procedural risk is not high, and there is no evidence that any intervention will improve her outcome. Fleisher LA, Fleischmann KE, Auerbach AD, et al. 2014 ACC/AHA Guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation.* 2014;130:2215-2245.

4. A 70-year-old man is scheduled to undergo an elective total hip replacement for severe osteoarthritis, which limits his activity. He has coronary artery disease and had a drug eluting stent (DES) placed in his proximal left anterior descending coronary artery 3 months ago. He is on aspirin and clopidogrel and has had no chest pain or dyspnea since the stent placement. Which of the following would you recommend?

A. Proceed to surgery; continue aspirin and clopidogrel
B. Proceed to surgery; continue aspirin; stop clopidogrel 5-7 days preoperatively
C. Proceed to surgery; continue clopidogrel; stop aspirin 5-7 days before surgery
D. Proceed to surgery; stop both aspirin and clopidogrel 5-7 days before
E. Postpone surgery for at least 3 months; continue aspirin and clopidogrel

Answer: E Current guidelines recommend continuing uninterrupted dual antiplatelet therapy for at least 6 if not 12 months after placement of a DES to minimize the chance of in-stent thrombosis. Although the newest generation of DES may require shorter durations of antiplatelet therapy, at the current time elective surgery should be postponed for at least 6 and preferably 12 months in order to complete the recommended course of uninterrupted dual antiplatelet therapy. At that time, a decision can be made about whether to continue both drugs or to continue just aspirin, which is continued for life. (Darvish-Kazem S, Gandhi M, Marcucci M, Douketis JD. Perioperative management of antiplatelet therapy in patients with a coronary stent who need non-cardiac surgery: a systematic review of clinical practice guidelines. *Chest.* 2013;144:1848-1856.)

5. A 65-year-old obese woman is seen preoperatively before a total abdominal hysterectomy and bilateral oophorectomy under general anesthesia scheduled for next week. She has a history of asthma, with no recent exacerbations, and cigarette smoking. She has no cough, shortness of breath, or wheezing. She uses albuterol and ipratropium inhalers as needed. Her chest is clear on examination. Which of the following is most likely to reduce her risk for developing postoperative pulmonary complications?

A. 10% weight reduction
B. 5-day course of broad-spectrum antibiotics
C. Incentive spirometry
D. Prophylactic corticosteroids
E. Smoking cessation

Answer: C This patient's risk factors for postoperative pulmonary complications include age, abdominal surgery, general anesthesia, and cigarette smoking. Her asthma is asymptomatic and should not increase her risk. There is no evidence that antibiotics or steroids are helpful in this case. Obesity may increase pulmonary risk, but she will not lose a significant amount of weight in the next week. In order for smoking cessation to improve pulmonary function and decrease complications, she would have to stop at least 4 to 8 weeks before surgery. Lung expansion maneuvers have been show to reduce pulmonary complications and were the only modality given an A level recommendation by the American College of Physicians guidelines. (Qaseem A, Snow V, Fitterman N, et al; Clinical Efficacy Assessment Subcommittee of the American College of Physicians. Risk assessment for and strategies to reduce perioperative pulmonary complications for patients undergoing noncardiothoracic surgery: a guideline from the American College of Physicians. *Ann Intern Med.* 2006;144:575-80.)

432 OVERVIEW OF ANESTHESIA

JEANINE P. WIENER-KRONISH AND LEE A. FLEISHER

In the United States, more than 40 million procedures, including outpatient procedures that require an anesthetic, are performed annually. Additionally, many invasive procedures outside of the operating room, such as in the gastrointestinal endoscopy and electrophysiology suites, are performed using deep sedation or general anesthesia. With modern techniques, anesthesia causes or contributes to mortality in about 1 per 20,000 healthy patients. Although the worldwide perioperative mortality attributable to anesthesia has declined by more than 90% in the past several decades,[1] the overall inpatient postoperative mortality rate remains about 4%, with large variations even among developed countries.[2]

PREOPERATIVE ASSESSMENT

Important aspects of preoperative risk assessment include the type of surgery to be performed, the patient's underlying medical condition, and the particular demands for anesthesia (Chapter 431). In addition, a number of other issues are relevant to management and anesthetic evaluation.

Airway Assessment

Assessment of the airway is always necessary, even if regional anesthesia or monitored anesthesia care (local anesthesia with sedation) is planned, because unexpected complications or compromise of airway reflexes may lead to an emergent need to support ventilation. The laryngeal mask airway device allows many patients to be ventilated easily, but it is important to assess the ability to intubate the patient as well as the ability to ventilate. The prevalence of difficult intubation is about 6% for nonobese patients, and reasons for difficulty include airway pathology (e.g., tumors, previous surgery), reduced mobility of the cervical spine, obstructive sleep apnea, or the anatomic relationship between the larynx and trachea.[3] In many such patients, endotracheal intubation can be accomplished by using a fiberoptic bronchoscope to place the endotracheal tube through either the nose or mouth. If this approach is not successful, a surgical airway must be created.

Criteria for extubation in postoperative patients are similar to those in other patients who receive mechanical ventilation (Chapter 105). Older patients with more severe, comorbid diseases, especially underlying cardiac or pulmonary disease, are more likely to require postoperative reintubation, which is associated with a nine-fold increase in mortality.[4]

MEDICATION REACTIONS

Malignant Hyperthermia

Malignant hyperthermia (Chapter 434) is characterized by acute hyperpyrexia developing during or immediately after general anesthesia.[5] The channels that regulate the duration and amplitude of calcium efflux from the sarcoplasmic reticulum are the ryanodine receptors, which exist as three isoforms. Gain-of-function mutations affecting RyR1, the receptor expressed primarily in skeletal muscle, are present in 1 per 15,000 to 50,000 people and are associated with enhanced sensitivity to halothane and caffeine and with malignant hyperthermia and central core disease. More than 80 distinct mutations have been detected, and mutation of the adult skeletal muscle sodium channel, SCN4A, may also cause the syndrome. Patients with mutations predisposing to malignant hyperthermia function normally at resting conditions, but exposure to volatile anesthetics, including halothane, isoflurane, enflurane, desflurane, and sevoflurane, or exposure to a depolarizing muscle relaxant, succinylcholine, can precipitate life-threatening muscle contractures, increases in heart rate and body temperature, rhabdomyolysis, myoglobinuria, and metabolic acidosis. The mortality rate is 80% in untreated patients but about 5% with current treatment. Note that succinylcholine causes a release of myoglobin from muscle in small amounts even in normal

patients. Patients with malignant hyperthermia do not predictably respond to triggering agents, and some patients with malignant hyperthermia have had milder symptoms of malignant hyperthermia after the administration of nontriggering agents. Malignant hyperthermia now often occurs in muted forms, probably because of the decreased use of succinylcholine by anesthesiologists, the diagnostic awareness of malignant hyperthermia by anesthesiologists, the routine use of carbon dioxide monitors so that increases in end-expiratory carbon dioxide are detected quickly, and the availability of dantrolene. If malignant hyperthermia is suspected by obtaining a family history of adverse events with the administration of anesthesia or when a patient has a reaction suspicious for malignant hyperthermia, a muscle biopsy is usually obtained for in vitro contracture testing, which evaluates the muscle contracture responses to caffeine or halothane. Genetic investigations are also recommended, but malignant hyperthermia cannot be excluded on the basis of genetic testing alone because of the diversity of mutations and genes that can be involved in this syndrome. The Malignant Hyperthermia Association of the United States, www.mhaus.org, is available for information to the public, and all medical personnel can get information 24 hours every day on the malignant hyperthermia hotline, 1-800-MHHYPER or 1-800-644-9737.

Dantrolene is the drug of choice to prevent and to reverse the symptoms of malignant hyperthermia. Dantrolene decreases muscle sensitivity to caffeine, reduces the calcium release from the sarcoplasmic reticulum, and produces some muscle weakness. Dantrolene comes in 20-mg bottles and must be dissolved in sterile water; the recommended dose is 2.5 mg/kg given rapidly up to 10 mg/kg. Drug should be given every 5-10 minutes until symptoms subside. Other treatments for malignant hyperthermia include: discontinuing the use of any volatile anesthetics; hyperventilating the patient and administering 100% oxygen; administering bicarbonate for severe acidosis; controlling fevers; and maintaining a temperature below 39°C, without causing hypothermia by using iced fluids, surface cooling, and cooling of body cavities if necessary. Monitoring of temperature and vital signs, urinary output, muscle enzymes, glucose, coagulation studies, acid-base status, and gas exchange is recommended.

Two other rare congenital myopathies associated with mutations of the RyR1 include central core disease and multiminicore disease. Patients with central core disease present with infantile hypotonia; a muscle biopsy is needed for definitive diagnosis. Multiminicore disease is a nonprogressive congenital myopathy in which infants present with hypotonia, ophthalmoplegia, and arthrogryposis. These children develop scoliosis and eventually may require chronic ventilation. Avoidance of triggering agents is advised for these syndromes and for patients with other myopathies. Both malignant hyperthermia and central core disease are thought to be inherited as autosomal dominant diseases, but extensive genetic analysis has revealed overlapping phenotypes.

Monoamine Oxidase Inhibitors and Serotonin Toxicity

Anesthesiologists routinely ask if patients are taking a monoamine oxidase (MAO) inhibitor because of their many drug interactions with analgesics in perioperative patients. Also, serotonin toxicity has features similar to malignant hyperthermia and must be distinguished from it. Serotonin toxicity, characterized as a triad of neuromuscular hyperactivity (tremor, clonus, myoclonus, hyperreflexia, and pyramidal rigidity), autonomic hyperactivity (diaphoresis, fever, tachycardia, and tachypnea), and altered mental status (agitation, excitement, and confusion) can be precipitated by the coadministration of MAO inhibitors and selective serotonin reuptake inhibitors (SSRIs). Patients who are taking SSRIs have a higher overall perioperative mortality, a higher 30-day readmission rate, and a higher likelihood of bleeding.[6] Rigidity, increasing arterial carbon dioxide levels, and fever above 38.5°C are associated with life-threatening toxicity. Ecstasy, or 3,4-met hylenedioxymethamphetamine (MDMA), combined with MAO inhibitors, including moclobemide, can lead to fatalities because it acts as a serotonin releaser. Tramadol, used for pain relief, and venlafaxine, an antidepressant, act as serotonin releasers and are associated with toxicity when used in patients who are taking MAO inhibitors.

Anaphylaxis in the Perioperative Period

The incidence of life-threatening hypersensitivity reactions during anesthesia is 1 : 4000 to 1 : 25,000. Anaphylaxis is caused by immunoglobulin E (IgE)-mediated reactions (Chapter 253), whereas anaphylactoid reactions produce the same clinical picture but are not mediated by IgE. Anaphylaxis during anesthesia can present as cardiovascular collapse, airway obstruction, flushing, or edema of the skin, singly or in combination, so a careful history of any previous allergic reactions to medications and the nature of the reaction must be obtained by the anesthesiologist and other members of the perioperative team. Neuromuscular blocking agents, such as succinylcholine, and opioid analgesics can cause nonimmunologic release of histamine from mast cells and produce a similar clinical syndrome. Antibiotics, protamine, and blood transfusions (Chapter 177), all given routinely during operations, also can elicit a variety of systemic reactions. About 75% of perioperative hypersensitivity reactions appear to be due to muscle relaxants, especially rocuronium and vecuronium, with a mortality of 3 to 6%. In patients with apparent allergic reactions, skin testing is usually performed, and IgE levels are usually obtained to determine whether the patient had an allergic reaction to a perioperative medication.

Latex Allergies

For sensitized patients (Chapter 253), exposure to even low amounts of latex-containing particles is sufficient to induce a severe anaphylactic reaction. A latex-free operating environment, in which no latex gloves or latex accessories are used, is key in patients with known allergy. Skin prick tests with latex extracts should be considered in patients at high risk for latex allergy. Early aggressive treatment with epinephrine is critical if severe anaphylaxis occurs.

INTRAOPERATIVE MANAGEMENT

There are three general classes of anesthesia: general, regional, and monitored anesthesia care. The same drugs are often used for general anesthesia and monitored anesthesia care; achievement of the two different conditions requires knowledge of the pharmacokinetics of the drugs (Table 432-1).

General Anesthesia

General anesthesia can be achieved with a balanced drug regimen that induces a loss of consciousness, which can range from a deep sedation requiring only airway support to states requiring full ventilatory support because of weakness and loss of respiratory drive. Both intravenous and inhalational drugs can be used to induce and maintain general anesthesia. In contrast, monitored anesthesia care denotes a state in which patients can still control their airway, do not require ventilatory support, but are sleepy, have less pain, and may be amnestic.

Propofol

Propofol, an alkylphenol, is perhaps the most frequently used intravenous anesthetic for induction of anesthesia and is often used for maintenance of anesthesia during short procedures or to achieve deep sedation during monitored anesthesia care. It is lipid soluble and quickly cleared from the central

TABLE 432-1 COMMON ANESTHETIC APPROACHES FOR VARIOUS TYPES OF SURGERY

SURGERY ON INTRA-ABDOMINAL OR INTRATHORACIC ORGANS
Examples: cardiac surgery, lung resections, gastric bypass
General anesthesia usually administered because mechanical ventilation is often required
Drugs include premedication for anxiety with midazolam, general anesthesia with volatile anesthetics (desflurane, sevoflurane, nitrous oxide), neuromuscular blockade, and opioid analgesics*
Epidural anesthesia and analgesia also used; examples include ropivacaine, lidocaine, with fentanyl
SURGERY ON LIMBS
Examples: hip replacement, knee replacement, foot or arm surgery
Can perform with epidural or spinal anesthesia, depending on the limb. Examples of medications would include tetracaine, lidocaine, ropivacaine, and fentanyl or morphine.
Can perform axillary or scalene block; examples include lidocaine and ropivacaine
For postoperative pain control: can perform regional blocks that leave the catheter in place, including femoral nerve block, axillary nerve blocks
CATARACT SURGERY—LOCAL ANESTHESIA ON EYE WITH OR WITHOUT SEDATION
Examples of drugs used for sedation include midazolam and fentanyl

*Includes the use of opioids given intraoperatively with effects that extend into the postanesthesia care unit (PACU) or postoperative period, opioids given in the PACU, or opioids given or intended to be given after discharge from the PACU.

compartment, so it is rapidly eliminated even after long periods of continuous infusion. However, the clearance of propofol is changed by gender (men have lower clearance rates than women), size (children require higher doses), age (elderly patients have decreased clearance rates and experience increased effects with the drug), and narcotics, which decrease its clearance. Because of its predilection for causing apnea, propofol should be administered only by someone with expertise in airway management. Propofol also decreases arterial blood pressure, causes pain with injection, and can precipitate myoclonus. Large quantities of propofol can cause the propofol infusion syndrome, which is associated with cardiomyopathy, metabolic acidosis, skeletal myopathy, hyperkalemia, hepatomegaly, and lipemia. Despite these issues, propofol is frequently used because the recovery from propofol is within minutes, even after it is given as a prolonged continuous infusion, in contrast to the longer duration of drug effects seen after the administration of other intravenous sedatives.

Propofol is the preferred agent for healthy outpatients undergoing colonoscopy because it leads to earlier discharge and higher patient satisfaction compared with other agents.[■] When combined with midazolam, sedation is deeper, patient satisfaction is higher, and time to discharge is no longer,[■] even when administered by nonanesthesiologists.[■]

Midazolam

Midazolam, a benzodiazepine that produces muscle relaxation through a central mechanism, is hypnotic, sedative, anxiolytic, amnesic, and anticonvulsant. Its amnesic and anticonvulsant effects are mediated through α_1-subunit-containing γ-aminobutyric acid A ($GABA_A$) receptors, and the anxiolytic and muscle relaxation are mediated through α_2-subunit-containing $GABA_A$ receptors. Only 20% receptor occupancy is needed to produce anxiolysis, whereas unconsciousness requires 60%. Long-term administration of benzodiazepines produces tolerance, which appears to decrease receptor binding and function. Benzodiazepines cause dose-related depression of the respiratory system, with a peak effect at 3 minutes and significant depression persisting for 60 to 120 minutes. The rate of administration of the drug affects the onset of depression: the faster the drug is given, the quicker the respiratory depression occurs. Benzodiazepines and opioids appear to produce additive respiratory depression, including apnea. Unlike propofol, benzodiazepines used alone decrease blood pressure only modestly. Other drugs, particularly drugs that affect the cytochrome P-450 3A4 enzyme (including azole antifungals, human immunodeficiency virus [HIV] protease inhibitors, and calcium-channel blockers), affect the clearance of midazolam and prolong its half-life significantly. There are several reports of prolonged amnesia in HIV patients who received midazolam for conscious sedation. Midazolam also has an active metabolite and is often associated with delirium in elderly patients (Chapter 28), perhaps because it impairs both implicit and relational memory.

Opioids

Opioids are classified as naturally occurring (morphine, codeine), semisynthetic (heroin), and synthetic (methadone, fentanyl, remfentanil). They can be administered both intravenously and in the neuraxial space (epidural or spinal). There are four opiate receptors (mu, kappa, delta, and nociceptin receptors), which are G protein–coupled receptors. Chronic exposure to agonists leads to cellular adaptation mechanisms that probably are involved in tolerance, dependence, and withdrawal. Clinically, mu agonists are used almost exclusively; mu agonists include morphine, fentanyl, and meperidine. Opioid analgesics are administered because they relieve pain, but they have other important effects, including respiratory depression, decreased gastric emptying, nausea and vomiting, sedation, constipation, pruritus, dependence, and tolerance, when given repeatedly. When opioids are given with propofol or benzodiazepines, there is a synergistic depressive effect on respiration, hence the rationale for monitoring patients who receive medications for conscious sedation.

Ketamine

Ketamine is unique among the intravenous agents because it has analgesic properties and decreases tolerance to opiates. Ketamine produces dose-related analgesia, which may be profound even when patients can keep their eyes open, breathe spontaneously, and protect their own airway with conserved swallowing and cough reflex. Side effects include increased lacrimation, salivation, and muscle tone. Ketamine increases cerebral blood flow, can increase seizure activity, and can produce undesirable psychological reactions; these side effects are dose related and may be minimized by the concomitant use of benzodiazepines. Ketamine is also a bronchial smooth muscle relaxant and can prevent experimentally induced bronchospasm. Ketamine is usually associated with an increase in blood pressure, heart rate, and cardiac output. These features make ketamine a useful drug for sedating patients with hemodynamic instability.

Dexmedetomidine

Dexmedetomidine is a highly selective α_2-agonist that is associated with less respiratory depression and more cooperative behavior than is propofol. Dexmedetomidine also causes hypnosis, analgesia, sympatholysis, and inhibition of insulin secretion. Dexmedetomidine induces sedation with a respiratory pattern and electroencephalographic changes similar to natural sleep. Even high concentrations of dexmedetomidine are associated with preservation of spontaneous respiration; however, when dexmedetomidine is administered in combination with sympatholytic or cholinergic agents, there is a high risk for extreme bradycardia and sinus arrest. Dexmedetomidine is associated with less amnesia than are benzodiazepines. Although propofol and benzodiazepines commonly have been used in critically ill patients to achieve sedation for procedures or for maintenance of mechanical ventilation, dexmedetomidine appears to have significant advantages over benzodiazepines because it causes less delirium and decreases the time that critical care patients spend on ventilators.[■]

Volatile Anesthetics

Volatile (inhalational) anesthetics include desflurane, sevoflurane, isoflurane, and nitrous oxide, as well as halothane, which now is rarely used in the United States. Inhaled anesthetics are absorbed through the respiratory epithelium and mucous membranes of the respiratory tract, and they are excreted mainly by exhalation. Access to the circulation is almost instantaneous, owing to the large pulmonary surface area. The pharmacologic effects of inhaled anesthetics depend primarily on alveolar ventilation, the ventilation-perfusion ratio, coadministered gases, gas flow, and the physicochemical properties of the anesthetic gas rather than on the quantity of drug administered, the extent and rate of absorption, protein binding, excretion, secretion, or metabolism. Based on the available evidence, no inhalational agent appears to be superior to any other.

All inhalational agents, with the exception of nitrous oxide, cause dose-dependent cardiovascular depression. Severe hepatotoxicity, which led to the discontinuation of the use of chloroform, carbon tetrachloride, and trichloroethylene anesthetics, is seen as fatal hepatic necrosis with in 1 in 10,000 halothane anesthetics. This problem appears to occur much less frequently with isoflurane and desflurane. Mild halothane hepatoxicity is self-limited and can occur with a single exposure, whereas fulminant halothane hepatitis occurs only after multiple exposures to the drug, has a high mortality rate (50%), and is associated with antibodies to halothane-altered antigens.

Nitrous oxide, which is the only nonhalogenated agent still used, is not metabolized in human tissues. It irreversibly oxidizes the cobalt atom of vitamin B_{12}, thereby inhibiting the activity of the cobalamin-dependent enzyme methionine synthase. Individuals with vitamin B_{12} deficiency or with mutations of methionine synthase may be at risk for neurologic injury from nitrous oxide, which should not be used in patients at risk. Exposure to high concentrations of more than 10^3 ppm may be associated with an increased incidence of abortions and decreased fertility, so exposure should be avoided in patients and personnel at risk. Nitrous oxide is safe in major non-cardiac surgery.[■] General anesthesia can be achieved only by giving combinations of drugs along with nitrous oxide to achieve the desired effects. Based on the available evidence, no one general anesthetic appears to be superior to any other.

Neuromuscular Blockers

Neuromuscular blockers are used to paralyze muscles to facilitate endotracheal intubation and mechanical ventilation, to decrease shivering during induced hypothermia, or to improve conditions for optimal surgery. Succinylcholine causes prolonged depolarization of the neuromuscular junction, thereby resulting in failure to generate an action potential. Within 9 to 13 minutes after 1 mg/kg of succinylcholine, 90% of muscle strength is restored. The very rapid onset and rapid return of muscle function make succinylcholine a useful drug for difficult intubations. Side effects of succinylcholine include hyperkalemia, myalgia, masseter spasm, sinus bradycardia and nodal rhythms, and increased intraocular pressure.

Most of the other neuromuscular drugs used by anesthesiologists are nondepolarizing in that they compete with acetylcholine for the neuromuscular junction and can be reversed by increasing the quantity of acetylcholine.

These drugs are categorized by their chemical makeup: steroidal compounds, benzylisoquinolinium compounds, and other chemical compounds. Clinically, a drug is often chosen for its duration of action. Intermediate agents, which act for 20 to 50 minutes and are used most frequently, include vecuronium, rocuronium, atracurium, and cisatracurium. These drugs have different routes of metabolism, so the choice of agent depends in part on the presence of coexisting disease. Sugammadex, a recently developed direct antagonist to the neuromuscular blocking agents, is not approved for use in the United States at the time of this writing.

The chronic administration of neuromuscular blocking agents is associated with prolonged paralysis, particularly in patients given concomitant steroids. Other notable interactions with nondepolarizing agents include that antibiotics can increase neuromuscular blockade; magnesium sulfate potentiates neuromuscular blockade; lithium can potentiate neuromuscular blockade with succinylcholine and with pipecuronium; and antiepileptic drugs cause resistance to nondepolarizing muscle blockade so that larger doses must be administered to achieve paralysis; and patients receiving anticonvulsants have accelerated recovery from neuromuscular blockade.

Regional Anesthesia

Regional anesthesia involves the deposition of local anesthetics near nerves, including the deposition of local anesthetics in the epidural space and into the cerebral spinal fluid (CSF). Local anesthetics, which are aminoesters or aminoamides, affect cardiac function, as well as central nervous system function when administered systemically.

The binding of the local anesthetic to the sodium channels in the axoplasm prevents opening of the channels and conduction of nerve impulses. The rates of onset and recovery from nerve blockade are controlled by the diffusion of the local anesthetic into and out of the whole nerve.

Examples of regional anesthesia include neuraxial techniques, the deposition of local anesthetics near the brachial plexus to anesthetize the arms (axillary or intrascalene blocks), deposition near the femoral or sciatic nerves to anesthetize the legs, deposition near ulnar or radial nerves for lower arm blocks, deposition near the pudenal nerves for groin procedures, and deposition of local anesthesia in the caudal space for groin surgeries. Dentists employ this technique frequently when they inject local anesthesia near various nerves in the oral cavity. Many surgeries, including carotid surgery and the placement of fistulas for dialysis, can be performed with regional anesthesia. Regional anesthetics may also require supplementation with sedation or general anesthesia.

The dangers of regional anesthesia include the injection of local anesthesia into the systemic circulation. Systemic toxicity is manifested as convulsions and respiratory depression, which can require assisted ventilation. Tinnitus, visual and auditory disturbances, and dizziness are signs of milder central nervous toxicity. Cardiac toxicity can be manifested by decreases in heart rate, prolonged conduction times, and negative inotropic effects. Bupivacaine toxicity is associated with ventricular fibrillation. Intralipid 20% at various doses (1.5 ml/kg rapid bolus [~100mL in average adult] followed by infusion of 0.25 ml/kg/min for 10 minutes) has been reported in case reports and in animal studies to reverse these toxic effects, although the optimal dosing has yet to be determined. Furthermore, the prolonged duration of many of the local anesthetics may require the institution of cardiopulmonary bypass until the drugs are metabolized.

Neuraxial (Spinal and Epidural) Anesthesia and Analgesia

Spinal anesthesia is the instillation of local anesthetics into the CSF. Epidural anesthesia is the instillation of larger volumes of local anesthetics into the epidural space, which is the potential space that exists just before the CSF. Spinal anesthesia is associated with an increased incidence of headache in younger patients, so epidural anesthesia is often used in younger patients. Complications of epidural and spinal anesthesia and analgesia include failed blocks, postdural puncture headaches, and toxicity from the local anesthetics. Another major concern of neuraxial anesthesia is that patients on antiplatelet agents may develop epidural hematomas, although epidural hematoma remains a rare event, occurring in fewer than 1 in 150,000 operations even in the presence of potent antiplatelet agents. Other more rare complications of epidural and spinal anesthetics, in addition to the effects of local anesthesia outlined previously, include intracranial subdural hematoma, transverse myelitis, hypotension, and cardiac arrest.

Postoperative epidural analgesia, by which either local anesthesia or local anesthesia and narcotics are instilled into the epidural space for postoperative pain control, is associated with superior pain control, lower doses of opioids, improved bowel mobility, slightly decreased length of stay in the intensive care unit, and a slight decrease in the requirement for mechanical ventilation.[A]

GENERAL VERSUS REGIONAL ANESTHESIA

The decision regarding what type of anesthesia should be administered often depends on the requirements of the surgery. For example, laparoscopic surgery requires general anesthesia because the insufflations of gases impair the ability to breathe adequately. General anesthesia is also required for surgeries on the airway or thorax because mechanical ventilation is usually needed to sustain adequate respiration. Low tidal volume and low positive end-expiratory pressure are preferred.[A] Procedures that do not allow any movement (e.g., precise procedures in the brain) often require general anesthesia and paralysis. For patients in whom the intraoperative technique could include general anesthesia, regional anesthesia, or a combination of the two, regional anesthesia may minimize pulmonary complications, but the data are conflicting.

Side effects of general anesthesia depend on the drugs used to achieve anesthesia, whether neuromuscular blockade is administered, and whether mechanical ventilation is used. Complications of endotracheal intubation include local pain, trauma to the airway, swelling, vocal cord paralysis, increased bronchospasm, and death from improper placement. Volatile anesthetics are associated with postoperative atelectasis (Chapter 90), whereas regional anesthesia helps preserve respiratory dynamics. Postoperative cognitive dysfunction (Chapter 28) does not seem to depend on the type of anesthesia administered.

NAUSEA AND VOMITING

Postoperative nausea and vomiting are more likely with volatile anesthetics but also are common when perioperative opioids are administered. Prophylactic ondansetron, dexamethasone, and droperidol each reduce postoperative nausea and vomiting, independently, by about 26%, with the main predictor for efficacy being the patient's risk for nausea and vomiting.[A] It should be noted that droperidol has received a "black box" warning from the U.S. Food and Drug Administration, so it is not used very often in the United States. Total intravenous anesthesia with propofol reduces postoperative nausea and vomiting by only about 20%, often because narcotics are still administered. The use of spinal or epidural anesthesia may decrease the incidence of nausea and vomiting. In addition to general anesthesia, risk factors for postoperative nausea and vomiting include female gender, a prior history of nausea and vomiting, a history of motion sickness, nonsmoking, and intended administration of opioids for postoperative analgesia. If three or more risk factors are present, patients generally are recommended to receive at least two prophylactic pharmacologic antiemetic agents of different classes (e.g., selected among ondansetron or another 5-HT_3 antagonist, droperidol, dexamethasone, scopolamine, or phenothiazides) preoperatively for the prevention of nausea and vomiting.

Grade A References

A1. Wang D, Chen C, Chen J, et al. The use of propofol as a sedative agent in gastrointestinal endoscopy: a meta-analysis. *PLoS ONE*. 2013;8:e53311.

A2. Wang D, Wang S, Chen J, et al. Propofol combined with traditional sedative agents versus propofol-alone sedation for gastrointestinal endoscopy: a meta-analysis. *Scand J Gastroenterol*. 2013;48:101-110.

A3. Molina-Infante J, Dueñas-Sadornil C, Mateos-Rodríguez JM, et al. Nonanesthesiologist-administered propofol versus midazolam and propofol, titrated to moderate sedation, for colonoscopy: a randomized controlled trial. *Dig Dis Sci*. 2012;57:2385-2393.

A4. Riker RR, Shehabi Y, Bokesch PM, et al. Dexmedetomidine vs midazolam for sedation of critically ill patients. *JAMA*. 2009;301:489-499.

A5. Myles PS, Leslie K, Chan MT, et al. The safety of addition of nitrous oxide to general anaesthesia in at-risk patients having major non-cardiac surgery (ENIGMA-II): a randomised, single-blind trial. *Lancet*. 2014;384:1446-1454.

A6. Popping DM, Elia N, Van Aken HK, et al. Impact of epidural analgesia on mortality and morbidity after surgery: systematic review and meta-analysis of randomized controlled trials. *Ann Surg*. 2013;259:1056-1067.

A7. Hemmes SN, Gama de Abreu M, Pelosi P, et al. High versus low positive end-expiratory pressure during general anaesthesia for open abdominal surgery (PROVHILO trial): a multicentre randomised controlled trial. *Lancet*. 2014;384:495-503.

A8. Apfel CC, Korttila K, Abdalla M, et al. A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med*. 2004;350:2441-2451.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bainbridge D, Martin J, Arango M, et al. Perioperative and anaesthetic-related mortality in developed and developing countries: a systematic review and meta-analysis. *Lancet*. 2012;380:1075-1081.
2. Pearse RM, Moreno RP, Bauer P, et al. Mortality after surgery in Europe: a 7 day cohort study. *Lancet*. 2012;380:1059-1065.
3. De Jong A, Molinari N, Terzi N, et al. Early identification of patients at risk for difficult intubation in the intensive care unit: development and validation of the MACOCHA score in a multicenter cohort study. *Am J Respir Crit Care Med*. 2013;187:832-839.
4. Ramachandran SK, Nafiu OO, Ghaferi A, et al. Independent predictors and outcomes of unanticipated early postoperative tracheal intubation after nonemergent, noncardiac surgery. *Anesthesiology*. 2011;115:44-53.
5. Riazi S, Larach MG, Hu C, et al. Malignant hyperthermia in Canada: characteristics of index anesthetics in 129 malignant hyperthermia susceptible probands. *Anesth Analg*. 2014;118:381-387.
6. Auerbach AD, Vittinghoff E, Maselli J, et al. Perioperative use of selective serotonin reuptake inhibitors and risks for adverse outcomes of surgery. *JAMA Intern Med*. 2013;173:1075-1081.

REVIEW QUESTIONS

1. A 75-year-old healthy woman asks your opinion if she should undergo an elective back operation. She is concerned about her perioperative risks for mortality. You tell her that her that perioperative mortality is mostly determined by which of the following?

A. Patient's underlying medical problems
B. The surgical procedure
C. The anesthetic
D. The volume of procedures done at a hospital

Answer: A.

2. A 55-year-old woman comes to you for a preoperative assessment. She has a long history of refractory depression. You inquire whether she is on monoamine oxidase inhibitors, so your list of medications should include which of the following?

A. Selegiline
B. Moclobemide
C. Pargyline
D. Isoniazid

Answer: All are correct. See Gillman PK. Monoamine oxidase inhibitors, opioid analgesics and serotonin toxicity. *Br J Anaesth.* 2005;95:434-441.

3. You are treating a 35-year-old man who is on chronic opioids for pain. He comes into the hospital with a broken limb, and his pain is not adequately treated despite increasing doses of opioid analgesics. You suggest ketamine because it will do all of the following except which one?

A. Decrease the need for opioids
B. Provide analgesia
C. Improve sedation
D. Decrease respiration

Answer is D. Ketamine preserves respiration.

433

POSTOPERATIVE CARE AND COMPLICATIONS

DONALD A. REDELMEIER

POSTOPERATIVE CARE

Overview

Postoperative medical complications are common, potentially fatal, and variable across different settings. Large national studies show about a two-fold difference in risk for mortality between high- and low-ranked hospitals. However, analyses disagree about how much these differences reflect a greater incidence of each complication (failure of prevention around the time of surgery), a heightened lethality of each complication (failure to rescue in the aftermath of surgery), or the differences in the severity of disease or surgical skill. Regardless of the explanation, the purpose of medical consultation is to relieve human suffering by the prevention, detection, and correction of postoperative complications. The main constraint is that the consultant often has limited ongoing direct contact with the patient before or after the perioperative interval.

Effective Teamwork

The medical consultant (Chapter 430) in the postoperative setting must have both a knowledge of medicine and an appreciation of the team psychology that can improve the outcomes of patients.[1] In contrast to other settings, the internist is not the team leader, often does not maintain an ongoing relationship with the patient, and does not have the authority of the most responsible physician. Moreover, patients may be dispersed across diverse surgical services, each with its own orientation and culture. The challenges of coordination and communication are enormous, particularly given the multiple other health care professionals involved in complex surgical cases. Considerable tact is often needed to avoid antagonizing the surgeon, disrupting the team's dynamics, or inducing a cascade of cumbersome inopportune testing. The development and use of safety checklists can be an effective way for teamwork to improve outcomes.[2]

Focusing on Recovery

Facilitating the patient's recovery from surgery differs conceptually from managing patients with acute exacerbations of chronic disease. In the postoperative setting, many therapies need to be stopped at some point because the patient has recovered, such as discontinuing a urinary catheter because the patient can now void spontaneously or discontinuing a major tranquilizer because the patient is now oriented and coherent. Discontinuation of many other interventions requires substantial judgment, such as the decision when to discontinue intravenous access, supplemental oxygen, and intermittent laxatives. Much depends on experience and reconsideration of an individual patient's situation on a regular basis.

Reading Anesthesia Records

A focused review of the anesthesia record is essential because the consultant is rarely present during the operation. Perhaps the most basic information to identify is the date of surgery because the time elapsed helps in interpreting the patient's current state of recovery. Sometimes the date is not immediately evident if more than one surgery has been performed, a planned operation was canceled, or misquotations have arisen. Data about the duration of surgery, type of anesthesia (e.g., regional, spinal, or general [Chapter 432]), and major intraoperative events help establish reasonable expectations about the future course as well as the possibility of specific complications (e.g., epidural hematoma after spinal anesthesia). Sharing some of the basic data with the patient is often helpful because many individuals either benefit from repetition or are not otherwise informed.

Patterns of Mistakes

Medical errors (Chapter 12) that arise in postoperative care often seem mundane in retrospect yet can be lethal if undetected. Some patterns of mistakes have the feature of "double trouble," such as when a patient has both a potassium level of 2.0 mEq/L and an international normalized ratio of 2.0 but care focuses on only one of these abnormalities. Other mistakes occur because a single problem arises at an awkward moment, such as a patient in whom acute dyspnea develops when another patient is having a seizure. Still other mistakes relate to the fallibility of human memory and attention, such as when a normal blood glucose value in the morning leads clinicians to presume that the level is still normal at night. These errors can result in substantial harm, failures of clinicians to learn from past mistakes, and unprofessional reactions related to embarrassment. None of these patterns are unique to postoperative care, yet the fast and unfamiliar terrain of surgical settings can make even simple mistakes difficult to avoid.

Checking Orders

The first method for reducing errors after surgery is to check the postoperative orders already written for the patient. Such double-checking is a tedious task, and clinicians often direct insufficient attention to this review in the faulty belief that most of the work is already done. Ironically, checking orders written by another clinician requires more than customary attention because of the challenges of following someone else's legibility, sequencing, and preferences. The set of orders may need to be read twice: once for errors of commission (e.g., a calcium-channel blocker ordered at the wrong dose) and once for errors of omission (e.g., a β-blocker inadvertently not reordered after surgery). A classic mistake on postoperative orders is failure to follow through on interventions initiated immediately before surgery (e.g., delirium tremens prophylaxis). A particularly vexing issue is the need for repeated rechecking on subsequent days (e.g., new orders for sedative drugs).

Recommended Prophylaxis

Some complications are sufficiently frequent and serious that routine prophylaxis is merited in the postoperative setting. For example, systemic anticoagulation is indicated for most patients at risk for postoperative deep venous thrombosis (Chapters 38 and 81). The 2012 American College of Chest Physicians evidence-based clinical practice guidelines provide specific recommendations for both nonorthopedic and orthopedic surgery patients. Gastric acid suppression (Chapter 139) is justified for patients at high risk for postoperative gastric bleeding. Parenteral antibiotics are indicated for patients undergoing prosthetic joint replacement. In contrast, antibiotic prophylaxis is indicated only for selected patients who are at high risk for endocarditis (Chapter 76). The optimal method for gauging whether a patient is at high risk for each complication is contentious and thereby leads to variation in practice patterns across different settings.

Future Prevention

A postoperative consultant who maintains communication, facilitates the patient's recovery, and avoids postoperative mistakes also has the chance to initiate medical interventions for general medical care. Such opportunities for prevention might include influenza vaccination, colon cancer screening, and cholesterol reduction. The main advantage of such comprehensive care is that it conforms to the ideal of providing all services possible to the individual. The main disadvantage of such comprehensive care is the potential for creating unintended chaos, confusion, or misquotation (Chapter 430). Such unintended consequences distract the surgical team from the primary goal and also carry some risk for side effects at a time when the patient is trying to recover from surgery. Many effective postoperative consultants will defer such opportunities for prevention to the physicians who assume long-term responsibility for the patient's care.

COMPLICATIONS

Symptoms

Chest Pain

Chest pain is a common problem after surgery and has an extensive differential diagnosis (Chapter 51). In the postoperative setting, the immediate consideration is an acute ischemic myocardial event. The diagnosis of a perioperative myocardial infarction (MI) differs somewhat from community-acquired MI (Table 433-1). Interpretation of a patient's symptoms, examination findings, and electrocardiogram is often problematic because of changes related to surgery and anesthesia. Instead, diagnosis is heavily dependent on biomarkers, such as an elevated troponin level, especially because many postoperative MIs are painless. Among patients undergoing noncardiac surgery, the peak postoperative troponin level during the first 3 days after surgery is significantly associated with 30-day mortality, even if patients do not have any other evidence of an acute MI.[3] Management priorities include supplemental oxygen, heart rate control, and correction of severe anemia. Thrombolysis is often contraindicated, but percutaneous coronary intervention may

be considered. In the absence of data from randomized trials, other therapies, such as aspirin, clopidogrel, nitrates, statins, and angiotensin-converting enzyme inhibitors, should be used on a case-by-case basis (Chapter 73).

Dyspnea

Shortness of breath (Chapter 83) after surgery has an extensive differential diagnosis (Table 433-2). The three key considerations are fluid overload/heart failure (Chapter 58), pulmonary embolism (Chapter 98), and air space disease (a continuum encompassing atelectasis [Chapter 90], bronchitis [Chapter 96], aspiration [Chapter 94], mucous plugging, and pneumonia). Distinguishing among these considerations requires focusing on the speed of onset, timing relative to surgery, vital signs, findings on oximetry, and physical examination findings (Chapter 83). Fluid overload is most commonly seen soon after the cessation of positive-pressure ventilation or vasodilating analgesia. It is also common 3 to 5 days postoperatively when fluid that had been "third spaced" is mobilized into the intravascular compartment. Interventions that are safe in most situations include administration of oxygen and withholding of sedation. The use of continuous positive airway pressure can reduce the rate of reintubation in hypoxemic postoperative patients. Other interventions that will be helpful or harmful, depending on the specific situation, include diuretics, opioids, elaborate medical imaging, and vigorous physiotherapy.

TABLE 433-1 CRITERIA FOR DIAGNOSIS OF POSTOPERATIVE MYOCARDIAL INFARCTION

The diagnosis of perioperative MI requires any one of the following criteria.
Criterion 1: A typical rise in the troponin level or a typical fall in an elevated troponin level detected at its peak after surgery in a patient without a documented alternative explanation for an elevated troponin level (e.g., pulmonary embolism) or a rapid rise and fall in CK-MB only if troponin measurement is unavailable.* This criterion requires that one of the following criteria be met: Ischemic signs or symptoms (e.g., chest, arm, or jaw discomfort; shortness of breath; pulmonary edema) Development of pathologic Q waves on an ECG Changes on an ECG indicative of ischemia Coronary artery intervention New or presumed new cardiac wall motion abnormality on echocardiography or new or presumed new fixed defect on radionuclide imaging
Criterion 2: Pathologic findings of acute or healing MI
Criterion 3: Development of new pathologic Q waves on an ECG if troponin levels were not obtained or were obtained at times that could have missed the clinical event

*Because CK-MB is both less sensitive and less specific than troponin levels in the perioperative setting than in other settings, it should be used for diagnostic purposes only when troponin levels are not obtainable.
CK-MB = creatine kinase MB isoenzyme; ECG = electrocardiogram; MI = myocardial infarction.
Reproduced with permission from Devereaux PJ, Goldman L, Yusuf S, et al. Surveillance and prevention of major perioperative ischemic cardiac events in patients undergoing noncardiac surgery: a review. CMAJ. 2005;173:779-788.

Anorexia

Loss of appetite (Chapter 132) after surgery has an extensive differential diagnosis that can be narrowed substantially if the patient was eating properly before surgery. The immediate priority is to search for and correct underlying contributors. Oral, enteral, or parenteral support is not the priority initially, although such support may become necessary. Drug toxicity is a particularly common, easily detected when considered, and a rapidly reversible contributor to postoperative anorexia. Anatomic abnormalities are usually evident by medical imaging studies. Other common metabolic contributors include abnormalities in electrolytes, calcium, phosphorus, and magnesium. Acalculous cholecystitis (Chapter 155) is an important postoperative complication that must be considered in a patient with right upper quadrant tenderness.

Vomiting

Vomiting is the extreme form of nausea in the postoperative setting, and the two symptoms share the same differential diagnosis. In most patients, vomiting is unexpected and merits immediate attention. Initial management is to ensure that the patient's airway is protected, to discontinue oral medications (and find parenteral substitutes if necessary), and to consider insertion of a nasogastric tube. In patients after gastrointestinal surgery, the priority considerations include the possibility of an anastomotic leak, peritoneal abscess, and other anatomic abnormality. In patients after operations on more remote parts of the body, the priorities are emetogenic medications (such

TABLE 433-2 DISTINGUISHING AMONG COMMON CAUSES OF ACUTE POSTOPERATIVE DYSPNEA

	PULMONARY AIR SPACE	FLUID OVERLOAD/HEART FAILURE	PULMONARY THROMBOEMBOLISM
CHARACTERISTICS OF TIMING			
Days since surgery	1-7 days	0-5 days	5-28 days
Speed of onset	1-3 days	1-24 hours	1-5 minutes
PREVIOUS HISTORY			
Previous lung disease	++		
Previous heart failure		++	
Previous venous thrombosis			++
ABNORMAL VITAL SIGNS			
Temperature			
Heart rate	+	+	+
Blood pressure	+	+	++
Respiratory rate	+	++	+
Oximetry	++	+	+
PHYSICAL EXAMINATION			
Jugular venous distention		+	+
Pulmonary rales	+	++	
S_3 gallop		++	
RESPONSE TO TREATMENT			
Oxygen	+	+	+
Anticholinergic bronchodilators	+	+	+
Withdrawal of sedatives	+	+	+
Aggressive physiotherapy	++		
Diuretics/afterload reduction		++	

as postoperative chemotherapy), gastroparesis associated with autonomic neuropathy, and fecal impaction. Multiple antinausea medications are available for symptomatic relief (e.g., prochlorperazine, ondansetron, dexamethasone, droperidol) and act in an additive manner when they are used in combination (e.g., prochlorperazine 5 mg intramuscularly plus ondansetron 4 mg intramuscularly). If no reversible contributor is identified, the default diagnosis is prolonged idiopathic ileus, and a therapeutic trial of intravenous neostigmine can be considered (e.g., neostigmine 2.5 mg intravenously during 5 minutes).

Diarrhea

Diarrhea (Chapter 140) is relatively rare after surgery and involves a limited number of possibilities if the patient's bowel movements were normal before surgery. In such cases, the situation represents an acute-onset diarrhea that is usually secretory in nature. The immediate priority is to exclude toxic megacolon, which is caused by overgrowth of toxigenic *Clostridium difficile* and is a potential emergency. Clinical evaluation for toxic megacolon requires assessment for tachycardia, hypotension, delirium, and other signs of sepsis rather than waiting on initial stool studies for confirmation of infection with *C. difficile*. Risk factors for antibiotic-associated diarrhea include advanced age, use of broad-spectrum antibiotics (e.g., third-generation cephalosporins), and unknown host susceptibility factors (e.g., past episodes of pseudomembranous colitis). A definitive diagnosis is frequently never established, and treatment focuses on feeding the patient a lactose-free diet while avoiding intestinal paralytics. Complete resolution is typical, provided adequate fluid and electrolyte levels are maintained.

Weakness

Generalized weakness after surgery is almost inevitable, but focal weakness may reflect nerve damage caused by intraoperative positioning (e.g., damage to the facial nerve after carotid endarterectomy) and rarely indicates a new intracranial event (e.g., intracerebral bleeding [Chapter 408] secondary to anticoagulation). Neurologic deficits are often overlooked during the initial postoperative interval and may become apparent only after the patient has regained strength elsewhere in the body. Conversely, new deficits that are evident early after surgery and resolve rapidly thereafter may reflect an old stroke that was fully compensated during less stressful circumstances. Medical imaging of the brain is worthwhile if no explanation is apparent on initial assessment. Nonfocal weakness commonly responds to physical therapy.

Delirium

Changes in mental status after surgery are common, especially in elderly patients (Chapter 28), and can be remarkably difficult to correct. The immediate priorities are to determine whether the impairment is acute or chronic and to detect easily reversible contributors (such as infection, hypoglycemia, and alkalosis). A complete assessment is often unnecessary if the patient had normal mental status before surgery because many dementia syndromes are thereby excluded (such as vitamin B_{12} deficiency, tertiary syphilis, and Alzheimer's disease). Immediate treatment usually focuses on discontinuation of medications such as anticholinergics, narcotics, and tranquilizers. The patient may also benefit from the continuous presence of friends and family members, who can provide frequent orientation and constant attention. A sense of patience is necessary because delirium rarely resolves instantly. The benefit of low-dose neuroleptics (e.g., risperidone, 0.5 mg by mouth twice daily) remains uncertain. Delirium after surgery is associated with a significant decline in cognitive ability during the next year, with a trajectory characterized by an initial decline and prolonged impairment.[4]

Seizures

The development of uncontrolled seizures (Chapter 403) after surgery is rare. The immediate priorities are to exclude status epilepticus, to uncover any past history of a seizure disorder, and to identify provocative factors. Neurosurgical patients typically undergo a standardized treatment protocol, including steroids and imaging. Other conditions that can cause abnormal movements must be excluded, such as septic rigors, delirium tremens, Parkinson's disease, major psychopathology, hypothermic shivering, and hypercapnic asterixis. Additional considerations include detection and correction of any underlying metabolic abnormalities, such as hypocalcemia, hypoxemia, hyponatremia, hypophosphatemia, and drug toxicity. Treatment focuses primarily on reversing the underlying precipitating cause and providing nonspecific care with benzodiazepines, phenytoin, and ongoing monitoring.

Signs

Hypertension

Hypertension may reflect a variety of disorders and must be treated in a manner that neither overreacts nor underreacts to the situation. Hypertension is particularly common after neurosurgical procedures or carotid endarterectomy. The initial assessment focuses on whether the patient has chronic hypertension based on the past history, current electrocardiogram, or findings on funduscopy. Other potential causes include undertreated pain, agitated delirium, fluid overload, alcohol withdrawal, and inadvertent discontinuation of chronic antihypertensive medications. In uncertain cases, systemic analgesia is often helpful, along with nitrates (e.g., nitroglycerin, 0.4 mg/hour transdermally) and β-blockers (e.g., metoprolol, 5 mg intravenously). The major complication of treatment is the potential for overcorrection and inadvertent hypotension; such errors are particularly common in patients with no evidence of past hypertension.

Hypotension

Hypotension (Chapter 8) after surgery is generally an emergency, and the immediate concern is internal bleeding, especially after intra-abdominal operations or when anticoagulation is used to prevent venous thrombosis. The initial stages of hypotension are frequently unrecognized because of biologic stress responses by patients, psychological denial by clinicians, and misattribution to the concurrent use of analgesia. Early hypotension is particularly easy to overlook if the patient has coexisting chronic hypertension and the apparently "normal" blood pressure is dismissed as unremarkable. Treatment usually entails volume supplementation,[5] vasopressors as needed (Chapters 107 and 108), serial assessments, and a search for underlying causes. An extensive differential diagnosis sometimes needs to be considered if no anatomic cause related to surgery is evident (Chapter 106). Routine use of pulmonary artery catheters to guide therapy is not helpful, whereas prompt echocardiography can almost always find the cause of severe hemodynamic instability.

Tachycardia

Tachycardia after surgery can be caused by myriad arrhythmias (Chapters 64 and 65) and may contribute to postoperative cardiac ischemia. Distinguishing between newly detected and newly incident tachycardia can sometimes be accomplished by determining whether the patient does or does not complain of palpitations. An initial assessment also requires review of the electrocardiogram to distinguish atrial fibrillation from other disorders. The goal of treatment is to identify and to correct precipitating factors, such as pain, blood loss, hypoxia, electrolyte abnormalities, fluid overload, volume depletion, pulmonary embolism, and drug withdrawal. Most arrhythmias respond to correction of the underlying abnormality. Specific antiarrhythmic treatment, when needed, is generally similar to that used in the nonoperative setting (Chapters 64 and 65). For atrial fibrillation, anticoagulation is sometimes contraindicated; in such situations, cardioversion within 48 hours merits consideration. Postoperative atrial fibrillation portends a 2-fold higher long-term risk of ischemic stroke.[6]

Fever

Fever (Chapter 280) after surgery is common, frequently perplexing, and often multifactorial. Worrisome possibilities include transfusion reactions (Chapter 177), hospital-acquired pneumonia (Chapter 97), urinary tract infection (Chapter 284), line sepsis (Chapter 282), and wound infection. In many cases, no definitive cause is found, the patient recovers spontaneously, and the default diagnosis is atelectasis. Detailed evaluation, when necessary, requires culture of blood, urine, and the surgical site to identify specific microbiologic organisms. Selection of empirical antibiotics is usually based on local practice patterns and hospital ecology, with the disadvantages of breeding resistant organisms. Hydration, nutrition, and general supportive care are important yet frequently neglected needs of patients with prolonged elevations in body temperature. Selective decontamination of the digestive tract and oropharynx appears to be beneficial but chlorhexidine is not.

Edema

Peripheral edema (Chapter 51), which is often first noticed by nursing staff after surgery, is rarely life-threatening unless it is treated with excessive diuretics. The cause is usually multifactorial and includes increased hydrostatic pressure (including heart failure and gravity from intraoperative positioning), decreased oncotic pressure (related to hypoalbuminemia from decreased

liver production or increased losses), and capillary leak (potentially caused by medications or tissue reactions). Treatment focuses on correction of underlying abnormalities, maintenance of nutrition, judicious use of diuretics, monitoring of renal function, provision of systemic anticoagulation against deep venous thrombosis, and efforts toward mobilization of the patient. A low-salt diet, afterload reduction, and aldosterone antagonists (e.g., oral spironolactone, 25 mg once daily) may be helpful in patients in whom heart failure (Chapter 59) is the dominant mechanism.

Laboratory

Leukocytosis

An elevated white blood cell count (Chapter 167) can have many causes, but the immediate priority is to exclude a life-threatening septic process (Chapters 106 and 108). Direct microscopic examination of the peripheral blood smear (Chapter 157) can be helpful to check for toxic granulations (see Fig. 157-18), Döhle's bodies (see Fig. 157-19), and a shift toward primitive band cells. Many cases are due to noninfectious causes, including infarcted tissue (skin, heart, intestinal tract), inflammatory conditions (renal insufficiency, diabetic ketoacidosis, lupus erythematosus), and demargination stress reactions (dehydration, systemic corticosteroids, inotropic medications). In the absence of direct evidence, some clinicians may initiate antibiotics empirically, whereas others may elect waiting. Substantial controversy remains about the proper duration of an empirical trial of antibiotics when no cause is discovered and the patient is otherwise recovering.

Anemia

Anemia (Chapter 158) is common and sometimes underappreciated because coexisting volume depletion causes the blood hemoglobin concentration to underestimate the degree of blood loss. Major perioperative hemorrhage is associated with subsequent stroke and MI in patients undergoing noncardiac, non-neurologic surgery.[7] The initial priority is to differentiate bleeding at the surgical site from other causes. In many cases, the exact cause is unclear, and substantial uncertainty may arise over the need to initiate gastric acid suppression therapy or to interrupt systemic anticoagulation against venous thrombosis. In a large randomized trial, transfusion at a hemoglobin threshold of 10 g/dL was no better than transfusion for symptoms of anemia or at the physician's discretion for a hemoglobin level of below 8 g/dL.■ Guidelines for transfusion therapy depend on the patient's cardiac reserve as well as on the available blood bank supply at the particular medical center. A reasonable goal is to maintain a hemoglobin level of 7 mg/dL or higher, except in patients with cardiovascular disease, in whom a hemoglobin goal of 10 mg/dL is reasonable. The immediate postoperative interval is not usually the appropriate time to initiate erythropoietin, oral iron replacement, or detailed evaluations for other hematologic abnormalities.

Abnormalities in Platelet Count

Patients often have abnormal platelet counts after surgery yet rarely require further evaluation or treatment. In most cases, the thrombocytopenia is mild, does not require transfusion therapy, resolves in a few weeks, and is not a sign of an ominous disorder (e.g., sepsis or heparin-induced thrombocytopenia). Platelet transfusions are indicated if the decrease in platelet count is extreme, accompanied by evidence of major blood loss, or related to recent surgery on the central nervous system (including the eye). Thrombocytosis is also common about a month after surgery and is occasionally extreme. However, even postoperative thrombocytosis exceeding 1,000,000/mL rarely necessitates treatment, does not predispose patients to unwanted clotting disorders, and typically resolves spontaneously after a few weeks.

Abnormal Sodium Concentration

Both hyponatremia and hypernatremia (Chapter 116) are frequent complications in the postoperative setting. The immediate priorities are to assess the patient's intravascular volume status and to correct possible volume depletion. The causes of hyponatremia are multifactorial, including excessive use of diuretics, high levels of intrinsic antidiuretic hormone (as a result of factors such as drugs, pain, and mechanical ventilation), and unmeasured osmoles (e.g., intravenous contrast agents). This risk for postoperative hyponatremia may be reduced by the administration of isotonic saline rather than by water restriction.■ Once hyponatremia develops, vasopressin antagonists (Chapter 116) are effective for both hypervolemic and euvolemic hyponatremia.■ Hypernatremia is always due to free water deficiency, which may indicate severe cognitive impairment or other factors interfering with the ability to express thirst or to ingest water. Correction is similar to that in the nonoperative setting. Abnormalities in sodium concentration require careful follow-up, can recur at any point after an operation, yet are rarely the root cause of a patient's inability to recover from surgery.

Abnormal Serum Potassium Concentration

Hyperkalemia and hypokalemia (Chapter 117) are also frequent postoperative complications. The immediate priority is to assess and to stabilize the patient's electrocardiographic findings. Hyperkalemia is usually due to cellular shifts, renal failure, and tissue destruction (including hemolysis). Hyperkalemia will generally be corrected with treatments that shift potassium into cells (e.g., intravenous glucose with or without insulin) and enhance total excretion (e.g., gastrointestinal binding agents). Hypokalemia is usually due to inadequate intake, excessive loss, or cellular shifts. Hypokalemia will generally be corrected with replacement therapy and rarely requires aldosterone antagonism. Both abnormalities can usually be treated as in the nonoperative setting (Chapter 117). The prognosis is favorable if the patient's electrocardiogram shows no major dysrhythmias and if renal function is preserved.

Alkalosis

Systemic alkalosis (Chapter 118) typically requires volume supplementation because the cause is generally intravascular volume depletion. Blood gas determinations may be necessary in some cases to exclude the possibility of concurrent carbon dioxide retention with compensatory metabolic alkalosis. Untreated, alkalosis can result in altered mentation, cardiac arrhythmias, and delayed mobilization. Most patients with postoperative alkalosis do not require carbonic anhydrase inhibitors or intravenous acid. The prognosis is usually favorable, with gradual correction during a period of several days. Rapid correction of alkalosis, unlike rapid correction of hyponatremia, is not known to cause neurologic injury.

Azotemia

The initial assessment of an elevated serum creatinine concentration (Chapter 114) focuses on reviewing previous values (to distinguish acute from chronic renal insufficiency) and identifying contributing factors (such as prerenal volume depletion, intrarenal nephrotoxins, or postrenal urethral obstruction). A trial of intravenous fluids may be useful on both a diagnostic and therapeutic basis. Treatment is the same as in the nonoperative setting. Subsequent monitoring is always necessary, with serum creatinine measurements obtained on a daily basis. Serial measurements of urinary volume and body weight as well as a urine culture are occasionally helpful in selected cases. The prognosis is dependent on the underlying factors and is less favorable after cardiac surgery.

Hyperbilirubinemia

Elevations in serum bilirubin (Chapter 147) are rare after surgery, even though abnormalities in liver enzyme levels occur frequently with general anesthesia. The most benign explanation is Gilbert's syndrome, but the immediate priority is to assess for possible hepatic failure (especially in patients who have received halothane). As in the nonoperative setting, treatment involves withdrawing potential hepatotoxins, supporting the patient, and allowing time for liver function to recover (Chapter 153). Treatment of hepatic encephalopathy is particularly important because of the constipation and generalized catabolic state that also follow major surgery. Monitoring should include serial measurement of liver function on a daily basis because each component (e.g., bilirubin, albumen, prothrombin time) can be altered by factors unrelated to the liver. The prognosis is unfavorable if the patient's liver function fails to recover quickly.

Hypoalbuminemia

Reductions in serum albumin are common after surgery and are an ominous prognostic finding. The cause is rarely decreased production if the reduction in the albumin level occurs rapidly. Possible explanations include nephrotic syndrome, capillary leak into extravascular spaces, and occult catabolism in unrecognized sites. Treatment with albumin infusions does not usually normalize the biochemical abnormality and does not seem to improve patient mortality rates significantly. The main priorities are to continue nutritional support, to preserve skin integrity, to minimize the use of systemic diuretics, to correct any contributing factors, and to consider correlated serum protein deficiencies (such as reduced levels of immunoglobulins or antithrombin III). Hypoalbuminemia can also cause indirect harm by the loss of carrier proteins, which thereby predisposes patients to potential drug toxicity. The

long-term prognosis is favorable for survivors because the serum albumin level will eventually fully return to normal.

Abnormalities in Blood Glucose Concentration

In patients with diabetes mellitus (Chapter 229), serum blood glucose concentrations often become unstable after surgery because of altered dietary intake, decreased physical activity, and the release of counter-regulatory hormones. The priority is to avoid hypoglycemia, severe hyperglycemia, diabetic ketoacidosis, cerebral damage, and repeated events. Intensive control of the blood glucose level increases the risk for severe hypoglycemia and death, so a target glucose level of about 140 to 200 mg/dL is recommended. Rapid reversal of sepsis or focal infection can lead to a precipitous decrease in insulin requirements; in such cases, vigilance is required because unsuspected hypoglycemia may cause permanent damage or be fatal in a patient who may otherwise seem to be sleeping. Patients need to be forewarned that temporary doses of subcutaneous insulin may be required but do not commit the patient to chronic insulin therapy. Monitoring involves serial measurement of blood glucose concentration until the patient is eating in a reliable manner.

Special Situations

Multiplicity

Some postoperative complications are difficult to classify because no single dominant problem is apparent by symptoms, signs, or laboratory test results. Instead, patients may have multiple problems that need to be addressed simultaneously. The immediate goal is to set priorities and to avoid the temptation to try to eliminate every possibility on the first day. The corollary is to continue to check progress during the subsequent days needed for a complete diagnosis and successful therapy. Because so many concerns require attention in the postoperative interval, the risk is that clinicians will lose track of a secondary issue and make an error that seems obvious in retrospect.

Redundancy

Postoperative complications sometimes generate multiple consultations with physicians who have overlapping abilities. An example might be a patient with a postoperative fever that prompts consultations from pulmonology, nephrology, dermatology, general medicine, and infectious disease specialists. In theory, gathering a critical mass of medical experts together should increase the likelihood of accurate diagnosis, timely treatment, and foolproof follow-up. In reality, however, coordination and communication are never perfect. Personal rivalries, diffusion of responsibility, and many other psychological factors may impede interactions among consultants. Opportunities for miscommunication may be further accentuated if the patient has an exotic diagnosis that is a special draw on the consultant's attention (e.g., pheochromocytoma). Arguing in front of the patient, in view of other professionals, or in the medical record can be demoralizing. The priority is to communicate effectively with the surgical team responsible for the patient and to encourage that team to make final decisions.

Ambiguity

Another vexation occurs when an urgent request is not connected to a clear rationale (Chapter 430). Diplomacy is needed to establish whether the motivation reflects medicolegal concerns rather than biologic changes in the patient. Sometimes the stimulus that drives the consultation can best be addressed by providing reassurance and confirmation. Sometimes the stimulus is an obscure preexisting disorder (e.g., moyamoya disease), and the surgical team has neither the experience nor the time to investigate how this unrelated medical condition can influence recovery from surgery. Sometimes the stimulus is an unspoken political wish to transfer the care of a burdensome patient from one physician to another. A consultant should develop an understanding of how to interact with other clinicians under such ambiguous circumstances.

Setting Priorities

Requests for consultation often arrive outside conventional working hours, are usually tinged with a sense of urgency, and sometimes cluster to encompass more than one patient. Developing an effective method for prioritizing patients is a crucial clinical skill. One communication strategy is to provide an objective estimated time of arrival for the initial request from the surgical team. An often helpful treatment strategy is to make some safe suggestions at the time of the initial request so that clear-cut recommendations can be instituted during the interval before the patient is seen and used later to help evaluate the patient's status and course.

Aftermath

In many postoperative cases, the original reason for consultation may resolve and no major issues remain. The situation now provides an opportunity to review the patient, particularly for the appropriate use of unrelated medications. The consultant may often detect excessive medications that were appropriate early in the hospital course but have ceased to be necessary, thereby justifying discontinuation (e.g., diuretics, antibiotics, bronchodilators). Discontinuing medications that have become superfluous requires initiative and wisdom, and the common mistake is to propagate unnecessary medications in stable patients under the rationale of "don't mess with success." The ability to watch the patient for several hours or for a day or two often presents an ideal opportunity for the safe withdrawal of medications. Ironically, discontinuing a treatment sometimes requires more skill, time, and initiative than starting it.

Grade A References

A1. Gould MK, Garcia DA, Wren SM, et al. Prevention of VTE in nonorthopedic surgical patients. Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e227S-e277S.

A2. Falck-Ytter Y, Francis CW, Johanson NA, et al. Prevention of VTE in orthopedic surgery patients. Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141:e278S-325S.

A3. Squadrone V, Coha M, Cerutti E, et al. Continuous positive airway pressure for treatment of postoperative hypoxemia: a randomized controlled trial. *JAMA*. 2005;293:589-595.

A4. Apfel CC, Korttila K, Abdalla M, et al. A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med*. 2004;350:2441-2451.

A5. Sandham JD, Jull RD, Brant RF, et al. A randomized, controlled trial of the use of pulmonary-artery catheters in high-risk surgical patients. *N Engl J Med*. 2003;348:5-14.

A6. Price R, MacLennan G, Glen J. SuDDICU Collaboration. Selective digestive or oropharyngeal decontamination and topical oropharyngeal chlorhexidine for prevention of death in general intensive care: systematic review and network meta-analysis. *BMJ*. 2014;348:g2197.

A7. Carson JL, Terrin ML, Noveck H, et al. Liberal or restrictive transfusion in high-risk patients after hip surgery. *N Engl J Med*. 2011;365:2453-2462.

A8. Neville KA, Sandeman DJ, Rubinstein A, et al. Prevention of hyponatremia during maintenance intravenous fluid administration: a prospective randomized study of fluid type versus fluid rate. *J Pediatr*. 2010;156:313-319.

A9. Rozen-Svi B, Yahav D, Gheorghiade M, et al. Vasopressin receptor antagonists for the treatment of hyponatremia: systematic review and meta-analysis. *Am J Kidney Dis*. 2010;56:325-337.

A10. Kansagara D, Fu R, Freeman M, et al. Intensive insulin therapy in hospitalized patients: a systematic review. *Ann Intern Med*. 2011;154:268-282.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. de Vries EN, Prins HA, Crolla RM, et al. Effect of a comprehensive surgical safety system on patient outcomes. *N Engl J Med*. 2010;363:1928-1937.
2. Urbach DR, Govindarajan A, Saskin R, et al. Introduction of surgical safety checklists in Ontario, Canada. *N Engl J Med*. 2014;370:1029-1038.
3. Botto F, Alonso-Coello P, Chan MT, et al. Myocardial injury after noncardiac surgery: a large, international, prospective cohort study establishing diagnostic criteria, characteristics, predictors, and 30-day outcomes. *Anesthesiology*. 2014;120:564-578.
4. Saczynski JS, Marcantonio ER, Quach L, et al. Cognitive trajectories after postoperative delirium. *N Engl J Med*. 2012;367:30-39.
5. Pearse RM, Ackland GL. Perioperative fluid therapy. *BMJ*. 2012;344:e2865.
6. Gialdini G, Nearing K, Bhave PD, et al. Perioperative atrial fibrillation and the long-term risk of ischemic stroke. *JAMA*. 2014;312:616-622.
7. Kamel H, Johnston SC, Kirkham JC, et al. Association between major perioperative hemorrhage and stroke or Q-wave myocardial infarction. *Circulation*. 2012;126:207-212.

REVIEW QUESTIONS

1. An 83-year-old woman who fell in a nursing home is admitted because she now is unable to walk. The diagnosis is a hip fracture, and the plan is for surgical repair. She is taking multiple chronic medications that might be sources of drug toxicity if surgery leads to changes in liver function, kidney function, oral intake, or drug-drug interactions. Which of the following medications, however, would not merit careful monitoring and dose adjustment during the perioperative interval?

A. L-Thyroxine
B. Lithium
C. Phenytoin
D. Cyclosporine
E. Metoprolol

Answer: A Changes in drug absorption and elimination lead to large changes in the pharmacodynamics of most medications. As a consequence, routine measurement of lithium levels, phenytoin levels, and cyclosporine levels is indicated in patients receiving such medications. Doses of β-blockers, such a metoprolol, may also need to be adjusted on the basis of the patient's heart rate and blood pressure. One exception is that thyroid supplements, such as L-thyroxine, do not typically require monitoring and adjustment.

2. A 72-year-old man has been diagnosed with ocular cataracts and is awaiting carotid endarterectomy surgery. His list of medications needs to be reviewed to distinguish drugs that are essential in the postoperative setting from other drugs that are counterproductive (or superfluous) in the postoperative setting. Which of the following chronic medications should *not* be routinely discontinued in this patient?

A. Alendronate
B. Calcium carbonate
C. Iron sulfate
D. Ginkgo biloba
E. Aspirin

Answer: E The patient is receiving multiple medications that are often available without a prescription in some countries. Each is relatively safe among otherwise healthy outpatients, can cause problems in patients who are recovering from surgery, and is a long-acting metabolic agent that could be discontinued with no immediate withdrawal syndrome. Calcium carbonate is typically formulated in large pills that might be aspirated in the immediate postoperative setting. Iron sulfate might worsen postoperative gastritis even though it was tolerated on an outpatient basis. Ginkgo biloba can increase the risk of postoperative hepatitis when it is taken in the perioperative setting. Alendronate can lead to hemorrhagic esophagitis arising from impaired swallowing after surgery. Aspirin is the major exception with this particular operation, in which randomized trials suggest a significantly reduced risk of a postoperative stroke after carotid surgery that outweighs the risk of a secondary gastrointestinal tract hemorrhage.

3. A 94-year-old woman presents with postmenopausal painless vaginal bleeding. Her baseline functional status is excellent, and she walks half an hour each day for exercise. The diagnosis is endometrial cancer, and she undergoes a total abdominal hysterectomy. On the first day after surgery, she develops delirium characterized as a hypoactive state with no oral intake, simple one-word verbal responses to questioning, and an inability to follow simple commands. Which of the following would *not* be a contributing cause to her delirium?

A. Carbon dioxide narcosis
B. Adverse drug reaction
C. Hypothyroidism
D. Alcohol withdrawal syndrome
E. Surgical site infection

Answer: C The patient has developed an acute postoperative delirium with no clear localizing findings. An adverse drug reaction is a leading consideration because of drug-drug interactions, altered pharmacokinetics, or other mechanisms. Carbon dioxide narcosis is a possibility, particularly if the patient has undiagnosed sleep apnea, respiratory depression with analgesia, and pulmonary atelectasis after abdominal surgery. Alcohol withdrawal with hypoactive delirium tremens is also a potential, as is an occult surgical site infection. Hypothyroidism, however, rarely is a major contributor to acute delirium. Therefore, hypothyroidism is one factor that is unlikely to explain postoperative delirium.

434

MEDICAL CONSULTATION IN PSYCHIATRY

PETER MANU

HEALTH STATUS IN PSYCHIATRIC PATIENTS

The physical health of psychiatric patients is poor, with a higher risk for death at an early age than in the mentally sane. Poverty, social neglect, substandard medical care, unhealthy life habits, and complications of psychiatric treatments are major contributors to the increased morbidity and mortality of patients with chronic psychiatric disorders, who have a 20-year average decline in life expectancy. Integrated care models that coordinate general medical care with psychiatric care can help meet the needs of patients with chronic psychiatric problems.[1]

MEDICAL EVALUATION IN PSYCHIATRIC SETTINGS

Inpatient psychiatric care is provided in the United States in 444 non-federal, self-standing psychiatric hospitals with a total of 101,500 beds and 1373 general hospitals with inpatient psychiatric units. Requirements for medical training in psychiatry vary. In the United States, psychiatric residents have up

TABLE 434-1 COMMON REASONS FOR TRANSFERRING PSYCHIATRIC INPATIENTS WITH ACUTE MEDICAL DETERIORATION TO A GENERAL HOSPITAL

Fever	17%
Neurologic deficits, seizures, alteration of consciousness	14%
Fall, head trauma	13%
Abdominal pain, gastrointestinal bleeding	10%
Dyspnea, hypoxia	10%
Chest pain	8%
Urinary retention, azotemia, electrolyte imbalance	7%
Arrhythmia, hypotension, syncope	6%
Edema, cellulitis	5%
All others	10%

From Manu P, Asif M, Khan S, et al. Risk factors for medical deterioration of psychiatric inpatients: opportunities for early recognition and prevention. *Compr Psychiatry.* 2012;53:968-974.

to 4 months of inpatient medical training. With such limited medical training, psychiatrists cannot be expected to be experts in assessing nonpsychiatric problems. Free-standing psychiatric facilities employ consultants to evaluate medical issues and to manage outpatient-level conditions, but they are not equipped to provide inpatient-level medical or surgical services. Patients who require a higher level of care are transferred to a medical, surgical, or medical/psychiatric unit of a general hospital.

As many as 15% of patients who are admitted for psychiatric care may be transferred to a general hospital for medical conditions that arise or deteriorate during inpatient psychiatric care.[2] Febrile illnesses, acute neurologic changes, and falls account for nearly half of the transfers (Table 434-1). Nearly half of the patients admitted for dementia with behavioral disturbance will develop a medical complication during an inpatient psychiatric stay, a rate that is two- to three-fold higher than for other psychiatric patients. Renal insufficiency, anemia, poor nutritional status, and older age are also independent predictors of medical deterioration.

Medical deterioration can have major adverse consequences for psychiatric inpatients. First, it may lead to life-threatening complications if the condition is not rapidly diagnosed and treated. Second, it interrupts behavioral interventions and may require the discontinuation of psychotropic drug treatment or electroconvulsive therapy (ECT). Third, it prolongs the length of stay and can add considerable expense to the episode of psychiatric illness, especially when patients are transferred from locked psychiatric units to a general hospital, where they require constant observation by qualified personnel.

EVALUATION OF CHIEF COMPLAINTS

Medical consultation for psychiatric patients creates unique challenges in evaluation of the chief complaint. Many patients with outpatient psychiatric disorders have somatic symptoms, such as fatigue, weakness, dizziness, headache, insomnia, widespread pain, and constipation. In most of these patients, the underlying mental illnesses are mood disorders (unipolar depression and dysthymia), anxiety disorders (panic disorder and generalized anxiety disorder), somatoform disorders, substance use disorders (most often alcohol, opiates, cocaine, and benzodiazepine), and borderline personality disorder. As a group, these patients have many physical complaints and resist a psychological explanation for their symptoms even when the medical evaluation fails to identify objective abnormalities. Standardized screening approaches can help clinicians and patients alike.[3]

In contrast, patients admitted for inpatient psychiatric treatment often have psychotic disorders, developmental abnormalities, or dementia with behavioral disturbance; they are frequently vague or silent about physical suffering and only rarely voice somatic delusions. Such patients may deny pain even after bowel perforation or myocardial infarction.

Medical consultation in psychiatry is not more difficult than in other clinical settings but must be informed by knowledge about the serious complications of psychiatric treatments (Table 434-2). A rigorous diagnostic evaluation is therefore required to avoid the errors of omission created by the weak correlation between complaints and disease. This evaluation should include consideration of physical disorders, drug effects (adverse reaction, toxicity, or withdrawal), cognitive impairment (delirium), sensory impairment (loss of vision, hearing, speech, or postural balance), and situational maladjustment (isolation, overload, or loss of privacy). Attribution of symptoms to the patient's psychiatric condition should remain a diagnosis of exclusion.

TABLE 434-2 MAJOR MEDICAL COMPLICATIONS OF PSYCHOTROPIC DRUGS

CARDIOVASCULAR	
Cardiomyopathy	Clozapine
Hypertension	MAO inhibitors, venlafaxine
Myocarditis	Clozapine
Orthostatic hypotension	Tricyclics, trazodone, antipsychotics
QTc prolongation	Antipsychotics, tricyclics
Venous thromboembolism	Clozapine, risperidone, phenothiazines
RESPIRATORY	
Choking	Antipsychotics, tricyclics
Laryngospasm	Antipsychotics
Respiratory depression	Benzodiazepines, barbiturates, methadone, antidepressants, atypical antipsychotics
GASTROINTESTINAL	
Bowel obstruction	Tricyclics, antipsychotics
Dysphagia	Tricyclics, antipsychotics
Hepatic impairment	Carbamazepine, valproic acid, phenothiazines, mirtazapine, nefazodone, quetiapine, olanzapine, clozapine, MAO inhibitors, naltrexone
Pancreatitis	Carbamazepine, valproic acid, clozapine, olanzapine
KIDNEY AND URINARY TRACT	
Renal insufficiency	Lithium, clozapine
Urinary retention	Antipsychotics, tricyclics
ENDOCRINE	
Hyperprolactinemia	First-generation antipsychotics, risperidone
Hypothyroidism	Lithium, quetiapine
Inappropriate ADH secretion	Serotonin reuptake inhibitors, methadone, tricyclics
Metabolic syndrome	Clozapine, olanzapine, risperidone, quetiapine
HEMATOLOGIC	
Leukocytosis	Lithium
Neutropenia	Clozapine, olanzapine, risperidone, carbamazepine, valproate, mirtazapine
Thrombocytopenia	Carbamazepine, valproate
MUSCULOSKELETAL	
Rhabdomyolysis*	Antipsychotics, serotonin reuptake inhibitors, MAO inhibitors
SKIN	
Stevens-Johnson syndrome	Lamotrigine, carbamazepine, barbiturates
OTHER	
Fever*	Antipsychotics, serotonin reuptake inhibitors, MAO inhibitors
Seizure	Bupropion, MAO inhibitors, tricyclics, phenothiazines

*Includes neuroleptic malignant syndrome and serotonin syndrome.
ADH = antidiuretic hormone; MAO = monoamine oxidase.

MEDICAL COMPLICATIONS OF PSYCHIATRIC TREATMENTS

Antipsychotic-Induced Metabolic Syndrome

Metabolic syndrome is more prevalent (range, 29 to 63%) in schizophrenic and other psychiatric patients treated with second-generation antipsychotics, especially clozapine but also olanzapine, risperidone, and quetiapine because they induce substantial weight gain. The mechanism of weight gain centers on the drug's affinity for the histamine$_1$ (H_1)–receptor and the neurobiologic mechanisms that regulate appetite and metabolism through the production and activity of serotonin, leptin, and tumor necrosis factor-α. As a result, up to 40% of patients receiving long-term treatment with antipsychotics have prediabetes, and about 10% have diabetes.[4]

Because the glucose intolerance often seen in patients treated with antipsychotic agents is due to insulin resistance (Chapter 229), it is best treated by aggressive weight reduction (Chapter 213), increasing physical activity (Chapter 16), and metformin (Chapter 229) or a combination of these. For example, in randomized trials, metformin, 850 mg twice daily, significantly reduced the weight gain and reversed the metabolic abnormalities associated with the initiation of second-generation antipsychotic drugs,[■] and a single daily dose of 750 mg may prevent weight gain if it is initiated at the onset of

olanzapine treatment. Another option is to switch from olanzapine, quetiapine, or risperidone to aripiprazole. In treating hypertension, β-adrenergic blockers should be used cautiously in patients receiving neuroleptics because they increase the potential for orthostatic hypotension, syncope, and falls. Smoking cessation programs combining transdermal or transmucosal nicotine replacement, bupropion, and cognitive-behavioral therapy should be strongly recommended despite the dismally high failure rates that have been reported in patients with chronic schizophrenia (Chapter 32).

Antipsychotic-Induced Myocarditis and Cardiomyopathy

Antipsychotic-induced myocarditis and cardiomyopathy (Chapter 60) are most common in patients treated with clozapine (0.9%) and fluphenazine (0.4%). In contrast, the risk for these complications is only 0.1% in patients receiving haloperidol, thioridazine, and risperidone.

The accepted pathophysiologic explanation for myocarditis is an immunoglobulin E–mediated acute hypersensitivity reaction, similar to the allergic myocarditis produced by penicillins, sulfonamides, and methyldopa. In a small number of patients, a competing hypothesis proposes that clozapine induces hypereosinophilic myocarditis, colitis, hepatitis, pancreatitis, alveolitis, and interstitial nephritis. A direct cardiotoxic effect of drug metabolites cannot be excluded.

In patients in whom myocarditis develops, the mortality rate is as high as 50%, with almost half the deaths occurring suddenly and unexpectedly. The average duration of exposure to clozapine before diagnosis or death is 21 days, and the dosage range is 50 to 725 mg/day. Common symptoms are fever (48%), dyspnea (35%), influenza-like illness (30%), chest pain (22%), and fatigue (17%). Laboratory features include left ventricular hypokinesia or reduced ejection fraction (48%) or pericardial effusion (17%) on echocardiography, nonspecific repolarization abnormalities on electrocardiography (35%), peripheral eosinophilia (35%), elevated creatine kinase and troponin levels (22%), and radiographic evidence of heart failure (13%). The diagnosis can be confirmed by endomyocardial biopsy showing fraying of myocytes and perivascular infiltrates with degranulated eosinophils. Among survivors, symptoms resolve or substantially improve after discontinuation of clozapine and treatment with high-dose corticosteroids (e.g., prednisone, 1 mg/kg/day for 4 days, tapered to 0.33 mg/kg/day for the following 4 days).

Clozapine-induced dilated cardiomyopathy may be caused by an evolving myocarditis or by chronic injury mediated by free radicals, similar to the myocarditis produced by doxorubicin (Chapter 60). The demographic features are similar to those of myocarditis, but the mean duration of treatment before diagnosis is much longer (9 months vs. 3 weeks), and the mortality rate is lower (22% vs. 51%). Patients have clinical or echocardiographic evidence of left ventricular dysfunction without eosinophilia or enzymatic evidence of myocardial necrosis.

Prolonged QTc Interval and Sudden Death

Significant prolongation of the QTc interval (Chapter 65) leading to ventricular tachyarrhythmias and sudden cardiac death (Chapter 63) can occur after antipsychotic treatment with the usual doses of thioridazine, haloperidol, and sertindole. Abnormal myocardial repolarization has been observed during treatment with most antipsychotic medications and after intentional or accidental overdoses of tricyclic antidepressants, lithium, and methadone (Chapter 110). All antipsychotics affect the cardiac potassium channel by blocking the rapidly activating component of the rectifier potassium current (Chapter 61). This effect translates into a dose-dependent increase in the duration of phase 3 of the action potential. The drug concentration that produces 50% inhibition of rapid potassium outflow varies for each drug (e.g., 1 nmol/L for haloperidol and 6 nmol/L for olanzapine). Compared with nonusers of antipsychotic drugs, the risk for sudden death is twice as high for current users of conventional (first-generation) antipsychotics and is 2.25 times higher for current users of atypical (second-generation) antipsychotics. All patients about to start antipsychotic drugs should be asked about a personal history of syncope and a family history of long QT syndrome or sudden death at a young age. A baseline electrocardiogram and serum electrolyte values should be obtained before starting of antipsychotic drug therapy, tricyclic antidepressants, and methadone. Interval electrocardiograms should be obtained after each increase in medication in older patients, patients with known heart disease, and those starting other drugs known to produce QTc prolongation or hypokalemia (Chapter 65). A QTc interval of 500 milliseconds or longer requires the discontinuation of all drugs that affect membrane repolarization. QTc intervals longer than 450 milliseconds in men and 470 milliseconds in women, QTc dispersion (difference between the longest and shortest QTc on a 12-lead electrocardiogram) longer than 100 milliseconds, and increase in QTc duration of more than 60 milliseconds in comparison to the baseline measurement should prompt re-evaluation of the risks and benefits associated with the drugs in question.

Choking and Laryngeal Dystonia

Asphyxia deaths from choking occur at a rate of 0.8% per 1000 psychiatric patients each year, a frequency that is more than 100 times greater than in the general population. In addition, videofluoroscopy demonstrates silent aspiration in 38% of psychiatric patients who survive a choking incident. Half the psychiatric patients with dysphagia have a fast-eating syndrome seen in association with restlessness, poor chewing skills, food pocketing in the cheeks, and attention deficits that characterize psychotic disorders and mental retardation. Bradykinetic dysphagia, which is seen in 25% of psychiatric patients with choking episodes, is due to the antidopaminergic and anticholinergic effects of psychotropic medications. This condition, which features reduced lingual range of motion, increased oral transit time, decreased pharyngeal peristalsis, and delayed initiation of the swallowing reflex, is seen in patients with neurologic features of drug-induced Parkinsonism (Chapter 409). Dyskinetic dysphagia (7% of choking cases), which generally occurs in patients maintained with long-term antipsychotic medication, is part of the clinical spectrum of tardive dyskinesia (Chapter 410). The examination reveals involuntary contractions of the tongue and perioral musculature, clumsiness of voluntary movements of the tongue, and discontinuous bolus propulsion in the oral stage. In the remaining patients, the dysphagia is due to cerebrovascular disease (11%) or to pharyngeal or esophageal disease (7%).

Laryngeal dystonia, which is a life-threatening complication of antipsychotic drug therapy, primarily with haloperidol and phenothiazines, is produced by acute spasmodic contraction of the adductor laryngeal muscles. Symptoms include respiratory distress, dysphonia, and stridor. Neuroleptic-induced bronchospasm may precede the onset of stridor. Patients typically indicate extreme subjective distress by clutching their anterior cervical area. Most patients also have other dystonias involving the head and neck, including torticollis, retrocollis, trismus, tongue protrusion, and deviation of the eyes (up, down, or sideward). In general, the symptoms and signs develop in the first week after starting or rapidly increasing the dose of neuroleptic medications. A reduction in the dose of anticholinergic or antiparkinsonian medication used to prevent or to treat extrapyramidal symptoms can also precipitate laryngeal dystonia. The condition is more common in young men and must be distinguished from epiglottitis, allergic or anaphylactic laryngeal edema or laryngospasm, mechanical obstruction, and psychogenic stridor. Intravenous administration of diphenhydramine (initial dose, 25 mg; may repeat after 5 minutes if symptoms persist) is the treatment of choice, and endotracheal intubation is seldom required.

Drug-Induced Neutropenia and Agranulocytosis

Drug-induced neutropenia with absolute neutrophil counts of less than 1500/μL has been observed during treatment with most second-generation antipsychotics (clozapine, olanzapine, risperidone, and quetiapine) and mood stabilizers (carbamazepine, valproic acid, and lamotrigine) as well as with some antidepressant drugs (tricyclic antidepressants and mirtazapine). Clozapine-induced neutropenia occurs in 4 to 5% of patients within 6 months after treatment is started and progresses to agranulocytosis in 10% or more of neutropenic patients if the drug is continued. In vitro, clozapine toxicity requires peroxide and peroxidase, and the defect in oxidation is related to abnormalities in the *NQO2* (quinone oxidoreductase) gene involved in drug detoxification. Treatment with clozapine should be started only if the baseline absolute neutrophil count is higher than 1500/μL. The concomitant use of carbamazepine, angiotensin-converting enzyme inhibitors, sulfonamides, propylthiouracil, and mirtazapine should be avoided because they can produce neutropenia and increase the risk for agranulocytosis. Clozapine should be stopped and the patient evaluated immediately for fever, oral ulcerations, and symptoms or signs of infection. Complete blood counts should be obtained once a week for the first 26 weeks and every other week thereafter, and clozapine should be stopped and all medications reassessed if the absolute neutrophil count drops below 1500/μL. Clozapine-related agranulocytosis has been treated successfully with colony-stimulating factors (either granulocyte or granulocyte-macrophage colony-stimulating factor).

Neutropenia has also been associated with olanzapine, risperidone, and quetiapine in patients who have never received clozapine. Treatment with anticonvulsant mood stabilizers, particularly carbamazepine, is associated with a dose-dependent neutropenia and thrombocytopenia in approximately

TABLE 434-3 DIFFERENTIAL DIAGNOSIS OF NEUROLEPTIC MALIGNANT SYNDROME

Infection of the central nervous system
Infection in patients with drug-induced Parkinsonism
Drug overdose (psychostimulants, antidepressants, lithium, anticholinergics)
Alcohol or drug withdrawal (benzodiazepines, barbiturates, antiparkinsonian drugs)
Side effects of nonpsychotropic dopamine-depleting drugs (reserpine, metoclopramide, prochlorperazine, promethazine)
Cholinergic rebound
Serotonin syndrome
Thyrotoxicosis
Malignant hyperthermia

10% of patients in the first 6 months of treatment and should be monitored with complete blood counts twice each month during this period.

Neuroleptic Malignant Syndrome

Neuroleptic malignant syndrome, which occurs in approximately 0.2% of patients receiving neuroleptics, must be part of the differential diagnosis of fever and rhabdomyolysis (Chapter 113) in a psychiatric patient (Table 434-3). The frequency is greater in young men and patients who are malnourished or dehydrated, have Parkinson's disease, or are treated parenterally with large doses of neuroleptics during short periods. The main diagnostic criteria are elevated temperature (higher than 104°F [40°C] in 40% of patients) and diffuse muscle rigidity (ranging from mild hypertonicity to severe "lead pipe" stiffness). In addition, two or more of the following are required for a definitive diagnosis: (1) autonomic instability (tachycardia, elevated or labile blood pressure, postural hypotension, diaphoresis, sialorrhea, and urinary incontinence), (2) changes in mental status (ranging from confusion to mutism or coma), (3) leukocytosis (up to 20,000/mL), and (4) elevated creatine kinase (up to 100,000 IU/L). Other clinical manifestations include bradykinesia, chorea, dystonias, dysphagia, dysarthria or aphonia, seizures, and tremor. The severity of rhabdomyolysis correlates with the creatine kinase level and with the presence of myoglobinemia, myoglobinuria, metabolic acidosis, and azotemia. The electroencephalogram shows nonspecific slowing in slightly more than half of patients.

The time lag from starting of the drug to the onset of neuroleptic malignant syndrome is generally short, with 30% of cases developing within 48 hours and 96% within the first month of treatment. The exception appears to be clozapine-associated neuroleptic malignant syndrome, which has an average time lag of 50 days. Neuroleptic syndrome is sometimes confused with severe catatonia (Chapter 397), but the catatonic signs in neuroleptic malignant syndrome are usually restricted to mutism and akinesia. Furthermore, hyperthermia, rigidity, tremor, and rhabdomyolysis are not present in patients with catatonia. Nonetheless, close medical follow-up of severely catatonic patients is warranted because they are at very high risk (22%) for neuroleptic malignant syndrome.

Untreated, neuroleptic malignant syndrome has a mortality rate of 10% as a result of acute renal failure, aspiration pneumonia, adult respiratory distress syndrome, disseminated intravascular coagulation, and cerebellar neuronal degeneration. Most fatalities are avoidable if the diagnosis is made early, the neuroleptic agent is discontinued rapidly, and the patient is immediately transferred to an intensive care setting for supportive and specific therapy. Bromocriptine (2.5 mg three times daily orally or through a nasogastric tube; may increase by 2.5 mg three times daily to a maximal daily dose of 40 mg) or amantadine (100 mg orally or through a nasogastric tube twice daily; may increase to 300 mg/day in divided doses) should be used in moderately severe cases and continued until the muscle rigidity and metabolic abnormalities have significantly improved. The skeletal muscle relaxant dantrolene (starting with a dose of 1 mg/kg intravenously and titrated to a maximal dose of 10 mg/kg/day divided into three intravenous or oral doses) should be added to bromocriptine or amantadine in patients with fulminant hypermetabolic features and those with persistent muscle rigidity despite treatment with dopamine agonists. Refractory neuroleptic malignant syndrome improves after ECT.

Serotonin Syndrome

Serotonin syndrome (Chapter 432) is an adverse drug reaction primarily produced by excess serotonergic agonism of central nervous system and peripheral serotonin receptors by selected drugs (Table 434-4).[5] In postmarketing surveillance studies of the newer antidepressants, the syndrome has an incidence of 4 cases per 10,000 patient-months in patients who start taking nefazodone, a drug that inhibits neuronal uptake of serotonin and norepinephrine and also acts as a 5-hydroxytryptamine type 2 (5-HT_2) receptor antagonist. The syndrome also occurs in 15% of patients with intentional overdose of selective serotonin reuptake inhibitors (SSRIs). The serotonin syndrome is caused by overstimulation of 5-HT_{1A} and possibly also 5-HT_2 receptors through excess of serotonin precursors or agonists, increased serotonin release, reduced serotonin uptake, and decreased serotonin metabolism. Severe cases of the syndrome have been more frequently reported in patients treated with monoamine oxidase inhibitors who took over-the-counter dextromethorphan or the illegal methylenedioxymethamphetamine (Ecstasy) or who started treatment with serotonin reuptake inhibitors, meperidine, or atypical antipsychotics such as aripiprazole.

TABLE 434-4 CLASSES OF MEDICATIONS THAT PRODUCE SEROTONIN SYNDROME IN PSYCHIATRIC PATIENTS

Selective serotonin reuptake inhibitors
Monoamine oxidase inhibitors
Atypical antipsychotics
Heterocyclic antidepressants
Trazodone
Dual-uptake inhibitors
Psychostimulants
Buspirone
Mood stabilizers
Analgesics
Antiemetics
Cough suppressants
Dietary supplements

Potentially life-threatening, the syndrome is characterized by changes in mental status (ranging from agitation to confusion and coma), autonomic instability (tachycardia, labile or high blood pressure, diaphoresis, and diarrhea), neuromuscular abnormalities (myoclonus, mydriasis, ocular clonus, rigidity, hyperreflexia, tremors, and shivering), and hyperthermia. The symptoms occur within the first 24 hours and sometimes within minutes after the initial use of medication, a change in dose, addition of a new drug, or overdose attempt. Death may occur as a consequence of rhabdomyolysis with renal failure, hyperkalemia, disseminated intravascular coagulation, and acute respiratory distress syndrome. The differential diagnosis includes neuroleptic malignant syndrome, viral or bacterial meningitis or encephalitis, heat stroke (Chapter 109), anticholinergic toxidrome (Chapter 110), and drug (Chapter 34) or alcohol (Chapter 33) withdrawal.

General management includes immediate discontinuation of serotonergic drugs, comprehensive supportive therapy, and benzodiazepines for control of agitation and myoclonus. Specific therapy relies on the use of cyproheptadine (an H_1-receptor antagonist with antiserotonergic and anticholinergic properties) and chlorpromazine (a 5-HT_{1A} and 5-HT_2 receptor antagonist). Cyproheptadine should be started at a dose of 12 mg administered orally or through a nasogastric tube, with additional 2-mg doses given every 2 hours until symptoms improve or the maximal dose of 32 mg has been reached. The usual maintenance dose of cyproheptadine is 8 mg three times daily. Chlorpromazine (50 mg intramuscularly; may repeat three or four times daily and increase gradually to 400 mg/day in divided doses) is indicated in patients with severe symptoms who must be treated parenterally. Rapid improvement has also been observed after single doses of olanzapine (10 mg administered sublingually). Chlorpromazine and olanzapine should be used only after the possibility of neuroleptic malignant syndrome has been excluded.

Antipsychotic-Induced Hyperprolactinemia

Drug-induced hyperprolactinemia is produced by first-generation antipsychotic medications and by risperidone, but it is rare with other atypical antipsychotics such as aripiprazole, olanzapine, and ziprasidone. In patients treated with prolactin-raising antipsychotic medications, hormone levels are above the normal limit in 60% of women and 40% of men. Symptomatic

hyperprolactinemia (Chapter 224) occurs in about one third of these patients and is generally associated with a 10-fold increase above baseline levels. Excess prolactin leads to dysfunction of target tissues (galactorrhea, oligomenorrhea and amenorrhea, infertility, sexual impairment, and gynecomastia) as well as an increased risk for breast cancer, osteoporosis, and cardiovascular disease. The mechanism of antipsychotic-related hyperprolactinemia is suppression of dopamine inhibition of lactotroph cells in the hypothalamus. Brain imaging is required in symptomatic patients and those with significant elevated prolactin levels to exclude tumors of the pituitary and hypothalamus.

Psychogenic Polydipsia and Drug-Induced Hyponatremia

Hyponatremia, which is present in approximately 6% of patients at the time of admission to inpatient psychiatric care and is more common in older patients and those with hypertension, approximately doubles the risk of medical deterioration.[6] The diagnosis is often delayed because traditional manifestations of hyponatremia, such as lethargy, restlessness, weakness, and disorientation, overlap with features of psychiatric disorders.

Patients with psychogenic polydipsia typically have serum hypo-osmolality and a maximally dilute urine (urine osmolality <100 mOsm/L).[7] The incidence of polydipsia is 20% and the incidence of water intoxication is 5% in inpatient psychiatric facilities. Urinary incontinence and nocturnal enuresis may be part of the clinical manifestation. The mechanism of increased thirst is poorly understood but may involve incomplete suppression of antidiuretic hormone (ADH) by the hypothalamus as well as response to the mouth dryness produced by the anticholinergic effect of many psychotropic drugs. The differential diagnosis includes diuretic effect, renal insufficiency, glucocorticoid deficiency, and hypothyroidism. Stringent measures to restrict fluid intake are generally effective in patients with moderate hyponatremia, and clozapine limits polydipsia and improves water intoxication in refractory cases.

Drug-induced syndrome of inappropriate antidiuretic hormone (SIADH) is predominantly related to SSRIs, and animal experiments have suggested that the excess serotonin stimulates release of ADH and will lead to hyponatremia, provided water intake is sufficient. Other drugs that are commonly used by psychiatric patients and that may produce SIADH include tricyclic antidepressants, monoamine oxidase inhibitors, carbamazepine, conventional and second-generation antipsychotics, benzodiazepines, methadone, and nicotine. Elderly patients, patients with a lower body mass index, and those with a baseline plasma sodium level of less than 138 mEq/L are at higher risk. The median time to diagnosis after SSRIs are started is 9 days. Urinary excretion of sodium usually is more than 20 mEq/L, and urine osmolality is higher than 300 mOsm/L. In patients with mild asymptomatic hyponatremia, SSRIs can be continued with careful monitoring while the patient is placed on supervised fluid restriction. For patients who cannot tolerate fluid restriction or have symptomatic hyponatremia with serum sodium levels of less than 125 mEq/L, the use of tolvaptan (initially 15 mg, titratable to 30 or 60 mg), a vasopressin V_2-receptor antagonist, is effective.

Risk Assessment before Electroconvulsive Therapy

ECT is highly effective for the treatment of drug-refractory major depression and other psychiatric disorders. The procedure requires a brief period of general anesthesia with sodium pentothal, etomidate, or propofol as well as muscle paralysis with succinylcholine, during which the patient receives bag-valve-mask ventilation with supplemental oxygen and is monitored with continuous electrocardiography and pulse oximetry. Bronchospasm may follow induction of anesthesia, particularly in patients already at risk for respiratory compromise. During electrical stimulation of the brain, parasympathetic activation can lead to bradycardia or several seconds of asystole, which can be avoided by premedicating the patient with intravenous glycopyrrolate. The parasympathetic effect lasts until onset of the akinetic seizure, when sympathetic tone increases and produces tachycardia, an elevation in blood pressure, and increased myocardial demand for oxygen. These changes can be corrected with intravenous esmolol. At the same time, because of the enhanced neuronal metabolic rate, augmented blood flow to the brain increases intracranial pressure. This elevated sympathetic tone causes ECT-related myocardial ischemia, tachyarrhythmias, and potential rupture of aortic or intracranial aneurysms. The elevated intracranial pressure may lead to brain herniation in patients with a space-occupying lesion. Older patients have a high rate of prolonged confusion, arrhythmias, and falls after ECT. A structured medical evaluation before ECT can address risks and the potential for complications (Fig. 434-1). Every effort must be made to optimize the patient's active medical conditions before ECT. For high-risk patients, one must require that ECT be performed in a setting that allows immediate access to an intensive care unit rather than in the ECT suite of a self-standing psychiatric hospital. Essential medications should be administered with a small amount of fluid 6 hours before ECT. Drugs that can increase or decrease the seizure threshold, such as lidocaine, theophylline, phenothiazine, tricyclic antidepressants, and benzodiazepines, must be discontinued before ECT. Optimal pre-ECT risk assessment, careful anesthesia, and post-ECT monitoring result in a serious complication rate of only 0.9% and essentially no fatalities.

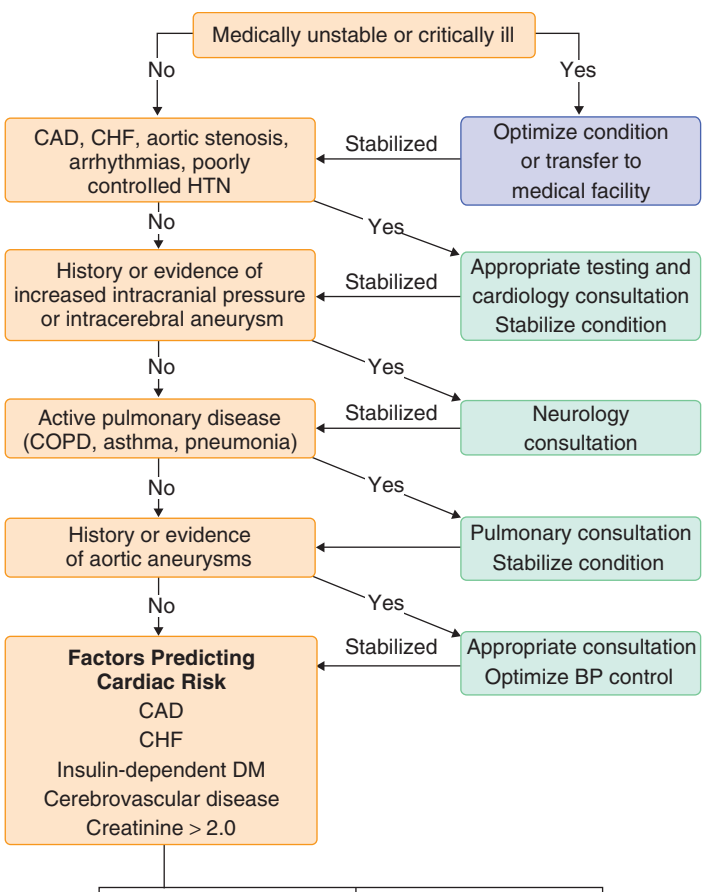

FIGURE 434-1. Risk assessment before electroconvulsive therapy. BP = blood pressure; CAD = coronary artery disease; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; DM = diabetes mellitus; HTN = hypertension. (Modified from Frederickson A, Manu P. Risk assessment prior to electroconvulsive therapy. In Manu P, Suarez RE, Barnett BJ, eds. Handbook of Medicine in Psychiatry. Washington, DC: American Psychiatric Publishing; 2006:687-700.)

Grade A References

A1. Chen CH, Huang MC, Kao CF, et al. Effects of adjunctive metformin on metabolic traits in nondiabetic clozapine-treated patients with schizophrenia and the effect of metformin discontinuation on body weight: a 24-week, randomized, double-blind, placebo-controlled study. *J Clin Psychiatry*. 2013;74:e424-e430.

A2. Wu RR, Zhao JP, Guo XF, et al. Metformin addition attenuates olanzapine-induced weight gain in drug-naïve first-episode schizophrenia patients: a double-blind, placebo-controlled study. *Am J Psychiatry*. 2008;165:352-358.

A3. Stroup TS, Byerly MJ, Nasrallah HA, et al. Effects of switching from olanzapine, quetiapine, and risperidone to aripiprazole on 10-year coronary heart disease risk and metabolic syndrome status: results from a randomized controlled trial. *Schizophr Res*. 2013;146:190-195.

A4. Josiassen RC, Goldman M, Jessani M, et al. Double-blind, placebo-controlled, multicenter trial of a vasopressin V_2-antagonist in patients with schizophrenia and hyponatremia. *Biol Psychiatry*. 2008;54:1097-1100.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Bradford DW, Cunningham NT, Slubicki MN, et al. An evidence synthesis of care models to improve general medical outcomes for individuals with serious mental illness: a systematic review. *J Clin Psychiatry*. 2013;74:e754-e764.
2. Manu P, Asif M, Khan S, et al. Risk factors for medical deterioration of psychiatric inpatients: opportunities for early recognition and prevention. *Compr Psychiatry*. 2012;53:968-974.
3. Rayner L, Matcham F, Hutton J, et al. Embedding integrated mental health assessment and management in general hospital settings: feasibility, acceptability and the prevalence of common mental disorder. *Gen Hosp Psychiatry*. 2014;36:318-324.
4. Manu P, Correll CU, van Winkel R, et al. Prediabetes in patients treated with antipsychotic drugs. *J Clin Psychiatry*. 2012;73:460-466.
5. Alusik S, Kalatova D, Paluch Z. Serotonin syndrome. *Neuro Endocrinol Lett*. 2014;35:265-273.
6. Manu P, Ray K, Rein JL, et al. Medical outcome of psychiatric inpatients with admission hyponatremia. *Psychiatry Res*. 2012;198:24-27.
7. Atsariyasing W, Goldman MB. A systematic review of the ability of urine concentration to distinguish antipsychotic- from psychosis-induced hyponatremia. *Psychiatry Res*. 2014;217:129-133.

REVIEW QUESTIONS

1. Medical deterioration during psychiatric admissions may lead to life-threatening complications, interrupt biobehavioral interventions, and require transfer to medical facilities. Older age and a diagnosis of dementia with behavioral disturbance are nonmodifiable correlates of an increased rate of acute changes in medical conditions in psychiatric patients. Which of the following is an independent risk factor for medical deterioration in psychiatric settings?

A. Blood urea nitrogen concentration (higher)
B. Congestive heart failure
C. Serum potassium concentration (lower)
D. Serum sodium concentration (lower)
E. Substance use disorder

Answer: A Among 1000 adults consecutively admitted to a free-standing psychiatric hospital, 14% were transferred to a general hospital for acute medical deterioration. The independent predictors of medical deteriorations were higher blood urea nitrogen concentration (odds ratio, 63.2), lower hemoglobin level (odds ratio, 35.3), and lower albumin level (odds ratio, 7.3). (Manu P, Asif M, Khan S, et al. Risk factors for medical deterioration of psychiatric inpatients: opportunities for early recognition and prevention. *Compr Psychiatry.* 2012;53:968-974.)

2. The 27-year-old with clozapine-treated schizophrenia admitted for suicidal ideation is prescribed venlafaxine 37.5 mg/day, which is rapidly increased to 150 mg/day without adverse drug reactions. An episode of severe agitation is treated with aripiprazole 9.75 mg administered intramuscularly. The following morning, the patient has good behavioral control and is allowed to play basketball on the outdoor court for 45 minutes. He then becomes febrile (39.1° C) and has two episodes of nonbloody diarrhea. The physical examination is significant for mild and diffuse muscle rigidity, diaphoresis, and ankle clonus. The white blood cell count is 10,200 with an absolute neutrophil count of 4800; the creatine kinase level is 380. The electrolyte values and urinalysis are normal. What is the most likely cause of the febrile syndrome?

A. Neuroleptic malignant syndrome
B. Eosinophilic colitis
C. Viral gastroenteritis and neuroleptic-induced Parkinsonism
D. Serotonin syndrome
E. Heat stroke

Answer: D Serotonin syndrome may develop rapidly after the addition of an antipsychotic (aripiprazole) with serotonergic enhancing properties in patients treated with a dual (serotonin and norepinephrine) reuptake antidepressant drug (venlafaxine). Some of its clinical features, (e.g., diarrhea and clonus) are not seen in neuroleptic malignant syndrome. (Perry PJ, Wilborn CA. Serotonin syndrome vs neuroleptic malignant syndrome: a contrast of causes, diagnoses, and management. *Ann Clin Psychiatry.* 2012;24:155-162.)

435

STRUCTURE AND FUNCTION OF THE SKIN

DAVID H. CHU

The skin, as the largest organ of the body, is a complex multifunctional entity with many regional specializations that provide its host both protection from, and interaction with, its environment. The skin is not just an impenetrable shield against external injury but rather a dynamic, intricate, integrated arrangement of cells, tissues, and matrix elements that together perform a variety of functions, as follows:

- *Physical/mechanical protection:* The epidermis, which is the selectively permeable layer outermost covering of the skin, helps regulate water loss and provides a physical barrier against external insults. The underlying dermis provides mechanical strength to the skin.
- *Thermoregulation:* Eccrine glands mediate the excretion of water, evaporation of which is critical for thermoregulation. The cutaneous vasculature also helps regulate heat exchange via vasoconstriction and vasodilatation.
- *Immunologic surveillance:* Resident and transient cells of the immune system provide some of the earliest defenses against infectious diseases, toxins, and malignantly transformed cells.
- *Sensation:* Nerve endings and specialized sensory apparatus embedded within the skin provide sensory inputs such as pain and itch.
- *External appearance:* The contours, texture, pigment, and regional variations of the skin are keys to an individual's external appearance and perception by others.

These various functions of the skin are mediated by one or more of its major regions—the epidermis, dermis, and subcutaneous fat (Fig. 435-1). These anatomic subdivisions are interdependent, functional units, each relying on and connected with its surrounding tissue for regulation and modulation of normal structure and function at molecular, cellular, and tissue levels of organization.

Whereas the epidermis and its outer stratum corneum constitute a large part of the physical barrier provided by the skin, the structural integrity of skin as a whole is primarily attributable to the dermis and subcutaneous fat. Antimicrobial activities are provided by the innate immune system and antigen-presenting Langerhans cells of the epidermis, by circulating immune cells that migrate from the blood vessels in the dermis, and by antigen-presenting dendritic cells of the dermis. The most superficial cells of the epidermis provide most of the protection from ultraviolet (UV) irradiation (Chapter 20). Sensation emanates from nerves that initially traverse the subcutaneous fat to the dermis and epidermis before ending in specialized receptive organs or free nerve endings. The largest blood vessels of the skin are found in the subcutaneous fat, where they transport nutrients and circulating immune cells. The cutaneous lymphatics, which also course through the dermis and subcutaneous fat, filter debris and regulate tissue hydration. The skin's pigmentation is regulated by melanocytes, whereas actinic damage and aging are influenced by the epidermis, the dermis, and the subcutaneous fat.

EPIDERMIS

Epidermal Differentiation

The epidermis, which is the outermost layer of the skin, is composed primarily of keratinocytes that differentiate to form a stratified squamous epithelium.[1] The epidermis is a continually renewing structure that gives rise to specialized derivatives called *appendages*—pilosebaceous units, sweat glands, and nails. The keratinocytes within the epidermis are organized into four layers, named for either their position or structural property (Fig. 435-2). Cells develop from the basal layer and then undergo programmed biochemical and morphologic changes that ultimately result in formation of the outermost stratum corneum, which serves as a hardened shield against the host environment. Melanocytes, Langerhans cells, and Merkel cells form the bulk of cells that reside within different layers of the epidermis. Some regional differences in the epidermis and its appendages are readily apparent (e.g., the thickness of palmoplantar and truncal skin compared with eyelid skin), whereas other differences are evident only at the microscopic or biochemical level (e.g., differential expression of keratins in dorsal compared with volar skin).

The Basal Layer

The basal layer, which is the germinative layer of the epidermis, contains proliferative cells that give rise to the more differentiated levels of the epidermis. Keratinocyte differentiation (keratinization) is a genetically programmed, regulated, complex series of morphologic and metabolic changes whose end point is a terminally differentiated, cornified envelope that consists of dead keratinocytes (corneocytes) and contains a protein-reinforced plasma membrane with surface-associated lipids.

All epithelial cells have intermediate filaments known as keratins. Keratins primarily serve a structural role, and they are expressed in region-specific and function-specific obligate heteropolymer pairs in a pattern that is determined by cell type, tissue type, developmental stage, differentiation stage, and disease condition (E-Table 435-1). The critical role of these molecules is underscored by the numerous manifestations of disease that arise because of mutations in their genes.

Basal cells attach to their underlying basal lamina via keratin filaments at hemidesmosomes (E-Fig. 435-1) in the basement membrane zone. They also attach to each other and to other keratinocytes via desmosomes, which are specialized junctions required for epidermal integrity. Keratinocytes also contain melanosomes—vacuoles that contain pigment that is synthesized by melanocytes and transferred to keratinocytes via phagocytosis. This pigment helps protect against UV damage and contributes to macroscopic skin coloring.

In the epidermis, the greatest amount of mitotic activity occurs in the basal layer. Cell kinetic studies suggest that different basal keratinocytes have different proliferative potentials, and in vivo and in vitro studies suggest the existence of long-lived epidermal stem cells. Because basal cells can be expanded in tissue culture and used to reconstitute enough epidermis to cover the entire skin surface of burn patients, these long-lived stem cells appear to have extensive proliferative potential.

The Spinous Layer

The spinous layer is characterized by the presence of abundant desmosomes, which are critical for tight epidermal adhesion and resemble intercellular "spines" under light microscopy. The classic vesiculobullous disease pemphigus vulgaris (Chapter 439) results from desmosomal disruption.

The Granular Layer

The granular layer, so named because of the abundant keratohyalin granules contained in the cells that constitute it, represents an intermediate phase of keratinocyte differentiation. This layer is where a number of the structural components will form the cornified cell envelope of the epidermal barrier, as well as a number of proteins that process these components. Keratohyalin granules are composed primarily of profilaggrin, keratin filaments, and loricrin. Profilaggrin is cleaved into filaggrin monomers, which aggregate with keratin to form macrofilaments. Eventually, filaggrin is degraded into molecules that contribute to hydration of the stratum corneum and help filter UV radiation. Loricrin, which is a cysteine-rich protein, is released from keratohyalin granules and then binds to desmosomal structures. Loricrin is subsequently cross-linked to the plasma membrane by tissue transglutaminases to help form the cornified cell envelope.

The Stratum Corneum

The final stage of granular cell differentiation into a corneocyte involves the cell's own programmed destruction, a process during which almost all cellular contents are destroyed, with the exception of the keratin filaments and filaggrin matrix. The cells eventually extrude their nuclei and become a flattened stratum corneum that provides mechanical protection and a barrier to the loss of water or the entry of soluble substances from the environment.

The stratum corneum barrier is formed by a two-compartment system of lipid-depleted, protein-enriched corneocytes surrounded by a continuous extracellular lipid matrix. The extracellular lipid matrix is primarily responsible for the regulation of permeability, desquamation, antimicrobial peptide activity, toxin exclusion, and selective chemical absorption. By comparison, corneocytes provide mechanical reinforcement, hydration, cytokine-mediated initiation of inflammation, and protection from UV damage.

Mutations in the transglutaminase enzymes and in structural proteins such as filaggrin and loricrin cause the ichthyoses and keratodermas, which result when a poorly formed epidermal barrier, both functionally and structurally,

E-TABLE 435-1 KERATIN EXPRESSION AND ASSOCIATED DISEASES

KERATIN PAIRS	EXPRESSION PATTERN	DISEASE PHENOTYPE
K1/K10	Suprabasal keratinocytes	Bullous congenital ichthyosiform erythroderma; diffuse nonepidermolytic PPK
K1/K9	Palmoplantar suprabasilar keratinocytes	Epidermolytic PPK
K2e/K10	Upper spinous and granular layers	Ichthyosis bullosa of Siemens
K3/K12	Cornea	Meesmann's corneal dystrophy
K4/K13	Mucosal epithelium	White sponge nevus
K5/K14	Basal keratinocytes	Epidermolysis bullosa simplex
K6a/K16	Outer root sheath (hair), hyperproliferative keratinocytes, palmoplantar keratinocytes	Pachyonychia congenita type I; focal nonepidermolytic PPK
K6b/K17	Nail bed, epidermal appendages	Pachyonychia congenita type II; steatocystoma multiplex
K8/K18	Simple epithelium	Cryptogenic cirrhosis

PPK = palmoplantar keratoderma

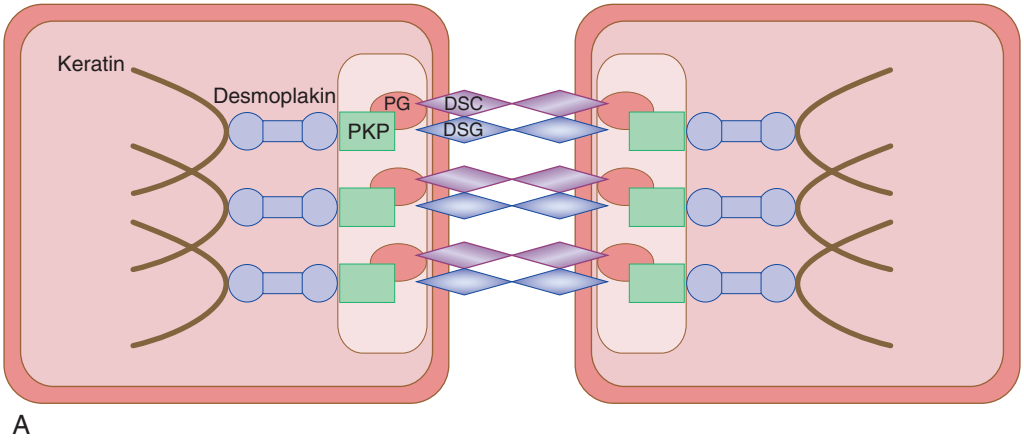

A

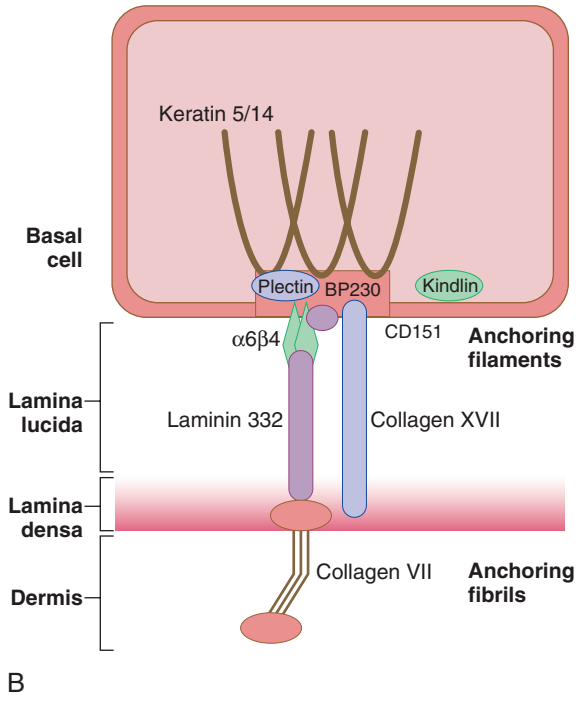

B

E-FIGURE 435-1. Adhesion in the epidermis. Cell-cell adhesion between keratinocytes is achieved via (**A**) desmosomes, whereas the basal cell keratinocytes attach to the basal lamina via (**B**) hemidesmosomes. alpha6β4 = alpha6β4 integrin; DSC = desmocollin; DSG = desmoglein; PG = plakoglobin; PKP = plakophilin. Note that kindlin-1 is an adhesion protein that has been localized not to hemidesmosomes but rather to focal contacts, adhesion complexes that facilitate association of the actin cytoskeleton to the basement membrane.

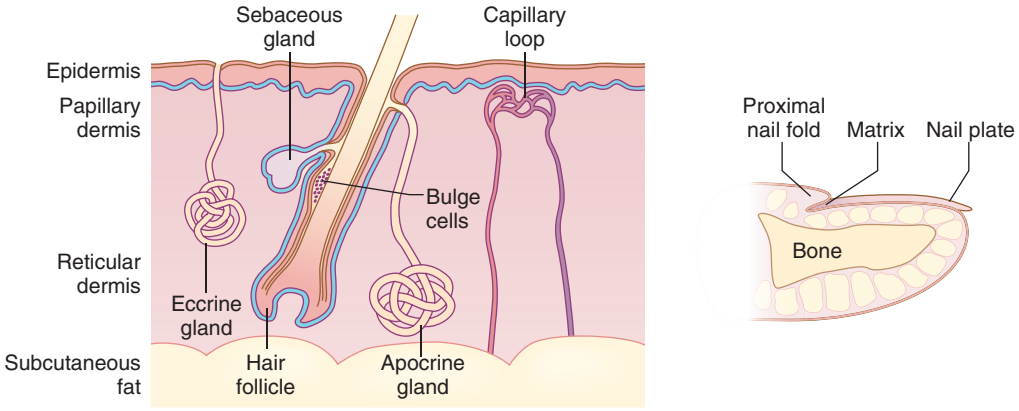

FIGURE 435-1. General structure of the skin and epidermal appendages. The epidermis is the outermost keratinized layer of skin, resting on top of the dermis via a basement membrane zone. The epidermal appendages are specialized structures and include the eccrine and apocrine sweat glands, sebaceous glands, hair, and nails. Capillary loops traverse the dermis to its most superficial aspect for diffusion and transport.

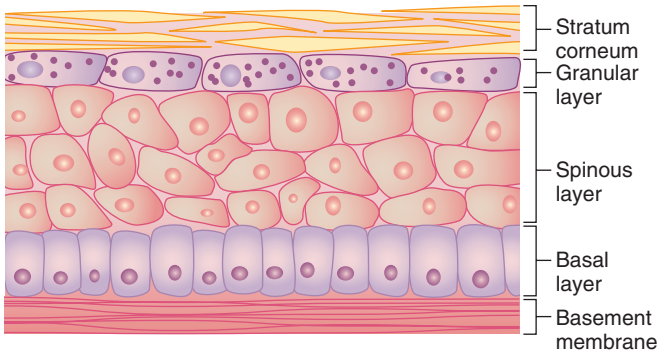

FIGURE 435-2. The epidermis and its layers of differentiation.

causes hyperkeratosis and scaling, derangements in water permeability and temperature regulation, and increased susceptibility to infections. Patients with filaggrin mutations are more susceptible to developing atopic dermatitis.[2] The sequelae of these mutations highlight the importance of the epidermis and stratum corneum for maintaining tissue and organism homeostasis.

Epidermal Adhesion

Keratinocytes are held together by a number of intercellular junctions, including tight junctions, gap junctions, adherens junctions, and desmosomes. Complex protein networks form each of these types of junctions, and genetic mutations can have significant effects on the function of the epidermis (see E-Table 435-1).

Tight junctions determine epidermal permeability by regulating the distribution of water-soluble molecules between adjacent cells. The primary structural proteins in these junctions are the claudins, and mutations in claudin 1 can cause a syndromic form of ichthyosis.

Gap junctions, formed primarily by connexins, allow small molecules and electric current to pass between adjacent cells. Connexin mutations can cause a number of phenotypes in genetic syndromes, such as the keratitis-ichthyosis deafness syndrome.

Adherens junctions are transmembrane structures that associate with the actin cytoskeleton. The extracellular interactions are calcium-dependent reactions between cadherins. These cadherin proteins associate with an intracellular complex including p120ctn, β-catenin, and plakoglobin. Plakoglobin and p-cadherin have been identified to be mutated in two genetic conditions, Naxos disease (plakoglobin) and hypotrichosis with juvenile macular dystrophy (p-cadherin).

Desmosomes are cell-cell junctions that provide strength and rigidity to cells (see E-Fig. 435-1). They also participate in skin and tissue differentiation and development. Three protein families make up desmosomes—the armadillo proteins (such as plakoglobin and plakophilins), cadherins (including desmocollins and desmogleins), and plakins (desmoplakin, envoplakin, periplakin, plectin, BPAG1, corneodesmosin, and microtubule actin cross-linking factor). A number of genetic conditions involve mutations within these protein-encoding genes, such as ectodermal dysplasia–skin fragility syndrome (plakophilin 1). In addition, autoantibody diseases (e.g., bullous pemphigoid and pemphigus vulgaris) and bacterial infections (e.g., bullous impetigo and staphylococcal scalded skin syndrome) can inactivate these proteins, thereby resulting in numerous adhesion disorders that are manifested clinically as skin fragility and blistering conditions.

Calcium is an important mediator of cell adhesion, as evidenced by Hailey-Hailey and Darier's diseases, in which mutations in genes encoding calcium transporters disrupt calcium distribution, thereby resulting in characteristic epidermal dyscohesion and blistering. Many of the proteins involved in cell-cell adhesion also require calcium for their binding or adhesive properties.

Nonkeratinocytes of the Epidermis

Melanocytes synthesize the pigment molecule melanin, which performs a critical role in absorbing and mitigating UV damage to the skin and underlying tissues.[3] A number of enzymes convert the amino acid tyrosine into different forms of melanin, which then are sorted into membrane-bound organelles known as melanosomes. These melanosomes are then transported to the tips of melanocytic dendrites and subsequently transferred to adjacent keratinocytes. Melanocytes are neuroendocrine derived cells that are not intrinsic cells of the skin but migrate there early in development.

The function of melanocytes has been highlighted by disorders in melanocyte number or function. The classic dermatologic disease, vitiligo (Chapter 441), is caused by the autoimmune depletion of melanocytes, but disorders of pigmentation can be caused by dysfunction at any of the steps of melanogenesis. For example, various forms of oculocutaneous albinism result from defects in melanin synthesis, whereas the subtypes of Griscelli's syndrome (skin and hair hypopigmentation, ocular albinism, neurologic disorders) result from mutations involved in lysosomal and melanosomal transfer of pigment to keratinocytes. Keratinocyte-melanocyte interactions are critical for melanocyte homeostasis and differentiation, influencing proliferation, dendricity, and melanization.

Merkel cells, which are mechanoreceptors, are located in sites of high tactile sensitivity, including skin of the digits, lips, regions of the oral cavity, and the hair follicle. Ultrastructurally, Merkel cells are easily identified by the membrane-bounded, dense-core granules that contain neurotransmitter-like substances and markers of neuroendocrine cells. Merkel cell–derived neoplasms are particularly aggressive and difficult to treat.

Langerhans cells are dendritic antigen-processing and antigen-presenting cells in the epidermis, mostly in a suprabasal position. Although they are not unique to the epidermis, they form 2 to 8% of the total epidermal cell population. Langerhans cells, which present antigens to T cells of the epidermis, are implicated in allergic contact dermatitis, cutaneous leishmaniasis, and human immunodeficiency virus (HIV) infection. Langerhans cells are reduced in the epidermis of patients with conditions such as psoriasis, sarcoidosis, and contact dermatitis, and they are functionally impaired by UV radiation, especially UVB.

Nonresident Cells of the Epidermis

Circulating immune cells that pass through the skin play key roles in infection control and immune surveillance. The activation of γ-δ T cells in the

epidermis is an important step in wound healing. Recent advances in melanoma therapy (Chapter 203) target T-cell costimulatory molecules, thereby underscoring the importance of this subset of immune cells in host responses to malignancy.

Epidermal Appendages

The epidermal appendages, including the hair follicle, nails, sebaceous gland, eccrine gland, and apocrine gland (see Fig. 435-1), are specialized structures that have particular unique functions.

Eccrine glands occur throughout the skin, but they are found in the highest concentrations on the palms, soles, and head. Eccrine glands, which are critical for thermoregulation, secrete a watery solution that includes sodium chloride, urea, uric acid, potassium, and immunoglobulins. Secretion is controlled by cholinergic sympathetic nerves.

Apocrine glands, which are scent glands, are found in the axilla, external genitalia, areola, and perianal areas. Secretion by these glands is increased by fear, sexual excitement, and other situations that result in heightened tension.

The nail unit (Chapter 442) consists of a nail matrix, nail plate, and nail bed. The proximal nail matrix is an organized germinative epithelium that produces the keratinized nail plate. The nail bed, which is a specialized epithelium located under the nail plate, attaches the nail plate to the digit.

Sebaceous glands are found throughout the dermis except for the palms and soles. Many of these glands empty their contents through a duct into the lumen of hair follicles. In certain areas of the body, the glands occur in the absence of hairs (glans penis, lips, labia minora, and eyelids). Sebaceous secretions include triglycerides, waxy esters, squalene, cholesterol, and fatty acids.

Hair follicles undergo a carefully orchestrated program of cyclical differentiation and proliferation. The hair follicle bulge consists of pluripotent skin stem cells that can give rise to all the layers of the follicle and hair shaft, as well as the epidermis under certain conditions of wound repair. The hair cycle (Chapter 442) consists of the phases of growth (anagen), involution (catagen), resting (telogen), and eventual release (exogen) of the mature hair follicle.

THE DERMAL-EPIDERMAL JUNCTION

The dermal-epidermal junction is a basement membrane zone that forms the interface between the epidermis and dermis. The major functions of the dermal-epidermal junction are to attach the epidermis and dermis to each other and provide resistance against external shearing forces. This junction serves as a support for the epidermis, determines the polarity of growth, directs the organization of the cytoskeleton in basal cells, provides developmental signals, and serves as a semipermeable barrier.

The dermal-epidermal junction can be subdivided into three supramolecular networks: the hemidesmosome-anchoring filament complex, the basement membrane, and the anchoring fibrils. The critical role of this region in maintaining skin structural integrity is exemplified by the large number of mutations or functional inactivations that cause blistering diseases[4,5] (E-Table 435-2). For example, basal keratinocyte cleavage within the superficial dermal-epidermal junction causes epidermoloysis bullosa simplex. Abnormalities within the lamina lucida and lamina densa regions cause junctional epidermolysis bullosa, and abnormalities within the sublamina densa/anchoring filaments cause the deep blistering of dystrophic epidermolysis bullosa.

THE DERMIS

The dermis, which constitutes the bulk of the skin and provides its flexibility and mechanical strength, is an integrated system of connective tissue elements that accommodate networks of nerves and blood vessels. The dermis has two major regions: the upper papillary dermis and the deeper reticular dermis. The papillary dermis, which is usually no more than twice the thickness of the epidermis, runs just beneath it and undulates its contours. The reticular dermis, which forms the bulk of the dermal tissue, is composed primarily of collagen and elastin. The subpapillary plexus, a horizontal plane of vessels, marks the boundary between the papillary and reticular dermis. The lowest boundary of the reticular dermis is defined by the transition of its fibrous connective tissue to the adipose connective tissue of the subcutaneous fat.

Connective Tissue Matrix of the Dermis

Collagen, which forms the bulk of the acellular portion of the dermis, provides both tensile strength and elasticity. The periodically banded, interstitial collagens account for most of the collagen in the adult dermis (type I, 80 to 90%; III, 8 to 12%; and V, < 5%). Type IV collagen is confined to the basal lamina of the dermal-epidermal junction, blood vessels, and epidermal appendages. Type VI collagen is associated with fibrils and interfibrillar spaces. Type VII collagen forms anchoring fibrils at the dermal-epidermal junction. The classic diseases of collagen function are the various subtypes of Ehlers-Danlos syndrome (Chapter 260),[6] a heterogeneous collection of disorders characterized by joint hypermobility, skin extensibility, abnormal scarring, and tissue friability.

Elastic connective tissue is a complex molecular mesh that includes elastin and fibrillin and that extends from the lamina densa of the dermal-epidermal junction throughout the dermis and into the connective tissue of the subcutaneous fat. Elastic fibers return the skin to its normal configuration after it has been stretched or deformed. Mutations in elastin, which is the elastic fiber matrix component, causes the disease cutis laxa, a condition that is characterized by sagging skin that has little extensibility and hangs in loose folds, especially noticeable on the neck and in the axillae and groin. Mutations in the gene encoding fibrillin, a microfibril component, are implicated in Marfan syndrome (Chapter 260). Elastic fibers are normally located between bundles of collagen fibers, although in certain pathologic conditions, such as Buschke-Ollendorff syndrome (dermatofibrosis lenticularis disseminata), both elastic and collagen fibers become assembled within the same bundle, thereby resulting in the formation of characteristic benign connective tissue nevi and osteopoikilosis.

Pseudoxanthoma elasticum (Chapter 260), which is characterized by loss of skin elasticity and calcified elastic fibers, results from mutations in ABCC6, a transmembrane transporter in the ATP-binding cassette transporter family, although the mechanisms responsible for the observed elastic fiber defects are not understood. In addition to genetic mutations, solar radiation and aging also damage elastic fibers.

The fibrous and cellular elements of the dermis are embedded within a more amorphous matrix that can bind water and regulate the compressibility of the dermis. Various glycoproteins interact with other matrix components via integrin receptors to facilitate cell migration, adhesion, morphogenesis, and differentiation. Fibronectin, which is synthesized by both epithelial and mesenchymal cells, covers collagen bundles and the elastic network. Vitronectin is present on all elastic fibers except for oxytalan. Tenascin is found around the smooth muscle of blood vessels, arrector pili muscles, and appendages such as sweat glands.

Cellular Components of the Dermis

Fibroblasts, macrophages, and mast cells are the regular residents of the dermis. Fibroblasts migrate through the tissue and are responsible for the synthesis and degradation of matrix proteins and a number of soluble factors in the fibrous and nonfibrous connective tissue. Fibroblasts provide a structural extracellular matrix framework and also promote interactions between the epidermis and the dermis. Fibroblasts are also instrumental in wound healing and scarring, increasing their proliferative and synthetic activity during these processes.

Monocytes, macrophages, and dermal dendrocytes constitute the mononuclear phagocytic system of cells in the skin. Macrophages are derived from precursors in the bone marrow, differentiate into circulating monocytes, and then migrate into the dermis to differentiate. These cells are phagocytic, antigen-processing and antigen-presenting, microbicidal, tumoricidal, secretory, and hematopoietic. They are also involved in coagulation, atherogenesis, wound healing, and tissue remodeling. The dermal dendrocyte is a phagocytic fixed connective tissue cell in the dermis of normal skin. These cells are particularly abundant in the papillary dermis and upper reticular dermis, where they function as antigen-presenting cells. They are also likely the cell of origin of a number of benign fibrotic proliferative conditions in the skin, such as dermatofibromas and fibroxanthomas.

Mast cells, which are the specialized secretory cells responsible for the immediate-type hypersensitivity reaction in skin, are also involved in subacute and chronic inflammatory processes. These processes are mediated by the histamine, heparin, tryptase, chymase, carboxypeptidase, neutrophil chemotactic factor, and eosinophilic chemotactic factor of anaphylaxis that are synthesized in their granules and released in response to various stimuli. Mast cells also can become hyperplastic and hyperproliferative in mastocytosis (Chapter 255).

SUBCUTANEOUS FAT

The subcutaneous fat insulates the body, serves as a reserve energy supply, cushions and protects the skin, and allows for its mobility over underlying structures. This fat is a key determinant of body contours and therefore plays

E-TABLE 435-2 HEMIDESMOSOMAL AND BASEMENT MEMBRANE ZONE COMPONENTS IN SKIN DISEASE

BMZ REGION	PROTEINS	ACQUIRED DISEASES	HEREDITARY DISEASES
Cytoskeletal proteins	Keratin 5 and 14		EB simplex
Hemidesmosomal plaque protein	BPAG11/BP230 Plectin	Bullous pemphigoid Bullous pemphigoid, cicatricial pemphigoid	EB simplex
Other intracellular adhesion complex proteins	Kindlin-1		Kindler syndrome
Hemidesmosomal transmembrane components	Collagen XVII/BP180 alpha6β4 integrin CD151	Bullous pemphigoid, cicatricial pemphigoid, linear IgA dermatosis, pemphigoid getationis Bullous pemphigoid, cicatricial pemphigoid	JEB-non-Herlitz JEB w/pyloric atresia Pretibial EB, nephritis, deafness, β-thalassemia minor
Anchoring filament proteins	Laminin 332 Ectodomain of collagen XVII	Cicatricial pemphigoid Linear IgA dermatosis, bullous pemphigoid	JEB-Herlitz JEB-non-Herlitz JEB-non-Herlitz
Anchoring fibril proteins	Collagen VII	EB acquisita	Dystrophic EB

BMZ = basement membrane zone; EB = epidermolysis bullosa; JEB = junctional EB.

an important cosmetic role as well. At the junction of the deep reticular dermis and the underlying fat layer is an abrupt transition from predominantly fibrous dermal connective tissue to primarily adipose tissue. Despite this anatomic contrast, the dermis and the subcutaneous fat are structurally and functionally integrated through networks of nerves and vessels, as well as in the continuity of epidermal appendages such as growing hair follicles and sweat glands.

Adipocytes, which form the bulk of the cells in the subcutaneous fat layer, are organized into lobules defined by septa of fibrous connective tissue. Within the septa run the nerves and vascular structures that supply the region. Fat synthesis and storage can occur by enhanced accumulation of lipid within, or proliferation of, existing adipocytes or by the creation of new adipocytes from undifferentiated mesenchyme. The hormone leptin (Chapter 223), secreted by adipocytes, regulates fat homeostasis. Leptin levels are higher in subcutaneous than omental adipose tissue, thereby suggesting a role for leptin in the distribution of adipose tissue.

The importance of the subcutaneous tissue is highlighted in conditions in which it is absent or abnormal. In Werner's syndrome (Chapter 205), the absence of subcutaneous fat over bone lesions results in ulcers and poor wound healing. In scleroderma (Chapter 267), the replacement of subcutaneous fat with dense fibrous connective tissue results in taut and painful skin. In both the hereditary and acquired lipodystrophies (Chapter 392), loss of subcutaneous fat disrupts glucose, triglyceride, and cholesterol regulation and can cause significant cosmetic alteration. Inflammation of the subcutaneous fat, termed panniculitis (Chapters 142 and 266), can be caused by many different tissue derangements and systemic conditions.

CUTANEOUS VASCULATURE

Blood Vessels

The rich vascular network of the skin is located at boundaries within the dermis and supplies the epidermal appendages. Dermal vessels branch from musculocutaneous arteries that penetrate the subcutaneous fat and enter the deep reticular dermis. At this point, they are organized into a horizontal arteriolar plexus. From this plexus, ascending arterioles extend toward the epidermis. At the junction between the papillary and reticular dermis, terminal arterioles form the subpapillary plexus. Capillary loops then extend from these terminal arterioles of the plexus into the more superficial papillary dermis. At the apex of each capillary loop, the thinnest portion of the vessel allows for diffusion and transport of material out of the capillary. The descending limbs of capillary loops drain into venous channels of the subpapillary plexus. The postcapillary venules of the subpapillary plexus are responsive to histamine and are therefore often the sites of inflammatory cells.

In the adult, the cutaneous vasculature normally remains quiescent, in part owing to inhibition of angiogenesis by factors such as thrombospondin. Pathogenic stimuli, such as from tumors or after a wound, sometimes result in secondary angiogenesis. One of the key mediators of such angiogenesis is vascular endothelial growth factor, which often is secreted by tumors or by keratinocytes.

Numerous disorders can manifest themselves within the cutaneous vasculature.[7] Leukocytoclastic vasculitis (Chapters 270 and 439), also called cutaneous necrotizing venulitis, occurs within the venules in response to a number of potential pathogenic mechanisms, including medication reactions, infections, neoplasms, and systemic inflammatory conditions. Stasis dermatitis (Chapter 436), urticaria (Chapters 252 and 440), polyarteritis nodosa (Chapter 270), thrombosis (Chapter 70), and thrombophlebitis (Chapter 81) all affect different-sized vessels in the skin, some by occlusion of vessels (vasculopathy) and others by inflammation of the vessels (vasculitis).

Lymphatics

The lymph channels of the skin are responsible for resorbing fluid released by vessels and clearing the tissues of cells, proteins, lipids, bacteria, and degraded substances.[8] These vessels begin in blind-ending loops in the papillary dermis and then continue to drain into successively larger plexuses deeper in the tissue. Lymphatic flow in the skin is propelled by arterial pulsations, muscle contractions, and movement of the body. Bicuspid-like valves within the lymphatic vessels promote unidirectional flow. Because lymphatic vessels are often collapsed in skin and because lymphatic vessels have thinner walls than blood vessels, they are rarely seen on routine histologic section.

Pathologic conditions that involve or highlight the function of lymphatic vessels include lymphedema, lymphangioma circumscriptum (clustered, deep vesicles on the skin resulting from lymphatic dilation and malformation), and stasis dermatitis. Lymphatics are also important for the progression and spread of cancer. For example, melanoma cells destroy the endothelial cells of the local lymphatics to gain entry to the lymphatic circulation, and tumors themselves can promote lymphangiogenesis as part of their process of metastasis.

CUTANEOUS NERVES AND RECEPTORS

The nerve networks of the skin contain somatic sensory and sympathetic autonomic fibers. The sensory fibers alone (free nerve endings) or in conjunction with specialized structures (corpuscular receptors) function as receptors of touch, pain, temperature, itch, and mechanical stimuli. The differing density and types of receptors in different body regions account for the variation in sensory acuity at different body sites. Receptors are particularly dense in hairless areas such as the areola, labia, and glans penis. Sympathetic motor fibers run with the sensory nerves in the dermis until they branch to innervate the sweat glands, vascular smooth muscle, the arrector pili muscle of hair follicles, and sebaceous glands.

The nerves of skin branch from musculocutaneous nerves that arise segmentally from spinal nerves. The pattern of nerve fibers in skin is similar to the vascular patterns—nerve fibers form a deep plexus, then ascend to a superficial, subpapillary plexus.

Free nerve endings, which are the most widespread sensory receptors in skin, are particularly common in the papillary dermis. The penicillate fibers, which are the primary nerve fibers found subepidermally in haired skin, are rapidly adapting receptors that function in the perception of touch, temperature, pain, and itch. Papillary nerve endings are found at the orifice of a follicle and are thought to be particularly receptive to cold sensation.

Corpuscular receptors, which contain a capsule and inner core, are composed of both neural and non-neural components. Meissner corpuscles are elongated or ovoid mechanoreceptors that are located in the dermal papillae of digital skin and oriented vertically toward the epidermal surface. Pacinian corpuscles lie in the deep dermis and subcutaneous tissue of skin that covers weight-bearing surfaces of the body; they have a characteristic capsule and lamellar wrappings and serve as rapidly adapting mechanoreceptors that respond to vibrational stimuli.

Pathophysiology of Pruritus

Pruritus (itching) is mediated by unmyelinated C fibers, the same fibers that are responsible for the transmission of pain.[9] The receptors for these fibers are probably the free nerve endings in the dermis and epidermis. Stimuli that trigger these nerve fibers result in the transmission of signals along peripheral nerves to the dorsal root ganglion, spinal cord, and finally to the thalamus and other parts of the brain (E-Fig. 435-2).

Pruritus, which can be subdivided into clinical categories based on the presumptive mechanism (Table 435-1),[10] is associated with a number of systemic as well as skin-specific diseases (Table 435-2). Antihistamines, although commonly prescribed, are ineffective for many causes of pruritus—with the exception of urticaria—because they primarily cause sedation. Depending on the suspected mechanism of pruritus, classes of medication as diverse as the opiate antagonists (e.g., naltrexone), antidepressants (e.g., mirtazapine), anticonvulsants (e.g., gabapentin), and substance P antagonists (e.g., aprepitant) have proved to be useful (Tables 435-3 and 435-4). Although the precise targets of these classes of medications are poorly understood in pruritus, the clinical effectiveness of these agents may offer insight regarding the molecular and cellular pathways involved in these conditions.[11] Ultraviolet

TABLE 435-1 PATHOPHYSIOLOGY OF PRURITUS

PRURITUS CATEGORY	PRURITUS MEDIATOR	CLINICAL EXAMPLES
Pruritoceptive	Generated in skin, usually inflammatory	Atopic dermatitis, allergic contact dermatitis, lichen planus
Neurogenic (systemic)	Generated in central nervous system but without any neural pathology	Renal failure, liver disease, lymphoproliferative disorders, malignancy
Neuropathic	Neuronal pathology along afferent pathway	Brachioradial pruritus, notalgia paresthetica
Psychogenic	Caused by psychological disorder	Delusions of parasitosis

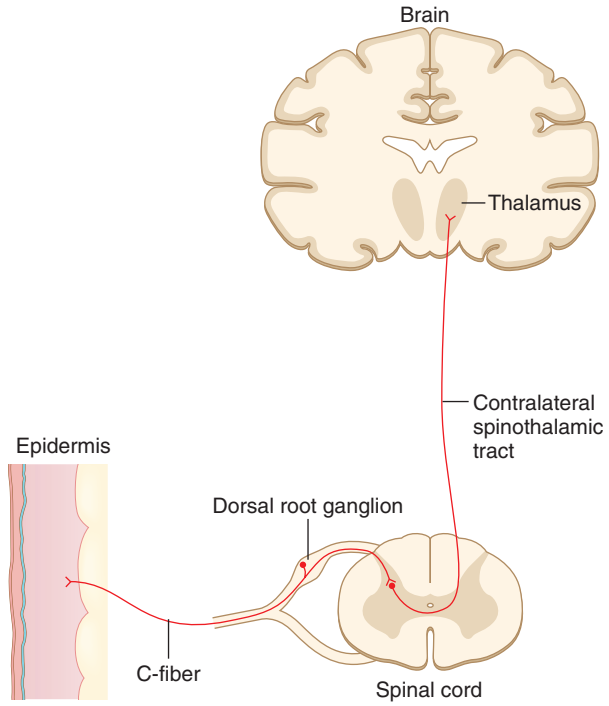

E-FIGURE 435-2. Pathways of pruritus. Itch triggered in the skin sends signals via nerve fibers to the dorsal root ganglion, spinal cord, and ascends via the contralateral spinothalamic tract to project to the thalamus. From there, itch is transmitted to several other regions of the brain.

TABLE 435-2 TERMINOLOGY TO DESCRIBE THE MORPHOLOGY OF INDIVIDUAL SKIN LESIONS

TERM	DEFINITION	EXAMPLE
PRIMARY SKIN LESIONS: INITIAL PATHOLOGIC CHANGE		
Macule	Circumscribed change in skin color that is flush with the surrounding skin. Lesion is <1 cm in diameter	Solar lentigo Traumatic purpura
Patch	Circumscribed change in skin color that is flush with the surrounding skin. Lesion is ≥1 cm in diameter	Café au lait spot Vitiligo
Papule	A solid or cystic elevation <1 cm in diameter	Acne Eruptive xanthoma
Nodule	A solid or cystic elevation >1 cm but <2 cm in diameter	Dermatofibroma
Tumor	A solid or cystic elevation >2 cm in diameter	Follicular cyst
Plaque	An elevated lesion that is >1 cm in diameter	Psoriasis
Scale	Desiccated, thin plates of cornified epidermal cells that form flakes on the skin surface	Ichthyosis
Wheal	Circumscribed, flat-topped, firm elevation of skin with a well-demarcated and palpable margin	Urticaria
Vesicle	Circumscribed, elevated lesion containing clear serous or hemorrhagic fluid that is <1 cm in diameter	Contact dermatitis Herpes simplex
Bulla	Circumscribed, elevated lesion containing clear serous or hemorrhagic fluid that is >2 cm in diameter	Bullous pemphigoid
Pustule	A vesicle containing purulent exudate	Folliculitis
Atrophy	A depression from the surface of the skin with underlying loss of epidermal or dermal substance	Lichen sclerosis et atrophicus
Erosion	A depression from the surface of the skin with a loss of all or part of the epidermis Can be a secondary lesion	Burn Ruptured bulla
Ulceration	A depression from the surface of the skin with a loss of the entire epidermis and at least some of the dermis Can be a secondary lesion	Ecthyma Excoriation of acne papule
SECONDARY SKIN LESIONS: RESULT FROM EXTERNAL FORCES SUCH AS SCRATCHING, PICKING, INFECTION, OR HEALING OF PRIMARY LESIONS		
Lichenification	Dry, leathery thickening of the skin with exaggerated skin markings	Chronic eczema
Scar	An elevated or depressed area of fibrosis of the dermis or subcutaneous tissue resulting from an antecedent destructive process	Healing wound
Fissure	A deep linear split in the skin extending through the epidermis	Traumatized eczema
Crust	Dried exudates of serum, blood, sebum, or purulent material on the surface of the skin	Impetigo

Modified from Armstrong CA. Examination of the skin and approach to diagnosing skin diseases. In: Goldman L, Schafer AI, eds. *Goldman's Cecil Medicine.* 24th ed. Philadelphia: Saunders; 2012: Table 444-1.

TABLE 435-3 TOPICAL TREATMENTS FOR PRURITUS

MEDICATION	DOSE	COMMENTS
Emollients	Variable	For skin barrier damage, dry skin itch
Corticosteroids	Variable	Useful in pruritus due to inflammatory skin dermatitides. Low-potency agents safest for use in children and on face and in skin folds
Calcineurin inhibitors	Tacrolimus 0.03 and 0.1% ointment Pimecrolimus 1% cream	For use in atopic dermatitis and contact dermatitis. Particularly useful in facial or anogenital pruritus. May cause transient burning and stinging
Doxepin	5% cream	Topical formulation of tricyclic antidepressant; 20-25% risk for sedation owing to systemic absorption
Menthol	1-5% cream	Useful in patients who report cooling as an alleviating factor. Higher concentrations can cause hypersensitivity reactions and burning sensation
Anesthetic agents		
Lidocaine/prilocaine	2.5-5%	Useful for neuropathic and postburn itch. Risk for methemoglobinemia
Capsaicin	0.025-0.1% cream	Particularly useful in neuropathic itch and itch caused by chronic kidney disease. May cause transient burning
Pramoxine	1-2.5%	Useful on face and genitals, for chronic kidney disease, and for neuropathic itch

Data from Yosipovitch G, Bernhard JD. Chronic pruritus. *N Engl J Med.* 2013;368:1625-1634.

TABLE 435-4 SYSTEMIC TREATMENTS FOR PRURITUS

MEDICATION	DOSE	COMMENTS
Antihistamines	Hydroxyzine 10-50 mg four times daily Cetirizine 10 mg/day Fexofenadine 60-180 mg/day	Sedating, but no direct effect on pruritus, except in urticaria
Antidepressants		
Tricyclic antidepressants	Amitriptyline 25-150 mg/day	Neuropathic itch. May cause drowsiness, dizziness, constipation, urinary retention, blurred vision, palpitations, low blood pressure
Noradrenergic and specific serotonergic antidepressants	Mirtazapine 7.5-15 mg PO at bedtime	Nocturnal pruritus. May increase appetite and weight
Selective serotonin reuptake inhibitors	Paroxetine 10-40 mg/day PO Fluvoxamine 25-150 mg/day PO Sertraline 50-200 mg/day PO	Psychiatric patients and paraneoplastic pruritus Psychiatric patients and paraneoplastic pruritus Cholestatic pruritus
Opioids		
μ Antagonist	Naltrexone 25-50 mg/day PO	Pruritus associated with cholestatic or chronic kidney disease. Can cause nausea, vomiting, and drowsiness
κ Agonist and μ antagonist	Butorphanol, 1-4 mg inhaled at bedtime	Intractable itch. May cause nausea, vomiting, and drowsiness
Anticonvulsants	Gabapentin 100-1200 mg PO three times per day Pregabalin 25-200 mg PO twice per day	Neuropathic itch and pruritus from chronic kidney disease. Can cause drowsiness, weight gain, and leg swelling
Substance P antagonist	Aprepitant 80 mg/day PO	Sézary syndrome
Immunosuppressants	Cyclosporin 2.5-5 mg/kg/day PO Azathioprine 2.5 mg/kg/day PO	Short-term use for refractory atopic dermatitis. Monitor blood pressure and renal function Refractory atopic dermatitis. Monitor for myelosuppression
Ultraviolet B radiation (broad and narrow band)	Three times per week	Atopic dermatitis, psoriasis, pruritus from chronic kidney disease

Data from Yosipovitch G, Bernhard JD. Chronic pruritus. *N Engl J Med.* 2013;368:1625-1634.
PO = orally.

phototherapy is also effective for multiple forms of pruritus, including pruritus that results from chronic kidney disease, as well as more inflammatory conditions such as psoriasis and atopic dermatitis; however, the mechanism underlying the resulting improvement is poorly understood.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Eckhart L, Lippens S, Tschachler E, et al. Cell death by cornification. *Biochim Biophys Acta*. 2013;1833:3471-3480.
2. Kubo A, Nagao K, Amagai M. Epidermal barrier dysfunction and cutaneous sensitization in atopic diseases. *J Clin Invest*. 2012;122:440-447.
3. Bonaventure J, Domingues MJ, Larue L. Cellular and molecular mechanisms controlling the migration of melanocytes and melanoma cells. *Pigment Cell Melanoma Res*. 2013;26:316-325.
4. Baum S, Sakka N, Artsi O, et al. Diagnosis and classification of autoimmune blistering diseases. *Autoimmun Rev*. 2014;13:482-489.
5. Kershenovich R, Hodak E, Mimouni D. Diagnosis and classification of pemphigus and bullous pemphigoid. *Autoimmun Rev*. 2014;13:477-481.
6. Malfait F, De Paepe A. The Ehlers-Danlos syndrome. *Adv Exp Med Biol*. 2014;802:129-143.
7. Wollina U, Unger L, Haroske G, et al. Classification of vascular disorders in the skin and selected data on new evaluation and treatment. *Dermatol Ther*. 2012;25:287-296.
8. Kesler CT, Liao S, Munn LL, et al. Lymphatic vessels in health and disease. *Wiley Interdiscip Rev Syst Biol Med*. 2013;5:111-124.
9. Misery L, Brenaut E, Le Garrec R, et al. Neuropathic pruritus. *Nat Rev Neurol*. 2014;10:408-416.
10. Yosipovitch G, Bernhard JD. Clinical practice: chronic pruritus. *N Engl J Med*. 2013;368:1625-1634.
11. Misery L, Brenaut E, Le Garrec R, et al. Neuropathic pruritus. *Nat Rev Neurol*. 2014;10:408-416.

REVIEW QUESTIONS

1. The proliferative layer of the epidermis is the _____.

A. Basal layer.
B. Spinous layer.
C. Granular layer.
D. Stratum corneum.
E. Stratum lucidum

Answer: A The basal layer is the most undifferentiated layer of the epidermis and the layer in which most proliferation occurs. The other, more superficial layers of the epidermis are formed from keratinocytic differentiation. The stratum lucidum is a compact layer between the granular layer and stratum corneum and is present only in palmoplantar skin.

2. Marfan syndrome is caused by a mutation in the gene encoding _____.

A. Elastin
B. Fibrillin-1
C. Elaunin
D. Type I procollagen
E. Type V collagen

Answer: B Mutations in fibrillin-1 cause Marfan syndrome. Mutations in elastin result in cutis laxa. Elaunin is an elastic fiber component that does not have a described genodermatosis. Mutations in type I procollagen cause osteogenesis imperfecta, whereas mutations in type V collagen can result in Ehlers-Danlos syndrome.

3. Which of the following does *not* play a role in the pathogenesis of itch?

A. Dorsal root ganglion
B. Thalamus
C. Pituitary gland
D. Spinothalamic tract neurons
E. Free nerve endings

Answer: C The pituitary gland is not involved in the itch pathway. The other choices are all found along the pathway for itch perception between the free nerve endings of the skin and the processing centers in the brain.

4. Defects in all of the following proteins can cause blistering diseases *except:*

A. Keratin 5
B. Laminin 332
C. Filaggrin
D. Collagen VII
E. Plectin

Answer: C Defects in filaggrin can lead to ichthyosis and atopic dermatitis, but not a blistering disease. Defects in keratin 5 and plectin can cause epidermolysis bullosa simplex. Mutations in laminin 332 can cause junctional epidermolysis bullosa. Collagen VII forms anchoring fibrils, defects in which can cause dystrophic epidermolysis bullosa.

436

EXAMINATION OF THE SKIN AND AN APPROACH TO DIAGNOSING SKIN DISEASES

JAMES C. SHAW

Dermatology encompasses well over a thousand disease entities, many of which expand further into multiple subclassifications, variants, and etiologies. The first goal of a clinician is quickly to recognize the few skin diseases that can rapidly cause severe morbidity or even kill the patient. For the remaining hundreds of non–life-threatening dermatoses, a careful history and physical examination, which is often performed before or simultaneously with the history, can help the clinician make an accurate diagnosis or obtain expert dermatologic referral for doing so.[1]

THE SKIN EXAMINATION

The basic requirements for good examination of the skin are lighting and magnification. The best lighting is natural daylight or window light, but bright fluorescent ceiling lights, surgical lamps, or specialized magnifying lights all suffice. For simple magnification, a 4× hand lens is highly effective, but a 10× handheld polarized dermatoscope is even better, especially for diagnosing melanoma[2] and basal cell carcinoma.

Color

Attention to the color of the skin lesions can be key to making a correct diagnosis (Table 436-1). The color of the adjacent normal skin also can influence the appearance of dermatoses. Subtle colors such as violaceous or yellow or even simple erythema can be more difficult to appreciate in darker skin types than in light-colored skin (Fig. 436-1). The lighter an individual's skin, the higher the risk is for developing sun-induced skin cancer (Chapter 203).

Palpation

Palpation with an ungloved hand is often important in dermatologic diagnosis. Examples include the skin fibrosis and sclerosis of scleroderma (Chapter 267), the induration and firmness of cellulitis (Chapter 290), the roughness of subtle actinic keratoses (Chapter 440), the firmness of a dermatofibroma (Chapter 440), the palpable purpura of leukocytoclastic vasculitis (Chapter 439), and the difference between a lipoma (soft, subcutaneous) and an epidermoid cyst (firmer, intradermal) or between a dermal nevus (somewhat firm) and a solitary neurofibroma (rubbery soft).

A glove should be worn for protection when examining patients who may have blood-borne diseases, such as hepatitis B (Chapters 148 and 149) and human immunodeficiency virus (HIV) infections, or patients who may have secondary syphilis (Chapter 319) or herpes simplex (Chapter 374). Gloves are also essential when the clinician wishes to examine mucosae, denuded skin, blood, or exudates. Routine skin infections such as scabies (Chapter 441), human papillomavirus infection (Chapter 373), and superficial fungal infections (Chapter 438) are not highly contagious, but handwashing after contact is important to prevent transmission.

Another clue to diagnosis is determining the depth of the lesion (Chapter 435) and whether it involves the epidermis, the dermis, the fat layer below the dermis, or more than one layer (Table 436-2). Epidermal pathology includes the disorderly yet benign growth of keratinocytes resulting in flaking (scaling) and a thickened epidermis (psoriasis); the presence of microvesiculation (histologic term: spongiosis) that leads to oozing of serum (contact dermatitis); and hyperplasia of the epidermis in benign lesions (seborrheic keratoses), infections (warts), or malignancies (basal cell carcinomas and squamous cell carcinomas).

Dermal pathology often leads to inflammation because of the recruitment of lymphocytes, neutrophils, and histiocytes, as well as infiltration of malignant cells (lymphoma), antibody deposition (lupus erythematosus, autoimmune blistering diseases, drug eruptions), infections, or vascular damage. Erythema, which is caused by vasodilation of vasculature, is the most common dermal finding, even in diseases that are primarily epidermal. Abnormalities are less common in the subcutaneous fat than in the epidermis or dermis, but benign hyperplasia of the fat layer results in benign lipomas. Vasculitis of midsize vessels will often extend into the fat layer in polyarteritis nodosa (Chapter 270), and several panniculitides (e.g., erythema nodosum) also involve the fat.

HISTORY

The two most important facts to establish immediately from the history are whether the problem is acute (new onset within the last few days or hours) and whether the patient is systemically ill. Key clues to the diagnosis come from whether the lesions are associated with pruritus or pain, and what medications, especially new medications, the patient is taking (Table 436-3). For example, the absence of pruritus makes the diagnosis of allergic contact dermatitis (Chapter 440) highly unlikely. In the setting of unilateral acute severe pain, burning, and itching, herpes zoster (Chapter 439) should always be considered, even in the absence of skin findings.

DIFFERENTIAL DIAGNOSIS

Morphology

In the ambulatory outpatient who does not have a life-threatening skin problem, lesions are commonly classified as macules, patches, papules, plaques, nodules, tumors, vesicles, bullae, or pustules (Table 436-4). However, plain language can be just as effective in arriving at an accurate diagnosis. For example, a description of "small, raised solid bumps 0.4 cm in diameter" is as clear as "multiple papules." Furthermore, if dermatologic terms are used incorrectly, clinicians may be misled toward an incorrect diagnosis. Although primary lesions (which develop de novo) are most helpful in leading to a correct diagnosis, secondary features (which are altered by superinfection, manipulation, healing, etc.) are also important.

Text continued on p. 2644

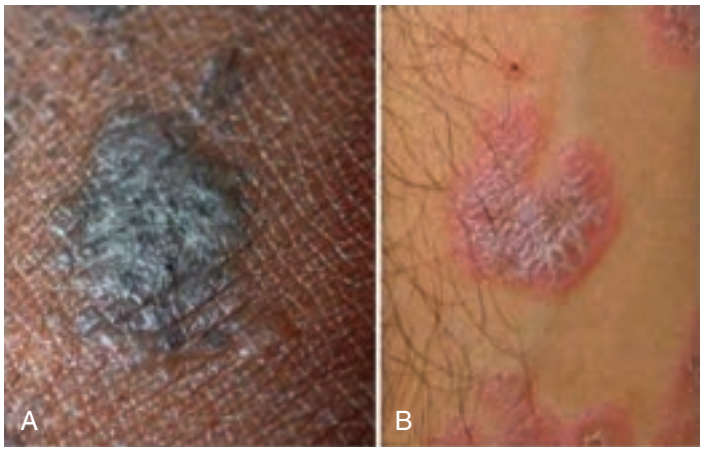

FIGURE 436-1. Skin pigmentation. Lichen planus presents differently in darkly pigmented (**A**) versus lightly pigmented (**B**) skin. The violaceous hue seen in **B** is more muted in **A**, and these lesions appear brown-black in color. Wickham striae (lacy white pattern) are more easily seen in **B**. (From Bolognia JL, Jorizzo JL, Schaffer JV. *Dermatology,* 3rd ed. Philadelphia: Saunders; 2012:8.)

TABLE 436-1 COLOR CLUES IN DIAGNOSING SKIN DISEASE

COLOR	CLINICAL EXAMPLE
Erythema: Pink Pink is a common finding when vasodilation (with minimal cellular infiltrate) is the primary dermal pathology.	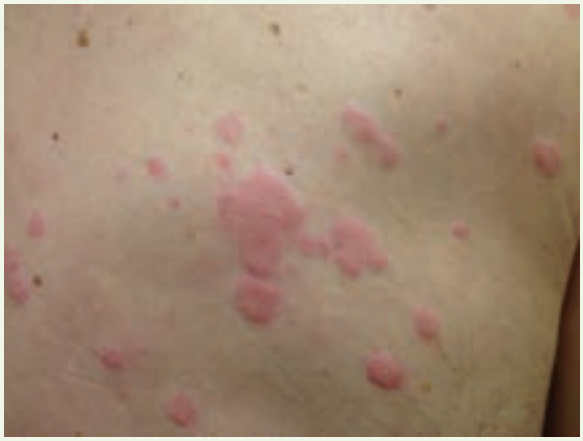 **FIGURE 436-2.** Urticaria (Chapter 440).
Erythema: Pink-Red Although many dermatoses can be pink-red, pityriasis rosea is classic.	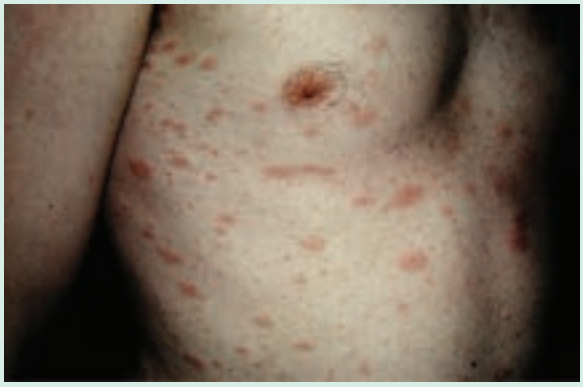 **FIGURE 436-3.** Pityriasis rosea (Chapter 438).
Erythema: Orange-Red (with Scale) Color in psoriasis ranges from orange-red to purple-red. A related condition, pityriasis rubra pilaris, also is orange-red.	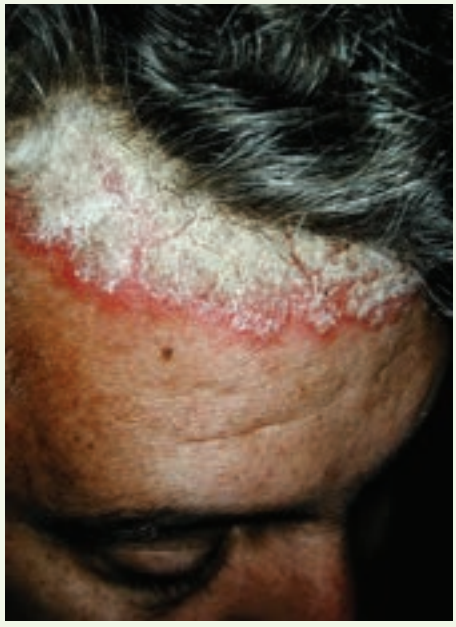 **FIGURE 436-4.** Psoriasis (Chapter 438).

TABLE 436-1 COLOR CLUES IN DIAGNOSING SKIN DISEASE—cont'd

COLOR	CLINICAL EXAMPLE
Erythema: Red-Orange-Pink A mixture of erythematous hues is common and makes diagnosis difficult by color alone.	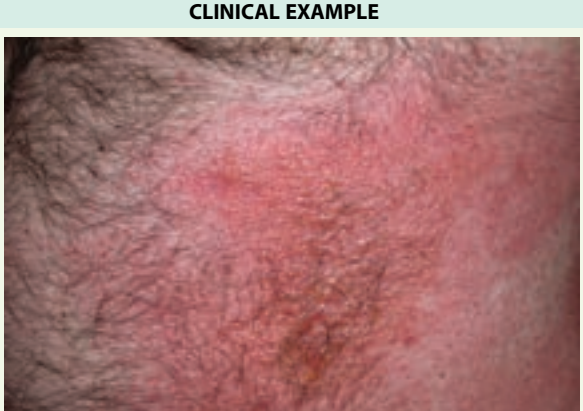 **FIGURE 436-5.** Allergic contact dermatitis (Chapter 438).
Erythema: Copper-Red Mixed cellular infiltrates with plasma cells, classically seen in syphilis.	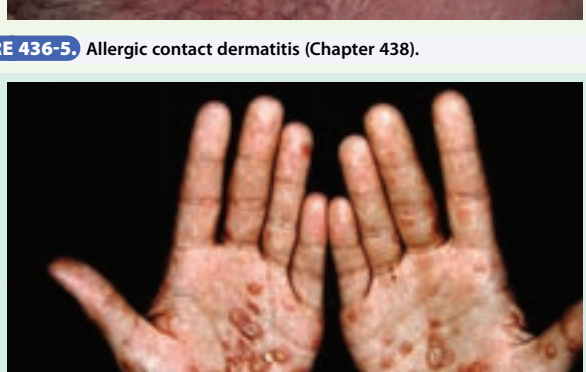 **FIGURE 436-6.** Secondary syphilis (Chapter 319).
Erythema: Red-Brown The color of histiocytic infiltrates, seen commonly in granulomatous diseases (e.g., cutaneous tuberculosis, deep fungal infections, atypical mycobacteria, sarcoidosis, granuloma annulare, leprosy, necrobiosis lipoidica) depends on depth of the infiltrates.	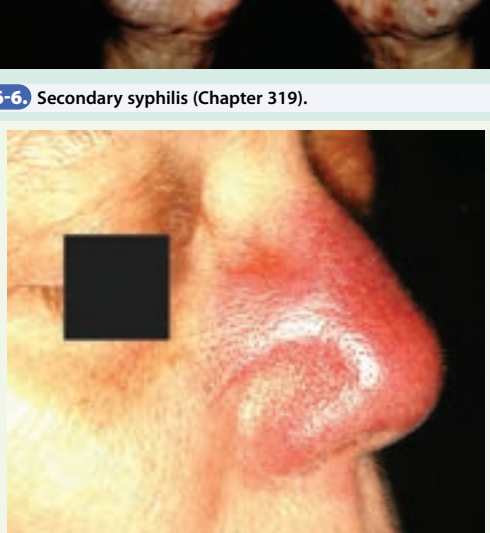 **FIGURE 436-7.** Sarcoidosis (Chapter 95).

TABLE 436-1 COLOR CLUES IN DIAGNOSING SKIN DISEASE—cont'd

COLOR	CLINICAL EXAMPLE
	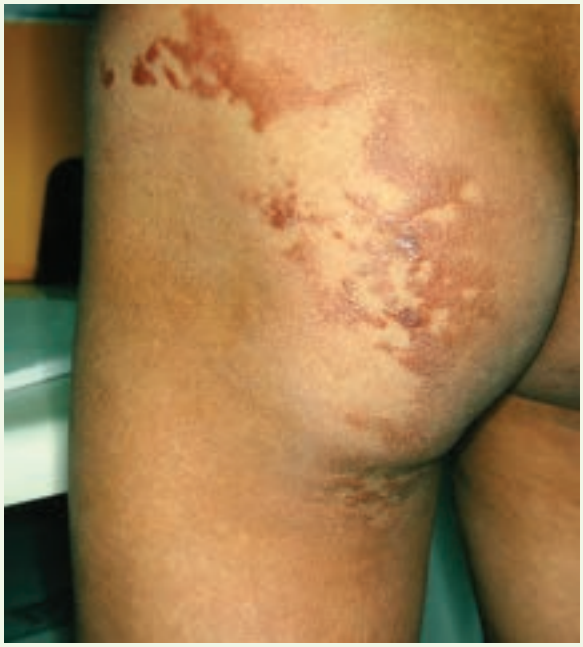
	FIGURE 436-8. Leprosy (Chapter 326).
	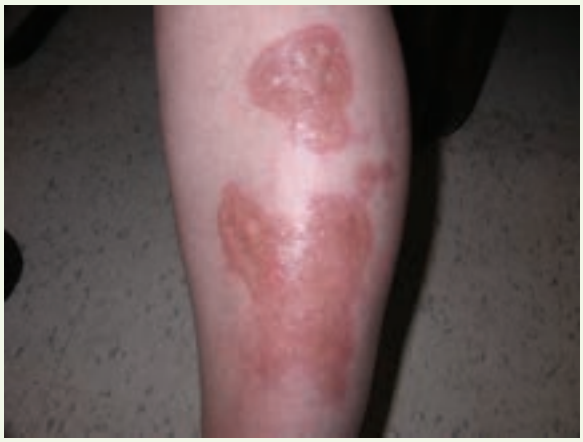
	FIGURE 436-9. Necrobiosis lipoidica (note the characteristic slight yellow tinge; Chapter 440).
Violaceous: Purple Lymphocytes in the dermis impart a violaceous color in lichen planus, lupus erythematosus, lymphoma cutis, and pseudolymphoma.	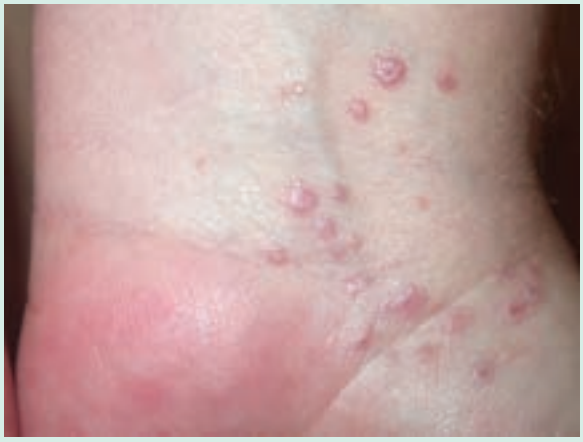
	FIGURE 436-10. Lichen planus (pruritic purple polygonal papules; Chapter 438).

TABLE 436-1 COLOR CLUES IN DIAGNOSING SKIN DISEASE—cont'd

COLOR	CLINICAL EXAMPLE
	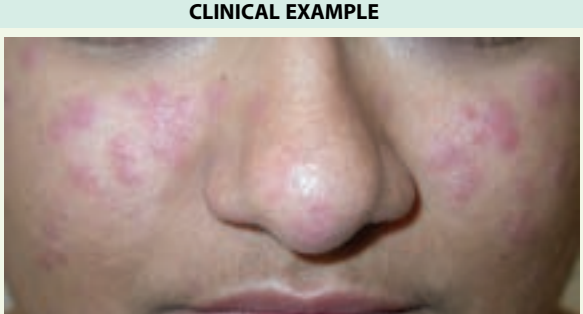 **FIGURE 436-11.** Lupus erythematosus on sun-exposed surface (the "butterfly" rash when more confluent; Chapter 266).
Violaceous: Purple to Black Blood in the dermis (e.g., ecchymoses, vasculitis, subungual hemorrhage) leads to a dark purple color with blue and black overtones. Unlike blood in vessels, lesions do not blanch with pressure.	 **FIGURE 436-12.** Purpura. Small vessel vasculitis with small lesions coalescing into larger plaque (Chapter 439).
	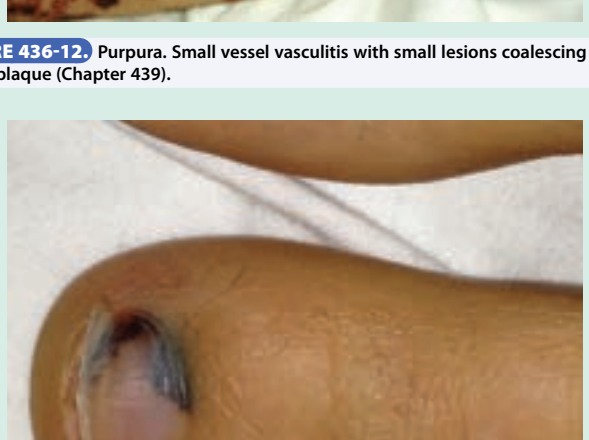 **FIGURE 436-13.** Subungual hemorrhage can mimic melanoma.

TABLE 436-1 COLOR CLUES IN DIAGNOSING SKIN DISEASE—cont'd

COLOR	CLINICAL EXAMPLE
Gray Necrotic keratinocytes (e.g., center of target lesions of erythema multiforme and the border of pyoderma gangrenosum) impart a grayish color. Gray color also is seen with melanin in the dermis (e.g., postinflammatory pigmentary change).	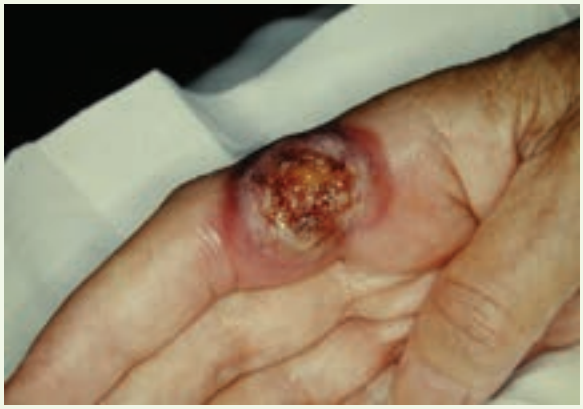 **FIGURE 436-14.** Border of pyoderma gangrenosum (Chapter 440); note gray representing keratinocyte necrosis. 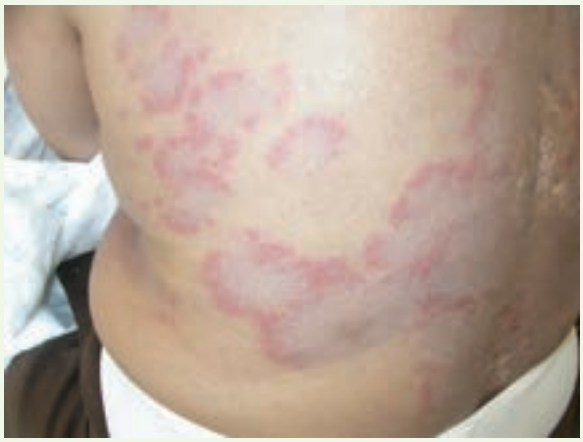 **FIGURE 436-15.** Postinflammatory gray centrally in subacute cutaneous lupus (Chapter 266) in person with olive skin.
Black Black is a sign of dermal and epidermal necrosis (e.g., polyarteritis nodosa, antiphospholipid antibody syndrome, warfarin necrosis, heparin necrosis, calciphylaxis, ANCA-positive vasculitis) from vascular compromise, infection (e.g., mucormycosis, aspergillosis, anthrax, DIC from meningococcemia), or from melanin (the more superficial the melanin, the darker the color).	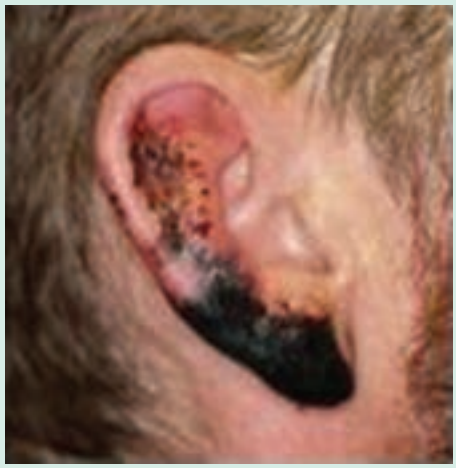 **FIGURE 436-16.** Ischemic necrosis secondary to vasculopathy from levamisole-contaminated cocaine. (From Bolognia JL, Jorizzo JL, Schaffer JV. *Dermatology,* 3rd ed. Philadelphia: Saunders; 2012: Table 0.4.)

TABLE 436-1 COLOR CLUES IN DIAGNOSING SKIN DISEASE—cont'd

COLOR	CLINICAL EXAMPLE
	
	FIGURE 436-17. Melanoma (Chapter 203) demonstrating superficial deposition of melanin in the nail plate.
Brown Brown is a result of pathologic pigments (e.g., hemosiderin; hemochromatosis; drug induced) in the dermis or normal epidermal melanin in individuals of color.	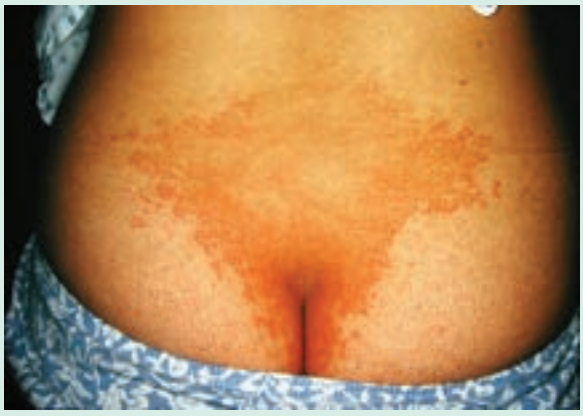
	FIGURE 436-18. Tinea versicolor (Chapter 438); note bran-colored scaling.
	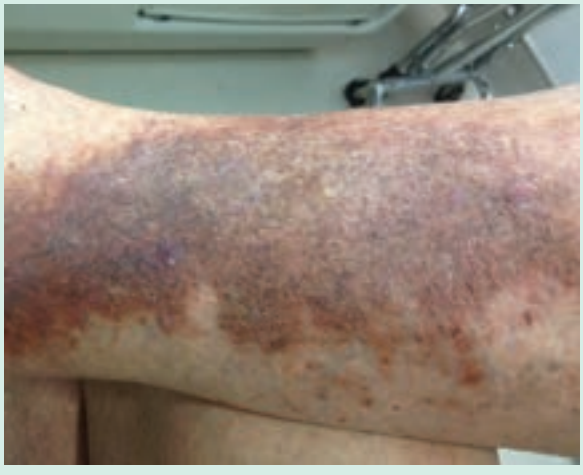
	FIGURE 436-19. Stasis dermatitis (Chapter 440); note brown superiorly from hemosiderin.

TABLE 436-1 COLOR CLUES IN DIAGNOSING SKIN DISEASE—cont'd

COLOR	CLINICAL EXAMPLE
White White can be seen in hyperplastic mucosal epithelium, sclerosis of dermis, scar tissue, loss of pigmentation, or small epidermal cysts.	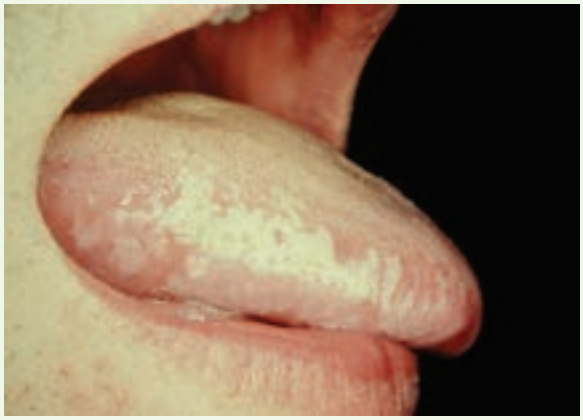

FIGURE 436-20. Oral hairy leukoplakia (Chapter 425).

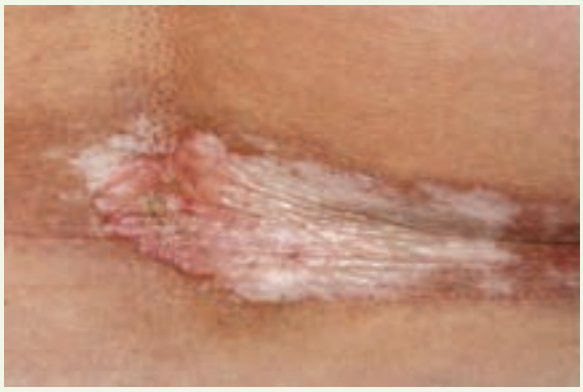

FIGURE 436-21. Lichen sclerosis (Chapter 440) (From Bolognia JL, Jorizzo JL, Schaffer JV. *Dermatology,* 3rd ed. Philadelphia: Saunders; 2012: Fig. 44-12).

ANCA = antineutrophil cytoplasmic antibodies; DIC = disseminated intravascular coagulation.

Purpura refers to the appearance of skin into which red blood cells have extravasated, thereby producing a bluish to dark purple to black color. Purpura is macular in cases of tiny petechiae and ecchymoses from minor trauma. In small vessel leukocytoclastic vasculitis, purpura becomes palpable, usually as small papules. In deeper vasculopathies, purpura can be nodular, linear, stellate or associated with ulceration because of damage to vascular structures in the dermis or subcutaneous fat.

Scaling, which is a manifestation of abnormal production of stratum corneum, is frequently an important clue to the correct diagnosis. It is commonly seen as a primary component of conditions such as psoriasis, tinea, and pityriasis rosea.

Secondary characteristics include crusting, fissures, erosion, ulceration, excoriation, atrophy, and lichenification (Table 436-5). Each may be helpful in diagnosing a skin disorder depending on whether the finding represents a normal evolution of the disease, such as ulcer formation in pyoderma gangrenosum, or an extraneous influence, such as lichenification from scratching in a patient with atopic dermatitis.

Distribution

Some dermatologic diseases can be recognized by where they manifest on the body (Table 436-6). Prime examples include herpes zoster (unilateral dermatomal distribution of vesicles), dermatitis herpetiformis (pruritic vesicles over extensor areas of elbows, knees, and sacral skin), and hidradenitis suppurativa (inflammatory acne-like nodules and cysts in the axillae, inguinal, and intergluteal areas). Other sites manifestation include sun-exposed skin for lupus erythematosus and dermatomyositis, facial distribution for acne vulgaris and rosacea, and penile involvement for psoriasis and lichen planus.

Sun-exacerbated diseases spare areas under the nose and chin, inner arms, intertriginous regions, the mid and distal phalanges, and clothed areas. In a patient with bilateral erythema and swelling of the legs with dermatitis, the odds of cellulitis are extremely low, because cellulitis of the extremities is a unilateral disease.

Life-Threatening Emergencies

In the acutely ill patient with a dermatosis, a clinician must determine whether the patient has a potentially rapidly fatal skin disease—typically a systemic infection or severe drug reaction—and if so, what treatment to initiate urgently. The most common rapidly fatal diseases with skin involvement are disseminated herpes zoster infection[3] (mortality up to 30%, Chapter 375); disseminated herpes simplex infection (mortality up to 80% in the presence of encephalitis, Chapter 374); toxic epidermal necrolysis[4] or its variant Stevens-Johnson syndrome (mortality up to 20% for toxic epidermal necrolysis Chapter 440); meningococcemia (mortality up to 35%, Chapter 298); toxic shock syndrome (mortality up to 30%, Chapters 288 and 290); necrotizing fasciitis (mortality 20 to 40%, Chapters 288 and 290); and disseminated fungal diseases including candidiasis (Chapter 338), histoplasmosis (Chapter 332), cryptococcosis (Chapter 336).

In the acutely ill patient, physical findings that can often help make the correct diagnosis include the following:

- Multiple small, discrete lesions (10 to several hundred; <1 cm diameter) scattered over the body that do not become confluent into large sheets imply a hematogenously spread infectious disease, such as seen in

Text continued on p. 2654

TABLE 436-2 REPRESENTATIVE DIAGNOSES BASED ON DEPTH OF PATHOLOGY IN THE SKIN

DEPTH	CHARACTERISTICS	DISORDERS		CLINICAL EXAMPLE
Epidermal	Distinct, even sharp margins Scaling Epidermal thickening Absence of induration May have mild erythema as well (some dermal involvement)	Seborrheic keratosis Warts In situ squamous cell carcinoma (Bowen's disease) Ichthyosis Bullous impetigo Superficial basal cell carcinoma Acanthosis nigricans	Seborrheic dermatitis Thin-plaque psoriasis Mild contact dermatitis Pityriasis rosea Tinea versicolor Tinea corporis Actinic keratosis Patch stage mycosis fungoides	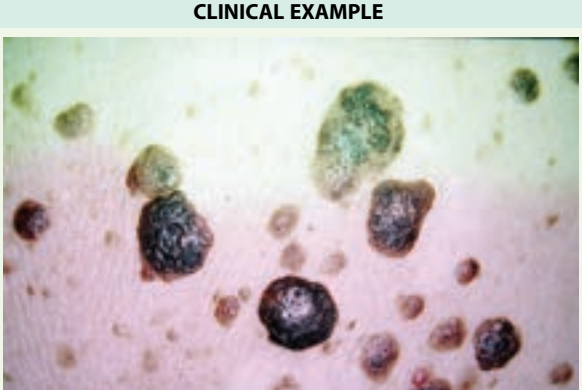 **FIGURE 436-22.** Seborrheic keratosis (Chapter 440).
Epidermal and dermal Most skin diseases will have some degree of dermal and epidermal involvement	Margins moderately well defined Scale can be present Inflammation (erythema) Usually palpable lesions (raised)	Lichen planus Systemic lupus erythematosus Chronic cutaneous lupus Nummular (discoid) dermatitis Plaque-stage mycosis fungoides Pyoderma gangrenosum	Atopic dermatitis Small vessel vasculitis Secondary syphilis Cellulitis/erysipelas Nodular basal cell carcinoma Squamous cell carcinoma Melanoma Most vesiculobullous diseases (HSV, autoimmune)	 **FIGURE 436-23.** Lichen planus (Chapter 438).
Dermal Diseases with only dermal involvement usually consist of infiltrates with inflammatory or malignant cells that do not influence epidermal function	Margins moderately defined No scale Smooth surface No ulcer Variable inflammation Usually palpable	Urticaria Sarcoidosis Granuloma annulare Leprosy Necrobiosis lipoidica Morphea Scleroderma Plaque- or tumor-stage mycosis fungoides	Cutaneous metastases Most cysts Dermatofibromas Dermal melanocytic nevi Pretibial myxedema Other cutaneous lymphomas	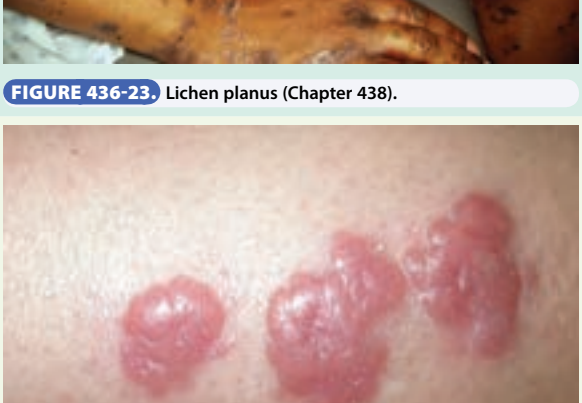 **FIGURE 436-24.** Sarcoidosis: Histiocyte collections in the upper dermis (Chapter 95).

TABLE 436-2 REPRESENTATIVE DIAGNOSES BASED ON DEPTH OF PATHOLOGY IN THE SKIN—cont'd

DEPTH	CHARACTERISTICS	DISORDERS	CLINICAL EXAMPLE
Subcutaneous fat	Margins rounded and/or poorly defined Variable inflammation Smooth overlying skin	Large cysts, lipoma Erythema nodosum Panniculitis of any etiology Some lymphomas Medium-vessel vasculitis (polyarteritis nodosa)	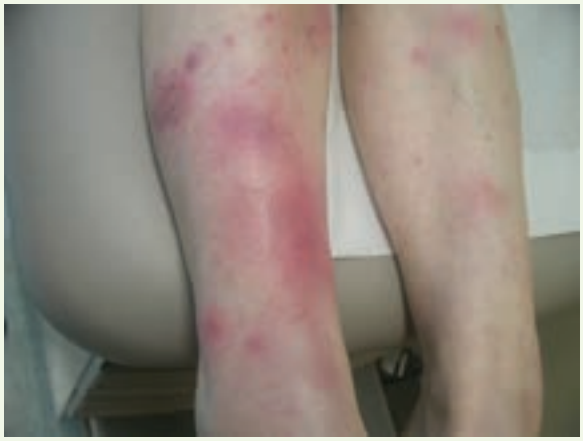

FIGURE 436-25. Erythema nodosum (Chapter 440).

HSV = herpes simplex virus.

TABLE 436-3 IMPORTANT CLUES FROM THE HISTORY AND GENERAL HEALTH STATUS

CLUES	POSSIBLE DIAGNOSES	
Pruritus (itch)	Atopic dermatitis (eczema) Allergic contact dermatitis Urticaria Scabies	Bullous pemphigoid Lichen planus Dermatitis herpetiformis Inflammatory tinea pedis
Absence of pruritus	Acne vulgaris Rosacea Syphilis	Skin malignancies Lupus Pemphigus vulgaris Erythema multiforme
Pain	Herpes zoster Ischemic necrosis from all causes Cellulitis Furunculosis	Carbuncles Pyoderma gangrenosum Severe systemic illnesses
Medications (especially new)	Severe drug eruptions Fixed drug eruptions Morbilliform drug eruptions Immunosuppression	Subacute cutaneous lupus Psoriasis exacerbation Urticaria Drug induced hair loss
Cachexia/malnourished	Paraneoplastic dermatoses Skin malignancy Alcohol exacerbated psoriasis	Nutritional deficiencies Metastatic disease Eating disorders leading to nutritional deficiencies
Obesity or weight gain	Dermatoses of diabetes Thyroid disease Cushing's disease Polycystic ovarian syndrome Acne	Hirsutism Striae Acanthosis nigricans Eruptive xanthomas
Poor hygiene	Bacterial skin infections Infestations	Substance abuse
Psychiatric illness	Self-inflicted skin disease	Drug-induced dermatoses

TABLE 436-4 MORPHOLOGIC TERMS FOR PRIMARY LESIONS

TERMS	ILLUSTRATIONS
Macule: Flat, nonpalpable, <1 cm diameter	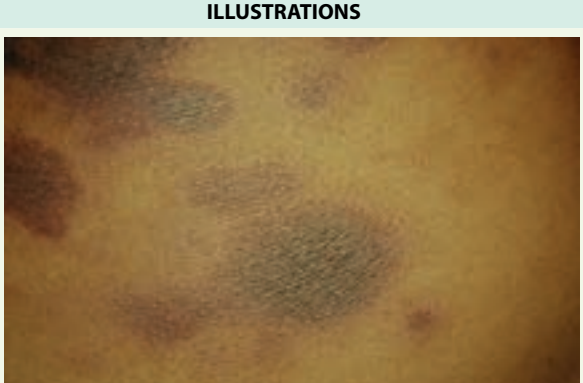
	FIGURE 436-26. Fixed drug eruption (Chapter 440).
	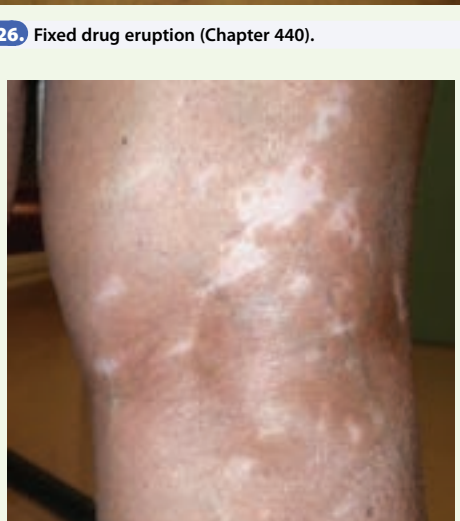
	FIGURE 436-27. Vitiligo (Chapter 441).
Patch: Large macules >1 cm diameter	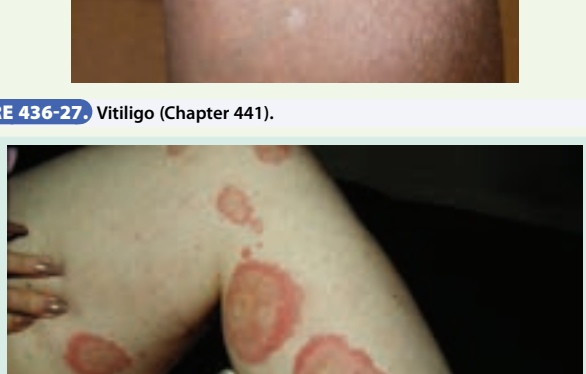
	FIGURE 436-28. Erythema multiforme (Chapter 439).

TABLE 436-4 MORPHOLOGIC TERMS FOR PRIMARY LESIONS—cont'd

TERMS	ILLUSTRATIONS
	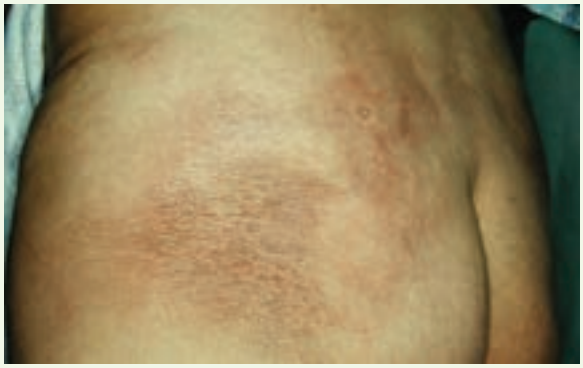
	FIGURE 436-29. Mycosis fungoides (cutaneous T-cell lymphoma [CTCL]; Chapter 185).
Papule: Superficial, raised, palpable, <1 cm diameter	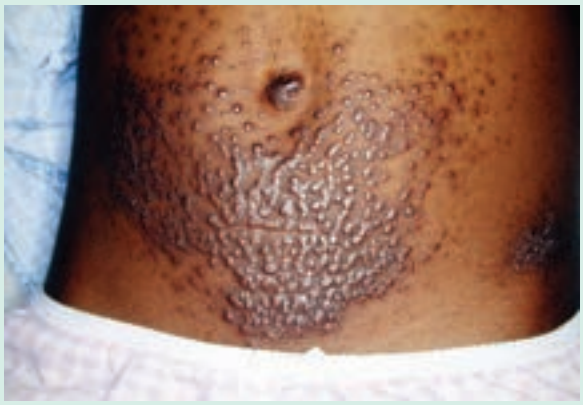
	FIGURE 436-30. Atopic dermatitis with follicular accentuation (Chapter 438).
Plaque: Raised, palpable, >1 cm diameter	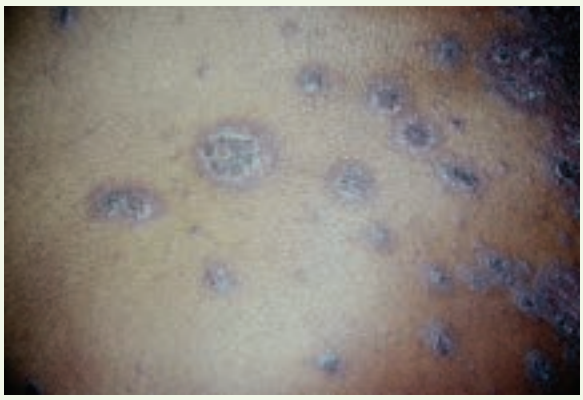
	FIGURE 436-31. Psoriasis (Chapter 438).
	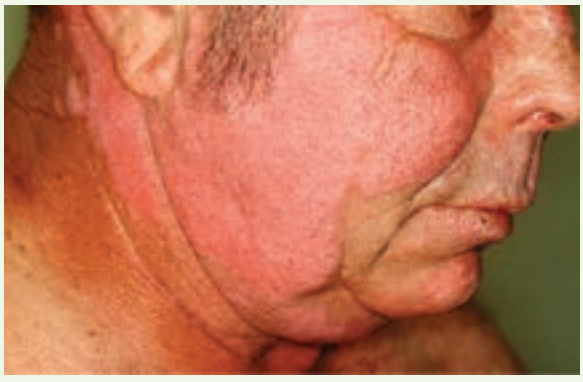
	FIGURE 436-32. Discoid lupus (Chapter 266).

TABLE 436-4 MORPHOLOGIC TERMS FOR PRIMARY LESIONS—cont'd

TERMS	ILLUSTRATIONS
Nodule: Deeper than a papule, usually <1 cm diameter	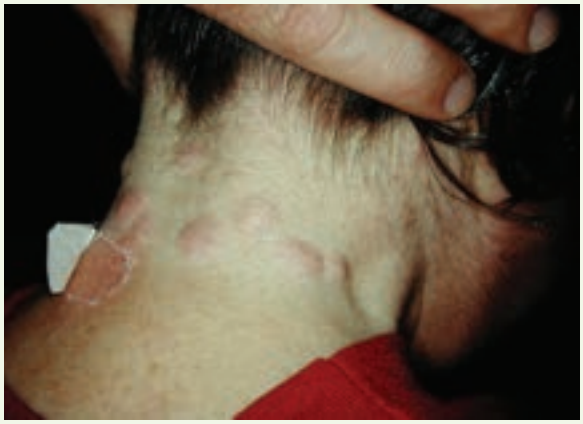
	FIGURE 436-33. Metastatic breast carcinoma (Chapter 198).
Tumor: Large nodule, >1 cm diameter	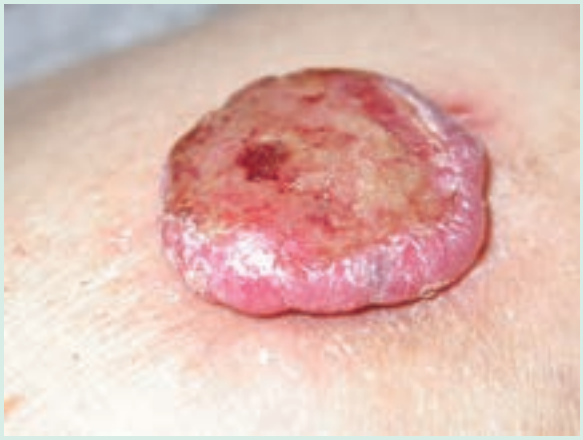
	FIGURE 436-34. Amelanotic melanoma (Chapter 203).
	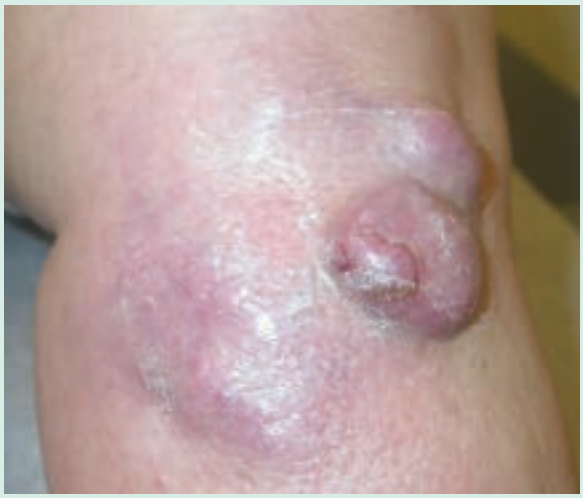
	FIGURE 436-35. Post-transplant lymphoproliferative disorder (Chapter 185).

TABLE 436-4 MORPHOLOGIC TERMS FOR PRIMARY LESIONS—cont'd

TERMS	ILLUSTRATIONS
Vesicle: <1cm fluid-filled, may be umbilicated or contain pus or blood	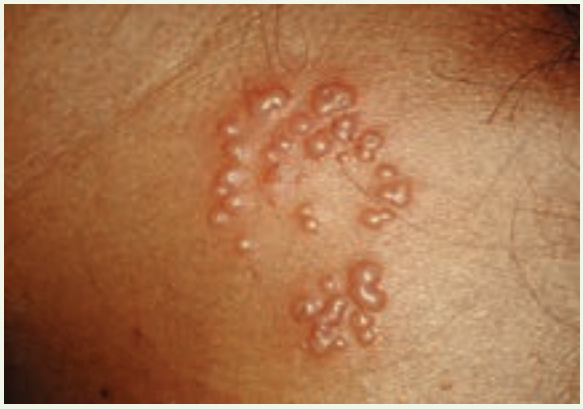
	FIGURE 436-36. Herpes simplex (Chapter 374).
Bulla: >1-2 cm, filled with clear fluid, pus, or blood	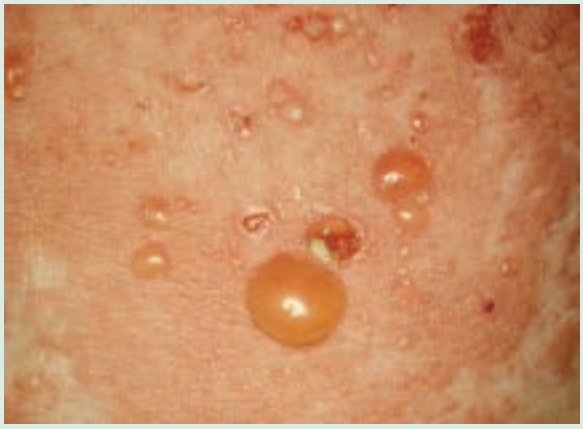
	FIGURE 436-37. Bullous pemphigoid with small and larger bullae (Chapter 439).
Pustule: White, pus-filled, small, raised, <1 cm diameter	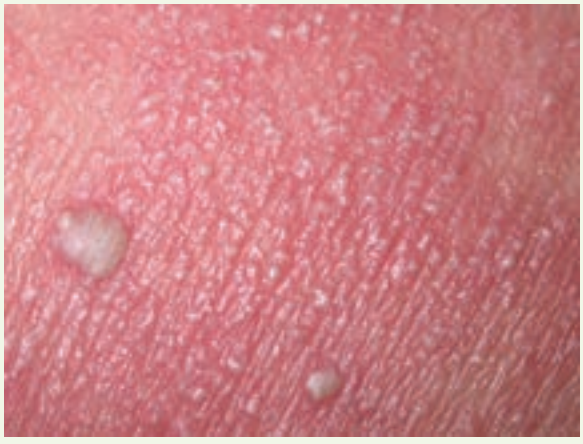
	FIGURE 436-38. Acute generalized exanthematous pustulosis (Chapter 439).

TABLE 436-4 MORPHOLOGIC TERMS FOR PRIMARY LESIONS—cont'd

TERMS	ILLUSTRATIONS
Scale: White keratinous flakes of retained stratum corneum	

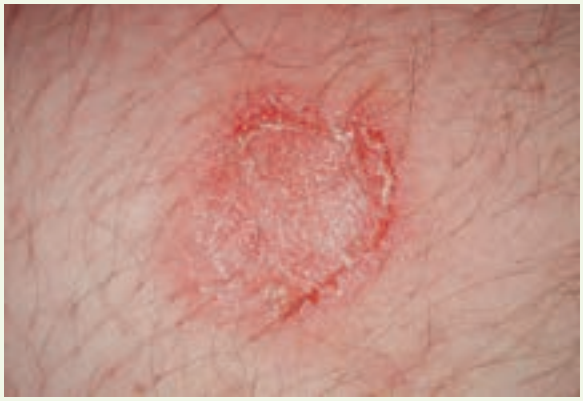

FIGURE 436-39. Nummular dermatitis (Chapter 438).

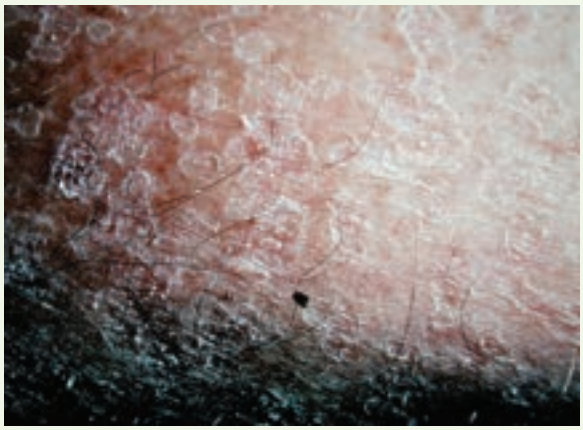

FIGURE 436-40. Disseminated superficial actinic porokeratosis (Chapter 440).

TABLE 436-5 MORPHOLOGIC TERMS FOR SECONDARY FEATURES

Crust: Dried serum, yellow to hemorrhagic depending on number of cells in exudate

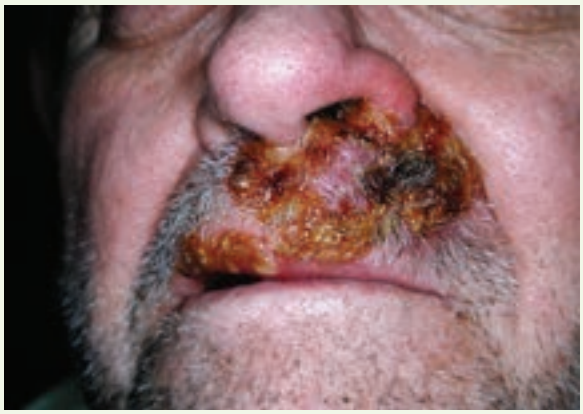

FIGURE 436-41. Impetigo (Chapter 441).

TABLE 436-5 MORPHOLOGIC TERMS FOR SECONDARY FEATURES—cont'd

Fissure: Small linear splits in skin, usually superficial

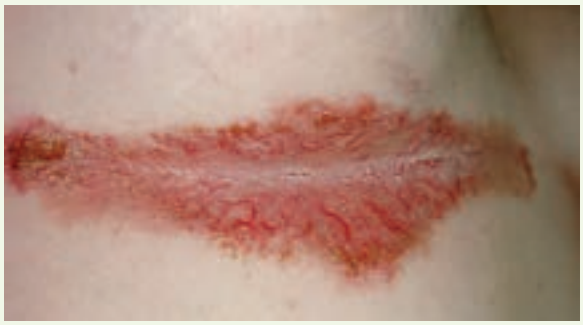

FIGURE 436-42. Hailey-Hailey disease.

Erosion: Loss of part or all of epidermis into upper dermis at most

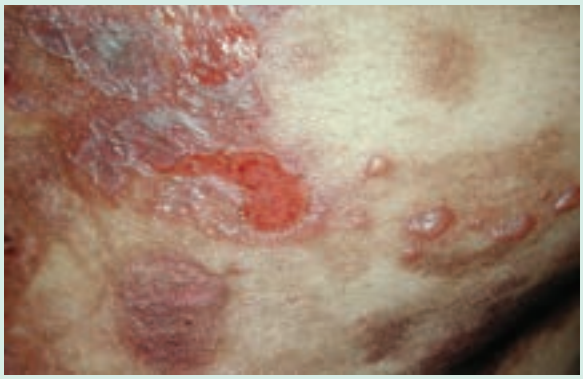

FIGURE 436-43. Pemphigus vulgaris (Chapter 439).

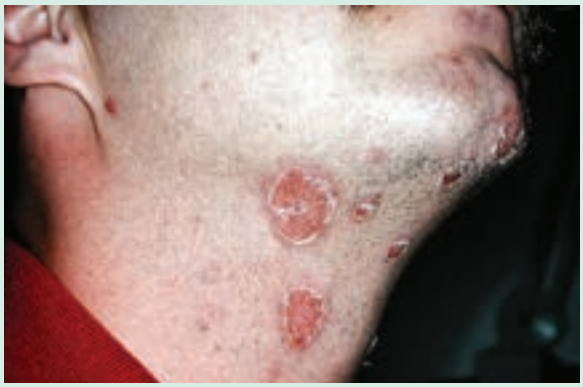

FIGURE 436-44. Staphylococcal bullous impetigo (Chapter 441).

Ulceration: Loss of epidermis into deep dermis or fat

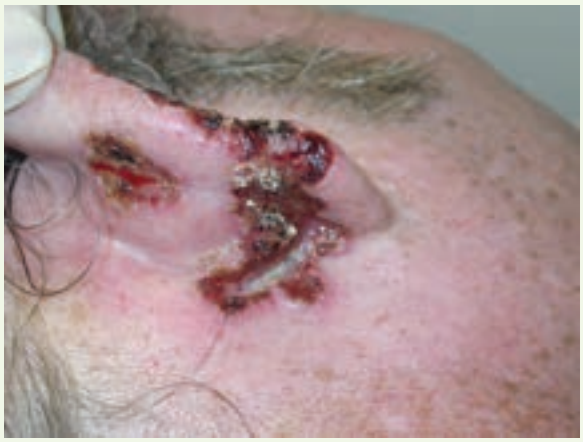

FIGURE 436-45. Chronic erosive herpes simplex (Chapter 374).

TABLE 436-5 MORPHOLOGIC TERMS FOR SECONDARY FEATURES—cont'd

Excoriation: Self-inflicted erosion or ulcer

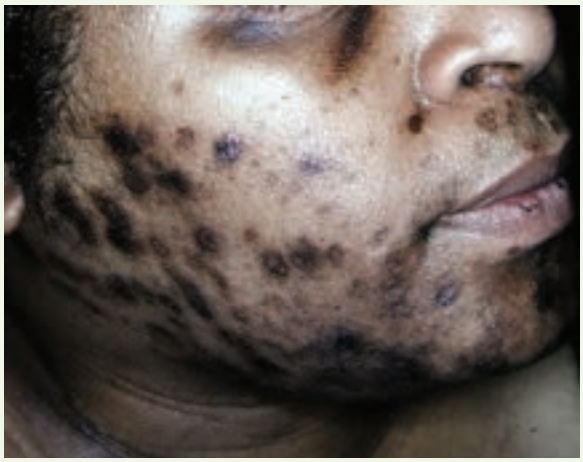

FIGURE 436-46. Factitial dermatitis (dermatitis artefacta).

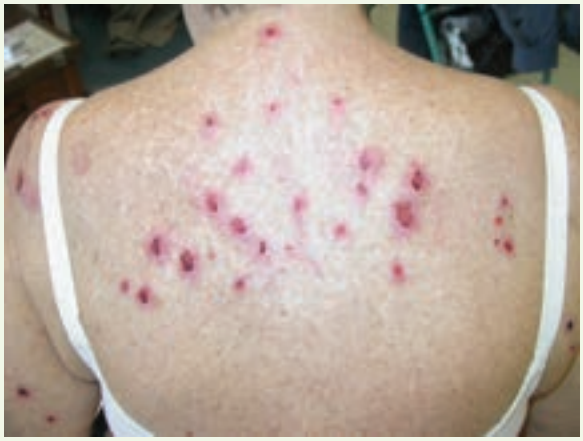

FIGURE 436-47. Neurodermatitis; note inferior cutoff where patient cannot reach.

Atrophy: Thinning or loss of dermal structures or fat usually; may apply to any tissue

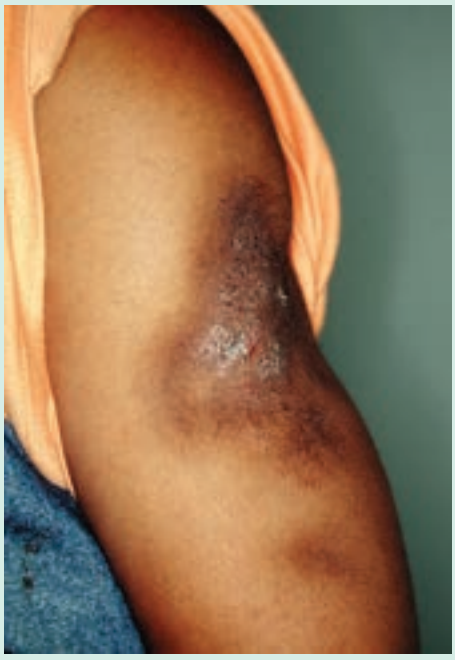

FIGURE 436-48. Fat atrophy: Lupus panniculitis.

TABLE 436-5 MORPHOLOGIC TERMS FOR SECONDARY FEATURES—cont'd

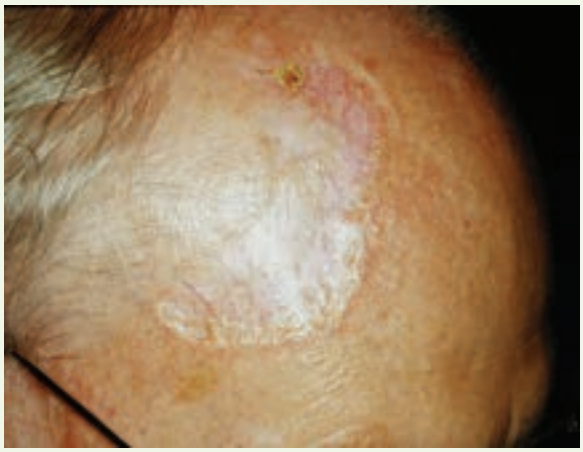

FIGURE 436-49. Central atrophy: Annular elastotic giant cell granuloma.

Lichenification: Thickened skin from chronic rubbing. Poorly demarcated, skin lines accentuated.

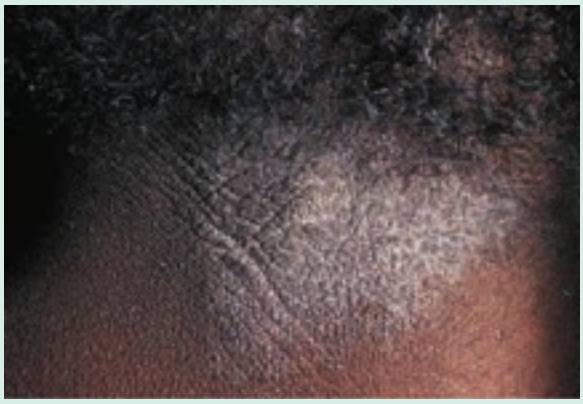

FIGURE 436-50. Lichen simplex chronicus.

TABLE 436-6 LOCATION OF SKIN LESIONS AS A DIAGNOSTIC CLUE

LOCATION	POSSIBLE DIAGNOSES	
Unilateral	Herpes zoster Herpes simplex Deep vein thrombosis Embolic disease	Contact dermatitis where exposed Cellulitis Necrotizing fasciitis Solitary skin malignancies
Bilateral, widespread	Hematogenously spread process Severe drug eruption Disseminated viral, bacterial, or fungal infection Vasculitis Autoimmune disease	Psoriasis Cutaneous T-cell lymphoma Scabies Urticaria Atopic dermatitis (eczema) Other drug eruptions
Bilateral, limited sites	Stasis dermatitis Contact dermatitis	Vasculitis Erythema multiforme

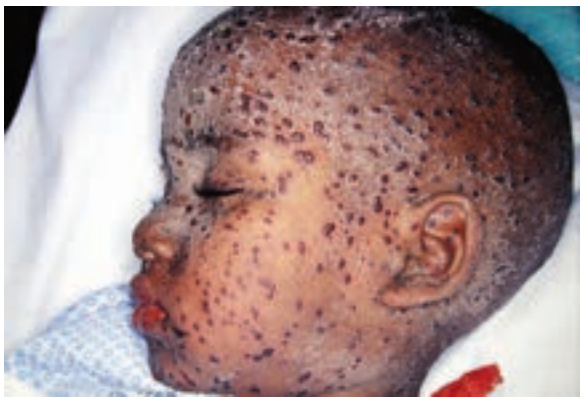

FIGURE 436-51. Chickenpox (Chapter 375). Multiple small vesicles that do not coalesce into larger bullae. Several lesions demonstrate central umbilication.

chickenpox (Fig. 436-51), disseminated herpes zoster (herpes simplex virus) (Fig. 436-52), disseminated fungal disease (Fig. 436-53), or a vascular immune process, such as autoimmune vasculitis (Fig. 436-54).

- Confluence of individual red lesions into sheets of erythema, especially in the presence of skin detachment or fluid-filled bullae, suggests a severe drug eruption such as toxic epidermal necrolysis or Stevens-Johnson syndrome (Fig. 436-55; Chapter 440).
- Mucositis, which is inflammation and erosions of conjunctivae or of the oral or anogenital mucosae, in the acutely ill patient also strongly suggests toxic epidermal necrolysis or Stevens-Johnson syndrome (Chapter 440). Although some viral infections, such as herpesvirus and coxsackievirus, can cause multiple discrete mucosal lesions, severe mucositis is rarely a presenting finding in patients with disseminated viral, fungal, or bacterial infections.
- Widespread vesicles (Fig. 436-56) suggest disseminated herpesvirus infection, either varicella (Chapter 375) or simplex (Chapter 374). The vesicles often become hemorrhagic or pustular 2 to 5 days later when red blood cells or neutrophils infiltrate them.
- Widespread bullae (Fig. 436-57) with and without skin detachment suggest severe drug eruptions, as in toxic epidermal necrolysis and the

Stevens-Johnson syndrome (Chapter 440). By comparison, bullae accompanied by intense pruritus suggest autoimmune bullous pemphigoid (Fig. 436-58; Chapter 439) or severe allergic contact dermatitis (Fig. 436-59; Chapter 438).

- Widespread purpura (see Fig. 436-12) in an acutely ill patient may be a sign of disseminated intravascular coagulation (Chapter 174), which is seen in severe systemic infections, such as meningococcemia (Chapter 298). In less urgent settings, purpura can be caused by localized coagulation disorders (reactions to warfarin or heparin) or can be palpable and a sign of small vessel leukocytoclastic vasculitis[5] (Chapter 270). Purpura above the waist or on mucosal surfaces strongly implies systemic vasculitis or coagulation disorder, whereas purpura below the waist, especially below the knees, can be caused by the extravasation of red blood cells owing to increased hydrostatic pressure in association with any cause of tissue inflammation.
- Acute erythema of the hands, feet, and pelvic groin region (Fig. 436-60) associated with systemic illness can suggest a bacterial exotoxin as seen in toxic shock syndrome (Chapters 288 and 290). Peeling of palmar and plantar skin is not seen until 2 weeks later, so early diagnosis and treatment during the erythematous phase is critical. By comparison, erythroderma, which is confluent erythema without individual lesions, usually is not immediately life-threatening and can be seen in severe forms of psoriasis (Chapter 438), cutaneous T-cell lymphoma (mycosis fungoides, Chapter 185), and erythrodermic drug eruptions (Chapter 440).

Diagnostic Tests

Diascopy is merely pressing carefully on the skin with something transparent to look at the vasculature. Options include a clear plastic diascope, a glass slide, or any other lens. Vascular malformations and small hemangiomas can be diagnosed this way, and careful examination may reveal pulsations in spider telangiectasias.

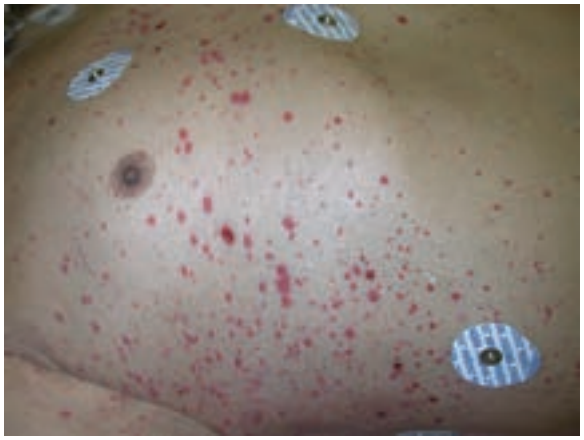

FIGURE 436-52. Disseminated varicella-zoster virus. Widespread, individual papules and vesicles, many with hemorrhage. Note the absence of coalescence into larger bullae.

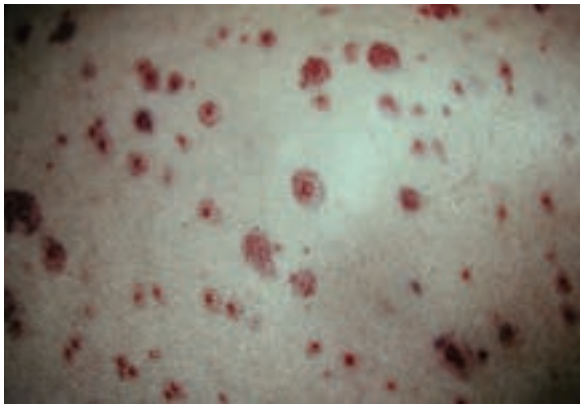

FIGURE 436-53. Disseminated candidiasis. Multiple individual small lesions, papules mostly, some with surrounding purpura, consistent with hematogenously spread disease.

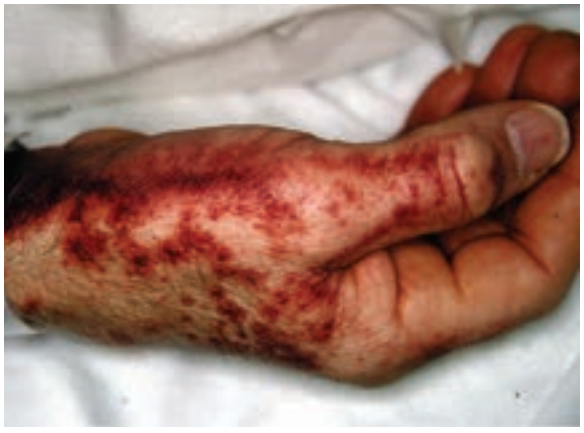

FIGURE 436-54. Autoimmune vasculitis. Purpura with individual small petechiae and coalescence into linear purpura.

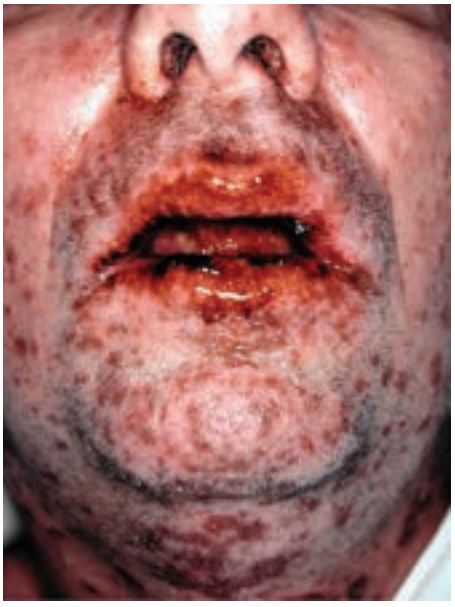

FIGURE 436-55. Mucositis (Stevens-Johnson syndrome). Extensive erosive changes on lips and oral mucosa are highly suggestive of a severe drug eruption such as Stevens-Johnson syndrome or toxic epidermal necrolysis.

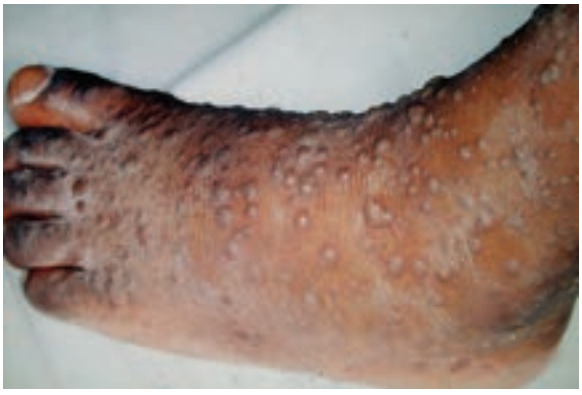

FIGURE 436-56. Varicella-zoster virus. Multiple vesicles characteristic of severe varicella (chickenpox).

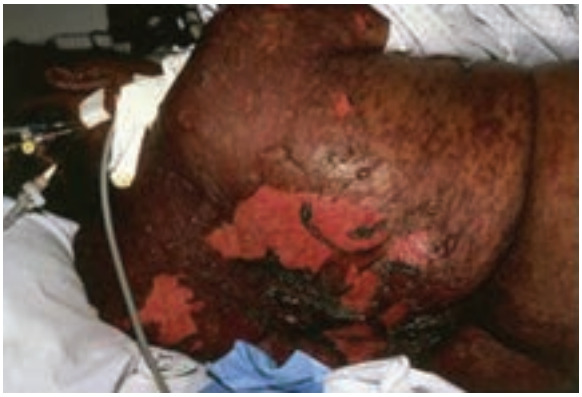

FIGURE 436-57. Bullae (toxic epidermal necrolysis). Detachment of large sheets of necrolytic epidermis (>30% body, surface area), leading to extensive areas of denuded skin. A few intact bullae are still present. (From Bolognia JL, Jorizzo JL, Schaffer JV. *Dermatology,* 3rd ed. Philadelphia: Saunders; 2012:328, Fig. 20.10A.)

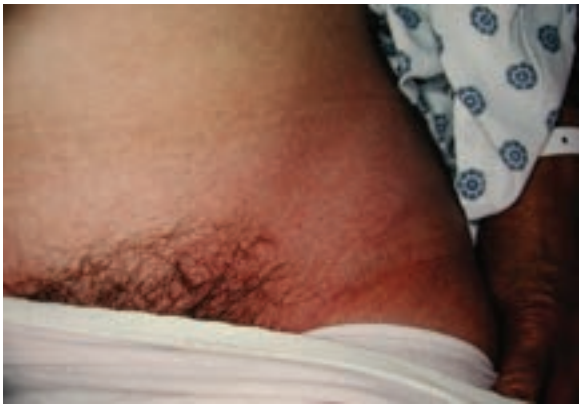

FIGURE 436-60. Toxic shock syndrome. Acute erythema in a swim-trunk distribution can be an early sign of toxic shock syndrome. Similar erythema occurs on dorsal hands and feet.

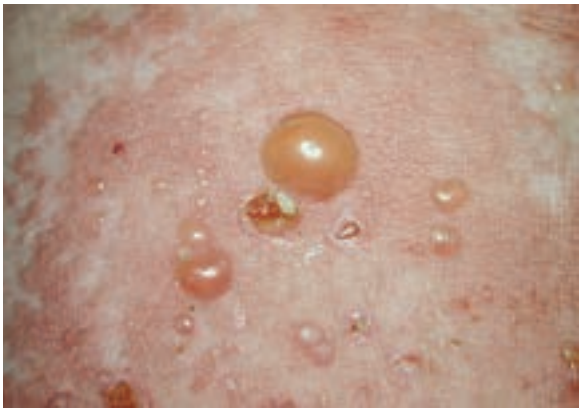

FIGURE 436-58. Vesicles and bullae (bullous pemphigoid). Large bullae as well as smaller bullae and vesicles on a background of urticarial-like erythema.

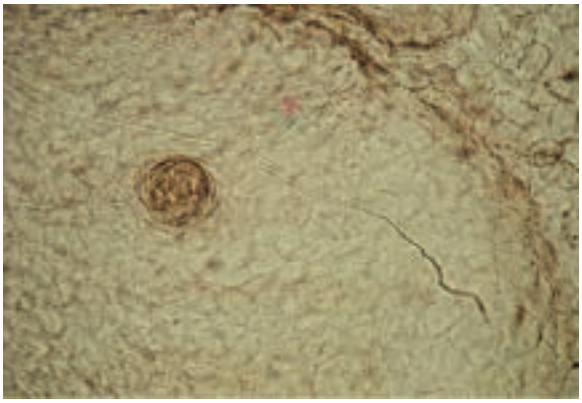

FIGURE 436-61. Potassium hydroxide examination at 40× demonstrating linear fungal hyphae.

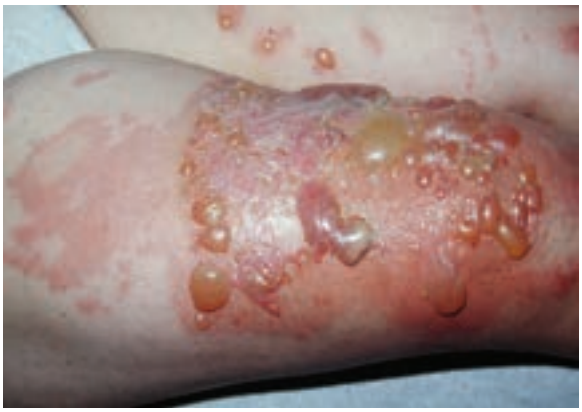

FIGURE 436-59. Severe allergic contact dermatitis. Bullae and vesicles on a background of erythema. Severe cases such as this can be mistaken for the autoimmune disease bullous pemphigoid.

Dermoscopy, also known as dermatoscopy, uses a polarized magnification light source to reveal otherwise nonvisible features of skin lesions. Dermoscopy requires specialized training not usually available to the nondermatologist.

The scaly portion of a lesion can be scraped with a scalpel blade, held perpendicular to the skin to avoid lacerations, to obtain a specimen to be analyzed using 10% solution potassium hydroxide (KOH), which dissolves keratin, thereby exposing microscopic nonkeratinized structures (primarily fungal hyphae). The scales should be placed on a glass slide, and the KOH solution should then be applied and covered with a coverslip. The slide is exposed briefly to flame heat (a Bunsen or alcohol burner are best) while applying gentle pressure. The specimen is then examined under low and medium power (Fig. 436-61). Finding and interpreting fungal elements requires some training. The same approach is also highly effective for identifying scabies mites (Fig. 436-62) and eggs, although mineral oil can be used instead of KOH.

The Tzanck smear, which is a time-honored test for herpes zoster and herpes simplex, does not differentiate between the two. Other methods for detecting herpesvirus include viral culture (best for herpes simplex), direct fluorescent antibody staining, polymerase chain reaction testing, and electron microscopy. The preferred method usually depends on local availability.

For the Tzanck smear, any nuclear stain (methylene blue, alanine, hematoxylin, etc.) can be used to identify the pathognomonic multinucleated keratinocytes, often called giant cells (Fig. 436-63). To obtain the specimen, the clinician should unroof and scrape the base of a vesicle, fix the smeared material to the slide with mild heat, place several drops of stain, rinse with water after a few seconds, lightly blot the slide dry, and examine it under medium power. Immersion oil is necessary for good optics, with or without coverslip. To perform a Gram stain to look for bacteria or fungal forms, the clinician should fix the material onto the slide, apply each stain (crystal violet, iodine, alcohol, safranin) in order for a few seconds, rinse, gently blot dry, mount under immersion or mineral oil, and examine under high power.

A skin biopsy can add valuable information but does not always provide a definitive diagnosis. The vast majority of skin lesions should be biopsied at the site of the most obvious pathologic process, not at the periphery. The biopsy should be performed on the periphery only of a lesion in the setting of an ulcer, where the center may show only nondiagnostic granulation

FIGURE 436-62. A scabies mite is seen on this microscopic examination of an oil mount of a scraping taken from the end of a small burrow on the wrist.

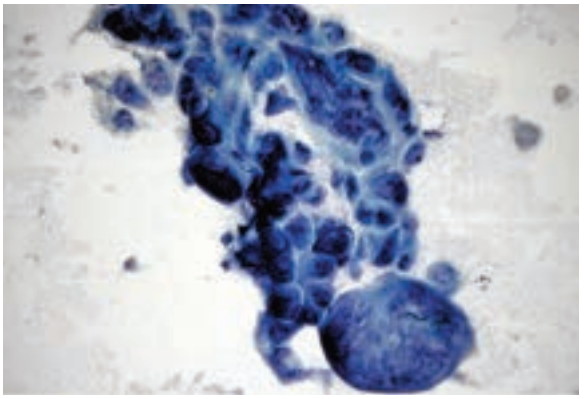

FIGURE 436-63. Multinucleated keratinocytes in a Tzanck preparation for herpes simplex, zoster, or varicella. Sample taken from the base of a vesicle (40×).

tissue[6]; for bullae, in which the edge at the junction of a bulla and normal skin is preferable; and when obtaining perilesional normal skin for immunofluorescence studies. Otherwise, there is no need to have normal skin in a biopsy specimen.

A skin biopsy is most commonly performed as a punch biopsy, which cuts a circular piece of skin or drills a core of skin down to subcutaneous fat. Biopsy specimens range from 2 to 8 mm in diameter. Infiltration of the soft tissue with lidocaine plus epinephrine 5 to 10 minutes before the procedure can minimize bleeding. For biopsies of 4 mm and greater, sutures control bleeding and speed healing.

When the diagnosis is probable melanoma, excision, not punch biopsy, should be performed by a dermatologist or surgeon. If the depth of involvement is not essential for diagnosis or treatment decisions, shave biopsies are useful for removing small raised lesions at the skin surface. The snip excision/biopsy can be particularly useful on the flaccid skin of eyelids or genitalia.

WHEN TO REFER

Most physicians know when a medical problem goes beyond their level of expertise or knowledge. Because cutaneous manifestations of life-threatening disease must be diagnosed quickly, an urgent consultation is appropriate for inpatients who develop new skin lesions or outpatients with severe underlying illnesses. Early referral also is indicated when a malignancy, especially melanoma, is suspected.

Grade A Reference

A1. Arakaki RY, Strazzula L, Woo E, et al. The impact of dermatology consultation on diagnostic accuracy and antibiotic use among patients with suspected cellulitis seen at outpatient internal medicine offices: a randomized clinical trial. *JAMA Dermatol.* 2014;150:1056-1061.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Storan ER, McEvoy MT, Wetter DA, et al. Experience of a year of adult hospital dermatology consultations. *Int J Dermatol*. 2014;[Epub ahead of print].
2. Haenssle HA, Korpas B, Hansen-Hagge C, et al. Seven-point checklist for dermatoscopy: performance during 10 years of prospective surveillance of patients at increased melanoma risk. *J Am Acad Dermatol*. 2010;62:785-793.
3. Rommelaere M, Maréchal C, Yombi JC, et al. disseminated varicella zoster virus infection in adult renal transplant recipients: outcome and risk factors. *Transplant Proc*. 2012;44:2814-2817.
4. Schwartz RA, McDonough PH, Lee BW. Toxic epidermal necrolysis: Part I. Introduction, history, classification, clinical features, systemic manifestations, etiology, and immunopathogenesis. *J Am Acad Dermatol*. 2013;69:173, e171-e113.
5. Ricketts JR, Rothe MJ, Grant-Kels JM. Cutaneous simulants of infectious disease. *Int J Dermatol*. 2011;50:1043-1057.
6. Senet P, Combemale P, Debure C, et al. Malignancy and chronic leg ulcers: the value of systematic wound biopsies: a prospective, multicenter, cross-sectional study. *Arch Dermatol*. 2012;148:704-708.

REVIEW QUESTIONS

1. A 48-year-old man comes to the emergency room complaining of 2 days of excruciating stabbing pain and tenderness behind his right ear. There is no history of trauma. The pain keeps him up at night. He is healthy otherwise. Your examination reveals exquisite tenderness on the occipital scalp behind his right ear and extending to parietal scalp, but there are no other cutaneous findings. The most likely diagnosis is:

A. Erysipelas.
B. Herpes simplex.
C. Herpes zoster.
D. Malingering.
E. Meningitis.

Answer: C Pain involving the skin or scalp eliminates intracranial disease. Herpes zoster classically causes pain along part or all of a dermatome several days before lesions become apparent on the skin. Malingering is always a consideration when no physical findings are present, but careful examination usually can delineate the margins of tenderness accurately.

2. A 52-year-old man is transferred to your hospital with fever, severe malaise, and sharp right upper quadrant pain. For the last 4 years, he has been on low-dose methotrexate for rheumatoid arthritis. He also has taken occasional nonsteroidal anti-inflammatory drugs for pain. Imaging at a community hospital revealed no anatomic liver pathology, but blood work showed elevated liver enzymes. On examination he has ill-defined right upper quadrant tenderness with no obvious hepatomegaly and a nonsurgical abdomen. On examination of his skin, you notice approximately 50 individual raised vesicles, each the size of a pea, on his trunk and extremities. Some lesions are crusted, some appear hemorrhagic, some contain clear fluid, and some have an umbilicated appearance. Mucosal surfaces are normal.

3. In the patient in question 2, what diagnosis is *least* likely?

A. Disseminated candidiasis
B. Disseminated varicella-zoster virus
C. Rocky Mountain spotted fever
D. Erythema multiforme
E. Toxic epidermal necrolysis

Answer: E Toxic epidermal necrolysis (TEN) could manifest with symptoms of fever and malaise, but the morphology of multiple individual discrete lesions and the absence of mucosal involvement make TEN highly unlikely.

4. In the patient in question 2, what diagnosis is *most* likely?

A. Disseminated candidiasis
B. Disseminated varicella-zoster virus
C. Rocky Mountain spotted fever
D. Erythema multiforme
E. Toxic epidermal necrolysis

Answer: B Disseminated varicella-zoster virus is most likely. The multiple widely spread lesions, some of which are umbilicated, favor a disseminated viral infection. The right upper quadrant pain represents dermatomal pain from herpes-zoster viral reactivation, but without dermatomal skin findings.

5. A 25-year-old woman comes to your office complaining of 2 days of sore throat, as well as severe redness and soreness of her lips, the inside of her mouth, her eyes, and her anogenital region. Painful sores on her palms and soles developed in the last 24 hours. She denies fever but complains of coughing. She attributes the skin problem to her recent urinary tract infection, for which she was given trimethoprim-sulfamethoxazole 7 days earlier. Your examination shows confluent erosions on the oral, anogenital, and conjunctival mucosae. Palm lesions consist of numerous tender erythematous macules. What is the *most* threatening diagnosis to consider and treat immediately?

A. Kawasaki's disease
B. Toxic shock syndrome
C. Toxic epidermal necrolysis
D. Staphylococcal scalded skin syndrome
E. Glucagonoma syndrome

Answer: C Erosive mucositis is the key morphology here. In patients with acute erosive mucositis, especially at multiple sites, the first diagnosis to consider is toxic epidermal necrolysis or its slightly less severe variant, Stevens-Johnson syndrome. In this case, in addition to mucositis, the drug history points to a possible severe drug eruption. Glucagonoma syndrome can manifest with mucosal erosions, but they are not acute. Kawasaki's disease and toxic shock syndrome manifest with erythema of the mucosae, but they rarely become erosive.

437

PRINCIPLES OF THERAPY OF SKIN DISEASES

VICTORIA P. WERTH

The goal of therapy is to improve a skin condition with the least toxic and most specific approach. Because many treatments or medications can be applied directly to the skin, the option for topical therapy is attractive for treating many dermatologic diseases. However, many diseases require systemic therapy, particularly when patients have widespread involvement of the skin or a disease that cannot be improved with topical therapy. Therapies work by improving barrier function, removing scale, and altering inflammation in the skin, altering blood flow, providing antimicrobial effects, or affecting proliferating cells. Recent advances in the understanding of cutaneous biology have not been routinely accompanied by evidence-based documentation of the benefits of many specific therapies.

PRINCIPLES OF TOPICAL THERAPY

Soaks and Dressings

Water or saline applied by soaks and wet dressings can be beneficial for many skin conditions, including ulcers, by promoting healing of the epidermis and débridement of crusts. Soaked gauze is applied to involved areas for 15 to 30 minutes several times per day, and care should be taken not to allow the gauze to dry and adhere. If adherence occurs, the gauze should be soaked before the dressing is removed. Use of strong antiseptic solutions, including hydrogen peroxide, is not recommended because of toxicity to cells. Whirlpool action can enhance débridement. When large areas of skin are involved, baths are a convenient way to treat the skin with medications that reduce itching and inflammation. The best time to apply moisturizers that help trap water in the upper layers of skin is immediately after a bath or shower.

Wet-to-dry dressings are rarely used, except when initial vigorous wound débridement is necessary. Continued use after wounds are débrided traumatizes wounds and delays healing. Moist wound healing, which is often ideal, can be accomplished with a topical antibiotic such as a combination of polymyxin B and bacitracin (Polypore) or mupirocin (Bactrian), gauze impregnated with petrolatum (Vaseline), or an occlusive hydrocolloid dressing. Little evidence indicates that débriding enzymes are beneficial. Compression with an Unna or multilayered boot, which includes an elastic dressing such as Coban, can decrease local edema and facilitate wound healing. Polysporin/petrolatum gauze or occlusive dressings are placed underneath, an approach that is helpful for chronic venous, diabetic, and pressure ulcers, as well as for acute wounds. Closed wet dressings, in which gauze is soaked and then covered with an impervious material, can help when maceration and heat retention are needed. Biologic dressings with skin substitutes or keratinocytes can be beneficial for wounds that are resistant to healing. Skin grafts also can facilitate healing of otherwise nonhealing wounds. Platelet-derived growth factor, which is approved for use on diabetic ulcers, can modestly improve wound healing.

Topical Medications

Topical medications mix an active drug with preservatives, emulsifying agents, and an appropriate base or vehicle. Systemic absorption varies among patients, sites, and vehicles. Topically applied drugs are absorbed more readily through inflamed, thin skin. The base can be any of the following: a *powder,* which promotes dryness and is used to reduce maceration in intertriginous areas; a *lotion,* which is a suspension of oil in water; *solutions,* which include water, alcohol, and propylene glycol, but not oil; *gels,* which are solid at room temperature but melt on contact with the skin; a *cream,* which is an emulsion of oil in water that leaves a thin oil coating as the water evaporates; an *ointment,* which combines oils, such as petrolatum (Vaseline), with small amounts of water and which is more occlusive and hence increases the absorption of medication but also results in a greasier appearance; a *paste,* which is a mixture of powder and ointment; or a *spray*. Lotions, solutions, and gels provide less penetration than ointments do, but they are especially useful for the treatment of hair-bearing areas such as the scalp, where

greasiness is displeasing. Creams are less greasy than ointments and are useful for the face, groin, and intertriginous areas. Ointments are often more effective for dry, scaly conditions such as eczema and psoriasis and are helpful in areas that have thick skin, such as the palms and soles, but they should be avoided in infected or intertriginous areas. The choice of base is determined by the skin condition and location. Impregnated tapes are another delivery method to provide occlusion and protect the skin from manipulation.

ANTI-INFLAMMATORY AGENTS

Glucocorticoids

Topical glucocorticoids work because of their effects on vasoconstriction, proliferation, immunosuppression, and inflammation. Assays related to the ability to vasoconstrict and clinical trials of efficacy have allowed glucocorticoids to be divided into various classes based on potency (Table 437-1). These medications are typically used twice per day. Side effects include atrophy of the skin, telangiectases, purpura, striae, local skin infections (e.g., folliculitis, tinea, and candidiasis), hypopigmentation, hypertrichosis, systemic adrenal suppression when these agents are used on as little as 20% of the skin's surface area, and glaucoma when they are used around the eye. Side effects are especially prevalent when fluorinated steroids are used on thin skin (e.g., face, groin, or scrotum), and prolonged use on the face can result in facial dermatitis, acne, and an eruption resembling acne rosacea that is often exacerbated when use of the steroid is terminated. Certain conditions are more responsive to steroids, and the potency of the steroid chosen must be based on the condition and its location (Table 437-2). The superpotent class I agents should be restricted to patients with severe dermatoses, and their use normally should not exceed 2 weeks. Patients who receive these potent agents require frequent follow-up and must be carefully evaluated for the need to continue strong topical steroids. Use of any fluorinated steroid on the face requires an exact diagnosis and should be limited in the extent of application and duration of use. Intralesional glucocorticoids can be injected into individual lesions to improve delivery of the medication, and this method is commonly used to treat patients with acne cysts, hypertrophic scars, keloids, alopecia areata, granuloma annulare, discoid and panniculitic lupus erythematosus (Chapter 266), psoriasis, and lichen simplex chronicus. Triamcinolone acetonide is most frequently used, followed by the longer acting triamcinolone hexacetonide. It is important to use proper dilutions, such as 2.5 mg/mL on the face and 5 mg/mL elsewhere, to avoid local skin atrophy.

TABLE 437-1 RANKING OF SOME COMMONLY USED TOPICAL STEROIDS BY POTENCY

Super potency	Clobetasol propionate (Temovate ointment and cream), betamethasone dipropionate (Diprolene cream and ointment), diflorasone diacetate (Psorcon E ointment), halobetasol propionate (Ultravate ointment)
High potency	Amcinonide, mometasone furoate ointment, diflorasone diacetate (Florone ointment), halcinonide 0.1% cream, fluocinonide, desoximetasone, triamcinolone acetonide, diflorasone diacetate ointment and cream, betamethasone dipropionate (Diprosone), betamethasone benzoate and valerate
Medium potency	Fluticasone propionate; mometasone furoate cream; halcinonide 0.25% ointment; triamcinolone acetonide 0.1% cream and lotion; fluocinolone acetonide 0.2%, 0.25%, and 0.1% cream and 0.25% ointment and 0.5% solution; hydrocortisone valerate 0.2% ointment and cream; alclometasone dipropionate 0.5% ointment; betamethasone dipropionate 0.5% lotion; hydrocortisone butyrate 0.1% cream; betamethasone benzoate 0.25% cream; betamethasone valerate 0.1% cream and 0.5% lotion; flumethasone pivalate 0.3% cream; desonide 0.5% cream
Low potency	Hydrocortisone 1% cream

TABLE 437-2 CLINICAL APPLICATION OF TOPICAL GLUCOCORTICOIDS

Super potency and high potency	Plaque and palmoplantar psoriasis, lichen planus, dyshidrotic eczema, lichen simplex chronicus, granuloma annulare, sarcoidosis
Medium potency	Dermatitis: allergic contact, atopic, neurodermatitis
Low potency	Intertrigo, pruritus ani, seborrheic dermatitis

Systemic glucocorticoids are used for acute and chronic conditions in dermatology, but they should be avoided, if possible, or minimized because of their well-known side effects (Chapter 35). Acute conditions that commonly require systemic steroids include severe contact dermatitis such as poison ivy, photodermatitis, severe atopic dermatitis, and acute urticaria. Many skin conditions such as psoriasis and eczema become exacerbated when use of the steroids is tapered, so steroids should be avoided when possible in these conditions. The dose of steroid must be individualized to the condition and its severity. Steroid-sparing drugs, such as immunosuppressive agents, can be used to minimize the long-term use of steroids for selected conditions.

Nonsteroidal Anti-Inflammatory Agents

Psoriasis Therapies

Tars and anthralin are used for psoriasis (Chapter 438). Tars are most commonly used in conjunction with ultraviolet B (UVB) light. Tars also are used in shampoos and bath oils to treat seborrhea and psoriasis. Anthralin is a synthetic hydroxyanthrone that inhibits keratinocyte proliferation; it stains and can be irritating, but it can be effective therapy (Chapter 438).

Calcipotriol

Calcipotriol is a vitamin D derivative that has antiproliferative and immunomodulatory effects on skin. Hypercalcemia can occur if more than 100 g/week is used, so this agent cannot be used for widespread disease. It is applied twice daily, can be irritating on thin skin, and takes 6 to 8 weeks to be effective.

Retinoids

The retinoids are a group of compounds that include vitamin A and its derivatives. Their effects are mediated through several different classes of receptors, and the receptor-drug complex has effects on other regulatory proteins that affect growth factors, oncogenes, keratins, or transglutaminases. Retinoids affect cell growth, differentiation, and morphogenesis; inhibit tumor promotion and malignant cell growth; have immunomodulatory effects; and alter cell cohesion.

Topical retinoids include all-*trans*-retinoic acid (tretinoin), which is approved for acne (Retin-A) and photoaging (Renova 0.05% cream) and is also useful for hyperpigmentation, steroid-induced atrophy, and early stretch marks. Tretinoin is available as a cream (0.025%, 0.05%, 0.1%), a gel (0.01%, 0.025%), and a solution (0.05%). Adapalene (Differin 0.1% gel) and tazarotene (Tazorac) are used for acne (Chapter 439). Tazarotene is also used for psoriasis, often in combination with topical steroids to minimize irritation and chronic photodamage. Bexarotene (Targretin) 1% gel is used for the topical treatment of cutaneous lesions in patients who have refractory or persistent stage IA or IB cutaneous T-cell lymphoma. Topical retinoids can be irritating and frequently cause an exacerbation before improvement. However, they should be used regularly on lesion-prone skin to produce improvement. Moisturizers may be needed to minimize drying effects.

Systemic retinoids commonly used for the skin include isotretinoin, acitretin, and bexarotene (Targretin). They have many applications, but most frequently isotretinoin (Accutane) is used for cystic and conglobate acne, acitretin for severe psoriasis (especially the erythrodermic and pustular forms), and bexarotene for cutaneous T-cell lymphoma. Isotretinoin and acitretin also have been used to treat several forms of ichthyosis and lupus erythematosus and for the chemoprevention of skin cancers, particularly in immunosuppressed transplant recipients. The many side effects of systemic retinoids include teratogenicity, cheilitis, hair loss, headaches, hyperlipidemia, abnormal liver enzyme levels, vertebral hyperostosis, tendon and ligament calcification, osteoporosis, and central hypothyroidism with bexarotene.[1] Pregnancy must be avoided, and the use of retinoids in women of childbearing age therefore requires careful monitoring. Treatment of acne is reserved for cases of cystic acne not responding to less toxic therapies; in this setting, a 4- to 5-month course of Accutane, 0.5 to 1 mg/kg/day, is curative in 85 to 90% of patients.

Brimonidine

This selective α_2-adrenergic receptor agonist with vasoconstrictive activity is used in rosacea.

Antimalarial Drugs

Aminoquinolines include hydroxychloroquine, quinacrine, and chloroquine. These agents have inhibitory effects on pro-inflammatory cytokine

production, DNA replication, and chemotaxis. They are useful in patients with connective tissue diseases, polymorphous light eruption, sarcoidosis (Chapter 95), porphyria cutanea tarda (Chapters 210 and 439), sclerosing conditions, and vasculitis. Side effects include diarrhea, headache, irritability, psychosis, skin dyspigmentation, and, rarely, retinopathy. Retinopathy is rare if doses of chloroquine are 3.5 mg/kg/day or less and doses of hydroxychloroquine are 6.5 mg/kg/day or less. Combinations of hydroxychloroquine or chloroquine with quinacrine are frequently helpful when a solitary agent is inadequate.[2] The combination of hydroxychloroquine and chloroquine should not be used because of the additive risk for retinopathy.

Dapsone

Dapsone is a sulfone that inhibits the response of neutrophils and possibly eosinophils to chemotactic stimuli. It is useful for dermatitis herpetiformis (Chapter 439), cutaneous vasculitis (Chapter 440), pyoderma gangrenosum (Chapter 440), bullous lupus erythematosus (Chapter 266), Behçet's disease (Chapter 270), and autoimmune bullous diseases (Chapter 439). Topical 5% gel is approved for use in acne. Systemic use can be associated with side effects that include hemolysis, methemoglobinemia, peripheral neuropathy, agranulocytosis, and, rarely, a hypersensitivity syndrome with hepatitis, fevers, and rash. The glucose-6-dehydrogenase level should be checked before starting the drug, and it is common for patients with a normal glucose-6-dehydrogenase level to experience a 2-g/dL decrease in hemoglobin after achieving therapeutic doses of 100 to 200 mg/day.

Thalidomide

Thalidomide has potent anti-inflammatory effects, probably resulting from inhibition of tumor necrosis factor-α (TNF-α). It also modifies adhesion molecules on circulating leukocytes. Thalidomide is a serious teratogen, and patients must comply with strict birth control and monitoring. It is effective at a dose of 50 to 100 mg/day, with improvement beginning in 2 weeks and a full clinical response seen in 2 to 3 months in patients with severe cutaneous lupus erythematosus (Chapter 266), erythema nodosum leprosum (Chapter 438), aphthae, Behçet's disease (Chapter 270), scleromyxedema, actinic prurigo, chronic graft-versus-host disease (Chapter 178), multiple myeloma (Chapter 187), and numerous other inflammatory dermatoses. Besides teratogenesis, the main side effects include peripheral neuropathy, constipation, sedation, and, rarely, amenorrhea.

Colchicine

Colchicine, usually at a dose of 0.6 mg twice daily, is used for leukocytoclastic vasculitis (Chapter 439) and Behçet's disease, as well as some patients with epidermolysis bullosa acquisita. The main side effect of this low oral dose is diarrhea.

ANTIMICROBIAL AGENTS

Antibacterials

Topical antibiotics are used to treat superficial skin diseases, such as acne and folliculitis, as well as skin wounds and ulcers. They may work by decreasing neutrophil chemotaxis and other anti-inflammatory mechanisms. Topical solutions, gels, pledgets, and ointments are available, depending on the agent, and antibiotics include erythromycin, clindamycin, tetracycline, and metronidazole.[3,4] Benzoyl peroxide also has antibacterial properties and is quite effective for mild-to-moderate acne, while also minimizing bacterial resistance when topical antibiotics are also used (Chapter 439). Bacitracin and Polysporin ointments are typically used for wounds, but they can cause contact hypersensitivity; neomycin should be avoided because of the high incidence of allergic reactions. Mupirocin is particularly effective against *Staphylococcus* and *Streptococcus* spp., and it can be used in the nose for carriers of staphylococci. Systemic antibiotics such as penicillins, cephalosporins, and erythromycin are used in patients with soft tissue infections such as impetigo, folliculitis, furuncles, carbuncles, cellulitis, ecthyma, erysipelas, postoperative wound infections, and necrotizing fasciitis. Tetracycline, doxycycline, and minocycline are used for acne, rosacea, and perioral dermatitis.[5] Fluoroquinolones such as ciprofloxacin are useful for the treatment of gram-negative soft tissue infections.

Antifungals

Topical antifungal agents are used in patients with limited superficial fungal infections of the skin (Chapter 438). The numerous topical antifungal drugs available include the azoles (clotrimazole, econazole, ketoconazole, oxiconazole, and miconazole), which are available as creams and lotions applied once or twice daily. The creams tend to be more effective. Topical agents used for dermatophytes, but not *Candida,* are haloprogin and tolnaftate. The newer allylamine antifungals naftifine and terbinafine have fungicidal effects. Ciclopirox 8% topical solution was recently approved for use in treating and preventing relapses of onychomycosis. Nystatin creams, oral suspensions, and vaginal tablets are effective for the treatment of *Candida* infections. The combination of antifungals with potent topical steroids such as betamethasone dipropionate is not advised because of increased side effects from the steroid and decreased efficacy of the antifungal as a result of the concomitant steroid.

Systemic antifungal agents include griseofulvin, terbinafine (allylamine), ketoconazole (imidazole), itraconazole, and fluconazole. These agents are used for extensive or severe superficial skin fungal infections (Chapter 438) caused by dermatophytes, *Candida,* or *Malassezia furfur* or for local infections not responsive to topical drugs, such as those found in the nails and scalp. Itraconazole and terbinafine are the only oral antifungals approved in the United States for the treatment of onychomycosis,[6] and griseofulvin is the only oral agent approved for tinea capitis. Griseofulvin is best taken with a fatty meal to improve absorption and is the only antifungal drug not requiring regular monitoring of liver enzymes. Griseofulvin shows weak affinity for keratin, and thus it must be used for 18 months for onychomycosis of the toenails and 6 months for the fingernails to achieve even relatively poor cure rates. Terbinafine is the only fungicidal drug; the rest are fungistatic. The number of interactions with medications is lower with terbinafine than with the triazole antifungals and ketoconazole because terbinafine does not inhibit or induce hepatic isoenzyme cytochrome P (CYP3A4) (Chapter 29). However, terbinafine affects CYP2D6, another hepatic isoenzyme, so it is relatively contraindicated in patients who are taking cyclosporine or rifampin. Itraconazole and fluconazole have been used in pulse-dosing regimens for the treatment of onychomycosis. Fluconazole and terbinafine are not dependent on gastric acidity for optimal gastrointestinal absorption. Overall, the side effects of the systemic antifungal agents are similar and include headache and gastrointestinal symptoms (griseofulvin, terbinafine), nausea and vomiting (itraconazole, fluconazole, ketoconazole), hepatitis, and lupus-like syndromes (terbinafine).

Antivirals

Verrucae are treated with various destructive modalities, including 50 to 80% dichloroacetic acid and trichloroacetic acid solutions, podophyllin resin, and podofilox. Topical antiviral creams such as penciclovir and acyclovir do not significantly shorten the course of herpes simplex. Systemic antiviral drugs include acyclovir, valacyclovir, famciclovir, and foscarnet, which are used to treat primary and recurrent herpes simplex (Chapter 374) and herpes zoster (Chapter 375), although only acyclovir is approved for treatment of herpes zoster. These agents specifically block the function of herpesvirus DNA polymerase. Valacyclovir and famciclovir are available only orally in the United States, but their prolonged intracellular half-life allows less frequent dosing than with acyclovir. Patients with herpes zoster require higher doses than do patients with herpes simplex. Side effects include nausea and headaches. The varicella-zoster vaccine has significantly decreased the incidence of herpes zoster in individuals older than 50 years of age, and current recommendation is for vaccination after the age of 60 (Chapter 18). If possible, it should be administered before starting oral immunosuppressive agents.

Antiparasitics

Topical antiparasitic medications are used to treat pediculosis capitis, pediculosis pubis, and scabies. In addition, topical metronidazole has anti-inflammatory properties and is used to treat rosacea. For pediculosis and scabies, clothing and bedding must be washed and all family members must be treated. Effective treatment includes 1% γ-benzene hexachloride (lindane), a chlorinated hydrocarbon pesticide that should not be used on young children or on pregnant or lactating women. It is ineffective against nits and thus must be reapplied after 1 week. Permethrin 5% cream (Elimite) for scabies or 1% cream rinse is particularly effective for head lice and requires just one application; 10% crotamiton (Eurax) and 5% topical sulfur ointments are less effective. Malathion is a moderately toxic organophosphate insecticide, but it must be applied overnight to treat lice. Pyrethrins (RID, Nix) are best used twice, 1 week apart, to treat head lice and nits.

ANTIPRURITIC OR ANESTHETIC AGENTS

Topical Analgesics

Capsaicin, an active ingredient of cayenne peppers and other plants of the genus *Capsicum* is used for postherpetic neuralgia and other painful

nerve-related conditions. It causes excitation of neural afferent C fibers and reduces substance P levels. Capsaicin causes a burning sensation and is applied four or five times daily for 5 to 6 weeks. Eutectic mixture of local anesthetic (EMLA) is a mixture of lidocaine and prilocaine that is used under occlusion to induce cutaneous anesthesia before a procedure. Lidocaine can be used as a topical anesthetic, but benzocaine should be avoided because it is a sensitizer.

Antipruritic Agents

Doxepin 5% cream is used for localized pruritus. Menthol is a cyclic terpene plant alcohol used for non–histamine-related itching. Pramoxine hydrochloride is a topical anesthetic used for mild-to-moderate itching. Oral antihistamines play an important role in controlling pruritus in many skin conditions (Table 437-3), especially those mediated by histamine, such as urticaria,[7] angioedema, and urticaria pigmentosa. The sedating and anticholinergic properties of many H_1-receptor antihistamines probably account for some of their efficacy. H_1-receptor antihistamines are the cornerstones of routine therapy, and if an agent from one group of H_1-receptor antihistamines is ineffective, an agent from a different class should be administered or combined. Second-generation H_1-receptor antihistamines are less sedating and are used if patients cannot tolerate or do not improve after taking first-generation agents. The combination of two different H_1-receptor antihistamines can be used when a solitary agent does not work; in particular, use of a sedating antihistamine at night and a second-generation antihistamine during the day can be helpful. The skin contains both H_1- and H_2-receptors, and occasionally combining H_1- and H_2-receptor antagonists can be beneficial. Usually, first-generation agents (e.g., hydroxyzine, 10 to 25 mg every 6 hours) are started at low doses and increased as tolerated, and regular continuous dosing is recommended. The tricyclic antidepressant doxepin, normally started at 10 to 25 mg at bedtime, has both anti–H_1- and H_2-receptor activity, but it interacts with drugs metabolized by the CYP450 pathway. Side effects of commonly used first-generation antihistamines include sedation, dry mouth, blurred vision, constipation, and urinary retention, and lower doses may be required in elderly patients. The recommended dose for second-generation antihistamines (e.g., fexofenadine, 60 mg twice daily) should not be exceeded.

TABLE 437-3 OVERVIEW OF ANTIHISTAMINES

ANTIHISTAMINE GROUP	GENERIC NAME	AVERAGE ORAL ADULT DOSES
FIRST-GENERATION H_1-TYPE ANTIHISTAMINES		
Alkylamine	Brompheniramine (Dimetapp)	4 mg q4-6h
	Chlorpheniramine (Chlor-Trimeton)	4 mg q4-6h (short acting); 8-12 mg q8-12h (long acting)
Amino alkyl ether (ethanolamine)	Clemastine fumarate	1.34 mg bid or 2.68 mg qd-tid
	Diphenhydramine (Benadryl)	25-50 mg q4-6h
Ethylenediamine	Pyrilamine (Triaminic)	30 mg bid
Phenothiazine	Promethazine (Phenergan)	10-12.5 mg qid
	Trimeprazine (Temaril)	2.5 mg q6h
Piperidine	Azatadine	1-2 mg q8-12h
	Cyproheptadine	4 mg q8h
	Diphenylpyraline	2 mg tid-qid
Piperazine	Hydroxyzine (Atarax)	25-100 mg tid-qid
SECOND-GENERATION H_1-TYPE ANTIHISTAMINES		
Alkylamine	Acrivastine (combined with pseudoephedrine in allergy medication)	8 mg qid
Piperidine	Astemizole (Hismanal)	10 mg qd
	Loratadine (Claritin)	10 mg qd
	Fexofenadine (Allegra)	60 mg bid or 180 mg qd
Piperazine	Cetirizine (Zyrtec)	5-10 mg/day
H_2-TYPE ANTIHISTAMINES		
	Cimetidine (Tagamet)	400 mg bid
	Ranitidine (Zantac)	150 mg bid
	Famotidine	10 mg bid
	Nizatidine	300 mg hs
H_1- AND H_2-TYPE ANTIHISTAMINES		
	Doxepin (Sinequan)	10-25 mg hs

bid = twice daily; hs = at bedtime; qd = once daily; qid = four times daily; tid = three times daily.

AGENTS THAT IMPROVE SURFACE FUNCTIONS (LUBRICATION, SCALE)

Moisturizers

Moisturizers improve skin by diminishing scale and increasing water content. They usually contain mixtures of water and fatty substances such as petrolatum, lanolin, lanolin derivatives, and fatty alcohols. Greasy moisturizers tend to function better, but they are less acceptable cosmetically.

Keratolytics

α-Hydroxy acids (lactic acid, glycolic acid, citric acid, glucuronic acid, pyruvic acid) are extremely effective keratolytics. They are helpful in treating disorders of keratinization and photoaging, as well as acne. Propylene glycol, used in 40 to 60% aqueous solutions, can decrease scaling. Salicylic acid, which works by decreasing keratinocyte adhesion and hydrating keratins, is used in a range of concentrations with many different bases to remove scale, to soften the stratum corneum, or as destructive therapy to remove warts and calluses. Urea is used in varying concentrations to treat scaling.

IMMUNE THERAPIES

Hormonal Therapies

Systemic therapies to modulate androgen production can be beneficial in patients with acne and hidradenitis suppurativa. Such treatments for moderate acne include spironolactone and, in women, U.S. Food and Drug Administration (FDA)-approved combination oral contraceptives, such as Ortho Tri-Cyclen (norgestimate and ethinyl estradiol), Estrostep (norethindrone acetate and ethinyl estradiol), and Yaz (drospirenone and ethinyl estradiol), as they would be prescribed for contraception (Chapter 238).

Immunosuppressive Agents

Topical cytotoxic drugs include 5-fluorouracil, mechlorethamine (nitrogen mustard), carmustine (BCNU), bleomycin, and the calcineurin inhibitors tacrolimus and pimecrolimus. Topical 5-fluorouracil interferes with pyrimidine metabolism and action and blocks DNA synthesis. It is used to treat actinic keratosis, superficial basal cell cancer, Bowen's disease, bowenoid papulosis, actinic cheilitis, and warts. Topical use does not cause systemic toxicity, but expected side effects include local irritation, erythema, and pain. Nitrogen mustard and BCNU, which have alkylating agents that inhibit DNA, RNA, and protein synthesis, are used to treat cutaneous T-cell lymphoma (Chapter 185); they can cause cutaneous reactions and myelosuppression, and nitrogen mustard commonly causes a cutaneous hypersensitivity reaction.

Intralesional bleomycin, which disrupts DNA synthesis, has been used to treat warts. Topical tacrolimus (Prograf) and pimecrolimus, immunosuppressive macrolides that act on T lymphocytes to inhibit interleukin-2 (IL-2) transcription, are used for atopic dermatitis, allergic contact dermatitis, psoriasis, and several other inflammatory skin conditions. They frequently cause a burning sensation in the skin; although systemic absorption is minimal, ongoing studies are assessing whether the risk for cancer is increased by their topical use. Systemic immunosuppressives such as methotrexate, azathioprine, thioguanine, hydroxyurea, mycophenolate (CellCept), cyclophosphamide, chlorambucil, rapamycin, and cyclosporine are used for numerous inflammatory or immunologically mediated skin conditions, particularly for widespread psoriasis (Chapter 438) and as glucocorticoid-sparing agents for autoimmune blistering diseases.

Immunomodulatory Therapies

Imiquimod, available as a 5% cream, is an imidazoquinolinamine that has antitumor and antiviral activity. It induces local production of interferon-γ and is used to treat warts and superficial skin cancers. Topical 3% diclofenac in hyaluronic acid, which blocks the induced cyclooxygenase-2 found in precancerous lesions, is approved to treat actinic keratosis. Many systemic immunomodulatory drugs are currently used in dermatology. These include interferons and TNF-α inhibitors such as etanercept, adalimumab, and infliximab (Remicade).[8] Anti-IL-12/23 therapy and anti-IL-17 therapy are effective for psoriasis. Ipilimumab, an antibody against T-lymphocyte-associated antigen 4 is effective for advanced melanoma. Anti-CD20 antibodies have been effective in the treatment of autoimmune skin-blistering diseases. Interferon alfa-2b is used both intralesionally and subcutaneously to treat genital warts, high-risk

melanoma, Kaposi's sarcoma, hemangiomas, cutaneous T-cell lymphoma, keloids, Behçet's disease, cryoglobulinemia and vasculitis from hepatitis C (Chapter 149), and perhaps basal cell and squamous cell carcinoma. Total interferon doses are generally 3 million IU or less per session, and systemic doses are usually administered 3 days per week. Side effects include flulike symptoms, leukopenia, anemia, and hepatitis. Denileukin diftitox is a fusion protein consisting of a fragment of diphtheria toxin genetically fused to IL-2. It targets IL-2 receptors on the surface of malignant cells and is approved for use in patients with resistant or recurrent cutaneous T-cell lymphoma. Alefacept, which is used to treat psoriasis, affects T cells by blockade of costimulatory signals. Intravenous immunoglobulin, which is used to treat certain autoimmune skin diseases, including pemphigus vulgaris, cicatricial pemphigoid, and dermatomyositis, probably works through Fc receptor modulation and anti-idiotype interactions. Antibodies against the programmed cell death 1 (PD-1) receptor are effective in advanced melanoma.

Histone deacetylase inhibition increases acetylation of lysine residues that form the octomeric histone core of chromatin, thereby decreasing the ability of the histones to bind to DNA. This decreased binding allows chromatin expansion, permitting transcription of the tumor suppressor genes. However, histone deacetylase inhibitors affect acetylation globally and may have wider effects on various cellular functions. Two novel inhibitors (vorinostat and romidepsin) are approved by the FDA for use in patients with cutaneous T-cell lymphoma.[9]

Extracorporeal photochemotherapy (photopheresis), which combines 8-methoxypsoralen and ultraviolet A (UVA) irradiation of lymphocytes, is used for Sézary's syndrome, the leukemic form of cutaneous T-cell lymphoma (Chapter 185). Plasmapheresis, used in combination with other immunosuppressive therapies, can remove autoantibodies and immune complexes in patients in whom autoimmune disease or cryoglobulinemia is resistant to other therapies.

TARGETED PATHWAYS FOR CANCER TREATMENT

Hedgehog Signaling Pathway Inhibition

The hedgehog pathway is activated in most basal cell carcinomas. The central component of the hedgehog signaling pathway is smoothened (SMO), a transmembrane protein that initiates a signaling cascade. Vismodegib, an antismoothen inhibitor, has been approved for treatment of patients with advanced basal cell carcinomas.

BRAF Kinase Inhibition

Vemurafenib is a potent inhibitor of mutated BRAF. It has antitumor effects against the BRAF V600E mutation, and is used for advanced melanoma.

Other Therapies

Phototherapy and Laser

Ultraviolet treatments are given with different wavelengths, depending on the condition and the response to treatment. Currently, clinicians use broadband UVB (290 to 320 nm), narrow-band 311-nm UVB, PUVA (psoralen with 320- to 400-nm UVA), and UVA1 (340 to 400 nm). Both forms of UVB and PUVA are used for psoriasis and vitiligo, but other conditions such as nummular and atopic dermatitis, pruritus resulting from uremia, and cutaneous T-cell lymphoma are treated in this way. High-dose UVA1 is used, mainly in Europe, to treat atopic dermatitis, localized scleroderma, and mastocytosis. PUVA is associated with increased risk for skin cancers, including melanoma. The risks related to long-term UVA therapy are currently unknown, but photoaging is associated with UVA and there have been reports of an increase in melanoma associated with use of suntanning beds, in which much of the exposure is to UVA. Laser therapy is used to treat vascular lesions such as port-wine stains, tattoos, psoriasis, benign skin tumors, and photodamage, as well as to remove hair. Photodynamic therapy involves activation of a photosensitizer by illumination with visible light, which leads to photochemical tissue destruction or immunomodulation. Photodynamic therapy can be used to treat actinic keratosis, Bowen's disease, and superficial basal cell carcinoma by causing selective tissue necrosis and tumor destruction. Fractional lasers have been successfully used for treatment of actinic keratoses, photodamage, and scars.

Dermatologic Surgery

Although approaches such as desiccation and curettage can be used for some skin tumors, others require excisional surgery or Mohs' microscopic controlled surgery to ensure complete removal of lesions. If the tumors are recurrent, of a pathologic type that increases the likelihood for recurrence, or large and requiring clearance of the tumor before repair, the Mohs approach can provide rapid documentation of full removal while sparing as much normal tissue as possible. After the margins have been cleared of tumor, flaps and grafts can be used immediately for repair of the resultant defects.

Patients with extensive actinic damage resulting in either large numbers of actinic keratoses or photodamage can be treated with various ablative approaches that use either chemical peels or laser resurfacing with the carbon dioxide laser. Chemical peels can be performed at different depths and intensities, and agents can include glycolic acid, acetic acid, or even phenol. Lasers used to remove sun-induced lentigines include Q-switched lasers such as the neodymium:yttrium-aluminum-garnet, ruby, and alexandrite lasers. Many patients seek treatment of wrinkles with soft tissue augmentation that uses human-derived collagen and hyaluronic acid or with a muscle relaxer, botulinum type A exotoxin.

Hair transplants are a surgical approach to the problem of hair loss. The process includes harvesting hair grafts from the posterior of the scalp and placing the grafts in areas of alopecia.

Sunscreen

Transparent sunscreens absorb photons of light. They are rated by the sun protection factor (SPF), which is determined by the ratio of ultraviolet exposure needed to cause erythema in protected versus unprotected skin. Most sunscreens work in the UVB range or shorter UVA wavelengths. Examples of UVB-absorptive compounds include aminobenzoates, cinnamates, salicylates, and benzophenones. Short-wavelength UVA-absorptive compounds include benzophenones and anthranilates. The best UVA blocking agent in the United States is avobenzone (Parsol 1789), which can be combined with UVB screens.[10] Some sunscreens are water resistant or waterproof, as determined by the substantivity of the sunscreen, and these agents provide continued protection after sweating or swimming. Sunscreens can cause irritation and, rarely, contact allergic reactions. Physical sunscreens, such as zinc oxide and titanium dioxide, reflect light from the skin and include newer micronized reflecting powders that provide broad-spectrum (UVB and UVA) protection. Sunscreens decrease skin cancers and photodamage. UVB is partially reflected by clothing, and sun-protective clothing can provide substantial protection (Solumbra, SPF of 30).

Cosmetics: Camouflage, Bleaching, and Hair Loss

Patients with numerous skin conditions benefit from camouflage cosmetics, which can also cause contact hypersensitivity. Products such as Dermablend can be blended to match skin colors, are thicker, can cover disfiguring lesions, and can be fixed with powder. Hydroquinones, topical retinoic acid, and azelaic acid (inhibits tyrosinase) are used to treat hyperpigmented conditions such as melasma and lentigines; these agents can be irritating and cause dyspigmentation. Topical minoxidil, 2% (available over the counter) and 5% solutions, are used for androgenic alopecia and alopecia areata. Finasteride, a 5α-reductase inhibitor, is effective in men with androgenic alopecia.

Grade A References

A1. Mason AR, Mason J, Cork M, et al. Topical treatments for chronic plaque psoriasis. *Cochrane Database Syst Rev*. 2013;3:CD005028.

A2. Jackson JM, Fowler J, Moore A, et al. Improvement in facial erythema within 30 minutes of initial application of brimonidine tartrate in patients with rosacea. *J Drugs Dermatol*. 2014;13:699-704.

A3. El-Gohary M, van Zuuren EJ, Fedorowicz Z, et al. Topical antifungal treatments for tinea cruris and tinea corporis. *Cochrane Database Syst Rev*. 2014;8:CD009992.

A4. Arowojolu AO, Gallo MF, Lopez LM, et al. Combined oral contraceptive pills for treatment of acne. *Cochrane Database Syst Rev*. 2012;7:CD004425.

A5. Lessin SR, Duvic M, Guitart J, et al. Topical chemotherapy in cutaneous T-cell lymphoma: positive results of a randomized, controlled, multicenter trial testing the efficacy and safety of a novel mechlorethamine, 0.02%, gel in mycosis fungoides. *JAMA Dermatol*. 2013;149:25-32.

A6. Gupta AK, Paquet M, Villanueva E, et al. Interventions for actinic keratoses. *Cochrane Database Syst Rev*. 2012;12:CD004415.

A7. Robert C, Long GV, Brady B, et al. Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med*. 2015;372:320-330.

A8. Sekulic A, Migden MR, Oro AE, et al. Efficacy and safety of vismodegib in advanced basal-cell carcinoma. *N Engl J Med*. 2012;366:2171-2179.

A9. Chapman PB, Hauschild A, Robert C, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med*. 2011;364:2507-2516.

A10. Togsverd-Bo K, Haak CS, Thaysen-Petersen D, et al. Intensified photodynamic therapy of actinic keratoses with fractional CO2 laser: a randomized clinical trial. *Br J Dermatol*. 2012;166:1262-1269.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Scarisbrick JJ, Morris S, Azurdia R, et al. U.K. consensus statement on safe clinical prescribing of bexarotene for patients with cutaneous T-cell lymphoma. *Br J Dermatol*. 2013;168:192-200.
2. Chang AY, Piette EW, Foering KP, et al. Response to antimalarials in cutaneous lupus erythematosus: a prospective analysis. *Arch Dermatol*. 2011;147:1261-1267.
3. Dawson AL, Dellavalle RP. Acne vulgaris. *BMJ*. 2013;346:f2634.
4. Lynde C, Tan J, Andriessen A, et al. A consensus on acne management focused on specific patient features. *J Cutan Med Surg*. 2014;18:243-255.
5. Chang BP, Kurian A, Barankin B. Rosacea: an update on medical therapies. *Skin Therapy Lett*. 2014;19:1-4.
6. Gupta AK, Simpson FC. New pharmacotherapy for the treatment of onychomycosis: an update. *Expert Opin Pharmacother*. 2015;16:227-236.
7. Yosipovitch G, Bernhard JD. Clinical practice: chronic pruritus. *N Engl J Med*. 2013;368:1625-1634.
8. Armstrong AW, Bagel J, Van Voorhees AS, et al. Combining Biologic Therapies With Other Systemic Treatments in Psoriasis: Evidence-Based, Best-Practice Recommendations From the Medical Board of the National Psoriasis Foundation. *JAMA Dermatol*. 2014 [Epub ahead of print].
9. Foss FM. Treatment strategies for peripheral T-cell lymphomas. *Best Pract Res Clin Haematol*. 2013;26:43-56.
10. Jou PC, Feldman RJ, Tomecki KJ. UV protection and sunscreens: what to tell patients. *Cleve Clin J Med*. 2012;79:427-436.

REVIEW QUESTIONS

1. A clean ulcer that needs to granulate and re-epithelialize should be treated by:

A. Moist wound healing with a topical ointment and dressing.
B. Wet-to-dry dressing with dry gauze and no ointment.
C. Being left open to heal.
D. Oral antibiotics.
E. Oral steroids.

Answer: A Moist wound healing is the standard for chronic wound management. Moist wounds heal two to three times faster than dry wounds. The primary purpose of wet-to-dry dressings is the mechanical débridement of necrotic tissue. Dried tissue is more prone to infection and pain, and it heals more slowly than if kept moist. A moist wound environment physiologically favors cell migration and matrix formation, while accelerating healing of wounds by promoting autolytic débridement.

2. The comination clotrimazole and betamethasone dipropionate topical therapy (lotrisone) is:

A. More effective for fungal infections than clotrimazole alone.
B. Safe for use in the groin and face
C. Approved for use in children younger than 12 years of age.
D. Should be used sparingly because of risks for thinning the skin and striae.
E. Is effective against bacteria.

Answer: D Fluorinated topical steroids are potent and cause thinning of the skin and striae. As a result, they should not be used in areas where the skin is thin, such as the face, or where the area is occluded and increases absorption in the skin, such as the groin. It is preferable to treat fungal infections with an antifungal agent, and if a topical anti-inflammatory steroid is also desired, a separate treatment with a mild lower class steroid may be used.

3. Topical antibiotics as a solitary agent for acne for more than 3 months is not recommended because:

A. Increased risk of drug toxicity.
B. Higher incidence of bacterial resistance after that time.
C. Effects on thinning of skin.
D. It should not be combined with other therapies.
E. It causes photosensitivity.

Answer: B When topical antibiotics are used in combination with benzoyl peroxide, there is reduced emergence of resistant strains of *Propionibacterium acnes*.

4. People starting a TNF blocker or other biologic therapy should be tested to exclude:

A. Heart disease.
B. Underlying malignancy.
C. Lung disease.
D. Atherosclerosis.
E. Tuberculosis.

Answer: E The biologics are immunosuppressants that facilitate reactivation of tuberculosis, which can then disseminate. Unless a person has a history of a positive test followed by appropriate treatment, routine tuberculosis testing is recommended, either by PPD or quantiferon gold testing.

438

ECZEMAS, PHOTODERMATOSES, PAPULOSQUAMOUS (INCLUDING FUNGAL) DISEASES, AND FIGURATE ERYTHEMAS

HENRY W. LIM

ECZEMA

The more commonly encountered eczemas (Table 438-1) share similar histologic characteristics. However, they have varying degrees of edema within the epidermis (spongiosis) and of infiltration with lymphocytes and macrophages in the superficial dermis.

Nummular Dermatitis

Nummular dermatitis occurs most frequently in patients who are in their 50s to 60s. Both sexes are affected; in temperate climates, this condition is most frequently seen in the winter. The condition appears to be more frequent and severe among Asians. The pathogenesis is unclear, although xerosis plays a significant role.

Patients usually present with pruritic, coin-shaped, erythematous patches with some scales and occasionally with pinhead-sized vesicles (Figs. 438-1 and 438-2). Lesions may be excoriated and lichenified (i.e., thickened skin with accentuation of skin markings). Legs and arms are commonly affected, and less so on the trunk; facial involvement is uncommon. All patients should be educated about the care of dry skin, such as the use of emollients and moisturizing soaps, and avoidance of long, hot showers. Topical corticosteroid ointments (e.g., triamcinolone ointment, 0.1% twice daily for 1 to 2 weeks) are helpful for active lesions, and oral antihistamines (e.g., fexofenadine, 180 mg every morning, and hydroxyzine, 25 to 50 mg at bedtime, as needed) are useful for pruritus. In severe cases, narrow-band ultraviolet B (NB-UVB) phototherapy, a short course of oral corticosteroids (prednisone, 0.5 to 1 mg/kg/day, with a maximal dose of 60 mg/day, for 1 to 2 weeks, then taper in 10 to 14 days) or day hospitalization for intensive topical and NB-UVB therapy is beneficial.

TABLE 438-1 ECZEMAS

Nummular dermatitis
Dyshidrosis
Atopic dermatitis
Seborrheic dermatitis
Allergic contact dermatitis
Irritant contact dermatitis

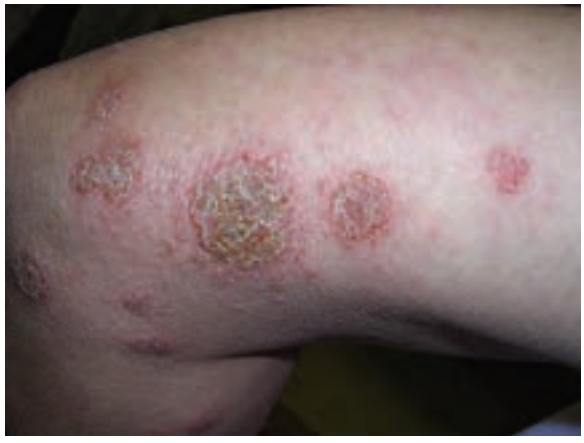

FIGURE 438-1. Nummular dermatitis. Coin-shaped erythematous patches.

Dyshidrosis

Dyshidrosis manifests as deep-seated, pinhead-sized vesicles, most commonly along the sides of the fingers (Figs. 438-3 and 438-4). In severe cases, the palms and soles also may be involved. Lesions are usually pruritic, associated with xerosis, scaliness, and fissures. Dyshidrosis is seen in individuals who wash their hands frequently, such as restaurant workers and mothers of young infants. With the widespread use of hand sanitizers, dyshidrosis has become less common in health care workers. Treatment follows a sequential

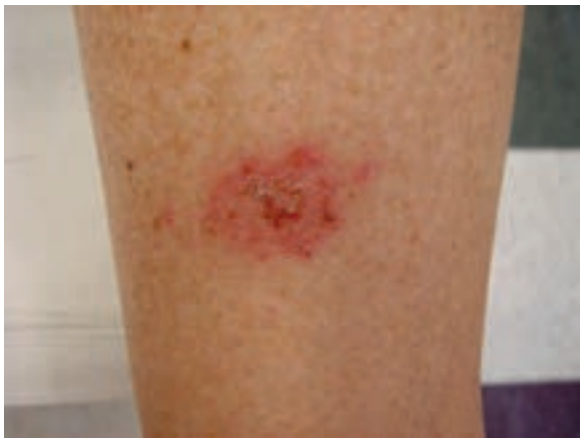

FIGURE 438-2. Nummular dermatitis. Coin-shaped erythematous patch.

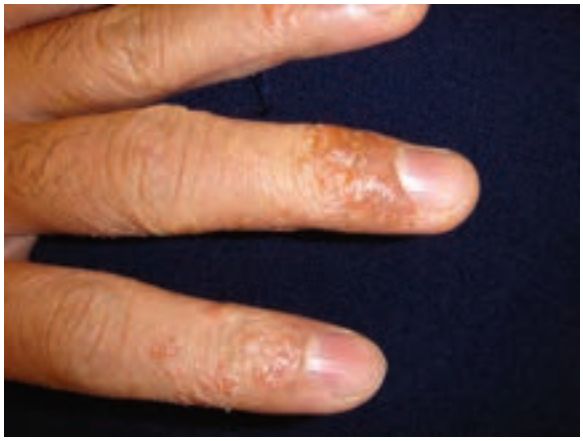

FIGURE 438-3. Dyshidrosis. Deep-seated vesicles and scaliness on fingers.

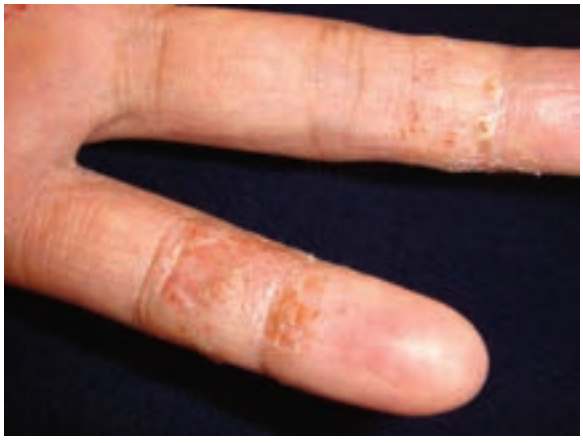

FIGURE 438-4. Dyshidrosis. Deep-seated vesicles and scaliness on fingers.

order: (1) replacing soap-and-water handwashing with the use of hand sanitizers, (2) liberal use of emollients, (3) topical corticosteroid ointments (e.g., fluocinonide ointment, 0.05% twice daily for 2 weeks), and (4) oral antihistamines (e.g., fexofenadine, 180 mg every morning, and hydroxyzine, 25 to 50 mg at bedtime, as needed).

Atopic Dermatitis

Atopic dermatitis is most commonly seen among young children, but severe cases persist into adulthood. In 90% of patients, the disease starts before the age of 5 years. The prevalence has been estimated at between 15 and 23%. Patients usually present with xerosis, erythematous scaly patches, small vesicles, excoriations, crusting, and, not infrequently, impetiginization (Fig. 438-5). In dark-skinned patients, a papular variant is commonly seen (Fig. 438-6). With chronic scratching and rubbing, hyperpigmentation and lichenification occur. Commonly affected sites include the periorbital area and flexor areas such as the neck, antecubital fossa, and popliteal fossa. In severe cases, the entire skin surface may be involved. Diagnosis is made by the typical morphology, the distribution of lesions, and family and personal history of atopy. The therapeutic ladder consists of (1) emollients; (2) topical corticosteroid ointments (e.g., triamcinolone ointment, 0.1% twice daily for 1 to 2 weeks), or topical calcineurin inhibitors (tacrolimus ointment, 0.1% for 3 to 4 weeks, or pimecrolimus cream, 1% for 3 to 4 weeks); (3) oral antihistamines (e.g., fexofenadine, 180 mg every morning, and hydroxyzine, 25 to 50 mg at bedtime, as needed); and (4) NB-UVB phototherapy. Although topical tacrolimus and pimecrolimus have "black box" warnings from the U.S. Food and Drug Administration for their potential association with the development of malignancy, a safety study of tacrolimus ointment for up to 4 years in patients with pediatric atopic dermatitis showed no immunosuppressive side effects. In patients with recalcitrant cases, day hospitalization, oral prednisone (0.5 to 1 mg/kg/day), cyclosporine (3 to 5 mg/kg/day), mycophenolate mofetil (1 to 2 g/day), and dupilumab (an interleukin-4 and -13 blocker at 300 mg subcutaneously weekly) have been successful.

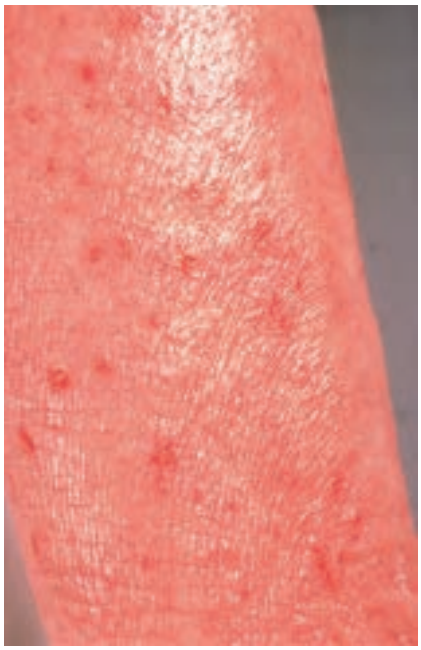

FIGURE 438-5. Atopic dermatitis. Note the erythema, excoriation, and lichenification.

FIGURE 438-6. Atopic dermatitis in a dark-skinned patient. Note the typical papular variant commonly seen in dark-skinned individuals.

Seborrheic Dermatitis

Seborrheic dermatitis is a common condition that occurs as erythematous patches with fine, greasy-appearing scales, most commonly on the malar area, midforehead, midchest, and scalp (Fig. 438-7). In dark-skinned individuals, lesions may be hypopigmented (Fig. 438-8). The pathogenesis is unknown, although *Pityrosporum ovale* is believed to play a role. Lesions are common in patients with human immunodeficiency virus (HIV) infection (Chapter 392). The diagnosis is made clinically. Topical corticosteroids (e.g., hydrocortisone cream 2.5%, twice daily for 1 to 2 weeks for facial lesions, fluocinolone acetonide 0.01% solution for the scalp, twice daily for 3 to 4 weeks) can rapidly reduce the inflammation; then topical ketoconazole cream, 2% twice daily, as needed (or shampoo 2% daily or every other day for the scalp), is safe for long-term treatment.

Allergic Contact Dermatitis and Irritant Contact Dermatitis

Allergic contact dermatitis and irritant contact dermatitis are induced by exogenous agents. Allergic contact dermatitis is a delayed hypersensitivity response to external allergens, whereas irritant contact dermatitis is a nonspecific toxic response to contact irritants. In both conditions, lesions occur in the exposed area. In severe cases, however, nonexposed areas may be less intensely involved. Patients with allergic contact dermatitis present with

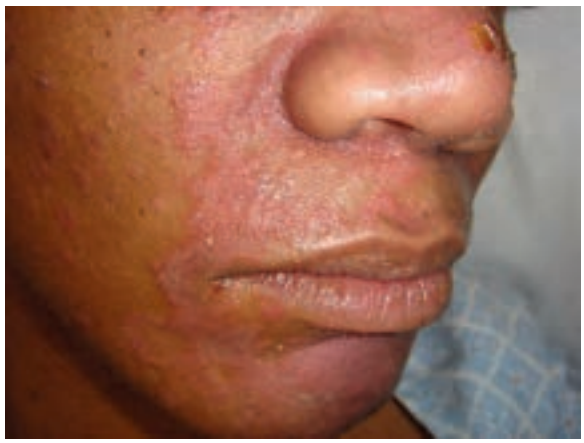

FIGURE 438-7. Seborrheic dermatitis presenting as erythematous patches and plaques with fine scales on the malar area of an HIV-positive patient.

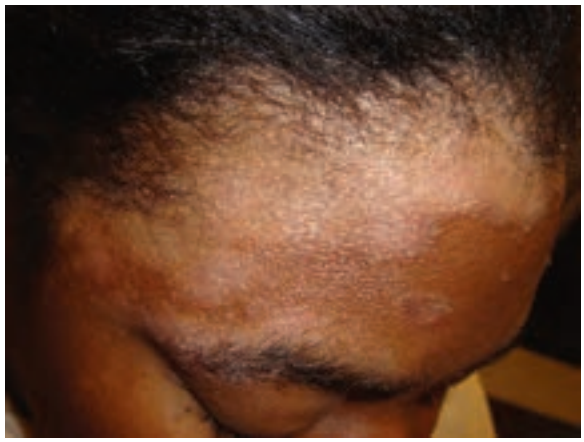

FIGURE 438-8. Seborrheic dermatitis. Hypopigmentation with fine scales on the forehead and scalp.

erythematous pruritic papules and vesicles. Lesions resolve with fine scales. Postinflammatory hyperpigmentation may be observed, especially in dark-skinned individuals. Histologically, epidermal edema and dermal histiocytic infiltrates are observed. Irritant contact dermatitis manifests with lesions morphologically similar to those of allergic contact dermatitis. However, irritant contact dermatitis is usually associated with a burning sensation rather than with pruritus. Postinflammatory hyperpigmentation is frequently observed. Histologic changes consist of necrotic keratinocytes, epidermal necrosis, and neutrophilic infiltrates. Management includes identification and removal of the offending agent, as well as symptomatic treatments such as topical corticosteroids and oral antihistamines.

PHOTODERMATOSES

Photodermatoses are cutaneous eruptions secondary to exposure to sunlight (Table 438-2). By convention, electromagnetic radiation in the UV region is divided into UVC (200 to 290 nm), UVB (290 to 320 nm), UVA-2 (320 to 340 nm), and UVA-1 (340 to 400 nm). Visible light extends from 400 to 760 nm. Because UVC emitted by the sun is absorbed by ozone in the stratosphere, UVC does not reach the Earth's surface. UVB, UVA, and less frequently, visible light are the relevant spectra in photodermatoses.

Polymorphic Light Eruption

Polymorphic light eruption, the most common immunologically mediated photodermatosis, occurs in 10 to 20% of the general population. It usually occurs in young adults, has a slight female predominance, and is seen worldwide.[1] Affected individuals are less susceptible to cutaneous photoimmunosuppression and hence have an enhanced response to UV-induced neoantigens in the skin. Lesions usually occur in early spring, within a few hours of exposure to sunlight. Lesions can manifest as pinhead papules (common among dark-skinned patients), papules, papulovesicles, or, less commonly, vesicles (Fig. 438-9); they can also be pruritic. Usually, lesions persist for several days and resolve spontaneously. The condition tends to improve as the sunny season progresses, a phenomenon known as hardening.

The course is chronic; only 11% of patients have complete resolution of the disease in 16 years and 24% in 32 years. Diagnosis is based on the typical history and morphologic features of the lesion; the diagnosis can be confirmed by the induction of lesions with provocative phototesting. When lesions occur primarily on the face, a diagnosis of lupus must be excluded. Management consists of sun avoidance and the use of broad-spectrum sunscreens, topical corticosteroids, and oral antihistamines. In severe cases, desensitization treatment using NB-UVB has been successful. Desensitization is usually performed in early spring by exposing patients to increasing doses of NB-UVB three times weekly for 15 treatments.

TABLE 438-2 SELECTED PHOTODERMATOSES

Polymorphic light eruption
Chronic actinic dermatitis
Phototoxicity and photoallergy
Porphyrias

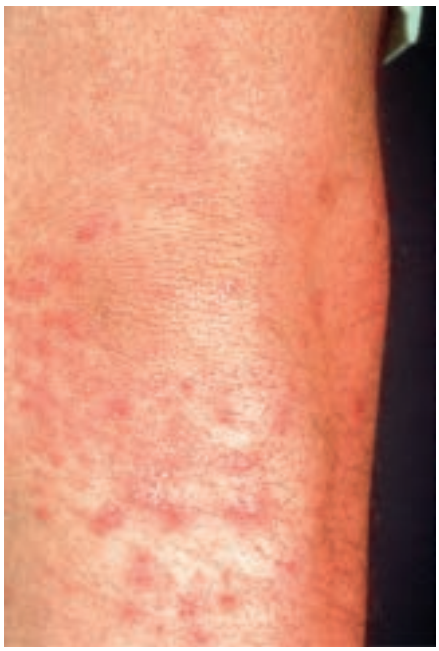

FIGURE 438-9. Polymorphic light eruption. Erythematous papules a few hours after exposure to sunlight.

Chronic Actinic Dermatitis

Chronic actinic dermatitis is a chronic photodermatosis that occurs most commonly in men in their 60s and 70s. It occurs in patients of all ethnic groups, but in the United States it is more commonly seen in dark-skinned individuals. It is seen in 5 to 17% of patients referred for evaluation of photosensitivity. Chronic actinic dermatitis can evolve from photoallergic contact dermatitis, allergic contact dermatitis, or exposure to a known photosensitizing agent; however, it also can arise de novo. Investigators have postulated that this condition represents a delayed hypersensitivity response to an unidentified antigen.

Patients present with lichenified plaques on sun-exposed areas (Figs. 438-10 and 438-11). Typically, sun-protected areas, such as the postauricular area, the area underneath the chin, the area above the eyes, and the trunk, are spared. Histologically, a dermal lymphohistiocytic infiltrate is seen and atypical mononuclear cells may be observed. On phototesting, patients have increased sensitivity to UVA, UVB, or visible light or a combination of these. In a study of 178 cases, 10% resolved in 5 years and 50% in 15 years. An association with HIV infection (Chapter 392) has been reported.

The diagnosis is based on the patient's history and on the morphologic features and distribution of the lesions. It is confirmed by phototesting.

Management is challenging. During the sunny season, it is critical that patients practice maximal photoprotection, consisting of staying in the shade, using broad-spectrum sunscreens with high sun protection factor, wearing

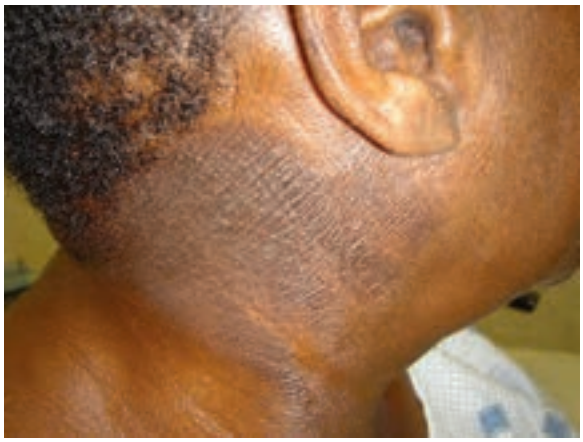

FIGURE 438-10. Chronic actinic dermatitis. Hyperpigmentation and lichenification; note sparing of the sun-protected areas of the neck and infra-auricular area.

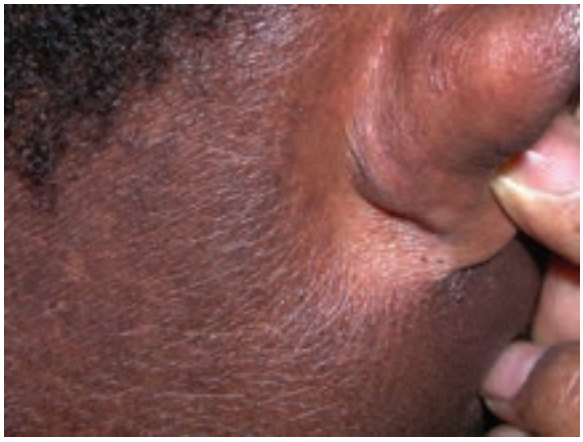

FIGURE 438-11. Hyperpigmentation and lichenification in a patient with chronic actinic dermatitis. Note sparing of the sun-protected postauricular area.

appropriate clothing, and wearing a wide-brimmed hat. Other treatment modalities, in approximate sequential order, are topical corticosteroids (fluocinonide ointment 0.05% twice daily), tacrolimus ointment (0.1% twice daily), oral mycophenolate mofetil (1 to 2 g/day), oral cyclosporine (3 to 5 mg/kg/day), and azathioprine (up to 2 to 2.5 mg/kg/day). Treatment with oral corticosteroids (e.g., prednisone, 1 mg/kg/day) may be needed for acute flare. In recalcitrant cases, the following may be used: PUVA in conjunction with oral corticosteroids or a combination of mycophenolate mofetil, PUVA, and oral corticosteroids.

Phototoxicity and Photoallergy

The terms *phototoxicity* and *photoallergy* refer to the development of skin lesions after combined exposure to an oral or topical photosensitizer and electromagnetic radiation. Phototoxicity is a nonspecific cutaneous toxic reaction, whereas photoallergy is a delayed hypersensitivity response. For most photosensitizers, the action spectrum for both lies in the UVA range (Table 438-3).

Porphyrias

The most common cutaneous porphyria is porphyria cutanea tarda, in which patients present with skin fragility and blister formation on sun-exposed areas, most commonly the dorsum of the hands and the forearms (Fig. 438-12; Chapter 210). Patients usually have periorbital hypertrichosis and, less frequently, periorbital mottled hyperpigmentation and hypopigmentation. Sclerodermoid skin changes can occur in both sun-exposed and sun-protected areas. The defective enzyme is uroporphyrinogen decarboxylase. Porphyria cutanea tarda is associated with excessive alcohol intake, exposure to estrogens, hepatitis C infection (Chapter 149), HIV infection (Chapter 392), and hemochromatosis (Chapter 212). Patients invariably have an elevated level of ferritin and frequently have elevated liver enzyme values.

The diagnosis is suggested by the typical clinical appearance and is confirmed by the characteristic porphyrin profile (elevated levels of 8-, 7-, 6-, 5-, and 4-carboxyl porphyrins in the urine and isocoproporphyrin in feces; Chapter 210). Management consists of avoidance of precipitating factors (alcohol, iron-containing vitamins, estrogen-containing birth control pills) and weekly phlebotomy. In patients who are anemic (e.g., those with HIV infection), low-dose hydroxychloroquine (200 mg/week) is beneficial.

TABLE 438-3 PHOTOTOXICITY AND PHOTOALLERGY

FEATURES	PHOTOTOXICITY	PHOTOALLERGY
Lesions after first exposure	Yes	No
Onset	Minutes after sun exposure	Delayed (24-48 hr after sun exposure)
Common offending agents	Systemic medications	Sunscreen agents
Morphology	Vesicles, bulla, hyperpigmentation	Eczematous (erythema, scaliness)
Management	Symptomatic (topical corticosteroids, antihistamine) Removal of the offending agent	

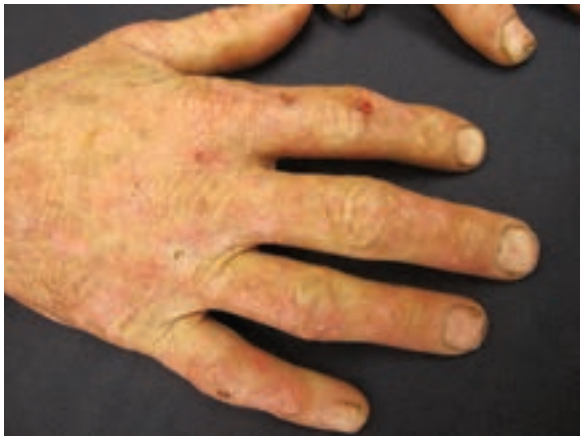

FIGURE 438-12. Erosion, crusting, and vesicles on the dorsum of the hand of a patient with porphyria cutanea tarda.

PAPULOSQUAMOUS (INCLUDING FUNGAL) DISEASES

Common papulosquamous diseases are listed in Table 438-4.

Psoriasis

EPIDEMIOLOGY

Psoriasis is the most commonly recognized papulosquamous disease. It occurs in 2 to 3% of the general population, with considerable variation in different parts of the world. It affects male and female patients equally. Approximately one third of the patients have a positive family history. Psoriasis has a bimodal peak of onset, at 22.5 years of age and again at age 55 years. The onset of psoriasis before the age of 15 years is associated with a higher prevalence of positive family history of psoriasis and with more severe disease.

PATHOBIOLOGY

Psoriasis involves the innate and adaptive immune systems, with abnormal keratinocyte proliferation. Factors playing a role in the pathogenesis include activation of antigen-presenting cells and development of T_H1 and T_H17 cells. Mediators include interleukin-12 (IL-12), IL-23, tumor necrosis factor-α (TNF-α), and interferon-γ.[2] In genetically susceptible individuals, exposure to precipitating factors such as infections (e.g., streptococcal or HIV infections), stress, or physical injury may activate T cells and stimulate the influx of neutrophils and the subsequent release of inflammatory mediators, which lead to the development of cutaneous lesions.

Psoriasis has a complex, polygenetic inheritance.[3] Cutaneous psoriasis is strongly associated with human leukocyte antigen–Cw6 (HLA-Cw6), whereas psoriatic arthritis can be associated with HLA-Cw6, HLA-B38/39, or HLA-B27.[4] Approximately 40 non–major histocompatibility complex loci have been reported by genome-wide studies to increase the risk for psoriasis, with the number expected to increase as larger cohorts are studied. Many of these loci contain genes involved in signaling pathways targeted by highly effective biologic therapies. Psoriasis also is associated with ulcerative colitis, lymphoma, the metabolic syndrome, heart disease, depression, smoking, and alcohol consumption.[5,6]

CLINICAL MANIFESTATIONS

Psoriasis can involve the skin, scalp, and nails. Skin lesions are characterized by erythematous macules, papules, or plaques that are usually covered with silvery scales (Fig. 438-13). On removal of the scales, pinpoint bleeding may be observed (the Auspitz sign), a finding reflecting the proliferation of blood vessels in the superficial dermis. Nail involvement includes pittings, yellowish macules underneath the nail plate ("oil drop" sign), and thickening of the nail (onychodystrophy) (Fig. 438-14). Minor injury to the skin can result in the development of psoriatic lesions (Koebner phenomenon). The association of psoriasis with HIV infection (Chapter 392) has been well documented.

Several distinct forms of psoriasis are recognized. *Psoriasis vulgaris,* the most common type, appears as a persistent erythematous scaly papule and plaque most commonly on elbows, knees, and scalp, where it can be associated with erythematous patches with pustules (Fig. 438-15). *Guttate psoriasis* usually occurs after viral or bacterial (most commonly streptococcal) infection; it appears as small, erythematous, scaly papules scattered over a large area of the body in a raindrop distribution (*guttate* means "droplike"). *Inverse psoriasis* refers to psoriasis that occurs in skin-fold areas such as the groin, axilla, and inframammary folds. It appears as an erythematous, somewhat shiny patch; because of the constant friction in the involved areas, scales are usually absent. *Erythrodermic psoriasis* appears as widespread erythroderma

TABLE 438-4 PAPULOSQUAMOUS DISEASES

Psoriasis
Pityriasis rubra pilaris
Pityriasis rosea
Lichen planus
Lichen nitidus
Secondary syphilis
Pityriasis lichenoides
Parapsoriasis
Mycosis fungoides
Acrokeratosis paraneoplastica of Bazex
Necrolytic acral erythema
Dermatophytosis
Tinea versicolor

with fine silvery scales. *Palmoplantar psoriasis* manifests as keratotic scaly patches and plaques on the palms and soles, very frequently with accompanying fissures. *Pustular psoriasis of von Zumbusch* is a rare variant of psoriasis occurring with generalized pustules that are 2 to 3 mm in diameter and associated with the onset of fever.

Of patients with psoriasis, 5 to 30% also may have psoriatic arthritis, which may precede the appearance of cutaneous lesions (Chapter 265). Approximately 95% of these patients present with peripheral asymmetrical oligoarthritis involving the interphalangeal joints of the hands and feet, whereas 5% have exclusively axial or skeletal disease.

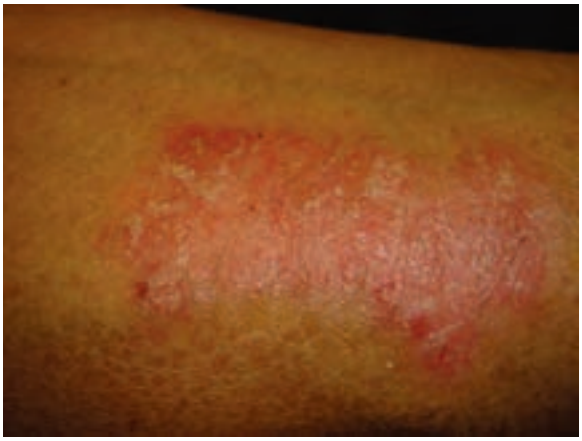

FIGURE 438-13. Psoriasis. Erythematous plaques with silvery scales.

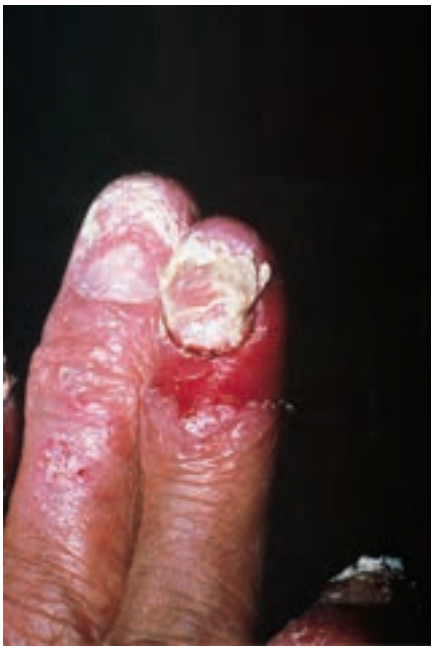

FIGURE 438-14. Psoriasis. Thickening and crumbling of the nail plate (onychodystrophy). Note the erythematous patches with silvery scales in the periungual area.

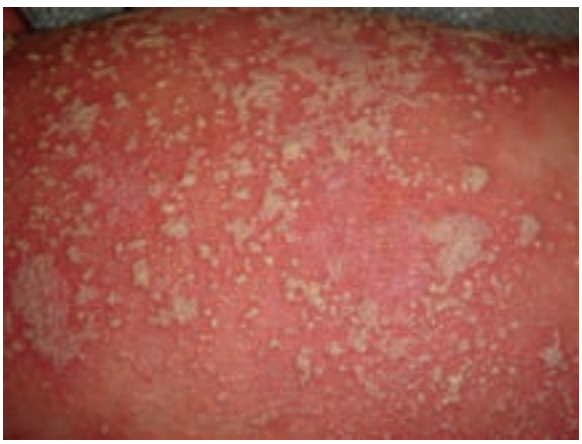

FIGURE 438-15. Psoriasis. Erythematous patch with pustules in a patient with active disease.

DIAGNOSIS

In most cases, the diagnosis of psoriasis can be made based on the history and physical examination alone. However, in patients with erythrodermic psoriasis, skin biopsy is needed to exclude other causes of generalized erythroderma, such as drug eruption, cutaneous T-cell lymphoma (Chapter 185), and pityriasis rubra pilaris.

TREATMENT

Rx

Sequential treatment modalities (Table 438-5) start with topical therapy, UV-based therapy, and then traditional systemic therapy, biologics, or an oral phosphodiesterase inhibitor.[7] For recalcitrant cases, combination therapy is frequently used. Oral corticosteroids should not be used because psoriasis may worsen when they are discontinued.

Pityriasis Rubra Pilaris

EPIDEMIOLOGY

Pityriasis rubra pilaris occurs equally in men and women; the incidence ranges from 1 in 5000 new dermatology patients in Great Britain to 1 in 50,000 in India. This disease most frequently occurs as the acquired form, although a familial form (autosomal dominant with variable expression) occasionally has been reported. Abnormal vitamin A metabolism and autoimmunity have been postulated as possible precipitants.

CLINICAL MANIFESTATIONS

The most common form of pityriasis rubra pilaris is type I, which is characterized by widespread salmon-colored plaques with fine scales, islands of sparing, scaliness on the scalp, waxy keratoderma of the palms and soles, and follicular hyperkeratosis (Figs. 438-16 and 438-17). In adult patients, the condition typically starts on the face and moves to the lower extremities; in the juvenile form, it usually starts in the lower half of the body. Ectropion and pruritus may occur. Atypical adult pityriasis rubra pilaris can manifest with palmar plantar keratoderma with coarse scales and alopecia.

DIAGNOSIS

Diagnosis is based on the clinical presentations and by the characteristic histologic findings of alternating vertical and horizontal parakeratosis in the stratum corneum.

TABLE 438-5 SEQUENTIAL THERAPEUTIC APPROACH IN PSORIASIS

Topical agents	Corticosteroids (e.g., triamcinolone ointment 0.1%) Vitamin D analogues (e.g., calcinotriene cream 0.005%) Retinoids (e.g., tazarotene cream 0.1%)
Phototherapy	Narrow band ultraviolet B (three times/wk) PUVA (psoralen and ultraviolet A, three times/wk)
Traditional systemic therapy	Methotrexate (10-20 mg/wk) Cyclosporine (3-5 mg/kg/day) Mycophenolate mofetil (1.5-2 g/day) Acitretin (25-50 mg/day)
Biologics	TNF-α inhibitors • Etanercept (50 mg/wk SC) • Adalimumab (40 mg every 2 wk SC) • Infliximab (5-10 mg/kg every 8 wk IV) Anti–IL-12/23 • Ustekinumab (45-90 mg every 12 wk SC)
Oral phosphodiesterase-4 inhibitor	Apremilast (30 mg twice/day)
Potential future treatments*	Anti–IL-17 receptor antibody • Brodalumab Anti–IL-17 antibody • Secukinumab • Ixekizumab Oral Janus kinase inhibitor • Tofacitinib

IL = interleukin; IV = intravenously; SC = subcutaneously; TNF = tumor necrosis factor.
*Doses may change if and when FDA approved.

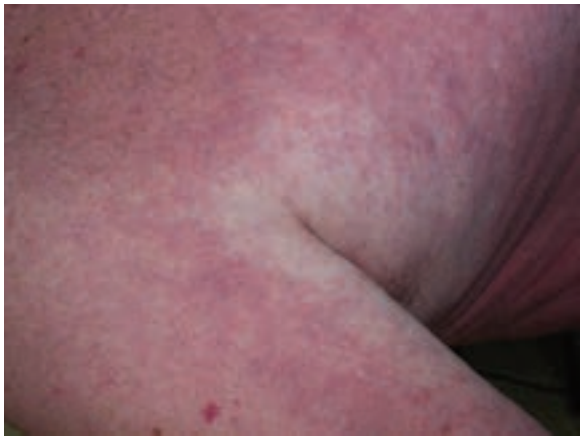

FIGURE 438-16. Pityriasis rubra pilaris. Note the erythematous orange plaques with islands of sparing.

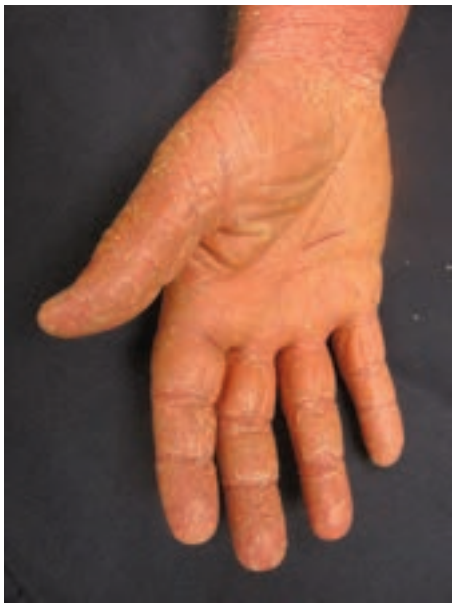

FIGURE 438-17. Pityriasis rubra pilaris. Palmar hyperkeratosis with waxy scales.

TREATMENT Rx

The most effective treatment is with oral retinoids (acitretin, 25 to 50 mg/day for 2 to 4 months). Some patients benefit from methotrexate (7.5 to 15 mg/week) or cyclosporine (3 to 5 mg/kg/day). TNF antagonists, used at the same doses as for psoriasis, are helpful for patients with recalcitrant type I disease. Topical keratolytic agents, such as ammonium lactate lotion 12%, twice daily, are helpful as adjunctive therapy.

Pityriasis Rosea

EPIDEMIOLOGY AND PATHOBIOLOGY

The incidence of pityriasis rosea has been reported as 3 to 30 per 1000 patients. It occurs in all ethnic groups, most commonly in the third and fourth decades of life, with a slight female predominance. The cause is not known, and a possible association with human herpesvirus types 6 and 7 has been reported.[8]

CLINICAL MANIFESTATIONS

In 50 to 90% of patients, pityriasis rosea starts with a primary lesion (herald patch), which is an erythematous, scaly, oval patch a few centimeters in diameter (Fig. 438-18). This lesion is usually followed within a few days by smaller, minimally pruritic, erythematous scaly patches on the trunk, less commonly on the proximal extremities. As a rule, the palms and soles are spared. The distribution of the eruption, especially on the back, tends to follow the lines of cleavage of the skin, with a resulting "Christmas tree" distribution. The eruption is self-limited and resolves within 6 to 8 weeks. In rare instances, lesions may persist.

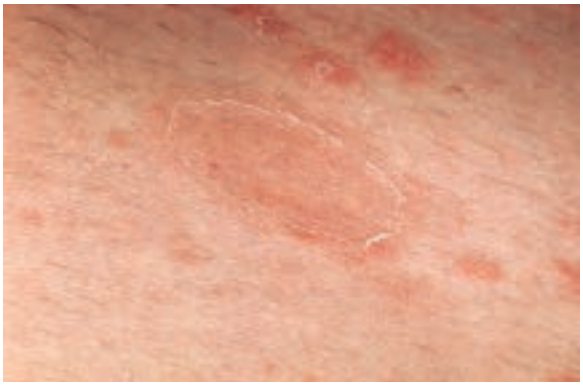

FIGURE 438-18. Pityriasis rosea. Large erythematous oval patch (herald patch) accompanied by smaller erythematous patches.

DIAGNOSIS

The diagnosis usually can be made clinically. The most important differential diagnosis is secondary syphilis, which, in contrast to pityriasis rosea, usually involves the palms and soles. Serology testing to exclude syphilis (Chapter 319) is advisable.

TREATMENT Rx

Treatment is primarily symptomatic, including topical corticosteroids and oral antihistamines. NB-UVB phototherapy should be reserved for severe, recalcitrant cases.

Lichen Planus

EPIDEMIOLOGY

Lichen planus occurs most commonly in patients between 30 and 60 years of age. Women are affected more frequently than men and tend to be somewhat older at the onset of disease. The prevalence in the general population is approximately 1%.

PATHOBIOLOGY

Histologically, lichen planus is characterized by dense lymphocytic infiltrate at the dermal-epidermal junction. The infiltrates consist of predominantly T cells, a finding suggesting the pathogenic role of cell-mediated immunity. Because lichen planus or lichen planus–like eruptions can occur after exposure to drugs or chemicals (e.g., color film developer), the role of drugs and chemicals in inducing a T-cell–mediated response against the epidermis has been postulated. Lichen planus may be associated with hepatitis C infection (Chapter 149).

CLINICAL MANIFESTATIONS

Patients present with erythematous-to-violaceous flat-topped papules, often with white lacy lines (Wickham striae) on the wrists, forearms, and genitalia (Fig. 438-19).[9] *Oral lichen planus* occurs as white papules and plaques with a reticulated appearance, most commonly along the bite line on the buccal mucosa. Similar lesions can be seen on the tongue (Fig. 438-20) and genital mucosa. Painful erosion may occur. *Hypertrophic lichen planus* usually occurs on the lower extremities as lichenified, violaceous plaques, probably secondary to chronic rubbing and scratching of the lichen planus lesions.

DIAGNOSIS

The diagnosis is made clinically and is confirmed by the characteristic histologic findings of infiltration by the lichenoid infiltrate of lymphocytes at the dermal-epidermal junction.

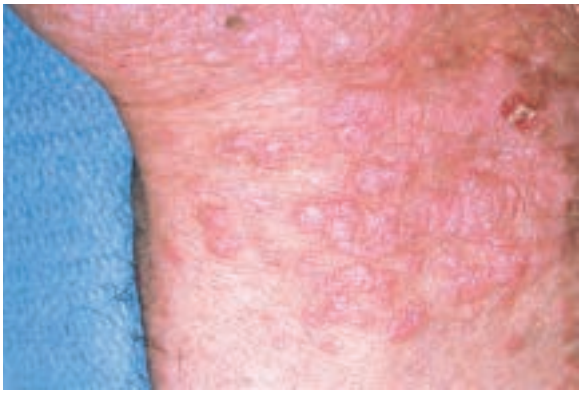

FIGURE 438-19. Lichen planus. Erythematous flat-topped papules on the wrist.

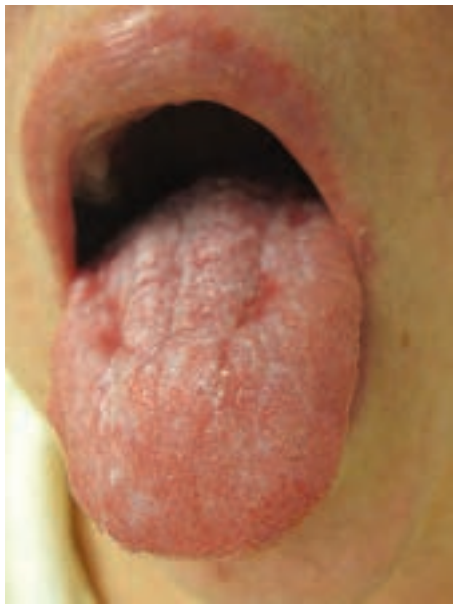

FIGURE 438-20. Lichen planus of the tongue. Note white plaques on dorsal surface of the tongue, with reticulated white line on the distal aspect of the tongue.

TABLE 438-6 THERAPEUTIC OPTIONS FOR LICHEN PLANUS

CUTANEOUS LESIONS
Topical corticosteroids (triamcinolone ointment 0.1% twice/day)
HYPERTROPHIC LESIONS
Intralesional corticosteroids (triamcinolone suspension 3-5 mg/mL)
ORAL LESIONS
Corticosteroid paste (triamcinolone 0.1% paste twice/day) or cyclosporine solution (100 mg/mL, 2 mL, twice/day, swish and spit).
GENERALIZED LICHEN PLANUS, PAINFUL ORAL/GENITAL EROSIONS, RECALCITRANT DISEASE
Narrow band UVB phototherapy (2-3 times/wk) Oral prednisone (0.5-1 mg/kg, taper in 6-8 wk) Mycophenolate mofetil (1-2 g/day) Cyclosporin (3-5 mg/kg) Tumor necrosis factor-α inhibitors (see Table 438-5)

TREATMENT

Rx

Therapeutic options depend on the location of lesions (Table 438-6). Without treatment, cutaneous lesions usually resolve in approximately 1 year, whereas oral and hypertrophic lesions tend to be much more chronic, persisting for an average of 4.5 years and 8.5 years, respectively.

FIGURE 438-21. Lichen nitidus. Note skin-colored fine papules on the upper back.

Lichen Nitidus

Lichen nitidus is a rather uncommon condition that usually occurs in children or young adults. The incidence has been estimated to be 3.4 cases per 10,000 persons. It is more commonly observed in dark-skinned individuals. The cause is unclear.

The lesions are asymptomatic, 1- to 2-mm, shiny, skin-colored discrete papules, sometimes with fine scales on their surface, occurring most commonly on the genitalia or forearms and occasionally on the trunk (Fig. 438-21). A generalized form has rarely been reported. Histologically, a dense lymphocytic infiltrate can be seen in the superficial dermis and at the dermal-epidermal junction. In contrast to lichen planus, in which the infiltrate tends to involve the entire dermal-epidermal junction, the infiltrate in lichen nitidus tends to be much more focal.

The diagnosis can be confirmed from the typical clinical appearance and the characteristic histologic changes. The condition tends to remit spontaneously in a few years, so therapy, with topical corticosteroids (e.g., triamcinolone ointment, 0.1% twice daily for 2 weeks) and oral antihistamines (e.g., fexofenadine, 180 mg every morning, and hydroxyzine, 25 to 50 mg at bedtime, as needed), should be reserved for symptomatic cases only.

Secondary Syphilis

Lesions typically occur 1 to 2 months after the development of a primary chancre lesion (Chapter 319). However, up to 25% of patients may not remember having a chancre. Once the eruption occurs, it lasts for 1 to 3 months.

Clinically, secondary syphilis may appear as erythematous macules (roseola syphilitica), erythematous-to-hyperpigmented oval or circular papules and plaques covered with scales, or a maculopapular eruption (Fig. 438-22). Nodular eruption also may occur occasionally. The lesions tend to be widespread, and the palms and soles are very frequently involved (Fig. 438-23). The diagnosis is made based on the history, physical examination, and a positive serology. Skin biopsy shows the proliferation of endothelial cells in the dermis and a dense dermal infiltrate containing many plasma cells. Intramuscular benzathine penicillin G (2.4 million U intramuscularly in a single dose) is currently the recommended treatment.

Pityriasis Lichenoides

Pityriasis lichenoides occurs as erythematous papules that may be minimally pruritic and covered with scales, scattered on all parts of the body. In the acute form (pityriasis lichenoides et varioliformis acuta [PLEVA]), the central part of the lesions develops vesicles, pustules, and hemorrhages, with eventual crusting of the lesions. The patient may have mild constitutional symptoms of fever and malaise. The chronic form (pityriasis lichenoides chronica [PLC]) occurs as asymptomatic erythematous-to-hyperpigmented papules and plaques covered with fine scales; the trunk and extremities are common

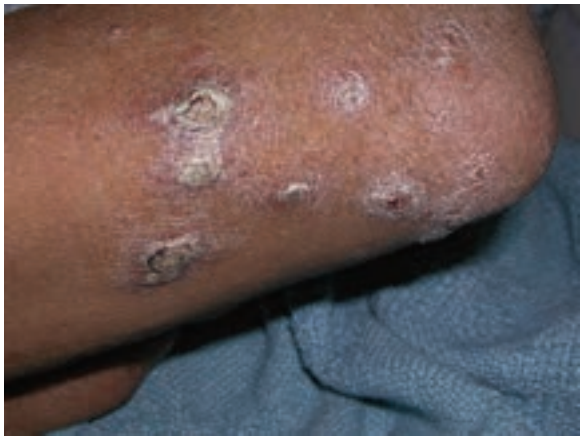

FIGURE 438-22. Secondary syphilis. Papules with crust on elbow.

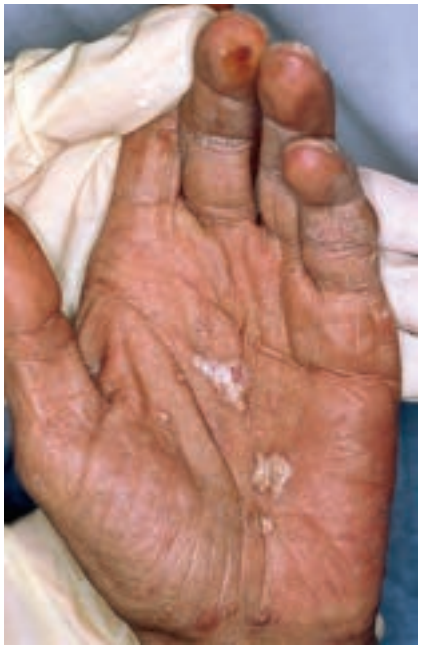

FIGURE 438-23. Secondary syphilis. Scaly papules and plaques on the palm.

sites. Histologically, both PLEVA and PLC are characterized by dense lymphocytic infiltrates in the dermis, with CD8 lymphocytes predominating in PLEVA and CD4 lymphocytes in PLC.

PLEVA usually resolves in a few months, although it can persist. PLC usually lasts for a few years. Both disorders affect patients of all ages, with a slight male predominance.

Treatment generally follows a sequential order: (1) topical corticosteroids (e.g., triamcinolone ointment, 0.1% twice daily for 1 to 2 weeks) and antihistamines, (2) doxycycline (100 mg twice daily) or erythromycin (1 to 2 g/day), (3) NB-UVB phototherapy (three times weekly for 8 to 10 weeks with increasing doses of NB-UVB), and (4) methotrexate (7.5 to 15 mg/week).

Parapsoriasis

The two common variants of parapsoriasis are large plaque parapsoriasis and small plaque parapsoriasis. The peak incidence is in the fifth decade, although rare cases may begin in childhood. Large plaque parapsoriasis appears as minimally pruritic, oval-to-circular, erythematous-to-hyperpigmented macules and patches with fine scales and superficial atrophy (crinkling atrophy) scattered on all parts of the body (Fig. 438-24). These lesions are usually larger than 5 cm. Large plaque parapsoriasis is considered by some to be a less aggressive variant of mycosis fungoides (see later). Small plaque parapsoriasis appears as circular-to-oval, erythematous-to-hyperpigmented

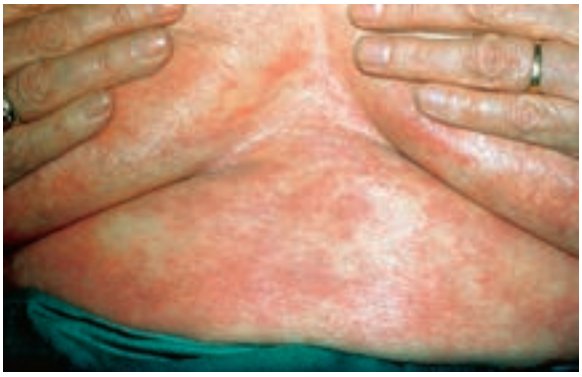

FIGURE 438-24. Large plaque parapsoriasis. Erythematous patches with fine scales.

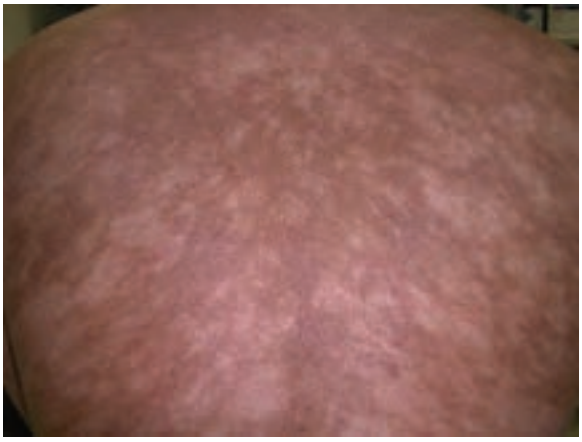

FIGURE 438-25. Mycosis fungoides. Hypopigmented patches in a dark-skinned patient.

patches or minimally elevated plaques, with lesions smaller than 5 cm in diameter and usually covered with fine scales. Digitate dermatosis is a distinct variant of small plaque parapsoriasis in which lesions appear along the lines of cleavage, usually on the lateral aspect of the trunk in the shape of fingerprints. Histologically, large plaque parapsoriasis is characterized by a dermal lymphocytic infiltrate, which may extend into the epidermis whereas small plaque parapsoriasis is characterized by spongiotic dermatitis, with a mild superficial lymphocytic infiltrate in the dermis. In up to one third of patients, large plaque parapsoriasis may evolve into mycosis fungoides. As a result, treatment of large plaque parapsoriasis is similar to that of early-stage mycosis fungoides: high-potency topical corticosteroids, topical nitrogen mustard, NB-UVB phototherapy, and psoralen and UVA (PUVA). By comparison, patients with small plaque parapsoriasis have a benign course and management should be symptomatic only, with emollients, topical corticosteroids, and NB-UVB phototherapy.

Mycosis Fungoides

Mycosis fungoides is the most common variant of cutaneous T-cell lymphoma (Chapter 185). The four types of cutaneous manifestations are patch, plaque, tumor, and erythrodermic. Patch-stage disease manifests as skin-colored or minimally erythematous patches with fine "cigarette paper" wrinkling of the epidermis; hyperpigmented or hypopigmented lesions are frequently seen in dark-skinned patients. The patches can vary from a few millimeters to a few centimeters in diameter; they are more common on sun-protected areas such as the buttocks (Fig. 438-25). The patches are usually asymptomatic, although they occasionally may be mildly pruritic. Lesions may be present for years. As the disease progresses, some of the patches may become more indurated and may evolve into plaques (Fig. 438-26). Nodular lesions may occur in patients without any patch or plaque lesions, although more commonly these lesions occur in conjunction with patches and plaques. Erythrodermic mycosis fungoides occurs as a generalized erythroderma with significant scaling and pruritus. Hyperkeratosis of the palms and soles, as well as fissuring of hands and feet, are quite common.

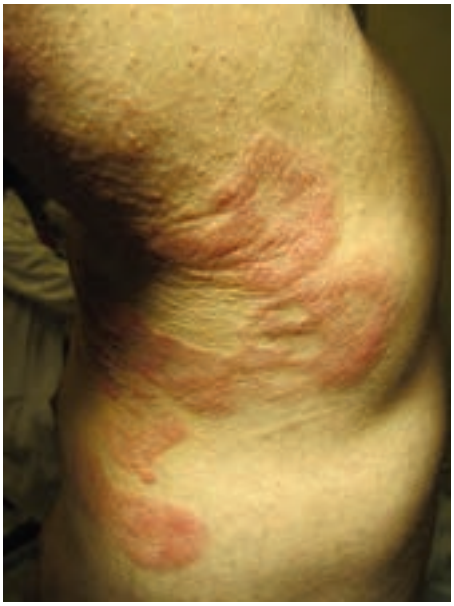

FIGURE 438-26. Mycosis fungoides. Plaque-stage disease.

TABLE 438-7 TREATMENT FOR MYCOSIS FUNGOIDES

CLINICAL TYPE	TREATMENT
Patch and localized plaque	Topical corticosteroids (e.g., triamcinolone ointment 0.1% q12h) Topical nitrogen mustard (mechlorethamine 0.01% in aquaphor, qd) Topical retinoids (e.g., bexarotene gel 1%, 1-4 times/day) Narrow band UVB (2-3 times/wk)
Extensive plaques and Tumors	Psoralen and UVA (PUVA; 2-3 times/wk) Oral bexarotene (300 mg/m^2/day) Methotrexate (10-15 mg/wk) Interferon-α (1-5 million U 3-5 times/wk SC) Total skin electron beam therapy Histone deacetylase inhibitors: Vorinostat (400 mg/day PO); romidepsin (14 mg/m^2 IV on days 1, 8, and 15 of a 28-day cycle) Anti-CD52 antibody: Alemtuzumab (10-15 mg three times/wk SC) Denileukin diftitox (9 or 18 µg/kg/day IV for 5 consecutive days every 21 days for 8 cycles Radiation therapy for localized tumors
Erythrodermic	Extracorporeal photopheresis (2 consecutive days every 2-4 wk)

IV = intravenously; PO = orally; qd = daily; UVA = ultraviolet A; UVB = ultraviolet B.

The diagnosis is confirmed by histologic demonstration of atypical mononuclear cells both in the epidermis and in the dermis, as well as immunophenotypic markers showing predominance of CD4 cells in the infiltrate.[10] Treatment options are summarized in Table 438-7.

Bazex Syndrome and Necrolytic Acral Erythema

Patients with Bazex syndrome (acrokeratosis neoplastica) present with symmetrical, scaly, erythematous-to-violaceous hyperkeratotic plaques on acral areas, such as digits, palms, soles, nose, and ears. Almost all have involvement of the ears and ridging of the nails.[11] Bazex syndrome is associated with malignancy, especially of lips, tongue, larynx, pharynx, and esophagus, perhaps because of cross reactivity between tumor antigens and normal keratinocytic antigens.

Necrolytic acral erythema is a marker of chronic hepatitis C infection (Chapter 149).[12] It manifests with well-defined hyperkeratotic, lichenified plaques on dorsum of hands and feet (Fig 438-27). Low serum zinc levels have been reported in some patients whose disease improved following oral zinc therapy.

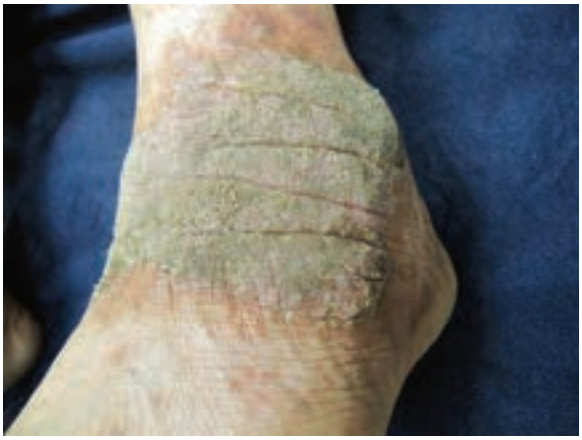

FIGURE 438-27. Necrolytic acral erythema. Lichenified plaques with fine scales on anterior lateral ankle.

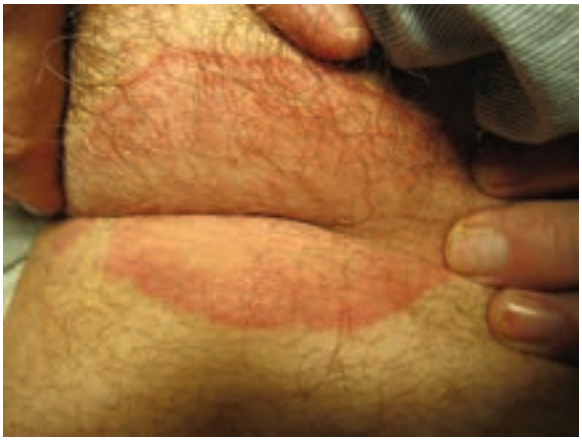

FIGURE 438-28. Tinea cruris. Erythematous patch with erythematous papules and scales at the periphery.

Dermatophytoses

Fungal infections that occur as papulosquamous eruptions include tinea corporis, tinea manuum, tinea cruris, and tinea pedis. *Tinea corporis* manifests as a polycyclic erythematous scaly patch that has elevated borders and consists of papules and sometimes pustules. As the lesion progresses, the border advances centrifugally. The trunk is the most common site. *Tinea cruris* has similar morphology, except it is located in the inguinal folds (Fig. 438-28). *Tinea manuum* presents as an erythematous scaly patch with an advancing active border, usually located on the dorsum of the hands, or it may occur as diffuse scaly patches with mild hyperkeratosis involving part or the entire surface of the palm and palmar aspect of the fingers. *Tinea pedis* has two clinical manifestations: it can occur as scaly macerated lesions with erythema in the toe webs or as patchy or diffuse scaliness on the sole extending to the medial and lateral aspect of the foot (moccasin distribution). The latter presentation can be associated with diffuse scaliness of one but not both palms, a condition known as the "one-hand, two-feet syndrome." The diagnosis can be confirmed by examination of skin scrapings using 10% potassium hydroxide preparation or by fungal culture. Treatment consists of topical or oral antifungal medications (e.g., clotrimazole cream, 1%, twice daily for 2 to 4 weeks, or terbinafine, 250 mg for 2 to 12 weeks), depending on the site involved.

TINEA VERSICOLOR

Tinea versicolor is a fungal infection of the skin caused by *Malassezia furfur.* It occurs in otherwise healthy young individuals, especially in warm and moist environments during the summer. The prevalence is estimated to be 2 to 8% in the United States and up to 50% in the tropical countries. Clinically, it appears as macules and patches with very fine scales; the color can be hypopigmented, skin colored, minimally erythematous, or light brown (Fig. 438-29). The patches start as perifollicular macules, with the midchest and midback the most common sites. As the lesions progress, hypopigmentation of the skin also may occur. The lesion usually is asymptomatic. The diagnosis

FIGURE 438-29. Tinea versicolor. Hypopigmented patches on the trunk.

is confirmed by the characteristic appearance of the fungal elements on a 10% potassium hydroxide preparation—grapelike clusters of yeast and short, septate branching hyphae ("spaghetti and meatballs" appearance).[13] Treatment is with 2.5% selenium sulfide shampoo (applied for 10 minutes then wash off, 5 times weekly for 4 to 6 weeks), topical antifungal preparations (e.g., clotrimazole cream, 1% twice daily for 4 weeks), or a 1- to 3-day course of oral ketoconazole (200 mg/day).

FIGURATE ERYTHEMIAS

The figurate erythemas (which include erythema annulare centrifugum, erythema gyratum repens, and erythema chronicum migrans) appear as erythematous circular or polycyclic plaques with central clearing and, frequently, a centrifugally migrating border. Occasionally, fine scaling also may be observed. The extremities are the most common sites. The diagnosis frequently can be made by the typical history and morphologic features.

Erythema annulare centrifugum is most commonly idiopathic; however, it also can be a manifestation of a hypersensitivity response to medications. Management includes identification of a precipitating agent (if possible) and treatment with topical or systemic corticosteroids. *Erythema gyratum repens* occurs as concentric erythematous plaques with fine scales, resembling a wood-grain pattern. This unusual form of figurate erythema has been associated with malignant hematologic diseases and with carcinomas of the breast, lung, gastrointestinal tract, prostate, and cervix. Treatment of the underlying malignant disease results in the resolution of the skin lesion in a few months. *Erythema chronicum migrans,* which is a cutaneous manifestation of Lyme disease and is caused by the spirochete *Borrelia burgdorferi* (Chapter 321), appears as a concentric ring of erythema that progresses centrifugally from the site of a tick bite. Occasionally, it may appear as a circular erythematous patch. The diagnosis is made by a history of a tick bite, the characteristic cutaneous lesion, or elevated serum antibodies to *B. burgdorferi*. Management is the same as for Lyme disease.

Grade A References

A1. Paller AS, Lebwohl M, Fleischer AB Jr, et al. Tacrolimus ointment is more effective than pimecrolimus cream with a similar safety profile in the treatment of atopic dermatitis: results from 3 randomized, comparative studies. *J Am Acad Dermatol*. 2005;52:810-822.

A2. Beck LA, Thaci D, Hamilton JD, et al. Dupilumab treatment in adults with moderate-to-severe atopic dermatitis. *N Engl J Med*. 2014;371:130-139.

A3. Ohtsuki M, Terui T, Ozawa A, et al. Japanese guidance for use of biologics for psoriasis (the 2013 version). *J Dermatol*. 2013;40:683-695.

A4. McInnes IB, Kavanaugh A, Gottlieb AB, et al. Efficacy and safety of ustekinumab in patients with active psoriatic arthritis: 1 year results of the phase 3, multicentre, double-blind, placebo-controlled PSUMMIT 1 trial. *Lancet*. 2013;382:780-789.

A5. Papp K, Cather JC, Rosoph L, et al. Efficacy of apremilast in the treatment of moderate to severe psoriasis: a randomised controlled trial. *Lancet*. 2012;380:738-746.

A6. Papp KA, Leonardi C, Menter A, et al. Brodalumab, an anti-interleukin-17-receptor antibody for psoriasis. *N Engl J Med*. 2012;366:1181-1189.

A7. Langley RG, Elewski BE, Lebwohl M, et al. Secukinumab in plaque psoriasis—results of two phase 3 trials. *N Engl J Med*. 2014;371:326-338.

A8. Leonardi C, Matheson R, Zachariae C, et al. Anti-interleukin-17 monoclonal antibody ixekizumab in chronic plaque psoriasis. *N Engl J Med*. 2012;366:1190-1199.

A9. Menter A, Papp KA, Tan H, et al. Efficacy of tofacitinib, an oral janus kinase inhibitor, on clinical signs of moderate-to-severe plaque psoriasis in different body regions. *J Drugs Dermatol*. 2014;13:252-256.

A10. Petrof G, Almaani N, Archer CB, et al. A systematic review of the literature on the treatment of pityriasis rubra pilaris type 1 with TNF-antagonists. *J Eur Acad Dermatol Venereol*. 2013;27:e131-e135.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Wadhwani AR, Sharma VK, Ramam M, et al. A clinical study of the spectrum of photodermatoses in dark-skinned populations. *Clin Exp Dermatol*. 2013;38:823-829.
2. Swindell WR, Xing X, Stuart PE, et al. Heterogeneity of inflammatory and cytokine networks in chronic plaque psoriasis. *PLoS ONE*. 2012;7:e34594.
3. Mahil SK, Capon F, Barker JN. Genetics of psoriasis. *Dermatol Clin*. 2015;33:1-11.
4. Eder L, Chandran V, Pellett F, et al. Differential human leucocyte allele association between psoriasis and psoriatic arthritis: a family-based association study. *Ann Rheum Dis*. 2012;71:1361-1365.
5. Armstrong AW, Harskamp CT, Armstrong EJ. Psoriasis and metabolic syndrome: a systematic review and meta-analysis of observational studies. *J Am Acad Dermatol*. 2013;68:654-662.
6. Pouplard C, Brenaut E, Horreau C, et al. Risk of cancer in psoriasis: a systematic review and meta-analysis of epidemiological studies. *J Eur Acad Dermatol Venereol*. 2013;27(suppl 3):36-46.
7. Sobell JM, Leonardi CL. Therapeutic development in psoriasis. *Semin Cutan Med Surg*. 2014;33:S69-S72.
8. Rebora A, Drago F, Broccolo F. Pityriasis rosea and herpesviruses: facts and controversies. *Clin Dermatol*. 2010;28:497-501.
9. Le Cleach L, Chosidow O. Clinical practice: lichen planus. *N Engl J Med*. 2012;366:723-732.
10. Song SX, Willemze R, Swerdlow SH, et al. Mycosis fungoides: report of the 2011 Society for Hematopathology/European Association for Haematopathology workshop. *Am J Clin Pathol*. 2013;139:466-490.
11. Shah KR, Boland CR, Patel M, et al. Cutaneous manifestations of gastrointestinal disease: part I. *J Am Acad Dermatol*. 2013;68:189-209.
12. Moneib HA, Salem SA, Darwish MM. Evaluation of zinc level in skin of patients with necrolytic acral erythema. *Br J Dermatol*. 2010;163:476-480.
13. Moriarty B, Hay R, Morris-Jones R. The diagnosis and management of tinea. *BMJ*. 2012;345:e4380.

REVIEW QUESTIONS

1. A 16-year-old high school student presented with pruritic, lichenified plaques on her antecubital and popliteal fossa and dryness of skin. The first-line treatment is:

A. Cyclosporin 100 mg twice daily.
B. Narrow band ultraviolet light B therapy.
C. Oral prednisone 40 mg/day.
D. Psoralen and UVA (PUVA) photochemotherapy.
E. Triamcinolone ointment 0.1% twice daily.

Answer: E This is a case of atopic dermatitis, for which topical corticosteroids are first-line therapy. The next steps in the therapeutic ladder are narrow band UVB phototherapy (a benign treatment, but the patient needs to go to phototherapy center 2 to 3 times per week) and PUVA (potential photocarcinogenesis with long-term treatment is a drawback, especially for a 16-year-old patient). Cyclosporine and oral prednisone should be reserved for recalcitrant cases because of their side effects.

2. A 23-year-old otherwise healthy African American woman sees you 2 days after she spent a week in Hawaii. A few days after she arrived in Hawaii, she developed minimally pruritic 1- to 2-mm papules on her extensor forearm and on the dorsum of her hands. The most likely diagnosis is:

A. Chronic actinic dermatitis.
B. Drug-induced phototoxicity.
C. Erythropoeitic protoporphyria.
D. Polymorphic light eruption.
E. Porphyria cutanea tarda.

Answer: D Pinhead popular eruption is the typical morphology of polymorphic light eruption in dark-skinned individuals. Chronic actinic dermatitis presents with chronic photosensitivity and lichenification on sun-exposed areas. Drug-induced phototoxicity can be excluded by the history. Erythropoietic protoporphyria is a chronic photosensitivity, which typically starts in childhood. Porphyria cutanea tarda manifests with skin fragility, blisters, and erosions on a sun-exposed area.

3. A 55-year-old man with a long history of psoriasis presents with a 2-week history of worsening skin lesions and generalized erythroderma. The single most appropriate treatment is:

A. Calcipotriene cream 0.005% twice daily.
B. Cyclosporin 100 mg twice daily.
C. Tazarotene cream 0.1% twice daily.
D. Triamcinolone ointment 0.1% daily.
E. Prednisone 60 mg/day.

Answer: B Topical treatments (calcipotriene cream, tazarotene cream, triamcinolone ointment) are not sufficient to suppress active disease in rapidly progressing, erythrodermic psoriasis. Oral cyclosporine is a fast-acting medication with a well-documented efficacy for this condition. Oral prednisone should not be used because psoriasis can flare significantly when it is tapered.

4. An otherwise healthy 38-year-old man presents with a 1-month history of generalized erythroderma with salmon-colored and fine scaling, several areas of sparing on his abdomen and upper back, ectropion and keratoderma of palms and soles. The most likely diagnosis is:

A. Erythrodermic mycosis fungoides.
B. Pityriasis lichenoides chronica.
C. Pityriasis rosea.
D. Pityriasis rubra pilaris.
E. Erythrodermic psoriasis.

Answer: D Salmon-colored erythroderma, "islands of sparing," and keratoderma of the palms and soles are the characteristic morphology of pityriasis rubra pilaris. Erythrodermic mycosis fungoides manifests with reddish-to-violaceous erythroderma; it does not typically have islands of sparing. Pityriasis lichenoides chronica manifests as scattered erythematous papules and plaques; it does not manifest as generalized erythroderma. Pityriasis rosea manifests with minimally erythematous patches with fine scales. Patients with erythrodermic psoriasis usually have areas of typical psoriasis (papules and plaques with micaceous scales on elbows, knees, scalp); they do not have islands of sparing.

5. A 66-year-old man presents with a 1-year history of several asymptomatic patches, 5 to 6 cm in diameter, epidermal atrophy on his buttocks, and several semicircular erythematous plaques with no scales on his arms and thighs. Biopsy of a plaque on his thigh will most like show:

A. Atypical mononuclear cells in the epidermis, with a CD4 predominant infiltrate.
B. Dense infiltrate of lymphocytes at dermal-epidermal junction (lichenoid infiltrate).
C. Dermal-epidermal separation.
D. Edema in the epidermis (spongiosis) and in the dermis.
E. Necrotic keratinocytes in the epidermis.

Answer: A The other biopsy findings are typical of lichen planus (B), blistering skin diseases (C), dyshidrosis (D), and erythema multiforme (E).

439

MACULAR, PAPULAR, VESICULOBULLOUS, AND PUSTULAR DISEASES

NEIL J. KORMAN

MACULAR AND PAPULAR EXANTHEMS

An exanthem is an acute generalized eruption of the skin, and there are two major types, scarlatiniform eruptions and morbilliform eruptions. Scarlatiniform eruptions consist of confluent blanching erythema; their name was derived from their similarity to the eruption of scarlet fever (Table 439-1). Morbilliform eruptions consist of erythematous macules and papules; they are named for their resemblance to the measles eruption. Morbilliform eruptions can be caused by exposure to medications (Chapter 440) or viral infections.

Scarlatiniform Eruptions

SCARLET FEVER

Scarlet fever is caused by infection of the ears, nose, throat, and skin with toxin-producing β-hemolytic streptococci (Chapter 290). It most commonly occurs in children after streptococcal wound infections, burns, and upper respiratory tract infections. Occasional cases of scarlet fever can also be caused by infection with *Staphylococcus aureus* (Chapter 288), *Haemophilus influenzae* (Chapter 300), and *Clostridium* spp. (Chapter 296). The rash is caused by a circulating toxin that induces local production of inflammatory mediators and alteration of cutaneous cytokines. Patients may have an abrupt onset of fever, headache, vomiting, malaise, chills, and sore throat. The mucous membranes are usually erythematous with petechiae, and the tongue commonly has a white membrane. Red, exudative tonsils are present with pharyngeal infections. The skin eruption appears after the fever and is characterized by fine erythematous papules, first on the upper part of the trunk and then in a more general distribution. The face is flushed, and circumoral pallor is seen. This eruption lasts 4 to 5 days followed by fine desquamation, the extent and duration of which are related to the severity of the eruption. Treatment is 1.2 million units of benzathine penicillin G given intramuscularly or oral penicillin VK, 1000 mg twice daily for 10 days. Most patients recover after 4 to 5 days, and the rash usually resolves completely over a period of several weeks.

TABLE 439-1 MACULAR AND PAPULAR ERUPTIONS

SCARLATINIFORM ERUPTIONS
Scarlet fever
Toxic shock syndrome
Kawasaki disease
MORBILLIFORM ERUPTIONS
Measles
Rubella
Erythema infectiosum
Roseola
PAPULAR ERUPTIONS
Molluscum contagiosum
Warts

TOXIC SHOCK SYNDROME

DEFINITION

Toxic shock syndrome is an acute febrile illness caused by toxin-producing strains of *S. aureus* (Chapter 288) or, less commonly, *Streptococcus* spp. (toxic shock–like syndrome [Chapter 290]).

PATHOBIOLOGY

Most cases of staphylococcal toxic shock syndrome or streptococcal toxic shock–like syndrome occur in young healthy persons age 20 to 50 years. These toxins cause massive release of tumor necrosis factor-α and interleukin-1, cytokines that mediate fever, rash, hypotension, tissue injury, and shock.

CLINICAL MANIFESTATIONS

The hallmarks of the disease are fever, rash (Fig. 436-12 in Chapter 436), hypotension, and involvement of multiple organs, including the lungs, kidneys, liver, and gastrointestinal tract. Desquamation of the palms and soles follows onset of the illness by 1 to 2 weeks. There is diffuse macular erythema with flexural accentuation, mucous membrane erythema, and severe conjunctival involvement. Blood cultures are positive in 5% to 15% of patients with staphylococcal toxic shock syndrome and approximately 50% of those with streptococcal toxic shock–like syndrome.

TREATMENT Rx

Treatment is supportive and includes hydration, vasopressors, appropriate antibiotics, and drainage of infected sites. Patients with staphylococcal toxic shock should be treated with intravenous (IV) vancomycin to cover methicillin-resistant staphylococci, 1 g every 12 hours for 10 to 15 days, with dose adjustment based on creatinine clearance. Patients with streptococcal toxic shock should be treated with both IV penicillin G, 3 to 4 million units every 4 hours, and IV clindamycin, 600 to 900 mg every 8 hours for 10 to 15 days, followed by oral therapy. Double antibiotic coverage is the standard of care for streptococcal toxic shock syndrome because this infection is characterized by extremely large numbers of stationary bacteria and penicillin alone is not effective in this scenario inasmuch as penicillin-binding proteins are not expressed during the stationary group phase of streptococci. Silver sulfadiazine cream may lead to increased toxin production; therefore, mupirocin ointment should be used for infected sites.

PROGNOSIS

The mortality rate in patients with staphylococcal toxic shock syndrome is 5% to 15%; that for streptococcal toxic shock–like syndrome may be five times higher.

KAWASAKI DISEASE

EPIDEMIOLOGY

Kawasaki disease, a systemic vasculitis of unknown etiology, occurs in children of all races but is up to 20 times more common in North East Asians than in whites. Although primarily an illness of children younger than 5 years, Kawasaki disease also occurs in adults. Its epidemiologic and clinical manifestations imply that an infection is the cause, but bacterial, viral, and serologic studies have yet to confirm such an etiology.

CLINICAL MANIFESTATIONS

The clinical hallmarks are fever lasting up to 2 weeks, with spikes to 40° C (104° F), and a toxic-appearing patient.[1] During the acute phase, the polymorphic eruption may be scarlatiniform, urticarial, morbilliform, or targetoid. Desquamation occurs in the perianal area 2 days after the onset of fever and on the extremities 2 to 3 weeks later. Patients often have hemorrhagic, dry fissured lips; conjunctival injection; a "strawberry tongue;" and cervical lymphadenitis. Myocarditis and coronary artery aneurysms may develop in untreated patients, so prompt diagnosis is critical. Other findings include arthralgias and arthritis, urethritis, aseptic meningitis, pneumonitis, and diarrhea.

DIAGNOSIS

Despite the lack of specific diagnostic tests for this syndrome, the typical skin eruption accompanied by myocarditis is characteristic.

TREATMENT Rx

Recommended therapy in the acute phase of Kawasaki disease is a single infusion of 2 g/kg of IV gamma globulin. When this therapy is given within 5 to 10 days after the onset of fever, 85% to 90% of patients will become afebrile within 36 hours, and the risk of developing coronary artery aneurysms is significantly reduced.■ The addition of prednisolone (2 mg/kg/day for 15 days after the C-reactive protein normalizes) further reduces the risk of developing coronary artery abnormalities.■ During the subacute and convalescent phase, acetylsalicylic acid is usually given at 3 to 8 mg/kg for 6 to 8 weeks. The optimal salicylate regimen for Kawasaki disease remains uncertain, and there are no controlled trials to prove that aspirin reduces coronary artery aneurysms.

Morbilliform Eruptions

MEASLES

Measles is caused by a paramyxovirus (Chapter 367) that infects respiratory epithelium and is highly transmissible. The incubation period is 7 to 14 days. The prodrome consists of cough, coryza, and conjunctivitis. The enanthem, or Koplik spots (Fig. 367-1 in Chapter 367), predates the exanthem by 1 to 2 days and lasts 2 to 4 days. These blue-white spots surrounded by a red halo appear on the buccal mucosa and are pathognomonic for measles. The exanthem (Fig. 367-2 in Chapter 367) begins on the fourth or fifth day as papules on the face and behind the ears; it then spreads to the trunk and extremities. Measles is diagnosed on clinical grounds. Active immunization with live attenuated virus has dramatically reduced the incidence of measles infection (Chapter 18) and is the most important preventive measure. Treatment consists of supportive care, with attention to maintaining good hydration.

RUBELLA

Rubella is an RNA virus of the Togaviridae family (Chapter 368). Infection with this virus leads to an illness involving the skin, lymph nodes, and occasionally the joints, primarily in young children. The disease is spread by nasal droplet infection and has an incubation period of 14 to 21 days. Patients are most contagious when the rash is erupting. In children, there may be no prodrome; in adults, however, fever, sore throat, and rhinitis may be present. The exanthem (Fig. 368-1 in Chapter 368) begins as pink macules and papules on the face that spread to the trunk and extremities; it lasts 1 to 3 days. Generalized tender lymphadenopathy, especially the suboccipital, postauricular, and cervical nodes, is the hallmark of rubella. In normal children and adolescents, the diagnosis is made clinically, and laboratory work is unnecessary. If the diagnosis is questioned, a rising immunoglobulin M (IgM) antibody titer over a 2-week period indicates recent infection. No treatment exists, and the disease is usually self-limited. Rest and fluids are appropriate. The best protection is vaccination given with measles and mumps vaccine (i.e., measles, mumps, and rubella) at 12 to 15 months and again at 4 to 6 years (Chapter 18).

ERYTHEMA INFECTIOSUM

Erythema infectiosum is an exanthem caused by human parvovirus B19 (Chapter 371). It has a 4- to 14-day incubation period and is spread by aerosolized respiratory droplets. Acute infection leads to the production of IgM antibodies and the formation of immune complexes, which are deposited in the skin and joints. Bright red erythema appears abruptly over the cheeks (Fig. 371-1 in Chapter 371). Within 1 to 4 days, an erythematous morbilliform eruption occurs on the extremities; it fades within several days to a reticulate pattern (Fig. 439-1). There may also be malar erythema or a reticulate eruption on the extremities. Exposure of adults to parvovirus B19 leads to an acute polyarthropathy of the hands, wrists, knees, and ankles. Parvovirus B19 can interfere with erythropoiesis (Chapter 165) and can cause aplastic crisis in patients with sickle cell disease (Chapter 163). The diagnosis is clinical, and further testing is not generally necessary. Erythema infectiosum is usually a benign, self-limited disease.

ROSEOLA

Roseola (exanthem subitum) is caused by human herpesvirus 6 (Chapter 374). Virus replication occurs in leukocytes and salivary glands. Early invasion of the central nervous system (CNS) may lead to seizures. The classic patient is a healthy 9- to 12-month-old infant with an abrupt onset of high fever (40° C [104° F]) lasting 3 days. Febrile seizures occur in 15% of cases. Its rapid defervescence is striking, with the onset of a generalized pink mor-

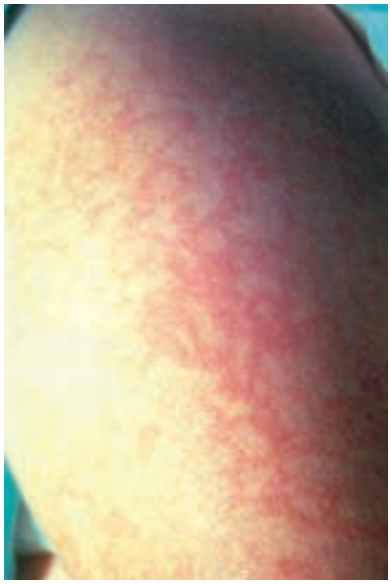

FIGURE 439-1. Reticulate macular erythema on the thigh of a patient with erythema infectiosum.

billiform exanthem. The eruption lasts 2 days and consists of pink papules or blanchable macular erythema. The lack of symptoms during the febrile phase and appearance of the exanthem as the fever subsides help with the diagnosis, but rubella and measles must also be considered. In an immunocompromised child or adult, there is usually an abrupt onset of fever, malaise, and sometimes CNS involvement. Virus isolation, seroconversion (IgM), or detection of viral DNA sequences in peripheral blood mononuclear cells can confirm the diagnosis. No antiviral therapy is available for roseola, and treatment is supportive. Practically all immunocompetent patients recover from roseola without sequelae, but chronic infection with multisystem complications may develop in immunocompromised patients.

Papular Eruptions

MOLLUSCUM CONTAGIOSUM

Molluscum contagiosum is a cutaneous infection caused by a large DNA poxvirus (Chapter 372) that affects both children and adults. Firm, smooth, umbilicated papules, usually 2 to 6 mm in diameter, are present in groups or widely disseminated on the skin and mucosal surfaces. Patients infected with human immunodeficiency virus (HIV) may have hundreds of lesions, and some lesions can be larger than 15 mm (Chapter 392). The diagnosis is made on clinical grounds. Molluscum contagiosum is self-limited, with the goal of treatment being destruction of the lesions. Commonly used treatments are all topical and include cryotherapy with liquid nitrogen, curettage, cantharidin, podophyllin, and tretinoin.

WARTS

Warts are benign proliferations of skin and mucosa caused by human papillomaviruses (HPVs) (Chapter 373). More than 150 types of HPV have been identified. Certain types of HPV occur at particular anatomic sites; however, warts of any HPV type may be found at any site. Variants include common warts, genital warts, flat warts, and deep palmoplantar warts. Common warts, known as verruca vulgaris, are hard papules that range in size from 1 mm to more than 1 cm with a rough scaly surface (Fig. 439-2) and can occur anywhere on the body. Warts are transmitted by direct contact, and disruption of the epithelial barrier is a predisposing factor. HPV subtypes 6, 11, 16, 18, 31, and 35 may be associated with malignancies (Chapter 373). Malignant transformation, although uncommon, can occur in patients with genital warts or in immunocompromised patients. Infection is confined to the epithelium and does not result in systemic viral dissemination. The diagnosis is made on clinical grounds. Viral DNA identification by Southern blot hybridization is used to identify specific HPV subtypes. All therapies are methods of physical destruction of the skin where the virus is located because there are no specific anti-HPV medications. Common topical therapies include in-office treatment with liquid nitrogen, cantharidin, or podophyllin, as well as prescription-strength tretinoin or imiquimod. Over-the-counter liquid nitrogen is neither as cold nor as effective as in-office treatment.

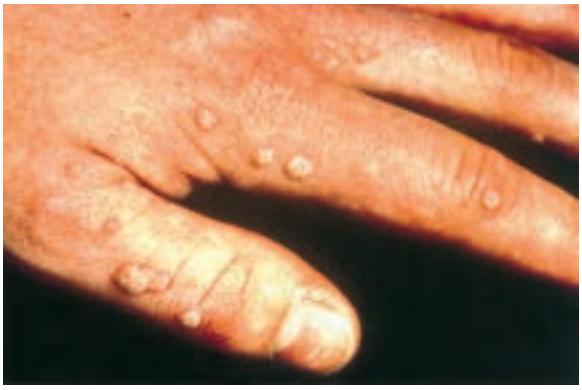

FIGURE 439-2. Hand of a patient with verruca vulgaris revealing many verrucous papules.

TABLE 439-2 PURPURIC ERUPTIONS

NONPALPABLE PURPURA
Cutaneous Disorders
Solar purpura
Steroid purpura
Pigmented purpuric dermatosis
Systemic Disorders
Idiopathic thrombocytopenic purpura
Abnormal platelet function in renal or hepatic disease
Thrombocytosis in myeloproliferative neoplasms
Clotting factor abnormalities
Ehlers-Danlos syndrome
Scurvy
Amyloidosis
Disseminated intravascular coagulation
Thrombotic thrombocytopenic purpura
Monoclonal cryoglobulinemia
Warfarin necrosis
Emboli
Cholesterol emboli
Fat emboli
Tumor emboli from atrial myxomas
Emboli from endocarditis
PALPABLE PURPURA
Vasculitis
Leukocytoclastic vasculitis
Henoch-Schönlein purpura
Urticarial vasculitis
Polyarteritis nodosa
Infectious Emboli
Meningococcemia
Gonococcemia
Rocky Mountain spotted fever
Ecthyma gangrenosum

Purpuric Eruptions

Purpura occurs when red blood cell extravasation leads to visible hemorrhage in the skin (Table 439-2). Petechiae (<3 mm) and purpura (>3 mm) may be nonpalpable or palpable. When the condition is severe, petechiae and purpuric lesions may become confluent and form ecchymoses larger than 1 cm.

NONPALPABLE PURPURA

Dermal Causes

Frequent causes of nonpalpable purpura (Fig. 436-11 in Chapter 436) include solar purpura, steroid purpura, and Schamberg disease. Solar purpura, caused by chronic sun exposure and aging, is usually found on the forearms. Steroid purpura, which is caused by prolonged use of topical or systemic steroids, can occur in any location (Chapter 35). Both conditions are caused by changes in the dermal connective tissue surrounding blood vessels. Schamberg disease, or pigmented purpuric dermatosis, is a capillaritis with yellow-brown macules and petechiae on the lower part of the legs. This capillaritis occurs as a result of red blood cell extravasation secondary to perivascular lymphocyte inflammation.

Systemic Causes

Systemic causes of nonpalpable purpura include idiopathic thrombocytopenic purpura (Chapter 172), abnormal platelet function as a result of renal or hepatic insufficiency (Chapters 130 and 153) or thrombocytosis as seen in myeloproliferative diseases (Chapter 166), and clotting factor abnormalities (Chapter 174). Fragility of the blood vessels, especially the capillaries, is found in Ehlers-Danlos syndrome (Chapter 260), scurvy (Chapter 213), and systemic amyloidosis (Chapter 188).

Thrombi

Thrombus formation within skin blood vessels also leads to purpura in patients with disseminated intravascular coagulation (DIC) (Chapter 175), thrombotic thrombocytopenic purpura (Chapter 172), monoclonal cryoglobulinemia (Chapter 187), and drug reactions to warfarin (Chapter 38). DIC may be caused by infectious agents (bacterial, particularly meningococcemia, viral, or rickettsial) and by malignancies such as leukemia (Chapter 175). Purpura fulminans is a type of DIC associated with fever and hypotension; it is usually found in children after a bacterial or viral infection. Widespread purpura and hemorrhagic bullae can be seen in DIC and purpura fulminans (Fig. 439-3). Thrombotic thrombocytopenic purpura (Chapter 172) is manifested as fever, purpura, renal failure, microangiopathic hemolytic anemia, and neurologic disease. Monoclonal cryoglobulinemia may be associated with leukemia, lymphoma, multiple myeloma, and Waldenström macroglobulinemia. Mixed cryoglobulinemia is frequently associated with hepatitis C infection (Chapter 149). Widespread purpura, along with ulcerations limited to the lower extremities or fingers and toes, can occur. Skin biopsy specimens may reveal intracapillary deposits of precipitated cryoglobulins. Disease can be worsened by exposure to cold. The vessels of the lungs, brain, and kidneys may be involved. Warfarin necrosis of the skin (Chapter 38) is an uncommon reaction that occurs between the third and tenth day of therapy and is characterized by painful erythematous to purpuric plaques in which hemorrhagic bullae develop. The most common sites include the breasts, thighs, and buttocks. The onset of disease is unrelated to the dose of warfarin, and continued warfarin therapy does not alter the course of the disease.

Emboli

Cholesterol emboli are found in the lower extremities of patients with severe atherosclerosis as a result of occlusion of small- and medium-caliber arteries by cholesterol crystals (Chapters 80 and 125). Cholesterol emboli are triggered by vascular procedures or thrombolytic therapy, but they can also occur spontaneously. Other sources of emboli that may cause petechiae or purpura include fat emboli occurring after major injury (Chapters 98 and 111), tumor emboli from atrial myxomas (Chapter 60), emboli from infective endocarditis (Chapter 76), or nonbacterial thrombotic endocarditis (Chapters 60 and 179).

PALPABLE PURPURA

Vasculitis

Palpable purpura results from inflammatory damage to cutaneous blood vessels.[2] Leukocytoclastic vasculitis, which is manifested as palpable purpura (Fig. 439-4), may be idiopathic or associated with sepsis, drug reactions, connective tissue diseases, cryoglobulinemia, hepatitis B or C infection, or underlying malignancies. Evaluation of patients with palpable purpura should always include histopathologic evaluation to confirm the diagnosis. Skin biopsy specimens from patients with leukocytoclastic vasculitis reveal angiocentric inflammation with endothelial cell swelling, fibrinoid necrosis of blood vessel walls, a neutrophilic cellular infiltrate with fragmentation of nuclei (karyorrhexis or leukocytoclasia) around and within blood vessel walls, and extravasated red blood cells (Fig. 439-5). Fresh skin biopsy specimens processed for direct immunofluorescence reveal deposits of immunoglobulins and complement in blood vessel walls. After the diagnosis is confirmed, renal and liver function testing and urinalysis should be performed in all patients, but specialized tests should be targeted to specific patients with suggestive findings. If an etiologic agent can be identified and treated, the vasculitis will often resolve. Patients with idiopathic leukocytoclastic vasculitis can be treated with oral colchicine (0.6 mg twice daily); oral dapsone (up to 200 mg once daily); or in the most severe cases, immunosuppressive agents such as mycophenolate mofetil (up to 45 mg/kg for as long as the disease is active), azathioprine (up to 2.5 mg/kg for as long as the disease is active, as guided by the thiopurine methyltransferase level), or cyclophosphamide (up to 2.5 mg/kg for as long as the disease is active).

Henoch-Schönlein purpura is an IgA mediated leukocytoclastic vasculitis that affects children and young adults; it is often preceded by an upper respiratory infection with associated fever, arthralgias, abdominal pain, and renal vasculitis (Chapter 121).

Urticarial or hypocomplementemic vasculitis (Chapter 270) is characterized by urticarial lesions that last longer than 24 hours; arthritis, facial, and laryngeal edema; and low serum complement levels. In some patients, systemic lupus erythematosus (Chapter 266) may develop. In polyarteritis nodosa (Chapter 270), vasculitis of the arterial blood vessels leads to ischemia of the skin. Skin lesions usually include ulcerated nodules and ecchymoses.

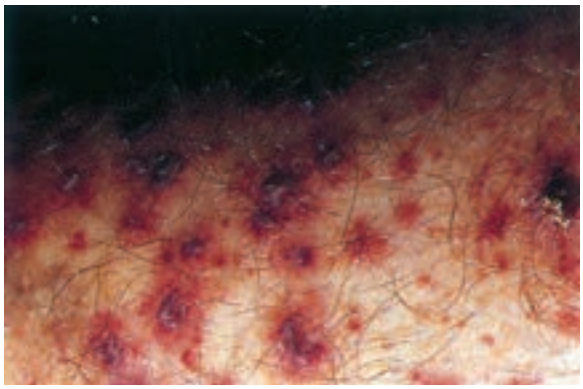

FIGURE 439-4. Palpable purpura. Leukocytoclastic vasculitis commonly causes raised purpuric and ulcerated lesions on the legs.

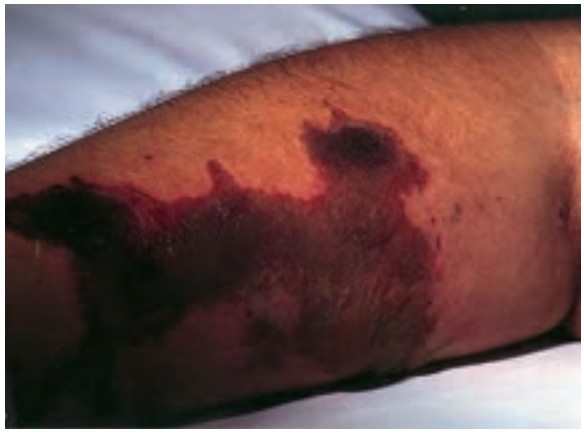

FIGURE 439-3. Purpura fulminans. Purpura and hemorrhagic blisters are seen on the arm of this patient.

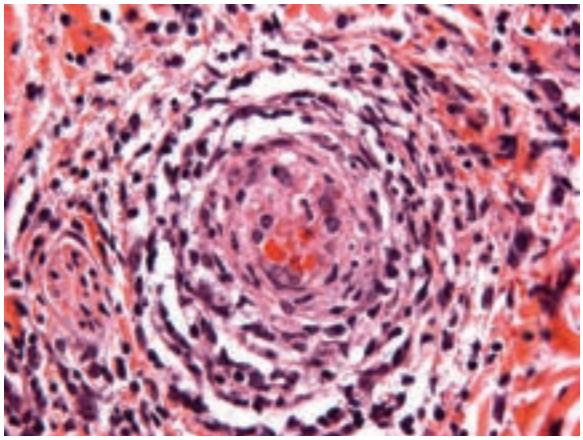

FIGURE 439-5. Leukocytoclastic vasculitis. Histologic evaluation reveals a smudged blood vessel in the dermis with neutrophils, neutrophilic dust, and red blood cells.

Cutaneous Emboli

In addition to vasculitis, cutaneous emboli can also lead to the development of palpable purpura. Infectious emboli (Chapter 76) can be caused by gram-negative cocci, gram-negative rods, *Rickettsia* spp., and in immunocompromised patients, *Candida* spp. and opportunistic fungi. Acute meningococcemia (Chapter 298) occurs after an upper respiratory tract infection and is associated with headache, fever, meningitis, hypotension, and DIC. The embolic lesions, which are found on the trunk and lower extremities, can range from 1 mm up to several centimeters. Disseminated gonococcal infection (Chapter 299) is accompanied by fever, arthralgias, tenosynovitis, and a small number of vesiculopustules with purpura or hemorrhagic necrosis over the distal ends of the extremities. Rocky Mountain spotted fever (Chapter 327) is a tick-borne disease characterized by headache, fever, chills, photophobia, and myalgias. The cutaneous eruption starts acrally and spreads centripetally as small, erythematous, blanchable macules that evolve into petechiae, palpable purpura, and ecchymoses. Ecthyma gangrenosum is manifested as erythematous papules and plaques in which central purpura and hemorrhagic necrosis develop; *Pseudomonas aeruginosa* (Chapter 306) is the most common organism, but *Klebsiella* spp. (Chapter 305), *Escherichia coli* (Chapter 304), and *Serratia* spp. (Chapter 305) have also been implicated. In immunocompromised patients, ecthyma gangrenosum may develop as a result of infection with *Candida* spp. or opportunistic fungi.

VESICULOBULLOUS DISEASES

Vesicles are clear, fluid-filled lesions measuring smaller than 5 mm; bullae or blisters are clear, fluid-filled lesions larger than 5 mm. Vesiculobullous lesions in the skin may be caused by immunologically mediated mechanisms, hypersensitivity reactions, metabolic disorders, inherited genetic defects, and infections (Table 439-3).

Immunologically Mediated Blistering Diseases

BULLOUS PEMPHIGOID

DEFINITION

Bullous pemphigoid is an autoimmune blistering disease that occurs in elderly adults.[3] Tense blisters and urticarial plaques occur on the flexor surfaces of the arms and legs, axilla, groin, and abdomen (Fig. 439-6). The incidence is 162 cases per million per year.

TABLE 439-3 VESICULOBULLOUS DISEASES

IMMUNOLOGICALLY MEDIATED DISEASES
Bullous pemphigoid
Herpes gestationis
Mucous membrane pemphigoid
Epidermolysis bullosa acquisita
Dermatitis herpetiformis
Linear immunoglobulin A bullous dermatosis
Pemphigus
Vulgaris
Foliaceus
Paraneoplastic
HYPERSENSITIVITY DISEASES
Erythema multiforme minor
Erythema multiforme major (Stevens-Johnson syndrome)
Toxic epidermal necrolysis
METABOLIC DISEASES
Porphyria cutanea tarda
Pseudoporphyria
Diabetic blisters
INHERITED GENETIC DISORDERS
Epidermolysis bullosa
Simplex
Junctional
Dystrophic
INFECTIOUS DISEASES
Impetigo
Staphylococcal scalded skin syndrome
Herpes simplex
Varicella
Herpes zoster

PATHOBIOLOGY AND DIAGNOSIS

IgG autoantibodies bind to the epidermal basement membrane and activate complement, which attracts inflammatory cells. These inflammatory cells release proteases that degrade basement membrane proteins and lead to blister formation. Histology reveals a subepidermal blister with an eosinophilic infiltrate. Direct immunofluorescence shows linear deposits of IgG and C3 at the basement membrane. Indirect immunofluorescence studies using salt-split skin demonstrate circulating IgG antibodies that bind to the epidermal side.[4]

TREATMENT Rx

Treatment is dictated by the degree of involvement and the rate of progression of the disease. The ultrapotent topical steroid clobetasol (10-30 g/day), applied to the skin of patients with bullous pemphigoid until 15 days after disease control is obtained and then tapered over 4 months, is effective, superior to oral corticosteroid treatment, and is associated with fewer side effects than higher dose clobetasol in patients with both moderate and severe disease.[5] Nevertheless, the practicality of such potent topical steroids is controversial, and many experts use oral prednisone, 1 mg/kg in an early morning daily dose, as a foundation of therapy for patients with generalized disease. Other treatments include dapsone (up to 200 mg/day), azathioprine (up to 2.5 mg/kg as guided by the thiopurine methyltransferase level), methotrexate (up to 15 mg weekly), mycophenolate mofetil (35-45 mg/kg divided into two daily doses), and cyclophosphamide (up to about 2.5 mg/kg). The duration of therapy with these medications varies, depending on the activity of the disease. Left untreated, bullous pemphigoid generally persists for months to years. Spontaneous remissions and exacerbations occur. The 1-year mortality rate is about six times worse than for the general population.

Herpes Gestationis

Herpes gestationis is a rare autoimmune dermatosis of pregnancy. Despite the name, herpes gestationis has no relationship to herpesvirus infection. Most patients have intense pruritus. Periumbilical urticarial plaques progress to vesicles and blisters. The eruption spreads peripherally (Fig. 439-7), typically sparing the face, palms, soles, and mucous membranes. IgG autoantibodies are produced against bullous pemphigoid antigen II, which is critical in epidermal-dermal adhesion. Antibody binding triggers an immune response leading to blister formation. Histology reveals a subepidermal blister, and direct immunofluorescence shows linear C3 deposits at the basement membrane.

Corticosteroids are the mainstay of therapy. Patients with mild disease are treated with moderate- to high-potency topical corticosteroids, such as 0.05% fluocinonide ointment applied twice daily to affected areas, but patients with extensive disease require systemic corticosteroids, such as prednisone, 1 mg/kg given once daily as an early morning dose. Disease clears within 1 to 2 weeks after initiation of treatment. There is an increased risk for premature delivery and birth of infants who are small for gestational age, suggesting that these women should be managed by obstetricians experienced in high-risk pregnancies.

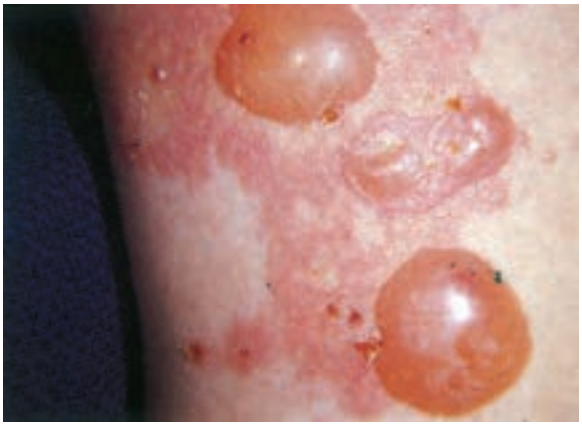

FIGURE 439-6. Bullous pemphigoid. Tense subepidermal bullae are seen on an erythematous base.

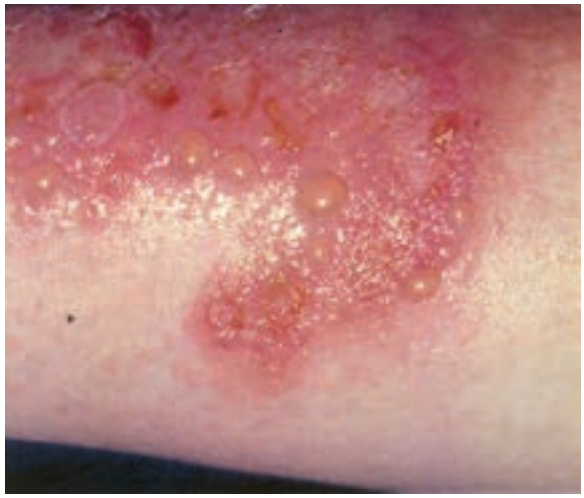

FIGURE 439-7. Herpes gestationis. Multiple tense blisters and erosions on an erythematous base are present.

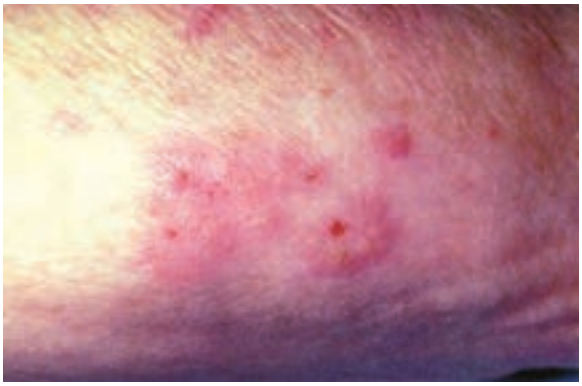

FIGURE 439-8. Dermatitis herpetiformis. The elbow of a patient has eroded erythematous papules and papulovesicles.

Mucous Membrane Pemphigoid

DEFINITION

Mucous membrane pemphigoid, previously called cicatricial pemphigoid, is a group of subepithelial blistering diseases that involve the mucosal surfaces. Patients have blisters of the oral, ocular, nasopharyngeal, laryngeal, anogenital, and esophageal mucosa that heal with scarring, which causes the major morbidity associated with the disease.

PATHOBIOLOGY

There are several subgroups of mucous membrane pemphigoid. Some patients have circulating IgG autoantibodies that bind to the dermal side of salt-split skin and recognize laminin 332. A second subgroup includes patients who have pure ocular disease and IgG antibodies directed against β_4 integrin. A third subgroup has mucosal disease and skin lesions. The fourth variant includes patients with oral disease but without skin disease. Histology reveals a subepidermal blister with an inflammatory cell infiltrate. Direct immunofluorescence shows linear IgG, IgA, and C3 deposits at the basement membrane. Indirect immunofluorescence reveals circulating IgG or IgA antibodies, or both.

TREATMENT Rx

Treatment is dictated by the extent, severity, and location of disease; it ranges from topical corticosteroids, such as 0.05% fluocinonide ointment applied twice daily to affected areas under occlusion in patients with only oral disease; to prednisone, 1 mg/kg given once daily as an early morning dose; and cyclophosphamide, up to 2.5 mg/kg given as an early morning dose for up to 2 years, for severe ocular disease. Management teams should include ophthalmologists, otolaryngologists, dermatologists, and internists for patients with severe disease. Patients with both circulating IgG and IgA antibodies tend to have more severe disease.

PROGNOSIS

Mucous membrane pemphigoid is a chronic disease. Untreated ocular disease can result in blindness.

EPIDERMOLYSIS BULLOSA ACQUISITA

Epidermolysis bullosa acquisita is an acquired autoimmune blistering disease that generally occurs in middle age.[5] There are two types of skin lesions, noninflammatory acral blisters that heal with scarring and milia formation and widespread inflammatory vesiculobullous disease.

PATHOBIOLOGY

Epidermolysis bullosa acquisita is characterized by IgG autoantibodies that target collagen VII, the major protein of anchoring fibrils. These autoantibodies alter dermal-epidermal adhesion and lead to blister formation. Histology reveals a subepidermal blister containing few inflammatory cells when mechanobullous lesions are sampled or a neutrophil-rich infiltrate when inflammatory blisters are sampled. Direct immunofluorescence reveals linear IgG deposits at the basement membrane. Indirect immunofluorescence shows circulating IgG autoantibodies that bind the dermal side of salt-split skin.

TREATMENT Rx

Epidermolysis bullosa acquisita is a chronic disease that is very difficult to treat. Dapsone (at doses up to 200 mg/day), colchicine (0.6 mg twice daily), azathioprine (up to 2.5 mg/kg as guided by the thiopurine methyltransferase level), or cyclophosphamide (2.5 mg/kg given as an early morning dosage) alone or along with prednisone (60 mg given once daily as an early morning dose) is only occasionally successful. Cyclosporine at doses of up to 5 mg/kg, extracorporeal photopheresis, and IV immunoglobulin (at doses of 2 gm/kg given monthly) have been used successfully in patients with severe disease. Because of nephrotoxicity, cyclosporine should be reserved for crisis management of patients who have severe disease and are experiencing a major flare.

DERMATITIS HERPETIFORMIS

Dermatitis herpetiformis is an immune-mediated vesicular disease that occurs in young to middle-aged patients.[6] Skin lesions are extremely pruritic; grouped vesicles and erosions located on the scalp; posterior of the neck; and extensor surfaces of the elbows, knees, and buttocks (Fig. 439-8). Most patients have a subclinical gluten-sensitive enteropathy (Chapter 140) that is reversible with a gluten-free diet. Diet alone can sometimes control the skin disease, with clearance of the cutaneous granular IgA deposits at the basement membrane. Biopsy specimens of skin lesions reveal dermal papillary neutrophilic microabscesses. Direct immunofluorescence shows dermal papillary granular IgA deposits. Most patients with dermatitis herpetiformis have circulating IgA antibodies directed against tissue transglutaminase. Dermatitis herpetiformis can be treated with dapsone, up to 200 mg/day given chronically. Dermatitis herpetiformis is a lifelong disease.

LINEAR IMMUNOGLOBULIN A BULLOUS DERMATOSIS

Linear IgA bullous dermatosis is an acquired autoimmune blistering disease. Primary lesions are papulovesicles, and involvement of the oral mucous membranes is common. The disease occurs throughout adulthood. Deposition of IgA antibody specific for a portion of bullous pemphigoid antigen II leads to complement activation and neutrophil chemotaxis. Proteolytic enzymes are released, destroy the dermal-epidermal junction, and cause blister formation. Histology reveals a subepidermal vesicle with neutrophil predominance. Direct immunofluorescence shows linear deposits of IgA at the basement membrane. Indirect immunofluorescence demonstrates circulating IgA antibodies. Most patients respond to dapsone, up to 200 mg/day, given chronically. Patients whose disease is not controlled with dapsone may benefit from the addition of oral prednisone, 1 mg/kg once daily as an early morning dose. The disease tends to be chronic in adults, but the childhood version (called chronic bullous disease of childhood) may run a several-year course and then remit.

PEMPHIGUS

DEFINITION

Pemphigus refers to a group of autoimmune blistering intraepidermal diseases of the skin and mucous membranes that are most common in middle age.[7]

PATHOBIOLOGY

Autoantibodies in pemphigus vulgaris target desmoglein III, and autoantibodies in pemphigus foliaceus target desmoglein I. Circulating antibodies in paraneoplastic pemphigus recognize a complex of proteins, including desmoplakin I and II, bullous pemphigoid antigen I, envoplakin, periplakin, and desmoglein I and III.

CLINICAL MANIFESTATIONS

Patients with pemphigus vulgaris have flaccid blisters and erosions in the oropharynx (Fig. 439-9), trunk, head, neck, and intertriginous areas. Pemphigus foliaceus is accompanied by erythema, scaling, and crusting of the face, scalp, and upper part of the trunk. Patients with paraneoplastic pemphigus have ocular and oral blisters and erosions along with skin lesions resembling erythema multiforme and an associated underlying malignancy that is generally lymphoreticular in origin (Chapter 179).

DIAGNOSIS

Skin biopsy specimens from patients with pemphigus vulgaris reveal suprabasilar acantholysis, but pemphigus foliaceus biopsy specimens demonstrate subcorneal acantholysis. Biopsy specimens of paraneoplastic pemphigus show suprabasilar acantholysis and dyskeratotic keratinocytes with basal cell vacuolization.

Whereas direct immunofluorescence demonstrates cell surface deposits of IgG in patients with pemphigus vulgaris and foliaceus, indirect immunofluorescence reveals circulating IgG antibodies that recognize molecules on the epidermal cell surface. Patients with paraneoplastic pemphigus have circulating and tissue-bound IgG antibodies that are indistinguishable from those in pemphigus vulgaris and that also recognize the cell surface of simple epithelia, including the liver and heart.

TREATMENT

Rx

Treatment regimens depend on the patient's age, the degree of involvement, the rate of disease progression, and the subtype of pemphigus.[8] Whereas systemic corticosteroids (e.g., oral prednisone, 1 mg/kg once daily as an early morning dose) are required for the treatment of patients with pemphigus vulgaris, topical corticosteroids (e.g., 0.05% fluocinonide ointment applied twice daily to affected areas) may occasionally control pemphigus foliaceus. One course of IV immunoglobulin (400 mg/kg/day for 5 days) appears to be a safe and effective treatment for patients whose disease is resistant to steroids. Other agents include dapsone (up to 200 mg/day), hydroxychloroquine (administered at less than 6 mg/kg of lean body mass divided into two daily doses), mycophenolate mofetil (35-45 mg/kg/day divided into two daily doses), azathioprine (up to 2.5 mg/kg as guided by the thiopurine methyltransferase level), cyclophosphamide (up to 2.5 mg/kg given as an early morning dose), and rituximab (four weekly doses of 375 mg/m^2). The duration of therapy with each of these medications varies according to the level of disease activity. Although the use of steroid-sparing agents is supported by clinical experience, few controlled studies have demonstrated their benefit. For paraneoplastic pemphigus caused by benign tumors such as Castleman disease (Chapter 185), tumor removal is often curative. Patients with associated malignant tumors have recalcitrant disease, although there are occasional successes with pulse corticosteroids (methylprednisolone, 1000 mg given daily for 3 consecutive days), pulse cyclophosphamide (500-1000 mg given monthly for 6 months to 1 year, often along with varying doses of prednisone), immunoapheresis, immunoablative high-dose cyclophosphamide (50 mg/kg/day for 4 days), and rituximab (four weekly doses of 375 mg/m^2).

PROGNOSIS

Before the availability of corticosteroids, 60% to 90% of patients with pemphigus vulgaris died; the mortality rate has now decreased to the 5% to 10% range. Overall, however, the mortality rate is about twice that of the general population. The prognosis of patients with paraneoplastic pemphigus is related to the type of associated neoplasm. Patients with benign tumors usually experience clearance of their lesions after tumor resection, but those with malignant tumors generally have a poor prognosis.

Hypersensitivity Reactions That Cause Blisters

ERYTHEMA MULTIFORME

DEFINITION

Erythema multiforme is an acute blistering eruption that occurs in all age groups.[9] Erythema multiforme minor is localized, with minimal or no mucosal involvement. Erythema multiforme major, also known as Stevens-Johnson syndrome, is a more severe mucosal and skin disease characterized by signs and symptoms reminiscent of serum sickness (Chapter 47).

CLINICAL MANIFESTATIONS

Toxic epidermal necrolysis is at the most severe end of the erythema multiforme spectrum. The primary lesions of erythema multiforme minor are erythematous macules and edematous papules with vesicular centers that become dusky violet. Target lesions are found on extensor surfaces of the extremities and spread centripetally (Fig. 439-10). The skin lesions of Stevens-Johnson syndrome resemble those of erythema multiforme minor but are likely to be generalized and show confluent erythema with urticarial and purpuric lesions. Erosions of two or more mucosal surfaces occur in Stevens-Johnson syndrome and may include hemorrhagic crusting of the lips, ulceration of the ocular mucosa, and genital involvement. Patients with erythema multiforme major have a 1- to 14-day prodrome that includes fever, cough, sore throat, vomiting, and diarrhea. Patients with toxic epidermal necrolysis may have a similar prodrome, rapidly followed by generalized macular erythema that progresses to confluent erythema with skin tenderness. Large blisters follow soon afterward, and then skin sloughing occurs as the large blisters break and leave denuded skin.

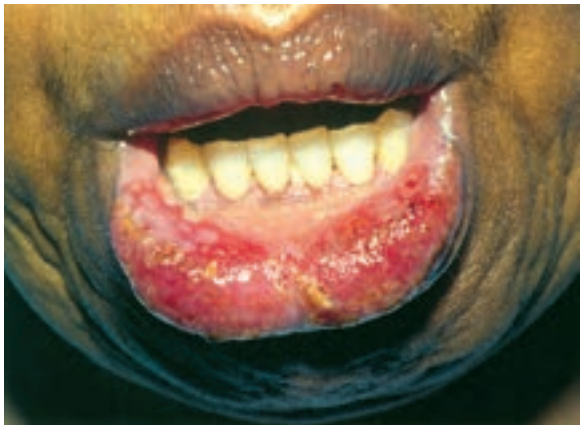

FIGURE 439-9. Pemphigus vulgaris. The lower lip of a patient has confluent erosions with scattered crusting.

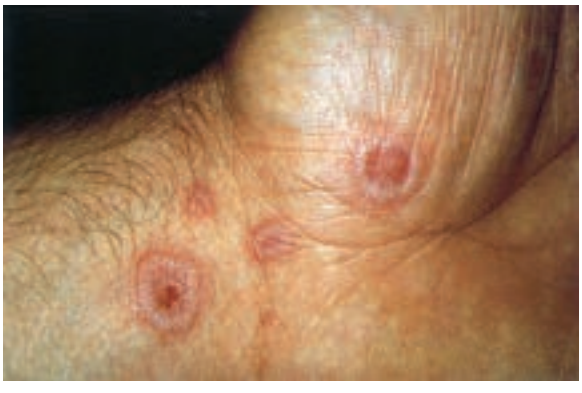

FIGURE 439-10. Erythema multiforme. Target or "bull's-eye" annular lesions with central vesicles and bullae are characteristic of erythema multiforme.

PATHOBIOLOGY

Common etiologic associations of erythema multiforme include infections such as herpes simplex (Chapter 374) (especially recurrent erythema multiforme minor), *Mycoplasma pneumoniae* (Chapter 317), and drug reactions (Chapter 254). Sulfonamides, penicillins, barbiturates, carbamazepine, phenytoin, allopurinol, and nonsteroidal anti-inflammatory drugs (NSAIDs) are the most common drugs implicated in Stevens-Johnson syndrome and toxic epidermal necrolysis.

DIAGNOSIS

The diagnosis of erythema multiforme is clinical.

TREATMENT Rx

Chronic antiviral treatment with acyclovir (400 mg twice daily for 6 months) decreases outbreaks in a subset of patients with recurrent erythema multiforme minor. Treatment of erythema multiforme major is otherwise nonspecific, with attention to fluid and electrolyte balance and eye disease being critical. If a drug is suspected, it must be withdrawn. Systemic corticosteroid treatment is contraindicated in Stevens-Johnson syndrome. Treatment of toxic epidermal necrolysis is very difficult, but uncontrolled studies suggest that IV immune globulin (2-3 g/kg over a period of 2-5 days) may improve the prognosis.

PROGNOSIS

Erythema multiforme minor usually subsides within 2 to 3 weeks. Erythema multiforme major takes 3 to 6 weeks to clear and has less than a 5% mortality rate. Toxic epidermal necrolysis has a mortality rate approaching 30%, and patients are best managed in an intensive care or burn unit.

Metabolic Disorders That Cause Blisters

PORPHYRIA CUTANEA TARDA

Porphyria cutanea tarda is caused by deficient activity of the heme synthetic enzyme uroporphyrinogen decarboxylase (Chapter 210).[10] The fragility of sun-exposed skin leads to erosions and bullae, which are worst on the dorsal surface of the hands (Fig. 439-11), forearms, and face. Healing of crusted erosions and blisters leaves scars, milia, and hyperpigmented and hypopigmented atrophic patches. Hypertrichosis is common and most florid over the temporal and malar areas. Urinary porphyrin levels are abnormally high. Histology reveals a subepidermal blister with minimal dermal infiltrate, and direct immunofluorescence demonstrates immunoglobulin and complement deposition in dermal capillaries and at the basement membrane. Phlebotomy is the standard treatment, and the goal is to reduce serum ferritin to the lower limit of the normal range. Another treatment option is oral hydroxychloroquine at a dose (200 mg twice weekly until the disease is under control) that is much lower than that used for photoprotective indications. Alcohol and estrogen use should be discontinued because they can cause the disease to flare.

PSEUDOPORPHYRIA

Pseudoporphyria is a bullous eruption that mimics porphyria cutanea tarda clinically and histologically without porphyrin abnormalities. Many medications can cause pseudoporphyria, including propionic acid–derivative NSAIDs (e.g., naproxen, diflunisal, ketoprofen, nabumetone, oxaprozin, and mefenamic acid), furosemide, tetracycline, fluoroquinolones, amiodarone, cyclosporine, dapsone, etretinate, and flutamide. The prognosis is good for patients with pseudoporphyria after the offending agent has been discontinued. However, resolution of the disease may take several months. In patients with chronic renal failure treated by hemodialysis, true porphyria cutanea tarda or pseudoporphyria may develop and is very difficult to treat.

DIABETES MELLITUS

Distal extremity blisters may occasionally develop in patients with diabetes mellitus (Chapter 229). There is no correlation between the development of blisters and the severity, duration, or complications of diabetes. The mechanism of blister formation is not understood.

Inherited Genetic Disorders That Cause Blisters

Epidermolysis bullosa is a group of inherited bullous disorders characterized by blister formation in response to mechanical trauma. Subtypes include epidermolysis bullosa simplex (intraepidermal skin separation) (Fig. 439-12), junctional epidermolysis bullosa (skin separation in the lamina lucida), and dystrophic epidermolysis bullosa (sublamina densa separation). Infancy is an especially difficult time, and epidermolysis bullosa may be accompanied by blistering that is complicated by infection and sepsis. Many patients with junctional epidermolysis bullosa have severe disease that can lead to death, usually secondary to infection, and metastatic squamous cell carcinoma may develop in some patients with recessive dystrophic epidermolysis bullosa and can also lead to death. In contrast, epidermolysis bullosa simplex, milder forms of junctional epidermolysis bullosa, and dominant dystrophic epidermolysis bullosa do not usually affect life expectancy. Epidermolysis bullosa simplex is caused by mutations of the genes coding for keratins 5 and 14. Junctional epidermolysis bullosa has a variable molecular etiology, and mutations in genes coding for laminin 5 subunits, bullous pemphigoid antigen I, α_6 integrin, and β_4 integrin have been demonstrated. Dystrophic epidermolysis bullosa is caused by mutations of the type VII collagen gene. Epidermolysis bullosa is a lifelong disease. Some subtypes, especially the milder forms, improve with age. No medications are known to correct the underlying molecular defects, but gene therapy is being actively pursued.

Infectious Diseases That Cause Blisters

IMPETIGO

Impetigo is a bacterial infection of the superficial layers of the epidermis. Bullous impetigo is caused by *S. aureus* (Chapter 288), and nonbullous impetigo is caused by group A β-hemolytic streptococci (Chapter 290). Bullous impetigo is manifested as vesicles and bulla (Fig. 439-13). Lesions are common on the face but may appear anywhere. Nonbullous impetigo is characterized by fragile vesicles or pustules that rupture and leave honey-colored crusted papules or plaques, especially near the nose and mouth and on the extremities. Lesions develop on normal or traumatized skin or are superimposed on preexisting conditions, including scabies, varicella, or atopic dermatitis. The causative agent of bullous impetigo is coagulase-positive *S. aureus,*

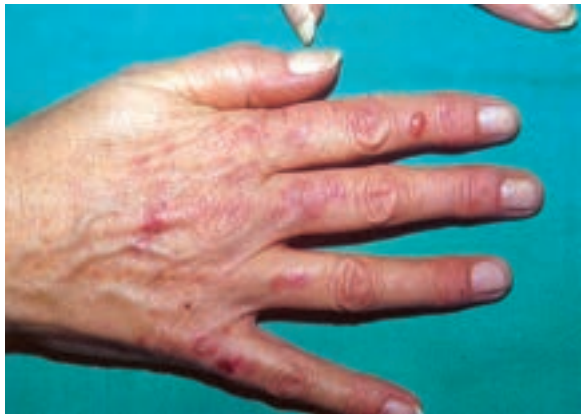

FIGURE 439-11. Porphyria cutanea tarda. A blister and erosions are present on the dorsal surface of the hand.

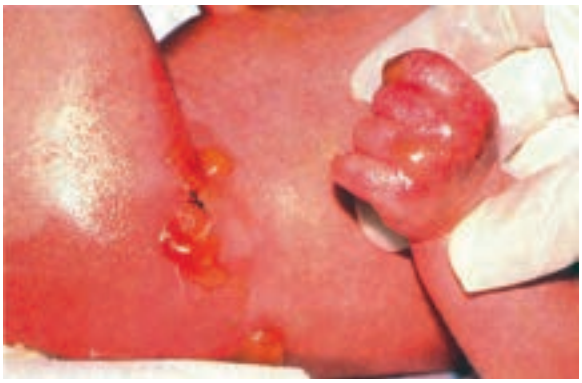

FIGURE 439-12. Epidermolysis bullosa simplex. Tense blisters and erosions are present on the trunk and extremities of a newborn.

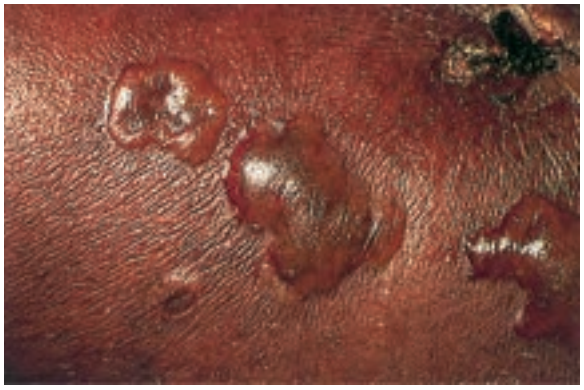

FIGURE 439-13. Bullous impetigo. Multiple blisters are present on the trunk of this patient.

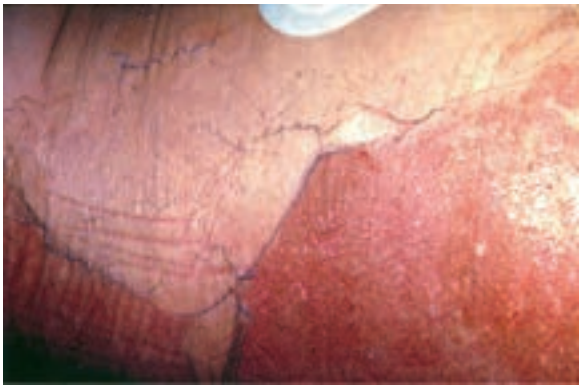

FIGURE 439-14. Staphylococcal scalded skin syndrome. Confluent erythema with exfoliation of skin is seen on the trunk.

which produces exfoliatins A and B. The toxins cause cleavage within or below the stratum granulosum. Impetigo is diagnosed clinically. Culture and sensitivity studies are recommended if topical or oral treatment is ineffective. Oral antibiotics, including dicloxacillin (500 mg four times daily for 7 days) or cephalexin (500 mg three times daily for 7 days), are used for extensive disease or in patients refractory to topical mupirocin. Gentle débridement of crusts is recommended. Lesions resolve after 7 to 10 days of treatment. Acute glomerulonephritis (Chapter 121) develops in 2% to 5% of young children with nonbullous impetigo, usually within 10 days after the skin lesions appear.

STAPHYLOCOCCAL SCALDED SKIN SYNDROME

Staphylococcal scalded skin syndrome is a blistering disease caused by an exotoxin produced by *S. aureus* (Chapter 288). It is most common in young children but may occur in adults who have renal insufficiency or are immunocompromised. The site of the staphylococcal infection is usually extracutaneous. Staphylococcal scalded skin syndrome is manifested as a sudden onset of fever and tender, blanchable erythema. It starts on the central part of the face, neck, and intertriginous areas and rapidly generalizes. The palms, soles, and mucous membranes are spared. Flaccid blisters occur within 1 to 2 days and soon exfoliate in large sheets, with superficially denuded skin remaining (Fig. 439-14). The disease must be distinguished from toxic epidermal necrolysis by skin biopsy. Whereas in staphylococcal scalded skin syndrome, there is an upper epidermal blister, toxic epidermal necrolysis causes a dermal-epidermal blister. Patients with the most severe staphylococcal scalded skin syndrome should be treated with IV nafcillin or oxacillin, 2 g every 4 to 6 hours for 10 to 14 days. If the patient is found to have methicillin-resistant staphylococci, vancomycin, 1 g every 12 hours (with dose adjustment based on creatinine clearance), should be given for 10 to 14 days. Patients with mild disease may be treated with oral dicloxacillin, 500 mg four times daily for 10 to 14 days, unless the staphylococcal isolate is methicillin resistant, in which case the choice of antibiotics should be guided by the results of sensitivity testing. For severe methicillin-resistant gram-positive skin infections, one dose of intravenous oritavancin (1200 mg) or two doses of intravenous dalbavancin (1 g on day 1, 500 mg on day 8) is as efficacious as twice daily intravenous vancomycin for up to 7-10 days.

HERPES SIMPLEX VIRUS INFECTION

PATHOBIOLOGY

Herpes simplex virus (HSV) infection (Chapter 374) may be caused by type 1 or type 2 HSV. The hallmark of HSV infection is its ability to establish latent infection. Disease commonly occurs as a recurrent vesicular eruption of the oral, perioral (typically HSV-1), or genital (typically HSV-2) regions, although primary gingivostomatitis (typically in children and young adults and caused by HSV-1) and primary genital herpes (typically HSV-2) are less common.

CLINICAL MANIFESTATIONS

Patients with primary gingivostomatitis have high fever, regional lymphadenopathy, and malaise. Patients with primary genital herpes have fever, flulike symptoms, tender inguinal adenopathy, and aseptic meningitis. These infections all reveal grouped vesicles on an erythematous base. Recurrent eruptions can be triggered by skin trauma, cold or heat, concurrent infection, and menstruation. Chronic erosive ulcers of the face and anogenital areas may develop in immunocompromised patients.

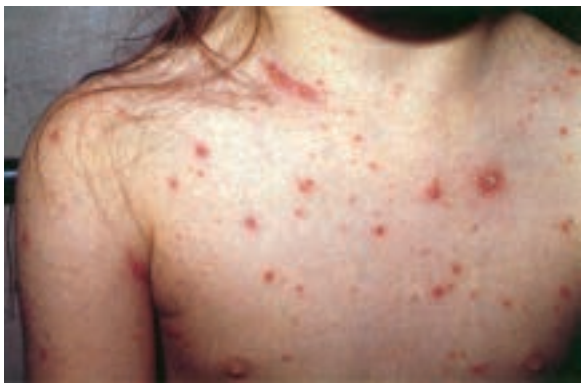

FIGURE 439-15. Erythematous macules and vesicles with crusted erosions on the chest of a patient with varicella.

DIAGNOSIS

Tzanck smear of fluid from the roof of a vesicle can be helpful in confirming the diagnosis (see Fig. 436-15 in Chapter 436), but viral culture is the diagnostic "gold standard." The direct fluorescent antibody test is an antigen-based technique that not only yields same-day results but can also distinguish HSV from varicella-zoster virus (VZV) and is becoming widely used.

TREATMENT

Rx

In healthy individuals, HSV infection is self-limited. The goal of treatment is to shorten the current attack and prevent recurrences. Acyclovir is effective in the treatment of HSV infections; valacyclovir and famciclovir are closely related, effective medications with improved oral bioavailability. The doses and duration of therapy vary depending on whether the infection is limited to the oral or genital mucous membranes or is disseminated and on whether it is primary or recurrent disease (Chapter 374).

CHICKENPOX

Chickenpox is caused by VZV (Chapter 375). It is usually a childhood disease, but affected adults have more morbidity. Skin lesions occur 10 to 21 days after exposure to VZV. Erythematous macules appear on the scalp, face, trunk, and proximal ends of the limbs, with rapid progression to papules, vesicles, pustules, and crusting (Fig. 439-15). Adults may experience a more widespread eruption, prolonged fever, and pneumonia. The diagnosis is usually made clinically, but direct fluorescent antibody or culture confirmation is sometimes needed. Treatment of healthy children is unnecessary because the disease is self-limited. Adults should be treated with oral acyclovir, 800 mg five times a day for 7 days. The varicella vaccine given once to

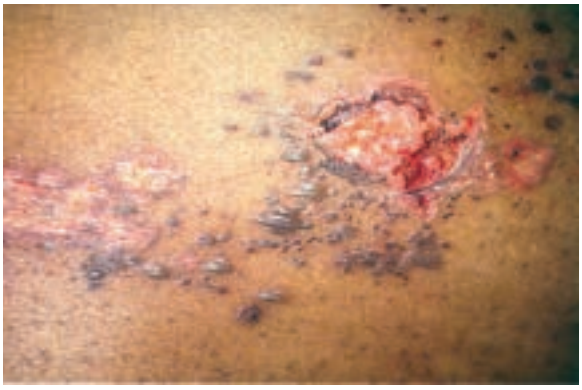

FIGURE 439-16. Herpes zoster. Necrotic blisters and erosions in a dermatomal pattern are seen on the trunk of this patient.

healthy children 12 to 18 months of age and twice, in a 4- to 8-week interval, to susceptible persons older than 13 years (Chapter 18) is highly effective.

HERPES ZOSTER

PATHOBIOLOGY

Herpes zoster (Chapter 375) is caused by reactivation of VZV from a previous chickenpox infection. The disease is more common in older and immunocompromised patients.

CLINICAL MANIFESTATIONS

The typical manifestation is painful grouped herpetiform vesicles on an erythematous base confined to cutaneous surfaces innervated by one sensory nerve and preceded by radicular pain (Fig. 439-16). The major morbidity, which is pain within the affected dermatome, can be severe and persist after the skin lesions have resolved (postherpetic neuralgia). Immunocompromised patients have an increased risk for cutaneous dissemination and visceral involvement of the bladder, lungs, and CNS.

TREATMENT Rx

Acyclovir (800 mg orally five times daily for 7 days) and its derivatives valacyclovir (500 mg orally three times daily for 7 days) and famciclovir (500 mg orally three times daily for 7 days) are safe and effective in the treatment of active disease and prevention of postherpetic neuralgia. In an immunocompromised patient, acyclovir should be given intravenously (10 mg/kg every 8 hours for 7-10 days). The earlier antiviral medications are started, the more effective they are in shortening the duration of herpes zoster and in preventing or decreasing the severity of postherpetic neuralgia. Oral steroids are not effective in reducing the incidence of acute pain or postherpetic neuralgia. Patients with postherpetic neuralgia of more than 3 months' duration benefit from the use of gabapentin, 1600 or 2400 mg/day, with a significant reduction in pain. VZV vaccine markedly reduces both the morbidity from herpes zoster and postherpetic neuralgia in healthy adults older than 60 years (Chapter 375).

PUSTULAR ERUPTIONS

Acne Vulgaris

DEFINITION

Acne vulgaris is the most common pustular skin condition. Teenagers are usually affected, but the disease may persist into adulthood. The comedo, which is the primary lesion, can be either closed (whitehead) or open (blackhead).

PATHOBIOLOGY

Androgen production after puberty stimulates the release of sebum by sebaceous glands. Sebum flow is impeded because of abnormal keratinization within the pilosebaceous canal, a process that leads to the formation of comedones. Bacterial (*Propionibacterium acnes*) proliferation within the comedo predisposes to rupture of the pilosebaceous unit with extravasation into the surrounding dermis, which results in papules, pustules, and cysts.

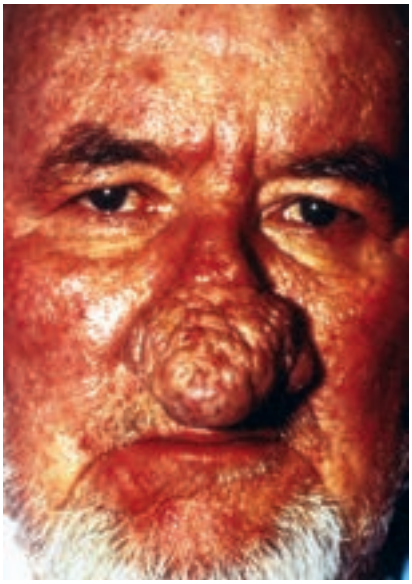

FIGURE 439-17. Sebaceous gland hyperplasia. A large red bulbous nose known as rhinophyma is characteristic of late-stage rosacea.

TREATMENT Rx

Patients with mild disease are treated topically with benzoyl peroxide, tretinoin, adapalene, or tazarotene, which normalize follicular keratinization. In patients with mild to moderate disease, treatment with benzoyl peroxide is the most cost-effective therapy.[■] The addition of topical antibiotics helps control inflammatory papules and pustules. More significant disease is often treated with oral tetracycline (250-1000 mg/day), doxycycline (200 mg/day), or minocycline (200 mg/day). Dapsone gel (5%) is safe and effective.[■] Another approach is oral contraceptives containing either ethinyl estradiol and norgestimate or ethinyl estradiol and drospirenone, both of which are superior to placebo in the treatment of acne in women.[■] Isotretinoin (1 mg/kg given for 5 months), which decreases sebaceous gland size and sebum production, is reserved for severe cystic disease because of its teratogenicity, possible associated risk for depression, and other significant side effects. Acne may be exacerbated by the use of oil-based cosmetics or hair preparations. Androgenic hormones, systemic corticosteroids, lithium, phenytoin, phenobarbital, isoniazid, and endocrinologic conditions such as polycystic ovary disease and adrenal or ovarian tumors may produce acneiform eruptions or aggravate preexisting acne.

ROSACEA

Rosacea, which is a chronic inflammatory disease of the face, affects the pilosebaceous units and blood vessels and generally occurs in middle age. Erythema, telangiectases, erythematous papules, and pustules occur on the central part of the face. Ocular rosacea can lead to keratitis, iritis, blepharitis, and recurrent chalazion; it should be managed by an ophthalmologist. In its most severe form, rosacea can cause sebaceous gland hyperplasia leading to a large red bulbous nose known as rhinophyma (Fig. 439-17). Rosacea is more likely to develop in patients with a tendency toward facial flushing. Flushing[11] can be caused by heat, spicy foods, hot drinks, alcohol, or emotional stimuli. With time, the flushing reaction lasts longer and longer until it persists. Topical antibiotics, including 0.75% to 1% metronidazole, are helpful in mild disease, and the combination of a subantimicrobial dose (20 mg twice daily) of doxycycline plus topical 0.75% metronidazole is more efficacious than topical metronidazole alone.[■] Patients with more severe disease require oral tetracycline 500 mg twice daily for 3 to 4 months, oral doxycycline 100 mg twice daily, oral minocycline 100 mg twice daily, or low-dose isotretinoin 0.3 mg/kg.[■] Topical azelaic acid gel (15%) can significantly improve papulopustular rosacea.

PERIORAL DERMATITIS

Perioral dermatitis is characterized by erythematous papules and pustules as well as scaling patches around the mouth and eyes (Fig. 439-18). Most

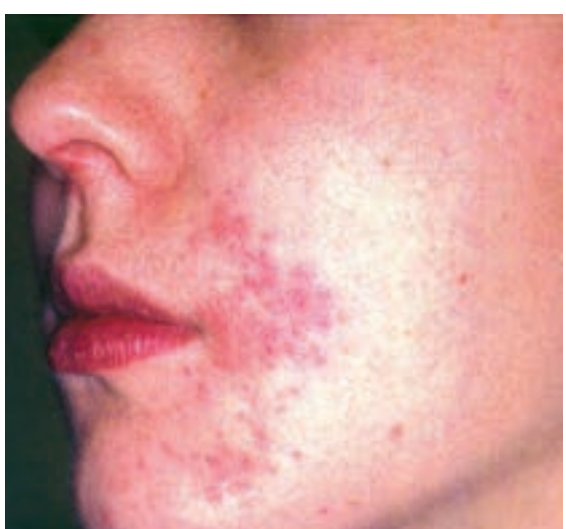

FIGURE 439-18. Perioral dermatitis. Erythematous papules and pinpoint pustules are evident around the mouth.

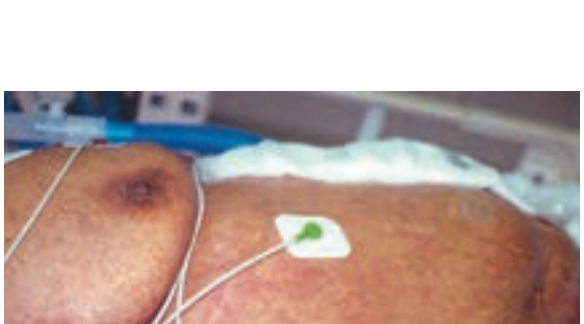

FIGURE 439-19. Acute generalized exanthematous pustulosis. Erythematous macules and numerous superficial pustules are present on the trunk of this patient.

patients have used potent topical corticosteroids inappropriately for long periods. The eruption generally clears after stopping the corticosteroid and using tetracycline, 250 mg orally twice daily for 6 weeks.

ACUTE GENERALIZED EXANTHEMATOUS PUSTULOSIS

Acute exanthematous pustulosis is a generalized pustular eruption that is associated with fever and frequently caused by antibiotics. Pustules develop within 2 days of drug administration, start on the face or in flexural areas, and rapidly disseminate (Fig. 439-19). Spontaneous resolution occurs in less than 2 weeks.

PUSTULAR PSORIASIS

Pustular psoriasis is a variant of psoriasis (Chapter 438) that localizes to the palms and soles or generalizes over the entire body. Patients with generalized disease have fever and leukocytosis and require systemic therapy. Pustules can also be seen in patients with septic emboli of bacterial or fungal origin, including gonococcemia and systemic candidiasis (see Purpuric Eruptions).

FOLLICULITIS

Folliculitis is inflammation of the hair follicles caused by infection with staphylococci. It is caused by obstruction of individual hair follicles and associated pilosebaceous units. Folliculitis is more common in patients with diabetes mellitus, obesity, or immunocompromised states. The primary lesion is a pustule with a central hair. Typical affected sites are the scalp, thighs, trunk, axilla, and inguinal area. Sometimes the infection can extend deeper into the

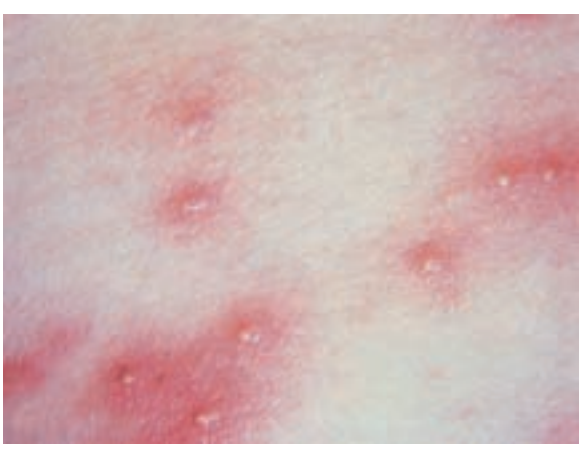

FIGURE 439-20. *Pseudomonas* species folliculitis. The trunk of this patient has numerous pustules on an erythematous base.

dermis and form larger erythematous nodules from one (furuncle) or more (carbuncle) follicles. Treatment with oral antibiotics such as cephalexin, 500 mg twice daily for 14 days, clears extensive infections, but topical antibiotics (e.g., clindamycin solution applied twice daily for 2 weeks or longer) and antibacterial soaps (e.g., Dial or Lever 2000) help in milder disease.

Pseudomonas Folliculitis

Pseudomonas folliculitis is acquired from hot tubs contaminated with *P. aeruginosa* (Chapter 306). The typical finding is papules and pustules in areas of skin occluded by a bathing suit (Fig. 439-20). Treatment with ciprofloxacin, 500 mg twice daily for 10 to 14 days, is usually curative.

Pityrosporum folliculitis is a pruritic, acne-like eruption that occurs on the upper part of the back and the chest, arms, and face and is caused by *Pityrosporum ovale.* Treatment is with oral itraconazole, 200 mg/day for 1 week, and 2% ketoconazole shampoo applied to the affected area daily for 1 month.

Eosinophilic Pustular Folliculitis

Eosinophilic pustular folliculitis is a sterile, intensely pruritic folliculitis usually found on the faces, chests, and backs of patients who are positive for HIV (Chapter 392). Skin biopsy is needed to confirm the diagnosis. Treatment is difficult, but options include potent topical corticosteroids, such as 0.1% triamcinolone ointment applied twice daily for several months (although this class 4 topical corticosteroid should not be used on the face); tacrolimus ointment 0.1% is another topical option that is safe for facial use. Systemic options include antihistamines, such as hydroxyzine, 25 mg every 8 hours as needed, and ultraviolet light therapy.

HIDRADENITIS SUPPURITIVA

Hidradenitis suppurativa is a chronic recurring inflammatory disease that occurs in skin-bearing apocrine glands and is characterized by painful deep-seated nodules and abscesses.

EPIDEMIOLOGY AND PATHOBIOLOGY

Hidradenitis most commonly occurs in the early 20s, and its incidence declines after age 50 years. There is a 3 : 1 predominance of women to men. Cigarette smoking and obesity are risk factors that also correlate with the severity of disease.

The pathogenesis of hidradenitis remains poorly understood. Histologic studies suggest a multifocal nature, with early atrophy of the sebaceous glands followed by lymphocytic inflammation of the pilosebaceous unit and later by destruction of the hair follicles and the formation of granulomas. More recent evidence supports the concept that hidradenitis suppurativa is an inflammatory or immune disease.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Hidradenitis is characterized by its chronicity, with periods of flares and remissions. Although the typical presentation is with a painful, inflamed nodule in the axilla or the groin, lesions also are often found in the inframammary, genital, and perineal regions. The initial inflamed nodule may either resolve or develop into a persisting painless nodule that intermittently flares (often related to menses) or into a discharging abscess with substantial pain.

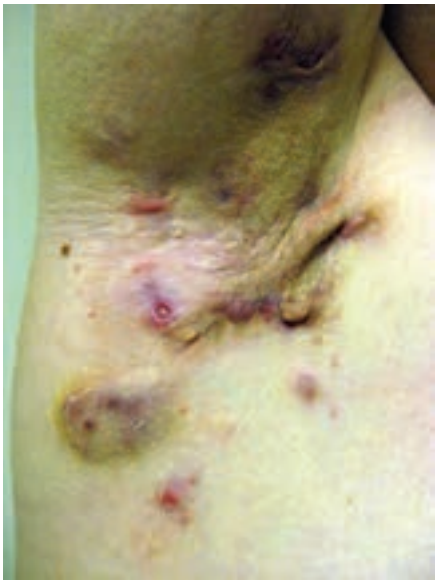

FIGURE 439-21. Hidradenitis suppurativa. The axilla of this patient has several erythematous nodules with draining sinus tracts

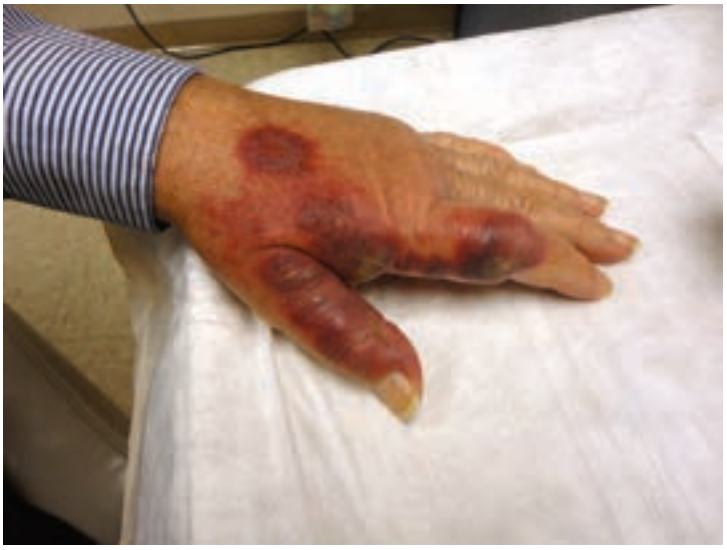

FIGURE 439-22. Sweet syndrome. The distal extremity of this patient has very large edematous indurated plaques and nodules

Lesions typically recur at the same sites or very nearby despite treatment by incision and drainage as well as with antibiotics. Over time, more sites may become affected, and patients sometimes develop sinus tracts and cordlike scarring (Fig 439-21).

TREATMENT Rx

Treatment is determined by the staging of disease. Localized (Hurley stage 1) disease should be managed with topical clindamycin (1% solution) used twice daily or intralesional triamcinolone (5-10 mg/cc). Patients with Hurley stage 2 (which is defined by the presence of one or more widely separated recurrent abscesses with tract formation and scars) and Hurley stage 3 (which is defined by the presence of multiple interconnected tracts and abscesses throughout an entire area) require more aggressive treatment. A variety of oral antibiotics, systemic corticosteroids, antiandrogens, retinoids, dapsone, and cyclosporine have been used but with minimal evidence to support their efficacy. Recent studies demonstrate improvement with the tumor necrosis factor inhibitors adalimumab or infliximab.[A12,A13]

SWEET SYNDROME

Sweet syndrome (also known as acute febrile neutrophilic dermatosis) is characterized by fever; peripheral neutrophilia; and painful papules, nodules, or plaques with a dense neutrophilic infiltrate but without any frank vasculitis[12] Although the pathogenesis of this syndrome is unknown, it may represent a hypersensitivity reaction occurring in response to either underlying systemic disease or antigenic stimulation

CLINICAL MANIFESTATIONS

Sweet syndrome may be subdivided into three distinct subtypes: classic or idiopathic, paraneoplastic, and drug-induced. Classic Sweet is more common in women, especially between the ages 30 and 60 years. It is often preceded by a flulike (respiratory or gastrointestinal) syndrome and then presents as multiple firm, tender, deeply erythematous papules or nodules, which may turn into edematous plaques. Common locations include the head, neck, and upper and lower extremities (Fig. 439-22 and Fig. 440-23 in Chapter 440) Patients may have fever, arthralgias, arthritis, and myalgias when the disease is active.

Sweet syndrome can be the presenting manifestation of a malignancy, and recurrent episodes of the syndrome can be a sign of recurrent cancer. Malignancy associated Sweet syndrome affects men and women equally. Hematologic malignancies are the most common paraneoplastic association, and acute myelogenous leukemia accounts for the majority of these cases. Drug-induced Sweet syndrome most typically occurs in patients receiving granulocyte colony-stimulating factor.

TREATMENT Rx

Oral corticosteroids (e.g., prednisone, 1 mg/kg/day given over a 4- to6-week period with tapering at that time) are the standard and most effective therapy for Sweet syndrome, but some patients may require longer low-dose prednisone to prevent recurrences. The two other most effective therapies are potassium iodide (300 mg three times daily) and colchicine (0.6 mg three times daily).

Grade A References

A1. Oates-Whitehead RM, Baumer JH, Haines L, et al. Intravenous immunoglobulin for the treatment of Kawasaki disease in children. *Cochrane Database Syst Rev.* 2003;4:CD004000.

A2. Kobayashi T, Saji T, Otani T, et al. Efficacy of immunoglobulin plus prednisolone for prevention of coronary artery abnormalities in severe Kawasaki disease (RAISE study): a randomised, open-label, blinded-endpoints trial. *Lancet.* 2012;379:1613-1620.

A3. Kirtschig G, Middleton P, Bennett C, et al. Interventions for bullous pemphigoid. *Cochrane Database Syst Rev.* 2010;10:CD002292.

A4. Amagai M, Ikeda S, Shimizu H, et al. A randomized double-blind trial of intravenous immunoglobulin for pemphigus. *J Am Acad Dermatol.* 2009;60:595-603.

A5. Corey GR, Kabler H, Mehra P, et al. Single-dose oritavancin in the treatment of acute bacterial skin infections. *N Engl J Med.* 2014;370:2180-2190.

A6. Boucher HW, Wilcox M, Talbot GH, et al. Once-weekly dalbavancin versus daily conventional therapy for skin infection. *N Engl J Med.* 2014;370:2169-2179.

A7. Ozolins M, Eady EA, Avery AJ, et al. Comparison of five antimicrobial regimens for the treatment of mild to moderate inflammatory facial acne vulgaris in the community: randomized controlled trial. *Lancet.* 2004;364:2188-2195.

A8. Draelos ZD, Carter E, Maloney JM, et al. Two randomized studies demonstrate the efficacy and safety of dapsone gel 5% for the treatment of acne vulgaris. *J Am Acad Dermatol.* 2007;56:439.

A9. Koltun W, Lucky AW, Thiboutot D, et al. Efficacy and safety of 3 mg drospirenone/20 mcg ethinylestradiol oral contraceptive administered in 24/4 regimen in the treatment of acne vulgaris: a randomized double-blind, placebo-controlled trial. *Contraception.* 2008;77:249-256.

A10. Sanchez J, Somolinos AL, Almodovar PI, et al. A randomized, double-blind, placebo-controlled trial of the combined effect of doxycycline hyclate 20-mg tablets and metronidazole 0.75% topical lotion in the treatment of rosacea. *J Am Acad Dermatol.* 2005;53:791-797.

A11. Gollnick H, Blume-Peytavi U, Szabo EL, et al. Systemic isotretinoin in the treatment of rosacea-doxycycline- and placebo-controlled, randomized clinical study. *J Dtsch Dermatol Ges.* 2010;8:505-515.

A12. Kimball AB, Kerdel F, Adams D, et al. Adalimumab for the treatment of moderate to severe hidradenitis suppurativa: a parallel randomized trial. *Ann Intern Med.* 2012;157:846-855.

A13. Grant A, Gonzalez T, Montgomery MO, et al. Infliximab therapy for patients with moderate to severe hidradenitis suppurativa: a randomized, double-blind, placebo-controlled crossover trial. *J Am Acad Dermatol.* 2010;62:205-217.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Yim D, Curtis N, Cheung M, et al. An update on Kawasaki disease II: clinical features, diagnosis, treatment and outcomes. *J Paediatr Child Health*. 2013;49:614-623.
2. Marzano AV, Vezzoli P, Berti E. Skin involvement in cutaneous and systemic vasculitis. *Autoimmun Rev*. 2013;12:467-476.
3. Schmidt E, Zillikens D. Pemphigoid diseases. *Lancet*. 2013;381:320-332.
4. Kershenovich R, Hodak E, Mimouni D. Diagnosis and classification of pemphigus and bullous pemphigoid. *Autoimmun Rev*. 2014;13:477-481.
5. Intong LR, Murrell DF. Inherited epidermolysis bullosa: new diagnostic criteria and classification. *Clin Dermatol*. 2012;30:70-77.
6. Yost JM, Hale CS, Meehan SA, et al. Dermatitis herpetiformis. *Dermatol Online J*. 2014;20.
7. Huang YH, Kuo CF, Chen YH, et al. Incidence, mortality, and causes of death of patients with pemphigus in Taiwan: a nationwide population-based study. *J Invest Dermatol*. 2012;132:92-97.
8. Sinha AA, Hoffman MB, Janicke EC. Pemphigus vulgaris: approach to treatment. *Eur J Dermatol*. 2014 [Epub ahead of print].
9. Mockenhaupt M. Stevens-Johnson syndrome and toxic epidermal necrolysis: clinical patterns, diagnostic considerations, etiology, and therapeutic management. *Semin Cutan Med Surg*. 2014;33:10-16.
10. Schulenburg-Brand D, Katugampola R, Anstey AV, et al. The cutaneous porphyrias. *Dermatol Clin*. 2014;32:369-384.
11. Ikizoglu G. Red face revisited: Flushing. *Clin Dermatol*. 2014;32:800-808.
12. Rochet NM, Chavan RN, Cappel MA, et al. Sweet syndrome: clinical presentation, associations, and response to treatment in 77 patients. *J Am Acad Dermatol*. 2013;69:557-564.

REVIEW QUESTIONS

1. Treatment of severe nodulocystic acne with which of the following vitamin derivatives may completely arrest the disease process through decreasing *Propionibacterium acnes*?

A. Vitamin D
B. Vitamin A
C. Vitamin K
D. Vitamin E
E. Vitamin B3

Answer: B 13-cis-Retinoic acid (isotretinoin/Accutane) is a derivative of vitamin A. In acne, it rapidly suppresses sebum production, causing a decrease in *P. acnes* populations. It also decreases follicular plugging. Vitamins D, K, and E are fat-soluble vitamins that do not completely arrest the disease process in acne.

2. The most common extracutaneous complications of varicella zoster virus is

A. lymphoreticular.
B. musculoskeletal.
C. cardiovascular.
D. central nervous system.
E. postherpetic neuralgia.

Answer: E Zoster usually resolves without sequelae in children and young adults with intact immune systems. However, the pain, cutaneous eruption, and complications of zoster become more severe with increasing age and immune compromise. Complications of zoster include postherpetic neuralgia, secondary bacterial infection, scarring, ophthalmic zoster, Ramsay-Hunt syndrome, meningoencephalitis, motor paralysis, pneumonitis, and hepatitis.

3. What is the following is the best first line choice for treatment of herpes zoster?

A. Ganciclovir
B. Foscarnet
C. Valacyclovir
D. Gabapentin
E. Cidofovir

Answer: C First-line treatment of herpes zoster (shingles) is with valacyclovir, acyclovir, or famciclovir. Gabapentin may be used for postherpetic neuralgia. Cidofovir and ganciclovir are reasonable treatment options for cytomegalovirus. Foscarnet is used to treat acyclovir resistant herpes simplex infections.

4. What is the most common cause of erythema multiforme?

A. Herpes simplex virus
B. Mycoplasma pneumonia
C. Amoxicillin
D. Ibuprofen
E. Cytomegalovirus

Answer: A The most common cause of erythema multiforme is herpes simplex virus, which may not be active at the time of the erythema multiforme eruption. Patients with recurrent erythema multiforme are typically treated with acyclovir or valacyclovir. Mycoplasma pneumonia is a less common cause of erythema multiforme. Amoxicillin, ibuprofen, and cytomegalovirus may cause erythema multiforme but are much less commonly associated with it than is herpes simplex virus.

440 URTICARIA, DRUG HYPERSENSITIVITY RASHES, NODULES AND TUMORS, AND ATROPHIC DISEASES

MADELEINE DUVIC

URTICARIA

DEFINITION AND EPIDEMIOLOGY

Urticaria, also known as hives, is one of the most common cutaneous reaction patterns (Fig. 440-1). It is triggered by a wide variety of antigens or by physical stimuli, including cold, pressure, and sunlight (Table 440-1). The term *urticaria* refers to a disease spectrum ranging from simple wheals to angioedema. Clinical distinction between acute and chronic urticaria is important for diagnosis and treatment. Chronic urticaria, defined as urticaria that recurs over a period of 6 weeks or more, is often of unknown cause.

Urticaria is common worldwide in persons of all ages, although certain types of urticaria have a predilection for certain age groups. For example, whereas acute urticaria is often seen in children with atopic dermatitis, chronic urticaria peaks in the fourth decade.

PATHOBIOLOGY

Urticaria can be caused by immunologic (autoimmune, immunoglobulin E [IgE] –dependent, immune complex–mediated, complement-kinin dependent) or nonimmunologic (direct mast cell–releasing agents, vasoactive stimuli, drugs) reactions. Local degranulation of mast cells with the release of histamine and other factors, such as slow-reacting substance of anaphylaxis, precipitates urticaria. Functional IgG autoantibodies, which release histamine from mast cells and basophils, are commonly found in the blood of patients with chronic urticaria. Additionally, basophils are recruited into the wheals, where they sustain the response by releasing histamine. Eosinophils also contribute through leukotriene C_4 (LTC_4), leukotriene D_4 (LTD_4), leukotriene E_4 (LTE_4), and major basic protein. Urticaria is typically transient and self-limited, without leakage of blood cells into the skin or damage to the blood vessels. Leakage of plasma into the dermis from capillaries and small postcapillary venules correlates clinically with development of a demarcated, pink, raised lesion (hive).

CLINICAL MANIFESTATIONS

Urticarial lesions are pink to light red, blanch with pressure, and are raised above the surface of the skin. The center of the lesions may be paler than the leading edge. By definition, individual wheals come and go within 24 hours. A mosquito bite (Chapter 253) is the archetypal urticarial lesion. Individual hives can coalesce into giant plaques or annular rings called giant urticaria as is seen in serum sickness, in which they are accompanied by arthralgias and fever. Confluent urticaria may also be accompanied by swelling of the underlying soft tissue or the mucous membranes (angioedema) as well as by anaphylaxis with laryngeal edema, a life-threatening emergency. In otherwise normal individuals, pressure or writing on the skin will cause spontaneous local release of histamine, which induces a wheal-and-flare reaction known as dermatographism. Deep swelling develops at the site of sustained pressure and may remain for days in a condition known as delayed pressure urticaria. Affected individuals may develop lesions from tight-fitting clothing, shoes, socks, or sexual intercourse.

Other physical stimuli such as cold, heat, sun, or exercise may induce urticaria. Cold urticaria may be precipitated by putting an ice cube on the skin; the interval until hives develop and the duration of the hives correlate with the severity of the condition, which can be life threatening if the patient is suddenly immersed in cold water. Heat, exercise, or exertion may be accompanied by small, 2- to 3-mm urticarial lesions in a condition called cholinergic urticaria. Exercise-induced anaphylaxis may be hereditary, but the defect is unknown. Patients with vibratory angioedema develop swelling and erythema within a few minutes of exposure to a vibratory stimulus. Lesions persist for about 30 minutes. Other forms of physically induced urticaria include solar urticaria and aquagenic urticaria (urticarial lesions caused by exposure to sun and water, respectively).

Food and exercise–induced anaphylaxis is a syndrome in which a few minutes of exercise after ingestion of specific foods results in angioedema or anaphylaxis. The cause of this syndrome is still controversial, but reduced gastric acid secretion may be involved in food and exercise-induced anaphylaxis.

DIAGNOSIS

A detailed history (duration, occupation, medications, frequency of episodes, associated illness) is critical. Urticaria typically results from exposure to antigen only minutes to a few hours before the onset of the lesions. In many cases, pruritus may precede onset of the rash. The most common triggers of IgE-mediated allergic urticarial reactions are drugs (especially penicillin, sulfa drugs, antibiotics, and contrast dye), foods (shellfish, salicylates in berries, tomatoes, yeast, and penicillin in blue cheese), food additives (sodium benzoate), nuts (especially peanuts), latex, and insect bites.

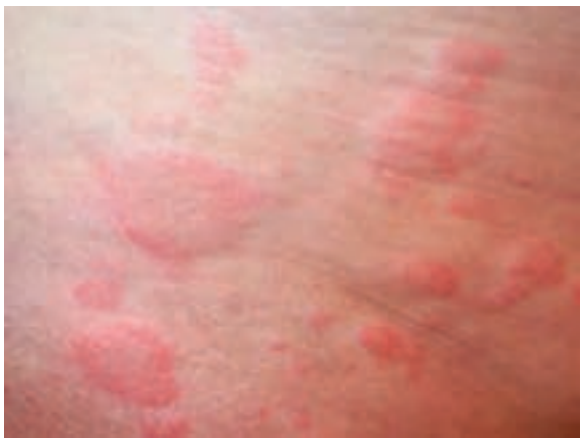

FIGURE 440-1 Urticaria. (From DermNet. Urticaria; 2014. http://www.dermnetnz.org/reactions/urticaria.html. Accessed October 23, 2014.)

TABLE 440-1 COMMON CAUSES OF URTICARIA

URTICARIA MAY BE ACCOMPANIED BY ANGIOEDEMA AND ANAPHYLAXIS
Blood products: red blood cells, platelets, gamma globulin
Drugs
Antibiotics: penicillins, cephalosporins, sulfonamides, isoniazid
Aspirin: salicylates, benzoates, phenylbutazone
Anticonvulsants: hydantoin
Chemotherapy: doxorubicin, daunorubicin, L-asparaginase, chlorambucil, cyclophosphamide, melphalan, methotrexate, nitrogen mustard, procarbazine
Dextran
Nonsteroidal anti-inflammatory drugs
Opiates
Quinidine
Radiocontrast dyes, iodine
Environmental: animal dander or proteins, formaldehyde, pollen, mold, plants, latex, plastic tubing, exercise, heat, cold, sunlight
Foods: berries, eggs, milk, nuts, tomatoes, shellfish, soy
Food additives: sodium benzoate, tartrazine (yellow dye #5)
Hormones
Infections: streptococcal, staphylococcal, sinusitis or abscesses, viral hepatitis, Epstein-Barr virus mononucleosis, *Candida* spp.
Insect bites or venom: Hymenoptera, mosquitoes, mites, scabies
Mechanical stimuli (dermographism, vibratory angioedema, delayed-pressure urticaria)
Vaccines
URTICARIA-LIKE ERUPTIONS AND REACTIVE ERYTHEMAS
Erythema multiforme: herpes simplex, DNA viruses, *Mycoplasma pneumoniae,* drugs
Erythema marginatum: streptococcal rheumatic fever
Juvenile rheumatoid arthritis
Erythema chronicum migrans: *Borrelia* spp. infections
Erythema annulare centrifugum: tinea, drugs
Figurate erythemas: erythema repens (often with underlying carcinoma)
Urticaria pigmentosa (mastocytosis)

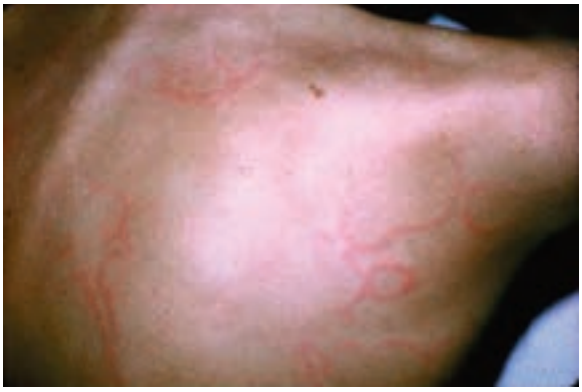

FIGURE 440-2. Erythema marginatum. (From Medscape. Urticaria; n.d.. http://www.medscape.com/content/1998/00/41/73/417394/art-m5649.fig2.jpg. http://www.dermnetnz.org/reactions/urticaria.html. Accessed October 23, 2014.)

Nonimmunologic mediators of urticaria include aspirin and opiates as well as physical agents that work through the prostaglandin pathway or degranulate mast cells.

Acute urticaria can also be triggered by skin contact with an antigen, such as latex, and can progress to anaphylaxis. In addition, urticaria can be a sign or prodrome of a latent infection, especially streptococcal pharyngitis in children or viral hepatitis in adults. The migratory urticarial rash accompanying rheumatic fever, erythema marginatum (Fig. 440-2), is characterized by evanescent, scalloped lesions that change location over the course of hours.

When urticarial lesions are present for more than 24 hours, underlying urticarial vasculitis should be suspected. A skin biopsy is required to distinguish urticarial vasculitis from urticaria in which no damage to the blood vessels is evident. When vascular damage is present, the lesion is termed *leukocytoclastic vasculitis,* the most severe expression of hypersensitivity reactions involving cutaneous blood vessels.

Chronic urticaria can be caused by occult infections (sinusitis, gallbladder disease, *Helicobacter pylori,* yeast infections, tooth abscesses, or silent hepatitis) as well as by collagen vascular diseases and tumors, especially Hodgkin lymphoma. Deficiency of the C1 esterase inhibitor can be manifested as chronic urticaria with angioedema. Allergy testing is recommended if the history is unrevealing. If lesions persist for more than 24 hours, skin biopsy is indicated to determine whether vasculitis or mastocytosis is present. If infection, collagen vascular disease, or a tumor is suspected, a full serologic evaluation should be undertaken. Although a thorough medical evaluation may aid in diagnosis, the cause of chronic urticaria may remain uncertain. In the absence of a known antigen, stress is often invoked as the underlying cause of chronic recurrent idiopathic urticaria.

Systemic mastocytosis (Chapter 255) may be accompanied by urticarial lesions and gastrointestinal symptoms. In the form of mastocytosis known as *urticaria pigmentosa,* stroking the lesions produces urticaria, known as *Darier sign.* A skin biopsy shows an increased number of dermal mast cells. Serum tryptase and histamine levels may be elevated during an attack.

TREATMENT

Rx

Management of urticaria depends on its severity and the duration of the problem (Chapter 252). For mild urticaria limited to the skin, traditional antihistamines (diphenhydramine) or the newer nonsedating agents (terfenadine, cetirizine, loratadine) can be administered by mouth intermittently as needed (Table 440-2). Acute urticaria is often treated with diphenhydramine orally. If the urticaria is severe, short-term corticosteroids, up to 1 mg/kg, can be used. For urticaria associated with wheezing or anaphylaxis, subcutaneous epinephrine, intravenous (IV) corticosteroids, and oxygen should be administered immediately. For chronic urticaria that persists despite antihistamines, cyclosporine at doses of 3 mg/kg or higher for 8 to 16 weeks may be helpful. More recently, omalizumab, an anti-IgE monoclonal antibody, has been shown to significantly reduce symptoms when given as three doses of 150 or 300 mg at 4-week intervals in patients with moderate to severe chronic urticaria. Finding the cause and removing the antigen of chronic recurrent urticaria is highly preferable to chronic administration of corticosteroids or antihistamines.[1] The patient should avoid aspirin compounds and other drugs that could be the cause.

TABLE 440-2 TREATMENT OF URTICARIA

1. Avoid the inciting agent!
2. Medications based on severity
   - A. Mild to moderate, acute urticaria
     Oral antihistamines, e.g., diphenhydramine (Benadryl), 10-50 mg PO q12h, or hydroxyzine, 10-25 mg PO q8h; nonsedating alternatives include cetirizine (Zyrtec), 5-10 mg, or loratadine (Claritin), 10 mg/day
   - B. Severe urticaria with or without angioedema
     Antihistamines, e.g., diphenhydramine (Benadryl), 25-50 mg PO q6-8h or 10-50 mg IV q2-4h, not to exceed 400 mg/24 hr
     Corticosteroids, e.g., prednisone, 10-60 mg PO every morning with tapering over a 2-wk period; triamcinolone (Kenalog), 40 mg IM for one dose; or dexamethasone, 0.6-0.75 mg/m^2/day IV in divided doses q6-12h, depending on severity
   - C. Anaphylaxis
     A—Airway (intubation)
     B—Breathing (oxygen)
     C—Circulation: parenteral aqueous epinephrine, 1 : 1000 IV, saline or volume expanders
     IV corticosteroids (e.g., methylprednisolone, 125 mg)
     Histamine H_1- and H_2-antagonists (50 mg each of diphenhydramine and ranitidine)
   - D. Chronic idiopathic urticaria—combination therapy
     Nonsedating antihistamine: cetirizine, 10 mg/day, or fexofenadine, 30-180 mg twice daily, alone or with montelukast, 10 mg/day, or H_1 and H_2 antagonists (50 mg each of diphenhydramine and ranitidine) and/or low-dose corticosteroids (if unavoidable)

IV, Intravenous; *PO,* oral; *q,* every.

DRUG RASHES

DEFINITION

Drugs have been associated with every type of cutaneous reaction pattern ranging from mild and self-limited to severe and life threatening. Urticaria and exanthematous eruptions are common manifestations of cutaneous drug reactions. Less commonly seen are fixed drug, lichenoid, pustular, phototoxic, bullous, or vasculitic reactions, as well as Stevens-Johnson syndrome and toxic epidermal necrolysis.

PATHOBIOLOGY

Drug rashes, which result from drug toxicity, overdose, drug–drug interactions, or products of metabolism, may be caused by immunologic or nonimmunologic mechanisms. Drugs or their metabolites can act as haptens and induce cell-mediated or humoral responses. Mechanisms include IgE-dependent anaphylaxis and urticaria, cytotoxic reactions resulting in thrombocytopenia and resultant petechiae, immune complex–mediated serum sickness, and delayed-type hypersensitivity resulting in exanthematous or fixed drug eruptions or Stevens-Johnson syndrome.

CLINICAL MANIFESTATIONS AND DIAGNOSIS

Most drug rashes are either immediate (urticaria) or delayed hypersensitivity reactions (exanthems). Immediate reactions such as pruritus, hives, angioedema, and anaphylaxis occur within minutes to a few hours after the drug is taken. The most common drug-related rash (Table 440-3) is a T cell–mediated hypersensitivity reaction manifested as a macular, bright pink to salmon-colored exanthem that appears as early as 7 to 10 days and as late as 14 days after a drug is first administered. Delayed hypersensitivity reactions can be macular or papular exanthems (or both), morbilliform eruptions, annular erythema, or confluent erythema (Fig. 440-3). After sensitization to a particular drug has occurred, readministration of the same drug may trigger an eruption within 24 to 72 hours. Drug hypersensitivity reactions are typically symmetrical. They characteristically begin on the face and upper trunk and progress to the lower extremities, where they may become purpuric. Exanthems secondary to drugs most often become confluent erythematous patches after several days.

Pruritus is the most common symptom. The differential diagnosis for drug rashes includes viral exanthems (Chapter 439), graft-versus-host disease or the leukocyte recovery rash after allogeneic bone marrow transplantation, erythematous exanthems that accompany streptococcal (scarlet fever

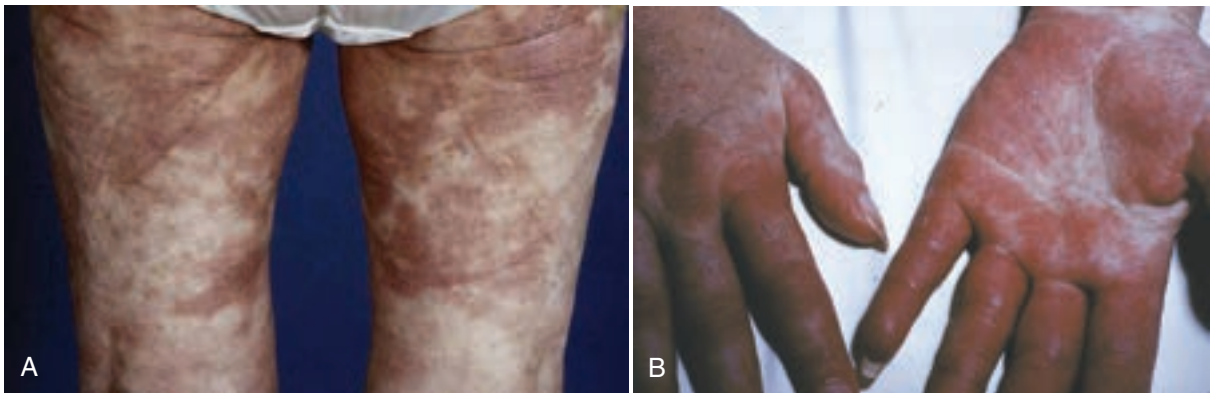

FIGURE 440-3. Delayed hypersensitivity reaction. A, Drug reaction. B, Acral erythema.

TABLE 440-3 DELAYED HYPERSENSITIVITY DRUG RASHES BY CATEGORY

MACULOPAPULAR EXANTHEMS—ANY DRUG CAN PRODUCE A RASH 7-10 DAYS AFTER THE FIRST DOSE
Allopurinol Antibiotics: penicillin, sulfonamides Antiepileptics: phenytoin, phenobarbital Antihypertensives: captopril, thiazide diuretics Contrast dye: iodine Gold salts Hypoglycemic drugs Meprobamate Phenothiazines Quinine
DRUG RASH WITH EOSINOPHILIA AND SYSTEMIC SYMPTOMS (DRESS)
Anticonvulsants: phenytoin, phenobarbital, valproate, lamotrigine Antibiotics: sulfonamides, minocycline, dapsone, ampicillin, ethambutol, isoniazid, linezolid, metronidazole, rifampin, streptomycin, vancomycin Antihypertensives: amlodipine, captopril Antidepressants: bupropion, fluoxetine Allopurinol Celecoxib Ibuprofen Phenothiazines
ERYTHEMA MULTIFORME/STEVENS-JOHNSON SYNDROME
Sulfonamides, phenytoin, barbiturates, carbamazepine, allopurinol, amikacin, phenothiazines Toxic epidermal necrolysis: same as for erythema multiforme but also acetazolamide, gold, nitrofurantoin, pentazocine, tetracycline, quinidine
ACUTE GENERALIZED EXANTHEMIC PUSTULOSIS
Antibiotics: penicillins, macrolides, cephalosporins, clindamycin, imipenem, fluoroquinolones, isoniazid, vancomycin, minocycline, doxycycline, linezolid Antimalarials: chloroquine, hydroxychloroquine Antifungals: terbinafine, nystatin Anticonvulsants: carbamazepine Calcium-channel blockers Furosemide Systemic corticosteroids Protease inhibitors
COLLAGEN VASCULAR OR LUPUS-LIKE REACTIONS
Procainamide, hydralazine, phenytoin, penicillamine, trimethadione, methyldopa, carbamazepine, griseofulvin, nalidixic acid, oral contraceptives, propranolol
ERYTHEMA NODOSUM
Oral contraceptives, penicillin, sulfonamides, diuretics, gold, clonidine, propranolol, opiates Fixed drug reactions: phenolphthalein, barbiturates, gold, sulfonamides, meprobamate, penicillin, tetracycline, analgesics

[Chapter 290]) or staphylococcal (toxic shock syndrome [Chapter 288]) infections, and the acute manifestation of collagen vascular diseases.[2] A similar exanthem occurs when ampicillin is administered to patients who have infectious mononucleosis. A careful drug history is critical in diagnosis and treatment.

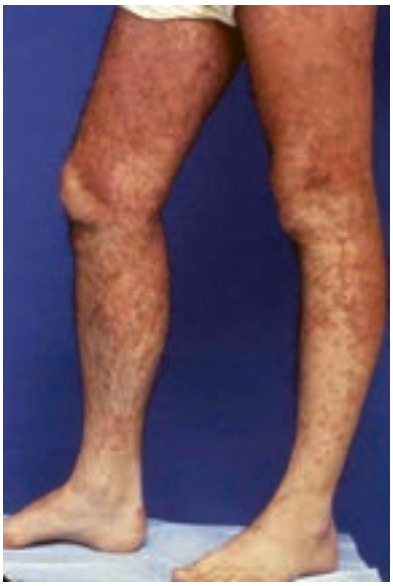

FIGURE 440-4. Hypersensitivity drug rash caused by phenytoin.

TREATMENT AND PROGNOSIS

Rx

After the offending drug is discontinued, delayed hypersensitivity reactions resolve in about 1 week. Therapy is mostly supportive. Corticosteroids, such as 0.01% triamcinolone cream, applied several times per day to the affected area and antihistamines given orally three to four times daily are helpful in reducing the itching and shortening the course.

Specific Syndromes

DRUG RASH WITH EOSINOPHILIA AND SYSTEMIC SYMPTOMS

An especially severe hypersensitivity *drug rash with eosinophilia and systemic symptoms* (DRESS) is most frequently seen with sulfonamides and anticonvulsants (Fig. 440-4). This condition is thought to be caused by an alteration in drug metabolism. Activated T lymphocytes release interleukin-5 (IL-5), leading to characteristic eosinophilia. DRESS may be delayed in onset by 2 to 6 weeks, persists longer than classic drug-induced eruptions, and becomes generalized and severe even when use of the agent is discontinued. It typically begins as a morbilliform eruption, which later evolves into an edematous pustular eruption with erythroderma and purpura. The rash begins on the face, upper trunk, and extremities and later becomes generalized.[3] Continued administration of the drug can result in exfoliative erythroderma; toxic necrolysis; and systemic hypersensitivity, including hepatitis (50%), nephritis (10%); or atypical lymphocytosis and lymphadenopathy mimicking mononucleosis or T-cell lymphoma. Less commonly seen are pneumonitis,

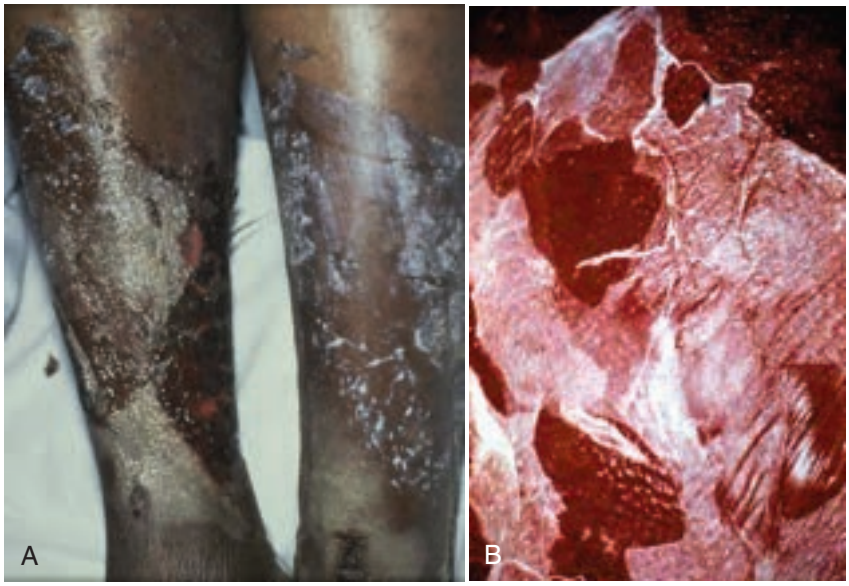

FIGURE 440-5. **Toxic epidermal necrolysis. A,** Clinical appearance. **B,** Close-up appearance of epidermal sheets.

myocarditis, and pericarditis. With visceral involvement, there is a 10% mortality rate, usually from hepatic failure. The initial step in management is the immediate withdrawal of the suspected drug. Systemic corticosteroids (e.g., oral prednisone at 1.0 mg/kg/day and tapered over 3 to 6 months or IV methylprednisolone at 30 mg/kg for 3 days) should be started as early as possible. For patients who develop exfoliative dermatitis, admission to a specialized unit such as a burn unit or intensive care unit is critical.

ERYTHEMA MULTIFORME, STEVENS-JOHNSON SYNDROME, AND TOXIC EPIDERMAL NECROLYSIS

Stevens-Johnson syndrome and toxic epidermal necrolysis represent a spectrum of the same disease, of which erythema multiforme (see Fig. 439-10 in Chapter 439) is the least severe. Stevens-Johnson syndrome is defined as less than 10% body surface area involvement, and toxic epidermal necrolysis is defined as greater than 30% body surface area involvement. Erythema multiforme, a hybrid of urticaria and vasculitis, consists of symmetrically distributed red macules or papules that evolve into classic target or bull's-eye lesions with deep red centers and pink urticarial rims. It is commonly precipitated by herpes simplex infections, other DNA viruses, or drugs. Stevens-Johnson syndrome is characterized by severe mucosal involvement with purpuric lesions. Widespread epidermal necrosis resulting from cell apoptosis is seen in toxic epidermal necrolysis.[4] Drugs are almost always implicated when Stevens-Johnson syndrome or toxic epidermal necrolysis develops in adults (Fig. 440-5). Commonly implicated medications include nonsteroidal anti-inflammatory drugs (NSAIDs), acetaminophen, allopurinol, phenytoin, and sulfa drugs. Alterations in drug metabolism (i.e., slow acetylators of sulfonamides) are often implicated. Full-thickness keratinocyte necrosis or cell apoptosis leads to separation at the dermal-epidermal junction.

Each of these conditions may start as a morbilliform drug rash and progress. Symptoms include fever, severe pain, or sometimes asthenia. The condition can progress rapidly, so it is critical to determine and discontinue the causative agent immediately. Superinfection and fluid and electrolyte imbalances can lead to death in 5% of cases of Stevens-Johnson syndrome and 30% of cases of toxic epidermal necrolysis.

Management includes supportive care such as fluid and electrolyte replacement, transfer to a burn unit, and ophthalmologic evaluation. The use of corticosteroids in Stevens-Johnson syndrome and toxic epidermal necrolysis remains controversial. In Stevens-Johnson syndrome, corticosteroids (e.g., IV methylprednisolone at 60 mg every 6 hours or 1-2 mg/kg for a short course) are frequently used and may decrease the duration of fever and slow eruptions. In toxic epidermal necrolysis, however, retrospective studies suggest that corticosteroids may increase mortality. IV immunoglobulin and tumor necrosis factor-α inhibitors, such as infliximab, may reduce the severity of toxic epidermal necrolysis, but no randomized control trials have been done to date.

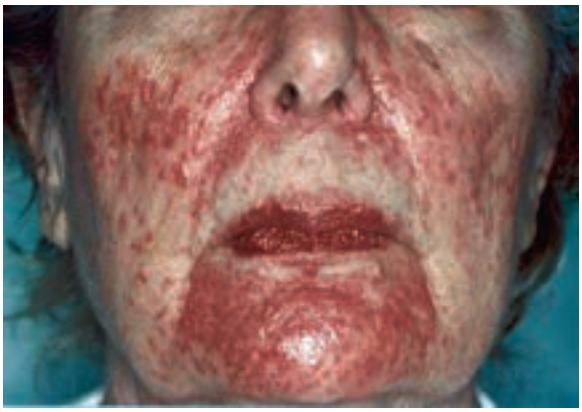

FIGURE 440-6. **Epidermal growth factor receptor inhibitor–associated rash.**

LEUKOCYTOCLASTIC VASCULITIS

Severe drug reactions can also be manifested as vasculitis, neutrophilic eruptions, and ulcerations. Vasculitis is further categorized by the size of the involved vessel and the nature of the cellular reaction and immune complexes. Leukocytoclastic vasculitis, which is the most common form of vasculitis induced by drugs, is manifested as palpable purpura, usually on the lower extremities (see Figs. 439-4 and 439-5 in Chapter 439).[5]

NEUTROPHILIC DRUG REACTIONS

Neutrophilic drug reactions include iododermas, bromodermas, acute generalized exanthematous pustulosis, and acneiform folliculitis. Sweet syndrome (acute febrile neutrophilic dermatosis; see later) also can be drug related. Epidermal growth factor receptor inhibitors and protein kinase inhibitors cause an acneiform facial or chest eruption that is associated with drugs (Fig. 440-6).

Acute generalized exanthematous pustulosis is characterized by numerous (>100), small (<5 mm), nonfollicular subcorneal pustules that arise on erythematous skin, often beginning in skin creases or on the face. High fever and peripheral neutrophilia may precede or accompany the eruption. The pustules are sterile, present for 5 to 10 days, and followed by desquamation. They usually appear less than 2 days after the administration of the causative drug. Ninety percent of cases are due to drugs, most commonly β-lactam antibiotics, macrolides, and calcium channel blockers. This syndrome has also been called pustular drug rash, pustular psoriasis after corticosteroid withdrawal, and toxic pustuloderma. When severe, it may be confused with toxic epidermal necrolysis, but the mortality rate is only 1% to 2%. Skin patch testing is frequently positive.

TABLE 440-4 DRUGS ASSOCIATED WITH SUN SENSITIVITY

PHOTOTOXIC
Chlorpromazine
Hydralazine
Levaquin
Procainamide
Psoralens
Porphyrins
Sulfonamides
Tetracyclines
Thiazide diuretics
PHOTOALLERGIC
Chlorothiazide
Griseofulvin
Hypoglycemic drugs
Promethazine

FIXED DRUG ERUPTIONS

A fixed drug reaction appears at the same location 1 to 2 weeks after first drug exposure and within 24 hours of repeat exposure. The lips, hands, face, feet, and genitalia are most commonly involved. The lesion may begin as erythema and then become gray, brown, or violaceous. Trimethoprim–sulfamethoxazole, NSAIDs, tetracyclines, and pseudoephedrine are common causes.

PHOTOSENSITIVITY AND WITHDRAWAL REACTIONS

Light combined with drugs (Table 440-4) can produce photosensitivity reactions that can be quite severe and mimic sunburn. Ultraviolet radiation interacts with a drug or its metabolite to generate reactive oxygen species, leading to cellular damage. Tetracyclines, sulfa drugs, NSAIDs, and fluoroquinolones are often implicated in photosensitivity reactions. Photosensitizing drugs may exacerbate lupus erythematosus (Chapter 266) or porphyria cutanea tarda (Chapter 210).

CONTACT DERMATITIS

Allergic contact dermatitis is a T cell–mediated delayed hypersensitivity reaction that occurs after topical drug application or exposure to poison ivy, oak, or sumac. It is manifested by erythema and microvesiculation and may spread beyond the area of application (id reaction). Common contact sensitizers include Neosporin (polymyxin B, neomycin, and bacitracin), bacitracin, diphenhydramine, doxepin, lidocaine, lanolin, mercury, henna, ethyl cyanoacrylate (eyelash adhesive), nickel, hair dyes, latex, and *p*-aminobenzoic acid.

BENIGN PAPULES, NODULES, AND TUMORS

The skin is heterogeneous, composed of epidermis, dermis, subcutaneous compartments, and blood vessels. The skin hosts a number of migrating cells (Chapter 435), all of which can give rise to benign or malignant tumors. Lesions that arise from epidermal keratinocytes are usually papules (warts, sebaceous hyperplasia) or plaques (psoriasis, Bowen disease). Nodules are deeper lesions that may be tender or asymptomatic and single or multiple, and they break down to form ulcers. Nodules are classified as inflammatory (granulomas, vasculitis, or panniculitis), infectious, vascular, or metabolic; they can be benign or malignant tumors that arise from skin cells or migrant cells (Table 440-5). Nodules that are smaller and symmetrical are more likely to be benign than lesions that grow rapidly, are larger, or invade surrounding tissue. Any rapidly changing skin nodule should be investigated with an excisional biopsy to the level of fat and sent for histology as well as for bacterial, fungal, and acid-fast cultures.

Benign Epidermal Tumors

The top layer of skin is the avascular epidermis composed of keratinocytes that undergo apoptosis from the stratum corneum. Melanocytes, Langerhans cells, and inflammatory cells may enter the epidermis. Epidermal stem cells form adnexal organs such as hair follicles and sebaceous, eccrine, and apocrine glands that can give rise to tumors.

ACTINIC KERATOSES

Actinic keratoses are pink, scaly macules composed of sun-damaged keratinocytes and are precursors of in situ squamous cell carcinoma (Bowen disease) or invasive squamous carcinomas (Chapter 203). Actinic keratoses are 0.1 cm to 1.0 cm large and are found on sun-exposed areas such as the forearms, hands, face, and scalp (Fig. 440-7). Lesions with induration, thick crusts, ulceration, or pain are excised for biopsy to exclude invasive squamous cell carcinoma. In randomized studies, ingenol mebutate gel (self-applied to a 25-cm^2 contiguous field once daily for 3 consecutive days for lesions on the face or scalp or for 2 consecutive days for the trunk or extremities) led to complete clearance in 42% of treated patients versus 4% of control participants on the face and scalp and 34% versus 5% for the trunk and extremities.

TABLE 440-5 TUMORS AND NODULES OF THE SKIN

Benign, nonpigmented tumors and nodules
Epidermal: warts, acrochordons, tricholemmomas, sebaceous hyperplasia
Adnexal: epidermal cysts, syringomas, follicular cysts, pilomatricoma, apocrine or eccrine adenomas
Dermal and subcutaneous: lipomas, angiolipomas, neurofibromas, leiomyomas
Benign, pigmented tumors and nodules
Epidermal: seborrheic keratoses
Melanocytic compound nevi (junctional nevi are flat)
Spitz nevus
Blue nevus
Dermatofibromas
Malignant, nonpigmented tumors and nodules
Basal cell carcinoma (nodular, superficial, morpheaform, pigmented)
Squamous cell carcinoma (actinic keratoses, Bowen disease, keratoacanthomas)
Cutaneous T- and B-cell lymphomas
Amelanotic melanomas
Merkel cell carcinomas
Adnexal carcinomas of the sebaceous and apocrine glands
Malignant, pigmented tumors and nodules
Pigmented basal cell carcinoma
Malignant melanoma: in situ, superficial spreading, nodular, acral lentiginous
Dermatofibrosarcoma protuberans
Inflammatory nodules over joints
Gottron papules (dermatomyositis)
Gouty tophi
Heberden nodes (osteoarthritis)
Multicentric reticulohistiocytosis (paraneoplastic syndrome)
Rheumatoid nodules
Granuloma annulare
Inflammatory nodules of the lower extremities
Panniculitis
Vasculitis: periarteritis nodosa
Metabolic nodules of the skin
Amyloidosis
Gouty tophi
Xanthomas, necrobiotic xanthogranuloma
Xanthelasma
Vascular lesions
Benign: nevus flammeus, angiokeratomas, spider hemangiomas, capillary hemangiomas, cavernous hemangiomas, blue rubber bleb nevi, pyogenic granulomas
Malignant: Kaposi sarcoma, angiosarcoma

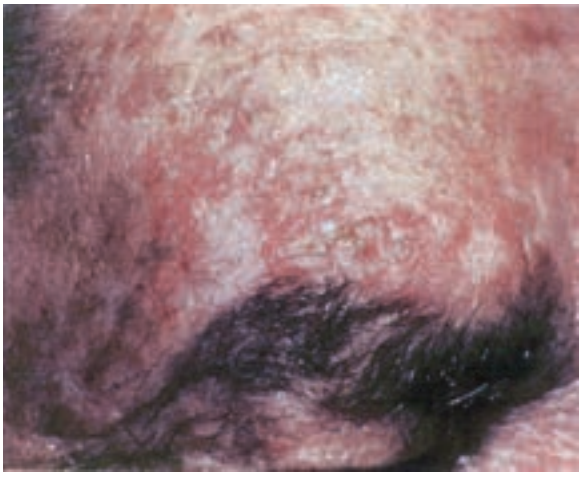

FIGURE 440-7. Actinic keratoses.

Actinic keratoses also can be treated with cryotherapy and topical fluorouracil, retinoids, or imiquimod. To treat and prevent actinic keratoses, sun-exposed areas may be treated topically with 5-fluorouracil cream (5%) applied daily for 2 weeks or twice weekly for 8 weeks.

SEBORRHEIC KERATOSES

Seborrheic keratoses are common verrucous or stuck-on epidermal papules of various colors (Fig. 440-8). They are commonly seen with advancing age but may arise suddenly (sign of Leser-Trélat) in association with internal malignancy. Seborrheic keratoses consist of a single clone of keratinocytes and inherited as an autosomal dominant trait. Seborrheic keratoses have FGFR3, PIK3CA, KRAS, EGFR, HRAS, and AKT mutations but remain clinically benign. Their surface may be friable, and lesions can be scraped off. Seborrheic keratoses spare the palms, soles, and mucosal surfaces. Although benign, seborrheic keratoses must be differentiated from melanocytic nevi, melanomas, and pigmented basal cell carcinoma, usually by the presence of white to yellow horn cysts on their surface, best appreciated with dermoscopy.

WARTY LESIONS

Epidermal papules include common warts (see Fig. 439-2 in Chapter 439) caused by human papillomavirus (HPV). HPV can also be detected in squamous carcinomas arising on the digits and in keratoacanthomas, which are low-grade, well-demarcated, dome-shaped papules or nodules that grow rapidly and spontaneously involute in 6 to 8 weeks. Acrodermatitis verruciformis is characterized by multiple warts with the appearance of seborrheic keratoses on the dorsal extremities and gives rise to squamous cell carcinomas. Molluscum contagiosum (Chapter 439) (Fig. 440-9), caused by a DNA virus, are small, shiny, domed-shaped, 1- to 5-mm papules with a central depression. Molluscum contagiosum lesions are common in children and immunocompromised patients. Treatment with anti-cancer BRAF inhibitors can cause warty papules, keratoacanthomas, and eruptive keratosis pilaris.

Cowden syndrome (Chapter 193), caused by mutations in *PTEN* gene, is associated with warty papules (tricholemmomas; Fig. 440-10); cobblestone papules on the gums and tongue; fibrous papules; and multiple hamartomas involving the breast, thyroid, intestines, ovary, and cerebellum.

ADNEXAL TUMORS

Adnexal tumors arise from hair follicles or glands and are commonly found on the face or scalp. Trichoepitheliomas resemble basal cell carcinomas. Sebaceous hyperplasia consists of small yellow papules with a central depression. Adnexal tumors of sebaceous origin, including sebaceous adenomas and sebaceous carcinomas, can also occur on the face, where they may be markers of the Muir-Torre syndrome of familial breast and colon cancer. Epidermal or sebaceous cysts, which are found in acne or as single firm nodules with a central pore, are filled with sebum or keratin. Epidermoid tumors of the scalp and a family history of colon cancer raises question of a diagnosis of Gardner syndrome (Chapter 193).[6]

DERMATOFIBROMAS AND COLLAGENOMAS

Fibroblasts, the resident cells of the dermis, produce collagen, elastin, and mucopolysaccharides. Accumulation of these products results in sclerosis, papules, or nodules. Fibroblasts in small, dense clusters form firm brown or tan papules known as dermatofibromas. Dermatofibromas are hypertrophic scars that are commonly found on the extremities and that may form after insect bites or trauma. They are firm and well-demarcated papules, and the skin puckers when lateral pressure is applied. Dermatofibromas can be treated, but increased scarring may occur. The malignant counterpart is dermatofibroma sarcoma protuberans, which is a poorly defined, rapidly expanding, dermal, malignant tumor. Overlying erythema or hyperpigmentation is often present.

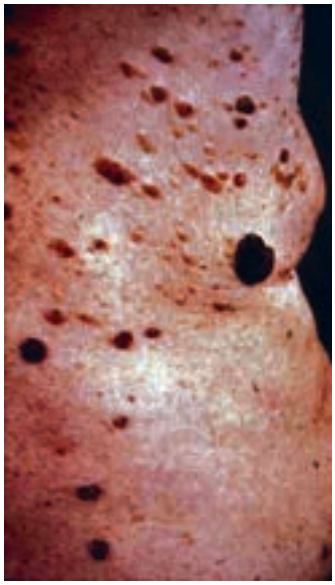

FIGURE 440-8. Seborrheic keratoses.

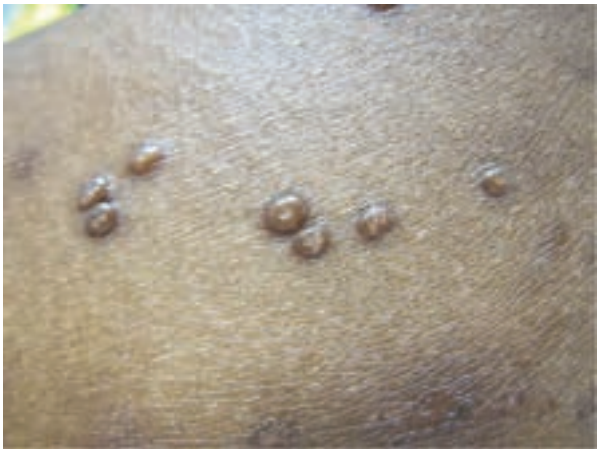

FIGURE 440-9. Molluscum contagiosum.

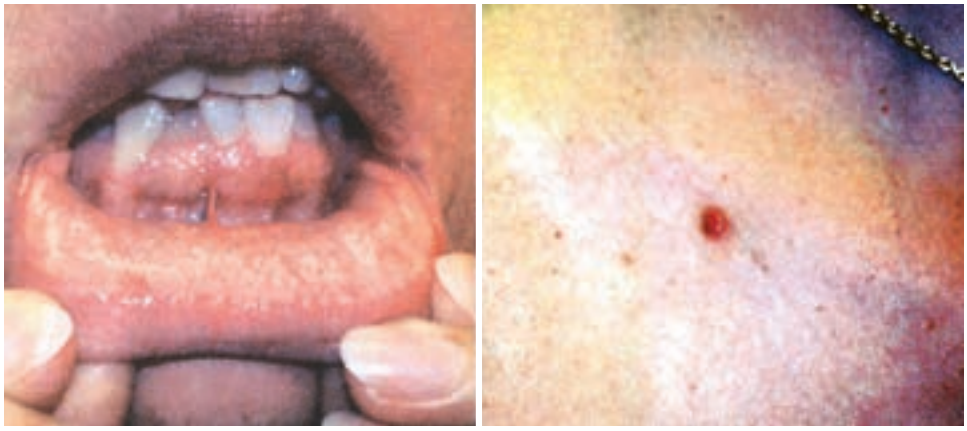

FIGURE 440-10. Cowden syndrome: cobblestone gums *(left)* and tricholemmoma *(right)*.

Some individuals have more pronounced and hypertrophic scar formation known as keloids, with an autosomal dominant or autosomal recessive inheritance pattern. Keloids, which are shiny and firm in appearance, result from an overproduction of collagen. They are especially common on the anterior chest, neck, and earlobes and may require antineoplastic treatment, such as interferon alfa 2b, mitomycin C, bleomycin, and 5-fluorouracil (Chapter 179), and laser therapy.[7]

Collagenomas and elastic tumors with the appearance of small white to yellow papules are found in the skin and bone of patients with Buschke-Ollendorff syndrome. *Pseudoxanthoma elasticum* (Chapter 260), an autosomal recessive disorder, is typically manifested as cutaneous yellow plaques on the neck or antecubital fossa from damaged elastin tissue. Mucin cysts are gray, shiny, well-demarcated, round nodules that generally arise on the mucosa or on the digits, where they may have an underlying connection to the joint space.

NEURAL CREST CELL TUMORS

Benign tumors in the dermis arising from neural crest cells include neurofibromas (soft, flesh-colored papules; Fig. 440-11), schwannomas (larger subcutaneous soft tumors or plaques; Fig. 440-12), and melanocytic lesions. Although solitary neurofibromas may occur, multiple lesions with cafe au lait spots (tan macules) or axillary freckling (Crowe sign) are diagnostic of neurofibromatosis type I, an autosomal dominant disorder caused by mutations in neurofibromin (Chapter 417). Schwannomas can become malignant and be manifested as dermal nodules. Merkel cell carcinoma, which is a neuroendocrine carcinoma of the skin, is a particularly aggressive small cell tumor arising from the cutaneous nerve endings or Meissner corpuscles. Merkel cell polyomavirus is associated with development of Merkel cell carcinoma.[8] Merkel cell carcinoma may present as a solitary pink to purple dome-shaped papule on the head or neck (Fig. 440-13). Sentinel lymph node biopsy is recommended in all patients with primary Merkel cell carcinoma. Treatment requires full excision, radiation therapy, and often chemotherapy because the cancer tends to recur and metastasize.

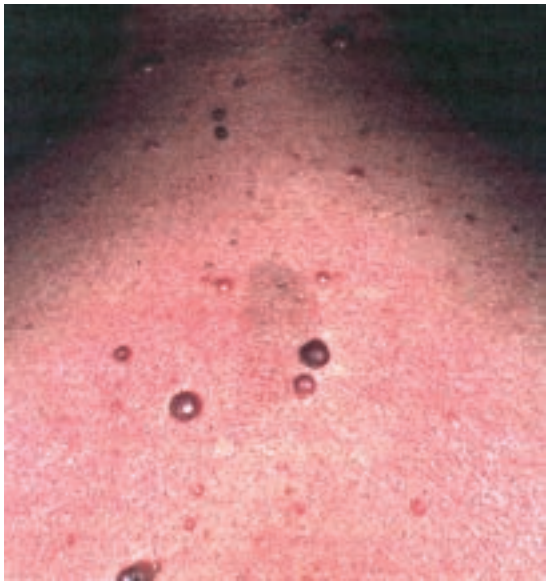

FIGURE 440-11. Neurofibromatosis with cafe au lait spots and neurofibromas.

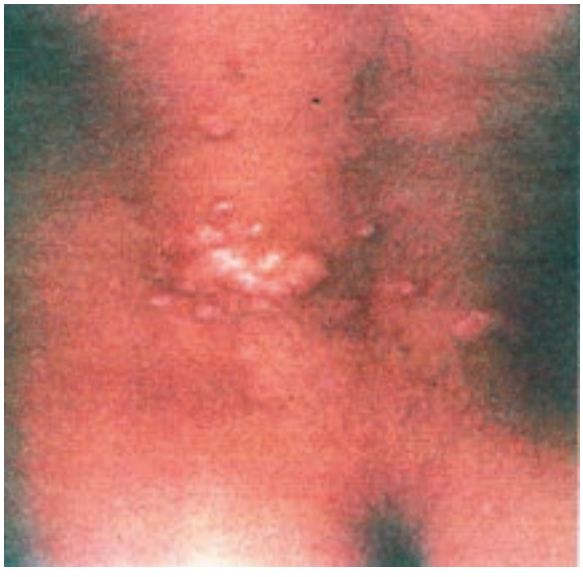

FIGURE 440-12. Schwannoma.

Melanocytic Lesions

Benign melanocytic moles or nevi (new) are discrete nests of melanocytes acquired during childhood and young adulthood, stimulated by sun exposure. Nevi are benign and composed of melanocytes (Chapter 203). They regress with age and change in color during pregnancy. Benign melanocytic nevi are formed by nests of melanocytes at the epidermal junction (junctional nevi), in the dermis (intradermal nevi), or in both compartments (compound nevi). Their appearance depends on type and age of the lesion. Junctional nevi (Fig. 440-14) are small, flat, and light to dark brown. Intradermal nevi are soft, flesh-colored to pink papules with smooth regular borders and surface. Compound nevi are globular papules with brown pigmentation. Whereas blue nevi (Fig. 440-15) are flat, grayish blue, and regular. Small congenital nevi are dark brown, dysplastic nevi (Fig. 440-16) have variegated colors and may transform into melanoma. More than 10 large, atypical moles with irregular borders and colors confer higher risk of developing melanoma, especially with a positive family history. Other recognized risk factors for melanoma include having more than 50 small nevi, red or blonde hair, or fair skin that burns and a history of blistering sunburns as a child. Patients with higher risk for melanoma should have surveillance and regular skin examinations.

Langerhans Cell Histiocytosis

Skin surveillance is mediated by antigen-presenting cells: Langerhans cells, dermal dendritic cells, and skin-homing T lymphocytes. Proliferation of Langerhans cells is called histiocytosis. Childhood histiocytosis X is manifested as severe seborrheic dermatitis of the scalp and gluteal areas with

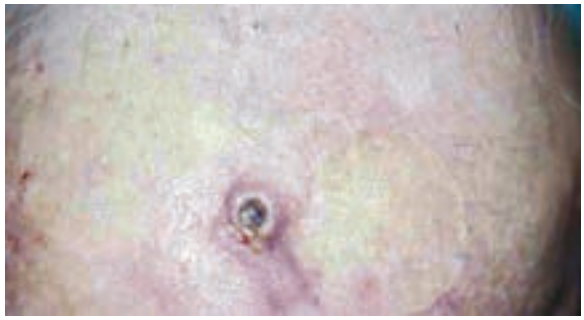

FIGURE 440-13. Merkel cell tumor.

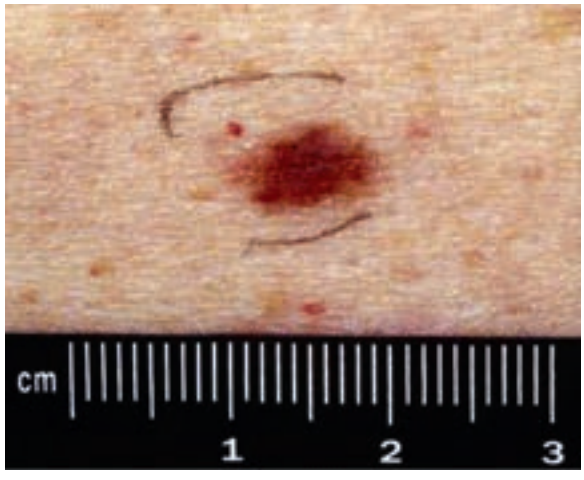

FIGURE 440-14. Junctional nevus.

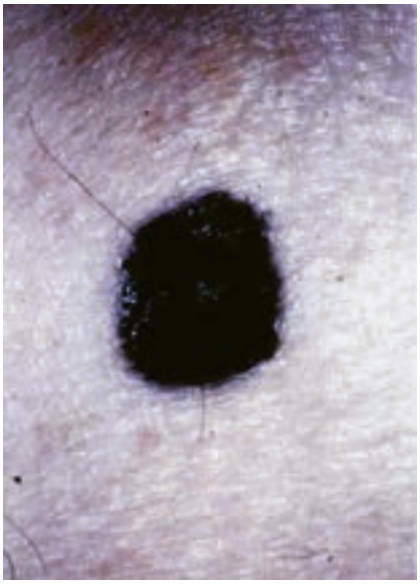

FIGURE 440-15. Benign blue nevus.

FIGURE 440-16. Dysplastic nevus syndrome.

underlying purpura and may result in the hemophagocytic syndrome. In adults, lesions appear in the intertriginous areas (Fig. 440-17). Patients with the non–Langerhans cell histiocytosis have lytic bone involvement (eosinophilic granulomas) or diabetes insipidus (Hand-Schüller-Christian syndrome).

Vascular Lesions

HEMANGIOMAS

Benign capillary, or cherry, hemangiomas are bright cherry-red to purple papules, generally less than 5 mm in diameter. They appear on the trunk with aging and may be numerous (Fig. 440-18). Pyogenic granulomas can resemble hemangiomas but contain polymorphonuclear leukocytes, are friable, and bleed easily. Multiple pyogenic granulomas are seen in infectious bacillary angiomatosis in immunocompromised hosts (Chapter 315). Cavernous or strawberry hemangiomas can also appear in the neonatal period as rapidly growing vascular tumors; they may obstruct the eye or the pharynx before regressing. Propranolol 2 mg/kg/day in two or three divided doses for variable duration can decrease the hemangioma's volume, color, and elevation in children younger than 5 years of age.[■] Corticosteroids, interferon, or antiangiogenic factors also can treat these lesions if propranolol is not successful. Cavernous hemangiomas are deeper and less likely to resolve than smaller lesions. When associated with platelet consumption, Kasabach-Merritt syndrome is present (Chapter 171).

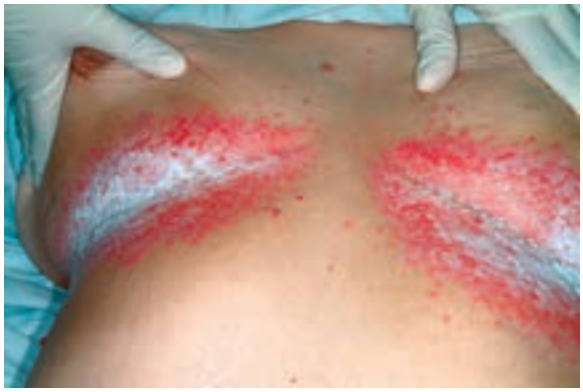

FIGURE 440-17. Histiocytosis X.

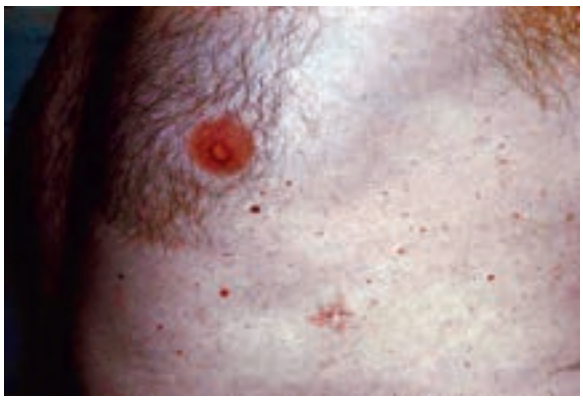

FIGURE 440-18. Benign capillary hemangioma.

KAPOSI SARCOMA

Kaposi sarcoma (Chapter 392) is a disseminated angiomatosis that arises from viral IL-8 production by herpesvirus 8. Lesions are symmetrical purple, red, gray, or brown patches, papules, nodules, or ulcers (Fig. 440-19). Mucosal involvement is more common in advanced disease. Kaposi sarcoma in young African adults and Kaposi sarcoma associated with human immunodeficiency virus (HIV) infection often have a more aggressive course than Kaposi sarcoma in elderly men of Mediterranean background, whose disease is indolent and often confined to the lower extremities. Treatment of HIV disease with highly active antiretroviral therapy has been associated with a marked decreased incidence and severity of HIV-associated Kaposi sarcoma. Angiosarcomas are malignant purple to red vascular tumor nodules that are more common in elderly individuals or on the extremities of patients with chronic lymphedema.[9] To help distinguish cutaneous angiosarcomas from benign histologic mimics, immunohistochemistry staining for ERG, an Ets family transcription factor, is a specific and sensitive marker for endothelial differentiation.

Inflammatory and Hematopoietic Papules and Tumors

Inflammatory diseases of the skin involve the superficial or dermal vessels or subcutaneous tissue. Inflammatory infiltrates can be mixed or restricted in nature. Lymphocytes, polymorphonuclear leukocytes, histocytes, eosinophils, and plasma cells are involved in the most common inflammatory reactions. Hematologic malignancies can present with secondary skin lesions, including patches, nodules, papules, or vasculitic lesions. Skin-homing $CD4^+$ T cells give rise to cutaneous T-cell lymphomas. Mycosis fungoides (Chapter 185) lesions are pleomorphic pink, white, or brown patches or plaques, alopecia, or diffuse erythroderma with blood involvement (Sézary syndrome). Early patch or plaque mycosis fungoides is indistinguishable from chronic eczematous or psoriasiform dermatitis. Tumors occur late in mycosis fungoides and can transform to a large cell lymphoma phenotype, with or without expression of CD30. Peripheral cutaneous T-cell lymphomas may also be found in subcutaneous tissue as panniculitic lesions. Lymphomatoid papulosis is characterized by crops of red to pink self-regressing papules with

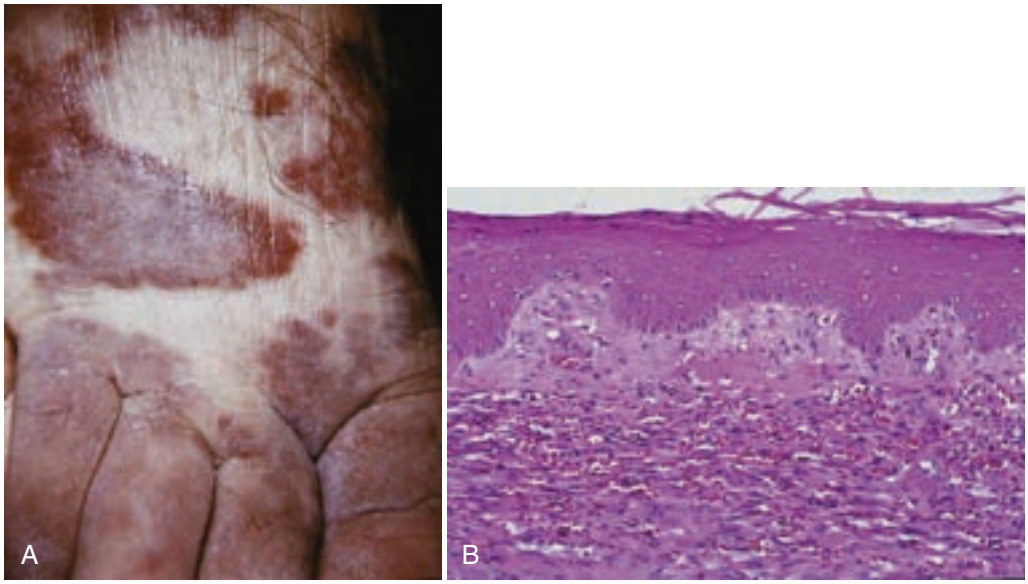

FIGURE 440-19. Kaposi sarcoma. A, Involvement of the lower extremity (Mediterranean Kaposi sarcoma). **B,** Histology.

histologic findings similar to those of anaplastic large cell lymphoma (Chapter 185), including expression of CD30 antigen. Clinically, lymphomatoid papulosis presents as self-regressing papules, $CD3^+$ anaplastic T-cell lymphomas can present as tumors, and patients with transformed $CD30^+$ mycosis fungoides initially have patch or plaque stage disease before developing tumors. Treatment of lymphomatoid papulosis consists of topical corticosteroids, methotrexate, or bexarotene. For $CD30^+$ anaplastic T-cell lymphoma, therapeutic options include radiation therapy, methotrexate, bexarotene, brentuximab vedotin, and chemotherapy (Chapter 185). Natural killer T-cell lymphomas, immunoblastic lymphomas, and plasmacytoid dendritic cell tumors appear as brown to purple dermal nodules often with purpura.

Cutaneous B-cell lymphomas present as pink, infiltrated, dome-shaped shiny papules or tumors. Whereas follicular B-cell lymphoma is commonly located on the face, scalp, or upper part of the back, mucosa-associated B-cell tumors are more common on the trunk. With the exception of large B-cell lymphoma, follicular and MALT (mucosa-associated lymphoid tissue) B-cell lymphomas of the skin are indolent. Cutaneous MALT lymphomas (Fig. 440-20) have been associated with *Borrelia* spp. infection (Chapter 321), *Helicobacter pylori* infection, and chronic inflammation.[10] Plasmacytomas can arise in the skin, in bone, with multiple myeloma (Chapter 187), or independently. Extramedullary hematopoiesis or endometriosis can be associated with red or brown nodules in the dermis.

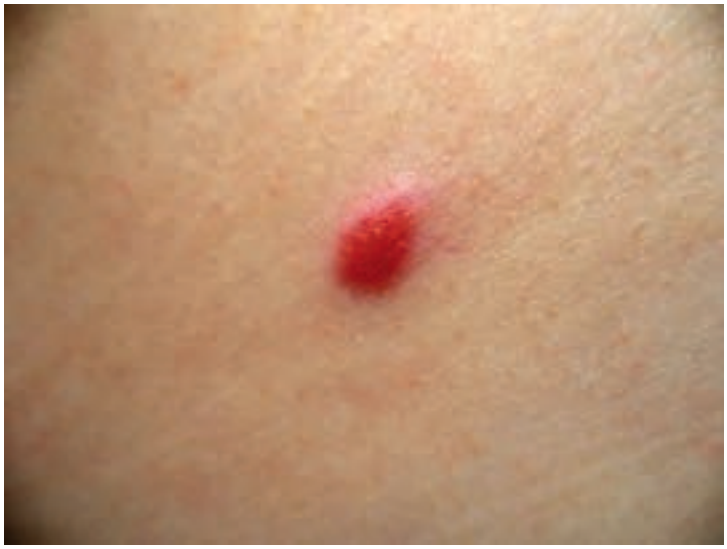

FIGURE 440-20. Cutaneous MALT (mucosa-associated lymphoid tissue) lymphoma.

Granulomatous Diseases

Sarcoidosis is an inflammatory granulomatous process manifested as ichthyosis, papules, plaques, or tumors with an apple-jelly color (Fig. 440-21). Patients with lepromatous leprosy also can have histiocytic plaques or tumors (Fig. 440-22); treatment of leprosy may induce an inflammatory reaction called erythema nodosum leprosum. Granulomatous mycosis fungoides, which is a variant of cutaneous T-cell lymphoma, is difficult to diagnose and treat. Granulomatous inflammation within the dermis can result in damage to collagen, as seen in granuloma annulare (ringlike pink to red infiltrated lesions, often on the hands or elbows), rheumatoid nodules that occur on the extensor surface of the arms, and necrobiosis lipoidica on the shins of patients with diabetes. All three lesions typically include fibrin deposits within dermal blood vessels. Multicentric reticulohistiocytosis is a rare paraneoplastic syndrome in which histiocytic nodules form over joints with associated arthritis.

Inflammatory Skin Lesions and Nodules

Inflammatory skin nodules arise from inflamed blood vessels (vasculitis) or adipose tissue (panniculitis). Either can arise in response to underlying infection or antigen stimulation with influx of inflammatory cells. Vasculitis is categorized by vessel size and circulating immune complexes. Damage to blood vessels results in leakage of red blood cells with the development of purpura (nonblanching red to purple lesions; Chapter 270).

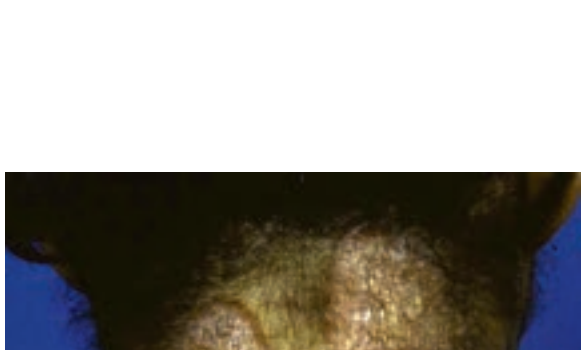

FIGURE 440-21. Cutaneous sarcoidosis.

Sweet syndrome, also called febrile neutrophilic dermatosis (Fig. 440-23), is accompanied by fever, leukocytosis, and tender reddish skin plaques. Some patients also have arthralgia. Biopsy shows sheets of leukocytes filling the upper dermis in the absence of infection. It can be idiopathic; drug induced; or associated with an underlying disease, typically with an underlying streptococcal infection, acute myelogenous leukemia, other malignancies, inflammatory bowel disease, or rheumatoid arthritis.[11] Use of hematopoietic growth factors also can precipitate Sweet syndrome. Sweet syndrome, but not erythema elevatum diutinum, is highly responsive to corticosteroids (oral prednisone 1-2 mg/kg/day gradually tapered over 6 weeks to 3 months) or indomethacin (150 mg per day for 1 week; then 100 mg per day for 2 weeks), but oral dapsone (100-200 mg/day) can improve both conditions.

Erythema elevatum diutinum is manifested as multiple, infiltrated pink, yellow, red, or violaceous nodules or papules that may be painful or asymptomatic. The lesions can coalesce to form gyrate lesions on the dorsum of the hands or extensor surfaces similar to granuloma annulare. Erythema elevatum diutinum is associated with upper respiratory infections (especially *Streptococcus* spp.), HIV infection, and inflammatory bowel disease. Clinically, the lesions look similar to Sweet syndrome, but their underlying histopathology (a necrotizing vasculitis with neutrophils and hyalinization of the vessels) can be distinguished from the neutrophils seen in the upper dermis in Sweet syndrome on biopsy.

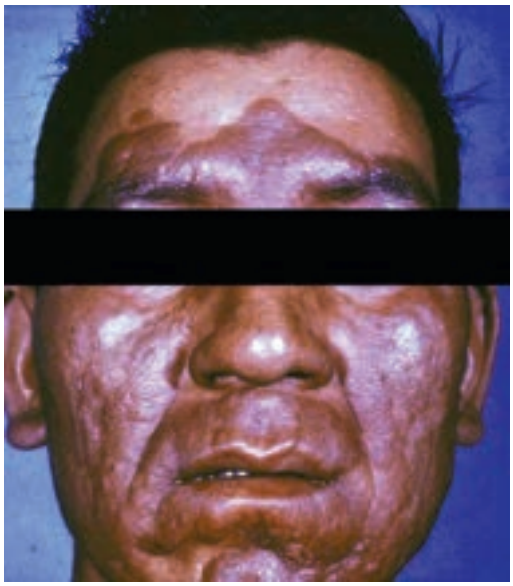

FIGURE 440-22. Leonine facies associated with lepromatous leprosy.

Polyarteritis Nodosa and Panniculitis

Polyarteritis nodosa (Chapter 270) arises in larger arterioles and may be associated with hepatitis C infection, mesenteric aneurysms, cryoglobulinemia, cutaneous ulceration, and livedo reticularis. Polyarteritis nodosa is distinct from small vessel leukocytoclastic vasculitis, which is characterized by smaller areas (a few millimeters) of purpura.

In the clinical setting, *panniculitis* occurs more frequently than nodular vasculitis. The diagnosis of vasculitis versus septal or lobular panniculitis requires an excisional biopsy, including fat, with appropriate cultures and stains.

Erythema nodosum (Fig. 440-24) is a septal panniculitis characterized by tender nodules that are 1 to 2 cm in diameter with warm, pink, overlying epidermis. They appear in crops on the extremities. A perivascular inflammatory infiltrate is present around small intralobular vessels without vasculitis.

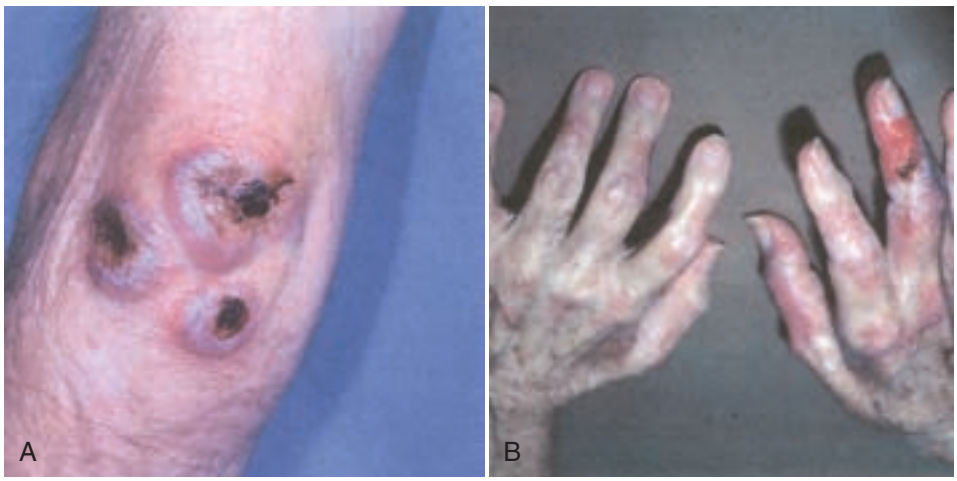

FIGURE 440-23. Sweet syndrome. **A** and **B**, Sweet syndrome in patients with leukemia.

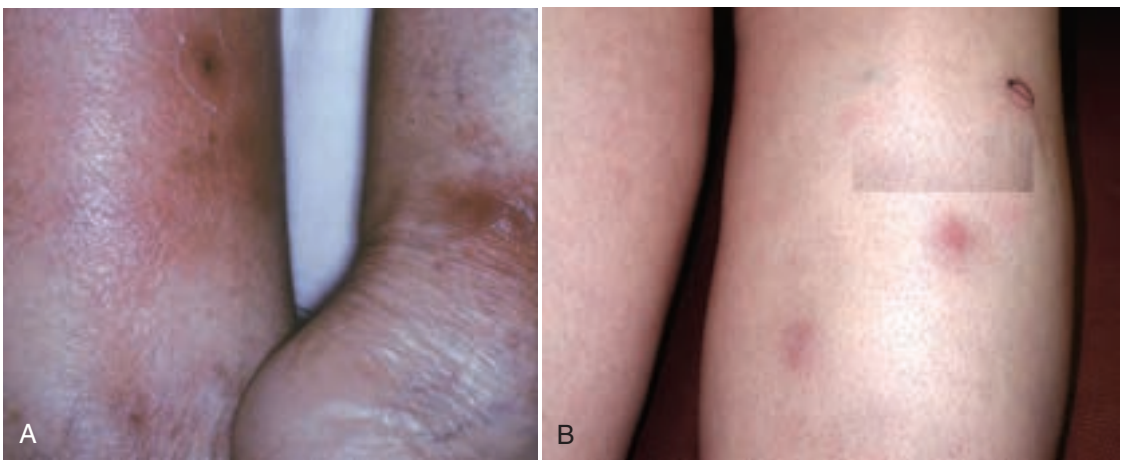

FIGURE 440-24. Erythema nodosum. **A** and **B**, Erythema nodosum and septal panniculitis of the lower extremities.

Erythema nodosum arises frequently in response to sarcoidosis (Chapter 95), various infections, inflammatory bowel disease, or drug use and less commonly in patients with azathioprine-induced pancreatitis or primary biliary cirrhosis. Often, however, the underlying cause remains unknown (Table 440-6).

Lobular panniculitis with necrosis and purpura is called nodular vasculitis or erythema induratum. Nodular vasculitis is characterized by painful, chronic recurrent nodules on the shin or thighs that become bluish, ulcerate, and heal with scarring. Erythema induratum (Fig. 440-25) is exacerbated by cold exposure and is sometimes associated with *Mycobacterium tuberculosis* (Chapter 325). True lobular panniculitis, with or without fat necrosis, is more frequent in men with underlying pancreatitis (Chapter 144) and may precede pancreatic cancer (Chapter 194). The lesions have a predilection for the anterior aspect of the shins and may be fluctuant as a result of fat necrosis. *Lupus panniculitis,* or lupus profundus, which involves the fat, is diagnosed by overlying granular immune complex deposition of IgM along the dermal-epidermal junction and is often difficult to distinguish from subcutaneous panniculitic T-cell lymphoma. Subcutaneous panniculitic γ/δ T-cell lymphoma is more aggressive and has a poor prognosis compared with α/β panniculitic T-cell lymphoma. Lupus panniculitis of the breast, which can be mistaken for adenocarcinoma, is treated with antimalarials or corticosteroids. Lobular panniculitis with calcification of the small arterioles, which occurs in the setting of renal failure with hyperparathyroidism, is called calciphylaxis (Chapter 130). Granulomatous lobular panniculitis may also arise in the setting of schistosomiasis (Chapter 355), Sjögren syndrome (Chapter 268), Crohn disease (Chapter 141), sarcoidosis (Chapter 95), ruptured epidermal cysts, atypical mycobacterial infection (Chapter 325), or tuberculosis (Chapter 324).

Fungal Infections

In immunocompromised patients, necrotic or granulomatous lobular panniculitis can be caused by disseminated fungal infections with *Candida* spp., *Sporothrix schenckii, Cryptococcus* spp., *Histoplasma* spp., *Nocardia* spp., *Rhizopus* spp., *Aspergillus* spp., *Fusarium* spp., or chromomycosis. Fungal mycelia invade vessel walls, where they produce purpuric and painful lesions that may ulcerate. Osler nodes, which are tender nodular vasculitic lesions on the extremities, occur in the setting of bacterial endocarditis (Chapter 76). Staphylococcal or streptococcal sepsis may be manifested as pustules, papules, or panniculitic lesions.

TABLE 440-6 TRIGGER FACTORS ASSOCIATED WITH ERYTHEMA NODOSUM

Infections
Bacterial: *Streptococcus* spp., tuberculosis, leprosy, *Mycoplasma* spp., *Yersinia* spp., *Salmonella* spp., leptospirosis, tularemia
Fungal: coccidioidomycosis, blastomycosis, histoplasmosis, dermatophytosis
Viruses and *Chlamydia:* paravaccinia, Epstein-Barr virus, lymphogranuloma venereum, cat-scratch disease, psittacosis, hepatitis B
Drugs: sulfonamides, bromides, oral contraceptives
Malignancies: lymphoma, leukemia, carcinoma, after tumor radiation
Inflammatory: ulcerative colitis, Crohn disease, Whipple disease, Behçet syndrome, Sweet syndrome, collagen vascular diseases
Pregnancy

ATROPHIC AND SCLEROTIC LESIONS

Atrophic Lesions

Atrophic lesions result from thinning or loss of the epidermal and dermal layers (Table 440-7). Examples are photoaging caused by loss of epidermal thickness and collagen, discoid lupus, and genetic disorders of collagen production (e.g., Ehlers-Danlos syndrome; Fig. 440-26). Epidermal wrinkling can result in a cigarette paper appearance with prominence of the underlying blood vessels. High-potency topical corticosteroids cause loss of collagen,

TABLE 440-7 ATROPHIC SKIN CONDITIONS WITH SCARRING, ULCERATIONS, OR TELANGIECTASES

ATROPHY
Epidermal: chronic corticosteroid use, photoaging, mycosis fungoides
Dermal elastin: anetoderma, cutis laxa, intrinsic aging
Dermal collagen: Ehlers-Danlos syndrome, aging
Subcutaneous: granulomatous slack skin (a mycosis fungoides variant)
Lipodystrophy (loss of fat)
SCARRING OR ATROPHY WITH TELANGIECTASIAS
Discoid and subacute cutaneous lupus erythematosus
Dermatomyositis
Keloid formation
Large plaque parapsoriasis (poikiloderma vasculare atrophicans variant of mycosis fungoides)
Photoaging
Necrobiosis lipoidica diabeticorum
Radiation dermatitis
Porphyrias
Thermal burns (erythema ab igne)
SCLEROSIS OR INFILTRATIVE PROCESSES
Amyloidosis
Systemic sclerosis, scleroderma
Localized sclerosis, morphea
Lichen sclerosis et atrophicus
Lichen myxedematosus or papular mucinosis (mucopolysaccharide deposition with paraproteinemia)
Myxedema (mucin deposits with anti–thyroid-stimulating hormone receptor antibodies)
ULCERATIONS
Secondary breakdown of any blister or nodule: infectious, inflammatory, tumor, vasculitis
Decubitus or pressure ulcers
Genital ulcers: syphilis, herpes simplex, chancroid, lymphogranuloma venereum, Behçet syndrome
Pyoderma gangrenosum, Sweet syndrome

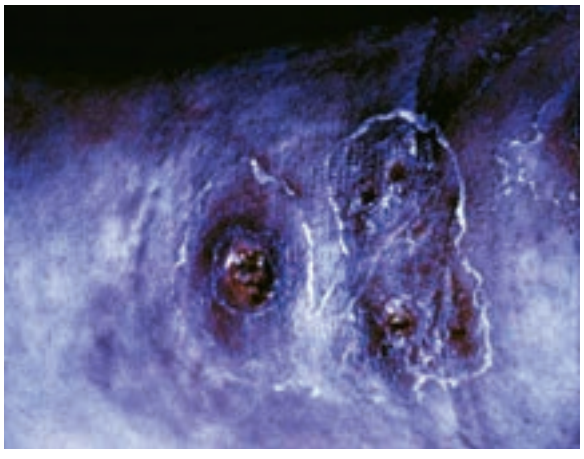

FIGURE 440-25. Erythema induratum.

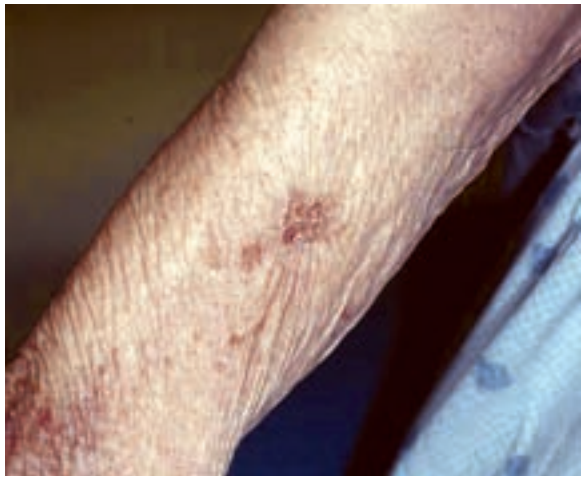

FIGURE 440-26. Atrophic skin in Ehlers-Danlos syndrome type 2.

resulting in atrophy. In Cushing syndrome (Chapter 227), striae appear as red or purple streaks because the underlying dermis can be seen through the epidermis.

Aging skin is most pronounced in sun-exposed areas, but intrinsic aging beginning as early as 30 years of age is characterized by abnormalities in the formation of elastin fibers. Aging of the skin is accompanied by decreased rete ridges and diminished circulation. Sunlight ages the skin by inducing proteolytic enzymes that digest the underlying collagen and elastin (wrinkles). In addition, sun exposure induces pigment incontinence (freckling), increased junctional nevi, and proliferation of benign keratinocyte growths (seborrheic keratoses).

Atrophy can also result from ongoing inflammatory processes that cause scarring, such as collagen vascular disease or mycosis fungoides. The cutaneous and discoid forms of lupus erythematosus (Chapter 266) are manifested as scaly plaques with atrophy or alopecia on sun-exposed areas; the systemic form is characterized by malar rash, urticaria, or vasculitic lesions. Dermatomyositis (Chapter 269) can be associated with collagen vascular disease or malignancy; periorbital suffusion, telangiectasia of the nail beds, and Gottron papules or scaly lesions over the joints are the skin manifestations. Anetodermas are localized sclerotic lesions (Fig. 440-27) with distinctive clinical features from underlying inflammation.

Eosinophilic fasciitis is accompanied by nodules or sclerosis of the lower extremities, myopathy, pulmonary disease, and eosinophilia. This syndrome, which follows the ingestion of L-tryptophan or its contaminants, resembles the panniculitis seen in systemic sclerosis, in which fat lobules are replaced by new collagen formation. Eosinophilic cellulitis, or Wells syndrome, is manifested as nodules, papules, or ulcerative lesions, as well as red plaques in which eosinophils infiltrate the area between collagen fibers.

Sclerotic Lesions

Sclerotic lesions are accompanied by more collagen production, which results in skin with a glossy appearance. Sclerosis may also result from the accumulation of mucopolysaccharides in scleromyxedema (lichen myxedematosus) or from amyloid deposits. Papular mucinosis, lichen myxedematosus (Fig. 440-28), and scleromyxedema are a spectrum of diseases that are caused by deposition of hyaluronic acid. An entity associated with renal failure and gadolinium exposure, nephrogenic fibrosing dermopathy[12] (Fig. 440-29), is also characterized by acral fibrosis and deposition of hyaluronate in the skin. In scleroderma (Chapter 267), increased collagen deposition may be associated with Raynaud syndrome, calcinosis, and telangiectasia. A localized form of scleroderma, termed *morphea,* may occur down the center of the face (coup de sabre) or as plaques on the extremities (Fig. 440-30), after radiation exposure, or with *Borrelia* spp. infection. Lichen sclerosis et atrophicus is a superficial inflammatory morphea characterized by white atrophic patches, especially in the genital region. Widespread systemic sclerosis may also follow bone marrow transplantation in the setting of chronic graft-versus-host disease.

Telangiectasia

Telangiectasia is prominence of skin blood vessels that frequently accompanies atrophic as well as sclerotic processes and is common in photoaged skin and after radiation therapy. Telangiectasia of the mucous membranes is found in Osler-Weber-Rendu syndrome (Chapter 173), and vascular spiders are

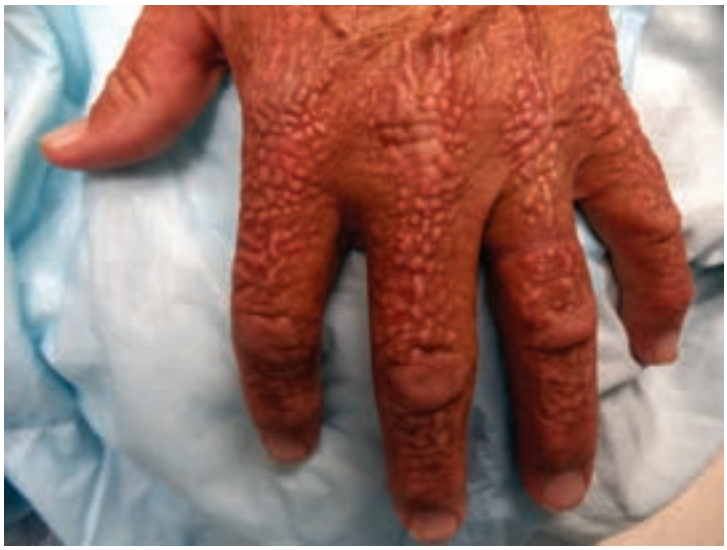

FIGURE 440-28. Lichen myxedematosus.

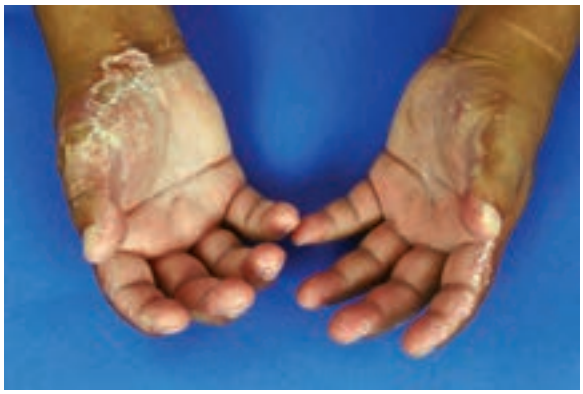

FIGURE 440-29. Nephrogenic fibrosing dermopathy.

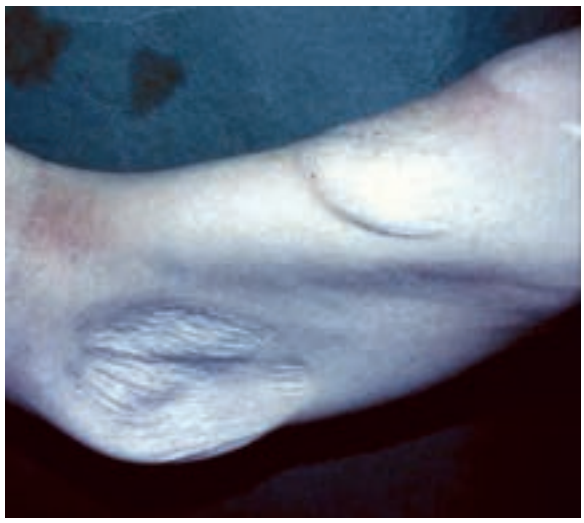

FIGURE 440-27. Anetoderma.

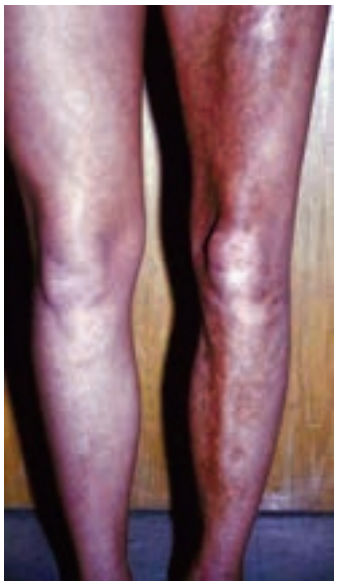

FIGURE 440-30. Linear morphea.

found both in α_1-antitrypsin deficiency and alcoholic liver disease. The presence of telangiectasia, hyperpigmentation, and hypopigmentation (poikiloderma) in sun-shielded areas of the body should alert the clinician to the diagnosis of early mycosis fungoides.

Ulcers

Ulcers are secondary skin lesions that may arise from trauma, loss of proper blood supply, aging, vasculitis, blister formation, infection, or underlying neoplasia. Ulcers may be shallow erosions (loss of the epidermis) or may be deeper and involve the dermis and underlying subcutaneous structures. Ulcers most commonly appear on the lower extremities, where they result from stasis dermatitis and venous insufficiency, arteriolar insufficiency, diabetic neuropathy, or vasculitis. *Pyoderma gangrenosum* is a trauma-induced ulcer that is part of the spectrum of Sweet syndrome, accompanies other conditions, and may require immunosuppressive therapy. Diagnosis requires skin biopsy, cultures, and serologic testing for other associated diseases. In contrast, decubitus ulcers require débridement, elimination of local pressure, and attention to nutrition.

Grade A References

A1. Sharma M, Bennett C, Cohen SN, et al. H1-antihistamines for chronic spontaneous urticaria. *Cochrane Database Syst Rev*. 2014;11:CD006137.

A2. Vena GA, Cassano N, Colombo D, et al. Cyclosporine in chronic idiopathic urticaria: a double-blind, randomized, placebo-controlled trial. *J Am Acad Dermatol*. 2006;55:705-709.

A3. Maurer M, Rosen K, Hsieh H, et al. Omalizumab for the treatment of chronic idiopathic or spontaneous urticaria. *N Engl J Med*. 2013;368:924-935.

A4. Lebwohl M, Swanson N, Anderson LL, et al. Ingenol mebutate gel for actinic keratosis. *N Engl J Med*. 2012;366:1010-1019.

A5. Hogeling M, Adams S, Wargon O. A randomized controlled trial of propranolol for infantile hemangiomas. *Pediatrics*. 2011;128:e259-e266.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Fonacier L, Aquino M, Kim B. Clinical evaluation and treatment of chronic urticaria. *Postgrad Med*. 2010;122:148-156.
2. Stern RS. Clinical practice. Exanthematous drug eruptions. *N Engl J Med*. 2012;366:2492-2501.
3. Husain Z, Reddy BY, Schwartz RA. DRESS syndrome: part I. Clinical perspectives. *J Am Acad Dermatol*. 2013;68:693.
4. Schwartz RA, McDonough PH, Lee BW. Toxic epidermal necrolysis: part I. Introduction, history, classification, clinical features, systemic manifestations, etiology, and immunopathogenesis. *J Am Acad Dermatol*. 2013;69:173.
5. Goeser MR, Laniosz V, Wetter DA. A practical approach to the diagnosis, evaluation, and management of cutaneous small-vessel vasculitis. *Am J Clin Dermatol*. 2014;15:299-306.
6. Juhn E, Khachemoune A. Gardner syndrome: skin manifestations, differential diagnosis and management. *Am J Clin Dermatol*. 2010;11:117-122.
7. Arno AI, Gauglitz GG, Barret JP, et al. Up-to-date approach to manage keloids and hypertrophic scars: a useful guide. *Burns*. 2014;40:1255-1266.
8. Amber K, McLeod MP, Nouri K. The Merkel cell polyomavirus and its involvement in Merkel cell carcinoma. *Dermatol Surg*. 2013;39:232-238.
9. Young RJ, Brown NJ, Reed MW, et al. Angiosarcoma. *Lancet Oncol*. 2010;11:983-991.
10. Fernandez-Flores A. Current concepts on cutaneous MALT lymphomas. *Am J Dermatopathol*. 2013;35:477-484.
11. Anzalone CL, Cohen PR. Acute febrile neutrophilic dermatosis (Sweet's syndrome). *Curr Opin Hematol*. 2013;20:26-35.
12. Daftari Besheli L, Aran S, Shaqdan K, et al. Current status of nephrogenic systemic fibrosis. *Clin Radiol*. 2014;69:661-668.

REVIEW QUESTIONS

1. A 45-year-old man with a history of atopic dermatitis in childhood presents with faint pink, slightly raised lesions for the past 10 weeks. He complains of epigastric pain and burning in his chest, typically occurring after eating a meal, for the past 3 months. He reports that the lesions come and go and do not persist for more than 24 hours. On physical examination, he has urticaria that blanches with pressure on his chest, back, and upper and lower extremities. What is the next step in management?

A. Skin biopsy
B. Begin intravenous (IV) methylprednisolone
C. Begin cyclosporine
D. Allergy testing
E. *Helicobacter pylori* infection testing

Answer: E The lesions described above are classic for urticaria and would be considered chronic given the duration of more than 6 weeks. Chronic urticaria can be triggered by occult infections, such as *H. pylori*, so *H. pylori* infection should be evaluated first in this case. The patient's complaints of epigastric pain and burning suggest the possibility of gastroesophageal reflux disorder. He should be evaluated for *H. pylori* infection. Skin biopsy should be performed if urticarial lesions present for more than 24 hours to evaluate him for urticarial vasculitis. IV corticosteroids can be considered if the urticaria is severe or if it is associated with anaphylaxis or wheezing. If the history is not suggestive of a trigger, allergy testing is recommended.

Chronic urticaria (Chapter 440).

2. A 25-year-old white woman with seizures presents to her primary care physician with a generalized symmetrical rash that started 1 week ago. She had her first seizure 6 weeks ago, at which time she was started on carbamazepine, which controlled her seizures. Vital signs reveal blood pressure, 119/60; heart rate, 105 beats/min; respiratory rate, 14 breaths/min; temperature, 100.9° F; and oxygen saturation, 99% on room air. Physical examination reveals normal conjunctivae and oropharynx, but she has erythroderma, and pustules present diffusely. Laboratory data include:

White blood cell count, 9600/μL
Hemoglobin, 14.2 g/dL
Hematocrit, 43%
Platelets, 300,000/μL
Neutrophils, 35%
Lymphocytes, 25%
Monocytes, 8%
Eosinophils, 12%
Basophils, 0%
What is the initial step in management?

A. Discontinue carbamazepine.
B. Begin intravenous vancomycin.
C. Begin prednisone.
D. Perform a skin biopsy.
E. Begin intravenous immunoglobulin therapy (IVIG).

Answer: A This young woman is febrile, has eosinophilia, and has a rash that developed 6 weeks after starting carbamazepine. This clinical presentation is suggestive of drug rash with eosinophilia and systemic symptoms (DRESS). The next step in management is to remove the offending agent. Carbamazepine is a commonly reported medication that causes DRESS. Because the patient is febrile, an infectious etiology should be excluded. However, antibiotics are not indicated at this time. Systemic corticosteroids are recommended as part of the treatment of DRESS, but removal of the offending agent is crucial. In patients who do not respond to systemic steroids, IVIG can be considered. Skin biopsy of these lesions shows perivascular lymphocytic infiltrate in the papillary dermis with eosinophils. Although a skin biopsy would be helpful in the diagnosis of DRESS, the first step should be to discontinue carbamazepine, since the clinical presentation is highly suggestive of DRESS.

Husain Z, Reddy BY, Schwart RA. DRESS syndrome: part II. Management and therapeutics. *J Am Acad Dermatol*. 2013;68:709-720.

3. A 34-year-old Asian woman presented to her primary care physician with dysuria and hematuria 2 weeks ago. She was treated with trimethoprim–sulfamethoxazole for 3 days for a urinary tract infection. She then presented to the emergency department this morning with fever, oral pain, and a rash. On physical examination, she has oral erosions with hemorrhage, as well as purpura on her lower extremities. With gentle pressure, she has sloughing of the skin. What is the next step in management?

A. Provide supportive care.
B. Begin vancomycin and piperacillin–tazobactam.
C. Initiate intravenous immunoglobulin (IVIG) therapy.
D. Begin infliximab.
E. Begin corticosteroids.

Answer: A This young woman has clinical findings suggestive of Stevens-Johnson syndrome. Sloughing of the skin with gentle pressure is known as Nikolsky sign. The next step in management is supportive care, which includes transfer to a burn unit, fluid and electrolyte repletion, and ophthalmologic evaluation. She is at increased risk for bacterial superinfections, but antibiotics are not indicated for prevention of a secondary infection. The use of IVIG remains controversial, and randomized controlled trials have not been done yet. Tumor necrosis factor-α inhibitors, including infliximab, have been shown to decrease the time to reepithelialization, but starting infliximab would not be the next step in management. Corticosteroids have been frequently used as part of the management, but there no randomized controlled trials have evaluated corticosteroid use in Stevens-Johnson syndrome.

Worswick S, Cotliar J. Stevens-Johnson syndrome and toxic epidermal necrolysis: a review of treatment options. *Dermatol Ther*. 2011;24:207-218.

4. A 35-year-old woman noted a new onset of multiple pink papules that appeared suddenly and looked like mosquito bites. Three months later, most of these lesions had resolved, but new ones appeared. On examination, there are pink to red papules, some with ulceration and some with scarring. Biopsy of a papule shows a $CD30^+$ lymphoproliferative disorder. All of the following are acceptable treatment options except:

A. Oral low-dose methotrexate
B. Systemic evaluation for lymphoma
C. Multidrug chemotherapy
D. Topical clobetasol
E. None of the above

Answer: C These crops of pink papules that regress spontaneously are typical of lymphomatoid papulosis. Patients with lymphomatoid papulosis can also get mycosis fungoides or anaplastic large T-cell lymphoma. Oral low-dose methotrexate and topical corticosteroids, including clobetasol, are acceptable treatment options for lymphomatoid papulosis. Systemic evaluation for lymphoma should also be considered. It is important not to treat this patient too aggressively.

Duvic M. CD30+ Neoplasms of the skin. *Curr Hematol Malig Rep*. 2011;6:245-250.

5. A 56-year-old white man with well-controlled Crohn disease on oral mesalamine presents with an ulcer on his left lower extremity for the past 2 months. He denies any associated fevers or chills. Vital signs are blood pressure, 125/72 mm Hg; heart rate, 75 beats/min; respiratory rate, 14 beats/min; and temperature, 98.4° F. Physical examination reveals warm lower extremities with good hair growth, palpable distal pulses, and no edema bilaterally. Pinprick and vibratory sensation are intact on the plantar aspects of both feet. There is a deep ulcer with a violaceous border. Laboratory data include:

White blood cell, 7600/μL
Hemoglobin, 13.5 g/dL
Hematocrit, 46%
Platelets, 250,000/μL
Neutrophils, 57%
Lymphocytes, 37%
Monocytes, 5%
Eosinophils, 2%
Basophils, 0%
Hemoglobin A1c, 5.9%
What is the next step in management?

A. Débridement
B. Begin intravenous vancomycin therapy
C. Ankle-brachial index testing
D. Begin oral prednisone
E. All of the above

Answer: D The description of the ulcer and the patient's history of Crohn disease are suggestive of pyoderma gangrenosum. Oral prednisone is one of the initial therapies for pyoderma gangrenosum. A wound culture of the ulcer should be obtained to exclude an infectious etiology, but antibiotics are not indicated at this time. Débridement is contraindicated and can result in a larger wound. Because his pulses are palpable, ankle-brachial index testing is not indicated.

Ahronowitz I, Harp J, Shinkai K. Etiology and management of pyoderma gangrenosum: a comprehensive review. *Am J Clin Dermatol*. 2012;13: 191-211.

TABLE 441-1 SKIN INFECTIONS

BACTERIAL DISEASES
Impetigo
Ecthyma
Folliculitis
Furuncle/carbuncle
Abscess
Erysipelas
Cellulitis
Necrotizing fasciitis
Ecthyma gangrenosum
Other
Gram-negative cocci: Meningococcemia, gonococcemia
Gram-positive bacilli: Erythrasma, anaerobic cellulitis
Spirochetes: Lyme disease, syphilis, endemic treponematoses
Mycobacteria
Rickettsia
VIRAL DISEASES
Herpes simplex virus: Oral, genital
Human papillomavirus: Common warts, condyloma acuminata
Poxvirus: Molluscum contagiosum
Varicella-zoster virus
Viral exanthems (e.g., enteroviruses, rubeola, rubella, parvovirus, Epstein-Barr virus, adenovirus, dengue virus, human immunodeficiency virus [seroconversion])
FUNGAL DISEASES
Candidiasis
Tinea (dermatophytoses): pedis, corporis, cruris, manuum, capitis
Pityriasis (tinea) versicolor
Emboli (e.g., *Aspergillus, Mucor* spp)
ECTOPARASITES/PARASITES
Scabies
Lice: Scalp, pubic, body
Leishmaniasis
Schistosomiasis, human and animal
Onchocerciasis
Strongyloidiasis
Amebiasis
Trypanosomiasis
Hookworm infections, human and animal
Filariasis
Acanthamoebiasis

441

INFECTIONS, HYPERPIGMENTATION AND HYPOPIGMENTATION, REGIONAL DERMATOLOGY, AND DISTINCTIVE LESIONS IN BLACK SKIN

JEAN BOLOGNIA

INFECTIONS, INCLUDING CELLULITIS

Cutaneous infections can be divided into four major categories: bacterial, fungal (Chapter 438), viral, and parasitic (Table 441-1).

Bacterial Infections

Of the cutaneous bacterial infections, impetigo, folliculitis, furuncles, and cellulitis are most commonly encountered.

IMPETIGO

Impetigo, which is caused by *Staphylococcus aureus* or group A β-hemolytic streptococci, is usually seen as honey-colored crusts (Fig. 441-1); less often, subcorneal (superficial) bullae are present. This infection is most commonly found on the face in children, but it can develop at any site where the cutaneous barrier has been disrupted (e.g., areas of dermatitis, sites of trauma, or arthropod bites). A deeper, but less common, bacterial infection of the skin is ecthyma, which is most frequently streptococcal in origin; it is characterized by thick hemorrhagic crusts overlying erosions or ulcerations, usually 0.5 to 1.5 cm in diameter. These lesions favor the extremities, especially in the setting of lymphedema. Ecthyma should not be confused with ecthyma gangrenosum, which represents an embolic phenomenon most often caused by bacteremia with gram-negative bacilli. Although mild cases of impetigo usually respond to topical 2% mupirocin three times daily or 1% retapamulin twice daily, more severe impetigo and ecthyma require oral antibiotics that cover *S. aureus* (e.g., dicloxacillin, 250 mg orally [PO] four times daily, or

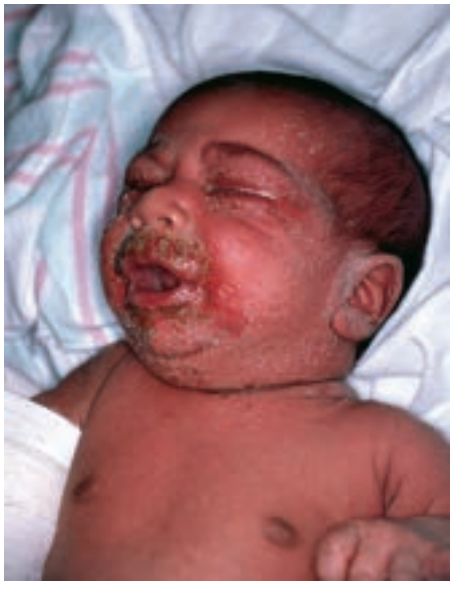

FIGURE 441-1. Impetigo in an infant and marked involvement of the face with honey-colored crusts and superficial erosions. (Courtesy Yale Dermatology Residents' Slide Collection.)

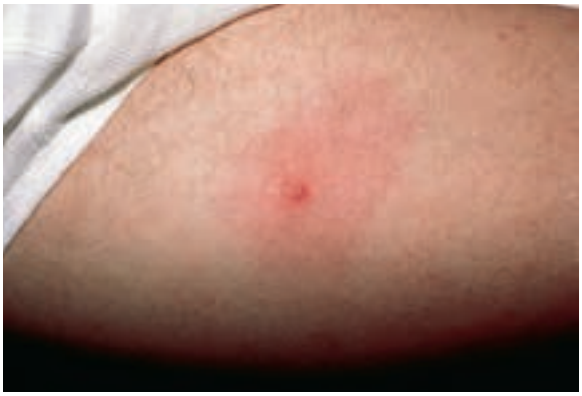

FIGURE 441-2. Furuncle with surrounding cellulitis. This common presentation of methicillin-resistant *Staphylococcus aureus* should be treated with incision and drainage, as well as the administration of systemic antibiotics. (Courtesy Yale Dermatology Residents' Slide Collection.)

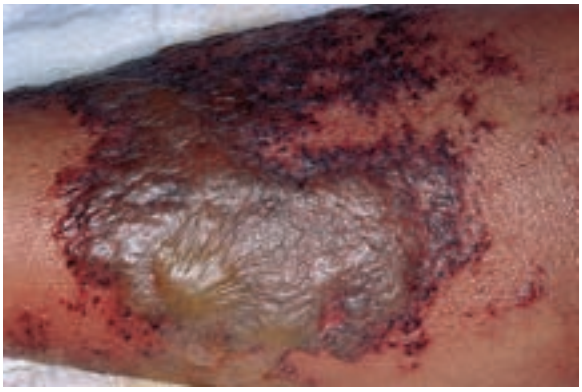

FIGURE 441-3. Bullous and hemorrhagic cellulitis of the shin. (Courtesy University of Southern California Dermatology Residents' Slide Collection.)

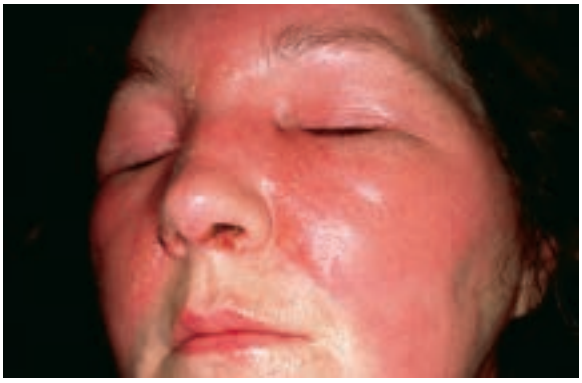

FIGURE 441-4. Erysipelas of the face with well-demarcated erythematous plaques. (Courtesy Yale Dermatology Residents' Slide Collection.)

cephalexin, 250 mg PO four times daily). Compared with furuncles and abscesses, impetigo is less often due to methicillin-resistant *S. aureus* (MRSA).

FOLLICULITIS

The initial lesions of folliculitis are follicular pustules that are often surrounded by a rim of erythema (Chapter 439). *Pseudomonas* folliculitis, which favors the trunk, is usually associated with the use of hot tubs or whirlpools because their higher temperatures (vs. swimming pools) make eradication of *Pseudomonas* more difficult (see Fig. 439-20).

FURUNCLES

Furuncles, also called *boils*, represent *S. aureus* cutaneous infections that are localized primarily within the dermis. In contrast to folliculitis, the lesions are larger and manifest as tender erythematous nodules (Fig. 441-2). A central follicular structure may be noted, as may a central pustule ("pointing"). Because a furuncle is an abscess, the preferred treatment is incision and drainage followed by oral antistaphylococcal antibiotics (e.g., dicloxacillin, 250 mg PO four times daily, or cephalexin, 250 mg PO four times daily); if MRSA is likely (e.g., use of health care facilities such as dialysis units, participation in skin-to-skin contact sports, an elevated prevalence of methicillin-resistant isolates in the local community), the antibiotic should be changed to clindamycin (300 to 600 mg PO three times daily), doxycycline (100 mg PO twice daily), minocycline (100 mg PO twice daily), trimethoprim-sulfamethoxazole (160 mg/800 mg PO twice daily), or linezolid (600 mg PO twice daily), depending on local sensitivity patterns.[1] The duration of therapy is usually 10 to 14 days. Carbuncles, which are larger, more complex, and more extensive versions of furuncles, may be accompanied by systemic symptoms such as fever. In addition to incision and drainage, they may require a more prolonged course of antibiotic therapy.

CELLULITIS

Cellulitis is a fairly common cutaneous infection that occurs most often on the lower extremities. Locally, it manifests as erythema, edema, warmth, and tenderness; systemic findings can include fever, malaise, and leukocytosis. Most cases are bacterial in origin, but some are caused by fungal infections (e.g., *Cryptococcus* spp) or chemical reactions (e.g., extravasated oxacillin or calcium salts). Bacterial cellulitis is most commonly caused by group A β-hemolytic streptococci and *S. aureus*, with the former being associated with the more severe, necrotizing variant. In patients who have diabetes or are immunocompromised, cellulitis can be caused by gram-negative bacilli or atypical mycobacteria. Risk factors include a preceding break in the skin barrier, edema secondary to venous hypertension, lymphedema, and previous bouts of cellulitis.

Although the diagnosis of cellulitis is usually fairly straightforward (Fig. 441-3), it can sometimes be difficult in patients with chronic lower extremity edema, especially in those who are afebrile and have persistent discoloration. One complication of chronic lower extremity edema is lipodermatosclerosis (i.e., inflammation followed by fibrosis of subcutaneous fat), which is seen acutely as erythema, warmth, and tenderness and is easily confused with cellulitis. The skin above the medial malleolus is often the initial site of involvement for lipodermatosclerosis, but the inflammation can extend onto the shin and calf. The chronic phase of lipodermatosclerosis is characterized by induration, a permanent brown-red to violet discoloration of the skin, and an "inverted wine bottle" appearance of the distal end of the lower extremity. It is important for the clinician to realize that in patients with chronic lipodermatosclerosis and superimposed cellulitis, the skin will never return to the color of uninvolved skin, even after adequate antibiotic therapy.

Unless there is an associated bacteremia, the diagnosis of cellulitis is primarily clinical.[2] In immunocompromised hosts, a saline injection followed by aspiration and culture can be helpful. Histologically, cellulitis is characterized by an infiltrate of neutrophils within the dermis. Skin biopsy can exclude disorders that may be confused with cellulitis, such as contact dermatitis, erythema migrans, inflammatory carcinoma, toxic erythema of chemotherapy, and Wells syndrome (an idiopathic disorder in which eosinophils infiltrate the dermis). Treatment of cellulitis varies from oral cephalexin (250-500 mg four times daily for 10 to 14 days) to intravenous vancomycin plus intravenous ceftazidime (15 mg/kg twice daily and 0.5 to 1 g three times daily, respectively, until clinical response allows transition to oral medications), depending on the suspected pathogens, the host, and the severity of systemic toxicity. Penicillin (250 mg twice daily) is effective for preventing recurrent cellulitis.

Cellulitis resides in the middle of a spectrum of soft tissue infections that includes erysipelas (more superficial and more sharply demarcated; Fig. 441-4) at one end and necrotizing fasciitis (deeper, more necrotic, and undermining) at the other. In healthy adults, erysipelas can be treated with oral penicillin (200,000 units four times daily) or, if methicillin-sensitive *S. aureus* (MSSA) is of concern, oral dicloxacillin (500 mg four times daily) for a 10-day course. Necrotizing fasciitis is usually caused by multiple organisms, including anaerobic streptococci; its diagnosis requires a high index of suspicion, and it must be considered when there are areas of painful violaceous induration or a foul-smelling discharge. Prompt surgical débridement and broad-spectrum systemic antibiotics (e.g., a β-lactam/β-lactamase inhibitor such as intravenous piperacillin-tazobactam, 4.5 g every 6 hours for a total of 18 g/day [16 g piperacillin/2 g tazobactam]) for at least 2 weeks, are mandatory; addition of ciprofloxacin (500 mg PO or 400 mg intravenously [IV] twice daily), metronidazole (500 mg IV three times daily), and vancomycin (15 mg/kg IV twice daily) depends on suspected pathogens. Unless only a

single organism is seen on Gram stain and isolated on culture, broad-spectrum antibiotic coverage should be continued because of the polymicrobial nature of necrotizing fasciitis and the difficulty of culturing anaerobes.

Although *Clostridium perfringens* can cause anaerobic cellulitis and gas gangrene, the most common cutaneous infection by gram-positive bacilli is erythrasma, which manifests as interdigital toe web maceration with fissures, as well as shiny or scaly brown-red patches in the axillae and groin. The latter is often confused with tinea cruris (Chapter 438) and seborrheic dermatitis. A diagnostic finding is the presence of coral (orange-pink) fluorescence on Wood lamp illumination (ultraviolet A). The responsible organism is *Corynebacterium minutissimum*. Treatment options include topical and oral erythromycin (e.g., 333 mg three times daily for 7 to 14 days).

TOXIC ERYTHEMAS

Eruptions caused by the release of toxins (e.g., exfoliative toxins ET-A and ET-B, erythrogenic toxin) produced by *S. aureus* and streptococci include staphylococcal scalded skin syndrome (Chapters 288 and 439), scarlet fever (Chapter 290), and toxic shock syndrome (Chapter 439). Staphylococcal scalded skin syndrome (see Fig. 439-14) is characterized by large areas of tender erythema in which superficial desquamation (peeling) develops, often with scaling and crusting in a radial array around the mouth. The areas of erythema are sterile; the conjunctivae, nasopharynx, or a distant site on the skin is the usual site of the primary staphylococcal infection. A clue to the diagnosis of scarlet fever is the presence of a strawberry tongue with prominent red papillae. Management involves treatment of the systemic infection (Chapters 288 and 439).

NEISSERIA INFECTIONS

Both gonococcemia (Chapter 299) and meningococcemia (Chapter 298) can manifest with cutaneous lesions. The former gives rise to a small number of vesicopustules on an erythematous base, generally acral in location (Fig. 441-5); these lesions represent septic emboli and are accompanied by fever, arthritis, and tenosynovitis. The earliest lesions of acute meningococcemia may be subtle (macular areas of erythema), but central hemorrhage (petechiae and purpura) and necrosis (gun-metal gray color) soon follow (Fig. 441-6). When accompanied by disseminated intravascular coagulation, large areas of retiform purpura and severe peripheral ischemia may develop. Cutaneous involvement in chronic meningococcemia is a reflection of lymphocytic or leukocytoclastic vasculitis. Management is systemic treatment (Chapters 298 and 299).

PSEUDOMONAS INFECTIONS

Pseudomonas infections of the skin vary from "hot tub" folliculitis (Chapter 439) to soft tissue infections of the external ear. Interdigital toe web infections that begin as simple tinea pedis can be complicated by superimposed *Pseudomonas* infection and result in erythema, swelling, tenderness, and drainage. Depending on its severity, treatment varies from topical antiseptics to oral or intravenous fluoroquinolones (e.g., ciprofloxacin, 500 mg PO twice daily for 7 to 14 days). In immunocompromised hosts, *Pseudomonas* and other gram-negative bacilli can produce cellulitis and secondary septic emboli in the skin. The latter begin as purpura or purpuric bullae in which central necrosis then develops. These lesions, which arise as a result of ischemic infarction of the skin, are termed *ecthyma gangrenosum*. Management is treatment of the systemic pseudomonal disease (Chapter 306).

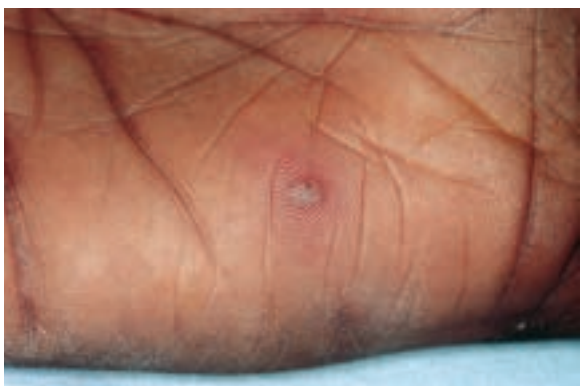

FIGURE 441-5. Disseminated gonococcal infection with an acral pustule on a red-violet base. (Courtesy Yale Dermatology Residents' Slide Collection.)

SPIROCHETES

Spirochetal infections have a wide range of skin findings, from erythema migrans secondary to *Borrelia burgdorferi* (Chapter 321), to endemic treponematoses such as yaws and pinta (Chapter 320), to the cutaneous manifestations of the three stages of syphilis (Chapter 319). Syphilitic lesions include a firm, generally nontender ulceration (chancre) in primary syphilis; a generalized papulosquamous eruption (Chapter 438) plus alopecia, oral ulcers, and condylomata lata in secondary syphilis; and thick plaques and ulcers in tertiary disease. Management involves treatment of the systemic disease.

MYCOBACTERIA INFECTIONS

Infections with *Mycobacterium tuberculosis* and mycobacteria other than *M. tuberculosis* (atypical mycobacteria) are associated with skin lesions, including verrucous or crusted papules, scarring granulomatous plaques, and draining ulcers. In immunocompetent hosts in developed countries, *Mycobacterium marinum* (Chapter 325) is most commonly associated with skin disease, which is usually manifested in a lymphocutaneous (i.e., sporotrichoid) pattern. Lower extremity furunculosis due to atypical mycobacteria can occur following pre-pedicure footbaths, and injection of tattoo ink contaminated with *Mycobacterium chelonae* can lead to erythematous papules.[3] Treatment of cutaneous mycobacterial disease is the same as for systemic disease (Chapters 324 and 325).

Viral Infections

The most common viral infections of the skin are verrucae (warts; see Fig. 439-2), recurrent oral and genital herpes simplex (Chapters 374 and 439), molluscum contagiosum (see Fig. 440-9), and exanthems (Chapter 439). Varicella and herpes zoster are seen less frequently (Chapter 375).

Fungal Infections

A variety of fungal infections involve the skin and nails and are most commonly due to dermatophytes (tinea), *Candida* spp, and *Malassezia* spp (pityriasis versicolor, also referred to as tinea versicolor) (Chapter 438; also see Table 441-1). Although both dermatophyte infections and pityriasis versicolor are associated with scaling, cutaneous candidiasis is characterized by erythema, a more erosive appearance, and satellite pustules. Treatment is described in Chapter 438.

Septic emboli caused by *Candida* or other opportunistic fungi such as *Aspergillus* (Chapter 339) or *Fusarium* often have a clinical appearance similar to that of ecthyma gangrenosum secondary to gram-negative rods such as *Pseudomonas*. The responsible organisms can be detected histologically in biopsy specimens or by bedside examination of dermal scrapings; culture confirms the specific organism. While rare, cutaneous plaques secondary to *Pneumocystis jiroveci* favor the internal ear. Treatment is for the underlying fungal infection.

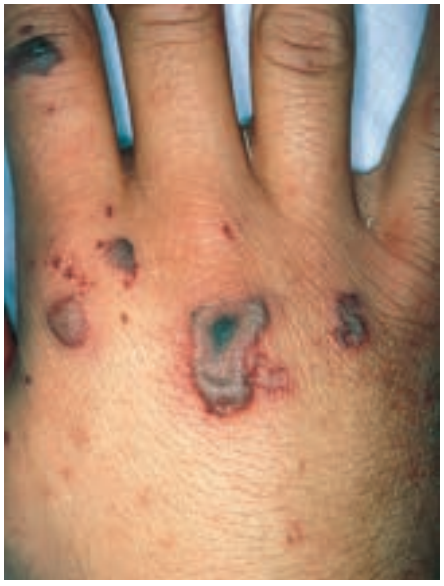

FIGURE 441-6. Purpuric and necrotic embolic lesions of meningococcemia. (Courtesy Yale Dermatology Residents' Slide Collection.)

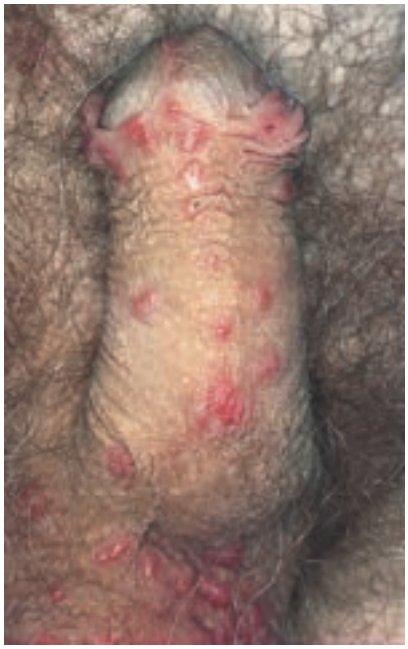

FIGURE 441-7. Scabies with involvement of the penis. (From Bolognia JL, Jorizzo JL, Schaffer JV, eds. *Dermatology,* 3rd ed. London: Elsevier; 2012.)

Ectoparasites and Parasites

ECTOPARASITES: SCABIES AND LICE

The most common ectoparasitic cutaneous infestations are (1) scabies from the human variant of the *Sarcoptes* mite; and (2) lice, of which there are three subtypes: head, body, and pubic. Scabies is characterized by pruritus in association with papules, papulovesicles, and linear burrows, as well as signs of scratching, such as excoriations and areas of dermatitis. Sites of predilection include the wrists, ankles, fingers, and toes (including the web spaces), areolae, and genitalia (especially the penis) (Fig. 441-7). The number of mites living within the stratum corneum is limited in immunocompetent hosts; when scraped and examined microscopically, linear burrows provide the highest yield of mites and eggs. In elderly and immunosuppressed patients, a form of scabies known as crusted (previously Norwegian) scabies manifests as multiple areas of scaling and crusting that are teeming with mites.

Infestations with scalp lice are seen most commonly in children, who may be asymptomatic or have marked pruritus. In addition to the lice, multiple egg casings ("nits") are attached to the proximal portions of scalp hairs. In developed countries, body lice are seen primarily in homeless individuals; patients typically have multiple erythematous papules at the sites of bites, as well as signs of scratching. The lice and their eggs are found in the patient's clothing. Pubic lice are sometimes called "crabs" because their bodies are shorter and broader than those of scalp or body lice and thus resemble the shape of a crab. Because of their leg span, these lice reside primarily on pubic hairs and less often on axillary hairs or eyelashes.

The first-line U.S. Food and Drug Administration–approved treatments for head lice, pubic lice, and routine scabies are topical 0.5% malathion lotion or gel, 5% permethrin cream, and 5% permethrin cream, respectively[4]; each of these topical medications is applied for 8 to 12 hours on days 1 and 8. Recently, topical ivermectin 0.5% lotion was also approved for the treatment of head lice.[5] For crusted scabies, epidemic outbreaks of scabies (e.g., in nursing homes), or difficult-to-treat head lice, oral ivermectin (250 to 400 μg/kg; off-label use) can eradicate the infestation. Treatment of body lice involves discarding egg- and lice-infested clothing; for head lice, potential sources of reinfection, such as hairbrushes, should be discarded. Sexual and household contacts of patients with pubic lice and scabies, respectively, must be treated similarly to the symptomatic patient.

OTHER PARASITES

Cutaneous lesions are seen in leishmaniasis (Chapter 348), amebiasis (Chapter 352), schistosomiasis (Chapter 355), onchocerciasis (Chapter 358), strongyloidiasis (Chapter 357), and hookworm infections (Chapter 357). Exposure to water infested with the cercariae of animal schistosomes results in multiple erythematous papules, which occur most commonly on the feet and are termed *swimmer's itch*. Dog and cat hookworm infections lead to cutaneous larva migrans, with serpiginous erythematous tracks that correspond to the path of migration of the hookworm larvae in sites where there has been direct contact with infected sand, most commonly the feet. Both of these infections are self-limited because the parasite's life cycle cannot be completed in humans. In immunocompromised hosts, cutaneous plaques can occasionally develop from free-living amebae such as *Acanthamoeba*.

TABLE 441-2 DISORDERS OF PIGMENTATION

HYPOPIGMENTATION
Diffuse (Pigmentary Dilution)
Oculocutaneous albinism Hermansky-Pudlak syndrome Chédiak-Higashi syndrome Generalized (total) vitiligo Inborn errors of metabolism (e.g., phenylketonuria)
Circumscribed
Decrease in pigment Acquired: Postinflammatory hypopigmentation (e.g., atopic dermatitis, sarcoidosis, systemic lupus erythematosus, mycosis fungoides), pityriasis (tinea) versicolor secondary to *Malassezia* spp infection Congenital: Nevus depigmentosus, ash-leaf spots of tuberous sclerosis Absence of pigment Acquired: Vitiligo, chemical leukoderma, leukoderma of scleroderma, leukoderma of melanoma Congenital: Piebaldism
Linear
Linear nevoid hypopigmentation, segmental nevus depigmentosus
Guttate
Idiopathic guttate hypomelanosis Confetti macules of tuberous sclerosis
HYPERPIGMENTATION
Diffuse
Drug reactions (e.g., cyclophosphamide, busulfan) Addison disease Ectopic adrenocorticotropic hormone production (e.g., small cell lung cancer) Hemochromatosis Scleroderma Primary biliary cirrhosis Hyperthyroidism Vitamin B_{12} or folate deficiency Porphyria cutanea tarda POEMS syndrome (see Table 441-3) Melanosis secondary to metastatic melanoma Argyria (gray hue)
Circumscribed
Postinflammatory hyperpigmentation (e.g., acne vulgaris, arthropod bites, dermatitis, lichen planus) Melasma Pityriasis (tinea) versicolor Mastocytosis Fixed drug reactions Deposits of drugs and their metabolites
Linear
Exposure to psoralen-containing plants (e.g., limes) plus ultraviolet A light Drug reactions (e.g., bleomycin) Linear nevoid hyperpigmentation Genodermatoses (e.g., incontinentia pigmenti)
Reticulated
Erythema ab igne Genodermatoses

DISORDERS OF HYPOPIGMENTATION AND HYPERPIGMENTATION

Disorders of pigmentation can be divided into four major categories: diffuse, linear, circumscribed, and either reticulated (in the case of hyperpigmentation) or guttate (in the case of hypopigmentation) (Table 441-2).

Hypopigmentation

ALBINISM

The primary disorder of diffuse hypopigmentation is oculocutaneous albinism, an autosomal recessive disorder in which there is a pigmentary dilution

of melanin-containing structures (i.e., the eyes, hair, and skin). The phenotype varies from total absence of melanin pigment to a subtle decrease whose recognition requires comparison with first-degree relatives; the density of melanocytes in skin is normal, but their ability to produce pigment is absent or decreased. Ninety percent of patients with oculocutaneous albinism have mutations in the genes that encode either tyrosinase (type I) or P protein (type II). Complications of oculocutaneous albinism include decreased visual acuity, nystagmus, photophobia, and an increase in cutaneous carcinomas, especially squamous cell carcinoma. These signs and symptoms are most severe in those who produce the least pigment and have the greatest amount of cumulative sun exposure. The differential diagnosis includes total vitiligo (absence of melanocytes histologically) and a few inborn errors of metabolism (e.g., phenylketonuria). Treatment consists of longitudinal ophthalmologic care and minimizing sun exposure.

LINEAR HYPOPIGMENTATION

Disorders of linear hypopigmentation consist primarily of nevoid conditions (e.g., linear nevoid hypopigmentation, segmental nevus depigmentosus), in which streaks of hypomelanosis follow the Blaschko lines, or patients have blocklike hypopigmented patches due to mosaicism. A minority of patients have associated central nervous system and musculoskeletal abnormalities (hypomelanosis of Ito).

CIRCUMSCRIBED (PATCHY) HYPOPIGMENTATION

Vitiligo

Vitiligo (Fig. 441-8) is usually slowly progressive and occurs principally in periorificial areas (around the eyes, nose, lips, genitalia) and the hands, feet, flexor surface of the wrists, ankles, elbows, knees, and major body folds. Vitiligo, which is caused by a loss of melanocytes within the skin, is also associated with autoimmune endocrinopathies and alopecia areata. T cells that recognize antigens on the surface of melanocytes (and melanoma cells) are found within the skin and in peripheral blood. Treatment includes topical corticosteroids, topical immunomodulators (e.g., tacrolimus), and phototherapy. The differential diagnosis is primarily chemical leukoderma secondary to compounds that are cytotoxic to melanocytes (e.g., catechols, phenols), the leukoderma of melanoma (a good prognostic sign if seen in association with immunotherapy but an indication to exclude metastases if occurs spontaneously), and the leukoderma of scleroderma with retention of perifollicular pigmentation.

Guttate Hypopigmentation

Idiopathic guttate hypomelanosis, in which there are well-demarcated hypopigmented macules usually measuring 2 to 4 mm in diameter, is the most common cause of guttate ("raindrop") leukoderma (Fig. 441-9). The favored sites for this common age-related disorder, which may be related to chronic sun exposure, are the shins and the extensor surface of the forearms.

Other Acquired Causes

Circumscribed hypomelanosis is seen in patients with pityriasis (tinea) versicolor (see Fig. 438-29) and postinflammatory hypopigmentation. Although postinflammatory hypopigmentation is most often associated with atopic dermatitis, it also can occur with sarcoidosis (Chapter 95), lupus erythematosus (Chapter 266), and mycosis fungoides (Chapter 185).

Congenital Causes

Congenital circumscribed areas of hypomelanosis include *nevus depigmentosus*, a common tan "birthmark" seen in 1 in 50 infants in whom there is a partial decrease in pigment; *piebaldism*, an unusual autosomal dominant disorder with areas of complete absence of pigment caused by mutations in the *KIT* gene; *nevus anemicus*, a localized area of vasoconstriction; and the *ash-leaf spots* of tuberous sclerosis (Chapter 417), with a partial decrease in pigment.

Hyperpigmentation

DIFFUSE HYPERPIGMENTATION

Diffuse hyperpigmentation is most commonly due to drugs (e.g., cyclophosphamide, zidovudine) and endocrinopathies associated with increased circulating levels of adrenocorticotropic hormone (ACTH) (e.g., Addison disease [Chapter 227], ectopic ACTH production by tumors such as small cell lung carcinoma [Chapter 191]). ACTH, as well as melanocyte-stimulating hormone, can bind and activate the melanocortin-1 receptors on melanocytes, thereby leading to increased melanin production. Additional causes include hemochromatosis (Chapter 212), scleroderma (Chapter 267), primary biliary cirrhosis (Chapter 155), POEMS syndrome (*p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathy, *m*onoclonal protein, *s*kin changes (Table 441-3), and hyperthyroidism (Chapter 226). Systemic exposure to silver (argyria; Chapter 22) can lead to a slate-gray color.

TABLE 441-3 CUTANEOUS FINDINGS IN POLYNEUROPATHY, ORGANOMEGALY, ENDOCRINOPATHY, MONOCLONAL (M) PROTEIN, AND SKIN CHANGES (POEMS) SYNDROME

Diffuse hyperpigmentation
Vascular tumors, including glomeruloid hemangiomas
Peripheral edema
Induration (sclerodermoid)
Hypertrichosis
Hyperhidrosis
Acrocyanosis
Clubbing and/or leukonychia
Acquired facial lipoatrophy
Livedo reticularis

POEMS = *p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathy, *m*onoclonal protein, *s*kin changes.

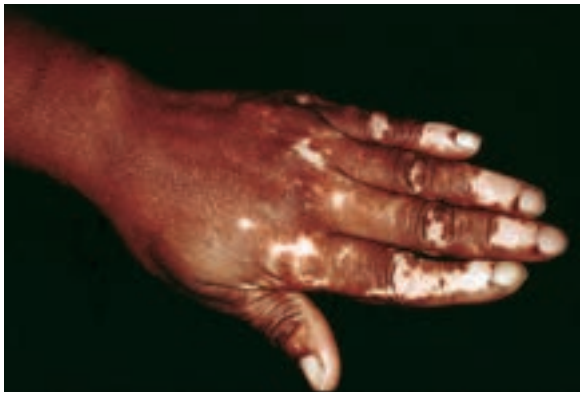

FIGURE 441-8. Striking leukoderma of the hand in a patient with vitiligo. In well-developed lesions, the skin is white, not tan, in color.

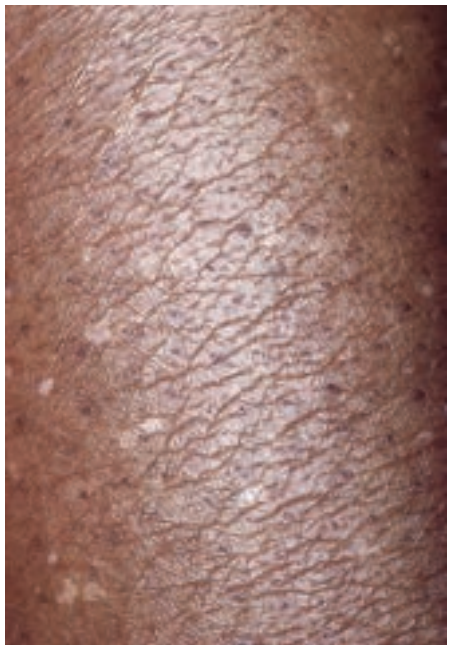

FIGURE 441-9. Idiopathic guttate hypomelanosis with small, well-demarcated hypopigmented macules on the shin.

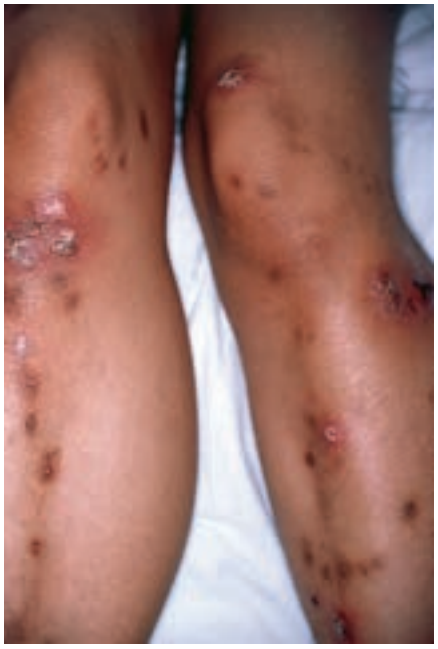

FIGURE 441-10. Postinflammatory hyperpigmentation secondary to arthropod bites. (Courtesy Yale Dermatology Residents' Slide Collection.)

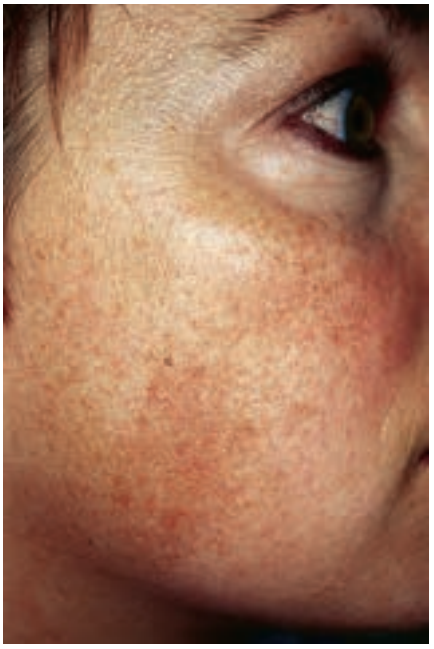

FIGURE 441-11. Hyperpigmented patches on the cheek in a patient with melasma.

LINEAR AND RETICULATED HYPERPIGMENTATION

Linear streaks of hyperpigmentation can be due to nevoid conditions that reflect cutaneous mosaicism, as in linear hypopigmentation (see earlier) and in genodermatoses (e.g., incontinentia pigmenti, which is an X-linked dominant disorder caused by mutations in the gene *NEMO*), or due to exposure to either plant-derived psoralens (e.g., from limes) plus ultraviolet A irradiation or systemic bleomycin (flagellate pigmentation). Reticulated hypermelanosis is also seen in several genodermatoses (e.g., dyskeratosis congenita) and after chronic exposure to heat (erythema ab igne). The latter corresponds to the cutaneous venous plexus and is seen most commonly in the lumbosacral region where heating pads have been applied or on the anterior thighs from laptop computers.

CIRCUMSCRIBED (PATCHY) HYPERPIGMENTATION

The most common causes of circumscribed hypermelanosis are pityriasis (tinea) versicolor (which can present as both hypopigmentation and hyperpigmentation, hence its name), postinflammatory hyperpigmentation, and melasma. Postinflammatory hyperpigmentation (Fig. 441-10) is observed more frequently in darkly pigmented individuals and often follows acne vulgaris,[6] arthropod bites, chronic dermatitis, and lichen planus. Additional causes of circumscribed darkening of the skin are cutaneous mastocytosis (urticaria pigmentosa; Chapter 255), deposits of drugs such as antimalarials and minocycline (blue-gray discoloration), and medications that produce fixed drug reactions, most frequently trimethoprim-sulfamethoxazole and nonsteroidal anti-inflammatory drugs. In melasma (Fig. 441-11), symmetrical hyperpigmented patches are seen on the lateral aspect of the forehead, upper part of the cheek, and mandibular area. At least 90% of patients with melasma are women. The lesions are exacerbated by ultraviolet light and estrogen (oral contraceptives, pregnancy). Melasma is treated with daily broad-spectrum sunscreens plus lightening agents such as hydroquinone (4% cream) and retinoic acid (0.025 to 0.10% cream) for 3 to 4 months, often in combination with mild topical corticosteroids to reduce irritation[7]; if irritation develops, the creams are discontinued temporarily and then restarted at a reduced frequency.

DISTINCTIVE LESIONS IN BLACK SKIN

Although some diseases are more common in patients of African ancestry (e.g., tinea capitis, pseudofolliculitis barbae, dissecting cellulitis), others are simply more noticeable (e.g., vitiligo and postinflammatory hypopigmentation) (Table 441-4). The explanation for the increased incidence is speculative in most instances, with the exception of curled hairs leading to pseudofolliculitis barbae. Tightly curled hairs, when shaved, are usually cut at an oblique angle, which results in a sharp tip at the distal end of the hair shaft that allows penetration of the skin adjacent to the hair follicle and subsequent inflammation. Some cutaneous disorders are seen less commonly in black skin (e.g., acne rosacea and scabies).

TABLE 441-4 DISORDERS SEEN MORE COMMONLY IN PATIENTS OF AFRICAN ANCESTRY

HEAD AND NECK
Folliculitis decalvans/dissecting cellulitis
Tinea capitis due to *Trichophyton tonsurans*
Traction alopecia
Central centrifugal cicatricial alopecia*
Acne keloidalis nuchae
Pseudofolliculitis barbae
Pomade acne
Dermatosis papulosa nigra
Inherited patterned lentiginosis
Melasma
Discoid lupus erythematosus
PALMAR
Keratosis punctata of the palmar creases
LOWER EXTREMITIES
Ulcers secondary to sickle cell anemia
GENERALIZED
Keloids
Cutaneous sarcoidosis
Papular eczema and follicular-based inflammation

*Also termed *follicular degeneration syndrome or hot comb alopecia.*

Another entity seen more commonly in individuals of African descent is keloids (Fig. 441-12). Keloids generally appear at sites of trauma (e.g., ear piercing) but can occasionally develop spontaneously, especially on the trunk. In the former situation, they are thought to represent an exaggerated response to wound healing, with increased formation of collagen not only at the site of the trauma (as in hypertrophic scars) but also in adjacent, previously uninvolved skin. Treatment options include intralesional corticosteroids, intralesional interferon, pulsed dye laser, or excision followed by radiation therapy.[8]

REGIONAL DERMATOSES

Some common dermatoses have a predilection for particular anatomic sites (Fig. 441-13 and Table 441-5). These locations can help narrow the differential and guide further diagnostic testing and therapy in many patients.

One regional dermatosis is acanthosis nigricans (Fig. 441-14). Because of its potential systemic implications, careful medical evaluation is required (Table 441-6).

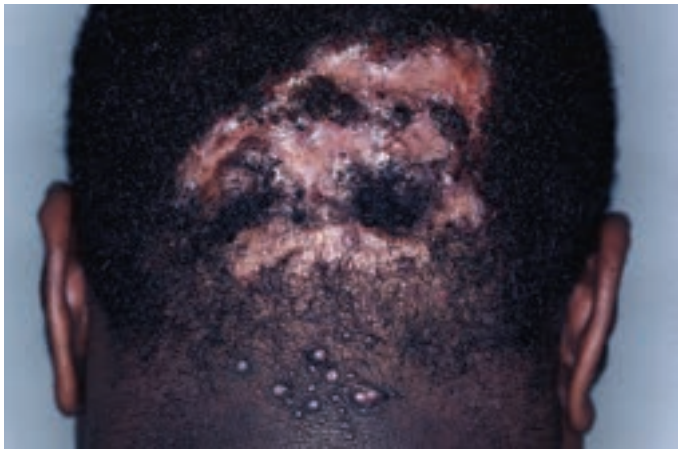

FIGURE 441-12. Acne keloidalis in an African American man. (Courtesy Kalman Watsky, MD.)

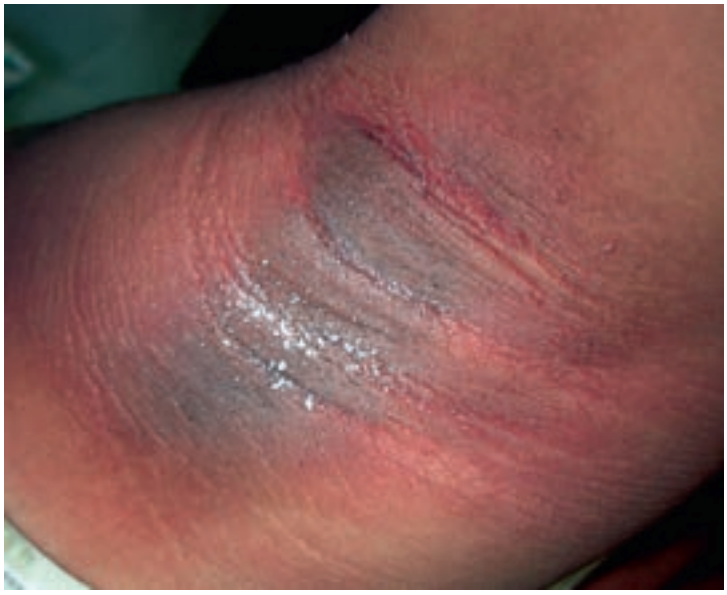

FIGURE 441-14. Acanthosis nigricans of the axilla. Note the velvet-like appearance of the skin. (Courtesy Yale Dermatology Residents' Slide Collection.)

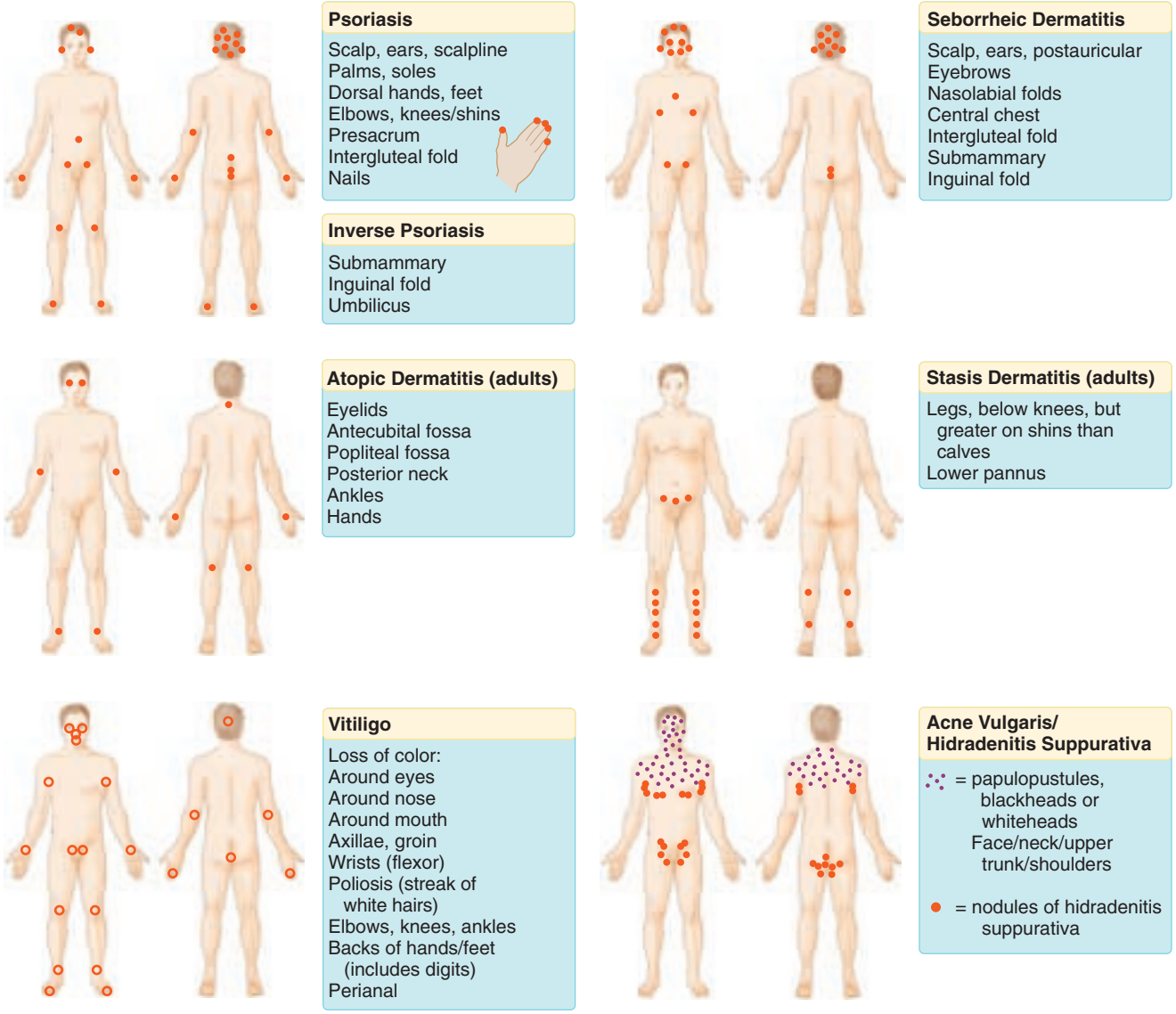

FIGURE 441-13. Regional involvement of specific skin diseases.

TABLE 441-5 REGIONAL DERMATOLOGY

REGION OF SKIN	TYPE OF SKIN LESION	DISEASE PROCESS
Scalp	Papulosquamous and eczematous	Seborrheic dermatitis, psoriasis, tinea capitis, eczema (atopic, contact)
	Pustular	Folliculitis, kerion
	Papulonodular	Melanocytic nevi, seborrheic keratoses, pilar cysts, verrucae, hemangiomas, actinic keratoses (bald scalp)
Face	Pustular	Acne, rosacea, folliculitis (beard), tinea
	Papulosquamous and eczematous	Seborrheic dermatitis, psoriasis (hairline), contact dermatitis (e.g., cosmetics), atopic dermatitis, impetigo, systemic lupus erythematosus, photodermatitis
	Vesicular	Herpes simplex, herpes zoster, bullous impetigo
	Papulonodular	Melanocytic nevi, actinic keratoses, seborrheic keratoses, sebaceous hyperplasia, basal cell carcinomas, squamous cell carcinomas, melanomas
	Atrophic and telangiectatic	Discoid lupus erythematosus
Trunk	Papulosquamous and eczematous	Psoriasis, atopic dermatitis, contact dermatitis, tinea versicolor, pityriasis rosea, scabies, secondary syphilis, subacute cutaneous lupus (upper trunk)
	Vesiculobullous	Bullous pemphigoid, pemphigus, erythema multiforme/Stevens-Johnson syndrome, herpes zoster
	Maculopapular	Morbilliform drug reactions, viral exanthems
	Papulonodular	Melanocytic nevi, seborrheic keratoses, angiomas, lipomas, epidermoid inclusion cysts, basal and squamous cell carcinomas, keloids, neurofibromas, melanoma
	Pustular	Acne, folliculitis
	Urticarial	Hives, drug reactions, early herpes zoster
Arms and forearms	Eczematous and papulosquamous	Contact dermatitis (e.g., plants), atopic dermatitis, psoriasis, lichen planus, photodermatitis (drugs, contactants, dermatomyositis, subacute cutaneous lupus)
	Papulonodular	Melanocytic nevi, verrucae, seborrheic keratoses, actinic keratoses, squamous cell carcinomas, polymorphic light eruption, rheumatoid nodules (elbows), xanthomas (elbows)
	Purpuric	Actinic (solar) purpura
	Annular	Granuloma annulare, subacute cutaneous lupus
Legs	Eczematous and papulosquamous	Stasis dermatitis, eczema craquelé (xerotic eczema), contact dermatitis, atopic dermatitis, psoriasis, lichen planus
	Papulonodular	Melanocytic nevi, dermatofibromas, erythema nodosum, melanoma, xanthomas (knees, Achilles tendon), Kaposi sarcoma
	Purpuric	Schamberg disease (capillaritis), vasculitis
	Ulcerative	Stasis ulcers, arterial insufficiency, pyoderma gangrenosum, trauma, livedoid vasculopathy, squamous cell carcinoma, diffuse dermal angiomatosis
Genitalia and groin	Eczematous and papulosquamous	Seborrheic dermatitis, tinea, psoriasis, contact dermatitis, scabies, reactive arthritis (Reiter syndrome), erythrasma, candidiasis, lichen planus, lichen simplex chronicus
	Vesiculobullous	Herpes simplex, Stevens-Johnson syndrome
	Ulcerative	Herpes simplex, syphilis, chancroid, Behçet disease, squamous cell carcinoma, trauma
	Papulonodular	Condyloma acuminata, molluscum contagiosum, angiokeratomas, epidermoid inclusion cysts, hidradenitis suppurativa, squamous cell carcinoma
	Pustular	Folliculitis, candidiasis, hidradenitis suppurativa
Hands	Eczematous and papulosquamous	Irritant and allergic contact dermatitis, atopic dermatitis, tinea, scabies, secondary syphilis
	Vesiculobullous, pustular	Dyshidrotic eczema, erythema multiforme, hand-foot-and-mouth disease (palmar), porphyria cutanea tarda (dorsal), epidermolysis bullosa acquisita, herpetic whitlow, blistering dactylitis, psoriasis (palmar)
	Papulonodular	Warts, actinic keratoses (dorsal), squamous cell carcinomas (dorsal), pyogenic granuloma, granuloma annulare, digital mucous cysts (dorsal fingers)
	Depigmentation	Vitiligo, chemical leukoderma
	Cuticular telangiectases	Scleroderma, dermatomyositis, systemic lupus erythematosus, Osler-Weber-Rendu disease
Feet	Eczematous and papulosquamous	Tinea, psoriasis, contact dermatitis, atopic dermatitis, syphilis (plantar)
	Vesiculobullous	Tinea, arthropod bites, epidermolysis bullosa (inherited and acquired), erythema multiforme, hand-foot-and mouth disease (plantar)
	Papules	Verrucae (plantar), corns, perniosis
	Ulcerative	Neuropathic ulcers (plantar)

TABLE 441-6 CLUES TO UNDERLYING SYSTEMIC CONDITIONS ASSOCIATED WITH ADULT-ONSET ACANTHOSIS NIGRICANS

Polycystic ovary syndrome	Women with acne, hirsutism, and/or menstrual irregularities
Malignancy	Sudden onset, weight loss, inflamed seborrheic keratoses, tripe palms
Endocrinopathy	Consider type 2 diabetes, Cushing syndrome (striae, hypertension, central obesity, buffalo hump), hypothyroidism
Drug-induced	Especially niacin, human growth hormone, oral contraceptive agents, corticosteroids, protease inhibitors

Grade A References

A1. Pallin DJ, Binder WD, Allen MB, et al. Clinical trial: comparative effectiveness of cephalexin plus trimethoprim-sulfamethoxazole versus cephalexin alone for treatment of uncomplicated cellulitis—a randomized controlled trial. *Clin Infect Dis.* 2013;56:1754-1762.

A2. Thomas KS, Crook AM, Nunn AJ, et al. Penicillin to prevent recurrent leg cellulitis. *N Engl J Med.* 2013;368:1695-1703.

A3. Chosidow O, Giraudeau B, Cottrell J, et al. Oral ivermectin versus malathion lotion for difficult-to-treat head lice. *N Engl J Med.* 2010;362:896-905.

A4. Whitton ME, Pinart M, Batchelor J, et al. Interventions for vitiligo. *Cochrane Database Syst Rev.* 2010;1:CD003263.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Singer AJ, Talan DA. Management of skin abscesses in the era of methicillin-resistant Staphylococcus aureus. *N Engl J Med*. 2014;370:1039-1047.
2. Phoenix G, Das S, Joshi M. Diagnosis and management of cellulitis. *BMJ*. 2012;345:e4955.
3. Kennedy BS, Bedard B, Younge M, et al. Outbreak of *Mycobacterium chelonae* infection associated with tattoo ink. *N Engl J Med*. 2012;367:1020-1024.
4. Currie BJ, McCarthy JS. Permethrin and ivermectin for scabies. *N Engl J Med*. 2010;362:717-725.
5. Pariser DM, Meinking TL, Bell M, et al. Topical 0.5% ivermectin lotion for treatment of head lice. *N Engl J Med*. 2012;367:1687-1693.
6. Silverberg NB. A brief primer on acne therapy for adolescents with skin of color. *Cutis*. 2013;92:20-26.
7. Rivas S, Pandya AG. Treatment of melasma with topical agents, peels and lasers: an evidence-based review. *Am J Clin Dermatol*. 2013;14:359-376.
8. Song C, Wu HG, Chang H, et al. Adjuvant single-fraction radiotherapy is safe and effective for intractable keloids. *J Radiat Res*. 2014;55:912-916.

REVIEW QUESTIONS

1. Acute lipodermatosclerosis is most commonly misdiagnosed as which of the following?

A. Allergic contact dermatitis
B. Irritant contact dermatitis
C. Cellulitis
D. A spider bite
E. Tinea pedis with extension on the ankle

Answer: C Acute lipodermatosclerosis is a form of panniculitis (inflammation of the subcutaneous fat) that manifests as tender, warm, erythematous plaques in patients with venous hypertension. The most common initial location is the lower extremity, superior to the medial malleolus. Because of its clinical presentation, acute lipodermatoslerosis is often misdiagnosed as cellulitis, but there is no fever or leukocytosis. Dermatitis has overlying scale-crust and is usually pruritic, whereas tinea has an active border with scale.

2. Disseminated gonococcal infection most commonly presents as tenosynovitis plus which of the following?

A. Vesiculopustules on an erythematous base, scattered acrally
B. Large areas of erythema studded with pustules on the trunk
C. Retiform purpura with central necrosis on the distal extremities
D. Morbilliform eruption on the trunk and extremities
E. Multiple erythematous subcutaneous nodules on the shins

Answer: A The other descriptions are of (B) acute generalized exanthematous pustulosis, also referred to as a pustular drug eruption; (C) disseminated intravascular coagulation that can occur in the setting of several infections, including acute meningococcemia; (D) viral exanthem or the most common form of drug eruption; and (E) erythema nodosum.

3. Of the following entities, which is seen more commonly in African Americans than in Asians or Caucasians?

A. Atopic dermatitis involving the antecubital and popliteal fossae
B. Psoriasis of the elbows and knees
C. Scabies infestation
D. Pseudofolliculitis barbae
E. Tinea (pityriasis) versicolor

Answer: D Entities that are seen more commonly in African Americans are outlined in Table 441-3.

4. A 60-year-old woman presents with an area of reticulated hyperpigmentation overlying the lumbosacral spine. She states that she has been suffering from chronic lower back pain. What is the most likely diagnosis?

A. Dyskeratosis congenita
B. Melasma
C. Erythema ab igne
D. Hyperpigmentation due to cupping and moxibustion
E. Fixed drug eruption secondary to nonsteroidal anti-inflammatory drugs

Answer: C Erythema ab igne is seen at sites of repeated heat exposure, such as from heating pads and computer batteries. Hyperpigmentation due to cupping and moxibustion (traditional Chinese medicine) or a fixed drug eruption is circular in shape, not reticulated. Melasma can affect extrafacial sites, but they are primarily the extensor forearms and upper chest, and the lesions are circumscribed, not reticulated. Patients with the genodermatosis dyskeratosis congenita have reticulated pigmentation and pancytopenia, but lesions would not be limited to the lower back.

442 DISEASES OF HAIR AND NAILS

ANTONELLA TOSTI

HAIR DISORDERS

Normal Hair

The hair shaft is a fully keratinized structure that is produced by the hair follicle. The entire skin, with the exception of the palms and soles, contains hair follicles. Hair follicles are of two types: terminal follicles and vellus follicles. Terminal follicles, which reach the hypodermis, produce terminal hairs, which are long, thick (60 to 80 µm), and pigmented. Terminal hairs are present since birth on the scalp, eyebrows, and eyelashes and later develop after puberty on the axillae, pubis, and beard region in males. Vellus follicles are small and localized to the superficial dermis and mid-dermis, where they produce vellus hairs, which are thin (<30 µm), short (<2 cm), and not pigmented and cover all the glabrous skin.

The hair follicle is formed by an upper permanent portion and a lower dynamic transient portion that migrates during the hair cycle. The transient portion includes the hair bulb, which is surrounded by the dermal papilla and contains the hair matrix that produces the hair shaft and its sheaths. The anatomic division between the permanent and the transient portion is just below the bulge region, which corresponds to the insertion of the erector pili muscle. The bulge region contains the epithelial stem cells that regenerate the follicle in each hair growth cycle; its damage results in stem cell destruction and cicatricial alopecia.

Hair Cycle

Hair follicles have a cyclic activity, characterized by alternating periods of hair shaft production and periods of resting (anagen, catagen, telogen). During the anagen phase, the follicles produce the hair shaft. Duration of anagen, which in the scalp ranges from 2 to 7 years, determines hair shaft length. Maximal length and growth rate of terminal hair varies in the different body regions. Scalp hair grows approximately 0.4 mm/day and may reach a length of more than 1 m. Maximal hair length decreases with age. During telogen, hair production is absent, even if the shaft remains within the follicle to be shed only when, after 3 months, the follicle reenters the anagen phase.

The hair cycle of adjacent scalp follicles is not synchronized. In normal conditions, approximately 85 to 90% of follicles are in anagen and 10 to 15% in telogen.

Hair Loss and Alopecias

Hair loss distresses most patients, independently from its severity and pattern. In some cases, the decrement in quality of life attributable to hair loss is comparable with that caused by major chronic diseases.

The first diagnostic step is to assess family history, drug intake, systemic illness, and severity and duration of hair loss (Table 442-1).[1] The second step is to establish whether the hair density is normal or decreased. The third step is to evaluate whether the rate of hair shedding is normal or increased. Acute and severe hair loss is typical of diseases that interrupt the mitotic activity of anagen follicles (drugs, alopecia areata). A normal hair density suggests telogen effluvium, which may be acute or chronic. A reduced hair density may involve the whole scalp (diffuse alopecia), may be limited to specific scalp regions (patterned alopecia), or may manifest with bald patches (patchy alopecia). In patchy alopecias, the scalp may show patches of alopecia that are completely devoid of hairs (alopecia areata, cicatricial alopecia) or have short broken hairs (trichotillomania, hair shaft disorders). Dermatoscopy is a rapid, noninvasive technique that greatly improves the clinical diagnosis of alopecias and hair shaft disorders in adults and children (Table 442-2).[2]

TABLE 442-1 CAUSES OF HAIR LOSS

DIFFUSE ALOPECIA
Telogen effluvium (e.g., after illness or stress)
Anagen effluvium (e.g., after chemotherapy or radiation therapy)
Medications (see Table 442-3)
Nutritional deficiency
Hair treatments
Androgenetic alopecia (in women)
Hormonal changes (e.g., menopause, discontinuation of oral contraceptives, hypothyroidism)
PATCHY ALOPECIA
Alopecia areata (probably autoimmune)
Cicatricial (scarring) alopecia (e.g., lichen planopilaris, discoid lupus erythematosus, folliculitis decalvans, central centrifugal cicatricial alopecia)
Traction alopecia (e.g., excessive hair straightening or braiding)
Trichotillomania (hair pulling)
Scalp infection (e.g., ringworm)

TELOGEN EFFLUVIUM

Acute Telogen Effluvium

Acute telogen effluvium results from noxious events that precipitate the entry of a large number of follicles into their resting phase (telogen). Possible causes include systemic diseases, drugs (Table 442-3), fever, stress, weight loss, delivery, iron deficiency, and inflammatory scalp disorders. For drugs, the severity of hair loss depends on the drug, its dosage, and the patient's susceptibility.[3]

Hair loss starts approximately 3 months after the causative event, a time frame that corresponds with the duration of the telogen phase. Telogen hairs are retained within the follicle during telogen, to be shed when the follicle produces a new anagen hair. Hair loss is severe when 100 to 200 hairs are shed daily. The patient usually remembers quite precisely when the increased shedding began. Acute telogen effluvium does not usually produce visible alopecia because approximately 50% of hairs need to be lost before reduction of hair density is evident.

TREATMENT AND PROGNOSIS Rx

Acute telogen effluvium subsides spontaneously in a few months after removal of the cause. It may, however, unmask or aggravate androgenetic alopecia.

Chronic Telogen Effluvium

Chronic telogen effluvium, which is characterized by increased hair shedding lasting for more than 6 months, mostly affects middle-aged women and frequently remains unexplained. The daily shedding is mild (<100 hairs daily), but patients are distressed and complain of progressive temporal thinning and decreased hair mass. Scalp pain (trichodynia) is frequently reported. Patients who have a high hair density may bring in envelopes of shed hairs to prove the amount of hair loss. There is no effective treatment. Chronic telogen effluvium has a chronic course with periodic exacerbations.

DIFFUSE ALOPECIA

Anagen Effluvium

Acute hair shedding leading to diffuse alopecia is a typical side effect of cancer chemotherapy and scalp radiation. Hair loss is acute and severe and may

TABLE 442-2 DERMATOSCOPIC SIGNS IN HAIR AND SCALP DISORDERS

Alopecia areata	Yellow dots, exclamation mark hairs
Androgenetic alopecia	>20% variability in the hair diameter
Lichen planopilaris/frontal fibrosing alopecia	Peripilar casts; loss of follicular openings
Trichotillomania	Broken hairs, question mark hairs
Tinea capitis	Comma hairs, corkscrew hairs
Traction alopecia	Hair casts
Discoid lupus erythematosus	Red dots, follicular plugs
Folliculitis decalvans	Hair tufts
Scalp psoriasis	Coiled capillaries
Seborrheic dermatitis	Arborizing vessels

TABLE 442-3 DRUGS REPORTED TO INDUCE HAIR LOSS

ACE inhibitors (captopril, enalapril, moexipril, ramipril)	Immunoglobulins
Allopurinol	Indanediones
Amiodarone	Indinavir*
Amphetamines*,†	Interferons*,†
Analgesics, anti-inflammatories (ibuprofen, indomethacin, naproxen)	Isonicotinic acid hydrazide‖
Androgens*,†	Leflunomide*
Anticoagulants (coumarin, dextran, heparin/heparinoids)*,†	Levodopa
Antiepileptics (carbamazepine, hydantoins, lamotrigine, troxidone, valproic acid, vigabatrin)*,†	Lithium*
Antipsychotics (flupenthixol decanoate, fluphenazine decanoate)	Maprotiline
Antithyroid drugs (carbimazole, iodine, thiouracil)*	Mesalazine
Appetite suppressants	Methyldopa
Aromatase inhibitors (fadrozole, formestane [4-OHA], vorozole)*,‡	Methysergide
Benzimidazoles (albendazole, mebendazole)	Metyrapone
β-Blockers (levobunolol, metoprolol, nadolol, propranolol, timolol)*	Minoxidil¶
Bromocriptine	Nicotinic acid
Buspirone	Nitrofurantoin
Butyrophenones	Octreotide
Cantharidin	Olanzapine
Cholestyramine	Pentosan polysulfate
Chloramphenicol	Phenindione
Cidofovir	Piroxicam
Cimetidine	Potassium thiocyanate
Clonazepam	Pyridostigmine
Clotrimazole	Radiation (<700 Gy)*,‖
Colchicine	Retinol (vitamin A)*
Contraceptives (oral)§	Retinoids (acitretin, etretinate, isotretinoin)*,†
Danazol	Risperidone
Diclofenac	Salicylates
Dixyrazine	Serotonin reuptake inhibitors (fluoxetine, fluvoxamine, paroxetine, sertraline)*
Diazoxide	Sorafenib
Ethambutol	Strontium ranelate*,†
Ethionamide	Sulfasalazine
Fibrates (clofibrate, fenofibrate)	Tamoxifen
Gentamicin	Terbinafine
Gefitinib‖	Terfenadine
Glatiramer acetate	Thiamphenicol
Glibenclamide	Thyroxine
Gold salts	Tocopherol (vitamin E)
Granulocyte-colony stimulating factor	Trazodone
Haloperidol	Triazoles (fluconazole, itraconazole)
	Tricyclic antidepressants (amitriptyline, desipramine, doxepin, imipramine, maprotiline)
	Trimethadione
	Triparanol
	Vasopressin
	Vismodegib*,†
	Spironolactone

*Established by multiple reports or proved by rechallenge.
†Hair loss usually severe.
‡May produce androgenetic alopecia.
§May produce permanent alopecia.
‖May produce anagen effluvium.
¶May produce telogen effluvium 3 mo after discontinuation.
ACE = angiotensin-converting enzyme inhibitor.

produce loss of most of the scalp hair, eyebrows, and eyelashes; other body hairs are less commonly involved. Hair shedding usually starts 4 to 6 weeks after drug intake with up to 1000 hairs shed daily. Regrowth is usually fast after discontinuation of therapy, but hair shape and color may be different. Permanent alopecia may occur with high-dose radiation and certain drug regimens, such as busulfan and taxanes. Scalp hypothermia prevents or reduces hair loss during chemotherapy except in patients treated with the combined taxotere, doxorubicin (Adriamycin), and cyclophosphamide regimen. Topical minoxidil accelerates hair regrowth but does not prevent hair loss.

Patterned Alopecia (Androgenetic Alopecia)

Androgenetic alopecia, which is the most common form of hair loss, affects up to 80% of men and 50% of women in the course of their life. Androgenetic alopecia is caused by a progressive reduction in the diameter, length, and pigmentation of the hair. Hair thinning is not diffuse, but rather is limited to the frontal, temporal, and vertex areas, where hair follicles are sensitive to the

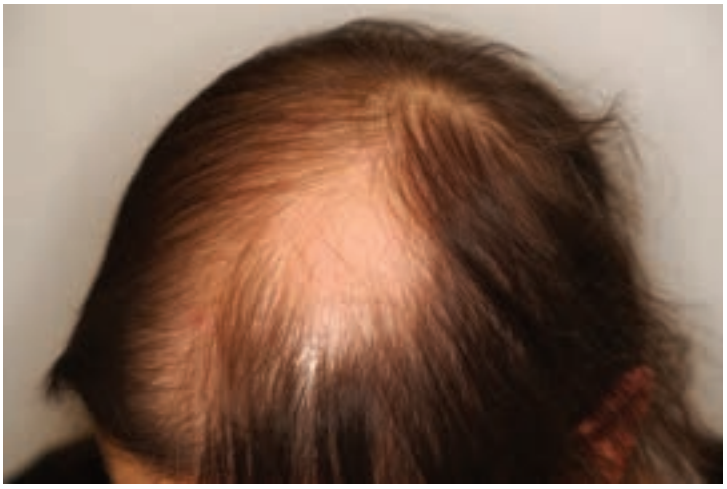

FIGURE 442-1. Alopecia areata: patchy hair loss. The alopecic area is devoid of hairs, and the scalp does not present inflammatory changes. Note diffuse thinning of the scalp surrounding the patch.

effects of the testosterone metabolite dehydrotestosterone (DHT). This sensitivity to DHT, which requires the 5α-reductase enzymes, is genetically determined. In men, androgenetic alopecia involves the frontotemporal areas and the vertex, following a pattern that corresponds to the Hamilton-Norwood scale. In women, androgenetic alopecia produces diffuse thinning of the crown region with maintenance of the frontal hairline (Ludwig pattern), a pattern that can easily be appreciated by making a central parting and comparing hair density at the top with hair density at the occipital region. In premenopausal women, androgenetic alopecia can be a sign of hyperandrogenism, together with hirsutism and acne. In most women, however, it occurs in the absence of biochemical and clinical evidence of androgen excess and may be due to excessive follicular sensitivity to androgens. Recent data show that patients with androgenetic alopecia may be at increased risk for dying from diabetes and heart disease.[4]

Androgenetic alopecia is a progressive disease that tends to worsen with time. Medical treatments include 2% topical minoxidil in women and 5% topical minoxidil and/or oral finasteride, 1 mg/day, in men. Clinical improvement is mostly due to thickening of the preexisting hair.

Treatments for androgenetic alopecia should be continued for at least 6 months before assessing efficacy, and regular drug use is mandatory for maintaining results. Interruption of minoxidil produces an acute telogen effluvium, which becomes evident 3 to 4 months after interruption and cannot be prevented by concomitant finasteride treatment. Interruption of finasteride is followed by gradual hair loss, with return to the pretreatment status after 1 year. Hair transplantation is a good option for men with severe androgenetic alopecia, and treatment with finasteride, 1 mg/day, improves the long-term results of surgery. Hair transplantation in women is more complicated because the hair thinning is often diffuse in the parietal and occipital regions, so there is not a good hair donor area.

PATCHY ALOPECIA

Alopecia Areata

Alopecia areata is a common form of nonscarring, usually patchy hair loss affecting up to 2% of the population.[5] The etiology is unknown, but evidence is consistent with an autoimmune disease to which both a genetic predisposition and environmental factors contribute.[6] In genetically predisposed individuals, various triggering factors cause a predominantly CD8-driven, T_H1-type T-cell autoimmune reaction against the hair follicles, thereby resulting in acute hair loss.

Alopecia areata can start at any age, but severe forms often start during childhood and are more frequent in males. Clinical examination reveals one or multiple well-circumscribed smooth patches of nonscarring absence of hair that enlarge in a centrifugal way (Fig. 442-1). The margins of the patches often show 3-mm-long broken hairs with a pigmented tip (exclamation mark hairs), which indicates disease progression.

Alopecia areata may affect all hairy body areas, including the eyebrows and eyelashes (Fig 442-2). Severe forms involve the entire scalp (alopecia totalis) or all body hair (alopecia universalis). Involvement of the scalp margins

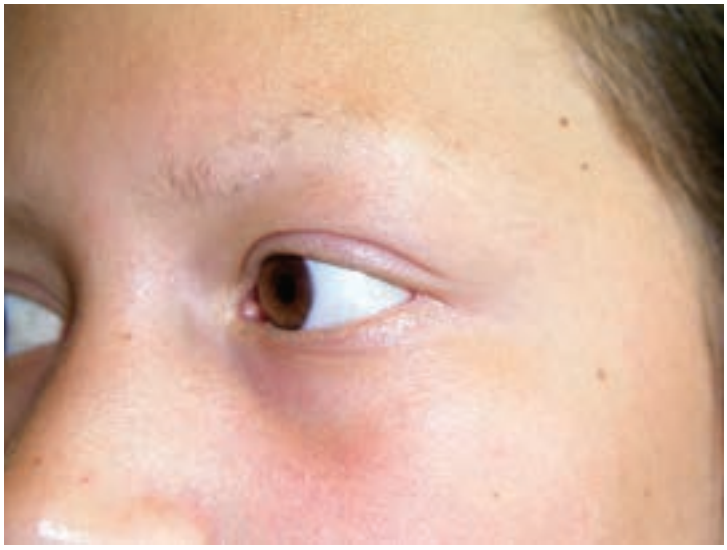

FIGURE 442-2. Alopecia areata: patchy hair loss. Hair loss involves the eyelashes and eyebrows.

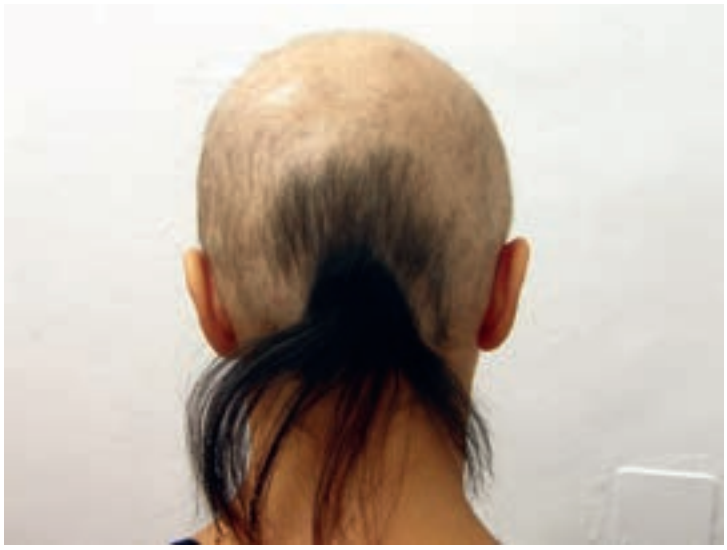

FIGURE 442-3. Trichotillomania: patchy hair loss. The alopecic area presents hairs broken at different lengths.

(ophiasis) is associated with a poor prognosis. Alopecia areata may be associated with other autoimmune diseases, most commonly thyroid diseases (Chapter 226). Other possible associations include celiac disease (Chapter 140), vitiligo (Chapter 441), and atopy (Chapter 438). Nail abnormalities are common, especially in children.

The natural history of alopecia areata is unpredictable. Studies indicate that 34 to 50% of patients will recover within 1 year and 15 to 25% will progress to alopecia totalis/universalis, a condition from which full recovery is only approximately 10%. To date, available treatments may induce temporary hair regrowth but do not change the long-term prognosis.[7]

Relapses occur in a high percentage of patients, even during therapy. High-dose pulse corticosteroid therapy is effective in approximately 60% of patients with acute alopecia areata but is not useful in ophiasis or long-standing alopecia totalis/universalis. Intralesional steroids can be used in localized areas, including the eyebrows. High-potency topical steroids (clobetasol propionate cream 0.05%) under occlusion induce regrowth in approximately 25% of patients with long-standing alopecia totalis/universalis, but relapses are common. Clobetasol propionate 0.05% in foam can be used in patchy alopecia areata. Topical immunotherapy with diphenylcyclopropenone (DPCP) or squaric acid dibutylester (SADBE), which is not approved by the U.S. Food and Drug Administration, induces hair regrowth in approximately 20% of patients with severe alopecia totalis/universalis. Medium- and low-potency topical steroids and topical minoxidil are probably only placebo treatments. Recent data suggest potential benefit from Janus kinase inhibitors, such as ruxolitinib.[8]

Trichotillomania

Trichotillomania is a compulsive disorder that is more common in children. Repetitive hair pulling and plucking produces patches of irregular alopecia. The scalp is not completely bald, but rather shows broken hairs of various lengths (Fig. 442-3). The frontal, parietal, and occipital scalp are most commonly affected, but other terminal hairs can be involved, especially the upper eyelashes.

Occasionally, patients develop the habit of chewing or eating the pulled hairs. Patients with trichotillomania frequently do not admit their habit, and parents of affected children may be recalcitrant to accept the diagnosis. Psychiatric referral is indicated in adults.

Cicatricial Alopecias (Scarring Alopecia)

The hallmark of cicatricial alopecias is loss of follicular ostia. Cicatricial alopecias include diseases that primarily affect the hair follicles and diseases that affect the dermis and secondarily cause follicular destruction.

Primary cicatricial alopecias are classified as lymphocytic or neutrophilic, based on the principal inflammatory cell type seen on pathologic examination. Lymphocytic cicatricial alopecias include lichen planopilaris, frontal fibrosing alopecia, and discoid lupus erythematosus. Neutrophilic cicatricial alopecias include folliculitis decalvans and dissecting cellulitis.

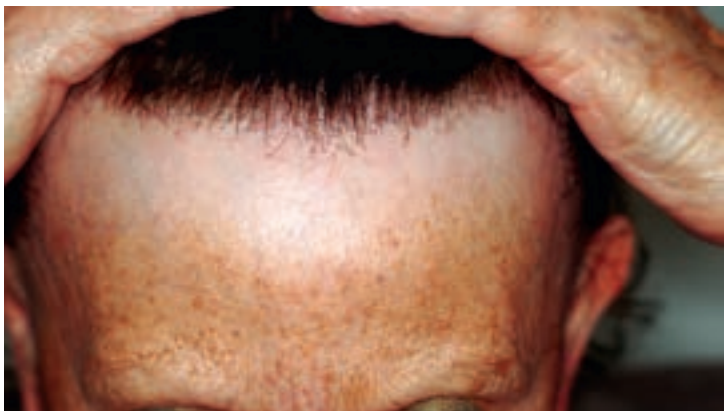

FIGURE 442-4. Frontal fibrosing alopecia. Scarring alopecia of the hair margin with recession of the frontal hairline. The alopecic area can be easily distinguished because it shows no photoaging compared with the normal forehead. Note alopecia of the eyebrows.

Secondary cicatricial alopecias, which result from disorders that cause diffuse scarring of the dermis, including burns, radiation, severe skin infections, localized scleroderma, and scalp tumors. The diagnosis of primary scarring alopecia requires a scalp biopsy, which should be taken from areas with active evidence of inflammation because biopsy of atrophic scalp will typically show only follicular or dermal fibrosis. In scarring alopecias, hair loss is permanent; treatment can prevent progression but not induce hair regrowth.

LICHEN PLANOPILARIS

In lichen planopilaris, which is the most common form of cicatricial alopecia, the hair follicles surrounding the alopecic patches show perifollicular erythema and scaling. The patient usually complains of severe itching. A variant of lichen planopilaris is frontal fibrosing alopecia, which typically affects the frontal hairline of postmenopausal women with a bandlike scarring alopecia, often associated with marked decrease or complete loss of the eyebrows (Fig. 442-4). No therapy is consistently effective, but options include intralesional or systemic steroids, systemic antimalarial agents, and excimer laser treatment.

DISCOID LUPUS ERYTHEMATOSUS

In discoid lupus erythematosus, the alopecic area shows active inflammation with erythema, edema, scaling, and follicular plugging, as well as atrophy with variable degrees of telangiectasia and dyspigmentation (Fig. 442-5).

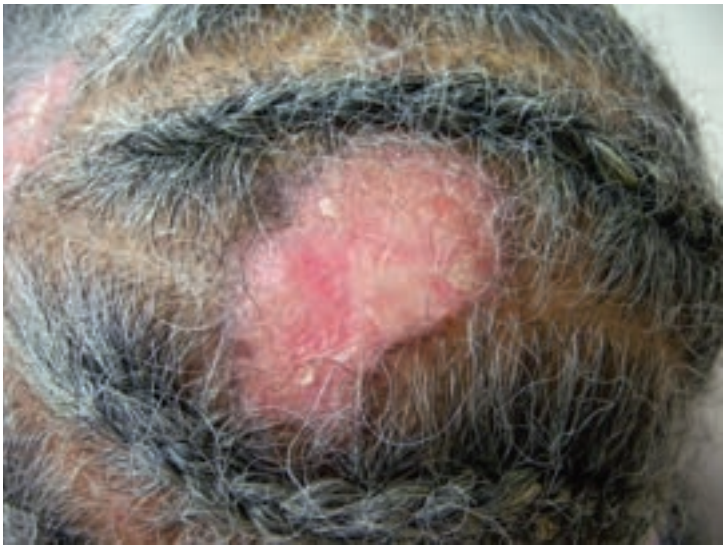

FIGURE 442-5. Discoid lupus erythematosus. The alopecic patch shows erythema, scaling, and depigmentation.

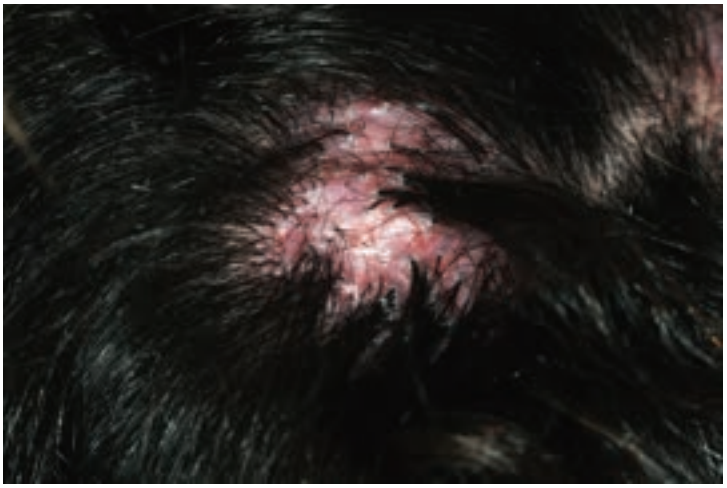

FIGURE 442-6. Folliculitis decalvans. Note alopecia with tufts of six or more hairs emerging together.

The disease is more common in African American women. Approximately 5 to 10% of adults with discoid lupus erythematosus will develop systemic lupus erythematosus (Chapter 266), especially those with widespread discoid lesions. In localized lesions, treatment with high-potency topical steroids (see Table 437-1) is usually effective. Antimalarials are a second-line treatment. Hair regrowth may occur when treatment is promptly started. Therapeutic strategy should include photoprotection of the involved area.

FOLLICULITIS DECALVANS

In folliculitis decalvans, the scalp shows papulopustular lesions that often coalesce to form exudative crusted areas that result in cicatricial alopecia. A typical finding is tufted folliculitis, in which tufts of 6 to 15 hairs emerge together from the scalp (Fig. 442-6). The cause is unknown, but the condition may reflect an abnormal host response to bacterial antigens. *Staphylococcus aureus* often is isolated from active lesions. Folliculitis decalvans responds to oral antibiotics (e.g., trimethoprim-sulfamethoxazole [one tablet, 160 mg/800 mg/day] or clindamycin [300 mg twice daily] with or without rifampin [300 mg twice daily] for 8 to 10 weeks) but usually relapses after interruption of therapy.

DISSECTING CELLULITIS

Dissecting cellulitis of the scalp is a follicular occlusion disorder that may progress to scarring alopecia. Patients typically complain of relapsing multifocal alopecic painful nodules, boggy plaques, and draining tracts. Possible treatments include systemic antibiotics, isotretinoin, and tumor necrosis factor-α inhibitors, but data are limited to small case series.

HAIR DISORDERS IN CHILDREN

Hereditary and congenital alopecias are present since birth or appear in the first years of life. Aplasia cutis congenita is the most common form of focal alopecia in newborns.

Hereditary hair shaft abnormalities associated with increased hair fragility produce diffuse or patchy alopecia that appears during childhood. The most common hereditary hair shaft disorder is monilethrix, in which hair shaft fragility is associated with follicular hyperkeratosis. The hair is dull and fragile and breaks easily, especially in the nape and occipital areas. Diagnosis is confirmed by finding hair beading on dermoscopy or microscopic examination.

Congenital triangular alopecia is usually noticed by the age of 6 years as an irregularly triangular patch of alopecia with vellus hairs on the frontotemporal region.

The loose anagen hair syndrome is characterized by a defective anchorage of the hair to the follicle, resulting in hairs that are easily and painlessly pulled out from the scalp. The condition is typical of children and may manifest with patchy hair loss due to hair pulling while playing.

Short anagen syndrome is a congenital disease characterized by short (<6 cm) fine hair and increased hair shedding. The condition is due to a decreased duration of the anagen phase.

FIGURE 442-7. Traction alopecia. Hair loss involves the temporal scalp. Note presence of remaining hairs along the hairline.

RACIAL DIFFERENCES

The frequency and clinical aspects of hair disorders vary in different races.[9] Androgenetic alopecia, for instance, is more frequent in whites than in blacks and Asians, whereas black and Asian hair is more susceptible to weathering and fragility.

The black hair shaft is flat, highly twisted, and difficult to manage without strong chemical or hair styling procedures, which often cause considerable damage. Hair straightening and braiding are responsible for traction alopecia, which is common and typically produces cicatricial alopecia of the frontal and lateral margins (Fig 442-7). Central centrifugal cicatricial alopecia is a very common cause of scarring alopecia in black women (Fig 442-8). It manifests as slowly progressive scarring hair loss in the vertex or crown that spreads in a centrifugal pattern.

In Asians, the hair shaft is round, thick, robust, and straight. Asian hair is very difficult to style and dye, and it is often damaged by its treatment with high concentrations or long exposure to chemicals.

EXCESSIVE HAIR GROWTH

Hirsutism describes excessive terminal hair with male distribution in a female. Hirsutism should be distinguished from hypertrichosis, which is characterized by the presence of an excessive amount of hair in a non–androgen-dependent area.

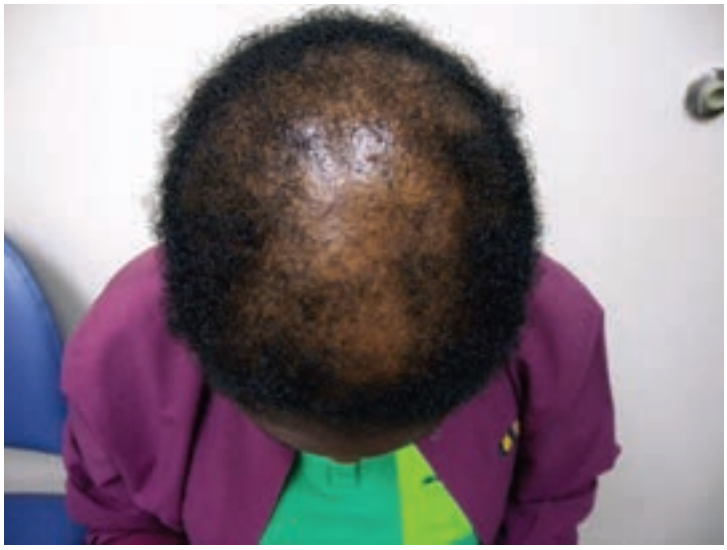

FIGURE 442-8. Central centrifugal cicatricial alopecia. The alopecic area involves the central portion of the scalp and expands centrifugally.

FIGURE 442-9. Spooning.

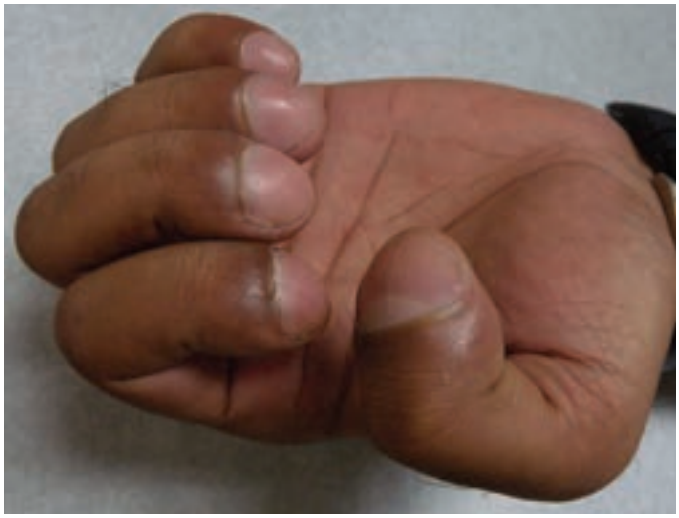

FIGURE 442-10. Clubbing.

Hirsutism

Hirsutism is a common condition that affects up to 10% of women and is more frequent in Hispanic and Mediterranean women.[10] Hirsutism may be associated with hyperandrogenism, but it is idiopathic in approximately 15% of cases.

Hypertrichosis

Hypertrichosis results from the presence of terminal hairs in anatomic areas that are normally characterized by vellus hair. Hypertrichoses can be congenital or acquired, localized or generalized. Acquired hypertrichoses are most commonly iatrogenic, metabolic (e.g., Cushing syndrome, porphyria, hyperthyroidism), nutritional (e.g., anorexia nervosa), or paraneoplastic (Chapter 179).

NAIL DISORDERS

Normal Nail

The nail plate is a keratinized hard structure produced by the nail matrix, which is a specialized epithelium located above the distal phalanx of the finger. In longitudinal sections, the matrix consists of a dorsal, an apical, and a ventral portion. The proximal matrix (dorsal portion and apex) produces the dorsal nail plate (two thirds upper plate), and the distal matrix (ventral portion) produces the ventral plate (one third lower plate). The nail plate is produced continuously throughout life. Nails grow slowly, and the nail plate can reflect illnesses that occurred several months earlier. Complete replacement takes approximately 6 months for fingernails and 12 to 18 months for toenails. Many medications (e.g., anticonvulsants, neuroleptics, antifungals), drugs (e.g., amphetamine, cocaine, doping substances), and poisons (e.g., mercury, arsenic) are retained in the nails, whose fragments (nail clippings) can be used to monitor previous exposure.

NAIL DISORDERS

Nail abnormalities can be congenital or acquired and may caused by developmental, traumatic, inflammatory, infective, and neoplastic disorders or by medications. The diagnosis of nail dystrophies usually relies on a careful clinical examination and an accurate history, but radiographic or magnetic resonance imaging investigation of the digit and pathology may be required.

Koilonychia (Spoon Nails)

In koilonychia (Fig. 442-9), which is also called spooning of the nails, the nail plate is thin and has a concave appearance. Koilonychia is physiologic in children. In adults, it can be occupational or, more rarely, a sign of iron deficiency (Chapter 159).

Clubbing

Clubbing (Fig. 442-10) develops when enlargement of the soft tissue of the distal digit causes a bulbous digit with an enlarged and overcurved nail plate. The angle between the proximal nail fold and the nail plate (Lovibond angle) is greater than 180 degrees. Clubbing may be congenital (i.e., in congenital heart disease; Chapter 69) or acquired. Other causes of acquired clubbing include intrathoracic (Chapter 191) and gastrointestinal neoplasms (Chapters 192 to 196), chronic intrathoracic suppurative disease (Chapters 90 and 99), inflammatory bowel disease (Chapter 141), and liver disorders.

Beau Lines and Onychomadesis

Beau lines (Fig. 442-11) and onychomadesis are due to a temporary reduction or arrest of nail growth. Beau lines appear as transverse grooves of various depth; onychomadesis as a full-thickness transverse groove of the proximal nail plate. Causes include trauma, skin diseases involving the proximal nail fold and the matrix, drugs, and systemic diseases (Table 442-4). In the latter case, Beau lines or onychomadesis involve all the nails and are localized at the same level.

Pitting

Pitting (Fig. 442-12) appears as punctate depressions of the dorsal nail plate, with variable size and depth. Pitting is caused by inflammatory skin disorders such as psoriasis (Chapter 438), alopecia areata, and eczema (Chapter 438).

Longitudinal Grooves and Striations

Normal nails often show superficial thin longitudinal ridges that increase in number with aging. Deep longitudinal fissures, which indicate damage to the proximal matrix, can be caused by nail lichen planus, vascular insufficiency, trauma, and tumors involving or compressing the matrix.

TABLE 442-4 CAUSES OF BEAU LINES AND ONYCHOMADESIS

SYSTEMIC
Acrodermatitis enteropathica
Severe metabolic stress
High fever
Viral infections (Kawasaki disease, measles, hand-foot-and-mouth syndrome)
Typhus
Stevens-Johnson syndrome
Drugs
Bullous disorders (pemphigus, pemphigoids)
Deep saturation, high altitude
Hemodialysis
Myocardial infarction
LOCAL
Trauma (including manicure)
Paronychia
Congenital malalignment (big toenails)

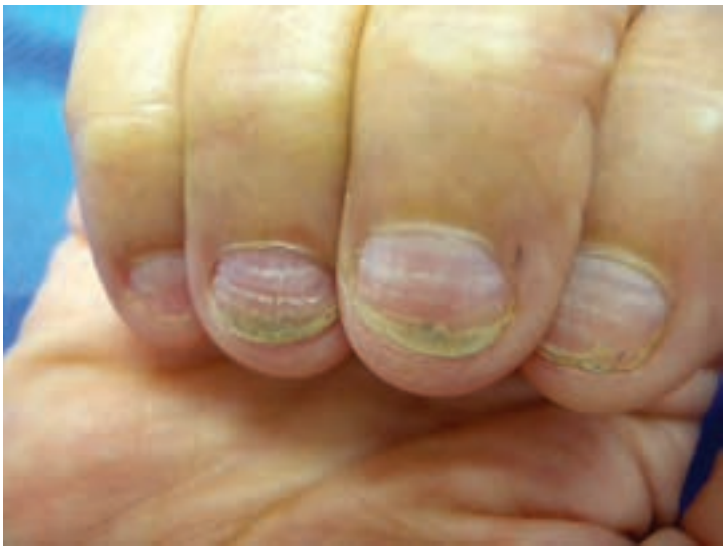

FIGURE 442-11. Beau lines due to chemotherapy. The lines affect all the nails at the same level.

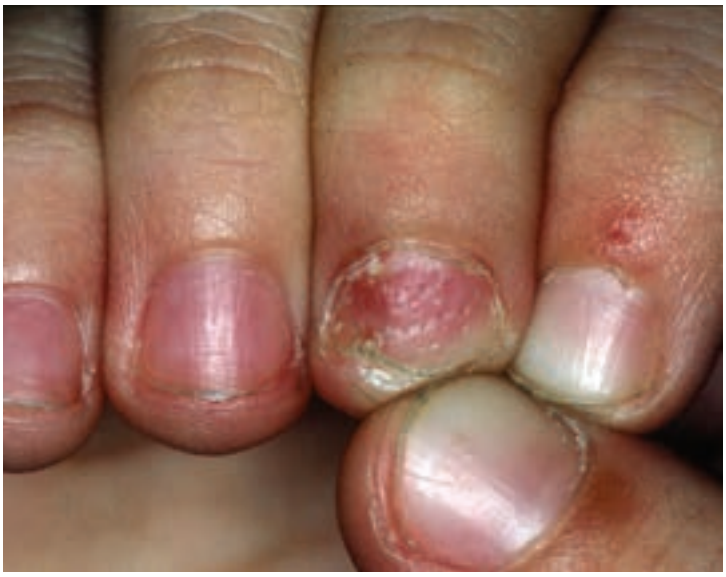

FIGURE 442-12. Pitting. Multiple punctate depressions on the nail plate surface.

Leukonychia

Leukonychia describes a whitish discoloration of the nail, which may be due to persistence of nuclei in the cells of the ventral nail plate (true leukonychia) or to a pallor of the nail bed (apparent leukonychia). True leukonychia, including the Mees lines of arsenic exposure (Chapter 22), does not fade with pressure and moves distally with nail growth; it is most commonly caused by trauma. Apparent leukonychia, which does not follow nail growth and fades with pressure, may be a sign of systemic diseases such as liver cirrhosis (Terry nails; Chapter 146), chronic renal diseases (half-and-half nails, characterized by apparent leukonychia of the proximal half of the nail; Chapter 130), hypoalbuminemia (Chapter 121), and systemic chemotherapy (Muehrcke lines; Fig. 442-13).

Yellow Nail Syndrome

The yellow nail syndrome is a chronic nail disorder characterized by an arrest or a reduction of nail growth, resulting in nail thickening and hardening and yellow discoloration.[11] Fingernails and toenails are excessively curved from side to side, and cuticles are absent (Fig. 442-14). Yellow nail syndrome occasionally may be paraneoplastic (Chapter 179). The pathogenesis of yellow nail syndrome is unknown, but a congenital abnormality of the lymphatic vessels may be involved. Typical cases have associated lymphedema or respiratory disturbances. The nail abnormalities improve with treatment of the associated respiratory disorders. Oral vitamin E (1200 mg/day for several months) is useful in some cases.

Splinter Hemorrhages

Splinter hemorrhages (see Fig. 51-11) appear as longitudinal thin red-brown lines of variable length. Splinter hemorrhages are usually localized in the distal nail and are commonly seen in inflammatory diseases, including eczema (Chapter 438), psoriasis (Chapter 438), and onychomycosis. Multiple splinter hemorrhages localized in the proximal nail plate can be a sign of systemic

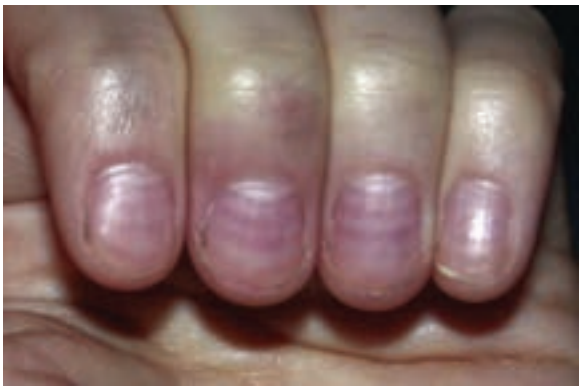

FIGURE 442-13. Muehrcke lines. (From Wikipedia Commons.)

FIGURE 442-14. Yellow nail syndrome. The nails are yellow, overcurved, and do not grow.

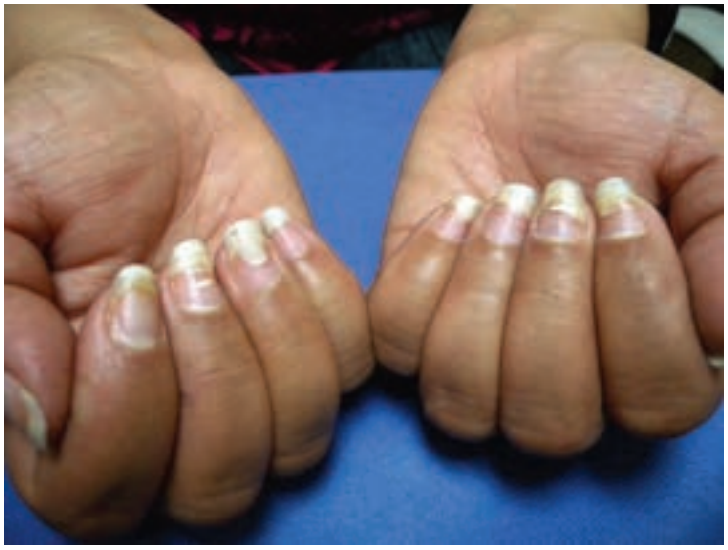

FIGURE 442-15. Onycholysis. The detached nail plate is white in color.

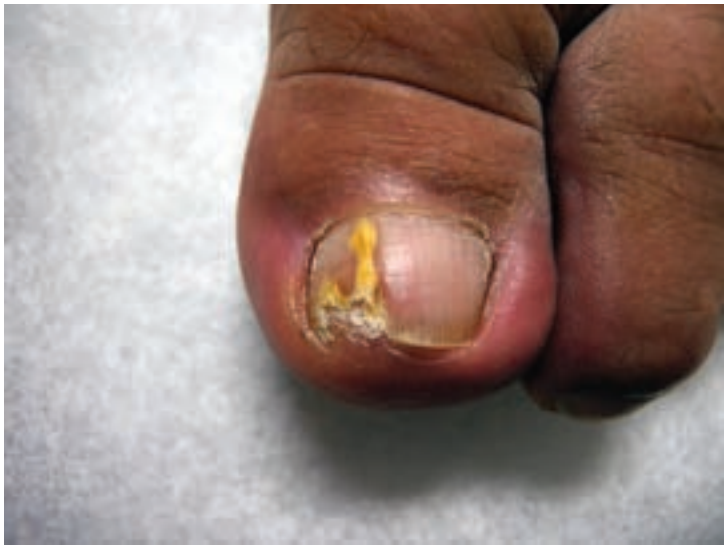

FIGURE 442-16. Distal subungual onychomycosis. The nail shows subungual hyperkeratosis and a yellow streak.

diseases, including infectious or marantic endocarditis (Chapter 76), trichinosis (Chapter 357), and the antiphospholipid syndrome.

Onycholysis

Onycholysis (Fig. 442-15) describes detachment of the nail plate from the bed. The detachment usually occurs at the free lateral margins of the nail. The onycholytic area is white owing to the presence of air, but it may acquire a green-brown color if the space if colonized by bacteria, such as *Pseudomonas aeruginosa*.

Onycholysis of the fingernails is a common sign of nail psoriasis (Chapter 438). It also may be due to prolonged and frequent contact with water, detergents, or irritants (idiopathic onycholysis). Toenail onycholysis is almost exclusively caused by trauma or onychomycosis. When onycholysis is limited to one digit, the possibility of a nail tumor always should be considered.

Paronychia

Paronychia, which describes the acute or chronic inflammation of the proximal and lateral nail folds, is common in the fingernails at any age. In acute paronychia, the affected digit is painful, with erythema, swelling, and pus discharge localized to one corner of the proximal nail fold. Acute paronychia usually follows a trauma to the nail fold, as in children who pick or bite the cuticles or in women after a manicure.

In chronic paronychia, prolonged mechanical or environmental trauma (such as contact with water and irritants) damages the cuticle, allowing penetration of dirt, bacteria, and other particles under the proximal nail fold. The result is an inflammatory reaction of the proximal nail fold and nail matrix, with edema and redness of the fold, absence of the cuticles, Beau lines, and abnormalities of the nail plate surface. Treatment includes protective measures, such as use of cotton and rubber gloves to avoid contact with irritants, as well as topical steroids and topical antimicrobials.

Onychomycosis

Fungal nail infections most commonly affect the toenails of adults. Dermatophytes (particularly *Trichophyton rubrum*) are responsible for most infections. The clinical presentation varies depending on the modality of the nail invasion. In distal subungual onychomycosis, the most common form, fungi spread from plantar skin and invade the nail bed. The affected nail shows subungual hyperkeratosis, onycholysis, and yellow streaks (Fig. 442-16). In white superficial onychomycosis, which only affects toenails, fungi colonize the surface of the nail plate, where they cause multiple white friable patches (Fig. 442-17). Proximal subungual onychomycosis produces a true leukonychia, owing to the presence of fungal hyphae in the deep layers of the plate. Proximal subungual onychomycosis caused by *T. rubrum* is typical in immunosuppressed patients. Diagnosis of onychomycosis must always be confirmed by mycologic examination. Treatment depends on clinical type, number of affected nails, and severity of nail involvement. A systemic treatment is always required for proximal subungual onychomycosis and for distal subungual onychomycosis involving the proximal nail. Terbinafine (250 mg/day) for 2 (fingernails) or 3 (toenails) months is the most effective treatment for dermatophyte infections.

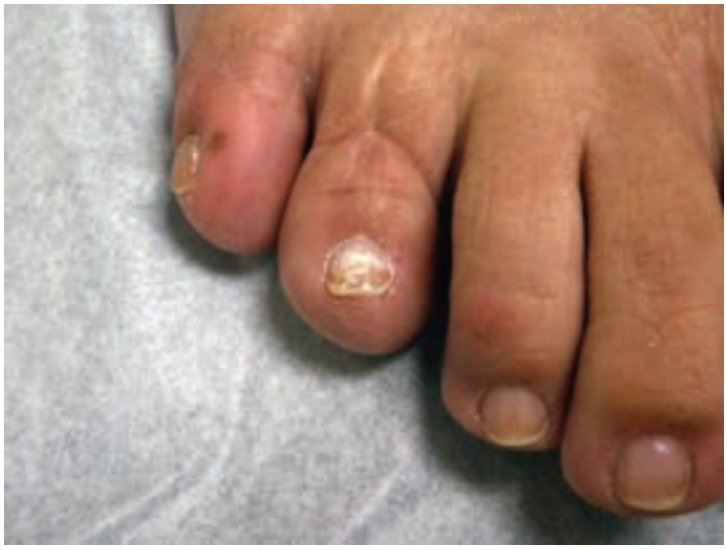

FIGURE 442-17. White superficial onychomycosis. The affected nail shows white superficial patches.

Ingrowing Toenails

Ingrown toenails are a common condition, especially in young patients. Nail ingrowing most commonly affects one or both the big toes and is related to genetic factors, hyperhidrosis, and poorly fitting shoes. Ingrowing is usually precipitated by incorrect nail trimming, with formation of a sharp edge (spicule) of the lateral nail plate that penetrates and injures the soft tissues of the lateral nail fold (Fig. 442-18). Depending on severity, treatment varies from simple disembedding of the spicule to chemical destruction of the lateral nail matrix by phenolization.[12]

Nail Pigmentation

Nail pigmentation is usually caused by staining from external agents, such as nicotine or hair dyes. It rarely may be due to drug deposition in the nail plate or into the nail bed (i.e., antimalarials) or systemic diseases (argyria). In these cases, the proximal margin of the pigmentation follows the shape of the lunula.

Melanonychia

Melanonychia is defined by the presence of melanin within the nail plate. It appears more often as a longitudinal brown-black band starting from the matrix and extending to the free edge of the nail plate (Fig. 442-19).

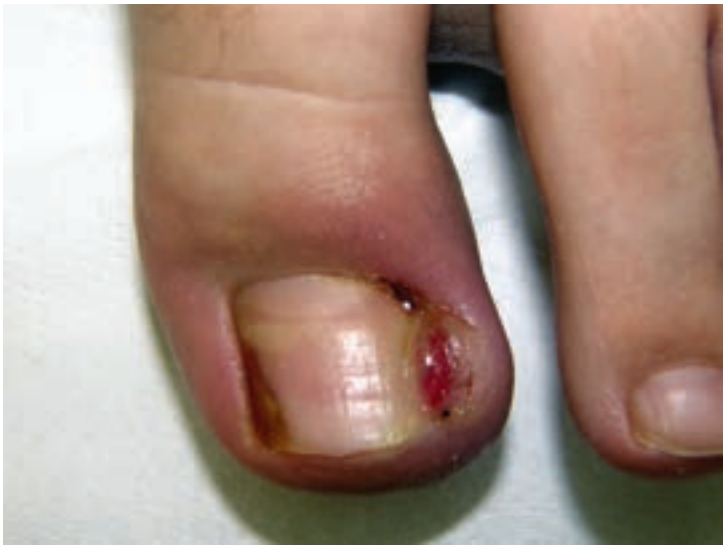

FIGURE 442-18. Ingrowing toenail. Penetration of the nail spicule causes inflammation and granulomatous reaction.

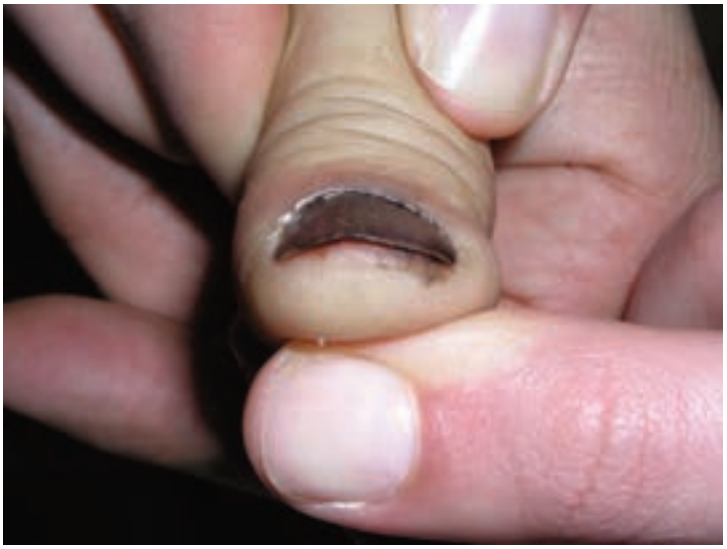

FIGURE 442-20. Nail melanoma. Melanonychia and periungual pigmentation (Hutchinson sign).

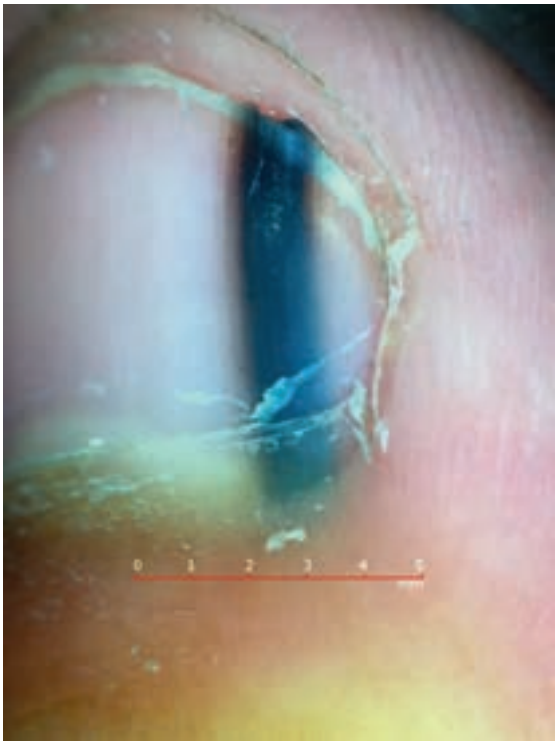

FIGURE 442-19. Longitudinal melanonychia. Black longitudinal pigmented band extending from the proximal nail fold to the free edge.

Melanonychia results from production of melanin by melanocytes of the nail matrix. Melanonychia has three main causes: simple melanocytic activation, benign melanocyte proliferations (lentigo, nevus), and malignant melanocyte proliferation (melanoma; Chapter 203).

Common causes of longitudinal melanonychia due to melanocytic activation include inflammatory and traumatic nail disorders, drugs (chemotherapy, azidothymidine, antimalarials, psoralen and ultraviolet A [PUVA] therapy), and systemic diseases (acquired immunodeficiency syndrome [Chapter 392]; Addison disease [Chapter 227]).

Nail melanoma is rare and most frequently involves the thumb of middle-aged individuals. Diagnosis is often delayed, and the 5-year survival is only 15%. Hutchinson sign, extension of the pigmentation to the proximal or lateral nail folds, is an important indicator of nail melanoma (Fig. 442-20).

Grade A References

A1. Shin H, Jo SJ, Kim do H, et al. Efficacy of interventions for prevention of chemotherapy-induced alopecia: a systematic review and meta-analysis. *Int J Cancer*. 2015;136:E442-E454.

A2. van Zuuren EJ, Fedorowicz Z, Carter B. Evidence-based treatments for female pattern hair loss: a summary of a Cochrane systematic review. *Br J Dermatol*. 2012;167:995-1010.

A3. Varothai S, Bergfeld WF. Androgenetic alopecia: an evidence-based treatment update. *Am J Clin Dermatol*. 2014;15:217-230.

A4. de Sá DC, Lamas AP, Tosti A. Oral therapy for onychomycosis: an evidence-based review. *Am J Clin Dermatol*. 2014;15:17-36.

GENERAL REFERENCES

For the General References and other additional features, please visit Expert Consult at https://expertconsult.inkling.com.

GENERAL REFERENCES

1. Mubki T, Rudnicka L, Olszewska M, et al. Evaluation and diagnosis of the hair loss patient: part I. History and clinical examination. *J Am Acad Dermatol*. 2014;71:415.e1-415.e15.
2. Mubki T, Rudnicka L, Olszewska M, et al. Evaluation and diagnosis of the hair loss patient: part II. Trichoscopic and laboratory evaluations. *J Am Acad Dermatol*. 2014;71:431.e1-431.e11.
3. Betticher DC, Delmore G, Breitenstein U, et al. Efficacy and tolerability of two scalp cooling systems for the prevention of alopecia associated with docetaxel treatment. *Support Care Cancer*. 2013;21: 2565-2573.
4. Su LH, Chen LS, Lin SC, et al. Association of androgenetic alopecia with mortality from diabetes mellitus and heart disease. *JAMA Dermatol*. 2013;149:601-606.
5. Gilhar A, Etzioni A, Paus R. Alopecia areata. *N Engl J Med*. 2012;366:1515-1525.
6. Islam N, Leung PS, Huntley AC, et al. The autoimmune basis of alopecia areata: a comprehensive review. *Autoimmun Rev*. 2015;14:81-89.
7. Hordinsky M, Donati A. Alopecia areata: an evidence-based treatment update. *Am J Clin Dermatol*. 2014;15:231-246.
8. Xing L, Dai Z, Jabbari A, et al. Alopecia areata is driven by cytotoxic T lymphocytes and is reversed by JAK inhibition. *Nat Med*. 2014;20:1043-1049.
9. Rodney IJ, Onwudiwe OC, Callender VD, et al. Hair and scalp disorders in ethnic populations. *J Drugs Dermatol*. 2013;12:420-427.
10. Loriaux DL. An approach to the patient with hirsutism. *J Clin Endocrinol Metab*. 2012;97: 2957-2968.
11. Piraccini BM, Urciuoli B, Starace M, et al. Yellow nail syndrome: clinical experience in a series of 21 patients. *J Dtsch Dermatol Ges*. 2014;12:131-137.
12. Di Chiacchio N, Belda W Jr, Di Chiacchio NG, et al. Nail matrix phenolization for treatment of ingrowing nail: technique report and recurrence rate of 267 surgeries. *Dermatol Surg*. 2010;36: 534-537.

REVIEW QUESTIONS

1. A 57-year-old woman complains of progressive hair loss involving the frontotemporal scalp and the eyebrows. She states that her hairline has been receding. Clinical examination shows a band of scarring alopecia on the frontotemporal scalp, as well as alopecia of the eyebrows. Perifollicular scaling is evident around the terminal follicles at the hairline, and dermoscopy shows perifollicular casts. A scalp biopsy shows loss of follicles and a perifollicular, lichenoid lymphocytic infiltrate with fibrosis. Which of the following is the *most* appropriate diagnosis?

A. Traction alopecia
B. Alopecia areata
C. Androgenetic alopecia (female pattern hair loss)
D. Frontal fibrosing alopecia
E. Trichotillomania

Answer: D This patient has frontal fibrosing alopecia, which is a variant of lichen planopilaris and typically affects the frontal hairline of postmenopausal women with a bandlike scarring alopecia. It often is associated with marked decrease or complete loss of the eyebrows. Frontal fibrosing alopecia is a relatively newly defined condition that is becoming increasingly common worldwide. The cause of this condition is unknown. Possible treatments include intralesional or systemic steroids, systemic antimalarial agents, and excimer laser therapy.

2. An 18-year-old man complains of alopecia involving his entire scalp, eyelashes, and eyebrows since the age of 14. Family history is positive for a sister with type I diabetes. Dermoscopy shows yellow dots corresponding to empty follicular openings. Diagnosis of alopecia areata universalis is established. The patient is otherwise healthy and is seeking a rapidly effective treatment. Which of the following is the *most* appropriate treatment regimen?

A. High-dose pulse corticosteroid therapy
B. Intralesional steroid injections
C. Clobetasol propionate cream 0.05% under occlusion
D. Topical minoxidil 5% twice per day
E. Topical immunotherapy with diphenylcyclopropenone

Answer: C High-dose pulse steroids are effective in patients with acute hair shedding but not in long-standing alopecia universalis, and intralesional steroids are an option only in circumscribed disease. Topical immunotherapy can be effective, but regrowth usually is not observed for 12 to 18 months. Topical minoxidil is not effective. The best choice for this patient is clobetasol propionate cream 0.05% under occlusion. Complete regrowth will be observed in 25% of patients with severe alopecia areata after 3 to 4 months of therapy.

3. A 32-year-old man complains of progressive thickening and yellow discoloration of his big toenails. Clinical examination shows dryness and desquamation of the plantar skin, subungual hyperkeratosis with onycholysis, and longitudinal yellow streaks of the first toenail bilaterally. Which of the following is the *most* appropriate recommendation?

A. Start terbinafine 250 mg/day for 3 months
B. Start terbinafine 250 mg/day for 6 weeks
C. Take a nail clipping for KOH testing or culture
D. Start itraconazole 200 mg/day for 3 months
E. Start fluconazole 150 mg/week for 9 months

Answer: C The diagnosis of onychomycosis always requires laboratory confirmation. Numerous common nail diseases, including psoriasis and trauma, can appear very similar to onychomycosis on clinical examination. In this patient, culture isolated *T. rubrum.* Terbinafine 250 mg/day is started if laboratory testing confirms normal baseline liver enzymes.

4. A 64-year-old woman complains of progressive pigmentation of her left thumbnail over the past 2 years. Clinical examination shows a 5-mm large band of longitudinal melanonychia associated with pigmentation of the proximal nail fold. Which of the following is the *most* appropriate recommendation?

A. Take a punch biopsy of the band
B. Follow up after 6 months
C. Take a nail clipping for pathologic examination
D. Refer for a nail matrix biopsy
E. Reassure her

Answer: D Longitudinal melanonychia with periungual pigmentation is highly suspicious for nail melanoma. Failure to diagnose nail melanoma can have fatal consequences for the patient. Diagnosis requires a nail matrix biopsy. A biopsy from the band is not appropriate for diagnosis because it will sample the nail bed, which is not the site of the tumor.

APPENDIX

REFERENCE INTERVALS AND LABORATORY VALUES

RONALD J. ELIN

Reference intervals are invaluable guidelines based upon the methodology of the laboratory for the clinician to assess health and disease, but they should not be used as absolute indicators of health and disease. For essentially every test, significant overlap exists between healthy and diseased populations. Many factors may influence the determination of the reference interval. The method and mode of standardization are variables for the reference interval, particularly for immunologic and enzymatic tests. The selection of the "normal" population is also important because factors such as age, gender, race, diet, personal habits (e.g., alcohol consumption, smoking, etc.), and exercise may influence the reference interval for a given analyte. Finally, the statistics chosen to define the reference interval are also a factor. These multiple variables for determining the reference interval indicate why differences exist among institutions for the same analyte.

The values in this appendix are primarily for adults in the fasting state.[1,2] Values for other groups, when included, are clearly identified. Prefixes and abbreviations are listed in Table Appendix-1. Clinical chemistry, toxicology, and serology are summarized in Table Appendix-2; hematology and coagulation in Table Appendix-3; and drugs, therapeutic and toxic, in Table Appendix-4. The lists include reference intervals for the most common tests used in the practice of internal medicine.

All laboratory values are given in conventional and international units. If the value and units for a reference interval are the same for conventional and international units, the interval is listed only in the column for international units. The temperature for all enzyme assays listed is 37° C. The pertinent prefixes denoting the decimal factors and abbreviations are listed in Table Appendix-1.

REFERENCES

1. Burtis CA, Ashwood ER, Bruns DE, eds. *Tietz Textbook of Clinical Chemistry and Molecular Diagnosis.* 5th ed. St. Louis: Elsevier Saunders; 2012. *A comprehensive text of clinical chemistry.*
2. Wu AHB, ed. *Tietz Clinical Guide to Laboratory Tests.* 4th ed. St. Louis: Elsevier Saunders; 2006. *A comprehensive text of reference intervals.*

TABLE APPENDIX-1 PREFIXES AND ABBREVIATIONS

PREFIXES DENOTING DECIMAL FACTORS

PREFIX	SYMBOL	FACTOR
mega	M	10^{6}
kilo	k	10^{3}
hecto	h	10^{2}
deca	da	10^{1}
deci	d	10^{-1}
centi	c	10^{-2}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}

ABBREVIATIONS

AU	Arbitrary units
CSF	Cerebrospinal fluid
EDTA	Ethylenediaminetetraacetic acid
EU	Ehrlich unit
F	Female
Hb	Hemoglobin
hpf	High-power field
IFA	Immunofluorescent assay
IFCC	International Federation of Clinical Chemistry
IU	International unit (of hormone activity)
Kat	Katal
M	Male
Pa	Pascal
RBC	Red blood cell
RID	Radial immunodiffusion
S	Substrate
U	International unit (of enzyme activity)
WBC	White blood cell

TABLE APPENDIX-2 CLINICAL CHEMISTRY, TOXICOLOGY, AND SEROLOGY

TEST	SPECIMEN	REFERENCE INTERVAL (CONVENTIONAL UNITS)	REFERENCE INTERVAL (INTERNATIONAL UNITS)
Acetoacetate Semiquantitative	Serum or plasma (fluoride/oxalate)	Negative (<1 mg/dL)	Negative (<0.1 mmol/L)
Acetone Semiquantitative	Urine Serum or plasma (fluoride or oxalate)	Negative Negative (<1 mg/dL)	Negative Negative (0.17 mmol/L)
Acid phosphatase (S:p-nitrophenylphosphate)	Serum		M: 2.5-11.7 U/L F: 0.3-9.2 U/L
Adrenocorticotropic hormone (ACTH)	Plasma (EDTA)	0800 hr: <120 pg/mL 24-hr, supine: <85 pg/mL	<26 pmol/L <19 pmol/L
Alanine aminotransferase (ALT, SGPT)	Serum	M: <45 U/L F: <34 U/L	<0.77 μKat/L <0.58 μKat/L
Albumin Nephelometric, colorimetric	Serum CSF Urine	3.4-4.8 g/dL 17.7-25.1 mg/dL	34-48 g/L 177-251 mg/L 3.9-24.4 mg/day
Aldolase	Serum	2.5-10.0 U/L	0.04-0.17 μKat/L
Aldosterone	Plasma (heparin EDTA) or serum	Adult, average sodium diet Supine: 3-16 ng/dL Upright: 7-30 ng/dL	 0.8-0.44 nmol/L 0.19-0.83 nmol/L

TABLE APPENDIX-2 CLINICAL CHEMISTRY, TOXICOLOGY, AND SEROLOGY—cont'd

TEST	SPECIMEN	REFERENCE INTERVAL (CONVENTIONAL UNITS)	REFERENCE INTERVAL (INTERNATIONAL UNITS)
Alkaline phosphatase, IFCC	Serum	M: (>20 yr): 53-128 U/L F: (>20 yr): 42-98 U/L	0.90-2.18 µKat/L 0.71-1.67 µKat/L
Aluminum	Serum	<5.41 µg/L	<0.2 µmol/L
δ-Aminolevulinic acid (δ-ALA)	Serum Urine	15-23 µg/dL 1.5-7.5 mg/day	1.1-8 µmol/L 11.4-57.2 µmol/day
Ammonia nitrogen	Serum or plasma (Na-heparin) Urine, 24-hr	15-45 µg N/dL 140-1500 mg N/day	11-32 µmol/L 10-107 mmol N/day
Amylase, IFCC	Serum Urine, timed specimen	28-100 U/L	0.48-1.70 µKat/L 1-17 U/hr
Angiotensin I	Peripheral venous plasma (EDTA)	<25 pg/mL	<25 ng/L
Angiotensin II	Arterial blood plasma (EDTA)	10-60 pg/mL	10-60 ng/L
α_1-Antitrypsin	Serum	90-200 mg/dL	0.9-2.0 g/L
Anion gap $[Na^+ - (Cl^- + HCO_3^-)]$	Serum or plasma (heparin)	7-16 mEq/L	7-16 mmol/L
Arsenic	Whole blood (heparin)	0.2-2.3 µg/dL Chronic poisoning: 10-50 µg/dL Acute poisoning: 60-930 µg/dL	0.3-0.31 µmol/L 1.33-6.65 µmol/L 7.98-124 µmol/L
Ascorbic acid (see Vitamin C)			
Aspartate aminotransferase (AST, SGOT)	Serum	M: <35 U/L F: <31 U/L	<0.60 µKat/L <0.53 µKat/L
Bicarbonate	Serum	22-29 mEq/L	22-29 mmol/L
Bilirubin, conjugated (direct)	Serum	0-0.2 mg/dL	0-3.4 µmol/L
Bilirubin, total	Serum Urine	0-2.0 mg/dL	0-34 µmol/L Negative
B-type natriuretic peptide (BNP)	Whole blood or plasma (EDTA)	<100 pg/mL	<28.8 pmol/L
Calcium, ionized (iCa)	Serum	4.65-5.28 mg/dL	1.16-1.32 mmol/L
Calcium, total	Serum Urine, 24-hr	8.6-10.2 mg/dL 100-300 mg/day	2.15-2.55 mmol/L 2.5-7.5 mmol/day
Cancer antigen 125 (CA 125)	Serum	<35 U/mL	<35 kU/L
Cancer antigen 15-3 (CA 15-3)	Serum	<30 U/mL	<30 kU/L
Carbohydrate antigen 19-9 (CA 19-9)	Serum	<37 U/mL	<37 kU/L
Carbon dioxide, partial pressure (P_{CO_2})	Whole blood, arterial (heparin)	M: 35-48 mm Hg F: 32-45 mm Hg	4.66-6.38 kPa 4.26-5.99 kPa
Carbon dioxide, total (T_{CO_2})	Serum or plasma (heparin)	23-29 mEq/L	23-29 mmol/L
Carcinoembryonic antigen (CEA)	Serum	Nonsmokers: <3 ng/mL	<3 µg/L
β-Carotene	Serum	10-85 µg/dL	0.19-1.58 µmol/L
Catecholamines, total	Urine, 24-hr	<100 µg/day	<5.91 nmol/day
Ceruloplasmin	Serum	20-60 mg/dL	0.2-0.6 g/L
Chloride	Serum or plasma (heparin) Urine, 24-hr	98-107 mEq/L 110-250 mEq/day	98-107 mmol/L 110-250 mmol/day
Cholesterol, total	Serum or plasma (EDTA)	Recommended: <200 mg/dL Moderate risk: 200-239 mg/dL High risk: >239 mg/dL	<5.18 mmol/L 5.18-6.19 mmol/L >6.19 mmol/L
Chorionic gonadotropin, intact	Serum or plasma (EDTA)	M and nonpregnant F:	<5.0 IU/L
Complement C3	Serum	90-180 mg/dL	0.9-1.8 g/L
Complement C4	Serum	10-40 mg/dL	0.1-0.4 g/L
Copper	Serum Urine, 24-hr	M: 70-140 µg/dL F: 80-155 µg/dL <60 µg/day	10.99-21.98 µmol/L 12.56-24.34 µmol/L <1.0 µmol/day
Cortisol, total	Serum or plasma (heparin)	0800 hr: 5-23 µg/dL 1600 hr: 3-16 µg/dL 2000 hr: ≤50% of 0800 hr	138-635 nmol/L 83-441 nmol/L Fraction of 0800 hr: ≤0.50
Cortisol, free	Urine, 24-hr	20-90 µg/day	55-248 nmol/day
C-peptide	Serum	0.78-1.89 ng/mL	0.26-0.62 nmol/L
C-reactive protein (CRP)	Serum	<0.5 mg/dL	<5 mg/L
Creatine kinase (CK) IFCC	Serum	M: 46-171 U/L F: 34-145 U/L	M: 0.78-2.90 µKat/L F: 0.58-2.47 µKat/L

TABLE APPENDIX-2 CLINICAL CHEMISTRY, TOXICOLOGY, AND SEROLOGY—cont'd

TEST	SPECIMEN	REFERENCE INTERVAL (CONVENTIONAL UNITS)	REFERENCE INTERVAL (INTERNATIONAL UNITS)
Creatinine	Serum	M: 0.62-1.10 mg/dL	55-96 µmol/L
Enzymatic		F: 0.45-0.75 mg/dL	40-66 µmol/L
	Urine, 24-hr	M: 14-26 mg/kg/day	124-230 µmol/kg/day
		F: 11-20 mg/kg/day	97-177 µmol/kg/day
Creatinine clearance (glomerular filtration rate endogenous)	Serum or plasma, and urine	M: 90-139 mL/min/1.73 m^2	0.87-1.34 mL/sec/m^2
		F: 80-125 mL/min/1.73 m^2	0.77-1.20 mL/sec/m^2
Dehydroepiandrosterone (DHEA), unconjugated	Serum	M: 1.8-12.5 ng/mL	6.25-43.4 nmol/L
		F: 1.3-9.8 ng/mL	4.51-34.0 nmol/L
11-Deoxycortisol (compound S)	Serum	20-158 ng/dL	0.6-4.6 nmol/L
Erythropoietin	Serum		5-36 U/L
Estradiol	Serum	M: 10-50 pg/mL	37-184 pmol/L
		F, cycle:	
		Follicular phase: 20-350 pg/mL	73-1285 pmol/L
		Luteal phase: 30-450 pg/mL	110-1652 pmol/L
		Postmenopausal: <21 pg/mL	<74 pmol/L
Fat, fecal	Feces, 72-hr		<7 g/day
Fatty acids, nonesterified (free)	Serum or plasma (heparin)	8-25 mg/dL	0.28-0.89 mmol/L
Ferritin	Serum	M: 20-250 ng/mL	20-250 µg/L
		F: 10-120 ng/mL	10-120 µg/L
α_1-Fetoprotein (AFP)	Serum	<15 ng/mL	<15 µg/L
Fibrinogen (see Table Appendix-3)			
Folate	Serum	2.6-12.2 ng/mL	6.0-28.0 nmol/L
	Erythrocytes (EDTA)	103-411 ng/mL packed cells	237-948 nmol/L packed cells
Follitropin (FSH)	Serum or plasma (heparin)	M: 1.4-15.4 mIU/mL	1.4-15.4 IU/L
		F: Follicular phase: 1.4-9.9 mIU/L	1.4-9.9 IU/L
		Luteal phase: 1.1-9.2 mIU/mL	1.1-9.2 IU/L
		Postmenopausal: 19.3-100.6 mIU/mL	19.3-100.6 IU/L
Fructosamine	Serum		205-285 µmol/L
Gastrin	Serum	25-90 pg/mL	25-90 ng/L
Glucagon	Plasma (heparin or EDTA)		70-180 ng/L
Glucose	Serum, fasting	Adult: 74-100 mg/dL	4.1-5.6 mmol/L
		>60 yr: 82-115 mg/dL	4.6-6.4 mmol/L
	Whole blood (heparin)	65-95 mg/dL	3.6-5.3 mmol/L
	CSF	40-70 mg/dL	2.2-3.9 mmol/L
Quantitative, enzymatic	Urine	<0.5 g/day	<2.8 mmol/day
Qualitative	Urine		Negative
Glucose, 2-hr postprandial	Serum	<126 mg/dL	<7.0 mmol/L
Glucose tolerance test (GTT), oral	Serum	mg/dL: *Normal* / *Diabetic*; mmol/L: *Normal* / *Diabetic*	
		Fasting: 70-105 / >140; 3.9-5.8 / >7.8	
		60 min: 120-170 / ≥200; 6.7-9.4 / ≥11	
		90 min: 100-140 / ≥200; 5.6-7.8 / ≥11	
		120 min: 70-120 / ≥140; 3.9-6.7 / ≥7.8	
γ-Glutamyltransferase (GGT) IFCC	Serum	M: <55 U/L	<0.94 µKat/L
		F: <38 U/L	<0.65 µKat/L
Glycated hemoglobin (HbA_{1c})	Whole blood (EDTA or heparin)	Diagnosis: <6.5% total Hb (NGSP)	<6.5 Hb fraction
Growth hormone (hGH, somatotropin)	Serum	Adult, M: 0-4 ng/mL	0-4 µg/L
		F: 0-18 ng/mL	0-18 µg/L
		>60 yr, M: 1-9 ng/mL	1-9 µg/L
		F: 1-16 ng/mL	1-16 µg/L
Haptoglobin (see Table Appendix-3)			
HDL cholesterol (HDLC) (5th percentile from Lipid Research Clinics)	Serum or plasma (EDTA)	M: >29 mg/dL	>0.75 mmol/L
		F: >35 mg/dL	>0.91 mmol/L
5-Hydroxyindole acetic acid (5- HIAA)	Plasma	5.2-13.4 ng/L	27-70 nmol/L
	Urine	<16 mg/g creatinine	<13 mmol/mol creatinine
17-Hydroxyprogesterone (17-OHP)	Serum	M: 27-199 ng/dL	0.8-6.0 nmol/L
		F, Follicular: 15-70 ng/dL	0.4-2.1 nmol/L
		Luteal: 35-290 ng/dL	1.0-8.7 nmol/L
		Postmenopausal: ≤0.70 ng/dL	≤2.1 nmol/L
Immunoglobulin A (IgA)	Serum	90-410 mg/dL	0.9-4.1 g/L
Immunoglobulin D (IgD)	Serum	0-384 ng/mL	0-384 µg/L

TABLE APPENDIX-2 CLINICAL CHEMISTRY, TOXICOLOGY, AND SEROLOGY—cont'd

TEST	SPECIMEN	REFERENCE INTERVAL (CONVENTIONAL UNITS)	REFERENCE INTERVAL (INTERNATIONAL UNITS)
Immunoglobulin E (IgE)	Serum	0-160 kIU/L	0-380 µg/L
Immunoglobulin G (IgG)	Serum CSF	700-1,600 mg/dL 0-5.5 mg/dL	7.0-16 g/L 0-55 mg/L
Immunoglobulin M (IgM)	Serum	40-230 mg/dL	0.4-2.3 g/L
Insulin	Serum	2-25 µIU/mL	12-150 pmol/L
Intrinsic factor (see Vitamin B_{12})			
Iron	Serum	M: 65-175 µg/dL F: 50-170 µg/dL	11.6-31.3 µmol/L 9.0-30.4 µmol/L
Iron-binding capacity, total (TIBC)	Serum	250-450 µg/dL	44.8-80.6 µmol/L
Iron saturation	Serum	 M: 20-50 F: 15-50	Fraction of iron saturation: 0.20-0.50 0.15-0.50
Ketone bodies Qualitative	 Serum Urine, random	 Negative	 Negative Negative
L-Lactate	Whole blood (heparin)	Venous: 3-7 mg/dL Arterial: 16-17 mg/dL	0.36-0.75 mmol/L 1.78-1.88 mmol/L
Lactate dehydrogenase (LDH) IFCC	Serum	125-220 U/L	2.1-3.7 µKat/L
Lead	Whole blood (heparin) Urine	<25 µg/dL Toxic: ≥100 µg/dL <80 µg/dL	<1.21 µmol/L ≥4.83 µmol/L <0.39 µmol/L
Lipase	Serum	<38 U/L	<0.65 µKat/L
Low-density lipoprotein cholesterol (LDLC)	Serum or plasma (EDTA)	Recommended: <100 mg/dL Mild risk: 100-129 mg/dL Moderate risk: 130-159 mg/dL High risk: ≥160 mg/dL	<2.59 mmol/L 2.59-3.34 mmol/L 3.37-4.12 mmol/L ≥4.14 mmol/L
Luteinizing hormone (LH)	Serum or plasma (heparin)	M: 1.2-7.8 mIU/mL F, Follicular phase: 1.7-15.0 mIU/mL Midcycle: 21.9-56.5 mIU/mL Luteal: 0.6-16.3 mIU/mL Postmenopausal: 14.2-52.3 mIU/mL	1.2-7.8 IU/L 1.7-15.0 IU/L 21.9-56.6 IU/L 0.6-16.3 IU/L 14.2-52.3 IU/L
Magnesium AAS	Serum Urine, 24-hr	1.6-2.6 mg/dL 6.0-10.0 mEq/day	0.66-1.07 mmol/L 3.00-5.00 mmol/day
Magnesium, free	Serum		0.45-0.60 mmol/L
Manganese	Whole blood (heparin) Serum Urine	5-15 µg/L 0.5-1.3 µg/L 0.5-9.8 µg/L	90-270 nmol/L 9-24 nmol/L 9-178 nmol/L
Mercury	Whole blood (EDTA) Urine, 24-hr	<5.0 µg/dL <20 µg/L Toxic: >150 µg/L	<0.25 µmol/L <0.1 µmol/L >0.75 µmol/L
Nickel	Whole blood Serum Urine, 24-hr	1-28 µg/L 0.14-1.0 µg/L 0.1-10 µg/day	17-476 nmol/L 2.4-17.0 nmol/L 2-170 nmol/day
Osmolality	Serum		275-295 mOsmol/kg
Oxalate	Serum	1-2.4 µg/mL Ethylene glycol poisoning: >20 µg/mL	11-27 µmol/L Ethylene glycol poisoning: >228 µmol/L
Oxygen (Po_2)	Whole blood, arterial (heparin)	83-108 mm Hg	11-14.4 kPa
Oxygen saturation	Whole blood, arterial (heparin)	94-98%	Fraction saturated: 0.94-0.98
Parathyroid hormone, intact	Serum	10-65 pg/mL	10-65 ng/L
Parathyroid hormone, (1-84)	Serum	6-40 pg/mL	6-40 ng/L
pH (37° C)	Whole blood, arterial (heparin) Serum	 2.5-4.5 mg/dL	7.35-7.45 0.81-1.45 nmol/L
Phosphate	 Urine, 24-hr	 0.4-1.3 g/day	 13-42 mmol/day
Porphobilinogen (PBG) Quantitative Qualitative	 Urine, 24-hr Urine, fresh random	 0-2.0 mg/day	0-8.8 µmol/day 0-8.8 µmol/day Negative
Potassium	Serum Plasma (heparin) Urine, 24-hr	3.5-5.1 mEq/L 3.5-4.5 mEq/L 25-125 mEq/day	3.5-5.1 mmol/L 3.5-4.5 mmol/L 25-125 mmol/day

TABLE APPENDIX-2 CLINICAL CHEMISTRY, TOXICOLOGY, AND SEROLOGY—cont'd

TEST	SPECIMEN	REFERENCE INTERVAL (CONVENTIONAL UNITS)	REFERENCE INTERVAL (INTERNATIONAL UNITS)
Proinsulin	Serum		1.1-6.9 pmol/L
Prolactin	Serum	M: 3.0-14.7 ng/mL	3.0-14.7 µg/L
Prostate-specific antigen (PSA)	Serum	F: 3.8-23.2 ng/mL M: 40-49 yr 0-2.5 ng/mL 50-59 yr 0-3.5 ng/mL 60-69 yr 0-4.5 ng/mL 70-79 yr 0-6.5 ng/mL	3.8-23.2 µg/L 0-2.5 µg/L 0-3.5 µg/L 0-4.5 µg/L 0-6.5 µg/L
Protein			
Total	Serum	6.4-8.3 g/dL	64.0-83.0 g/L
Electrophoresis (cellulose acetate)	Serum	Albumin: 3.5-5.0 g/dL α_1-Globulin: 0.1-0.3 g/dL α_2-Globulin: 0.6-1.0 g/dL β-Globulin: 0.7-1.1 g/dL γ-Globulin: 0.8-1.6 g/dL	35-50 g/L 1-3 g/L 6-10 g/L 7-11 g/L 8-16 g/L
Total	Urine, 24-hr	<100 mg/day	<0.1g/day
Total	CSF	Lumbar: 15-25 mg/dL	150-250 mg/L
Pyruvic acid	Whole blood (heparin)	0.3-0.9 mg/dL	0.03-0.10 mmol/L
Renin (normal diet)	Plasma (EDTA)	Supine: 0.2-1.6 ng/mL/hr Standing: 0.7-3.3 ng/mL/hr	0.2-1.6 µg/L/hr 0.7-3.3 µg/L/hr
Riboflavin (see Vitamin B_2)			
Sediment	Urine, fresh, random		Hyaline: occasional(0-1) casts/hpf
Casts			RBC: not seen WBC: not seen Tubular epithelial: not seen Transitional and squamous epithelial: not seen
Cells			RBC: 0-2/hpf WBC: M: 0-3/hpf F: 0-5/hpf Epithelial: few Bacteria: Unspun: no organisms/oil immersion field Spun: <20 organisms/hpf
Selenium	Serum Whole blood Urine, 24-hr	63-160 µg/L 58-234 µg/L 7-160 µg/L	0.8-2.0 µmol/L 0.74-2.97 µmol/L 0.09-2.03 µmol/L
Sodium	Serum or plasma (heparin) Urine, 24-hr	136-145 mEq/L 40-220 mEq/day	136-145 mmol/L 40-220 mmol/day
Specific gravity	Urine, random Urine, 24-hr		1.002-1.030 1.015-1.025
Testosterone, free	Serum	M: 50-210 pg/mL F: 1.0-8.5 pg/mL	174-729 pmol/L 3.5-29.5 pmol/L
Testosterone, total	Serum	M: 260-1000 ng/dL F: 15-70 ng/dL	9.0-34.7 nmol/L 0.5-2.4 nmol/L
Thiamine (see Vitamin B_1)			
Thyroglobulin (Tg)	Serum	3-42 ng/mL	3-42 µg/L
Thyroglobulin antibodies	Serum		<1 : 10
Thyroid microsomal antibodies	Serum		Nondetectable
Thyrotropin (hTSH)	Serum or plasma	0.4-4.2 µU/mL	0.4-4.2 mU/L
Thyrotropin-releasing hormone	Plasma	5-60 pg/mL	5-60 ng/L
Thyroxine, free (FT_4)	Serum	0.8-2.7 ng/dL	10.3-34.7 pmol/L
Thyroxine (T_4), total	Serum	M: 4.6-10.5 µg/dL F: 5.5-11.0 µg/dL	59-135 nmol/L 71-142 nmol/L
Thyroxine-binding globulin (TBG)	Serum	15.0-34.0 µg/mL	15.0-34.0 mg/L
Thyroxine index, free	Serum	4.2-13.0 µg/dL	54-168 nmol/L
Transcortin	Serum	M: 18.8-25.2 mg/L	323-433 nmol/L
Transferrin	Serum	F: 14.9-22.9 mg/L 200-360 mg/dL >60 yr: 160-340 mg/dL	256-393 nmol/L 2.0-3.6 g/L 1.6-3.4 g/L
Transthyretin (prealbumin)	Serum	20-40 mg/dL	200-400 mg/L
Triglycerides (TG)	Serum, after ≥ 12-hr fast	Recommended: <150 mg/dL	1.7 mmol/L
Triiodothyronine, free	Serum	210-440 pg/dL	3.2-6.8 pmol/L
Triiodothyronine, total (T_3)	Serum	70-204 ng/dL	1.08-3.14 mmol/L
Troponin-I	Serum		<10 µg/L

TABLE APPENDIX-2 CLINICAL CHEMISTRY, TOXICOLOGY, AND SEROLOGY—cont'd

TEST	SPECIMEN	REFERENCE INTERVAL (CONVENTIONAL UNITS)	REFERENCE INTERVAL (INTERNATIONAL UNITS)
Troponin-T	Serum		0-0.1 µg/L
Urea nitrogen	Serum or plasma Urine, 24-hr	6-20 mg/dL 12-20 g/day	2.1-7.1 mmol/L 0.43-0.71 mol/day
Urea nitrogen/creatinine ratio	Serum		12/1-20/1
Uric acid (uricase)	Serum Urine, 24-hr	M: 3.5-7.2 mg/dL F: 2.6-6.0 mg/dL 250-750 mg/day	0.21-0.42 mmol/L 0.15-0.35 mmol/L 1.48-4.43 mmol/day
Urinary sediment (see Sediment)			
Vanillylmandelic acid (VMA)	Urine, 24-hr	2-7 mg/day	10.1-35.4 µmol/L
Viscosity	Serum		1.10-1.22 centipoise
Vitamin A	Serum	30-80 µg/dL	1.05-2.8 µmol/L
Vitamin B_1 (thiamine)	Whole blood		90-140 nmol/L
Vitamin B_2 (riboflavin)	Serum	4-24 µg/dL	106-638 nmol/L
Vitamin B_6	Plasma (EDTA)	5-30 ng/mL	20-121 nmol/L
Vitamin B_{12}	Serum	206-678 pg/mL	151-497 pmol/L
Vitamin C	Serum	0.4-1.5 mg/dL	23-85 µmol/L
Vitamin D_3, 1,25-dihydroxy	Serum	15-60 pg/mL	36-144 pmol/L
Vitamin D_3, 25-hydroxy	Serum	10-65 ng/mL	25-162 nmol/L
Vitamin E	Serum	5.0-18.0 µg/mL	12-42 µmol/L
Zinc	Serum	80-120 µg/dL	12-18 µmol/L

TABLE APPENDIX-3 HEMATOLOGY AND COAGULATION

TEST	SPECIMEN	REFERENCE INTERVAL (CONVENTIONAL UNITS)	REFERENCE INTERVAL (INTERNATIONAL UNITS)
Activated partial thromboplastin time (APTT)	Plasma (Na citrate)		25-35 sec
Antithrombin III	Plasma (Na citrate)	85-115% of normal human plasma	0.85-1.15
Bleeding time (BT) Ivy Simplate (G-D)	Blood from skin		Normal: 2-7 min Borderline: 7-11 min 2.75-8 min
Blood volume	Whole blood (heparin)		M: 52-83 mL/kg F: 50-75 mL/kg
Bone marrow Differential count Myeloblasts Promyelocytes Myelocytes Neutrophilic Eosinophilic Basophilic Metamyelocytes Polymorphonuclear neutrophils Polymorphonuclear eosinophils Polymorphonuclear basophils Lymphocytes Plasma cells Monocytes Reticulocytes Megakaryocytes Pronormoblasts Normoblasts	Bone marrow aspirate	% (mean) 0.3-5.0 (2.0) 1.0-8.0 (5.0) 5.0-19.0 (12.0) 0.5-3.0 (1.5) 0.0-0.5 (0.3) 13.0-32.0 (22.0) 0.7-3.0 (2.0) 0.5-4.0 (2.0) 0.0-0.7 (0.2) 3.0-17.0 (10.0) 0.0-2.0 (0.4) 0.5-5.0 (2.0) 0.1-2.0 (0.2) 0.3-3.0 (1.0) 1.0-8.0 (4.0) 7.0-32.0 (18.0)	Number fraction (mean) 0.003-0.05 (0.02) 0.01-0.08 (0.05) 0.05-0.19 (0.12) 0.005-0.03 (0.015) 0.00-0.005 (0.003) 0.13-0.32 (0.22) 0.007-0.03 (0.02) 0.005-0.04 (0.02) 0.0-0.007 (0.002) 0.03-0.17 (0.10) 0.00-0.02 (0.004) 0.005-0.05 (0.02) 0.001-0.02 (0.002) 0.003-0.03 (0.01) 0.01-0.08 (0.04) 0.07-0.32 (0.18)
Clot lysis, 37° C	Whole clotted blood		47-72 hr
Clot retraction screen	Whole blood (no anticoagulant)		Retraction begins at 1 hr maximum at 24 hr
Clotting time, Lee-White, 37° C	Whole blood (no anticoagulant)		5-8 min
Differential count (see bone marrow differential count or leukocyte differential count)			
Eosinophil count	Whole blood (EDTA); capillary blood	50-400 cells/µL (mm^3)	50-400 × 10^6 cells/L
Erythrocyte count (RBC count)	Whole blood (EDTA)	×10^6 cells/µL M: 4.3-5.7 F: 3.8-5.1	×10^{12} cells/L 4.3-5.7 3.8-5.1

TABLE APPENDIX-3 HEMATOLOGY AND COAGULATION—cont'd

TEST	SPECIMEN	REFERENCE INTERVAL (CONVENTIONAL UNITS)	REFERENCE INTERVAL (INTERNATIONAL UNITS)
Erythrocyte sedimentation rate (ESR), Wintrobe			M: 0-15 mm/hr F: 0-20 mm/hr
Ferritin (see Table Appendix-2)			
Fibrin degradation products (agglutination, Thrombo-Wellco test)	Whole blood: special tube containing thrombin and proteolytic inhibitor	<10 µg/mL	<10 mg/L
Fibrinogen	Plasma (Na citrate)	200-400 mg/dL	2.00-4.00 g/L
Glucose-6-phosphate dehydrogenase (G6PD) in erythrocytes	Whole blood (ACD, EDTA, or heparin)	12.1 ± 2.09 U/g Hb (1 SD)	0.78 ± 0.13 MU/mol Hb (1 SD)
Haptoglobin (Hp)	Serum; avoid hemolysis	26-85 mg/dL	260-850 mg/L
Hematocrit (HCT, Hct)	Whole blood (EDTA)		
Calculated from MCV and RBC (electronic displacement or laser)		M: 39-49% F: 35-45%	0.39-0.49 volume fraction 0.35-0.45 volume fraction
Hemoglobin (Hb)	Whole blood (EDTA) Plasma (heparin, ACD) Urine, fresh, random	M: 13.5-17.5 g/dL F: 12.0-16.0 g/dL <3 mg/dL	2.09-2.71 mmol/L 1.86-2.48 mmol/L <0.47 mmol/L Negative
Hemoglobin electrophoresis	Whole blood (EDTA, citrate, or heparin)	HbA: >95% HbA_2: 1.5-3.5% HbF: <2%	Mass function >0.95 0.015-0.035 <0.02
Leukocyte count (WBC count)	Whole blood (EDTA) CSF	4.5-11.0 × 10^3 cells/µL (mm^3) 0.5 mononuclear cells/µL	4.5-11.0 × 10^9 cells/L 0.5 × 10^6 cells/L
Leukocyte Differential count Myelocytes Neutrophils—bands Neutrophils—segmented Lymphocytes Monocytes Eosinophils Basophils	Whole blood (EDTA)	% Cells/µL (mm^3) 0 0 3-5 150-400 54-62 3000-5800 23-33 1500-3000 3-7 285-500 1-3 50-250 0-0.75 15-50	Number fraction cells × 10^6/L 0 0 0.03-0.05 150-400 0.54-0.62 3000-5800 0.23-0.33 1500-3000 0.03-0.07 285-500 0.01-0.03 50-250 0-0.0075 15-50
Leukocyte Differential count Lymphocytes Monocytes (includes pia- arachnoid mesothelial cells) Neutrophils Histiocytes Ependymal cells Eosinophils	CSF	% 62 ± 34 36 ± 20 2 ± 5	Number fraction 0.62 ± 0.34 0.36 ± 0.20 0.02 ± 0.05 Rare Rare Rare
Mean corpuscular hemoglobin (MCH)	Whole blood (EDTA)	26-34 pg/cell	0.40-0.53 fmol/cell
Mean corpuscular hemoglobin centration (MCHC)	Whole blood (EDTA)	31-37% Hb/cell or g Hb/dL RBC	4.81-5.74 mmol Hb/L RBC
Mean corpuscular volume (MCV)	Whole blood (EDTA)		80-100 fL
Methemoglobin (MetHb)	Whole blood (EDTA, heparin, or ACD)	0.06-0.24 g/dL	9.3-37.2 mmol/L
Plasma volume	Plasma (heparin)	M: 25-43 mL/kg F: 28-45 mL/kg	0.025-0.043 L/kg 0.028-0.045 L/kg
Platelet count (thrombocyte count)	Whole blood (EDTA)	150-450 × 10^3/µL (mm^3)	150-450 × 10^9/L
Prothrombin time, one-stage	Plasma (Na citrate)	Reference values will vary with the type of thromboplastin	In general: 11-16 sec
International normalized ratio (INR)		Relevant only in patients on warfarin $INR = \left[\frac{\text{Patient PT}}{\text{Normal mean PT}}\right]^{ISI}$ ISI = International Sensitivity Index of thromboplastin	INR 2.0-3.0 (for most indications)
Red cell volume	Whole blood (heparin)	M: 20-36 mL/kg F: 19-31 mL/kg	M: 0.020-0.036 L/kg F: 0.019-0.031 L/kg
Reticulocyte count	Whole blood (EDTA, heparin, or oxalate)	0.5-1.5% of erythrocytes	0.005-0.015 (number fraction)
Sulfhemoglobin	Whole blood (EDTA, heparin)	≤1.0% of total Hb	<0.010 of total Hb (mass fraction)
Thrombin time	Whole blood (Na citrate)		Time of control ±25 when control is <22 sec

TABLE APPENDIX-4 DRUGS: THERAPEUTIC AND TOXIC

DRUG	SPECIMEN	REFERENCE INTERVAL (CONVENTIONAL UNITS)		REFERENCE INTERVAL (INTERNATIONAL UNITS)
Acetaminophen	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	10-30 µg/mL >200 µg/mL	66-199 µmol/L >1324 µmol/L
Amikacin	Serum or plasma (EDTA)	Therapeutic: Peak Trough (severe infection) Toxic: Peak Trough	 25-35 µg/mL 4-8 µg/mL >40 µg/mL >10 µg/mL	 43-60 µmol/L 6.8-13.7 µmol/L >68 µmol/L >17 µmol/L
ε-Aminocaproic acid	Serum or plasma (heparin or EDTA); trough	Therapeutic:	100-400 µg/mL	0.76-3.05 mmol/L
Amitriptyline + nortriptyline	Serum or plasma (heparin or EDTA); trough (>12 hr after dose)	Therapeutic: Toxic:	80-200 ng/mL >500 ng/mL	289-722 nmol/L >1805 nmol/L
Amobarbital	Serum	Therapeutic: Toxic:	1-5 µg/mL >10 µg/mL	4-22 µmol/L >44 µmol/L
Amphetamine	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	20-30 ng/mL >200 ng/mL	148-222 nmol/L >1480 nmol/L
Bromide	Serum	Toxic:	>1250 µg/mL	>15.6 µmol/L
Caffeine	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	8-20 µg/mL >20 µg/mL	41-103 µmol/L >103 µmol/L
Carbamazepine	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	4-12 µg/mL >15 µg/mL	17-51 µmol/L >63 µmol/L
Carbenicillin	Serum or plasma	Therapeutic: Toxic:	Dependent on inhibitory concentration of specific organism >250 µg/mL	Same >660 µmol/L
Chloramphenicol	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	10-25 µg/L >25 µg/mL	31-77 µmol/L >77 µmol/L
Chlordiazepoxide	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	700-1000 ng/mL >5000 ng/mL	2.34-3.34 µmol/L >16.7 µmol/L
Chlorpromazine	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	30-300 ng/mL >750 ng/mL	94-942 nmol/L >2355 nmol/L
Cimetidine	Serum or plasma (heparin or EDTA); trough	Therapeutic:	0.5-1.2 µg/mL	2-5 µmol/L
Ciprofloxacin	Serum	Therapeutic (oral): Toxic:	0.5-1.5 µg/mL >5.0 µg/mL	2-5 µmol/L >15 µmol/mL
Clonazepam	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	20-70 ng/mL >80 ng/mL	63-222 nmol/L >254 nmol/L
Clonidine	Serum or plasma (heparin or EDTA)	Therapeutic:	1.0-2.0 ng/mL	4.4-8.7 nmol/L
Cocaine	Serum or plasma (heparin or EDTA) on ice	Therapeutic: Toxic:	100-500 ng/mL >1000 ng/mL	330-1650 nmol/L >3300 nmol/L
Codeine	Serum	Therapeutic: Toxic:	10-100 ng/mL >1100 ng/mL	33-334 nmol/L >3340 nmol/L
Cyclosporine A	Serum (12-hr after dose)	Therapeutic: Toxic:	50-350 ng/mL >350 ng/mL	42-291 nmol/L >291 nmol/L
Desipramine	Serum or plasma (heparin or EDTA); trough (>12 hr after dose)	Therapeutic: Toxic:	100-300 ng/mL >400 ng/mL	375-1125 nmol/L >1500 nmol/L
Diazepam	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	100-1000 ng/mL >5000 ng/mL	0.35-3.51 µmol/L >17.55 µmol/L
Digitoxin	Serum or plasma (heparin or EDTA) >6 hr after dose	Therapeutic: Toxic:	10-30 ng/mL >45 ng/mL	13-39 nmol/L >59 nmol/L
Digoxin	Serum or plasma (heparin or EDTA) trough (>12 hr after dose)	Therapeutic: Toxic:	0.5-2.0 ng/mL >1.5 ng/mL	0.6-3.0 nmol/L >2 nmol/L
Diphenylhydantoin (see Phenytoin)				
Disopyramide	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	2.8-7.5 µg/mL >5 µg/mL	8-22 µmol/L 15 µmol/L
Doxepin	Serum or plasma (heparin or EDTA); trough (>12 hr after dose)	Therapeutic: Toxic:	50-150 ng/mL >500 ng/mL	179-537 nmol/L >1790 nmol/L
Ephedrine	Serum	Therapeutic: Toxic:	0.05-0.10 µg/mL >2 µg/mL	0.30-0.61 µmol/L >12.1 µmol/L
Ethchlorvynol	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	2-8 µg/mL >20 µg/mL	14-55 µmol/L >138 µmol/L
Ethosuximide	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	40-100 µg/mL >150 µg/mL	283-708 µmol/L >1062 µmol/L

TABLE APPENDIX-4 DRUGS: THERAPEUTIC AND TOXIC—cont'd

DRUG	SPECIMEN	REFERENCE INTERVAL (CONVENTIONAL UNITS)		REFERENCE INTERVAL (INTERNATIONAL UNITS)
Everolimus	Whole blood	Therapeutic: Toxic:	3-15 ng/mL >15 ng/mL	3-16 nmol/L >16 nmol/L
Fenoprofen	Plasma (EDTA)	Therapeutic:	20-65 µg/mL	82-268 µmol/L
Flecainide	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	0.2-1.0 µg/mL >1.0 µg/mL	0.5-2.4 µmol/L >2.4 µmol/L
Fluoxetine	Serum or plasma	Therapeutic: Toxic:	120-300 ng/mL >1000 ng/mL	388-969 nmol/L >3230 nmol/L
Flurazepam	Serum or plasma (EDTA)	Therapeutic: Toxic:	Not well defined >0.2 µg/mL	 >0.5 µmol/L
Gabapentin	Serum or plasma	Therapeutic: Toxic:	2-20 µg/mL >12 µg/mL	12-117 µmol/L >70 µmol/L
Gentamicin	Serum or plasma (EDTA)	Therapeutic: Peak (severe infection) Trough (severe infection) Toxic: Peak Trough	 8-10 µg/mL <4 µg/mL >10 µg/mL >2 µg/mL	 16.7-20.9 µmol/L <8 µmol/L >21 µmol/L >4 µmol/L
Glutethimide	Serum	Therapeutic: Toxic:	2-6 µg/mL >5 µg/mL	9-28 µmol/L >23 µmol/L
Haloperidol	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	5-17 ng/mL >42 ng/mL	13-45 nmol/L >112 nmol/L
Ibuprofen	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	10-50 µg/mL >200 µg/mL	49-243 µmol/L >970 µmol/L
Imipramine + desipramine	Serum or plasma (heparin or EDTA); trough (>12 hr after dose)	Therapeutic: Toxic:	150-300 ng/mL >400 ng/mL	536-1071 nmol/L >1428 nmol/L
Isoniazid	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	1-7 µg/mL >20 µg/mL	7-51 µmol/L >146 µmol/L
Itraconazole	Serum or plasma	Therapeutic:	>1.5 µg/mL	>2 µmol/L
Kanamycin	Serum or plasma (EDTA)	Therapeutic: Peak Trough (severe infection) Toxic: Peak Trough	 25-35 µg/mL 4-8 µg/mL >35 µg/mL >10 µg/mL	 52-72 µmol/L 8-16 µmol/L >72 µmol/L >21 µmol/L
Lidocaine	Serum or plasma (heparin or EDTA); >45 min following bolus dose	Therapeutic: Toxic:	1.5-5 µg/mL >6 µg/mL	6-21 µmol/L >26 µmol/L
Lithium	Serum or plasma (heparin or EDTA); >12 hr after last dose	Therapeutic: Toxic:	0.5-1.2 mEq/L >2 mEq/L	0.5-1 nmol/L >2 µmol/L
Lorazepam	Serum or plasma (heparin or EDTA)	Therapeutic:	50-240 ng/mL	156-746 nmol/L
Meperidine	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	70-500 ng/mL >1 µg/mL	283-2020 nmol/L >4040 nmol/L
Meprobamate	Serum	Therapeutic: Toxic:	6-12 µg/mL >60 µg/mL	28-55 µmol/L >275 µmol/L
Methadone	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	100-400 ng/mL >2000 ng/mL	0.32-1.29 µmol/L >6.46 µmol/L
Methamphetamine	Serum	Therapeutic: Toxic:	0.01-0.05 µg/mL >0.5 µg/mL	0.07-0.34 µmol/L >3.35 µmol/L
Methaqualone	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	2-3 µg/mL >10 µg/mL	8-12 µmol/L >40 µmol/L
Methsuximide (*N*-desmethyl methsuximide)	Serum	Therapeutic: Toxic:	10-40 µg/mL >40 µg/mL	53-212 µmol/L >212 µmol/L
Methyldopa	Plasma (EDTA)	Therapeutic: Toxic:	1-5 µg/mL >7 µg/mL	4.7-23.7 µmol/L >33 µmol/L
Methyprylon	Serum	Therapeutic: Toxic:	8-10 µg/mL >50 µg/mL	43-55 µmol/L >273 µmol/L
Morphine	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	10-80 ng/mL >200 ng/mL	35-280 nmol/L >700 nmol/L
Netilmicin	Serum or plasma (EDTA)	Therapeutic: Peak (severe infection) Trough (severe infection) Toxic: Peak Trough	 8-10 µg/mL <4 µg/mL >10 µg/mL >2 µg/mL	 17-21 µmol/L <8 µmol/L >21 µmol/L >4 µmol/L

TABLE APPENDIX-4 DRUGS: THERAPEUTIC AND TOXIC—cont'd

DRUG	SPECIMEN	REFERENCE INTERVAL (CONVENTIONAL UNITS)		REFERENCE INTERVAL (INTERNATIONAL UNITS)
Nortriptyline	Serum or plasma (heparin or EDTA); trough (>12 hr after dose)	Therapeutic: Toxic:	70-170 ng/mL >500 ng/mL	266-646 nmol/L >1900 nmol/L
Oxazepam	Serum or plasma (heparin or EDTA)	Therapeutic:	0.2-1.4 µg/mL	0.70-4.9 µmol/L
Oxycodone	Serum	Therapeutic: Toxic:	10-100 ng/mL >200 ng/mL	32-317 nmol/L >634 nmol/L
Paroxetine	Serum or plasma	Therapeutic:	70-120 ng/mL	213-365 nmol/L
Pentazocine	Serum or plasma (EDTA)	Therapeutic: Toxic:	0.05-0.2 µg/mL >1 µg/mL	0.2-0.7 µmol/L >3.5 µmol/L
Pentobarbital	Serum or plasma (heparin or EDTA); trough	Therapeutic, hypnotic: Therapeutic, coma: Toxic:	1-5 µg/mL 20-50 µg/mL >10 µg/mL	4-22 µmol/L 88-221 µmol/L >44 µmol/L
Phenacetin	Plasma (EDTA)	Therapeutic: Toxic:	1-30 µg/mL 50-250 µg/mL	6-167 µmol/L 279-1395 µmol/L
Phenobarbital	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic: Slowness, ataxia, nystagmus Coma with reflexes Coma without reflexes	10-40 µg/mL 35-80 µg/mL 65-117 µg/mL >100 µg/mL	43-173 µmol/L 151-345 µmol/L 280-504 µmol/L >430 µmol/L
Phensuximide + norphensuximide	Serum or plasma (heparin or EDTA)	Therapeutic:	40-60 µg/mL	212-317 µmol/L
Phenylbutazone	Plasma (EDTA)	Therapeutic: Toxic:	50-100 µg/mL >100 µg/mL	162-324 µmol/L >324 µmol/mL
Phenytoin	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	10-20 µg/mL >20 µg/mL	40-79 µmol/L >79 µmol/L
Primidone + phenobarbital	Serum or plasma (heparin or EDTA) trough	Therapeutic: Toxic:	5-10 µg/mL >15 µg/mL	23-46 µmol/L >69 µmol/L
Procainamide	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	4-10 µg/mL >12 µg/mL	17-42 µmol/L >51 µmol/L
Propoxyphene	Plasma (EDTA)	Therapeutic: Toxic:	0.1-0.4 µg/mL >0.5 µg/mL	0.3-1.2 µmol/L >1.5 µmol/L
Propranolol	Serum or plasma (heparin or EDTA); trough	Therapeutic:	20-100 ng/mL	77-386 nmol/L
Protriptyline	Serum or plasma (heparin or EDTA); trough (>12 hr after dose)	Therapeutic: Toxic:	70-260 ng/mL >500 ng/mL	266-988 nmol/L >1900 nmol/L
Quinidine	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	2-5 µg/mL >6 µg/mL	6-15 µmol/L >19 µmol/L
Risperidone + 9-hydroxyrisperidone	Serum or plasma	Therapeutic:	20-60 ng/mL	49-146 nmol/L
Salicylates	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	150-300 µg/mL >300 µg/mL	1086-2172 µmol/L >2172 µmol/L
Secobarbital	Serum	Therapeutic: Toxic:	1-2 µg/mL >5 µg/mL	4.2-8.4 µmol/L >21.0 µmol/L
Sirolimus	Whole blood	Therapeutic: Toxic:	4-20 ng/mL >20 ng/mL	4-22 nmol/L >22 nmol/L
Sulfonamides as sulfanilamide	Serum or plasma	Therapeutic: Toxic:	5-15 mg/mL >20 mg/mL	29-87 mmol/L >116 mmol/L
Theophylline	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	8-20 µg/mL >20 µg/mL	44-111 µmol/L >111 µmol/L
Thiopental	Serum or plasma (heparin or EDTA); trough	Hypnotic: Coma: Anesthesia: Toxic:	1-5 µg/mL 30-100 µg/mL 7-130 µg/mL >10 µg/mL	4.1-20.7 µmol/L 124-413 µmol/L 29-536 µmol/L >41 µmol/L
Thioridazine	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	0.2-2.0 µg/mL >10 µg/mL	0.5-5 µmol/L >27 µmol/L
Tobramycin	Serum or plasma (heparin or EDTA)	Therapeutic: Peak (severe infection) Trough (severe infection) Toxic: Peak Trough	 8-10 µg/mL <4 µg/mL >10 µg/mL >2 µg/mL	 17-21 µmol/L <9 µmol/L >21 µmol/L >4 µmol/L

TABLE APPENDIX-4 DRUGS: THERAPEUTIC AND TOXIC—cont'd

DRUG	SPECIMEN	REFERENCE INTERVAL (CONVENTIONAL UNITS)		REFERENCE INTERVAL (INTERNATIONAL UNITS)
Tocainide	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	6-15 µg/mL >15 µg/mL	31-78 µmol/L >78 µmol/L
Tolbutamide	Serum	Therapeutic: Toxic:	90-240 µg/mL >640 µg/mL	333-888 µmol/L >2368 µmol/L
Valproic acid	Serum or plasma (heparin or EDTA); trough	Therapeutic: Toxic:	50-100 µg/mL >100 µg/mL	347-693 µmol/L >693 µmol/L
Vancomycin	Serum or plasma (heparin or EDTA); trough	Therapeutic: Peak: Trough: Toxic: (not well established)	 20-40 µg/mL >10 µg/mL >80 µg/mL	 14-28 µmol/L >7 µmol/L >55µmol/L
Warfarin	Serum or plasma (heparin or EDTA)	Therapeutic: Toxic:	1-10 µg/mL >10 µg/mL	3-32 µmol/L >32 µmol/L
Zidovudine	Serum or plasma	Therapeutic:	>0.2 µg/mL	>0.8 µmol/L

GUIDE TO THE APPROACH TO COMMON SYMPTOMS, SIGNS, AND LABORATORY ABNORMALITIES

	CHAPTER	SPECIFIC TABLES OR FIGURES
SYMPTOMS		
Constitutional		
Fever	280	Tables 280-1 to 280-8
Fatigue	274	E-Table 274-1
Poor appetite	132	Table 132-1
Weight loss	132, 219	Figure 132-4; Tables 132-4, 219-1, 219-2
Obesity	220	Figure 220-1
Snoring, sleep disturbances	100, 405	Table 405-6
Head, Eyes, Ears, Nose, Throat		
Headache	398	Tables 398-1, 398-2
Visual loss, transient	423, 424	Tables 423-2, 424-1
Ear pain	426	Table 426-3
Hearing loss	428	Figure 428-1
Ringing in ears (tinnitus)	428	Figure 428-2
Vertigo	428	Figure 428-3
Nasal congestion, rhinitis, or sneezing	251, 426	Figure 251-1; Table 251-2
Loss of smell or taste	427	Table 427-1
Dry mouth	425	Table 425-7
Sore throat	429	Figure 429-2; Table 429-1
Hoarseness	429	
Cardiopulmonary		
Chest pain	51, 137	Tables 51-2, 137-5, 137-6
Bronchitis	96	
Shortness of breath	51, 83	Figure 83-3
Palpitations	51, 62	Figure 62-1; Tables 51-4, 62-5
Dizziness	51, 62, 428	Figure 62-1; Table 428-1
Syncope	62	Figure 62-1; Tables 62-1, 62-2, 62-4
Cardiac arrest	63	Figures 63-2, 63-3
Cough	83	Figure 83-1; Tables 83-2, 83-3
Hemoptysis	83	Tables 83-6, 83-7
Gastrointestinal		
Nausea and vomiting	132	Figure 132-5; Table 132-5
Dysphagia, odynophagia	132, 138	Table 132-1
Hematemesis	135, 153	Figure 135-3; Table 135-1
Heartburn/dyspepsia	132, 137, 138, 139	Figures 132-6, 138-2; Tables 137-3, 137-4, 139-1
Abdominal pain		
Acute	132, 142	Figures 132-1, 132-2; Tables 132-2, 132-3, 142-1
Chronic	132, 137	Figure 132-3; Tables 132-2, 137-1
Diarrhea	137, 140	Figures 137-1, 140-1 to 140-4
Melena, blood in stool	135	Figures 135-3, 135-4, 135-6; Table 135-4
Constipation	136, 137	Figures 136-3, 136-5, 137-1; Table 136-2
Fecal incontinence	145	Figure 145-5
Anal pain	145	
Genitourinary		
Dysuria	284, 285	Tables 284-3, 284-5, 285-2
Frequency	284	Table 284-3
Incontinence	26	Tables 26-1 to 26-3
Urinary obstruction	123	Tables 123-1 to 123-3
Renal colic	126	Figure 126-1
Vaginal discharge	285	
Menstrual irregularities	236	Figure 236-3; Tables 236-3, 236-4
Female infertility	236	Table 236-5
Hot flushes	240	Table 240-1
Erectile dysfunction	234	Figure 234-10
Male infertility	234	Figures 234-8, 234-9; Table 234-7
Scrotal mass	200	Figure 200-1
Genital ulcers or warts	285	Table 285-1

GUIDE TO THE APPROACH TO COMMON SYMPTOMS, SIGNS, AND LABORATORY ABNORMALITIES—cont'd

	CHAPTER	SPECIFIC TABLES OR FIGURES
Musculoskeletal		
Neck or back pain	400	Figures 400-4, 400-5, 400-6; Tables 400-3 to 400-5
Painful joints	256	Figure 256-1; Tables 256-1, 256-3
Extremities		
Swollen feet, ankles, or legs		
Bilateral	51	Figure 51-8
Unilateral	81	Figure 81-2; Table 81-2
Claudication	79	Table 79-3
Acute limb ischemia	79	Figure 79-5; Table 79-1
Neurologic		
Weakness	396, 420, 421, 422	Tables 396-1, 420-2, 421-2, 421-4
Sensory loss	396, 420	Figure 420-1; Tables 420-1, 420-3 to 420-5
Memory loss	402	Figures 402-1, 402-2; Tables 402-1 to 402-6
Abnormal gait	396	Table 396-2
Seizures	403	Tables 403-1 to 403-6
Integumentary		
Abnormal bleeding	171	Table 171-1
Rash	436, 441	Figure 436-1; Tables 436-1 to 436-6, 441-5
Hives	252, 440	Figure 252-2; Tables 252-1, 440-1, 440-2
Abnormal pigmentation	441	Table 441-2
Alopecia and hirsutism	442	Tables 442-1, 442-3
Nail disorders	442	Table 442-4
SIGNS		
Vital Signs		
Fever	280, 281	Figure 281-1; Tables 280-1 to 280-8, 281-2
Hypothermia	8, 109	Table 109-4
Tachycardia/bradycardia	8, 62, 64, 65	Figures 62-2, 62-3; Tables 64-4, 65-2
Hypertension	67	Table 67-5
Hypotension/shock	8, 106	Figures 106-3, 108-1; Tables 106-1, 107-1, 107-2
Altered respiration	8, 86, 104	Tables 86-1, 86-2, 104-2
Head, Eyes, Ears, Nose, Throat		
Eye pain	423	Table 423-3
Red eye	423	Tables 423-4, 423-6
Dilated pupil	424	Figure 424-4
Nystagmus	424	Table 424-5
Papilledema	424	Table 424-2
Strabismus	424	Figure 424-6
Jaundice	147	Figure 147-2; Tables 147-1 to 147-3
Otitis	426	Table 426-3
Sinusitis	251, 426	Tables 251-3, 426-1, 426-2
Oral ulcers and discolorations	425	Tables 425-1 to 425-4
Salivary gland enlargement	425	Table 425-6
Neck		
Neck mass	190	Figure 190-3
Lymphadenopathy	168	Tables 168-1 to 168-6
Thyroid nodule	226	Figure 226-4
Thyromegaly/goiter	226	Figures 226-1, 226-3
Breast		
Breast mass	198	
Lungs		
Wheezes	83	Table 83-4
Cardiac		
Heart murmur or extra sounds	51	Figure 51-6; Tables 51-7, 51-8
Jugular venous distention	51	Table 51-6
Carotid pulse abnormalities	51	Figure 51-5

GUIDE TO THE APPROACH TO COMMON SYMPTOMS, SIGNS, AND LABORATORY ABNORMALITIES—cont'd

	CHAPTER	SPECIFIC TABLES OR FIGURES
Abdomen		
Hepatomegaly	146	Figure 146-5
Splenomegaly	168	Tables 168-7, 168-9
Acute abdomen	142, 143	Figure 143-1; Table 142-1
Abdominal swelling/ascites	142, 153	Table 153-3
Rectal bleeding/positive stool	135, 193	Figures 135-3, 135-4, 135-6; Table 135-4
Hemorrhoids	145	Table 145-1
Musculoskeletal/Extremities		
Arthritis	256	Figure 256-1
Edema	51	Figure 51-8
Cyanosis	51	
Clubbing	51	Figure 51-10
Neurologic		
Delirium	28	Figure 28-1; Tables 28-1, 28-2
Psychiatric disturbances	397	Tables 397-1 to 397-4, 397-6 to 397-8, 397-10, 397-11, 397-13, 397-14
Coma	404	Tables 404-1 to 404-4
Stroke	407, 408	Figure 407-1; Tables 407-2, 407-3, 407-5, 407-6, 408-5, 408-6
Movement disorders	409, 410	Tables 409-4, 410-1 to 410-8
Neuropathy	420	Figure 420-1; Tables 420-1 to 420-5, E-Table 420-1
Skin and Nails		
Suspicious mole	203	Table 203-1
Nail diseases	442	Table 442-4
COMMON LABORATORY ABNORMALITIES		
Hematology/Urinalysis		
Anemia	158	Tables 158-2 to 158-6
Polycythemia	166	Figure 166-2; Table 166-4
Leukocytosis	167	Figure 167-4; Table 167-1
Lymphocytosis	167	Table 167-3
Monocytosis	167	Table 167-2
Eosinophilia	170	Figure 170-2; Table 170-1
Neutropenia	167	Figure 167-7; Tables 167-4 and 167-5
With fever	281	Figure 281-1
Thrombocytosis	166	Figure 166-6; Table 166-6
Thrombocytopenia	172	Figure 172-1; Tables 172-1, 172-3
Prolonged PT or PTT	171	Figure 171-3
Urinalysis	114, 120	Tables 114-2, 120-6
Chemistries		
Abnormal liver enzymes	147	Figures 147-2 to 147-4
Elevated BUN/creatinine		
Acute	120	Figure 120-1; Tables 120-1 to 120-5
Chronic	130	Table 130-1
Hyperglycemia	229	Tables 229-1, 229-2
Hypoglycemia	230	Tables 230-1, 230-2
Electrolyte abnormalities	116, 117	Figure 116-4; Tables 116-6, 116-7, 117-2, 117-3
Acid-base disturbances	118	Figures 118-1, 118-2; Tables 118-1 to 118-6
Hypercalcemia	245	Figure 245-3; Tables 245-2 to 245-4
Hypocalcemia	245	Figure 245-4; Table 245-6
Hypo- and hyperphosphatemia	119	Tables 119-2, 119-3
Magnesium deficiency	119	Table 119-1
Elevated P_{CO_2}	86	Figure 86-2
Chest Radiograph/ECG		
Solitary pulmonary nodule	191	Figure 191-2
Pleural effusion	99	Tables 99-4 to 99-6
ECG abnormalities	54	Tables 54-2 to 54-5

BUN = blood urea nitrogen; ECG = electrocardiogram; PT = prothrombin time; PTT = partial thromboplastin time.